Symmetries and Gradient Flows in the Deep Linear Network

by

Tejas Kotwal

Five-Year Integrated M.Sc., Indian Institute of Technology Bombay, Mumbai, 2019

ScM, Brown University, Providence, 2020

Thesis

Submitted in partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

in the Division of Applied Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

May 2026

This dissertation by Tejas Kotwal is accepted in its present form
by the Division of Applied Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date________________ ______________________________

Govind Menon, Ph.D., Advisor

Recommended to the Graduate Council

Date________________ ______________________________

Jérôme Darbon, Ph.D., Reader

Date________________ ______________________________

Björn Sandstede, Ph.D., Reader

Approved by the Graduate Council

Date________________ ______________________________

David P. Lindstrom, Dean of the Graduate School

To my grandparents

Vasanthi and Nooyi Narayana

Yallamma and Joteppa Kotwal

Acknowledgments

Above all, I thank Govind Menon. From him I learned to care about taste, simplicity of ideas, honesty in science and scholarship, and clear mathematical communication. He also taught me a certain fearlessness about hard problems, while giving me a great deal of freedom in how I approached them. Mathematics has always felt to me like an artistic and spiritual endeavour, and I am grateful that there was room for that during my PhD. My relationship with mathematics was tested several times during these years. I am glad to have come through them with my love for the subject intact, and Govind's guidance is a large part of the reason why. His breadth, originality, and conviction continue to inspire me, and I feel extraordinarily fortunate to have learned from him.

I thank Jérôme Darbon and Björn Sandstede for agreeing to read this thesis and serve on my committee. I have also learned a great deal from Paul Dupuis, Stuart Geman, Basilis Gidas, Javier Gómez-Serrano, Kavita Ramanan, Richard Schwartz, and others at Brown whose courses and conversations shaped my mathematical education. I am grateful to Tony Bloch for conversations about metrics on adjoint orbits, and to Xiangxiong Zhang for hosting me at Purdue during a week of work on the graph embedding problem.

Jake Mundo, Zsolt Veraszto, and Tianmin Yu were wonderful companions throughout graduate school, and many of my first glimpses of the broader landscape around Govind's research came through conversations with them.

Long before Brown, Gopala Krishna Srinivasan shaped much of my taste and attitude toward mathematics. He has been an undergraduate advisor, a mentor, and a friend, and I owe him more than I can summarize here. I am equally grateful to Sudheer and Gayathri Murthy, and to the FSnT teachers, for making the pursuit of physics and mathematics feel joyful long before I had any clear idea what a life in either subject might look like.

A great deal of this thesis depended on people whose work is often invisible in the final document. I thank the APMA staff, especially Candy, Jean, Rosanna, Rachel, and others who helped me navigate the administrative life of the department with warmth and generosity. I am also deeply grateful to my counselor Madhuri Dugyala, to CAPS, and to Dean Maria Suarez for their help and accommodation during difficult stretches of graduate school. They made it possible for me to do this work under far better conditions than I otherwise could have.

I have been sustained throughout these years by friends. Raunaq and Roshail have been present at every important decision in my adult life, including the decision to move into mathematics, and their faith in my instincts has carried me more often than they know. Aakriti, Basuhi, Deepanshu, Mishal, and others have given me the rare comfort of friendships that survived distance, time, and repeated reinvention.

At Brown, Gopal made my PhD years much richer, and I will miss our long conversations about philosophy, AI, and the many shows we watched together. Akash and Rahul gave me some of my happiest memories from these years, through daily acts of kindness as well as trips and adventures I will remember for a long time. Faraz, Nihal, and Sayan added many conversations and moments of camaraderie that made these years lighter and more memorable. I thank Saket for opening a window onto a different research culture through reinforcement learning. The friends I met online during the pandemic, especially through Discord and far too much Civilization VI, showed me that friendship can arrive from unexpected places.

I am also grateful to my landlord Alex and his mother Raisa for their warmth and kindness, which made Providence feel more like home. During my years at Brown, I was also fortunate to study tabla with Nitin Mitta, a kind person and an excellent teacher.

Another striking feature of these years was the emergence of large language models. I had the unusual experience of watching these models improve dramatically while learning to use them as a way to work through ideas. Interacting with them often became a kind of conversational exploration: a way to test questions, challenge intuitions, and accelerate the back-and-forth through which mathematical understanding develops. At the same time, the responsibility for judgment and scrutiny still rests entirely with the human mathematician. It seems certain that tools of this kind will shape how mathematics and science are practiced in the years ahead. I am grateful to Andrew Chen for giving me access to one of these models at a moment when I was beginning to seriously experiment with this way of working.

My sibling Veer has been a constant source of inspiration. Their seriousness about the arts, and attention to questions larger than science, have widened my world in lasting ways. I hope that one day we find a way to create something together. To my parents, Archana and Suresh Kotwal, I owe more than I can sensibly put into words. They encouraged me to follow my instincts, gave me room to make mistakes and change course, and supported me quietly through every stage of this journey. Everything that matters in this thesis rests on that foundation.

I would be remiss not to thank the dogs who have accompanied so much of my life: Simba, Donna, DP, and now Coco. I miss Simba and Donna dearly. Watching them grow from puppies into old companions taught me more than I understood at the time about affection, and the quiet fact that nothing living stays young forever.

Finally, I dedicate this thesis to my grandparents. They came from very different walks of life, and it moves me to think that two generations later their grandson would one day defend a thesis in applied mathematics. Whatever I have managed to do here is inseparable from the values they passed down: discipline, generosity, curiosity, and care.

Contents

List of Tables

List of Illustrations

Part I

Foundations

Chapter 1

Introduction

1.1 Deep learning and overparametrization

Deep learning is function approximation by composition. An architecture specifies a sequence of maps, depending on trainable parameters θ, whose composition defines a model $f(\cdot;\theta)$. Given training data $\{(x_k, y_k)\}_{k=1}^n$, one fits θ by minimizing an empirical loss. In continuous time, the resulting training dynamics is modeled by a gradient flow.

We study the non-uniqueness that arises when a single map is realized as a composition of many layers. In the linear model, this amounts to factoring a matrix X as a product of layer matrices. Such a factorization is highly non-unique: many different collections of matrices yield the same product. In this thesis, overparametrization refers to this non-uniqueness, rather than to the relation between the dimension of the parameter space and sample size that is common in the machine learning literature. Depth thus matters not only because it changes the parametrization, but also because it enlarges the parameter space and alters the geometry of optimization. We study the deep linear network because it isolates this effect in the simplest possible setting [BH89; SMG14].

Fix a depth $N \geq 2$, a width $d \geq 1$, and $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A point in parameter space and its end-to-end matrix are related by the product map

$$\phi : \mathrm{M}_d(\mathbb{F})^N \to \mathrm{M}_d(\mathbb{F}), \qquad \mathbf{W} = (W_N, \dots, W_1) \mapsto X = W_N \cdots W_1. \tag{1.1}$$

Unless stated otherwise, we assume that each layer matrix is invertible,

$$W_k \in \mathrm{GL}(d, \mathbb{F}), \qquad 1 \le k \le N, \tag{1.2}$$

so in particular $X = \phi(\mathbf{W}) \in \mathrm{GL}(d, \mathbb{F})$. If $E : \mathrm{M}_d(\mathbb{F}) \to \mathbb{R}$ is a smooth loss on the space of represented matrices, the lifted objective is

$$L(\mathbf{W}) = E(\phi(\mathbf{W})), \tag{1.3}$$

and its Euclidean gradient flow on parameter space is

$$\dot{\mathbf{W}} = -\operatorname{grad} L(\mathbf{W}). \tag{1.4}$$

The represented map $x \mapsto Xx$ is linear, but the lifted objective L is nonconvex in the parameters because X is their product. This makes the DLN a useful model of deep learning: it retains composition and overparametrization while isolating them from the additional complications of nonlinear activations. The central problem is to understand how the geometry of the fibres of ϕ controls the dynamics of the end-to-end matrix X. For a fixed X, this raises three questions: how large is the family of factorizations realizing X, how does gradient flow move through that family, and what geometry does depth induce on the space of end-to-end maps?

1.2 Symmetry, balancedness, and entropy

The product map ϕ distinguishes the space of factorizations $\mathrm{M}_d(\mathbb{F})^N$ from the space of represented matrices $\mathrm{M}_d(\mathbb{F})$. For fixed X, the fibre

$$\mathcal{F}_X = \phi^{-1}(X) \tag{1.5}$$

consists of all depth-N factorizations of X.

This nonuniqueness is organized by a natural group action. If $A = (A_{N-1}, \ldots, A_1) \in \mathrm{GL}(d, \mathbb{F})^{N-1}$, define

$$A \cdot \mathbf{W} = \Big(W_N A_{N-1}^{-1},\ A_{N-1} W_{N-1} A_{N-2}^{-1},\ \ldots,\ A_1 W_1\Big). \tag{1.6}$$

Then $\phi(A \cdot \mathbf{W}) = \phi(\mathbf{W})$, so the action (1.6) preserves the product map (1.1). In particular, each fibre (1.5) is invariant under this $\mathrm{GL}(d, \mathbb{F})^{N-1}$-action. On the full-rank locus, this action organizes each fibre. In fact, each fibre over $X \in \mathrm{GL}(d, \mathbb{F})$ is a single $\mathrm{GL}(d, \mathbb{F})^{N-1}$-orbit.

A factorization is called balanced when

$$W_{k+1}^* W_{k+1} = W_k W_k^*, \qquad 1 \le k \le N-1. \tag{1.7}$$

The balanced factorizations form the balanced manifold $\mathcal{M}$. Balancedness is central because it identifies the part of parameter space on which the geometry simplifies. It is preserved by the Euclidean gradient flow of $E \circ \phi$ [ACH18], and on each fibre it also characterizes the preferred factorizations selected by natural regularization principles [LM25].

On $\mathcal{M}$, the map ϕ becomes a Riemannian submersion. Equivalently, the Euclidean metric on parameter space descends through ϕ to a metric g_N on the space of end-to-end matrices, and along balanced trajectories the represented matrix satisfies

$$\dot{X} = -\operatorname{grad}_{g_N} E(X). \tag{1.8}$$

Thus depth enters the end-to-end dynamics not through the loss E, but through the metric g_N [Bah+22; MY26b].

For fixed invertible X, the set of balanced factorizations

$$\mathcal{O}_X = \mathcal{F}_X \cap \mathcal{M}$$

is a compact K-orbit, where $K = \mathrm{O}(d)^{N-1}$ or $K = \mathrm{U}(d)^{N-1}$ according to whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Its log-volume

$$S_N(X) = \log \operatorname{vol}(\mathcal{O}_X) \tag{1.9}$$

taken with respect to the Riemannian volume induced by the Euclidean metric on parameter space, defines a Boltzmann entropy on the space of represented matrices. It measures the size of the family of balanced factorizations that represent the same matrix X [MY26b]. This entropy also appears in a stochastic description of training. In the Riemannian Langevin equation (RLE), the projected drift contains the gradient of this entropy, and the loss and entropy combine to form the free energy

$$F_\beta(X) = E(X) - \beta^{-1} S_N(X), \tag{1.10}$$

where $\beta > 0$ is an inverse temperature parameter [MY26b; MY26a].

A striking feature of the DLN is that it connects to several areas of mathematics, including symplectic geometry, random matrix theory, stochastic processes, and gauge theory. These themes recur throughout the thesis and remain tightly linked because they arise from the same overparametrized matrix factorization.

1.3 Main results

The main results of the thesis describe how depth shapes the geometry of the deep linear network. Starting from the nonuniqueness of matrix factorizations, we show that symmetry organizes the fibres, that balancedness selects distinguished factorizations, and that the resulting quotient geometry determines both the dynamics and the associated entropy on the space of end-to-end matrices. We then pass to an infinite-depth limit, where this finite-dimensional geometry gives rise to a continuum model with gauge symmetry and a renormalized quotient geometry. The aim throughout is to identify the geometry on parameter space that descends to the space of end-to-end matrices and governs the dynamics.

In Chapter 2, we fix notation and record the structural facts used throughout the thesis, including conserved quantities, balancedness, the metric g_N, the entropy S_N, and the fibrewise regularizers.

In Chapter 3, we analyze the real DLN at finite depth. For spectral energies, the g_N-gradient flow of the free energy closes on the chamber of singular values. We prove that the entropy S_N extends real-analytically through singular-value collisions. We also analyze its concavity properties with respect to both the Euclidean and Riemannian geometries of the chamber. Under the stated convexity hypotheses on the spectral energy, the free energy has a unique minimizer on the chamber, and this minimizer lies on the diagonal. For Schatten energies, the diagonal flow reduces to an explicitly solvable scalar equation.

In Chapters 4 and 5, we study the complex DLN. We identify the balanced manifold as the zero level set of the moment map for the unitary symmetry. We prove that, for each invertible matrix X, the set of balanced factorizations of X is a single compact unitary orbit. We then compute the quotient metric on $\mathrm{GL}(d, \mathbb{C})$ and derive explicit formulas for the Kähler potential, the entropy, and the Ricci form.

In Chapters 6 and 7, we study nonzero level sets of the moment map. For fixed X, we describe the quotient of the fibre by the compact symmetry using the Gram matrices of partial products, and hence obtain coordinates on a product of positive-definite cones. In these coordinates, the balanced point is the geodesic from I to X^*X with respect to the trace metric. We also compute the mean-curvature drift arising from Brownian motion on general levels sets of the moment map and its projection to the space of end-to-end matrices.

In Chapters 8–10, we pass from a finite sequence of layers to a continuum field on an interval. We rewrite the finite-depth fibre geometry in a form adapted to renormalization, identify the Sobolev spaces and gauge symmetry of the continuum model, and express balancedness through a continuum moment map. We also show that the fibrewise regularizing dynamics take the form of heat-type flows along the depth direction, dissipating toward balancedness. In Chapter 9, we give a rigorous formulation of the continuum model using Sobolev spaces. In Chapter 10 we derive the limiting Hermitian metric on the quotient $\mathrm{GL}(d, \mathbb{C})$ together with the renormalized Kähler potential, entropy, and Ricci formula. These describe the geometric quantities that survive after renormalization in the infinite-depth setting. We do not prove here a full convergence theory from the finite-depth model to the continuum limit.

1.4 On the use of large language models

Mathematics does not advance through finished theorem statements alone. It proceeds through examples, failed calculations, conversations, and the gradual clarification of what a proof ought to explain. Lakatos [Lak63], Thurston [Thu06], and Gowers [Gow23], in different ways, all stress that mathematical work contains much more than the final polished argument: it includes search, revision, and judgment long before the proof is written in its final form. The emergence of large language models changes this exploratory layer of mathematical work, but it does not erase the distinction between exploration and proof.

I used ChatGPT during both the research and writing of this thesis.[1] Most often, I used it to test an example, probe an edge case, rephrase a definition or paragraph, or check whether a local step was stated clearly. These systems were treated as fallible tools, not as mathematical authorities, and no mathematical claim in this thesis is included on the authority of model output alone. Their responses were often useful, but they were also frequently incomplete, overconfident, or simply wrong. At times, they also helped clarify or reorganize the presentation of an argument, though the underlying mathematics and its validation remained my responsibility.

Another limitation is what is often called *sycophancy*: the tendency of these systems to accommodate the direction of a conversation rather than challenge it. This is most pronounced in subjective or open-ended settings, where there may be no clear notion of correctness and the model can elaborate on a weak premise without resistance. In mathematics, this tendency is checked more sharply by explicit standards of correctness. A model may present an argument fluently, but agreement or fluency is not evidence of

[1] For more demanding mathematical exchanges and line-by-line editorial work, I generally used ChatGPT Pro with GPT-5.4 Pro, GPT-5.3, or GPT-5.2. Lighter ChatGPT models were sometimes used for lower-stakes drafting, formatting, or exploratory prompts. Where recoverable, model/version metadata will be recorded in the companion repository.

truth. Its claims can be tested against axioms, definitions, computations, and standard results. In practice, I therefore treated agreement or fluency as insufficient, and instead asked for derivations, checked edge cases, and verified each step independently. The model's role was not to confirm correctness, but to surface candidate arguments that could then be accepted or rejected on mathematical grounds.

Originality here does not mean that every intermediate sentence or heuristic arose in isolation. That would misdescribe not only AI-assisted work but much of ordinary mathematics, which is already shaped by seminars, correspondence, textbooks, examples, collaborators, and inherited language. The relevant standard is one of judgment and responsibility. The choice of problems, the judgment of what mattered, the recognition of which arguments were sound, the decision to discard false lines of thought, and the synthesis of the material into a coherent mathematical contribution remained mine. In this sense, the originality of the thesis lies not in solitude, but in intellectual direction and mathematical responsibility.

This view of originality depends on how one understands the development of mathematical ideas. Recent work of Asvin G offers a broader account of that process. In [G25], mathematics is understood as organizing and explaining patterns in *computational traces*, meaning the concrete outputs of calculation, symbolic manipulation, proof search, algorithms, and related formal procedures, by means of axiomatic frameworks. Problem-solving is described as a loop of belief revision, and a distinction is drawn between *implicit* concepts, which sharpen search within an existing language, and *explicit* concepts, which change the language of possible moves [G26]. Whether or not one accepts this framework in full, it captures something important about my own experience with these tools. If mathematical work develops through search, revision, and the interpretation of such traces, then originality need not mean that every intermediate step arose in isolation. It lies instead in directing, interpreting, and validating the process. Present-day LLMs can assist with local search, reformulation, comparison, and heuristic testing, but they are much less reliable at the deeper task of determining which concepts, theorems, or explanations are the right ones. These tools can accelerate parts of the back-and-forth through which understanding develops, but they do not by themselves determine what deserves to enter the literature.

Whenever a conversation with a model suggested a claim, derivation, or reformulation that seemed useful, I checked it separately. Sometimes I rederived the point from first principles, sometimes I traced it back to standard sources, and sometimes I discarded it. Model output was therefore part of the research process, not a substitute for proof or for standard mathematical sources.

To make that role concrete, I summarize three representative conversations below. Additional annotated transcripts are provided in a companion repository.[2]

Example 1: debugging moment-map geometry. In one exchange, I used the model to check a draft discussion of ambient Kähler geometry, moment maps, and imbalanced level sets. The main mathematical structure was already in place: the group action, the description of the normal space via J-rotated orbit directions, and the claim that nonzero imbalance produces a symplectic tilt of the orbit relative to the tangent space of the level set. The model was helpful in surfacing places where conventions might be muddled, especially signs, pairings, and equivariance. Some suggestions survived, some had to be corrected, and some were discarded after I checked the calculations directly. What survived into the thesis was a clearer and independently verified presentation.

Example 2: provenance and overclaiming in the balanced fibre discussion. A second exchange concerned claims about balanced fibres, moment maps, and the

[2]A companion repository containing annotated transcripts, curation notes, and model/version metadata accompanies this thesis.

isotropic or Lagrangian character of the relevant unitary orbit. Here the main issue was provenance. I needed to separate what was explicit in the literature, what belonged to standard symplectic background, and what required a fresh argument in the deep linear network setting. The model helped mainly by forcing that separation: vague statements had to be made precise, and overstatements had to be justified or removed. In particular, claims about Lagrangian structure were kept only after separate checking and reformulation. The result was a cleaner account of scope and sources.

Example 3: finding a visual language for group orbits and moment-map level sets. A third exchange was mainly about exposition. I was trying to find a clear visual and conceptual way to explain how group orbits sit relative to level sets of the moment map, especially the difference between the balanced set and generic nonzero level sets. The mathematical point itself was already settled. The balanced set is preserved by the group action, whereas a generic nonzero level set is not invariant and is typically cut across by group orbits. What I lacked was a clean presentation of that distinction. The model helped by offering rough pictures, captions, and local descriptions that clarified the separation between global equivariance and local tangent/normal geometry. Some of those pictures and analogies were misleading and had to be rejected or redrawn. What remained was a clearer exposition of the same argument.

The repository is selective rather than exhaustive and is limited to thesis-related material. The aim is not to archive every prompt, but to show representative modes of use. Each public transcript is accompanied by a short note stating the mathematical question, what I had before the exchange, what the model contributed, where it failed or overreached, and what remained after independent checking. Exhaustiveness here would obscure rather than clarify the role of the tool.

I expect some version of this workflow to become increasingly normal in mathematics. If Thurston is right that proof is only one part of mathematical progress [Thu06], and if Gowers is right that research depends on intermediate judgments, search strategies, and changing levels of confidence [Gow23], then tools that accelerate these pre-proof stages will change the practice of the subject. Recent work on AI and mathematical practice points in a similar direction [Hen25; Pan25]. The central question is how such systems can be used well without blurring the line between assistance and mathematical judgment. My own view is that mathematics in the coming years will be carried out in a richer ecology of thought, involving human conversation, computation, proof, and increasingly sophisticated machine assistance. That makes human judgment all the more important, not less. Someone must still decide what is true, what is interesting, what is explained, and what is worth preserving. For this thesis, that responsibility is mine.

Chapter 2

The Deep Linear Network

In this chapter, we fix notation for the DLN and record the structural facts used throughout the thesis: the product map, the group action on fibres, the conserved quantities, the balanced manifold, the quotient metric, the entropy, and the fibrewise regularizer flows.

2.1 The model

We write

$$\mathbf{W} = (W_N, \ldots, W_1), \qquad W_k \in \mathrm{M}_d(\mathbb{F}).$$

The associated end-to-end map is defined by

$$\phi : \mathrm{M}_d(\mathbb{F})^N \to \mathrm{M}_d(\mathbb{F}), \qquad \phi(\mathbf{W}) = X = W_N \cdots W_1.$$

Let

$$E : \mathrm{M}_d(\mathbb{F}) \to \mathbb{R}$$

be a smooth loss on the space of end-to-end matrices, and let

$$L(\mathbf{W}) = E(\phi(\mathbf{W}))$$

be the lifted loss. The Euclidean gradient flow on parameter space is

$$\dot{W}_k = -(W_N \cdots W_{k+1})^* \mathrm{d}E(X)(W_{k-1} \cdots W_1)^*, \qquad 1 \le k \le N, \tag{2.1}$$

with the convention that empty products are the identity. Unless stated otherwise, we work on the full-rank locus, so $W_k \in \mathrm{GL}(d, \mathbb{F})$ and $X \in \mathrm{GL}(d, \mathbb{F})$.

For $X \in \mathrm{M}_d(\mathbb{F})$, the fibre over X is

$$\mathcal{F}_X = \phi^{-1}(X).$$

For $A = (A_{N-1}, \ldots, A_1) \in \mathrm{GL}(d, \mathbb{F})^{N-1}$, define

$$A \cdot (W_N, \ldots, W_1) = (W_N A_{N-1}^{-1},\ A_{N-1} W_{N-1} A_{N-2}^{-1},\ \ldots,\ A_1 W_1). \tag{2.2}$$

The group $\mathrm{GL}(d, \mathbb{F})^{N-1}$ acts by changing bases at the intermediate layers. These changes cancel in the full product, so $\phi(A \cdot \mathbf{W}) = \phi(\mathbf{W})$ and every fibre is invariant under this action.

Define the Hermitian matrices

$$G_k = W_k W_k^* - W_{k+1}^* W_{k+1}, \qquad 1 \le k \le N-1.$$

Thus G_k compares the outgoing Gram matrix of layer k with the incoming Gram matrix of layer $k+1$. We collect these matrices into the tuple

$$\mathbf{G} = (G_{N-1}, \ldots, G_1),$$

ordered to match the coordinates of the group $\mathrm{GL}(d, \mathbb{F})^{N-1}$.

For a prescribed tuple $\mathbf{G} = (G_{N-1}, \ldots, G_1)$, let

$$\mathcal{M}_{\mathbf{G}} = \left\{ \mathbf{W} \in \mathrm{M}_d(\mathbb{F})^N : W_k W_k^* - W_{k+1}^* W_{k+1} = G_k \text{ for } 1 \leq k \leq N-1 \right\}.$$

The case $\mathbf{G} = \mathbf{0}$ is the balanced variety. Its full-rank stratum is the balanced manifold $\mathcal{M}$.

Choose a singular value decomposition $X = Q_N \Sigma Q_0^*, \qquad \Lambda = \Sigma^{1/N}$. The tuple $C_X = (Q_N \Lambda, \Lambda, \ldots, \Lambda Q_0^*)$ lies in $\mathcal{F}_X$ and is balanced. Thus C_X is the balanced representative determined by the chosen singular value decomposition.

If $K = \mathrm{O}(d)^{N-1} \quad \text{or} \quad \mathrm{U}(d)^{N-1}$ according to whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, then for $X \in \mathrm{GL}(d, \mathbb{F})$ we set

$$\mathcal{O}_X = \mathcal{F}_X \cap \mathcal{M}.$$

Proposition 2.1 ([MY26b; LM25])**.** *Let $X \in \mathrm{GL}(d, \mathbb{F})$, and choose a singular value decomposition defining the balanced representative C_X. Then*

$$\mathcal{F}_X = \mathrm{GL}(d, \mathbb{F})^{N-1} \cdot C_X, \qquad \mathcal{O}_X = K \cdot C_X.$$

In particular, $\mathcal{O}_X$ is compact.

Thus every full-rank factorization of X lies in the $\mathrm{GL}(d, \mathbb{F})^{N-1}$-orbit of C_X, and the balanced factorizations are exactly the K-orbit of C_X.

2.2 Invariant sets

The tuple $\mathbf{G}$ is the conserved quantity of the Euclidean gradient flow.

Theorem 2.2 ([ACH18])**.** *Let $\mathbf{W}(t) = (W_N(t), \ldots, W_1(t))$ solve the parameter flow*

$$\dot{W}_k = -(W_N \cdots W_{k+1})^* \mathrm{d}E(X)(W_{k-1} \cdots W_1)^*, \qquad X = W_N \cdots W_1, \qquad 1 \le k \le N, \tag{2.3}$$

where empty products are the identity. For

$$G_k(t) = W_k(t)W_k(t)^* - W_{k+1}(t)^* W_{k+1}(t), \qquad 1 \le k \le N-1,$$

the tuple

$$\mathbf{G}(t) = \Big(G_{N-1}(t), \ldots, G_1(t)\Big)$$

is constant in t. Equivalently, each level set $\mathcal{M}_{\mathbf{G}}$ is invariant under the flow (2.3)*: if $\mathbf{W}(0) \in \mathcal{M}_{\mathbf{G}}$, then $\mathbf{W}(t) \in \mathcal{M}_{\mathbf{G}}$ for as long as the solution exists.*

Proposition 2.3 ([ACH18])**.** *Let $\mathbf{W}(t)$ solve the flow* (2.3)*. Define the partial products*

$$P_k = W_N \cdots W_k, \qquad Q_k = W_k \cdots W_1, \qquad 1 \le k \le N,$$

and set $P_{N+1} = Q_0 = I$. Then the end-to-end matrix $X(t) = \phi(\mathbf{W}(t))$ satisfies

$$\dot{X} = -\sum_{k=1}^{N} (P_{k+1}P_{k+1}^*)\, \mathrm{d}E(X)\, (Q_{k-1}^* Q_{k-1}). \tag{2.4}$$

The flow (2.4) is not closed in X in general, because the Gram matrices of the partial products depend on the factorization. On the balanced manifold, however, these matrices are determined by X alone.

Lemma 2.4. *Let $\mathbf{W} \in \mathcal{M}$ and $\phi(\mathbf{W}) = X$. Then*

$$P_{k+1}P_{k+1}^* = (XX^*)^{\frac{N-k}{N}}, \qquad Q_{k-1}^* Q_{k-1} = (X^*X)^{\frac{k-1}{N}}, \qquad 1 \le k \le N.$$

For $X \in \mathrm{GL}(d, \mathbb{F})$, define the linear operator

$$\mathcal{A}_{N,X}(Z) = \sum_{k=1}^{N} (XX^*)^{\frac{N-k}{N}} \, Z \, (X^*X)^{\frac{k-1}{N}}, \qquad Z \in \mathrm{M}_d(\mathbb{F}). \tag{2.5}$$

The operator $\mathcal{A}_{N,X}$ is positive definite with respect to the Frobenius pairing, and therefore defines a Riemannian metric on $\mathrm{GL}(d, \mathbb{F})$. In the real case this metric is

$$g_N(X)(Z_1, Z_2) = \mathrm{tr}\Big(Z_1^T \mathcal{A}_{N,X}^{-1}(Z_2)\Big), \qquad Z_1, Z_2 \in \mathrm{M}_d(\mathbb{R}), \tag{2.6}$$

while in the complex case we take the real part of the Hermitian pairing:

$$g_N(X)(Z_1, Z_2) = \Re\, \mathrm{tr}\Big(Z_1^* \mathcal{A}_{N,X}^{-1}(Z_2)\Big), \qquad Z_1, Z_2 \in \mathrm{M}_d(\mathbb{C}). \tag{2.7}$$

Theorem 2.5 ([ACH18; Bah+22]). *Assume* $\mathbf{W}(0) \in \mathcal{M}$*, and let* $X(t) = \phi(\mathbf{W}(t))$*. Then*

$$\dot{X} = -\sum_{k=1}^{N} (XX^*)^{\frac{N-k}{N}} \, \mathrm{d}E(X) \, (X^*X)^{\frac{k-1}{N}}. \tag{2.8}$$

Equivalently,

$$\dot{X} = -\, \mathrm{grad}_{g_N} E(X). \tag{2.9}$$

Equations (2.8) and (2.9) give, respectively, the closed evolution of the end-to-end matrix and the Riemannian gradient-flow formulation of the same dynamics. The loss does not change with the depth, but the metric does.

2.3 Entropy and Riemannian submersion

The quotient geometry of $\mathcal{M}$ yields both the metric g_N on the space of end-to-end matrices and the volume of the compact orbit $\mathcal{O}_X$. For the real model, the volume of the orbit is given by the following explicit formula.

Write

$$\operatorname{van}(\Sigma) = \prod_{1\le i<j\le d} (\sigma_i - \sigma_j).$$

Theorem 2.6 ([MY26b]). *Let $X \in \mathrm{GL}(d,\mathbb{R})$ have singular value decomposition*

$$X = Q_N \Sigma Q_0^T, \qquad \Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_d),$$

and assume that the singular values are distinct. Then

$$S_N(X) = \log \operatorname{vol}(\mathcal{O}_X) = (N-1)\log c_d + \frac{1}{2}\log \frac{\operatorname{van}(\Sigma^2)}{\operatorname{van}(\Sigma^{2/N})},$$

equivalently,

$$S_N(X) = (N-1)\log c_d + \frac{1}{2}\sum_{1\le i<j\le d} \log\left(\frac{\sigma_i^2 - \sigma_j^2}{\sigma_i^{2/N} - \sigma_j^{2/N}} \right),$$

where c_d is the volume of $\mathrm{O}(d)$ with Haar measure.

Thus S_N depends only on the singular values of X. It appears as the entropic term in the drift of the Riemannian Langevin equation. The corresponding free energy is

$$F_\beta(X) = E(X) - \beta^{-1} S_N(X),$$

which appears in the stochastic description of training [MY26b].

Let ι be the metric induced on the real balanced manifold $\mathcal{M}$ by the Euclidean metric on $\mathrm{M}_d(\mathbb{R})^N$.

Theorem 2.7 ([MY26b]). *The map*

$$\phi : (\mathcal{M}, \iota) \to (\mathrm{GL}(d,\mathbb{R}), g_N)$$

is a Riemannian submersion.

Thus g_N is the quotient metric obtained from $(\mathcal{M}, \iota)$ by collapsing the directions tangent to $\mathcal{O}_X$. In Chapter 5, we prove the complex analogues of Theorems 2.6 and 2.7.

2.4 Fibrewise regularizers and flows

We now record the functionals on a fixed fibre $\mathcal{F}_X$ that single out the balanced set $\mathcal{O}_X$ together with their associated gradient flows.

In the complex model, these functionals are closely related to the geometry of the moment map. The group $K = \mathrm{U}(d)^{N-1}$ acts on $\mathrm{M}_d(\mathbb{C})^N$ by (2.2). With the standard Kähler structure, the trace pairing on $\mathfrak{u}(d)^{N-1}$, and our sign convention, the moment map is

$$\mu(\mathbf{W}) = \mathbf{G}(\mathbf{W}) = (G_{N-1}, \ldots, G_1).$$

We prove this identification in Chapter 4. On the full-rank locus, the balanced manifold is the zero level set of μ:

$$\mathcal{M} = \mu^{-1}(\mathbf{0}) \cap \mathrm{GL}(d, \mathbb{C})^N.$$

Let ι_X denote the metric on $\mathcal{F}_X$ induced by the Euclidean metric on $\mathrm{M}_d(\mathbb{F})^N$. For $\mathbf{W} = (W_N, \ldots, W_1) \in \mathrm{M}_d(\mathbb{F})^N$, define the ridge functional

$$\mathcal{R}(\mathbf{W}) := \frac{1}{2} \sum_{k=1}^{N} \mathrm{tr}(W_k^* W_k).$$

For $1 < p < \infty$, define the Schatten-p quantity

$$\|\mathbf{W}\|_p := \sum_{k=1}^{N} \|W_k\|_p,$$

where $\|\cdot\|_p$ is the Schatten p-norm. In particular, the case $p = 2$ corresponds to the Frobenius norm on each layer, so that $\|\mathbf{W}\|_2$ is the sum of Frobenius norms.

Also define the functional

$$\mathcal{E}(\mathbf{W}) := \frac{1}{2}\|\mu(\mathbf{W})\|^2 = \frac{1}{2}\sum_{k=1}^{N-1} \operatorname{tr}(G_k^* G_k) = \frac{1}{2}\sum_{k=1}^{N-1} \operatorname{tr}(G_k^2),$$

since each G_k is Hermitian. This is the Kirwan–Ness functional, the squared norm of the moment map. On a fixed fibre, its zero set is the balanced set $\mathcal{O}_X$.

The associated fibrewise gradient flows are

$$\dot{\mathbf{W}} = -\operatorname{grad}_{\iota_X} \mathcal{R}(\mathbf{W}), \qquad \mathbf{W}(t) \in \mathcal{F}_X, \tag{2.10}$$

$$\dot{\mathbf{W}} = -\operatorname{grad}_{\iota_X} \mathcal{E}(\mathbf{W}), \qquad \mathbf{W}(t) \in \mathcal{F}_X. \tag{2.11}$$

The first is the flow generated by the ridge functional on a fixed fibre, and the second is the Kirwan–Ness flow on a fixed fibre.

Theorem 2.8 ([LM25])**.** *Assume X has full rank. Then*

$$\underset{\mathbf{W}\in\mathcal{F}_X}{\arg\min}\, \|\mathbf{W}\|_2 = \mathcal{F}_X \cap \mathcal{M} = \mathcal{O}_X.$$

Moreover, for every $1 < p < \infty$,

$$\underset{\mathbf{W}\in\mathcal{F}_X}{\arg\min}\, \|\mathbf{W}\|_p = \mathcal{F}_X \cap \mathcal{M} = \mathcal{O}_X.$$

Balancedness is therefore characterized by a minimum principle on each fibre. In the proof for the complex DLN, this is exactly the content of the Kempf–Ness theorem [KN79]. The flow generated by the ridge functional does not merely converge to balancedness. It induces an exact evolution for the moment tuple $\mathbf{G}$, which decays exponentially.

Theorem 2.9 ([LM25]). *Assume X has full rank, and let $\mathbf{W}(t) \in \mathcal{F}_X$ solve the flow generated by the ridge functional on the fixed fibre $\mathcal{F}_X$,*

$$\dot{\mathbf{W}} = -\operatorname{grad}_{\iota_X} \mathcal{R}(\mathbf{W}), \qquad \mathcal{R}(\mathbf{W}) = \frac{1}{2}\sum_{k=1}^{N} \operatorname{tr}(W_k^* W_k).$$

If $\mathbf{G}(t) := \mathbf{G}(\mathbf{W}(t))$, then

$$\mathbf{G}(t) = e^{-2t}\mathbf{G}(0), \qquad t \geq 0.$$

In Chapters 8 and 9, we reinterpret the flows (2.10) and (2.11) as heat-type flows on the fibre and develop their continuum analogues.

Bibliographical notes

The conservation of moments, the end-to-end dynamics, and the metric formulation of Euclidean gradient flow on the balanced manifold go back to [ACH18] and [Bah+22]. For an expository account of the geometry of deep linear networks, see [Men25]. The entropy formula for the group orbit and the Riemannian submersion are from [MY26b]. The fibrewise minimum principle and the explicit regularizer flow are from [LM25].

For complementary results on the global algebraic geometry of the fibres of the multiplication map, studied via quiver representations, see [LR24].

For the role of $\|\mu\|^2$ in geometric invariant theory, see [MFK94]. In particular, the associated gradient flow was used by Ness to study the stratification of the null cone [Nes84] and by Kirwan in the Morse theory of symplectic quotients [Kir84]. For surveys and related results on this correspondence, see [Woo10; Ler05]. For the infinite-dimensional analogue arising from Yang–Mills theory, see [AB83; Tra17].

Chapter 3

The Real DLN and Entropic Regularization

On the balanced manifold, Euclidean gradient flow in parameter space induces the Riemannian gradient flow $\dot{X} = -\operatorname{grad}_{g_N} E(X)$ on $\mathrm{GL}(d, \mathbb{R})$.

We study the free energy

$$F_\beta(X) = E(X) - \frac{1}{\beta} S_N(X), \tag{3.1}$$

where $S_N(X) = \log \operatorname{vol}(\mathcal{O}_X)$ is the entropy associated with the balanced factorizations of X.

The key simplification is to restrict to spectral energies. In that case, both the metric g_N and the entropy S_N descend to the singular values of X, and the dynamics close on the singular-value chamber. This reduces the problem for matrices to a system of interacting scalar variables.

3.1 Main results

Throughout this chapter, unless explicitly stated otherwise, assume $N \geq 2$, $d \geq 2$, and $\beta > 0$.

The gradient flow on the space of end-to-end matrices is $\dot{X} = -\operatorname{grad}_{g_N} E(X)$, with $X \in \mathrm{GL}(d, \mathbb{R})$. In Chapter 2, we showed that this is the flow induced by Euclidean gradient flow in parameter space from balanced initial data. In this chapter we replace E by the free energy (3.1).

For $X \in \mathrm{GL}(d, \mathbb{R})$ with singular values $\sigma = (\sigma_1, \ldots, \sigma_d)$, the entropy is [MY26b, Theorem 4]

$$S_N(X) \;=\; (N-1)\log c_d + \frac{1}{2}\sum_{1\leq i<j\leq d} \log\left(\frac{\sigma_i^2 - \sigma_j^2}{\sigma_i^{2/N} - \sigma_j^{2/N}}\right), \tag{3.2}$$

where c_d is the volume of $\mathrm{O}(d)$. We study the gradient flow

$$\dot{X} = -\operatorname{grad}_{g_N} F_\beta(X), \qquad X \in \mathrm{GL}(d, \mathbb{R}). \tag{3.3}$$

Equivalently,

$$\dot{X} \;=\; -\sum_{k=1}^{N} (XX^T)^{\frac{N-k}{N}}\, \mathrm{d}F_\beta(X)\, (X^T X)^{\frac{k-1}{N}}, \tag{3.4}$$

where $\mathrm{d}F_\beta$ denotes the differential of F_β.

Let $\mathcal{S}_d = \{\sigma \in \mathbb{R}^d :\; \sigma_1 \geq \cdots \geq \sigma_d > 0\}$ be the chamber of singular values in decreasing order, with interior $\mathcal{S}_d^\circ = \{\sigma \in \mathbb{R}^d :\; \sigma_1 > \cdots > \sigma_d > 0\}$.

Entropy. In Section 3.2, we show that g_N pushes forward to a metric $g_{N,\sigma}$ on $\mathcal{S}_d^\circ$, so the chamber of singular values is the natural reduced space of coordinates for the real problem. The first group of results concerns the entropy on this chamber. Although the formula (3.2) is written as a ratio of Vandermonde factors, its singularities at

$\sigma_i = \sigma_j$ are removable.

Theorem 3.1. *The entropy S_N is real-analytic on $\mathcal{S}_d$.*

Let $(\mathcal{S}_d, \iota)$ denote the chamber equipped with the Euclidean metric inherited from $\mathbb{R}^d$.

Theorem 3.2. *The entropy S_N is strictly concave on $(\mathcal{S}_d, \iota)$, except in the case $(N, d) = (2, 2)$ where its Hessian has rank one.*

Theorem 3.3. *The entropy S_N is not concave on $(\mathcal{S}_d, g_{N,\sigma})$: at every point with $\sigma_1 = \cdots = \sigma_d$ the Hessian is indefinite.*

Free-energy dynamics. We now restrict to spectral energies.

Definition 3.4. We say that $E : \mathrm{M}_d(\mathbb{R}) \to \mathbb{R}$ is a *spectral energy* if

$$E(X) = E(\sigma(X)), \qquad E(\sigma) = g\Big(\sum_{i=1}^{d} f(\sigma_i)\Big), \tag{3.5}$$

where $g : \mathbb{R} \to \mathbb{R}$ is nondecreasing and $f : (0, \infty) \to \mathbb{R}$ is convex.

For the equilibrium and linearization results we impose the standing hypotheses

$$g \in C^2(\mathbb{R}), \qquad g' > 0, \qquad g'' \geq 0, \qquad f \in C^2((0,\infty)), \qquad f' \geq 0, \qquad f'' > 0. \tag{3.6}$$

These assumptions guarantee the monotonicity and convexity used in Lemmas 3.26, 3.27, and 3.28, and in the proof of Theorem 3.6. The Schatten energies with $p > 1$ satisfy (3.6). The diagonal reduction in Section 3.5 remains valid for $p = 1$ as well.

Remark 3.5. For energies without symmetry the singular values do not evolve autonomously. A basic example is matrix completion: for $\Omega \subset \{1, \dots, d\}^2$ and observed entries a_{ij},

$$E(X) = \frac{1}{2} \sum_{(i,j)\in\Omega} (X_{ij} - a_{ij})^2.$$

This loss depends on the entries of X, not just its singular values. Its set of minimizers is typically an affine space containing matrices of different rank. Understanding how the entropy S_N interacts with this geometry remains open.

For spectral energies, F_β depends only on σ: $F_\beta(\sigma) = E(\sigma) - \beta^{-1} S_N(\sigma)$. In Section 3.2, we then show that the reduced gradient flow is

$$\dot{\sigma}_i = -N\, \sigma_i^{2-2/N}\, \partial_{\sigma_i} F_\beta(\sigma), \qquad i = 1, \ldots, d. \tag{3.7}$$

The right-hand side extends continuously to $\mathcal{S}_d$, and we use the same formula there whenever no differentiability issue is involved.

Under (3.6), the free energy has a unique critical point, and it lies on the diagonal.

Theorem 3.6. *There exists a unique equilibrium $\sigma \in \mathcal{S}_d$ of F_β, and it has the form $\sigma_1 = \cdots = \sigma_d = \sigma_\star > 0$, where $\sigma_\star$ is the unique solution of*

$$g'\Big(d\, f(\sigma_\star)\Big)\, f'(\sigma_\star) \;=\; \beta^{-1}\, \frac{d-1}{2\sigma_\star} \Big(1 - \frac{1}{N}\Big). \tag{3.8}$$

Moreover, this equilibrium is the unique minimizer of F_β on $\mathcal{S}_d$.

Remark 3.7. The entropy S_N has a Vandermonde-type structure, and the corresponding equilibrium equations resemble the Coulomb–gas conditions from random matrix theory. However, the interaction differs in a fundamental way. The logarithmic repulsion present in random matrix ensembles is replaced by an expression that extends real-analytically across collisions of singular values. In particular, there is no singular-value repulsion, and coincident singular values are not penalized by the entropy. As a consequence, the unique equilibrium lies on the diagonal $\sigma_1 = \cdots = \sigma_d$.

Write $\vec{\sigma}_\star = (\sigma_\star, \dots, \sigma_\star)$.

The second main conclusion is that the linearization splits according to the decomposition

$$\mathbb{R}^d = \operatorname{span}\{\mathbf{1}\} \oplus \operatorname{span}\{\mathbf{1}\}^\perp.$$

At the equilibrium $\vec{\sigma}_\star$, the flow on the chamber

$$\dot{\sigma}_i = -N\,\sigma_i^{2-2/N}\,\partial_{\sigma_i} F_\beta(\sigma), \qquad i = 1, \dots, d,$$

linearizes to

$$\dot{\eta} = J\eta, \qquad J = -N\,\sigma_\star^{2-2/N}\,\nabla_\sigma^2 F_\beta(\vec{\sigma}_\star).$$

Theorem 3.8. *Let $H_E = \nabla_\sigma^2 E(\vec{\sigma}_\star)$ and $H_S = \nabla_\sigma^2 S_N(\vec{\sigma}_\star)$ denote the Euclidean Hessians at the stationary point. Write $\theta_{\mathbf{1}}(E)$ and $\theta_\perp(E)$ for the eigenvalues of H_E on* span$\{\mathbf{1}\}$ *and its orthogonal complement, and define $\theta_{\mathbf{1}}(S_N)$ and $\theta_\perp(S_N)$ similarly. Then the linearized system $\dot{\eta} = J\eta$ diagonalizes in the splitting*

$$\mathbb{R}^d = \operatorname{span}\{\mathbf{1}\} \oplus \operatorname{span}\{\mathbf{1}\}^\perp$$

with eigenvalues

$$\begin{aligned} \rho_{\mathbf{1}} &= -N\,\sigma_\star^{2-2/N}\Big(\theta_{\mathbf{1}}(E) - \beta^{-1}\theta_{\mathbf{1}}(S_N)\Big), \\ \rho_\perp &= -N\,\sigma_\star^{2-2/N}\Big(\theta_\perp(E) - \beta^{-1}\theta_\perp(S_N)\Big), \qquad (\textit{multiplicity } d-1). \end{aligned}$$

Remark 3.9 (Infinite depth). The entropy S_N and the rescaled metrics Ng_N and $Ng_{N,\sigma}$ have well-defined limits as $N \to \infty$. The renormalized entropy is

$$S_\infty(X) \;=\; \frac{1}{2}\sum_{1\le i<j\le d} \log\!\left(\frac{\sigma_i^2 - \sigma_j^2}{\log\sigma_i^2 - \log\sigma_j^2}\right). \tag{3.9}$$

Likewise, Ng_N converges to a limiting metric g_∞ [CMV23], and the corresponding metric on $\mathcal{S}_d$ also has a limit. The statements in this chapter have direct analogues after replacing S_N, g_N, and $g_{N,\sigma}$ by these limits.

Since $F_\beta(X)$ depends only on singular values, the minimizers in $\mathrm{GL}(d,\mathbb{R})$ form the orbit $\mathcal{O}_\star := \{\sigma_\star Q : Q \in \mathrm{O}(d)\}$.

Corollary 3.10. *The set of minimizers of F_β on $\mathrm{GL}(d,\mathbb{R})$ is $\mathcal{O}_\star$.*

If $X_\star = \sigma_\star Q \in \mathcal{O}_\star$, the kernel of the linearization is the tangent space to this orbit, $T_{X_\star}\mathcal{O}_\star = \{X_\star A : A^T = -A\}$.

Corollary 3.11. *The linearization of the flow (3.3) at $X_\star$ diagonalizes in the splitting*

$$\mathrm{M}_d(\mathbb{R}) = T_{X_\star}\mathcal{O}_\star \oplus \mathrm{span}\{X_\star\} \oplus \{QS : S^T = S,\ \mathrm{tr}\, S = 0\}, \tag{3.10}$$

with eigenvalues

$$\begin{aligned} 0 \quad & \textit{on } T_{X_\star}\mathcal{O}_\star, \\ \rho_{\mathbf{1}} \quad & \textit{on } \mathrm{span}\{X_\star\}, \\ \rho_\perp \quad & \textit{on } \{QS : S^T = S,\ \mathrm{tr}\, S = 0\}, \end{aligned}$$

where

$$\begin{aligned} \rho_{\mathbf{1}} &= -N\,\sigma_\star^{2-2/N}\left(\theta_{\mathbf{1}}(E) - \beta^{-1}\theta_{\mathbf{1}}(S_N)\right), \\ \rho_\perp &= -N\,\sigma_\star^{2-2/N}\left(\theta_\perp(E) - \beta^{-1}\theta_\perp(S_N)\right). \end{aligned}$$

The Schatten energies provide a concrete class of examples. For

$$E_p(X) = \frac{1}{p}\sum_{i=1}^{d}\sigma_i^p, \qquad 1 \le p < \infty, \tag{3.11}$$

Theorem 3.6 gives

$$\sigma_\star = \left(\frac{d-1}{2\beta}\left(1-\frac{1}{N}\right)\right)^{1/p}. \tag{3.12}$$

Along the diagonal set $\mathcal{D} = \{\sigma \in \mathcal{S}_d : \ \sigma_1 = \cdots = \sigma_d\}$ the flow reduces to a scalar ODE.

On the diagonal set $\mathcal{D}$, write

$$\sigma_i = s^N.$$

Set

$$\nu := Np, \qquad s_\star := \sigma_\star^{1/N},$$

and let

$$I_- = (0, s_\star), \qquad I_+ = (s_\star, \infty).$$

On each component $I_\pm$ define a real primitive $\mathcal{T}_\pm$ by

$$\frac{d}{ds}\mathcal{T}_\pm(s) = \frac{s}{s_\star^\nu - s^\nu}, \qquad s \in I_\pm. \tag{3.13}$$

The additive constants in $\mathcal{T}_\pm$ are immaterial. On I_- we use the standard real hypergeometric branch

$$\mathcal{T}_-(s) = \frac{s^2}{2s_\star^\nu}\,{}_2F_1\!\left(1, \frac{2}{\nu}; 1+\frac{2}{\nu}; \left(\frac{s}{s_\star}\right)^\nu\right), \qquad 0 < s < s_\star, \tag{3.14}$$

where ${}_2F_1$ denotes the Gauss hypergeometric function [Olv+10]. On I_+, $\mathcal{T}_+$ denotes the corresponding real branch of the same antiderivative; equivalently, it is obtained by integrating $s/(s_\star^\nu - s^\nu)$ on I_+, with the logarithmic term in the analytic continuation read as $\log|1-(s/s_\star)^\nu|$.

Theorem 3.12 (Exact solution on $\mathcal{D}$)**.** *For the Schatten–p energy, the diagonal* *variable* $s(t)$ *satisfies*

$$\dot{s} = -s^{\nu-1} + \frac{s_\star^\nu}{s}. \tag{3.15}$$

If $s(t_0) = s_\star$, *then* $s(t) \equiv s_\star$. *If* $s(t_0) = s_0 \neq s_\star$, *then the solution remains in the component* I_- *or* I_+ *containing* s_0 *and obeys the real quadrature*

$$t - t_0 = \int_{s_0}^{s(t)} \frac{\tau \, \mathrm{d}\tau}{s_\star^\nu - \tau^\nu} = \mathcal{T}_\pm\big(s(t)\big) - \mathcal{T}_\pm(s_0), \tag{3.16}$$

where the sign is chosen so that $s_0, s(t) \in I_\pm$.

3.2 The chamber of singular values

We now compute the metric induced by g_N on the singular-value chamber. This is the basic reduction step: once the metric is known on the chamber, the spectral flow is obtained by an elementary gradient calculation.

In Chapter 2, we showed that for $X \in \mathrm{GL}(d, \mathbb{R})$ the operator $\mathcal{A}_{N,X} : T_X\mathrm{GL}(d, \mathbb{R})^* \to T_X\mathrm{GL}(d, \mathbb{R})$ is

$$\mathcal{A}_{N,X}(P) := \sum_{k=1}^{N} (XX^T)^{\frac{N-k}{N}} \, P \, (X^TX)^{\frac{k-1}{N}}, \tag{3.17}$$

and the DLN metric is

$$\|Z\|_{g_N}^2 = \mathrm{tr}\Big(Z^T \mathcal{A}_{N,X}^{-1} Z\Big), \qquad Z \in T_X\mathrm{GL}(d, \mathbb{R}). \tag{3.18}$$

We begin with the spectral decomposition of $\mathcal{A}_{N,X}$.

Lemma 3.13 ([Men25])**.** *Let* $X = U\Sigma V^T$ *be the singular value decomposition of* X. *The operator* $\mathcal{A}_{N,X} : T_X\mathrm{GL}(d, \mathbb{R})^* \to T_X\mathrm{GL}(d, \mathbb{R})$ *is symmetric and positive definite* *with respect to the Frobenius inner product. It has the spectral decomposition*

$$\mathcal{A}_{N,X}\, u_k v_\ell^T = \frac{\sigma_k^2 - \sigma_\ell^2}{\sigma_k^{2/N} - \sigma_\ell^{2/N}}\, u_k v_\ell^T, \qquad 1 \leq k, \ell \leq d,$$

when $k \neq \ell$, and

$$\mathcal{A}_{N,X}\, u_k v_k^T = N\, \sigma_k^{2-2/N}\, u_k v_k^T, \qquad 1 \leq k \leq d,$$

where u_k and v_ℓ are the columns of U and V, respectively.

We also use the Riemannian submersion from Chapter 2.

Theorem 3.14 ([MY26b])**.** *The map*

$$\phi : (\mathcal{M}, \iota) \longrightarrow (\mathrm{GL}(d, \mathbb{R}), g_N)$$

is a Riemannian submersion.

On the open set where the singular values are distinct, the singular-value map is smooth. The induced metric on $\mathcal{S}_d^\circ$ is

$$g_{N,\sigma}(\dot{\sigma}, \dot{\sigma}') = \sum_{i=1}^{d} \frac{1}{N}\, \sigma_i^{2/N-2}\, \dot{\sigma}_i\, \dot{\sigma}_i'. \tag{3.19}$$

Lemma 3.15. *On the open set*

$$\{X \in \mathrm{GL}(d, \mathbb{R}) : \ \sigma_1(X) > \cdots > \sigma_d(X) > 0\},$$

the singular-value map

$$X \longmapsto (\sigma_1(X), \ldots, \sigma_d(X))$$

is a Riemannian submersion from $(\mathrm{GL}(d, \mathbb{R}), g_N)$ *to* $(\mathcal{S}_d^\circ, g_{N,\sigma})$*.*

Proof. Let $X \in \mathrm{GL}(d, \mathbb{R})$ have distinct singular values and write $X = U\Sigma V^T$. Set $E_{k\ell} := u_k v_\ell^T$. By Lemma 3.13,

$$\mathcal{A}_{N,X} E_{k\ell} = \begin{cases} \dfrac{\sigma_k^2 - \sigma_\ell^2}{\sigma_k^{2/N} - \sigma_\ell^{2/N}}\, E_{k\ell}, & k \neq \ell, \\ N\, \sigma_k^{2-2/N}\, E_{kk}, & k = \ell. \end{cases}$$

Hence $\mathcal{A}_{N,X}^{-1} E_{k\ell} = \mu_{k\ell} E_{k\ell}$ with $\mu_{kk} = \frac{1}{N}\sigma_k^{2/N-2}$ and

$$\mu_{k\ell} = \frac{\sigma_k^{2/N} - \sigma_\ell^{2/N}}{\sigma_k^2 - \sigma_\ell^2}, \qquad k \neq \ell.$$

Therefore $T_X\mathrm{GL}(d, \mathbb{R})$ decomposes as

$$T_X\mathrm{GL}(d, \mathbb{R}) = \underbrace{\mathrm{span}\{E_{kk}\}_{k=1}^{d}}_{\mathcal{H}_X} \oplus \underbrace{\mathrm{span}\{E_{k\ell} : k \neq \ell\}}_{\mathcal{V}_X},$$

and $\mathcal{H}_X$ and $\mathcal{V}_X$ are g_N-orthogonal.

The first-order perturbation formula for simple singular values gives

$$\mathrm{d}\sigma_k(X)[Z] = u_k^T Z v_k$$

[Kat13, Theorem II-5.4]. Hence $\ker \mathrm{d}\sigma(X) = \mathcal{V}_X$, and $\mathrm{d}\sigma(X)$ maps $\mathcal{H}_X$ isomorphically onto $T_{\sigma(X)}\mathcal{S}_d^\circ \cong \mathbb{R}^d$ since $\mathrm{d}\sigma(X)[E_{kk}] = e_k$. Therefore the singular-value map is a smooth submersion on this open set.

For $\dot\sigma, \dot\sigma' \in T_{\sigma(X)}\mathcal{S}_d^\circ \cong \mathbb{R}^d$, the horizontal lifts are

$$Z^{\mathrm{hor}} = \sum_i \dot\sigma_i E_{ii} = U\,\mathrm{diag}(\dot\sigma) V^T$$

and similarly for $\dot\sigma'$. Using $g_N(X)(E_{ii}, E_{jj}) = \mu_{ii}\,\delta_{ij}$, we obtain

$$g_N(X)\Big(Z^{\mathrm{hor}}, (Z')^{\mathrm{hor}}\Big) = \sum_{i=1}^{d} \mu_{ii}\,\dot\sigma_i\,\dot\sigma_i' = \sum_{i=1}^{d} \frac{1}{N}\,\sigma_i^{2/N-2}\,\dot\sigma_i\,\dot\sigma_i' = g_{N,\sigma}(\dot\sigma, \dot\sigma').$$

Thus $\mathrm{d}\sigma(X) : (\mathcal{H}_X, g_N) \to (T_{\sigma(X)}\mathcal{S}_d^\circ, g_{N,\sigma})$ is an isometry, which is the Riemannian submersion condition.

The choice of U and V does not affect $\mathcal{H}_X$ or the value of $g_{N,\sigma}$: when singular values are simple, the singular vectors are unique up to signs, and $\mathrm{span}\{u_k v_k^T\}$ is sign invariant. $\square$

Remark 3.16. The metric (3.19) extends continuously from $\mathcal{S}_d^\circ$ to all of $\mathcal{S}_d$. The singular-value map itself is smooth only when the singular values are distinct.

With this metric, the spectral gradient flow is immediate.

Lemma 3.17. *On $(\mathcal{S}_d, g_{N,\sigma})$ the gradient flow of the free energy $F_\beta(\sigma) = E(\sigma) - \beta^{-1}S_N(\sigma)$ has components*

$$\dot\sigma_i = -N\,\sigma_i^{2-2/N}\,\partial_{\sigma_i}F_\beta(\sigma), \qquad i = 1, \ldots, d.$$

Equivalently, if $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$ then

$$\dot\Sigma = -N\,\Sigma^{2-2/N}\ \mathrm{diag}\Big(\partial_{\sigma_i}F_\beta(\sigma)\Big). \tag{3.20}$$

Proof. By definition of the gradient, for any $\xi \in T_\sigma\mathcal{S}_d \cong \mathbb{R}^d$,

$$g_{N,\sigma}\Big(\mathrm{grad}_{g_{N,\sigma}} F_\beta, \xi\Big) = \mathrm{d}F_\beta(\sigma)[\xi] = \sum_{i=1}^{d} \frac{\partial F_\beta}{\partial \sigma_i}\,\xi_i.$$

Since $g_{N,\sigma}$ is diagonal with $g_{ii} = \frac{1}{N}\sigma_i^{2/N-2}$, its inverse has $g^{ii} = N\sigma_i^{2-2/N}$. Therefore

$$\operatorname{grad}_{g_{N,\sigma}} F_\beta = \left(g^{ii}\,\partial_{\sigma_i} F_\beta\right)_{i=1}^d = \left(N\sigma_i^{2-2/N}\,\partial_{\sigma_i} F_\beta\right)_{i=1}^d,$$

and the gradient flow $\dot\sigma = -\operatorname{grad}_{g_{N,\sigma}} F_\beta$ is as stated. □

Figure 3.1 shows representative trajectories of the gradient flow of the free energy on the singular-value chamber.

3.3 Analyticity and concavity of the entropy

We now study the entropy itself. The formula (3.2) contains quotients that vanish when $\sigma_i = \sigma_j$, so the first task is to show that these singularities are removable.

Proof of Theorem 3.1. Write $\lambda_i = \sigma_i^{1/N}$, $i = 1, \ldots, d$, so that $\lambda_i > 0$ whenever $\sigma_i > 0$. In these variables the entropy has the representation

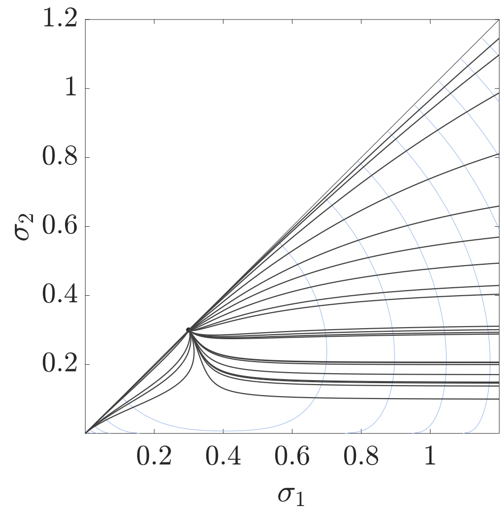
$$S_N(\lambda) = \widetilde{C}_N + \frac{1}{2}\sum_{1\le j<k\le d} \log\left(\frac{\lambda_j^{2N} - \lambda_k^{2N}}{\lambda_j^2 - \lambda_k^2}\right), \tag{3.21}$$

where $\widetilde{C}_N$ depends only on N and d. Introduce

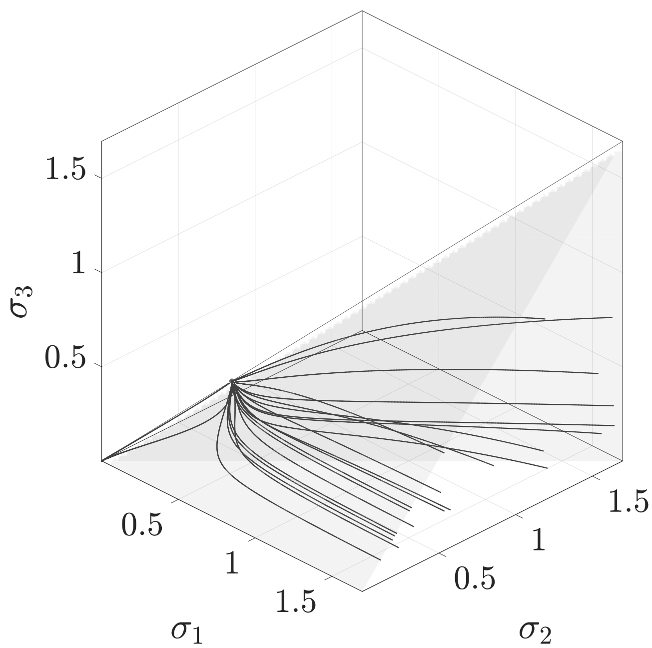
$$\Phi_N(a,b) := \frac{a^{2N} - b^{2N}}{a^2 - b^2}, \qquad a, b > 0, \tag{3.22}$$

so that (3.21) becomes

$$S_N(\lambda) = \widetilde{C}_N + \frac{1}{2}\sum_{1\le j<k\le d} \log \Phi_N(\lambda_j, \lambda_k). \tag{3.23}$$



(a) $d = 2$.



(b) $d = 3$.

Figure 3.1: Gradient flow of the free energy F_β on the singular-value chamber ($\beta = 5$). Black curves are trajectories of $\dot{\sigma}_i = -N\sigma_i^{2-2/N}\partial_{\sigma_i}F_\beta(\sigma)$. In panel (a), the blue curves indicate level sets of F_β and the flow is driven toward the equilibrium on the diagonal. Panel (b) shows the analogous dynamics for $d = 3$.

The quotient in (3.22) satisfies the algebraic identity

$$\frac{a^{2N} - b^{2N}}{a^2 - b^2} = \sum_{m=0}^{N-1} a^{2(N-1-m)} b^{2m}, \tag{3.24}$$

valid for all $a, b \in \mathbb{R}$. Thus Φ_N is a polynomial in (a, b) and hence real analytic on $\mathbb{R}^2$. In particular,

$$\Phi_N(a, a) = N\, a^{2N-2}, \tag{3.25}$$

so there is no singularity at $a = b$.

For $a, b > 0$, every term in (3.24) is nonnegative and at least one is strictly positive, so

$$\Phi_N(a, b) > 0 \qquad \text{for all } a, b > 0. \tag{3.26}$$

The logarithm is real analytic on $(0, \infty)$, hence the map

$$(a, b) \longmapsto \log \Phi_N(a, b) \tag{3.27}$$

is real analytic on $(0, \infty)^2$. Therefore each term $\log \Phi_N(\lambda_j, \lambda_k)$ in (3.23) is real analytic on $(0, \infty)^d$, and finite sums preserve analyticity. Thus $S_N(\lambda)$ is real analytic for all $\lambda \in (0, \infty)^d$.

The change of variables $\sigma_i = \lambda_i^N$ is real analytic on $(0, \infty)^d$ in each coordinate. Since $\mathcal{S}_d \subset (0, \infty)^d$, it follows that S_N is real analytic on $\mathcal{S}_d$.

Finally, the polynomial identity (3.24) shows that Φ_N is analytic at $a = b > 0$, so the expression (3.23) extends real analytically to points where $\lambda_j = \lambda_k > 0$. Via the change of variables $\sigma_i = \lambda_i^N$, this gives a real analytic extension of S_N across the sets $\sigma_i = \sigma_j > 0$. $\square$

Gradient and Hessian formulas. Lemmas 3.18 and 3.19 record the first and second derivatives that will be used in the concavity and equilibrium analysis.

Lemma 3.18. *The gradient of S_N has components*

$$\frac{\partial S_N}{\partial \sigma_i} = \sum_{k \neq i} \left(\frac{\sigma_i}{\sigma_i^2 - \sigma_k^2} - \frac{\sigma_i^{2/N-1}}{N\left(\sigma_i^{2/N} - \sigma_k^{2/N}\right)} \right). \tag{3.28}$$

For each fixed i and $k \neq i$ the summand has a finite limit as $\sigma_k \to \sigma_i = \sigma_\star > 0$, namely

$$\frac{\sigma_i}{\sigma_i^2 - \sigma_k^2} - \frac{\sigma_i^{2/N-1}}{N\left(\sigma_i^{2/N} - \sigma_k^{2/N}\right)} \longrightarrow \frac{1}{2\sigma_\star}\left(1 - \frac{1}{N}\right). \tag{3.29}$$

Proof. We start from the representation (3.21) in the variables $\lambda_i = \sigma_i^{1/N}$,

$$S_N(\lambda) = \widetilde{C}_N + \frac{1}{2} \sum_{1 \le j < k \le d} \log\left(\frac{\lambda_j^{2N} - \lambda_k^{2N}}{\lambda_j^2 - \lambda_k^2} \right), \tag{3.30}$$

valid for $\lambda_i > 0$. Differentiating (3.30) with respect to λ_i and noting that only pairs containing i contribute gives

$$\frac{\partial S_N}{\partial \lambda_i} = \sum_{k \neq i} \left(N \frac{\lambda_i^{2N-1}}{\lambda_i^{2N} - \lambda_k^{2N}} - \frac{\lambda_i}{\lambda_i^2 - \lambda_k^2} \right). \tag{3.31}$$

The change of variables $\sigma_i = \lambda_i^N$ implies

$$\frac{\partial S_N}{\partial \sigma_i} = \frac{1}{N\lambda_i^{N-1}} \frac{\partial S_N}{\partial \lambda_i}. \tag{3.32}$$

Substituting (3.31) into (3.32) yields

$$\begin{aligned}\frac{\partial S_N}{\partial \sigma_i} &= \sum_{k\neq i}\left(\frac{\lambda_i^N}{\lambda_i^{2N}-\lambda_k^{2N}} - \frac{1}{N}\frac{\lambda_i^{2-N}}{\lambda_i^2-\lambda_k^2}\right) \\ &= \sum_{k\neq i}\left(\frac{\sigma_i}{\sigma_i^2-\sigma_k^2} - \frac{\sigma_i^{2/N-1}}{N\left(\sigma_i^{2/N}-\sigma_k^{2/N}\right)}\right).\end{aligned} \tag{3.33}$$

This is exactly (3.28).

For the limit (3.29), set $\lambda_i = \lambda_\star$ and $\lambda_k = \lambda_\star(1-\varepsilon)$ with $\varepsilon \downarrow 0$. Then

$$(1-\varepsilon)^{2N} = 1 - 2N\varepsilon + N(2N-1)\varepsilon^2 + O(\varepsilon^3). \tag{3.34}$$

First expand the two terms in the λ_i-derivative:

$$N\frac{\lambda_i^{2N-1}}{\lambda_i^{2N}-\lambda_k^{2N}} = \frac{1}{2\lambda_\star\varepsilon} + \frac{2N-1}{4\lambda_\star} + O(\varepsilon), \qquad \frac{\lambda_i}{\lambda_i^2-\lambda_k^2} = \frac{1}{2\lambda_\star\varepsilon} + \frac{1}{4\lambda_\star} + O(\varepsilon). \tag{3.35}$$

Subtracting these expressions cancels the $1/\varepsilon$ term and yields the limit of the corresponding summand in $\partial S_N/\partial\lambda_i$:

$$N\frac{\lambda_i^{2N-1}}{\lambda_i^{2N}-\lambda_k^{2N}} - \frac{\lambda_i}{\lambda_i^2-\lambda_k^2} \longrightarrow \frac{N-1}{2\lambda_\star} \qquad (\varepsilon\downarrow 0). \tag{3.36}$$

To obtain the limit of the summand in $\partial S_N/\partial\sigma_i$, we must still apply the chain-rule factor in (3.32). Hence

$$\frac{1}{N\lambda_\star^{N-1}}\left(N\frac{\lambda_i^{2N-1}}{\lambda_i^{2N}-\lambda_k^{2N}} - \frac{\lambda_i}{\lambda_i^2-\lambda_k^2}\right) \longrightarrow \frac{1}{N\lambda_\star^{N-1}}\frac{N-1}{2\lambda_\star} = \frac{N-1}{2N\lambda_\star^N}. \tag{3.37}$$

Since $\lambda_\star^N = \sigma_\star$, this becomes

$$\frac{N-1}{2N\lambda_\star^N} = \frac{1}{2\sigma_\star}\left(1-\frac{1}{N}\right), \tag{3.38}$$

which is exactly (3.29).

Thus each summand in (3.28) has a finite limit as $\sigma_k \to \sigma_i > 0$, and the sum over $k \neq i$ extends continuously to points with repeated singular values. $\square$

For the renormalized entropy (3.9), the same argument gives

$$\frac{\partial S_\infty}{\partial \sigma_i} = \sum_{k \neq i} \left(\frac{\sigma_i}{\sigma_i^2 - \sigma_k^2} - \frac{\sigma_i^{-1}}{\log \sigma_i^2 - \log \sigma_k^2} \right), \tag{3.39}$$

and each summand has limit $\frac{1}{2\sigma_i}$ as $\sigma_k \to \sigma_i$.

Lemma 3.19. *For S_N the second derivatives in the coordinates σ are*

$$\frac{\partial^2 S_N}{\partial \sigma_i \partial \sigma_j} = \begin{cases} \displaystyle\sum_{k \neq i} \left(\frac{-\sigma_i^2 - \sigma_k^2}{(\sigma_i^2 - \sigma_k^2)^2} + \frac{\sigma_i^{2/N-2}\Big(N(\sigma_i^{2/N} - \sigma_k^{2/N}) + 2\sigma_k^{2/N}\Big)}{\Big(N(\sigma_i^{2/N} - \sigma_k^{2/N})\Big)^2} \right), & i = j, \\ \displaystyle\frac{2\sigma_i\sigma_j}{(\sigma_i^2 - \sigma_j^2)^2} - \frac{2\sigma_i^{2/N-1}\sigma_j^{2/N-1}}{\Big(N(\sigma_i^{2/N} - \sigma_j^{2/N})\Big)^2}, & i \neq j, \end{cases} \tag{3.40}$$

and $\nabla_\sigma^2 S_N$ extends continuously to all of $\mathcal{S}_d$.

Proof. We start from the expression for the gradient in σ coordinates (Lemma 3.18),

$$\frac{\partial S_N}{\partial \sigma_i} = \sum_{k \neq i} \left(\frac{\sigma_i}{\sigma_i^2 - \sigma_k^2} - \frac{\sigma_i^{2/N-1}}{N(\sigma_i^{2/N} - \sigma_k^{2/N})} \right). \tag{3.41}$$

Differentiating the kth summand in σ_j for $j \neq i$ gives the off-diagonal entries,

$$\frac{\partial^2 S_N}{\partial \sigma_i \partial \sigma_j} = \frac{2\sigma_i\sigma_j}{(\sigma_i^2 - \sigma_j^2)^2} - \frac{2\sigma_i^{2/N-1}\sigma_j^{2/N-1}}{\Big(N(\sigma_i^{2/N} - \sigma_j^{2/N})\Big)^2}, \qquad i \neq j,$$

and differentiating in σ_i and summing over $k \neq i$ gives the diagonal entries,

$$\frac{\partial^2 S_N}{\partial \sigma_i^2} = \sum_{k \neq i} \left(\frac{-\sigma_i^2 - \sigma_k^2}{(\sigma_i^2 - \sigma_k^2)^2} + \frac{\sigma_i^{2/N-2}\Big(N(\sigma_i^{2/N} - \sigma_k^{2/N}) + 2\sigma_k^{2/N}\Big)}{\Big(N(\sigma_i^{2/N} - \sigma_k^{2/N})\Big)^2} \right),$$

which is exactly (3.40).

Each off-diagonal entry is $p_N(\sigma_i, \sigma_j)$ and each diagonal summand is $q_N(\sigma_i, \sigma_k)$ in the notation of (3.43)–(3.44). Lemma 3.21 shows that p_N and q_N have finite limits as $\sigma_k \to \sigma_i > 0$, so all entries extend continuously to $\mathcal{S}_d$. $\square$

For the renormalized entropy (3.9), differentiating the gradient in σ coordinates gives

$$\frac{\partial^2 S_\infty}{\partial \sigma_i \partial \sigma_j} = \begin{cases} \displaystyle\sum_{k \neq i} \left(\frac{-\sigma_i^2 - \sigma_k^2}{(\sigma_i^2 - \sigma_k^2)^2} + \frac{\sigma_i^{-2}\Big(\log(\sigma_i/\sigma_k) + 1\Big)}{2\Big(\log(\sigma_i/\sigma_k)\Big)^2} \right), & i = j, \\ \displaystyle\frac{2\sigma_i\sigma_j}{(\sigma_i^2 - \sigma_j^2)^2} - \frac{\sigma_i^{-1}\sigma_j^{-1}}{2\Big(\log(\sigma_i/\sigma_j)\Big)^2}, & i \neq j, \end{cases} \tag{3.42}$$

and each summand again has a finite limit as $\sigma_j \to \sigma_i > 0$, so $\nabla_\sigma^2 S_\infty$ also extends continuously to $\mathcal{S}_d$.

Euclidean concavity. To analyze concavity with respect to the Euclidean metric, it is convenient to express the Hessian as a sum of embedded 2×2 blocks. Define the kernels

$$p_N(a, b) := \frac{2ab}{(a^2 - b^2)^2} - \frac{2a^{\frac{2}{N}-1}b^{\frac{2}{N}-1}}{\Big(N(a^{\frac{2}{N}} - b^{\frac{2}{N}})\Big)^2}, \tag{3.43}$$

$$q_N(a, b) := -\frac{a^2 + b^2}{(a^2 - b^2)^2} + \frac{a^{\frac{2}{N}-2}\Big(N(a^{\frac{2}{N}} - b^{\frac{2}{N}}) + 2b^{\frac{2}{N}}\Big)}{\Big(N(a^{\frac{2}{N}} - b^{\frac{2}{N}})\Big)^2}, \tag{3.44}$$

and for $1 \leq i < j \leq d$ let

$$\iota_{ij} : \mathbb{R}^2 \hookrightarrow \mathbb{R}^d, \qquad \iota_{ij}(u, v) = ue_i + ve_j, \tag{3.45}$$

with

$$B_N^{(ij)}(\sigma) = \begin{pmatrix} q_N(\sigma_i, \sigma_j) & p_N(\sigma_i, \sigma_j) \\ p_N(\sigma_i, \sigma_j) & q_N(\sigma_j, \sigma_i) \end{pmatrix}. \tag{3.46}$$

Lemma 3.20. *For every $\sigma \in \mathcal{S}_d$,*

$$\nabla_\sigma^2 S_N(\sigma) = \sum_{1 \le i < j \le d} \iota_{ij} B_N^{(ij)}(\sigma) \iota_{ij}^T. \tag{3.47}$$

Equivalently,

$$(\nabla_\sigma^2 S_N)_{ij} = p_N(\sigma_i, \sigma_j) \ (i \neq j), \qquad (\nabla_\sigma^2 S_N)_{ii} = \sum_{k \neq i} q_N(\sigma_i, \sigma_k).$$

Proof. By Lemma 3.19, for $i \neq j$ one has

$$\frac{\partial^2 S_N}{\partial \sigma_i \partial \sigma_j} = \frac{2\sigma_i \sigma_j}{(\sigma_i^2 - \sigma_j^2)^2} - \frac{2\sigma_i^{2/N-1} \sigma_j^{2/N-1}}{\left(N(\sigma_i^{2/N} - \sigma_j^{2/N})\right)^2} = p_N(\sigma_i, \sigma_j),$$

and for $i = j$,

$$\frac{\partial^2 S_N}{\partial \sigma_i^2} = \sum_{k \neq i} \left(\frac{-\sigma_i^2 - \sigma_k^2}{(\sigma_i^2 - \sigma_k^2)^2} + \frac{\sigma_i^{2/N-2}\left(N(\sigma_i^{2/N} - \sigma_k^{2/N}) + 2\sigma_k^{2/N}\right)}{\left(N(\sigma_i^{2/N} - \sigma_k^{2/N})\right)^2} \right) = \sum_{k \neq i} q_N(\sigma_i, \sigma_k).$$

On the indices $\{i, j\}$ the principal 2×2 block of $\nabla_\sigma^2 S_N$ is therefore $B_N^{(ij)}(\sigma)$, and composing with the injections ι_{ij} gives (3.47). $\square$

To study each block, write $\lambda_\ell = \sigma_\ell^{1/N}$ and $r = \lambda_i / \lambda_j > 1$.

Lemma 3.21. *For $i < j$ and $r = \lambda_i / \lambda_j > 1$,*

$$p_N(\sigma_i, \sigma_j) = \frac{1}{\sigma_j^2} \left(\frac{2r^N}{(r^{2N} - 1)^2} - \frac{2}{N^2} \frac{r^{2-N}}{(r^2 - 1)^2} \right), \tag{3.48}$$

$$q_N(\sigma_i,\sigma_j) = \frac{1}{\sigma_j^2}\left(-\frac{r^{2N}+1}{(r^{2N}-1)^2} + \frac{r^{2-2N}}{N^2}\frac{N(r^2-1)+2}{(r^2-1)^2}\right), \tag{3.49}$$

and in particular $p_N(\sigma_i,\sigma_j) < 0$ and $q_N(\sigma_i,\sigma_j) < 0$. As $r \downarrow 1$ (equivalently $\sigma_i \to \sigma_j = \sigma$),

$$p_N(\sigma_i,\sigma_j) \to -\frac{1}{6\sigma^2}\Big(1-\frac{1}{N^2}\Big), \qquad q_N(\sigma_i,\sigma_j) \to -\frac{1}{3\sigma^2}\Big(1-\frac{3}{2N}+\frac{1}{2N^2}\Big). \tag{3.50}$$

Proof. Substitute $\sigma_\ell = \lambda_\ell^N$ into Lemma 3.20. For the off-diagonal entry,

$$p_N(\sigma_i,\sigma_j) = \frac{2\lambda_i^N\lambda_j^N}{(\lambda_i^{2N}-\lambda_j^{2N})^2} - \frac{2\lambda_i^{2-N}\lambda_j^{2-N}}{N^2(\lambda_i^2-\lambda_j^2)^2}.$$

Factoring λ_j and setting $r = \lambda_i/\lambda_j$ gives

$$p_N(\sigma_i,\sigma_j) = \frac{1}{\lambda_j^{2N}}\left(\frac{2r^N}{(r^{2N}-1)^2} - \frac{2}{N^2}\frac{r^{2-N}}{(r^2-1)^2}\right) = \frac{1}{\sigma_j^2}\left(\frac{2r^N}{(r^{2N}-1)^2} - \frac{2}{N^2}\frac{r^{2-N}}{(r^2-1)^2}\right),$$

which is (3.48).

Writing $r = e^t$ ($t > 0$) and using

$$r^{2m}-1 = 2e^{mt}\sinh(mt), \qquad r^2-1 = 2e^t\sinh(t),$$

we obtain

$$\frac{2r^N}{(r^{2N}-1)^2} = \frac{1}{2}\frac{e^{-Nt}}{\sinh^2(Nt)}, \qquad \frac{2}{N^2}\frac{r^{2-N}}{(r^2-1)^2} = \frac{1}{2}\frac{e^{-Nt}}{N^2\sinh^2 t}.$$

Hence

$$p_N(\sigma_i,\sigma_j) = \frac{e^{-Nt}}{2\sigma_j^2}\left(\frac{1}{\sinh^2(Nt)} - \frac{1}{N^2\sinh^2 t}\right),$$

and $\sinh(Nt) > N\sinh t$ for $t > 0$, so $p_N(\sigma_i,\sigma_j) < 0$.

For the diagonal summand,

$$q_N(\sigma_i, \sigma_j) = -\frac{\lambda_i^{2N} + \lambda_j^{2N}}{(\lambda_i^{2N} - \lambda_j^{2N})^2} + \frac{\lambda_i^{2-2N}\Big(N(\lambda_i^2 - \lambda_j^2) + 2\lambda_j^2\Big)}{N^2(\lambda_i^2 - \lambda_j^2)^2}.$$

The same substitution yields (3.49) after factoring $\lambda_j^{2N} = \sigma_j^2$. Expressing again in $t = \log r > 0$ shows

$$q_N(\sigma_i, \sigma_j) = \frac{e^{-Nt}}{2\sigma_j^2}\left(-\frac{\cosh(Nt)}{\sinh^2(Nt)} + \frac{e^{-Nt}}{N^2}\frac{Ne^t \sinh t + 1}{\sinh^2 t}\right),$$

and a direct comparison using $\sinh(Nt) > N \sinh t$ and $\cosh(Nt) \geq 1$ yields strict negativity for all $t > 0$ and $N \geq 2$.

The limits in (3.50) as $r \downarrow 1$ follow by Taylor expansion. Writing $r = e^t$ with $t \downarrow 0$ and using

$$\sinh t = t + \tfrac{1}{6}t^3 + O(t^5),$$

$$\sinh(Nt) = Nt + \tfrac{N^3}{6}t^3 + O(t^5),$$

$$\cosh(Nt) = 1 + \tfrac{N^2}{2}t^2 + O(t^4)$$

one obtains the stated limits after a straightforward calculation. $\square$

Lemma 3.22. *For $i < j$:*

1. *If $N = 2$ and $\sigma_i \neq \sigma_j$, then*

$$B_2^{(ij)}(\sigma) = -\frac{1}{2(\sigma_i + \sigma_j)^2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \tag{3.51}$$

   and $B_2^{(ij)}$ is rank-one negative semidefinite.

2. *If $N > 2$ and $\sigma_i \neq \sigma_j$, then $B_N^{(ij)}(\sigma) \prec 0$.*

Proof. (1) For $N = 2$, insert $N = 2$ into (3.43)–(3.44). Using $(\sigma_i^2 - \sigma_j^2)^2 = (\sigma_i -$

$\sigma_j)^2(\sigma_i+\sigma_j)^2$,

$$\frac{2\sigma_i\sigma_j}{(\sigma_i^2-\sigma_j^2)^2}-\frac{1}{2(\sigma_i-\sigma_j)^2}=-\frac{1}{2(\sigma_i+\sigma_j)^2},$$

$$-\frac{\sigma_i^2+\sigma_j^2}{(\sigma_i^2-\sigma_j^2)^2}+\frac{1}{2(\sigma_i-\sigma_j)^2}=-\frac{1}{2(\sigma_i+\sigma_j)^2},$$

so $p_2(\sigma_i,\sigma_j)=q_2(\sigma_i,\sigma_j)=-1/(2(\sigma_i+\sigma_j)^2)$, which gives (3.51).

(2) For $N>2$, Lemma 3.21 gives

$$p_N(\sigma_i,\sigma_j)<0, \qquad q_N(\sigma_i,\sigma_j)<0, \qquad q_N(\sigma_j,\sigma_i)<0,$$

so $\operatorname{tr} B_N^{(ij)}<0$. It remains to show $\det B_N^{(ij)}>0$.

Set $r:=\lambda_i/\lambda_j>1$. From Lemma 3.21 and the same formula with the two arguments interchanged, write

$$p_N(\sigma_i,\sigma_j)=\frac{1}{\sigma_j^2}P_N(r), \qquad q_N(\sigma_i,\sigma_j)=\frac{1}{\sigma_j^2}Q_N^+(r), \qquad q_N(\sigma_j,\sigma_i)=\frac{1}{\sigma_j^2}Q_N^-(r),$$

where

$$\begin{aligned}
P_N(r)&=\frac{2r^N}{(r^{2N}-1)^2}-\frac{2}{N^2}\frac{r^{2-N}}{(r^2-1)^2},\\
Q_N^+(r)&=-\frac{r^{2N}+1}{(r^{2N}-1)^2}+\frac{r^{2-2N}}{N^2}\frac{N(r^2-1)+2}{(r^2-1)^2},\\
Q_N^-(r)&=-\frac{r^{2N}+1}{(r^{2N}-1)^2}+\frac{N(1-r^2)+2r^2}{N^2(r^2-1)^2}.
\end{aligned}$$

Thus

$$\det B_N^{(ij)}=\frac{1}{\sigma_j^4}\,\Delta_N(r), \qquad \Delta_N(r):=Q_N^+(r)Q_N^-(r)-P_N(r)^2.$$

From the limits in Lemma 3.21 we obtain

$$\Delta_N(1) = \left(-\frac{1}{3}\left(1 - \frac{3}{2N} + \frac{1}{2N^2}\right)\right)^2 - \left(-\frac{1}{6}\left(1 - \frac{1}{N^2}\right)\right)^2 = \frac{(N-2)(N-1)^2}{12N^3} > 0.$$

Set $u = r^2$ and

$$G_N(u) := \sum_{m=0}^{N-1} u^m.$$

A direct simplification of the explicit formulas in Lemma 3.21 gives

$$\Delta_N(r) = \frac{H_N(u)}{N^3 u^{N-1} G_N(u)^2}, \tag{3.52}$$

where

$$H_N(u) = \sum_{k=0}^{N-3} c_{N,k}\left(u^k + u^{2N-4-k}\right) - \binom{N}{3} u^{N-2}, \tag{3.53}$$

with

$$c_{N,k} = \binom{k+2}{2}\left(N(N-k-3) + \frac{2(k+3)}{3}\right). \tag{3.54}$$

For $0 \le k \le N-3$ these coefficients are positive. Moreover, for $u > 0$,

$$u^k + u^{2N-4-k} \ge 2u^{N-2}$$

by the arithmetic–geometric mean inequality. Hence

$$\begin{aligned} H_N(u) &\ge \left(2\sum_{k=0}^{N-3} c_{N,k} - \binom{N}{3}\right) u^{N-2} \\ &= \frac{N^2(N-1)^2(N-2)}{12} u^{N-2} > 0, \end{aligned}$$

where the equality follows by summing the explicit polynomial expression for $c_{N,k}$. Therefore $\Delta_N(r) > 0$ for all $r > 1$. The same representation shows $\Delta_N(r) \to 0^+$ as

$r \to \infty$. Thus $\det B_N^{(ij)} > 0$, and with negative trace we conclude $B_N^{(ij)} \prec 0$. $\square$

Lemma 3.23. *Let*

$$A = \sum_{1 \le i < j \le d} \iota_{ij} B^{(ij)} \iota_{ij}^T$$

with each $B^{(ij)}$ symmetric. Then:

1. *If $B^{(ij)} \prec 0$ for all $i < j$, then $A \prec 0$.*

2. *If each $B^{(ij)} = -\gamma_{ij} v v^T$ with $\gamma_{ij} > 0$ and $v = (1,1)^T$, then $A \preceq 0$, with strict negativity when $d \ge 3$ and rank one when $d = 2$.*

Proof. For $x \in \mathbb{R}^d$ set $y_{ij} := \iota_{ij}^T x = (x_i, x_j)^T \in \mathbb{R}^2$. Then

$$x^T A x = \sum_{1 \le i < j \le d} y_{ij}^T B^{(ij)} y_{ij}.$$

(1) If each $B^{(ij)} \prec 0$, then for any nonzero x, pick i with $x_i \neq 0$ and some $j \neq i$. Then $y_{ij} \neq 0$ and $y_{ij}^T B^{(ij)} y_{ij} < 0$, while all other terms are ≤ 0. Thus $x^T A x < 0$ for all $x \neq 0$, so $A \prec 0$.

(2) If each $B^{(ij)} = -\gamma_{ij} v v^T$ with $\gamma_{ij} > 0$ and $v = (1,1)^T$, then

$$y_{ij}^T B^{(ij)} y_{ij} = -\gamma_{ij} (x_i + x_j)^2 \le 0,$$

so $A \preceq 0$. If $x^T A x = 0$, then $x_i + x_j = 0$ for all pairs $i < j$.

For $d \ge 3$, this system forces $x = 0$, so $A \prec 0$. For $d = 2$, the single condition is $x_1 + x_2 = 0$, so $\ker A = \operatorname{span}\{(1,-1)^T\}$ and $\operatorname{rank}(A) = 1$. $\square$

We can now prove Euclidean concavity.

Proof of Theorem 3.2. Fix $\sigma \in \mathcal{S}_d$. By Lemma 3.20, the Hessian admits the block decomposition

$$\nabla_\sigma^2 S_N(\sigma) = \sum_{1\le i<j\le d} \iota_{ij} B_N^{(ij)}(\sigma)\iota_{ij}^T, \qquad B_N^{(ij)}(\sigma) = \begin{pmatrix} q_N(\sigma_i,\sigma_j) & p_N(\sigma_i,\sigma_j) \\ p_N(\sigma_i,\sigma_j) & q_N(\sigma_j,\sigma_i) \end{pmatrix}.$$

Case $N = 2$. Lemma 3.22 gives, for every unordered pair $\{i,j\}$ (including $\sigma_i = \sigma_j$ via the limits in Lemma 3.21),

$$B_2^{(ij)}(\sigma) = -\frac{1}{2(\sigma_i+\sigma_j)^2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} =: -\gamma_{ij} v v^T, \qquad \gamma_{ij} > 0,\ v = (1,1)^T.$$

Thus $\nabla_\sigma^2 S_N(\sigma)$ is a sum of embedded rank-one negative semidefinite blocks of the form $-\gamma_{ij} v v^T$. By Lemma 3.23, the sum is negative semidefinite for all d. When $d \ge 3$, the embedded directions $\iota_{ij} v = e_i + e_j$ span $\mathbb{R}^d$, so the Hessian is negative definite. When $d = 2$, there is a one-dimensional kernel $\operatorname{span}\{(1,-1)^T\}$ and $\nabla_\sigma^2 S_N(\sigma)$ has rank one.

Case $N > 2$. First suppose $\sigma_i \ne \sigma_j$ for all $i \ne j$. Lemma 3.22 shows that each block $B_N^{(ij)}(\sigma)$ is negative definite. Applying Lemma 3.23 to the block sum yields

$$\nabla_\sigma^2 S_N(\sigma) \prec 0$$

at every point with distinct singular values.

It remains to treat points with $\sigma_i = \sigma_j = \sigma$ for some $i \ne j$. By Lemma 3.21, as $\sigma_i \to \sigma_j = \sigma$ one has

$$p_N(\sigma_i,\sigma_j) \longrightarrow -\frac{1}{6\sigma^2}\Big(1 - \frac{1}{N^2}\Big), \qquad q_N(\sigma_i,\sigma_j) \longrightarrow -\frac{1}{3\sigma^2}\Big(1 - \frac{3}{2N} + \frac{1}{2N^2}\Big),$$

so the limiting 2×2 block is

$$\widehat{B}_N := \begin{pmatrix} q & p \\ p & q \end{pmatrix}, \qquad p = -\frac{1}{6\sigma^2}\Big(1 - \frac{1}{N^2}\Big), \quad q = -\frac{1}{3\sigma^2}\Big(1 - \frac{3}{2N} + \frac{1}{2N^2}\Big).$$

The eigenvalues of $\widehat{B}_N$ are $q \pm p$, and a direct calculation gives

$$q + p = -\frac{1}{2\sigma^2}\Big(1 - \frac{1}{N}\Big) < 0$$

and

$$\begin{aligned} q - p &= -\frac{1}{3\sigma^2}\Big(1 - \frac{3}{2N} + \frac{1}{2N^2}\Big) + \frac{1}{6\sigma^2}\Big(1 - \frac{1}{N^2}\Big) \\ &= -\frac{1}{6\sigma^2}\Big(1 - \frac{3}{N} + \frac{2}{N^2}\Big) \\ &= -\frac{(N-1)(N-2)}{6N^2\sigma^2} < 0 \qquad \text{for } N > 2. \end{aligned}$$

Thus $\widehat{B}_N \prec 0$, and by continuity this is the value of $B_N^{(ij)}(\sigma)$ on $\{\sigma_i = \sigma_j\}$. Hence each block $B_N^{(ij)}(\sigma)$ is negative definite for all $\sigma \in \mathcal{S}_d$ when $N > 2$. Lemma 3.23 then implies that $\nabla^2_\sigma S_N(\sigma) \prec 0$ on $(\mathcal{S}_d, \iota)$.

Combining the two cases, we obtain that S_N has negative definite Hessian on $(\mathcal{S}_d, \iota)$ for all $(N, d) \neq (2, 2)$, and in the exceptional case $(N, d) = (2, 2)$ the Hessian has rank one. $\square$

The DLN metric. We now compute the Hessian with respect to the metric $g_{N,\sigma}$. For a smooth function $f : \mathcal{S}_d \to \mathbb{R}$, let Γ^k_{ij} denote the Christoffel symbols of $g_{N,\sigma}$ in the coordinates σ. Then

$$(\nabla^2_{g_{N,\sigma}} f)_{ij} = \partial^2_{ij} f - \sum_{k=1}^{d} \Gamma^k_{ij}\, \partial_k f. \tag{3.55}$$

Lemma 3.24. *For any smooth $f : \mathcal{S}_d \to \mathbb{R}$ one has*

$$\left(\nabla^2_{g_{N,\sigma}} f\right)_{ij} = \frac{\partial^2 f}{\partial\sigma_i \partial\sigma_j} + \delta_{ij}\, \frac{N-1}{N}\, \frac{1}{\sigma_i}\, \frac{\partial f}{\partial \sigma_i}. \tag{3.56}$$

In particular, if the Euclidean gradient and Hessian of f extend continuously across the sets $\{\sigma_i = \sigma_j\}$, then so does $\nabla^2_{g_{N,\sigma}} f$.

Proof. The metric $g_{N,\sigma}$ is diagonal in the σ coordinates with

$$g_{ii}(\sigma) = \frac{1}{N}\, \sigma_i^{2/N-2}, \qquad g_{ij}(\sigma) = 0 \quad (i \neq j),$$

so

$$g^{ii}(\sigma) = N\, \sigma_i^{2-2/N}, \qquad g^{ij}(\sigma) = 0 \quad (i \neq j).$$

For a diagonal metric the only nonzero Christoffel symbols are

$$\Gamma^i_{ii} = \frac{1}{2} g^{ii}\, \partial_i g_{ii}, \qquad \Gamma^k_{ij} = 0 \quad \text{if } k \neq i \text{ or } \ i \neq j.$$

A direct computation gives

$$\partial_i g_{ii} = \frac{1}{N}\Big(\frac{2}{N} - 2\Big)\sigma_i^{2/N-3} = \frac{2}{N}\Big(\frac{1}{N} - 1\Big)\sigma_i^{2/N-3},$$

and hence

$$\Gamma^i_{ii} = \frac{1}{2} N \sigma_i^{2-2/N} \cdot \frac{2}{N}\Big(\frac{1}{N} - 1\Big)\sigma_i^{2/N-3} = \Big(\frac{1}{N} - 1\Big)\frac{1}{\sigma_i} = -\frac{N-1}{N}\frac{1}{\sigma_i}.$$

All other Γ^k_{ij} vanish. Substituting into (3.55) yields

$$(\nabla^2_{g_{N,\sigma}} f)_{ij} = \partial^2_{ij} f - \Gamma^i_{ij}\, \partial_i f = \partial^2_{ij} f + \delta_{ij}\, \frac{N-1}{N}\frac{1}{\sigma_i}\, \partial_i f,$$

which is (3.56). The continuity statement follows immediately from the continuity of the Euclidean derivatives and the explicit factor $1/\sigma_i$. $\square$

At a point $\vec{\sigma}_\star = (\sigma_\star, \ldots, \sigma_\star)$, set

$$p_\star := -\frac{1}{6\sigma_\star^2}\Big(1 - \frac{1}{N^2}\Big), \qquad q_\star := -\frac{1}{3\sigma_\star^2}\Big(1 - \frac{3}{2N} + \frac{1}{2N^2}\Big). \tag{3.57}$$

Lemma 3.25. *At $\vec{\sigma}_\star$ the matrix $\nabla^2_{g_{N,\sigma}} S_N$ has constant entries*

$$\Big(\nabla^2_{g_{N,\sigma}} S_N\Big)_{ij} = \begin{cases} (d-1)q_\star + (d-1)\chi_N, & i = j, \\ p_\star, & i \neq j, \end{cases} \tag{3.58}$$

where

$$\chi_N := \frac{(N-1)^2}{2N^2\sigma_\star^2}.$$

Consequently, the eigenvalues of $\nabla^2_{g_{N,\sigma}} S_N(\vec{\sigma}_\star)$ are

$$\theta_{\mathbf{1}}(S_N) = (d-1)\Big(q_\star + p_\star + \chi_N\Big) = -\frac{d-1}{2\sigma_\star^2} \cdot \frac{N-1}{N^2} < 0, \tag{3.59}$$

$$\theta_\perp(S_N) = (d-1)\Big(q_\star + \chi_N\Big) - p_\star = \frac{1}{\sigma_\star^2}\left(\frac{d}{6} - \frac{d-1}{2N} + \frac{2d-3}{6N^2}\right) > 0, \tag{3.60}$$

where $\theta_{\mathbf{1}}(S_N)$ corresponds to the eigenvector $\mathbf{1} = (1, \ldots, 1)$ and $\theta_\perp(S_N)$ is the common eigenvalue on $\operatorname{span}\{\mathbf{1}\}^\perp$ with multiplicity $d-1$.

Proof. From Lemma 3.19, at $\vec{\sigma}_\star$ the Euclidean Hessian has off-diagonal entries $p_\star$ and diagonal entries

$$\Big(\nabla^2_\sigma S_N(\vec{\sigma}_\star)\Big)_{ii} = \sum_{k \neq i} q_\star = (d-1)q_\star.$$

The correction term in (3.56) contributes only on the diagonal. Using (3.29) at

$\sigma_i = \sigma_k = \sigma_\star$,

$$\frac{\partial S_N}{\partial \sigma_i}(\vec{\sigma}_\star) = \sum_{k \neq i} \frac{1}{2\sigma_\star}\Big(1 - \frac{1}{N}\Big) = \frac{d-1}{2\sigma_\star}\Big(1 - \frac{1}{N}\Big), \tag{3.61}$$

independent of i. Therefore

$$\frac{N-1}{N}\frac{1}{\sigma_i}\frac{\partial S_N}{\partial \sigma_i}(\vec{\sigma}_\star) = \frac{N-1}{N}\frac{1}{\sigma_\star} \cdot \frac{d-1}{2\sigma_\star}\Big(1 - \frac{1}{N}\Big) = (d-1)\chi_N,$$

which is again independent of i. Thus

$$\Big(\nabla^2_{g_{N,\sigma}} S_N(\vec{\sigma}_\star)\Big)_{ii} = (d-1)q_\star + (d-1)\chi_N, \qquad \Big(\nabla^2_{g_{N,\sigma}} S_N(\vec{\sigma}_\star)\Big)_{ij} = p_\star \quad (i \neq j),$$

giving (3.58).

A matrix with constant diagonal entry a and constant off-diagonal entry b has eigenvalues

$$a + (d-1)b \quad \text{on } \mathbf{1}, \qquad a - b \quad \text{with multiplicity } d-1$$

on $\operatorname{span}\{\mathbf{1}\}^\perp$. Here

$$a = (d-1)(q_\star + \chi_N), \qquad b = p_\star.$$

Substituting (3.57)–(3.58) and simplifying yields (3.59) and (3.60). $\square$

Proof of Theorem 3.3. By Lemma 3.25, at any point $\vec{\sigma}_\star = (\sigma_\star, \ldots, \sigma_\star)$ the Hessian $\nabla^2_{g_{N,\sigma}} S_N(\vec{\sigma}_\star)$ has one negative eigenvalue $\theta_{\mathbf{1}}(S_N)$ and $d-1$ positive eigenvalues $\theta_\perp(S_N)$. Thus the Hessian is indefinite at every such point, so S_N is not concave on $(\mathcal{S}_d, g_{N,\sigma})$. $\square$

3.4 Equilibria and local rates

We now turn from the entropy alone to the free energy $F_\beta = E - \beta^{-1} S_N$. The key point is that the first-order conditions force all singular values at equilibrium to coincide.

For brevity, set

$$r_N(a,b) := \frac{a}{a^2-b^2} - \frac{a^{2/N-1}}{N\Big(a^{2/N}-b^{2/N}\Big)}, \tag{3.62}$$

so that

$$\partial_i S_N(\sigma) = \sum_{k\neq i} r_N(\sigma_i, \sigma_k).$$

Lemma 3.26. *For each fixed $b > 0$, the map $a \mapsto r_N(a,b)$ is strictly decreasing on $(0,\infty)$.*

Proof. Differentiating (3.62) in a gives the kernel $q_N(a,b)$ from Lemma 3.20. By Lemma 3.21, $q_N(a,b) < 0$ for $a \neq b$, hence $r_N(\cdot, b)$ is strictly decreasing. □

Lemma 3.27. *For $a \geq b > 0$ one has*

$$r_N(a,b) - r_N(b,a) \leq 0, \tag{3.63}$$

with equality if and only if $N = 2$ or $a = b$.

Proof. Let $\alpha := 2/N \in (0,1]$ and write $a = rb$ with $r > 1$. Using

$$\frac{a+b}{a^2-b^2} = \frac{1}{a-b} = \frac{1}{b(r-1)}, \qquad \frac{a^{\alpha-1}+b^{\alpha-1}}{a^\alpha - b^\alpha} = \frac{1}{b}\frac{r^{\alpha-1}+1}{(r-1)h_\alpha(r)}, \tag{3.64}$$

where $h_\alpha(r) := \dfrac{r^\alpha - 1}{r-1}$, we obtain

$$r_N(a,b) - r_N(b,a) = \frac{1}{b(r-1)}\left(1 - \frac{1}{N}\frac{r^{\alpha-1}+1}{h_\alpha(r)}\right). \tag{3.65}$$

Since $t \mapsto t^{\alpha-1}$ is decreasing on $[1,\infty)$ and

$$h_\alpha(r) = \frac{1}{r-1}\int_1^r \alpha t^{\alpha-1}\,\mathrm{d}t,$$

the trapezoid bound gives

$$h_\alpha(r) \le \frac{\alpha}{2}\Big(1 + r^{\alpha-1}\Big). \tag{3.66}$$

Thus $\dfrac{r^{\alpha-1}+1}{h_\alpha(r)} \ge \dfrac{2}{\alpha} = N$, so the bracket in (3.65) is ≤ 0, with equality only when $\alpha = 1$ (that is, $N = 2$) or $r = 1$ (that is, $a = b$). $\square$

Lemma 3.28. *The equation*

$$g'\Big(d\,f(\sigma)\Big)f'(\sigma) = \beta^{-1}\,\frac{d-1}{2\sigma}\Big(1 - \frac{1}{N}\Big) \tag{3.67}$$

has a unique solution $\sigma_\star > 0$.

Proof. Define

$$L(\sigma) = g'\Big(d\,f(\sigma)\Big)f'(\sigma), \tag{3.68}$$

and

$$R(\sigma) = \beta^{-1}\,\frac{d-1}{2\sigma}\Big(1 - \frac{1}{N}\Big). \tag{3.69}$$

By (3.6),

$$L'(\sigma) = d\,g''\Big(d\,f(\sigma)\Big)\,[f'(\sigma)]^2 + g'\Big(d\,f(\sigma)\Big)f''(\sigma) > 0, \tag{3.70}$$

so L is strictly increasing on $(0,\infty)$. Since $g' > 0$ and $f' \ge 0$, we have $L(\sigma) > 0$ for every $\sigma > 0$. On the other hand,

$$R'(\sigma) = -\beta^{-1}\,\frac{d-1}{2\sigma^2}\Big(1 - \frac{1}{N}\Big) < 0, \tag{3.71}$$

so R is strictly decreasing. Moreover,

$$\lim_{\sigma \downarrow 0} R(\sigma) = +\infty, \qquad \lim_{\sigma \to \infty} R(\sigma) = 0.$$

Thus $R(\sigma) > L(\sigma)$ for σ sufficiently small. Fix any $\sigma_0 > 0$. Since $L(\sigma_0) > 0$ and $R(\sigma) \to 0$ as $\sigma \to \infty$, there exists $\sigma_1 \geq \sigma_0$ such that $R(\sigma_1) < L(\sigma_0) \leq L(\sigma_1)$. Therefore L and R cross at least once. Strict monotonicity of L and R shows that the crossing is unique. $\square$

Proof of Theorem 3.6. Let $\sigma \in \mathcal{S}_d$ be an equilibrium of F_β. Since the coefficients $N\sigma_i^{2-2/N}$ in (3.7) are strictly positive, stationarity of (3.7) is equivalent to

$$\partial_{\sigma_i} F_\beta(\sigma) = 0, \qquad i = 1, \ldots, d. \tag{3.72}$$

Using $F_\beta(\sigma) = E(\sigma) - \beta^{-1} S_N(\sigma)$ and (3.5), together with Lemma 3.18 and the definition (3.62), the condition (3.72) becomes

$$g'\Big(\sum_{k=1}^{d} f(\sigma_k)\Big) f'(\sigma_i) = \beta^{-1} \sum_{k \neq i} r_N(\sigma_i, \sigma_k), \qquad i = 1, \ldots, d. \tag{3.73}$$

Fix $i \neq j$ and subtract the jth equation in (3.73) from the ith to obtain

$$\begin{aligned} g'\Big(\sum_{k=1}^{d} f(\sigma_k)\Big)\Big(f'(\sigma_i) - f'(\sigma_j)\Big) &= \beta^{-1}\Big(r_N(\sigma_i, \sigma_j) - r_N(\sigma_j, \sigma_i)\Big) \\ &\quad + \beta^{-1} \sum_{k \neq i,j} \Big(r_N(\sigma_i, \sigma_k) - r_N(\sigma_j, \sigma_k)\Big). \end{aligned} \tag{3.74}$$

If $\sigma_i > \sigma_j$, then the left-hand side of (3.74) is > 0 by (3.6), while the right-hand side is ≤ 0 by Lemmas 3.27 and 3.26. This contradiction shows that no strict inequality among the singular values is possible. Hence $\sigma_1 = \cdots = \sigma_d =: \sigma_\star > 0$.

Substituting this equality into (3.73) and interpreting $r_N(\sigma_\star, \sigma_\star)$ by the limit (3.29) yields exactly (3.8). By Lemma 3.28, equation (3.67) has a unique solution $\sigma_\star > 0$, hence the equilibrium $\sigma = (\sigma_\star, \dots, \sigma_\star)$ in $\mathcal{S}_d$ is unique.

Finally, (3.6) implies that E is convex on $\mathcal{S}_d$: the map $\sigma \mapsto \sum_{k=1}^d f(\sigma_k)$ is convex, and composition with the convex nondecreasing function g preserves convexity. Theorem 3.2 shows that S_N is concave on $(\mathcal{S}_d, \iota)$. Therefore F_β is convex on $\mathcal{S}_d$, so its unique critical point is its unique minimizer. $\square$

Remark 3.29. The proof of Theorem 3.6 uses the first-order conditions and the monotonicity of the kernel in the entropy. A shorter argument is available when $g' > 0$ and f is strictly convex: then F_β is strictly convex in the variables $\sigma = (\sigma_1, \dots, \sigma_d)$, and symmetry under permutations of the σ_i forces the unique minimizer to lie on the diagonal.

We now linearize the flow at $\vec{\sigma}_\star$. For convenience, recall that

$$p_\star := -\frac{1}{6\sigma_\star^2}\Big(1 - \frac{1}{N^2}\Big), \qquad q_\star := -\frac{1}{3\sigma_\star^2}\Big(1 - \frac{3}{2N} + \frac{1}{2N^2}\Big). \tag{3.75}$$

Lemma 3.30. *Let $H_S := \nabla_\sigma^2 S_N(\vec{\sigma}_\star)$. Then*

$$(H_S)_{ij} = \begin{cases} (d-1)q_\star, & i = j, \\ p_\star, & i \neq j, \end{cases} \tag{3.76}$$

hence

$$\theta_{\mathbf{1}}(S_N) = (d-1)\Big(q_\star + p_\star\Big), \tag{3.77}$$

$$\theta_\perp(S_N) = (d-1)q_\star - p_\star, \tag{3.78}$$

where $\theta_\perp(S_N)$ has multiplicity $d-1$.

Proof. The limits in Lemma 3.19 give

$$(H_S)_{ij} = p_\star \quad (i \neq j), \qquad (H_S)_{ii} = \sum_{k \neq i} q_\star = (d-1)q_\star,$$

which yields (3.76). The eigenvalue formulas follow from the standard spectrum of a matrix with constant diagonal and constant off-diagonal entries. $\square$

Lemma 3.31. *Let*

$$E(\sigma) = g\Big(\sum_{k=1}^{d} f(\sigma_k)\Big)$$

be a spectral energy with $g \in C^2(\mathbb{R})$ *and* $f \in C^2((0,\infty))$, *and let* $H_E := \nabla_\sigma^2 E(\vec{\sigma}_\star)$. *Then*

$$(H_E)_{ii} = h_1, \qquad (H_E)_{ij} = h_2 \quad (i \neq j), \tag{3.79}$$

where

$$h_1 = g''\Big(d\,f(\sigma_\star)\Big)[f'(\sigma_\star)]^2 + g'\Big(d\,f(\sigma_\star)\Big)f''(\sigma_\star), \qquad h_2 = g''\Big(d\,f(\sigma_\star)\Big)[f'(\sigma_\star)]^2. \tag{3.80}$$

Consequently,

$$\theta_{\mathbf{1}}(E) = h_1 + (d-1)h_2, \tag{3.81}$$

$$\theta_\perp(E) = h_1 - h_2, \tag{3.82}$$

where $\theta_\perp(E)$ *has multiplicity* $d-1$.

Proof. Write $H(\sigma) := \sum_{k=1}^{d} f(\sigma_k)$. Then

$$\partial_i E = g'(H)f'(\sigma_i), \qquad \partial_{ij}^2 E = g''(H)f'(\sigma_i)f'(\sigma_j) + g'(H)f''(\sigma_i)\delta_{ij}. \tag{3.83}$$

Evaluating (3.83) at $\vec{\sigma}_\star$ yields (3.79)–(3.80). $\square$

Proof of Theorem 3.8. At $\vec{\sigma}_\star$ one has $\nabla_\sigma F_\beta(\vec{\sigma}_\star) = 0$. Linearizing (3.7) at $\vec{\sigma}_\star$ gives the Jacobian

$$J = -N\,\sigma_\star^{2-2/N}\,\nabla_\sigma^2 F_\beta(\vec{\sigma}_\star) = -N\,\sigma_\star^{2-2/N}\left(H_E - \beta^{-1} H_S\right). \tag{3.84}$$

By Lemmas 3.30 and 3.31, H_E and H_S share the invariant splitting $\mathrm{span}\{\mathbf{1}\} \oplus \mathrm{span}\{\mathbf{1}\}^\perp$ and have eigenvalues (3.81)–(3.82) and (3.77)–(3.78) on the respective subspaces. Substituting into (3.84) yields the claimed eigenvalues. □

Remark 3.32. Substituting the eigenvalues from Lemma 3.30 gives

$$\rho_{\mathbf{1}} = -N\,\sigma_\star^{2-2/N}\,\theta_{\mathbf{1}}(E) - N\,\sigma_\star^{-2/N}\,\frac{d-1}{2\beta}\left(1 - \frac{1}{N}\right), \tag{3.85}$$

$$\rho_\perp = -N\,\sigma_\star^{2-2/N}\,\theta_\perp(E) - N\,\sigma_\star^{-2/N}\,\frac{1}{6\beta}\left(2d - 3 - \frac{3(d-1)}{N} + \frac{d}{N^2}\right). \tag{3.86}$$

The dependence on the energy enters only through $\theta_{\mathbf{1}}(E)$ and $\theta_\perp(E)$.

Remark 3.33. The splitting $\mathbb{R}^d = \mathrm{span}\{\mathbf{1}\} \oplus \mathrm{span}\{\mathbf{1}\}^\perp$ diagonalizes the linearization of the flow at $\vec{\sigma}_\star$. Under the additional hypotheses $g'' \geq 0$ and $f' \geq 0$,

$$\theta_\perp(E) \leq \theta_{\mathbf{1}}(E) \qquad \text{and} \qquad \theta_\perp(S_N) > \theta_{\mathbf{1}}(S_N). \tag{3.87}$$

Hence $\rho_\perp$ is the least negative eigenvalue: perturbations that change the singular values relative to one another decay slowest, whereas the scaling direction $\mathbf{1}$ relaxes faster.

This description on $\mathcal{S}_d$ translates directly to $\mathrm{GL}(d, \mathbb{R})$.

Proof of Corollary 3.10. By Theorem 3.6, the unique minimizer of F_β on $\mathcal{S}_d$ has singular values $(\sigma_\star, \ldots, \sigma_\star)$. Since $F_\beta(X)$ depends only on the singular values of X, every matrix in $\mathrm{GL}(d, \mathbb{R})$ with those singular values is a minimizer, and every such matrix has the form $\sigma_\star Q$ with $Q \in \mathrm{O}(d)$. This is exactly $\mathcal{O}_\star$. □

Proof of Corollary 3.11. Because both F_β and g_N are bi-orthogonally invariant, it is enough to analyze the linearization at $X_\star = \sigma_\star I$ and then transport the result by left multiplication with Q.

At $X_\star = \sigma_\star I$, the orbit $\mathcal{O}_\star = \sigma_\star \mathrm{O}(d)$ has tangent space $\{X_\star A : A^T = -A\}$. Since F_β is constant on $\mathcal{O}_\star$, the linearization vanishes on this subspace.

The orthogonal complement of the skew directions is the space of symmetric matrices. Inside it, the splitting

$$\mathrm{Sym}_d(\mathbb{R}) = \mathrm{span}\{I\} \oplus \{S : S^T = S,\ \mathrm{tr}\, S = 0\}$$

corresponds exactly to $\mathrm{span}\{\mathbf{1}\} \oplus \mathrm{span}\{\mathbf{1}\}^\perp$ under the singular-value map. The linearization on $\mathrm{span}\{I\}$ is therefore $\rho_{\mathbf{1}}$, and the linearization on symmetric trace-free matrices is $\rho_\perp$. Transporting this decomposition back to $X_\star = \sigma_\star Q$ by left multiplication with Q yields (3.10). □

3.5 Schatten energies

We now specialize to the Schatten energies (3.11). They provide the simplest explicit model for the regularized dynamics and lead to an exact scalar reduction on the diagonal.

Introduce the variables

$$\lambda_i = \sigma_i^{1/N}, \qquad i = 1, \dots, d, \qquad \Lambda = \Sigma^{1/N} = \mathrm{diag}(\lambda_1, \dots, \lambda_d),$$

so that $\mathrm{d}\sigma_i = N\lambda_i^{N-1}\,\mathrm{d}\lambda_i$. In these coordinates the metric on the chamber becomes flat:

$$g_{N,\sigma} = N \sum_{i=1}^{d} (\mathrm{d}\lambda_i)^2. \tag{3.88}$$

Lemma 3.34. *Let*

$$\dot\sigma_i = -N\,\sigma_i^{2-2/N}\,\partial_{\sigma_i} F_\beta(\sigma), \qquad i = 1,\dots,d,$$

be the chamber flow, and set $\lambda_i = \sigma_i^{1/N}$. *Then*

$$\dot\lambda_i = -\frac{1}{N}\frac{\partial}{\partial\lambda_i} F_\beta\big(\sigma(\lambda)\big), \qquad i = 1,\dots,d. \tag{3.89}$$

For the Schatten–p energy

$$E_p(\sigma) = \frac{1}{p}\sum_{i=1}^{d}\sigma_i^p,$$

this becomes

$$\dot\lambda_i = -\lambda_i^{Np-1} + \frac{1}{\beta}\sum_{k\neq i}\left(\frac{\lambda_i^{2N-1}}{\lambda_i^{2N}-\lambda_k^{2N}} - \frac{\lambda_i}{N(\lambda_i^2-\lambda_k^2)}\right), \qquad i = 1,\dots,d. \tag{3.90}$$

Proof. By (3.88), $g_{N,\sigma}$ is a constant multiple of the Euclidean metric in λ, so (3.89) follows from the definition of the gradient. Using $\partial_{\lambda_i} = N\lambda_i^{N-1}\partial_{\sigma_i}$ and $\sigma_i = \lambda_i^N$ gives $\dot\lambda_i = -\lambda_i^{N-1}\partial_{\sigma_i}F_\beta$. For (3.11) one has $\partial_{\sigma_i}E = \sigma_i^{p-1}$, and substituting $\partial_{\sigma_i}S_N$ from Lemma 3.18 yields (3.90). □

To separate the common scale from the relative singular values, write

$$\lambda_d = s > 0, \qquad \lambda_i = u_i s, \quad i = 1,\dots,d-1, \qquad u_1 \geq \cdots \geq u_{d-1} \geq 1. \tag{3.91}$$

For $d = 2$, this becomes $\lambda_1 = us$, $\lambda_2 = s$.

Lemma 3.35. *For the Schatten–p flow*

$$\dot\lambda_i = -\lambda_i^{Np-1} + \frac{1}{\beta}\sum_{k\neq i}\left(\frac{\lambda_i^{2N-1}}{\lambda_i^{2N}-\lambda_k^{2N}} - \frac{\lambda_i}{N(\lambda_i^2-\lambda_k^2)}\right), \qquad i = 1,\dots,d,$$

the substitution $\lambda_1 = us$, $\lambda_2 = s$ yields

$$\dot{s} = -s^{Np-1} + \frac{1}{\beta}s^{-1}\left(-\frac{1}{u^{2N}-1} + \frac{1}{N(u^2-1)}\right), \tag{3.92}$$

$$\dot{u} = s^{Np-2}\left(u - u^{Np-1}\right) + \frac{1}{\beta}s^{-2}\left(\frac{u^{2N-1}+u}{u^{2N}-1} - \frac{2u}{N(u^2-1)}\right). \tag{3.93}$$

Proof. Insert $\lambda_1 = us$, $\lambda_2 = s$ into (3.90) and use $\dot{u} = (\dot{\lambda}_1 s - \lambda_1 \dot{s})/s^2$. $\square$

In the general case $d \geq 2$, define

$$\varphi_N(a,b) := \frac{a^{2N-1}}{a^{2N}-b^{2N}} - \frac{a}{N(a^2-b^2)}, \qquad a \neq b > 0. \tag{3.94}$$

Lemma 3.36. *For the Schatten–p flow*

$$\dot{\lambda}_i = -\lambda_i^{Np-1} + \frac{1}{\beta}\sum_{k\neq i}\left(\frac{\lambda_i^{2N-1}}{\lambda_i^{2N}-\lambda_k^{2N}} - \frac{\lambda_i}{N(\lambda_i^2-\lambda_k^2)}\right), \qquad i = 1,\dots,d,$$

the change of variables

$$\lambda_d = s > 0, \qquad \lambda_i = u_i s, \qquad u_1 \geq \dots \geq u_{d-1} \geq 1,$$

is equivalent to

$$\dot{s} = -s^{Np-1} + \frac{1}{\beta}s^{-1}\sum_{j=1}^{d-1}\left(-\frac{1}{u_j^{2N}-1} + \frac{1}{N(u_j^2-1)}\right), \tag{3.95}$$

$$\begin{aligned}\dot{u}_i = s^{Np-2}\left(u_i - u_i^{Np-1}\right) + \frac{1}{\beta}s^{-2}\Bigg(&\sum_{\substack{k=1\\k\neq i}}^{d-1}\varphi_N(u_i,u_k) + \varphi_N(u_i,1)\\ &- u_i\sum_{j=1}^{d-1}\left(-\frac{1}{u_j^{2N}-1} + \frac{1}{N(u_j^2-1)}\right)\Bigg),\end{aligned} \tag{3.96}$$

for $i = 1,\dots,d-1$.

Proof. Use $\dot{s} = \dot{\lambda}_d$ and $\dot{u}_i = (\dot{\lambda}_i s - \lambda_i \dot{s})/s^2$, and simplify the pair terms in (3.90) using (3.91) and (3.94). $\square$

Lemma 3.37. *Let*

$$\varphi_N(a,b) := \frac{a^{2N-1}}{a^{2N} - b^{2N}} - \frac{a}{N(a^2 - b^2)}, \qquad a \neq b > 0.$$

Then

$$\lim_{b \to a} \varphi_N(a,b) = \frac{N-1}{2Na}, \qquad a > 0. \tag{3.97}$$

In particular,

$$\lim_{u \to 1} \left(-\frac{1}{u^{2N} - 1} + \frac{1}{N(u^2 - 1)} \right) = \frac{N-1}{2N}. \tag{3.98}$$

Consequently $u_1 = \cdots = u_{d-1} = 1$ *is an invariant set for* (3.96).

Proof. Write

$$\varphi_N(a,b) = \frac{1}{a} \left(\frac{1}{1 - r^{2N}} - \frac{1}{N} \frac{1}{1 - r^2} \right), \qquad r = b/a,$$

and expand at $r = 1$ to obtain (3.97), hence (3.98). Substituting $u_i \equiv 1$ into (3.96) and using (3.98) gives $\dot{u}_i = 0$. $\square$

Proof of Theorem 3.12. By Lemma 3.37, the set $u_1 = \cdots = u_{d-1} = 1$ is invariant for (3.96). Along this set, (3.95) and (3.98) give

$$\dot{s} = -s^{Np-1} + \frac{1}{\beta} s^{-1} \sum_{j=1}^{d-1} \frac{N-1}{2N} = -s^{\nu - 1} + \beta^{-1} \frac{d-1}{2} \left(1 - \frac{1}{N}\right) \frac{1}{s}, \tag{3.99}$$

where $\nu = Np$. For the Schatten energy, (3.8) reduces to

$$\sigma_\star^p = \beta^{-1} \frac{d-1}{2} \left(1 - \frac{1}{N}\right),$$

hence $s_\star^\nu = \sigma_\star^p$. Therefore (3.99) is exactly (3.15).

If $s_0 = s_\star$, then the right-hand side of (3.15) vanishes and the solution is constant. Suppose $s_0 \neq s_\star$. Since the vector field in (3.15) is smooth on $(0, \infty)$ and has the unique zero $s_\star$, uniqueness prevents the solution from crossing $s_\star$. Thus $s(t)$ remains in the component I_- or I_+ containing s_0.

Separating variables in (3.15) on that component gives

$$t - t_0 = \int_{s_0}^{s(t)} \frac{\tau \, \mathrm{d}\tau}{s_\star^\nu - \tau^\nu}, \qquad s_0 = s(t_0). \tag{3.100}$$

This proves the real quadrature in (3.16).

It remains only to identify the hypergeometric primitive on the chosen real branch. With $z = (s/s_\star)^\nu$ one has $s \, \mathrm{d}s = \frac{s_\star^2}{\nu} z^{\frac{2}{\nu}-1} \, \mathrm{d}z$, so (3.100) becomes

$$t - t_0 = \frac{s_\star^{2-\nu}}{\nu} \int_{z_0}^{z(t)} \frac{z^{\frac{2}{\nu}-1}}{1-z} \, \mathrm{d}z, \qquad z_0 = \left(\frac{s_0}{s_\star}\right)^\nu. \tag{3.101}$$

For $0 < z < 1$, the standard real primitive [Olv+10, §8.17] is

$$\int \frac{z^{a-1}}{1-z} \, \mathrm{d}z = \frac{z^a}{a} \, {}_2F_1(a, 1; a+1; z) + \text{const}, \qquad a > 0. \tag{3.102}$$

Taking $a = \frac{2}{\nu}$ and substituting back $z = (s/s_\star)^\nu$ gives (3.14) on I_-. On I_+, the same antiderivative is understood on the real branch determined by the integral in (3.100), equivalently with the logarithmic term in its analytic continuation read as $\log|1 - z|$. $\square$

Remark 3.38. The same equilibrium may be recovered from the constrained dual problem

$$\max_X \; S_N(X) \qquad \text{subject to} \qquad E_p(X) = 1, \tag{3.103}$$

where

$$E_p(X) = \frac{1}{p}\sum_{i=1}^{d} \sigma_i^p.$$

At a maximizer with all singular values equal, $\vec{\sigma}_\star = (\sigma_\star, \dots, \sigma_\star)$, the constraint gives

$$E_p(X_\star) = \frac{d}{p}\sigma_\star^p = 1, \qquad \sigma_\star^p = \frac{p}{d}.$$

We use the Lagrange-multiplier convention

$$\mathrm{d}S_N(X_\star) = \lambda_\star \, \mathrm{d}E_p(X_\star).$$

At this point,

$$\partial_{\sigma_i} S_N(\vec{\sigma}_\star) = \frac{d-1}{2\sigma_\star}\left(1 - \frac{1}{N}\right), \qquad \partial_{\sigma_i} E_p(\vec{\sigma}_\star) = \sigma_\star^{p-1}.$$

Therefore stationarity gives

$$\frac{d-1}{2\sigma_\star}\left(1 - \frac{1}{N}\right) = \lambda_\star \sigma_\star^{p-1},$$

and hence

$$\lambda_\star = \frac{d-1}{2\sigma_\star^p}\left(1 - \frac{1}{N}\right) = \frac{d(d-1)}{2p}\left(1 - \frac{1}{N}\right) = \frac{1}{p}\binom{d}{2}\left(1 - \frac{1}{N}\right). \tag{3.104}$$

Bibliographical notes

This chapter is based on *Entropic Regularization in the Deep Linear Network* [CKM25]. The background geometry is inherited from Chapter 2. In particular, the operator $\mathcal{A}_{N,X}$, the metric g_N, the Riemannian submersion statement, and the real entropy formula (3.2) are taken from [Men25; MY26b].

Part II

Kähler Geometry of the DLN

Chapter 4

The Complex DLN and the Moment Map

Over $\mathbb{C}$, the conserved quantity $\mathbf{G}$ of Chapter 2 is the moment map μ for the unitary symmetry. We prove that $\mathcal{M} = \mu^{-1}(0)$, that each balanced set of a fibre is a compact unitary orbit, and that the quotient $\mathcal{M}/K$ is biholomorphic to $\mathrm{GL}(d, \mathbb{C})$.

4.1 Main results

We begin with the ambient Kähler geometry on parameter space. Equip $\mathrm{GL}(d, \mathbb{C})^N$ with its standard flat Kähler structure.

$$h(Z, Z') = \sum_{k=1}^{N} \mathrm{tr}(Z_k^* Z_k'),$$

with associated Riemannian metric and symplectic form

$$g = \Re h, \qquad \omega = \Im h,$$

and complex structure J given by multiplication by i, acting componentwise on $\mathrm{M}_d(\mathbb{C})^N$.

Proposition 4.1. *The action of $K = \mathrm{U}(d)^{N-1}$ on $\mathrm{GL}(d,\mathbb{C})^N$ given by*

$$U \cdot \mathbf{W} = (W_N U_{N-1}^{-1},\, U_{N-1} W_{N-1} U_{N-2}^{-1},\, \ldots,\, U_1 W_1) \tag{4.1}$$

is smooth, holomorphic, and preserves the Hermitian form h. In particular, it preserves g, J, and ω. Moreover, $\phi(U \cdot \mathbf{W}) = \phi(\mathbf{W})$ for all $U \in K$ and $\mathbf{W} \in \mathrm{GL}(d,\mathbb{C})^N$.

Hamiltonian symmetry. We now identify the moment map for this action. For $\mathbf{W} \in \mathrm{GL}(d,\mathbb{C})^N$, define

$$G_k(\mathbf{W}) = W_k W_k^* - W_{k+1}^* W_{k+1}, \qquad 1 \le k \le N-1. \tag{4.2}$$

We order the tuple to match the coordinates of $K = \mathrm{U}(d)^{N-1}$, and set $G(\mathbf{W}) = (G_{N-1}(\mathbf{W}), \ldots, G_1(\mathbf{W}))$.

Let $\mathfrak{k} = \mathfrak{u}(d)^{N-1}$. We identify $\mathfrak{k}^*$ with $\mathrm{Herm}(d)^{N-1}$ by the real pairing

$$\langle H, \xi \rangle = -\frac{1}{2} \sum_{k=1}^{N-1} \Im \operatorname{tr}(H_k \xi_k), \qquad H \in \mathrm{Herm}(d)^{N-1}, \quad \xi \in \mathfrak{k}. \tag{4.3}$$

With this identification, define the moment map $\mu : \mathrm{GL}(d,\mathbb{C})^N \to \mathrm{Herm}(d)^{N-1}$ by $\mu(\mathbf{W}) = G(\mathbf{W})$.

Theorem 4.2. *The K-action on $(\mathrm{GL}(d,\mathbb{C})^N, \omega)$ is Hamiltonian, and the map*

$$\mu(\mathbf{W}) = (G_{N-1}(\mathbf{W}), \ldots, G_1(\mathbf{W}))$$

is a K-equivariant moment map. Equivalently, for every $\xi \in \mathfrak{k}$,

$$\iota_{K_\xi} \omega = \mathrm{d}\langle \mu, \xi \rangle,$$

where K_ξ is the fundamental vector field generated by ξ.

Balancedness. The balanced manifold,

$$\mathcal{M} = \{\mathbf{W} \in \mathrm{GL}(d,\mathbb{C})^N : W_k W_k^* = W_{k+1}^* W_{k+1} \text{ for } 1 \le k \le N-1\}, \tag{4.4}$$

is the zero level set of the moment map.

Corollary 4.3. *The balanced manifold satisfies*

$$\mathcal{M} = \mu^{-1}(0),$$

and is K-invariant.

We next establish the regularity needed for the quotient.

Proposition 4.4. *The action of K on $\mathrm{GL}(d,\mathbb{C})^N$ is free. In particular, its restriction to $\mathcal{M}$ is free.*

Proposition 4.5. *The value 0 is regular for μ. Hence $\mathcal{M}$ is a smooth submanifold of real dimension $(N+1)d^2$.*

We next describe the balanced factorizations in each fibre. For $X \in \mathrm{GL}(d,\mathbb{C})$, write its polar decomposition as

$$X = U(X)P(X), \qquad P(X) = (X^*X)^{1/2} > 0, \qquad U(X) = X\,P(X)^{-1} \in \mathrm{U}(d).$$

Define

$$s(X) = \Big(U(X)P(X)^{1/N},\, P(X)^{1/N},\, \ldots,\, P(X)^{1/N}\Big) \in \mathrm{GL}(d,\mathbb{C})^N. \tag{4.5}$$

Also write

$$\mathcal{F}_X = \phi^{-1}(X), \qquad \mathcal{O}_X = \mathcal{F}_X \cap \mathcal{M}. \tag{4.6}$$

The section $s(X)$ is defined by polar decomposition and functional calculus on positive Hermitian matrices, so it is well defined without assuming that the singular values are distinct.

Proposition 4.6. *For every $X \in \mathrm{GL}(d, \mathbb{C})$,*

$$\mathcal{O}_X = K \cdot s(X).$$

In particular, $\mathcal{O}_X$ is nonempty and is a single compact K-orbit.

Since K is compact and acts freely on $\mathcal{M}$, the quotient $M_{\mathrm{red}} = \mathcal{M}/K = \mu^{-1}(0)/K$ is a smooth manifold, and $\dim_{\mathbb{R}} M_{\mathrm{red}} = 2d^2$.

The quotient. We now describe the three quotient constructions. Let

$$\pi : \mathcal{M} \to M_{\mathrm{red}}$$

be the quotient map, and let

$$j : \mathcal{M} \hookrightarrow \mathrm{GL}(d, \mathbb{C})^N$$

be the inclusion.

Theorem 4.7 (Marsden–Weinstein reduction)**.** *There is a unique symplectic form ω_{red} on M_{red} satisfying*

$$\pi^* \omega_{\mathrm{red}} = j^* \omega.$$

For $\mathbf{W} \in \mathcal{M}$, set

$$\mathcal{V}_{\mathbf{W}} = T_{\mathbf{W}}(K \cdot \mathbf{W}), \qquad \mathcal{H}_{\mathbf{W}} = T_{\mathbf{W}}\mathcal{M} \cap \mathcal{V}_{\mathbf{W}}^{\perp_g}.$$

Theorem 4.8 (Riemannian submersion)**.** *There is a unique Riemannian metric g_{red} on M_{red} such that*

$$\pi : (\mathcal{M}, g|_{\mathcal{M}}) \longrightarrow (M_{\mathrm{red}}, g_{\mathrm{red}})$$

is a Riemannian submersion. Equivalently, if $\bar{Y}, \bar{Z} \in T_{[\mathbf{W}]}M_{\mathrm{red}}$ and $Y^{\mathrm{hor}}, Z^{\mathrm{hor}} \in \mathcal{H}_{\mathbf{W}}$ are their horizontal lifts, then

$$g_{\mathrm{red}}(\bar{Y}, \bar{Z}) = g(Y^{\mathrm{hor}}, Z^{\mathrm{hor}}).$$

Theorem 4.9 (Kähler reduction)**.** *The horizontal distribution $\mathcal{H}$ is J-invariant. Consequently M_{red} carries a natural Kähler structure*

$$(g_{\mathrm{red}}, J_{\mathrm{red}}, \omega_{\mathrm{red}}),$$

where ω_{red} is the Marsden–Weinstein reduced symplectic form, g_{red} is the Riemannian quotient metric, and J_{red} is the complex structure induced by the J-invariant horizontal distribution.

The quotient admits a natural identification with the space of end-to-end matrices.

Theorem 4.10. *The end-to-end map restricted to the balanced manifold,*

$$\phi|_{\mathcal{M}} : \mathcal{M} \to \mathrm{GL}(d, \mathbb{C}),$$

descends to a smooth bijection

$$\bar{\phi} : M_{\mathrm{red}} \longrightarrow \mathrm{GL}(d, \mathbb{C}), \qquad \bar{\phi}([\mathbf{W}]) = \phi(\mathbf{W}).$$

Its inverse is

$$\bar{s} : \mathrm{GL}(d, \mathbb{C}) \longrightarrow M_{\mathrm{red}}, \qquad \bar{s}(X) = \pi(s(X)).$$

Hence $\bar{\phi}$ is a diffeomorphism. Moreover, with the reduced complex structure on M_{red} and the standard complex structure on $\mathrm{GL}(d, \mathbb{C})$, the map $\bar{\phi}$ is biholomorphic.

4.2 The ambient Kähler geometry

We equip $\mathrm{GL}(d, \mathbb{C})^N$ with its standard flat Kähler structure, which will be used throughout this section. Since $\mathrm{GL}(d, \mathbb{C})^N$ is an open subset of $\mathrm{M}_d(\mathbb{C})^N$, every tangent space is canonically identified with $\mathrm{M}_d(\mathbb{C})^N$. For tangent vectors $Z = (Z_N, \ldots, Z_1)$ and $Z' = (Z'_N, \ldots, Z'_1)$, set

$$h(Z, Z') = \sum_{k=1}^{N} \mathrm{tr}(Z_k^* Z'_k), \qquad g = \Re h, \qquad \omega = \Im h. \tag{4.7}$$

The complex structure J is multiplication by i in each factor.

In the coordinates W_k on the kth layer, the symplectic form is

$$\omega = \frac{1}{2\mathrm{i}} \sum_{k=1}^{N} \mathrm{tr}\Big(dW_k^* \wedge dW_k\Big). \tag{4.8}$$

Here dW_k denotes the matrix of complex coordinate one-forms, and the trace is applied after taking the wedge product entrywise. Equivalently, for tangent vectors

$Z, Z' \in T_{\mathbf{W}}\mathrm{GL}(d,\mathbb{C})^N \cong \mathrm{M}_d(\mathbb{C})^N$,

$$\omega_{\mathbf{W}}(Z, Z') = \frac{1}{2\mathrm{i}} \sum_{k=1}^{N} \mathrm{tr}\Big(Z_k^* Z_k' - Z_k'^* Z_k\Big) = \sum_{k=1}^{N} \Im\,\mathrm{tr}(Z_k^* Z_k'). \tag{4.9}$$

With these conventions,

$$\omega(Z, Z') = g(JZ, Z'). \tag{4.10}$$

The map ϕ is holomorphic because its entries are polynomial in the matrix entries. Its differential is

$$(\mathrm{d}\phi)_{\mathbf{W}}(Z) = \sum_{k=1}^{N} W_N \cdots W_{k+1}\, Z_k\, W_{k-1} \cdots W_1. \tag{4.11}$$

The formula follows by the Leibniz rule. Since each W_k is invertible, ϕ is a submersion: given $Y \in T_{\phi(\mathbf{W})}\mathrm{GL}(d,\mathbb{C}) \cong \mathrm{M}_d(\mathbb{C})$, take

$$Z_N = Y\,(W_{N-1} \cdots W_1)^{-1}, \qquad Z_k = 0 \quad (1 \le k < N).$$

Then $(\mathrm{d}\phi)_{\mathbf{W}}(Z) = Y$.

Proof of Proposition 4.1. For fixed $U \in K$, the map $\mathbf{W} \mapsto U \cdot \mathbf{W}$ is complex-linear in the layer variables, hence holomorphic. Its smooth dependence on U follows from matrix multiplication and inversion in the unitary group.

We use the conventions $U_N = U_0 = I$. The differential of the action sends tangent vectors $Z = (Z_N, \dots, Z_1)$ and $Z' = (Z_N', \dots, Z_1')$ to

$$(U \cdot Z)_k = U_k Z_k U_{k-1}^{-1}, \qquad (U \cdot Z')_k = U_k Z_k' U_{k-1}^{-1}.$$

Since U_k and U_{k-1} are unitary, left and right multiplication by these matrices preserve the Frobenius Hermitian inner product. Hence

$$\begin{aligned} h(U\cdot Z, U\cdot Z') &= \sum_{k=1}^{N} \operatorname{tr}\Big((U_k Z_k U_{k-1}^{-1})^*(U_k Z_k' U_{k-1}^{-1})\Big) \\ &= \sum_{k=1}^{N} \operatorname{tr}\Big(U_{k-1} Z_k^* Z_k' U_{k-1}^{-1}\Big) \\ &= \sum_{k=1}^{N} \operatorname{tr}(Z_k^* Z_k') = h(Z, Z'). \end{aligned}$$

Thus the action preserves h. Taking real and imaginary parts shows that it preserves g and ω. Since the action is complex-linear, it also preserves J.

It remains to check that the action preserves the map ϕ. By (4.1),

$$\phi(U\cdot \mathbf{W}) = (W_N U_{N-1}^{-1})(U_{N-1} W_{N-1} U_{N-2}^{-1})\cdots(U_1 W_1).$$

All intermediate unitary factors cancel, so $\phi(U\cdot \mathbf{W}) = W_N \cdots W_1 = \phi(\mathbf{W})$. □

4.3 The moment map and the balanced manifold

We now prove that the action of K is Hamiltonian, with moment map $\mu = \mathbf{G}$.

Let $\mathfrak{k} = \mathfrak{u}(d)^{N-1}$. For $\xi = (\xi_{N-1}, \ldots, \xi_1) \in \mathfrak{k}$, set $\xi_0 = \xi_N = 0$.

Lemma 4.11. *The fundamental vector field generated by $\xi \in \mathfrak{k}$ is*

$$(K_\xi(W))_k = \xi_k W_k - W_k \xi_{k-1}, \qquad 1 \le k \le N.$$

Proof. Differentiate the curve $t \mapsto \exp(t\xi)\cdot W$ at $t = 0$. By the group action (4.1),

$$(\exp(t\xi)\cdot W)_k = \exp(t\xi_k)\, W_k\, \exp(-t\xi_{k-1}).$$

Therefore

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}(\exp(t\xi)\cdot W)_k = \xi_k W_k - W_k\xi_{k-1}. \qquad \square$$

Proof of Theorem 4.2. The only point requiring care is the sign convention in the pairing (4.3). With this normalization, left multiplication on a matrix factor has moment map $W \mapsto WW^*$, while right multiplication has moment map $W \mapsto -W^*W$.

We verify the identity

$$\iota_{K_\xi}\omega = \mathrm{d}\langle\mu,\xi\rangle \qquad (\xi \in \mathfrak{k})$$

and then prove equivariance.

First consider a single factor $\mathrm{M}_d(\mathbb{C})$, with

$$\omega(U,V) = \Im\,\mathrm{tr}(U^*V).$$

Fix $\xi \in \mathfrak{u}(d)$. For $W \in \mathrm{M}_d(\mathbb{C})$ and $Z \in T_W\mathrm{M}_d(\mathbb{C}) \cong \mathrm{M}_d(\mathbb{C})$,

$$\begin{aligned}
\mathrm{d}\Big(-\frac{1}{2}\Im\,\mathrm{tr}(WW^*\xi)\Big)_W(Z) &= -\frac{1}{2}\Im\,\mathrm{tr}(ZW^*\xi + WZ^*\xi)\\
&= -\Im\,\mathrm{tr}(ZW^*\xi)\\
&= -\Im\,\mathrm{tr}(W^*\xi Z)\\
&= \Im\,\mathrm{tr}((\xi W)^*Z)\\
&= \omega(\xi W, Z).
\end{aligned}$$

Here we used $\xi^* = -\xi$, which gives

$$\mathrm{tr}(WZ^*\xi) = -\overline{\mathrm{tr}(ZW^*\xi)}.$$

Likewise,

$$\begin{aligned}
\mathrm{d}\Big(\frac{1}{2}\Im\operatorname{tr}(W^*W\xi)\Big)_W(Z) &= \frac{1}{2}\Im\operatorname{tr}(Z^*W\xi + W^*Z\xi) \\
&= \Im\operatorname{tr}(W^*Z\xi) \\
&= \Im\operatorname{tr}((-W\xi)^*Z) \\
&= \omega(-W\xi, Z),
\end{aligned}$$

because

$$\operatorname{tr}(Z^*W\xi) = -\overline{\operatorname{tr}(W^*Z\xi)}.$$

Thus, with the pairing (4.3), left multiplication by $\mathrm{U}(d)$ has moment map $W \mapsto WW^*$, and right multiplication has moment map $W \mapsto -W^*W$.

We now apply these single-factor computations to $\mathrm{GL}(d,\mathbb{C})^N$. By Lemma 4.11, the component ξ_k acts on the kth factor by

$$W_k \longmapsto \xi_k W_k$$

and on the $(k+1)$st factor by

$$W_{k+1} \longmapsto -W_{k+1}\xi_k.$$

Hence

$$\begin{aligned}
\mathrm{d}\langle \mu, \xi\rangle &= \sum_{k=1}^{N-1} \mathrm{d}\Big(-\frac{1}{2}\Im\operatorname{tr}(W_k W_k^* \xi_k) + \frac{1}{2}\Im\operatorname{tr}(W_{k+1}^* W_{k+1}\xi_k)\Big) \\
&= \iota_{K_\xi}\omega.
\end{aligned}$$

Thus μ is a moment map.

It remains to check equivariance. Let $U \in K$, and set $U_0 = U_N = I$. Since

$$(U \cdot W)_k = U_k W_k U_{k-1}^{-1},$$

we have, for $1 \leq k \leq N-1$,

$$\begin{aligned}
\Big(\mu(U \cdot W)\Big)_k &= (U_k W_k U_{k-1}^{-1})(U_k W_k U_{k-1}^{-1})^* - (U_{k+1} W_{k+1} U_k^{-1})^* (U_{k+1} W_{k+1} U_k^{-1}) \\
&= U_k W_k W_k^* U_k^* - U_k W_{k+1}^* W_{k+1} U_k^* \\
&= U_k \Big(W_k W_k^* - W_{k+1}^* W_{k+1} \Big) U_k^* \\
&= U_k G_k(W) U_k^*.
\end{aligned}$$

Under the identification (4.3), this is the coadjoint action of K on $\mathfrak{k}^*$. Therefore μ is K-equivariant. $\square$

Proof of Corollary 4.3. By definition,

$$\mu(W) = 0$$

if and only if

$$G_k(W) = W_k W_k^* - W_{k+1}^* W_{k+1} = 0, \qquad 1 \leq k \leq N-1.$$

These are exactly the equations for balancedness. Hence

$$\mathcal{M} = \mu^{-1}(0).$$

The K-invariance follows from the equivariance of μ, since the coadjoint action fixes 0. $\square$

Proof of Proposition 4.4. Suppose $U \cdot W = W$. From the first layer we get

$$U_1 W_1 = W_1.$$

Since W_1 is invertible, $U_1 = I$. Now assume $U_{k-1} = I$. The equality of the kth layers gives

$$U_k W_k U_{k-1}^{-1} = W_k.$$

Since $U_{k-1} = I$ and W_k is invertible, it follows that $U_k = I$. By induction,

$$U_1 = \cdots = U_{N-1} = I.$$

Thus $U = I$, so the action is free on $\mathrm{GL}(d, \mathbb{C})^N$. Its restriction to $\mathcal{M}$ is therefore free as well. $\square$

Proof of Proposition 4.5. We prove the stronger statement that $(\mathrm{d}\mu)_W$ is surjective for every $W \in \mathrm{GL}(d, \mathbb{C})^N$.

Fix $W \in \mathrm{GL}(d, \mathbb{C})^N$, and let $\xi \in \mathfrak{k}$ annihilate the image of $(\mathrm{d}\mu)_W$ under the pairing (4.3). Then

$$\mathrm{d}\langle \mu, \xi \rangle_W = 0.$$

By Theorem 4.2,

$$\omega_W(K_\xi(W), Z) = 0 \qquad \text{for all } Z \in T_W \mathrm{GL}(d, \mathbb{C})^N.$$

Since ω is nondegenerate, $K_\xi(W) = 0$.

The kernel of the infinitesimal action at W is the Lie algebra of the stabilizer of W. Proposition 4.4 shows that this stabilizer is trivial, so its Lie algebra is zero. Hence $\xi = 0$. Therefore the annihilator of $\mathrm{im}(\mathrm{d}\mu)_W$ is trivial. Since $\mathfrak{k}^*$ is finite dimensional,

$(\mathrm{d}\mu)_W$ is surjective.

Thus 0 is a regular value of μ. By the regular-value theorem,

$$\dim_{\mathbb{R}} \mathcal{M} = \dim_{\mathbb{R}} \mathrm{GL}(d, \mathbb{C})^N - \dim_{\mathbb{R}} \mathrm{Herm}(d)^{N-1}.$$

Since

$$\dim_{\mathbb{R}} \mathrm{GL}(d, \mathbb{C})^N = 2Nd^2, \qquad \dim_{\mathbb{R}} \mathrm{Herm}(d)^{N-1} = (N-1)d^2,$$

we obtain

$$\dim_{\mathbb{R}} \mathcal{M} = 2Nd^2 - (N-1)d^2 = (N+1)d^2.$$

□

4.4 Balanced factorizations in a fibre

The condition $\mu(\mathbf{W}) = 0$ defines $\mathcal{M}$. To identify the quotient $\mathcal{M}/K$, we must understand how $\mathcal{M}$ meets each fibre of ϕ. In each fibre, the intersection is a single K-orbit.

We now prove Proposition 4.6. The key point is that balancedness aligns the singular-value data across the layers. Once this alignment is fixed, the only remaining freedom is the choice of unitary bases at the intermediate layers.

The polar representative used in Proposition 4.6 depends smoothly on X. Indeed, the map $X \mapsto X^*X$ is smooth from $\mathrm{GL}(d, \mathbb{C})$ to the cone of positive Hermitian matrices, and functional calculus gives smooth maps

$$H \longmapsto H^{1/2}, \qquad H \longmapsto H^{1/N}$$

on that cone. Hence $P(X)$, $P(X)^{1/N}$, and $U(X) = XP(X)^{-1}$ are smooth functions of

X.

Proof of Proposition 4.6. Fix $X \in \mathrm{GL}(d, \mathbb{C})$, and write its polar decomposition as

$$X = U_X P_X, \qquad P_X = (X^* X)^{1/2} > 0, \qquad U_X = X P_X^{-1} \in \mathrm{U}(d).$$

For the factors of $s(X)$, we have

$$W_N = U_X P_X^{1/N}, \qquad W_k = P_X^{1/N} \quad \text{for } 1 \leq k \leq N-1.$$

Thus

$$\phi(s(X)) = U_X P_X^{1/N} P_X^{(N-1)/N} = U_X P_X = X.$$

Moreover, for $1 \leq k \leq N-2$,

$$W_k W_k^* = P_X^{2/N} = W_{k+1}^* W_{k+1},$$

and at the last interface,

$$W_{N-1} W_{N-1}^* = P_X^{2/N} = P_X^{1/N} U_X^* U_X P_X^{1/N} = W_N^* W_N.$$

Hence $s(X) \in \mathcal{F}_X \cap \mathcal{M} = \mathcal{O}_X$.

Since the K-action preserves ϕ by Proposition 4.1 and preserves $\mathcal{M}$ by Corollary 4.3, it follows that

$$K \cdot s(X) \subseteq \mathcal{O}_X.$$

We prove the opposite inclusion. Let $\mathbf{W} = (W_N, \ldots, W_1) \in \mathcal{O}_X$. Choose a singular value decomposition of the first layer,

$$W_1 = Q_1 \Lambda Q_0^*,$$

where $Q_0, Q_1 \in \mathrm{U}(d)$ and

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d), \qquad \lambda_j > 0.$$

We first align the remaining layers with the same positive diagonal matrix Λ. Suppose that, for some $1 \le k \le N-1$,

$$W_k = Q_k \Lambda Q_{k-1}^*$$

with $Q_{k-1}, Q_k \in \mathrm{U}(d)$. Then

$$W_k W_k^* = Q_k \Lambda^2 Q_k^*.$$

Balancedness gives

$$W_{k+1}^* W_{k+1} = W_k W_k^* = Q_k \Lambda^2 Q_k^*.$$

Define

$$Q_{k+1} := W_{k+1} Q_k \Lambda^{-1}.$$

Then

$$Q_{k+1}^* Q_{k+1} = \Lambda^{-1} Q_k^* W_{k+1}^* W_{k+1} Q_k \Lambda^{-1} = I,$$

so $Q_{k+1} \in \mathrm{U}(d)$, and therefore

$$W_{k+1} = Q_{k+1} \Lambda Q_k^*.$$

Induction gives unitary matrices $Q_0, \ldots, Q_N$ such that

$$W_k = Q_k \Lambda Q_{k-1}^*, \qquad 1 \leq k \leq N.$$

Multiplying the layers yields

$$X = \phi(\mathbf{W}) = Q_N \Lambda^N Q_0^*.$$

Consequently,

$$X^* X = Q_0 \Lambda^{2N} Q_0^*, \qquad P_X = Q_0 \Lambda^N Q_0^*, \qquad P_X^{1/N} = Q_0 \Lambda Q_0^*,$$

and

$$U_X = X P_X^{-1} = Q_N Q_0^*.$$

Now define $V = (V_{N-1}, \ldots, V_1) \in K$ by

$$V_k := Q_0 Q_k^*, \qquad 1 \leq k \leq N-1.$$

Then the last layer transforms as

$$(V \cdot \mathbf{W})_N = W_N V_{N-1}^{-1} = Q_N \Lambda Q_0^* = U_X P_X^{1/N}.$$

For $2 \leq k \leq N-1$,

$$(V \cdot \mathbf{W})_k = V_k W_k V_{k-1}^{-1} = Q_0 \Lambda Q_0^* = P_X^{1/N},$$

and for the first layer,

$$(V \cdot \mathbf{W})_1 = V_1 W_1 = Q_0 \Lambda Q_0^* = P_X^{1/N}.$$

Thus

$$V \cdot \mathbf{W} = s(X).$$

Hence $\mathbf{W} = V^{-1} \cdot s(X)$, so

$$\mathcal{O}_X \subseteq K \cdot s(X).$$

Together with the first inclusion, this proves

$$\mathcal{O}_X = K \cdot s(X).$$

Since K is compact, $\mathcal{O}_X$ is a compact K-orbit. □

Remark 4.12. The set $\mathcal{O}_X$ is the balanced part of the fibre defined in Chapter 2. Proposition 4.6 gives a representative of this orbit that varies smoothly with X.

4.5 The geometry of the quotient

We now combine Theorem 4.2, equivalently (4.4), with Proposition 4.6. The first identifies $\mathcal{M}$ with $\mu^{-1}(0)$, and the second shows that each fibre meets $\mathcal{M}$ in exactly one compact K-orbit. Thus quotienting the balanced manifold by K removes precisely the ambiguity left by balanced factorizations. What remains is the end-to-end matrix.

Set $M_{\mathrm{red}} := \mathcal{M}/K = \mu^{-1}(0)/K$. Let

$$\pi : \mathcal{M} \to M_{\mathrm{red}}, \qquad j : \mathcal{M} \hookrightarrow \mathrm{GL}(d, \mathbb{C})^N \tag{4.12}$$

be the quotient map and the inclusion. For $\mathbf{W} \in \mathcal{M}$, write

$$\mathcal{V}_{\mathbf{W}} := T_{\mathbf{W}}(K \cdot \mathbf{W}), \qquad \mathcal{H}_{\mathbf{W}} := T_{\mathbf{W}}\mathcal{M} \cap \mathcal{V}_{\mathbf{W}}^{\perp_g}. \tag{4.13}$$

The three reductions in Theorems 4.7, 4.8, and 4.9 descend different pieces of the same ambient Kähler structure. The Marsden–Weinstein reduction descends the symplectic form. The Riemannian quotient descends the metric through horizontal lifts. The Kähler reduction says that these two descended structures remain compatible because the horizontal distribution is J-invariant. Finally, Theorem 4.10 identifies the quotient with $\mathrm{GL}(d, \mathbb{C})$ by the end-to-end map.

Proof of Theorem 4.7. By Proposition 4.5, the value 0 is regular for μ. Hence

$$\mathcal{M} = \mu^{-1}(0)$$

is a smooth submanifold. By Proposition 4.4, the action of K on the ambient space is free, so its restriction to $\mathcal{M}$ is free. Since K is compact, this restricted action is proper. Therefore $M_{\mathrm{red}} = \mathcal{M}/K$ is a smooth manifold, and $\pi : \mathcal{M} \to M_{\mathrm{red}}$ is a smooth quotient map.

We next descend the symplectic form. For $\xi \in \mathfrak{k}$, Theorem 4.2 gives

$$\iota_{K_\xi}(j^*\omega) = j^*(\iota_{K_\xi}\omega) = j^*\mathrm{d}\langle \mu, \xi \rangle = 0,$$

because $\mu = 0$ on $\mathcal{M}$. The K-action preserves ω, so $j^*\omega$ is also K-invariant. Hence $j^*\omega$ is basic. Since π is the quotient map, there is a unique two-form ω_{red} on M_{red} satisfying

$$\pi^*\omega_{\mathrm{red}} = j^*\omega. \tag{4.14}$$

Moreover, $j^*\omega$ is closed. Since π is a surjective submersion, (4.14) implies that ω_{red} is closed.

It remains to prove nondegeneracy. At $\mathbf{W} \in \mathcal{M}$, the moment-map identity gives

$$T_{\mathbf{W}}\mathcal{M} = \{Y : \omega(K_\xi(\mathbf{W}), Y) = 0 \text{ for all } \xi \in \mathfrak{k}\} = \mathcal{V}_{\mathbf{W}}^{\omega},$$

where $\mathcal{V}_{\mathbf{W}}^{\omega}$ denotes the symplectic orthogonal of $\mathcal{V}_{\mathbf{W}}$. Hence

$$(T_{\mathbf{W}}\mathcal{M})^{\omega} = (\mathcal{V}_{\mathbf{W}}^{\omega})^{\omega} = \mathcal{V}_{\mathbf{W}}.$$

Let $\bar{Y} \in T_{[\mathbf{W}]}M_{\mathrm{red}}$ lie in the kernel of ω_{red}, and choose a lift $Y \in T_{\mathbf{W}}\mathcal{M}$. Then

$$\omega(Y, Z) = 0 \qquad \text{for all } Z \in T_{\mathbf{W}}\mathcal{M}.$$

Thus

$$Y \in (T_{\mathbf{W}}\mathcal{M})^{\omega} = \mathcal{V}_{\mathbf{W}}.$$

Therefore $(\mathrm{d}\pi)_{\mathbf{W}}Y = 0$, and hence $\bar{Y} = 0$. This proves that ω_{red} is nondegenerate. $\square$

Proof of Theorem 4.8. The action of K preserves g by Proposition 4.1. Since the action on $\mathcal{M}$ is free and proper, the quotient map π is a submersion. The g-orthogonal complement of the vertical distribution therefore gives the smooth horizontal distribution

$$\mathcal{H}_{\mathbf{W}} = T_{\mathbf{W}}\mathcal{M} \cap \mathcal{V}_{\mathbf{W}}^{\perp_g}.$$

For $\bar{Y}, \bar{Z} \in T_{[\mathbf{W}]}M_{\mathrm{red}}$, let $Y^{\mathrm{hor}}, Z^{\mathrm{hor}} \in \mathcal{H}_{\mathbf{W}}$ be the unique horizontal lifts. Define

$$g_{\mathrm{red}}(\bar{Y}, \bar{Z}) := g(Y^{\mathrm{hor}}, Z^{\mathrm{hor}}). \tag{4.15}$$

The definition is independent of the representative $\mathbf{W}$ because the K-action preserves g and carries horizontal spaces to horizontal spaces. It is smooth because the horizontal distribution is smooth. It is positive definite because a nonzero tangent vector downstairs has a nonzero horizontal lift. Formula (4.15) is forced by the Riemannian submersion condition, so the metric is unique with this property. Thus

$$\pi : (\mathcal{M}, g|_{\mathcal{M}}) \to (M_{\mathrm{red}}, g_{\mathrm{red}})$$

is a Riemannian submersion. □

Proof of Theorem 4.9. We first prove the J-invariance of the horizontal distribution. Let $\xi \in \mathfrak{k}$ and $Y \in T_{\mathbf{W}}\mathcal{M}$. Since Y is tangent to $\mu^{-1}(0)$, Theorem 4.2 gives

$$0 = \mathrm{d}\langle \mu, \xi \rangle(Y) = \omega(K_\xi(\mathbf{W}), Y) = g(JK_\xi(\mathbf{W}), Y).$$

Thus

$$JK_\xi(\mathbf{W}) \in (T_{\mathbf{W}}\mathcal{M})^{\perp_g},$$

and therefore

$$J\mathcal{V}_{\mathbf{W}} \subset (T_{\mathbf{W}}\mathcal{M})^{\perp_g}. \tag{4.16}$$

Now let $H \in \mathcal{H}_{\mathbf{W}}$. For every $\xi \in \mathfrak{k}$,

$$\mathrm{d}\langle \mu, \xi \rangle(JH) = \omega(K_\xi(\mathbf{W}), JH) = g(JK_\xi(\mathbf{W}), JH) = g(K_\xi(\mathbf{W}), H) = 0,$$

because H is horizontal. Hence $JH \in T_{\mathbf{W}}\mathcal{M}$. Moreover,

$$g(JH, K_\xi(\mathbf{W})) = -g(H, JK_\xi(\mathbf{W})) = 0$$

by (4.16), since $H \in T_{\mathbf{W}}\mathcal{M}$. Thus JH is both tangent to $\mathcal{M}$ and orthogonal to $\mathcal{V}_{\mathbf{W}}$. Hence

$$J\mathcal{H}_{\mathbf{W}} \subset \mathcal{H}_{\mathbf{W}}.$$

Since $J^2 = -I$, equality holds, so $\mathcal{H}_{\mathbf{W}}$ is J-invariant.

The J-invariant horizontal distribution defines an almost complex structure downstairs by

$$J_{\mathrm{red}}\bar{Y} := (\mathrm{d}\pi)_{\mathbf{W}}(JY^{\mathrm{hor}}), \tag{4.17}$$

where $Y^{\mathrm{hor}} \in \mathcal{H}_{\mathbf{W}}$ is the horizontal lift of $\bar{Y}$. The definition is independent of the representative because the K-action is holomorphic and preserves the horizontal distribution. Since $J^2 = -I$, one has $J_{\mathrm{red}}^2 = -I$.

Let $\bar{Y}, \bar{Z} \in T_{[\mathbf{W}]}M_{\mathrm{red}}$, and let $Y^{\mathrm{hor}}, Z^{\mathrm{hor}}$ be their horizontal lifts. From (4.14),

$$\omega_{\mathrm{red}}(\bar{Y}, \bar{Z}) = \omega(Y^{\mathrm{hor}}, Z^{\mathrm{hor}}).$$

Using the convention $\omega(\cdot,\cdot) = g(J\cdot,\cdot)$, we obtain

$$\omega_{\mathrm{red}}(\bar{Y}, \bar{Z}) = g(JY^{\mathrm{hor}}, Z^{\mathrm{hor}}) = g_{\mathrm{red}}(J_{\mathrm{red}}\bar{Y}, \bar{Z}).$$

Equivalently,

$$\omega_{\mathrm{red}}(\bar{Y}, J_{\mathrm{red}}\bar{Z}) = g_{\mathrm{red}}(\bar{Y}, \bar{Z}). \tag{4.18}$$

In particular,

$$\omega_{\mathrm{red}}(\bar{Y}, J_{\mathrm{red}}\bar{Y}) = g_{\mathrm{red}}(\bar{Y}, \bar{Y}) > 0$$

for $\bar{Y} \neq 0$. The standard Kähler reduction theorem now implies that J_{red} is integrable and that

$$(M_{\mathrm{red}}, g_{\mathrm{red}}, J_{\mathrm{red}}, \omega_{\mathrm{red}})$$

is Kähler; see, for example, [MS17, Chapter 5]. $\square$

Proof of Theorem 4.10. By Proposition 4.1, the restriction $\phi|_{\mathcal{M}}$ is K-invariant. Hence it factors through a smooth map

$$\bar{\phi} : M_{\mathrm{red}} \to \mathrm{GL}(d, \mathbb{C}), \qquad \bar{\phi} \circ \pi = \phi|_{\mathcal{M}}.$$

The polar-decomposition formula for s shows that s is smooth, and therefore

$$\bar{s} := \pi \circ s : \mathrm{GL}(d, \mathbb{C}) \to M_{\mathrm{red}}$$

is smooth. For $X \in \mathrm{GL}(d, \mathbb{C})$,

$$\bar{\phi}(\bar{s}(X)) = \bar{\phi}(\pi(s(X))) = \phi(s(X)) = X.$$

Thus

$$\bar{\phi} \circ \bar{s} = \mathrm{id}_{\mathrm{GL}(d,\mathbb{C})}. \tag{4.19}$$

Conversely, let $[\mathbf{W}] \in M_{\mathrm{red}}$, and set

$$X := \bar{\phi}([\mathbf{W}]).$$

Then $\mathbf{W} \in \phi^{-1}(X) \cap \mathcal{M} = \mathcal{O}_X$. By Proposition 4.6,

$$\mathcal{O}_X = K \cdot s(X).$$

Hence $[\mathbf{W}] = [s(X)]$, and therefore

$$[\mathbf{W}] = \bar{s}(X) = \bar{s}(\bar{\phi}([\mathbf{W}])).$$

Thus

$$\bar{s} \circ \bar{\phi} = \mathrm{id}_{M_{\mathrm{red}}}. \tag{4.20}$$

Equations (4.19) and (4.20) show that $\bar{\phi}$ is a diffeomorphism with inverse $\bar{s}$.

It remains to identify the complex structure. Let J_{std} denote the standard complex structure on $\mathrm{GL}(d, \mathbb{C})$, namely multiplication by i on tangent matrices. Let

$$\bar{Y} \in T_{[\mathbf{W}]} M_{\mathrm{red}},$$

and choose its horizontal lift $Y^{\mathrm{hor}} \in \mathcal{H}_{\mathbf{W}}$, so that

$$\bar{Y} = (\mathrm{d}\pi)_{\mathbf{W}} Y^{\mathrm{hor}}.$$

Then

$$\begin{aligned} (\mathrm{d}\bar{\phi})_{[\mathbf{W}]}(J_{\mathrm{red}} \bar{Y}) &= (\mathrm{d}\bar{\phi})_{[\mathbf{W}]}(\mathrm{d}\pi)_{\mathbf{W}}(J Y^{\mathrm{hor}}) \\ &= (\mathrm{d}\phi)_{\mathbf{W}}(J Y^{\mathrm{hor}}) \\ &= J_{\mathrm{std}}(\mathrm{d}\phi)_{\mathbf{W}}(Y^{\mathrm{hor}}) \\ &= J_{\mathrm{std}}(\mathrm{d}\bar{\phi})_{[\mathbf{W}]} \bar{Y}, \end{aligned}$$

because ϕ is holomorphic. Thus $\bar{\phi}$ is holomorphic. Since $\bar{\phi}$ is a diffeomorphism and its differential is complex-linear everywhere, the differential of $\bar{s} = \bar{\phi}^{-1}$ is also complex-linear. Hence $\bar{s}$ is holomorphic, and $\bar{\phi}$ is biholomorphic. □

Remark 4.13. The biholomorphism in Theorem 4.10 identifies the reduced complex structure on M_{red} with the standard complex structure on $\mathrm{GL}(d, \mathbb{C})$. Thus the operators ∂ and $\bar{\partial}$ used on $\mathrm{GL}(d, \mathbb{C})$ in Chapter 5 are the standard ones. The quotient contributes the Kähler metric together with its associated symplectic form. It does not introduce a new complex structure on $\mathrm{GL}(d, \mathbb{C})$.

Bibliographical notes

The general moment-map and Kähler-reduction framework used in this chapter is classical. For the original symplectic reduction construction, see [MW74]. For background on moment maps, symplectic reduction, and Kähler reduction, see [GS84] and [MS17, Chapter 5]. In this framework, balancedness is the zero level set of the moment map. A variational characterization of balancedness on each fibre is given in [LM25].

Chapter 5

Kähler Potential, Metric, and Entropy

Having identified $\mathrm{GL}(d,\mathbb{C})$ with the quotient $\mathcal{M}/K$, we now compute the quotient metric, its Kähler potential, the entropy, and the Ricci form. The calculations make use of an adapted basis: horizontal directions give the quotient metric, while vertical directions give the volume of the orbit.

Compared with the argument of [MY26b] for the real DLN, we reverse the order of proof. In that setting, one first constructs the adapted basis and then uses it to prove the submersion theorem. In Chapter 4 we identified the quotient by Kähler reduction. We first recover the quotient metric and its Kähler potential by constructing explicit horizontal lifts. We then return to the full basis on $\mathcal{M}$, where the vertical directions determine the volume of the orbit $\mathcal{O}_X$ and hence the entropy.

5.1 Main results

Throughout this chapter, $K = \mathrm{U}(d)^{N-1}$, $S_N(X) := \log \mathrm{vol}(\mathcal{O}_X)$, where the volume of $\mathcal{O}_X$ is computed with the metric induced from $\mathcal{M}$, and $u_d := \mathrm{vol}(\mathrm{U}(d))$ denotes the volume of $\mathrm{U}(d)$ with respect to the bi-invariant metric induced by the Frobenius inner product on $\mathfrak{u}(d)$.

The quotient metric. We transport the Kähler structure on $\mathcal{M}/K$ from Chapter 4 to $\mathrm{GL}(d,\mathbb{C})$ via the map

$$\bar{\phi} : \mathcal{M}/K \to \mathrm{GL}(d,\mathbb{C}).$$

This defines a Hermitian metric on $\mathrm{GL}(d,\mathbb{C})$.

For $X \in \mathrm{GL}(d,\mathbb{C})$ define

$$\mathcal{A}_{N,X}(Z) := \sum_{k=1}^{N} (XX^*)^{\frac{N-k}{N}} Z (X^*X)^{\frac{k-1}{N}}, \qquad Z \in \mathrm{M}_d(\mathbb{C}). \tag{5.1}$$

Theorem 5.1. *The Hermitian metric on* $\mathrm{GL}(d,\mathbb{C})$ *transported from the quotient* $\mathcal{M}/K$ *is*

$$h_N(X)(Z_1, Z_2) = \mathrm{tr}\Big(Z_1^* \mathcal{A}_{N,X}^{-1}(Z_2)\Big), \qquad Z_1, Z_2 \in T_X\mathrm{GL}(d,\mathbb{C}) \cong \mathrm{M}_d(\mathbb{C}). \tag{5.2}$$

Its real and imaginary parts define

$$g_N = \Re h_N, \qquad \omega_N = \Im h_N,$$

where g_N *is the Riemannian metric and* ω_N *is the associated Kähler form.*

The Kähler potential. The quotient metric is Kähler for the standard complex structure on $\mathrm{GL}(d,\mathbb{C})$ and admits a global potential that depends only on X^*X.

Theorem 5.2. *The function*

$$\Psi_N(X) = \frac{N}{2}\,\mathrm{tr}\Big((X^*X)^{1/N}\Big) \tag{5.3}$$

is a Kähler potential for the Kähler form ω_N, namely $\omega_N = \mathrm{i}\,\partial\bar{\partial}\Psi_N$.

The function Ψ_N also admits a variational characterization on the fibres of the map ϕ. On the parameter space $\mathrm{GL}(d,\mathbb{C})^N$, define the ridge functional

$$\mathcal{R}(\mathbf{W}) := \frac{1}{2}\sum_{k=1}^{N}\|W_k\|^2, \qquad \mathbf{W} = (W_N, \dots, W_1). \tag{5.4}$$

This functional is a Kähler potential for the ambient metric on $\mathrm{GL}(d,\mathbb{C})^N$. The quotient potential Ψ_N is obtained by minimizing $\mathcal{R}$ along each fibre $\mathcal{F}_X$.

Proposition 5.3. *For every $X \in \mathrm{GL}(d,\mathbb{C})$,*

$$\Psi_N(X) = \min_{\mathbf{W}\in\mathcal{F}_X} \mathcal{R}(\mathbf{W}). \tag{5.5}$$

Moreover, the minimum is attained precisely on the set of balanced factorizations in $\mathcal{F}_X$:

$$\operatorname*{arg\,min}_{\mathbf{W}\in\mathcal{F}_X} \mathcal{R}(\mathbf{W}) = \mathcal{O}_X. \tag{5.6}$$

Riemannian submersion. The quotient metric on $\mathrm{GL}(d,\mathbb{C})$ is induced from the metric on the balanced manifold $\mathcal{M}$ through the map ϕ. The key tool in the chapter is an adapted orthonormal basis on $\mathcal{M}$ that separates directions tangent to the orbit $\mathcal{O}_X$ from directions that project to $T_X\mathrm{GL}(d,\mathbb{C})$. In this basis, horizontal vectors determine the quotient metric, while vertical vectors control the volume of $\mathcal{O}_X$ and hence the entropy.

Theorem 5.4. *Let ι be the metric on $\mathcal{M}$ induced from the ambient Frobenius metric on $\mathrm{GL}(d,\mathbb{C})^N$. Then the map*

$$\phi : (\mathcal{M}, \iota) \longrightarrow (\mathrm{GL}(d,\mathbb{C}), g_N)$$

is a Riemannian submersion.

Entropy. The entropy is the log-volume of the compact orbit $\mathcal{O}_X$ with the metric induced from $\mathcal{M}$. It depends only on the singular values of X.

Theorem 5.5 (Entropy formula)**.** *Let $X \in \mathrm{GL}(d,\mathbb{C})$ have singular value decomposition*

$$X = Q\Sigma R^*, \qquad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d),$$

with distinct singular values. Then

$$\mathrm{vol}(\mathcal{O}_X) = u_d^{N-1} N^{d/2} \det(\Sigma^2)^{\frac{N-1}{2N}} \frac{\mathrm{van}(\Sigma^2)}{\mathrm{van}(\Sigma^{2/N})}. \tag{5.7}$$

Equivalently,

$$S_N(X) = (N-1)\log u_d + \frac{d}{2}\log N + \frac{N-1}{2N}\log\det(\Sigma^2) + \log\frac{\mathrm{van}(\Sigma^2)}{\mathrm{van}(\Sigma^{2/N})}. \tag{5.8}$$

The formula extends to repeated singular values by continuity.

Corollary 5.6. *For every $X \in \mathrm{GL}(d,\mathbb{C})$,*

$$\mathrm{vol}(\mathcal{O}_X)\ \det(\mathcal{A}_{N,X}^{-1})^{1/2} = u_d^{N-1}, \tag{5.9}$$

where the determinant is taken for the complex-linear operator $\mathcal{A}_{N,X}^{-1}$ on $T_X\mathrm{GL}(d,\mathbb{C}) \cong \mathrm{M}_d(\mathbb{C})$. Equivalently,

$$S_N(X) = (N-1)\log u_d - \frac{1}{2}\log\det(\mathcal{A}_{N,X}^{-1}). \tag{5.10}$$

Ricci curvature. The identity (5.10) expresses $\det(\mathcal{A}_{N,X}^{-1})$ in terms of the entropy. Since the Hermitian matrix of h_N is given by $\mathcal{A}_{N,X}^{-1}$, we have $\det h_N = \det(\mathcal{A}_{N,X}^{-1})$.

Substituting into $\rho_N = -\mathrm{i}\,\partial\bar{\partial}\log\det h_N$ shows that the entropy is a Ricci potential for the Kähler metric h_N.

Corollary 5.7. *Let ρ_N be the Ricci form of the Kähler metric h_N on $\mathrm{GL}(d,\mathbb{C})$. Then*

$$\rho_N = -\mathrm{i}\,\partial\bar{\partial}\log\det h_N = 2\mathrm{i}\,\partial\bar{\partial}S_N. \tag{5.11}$$

Thus $2S_N$ is a Ricci potential for the Kähler metric h_N.

5.2 The metric on $\mathrm{GL}(d,\mathbb{C})$

In Chapter 4, we identified the quotient $\mathcal{M}/K$ with $\mathrm{GL}(d,\mathbb{C})$. We now compute the Hermitian metric carried by this quotient. The calculation has two parts. First, bi-unitary symmetry reduces the problem to a positive diagonal matrix. Second, we construct horizontal lifts of the matrix units and use them to compute the metric.

Let (g, J, ω) be the ambient Kähler structure on $\mathrm{GL}(d,\mathbb{C})^N$, and let h be the associated Frobenius Hermitian form. Let ι be the metric on $\mathcal{M}$ induced by g. By Theorem 4.9, the quotient $\mathcal{M}/K$ carries a Kähler structure. By Theorem 4.10,

$$\bar{\phi} : \mathcal{M}/K \longrightarrow \mathrm{GL}(d,\mathbb{C})$$

is a diffeomorphism. We transport the quotient Kähler structure to $\mathrm{GL}(d,\mathbb{C})$ through $\bar{\phi}$.

The operator $\mathcal{A}_{N,X}$ already appeared in Theorem 2.5. Here we recover the same operator from Kähler reduction in the complex DLN.

Proposition 5.8. *Let $Q, R \in \mathrm{U}(d)$, and define*

$$(Q, R)\cdot(W_N, \ldots, W_1) = (QW_N,\, W_{N-1},\, \ldots,\, W_2,\, W_1R^*).$$

This action is holomorphic and preserves the Hermitian form h. It also preserves $\mathcal{M}$, commutes with the K-action, and descends under $\bar{\phi}$ to

$$X \longmapsto QXR^*$$

on $\mathrm{GL}(d, \mathbb{C})$.

Proof. Left and right multiplication by fixed unitary matrices are complex-linear and preserve the Frobenius Hermitian form. Hence the action is holomorphic and preserves h.

The action changes only the first and last factors in the product. Since

$$(QW_N)^*(QW_N) = W_N^* W_N, \qquad (W_1 R^*)(W_1 R^*)^* = W_1 W_1^*,$$

the equations for balancedness are unchanged. Thus the action preserves $\mathcal{M}$.

Let $U = (U_{N-1}, \dots, U_1) \in K$. Applying U and then (Q, R), or applying (Q, R) and then U, gives

$$(QW_N U_{N-1}^{-1},\, U_{N-1} W_{N-1} U_{N-2}^{-1},\, \dots,\, U_1 W_1 R^*).$$

Thus the two actions commute. Finally,

$$\phi\Big((Q, R) \cdot (W_N, \dots, W_1)\Big) = QW_N \cdots W_1 R^* = Q\phi(W_N, \dots, W_1)R^*.$$

Therefore the induced action on the quotient is $X \mapsto QXR^*$. $\square$

Fix a positive diagonal matrix

$$\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_d), \qquad \sigma_i > 0,$$

and set

$$\Lambda = \Sigma^{1/N} = \operatorname{diag}(\lambda_1, \dots, \lambda_d), \qquad \lambda_i^N = \sigma_i.$$

Define

$$\mathbf{W}^\Sigma = (\Lambda, \dots, \Lambda) \in \mathcal{M}.$$

For $1 \le i, j \le d$, let E_{ij} denote the standard matrix units, and define

$$\alpha_{ij} := \sum_{k=1}^{N} \sigma_i^{\frac{2(N-k)}{N}} \sigma_j^{\frac{2(k-1)}{N}} = \sum_{k=1}^{N} \lambda_i^{2(N-k)} \lambda_j^{2(k-1)}. \tag{5.12}$$

Lemma 5.9. *At $X = \Sigma$, the operator $\mathcal{A}_{N,X}$ satisfies*

$$\mathcal{A}_{N,X}(E_{ij}) = \alpha_{ij} E_{ij}.$$

Proof. By the definition of $\mathcal{A}_{N,X}$ in (5.1),

$$\mathcal{A}_{N,X}(E_{ij}) = \sum_{k=1}^{N} (\Sigma\Sigma^*)^{\frac{N-k}{N}} E_{ij} (\Sigma^*\Sigma)^{\frac{k-1}{N}}.$$

Since

$$\Sigma\Sigma^* = \Sigma^2 = \operatorname{diag}(\sigma_1^2, \dots, \sigma_d^2),$$

we have

$$(\Sigma\Sigma^*)^{\frac{N-k}{N}} E_{ij} (\Sigma^*\Sigma)^{\frac{k-1}{N}} = \sigma_i^{\frac{2(N-k)}{N}} \sigma_j^{\frac{2(k-1)}{N}} E_{ij}.$$

Summing over k gives the formula. $\square$

For $1 \le i, j \le d$, define

$$c_k^{(ij)} := \lambda_i^{N-k} \lambda_j^{k-1}, \qquad 1 \le k \le N, \tag{5.13}$$

and

$$\Gamma_{ij}^{\Sigma} := \Big(c_N^{(ij)} E_{ij},\, c_{N-1}^{(ij)} E_{ij},\, \ldots,\, c_1^{(ij)} E_{ij}\Big) \in T_{\mathbf{W}^{\Sigma}} \mathrm{GL}(d, \mathbb{C})^N. \tag{5.14}$$

Thus the component of Γ_{ij}^{Σ} in the k-th layer is $c_k^{(ij)} E_{ij}$. The recurrence

$$\lambda_i c_{k+1}^{(ij)} = \lambda_j c_k^{(ij)}, \qquad 1 \le k \le N-1,$$

is the reason these vectors lie in $T_{\mathbf{W}^{\Sigma}} \mathcal{M}$ and are horizontal.

Lemma 5.10. *For every i, j, the vector Γ_{ij}^{Σ} lies in $T_{\mathbf{W}^{\Sigma}} \mathcal{M}$, is g-orthogonal to $T_{\mathbf{W}^{\Sigma}}(K \cdot \mathbf{W}^{\Sigma})$, and satisfies*

$$(\mathrm{d}\phi)_{\mathbf{W}^{\Sigma}}(\Gamma_{ij}^{\Sigma}) = \alpha_{ij} E_{ij}, \qquad \|\Gamma_{ij}^{\Sigma}\|_g^2 = \alpha_{ij}. \tag{5.15}$$

Proof. For $1 \le k \le N-1$, let

$$F_k(\mathbf{W}) := W_{k+1}^* W_{k+1} - W_k W_k^*.$$

Then $\mathcal{M} = \{F_1 = \cdots = F_{N-1} = 0\}$. At $\mathbf{W}^{\Sigma} = (\Lambda, \ldots, \Lambda)$,

$$(\mathrm{d}F_k)_{\mathbf{W}^{\Sigma}}(Z) = Z_{k+1}^* \Lambda + \Lambda Z_{k+1} - Z_k \Lambda - \Lambda Z_k^*.$$

Substituting $Z = \Gamma_{ij}^{\Sigma}$ gives

$$(\mathrm{d}F_k)_{\mathbf{W}^{\Sigma}}(\Gamma_{ij}^{\Sigma}) = \Big(\lambda_i c_{k+1}^{(ij)} - \lambda_j c_k^{(ij)}\Big) E_{ij} + \Big(\lambda_i \overline{c_{k+1}^{(ij)}} - \lambda_j \overline{c_k^{(ij)}}\Big) E_{ji}.$$

By (5.13),

$$\lambda_i c_{k+1}^{(ij)} = \lambda_i^{N-k} \lambda_j^k = \lambda_j c_k^{(ij)}.$$

Hence $(\mathrm{d}F_k)_{\mathbf{W}^{\Sigma}}(\Gamma_{ij}^{\Sigma}) = 0$ for every k, so $\Gamma_{ij}^{\Sigma} \in T_{\mathbf{W}^{\Sigma}} \mathcal{M}$.

Next let $\xi = (\xi_{N-1}, \ldots, \xi_1) \in \mathfrak{k}$, and set $\xi_0 = \xi_N = 0$. The fundamental vector field generated by ξ is

$$K_\xi(\mathbf{W}^\Sigma)_k = \xi_k \Lambda - \Lambda \xi_{k-1}.$$

Therefore

$$\begin{aligned} g\Big(\Gamma_{ij}^\Sigma, K_\xi(\mathbf{W}^\Sigma)\Big) &= \Re \sum_{k=1}^{N} \operatorname{tr}\Big((c_k^{(ij)} E_{ij})^* (\xi_k \Lambda - \Lambda \xi_{k-1})\Big) \\ &= \Re \sum_{k=1}^{N-1} \Big(\overline{c_k^{(ij)}} \lambda_j - \overline{c_{k+1}^{(ij)}} \lambda_i\Big) (\xi_k)_{ij} = 0. \end{aligned}$$

Thus Γ_{ij}^Σ is orthogonal to the tangent space of the K-orbit.

The differential of ϕ at $\mathbf{W}^\Sigma$ is

$$(\mathrm{d}\phi)_{\mathbf{W}^\Sigma}(Z) = \sum_{k=1}^{N} \Lambda^{N-k} Z_k \Lambda^{k-1}.$$

Substituting $Z = \Gamma_{ij}^\Sigma$, we obtain

$$(\mathrm{d}\phi)_{\mathbf{W}^\Sigma}(\Gamma_{ij}^\Sigma) = \sum_{k=1}^{N} c_k^{(ij)} \lambda_i^{N-k} \lambda_j^{k-1} E_{ij} = \alpha_{ij} E_{ij}.$$

Finally,

$$\|\Gamma_{ij}^\Sigma\|_g^2 = \sum_{k=1}^{N} \|c_k^{(ij)} E_{ij}\|_F^2 = \sum_{k=1}^{N} |c_k^{(ij)}|^2 = \alpha_{ij}.$$

This proves (5.15). $\square$

Proof of Theorem 5.1. Let h_N^{red} denote the Hermitian metric transported from the quotient $\mathcal{M}/K$, and let $\widetilde{h}_N$ denote the Hermitian form on the right-hand side of (5.2). By Proposition 5.8, the form h_N^{red} is invariant under

$$X \longmapsto Q X R^*, \qquad Q, R \in \mathrm{U}(d).$$

The same invariance holds for $\widetilde{h}_N$. Indeed, from the definition (5.1),

$$\mathcal{A}_{N,QXR^*}(QZR^*) = Q\,\mathcal{A}_{N,X}(Z)\,R^*, \qquad Q,R \in \mathrm{U}(d).$$

Hence

$$\mathcal{A}_{N,QXR^*}^{-1}(QZR^*) = Q\,\mathcal{A}_{N,X}^{-1}(Z)\,R^*.$$

Therefore

$$\begin{aligned}
\widetilde{h}_N(QXR^*)(QZ_1R^*, QZ_2R^*) &= \mathrm{tr}\Big((QZ_1R^*)^*\mathcal{A}_{N,QXR^*}^{-1}(QZ_2R^*)\Big) \\
&= \mathrm{tr}\Big(RZ_1^*Q^*Q\mathcal{A}_{N,X}^{-1}(Z_2)R^*\Big) \\
&= \mathrm{tr}\Big(Z_1^*\mathcal{A}_{N,X}^{-1}(Z_2)\Big) \\
&= \widetilde{h}_N(X)(Z_1, Z_2).
\end{aligned}$$

Thus both Hermitian forms are bi-unitarily invariant. It is therefore enough to compare them at a positive diagonal matrix Σ.

At $\mathbf{W}^\Sigma$, define

$$H_{ij}^\Sigma := \alpha_{ij}^{-1}\Gamma_{ij}^\Sigma.$$

By Lemma 5.10, H_{ij}^Σ is the horizontal lift of E_{ij}. Hence

$$h_N^{\mathrm{red}}(\Sigma)(E_{ij}, E_{k\ell}) = h\Big(H_{ij}^\Sigma, H_{k\ell}^\Sigma\Big).$$

The vectors Γ_{ij}^Σ and $\Gamma_{k\ell}^\Sigma$ are h-orthogonal when $(i,j) \neq (k,\ell)$, because their layerwise components are scalar multiples of distinct matrix units. Using (5.15), we get

$$h_N^{\mathrm{red}}(\Sigma)(E_{ij}, E_{k\ell}) = \delta_{ik}\delta_{j\ell}\,\alpha_{ij}^{-1}.$$

By Lemma 5.9, the same basis diagonalizes $\mathcal{A}_{N,X}$ with eigenvalues α_{ij}. Therefore

$$\widetilde{h}_N(\Sigma)(E_{ij}, E_{k\ell}) = \operatorname{tr}\Big(E_{ij}^* \mathcal{A}_{N,X}^{-1}(E_{k\ell})\Big) = \delta_{ik}\delta_{j\ell}\, \alpha_{ij}^{-1}.$$

Thus $h_N^{\mathrm{red}} = \widetilde{h}_N$ at Σ. Bi-unitary invariance then gives the equality for every $X \in \mathrm{GL}(d, \mathbb{C})$.

Taking real and imaginary parts gives the Riemannian metric g_N and the Kähler form ω_N. Equivalently,

$$\omega_{N,X}(Z_1, Z_2) = \Im \operatorname{tr}\Big(Z_1^* \mathcal{A}_{N,X}^{-1}(Z_2)\Big). \tag{5.16}$$

Since the Hermitian form is conjugate symmetric, this is the same as

$$\omega_{N,X}(Z_1, Z_2) = \frac{1}{2\mathrm{i}} \left[\operatorname{tr}\Big(Z_1^* \mathcal{A}_{N,X}^{-1}(Z_2)\Big) - \operatorname{tr}\Big(Z_2^* \mathcal{A}_{N,X}^{-1}(Z_1)\Big)\right]. \tag{5.17}$$

□

Corollary 5.11. *Let $X = Q\Sigma R^*$ be a singular value decomposition, and let u_i and v_j be the columns of Q and R. Then*

$$\mathcal{A}_{N,X}(u_i v_j^*) = \alpha_{ij} u_i v_j^*, \qquad \alpha_{ij} = \sum_{k=1}^{N} \sigma_i^{\frac{2(N-k)}{N}} \sigma_j^{\frac{2(k-1)}{N}}.$$

Equivalently,

$$h_N(X)(u_i v_j^*, u_k v_\ell^*) = \delta_{ik}\delta_{j\ell}\, \alpha_{ij}^{-1}.$$

If

$$Z = Q\zeta R^*, \qquad Z' = Q\eta R^*,$$

then

$$\omega_{N,X}(Z,Z') = \sum_{i,j=1}^{d} \alpha_{ij}^{-1}\, \Im\Big(\overline{\zeta_{ij}}\eta_{ij}\Big). \tag{5.18}$$

Proof. The formula for $\mathcal{A}_{N,X}$ is Lemma 5.9 transported by the bi-unitary change of variables $X = Q\Sigma R^*$. The formula for h_N then follows from Theorem 5.1. Finally,

$$\operatorname{tr}\Big(Z^* \mathcal{A}_{N,X}^{-1}(Z')\Big) = \sum_{i,j=1}^{d} \alpha_{ij}^{-1}\overline{\zeta_{ij}}\eta_{ij},$$

so taking the imaginary part gives (5.18). □

5.3 The Kähler potential

The formula for the Hermitian metric identifies the quotient tensor. We now show that its Kähler form is generated by the scalar potential Ψ_N in (5.3). The proof is a calculation at a positive diagonal matrix, followed by transport through the bi-unitary symmetry from Proposition 5.8.

Lemma 5.12. *For $r, s > 0$,*

$$\frac{r^{1/N} - s^{1/N}}{r - s} = \left(\sum_{k=1}^{N} r^{\frac{N-k}{N}} s^{\frac{k-1}{N}}\right)^{-1}, \tag{5.19}$$

where the left-hand side is interpreted by continuity when $r = s$. Consequently, if $r_i = \sigma_i^2$, then

$$\alpha_{ij}^{-1} = \begin{cases} \dfrac{r_i^{1/N} - r_j^{1/N}}{r_i - r_j}, & r_i \neq r_j, \\[2ex] \dfrac{1}{N} r_i^{1/N-1}, & r_i = r_j. \end{cases} \tag{5.20}$$

Proof. The finite geometric series gives

$$r - s = \left(r^{1/N} - s^{1/N}\right) \sum_{k=1}^{N} r^{\frac{N-k}{N}} s^{\frac{k-1}{N}}.$$

Dividing by the nonzero factor and then passing to the limit $r = s$ proves (5.19). Taking $r = r_i$ and $s = r_j$ gives (5.20). $\square$

Proof of Theorem 5.2. Let $\omega_N^{\Psi} := \mathrm{i}\,\partial\bar{\partial}\Psi_N$, and let h_N^{Ψ} be the Hermitian form determined by ω_N^{Ψ}. With the convention $h = g + \mathrm{i}\omega$, the coefficient of h_N^{Ψ} in a holomorphic coordinate direction Z is

$$h_N^{\Psi}(Z, Z) = 2\,\partial_z \partial_{\bar{z}}\Big|_{z=0} \Psi_N(X + zZ). \tag{5.21}$$

Indeed, $\mathrm{i}\,\partial\bar{\partial} f$ evaluated on the real plane spanned by Z and $\mathrm{i}Z$ gives $2\partial_z \partial_{\bar{z}} f$.

Both ω_N and Ψ_N are invariant under the left-right action

$$X \longmapsto QXR^*, \qquad Q, R \in \mathrm{U}(d),$$

and so is ω_N^{Ψ}. It is therefore enough to work at a positive diagonal matrix

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d), \qquad r_i = \sigma_i^2.$$

We first assume that the numbers $r_1, \ldots, r_d$ are pairwise distinct. The general case follows by continuity at the end of the proof.

The stabilizer of Σ contains the diagonal subgroup

$$D \longmapsto (D, D), \qquad D = \mathrm{diag}(e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_d}).$$

Its differential acts on the matrix units by

$$E_{ij} \longmapsto e^{\mathrm{i}(\theta_i - \theta_j)} E_{ij}.$$

Since h_N^Ψ is invariant under this stabilizer, the off-diagonal directions are orthogonal to one another and to the diagonal subspace:

$$h_N^\Psi(E_{ij}, E_{k\ell}) = 0 \quad \text{if } i \neq j \text{ and } (k,\ell) \neq (i,j).$$

It remains to compute the off-diagonal coefficients and the diagonal block.

First let $i \neq j$ and set

$$X(z) = \Sigma + zE_{ij}.$$

Then

$$X(z)^* X(z) = \Sigma^2 + z\sigma_i E_{ij} + \bar{z}\sigma_i E_{ji} + |z|^2 E_{jj}.$$

The only nontrivial block is the block on $\mathrm{span}\{e_i, e_j\}$:

$$B_{ij}(z) = \begin{pmatrix} r_i & \sigma_i z \\ \sigma_i \bar{z} & r_j + |z|^2 \end{pmatrix}.$$

Write $t = |z|^2$, and let $\rho_i(t)$ and $\rho_j(t)$ be the two eigenvalue branches of $B_{ij}(z)$ satisfying

$$\rho_i(0) = r_i, \qquad \rho_j(0) = r_j.$$

Since

$$\rho_i(t) + \rho_j(t) = r_i + r_j + t, \qquad \rho_i(t)\rho_j(t) = r_i r_j,$$

we obtain

$$\rho_i'(0) + \rho_j'(0) = 1, \qquad r_j\rho_i'(0) + r_i\rho_j'(0) = 0.$$

Thus

$$\rho_i'(0) = \frac{r_i}{r_i - r_j}, \qquad \rho_j'(0) = -\frac{r_j}{r_i - r_j}.$$

The remaining eigenvalues of $X(z)^*X(z)$ are constant, so

$$\Psi_N(X(z)) = \Psi_N(\Sigma) + \frac{N}{2}\Big(\rho_i(t)^{1/N} + \rho_j(t)^{1/N} - r_i^{1/N} - r_j^{1/N}\Big).$$

Therefore

$$\begin{aligned} 2\,\partial_z\partial_{\bar z}\Big|_{z=0}\Psi_N(X(z)) &= N\Big(\frac{1}{N}r_i^{1/N-1}\rho_i'(0) + \frac{1}{N}r_j^{1/N-1}\rho_j'(0)\Big) \\ &= \frac{r_i^{1/N} - r_j^{1/N}}{r_i - r_j} = \alpha_{ij}^{-1}, \end{aligned}$$

where the last equality is (5.20). Hence

$$h_N^\Psi(E_{ij}, E_{ij}) = \alpha_{ij}^{-1} \qquad (i \neq j).$$

We now compute the diagonal block. If $i \neq j$, set

$$X(z, w) = \Sigma + zE_{ii} + wE_{jj}.$$

Then $X(z,w)^*X(z,w)$ is diagonal, and the only entries depending on z and w are

$$|\sigma_i + z|^2, \qquad |\sigma_j + w|^2.$$

Thus $\Psi_N(X(z,w))$ is the sum of a function of $z, \bar z$, a function of $w, \bar w$, and a constant.

Therefore

$$\partial_z \partial_{\bar{w}} \Big|_{z=w=0} \Psi_N(X(z,w)) = 0,$$

and hence

$$h_N^{\Psi}(E_{ii}, E_{jj}) = 0 \qquad (i \neq j).$$

It remains to compute the diagonal coefficient. Set

$$X(z) = \Sigma + zE_{ii}.$$

Then

$$X(z)^* X(z) = \operatorname{diag}\Big(r_1, \dots, |\sigma_i + z|^2, \dots, r_d\Big),$$

and hence

$$\Psi_N(X(z)) = \Psi_N(\Sigma) + \frac{N}{2}\Big(|\sigma_i + z|^{2/N} - \sigma_i^{2/N}\Big).$$

If

$$u(z, \bar{z}) = |\sigma_i + z|^2,$$

then

$$\partial_z u\Big|_0 = \sigma_i, \qquad \partial_{\bar{z}} u\Big|_0 = \sigma_i, \qquad \partial_z \partial_{\bar{z}} u\Big|_0 = 1.$$

The chain rule gives

$$\begin{aligned} 2\, \partial_z \partial_{\bar{z}} \Big|_{z=0} \Psi_N(X(z)) &= N\Big(\frac{1}{N}\Big(\frac{1}{N} - 1\Big) r_i^{1/N-2} \sigma_i^2 + \frac{1}{N} r_i^{1/N-1}\Big) \\ &= \frac{1}{N} r_i^{1/N-1} = \alpha_{ii}^{-1}. \end{aligned}$$

Thus

$$h_N^{\Psi}(E_{ij}, E_{k\ell}) = \delta_{ik}\delta_{j\ell}\, \alpha_{ij}^{-1}.$$

By Corollary 5.11, these are exactly the coefficients of the Hermitian metric h_N at Σ. Therefore

$$h_N^{\Psi} = h_N \qquad \text{and hence} \qquad \omega_N^{\Psi} = \omega_N$$

at every positive diagonal matrix with pairwise distinct entries.

Both sides are smooth, bi-unitarily invariant real $(1,1)$-forms on $\mathrm{GL}(d,\mathbb{C})$. Since positive diagonal matrices with pairwise distinct entries are dense in the positive diagonal cone, and every matrix in $\mathrm{GL}(d,\mathbb{C})$ has a singular value decomposition, the identity extends to all of $\mathrm{GL}(d,\mathbb{C})$. □

5.3.1 The potential as a minimum on each fibre

We next prove the variational characterization of Ψ_N. Throughout this subsection, $\mathcal{R}$ denotes the ridge functional in (5.4). The function $\mathcal{R}$ is also a Kähler potential for the ambient Frobenius metric on $\mathrm{GL}(d,\mathbb{C})^N$. After reduction, its minimum on each fibre gives the quotient potential.

Proof of Proposition 5.3. The fibrewise minimum principle, Theorem 2.8 from Chapter 2, applied to the ridge functional, gives

$$\operatorname*{arg\,min}_{\mathbf{W}\in\mathcal{F}_X} \mathcal{R}(\mathbf{W}) = \mathcal{O}_X.$$

Thus it remains only to compute the minimum value.

Write a singular value decomposition

$$X = Q\Sigma R^*, \qquad \Sigma = \mathrm{diag}(\sigma_1,\ldots,\sigma_d),$$

and set

$$\Lambda = \Sigma^{1/N}.$$

Then

$$\mathbf{W}^X = \Big(Q\Lambda, \underbrace{\Lambda, \dots, \Lambda}_{N-2 \text{ factors}}, \Lambda R^*\Big) \in \mathcal{O}_X,$$

with the evident interpretation when $N = 2$. Since the K-action is unitary on each layer, $\mathcal{R}$ is constant on $\mathcal{O}_X$. Therefore

$$\min_{\mathbf{W} \in \mathcal{F}_X} \mathcal{R}(\mathbf{W}) = \mathcal{R}(\mathbf{W}^X) = \frac{1}{2} \sum_{k=1}^{N} \operatorname{tr}(\Lambda^2) = \frac{N}{2} \operatorname{tr}(\Sigma^{2/N}).$$

On the other hand,

$$X^* X = R \Sigma^2 R^*, \qquad (X^* X)^{1/N} = R \Sigma^{2/N} R^*.$$

Taking traces gives

$$\frac{N}{2} \operatorname{tr}(\Sigma^{2/N}) = \frac{N}{2} \operatorname{tr}\Big((X^* X)^{1/N}\Big) = \Psi_N(X),$$

by (5.3). This proves the asserted minimum formula and the asserted description of its minimizers. $\square$

The equality in Proposition 5.3 is structural. It follows from how the ridge functional behaves under Kähler reduction.

Indeed, $\mathcal{R}$ is K-invariant and satisfies

$$\mathrm{i}\, \partial \bar{\partial} \mathcal{R} = \omega$$

on the ambient Kähler manifold $\mathrm{GL}(d, \mathbb{C})^N$. Hence the restriction of $\mathcal{R}$ to $\mathcal{M} = \mu^{-1}(0)$

descends to a function

$$\psi_{\mathrm{red}} : \mathcal{M}/K \longrightarrow \mathbb{R}, \qquad \psi_{\mathrm{red}}([W]) = \mathcal{R}(W).$$

If $Y, Z \in T_{[W]}(\mathcal{M}/K)$, and if $\widetilde{Y}, \widetilde{Z}$ are their horizontal lifts at $W \in \mathcal{M}$, then

$$\left(\mathrm{i}\,\partial\bar{\partial}\psi_{\mathrm{red}}\right)_{[W]}(Y, Z) = \left(\mathrm{i}\,\partial\bar{\partial}\mathcal{R}\right)_{W}(\widetilde{Y}, \widetilde{Z}) = \omega_W(\widetilde{Y}, \widetilde{Z}) = (\omega_{\mathrm{red}})_{[W]}(Y, Z).$$

Thus ψ_{red} is a Kähler potential for the reduced Kähler form on $\mathcal{M}/K$. After transport by $\bar{\phi}$, it becomes a Kähler potential for ω_N on $\mathrm{GL}(d, \mathbb{C})$.

Under the identification

$$\bar{\phi} : \mathcal{M}/K \longrightarrow \mathrm{GL}(d, \mathbb{C}),$$

this descended function becomes

$$(\psi_{\mathrm{red}} \circ \bar{\phi}^{-1})(X) = \mathcal{R}(W), \qquad W \in \mathcal{O}_X.$$

By the fibrewise minimum principle, $\mathcal{O}_X$ is precisely the set of minimizers of $\mathcal{R}$ on $\mathcal{F}_X$. Therefore

$$(\psi_{\mathrm{red}} \circ \bar{\phi}^{-1})(X) = \min_{\mathbf{W} \in \mathcal{F}_X} \mathcal{R}(\mathbf{W}) = \Psi_N(X).$$

Thus the quotient Kähler potential Ψ_N is the minimum value of the ambient Kähler potential $\mathcal{R}$ on the fibre $\mathcal{F}_X$. Equivalently, the argument in this subsection gives a second proof of Theorem 5.2 from Proposition 5.3.

Remark 5.13. Proposition 5.3 concerns the ridge functional. The Kirwan–Ness functional

$$\mathcal{E}(\mathbf{W}) := \frac{1}{2}\|\mu(\mathbf{W})\|^2$$

has a different role. On a fibre $\mathcal{F}_X$,

$$\min_{\mathbf{W}\in\mathcal{F}_X} \mathcal{E}(\mathbf{W}) = 0, \qquad \operatorname*{arg\,min}_{\mathbf{W}\in\mathcal{F}_X} \mathcal{E}(\mathbf{W}) = \mathcal{O}_X,$$

because $\mathcal{M} = \mu^{-1}(0)$ and $\mathcal{O}_X = \mathcal{F}_X \cap \mathcal{M}$. Thus $\mathcal{E}$ drives the flow toward balancedness, while $\mathcal{R}$ produces the nonconstant scalar potential Ψ_N on the quotient.

5.4 An orthonormal basis on $\mathcal{M}$

For $\mathbf{W} \in \mathcal{M}$ with $X = \phi(\mathbf{W})$, the tangent space $T_{\mathbf{W}}\mathcal{M}$ splits into directions tangent to the K-orbit through $\mathbf{W}$ and directions orthogonal to that orbit. The orbit directions are precisely the vertical directions for the quotient map, and they lie in the kernel of $\mathrm{d}(\phi|_{\mathcal{M}})_{\mathbf{W}}$. We write

$$\mathcal{H}_{\mathbf{W}} := T_{\mathbf{W}}\mathcal{M} \cap T_{\mathbf{W}}(K\cdot\mathbf{W})^{\perp_g}.$$

Then

$$T_{\mathbf{W}}\mathcal{M} = T_{\mathbf{W}}(K\cdot\mathbf{W}) \oplus \mathcal{H}_{\mathbf{W}}$$

as a g-orthogonal direct sum. We construct an orthonormal basis adapted to this splitting. Its vertical part describes the orbit geometry of $\mathcal{O}_X$, and its horizontal part projects to an orthonormal basis of $T_X\mathrm{GL}(d,\mathbb{C})$ for the quotient metric.

The construction is most transparent at a diagonal representative

$$\mathbf{W}^{\Sigma} = (\Lambda,\ldots,\Lambda), \qquad \Lambda = \Sigma^{1/N}.$$

At this point the fundamental vector field has the layerwise form

$$K_\xi(\mathbf{W}^{\Sigma})_k = \xi_k\Lambda - \Lambda\xi_{k-1}.$$

After choosing the standard real basis of $\mathfrak{u}(d)$ used in Proposition 5.16, each vertical component is described by a string of coefficients along the depth direction. The zero endpoint conditions $\xi_0 = \xi_N = 0$ determine the vertical modes. In the off-diagonal components, restoring the two endpoint coefficients gives a second-order recurrence whose two geometric solutions are the horizontal lifts constructed in Section 5.2. Thus the vertical and horizontal bases arise from the same layerwise recurrence, with different endpoint conditions.

5.4.1 Vertical directions

Let

$$X = Q\Sigma R^*, \qquad \Lambda = \Sigma^{1/N}, \qquad \mathbf{W}^X = (Q\Lambda, \Lambda, \ldots, \Lambda, \Lambda R^*) \in \mathcal{M}.$$

For $X = \Sigma$, write

$$\mathbf{W}^\Sigma = (\Lambda, \ldots, \Lambda).$$

By Proposition 4.6,

$$\mathcal{O}_X = K \cdot \mathbf{W}^X.$$

Let $\pi : \mathcal{M} \to \mathcal{M}/K$ be the quotient map. Since $\phi|_\mathcal{M} = \bar{\phi} \circ \pi$ and $\bar{\phi}$ is a diffeomorphism, the kernel of $\mathrm{d}(\phi|_\mathcal{M})$ is the vertical space of π.

Lemma 5.14. *At* $\mathbf{W}^\Sigma$*,*

$$T_{\mathbf{W}^\Sigma}(K \cdot \mathbf{W}^\Sigma) = \ker\Big(\mathrm{d}(\phi|_\mathcal{M})\Big)_{\mathbf{W}^\Sigma} = T_{\mathbf{W}^\Sigma}\mathcal{M} \cap \ker(\mathrm{d}\phi)_{\mathbf{W}^\Sigma}.$$

Proof. Since $\phi|_\mathcal{M} = \bar{\phi} \circ \pi$ and $\bar{\phi}$ is a diffeomorphism,

$$\ker\Big(\mathrm{d}(\phi|_\mathcal{M})\Big)_{\mathbf{W}^\Sigma} = \ker(\mathrm{d}\pi)_{\mathbf{W}^\Sigma}.$$

The kernel of $(\mathrm{d}\pi)_{\mathbf{W}^\Sigma}$ is the tangent space to the K-orbit through $\mathbf{W}^\Sigma$. This proves the first equality. The second equality follows from the definition of the differential of a restriction. □

For $1 \le i < j \le d$, define

$$A_{ij} := \frac{1}{\sqrt{2}}(E_{ij} - E_{ji}), \qquad S_{ij} := \frac{\mathrm{i}}{\sqrt{2}}(E_{ij} + E_{ji}), \tag{5.22}$$

and for $1 \le i \le d$ define

$$D_i := \mathrm{i}E_{ii}. \tag{5.23}$$

The vectors in (5.22) and (5.23) form a real orthonormal basis of $\mathfrak{u}(d)$ for the Frobenius inner product.

For $\xi = (\xi_{N-1}, \dots, \xi_1) \in \mathfrak{k} = \mathfrak{u}(d)^{N-1}$, set $\xi_0 = \xi_N = 0$. At the diagonal representative,

$$K_\xi(\mathbf{W}^\Sigma)_k = \xi_k\Lambda - \Lambda\xi_{k-1}, \qquad 1 \le k \le N. \tag{5.24}$$

For $1 \le m \le N-1$, let

$$s_m(k) := \sqrt{\frac{2}{N}} \sin\frac{m\pi k}{N}, \qquad 1 \le k \le N-1. \tag{5.25}$$

The vectors $s_m = (s_m(1), \dots, s_m(N-1))^\top$ form an orthonormal basis of $\mathbb{R}^{N-1}$ and diagonalize the tridiagonal matrix

$$C_{N-1} = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}. \tag{5.26}$$

Indeed,

$$C_{N-1}s_m = \kappa_m s_m, \qquad \kappa_m = 2 - 2\cos\frac{m\pi}{N} = 4\sin^2\frac{m\pi}{2N}. \tag{5.27}$$

Lemma 5.15. *Let $a = (a_1, \dots, a_{N-1}) \in \mathbb{R}^{N-1}$, and set $a_0 = a_N = 0$.*

1. *For $1 \le i < j \le d$, if $\xi_k = a_k A_{ij}$ or $\xi_k = a_k S_{ij}$, then*

$$\|K_\xi(\mathbf{W}^\Sigma)\|_g^2 = a^\top\Big((\lambda_i - \lambda_j)^2 I_{N-1} + \lambda_i\lambda_j C_{N-1}\Big)a. \tag{5.28}$$

2. *For $1 \le i \le d$, if $\xi_k = a_k D_i$, then*

$$\|K_\xi(\mathbf{W}^\Sigma)\|_g^2 = \lambda_i^2 a^\top C_{N-1} a. \tag{5.29}$$

3. *The sectors indexed by different pairs (i, j) are mutually orthogonal. The families generated by A_{ij}, S_{ij}, and D_i are also mutually orthogonal.*

Proof. We use (5.24). For A_{ij},

$$A_{ij}\Lambda = \frac{1}{\sqrt{2}}\Big(\lambda_j E_{ij} - \lambda_i E_{ji}\Big), \qquad \Lambda A_{ij} = \frac{1}{\sqrt{2}}\Big(\lambda_i E_{ij} - \lambda_j E_{ji}\Big).$$

Thus

$$K_\xi(\mathbf{W}^\Sigma)_k = \frac{1}{\sqrt{2}}\Big((\lambda_j a_k - \lambda_i a_{k-1})E_{ij} - (\lambda_i a_k - \lambda_j a_{k-1})E_{ji}\Big).$$

Taking the Frobenius norm and summing over k gives

$$\begin{aligned}
\|K_\xi(\mathbf{W}^\Sigma)\|_g^2 &= \sum_{k=1}^{N} \frac{1}{2}\Big((\lambda_j a_k - \lambda_i a_{k-1})^2 + (\lambda_i a_k - \lambda_j a_{k-1})^2\Big) \\
&= (\lambda_i^2 + \lambda_j^2)\sum_{k=1}^{N-1} a_k^2 - 2\lambda_i\lambda_j \sum_{k=1}^{N-2} a_k a_{k+1} \\
&= a^\top\Big((\lambda_i - \lambda_j)^2 I_{N-1} + \lambda_i\lambda_j C_{N-1}\Big)a.
\end{aligned}$$

The same calculation applies to S_{ij}, since

$$S_{ij}\Lambda = \frac{\mathrm{i}}{\sqrt{2}}\Big(\lambda_j E_{ij} + \lambda_i E_{ji}\Big), \qquad \Lambda S_{ij} = \frac{\mathrm{i}}{\sqrt{2}}\Big(\lambda_i E_{ij} + \lambda_j E_{ji}\Big).$$

For D_i, one has $D_i\Lambda = \Lambda D_i = \mathrm{i}\lambda_i E_{ii}$. Hence

$$K_\xi(\mathbf{W}^\Sigma)_k = \mathrm{i}\lambda_i(a_k - a_{k-1})E_{ii},$$

and therefore

$$\|K_\xi(\mathbf{W}^\Sigma)\|_g^2 = \lambda_i^2 \sum_{k=1}^{N}(a_k - a_{k-1})^2 = \lambda_i^2 a^\top C_{N-1} a.$$

The orthogonality follows because different sectors involve orthogonal matrix units in each layer. □

For $1 \le i < j \le d$ and $1 \le m \le N-1$, define

$$\mu_{ij,m} := \lambda_i^2 + \lambda_j^2 - 2\lambda_i\lambda_j \cos\frac{m\pi}{N}, \qquad \nu_{i,m} := 4\lambda_i^2 \sin^2\frac{m\pi}{2N}. \tag{5.30}$$

Also define elements of $\mathfrak{k}$ by

$$(\xi^A_{ij,m})_k := s_m(k)A_{ij}, \qquad (\xi^S_{ij,m})_k := s_m(k)S_{ij}, \qquad (\xi^D_{i,m})_k := s_m(k)D_i. \tag{5.31}$$

Proposition 5.16. *The vectors*

$$\mu_{ij,m}^{-1/2}K_{\xi^A_{ij,m}}(\mathbf{W}^\Sigma), \qquad \mu_{ij,m}^{-1/2}K_{\xi^S_{ij,m}}(\mathbf{W}^\Sigma), \qquad \nu_{i,m}^{-1/2}K_{\xi^D_{i,m}}(\mathbf{W}^\Sigma),$$

where $1 \le i < j \le d$, $1 \le i \le d$, and $1 \le m \le N-1$, form a g-orthonormal basis of

$$T_{\mathbf{W}^\Sigma}(K \cdot \mathbf{W}^\Sigma).$$

Proof. By Lemma 5.14, the vertical space is the tangent space to the K-orbit. By Lemma 5.15, the metric on this space splits as the orthogonal sum of the diagonal blocks $\lambda_i^2 C_{N-1}$ and the off-diagonal blocks

$$(\lambda_i - \lambda_j)^2 I_{N-1} + \lambda_i \lambda_j C_{N-1}.$$

The vectors s_m diagonalize C_{N-1} by (5.27). Therefore the vectors in (5.31) diagonalize all blocks. The corresponding eigenvalues are $\mu_{ij,m}$ and $\nu_{i,m}$ from (5.30). The count

$$2(N-1)\binom{d}{2} + d(N-1) = (N-1)d^2 = \dim_{\mathbb{R}} T_{\mathbf{W}^\Sigma}(K \cdot \mathbf{W}^\Sigma)$$

proves completeness. $\square$

5.4.2 Horizontal directions and endpoint coefficients

The horizontal lifts Γ_{ij}^Σ were constructed in Section 5.2. We now record their role in the basis of $T_{\mathbf{W}^\Sigma}\mathcal{M}$ and relate them to the two coefficients excluded by the zero endpoint conditions $a_0 = a_N = 0$.

Proposition 5.17. *Let*

$$\mathcal{H}_{\mathbf{W}^\Sigma} := T_{\mathbf{W}^\Sigma}\mathcal{M} \cap T_{\mathbf{W}^\Sigma}(K \cdot \mathbf{W}^\Sigma)^{\perp_g}.$$

The vectors

$$h_{ij}^\Sigma := \alpha_{ij}^{-1/2}\Gamma_{ij}^\Sigma, \qquad 1 \le i, j \le d,$$

form an h-orthonormal complex basis of $\mathcal{H}_{\mathbf{W}^\Sigma}$. Equivalently,

$$\{h_{ij}^\Sigma, Jh_{ij}^\Sigma : 1 \le i, j \le d\}$$

is a g-orthonormal real basis of $\mathcal{H}_{\mathbf{W}^\Sigma}$. Moreover,

$$(\mathrm{d}\phi)_{\mathbf{W}^\Sigma}(h^\Sigma_{ij}) = \alpha_{ij}^{1/2} E_{ij}. \tag{5.32}$$

Proof. By Lemma 5.10, each Γ^Σ_{ij} lies in $T_{\mathbf{W}^\Sigma}\mathcal{M}$, is orthogonal to the vertical space, and satisfies

$$\|\Gamma^\Sigma_{ij}\|_g^2 = \alpha_{ij}.$$

The same layerwise calculation gives

$$h(\Gamma^\Sigma_{ij}, \Gamma^\Sigma_{k\ell}) = \delta_{ik}\delta_{j\ell}\,\alpha_{ij}.$$

After normalization, the vectors h^Σ_{ij} are therefore h-orthonormal and lie in $\mathcal{H}_{\mathbf{W}^\Sigma}$.

By Theorem 4.9, the horizontal distribution is J-invariant. Hence $\mathcal{H}_{\mathbf{W}^\Sigma}$ is a complex vector space. Its complex dimension is

$$\frac{1}{2}\Big(\dim_{\mathbb{R}} \mathcal{M} - \dim_{\mathbb{R}}(K \cdot \mathbf{W}^\Sigma)\Big) = \frac{1}{2}\Big((N+1)d^2 - (N-1)d^2\Big) = d^2.$$

Thus the d^2 vectors h^Σ_{ij} form a complex basis. The real orthonormal-basis statement follows by adjoining their J-images. Finally, (5.32) follows from (5.15). $\square$

Theorem 5.18. *At $\mathbf{W}^\Sigma$,*

$$T_{\mathbf{W}^\Sigma}\mathcal{M} = T_{\mathbf{W}^\Sigma}(K \cdot \mathbf{W}^\Sigma) \oplus \mathcal{H}_{\mathbf{W}^\Sigma}$$

as a g-orthogonal direct sum. A g-orthonormal real basis is obtained by taking the normalized vertical modes from Proposition 5.16 together with

$$h^\Sigma_{ij}, \qquad Jh^\Sigma_{ij}, \qquad 1 \le i, j \le d.$$

Proof. The horizontal space is the g-orthogonal complement of the vertical space inside $T_{\mathbf{W}^\Sigma}\mathcal{M}$. The vertical basis is orthonormal by Proposition 5.16. The horizontal basis is orthonormal by Proposition 5.17. The two families are orthogonal by construction and have total real dimension

$$(N-1)d^2 + 2d^2 = (N+1)d^2 = \dim_{\mathbb{R}} \mathcal{M}.$$

$\square$

The basis in Theorem 5.18 uses the endpoint conditions $a_0 = a_N = 0$ in each off-diagonal sector. We now restore the coefficients a_0 and a_N and keep $a_1, \ldots, a_{N-1}$ as interior variables.

Lemma 5.19. *Fix $1 \leq i < j \leq d$, and let $a = (a_0, \ldots, a_N) \in \mathbb{R}^{N+1}$. In the A_{ij}- or S_{ij}-sector, allowing the endpoint coefficients a_0 and a_N to vary gives the quadratic form*

$$a^\top \widetilde{h}_{ij}^{(N)} a,$$

where

$$\widetilde{h}_{ij}^{(N)} = \begin{pmatrix} \frac{1}{2}(\lambda_i^2 + \lambda_j^2) & -\lambda_i\lambda_j & & & \\ -\lambda_i\lambda_j & \lambda_i^2 + \lambda_j^2 & -\lambda_i\lambda_j & & \\ & \ddots & \ddots & \ddots & \\ & & -\lambda_i\lambda_j & \lambda_i^2 + \lambda_j^2 & -\lambda_i\lambda_j \\ & & & -\lambda_i\lambda_j & \frac{1}{2}(\lambda_i^2 + \lambda_j^2) \end{pmatrix}. \tag{5.33}$$

The restriction $a_0 = a_N = 0$ is the off-diagonal block in (5.28).

Proof. Repeat the calculation in Lemma 5.15 without imposing $a_0 = a_N = 0$. The interior coefficients appear twice in the sum over layers, while a_0 and a_N appear once. This gives the first and last diagonal entries in (5.33). Setting $a_0 = a_N = 0$ removes the first and last rows and columns, and the remaining block is the one in (5.28). $\square$

Proposition 5.20. *Fix $1 \le i < j \le d$. For the block $\widetilde{h}_{ij}^{(N)}$ in (5.33), the stationarity equations in the interior variables $a_1, \dots, a_{N-1}$ are*

$$-\lambda_i\lambda_j a_{k-1} + (\lambda_i^2 + \lambda_j^2)a_k - \lambda_i\lambda_j a_{k+1} = 0, \qquad 1 \le k \le N-1. \tag{5.34}$$

This recurrence has the two geometric solutions

$$a_{ij,k}^{(0)} := \lambda_i^k \lambda_j^{N-k}, \qquad a_{ij,k}^{(N)} := \lambda_i^{N-k}\lambda_j^k, \qquad 0 \le k \le N. \tag{5.35}$$

These two solutions give the horizontal lifts h_{ji}^{Σ} and h_{ij}^{Σ}, respectively.

Proof. Treat a_0 and a_N as fixed and differentiate $a^\top \widetilde{h}_{ij}^{(N)} a$ with respect to a_k for $1 \le k \le N-1$. The factor 2 from differentiating the quadratic form is irrelevant, and the resulting equations are exactly (5.34).

Substituting either sequence in (5.35) into (5.34) verifies the recurrence. If

$$(\Gamma_{ij}^{\Sigma})_k = c_k^{(ij)} E_{ij}, \qquad c_k^{(ij)} = \lambda_i^{N-k}\lambda_j^{k-1},$$

then, for $1 \le k \le N$,

$$a_{ij,k-1}^{(N)} = \lambda_i c_k^{(ij)}, \qquad a_{ij,k-1}^{(0)} = \lambda_j c_k^{(ji)}.$$

Thus $a_{ij}^{(N)}$ is the sequence of coefficients of Γ_{ij}^{Σ}, while $a_{ij}^{(0)}$ is the sequence of coefficients of Γ_{ji}^{Σ}. After normalization these are the horizontal vectors h_{ij}^{Σ} and h_{ji}^{Σ}. $\square$

Remark 5.21. The vertical modes satisfy the endpoint conditions $a_0 = a_N = 0$. Restoring those two coefficients enlarges the off-diagonal block to $\widetilde{h}_{ij}^{(N)}$, and the two geometric solutions in (5.35) give the missing horizontal directions.

If $\lambda_i = \lambda_j$, then

$$\widetilde{h}_{ij}^{(N)} = \lambda_i^2 \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}.$$

This matrix is singular, with kernel spanned by $(1, \ldots, 1)^\top$. Thus the enlarged off-diagonal block acquires a zero mode when singular values collide, as noted in [MY26b, Remark 12].

Corollary 5.22. *Let $X = Q\Sigma R^*$ have distinct singular values, and let*

$$\mathbf{W}^X = (Q\Lambda, \Lambda, \ldots, \Lambda, \Lambda R^*) \in \mathcal{M}.$$

Define

$$A_{Q,R}(W_N, \ldots, W_1) := (QW_N,\, W_{N-1},\, \ldots,\, W_2,\, W_1 R^*).$$

Then $A_{Q,R}$ is a complex-linear isometry and $A_{Q,R}(\mathbf{W}^\Sigma) = \mathbf{W}^X$. Applying $(\mathrm{d}A_{Q,R})_{\mathbf{W}^\Sigma}$ to the basis in Theorem 5.18 gives a g-orthonormal basis of $T_{\mathbf{W}^X}\mathcal{M}$. In particular, if h_{ij}^X denotes the transported horizontal vector, then

$$(\mathrm{d}\phi)_{\mathbf{W}^X}(h_{ij}^X) = \alpha_{ij}^{1/2} u_i v_j^*,$$

where u_i and v_j are the singular vectors of X.

Proof. The map $A_{Q,R}$ is a complex-linear isometry by Proposition 5.8. Hence its differential sends orthonormal bases to orthonormal bases. The formula for the projected horizontal vectors follows from (5.32) and bi-unitary invariance. $\square$

5.4.3 Riemannian submersion

The basis gives a concrete form of the submersion stated in Theorem 5.4: the vertical vectors span the kernel of $\mathrm{d}(\phi|_{\mathcal{M}})$, and the horizontal vectors project to an h_N-orthonormal complex basis of $T_X\mathrm{GL}(d,\mathbb{C})$.

Proof of Theorem 5.4. Let $\pi : \mathcal{M} \to \mathcal{M}/K$ be the quotient map. By Theorem 4.8, the map

$$\pi : (\mathcal{M}, \iota) \longrightarrow \mathcal{M}/K$$

is a Riemannian submersion for the reduced metric on $\mathcal{M}/K$. By Theorem 5.1, the metric transported from $\mathcal{M}/K$ to $\mathrm{GL}(d,\mathbb{C})$ through $\bar{\phi}$ has real part g_N. Hence

$$\bar{\phi} : \mathcal{M}/K \longrightarrow (\mathrm{GL}(d,\mathbb{C}), g_N)$$

is an isometry. Since

$$\phi|_{\mathcal{M}} = \bar{\phi} \circ \pi,$$

the map

$$\phi : (\mathcal{M}, \iota) \longrightarrow (\mathrm{GL}(d,\mathbb{C}), g_N)$$

is a Riemannian submersion. $\square$

5.5 Entropy and Ricci curvature

5.5.1 Entropy formula

We now return to the vertical directions in the adapted basis constructed in Section 5.4. The horizontal directions determine the metric on $\mathrm{GL}(d,\mathbb{C})$ obtained from the quotient. The vertical directions determine the metric induced on the orbit $\mathcal{O}_X$, and therefore its volume.

Recall that

$$S_N(X) = \log \operatorname{vol}(\mathcal{O}_X),$$

where the volume is computed using the metric induced from $\mathcal{M}$.

Let

$$X = Q\Sigma R^*, \qquad \Lambda = \Sigma^{1/N},$$

be a singular value decomposition of X, and set

$$\mathbf{W}^X := (Q\Lambda, \Lambda, \ldots, \Lambda, \Lambda R^*) \in \mathcal{O}_X.$$

By Proposition 4.6,

$$\mathcal{O}_X = K \cdot \mathbf{W}^X.$$

Proposition 5.23. *The map*

$$\mathfrak{y}_X : K \to \mathcal{O}_X, \qquad \mathfrak{y}_X(U) = U \cdot \mathbf{W}^X, \tag{5.36}$$

is a diffeomorphism.

Proof. Proposition 4.6 gives surjectivity. By Proposition 4.4, the stabilizer of $\mathbf{W}^X$ is trivial, so $\mathfrak{y}_X$ is injective.

At the identity,

$$(\mathrm{d}\mathfrak{y}_X)_e(\xi) = K_\xi(\mathbf{W}^X), \qquad \xi \in T_e K \cong \mathfrak{u}(d)^{N-1}.$$

If $(\mathrm{d}\mathfrak{y}_X)_e(\xi) = 0$, then $K_\xi(\mathbf{W}^X) = 0$. Freeness gives $\xi = 0$, so $(\mathrm{d}\mathfrak{y}_X)_e$ is injective.

For $V \in K$, let ℓ_V denote left translation on K and let L_V denote the action of V

on $\mathcal{O}_X$. Then

$$\mathfrak{y}_X \circ \ell_V = L_V \circ \mathfrak{y}_X.$$

Since ℓ_V and L_V are diffeomorphisms, injectivity of $(\mathrm{d}\mathfrak{y}_X)_e$ implies injectivity of $(\mathrm{d}\mathfrak{y}_X)_U$ for every $U \in K$. Thus $\mathfrak{y}_X$ is an immersion.

The domain K is compact and $\mathcal{O}_X$ is Hausdorff. Hence a bijective immersion $K \to \mathcal{O}_X$ is an embedding, and therefore a diffeomorphism. □

Let

$$\gamma_X := \mathfrak{y}_X^*(\iota|_{\mathcal{O}_X}),$$

where ι is the metric on $\mathcal{M}$ induced by the ambient Frobenius metric.

In Chapters 2 and 3, we used c_d for $\mathrm{vol}(\mathrm{O}(d))$. In the complex case we write

$$u_d := \mathrm{vol}\Big(\mathrm{U}(d)\Big),$$

where the volume is taken with respect to the bi-invariant metric induced by the Frobenius inner product on $\mathfrak{u}(d)$. Thus

$$\mathrm{vol}(K) = u_d^{N-1}.$$

Proposition 5.24. *The pullback metric γ_X is left-invariant. Consequently,*

$$\mathrm{vol}(\mathcal{O}_X) = u_d^{N-1} \det \gamma_X(e)^{1/2}, \tag{5.37}$$

where the determinant is computed in any basis of $T_eK \cong \mathfrak{u}(d)^{N-1}$ that is orthonormal for the product bi-invariant Frobenius metric on K.

Proof. For $V \in K$, let $\ell_V : K \to K$ be left translation and let $L_V : \mathcal{O}_X \to \mathcal{O}_X$ be the restriction of the action of K:

$$L_V(\mathbf{W}) = V \cdot \mathbf{W}.$$

Then

$$\mathfrak{y}_X \circ \ell_V = L_V \circ \mathfrak{y}_X.$$

By Proposition 4.1, the action of K is isometric for ι. Therefore

$$\ell_V^* \gamma_X = \ell_V^* \mathfrak{y}_X^* (\iota|_{\mathcal{O}_X}) = \mathfrak{y}_X^* L_V^* (\iota|_{\mathcal{O}_X}) = \mathfrak{y}_X^* (\iota|_{\mathcal{O}_X}) = \gamma_X.$$

Thus γ_X is left-invariant.

Choose a basis of T_eK that is orthonormal for the product bi-invariant Frobenius metric, and left-translate it to a global frame on K. Since both the product metric and γ_X are left-invariant, the matrix of γ_X in this frame is constant and equals the matrix of $\gamma_X(e)$. Integrating the resulting volume form gives (5.37). $\square$

Proposition 5.25. *In the basis*

$$\left\{\xi^A_{ij,m},\ \xi^S_{ij,m},\ \xi^D_{i,m}\right\}$$

of T_eK corresponding to the vertical part of the adapted basis, the metric $\gamma_X(e)$ is diagonal. Its diagonal entries are

$$\gamma_X(e)(\xi^A_{ij,m}, \xi^A_{ij,m}) = \mu_{ij,m}, \qquad \gamma_X(e)(\xi^S_{ij,m}, \xi^S_{ij,m}) = \mu_{ij,m},$$

and

$$\gamma_X(e)(\xi^D_{i,m}, \xi^D_{i,m}) = \nu_{i,m}.$$

Consequently,

$$\det \gamma_X(e) = \prod_{i=1}^{d} \prod_{m=1}^{N-1} \nu_{i,m} \prod_{1 \le i < j \le d} \prod_{m=1}^{N-1} \mu_{ij,m}^2. \tag{5.38}$$

Proof. By the definition of the pullback metric,

$$\gamma_X(e)(\xi, \eta) = \iota_{\mathbf{W}^X}\Big(K_\xi(\mathbf{W}^X), K_\eta(\mathbf{W}^X)\Big).$$

The map in Proposition 5.8 sends $\mathbf{W}^\Sigma$ to $\mathbf{W}^X$, is an isometry, and commutes with the action of K. Hence the Gram matrix at $\mathbf{W}^X$ is the same as the Gram matrix at $\mathbf{W}^\Sigma$.

Lemma 5.15 diagonalizes the vertical blocks at $\mathbf{W}^\Sigma$. In the notation of Section 5.4,

$$C_{N-1} s_m = \kappa_m s_m, \qquad \kappa_m = 4 \sin^2 \frac{m\pi}{2N}.$$

Hence the eigenvalues of the A_{ij}- and S_{ij}-blocks are

$$(\lambda_i - \lambda_j)^2 + \lambda_i \lambda_j \kappa_m = \lambda_i^2 + \lambda_j^2 - 2\lambda_i\lambda_j \cos \frac{m\pi}{N} = \mu_{ij,m},$$

and the eigenvalues of the diagonal phase block are

$$\lambda_i^2 \kappa_m = \nu_{i,m}.$$

This proves the diagonalization. Taking the product of the diagonal entries gives (5.38). $\square$

Lemma 5.26. *The eigenvalues in Proposition 5.25 satisfy*

$$\prod_{m=1}^{N-1} \nu_{i,m} = N\lambda_i^{2(N-1)} = N\sigma_i^{\frac{2(N-1)}{N}}, \tag{5.39}$$

$$\prod_{m=1}^{N-1} \mu_{ij,m} = \sum_{\ell=0}^{N-1} \lambda_i^{2(N-1-\ell)} \lambda_j^{2\ell}. \tag{5.40}$$

When $\lambda_i \neq \lambda_j$, the second identity is equivalently

$$\prod_{m=1}^{N-1} \mu_{ij,m} = \frac{\lambda_i^{2N} - \lambda_j^{2N}}{\lambda_i^2 - \lambda_j^2} = \frac{\sigma_i^2 - \sigma_j^2}{\sigma_i^{2/N} - \sigma_j^{2/N}}. \tag{5.41}$$

At $\lambda_i = \lambda_j$, the quotient in (5.41) *is interpreted through the polynomial extension in* (5.40).

Proof. For (5.39), use

$$\nu_{i,m} = \lambda_i^2 \kappa_m.$$

The numbers κ_m are the eigenvalues of C_{N-1}, and

$$\prod_{m=1}^{N-1} \kappa_m = \det C_{N-1} = N.$$

Therefore

$$\prod_{m=1}^{N-1} \nu_{i,m} = \lambda_i^{2(N-1)} \prod_{m=1}^{N-1} \kappa_m = N\lambda_i^{2(N-1)}.$$

Since $\lambda_i^N = \sigma_i$, this proves (5.39).

For (5.40), write

$$\mu_{ij,m} = \lambda_i^2 - 2\lambda_i\lambda_j \cos\frac{m\pi}{N} + \lambda_j^2.$$

The identity

$$\prod_{m=1}^{N-1} \left(z^2 - 2z\cos\frac{m\pi}{N} + 1\right) = \frac{z^{2N} - 1}{z^2 - 1} \tag{5.42}$$

holds as a polynomial identity, with the right-hand side interpreted by its polynomial extension at $z = \pm 1$. Substituting $z = \lambda_i/\lambda_j$ and multiplying by $\lambda_j^{2(N-1)}$ gives

$$\prod_{m=1}^{N-1} \mu_{ij,m} = \frac{\lambda_i^{2N} - \lambda_j^{2N}}{\lambda_i^2 - \lambda_j^2}$$

when $\lambda_i \neq \lambda_j$. The algebraic identity

$$\frac{a^{2N} - b^{2N}}{a^2 - b^2} = \sum_{\ell=0}^{N-1} a^{2(N-1-\ell)} b^{2\ell}$$

gives the polynomial extension. Finally, $\lambda_i^N = \sigma_i$ gives (5.41). $\square$

Proof of Theorem 5.5. By Proposition 5.24,

$$\operatorname{vol}(\mathcal{O}_X) = u_d^{N-1} \det \gamma_X(e)^{1/2}.$$

By Proposition 5.25 and Lemma 5.26,

$$\begin{aligned} \det \gamma_X(e) &= \prod_{i=1}^{d} \prod_{m=1}^{N-1} \nu_{i,m} \prod_{1 \le i < j \le d} \prod_{m=1}^{N-1} \mu_{ij,m}^2 \\ &= N^d \det(\Sigma^2)^{\frac{N-1}{N}} \left(\frac{\operatorname{van}(\Sigma^2)}{\operatorname{van}(\Sigma^{2/N})} \right)^2 . \end{aligned} \tag{5.43}$$

On the locus of distinct singular values, the ratio $\operatorname{van}(\Sigma^2)/\operatorname{van}(\Sigma^{2/N})$ is positive because the map $t \mapsto t^{1/N}$ is strictly increasing on $(0, \infty)$. Taking the positive square root in (5.43) gives (5.7). Taking logarithms gives (5.8).

It remains to justify the extension across repeated singular values. In the variables $\lambda_i = \sigma_i^{1/N}$, each factor

$$\frac{\lambda_i^{2N} - \lambda_j^{2N}}{\lambda_i^2 - \lambda_j^2}$$

is the polynomial

$$\sum_{\ell=0}^{N-1} \lambda_i^{2(N-1-\ell)} \lambda_j^{2\ell}.$$

This polynomial is positive for $\lambda_i, \lambda_j > 0$. Hence the ratio in (5.7) has a positive continuous extension through collisions of singular values. Formula (5.43), and therefore the entropy formula, extend by continuity. $\square$

The same coefficients α_{ij} that determine the Hermitian metric on $\mathrm{GL}(d,\mathbb{C})$ also determine the determinant of $\mathcal{A}_{N,X}^{-1}$. Since

$$h_N(X)(Z_1,Z_2) = \mathrm{tr}\Big(Z_1^*\mathcal{A}_{N,X}^{-1}(Z_2)\Big),$$

we compare the volume of $\mathcal{O}_X$ with the determinant of the complex-linear operator

$$\mathcal{A}_{N,X}^{-1} : \mathrm{M}_d(\mathbb{C}) \to \mathrm{M}_d(\mathbb{C}).$$

By Corollary 5.11,

$$\det(\mathcal{A}_{N,X}^{-1}) = \prod_{i,j=1}^{d} \alpha_{ij}^{-1}. \tag{5.44}$$

Proof of Corollary 5.6. By Corollary 5.11, the basis associated with a singular value decomposition of X diagonalizes $\mathcal{A}_{N,X}$ with eigenvalues α_{ij}. Therefore $\mathcal{A}_{N,X}^{-1}$ has eigenvalues α_{ij}^{-1}, and (5.44) holds.

When the singular values are distinct, (5.20) gives

$$\det(\mathcal{A}_{N,X}^{-1}) = N^{-d}\det(\Sigma^2)^{\frac{1}{N}-1}\left(\frac{\mathrm{van}(\Sigma^{2/N})}{\mathrm{van}(\Sigma^2)}\right)^2.$$

Multiplying this identity by the square of (5.7) gives the square of (5.9). Both sides of (5.9) are positive, so (5.9) holds on the locus of distinct singular values.

The two sides of (5.9) are continuous on $\mathrm{GL}(d,\mathbb{C})$, and the locus of matrices with distinct singular values is dense. Thus (5.9) holds for every $X \in \mathrm{GL}(d,\mathbb{C})$. Taking logarithms gives (5.10). $\square$

5.5.2 Ricci curvature

We now use the determinant identity (5.10) to derive the Ricci form. For a Kähler metric h with Kähler form ω, we use the convention

$$\rho_h = -\mathrm{i}\,\partial\bar{\partial}\log\det(h_{a\bar{b}})$$

for the Ricci form in a local holomorphic frame. If $g = \Re h$ is the underlying Riemannian metric, then

$$\mathrm{Ric}(U, V) = \rho_h(U, JV).$$

A local real-valued function f satisfying

$$\rho_h = \mathrm{i}\,\partial\bar{\partial} f$$

is a Ricci potential.

Because $\mathrm{GL}(d, \mathbb{C})$ is an open subset of the complex vector space $\mathrm{M}_d(\mathbb{C})$, the constant vector fields $\{E_{ij}\}_{i,j}$ form a global holomorphic frame of $T^{1,0}\mathrm{GL}(d, \mathbb{C})$. In this frame define

$$\det h_N(X) := \det\Big(h_N(X)(E_{ij}, E_{k\ell})\Big)_{(ij),(k\ell)}. \tag{5.45}$$

Lemma 5.27. *For every $X \in \mathrm{GL}(d, \mathbb{C})$,*

$$\det h_N(X) = \det(\mathcal{A}_{N,X}^{-1}). \tag{5.46}$$

Proof. The matrix units are orthonormal for the Frobenius Hermitian form on $\mathrm{M}_d(\mathbb{C})$.

By Theorem 5.1,

$$h_N(X)(E_{ij}, E_{k\ell}) = \operatorname{tr}\Big(E_{ij}^* \mathcal{A}_{N,X}^{-1}(E_{k\ell})\Big).$$

Thus the Hermitian matrix of $h_N(X)$ in the frame $\{E_{ij}\}$ is the matrix of the positive operator $\mathcal{A}_{N,X}^{-1}$ in the same orthonormal basis. Their determinants agree. $\square$

Proof of Corollary 5.7. By the definition of the Ricci form and by Lemma 5.27,

$$\rho_N = -\mathrm{i}\,\partial\bar{\partial} \log \det h_N = -\mathrm{i}\,\partial\bar{\partial} \log \det(\mathcal{A}_{N,X}^{-1}).$$

By Corollary 5.6,

$$-\log \det(\mathcal{A}_{N,X}^{-1}) = 2S_N - 2(N-1)\log u_d.$$

The final term is constant. Applying $\mathrm{i}\,\partial\bar{\partial}$ gives (5.11). Hence $2S_N$ is a Ricci potential.

$\square$

Remark 5.28. The entropy formula in Theorem 5.5 splits as

$$S_N(X) = (N-1)\log u_d + \frac{d}{2}\log N + \frac{N-1}{2N}\log\det(X^*X) + \sum_{1\le i<j\le d} \log\left(\frac{\sigma_i^2 - \sigma_j^2}{\sigma_i^{2/N} - \sigma_j^{2/N}}\right),$$

where $\sigma_1, \ldots, \sigma_d$ are the singular values of X, and each quotient is interpreted by the extension described in Remark 5.29 when singular values coincide. The determinant term does not contribute to the Ricci form. Indeed, locally on $\mathrm{GL}(d, \mathbb{C})$ one can choose a branch of $\log\det X$, and then

$$\log\det(X^*X) = \log\det X + \overline{\log\det X}.$$

Thus $\log\det(X^*X)$ is pluriharmonic. The Ricci form is therefore determined by the pairwise singular-value term.

Remark 5.29. At a collision of singular values, every quotient of the form

$$\frac{\sigma_i^2 - \sigma_j^2}{\sigma_i^{2/N} - \sigma_j^{2/N}}$$

is interpreted through the polynomial extension

$$\sum_{\ell=0}^{N-1} \lambda_i^{2(N-1-\ell)} \lambda_j^{2\ell}, \qquad \lambda_i = \sigma_i^{1/N}, \quad \lambda_j = \sigma_j^{1/N}.$$

With this convention, (5.7), (5.8), and (5.10) hold on all of $\mathrm{GL}(d, \mathbb{C})$.

Remark 5.30. The complex entropy formula above is the $\mathbb{F} = \mathbb{C}$ member of a uniform family over the real division algebras [CN26]. Let

$$\mathbb{F} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}, \qquad \beta = \dim_{\mathbb{R}} \mathbb{F} \in \{1, 2, 4\}.$$

Let

$$K_{\mathbb{F},d} = \begin{cases} O(d), & \mathbb{F} = \mathbb{R}, \\ U(d), & \mathbb{F} = \mathbb{C}, \\ \mathrm{Sp}(d), & \mathbb{F} = \mathbb{H}, \end{cases}$$

and write

$$c_{\mathbb{F},d} = \mathrm{vol}(K_{\mathbb{F},d}),$$

with the Haar-volume normalization induced by the corresponding Frobenius metric. If $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_d)$ is the positive diagonal matrix of singular values of X, then the entropy of the balanced compact orbit in the full-rank square DLN over $\mathbb{F}$ is

$$S_N^{(\beta)}(X) = (N-1) \log c_{\mathbb{F},d} + \frac{\beta}{2} \log \frac{\mathrm{van}(\Sigma^2)}{\mathrm{van}(\Sigma^{2/N})} + (\beta - 1) \left(\frac{d}{2} \log N + \log \frac{\det \Sigma}{\det(\Sigma^{1/N})} \right).$$

Bibliographical notes

This chapter is the complex counterpart of the real quotient geometry developed in Chapters 2 and 3. The operator $\mathcal{A}_{N,X}$ and the metric g_N on $\mathrm{GL}(d,\mathbb{C})$ coincide with the tensors that arise in the real setting. The new ingredient here is the Kähler framework, obtained from the moment-map description of Chapter 4. It is natural to ask whether analogous structures exist in the real and quaternionic settings; this remains open.

The entropy and submersion formulas in [MY26b, Theorem 4 and Theorem 8] provide the real precursors of Section 5.5. The complex entropy formula proved here has the same Vandermonde structure and corresponds to the $\beta = 2$ case of the real, complex, and quaternionic family [CN26].

Chapter 6

Geometry of a Fibre $\mathcal{F}_X$

The balanced quotient satisfies $\mathcal{M}/K \cong \mathrm{GL}(d, \mathbb{C})$. In this chapter, we fix $X \in \mathrm{GL}(d, \mathbb{C})$ and study the fibre $\mathcal{F}_X = \phi^{-1}(X)$. For a factorization $\mathbf{W} = (W_N, \ldots, W_1)$, set $Y_k = W_k \cdots W_1$ and $H_k = Y_k^* Y_k$. Then $H_0 = \mathrm{Id}$ and $H_N = X^* X$, and $K = \mathrm{U}(d)^{N-1}$ acts by unitary changes of intermediate coordinates. The collection $(H_1, \ldots, H_{N-1})$ identifies the quotient $\mathcal{F}_X/K$ with $\mathbb{P}_d^{N-1}$. Under this identification, the balanced condition means that the sequence (H_k) consists of the points at times k/N along the geodesic, with respect to the trace metric on the positive-definite cone, from Id to $X^* X$.

6.1 Main results

Throughout this chapter $X \in \mathrm{GL}(d, \mathbb{C})$ is fixed. For $\mathbf{W} \in \mathcal{F}_X$, set $Y_k = W_k \cdots W_1$ and $H_k = Y_k^* Y_k$. We also set $H_0 = \mathrm{Id}$ and $H_N = X^* X$.

Let $\mathbb{P}_d = \{H \in \mathrm{Herm}(d) : H > 0\}$ be the cone of positive-definite Hermitian matrices.

In this chapter, we identify $\mathcal{F}_X/K \cong \mathbb{P}_d^{N-1}$, rewrite the moment equations in these coordinates, characterize the balanced path as the geodesic from Id to $X^* X$ for the trace metric, and compute the ridge functional on the quotient.

The quotient of a fixed fibre. Define

$$q_X : \mathcal{F}_X \longrightarrow \mathbb{P}_d^{N-1}, \qquad q_X(\mathbf{W}) = (H_1, \ldots, H_{N-1}). \tag{6.1}$$

Proposition 6.1. *The map q_X is constant on K-orbits and descends to a diffeomorphism*

$$\mathcal{F}_X/K \cong \mathbb{P}_d^{N-1}. \tag{6.2}$$

Thus, modulo unitary changes of intermediate coordinates, a factorization of X is the same as a sequence $\mathrm{Id} = H_0,\ H_1, \ldots, H_{N-1},\ H_N = X^*X$ in $\mathbb{P}_d$ with fixed endpoints.

Level sets of the moment map in a fibre. Recall that the moment map is

$$\mu(\mathbf{W}) = (G_{N-1}(\mathbf{W}), \ldots, G_1(\mathbf{W})), \qquad G_k(\mathbf{W}) = W_k W_k^* - W_{k+1}^* W_{k+1}. \tag{6.3}$$

For a prescribed value

$$\mathbf{G} = (G_{N-1}, \ldots, G_1) \in \mathrm{Herm}(d)^{N-1},$$

set

$$\mathcal{M}_{\mathbf{G}} := \mu^{-1}(\mathbf{G}) = \Big\{ \mathbf{W} \in \mathrm{GL}(d, \mathbb{C})^N : G_k(\mathbf{W}) = G_k \text{ for } 1 \le k \le N-1 \Big\}. \tag{6.4}$$

For $X \in \mathrm{GL}(d, \mathbb{C})$, set

$$\mathcal{O}_{X,\mathbf{G}} := \mathcal{F}_X \cap \mathcal{M}_{\mathbf{G}}. \tag{6.5}$$

When $\mathbf{G} = 0$, this is the group orbit $\mathcal{O}_X$, namely the set of balanced factorizations over X:

$$\mathcal{O}_{X,0} = \mathcal{O}_X = \mathcal{F}_X \cap \mathcal{M}. \tag{6.6}$$

For a general value $\mathbf{G}$, the full group K need not preserve $\mathcal{M}_{\mathbf{G}}$. The subgroup that always preserves it is

$$K_{\mathbf{G}} := \prod_{k=1}^{N-1} \mathrm{Stab}_{\mathrm{U}(d)}(G_k), \qquad \mathrm{Stab}_{\mathrm{U}(d)}(G_k) := \{U \in \mathrm{U}(d) : UG_kU^* = G_k\}. \tag{6.7}$$

Proposition 6.2. *The moment map*

$$\mu : \mathrm{GL}(d, \mathbb{C})^N \longrightarrow \mathrm{Herm}(d)^{N-1}$$

is a surjective submersion. Consequently every value of μ is regular, and for every $\mathbf{G} \in \mathrm{Herm}(d)^{N-1}$, the level set $\mathcal{M}_{\mathbf{G}}$ is a smooth submanifold of $\mathrm{GL}(d, \mathbb{C})^N$ of real dimension $(N+1)d^2$.

Proposition 6.3. *The group $K_{\mathbf{G}}$ acts freely on $\mathcal{M}_{\mathbf{G}}$. Moreover, for every $U \in K_{\mathbf{G}}$ and every $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$,*

$$\phi(U \cdot \mathbf{W}) = \phi(\mathbf{W}).$$

Hence $K_{\mathbf{G}}$ preserves $\mathcal{O}_{X,\mathbf{G}}$ for every $X \in \mathrm{GL}(d, \mathbb{C})$.

The reduced moment equations. We now write the moment equations in the quotient coordinates. Let $Y_k = V_k H_k^{1/2}$ be the polar decomposition of the kth partial product. For $1 \le k \le N-1$, define

$$\Phi_k(H) := H_k^{1/2} H_{k-1}^{-1} H_k^{1/2} - H_k^{-1/2} H_{k+1} H_k^{-1/2}. \tag{6.8}$$

This depends only on the three consecutive terms H_{k-1}, H_k, and H_{k+1}.

Proposition 6.4. *For every* $\mathbf{W} \in \mathcal{F}_X$ *and every* $1 \le k \le N-1$,

$$G_k(\mathbf{W}) = V_k \Phi_k(H) V_k^*. \tag{6.9}$$

For $A \in \mathrm{Herm}(d)$, write

$$\mathrm{Ad}_{\mathrm{U}(d)}(A) := \{UAU^* : U \in \mathrm{U}(d)\}.$$

Define the reduced moment set

$$R_{X,\mathbf{G}} := q_X(\mathcal{O}_{X,\mathbf{G}}) \subset \mathbb{P}_d^{N-1}.$$

Theorem 6.5. *For every* $X \in \mathrm{GL}(d, \mathbb{C})$ *and every*

$$\mathbf{G} = (G_{N-1}, \ldots, G_1) \in \mathrm{Herm}(d)^{N-1},$$

the image of $\mathcal{O}_{X,\mathbf{G}}$ *under the quotient map* q_X *is*

$$R_{X,\mathbf{G}} = \Big\{ H \in \mathbb{P}_d^{N-1} : \Phi_k(H) \in \mathrm{Ad}_{\mathrm{U}(d)}(G_k) \text{ for every } 1 \le k \le N-1 \Big\}. \tag{6.10}$$

Moreover, the fibres of the restricted map

$$q_X\Big|_{\mathcal{O}_{X,\mathbf{G}}} : \mathcal{O}_{X,\mathbf{G}} \longrightarrow R_{X,\mathbf{G}}$$

are precisely the $K_{\mathbf{G}}$*-orbits.*

For fixed X and $\mathbf{G}$, the slice $\mathcal{O}_{X,\mathbf{G}}$ may be empty. In that case $R_{X,\mathbf{G}} := q_X(\mathcal{O}_{X,\mathbf{G}})$ is empty. The surjectivity of μ ensures that $\mathcal{M}_{\mathbf{G}}$ itself is nonempty and smooth. It does not imply that every fixed fibre $\mathcal{F}_X$ meets $\mathcal{M}_{\mathbf{G}}$.

The path selected by balancedness. Introduce the multiplicative increments $\Pi_k := H_{k-1}^{-1}H_k$, $1 \le k \le N$. These matrices need not be Hermitian, but each is similar to a positive-definite Hermitian matrix.

Lemma 6.6. *For every* $1 \le k \le N-1$,

$$\Phi_k(H) = H_k^{1/2}(\Pi_k - \Pi_{k+1})H_k^{-1/2}. \tag{6.11}$$

Consequently,

$$\Phi_k(H) = 0 \qquad \Longleftrightarrow \qquad \Pi_k = \Pi_{k+1}.$$

Proposition 6.7. *For* $H \in \mathbb{P}_d^{N-1}$*, the following are equivalent:*

1. $\Phi_k(H) = 0$ *for every* $1 \le k \le N-1$.

2. $\Pi_1 = \Pi_2 = \cdots = \Pi_N$*, where* $\Pi_k = H_{k-1}^{-1}H_k$.

Corollary 6.8. *For* $\mathbf{G} = 0$,

$$R_{X,0} = \Big\{\Big((X^*X)^{1/N}, \ldots, (X^*X)^{(N-1)/N}\Big)\Big\}.$$

Equivalently, the unique path corresponding to balancedness is

$$H_k = (X^*X)^{k/N}, \qquad 0 \le k \le N. \tag{6.12}$$

In particular,

$$\mathcal{O}_{X,0} = \mathcal{O}_X$$

is a single K*-orbit.*

We equip $\mathbb{P}_d$ with the trace metric

$$g_H^{\mathrm{tr}}(A,B) := \mathrm{tr}\Big(H^{-1}AH^{-1}B\Big), \qquad A,B \in T_H\mathbb{P}_d \cong \mathrm{Herm}(d).$$

Proposition 6.9 ([Bha07])**.** *Let $A, B \in \mathbb{P}_d$. The geodesic of the trace metric joining A to B is*

$$\gamma(t) = A^{1/2}\Big(A^{-1/2}BA^{-1/2}\Big)^t A^{1/2}, \qquad 0 \le t \le 1.$$

Along this geodesic,

$$\frac{d}{dt}\Big(\gamma^{-1}\dot{\gamma}\Big) = 0.$$

Corollary 6.10. *The path corresponding to balancedness*

$$H_k = (X^*X)^{k/N}, \qquad 0 \le k \le N,$$

is obtained by sampling the geodesic of the trace metric from Id *to X^*X at the times $t = k/N$.*

The ridge functional and the Kähler potential. Let

$$\mathcal{R}(\mathbf{W}) := \frac{1}{2}\sum_{k=1}^{N} \mathrm{tr}(W_k^* W_k)$$

be the ridge functional on the fixed fibre.

Proposition 6.11. *For every $\mathbf{W} \in \mathcal{F}_X$ with $H = q_X(\mathbf{W})$,*

$$\mathcal{R}(\mathbf{W}) = \frac{1}{2}\sum_{k=1}^{N} \mathrm{tr}\Big(H_k H_{k-1}^{-1}\Big). \tag{6.13}$$

Here $H_0 = \mathrm{Id}$ and $H_N = X^*X$. In particular, $\mathcal{R}$ is constant on K-orbits and descends *to the smooth function*

$$\mathcal{R}_X : \mathbb{P}_d^{N-1} \longrightarrow \mathbb{R}, \qquad \mathcal{R}_X(H_1, \ldots, H_{N-1}) = \frac{1}{2} \sum_{k=1}^{N} \operatorname{tr}\Big(H_k H_{k-1}^{-1}\Big).$$

Proposition 6.12. *The critical points of*

$$\mathcal{R}_X(H_1, \ldots, H_{N-1}) = \frac{1}{2} \sum_{k=1}^{N} \operatorname{tr}\Big(H_k H_{k-1}^{-1}\Big)$$

with $H_0 = \mathrm{Id}$ *and* $H_N = X^*X$ *are precisely the paths satisfying*

$$H_{k-1}^{-1} H_k = H_k^{-1} H_{k+1}, \qquad 1 \leq k \leq N-1.$$

Hence the unique critical point of $\mathcal{R}_X$ *is the path* $H_k = (X^*X)^{k/N}$, $0 \leq k \leq N$. *This path is the unique minimizer of* $\mathcal{R}_X$ *on* $\mathbb{P}_d^{N-1}$.

Corollary 6.13. *On the path* $H_k = (X^*X)^{k/N}$, $0 \leq k \leq N$, *one has*

$$\mathcal{R}_X(H_1, \ldots, H_{N-1}) = \frac{N}{2} \operatorname{tr}\Big((X^*X)^{1/N}\Big). \tag{6.14}$$

Consequently,

$$\min_{\mathbf{W} \in \mathcal{F}_X} \mathcal{R}(\mathbf{W}) = \frac{N}{2} \operatorname{tr}\Big((X^*X)^{1/N}\Big) = \Psi_N(X). \tag{6.15}$$

Barrier, entropy, and Ricci curvature. On the positive-definite cone $\mathbb{P}_d$, consider the logarithmic barrier

$$B : \mathbb{P}_d \longrightarrow \mathbb{R}, \qquad B(H) := -\log\det H. \tag{6.16}$$

For $H \in \mathbb{P}_d$ and $U, V \in T_H\mathbb{P}_d \cong \mathrm{Herm}(d)$, one has

$$\mathrm{D}^2 B_H(U,V) = \mathrm{tr}(H^{-1}UH^{-1}V).$$

This is exactly the trace metric on $\mathbb{P}_d$:

$$g_H^{\mathrm{tr}}(U,V) = \mathrm{tr}(H^{-1}UH^{-1}V).$$

Thus the trace metric is the Hessian metric associated with the logarithmic barrier. In the coordinates on the quotient $\mathcal{F}_X/K$, the path corresponding to balancedness is the geodesic of the trace metric joining Id to X^*X.

Together with Chapter 5, we therefore have three scalar potentials:

$$B(H) = -\log\det H, \qquad \Psi_N(X) = \frac{N}{2}\,\mathrm{tr}\Big((X^*X)^{1/N}\Big), \qquad S_N(X) = \log\mathrm{vol}(\mathcal{O}_X). \tag{6.17}$$

They play different roles. The barrier B generates the trace metric on $\mathbb{P}_d$. The function Ψ_N generates the Kähler form of the quotient metric on $\mathrm{GL}(d,\mathbb{C})$:

$$\omega_N = i\,\partial\bar{\partial}\Psi_N.$$

The entropy S_N generates the Ricci form of the same quotient metric:

$$\rho_N = 2i\,\partial\bar{\partial}S_N. \tag{6.18}$$

Here ρ_N is the Ricci form of h_N. Equivalently, $2S_N$ is a Ricci potential for h_N.

This separation is worth comparing with barrier geometry in conic optimization. In optimization theory, the Hessian of a barrier function defines a Riemannian metric that governs the local geometry used by Newton and interior-point methods [NN94; NT97; Gül96]. Hildebrand's canonical barrier construction goes further. For a regular convex cone, the canonical barrier is characterized, after normalization, by a Monge–Ampère equation of the form

$$\log\det D^2F = 2F + \text{constant}.$$

Since the Hessian of F defines the metric and the Ricci form is obtained from $-i\,\partial\bar{\partial}\log\det D^2F$, this equation implies that the associated metric is Kähler–Einstein [Hil14]. In that setting, a single potential determines both the metric and the Ricci form.

In the DLN, the metric and the Ricci form are not determined by a single potential. To see the obstruction, write

$$H = X^*X.$$

The entropy formula from Chapter 5 has the form

$$S_N(X) = C_{N,d} + \frac{N-1}{2N}\log\det H + \mathcal{I}_N(X),$$

where

$$\mathcal{I}_N(X) := \sum_{i<j}\log\left(\frac{\sigma_i^2 - \sigma_j^2}{\sigma_i^{2/N} - \sigma_j^{2/N}}\right).$$

The term $\log\det H$ is the contribution from the barrier. But locally on $\mathrm{GL}(d,\mathbb{C})$,

$$\log\det(X^*X) = \log\det X + \overline{\log\det X},$$

so

$$i\,\partial\bar{\partial}\log\det(X^*X) = 0.$$

Therefore the Ricci form is determined entirely by the singular-value interaction:

$$\rho_N = 2i\,\partial\bar{\partial}\mathcal{I}_N.$$

Consequently, the cone barrier controls the trace geometry of the fixed-fibre quotient, but it is invisible to the Ricci form on the space of end-to-end matrices. The curvature of the DLN quotient metric comes from the interaction term $\mathcal{I}_N$. If h_N were Kähler–Einstein, then locally one would have

$$2S_N - \lambda\Psi_N$$

pluriharmonic for some constant λ, since $\rho_N = \lambda\omega_N$. The entropy formula shows that any such identity must arise from the interaction term, not from the logarithmic barrier. This is the obstruction to a Kähler–Einstein interpretation. The metric is generated by Ψ_N, while the Ricci form is generated by S_N, and these do not coincide. In contrast, in the canonical barrier setting of Hildebrand, a single potential determines both the metric and the Ricci form, which yields a Kähler–Einstein metric.

Coordinates from the left endpoint. Define the left-end partial products

$$Z_k := W_N \cdots W_{k+1}, \qquad 0 \le k \le N, \tag{6.19}$$

with $Z_0 = X$ and $Z_N = \mathrm{Id}$, and set

$$\widetilde{H}_k := Z_k Z_k^*, \qquad 0 \le k \le N. \tag{6.20}$$

Then

$$\widetilde{H}_0 = XX^*, \qquad \widetilde{H}_N = \mathrm{Id}\,. \tag{6.21}$$

Proposition 6.14. *For every* $\mathbf{W} \in \mathcal{F}_X$ *and every* $0 \le k \le N$*,*

$$\widetilde{H}_k = X H_k^{-1} X^*. \tag{6.22}$$

Consequently, the left-end variables give the same quotient $\mathcal{F}_X/K$ *as the variables* $(H_1, \dots, H_{N-1})$*, with the two endpoint conventions related by* (6.22)*.*

Corollary 6.15. *In the left-end variables, the ridge functional descends to*

$$\mathcal{R}_X = \frac{1}{2} \sum_{k=1}^{N} \operatorname{tr}\Big(\widetilde{H}_{k-1} \widetilde{H}_k^{-1}\Big), \qquad \widetilde{H}_0 = XX^*, \quad \widetilde{H}_N = \mathrm{Id}\,. \tag{6.23}$$

On the balanced path,

$$\widetilde{H}_k = (XX^*)^{1-k/N}, \qquad 0 \le k \le N. \tag{6.24}$$

Thus the left-end variables sample the geodesic of the trace metric from XX^* *to* Id *at the times* $t = k/N$*.*

6.2 Level sets of the moment map

Recall that $\mathbf{G} = (G_{N-1}, \dots, G_1)$ denotes a prescribed value of the moment map, $\mathcal{M}_{\mathbf{G}} = \mu^{-1}(\mathbf{G})$, and

$$\mathcal{O}_{X,\mathbf{G}} = \mathcal{F}_X \cap \mathcal{M}_{\mathbf{G}}.$$

The quotient $\mathcal{F}_X/K$ removes the unitary factors at the intermediate layers. The level set $\mathcal{M}_{\mathbf{G}}$, by contrast, fixes the values of the moment map. For $\mathbf{G} = 0$, these two sets meet in the group orbit $\mathcal{O}_X$. For a general value $\mathbf{G}$, only the stabilizer $K_{\mathbf{G}}$ is forced to preserve the level set $\mathcal{M}_{\mathbf{G}}$.

Proof of Proposition 6.2. We first prove that μ is onto. Fix

$$\mathbf{G} = (G_{N-1}, \ldots, G_1) \in \mathrm{Herm}(d)^{N-1}.$$

Choose $c > 0$ large enough that the Hermitian matrices

$$A_N := c\,\mathrm{Id}, \qquad A_k := c\,\mathrm{Id} + \sum_{\ell=k}^{N-1} G_\ell, \qquad 1 \leq k \leq N-1,$$

are all positive definite. Set

$$W_k := A_k^{1/2}, \qquad 1 \leq k \leq N,$$

and let $\mathbf{W} = (W_N, \ldots, W_1)$. Then each W_k is invertible, and for $1 \leq k \leq N-1$,

$$G_k(\mathbf{W}) = W_k W_k^* - W_{k+1}^* W_{k+1} = A_k - A_{k+1} = G_k.$$

Hence $\mu(\mathbf{W}) = \mathbf{G}$.

It remains to prove that $(\mathrm{d}\mu)_{\mathbf{W}}$ is surjective at every $\mathbf{W} \in \mathrm{GL}(d, \mathbb{C})^N$. Suppose

$$\xi = (\xi_{N-1}, \ldots, \xi_1) \in \mathfrak{u}(d)^{N-1}$$

annihilates $\mathrm{im}(\mathrm{d}\mu)_{\mathbf{W}}$. Then

$$\mathrm{d}\langle \mu, \, \xi \rangle_{\mathbf{W}} = 0.$$

By the identity for the moment map,

$$\iota_{K_\xi} \omega \Big|_{\mathbf{W}} = \mathrm{d}\langle \mu, \, \xi \rangle_{\mathbf{W}}.$$

Thus $\iota_{K_\xi}\omega|_{\mathbf{W}} = 0$. Since ω is nondegenerate,

$$K_\xi(\mathbf{W}) = 0.$$

The fundamental vector field has components

$$\Big(K_\xi(\mathbf{W})\Big)_k = \xi_k W_k - W_k \xi_{k-1}, \qquad 1 \leq k \leq N,$$

where $\xi_0 = \xi_N = 0$. Since W_1 is invertible, the equation

$$\xi_1 W_1 = 0$$

gives $\xi_1 = 0$. Inductively, if $\xi_{k-1} = 0$, then

$$\xi_k W_k - W_k \xi_{k-1} = 0$$

gives $\xi_k = 0$. Hence $\xi = 0$.

Therefore the annihilator of $\operatorname{im}(\mathrm{d}\mu)_{\mathbf{W}}$ is trivial. Since the target is finite dimensional and the pairing between $\operatorname{Herm}(d)^{N-1}$ and $\mathfrak{u}(d)^{N-1}$ is nondegenerate, $(\mathrm{d}\mu)_{\mathbf{W}}$ is surjective. Thus μ is a surjective submersion.

The dimension follows from the regular-value theorem:

$$\dim_{\mathbb{R}} \mathcal{M}_{\mathbf{G}} = \dim_{\mathbb{R}} \mathrm{GL}(d, \mathbb{C})^N - \dim_{\mathbb{R}} \operatorname{Herm}(d)^{N-1} = 2Nd^2 - (N-1)d^2 = (N+1)d^2.$$

$\square$

Proof of Proposition 6.3. Let

$$U = (U_{N-1}, \dots, U_1) \in K_{\mathbf{G}}.$$

Set $U_0 = U_N = \mathrm{Id}$. The action is

$$(U \cdot \mathbf{W})_k = U_k W_k U_{k-1}^{-1}, \qquad 1 \le k \le N.$$

The product telescopes:

$$\phi(U \cdot \mathbf{W}) = (W_N U_{N-1}^{-1})(U_{N-1} W_{N-1} U_{N-2}^{-1}) \cdots (U_1 W_1) = W_N \cdots W_1 = \phi(\mathbf{W}).$$

Thus the action preserves every fibre $\mathcal{F}_X$.

A direct substitution in (6.3) gives

$$G_k(U \cdot \mathbf{W}) = U_k G_k(\mathbf{W}) U_k^*, \qquad 1 \le k \le N-1.$$

If $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$, then $G_k(\mathbf{W}) = G_k$. Since $U_k \in \mathrm{Stab}_{\mathrm{U}(d)}(G_k)$, we obtain

$$G_k(U \cdot \mathbf{W}) = U_k G_k U_k^* = G_k.$$

Hence $U \cdot \mathbf{W} \in \mathcal{M}_{\mathbf{G}}$. Therefore $K_{\mathbf{G}}$ preserves $\mathcal{M}_{\mathbf{G}}$, and since it also preserves $\mathcal{F}_X$, it preserves

$$\mathcal{O}_{X,\mathbf{G}} = \mathcal{F}_X \cap \mathcal{M}_{\mathbf{G}}.$$

It remains to prove freeness. Suppose $U \cdot \mathbf{W} = \mathbf{W}$. From the first layer,

$$U_1 W_1 = W_1.$$

Since W_1 is invertible, $U_1 = \mathrm{Id}$. If $U_{k-1} = \mathrm{Id}$, then the kth layer equation gives

$$U_k W_k = W_k,$$

and hence $U_k = \mathrm{Id}$. Induction gives

$$U_1 = \cdots = U_{N-1} = \mathrm{Id}\,.$$

Thus the action of $K_{\mathbf{G}}$ on $\mathcal{M}_{\mathbf{G}}$ is free. $\square$

6.3 Polar factorizations and coordinates on $\mathcal{F}_X/K$

6.3.1 Polar factorizations

Let

$$\mathbf{W} = (W_N, \dots, W_1) \in \mathrm{GL}(d, \mathbb{C})^N.$$

For each layer we use both polar decompositions:

$$W_k = Q_k P_k = R_k S_{k-1}^*, \tag{6.25}$$

where

$$P_k = (W_k^* W_k)^{1/2}, \qquad R_k = (W_k W_k^*)^{1/2}, \qquad Q_k, S_{k-1} \in \mathrm{U}(d). \tag{6.26}$$

Thus P_k is the positive factor in the right polar decomposition, and R_k is the positive factor in the left polar decomposition. Equivalently,

$$P_k^2 = W_k^* W_k, \qquad R_k^2 = W_k W_k^*. \tag{6.27}$$

At the interface between W_k and W_{k+1}, the kth moment compares the left positive factor of W_k with the right positive factor of W_{k+1}:

$$G_k(\mathbf{W}) = W_k W_k^* - W_{k+1}^* W_{k+1} = R_k^2 - P_{k+1}^2. \tag{6.28}$$

Hence balancedness is the matching condition

$$R_k^2 = P_{k+1}^2, \qquad 1 \leq k \leq N-1. \tag{6.29}$$

The same distinction explains the point of view based on polar factorizations in [Men25]. Suppose that $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$. Then, for $1 \leq k \leq N-1$,

$$R_k^2 = W_k W_k^*, \qquad P_{k+1} = (R_k^2 - G_k)^{1/2}, \qquad W_{k+1} = Q_{k+1} P_{k+1}. \tag{6.30}$$

Since $R_k^2 = Q_k P_k^2 Q_k^*$, this recursion can also be written as

$$P_1^2 = W_1^* W_1, \qquad P_{k+1} = (Q_k P_k^2 Q_k^* - G_k)^{1/2}, \qquad 1 \leq k \leq N-1. \tag{6.31}$$

For an actual point of $\mathcal{M}_{\mathbf{G}}$, the matrices under the square roots in (6.30) and (6.31) are positive definite by construction. Conversely, these positivity conditions are part of the constraint if one uses (6.31) to construct a point of $\mathcal{M}_{\mathbf{G}}$.

The recursion (6.31) starts from W_1 and advances with increasing depth. Since the product is written as

$$X = W_N \cdots W_1,$$

this advance runs from right to left in the displayed product.

There is a dual recursion from the other end. If $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$, then for $1 \le k \le N-1$,

$$P_{k+1}^2 = W_{k+1}^* W_{k+1}, \qquad R_k = (P_{k+1}^2 + G_k)^{1/2}, \qquad W_k = R_k S_{k-1}^*. \tag{6.32}$$

This recursion starts from W_N and advances with decreasing depth. In the displayed product it therefore runs from left to right.

The forward recursion is naturally expressed in the right polar decomposition, whereas the backward recursion is naturally expressed in the left polar decomposition. A fixed fibre has no distinguished intermediate layer. We therefore pass from the individual layers to the partial products, and then to their Gram matrices. These matrices remove the unitary frames introduced by the K-action.

6.3.2 Coordinates on $\mathcal{F}_X/K$

Fix $X \in \mathrm{GL}(d, \mathbb{C})$, and let

$$\mathbf{W} = (W_N, \ldots, W_1) \in \mathcal{F}_X.$$

Define the forward partial products by

$$Y_k := W_k \cdots W_1, \qquad 0 \le k \le N, \tag{6.33}$$

with

$$Y_0 = \mathrm{Id}, \qquad Y_N = X.$$

Then the layers are recovered from the partial products by

$$W_k = Y_k Y_{k-1}^{-1}, \qquad 1 \le k \le N. \tag{6.34}$$

Let

$$K = \mathrm{U}(d)^{N-1}.$$

For $U = (U_{N-1}, \dots, U_1) \in K$, set $U_0 = U_N = \mathrm{Id}$. Under the K-action on the fibre, the partial products transform as

$$Y_k \longmapsto U_k Y_k, \qquad 1 \le k \le N-1. \tag{6.35}$$

It follows that the matrices

$$H_k := Y_k^* Y_k, \qquad 1 \le k \le N-1, \tag{6.36}$$

are invariant under the K-action. We also use the endpoint convention

$$H_0 = \mathrm{Id}, \qquad H_N = X^* X. \tag{6.37}$$

We write

$$\mathbb{P}_d \subset \mathrm{Herm}(d)$$

for the cone of positive-definite Hermitian matrices. With Y_k and H_k as in (6.33) and (6.36), the map from Proposition 6.1 is

$$q_X : \mathcal{F}_X \longrightarrow \mathbb{P}_d^{N-1}, \qquad q_X(\mathbf{W}) = (H_1, \dots, H_{N-1}). \tag{6.38}$$

Proof of Proposition 6.1. By (6.35), each matrix $H_k = Y_k^* Y_k$ is K-invariant. Hence q_X is constant on K-orbits and descends to a smooth map

$$\overline{q}_X : \mathcal{F}_X / K \longrightarrow \mathbb{P}_d^{N-1}.$$

$$\mathbf{W} \in \mathcal{F}_X \quad \longmapsto \quad (Y_1, \ldots, Y_{N-1}) \quad \longmapsto \quad (H_1, \ldots, H_{N-1}) \in \mathbb{P}_d^{N-1}$$

$$Y_k = W_k \cdots W_1, \qquad H_k = Y_k^* Y_k.$$

Figure 6.1: The positive-definite coordinates attached to a factorization in $\mathcal{F}_X$. The partial products Y_k change by left multiplication by unitaries under the K-action; the matrices $H_k = Y_k^* Y_k$ do not.

We first prove surjectivity. Let

$$(H_1, \ldots, H_{N-1}) \in \mathbb{P}_d^{N-1}$$

be given. Define

$$W_1 := H_1^{1/2}, \qquad W_k := H_k^{1/2} H_{k-1}^{-1/2} \quad (2 \leq k \leq N-1), \qquad W_N := X H_{N-1}^{-1/2}. \quad (6.39)$$

Then the product telescopes:

$$Y_k = W_k \cdots W_1 = H_k^{1/2}, \qquad 1 \leq k \leq N-1,$$

and

$$Y_N = W_N Y_{N-1} = X H_{N-1}^{-1/2} H_{N-1}^{1/2} = X.$$

Thus $\mathbf{W} \in \mathcal{F}_X$, and

$$q_X(\mathbf{W}) = (H_1, \ldots, H_{N-1}).$$

We next prove injectivity on the quotient. Suppose that $\mathbf{W}, \mathbf{W}' \in \mathcal{F}_X$ satisfy

$$q_X(\mathbf{W}) = q_X(\mathbf{W}').$$

Let Y_k and Y'_k be the corresponding partial products. Then

$$Y_k^* Y_k = (Y'_k)^* Y'_k = H_k, \qquad 1 \le k \le N-1.$$

Define

$$U_k := Y'_k Y_k^{-1}, \qquad 1 \le k \le N-1. \tag{6.40}$$

Then

$$U_k^* U_k = Y_k^{-*} (Y'_k)^* Y'_k Y_k^{-1} = Y_k^{-*} H_k Y_k^{-1} = Y_k^{-*} Y_k^* Y_k Y_k^{-1} = \mathrm{Id}\,.$$

Hence each U_k is unitary. Set $U_0 = U_N = \mathrm{Id}$. Using (6.34), we obtain

$$W'_k = Y'_k (Y'_{k-1})^{-1} = U_k Y_k Y_{k-1}^{-1} U_{k-1}^{-1} = U_k W_k U_{k-1}^{-1}, \qquad 1 \le k \le N,$$

where the cases $k = 1$ and $k = N$ use $U_0 = \mathrm{Id}$ and $U_N = \mathrm{Id}$, respectively. Therefore

$$\mathbf{W}' = U \cdot \mathbf{W}$$

for $U = (U_{N-1}, \ldots, U_1) \in K$. Thus q_X separates K-orbits.

It remains only to record smoothness of the inverse. The construction (6.39) gives a smooth section

$$\mathbb{P}_d^{N-1} \longrightarrow \mathcal{F}_X,$$

because inversion and the positive square-root map are smooth on $\mathbb{P}_d$. Passing this section to K-orbits gives a smooth inverse to $\overline{q}_X$. Hence

$$\overline{q}_X : \mathcal{F}_X / K \longrightarrow \mathbb{P}_d^{N-1}$$

is a diffeomorphism. $\square$

We shall therefore represent a point of $\mathcal{F}_X/K$ by the discrete path

$$\mathrm{Id} = H_0, \quad H_1, \quad \dots, \quad H_{N-1}, \quad H_N = X^*X \tag{6.41}$$

in $\mathbb{P}_d$.

6.4 The moment equations on the quotient

Write the polar decomposition of each partial product as

$$Y_k = V_k H_k^{1/2}, \qquad V_k \in \mathrm{U}(d), \qquad 0 \le k \le N, \tag{6.42}$$

where

$$V_0 = \mathrm{Id}, \qquad V_N = U_X, \qquad X = U_X(X^*X)^{1/2}$$

is the polar decomposition of X. Define

$$C_k := H_k^{1/2} H_{k-1}^{-1/2}, \qquad 1 \le k \le N. \tag{6.43}$$

Since $W_k = Y_k Y_{k-1}^{-1}$, we have

$$W_k = V_k C_k V_{k-1}^*, \qquad 1 \le k \le N. \tag{6.44}$$

The following auxiliary computation records the two polar squares of a layer in the H-coordinates.

Lemma 6.16. *For $1 \leq k \leq N$,*

$$W_k^* W_k = V_{k-1}\Big(H_{k-1}^{-1/2} H_k H_{k-1}^{-1/2}\Big)V_{k-1}^*, \tag{6.45}$$

$$W_k W_k^* = V_k\Big(H_k^{1/2} H_{k-1}^{-1} H_k^{1/2}\Big)V_k^*. \tag{6.46}$$

Proof. By (6.44) and unitarity of V_k and V_{k-1},

$$W_k^* W_k = V_{k-1} C_k^* C_k V_{k-1}^*, \qquad W_k W_k^* = V_k C_k C_k^* V_k^*.$$

Using (6.43),

$$C_k^* C_k = H_{k-1}^{-1/2} H_k H_{k-1}^{-1/2}, \qquad C_k C_k^* = H_k^{1/2} H_{k-1}^{-1} H_k^{1/2}.$$

Substitution gives (6.45) and (6.46). □

The two terms in the kth component of the moment map are now written in the same unitary frame V_k.

Proof of Proposition 6.4. By (6.46),

$$W_k W_k^* = V_k\Big(H_k^{1/2} H_{k-1}^{-1} H_k^{1/2}\Big)V_k^*.$$

Applying (6.45) to the layer W_{k+1} gives

$$W_{k+1}^* W_{k+1} = V_k\Big(H_k^{-1/2} H_{k+1} H_k^{-1/2}\Big)V_k^*.$$

Therefore

$$G_k(\mathbf{W}) = W_k W_k^* - W_{k+1}^* W_{k+1} = V_k\Big(H_k^{1/2} H_{k-1}^{-1} H_k^{1/2} - H_k^{-1/2} H_{k+1} H_k^{-1/2}\Big)V_k^*.$$

The term in parentheses is $\Phi_k(H)$ by (6.8). Hence

$$G_k(\mathbf{W}) = V_k \Phi_k(H) V_k^*, \qquad 1 \le k \le N-1.$$

$\square$

Thus the quotient coordinates determine each moment component up to unitary conjugacy. The proof of Theorem 6.5 identifies the exact reduced set.

Proof of Theorem 6.5. Let $\mathbf{W} \in \mathcal{O}_{X,\mathbf{G}}$, and set

$$H = q_X(\mathbf{W}).$$

Since $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$, one has

$$G_k(\mathbf{W}) = G_k, \qquad 1 \le k \le N-1.$$

By Proposition 6.4,

$$G_k = V_k \Phi_k(H) V_k^*.$$

Equivalently,

$$\Phi_k(H) = V_k^* G_k V_k \in \mathrm{Ad}_{\mathrm{U}(d)}(G_k).$$

Hence H belongs to the right-hand side of (6.10). This proves the inclusion

$$R_{X,\mathbf{G}} \subseteq \Big\{ H \in \mathbb{P}_d^{N-1} : \Phi_k(H) \in \mathrm{Ad}_{\mathrm{U}(d)}(G_k) \text{ for } 1 \le k \le N-1 \Big\}.$$

Conversely, let $H \in \mathbb{P}_d^{N-1}$ satisfy

$$\Phi_k(H) \in \mathrm{Ad}_{\mathrm{U}(d)}(G_k), \qquad 1 \le k \le N-1.$$

For each k, choose $V_k \in \mathrm{U}(d)$ such that

$$G_k = V_k \Phi_k(H) V_k^*.$$

Set

$$V_0 = \mathrm{Id}, \qquad V_N = U_X,$$

where $X = U_X (X^* X)^{1/2}$. Define

$$Y_k := V_k H_k^{1/2}, \qquad 1 \leq k \leq N-1,$$

and set

$$Y_0 := \mathrm{Id}, \qquad Y_N := X.$$

Now define the layers by

$$W_k := Y_k Y_{k-1}^{-1}, \qquad 1 \leq k \leq N.$$

The product telescopes, so $\mathbf{W} \in \mathcal{F}_X$, and by construction $q_X(\mathbf{W}) = H$. Proposition 6.4 gives

$$G_k(\mathbf{W}) = V_k \Phi_k(H) V_k^* = G_k, \qquad 1 \leq k \leq N-1.$$

Thus $\mathbf{W} \in \mathcal{O}_{X,\mathbf{G}}$, and therefore

$$H \in q_X(\mathcal{O}_{X,\mathbf{G}}) = R_{X,\mathbf{G}}.$$

This proves the equality (6.10).

It remains to identify the fibres of the restricted map. Let

$$\mathbf{W}, \mathbf{W}' \in \mathcal{O}_{X,\mathbf{G}}$$

satisfy

$$q_X(\mathbf{W}) = q_X(\mathbf{W}').$$

By Proposition 6.1, there exists

$$U = (U_{N-1}, \dots, U_1) \in K$$

such that

$$\mathbf{W}' = U \cdot \mathbf{W}.$$

For the K-action, the moment components transform as

$$G_k(U \cdot \mathbf{W}) = U_k G_k(\mathbf{W}) U_k^*, \qquad 1 \le k \le N-1.$$

Since both $\mathbf{W}$ and $\mathbf{W}'$ lie in $\mathcal{M}_{\mathbf{G}}$,

$$G_k = G_k(\mathbf{W}') = U_k G_k(\mathbf{W}) U_k^* = U_k G_k U_k^*.$$

Hence

$$U_k \in \mathrm{Stab}_{\mathrm{U}(d)}(G_k), \qquad 1 \le k \le N-1,$$

so $U \in K_{\mathbf{G}}$. Thus each fibre of

$$q_X\Big|_{\mathcal{O}_{X,\mathbf{G}}}$$

is contained in a $K_{\mathbf{G}}$-orbit.

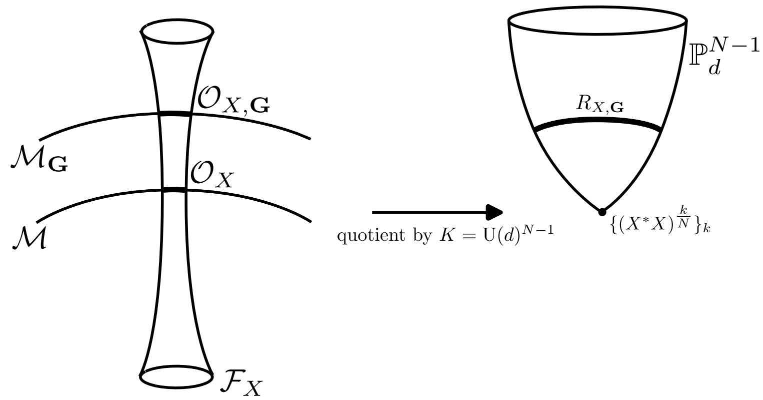


Figure 6.2: Level sets of the moment map in a fixed fibre and their images under q_X. At $\mathbf{G} = 0$, the image of the balanced part of $\mathcal{F}_X$ is the single point $\{(X^*X)^{k/N}\}_{k=1}^{N-1}$. For $\mathbf{G} \neq 0$, the image is the reduced set $R_{X,\mathbf{G}} \subset \mathbb{P}_d^{N-1}$ defined by the constraints $\Phi_k(H) \in \mathrm{Ad}_{U(d)}(G_k)$ whenever $\mathcal{O}_{X,\mathbf{G}}$ is nonempty.

The reverse inclusion follows because $K_{\mathbf{G}}$ preserves $\mathcal{O}_{X,\mathbf{G}}$ by Proposition 6.3, and because q_X is invariant under the whole group K. Therefore the fibres of $q_X\big|_{\mathcal{O}_{X,\mathbf{G}}}$ are precisely the $K_{\mathbf{G}}$-orbits. $\square$

Figure 6.2 records the geometry described by Theorem 6.5.

Remark 6.17. Theorem 6.5 shows that, for $\mathbf{G} \neq 0$, the reduced object is the set

$$R_{X,\mathbf{G}} \subset \mathbb{P}_d^{N-1},$$

not a single compact orbit. Any entropy attached to $\mathcal{O}_{X,\mathbf{G}}$ would therefore have to combine two pieces of geometry: the volumes of the $K_{\mathbf{G}}$-orbits and a measure on the reduced set $R_{X,\mathbf{G}}$. We do not define such an entropy here. The theorem only identifies the reduced finite-dimensional set on which such a construction would have to live. When $\mathbf{G} = 0$, the reduced set collapses to a point and one recovers the entropy formula for a single orbit studied in Chapter 5.

The balanced case is obtained by imposing $\Phi_k(H) = 0$ for every $1 \leq k \leq N-1$. In terms of the multiplicative increments $\Pi_k = H_{k-1}^{-1} H_k$, these equations become the first-order recurrence $\Pi_k = \Pi_{k+1}$.

Proof of Lemma 6.6. Using

$$\Pi_k = H_{k-1}^{-1} H_k, \qquad \Pi_{k+1} = H_k^{-1} H_{k+1},$$

we compute

$$H_k^{1/2} \Pi_k H_k^{-1/2} = H_k^{1/2} H_{k-1}^{-1} H_k^{1/2},$$

and

$$H_k^{1/2} \Pi_{k+1} H_k^{-1/2} = H_k^{1/2} H_k^{-1} H_{k+1} H_k^{-1/2} = H_k^{-1/2} H_{k+1} H_k^{-1/2}.$$

Subtracting the two identities gives

$$\Phi_k(H) = H_k^{1/2} (\Pi_k - \Pi_{k+1}) H_k^{-1/2}.$$

Since H_k is invertible, this expression vanishes if and only if $\Pi_k - \Pi_{k+1} = 0$. Hence

$$\Phi_k(H) = 0 \qquad \Longleftrightarrow \qquad \Pi_k = \Pi_{k+1}.$$

$\square$

Proof of Proposition 6.7. For each $1 \leq k \leq N-1$, Lemma 6.6 gives

$$\Phi_k(H) = 0 \qquad \Longleftrightarrow \qquad \Pi_k = \Pi_{k+1}.$$

Thus imposing $\Phi_k(H) = 0$ for all k is exactly the same as imposing

$$\Pi_1 = \Pi_2 = \cdots = \Pi_N.$$

$\square$

Proof of Corollary 6.8. For $\mathbf{G} = 0$, the adjoint orbit of each component $G_k = 0$ is the singleton $\{0\}$. Hence Theorem 6.5 gives

$$R_{X,0} = \Big\{H \in \mathbb{P}_d^{N-1} : \Phi_k(H) = 0 \text{ for } 1 \leq k \leq N-1\Big\}.$$

By Proposition 6.7, this is equivalent to

$$\Pi_1 = \Pi_2 = \cdots = \Pi_N.$$

Write the common value as Π. Since

$$\Pi_1 = H_0^{-1} H_1 = H_1$$

and $H_1 \in \mathbb{P}_d$, the matrix Π is positive definite. The recurrence $H_k = H_{k-1}\Pi$, together with $H_0 = \mathrm{Id}$, gives

$$H_k = \Pi^k, \qquad 0 \leq k \leq N.$$

The endpoint condition $H_N = X^*X$ gives

$$\Pi^N = X^*X.$$

Because Π is positive definite, it is the positive Nth root of X^*X:

$$\Pi = (X^*X)^{1/N}.$$

Therefore

$$H_k = (X^*X)^{k/N}, \qquad 0 \leq k \leq N.$$

This proves the formula for $R_{X,0}$ and the uniqueness of the balanced path.

Finally, when $\mathbf{G} = 0$, the stabilizer $K_{\mathbf{G}}$ is the whole group K. Since the reduced set consists of one point, Theorem 6.5 shows that $\mathcal{O}_{X,0}$ is a single K-orbit. By (6.6), this orbit is $\mathcal{O}_X$. $\square$

6.5 The trace metric and balancedness

The coordinates on $\mathcal{F}_X/K$ turn a factorization into a path $\mathrm{Id} = H_0,\ H_1, \dots, H_{N-1},\ H_N = X^*X$ in $\mathbb{P}_d$, with fixed endpoints. The metric used in this section is the affine-invariant trace metric on $\mathbb{P}_d$:

$$g_H^{\mathrm{tr}}(U,V) := \mathrm{tr}\Big(H^{-1}UH^{-1}V\Big), \qquad H \in \mathbb{P}_d, \quad U,V \in T_H\mathbb{P}_d \cong \mathrm{Herm}(d). \tag{6.47}$$

Equivalently, if

$$S_H(U) := H^{-1/2}UH^{-1/2},$$

then

$$g_H^{\mathrm{tr}}(U,V) = \mathrm{tr}(S_H(U)S_H(V)). \tag{6.48}$$

The trace metric enters here only as a metric on the cone $\mathbb{P}_d$. We are not claiming that the diffeomorphism

$$\mathcal{F}_X/K \cong \mathbb{P}_d^{N-1}$$

is an isometry for the product of trace metrics. The point is more specific: the path selected by balancedness is the geodesic of the trace metric joining the two endpoints in $\mathbb{P}_d$.

For a smooth curve $\gamma(t) \in \mathbb{P}_d$, the geodesic equation for (6.47), written in the affine coordinates of Herm(d), is

$$\ddot{\gamma} = \dot{\gamma}\,\gamma^{-1}\dot{\gamma}. \tag{6.49}$$

Equivalently,

$$\frac{d}{dt}\Big(\gamma^{-1}\dot{\gamma}\Big) = 0. \tag{6.50}$$

Proof of Proposition 6.9, [Bha07]. Set

$$S := A^{-1/2}BA^{-1/2}, \qquad \gamma(t) := A^{1/2}S^tA^{1/2}.$$

Then $\gamma(0) = A$ and $\gamma(1) = B$. Since $S \in \mathbb{P}_d$, the logarithm $\log S$ is Hermitian and commutes with S^t. Hence

$$\dot{\gamma}(t) = A^{1/2}S^t\log(S)A^{1/2}, \qquad \ddot{\gamma}(t) = A^{1/2}S^t(\log(S))^2A^{1/2}.$$

Also

$$\gamma(t)^{-1} = A^{-1/2}S^{-t}A^{-1/2}.$$

Therefore

$$\dot{\gamma}(t)\,\gamma(t)^{-1}\dot{\gamma}(t) = A^{1/2}S^t(\log(S))^2A^{1/2} = \ddot{\gamma}(t).$$

Thus γ satisfies the geodesic equation (6.49).

Moreover,

$$\gamma(t)^{-1}\dot{\gamma}(t) = A^{-1/2}\log(S)A^{1/2},$$

which is independent of t. This proves (6.50). Taking $A = \mathrm{Id}$ and $B = X^*X$ gives the geodesic

$$\gamma(t) = (X^*X)^t.$$

$\square$

Proof of Corollary 6.10. By Corollary 6.8, the path selected by balancedness is

$$H_k = (X^*X)^{k/N}, \qquad 0 \leq k \leq N.$$

By Proposition 6.9, the geodesic of the trace metric from Id to X^*X is

$$\gamma(t) = (X^*X)^t, \qquad 0 \leq t \leq 1.$$

Thus

$$H_k = \gamma(k/N), \qquad 0 \leq k \leq N,$$

so the balanced path is obtained by sampling this geodesic at the times k/N. $\square$

6.6 The ridge regularizer on the quotient

Proof of Proposition 6.11. Let

$$Y_k = V_k H_k^{1/2}$$

be the polar decomposition of the kth partial product. By Lemma 6.16,

$$W_k^* W_k = V_{k-1}\Big(H_{k-1}^{-1/2} H_k H_{k-1}^{-1/2}\Big)V_{k-1}^*.$$

Taking traces and using cyclicity gives

$$\operatorname{tr}(W_k^* W_k) = \operatorname{tr}\Big(H_{k-1}^{-1/2} H_k H_{k-1}^{-1/2}\Big) = \operatorname{tr}\Big(H_k H_{k-1}^{-1}\Big).$$

Summing over $1 \leq k \leq N$ gives (6.13).

The right-hand side of (6.13) depends only on

$$q_X(\mathbf{W}) = (H_1, \ldots, H_{N-1}),$$

so $\mathcal{R}$ is constant on K-orbits. The map q_X identifies $\mathcal{F}_X/K$ with $\mathbb{P}_d^{N-1}$ by Proposition 6.1. Therefore $\mathcal{R}$ descends to the function $\mathcal{R}_X$ on $\mathbb{P}_d^{N-1}$. Smoothness follows because the map

$$(A, B) \longmapsto \operatorname{tr}(AB^{-1})$$

is smooth on $\mathbb{P}_d \times \mathbb{P}_d$. $\square$

The Euler–Lagrange equations for this energy force the multiplicative increments to be constant.

Proof of Proposition 6.12. Fix $1 \leq k \leq N-1$, let $U_k \in \operatorname{Herm}(d)$, and vary only the kth interior matrix:

$$H_k(\varepsilon) = H_k + \varepsilon U_k.$$

Since $\mathbb{P}_d \subset \operatorname{Herm}(d)$ is open, this is an admissible variation for ε sufficiently small.

Only two terms in $\mathcal{R}_X$ contain H_k:

$$\frac{1}{2} \operatorname{tr}\Big(H_k H_{k-1}^{-1}\Big) + \frac{1}{2} \operatorname{tr}\Big(H_{k+1} H_k^{-1}\Big).$$

Using

$$\mathrm{D}(H^{-1})_H[U] = -H^{-1} U H^{-1},$$

we obtain

$$\frac{d}{d\varepsilon}\bigg|_{\varepsilon=0}\mathcal{R}_X = \frac{1}{2}\operatorname{tr}\Big(U_k H_{k-1}^{-1}\Big) - \frac{1}{2}\operatorname{tr}\Big(H_{k+1}H_k^{-1}U_k H_k^{-1}\Big)$$
$$= \frac{1}{2}\operatorname{tr}\Big(U_k\Big(H_{k-1}^{-1} - H_k^{-1}H_{k+1}H_k^{-1}\Big)\Big).$$

Since $U_k \in \mathrm{Herm}(d)$ is arbitrary and the trace pairing is nondegenerate on $\mathrm{Herm}(d)$, the first variation vanishes in the kth variable if and only if

$$H_{k-1}^{-1} = H_k^{-1}H_{k+1}H_k^{-1}.$$

Right multiplication by H_k gives

$$H_{k-1}^{-1}H_k = H_k^{-1}H_{k+1}.$$

Thus the critical point equations are exactly

$$\Pi_k = \Pi_{k+1}, \qquad 1 \le k \le N-1,$$

with $\Pi_k = H_{k-1}^{-1}H_k$.

By Proposition 6.7, these equations are equivalent to the reduced balancedness equations. Corollary 6.8 then gives the unique critical point:

$$H_k = (X^*X)^{k/N}, \qquad 0 \le k \le N.$$

It remains to see that this critical point is the minimizer. The function $\mathcal{R}_X$ is proper on $\mathbb{P}_d^{N-1}$. Indeed, suppose $\mathcal{R}_X(H) \leq C$. Then for every k,

$$\operatorname{tr}\Big(H_k H_{k-1}^{-1}\Big) \leq 2C.$$

Starting from $H_0 = \mathrm{Id}$, this gives upper bounds for $H_1, \ldots, H_{N-1}$ by induction. Namely, if $H_{k-1} \leq M_{k-1}\,\mathrm{Id}$, then $H_{k-1}^{-1} \geq M_{k-1}^{-1}\,\mathrm{Id}$, and hence

$$M_{k-1}^{-1}\operatorname{tr}(H_k) \leq \operatorname{tr}\Big(H_k H_{k-1}^{-1}\Big) \leq 2C.$$

Thus $H_k \leq \operatorname{tr}(H_k)\,\mathrm{Id} \leq 2CM_{k-1}\,\mathrm{Id}$.

Similarly, since $H_N = X^*X$ is fixed and positive definite, there is $m_N > 0$ such that $H_N \geq m_N\,\mathrm{Id}$. The same estimate applied backward gives upper bounds for $H_{N-1}^{-1}, \ldots, H_1^{-1}$. Thus every sublevel set of $\mathcal{R}_X$ lies in a set of the form

$$\{(H_1, \ldots, H_{N-1}) : m_k\,\mathrm{Id} \leq H_k \leq M_k\,\mathrm{Id} \;\text{ for } 1 \leq k \leq N-1\},$$

which is compact in $\mathrm{Herm}(d)^{N-1}$. Therefore $\mathcal{R}_X$ attains its minimum. Any minimizer is an interior critical point, and the critical point is unique. Hence the balanced path is the unique minimizer. $\square$

Proof of Corollary 6.13. On the balanced path,

$$H_k = (X^*X)^{k/N}, \qquad 0 \leq k \leq N.$$

Therefore

$$H_k H_{k-1}^{-1} = (X^*X)^{1/N}$$

for every $1 \le k \le N$. Proposition 6.11 gives

$$\mathcal{R}_X(H_1, \dots, H_{N-1}) = \frac{1}{2}\sum_{k=1}^{N} \operatorname{tr}\Big((X^*X)^{1/N}\Big) = \frac{N}{2}\operatorname{tr}\Big((X^*X)^{1/N}\Big).$$

Since Proposition 6.12 identifies this path as the unique minimizer of $\mathcal{R}_X$, and since $\mathcal{R}_X$ is the function induced by $\mathcal{R}$ under $q_X : \mathcal{F}_X \to \mathbb{P}_d^{N-1}$, we obtain

$$\min_{\mathbf{W}\in\mathcal{F}_X} \mathcal{R}(\mathbf{W}) = \frac{N}{2}\operatorname{tr}\Big((X^*X)^{1/N}\Big).$$

The last expression is $\Psi_N(X)$ by (5.3). □

6.7 Coordinates from the left endpoint on the quotient

The quotient coordinates in (6.38) follow the product from W_1 toward W_N. Namely,

$$Y_k = W_k \cdots W_1, \qquad H_k = Y_k^* Y_k, \qquad 0 \le k \le N,$$

with

$$H_0 = \mathrm{Id}, \qquad H_N = X^*X.$$

We now record the same quotient information from the other end of the product.

For

$$\mathbf{W} = (W_N, \dots, W_1) \in \mathcal{F}_X,$$

define

$$Z_k := W_N \cdots W_{k+1}, \qquad 0 \le k \le N, \tag{6.51}$$

with the convention

$$Z_0 = X, \qquad Z_N = \mathrm{Id}\,.$$

Set

$$\widetilde{H}_k := Z_k Z_k^*, \qquad 0 \le k \le N. \tag{6.52}$$

Thus

$$\widetilde{H}_0 = XX^*, \qquad \widetilde{H}_N = \mathrm{Id}\,. \tag{6.53}$$

Let $U = (U_{N-1}, \dots, U_1) \in K$, and set $U_0 = U_N = \mathrm{Id}$. Under the K-action on $\mathcal{F}_X$, the products Z_k transform by

$$Z_k \longmapsto Z_k U_k^{-1}, \qquad 0 \le k \le N. \tag{6.54}$$

Consequently each $\widetilde{H}_k$ is K-invariant. The associated map

$$\widetilde{q}_X : \mathcal{F}_X \longrightarrow \mathbb{P}_d^{N-1}, \qquad \widetilde{q}_X(\mathbf{W}) = \Big(\widetilde{H}_1, \dots, \widetilde{H}_{N-1}\Big), \tag{6.55}$$

is therefore constant on K-orbits.

The following proof identifies $\widetilde{q}_X$ with the quotient map q_X through the change of variables

$$H_k \longmapsto X H_k^{-1} X^*.$$

Proof of Proposition 6.14. For every $0 \le k \le N$, the partial products satisfy

$$X = Z_k Y_k.$$

Hence

$$Z_k = X Y_k^{-1}.$$

Using (6.52), we obtain

$$\widetilde{H}_k = Z_k Z_k^* = XY_k^{-1}Y_k^{-*}X^* = X(Y_k^*Y_k)^{-1}X^* = XH_k^{-1}X^*.$$

Taking $k = 0$ and $k = N$ gives

$$\widetilde{H}_0 = XX^*, \qquad \widetilde{H}_N = X(X^*X)^{-1}X^* = \mathrm{Id},$$

as claimed. $\square$

Proof of Corollary 6.15. By Proposition 6.14,

$$\widetilde{H}_k = XH_k^{-1}X^*, \qquad 0 \le k \le N.$$

Therefore, for $1 \le k \le N$,

$$\widetilde{H}_{k-1}\widetilde{H}_k^{-1} = XH_{k-1}^{-1}X^*\Big(XH_k^{-1}X^*\Big)^{-1} = XH_{k-1}^{-1}H_kX^{-1}.$$

Taking traces and using cyclicity gives

$$\mathrm{tr}\Big(\widetilde{H}_{k-1}\widetilde{H}_k^{-1}\Big) = \mathrm{tr}\Big(H_{k-1}^{-1}H_k\Big) = \mathrm{tr}\Big(H_kH_{k-1}^{-1}\Big).$$

The formula for $\mathcal{R}_X$ now follows from Proposition 6.11.

On the balanced path, Corollary 6.8 gives

$$H_k = (X^*X)^{k/N}, \qquad 0 \le k \le N.$$

Hence

$$\widetilde{H}_k = X(X^*X)^{-k/N}X^*.$$

Write the polar decomposition of X as

$$X = U_X(X^*X)^{1/2}.$$

Then

$$\widetilde{H}_k = U_X(X^*X)^{1-k/N}U_X^* = (XX^*)^{1-k/N}.$$

By Proposition 6.9, the curve

$$t \longmapsto (XX^*)^{1-t}, \qquad 0 \leq t \leq 1,$$

is the geodesic of the trace metric from XX^* to Id. Thus the matrices $\widetilde{H}_k$ are obtained by sampling this geodesic at $t = k/N$. $\square$

The two sequences therefore give the same point of the quotient in two endpoint conventions. The forward variables describe the path

$$\mathrm{Id} = H_0,\ H_1, \ldots, H_N = X^*X,$$

whereas the variables from the left end describe the path

$$XX^* = \widetilde{H}_0,\ \widetilde{H}_1, \ldots, \widetilde{H}_N = \mathrm{Id}\,.$$

The conversion between them is

$$\widetilde{H}_k = XH_k^{-1}X^*.$$

This is the quotient form of the two polar recursions: (6.31) follows the product from W_1 toward W_N, while (6.32) follows it from W_N toward W_1.

Bibliographical notes

The barrier $B(H) = -\log\det H$ and its role in interior-point methods come from the theory of convex cones [NN94; NT97; Gül96; Hil14]. The trace metric on $\mathbb{P}_d$ is standard in matrix analysis [Bha07]. The local Ricci-potential identity used in the comparison with Chapter 5 is standard in Kähler geometry [Sun17]. The polar-factorization viewpoint appeared earlier in [Men25]. We recast it here in quotient form.

Chapter 7

Mean Curvature and Brownian Motion

In this chapter we study how the level set $\mathcal{M}_{\mathbf{G}} = \mu^{-1}(\mathbf{G})$ sits inside $\mathrm{GL}(d,\mathbb{C})^N$, and how Brownian motion on $\mathcal{M}_{\mathbf{G}}$ behaves under the end-to-end map. At a regular point, the normal space to $\mathcal{M}_{\mathbf{G}}$ is J applied to the tangent space of the compact group orbit. This identity is the main geometric input: it turns the equation $\mu(\mathbf{W}) = \mathbf{G}$ into a formula for the mean-curvature vector of $\mathcal{M}_{\mathbf{G}}$.

We then apply Itô's formula to the end-to-end map. The mean-curvature drift lies in a direction annihilated by the differential of this map, so only the Itô correction remains. For a general value of $\mathbf{G}$, the resulting drift may depend on the factorization $\mathbf{W}$, not only on the product $X = \phi(\mathbf{W})$. On the balanced manifold $\mathcal{M}$, this drift is basic and descends to $\frac{1}{2}\operatorname{grad}_{g_N} S_N$ on $\mathrm{GL}(d,\mathbb{C})$.

7.1 Main results

Fix $\mathbf{G} \in \mathrm{Herm}(d)^{N-1}$, $\quad \mathcal{M}_{\mathbf{G}} = \mu^{-1}(\mathbf{G})$, and set $K = U(d)^{N-1}$, $\quad \mathfrak{k} = \mathfrak{u}(d)^{N-1}$. For $\mathbf{W} \in \mathrm{GL}(d,\mathbb{C})^N$, write

$$\mathcal{O}_{\mathbf{W}} := K \cdot \mathbf{W}, \qquad \mathcal{V}_{\mathbf{W}} := T_{\mathbf{W}}\mathcal{O}_{\mathbf{W}}.$$

For $\xi \in \mathfrak{k}$, the fundamental vector field is

$$(K_\xi(\mathbf{W}))_k = \xi_k W_k - W_k \xi_{k-1}, \qquad 1 \le k \le N,$$

with the boundary convention $\xi_0 = \xi_N = 0$.

Proposition 7.1. *For every* $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$*,*

$$T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = \ker(\mathrm{d}\mu)_{\mathbf{W}}, \qquad \mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = J\mathcal{V}_{\mathbf{W}},$$

where $\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}$ *is the* g*-orthogonal complement of* $T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}$ *in* $T_{\mathbf{W}}\mathrm{GL}(d,\mathbb{C})^N$*. Equivalently,*

$$T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = (J\mathcal{V}_{\mathbf{W}})^{\perp_g}.$$

Choose a real basis $\{\xi_a\}_{a=1}^{(N-1)d^2}$ of $\mathfrak{k}$, and set

$$K_a := K_{\xi_a}, \qquad \mu_a(\mathbf{W}) := \langle \mu(\mathbf{W}),\, \xi_a \rangle.$$

The Gram matrix of the orbit $K \cdot \mathbf{W}$ is

$$h_{ab}(\mathbf{W}) := g\Big(K_a(\mathbf{W}), K_b(\mathbf{W})\Big), \qquad h^{ab}(\mathbf{W}) := (h(\mathbf{W})^{-1})^{ab}.$$

Repeated Lie-algebra indices are summed.

Let

$$S_{\mathrm{orb}}(\mathbf{W}) := \log \mathrm{vol}(K \cdot \mathbf{W}), \tag{7.1}$$

where the volume is computed using the metric induced by g on the orbit. For a smooth function f on $\mathrm{GL}(d, \mathbb{C})^N$, define

$$\nabla^{\perp} f := h^{ab}\Big((JK_a)f\Big)\, JK_b. \tag{7.2}$$

By Proposition 7.1, $\nabla^{\perp} f$ is a section of the normal bundle of $\mathcal{M}_{\mathbf{G}}$.

Define the normal vector field

$$\mathcal{C}_{\mathbf{G}}(\mathbf{W}) := h^{ad}(\mathbf{W}) h^{bc}(\mathbf{W})\, \langle \mathbf{G},\, [[\xi_a, \xi_b], \xi_c]\rangle\, JK_d(\mathbf{W}). \tag{7.3}$$

This term vanishes when $\mathbf{G} = \mathbf{0}$. Although the formula uses the chosen basis $\{\xi_a\}$, it defines a normal vector field that is independent of that basis.

Theorem 7.2. *For every $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$, the mean curvature vector of $\mathcal{M}_{\mathbf{G}}$ in the ambient metric g is*

$$\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W}) = \nabla^{\perp} S_{\mathrm{orb}}(\mathbf{W}) + \mathcal{C}_{\mathbf{G}}(\mathbf{W}). \tag{7.4}$$

Thus the first term is determined by the variation of the volume of $K \cdot \mathbf{W}$. The second term contains the dependence on the value $\mathbf{G}$ through Lie brackets. The whole expression is independent of the chosen Lie-algebra basis; the inverse Gram matrix performs the trace over the orbit directions.

Let $\nabla^2 \phi$ denote the Hessian of

$$\phi : \mathrm{GL}(d, \mathbb{C})^N \longrightarrow \mathrm{GL}(d, \mathbb{C}) \subset \mathrm{M}_d(\mathbb{C})$$

with respect to the flat connections on the source and target. Define the section of $(\phi|_{\mathcal{M}_{\mathbf{G}}})^{*}T\mathrm{GL}(d,\mathbb{C})$

$$\mathcal{D}_{\mathbf{G}}(\mathbf{W}) := h^{ab}(\mathbf{W})\,\nabla^2\phi_{\mathbf{W}}\Big(K_a(\mathbf{W}),K_b(\mathbf{W})\Big). \tag{7.5}$$

All stochastic statements in this chapter are understood up to the lifetime of the relevant process.

Theorem 7.3. *Let $\mathbf{W}_t$ be Brownian motion on $(\mathcal{M}_{\mathbf{G}},\iota_{\mathbf{G}})$, and set $X_t := \phi(\mathbf{W}_t)$. Then X_t is a semimartingale with drift $\frac{1}{2}\mathcal{D}_{\mathbf{G}}(\mathbf{W}_t)\,\mathrm{d}t$. Equivalently,*

$$\mathrm{d}X_t = \mathrm{d}M_t + \frac{1}{2}\,\mathcal{D}_{\mathbf{G}}(\mathbf{W}_t)\,\mathrm{d}t, \tag{7.6}$$

where M_t is a local martingale.

The mean-curvature drift does not contribute after applying ϕ. Indeed, ϕ is K-invariant and holomorphic, hence

$$(\mathrm{d}\phi)_{\mathbf{W}}(J\mathcal{V}_{\mathbf{W}}) = 0.$$

Since $\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W}) \in \mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}$, Proposition 7.1 gives

$$(\mathrm{d}\phi)_{\mathbf{W}}\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W})\Big) = 0.$$

The drift in (7.6) is therefore the Itô correction.

For a general value $\mathbf{G}$, the section $\mathcal{D}_{\mathbf{G}}$ need not be projectable under $\phi|_{\mathcal{M}_{\mathbf{G}}}$. It defines a vector field on $\phi(\mathcal{M}_{\mathbf{G}})$ if and only if $\mathcal{D}_{\mathbf{G}}(\mathbf{W}) = \mathcal{D}_{\mathbf{G}}(\mathbf{W}')$ whenever

$$\mathbf{W},\mathbf{W}' \in \mathcal{M}_{\mathbf{G}}, \qquad \phi(\mathbf{W}) = \phi(\mathbf{W}').$$

Proposition 7.4. *For every* $\mathbf{W} \in \mathcal{M}$,

$$K \cdot \mathbf{W} = (\phi|_{\mathcal{M}})^{-1}\Big(\phi(\mathbf{W})\Big) = \mathcal{O}_{\phi(\mathbf{W})}.$$

Consequently,

$$S_{\mathrm{orb}}(\mathbf{W}) = S_N(\phi(\mathbf{W})).$$

On $\mathcal{M}$, the restriction

$$\phi : (\mathcal{M}, \iota) \longrightarrow (\mathrm{GL}(d, \mathbb{C}), g_N)$$

is the Riemannian submersion from Chapter 5.

Theorem 7.5. *For every* $\mathbf{W} \in \mathcal{M}$,

$$\mathcal{D}_{\mathbf{0}}(\mathbf{W}) = \Big(\mathrm{grad}_{g_N} S_N\Big)_{\phi(\mathbf{W})}. \tag{7.7}$$

Corollary 7.6. *Let* $\mathbf{W}_t$ *be Brownian motion on* $(\mathcal{M}, \iota)$*, and set* $X_t := \phi(\mathbf{W}_t)$*. Then* X_t *is a diffusion on* $(\mathrm{GL}(d, \mathbb{C}), g_N)$ *with generator*

$$\mathcal{L}f = \frac{1}{2}\,\Delta_{g_N} f + \frac{1}{2}\Big\langle \mathrm{grad}_{g_N} S_N, \mathrm{grad}_{g_N} f \Big\rangle_{g_N}. \tag{7.8}$$

Equivalently, in local g_N*-orthonormal frames,*

$$\mathrm{d}X_t = \mathrm{d}\widetilde{B}_t + \frac{1}{2}\,\mathrm{grad}_{g_N} S_N(X_t)\,\mathrm{d}t, \tag{7.9}$$

where $\widetilde{B}_t$ *denotes the martingale part with covariance* g_N*.*

7.2 Orbits and the Level Set $\mathcal{M}_{\mathbf{G}}$

Fix $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$ and write

$$\mathcal{O}_{\mathbf{W}} := K \cdot \mathbf{W}, \qquad \mathcal{V}_{\mathbf{W}} := T_{\mathbf{W}}\mathcal{O}_{\mathbf{W}} \subset T_{\mathbf{W}}\mathrm{GL}(d,\mathbb{C})^N.$$

By Proposition 4.4, the K-action is free on $\mathrm{GL}(d,\mathbb{C})^N$. Hence the map

$$\mathfrak{k} \longrightarrow \mathcal{V}_{\mathbf{W}}, \qquad \xi \longmapsto K_\xi(\mathbf{W}),$$

is an isomorphism of real vector spaces, and

$$\dim_{\mathbb{R}} \mathcal{V}_{\mathbf{W}} = \dim_{\mathbb{R}} K = (N-1)d^2.$$

For $\xi \in \mathfrak{k} = \mathfrak{u}(d)^{N-1}$, Lemma 4.11 gives the fundamental vector field

$$(K_\xi(\mathbf{W}))_k = \xi_k W_k - W_k \xi_{k-1}, \qquad 1 \le k \le N,$$

with the boundary convention $\xi_0 = \xi_N = 0$.

For $\xi \in \mathfrak{k}$, set

$$\mu_\xi(\mathbf{W}) := \langle \mu(\mathbf{W}),\, \xi \rangle.$$

Lemma 7.7. *For every $\xi \in \mathfrak{k}$,*

$$\operatorname{grad} \mu_\xi(\mathbf{W}) = J K_\xi(\mathbf{W}) \qquad \textit{for every } \mathbf{W} \in \mathrm{GL}(d,\mathbb{C})^N.$$

Proof. By Theorem 4.2,

$$\mathrm{d}\mu_\xi = \iota_{K_\xi}\omega.$$

Since $\omega(U, V) = g(JU, V)$, for every tangent vector Z we have

$$g(\operatorname{grad} \mu_\xi, Z) = \mathrm{d}\mu_\xi(Z) = \omega(K_\xi, Z) = g(JK_\xi, Z).$$

Thus $\operatorname{grad} \mu_\xi = JK_\xi$. $\square$

Proof of Proposition 7.1. By Proposition 6.2, every value of μ is regular. Therefore

$$T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = \ker(\mathrm{d}\mu)_{\mathbf{W}}.$$

Choose a real basis $\{\xi_a\}$ of $\mathfrak{k}$ and set

$$\mu_a(\mathbf{W}) := \langle \mu(\mathbf{W}),\, \xi_a \rangle.$$

Near $\mathbf{W}$, the level set $\mathcal{M}_{\mathbf{G}}$ is cut out by the independent equations

$$\mu_a(\mathbf{W}) = \langle \mathbf{G},\, \xi_a \rangle.$$

Hence the normal space is the span of the gradients of these defining functions:

$$\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = \operatorname{span}\{\operatorname{grad} \mu_a(\mathbf{W})\}.$$

By Lemma 7.7,

$$\operatorname{grad} \mu_a(\mathbf{W}) = JK_{\xi_a}(\mathbf{W}).$$

The vectors $K_{\xi_a}(\mathbf{W})$ span $\mathcal{V}_{\mathbf{W}}$, so the vectors $JK_{\xi_a}(\mathbf{W})$ span $J\mathcal{V}_{\mathbf{W}}$. Thus

$$\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = J\mathcal{V}_{\mathbf{W}}.$$

Taking orthogonal complements gives

$$T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = (J\mathcal{V}_{\mathbf{W}})^{\perp_g}.$$

□

The next identity gives the infinitesimal form of the fact that a nonzero moment value is not fixed by the whole group K.

Lemma 7.8. *For every $\mathbf{W} \in \mathrm{GL}(d,\mathbb{C})^N$ and every $\xi, \eta \in \mathfrak{k}$,*

$$g\Big(K_\xi(\mathbf{W}), JK_\eta(\mathbf{W})\Big) = -\langle \mu(\mathbf{W}),\, [\xi,\eta]\rangle.$$

In particular, if $\mathbf{W} \in \mathcal{M}$, then

$$g\Big(K_\xi(\mathbf{W}), JK_\eta(\mathbf{W})\Big) = 0.$$

Proof. Using $\omega(U,V) = g(JU,V)$ and the symmetry of g, we obtain

$$g(K_\xi, JK_\eta) = g(JK_\eta, K_\xi) = \omega(K_\eta, K_\xi).$$

The moment-map identity gives

$$\omega(K_\eta, K_\xi) = \mathrm{d}\mu_\eta(K_\xi) = K_\xi(\mu_\eta).$$

By infinitesimal equivariance of the moment map,

$$K_\xi(\mu_\eta) = -\mu_{[\xi,\eta]} = -\langle \mu,\, [\xi,\eta]\rangle.$$

This proves the identity. If $\mathbf{W} \in \mathcal{M}$, then $\mu(\mathbf{W}) = \mathbf{0}$, and the right-hand side vanishes. $\square$

The next two lemmas record the infinitesimal invariance of the end-to-end map. They will be used in the submersion argument and in the stochastic calculation.

Lemma 7.9. *For every $\mathbf{W} \in \mathrm{GL}(d, \mathbb{C})^N$ and every $\xi \in \mathfrak{k}$,*

$$(\mathrm{d}\phi)_{\mathbf{W}}\Big(K_\xi(\mathbf{W})\Big) = 0.$$

Equivalently,

$$\mathcal{V}_{\mathbf{W}} \subset \ker(\mathrm{d}\phi)_{\mathbf{W}}.$$

Proof. The map ϕ is invariant under the action of K. Hence

$$\phi(\exp(t\xi) \cdot \mathbf{W}) = \phi(\mathbf{W})$$

for every t. Differentiating at $t = 0$ gives the identity. $\square$

Lemma 7.10. *For every $\mathbf{W} \in \mathrm{GL}(d, \mathbb{C})^N$,*

$$(\mathrm{d}\phi)_{\mathbf{W}}\Big(J\mathcal{V}_{\mathbf{W}}\Big) = 0.$$

Consequently, for every $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$,

$$(\mathrm{d}\phi)_{\mathbf{W}}\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W})\Big) = 0. \tag{7.10}$$

Proof. The map ϕ is holomorphic. Therefore

$$(\mathrm{d}\phi)_{\mathbf{W}}(JZ) = \mathrm{i}\,(\mathrm{d}\phi)_{\mathbf{W}}(Z)$$

for every tangent vector Z. Apply this identity to $Z \in \mathcal{V}_{\mathbf{W}}$, and use Lemma 7.9. This proves the first claim.

If $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$, then Proposition 7.1 gives

$$\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = J\mathcal{V}_{\mathbf{W}}.$$

The mean curvature vector of $\mathcal{M}_{\mathbf{G}}$ is normal to $\mathcal{M}_{\mathbf{G}}$, so (7.10) follows. □

For $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$, Lemma 7.9, Proposition 7.1, and Lemma 7.10 give the structural triad

$$\mathcal{V}_{\mathbf{W}} \subset \ker(\mathrm{d}\phi)_{\mathbf{W}}, \qquad J\mathcal{V}_{\mathbf{W}} = \mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}, \qquad (\mathrm{d}\phi)_{\mathbf{W}}(J\mathcal{V}_{\mathbf{W}}) = 0.$$

Thus the orbit through $\mathbf{W}$ controls the constraint in two different ways. The tangent space $\mathcal{V}_{\mathbf{W}}$ consists of the symmetry directions along which the end-to-end map ϕ is constant. Proposition 7.1 shows that the rotated space $J\mathcal{V}_{\mathbf{W}}$ is exactly the normal space to $\mathcal{M}_{\mathbf{G}}$.

The whole orbit $K \cdot \mathbf{W}$ need not remain inside $\mathcal{M}_{\mathbf{G}}$. Indeed, moment-map equivariance gives, for $U = (U_{N-1}, \dots, U_1) \in K$,

$$\mu(U \cdot \mathbf{W})_k = U_k \, \mu(\mathbf{W})_k \, U_k^*, \qquad 1 \le k \le N-1.$$

If $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$, then $U \cdot \mathbf{W}$ belongs to $\mathcal{M}_{\mathbf{G}}$ exactly when

$$U_k \, G_k \, U_k^* = G_k, \qquad 1 \le k \le N-1.$$

Equivalently, U belongs to the stabilizer subgroup $K_{\mathbf{G}}$ introduced in Chapter 6. Thus only the smaller orbit $K_{\mathbf{G}} \cdot \mathbf{W}$ is guaranteed to remain inside the fixed level set. Lemma 7.8 is the infinitesimal version of this statement. The balanced case is special

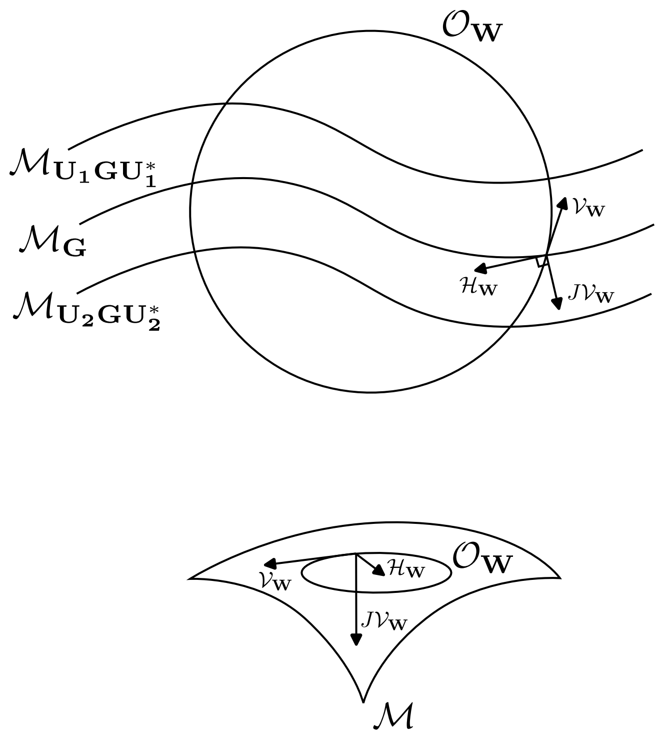


Figure 7.1: Group orbits and level sets of the moment map. Top: for a nonzero moment value $\mathbf{G} \neq \mathbf{0}$, the orbit $\mathcal{O}_{\mathbf{W}} = K \cdot \mathbf{W}$ usually does not remain inside $\mathcal{M}_{\mathbf{G}}$; instead it moves through the conjugate level sets $\mu^{-1}(\mathbf{UGU}^*)$. The tangent space $\mathcal{V}_{\mathbf{W}} = T_{\mathbf{W}}\mathcal{O}_{\mathbf{W}}$ consists of directions tangent to the orbit, while $J\mathcal{V}_{\mathbf{W}}$ is the normal space to $\mathcal{M}_{\mathbf{G}}$. Bottom: in the balanced case $\mathbf{G} = \mathbf{0}$, the whole orbit lies in $\mathcal{M}$.

because $\mathbf{0}$ is fixed by every element of K. Hence the whole orbit lies in $\mathcal{M}$.

Figure 7.1 illustrates this distinction.

The normal-bundle description also shows that ϕ remains a submersion after restriction to any regular moment level. This is the smooth analogue, for $\mathcal{M}_{\mathbf{G}}$, of the Riemannian submersion used on $\mathcal{M}$ in Chapter 5.

Theorem 7.11. *For every* $\mathbf{G} \in \mathrm{Herm}(d)^{N-1}$*, the restriction*

$$\phi|_{\mathcal{M}_{\mathbf{G}}} : \mathcal{M}_{\mathbf{G}} \to \mathrm{GL}(d, \mathbb{C})$$

is a smooth submersion. Consequently, its image is open in $\mathrm{GL}(d, \mathbb{C})$*, and for every*

$X \in \phi(\mathcal{M}_{\mathbf{G}})$ *the fibre*

$$\mathcal{F}_X \cap \mathcal{M}_{\mathbf{G}}$$

is a smooth submanifold of $\mathcal{M}_{\mathbf{G}}$ *of real dimension* $(N-1)d^2$.

Proof. Fix $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$. Since all layers are invertible, the differential

$$(\mathrm{d}\phi)_{\mathbf{W}} : T_{\mathbf{W}}\mathrm{GL}(d, \mathbb{C})^N \to T_{\phi(\mathbf{W})}\mathrm{GL}(d, \mathbb{C})$$

is surjective by the submersion argument following (4.11).

By Lemma 7.10, $(\mathrm{d}\phi)_{\mathbf{W}}$ vanishes on $J\mathcal{V}_{\mathbf{W}}$. By Proposition 7.1,

$$\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = J\mathcal{V}_{\mathbf{W}}.$$

Thus $(\mathrm{d}\phi)_{\mathbf{W}}$ vanishes on $\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}$.

Let $Y \in T_{\phi(\mathbf{W})}\mathrm{GL}(d, \mathbb{C})$ be arbitrary. Choose $Z \in T_{\mathbf{W}}\mathrm{GL}(d, \mathbb{C})^N$ such that

$$(\mathrm{d}\phi)_{\mathbf{W}}(Z) = Y.$$

Decompose Z into its tangent and normal parts:

$$Z = Z^{\top} + Z^{\perp}, \qquad Z^{\top} \in T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}, \quad Z^{\perp} \in \mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}.$$

Since $(\mathrm{d}\phi)_{\mathbf{W}}(Z^{\perp}) = 0$, we get

$$Y = (\mathrm{d}\phi)_{\mathbf{W}}(Z) = (\mathrm{d}\phi)_{\mathbf{W}}(Z^{\top}).$$

Thus the restriction of $(\mathrm{d}\phi)_{\mathbf{W}}$ to $T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}$ is surjective. Hence $\phi|_{\mathcal{M}_{\mathbf{G}}}$ is a submersion.

The image of a submersion is open. If $X \in \phi(\mathcal{M}_{\mathbf{G}})$, then the rank theorem gives

$$\dim_{\mathbb{R}}(\mathcal{F}_X \cap \mathcal{M}_{\mathbf{G}}) = \dim_{\mathbb{R}} \mathcal{M}_{\mathbf{G}} - \dim_{\mathbb{R}} \mathrm{GL}(d, \mathbb{C}).$$

By Proposition 6.2,

$$\dim_{\mathbb{R}} \mathcal{M}_{\mathbf{G}} = (N+1)d^2,$$

while

$$\dim_{\mathbb{R}} \mathrm{GL}(d, \mathbb{C}) = 2d^2.$$

Therefore

$$\dim_{\mathbb{R}}(\mathcal{F}_X \cap \mathcal{M}_{\mathbf{G}}) = (N+1)d^2 - 2d^2 = (N-1)d^2. \qquad \square$$

7.2.1 Volume of the orbit and entropy

For every $\mathbf{W} \in \mathrm{GL}(d, \mathbb{C})^N$, the orbit

$$\mathcal{O}_{\mathbf{W}} = K \cdot \mathbf{W}$$

is a compact submanifold of $\mathrm{GL}(d, \mathbb{C})^N$, diffeomorphic to K. Its Riemannian volume, computed with the metric induced from the ambient flat metric, defines the smooth function

$$S_{\mathrm{orb}}(\mathbf{W}) := \log \mathrm{vol}(\mathcal{O}_{\mathbf{W}}). \tag{7.11}$$

Proof of Proposition 7.4. Let $\mathbf{W} \in \mathcal{M}$ and set $X = \phi(\mathbf{W})$. By Proposition 4.6, the balanced factorizations over X form a single K-orbit. Since

$$\mathcal{O}_X = \mathcal{F}_X \cap \mathcal{M} = (\phi|_{\mathcal{M}})^{-1}(X),$$

we have

$$\mathcal{O}_X = K \cdot \mathbf{W} = \mathcal{O}_{\mathbf{W}}.$$

Taking logarithms of the induced volumes gives

$$S_{\text{orb}}(\mathbf{W}) = \log \operatorname{vol}(\mathcal{O}_{\mathbf{W}}) = \log \operatorname{vol}(\mathcal{O}_X) = S_N(X).$$

$\square$

Remark 7.12. Proposition 7.4 identifies the restriction of the orbit-volume function S_{orb} to the balanced manifold with the entropy S_N on end-to-end space.

For a general value $\mathbf{G}$, the set

$$\mathcal{O}_{X,\mathbf{G}} = \mathcal{F}_X \cap \mathcal{M}_{\mathbf{G}}$$

may be empty. When it is nonempty, it need not be a single K-orbit, so its volume is not determined by $\operatorname{vol}(\mathcal{O}_{\mathbf{W}})$ for one point $\mathbf{W} \in \mathcal{O}_{X,\mathbf{G}}$.

A natural open problem is to find a closed formula for

$$\log \operatorname{vol}(\mathcal{O}_{X,\mathbf{G}})$$

in terms of the singular values of X and the eigenvalues of the prescribed moment matrices G_k, in the spirit of the entropy formula [MY26b, Remark 7]. We do not derive such a formula here.

7.3 The mean curvature of $\mathcal{M}_{\mathbf{G}}$

We now prove Theorem 7.2. We keep the notation introduced in Section 7.1. Thus

$$K_a := K_{\xi_a}, \qquad \mu_a(\mathbf{W}) := \langle \mu(\mathbf{W}), \xi_a \rangle,$$

and

$$h_{ab}(\mathbf{W}) = g\Big(K_a(\mathbf{W}), K_b(\mathbf{W})\Big), \qquad h^{ab}(\mathbf{W}) = (h(\mathbf{W})^{-1})^{ab}.$$

Repeated Lie-algebra indices are summed. Freeness of the K-action implies that $h_{ab}(\mathbf{W})$ is symmetric and positive definite. Since J is a g-isometry,

$$g\Big(JK_a(\mathbf{W}), JK_b(\mathbf{W})\Big) = h_{ab}(\mathbf{W}).$$

The proof has three parts. First we express the mean curvature through the Hessians of the scalar functions μ_a. Then we convert the trace over the normal directions $J\mathcal{V}_{\mathbf{W}}$ into a trace over the directions tangent to the K-orbit. Finally, we identify the resulting contraction with the first variation of S_{orb}, up to the Lie bracket correction that appears at nonzero moment value.

Lemma 7.13. *Let $B : T_{\mathbf{W}}\mathrm{GL}(d,\mathbb{C})^N \times T_{\mathbf{W}}\mathrm{GL}(d,\mathbb{C})^N \to E$ be a bilinear map with values in a real vector space E. Then*

$$\mathrm{Tr}_{\mathcal{V}_{\mathbf{W}}} B = h^{ab}(\mathbf{W})\, B\Big(K_a(\mathbf{W}), K_b(\mathbf{W})\Big),$$

and

$$\mathrm{Tr}_{J\mathcal{V}_{\mathbf{W}}} B = h^{ab}(\mathbf{W})\, B\Big(JK_a(\mathbf{W}), JK_b(\mathbf{W})\Big).$$

Proof. Set $e_a := K_a(\mathbf{W})$. If $u_i = c_i^{\,a} e_a$ is a g-orthonormal basis of $\mathcal{V}_{\mathbf{W}}$, then the matrix

$C = (c_i^a)$ satisfies

$$ChC^\top = I.$$

Hence $C^\top C = h^{-1}$, and therefore

$$\sum_i B(u_i, u_i) = \sum_i c_i^a c_i^b B(e_a, e_b) = h^{ab} B(e_a, e_b).$$

The second identity follows by applying the same argument to the basis Je_a of $J\mathcal{V}_{\mathbf{W}}$. $\square$

Proposition 7.14 ([Bes87; Lee18]). *Let Σ be an embedded submanifold of $\mathrm{GL}(d, \mathbb{C})^N$, and let $f : \mathrm{GL}(d, \mathbb{C})^N \to \mathbb{R}$ be smooth. If f is constant on Σ, then*

$$g\Big(\vec{H}_\Sigma, \operatorname{grad} f\Big) = -\mathrm{Tr}_{T\Sigma} \nabla^2 f,$$

where $\vec{H}_\Sigma = \mathrm{Tr}_{T\Sigma} B_\Sigma$ is the mean curvature vector of Σ in the ambient metric g.

Proof. Let $\{e_i\}$ be a local g-orthonormal frame for $T\Sigma$. Since the ambient connection is flat,

$$\Delta_\Sigma(f|_\Sigma) = \sum_i \Big(e_i(e_i f) - (\nabla^\Sigma_{e_i} e_i) f \Big).$$

The Gauss formula gives

$$\nabla_{e_i} e_i = \nabla^\Sigma_{e_i} e_i + B_\Sigma(e_i, e_i).$$

Hence

$$e_i(e_i f) - (\nabla^\Sigma_{e_i} e_i) f = \nabla^2 f(e_i, e_i) + g\Big(\operatorname{grad} f, B_\Sigma(e_i, e_i)\Big).$$

After summing over i, we obtain

$$\Delta_\Sigma(f|_\Sigma) = \mathrm{Tr}_{T\Sigma}\nabla^2 f + g\Big(\operatorname{grad} f, \vec{H}_\Sigma\Big).$$

The left-hand side vanishes because f is constant on Σ. $\square$

Lemma 7.15. *For every a,*

$$\Delta\mu_a = 0$$

with respect to the ambient metric g.

Proof. Write

$$\xi_a = (\xi_{a,N-1}, \dots, \xi_{a,1}) \in \mathfrak{u}(d)^{N-1}, \qquad \xi_{a,0} = \xi_{a,N} = 0.$$

Using the pairing (4.3),

$$\mu_a(\mathbf{W}) = -\frac{1}{2}\sum_{k=1}^{N-1} \Im\operatorname{tr}\Big((W_k W_k^* - W_{k+1}^* W_{k+1})\xi_{a,k}\Big).$$

We first record the one-layer computation. For $\eta \in \mathfrak{u}(d)$, the real Laplacian associated with $g = \Re\operatorname{tr}(\cdot^*\cdot)$ satisfies

$$\Delta_W\left(-\frac{1}{2}\Im\operatorname{tr}(WW^*\eta)\right) = -2d\,\Im\operatorname{tr}(\eta),$$

and

$$\Delta_W\left(\frac{1}{2}\Im\operatorname{tr}(W^*W\eta)\right) = 2d\,\Im\operatorname{tr}(\eta).$$

Indeed, by unitary invariance we may assume $\eta = \mathrm{i}\operatorname{diag}(\lambda_1, \dots, \lambda_d)$. Then

$$\Im\operatorname{tr}(WW^*\eta) = \sum_{i=1}^{d}\lambda_i\sum_{j=1}^{d}|W_{ij}|^2,$$

and the real Laplacian of $|z|^2$ on $\mathbb{C} \cong \mathbb{R}^2$ is 4. The first identity follows. The second is the same computation applied to the columns of W.

Now fix a layer index ℓ. Only the two terms involving W_ℓ contribute to $\Delta_{W_\ell}\mu_a$, so

$$\Delta_{W_\ell}\mu_a = 2d\,\Im\operatorname{tr}(\xi_{a,\ell-1}) - 2d\,\Im\operatorname{tr}(\xi_{a,\ell}).$$

Summing over $\ell = 1, \dots, N$ gives the telescoping cancellation

$$\Delta\mu_a = 2d\sum_{\ell=1}^{N}\Big(\Im\operatorname{tr}(\xi_{a,\ell-1}) - \Im\operatorname{tr}(\xi_{a,\ell})\Big) = 2d\,\Im\operatorname{tr}(\xi_{a,0}) - 2d\,\Im\operatorname{tr}(\xi_{a,N}) = 0.$$

$\square$

Fix $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$. Since μ_a is constant on $\mathcal{M}_{\mathbf{G}}$, Proposition 7.14 gives

$$g\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}, JK_a\Big) = -\operatorname{Tr}_{T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}\nabla^2\mu_a.$$

By Lemma 7.15,

$$0 = \Delta\mu_a = \operatorname{Tr}_{T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}\nabla^2\mu_a + \operatorname{Tr}_{\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}\nabla^2\mu_a.$$

Therefore

$$g\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}, JK_a\Big) = \operatorname{Tr}_{\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}\nabla^2\mu_a.$$

Using Proposition 7.1,

$$\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = J\mathcal{V}_{\mathbf{W}},$$

and Lemma 7.13 gives

$$g\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}, JK_a\Big) = h^{bc}\,\nabla^2\mu_a(JK_b, JK_c). \tag{7.12}$$

Lemma 7.16. *For every a and all tangent vectors Z, Z',*

$$\nabla^2\mu_a(JZ, JZ') = \nabla^2\mu_a(Z, Z').$$

Proof. In one matrix factor, the second differential of $W \mapsto WW^*$ in the directions Z, Z' is

$$ZZ'^* + Z'Z^*,$$

and the second differential of $W \mapsto W^*W$ is

$$Z^*Z' + Z'^*Z.$$

Replacing Z and Z' by $JZ = \mathrm{i}Z$ and $JZ' = \mathrm{i}Z'$ leaves both expressions unchanged:

$$(\mathrm{i}Z)(\mathrm{i}Z')^* = ZZ'^*, \qquad (\mathrm{i}Z)^*(\mathrm{i}Z') = Z^*Z'.$$

Since μ_a is a real linear combination of such quadratic terms, its Hessian is unchanged. $\square$

Combining (7.12) with Lemma 7.16, we obtain

$$g\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}, JK_a\Big) = h^{bc}\,\nabla^2\mu_a(K_b, K_c). \tag{7.13}$$

The remaining step is to identify the contraction on the right with the variation of the volume of the K-orbit.

Lemma 7.17. *For every a,*

$$h^{bc}\,\nabla^2\mu_a(K_b, K_c) = (JK_a)S_{\mathrm{orb}} + h^{bc}\,\langle \mu(\mathbf{W}),\, [[\xi_a, \xi_b], \xi_c]\rangle.$$

Proof. All quantities in the proof are evaluated at the fixed point $\mathbf{W}$, unless the basepoint is displayed.

Differentiate

$$h_{bc} = g(K_b, K_c)$$

in the direction JK_a. The K-action is holomorphic and isometric, so each K_b is a holomorphic Killing field. On a Kähler manifold,

$$\nabla_{JK_a} K_b = J\nabla_{K_a} K_b.$$

Thus

$$\begin{aligned}(JK_a)h_{bc} &= g\Big(J\nabla_{K_a} K_b, K_c\Big) + g\Big(J\nabla_{K_a} K_c, K_b\Big)\\ &= \nabla^2\mu_b(K_a, K_c) + \nabla^2\mu_c(K_a, K_b),\end{aligned}$$

where we used $\operatorname{grad}\mu_b = JK_b$ from Lemma 7.7.

Since the ambient connection is torsion-free,

$$\nabla_{K_a} K_b = \nabla_{K_b} K_a + [K_a, K_b].$$

For a left action, the fundamental vector fields satisfy

$$[K_a, K_b] = -K_{[\xi_a,\xi_b]}.$$

Therefore

$$\nabla^2\mu_b(K_a, K_c) = \nabla^2\mu_a(K_b, K_c) - g\Big(JK_{[\xi_a,\xi_b]}, K_c\Big),$$

and, with b and c interchanged,

$$\nabla^2\mu_c(K_a, K_b) = \nabla^2\mu_a(K_b, K_c) - g\Big(JK_{[\xi_a,\xi_c]}, K_b\Big).$$

By Lemma 7.8 and skew-symmetry of the Lie bracket,

$$g\Big(JK_{[\xi_a,\xi_b]}, K_c\Big) = \langle \mu,\ [[\xi_a, \xi_b], \xi_c]\rangle.$$

The same identity applies to the second bracket term. Hence

$$\begin{aligned}(JK_a)h_{bc} = 2\nabla^2\mu_a(K_b, K_c)& \\ &- \langle \mu,\ [[\xi_a, \xi_b], \xi_c]\rangle - \langle \mu,\ [[\xi_a, \xi_c], \xi_b]\rangle.\end{aligned}$$

Contracting with h^{bc}, and using the symmetry of h^{bc}, gives

$$h^{bc}(JK_a)h_{bc} = 2h^{bc}\nabla^2\mu_a(K_b, K_c) - 2h^{bc}\langle \mu,\ [[\xi_a, \xi_b], \xi_c]\rangle.$$

Thus

$$h^{bc}\nabla^2\mu_a(K_b, K_c) = \frac{1}{2}h^{bc}(JK_a)h_{bc} + h^{bc}\langle \mu,\ [[\xi_a, \xi_b], \xi_c]\rangle. \tag{7.14}$$

It remains to identify the first term with $(JK_a)S_{\text{orb}}$. The orbit map

$$\mathfrak{y}_{\mathbf{W}} : K \to \mathcal{O}_{\mathbf{W}}, \qquad \mathfrak{y}_{\mathbf{W}}(U) = U \cdot \mathbf{W},$$

is a diffeomorphism because the action is free. In the invariant frame on K determined by the fixed basis $\{\xi_a\}$, the Gram matrix of the pulled-back metric is $h_{ab}(\mathbf{W})$. Hence

$$\operatorname{vol}(\mathcal{O}_{\mathbf{W}}) = C_K\ \det(h_{ab}(\mathbf{W}))^{1/2},$$

where C_K is independent of $\mathbf{W}$. Taking logarithms and applying Jacobi's formula yields

$$(JK_a)S_{\text{orb}} = (JK_a)\log\operatorname{vol}(\mathcal{O}_{\mathbf{W}}) = \frac{1}{2}h^{bc}(JK_a)h_{bc}.$$

Substitution into (7.14) proves the claim. $\square$

Proof of Theorem 7.2. Fix $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$. By Proposition 7.1,

$$\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W}) \in \mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = J\mathcal{V}_{\mathbf{W}}.$$

Therefore

$$\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W}) = H^d JK_d(\mathbf{W})$$

for some coefficients H^d. Pairing with $JK_a(\mathbf{W})$ gives

$$h_{ad}H^d = g\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}, JK_a\Big).$$

By (7.13) and Lemma 7.17,

$$g\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}, JK_a\Big) = (JK_a)S_{\mathrm{orb}} + h^{bc}\langle \mu(\mathbf{W}),\, [[\xi_a, \xi_b], \xi_c]\rangle.$$

On $\mathcal{M}_{\mathbf{G}}$, we have $\mu(\mathbf{W}) = \mathbf{G}$. Hence

$$H^d = h^{ad}(JK_a)S_{\mathrm{orb}} + h^{ad}h^{bc}\langle \mathbf{G},\, [[\xi_a, \xi_b], \xi_c]\rangle.$$

Multiplying by $JK_d(\mathbf{W})$, we obtain

$$\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W}) = h^{ad}\Big((JK_a)S_{\mathrm{orb}}\Big)\, JK_d(\mathbf{W}) + h^{ad}h^{bc}\langle \mathbf{G},\, [[\xi_a, \xi_b], \xi_c]\rangle\, JK_d(\mathbf{W}).$$

The first term is $\nabla^{\perp}S_{\mathrm{orb}}(\mathbf{W})$ by (7.2), and the second is $\mathcal{C}_{\mathbf{G}}(\mathbf{W})$ by (7.3). This proves

$$\vec{H}_{\mathcal{M}_{\mathbf{G}}} = \nabla^{\perp}S_{\mathrm{orb}} + \mathcal{C}_{\mathbf{G}}.$$

$\square$

When $\mathbf{G} = \mathbf{0}$, the bracket correction $\mathcal{C}_{\mathbf{0}}$ vanishes. Thus Theorem 7.2 gives, on the balanced manifold,

$$\vec{H}_{\mathcal{M}} = \nabla^{\perp} S_{\mathrm{orb}}.$$

By Proposition 7.4, this is the same orbit volume whose logarithm descends to the entropy S_N on end-to-end space.

Remark 7.18. The term $\mathcal{C}_{\mathbf{G}}$ is a Lie-bracket correction. It enters because the trace over the directions tangent to the K-orbit is related to the first variation of the orbit volume by the bracket relation for fundamental vector fields. Its disappearance at $\mathbf{G} = \mathbf{0}$ reflects the special role of the balanced manifold, where the whole K-orbit remains in the fixed level set of the moment map.

A coordinate formula for

$$\nabla^{\perp} S_{\mathrm{orb}} \qquad \text{and} \qquad \mathcal{C}_{\mathbf{G}}$$

would require expressing h^{ab}, the derivatives $(JK_a)S_{\mathrm{orb}}$, and the bracket contraction in spectral or quotient coordinates. The formula in (7.4) is intrinsic to the ambient level set and does not carry out that reduction.

7.4 Brownian motion on $\mathcal{M}_{\mathbf{G}}$ and its pushforward

We now prove Theorem 7.3. Brownian motion on the level set $\mathcal{M}_{\mathbf{G}} \subset \mathrm{GL}(d, \mathbb{C})^N$ has Euclidean Itô drift given by the mean curvature vector of that level set. The end-to-end map ϕ kills this normal drift, so the drift of the pushforward is the Itô correction coming from the Hessian of ϕ.

All stochastic equations in this section are understood up to the lifetime of the process in $\mathrm{GL}(d,\mathbb{C})^N$.

7.4.1 Brownian motion on $\mathcal{M}_{\mathbf{G}}$

Let $\mathbf{B}_t$ be standard Brownian motion in the real Euclidean vector space underlying $\mathrm{M}_d(\mathbb{C})^N$. For $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$, let

$$\Pi_{\mathbf{W}} : T_{\mathbf{W}}\mathrm{M}_d(\mathbb{C})^N \longrightarrow T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}$$

be the g-orthogonal projection. Brownian motion $\mathbf{W}_t$ on $(\mathcal{M}_{\mathbf{G}}, \iota_{\mathbf{G}})$, with generator $\frac{1}{2}\Delta_{\mathcal{M}_{\mathbf{G}}}$, has the extrinsic Stratonovich form

$$\mathrm{d}\mathbf{W}_t = \Pi_{\mathbf{W}_t} \circ \mathrm{d}\mathbf{B}_t. \tag{7.15}$$

Since the ambient connection is flat, the corresponding Itô equation is

$$\mathrm{d}\mathbf{W}_t = \Pi_{\mathbf{W}_t}\,\mathrm{d}\mathbf{B}_t + \frac{1}{2}\,\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W}_t)\,\mathrm{d}t. \tag{7.16}$$

By Theorem 7.2, the drift in (7.16) is

$$\mathrm{d}\mathbf{W}_t = \Pi_{\mathbf{W}_t}\,\mathrm{d}\mathbf{B}_t + \frac{1}{2}\,\nabla^{\perp} S_{\mathrm{orb}}(\mathbf{W}_t)\,\mathrm{d}t + \frac{1}{2}\,\mathcal{C}_{\mathbf{G}}(\mathbf{W}_t)\,\mathrm{d}t. \tag{7.17}$$

In the balanced case $\mathbf{G} = \mathbf{0}$, the bracket correction vanishes, and (7.17) becomes

$$\mathrm{d}\mathbf{W}_t = \Pi_{\mathbf{W}_t}\,\mathrm{d}\mathbf{B}_t + \frac{1}{2}\,\nabla^{\perp} S_{\mathrm{orb}}(\mathbf{W}_t)\,\mathrm{d}t. \tag{7.18}$$

The drift in (7.17) is normal to $\mathcal{M}_{\mathbf{G}}$. Indeed, by Proposition 7.1,

$$\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = J\mathcal{V}_{\mathbf{W}},$$

and both terms in the right-hand side of (7.4) take values in this normal bundle. By Lemma 7.10, $(\mathrm{d}\phi)_{\mathbf{W}}$ annihilates this normal space. Thus the mean curvature drift of $\mathbf{W}_t$ has no first-order image under ϕ.

7.4.2 The end-to-end process

Let $\mathbf{W}_t$ be Brownian motion on $(\mathcal{M}_{\mathbf{G}}, \iota_{\mathbf{G}})$, and set

$$X_t := \phi(\mathbf{W}_t).$$

Lemma 7.10 removes the first-order drift in (7.16). The drift of X_t is therefore the second-order term in Itô's formula.

Proposition 7.19. *Let $\mathbf{W}_t$ solve* (7.16)*, and set $X_t = \phi(\mathbf{W}_t)$. Then*

$$\mathrm{d}X_t = \mathrm{d}M_t + \frac{1}{2}\,\mathrm{Tr}_{T_{\mathbf{W}_t}\mathcal{M}_{\mathbf{G}}}(\nabla^2\phi)_{\mathbf{W}_t}\,\mathrm{d}t, \tag{7.19}$$

where the local martingale M_t is given by

$$\mathrm{d}M_t := (\mathrm{d}\phi)_{\mathbf{W}_t}\Big(\Pi_{\mathbf{W}_t}\,\mathrm{d}\mathbf{B}_t\Big).$$

Proof. Apply the Euclidean Itô formula to

$$\phi : \mathrm{GL}(d,\mathbb{C})^N \longrightarrow \mathrm{GL}(d,\mathbb{C}) \subset \mathrm{M}_d(\mathbb{C})$$

along (7.16). The first-order terms are

$$(\mathrm{d}\phi)_{\mathbf{W}_t}\Big(\Pi_{\mathbf{W}_t}\,\mathrm{d}\mathbf{B}_t\Big)$$

and

$$\frac{1}{2}\,(\mathrm{d}\phi)_{\mathbf{W}_t}\Big(\vec{H}_{\mathcal{M}_{\mathbf{G}}}(\mathbf{W}_t)\Big)\,\mathrm{d}t.$$

The second term vanishes by Lemma 7.10.

Let $\{E_\alpha\}$ be a fixed g-orthonormal basis of the ambient real Euclidean vector space. The quadratic variation of the martingale part contributes

$$\frac{1}{2}\sum_{\alpha}(\nabla^2\phi)_{\mathbf{W}_t}\Big(\Pi_{\mathbf{W}_t}E_\alpha,\Pi_{\mathbf{W}_t}E_\alpha\Big)\,\mathrm{d}t.$$

Since $\Pi_{\mathbf{W}_t}$ is the orthogonal projection onto $T_{\mathbf{W}_t}\mathcal{M}_{\mathbf{G}}$, the projected family $\{\Pi_{\mathbf{W}_t}E_\alpha\}$ is a Parseval frame for $T_{\mathbf{W}_t}\mathcal{M}_{\mathbf{G}}$. The last sum is therefore the metric trace over $T_{\mathbf{W}_t}\mathcal{M}_{\mathbf{G}}$, which proves (7.19). □

It remains to identify the trace in (7.19). The ambient Laplacian of ϕ vanishes, and the normal trace can be converted into a trace over the directions tangent to the K-orbit.

Lemma 7.20. *With respect to the ambient flat metric,*

$$\Delta\phi = 0$$

as a $\mathrm{M}_d(\mathbb{C})$*-valued function.*

Proof. Use a constant ambient orthonormal basis adapted to the factors of $\mathrm{M}_d(\mathbb{C})^N$.

If Z is supported in one factor, then along the affine line

$$\mathbf{W}(t) = \mathbf{W} + tZ$$

the product $\phi(\mathbf{W}(t))$ is affine in t, because only one factor varies. Hence

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\bigg|_{t=0} \phi(\mathbf{W} + tZ) = 0.$$

Summing these second derivatives over the constant orthonormal basis gives $\Delta\phi = 0$. $\square$

Lemma 7.21. *For every* $\mathbf{W} \in \mathrm{GL}(d, \mathbb{C})^N$ *and all* $Z, Z' \in T_{\mathbf{W}}\mathrm{GL}(d, \mathbb{C})^N$*,*

$$(\nabla^2\phi)_{\mathbf{W}}(JZ, JZ') = -(\nabla^2\phi)_{\mathbf{W}}(Z, Z').$$

Proof. The source and target have their standard flat Kähler connections, and ϕ is holomorphic. Thus

$$(\mathrm{d}\phi) \circ J = J_{\mathrm{tgt}} \circ (\mathrm{d}\phi), \qquad J_{\mathrm{tgt}}(Y) = \mathrm{i}Y.$$

Differentiating this identity with respect to the flat connections gives

$$(\nabla^2\phi)(JZ, Z') = J_{\mathrm{tgt}}(\nabla^2\phi)(Z, Z'), \qquad (\nabla^2\phi)(Z, JZ') = J_{\mathrm{tgt}}(\nabla^2\phi)(Z, Z').$$

Applying these identities in both arguments yields

$$(\nabla^2\phi)(JZ, JZ') = J_{\mathrm{tgt}}^2(\nabla^2\phi)(Z, Z') = -(\nabla^2\phi)(Z, Z').$$

$\square$

Proof of Theorem 7.3. Fix $\mathbf{W} \in \mathcal{M}_{\mathbf{G}}$. By Lemma 7.20 and the orthogonal decomposition

$$T_{\mathbf{W}}\mathrm{GL}(d, \mathbb{C})^N = T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} \oplus \mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}},$$

we have

$$0 = \mathrm{Tr}_{T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}(\nabla^2\phi)_{\mathbf{W}} + \mathrm{Tr}_{\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}(\nabla^2\phi)_{\mathbf{W}}. \tag{7.20}$$

By Proposition 7.1,

$$\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}} = J\mathcal{V}_{\mathbf{W}}.$$

Lemma 7.13 therefore gives

$$\mathrm{Tr}_{\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}(\nabla^2\phi)_{\mathbf{W}} = h^{ab}(\mathbf{W})(\nabla^2\phi)_{\mathbf{W}}\Big(JK_a(\mathbf{W}), JK_b(\mathbf{W})\Big).$$

By Lemma 7.21,

$$\mathrm{Tr}_{\mathcal{N}_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}(\nabla^2\phi)_{\mathbf{W}} = -h^{ab}(\mathbf{W})(\nabla^2\phi)_{\mathbf{W}}\Big(K_a(\mathbf{W}), K_b(\mathbf{W})\Big).$$

Substituting this identity into (7.20) yields

$$\mathrm{Tr}_{T_{\mathbf{W}}\mathcal{M}_{\mathbf{G}}}(\nabla^2\phi)_{\mathbf{W}} = h^{ab}(\mathbf{W})(\nabla^2\phi)_{\mathbf{W}}\Big(K_a(\mathbf{W}), K_b(\mathbf{W})\Big) = \mathcal{D}_{\mathbf{G}}(\mathbf{W}), \tag{7.21}$$

where the last equality is (7.5). Applying (7.21) along $\mathbf{W}_t$ in Proposition 7.19 proves (7.6). □

Remark 7.22. Theorem 7.3 identifies the drift of the pushed-forward process. The local martingale part is still expressed through the projected ambient Brownian motion,

$$\mathrm{d}M_t = (\mathrm{d}\phi)_{\mathbf{W}_t}\Big(\Pi_{\mathbf{W}_t}\,\mathrm{d}\mathbf{B}_t\Big).$$

A formula in coordinates would require an explicit expression for the covariance of this projected noise after applying $\mathrm{d}\phi$. We do not use such coordinates in this chapter.

Remark 7.23. For a general value $\mathbf{G}$, the drift $\mathcal{D}_{\mathbf{G}}(\mathbf{W}_t)$ is a section along $\phi|_{\mathcal{M}_{\mathbf{G}}}$. It need not depend only on $X_t = \phi(\mathbf{W}_t)$. Equivalently, $\mathcal{D}_{\mathbf{G}}$ need not be constant on the fibres of $\phi|_{\mathcal{M}_{\mathbf{G}}}$. This is the stochastic counterpart of the geometric fact that

$$\mathcal{O}_{X,\mathbf{G}} = \mathcal{F}_X \cap \mathcal{M}_{\mathbf{G}}$$

is usually not a single compact orbit. The balanced manifold is special: its fibres are the orbits described in Proposition 7.4. The next section proves that the drift descends to the entropy drift on the space of end-to-end matrices.

7.5 End-to-end process induced by Brownian motion on $\mathcal{M}$

We now specialize the preceding construction to the balanced level

$$\mathcal{M} = \mu^{-1}(\mathbf{0}).$$

On this level the compact orbit through $\mathbf{W}$ lies inside $\mathcal{M}$, and Theorem 5.4 gives the Riemannian submersion

$$\phi : (\mathcal{M}, \iota) \longrightarrow (\mathrm{GL}(d, \mathbb{C}), g_N).$$

This is the point at which the section $\mathcal{D}_{\mathbf{G}}$ from Theorem 7.3 becomes a vector field on end-to-end space.

There are two mean-curvature vectors in play. The vector $\vec{H}_{\mathcal{M}}$ is the ambient mean curvature of the embedded submanifold $\mathcal{M} \subset \mathrm{GL}(d, \mathbb{C})^N$. It is normal to $\mathcal{M}$, and Lemma 7.10 shows that $\mathrm{d}\phi$ kills it. The vector $\vec{H}_{\mathrm{fib}}$ is the mean curvature of the compact fibre $\mathcal{O}_X \subset \mathcal{M}$. Its pushforward is $-\operatorname{grad}_{g_N} S_N$. The entropy drift comes from the fibre mean curvature, not from pushing forward the ambient normal drift.

For $\mathbf{W} \in \mathcal{M}$, write

$$\mathcal{V}_{\mathbf{W}} := T_{\mathbf{W}}(K \cdot \mathbf{W}) \subset T_{\mathbf{W}}\mathcal{M}, \qquad \mathcal{H}_{\mathbf{W}} := \mathcal{V}_{\mathbf{W}}^{\perp_\iota}.$$

Then $\mathcal{V}$ is the vertical distribution of $\phi : \mathcal{M} \to \mathrm{GL}(d, \mathbb{C})$, and $\mathcal{H}$ is the horizontal distribution.

Definition 7.24. For $\mathbf{W} \in \mathcal{M}$, the mean curvature vector of the fibre of $\phi|_{\mathcal{M}}$ through $\mathbf{W}$ is

$$\vec{H}_{\mathrm{fib}}(\mathbf{W}) := h^{ab}(\mathbf{W}) \left(\nabla^{\mathcal{M}}_{K_a} K_b\right)^{\mathrm{hor}}(\mathbf{W}),$$

where $(\cdot)^{\mathrm{hor}}$ denotes ι-orthogonal projection onto $\mathcal{H}_{\mathbf{W}}$.

This vector is the trace of the second fundamental form of the fibre inside $(\mathcal{M}, \iota)$. The expression in Definition 7.24 is independent of the chosen basis $\{\xi_a\}$ of $\mathfrak{k}$, since h^{ab} contracts the trace over the vertical tangent space.

Lemma 7.25. *For every* $\mathbf{W} \in \mathcal{M}$*,*

$$\mathcal{D}_{\mathbf{0}}(\mathbf{W}) = -(\mathrm{d}\phi)_{\mathbf{W}}\Big(\vec{H}_{\mathrm{fib}}(\mathbf{W})\Big).$$

Proof. All quantities are evaluated at the fixed point $\mathbf{W}$. By Lemma 7.9,

$$(\mathrm{d}\phi)(K_a) = 0 \qquad \text{for every } a.$$

With the flat connections on source and target, this gives

$$\nabla^2\phi(K_a, K_b) = -(\mathrm{d}\phi)(\nabla_{K_a} K_b).$$

Use the Gauss formula for the embedding $\mathcal{M} \subset \mathrm{GL}(d, \mathbb{C})^N$:

$$\nabla_{K_a} K_b = \nabla^{\mathcal{M}}_{K_a} K_b + B_{\mathcal{M}}(K_a, K_b).$$

The second term is normal to $\mathcal{M}$. By Proposition 7.1, with $\mathbf{G} = \mathbf{0}$, this normal space is $J\mathcal{V}_{\mathbf{W}}$. Hence Lemma 7.10 gives

$$(\mathrm{d}\phi)\Big(B_{\mathcal{M}}(K_a, K_b)\Big) = 0.$$

Therefore

$$\nabla^2\phi(K_a, K_b) = -(\mathrm{d}\phi)\Big(\nabla^{\mathcal{M}}_{K_a} K_b\Big).$$

The vertical part of $\nabla^{\mathcal{M}}_{K_a} K_b$ is also annihilated by $\mathrm{d}\phi$. Thus only its horizontal part contributes:

$$\nabla^2\phi(K_a, K_b) = -(\mathrm{d}\phi)\Big((\nabla^{\mathcal{M}}_{K_a} K_b)^{\mathrm{hor}}\Big).$$

Contracting with h^{ab} and using (7.5) with $\mathbf{G} = \mathbf{0}$ proves the identity. □

The next lemma identifies this mean curvature vector with the gradient of the entropy on the base.

Lemma 7.26. *Let $\mathbf{W} \in \mathcal{M}$, and set $X = \phi(\mathbf{W})$. Then*

$$(\mathrm{d}\phi)_{\mathbf{W}}\Big(\vec{H}_{\mathrm{fib}}(\mathbf{W})\Big) = -\Big(\mathrm{grad}_{g_N} S_N\Big)_X.$$

Proof. Let Y be a smooth vector field on $\mathrm{GL}(d, \mathbb{C})$, and let Y^{hor} be its horizontal lift to $\mathcal{M}$. Since Y^{hor} is ϕ-related to Y, its flow maps each fibre of $\phi|_{\mathcal{M}}$ to a nearby fibre.

The first variation formula for the volume of a compact submanifold gives

$$Y \ \mathrm{vol}(\mathcal{O}_X) = -\int_{\mathcal{O}_X} \iota\Big(\vec{H}_{\mathrm{fib}}, Y^{\mathrm{hor}}\Big) \, \mathrm{d}\,\mathrm{vol}_{\mathcal{O}_X} .$$

The K-action preserves ι, preserves the horizontal distribution, and acts transitively on $\mathcal{O}_X$. Hence $\vec{H}_{\mathrm{fib}}$ and Y^{hor} are K-equivariant along the fibre, and their inner product is constant on $\mathcal{O}_X$. Therefore

$$Y \, S_N(X) = -\iota_{\mathbf{W}}\Big(\vec{H}_{\mathrm{fib}}(\mathbf{W}), Y^{\mathrm{hor}}(\mathbf{W})\Big).$$

Because $\phi : (\mathcal{M}, \iota) \to (\mathrm{GL}(d, \mathbb{C}), g_N)$ is a Riemannian submersion,

$$\iota_{\mathbf{W}}\Big(\vec{H}_{\mathrm{fib}}(\mathbf{W}), Y^{\mathrm{hor}}(\mathbf{W})\Big) = \Big(g_N\Big)_X\Big((\mathrm{d}\phi)_{\mathbf{W}}\vec{H}_{\mathrm{fib}}(\mathbf{W}), Y(X)\Big).$$

Thus

$$Y \, S_N(X) = -\Big(g_N\Big)_X\Big((\mathrm{d}\phi)_{\mathbf{W}}\vec{H}_{\mathrm{fib}}(\mathbf{W}), Y(X)\Big)$$

for every vector field Y. This is exactly the stated gradient identity. $\square$

Proof of Theorem 7.5. Let $X = \phi(\mathbf{W})$. Lemma 7.25 gives

$$\mathcal{D}_{\mathbf{0}}(\mathbf{W}) = -(\mathrm{d}\phi)_{\mathbf{W}}\Big(\vec{H}_{\mathrm{fib}}(\mathbf{W})\Big).$$

Lemma 7.26 identifies the right-hand side with $(\mathrm{grad}_{g_N} S_N)_X$. This proves (7.7). $\square$

Proof of Corollary 7.6. Let $f \in C^{\infty}(\mathrm{GL}(d, \mathbb{C}))$, and set

$$u := f \circ \phi.$$

Then u is constant on the fibres of $\phi|_{\mathcal{M}}$.

Choose a local g_N-orthonormal frame

$$e_1, \dots, e_{2d^2}$$

on $\mathrm{GL}(d, \mathbb{C})$, and let

$$E_1, \dots, E_{2d^2}$$

be the corresponding horizontal lifts to $\mathcal{M}$. Complete these vectors to a local ι-orthonormal frame

$$\{E_1, \dots, E_{2d^2}, V_1, \dots, V_{(N-1)d^2}\},$$

where the V_a are vertical. Since $V_a u = 0$, the Laplacian of u is

$$\begin{aligned} \Delta_{\mathcal{M}} u = \sum_i \Big(E_i(E_i u) - (\nabla^{\mathcal{M}}_{E_i} E_i) u \Big) \\ - \sum_a (\nabla^{\mathcal{M}}_{V_a} V_a) u. \end{aligned}$$

The first sum equals $(\Delta_{g_N} f) \circ \phi$, because the E_i are horizontal lifts and ϕ is a Riemannian submersion.

For the second sum, only the horizontal part of $\sum_a \nabla^{\mathcal{M}}_{V_a} V_a$ acts on u. Hence, by Definition 7.24,

$$\sum_a (\nabla^{\mathcal{M}}_{V_a} V_a) u = \mathrm{d}u\Big(\vec{H}_{\mathrm{fib}}\Big).$$

Using $u = f \circ \phi$ and Lemma 7.26,

$$\mathrm{d}u\Big(\vec{H}_{\mathrm{fib}}\Big) = \mathrm{d}f\Big((\mathrm{d}\phi)\vec{H}_{\mathrm{fib}}\Big) = -\Big\langle \mathrm{grad}_{g_N} S_N, \mathrm{grad}_{g_N} f \Big\rangle_{g_N} \circ \phi.$$

Therefore

$$\Delta_{\mathcal{M}}(f \circ \phi) = \Big(\Delta_{g_N} f + \Big\langle \operatorname{grad}_{g_N} S_N, \operatorname{grad}_{g_N} f \Big\rangle_{g_N}\Big) \circ \phi.$$

Brownian motion on $(\mathcal{M}, \iota)$ has generator $\frac{1}{2}\Delta_{\mathcal{M}}$. The generator of $X_t = \phi(\mathbf{W}_t)$ is therefore the operator in (7.8). The local Itô form (7.9) is the standard expression for a diffusion with this generator. □

Remark 7.27. Corollary 7.6 gives an intrinsic description of the diffusion on $(\mathrm{GL}(d, \mathbb{C}), g_N)$. A formula in a chosen coordinate system would require explicit g_N-orthonormal frames and the corresponding connection coefficients. We do not pursue such a formula for the closure result.

Bibliographical notes

The stochastic and submanifold identities used here are standard. For Brownian motion on Riemannian manifolds and embedded submanifolds, including the extrinsic Itô description, see Hsu [Hsu02]. For the mean-curvature identities and the first variation of submanifold volume used in the fibre-volume argument, see Besse [Bes87] and Lee [Lee18].

The stochastic motivation for DLNs comes from the entropy and the quotient geometry developed earlier, especially in Chapter 5, and from the Riemannian Langevin formulation of Menon and Yu [MY26a]. Related work includes Huang, Inauen, and Menon [HIM23], where Itô drift is identified as motion by mean curvature of the orbit.

Part III

Infinite-Depth Limit

Chapter 8

Discrete Exterior Calculus and Flows on the Fibre

We recast the geometry of a fibre at finite depth as a discrete Hodge theory on the layer path $0, 1, \dots, N$. The moment tuple $\mathbf{G}$ becomes the codifferential $D^*_{\mathbf{W}}\mathbf{W}$, balancedness becomes $D^*_{\mathbf{W}}\mathbf{W} = 0$, and the ridge and Kirwan–Ness flows become heat-type equations along the path. The final section records the formal scaling laws leading to the interval model of Chapter 9.

8.1 Main results

Fix $N \geq 2$, $d \geq 1$, and $X \in \mathrm{GL}(d, \mathbb{C})$. Let

$$\mathbf{W} = (W_N, \dots, W_1) \in \mathcal{F}_X$$

be a full-rank factorization. We use the cochain spaces

$$\mathcal{C}^0_N = \{a = (a_{N-1}, \dots, a_1) : a_k \in \mathfrak{gl}(d, \mathbb{C})\}, \qquad \mathcal{C}^1_N = \{\xi = (\xi_N, \dots, \xi_1) : \xi_k \in \mathrm{M}_d(\mathbb{C})\},$$

with the boundary convention $a_0 = a_N = 0$. The discrete differential at $\mathbf{W}$ is

$$D_{\mathbf{W}} : \mathcal{C}^0_N \to \mathcal{C}^1_N, \qquad (D_{\mathbf{W}} a)_k = a_k W_k - W_k a_{k-1}, \qquad 1 \le k \le N.$$

Its adjoint is taken with respect to the Frobenius Hermitian products on $\mathcal{C}^0_N$ and $\mathcal{C}^1_N$. In coordinates,

$$(D^*_{\mathbf{W}} \xi)_k = \xi_k W^*_k - W^*_{k+1} \xi_{k+1}, \qquad 1 \le k \le N-1.$$

The first result identifies the fibre with the image of this differential and identifies the moment tuple with its codifferential.

Theorem 8.1. *Let $X \in \mathrm{GL}(d, \mathbb{C})$ and let $\mathbf{W} \in \mathcal{F}_X$. Then every vertical tangent vector at $\mathbf{W}$ has a unique representation*

$$v = D_{\mathbf{W}} a, \qquad a \in \mathcal{C}^0_N.$$

Equivalently,

$$T_{\mathbf{W}} \mathcal{F}_X = D_{\mathbf{W}}(\mathcal{C}^0_N), \qquad D_{\mathbf{W}} \textit{ is injective.}$$

Moreover,

$$\left(D^*_{\mathbf{W}} \mathbf{W}\right)_k = W_k W^*_k - W^*_{k+1} W_{k+1}, \qquad 1 \le k \le N-1.$$

Thus

$$D^*_{\mathbf{W}} \mathbf{W} = \mathbf{G}(\mathbf{W}),$$

and $\mathbf{W}$ is balanced if and only if $D^*_{\mathbf{W}} \mathbf{W} = 0$.

Define the degree-0 and degree-1 Hodge Laplacians by $\Delta^{(0)}_{\mathbf{W}} := D^*_{\mathbf{W}} D_{\mathbf{W}}$ and $\Delta^{(1)}_{\mathbf{W}} := D_{\mathbf{W}} D^*_{\mathbf{W}}$. The next result gives the coordinate form of the vertical Laplacian.

Theorem 8.2. *For $a \in \mathcal{C}^0_N$,*

$$(\Delta^{(0)}_{\mathbf{W}} a)_k = -W_k a_{k-1} W_k^* + a_k W_k W_k^* + W_{k+1}^* W_{k+1} a_k - W_{k+1}^* a_{k+1} W_{k+1},$$

for $1 \le k \le N-1$. In particular,

$$\langle a,\, \Delta^{(0)}_{\mathbf{W}} a\rangle_0 = \|D_{\mathbf{W}} a\|_1^2.$$

Since $D_{\mathbf{W}}$ is injective on a full-rank fibre, $\Delta^{(0)}_{\mathbf{W}}$ is positive definite on $\mathcal{C}^0_N$.

Thus the block tridiagonal operator from the finite-depth fibre geometry is the Hodge Laplacian $D^*_{\mathbf{W}} D_{\mathbf{W}}$. The induced metric on the fibre is its Dirichlet form:

$$g^{\iota}_{\mathbf{W}}(D_{\mathbf{W}} a, D_{\mathbf{W}} c) = \Re\,\langle a,\, \Delta^{(0)}_{\mathbf{W}} c\rangle_0.$$

The normal metric is the coefficient metric

$$g^{\mathrm{nor}}_{\mathbf{W}}(D_{\mathbf{W}} a, D_{\mathbf{W}} c) = \Re\,\langle a,\, c\rangle_0.$$

Let

$$G := \mathbf{G}(\mathbf{W}) = D^*_{\mathbf{W}} \mathbf{W}.$$

On the fixed fibre define

$$\mathcal{R}(\mathbf{W}) = \frac{1}{2}\sum_{k=1}^{N} \mathrm{tr}(W_k^* W_k), \qquad \mathcal{E}(\mathbf{W}) = \frac{1}{2}\sum_{k=1}^{N-1} \mathrm{tr}(G_k^2).$$

Here $\mathcal{R}$ is the fibrewise ridge regularizer and $\mathcal{E}$ is the Kirwan–Ness functional.

The heat equations follow from one identity.

Functional and metric	**Coefficient α in $\dot{\mathbf{W}} = -D_{\mathbf{W}}\alpha$**	**Moment evolution**
$\mathcal{R}$ with g^ι	u, where $\Delta^{(0)}_{\mathbf{W}} u = G$	$\dot{G} = -2G$
$\mathcal{R}$ with g^{nor}	G	$\dot{G} = -\left(\Delta^{(0)}_{\mathbf{W}} G + (\Delta^{(0)}_{\mathbf{W}} G)^*\right)$
$\mathcal{E}$ with g^ι	$2G$	$\dot{G} = -2\left(\Delta^{(0)}_{\mathbf{W}} G + (\Delta^{(0)}_{\mathbf{W}} G)^*\right)$
$\mathcal{E}$ with g^{nor}	$2\Delta^{(0)}_{\mathbf{W}} G$	$\dot{G} = -2\left((\Delta^{(0)}_{\mathbf{W}})^2 G + ((\Delta^{(0)}_{\mathbf{W}})^2 G)^*\right)$

Table 8.1: Discrete fibrewise gradient flows and their corresponding moment evolutions.

Theorem 8.3. *Let $\mathbf{W}(t) \in \mathcal{F}_X$ be a smooth curve of full-rank factorizations. Suppose*

$$\dot{\mathbf{W}} = -D_{\mathbf{W}}\alpha, \qquad \alpha(t) \in \mathcal{C}^0_N.$$

Then the moment tuple $G(t) = \mathbf{G}(\mathbf{W}(t))$ satisfies

$$\dot{G} = -\left(\Delta^{(0)}_{\mathbf{W}}\alpha + \left(\Delta^{(0)}_{\mathbf{W}}\alpha\right)^*\right).$$

The finite-dimensional gradient flows computed in Section 8.3 are summarized in Table 8.1. In the first row, u denotes the unique solution of $\Delta^{(0)}_{\mathbf{W}} u = G$.

Theorem 8.4 (One heat equation, two gradient structures)**.** *On the full-rank fibre $\mathcal{F}_X$, with $G = D^*_{\mathbf{W}}\mathbf{W}$, one has*

$$-\operatorname{grad}_{g^{\mathrm{nor}}} \mathcal{R}(\mathbf{W}) = -\frac{1}{2}\operatorname{grad}_{g^\iota} \mathcal{E}(\mathbf{W}) = -D_{\mathbf{W}} G = -\Delta^{(1)}_{\mathbf{W}}\mathbf{W}. \tag{8.1}$$

Consequently, the g^{nor}-gradient flow of $\mathcal{R}$ and the g^ι-gradient flow of $\mathcal{E}$ trace the same curves on the fibre, with the constant time change $\tau = 2t$. This common equation is the discrete Hodge heat equation on the fibre.

We now record the formal scaling laws. Set $I = [0,1]$, $h = 1/N$, and $x_k = kh$, and assume $W_k^h = \exp\Big(h\, b(x_k)\Big)$, with $b : I \to \mathfrak{gl}(d, \mathbb{C})$. The partial products then satisfy the formal transport equation

$$X'(x) = b(x)X(x), \qquad X(0) = I.$$

For a fixed end-to-end matrix $X \in \mathrm{GL}(d, \mathbb{C})$, the continuum fibre is obtained by imposing the endpoint condition $X(1) = X$.

The next two statements are formal continuum statements. They are Taylor-expansion limits of the discrete expressions. The stated $O(h)$ remainders are asserted under C^2 regularity. With only C^1 data, the same limiting expressions hold with $o(1)$ errors, or with errors controlled by the modulus of continuity of the first derivatives.

Proposition 8.5. *Assume that a, b, ξ are C^2 on I, and assume the based boundary condition*

$$a(0) = a(1) = 0.$$

Set

$$a_k = a(x_k), \qquad \xi_k = \xi(x_k).$$

Under the scaling $W_k^h = \exp(hb(x_k))$,

$$\frac{1}{h}(D_{\mathbf{W}^h} a)_k = a'(x_k) + [a(x_k), b(x_k)] + O(h),$$

and

$$\frac{1}{h}(D^*_{\mathbf{W}^h} \xi)_k = -\xi'(x_k) + [\xi(x_k), b(x_k)^*] + O(h),$$

uniformly in k. Thus the formal limiting operators are

$$\nabla_b a = a' + [a, b], \qquad \nabla_b^* \xi = -\xi' + [\xi, b^*].$$

The operator ∇_b^* is the L^2-adjoint of ∇_b under the based condition $a(0) = a(1) = 0$. The corresponding continuum Laplacians are

$$\Delta_b^{(0)} = \nabla_b^* \nabla_b, \qquad \Delta_b^{(1)} = \nabla_b \nabla_b^*.$$

Proposition 8.6. *Assume that b is C^2 on I. Define the renormalized discrete moment by*

$$\omega_k^h := \frac{1}{h^2}\Big(W_k^h (W_k^h)^* - (W_{k+1}^h)^* W_{k+1}^h\Big), \qquad 1 \le k \le N-1.$$

Then

$$\omega_k^h = -(b+b^*)'(x_k) + [b(x_k), b(x_k)^*] + O(h),$$

uniformly in k. Hence the formal continuum moment field is

$$\mu_\infty(b) = -(b+b^*)' + [b, b^*].$$

Equivalently,

$$\mu_\infty(b) = \nabla_b^*(b+b^*).$$

The continuum equation for balancedness suggested by the discrete equation $D_{\mathbf{W}}^* \mathbf{W} = 0$ is therefore

$$\mu_\infty(b) = 0.$$

The ridge regularizer itself does not produce a nontrivial continuum fibre functional under this scaling. The object that survives is the renormalized moment energy

$$\frac{1}{2}\int_0^1 \mathrm{tr}\Big(\mu_\infty(b)(x)^* \mu_\infty(b)(x)\Big)\, dx,$$

which becomes the continuum Kirwan–Ness functional in Chapter 9.

8.2 A layerwise discrete exterior calculus

We now prove the finite-depth Hodge statements from Section 8.1. Fix a depth $N \geq 2$ and a width $d \geq 1$. The oriented path has vertices $0, 1, \ldots, N$, and edge k is oriented from $k-1$ to k.

A matrix $a_k \in \mathfrak{gl}(d, \mathbb{C})$ at the internal vertex k represents an infinitesimal change of frame at that layer. A variation of the tuple of weights assigns a matrix to each edge. Thus the natural cochain spaces are

$$\begin{aligned} \mathcal{C}_N^0 &:= \{a = (a_{N-1}, \ldots, a_1) : a_k \in \mathfrak{gl}(d, \mathbb{C})\}\,, \\ \mathcal{C}_N^1 &:= \{\xi = (\xi_N, \ldots, \xi_1) : \xi_k \in \mathrm{M}_d(\mathbb{C})\}\,. \end{aligned} \tag{8.2}$$

We impose the based boundary condition $a_0 = a_N = 0$. These conditions at the endpoints express that a fibre fixes the input and output frames.

Equip the two cochain spaces with the Frobenius Hermitian products

$$\langle a,\, b\rangle_0 = \sum_{k=1}^{N-1} \operatorname{tr}(a_k^* b_k), \qquad \langle \xi,\, \eta\rangle_1 = \sum_{k=1}^{N} \operatorname{tr}(\xi_k^* \eta_k). \tag{8.3}$$

Definition 8.7. For a tuple of weights

$$\mathbf{W} = (W_N, \ldots, W_1) \in \mathrm{GL}(d, \mathbb{C})^N,$$

the discrete differential at $\mathbf{W}$ is

$$D_{\mathbf{W}} : \mathcal{C}_N^0 \to \mathcal{C}_N^1, \qquad (D_{\mathbf{W}} a)_k = a_k W_k - W_k a_{k-1}, \qquad 1 \leq k \leq N. \tag{8.4}$$

Its adjoint

$$D_{\mathbf{W}}^* : \mathcal{C}_N^1 \to \mathcal{C}_N^0$$

is defined by

$$\langle D_{\mathbf{W}} a,\, \xi \rangle_1 = \langle a,\, D_{\mathbf{W}}^* \xi \rangle_0. \tag{8.5}$$

Remark 8.8. If $W_k = I$ for every k, then

$$(D_{\mathbf{W}} a)_k = a_k - a_{k-1}, \qquad (D_{\mathbf{W}}^* \xi)_k = \xi_k - \xi_{k+1}. \tag{8.6}$$

Thus $D_{\mathbf{W}}$ is the ordinary discrete differential on the path, and $D_{\mathbf{W}}^*$ is the corresponding codifferential. For a general tuple $\mathbf{W}$, the matrix W_k transports the coefficient at vertex $k-1$ across edge k. In this sense, $D_{\mathbf{W}}$ is a covariant differential along the path.

Lemma 8.9. *The adjoint of $D_{\mathbf{W}}$ is*

$$(D_{\mathbf{W}}^* \xi)_k = \xi_k W_k^* - W_{k+1}^* \xi_{k+1}, \qquad 1 \le k \le N-1. \tag{8.7}$$

Proof. Using (8.4), we expand

$$\langle D_{\mathbf{W}} a,\, \xi \rangle_1 = \sum_{k=1}^{N} \operatorname{tr}\Big((a_k W_k - W_k a_{k-1})^* \xi_k \Big).$$

Since

$$(a_k W_k - W_k a_{k-1})^* = W_k^* a_k^* - a_{k-1}^* W_k^*,$$

cyclic invariance of trace gives

$$\langle D_{\mathbf{W}} a,\, \xi \rangle_1 = \sum_{k=1}^{N} \operatorname{tr}(a_k^* \xi_k W_k^*) - \sum_{k=1}^{N} \operatorname{tr}(a_{k-1}^* W_k^* \xi_k).$$

Reindex the second sum:

$$\sum_{k=1}^{N} \operatorname{tr}(a_{k-1}^* W_k^* \xi_k) = \sum_{k=0}^{N-1} \operatorname{tr}(a_k^* W_{k+1}^* \xi_{k+1}).$$

The terms with a_0 and a_N vanish by the based boundary condition $a_0 = a_N = 0$. Therefore

$$\langle D_{\mathbf{W}} a,\, \xi \rangle_1 = \sum_{k=1}^{N-1} \operatorname{tr}\Big(a_k^* (\xi_k W_k^* - W_{k+1}^* \xi_{k+1}) \Big),$$

which is exactly (8.7). □

Remark 8.10. If one inserts the mesh factor $h = 1/N$ into both inner products,

$$\langle a,\, b \rangle_{0,h} = h \sum_{k=1}^{N-1} \operatorname{tr}(a_k^* b_k), \qquad \langle \xi,\, \eta \rangle_{1,h} = h \sum_{k=1}^{N} \operatorname{tr}(\xi_k^* \eta_k), \tag{8.8}$$

then the formula (8.7) is unchanged, because the common factor h cancels in the adjoint relation. These are the Riemann-sum inner products used in the continuum limit.

Lemma 8.11. *Let $a \in \mathcal{C}_N^0$, with the convention $a_0 = a_N = 0$. The vector generated by the curve $e^{ta} \cdot \mathbf{W}$, where $e^{ta} := (e^{ta_{N-1}}, \dots, e^{ta_1})$, is*

$$\left.\frac{d}{dt}\right|_{t=0} \Big(e^{ta} \cdot \mathbf{W} \Big) = D_{\mathbf{W}} a. \tag{8.9}$$

Proof. The group action has components

$$(e^{ta} \cdot \mathbf{W})_k = e^{ta_k} W_k e^{-ta_{k-1}}, \qquad 1 \le k \le N,$$

where $a_0 = a_N = 0$. Differentiating at $t = 0$ gives

$$\left.\frac{d}{dt}\right|_{t=0} (e^{ta_k} W_k e^{-ta_{k-1}}) = a_k W_k - W_k a_{k-1} = (D_{\mathbf{W}} a)_k.$$

□

Proof of Theorem 8.1. Let

$$Y_k := W_k W_{k-1} \cdots W_1, \qquad Y_0 := I.$$

For a variation $v = (v_N, \dots, v_1) \in \mathcal{C}_N^1$, let δY_k denote the induced variation of Y_k. Then

$$\delta Y_k = v_k Y_{k-1} + W_k \delta Y_{k-1}, \qquad \delta Y_0 = 0. \tag{8.10}$$

The variation v is tangent to the fibre $\mathcal{F}_X$ if and only if $\delta Y_N = 0$.

First suppose that $v = D_{\mathbf{W}} a$. We claim that

$$\delta Y_k = a_k Y_k, \qquad 0 \le k \le N.$$

Indeed, using (8.10) and (8.4),

$$\delta Y_k = (a_k W_k - W_k a_{k-1}) Y_{k-1} + W_k a_{k-1} Y_{k-1} = a_k Y_k.$$

Since $a_N = 0$, this gives $\delta Y_N = 0$. Hence $D_{\mathbf{W}}(\mathcal{C}_N^0) \subset T_{\mathbf{W}} \mathcal{F}_X$.

Conversely, let $v \in T_{\mathbf{W}} \mathcal{F}_X$. Since X is invertible, every W_k and every Y_k is invertible. Define

$$a_k := \delta Y_k Y_k^{-1}, \qquad 0 \le k \le N.$$

Then $a_0 = 0$, and the vertical condition $\delta Y_N = 0$ gives $a_N = 0$. Multiplying (8.10) on the right by Y_{k-1}^{-1} gives

$$a_k W_k = v_k + W_k a_{k-1}.$$

Thus

$$v_k = a_k W_k - W_k a_{k-1} = (D_{\mathbf{W}} a)_k.$$

Therefore $T_{\mathbf{W}}\mathcal{F}_X = D_{\mathbf{W}}(\mathcal{C}_N^0)$.

The same construction proves uniqueness. For a given vertical vector v, the quantities δY_k are determined by (8.10), and hence $a_k = \delta Y_k Y_k^{-1}$ is forced. Equivalently, if $D_{\mathbf{W}}a = 0$, then $a_1 W_1 = 0$, so $a_1 = 0$, and induction in k gives $a_k = 0$ for all $1 \leq k \leq N-1$. Thus $D_{\mathbf{W}}$ is injective.

It remains to identify the moment tuple. Apply (8.7) with $\xi = \mathbf{W}$. For $1 \leq k \leq N-1$,

$$(D_{\mathbf{W}}^*\mathbf{W})_k = W_k W_k^* - W_{k+1}^* W_{k+1} = G_k(\mathbf{W}). \tag{8.11}$$

Thus

$$D_{\mathbf{W}}^*\mathbf{W} = \mathbf{G}(\mathbf{W}).$$

By definition, $\mathbf{W}$ is balanced precisely when all the matrices $G_k(\mathbf{W})$ vanish. Hence balancedness is equivalent to $D_{\mathbf{W}}^*\mathbf{W} = 0$. □

Definition 8.12. The degree-0 and degree-1 Hodge Laplacians are $\Delta_{\mathbf{W}}^{(0)} := D_{\mathbf{W}}^* D_{\mathbf{W}} : \mathcal{C}_N^0 \to \mathcal{C}_N^0$ and $\Delta_{\mathbf{W}}^{(1)} := D_{\mathbf{W}} D_{\mathbf{W}}^* : \mathcal{C}_N^1 \to \mathcal{C}_N^1$.

Proof of Theorem 8.2. Using (8.7) with $\xi = D_{\mathbf{W}}a$, we have

$$(\Delta_{\mathbf{W}}^{(0)} a)_k = (D_{\mathbf{W}}a)_k W_k^* - W_{k+1}^*(D_{\mathbf{W}}a)_{k+1}.$$

Substitute

$$(D_{\mathbf{W}}a)_k = a_k W_k - W_k a_{k-1}, \qquad (D_{\mathbf{W}}a)_{k+1} = a_{k+1} W_{k+1} - W_{k+1} a_k.$$

Then, for $1 \le k \le N-1$,

$$\begin{aligned}(\Delta^{(0)}_{\mathbf{W}} a)_k &= (a_k W_k - W_k a_{k-1}) W_k^* - W_{k+1}^* (a_{k+1} W_{k+1} - W_{k+1} a_k) \\ &= - W_k a_{k-1} W_k^* + a_k W_k W_k^* + W_{k+1}^* W_{k+1} a_k - W_{k+1}^* a_{k+1} W_{k+1}. \end{aligned} \tag{8.12}$$

This is the coordinate formula for $\Delta^{(0)}_{\mathbf{W}}$.

The Dirichlet identity follows from adjointness:

$$\langle a,\, \Delta^{(0)}_{\mathbf{W}} a\rangle_0 = \langle a,\, D_{\mathbf{W}}^* D_{\mathbf{W}} a\rangle_0 = \langle D_{\mathbf{W}} a,\, D_{\mathbf{W}} a\rangle_1 = \|D_{\mathbf{W}} a\|_1^2. \tag{8.13}$$

The right-hand side is nonnegative. If it vanishes, then $D_{\mathbf{W}} a = 0$, and Theorem 8.1 gives $a = 0$. Therefore $\Delta^{(0)}_{\mathbf{W}}$ is positive definite on $\mathcal{C}^0_N$. □

Remark 8.13. The block tridiagonal operator in (8.12) is the same operator C_{N-1} from (5.26). This coordinate formula identifies it intrinsically as $\Delta^{(0)}_{\mathbf{W}} = D^*_{\mathbf{W}} D_{\mathbf{W}}$. Thus the operator is not an additional structure on the fibre; it is the degree-0 Hodge Laplacian determined by the layerwise differential.

8.3 Heat flows on the fibre

Fix $X \in \mathrm{GL}(d, \mathbb{C})$ and

$$\mathbf{W} = (W_N, \ldots, W_1) \in \mathcal{F}_X.$$

Throughout this section we write

$$\mathbf{G}(\mathbf{W}) = D^*_{\mathbf{W}} \mathbf{W}.$$

The purpose of the section is to prove the finite-dimensional heat-flow assertions stated in Section 8.1. The vertical tangent space has already been identified with $D_{\mathbf{W}}(\mathcal{C}_N^0)$, and the moment tuple with $D_{\mathbf{W}}^*\mathbf{W}$. We now use these identifications to define the two metrics on the fibre and compute the four fibrewise gradients.

Definition 8.14 (Induced and normal metrics on the fibre)**.** For $a, c \in \mathcal{C}_N^0$, define

$$g_{\mathbf{W}}^{\iota}(D_{\mathbf{W}}a, D_{\mathbf{W}}c) := \Re\,\langle D_{\mathbf{W}}a,\ D_{\mathbf{W}}c\rangle_1, \tag{8.14}$$

$$g_{\mathbf{W}}^{\mathrm{nor}}(D_{\mathbf{W}}a, D_{\mathbf{W}}c) := \Re\,\langle a,\ c\rangle_0. \tag{8.15}$$

The first metric is the one induced by the Frobenius metric on $\mathrm{M}_d(\mathbb{C})^N$. The second metric is obtained by transporting the Frobenius metric on the coefficient space $\mathcal{C}_N^0$ through the isomorphism

$$a \longmapsto D_{\mathbf{W}}a.$$

We call it the *normal metric*.

Proposition 8.15. *For $a, c \in \mathcal{C}_N^0$,*

$$g_{\mathbf{W}}^{\iota}(D_{\mathbf{W}}a, D_{\mathbf{W}}c) = \Re\,\langle a,\ \Delta_{\mathbf{W}}^{(0)}c\rangle_0. \tag{8.16}$$

Proof. By adjointness,

$$g_{\mathbf{W}}^{\iota}(D_{\mathbf{W}}a, D_{\mathbf{W}}c) = \Re\,\langle D_{\mathbf{W}}a,\ D_{\mathbf{W}}c\rangle_1 = \Re\,\langle a,\ D_{\mathbf{W}}^*D_{\mathbf{W}}c\rangle_0 = \Re\,\langle a,\ \Delta_{\mathbf{W}}^{(0)}c\rangle_0.$$

□

Definition 8.16. On the fixed fibre $\mathcal{F}_X$ define

$$\mathcal{R}(\mathbf{W}) := \frac{1}{2}\,\langle \mathbf{W},\, \mathbf{W}\rangle_1 = \frac{1}{2}\sum_{k=1}^{N} \operatorname{tr}(W_k^* W_k), \tag{8.17}$$

$$\mathcal{E}(\mathbf{W}) := \frac{1}{2}\,\langle \mathbf{G},\, \mathbf{G}\rangle_0 = \frac{1}{2}\sum_{k=1}^{N-1} \operatorname{tr}\Big(G_k(\mathbf{W})^2\Big). \tag{8.18}$$

Here $\mathbf{G}(\mathbf{W}) = D_{\mathbf{W}}^* \mathbf{W}$. Thus $\mathcal{R}$ is the fibrewise ridge regularizer, and $\mathcal{E}$ is the fibrewise Kirwan–Ness functional introduced in Chapter 2.

Proposition 8.17. *For $a \in \mathcal{C}_N^0$,*

$$\mathrm{d}\mathcal{R}_{\mathbf{W}}[D_{\mathbf{W}} a] = \Re\,\langle G,\, a\rangle_0. \tag{8.19}$$

Proof. The differential of $\mathcal{R}$ in the direction $D_{\mathbf{W}} a$ is

$$\mathrm{d}\mathcal{R}_{\mathbf{W}}[D_{\mathbf{W}} a] = \Re\,\langle \mathbf{W},\, D_{\mathbf{W}} a\rangle_1.$$

Using adjointness and the identity $D_{\mathbf{W}}^* \mathbf{W} = G$,

$$\Re\,\langle \mathbf{W},\, D_{\mathbf{W}} a\rangle_1 = \Re\,\langle D_{\mathbf{W}}^* \mathbf{W},\, a\rangle_0 = \Re\,\langle G,\, a\rangle_0.$$

□

Proposition 8.18. *For $a \in \mathcal{C}_N^0$,*

$$\mathrm{d}\mathbf{G}_{\mathbf{W}}[D_{\mathbf{W}} a] = \Delta_{\mathbf{W}}^{(0)} a + \Big(\Delta_{\mathbf{W}}^{(0)} a\Big)^*. \tag{8.20}$$

Consequently,

$$\mathrm{d}\mathcal{E}_{\mathbf{W}}[D_{\mathbf{W}} a] = 2\,\Re\,\langle G,\, \Delta_{\mathbf{W}}^{(0)} a\rangle_0. \tag{8.21}$$

Proof. Set

$$v = D_{\mathbf{W}} a.$$

Since

$$G_k(\mathbf{W}) = W_k W_k^* - W_{k+1}^* W_{k+1},$$

we have

$$\mathrm{d}G_{k,\mathbf{W}}[v] = v_k W_k^* + W_k v_k^* - v_{k+1}^* W_{k+1} - W_{k+1}^* v_{k+1}.$$

On the other hand,

$$(\Delta_{\mathbf{W}}^{(0)} a)_k = (D_{\mathbf{W}} a)_k W_k^* - W_{k+1}^* (D_{\mathbf{W}} a)_{k+1} = v_k W_k^* - W_{k+1}^* v_{k+1}.$$

Taking the adjoint gives

$$\Big((\Delta_{\mathbf{W}}^{(0)} a)_k \Big)^* = W_k v_k^* - v_{k+1}^* W_{k+1}.$$

Adding these two identities proves (8.20).

Now differentiate (8.18). Since each G_k is Hermitian,

$$\mathrm{d}\mathcal{E}_{\mathbf{W}}[D_{\mathbf{W}} a] = \Re \langle \mathbf{G},\, \Delta_{\mathbf{W}}^{(0)} a + \Big(\Delta_{\mathbf{W}}^{(0)} a \Big)^* \rangle_0.$$

For Hermitian $\mathbf{G}$,

$$\Re \langle \mathbf{G},\, \Big(\Delta_{\mathbf{W}}^{(0)} a \Big)^* \rangle_0 = \Re \langle \mathbf{G},\, \Delta_{\mathbf{W}}^{(0)} a \rangle_0.$$

Hence

$$\mathrm{d}\mathcal{E}_{\mathbf{W}}[D_{\mathbf{W}} a] = 2\, \Re \langle \mathbf{G},\, \Delta_{\mathbf{W}}^{(0)} a \rangle_0.$$

□

Proof of Theorem 8.3. At each time t, apply Proposition 8.18 at $\mathbf{W}(t)$. If

$$\dot{\mathbf{W}} = -D_{\mathbf{W}}\alpha,$$

then

$$\dot{\mathbf{G}} = \mathrm{d}\mathbf{G}_{\mathbf{W}}[\dot{\mathbf{W}}] = -\mathrm{d}\mathbf{G}_{\mathbf{W}}[D_{\mathbf{W}}\alpha].$$

Using (8.20) with $a = \alpha$ gives

$$\dot{\mathbf{G}} = -\left(\Delta^{(0)}_{\mathbf{W}}\alpha + \left(\Delta^{(0)}_{\mathbf{W}}\alpha\right)^*\right).$$

□

Lemma 8.19. *Let F be a smooth function on $\mathcal{F}_X$, and write every tangent vector as $D_{\mathbf{W}}a$. Let $r \in \mathcal{C}^0_N$.*

1. *If*

$$\mathrm{d}F_{\mathbf{W}}[D_{\mathbf{W}}a] = \Re\langle r,\, a\rangle_0 \qquad \text{for all } a \in \mathcal{C}^0_N,$$

then

$$\operatorname{grad}_{g^{\mathrm{nor}}} F(\mathbf{W}) = D_{\mathbf{W}}r,$$

and

$$\operatorname{grad}_{g^{\iota}} F(\mathbf{W}) = D_{\mathbf{W}}\beta, \qquad \Delta^{(0)}_{\mathbf{W}}\beta = r.$$

2. *If*

$$\mathrm{d}F_{\mathbf{W}}[D_{\mathbf{W}}a] = \Re\langle r,\, \Delta^{(0)}_{\mathbf{W}}a\rangle_0 \qquad \text{for all } a \in \mathcal{C}^0_N,$$

then

$$\operatorname{grad}_{g^\iota} F(\mathbf{W}) = D_{\mathbf{W}} r,$$

and

$$\operatorname{grad}_{g^{\mathrm{nor}}} F(\mathbf{W}) = D_{\mathbf{W}}\Big(\Delta^{(0)}_{\mathbf{W}} r\Big).$$

Proof. By Theorem 8.1, every tangent vector has a unique coefficient. Thus each gradient has the form $D_{\mathbf{W}}\beta$ for a unique $\beta \in \mathcal{C}^0_N$.

For the normal metric,

$$g^{\mathrm{nor}}_{\mathbf{W}}(D_{\mathbf{W}}\beta, D_{\mathbf{W}}a) = \Re\,\langle \beta,\, a\rangle_0.$$

Therefore the first differential identity gives $\beta = r$. If the second identity holds, then self-adjointness of $\Delta^{(0)}_{\mathbf{W}}$ gives

$$\Re\,\langle r,\, \Delta^{(0)}_{\mathbf{W}} a\rangle_0 = \Re\,\langle \Delta^{(0)}_{\mathbf{W}} r,\, a\rangle_0,$$

so $\beta = \Delta^{(0)}_{\mathbf{W}} r$.

For the induced metric, Proposition 8.15 gives

$$g^{\iota}_{\mathbf{W}}(D_{\mathbf{W}}\beta, D_{\mathbf{W}}a) = \Re\,\langle \beta,\, \Delta^{(0)}_{\mathbf{W}} a\rangle_0 = \Re\,\langle \Delta^{(0)}_{\mathbf{W}}\beta,\, a\rangle_0.$$

Thus the first differential identity gives

$$\Delta^{(0)}_{\mathbf{W}}\beta = r.$$

The operator $\Delta^{(0)}_{\mathbf{W}}$ is positive definite by Theorem 8.2, so this equation has a unique solution. If the second differential identity holds, then

$$\Re \langle \beta - r,\, \Delta^{(0)}_{\mathbf{W}} a \rangle_0 = 0 \qquad \text{for all } a.$$

Since $\Delta^{(0)}_{\mathbf{W}}$ is invertible, this implies $\beta = r$. □

Proof of the four finite-dimensional flow formulas in Section 8.1. We use the notation $\mathbf{G} = \mathbf{G}(\mathbf{W})$.

First consider $\mathcal{R}$. By Proposition 8.17,

$$\mathrm{d}\mathcal{R}_{\mathbf{W}}[D_{\mathbf{W}} a] = \Re \langle \mathbf{G},\, a \rangle_0.$$

The first part of Lemma 8.19 gives the g^ι-gradient in the form $D_{\mathbf{W}} u$, where

$$\Delta^{(0)}_{\mathbf{W}} u = \mathbf{G}.$$

Hence the negative g^ι-gradient flow of $\mathcal{R}$ is

$$\dot{\mathbf{W}} = -D_{\mathbf{W}} u, \qquad \Delta^{(0)}_{\mathbf{W}} u = \mathbf{G}. \tag{8.22}$$

Applying Theorem 8.3 with $\alpha = u$ gives

$$\dot{\mathbf{G}} = -\left(\Delta^{(0)}_{\mathbf{W}} u + \left(\Delta^{(0)}_{\mathbf{W}} u \right)^* \right).$$

Since $\mathbf{G}$ is Hermitian, (8.22) yields

$$\dot{\mathbf{G}} = -2\mathbf{G}. \tag{8.23}$$

For the normal metric, the first part of Lemma 8.19 gives

$$\operatorname{grad}_{g^{\mathrm{nor}}} \mathcal{R}(\mathbf{W}) = D_{\mathbf{W}}\mathbf{G}.$$

Since $\mathbf{G} = D^*_{\mathbf{W}}\mathbf{W}$,

$$D_{\mathbf{W}}\mathbf{G} = D_{\mathbf{W}}D^*_{\mathbf{W}}\mathbf{W} = \Delta^{(1)}_{\mathbf{W}}\mathbf{W}.$$

Thus the negative g^{nor}-gradient flow of $\mathcal{R}$ is

$$\dot{\mathbf{W}} = -D_{\mathbf{W}}\mathbf{G} = -\Delta^{(1)}_{\mathbf{W}}\mathbf{W}. \tag{8.24}$$

Theorem 8.3, with $\alpha = \mathbf{G}$, gives

$$\dot{\mathbf{G}} = -\left(\Delta^{(0)}_{\mathbf{W}}\mathbf{G} + \left(\Delta^{(0)}_{\mathbf{W}}\mathbf{G}\right)^*\right). \tag{8.25}$$

Now consider $\mathcal{E}$. By Proposition 8.18,

$$\mathrm{d}\mathcal{E}_{\mathbf{W}}[D_{\mathbf{W}}a] = 2\,\Re\,\langle \mathbf{G},\, \Delta^{(0)}_{\mathbf{W}}a\rangle_0.$$

The second part of Lemma 8.19, with $r = 2\mathbf{G}$, gives

$$\operatorname{grad}_{g^{\iota}} \mathcal{E}(\mathbf{W}) = 2D_{\mathbf{W}}\mathbf{G}.$$

Therefore the negative g^{ι}-gradient flow of $\mathcal{E}$ is

$$\dot{\mathbf{W}} = -2D_{\mathbf{W}}\mathbf{G} = -2\Delta^{(1)}_{\mathbf{W}}\mathbf{W}. \tag{8.26}$$

Theorem 8.3, with $\alpha = 2\mathbf{G}$, gives

$$\dot{\mathbf{G}} = -2\left(\Delta^{(0)}_{\mathbf{W}}\mathbf{G} + \left(\Delta^{(0)}_{\mathbf{W}}\mathbf{G}\right)^*\right). \tag{8.27}$$

Finally, the normal metric and the second part of Lemma 8.19 give

$$\operatorname{grad}_{g^{\mathrm{nor}}} \mathcal{E}(\mathbf{W}) = 2D_{\mathbf{W}}\left(\Delta^{(0)}_{\mathbf{W}}\mathbf{G}\right).$$

Since $\mathbf{G} = D^*_{\mathbf{W}}\mathbf{W}$,

$$D_{\mathbf{W}}\left(\Delta^{(0)}_{\mathbf{W}}\mathbf{G}\right) = D_{\mathbf{W}}D^*_{\mathbf{W}}D_{\mathbf{W}}D^*_{\mathbf{W}}\mathbf{W} = \left(\Delta^{(1)}_{\mathbf{W}}\right)^2\mathbf{W}.$$

Thus the negative g^{nor}-gradient flow of $\mathcal{E}$ is

$$\dot{\mathbf{W}} = -2D_{\mathbf{W}}\left(\Delta^{(0)}_{\mathbf{W}}\mathbf{G}\right) = -2\left(\Delta^{(1)}_{\mathbf{W}}\right)^2\mathbf{W}. \tag{8.28}$$

Theorem 8.3, with

$$\alpha = 2\Delta^{(0)}_{\mathbf{W}}\mathbf{G},$$

gives

$$\dot{\mathbf{G}} = -2\left((\Delta^{(0)}_{\mathbf{W}})^2\mathbf{G} + \left((\Delta^{(0)}_{\mathbf{W}})^2\mathbf{G}\right)^*\right). \tag{8.29}$$

These four computations give the four finite-dimensional flow formulas stated in Section 8.1. □

Proof of Theorem 8.4. Equation (8.24) gives

$$-\operatorname{grad}_{g^{\mathrm{nor}}} \mathcal{R}(\mathbf{W}) = -D_{\mathbf{W}}\mathbf{G}.$$

Equation (8.26) gives

$$-\operatorname{grad}_{g^\iota} \mathcal{E}(\mathbf{W}) = -2D_{\mathbf{W}}\mathbf{G},$$

and therefore

$$-\frac{1}{2}\operatorname{grad}_{g^\iota} \mathcal{E}(\mathbf{W}) = -D_{\mathbf{W}}\mathbf{G}.$$

Finally, since $\mathbf{G} = D^*_{\mathbf{W}}\mathbf{W}$,

$$D_{\mathbf{W}}\mathbf{G} = D_{\mathbf{W}}D^*_{\mathbf{W}}\mathbf{W} = \Delta^{(1)}_{\mathbf{W}}\mathbf{W}.$$

This proves (8.1). The two vector fields differ only by the constant factor 2 before the normalization in (8.1). Hence their integral curves agree after the time change $\tau = 2t$. The moment equations (8.25) and (8.27) differ by the same factor, so the same time change identifies the corresponding evolutions of $\mathbf{G}$. □

Remark 8.20 (Two gradient structures for one heat equation)*.* Theorem 8.4 is analogous to the classical heat equation: in Euclidean space the heat flow is the L^2-gradient flow of the Dirichlet energy, while in Wasserstein geometry it is the gradient flow of entropy [JKO98; Ott01]. Here the same discrete equation along the layer path arises from two pairs,

$$(\mathcal{R}, g^{\mathrm{nor}}) \qquad \text{and} \qquad (\mathcal{E}, g^{\iota}).$$

Changing the metric changes the functional, but not the underlying dissipative dynamics, except for a constant rescaling of time.

Equations (8.24) and (8.26) give the precise heat-equation interpretation. The layer index is the base variable, $G = D^*_{\mathbf{W}}\mathbf{W}$ is the discrete divergence, and balancedness is the zero divergence condition. Thus the fibrewise regularizer dynamics dissipate imbalance along depth.

8.4 The formal continuum limit

In Sections 8.2 and 8.3, we rewrote the geometry of a fibre at depth N as a discrete Hodge theory on the path $0, 1, \ldots, N$. We now let this path converge formally to the interval. The calculations in this section are Taylor expansions under the smoothness assumptions stated below. Whenever an $O(h)$ remainder is claimed after division by h, we assume enough regularity to have second-order Taylor remainders; C^2 is the clean hypothesis. With only C^1 data, the same limiting equations hold, but the corresponding errors are $o(1)$, or more explicitly controlled by the modulus of continuity of the first derivative. In Chapter 9, we place the corresponding continuum objects in Sobolev spaces and give the gauge-theoretic model.

Set $I = [0, 1]$, $h := 1/N$, and $x_k := kh$ for $0 \le k \le N$. Thus the index of the layer becomes the point x_k of the mesh.

8.4.1 Scaling and transport

A nontrivial limit requires each layer to be a small step over an interval of length h. We impose the scaling $W_k^h = \exp\Big(h\, b(x_k)\Big)$, with $b : I \to \mathfrak{gl}(d, \mathbb{C})$. Equivalently,

$$W_k^h = I + h\, b(x_k) + O(h^2),$$

but the exponential form preserves invertibility at every depth.

Let

$$\mathbf{W}^h := (W_N^h, \ldots, W_1^h),$$

and define the partial products

$$X_0^h := I, \qquad X_k^h := W_k^h W_{k-1}^h \cdots W_1^h, \qquad 1 \le k \le N. \tag{8.30}$$

The matrix X_k^h transports the frame at the left endpoint to the point x_k. This is the source of the continuum transport equation.

Write

$$b_k := b(x_k).$$

Since b is bounded on I,

$$\|W_k^h\| \leq \exp(h\|b\|_{L^\infty}), \qquad \|X_k^h\| \leq \exp(\|b\|_{L^\infty}),$$

so the following error estimates are uniform in k. From the Taylor expansion of the exponential,

$$W_k^h = I + h\, b_k + O(h^2).$$

Using $X_k^h = W_k^h X_{k-1}^h$, we get

$$\frac{X_k^h - X_{k-1}^h}{h} = b_k X_{k-1}^h + O(h). \tag{8.31}$$

Thus the formal continuum equation is

$$X'(x) = b(x)X(x), \qquad X(0) = I. \tag{8.32}$$

Equation (8.32) fixes only the left endpoint. For a prescribed end-to-end matrix $X_{\text{end}} \in \mathrm{GL}(d, \mathbb{C})$, the formal continuum fibre is obtained by imposing $X(1) = X_{\text{end}}$. This is the continuum analogue of fixing the product $W_N \cdots W_1$ at depth N.

8.4.2 Operators after rescaling

The operators $D_{\mathbf{W}^h}$ and $D_{\mathbf{W}^h}^*$ are of order h under this near-identity scaling. Their continuum limits therefore come from $h^{-1}D_{\mathbf{W}^h}$ and $h^{-1}D_{\mathbf{W}^h}^*$.

Let

$$a_k := a(x_k), \qquad \xi_k := \xi(x_k),$$

where

$$a : I \to \mathfrak{gl}(d, \mathbb{C}), \qquad \xi : I \to \mathrm{M}_d(\mathbb{C}).$$

We use the based condition

$$a(0) = a(1) = 0,$$

which matches the finite-depth convention $a_0 = a_N = 0$.

Definition 8.21. For a smooth coefficient field $b : I \to \mathfrak{gl}(d, \mathbb{C})$, define

$$\nabla_b a := a' + [a, b], \qquad \nabla_b^* \xi := -\xi' + [\xi, b^*]. \tag{8.33}$$

The associated continuum Laplacians are

$$\Delta_b^{(0)} := \nabla_b^* \nabla_b, \qquad \Delta_b^{(1)} := \nabla_b \nabla_b^*. \tag{8.34}$$

Proof of Proposition 8.5. All error terms below are uniform in the index k, because the functions involved are C^2 on the compact interval I. Write

$$b_k := b(x_k), \qquad W_k^h = I + h\, b_k + O(h^2).$$

For the differential,

$$(D_{\mathbf{W}^h} a)_k = a_k W_k^h - W_k^h a_{k-1}.$$

Taylor's theorem gives

$$a_{k-1} = a_k - h\, a'(x_k) + O(h^2).$$

Hence

$$\begin{aligned}(D_{\mathbf{W}^h} a)_k &= a_k(I + h\, b_k + O(h^2)) - (I + h\, b_k + O(h^2))(a_k - h\, a'(x_k) + O(h^2)) \\ &= h\Big(a'(x_k) + a_k b_k - b_k a_k\Big) + O(h^2) \\ &= h\Big(a'(x_k) + [a(x_k), b(x_k)]\Big) + O(h^2).\end{aligned}$$

Dividing by h gives the first limit.

For the codifferential,

$$(D^*_{\mathbf{W}^h} \xi)_k = \xi_k (W^h_k)^* - (W^h_{k+1})^* \xi_{k+1}, \qquad 1 \leq k \leq N-1.$$

We have

$$(W^h_k)^* = I + h\, b^*_k + O(h^2), \qquad (W^h_{k+1})^* = I + h\, b^*_{k+1} + O(h^2),$$

and

$$\xi_{k+1} = \xi_k + h\, \xi'(x_k) + O(h^2), \qquad b^*_{k+1} = b^*_k + O(h).$$

Therefore

$$\begin{aligned}(D^*_{\mathbf{W}^h} \xi)_k &= \xi_k(I + h\, b^*_k + O(h^2)) \\ &\qquad - (I + h\, b^*_{k+1} + O(h^2))(\xi_k + h\, \xi'(x_k) + O(h^2)) \\ &= h\Big(-\xi'(x_k) + \xi_k b^*_k - b^*_k \xi_k\Big) + O(h^2) \\ &= h\Big(-\xi'(x_k) + [\xi(x_k), b(x_k)^*]\Big) + O(h^2).\end{aligned}$$

Dividing by h gives the second limit. □

The sign in (8.33) is fixed by the L^2 adjoint relation. For smooth a with $a(0) =$

$a(1) = 0$ and smooth ξ, one has

$$\int_0^1 \mathrm{tr}\Big((\nabla_b a)^*(x)\,\xi(x)\Big)\,dx = \int_0^1 \mathrm{tr}\Big(a(x)^*(\nabla_b^*\xi)(x)\Big)\,dx. \tag{8.35}$$

Adjointness calculation. Expand

$$(\nabla_b a)^* = a'^* + [a, b]^* = a'^* + [b^*, a^*].$$

Then

$$\int_0^1 \mathrm{tr}\Big((\nabla_b a)^*\xi\Big)\,dx = \int_0^1 \mathrm{tr}(a'^*\xi)\,dx + \int_0^1 \mathrm{tr}([b^*, a^*]\xi)\,dx.$$

The boundary condition on a gives

$$\int_0^1 \mathrm{tr}(a'^*\xi)\,dx = -\int_0^1 \mathrm{tr}(a^*\xi')\,dx.$$

For the commutator term, cyclicity of trace gives

$$\mathrm{tr}([b^*, a^*]\xi) = \mathrm{tr}(b^*a^*\xi - a^*b^*\xi) = \mathrm{tr}(a^*(\xi b^* - b^*\xi)) = \mathrm{tr}(a^*[\xi, b^*]).$$

Combining these identities yields (8.35). □

8.4.3 The renormalized moment

Balancedness at finite depth is the equation $D^*_{\mathbf{W}}\mathbf{W} = 0$. Under this scaling, the corresponding moment is not of order h but of order h^2. The first-order terms cancel because $W_k^h(W_k^h)^*$ and $(W_k^h)^*W_k^h$ have the same linear part. Thus the continuum field of moments appears only after division by h^2.

Definition 8.22. For the scaled tuple $\mathbf{W}^h = (W_N^h, \ldots, W_1^h)$, define

$$\omega_k^h := \frac{1}{h^2} G_k(\mathbf{W}^h) = \frac{1}{h^2}\Big(W_k^h (W_k^h)^* - (W_{k+1}^h)^* W_{k+1}^h\Big), \qquad 1 \le k \le N-1. \tag{8.36}$$

Definition 8.23. The formal field of moments in the continuum is

$$\mu_\infty(b) := -(b + b^*)' + [b, b^*]. \tag{8.37}$$

Proof of Proposition 8.6. Write

$$b_k := b(x_k), \qquad b_{k+1} := b(x_{k+1}).$$

The exponential expansion gives, uniformly in k,

$$e^{hb_k} e^{hb_k^*} = I + h(b_k + b_k^*) + h^2 \left(\frac{1}{2} b_k^2 + b_k b_k^* + \frac{1}{2}(b_k^*)^2\right) + O(h^3),$$

and

$$e^{hb_{k+1}^*} e^{hb_{k+1}} = I + h(b_{k+1} + b_{k+1}^*) + h^2 \left(\frac{1}{2}(b_{k+1}^*)^2 + b_{k+1}^* b_{k+1} + \frac{1}{2} b_{k+1}^2\right) + O(h^3).$$

Since $b_{k+1} - b_k = O(h)$, the square terms

$$\frac{1}{2} b_k^2 - \frac{1}{2} b_{k+1}^2, \qquad \frac{1}{2}(b_k^*)^2 - \frac{1}{2}(b_{k+1}^*)^2$$

contribute only $O(h^3)$ to the difference. Hence

$$\begin{aligned} G_k(\mathbf{W}^h) = h\Big((b_k + b_k^*) - (b_{k+1} + b_{k+1}^*)\Big) \\ + h^2\Big(b_k b_k^* - b_{k+1}^* b_{k+1}\Big) + O(h^3). \end{aligned} \tag{8.38}$$

Taylor's theorem gives

$$b_{k+1} = b_k + h\, b'(x_k) + O(h^2), \qquad b^*_{k+1} = b^*_k + h\,(b^*)'(x_k) + O(h^2).$$

Therefore

$$(b_k + b_k^*) - (b_{k+1} + b^*_{k+1}) = -h(b + b^*)'(x_k) + O(h^2),$$

and

$$b_k b_k^* - b^*_{k+1} b_{k+1} = b_k b_k^* - b_k^* b_k + O(h) = [b(x_k), b(x_k)^*] + O(h).$$

Dividing (8.38) by h^2 yields

$$\omega_k^h = -(b + b^*)'(x_k) + [b(x_k), b(x_k)^*] + O(h) = \mu_\infty(b)(x_k) + O(h),$$

which is the desired formal limit.

Finally, by (8.33),

$$\mu_\infty(b) = \nabla_b^*(b + b^*). \tag{8.39}$$

Indeed,

$$\nabla_b^*(b + b^*) = -(b + b^*)' + [b + b^*, b^*] = -(b + b^*)' + [b, b^*].$$

□

Thus the finite-depth balancedness equation $D^*_{\mathbf{W}}\mathbf{W} = 0$ formally becomes

$$\mu_\infty(b) = 0. \tag{8.40}$$

This observation anticipates the gauge-theoretic formulation of Chapter 9.

Remark 8.24 (Restricted Nahm equation)*.* Write

$$b = -T_0 - \mathrm{i}T_1, \qquad T_0, T_1 : I \to \mathfrak{u}(d).$$

Then

$$b + b^* = -2\mathrm{i}T_1, \qquad [b, b^*] = 2\mathrm{i}[T_0, T_1].$$

Hence

$$\mu_\infty(b) = 2\mathrm{i}\Big(T_1' + [T_0, T_1]\Big).$$

The continuum balancedness equation (8.40) is therefore equivalent to

$$T_1' + [T_0, T_1] = 0.$$

This is the one-component reduction of Nahm's equations obtained by setting the remaining two components to zero [Don84].

Remark 8.25*.* The ridge regularizer itself does not produce a nontrivial functional on a fixed continuum fibre at the natural scale. Indeed, under the near-identity scaling $W_k^h = \exp\Big(h\, b(x_k)\Big)$,

$$(W_k^h)^* W_k^h = I + h(b_k + b_k^*) + O(h^2), \tag{8.41}$$

and therefore

$$\begin{aligned} \mathcal{R}(\mathbf{W}^h) &= \frac{1}{2}\sum_{k=1}^{N} \mathrm{tr}\Big((W_k^h)^* W_k^h\Big) \\ &= \frac{Nd}{2} + \frac{h}{2}\sum_{k=1}^{N} \mathrm{tr}(b_k + b_k^*) + O(h). \end{aligned} \tag{8.42}$$

The leading term diverges. The next term has the formal limit

$$\frac{1}{2}\int_0^1 \mathrm{tr}(b + b^*)\, dx = \log|\det X_{\mathrm{end}}|,$$

where X_{end} is the endpoint satisfying $X(1) = X_{\mathrm{end}}$. Thus this term is fixed on the fibre. Consequently, the ridge regularizer carries no nontrivial fibre dependence in the continuum limit considered here.

By contrast, the renormalized energy of the moment has the formal Riemann-sum limit

$$\frac{1}{2}\, h \sum_{k=1}^{N-1} \mathrm{tr}\Big((\omega_k^h)^* \omega_k^h\Big) \longrightarrow \frac{1}{2} \int_0^1 \mathrm{tr}\Big(\mu_\infty(b)(x)^* \mu_\infty(b)(x)\Big)\, dx. \tag{8.43}$$

This is the continuum Kirwan–Ness functional used in Chapter 9.

The formal limit has therefore isolated the continuum structures that survive the near-identity regime. The discrete differential $D_{\mathbf{W}}$ becomes the covariant derivative ∇_b, while the finite moment $D_{\mathbf{W}}^* \mathbf{W}$, after the h^{-2} renormalization in (8.36), becomes the continuum moment field $\mu_\infty(b)$. Chapter 9 turns this formal picture into a gauge theory on the interval: it chooses the Sobolev configuration space, defines the based gauge action, identifies the continuum moment map, and returns to the fibre metrics and heat flows in that analytic setting.

Bibliographical notes

For the general framework of discrete exterior calculus, including cochain complexes, discrete differentials, codifferentials, and Hodge Laplacians, see [Hir03; Des+05].

For normal metrics on adjoint orbits and related isospectral manifolds, see [Ati82; BMR13; BK23]. The fibrewise minimization theorem and the explicit regularizer flow for the DLN are from [LM25].

The near-identity scaling and transport viewpoint are related to the infinite-depth DLN limit studied by Cohen, Menon, and Veraszto [CMV23]. For the Nahm-equation background used in Remark 8.24, see Donaldson [Don84].

Chapter 9

Gauge Theory on the Interval

We construct the infinite-depth model on the interval within a Sobolev framework. A coefficient $b \in H^1([0,1], \mathfrak{gl}(d,\mathbb{C}))$ determines a path X_b by the transport equation $X_b' = bX_b$, $X_b(0) = I$. For a given matrix $X \in \mathrm{GL}(d,\mathbb{C})$, the corresponding fibre is defined by the endpoint condition $X_b(1) = X$.

The symmetry group consists of gauge transformations that are equal to the identity at both endpoints. After setting up the Sobolev spaces, the final section gives a formal derivation of the heat flows on the fibre.

Remark 9.1. Karen Uhlenbeck observed that the infinite-depth DLN should be interpreted through the lens of Yang–Mills theory [Uhl26]. In this interpretation, the continuum model on the interval is viewed as a one-dimensional gauge theory, with the moment field playing the role of curvature and balancedness corresponding to the vanishing of this field.

9.1 Main results

Set $\mathcal{B} := H^1\Big([0,1], \mathfrak{gl}(d,\mathbb{C})\Big)$. For $b \in \mathcal{B}$, let X_b solve the transport equation, and define the endpoint map by $\phi_\infty(b) := X_b(1)$. For $X \in \mathrm{GL}(d,\mathbb{C})$, the continuum fibre over X is $\mathcal{F}_X^\infty := \phi_\infty^{-1}(X)$.

The based complex and unitary gauge groups are

$$\mathcal{G} := \{g \in H^2([0,1], \mathrm{GL}(d,\mathbb{C})) : \; g(0) = g(1) = I\},$$

and

$$\mathcal{K} := \{u \in H^2([0,1], \mathrm{U}(d)) : \; u(0) = u(1) = I\}.$$

The action of $\mathcal{G}$ on $\mathcal{B}$ is

$$b^{(g)} := gbg^{-1} + g'g^{-1}. \tag{9.1}$$

For either $\mathfrak{g} = \mathfrak{gl}(d,\mathbb{C})$ or $\mathfrak{g} = \mathfrak{u}(d)$, write

$$H_0^2([0,1], \mathfrak{g}) := \{a \in H^2([0,1], \mathfrak{g}) : \; a(0) = a(1) = 0\}.$$

The infinitesimal gauge operator at b is $\nabla_b a := a' + [a,b]$. Its formal L^2 adjoint is $\nabla_b^* Z := -Z' + [Z, b^*]$.

Remark 9.2. In Yang–Mills theory on a circle, the symmetry is the full loop group, and the natural gauge-invariant object is the conjugacy class of the holonomy. If one instead fixes a frame at a base point, this symmetry is reduced to the based loop group, and the holonomy element itself is well defined.

The DLN interval model is framed at both endpoints. The conditions $X_b(0) = I$ and $X_b(1) = Y$ fix the input and output frames, so the endpoint matrix is not taken modulo conjugacy. The relevant symmetry group therefore consists of gauge transformations that are the identity at the endpoints. In this sense, the continuum fibre is a one-dimensional gauge model for a factorization with prescribed product.

The first result identifies the continuum fibre with the orbit of the based complex gauge group.

Theorem 9.3. *The endpoint map*

$$\phi_\infty : \mathcal{B} \to \mathrm{GL}(d, \mathbb{C})$$

is a smooth submersion. Hence, for every $X \in \mathrm{GL}(d, \mathbb{C})$, the fibre $\mathcal{F}_X^\infty$ is a smooth Hilbert submanifold of $\mathcal{B}$. Moreover, $\mathcal{F}_X^\infty$ is a single $\mathcal{G}$-orbit under (9.1)*, and for every $b \in \mathcal{F}_X^\infty$,*

$$T_b\mathcal{F}_X^\infty = \nabla_b H_0^2\Big([0,1], \mathfrak{gl}(d, \mathbb{C})\Big).$$

We next record the Hamiltonian structure of the based unitary action. On $\mathcal{B}$ we use the L^2 Hermitian form

$$h_{L^2}(\alpha, \beta) := \int_0^1 \mathrm{tr}(\alpha(x)^*\beta(x))\, dx,$$

with symplectic form

$$\omega(\alpha, \beta) := \Im h_{L^2}(\alpha, \beta).$$

For $R \in L^2([0,1], \mathrm{Herm}(d))$ and $\xi \in H_0^2([0,1], \mathfrak{u}(d))$, define the weak pairing

$$\langle R,\, \xi\rangle := -\frac{1}{2}\int_0^1 \Im\, \mathrm{tr}(R(x)\xi(x))\, dx. \tag{9.2}$$

The continuum moment field is $\mu_\infty(b) := -(b + b^*)' + [b, b^*]$.

Theorem 9.4. *The action of $\mathcal{K}$ on $(\mathcal{B}, \omega)$ is Hamiltonian. With respect to the pairing* (9.2)*, a moment map is*

$$\mu_\infty : \mathcal{B} \to L^2([0,1], \mathrm{Herm}(d)), \qquad \mu_\infty(b) = -(b + b^*)' + [b, b^*].$$

Equivalently, for every $\xi \in H^2_0([0,1],\mathfrak{u}(d))$,

$$\mathrm{d}(\langle \mu_\infty(\cdot),\, \xi\rangle)_b\,[\alpha] = \omega(\nabla_b \xi, \alpha) \qquad \textit{for all } b, \alpha \in \mathcal{B}.$$

The balanced set is the zero level set of this moment map: $\mathcal{M}_\infty := \{b \in \mathcal{B} : \mu_\infty(b) = 0\}$. The following regular-value statement gives this set its Hilbert-manifold structure.

Proposition 9.5. *The map*

$$\mu_\infty : \mathcal{B} = H^1([0,1], \mathfrak{gl}(d,\mathbb{C})) \longrightarrow L^2([0,1], \mathrm{Herm}(d))$$

is smooth as a map of real Hilbert spaces. Moreover, for every $b \in \mathcal{B}$*, the differential* $(D\mu_\infty)_b$ *is surjective. Hence* 0 *is a regular value of* μ_∞*, and*

$$\mathcal{M}_\infty = \mu_\infty^{-1}(0)$$

is a smooth Hilbert submanifold of $\mathcal{B}$*, with*

$$T_b\mathcal{M}_\infty = \ker(D\mu_\infty)_b.$$

For fixed $X \in \mathrm{GL}(d,\mathbb{C})$, the balanced part of the continuum fibre is $\mathcal{O}^\infty_X := \mathcal{F}^\infty_X \cap \mathcal{M}_\infty$.

The next result rewrites balancedness in the positive-definite cone. For $b \in \mathcal{B}$, set $H_b(x) := X_b(x)^* X_b(x) \in \mathrm{Herm}^+(d)$. The trace metric on $\mathrm{Herm}^+(d)$ is

$$g_{\mathrm{tr},H}(U,V) := \mathrm{tr}(H^{-1}UH^{-1}V), \qquad U, V \in T_H\mathrm{Herm}^+(d) \simeq \mathrm{Herm}(d). \tag{9.3}$$

Its geodesic equation is $(H^{-1}H')' = 0$.

Theorem 9.6. *For every $b \in \mathcal{B}$,*

$$(H_b^{-1}H_b')' = -\,X_b^{-1}\mu_\infty(b)X_b. \tag{9.4}$$

Consequently, $b \in \mathcal{M}_\infty$ if and only if $(H_b^{-1}H_b')' = 0$.

On a fixed fibre, the endpoints of H_b are fixed:

$$H_b(0) = I, \qquad H_b(1) = X^*X.$$

The geodesic equation therefore has a unique solution with these endpoints.

Corollary 9.7. *Fix $X \in \mathrm{GL}(d,\mathbb{C})$ and let $b \in \mathcal{F}_X^\infty$. Then*

$$b \in \mathcal{O}_X^\infty \quad \Longleftrightarrow \quad H_b(x) = \exp\Big(x\log(X^*X)\Big), \qquad 0 \le x \le 1.$$

An H^2 path $Y : [0,1] \to \mathrm{GL}(d,\mathbb{C})$ with $Y(0) = I$ determines the coefficient

$$b_Y := Y'Y^{-1} \in \mathcal{B}.$$

We call Y balanced if $b_Y \in \mathcal{M}_\infty$.

Corollary 9.8. *Fix $X \in \mathrm{GL}(d,\mathbb{C})$, and write its polar decomposition as*

$$X = U_X(X^*X)^{1/2}.$$

An H^2 path $Y : [0,1] \to \mathrm{GL}(d,\mathbb{C})$ with

$$Y(0) = I, \qquad Y(1) = X$$

is balanced if and only if

$$Y(x) = U(x)\exp\Big(\frac{x}{2}\log(X^*X)\Big),$$

where

$$U \in H^2([0,1],\mathrm{U}(d)), \qquad U(0) = I, \qquad U(1) = U_X.$$

Corollary 9.8 shows that two balanced paths with the same endpoint differ by left multiplication by a based unitary gauge transformation.

Corollary 9.9. *For every $X \in \mathrm{GL}(d,\mathbb{C})$, the balanced fibre $\mathcal{O}_X^\infty$ is a single $\mathcal{K}$-orbit.*

It remains to connect this geometry to the fibrewise heat flows suggested by the discrete theory. Theorems 9.11 and 9.12 identify the formal interior equations.

The continuum Kirwan–Ness energy is

$$\mathcal{E}_\infty(b) := \frac{1}{2}\int_0^1 \mathrm{tr}\Big(\mu_\infty(b)(x)^*\mu_\infty(b)(x)\Big)\,dx. \tag{9.5}$$

For tangent vectors written as $\nabla_b a$ and $\nabla_b c$, define the induced fibre metric by

$$g_b^\iota(\nabla_b a, \nabla_b c) := \int_0^1 \Re\,\mathrm{tr}\Big((\nabla_b a)^*\nabla_b c\Big)\,dx, \tag{9.6}$$

and the normal fibre metric by

$$g_b^{\mathrm{nor}}(\nabla_b a, \nabla_b c) := \int_0^1 \Re\,\mathrm{tr}(a^*c)\,dx. \tag{9.7}$$

These definitions do not depend on the chosen based potentials. Indeed, if $\nabla_b a = 0$ and $a(0) = 0$, then

$$(X_b^{-1}aX_b)' = X_b^{-1}(\nabla_b a)X_b = 0.$$

Thus $X_b^{-1} a X_b$ is constant, and since $a(0) = 0$, we have $a \equiv 0$. Hence $a \mapsto \nabla_b a$ is injective on based fields.

For the formal Hodge notation, set

$$\Delta_b^{(0)} := \nabla_b^* \nabla_b. \tag{9.8}$$

We also write

$$\Delta_b^{(1)} b := \nabla_b \mu_\infty(b), \qquad (\Delta_b^{(1)})^2 b := \nabla_b \Delta_b^{(0)} \mu_\infty(b). \tag{9.9}$$

Let Y be the normalized transport path for b, so that

$$Y_x = bY, \qquad Y(0) = I,$$

and set

$$H := Y^* Y.$$

Define

$$P := H^{-1} H_x, \qquad Q := P_x, \qquad \tau_H := H_{xx} - H_x H^{-1} H_x, \qquad \Gamma := -Q_{xx} + [Q_x, P]. \tag{9.10}$$

Here $\tau_H = H(H^{-1} H_x)_x$ is the geodesic residual for the trace metric.

Proposition 9.10. *For smooth b,*

$$\mathcal{E}_\infty(b) = \frac{1}{2} \int_0^1 \operatorname{tr}\Big(H^{-1} \tau_H \, H^{-1} \tau_H \Big) \, dx. \tag{9.11}$$

Thus the continuum Kirwan–Ness energy is the squared norm of the geodesic residual of $H = Y^* Y$ for the trace metric.

Theorem 9.11. *For smooth b, the formal negative gradient of $\mathcal{E}_\infty$ with respect to g^ι*

has the interior form

$$b_t = -2\Delta_b^{(1)} b = -2\nabla_b \mu_\infty(b). \tag{9.12}$$

In the variable $H = Y^*Y$, the corresponding interior equation is

$$H_t = 4\Big(H_{xx} - H_x H^{-1} H_x\Big) = 4\tau_H. \tag{9.13}$$

Theorem 9.12. *For smooth b, the formal negative gradient of $\mathcal{E}_\infty$ with respect to g^{nor} has the interior form*

$$b_t = -2(\Delta_b^{(1)})^2 b = -2\nabla_b \Delta_b^{(0)} \mu_\infty(b). \tag{9.14}$$

In the variable $H = Y^*Y$, the corresponding interior equation is

$$H_t = 2\Big(\Gamma^* H + H\Gamma\Big). \tag{9.15}$$

The corresponding problem on a fixed fibre is obtained by coupling these interior equations to the endpoint conditions

$$Y(0,t) = I, \qquad Y(1,t) = X, \tag{9.16}$$

or, after passing to $H = Y^*Y$, to

$$H(0,t) = I, \qquad H(1,t) = X^*X.$$

9.2 The interval model and its gauge symmetry

The continuum fibre is an interval problem with fixed endpoint frames. The coefficient b records infinitesimal transport along the variable x, while the endpoint value $X_b(1)$ records the analogue of the finite product $W_N \cdots W_1$. Since the frames at $x = 0$ and $x = 1$ are part of the data, a gauge transformation may change the frame only in the interior of the interval.

Recall from Section 9.1 that $\mathcal{B} = H^1\Big([0,1], \mathfrak{gl}(d,\mathbb{C})\Big)$, and that the based complex and unitary gauge groups are

$$\mathcal{G} = \{g \in H^2([0,1], \mathrm{GL}(d,\mathbb{C})) : \ g(0) = g(1) = I\},$$

$$\mathcal{K} = \{u \in H^2([0,1], \mathrm{U}(d)) : \ u(0) = u(1) = I\}.$$

A coefficient $b \in \mathcal{B}$ defines the covariant derivative

$$D_x = \frac{d}{dx} - b(x),$$

and the transport equation is the parallel-transport equation for D_x. With this sign convention, a change of frame by g sends b to

$$b^{(g)} = gbg^{-1} + g'g^{-1},$$

as in (9.1). This is the usual one-dimensional connection law [KN63; FU91].

The Sobolev orders are chosen to make this formula closed on $\mathcal{B}$. In one dimension,

$$H^1([0,1]) \hookrightarrow C^0([0,1]), \qquad H^2([0,1]) \hookrightarrow C^1([0,1])$$

continuously [AF03]. Thus coefficients in $\mathcal{B}$ are continuous, while gauge transformations in $\mathcal{G}$ are continuously differentiable. Moreover, $H^1([0,1])$ is a Banach algebra, so pointwise matrix products of H^1 functions again belong to H^1.

Lemma 9.13. *The formula*

$$g \cdot b := b^{(g)} = gbg^{-1} + g'g^{-1}$$

defines a smooth left action of $\mathcal{G}$ on $\mathcal{B}$.

Proof. Let $g \in \mathcal{G}$. Since $g \in H^2([0,1], \mathrm{GL}(d,\mathbb{C}))$ and inversion is smooth on $\mathrm{GL}(d,\mathbb{C})$, the inverse g^{-1} also belongs to $H^2([0,1], \mathrm{GL}(d,\mathbb{C}))$. Hence

$$g,\ g^{-1},\ g' \in H^1([0,1]) \cap C^0([0,1]).$$

Since $H^1([0,1])$ is a Banach algebra, both products

$$gbg^{-1} \qquad \text{and} \qquad g'g^{-1}$$

belong to $H^1([0,1], \mathfrak{gl}(d,\mathbb{C}))$. Therefore $b^{(g)} \in \mathcal{B}$.

The identity element acts trivially. If $g, h \in \mathcal{G}$, then

$$h \cdot (g \cdot b) = h\Big(gbg^{-1} + g'g^{-1}\Big)h^{-1} + h'h^{-1} = (hg)b(hg)^{-1} + (hg)'(hg)^{-1}.$$

Thus

$$h \cdot (g \cdot b) = (hg) \cdot b,$$

so the action is a left action. Smoothness follows from the continuity of multiplication in the Sobolev algebras used above and from smoothness of inversion on the open set of invertible H^2 paths. $\square$

Remark 9.14. As explained at the beginning of the chapter, the endpoint framing is part of the continuum fibre. Thus $\mathcal{G}$ and $\mathcal{K}$ are based path groups.

9.2.1 The endpoint map

For $b \in \mathcal{B}$, let X_b denote the solution of the transport equation, and recall that $\phi_\infty(b) = X_b(1)$. The next lemma supplies the analytic input for Theorem 9.3.

Lemma 9.15. *For every $b \in \mathcal{B}$, the transport equation*

$$X_b' = bX_b, \qquad X_b(0) = I,$$

has a unique solution

$$X_b \in H^2([0,1], \mathrm{GL}(d,\mathbb{C})).$$

Moreover, the solution map

$$b \longmapsto X_b$$

is smooth as a map from $\mathcal{B}$ to $H^2([0,1], \mathrm{M}_d(\mathbb{C}))$. Consequently

$$\phi_\infty : \mathcal{B} \to \mathrm{GL}(d,\mathbb{C})$$

is smooth, and

$$(D\phi_\infty)_b[\alpha] = X_b(1) \int_0^1 X_b(t)^{-1}\alpha(t)X_b(t)\,dt. \tag{9.17}$$

Proof. Since $b \in H^1([0,1]) \subset C^0([0,1])$, the ordinary differential equation has a unique C^1 solution on $[0,1]$. Differentiating once more in the weak sense gives

$$X_b'' = b'X_b + bX_b'.$$

Here $b' \in L^2$, $X_b \in C^0$, $b \in C^0$, and $X_b' = bX_b \in L^2$. Therefore $X_b'' \in L^2$, so $X_b \in H^2$.

The solution remains invertible. Indeed, if Y solves

$$Y' = -Yb, \qquad Y(0) = I,$$

then $(YX_b)' = 0$, and hence $Y(x)X_b(x) = I$ for all x. Thus $X_b(x) \in \mathrm{GL}(d, \mathbb{C})$ for every x, and

$$X_b \in H^2([0,1], \mathrm{GL}(d, \mathbb{C})).$$

We next prove smooth dependence on b. Consider

$$\mathcal{F}(b, X) := \Big(X' - bX,\ X(0)\Big),$$

as a map

$$\mathcal{F} : \mathcal{B} \times H^2([0,1], \mathrm{M}_d(\mathbb{C})) \longrightarrow H^1([0,1], \mathrm{M}_d(\mathbb{C})) \times \mathrm{M}_d(\mathbb{C}).$$

This map is smooth because multiplication

$$H^1 \times H^2 \longrightarrow H^1$$

is continuous in one dimension. At a solution X_b, the derivative with respect to X is

$$\eta \longmapsto \Big(\eta' - b\eta,\ \eta(0)\Big).$$

This operator is an isomorphism: for every

$$(f, C) \in H^1([0,1], \mathrm{M}_d(\mathbb{C})) \times \mathrm{M}_d(\mathbb{C}),$$

the linear problem

$$\eta' = b\eta + f, \qquad \eta(0) = C,$$

has a unique solution $\eta \in H^2([0,1], \mathrm{M}_d(\mathbb{C}))$. The inverse is bounded by the usual linear ODE estimates on the compact interval. The implicit function theorem therefore gives smoothness of $b \mapsto X_b$.

Now fix $b, \alpha \in \mathcal{B}$, and let

$$Y = (DX)_b[\alpha].$$

Differentiating the transport equation in the direction α gives

$$Y' = bY + \alpha X_b, \qquad Y(0) = 0. \tag{9.18}$$

Write $Y = X_b Z$. Since $X_b' = bX_b$, equation (9.18) gives

$$Z' = X_b^{-1}\alpha X_b, \qquad Z(0) = 0.$$

Therefore

$$Z(1) = \int_0^1 X_b(t)^{-1}\alpha(t)X_b(t)\,dt,$$

and hence

$$Y(1) = X_b(1)\int_0^1 X_b(t)^{-1}\alpha(t)X_b(t)\,dt.$$

Since $(D\phi_\infty)_b[\alpha] = Y(1)$, this proves (9.17). □

Proof of Theorem 9.3. By Lemma 9.15, the endpoint map

$$\phi_\infty : \mathcal{B} \to \mathrm{GL}(d, \mathbb{C})$$

is smooth. We first prove that it is a submersion. Fix $b \in \mathcal{B}$ and let

$$M \in T_{\phi_\infty(b)}\mathrm{GL}(d,\mathbb{C}) \simeq \mathrm{M}_d(\mathbb{C})$$

be arbitrary. Set

$$R := X_b(1)^{-1}M, \qquad \alpha(t) := X_b(t)RX_b(t)^{-1}.$$

Since $X_b, X_b^{-1} \in H^2$, we have $\alpha \in H^1 = \mathcal{B}$. Moreover,

$$X_b(t)^{-1}\alpha(t)X_b(t) = R$$

for every t. Using (9.17), we obtain

$$(D\phi_\infty)_b[\alpha] = X_b(1)\int_0^1 R\,dt = X_b(1)R = M.$$

Thus $(D\phi_\infty)_b$ is surjective for every $b \in \mathcal{B}$.

Since the target is finite-dimensional, the kernel of $(D\phi_\infty)_b$ is complemented. The Banach-space implicit function theorem therefore implies that, for every $X \in \mathrm{GL}(d,\mathbb{C})$,

$$\mathcal{F}_X^\infty = \phi_\infty^{-1}(X)$$

is a smooth Hilbert submanifold of $\mathcal{B}$.

We next prove that each fibre is a single $\mathcal{G}$-orbit. Let $g \in \mathcal{G}$. By Lemma 9.13, $b^{(g)} \in \mathcal{B}$. Differentiate the product gX_b:

$$(gX_b)' = g'X_b + gX_b' = g'X_b + gbX_b = \Big(gbg^{-1} + g'g^{-1}\Big)(gX_b).$$

Thus gX_b solves the transport equation with coefficient $b^{(g)}$. Since

$$(gX_b)(0) = g(0)X_b(0) = I,$$

uniqueness gives

$$X_{b^{(g)}}(x) = g(x)X_b(x), \qquad 0 \le x \le 1. \tag{9.19}$$

Evaluating at $x = 1$ and using $g(1) = I$, we get

$$\phi_\infty(b^{(g)}) = X_{b^{(g)}}(1) = g(1)X_b(1) = X_b(1) = \phi_\infty(b).$$

Therefore the $\mathcal{G}$-orbit of b lies in the fibre through b.

Conversely, let $b, c \in \mathcal{F}_X^\infty$. Define

$$g(x) := X_c(x)X_b(x)^{-1}.$$

Since $X_b, X_c \in H^2([0,1], \mathrm{GL}(d,\mathbb{C}))$, we have

$$g \in H^2([0,1], \mathrm{GL}(d,\mathbb{C})).$$

The endpoint conditions are

$$g(0) = X_c(0)X_b(0)^{-1} = I, \qquad g(1) = X_c(1)X_b(1)^{-1} = XX^{-1} = I.$$

Thus $g \in \mathcal{G}$. Differentiating $g = X_cX_b^{-1}$ gives

$$g' = cX_cX_b^{-1} - X_cX_b^{-1}b = cg - gb.$$

Rearranging,

$$c = gbg^{-1} + g'g^{-1} = b^{(g)}.$$

Hence any two points in $\mathcal{F}_X^\infty$ lie in the same $\mathcal{G}$-orbit.

It remains to identify the tangent space. Since $\mathcal{F}_X^\infty$ is the level set of the submersion ϕ_∞,

$$T_b\mathcal{F}_X^\infty = \ker(D\phi_\infty)_b.$$

First let

$$a \in H_0^2\Big([0,1], \mathfrak{gl}(d,\mathbb{C})\Big).$$

The path

$$g_s = \exp(sa)$$

lies in $\mathcal{G}$ for all real s, because $a(0) = a(1) = 0$. Differentiating the gauge action at $s = 0$ gives

$$\frac{d}{ds}\bigg|_{s=0} b^{(g_s)} = a' + [a,b] = \nabla_b a.$$

The endpoint map is invariant under $\mathcal{G}$, so

$$(D\phi_\infty)_b[\nabla_b a] = 0.$$

Therefore

$$\nabla_b H_0^2\Big([0,1], \mathfrak{gl}(d,\mathbb{C})\Big) \subseteq T_b\mathcal{F}_X^\infty.$$

For the reverse inclusion, let

$$\alpha \in T_b\mathcal{F}_X^\infty = \ker(D\phi_\infty)_b.$$

Define

$$m(x) := \int_0^x X_b(t)^{-1}\alpha(t)X_b(t)\,dt, \qquad a(x) := X_b(x)m(x)X_b(x)^{-1}.$$

The integrand $X_b^{-1}\alpha X_b$ lies in H^1, so $m \in H^2$ and $m(0) = 0$. Hence $a \in H^2$ and $a(0) = 0$.

Using (9.17) and $\alpha \in \ker(D\phi_\infty)_b$, we find

$$0 = (D\phi_\infty)_b[\alpha] = X_b(1)m(1).$$

Since $X_b(1)$ is invertible, $m(1) = 0$, and therefore $a(1) = 0$. Thus

$$a \in H_0^2\Big([0,1], \mathfrak{gl}(d,\mathbb{C})\Big).$$

Finally,

$$a' = bX_bmX_b^{-1} + X_bm'X_b^{-1} - X_bmX_b^{-1}b.$$

Since $m' = X_b^{-1}\alpha X_b$, this becomes

$$a' = ba + \alpha - ab.$$

Consequently,

$$\nabla_b a = a' + [a,b] = \alpha.$$

Thus

$$T_b\mathcal{F}_X^\infty \subseteq \nabla_b H_0^2\Big([0,1], \mathfrak{gl}(d,\mathbb{C})\Big),$$

and the reverse inclusion proved above gives equality. $\square$

9.3 The continuum moment map and balancedness

In Chapter 4, we identified balancedness with the zero level set of the moment map for the finite-dimensional unitary symmetry. We now prove the corresponding statement on the interval.

The finite-depth moment tuple is a codifferential. In the interval model, the same role is played by applying the formal adjoint $\nabla_b^* Z = -Z' + [Z, b^*]$ to the Hermitian part of the coefficient. Indeed,

$$\nabla_b^*(b + b^*) = -(b + b^*)' + [b + b^*, b^*] = -(b + b^*)' + [b, b^*] = \mu_\infty(b). \tag{9.20}$$

Thus the field $\mu_\infty(b)$ is the continuum analogue of the discrete codifferential of the weight tuple.

We use the weak L^2 Hermitian form

$$h_{L^2}(\alpha, \beta) := \int_0^1 \operatorname{tr}(\alpha(x)^* \beta(x))\, dx$$

on $\mathcal{B}$, and write

$$g_{L^2}(\alpha, \beta) := \Re h_{L^2}(\alpha, \beta), \qquad \omega(\alpha, \beta) := \Im h_{L^2}(\alpha, \beta), \qquad J\alpha := \mathrm{i}\alpha.$$

The form ω is constant on the affine space $\mathcal{B}$, hence closed. It is weakly nondegenerate because $h_{L^2}(\alpha, \alpha) > 0$ for every nonzero $\alpha \in \mathcal{B}$.

Lemma 9.16. *The action of $\mathcal{K}$ on $\mathcal{B}$ preserves g_{L^2}, J, and ω.*

Proof. Fix $u \in \mathcal{K}$. The action is

$$b \longmapsto b^{(u)} := ubu^{-1} + u'u^{-1}.$$

Its differential at any $b \in \mathcal{B}$ is

$$\alpha \longmapsto u\alpha u^{-1}.$$

Since $u(x)$ is unitary for every $x \in [0,1]$, pointwise conjugation preserves the Frobenius Hermitian form:

$$\operatorname{tr}\Big((u\alpha u^{-1})^*(u\beta u^{-1})\Big) = \operatorname{tr}(\alpha^*\beta).$$

After integration over $[0,1]$, this gives

$$h_{L^2}(u\alpha u^{-1}, u\beta u^{-1}) = h_{L^2}(\alpha, \beta).$$

Therefore the action preserves g_{L^2} and ω. It also preserves J, since complex scalar multiplication commutes with conjugation. $\square$

Lemma 9.17. *The formula for μ_∞ defines a smooth map*

$$\mu_\infty : \mathcal{B} \longrightarrow L^2([0,1], \mathrm{Herm}(d)).$$

Its differential at $b \in \mathcal{B}$ is

$$(D\mu_\infty)_b[\alpha] = -(\alpha + \alpha^*)' + [\alpha, b^*] + [b, \alpha^*]. \tag{9.21}$$

Proof. If $b \in \mathcal{B} = H^1([0,1], \mathfrak{gl}(d, \mathbb{C}))$, then

$$b + b^* \in H^1, \qquad (b + b^*)' \in L^2.$$

Since $H^1([0,1])$ is a Banach algebra, the commutator

$$[b, b^*] = bb^* - b^*b$$

belongs to $H^1([0,1],\mathfrak{gl}(d,\mathbb{C}))$, hence to L^2. Thus $\mu_\infty(b) \in L^2$.

The field is Hermitian. Indeed, $b+b^*$ is Hermitian, so its weak derivative is Hermitian as an L^2 field, and

$$[b,b^*]^* = (bb^* - b^*b)^* = bb^* - b^*b = [b,b^*].$$

Hence $\mu_\infty(b)$ takes values in $\mathrm{Herm}(d)$.

Smoothness follows from the smoothness of the linear map $b \mapsto -(b+b^*)'$ from H^1 to L^2 and the smoothness of the quadratic map $b \mapsto [b,b^*]$ from H^1 to $H^1 \subset L^2$. Differentiating the formula for μ_∞ in the direction α gives (9.21). $\square$

Proof of Theorem 9.4. By Lemma 9.16, the based unitary gauge group acts on $(\mathcal{B},\omega)$ by symplectomorphisms. By Lemma 9.17, the field μ_∞ is a smooth map from $\mathcal{B}$ to $L^2([0,1],\mathrm{Herm}(d))$. It remains to verify the moment-map identity.

Let

$$\xi \in H^2_0([0,1],\mathfrak{u}(d)).$$

The corresponding fundamental vector field is obtained by differentiating $b^{(\exp(t\xi))}$ at $t=0$:

$$K_\xi(b) = \xi' + [\xi,b] = \nabla_b\xi.$$

Fix $b,\alpha \in \mathcal{B}$. Using (9.21) and the pairing (9.2), we get

$$\mathrm{d}(\langle\mu_\infty(\cdot),\,\xi\rangle)_b\,[\alpha] = -\frac{1}{2}\int_0^1 \Im\,\mathrm{tr}\Big(\big(-(\alpha+\alpha^*)' + [\alpha,b^*] + [b,\alpha^*]\big)\xi\Big)\,dx.$$

We compute the three contributions. For the derivative term, integration by parts and the endpoint condition $\xi(0) = \xi(1) = 0$ give

$$-\frac{1}{2}\int_0^1 \Im \operatorname{tr}\Big(-(\alpha+\alpha^*)'\xi\Big)\,dx = -\int_0^1 \Im \operatorname{tr}(\alpha\xi')\,dx.$$

Here we used $\xi'^* = -\xi'$.

For the first commutator term, cyclicity of trace gives

$$\operatorname{tr}([\alpha, b^*]\xi) = -\operatorname{tr}(\alpha[\xi, b^*]).$$

Hence

$$-\frac{1}{2}\int_0^1 \Im \operatorname{tr}([\alpha, b^*]\xi)\,dx = \frac{1}{2}\int_0^1 \Im \operatorname{tr}(\alpha[\xi, b^*])\,dx.$$

For the second commutator term,

$$\operatorname{tr}([b, \alpha^*]\xi) = \operatorname{tr}(\alpha^*[\xi, b]).$$

Since $[\xi, b]^* = [\xi, b^*]$, we have

$$\Im \operatorname{tr}(\alpha^*[\xi, b]) = -\Im \operatorname{tr}(\alpha[\xi, b^*]).$$

Therefore

$$-\frac{1}{2}\int_0^1 \Im \operatorname{tr}([b, \alpha^*]\xi)\,dx = \frac{1}{2}\int_0^1 \Im \operatorname{tr}(\alpha[\xi, b^*])\,dx.$$

Adding the three terms yields

$$\mathrm{d}(\langle \mu_\infty(\cdot),\, \xi\rangle)_b\,[\alpha] = \int_0^1 \Im \operatorname{tr}\Big(\alpha(-\xi' + [\xi, b^*])\Big)\,dx.$$

On the other hand,

$$(\nabla_b \xi)^* = (\xi' + [\xi, b])^* = -\xi' + [\xi, b^*],$$

because $\xi^* = -\xi$. Thus

$$\mathrm{d}(\langle \mu_\infty(\cdot),\, \xi \rangle)_b\,[\alpha] = \int_0^1 \Im\,\mathrm{tr}\Big((\nabla_b \xi)^* \alpha\Big)\,dx = \omega(\nabla_b \xi, \alpha).$$

This is the desired moment-map identity. $\square$

Proof of Proposition 9.5. Smoothness follows from Lemma 9.17; equivalently, it follows from the fact that in one dimension H^1 is a Banach algebra and $H^1 \hookrightarrow C^0$. Differentiating the moment field gives

$$(D\mu_\infty)_b[\alpha] = -(\alpha + \alpha^*)' + [\alpha, b^*] + [b, \alpha^*].$$

We prove surjectivity. Let

$$R \in L^2([0,1], \mathrm{Herm}(d))$$

be arbitrary. Write

$$a := \frac{1}{2}(b - b^*) \in H^1([0,1], \mathfrak{u}(d)).$$

We look for a Hermitian variation $\alpha = s = s^*$. For such s,

$$(D\mu_\infty)_b[s] = -2s' + [s, b^*] + [b, s] = -2s' - 2[s, a].$$

Thus it is enough to solve

$$s' + [s, a] = -\frac{1}{2}R, \qquad s(0) = 0.$$

This is a first-order linear equation with L^2 forcing and continuous coefficients. It has a unique H^1 solution. Since the coefficient $s \mapsto [s, a]$ preserves Hermitian matrices, and since the forcing and initial condition are Hermitian, the solution remains Hermitian. Therefore

$$(D\mu_\infty)_b[s] = R.$$

So $(D\mu_\infty)_b$ is surjective.

Because $(D\mu_\infty)_b$ is a surjective bounded linear map between Hilbert spaces, its kernel is complemented. The Hilbert-space regular-value theorem therefore implies that $\mathcal{M}_\infty = \mu_\infty^{-1}(0)$ is a smooth Hilbert submanifold and that

$$T_b\mathcal{M}_\infty = \ker(D\mu_\infty)_b.$$

□

The definitions of $\mathcal{M}_\infty$ and $\mathcal{O}_X^\infty$ are therefore the direct continuum analogues of the finite-dimensional definitions. By (9.20), balancedness is the covariant codifferential equation

$$\nabla_b^*(b + b^*) = 0.$$

This is the interval version of the discrete identity

$$\mathbf{G}(\mathbf{W}) = D_{\mathbf{W}}^*\mathbf{W}.$$

Remark 9.18 (Yang–Mills form, [Uhl26]). Assume for this remark that b is smooth, and write

$$b = u - h, \qquad u : [0,1] \to \mathfrak{u}(d), \qquad h : [0,1] \to \mathrm{Herm}(d).$$

Equivalently,

$$u = \frac{1}{2}(b - b^*), \qquad h = -\frac{1}{2}(b + b^*).$$

Then

$$b + b^* = -2h, \qquad [b, b^*] = -2[u, h],$$

and hence

$$\mu_\infty(b) = 2\Big(h' - [u, h]\Big).$$

Now place the pair (u, h) on the cylinder $[0, 1] \times S^1$ by taking the connection one-form

$$A = -u\, dx + \mathrm{i}h\, dy.$$

It is independent of the circle variable y. Its mixed curvature component is

$$F_{xy} = \partial_x(\mathrm{i}h) + [-u, \mathrm{i}h] = \mathrm{i}\Big(h' - [u, h]\Big).$$

Thus

$$\mu_\infty(b) = -2\mathrm{i}F_{xy}.$$

Consequently, the continuum Kirwan–Ness energy (9.5) is, up to the constant factor fixed by these conventions, the Yang–Mills energy of this one-dimensional reduction. In this form, the balanced equation says that the mixed curvature component vanishes.

9.4 The path in the positive-definite cone

In Chapter 6, we used the Gram matrices of the partial products to reduce a fixed finite-depth fibre to a discrete path in the positive-definite cone. In the continuum limit, this discrete path becomes the actual path $H_b(x) = X_b(x)^* X_b(x)$, where X_b solves the transport equation. Thus H_b is the positive part of the normalized transport.

For $b \in \mathcal{B}$, the initial point of this path is fixed:

$$H_b(0) = I.$$

If $b \in \mathcal{F}_X^\infty$, then $\phi_\infty(b) = X$, and hence

$$H_b(1) = X_b(1)^* X_b(1) = X^* X.$$

Therefore a point of the continuum fibre determines a path in $\mathrm{Herm}^+(d)$ from I to $X^* X$. The main point of this section is that balancedness is exactly the condition that this path be the geodesic between these two endpoints for the trace metric.

For $b \in \mathcal{B}$, define the logarithmic velocity of H_b by

$$P_b := H_b^{-1} H_b'.$$

All identities below are equalities in the weak sense for $b \in \mathcal{B}$. For smooth fields, the same calculations are pointwise.

Proposition 9.19. *For every $b \in \mathcal{B}$,*

$$P_b = X_b^{-1}(b + b^*) X_b. \tag{9.22}$$

Proof of Proposition 9.19. Differentiate $H_b = X_b^* X_b$. Using $X_b' = b X_b$, we obtain

$$H_b' = X_b'^* X_b + X_b^* X_b' = X_b^* b^* X_b + X_b^* b X_b = X_b^*(b^* + b) X_b.$$

Since

$$H_b^{-1} = X_b^{-1}(X_b^*)^{-1},$$

it follows that

$$P_b = H_b^{-1} H_b' = X_b^{-1}(b + b^*) X_b.$$

$\square$

Proof of Theorem 9.6. By Proposition 9.19,

$$H_b^{-1} H_b' = X_b^{-1}(b + b^*) X_b.$$

Differentiate this identity:

$$(H_b^{-1} H_b')' = (X_b^{-1})'(b + b^*) X_b + X_b^{-1}(b + b^*)' X_b + X_b^{-1}(b + b^*) X_b'.$$

Since

$$(X_b^{-1})' = -X_b^{-1} b, \qquad X_b' = b X_b,$$

we obtain

$$\begin{aligned}
(H_b^{-1} H_b')' &= -X_b^{-1} b(b + b^*) X_b + X_b^{-1}(b + b^*)' X_b \\
&\quad + X_b^{-1}(b + b^*) b X_b \\
&= X_b^{-1}\Big((b + b^*)' + [b^*, b]\Big) X_b \\
&= -X_b^{-1}\Big(-(b + b^*)' + [b, b^*]\Big) X_b \\
&= -X_b^{-1} \mu_\infty(b) X_b,
\end{aligned}$$

where the last equality uses the definition of μ_∞. This proves (9.4).

Since $X_b(x)$ is invertible for every x, the identity $(H_b^{-1} H_b')' = 0$ is equivalent to $\mu_\infty(b) = 0$. By the definition of $\mathcal{M}_\infty$, this is exactly the condition $b \in \mathcal{M}_\infty$. $\square$

The equation $(H^{-1} H')' = 0$ is the geodesic equation for the trace metric (9.3).

Thus the moment equation $\mu_\infty(b) = 0$ becomes, after passing to the positive part of the transport, the geodesic equation in $\mathrm{Herm}^+(d)$.

Proof of Corollary 9.7. Let $b \in \mathcal{F}_X^\infty$. Since $\mathcal{O}_X^\infty = \mathcal{F}_X^\infty \cap \mathcal{M}_\infty$, Theorem 9.6 gives $b \in \mathcal{O}_X^\infty$ if and only if $(H_b^{-1}H_b')' = 0$.

Assume first that $(H_b^{-1}H_b')' = 0$. Set

$$P := H_b^{-1}H_b'.$$

Then P is independent of x. Since H_b and H_b' are Hermitian,

$$H_bP = H_b' = (H_b')^* = P^*H_b.$$

Evaluating at $x = 0$, where $H_b(0) = I$, gives $P = P^*$. Thus P is Hermitian.

The equation $H_b' = H_bP$, together with $H_b(0) = I$, has the solution

$$H_b(x) = \exp(xP).$$

Because $b \in \mathcal{F}_X^\infty$,

$$H_b(1) = X^*X.$$

The matrix X^*X is positive definite, so it has a unique Hermitian logarithm. Hence

$$P = \log(X^*X),$$

and therefore

$$H_b(x) = \exp\Big(x\log(X^*X)\Big).$$

Conversely, if

$$H_b(x) = \exp\Big(x \log(X^*X)\Big),$$

then

$$H_b^{-1} H_b' = \log(X^*X),$$

so $(H_b^{-1} H_b')' = 0$. The theorem then gives $b \in \mathcal{O}_X^\infty$. $\square$

For an H^2 path $Y : [0,1] \to \mathrm{GL}(d, \mathbb{C})$ with $Y(0) = I$, write

$$b_Y := Y' Y^{-1} \in \mathcal{B}.$$

Then Y is balanced precisely when $b_Y \in \mathcal{M}_\infty$.

Proof of Corollary 9.8. Let

$$L_X := \log(X^*X).$$

Assume first that Y is balanced, and set

$$b_Y := Y' Y^{-1}.$$

Since $Y(0) = I$ and $Y(1) = X$, we have $b_Y \in \mathcal{F}_X^\infty$. Since Y is balanced, $b_Y \in \mathcal{O}_X^\infty$. By Corollary 9.7,

$$Y(x)^* Y(x) = \exp(x L_X).$$

Define

$$U(x) := Y(x) \exp\left(-\frac{x}{2} L_X\right).$$

Then

$$U(x)^* U(x) = \exp\left(-\frac{x}{2} L_X\right) Y(x)^* Y(x) \exp\left(-\frac{x}{2} L_X\right) = I,$$

so $U(x) \in \mathrm{U}(d)$ for every x. Since $Y \in H^2$ and $x \mapsto \exp(-xL_X/2)$ is smooth, we have $U \in H^2([0,1], \mathrm{U}(d))$. The endpoints are

$$U(0) = I, \qquad U(1) = X(X^*X)^{-1/2} = U_X,$$

where

$$X = U_X(X^*X)^{1/2}$$

is the polar decomposition of X. Thus Y has the asserted form.

Conversely, suppose Y has the asserted form:

$$Y(x) = U(x) \exp\left(\frac{x}{2} L_X\right), \qquad U \in H^2([0,1], \mathrm{U}(d)).$$

Then

$$Y(x)^* Y(x) = \exp(xL_X).$$

By Corollary 9.7, the coefficient $b_Y = Y'Y^{-1}$ belongs to $\mathcal{O}_X^\infty$, and hence Y is balanced.

□

Proof of Corollary 9.9. First let $b \in \mathcal{O}_X^\infty$ and $u \in \mathcal{K}$. By the gauge covariance of transport,

$$X_{b^{(u)}} = uX_b.$$

Since $u(x)$ is unitary,

$$H_{b^{(u)}} = X_{b^{(u)}}^* X_{b^{(u)}} = X_b^* X_b = H_b.$$

Corollary 9.7 therefore implies that $b^{(u)} \in \mathcal{O}_X^\infty$. Thus the based unitary gauge action preserves the balanced fibre.

Now let $b, c \in \mathcal{O}_X^\infty$. Set

$$Y_b := X_b, \qquad Y_c := X_c, \qquad u := Y_c Y_b^{-1}.$$

The endpoint conditions give

$$u(0) = I, \qquad u(1) = XX^{-1} = I.$$

Moreover, Corollary 9.7 gives

$$Y_b^* Y_b = \exp\Big(x \log(X^* X)\Big) = Y_c^* Y_c.$$

Hence

$$u^* u = (Y_b^*)^{-1} Y_c^* Y_c Y_b^{-1} = (Y_b^*)^{-1} Y_b^* Y_b Y_b^{-1} = I.$$

Thus $u \in \mathcal{K}$.

Finally, since $Y_c = uY_b$, differentiating gives

$$cY_c = Y_c' = u'Y_b + uY_b' = \Big(u'u^{-1} + ubu^{-1}\Big)Y_c.$$

Therefore

$$c = ubu^{-1} + u'u^{-1} = b^{(u)}.$$

Thus any two points of $\mathcal{O}_X^\infty$ lie in the same $\mathcal{K}$-orbit. $\square$

Corollary 9.20. *Let $Y \in H^2([0,1], \mathrm{GL}(d, \mathbb{C}))$ satisfy $Y(0) = I$ and $Y(1) = X$. If Y is balanced, then, after ordering singular values decreasingly,*

$$\sigma_i(Y(x)) = \sigma_i(X)^x, \qquad 1 \le i \le d, \quad 0 \le x \le 1.$$

Equivalently, the logarithms of the singular values vary linearly in the continuum depth coordinate x.

Proof of Corollary 9.20. By Corollary 9.7,

$$Y(x)^*Y(x) = \exp\Big(x\log(X^*X)\Big).$$

The eigenvalues of X^*X are $\sigma_i(X)^2$. Therefore the eigenvalues of

$$\exp\Big(x\log(X^*X)\Big)$$

are $\sigma_i(X)^{2x}$. These are the squares of the singular values of $Y(x)$. Taking positive square roots gives

$$\sigma_i(Y(x)) = \sigma_i(X)^x.$$

$\square$

Remark 9.21. In Chapter 6, we used the matrices

$$H_k = Y_k^*Y_k, \qquad 0 \le k \le N,$$

to describe a finite-depth factorization as a discrete path in the positive-definite cone, with endpoints

$$H_0 = I, \qquad H_N = X^*X.$$

The finite-depth balanced equations are equivalent to the constancy of the multiplicative increments

$$\Pi_k := H_{k-1}^{-1}H_k, \qquad \Pi_{k+1} = \Pi_k.$$

The continuum equation $(H^{-1}H')' = 0$ is the corresponding constancy condition for the logarithmic velocity $H^{-1}H'$. Thus the finite-depth balanced path is a sampled geodesic for the trace metric, while the infinite-depth balanced profile is the geodesic itself.

9.5 Continuum fibre metrics and formal heat flows

In Chapter 8 we rewrote the geometry of a fibre at finite depth as a discrete Hodge theory on the path of layers and then passed formally to the interval. We now prove the flow statements from Section 9.1. The energy, the two metrics, the degree-0 operator, and the degree-1 notation are those fixed in (9.5), (9.6), (9.7), (9.8), and (9.9).

This section has a narrow purpose. We compute the first variation of $\mathcal{E}_\infty$ along tangent directions of a fixed fibre, identify the formal interior vector fields selected by the two metrics, and then translate those vector fields to the transport variable Y and to the path $H = Y^*Y$ in $\mathrm{Herm}^+(d)$. The word *formal* remains essential: the calculation identifies interior differential equations. The problem on a fixed fibre is obtained only after imposing the conditions at the endpoints.

9.5.1 First variation on a fixed fibre

Fix $X \in \mathrm{GL}(d, \mathbb{C})$ and $b \in \mathcal{F}_X^\infty$. By Theorem 9.3, tangent vectors to the fibre have the form $\nabla_b a = a' + [a, b]$, with $a \in H_0^2\big([0,1], \mathfrak{gl}(d,\mathbb{C})\big)$. The condition $a(0) = a(1) = 0$ is part of the description of the tangent space. It is the source of the distinction between an interior equation and the problem with fixed endpoints. The potential a is unique: if $\nabla_b a = 0$ and $a(0) = 0$, then

$$(X_b^{-1} a X_b)' = X_b^{-1}(\nabla_b a) X_b = 0,$$

so $X_b^{-1} a X_b$ is constant. Since $a(0) = 0$, this constant is zero, and hence $a \equiv 0$. This injectivity is what makes the normal fibre metric (9.7) well defined.

For a smooth coefficient field a, the degree-0 operator is

$$\Delta_b^{(0)} a = \nabla_b^* \nabla_b a = -\Big(a' + [a, b]\Big)' + \Big[a' + [a, b], b^*\Big]. \tag{9.23}$$

Proposition 9.22. *For $a, c \in H_0^2([0, 1], \mathfrak{gl}(d, \mathbb{C}))$,*

$$g_b^\iota(\nabla_b a, \nabla_b c) = \Re \int_0^1 \operatorname{tr}\Big(a(x)^* (\Delta_b^{(0)} c)(x)\Big)\, dx. \tag{9.24}$$

Proof. Using the formal adjoint $\nabla_b^* Z = -Z' + [Z, b^*]$ and the endpoint values of a, we obtain

$$\int_0^1 \operatorname{tr}\Big((\nabla_b a)^* \nabla_b c\Big)\, dx = \int_0^1 \operatorname{tr}\Big(a^* \nabla_b^* \nabla_b c\Big)\, dx.$$

Since $\Delta_b^{(0)} = \nabla_b^* \nabla_b$, taking real parts gives (9.24). □

Proposition 9.23. *Let b and a be smooth, and set*

$$\eta := \nabla_b a = a' + [a, b].$$

Then

$$\mathrm{d}(\mu_\infty)_b[\eta] = \Delta_b^{(0)} a + \Big(\Delta_b^{(0)} a\Big)^*. \tag{9.25}$$

Proof. Differentiating $\mu_\infty(b) = -(b + b^*)' + [b, b^*]$ in the direction η gives

$$\mathrm{d}(\mu_\infty)_b[\eta] = -(\eta + \eta^*)' + [\eta, b^*] + [b, \eta^*].$$

On the other hand,

$$\Delta_b^{(0)} a = \nabla_b^* (\nabla_b a) = -\eta' + [\eta, b^*].$$

Taking adjoints gives

$$\left(\Delta_b^{(0)} a\right)^* = -(\eta^*)' + [b, \eta^*].$$

Adding the two identities proves (9.25). $\square$

Corollary 9.24. *For smooth b and $a \in H_0^2([0,1], \mathfrak{gl}(d, \mathbb{C}))$,*

$$\mathrm{d}\mathcal{E}_{\infty,b}[\nabla_b a] = 2\,\Re \int_0^1 \mathrm{tr}\Big(\mu_\infty(b)^* \Delta_b^{(0)} a\Big)\, dx. \tag{9.26}$$

Proof. By differentiating (9.5),

$$\mathrm{d}\mathcal{E}_{\infty,b}[\eta] = \Re \int_0^1 \mathrm{tr}\Big(\mu_\infty(b)^*\, \mathrm{d}(\mu_\infty)_b[\eta]\Big)\, dx.$$

Now take $\eta = \nabla_b a$ and use Proposition 9.23. Since $\mu_\infty(b)$ is Hermitian by Theorem 9.4, the two terms in (9.25) contribute the same real part. This gives (9.26). $\square$

For the metric induced from the ambient L^2 structure, equations (9.24) and (9.26) show that the first variation is represented by the coefficient $2\mu_\infty(b)$:

$$\mathrm{d}\mathcal{E}_{\infty,b}[\nabla_b a] = \Re \int_0^1 \mathrm{tr}\Big((2\mu_\infty(b))^* \Delta_b^{(0)} a\Big)\, dx. \tag{9.27}$$

For the metric transported from L^2 on the coefficients, one formally moves $\Delta_b^{(0)}$ from the test coefficient onto $\mu_\infty(b)$:

$$\mathrm{d}\mathcal{E}_{\infty,b}[\nabla_b a] = \Re \int_0^1 \mathrm{tr}\Big((2\Delta_b^{(0)} \mu_\infty(b))^* a\Big)\, dx. \tag{9.28}$$

Thus the two coefficient fields selected by the first variation are

$$\alpha_\iota(b) = 2\mu_\infty(b), \qquad \alpha_{\mathrm{nor}}(b) = 2\Delta_b^{(0)} \mu_\infty(b). \tag{9.29}$$

These coefficients need not vanish at $x = 0$ and $x = 1$. Therefore (9.29) gives the interior operators; the conditions at the endpoints must be imposed separately.

Proposition 9.25. *Suppose $b(x, t)$ is smooth and evolves by*

$$b_t = -\nabla_b \alpha$$

for a smooth matrix field $\alpha(x, t)$. Then

$$(\mu_\infty)_t = -\Big(\Delta_b^{(0)}\alpha + \big(\Delta_b^{(0)}\alpha\big)^*\Big). \tag{9.30}$$

Proof. Apply Proposition 9.23 to $\eta = b_t = -\nabla_b \alpha$. $\square$

9.5.2 Transport variables and the conditions at the endpoints

The equations for b have a simple expression in a transport variable. This also makes clear where the constraints on the fixed fibre enter.

Lemma 9.26. *Suppose $Y(x, t)$ solves*

$$Y_x = bY.$$

If

$$b_t = -\nabla_b \alpha = -(\alpha_x + [\alpha, b]),$$

then there is an x-independent matrix $C(t)$ such that

$$Y_t = -\alpha Y + YC(t). \tag{9.31}$$

Conversely, (9.31) *implies $b_t = -\nabla_b \alpha$.*

Proof. Set

$$Z := Y^{-1}Y_t.$$

Differentiating $b = Y_x Y^{-1}$ gives

$$b_t = Y Z_x Y^{-1}.$$

A direct calculation also gives

$$\left(Y^{-1}\alpha Y\right)_x = Y^{-1}(\nabla_b \alpha)Y.$$

Hence $b_t = -\nabla_b \alpha$ is equivalent to

$$Z_x = -\left(Y^{-1}\alpha Y\right)_x.$$

Thus

$$Z = -Y^{-1}\alpha Y + C(t),$$

where $C(t)$ is independent of x. Multiplying by Y gives (9.31). The converse follows by starting with (9.31) and differentiating $b = Y_x Y^{-1}$. $\square$

Corollary 9.27. *If $Y(0,t) = I$, then*

$$Y_t = -\alpha Y + Y\alpha(0,t). \tag{9.32}$$

In particular, if $\alpha(0,t) = \alpha(1,t) = 0$, then the endpoints of the normalized path are constant in t.

Proof. Putting $x = 0$ in (9.31) gives $0 = Y_t(0,t) = -\alpha(0,t) + C(t)$. Hence $C(t) = \alpha(0,t)$, which proves (9.32). The endpoint claim follows by evaluating (9.32) at

$x = 1$. ☐

Remark 9.28. If α is based, the normalized transport path satisfies $Y_t = -\alpha Y$ and keeps both endpoints fixed. The coefficients in (9.29) are not generally based. For them, the equation $Y_t = -\alpha Y$ is a convenient gauge for the interior calculation. The problem on a fixed fibre is the same interior equation coupled to (9.16), or equivalently to

$$H(0,t) = I, \qquad H(1,t) = X^*X.$$

9.5.3 The geodesic residual in the positive cone

Let Y be the normalized transport path for b and set $H = Y^*Y$. We use

$$P := H^{-1}H_x, \qquad Q := P_x, \qquad \tau_H := H_{xx} - H_x H^{-1} H_x,$$

as in (9.10). By Theorem 9.6,

$$Q = (H^{-1}H_x)_x = -Y^{-1}\mu_\infty(b)Y. \tag{9.33}$$

Proof of Proposition 9.10. Equation (9.33) gives

$$\mu_\infty(b) = -YQY^{-1}.$$

Therefore

$$\mu_\infty(b)^*\mu_\infty(b) = Y^{-*}Q^*HQY^{-1}.$$

Taking the trace and using cyclicity gives

$$\operatorname{tr}\Big(\mu_\infty(b)^*\mu_\infty(b)\Big) = \operatorname{tr}\Big(Q^*HQH^{-1}\Big).$$

Since

$$Q = (H^{-1}H_x)_x = H^{-1}H_{xx} - H^{-1}H_xH^{-1}H_x,$$

we have

$$HQ = H_{xx} - H_xH^{-1}H_x = \tau_H.$$

The matrix τ_H is Hermitian, so $Q^*H = (HQ)^* = \tau_H$. Hence

$$\operatorname{tr}\Big(Q^*HQH^{-1}\Big) = \operatorname{tr}\Big(H^{-1}\tau_H H^{-1}\tau_H\Big).$$

Integrating over $[0,1]$ and multiplying by $1/2$ proves (9.11). $\square$

Remark 9.29. The trace metric on $\mathrm{Herm}^+(d)$ has path energy

$$\mathcal{L}(H) := \frac{1}{2}\int_0^1 \operatorname{tr}\Big(H^{-1}H_x\, H^{-1}H_x\Big)\, dx,$$

whose Euler–Lagrange equation is $(H^{-1}H_x)_x = 0$. Balancedness is this geodesic equation. The functional $\mathcal{E}_\infty$ is different: by Proposition 9.10, it is the squared L^2 norm of the geodesic residual τ_H. Thus the continuum Kirwan–Ness energy is a bienergy for the trace metric, not the ordinary energy of a path.

9.5.4 Proofs of the formal heat-flow statements

Proof of Theorem 9.11. For the metric induced from the ambient L^2 structure, the first variation is represented by the first coefficient in (9.29). Thus the formal gradient is $\nabla_b(2\mu_\infty(b))$, and the negative gradient equation is (9.12).

The corresponding equation for the moment follows from Proposition 9.25 with $\alpha = 2\mu_\infty(b)$:

$$(\mu_\infty)_t = -2\Big(\Delta_b^{(0)}\mu_\infty(b) + \Big(\Delta_b^{(0)}\mu_\infty(b)\Big)^*\Big). \tag{9.34}$$

It remains to compute the equation for H. Use the gauge used for the interior calculation

$$Y_t = -2\mu_\infty(b)Y.$$

By (9.33),

$$\mu_\infty(b) = -YQY^{-1}, \qquad Q = (H^{-1}H_x)_x.$$

Therefore

$$Y_t = 2YQ.$$

Differentiating $H = Y^*Y$ gives

$$H_t = Y_t^*Y + Y^*Y_t = 2Q^*H + 2HQ.$$

But

$$HQ = H(H^{-1}H_x)_x = H_{xx} - H_xH^{-1}H_x = \tau_H,$$

and τ_H is Hermitian. Hence $Q^*H = HQ$, and

$$H_t = 4\Big(H_{xx} - H_xH^{-1}H_x\Big) = 4\tau_H.$$

This is (9.13). □

Proof of Theorem 9.12. For the metric transported from L^2 on coefficients, the formal integration by parts in (9.28) shows that the first variation is represented by the second coefficient in (9.29). Thus the formal negative gradient equation is (9.14).

The moment equation follows from Proposition 9.25 with $\alpha = 2\Delta_b^{(0)}\mu_\infty(b)$:

$$(\mu_\infty)_t = -2\Big((\Delta_b^{(0)})^2\mu_\infty(b) + \big((\Delta_b^{(0)})^2\mu_\infty(b)\big)^*\Big). \tag{9.35}$$

We now compute the equation for H. For a matrix field Z, set

$$\widetilde{Z} := Y^{-1} Z Y.$$

A direct differentiation gives

$$Y^{-1}(\nabla_b Z)Y = (\widetilde{Z})_x \tag{9.36}$$

and

$$Y^{-1}(\nabla_b^* Z)Y = -(\widetilde{Z})_x + [\widetilde{Z}, P], \qquad P = H^{-1}H_x. \tag{9.37}$$

Apply (9.36) to $Z = \mu_\infty(b)$. By (9.33), $\widetilde{Z} = -Q$, and so

$$Y^{-1}(\nabla_b \mu_\infty(b))Y = -Q_x.$$

Applying (9.37) to $Z = \nabla_b \mu_\infty(b)$ gives

$$Y^{-1}\Big(\Delta_b^{(0)} \mu_\infty(b)\Big)Y = Q_{xx} - [Q_x, P].$$

In the gauge used for the interior calculation

$$Y_t = -2\Big(\Delta_b^{(0)} \mu_\infty(b)\Big)Y,$$

we therefore have

$$Y_t = -2Y\Big(Q_{xx} - [Q_x, P]\Big) = 2Y\Gamma, \qquad \Gamma = -Q_{xx} + [Q_x, P].$$

Finally,

$$H_t = Y_t^* Y + Y^* Y_t = 2\Gamma^* H + 2H\Gamma,$$

which is (9.15). $\square$

Remark 9.30. The continuum fibre $\mathcal{F}_X^\infty$ is a Hilbert submanifold of $\mathcal{B} = H^1([0,1], \mathfrak{gl}(d, \mathbb{C}))$. Its topology is therefore the H^1 topology inherited from $\mathcal{B}$. The fibre metrics introduced above should be viewed as different ways of measuring tangent vectors on this same H^1 fibre, not as different choices of topology.

In the Hodge–Laplacian notation, the metric induced by the ambient L^2 pairing gives the heat flow $\dot{b} = -2\Delta_b^{(1)} b$, while the normal metric gives the biharmonic heat flow $\dot{b} = -2\left(\Delta_b^{(1)}\right)^2 b$. Equivalently, the moment evolves by a second-order equation in the first case and by a fourth-order equation in the second. Linearized at $b = 0$, these become

$$(\mu_\infty)_t = 4\partial_x^2 \mu_\infty \qquad \text{and} \qquad (\mu_\infty)_t = -4\partial_x^4 \mu_\infty,$$

respectively. Thus the normal metric is weaker than the induced metric when viewed on the field b. Formally, it loses one additional derivative. This distinction does not occur at finite depth, since all norms on a finite-dimensional space are equivalent.

Remark 9.31. Remark 9.30 compares two local continuum heat flows obtained from the ambient L^2 metric and from the coefficient L^2 metric. Uhlenbeck's discussion of the infinite-depth DLN points to another natural continuum geometry [Uhl26]. This geometry is compatible with the natural H^1 topology on the coefficient field b, in the sense that it induces the H^1 topology on the configuration space $\mathcal{B} = H^1([0,1], \mathfrak{gl}(d, \mathbb{C}))$.

In this geometry the moment is normalized to decay linearly,

$$(\mu_\infty)_t = -\mu_\infty,$$

at least formally. The corresponding gradient flow is therefore a smooth ODE on the H^1 configuration space, rather than a local second- or fourth-order PDE.

Although the ridge functional $\mathcal{R}$ does not admit a nontrivial continuum limit on the fibre, its gradient vector field does. It is therefore natural to study this flow in the continuum. In particular, the comparison between the different geometries remains meaningful at the level of their induced dynamics.

Remark 9.32. In the scalar case $d = 1$, all commutators vanish. If $u = \log H$, then

$$P = u_x, \qquad Q = u_{xx}, \qquad \Gamma = -u_{xxxx}.$$

Equation (9.15) becomes

$$H_t = -4H(\log H)_{xxxx},$$

or, equivalently,

$$u_t = -4u_{xxxx}.$$

For comparison, the interior equations from Theorems 9.11 and 9.12 are collected in Table 9.1. The table is not a separate statement of well-posedness. The problem on a fixed fibre is obtained by adding the endpoint data (9.16), or equivalently

$$H(0,t) = I, \qquad H(1,t) = X^*X.$$

The first row is the degree-1 heat equation that drives H toward the geodesic equation $(H^{-1}H_x)_x = 0$. The second row is the corresponding biharmonic equation driven by the bienergy $\mathcal{E}_\infty$. In this sense the continuum fibre dynamics are heat flows on the interval: the variable $x \in [0,1]$ is the depth variable, the moment $\mu_\infty(b)$ is the continuum divergence, and balancedness is the zero-divergence condition.

Table 9.1: Continuum interior equations for the induced and normal metrics on the fibre.

Functional and metric	Interior equation in b	Equation for μ_∞	Equation for $H = Y^*Y$
$\mathcal{E}_\infty$ with g^ι	$b_t = -2\Delta_b^{(1)} b$	$(\mu_\infty)_t = -2\Big(\Delta_b^{(0)}\mu_\infty + (\Delta_b^{(0)}\mu_\infty)^*\Big)$	$H_t = 4\big(H_{xx} - H_x H^{-1} H_x\big)$
$\mathcal{E}_\infty$ with g^{nor}	$b_t = -2(\Delta_b^{(1)})^2 b$	$(\mu_\infty)_t = -2\Big((\Delta_b^{(0)})^2\mu_\infty + ((\Delta_b^{(0)})^2\mu_\infty)^*\Big)$	$H_t = 2(\Gamma^* H + H\Gamma)$

Bibliographical notes

Background on gauge theory, Sobolev spaces, and the one-dimensional connection law may be found in [KN63; FU91; AF03]. The Hamiltonian interpretation follows the Atiyah–Bott moment-map framework [AB83]. The Yang–Mills/GIT perspective is surveyed in [Tra17]. The continuum formulation of deep linear networks developed here originates in Uhlenbeck's lecture [Uhl26].

Among existing infinite-depth DLN results, the closest antecedent is [CMV23], which studies the infinite-depth limit of real deep linear networks in a matrix-completion setting and identifies the limiting end-to-end metric.

Chapter 10

Quotient Geometry at Infinite Depth

We now identify the quotient geometry that survives at infinite depth. The space $\mathrm{GL}(d, \mathbb{C})$ is obtained by taking continuum coefficients modulo based unitary gauge transformations. Each matrix arises from such a coefficient, and equivalent coefficients determine the same matrix.

At infinite depth, the quantities defined at finite depth no longer remain finite without modification. After introducing the appropriate renormalizations, the metric on $\mathrm{GL}(d, \mathbb{C})$, the Kähler potential, the entropy, and the Ricci form admit explicit formulas.

The order of construction in this chapter is deliberate. In finite depth, the balanced manifold and the quotient by the compact unitary group are finite-dimensional objects, so the quotient geometry can be formed directly and then computed. A literal continuum analogue would begin from an infinite-dimensional orbit space and require endowing it with a Hilbert manifold structure, a step that is technically delicate and not intrinsic to the geometric formulas we seek. We therefore avoid this route. The endpoint classification from Chapter 9 first identifies the set of balanced based-gauge orbits with the finite-dimensional space $\mathrm{GL}(d, \mathbb{C})$. We then define the candidate

Hermitian tensor directly on $\mathrm{GL}(d, \mathbb{C})$, through the explicit operator $\mathcal{A}_{\infty,X}$. The submersion interpretation is recovered only a posteriori, by constructing the L^2-horizontal lifts and proving that their quotient tensor is exactly h_∞. In the same spirit, the renormalized entropy is first expressed as a determinant of $\mathcal{A}_{\infty,X}^{-1}$, and only afterward converted into the Ricci-potential formula. Thus the total space, base, fibres, and metric are all given explicitly, without invoking an abstract Hilbert quotient.

10.1 Main results

We begin by describing the set of balanced coefficients up to based unitary gauge transformations. The notation $\mathcal{M}_\infty/\mathcal{K}$ denotes the set of $\mathcal{K}$-orbits in $\mathcal{M}_\infty$. No manifold structure is assumed at this stage. Theorem 10.1 shows that the endpoint matrix determines the orbit. For each X, all balanced coefficients with endpoint X lie in a single based unitary gauge orbit. This is the continuum analogue of the finite-depth result that balanced factorizations, taken up to the compact unitary symmetry, parametrize the space of end-to-end matrices.

Theorem 10.1. *The restriction $\phi_\infty|_{\mathcal{M}_\infty} : \mathcal{M}_\infty \to \mathrm{GL}(d, \mathbb{C})$ is constant on $\mathcal{K}$-orbits and induces a bijection $\bar{\phi}_\infty : \mathcal{M}_\infty/\mathcal{K} \to \mathrm{GL}(d, \mathbb{C})$, $[b] \mapsto \phi_\infty(b)$. Equivalently, for every $X \in \mathrm{GL}(d, \mathbb{C})$, the balanced fibre $\mathcal{O}_X^\infty := \mathcal{F}_X^\infty \cap \mathcal{M}_\infty$ is a single $\mathcal{K}$-orbit.*

The metric obtained by taking L^2-orthogonal complements to the gauge directions can be written directly on $\mathrm{GL}(d, \mathbb{C})$. The relevant continuum operator is obtained by replacing the finite layer sum with an integral. For $X \in \mathrm{GL}(d, \mathbb{C})$, define

$$\mathcal{A}_{\infty,X}(Z) := \int_0^1 (XX^*)^{1-t}\, Z\, (X^*X)^t\, dt, \qquad Z \in \mathrm{M}_d(\mathbb{C}). \tag{10.1}$$

Proposition 10.2. *For every $X \in \mathrm{GL}(d,\mathbb{C})$, the operator $\mathcal{A}_{\infty,X}$ is Hermitian, positive definite, and invertible with respect to the Frobenius Hermitian form on $\mathrm{M}_d(\mathbb{C})$.*

Thus the formula

$$h_\infty(X)(Z_1,Z_2) := \operatorname{tr}\Big(Z_1^* \mathcal{A}_{\infty,X}^{-1}(Z_2)\Big) \tag{10.2}$$

defines a Hermitian metric on $\mathrm{GL}(d,\mathbb{C})$. We write $g_\infty := \Re h_\infty$ and $\omega_\infty := \Im h_\infty$.

The quotient tensor is defined by horizontal lift. For $b \in \mathcal{M}_\infty$, set $\mathcal{V}_b := T_b(\mathcal{K}\cdot b)$ and $\mathcal{H}_b := T_b\mathcal{M}_\infty \cap \mathcal{V}_b^{\perp_{L^2}}$. Here $\langle \alpha,\beta\rangle_{L^2} := \int_0^1 \operatorname{tr}(\alpha(x)^*\beta(x))\,dx$. If $X = \phi_\infty(b)$, and if $(D\phi_\infty)_b : \mathcal{H}_b \to T_X\mathrm{GL}(d,\mathbb{C})$ is an isomorphism, define $h_{\mathrm{quot}}(X)(Z_1,Z_2) := \langle \xi_1,\xi_2\rangle_{L^2}$, where $\xi_a \in \mathcal{H}_b$ is the unique horizontal lift satisfying $(D\phi_\infty)_b\xi_a = Z_a$. The based gauge action is isometric and leaves ϕ_∞ unchanged, so this definition is independent of the balanced representative once the horizontal-lift isomorphism is known. Theorem 10.5 proves the required isomorphism at a diagonal representative, and bi-unitary invariance transports it to all of $\mathrm{GL}(d,\mathbb{C})$.

Theorem 10.3. *Under the identification $\bar{\phi}_\infty : \mathcal{M}_\infty/\mathcal{K} \to \mathrm{GL}(d,\mathbb{C})$, the Hermitian tensor obtained from L^2-orthogonal horizontal lifts is h_∞. Hence the induced Riemannian metric and Kähler form on the space of end-to-end matrices are $g_\infty = \Re h_\infty$ and $\omega_\infty = \Im h_\infty$.*

The weights in this metric are logarithmic means. If $\Sigma = \operatorname{diag}(\sigma_1,\dots,\sigma_d)$ and $r_i := \sigma_i^2$, set

$$\ell_{ij} := \begin{cases} \dfrac{r_i - r_j}{\log r_i - \log r_j}, & r_i \neq r_j, \\ \\ r_i, & r_i = r_j. \end{cases} \tag{10.3}$$

Then $\mathcal{A}_{\infty,\Sigma}(E_{ij}) = \ell_{ij}E_{ij}$ and $h_\infty(\Sigma)(E_{ij},E_{k\ell}) = \delta_{ik}\delta_{j\ell}\,\ell_{ij}^{-1}$. Thus, if $X = Q\Sigma R^*$ and u_i, v_j are the corresponding singular vectors, then $e_{ij}^\infty := \ell_{ij}^{1/2}u_iv_j^*$ is an h_∞-orthonormal basis of $T_X\mathrm{GL}(d,\mathbb{C})$.

The metric has a simple Kähler potential after the finite-depth potential is renormalized.

Theorem 10.4. *Let*

$$\Psi_\infty(X) := \frac{1}{4}\operatorname{tr}\Big((\log(X^*X))^2\Big).$$

Then $\omega_\infty = \mathrm{i}\,\partial\bar{\partial}\Psi_\infty$.

The endpoint modes that enter the horizontal lift calculation are explicit. At a diagonal endpoint $X = \Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_d)$, the balanced representative is $Y_\Sigma(x) = \Sigma^x$, with $b_\Sigma = \log\Sigma$. For $1 \le i, j \le d$, set $c^R_{ij}(x) := \sigma_i^{1-x}\sigma_j^x$ and $c^L_{ij}(x) := \sigma_i^x\sigma_j^{1-x}$. Their boundary values are $c^R_{ij}(0) = \sigma_i$, $c^R_{ij}(1) = \sigma_j$, $c^L_{ij}(0) = \sigma_j$, and $c^L_{ij}(1) = \sigma_i$. The corresponding right-endpoint tangent field is $\Gamma^\infty_{ij}(x) := c^R_{ij}(x)E_{ij}$.

Theorem 10.5. *For every* $1 \le i, j \le d$, $\Gamma^\infty_{ij} \in T_{b_\Sigma}\mathcal{M}_\infty \cap T_{b_\Sigma}(\mathcal{K}\cdot b_\Sigma)^{\perp_{L^2}}$. *Moreover,* $(D\phi_\infty)_{b_\Sigma}[\Gamma^\infty_{ij}] = \ell_{ij}E_{ij}$ *and* $\|\Gamma^\infty_{ij}\|^2_{L^2} = \ell_{ij}$. *Consequently, the normalized lifts* $\widehat{\Gamma}^\infty_{ij} := \ell_{ij}^{-1/2}\Gamma^\infty_{ij}$ *form an* L^2*-orthonormal basis of the complexified horizontal space* $\mathcal{H}^{\mathbb{C}}_{b_\Sigma} := (T_{b_\Sigma}\mathcal{M}_\infty \cap T_{b_\Sigma}(\mathcal{K}\cdot b_\Sigma)^{\perp_{L^2}}) \otimes_{\mathbb{R}} \mathbb{C}$ *and project to the* $h_\infty(\Sigma)$*-orthonormal basis* $\ell_{ij}^{1/2}E_{ij}$.

The entropy that survives at infinite depth is a renormalized entropy rather than the logarithm of a finite-dimensional orbit volume. With the logarithmic mean ℓ_{ij} defined by (10.3), set

$$S_\infty(X) := \sum_{i=1}^{d}\log\sigma_i + \sum_{1\le i<j\le d}\log\ell_{ij}, \tag{10.4}$$

where $\sigma_1, \dots, \sigma_d$ are the singular values of X. This formula is symmetric in the singular values and extends continuously across collisions.

Theorem 10.6. *For every* $X \in \mathrm{GL}(d,\mathbb{C})$, $S_\infty(X) = -\frac{1}{2}\log\det(\mathcal{A}^{-1}_{\infty,X})$, *where the determinant is taken for the complex-linear operator* $\mathcal{A}^{-1}_{\infty,X}$ *on* $\mathrm{M}_d(\mathbb{C})$.

Since $X \mapsto \mathcal{A}_{\infty,X}$ is smooth and positive definite, the determinant identity shows that S_∞ is globally smooth on $\mathrm{GL}(d, \mathbb{C})$. In particular, the singular-value expression in (10.4) defines a smooth bi-unitarily invariant function on $\mathrm{GL}(d, \mathbb{C})$, including across singular-value collisions.

For the Ricci form we use the convention $\rho_h = -\mathrm{i}\,\partial\bar{\partial}\log\det(h_{\alpha\bar{\beta}})$ in a local holomorphic frame.

Theorem 10.7. *The Ricci form of h_∞ is*

$$\rho_\infty = 2\mathrm{i}\,\partial\bar{\partial}S_\infty.$$

Equivalently, $2S_\infty$ is a Ricci potential for the infinite-depth metric.

The Ricci-potential formula also shows that the infinite-depth metric is not Einstein, even though both ω_∞ and ρ_∞ are determined by singular values.

Corollary 10.8. *If $d \geq 2$, then the infinite-depth metric h_∞ is not Kähler–Einstein.*

These statements construct the tensors on the set of balanced gauge orbits by identifying that set with $\mathrm{GL}(d, \mathbb{C})$. They do not prove that the rescaled finite-depth metrics Nh_N, their geodesics, or the finite-depth quotient spaces converge to the objects constructed here. The estimates needed for such a convergence theorem are separate from the quotient identification and curvature computations carried out in this chapter.

10.2 The quotient and the Hermitian metric

At finite depth, we first identify the balanced quotient with $\mathrm{GL}(d, \mathbb{C})$ and only then compute the metric on that quotient. We use the same order here. In Chapter 9 we proved that, once the endpoint is fixed, balanced transport paths differ only by transformations in the based unitary gauge group. Thus the endpoint matrix is the invariant that remains after quotienting by $\mathcal{K}$.

Proof of Theorem 10.1. By (9.19) and the based endpoint condition $g(1) = I$, the endpoint map ϕ_∞ is constant on based gauge orbits. Hence its restriction to $\mathcal{M}_\infty$ induces a well-defined map

$$\bar{\phi}_\infty : \mathcal{M}_\infty/\mathcal{K} \to \mathrm{GL}(d, \mathbb{C}), \qquad [b] \longmapsto \phi_\infty(b).$$

We first prove surjectivity. Fix $X \in \mathrm{GL}(d, \mathbb{C})$. By Corollary 9.8, there exists a balanced transport path from I to X. Equivalently,

$$\mathcal{O}_X^\infty = \mathcal{F}_X^\infty \cap \mathcal{M}_\infty$$

is nonempty. Thus X lies in the image of $\bar{\phi}_\infty$.

We next prove injectivity. Suppose $b, c \in \mathcal{M}_\infty$ satisfy

$$\phi_\infty(b) = \phi_\infty(c) = X.$$

Then $b, c \in \mathcal{O}_X^\infty$. By Corollary 9.9, the set $\mathcal{O}_X^\infty$ is a single $\mathcal{K}$-orbit. Hence $c = b^{(u)}$ for some $u \in \mathcal{K}$, and therefore $[b] = [c]$ in $\mathcal{M}_\infty/\mathcal{K}$. □

Theorem 10.1 identifies the set of balanced gauge orbits with $\mathrm{GL}(d, \mathbb{C})$. It does not require, and does not assert, any additional Hilbert-manifold structure on the orbit set. The tensor on the quotient is defined by taking L^2-orthogonal complements to the based gauge directions. Namely, for $b \in \mathcal{M}_\infty$, set

$$\mathcal{V}_b := T_b(\mathcal{K} \cdot b), \qquad \mathcal{H}_b := T_b\mathcal{M}_\infty \cap \mathcal{V}_b^{\perp_{L^2}}.$$

Here $\langle \alpha, \beta \rangle_{L^2} := \int_0^1 \operatorname{tr}(\alpha(x)^* \beta(x))\, dx$. If $X = \phi_\infty(b)$ and $(D\phi_\infty)_b$ restricts to an isomorphism $\mathcal{H}_b \to T_X \mathrm{GL}(d, \mathbb{C})$, the quotient Hermitian tensor is

$$h_{\mathrm{quot}}(X)(Z_1, Z_2) := \langle \xi_1, \xi_2 \rangle_{L^2}, \tag{10.5}$$

where $\xi_a \in \mathcal{H}_b$ is the unique horizontal lift of Z_a. The based unitary gauge action is isometric for the L^2-metric and preserves ϕ_∞, so (10.5) is independent of the chosen balanced representative. The completeness statement in Theorem 10.5 verifies the required horizontal-lift isomorphism at diagonal representatives, and bi-unitary invariance then transports the construction to every point of $\mathrm{GL}(d, \mathbb{C})$.

The remaining task is to compute this tensor explicitly. That calculation is carried out after the horizontal lifts have been constructed. We first verify that the operator in (10.1) defines a Hermitian metric and record the diagonal form used throughout the rest of the chapter.

We write

$$\langle Y, Z \rangle_F := \operatorname{tr}(Y^* Z)$$

for the Frobenius Hermitian form on $\mathrm{M}_d(\mathbb{C})$.

Proof of Proposition 10.2. Let

$$P := XX^*, \qquad Q := X^*X.$$

Both P and Q are positive definite. For $0 \le t \le 1$, define

$$T_t(Z) := P^{1-t} Z Q^t.$$

For $Y, Z \in \mathrm{M}_d(\mathbb{C})$, cyclicity of the trace gives

$$\begin{aligned}\langle Y, T_t Z\rangle_F &= \mathrm{tr}\Big(Y^* P^{1-t} Z Q^t\Big)\\ &= \mathrm{tr}\Big(Q^t Y^* P^{1-t} Z\Big)\\ &= \langle T_t Y, Z\rangle_F.\end{aligned}$$

Thus T_t is Hermitian with respect to $\langle\cdot,\cdot\rangle_F$. Hence $\mathcal{A}_{\infty,X} = \int_0^1 T_t\, dt$ is Hermitian.

For $Z \neq 0$,

$$\langle Z, T_t Z\rangle_F = \mathrm{tr}\Big(Z^* P^{1-t} Z Q^t\Big) = \Big\| P^{(1-t)/2} Z Q^{t/2} \Big\|_F^2 > 0.$$

Therefore

$$\langle Z, \mathcal{A}_{\infty,X} Z\rangle_F = \int_0^1 \Big\| P^{(1-t)/2} Z Q^{t/2} \Big\|_F^2 \, dt > 0$$

for every nonzero Z. Thus $\mathcal{A}_{\infty,X}$ is positive definite. Since $\mathrm{M}_d(\mathbb{C})$ is finite-dimensional, a positive definite Hermitian operator is invertible. This proves the proposition. $\square$

The same argument shows that $X \mapsto \mathcal{A}_{\infty,X}$ is smooth on $\mathrm{GL}(d,\mathbb{C})$: the functional calculus is smooth on the positive-definite cone, the integral is taken over a compact interval, and inversion is smooth on the open set of invertible operators. Hence (10.2) defines a smooth Hermitian metric.

Lemma 10.9. *For $Q, R \in \mathrm{U}(d)$, $X \in \mathrm{GL}(d,\mathbb{C})$, and $Z \in \mathrm{M}_d(\mathbb{C})$,*

$$\mathcal{A}_{\infty,QXR^*}(QZR^*) = Q\,\mathcal{A}_{\infty,X}(Z)\,R^*.$$

Consequently,

$$h_\infty(QXR^*)(QZ_1R^*, QZ_2R^*) = h_\infty(X)(Z_1, Z_2).$$

Proof. We have

$$(QXR^*)(QXR^*)^* = QXX^*Q^*, \qquad (QXR^*)^*(QXR^*) = RX^*XR^*.$$

Using (10.1),

$$\begin{aligned}
\mathcal{A}_{\infty,QXR^*}(QZR^*) &= \int_0^1 (QXX^*Q^*)^{1-t}(QZR^*)(RX^*XR^*)^t\,dt \\
&= \int_0^1 Q(XX^*)^{1-t}Z(X^*X)^tR^*\,dt \\
&= Q\,\mathcal{A}_{\infty,X}(Z)\,R^*.
\end{aligned}$$

The first identity implies

$$\mathcal{A}^{-1}_{\infty,QXR^*}(QWR^*) = Q\,\mathcal{A}^{-1}_{\infty,X}(W)\,R^*$$

for every $W \in \mathrm{M}_d(\mathbb{C})$. Substituting $W = Z_2$ into (10.2) gives

$$\begin{aligned}
h_\infty(QXR^*)(QZ_1R^*, QZ_2R^*) &= \operatorname{tr}\Big((QZ_1R^*)^*Q\,\mathcal{A}^{-1}_{\infty,X}(Z_2)\,R^*\Big) \\
&= \operatorname{tr}\Big(RZ_1^*\mathcal{A}^{-1}_{\infty,X}(Z_2)R^*\Big) \\
&= \operatorname{tr}\Big(Z_1^*\mathcal{A}^{-1}_{\infty,X}(Z_2)\Big)\,.
\end{aligned}$$

This is the desired invariance. $\square$

The diagonal calculation does not require distinct singular values. The continuous value of the logarithmic mean in (10.3) is essential when singular values collide.

Proposition 10.10. *Let*

$$\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_d), \qquad r_i := \sigma_i^2.$$

Then the matrix units diagonalize $\mathcal{A}_{\infty,\Sigma}$, and

$$\mathcal{A}_{\infty,\Sigma}(E_{ij}) = \ell_{ij} E_{ij},$$

with ℓ_{ij} defined in (10.3)*. Consequently,*

$$h_\infty(\Sigma)(E_{ij}, E_{k\ell}) = \delta_{ik}\delta_{j\ell}\, \ell_{ij}^{-1}.$$

Proof. Since $\Sigma\Sigma^* = \Sigma^*\Sigma = \mathrm{diag}(r_1, \dots, r_d)$, we have

$$(\Sigma\Sigma^*)^{1-t} E_{ij} (\Sigma^*\Sigma)^t = r_i^{1-t} r_j^t E_{ij}.$$

Therefore

$$\mathcal{A}_{\infty,\Sigma}(E_{ij}) = \left(\int_0^1 r_i^{1-t} r_j^t \, dt \right) E_{ij}.$$

If $r_i \neq r_j$, then

$$\int_0^1 r_i^{1-t} r_j^t \, dt = r_i \int_0^1 \left(\frac{r_j}{r_i} \right)^t \, dt = \frac{r_i - r_j}{\log r_i - \log r_j}.$$

If $r_i = r_j$, the integrand is constant and the integral equals r_i. Thus the integral is exactly ℓ_{ij} as defined in (10.3).

The matrix units are orthonormal for $\langle \cdot, \cdot \rangle_F$. Hence

$$\begin{aligned} h_\infty(\Sigma)(E_{ij}, E_{k\ell}) &= \operatorname{tr}\Big(E_{ij}^* \mathcal{A}_{\infty,\Sigma}^{-1}(E_{k\ell})\Big) \\ &= \delta_{ik}\delta_{j\ell}\, \ell_{ij}^{-1}. \end{aligned}$$

□

Proof of the singular-vector basis formula in Section 10.1. Let

$$X = Q\Sigma R^*$$

be a singular value decomposition, and write $u_i = Qe_i$ and $v_j = Re_j$. Then

$$u_i v_j^* = QE_{ij}R^*.$$

By Lemma 10.9 and Proposition 10.10,

$$\begin{aligned} h_\infty(X)(u_i v_j^*, u_k v_\ell^*) &= h_\infty(Q\Sigma R^*)(QE_{ij}R^*, QE_{k\ell}R^*) \\ &= h_\infty(\Sigma)(E_{ij}, E_{k\ell}) \\ &= \delta_{ik}\delta_{j\ell}\, \ell_{ij}^{-1}. \end{aligned}$$

Therefore the vectors

$$e_{ij}^\infty := \ell_{ij}^{1/2} u_i v_j^*$$

form an h_∞-orthonormal basis of $T_X\mathrm{GL}(d, \mathbb{C})$. □

Remark 10.11. For fixed X, the operators $N^{-1}\mathcal{A}_{N,X}$ form a Riemann sum for $\mathcal{A}_{\infty,X}$:

$$\frac{1}{N}\mathcal{A}_{N,X}(Z) = \frac{1}{N}\sum_{k=1}^{N} (XX^*)^{\frac{N-k}{N}} Z (X^*X)^{\frac{k-1}{N}} \longrightarrow \int_0^1 (XX^*)^{1-t} Z (X^*X)^t \, dt.$$

This observation explains the normalization behind the candidate limit $Nh_N \to h_\infty$. It is not a convergence theorem for the quotient tensors. Such a theorem would require uniform control in X, after inversion, and for whatever derivatives are required by the chosen topology.

Remark 10.12. On $\mathrm{GL}(d, \mathbb{R})$, the real part $g_\infty = \Re h_\infty$ has the weights appearing in the infinite-depth metric of Cohen, Menon, and Veraszto [CMV23]. Indeed, at $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_d)$, for $r_i \neq r_j$,

$$\ell_{ij}^{-1} = \frac{\log r_i - \log r_j}{r_i - r_j} = \frac{2\log(\sigma_i/\sigma_j)}{\sigma_i^2 - \sigma_j^2}.$$

When $r_i = r_j$, the continuous value is

$$\ell_{ij}^{-1} = r_i^{-1} = \sigma_i^{-2}.$$

10.3 The renormalized Kähler potential

In Chapter 5 we proved that

$$\Psi_N(X) = \frac{N}{2} \operatorname{tr}\Big((X^*X)^{1/N}\Big)$$

is a Kähler potential for h_N. Since the infinite-depth metric is the object suggested by the rescaling Nh_N, the quantity to renormalize is $N\Psi_N$. For

$$H := X^*X,$$

the functional calculus expansion

$$H^{1/N} = \exp\Big(\frac{1}{N}\log H\Big) = I + \frac{1}{N}\log H + \frac{1}{2N^2}(\log H)^2 + O(N^{-3})$$

formally gives

$$N\Psi_N(X) = \frac{N^2 d}{2} + \frac{N}{2}\log\det(X^*X) + \frac{1}{4}\operatorname{tr}\Big((\log(X^*X))^2\Big) + O(N^{-1}). \tag{10.6}$$

The first term in (10.6) is constant. The second term is pluriharmonic: locally on $\mathrm{GL}(d,\mathbb{C})$,

$$\log\det(X^*X) = \log\det X + \overline{\log\det X}.$$

Thus neither term contributes to $\mathrm{i}\,\partial\bar\partial$. The finite term left by this subtraction is the potential

$$\Psi_\infty(X) := \frac{1}{4}\operatorname{tr}\Big((\log(X^*X))^2\Big), \tag{10.7}$$

which was stated in Theorem 10.4. The proof below uses Ψ_∞ directly; it does not use a convergence theorem for the family of renormalized finite-depth potentials.

We use the convention $h_\infty = g_\infty + \mathrm{i}\omega_\infty$ on real tangent vectors. Equivalently, if $h_{\alpha\bar\beta} = h_\infty(\partial_{z_\alpha}, \partial_{z_\beta})$ in holomorphic coordinates, then

$$\omega_\infty = \frac{\mathrm{i}}{2}\sum_{\alpha,\beta} h_{\alpha\bar\beta}\, dz_\alpha \wedge d\bar z_\beta.$$

This is the convention responsible for the factor $1/2$ in the Hessian calculation below.

Proof of Theorem 10.4. Set

$$\varphi(r) := \frac{1}{4}(\log r)^2, \qquad \Psi_\infty(X) = \operatorname{tr}\Big(\varphi(X^*X)\Big).$$

Both Ψ_∞ and h_∞ are invariant under the bi-unitary action

$$X \longmapsto QXR^*, \qquad Q, R \in \mathrm{U}(d).$$

It therefore suffices to prove the identity at a positive diagonal matrix

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d), \qquad R := \Sigma^2 = \mathrm{diag}(r_1, \ldots, r_d), \qquad r_i = \sigma_i^2.$$

Use the holomorphic coordinates inherited from $\mathrm{M}_d(\mathbb{C})$:

$$X = \Sigma + \sum_{a,b} z_{ab} E_{ab}.$$

Let

$$\mathcal{F}(H) := \mathrm{tr}\Big(\varphi(H)\Big), \qquad H = X^* X.$$

At $X = \Sigma$,

$$\partial_{z_{ij}} H\Big|_\Sigma = A_{ij} := \Sigma E_{ij} = \sigma_i E_{ij},$$

$$\partial_{\bar{z}_{k\ell}} H\big|_\Sigma = A_{k\ell}^* = \sigma_k E_{\ell k},$$

and

$$\partial_{z_{ij}} \partial_{\bar{z}_{k\ell}} H\Big|_\Sigma = E_{k\ell}^* E_{ij} = \delta_{ik} E_{\ell j}.$$

For the spectral trace $\mathcal{F}$, the standard divided-difference formulas for derivatives of spectral trace functions [Bha07, Chapter V] at the diagonal matrix R give

$$D\mathcal{F}_R[U] = \sum_a \varphi'(r_a) U_{aa},$$

and, after complex-bilinear extension of the second differential,

$$D^2\mathcal{F}_R[U,V] = \sum_{a,b} \psi(r_a, r_b)\, U_{ab}V_{ba},$$

where

$$\psi(x,y) := \begin{cases} \dfrac{\varphi'(x)-\varphi'(y)}{x-y}, & x \neq y, \\ \\ \varphi''(x), & x = y. \end{cases}$$

The chain rule therefore gives

$$\begin{aligned} \partial_{z_{ij}}\partial_{\bar{z}_{k\ell}}\Psi_\infty\Big|_\Sigma &= D^2\mathcal{F}_R[A_{ij}, A^*_{k\ell}] + D\mathcal{F}_R[\delta_{ik}E_{\ell j}] \\ &= \delta_{ik}\delta_{j\ell}\left(r_i\psi(r_i,r_j) + \varphi'(r_j)\right). \end{aligned} \tag{10.8}$$

We now evaluate the coefficient in (10.8). Since

$$\varphi'(r) = \frac{\log r}{2r},$$

if $r_i \neq r_j$, then

$$\begin{aligned} r_i\psi(r_i,r_j) + \varphi'(r_j) &= r_i\frac{\varphi'(r_i)-\varphi'(r_j)}{r_i - r_j} + \varphi'(r_j) \\ &= \frac{\log r_i - \log r_j}{2(r_i - r_j)} = \frac{1}{2\ell_{ij}}. \end{aligned}$$

If $r_i = r_j$, then $\psi(r_i, r_i) = \varphi''(r_i)$, and

$$r_i\varphi''(r_i) + \varphi'(r_i) = \frac{1}{2r_i} = \frac{1}{2\ell_{ij}},$$

where the last equality uses the continuous extension $\ell_{ij} = r_i$ when $r_i = r_j$. Hence, for all indices,

$$\partial_{z_{ij}} \partial_{\bar{z}_{k\ell}} \Psi_\infty \Big|_\Sigma = \frac{1}{2}\, \delta_{ik} \delta_{j\ell}\, \ell_{ij}^{-1}. \tag{10.9}$$

By Proposition 10.10,

$$h_\infty(\Sigma)(E_{ij}, E_{k\ell}) = \delta_{ik} \delta_{j\ell}\, \ell_{ij}^{-1}.$$

Thus, in the coordinates z_{ij},

$$\omega_\infty(\Sigma) = \frac{\mathrm{i}}{2} \sum_{i,j} \ell_{ij}^{-1}\, dz_{ij} \wedge d\bar{z}_{ij} = \mathrm{i}\, \partial\bar{\partial} \Psi_\infty(\Sigma).$$

Since every point of $\mathrm{GL}(d, \mathbb{C})$ lies in the bi-unitary orbit of a positive diagonal matrix, Lemma 10.9 and the bi-unitary invariance of Ψ_∞ give the identity on all of $\mathrm{GL}(d, \mathbb{C})$. ☐

Corollary 10.13. *For every $X \in \mathrm{GL}(d, \mathbb{C})$,*

$$\Psi_\infty(X) = \frac{1}{4}\, d_{\mathrm{tr}}\Big(\mathrm{Id}, X^* X\Big)^2.$$

Proof. For the trace metric on $\mathbb{P}_d$,

$$d_{\mathrm{tr}}(\mathrm{Id}, H)^2 = \mathrm{tr}\Big((\log H)^2\Big)$$

[Bha07]. Taking $H = X^* X$ and using (10.7) gives the claimed identity. ☐

Corollary 10.14. *For every $Q, U \in \mathrm{U}(d)$, the maps*

$$L_{Q,U}(X) := Q X U^*, \qquad J_{Q,U}(X) := Q (X^{-1})^* U^*$$

are Riemannian isometries of $(\mathrm{GL}(d, \mathbb{C}), g_\infty)$. *The maps $L_{Q,U}$ are holomorphic, and* *the maps $J_{Q,U}$ are anti-holomorphic.*

Proof. For $L_{Q,U}(X) = QXU^*$,

$$L_{Q,U}(X)^* L_{Q,U}(X) = U(X^*X)U^*.$$

For $J_{Q,U}(X) = Q(X^{-1})^*U^*$,

$$J_{Q,U}(X)^* J_{Q,U}(X) = U(X^*X)^{-1}U^*.$$

Since functional calculus commutes with unitary conjugation and

$$\log\Big((X^*X)^{-1}\Big) = -\log(X^*X),$$

both families preserve Ψ_∞.

If F is holomorphic and $F^*\Psi_\infty = \Psi_\infty$, then Theorem 10.4 gives

$$F^*\omega_\infty = F^*(\mathrm{i}\,\partial\bar{\partial}\Psi_\infty) = \mathrm{i}\,\partial\bar{\partial}(\Psi_\infty \circ F) = \omega_\infty.$$

Thus each $L_{Q,U}$ is a Kähler isometry.

If F is anti-holomorphic and $F^*\Psi_\infty = \Psi_\infty$, then

$$F^*\omega_\infty = -\omega_\infty.$$

Since $F_*J = -JF_*$, we obtain, for real tangent vectors V, W,

$$g_\infty(F_*V, F_*W) = \omega_\infty(F_*V, JF_*W) = -\omega_\infty(F_*V, F_*JW) = \omega_\infty(V, JW) = g_\infty(V, W).$$

Thus each $J_{Q,U}$ is a Riemannian isometry. The formulas defining $L_{Q,U}$ and $J_{Q,U}$ show their holomorphic and anti-holomorphic character, respectively. □

Remark 10.15. At finite depth, the Kähler potential is the minimum of the ridge regularizer on the fibre:

$$\Psi_N(X) = \min_{W \in \mathcal{F}_X} \mathcal{R}(W), \qquad \mathcal{R}(W) := \frac{1}{2} \sum_{k=1}^{N} \|W_k\|_F^2.$$

The function Ψ_∞ records the finite part of these minimum values after subtracting the divergent constant in (10.6) and the pluriharmonic term involving $\log\det(X^*X)$. We do not interpret Ψ_∞ as the minimum of a ridge functional on the continuum fibre $\mathcal{F}_X^\infty$, because the ridge regularizer itself has no nontrivial limit on that fibre.

10.4 The adapted basis in the continuum

We now build the objects that occur in Theorem 10.5. The calculation is the continuum analogue of the adapted basis from Chapter 5. At a diagonal endpoint, the matrix sectors decouple. The based gauge condition imposes Dirichlet boundary conditions on the vertical directions, while the horizontal directions come from the two endpoint solutions of the same scalar differential equation.

No distinctness assumption on the singular values is used in this section.

10.4.1 The balanced diagonal representative

Let

$$X = \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d), \qquad \sigma_i > 0,$$

and let

$$Y_\Sigma(x) = \Sigma^x = \exp(x \log \Sigma), \qquad 0 \le x \le 1, \tag{10.10}$$

be the balanced transport path from I to Σ. Its coefficient field is the constant diagonal matrix

$$b_\Sigma := Y_\Sigma' Y_\Sigma^{-1} = \log \Sigma = \operatorname{diag}(\log \sigma_1, \ldots, \log \sigma_d). \tag{10.11}$$

For $1 \le i, j \le d$, set

$$\gamma_{ij} := \log \sigma_j - \log \sigma_i. \tag{10.12}$$

If

$$a(x) = f(x) E_{ij},$$

then

$$\nabla_{b_\Sigma} a = \Big(f'(x) + \gamma_{ij} f(x)\Big) E_{ij}. \tag{10.13}$$

Thus each E_{ij}-sector is governed by one scalar differential expression on the interval.

Define

$$L_{ij} := -\frac{d^2}{dx^2} + \gamma_{ij}^2 = \left(-\frac{d}{dx} + \gamma_{ij}\right)\left(\frac{d}{dx} + \gamma_{ij}\right). \tag{10.14}$$

We write L_{ij}^D for the Dirichlet realization of L_{ij}, with domain

$$\operatorname{Dom}(L_{ij}^D) = H^2([0,1], \mathbb{C}) \cap H_0^1([0,1], \mathbb{C}).$$

Proposition 10.16. *For $f \in H_0^1([0,1], \mathbb{C})$,*

$$\Big\|\nabla_{b_\Sigma}(f E_{ij})\Big\|_{L^2}^2 = \int_0^1 \Big(|f'(x)|^2 + \gamma_{ij}^2 |f(x)|^2\Big)\, dx. \tag{10.15}$$

The quadratic form on the right is the closed form associated with L_{ij}^D. In particular, *if* $f \in \mathrm{Dom}(L^D_{ij})$*, then*

$$\left\|\nabla_{b_\Sigma}(fE_{ij})\right\|^2_{L^2} = \int_0^1 \overline{f(x)}\,(L^D_{ij}f)(x)\,dx. \tag{10.16}$$

Proof. Since $\gamma_{ij} \in \mathbb{R}$,

$$\left\|\nabla_{b_\Sigma}(fE_{ij})\right\|^2_{L^2} = \int_0^1 |f' + \gamma_{ij} f|^2\,dx.$$

Expanding gives

$$\int_0^1 \left(|f'|^2 + \gamma_{ij}\overline{f'}\,f + \gamma_{ij}\overline{f}\,f' + \gamma_{ij}^2|f|^2\right)\,dx.$$

The mixed terms vanish because f has zero trace at both endpoints:

$$\int_0^1 \left(\overline{f'}\,f + \overline{f}\,f'\right)\,dx = \int_0^1 (|f|^2)'\,dx = |f(1)|^2 - |f(0)|^2 = 0.$$

This proves (10.15). If $f \in H^2([0,1],\mathbb{C}) \cap H^1_0([0,1],\mathbb{C})$, one further integration by parts gives

$$\int_0^1 \left(|f'|^2 + \gamma_{ij}^2|f|^2\right)dx = \int_0^1 \overline{f}\,(-f'' + \gamma_{ij}^2 f)\,dx,$$

which proves (10.16). $\square$

Proposition 10.17. *For* $m \geq 1$*, let*

$$s_m(x) := \sqrt{2}\sin(m\pi x), \qquad 0 \leq x \leq 1.$$

Then

$$L^D_{ij}s_m = \lambda_{ij,m}s_m, \qquad \lambda_{ij,m} := (m\pi)^2 + \gamma_{ij}^2.$$

Consequently, in the complexified E_{ij}-sector,

$$V^{\infty}_{ij,m} := \lambda_{ij,m}^{-1/2} \nabla_{b_\Sigma}(s_m E_{ij}) \tag{10.17}$$

has unit L^2-norm, and the family $\{V^{\infty}_{ij,m}\}_{m\geq 1}$ is orthonormal.

Proof. The functions s_m satisfy

$$-s_m'' = (m\pi)^2 s_m, \qquad s_m(0) = s_m(1) = 0, \qquad \|s_m\|_{L^2} = 1.$$

Hence

$$L^D_{ij} s_m = -s_m'' + \gamma_{ij}^2 s_m = \Big((m\pi)^2 + \gamma_{ij}^2\Big) s_m.$$

Proposition 10.16 gives

$$\langle \nabla_{b_\Sigma}(s_m E_{ij}), \nabla_{b_\Sigma}(s_n E_{ij}) \rangle_{L^2} = \langle s_m, L^D_{ij} s_n \rangle_{L^2} = \lambda_{ij,n} \delta_{mn}.$$

The normalization in (10.17) therefore gives an orthonormal family. □

Remark 10.18. The calculation in the proof of Proposition 10.17 is written in the complexified E_{ij}-sector. The real vertical tangent space is obtained from skew-Hermitian generators, for example

$$E_{ij} - E_{ji}, \qquad \mathrm{i}(E_{ij} + E_{ji}), \qquad \mathrm{i}E_{ii}.$$

Since the scalar operator is the same in each sector, the real vertical modes are obtained by taking the corresponding real combinations of the complexified modes.

Remark 10.19 (Relation with the finite-depth blocks)*.* For fixed i, j, the Dirichlet operator L_{ij}^D is the continuum counterpart of the block at finite depth

$$L_{ij}^{(N)} := N^2\Big((\lambda_i - \lambda_j)^2 I_{N-1} + \lambda_i\lambda_j C_{N-1}\Big), \qquad \lambda_\nu := \sigma_\nu^{1/N}.$$

The based gauge condition is the source of the Dirichlet data: infinitesimal based gauge fields vanish at $x = 0$ and $x = 1$.

We do not prove an operator convergence theorem here. Such a theorem would require explicit sampling and interpolation maps between $\mathbb{C}^{N-1}$ and functions on $[0, 1]$, a consistency estimate, and a compactness or resolvent argument. For example, if $f \in C^4([0, 1], \mathbb{C})$ satisfies $f(0) = f(1) = 0$ and

$$f_k^{(N)} := f(k/N), \qquad 1 \le k \le N - 1,$$

then the first step would be an estimate of the form

$$\max_{1\le k\le N-1} \Big|(L_{ij}^{(N)} f^{(N)})_k - (L_{ij} f)(k/N)\Big| \le \frac{C_{f,i,j}}{N}. \tag{10.18}$$

One would then combine such estimates with coercivity to prove convergence of eigenvalues, eigenmodes, or resolvents. The present section only identifies the continuum operator and the endpoint solutions that such a theorem would target.

10.4.2 Endpoint solutions and horizontal lifts

The vertical calculation fixes both endpoints. To obtain the missing endpoint directions, we use the same scalar differential expression L_{ij} but do not impose Dirichlet conditions.

For $1 \le i, j \le d$, define

$$c_{ij}^L(x) := \sigma_i^x \sigma_j^{1-x}, \qquad c_{ij}^R(x) := \sigma_i^{1-x}\sigma_j^x. \tag{10.19}$$

The endpoint values are

$$c_{ij}^{R}(0) = \sigma_i, \qquad c_{ij}^{R}(1) = \sigma_j, \qquad c_{ij}^{L}(0) = \sigma_j, \qquad c_{ij}^{L}(1) = \sigma_i. \tag{10.20}$$

Thus c_{ij}^{R} carries the ordered pair (σ_i, σ_j) from left to right, while c_{ij}^{L} carries the same pair from right to left.

Proposition 10.20. *The functions in* (10.19) *satisfy*

$$(c_{ij}^{L})' = -\gamma_{ij} c_{ij}^{L}, \qquad (c_{ij}^{R})' = \gamma_{ij} c_{ij}^{R},$$

and hence

$$L_{ij} c_{ij}^{L} = 0, \qquad L_{ij} c_{ij}^{R} = 0$$

as differential equations on $[0, 1]$*. When* $i = j$*, both functions reduce to the constant function*

$$c_{ii}^{L}(x) = c_{ii}^{R}(x) = \sigma_i.$$

Proof. For the right endpoint solution,

$$c_{ij}^{R}(x) = \exp\Big((1-x)\log\sigma_i + x\log\sigma_j\Big),$$

so

$$(c_{ij}^{R})' = \gamma_{ij} c_{ij}^{R}, \qquad (c_{ij}^{R})'' = \gamma_{ij}^{2} c_{ij}^{R}.$$

Thus

$$L_{ij} c_{ij}^{R} = -(c_{ij}^{R})'' + \gamma_{ij}^{2} c_{ij}^{R} = 0.$$

The proof for c_{ij}^{L} is the same, with $(c_{ij}^{L})' = -\gamma_{ij} c_{ij}^{L}$. $\square$

Table 10.1: Finite-depth objects and their continuum counterparts.

Object at finite depth from Chapter 5	Continuum counterpart	Boundary behavior	Role
$N^2\Big((\lambda_i - \lambda_j)^2 I_{N-1} + \lambda_i\lambda_j C_{N-1}\Big)$	L^D_{ij}	Dirichlet	vertical operator
$s_m(k)$ from (5.25)	$s_m(x) = \sqrt{2}\sin(m\pi x)$	vanishes at $x = 0, 1$	vertical eigenmode
$\widetilde{h}^{(N)}_{ij}$ from (5.33)	$L_{ij}c = 0$ before imposing endpoint data	endpoints free	enlarged solution space
$a^{(N)}_{ij}$, $a^{(0)}_{ij}$ from (5.35)	$c^R_{ij}(x)$, $c^L_{ij}(x)$	nonzero endpoint values	endpoint solutions
Γ^Σ_{ij}	$\Gamma^\infty_{ij}(x) = c^R_{ij}(x)E_{ij}$	selected by right endpoint data	horizontal lift of E_{ij}
Γ^Σ_{ii}	$\Gamma^\infty_{ii}(x) = \sigma_i E_{ii}$	constant in x	diagonal horizontal lift

Remark 10.21. The distinction between vertical directions and endpoint directions is encoded in the boundary data. Vertical coefficient fields lie in H^1_0 and vanish at both endpoints. The endpoint solutions solve the same homogeneous differential equation but need not vanish at the endpoints.

Set

$$r_i := \sigma_i^2.$$

We use the logarithmic mean ℓ_{ij} defined in (10.3).

Lemma 10.22. *For all* $1 \le i, j \le d$,

$$\int_0^1 |c^R_{ij}(x)|^2\,dx = \ell_{ij}, \qquad \int_0^1 |c^L_{ij}(x)|^2\,dx = \ell_{ij}.$$

Proof. Since

$$|c^R_{ij}(x)|^2 = \sigma_i^{2(1-x)}\sigma_j^{2x} = r_i^{1-x}r_j^x,$$

the definition (10.3) gives

$$\int_0^1 |c_{ij}^R(x)|^2 \, dx = \ell_{ij}.$$

For c_{ij}^L, the integrand is $r_i^x r_j^{1-x}$, and the change of variables $x \mapsto 1 - x$ gives the same integral. $\square$

The right endpoint solution gives the horizontal lift used in Theorem 10.5. Recall that

$$\Gamma_{ij}^\infty(x) := c_{ij}^R(x) E_{ij}.$$

Proof of Theorem 10.5. In the E_{ij}-sector, the formal adjoint of ∇_{b_Σ} is

$$\nabla_{b_\Sigma}^* = -\frac{d}{dx} + \gamma_{ij}.$$

Since $(c_{ij}^R)' = \gamma_{ij} c_{ij}^R$, we have

$$\nabla_{b_\Sigma}^* \Gamma_{ij}^\infty = \Big(-(c_{ij}^R)' + \gamma_{ij} c_{ij}^R\Big) E_{ij} = 0.$$

Let $a \in H_0^2([0,1], \mathfrak{gl}(d, \mathbb{C}))$. By the formal adjoint formula and integration by parts,

$$\int_0^1 \operatorname{tr}\Big((\Gamma_{ij}^\infty)^* \, \nabla_{b_\Sigma} a\Big) \, dx = \int_0^1 \operatorname{tr}\Big((\nabla_{b_\Sigma}^* \Gamma_{ij}^\infty)^* a\Big) \, dx = 0.$$

Thus Γ_{ij}^∞ is orthogonal to the complexified gauge directions, and hence to the real vertical space $T_{b_\Sigma}(\mathcal{K} \cdot b_\Sigma)$.

It remains to check linearized balancedness and the endpoint projection. The linearization of the moment field at b_Σ, with the sign convention of this chapter, is

$$D(\mu_\infty)_{b_\Sigma}[\alpha] = -(\alpha + \alpha^*)' + [\alpha, b_\Sigma^*] + [b_\Sigma, \alpha^*].$$

Here $b_\Sigma = b_\Sigma^*$ is diagonal and c_{ij}^R is real-valued. If $i \neq j$, then

$$D(\mu_\infty)_{b_\Sigma}[\Gamma_{ij}^\infty] = \Big(-(c_{ij}^R)' + \gamma_{ij} c_{ij}^R\Big) E_{ij} + \Big(-(c_{ij}^R)' + \gamma_{ij} c_{ij}^R\Big) E_{ji} = 0.$$

If $i = j$, then $c_{ii}^R \equiv \sigma_i$ and $\gamma_{ii} = 0$, so

$$D(\mu_\infty)_{b_\Sigma}[\Gamma_{ii}^\infty] = -2(c_{ii}^R)' E_{ii} = 0.$$

Therefore

$$\Gamma_{ij}^\infty \in T_{b_\Sigma}\mathcal{M}_\infty \cap T_{b_\Sigma}(\mathcal{K} \cdot b_\Sigma)^{\perp_{L^2}}.$$

By (9.17),

$$(D\phi_\infty)_{b_\Sigma}[\Gamma_{ij}^\infty] = Y_\Sigma(1) \int_0^1 Y_\Sigma(t)^{-1} \Gamma_{ij}^\infty(t) Y_\Sigma(t)\, dt.$$

Using $Y_\Sigma(1) = \Sigma$, $Y_\Sigma(t) = \Sigma^t$, and

$$\Sigma^{-t} E_{ij} \Sigma^t = \sigma_i^{-t} \sigma_j^t E_{ij},$$

we obtain

$$\begin{aligned}
(D\phi_\infty)_{b_\Sigma}[\Gamma_{ij}^\infty] &= \Sigma \int_0^1 c_{ij}^R(t) \sigma_i^{-t} \sigma_j^t\, E_{ij}\, dt \\
&= \int_0^1 \sigma_i^{2(1-t)} \sigma_j^{2t}\, dt\, E_{ij} \\
&= \int_0^1 r_i^{1-t} r_j^t\, dt\, E_{ij} \\
&= \ell_{ij} E_{ij}.
\end{aligned}$$

The norm identity follows from Lemma 10.22:

$$\|\Gamma_{ij}^\infty\|_{L^2}^2 = \int_0^1 |c_{ij}^R(x)|^2\, dx = \ell_{ij}.$$

Define

$$\widehat{\Gamma}_{ij}^{\infty} := \ell_{ij}^{-1/2}\Gamma_{ij}^{\infty}.$$

After normalization,

$$\|\widehat{\Gamma}_{ij}^{\infty}\|_{L^2} = 1, \qquad (D\phi_{\infty})_{b_\Sigma}[\widehat{\Gamma}_{ij}^{\infty}] = \ell_{ij}^{1/2}E_{ij}. \tag{10.21}$$

Equation (10.21) shows that each normalized lift has unit L^2-norm and projects to $\ell_{ij}^{1/2}E_{ij}$. The family is L^2-orthonormal because the matrix units are orthonormal for the Frobenius Hermitian form.

It remains to show that no horizontal directions are missing. We work after complexifying the tangent spaces; the real statement is obtained by taking the corresponding real form. Let

$$\eta \in \Big(T_{b_\Sigma}\mathcal{M}_\infty \cap T_{b_\Sigma}(\mathcal{K}\cdot b_\Sigma)^{\perp_{L^2}}\Big) \otimes_{\mathbb{R}} \mathbb{C}.$$

Orthogonality to the complexified vertical directions gives, for every $a \in H_0^2([0,1], \mathfrak{gl}(d,\mathbb{C}))$,

$$0 = \langle \eta, \nabla_{b_\Sigma} a\rangle_{L^2} = \langle \nabla_{b_\Sigma}^{*}\eta, a\rangle_{L^2}.$$

Hence $\nabla_{b_\Sigma}^{*}\eta = 0$ in the distributional sense. In the E_{pq}-sector this equation is

$$-\eta_{pq}'(x) + \gamma_{pq}\eta_{pq}(x) = 0,$$

so

$$\eta_{pq}(x) = A_{pq}\,\sigma_p^{1-x}\sigma_q^{x}$$

for constants $A_{pq} \in \mathbb{C}$. Therefore

$$\eta = \sum_{p,q} A_{pq} \Gamma^{\infty}_{pq}.$$

The normalized fields $\widehat{\Gamma}^{\infty}_{ij}$ therefore form an L^2-orthonormal basis of the complexified horizontal space.

Finally, at the diagonal point,

$$\mathcal{A}_{\infty,\Sigma}(E_{ij}) = \left(\int_0^1 r_i^{1-t} r_j^t \, dt\right) E_{ij} = \ell_{ij} E_{ij}.$$

Together with the definition of h_∞ in (10.2), this gives

$$h_\infty(\Sigma)\left(\ell_{ij}^{1/2} E_{ij}, \ell_{k\ell}^{1/2} E_{k\ell}\right) = \delta_{ik}\delta_{j\ell}.$$

Thus the normalized lifts project to the $h_\infty(\Sigma)$-orthonormal basis

$$\ell_{ij}^{1/2} E_{ij}.$$

□

Lemma 10.23. *At the diagonal balanced representative b_Σ,*

$$\ker(D\phi_\infty)_{b_\Sigma} \cap T_{b_\Sigma}\mathcal{M}_\infty = T_{b_\Sigma}(\mathcal{K} \cdot b_\Sigma).$$

Proof. The inclusion

$$T_{b_\Sigma}(\mathcal{K} \cdot b_\Sigma) \subset \ker(D\phi_\infty)_{b_\Sigma} \cap T_{b_\Sigma}\mathcal{M}_\infty$$

follows because the based gauge action preserves $\mathcal{M}_\infty$ and leaves ϕ_∞ unchanged.

For the reverse inclusion, let $\alpha \in \ker(D\phi_\infty)_{b_\Sigma} \cap T_{b_\Sigma}\mathcal{M}_\infty$. The Dirichlet operators L_{ij}^D have a spectral gap, so the vertical space is closed and the L^2-orthogonal decomposition gives

$$\alpha = v + h, \qquad v \in T_{b_\Sigma}(\mathcal{K} \cdot b_\Sigma), \qquad h \in T_{b_\Sigma}\mathcal{M}_\infty \cap T_{b_\Sigma}(\mathcal{K} \cdot b_\Sigma)^{\perp_{L^2}}.$$

By the basis statement in Theorem 10.5,

$$h = \sum_{i,j} a_{ij} \Gamma_{ij}^\infty$$

for constants a_{ij}. Since based gauge directions do not change the endpoint, $(D\phi_\infty)_{b_\Sigma} v = 0$. The assumptions on α, together with Theorem 10.5, then give

$$0 = (D\phi_\infty)_{b_\Sigma} h = \sum_{i,j} a_{ij} \ell_{ij} E_{ij}.$$

All ℓ_{ij} are positive and the matrix units are linearly independent, so $a_{ij} = 0$ for all i, j. Hence $h = 0$ and $\alpha = v$ is vertical. $\square$

10.5 The renormalized entropy

At finite depth, the entropy is the logarithm of the volume of the compact set of balanced factorizations in a fixed fibre. In the continuum model there is no finite-dimensional compact orbit whose Riemannian volume is being computed. Nevertheless, the formula at finite depth has an X-dependent part that remains after subtracting terms independent of X. This remaining part is the renormalized entropy S_∞ defined in (10.4).

In Chapter 5 we computed the entropy at depth N. If $X \in \mathrm{GL}(d, \mathbb{C})$ has singular values $\sigma_1, \ldots, \sigma_d$, then

$$
\begin{aligned}
S_N(X) = (N-1)\log u_d + \frac{d}{2}\log N + \frac{N-1}{N}\sum_{i=1}^{d}\log\sigma_i \\
+ \sum_{1\le i<j\le d}\log\left(\frac{\sigma_i^2-\sigma_j^2}{\sigma_i^{2/N}-\sigma_j^{2/N}}\right),
\end{aligned}
\tag{10.22}
$$

where u_d is the Haar volume of $\mathrm{U}(d)$ with the normalization used there.

At a regular point of the singular-value chamber, for each $i<j$,

$$
\sigma_i^{2/N}-\sigma_j^{2/N} = \frac{1}{N}\Big(\log\sigma_i^2-\log\sigma_j^2\Big) + O_X(N^{-2}).
$$

Hence

$$
\log\left(\frac{\sigma_i^2-\sigma_j^2}{\sigma_i^{2/N}-\sigma_j^{2/N}}\right) = \log N + \log\left(\frac{\sigma_i^2-\sigma_j^2}{\log\sigma_i^2-\log\sigma_j^2}\right) + O_X(N^{-1}).
$$

Since there are $\binom{d}{2}$ unordered pairs, the $\log N$-terms in (10.22) combine to give

$$
\frac{d}{2}\log N + \binom{d}{2}\log N = \frac{d^2}{2}\log N.
$$

Thus, at a regular point,

$$
S_N(X) - (N-1)\log u_d - \frac{d^2}{2}\log N = S_\infty(X) + O_X(N^{-1}).
$$

This calculation fixes the normalization of S_∞. We do not use it as a convergence theorem; in particular, no uniform estimate through singular-value collisions is asserted here.

For later use, we record the equivalent form of (10.4). If

$$r_i := \sigma_i^2,$$

and if ℓ_{ij} denotes the logarithmic mean from (10.3), then

$$S_\infty(X) = \frac{1}{2}\log\det(X^*X) + \sum_{1\le i<j\le d} \log \ell_{ij}. \tag{10.23}$$

Indeed, $\frac{1}{2}\log\det(X^*X) = \sum_i \log\sigma_i$.

The word entropy in S_∞ is inherited from the finite-depth theory. It should not be read as the logarithm of an actual volume of a $\mathcal{K}$-orbit. Instead, S_∞ is the renormalized function on $\mathrm{GL}(d,\mathbb{C})$ obtained from the finite-depth entropy after the X-independent terms have been removed.

The formula is written using singular values, but it does not require a genericity assumption. The logarithmic mean has the integral representation

$$\ell_{ij} = \int_0^1 r_i^{1-t} r_j^{t}\, dt,$$

so it extends across $r_i = r_j$ with value $\ell_{ii} = r_i$. Hence (10.4) extends continuously across singular-value collisions. The determinant identity in Theorem 10.6 upgrades this statement to global smoothness on $\mathrm{GL}(d,\mathbb{C})$.

Remark 10.24 (The formal β-family). The real and complex infinite-depth entropies fit the same β-pattern as the finite-depth formulas of Contreras and Nahas [CN26]. If $\beta \in \{1,2,4\}$ corresponds respectively to the real, complex, and quaternionic models, the formal infinite-depth renormalization gives

$$S_\infty^{(\beta)}(X) = (\beta-1)\sum_{i=1}^d \log\sigma_i + \frac{\beta}{2}\sum_{1\le i<j\le d} \log\left(\frac{\sigma_i^2-\sigma_j^2}{\log\sigma_i^2 - \log\sigma_j^2}\right),$$

after subtracting the X-independent terms, including the term coming from the Haar volume of the compact group at finite depth. This chapter proves the complex case $\beta = 2$.

Proof of Theorem 10.6. The left-hand side depends only on the singular values of X. We first reduce the right-hand side to the same diagonal calculation. Let

$$X = Q\Sigma R^*, \qquad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d),$$

be a singular value decomposition. For the complex-linear isomorphism

$$T : \mathrm{M}_d(\mathbb{C}) \to \mathrm{M}_d(\mathbb{C}), \qquad T(Z) = QZR^*,$$

the definition of $\mathcal{A}_{\infty,X}$ gives

$$\mathcal{A}_{\infty,X} \circ T = T \circ \mathcal{A}_{\infty,\Sigma}.$$

Thus $\mathcal{A}_{\infty,X}$ and $\mathcal{A}_{\infty,\Sigma}$ are conjugate as complex-linear operators on $\mathrm{M}_d(\mathbb{C})$, and so

$$\det(\mathcal{A}_{\infty,X}^{-1}) = \det(\mathcal{A}_{\infty,\Sigma}^{-1}).$$

It is therefore enough to compute at $X = \Sigma$.

Set $r_i = \sigma_i^2$. For the matrix unit E_{ij},

$$(\Sigma\Sigma^*)^{1-t} E_{ij} (\Sigma^*\Sigma)^t = r_i^{1-t} r_j^t E_{ij}.$$

Hence

$$\mathcal{A}_{\infty,\Sigma}(E_{ij}) = \left(\int_0^1 r_i^{1-t} r_j^t \, dt \right) E_{ij} = \ell_{ij} E_{ij}.$$

The matrix units form a complex basis of $\mathrm{M}_d(\mathbb{C})$, so

$$\det(\mathcal{A}_{\infty,\Sigma}^{-1}) = \prod_{i,j=1}^{d} \ell_{ij}^{-1}.$$

Since $\ell_{ii} = r_i = \sigma_i^2$ and $\ell_{ij} = \ell_{ji}$, this becomes

$$\det(\mathcal{A}_{\infty,\Sigma}^{-1}) = \left(\prod_{i=1}^{d} \sigma_i^{-2}\right)\left(\prod_{1\le i<j\le d} \ell_{ij}^{-2}\right).$$

Taking logarithms gives

$$-\frac{1}{2}\log\det(\mathcal{A}_{\infty,\Sigma}^{-1}) = \sum_{i=1}^{d}\log\sigma_i + \sum_{1\le i<j\le d}\log\ell_{ij}.$$

By (10.4), the right-hand side is $S_\infty(X)$. $\square$

Since $X \mapsto \mathcal{A}_{\infty,X}$ is smooth on $\mathrm{GL}(d,\mathbb{C})$, and since $\mathcal{A}_{\infty,X}$ is positive definite, the function $-\frac{1}{2}\log\det(\mathcal{A}_{\infty,X}^{-1})$ is smooth. The theorem therefore shows that the singular-value expression for S_∞ is not merely a continuous extension across collisions. In particular, the singular-value expression defines a smooth bi-unitarily invariant function on $\mathrm{GL}(d,\mathbb{C})$.

Remark 10.25. The proof identifies the renormalized entropy with the determinant of the metric matrix in the global Frobenius frame. Section 10.7 uses this determinant formula to compute the Ricci form.

10.6 Horizontal lifts and the quotient metric

We defined the Hermitian tensor h_∞ in Section 10.2 and constructed the horizontal endpoint modes in Section 10.4. We now show that these two constructions give the same metric on the quotient.

The horizontal-lift theorem was proved in Section 10.4. We now use those lifts to identify the quotient tensor with the explicit metric h_∞.

Proof of Theorem 10.3. Let h_{quot} denote the Hermitian tensor on $\mathrm{GL}(d,\mathbb{C})$ obtained by transporting the L^2-horizontal quotient tensor through

$$\bar{\phi}_\infty : \mathcal{M}_\infty/\mathcal{K} \to \mathrm{GL}(d,\mathbb{C}).$$

We prove that $h_{\mathrm{quot}} = h_\infty$.

First take $X = \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$. By Theorem 10.5, the vectors

$$e^\infty_{ij} := \ell_{ij}^{1/2} E_{ij}$$

are the images of an L^2-orthonormal horizontal basis. Therefore

$$h_{\mathrm{quot}}(\Sigma)(e^\infty_{ij}, e^\infty_{k\ell}) = \delta_{ik}\delta_{j\ell}.$$

Equivalently,

$$h_{\mathrm{quot}}(\Sigma)(E_{ij}, E_{k\ell}) = \delta_{ik}\delta_{j\ell}\, \ell_{ij}^{-1}.$$

Proposition 10.10 gives the same formula for $h_\infty(\Sigma)$. Hence

$$h_{\mathrm{quot}}(\Sigma) = h_\infty(\Sigma)$$

at every positive diagonal matrix Σ.

It remains to move this identity to an arbitrary point of $\mathrm{GL}(d,\mathbb{C})$. Let

$$X = Q\Sigma R^*$$

be a singular value decomposition. Choose a path $u \in H^2([0,1], \mathrm{U}(d))$ such that

$$u(0) = R, \qquad u(1) = Q.$$

If Y is a transport path from I to Σ, then

$$\widetilde{Y}(x) := u(x)Y(x)R^*$$

is a transport path from I to X. If $b = Y'Y^{-1}$, then the coefficient of $\widetilde{Y}$ is

$$\widetilde{b} = u'u^* + ubu^*.$$

A tangent field η at b is sent to $u\eta u^*$. This transformation preserves the L^2 Hermitian inner product, carries based gauge directions to based gauge directions, and therefore carries horizontal directions to horizontal directions. Thus h_{quot} is invariant under $X \mapsto QXR^*$.

The tensor h_∞ has the same bi-unitary invariance by Lemma 10.9. Since every $X \in \mathrm{GL}(d, \mathbb{C})$ admits a singular value decomposition, the equality at positive diagonal matrices implies

$$h_{\mathrm{quot}} = h_\infty$$

on all of $\mathrm{GL}(d, \mathbb{C})$. Taking real and imaginary parts gives the induced Riemannian metric and Kähler form:

$$g_\infty = \Re h_\infty, \qquad \omega_\infty = \Im h_\infty.$$

$\square$

Remark 10.26 (Scope of the quotient construction). The proof of Theorem 10.3 identifies the tensor obtained from L^2-horizontal lifts after the continuum quotient has been identified with $\mathrm{GL}(d, \mathbb{C})$. It is not a convergence theorem for the quotient tensors at finite depth.

Such a theorem would require uniform control, on compact subsets of $\mathrm{GL}(d, \mathbb{C})$, of

$$N\, \mathcal{A}_{N,X}^{-1} - \mathcal{A}_{\infty,X}^{-1}$$

together with corresponding estimates for derivatives. Convergence of geodesics would also require an argument showing that the relevant geodesics remain in a compact region on the time interval under consideration. A global Gromov–Hausdorff statement would require additional control near $\{\det X = 0\}$, such as completeness, curvature, or volume estimates. Thus the curvature formulas that follow should be read as identities for the limiting metric, not as compactness results for the quotients at finite depth.

10.7 Ricci curvature and entropy

We now use the determinant identity for S_∞ to compute the Ricci form. Throughout this section, for a Kähler metric h on a complex manifold, we use the convention

$$\rho_h = -\mathrm{i}\, \partial\bar{\partial} \log \det(h_{\alpha\bar{\beta}})$$

in any local holomorphic frame.

Proof of Theorem 10.7. Choose a basis

$$E_1, \ldots, E_{d^2}$$

of $\mathrm{M}_d(\mathbb{C})$ that is orthonormal for the Frobenius Hermitian form. We regard this basis as a global holomorphic frame of

$$T\mathrm{GL}(d,\mathbb{C}) \cong \mathrm{GL}(d,\mathbb{C}) \times \mathrm{M}_d(\mathbb{C}).$$

By (10.2), the Hermitian matrix

$$h_{\infty,\alpha\bar{\beta}}(X) := h_\infty(X)(E_\alpha, E_\beta)$$

is the matrix of the operator $\mathcal{A}_{\infty,X}^{-1}$ in this frame. Therefore

$$\det(h_{\infty,\alpha\bar{\beta}}(X)) = \det(\mathcal{A}_{\infty,X}^{-1}).$$

Using the definition of the Ricci form and Theorem 10.6, we obtain

$$\begin{aligned} \rho_\infty &= -\mathrm{i}\,\partial\bar{\partial} \log\det(h_{\infty,\alpha\bar{\beta}}) \\ &= -\mathrm{i}\,\partial\bar{\partial} \log\det(\mathcal{A}_{\infty,X}^{-1}) \\ &= 2\mathrm{i}\,\partial\bar{\partial} S_\infty. \end{aligned}$$

This proves the claimed Ricci-potential identity. □

Remark 10.27 (The pluriharmonic term). The formula (10.4) may be written as

$$S_\infty(X) = \frac{1}{2} \log\det(X^*X) + \sum_{1\le i<j\le d} \log \ell_{ij}.$$

Locally on $\mathrm{GL}(d,\mathbb{C})$,

$$\log\det(X^*X) = \log\det X + \overline{\log\det X}.$$

Hence

$$\mathrm{i}\,\partial\bar{\partial}\left(\frac{1}{2}\log\det(X^*X)\right) = 0.$$

Thus only the terms involving the logarithmic means contribute to ρ_∞.

The Ricci-potential formula also shows that the infinite-depth metric is not Einstein, even though both ω_∞ and ρ_∞ are determined by singular values.

Proof of Corollary 10.8. For a Kähler metric, the Einstein condition $\mathrm{Ric}(g_\infty) = \lambda g_\infty$ is equivalent to $\rho_\infty = \lambda\,\omega_\infty$ for some constant $\lambda \in \mathbb{R}$. Assume, for contradiction, that this identity holds.

Consider the holomorphic curve

$$\iota : \mathbb{C} \to \mathrm{GL}(d, \mathbb{C}), \qquad \iota(w) = \mathrm{diag}(e^{w/2}, e^{-w/2}, 1, \ldots, 1),$$

where the identity block is omitted when $d = 2$. Write

$$x := \Re w.$$

Then

$$\iota(w)^*\iota(w) = \mathrm{diag}(e^x, e^{-x}, 1, \ldots, 1),$$

and therefore, by Theorem 10.4,

$$(\Psi_\infty \circ \iota)(w) = \frac{1}{4}\,\mathrm{tr}\Big(\big(\log(\iota(w)^*\iota(w))\big)^2\Big) = \frac{1}{2}x^2.$$

The singular values of $\iota(w)$ are

$$e^{x/2},\ e^{-x/2},\ 1, \ldots, 1.$$

Using (10.4) and (10.3), we get

$$(S_\infty \circ \iota)(w) = \log\left(\frac{\sinh x}{x}\right) + (d-2)\log\left(\frac{4\sinh^2(x/2)}{x^2}\right),$$

with the smooth extension at $x = 0$.

The assumed Einstein condition, Theorem 10.7, and Theorem 10.4 imply

$$\mathrm{i}\,\partial\bar{\partial}\,(2S_\infty - \lambda\Psi_\infty) = 0.$$

Pulling back by ι, the function

$$2(S_\infty \circ \iota) - \lambda(\Psi_\infty \circ \iota)$$

is harmonic on $\mathbb{C}$. It depends only on $x = \Re w$, so it is affine in x. It is also even in x, hence it must be constant.

Its Taylor expansion at $x = 0$ is

$$2(S_\infty \circ \iota)(w) - \lambda(\Psi_\infty \circ \iota)(w) = \left(\frac{d}{6} - \frac{\lambda}{2}\right)x^2 - \frac{d+6}{720}x^4 + O(x^6).$$

The coefficient of x^4 is nonzero for every $d \geq 2$. The function is therefore not constant, contradicting the Einstein assumption. Hence h_∞ is not Kähler–Einstein. $\square$

Bibliographical notes

This chapter is the continuum counterpart of Chapter 5. The finite-depth quotient, metric, Kähler-potential, entropy, and Ricci computations that motivate the present construction come from Chapters 4 and 5, together with the papers cited there.

On the continuum side, the real part g_∞ of the Hermitian metric restricts to the infinite-depth metric of Cohen, Menon, and Veraszto [CMV23] on $\mathrm{GL}(d, \mathbb{R})$. The formal β-family in Remark 10.24 is motivated by the finite-depth complex and quaternionic formulas of Contreras and Nahas [CN26]. We developed the gauge-theoretic continuum model, the based gauge action, and the balanced classification in Chapter 9.

No convergence theorem is proved in this chapter. In particular, we do not prove convergence of the rescaled finite-depth quotient tensors, the corresponding off-diagonal tridiagonal blocks with Dirichlet endpoint conditions, geodesics, or Gromov–Hausdorff limits.

Bibliography

[AB83] M. F. Atiyah and R. Bott. "The Yang-Mills equations over Riemann surfaces". In: *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 308.1505 (1983), pp. 523–615. DOI: 10.1098/rsta.1983.0017.

[ACH18] Sanjeev Arora, Nadav Cohen, and Elad Hazan. "On the Optimization of Deep Networks: Implicit Acceleration by Overparameterization". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 244–253.

[AF03] Robert A. Adams and John J. F. Fournier. *Sobolev Spaces*. 2nd ed. Vol. 140. Pure and Applied Mathematics. Academic Press, 2003. ISBN: 9780120441433.

[Ati82] M. F. Atiyah. "Convexity and commuting Hamiltonians". In: *Bulletin of the London Mathematical Society* 14.1 (1982), pp. 1–15.

[Bah+22] Bubacarr Bah et al. "Learning deep linear neural networks: Riemannian gradient flows and convergence to global minimizers". In: *Information and Inference: A Journal of the IMA* 11.1 (2022), pp. 307–353. DOI: 10.1093/imaiai/iaaa039.

[Bes87] Arthur L. Besse. *Einstein Manifolds*. Berlin, Heidelberg: Springer, 1987. DOI: 10.1007/978-3-540-74311-8.

[BH89] Pierre Baldi and Kurt Hornik. "Neural networks and principal component analysis: Learning from examples without local minima". In: *Neural Networks* 2.1 (1989), pp. 53–58. DOI: 10.1016/0893-6080(89)90014-2.

[Bha07] Rajendra Bhatia. *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton, NJ: Princeton University Press, 2007. ISBN: 9780691129181.

[BK23] Anthony M. Bloch and Steven N. Karp. "Gradient Flows, Adjoint Orbits, and the Topology of Totally Nonnegative Flag Varieties". In: *Communications in Mathematical Physics* 398.3 (2023), pp. 1213–1289. DOI: 10.1007/s00220-022-04540-5.

[BMR13] Anthony M. Bloch, Philip J. Morrison, and Tudor S. Ratiu. "Gradient flows in the normal and Kähler metrics and triple bracket generated metriplectic systems". In: *Recent Trends in Dynamical Systems*. Vol. 35. Springer Proceedings in Mathematics & Statistics. Springer, 2013, pp. 371–415.

[CKM25] Alan Chen, Tejas Kotwal, and Govind Menon. "Entropic Regularization in the Deep Linear Network". In: *arXiv preprint arXiv:2512.06137* (2025).

[CMV23] Nadav Cohen, Govind Menon, and Zsolt Veraszto. "Deep linear networks for matrix completion—an infinite depth limit". In: *SIAM Journal on Applied Dynamical Systems* 22.4 (2023), pp. 3208–3232.

[CN26] Luis Contreras and Marco Nahas. *Complex and Quaternion Deep Linear Networks: Dynamics and Entropy*. Preprint. 2026.

[Des+05] Mathieu Desbrun et al. *Discrete Exterior Calculus*. 2005. arXiv: math/0508341 [math.DG]. URL: https://arxiv.org/abs/math/0508341.

[Don84] Simon K. Donaldson. “Nahm’s equations and the classification of monopoles”. In: *Communications in Mathematical Physics* 96.3 (1984), pp. 387–407. DOI: 10.1007/BF01214583.

[FU91] Daniel S. Freed and Karen K. Uhlenbeck. *Instantons and Four-Manifolds*. 2nd ed. Vol. 1. Mathematical Sciences Research Institute Publications. New York: Springer-Verlag, 1991.

[G25] Asvin G. *Computational Platonism*. 2025. arXiv: 2506.19787 [math.HO]. URL: https://arxiv.org/abs/2506.19787.

[G26] Asvin G. *On the mechanical creation of mathematical concepts*. 2026. arXiv: 2507.10179 [math.HO]. URL: https://arxiv.org/abs/2507.10179.

[Gow23] Timothy Gowers. “What makes mathematicians believe unproved mathematical statements”. In: *Annals of Mathematics and Philosophy* 1.1 (2023), pp. 57–110.

[GS84] Victor Guillemin and Shlomo Sternberg. *Symplectic Techniques in Physics*. Cambridge: Cambridge University Press, 1984.

[Gül96] Osman Güler. “Barrier functions in interior point methods”. In: *Mathematics of Operations Research* 21.4 (1996), pp. 860–885.

[Hen25] Jonas Henkel. “The mathematician’s assistant: integrating AI into research practice”. In: *Mathematische Semesterberichte* 72.2 (2025), pp. 117–144.

[Hil14] Roland Hildebrand. “Canonical barriers on convex cones”. In: *Mathematics of operations research* 39.3 (2014), pp. 841–850.

[HIM23] Ching-Peng Huang, Dominik Inauen, and Govind Menon. “Motion by mean curvature and Dyson Brownian motion”. In: *Electronic Communications in Probability* 28 (2023), pp. 1–10.

[Hir03] Anil N. Hirani. “Discrete Exterior Calculus”. PhD thesis. California Institute of Technology, 2003. DOI: 10.7907/ZHY8-V329.

[Hsu02] Elton P. Hsu. *Stochastic Analysis on Manifolds*. Vol. 38. Graduate Studies in Mathematics. Providence, RI: American Mathematical Society, 2002. ISBN: 9780821808023.

[JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. “The Variational Formulation of the Fokker–Planck Equation”. In: *SIAM Journal on Mathematical Analysis* 29.1 (1998), pp. 1–17. DOI: 10.1137/S0036141096303359.

[Kat13] Tosio Kato. *Perturbation Theory for Linear Operators*. Vol. 132. Springer Science & Business Media, 2013.

[Kir84] Frances Clare Kirwan. *Cohomology of Quotients in Symplectic and Algebraic Geometry*. Vol. 31. Mathematical Notes. Princeton, NJ: Princeton University Press, 1984. ISBN: 0691083703.

[KN63] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of Differential Geometry. Vol. I*. New York: Interscience Publishers, 1963.

[KN79] George Kempf and Linda Ness. “The length of vectors in representation spaces”. In: *Algebraic Geometry*. Ed. by Knud Lønsted. Vol. 732. Lecture Notes in Mathematics. Berlin, Heidelberg: Springer, 1979, pp. 233–243. DOI: 10.1007/BFb0066647.

[Lak63] Imre Lakatos. “Proofs and refutations (I)”. In: *The British Journal for the Philosophy of science* 14.53 (1963), pp. 1–25.

[Lee18] John M. Lee. *Introduction to Riemannian Manifolds*. Vol. 176. Graduate Texts in Mathematics. Springer, 2018. DOI: 10.1007/978-3-319-91755-9.

[Ler05] Eugene Lerman. "Gradient Flow of the Norm Squared of a Moment Map". In: *Enseignement Mathématique* 51.1-2 (2005), pp. 117–127.

[LM25] Kathryn Lindsey and Govind Menon. "Regularization Implies balancedness in the deep linear network". In: *arXiv preprint arXiv:2511.01137* (2025). DOI: 10.48550/arXiv.2511.01137.

[LR24] Simon Pepin Lehalleur and Richárd Rimányi. "Geometry of fibers of the multiplication map of deep linear neural networks". In: *arXiv preprint arXiv:2411.19920* (2024).

[Men25] Govind Menon. "The Geometry of the Deep Linear Network". In: *XIV Symposium on Probability and Stochastic Processes*. Ed. by Carmen Geraldi Higuera Chan et al. Vol. 81. Progress in Probability. Cham: Birkhäuser, 2025, pp. 1–47. DOI: 10.1007/978-3-031-96118-2_1.

[MFK94] David Mumford, John Fogarty, and Frances Kirwan. *Geometric Invariant Theory*. 3rd ed. Berlin: Springer, 1994.

[MS17] Dusa McDuff and Dietmar Salamon. *Introduction to symplectic topology*. Oxford University Press, 2017.

[MW74] Jerrold Marsden and Alan Weinstein. "Reduction of symplectic manifolds with symmetry". In: *Reports on mathematical physics* 5.1 (1974), pp. 121–130.

[MY26a] Govind Menon and Tianmin Yu. *A Riemannian Langevin equation for the deep linear network*. Preprint, forthcoming. 2026.

[MY26b] Govind Menon and Tianmin Yu. "An entropy formula for the Deep Linear Network". In: *arXiv preprint arXiv:2509.09088* (2026). DOI: 10.48550/arXiv.2509.09088.

[Nes84] Linda Ness. “A Stratification of the Null Cone via the Moment Map”. In: *American Journal of Mathematics* 106.6 (1984). With an appendix by David Mumford, pp. 1281–1329. DOI: 10.2307/2374395.

[NN94] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

[NT97] Yu E Nesterov and Michael J Todd. “Self-scaled barriers and interior-point methods for convex programming”. In: *Mathematics of Operations research* 22.1 (1997), pp. 1–42.

[Olv+10] Frank W. J. Olver et al., eds. *NIST Handbook of Mathematical Functions*. Cambridge: Cambridge University Press, 2010.

[Ott01] Felix Otto. “The Geometry of Dissipative Evolution Equations: The Porous Medium Equation”. In: *Communications in Partial Differential Equations* 26.1–2 (2001), pp. 101–174. DOI: 10.1081/PDE-100002243.

[Pan25] Markus Pantsar. “The need for ethical guidelines in mathematical research in the time of generative AI”. In: *AI and Ethics* 5.4 (2025), pp. 3657–3668.

[SMG14] Andrew M. Saxe, James L. McClelland, and Surya Ganguli. “Exact solutions to the nonlinear dynamics of learning in deep linear neural networks”. In: *International Conference on Learning Representations (ICLR)*. Also available as arXiv:1312.6120. 2014.

[Sun17] Song Sun. “Ricci curvature in Kähler geometry”. In: *International Journal of Mathematics* 28.09 (2017), p. 1740009.

[Thu06] William P Thurston. “On proof and progress in mathematics”. In: *18 Unconventional essays on the nature of mathematics*. Springer, 2006, pp. 37–55.

[Tra17] Samuel Trautwein. “A Survey of the GIT Picture for the Yang–Mills Equation over Riemann Surfaces”. In: *L'Enseignement Mathématique* 63.1-2 (2017), pp. 63–153. DOI: 10.4171/LEM/63-1/2-3.

[Uhl26] Karen Uhlenbeck. *A Sampling of Minimization Problems*. Abel at IAS lecture. Institute for Advanced Study. Jan. 28, 2026. URL: https://www.youtube.com/watch?v=rLCwqR6NWmM.

[Woo10] Christopher T. Woodward. “Moment Maps and Geometric Invariant Theory”. In: *Les cours du CIRM* 1.1 (2010), pp. 55–98. DOI: 10.5802/ccirm.4.