International Workshop Problems of Modern Mathematical Physics (PMMP'26)

Geometric Optimal Control Theory

T. Vetsov

Department of Theoretical Physics, Sofia University, Bulgaria

BLTP, JINR, Dubna, Russia

16-20 February 2026

Based on:

- M. Radomirov, R. C. Rashkov, G. S. Stoilov and T. Vetsov, Optimal Control Theory of the (2+1)-Dimensional BTZ Black Hole, [arXiv:2512.06931 [gr-qc]].
- V. Avramov, H. Dimov, M. Radomirov, R. C. Rashkov, and T. Vetsov, On Complexity and Supersymmetry, Phys. Part. Nucl. Lett., Vol. 22, No. 6 (December), pp. 1369–1373., (2025).
- V. Avramov, H. Dimov, M. Radomirov, R. C. Rashkov and T. Vetsov, Black Holes and Thermogeometric Optimization, Eur. Phys. J. C, vol: 85, issue: 587, (2025).
- V. Avramov, H. Dimov, M. Radomirov, R. C. Rashkov, T. Vetsov, On Thermodynamic Stability of Black Holes. Part I: Classical Stability, Eur. Phys. J. C 84, no.3, 281 (2024).
- H. Dimov, I. N. Iliev, M. Radomirov, R. C. Rashkov and T. Vetsov, Holographic Fisher information metric in Schrodinger spacetime, Eur. Phys. J. Plus 136, no. 11, 1128 (2021).

Outline

1. Classical Information Geometry
2. Quantum Information Geometry
3. Thermodynamic Information Geometry
4. Geometric Optimization (Optimal Control)

The Shortest Path to Optimality

System →

Parameter Space →

Information Metric →

Optimal Transformations Between States

What is the most efficient way to change our model?

What is Information Space?

- In physics, we often deal with a **Parameter Space** Θ.
- **Information Space** is the transformation of Θ into a Riemannian manifold $\mathcal{S}$.
- Each point is a probability distribution $p(x|\boldsymbol{\theta})$.
- The metric is not Euclidean; it is defined by **Statistical Distinguishability**.

The Statistical Manifold

Define a family of distributions:

$$\mathcal{S} = \left\{ p(x; \boldsymbol{\theta}) \mid \boldsymbol{\theta} = (\theta^1, \ldots, \theta^n) \in \mathbb{R}^n \right\}$$

Key Insight

The "distance" between two points $\boldsymbol{\theta}$ and $\boldsymbol{\theta} + d\boldsymbol{\theta}$ reflects how much the underlying probability distribution changes, not just the change in parameter values.

The Fisher Information Metric

The "**ruler**" of Information Space is the **Fisher Information Metric**:

$$g_{ij}(\boldsymbol{\theta}) = \int p(x;\boldsymbol{\theta}) \frac{\partial \ell(x;\boldsymbol{\theta})}{\partial \theta^i} \frac{\partial \ell(x;\boldsymbol{\theta})}{\partial \theta^j} dx,$$

where $\ell = \ln p$ is the **log-likelihood**.

• **Probability distribution** treats $\boldsymbol{\theta}$ as fixed and x varies. It asks: Given these fixed parameters ($\boldsymbol{\theta}$), what is the probability of observing different outcomes (x or data)?

• **Likelihood** treats the data x as fixed and varies the parameters $\boldsymbol{\theta}$. It asks: "Of all the points in my Information Space, which one is most likely to have generated this data?", i.e. "Given this result, which parameter is the best fit?"

Geometry and Distance

The distance in Information Space is the **Fisher-Rao Distance**:

$$ds^2 = \sum_{i,j} g_{ij}(\boldsymbol{\theta}) d\theta^i d\theta^j .$$

- **Large distance**: The distributions are distinct: they are easily distinguishable.
- **Small distance**: The two distributions overlap significantly: they are hard to tell apart.

For instance, change the mean (μ) and the standard deviation (σ) of a normal distribution slightly, the shape of the curve changes.
Fisher-Rao distance looks at how much the "information" or the "shape" of the distribution changes.

Relation to KL Divergence

Kullback-Leibler (KL) Divergence measures how much information is lost when we use one probability distribution (q) to approximate the "true" underlying distribution (p):

$$D_{KL}(p \parallel q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx, \quad D_{KL}(p\|q) \neq D_{KL}(q\|p).$$

- **Forward KL ($D_{KL}(p\|q)$):** It forces q to cover the entire support of p (it is "mean seeking", e.g. "**Don't miss anything**").
- **Reverse KL ($D_{KL}(q\|p)$):** It forces q to fit into a local mode (peak) on p (it is "mode seeking", e.g. "**Don't be wrong**").

Taylor Expansion and Relation to Fisher Metric

For two PDFs separated by $d\boldsymbol{\theta}$, KL is locally the Fisher Metric:

$$D_{KL}(p_{\boldsymbol{\theta}}\|p_{\boldsymbol{\theta}+d\boldsymbol{\theta}}) \approx \frac{1}{2} g_{ij}(\boldsymbol{\theta}) d\theta^i d\theta^j \quad \Rightarrow \quad g_{ij}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta'_i \partial\theta'_j} D_{KL}(p_{\boldsymbol{\theta}}\|p_{\boldsymbol{\theta}'})\Big|_{\boldsymbol{\theta}'=\boldsymbol{\theta}}.$$

Thus, the Fisher Information Metric g_{ij} is the local Hessian (the curvature) of the KL Divergence.

Shannon Entropy

Shannon Entropy $H(p)$ quantifies the **average uncertainty** or **information content** associated with a probability distribution $p(x)$ defining the state of the system.

$$H[p] = -\int_X p(x) \ln p(x)\, dx.$$

- **High Entropy:** The distribution is spread out (e.g., a flat line). We have little knowledge about the specific outcome of x.
- **Low Entropy:** The distribution is sharply peaked (e.g., a Dirac delta). We can predict the outcome with high confidence.
- **In the Manifold**: Entropy is a scalar field defined over the information space. Each point θ on your manifold has an associated entropy value $H(\theta)$.

Physical Intuition

In statistical mechanics, this is equivalent to the **Gibbs Entropy**. It counts the number of microscopic configurations consistent with a macroscopic parameters $\boldsymbol{\theta}$.

Comparison: Entropy vs. KL Divergence

Feature	**Shannon Entropy**	**KL Divergence**
Scope	Single distribution p	Two distributions p and q
Analogy	**Volume** of a point (the spread of a distribution)	**Distance** between points
Minimum	0 (for discrete delta)	0 (when $p = q$)
Geometric Role	Scalar field on $\mathcal{S}$	Generates the metric g_{ij}
Question	How much do I not know?	How wrong is my model?

The Connection

The KL Divergence is the *excess* entropy produced when we assume the system is in state q when it is actually in state p.

From Classical To Quantum Metric

Classical Fisher Information Metric

Fisher Information Metric (g_{ij}) defines the geometry of the manifold of probability distributions. It measures the sensitivity of a probability distribution $P(x|\theta)$ to changes in parameters θ. It quantifies how much information an observation carries about an unknown parameter (**Statistical Distinguishability**).

Quantum Fisher Information Metric

In QM, density matrices ρ replace probability distributions. The **Bures Metric** measures distances between quantum states ρ and σ:

$$d_B(\rho, \sigma) = \sqrt{2(1 - \sqrt{F(\rho, \sigma)})},$$

where F is the **Uhlmann Fidelity**. **The Link**: If we perturb a state ρ by $d\theta$, the distance is proportional to the **Quantum Fisher Information Metric** $\mathcal{F}_{ij}$:

$$d_B^2(\rho, \rho + d\rho) = \frac{1}{4} \sum_{i,j} \mathcal{F}_{ij} d\theta^i d\theta^j .$$

Classical vs. Quantum Information Geometry

Scale	Classical (p, q)	Quantum (ρ, σ)	Correspondence
Overlap	$BC = \int \sqrt{pq}dx$	$\sqrt{F} = \mathrm{Tr}\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}$	Stat. Overlap
Metric	$d_H^2 = 1 - BC$	$d_B^2 = 2(1 - \sqrt{F})$	Distinguishability
Divergence	$D_{KL} = \int p \ln \frac{p}{q} dx$	$S(\rho\|\|\sigma) = \mathrm{Tr}(\rho \ln \rho - \rho \ln \sigma)$	Info. Gain
Local	**Fisher** (g_{ij})	**Quantum Fisher** $(\mathcal{F}_{ij})$	Precision Limit

All measures converge to the Fisher Information Metric locally. If we let $q = p + dp$ and $\sigma = \rho + d\rho$, the Taylor expansions show:

- **Classical convergence (Fisher-Rao)**:

$$D_{KL}(p\|p + dp) \approx 4 \cdot d_H^2(p, p + dp) \approx \frac{1}{2} \sum g_{ij} d\theta^i d\theta^j .$$

- **Quantum convergence (Quantum Fisher)**:

$$S(\rho\|\rho + d\rho) \approx 2 \cdot d_B^2(\rho, \rho + d\rho) \approx \frac{1}{2} \sum \mathcal{F}_{ij} d\theta^i d\theta^j .$$

Takeaway

The Fisher metric is the Hessian of all standard information divergences.

* $BC = \langle \sqrt{p}, \sqrt{q} \rangle$ – Bhattacharyya Coefficient: Similarity (1 = identical PDFs)
* $d_B = -\ln BC$ is the Bhattacharyya Distance: Divergence (Not a metric)
* $d_H = \sqrt{1 - BC}$ is the Hellinger Distance: Metric Distance

Thermodynamic Connection: Ruppeiner Metric

• The **Ruppeiner Metric** measures distances between equilibrium states via entropy:

$$g^R_{ij} = -\partial_i \partial_j S,$$

• It is equivalent to the **Fisher Metric** when the "probability distribution" is the **Boltzmann distribution** (Gibbs state):

$$p(x|\theta) \propto e^{-\beta H}.$$

• The "**distinguishability**" of two thermal states is exactly the same as the "**fluctuation**" probability between them.

• **Quantum Link:** For quantum system at thermal equilibrium, described by the thermal density matrix:

$$\rho(\beta) = \frac{e^{-\beta \hat{H}}}{Z(\beta)},$$

the **Quantum Fisher Metric** (or the Bures Metric) for $\rho(\beta)$, reproduces the **thermodynamic metric**.

The Final Ingredient: The Length Functional $\mathcal{L}$

In information geometry, the **Length Functional** $\mathcal{L}$ represents the accumulated distinguishability along a path γ in the parameter space. Depending on whether the model is classical, quantum, or thermal, this integral provides a physical "**cost**" or "**limit** on a process."

$$\mathcal{L} = \int_\gamma \sqrt{g_{ij} d\theta^i d\theta^j} = \int_0^\tau \sqrt{g_{ij} \frac{d\theta^i}{dt} \frac{d\theta^j}{dt}}\, dt.$$

- **Information Content:** Quantifies how many "meaningfully different" states exist along γ.
- **Geodesics:** The path of least information change

  $\rightarrow$ The most efficient way to change our model

  $\rightarrow$ The core of **Geometric Optimization**.

$$\ddot{\theta}^k(t) + \Gamma^k_{ij}(g)\dot{\theta}^i(t)\dot{\theta}^j(t) = 0.$$

The Classical Length $\mathcal{L}$

In the classical manifold of probability distributions $p(x|\theta)$, the length $\mathcal{L}$ is the **Fisher-Rao distance**.

- It quantifies the **total number of "distinguishable" steps** between two distributions.
- If we are performing a statistical experiment, $\mathcal{L}$ represents the **total "amount of information"** gained as we move along the path.
- The **geodesic paths** minimizing $\mathcal{L}$ are the **most efficient way** to change the model parameters while maintaining the maximum possible statistical similarity between intermediate steps.

Statistical and Quantum Speed Limits

In the space of density matrices $\rho(\theta)$, using the Bures metric (or Quantum Fisher Metric), $\mathcal{L}$ is the **Bures Length**.

- It provides a **fundamental limit on Quantum Speed**: The length $\mathcal{L}$ is related to the time τ required to evolve from one state to another.
- According to the Anandan-Aharonov relation, the **time-energy uncertainty** is bounded by this geometric length:

$$\Delta E \cdot \tau \geq \hbar \mathcal{L}.$$

- **Geometric Phase**: For specific cyclic paths, this length functional is closely related to the **Berry phase** and the study of holonomies in quantum manifolds.
- Relation to **Nielsen Geometric Complexity**.

Thermodynamic Efficiency

In Ruppeiner geometry, $\mathcal{L}$ is the **Thermodynamic Length**.

- It measures the **dissipation** or **entropy production** in finite-time thermodynamics.
- **Minimum Dissipation**: If a system transitions from one state to another in a finite time τ, the excess work (dissipation) is bounded by the square of the thermodynamic length:

$$W_{ex} \gtrsim \mathcal{L}^2.$$

- A process that follows a **geodesic** (the path that minimizes $\mathcal{L}$) is **the most energy-efficient protocol** for a heat engine or a cooling cycle.
- **Practical Application:** Designing "**Optimal Protocols**" by following geodesics to minimize entropy production.

Physical Applications

• **The first information-theoretic confirmation of the holographic principle for systems with non-relativistic symmetries**:

H. Dimov, I. N. Iliev, M. Radomirov, R. C. Rashkov and T. Vetsov, Holographic Fisher information metric in Schrodinger spacetime, Eur. Phys. J. Plus 136, no. 11, 1128 (2021).

• **The first geometric optimal control theory for black holes**:

V. Avramov, H. Dimov, M. Radomirov, R. C. Rashkov and T. Vetsov, Black Holes and Thermogeometric Optimization, Eur. Phys. J. C, vol: 85, issue: 587, (2025).

M. Radomirov, R. C. Rashkov, G. S. Stoilov and T. Vetsov, Optimal Control Theory of the (2+1)-Dimensional BTZ Black Hole, [arXiv:2512.06931 [gr-qc]].

• **The first information-theoretic example that supersymmetry is hard to produce**:

V. Avramov, H. Dimov, M. Radomirov, R. C. Rashkov, and T. Vetsov, On Complexity and Supersymmetry, Physics of Particles and Nuclei Letters, Vol. 22, No. 6, pp. 1369–1373., (2025).

Summary and Open Problems

- Describe Non-Equilibrium Physics
- Noise-To-Signal-Ratio Reduction in Quantum Control Theory
- Optimal Accretion/Evaporation/Fluctuation in Black Hole Physics
- Definition of Complexity in Quantum and Holographic Systems
- Diagnostic of Chaos in Quantum Systems
- ...

Thank You!

- Special Gratitude to the PMMP'26 Organizers!
- Program "JINR-Bulgaria" of the Bulgarian Nuclear Regulatory Agency.
- Bulgarian NSF Grant KP-06-N88/3.