

The Geometry of Risk and Inference

Abstract

We integrate portfolio theory and statistical inference into a unified Riemannian geometric framework. In the cross-sectional setting regarding assets, the covariance matrix acts as the metric tensor, raising the distinction between portfolio “vectors” and alpha “covectors”. The Transfer Coefficient (TC) appears naturally as the cosine of the angle between the implemented portfolio and the ideal skill vector. In the time-series setting regarding estimation, the Delta method is revealed as the pushforward of the covariance tensor from the manifold of raw moments to the manifold of derived statistics. We derive the asymptotic variance of the Sharpe ratio difference, identifying the Jobson–Korkie result as a scalar contraction.¹

1 Introduction

Financial mathematics is often obscured by a forest of indices and sums. By adopting the language of differential geometry, we reveal the simple structures underneath. This note establishes two geometric settings:

1. **The Manifold of Assets:** Where portfolios live. Here, risk defines distances and angles.
2. **The Manifold of Statistics:** Where estimators live. Here, the Central Limit Theorem defines the local tangent structure.

Why Geometry? This approach offers three distinct advantages:

1. **Unified Intuition:** Both domains are governed by a *metric tensor* representing “cost”. In asset space, Σ measures the cost of risk; in statistical space, Γ measures the cost of uncertainty.
2. **Coordinate Invariance:** Financial quantities (Sharpe, PnL) are scalars and must not depend on arbitrary choices of basis (e.g., dollars vs. cents, central vs. raw moments). Geometry enforces this invariance by construction.
3. **Computational Power:** Complex derivations, such as the asymptotic variance of the Sharpe ratio, reduce from pages of algebra to simple tensor contractions.

2 Part I: The Riemannian Geometry of Assets

Consider the space of N assets. A portfolio is specified by its allocation vector w .

¹This note is inspired by Giuseppe Paleologo’s “IC, TC and All That” (<https://byfire.substack.com/p/information-coefficient-transfer>). We recast the core ideas using the language of differential geometry to highlight the underlying structure.

2.1 Vectors, Duals, and the Metric

We must rigorously distinguish between two types of objects in the coordinate basis:

- **Allocations are Contravariant Vectors (w^i):** They represent the state of the system (quantities held, e.g., number of shares). They have Volume units [V]. Under a change of basis (e.g., stock split $S' = \lambda S$), the components transform inversely to the basis: $w' = (1/\lambda)w$.
- **Alphas are Covariant Covectors (α_i):** They represent the gradient of expected return (force acting on the system). They have units of Pressure [P] so that the product $\alpha_i w^i$ is money (or unitless return). Under a change of basis, they transform covariantly with the basis: $\alpha' = \lambda \alpha$.

The expected PnL is the contraction $\mathbb{E}[\text{PnL}] = \alpha_i w^i$. To discuss risk, we introduce the covariance metric $g_{ij} = \Sigma_{ij}$.²

2.2 The Risk Frame (Vielbein)

The metric g_{ij} is generally not identity (assets are correlated). To reveal the simple geometry, we move to an **Orthonormal Frame** (or “Risk Frame”) where risk is uniform. We introduce the square-root matrix (Vielbein) $e^{\hat{a}}_i$ such that $\Sigma = e^T e$. In index notation, this decomposition expresses the covariance matrix as:

$$\Sigma_{ij} = \delta_{\hat{a}\hat{b}} e^{\hat{a}}_i e^{\hat{b}}_j. \quad (1)$$

In this flat frame, we define our primary geometric objects:

1. **The Risk Portfolio (p):** The allocation vector rescaled by risk.

$$p^{\hat{a}} = e^{\hat{a}}_i w^i \quad (\approx w \cdot \sigma). \quad (2)$$

Its length is the total volatility: $\|p\| = \sqrt{\delta_{\hat{a}\hat{b}} p^{\hat{a}} p^{\hat{b}}} = \sqrt{w^T \Sigma w} = \text{Risk}$.

2. **The Skill Vector (s):** The alpha covector rescaled by inverse risk (whitened).³ We obtain s by raising the index of α with the metric, then mapping to the frame:

$$s^{\hat{a}} = e^{\hat{a}}_k s^k = e^{\hat{a}}_k g^{ki} \alpha_i. \quad (3)$$

Its length is the **Information Coefficient (IC)**:

$$\text{IC} := \|s\| = \sqrt{\delta_{\hat{a}\hat{b}} s^{\hat{a}} s^{\hat{b}}} = \sqrt{\alpha_i g^{ij} \alpha_j}. \quad (4)$$

²**Thermodynamic Analogy:** A useful physical parallel is Thermodynamics. Allocations w are **Extensive** variables (they scale with system size, like Volume V). Alphas α are **Intensive** variables (they do not scale, like Pressure P). Their pairing PnL (like Energy) is naturally a scalar. The “density” interpretation arises because we often normalize w (setting $\sum w_i = 1$) to remove the extensive scale freedom.

³Computations are often simpler in this orthonormal frame. Here \hat{a}, \hat{b} are frame indices (flat/hatted) and i, j are coordinate indices (curved). The Sharpe ratio becomes the standard Euclidean dot product projection.

2.3 The Sharpe Ratio and Transfer Coefficient

In the Risk Frame, the metric is simply $\delta_{\hat{a}\hat{b}}$ (Euclidean). The Sharpe ratio is the projection of skill onto the implemented portfolio:

$$\eta(w) = \frac{\langle s, p \rangle}{\|p\|}. \quad (5)$$

This is maximized when p aligns with s . The **Transfer Coefficient (TC)** is the cosine of the angle θ between them:

$$\text{TC} := \cos(\theta) = \frac{\langle s, p \rangle}{\|s\| \|p\|}. \quad (6)$$

Thus we recover the fundamental law: Sharpe = IC \times TC.

Remark 1 (Scale Invariance and Projective Space). *The Sharpe ratio is invariant under the scaling of the portfolio vector $w \rightarrow \lambda w$ for $\lambda > 0$ (leverage). This scale invariance (homogeneity of degree 0) implies that the Sharpe ratio is not a function on the vector space of portfolios, but rather on the **Projective Space** (or the Sphere of Rays). Geometrically, maximizing the Sharpe ratio is equivalent to finding a point on the sphere S^{N-1} that minimizes the geodesic distance to the skill vector s .*

Lemma 2 (Orthogonal decomposition of Risk). *Decompose the risk portfolio p into a component aligned with skill and an orthogonal component: $p = p_{\parallel} + p_{\perp}$. Then:*

$$\frac{\|p_{\perp}\|^2}{\|p\|^2} = 1 - \text{TC}^2. \quad (7)$$

The orthogonal component p_{\perp} corresponds to variance consumed by positions that are uncorrelated with alpha. It is “wasted risk”.

2.4 Constrained Optimization: A Worked Example

In practice, portfolios face constraints (e.g., dollar-neutrality, sector limits). Geometrically, constraints define a **submanifold** $\mathcal{C} \subset \mathbb{R}^N$ on which the portfolio must live.

Setup: Consider a two-asset world with holdings (w^1, w^2) . Let $\alpha = (\alpha_1, \alpha_2)$ be the alphas and Σ the covariance matrix. The unconstrained optimal portfolio aligns with the skill vector $s^{\hat{a}} \propto (\Sigma^{-1}\alpha)^i$.

Constraint: Impose dollar-neutrality: $w^1 + w^2 = 0$. This defines a one-dimensional submanifold (a line through the origin).

Geometric Effect: In the Risk Frame, the constraint becomes a hyperplane (here, a line). The constrained optimum is the *projection* of the skill vector s onto this hyperplane. Let $n^{\hat{a}}$ be the unit normal to the constraint surface. The constrained skill vector is:

$$s_{\text{constrained}}^{\hat{a}} = s^{\hat{a}} - (s \cdot n)n^{\hat{a}}. \quad (8)$$

The Transfer Coefficient under the constraint becomes:

$$\text{TC}_{\text{constrained}} = \frac{\langle s_{\text{constrained}}, p \rangle}{\|s_{\text{constrained}}\| \|p\|}. \quad (9)$$

Since $\|s_{\text{constrained}}\| \leq \|s\|$, the constraint *reduces achievable IC*. Meanwhile, the constrained portfolio can now achieve $\text{TC} = 1$ (perfect alignment with $s_{\text{constrained}}$), but with a diminished target.

Example: Let $\Sigma = I$ (uncorrelated, unit variance), and $\alpha = (1, 0.5)$. The unconstrained skill vector is $s = \alpha = (1, 0.5)$ with $\|s\| = \sqrt{1.25} \approx 1.12$.

The dollar-neutral constraint $w^1 + w^2 = 0$ has normal $n = \frac{1}{\sqrt{2}}(1, 1)$. Projecting:

$$s_{\text{constrained}} = (1, 0.5) - \frac{1.5}{2}(1, 1) = (0.25, -0.25). \quad (10)$$

The constrained IC is $\|s_{\text{constrained}}\| = \sqrt{0.125} \approx 0.35$, a 69% reduction. The optimal constrained portfolio is $w \propto (1, -1)$, which aligns perfectly with $s_{\text{constrained}}$.

Remark 3 (Multiple Constraints). *With k linear constraints, the feasible set is an $(N - k)$ -dimensional affine subspace. The projection becomes orthogonal projection onto this subspace. Non-linear constraints (e.g., box constraints $|w^i| \leq 1$) yield curved submanifolds, and the projection becomes a geodesic projection—generally requiring numerical methods.*

3 Part II: The Geometry of Statistical Inference

We now shift focus from the geometry of assets to the geometry of probability distributions.

3.1 The Manifold of Moments

Let our statistical manifold \mathcal{Z} be coordinatized by the raw moments z^μ (e.g., sample means, sample second moments). The Central Limit Theorem (CLT) governs the fluctuations of these moments. It states that the fluctuations $\delta z^\mu = \hat{z}^\mu - z^\mu$ live in the tangent space $T_z \mathcal{Z}$ and are distributed as a multivariate Gaussian with covariance tensor $\Gamma^{\mu\nu}$:

$$\sqrt{T}\delta z \sim \mathcal{N}(0, \Gamma). \quad (11)$$

$\Gamma^{\mu\nu} = \mathbb{E}[\delta z^\mu \delta z^\nu]$ is a $(2, 0)$ tensor field on \mathcal{Z} .

3.2 The Delta Method as a Pushforward

We are interested in derived parameters θ^α (e.g., variances, Sharpe ratios) which are functions of z . Let $\phi : \mathcal{Z} \rightarrow \mathcal{P}$ be the map from moments to parameters. The linear map deriving the fluctuations in parameters ($\delta\theta$) from fluctuations in moments (δz) is the **Pushforward Map** (or Jacobian) $E : T_z \mathcal{Z} \rightarrow T_{\theta} \mathcal{P}$. Its components are:

$$E^\alpha{}_\mu = \frac{\partial \theta^\alpha}{\partial z^\mu}. \quad (12)$$

The covariance tensor on the parameter manifold \mathcal{P} is the pushforward of Γ :

$$\Omega^{\alpha\beta} = E^\alpha{}_\mu E^\beta{}_\nu \Gamma^{\mu\nu}. \quad (13)$$

In compact notation: $\Omega = \phi_* \Gamma$. This corresponds to the standard transformation rule for $(2, 0)$ tensors under a change of coordinates.

4 Part III: Application to Sharpe Difference

We apply this machinery to find the variance of the difference between two Sharpe ratios, $\Delta\eta = \eta_1 - \eta_2$.

4.1 Coordinates and Tensor Components

1. **Raw Coordinates** (z^μ): (μ_1, μ_2, q_1, q_2) where $q_i = \mathbb{E}[r_i^2]$.
2. **Parameter Coordinates** (θ^α): (μ_1, μ_2, v_1, v_2) where $v_i = \sigma_i^2 = q_i - \mu_i^2$.

The Jacobian E^α_μ transforming δz to $\delta\theta$ is:

$$E = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2\mu_1 & 0 & 1 & 0 \\ 0 & -2\mu_2 & 0 & 1 \end{pmatrix}. \quad (14)$$

The Covariance Tensor Ω in parameter space, assuming Gaussian returns (Basu's theorem implies independence of sample mean and variance), is block diagonal:

$$\Omega = \begin{pmatrix} v_1 & \rho\sqrt{v_1 v_2} & 0 & 0 \\ \rho\sqrt{v_1 v_2} & v_2 & 0 & 0 \\ 0 & 0 & 2v_1^2 & 2\rho^2 v_1 v_2 \\ 0 & 0 & 2\rho^2 v_1 v_2 & 2v_2^2 \end{pmatrix}. \quad (15)$$

4.2 The Gradient Covector

Our target function is the scalar field $f(\theta) = \eta_1 - \eta_2 = \frac{\mu_1}{\sqrt{v_1}} - \frac{\mu_2}{\sqrt{v_2}}$. Its differential df is a covector with components $g_\alpha = \frac{\partial f}{\partial \theta^\alpha}$:

$$g_\alpha = \begin{pmatrix} v_1^{-1/2} \\ -v_2^{-1/2} \\ -\frac{1}{2}\mu_1 v_1^{-3/2} \\ \frac{1}{2}\mu_2 v_2^{-3/2} \end{pmatrix}. \quad (16)$$

Geometric Interpretation (Foliation): The function $f(\theta)$ defines a *foliation* of the statistical manifold by level sets (hypersurfaces of constant Sharpe difference). The gradient covector g is the **normal vector** to these leaves. The asymptotic variance is simply the “squared length” of this normal vector, as measured by the covariance metric Ω . It represents the cost of moving orthogonal to the foliation—i.e., the uncertainty in the direction that actually changes the Sharpe ratio.

4.3 Metric Contraction (The Result)

The asymptotic variance is the scalar obtained by contracting the gradient covector with the covariance tensor:

$$\text{Var}(\Delta\eta) \approx \frac{1}{T} \Omega^{\alpha\beta} (df)_\alpha (df)_\beta. \quad (17)$$

Geometric Intuition: Just as the length of a vector v is $\|v\|^2 = g_{ij}v^i v^j$, the “uncertainty” of a scalar function f is the squared length of its gradient df measured by the covariance metric Ω . This quantity is a scalar invariant: it does not depend on which coordinate system (e.g., central moments vs. raw moments) we use to perform the calculation. Performing the matrix multiplication (see `derivations.ipynb` for symbolic verification):

$$\Omega^{\alpha\beta} (df)_\alpha (df)_\beta = \underbrace{v_1(df)_1^2 + v_2(df)_2^2 + 2\rho\sqrt{v_1 v_2}(df)_1(df)_2}_{\text{Mean terms}} \quad (18)$$

$$+ \underbrace{2v_1^2(df)_3^2 + 2v_2^2(df)_4^2 + 2(2\rho^2 v_1 v_2)(df)_3(df)_4}_{\text{Variance terms}}. \quad (19)$$

Substituting the components yields:

$$\text{Var}(\Delta\eta) \approx \frac{1}{T} \left[2(1 - \rho) + \frac{1}{2}(\eta_1^2 + \eta_2^2) - \rho^2\eta_1\eta_2 \right]. \quad (20)$$

This recovers the Jobson–Korkie result as a clean geometric contraction.⁴

⁴If returns are not Gaussian, the covariance tensor Ω acquires off-diagonal blocks relating means and variances (skewness). Geometrically, this introduces “torsion” to the statistical manifold, twisting the independence of the first and second moments.