LECTURE

Intrinsic First-Order Optimization

Differentials, Gradients, Constraints, and Multipliers

Antonio Franchi

University of Twente and Sapienza University of Rome

August 2026

Version 0.2.0 (Draft)

Guiding question

What belongs intrinsically to the cost, and what appears only after choosing a metric, coordinates, or constraints?

Why these notes are shared

A learner's reconstruction

These are evolving notes written while studying, conducting research, and preparing lectures.

They record an effort to:

- connect ideas often presented separately;
- expose hidden dependencies;
- make implicit reasoning explicit.

Expert accounts benefit from compression. A learning perspective may recover steps obscured by this *curse of knowledge*.

The other edge

These notes do not claim specialist authority.

A non-specialist reconstruction may offer a useful viewpoint, but it may also contain:

- imprecise statements;
- missing hypotheses;
- errors or incomplete arguments.

This is not peer-reviewed or definitive teaching material. Please read critically and verify independently.

Sharing and reuse

Corrections, counterexamples, missing hypotheses, and clearer explanations are welcome.

Lecture contents

Prerequisites and notation

This lecture assumes familiarity with the following basic concepts.

Linear algebra

- vector spaces and linear maps;
- kernels, images, and rank;
- dual spaces and covectors;
- inner products and positive-definite matrices.

Multivariable calculus

- directional derivatives;
- Jacobian matrices;
- the chain rule;
- gradients and Hessian matrices;
- equality constraints.

Smooth manifolds

- charts and coordinates;
- tangent and cotangent spaces;
- differentials of smooth maps;
- regular values and level sets;
- Riemannian metrics.

Scope of the lecture

The differential, metric gradient, constraint fiber, multiplier covector, tangent projection, and retraction will all be introduced explicitly. They are not assumed as prerequisites.

Companion references

The lecture is self-contained at the level of its main constructions. The following references provide broader treatments of the geometric and algorithmic background.

Manifold optimization

N. Boumal, *An Introduction to Optimization on Smooth Manifolds*, Cambridge University Press, 2023.

Riemannian gradients, retractions, and optimization algorithms

Geometric foundations

J. M. Lee, *Introduction to Smooth Manifolds*, 2nd ed., Springer, 2013.

tangent and cotangent spaces, differentials, and submanifolds

Matrix-manifold algorithms

P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2008.

matrix manifolds and numerical optimization methods

Orientation roadmap

The next three slides present the complete conceptual architecture of the lecture before its individual parts are developed.

What is expected now

Follow the distinction between:

- intrinsic objects,
- added structures,
- representations.

What is not expected yet

The detailed meaning of

$$\sharp_{g,v}, \qquad (\mathrm{d}f_v)^*, \qquad (\mathrm{T}_v\mathcal{F}_w)^\circ$$

will be developed later.

Takeaway

At this stage, the formulas are a map of the route ahead. They are not being used as prior knowledge.

Roadmap 1/3: one differential, two geometric developments

Let $\mathcal{V}$ and $\mathcal{W}$ be smooth finite-dimensional manifolds, and consider a smooth cost function

$$J : \mathcal{V} \to \mathbb{R}$$

At $v \in \mathcal{V}$, the intrinsic first-order object supplied by J is the covector

$$\boxed{\mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}.}$$

From this common starting point, two complementary paths emerge.

Metric realization

A Riemannian metric g converts the differential into a tangent vector:

$$\mathrm{d}J_v \xrightarrow{\sharp_{g,v}} \operatorname{grad}_g J(v).$$

Steepest descent, preconditioning, Newton.

Constraint realization

A smooth task map $f : \mathcal{V} \to \mathcal{W}$ and a prescribed value $w \in \mathcal{W}$ select the fiber

$$\mathcal{F}_w = f^{-1}(w).$$

At constrained stationary points,

$$\mathrm{d}J_v|_{\mathrm{T}_v\mathcal{F}_w} = 0 \iff \mathrm{d}J_v = (\mathrm{d}f_v)^*\mu.$$

Conormals, multipliers, feasible updates.

Roadmap 2/3: four levels of description

Intrinsic first-order objects

$$\mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}, \quad \text{for} \quad J : \mathcal{V} \to \mathbb{R}, \text{ and}$$

$$\mathrm{d}f_v : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_w\mathcal{W}, \quad \text{for} \quad f : \mathcal{V} \to \mathcal{W}$$

↓ enrich

Additional geometric structures

metric g_v, constraint fiber $\mathcal{F}_w$, tangent and conormal spaces

↓ represent

Coordinate representations

$$d_v, \qquad G_v, \qquad F_v, \qquad [\mu]_w$$

↓ compute

Numerical constructions

metric gradients, preconditioned directions,
tangent projections, and retraction updates

Conceptual warning

Coordinate arrays are numerical representations of geometric objects, not new geometric objects.

Roadmap 3/3: the central message

Takeaway

$$
\begin{aligned}
J : \mathcal{V} \to \mathbb{R} \quad &\longrightarrow \quad \mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}, \\
g_v \quad &\longrightarrow \quad \operatorname{grad}_g J(v) = \sharp_{g,v}(\mathrm{d}J_v), \\
f(v) = w \quad &\longrightarrow \quad \mathcal{F}_w = f^{-1}(w), \qquad \mathrm{T}_v\mathcal{F}_w = \ker \mathrm{d}f_v, \\
\text{constrained stationarity} \quad &\longrightarrow \quad \mathrm{d}J_v|_{\mathrm{T}_v\mathcal{F}_w} = 0 \iff \mathrm{d}J_v = (\mathrm{d}f_v)^*\mu.
\end{aligned}
$$

Coordinates

Coordinates represent these objects by arrays suitable for computation.

Algorithms

Projection and retraction turn tangent descent directions into numerical iterations.

The differential is the common first-order object.

Roadmap complete: we now develop each construction from first principles.

PART I

I

Foundations

From the abstract cost map to its intrinsic first-order information.

When we write J, are we writing a function or one of its formulas?

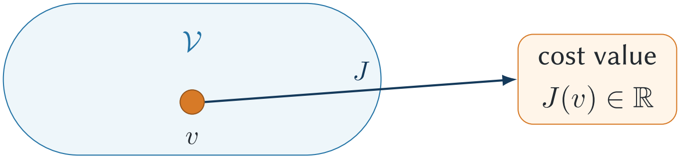
The function is intrinsic. A formula appears only after choosing coordinates.

The cost is an abstract map

The cost is a smooth map

$$\boxed{J : \mathcal{V} \to \mathbb{R}.}$$

Its argument is a point $v \in \mathcal{V}$, and its value is the scalar $J(v) \in \mathbb{R}$.



Conceptual warning

A point of $\mathcal{V}$ is not intrinsically a numerical column. Numerical components appear only after coordinates have been chosen.

A chart produces a coordinate formula

Choose a chart

$$\varphi : U \subset \mathcal{V} \longrightarrow \varphi(U) \subset \mathbb{R}^n.$$

The same cost is then represented by the coordinate map

$$\boxed{J_\varphi = J \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}.}$$

$$\begin{tikzcd}
\text{points } U \subset \mathcal{V} \arrow[r, "\varphi"] \arrow[dr, "J"'] & \text{coordinate arrays } \varphi(U) \subset \mathbb{R}^n \arrow[d, "J_\varphi"] \\
& \text{cost values } \mathbb{R}
\end{tikzcd}$$

Coordinate realization

J_φ is a numerical formula for J in the chart φ. It is not a new cost function on $\mathcal{V}$.

One function can have different formulas

Consider a one-dimensional manifold with coordinate x, in which the cost is represented by

$$J_x(x) = x^2.$$

Introduce the new coordinate

$$y = \arctan x, \qquad x = \tan y, \qquad y \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right).$$

The coordinate representative becomes

$$\boxed{J_y(y) = J_x(\tan y) = \tan^2 y.}$$

x-coordinate

$$J_x(x) = x^2$$

y-coordinate

$$J_y(y) = \tan^2 y$$

Key idea

The formulas and their numerical arguments differ, but both assign the same cost value to the same point of the manifold.

What is the primary first-order object supplied by the cost?

A covector that assigns a first-order rate of change to every tangent direction.

The differential is a covector

Let $J : \mathcal{V} \to \mathbb{R}$ be smooth and fix $v \in \mathcal{V}$. The differential of J at v is the linear map

$$\boxed{\mathrm{d}J_v : \mathrm{T}_v\mathcal{V} \to \mathbb{R}.}$$

For $\xi \in \mathrm{T}_v\mathcal{V}$, choose a smooth curve

$$\gamma : (-\varepsilon, \varepsilon) \to \mathcal{V}, \qquad \gamma(0) = v, \qquad \dot{\gamma}(0) = \xi.$$

Then

$$\boxed{\mathrm{d}J_v(\xi) = \frac{\mathrm{d}}{\mathrm{d}t} J\big(\gamma(t)\big)\Big|_{t=0}.}$$

Key idea

The differential is an element of the cotangent space:

$$\mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}.$$

Its value depends only on the tangent vector ξ, not on the particular curve used to represent ξ.

The differential predicts first-order change

Let

$$\gamma(0) = v, \qquad \dot{\gamma}(0) = \xi.$$

The cost along the curve satisfies

$$\boxed{J\big(\gamma(t)\big) = J(v) + t\,\mathrm{d}J_v(\xi) + o(t).}$$

Descent

If

$$\mathrm{d}J_v(\xi) < 0,$$

the cost decreases to first order along ξ.

Level

If

$$\mathrm{d}J_v(\xi) = 0,$$

the cost is constant to first order along ξ.

Ascent

If

$$\mathrm{d}J_v(\xi) > 0,$$

the cost increases to first order along ξ.

A covector measures tangent directions

At the point v, the differential accepts a tangent vector and returns its first-order cost rate:

$$\xi \in \mathrm{T}_v\mathcal{V} \quad \xmapsto{\mathrm{d}J_v} \quad \mathrm{d}J_v(\xi) \in \mathbb{R}.$$

$$\begin{tikzcd}[column sep=huge] \boxed{\begin{array}{c}\text{tangent vector}\\ \xi \in \mathrm{T}_v\mathcal{V}\end{array}} \arrow[r, "\text{is evaluated by}"] & \boxed{\begin{array}{c}\text{differential}\\ \mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}\end{array}} \arrow[r, "\text{returns}"] & \boxed{\begin{array}{c}\text{first-order rate}\\ \mathrm{d}J_v(\xi) \in \mathbb{R}\end{array}} \end{tikzcd}$$

Geometric interpretation

A tangent vector represents an infinitesimal motion. A covector is a linear measuring device that assigns a scalar to each such motion.

A differential is not a gradient

The differential and a metric gradient belong to different vector spaces:

$$\boxed{\mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}, \qquad \operatorname{grad}_g J(v) \in \mathrm{T}_v\mathcal{V}.}$$

Differential

$$\mathrm{d}J_v : \mathrm{T}_v\mathcal{V} \to \mathbb{R}$$

A covector supplied intrinsically by the cost.

Gradient

$$\operatorname{grad}_g J(v) = \sharp_{g,v}(\mathrm{d}J_v)$$

A tangent vector obtained only after choosing a metric.

Conceptual warning

Writing the partial derivatives of J in a column does not turn the differential into a tangent vector. The column still contains covector components.

What can the differential tell us before any metric is chosen?

It already identifies level, ascent, descent, and stationary behavior.

Three classes of tangent directions

Assume that $\mathrm{d}J_v \neq 0$. The differential partitions $\mathrm{T}_v\mathcal{V}$ according to the sign of the first-order cost change:

$$\mathcal{D}_v^- = \{\xi \in \mathrm{T}_v\mathcal{V} : \mathrm{d}J_v(\xi) < 0\}\,,$$

$$\ker \mathrm{d}J_v = \{\xi \in \mathrm{T}_v\mathcal{V} : \mathrm{d}J_v(\xi) = 0\}\,,$$

$$\mathcal{D}_v^+ = \{\xi \in \mathrm{T}_v\mathcal{V} : \mathrm{d}J_v(\xi) > 0\}\,.$$

Descent

$$\xi \in \mathcal{D}_v^-$$

The cost decreases to first order.

Level

$$\xi \in \ker \mathrm{d}J_v$$

The cost is constant to first order.

Ascent

$$\xi \in \mathcal{D}_v^+$$

The cost increases to first order.

The level hyperplane

Because the nonzero linear functional

$$\mathrm{d}J_v : \mathrm{T}_v\mathcal{V} \to \mathbb{R}$$

has rank one, its kernel is a codimension-one linear subspace:

$$\boxed{\ker \mathrm{d}J_v \subset \mathrm{T}_v\mathcal{V}.}$$

If $\mathrm{d}J_v \neq 0$, then $J(v)$ is a regular value locally at v, and the cost level set through v satisfies

$$\boxed{\mathrm{T}_v J^{-1}\big(J(v)\big) = \ker \mathrm{d}J_v.}$$

Geometric interpretation

The level set $J^{-1}(J(v))$ is generally nonlinear. Its tangent space at v is the linear hyperplane containing exactly the directions along which J remains constant to first order.

Key idea

The level hyperplane is determined by the differential alone. No metric, angle, orthogonality, or gradient has yet been introduced.

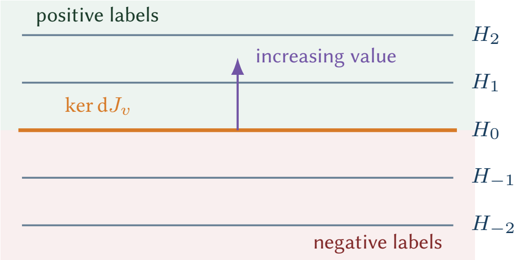
The labelled stack of affine hyperplanes

For every $a \in \mathbb{R}$, define

$$H_a = \{\xi \in \mathrm{T}_v\mathcal{V} : \mathrm{d}J_v(\xi) = a\}\,.$$

Each nonempty H_a is an affine hyperplane parallel to

$$H_0 = \ker \mathrm{d}J_v.$$



What the stack records

The kernel records only the zero level. The full family

$$\{H_a\}_{a\in\mathbb{R}}$$

also records the scalar assigned to every tangent direction.

Scaling the covector

The covectors α and 2α have the same kernel and sign half-spaces, but

$$(2\alpha)(\xi) = 2\alpha(\xi).$$

Stationarity is defined before the gradient

A point $v \in \mathcal{V}$ is an unconstrained stationary point of J if

$$\boxed{\mathrm{d}J_v = 0 \in \mathrm{T}_v^*\mathcal{V}.}$$

At such a point,

$$\mathrm{d}J_v(\xi) = 0 \qquad \forall \xi \in \mathrm{T}_v\mathcal{V},$$

so there is no tangent direction of strict first-order ascent or descent:

$$\mathcal{D}_v^- = \varnothing, \qquad \mathcal{D}_v^+ = \varnothing, \qquad \ker \mathrm{d}J_v = \mathrm{T}_v\mathcal{V}.$$

Stationarity is not optimality

The equation $\mathrm{d}J_v = 0$ identifies a critical point. First-order information alone does not determine whether the point is a minimum, a maximum, or a saddle.

Takeaway

Stationarity is intrinsic and metric-free. A metric will later provide a vector representation of the same condition through $\operatorname{grad}_g J(v) = 0$.

PART II

Metric and Coordinate Geometry

How metrics and coordinates represent and use the differential.

If the differential is a covector, where does the gradient come from?

A metric provides the covector-to-vector identification.

A metric adds geometric structure

A Riemannian metric g assigns to every $v \in \mathcal{V}$ an inner product on T_v

$$\boxed{g_v : \mathrm{T}_v\mathcal{V} \times \mathrm{T}_v\mathcal{V} \to \mathbb{R},}$$

varying smoothly with v.

Already determined by $\mathrm{d}J_v$

$$\ker \mathrm{d}J_v, \qquad \mathcal{D}_v^-, \qquad \mathcal{D}_v^+.$$

Level, descent, and ascent directions are metric-free.

Introduced by g_v

$$\|\xi\|_g, \qquad \langle \xi, \zeta \rangle_{g_v}, \qquad \perp_g.$$

Length, angle, orthogonality, and normal directions require a metric.

Takeaway

The differential identifies all strict descent directions, but a metric is required to compare their lengths and decide which one is steepest.

The metric connects vectors and covectors

At every $v \in \mathcal{V}$, the metric defines the flat map

$$\flat_{g,v} : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_v^*\mathcal{V}, \qquad \boxed{\flat_{g,v}(\xi) = g_v(\xi, \cdot).}$$

Because g_v is an inner product, $\flat_{g,v}$ is a linear isomorphism. Its inverse is the sharp map

$$\sharp_{g,v} : \mathrm{T}_v^*\mathcal{V} \to \mathrm{T}_v\mathcal{V}.$$

$$\begin{tikzcd}
\begin{array}{c}\text{tangent vectors}\\ \mathrm{T}_v\mathcal{V}\end{array} \arrow[r, shift left, "\flat_{g,v}"] & \begin{array}{c}\text{covectors}\\ \mathrm{T}_v^*\mathcal{V}\end{array} \arrow[l, shift left, "\sharp_{g,v}"]
\end{tikzcd}$$

Geometric interpretation

The sharp and flat maps are supplied by the chosen metric. They are not canonical identifications between $\mathrm{T}_v\mathcal{V}$ and $\mathrm{T}_v^*\mathcal{V}$.

The metric gradient represents the differential

The gradient of J associated with the metric g is

$$\boxed{\operatorname{grad}_g J(v) = \sharp_{g,v}(\mathrm{d}J_v) \in \mathrm{T}_v\mathcal{V}.}$$

Equivalently, it is the unique tangent vector satisfying

$$\boxed{g_v\big(\operatorname{grad}_g J(v), \xi\big) = \mathrm{d}J_v(\xi) \qquad \forall \xi \in \mathrm{T}_v\mathcal{V}.}$$

Differential

$$\mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}$$

Intrinsic first-order covector supplied by J.

Metric gradient

$$\operatorname{grad}_g J(v) \in \mathrm{T}_v\mathcal{V}$$

Metric-dependent vector associated with $\mathrm{d}J_v$.

Takeaway

Different metrics generally associate different gradients with the same differential. The differential is unique, while the gradient depends on the chosen metric.

The negative gradient is a descent direction

If $\mathrm{d}J_v \neq 0$, then the gradient is nonzero because $\sharp_{g,v}$ is an isomorphism. Evaluating the differential on the negative gradient gives

$$\begin{aligned} \mathrm{d}J_v\big(-\operatorname{grad}_g J(v)\big) &= g_v\big(\operatorname{grad}_g J(v), -\operatorname{grad}_g J(v)\big) \\ &= -\left\|\operatorname{grad}_g J(v)\right\|_g^2 < 0. \end{aligned}$$

Therefore,

$$\boxed{-\operatorname{grad}_g J(v) \in \mathcal{D}_v^-.}$$

For every metric

The negative of the induced gradient is a strict descent direction at every nonstationary point.

Stationarity

Since the sharp map is invertible,

$$\mathrm{d}J_v = 0 \iff \operatorname{grad}_g J(v) = 0.$$

Takeaway

Changing the metric changes the gradient, but it does not change the stationary set or the intrinsic descent half-space.

Does the metric create the level hyperplane, or only its normal representation?

The hyperplane already exists. The metric selects a perpendicular direction.

The gradient is normal to the level hyperplane

Assume that $\mathrm{d}J_v \neq 0$. From the defining identity

$$g_v\big(\operatorname{grad}_g J(v), \xi\big) = \mathrm{d}J_v(\xi),$$

a tangent vector $\xi \in \mathrm{T}_v\mathcal{V}$ belongs to the kernel precisely when

$$g_v\big(\operatorname{grad}_g J(v), \xi\big) = 0.$$

Therefore,

$$\boxed{\ker \mathrm{d}J_v = \operatorname{span}\{\operatorname{grad}_g J(v)\}^{\perp_g}.}$$

Equivalently,

$$\boxed{\operatorname{span}\{\operatorname{grad}_g J(v)\} = \big(\ker \mathrm{d}J_v\big)^{\perp_g}.}$$

Geometric interpretation

The gradient spans the one-dimensional g-normal space to the level hyperplane. Both the direction and the notion of perpendicularity depend on the chosen metric.

The hyperplane precedes its metric normal

The relation

$$\ker \mathrm{d}J_v = \operatorname{span}\{\operatorname{grad}_g J(v)\}^{\perp_g}$$

must be interpreted in the correct direction.

First: intrinsic hyperplane

The differential determines

$$\ker \mathrm{d}J_v = \{\xi \in \mathrm{T}_v\mathcal{V} : \mathrm{d}J_v(\xi) = 0\}\,.$$

This hyperplane exists independently of any metric.

Then: metric normal

After choosing g, the sharp map produces

$$\operatorname{grad}_g J(v) = \sharp_{g,v}(\mathrm{d}J_v).$$

This vector is g-orthogonal to the pre-existing hyperplane.

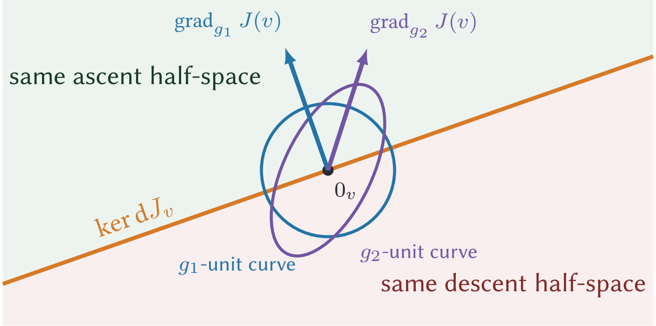
Conceptual warning

The right-hand side does not define the level hyperplane. It represents that intrinsic hyperplane as the g-orthogonal complement of a metric-dependent gradient direction.

Different metrics select different gradients

Let g_1 and g_2 be two Riemannian metrics. The same differential produces

$$\operatorname{grad}_{g_1} J(v) = \sharp_{g_1,v}(\mathrm{d}J_v), \qquad \operatorname{grad}_{g_2} J(v) = \sharp_{g_2,v}(\mathrm{d}J_v).$$



Key idea

Changing the metric changes the gradient direction and its norm. It does not change $\mathrm{d}J_v$, its kernel, its sign half-spaces, or the stationary set.

What does “steepest descent” actually mean?

It means minimizing the first-order cost change among tangent directions of unit metric length.

Steepest descent requires a metric

Assume that $\mathrm{d}J_v \neq 0$. Among all tangent vectors of unit g-norm, the steepest first-order descent direction solves

$$\min_{\substack{\xi \in \mathrm{T}_v\mathcal{V} \\ \|\xi\|_g = 1}} \mathrm{d}J_v(\xi).$$

Since

$$\mathrm{d}J_v(\xi) = g_v\big(\operatorname{grad}_g J(v), \xi\big),$$

the Cauchy–Schwarz inequality gives

$$\mathrm{d}J_v(\xi) \geq -\left\|\operatorname{grad}_g J(v)\right\|_g \|\xi\|_g .$$

For $\|\xi\|_g = 1$, equality is attained uniquely at

$$\boxed{\xi_{\mathrm{sd}} = -\frac{\operatorname{grad}_g J(v)}{\left\|\operatorname{grad}_g J(v)\right\|_g}.}$$

Key idea

The negative gradient is not merely a descent direction. After normalization, it is the direction of greatest first-order decrease among all tangent vectors of unit g-norm.

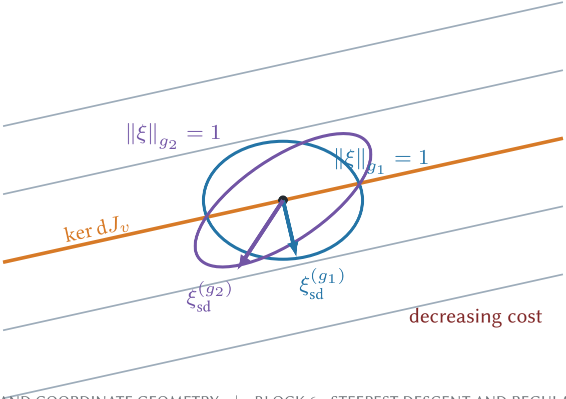
Different metrics define different steepest directions

The differential fixes the objective

$$\xi \longmapsto \mathrm{d}J_v(\xi),$$

but the unit constraint depends on the chosen metric:

$$\|\xi\|_{g_1} = 1 \qquad \text{or} \qquad \|\xi\|_{g_2} = 1.$$



A gradient step minimizes a regularized model

The steepest-direction problem fixes the metric length of ξ. A complete tangent displacement can instead be selected by solving

$$\min_{\xi \in \mathrm{T}_v \mathcal{V}} \left[\mathrm{d}J_v(\xi) + \frac{1}{2\eta} \|\xi\|_g^2 \right], \qquad \eta > 0.$$

The first term predicts the cost change, while the second penalizes the size of the tangent displacement.

The unique minimizer is

$$\boxed{\xi_\eta = -\eta \operatorname{grad}_g J(v).}$$

Indeed, the first-order condition is

$$\mathrm{d}J_v(\zeta) + \frac{1}{\eta} g_v(\xi_\eta, \zeta) = 0 \qquad \forall \zeta \in \mathrm{T}_v \mathcal{V},$$

which is equivalent to

$$\operatorname{grad}_g J(v) + \frac{1}{\eta} \xi_\eta = 0.$$

Key idea

The metric determines both the descent direction and the quadratic cost assigned to tangent displacement.

Direction selection versus displacement selection

The two variational problems answer different questions.

Steepest unit direction

$$\min_{\|\xi\|_g=1} \mathrm{d}J_v(\xi)$$

selects only the direction

$$-\frac{\operatorname{grad}_g J(v)}{\left\|\operatorname{grad}_g J(v)\right\|_g}.$$

The metric unit sphere fixes the length.

Regularized displacement

$$\min_{\xi\in \mathrm{T}_v\mathcal{V}} \left[\mathrm{d}J_v(\xi) + \frac{1}{2\eta}\|\xi\|_g^2\right]$$

selects the complete displacement

$$-\eta \operatorname{grad}_g J(v).$$

The parameter η controls its scale.

Takeaway

The gradient is simultaneously the metric representation of the differential and the solution-generating direction of a local optimization problem in the tangent space.

What is actually differentiated when numerical coordinates are used?

The coordinate representative is differentiated; the resulting array represents the intrinsic differential.

A chart assigns coordinates to manifold points

Choose a chart around $v \in \mathcal{V}$:

$$\varphi : U \subset \mathcal{V} \longrightarrow \varphi(U) \subset \mathbb{R}^n, \qquad \varphi(q) = \big(x^1(q), \dots, x^n(q)\big).$$

The chart induces the coordinate tangent basis

$$\boxed{\left\{ \frac{\partial}{\partial x^1}\bigg|_v, \dots, \frac{\partial}{\partial x^n}\bigg|_v \right\} \subset \mathrm{T}_v\mathcal{V}}$$

and the dual coordinate cotangent basis

$$\boxed{\left\{ \mathrm{d}x^1|_v, \dots, \mathrm{d}x^n|_v \right\} \subset \mathrm{T}^*_v\mathcal{V}.}$$

These bases satisfy

$$\mathrm{d}x^i|_v \left(\frac{\partial}{\partial x^j}\bigg|_v \right) = \delta^i_j.$$

Key idea

Coordinates provide numerical representations of points, tangent vectors, and covectors. They do not replace the underlying intrinsic objects.

The coordinate representative is differentiated

The chart represents the abstract cost $J : \mathcal{V} \to \mathbb{R}$ by

$$J_\varphi = J \circ \varphi^{-1} : \varphi(U) \subset \mathbb{R}^n \to \mathbb{R}.$$

Let $u^1, \ldots, u^n$ denote the standard numerical coordinates on $\varphi(U)$. Ordinary differentiation is applied to J_φ:

$$\left.\frac{\partial J_\varphi}{\partial u^i}\right|_{u=\varphi(v)} = \left.\frac{\partial (J \circ \varphi^{-1})}{\partial u^i}\right|_{u=\varphi(v)}.$$

These numbers are the coordinate components of the intrinsic differential:

$$\boxed{\mathrm{d}J_v = \sum_{i=1}^{n} \left.\frac{\partial J_\varphi}{\partial u^i}\right|_{u=\varphi(v)} \mathrm{d}x^i|_v.}$$

Coordinate realization

The abstract map J does not intrinsically possess numerical variables. Differentiation rules are applied to its coordinate representative J_φ.

What does the column d_v store?

Evaluating the differential on the i-th coordinate tangent vector gives

$$\mathrm{d}J_v \left(\left. \frac{\partial}{\partial x^i} \right|_v \right) = \left. \frac{\partial J_\varphi}{\partial u^i} \right|_{u=\varphi(v)} .$$

This scalar is conventionally abbreviated as

$$\frac{\partial J}{\partial x^i}(v).$$

Store the components of $\mathrm{d}J_v$ in the column

$$\boxed{d_v := \begin{bmatrix} \frac{\partial J}{\partial x^1}(v) & \cdots & \frac{\partial J}{\partial x^n}(v) \end{bmatrix}^\top .}$$

Conceptual warning

The column d_v contains covector components in the basis $\{\mathrm{d}x^i|_v\}$. Its vertical storage format does not make it a tangent vector.

The differential still acts on tangent vectors

Write an arbitrary tangent vector as

$$\xi = \sum_{i=1}^{n} \xi^i \left.\frac{\partial}{\partial x^i}\right|_v, \qquad [\xi]_v = \begin{bmatrix} \xi^1 \\ \vdots \\ \xi^n \end{bmatrix}.$$

Using the duality of the coordinate bases,

$$\mathrm{d}J_v(\xi) = \sum_{i=1}^{n} \frac{\partial J}{\partial x^i}(v)\, \xi^i = \boxed{d_v^\top [\xi]_v.}$$

Covector components

$$d_v \quad \text{represent} \quad \mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}.$$

Vector components

$$[\xi]_v \quad \text{represent} \quad \xi \in \mathrm{T}_v\mathcal{V}.$$

Takeaway

The matrix product $d_v^\top [\xi]_v$ is the coordinate representation of the intrinsic dual pairing $\mathrm{d}J_v(\xi)$.

How does the metric convert covector components into vector components?

Through the inverse metric matrix, which represents the sharp map in coordinates.

The metric is represented by a matrix

In the coordinate tangent basis, define $G_v = [g_{ij}(v)]$, where

$$\boxed{g_{ij}(v) = g_v\left(\frac{\partial}{\partial x^i}\bigg|_v, \frac{\partial}{\partial x^j}\bigg|_v \right).}$$

For tangent vectors

$$\xi = \sum_i \xi^i \frac{\partial}{\partial x^i}\bigg|_v, \qquad \zeta = \sum_j \zeta^j \frac{\partial}{\partial x^j}\bigg|_v,$$

the metric pairing is represented by

$$\boxed{g_v(\xi, \zeta) = [\xi]_v^\top G_v [\zeta]_v.}$$

Coordinate realization

Because g_v is an inner product,

$$G_v = G_v^\top \succ 0.$$

Thus G_v is symmetric, positive definite, and invertible.

The gradient identity becomes a linear system

Write the metric gradient as

$$\operatorname{grad}_g J(v) = \sum_{i=1}^{n} \zeta^i \left.\frac{\partial}{\partial x^i}\right|_v, \qquad [\operatorname{grad}_g J]_v = \begin{bmatrix} \zeta^1 \\ \vdots \\ \zeta^n \end{bmatrix}.$$

For an arbitrary tangent vector $\xi \in \mathrm{T}_v\mathcal{V}$, the defining identity

$$g_v\big(\operatorname{grad}_g J(v), \xi\big) = \mathrm{d}J_v(\xi)$$

becomes

$$[\operatorname{grad}_g J]_v^\top G_v [\xi]_v = d_v^\top [\xi]_v.$$

Since this holds for every component column $[\xi]_v$, and since G_v is symmetric,

$$\boxed{G_v [\operatorname{grad}_g J]_v = d_v.}$$

Key idea

The coordinate gradient is determined by a linear system whose right-hand side contains the components of the differential.

The inverse metric represents the sharp map

Solving

$$G_v[\operatorname{grad}_g J]_v = d_v$$

gives

$$\boxed{[\operatorname{grad}_g J]_v = G_v^{-1} d_v.}$$

$$\begin{array}{c}\text{covector components}\\ d_v\end{array} \xrightarrow[\text{coordinate sharp map}]{G_v^{-1}} \begin{array}{c}\text{vector components}\\ [\operatorname{grad}_g J]_v\end{array}$$

Coordinate realization

The matrix G_v^{-1} does not merely rescale a numerical column. It represents the linear map

$$\sharp_{g,v} : \mathrm{T}_v^*\mathcal{V} \to \mathrm{T}_v\mathcal{V}$$

in the selected coordinate bases.

The two columns represent different objects

Although d_v and $[\operatorname{grad}_g J]_v$ are both stored as numerical columns, they represent geometrically different objects.

Covector components

$$\mathrm{d}J_v = \sum_{i=1}^{n} \frac{\partial J}{\partial x^i}(v)\, \mathrm{d}x^i|_v.$$

The entries of d_v are called *covariant components.*

Vector components

$$\operatorname{grad}_g J(v) = \sum_{i=1}^{n} \zeta^i \left.\frac{\partial}{\partial x^i}\right|_v.$$

The entries of $[\operatorname{grad}_g J]_v$ are called *contravariant components.*

Conceptual warning

Covariant and contravariant refer to different transformation laws under a change of coordinates. The terms do not describe the physical orientation of the columns.

Takeaway

The equation

$$[\operatorname{grad}_g J]_v = G_v^{-1} d_v$$

converts the coordinate representation of a covector into the coordinate representation of a vector.

If the arrays change with coordinates, how does the geometry remain unchanged?

Vectors, covectors, and metrics transform differently but consistently, preserving all intrinsic pairings and identities.

A coordinate change transforms component columns

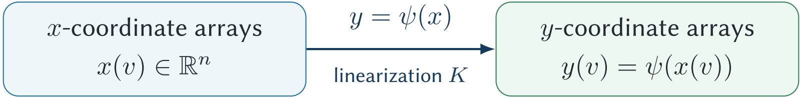
Let $x = (x^1, \dots, x^n)$ and $y = (y^1, \dots, y^n)$ be two coordinate systems around $v \in \mathcal{V}$. Their transition map is

$$y = \psi(x),$$

with Jacobian matrix

$$\boxed{K = \left.\frac{\partial y}{\partial x}\right|_{x(v)}.}$$

Since the coordinate transition is a local diffeomorphism, $K \in \mathbb{R}^{n\times n}$ is invertible.



Coordinate realization

The matrix K represents the differential of the coordinate transition. It determines how all coordinate components must change.

Vectors and covectors transform differently

The component column of a tangent vector transforms according to

$$\boxed{[\xi]_y = K[\xi]_x.}$$

The component column of a covector transforms according to

$$\boxed{[\alpha]_y = K^{-\top}[\alpha]_x.}$$

In particular,

$$\boxed{[\mathrm{d}J]_y = K^{-\top}[\mathrm{d}J]_x.}$$

Contravariant components

Vector components transform with K.

Covariant components

Covector components transform with $K^{-\top}$.

Conceptual warning

The different transformation laws reflect different tensor types. They are not consequences of storing one array as a row and another as a column.

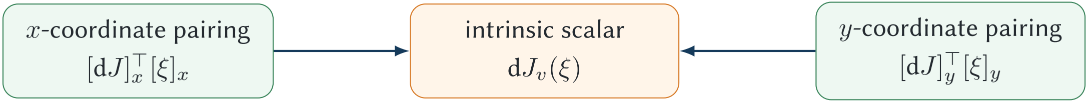
The vector–covector pairing is invariant

Although the two component columns transform differently, their scalar pairing remains unchanged:

$$
\begin{aligned}
[\mathrm{d}J]_y^\top [\xi]_y &= \left(K^{-\top}[\mathrm{d}J]_x\right)^\top K[\xi]_x \\
&= [\mathrm{d}J]_x^\top K^{-1} K[\xi]_x \\
&= \boxed{[\mathrm{d}J]_x^\top [\xi]_x.}
\end{aligned}
$$

Thus,

$$
[\mathrm{d}J]_y^\top [\xi]_y = [\mathrm{d}J]_x^\top [\xi]_x = \mathrm{d}J_v(\xi).
$$



Takeaway

The coordinate arrays change, but the number assigned by the covector to the vector does not.

The metric makes the gradient transform as a vector

The metric matrix transforms as $\boxed{G_y = K^{-\top} G_x K^{-1},}$

and therefore its inverse transforms as $\boxed{G_y^{-1} = K G_x^{-1} K^{\top}.}$

Using the coordinate gradient formula,

$$[\operatorname{grad}_g J]_y = G_y^{-1}[\mathrm{d}J]_y = \left(K G_x^{-1} K^{\top}\right)\left(K^{-\top}[\mathrm{d}J]_x\right) = K G_x^{-1}[\mathrm{d}J]_x = \boxed{K[\operatorname{grad}_g J]_x.}$$

Key idea

The coordinate gradient transforms exactly as a tangent vector. Thus the columns $[\operatorname{grad}_g J]_x$ and $[\operatorname{grad}_g J]_y$ represent the same intrinsic vector in two coordinate systems.

Conceptual warning

Changing coordinates transforms G_x into G_y. Setting the metric matrix to the identity again after a nonlinear coordinate change would generally select a different metric.

Why is the identity matrix not a coordinate-independent metric?

The same metric generally has different matrix representations in different coordinates.

The standard metric in Cartesian coordinates

Consider the function $J : \mathbb{R}^2 \to \mathbb{R}, \qquad J(x, y) = \frac{1}{2}\left(x^2 + y^2\right).$

In Cartesian coordinates, the differential is $\mathrm{d}J = x\,\mathrm{d}x + y\,\mathrm{d}y$, and its component column is

$$\boxed{[\mathrm{d}J]_{xy} = \begin{bmatrix} x \\ y \end{bmatrix}.}$$

The standard inner product is represented by

$$\boxed{G_{xy} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.}$$

Therefore,

$$\boxed{[\operatorname{grad} J]_{xy} = G_{xy}^{-1}[\mathrm{d}J]_{xy} = \begin{bmatrix} x \\ y \end{bmatrix}.}$$

Coordinate realization

The identity matrix represents the standard metric in Cartesian coordinates. This matrix representation is coordinate-dependent.

The same metric in polar coordinates

On $\mathbb{R}^2 \setminus \{0\}$, introduce local polar coordinates

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad r > 0.$$

The same intrinsic cost is represented by $J(r,\theta) = \frac{1}{2}r^2$, so

$$\mathrm{d}J = r\,\mathrm{d}r, \qquad \boxed{[\mathrm{d}J]_{r\theta} = \begin{bmatrix} r \\ 0 \end{bmatrix}.}$$

The standard metric is represented in the coordinate tangent basis $\left\{\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}\right\}$ by

$$\boxed{G_{r\theta} = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}.}$$

Geometric interpretation

A unit variation in θ produces a physical displacement whose length is proportional to r. The factor r^2 records this geometric scaling.

Different arrays represent the same gradient

Using the polar metric matrix,

$$G_{r\theta}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & r^{-2} \end{bmatrix}.$$

Hence

$$[\mathrm{grad}\, J]_{r\theta} = G_{r\theta}^{-1}[\mathrm{d}J]_{r\theta} = \begin{bmatrix} 1 & 0 \\ 0 & r^{-2} \end{bmatrix} \begin{bmatrix} r \\ 0 \end{bmatrix} = \boxed{\begin{bmatrix} r \\ 0 \end{bmatrix}.}$$

This component column represents $\boxed{\mathrm{grad}\, J = r\frac{\partial}{\partial r},}$

which is the same radial vector represented in Cartesian coordinates by $x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}$.

Conceptual warning

Replacing $G_{r\theta} = \mathrm{diag}(1, r^2)$ by I would not merely simplify the representation. It would select a different metric.

Takeaway

A coordinate change modifies the component columns and the metric matrix together, while preserving the intrinsic differential and gradient.

PART III

Metric Selection and Newton

How metrics and coordinates represent and use the differential.

What changes when we deliberately choose another metric?

The differential remains fixed, but the covector-to-vector conversion changes.

Two metrics are related by a tangent-space operator

Let g and h be two Riemannian metrics on $\mathcal{V}$. At each $v \in \mathcal{V}$, define

$$\boxed{A_v := \sharp_{g,v} \circ \flat_{h,v} : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_v\mathcal{V}.}$$

The construction is fully intrinsic:

$$\mathrm{T}_v\mathcal{V} \xrightarrow{\flat_{h,v}} \mathrm{T}_v^*\mathcal{V} \xrightarrow{\sharp_{g,v}} \mathrm{T}_v\mathcal{V}.$$

$$\begin{tikzcd} \begin{array}{c}\text{tangent vector}\\ \xi \in \mathrm{T}_v\mathcal{V}\end{array} \arrow[r, "\flat_{h,v}"] & \begin{array}{c}h\text{-dual covector}\\ \flat_{h,v}(\xi)\end{array} \arrow[r, "\sharp_{g,v}"] & \begin{array}{c}\text{tangent vector}\\ A_v(\xi) \in \mathrm{T}_v\mathcal{V}\end{array} \end{tikzcd}$$

Key idea

The operator A_v compares two covector-to-vector identifications while acting entirely within the tangent space $\mathrm{T}_v\mathcal{V}$.

The metric-change operator expresses h_v through g_v

The definition of A_v gives

$$\boxed{h_v(\xi,\zeta) = g_v\big(A_v(\xi),\zeta\big) \qquad \forall \xi,\zeta \in \mathrm{T}_v\mathcal{V}.}$$

Conversely, this identity uniquely determines A_v. Moreover,

$$g_v\big(A_v(\xi),\zeta\big) = g_v\big(\xi, A_v(\zeta)\big),$$

and

$$g_v\big(A_v(\xi),\xi\big) = h_v(\xi,\xi) > 0 \qquad \forall \xi \neq 0.$$

g_v-self-adjointness

The symmetry of h_v gives

$$g_v(A_v\xi,\zeta) = g_v(\xi, A_v\zeta).$$

g_v-positivity

The positive definiteness of h_v gives

$$g_v(A_v\xi,\xi) = h_v(\xi,\xi) > 0.$$

Takeaway

The operator A_v is invertible and records intrinsically how the two metrics assign different tangent vector representatives to covectors.

Changing the metric changes the gradient

The same differential is represented through the two metrics by

$$\mathrm{d}J_v = \flat_{g,v}\big(\operatorname{grad}_g J(v)\big) = \flat_{h,v}\big(\operatorname{grad}_h J(v)\big).$$

Applying $\sharp_{g,v}$ and using

$$A_v = \sharp_{g,v} \circ \flat_{h,v}$$

gives

$$\operatorname{grad}_g J(v) = A_v\big(\operatorname{grad}_h J(v)\big).$$

Therefore,

$$\boxed{\operatorname{grad}_h J(v) = A_v^{-1}\big(\operatorname{grad}_g J(v)\big).}$$

$$\begin{array}{c} g\text{-gradient} \\ \operatorname{grad}_g J(v) \end{array} \xrightarrow{\quad A_v^{-1} \quad} \begin{array}{c} h\text{-gradient} \\ \operatorname{grad}_h J(v) \end{array}$$

Takeaway

Changing the metric does not modify $\mathrm{d}J_v$. It changes the sharp map used to convert that fixed covector into a tangent vector.

Preconditioning selects an inverse metric

Return to coordinates. Let d_v contain the components of $\mathrm{d}J_v$, and choose $P_v = P_v^\top \succ 0$.
Define the tangent-vector component column $\boxed{[\xi]_v = -P_v d_v.}$
Interpreting $P_v = G_v^{-1},\ G_v = P_v^{-1}$, means that P_v represents the sharp map of a selected metric. Consequently,

$$[\xi]_v = -[\operatorname{grad}_g J]_v.$$

The intrinsic first-order variation along ξ is

$$\boxed{\mathrm{d}J_v(\xi) = d_v^\top [\xi]_v = -d_v^\top P_v d_v < 0 \qquad \text{if } d_v \neq 0.}$$

Type conversion

The matrix P_v maps covector components d_v to tangent-vector components $P_v d_v$.

Descent

Positive definiteness ensures that $-P_v d_v$ represents a strict descent direction whenever $\mathrm{d}J_v \neq 0$.

Takeaway

$\boxed{\text{choosing an SPD preconditioner} = \text{choosing an inverse metric matrix.}}$

Is Newton's method fundamentally different from gradient descent?

In a selected local model, its direction can be read as a gradient direction under a positive Hessian-based inner product.

Newton begins from a selected local coordinate model

Choose a chart around $v \in \mathcal{V}$:

$$\varphi : U \subset \mathcal{V} \longrightarrow \varphi(U) \subset \mathbb{R}^n.$$

The abstract cost is represented in this chart by

$$\boxed{J_\varphi = J \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}.}$$

At the coordinate point

$$x = \varphi(v),$$

define

$$\boxed{d_v = \nabla J_\varphi(x), \qquad H_v = \nabla^2 J_\varphi(x).}$$

Coordinate realization

Both d_v and H_v are computed from the selected coordinate representative J_φ. Their geometric meanings, however, are different.

The differential and coordinate Hessian have different status

The column d_v represents the intrinsic covector

$$\boxed{\mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}}$$

in the coordinate cotangent basis.
By contrast,

$$\boxed{H_v = \nabla^2 J_\varphi\big(\varphi(v)\big)}$$

is the matrix of ordinary second derivatives of the selected coordinate representative.

Intrinsic first-order object

Under a coordinate change, d_v transforms as a covector component column and continues to represent the same differential $\mathrm{d}J_v$.

Coordinate-local second-order object

Under a nonlinear coordinate change, H_v does not generally transform as the matrix of an intrinsic bilinear form.

Takeaway

The Newton construction developed next belongs to this selected local coordinate model. An intrinsic Hessian requires additional geometric structure.

Newton selects the inverse coordinate Hessian

The generic preconditioned direction has components

$$[\xi]_v = -P_v d_v, \qquad P_v = P_v^\top \succ 0.$$

In the selected local coordinate model, assume that

$$\boxed{H_v = \nabla^2 J_\varphi\big(\varphi(v)\big) = H_v^\top \succ 0.}$$

Newton selects

$$\boxed{P_v = H_v^{-1},}$$

and therefore

$$\boxed{[\xi_{\mathrm{N}}]_v = -H_v^{-1} d_v.}$$

Takeaway

Within the selected coordinate model, Newton is the preconditioned direction obtained by choosing the inverse positive coordinate Hessian.

The Newton direction solves a linear system

The Newton direction

$$[\xi_{\mathrm{N}}]_v = -H_v^{-1} d_v$$

is equivalently characterized by

$$\boxed{H_v[\xi_{\mathrm{N}}]_v = -d_v.}$$

Its intrinsic first-order cost variation is

$$\mathrm{d}J_v(\xi_{\mathrm{N}}) = d_v^\top [\xi_{\mathrm{N}}]_v = -d_v^\top H_v^{-1} d_v.$$

Since

$$H_v^{-1} \succ 0,$$

we obtain

$$\boxed{\mathrm{d}J_v(\xi_{\mathrm{N}}) < 0 \qquad \text{whenever } d_v \neq 0.}$$

Key idea

Positive definiteness makes the coordinate Newton direction a strict descent direction at every nonstationary point.

A positive coordinate Hessian defines a pointwise inner product

Fix the selected chart and assume

$$H_v \succ 0.$$

Then

$$\boxed{g^{\mathrm{N}}_{v,\varphi}(\xi,\zeta) := [\xi]_\varphi^\top H_v [\zeta]_\varphi, \qquad \xi,\zeta \in \mathrm{T}_v\mathcal{V},}$$

defines an inner product on the single tangent space $\mathrm{T}_v\mathcal{V}$.
Its matrix in the selected coordinate tangent basis is

$$\boxed{G^{\mathrm{N}}_{v,\varphi} = H_v.}$$

The corresponding pointwise sharp map is represented by

$$\boxed{\left(G^{\mathrm{N}}_{v,\varphi}\right)^{-1} = H_v^{-1}.}$$

Conceptual warning

This is a pointwise inner product relative to the selected coordinate model, not yet a coordinate-independent Riemannian metric on $\mathcal{V}$.

Newton is a gradient direction in the selected local geometry

The gradient associated with the pointwise inner product $g^{\mathrm{N}}_{v,\varphi}$ has components

$$[\operatorname{grad}_{g^{\mathrm{N}}_{\varphi}} J]_v = H_v^{-1} d_v.$$

The Newton direction has components

$$[\xi_{\mathrm{N}}]_v = -H_v^{-1} d_v.$$

Hence, in the selected local model,

$$\boxed{\xi_{\mathrm{N}} = -\operatorname{grad}_{g^{\mathrm{N}}_{\varphi}} J(v).}$$

Key idea

The coordinate Newton direction is the negative gradient associated with the pointwise inner product represented by H_v.

An intrinsic Hessian requires a chosen connection

Choose an affine connection ∇. The covariant Hessian of J at v is

$$\boxed{\mathrm{Hess}^{\nabla} J(v) := \nabla(\mathrm{d}J)_v : \mathrm{T}_v\mathcal{V} \times \mathrm{T}_v\mathcal{V} \to \mathbb{R}.}$$

When ∇ is torsion-free,

$$\mathrm{Hess}^{\nabla} J(v)$$

is symmetric.
If, in addition,

$$\boxed{\mathrm{Hess}^{\nabla} J(v)(\xi,\xi) > 0 \qquad \forall \xi \neq 0,}$$

then it defines an intrinsic pointwise inner product on $\mathrm{T}_v\mathcal{V}$.

Key idea

The covariant Hessian is an intrinsic bilinear form after the connection ∇ has been chosen.

The intrinsic Newton equation has the same algebraic pattern

The intrinsic Newton direction satisfies

$$\boxed{\operatorname{Hess}^{\nabla} J(v)\big(\xi_{\mathrm{N}}^{\nabla}, \cdot\big) = -\mathrm{d}J_v.}$$

Ordinary coordinate Hessian

$$H_v = \nabla^2 J_\varphi\big(\varphi(v)\big).$$

This matrix belongs to a selected local coordinate model.

Covariant Hessian

$$\operatorname{Hess}^{\nabla} J(v).$$

This is an intrinsic bilinear form after choosing the connection ∇.

Takeaway

The two Newton constructions have the same algebraic pattern, but only the covariant formulation is independent of the chosen chart.

What does a positive Hessian-based local geometry do to the cost model?

Within the selected model, it prices tangent displacements according to the represented second-order variation.

The identity matrix defines an isotropic coordinate penalty

Recall the coordinate representation of the regularized first-order model:

$$[\xi]_v \longmapsto d_v^\top [\xi]_v + \frac{1}{2\eta}[\xi]_v^\top G_v [\xi]_v.$$

With

$$\eta = 1, \qquad G_v = I,$$

this becomes

$$\boxed{[\xi]_v \longmapsto d_v^\top [\xi]_v + \frac{1}{2}[\xi]_v^\top [\xi]_v.}$$

Its unique minimizer is

$$\boxed{[\xi_I]_v = -d_v.}$$

Geometric interpretation

The penalty assigns the same weight to every coordinate direction, independently of the represented second-order variation of the cost.

$$[\xi]_v^\top [\xi]_v = \sum_{i=1}^{n} (\xi^i)^2.$$

The coordinate Hessian reproduces the local quadratic model

In the selected chart, assume

$$H_v = \nabla^2 J_\varphi\big(\varphi(v)\big) = H_v^\top \succ 0.$$

Choosing

$$\eta = 1, \qquad G_v = H_v,$$

gives

$$\boxed{[\xi]_v \longmapsto d_v^\top [\xi]_v + \frac{1}{2}[\xi]_v^\top H_v [\xi]_v.}$$

Apart from the constant term $J(v)$, this is the coordinate-local second-order model

$$J(v) + d_v^\top [\xi]_v + \frac{1}{2}[\xi]_v^\top H_v [\xi]_v.$$

Its unique minimizer is

$$\boxed{[\xi_{\mathrm{N}}]_v = -H_v^{-1} d_v.}$$

Takeaway

Within the selected coordinate model, the Newton displacement is the gradient displacement obtained when the quadratic penalty is chosen to match the positive local quadratic model of J_φ.

The selected Hessian prices its eigendirections differently

Because H_v is symmetric and positive definite, choose an orthonormal eigenbasis of the selected coordinate space:

$$H_v e_i = \lambda_i e_i, \qquad \lambda_i > 0.$$

For a component column

$$[\xi]_v = \sum_{i=1}^{n} a_i e_i,$$

the Hessian-weighted penalty becomes

$$\boxed{[\xi]_v^\top H_v [\xi]_v = \sum_{i=1}^{n} \lambda_i a_i^2.}$$

Large positive eigenvalue

If λ_i is large, the selected quadratic model assigns a larger penalty to displacement along e_i.

Small positive eigenvalue

If λ_i is small, the selected quadratic model assigns a smaller penalty to displacement along e_i.

Takeaway

The inverse Hessian compensates for the anisotropy represented by the selected positive quadratic model by scaling each eigendirection inversely to its eigenvalue.

An exact Euclidean quadratic makes the comparison intrinsic

Consider the Euclidean vector space $\mathbb{R}^n$ with its standard affine structure and the quadratic cost

$$J(z) = \frac{1}{2} z^\top Q z, \qquad Q = Q^\top \succ 0.$$

The differential is represented by $d_z = Qz$, and the constant positive-definite bilinear form underlying the quadratic is represented by $H_z = Q$.

With the standard Euclidean metric, $\boxed{-[\mathrm{grad}_I\, J]_z = -Qz.}$

With the metric represented by Q, $\boxed{-[\mathrm{grad}_Q\, J]_z = -Q^{-1} Q z = -z.}$

Key idea

In this vector-space example, Q represents a globally defined constant positive-definite bilinear form. The Hessian-metric interpretation therefore does not depend on choosing an arbitrary nonlinear chart.

A unit Newton step solves the exact quadratic

Because the domain is the vector space $\mathbb{R}^n$, the Newton update with unit step size is ordinary vector addition:

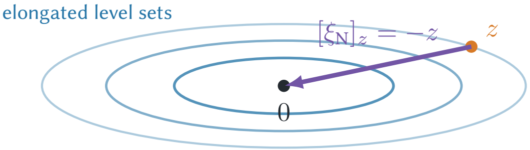
$$z_+ = z + [\xi_{\mathrm{N}}]_z.$$

For the exact positive-definite quadratic,

$$[\xi_{\mathrm{N}}]_z = -[\operatorname{grad}_Q J]_z = -z,$$

and therefore

$$\boxed{z_+ = z - z = 0.}$$



Takeaway

For an exact positive-definite quadratic on $\mathbb{R}^n$, the Q-metric normalizes the unequal positive scalings, and the unit Newton step reaches the unique minimizer.

IV

PART IV

Constraints and Multipliers

How equality constraints lead from feasible directions to conormal covectors.

What does an equality constraint change at first order?

It replaces all ambient tangent directions by the tangent directions of a regular fiber.

An equality constraint selects a fiber

Let

$$f : \mathcal{V} \to \mathcal{W}$$

be a smooth task map between smooth finite-dimensional manifolds, and let

$$w \in \mathcal{W}$$

be a prescribed task value.
The equality constraint

$$f(v) = w$$

selects the fiber

$$\boxed{\mathcal{F}_w := f^{-1}(w) = \{v \in \mathcal{V} : f(v) = w\}\,.}$$

Geometric interpretation

The ambient manifold $\mathcal{V}$ contains all candidate points. The fiber $\mathcal{F}_w \subset \mathcal{V}$ contains exactly the points satisfying the prescribed task value.

A regular value produces an embedded submanifold

Assume that w is a regular value of

$$f : \mathcal{V} \to \mathcal{W}.$$

Equivalently,

$$\boxed{\mathrm{d}f_v : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_w\mathcal{W} \quad \text{is surjective for every } v \in \mathcal{F}_w.}$$

The regular-value theorem then guarantees that

$$\boxed{\mathcal{F}_w = f^{-1}(w)}$$

is an embedded submanifold of $\mathcal{V}$.

Key idea

The regular-value theorem guarantees that $\mathcal{F}_w$ is an embedded submanifold of $\mathcal{V}$.

Feasible tangent directions form a kernel

Let $v \in \mathcal{F}_w$. A curve

$$\gamma : (-\varepsilon, \varepsilon) \to \mathcal{F}_w, \qquad \gamma(0) = v,$$

remains feasible because

$$f\big(\gamma(t)\big) = w.$$

Differentiating at $t = 0$ gives

$$\mathrm{d}f_v\big(\dot\gamma(0)\big) = 0.$$

The regular-fiber theorem yields

$$\boxed{\mathrm{T}_v\mathcal{F}_w = \ker \mathrm{d}f_v.}$$

Geometric interpretation

The nonlinear constraint $f(v) = w$ becomes the linear first-order feasibility condition

$$\mathrm{d}f_v(\xi) = 0.$$

The constrained cost is a restriction

Let

$$\iota_w : \mathcal{F}_w \hookrightarrow \mathcal{V}$$

denote the inclusion map. The constrained cost is

$$\boxed{J|_{\mathcal{F}_w} = J \circ \iota_w : \mathcal{F}_w \to \mathbb{R}.}$$

At $v \in \mathcal{F}_w$, the differential of the inclusion is

$$\mathrm{d}\iota_{w,v} : \mathrm{T}_v\mathcal{F}_w \to \mathrm{T}_v\mathcal{V}.$$

By the chain rule, the differential is the composition of the two differentials:

$$\boxed{\mathrm{d}(J|_{\mathcal{F}_w})_v = \mathrm{d}J_v \circ \mathrm{d}\iota_{w,v}}.$$

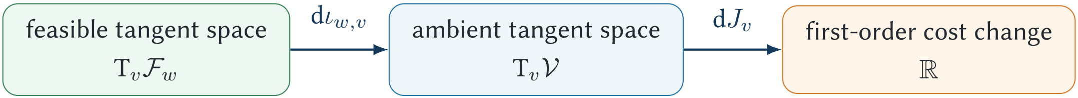
Applying the definition of the adjoint map, this composition is equivalent to the pullback of the covector $\mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}$:

$$\boxed{\mathrm{d}(J|_{\mathcal{F}_w})_v = \mathrm{d}J_v \circ \mathrm{d}\iota_{w,v} = (\mathrm{d}\iota_{w,v})^*(\mathrm{d}J_v) \in \mathrm{T}_v^*\mathcal{F}_w.}$$

The restricted differential acts on feasible directions

The tangent-space composition underlying the restricted differential is

$$\mathrm{T}_v\mathcal{F}_w \xrightarrow{\mathrm{d}\iota_{w,v}} \mathrm{T}_v\mathcal{V} \xrightarrow{\mathrm{d}J_v} \mathbb{R}.$$



For every $\xi \in \mathrm{T}_v\mathcal{F}_w$,

$$\mathrm{d}(J|_{\mathcal{F}_w})_v(\xi) = \mathrm{d}J_v\big(\mathrm{d}\iota_{w,v}(\xi)\big).$$

Coordinate realization

The covector $\mathrm{d}(J|_{\mathcal{F}_w})_v$ is the restriction of the ambient differential to feasible tangent directions.

Constrained stationarity tests only feasible directions

A point $v \in \mathcal{F}_w$ is stationary for the restricted cost if

$$\boxed{\mathrm{d}(J|_{\mathcal{F}_w})_v = 0 \in \mathrm{T}_v^*\mathcal{F}_w.}$$

Using

$$J|_{\mathcal{F}_w} = J \circ \iota_w,$$

we obtain

$$\begin{aligned} v \text{ stationary for } J|_{\mathcal{F}_w} &\iff \mathrm{d}(J|_{\mathcal{F}_w})_v = 0 \\ &\iff (\mathrm{d}\iota_{w,v})^*(\mathrm{d}J_v) = 0. \end{aligned}$$

Since

$$\mathrm{T}_v\mathcal{F}_w = \ker \mathrm{d}f_v,$$

this condition concerns exactly the feasible tangent directions.

Constrained stationarity requires restricted vanishing

The remaining equivalences are

$$\boxed{\begin{aligned}(\mathrm{d}\iota_{w,v})^*(\mathrm{d}J_v) = 0 &\iff \mathrm{d}J_v(\xi) = 0 \quad \forall \xi \in \mathrm{T}_v\mathcal{F}_w \\ &\iff \mathrm{d}J_v\big|_{\mathrm{T}_v\mathcal{F}_w} = 0.\end{aligned}}$$

Unconstrained stationarity

$$\mathrm{d}J_v(\xi) = 0 \qquad \forall \xi \in \mathrm{T}_v\mathcal{V}.$$

The differential vanishes on every ambient tangent direction.

Constrained stationarity

$$\mathrm{d}J_v(\xi) = 0 \qquad \forall \xi \in \mathrm{T}_v\mathcal{F}_w.$$

The differential vanishes only on feasible tangent directions.

Takeaway

Constrained stationarity does not require $\mathrm{d}J_v = 0$. It requires $\mathrm{d}J_v$ to annihilate the feasible tangent space.

Where does the Lagrange multiplier live geometrically?

In the cotangent space of the task-value manifold.

Constrained stationarity is a conormal condition

For a linear subspace of the ambient tangent space

$$S \subset \mathrm{T}_v\mathcal{V},$$

its annihilator is

$$S^\circ := \{\alpha \in \mathrm{T}_v^*\mathcal{V} : \alpha(\xi) = 0 \quad \forall \xi \in S\}.$$

Applying this definition to the feasible tangent space gives

$$\boxed{\mathrm{d}J_v|_{\mathrm{T}_v\mathcal{F}_w} = 0 \quad \Longleftrightarrow \quad \mathrm{d}J_v \in (\mathrm{T}_v\mathcal{F}_w)^\circ.}$$

The space

$$\boxed{N_v^*\mathcal{F}_w := (\mathrm{T}_v\mathcal{F}_w)^\circ \subset \mathrm{T}_v^*\mathcal{V}}$$

is the conormal space of the fiber $\mathcal{F}_w$ at v.

Takeaway

At a constrained stationary point, the cost differential is a conormal covector: it vanishes on every feasible infinitesimal motion.

The conormal bundle

For the embedded constraint fiber

$$\mathcal{F}_w \subset \mathcal{V},$$

the pointwise conormal spaces assemble into

$$\boxed{N^*\mathcal{F}_w := \bigsqcup_{v\in\mathcal{F}_w} (\mathrm{T}_v\mathcal{F}_w)^\circ \subset \mathrm{T}^*\mathcal{V}\big|_{\mathcal{F}_w}.}$$

Tangent bundle

$$\mathrm{T}\mathcal{F}_w = \bigsqcup_{v\in\mathcal{F}_w} \mathrm{T}_v\mathcal{F}_w.$$

Feasible infinitesimal motions.

Conormal bundle

$$N^*\mathcal{F}_w = \bigsqcup_{v\in\mathcal{F}_w} (\mathrm{T}_v\mathcal{F}_w)^\circ.$$

Covectors annihilating those motions.

$$\boxed{N_v^*\mathcal{F}_w = (\mathrm{T}_v\mathcal{F}_w)^\circ \subset \mathrm{T}_v^*\mathcal{V}.}$$

Stationarity is conormal-bundle membership

Globally, along the fiber, the differential defines a covector field (a differential 1-form), acting as a section of the restricted cotangent bundle:

$$\mathrm{d}J\big|_{\mathcal{F}_w} : \mathcal{F}_w \longrightarrow \mathrm{T}^*\mathcal{V}\big|_{\mathcal{F}_w}, \qquad v \longmapsto (v, \mathrm{d}J_v).$$

At $v \in \mathcal{F}_w$,

$$\begin{aligned} v \text{ stationary for } J|_{\mathcal{F}_w} &\iff \mathrm{d}J_v(\xi) = 0 \quad \forall \xi \in \mathrm{T}_v\mathcal{F}_w \\ &\iff \mathrm{d}J_v \in (\mathrm{T}_v\mathcal{F}_w)^\circ \\ &\iff \boxed{(v, \mathrm{d}J_v) \in N^*\mathcal{F}_w.} \end{aligned}$$

Dual maps generate annihilators of kernels

Let

$$L : E \to F$$

be a linear map between finite-dimensional vector spaces. Its dual map is

$$L^* : F^* \to E^*, \qquad L^*(\mu) = \mu \circ L.$$

The fundamental identity is

$$\boxed{\operatorname{Im} L^* = (\ker L)^\circ.}$$

Applying it to

$$L = \mathrm{d}f_v : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_w\mathcal{W}$$

and using

$$\mathrm{T}_v\mathcal{F}_w = \ker \mathrm{d}f_v$$

gives

$$\boxed{(\mathrm{T}_v\mathcal{F}_w)^\circ = \operatorname{Im}(\mathrm{d}f_v)^*.}$$

Geometric interpretation

Every covector that vanishes on feasible tangent directions is the pullback of a task-space covector.

The multiplier is a task-space covector

Combining constrained stationarity with the annihilator identity gives

$$
\begin{aligned}
v \text{ stationary for } J|_{\mathcal{F}_w} &\iff \mathrm{d}J_v \in (\mathrm{T}_v\mathcal{F}_w)^\circ \\
&\iff \mathrm{d}J_v \in \mathrm{Im}(\mathrm{d}f_v)^*.
\end{aligned}
$$

Therefore, there exists

$$\mu \in \mathrm{T}_w^*\mathcal{W}$$

such that

$$\boxed{\mathrm{d}J_v = (\mathrm{d}f_v)^*\mu.}$$

Because w is a regular value,

$$\mathrm{d}f_v : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_w\mathcal{W}$$

is surjective. Hence $(\mathrm{d}f_v)^*$ is injective, and μ is unique.

Key idea

The multiplier μ is a covector on the task-value space whose pullback equals the cost differential.

Conceptual warning

The multiplier μ is not initially a gradient vector.

The geometric and conventional multipliers differ by a sign (1/2)

When $\mathcal{W}$ is a vector space, the conventional plus-sign Lagrangian is

$$\boxed{\mathcal{L}(v, \lambda; w) = J(v) + \langle \lambda, f(v) - w \rangle\,, \qquad \lambda \in \mathcal{W}^*.}$$

Differentiation with respect to v gives

$$\boxed{\mathrm{d}_v\mathcal{L} = \mathrm{d}J_v + (\mathrm{d}f_v)^*\lambda.}$$

Thus Lagrangian stationarity is

$$\boxed{\mathrm{d}J_v + (\mathrm{d}f_v)^*\lambda = 0.}$$

The geometric and conventional multipliers differ by a sign (2/2)

Starting from the plus-sign Lagrangian stationarity equation

$$\mathrm{d}J_v + (\mathrm{d}f_v)^*\lambda = 0,$$

define

$$\boxed{\mu := -\lambda.}$$

The stationarity equation then becomes

$$\boxed{\mathrm{d}J_v = (\mathrm{d}f_v)^*\mu,}$$

which is precisely the intrinsic pullback equation.

Takeaway

The geometric multiplier and the conventional plus-sign Lagrangian multiplier satisfy

$$\mu = -\lambda.$$

Why does the Jacobian transpose appear, and what does “normal” require?

The transpose represents the pullback;
a metric turns conormal covectors into normal vectors.

The Jacobian transpose represents the pullback

Choose charts $\varphi : U \subset \mathcal{V} \to \mathbb{R}^n, \qquad \psi : Z \subset \mathcal{W} \to \mathbb{R}^m,$

around $v \in \mathcal{V}$ and $w = f(v) \in \mathcal{W}$.

Let F_v be the matrix representing $\mathrm{d}f_v : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_w\mathcal{W}$.

Then
$$[\mathrm{d}f_v(\xi)]_\psi = F_v[\xi]_\varphi.$$

For $\mu \in \mathrm{T}^*_w\mathcal{W}$,
$$\begin{aligned}\big((\mathrm{d}f_v)^*\mu\big)(\xi) = \mu\big(\mathrm{d}f_v(\xi)\big) &= [\mu]_\psi^\top F_v[\xi]_\varphi \\ &= \big(F_v^\top[\mu]_\psi\big)^\top[\xi]_\varphi.\end{aligned}$$

Therefore,
$$\boxed{[(\mathrm{d}f_v)^*\mu]_\varphi = F_v^\top[\mu]_\psi.}$$

Coordinate realization

The differential pushes tangent vectors forward through F_v. The dual map pulls covectors backward through $F_v^\top$.

Conormality is intrinsic; normality requires a metric

The conormal space of the fiber at $v \in \mathcal{F}_w$ is

$$\boxed{N_v^* \mathcal{F}_w = (\mathrm{T}_v \mathcal{F}_w)^\circ \subset \mathrm{T}_v^* \mathcal{V}.}$$

After choosing a Riemannian metric g, the metric normal space is

$$\boxed{N_v^g \mathcal{F}_w = (\mathrm{T}_v \mathcal{F}_w)^{\perp_g} \subset \mathrm{T}_v \mathcal{V}.}$$

The sharp map relates the two spaces:

$$\boxed{\sharp_{g,v}\left(N_v^* \mathcal{F}_w\right) = N_v^g \mathcal{F}_w.}$$

The sharp map converts conormals into metric normals

Let

$$\alpha \in N_v^* \mathcal{F}_w, \qquad u = \sharp_{g,v}(\alpha).$$

Since

$$\alpha \in (\mathrm{T}_v \mathcal{F}_w)^\circ,$$

for every $\xi \in \mathrm{T}_v \mathcal{F}_w$,

$$g_v(u, \xi) = \alpha(\xi) = 0.$$

Therefore,

$$u \in (\mathrm{T}_v \mathcal{F}_w)^{\perp_g} = N_v^g \mathcal{F}_w.$$

Takeaway

The primary metric-free statement is

$$\mathrm{d}J_v \in (\mathrm{T}_v \mathcal{F}_w)^\circ.$$

Only after choosing g can this be expressed as

$$\operatorname{grad}_g J(v) \in (\mathrm{T}_v \mathcal{F}_w)^{\perp_g}.$$

What does the multiplier say when the prescribed task value changes?

The multiplier measures how the optimal cost changes between nearby constraint fibers.

Varying the task value produces nearby fibers

Let

$$U \subset \mathcal{W}$$

be a neighborhood of the prescribed task value $w \in \mathcal{W}$.
For each $w' \in U$, define the corresponding constraint fiber

$$\boxed{\mathcal{F}_{w'} = f^{-1}(w').}$$

The associated constrained problem is

$$\boxed{\min_{v \in \mathcal{F}_{w'}} J(v).}$$

Thus, varying w' produces a family of optimization problems:

$$w' \quad \longmapsto \quad \mathcal{F}_{w'} \quad \longmapsto \quad \min_{v \in \mathcal{F}_{w'}} J(v).$$

Geometric interpretation

The task value w' varies in $\mathcal{W}$, while each fiber $\mathcal{F}_{w'}$ is a subset of the total-space manifold $\mathcal{V}$.

Select one minimizer on each nearby fiber

Assume that each nearby problem admits a selected minimizer and that these minimizers define a differentiable map

$$\boxed{\widehat{v} : U \to \mathcal{V}.}$$

For every $w' \in U$,

$$\widehat{v}(w') \in \mathcal{F}_{w'},$$

or equivalently,

$$\boxed{f\left(\widehat{v}(w')\right) = w'.}$$

The selection therefore satisfies

$$\boxed{f \circ \widehat{v} = \mathrm{Id}_U\,.}$$

Input

$$w' \in U \subset \mathcal{W}$$

prescribed task value

Output

$$\widehat{v}(w') \in \mathcal{F}_{w'}$$

selected constrained minimizer

The optimal value is a function of the task value

The selected minimizer induces the optimal-value function

$$\boxed{\mathcal{J}^{\mathrm{opt}} : U \to \mathbb{R}, \qquad \mathcal{J}^{\mathrm{opt}}(w') = J\big(\widehat{v}(w')\big).}$$

Equivalently,

$$\mathcal{J}^{\mathrm{opt}} = J \circ \widehat{v}.$$

Under the assumed selection of minimizers,

$$\boxed{\mathcal{J}^{\mathrm{opt}}(w') = \min_{v \in \mathcal{F}_{w'}} J(v).}$$

The optimizer and optimal value are different maps

Optimizer map

$$\widehat{v} : U \to \mathcal{V}$$

assigns a selected minimizing point to each task value.

Optimal-value map

$$\mathcal{J}^{\mathrm{opt}} : U \to \mathbb{R}$$

assigns the corresponding minimum cost.

The two maps are related by

$$\boxed{\mathcal{J}^{\mathrm{opt}} = J \circ \widehat{v}.}$$

Takeaway

The differential

$$\mathrm{d}\mathcal{J}^{\mathrm{opt}}_w \in \mathrm{T}^*_w\mathcal{W}$$

will measure the first-order sensitivity of the optimal cost to changes in the prescribed task value.

Why is the multiplier the differential of the optimal-value function?

Differentiated feasibility converts variations of the optimizer into variations of the prescribed task value.

Differentiate the feasibility identity

The selected minimizers satisfy

$$f \circ \widehat{v} = \mathrm{Id}_U \, .$$

At $w \in U$, the chain rule gives

$$\mathrm{d}f_{\widehat{v}(w)} \circ \mathrm{d}\widehat{v}_w = \mathrm{Id}_{\mathrm{T}_w \mathcal{W}} \, .$$

Therefore, for every task variation

$$\delta w \in \mathrm{T}_w \mathcal{W},$$

one has

$$\boxed{\mathrm{d}f_{\widehat{v}(w)} \left(\mathrm{d}\widehat{v}_w[\delta w]\right) = \delta w.}$$

$$\mathrm{T}_w \mathcal{W} \xrightarrow{\mathrm{d}\widehat{v}_w} \mathrm{T}_{\widehat{v}(w)} \mathcal{V} \xrightarrow{\mathrm{d}f_{\widehat{v}(w)}} \mathrm{T}_w \mathcal{W}.$$

Geometric interpretation

The variation $\mathrm{d}\widehat{v}_w[\delta w]$ of the selected optimizer is mapped by the task differential to the prescribed task variation δw.

Differentiate the optimal-value function

The optimal-value function is

$$\mathcal{J}^{\mathrm{opt}} = J \circ \widehat{v}.$$

Its differential at w is

$$\boxed{\mathrm{d}\mathcal{J}^{\mathrm{opt}}_w = (\mathrm{d}\widehat{v}_w)^* \left(\mathrm{d}J_{\widehat{v}(w)}\right) \in \mathrm{T}^*_w\mathcal{W}.}$$

Equivalently, for every $\delta w \in \mathrm{T}_w\mathcal{W}$,

$$\boxed{\mathrm{d}\mathcal{J}^{\mathrm{opt}}_w(\delta w) = \mathrm{d}J_{\widehat{v}(w)} \left(\mathrm{d}\widehat{v}_w[\delta w]\right).}$$

$$\delta w \xmapsto{\mathrm{d}\widehat{v}_w} \mathrm{d}\widehat{v}_w[\delta w] \xmapsto{\mathrm{d}J_{\widehat{v}(w)}} \mathrm{d}\mathcal{J}^{\mathrm{opt}}_w(\delta w).$$

Key idea

A task variation induces a tangent variation of the selected optimizer. The cost differential evaluates that induced tangent variation.

Insert the multiplier equation

At the selected constrained minimizer $\widehat{v}(w)$, let

$$\mu \in \mathrm{T}_w^* \mathcal{W}$$

be the geometric multiplier satisfying

$$\mathrm{d}J_{\widehat{v}(w)} = \left(\mathrm{d}f_{\widehat{v}(w)}\right)^* \mu.$$

For every $\delta w \in \mathrm{T}_w \mathcal{W}$,

$$\begin{aligned} \mathrm{d}\mathcal{J}_w^{\mathrm{opt}}(\delta w) &= \mathrm{d}J_{\widehat{v}(w)}\left(\mathrm{d}\widehat{v}_w[\delta w]\right) \\ &= \mu\left(\mathrm{d}f_{\widehat{v}(w)}\left(\mathrm{d}\widehat{v}_w[\delta w]\right)\right) \\ &= \boxed{\mu(\delta w).} \end{aligned}$$

The last equality follows from differentiated feasibility:

$$\mathrm{d}f_{\widehat{v}(w)}\left(\mathrm{d}\widehat{v}_w[\delta w]\right) = \delta w.$$

The multiplier is the sensitivity covector

Since

$$\mathrm{d}\mathcal{J}_w^{\mathrm{opt}}(\delta w) = \mu(\delta w) \qquad \forall \delta w \in \mathrm{T}_w\mathcal{W},$$

the two covectors are equal:

$$\boxed{\mathrm{d}\mathcal{J}_w^{\mathrm{opt}} = \mu = -\lambda.}$$

Thus,

$$\boxed{\mu(\delta w)}$$

is the first-order change in optimal cost caused by the task variation

$$\delta w \in \mathrm{T}_w\mathcal{W}.$$

Sensitivity interpretation

The geometric multiplier is not merely an algebraic variable used to express stationarity. It is the differential of the optimal-value function with respect to the prescribed task value.

Caveat: the sensitivity relation requires local regularity

The identity

$$\boxed{\mathrm{d}\mathcal{J}_w^{\mathrm{opt}} = \mu = -\lambda}$$

was derived from a differentiable selection

$$\widehat{v} : U \to \mathcal{V}$$

of minimizers satisfying

$$f\left(\widehat{v}(w')\right) = w' \qquad \forall w' \in U.$$

Required assumptions

The conclusion requires nearby selected minimizers and a differentiable selection

$$\widehat{v} : U \to \mathcal{V}.$$

A multiplier at one isolated constrained minimizer does not by itself guarantee differentiability of $\mathcal{J}^{\mathrm{opt}}$.

Takeaway

The multiplier equation gives a stationarity condition at one fiber. Its sensitivity interpretation additionally requires a differentiable family of nearby constrained minimizers.

PART V

Geometric Optimization Algorithms

From tangent descent directions to finite feasible iterations.

Why does a tangent descent direction not yet define the next iterate?

A tangent vector is an infinitesimal displacement,
not a new point of the manifold.

A descent direction is not a manifold point

At $v \in \mathcal{V}$, the negative metric gradient is a tangent vector:

$$\boxed{-\operatorname{grad}_g J(v) \in \mathrm{T}_v\mathcal{V}.}$$

It represents an infinitesimal descent direction, but it is not itself a point of the manifold:

$$-\operatorname{grad}_g J(v) \notin \mathcal{V} \qquad \text{in general.}$$

Likewise, on a nonlinear manifold the expression

$$v - \alpha \operatorname{grad}_g J(v)$$

is not an intrinsically defined addition of geometric objects.

A numerical update must connect different spaces

Current point

$$v \in \mathcal{V}$$

is a point of the manifold.

Descent displacement

$$-\alpha \operatorname{grad}_g J(v) \in \mathrm{T}_v\mathcal{V}$$

is a tangent vector at v.

The required update mechanism must therefore have the type

$$\boxed{\mathrm{T}_v\mathcal{V} \supset \mathcal{U}_v \longrightarrow \mathcal{V}.}$$

Takeaway

A numerical update requires an additional construction that maps a tangent displacement back to the manifold.

A retraction maps tangent displacements to points

Let

$$\mathcal{U} \subset \mathrm{T}\mathcal{V}$$

be an open neighborhood of the zero section.
A retraction on $\mathcal{V}$ is a smooth map

$$\boxed{R : \mathcal{U} \subset \mathrm{T}\mathcal{V} \to \mathcal{V}.}$$

For each $v \in \mathcal{V}$, its restriction to the tangent space at v is

$$\boxed{R_v : \mathcal{U}_v \subset \mathrm{T}_v\mathcal{V} \to \mathcal{V}, \qquad R_v(\xi) := R(v, \xi).}$$

Geometric interpretation

The map R_v converts a sufficiently small tangent displacement at v into a point of the manifold.

Takeaway

The retraction has a tangent vector as its input and a manifold point as its output.

A retraction reproduces tangent displacement to first order

For every $v \in \mathcal{V}$, a retraction satisfies

$$\boxed{R_v(0_v) = v,}$$

and

$$\boxed{\mathrm{d}(R_v)_{0_v} = \mathrm{Id}_{\mathrm{T}_v\mathcal{V}}\,.}$$

The canonical identification

$$\mathrm{T}_{0_v}(\mathrm{T}_v\mathcal{V}) \simeq \mathrm{T}_v\mathcal{V}$$

is understood in the second identity.

Zero displacement

$$R_v(0_v) = v.$$

Applying no tangent displacement leaves the point unchanged.

First-order agreement

$$\mathrm{d}(R_v)_{0_v} = \mathrm{Id}_{\mathrm{T}_v\mathcal{V}}\,.$$

The retraction preserves the tangent displacement to first order.

A retraction realizes the gradient step

In a local chart φ around v, the retraction satisfies

$$\varphi\big(R_v(\xi)\big) = \varphi(v) + [\xi]_\varphi + o\left(\|[\xi]_\varphi\|\right).$$

Thus the intrinsic metric-gradient iteration is

$$\boxed{v_{k+1} = R_{v_k}\Big(-\alpha_k \operatorname{grad}_g J(v_k)\Big).}$$

$$\begin{tikzcd} \begin{array}{c}\text{current point}\\ v_k \in \mathcal{V}\end{array} \arrow[r, "\text{select direction}"] & \begin{array}{c}\text{tangent displacement}\\ -\alpha_k \operatorname{grad}_g J(v_k)\end{array} \arrow[r, "R_{v_k}"] & \begin{array}{c}\text{next point}\\ v_{k+1} \in \mathcal{V}\end{array} \end{tikzcd}$$

Takeaway

The metric selects the tangent descent direction. The retraction realizes the finite displacement as a new manifold point.

How is an ambient descent direction made feasible to first order?

The ambient gradient is projected onto
the tangent space of the constraint fiber.

The metric splits ambient tangent vectors

Assume that the fiber $\mathcal{F}_w$ carries the metric induced by the ambient metric g. At every $v \in \mathcal{F}_w$,

$$\boxed{\mathrm{T}_v\mathcal{V} = \mathrm{T}_v\mathcal{F}_w \oplus (\mathrm{T}_v\mathcal{F}_w)^{\perp_g}.}$$

Consequently, every ambient tangent vector

$$\eta \in \mathrm{T}_v\mathcal{V}$$

admits a unique decomposition

$$\boxed{\eta = \eta^{\mathrm{T}} + \eta^{\mathrm{N}},}$$

where:

Tangent component

$$\eta^{\mathrm{T}} \in \mathrm{T}_v\mathcal{F}_w$$

is feasible to first order.

Normal component

$$\eta^{\mathrm{N}} \in (\mathrm{T}_v\mathcal{F}_w)^{\perp_g}$$

is g-orthogonal to every feasible tangent direction.

Tangent projection extracts the feasible component

The metric-orthogonal projection onto the feasible tangent space is the linear map

$$\boxed{\mathrm{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w} : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_v\mathcal{F}_w}$$

defined by

$$\boxed{\mathrm{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w}(\eta) = \eta^{\mathrm{T}}.}$$

Equivalently, it is characterized by

$$\boxed{\eta - \mathrm{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w}(\eta) \in (\mathrm{T}_v\mathcal{F}_w)^{\perp_g}.}$$

Thus,

$$\eta = \mathrm{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w}(\eta) + \eta^{\mathrm{N}}.$$

Takeaway

The projection depends on g, because the metric determines the normal complement

$$(\mathrm{T}_v\mathcal{F}_w)^{\perp_g}.$$

The feasible tangent space $\mathrm{T}_v\mathcal{F}_w$ itself is metric-independent.

The restricted gradient is the projected ambient gradient

The metric induced on $\mathcal{F}_w$ defines the gradient of the restricted cost

$$J|_{\mathcal{F}_w} : \mathcal{F}_w \to \mathbb{R}.$$

It is obtained by projecting the ambient gradient:

$$\boxed{\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) = \operatorname{Proj}^g_{\mathrm{T}_v\mathcal{F}_w}\big(\operatorname{grad}_g J(v)\big).}$$

Thus,

$$\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) \in \mathrm{T}_v\mathcal{F}_w.$$

The normal component discarded by the projection belongs to

$$(\mathrm{T}_v\mathcal{F}_w)^{\perp_g}.$$

The projected vector represents the restricted differential

For every feasible tangent vector

$$\xi \in \mathrm{T}_v\mathcal{F}_w,$$

the discarded normal component is g-orthogonal to ξ. Hence

$$\boxed{g_v\left(\mathrm{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v), \xi\right) = \mathrm{d}J_v(\xi).}$$

Since

$$\mathrm{d}(J|_{\mathcal{F}_w})_v(\xi) = \mathrm{d}J_v(\xi),$$

we obtain

$$\boxed{g_v\left(\mathrm{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v), \xi\right) = \mathrm{d}(J|_{\mathcal{F}_w})_v(\xi)}$$

for every $\xi \in \mathrm{T}_v\mathcal{F}_w$.

Key idea

Projection changes the ambient gradient into the unique feasible tangent vector representing the restricted differential under the induced metric.

The metric plays two compatible roles

For constrained optimization on

$$\mathcal{F}_w = f^{-1}(w), \qquad \mathrm{T}_v\mathcal{F}_w = \ker \mathrm{d} f_v,$$

an ambient metric g enters through two conceptually distinct constructions.

Representing the cost differential

The metric converts the ambient covector into a tangent vector:

$$\operatorname{grad}_g J(v) = \sharp_{g,v}(\mathrm{d}J_v).$$

This role concerns the cost J.

Splitting feasible and normal directions

The metric defines the orthogonal decomposition

$$\mathrm{T}_v\mathcal{V} = \mathrm{T}_v\mathcal{F}_w \oplus (\mathrm{T}_v\mathcal{F}_w)^{\perp_g}$$

and hence the tangent projection. This role concerns the constraint fiber.

Compatibility obtained from one metric

Using the same metric in both roles gives

$$\boxed{\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) = \mathrm{Proj}^g_{\mathrm{T}_v\mathcal{F}_w}\big(\operatorname{grad}_g J(v)\big).}$$

Two different metrics could define the two operations, but the resulting projected vector would not, in general, be the gradient of the restricted cost under either induced metric.

The negative restricted gradient is descending

If

$$\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) \neq 0,$$

then

$$\begin{aligned}
\mathrm{d}(J|_{\mathcal{F}_w})_v\left(-\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v)\right) &= -g_v\left(\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v), \operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v)\right) \\
&= \boxed{-\left\|\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v)\right\|_g^2 < 0.}
\end{aligned}$$

Descent

The restricted cost decreases to first order along the negative restricted gradient.

Takeaway

The restricted gradient is the intrinsic first-order search direction for optimization on the constraint fiber.

How do projected gradients and multipliers characterize the same stationary points?

The projected gradient supplies a search direction; the multiplier equation characterizes when that direction vanishes.

Stationarity means that the restricted gradient vanishes

A point $v \in \mathcal{F}_w$ is stationary for the restricted cost precisely when

$$\boxed{\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) = 0.}$$

Using the projected-gradient formula,

$$\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) = \operatorname{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w}\big(\operatorname{grad}_g J(v)\big),$$

this condition becomes

$$\boxed{\operatorname{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w}\big(\operatorname{grad}_g J(v)\big) = 0.}$$

Therefore, at a constrained stationary point, the ambient gradient has no feasible tangent component.

Takeaway

The restricted gradient is the feasible first-order search direction. Constrained stationarity occurs exactly when that direction vanishes.

Vanishing projection means metric normality

The tangent projection vanishes if and only if the ambient gradient belongs entirely to the metric normal space:

$$\boxed{\mathrm{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w}\big(\operatorname{grad}_g J(v)\big) = 0 \quad\Longleftrightarrow\quad \operatorname{grad}_g J(v) \in (\mathrm{T}_v\mathcal{F}_w)^{\perp_g}.}$$

Applying the flat map gives

$$\flat_{g,v}\big(\operatorname{grad}_g J(v)\big) \in (\mathrm{T}_v\mathcal{F}_w)^{\circ}.$$

Since

$$\flat_{g,v}\big(\operatorname{grad}_g J(v)\big) = \mathrm{d}J_v,$$

we obtain

$$\boxed{\operatorname{grad}_g J(v) \in (\mathrm{T}_v\mathcal{F}_w)^{\perp_g} \quad\Longleftrightarrow\quad \mathrm{d}J_v \in (\mathrm{T}_v\mathcal{F}_w)^{\circ}.}$$

Geometric interpretation

Metric normality is the vector representation of the intrinsic conormal condition.

Four equivalent characterizations of stationarity

For every $v \in \mathcal{F}_w$,

$$\boxed{\begin{aligned} \operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) = 0 &\iff \mathrm{d}J_v|_{\mathrm{T}_v\mathcal{F}_w} = 0 \\ &\iff \mathrm{d}J_v \in (\mathrm{T}_v\mathcal{F}_w)^\circ \\ &\iff \mathrm{d}J_v = (\mathrm{d}f_v)^*\mu \end{aligned}}$$

for a unique multiplier $\boxed{\mu \in \mathrm{T}_w^*\mathcal{W}.}$

These conditions determine the same constrained stationary points.

Algorithmic realization

$$\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) = 0.$$

The feasible gradient search direction vanishes.

Dual characterization

$$\mathrm{d}J_v = (\mathrm{d}f_v)^*\mu.$$

The cost differential is the pullback of a task-space covector.

Takeaway

Projected gradients and multiplier covectors are complementary realizations of the same intrinsic constrained-stationarity condition.

Does a feasible tangent direction automatically produce a feasible finite step?

No. Projection gives first-order feasibility;
a fiber retraction returns an actual feasible point.

Tangent feasibility is only a first-order property

Let

$$v \in \mathcal{F}_w, \qquad \xi \in \mathrm{T}_v\mathcal{F}_w.$$

The tangent vector ξ is feasible to first order:

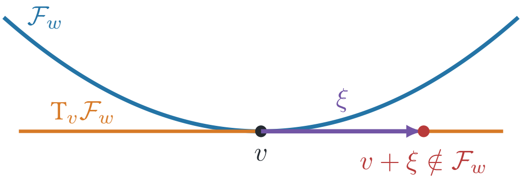
$$\boxed{\mathrm{d}f_v(\xi) = 0.}$$

On a curved fiber, however, a finite ambient addition generally gives

$$\boxed{v + \xi \notin \mathcal{F}_w.}$$

Thus, membership in $\mathrm{T}_v\mathcal{F}_w$ guarantees only first-order feasibility.

A finite tangent displacement can leave the fiber



Takeaway

The tangent space is a first-order approximation of the fiber. A finite tangent displacement must still be mapped back to $\mathcal{F}_w$.

Retractions on the constraint fiber

We apply the retraction construction to the constraint fiber $\mathcal{F}_w \subset \mathcal{V}$.

A retraction on $\mathcal{F}_w$ is a smooth map

$$\boxed{R^{\mathcal{F}_w} : \mathcal{U}^{\mathcal{F}_w} \to \mathcal{F}_w,}$$

where $\mathcal{U}^{\mathcal{F}_w} \subset \mathrm{T}\mathcal{F}_w$ is an open set containing all zero tangent vectors:

$$\{0_v : v \in \mathcal{F}_w\} \subset \mathcal{U}^{\mathcal{F}_w}.$$

For each $v \in \mathcal{F}_w$, its restriction is

$$\boxed{R_v^{\mathcal{F}_w} : \mathcal{U}_v^{\mathcal{F}_w} \subset \mathrm{T}_v\mathcal{F}_w \to \mathcal{F}_w.}$$

It satisfies

$$\boxed{R_v^{\mathcal{F}_w}(0_v) = v, \qquad \mathrm{d}(R_v^{\mathcal{F}_w})_{0_v} = \mathrm{Id}_{\mathrm{T}_v\mathcal{F}_w}.}$$

Geometric interpretation

The input of $R_v^{\mathcal{F}_w}$ is a feasible tangent displacement. Its output is an actual point of the constraint fiber.

The constrained update uses a fiber retraction

The negative restricted gradient is a feasible tangent direction:

$$-\operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v_k) \in \mathrm{T}_{v_k}\mathcal{F}_w.$$

A constrained metric-gradient iteration is

$$\boxed{v_{k+1} = R^{\mathcal{F}_w}_{v_k}\left(-\alpha_k \operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v_k)\right).}$$

Using the projected ambient gradient,

$$\boxed{v_{k+1} = R^{\mathcal{F}_w}_{v_k}\left(-\alpha_k \operatorname{Proj}^g_{\mathrm{T}_{v_k}\mathcal{F}_w}\left(\operatorname{grad}_g J(v_k)\right)\right).}$$

Tangent feasibility

$$\operatorname{Proj}^g_{\mathrm{T}_{v_k}\mathcal{F}_w}\left(\operatorname{grad}_g J(v_k)\right) \in \mathrm{T}_{v_k}\mathcal{F}_w.$$

Finite feasibility

$$v_{k+1} \in \mathcal{F}_w.$$

Projection and retraction have different types

The complete feasible-update chain is

$$\boxed{\operatorname{grad}_g J(v) \xmapsto{\mathrm{Proj}^g_{\mathrm{T}_v\mathcal{F}_w}} \operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) \xmapsto{-\alpha} \xi \xmapsto{R_v^{\mathcal{F}_w}} v_+.}$$

The objects belong to different spaces:

$$\operatorname{grad}_g J(v) \in \mathrm{T}_v\mathcal{V}, \qquad \operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v),\ \xi \in \mathrm{T}_v\mathcal{F}_w, \qquad \text{and} \quad v_+ \in \mathcal{F}_w.$$

The two geometric operations have different types:

$$\boxed{\mathrm{Proj}^g_{\mathrm{T}_v\mathcal{F}_w} : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_v\mathcal{F}_w,}$$

whereas

$$\boxed{R_v^{\mathcal{F}_w} : \mathcal{U}_v^{\mathcal{F}_w} \subset \mathrm{T}_v\mathcal{F}_w \to \mathcal{F}_w.}$$

Takeaway

Projection acts between tangent spaces at a fixed point. Retraction maps a tangent displacement to a new point of the constraint fiber. Neither operation replaces the other.

PART VI

Final Recap and Conclusions

One differential, several geometric and numerical realizations.

What has the intrinsic viewpoint established?

The differential is the common first-order object; metrics, constraints, and coordinates provide complementary realizations.

The cost first supplies a differential

A smooth cost

$$J : \mathcal{V} \to \mathbb{R}$$

supplies, at every $v \in \mathcal{V}$, the intrinsic covector

$$\boxed{\mathrm{d}J_v \in \mathrm{T}^*_v\mathcal{V}.}$$

Before any metric is chosen, the differential determines

$$\boxed{\ker \mathrm{d}J_v, \qquad \mathcal{D}_v^-, \qquad \mathcal{D}_v^+.}$$

It also defines stationarity intrinsically:

$$\boxed{v \text{ stationary} \quad \Longleftrightarrow \quad \mathrm{d}J_v = 0.}$$

Takeaway

Level, ascent, descent, and stationary behavior belong to the differential itself. They do not require a gradient or a metric.

A metric converts the differential into a vector

A Riemannian metric supplies the sharp map

$$\sharp_{g,v} : \mathrm{T}_v^*\mathcal{V} \to \mathrm{T}_v\mathcal{V}$$

and defines

$$\boxed{\operatorname{grad}_g J(v) = \sharp_{g,v}(\mathrm{d}J_v).}$$

In coordinates,

$$\boxed{[\operatorname{grad}_g J]_v = G_v^{-1} d_v.}$$

Thus G_v^{-1} represents the sharp map that converts the covector components d_v into the tangent-vector components $[\operatorname{grad}_g J]_v$.

Preconditioning and Newton select particular metrics

Preconditioning selects an inverse metric:

$$\boxed{[\xi]_v = -P_v d_v, \qquad P_v = G_v^{-1}.}$$

When $H_v \succ 0$, Newton selects

$$P_v = H_v^{-1},$$

and hence

$$\boxed{[\xi_{\mathrm{N}}]_v = -H_v^{-1} d_v = -[\operatorname{grad}_{g^{\mathrm{N}}} J]_v.}$$

Takeaway

Gradient descent, preconditioning, and positive-definite Newton directions convert the same differential into tangent vectors using different choices of geometry.

A constraint restricts the admissible directions

For a smooth task map

$$f : \mathcal{V} \to \mathcal{W}$$

and a regular value $w \in \mathcal{W}$, the constraint set is the fiber

$$\boxed{\mathcal{F}_w = f^{-1}(w).}$$

Its feasible tangent space is

$$\boxed{\mathrm{T}_v\mathcal{F}_w = \ker \mathrm{d}f_v.}$$

Constrained stationarity is the intrinsic condition

$$\boxed{\mathrm{d}J_v\big|_{\mathrm{T}_v\mathcal{F}_w} = 0.}$$

Thus the cost differential must vanish on every feasible infinitesimal motion.

Constrained stationarity produces a multiplier covector

The intrinsic stationarity condition is equivalent to conormality:

$$\boxed{\mathrm{d}J_v \in (\mathrm{T}_v\mathcal{F}_w)^\circ.}$$

Using

$$(\mathrm{T}_v\mathcal{F}_w)^\circ = \mathrm{Im}(\mathrm{d}f_v)^*,$$

we obtain

$$\boxed{\mathrm{d}J_v = (\mathrm{d}f_v)^*\mu, \qquad \mu \in \mathrm{T}_w^*\mathcal{W}.}$$

Takeaway

The multiplier is a task-space covector whose pullback equals the cost differential at a constrained stationary point.

The multiplier also measures optimal-value sensitivity

Let

$$\widehat{v} : U \subset \mathcal{W} \to \mathcal{V}$$

be a differentiable selection of constrained minimizers, and define

$$\mathcal{J}^{\text{opt}}(w') = J\big(\widehat{v}(w')\big).$$

Differentiated feasibility gives

$$\mathrm{d}f_{\widehat{v}(w)} \circ \mathrm{d}\widehat{v}_w = \mathrm{Id}_{\mathrm{T}_w \mathcal{W}}\,.$$

Combining this identity with

$$\mathrm{d}J_{\widehat{v}(w)} = \big(\mathrm{d}f_{\widehat{v}(w)}\big)^* \mu$$

yields

$$\boxed{\mathrm{d}\mathcal{J}^{\text{opt}}_w = \mu = -\lambda.}$$

The multiplier evaluates task-value variations

For every

$$\delta w \in \mathrm{T}_w\mathcal{W},$$

the optimal-value differential satisfies

$$\boxed{\mathrm{d}\mathcal{J}^{\mathrm{opt}}_w(\delta w) = \mu(\delta w).}$$

Geometric interpretation

For every $\delta w \in \mathrm{T}_w\mathcal{W}$,

$$\mu(\delta w)$$

is the first-order variation of the optimal cost caused by the prescribed task variation δw.

Takeaway

The geometric multiplier is the sensitivity covector

$$\mathrm{d}\mathcal{J}^{\mathrm{opt}}_w \in \mathrm{T}^*_w\mathcal{W}.$$

How does the intrinsic geometry become a numerical algorithm?

A metric selects a tangent direction, projection enforces first-order feasibility, and retraction creates a new point.

A retraction realizes a finite manifold displacement

A tangent descent direction satisfies

$$-\operatorname{grad}_g J(v) \in \mathrm{T}_v\mathcal{V},$$

but it is not itself a point of $\mathcal{V}$.
A retraction is a smooth map

$$R : \mathcal{U} \subset \mathrm{T}\mathcal{V} \to \mathcal{V}$$

whose restriction at v satisfies

$$\boxed{R_v(0_v) = v, \qquad \mathrm{d}(R_v)_{0_v} = \mathrm{Id}_{\mathrm{T}_v\mathcal{V}}\,.}$$

The unconstrained metric-gradient iteration is

$$\boxed{v_{k+1} = R_{v_k}\Big(-\alpha_k \operatorname{grad}_g J(v_k)\Big)\,.}$$

Takeaway

The gradient supplies an infinitesimal direction. The retraction realizes the corresponding finite displacement as a manifold point.

Projection produces a feasible tangent direction

Under the induced metric,

$$\mathrm{T}_v\mathcal{V} = \mathrm{T}_v\mathcal{F}_w \oplus (\mathrm{T}_v\mathcal{F}_w)^{\perp_g}.$$

The metric-orthogonal projection is

$$\boxed{\mathrm{Proj}^g_{\mathrm{T}_v\mathcal{F}_w} : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_v\mathcal{F}_w.}$$

The gradient of the restricted cost is

$$\boxed{\mathrm{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) = \mathrm{Proj}^g_{\mathrm{T}_v\mathcal{F}_w}\big(\mathrm{grad}_g\, J(v)\big).}$$

If it is nonzero, then

$$-\,\mathrm{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) \in \mathrm{T}_v\mathcal{F}_w$$

is a strict first-order descent direction for $J|_{\mathcal{F}_w}$.

Takeaway

Projection extracts the feasible tangent component of the ambient gradient. It does not yet create a new feasible point.

Constrained stationarity has equivalent realizations

For $v \in \mathcal{F}_w$,

$$\boxed{\begin{aligned} \operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) = 0 &\iff \mathrm{d}J_v|_{\mathrm{T}_v\mathcal{F}_w} = 0 \\ &\iff \mathrm{d}J_v \in (\mathrm{T}_v\mathcal{F}_w)^\circ \\ &\iff \mathrm{d}J_v = (\mathrm{d}f_v)^*\mu. \end{aligned}}$$

The same condition can therefore be read as:

Algorithmic

The feasible gradient vanishes.

Geometric

The cost differential is conormal.

Dual

The differential is a pulled-back multiplier.

Projection and retraction complete the feasible update

A retraction on the fiber is a smooth map

$$R^{\mathcal{F}_w} : \mathcal{U}^{\mathcal{F}_w} \subset \mathrm{T}\mathcal{F}_w \to \mathcal{F}_w.$$

The constrained metric-gradient iteration is

$$\boxed{v_{k+1} = R^{\mathcal{F}_w}_{v_k}\left(-\alpha_k \operatorname{Proj}^g_{\mathrm{T}_{v_k}\mathcal{F}_w}\left(\operatorname{grad}_g J(v_k)\right)\right).}$$

The complete typed chain is

$$\boxed{\mathrm{T}_v\mathcal{V} \xrightarrow{\operatorname{Proj}^g_{\mathrm{T}_v\mathcal{F}_w}} \mathrm{T}_v\mathcal{F}_w \xrightarrow{-\alpha} \mathrm{T}_v\mathcal{F}_w \xrightarrow{R^{\mathcal{F}_w}_v} \mathcal{F}_w.}$$

Projection

Acts between tangent spaces and enforces first-order feasibility.

Retraction

Maps a tangent displacement to an actual feasible point.

What should be retained from the complete construction?

Every familiar formula represents an intrinsic object, an added geometric structure, or a type-correct map between them.

Summary I: differential and metric gradient

Object or operation	Intrinsic statement	Coordinate realization
Cost differential	$\mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}$	$d_v = \big(\partial J/\partial x^i(v)\big)_i$
Differential evaluation	$\mathrm{d}J_v(\xi)$	$d_v^\top [\xi]_v$
Level directions	$\ker \mathrm{d}J_v$	$d_v^\top [\xi]_v = 0$
Metric gradient	$\operatorname{grad}_g J(v) = \sharp_{g,v}(\mathrm{d}J_v)$	$[\operatorname{grad}_g J]_v = G_v^{-1} d_v$
Steepest descent	$-\dfrac{\operatorname{grad}_g J(v)}{\left\lVert \operatorname{grad}_g J(v) \right\rVert_g}$	$\arg \min_{\xi^\top G_v \xi = 1} d_v^\top \xi$

Summary I: metric selection and covariance

Object or operation	Intrinsic statement	Coordinate realization
Preconditioned direction	negative sharp image under a selected metric	$-P_v d_v, \quad P_v = G_v^{-1}$
Newton direction	negative gradient under the Hessian metric	$-H_v^{-1} d_v, \quad H_v \succ 0$
Coordinate change	intrinsic pairings are invariant	$[\xi]_y = K[\xi]_x,\ [\mathrm{d}J]_y = K^{-\top}[\mathrm{d}J]_x$
Stationarity	$\mathrm{d}J_v = 0$	$d_v = 0 \iff G_v^{-1} d_v = 0$

Summary II: constraint geometry

Object or operation	Intrinsic statement	Coordinate realization
Constraint fiber	$\mathcal{F}_w = f^{-1}(w)$	$f(v) = w$
Feasible tangent space	$\mathrm{T}_v\mathcal{F}_w = \ker \mathrm{d} f_v$	$F_v[\xi]_v = 0$
Restricted differential	$\mathrm{d}(J\vert_{\mathcal{F}_w})_v = (\mathrm{d}\iota_{w,v})^*(\mathrm{d}J_v)$	evaluate $d_v^\top[\xi]_v$ on $\ker F_v$
Conormal space	$(\mathrm{T}_v\mathcal{F}_w)^\circ$	$\mathrm{Im}\, F_v^\top$

Summary II: multipliers and sensitivity

Object or operation	Intrinsic statement	Coordinate realization
Geometric multiplier	$\mathrm{d}J_v = (\mathrm{d}f_v)^* \mu$	$d_v = F_v^\top [\mu]_w$
Plus-sign multiplier	$\mathrm{d}J_v + (\mathrm{d}f_v)^* \lambda = 0$	$d_v + F_v^\top [\lambda]_w = 0$
Sign relation	$\mu = -\lambda$	$[\mu]_w = -[\lambda]_w$
Optimal-value sensitivity	$\mathrm{d}\mathcal{J}_w^{\mathrm{opt}} = \mu$	$\delta \mathcal{J}^{\mathrm{opt}} \approx [\mu]_w^\top [\delta w]_w$

Summary III: feasible gradients

Object or operation	Intrinsic statement	Role
Tangent projection	$\mathrm{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w} : \mathrm{T}_v\mathcal{V} \to \mathrm{T}_v\mathcal{F}_w$	extracts the feasible tangent component
Restricted gradient	$\mathrm{grad}_{\mathcal{F}_w}(J\|_{\mathcal{F}_w})(v) = \mathrm{Proj}^{g}_{\mathrm{T}_v\mathcal{F}_w}(\mathrm{grad}_g\, J(v))$	provides a feasible first-order search direction
Restricted stationarity	$\mathrm{grad}_{\mathcal{F}_w}(J\|_{\mathcal{F}_w})(v) = 0$	equivalent to the multiplier equation

Summary III: retractions and updates

Object or operation	Intrinsic statement	Role
Ambient retraction	$R_v : \mathcal{U}_v \subset \mathrm{T}_v\mathcal{V} \to \mathcal{V}$	maps a tangent displacement to a manifold point
Fiber retraction	$R_v^{\mathcal{F}_w} : \mathcal{U}_v^{\mathcal{F}_w} \subset \mathrm{T}_v\mathcal{F}_w \to \mathcal{F}_w$	maps a feasible tangent displacement to a feasible point
Unconstrained update	$v_{k+1} = R_{v_k}(-\alpha_k \operatorname{grad}_g J(v_k))$	realizes an ambient metric-gradient step
Constrained update	$v_{k+1} = R_{v_k}^{\mathcal{F}_w}(-\alpha_k \operatorname{grad}_{\mathcal{F}_w}(J\vert_{\mathcal{F}_w})(v_k))$	realizes a finite feasible step

One intrinsic structure, several realizations

Takeaway

The cost supplies the intrinsic first-order covector

$$\boxed{J \longrightarrow \mathrm{d}J_v \in \mathrm{T}_v^*\mathcal{V}.}$$

A metric turns that covector into a tangent vector:

$$\boxed{\mathrm{d}J_v \xrightarrow{\sharp_{g,v}} \operatorname{grad}_g J(v).}$$

A regular constraint restricts the admissible directions:

$$\boxed{\mathrm{d}J_v|_{\mathrm{T}_v\mathcal{F}_w} = 0 \iff \mathrm{d}J_v = (\mathrm{d}f_v)^*\mu.}$$

Projection and retraction turn the same geometry into an algorithm:

$$\boxed{\operatorname{grad}_g J(v) \xrightarrow{\mathrm{Proj}^g_{\mathrm{T}_v\mathcal{F}_w}} \operatorname{grad}_{\mathcal{F}_w}(J|_{\mathcal{F}_w})(v) \xrightarrow{R_v^{\mathcal{F}_w}} v_+ \in \mathcal{F}_w.}$$

One intrinsic first-order object, several geometric realizations.

Contact

Antonio Franchi

University of Twente

Sapienza University of Rome

`https://r-franchi.eu/`

Corrections, suggestions, and discussion

`schol@r-franchi.eu`