

LINMA2474 - High-Dimensional Data Analysis and Optimization

ISSAMBRE L'HERMITE DUMONT
SIMON DESMIDT

This summary may not be up-to-date, the newer version is available at this address:
<https://github.com/SimonDesmidt/Syntheses>

Academic year 2025-2026 - Q2



Contents

1 Optimization on manifolds	2
1.1 Introduction	2
1.2 Definitions	2

Optimization on manifolds

1.1 Introduction

Classical optimization methods like the gradient descent solve problems of the form

$$\min_{x \in \mathcal{M}} f(x) \quad (1.1)$$

for a set M . The methods rely on two key properties:

- Linearity: x_k and $\nabla f(x_k)$ belong to some vector space, in which they can be combined with linear operations;
- Inner product: $\nabla f(x_k)$ is the unique element of \mathbb{R}^D such that

$$\forall v \in \mathbb{R}^D, Df(x)[v] = \langle v, \nabla f(x) \rangle \quad (1.2)$$

where $Df(x)[v] = \lim_{t \rightarrow 0} \frac{f(x+tv)-f(x)}{t}$ is the directional derivative of f at x in the direction v .

There are two ways to see the problem 1.1: as a constrained optimization problem, or as an unconstrained optimization problem assuming that nothing else exists outside the set M . Optimization on manifolds extends the classical unconstrained optimization algorithms to problems whose search space is a manifold (will be defined later).

1.2 Definitions

1.2.1 Definition and properties of a manifold

Definition 1.1 (Optimisation on manifolds). To minimize a function f on a manifold M , we need several objects:

- A **local defining function** $h : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $h^{-1}(0) = M$;
- The **tangent space of M** at some point x is the local linearization of M at $x \in M$: $u \in$ tangent space of M at x iff $u \in \text{Ker}(Dh(x))$;
- An inner product on the tangent spaces to define a new notion of gradient:

$$Df(x)[v] = \langle \nabla f(x), v \rangle$$

- A **retraction function**, i.e. a tool that allows to make a step on a manifold in a given tangent direction.

Definition 1.2 (Smoothness). $F : \mathcal{U} \subseteq \mathcal{E} \rightarrow \mathcal{V} \subseteq \mathcal{E}'$, with \mathcal{U}, \mathcal{V} open, is said to be smooth if it is \mathcal{C}^∞ on its domain.

Let us explain the concepts needed for our optimization:

Definition 1.3 (Embedded submanifold and local defining function). Let \mathcal{E} be a linear space of dimension d . A non-empty subset \mathcal{M} of \mathcal{E} is a smooth embedded submanifold of \mathcal{E} of dimension n if either:

- $n = d$ and \mathcal{M} is open in \mathcal{E} (open submanifold);
- $n = d - k$ for some $k \geq 1$ and, for each $x \in \mathcal{M}$, there exists a neighbourhood \mathcal{U} of x in \mathcal{E} and a smooth function $h : \mathcal{U} \rightarrow \mathbb{R}^k$. In that case,
 - $\mathcal{M} \cap \mathcal{U} = h^{-1}(0) = \{y \in \mathcal{U} : h(y) = 0\}$ and
 - $\text{rank}(Dh(x)) = k$.

Such a function h is called a local defining function for \mathcal{M} at x .

Definition 1.4 (Tangent space). Let \mathcal{M} be a subset of \mathcal{E} . For all $x \in \mathcal{M}$, define

$$\mathcal{T}_x \mathcal{M} = \{c'(0) \mid c : \mathcal{I} \rightarrow \mathcal{M} \text{ is smooth and } c(0) = x\} \quad (1.3)$$

where \mathcal{I} is any open interval containing $t = 0$. That means that v is in $\mathcal{T}_x \mathcal{M}$ iff there exists a smooth curve on \mathcal{M} passing through x with velocity v .

Consider \mathcal{M} an embedded submanifold of \mathcal{E} , $x \in \mathcal{M}$ and the set $\mathcal{T}_x \mathcal{M}$.

- If \mathcal{M} is an open submanifold of \mathcal{E} , then $\mathcal{T}_x \mathcal{M} = \mathcal{E}$;
- Otherwise, $\mathcal{T}_x \mathcal{M} = \text{Ker}(Dh(x))$ with h any local defining function at x .

Definition 1.5 (Tangent bundle). The tangent bundle is the set of all tangent spaces: $\mathcal{T} \mathcal{M} = \{(x, v) : v \in \mathcal{T}_x \mathcal{M}\}$.

Definition 1.6 (Map between manifolds). Let \mathcal{M} and \mathcal{M}' be embedded submanifolds of \mathcal{E} and \mathcal{E}' . A map $F : \mathcal{M} \rightarrow \mathcal{M}'$ is smooth iff $F = \bar{F}|_{\mathcal{M}}$ where \bar{F} is some smooth map from a neighbourhood of \mathcal{M} in \mathcal{E} to \mathcal{E}' .

Definition 1.7 (Differential of a map between manifolds). The differential of $F : \mathcal{M} \rightarrow \mathcal{M}'$ at the point $x \in \mathcal{M}$ is the linear map $DF(x) : \mathcal{T}_x \mathcal{M} \rightarrow \mathcal{T}_{F(x)} \mathcal{M}'$ defined by

$$DF(x)[v] = \frac{d}{dt} F(c(t))|_{t=0} = (F \circ c)'(0) \quad (1.4)$$

where c is some smooth curve on \mathcal{M} passing through x at $t = 0$ with velocity $v \in \mathcal{T}_x \mathcal{M}$.

→ Note: the definition does not depend on the choice of the curve c : $DF(x) = D\bar{F}(x)|_{\mathcal{T}_x \mathcal{M}}$.

Definition 1.8 (Retraction function). A retraction on a manifold \mathcal{M} is a smooth map $R : \mathcal{T}\mathcal{M} \rightarrow \mathcal{M} : (x, v) \mapsto R_x(v)$ such that each curve $c(t) = R_x(tv)$ satisfies $c(0) = x$ and $c'(0) = v$.

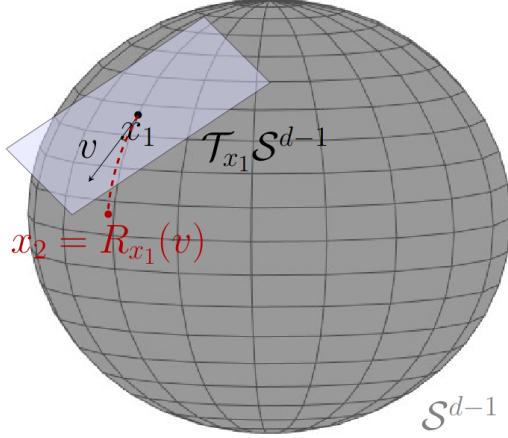


Figure 1.1: Retraction on a sphere.

1.2.2 Riemannian manifolds and metrics

Definition 1.9 (Inner product). As seen in previous courses, an inner product on $\mathcal{T}_x\mathcal{M}$ is a bilinear, symmetric, positive definite function $\langle \cdot, \cdot \rangle_x : \mathcal{T}_x\mathcal{M} \times \mathcal{T}_x\mathcal{M} \rightarrow \mathbb{R}$. Note that the inner product depends on the point of linearization. It induces some norm for tangent vectors: $\|u\|_x = \sqrt{\langle u, u \rangle_x}$. A metric on \mathcal{M} is a choice of inner product $\langle \cdot, \cdot \rangle_x$ for each \mathcal{M} .

Definition 1.10 (Metric). A metric $x \rightarrow \langle \cdot, \cdot \rangle_x$ on \mathcal{M} is a Riemannian metric if it varies smoothly with x , i.e. for all smooth vector fields V, W on \mathcal{M} , the function $x \rightarrow \langle V(x), W(x) \rangle_x$ is smooth from \mathcal{M} to \mathbb{R} .

Definition 1.11 (Riemannian manifold). A Riemannian manifold is a manifold with a Riemannian metric.

Definition 1.12 (Riemannian distance). Let \mathcal{M} be a Riemannian manifold. Given a smooth curve $c : [a, b] \rightarrow \mathcal{M}$, we define the length of c as

$$L(c) = \int_a^b \|c'(t)\|_{c(t)} dt \quad (1.5)$$

The Riemannian distance is then defined as $dist(x, y) = \inf_c L(c)$.

Definition 1.13 (Riemannian submanifolds). Let \mathcal{M} be an embedded submanifold of a Euclidean space \mathcal{E} . Equipped with the Riemannian metric obtained by restriction of the metric of \mathcal{E} , we call \mathcal{M} a Riemannian submanifold of \mathcal{E} .

1.2.3 Gradient on manifolds

Definition 1.14 (Riemannian gradient). Let $f : \mathcal{M} \rightarrow \mathbb{R}$ be smooth on a Riemannian manifold \mathcal{M} . The Riemannian gradient of f is the vector field $\text{grad}f$ on \mathcal{M} uniquely defined by the following identities:

$$\forall (x, v) \in \mathcal{T}\mathcal{M}, \quad Df(x)[v] = \langle v, \text{grad}f(x) \rangle_x \quad (1.6)$$

where $Df(x)$ is the differential of f and $\langle \cdot, \cdot \rangle_x$ is the Riemannian metric.

Theorem 1.15. Let \mathcal{M} be a Riemannian submanifold of \mathcal{E} endowed with the metric $\langle \cdot, \cdot \rangle$ and let $f : \mathcal{M} \rightarrow \mathbb{R}$ be a smooth function. The Riemannian gradient of f is given by

$$\text{grad}f(x) = \text{Proj}_x(\nabla \bar{f}(x)) \quad (1.7)$$

where \bar{f} is any smooth extension of f to a neighborhood of \mathcal{M} in \mathcal{E} , and $\nabla \bar{f}(x)$ is the Euclidean gradient of \bar{f} at x .

→ Note: for $\mathcal{E} = \mathbb{R}^d$ and using the usual metric $\langle u, v \rangle = u^T v$, the projection operator is

$$\text{Proj}_x(v) = v - (x^T v)x \quad (1.8)$$

Proposition 1.16. Let $f : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ and $G : \mathcal{M}_2 \rightarrow \mathcal{M}_3$ be smooth, where $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ are embedded submanifolds of $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ respectively. Then

$$G \circ F : \mathcal{M}_1 \rightarrow \mathcal{M}_3 : x \rightarrow G(F(x)) \quad (1.9)$$

is smooth and the chain rule applies:

$$D(G \circ F)(x)[v] = DG(F(x))[DF(x)[v]] \quad (1.10)$$

1.2.4 Taylor development of functions defined on manifolds

Let $f : \mathcal{M} \rightarrow \mathbb{R}$ be smooth and $c : \mathcal{I} \rightarrow \mathcal{M}$ be a smooth curve with $c(0) = x$ and $c'(0) = v$, with $\mathcal{I} \subseteq \mathbb{R}$ an open interval around $t = 0$ and $\|v\|_x = 1$. Let us define $g : \mathcal{I} \rightarrow \mathbb{R} : t \rightarrow g(t) = f(c(t))$. Since $g = f \circ c$ is smooth and maps real numbers to real numbers, it admits a Taylor expansion:

$$g(t) = g(0) + tg'(0) + \mathcal{O}(t^2) \quad (1.11)$$

By the chain rule,

$$g'(t) = Df(c(t))[c'(t)] = \langle \text{grad}f(c(t)), c'(t) \rangle_{c(t)} \quad (1.12)$$

and for $t = 0$,

$$g(0) = f(x) \quad g'(0) = \langle \text{grad}f(x), v \rangle_x \quad (1.13)$$

Therefore,

$$\begin{aligned} f(c(t)) &= f(x) + t\langle \text{grad}f(x), v \rangle_x + \mathcal{O}(t^2) \\ f(R_x(tv)) &= f(x) + \langle \text{grad}f(x), tv \rangle_x + \mathcal{O}(t^2) \end{aligned} \quad (1.14)$$

And defining $s := tv \in \mathcal{T}_x\mathcal{M}$,

$$f(R_x(s)) = f(x) + \langle \text{grad}f(x), s \rangle_x + \mathcal{O}(\|s\|_x^2) \quad (1.15)$$

This allows to define the Riemannian gradient descent:

Algorithm 1 Riemannian gradient descent

- 1: **Input:** $x_0 \in \mathcal{M}$;
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: Pick a step size $\alpha_k > 0$;
- 4:

$$x_{k+1} = R_{x_k}(-\alpha_k \text{grad}f(x_k)) \quad (1.16)$$

- 5: **end for**
-