THE LANDING ALGORITHM

FLORENTIN GOYENS, FLORIAN FEPPON

ABSTRACT. In recent years, an algorithm for optimization under constraints—sometimes called the landing algorithm—has received increased attention. In this work, we uncover several hidden connections between the landing and existing method. We also define the landing in its full generality, for any choice of metric in the ambient space. We provide the first adaptive step size selection procedure for the landing, with a line search on a merit function. The landing has mostly been considered for optimization under orthogonality constraints. We describe several choices of ambient metric for orthogonality constraints and relate it with existing version of the landing. We also numerically charactirize the relevance of the landing against feasible Riemannian optimization for orthogonality constraints.

keywords: landing algorihtm, constrained optimization, stiefel manifold, penalty method.

CONTENTS

1. INTRODUCTION

Under construction

Let $\mathcal{E}$ and $\mathcal{F}$ be finite-dimensional vector spaces of respective dimensions n and m with $m < n$, and let

$$f : \mathcal{E} \to \mathbb{R} \qquad \text{and} \qquad c : \mathcal{E} \to \mathcal{F}$$

Date: November 28, 2025.

be continuously differentiable mappings, possibly nonconvex. We consider the equality-constrained optimization problem

$$\text{(P)} \qquad \underset{x\in\mathcal{E}}{\text{minimize}}\ f(x) \qquad \text{subject to} \qquad c(x) = 0.$$

The Euclidean spaces $\mathcal{E}$ and $\mathcal{F}$ are equiped with inner products $\langle\cdot,\cdot\rangle_{\mathcal{E}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{F}}$, as well as their associated norms $\|\cdot\|_{\mathcal{E}}$ and $\|\cdot\|_{\mathcal{F}}$. The feasible set is denoted by

$$\mathcal{M} := \{x \in \mathcal{E} \colon c(x) = 0\}\,. \tag{1.1}$$

We also define the quadratic penalty function

$$\psi : \mathcal{E} \to \mathbb{R}, \qquad \psi(x) \ = \ \tfrac{1}{2}\,\|c(x)\|^2_{\mathcal{F}}. \tag{1.2}$$

In the applications of this paper, $\mathcal{E}$ and $\mathcal{F}$ are thought of as matrix subspaces, but at many places, we conveniently identify these spaces as $\mathcal{E} \simeq \mathbb{R}^n$ and $\mathcal{F} \simeq \mathbb{R}^m$ through a proper vectorial indexing.

Explain that landing-type methods are a special type of penalty methods that use the geometry of the feasible set explicitly. In this work, we propose a family of landing-like methods which subsumes several recent methods under a single framework. We show that existing algorithms correspond to various metric choices in the ambient space.

Algorithms for (P) fall into the two broad categories of feasible and infeasible methods. Infeasible methods—which are typically penalty methods—appeared in the 1980's. They constist in solving a succesion of unconstrained problems, whose solution eventually converges towards the solution of the constrained problem. Classical penalty methods include quadratic penalty methods and augmented Lagrangian methods. Nonsmooth penalty terms are also possible. Around the year 2000, Riemannian optimization methods appeared (Absil et al., 2008). This new framework introduced feasible algorithms for optimization problems on smooth Riemannian manifolds. They are a generalization of traditional unconstrained optimization methods, like gradient descent. More recently, various schemes introduced a bridge between the two approaches, combining the unconstrained approach with the geometric properties of Riemannian methods.

In this work, we show that several algorithms that have recently been proposed for (P) are all particular instances of a broader framework. The framework consists in a choice of metric in the ambient space.

Contributions and outline

In Section 2, we cover standard differential geometry concepts which connect with constrained optimization. In Section 3, we present a generalized framework for optimization with equality constraints, called the Riemannian Landing Method. In Section 3.3, we show that the Sequential Quadratic Programming method is a particular instance of the RLM, and we show that some augmented Lagrangian methods with a particular choice of mutlipliers are also part of the RLM framwork. In Section 4, we show that the RLM is globally convergent using adaptive step sizes based on a merit function. Finally, in Section **??**, we describe several choices of metric for optimization under orthogonality constraints and discuss their practicality. In the upcoming section, we review several algorithms recently proposed for (P), and show that many fall within the framework outline in this work, and correspond to particular choices of metric.

One thing that needs to be emphasized is where the novelties are

Comparison with the literature

FF: Add bibliography as exhaustive as possible

In this section, we examine various algorithms that have been proposed for (P), and show that they correspond to a specific choice of metric for the framework proposed in this paper.

Gao et al. (2022)

Cite (Mishra and Sepulchre, 2016), who uses the least-squares multipliers and makes the connection between SQP and Riemannian Newton method. He makes the connection on the feasible set (where the SQP iterate is a tangent step). But in fact we can characterize the normal part of the SQP update and make it fit into the landing framework.

Discuss the differences between penalty methods and Riemannian optimization. The latter's disadvantage is the cost of retractions, but it has the advantage of generating a sequence of feasible matrices. In order to ensure convergence to the feasible set, penalty methods may require to drive the penalty parameter towards infinity, which gives ill-conditioned subproblems that are difficult to solve.

With geometric infeasible method, we attempt to capture the best of each framework: no retractions and adding the geometry of the constraints to the subproblems and make them easier to solve.

Notations Let $\mathbb{R}_*^{n\times p}$ denote the matrices size $n \times p$ with full rank. We write $\mathrm{Sym}(p) = \{S \in \mathbb{R}^{p\times p} | S^\top = S\}$ and $\mathrm{Skew}(p) = \{\Omega \in \mathbb{R}^{p\times p} | \Omega^\top = -\Omega\}$, and $\mathrm{sym}(A) = (A+A^\top)/2$ and $\mathrm{skew}(A) = (A-A^\top)/2$.

2. Riemannian geometry and derivatives

2.1. Layered manifolds

Consider the following open set:

$$\mathcal{D} = \{x \in \mathcal{E} \colon \operatorname{rank}(\mathrm{D}c(x)) = m\}, \tag{2.1}$$

where we denote by $\mathrm{D}c$ the Fréchet derivative of c, namely $\mathrm{D}c(x)$ is the linear map such that

$$c(x+h) = c(x) + \mathrm{D}c(x)h + o(h) \text{ with } \frac{||o(h)||_{\mathcal{F}}}{||h||_{\mathcal{E}}} \to 0 \text{ as } ||h||_{\mathcal{E}} \to 0.$$

A central observation is that if the set

$$\mathcal{M}_x := \{y \in \mathcal{E} :\ c(y) = c(x)\}$$

is included in $\mathcal{D}$, then $\mathcal{M}_x$ is a smooth manifold. Every $\mathcal{M}_x$ is a level curve of the constraint function c, which we call a *layer manifold*. To every point $x \in \mathcal{D}$ is attached the tangent space

$$\mathrm{T}_x\mathcal{M}_x = \ker \mathrm{D}c(x). \tag{2.2}$$

2.2. Riemannian metric on $\mathcal{E}$

Let g be a Riemannian metric on $\mathcal{E}$, represented by a smoothly varying family of symmetric positive-definite operators $G(x) : \mathcal{E} \to \mathcal{E}$ such that

$$g(\xi, \zeta) = \langle G(x)\,\xi, \zeta\rangle_{\mathcal{E}}\,, \qquad \xi, \zeta \in \mathcal{E}. \tag{2.3}$$

The corresponding norm is $\|u\|_g = \sqrt{g(u,u)}$ for any $u \in \mathcal{E}$. When $G(x) = I_{\mathcal{E}}$ is the identity mapping, the metric coincides with the reference Euclidean inner product $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathcal{E}}$.

The metric g allows to define several important objects for optimization. First, the metric defines a *normal space* to the manifold $\mathcal{M}_x$ at $x \in \mathcal{D}$,

$$\mathrm{N}_x^g\mathcal{M}_x := \{v \in \mathcal{E} \colon g(v,\xi) = 0 \text{ for all } \xi \in \mathrm{T}_x\mathcal{M}_x\}\,. \tag{2.4}$$

Let $\mathrm{D}f(x) : \mathcal{E} \to \mathbb{R}$ denote the (Fréchet) differential of f. Through the celebrated Riesz identification theorem, there are several ways to identify the linear form $\mathrm{D}f(x)$ to a vector playing the role of a gradient. First, for any $x \in \mathcal{D}$, the *unconstrained Euclidean gradient* of f is the unique element $\nabla_{\mathcal{E}} f(x) \in \mathcal{E}$ such that

$$\langle \xi, \nabla_{\mathcal{E}} f(x)\rangle_{\mathcal{E}} = \mathrm{D}f(x)[\xi] \text{ for all } \xi \in \mathcal{E}.$$

Then, the unconstrained *Riemannian* gradient of f with respect to the metric g is written $\nabla_g f$ and is defined, for every $x \in \mathcal{D}$, as the unique element $\nabla_g f(x) \in \mathrm{T}_x\mathcal{D} \simeq \mathcal{E}$ that satisfies

$$g(\nabla_g f(x), \xi) = \mathrm{D}f(x)[\xi] \quad \text{ for all } \xi \in \mathrm{T}_x\mathcal{D} \simeq \mathcal{E}. \tag{2.5}$$

Finally, the Riemannian metric induces in turn a *constrained* Riemannian gradient on the manifold $\mathcal{M}_x$, denoted by $\mathrm{grad}^g_{\mathcal{M}_x} f$. Given $x \in \mathcal{D}$, it $\mathrm{grad}^g_{\mathcal{M}_x} f(x)$ is the unique vector in $\mathrm{T}_x\mathcal{M}_x$ satisfying

$$g(\mathrm{grad}^g_{\mathcal{M}_x} f(x), \xi) = \mathrm{D}f(x)[\xi], \quad \text{for all } \xi \in \mathrm{T}_x\mathcal{M}_x. \tag{2.6}$$

The unconstrained Euclidean and Riemannian gradients are related by the identity

$$\nabla_g f(x) = G(x)^{-1}\nabla_{\mathcal{E}} f(x). \tag{2.7}$$

Moreover, the constrained Riemannian gradient is the orthogonal projection of the unconstrained Riemannian gradient on the tangent space $T_x\mathcal{M}_x$:

$$\mathrm{grad}^g_{\mathcal{M}_x} f(x) = \mathrm{Proj}_{x,g}(\nabla_g f(x)), \tag{2.8}$$

where we denote by $\mathrm{Proj}_{x,g} \colon \mathcal{E} \to \mathrm{T}_x\mathcal{M}_x$ the orthogonal projection operator onto $\mathrm{T}_x\mathcal{M}_x \oplus N^g_x\mathcal{M}_x$. This operator is the unique linear operator satisfying

$$g(\xi, v - \mathrm{Proj}_{x,g}(v)) = 0 \quad \text{for all } \xi \in \mathrm{T}_x\mathcal{M}_x \text{ and } v \in \mathcal{E}, \tag{2.9}$$

see e.g., (Absil et al., 2008, (3.37)).

In the case where g is the Euclidean metric of $\mathcal{E}$, we denote by $\mathrm{grad}^{\mathcal{E}}_{\mathcal{M}_x} \equiv \mathrm{grad}^g_{\mathcal{M}_x}$ the associated constrained gradient which we call constrained Euclidean gradient.

2.3. Metric adjoints and projectors

Given a linear operator $A : \mathcal{E} \to \mathcal{F}$, the adjoint of A with respect to the inner products of $\mathcal{E}$ and $\mathcal{F}$ is the unique operator $A^{*,\mathcal{E}}$ defined by

$$\langle A\xi, y\rangle_{\mathcal{F}} \;=\; \left\langle \xi, A^{*,\mathcal{E}} y\right\rangle_{\mathcal{E}}, \qquad \text{for all } \xi \in \mathcal{E},\; y \in \mathcal{F}. \tag{2.10}$$

The adjoint with respect to the metric g is defined as the unique linear operator $A^{*,g} : \mathcal{F} \to \mathcal{E}$ satisfying

$$\langle A\xi, y\rangle_{\mathcal{F}} \;=\; g(\xi, A^{*,g} y), \qquad \text{for all } \xi \in \mathcal{E},\; y \in \mathcal{F}. \tag{2.11}$$

The adjoint operator $A^{*,g}$ can be obtained from $A^{*,\mathcal{E}}$ through the identity

$$A^{*,g} = G(x)^{-1}A^{*,\mathcal{E}}. \tag{2.12}$$

An important operator in optimization algorithms is the mapping

$$\mathrm{D}c(x)\,\mathrm{D}c(x)^{*,g} \colon \mathcal{F} \to \mathcal{F},$$

which is self-adjoint with respect to the inner product on $\mathcal{F}$ and positive definite for $x \in \mathcal{D}$. The right *g-pseudoinverse* of $\mathrm{D}c(x)$ is defined as the operator

$$\mathrm{D}c(x)^{\dagger,g} \;:=\; \mathrm{D}c(x)^{*,g}\,(\mathrm{D}c(x)\,\mathrm{D}c(x)^{*,g})^{-1} \;:\; \mathcal{F} \to \mathcal{E}. \tag{2.13}$$

The g-orthogonal projectors onto $\mathrm{T}_x\mathcal{M}_x$ and $\mathrm{N}_x\mathcal{M}_x$ read then respectively

$$\mathrm{Proj}_{x,g} \;=\; \mathrm{Id}_{\mathcal{E}} - (\mathrm{D}c(x))^{\dagger,g}\,\mathrm{D}c(x), \qquad \mathrm{Proj}^{\perp}_{x,g} \;=\; (\mathrm{D}c(x))^{\dagger,g}\,\mathrm{D}c(x). \tag{2.14}$$

Note that the g-normal space is $\mathrm{N}^g_x\mathcal{M}_x = \mathrm{Range}\,\mathrm{D}c(x)^{*,g}$. If $V \subset \mathcal{E}$ is a subspace of $\mathcal{E}$, we denote by $V^{\perp,\mathcal{E}} \subset \mathcal{E}$ its orthogonal subspace with respect to the Euclidean metric of $\mathcal{E}$.

In the particular case where $\mathcal{E} = \mathbb{R}^n$ and $\mathcal{F} = \mathbb{R}^m$, the differential $\mathrm{D}c(x)$ is represented by its Jacobian matrix

$$\mathrm{D}c(x) = \begin{pmatrix} \nabla c_1(x)^\top \\ \vdots \\ \nabla c_m(x)^\top \end{pmatrix} \in \mathbb{R}^{m\times n}, \tag{2.15}$$

and the Euclidean adjoint of any matrix $A \in \mathbb{R}^{m\times n}$ is given by the usual transpose $A^{*,\mathcal{E}} = A^\top$.

3. The Riemannian landing optimization algorithm

This section defines the landing algorithm as an iterative scheme involving two orthogonal tangent and normal steps, using the language of Riemannian geometry. We first define it in Section 3.1 by assuming that an ambient metric field g is given on $\mathcal{E}$. In Section 3.2, we provide some insights about the structure of the tangent and normal steps by expliciting their dependence to the orthogonal projectors on $\mathrm{T}_x\mathcal{M}_x$ and $\mathrm{N}_x\mathcal{M}_x$ and the restrictions of the metric to these spaces. Finally, we highlight some connections with the SQP and the Augmented Lagrangian methods in Section 3.3.

3.1. Definitions: tangent and normal steps by first specifying the metric

We now define the landing algorithm for solving (P), where $\mathcal{E}$ is viewed as an Euclidean space equiped with a Riemannian metric g. Given an arbitrary smooth symmetric positive-definite operator field

$$H(x)\colon \mathcal{F} \to \mathcal{F},$$

we consider the iterative scheme

$$x_{k+1} = x_k + \alpha_k(u(x_k) + v(x_k)), \tag{3.1}$$

where u and v are respectively the *tangent* and *normal* vector fields defined for $x \in \mathcal{D}$ by

$$u(x) := -\operatorname{grad}^g_{\mathcal{M}_x} f(x) \in \mathrm{T}_x\mathcal{M}_x, \tag{3.2}$$

$$v(x) := -\operatorname{D}c(x)^{\dagger,g}\, H(x)\, c(x) \in \mathrm{N}^g_x\mathcal{M}_x. \tag{3.3}$$

The tangent component $u(x)$ is the constrained Riemannian gradient of f on the manifold $\mathcal{M}_x$ with the metric g. Its role is to decrease the objective function f without increasing the violation of the constraints.

The normal component $v(x)$ is a vector orthogonal to $u(x)$, whose purpose is to 'land' the iterates x_k back towards feasibility by descending along the constraint violation $c(x)$. It is designed to be a descent direction for the infeasibility measure $\psi(x) = ||c(x)||^2_{\mathcal{F}}/2$ (eq. (1.2)). Indeed, if $c(x) \neq 0$, it holds

$$\begin{aligned} \operatorname{D}\psi(x)[v(x)] &= \langle \operatorname{D}c(x)[v(x)], c(x)\rangle_{\mathcal{F}} \\ &= -\left\langle \operatorname{D}c(x)\operatorname{D}c(x)^{*,g}\,(\operatorname{D}c(x)\operatorname{D}c(x)^{*,g})^{-1}\, H(x)\, c(x), c(x)\right\rangle_{\mathcal{F}} \\ &= -\langle H(x)c(x), c(x)\rangle_{\mathcal{F}} < 0. \end{aligned} \tag{3.4}$$

It is readily verified that the normal step $v(x)$ is the minimum norm solution of the undetermined system $\operatorname{D}c(x)[v] = -H(x)c(x)$, where the norm is taken with respect to g:

$$v(x) = \arg\min_{v\in\mathcal{E}} \frac{1}{2}\,\|v\|_g^2 \ \text{ such that } \operatorname{D}c(x)[v] = -H(x)c(x). \tag{3.5}$$

Hence the symmetric operator $H(x)$ gives an explicit control on the normal correction: on the one hand, it tunes the relative magnitude between the normal step $v(x_k)$ and the tangent step $u(x_k)$, on the other hand, its eigenvalues and eigenvectors determine the decay rate of the components of the constraint violation vector $c(x) \in \mathcal{F}$.

Two natural choices for $H(x)$ can be mentioned:

- $H(x) = \operatorname{D}c(x)\operatorname{D}c(x)^{*,g}$, whereby $v(x)$ becomes the negative unconstrained Riemannian gradient of the infeasibility measure $\psi(x) = ||c(x)||^2_{\mathcal{F}}/2$ relatively to the metric g:

$$v(x) = -\nabla_g\psi(x). \tag{3.6}$$

This equality can be inferred from identities (2.5) and (2.11):

$$g(\nabla_g\psi(x), \xi) = \operatorname{D}\psi(x)[\xi] = \langle \operatorname{D}c(x)\xi, c(x)\rangle_{\mathcal{F}} = g(\operatorname{D}c(x)^{*,g}c(x), \xi) \qquad \forall \xi \in \mathcal{E}.$$

- $H(x) = \mathrm{Id}_{\mathcal{F}}$, whereby $v(x)$ becomes the 'Newton-like' or 'pseudoinverse' step in the metric g:

$$v(x) = -\mathrm{D}c(x)^{\dagger,g} c(x) = -\mathrm{D}c(x)^{*,g} \left(\mathrm{D}c(x)\,\mathrm{D}c(x)^{*,g}\right)^{-1} c(x). \tag{3.7}$$

  This choice is natural because the iterative scheme $x_{k+1} = x_k - v(x_k)$ converges in that case quadratically to the feasible set $\mathcal{M}$ FF: REF? .

Formula (3.6) leads to the expression considered for the landing algorithm by Gao et al. (2022); Ablin and Peyré (2022) on the Stiefel manifold, while the 'Newton-like' formula (3.7) has been considered in the projected gradient flow of Yamashita (1980) and the null-space gradient flows of Feppon et al. (2020).

Although it is less obvious that $v(x)$ can be written as (3.6), we show in Proposition 3.2 below that it is always possible to redefine the metric g *without changing the values of* $u(x)$ *and* $v(x)$ so that $v(x)$ becomes the negative Riemannian unconstrained gradient of ψ in this redefined metric g, i.e.

$$v(x) = -\mathrm{D}c(x)^{\dagger,g} H(x) c(x) = -\nabla_g \psi(x).$$

In other words, defining the normal term as a negative unconstrained Riemannian gradient (3.6) or as the pseudo inverse step (3.3) leads to the same family of optimization algorithms.

The step length sequence $(\alpha_k)_{k\in\mathbb{N}}$ in (3.1) plays the role of an adaptive "time-step". If α_k is constant for any $k \in \mathbb{N}$, the iterative algorithm may be interpreted as the discretization of the ordinary differential equation

$$\begin{aligned} \dot{x} &= -u(x) - v(x) \\ &= (\mathrm{Id}_{\mathcal{E}} - \mathrm{D}c(x)^{*,g} \left(\mathrm{D}c(x)\,\mathrm{D}c(x)^{*,g}\right)^{-1} \mathrm{D}c(x))\nabla_g f(x) \\ &\quad - \mathrm{D}c(x)^{*,g} \left(\mathrm{D}c(x)\mathrm{D}c(x)^{*,g}\right)^{-1} H(x)\, c(x), \end{aligned} \tag{3.8}$$

which for $H(x) = \mathrm{Id}_{\mathcal{F}}$ and $g \equiv g^{\mathcal{E}}$ the Euclidean metric is exactly the projected gradient flow of Yamashita (1980) or null space gradient flow of Feppon et al. (2020).

FF: There is a natural condition that must be imposed on the time-step to ensure numerical stability: since

$$c(x_{k+1}) = (\mathrm{Id}_{\mathcal{F}} - \alpha_k H(x_k))c(x_k) + O(\alpha_k^2), \tag{3.9}$$

the condition

$$0 < \alpha_k < \frac{2}{\lambda_{\max}(H(x_k))} \text{ for all } k \in \mathbb{N},$$

must a priori be imposed on the step-length to ensure a decay at first order of the constraint violation vector. FF: In the case $H(x_k) = \mathrm{I}_{\mathcal{F}}$ corresponding to the 'Newton-like' pseudoinverse step, this condition reduces to $0 < \alpha_k < 2$, and the optimal decay predicted by (3.9) is $\alpha_k = 1$. We present a line search strategy for globablization in Section 4.

FF: I would mention the following results and comments somewhere around this remark, which will motivate further the globalization procedure:

- for a fixed step size $\alpha = \alpha_k$ and when removing the tangent space direction $u(x_k)$, the sequence $x_{k+1} = x_k - \alpha v(x_k)$ converges to the feasible domain in the sense at the geometric rate $O((1-\alpha\lambda)^k)$ with $\lambda \geq \sup_{x\in\mathcal{D}} \lambda_{\max}(H(x))$, for an initial point sufficiently close to the feasible domain.
- if $\alpha = 1$ and $H(x) = \mathrm{Id}_{\mathcal{F}}$, we have quadratic convergence close to the initial domain.
- if the tangent space term $u(x_k)$ is there, I think without further assumption we can only show

$$||c(x_k)|| \leq C \sup_k \alpha_k$$

  namely, the constraint is approached up to an error of the order of the step size.

The difficulty is thus to have global convergence both for the violation of the constraint and for the minimization. This may be solved by the analysis with the proper merit function ϕ after.

3.2. Definitions of the ambient metric g by first specifying the orthogonality

In the previous subsection, the metric g is defined *first*, which leads to the formulas (3.2) and (3.3) for the tangent and normal vector fields. Alternatively, the metric g can be constructed by first considering a smooth mapping of linear projectors

$$x \mapsto \mathrm{Proj}_x,$$

that is, for every $x \in \mathcal{D}$, Proj_x is a linear mapping on $\mathcal{E}$ satisfying $\mathrm{Proj}_x\mathrm{Proj}_x = \mathrm{Proj}_x$, $\mathrm{Range}(\mathrm{Proj}_x) = \mathrm{T}_x\mathcal{M}_x$ and $\mathrm{Proj}_x = \mathrm{Id}$ on $\mathrm{T}_x\mathcal{M}_x$. This viewpoint is equivalent to attaching to every $x \in \mathcal{M}_x$ a 'normal' space characterized by

$$\mathrm{N}_x\mathcal{M}_x := \ker(\mathrm{Proj}_x).$$

Then, we consider two symmetric operator fields

$$G_T(x) : \mathcal{E} \to \mathcal{E}, \qquad G_N(x) : \mathcal{E} \to \mathcal{E},$$

required to be positive definite on respectively $\mathrm{T}_x\mathcal{M}_x$ and $\mathrm{N}_x\mathcal{M}_x$: there exist $g_T, g_N > 0$ such that

$$\begin{aligned} \langle \xi, G_T(x)\xi\rangle_{\mathcal{E}} &\geq g_T\|\xi\|_{\mathcal{E}}^2, &\quad \forall \xi \in \mathrm{T}_x\mathcal{M}_x,\\ \langle \xi, G_N(x)\xi\rangle_{\mathcal{E}} &\geq g_N\|\xi\|_{\mathcal{E}}^2, &\quad \forall \xi \in \mathrm{N}_x\mathcal{M}_x. \end{aligned} \tag{3.10}$$

The ambient metric g can then be defined as (2.3) where $G(x) : \mathcal{E} \to \mathcal{E}$ is the block-diagonal operator

$$G(x) = \mathrm{Proj}_x^{*,\mathcal{E}} G_T(x)\mathrm{Proj}_x + (\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}} G_N(x)\mathrm{Proj}_x^{\perp}, \tag{3.11}$$

where we have denoted by $\mathrm{Proj}_x^{\perp} := \mathrm{Id}_{\mathcal{E}} - \mathrm{Proj}_x$ the associated projector on $\mathrm{N}_x\mathcal{M}_x$. Then, by the definition (3.11) of $G(x)$, it is clear that:

- G_T and G_N are the restrictions of the metric to the tangent space $\mathrm{T}_x\mathcal{M}_x$ and the normal space $\mathrm{N}_x\mathcal{M}_x$ respectively:

$$g(\xi, \zeta) = \langle \mathrm{Proj}_x\xi, G_T(x)\mathrm{Proj}_x\zeta\rangle_{\mathcal{E}} + \langle \mathrm{Proj}_x^{\perp}\xi, G_N(x)\mathrm{Proj}_x^{\perp}\zeta\rangle_{\mathcal{E}}. \tag{3.12}$$

- $\mathrm{T}_x\mathcal{M}_x$ and $\mathrm{N}_x\mathcal{M}_x$ are g-orthogonal subspaces, and $\mathrm{Proj}_x = \mathrm{Proj}_{x,g}$ and $\mathrm{Proj}_x^{\perp} = \mathrm{Proj}_{x,g}^{\perp}$ are the g-orthogonal projectors on the decomposition $\mathcal{E} = \mathrm{T}_x\mathcal{M}_x \oplus \mathrm{N}_x\mathcal{M}_x$.

We further note that $\mathrm{Proj}_x^{*,\mathcal{E}}$ and $(\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}}$ are the linear projectors associated to the decomposition $\mathcal{E} = \mathrm{N}_x\mathcal{M}_x^{\perp,\mathcal{E}} \oplus \mathrm{T}_x\mathcal{M}_x^{\perp,\mathcal{E}}$, and that the operators

$$\begin{aligned} \widetilde{G_T} &:= \mathrm{Proj}_x^{*,\mathcal{E}} G_T(x)\mathrm{Proj}_x &\quad &: \mathrm{T}_x\mathcal{M}_x \to (\mathrm{N}_x\mathcal{M}_x)^{\perp,\mathcal{E}},\\ \widetilde{G_N} &:= (\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}} G_N(x)\mathrm{Proj}_x^{\perp} &\quad &: \mathrm{N}_x\mathcal{M}_x \to (\mathrm{T}_x\mathcal{M}_x)^{\perp,\mathcal{E}}, \end{aligned} \tag{3.13}$$

are invertible due to (3.10). We denote by $\widetilde{G_T}(x)^{-1}$ and $\widetilde{G_N}(x)^{-1}$ their inverses, and we have then

$$G(x)^{-1} = \mathrm{Proj}_x\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}} + \mathrm{Proj}_x^{\perp}\widetilde{G_N}(x)^{-1}(\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}}. \tag{3.14}$$

Remark 3.1. *It is clear that only the restrictions of G_T and G_N to respectively the tangent space $\mathrm{T}_x\mathcal{M}_x$ and the normal space $\mathrm{N}_x\mathcal{M}_x$ matter in the definition of $G(x)$; the definitions of G_T and G_N as operators on the whole set $\mathcal{E}$ is motivated by the need to define a smooth metric field through (3.11).*

Remark 3.2. *This approach, whereby the differential structure of a manifold is specified by a differentiable mapping of projectors, has found other applications in Feppon and Lermusiaux (2019).*

The following proposition shows how the tangent and normal step of the landing algorithm depend on the family of projectors Proj_x and on the tangent metric $G_T(x)$. Remarkably, the normal step $v(x)$ depends only on the projector Proj_x, and thus on the choice of orthogonal normal space $\mathrm{N}_x\mathcal{M}_x$, but not on the particular metric G_N defined on $\mathrm{N}_x\mathcal{M}_x$.

Proposition 3.1. *The tangent and normal space steps of (3.2) and (3.3) can be rewritten in terms of the projectors and the tangent metric $G_T(x)$ as*

$$u(x) = -\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}\nabla_{\mathcal{E}} f(x), \tag{3.15}$$

$$v(x) = -\mathrm{Proj}_x^{\perp}\mathrm{D}c(x)^{\dagger,\mathcal{E}} H(x)c(x). \tag{3.16}$$

Proof. By (3.2), (2.7), (2.8) and (3.14),

$$\begin{aligned}
u(x) &= -\mathrm{grad}^g_{\mathcal{M}_x} f(x) \\
&= -\mathrm{Proj}_x(G(x)^{-1}\nabla_{\mathcal{E}} f(x)) \\
&= -\mathrm{Proj}_x(\mathrm{Proj}_x\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}} + \mathrm{Proj}_x^{\perp}\widetilde{G_N}(x)^{-1}(\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}})\nabla_{\mathcal{E}} f(x) \\
&= -\mathrm{Proj}_x\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}\nabla_{\mathcal{E}} f(x) = -\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}\nabla_{\mathcal{E}} f(x).
\end{aligned}$$

This proves (3.15). For (3.16), we start by recalling (eq. (2.14)) that

$$\mathrm{Proj}_x^{\perp} = \mathrm{D}c(x)^{\dagger,g}\mathrm{D}c(x).$$

Multiplying to the right by $\mathrm{D}c(x)^{\dagger,\mathcal{E}}$ and using $\mathrm{D}c(x)\mathrm{D}c(x)^{\dagger,\mathcal{E}} = \mathrm{Id}_{\mathcal{F}}$, we obtain the following fundamental identity relating the $g-$ and Euclidean right-pseudoinverses:

$$\mathrm{D}c(x)^{\dagger,g} = \mathrm{Proj}_x^{\perp}\mathrm{D}c(x)^{\dagger,\mathcal{E}}.$$

Formula (3.16) follows by substituting this formula into (3.3). □

Remark 3.3. *In particular, if the Riemannian Hessian* $\mathrm{Hess}_{\mathcal{E}} f(x)\colon \mathrm{T}_x\mathcal{M}_x \to \mathrm{T}_x\mathcal{M}_x$ *is positive definite for all* $x \in \mathcal{D}$, *then* $\widetilde{G_T}(x) = \mathrm{Hess}^g_{\mathcal{M}_x} f(x)$ *yields the Riemannian Newton step for the tangent component*

$$u(x) = -\mathrm{Hess}^g_{\mathcal{M}_x} f(x)^{-1}\mathrm{grad}^g_{\mathcal{M}_x} f(x). \tag{3.17}$$

This observation makes it possible to incorporate quasi-Newton updates such as BFGS updates for the tangent term.

Despite $v(x)$ does not depend on the metric G_N defined on the normal space $\mathrm{N}_x\mathcal{M}_x$, the following proposition shows that there is actually a natural choice of metric G_N for which (3.3) can be rewritten as a Riemannian gradient $v(x) = -\nabla_g\psi(x)$.

Proposition 3.2. *The pseudoinverse normal step (3.3) can be rewritten as*

$$v(x) = -\nabla_g\psi(x), \tag{3.18}$$

with $\psi(x) = ||c(x)||^2_{\mathcal{F}}/2$ *for the metric g defined through (2.3) and (3.11) by setting*

$$G_N(x) := \mathrm{D}c(x)^{*,\mathcal{E}} H(x)^{-1}\mathrm{D}c(x),$$

which defines a symmetric positive-definite operator on $\mathrm{N}_x\mathcal{M}_x$.

Proof. The operator $G_N(x)$ is clearly symmetric. To prove positive definiteness of $G_N(x)$ on $\mathrm{N}_x^{\mathcal{E}}\mathcal{M}_x$, consider any nonzero $\xi \in \mathrm{N}_x\mathcal{M}_x$. We find

$$\langle G_N(x)\xi,\, \xi\rangle_{\mathcal{E}} \;=\; \langle \mathrm{D}c(x)^{*,\mathcal{E}} H(x)^{-1}\mathrm{D}c(x)\xi,\, \xi\rangle_{\mathcal{E}} \;=\; \langle H(x)^{-1}\mathrm{D}c(x)\xi,\, \mathrm{D}c(x)\xi\rangle_{\mathcal{F}} > 0,$$

since $\xi \in \mathrm{N}_x^{\mathcal{E}}\mathcal{M}_x$, $\ker \mathrm{D}c(x) \cap \mathrm{N}_x = \{0\}$ and $H(x)^{-1}$ is symmetric definite positive. Thus, $G_N(x) \succ 0$ on $\mathrm{N}_x\mathcal{M}_x$ and (2.3) does define an metric g on $\mathcal{E}$.

Note now that since $\ker(\mathrm{D}c(x)) = \mathrm{T}_x\mathcal{M}_x$ and $\mathrm{Range}(\mathrm{D}c(x)^{*,\mathcal{E}}) = \mathrm{T}_x\mathcal{M}_x^{\perp,\mathcal{E}}$, we have the identities

$$\begin{aligned}
&\mathrm{D}c(x)\mathrm{Proj}_x^{\perp} = \mathrm{D}c(x), \qquad &&\mathrm{Proj}_x^{*,\mathcal{E}}\mathrm{D}c(x)^{*,\mathcal{E}} = 0 \\
&\mathrm{D}c(x)\mathrm{Proj}_x = 0, &&(\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}}\mathrm{D}c(x)^{*,\mathcal{E}} = \mathrm{D}c(x)^{*,\mathcal{E}}.
\end{aligned} \tag{3.19}$$

From (2.7), the unconstrained Riemannian gradient of $\psi(x) = ||c(x)||_{\mathcal{F}}^2/2$ is given by

$$
\begin{aligned}
\nabla_g \psi(x) &= G(x)^{-1} \nabla_{\mathcal{E}} \psi(x) = G(x)^{-1} \mathrm{D}c(x)^{*,\mathcal{E}} c(x) \\
&= \left(\mathrm{Proj}_x \widetilde{G_T}(x)^{-1} \mathrm{Proj}_x^{*,\mathcal{E}} + \mathrm{Proj}_x^{\perp} \widetilde{G_N}(x)^{-1} (\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}} \right) \mathrm{D}c(x)^{*,\mathcal{E}} c(x) \\
&= \widetilde{G_N}(x)^{-1} \mathrm{D}c(x)^{*,\mathcal{E}} c(x),
\end{aligned}
\tag{3.20}
$$

where we have used the second line of (3.19) in the last equality. It remains to express (3.20) in pseudoinverse form. We observe that

$$
\widetilde{G_N}(x) = (\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}} G_N(x) \mathrm{Proj}_x^{\perp} = (\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}} \mathrm{D}c(x)^{*,\mathcal{E}} H(x)^{-1} \mathrm{D}c(x) \mathrm{Proj}_x^{\perp} = \mathrm{D}c(x)^{*,\mathcal{E}} H(x)^{-1} \mathrm{D}c(x).
$$

Right multiplying by $\mathrm{Proj}_x^{\perp} \mathrm{D}c(x)^{\dagger,\mathcal{E}}$, this implies

$$
\begin{aligned}
\widetilde{G}_N(x) \mathrm{Proj}_x^{\perp} \mathrm{D}c(x)^{\dagger,\mathcal{E}} &= \mathrm{D}c(x)^{*,\mathcal{E}} H(x)^{-1} \mathrm{D}c(x) \mathrm{Proj}_x^{\perp} \mathrm{D}c(x)^{\dagger,\mathcal{E}} \\
&= \mathrm{D}c(x)^{*,\mathcal{E}} H(x)^{-1} \mathrm{D}c(x) \mathrm{D}c(x)^{\dagger,\mathcal{E}} = \mathrm{D}c(x)^{*,\mathcal{E}} H(x)^{-1},
\end{aligned}
$$

because of the first line of (3.19). Left multiplying by $\widetilde{G_N}(x)^{-1} : \mathrm{T}_x \mathcal{M}_x^{\perp,\mathcal{E}} \to \mathrm{N}_x \mathcal{M}_x$, which is allowed since $\mathrm{Range}(\mathrm{D}c(x)^{*,\mathcal{E}}) = \mathrm{T}_x \mathcal{M}_x^{\perp,\mathcal{E}}$, and right multiplying by $H(x)$, we obtain

$$
\widetilde{G_N}(x)^{-1} \, \mathrm{D}c(x)^{*,\mathcal{E}} = \mathrm{Proj}_x^{\perp} \mathrm{D}c(x)^{\dagger,\mathcal{E}} H(x). \tag{3.21}
$$

Combining (3.20) and (3.21) yields (3.18). □

3.3. Comparison with existing algorithms

3.3.1. *Link between the landing algorithm and Sequential Quadratic Programming*

A popular sequential quadratic programming (SQP) method considers the iterative sequence

$$
x_{k+1} = x_k + \alpha_k d_k, \tag{3.22}
$$

where d_k is obtained by solving at each iteration the quadratic program

$$
\begin{aligned}
d_k := \arg\min_{d \in \mathcal{E}} \quad & f(x_k) + \langle d, \nabla_{\mathcal{E}} f(x_k) \rangle_{\mathcal{E}} + \frac{1}{2} \langle d, B_k d \rangle_{\mathcal{E}} \\
\text{subject to} \quad & \mathrm{D}c(x_k) d + c(x_k) = 0,
\end{aligned}
\tag{3.23}
$$

for some symmetric positive definite operator $B_k : \mathcal{E} \to \mathcal{E}$ that approximates $\nabla_{\mathcal{E}}^2 \mathcal{L}(x_k, \lambda_k)$, where $\lambda_k \in \mathcal{F}$ are Lagrange multipliers and the Lagrangian is defined as

$$
\mathcal{L}(x, \lambda) = f(x) - \langle \lambda, c(x) \rangle_{\mathcal{F}}.
$$

This method is convergent under mild assumptions on the selection of symmetric positive-definite matrices B_k. In the particular case where $B_k = \nabla^2 \mathcal{L}(x, \lambda)$, it achieves quadratic convergence around a KKT point for a unit step size. Globalization procedures also exist, based on merit functions or filters (Nocedal and Wright, 2006, chap. 18).

The following proposition shows that the SQP iterations (3.22) form a particular instance of the Riemannian landing method (3.1). We note that this result can be found stated in different forms in the optimization literature. See for instance (Feppon, 2024, Proposition 1). The proof is provided for completeness.

Proposition 3.3. *Suppose that B_k is a symmetric positive definite operator on $\mathrm{T}_{x_k} \mathcal{M}_{x_k}$. Then, there exists a symmetric positive definite operator $G(x_k) : \mathcal{E} \to \mathcal{E}$ representing a metric inner produc g through (2.3), such that, if $x_k \in \mathcal{D}$, the SQP direction d_k of (3.23) is given by*

$$
d_k = -\mathrm{grad}_{\mathcal{M}_{x_k}}^g f(x_k) - \mathrm{D}c(x_k)^{\dagger,g} c(x_k). \tag{3.24}
$$

In other words, the SQP and the Riemannian landing steps (3.22) and (3.1) coincide at every step k for the metric induced by B_k on the tangent space $\mathrm{T}_{x_k} \mathcal{M}_{x_k}$, the choice $H(x_k) := \mathrm{Id}_{\mathcal{F}}$ for the normal component, and a unit step length $\alpha_k = 1$ for any $k \in \mathbb{N}$.

Proof. We construct the metric field g by following the construction of Section 3.2. Consider the space $\mathrm{N}_{x_k}\mathcal{M}_{x_k}$ defined as the $\mathcal{E}$-orthogonal space of $B_k\mathrm{T}_{x_k}\mathcal{M}_{x_k}$:

$$\mathrm{N}_{x_k}\mathcal{M}_{x_k} := (B_k\mathrm{T}_{x_k}\mathcal{M}_{x_k})^{\perp,\mathcal{E}}.$$

Because B_k is symmetric definite positive, we have the space decomposition $\mathcal{E} = \mathrm{T}_{x_k}\mathcal{M}_{x_k} \oplus \mathrm{N}_{x_k}\mathcal{M}_{x_k}$. Consider then Proj_{x_k} to be the projector on $\mathrm{T}_{x_k}\mathcal{M}_{x_k}$ whose kernel is $\mathrm{N}_{x_k}\mathcal{M}_{x_k}$. We consider g to be the metric satisfying $g(\xi,\zeta) := \langle G(x_k)\xi,\zeta\rangle_{\mathcal{E}}$, where $G(x_k)$ is given by (3.11) with $G_T(x_k) := B_k$ as the restricted metric on the tangent space $\mathrm{T}_{x_k}\mathcal{M}_{x_k}$, and any positive definite operator $G_N(x_k)$ on the normal space $\mathrm{N}_{x_k}\mathcal{M}_{x_k}$.

We observe that any d satisfying $\mathrm{D}c(x_k)d + c(x_k)$ can be written as

$$d = u + v(x_k),$$

with $v(x_k) = -\mathrm{D}c(x_k)^{\dagger,g}c(x_k)$ and $u \in \mathrm{T}_{x_k}\mathcal{M}_{x_k}$. The minimizer of (3.23) is $d_k = u_k + v(x_k)$ with u_k the solution to the quadratic unconstrained minimization problem

$$u_k = \arg\min_{u\in\mathrm{T}_{x_k}\mathcal{M}_{x_k}} f(x_k) + \langle u + v(x_k), \nabla_{\mathcal{E}} f(x_k)\rangle_{\mathcal{E}} + \frac{1}{2}\langle u + v(x_k), B_k(u + v(x_k))\rangle_{\mathcal{E}}. \tag{3.25}$$

We further note that for $u \in \mathrm{T}_{x_k}\mathcal{M}_{x_k}$,

$$\langle u, \nabla_{\mathcal{E}} f(x_k)\rangle_{\mathcal{E}} = \langle \mathrm{Proj}_{x_k} u, \nabla_{\mathcal{E}} f(x_k)\rangle_{\mathcal{E}} = \langle u, \mathrm{Proj}_{x_k}^{*,\mathcal{E}}\nabla_{\mathcal{E}} f(x_k)\rangle_{\mathcal{E}},$$

and since $v(x_k) \in \mathrm{N}_{x_k}\mathcal{M}_{x_k} = (G_T(x)\mathrm{T}_x\mathcal{M}_x)^{\perp,\mathcal{E}}$,

$$\begin{aligned}\langle u + v(x_k), B_k(u + v(x_k))\rangle_{\mathcal{E}} &= \langle u, B_k u\rangle_{\mathcal{E}} + 2\langle u, B_k v(x_k)\rangle_{\mathcal{E}} + \langle v(x_k), B_k v(x_k)\rangle_{\mathcal{E}}\\ &= \langle u, G_T(x_k)u\rangle_{\mathcal{E}} + 2\langle v(x_k), G_T(x_k)u(x_k)\rangle_{\mathcal{E}} + \langle v(x_k), B_k v(x_k)\rangle_{\mathcal{E}}\\ &= \langle u, G_T(x_k)u\rangle_{\mathcal{E}} + \langle v(x_k), B_k v(x_k)\rangle_{\mathcal{E}}.\end{aligned}$$

Consequently, by eliminating constant terms, the solution u_k to (3.25) is also the minimizer of the quadratic problem

$$u_k = \arg\min_{u\in\mathrm{T}_{x_k}\mathcal{M}_{x_k}} \langle u, \mathrm{Proj}_{x_k}^{*,\mathcal{E}}\nabla_{\mathcal{E}} f(x_k)\rangle_{\mathcal{E}} + \frac{1}{2}\langle u, \widetilde{G_T}(x_k)u\rangle_{\mathcal{E}}, \tag{3.26}$$

where $\widetilde{G_T}(x_k) : \mathrm{T}_{x_k}\mathcal{M}_{x_k} \to \mathrm{N}_{x_k}^{\perp,\mathcal{E}}$ is the operator defined in (3.13). The solution of this quadratic program is $u_k = -\widetilde{G_T}(x_k)^{-1}\mathrm{Proj}_{x_k}^{*,\mathcal{E}}\nabla_{\mathcal{E}} f(x_k)$, which is exactly $u(x_k)$ according to Proposition 3.1. □

3.3.2. *Interpretation of the landing algorithm as an augmented Lagrangian method*

We conclude this section by remarking that the landing algorithm (3.1) can also be interpreted as an augmented Lagrangian method with a particular choice of multipliers. Consider the iterative scheme

$$x_{k+1} = x_k - \alpha_k \nabla_g \mathcal{L}_\beta(x_k, \lambda_k), \tag{3.27}$$

with the augmented Lagrangian $\mathcal{L}_\beta$ given by

$$\mathcal{L}_\beta(x,\lambda) = f(x) + \langle \lambda, c(x)\rangle_{\mathcal{F}} + \frac{\beta}{2}\|c(x)\|_{\mathcal{E}}^2.$$

In (3.27), the unconstrained Riemannian gradient $\nabla_g\mathcal{L}_\beta$ is taken with respect to the variable x.

Proposition 3.4. *The Augmented Lagrangian iteration (3.27) coincides with the Riemannian landing iteration (3.1) by setting λ_k as the least-squares multiplier*

$$\lambda_k := \arg\min_{\lambda\in\mathcal{F}} \|\nabla_g f(x_k) - \mathrm{D}c(x_k)^{*,g}\lambda\|_{\mathcal{F}}^2 = (\mathrm{D}c(x_k)\mathrm{D}c(x_k)^{*,g})^{-1}\mathrm{D}c(x_k)\nabla_g f(x_k). \tag{3.28}$$

and $H(x) := \beta\mathrm{D}c(x)\mathrm{D}c(x)^{*,g}$.

Proof. The unconstrained Riemannian gradient of $\mathcal{L}_\beta$ with respect to x at $x = x_k$ reads

$$\begin{aligned}\nabla_g \mathcal{L}_\beta(x_k, \lambda_k) &= \nabla_g f(x_k) + \mathrm{D}c(x_k)^{*,g}\lambda_k + \beta \mathrm{D}c(x_k)^{*,g} c(x_k)\\ &= (\mathrm{Id}_{\mathcal{E}} - \mathrm{D}c(x_k)^{*,g}(\mathrm{D}c(x_k)\mathrm{D}c(x_k)^{*,g})^{-1}\mathrm{D}c(x_k))\nabla_g f(x_k) + \beta \mathrm{D}c(x_k)^{*,g} c(x_k).\end{aligned}$$

Using (2.8) and (3.3) with $H(x) = \beta \mathrm{D}c(x)\mathrm{D}c(x)^{*,g}$, we obtain $\nabla_g \mathcal{L}_\beta(x_k, \lambda_k) = \mathrm{grad}^g_{\mathcal{M}_x} f(x_k) + v(x_k)$. □

Thus, the landing algorithm with the Euclidean metric $g^{\mathcal{E}}$ and $H(x) = \beta \mathrm{D}c(x)\mathrm{D}c(x)^{*,g}$, as considered e.g. in Gao et al. (2022); Ablin and Peyré (2022), is an augmented Lagrangian method with the penalty parameter β is constant throughout iterations and the Lagrange multipliers λ_k being the least-squares multipliers, and for which the minimization of the augmented Lagrangian subproblem consists in a single gradient step.

4. Global convergence using adaptive step sizes

In this section, we design an inexact line search globalization procedure for the landing method. Namely, we propose an adaptive step size selection based on criteria similar to Armijo's rules, which guarantee global convergence on the scheme from an arbitrary initial point. Our approach builds upon the work of (Byrd et al., 2008) who designed a line search for the SQP method. This section exploits the connection between the landing algorithm and SQP highlighted in Section 3.3.1 to devise a globalization procedure for the landing algorithm (3.1).

In the previous proposition, the choice $\widetilde{G}_T(X)^{-1}$

4.1. Assumptions

We introduce a set of generic assumptions on the manifold and Lipschitz constants that will arise in our analysis. We first assume that there exists a constraint threshold $R > 0$ such that the following compactness assumption holds.

A1. *The set $\mathcal{R} = \{x \in \mathcal{E} : \|c(x)\|_{\mathcal{F}} \leq R\}$ is compact.*

We then require f and c to be Lipschitz continuous on $\mathcal{R}$.

A2. *f and c are Lipschitz continuous on $\mathcal{R}$. In particular, there exist constants $||\nabla_{\mathcal{E}} f||_\infty, ||\mathrm{D}c||_\infty > 0$ such that*

$$||\nabla_{\mathcal{E}} f(x)||_{\mathcal{E}} \leq ||\nabla_{\mathcal{E}} f||_\infty, \quad \sup_{v \in \mathcal{E}\setminus\{0\}} \frac{\|\mathrm{D}c(x)[v]\|_{\mathcal{F}}}{||v||_{\mathcal{E}}} \leq ||\mathrm{D}c||_\infty \qquad \forall x \in \mathcal{R}.$$

Third, we assume the classical linear independence constraint qualification (LICQ) in a neighborhood of the feasible set $\mathcal{M}$:

A3. *There exists a positive constants $\underline{\sigma}$ such that*

$$\inf_{x \in \mathcal{R}} \sigma_{\min}(\mathrm{D}c(x)) = \inf_{x \in \mathcal{R}} \sigma_m(\mathrm{D}c(x)) \geq \underline{\sigma} > 0,$$

where $\sigma_k(A)$ and $\sigma_{\min}(A)$ denote the kth and the smallest (Euclidean) singular value of a linear map A, respectively. Equivalently, the Euclidean pseudoinverse of $\mathrm{D}c(x)$ is bounded:

$$\sup_{w \in \mathcal{F}\setminus\{0\}} \frac{||\mathrm{D}c(x)^{\dagger,\mathcal{E}} w||_{\mathcal{E}}}{||w||_{\mathcal{F}}} \leq ||\mathrm{D}c^{\dagger,\mathcal{E}}||_\infty, \qquad \forall x \in \mathcal{R}.$$

These bounds will also imply the uniform boundedness of Riemannian gradients $\nabla_g f(x)$, adjoints $\mathrm{D}c^{*,g}(x)$ and pseudoinverses $\mathrm{D}c^{\dagger,g}$ by assuming uniform ellipticity of the metric.

A4. *There exist uniform constants $\underline{g} > 0$ and $\bar{g} > 0$ such that*

$$\forall x \in \mathcal{R}, \ \forall \xi \in \mathcal{E}, \ \underline{g}||\xi||^2_{\mathcal{E}} \leq g(\xi, \xi) = \langle G(x)\xi, \xi\rangle_{\mathcal{E}} \leq \bar{g}||\xi||^2_{\mathcal{E}}.$$

Finally, we also assume the uniform ellipticity of the operator field $H(x)$ involved in the normal space direction (3.3).

A5. *There exist uniform constants* $\lambda_{\max}(H) > 0$ *and* $\lambda_{\min}(H) > 0$ *such that*

$$\forall x \in \mathcal{R},\ \forall \zeta \in \mathcal{F},\ \lambda_{\min}(H)||\zeta||^2_{\mathcal{F}} \leq \langle \zeta, H(x)\zeta\rangle_{\mathcal{F}} \leq \lambda_{\max}(H)||\zeta||^2_{\mathcal{F}}.$$

Remark 4.1. *Assumption A5 holds in both cases* $H(x) = \mathrm{Id}_{\mathcal{F}}$ *and* $H(x) = \mathrm{D}c(x)\mathrm{D}c(x)^{*,g}$.

4.2. Line search procedure

The line search procedure is based on decreasing the ℓ_2-merit function

$$\phi_\mu(x) = f(x) + \mu \left\|c(x)\right\|_{\mathcal{F}} \tag{4.1}$$

for a large enough penalty parameter μ. We perform a backtracking line-search on the ℓ_2-merit $\phi_\mu(x)$ with Armijo parameters $\eta \in (0,1)$ and $\beta \in (0,1)$, and a nondecreasing update of μ that enforces a derivative decrease test.

Proposition 4.1. *If*

$$\mu \geq \frac{\mathrm{D}f(x)[v]}{\rho \left\|c(x)\right\|}, \tag{4.2}$$

with $\lambda_{\min}(H) \geq 2\rho$, *then*

$$\mathrm{D}\phi_\mu(x)[d] \leq -\|\mathrm{grad}^g_{\mathcal{M}_x} f(x)\|^2_g - \rho\mu\|c(x)\|_{\mathcal{F}}, \tag{4.3}$$

where $d = u + v = -\mathrm{grad}^g_{\mathcal{M}_x} f(x) - \mathrm{D}c(x)^{\dagger,g}H(x)c(x)$.

Proof. First note that for $\lambda_{\min}(H) \geq 2\rho$, we have

$$\mu \geq \frac{\mathrm{D}f(x)[v]}{\rho \left\|c(x)\right\|} \geq \frac{\mathrm{D}f(x)[v] \left\|c(x)\right\|}{(\lambda_{\min}(H(x)) - \rho) \left\|c(x)\right\|^2} \geq \frac{\mathrm{D}f(x)[v] \left\|c(x)\right\|}{\langle c(x), H(x)c(x)\rangle_{\mathcal{F}} - \rho \left\|c(x)\right\|^2}. \tag{4.4}$$

We have

$$\mathrm{D}\phi_\mu(x)[d] = \mathrm{D}f(x)[u] + \mathrm{D}f(x)[v] + \mu \left\langle \frac{c(x)}{\left\|c(x)\right\|}, \mathrm{D}c(x)[d] \right\rangle_{\mathcal{F}}. \tag{4.5}$$

Using

$$\mathrm{D}c(x)[d] = -H(x)c(x),$$

it gives

$$\begin{aligned}
\mathrm{D}\phi_\mu(x) &= g\left(\nabla_g f(x), -\mathrm{grad}^g_{\mathcal{M}_x} f(x)\right) + \mathrm{D}f(x)[v] - \mu\frac{\langle c(x), H(x)c(x)\rangle_{\mathcal{F}}}{\left\|c(x)\right\|} && (4.6)\\
&\leq - \left\|\mathrm{grad}^g f(x)\right\|^2_g + \frac{\mu}{\left\|c(x)\right\|}(c^\top Hc - \rho \left\|c(x)\right\|^2) - \mu\frac{\langle c(x), H(x)c(x)\rangle_{\mathcal{F}}}{\left\|c(x)\right\|} && (4.7)\\
&= - \left\|\mathrm{grad}^g f(x)\right\|^2_g - \mu\rho \left\|c(x)\right\|_{\mathcal{F}}. && (4.8)
\end{aligned}$$

□

Proposition 4.2.

$$\mu(x) = \frac{\mathrm{D}f(x)[v]}{\rho \left\|c(x)\right\|} \leq \bar{\mu} = \frac{\left\|\nabla f\right\|_\infty \left\|\mathrm{D}c^{\dagger,g}\right\| \lambda_{\max}(H)}{\rho} \tag{4.9}$$

Proof.

$$\begin{aligned}
\mathrm{D}f(x)[v] &= g\left(\nabla_g f(x), -\mathrm{D}c(x)^{\dagger,g}H(x)c(x)\right) && (4.10)\\
&\leq \left\|\nabla_g f\right\|_\infty \left\|\mathrm{D}c^{\dagger,g}\right\| \lambda_{\max}(H) \left\|c(x)\right\| && (4.11)
\end{aligned}$$

Florentin: Perhaps use $\bar{g}$ and change norm here? to check □

Proposition 4.3. *There exists* $\underline{\alpha} > 0$ *such that*

$$\alpha_k \geq \underline{\alpha} > 0 \qquad \forall k \in \mathbb{N},$$

and the Armijo linesearch satisfies, for some $\eta \in (0,1)$,

$$\phi_{\mu_k}(x_k) - \phi_{\mu_k}(x_{k+1}) \geq -\eta\alpha_k \mathrm{D}\phi_{\mu_k}(x_k)[d_k]. \tag{4.12}$$

Theorem 4.4. *Convergence*

Proof.

$$f(x_0) - f_{low} + \mu_0 \|c(x_0)\| \geq f(x_0) - f(x_{T+1}) + \mu_0 \|c(x_0)\| - \mu_T \|c(x_{T+1})\| + \sum \|c(x_k)\| (\mu_k - \mu_{k+1}) \tag{4.13}$$

$$= \sum_{k=0}^{T} [f(x_k) + \mu_k \|c(x_k)\| - f(x_{k+1}) - \mu_k \|c(x_{k+1})\|] \tag{4.14}$$

$$= \sum_{k=0}^{T} \phi_{\mu_k}(x_k) - \phi_{\mu_k}(x_{k+1}) \tag{4.15}$$

$$\geq \sum_{k=0}^{T} -\eta\alpha_k \mathrm{D}\phi_{\mu_k}(x_k)[d_k] \tag{4.16}$$

$$\geq \underline{\alpha}\|\mathrm{grad}^g_{\mathcal{M}_x} f(x)\|_g^2 + \underline{\alpha}\rho\mu\|c(x)\|_{\mathcal{F}} \tag{4.17}$$

□

Lemma 4.5. *Let $d(x) := u(x) + v(x) \in \mathcal{E}$ be the landing direction with $u(x) \in \mathrm{T}_x\mathcal{M}_x$ and normal component $v(x) = -\mathrm{D}c(x)^{\dagger,g} H(x)c(x) \in \mathrm{N}^g_x\mathcal{M}_x$. The merit function ϕ_μ is smooth on $\mathcal{R}\backslash\mathcal{M}$ and its derivative along the direction $d(x)$ is given by*

$$\mathrm{D}\phi_\mu(x)d(x) = d^\top \nabla f(x) - \mu\frac{\|\mathrm{D}c(x)^\top c(x)\|_2^2}{\|c(x)\|_2}, \qquad d \in \mathcal{E}. \tag{4.18}$$

Given $\rho \in (0, 1)$ such that

$$\|\mathrm{D}c(x)^\top c(x)\|_2^2 > 2\rho \|c(x)\|_2^2,$$

if

$$\mu \geq \frac{\left[d^\top \nabla f(x)\right]_+}{\rho \|c(x)\|_2}, \tag{4.19}$$

then we have the sufficient decrease for the merit function

$$\mathrm{D}\phi_\mu[d] \leq -\rho\mu \|c(x)\|_2 . \tag{4.20}$$

Proof. For $x \in \mathcal{R}$ and an arbitrary direction $d \in \mathcal{E}$, we have

$$\mathrm{D}\phi_\mu(x)d = \mathrm{D}f(x)d + \mu\langle \frac{c(x)}{||c(x)||_{\mathcal{F}}}, \mathrm{D}c(x)d\rangle_{\mathcal{F}}.$$

So for $d \equiv d(x)$, the identity $\mathrm{D}c(x)d(x) = -H(x)c(x)$ implies

$$\mathrm{D}\phi_\mu(x)d = \mathrm{D}f(x)d - \mu\langle \frac{c(x)}{||c(x)||_{\mathcal{F}}}, H(x)c(x)\rangle_{\mathcal{F}}.$$

$$\nabla_g\|c(x)\|_{\mathcal{E}} = \frac{\mathrm{D}c(x)^\top c(x)}{\|c(x)\|_2}.$$

Therefore, we have

$$\mathrm{D}\phi_\mu[d] = d^\top \nabla f(x) \ + \ \mu d^\top \nabla \|c(x)\|_2 \tag{4.21}$$

$$= d^\top \nabla f(x) \ + \ \mu v^\top \frac{\mathrm{D}c(x)^\top c(x)}{\|c(x)\|_2} \tag{4.22}$$

$$= d^\top \nabla f(x) - \mu\frac{\|\mathrm{D}c(x)^\top c(x)\|_2^2}{\|c(x)\|_2}. \tag{4.23}$$

Equation (4.20) is equivalent to

$$d^\top \nabla f(x) \le \mu \left(\frac{\|\mathrm{D}c(x)^\top c(x)\|_2^2}{\|c(x)\|_2} - \rho\, \|c(x)\|_2 \right), \tag{4.24}$$

which holds for any μ that satisfies (4.19) since

$$\mu \ge \frac{\left[d^\top \nabla f(x)\right]_+}{\rho\, \|c(x)\|_2} \ge \frac{\|c(x)\|_2 \left[d^\top \nabla f(x)\right]_+}{\|\mathrm{D}c(x)^\top c(x)\|_2^2 - \rho\, \|c(x)\|_2^2}. \tag{4.25}$$

□

Algorithm 1 Line-Search landing Algorithm (ℓ_2 merit-function)

Require: choose $\eta \in (0, \frac{1}{2})$, $\tau \in (0,1)$; initial pair $x_0 \in \mathbb{R}^n$
1: Evaluate f_0, ∇f_0, c_0, J_0
2: **repeat**
3: Compute d_k, the landing direction at x_k
4: Choose μ_k to satisfy (4.19) ▷ penalty for merit $\phi(\,\cdot\,; \mu_k)$
5: Set $\alpha_k \leftarrow 1$
6: **while** $\phi(x_k + \alpha_k d_k; \mu_k) > \phi(x_k; \mu_k) + \eta\, \alpha_k\, \nabla\phi(x_k; \mu_k)^\top d_k$ **do**
7: $\alpha_k \leftarrow \tau\, \alpha_k$ ▷ backtracking
8: **end while**
9: $x_{k+1} \leftarrow x_k + \alpha_k p_k$
10: Evaluate f_{k+1}, ∇f_{k+1}, c_{k+1}, J_{k+1}
11: **until** a convergence test is satisfied

A uniform choice of ρ. By Assumption 4(A2),

$$\|\mathrm{D}c(x)^\top c(x)\|_2 \;\ge\; \sigma_{\min}(\mathrm{D}c(x))\, \|c(x)\|_2 \;\ge\; \gamma\, \|c(x)\|_2,$$

hence $\|\mathrm{D}c(x)^\top c(x)\|_2^2 \ge \gamma^2 \|c(x)\|_2^2$. Therefore the premise $\|\mathrm{D}c(x)^\top c(x)\|_2^2 > 2\rho\, \|c(x)\|_2^2$ holds *uniformly* for any fixed

$$0 < \rho < \frac{\gamma^2}{2}. \tag{4.26}$$

4.3. Bounding the merit parameter

Theorem 4.6 (Uniform bound and eventual constancy of μ for the modified step)**.** *Let (A1), (A2), and (A3) hold, and choose any fixed $\rho \in (0, \gamma^2/2)$. Suppose μ is updated by a nondecreasing rule that enforces $\mu \ge \dfrac{[\nabla f(x)^\top d]_+}{\rho\, \|c(x)\|}$ whenever needed. Then*

$$\mu \le \bar{\mu} := \frac{C\, G}{\rho} \quad \textit{for all iterations in the neighborhood } \mathcal{N},$$

and therefore μ can increase only finitely many times and is eventually constant.

Proof. Note that

$$\nabla f(x)^\top u = \operatorname{grad} f(x)^\top u = -\operatorname{grad} f(x)^\top B_k \operatorname{grad} f(x) \;\le\; 0, \tag{4.27}$$

With $d = u + v$ and (4.27),

$$[\nabla f(x)^\top d]_+ = [\,\nabla f(x)^\top u + \nabla f(x)^\top v\,]_+ \le [\,\nabla f(x)^\top v\,]_+.$$

Since $v = -\mathrm{D}c(x)^\top c(x)$, we have

$$|\nabla f(x)^\top v| = \left|\, \nabla f(x)^\top \mathrm{D}c(x)^\top c(x) \,\right| = \left|\, c(x)^\top \mathrm{D}c(x)\, \nabla f(x) \,\right| \le \|c(x)\|\, \|\mathrm{D}c(x)\|\, \|\nabla f(x)\| \le C\, G\, \|c(x)\|,$$

using the bounds from (A1). Consequently,

$$\frac{[\nabla f(x)^\top d]_+}{\rho\,\|c(x)\|} \le \frac{CG\|c(x)\|}{\rho\|c(x)\|} \le \frac{C\,G}{\rho}. \tag{4.28}$$

Combine (4.28) with the update rule. Since the update is nondecreasing and there is a finite cap $\bar\mu$, only finitely many increases are possible. □

By Theorem 4.6, μ_k is uniformly bounded and eventually constant; denote this stabilized value by $\bar\mu$. We now analyze the backtracking Armijo line-search on $\phi_{\bar\mu}$ with parameters $\eta \in (0,1)$, $\beta \in (0,1)$.

4.4. Backtracking line search on $\phi_{\bar\mu}$

Lemma 4.7. *For $d_k = u(x_k)+v(x_k)$ with $u(x_k) = -\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)$ and $v(x_k) = -\mathrm{D}c(x_k)^{\dagger,g}H(x_k)c(x_k)$, and with*

$$\sigma_c := \bar\mu\,\lambda_{\min}(H) - ||\mathrm{D}f||_\infty||\mathrm{D}c^{\dagger,g}||_\infty\lambda_{\max}(H) > 0,$$

we have for any $k \in \mathbb{N}$ such that $c(x_k) \neq 0$:

$$\mathrm{D}\phi_{\bar\mu}(x_k)[d_k] \le -\underline{g}\|\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)\|^2_{\mathcal{E}} - \sigma_c\,\|c(x_k)\|_{\mathcal{F}}. \tag{4.29}$$

Proof. If $c(x_k) = 0$, then $v_k = 0$ and Lemma 4.5 gives

$$\mathrm{D}\phi_{\bar\mu}(x_k)[d_k] = \nabla f(x_k)^\top u_k \tag{4.30}$$
$$= -\,\mathrm{grad} f(x_k)^\top B_k\,\mathrm{grad} f(x_k) \tag{4.31}$$
$$\le -m\|\mathrm{grad} f(x_k)\|^2, \tag{4.32}$$

so (4.29) holds (the second term vanishes). FF: No, $\phi_{\bar\mu}$ is not differentiable if $c(x_k) = 0$ so the equation above does not make sense. Florentin: Don't we have a directional derivative?

By Lemma 4.5,

$$\mathrm{D}\phi_{\bar\mu}(x_k)[d_k] = \mathrm{D}f(x_k)[u(x_k)] + \mathrm{D}f(x_k)[v(x_k)] - \bar\mu\,\frac{\langle c(x_k), H(x_k)c(x_k)\rangle_{\mathcal{F}}}{\|c(x_k)\|_{\mathcal{F}}}.$$

For the tangential part, equation (2.6) and assumption A4 yields

$$\mathrm{D}f(x_k)[u(x_k)] = -g(\mathrm{grad}^g_{\mathcal{M}_x} f(x_k), \mathrm{grad}^g_{\mathcal{M}_x} f(x_k)) \le -\underline{g}||\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)||^2_{\mathcal{E}}.$$

For the normal part, we have, utilizing A5:

$$\mathrm{D}f(x_k)[v(x_k)] \;\le\; \left|\mathrm{D}f(x_k)[v(x_k)]\right| = \left|\mathrm{D}f(x_k)\mathrm{D}c(x_k)^{\dagger,g}H(x_k)c(x_k)\right| \le ||\mathrm{D}f||_\infty||\mathrm{D}c^{\dagger,g}||_\infty\lambda_{\max}(H)||c(x_k)||_{\mathcal{F}}.$$

Finally, assumption A5 again implies

$$\frac{\langle c(x_k), H(x_k)c(x_k)\rangle_{\mathcal{F}}}{||c(x_k)||_{\mathcal{F}}} \ge \lambda_{\min}(H)||c(x_k)||_{\mathcal{F}}.$$

Collecting the three bounds, we obtain

(4.33)
$$\begin{aligned}\mathrm{D}\phi_{\bar\mu}(x_k)[d_k] \;&\le\; -\,\underline{g}\,\|\mathrm{grad} f(x_k)\|^2 \;+\; ||\mathrm{D}f||_\infty||\mathrm{D}c^{\dagger,g}||_\infty\lambda_{\max}(H)\,\|c(x_k)\| \;-\; \bar\mu\,\lambda_{\min}(H)\,\|c(x_k)\| \\ &\le\; -\,\underline{g}\,\|\mathrm{grad} f(x_k)\|^2 \;-\; (\bar\mu\,\lambda_{\min}(H) - ||\mathrm{D}f||_\infty||\mathrm{D}c^{\dagger,g}||_\infty\lambda_{\max}(H))\,\|c(x_k)\|.\end{aligned}$$

FF: Check from here if consistent with before By construction of the merit-parameter update rule, we ensure $\bar\mu \ge C\,G/\rho$ for some $\rho \in (0, \gamma^2/2)$. Then

$$\bar\mu\,\gamma^2 - C\,G \;\ge\; C\,G\left(\frac{\gamma^2}{\rho} - 1\right) \;\ge\; C\,G \;>\; 0.$$

Hence, defining $\sigma_c := \bar\mu\,\lambda_{\min}(H) - ||\mathrm{D}f||_\infty||\mathrm{D}c^{\dagger,g}||_\infty\lambda_{\max}(H) > 0$ gives the result. □

Let L_ϕ denote a Lipschitz constant of $\nabla\phi_{\bar\mu}$ on the level set visited by the iterates with $c \neq 0$ (exists by (A1) and boundedness of the level set; see below). Standard backtracking then accepts some $\alpha_k = \beta^{j_k}$ satisfying the Armijo condition

$$\phi_{\bar\mu}(x_{k+1}) \;\le\; \phi_{\bar\mu}(x_k) + \eta\,\alpha_k\,\mathrm{D}\phi_{\bar\mu}(x_k)[d_k]. \tag{4.34}$$

Moreover, descent-direction smoothness yields the lower bound

$$\alpha_k \;\ge\; \min\Big\{1,\; \frac{2(1-\eta)\,|\mathrm{D}\phi_{\bar\mu}(x_k)[d_k]|}{L_\phi\,\|d_k\|^2}\Big\}. \tag{4.35}$$

Since $\|d_k\| \le \|u_k\| + \|v_k\| \le M\,\|\mathrm{grad} f(x_k)\| + C\,\|c(x_k)\|$ and both $\|\mathrm{grad} f(x_k)\|$ and $\|c(x_k)\|$ remain bounded on the level set, $\|d_k\|$ is uniformly bounded.

We also use that the iterates remain in a compact level set: by (4.34) and (4.29), $\{\phi_{\bar\mu}(x_k)\}$ is monotonically decreasing and bounded below, hence $\{x_k\}$ stays in the initial level set of $\phi_{\bar\mu}$, which is compact by (A1).

4.5. **Asymptotic convergence**

Lemma 4.8. *There exists* $\underline{\alpha} > 0$ *such that*

$$\alpha_k \ge \underline{\alpha} > 0 \qquad \forall k \in \mathbb{N}.$$

Proof. From (4.35), we see that

$$\alpha_k \ge \frac{2(1-\eta)}{L_\phi}\frac{||\mathrm{D}\phi_{\bar\mu}(x_k)[d_k]||_{\mathcal{F}}}{||d_k||^2_{\mathcal{E}}} \ge \frac{2(1-\eta)}{L_\phi}\frac{\underline{g}||\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)||^2_{\mathcal{E}} + \sigma_c||c(x_k)||_{\mathcal{F}}}{C(||\mathrm{grad} f(x_k)||^2 + ||c(x_k)||^2)} \ge \underline{\alpha}.$$

if $c(x_k)$ remains bounded (study the function $\beta(t) = \frac{a+t}{a+t^2}$). □

Lemma 4.9. *If f is bounded from below, then $c(x_k) \to 0$ and $||\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)|| \to 0$ as $k \to +\infty$.*

FF: Remark: if all iterates are in $\mathcal{R}$ then f is bounded from below...

Proof. From (4.34) and (4.29), we deduce that

$$\phi(x_{k+1}) \le \phi(x_k) - \eta\alpha_k(\sigma_c||c(x_k)||_{\mathcal{F}} + \underline{g}||\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)||^2_{\mathcal{E}}).$$

Moreover, from (4.35) and (4.29) we can find the existence of a constant $\underline{\alpha} > 0$ such that

$$\alpha_k \ge \underline{\alpha}$$

FF: Doubt about (4.35), to be derived more cleanly. This inequality implies that for any $k \in \mathbb{N}$,

$$\phi(x_k) \le \phi(x_0) - \eta\underline{\alpha}\left(\sigma_c\sum_{i=0}^{k-1}||c(x_k)||_{\mathcal{F}} + \sum_{i=0}^{k-1}\underline{g}||\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)||^2_{\mathcal{E}}\right).$$

Since f is bounded from below, ϕ is bounded from below and it follows from the above inequality that the series

$$\sum_{k=0}^{+\infty}||c(x_k)||_{\mathcal{F}}, \qquad \sum_{k=0}^{+\infty}||\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)||^2_{\mathcal{E}},$$

are convergent. The result follows. □

Corollary 4.10. *Let $x_\star$ be any accumulation point. Then* $\mathrm{grad} f(x_\star) = 0$ *and* $c(x_\star) = 0$.

FF: Due to compactness, there must exist an accumulation point, must there not? .

What happens at iterations before μ_k reaches $\bar\mu$?

5. Geometric design of tangent and normal terms for the landing algorithm with orthogonality constraints

A natural application for the landing algorithm is optimization with orthogonality constraints, such as

$$\underset{X\in\mathbb{R}^{n\times p}}{\text{minimize}}\ f(X) \text{ subject to } X^\top X = \mathrm{I_p}, \tag{5.1}$$

where $p \leq n$. The feasible set—called the Stiefel manifold—consists of rectangular matrices with orthonormal columns

$$\mathcal{M} \equiv \mathrm{St}(n,p) := \{X \in \mathbb{R}^{n\times p} \,|\, X^\top X = \mathrm{I_p}\}.$$

The function that defines the constraints is

$$c\colon \mathbb{R}^{n\times p} \to \mathrm{Sym}(p)\colon c(X) = \frac{1}{2}(X^\top X - \mathrm{I_p}). \tag{5.2}$$

Thus, we have an instance of (P) with $\mathcal{E} = \mathbb{R}^{n\times p}$ and $\mathcal{F} = \mathrm{Sym}(p)$, the set of symmetric matrices of size p. The spaces $\mathcal{E}$ and $\mathcal{F}$ are equiped with the standard Frobenius inner product $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathcal{E}}$ and $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathcal{F}}$. In all what follows, we drop the subscript notation $\cdot_{\mathcal{E}}$ and $\cdot_{\mathcal{F}}$ when writing matrix Frobenius inner products or norms, in particular, for two matrices of same size X and Y, $\langle X, Y\rangle = \mathrm{trace}(X^\top Y)$ and $||X||^2 = \langle X, X\rangle$. In this section the squared infeasibility reads

$$\psi(X) = \frac{1}{2}||c(X)||^2 = \frac{1}{4}||X^\top X - \mathrm{I}_p||^2,$$

where I_p is the identity matrix of $\mathbb{R}^{p\times p}$. The set $\mathcal{D}$ of (2.1) is the set of full rank matrices $X \in \mathbb{R}^{n\times p}_*$ and the layer manifold $\mathcal{M}_X$ is

$$\mathcal{M}_X \equiv \mathrm{St}_{X^\top X} := \left\{Y \in \mathbb{R}^{n\times p} : Y^\top Y = X^\top X\right\}.$$

In this section, we demonstrate how the choice of the family of normal spaces $X \to \mathrm{N}_X\mathrm{St}_{X^\top X}$ is crucial to lead to efficient computations of the tangent and normal terms of the landing algorithm (3.1). To illustrate this claim, we first consider the family of normal spaces orthogonal to the tangent spaces with respect to the Euclidean Frobenius inner product. Although this choice is natural, computing the corresponding orthogonal projections is expensive because it requires solving Sylvester equations. We then take the reverse approach of first introducing an alternative family of normal spaces with associated projection operators that can be computed explicitly, and second designing the tangent and the normal metric that lead to explicit and thus more exploitable formulas for the landing algorithm on the Stiefel manifold.

The starting point is to recall the characterization of the tangent space to $\mathrm{St}_{X^\top X}$ at $X \in \mathbb{R}^{n\times p}_*$. We emphasize that this set does not depend on the metric.

Proposition 5.1 (Gao et al. (2022); Goyens et al. (2025)). *The tangent space of* $\mathrm{St}_{X^\top X}$ *at* X *is the set*

$$\begin{aligned}
\mathrm{T}_X\mathrm{St}_{X^\top X} &= \{\xi \in \mathbb{R}^{n\times p} \,|\, \xi^\top X + X^\top \xi = 0\} && (5.3)\\
&= \{X(X^\top X)^{-1}\Omega + \Delta \,|\, \Omega \in \mathrm{Skew}(p),\ \Delta \in \mathbb{R}^{n\times p} \textit{ with } \Delta^\top X = 0\} && (5.4)\\
&= \{WX \,|\, W \in \mathrm{Skew}(n)\}, && (5.5)
\end{aligned}$$

with dimension $np - p(p+1)/2$.

5.1. Landing algorithm for the Euclidean metric

We first outline the derivation of the tangent and normal steps (3.2) and (3.3) when $g \equiv g^{\mathcal{E}}$ is the Euclidean or Frobenius inner product metric:

$$g^{\mathcal{E}}(\xi,\zeta) = \langle \xi, \zeta\rangle \qquad \forall \xi, \zeta \in \mathbb{R}^{n\times p}.$$

The tangent and normal vector fields (3.2) and (3.3), with respect to the Euclidean metric, are given by

$$u(X) = -\mathrm{Proj}_{X,\mathcal{E}}[\nabla_{\mathcal{E}} f(X)], \tag{5.6}$$

$$v(X) = -\mathrm{Dc}(X)^{\dagger,\mathcal{E}}[H(X)[c(X)]] \tag{5.7}$$

where brackets denote the action of linear operators on matrix spaces.

In what follows, we compute these two directions explicitly. Note that the orthogonal projector $\mathrm{Proj}_{X,\mathcal{E}}$ can be obtained through the characterization (2.9), as is commonly done in the literature on Riemannian optimization Absil et al. (2008). However, since the formula for $v(X)$ also requires the computation of the pseudoinverse $\mathrm{Dc}(X)^{\dagger,\mathcal{E}}$, we instead propose to compute it using the explicit relation $\mathrm{Proj}_{X,\mathcal{E}} = \mathrm{Id}_{\mathcal{E}} - \mathrm{Dc}(X)^{\dagger,\mathcal{E}}\mathrm{Dc}(X)$. As we shall see, the main difficulty in computing this projection lies in inverting the symmetric operator $(\mathrm{Dc}(X)\mathrm{Dc}(X)^{*,\mathcal{E}})^{-1}$ that appears in the pseudoinverse formula $\mathrm{Dc}(X)^{\dagger,\mathcal{E}} = \mathrm{Dc}(X)^{*,\mathcal{E}}(\mathrm{Dc}(X)\mathrm{Dc}(X)^{*,\mathcal{E}})^{-1}$. Let us start by evaluating $\mathrm{Dc}(X)$ and $\mathrm{Dc}(X)^{*,\mathcal{E}}$.

Proposition 5.2. *The operator $\mathrm{Dc}(X) : \mathbb{R}^{n\times p} \to \mathrm{Sym}(p)$ and its adjoint $\mathrm{Dc}(X)^{*,\mathcal{E}} : \mathrm{Sym}(p) \to \mathbb{R}^{n\times p}$ read:*

(i) $\mathrm{Dc}(X)[\xi] = \frac{1}{2}(X^\top\xi + \xi^\top X) = \mathrm{sym}(X^\top\xi)$ *for all* $\xi \in \mathbb{R}^{n\times p}$,
(ii) $\mathrm{Dc}(X)^{*,\mathcal{E}}[S] = XS$ *for all* $S \in \mathrm{Sym}(p)$.

Proof. The point (i) is obvious in view of (5.2). For (ii), we write, for any $S \in \mathrm{Sym}(p)$ and $\xi \in \mathbb{R}^{n\times p}$,

$$\langle \xi, \mathrm{Dc}(X)^{*,\mathcal{E}}[S]\rangle = \langle \mathrm{Dc}(X)[\xi], S\rangle = \langle \mathrm{sym}(X^\top\xi), S\rangle = \langle X^\top\xi, S\rangle = \langle \xi, XS\rangle.$$

□

We infer the following characterization of the normal space $\mathrm{N}^{\mathcal{E}}_X\mathrm{St}_{X^\top X} = \mathrm{Range}(\mathrm{Dc}(X)^{*,\mathcal{E}})$, and in the next proposition, that evaluating the pseudoinverse $\mathrm{Dc}(X)^{\dagger,\mathcal{E}}$ requires solving a Sylvester equation.

Corollary 5.3. *The normal space of $\mathrm{St}_{X^\top X}$ at $X \in \mathbb{R}^{n\times p}_*$ with respect to the Euclidean metric $g^{\mathcal{E}}$ is*

$$\mathrm{N}^{\mathcal{E}}_X\mathrm{St}_{X^\top X} = \{XS \mid S \in \mathrm{Sym}(p)\}. \tag{5.8}$$

Proposition 5.4. *The pseudoinverse $\mathrm{Dc}(X)^{\dagger,\mathcal{E}} : \mathrm{Sym}(p) \to \mathbb{R}^{n\times p}$ is the operator defined by*

$$\mathrm{Dc}(X)^{\dagger,\mathcal{E}}[T] = XS, \quad \forall T \in \mathrm{Sym}(p),$$

where S is the unique solution in $\mathrm{Sym}(p)$ to the Sylvester equation

$$\frac{1}{2}(X^\top XS + SX^\top X) = T. \tag{5.9}$$

Proof. From Proposition 5.2, it is readily seen that

$$\mathrm{Dc}(X)\mathrm{Dc}(X)^{*,\mathcal{E}}[S] = \frac{1}{2}(X^\top XS + SX^\top X),$$

from where these results follow easily. □

Consequently, the computation of the tangent and the normal terms of the landing algorithm require, in full generality, solving Sylvester equations.

Corollary 5.5. *(i) The tangent term (5.6) of the landing algorithm (3.2) with respect to the Euclidean metric given by*

$$u(X) = -\mathrm{Proj}_{X,\mathcal{E}}[\nabla_{\mathcal{E}}(f(X))] = -\nabla_{\mathcal{E}} f(X) + XS, \qquad \forall Z \in \mathbb{R}^{n\times p}, \tag{5.10}$$

where S is the unique solution in $\mathrm{Sym}(p)$ to the Sylvester equation (5.9) with $T = \mathrm{sym}(X^\top\nabla_{\mathcal{E}} f(X))$.
(ii) The normal vector field (5.7) is given by $v(X) = -XS$, where S is the solution to the Sylvester equation (5.9) with $T = \frac{1}{2}H(X)[X^\top X - \mathrm{I}_p]$.

Proof. These results are obtained by using Proposition 5.4 and formula (2.14) for the projection operator $\mathrm{Proj}_{X,\mathcal{E}}$. ☐

In the general case, we mention that computing the solution may require to calculate the singular value decomposition of X. The following result is classical and may be found in Horn and Johnson (2012).

Lemma 5.6. *Let $S \in \mathbb{R}^{p\times p}$ be a symmetric positive definite matrix and $T \in \mathbb{R}^{p\times p}$ be an arbitrary matrix. Let $X = \sum_{i=1}^{p} \sigma_i u_i v_i^\top$ be the singular value decomposition of X with singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p > 0$, left and right singular vectors $(u_i)_{1\leq i\leq p}$ and $(v_i)_{1\leq i\leq p}$. The unique solution $S \in \mathbb{R}^{p\times p}$ to the Sylvester equation (5.9) is given by*

$$S = \frac{1}{2} \sum_{1\leq i\leq j\leq p} \frac{v_i^\top T v_j}{\sigma_i^2 + \sigma_j^2} (v_i v_j^\top + v_j v_i^\top).$$

However, when $T = Q(X^\top X)$ with Q an analytic function, the solution to the Sylvester equation (5.9) is explicitly given by

$$S = Q(X^\top X)(X^\top X)^{-1}.$$

This observation leads to explicit formulas for the normal field $v(X)$ for two particular values of the operator $H(X) : \mathrm{Sym}(p) \to \mathrm{Sym}(p)$.

Proposition 5.7. *(i) If $H(X) = \mathrm{Id}_{\mathcal{F}}$, then $v(X) = -\frac{1}{2}X(\mathrm{I}_p - (X^\top X)^{-1})$.*
(ii) if $H(X) = \mathrm{Dc}(X)\mathrm{Dc}(X)^{*,\mathcal{E}}$, then $v(X) = -\nabla_{\mathcal{E}}\psi(X) = -X(X^\top X - \mathrm{I}_p)$.

Proof. For $H(X) = \mathrm{Id}_{\mathcal{F}}$, $v(X) = -XS$ with S being the solution to (5.9) with $T = Q(X^\top X)$ with $Q(x) = \frac{1}{2}(x-1)$. The solution is $S = (\mathrm{I}_p - (X^\top X)^{-1})/2$, which yields (i). For $H(X) = \mathrm{Dc}(X)\mathrm{Dc}(X)^{*,\mathcal{E}}$, (ii) can be found directly by using the identity $v(X) = -\nabla_{\mathcal{E}}\psi(X)$ with $\psi(X) = \frac{1}{4}||X^\top X - \mathrm{I}_p||_{\mathcal{F}}^2$. It is instructive, however, to retrieve this result from the result of Corollary 5.5. We have in this case that S is the solution to (5.9) the right-hand side T given by

$$\begin{aligned} T &= \frac{1}{2}\mathrm{Dc}(X)\mathrm{Dc}(X)^{*,\mathcal{E}}[X^\top X - \mathrm{I}_p] \\ &= \frac{1}{2}[X^\top(XX^\top X - X) + (X^\top XX^\top - X^\top)X] \\ &= (X^\top X)^2 - X^\top X \\ &= Q(X^\top X) \text{ with } Q(x) = x^2 - x. \end{aligned}$$

Hence, the solution to the Sylvester equation (5.9) is $S = (X^\top X - \mathrm{I}_p)$, which yields indeed (ii). ☐

To summarize, the execution of the landing algorithm (3.2) for minimizing a function on the Stiefel manifold using the Euclidean metric $g^{\mathcal{E}}$ to define the tangent and the normal terms $u(X)$ and $v(X)$ requires to solve a Sylvester equation for computing the tangent term at every iteration, which is potentially costly.

5.2. Metric design based on an explicit choice of projection operators

In this section, we adopt the reverse approach: instead of first defining the metric g and then deducing the associated tangent and normal terms $u(X)$ and $v(X)$ via (3.2) and (3.3), we begin by choosing *first* an explicit family of projection operators Proj_X onto the tangent space $\mathrm{T}_X\mathrm{St}_{X^\top X}$, or equivalently, a family of normal spaces $X \mapsto \mathrm{N}_X\mathrm{St}_{X^\top X}$. This step is outlined in details in Section 5.2.1. In the second step, we construct explicit operators

$$\widetilde{G_T}(X) : \mathrm{T}_X\mathrm{St}_{X^\top X} \to \mathrm{N}_X\mathrm{St}_{X^\top X}^{\perp,\mathcal{E}}, \qquad \widetilde{G_N}(X) : \mathrm{N}_X\mathrm{St}_{X^\top X} \to \mathrm{T}_X\mathrm{St}_{X^\top X}, \tag{5.11}$$

each admitting explicit inverses. This construction defines the metric *a posteriori* through (3.12):

$$g(\xi,\zeta) = \langle \widetilde{G_T}(X)\mathrm{Proj}_X[\xi], \zeta\rangle + \langle \widetilde{G_N}(X)\mathrm{Proj}_X^{\perp}[\xi], \zeta\rangle. \tag{5.12}$$

Moreover, these explicit inverses provide closed-form expressions for the tangent and normal terms of the landing algorithm via (3.15) and (3.16).

We consider two possible definitions of $\widetilde{G_T}$ and $\widetilde{G_N}$ for this second step. In **??**, we consider a canonical choice for these mappings, which leads to the definition of a new metric g on $\mathbb{R}^{n\times p}_*$ and new formulas for the landing algorithm for the optimization problem (5.1). Then, in Section 5.2.2, we consider instead the family of β-metrics g^β introduced in Goyens et al. (2025), which includes for $\beta = \frac{1}{2}$ the pulled-back canonical metric for the layered Stiefel manifold studied in Gao et al. (2022). We show that this family of metric is associated to the proposed family of normal spaces $X \mapsto \mathrm{N}_X\mathrm{St}_{X^\top X}$ considered in the first step, with a particular choice of $\widetilde{G_T}$ and $\widetilde{G_N}$. These operators turn to admit explicit inverses, thereby highlighting why closed-form formulas for $u(X)$ and $v(X)$ are also available in this case.

5.2.1. *Choice of oblique projection operators and associated normal spaces*

Recalling that any tangent vector $\xi \in \mathrm{T}_X\mathrm{St}_{X^\top X}$ can be written as

$$\xi = X(X^\top X)^{-1}\Omega + \Delta \text{ with } \Omega \in \mathrm{Skew}(p),$$

any matrix $Z \in \mathbb{R}^{n,p}$ can be naturally decomposed as

$$\begin{aligned} Z &= X(X^\top X)^{-1}X^\top Z + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)Z \\ &= X(X^\top X)^{-1}\,\mathrm{skew}(X^\top Z) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)Z + X(X^\top X)^{-1}\,\mathrm{sym}(X^\top Z) \\ &= \mathrm{Proj}_X[Z] + \mathrm{Proj}_X^\perp[Z], \end{aligned}$$

where Proj_X and $\mathrm{Proj}_X^\perp$ are the operator defined by

$$\mathrm{Proj}_X[Z] := X(X^\top X)^{-1}\,\mathrm{skew}(X^\top Z) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)Z, \tag{5.13}$$

$$\mathrm{Proj}_X^\perp[Z] := \mathrm{Id}_\mathcal{E} - \mathrm{Proj}_X(Z) = X(X^\top X)^{-1}\,\mathrm{sym}(X^\top Z). \tag{5.14}$$

It is straightforward to verify that Proj_X and $\mathrm{Proj}_X^\perp$ are linear (oblique) projectors. In view of (5.14), it is clear that this choice of projectors leads to the normal space

$$\mathrm{N}_X\mathrm{St}_{X^\top X} := \mathrm{Range}(\mathrm{Proj}_X^\perp) = \{X(X^\top X)^{-1}S \mid S \in \mathrm{Sym}(p)\}, \tag{5.15}$$

to be compared with (5.8). Given any symmetric operators $G_T(X) : \mathcal{E} \to \mathcal{E}$ and $G_N(X) : \mathcal{E} \to \mathcal{E}$, positive definite on respectively $\mathrm{T}_X\mathrm{St}_{X^\top X}$ and $\mathrm{N}_X\mathrm{St}_{X^\top X}$, Proj_X and $\mathrm{Proj}_X^\perp$ are orthogonal projectors for the metric (5.12) with

$$\widetilde{G_T}(X) = \mathrm{Proj}_X^{*,\mathcal{E}} G_T(X)\mathrm{Proj}_X, \quad \widetilde{G_N}(X) = (\mathrm{Proj}_X^\perp)^{*,\mathcal{E}} G_N(X)\mathrm{Proj}_X^\perp.$$

The following proposition provide expressions for the Euclidean adjoints of the tangent projection Proj_X and $\mathrm{Proj}_X^\perp$.

Proposition 5.8. *The adjoint operator* $\mathrm{Proj}_X^{*,\mathcal{E}}$ *of the projection operator* Proj_X*—with respect to the Euclidean inner product—is given by*

$$\mathrm{Proj}_X^{*,\mathcal{E}}[Z] = X\,\mathrm{skew}((X^\top X)^{-1}X^\top Z) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)Z, \qquad Z \in \mathbb{R}^{n\times p}. \tag{5.16}$$

The adjoint operator $(\mathrm{Proj}_X^\perp)^{*,\mathcal{E}}$ *of the projection operator* $\mathrm{Proj}_X^\perp$ *on the normal space is given by*

$$(\mathrm{Proj}_X^\perp)^{*,\mathcal{E}}[Z] = X\,\mathrm{sym}((X^\top X)^{-1}X^\top Z), \qquad Z \in \mathbb{R}^{n\times p}. \tag{5.17}$$

In particular, the Euclidean orthogonal of the tangent and normal spaces are

$$(\mathrm{T}_X\mathrm{St}_{X^\top X})^{\perp,\mathcal{E}} = \mathrm{Range}((\mathrm{Proj}_X^\perp)^{*,\mathcal{E}}) = \{XS \mid S \in \mathrm{Sym}(p)\},$$

$$(\mathrm{N}_X\mathrm{St}_{X^\top X})^{\perp,\mathcal{E}} = \mathrm{Range}(\mathrm{Proj}_X^{*,\mathcal{E}}) = \{X\Omega + \Delta \mid \Omega \in \mathrm{Skew}(p) \textit{ and } \Delta \in \mathbb{R}^{n\times p} \textit{ with } X^\top\Delta = 0\}.$$

Proof. By using the self-adjointness of the operators $X(X^\top X)^{-1}X^\top$ and skew, we have for any $\xi, \zeta \in \mathbb{R}^{n\times p}$,

$$\begin{aligned}\langle \mathrm{Proj}_X(\xi), \zeta\rangle &= \langle X(X^\top X)^{-1}\,\mathrm{skew}(X^\top\xi), \zeta\rangle + \langle (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\xi, \zeta\rangle \\ &= \langle \mathrm{skew}(X^\top\xi), (X^\top X)^{-1}X^\top\zeta\rangle + \langle \xi, (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta\rangle \\ &= \langle X^\top\xi, \mathrm{skew}((X^\top X)^{-1}X^\top\zeta)\rangle + \langle \xi, (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta\rangle \\ &= \langle \xi, X\,\mathrm{skew}((X^\top X)^{-1}X^\top\zeta) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta\rangle.\end{aligned}$$

Formula (5.17) can be obtained identically, or from $(\mathrm{Proj}_X^\perp)^{*,\mathcal{E}} = \mathrm{Id}_\mathcal{E} - \mathrm{Proj}_X^{*,\mathcal{E}}$. ☐

We mention the following characterization of the adjoint of the operator $\mathrm{D}c(X)$ with respect to the metric g defined in (5.12).

Proposition 5.9. *For any $S \in \mathrm{Sym}(p)$,*

$$\mathrm{D}c(X)^{*,g}[S] = \widetilde{G_N}(X)^{-1}[XS],$$

where $\widetilde{G_N}(X)^{-1} : (\mathrm{T}_X\mathrm{St}_{X^\top X})^{\perp,\mathcal{E}} \to \mathrm{N}_X\mathrm{St}_{X^\top X}$ is the inverse of the operator $\widetilde{G_N}$ defined in (3.13).

Proof. Due to (2.12) and Proposition 5.2:

$$\mathrm{D}c(X)^{*,g}[S] = G(X)^{-1}\mathrm{D}c(X)^{*,\mathcal{E}}[S] = G(X)^{-1}[XS]. \tag{5.18}$$

Then, in view of Proposition 5.8,

$$\mathrm{Proj}_X^{*,\mathcal{E}}[XS] = 0, \quad (\mathrm{Proj}_X^\perp)^{*,\mathcal{E}}[XS] = XS.$$

The result follows by applying (3.14) to (5.18). ☐

From Section 3.2, the tangent and normal steps of the landing algorithm (3.2) read

$$u(X) = -\widetilde{G_T}(X)^{-1}\mathrm{Proj}_X^{*,\mathcal{E}}(\nabla_\mathcal{E} f(X)), \tag{5.19}$$

$$v(X) = -\mathrm{Proj}_X^\perp \mathrm{D}c(X)^{\dagger,\mathcal{E}}(H(X)[c(X)]). \tag{5.20}$$

It is clear from these equations that the choice of the operators $\widetilde{G_T}(X)^{-1}$ and $H(X)$ offer some latitude for the definition of $u(X)$ and $v(X)$ respectively, with clear physical interpretations for the optimization scheme. Thanks to the results of the previous Section 5.1, we obtain the following formulas for $u(X)$ and $v(X)$.

Proposition 5.10. *The tangent and normal terms (eq. (5.19) and (5.20)) of the landing algorithm (3.2) relatively to the metric* **??** *are given by*

$$u(X) = -\widetilde{G_T}(X)^{-1}[X\,\mathrm{skew}((X^\top X)^{-1}X^\top\nabla_\mathcal{E} f(X)) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\nabla_\mathcal{E} f(X)], \tag{5.21}$$

$$v(X) = -X(X^\top X)^{-1}\,\mathrm{sym}(X^\top XS), \tag{5.22}$$

where S is the solution to the Sylvester equation (5.9) with T given by **??***. The expression of $v(X)$ becomes explicit in the following cases:*

(i) if $H(X) = \mathrm{Id}_\mathcal{E}$, it still reads $v(X) = -\frac{1}{2}X(\mathrm{I}_p - (X^\top X)^{-1})$;
(ii) if $H(X) = \mathrm{D}c(X)\mathrm{D}c(X)^{*,g}$, it reads $v(X) = -\frac{1}{2}\widetilde{G_N}(X)^{-1}[X(X^\top X - \mathrm{I}_p)]$;
(iii) if $H(X) = \mathrm{D}c(X)\mathrm{D}c(X)^{*,\mathcal{E}}$, it reads again $v(X) = -X(X^\top X - \mathrm{I}_p)$.

Proof. Formulas (5.21) follows immediately from (5.19) and (5.16). Formula (5.22) is obtained by combining (5.20) and the result of (5.6). For cases (i) and (iii), we observe that

$$v(X) = \mathrm{Proj}_X^\perp(v(X)^\mathcal{E})$$

where $v(X)^\mathcal{E}$ is the corresponding normal direction obtained in Proposition 5.7. Since

$$X(\mathrm{I}_p - (X^\top X)^{-1}) = X(X^\top X)^{-1}(X^\top X - \mathrm{I}_p) \in \mathrm{N}_X\mathrm{St}_{X^\top X},$$

$$X(X^\top X - \mathrm{I}_p) = X(X^\top X)^{-1}((X^\top X)^2 - X^\top X) \in \mathrm{N}_X\mathrm{St}_{X^\top X},$$

we have in both cases $v(X) = v(X)^{\mathcal{E}}$, which proves (i) and (iii). Finally, $H(X) = \mathrm{D}c(X)\mathrm{D}c(X)^{*,g}$ implies that $v(X) = -\mathrm{D}c(X)^{*,g}c(X)$, which yields (ii) after using Proposition 5.9. $\square$

5.2.2. *Construction of the tangent and normal steps with the β-metric*

while the influence of the choice of β is *a priori* less intuitive (we make it explicit in FF: proposition below).

It turns out that these projectors are exactly those derived in Goyens et al. (2025) where a family of β-metric g^β for the ambient space $\mathcal{E}$ was considered *first*,

$$g^\beta(\xi,\zeta) = \left\langle \left(\mathrm{I} - (1-\beta)X(X^\top X)^{-1}X^\top\right)\zeta(X^\top X)^{-1}, \xi\right\rangle, \tag{5.23}$$

from where Proj_X was shown to be the orthogonal projection on the tangent space. This β-metric is the pullback of a generalization of the canonical metric Edelman et al. (1998) on St to $\mathrm{St}_{X^\top X}$. We shall enlighten the definition of these metric by constructing it following the framework of Section 3.2. We first show that the β-metric (5.23) is the metric **??** with a particular choice of $G_T(X)$ and $G_N(X)$.

Proposition 5.11. *Let $\beta > 0$. The β-metric (5.23) is the metric* **??** *with $G_T(X) : \mathcal{E} \to \mathcal{E}$ and $G_N(X) : \mathcal{E} \to \mathcal{E}$ being the symmetric operators defined by*

$$G_T(X)\xi := ((\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\xi + \beta X(X^\top X)^{-1}X^\top\xi)(X^\top X)^{-1}, \tag{5.24}$$

$$G_N(X)\xi := \beta X(X^\top X)^{-1}X^\top\xi(X^\top X)^{-1}. \tag{5.25}$$

The operator $G_T(X)$ is positive definite on $\mathbb{R}^{n\times p}$ and the operator $G_N(X)$ is positive definite on $\mathrm{N}_X\mathrm{St}_{X^\top X}$.

Proof. Let us denote by $\Pi_X := X(X^\top X)^{-1}X^\top$ the projection matrix onto the image space of X. First, we observe that for $\xi,\zeta \in \mathbb{R}^{n\times p}$,

$$g^\beta(\xi,\zeta) = \beta\langle \Pi_X\xi(X^\top X)^{-1}, \Pi_X\zeta\rangle + \langle (\mathrm{I}_n - \Pi_X)\xi(X^\top X)^{-1}, (\mathrm{I}_n - \Pi_X)\zeta\rangle.$$

Consequently,

$$\begin{aligned} g^\beta(\mathrm{Proj}_X(\xi), \mathrm{Proj}_X^\perp(\zeta)) &= \langle X(X^\top X)^{-1}\operatorname{skew}(X^\top\xi)(X^\top X)^{-1}, X(X^\top X)^{-1}\operatorname{sym}(X^\top\zeta)\rangle \\ &= \langle \operatorname{skew}(X^\top\xi), (X^\top X)^{-1}\operatorname{sym}(X^\top\zeta)(X^\top X)^{-1}\rangle \\ &= 0, \end{aligned}$$

due to the Frobenius orthogonality between skew and symmetric matrices. This shows the orthogonality between the tangent space $\mathrm{T}_X\mathrm{St}_{X^\top X}$ and the normal space $\mathrm{N}_X\mathrm{St}_{X^\top X}$ of (5.15) for the metric g^β. Then, it follows that

$$g^\beta(\xi,\zeta) = g^\beta(\mathrm{Proj}_X(\xi), \mathrm{Proj}_X(\zeta)) + g^\beta(\mathrm{Proj}_X^\perp(\xi), \mathrm{Proj}_X^\perp(\zeta)),$$

with

$$\begin{aligned} &g^\beta(\mathrm{Proj}_X(\xi), \mathrm{Proj}_X(\zeta)) \\ &\qquad = \beta\langle \Pi_X \mathrm{Proj}_X(\xi)(X^\top X)^{-1}, \mathrm{Proj}_X(\zeta)\rangle_{\mathcal{E}} + \langle (\mathrm{I}_n - \Pi_X)\mathrm{Proj}_X(\xi)(X^\top X)^{-1}, \mathrm{Proj}_X(\zeta)\rangle_{\mathcal{E}} \\ &\qquad = \langle G_T(X)\mathrm{Proj}_X(\xi), \mathrm{Proj}_X(\zeta)\rangle_{\mathcal{E}}, \\ &g^\beta(\mathrm{Proj}_X^\perp(\xi), \mathrm{Proj}_X^\perp(\zeta)) = \beta\langle \Pi_X \mathrm{Proj}_X(\xi)(X^\top X)^{-1}, \mathrm{Proj}_X(\zeta)\rangle_{\mathcal{E}} \\ &\qquad\qquad = \langle G_N(X)\mathrm{Proj}_X(\xi), \mathrm{Proj}_X(\zeta)\rangle_{\mathcal{E}}. \end{aligned}$$

The positive definiteness of $G_T(X)$ follows from

$$\langle G_T(X)\xi, \xi\rangle_{\mathcal{E}} = \beta||\Pi_X\xi(X^\top X)^{-\frac{1}{2}}||_{\mathcal{E}}^2 + ||(\mathrm{I}_n - \Pi_X)\xi(X^\top X)^{-\frac{1}{2}}||_{\mathcal{E}}^2,$$

while for $G_N(X)$, we have

$$\begin{aligned}\forall S \in \mathrm{Sym}(p),\ &\langle G_N(X)X(X^\top X)^{-1}S, X(X^\top X)^{-1}S\rangle_{\mathcal{E}} \\ &= \beta\langle X(X^\top X)^{-1}S(X^\top X)^{-1}, X(X^\top X)^{-1}S\rangle_{\mathcal{E}} \\ &= \beta||(X^\top X)^{-\frac{1}{2}}S(X^\top X)^{-\frac{1}{2}}||^2_{\mathbb{R}^{p\times p}}.\end{aligned}$$

□

It remains to choose the operator $\widetilde{G_T}(X)^{-1}$ in (5.21), and the operator $\widetilde{G_N}(X)^{-1}$ if one chooses (ii) for the normal direction. There is actually a natural choice of operators $\widetilde{G_T}(X)$: $\mathrm{T}_X\mathrm{St}_{X^\top X} \to \mathrm{N}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X}$ and $\widetilde{G_N}(X)$: $\mathrm{N}_X\mathrm{St}_{X^\top X} \to \mathrm{T}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X}$ that lead to explicit inverses $\widetilde{G_T}(X)^{-1}$ and $\widetilde{G_N}(X)^{-1}$. We consider the mappings

$$\text{(5.26)}\qquad \widetilde{G_T}(X): \begin{array}{rcl} \mathrm{T}_X\mathrm{St}_{X^\top X} & \longrightarrow & \mathrm{N}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} \\ X(X^\top X)^{-1}\Omega + \Delta & \longmapsto & X\Omega + \Delta, \end{array}$$

$$\text{(5.27)}\qquad \widetilde{G_N}(X): \begin{array}{rcl} \mathrm{N}_X\mathrm{St}_{X^\top X} & \longrightarrow & \mathrm{T}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} \\ X(X^\top X)^{-1}S & \longmapsto & XS, \end{array}$$

where $\Omega \in \mathrm{Skew}(p)$, $\Delta \in \mathbb{R}^{n\times p}$ with $\Delta^T X = 0$ and $S \in \mathrm{Sym}(p)$. These mappings can be explicitly extended to the whole $\mathcal{E} = \mathbb{R}^{n\times p}$:

$$\text{(5.28)}\qquad G_T(X)Z := XX^\top Z + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)Z,$$

$$\text{(5.29)}\qquad G_N(X)Z := XX^\top Z.$$

Proposition 5.12. *The operators $G_T(X)$ and $G_N(X)$ of (5.28) and (5.29) are symmetric positive operators on $\mathcal{E}$, definite respectively on $\mathcal{E}$ and $\mathrm{N}_X\mathrm{St}_{X^\top X}$. Formula (3.12) defines thus the associated metric*

$$\text{(5.30)}\qquad \begin{aligned} g(\xi,\zeta) &= \langle \xi, \mathrm{Proj}_X^{*,\mathcal{E}} G_T(X)\mathrm{Proj}_X\zeta\rangle + \langle \xi, (\mathrm{Proj}_X^\perp)^{*,\mathcal{E}} G_N(X)\mathrm{Proj}_X^\perp\zeta\rangle \\ &= \langle \xi, (XX^\top + \mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta\rangle. \end{aligned}$$

Proof. The symmetry, the positivity of G_T and G_N is clear, as well as the definiteness of G_T on $\mathcal{E}$. The definiteness of G_N on $\mathrm{N}_X\mathrm{St}_{X^\top X}$ follows from the inequality

$$\langle X(X^\top X)^{-1}S, G_N(X)X(X^\top X)^{-1}S\rangle = \langle X(X^\top X)^{-1}S, XS\rangle = \langle S,S\rangle, \qquad \forall S\in\mathrm{Sym}(p).$$

We then find that

$$\begin{aligned}\mathrm{Proj}_X^{*,\mathcal{E}} G_T(X)\mathrm{Proj}_X\zeta &= \mathrm{Proj}_X^{*,\mathcal{E}}[X\,\mathrm{skew}(X^\top\zeta) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta] \\ &= X\,\mathrm{skew}(X^\top\zeta) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta,\end{aligned}$$

$$(\mathrm{Proj}_X^\perp)^{*,\mathcal{E}} G_N(X)\mathrm{Proj}_X^\perp\zeta = (\mathrm{Proj}_X^\perp)^{*,\mathcal{E}}[X\,\mathrm{sym}(X^\top\zeta)] = X\,\mathrm{sym}(X^\top\zeta).$$

With these formulas, we obtain thus

$$\begin{aligned} g(\xi,\zeta) &= \langle \xi, X\,\mathrm{skew}(X^\top\zeta) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta\rangle + \langle\xi, X\,\mathrm{sym}(X^\top\zeta)\rangle \\ &= \langle \xi, (XX^\top + \mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta\rangle. \end{aligned}$$

□

We observe that the metric $g(\xi,\zeta)$ is not a β-metric (5.23) for any value of β, despite both generate the same family of normal spaces $\mathrm{N}_X\mathrm{St}_{X^\top X}$. We can now provide a (new) explicit formula for the tangent step $u(X) = -\mathrm{grad}^g_{\mathrm{St}_{X^\top X}} f(X)$ and the normal step $v(X) = -\nabla_g\psi(X)$ in this metric (corresponding to $H(X) = \mathrm{D}c(X)\mathrm{D}c(X)^{*,g}$). Remarkably, $v(X)$, which is the Riemannian gradient of the penalty function $\psi(X)$, is also exactly the 'Newton-like' direction in this metric g.

Proposition 5.13. *With the metric g defined in (5.30), the tangent and normal directions of the landing algorithm (3.2) read*

$$u(X) = -X(X^\top X)^{-1}\operatorname{skew}((X^\top X)^{-1}X^\top \nabla_{\mathcal{E}} f(X)) - (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\nabla_{\mathcal{E}} f(X), \tag{5.31}$$

$$v(X) = -\frac{1}{2}X(\mathrm{I}_p - (X^\top X)^{-1}). \tag{5.32}$$

Proof. The inverse of the operators $\widetilde{G_T}(X)$ and $\widetilde{G_N}(X)$ read obviously

$$\begin{array}{rrcl} \widetilde{G_T}(X)^{-1} : & \mathrm{N}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} & \longrightarrow & \mathrm{T}_X\mathrm{St}_{X^\top X} \\ & X\Omega + \Delta & \longmapsto & X(X^\top X)^{-1}\Omega + \Delta, \end{array} \tag{5.33}$$

$$\begin{array}{rrcl} \widetilde{G_N}(X)^{-1} : & \mathrm{T}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} & \longrightarrow & \mathrm{N}_X\mathrm{St}_{X^\top X} \\ & XS & \longmapsto & X(X^\top X)^{-1}S. \end{array} \tag{5.34}$$

Formula (5.31) follow from (5.21), and combining (5.34) with the item (ii) of Proposition 5.10 yields

$$v(X) = -\frac{1}{2}X(X^\top X)^{-1}(X^\top X - \mathrm{I}_p) = -\frac{1}{2}X(\mathrm{I}_p - (X^\top X)^{-1}).$$

□

We conclude this section by deriving the analogous result for β-metric (5.23). The following proposition highlights that this metric leads to explicit inverses for the tangent and normal metric representers $\widetilde{G_T}(X)$ and $\widetilde{G_N}(X)$, which is the reason why this metric leads to explicit formulas for the tangent and the normal terms.

Proposition 5.14. *If $g \equiv g^\beta$ is the β-metric (5.23), the inverse operators $\widetilde{G_T}(X)^{-1}$ and $\widetilde{G_N}(X)^{-1}$ are explicit and given by*

$$\begin{array}{rrcl} \widetilde{G_T}(X)^{-1} : & \mathrm{N}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} & \longrightarrow & \mathrm{T}_X\mathrm{St}_{X^\top X} \\ & X\Omega + \Delta & \longmapsto & \frac{1}{\beta}X\Omega X^\top X + \Delta X^\top X, \end{array} \tag{5.35}$$

$$\begin{array}{rrcl} \widetilde{G_N}(X)^{-1} : & \mathrm{T}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} & \longrightarrow & \mathrm{N}_X\mathrm{St}_{X^\top X} \\ & XS & \longmapsto & \frac{1}{\beta}XS(X^\top X), \end{array} \tag{5.36}$$

for any $\Omega \in \mathrm{Skew}(p)$, $\Delta \in \mathbb{R}^{n\times p}$ with $\Delta^\top X = 0$ and $S \in \mathrm{Sym}(p)$. Consequently, the tangent and normal directions of the landing algorithm (3.2) with $H(X) = \mathrm{D}c(X)\mathrm{D}c(X)^{*,g^\beta}$ read

$$\begin{aligned} u(X) &= -\operatorname{grad}^{g^\beta}_{\mathrm{St}_{X^\top X}} f(X) \\ &= -\frac{1}{\beta}X \operatorname{skew}\big((X^\top X)^{-1}X^\top \nabla_{\mathcal{E}} f(X)\big)\, X^\top X - (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\nabla_{\mathcal{E}} f(X)\, X^\top X \end{aligned} \tag{5.37}$$

$$\begin{aligned} v(X) &= -\nabla_{g^\beta}\psi(X) \\ &= -\frac{1}{2\beta}X(X^\top X - \mathrm{I}_p)X^\top X \end{aligned} \tag{5.38}$$

Proof. Using (5.24) and (5.25), we see that $\widetilde{G_T}(X)$ and $\widetilde{G_N}(X)$ are the operators

$$\begin{array}{rrcl} \widetilde{G_T}(X) : & \mathrm{T}_X\mathrm{St}_{X^\top X} & \longrightarrow & \mathrm{N}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} \\ & X(X^\top X)^{-1}\Omega + \Delta & \longmapsto & \beta X(X^\top X)^{-1}\Omega(X^\top X)^{-1} + \Delta(X^\top X)^{-1}, \end{array}$$

$$\begin{array}{rrcl} \widetilde{G_N}(X) : & \mathrm{N}_X\mathrm{St}_{X^\top X} & \longrightarrow & \mathrm{T}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} \\ & X(X^\top X)^{-1}S & \longmapsto & \beta X(X^\top X)^{-1}S(X^\top X)^{-1}, \end{array}$$

whose inverses are obviously given by (5.35) and (5.36). Inserting these formulas into (5.19) and item (ii) of Proposition 5.10 yields the result. □

The reader may verify that (5.37) coincide with the formula found in (Goyens et al., 2025, eq. (35)). In the latter work, the normal term was considered to be the Euclidean gradient $v(X) = -\nabla_{\mathcal{E}}\psi(X)$, which corresponds to (5.20) with $H(X) = \mathrm{D}c(X)\mathrm{D}c(X)^{*,\mathcal{E}}$ (item (iii) of Proposition 5.10).

References

Ablin, P. and Peyré, G. (2022). Fast and accurate optimization on the orthogonal manifold without retraction. In *International Conference on Artificial Intelligence and Statistics*, pages 5636–5657. PMLR.

Absil, P.-A., Mahony, R., and Sepulchre, R. (2008). Optimization algorithms on matrix manifolds.

Byrd, R. H., Curtis, F. E., and Nocedal, J. (2008). An Inexact SQP Method for Equality Constrained Optimization. *SIAM Journal on Optimization*, 19(1):351–369.

Edelman, A., Arias, T. A., and Smith, S. T. (1998). The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353.

Feppon, F. (2024). Density-based topology optimization with the Null Space Optimizer: A tutorial and a comparison. *Structural and Multidisciplinary Optimization*, 67(1).

Feppon, F., Allaire, G., and Dapogny, C. (2020). Null space gradient flows for constrained optimization with applications to shape optimization. *ESAIM: Control, Optimisation and Calculus of Variations*, 26:90.

Feppon, F. and Lermusiaux, P. F. J. (2019). The Extrinsic Geometry of Dynamical Systems Tracking Nonlinear Matrix Projections. *SIAM Journal on Matrix Analysis and Applications*, 40(2):814–844.

Gao, B., Vary, S., Ablin, P., and Absil, P.-A. (2022). Optimization flows landing on the Stiefel manifold. *http://sites.uclouvain.be/absil/2022.03*, 55(30):25–30.

Goyens, F., Absil, P.-A., and Feppon, F. (2025). Geometric design of the tangent term in landing algorithms for orthogonality constraints. *arXiv preprint arXiv:2507.15638*.

Horn, R. A. and Johnson, C. R. (2012). *Matrix Analysis*. Cambridge University Press, Cambridge, 2 edition.

Mishra, B. and Sepulchre, R. (2016). Riemannian preconditioning. *SIAM Journal on Optimization*, 26(1):635–660.

Nocedal, J. and Wright, S. J. (2006). *Numerical optimization*. Springer.

Yamashita, H. (1980). A differential equation approach to nonlinear programming. *Mathematical Programming*, 18(1):155–168.