# THE RIEMANNIAN LANDING METHOD: FROM PROJECTED GRADIENT FLOWS TO SQP

FLORENTIN GOYENS[1*], FLORIAN FEPPON[2]

[1] *ICTEAM Institute, UCLouvain, Louvain-la-Neuve, Belgium*

[2] *NUMA Unit, Department of Computer Science, KU Leuven, Belgium.*

Abstract. Landing methods have recently emerged in Riemannian matrix optimization as efficient schemes for handling nonlinear equality constraints without resorting to costly retractions. These methods decompose the search direction into tangent and normal components, enabling asymptotic feasibility while maintaining inexpensive updates. In this work, we provide a unifying geometric framework that reveals the landing algorithm to encompass several classical optimization methods, under suitable choices of Riemannian metric, such as projected and null-space gradient flows, Sequential Quadratic Programming (SQP), and a certain form of the Augmented Lagrangian method. In particular, we show that a quadratically convergent landing method essentially reproduces the quadratically convergent SQP method. These connections also allow us to propose a globally convergent method using adaptive step sizes. The backtracking linesearch satisfies an Armijo condition on a merit function, and does not require a priori knowledge of Lipschitz constants.

Our second key contribution is to analyze landing methods through a geometric parameterization of the metric in terms of fields of oblique projectors and associated metric restrictions. This viewpoint disentangles the roles of orthogonality, tangent, and normal metrics, and elucidates how to design the metric so as to obtain explicit tangent and normal updates. For matrix optimization, this framework not only recovers recent constructions in the literature for problems with orthogonality constraints, but also provides systematic guidelines for designing new metrics that admit closed-form search directions.

**Keywords.** Landing algorithm, Constrained nonlinear optimization, Stiefel manifold, Projected gradient flows, Sequential Quadratic Programming, Oblique projections, Differential geometry in optimization.

**AMS 2020 Subject classifications.** 49M05, 65K10, 49M29, 90C30, 53C21

---

Contents



*Date*: January 26, 2026.

[*] Corresponding author. Email: `goyensflorentin@gmail.com`.

**Acknowledgements.** This work was supported by the Fonds de la Recherche Scientifique-FNRS under Grant no T.0001.23. F. Feppon was supported by the Flanders Research Foundation (FWO) under Grant G001824N.

## 1. Introduction

The optimization of functions subject to nonlinear constraints is a central problem across engineering, scientific computing, and machine learning. In recent years, the concept of *landing methods* has emerged in the Riemannian matrix optimization community after the works of Ablin and Peyré (2022) and Gao et al. (2022). The essential idea of the landing algorithm is to decompose the search direction

$$d = d_T + d_N \tag{1.1}$$

into a *tangent* component $d_T$, which is responsible for decreasing the objective function without worsening the constraint violation, and a *normal* component $d_N$, whose role is to progressively drive iterates toward the constraint manifold by reducing infeasibility.

In the context of matrix optimization, this approach enables one to enforce orthogonality constraints asymptotically, thus avoiding the computationally expensive *retraction* operations—i.e., nonlinear projections onto the feasible set—required by standard feasible Riemannian algorithms (Absil et al., 2008; Absil and Malick, 2012). Since evaluating the normal term $d_N$ is typically computationally cheaper than computing retractions, the method is appealing and makes geometric optimization potentially viable for large-scale problems, such as imposing the orthogonality of weight matrices during the training of Large Language Models (LLMs) or other resource-intensive machine learning tasks (Hu et al., 2025).

Consider an equality-constrained optimization problem of the form

$$\begin{cases} \min\limits_{x\in\mathcal{E}} & f(x) \\ \text{s.t.} & c(x) = 0, \end{cases} \tag{P}$$

where $f : \mathcal{E} \to \mathbb{R}$ and $c : \mathcal{E} \to \mathcal{F}$ are continuously differentiable—possibly nonconvex—mappings between finite-dimensional vector spaces $\mathcal{E}$ and $\mathcal{F}$ with $\dim(\mathcal{F}) < \dim(\mathcal{E})$. In its most general formulation, the landing algorithm can be written as

$$x_{k+1} = x_k + \alpha_k(d_T(x_k) + d_N(x_k)), \tag{1.2}$$

where $\alpha_k > 0$ is an adaptive step size, and

$$d_T(x) = -\operatorname{grad}^g_{\mathcal{M}_x} f(x), \tag{1.3}$$

$$d_N(x) = -\nabla_g \psi(x) \text{ with } \psi(x) = ||c(x)||^2_{\mathcal{F}}/2. \tag{1.4}$$

The tangent and normal directions $d_T(x)$ and $d_N(x)$ correspond respectively to the *Riemannian constrained gradient* of $f$ on the level curve

$$\mathcal{M}_x := \{y \in \mathcal{E} : \, c(y) = c(x)\}, \tag{1.5}$$

and the *unconstrained Riemannian gradient* of the infeasibility measure $\psi(x) := \|c(x)\|^2_{\mathcal{F}}/2$. The role of the tangent term is to decrease the objective function $f$ without increasing the violation of the constraints, while the purpose of the normal term is to ‘land’ the iterates $x_k$ back toward feasibility by descending along the constraint violation $c(x)$. Importantly, these gradients are computed with respect to a (user-defined) metric field $g$ that endows $\mathcal{E}$ with a Riemannian structure. When $\mathcal{E} = \mathbb{R}^n$ and $\mathcal{F} = \mathbb{R}^m$ are equipped with the standard Euclidean inner product, one obtains $d_N(x) = -\nabla c(x)\, c(x)$ and

$$d_T(x) = \big(\mathrm{I}_n - \nabla c(x)^\top (\nabla c(x) \nabla c(x)^\top)^{-1} \nabla c(x)\big) \nabla f(x), \tag{1.6}$$

the orthogonal projection of $\nabla f(x)$ onto the tangent space of $\mathcal{M}_x$.

The first contribution of this paper is to show that, *under suitable choices of the metric field $g$*, the landing method (1.2) is equivalent to existing algorithms that have emerged under different names in other areas of optimization, such as the *Corrective Gradient Projection* method (Frost, 1972; Gangi and Byun, 1976), *projected gradient flows* (Tanabe, 1980; Yamashita, 1980; Schropp and Singer, 2000), and *null-space gradient flows* (Feppon et al., 2020a; Feppon, 2024). A more detailed account of the historical development of methods exploiting the decomposition of the search direction into tangent and normal components is given below. In fact, we show in section 4.2 that the scheme (1.2) includes, as particular cases, the Sequential Quadratic Programming (SQP) algorithm (Boggs and Tolle, 1995) in its basic form (without trust regions), as well as an Augmented Lagrangian method (Tapia, 1977) with a particular update rule for the multipliers. Although the introduction of landing methods for matrix optimization can thus be viewed as a resurgence of older classical methods, the questions of how to choose the metric $g$ and how to derive global convergence guarantees for the iterative scheme (1.2), rather than for its interpretation as the discretization of an Ordinary Differential Equation (ODE), are relatively recent and open new perspectives on existing algorithms.

Building on these connections, we propose an adaptive procedure for selecting the step size $\alpha_k$, based on an Armijo condition for the decrease of a merit function, that guarantees global convergence of the algorithm (1.2) toward a feasible point satisfying first-order optimality conditions. The procedure and its analysis rely heavily on standart techniques for the globalization of SQP (Nocedal and Wright, 2006; Curtis et al., 2024). Prior works proving the asymptotic convergence of the discrete iterates assumed constant step sizes, where global convergence is obtained for a step size smaller than some unknown constant (Schechtman et al., 2023; Zhang et al., 2024; Song et al., 2025). The present work thus fills an important gap, since landing schemes with constant step sizes are very sensitive to parameter tuning. Incidentally, a landing scheme with adaptive step sizes has been recently proposed in the concurent work (Shi and Wang, 2025), but it relies on Lipschitz constants, while our linesearch does not rely on any problem constant.

Our second key contribution is to make explicit how (1.3) and (1.4) depend on the metric field $g$, and how to design it to (i) make (1.3) quadratically convergent, or (ii), to facilitate the computation of the tangent and normal steps (1.3) and (1.4).

Regarding the first goal, we realize that setting naturally the tangent term $d_T(x)$ to a Riemannian Newton step, the landing scheme (1.2) achieves quadratic convergence for unit step sizes *only* for a carefully designed normal step $d_N(x)$ such that (1.2) essentially reproduces SQP iterates (proposition 5.6). In reference to the leitmotiv "all roads lead to Newton" proclaimed in (Absil et al., 2009), we can thus state here that *all roads lead to SQP* for unfeasible constrained optimization methods.

Addressing the second goal leads us to introduce the following viewpoint to design the metric field $g$: *first*, associate a normal space $\mathrm{N}_x\mathcal{M}_x$ to every tangent space $\mathrm{T}_x\mathcal{M}_x$ of the constraint manifold $\mathcal{M}_x$. *Then*, define the value of the restriction of the metric to the tangent and normal spaces (see (4.6)). Mathematically, this is requires specifying a family of *oblique* projectors $x \mapsto \mathrm{Proj}_x$ on the tangent space at $x \in \mathcal{M}_x$. The corresponding tangent and normal directions $d_T(x_k)$ and $d_N(x_k)$ can be expressed explicitly in terms of the oblique projector and these metric restrictions (proposition 4.3). In the context of Riemannian matrix optimization with orthogonality constraint, computing the tangent direction requires in general solving a linear system involving the inversion of $\nabla c(x)\nabla c(x)^\top$, which translates to solving Sylvester equations.

Through a careful selection of the projector mapping and the tangent and normal metric restrictions, fully explicit formulas for $d_T(x)$ and $d_N(x)$ can be obtained, thereby enabling more efficient evaluations.

The article is organized as follows. Section 2 provides a detailed account of historical developments of constrained optimization methods closely related to the landing approach.

Section 3 reviews the background in differential geometry needed to introduce the notation conventions adopted in this work.

Section 4 establishes a formal link between the landing method and earlier projected gradient flows approaches. Such is achieved by exploiting the parameterization of the metric via a field of oblique projectors, together with the tangent and normal metric restrictions.

Section 5 establishes new connections between the landing methods and existing algorithms. We first show that (1.2) encompasses a version of the Augmented Lagrangian method with least-squares multiplier update, before proving that the Sequential Quadratic Programming (SQP) method is a particular instance of the Riemannian landing method. Then, we deepen known connections between SQP and Riemannian Newton method, formally establishing that a locally quadratically convergent landing method (1.2) is locally quadratically convergent essentially reproduces quadratically convergent SQP iterates, up to a negligible correction of the search direction.

In section 6, we establish global convergence of the Riemannian Landing Method in its most general form (1.2), with adaptive step sizes. The framework adapts well-known ideas for the gloablization of SQP methods based on a merit function.

Finally, section 7 presents several metric designs for optimization under orthogonality constraints, guided by a natural explicit choice of oblique projectors onto the tangent spaces of the constraint manifold. We detail the steps leading to explicit formulas for the normal and tangent components $d_T(x)$ and $d_N(x)$. This viewpoint clarifies why the metric fields considered in recent publications, e.g., (Ablin and Peyré, 2022; Goyens et al., 2026), lead to explicit formulas for $d_T(x)$ and $d_N(x)$, and it also allows us to determine how to design new metrics possessing these properties.

## 2. Historical developments of landing-related methods

To the best of our knowledge, the first trace of an algorithm decomposing the search direction as in (1.1) is the *Gradient Projection Method* formalized by Rosen (1960, 1961). For nonlinear constraints, this method projects at every step the gradient of the objective function onto the tangent constraint hyperplane—the null space of the linearized constraints—using (1.6), or onto the tangent cone in the case of inequality constraints. A similar algorithm was independently devised by Booker and Ong (1971) for geophysics signal processing with equality constraints. Shortly thereafter, the convergence rate of the method, considering geodesic retractions and equality constraints, was formally analyzed by Luenberger (1972).

However, finite steps on curved manifolds induce numerical infeasibility, which raised the need for an explicit correction term. This led to the development of the *Corrective Gradient Projection (CGP)* method (Frost, 1972; Gangi and Byun, 1976), which augments the tangent direction with a normal component correcting deviations from the constraints. This normal component,

$$d_N(x) = -\nabla c(x)^\top (\nabla c(x) \nabla c(x)^\top)^{-1} c(x), \tag{2.1}$$

acts as a 'pseudoinverse' or 'Newton-like' step for finding a zero of the constraint.

In parallel with the development of CGP, Tanabe (1974) proposed to interpret Rosen's gradient projection method as the discretization of the dynamical system $\dot{x} = d_T(x)$. Later, Yamashita (1980) introduced a corrected version of this projected gradient flow,

$$\begin{aligned}\dot{x} &= d_T(x) + d_N(x) \\ &= -(\mathrm{I}_n - \nabla c(x)^\top (\nabla c(x) \nabla c(x)^\top)^{-1} \nabla c(x)) \nabla f(x) - \nabla c(x)^\top (\nabla c(x) \nabla c(x)^\top)^{-1} c(x),\end{aligned} \tag{2.2}$$

rediscovering the pseudoinverse step of Frost (1972). This flow was later considered by Evtushenko and Zhadan (1994) to develop a barrier method for linear programming. Subsequently, Schropp and Singer (2000) proved that any bounded solution of (2.2) converges to a steady-state equilibrium, and stable equilibria coincide with local minimizers of (P). Schropp and Singer (2000) suggested that inequality constraints can be addressed with (2.2) through equalization via the introduction of additional slack variables; however, this produces an exponential growth in false equilibrium points as the number of inactive constraints increases (Jongen and Stein, 2003). To overcome this, Jongen and Stein (2004) and Shikhman and Stein (2009) proposed variations of (2.2) (without the corrective term), yielding modified gradient flows that implicitly incorporates the effects of slack variables.

A distinct differential-equation-based approach for problems with equality and inequality constraints emerged in shape and topology optimization under the name *null space gradient flows* or *Null Space Optimizer* (Feppon et al., 2020a). This method extends the projected flow (2.2) to general constrained problems and is compatible with the infinite-dimensional setting of shape optimization based on Hadamard's boundary variation method (Henrot and Pierre, 2018; Allaire et al., 2021). This compatibility is possible thanks to the central role of the metric in converting differential into gradients and the interpretation of shape updates as retractions on an abstract manifold. Its tangent direction is the projection of the objective gradient onto the tangent cone of active or violated constraints, akin to the gradient projection of Rosen (1961), here computed by solving a quadratic program. The absence of hyperparameters makes the method particularly effective for topology and shape optimization (Feppon et al., 2020b). An alternative construction of tangent and normal terms for equality-constrained shape optimization was proposed in Wegert et al. (2023). A variant of the Null Space Optimizer for dealing with constraints with sparse Jacobian matrix was later introduced in Feppon (2024). Before these lines of work, let us mention that related gradient projection ideas had appeared in topology optimization: Yulin and Xiaoming (2004) projected shape gradients on tangent cones, while Barbarosie et al. (2020) proposed a method close to the Null Space Optimizer for equality constraints, together with a heuristic active-set strategy to handle inequalities.

More recently, the *landing* method was introduced in Ablin and Peyré (2022); Gao et al. (2022) for matrix optimization with orthogonality constraints without the need for retractions. This contrasts with the so far predominant Riemannian optimization methods, such as gradient flows or Riemannian Newton methods on the constraint manifold, which leverage intrinsic differential geometry and retractions Absil et al. (2008); Absil and Malick (2012).

Although the landing method introduced in Ablin and Peyré (2022) turns out to be very close in spirit with the projected gradient flow (2.2) of Yamashita (1980), there is a key difference: the normal term in (1.2) is not a pseudoinverse step, but the unconstrained gradient of a penalty function. However, we show below that both coincide upon a suitable metric choice (proposition 4.4).

The landing algorithm was extended to general equality constraints and stochastic settings by Schechtman et al. (2023), who renamed it as *Orthogonal Directions Constrained Gradient Method*. The authors prove optimal complexity guarantees for the discretized algorithm, namely $O(\epsilon^{-2})$ iterations to reach an $\epsilon$-KKT point. To our knowledge, this is the first convergence proof for the discrete algorithm (1.2); earlier guarantees concerned the continuous system (2.2), assuming that the chosen discretization approximate the ODE trajectories sufficiently well. This analysis was further strengthened in Vary et al. (2024) in stochastic settings, using a smooth merit function to prove global convergence even in the presence of non-tangent random errors.

More recently, Zhang et al. (2024) applied the landing method to accelerate low-rank adaptation in fine-tuning large language models, proving convergence on the Stiefel manifold using only a sufficiently small constant step size. Goyens et al. (2026) derived landing flows on the Stiefel manifold for a family of metrics extending the $\beta$-metric of Hüper et al. (2021); Mataigne et al. (2025). Song et al. (2025) adapted the landing algorithm for distributed stochastic optimization on the Stiefel manifold. Lately, Shi and Wang (2025) proposed an extension called *Adaptive*

*Directional Decomposition Methods* to incorporate inequality constraints. The tangent term is obtained by solving a quadratic program, similarly to the null-space gradient-flow strategy of Feppon et al. (2020a) for equality and inequality constrained problems. They show global convergence guarantees using an adaptive step size rule, which relies on the knowledge of Lipschitz constants.

**Links with null-space projections and SQP**

The use of null-space projections to compute descent directions on the constraint manifold is certainly universal. It is therefore not surprising that formulas analogous to (1.6) and (2.1) have appeared in numerous alternative optimization frameworks, and even in robotics, where null-space control is used to exploit actuation redundancy and allow robots to perform multiple tasks simultaneously (Dietrich et al., 2015).

Most prominently, the decomposition (1.1) plays a central role in so-called *null-space methods*, where it is employed to isolate the objective descent within the null space and to construct reduced-Hessian approximations (Nocedal and Overton, 1985; Yuan, 2001; Nie, 2004; Biros and Ghattas, 2005; Nocedal and Wright, 2006; Berahas et al., 2024). These ideas lead to expressions directly comparable to the tangent and normal components in (2.2). Such decompositions into null-space and range-space steps have become standard in modern SQP implementations, most notably in the SNOPT software, where they are used both to reduce the dimension of the quadratic programming subproblem and to exploit reduced Hessians (Gill et al., 2002; Gill and Wong, 2012; Gill et al., 2015; Gharaei et al., 2023; Fang et al., 2024). They have also been exploited in interior-point methods (Liu and Yuan, 2010; Nocedal et al., 2014). Finally, we note that the least-squares multiplier $\lambda(x) = -(\nabla c(x)\nabla c(x)^\top)^{-1}\nabla c(x)\nabla f(x)$ is likewise employed in certain variants of the Augmented Lagrangian Method (Tapia, 1977; Conn et al., 1991), making these variants equivalent to landing-type algorithms (see proposition 5.2).

Finally, connections between SQP and differential-equation-based methods have also been pointed out repeatedly. For instance, Schropp and Singer (2000) showed that standard SQP schemes can be interpreted as variable step-size Euler discretizations of preconditioned versions of the flow (2.2). Precise relationships between the Riemannian Newton method and the quadratically convergent SQP method were established in Miller and Malick (2005) and Absil et al. (2009), and later exploited in Mishra and Sepulchre (2016) to design metric preconditioners accelerating the convergence of feasible Riemannian optimization methods on matrix manifolds. Recently, an analysis of the rate of convergence of the SQP algorithm with a constant quadratic term was derived by Bai and Mei (2018) through the interpretation as a Riemannian optimization method.

## 3. Riemannian geometry and notation conventions

We consider the equality-constrained optimization problem (P) on the Euclidean spaces $\mathcal{E}$ and $\mathcal{F}$. We denote by $\langle\cdot,\cdot\rangle_{\mathcal{E}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ their respective inner products, and by $\|\cdot\|_{\mathcal{E}}$ and $\|\cdot\|_{\mathcal{F}}$ their associated norms. The feasible set is denoted by $\mathcal{M} := \{x \in \mathcal{E} \colon c(x) = 0\}$. In the applicative section 7 of this paper, $\mathcal{E}$ and $\mathcal{F}$ are thought of as matrix subspaces, but in the other parts of this section, the reader may think of these spaces as $\mathcal{E} \simeq \mathbb{R}^n$ and $\mathcal{F} \simeq \mathbb{R}^m$ through a proper vectorial indexing. 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 $\mathcal{E}$.

Throughout the paper, we denote by $\mathrm{D}c$ the differential of the constraint function $c$: $\mathrm{D}c(x) : \mathcal{E} \to \mathcal{F}$ is the linear map such that

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

We denote by $\mathcal{D}$ the set of points $x \in \mathcal{E}$ at which the constraints satisfy the Linear Independence Constraint Qualification (LICQ) condition:

$$\mathcal{D} = \{x \in \mathcal{E} \colon \operatorname{rank}(\mathrm{D}c(x)) = m\}. \tag{3.1}$$

We recall that the set $\mathcal{M}_x$ defined in (1.5) is a smooth manifold when it is included in $\mathcal{D}$. Every set $\mathcal{M}_x$ is a level curve of the constraint function $c$, called a *layer manifold* (Goyens et al., 2024). The tangent space of $\mathcal{M}_x$ at $x \in \mathcal{D}$ is the null space of the differential of the constraint:

$$\mathrm{T}_x\mathcal{M}_x = \ker(\mathrm{D}c(x)). \tag{3.2}$$

### Riemannian metric and normal space

Let $g$ be a Riemannian metric on $\mathcal{E}$, that is a smooth family of inner products $x \mapsto g_x(\,\cdot\,,\,\cdot\,)$ on $\mathcal{E}$. With a small abuse of notation, we drop the subscript $_x$ when referring to the metric. The metric is represented by a smoothly varying family of operators $G(x) : \mathcal{E} \to \mathcal{E}$, symmetric positive definite with respect to the Euclidean product $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathcal{E}}$, such that

$$g(\xi, \zeta) = \langle G(x)\,\xi, \zeta\rangle_{\mathcal{E}}\,, \qquad \text{for all } \xi, \zeta \in \mathcal{E}. \tag{3.3}$$

The corresponding norm is denoted by $\|u\|_g = \sqrt{g(u,u)}$ for any $u \in \mathcal{E}$. We write occasionally $g^{\mathcal{E}} = \langle\,\cdot\,,\,\cdot\,\rangle_{\mathcal{E}}$ for the Euclidean metric.

The metric $g$ allows to define several important objects for optimization. First, it determines the *normal space* $\mathrm{N}^g_x$ to the manifold $\mathcal{M}_x$ at every $x \in \mathcal{D}$:

$$\mathrm{N}^g_x\mathcal{M}_x := \{v \in \mathcal{E} \colon g(v, \xi) = 0 \text{ for all } \xi \in \mathrm{T}_x\mathcal{M}_x\}\,. \tag{3.4}$$

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

$$g(\xi, v - \mathrm{Proj}_{x,g}(v)) = 0 \quad \text{for all } \xi \in \mathrm{T}_x\mathcal{M}_x \text{ and } v \in \mathcal{E}, \tag{3.5}$$

see e.g., (Absil et al., 2008, (3.37)). In section 3, we will adopt a point of view where the tangent projector is defined before the metric $g$; in that case, we will denote it simply by $\mathrm{Proj}_x$. The orthogonal projector associated to the Euclidean metric $g^{\mathcal{E}}$ will be occasionally denoted by $\Pi_x$.

The metric enables to define gradients and adjoints, whose definitions are recalled in the next paragraphs.

### Riemannian gradients on $\mathcal{E}$ and on $\mathcal{M}_x$

Let $\mathrm{D}f(x) : \mathcal{E} \to \mathbb{R}$ denote the differential of $f$ at $x \in \mathcal{D}$. Through the celebrated Riesz representation 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 denoted by $\nabla_{\mathcal{E}} f(x) \in \mathcal{E}$ such that

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

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

$$g(\nabla_g f(x), \xi) = \mathrm{D}f(x)[\xi] \quad \text{for all } \xi \in \mathrm{T}_x\mathcal{D} \simeq \mathcal{E}. \tag{3.6}$$

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 $x \in \mathcal{D}$, $\mathrm{grad}^g_{\mathcal{M}_x} f(x)$ is the unique vector in $\mathrm{T}_x\mathcal{M}_x$ satisfying

$$g(\mathrm{grad}^g_{\mathcal{M}_x} f(x), \xi) = \mathrm{D}f(x)[\xi], \quad \text{for all } \xi \in \mathrm{T}_x\mathcal{M}_x. \tag{3.7}$$

The unconstrained Euclidean and Riemannian gradients are related by the identity

$$\nabla_g f(x) = G(x)^{-1}\nabla_{\mathcal{E}} f(x). \tag{3.8}$$

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

$$\mathrm{grad}^g_{\mathcal{M}_x} f(x) = \mathrm{Proj}_{x,g}(\nabla_g f(x)). \tag{3.9}$$

### Metric adjoints

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

$$\langle A\xi, y\rangle_{\mathcal{F}} \;=\; g(\xi, A^{*,g}y), \qquad \text{for all } \ \xi \in \mathcal{E}, \ y \in \mathcal{F}. \tag{3.10}$$

We denote by $A^{*,\mathcal{E}}$ the adjoint operator with respect to the Euclidean metric, which satisfies:

$$\langle A\xi, y\rangle_{\mathcal{F}} \;=\; \left\langle \xi, A^{*,\mathcal{E}}y\right\rangle_{\mathcal{E}}, \qquad \text{for all } \ \xi \in \mathcal{E}, \ y \in \mathcal{F}. \tag{3.11}$$

The Riemannian and the Euclidean adjoint operators are related through the identity

$$A^{*,g} = G(x)^{-1}A^{*,\mathcal{E}}. \tag{3.12}$$

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

$$\mathrm{D}c(x) = \begin{pmatrix} \nabla c_1(x)^\top \\ \vdots \\ \nabla c_m(x)^\top \end{pmatrix} \in \mathbb{R}^{m\times n}, \tag{3.13}$$

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

**Metric pseudoinverse and formulas for $g$-orthogonal projectors**

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

$$\mathrm{D}c(x)^{\dagger,g} \;:=\; \mathrm{D}c(x)^{*,g}\,\left(\mathrm{D}c(x)\,\mathrm{D}c(x)^{*,g}\right)^{-1} \;:\; \mathcal{F} \to \mathcal{E}, \tag{3.14}$$

which satisfies $\mathrm{D}c(x)\mathrm{D}c(x)^{\dagger,g} = \mathrm{Id}_{\mathcal{F}}$. The $g$-orthogonal projectors onto $\mathrm{T}_x\mathcal{M}_x$ and $\mathrm{N}_x\mathcal{M}_x$ read then respectively

$$\mathrm{Proj}_{x,g} \;=\; \mathrm{Id}_{\mathcal{E}} - (\mathrm{D}c(x))^{\dagger,g}\,\mathrm{D}c(x), \qquad \mathrm{Proj}^{\perp}_{x,g} \;=\; (\mathrm{D}c(x))^{\dagger,g}\,\mathrm{D}c(x). \tag{3.15}$$

We note that that the $g$-normal space is $\mathrm{N}^g_x\mathcal{M}_x = \mathrm{Range}(\mathrm{D}c(x)^{*,g})$.

## 4. Alternative definitions of the landing algorithm and metric construction via normal bundles.

This section establishes the equivalence between the landing algorithm (1.2) and one closer to the projected gradient flow (2.2). Before this, we observe that there is no gain of generality in considering two possibly different metrics for the tangent and normal terms.

**Lemma 4.1.** *Consider the landing scheme (1.2) with the tangent and 'normal' terms calculated in two different metrics $g_1$ and $g_2$:*

$$d_T(x) = -\mathrm{grad}^{g_1}_{\mathcal{M}_x} f(x), \tag{4.1}$$

$$d_N(x) = -\nabla_{g_2}\psi(x) \ \textit{with} \ \psi(x) = ||c(x)||^2_{\mathcal{F}}/2. \tag{4.2}$$

*There exists a common metric $g$ such that $d_T(x) = -\mathrm{grad}^{g}_{\mathcal{M}_x} f(x)$ and $d_N(x) = -\nabla_g\psi(x)$.*

*Proof.* Consider the metric $g$ given by

$$g(\xi,\zeta) := g_1(\mathrm{Proj}_{x,g_2}\xi, \mathrm{Proj}_{x,g_2}\zeta) + g_2(\mathrm{Proj}^{\perp}_{x,g_2}\xi, \mathrm{Proj}^{\perp}_{x,g_2}\zeta).$$

It is clear that for any $\xi,\zeta \in \mathrm{T}_x\mathcal{M}_x$, $g(\xi,\zeta) = g_1(\xi,\zeta)$, which implies $\mathrm{grad}^{g_1}_{\mathcal{M}_x} f(x) = \mathrm{grad}^{g_2}_{\mathcal{M}_x} f(x)$. On the other hand, $\nabla_{g_2}\Psi(x)$ is the unique vector in $\mathcal{E}$ satisfying $g_2(\nabla_{g_2}\Psi(x),\xi) = \langle \mathrm{D}c(x)\xi, c(x)\rangle_{\mathcal{F}}$ for all $\xi \in \mathcal{E}$. But for any $\xi \in \mathcal{E}$, $g(\nabla_{g_2}\Psi(x),\xi) = g_2(\nabla_{g_2}\Psi(x),\xi) = \langle \mathrm{D}c(x)\xi, c(x)\rangle_{\mathcal{F}}$. This implies $\nabla_{g_2}\psi(x) = \nabla_g\psi(x)$. ◻

We note that the initial paper on the landing algorithm Gao et al. (2022) considered a Riemannian metric for the tangent term and the Euclidean metric for the computation of the normal term. The previous lemma shows that the metric in the normal space can always be redefined to infer both terms from a common metric.

The remainder of the section is structured as follows. We propose in section 4.1 a procedure for defining the ambient metric $g$ by first specifying the orthogonality, in other words, the normal bundle of $\mathcal{M}_x$ for every $x \in \mathcal{E}$. This enables to explicit the dependence of the the tangent and normal terms $d_T(x)$ and $d_N(x)$ to the orthogonal projectors on $\mathrm{T}_x\mathcal{M}_x$ and $\mathrm{N}_x\mathcal{M}_x$ and the restriction of the metric $g$ to these spaces.

Then, we show the equivalence between the landing algorithm (1.2) and a variation of it, where the normal step is replaced with a pseudoinverse or 'Newton'-like step of the form of (2.1), as considered in the projected gradient flow (2.2). More precisely, we show that the pseudoinverse step $d_N(x)$ depends only on the choice of the normal space and not on the restriction of the normal metric. Then, we prove that the normal metric can always be redefined so that the pseudoinverse step becomes the Riemannian gradient of the penalty functional $\psi(x) = ||c(x)||^2/2$ as in (1.4).

### 4.1. Construction of the ambient metric $g$ from a given normal bundle

In the definitions (1.3) and (1.4), the metric $g$ is defined *a priori*, which enables to compute the tangent and the normal terms from the formulas

$$d_T(x) = -\mathrm{Proj}_{x,g}(G(x)^{-1}\nabla_{\mathcal{E}} f(x)), \tag{4.3}$$

$$d_N(x) = -G(x)^{-1}\mathrm{D}c(x)^{*,\mathcal{E}}c(x). \tag{4.4}$$

The first formula follows from (3.8), (3.9) and (3.12), while the second can be inferred from the identities (3.6) and (3.10):

$$g(\nabla_g\psi(x),\xi) = \mathrm{D}\psi(x)[\xi] = \langle \mathrm{D}c(x)\xi, c(x)\rangle_{\mathcal{F}} = g(\mathrm{D}c(x)^{*,g}c(x),\xi) = g(G(x)^{-1}\mathrm{D}c(x)^{*,\mathcal{E}}c(x),\xi) \qquad \forall \xi \in \mathcal{E}.$$

An issue with formulas (4.3) and (4.4), in the context of Riemannian matrix optimization, is that closed-form expressions for that the projection mapping $\mathrm{Proj}_{x,g}$ or the inverse metric $G(x)^{-1}$ are not necessarily available.

Alternatively, the metric $g$ can be constructed by *first* considering a smooth mapping of (oblique) linear tangent projectors

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

that is, for every $x \in \mathcal{D}$, $\mathrm{Proj}_x$ is a given linear mapping on $\mathcal{E}$ satisfying

$$\mathrm{Proj}_x\mathrm{Proj}_x = \mathrm{Proj}_x, \quad \mathrm{Range}(\mathrm{Proj}_x) = \mathrm{T}_x\mathcal{M}_x \quad \text{ and } \mathrm{Proj}_x = \mathrm{Id} \text{ on } \mathrm{T}_x\mathcal{M}_x.$$

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, 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

$$\begin{aligned} \langle \xi, G_T(x)\xi\rangle_{\mathcal{E}} &\geq g_T||\xi||^2_{\mathcal{E}}, \qquad &\forall \xi \in \mathrm{T}_x\mathcal{M}_x,\\ \langle \xi, G_N(x)\xi\rangle_{\mathcal{E}} &\geq g_N||\xi||^2_{\mathcal{E}}, \qquad &\forall \xi \in \mathrm{N}_x\mathcal{M}_x. \end{aligned} \tag{4.5}$$

The ambient metric $g$ can finally be defined as

$$g(\xi,\zeta) = \langle \xi, G(x)\zeta\rangle\rangle = \langle \mathrm{Proj}_x\xi, G_T(x)\mathrm{Proj}_x\zeta\rangle_{\mathcal{E}} + \langle \mathrm{Proj}_x^{\perp}\xi, G_N(x)\mathrm{Proj}_x^{\perp}\zeta\rangle_{\mathcal{E}}, \tag{4.6}$$

denoting by $\mathrm{Proj}_x^{\perp} := \mathrm{Id}_{\mathcal{E}} - \mathrm{Proj}_x$ the oblique projector on $\mathrm{N}_x\mathcal{M}_x$. In other words, $g(\xi,\zeta) = \langle \xi, G(x)\zeta\rangle_{\mathcal{E}}$ where $G(x) : \mathcal{E} \to \mathcal{E}$ is the block-diagonal operator

$$G(x) = \mathrm{Proj}_x^{*,\mathcal{E}} G_T(x)\mathrm{Proj}_x + (\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}} G_N(x)\mathrm{Proj}_x^{\perp}. \tag{4.7}$$

Then, it is clear that:

(i) $G_T$ and $G_N$ are representers of the restriction of the metric to the tangent space $\mathrm{T}_x\mathcal{M}_x$ and the normal space $\mathrm{N}_x\mathcal{M}_x$;
(ii) $\mathrm{T}_x\mathcal{M}_x$ and $\mathrm{N}_x\mathcal{M}_x$ are $g$-orthogonal subspaces;
(iii) $\mathrm{Proj}_x = \mathrm{Proj}_{x,g}$ 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

$$\begin{aligned}\widetilde{G_T} &:= \mathrm{Proj}_x^{*,\mathcal{E}} G_T(x)\mathrm{Proj}_x \quad &: \mathrm{T}_x\mathcal{M}_x \to (\mathrm{N}_x\mathcal{M}_x)^{\perp,\mathcal{E}},\\ \widetilde{G_N} &:= (\mathrm{Proj}_x^\perp)^{*,\mathcal{E}} G_N(x)\mathrm{Proj}_x^\perp \quad &: \mathrm{N}_x\mathcal{M}_x \to (\mathrm{T}_x\mathcal{M}_x)^{\perp,\mathcal{E}},\end{aligned} \tag{4.8}$$

are invertible due to (4.5). Denoting by $\widetilde{G_T}(x)^{-1}$ and $\widetilde{G_N}(x)^{-1}$ their inverses, the inverse of $G(x)$ reads

$$G(x)^{-1} = \mathrm{Proj}_x\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}} + \mathrm{Proj}_x^\perp\widetilde{G_N}(x)^{-1}(\mathrm{Proj}_x^\perp)^{*,\mathcal{E}}. \tag{4.9}$$

**Remark 4.1.** *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 $G_T(x)$ and $G_N(x)$ as operators on the whole set $\mathcal{E}$ is motivated by the need to define a differentiable metric field through (4.7).*

**Remark 4.2.** *This approach, whereby the differential structure of a manifold is specified by a mapping of projectors, has been considered to analyze the geometric structure of certain algebraic matrix decompositions, see Feppon and Lermusiaux (2019).*

We can now express the tangent and normal steps (1.3) and (1.4) in terms of $\mathrm{Proj}_x$, $\widetilde{G_T}(x)^{-1}$ and $\widetilde{G_N}(x)^{-1}$.

**Proposition 4.2.** *The tangent and normal space steps of (1.3) and (1.4) can be rewritten in terms of the projectors, the tangent and the normal metric $G_T(x)$ and $G_N(x)$ as*

$$d_T(x) = -\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}\nabla_{\mathcal{E}}f(x), \tag{4.10}$$

$$d_N(x) = -\widetilde{G_N}(x)^{-1}\mathrm{D}c(x)^{*,\mathcal{E}}c(x). \tag{4.11}$$

*Proof.* By (3.8), (3.9) and (4.9),

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

This proves (4.10). For (4.11), we write

$$\begin{aligned}d_N(x) &= -\mathrm{D}c^{*,g}(x)c(x) = -G(x)^{-1}\mathrm{D}c^{*,\mathcal{E}}(x)c(x) = -G(x)^{-1}(\mathrm{Proj}_x^\perp)^{*,\mathcal{E}}\mathrm{D}c^{*,\mathcal{E}}(x)c(x)\\ &= -\widetilde{G_N}(x)^{-1}\mathrm{D}c(x)^{*,\mathcal{E}}c(x).\end{aligned}$$

□

**Remark 4.3.** *Although formula (4.10) suggests that $d_T(x)$ could depend on the choice of the normal space through the projector $\mathrm{Proj}_x^{*,\mathcal{E}}$, this is not the case. This is visible in the definition (3.7) of the constrained Riemannian gradient. The operator $\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}$ depends thus solely on the restriction of the metric $g$ to the tangent space $\mathrm{T}_x\mathcal{M}_x$, but not on the particular choice of normal space. In particular, the formula*

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

*does not depend on the choice of projection projector* $\mathrm{Proj}_x$ *so long as* $\mathrm{Range}(\mathrm{Proj}_x) = \mathrm{T}_x\mathcal{M}_x$*. In fact, one may verify that*

$$\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}\xi = \sum_{i=1}^{\dim(\mathrm{T}_x\mathcal{M}_x)} \langle \xi, u_i\rangle_{\mathcal{E}} u_i,$$

*where* $(u_i)_{1\leq i\leq \dim(\mathrm{T}_x\mathcal{M}_x)}$ *is any* $g$*-orthonormal basis of* $\mathrm{T}_x\mathcal{M}_x$*. However, (4.10) and (4.11) show that closed-form expressions are available for* $d_T(x)$ *and* $d_N(x)$ *if this is the case for (i) the projection operator* $\mathrm{Proj}_x$ *and (ii) the inverse of the restriction operators* $\widetilde{G_T}(x)$ *and* $\widetilde{G_N}(x)$*.*

### 4.2. **Equivalence between landing algorithms and projected gradient flows**

Given an arbitrary smooth symmetric positive-definite operator field

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

we consider the iterative scheme

$$x_{k+1} = x_k - \alpha_k(d_T(x_k) + d_N(x_k)), \tag{4.12}$$

where the tangent term $d_T(x) = -\mathrm{grad}^g_{\mathcal{M}_x} f(x) \in \mathrm{T}_x\mathcal{M}_x$ is unchanged while the normal vector field $d_N(x)$ is defined for $x \in \mathcal{D}$ by

$$d_N(x) := -\,\mathrm{D}c(x)^{\dagger,g}\, H(x)\, c(x) \in \mathrm{N}^g_x\mathcal{M}_x. \tag{4.13}$$

We refer to (4.13) as a 'pseudoinverse' step. The iterative algorithm (4.12) may be interpreted as the discretization of the ordinary differential equation

$$\begin{aligned}\dot{x} &= -d_T(x) - d_N(x)\\ &= (\mathrm{Id}_{\mathcal{E}} - \mathrm{D}c(x)^{*,g}\,(\mathrm{D}c(x)\,\mathrm{D}c(x)^{*,g})^{-1}\,\mathrm{D}c(x))\nabla_g f(x)\\ &\quad -\,\mathrm{D}c(x)^{*,g}\,(\mathrm{D}c(x)\mathrm{D}c(x)^{*,g})^{-1}\,H(x)\,c(x),\end{aligned} \tag{4.14}$$

When $H(x) = \mathrm{Id}$ and $g = g^{\mathcal{E}}$ is the Euclidean metric, (4.14) coincides with the 'Newton-like' or 'pseudoinverse' step of the projected gradient flow (2.2).

Formula (4.13) offers a somewhat more explicit control of how optimization trajectories land on the constraint manifold than (1.4) based on the unconstrained Riemannian gradient. It is a descent direction for the infeasibility measure $\psi(x) = ||c(x)||^2_{\mathcal{F}}/2$ with a descent value depending explicitly on $H(x)$ and independent of $g$:

$$\begin{aligned}\mathrm{D}\psi(x)[d_N(x)] &= \langle \mathrm{D}c(x)[d_N(x)], c(x)\rangle_{\mathcal{F}}\\ &= -\left\langle \mathrm{D}c(x)\mathrm{D}c(x)^{*,g}\,(\mathrm{D}c(x)\mathrm{D}c(x)^{*,g})^{-1}\,H(x)\,c(x), c(x)\right\rangle_{\mathcal{F}}\\ &= -\,\langle H(x)c(x), c(x)\rangle_{\mathcal{F}} < 0.\end{aligned} \tag{4.15}$$

More precisely, the normal step $d_N(x)$ is the minimum $g$-norm solution of the undetermined system $\mathrm{D}c(x)[d_N] = -H(x)c(x)$:

$$d_N(x) = \arg\min_{d_N\in\mathcal{E}} \frac{1}{2}\,\|d_N\|^2_g \;\text{ such that } \mathrm{D}c(x)[d_N] = -H(x)c(x). \tag{4.16}$$

This implies that, at the continuous level, the constraint values along the gradient flow trajectories $x(t)$ solving (4.14) satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}c(x(t)) = -H(x(t))c(x(t)),$$

which entails $c(x(t)) = c(x(0))\exp\left(-\int_0^t H(x(s))\mathrm{d}s\right)$. Hence, the eigenvalues and the eigenvectors of the symmetric operator $H(x)$ determine the instantaneous decay rate of the components of the constraint violation vector $c(x) \in \mathcal{F}$.

When $H(x) = \mathrm{D}c(x)\,\mathrm{D}c(x)^{*,g}$, we retrieve $d_N(x) = -\mathrm{D}c(x)^{*,g}c(x) = -\nabla_g\psi(x)$, the negative unconstrained Riemannian gradient of the infeasibility measure $\psi(x) = ||c(x)||^2_{\mathcal{F}}/2$. The following propositions show that the converse result is true: *firstly*, (4.13) does not depend on the choice of the normal metric $G_N(x)$; *secondly*, it is actually possible to redefine the normal metric $G_N(x)$ *without changing the values of* $d_T(x)$ *and* $d_N(x)$ so that $d_N(x)$ becomes the negative Riemannian unconstrained gradient of $\psi$ in the new metric.

**Proposition 4.3.** *The 'pseudoinverse' normal space step (4.13) depends only on the choice of orthogonal normal space* $\mathrm{N}_x\mathcal{M}_x$ *through the normal projector* $\mathrm{Proj}_x^\perp$*. More precisely, it holds*

$$d_N(x) = -\mathrm{Proj}_x^\perp\mathrm{D}c(x)^{\dagger,\mathcal{E}}H(x)c(x). \tag{4.17}$$

*Proof.* Recall that $\mathrm{Proj}_x^\perp = \mathrm{D}c(x)^{\dagger,g}\mathrm{D}c(x)$ (eq. (3.15)). 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 identity $\mathrm{D}c(x)^{\dagger,g} = \mathrm{Proj}_x^\perp\mathrm{D}c(x)^{\dagger,\mathcal{E}}$ relating the $g-$ and Euclidean right-pseudoinverses. Formula (4.17) follows by substituting into (4.13). □

**Proposition 4.4.** *The pseudoinverse normal step (4.13) (or (4.17)) can be rewritten as*

$$d_N(x) = -\nabla_g\psi(x), \tag{4.18}$$

*with* $\psi(x) = ||c(x)||^2_{\mathcal{F}}/2$ *for the metric field* $g$ *defined through (3.3) and (4.7) 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 for the Euclidean inner product. 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 (4.6) (leaving $G_T(x)$ unchanged) does define a metric $g$ on $\mathcal{E}$.
Let us recall (equation (4.11)) that the unconstrained Riemannian gradient of $\psi(x)$ is given by

$$\nabla_g\psi(x) = \widetilde{G_N}(x)^{-1}\mathrm{D}c(x)^{*,\mathcal{E}}c(x). \tag{4.19}$$

Let us express $\nabla_g\psi(x)$ in pseudoinverse form. Due to $\ker(\mathrm{D}c(x)) = \mathrm{T}_x\mathcal{M}_x$ and $\mathrm{Range}(\mathrm{D}c(x)^{*,\mathcal{E}}) = \mathrm{T}_x\mathcal{M}_x^{\perp,\mathcal{E}}$, we have

$$\begin{aligned} \mathrm{D}c(x)\mathrm{Proj}_x^\perp &= \mathrm{D}c(x), \qquad & \mathrm{Proj}_x^{*,\mathcal{E}}\mathrm{D}c(x)^{*,\mathcal{E}} &= 0 \\ \mathrm{D}c(x)\mathrm{Proj}_x &= 0, \qquad & (\mathrm{Proj}_x^\perp)^{*,\mathcal{E}}\mathrm{D}c(x)^{*,\mathcal{E}} &= \mathrm{D}c(x)^{*,\mathcal{E}}. \end{aligned} \tag{4.20}$$

These identities imply

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

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

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

invoking the first line of (4.20). 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 because $\mathrm{Range}(\mathrm{D}c(x)^{*,\mathcal{E}}) = \mathrm{T}_x\mathcal{M}_x^{\perp,\mathcal{E}}$, and right multiplying by $H(x)$, we obtain

$$\widetilde{G_N}(x)^{-1}\,\mathrm{D}c(x)^{*,\mathcal{E}} = \mathrm{Proj}_x^\perp\mathrm{D}c(x)^{\dagger,\mathcal{E}}H(x). \tag{4.21}$$

Combining (4.19) and (4.21) yields (4.18). □

In other words, defining the normal term as a negative unconstrained Riemannian gradient (1.4) or as the pseudo inverse step (4.13) leads to the same family of optimization algorithms.

## 5. Links between landing methods and existing constrained optimization algorithms

In this section, we establish explicit connections between the landing algorithm and several well-known constrained optimization algorithms from the literature. Specifically, in section 5.1, we demonstrate that the landing algorithm, as defined by (1.2), can be interpreted as an Augmented Lagrangian Method with the Lagrange multiplier assigned the least-squares multiplier value at every iteration.

In section 5.2, we prove that the most basic version of Sequential Quadratic Programming (SQP), without trust-regions, is a particular case of the landing algorithm.

In section 5.3, we deepen the connection—previously observed by multiple authors, e.g. Miller and Malick (2005); Absil et al. (2009); Mishra and Sepulchre (2016)—between the quadratically convergent version of SQP and the Riemannian-Newton method on a manifold. We show that when the tangent step in (1.2) is chosen to be the Riemannian-Newton direction, the landing method with a "Newton-like" normal update $d_N(x) = -\mathrm{D}c(x)^{\dagger,g}c(x)$ achieves quadratic convergence only for a distinguished choice of normal space, which precisely coincides with the one mandated by the SQP framework.

### 5.1. Interpretation of the landing algorithm as an augmented Lagrangian method

Consider the iterative scheme

$$x_{k+1} = x_k - \alpha_k \nabla_g \mathcal{L}_{\beta_k}(x_k, \lambda_k), \tag{5.1}$$

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

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

Here, $\beta_k$ is a penalty parameter which is not necessarily constant. In (5.1), the unconstrained Riemannian gradient $\nabla_g \mathcal{L}_{\beta_k}$ is taken with respect to the variable $x$.

**Proposition 5.1.** *The Augmented Lagrangian iteration (5.1) coincides with the Riemannian landing iteration (4.12) with pseudoinverse step (4.13), setting $\lambda_k$ as the least-squares multiplier*

$$\lambda_k := \arg\min_{\lambda\in\mathcal{F}} \|\nabla_g f(x_k) - \mathrm{D}c(x_k)^{*,g}\lambda\|_{\mathcal{F}}^2 = (\mathrm{D}c(x_k)\mathrm{D}c(x_k)^{*,g})^{-1}\mathrm{D}c(x_k)\nabla_g f(x_k). \tag{5.2}$$

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

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

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

Using (3.9) and (4.13) with $H(x) = \beta_k \mathrm{D}c(x)\mathrm{D}c(x)^{*,g}$, we obtain $\nabla_g \mathcal{L}_{\beta_k}(x_k,\lambda_k) = \mathrm{grad}^g_{\mathcal{M}_x} f(x_k) + d_N(x_k)$. □

Thus, the landing algorithm is an augmented Lagrangian method with the Lagrange multipliers $\lambda_k$ being the least-squares multipliers, and for which the minimization of the augmented Lagrangian subproblem consists in a single gradient step.

### 5.2. Sequential Quadratic Programming is a particular case of the landing algorithm

In its most basic form as described in Boggs and Tolle (1995), the Sequential Quadratic Programming (SQP) method considers the iterative sequence

$$x_{k+1} = x_k + \alpha_k d_k, \tag{5.3}$$

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

$$\begin{aligned} d_k := \arg\min_{d\in\mathcal{E}} \quad & f(x_k) + \langle d, \nabla_{\mathcal{E}} f(x_k)\rangle_{\mathcal{E}} + \frac{1}{2}\langle d, B_k d\rangle_{\mathcal{E}} \\ \text{subject to} \quad & \mathrm{D}c(x_k)d + c(x_k) = 0,\end{aligned} \tag{5.4}$$

given a sequence of symmetric positive definite operators $B_k : \mathcal{E} \to \mathcal{E}$. This method is locally convergent under mild assumptions on the selection of symmetric positive-definite matrices $B_k$. Setting $B_k = \nabla^2_{\mathcal{E}} \mathcal{L}(x, \lambda_k)$ ($\nabla^2_{\mathcal{E}}$ being the Euclidean Hessian in the $x$ variable), with the Lagrangian

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

SQP achieves quadratic convergence around a KKT point for a unit step size when $\lambda_k \in \mathcal{F}$ is set to the Lagrange multiplier associated to the equality constraint of (5.4) of the previous iteration (see Nocedal and Wright (2006)). Globalization procedures also exist, based on merit functions or filters (Boggs and Tolle, 1995; Fletcher et al., 2002; Wächter and Biegler, 2005; Obara et al., 2022).

The following proposition shows that the SQP algorithm (5.3) forms a particular instance of the Riemannian landing method (1.2).

**Proposition 5.2.** *Suppose that $B_k$ is a symmetric positive definite operator on $\mathrm{T}_{x_k}\mathcal{M}_{x_k}$. Consider the space $\mathrm{N}_{x_k}\mathcal{M}_{x_k}$ defined as the $\mathcal{E}$-orthogonal space of $B_k \mathrm{T}_{x_k}\mathcal{M}_{x_k}$:*

$$\mathrm{N}_{x_k}\mathcal{M}_{x_k} := (B_k \mathrm{T}_{x_k}\mathcal{M}_{x_k})^{\perp,\mathcal{E}}. \tag{5.5}$$

*Let $g$ be the metric defined by*

$$g(\xi, \zeta) := \langle \mathrm{Proj}_{x_k}\xi, B_k \mathrm{Proj}_{x_k}\zeta \rangle_{\mathcal{E}} + \langle \mathrm{Proj}^{\perp}_{x_k}\xi, G_N(x_k)\mathrm{Proj}^{\perp}_{x_k}\zeta \rangle_{\mathcal{E}},$$

*where $\mathrm{Proj}_{x_k}$ is the oblique projection on the decomposition $\mathcal{E} = \mathrm{T}_{x_k}\mathcal{M}_{x_k} \oplus \mathrm{N}_{x_k}\mathcal{M}_{x_k}$ and $G_N(x_k)$ any positive symmetric definite operator on $\mathrm{N}_{x_k}\mathcal{M}_{x_k}$. Then, if $x_k \in \mathcal{D}$, the SQP direction $d_k$ of (5.4) is given by*

$$d_k = -\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k) - \mathrm{D}c(x_k)^{\dagger,g} c(x_k). \tag{5.6}$$

*In other words, the SQP direction $d_k$ and the Riemannian landing direction $d_T(x_k) + d_N(x_k)$ based on the pseudoinverse formula with the choice $H(x_k) = \mathrm{Id}_{\mathcal{F}}$ for the normal component (eq. (1.3) and (4.13)) coincide at every step $k$ for the metric induced by $B_k$ on the tangent space $\mathrm{T}_{x_k}\mathcal{M}_{x_k}$ and the normal space (5.5).*

*Proof.* The space decomposition $\mathcal{E} = \mathrm{T}_{x_k}\mathcal{M}_{x_k} \oplus \mathrm{N}_{x_k}\mathcal{M}_{x_k}$ holds because $B_k$ is symmetric definite positive. Observe that any $d$ satisfying $\mathrm{D}c(x_k)d + c(x_k)$ can be written as

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

with $d_N(x_k) = -\mathrm{D}c(x_k)^{\dagger,g} c(x_k)$ and $u \in \mathrm{T}_{x_k}\mathcal{M}_{x_k}$. The minimizer of (5.4) is $d_k = u_k + d_N(x_k)$, where $u_k$ is the solution to the quadratic unconstrained minimization problem

$$u_k = \arg \min_{u \in \mathrm{T}_{x_k}\mathcal{M}_{x_k}} f(x_k) + \langle u + d_N(x_k), \nabla_{\mathcal{E}} f(x_k) \rangle_{\mathcal{E}} + \frac{1}{2} \langle u + d_N(x_k), B_k(u + d_N(x_k)) \rangle_{\mathcal{E}}. \tag{5.7}$$

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

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

and since $d_N(x_k) \in \mathrm{N}_{x_k}\mathcal{M}_{x_k} = (B_k \mathrm{T}_{x_k}\mathcal{M}_{x_k})^{\perp,\mathcal{E}}$,

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

where we denote $G(x_k) := B_k$. Consequently, by eliminating constant terms, the solution $u_k$ to (5.7) is also the minimizer of the quadratic problem

$$u_k = \arg \min_{u \in \mathrm{T}_{x_k}\mathcal{M}_{x_k}} \langle u, \mathrm{Proj}^{*,\mathcal{E}}_{x_k} \nabla_{\mathcal{E}} f(x_k) \rangle_{\mathcal{E}} + \frac{1}{2} \langle u, \widetilde{G_T}(x_k) u \rangle_{\mathcal{E}}, \tag{5.8}$$

where $\widetilde{G_T}(x_k) : \mathrm{T}_{x_k}\mathcal{M}_{x_k} \to \mathrm{N}^{\perp,\mathcal{E}}_{x_k}$ is the operator defined in (4.8). The solution to this quadratic program is $u_k = -\widetilde{G_T}(x_k)^{-1} \mathrm{Proj}^{*,\mathcal{E}}_{x_k} \nabla_{\mathcal{E}} f(x_k)$, which is exactly $d_T(x_k)$ according to (4.10). $\square$

**Remark 5.1.** *An alternative, more algebraic proof can be obtained if one assumes $B_k$ to be symmetric positive definite on the whole space $\mathcal{E}$ rather than only on $\mathrm{T}_{x_k}\mathcal{M}_{x_k}$. In this case, the result is well-known and can be found stated in different forms in the optimization literature; see for instance (Boggs and Tolle, 1995, eq. (3.8)-(3.10)) or more recently (Feppon, 2024, Proposition 1). The idea of the proof is recalled here to highlight the link with the landing method (4.12). The KKT condition for (5.4) states that there exists a Lagrange multiplier $\lambda_k \in \mathcal{F}$ such that the minimizer $d_k$ reads*

$$d_k = -B_k^{-1}\nabla_{\mathcal{E}} f(x_k) + B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}}\lambda_k. \tag{5.9}$$

*Inserting this expression into $\mathrm{D}c(x_k)d_k = -c(x_k)$, we infer that*

$$\mathrm{D}c(x_k)d_k = -c(x_k) = -\mathrm{D}c(x_k)B_k^{-1}\nabla_{\mathcal{E}} f(x_k) + \mathrm{D}c(x_k)B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}}\lambda_k.$$

*Since $\mathrm{D}c(x_k)$ is full rank and $B_k^{-1}$ is assumed to be invertible, the operator $\mathrm{D}c(x_k)B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}}$ is symmetric positive definite and $\lambda_k$ is given by*

$$\lambda_k = -(\mathrm{D}c(x_k)B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}})^{-1}c(x_k) + (\mathrm{D}c(x_k)B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}})^{-1}\mathrm{D}c(x_k)B_k^{-1}\nabla f(x_k).$$

*Substituting this expression into (5.9) yields*

$$\begin{aligned} d_k = &-\left(\mathrm{I}_n - B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}}\left(\mathrm{D}c(x_k)B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}}\right)^{-1}\mathrm{D}c(x_k)\right)B_k^{-1}\nabla_{\mathcal{E}} f(x_k) \\ &- B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}}\left(\mathrm{D}c(x_k)B_k^{-1}\mathrm{D}c(x_k)^{*,\mathcal{E}}\right)^{-1}c(x_k), \end{aligned} \tag{5.10}$$

*which is (5.6) with the metric $g(\xi,\zeta) := \langle \xi, B_k\zeta\rangle$.*

Conversely, given a metric $g$ at $x_k$, the proof of proposition 5.2 shows that the "landing" step (5.6) can be obtained as the SQP step $d_k$ solving (5.4) by choosing

$$B_k := \mathrm{Proj}_{x_k}^{*,\mathcal{E}}\, G_T(x_k)\,\mathrm{Proj}_{x_k}.$$

Hence, the SQP algorithm (5.3) and (5.4) constitutes a strict subclass of landing algorithms (1.2) to (1.4). In SQP, the iterate is completely determined by the choice of the normal space and the tangent metric, whereas the landing framework additionally allows one to tune the normal metric for the normal step (1.4)—or, equivalently, to choose the operator $H$ in (4.13).

Lastly, it is interesting to notice that, given a matrix $B_k$ generating the SQP direction $d_k$ through (5.4), the normal component can be modified without changing the tangent component by replacing $B_k$ with $\mathrm{Proj}_{x_k}^{*,\mathcal{E}}B_k\mathrm{Proj}_{x_k}$, where $\mathrm{Proj}_{x_k}$ is any linear projector on $\mathrm{T}_{x_k}\mathcal{M}_{x_k}$. After this change, the SQP direction reads (5.6) where $g$ is a metric making $\mathrm{T}_{x_k}\mathcal{M}_{x_k}$ and $\mathrm{N}_{x_k}\mathcal{M}_{x_k} := \ker(\mathrm{Proj}_{x_k})$ $g$-orthogonal and satisfying $g(\xi,\zeta) = \langle \xi, B_k\zeta\rangle_{\mathcal{E}}$ for any $\xi,\zeta \in \mathrm{T}_{x_k}\mathcal{M}_{x_k}$.

### 5.3. Link with the Riemannian Newton method: all roads lead to SQP

We have recalled earlier that the SQP method (5.4) is quadratically convergent with $B_k = \nabla^2_{\mathcal{E}}\mathcal{L}(x_k,\lambda_k)$ and $\lambda_k$ being the Lagrange multiplier of the quadratic program at the previous iteration. In fact, these settings make SQP equivalent to the Newton method for the functional $F(x,\lambda) := (\nabla_{\mathcal{E}}\mathcal{L}(x,\lambda), c(x))$ (Nocedal and Wright, 2006, Chapter 18). In view of proposition 5.2, this implies that in order to make the landing method

$$x_{k+1} = x_k - \mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k) - \mathrm{D}c(x_k)^{\dagger,g}c(x_k) \tag{5.11}$$

locally quadratically convergent, it is sufficient to choose a metric $g$ such that

(i) the normal space at $x_k$ is $\mathrm{N}_{x_k}\mathcal{M}_{x_k} := (\nabla^2_{\mathcal{E}}\mathcal{L}(x_k,\lambda_k)\mathrm{T}_{x_k}\mathcal{M}_{x_k})^{\perp,\mathcal{E}}$;
(ii) the tangent metric is $G_T(x_k) = \nabla^2_{\mathcal{E}}\mathcal{L}(x_k,\lambda_k)$.

On the other hand, it is known since Edelman et al. (1998); Absil et al. (2009) that when setting $G_T(x) = \nabla^2_{\mathcal{E}}\mathcal{L}(x,\lambda^*(x))$ and $\mathrm{Proj}_x = \Pi_x$ the $\mathcal{E}-$orthogonal projection, the tangent term becomes a Riemannian Newton step:

$$-\mathrm{grad}^g_{\mathcal{M}_x} f(x) = -\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_x} f(x)^{-1}\mathrm{grad}^{g^{\mathcal{E}}}_{\mathcal{M}_x} f(x).$$

Here, $\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_x} f(x) = \Pi_x \nabla^2_{\mathcal{E}} \mathcal{L}(x, \lambda(x))\Pi_x$ is the Riemannian Hessian with respect to the Riemannian connection associated to the Euclidean metric; this follows from the fact that $\xi := \mathrm{grad}^g_{\mathcal{M}_x} f(x)$ satisfies

$$\widetilde{G_T}(x)\xi = \Pi_x \nabla^2_{\mathcal{E}} \mathcal{L}(x, \lambda(x))\Pi_x \xi = -\Pi_x \nabla_{\mathcal{E}} f(x) = -\mathrm{grad}^{g^{\mathcal{E}}}_{\mathcal{M}_x} f(x).$$

In Absil et al. (2009), this property enabled to interpret the Feasibly Projected SQP (FP-SQP) method, i.e.

$$x_{k+1} = R_{x_k}(\alpha_k d_k), \tag{5.12}$$

with $d_k$ being the SQP direction (5.4) and $B_k = \nabla^2_{\mathcal{E}} \mathcal{L}(x_k, \lambda^*_k(x_k))$, as a Riemannian Newton method on the manifold $\mathcal{M}$. Here, the multiplier is not the SQP multiplier $\lambda_k$ as above but the least-squares multiplier $\lambda^*(x_k)$ where

$$\lambda^*(x) = (\mathrm{D}c(x)\mathrm{D}c(x)^{*,\mathcal{E}})^{-1}\mathrm{D}c(x)\nabla_{\mathcal{E}} f(x),$$

however, both choices are asymptotically equivalent since $\lambda_k \to \lambda^*(x^*)$ as $k \to +\infty$.

The FP-SQP method (5.12) combines the SQP step $d_k$ with $R_{x_k}$ a retraction on $\mathcal{M}$ to ensure the feasibility of the iterates at every iteration.

In what follows, we deepen this connection, by showing that the conditions (i) and (ii) are in some sense necessary to make (5.11) locally quadratically convergent. Consider the iterative scheme

$$x_{k+1} = x_k - \mathrm{Hess}^{g,A}_{\mathcal{M}_{x_k}} f(x_k)^{-1} \mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k) - \mathrm{D}c(x_k)^{\dagger,g} c(x_k), \tag{5.13}$$

which combines a Riemannian-Newton tangent step $-\mathrm{Hess}^{g,A}_{\mathcal{M}_x} f(x_k)^{-1}\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)$ on the manifold $\mathcal{M}_{x_k}$ with the 'Newton-like' pseudoinverse normal step $-\mathrm{D}c(x_k)^{\dagger,g}c(x_k)$ aiming to reduce the constraint violation. The scheme (5.13) can be seen as a particular case of the landing method (4.12), up to redefining the metric in the tangent space.

Let us recall that the Riemannian Hessian is defined with respect to the metric $g$ and the choice of an affine connection $\nabla^A$ by the formula

$$\mathrm{Hess}^{g,A}_{\mathcal{M}_x} f(x)\xi := \nabla^A_\xi \mathrm{grad}^g_{\mathcal{M}_x} f(x),$$

where an affine connection $\nabla^A$ is a bilinear operator on the tangent bundle that verifies Leibniz rule ($\nabla^A_\xi \zeta \in \mathrm{T}_x\mathcal{M}_x$ and $\nabla^A_\xi(f\zeta) = \mathrm{D}_\xi f \zeta + f\nabla^A_\xi \zeta$ for any smooth real function $f$ and any $\xi, \zeta \in \mathrm{T}_x\mathcal{M}_x$), see Absil et al. (2009), chapter 5. The superscript $\cdot^A$ in the notation $\nabla^A$ of the affine connection is suggested by the next lemma, which states that affine connections on a Riemannian submanifold can be parameterized by the choice of a tangent bilinear term $A(\xi, \zeta)$.

**Lemma 5.3.** *Any affine connection on $\mathcal{M}_x$ can be written as*

$$\nabla_\xi \eta \equiv \nabla^A_\xi \eta := \mathrm{Proj}_x(\mathrm{D}_\xi \eta) + A(\xi, \eta), \tag{5.14}$$

*where $A(\xi, \eta) : \mathrm{T}_x\mathcal{M}_x \times \mathrm{T}_x\mathcal{M}_x \to \mathrm{T}_x\mathcal{M}_x$ is a vectorial bilinear form with values on the tangent space $\mathrm{T}_x\mathcal{M}_x$, and $\mathrm{Proj}_x$ is any projection operator on $\mathrm{T}_x\mathcal{M}_x$. Reciprocally, (5.14) defines an affine connection.*

*Proof.* According to Gauss formula, any affine connection can be written as

$$\nabla_\xi \eta = \mathrm{d}_\xi \eta - \Gamma(\xi, \eta), \tag{5.15}$$

where $\Gamma(\xi, \eta)$ is the Christoffel symbol of the connection. Since $\nabla_\xi \eta \in \mathrm{T}_x\mathcal{M}_x$, we must have

$$\mathrm{Proj}^\perp_x(\Gamma(\xi, \eta)) = \mathrm{Proj}^\perp_x(\mathrm{D}_\xi \eta),$$

so that $\Gamma(\xi, \eta) = \mathrm{Proj}^\perp_x(\mathrm{D}_\xi \eta) - A(\xi, \eta)$ with $A(\xi, \eta) := -\mathrm{Proj}_x(\Gamma(\xi, \eta))$. Substituting this expression back into (5.15) yields (5.14). Conversely, we verify that for a given vector bilinear form $A(\xi, \eta)$, formula (5.14) satisfies all the axioms of an affine connection. ∎

Recall that on the one hand, the scheme $x_{k+1} = x_k - \mathrm{D}c(x_k)^{\dagger,g}c(x_k)$ generates iterates that approach the feasible set at a quadratic rate, and on the other hand, the Riemannian-Newton method (5.12) with $d_k = -\mathrm{Hess}^{A,g}_{\mathcal{M}_{x_k}}\mathrm{grad}^g_{\mathcal{M}_{x_k}}$ is locally quadratically convergent to a critical point; these properties hold for any choice of metric $g$ and any choice of affine connection $\nabla^A$ (Absil et al., 2008). It could therefore be tempting to think that the scheme (5.13) combining this two directions should be automatically locally quadratically convergent near critical points of $f$ on the manifold $\mathcal{M}$.

We show in the following that this is not true: the local quadratic convergence is obtained only by selecting a metric $g$ such that the normal space at $x_k$ is $\mathrm{N}_{x_k}\mathcal{M}_{x_k} = (\nabla^2_{\mathcal{E}}\mathcal{L}(x_k, \lambda^*(x_k))\mathrm{T}_x\mathcal{M}_x)^{\perp,\mathcal{E}}$ (or a convergent perturbation of it). This means that quadratic convergent iterates "land" on the manifold $\mathcal{M}$ only through a preferred distinguished direction of the space, which is exactly the one taken by SQP.

We prove this result in two steps. First, we observe that the choice of affine connection influences the Hessian only up to a term vanishing near critical points.

**Proposition 5.4.** *There exists a bilinear tangent mapping $\widetilde{A}_g : \mathrm{T}_x\mathcal{M}_x \times \mathrm{T}_x\mathcal{M}_x \to \mathrm{T}_x\mathcal{M}_x$ continuous with respect to $g$ and $x$ such that the Hessian $\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x) : \mathrm{T}_x\mathcal{M}_x \to \mathrm{T}_x\mathcal{M}_x$ can be rewritten as*

$$\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x)[\xi] = \widetilde{G_T}(x)^{-1}\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_x}f(x)[\xi] + \widetilde{A}_g(\xi, \Pi_x\nabla_{\mathcal{E}}f), \tag{5.16}$$

*where*

*(i)* $\Pi_x : \mathcal{E} \to \mathrm{T}_x\mathcal{M}_x$ *is the $\mathcal{E}$-orthogonal projection operator on* $\mathrm{T}_x\mathcal{M}_x$*;*

*(ii)* $\widetilde{G_T}(x)^{-1} : \mathrm{T}_x\mathcal{M}_x \to \mathrm{T}_x\mathcal{M}_x$ *is the mapping characterized by*

$$\widetilde{G_T}(x)^{-1}\xi \in \mathrm{T}_x\mathcal{M}_x \text{ and } g(\widetilde{G_T}(x)^{-1}\xi, \zeta) = \langle \xi, \zeta\rangle_{\mathcal{E}} \text{ for any } \xi, \zeta \in \mathrm{T}_x\mathcal{M}_x;$$

*(iii)* $\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_x}$ *is the Riemannian Hessian with respect to the Euclidean metric $g \equiv g^{\mathcal{E}}$ on $\mathcal{M}_x$. It reads*

$$\begin{aligned}\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_x}[\xi] &= \Pi_x\nabla^2_{\mathcal{E}}\mathcal{L}(x, \lambda^*(x))\Pi_x\xi \\ &= \Pi_x[\mathrm{d}_\xi\nabla_{\mathcal{E}}f(x) + \mathrm{D}_\xi\Pi_x((\mathrm{Id}_{\mathcal{E}} - \Pi_x)\nabla_{\mathcal{E}}f(x))], \qquad \forall \xi \in \mathrm{T}_x\mathcal{M}_x,\end{aligned}$$

*where $\lambda^*(x) := (\mathrm{D}c(x)\mathrm{D}c(x)^{*,\mathcal{E}})^{-1}\mathrm{D}c(x)\nabla_{\mathcal{E}}f(x)$ is the least-squares multiplier.*

*Proof.* Due to (5.14), the Hessian $\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x) : \mathrm{T}_x\mathcal{M}_x \to \mathrm{T}_x\mathcal{M}_x$ with respect to the connection $\nabla^A$ and the metric $g$ is the operator defined by

$$\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x)[\xi] := \nabla^A_\xi\mathrm{grad}^g_{\mathcal{M}_x}f = \mathrm{Proj}_x(\mathrm{D}_\xi\mathrm{grad}^g_{\mathcal{M}_x}f) + A(\xi, \mathrm{grad}^g_{\mathcal{M}_x}f), \qquad \forall\xi \in \mathrm{T}_x\mathcal{M}_x,$$

where for the moment, we don't specify the tangent projection operator $\mathrm{Proj}_x$. Since $\mathrm{grad}^g_{\mathcal{M}}f = \widetilde{G_T}(x)^{-1}\mathrm{Proj}^{*,\mathcal{E}}_x\nabla_{\mathcal{E}}f$, we find

$$\begin{aligned}\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x)[\xi] &= \mathrm{Proj}_x\left(\mathrm{d}_\xi(\widetilde{G_T}(x)^{-1}\mathrm{Proj}^{*,\mathcal{E}}_x\nabla_{\mathcal{E}}f)\right) + A(\xi, \widetilde{G_T}(x)^{-1}\mathrm{Proj}^{*,\mathcal{E}}_x\nabla_{\mathcal{E}}f) \\ &= \mathrm{Proj}_x(\widetilde{G_T}(x)^{-1}\mathrm{d}_\xi\left[\mathrm{Proj}^{*,\mathcal{E}}_x(\nabla_{\mathcal{E}}f)\right]) \\ &\quad + \mathrm{Proj}_x\left(\mathrm{D}_\xi(\widetilde{G_T}(x)^{-1}\mathrm{Proj}^{*,\mathcal{E}}_x)\mathrm{Proj}^{*,\mathcal{E}}_x\nabla_{\mathcal{E}}f\right) + A(\xi, \widetilde{G_T}(x)^{-1}\mathrm{Proj}^{*,\mathcal{E}}_x\nabla_{\mathcal{E}}f),\end{aligned} \tag{5.17}$$

where we note that $\widetilde{G_T}(x)^{-1}\mathrm{Proj}^{*,\mathcal{E}}_x = G(x)^{-1}\mathrm{Proj}^{*,\mathcal{E}}_x$ is differentiable as a mapping $\mathcal{E} \to \mathcal{E}$ since $G(x)^{-1}$ and $\mathrm{Proj}^{*,\mathcal{E}}_x$ are differentiable. Let us then recall the *Weingarten identities* for the dual projector $\mathrm{Proj}^{*,\mathcal{E}}_x$ (Feppon and Lermusiaux, 2019, Proposition 4):

$$\mathrm{D}_\eta\mathrm{Proj}^{*,\mathcal{E}}_x(\xi) = (\mathrm{Proj}^{\perp}_x)^{*,\mathcal{E}}(\mathrm{D}_\eta\xi) \text{ for any } \eta \in \mathrm{T}_x\mathcal{M}_x, \xi \in \mathcal{C}^\infty(\mathcal{M}_x, \mathrm{N}_x\mathcal{M}^{\perp,\mathcal{E}}_x),$$

$$\mathrm{D}_\eta\mathrm{Proj}^{*,\mathcal{E}}_x(\zeta) = -\mathrm{Proj}^{*,\mathcal{E}}_x(\mathrm{D}_\eta\zeta) \text{ for any } \eta \in \mathrm{T}_x\mathcal{M}_x, \zeta \in \mathcal{C}^\infty(\mathcal{M}_x, \mathrm{T}_x\mathcal{M}^{\perp,\mathcal{E}}_x).$$

These identities can be obtained by differentiating the identities $\mathrm{Proj}^{*,\mathcal{E}}_x\xi = \xi$ and $\mathrm{Proj}^{*,\mathcal{E}}_x\zeta = 0$ with respect to $\eta \in \mathrm{T}_x\mathcal{M}_x$ for any smooth vector field $\xi$ with values in $\mathrm{N}_x\mathcal{M}^{\perp,\mathcal{E}}_x$ and any smooth vector field $\zeta$ with values in $\mathrm{T}_x\mathcal{M}_x^{\perp,\mathcal{E}}$. Using $\mathrm{Range}(\widetilde{G_T}(x)^{-1}) = \mathrm{T}_x\mathcal{M}_x$ and the first Weingarten identity, we obtain

$$\begin{aligned}(5.18)\quad & \mathrm{Proj}_x(\widetilde{G_T}(x)^{-1}\mathrm{d}_\xi[\mathrm{Proj}_x^{*,\mathcal{E}}\nabla_{\mathcal{E}}f(x)]) \\ &\qquad = \widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}\mathrm{D}_\xi\mathrm{Proj}_x^{*,\mathcal{E}}\nabla_{\mathcal{E}}f(x) + \widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}(\mathrm{D}_\xi\nabla_{\mathcal{E}}f(x)).\end{aligned}$$

Introducing $\widetilde{A}_g(\xi,\zeta) := \mathrm{Proj}_x(\mathrm{D}_\xi(\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}})\zeta) + A(\xi,\widetilde{G_T}(x)^{-1}\zeta)$, (5.17) and (5.18) lead to

$$\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x)[\xi] = \widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}\left(\mathrm{D}_\xi\nabla_{\mathcal{E}}f(x) + \mathrm{D}_\xi\mathrm{Proj}_x^{*,\mathcal{E}}((\mathrm{Proj}_x^{\perp})^{*,\mathcal{E}}\nabla_{\mathcal{E}}f(x)\right) + \widetilde{A}_g(\xi,\Pi_x\nabla_{\mathcal{E}}f).$$

The projection perator $\mathrm{Proj}_x$ can be chosen freely in the definition of the covariant derivative by changing the operator $A$ according to lemma 5.3 and when writing $\widetilde{G_T}(x)^{-1}\mathrm{Proj}_x^{*,\mathcal{E}}$, according to remark 4.3. The result follows by choosing $\mathrm{Proj}_x = \Pi_x$ to be the $\mathcal{E}$-orthogonal projector $\Pi_x$ on $\mathrm{T}_x\mathcal{M}_x$, which is self-adjoint. The point (iii) can be found in Absil et al. (2009). ☐

The term in (5.16) featuring $A$ vanishes at critical points $x^*$ satisfying $\Pi_x\nabla_{\mathcal{E}}f(x) = 0$. This implies that the tangent Newton step in (5.13) does not significantly depend on the affine connection $\nabla^A$ or on the tangent metric $G_T(x)$.

**Lemma 5.5.** *In the vicinity of a critical point $x^*$ with $\mathrm{grad}^g_{\mathcal{M}_{x^*}}f(x^*) = 0$ and $\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_x}f(x^*)$ invertible, all Riemannian Newton steps agree up to a quadratically vanishing error:*

$$(5.19)\qquad -\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x)^{-1}\mathrm{grad}^g_{\mathcal{M}_x}f(x) = -(\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_x}f(x))^{-1}\Pi_x\nabla_{\mathcal{E}}f(x) + O(||\Pi_x\nabla_{\mathcal{E}}f(x)||^2_{\mathcal{E}}).$$

*Proof.* Let $d_T(x) = -(\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x))^{-1}\mathrm{grad}^g_{\mathcal{M}_x}f(x)$. It holds $\mathrm{Hess}^{g,A}_{\mathcal{M}_x}f(x)d_T(x) = -\mathrm{grad}^g_{\mathcal{M}_x}f(x) = -\widetilde{G_T}(x)\mathrm{Proj}_x^{*,\mathcal{E}}\nabla_{\mathcal{E}}f(x)$. Using (5.16), left multiplying by $\Pi_x\widetilde{G_T}(x)$ and using $\Pi_x\mathrm{Proj}_x^* = \Pi_x$, we find

$$(\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_x} + O(\Pi_x\nabla_{\mathcal{E}}f(x)))d_T(x) = -\Pi_x\nabla_{\mathcal{E}}f(x).$$

The result follows by a standard perturbation analysis. ☐

We are now in position to prove that Riemannian-Newton landing iterations (5.13) converge quadratically when choosing the normal space $\mathrm{N}_x\mathcal{M}_x = (\nabla^2_{\mathcal{E}}\mathcal{L}(x,\lambda^*(x)))^{\perp,\mathcal{E}}$, characterizing the normal step $d_N(x) = -\mathrm{D}c(x)^{\dagger,g}c(x) = -\mathrm{Proj}_x^{\perp}\mathrm{D}c(x)^{\dagger,\mathcal{E}}c(x)$. To make the proof more readable, we don't explicit here the various Lipschitz constants involved in the convergence estimates, absorbing them in the $O(\,\cdot\,)$ notation.

**Proposition 5.6.** *Assume that $c$ and $f$ are continuous twice differentiable mappings on $\mathcal{E}$. Let $x^* \in \mathcal{M}$ be a local minimizer of (P) with Riemannian Hessian $\mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_{x^*}}f(x^*)$ positive definite at $x^*$. If the metric $g$ is set such that the normal space is $\mathrm{N}_{x_k}\mathcal{M}_{x_k} = (\nabla^2_{\mathcal{E}}\mathcal{L}(x_k,\lambda^*(x_k)))^{\perp,\mathcal{E}}$, then, for any $x_0 \in \mathcal{E}$ sufficiently close to $x^*$, the Riemannian-Newton landing iterates (5.13) are well-defined and converge quadratically to $x^*$: there exists a constant $\gamma > 0$ independent of $k \in \mathbb{N}$ such that*

$$||x_{k+1} - x^*||_{\mathcal{E}} \leq \gamma||x_k - x^*||^2_{\mathcal{E}}, \qquad \forall k \in \mathbb{N}.$$

*Proof.* Due to (5.13) and (5.19).

$$\begin{aligned}(5.20)\quad & \\ x_{k+1} - x^* &= x_k - x^* - \mathrm{Hess}^{g,A}_{\mathcal{M}_{x_k}}f(x_k)^{-1}\mathrm{grad}^g_{\mathcal{M}_{x_k}}f(x_k) - \mathrm{D}c(x_k)^{\dagger,g}c(x_k) \\ &= x_k - x^* - \mathrm{Hess}^{\mathcal{E}}_{\mathcal{M}_{x_k}}f(x_k)^{-1}\Pi_{x_k}\nabla_{\mathcal{E}}f(x_k) - \mathrm{D}c(x_k)^{\dagger,g}c(x_k) + O(||\Pi_{x_k}\nabla_{\mathcal{E}}f(x_k)||^2_{\mathcal{E}}).\end{aligned}$$

Since $\Pi_{x^*}\nabla_{\mathcal{E}}f(x^*) = \mathrm{grad}^{\mathcal{E}}_{\mathcal{M}_{x^*}}f(x^*) = 0$ and $x \mapsto \Pi_x\nabla_{\mathcal{E}}f(x)$ is a Lipschitz map, it is clear that

$$(5.21)\qquad O(||\Pi_{x_k}\nabla_{\mathcal{E}}f(x_k)||^2_{\mathcal{E}}) = O(||x_k - x^*||^2_{\mathcal{E}}).$$

Then, by considering a Taylor expansion, we have on the one hand,

$$(5.22)\qquad 0 = c(x^*) = c(x_k) + \mathrm{D}c(x_k)[x^* - x_k] + O(||x_k - x^*||^2_{\mathcal{E}}),$$

which implies

$$\mathrm{Proj}_{x_k}^{\perp}(x^* - x_k) = \mathrm{D}c(x_k)^{\dagger,g}\mathrm{D}c(x_k)[x^* - x_k] = -\mathrm{D}c(x_k)^{\dagger,g}c(x_k) + O(\|x_k - x^*\|_{\mathcal{E}}^2).$$

Inserting into (5.20), we obtain

$$x_{k+1} - x^* = \mathrm{Proj}_{x_k}(x_k - x^*) - \mathrm{Hess}_{\mathcal{M}_{x_k}}^{\mathcal{E}} f(x_k)^{-1}\Pi_{x_k}\nabla_{\mathcal{E}} f(x_k) + O(\|x_k - x^*\|_{\mathcal{E}}^2). \tag{5.23}$$

On the other hand, performing a Taylor expansion of $x \mapsto \Pi_x \nabla_{\mathcal{E}} f(x)$ around $x_k$ and multiplying by $\Pi_{x_k}$,

$$0 = \Pi_{x_k}\Pi_{x^*}\nabla_{\mathcal{E}} f(x^*) = \Pi_{x_k}\Pi_{x_k}\nabla_{\mathcal{E}} f(x_k) + \Pi_{x_k}\mathrm{D}[\Pi_{\cdot}\nabla_{\mathcal{E}} f](x_k)[x^* - x_k] + O(\|x_k - x^*\|_{\mathcal{E}}^2). \tag{5.24}$$

Recall that for any $\xi \in \mathcal{E}$ (Absil et al., 2009),

$$\Pi_x \mathrm{D}_\xi(\Pi_{\cdot}\nabla_{\mathcal{E}} f)(x) = \Pi_x(\mathrm{D}_\xi \Pi_x[\Pi_x^{\perp}\nabla_{\mathcal{E}} f(x)] + \mathrm{D}_\xi \nabla_{\mathcal{E}} f(x)) = \Pi_x \nabla_{\mathcal{E}}^2 \mathcal{L}(x, \lambda^*(x))[\xi].$$

Consequently,

$$\begin{aligned}
\Pi_{x_k}\mathrm{D}(\Pi_{\cdot}\nabla_{\mathcal{E}} f)(x_k)[x^* - x_k] &= \Pi_{x_k}\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda^*(x_k))[x^* - x_k] \\
&= \Pi_{x_k}\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda^*(x_k))\mathrm{Proj}_{x_k}(x^* - x_k) + \Pi_{x_k}\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda^*(x_k))\mathrm{Proj}_x^{\perp}(x^* - x_k) \\
&= \mathrm{Hess}_{\mathcal{M}_{x_k}}^{\mathcal{E}} f(x_k)\mathrm{Proj}_{x_k}(x^* - x_k) + \Pi_{x_k}\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda^*(x_k))\mathrm{Proj}_x^{\perp}(x^* - x_k) \\
&\quad + O(\|x_k - x^*\|_{\mathcal{E}}^2).
\end{aligned}$$

Substituting into (5.23) after using (5.24), we infer

$$x_{k+1} - x^* = \mathrm{Hess}_{\mathcal{M}_{x_k}}^{\mathcal{E}} f(x_k)^{-1}\Pi_{x_k}\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda^*(x_k))\mathrm{Proj}_{x_k}^{\perp}(x^* - x_k) + O(\|x_k - x^*\|_{\mathcal{E}}^2). \tag{5.25}$$

Therefore, if $\mathrm{N}_{x_k}\mathcal{M}_{x_k} = (\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda^*(x_k)))^{\perp,\mathcal{E}}$, we have $\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda_k^*)\mathrm{N}_{x_k}\mathcal{M}_{x_k} = \mathrm{T}_{x_k}\mathcal{M}_{x_k}^{\perp,\mathcal{E}}$ and

$$\Pi_{x_k}\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda^*(x_k))\mathrm{Proj}_{x_k}^{\perp} = 0.$$

It follows that $x_{k+1} - x^* = O(\|x_k - x^*\|_{\mathcal{E}}^2)$, which completes the proof. □

If $\mathrm{N}_{x_k}\mathcal{M}_{x_k} \neq (\nabla_{\mathcal{E}}^2\mathcal{L}(x_k, \lambda^*(x_k)))^{\perp,\mathcal{E}}$, (5.25) shows that we only have a priori $x_{k+1} - x^* = O(\|x_k - x^*\|)$, which prevents quadratic convergence.

## 6. Global convergence using adaptive step sizes

In this section, we design an adaptive step size globalization procedure for the landing method. Namely, we propose a backtracking linesearch based on an Armijo condition for the decrease of a merit function. Our approach relies heavily on classical techniques for the globalization of SQP (Nocedal and Wright, 2006; Curtis et al., 2024), leveraging the fact—highlighted in section 5.2—that SQP is a particular instance of the Riemannian landing.

We consider the landing update with the normal term written in pseudoinverse form, given by

$$d(x) = d_T(x) + d_N(x) = -\mathrm{grad}_{\mathcal{M}_x}^g f(x) - \mathrm{D}c(x)^{\dagger,g}H(x)c(x) \tag{6.1}$$

for some metric $g$ on $\mathcal{E}$ and $H\colon \mathcal{F} \to \mathcal{F}$, as described in section 4.2.

### 6.1. Assumptions

We introduce a set of generic assumptions on the manifold and Lipschitz constants that will arise in our analysis. The assumptions are standard in the SQP literature (e.g., Berahas et al. (2021); Curtis et al. (2024)). We first assume that there exists a constraint threshold $R > 0$ such that the following compactness assumption holds.

**A1.** *There exists a threshold $R$ such that the set $\mathcal{R} = \{x \in \mathcal{E} : \|c(x)\|_{\mathcal{F}} \leq R\}$ is compact, contains all the iterates $(x_k)_{k\in\mathbb{N}}$ and trial points, and $f$ is bounded from below on $\mathcal{R}$:*

$$\inf_{x\in\mathcal{R}} f(x) > -\infty.$$

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

**A2.** *$f$ and $c$ are continuously differentiable on $\mathcal{D}$ with Lipschitz derivatives. In particular, there exist constants $||\nabla_{\mathcal{E}} f||_\infty, ||\mathrm{D}c||_\infty, L_f, L_c > 0$ such that*

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

$$\|\nabla_{\mathcal{E}} f(x) - \nabla_{\mathcal{E}} f(y)\|_{\mathcal{E}} \leq L_f \|x-y\|_{\mathcal{E}} \quad \forall x, y \in \mathcal{R},$$

$$\|\mathrm{D}c(x) - \mathrm{D}c(y)\|_{\mathcal{E}\to\mathcal{F}} \leq L_c \|x-y\|_{\mathcal{E}} \quad \forall x, y \in \mathcal{R},$$

*where $||A||_{\mathcal{E}\to\mathcal{F}} := \sup_{v\in\mathcal{E}\setminus\{0\}} ||Av||_{\mathcal{F}}/||v||$ denotes the operator norm for linear operators $A : \mathcal{E} \to \mathcal{F}$.*

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

**A3.** *There exists a positive constant $\underline{\sigma}$ 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 $k$th 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 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*

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

This assumption implies the following bound which will be useful in the proof of proposition 6.4.

**Lemma 6.1.** *Assuming A3, the $g$- pseudo inverse can be bounded in terms of the Euclidean pseudoinverse:*

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

*Proof.* This follows from the fact that $\mathrm{D}c(x)^{\dagger,g} w$ is the vector $\xi \in \mathcal{E}$ of smallest $g$-norm satisfying $\mathrm{D}c(x)\xi = w$. Since $\mathrm{D}c(x)^{\dagger,\mathcal{E}} w$ is another vector satisfying this property, this implies that

$$\sqrt{\underline{g}}||\mathrm{D}c(x)^{\dagger,g} w||_{\mathcal{E}} \leq ||\mathrm{D}c(x)^{\dagger,g} w||_g \leq ||\mathrm{D}c(x)^{\dagger,\mathcal{E}} w||_g \leq \sqrt{\bar{g}}||\mathrm{D}c(x)^{\dagger,\mathcal{E}} w||_{\mathcal{E}}.$$

□

Finally, we also assume the uniform ellipticity of the operator field $H(x)$ involved in the normal space direction (4.13).

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

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

**Remark 6.1.** *Assumption A5 holds in both cases $H(x) = \mathrm{Id}_{\mathcal{F}}$ and $H(x) = \mathrm{D}c(x)\mathrm{D}c(x)^{*,g}$.*

**Algorithm 1** Riemannian Landing Method

**Require:** choose $\eta \in (0, \frac{1}{2})$, $\beta \in (0,1)$; $\rho \in (0, \lambda_{\min}(H)/2)$; initial $x_0 \in \mathcal{E}$

$\mu_0 = 1$

$k = 1$

**repeat**

  Compute $d_k = d(x_k)$, the landing direction (6.1) at $x_k$

  Set $\mu_k = \mu_{k-1}$ if $c(x_k) = 0$ and

$$\mu_k = \max\left\{\mu_{k-1}, \frac{\mathrm{D}f(x_k)d_N(x_k)}{\rho\,\|c(x_k)\|_{\mathcal{F}}}\right\} \text{ if } c(x_k) \neq 0. \tag{6.3}$$

  Set $\alpha_k \leftarrow 1$

  **while** $\phi_{\mu_k}(x_k + \alpha_k d_k) > \phi_{\mu_k}(x_k) + \eta\,\alpha_k\,\mathrm{D}\phi_{\mu_k}(x_k)d_k$ **do**

    $\alpha_k \leftarrow \beta\,\alpha_k$   ▷ backtracking

  **end while**

  $x_{k+1} \leftarrow x_k + \alpha_k d_k$

**until** convergence

### 6.2. Line search procedure

The line search procedure is based on decreasing the $\ell_2$-merit function

$$\phi_\mu(x) = f(x) + \mu\,\|c(x)\|_{\mathcal{F}} \tag{6.2}$$

for a large enough penalty parameter $\mu$. We perform a backtracking line-search on the $\ell_2$-merit $\phi_\mu(x)$ with Armijo parameters $\eta \in (0,1)$ and $\beta \in (0,1)$ as described in algorithm 1.

The following lemma gives the directional derivative of the merit function $\phi_\mu$.

**Lemma 6.2.** *If $x \in \mathcal{E}$ is such that $\|c(x)\| \neq 0$, the merit function $\phi_\mu$ is differentiable at $x$ and*

$$\mathrm{D}\phi_\mu(x)d = \mathrm{D}f(x)d + \mu\left\langle \frac{c(x)}{\|c(x)\|}, \mathrm{D}c(x)d\right\rangle_{\mathcal{F}}. \tag{6.4}$$

*For $c(x) = 0$, the merit function is not differentiable at $x$; however, it admits the following directional derivative in any direction $d \in \mathcal{E}$:*

$$\mathrm{D}\phi_\mu(x)d := \lim_{t\to 0} \frac{\phi_\mu(x+td) - \phi_\mu(x)}{t} = \mathrm{D}f(x)d + \mu\,\|\mathrm{D}c(x)d\|_{\mathcal{F}}. \tag{6.5}$$

In what follows, for $x \in \mathcal{R}$, we denote by

$$d(x) = d_T(x) + d_N(x) \tag{6.6}$$

the landing step with $d_T(x) = -\mathrm{grad}^g_{\mathcal{M}_x} f(x)$ and $d_N(x) = -\mathrm{D}c(x)^{\dagger,g}H(x)c(x)$.

The following result shows that if the penalty parameter is large enough, the landing direction is a descent direction for the merit function $\phi_\mu$.

**Proposition 6.3** (Sufficient decrease)**.** *Let $x \in \mathcal{R}$. If $c(x) = 0$, or if $c(x) \neq 0$ and*

$$\mu \geq \frac{\mathrm{D}f(x)d_N(x)}{\rho\,\|c(x)\|_{\mathcal{F}}}, \tag{6.7}$$

*with $\rho \leq \lambda_{\min}(H(x))/2$, then the directional derivative of $\phi_\mu(x)$ in the direction $d(x)$ satisfies*

$$\mathrm{D}\phi_\mu(x)[d(x)] \leq -\|\mathrm{grad}^g_{\mathcal{M}_x} f(x)\|_g^2 - \rho\mu\|c(x)\|_{\mathcal{F}}. \tag{6.8}$$

*Proof.* For $\rho \leq \lambda_{\min}(H(x))/2$, we have

$$\frac{\mathrm{D}f(x)d_N(x)\,\|c(x)\|_{\mathcal{F}}}{\langle c(x), H(x)c(x)\rangle_{\mathcal{F}} - \rho\,\|c(x)\|^2_{\mathcal{F}}} \leq \frac{\mathrm{D}f(x)d_N(x)\,\|c(x)\|_{\mathcal{F}}}{(\lambda_{\min}(H(x)) - \rho)\,\|c(x)\|^2_{\mathcal{F}}} \leq \frac{\mathrm{D}f(x)d_N(x)}{\rho\,\|c(x)\|_{\mathcal{F}}} \leq \mu. \tag{6.9}$$

For $\|c(x)\| \neq 0$, (6.4) and $\mathrm{D}c(x)d(x) = -H(x)c(x)$ imply

$$\begin{aligned}
\mathrm{D}\phi_\mu(x)d(x) &= \mathrm{D}f(x)d_T(x) + \mathrm{D}f(x)d_N(x) + \mu\left\langle \frac{c(x)}{\|c(x)\|_{\mathcal{F}}}, \mathrm{D}c(x)d(x)\right\rangle_{\mathcal{F}} \\
&= g\left(\nabla_g f(x), -\mathrm{grad}^g_{\mathcal{M}_x} f(x)\right) + \mathrm{D}f(x)d_N(x) - \mu\frac{\langle c(x), H(x)c(x)\rangle_{\mathcal{F}}}{\|c(x)\|_{\mathcal{F}}} \\
&\leq -\left\|\mathrm{grad}^g f(x)\right\|_g^2 + \frac{\mu}{\|c(x)\|_{\mathcal{F}}}(\langle c(x), H(x)c(x)\rangle_{\mathcal{F}} - \rho\left\|c(x)\right\|_{\mathcal{F}}^2) - \mu\frac{\langle c(x), H(x)c(x)\rangle_{\mathcal{F}}}{\|c(x)\|_{\mathcal{F}}} \\
&= -\left\|\mathrm{grad}^g f(x)\right\|_g^2 - \mu\rho\left\|c(x)\right\|_{\mathcal{F}},
\end{aligned}$$

where we have used (6.9) in the third line. For $x \in \mathcal{M}$, where $c(x) = 0$, the landing direction $d$ only has a tangent component ($d_N(x) = 0$) and the directional derivative (6.5) becomes

$$\mathrm{D}\phi_\mu(x)d(x) = \mathrm{D}f(x)d_T(x) = -\left\|\mathrm{grad}^g f(x)\right\|_g^2.$$

Thus (6.8) holds for all $x \in \mathcal{R}$. □

The following result shows that the penalty parameter $\mu$ remains bounded in the region of interest $\mathcal{R}$.

**Proposition 6.4** (Upper bound on the penalty parameter)**.** *Under the above assumptions, the increasing merit parameter sequence $(\mu_k)_{k\geq 0}$ defined by (6.3) in algorithm 1 remains bounded:*

$$\mu(x) := \frac{\mathrm{D}f(x)d_N(x)}{\rho\left\|c(x)\right\|_{\mathcal{F}}} \leq \bar{\mu} := \frac{1}{\rho}\sqrt{\frac{\bar{g}}{\underline{g}}}\left\|\nabla_{\mathcal{E}} f\right\|_\infty \left\|\mathrm{D}c^{\dagger,\mathcal{E}}\right\|_\infty \lambda_{\max}(H), \qquad \forall x \in \mathcal{R}. \tag{6.10}$$

*In particular, the sequence $(\mu_k)_{k\in\mathbb{N}}$ is convergent.*

*Proof.* Recall that $\mu(x)$. Using the Cauchy-Schwartz inequality and lemma 6.1, we find

$$\begin{aligned}
|\mathrm{D}f(x)d_N(x)| = \langle \nabla_{\mathcal{E}} f(x), -\mathrm{D}c(x)^{\dagger,g}H(x)c(x)\rangle_{\mathcal{E}} &\leq \sqrt{\frac{\bar{g}}{\underline{g}}}||\nabla_{\mathcal{E}} f(x)||_{\mathcal{E}}\left\|\mathrm{D}c^{\dagger,\mathcal{E}}\right\|_\infty ||H(x)c(x)||_{\mathcal{E}} \\
&\leq \sqrt{\frac{\bar{g}}{\underline{g}}}\left\|\nabla_{\mathcal{E}} f\right\|_\infty \left\|\mathrm{D}c^{\dagger,\mathcal{E}}\right\|_\infty \lambda_{\max}(H)\left\|c(x)\right\|.
\end{aligned}$$

□

In what follows, we use the following inequality, actually valid for any Euclidean norm.

**Lemma 6.5.** *Let $v, w \in \mathcal{F}$ with $v \neq 0$. The following inequality holds:*

$$\|v + w\|_{\mathcal{F}} \leq \|v\|_{\mathcal{F}} + \left\langle \frac{v}{\|v\|_{\mathcal{F}}}, w\right\rangle_{\mathcal{F}} + \frac{\|w\|_{\mathcal{F}}^2}{2\|v\|_{\mathcal{F}}}. \tag{6.11}$$

*Proof.* We have

$$\|v + w\|_{\mathcal{F}}^2 = \|v\|_{\mathcal{F}}^2 + 2\langle v, w\rangle_{\mathcal{F}} + \|w\|_{\mathcal{F}}^2 = \|v\|_{\mathcal{F}}^2(1 + t). \tag{6.12}$$

with $t := (2\langle v, w\rangle + \|w\|_{\mathcal{F}}^2)/\|v\|_{\mathcal{F}}^2$. Since $1 + t = \|v + w\|_{\mathcal{F}}^2/\|v\|_{\mathcal{F}}^2 \geq 0$, it is clear that $t \geq -1$. Recalling the elementary inequality $\sqrt{1+t} \leq 1 + t/2$, valid for $t \geq -1$, this implies

$$\|v + w\|_{\mathcal{F}} = \|v\|_{\mathcal{F}}\sqrt{1+t} \leq \|v\|_{\mathcal{F}}\left(1 + \frac{t}{2}\right) = \|v\|_{\mathcal{F}}\left(1 + \frac{\langle v, w\rangle}{\|v\|_{\mathcal{F}}^2} + \frac{\|w\|_{\mathcal{F}}^2}{2\left\|v\right\|_{\mathcal{F}}^2}\right). \tag{6.13}$$

□

The following lemma gives an upper bound on the merit function along the landing direction, which we use below to show finite termination of the linesearch.

**Lemma 6.6** (Quadratic upper bound for the merit function)**.** *For any $x \in \mathcal{R}$ and $\alpha \geq 0$ such that $x + \alpha d(x) \in \mathcal{R}$ with $d(x)$ the landing step (6.6), it holds that*

$$\phi_\mu(x+\alpha d(x)) \leq \phi_\mu(x)+\alpha \mathrm{D}\phi_\mu(x)d(x)+\frac{\alpha^2}{2}\big(L_f+\mu L_c\big)\|d(x)\|_{\mathcal{E}}^2+\frac{\alpha^2}{2}\mu\lambda_{\max}(H)^2\,\|c(x)\|_{\mathcal{F}}, \tag{6.14}$$

*where $\mathrm{D}\phi_\mu(x)d$ denotes the one-sided directional derivative of $\phi_\mu$ at $x$ along $d$, given by Lemma 6.2.*

*Proof.* By the Lipschitz continuity of $\nabla_{\mathcal{E}} f$ (assumption A2), we can write

$$f(x+\alpha d(x)) \leq f(x) + \alpha \mathrm{D}f(x)d(x) + \tfrac{1}{2}L_f\alpha^2\|d(x)\|_{\mathcal{E}}^2. \tag{6.15}$$

By the Lipschitz continuity of $\mathrm{D}c$ and $\mathrm{D}c(x)d(x) = \mathrm{D}c(x)d_N(x) = -H(x)c(x)$, we also have

$$\begin{aligned} c(x+\alpha d(x)) &= c(x) + \alpha \mathrm{D}c(x)d(x) + r(\alpha), \\ &= (\mathrm{Id}_{\mathcal{F}} - \alpha H(x))c(x) + r(\alpha) \end{aligned} \tag{6.16}$$

with $\|r(\alpha)\|_{\mathcal{F}} \leq \frac{1}{2}L_c\,\alpha^2\|d(x)\|_{\mathcal{E}}^2$.
If $c(x) = 0$, it follows that $\|c(x+\alpha d)\|_{\mathcal{F}} \leq \|r(\alpha)\|_{\mathcal{F}}$, which imply

$$\begin{aligned} \phi_\mu(x+\alpha d(x)) &= f(x+\alpha d(x)) + \mu\|c(x+\alpha d(x))\|_{\mathcal{F}} \\ &\leq f(x) + \alpha \mathrm{D}f(x)d(x) + \tfrac{1}{2}L_f\alpha^2\|d(x)\|_{\mathcal{E}}^2 + \mu\tfrac{1}{2}L_c\,\alpha^2\|d(x)\|_{\mathcal{E}}^2. \end{aligned} \tag{6.17}$$

Moreover, (6.5) together with $\mathrm{D}c(x)d(x) = 0$ entail $\mathrm{D}f(x)d(x) = \mathrm{D}\phi_\mu(x)d(x)$. Therefore (6.14) holds if $c(x) = 0$.
We now consider the case $c(x) \neq 0$. Majoring the first line of section 6.2 using (6.11) with $v = c(x)$ and $w = \alpha\mathrm{D}c(x)d(x)$ gives

$$\begin{aligned} \|c(x+\alpha d(x))\|_{\mathcal{F}} &\leq \|c(x)\|_{\mathcal{F}} + \alpha\left\langle \frac{c(x)}{\|c(x)\|_{\mathcal{F}}}, \mathrm{D}c(x)d(x)\right\rangle_{\mathcal{F}} + \frac{\alpha^2}{2}\frac{\|\mathrm{D}c(x)d(x)\|_{\mathcal{F}}^2}{\|c(x)\|_{\mathcal{F}}} + \|r(\alpha)\|_{\mathcal{F}} \\ &\leq \|c(x)\|_{\mathcal{F}} + \alpha\left\langle \frac{c(x)}{\|c(x)\|_{\mathcal{F}}}, \mathrm{D}c(x)d(x)\right\rangle_{\mathcal{F}} + \frac{\alpha^2}{2}\lambda_{\max}(H)^2\,\|c(x)\|_{\mathcal{F}} + \frac{1}{2}L_c\,\alpha^2\|d\|_{\mathcal{E}}^2. \end{aligned} \tag{6.18}$$

Combining (6.18), (6.15) and (6.4) yields (6.14). □

The following result shows that the linesearch terminates in a finite number of steps, and gives an explicit lower bound on the stepsize.

**Proposition 6.7** (Lower bound on the step size)**.** *The sequence of step sizes $(\alpha_k)_{k\in\mathbb{N}}$ generated by algorithm 1 are bounded from below:*

$$\alpha_k \geq \underline{\alpha} := \underline{g}\frac{2\beta(1-\eta)}{(L_f+\bar{\mu}L_c)}\min\left\{1,\ \frac{\rho}{c_1}\right\} > 0 \qquad \forall k \in \mathbb{N}, \tag{6.19}$$

*where $c_1$ is the constant*

$$c_1 := \left(\bar{g}\left\|\mathrm{D}c^{\dagger,\mathcal{E}}\right\|_\infty^2 R + \frac{\underline{g}\bar{\mu}}{(L_f+L_c)}\right)\lambda_{\max}(H)^2. \tag{6.20}$$

*Proof.* Owing to (6.14), the Armijo condition

$$\phi_\mu(x+\alpha d(x)) \leq \phi_\mu(x) + \eta\,\alpha\, D\phi_\mu(x)d(x) \tag{6.21}$$

is satisfied whenever

$$\alpha D\phi_\mu(x)d(x) + \frac{\alpha^2}{2}\big(L_f+\mu L_c\big)\|d\|_{\mathcal{E}}^2 + \frac{\alpha^2}{2}\mu\lambda_{\max}(H)^2\,\|c(x)\|_{\mathcal{F}} \leq \eta\,\alpha\, D\phi_\mu(x)d(x), \tag{6.22}$$

or, equivalently,

$$\alpha \leq \alpha_{\min} \text{ with } \alpha_{\min} := \frac{2(1-\eta)|\mathrm{D}\phi_\mu(x)d(x)|}{(L_f+\mu L_c)\,\|d(x)\|_{\mathcal{E}}^2 + \mu\lambda_{\max}(H)^2\,\|c(x)\|_{\mathcal{F}}}. \tag{6.23}$$

This means that, at iteration $k$, the Armijo line search terminates with $\alpha_k > 0$ satisfying either $\alpha_k = 1$ or $\alpha_k/\beta \geq \alpha_{\min}$. The sufficient decrease condition (6.8) implies then

$$\alpha_k \geq \frac{2\beta(1-\eta)}{L_f + \mu_k L_c} \times \frac{\|\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)\|_g^2 + \rho\mu_k \|c(x_k)\|_{\mathcal{F}}}{\|d_k\|_{\mathcal{E}}^2 + \mu_k \lambda_{\max}(H)^2 \|c(x_k)\|_{\mathcal{F}} / (L_f + \mu_k L_c)}. \tag{6.24}$$

By using A4 and the $g$-orthogonality of the tangent and normal step, we descent direction $d_k$ can be estimated as

$$\begin{aligned}
||d_k||_{\mathcal{E}}^2 &\leq \frac{1}{\underline{g}} g(d_k, d_k) = \frac{1}{\underline{g}}(g(d_T(x_k), d_T(x_k)) + g(d_N(x_k), d_N(x_k))) \\
&\leq \frac{1}{\underline{g}} ||\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)||_g^2 + \frac{1}{\underline{g}} ||\mathrm{D}c(x_k)^{\dagger,\mathcal{E}} H(x_k) c(x_k)||_g^2) \\
&\leq \frac{1}{\underline{g}} \left( ||\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)||_g^2 + \bar{g} ||\mathrm{D}c^{\dagger,\mathcal{E}}||_\infty^2 \lambda_{\max}(H)^2 ||c(x_k)||_{\mathcal{F}}^2 \right),
\end{aligned}$$

where we use the property that $d_N(x_k)$ is the minimum $g$-norm vector $\xi \in \mathcal{E}$ satisfying $\mathrm{D}c(x_k)\xi = -H(x_k)c(x_k)$ and that $-\mathrm{D}c^{\dagger,\mathcal{E}}(x_k)H(x_k)c(x_k)$ is another such vector. Therefore, for $\|c(x)\|_{\mathcal{F}} \leq R$, the denominator of (6.24) is upper bounded as follows

$$\begin{aligned}
&\|d_k\|_{\mathcal{E}}^2 + \mu_k \lambda_{\max}(H)^2 \|c(x_k)\|_{\mathcal{F}} / (L_f + \mu_k L_c) \\
&\quad\leq \frac{1}{\underline{g}} \left\|\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)\right\|_g^2 + \frac{\bar{g}}{\underline{g}} \left\|\mathrm{D}c^{\dagger,\mathcal{E}}\right\|_\infty^2 \lambda_{\max}(H)^2 \|c(x_k)\|_{\mathcal{F}}^2 + \frac{\mu_k \lambda_{\max}(H)^2}{L_f + \mu_k L_c} \|c(x_k)\|_{\mathcal{F}} \\
&\quad\leq \frac{1}{\underline{g}} \left\|\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)\right\|_g^2 + \left( \frac{\bar{g}}{\underline{g}} \left\|\mathrm{D}c^{\dagger,\mathcal{E}}\right\|_\infty^2 R + \frac{\mu_k}{L_f + \mu_k L_c} \right) \lambda_{\max}(H)^2 \|c(x_k)\|_{\mathcal{F}} \\
&\quad\leq \frac{1}{\underline{g}} \left( \left\|\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)\right\|_g^2 + c_1 \|c(x_k)\|_{\mathcal{F}} \right),
\end{aligned} \tag{6.25}$$

where $c_1$ is defined in (6.20) and the last inequality follows from $1 \leq \mu_k \leq \bar{\mu}$ for all $k$. Thus we have

$$\alpha_k \geq \underline{g} \frac{2\beta(1-\eta)}{(L_f + \bar{\mu} L_c)} F(\|\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)\|_g^2, ||c(x_k)||_{\mathcal{F}}) \text{ with } F(t_1, t_2) := \frac{t_1 + \rho t_2}{t_1 + c_1 t_2}. \tag{6.26}$$

Finally, for $t_1, t_2 \geq 0$ with $t_1 + c_1 t_2 > 0$, it holds $F(t_1, t_2) \geq \min(1, \rho/c_1)$. This implies the bound. □

We are now in a position to prove our main convergence result for the Riemannian landing method. The convergence rate is consistent with an existing result for SQP (Curtis et al., 2024, Thm. 1).

**Theorem 6.8** (Global convergence of the landing iterates)**.** *Under assumptions A1 to A5, the iterates $(x_k)_{k\in\mathbb{N}}$ generated by algorithm 1 satisfy*

$$\|c(x_k)\|_{\mathcal{F}} \to 0 \text{ and } ||\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)||_{\mathcal{E}} \to 0 \text{ as } k \to +\infty. \tag{6.27}$$

*Additionally, the rate of convergences of these sequences can be estimated as follows:*

$$\min_{k=0,\ldots K-1} \|\mathrm{grad}^g_{\mathcal{M}_x} f(x_k)\|_g \leq \frac{C}{\sqrt{K}} \tag{6.28}$$

*and*

$$\min_{k=0,\ldots K-1} \|c(x_k)\|_{\mathcal{F}} \leq \frac{C}{K}. \tag{6.29}$$

*for a constant $C$ independent of $K$.*

*Proof.* For $K \geq 0$, the Armijo condition reads

$$\phi_{\mu_k}(x_{k+1}) \leq \phi_{\mu_k}(x_k) + \eta\alpha_k \mathrm{D}\phi_{\mu_k}(x_k)d_k,$$

which, thanks to (6.8), implies

$$f(x_{k+1}) + \mu_k||c(x_{k+1})||_{\mathcal{F}} \leq f(x_k) + \mu_k||c(x_k)||_{\mathcal{F}} - \eta\alpha_k||\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)||_g^2 - \eta\alpha_k\rho\mu_k||c(x_k)||_{\mathcal{F}}.$$

This can be rewritten as

$$\begin{aligned}(6.30)\quad f(x_k) - f(x_{k+1}) &+ (\mu_k||c(x_k)||_{\mathcal{F}} - \mu_{k+1}||c(x_{k+1})||_{\mathcal{F}}) \\ &\geq (\mu_k - \mu_{k+1})||c(x_{k+1})||_{\mathcal{F}} + \eta\alpha_k||\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)||_g^2 + \eta\alpha_k\rho\mu_k||c(x_k||_{\mathcal{F}}.\end{aligned}$$

Summing for $0 \leq k \leq K-1$ and using the lower bound (6.19), we obtain

$$\begin{aligned}&f(x_0) - f(x_T) + \mu_0||c(x_0)||_{\mathcal{F}} \\ &\geq \mu_T||c(x_T)||_{\mathcal{F}} + \sum_{k=1}^{K}(\mu_{k-1} - \mu_k)||c(x_k)||_{\mathcal{F}} + \sum_{k=0}^{K-1}\eta\alpha_k\rho\mu_k||c(x_k)||_{\mathcal{F}} + \sum_{k=0}^{K-1}\eta\alpha_k||\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)||_g^2 \\ &\geq \sum_{k=1}^{K-1}(\eta\underline{\alpha}\rho + \mu_{k-1} - \mu_k)||c(x_k)||_{\mathcal{F}} + \mu_{K-1}||c(x_{K-1})||_{\mathcal{F}} + \sum_{k=0}^{K-1}\eta\underline{\alpha}||\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)||_g^2.\end{aligned}$$

Since $\mu_{k-1} - \mu_k \to 0$ as $k \to +\infty$, we conclude that the series $\sum_k ||\mathrm{grad}^g_{\mathcal{M}_{x_k}} f(x_k)||_g^2$ and $\sum_k ||c(x_k)||_{\mathcal{F}}$ are convergent, which implies (6.27) to (6.29). □

Due to the boundedness of the sequence $(x_k)_{k\geq 0}$, there must be an accumulation point, which is necessary a feasible KKT point according to (6.27).

## 7. Geometric design of tangent and normal terms for matrix optimization with orthogonality constraints

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

$$\underset{X\in\mathbb{R}^{n\times p}}{\mathrm{minimize}}\ f(X) \text{ subject to } X^\top X = \mathrm{I_p}, \tag{7.1}$$

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

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

A function that defines the constraints is

$$c\colon \mathbb{R}^{n\times p} \to \mathrm{Sym}(p)\colon c(X) = \frac{1}{2}(X^\top X - \mathrm{I_p}). \tag{7.2}$$

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 $||X||^2 = \langle X, X\rangle$. 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 $\mathrm{skew}(A) = (A - A^\top)/2$.

The squared infeasibility for the problem (7.1) reads

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

where $\mathrm{I}_p$ is the identity matrix of $\mathbb{R}^{p\times p}$. The set $\mathcal{D}$ of (3.1) 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 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 (1.2). 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: we first introduce an alternative family of normal spaces with associated projection operators that can be computed explicitly, then we design tangent and normal metrics that lead to explicit—and thus more exploitable—formulas for the tangent and normal steps for matrix optimization with orthogonality constraints.

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 7.1** (Gao et al. (2022); Goyens et al. (2026))**.** *The tangent space of* $\mathrm{St}_{X^\top X}$ *at* $X$ *is the set*

$$\mathrm{T}_X\mathrm{St}_{X^\top X} = \{\xi \in \mathbb{R}^{n\times p} \,|\, \xi^\top X + X^\top \xi = 0\} \tag{7.3}$$

$$= \{X(X^\top X)^{-1}\Omega + \Delta \,|\, \Omega \in \mathrm{Skew}(p),\, \Delta \in \mathbb{R}^{n\times p} \text{ with } \Delta^\top X = 0\} \tag{7.4}$$

$$= \{WX \,|\, W \in \mathrm{Skew}(n)\}, \tag{7.5}$$

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

### 7.1. Landing algorithm for the Euclidean metric

We first outline the derivation of the tangent and normal steps (1.3) and (4.13) when $g \equiv g^{\mathcal{E}}$ is the Euclidean or Frobenius inner product metric:

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

We thus choose the formulation of the landing algorithm (4.12) based on the pseudoinverse formula for the normal step. The tangent and normal vector fields (1.3) and (4.13), with respect to the Euclidean metric, are given by

$$d_T(X) = -\mathrm{Proj}_{X,\mathcal{E}}[\nabla_{\mathcal{E}} f(X)], \tag{7.6}$$

$$d_N(X) = -\mathrm{D}c(X)^{\dagger,\mathcal{E}}[H(X)[c(X)]] \tag{7.7}$$

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

In what follows, we compute these two directions explicitly. The orthogonal projector $\mathrm{Proj}_{X,\mathcal{E}}$ can be obtained through the characterization (3.5), as is commonly done in the literature on Riemannian optimization Absil et al. (2008). However, since the formula for $d_N(X)$ also requires the computation of the pseudoinverse $\mathrm{D}c(X)^{\dagger,\mathcal{E}}$, we instead propose to compute it using the explicit relation $\mathrm{Proj}_{X,\mathcal{E}} = \mathrm{Id}_{\mathcal{E}} - \mathrm{D}c(X)^{\dagger,\mathcal{E}}\mathrm{D}c(X)$. As it becomes clear below, the main difficulty in computing this projection lies in inverting the symmetric operator $(\mathrm{D}c(X)\mathrm{D}c(X)^{*,\mathcal{E}})^{-1}$ that appears in the pseudoinverse formula $\mathrm{D}c(X)^{\dagger,\mathcal{E}} = \mathrm{D}c(X)^{*,\mathcal{E}}(\mathrm{D}c(X)\mathrm{D}c(X)^{*,\mathcal{E}})^{-1}$. Let us start by evaluating $\mathrm{D}c(X)$ and $\mathrm{D}c(X)^{*,\mathcal{E}}$.

**Proposition 7.2.** *The operator* $\mathrm{D}c(X) : \mathbb{R}^{n\times p} \to \mathrm{Sym}(p)$ *and its adjoint* $\mathrm{D}c(X)^{*,\mathcal{E}} : \mathrm{Sym}(p) \to \mathbb{R}^{n\times p}$ *read:*

*(i)* $\mathrm{D}c(X)[\xi] = \frac{1}{2}(X^\top\xi + \xi^\top X) = \mathrm{sym}(X^\top\xi)$ *for all* $\xi \in \mathbb{R}^{n\times p}$,

*(ii)* $\mathrm{D}c(X)^{*,\mathcal{E}}[S] = XS$ *for all* $S \in \mathrm{Sym}(p)$.

*Proof.* The point (i) is obvious in view of (7.2). For (ii), we write, for any $S \in \mathrm{Sym}(p)$ and $\xi \in \mathbb{R}^{n\times p}$,

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

□

We infer the following characterization of the normal space $\mathrm{N}^{\mathcal{E}}_X\mathrm{St}_{X^\top X} = \mathrm{Range}(\mathrm{D}c(X)^{*,\mathcal{E}})$ and of the Euclidean pseudoinverse $\mathrm{D}c(X)^{\dagger,\mathcal{E}}$.

**Corollary 7.3.** *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*

$$\mathrm{N}_X^{\mathcal{E}}\mathrm{St}_{X^\top X} = \{XS \mid S \in \mathrm{Sym}(p)\}. \tag{7.8}$$

**Proposition 7.4.** *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*

$$\frac{1}{2}(X^\top XS + SX^\top X) = T. \tag{7.9}$$

*Proof.* From proposition 7.2, 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 7.5.** *(i) The tangent term (7.6) of the landing algorithm (1.3) with respect to the Euclidean metric given by*

$$d_T(X) = -\mathrm{Proj}_{X,\mathcal{E}}[\nabla_{\mathcal{E}}(f(X))] = -\nabla_{\mathcal{E}}f(X) + XS, \qquad \forall Z \in \mathbb{R}^{n\times p}, \tag{7.10}$$

*where* $S$ *is the unique solution in* $\mathrm{Sym}(p)$ *to the Sylvester equation (7.9) with* $T = \mathrm{sym}(X^\top\nabla_{\mathcal{E}}f(X))$.
*(ii) The normal vector field (7.7) is given by* $d_N(X) = -XS$*, where* $S$ *is the solution to the Sylvester equation (7.9) with* $T = \frac{1}{2}H(X)[X^\top X - \mathrm{I}_p]$.

*Proof.* These results are obtained by using proposition 7.4 and formula (3.15) for the projection operator $\mathrm{Proj}_{X,\mathcal{E}}$. □

In the general case, we mention that the solution to (7.9) could be calculated from the singular value decomposition of $X$. The following result is classical and may be found e.g. in (Li and Zhou, 2018).

**Lemma 7.6.** *Let* $T \in \mathbb{R}^{p\times p}$ *be an arbitrary matrix. Let* $X = \sum_{i=1}^p \sigma_i u_i v_i^\top$ *be the singular value decomposition of a full rank matrix* $X \in \mathbb{R}^{n\times p}$ *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 (7.9) is given by*

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

When $T = Q(X^\top X)$ with $Q$ an analytic function, the solution to the Sylvester equation (7.9) is explicitly given by

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

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

**Proposition 7.7.** *(i) If* $H(X) = \mathrm{Id}_{\mathcal{F}}$*, then* $d_N(X) = -\frac{1}{2}X(\mathrm{I}_p - (X^\top X)^{-1})$.
*(ii) if* $H(X) = \mathrm{Dc}(X)\mathrm{Dc}(X)^{*,\mathcal{E}}$*, then* $d_N(X) = -\nabla_{\mathcal{E}}\psi(X) = -X(X^\top X - \mathrm{I}_p)$.

*Proof.* For $H(X) = \mathrm{Id}_{\mathcal{F}}$, $d_N(X) = -XS$ with $S$ being the solution to (7.9) 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 $d_N(X) = -\nabla_{\mathcal{E}}\psi(X)$ 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 7.5. We have in this case that $S$ is the solution to (7.9) the right-hand side $T$ given by

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

Hence, the solution to the Sylvester equation (7.9) is $S = (X^\top X - \mathrm{I}_p)$, which yields (ii). $\square$

To summarize, the execution of the landing algorithm (1.3) for minimizing a function on the Stiefel manifold using the Euclidean metric $g^{\mathcal{E}}$ to define the tangent and the normal terms $d_T(X)$ and $d_N(X)$ requires solving a Sylvester equation for computing the tangent term at every iteration, which is potentially costly for large matrices $X$.

### 7.2. 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 $d_T(X)$ and $d_N(X)$ via (4.3) and (4.13), we begin by choosing *first* an explicit family of projection operators $\mathrm{Proj}_X$ 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 7.2.1. In the second step, we construct explicit operators

$$\widetilde{G_T}(X) : \mathrm{T}_X\mathrm{St}_{X^\top X} \to \mathrm{N}_X\mathrm{St}_{X^\top X}^{\perp,\mathcal{E}}, \qquad \widetilde{G_N}(X) : \mathrm{N}_X\mathrm{St}_{X^\top X} \to \mathrm{T}_X\mathrm{St}_{X^\top X}, \tag{7.11}$$

each admitting explicit inverses. This construction defines the metric *a posteriori* through (4.6):

$$g(\xi,\zeta) = \langle \widetilde{G_T}(X)\mathrm{Proj}_X[\xi],\zeta\rangle + \langle \widetilde{G_N}(X)\mathrm{Proj}_X^\perp[\xi],\zeta\rangle. \tag{7.12}$$

Moreover, these explicit inverses provide closed-form expressions for the tangent and normal terms of the landing algorithm via (4.10) and (4.17).

We examine two possible choices for $\widetilde{G_T}$ and $\widetilde{G_N}$. In section 7.2.2, 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 (7.1). In section 7.2.3, we consider instead the family of $\beta$-metrics $g^\beta$ introduced in Goyens et al. (2026), 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 specific choice of $\widetilde{G_T}$ and $\widetilde{G_N}$. These operators turn to admit explicit inverses, thereby highlighting why closed-form formulas for $d_T(X)$ and $d_N(X)$ are also available in this case.

#### 7.2.1. *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 $\mathrm{Proj}_X$ and $\mathrm{Proj}_X^\perp$ are the operators defined by

$$\mathrm{Proj}_X[Z] := X(X^\top X)^{-1}\,\mathrm{skew}(X^\top Z) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)Z, \tag{7.13}$$

$$\mathrm{Proj}_X^\perp[Z] := \mathrm{Id}_{\mathcal{E}} - \mathrm{Proj}_X(Z) = X(X^\top X)^{-1}\,\mathrm{sym}(X^\top Z). \tag{7.14}$$

It is straightforward to verify that $\mathrm{Proj}_X$ and $\mathrm{Proj}_X^{\perp}$ are linear (oblique) projectors. In view of (7.14), it is clear that this choice of projectors corresponds to attaching to $\mathrm{T}_X\mathrm{St}_{X^\top X}$ the normal space

$$\mathrm{N}_X\mathrm{St}_{X^\top X} := \mathrm{Range}(\mathrm{Proj}_X^{\perp}) = \{X(X^\top X)^{-1}S \,|\, S \in \mathrm{Sym}(p)\}, \tag{7.15}$$

to be compared with (7.8). 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 $\mathrm{N}_X\mathrm{St}_{X^\top X}$, $\mathrm{Proj}_X$ and $\mathrm{Proj}_X^{\perp}$ are orthogonal projectors for the metric (7.12) 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 provides expressions for the Euclidean adjoints of the tangent projection $\mathrm{Proj}_X$ and $\mathrm{Proj}_X^{\perp}$.

**Proposition 7.8.** *The adjoint operators* $\mathrm{Proj}_X^{*,\mathcal{E}}$ *and* $(\mathrm{Proj}_X^{\perp})^{*,\mathcal{E}}$ *of the projection operators* $\mathrm{Proj}_X$ *and* $\mathrm{Proj}_X^{\perp}$*—with respect to the Euclidean inner product—are given by*

$$\mathrm{Proj}_X^{*,\mathcal{E}}[Z] = X\,\mathrm{skew}((X^\top X)^{-1}X^\top Z) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)Z, \qquad Z \in \mathbb{R}^{n\times p}, \tag{7.16}$$

$$(\mathrm{Proj}_X^{\perp})^{*,\mathcal{E}}[Z] = X\,\mathrm{sym}((X^\top X)^{-1}X^\top Z), \qquad Z \in \mathbb{R}^{n\times p}. \tag{7.17}$$

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

$$(\mathrm{T}_X\mathrm{St}_{X^\top X})^{\perp,\mathcal{E}} = \mathrm{Range}((\mathrm{Proj}_X^{\perp})^{*,\mathcal{E}}) = \{XS \,|\, S \in \mathrm{Sym}(p)\}, \tag{7.18}$$

(7.19)

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

*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 (7.17) 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 (7.12).

**Proposition 7.9.** *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* (4.8).

*Proof.* Due to (3.12) and proposition 7.2:

$$\mathrm{D}c(X)^{*,g}[S] = G(X)^{-1}\mathrm{D}c(X)^{*,\mathcal{E}}[S] = G(X)^{-1}[XS]. \tag{7.20}$$

Then, in view of proposition 7.8, $\mathrm{Proj}_X^{*,\mathcal{E}}[XS] = 0$ and $(\mathrm{Proj}_X^{\perp})^{*,\mathcal{E}}[XS] = XS$. The result follows by applying (4.9) to (7.20). □

From section 4.1, the tangent and normal steps of the landing algorithm (4.12) read

$$d_T(X) = -\widetilde{G_T}(X)^{-1}\mathrm{Proj}_X^{*,\mathcal{E}}[\nabla_{\mathcal{E}} f(X)], \tag{7.21}$$

$$d_N(X) = -\mathrm{Proj}_X^{\perp}\mathrm{D}c(X)^{\dagger,\mathcal{E}}H(X)[c(X)]. \tag{7.22}$$

The operators $\widetilde{G_T}(X)^{-1}$ and $H(X)$ can be freely chosen by the user which gives some latitude for the computation of $d_T(X)$ and $d_N(X)$, and which have rather clear physical interpretations. We obtain the following formulas for $d_T(X)$ and $d_N(X)$.

**Proposition 7.10.** *The tangent and normal terms (eq. (7.21) and (7.22)) of the landing algorithm (1.3) relatively to the metric (7.12) are given by*

$$d_T(X) = -\widetilde{G_T}(X)^{-1}[X\,\mathrm{skew}((X^\top X)^{-1}X^\top \nabla_{\mathcal{E}} f(X)) + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\nabla_{\mathcal{E}} f(X)], \tag{7.23}$$

$$d_N(X) = -X(X^\top X)^{-1}\,\mathrm{sym}(X^\top X S), \tag{7.24}$$

*where $S$ is the solution to the Sylvester equation (7.9) with $T = \frac{1}{2}H(X)[X^\top X - \mathrm{I}_p]$. The expression of $d_N(X)$ becomes explicit in the following cases:*

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

*Proof.* Formula (7.23) follows immediately from (7.21) and (7.16). Formula (7.24) is obtained by combining (7.22) and the result of (7.6). For the cases (i) and (iii), we observe that due to (7.22),

$$d_N(X) = \mathrm{Proj}_X^\perp[d_N(X)^{\mathcal{E}}]$$

where $d_N(X)^{\mathcal{E}}$ is the corresponding normal direction obtained in proposition 7.7. 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 $d_N(X) = d_N(X)^{\mathcal{E}}$, which proves (i) and (iii). Finally, $H(X) = \mathrm{D}c(X)\mathrm{D}c(X)^{*,g}$ implies that $d_N(X) = -\mathrm{D}c(X)^{*,g}c(X)$, which yields (ii) after using proposition 7.9. $\square$

It remains to choose the operator $\widetilde{G_T}(X)^{-1}$ in (7.23), and the operator $\widetilde{G_N}(X)^{-1}$ if one chooses (ii) for the normal direction.

### 7.2.2. *Construction of the tangent and normal steps based on canonical tangent and normal metric representers*

There is a natural choice of operators $\widetilde{G_T}(X) : \mathrm{T}_X\mathrm{St}_{X^\top X} \to \mathrm{N}_X\mathrm{St}_{X^\top X}^{\perp,\mathcal{E}}$ and $\widetilde{G_N}(X) : \mathrm{N}_X\mathrm{St}_{X^\top X} \to \mathrm{T}_X\mathrm{St}_{X^\top X}^{\perp,\mathcal{E}}$ with explicit inverses $\widetilde{G_T}(X)^{-1}$ and $\widetilde{G_N}(X)^{-1}$, leading to explicit expressions for the tangent and normal steps (7.23) and (7.24). Recalling (7.4) and (7.19), and (7.15) and (7.18), these 'canonical' operators read

$$\widetilde{G_T}(X) : \begin{array}{rcl} \mathrm{T}_X\mathrm{St}_{X^\top X} & \longrightarrow & \mathrm{N}_X\mathrm{St}_{X^\top X}^{\perp,\mathcal{E}} \\ X(X^\top X)^{-1}\Omega + \Delta & \longmapsto & X\Omega + \Delta, \end{array} \tag{7.25}$$

$$\widetilde{G_N}(X) : \begin{array}{rcl} \mathrm{N}_X\mathrm{St}_{X^\top X} & \longrightarrow & \mathrm{T}_X\mathrm{St}_{X^\top X}^{\perp,\mathcal{E}} \\ X(X^\top X)^{-1}S & \longmapsto & XS, \end{array} \tag{7.26}$$

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 have the following natural extensions to the whole $\mathcal{E} = \mathbb{R}^{n\times p}$:

$$G_T(X)[Z] := XX^\top Z + (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)Z, \qquad Z \in \mathbb{R}^{n\times p}, \tag{7.27}$$

$$G_N(X)[Z] := XX^\top Z, \qquad Z \in \mathbb{R}^{n\times p}. \tag{7.28}$$

With these definitions, we have indeed $G_T(X)\mathrm{Proj}_X = \widetilde{G_T}(X)\mathrm{Proj}_X$ and $G_N(X)\mathrm{Proj}_X^\perp = \widetilde{G_N}(X)\mathrm{Proj}_X^\perp$.

**Proposition 7.11.** *The operators $G_T(X)$ and $G_N(X)$ of (7.27) and (7.28) are symmetric positive operators on $\mathcal{E}$, definite respectively on $\mathcal{E}$ and $\mathrm{N}_X\mathrm{St}_{X^\top X}$. Formula (7.12) defines thus the associated metric*

$$
\begin{aligned}
g(\xi,\zeta) &= \langle \xi, \mathrm{Proj}_X^{*,\mathcal{E}} G_T(X) \mathrm{Proj}_X \zeta\rangle + \langle \xi, (\mathrm{Proj}_X^\perp)^{*,\mathcal{E}} G_N(X) \mathrm{Proj}_X^\perp \zeta\rangle \\
&= \langle \xi, (XX^\top + \mathrm{I}_n - X(X^\top X)^{-1}X^\top)\zeta\rangle.
\end{aligned} \tag{7.29}
$$

*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

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

□

Since $\widetilde{G_T}^{-1}$ and $\widetilde{G_N}^{-1}$ are explicit, we can provide explicit formulas for the tangent step $d_T(X) = -\mathrm{grad}^g_{\mathrm{St}_{X^\top X}} f(X)$ and the normal step $d_N(X) = -\nabla_g\psi(X)$ (corresponding to $H(X) = \mathrm{D}c(X)\mathrm{D}c(X)^{*,g}$) in the metric $g$ of (7.29).

**Proposition 7.12.** *With the metric $g$ defined in (7.29), the tangent and normal directions of the landing algorithm (1.3) read*

$$
d_T(X) = -X(X^\top X)^{-1}\,\mathrm{skew}((X^\top X)^{-1}X^\top\nabla_{\mathcal{E}} f(X)) - (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\nabla_{\mathcal{E}} f(X), \tag{7.30}
$$

$$
d_N(X) = -\frac{1}{2}X(\mathrm{I}_p - (X^\top X)^{-1}). \tag{7.31}
$$

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

$$
\begin{array}{rccc}
\widetilde{G_T}(X)^{-1} : & \mathrm{N}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} & \longrightarrow & \mathrm{T}_X\mathrm{St}_{X^\top X} \\
& X\Omega + \Delta & \longmapsto & X(X^\top X)^{-1}\Omega + \Delta,
\end{array} \tag{7.32}
$$

$$
\begin{array}{rccc}
\widetilde{G_N}(X)^{-1} : & \mathrm{T}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X} & \longrightarrow & \mathrm{N}_X\mathrm{St}_{X^\top X} \\
& XS & \longmapsto & X(X^\top X)^{-1}S.
\end{array} \tag{7.33}
$$

Formula (7.30) follow from (7.23), and combining (7.33) with the item (ii) of proposition 7.10 yields

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

□

**Remark 7.1.** *Here, $d_N(X)$ is both the Riemannian gradient of the penalty function $\psi(X)$ and the 'Newton-like' direction (7.22) with $H(X) = \mathrm{Id}_{\mathcal{E}}$. This happens, according to proposition 4.4, when $\widetilde{G_N}(X) = \mathrm{D}c(X)^{*,\mathcal{E}}\mathrm{D}c(X)$ as operators from $\mathrm{N}_X\mathrm{St}_{X^\top X}$ to $\mathrm{T}_X\mathrm{St}^{\perp,\mathcal{E}}_{X^\top X}$. This is incidentally indeed the case here: for any $S \in \mathrm{Sym}(p)$,*

$$
\begin{aligned}
\mathrm{D}c(X)^{*,\mathcal{E}}\mathrm{D}c(X)[X(X^\top X)^{-1}S] &= \mathrm{D}c(X)^{*,\mathcal{E}}[\mathrm{sym}(X^\top X(X^\top X)^{-1}S)] = \mathrm{D}c(X)^{*,\mathcal{E}}[S] \\
&= XS = \widetilde{G_N}(X)[X(X^\top X)^{-1}S].
\end{aligned}
$$

7.2.3. *Construction of the tangent and normal steps based on the $\beta$-metric*

It turns out that the projectors $\mathrm{Proj}_X$ and $\mathrm{Proj}_X^{\perp}$ of (7.13) and (7.14) are exactly the orthogonal projectors for the family of $\beta$-metric $g^\beta$ considered in Goyens et al. (2026):

$$g^\beta(\xi,\zeta) = \left\langle \left(\mathrm{I} - (1-\beta)X(X^\top X)^{-1}X^\top\right)\zeta(X^\top X)^{-1}, \xi \right\rangle, \tag{7.34}$$

see Goyens et al. (2026), Proposition 4. The $\beta$-metric $g^\beta$ is the pull-back of a generalization of the canonical metric Edelman et al. (1998) on the Stiefel manifold St to the layered manifold $\mathrm{St}_{X^\top X}$. We observe, however, that $g^\beta(\xi,\zeta)$ does not coincide with the metric $g$ found in in (7.29), despite both generate the same family of normal spaces $\mathrm{N}_X\mathrm{St}_{X^\top X}$.

In the next proposition, we show that the $\beta$-metric (7.34) is a special case of (7.12) for particular choices of $G_T(X)$ and $G_N(X)$. Then, we verify that the operators $\widetilde{G_T}$ and $\widetilde{G_N}$ have explicit inverses, which is the reason why this family of metrics leads to explicit formulas for the tangent and normal terms $d_T(X)$ and $d_N(X)$.

**Proposition 7.13.** *Let $\beta > 0$. The $\beta$-metric (7.34) is the metric (7.12) with $G_T(X) : \mathcal{E} \to \mathcal{E}$ and $G_N(X) : \mathcal{E} \to \mathcal{E}$ being the symmetric operators (with respect to the Euclidean inner product) defined by*

$$G_T(X)[\xi] := ((\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\xi + \beta X(X^\top X)^{-1}X^\top\xi)(X^\top X)^{-1}, \tag{7.35}$$

$$G_N(X)[\xi] := \beta X(X^\top X)^{-1}X^\top\xi(X^\top X)^{-1}. \tag{7.36}$$

*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 $X$. 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. We have thus retrieved the fact that the tangent space $\mathrm{T}_X\mathrm{St}_{X^\top X}$ and the normal space $\mathrm{N}_X\mathrm{St}_{X^\top X}$ of (7.15) are orthogonal for the metric $g^\beta$. This implies in particular 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

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

Thus, $g^\beta$ is the metric $g$ of (7.12) for $G_T$ and $G_N$ defined by (7.35) and (7.36). The positive definiteness of $G_T(X)$ and $G_N(X)$ are visible from

$$\begin{aligned} &\langle G_T(X)[\xi], \xi\rangle = \beta\|\Pi_X\xi(X^\top X)^{-\frac{1}{2}}\|^2 + \|(\mathrm{I}_n - \Pi_X)\xi(X^\top X)^{-\frac{1}{2}}\|^2, \\ &\forall S \in \mathrm{Sym}(p),\ \langle G_N(X)[X(X^\top X)^{-1}S], X(X^\top X)^{-1}S\rangle \\ &\qquad = \beta\langle X(X^\top X)^{-1}S(X^\top X)^{-1}, X(X^\top X)^{-1}S\rangle \\ &\qquad = \beta\|(X^\top X)^{-\frac{1}{2}}S(X^\top X)^{-\frac{1}{2}}\|^2. \end{aligned}$$

□

**Proposition 7.14.** *The inverse of the operators*

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

*associated to the $\beta$-metric $g^\beta$ are explicit and given by*

$$\widetilde{G_T}(X)^{-1} : \begin{array}{rcl} \mathrm{N}_X\mathrm{St}_{X^\top X}^{\perp,\mathcal{E}} & \longrightarrow & \mathrm{T}_X\mathrm{St}_{X^\top X} \\ X\Omega + \Delta & \longmapsto & \frac{1}{\beta}X\Omega X^\top X + \Delta X^\top X, \end{array} \tag{7.37}$$

$$\widetilde{G_N}(X)^{-1} : \begin{array}{rcl} \mathrm{T}_X\mathrm{St}_{X^\top X}^{\perp,\mathcal{E}} & \longrightarrow & \mathrm{N}_X\mathrm{St}_{X^\top X} \\ XS & \longmapsto & \frac{1}{\beta}XS(X^\top X), \end{array} \tag{7.38}$$

*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 (1.3) with $H(X) = \mathrm{Dc}(X)\mathrm{Dc}(X)^{*,g^\beta}$ read*

$$\begin{aligned} d_T(X) &= -\mathrm{grad}^{g^\beta}_{\mathrm{St}_{X^\top X}} f(X) \\ &= -\frac{1}{\beta}X\ \mathrm{skew}\big((X^\top X)^{-1}X^\top\nabla_{\mathcal{E}}f(X)\big)\, X^\top X - (\mathrm{I}_n - X(X^\top X)^{-1}X^\top)\nabla_{\mathcal{E}}f(X)\, X^\top X, \end{aligned} \tag{7.39}$$

$$d_N(X) = -\nabla_{g^\beta}\psi(X) = -\frac{1}{2\beta}X(X^\top X - \mathrm{I}_p)X^\top X. \tag{7.40}$$

*Proof.* Using (7.35) and (7.36), we see that $\widetilde{G_T}(X)$ and $\widetilde{G_N}(X)$ are the operators

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

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

whose inverses are obviously given by (7.37) and (7.38). Inserting these formulas into (7.21) and item (ii) of proposition 7.10 yields the result. □

The reader may verify that (7.39) coincide with the formula derived in Goyens et al. (2026). In this latter work, the normal term was considered to be the Euclidean gradient $d_N(X) = -\nabla_{\mathcal{E}}\psi(X)$, which corresponds to (7.22) with $H(X) = \mathrm{Dc}(X)\mathrm{Dc}(X)^{*,\mathcal{E}}$ (item (iii) of proposition 7.10).

## 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*. Princeton University Press.

Absil, P.-A. and Malick, J. (2012). Projection-like Retractions on Matrix Manifolds. *SIAM Journal on Optimization*, 22(1):135–158.

Absil, P.-A., Trumpf, J., Mahony, R., and Andrews, B. (2009). All roads lead to Newton: Feasible second-order methods for equality-constrained optimization. *https://sites.uclouvain.be/absil/2009.024*.

Allaire, G., Dapogny, C., and Jouve, F. (2021). Chapter 1 - Shape and topology optimization. In Bonito, A. and Nochetto, R. H., editors, *Handbook of Numerical Analysis*, volume 22 of *Geometric Partial Differential Equations - Part II*, pages 1–132. Elsevier.

Bai, Y. and Mei, S. (2018). Analysis of sequential quadratic programming through the lens of riemannian optimization. *arXiv preprint arXiv:1805.08756*.

Barbarosie, C., Toader, A.-M., and Lopes, S. (2020). A gradient-type algorithm for constrained optimization with application to microstructure optimization. *Discrete and Continuous Dynamical Systems - B*, 25(5):1729–1755.

Berahas, A. S., Curtis, F. E., O'Neill, M. J., and Robinson, D. P. (2024). A Stochastic Sequential Quadratic Optimization Algorithm for Nonlinear-Equality-Constrained Optimization with Rank-Deficient Jacobians. *Mathematics of Operations Research*, 49(4):2212–2248.

Berahas, A. S., Curtis, F. E., Robinson, D., and Zhou, B. (2021). Sequential quadratic optimization for nonlinear equality constrained stochastic optimization. *SIAM Journal on Optimization*, 31(2):1352–1379.

Biros, G. and Ghattas, O. (2005). Parallel Lagrange–Newton–Krylov–Schur Methods for PDE-Constrained Optimization. Part I: The Krylov–Schur Solver. *SIAM Journal on Scientific Computing*, 27(2):687–713.

Boggs, P. T. and Tolle, J. W. (1995). Sequential Quadratic Programming. *Acta Numerica*, 4:1–51.

Booker, A. and Ong, C.-y. (1971). Multiple-constraint adaptive filtering. *Geophysics*, 36(3):498–509.

Conn, A. R., Gould, N. I. M., and Toint, P. (1991). A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds. *SIAM Journal on Numerical Analysis*, 28(2):545–572.

Curtis, F. E., O'Neill, M. J., and Robinson, D. P. (2024). Worst-case complexity of an sqp method for nonlinear equality constrained stochastic optimization. *Mathematical Programming*, 205(1):431–483.

Dietrich, A., Ott, C., and Albu-Schäffer, A. (2015). An overview of null space projections for redundant, torque-controlled robots. *The International Journal of Robotics Research*, 34(11):1385–1400.

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.

Evtushenko, Y. G. and Zhadan, V. G. (1994). Stable barrier-projection and barrier-Newton methods in linear programming. *Computational Optimization and Applications*, 3(4):289–303.

Fang, Y., Na, S., Mahoney, M. W., and Kolar, M. (2024). Fully Stochastic Trust-Region Sequential Quadratic Programming for Equality-Constrained Optimization Problems. *SIAM Journal on Optimization*, 34(2):2007–2037.

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. (2020a). Null space gradient flows for constrained optimization with applications to shape optimization. *ESAIM: Control, Optimisation and Calculus of Variations*, page 90.

Feppon, F., Allaire, G., Dapogny, C., and Jolivet, P. (2020b). Topology optimization of thermal fluid–structure systems using body-fitted meshes and parallel computing. *Journal of Computational Physics*, 417:109574.

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.

Fletcher, R., Gould, N. I. M., Leyffer, S., Toint, P. L., and Wächter, A. (2002). Global Convergence of a Trust-Region SQP-Filter Algorithm for General Nonlinear Programming. *SIAM Journal on Optimization*, 13(3):635–659.

Frost, O. (1972). An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 60(8):926–935.

Gangi, A. F. and Byun, B. S. (1976). The corrective gradient projection method and related algorithms applied to seismic array processing. *Geophysics*, 41(5):970–984.

Gao, B., Vary, S., Ablin, P., and Absil, P. A. (2022). Optimization flows landing on the Stiefel manifold. *IFAC-PapersOnLine*, 55(30):25–30.

Gharaei, A., Amjadian, A., Amjadian, A., Shavandi, A., Hashemi, A., Taher, M., and Mohamadi, N. (2023). An integrated lot-sizing policy for the inventory management of constrained multi-level supply chains: Null-space method. *International Journal of Systems Science: Operations & Logistics*, 10(1):2083254.

Gill, P., Wong, E., Murray, W., and Saunders, M. (2015). User's Guide for SNOPT Version 7.4: Software for Large-Scale Nonlinear Programming.

Gill, P. E., Murray, W., and Saunders, M. A. (2002). SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization. *SIAM Journal on Optimization*, 12(4):979–1006.

Gill, P. E. and Wong, E. (2012). Sequential Quadratic Programming Methods. In Lee, J. and Leyffer, S., editors, *Mixed Integer Nonlinear Programming*, pages 147–224, New York, NY. Springer.

Goyens, F., Absil, P.-A., and Feppon, F. (2026). Geometric Design of the Tangent Term in Landing Algorithms for Orthogonality Constraints. In Nielsen, F. and Barbaresco, F., editors, *Geometric Science of Information*, pages 133–141, Cham. Springer Nature Switzerland.

Goyens, F., Eftekhari, A., and Boumal, N. (2024). Computing Second-Order Points Under Equality Constraints: Revisiting Fletcher's Augmented Lagrangian. *Journal of Optimization Theory and Applications*, 201(3):1198–1228.

Henrot, A. and Pierre, M. (2018). *Shape Variation and Optimization*, volume 28 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich.

Hu, X., Cheng, Y., Yang, D., Xu, Z., Yuan, Z., Yu, J., Xu, C., Jiang, Z., and Zhou, S. (2025). Ostquant: Refining large language model quantization with orthogonal and scaling transformations for better distribution fitting. *arXiv preprint arXiv:2501.13987*.

Hüper, K., Markina, I., and Leite, F. S. (2021). A Lagrangian approach to extremal curves on Stiefel manifolds. *Journal of Geometric Mechanics (JGM)*.

Jongen, H. T. and Stein, O. (2003). On the Complexity of Equalizing Inequalities. *Journal of Global Optimization*, 27(4):367–374.

Jongen, H. Th. and Stein, O. (2004). Constrained Global Optimization: Adaptive Gradient Flows. In Floudas, C. A. and Pardalos, P., editors, *Frontiers in Global Optimization*, pages 223–236, Boston, MA. Springer US.

Li, Z.-Y. and Zhou, B. (2018). Spectral decomposition based solutions to the matrix equation. *IET Control Theory & Applications*, 12(1):119–128.

Liu, X. and Yuan, Y. (2010). A null-space primal-dual interior-point algorithm for nonlinear optimization with nice convergence properties. *Mathematical programming*, 125(1):163–193.

Luenberger, D. G. (1972). The Gradient Projection Method Along Geodesics. *Management Science*, 18(11):620–631.

Mataigne, S., Zimmermann, R., and Miolane, N. (2025). An Efficient Algorithm for the Riemannian Logarithm on the Stiefel Manifold for a Family of Riemannian Metrics. *SIAM Journal on Matrix Analysis and Applications*, 46(2):879–905.

Miller, S. A. and Malick, J. (2005). Newton methods for nonsmooth convex minimization: Connections among -Lagrangian, Riemannian Newton and SQP methods. *Mathematical Programming*, 104(2):609–633.

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

Nie, P.-y. (2004). A null space method for solving system of equations. *Applied Mathematics and computation*, 149(1):215–226.

Nocedal, J. and Overton, M. L. (1985). Projected Hessian Updating Algorithms for Nonlinearly Constrained Optimization. *SIAM Journal on Numerical Analysis*, 22(5):821–850.

Nocedal, J., Öztoprak, F., and Waltz, R. A. (2014). An interior point method for nonlinear programming with infeasibility detection capabilities. *Optimization Methods and Software*, 29(4):837–854.

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

Obara, M., Okuno, T., and Takeda, A. (2022). Sequential Quadratic Optimization for Nonlinear Optimization Problems on Riemannian Manifolds. *SIAM Journal on Optimization*, 32(2):822–853.

Rosen, J. B. (1960). The Gradient Projection Method for Nonlinear Programming. Part I. Linear Constraints. *Journal of the Society for Industrial and Applied Mathematics*, 8(1):181–217.

Rosen, J. B. (1961). The Gradient Projection Method for Nonlinear Programming. Part II. Nonlinear Constraints. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):514–532.

Schechtman, S., Tiapkin, D., Muehlebach, M., and Moulines, É. (2023). Orthogonal Directions Constrained Gradient Method: From non-linear equality constraints to Stiefel manifold. In *Proceedings of Thirty Sixth Conference on Learning Theory*, pages 1228–1258. PMLR.

Schropp, J. and Singer, I. (2000). A dynamical systems approach to constrained minimization. *Numerical Functional Analysis and Optimization*, 21(3-4):537–551.

Shi, Q. and Wang, X. (2025). Adaptive directional decomposition methods for nonconvex constrained optimization.

Shikhman, V. and Stein, O. (2009). Constrained Optimization: Projected Gradient Flows. *Journal of Optimization Theory and Applications*, 140(1):117–130.

Song, Y., Li, P., Gao, B., and Yuan, K. (2025). Distributed retraction-free and communication-efficient optimization on the stiefel manifold. *arXiv preprint arXiv:2506.02879*.

Tanabe, K. (1974). An algorithm for constrained maximization in nonlinear programming. *Journal of the Operations Research Society of Japan*, 17:184–201.

Tanabe, K. (1980). A geometric method in nonlinear programming. *Journal of Optimization Theory and Applications*, 30(2):181–210.

Tapia, R. A. (1977). Diagonalized multiplier methods and quasi-Newton methods for constrained optimization. *Journal of Optimization Theory and Applications*, 22(2):135–194.

Vary, S., Ablin, P., Gao, B., and Absil, P.-A. (2024). Optimization without Retraction on the Random Generalized Stiefel Manifold. In *Proceedings of the 41st International Conference on Machine Learning*, pages 49226–49248. PMLR.

Wächter, A. and Biegler, L. T. (2005). Line Search Filter Methods for Nonlinear Programming: Motivation and Global Convergence. *SIAM Journal on Optimization*, 16(1):1–31.

Wegert, Z. J., Roberts, A. P., and Challis, V. J. (2023). A Hilbertian projection method for constrained level set-based topology optimisation. *Structural and Multidisciplinary Optimization*, 66(9):204.

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

Yuan, Y.-x. (2001). A null space algorithm for constrained optimization. *Advances in Scientific Computing, Science Press, Beijing*, pages 210–218.

Yulin, M. and Xiaoming, W. (2004). A level set method for structural topology optimization with multi-constraints and multi-materials. *Acta Mechanica Sinica*, 20(5):507–518.

Zhang, Y., Hu, J., Cui, J., Lin, L., Wen, Z., and Li, Q. (2024). Retraction-free optimization over the Stiefel manifold with application to the LoRA fine-tuning.