All roads lead to Newton: Feasible second-order methods for equality-constrained optimization *

P.-A. Absil† Jochen Trumpf‡ Robert Mahony‡ Ben Andrews§

August 31, 2009

Abstract

This paper considers the connection between the intrinsic Riemannian Newton method and other more classically inspired optimization algorithms for equality-constrained optimization problems. We consider the feasibly-projected sequential quadratic programming (FP-SQP) method and show that it yields the same update step as the Riemannian Newton, subject to a minor assumption on the choice of multiplier vector. We also consider Newton update steps computed in various 'natural' local coordinate systems on the constraint manifold and find simple conditions that guarantee that the update step is the Riemannian Newton update. In particular, we show that this is the case for projective local coordinates, one of the most natural choices that have been proposed in the literature. Finally we consider the case where the full constraints are approximated to simplify the computation of the update step. We show that if this approximation is good at least to second-order then the resulting update step is the Riemannian Newton update. The conclusion of this study is that the intrinsic Riemannian Newton algorithm is the archetypal feasible second order update for non-degenerate equality constrained optimisation problems.

Key words. Feasibly-projected sequential quadratic programming (FP-SQP); equality-constrained optimization; Riemannian manifold; Riemannian Newton method; sequential Newton method; retraction; osculating paraboloid; second-order correction; second fundamental form; Weingarten map

*This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. This research was supported by the Australian Research Council through the Discovery grant DP0987411.

†Département d'ingénierie mathématique, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (`http://www.inma.ucl.ac.be/~absil/`).

‡School of Engineering, Australian National University, Australia. `{firstname}.{lastname}@anu.edu.au`.

§Center for Mathematical Analysis, IAS, Australian National University, ACT, 0200, Australia (`Ben.Andrews@anu.edu.au`).

1 Introduction

We consider the problem of minimizing an objective function $f(x)$, $x \in \mathbb{R}^n$, subject to equality constraints $\Phi(x) = 0$. Recall that a point $x \in \mathbb{R}^n$ is termed *feasible* if $\Phi(x) = 0$, i.e., x belongs to the *feasible set* $\mathcal{M} := \Phi^{-1}(0)$. A point that is not feasible is termed *infeasible*.

For an arbitrary constraint function Φ, the problem of finding a *feasible* point close to a present *infeasible* estimate, is often of similar complexity to solving the full optimisation problem. Thus, most classical optimization methods solve the two parts of the problem concurrently, that is, they generate a sequence of infeasible estimates that both converge to the feasible set and to the function minimum at the same time; Sequential Quadratic Programming (SQP) being the archetypal example of such an algorithm.

However, in a number of important examples, notably when $\mathcal{M}$ is a "well-behaved" submanifold [AMS08], the cost of forcing iterates onto the feasible set is low and maintaining feasible iterates offers several advantages. There is no need for composite-step approaches that attempt to reduce infeasibility while preserving a decrease of the objective function, the objective function itself can be used as a merit function, and only the structure of the objective function on or close to the constraint set will be observed by the algorithm. Moreover, in case of early termination, the algorithm returns a feasible suboptimal point, which is often preferable to an infeasible suboptimum. Another motivation for enforcing feasibility arises when the objective function is smooth when restricted to a certain submanifold of the optimization domain and nonsmooth away from the manifold. This situation underlies the theory of $\mathcal{U}$-Lagrangians, and the related $\mathcal{VU}$-decompositions and fast tracks [LOS00, MS00], as well as the theory of partly smooth functions [Lew02], both of which coincide in the convex case [MM05, Th. 2.9].

Arguably the most natural approach to ensuring a feasible algorithm, at least when the constraint condition defines a submanifold, is obtained by undertaking the optimisation in local coordinates directly on the feasible set. One way of obtaining local coordinates, studied in the 1960s, is the method of direct elimination, which consists in dividing the variables into two groups such that the variables of the one are expressed (explicitly or implicitly) as a function of the others; see [Pol76, GL76] for an overview of the related literature. A Newton method based on direct partitioning was proposed and analyzed by Gabay and Luenberger [GL76]. The Riemannian Newton algorithm grew out of this work in an attempt to find a unifying framework for computation of the update step [Gab82, Shu86, Udr94, Smi94]. In the 1990s, the nature of the constraint sets were considered in more detail and a number of important examples with significant Lie group and homogeneous space structures were identified [HM94], and the explicit form of the Newton algorithm for several examples such as the Grassmann manifold were computed [EAS98]. More recent work [ADM$^+$02, HT04, AMS08] has developed a coherent theory of intrinsic Newton methods on structured manifolds with a particular emphasis on practical issues of implementation. The approach has led to a number of highly competitive algorithms for problems where the constraint set has suitable structure, e.g., algorithms for the extreme eigenvalue problem [ABG06, BAG08], rank reduction of correlation matrices [GP07], semidefinite programming [JBAS08], and stereo vision processing [HHLM07].

Another recent development concerning feasible second-order methods proceeds from the SQP approach. The classical SQP is not a feasible method: even if $\Phi(x_k) = 0$, it is likely that $\Phi(x_{k+1}) \neq 0$; see, e.g., [NW06, Ch. 18]. However, in recent work, Wright and Tenny [WT04] proposed a feasibly-projected SQP (FP-SQP) method, that generates a sequence of feasible iterates by solving a trust-region SQP subproblem at each iteration and perturbing the resulting step to retain feasibility. "By retaining feasibility, the algorithm avoids several complications of other trust-region SQP approaches: the objective function can be used as a merit function, and the SQP subproblems are feasible for all choices of the trust-region radius" [WT04]. This can be highly advantageous in situations where the computational cost of ensuring feasibility is negligible compared to the overall cost of the optimization algorithm.

Finally, we mention that the concepts of $\mathcal{U}$-Lagrangian and partly smooth functions led to several Newton-like algorithms whose iterates are constrained to a submanifold $\mathcal{M}$ such that the restriction $f_{|\mathcal{M}}$ is smooth. These algorithms are unified in [DHM06] under a common two-step, predictor-corrector form, and connections with SQP and Riemannian Newton are studied in [MM05].

In this paper, we investigate the similarities between feasible second-order optimisation algorithms that follow from an *extrinsic viewpoint*, that is an algorithm posed on the overarching space $\mathbb{R}^n$ with additional constraints embedded in this space, SQP being the archetypal example, as compared to algorithms that follow from an *intrinsic viewpoint*, that is algorithms posed explicitly on the constraint set, the Riemannian Newton being the archetypal example. The reason for the title, "All roads lead to Newton", is that this paper extends and reinforces the message, already implicitly present in the taxonomy constructed in [EAS98], that all feasible second-order methods for equality-constrained optimization are an intrinsic Newton algorithm, or at the very least, only vary from the Newton algorithm in non-substantive aspects.[1] It is important to note that our conclusions only hold on the intersection of the problem spaces for the different approaches, that is, where an intrinsic Newton is derived for an embedded submanifold (such algorithms can be applied to a wider class of problems [AMS08]) and where an extrinsic-viewpoint algorithm is constrained to have feasible iterates at every step.

In detail, we consider the Feasibly-Projected Sequential Quadratic Program (FP-SQP) [WT04] and show that (for a least-squares choice of the Lagrange multiplier λ_k and ignoring the trust region and other modifications) the algorithm is "intrinsic" to the feasible set and can be interpreted as a Riemannian Newton method. We also consider the intrinsic algorithm obtained by applying the Newton method in local coordinates on $\mathcal{M}$. We show that if the local coordinates chosen have a simple second-order rigidity condition at the iterate point then one recovers the Riemannian Newton as expected. A simple proof that the most natural choice of local coordinates, the projective coordinates (those coordinates obtained by projecting the tangent plane of a submanifold back onto the submanifold) satisfy the required rigidity conditions is provided. It is well known that the Newton method only uses second-order information from the cost. We demonstrate that it also uses second order information on the constraint set. This leads to a discussion of alternatives where the constraint set $\mathcal{M}$ is replaced by an approximation $\widetilde{\mathcal{M}}_{x_k}$ at each iteration. We propose ways of constructing approximations $\widetilde{\mathcal{M}}_x$ that admit a simple expression and that agree to second order with $\mathcal{M}$

[1] In fact, the proverb "all roads lead to Rome" was never itself a Roman proverb and the earliest reference we know of is [dL75] "mille viae ducunt homines per saecula Romam" or "a thousand roads lead men forever to Rome". In his discourse on the Astrolab [Cha91] Chaucer states "diverse pathes leden diverse folk the righte way to Rome," and the saying was popularised by Jean de La Fontaine ("Le Juge arbitre, l'Hospitalier, et le Solitaire", Fables, livre XII) "Tous chemins vont à Rome; ainsi nos concurrents crurent pouvoir choisir des sentiers différents." The historic versions of the quote are as well adapted to our message as the modern version; using our analogy they say roughly, however you derive a second-order feasible method for an equality constrained optimisation problem, and even if you believe you are doing something novel, you will still end up with an algorithm that is in essence an intrinsic Newton update.

at x.

The reader familiar with [MM05] will find an important overlap between the two works. This paper, however, departs from [MM05] in several ways. The convexity assumption is dropped. The Riemannian Newton method is considered in the general retraction-based setting of [ADM+02, §5], instead of being limited to the three tangential, exponential, and projection parameterizations. Proposition 4.2 shows that Sequential Newton [MM05, Alg. 1.8] yields the same Newton vector for all second-order retractions. The link between Riemannian Newton and SQP is extended to the feasibly-projected SQP of [WT04] and concerns the whole iteration rather than just the Newton vector. Equation 3.6 in [MM05] is used to obtain expression (34) showing that $\mathcal{M}$ enters the Riemannian Hessian through its tangent space and the second derivative of its tangential parameterization only. The extrinsic expression of the Riemannian Hessian in [MM05, Th. 3.4] is rewritten in (41) in terms of the derivative of the projector, which leads to rich interpretations in terms of the curvature of the manifold (see Section 6.1).

Even though some methods discussed in this paper stem from differential-geometric concepts, we do not assume any background in differential geometry beyond a basic knowledge in real analysis in the main body of the paper. In particular, the formulation of the methods solely relies on matrix computation and calculus. For completeness, we also provide a final section to the paper that provides the full geometric picture in the modern language of differential geometry. In particular, we show that the intrinsic Newton algorithm depends on, and only on, a second order approximation of the cost and the Riemannian curvature of the constraint set in its update.

Section 2 introduces notation and assumptions, Section 3 deals with SQP, Section 4 with intrinsically-built Newton methods, and Section 5 with approximations of $\mathcal{M}$. Section 6 studies the role of the curvature of $\mathcal{M}$ in the Riemannian Hessian, and the paper ends with concluding remarks.

2 Problem definition, concepts, and notation

In this section, we present the optimization problem considered and we explain in plain calculus language how it relates to optimization on manifolds.

We consider the general equality-constrained optimization problem

$$\min f(x) \text{ subject to } \Phi(x) = 0, \tag{1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ are C^2.

The Euclidean gradient of f at x will be denoted by $\nabla f(x) = \begin{bmatrix}\partial_1 f(x) & \dots & \partial_n f(x)\end{bmatrix}^T$ and its Hessian matrix by $\nabla^2 f(x)$. For a differentiable matrix-valued function F, we let $\mathrm{D}F(x)[z]$ denote the directional derivative of F at x along z. If F is vector valued, then $\mathrm{D}F(x)$ also denotes the Jacobian matrix of F at x and $\mathrm{D}F(x)z$ stands for $\mathrm{D}F(x)[z]$. We also find it convenient to use the notation $\nabla\Phi(x) = \mathrm{D}\Phi(x)^T = \begin{bmatrix}\nabla\Phi_1(x) & \dots & \nabla\Phi_m(x)\end{bmatrix}$.

For the purpose of relating SQP to Newton on manifolds, we assume that the linear independence constraint qualification (LICQ) holds at all points of $\Phi^{-1}(0)$, that is the columns of $\nabla\Phi(x)$ are linearly independent for all $x \in \Phi^{-1}(0)$; see, e.g., [BGLS03, §11.3]. The value 0 is then said to be a *regular value* of Φ, and a well-known theorem in differential geometry [AMR88, §3.5.3] ensures that

$$\mathcal{M} := \Phi^{-1}(0) = \{x \in \mathbb{R}^n : \Phi(x) = 0\}$$

is a d-dimensional *submanifold* (also called *embedded submanifold*) of $\mathbb{R}^n$, where

$$d = n - m.$$

This means that the subset $\mathcal{M}$ of $\mathbb{R}^n$ is locally a d-slice, that is, around every point $x \in \mathcal{M}$, there is an open set U in $\mathbb{R}^n$ and a coordinate system on U such that $\mathcal{M} \cap U$ corresponds to the points where all but the d first coordinates vanish [AMR88, §3.2.1]. The tangent space to $\mathcal{M}$ at a point x is $\ker(\nabla\Phi(x)^T)$ and is denoted by $T_x\mathcal{M}$. Observe that $T_x\mathcal{M}$ is a linear subspace of $\mathbb{R}^n$, hence its elements are naturally identified with elements of $\mathbb{R}^n$. We let

$$\mathcal{P}_x = I - \nabla\Phi(x)(\nabla\Phi(x)^T\nabla\Phi(x))^{-1}\nabla\Phi(x)^T \tag{2}$$

denote the orthogonal projector onto $T_x\mathcal{M}$. Observe that $\mathcal{P}$ is a smooth matrix-valued function. Note also that the expression (2) still makes sense when x is not in $\mathcal{M}$, provided that the columns of $\nabla\Phi(x)$ are linearly independent.

The manifold $\mathcal{M}$ is naturally a *Riemannian* submanifold of $\mathbb{R}^n$. This means that the tangent spaces to $\mathcal{M}$ are naturally equipped with an inner product that varies smoothly with the base point of the tangent space:

$$\langle z_1, z_2\rangle_x = z_1^T z_2, \quad z_1, z_2 \in T_x\mathcal{M}. \tag{3}$$

Since $\mathcal{M}$ is a Riemannian submanifold, there is a well-defined notion of gradient and Hessian of f in the sense of $\mathcal{M}$, which we denote by $\operatorname{grad} f_{|\mathcal{M}}(x)$ and $\operatorname{Hess} f_{|\mathcal{M}}(x)$; see Section 4.1 or [AMS08] for details.

An *iterative optimization method* A on $\mathcal{M}$ takes an objective function f to an iteration map $A_f : \mathcal{M} \to \mathcal{M}$. Given an objective function f and an initial point $x_0 \in \mathcal{M}$, the purpose of the method is to create a sequence on $\mathcal{M}$ defined by $x_{k+1} := A_f(x_k)$. We say that the iterative optimization method A is *intrinsic* to $\mathcal{M}$ if $f_{|\mathcal{M}} = \hat{f}_{|\mathcal{M}}$ implies that $A_f = A_{\hat{f}}$. In other words, the iteration map A_f does not change if f is modified outside $\mathcal{M}$. As we will see, an intrinsic method may have *extrinsic implementations*, i.e., implementations that use subroutines that are not intrinsic. Note also that if we are only given $f_{|\mathcal{M}}$, we can still use extrinsic implementations of intrinsic methods, provided that we have a routine that takes $f_{|\mathcal{M}}$ as input and extends it outside $\mathcal{M}$. How we extend it, apart from guaranteeing suitable smoothness, does not matter for an intrinsic method. Several of the algorithms studied in this paper have this property.

3 Sequential Quadratic Programming

Sequential Quadratic Programming (SQP) for problem (1) can be introduced as follows (see, e.g., [NW06, BGLS03]). The KKT necessary conditions for optimality are

$$\begin{cases} \nabla f(x) + \nabla\Phi(x)\lambda = 0 \\ \Phi(x) = 0, \end{cases} \tag{4}$$

$\lambda \in \mathbb{R}^m$. Points x^* for which there exists λ^* such that (4) is satisfied are termed *stationary points*. Linearizing the KKT system (4) at (x, λ) and denoting by $(z, \lambda_+ - \lambda)$ the variation of the variables x and λ respectively, we obtain the system

$$\left(\nabla^2 f(x) + \sum_{i=1}^{m} \nabla^2 \Phi_i(x)\lambda_i\right) z + \nabla f(x) + \nabla\Phi(x)\lambda_+ = 0 \tag{5a}$$

$$\nabla\Phi(x)^T z + \Phi(x) = 0. \tag{5b}$$

Following [AT07], we refer to (5) as the *Newton-KKT equations.* The pure Newton step for solving (4) is given by $(x, \lambda) \mapsto (x_+, \lambda_+) = (x + z, \lambda_+)$, where z and λ_+ solve (5).

The name *SQP* for methods based on (5) stem from the fact that the Newton-KKT equations can be viewed as the KKT equations of the quadratic problem (QP)

$$\begin{aligned} &\min_z z^T \nabla f(x) + \tfrac{1}{2} z^T (\nabla^2 f(x) + \sum_{i=1}^{m} \nabla^2 \Phi_i(x)\lambda_i) z \\ &\text{subject to } \nabla\Phi(x)^T z + \Phi(x) = 0, \end{aligned}$$

In preparation for the comparison with Newton's method on $\mathcal{M}$, observe that, when x is feasible (i.e., $x \in \mathcal{M}$), the Newton-KKT equations (5) are equivalent to

$$\mathcal{P}_x(\nabla^2 f(x) + \sum_{i=1}^{m} \nabla^2 \Phi_i(x)\lambda_i) z = -\mathcal{P}_x \nabla f(x) \tag{6a}$$

$$\nabla\Phi(x)^T z = 0 \tag{6b}$$

$$\nabla\Phi(x)^T (\nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 \Phi_i(x)) z + \nabla\Phi(x)^T \nabla\Phi(x)\lambda_+ = 0 \tag{6c}$$

where $\mathcal{P}_x$ denotes the orthogonal projector (2), and where we have used the identity $\mathcal{P}_x \nabla\Phi(x) = 0$. Equation (6) can be solved by obtaining z from (6a)-(6b) and then solving (6c) for λ_+.

Observe also that, given $x \in \mathcal{M}$, the LICQ condition ensures that the first equation in (4) has a unique least-squares solution

$$\lambda(x) = -(\nabla\Phi(x)^T \nabla\Phi(x))^{-1} \nabla\Phi(x)^T \nabla f(x). \tag{7}$$

3.1 FP-SQP

Most variants of SQP found in the literature are infeasible, i.e., the condition that the iterates x_k belong to $\mathcal{M}$ is not enforced. An exception is the feasibly-projected sequential quadratic programming (FP-SQP) algorithm of Wright and Tenny [WT04]. In that algorithm, a first step z is obtained by solving a problem of the form (5) at a feasible point $x \in \mathcal{M}$, under a trust-region constraint. Then, a perturbation $\tilde{z}$ of the step z is found with the following properties: *feasibility*, i.e.,

$$x + \tilde{z} \in \mathcal{M}, \tag{8}$$

and *asymptotic exactness*, that is, there is a continuous monotonically increasing function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ with $\phi(0) = 0$ such that

$$\|z - \tilde{z}\| \leq \phi(\|z\|_2)\|z\|_2. \tag{9}$$

(In [WT04], ϕ is required to be into $[0, 1/2]$, for the purpose of obtaining a global convergence result under the trust-region constraint. This restriction is not needed in the context of the

Algorithm 1 simple FS-SQP

Given a starting point $x_0 \in \mathcal{M}$;
for $k = 0, 1, 2, \ldots$ **do**
  Choose $\lambda_k \in \mathbb{R}^m$;
  Obtain z_k by solving the Newton-KKT equations (5);
  Seek $\tilde{z}_k$ satisfying feasibility (8) and asymptotic exactness (9);
  $x_{k+1} = x_k + \tilde{z}_k$.
end for

local convergence of the simplified SQP considered in this paper.) Note that (9) means that the mapping $\tilde{z} = z + o(z)$. One of the most natural ideas to obtain $\tilde{z}$ is to define $x + \tilde{z}$ as the projection of $x + z$ onto the feasible set $\mathcal{M}$ [WT04, Th. 2.2].

Algorithm 1 is statement of a simplified version of FP-SQP. The simplification comes by: (i) considering equality constraints only, because we are concerned with problem (1), (ii) ignoring the trust-region mechanism, in order to focus on the gist of the method for what concerns local convergence, (iii) assuming directly the classical choice for H as the Hessian of the Lagrangian [WT04, (4.7)].

4 Intrinsic Newton methods on the feasible set $\mathcal{M}$

The methods considered in this section proceed from an intrinsic viewpoint, in the following loosely defined sense: the way the method is constructed ensures that the iterates will depend on f only through the restriction $f_{|\mathcal{M}}$ of f to the manifold $\mathcal{M}$. Consequently, the methods are intrinsic, in the sense of Section 2. Recall however that certain implementations of the intrinsic methods can make use of extrinsic information; an example is the Riemannian Newton method when its Hessian is written as in (41) below. Note also that certain methods that proceed from an extrinsic viewpoint may turn out to be intrinsic, i.e., independent from modifications of f outside $\mathcal{M}$; see Section 5.

4.1 Riemannian Newton method on the feasible set $\mathcal{M}$

A goal of this section is to establish a link between FP-SQP and the Riemannian Newton method. Note that the section is accessible without a background in Riemannian geometry. In particular, the formulation of the Riemannian Newton method (Algorithm 2) solely relies on matrix computation and basic calculus. The interested reader can find the differential geometric foundations of Algorithm 2 in [AMS08].

Newton's method on Riemannian manifolds has a rich history, whose important milestones include [Gab82, Shu86, Smi94, EAS98, ADM$^+$02, HT04]. We rely on the formulation found in [ADM$^+$02, AMS08], and refer the reader to [AMS08, §6.6] for more detail on the historical developments of the Riemannian Newton algorithm. In this section, we present the fundamentals of the method in its application to the equality-constrained optimization problem (1).

The Riemannian Newton method, as formulated in [AMS08, §6.2], requires among other things the choice of an affine connection. Since here $\mathcal{M} = \Phi^{-1}(0)$ is a submanifold of the Euclidean space $\mathbb{R}^n$, a classical choice is the *Riemannian connection* (or *Levi-Civita connection*) [AMS08, §5.3.3] induced by the Riemannian submanifold structure of $\mathcal{M}$. This yields

Algorithm 2, for which some explanations are in order.

Algorithm 2 Riemannian Newton method on $\Phi^{-1}(0)$

1: Given a retraction R on $\Phi^{-1}(0)$ and a starting point $x_0 \in \Phi^{-1}(0)$;
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Solve, for the unknown $z_k \in \ker(\mathrm{D}\Phi(x_k))$, the Newton equation

$$\mathcal{P}_{x_k}\mathrm{D}(\mathcal{P}\nabla f)(x_k)[z_k] = -\mathcal{P}_{x_k}\nabla f(x_k), \tag{10}$$

where $\mathcal{P}_x$ denotes the orthogonal projector (2) onto $\ker(\mathrm{D}\Phi(x))$.
4: Set

$$x_{k+1} := R_{x_k}(z_k).$$

5: **end for**

Let

$$T\mathcal{M} = \bigcup_{x\in\mathcal{M}} (x, T_x\mathcal{M}) \tag{11}$$

denote the *tangent bundle* of $\mathcal{M}$. A *retraction* R [ADM+02] (or see [AMS08, §4.1] for the present equivalent formulation) is a smooth mapping from a neighborhood of $\mathcal{M}$ in $T\mathcal{M}$, taking its values in $\mathcal{M}$, and satisfying the following properties: Let R_x denote the restriction of R to $(x, T_x\mathcal{M}) \simeq T_x\mathcal{M}$, then

$$R_x(0_x) = x, \tag{12a}$$

where 0_x denotes the origin of the vector space $T_x\mathcal{M}$, and

$$\frac{\mathrm{d}}{\mathrm{d}t}R_x(t\xi_x)\bigg|_{t=0} = \xi_x, \text{ for all } \xi_x \in T_x\mathcal{M}. \tag{12b}$$

The notation

$$\mathrm{D}(\mathcal{P}\nabla f)(x_k)[z_k]$$

in (10) stands for the derivative at x_k along z_k of the function that maps x to $\mathcal{P}_x\nabla f(x)$. Because $\mathcal{P}_x\nabla f(x)$ is the gradient of f at x in the sense of the Riemannian submanifold $\mathcal{M}$ [AMS08, (3.37)], we adopt the notation

$$\operatorname{grad} f_{|\mathcal{M}}(x) := \mathcal{P}_x\nabla f(x). \tag{13}$$

For similar reasons [AMS08, (5.15)], we have

$$\begin{aligned}\operatorname{Hess} f_{|\mathcal{M}}(x) &:= \mathcal{P}_x\mathrm{D}\left(\operatorname{grad} f_{|\mathcal{M}}\right)(x) \\ &= \mathcal{P}_x\mathrm{D}(\mathcal{P}\nabla f)(x).\end{aligned} \tag{14}$$

The Riemannian Newton method (Algorithm 2) can thus be compactly written as the iteration $x \mapsto x_+$ computed as

$$\operatorname{Hess} f_{|\mathcal{M}}(x)z = -\operatorname{grad} f_{|\mathcal{M}}(x), \quad z \in T_x\mathcal{M} \tag{15a}$$

$$x_+ = R_x(z). \tag{15b}$$

Observe that the only freedom in Algorithm 2 is in the choice of the retraction R. In order to relate this algorithm with FP-SQP, we need to develop Newton's equation (15a), i.e., (10). We first do it for the case where $\mathcal{M}$ is a hypersurface, then for the general case.

4.1.1 Newton's method on hypersurfaces

We first consider the case $m = 1$, i.e., the constraint function Φ is a real-valued function and thus the manifold $\mathcal{M} = \Phi^{-1}(0)$ has codimension 1. Let $x \in \mathcal{M}$, i.e., $\Phi(x) = 0$. Let λ be the least-squares multipliers (7), i.e.,

$$\lambda(x) = -(\nabla\Phi(x)^T\nabla\Phi(x))^{-1}\nabla\Phi(x)^T\nabla f(x) \quad \in \mathbb{R}. \tag{16}$$

Then, from (13) and the expression of $\mathcal{P}_x$ (2),

$$\operatorname{grad} f_{|\mathcal{M}}(x) = \mathcal{P}_x\nabla f(x) = \nabla f(x) + \lambda(x)\nabla\Phi(x)$$

and, for all $z \in T_x\mathcal{M}$, the Riemannian Hessian $\operatorname{Hess} f_{|\mathcal{M}}$ satisfies

$$\operatorname{Hess} f_{|\mathcal{M}}(x)z = \mathcal{P}_x\mathrm{D}\operatorname{grad} f_{|\mathcal{M}}(x)z = \mathcal{P}_x(\nabla^2 f(x) + \lambda(x)\nabla^2\Phi(x))z, \tag{17}$$

where we have used $\mathcal{P}_x\nabla\Phi(x) = 0$. Newton's equation (15a) thus reads

$$\mathcal{P}_x(\nabla^2 f(x) + \lambda(x)\nabla^2\Phi(x))z = -\mathcal{P}_x\nabla f(x), \quad \nabla\Phi(x)^T z = 0. \tag{18}$$

This equation corresponds to FP-SQP's (6a)-(6b) in the particular case where λ is chosen according to (16).

4.1.2 General case

We now consider the general case $m \geq 1$, i.e., the constraint function Φ is in general a vector-valued function. Let $x \in \mathcal{M}$, i.e., $\Phi(x) = 0$. Let λ be as in (7), that is,

$$\lambda(x) = -(\nabla\Phi(x)^T\nabla\Phi(x))^{-1}\nabla\Phi(x)^T\nabla f(x) \quad \in \mathbb{R}^m. \tag{19}$$

We have, from (13) and the expression of $\mathcal{P}_x$ (2),

$$\operatorname{grad} f_{|\mathcal{M}}(x) = \mathcal{P}_x\nabla f(x) = \nabla f(x) + \nabla\Phi(x)\lambda(x)$$

and, for all $z \in T_x\mathcal{M}$, the Riemannian Hessian $\operatorname{Hess} f_{|\mathcal{M}}$ satisfies

$$\begin{aligned}\operatorname{Hess} f_{|\mathcal{M}}(x)z &= \mathcal{P}_x\mathrm{D}(\operatorname{grad} f_{|\mathcal{M}})(x)[z] \\ &= \mathcal{P}_x\nabla^2 f(x)z + \mathcal{P}_x\mathrm{D}\nabla\Phi(x)[z]\lambda(x) \qquad (20)\\ &= \mathcal{P}_x\nabla^2 f(x)z - \mathcal{P}_x\mathrm{D}\nabla\Phi(x)[z](\nabla\Phi(x)^T\nabla\Phi(x))^{-1}\nabla\Phi(x)^T\nabla f(x), \qquad (21)\end{aligned}$$

where we have used the identity $\mathcal{P}_x\nabla\Phi(x) = 0$. Newton's equation (15a) becomes

$$\mathcal{P}_x\nabla^2 f(x)z + \mathcal{P}_x\mathrm{D}\nabla\Phi(x)[z]\lambda(x) = -\mathcal{P}_x\nabla f(x), \quad \nabla\Phi(x)^T z = 0. \tag{22}$$

Observe that this equation is again FP-SQP's (6a)-(6b) for the special choice (7) of λ.

The next result formalizes the relation between FP-SQP and the Riemannian Newton method.

Proposition 4.1 (Riemannian Newton and FP-SQP) *Assume as always that* 0 *is a regular value of* Φ, *so that the set* $\mathcal{M}$ *is an embedded submanifold of* $\mathbb{R}^n$. *Then simple FP-SQP (Algorithm 1), with* λ *in Step 3 chosen as in* (7) *and with* $\tilde{z}$ *in Step 5 chosen as a smooth function of* (x, z), *is the Riemannian Newton method (Algorithm 2) with the retraction defined by* $R : z \mapsto x + \tilde{z}$. *Vice-versa, Algorithm 2 is Algorithm 1 with* $\tilde{z} := R(z) - x$.

Proof. For the first claim, it remains to show that that $R : z \mapsto x + \tilde{z}$ is a retraction. This directly follows from the smoothness assumption on $\tilde{z}$ and from (8)-(9). For the second claim, in view of (12), it is clear that $\tilde{z} := R(z) - x$ satisfies (8)-(9). □

4.1.3 Remarks

Let $\mathcal{L}(x,\lambda) = f(x) + \Phi(x)\lambda$, and observe that the left-hand side of (6a) is $\mathcal{P}_x\nabla^2\mathcal{L}(\cdot,\lambda)(x)z$. In the case where λ is chosen as in (7), the link

$$\mathcal{P}_x\nabla^2\mathcal{L}(\cdot,\lambda(x))(x)z = \operatorname{Hess} f_{|\mathcal{M}}(x)[z], \quad z \in T_x\mathcal{M}$$

between the Hessian of the Lagrangian and the Riemannian Hessian has been known since [EAS98, §4.9], and the relation at the critical points of $f_{|\mathcal{M}}$ can already be found in Gabay [Gab82]. Here, by relying on the retraction-based Newton method, we have shown that the overall algorithms overlap. A consequence is that the convergence analysis of the one can benefit the other.

Note that, even though they coincide under the above assumptions, FP-SQP and Newton on manifolds are not identical methods. FP-SQP makes no assumption on λ_k beyond the fact that it must converge to the correct value, whereas Newton on manifolds requires λ_k to be defined by (7). On the other hand, the intrinsic Newton algorithm on manifolds is not restricted to manifolds that come to us in the form $\mathcal{M} = \{x \in \mathbb{R}^n : \Phi(x) = 0\}$; see, e.g., quotient manifolds [AMS08]. Indeed, some of the more abstract manifolds provide the most successful intrinsic Newton algorithms developed to date [EAS98, ADM$^+$02, HT04].

We opted for the Levi-Civita connection on $\mathcal{M}$, hence the projector $\mathcal{P}_x$ in the definition of the Hessian (14) is the *orthogonal* projector onto the tangent space. However, the analysis of Newton's method on manifolds in [ADM$^+$02, AMS08] does not require any assumption on the choice of the affine connection. Choosing a non-orthogonal projector in (14) amounts to choosing a different affine connection to the Riemannian Levi-Civita connection, which leads to a different update step, However, the algorithm is still an intrinsic Newton algorithm, and the full convergence theory will still apply [AMS08].

There is a similar link between the trust-region FP-SQP [WT04] and the Riemannian trust-region method proposed in [ABG07].

4.2 Newton's method in coordinates: Sequential Newton

A natural approach to obtain an intrinsic second-order method is to perform a Newton method on the objective function expressed in local coordinates of $\mathcal{M}$. In this section, we show that for certain choices of the retraction R, the Riemannian Newton step (Algorithm 2) at any $x \in \mathcal{M}$—and thus the simple FP-SQP (Algorithm 1) under the conditions of Proposition 4.1—can be interpreted as a "classical" Newton step on the inner product space $T_x\mathcal{M}$ for the objective function $f \circ R_x$. (The inner product on $T_x\mathcal{M}$ comes naturally by viewing it as a linear subspace of $\mathbb{R}^n$.)

Newton's method in coordinates is defined in Algorithm 3, where $\nabla(f \circ R_x)$ denotes the gradient of $f \circ R_x$ and $\nabla^2(f \circ R_x)$ denotes its Hessian. They are well defined, since $T_x\mathcal{M}$ is a linear subspace of $\mathbb{R}^n$ and thus inherits an inner product. The name *Sequential Newton* is borrowed from [MM05, Alg. 1.8]; see remark in Section 4.2.2.

A sufficient condition for (23) to be equivalent to the Riemannian Newton method (Algorithm 2) with retraction R is that R be a *second-order retraction*, which means that, in addition to the properties (12), we have

$$\mathcal{P}_x \frac{\mathrm{d}^2}{\mathrm{d}t^2} R_x(t\xi_x)\bigg|_{t=0} = 0. \tag{24}$$

Algorithm 3 Sequential Newton

Given a retraction R on $\Phi^{-1}(0)$ and a starting point $x_0 \in \Phi^{-1}(0)$;
for $k = 0, 1, 2, \dots$ **do**
  Solve, for the unknown $z_k \in \ker(\mathrm{D}\Phi(x_k))$, the Newton equation

$$\nabla^2(f \circ R_x)(0_x) z_k = -\nabla(f \circ R_x)(0_x). \tag{23a}$$

  Set

$$x_{k+1} = R_{x_k}(z_k). \tag{23b}$$

end for

Indeed, because R is a retraction, we can appeal to [AMS08, (4.4)] to conclude that

$$\nabla(f \circ R_x)(0_x) = \operatorname{grad} f_{|\mathcal{M}}(x), \tag{25}$$

and when R is second-order, we can also appeal to [AMS08, Pr. 5.5.5] to conclude that

$$\nabla^2(f \circ R_x)(0_x) = \operatorname{Hess} f_{|\mathcal{M}}(x). \tag{26}$$

where $\operatorname{Hess} f_{|\mathcal{M}}$ is as in (14). From (25) and (26), it follows that (23) reduces to the Riemannian Newton method (15).

While (25) is direct, (26) requires more work. For the reader's convenience, we specialize the general proof [AMS08, Pr. 5.5.5] to the present case.

Proposition 4.2 (Riemannian Newton and Sequential Newton) *If R is a second-order retraction, then* (26) *holds, with* $\operatorname{Hess} f_{|\mathcal{M}}(x)$ *as in* (14)*. Consequently, given a second-order retraction R, Riemannian Newton (Algorithm 2) and Sequential Newton (Algorithm 3) are equivalent.*

Proof. Since the two sides of (26) are Hessians, thus symmetric operators, it is sufficient (in view of the polarization identity) to show that, for all $z \in T_x\mathcal{M}$, we have $z^T \nabla^2(f \circ R_x)(0_x) z = z^T \operatorname{Hess} f_{|\mathcal{M}}(x)[z]$. We develop the left-hand side:

$$\begin{aligned}
z^T \nabla^2(f \circ R_x)(0_x) z &= \frac{\mathrm{d}^2}{\mathrm{d}t^2} f \circ R_x(tz) \Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\mathrm{d}}{\mathrm{d}t} f \circ R_x(tz) \right) \Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \left(\nabla f(R_x(tz))^T \frac{\mathrm{d}}{\mathrm{d}t} R_x(tz) \right) \Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \left(\nabla f(R_x(tz))^T \mathcal{P}_{R_x(tz)} \frac{\mathrm{d}}{\mathrm{d}t} R_x(tz) \right) \Big|_{t=0}
\end{aligned}$$

since $\frac{\mathrm{d}}{\mathrm{d}t}R_x(tz) \in T_{R_x(tz)}\mathcal{M}$

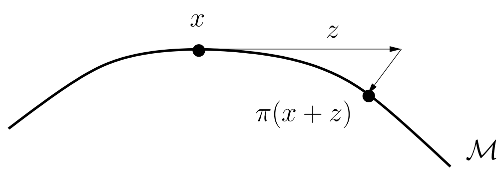
$$
\begin{aligned}
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathcal{P}_{R_x(tz)}\nabla f(R_x(tz))\right)^T\Big|_{t=0} \frac{\mathrm{d}}{\mathrm{d}t}R_x(z)\Big|_{t=0} + \nabla f(x)^T\mathcal{P}_x \frac{\mathrm{d}^2}{\mathrm{d}t^2}R_x(tz)\Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(\mathcal{P}_{R_x(tz)}\nabla f(R_x(tz))\right)^T\Big|_{t=0} z
\end{aligned}
$$

since $\frac{\mathrm{d}}{\mathrm{d}t}R_x(z)\big|_{t=0} = z$ and $\mathcal{P}_x \frac{\mathrm{d}^2}{\mathrm{d}t^2}R_x(tz)\Big|_{t=0} = 0$

$$
\begin{aligned}
&= \left(\mathrm{D}(P\nabla f)(R_x(0_x)) \frac{\mathrm{d}}{\mathrm{d}t}R_x(tz)\Big|_{t=0}\right)^T z \\
&= z^T\mathrm{D}(P\nabla f)(x)z \\
&= z^T\mathrm{Hess}\, f_{|\mathcal{M}}(x)[z].
\end{aligned}
$$

□

4.2.1 Newton's method in projective coordinates



In the context of FP-SQP, Wright and Tenny [WT04] mention the possibility of obtaining $\tilde{z}$ (Step 5 of Algorithm 1) from the projection of $x+z$ onto the feasible set $\mathcal{M}$. Transposed to the Riemannian Newton method (Algorithm 2), in view of Proposition 4.1, this yields the candidate retraction (we have yet to show that it is indeed a retraction)

$$R_x : z \mapsto \pi(x+z), \tag{27}$$

where $\pi : \mathbb{R}^n \to \mathcal{M}$ denotes the projection onto the manifold, i.e., π maps $x+z$ to the nearest point on $\mathcal{M}$ (recall that $\mathcal{M}$ is a submanifold of $\mathbb{R}^n$). By *projective coordinates*, we mean coordinates based on this R.

First, we show that (27) is locally well defined.

Proposition 4.3 (projection) *For all $x \in \mathcal{M}$, there is a neighborhood of 0_x in $T\mathcal{M}$ on which R (27) is well defined and smooth.*

Proof. The proof particularizes classical results of differential geometry pertaining to the normal exponential map and tubular neighborhoods. Let

$$T^{\perp}\mathcal{M} = \{(p,v) : p \in \mathcal{M}, v \perp T_p\mathcal{M}\}$$

denote the normal bundle of $\mathcal{M}$ and let $\mathrm{Exp}^{\perp}$ denote the normal exponential map, which in our case takes the form

$$\mathrm{Exp}^{\perp} : T^{\perp}\mathcal{M} \to \mathcal{M} : (p,v) \mapsto p+v.$$

It is known [Sak96, II.4.3] that for all $x \in \mathcal{M}$, $\mathrm{Exp}^\perp$ is a diffeomorphism (and thus a one-to-one map) on a neighborhood U of 0_x in $T \perp \mathcal{M}$ onto the neighborhood $\mathrm{Exp}^\perp(U)$ of x. Choose $\epsilon > 0$ smaller than half of the radius of a (nonempty) ball around x contained in $\mathrm{Exp}^\perp(U)$. Then, for all $q \in B_\epsilon$, $\pi(q)$ belongs to $\mathcal{M} \cap \mathrm{Exp}^\perp(U)$ and by a standard optimality argument must satisfy $\mathrm{Exp}^\perp(v_{\pi(q)}) = q$ for some $v_{\pi(q)} \in T^\perp_{\pi(q)}\mathcal{M}$. Since $\mathrm{Exp}^\perp$ is a diffeomorphism on U, it follows that $\pi(q) = \mathrm{pj}_1(\mathrm{Exp}^\perp)^{-1}(q)$, where $\mathrm{pj}_1((p,v)) := p$. Hence $\pi(p+z)$ is well defined and smooth for all $p \in \mathcal{M}$ and all $z \in T_p\mathcal{M}$ with $\|p-q\| < \epsilon/2$ and $\|z\| < \epsilon/2$. □

Proposition 4.4 (projective retraction) *The R mapping* (27) *is a retraction. Moreover, it is a second-order retraction.*

Proof. Let $x \in \mathcal{M}$, $z \in T_x\mathcal{M}$, and consider the curve $\gamma : t \mapsto \pi(x+tz)$. We need to show that $\gamma'(0) = z$ (hence R is a retraction) and that $\gamma''(0) \in (T_x\mathcal{M})^\perp$ (hence R is a second-order retraction). We have

$$\gamma(t) = x + tz - \underbrace{\|x + tz - \gamma(t)\|}_{:=\mu(t)} N_{\gamma(t)},$$

where $N_{\gamma(t)}$ belongs to $(T_{\gamma(t)}\mathcal{M})^\perp$. Thus

$$\gamma'(t) = z - \mu'(t)N_{\gamma(t)} - \mu(t)N'_{\gamma(t)},$$

and

$$\gamma''(t) = -\mu''(t)N_{\gamma(t)} - \mu'(t)N'_{\gamma(t)} - \mu'(t)N'_{\gamma(t)} - \mu(t)N''_{\gamma(t)}.$$

Observe that $\mu(0) = 0$, and since

$$\mu'(t) = \frac{1}{\mu(t)}\langle z - \gamma'(t), x + tz - \gamma(t)\rangle = \langle z, N_{\gamma(t)}\rangle,$$

we have $\mu'(0) = \langle z, N_x\rangle = 0$. Hence

$$\gamma'(0) = z \quad \text{and} \quad \gamma''(0) = -\mu''(t)N_x \in (T_x\mathcal{M})^\perp.$$

□

Proposition 4.5 (projective Newton) *Newton's method in projective coordinates (Algorithm 3 with R as in* (27)*) is equivalent to the Riemannian Newton method (Algorithm 2) with R as in* (27)*.*

Proof. Direct from Propositions 4.2 and 4.4. □

4.2.2 Remarks

Algorithm 3 is identical to [MM05, Alg. 1.8], except that the first-order condition (12b) is not enforced in the latter. Nevertheless, Algorithm 3 is not more restrictive than [MM05, Alg. 1.8]. Indeed, if φ_x satisfies the properties of a retraction except for (12b), then $R_x := \varphi_x \circ (\mathrm{D}\varphi_x(0))^{-1}$ is a retraction, and we have that the mapping $x \mapsto x_+$ defined by

$$\nabla^2(f \circ \varphi_x)(0_x)z = -\nabla(f \circ \varphi_x)(0_x) \tag{28a}$$

$$x_+ = \varphi_x(z) \tag{28b}$$

is the same as (23). To see this, observe that

$$\nabla^2(f\circ R_x)(0_x)z = \mathrm{D}\varphi_x(0_x)^{-T}\nabla^2(f\circ\varphi_x)(0_x)\mathrm{D}\varphi_x(0_x)^{-1}z$$

for all $z\in T_x\mathcal{M}$, and

$$\nabla(f\circ R_x)(0_x) = \mathrm{D}\varphi_x(0)^{-T}\nabla(f\circ\varphi_x)(0_x).$$

Note that the domain of the parameterizations φ_x is (a subset of) the tangent space $T_x\mathcal{M}$, and not (a subset of) $\mathbb{R}^d$. Strictly speaking, they are thus not parameterizations in the sense of differential geometry (see, e.g., [dC92]), because identifying $T_x\mathcal{M}$ with $\mathbb{R}^d$ requires to choose a basis of $T_x\mathcal{M}$. In [HT04], parameterizations are defined on open subsets of $\mathbb{R}^d$. Both approaches have advantages and drawbacks; see in particular the discussion in [AMS08, §4.1.3].

Several examples of second-order retractions can be found in [AM09].

The fact that the Riemannian Newton method converges quadratically does not follow straightforwardly from Proposition 4.5, because the objective function $f\circ R_x$ and its domain $T_x\mathcal{M}$ change at each step as x changes. Showing quadratic convergence requires some work, which can be found in [HT04, AMS08].

5 Methods based on approximations of $\mathcal{M}$

In this section, we tackle the general equality-constrained optimization problem (1) by considering at the current iterate x the subproblem

$$\min f(\tilde{x}) \text{ subject to } \tilde{x}\in\widetilde{\mathcal{M}}_x, \tag{29}$$

where the submanifold $\widetilde{\mathcal{M}}_x$ is some approximation of class C^2 of the submanifold $\mathcal{M}=\Phi^{-1}(0)$. This is thus an extrinsic approach: evaluating f on $\widetilde{\mathcal{M}}$ means in general evaluating f outside $\mathcal{M}$. We give conditions on $\widetilde{M}_x$, termed second-order agreement (defined next), such that the Newton vector at x for (29) and the one for the original problem (1) coincide.

We require that $\widetilde{\mathcal{M}}_x$ contains x and that the tangent spaces match, i.e., $T_x\widetilde{\mathcal{M}}_x = T_x\mathcal{M}$ thought of as subspaces in $\mathbb{R}^n$. To define the notion of second-order agreement at x between $\mathcal{M}$ and $\widetilde{\mathcal{M}}_x$, observe that every $z\in\mathbb{R}^n$ admits a unique decomposition $z=z_T+z_N$ where z_T belongs to $T_x\mathcal{M}$ and z_N to $T_x^\perp\mathcal{M}$. Since $\ker(\mathrm{D}\Phi(x)) = T_x\mathcal{M}$ and $\mathrm{D}\Phi(x)$ has full rank m (by LICQ), it follows that $\mathrm{D}_N\Phi(x)$ is invertible, where $\mathrm{D}_N\Phi(x)$ denotes $\mathrm{D}\Phi(x)$ restricted to act along $T_x^\perp\mathcal{M}$. Hence, by the implicit function theorem, there exists an open neighborhood U of 0 in $T_x\mathcal{M}$, an open neighborhood V of 0 in $T_x^\perp\mathcal{M}$, and a unique continuously differentiable function $z_N: U\to V$ such that $\{x+z_T+z_N(z_T) : z_T\in U\} = \mathcal{M}\cap(x+U+V)$. In other words, $\mathcal{M}-x$ is locally the graph of $z_N(\cdot)$. Since Φ is assumed to be C^2, it follows that z_N is also C^2. Since every C^2 submanifold is locally the zero set of a C^2 function (see [O'N83, Prop. 1.28]), the same reasoning shows that there is a unique C^2 (partial) function $\widetilde{z_N}$ such that $\widetilde{\mathcal{M}}_x - x$ is locally the graph of $\widetilde{z_N}$. We say that $\mathcal{M}$ and $\widetilde{\mathcal{M}}_x$ *agree to second order* if $\mathrm{D}^2 z_N(0) = \mathrm{D}^2\widetilde{z_N}(0)$. If $\widetilde{z_N}(z_T) = \frac{1}{2}\mathrm{D}^2 z_N(0)[z_T, z_T]$, then $\widetilde{\mathcal{M}}_x$ is termed the *osculating paraboloid* to $\mathcal{M}$ at x.

In [MM05], the (partial) function $z_T\mapsto x+z_T+z_N(z_T)$ is termed *tangential parameterization.* It defines a retraction (this follows, e.g., from [Lew02, Th. 6.1]), and even a second-order retraction [AM09].

5.1 Newton approach with $\widetilde{\mathcal{M}}$

We will see that if $\mathcal{M}$ and $\widetilde{\mathcal{M}}_x$ agree to second order at x and we compute the Riemannian Newton vector at x for problem (29), then we get the Riemannian Newton vector of the original problem (1).

Lemma 5.1 *Let x belong to $\mathcal{M} = \Phi^{-1}(0)$ and let $z_N : T_x\mathcal{M} \to T_x^{\perp}\mathcal{M}$ be such that $\mathcal{M} - x$ is locally the graph of $z_N(\cdot)$. Then*

$$\begin{aligned} \mathrm{D}^2 z_N(0)[u_1, u_2] &= -\mathrm{D}\Phi(x)^T(\mathrm{D}\Phi(x)\mathrm{D}\Phi(x)^T)^{-1}\mathrm{D}^2\Phi(x)[u_1, u_2] \\ &= -\nabla\Phi(x)(\nabla\Phi(x)^T\nabla\Phi(x))^{-1}\mathrm{D}(\nabla\Phi^T)(x)[u_2]u_1 \end{aligned} \tag{30}$$

for all $u_1, u_2 \in T_x\mathcal{M}$.

Proof. This is [MM05, (3.6)]. We give a sketch of proof for the reader's convenience. Differentiating the relation $\Phi(x + z_T + z_N(z_T)) = 0$ along u_1 yields

$$\mathrm{D}_T\Phi(x + z_T + z_N(z_T))[u_1] + \mathrm{D}_N\Phi(x + z_T + z_N(z_T))\mathrm{D}z_N(z_T)[u_1] = 0$$

for all $u_1 \in T_x\mathcal{M}$, where $\mathrm{D}_T\Phi$ denotes $\mathrm{D}\Phi$ restricted to act along $T_x\mathcal{M}$. Differentiating again along u_2 and evaluating the expression at $z_T = 0$ yields

$$\mathrm{D}_T^2\Phi(x)[u_1, u_2] + \mathrm{D}_N\Phi(x)\mathrm{D}^2 z_N(0)[u_1, u_2] = 0$$

for all $u_1, u_2 \in T_x\mathcal{M}$. This yields

$$\mathrm{D}^2 z_N(0)[u_1, u_2] = -(\mathrm{D}_N\Phi(x))^{-1}\mathrm{D}_T^2\Phi(x)[u_1, u_2] \tag{31}$$

$$= -(\mathrm{D}_N\Phi(x))^T \left(\mathrm{D}_N\Phi(x)(\mathrm{D}_N\Phi(x))^T\right)^{-1} \mathrm{D}_T^2\Phi(x)[u_1, u_2] \tag{32}$$

$$= -(\mathrm{D}\Phi(x))^T \left(\mathrm{D}\Phi(x)(\mathrm{D}\Phi(x))^T\right)^{-1} \mathrm{D}^2\Phi(x)[u_1, u_2] \tag{33}$$

since $\mathrm{D}_T\Phi(x) = 0$. □

Recall that $z \mapsto \mathrm{Hess}\, f_{|\mathcal{M}}(x)[z]$ is a mapping from $T_x\mathcal{M}$ to $T_x\mathcal{M}$, hence $\mathrm{Hess}\, f_{|\mathcal{M}}(x)$ is fully specified by the mapping $T_x\mathcal{M} \times T_x\mathcal{M} \ni (u, z) \mapsto u^T\mathrm{Hess}\, f_{|\mathcal{M}}(x)[z] \in \mathbb{R}$. Next we give an expression for this mapping that involves $\mathrm{D}^2 z_N(0)$.

Proposition 5.2 (Riemannian Hessian and tangential coordinates) *With the notation and assumptions of the previous lemma, for all $u, z \in T_x\mathcal{M}$,*

$$u^T\mathrm{Hess}\, f_{|\mathcal{M}}(x)[z] = u^T\nabla^2 f(x)z + \nabla f(x)^T\mathrm{D}^2 z_N(0)[u, z]. \tag{34}$$

Proof. By (21), the left-hand side is

$$u^T\nabla^2 f(x)z - u^T\mathrm{D}\nabla\Phi(x)[z](\nabla\Phi(x)^T\nabla\Phi(x))^{-1}\nabla\Phi(x)^T\nabla f(x),$$

and the result comes by Lemma 5.1. □

Proposition 5.3 (Riemannian Newton on $\widetilde{\mathcal{M}}_x$) *The Riemannian Newton vector z, solution of* (15a)*, is unchanged if the manifold $\mathcal{M}$ is replaced by a manifold $\widetilde{\mathcal{M}}_x$ that agrees with $\mathcal{M}$ to second order at x.*

Proof. Replacing $\mathcal{M}$ by $\widetilde{\mathcal{M}}_x$ preserves the tangent space at x and $\mathrm{D}^2 z_N(0)$ by definition. □

5.2 Obtaining $\widetilde{\mathcal{M}}$ from models of Φ

A way of obtaining an approximation $\widetilde{\mathcal{M}}_x$ of $\mathcal{M}$ is to approximate Φ by some $\widetilde{\Phi}$ and define $\widetilde{\mathcal{M}}_x = \widetilde{\Phi}^{-1}(0)$. We give necessary and sufficient conditions for this procedure to yield a manifold $\widetilde{\mathcal{M}}_x$ that agrees with $\mathcal{M}$ to second order at x.

Proposition 5.4 *Let Φ and $\widetilde{\Phi}$ be C^2 functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ with 0 as regular value. Then $\Phi^{-1}(0)$ and $\widetilde{\Phi}^{-1}(0)$ agree to second order at x if and only if $(\mathrm{D}_N\Phi(x))^{-1}\mathrm{D}^2_T\Phi(x) = (\mathrm{D}_N\widetilde{\Phi}(x))^{-1}\mathrm{D}^2_T\widetilde{\Phi}(x)$.*

Proof. Immediate from (31). □

The criterion in Proposition 5.4 is satisfied when the functions Φ and $\widetilde{\Phi}$ agree to second order, but this condition is not necessary. However, it is always possible to satisfy this condition by replacing one of the two functions by another function that admits the same zero level set, as the next result shows.

Proposition 5.5 *Let Φ and $\widetilde{\Phi}$ be C^2 functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ with 0 as regular value and let $x \in \Phi^{-1}(0)$. Then $\Phi^{-1}(0)$ and $\widetilde{\Phi}^{-1}(0)$ agree to second order at x if and only if there exists a function $\widehat{\Phi}$ such that $\widehat{\Phi}^{-1}(0) = \Phi^{-1}(0)$ and that $\widehat{\Phi}$ and $\widetilde{\Phi}$ agree to second order at x (i.e., $\mathrm{D}\widehat{\Phi}(x) = \mathrm{D}\widetilde{\Phi}(x)$ and $\mathrm{D}^2\widehat{\Phi}(x) = \mathrm{D}^2\widetilde{\Phi}(x)$).*

Proof. The condition is sufficient. If $\widetilde{\Phi}$ and $\widehat{\Phi}$ agree to second order at x, then from (31) it follows that $\widetilde{\Phi}^{-1}(0)$ and $\widehat{\Phi}^{-1}(0)$ agree to second order at x. Since $\widehat{\Phi}^{-1}(0) = \Phi^{-1}(0)$, the result follows.

The condition is necessary. It is enough to prove that there exists $M(\cdot)$ defined on a neighborhood of x in $\mathbb{R}^n$ such that $\widehat{\Phi}$ defined by

$$\widehat{\Phi}(y) := M(y)\Phi(y) \tag{35}$$

satisfies

$$\mathrm{D}\widehat{\Phi}(x) = \mathrm{D}\widetilde{\Phi}(x) \tag{36}$$

$$\mathrm{D}^2\widehat{\Phi}(x) = \mathrm{D}^2\widetilde{\Phi}(x). \tag{37}$$

The tangent part of (36) is trivially satisfied because $\mathrm{D}_T\Phi(0) = 0 = \mathrm{D}_T\widetilde{\Phi}(0)$. From (35), we have, for all $u \in \mathbb{R}^n$,

$$\mathrm{D}\widehat{\Phi}(y)[u] = \mathrm{D}M(y)[u]\Phi(y) + M(y)\mathrm{D}\Phi(y)[u]$$

and thus

$$\mathrm{D}\widehat{\Phi}(x)[u] = M(x)\mathrm{D}\Phi(x)[u].$$

Hence, the normal part of (36) is satisfied with

$$M(x) = \mathrm{D}_N\widetilde{\Phi}(x)\,(\mathrm{D}_N\Phi(x))^{-1}. \tag{38}$$

From (35), we have, for all $u, v \in \mathbb{R}^n$,

$$\mathrm{D}^2\widehat{\Phi}(x)[u,v] = \mathrm{D}M(x)[u]\mathrm{D}\Phi(x)[v] + \mathrm{D}M(x)[v]\mathrm{D}\Phi(x)[u] + M(x)\mathrm{D}^2\Phi(x)[u,v]. \tag{39}$$

The tangent-tangent part of $D^2\widehat{\Phi}(x)$ is given by

$$D^2\widehat{\Phi}(x)[u_T, v_T] = M(x)D^2\Phi(x)[u_T, v_T].$$

The tangent-tangent part of (37) is satisfied without any constraint on M further than (38), in view of Proposition 5.4. The tangent-normal part of $D^2\widehat{\Phi}(0)$ is given by

$$D^2\widehat{\Phi}(x)[u_T, v_N] = DM(x)[u_T]D\Phi(x)[v_N] + M(x)D^2\Phi(x)[u_T, v_N].$$

The tangent-normal part of (37) is satisfied if

$$D_T M(x)[u_T]D\Phi(x)[v_N] + M(x)D^2_{TN}\Phi(x)[u_T, v_N] = D^2_{TN}\widetilde{\Phi}(x)[u_T, v_N],$$

for all $u_T \in T_x\mathcal{M}$ and all $v_N \in T_x^\perp\mathcal{M}$. The application $v_N \mapsto D^2_{TN}\widetilde{\Phi}(x)[u_T, v_N]$ is a linear mapping from $T_x\mathcal{M} \simeq \mathbb{R}^m$ to $\mathbb{R}^m$; we denote its matrix expression by $D^2_{TN}\widetilde{\Phi}(x)[u_T, \cdot]$. This yields

$$D_T M(x)[u_T] = \left(D^2_{TN}\widetilde{\Phi}(x)[u_T, \cdot] - M(x)D^2_{TN}\Phi(x)[u_T, \cdot]\right) D_N\Phi(x)^{-1}.$$

This is well defined since, by the LICQ hypothesis and the fact that $D_T\Phi(x) = 0$, $D_N\Phi(x)$ is invertible. Finally, the normal-normal part of $D^2\widehat{\Phi}(x)$ is given by

$$D^2\widehat{\Phi}(x)[U_N, V_N] = D_N M(x)[u_N]D_N\Phi(x)[v_N]+D_N M(x)[v_N]D_N\Phi(x)[u_N]+M(x)D^2_{NN}\Phi(x)[u_N, v_N].$$

The normal-normal part of (37) is satisfied if

$$D_N M(x)[u_N] = \frac{1}{2}\left(D^2_{NN}\widetilde{\Phi}(x)[u_N, \cdot] - M(x)D^2_{NN}\Phi(x)[u_N, \cdot]\right) D_N\Phi(x)^{-1}.$$

This is readily checked by taking into account the symmetry of the second derivatives. In conclusion, $M(x)$ and $DM(x)$ have been fully specified, and this is enough for (36) and (37) to hold. A possible choice for $M(\cdot)$ is $M(y) = M(x) + DM(x)[y - x]$. □

5.3 $\widetilde{\mathcal{M}}$ and second-order corrections

Second-order corrections were proposed as a strategy to avoid the Maratos effect; see, e.g., [BGLS03, §15.3] or [NW06, §15.6]. Specifically, let $x \in \mathcal{M}$ and let K be a right inverse of $D\Phi(x)$. (For example, K can be chosen as the pseudo-inverse $D\Phi(x)^T(D\Phi(x)D\Phi(x)^T)^{-1}$.) Given $z \in T_x\mathcal{M}$, the *second-order correction step* is defined to be $-K\Phi(x + z)$, and the corrected step sends x to

$$y := x + z - K\Phi(x + z). \tag{40}$$

We show that the set of all second-order corrections is a manifold that agrees with $\mathcal{M}$ to second order at x.

Proposition 5.6 *Given $x \in \mathcal{M}$ and K such that $D\Phi(x)K = I$, the manifold*

$$\widetilde{\mathcal{M}}_x := \{x + z - K\Phi(x + z) : z \in T_x\mathcal{M}\}$$

agrees with $\mathcal{M}$ to second order at x.

Proof. Let P_{NT} denote the projection onto $T_x^\perp\mathcal{M}$ along $T_x\mathcal{M}$, and let P_{TK} denote the projection onto $T_x\mathcal{M}$ along the range of K. Since K is a right inverse of $\mathrm{D}\Phi(0)$, it follows that the range of K and $T_x\mathcal{M}$ are transverse, and thus P_{TK} is well defined. One can check that $P_{NT}K = (\mathrm{D}_N\Phi(x))^{-1}$ because $\mathrm{D}_T\Phi(x) = 0$. Applying P_{NT} to (40) thus yields

$$P_{NT}y = -(\mathrm{D}_N\Phi(0))^{-1}\Phi(x+z).$$

Since $P_{TK}K = 0$ and $P_{TK}z = z$, applying P_{TK} to (40) yields

$$P_{TK}(y-x) = z.$$

Replacing this equation into the previous one yields

$$0 = P_{NT}y + (\mathrm{D}_N\Phi(0))^{-1}\Phi(x + P_{TK}(y-x)).$$

Define $\widetilde{\Phi}$ by

$$\widetilde{\Phi}(y) = P_{NT}y + (\mathrm{D}_N\Phi(0))^{-1}\Phi(x + P_{TK}(y-x)).$$

Note that since $\widetilde{\Phi}$ is into $T_x^\perp\mathcal{M}$, it can be viewed as a function into $\mathbb{R}^m$. It can be shown that

$$\widetilde{\mathcal{M}}_x = \widetilde{\Phi}^{-1}(0).$$

(Indeed, the inclusion $\subseteq$ is direct and the inclusion $\supseteq$ comes because any $y \in \mathbb{R}^n$ is fully specified by $(P_{NT}y, P_{TK}y)$ hence no information is lost in the two projections of (40).) We have, for all $u \in T_x\mathcal{M}$ an all $v \in T_x^\perp\mathcal{M}$,

$$\begin{aligned}
\mathrm{D}_T\widetilde{\Phi}(y)[u] &= (\mathrm{D}_N\Phi(0))^{-1}\mathrm{D}\Phi(x + P_{TK}(y-x))P_{TK}u \\
\mathrm{D}^2_{TT}\widetilde{\Phi}(x)[u,u] &= (\mathrm{D}_N\Phi(0))^{-1}\mathrm{D}^2\Phi(x)[P_{TK}u, P_{TK}u] \\
\mathrm{D}_N\widetilde{\Phi}(x)[v] &= v + (\mathrm{D}_N\Phi(0))^{-1}\mathrm{D}\Phi(x)P_{TK}v.
\end{aligned}$$

Observe that $P_{TK}u = u$ and $\mathrm{D}\Phi(x)P_{TK} = 0$. Hence

$$(\mathrm{D}_N\widetilde{\Phi}(x))^{-1}\mathrm{D}^2_{TT}\widetilde{\Phi}(x) = (\mathrm{D}_N\Phi(x))^{-1}\mathrm{D}^2_{TT}\Phi(x).$$

The conclusion comes using Proposition 5.4. □

6 An extrinsic look at the Riemannian Hessian

Up to now we have obtained several expressions for the Riemannian Hessian of f on $\mathcal{M}$. In this section, we take a step back to interpret these expressions and explain how they relate to the curvature of $\mathcal{M}$.

We have, respectively from (14), (26), (21), (34), that for all $z \in T_x\mathcal{M}$,

$$\begin{aligned}
\operatorname{Hess} f_{|\mathcal{M}}(x)[z] &= \mathcal{P}_x\mathrm{D}\left(\operatorname{grad} f_{|\mathcal{M}}\right)(x)[z] = \mathcal{P}_x\mathrm{D}(\mathcal{P}\nabla f)(x)[z] \\
&= \nabla^2(f \circ R_x)(0_x)z
\end{aligned}$$

for all second-order retractions R,

$$\begin{aligned}
&= \mathcal{P}_x\nabla^2 f(x)z - \mathcal{P}_x\mathrm{D}\nabla\Phi(x)[z](\nabla\Phi(x)^T\nabla\Phi(x))^{-1}\nabla\Phi(x)^T\nabla f(x) \\
&= \mathcal{P}_x\nabla^2 f(x)z + (\mathrm{D}^2 z_N[\cdot, z])^T\nabla f(x),
\end{aligned}$$

and also, by Leibniz's law of derivation (product rule) applied to (14),

$$\text{Hess}\, f_{|\mathcal{M}}(x)[z] = \mathcal{P}_x \nabla^2 f(x) z + \mathcal{P}_x \mathrm{D}\mathcal{P}(x)[z]\, \nabla f(x). \tag{41}$$

The Riemannian Hessian is intrinsic to $\mathcal{M}$ (it is unaffected by modifications of f away from $\mathcal{M}$), as the second of the five expressions above confirms since the range of R_x is contained in $\mathcal{M}$. However, the last three expressions are the sum of two terms that, taken individually, are extrinsic to $\mathcal{M}$ (they depend on f outside $\mathcal{M}$). The first term in those expressions, $\mathcal{P}_x \nabla^2 f(x) z$, contains second-order information about f along the tangent space $T_x\mathcal{M}$ but only first-order information on $\mathcal{M}$. The second term involves second-order information about $\mathcal{M}$ but only first-order information about f. Observe also that the first-order information about f is only needed along the normal space, as the penultimate expression shows.

6.1 $\mathrm{D}\mathcal{P}$ and the second fundamental form

Expression (41) of the Riemannian Hessian involves the derivation of the projector field, $\mathrm{D}\mathcal{P}$. In this section, we show that $\mathrm{D}\mathcal{P}$ relates to the curvature of $\mathcal{M}$ through the so-called second fundamental form and Weingarten map. Whereas the relation between the second fundamental form and the Riemannian curvature tensor is well known, we could not find the link between $\mathrm{D}\mathcal{P}$ and the second fundamental form clearly spelled out in the literature. Contrary to the other sections, this section requires some background in differential geometry.

We opt for the more compact notation $\mathrm{D}_X F$ for the derivative of a (real-, vector-, or matrix-valued) function F along a vector field X tangent to $\mathcal{M}$. Recall that, for all $y \in \mathcal{M}$, $\mathcal{P}_y$ denotes the orthogonal projection onto the tangent space $T_y\mathcal{M}$. If we let N denote a matrix-valued function on $\mathcal{M}$ such that the columns of $N(y)$ form a basis of the normal space $T_y^\perp\mathcal{M}$ for all y in a neighborhood of x (such a N can always be constructed locally for any submanifold $\mathcal{M}$ of $\mathbb{R}^n$), then we have

$$\mathcal{P}_y = I - N(y)(N(y)^T N(y))^{-1} N(y)^T.$$

A possible choice is $N(y) = \nabla\Phi(y) = \mathrm{D}\Phi(y)^T$. We also let $\mathcal{P}_y^\perp$ denote the orthogonal projection onto $T_y^\perp\mathcal{M}$,

$$\mathcal{P}_y^\perp = N(y)(N(y)^T N(y))^{-1} N(y)^T. \tag{42}$$

Obviously, $\mathcal{P} = I - \mathcal{P}^\perp$.

The projection $\mathcal{P}$ defines a map from $\mathcal{M}$ to the Grassmannian of d-planes in $\mathbb{R}^n$, which we identify with the space of projections $\{P \in GL(n) :\ P^2 = P,\ P = P^T,\ \mathrm{tr}P = d\}$. This map is called the *Gauss map* of the submanifold $\mathcal{M}$ (in codimension 1 this is just the map to S^{n-1} given by the unit normal vector, up to sign). Since $\mathcal{P}\mathcal{P}^\perp = 0$ we have $(\mathrm{D}_X\mathcal{P})\mathcal{P}^\perp + \mathcal{P}(\mathrm{D}_X\mathcal{P}^\perp) = 0$, and hence

$$\mathcal{P}^\perp(\mathrm{D}_X\mathcal{P})\mathcal{P}^\perp = 0.$$

Similarly,

$$\mathcal{P}(\mathrm{D}_X\mathcal{P})\mathcal{P} = 0.$$

Thus we can write

$$\begin{aligned}\mathrm{D}_X\mathcal{P} &= (\mathcal{P}+\mathcal{P}^\perp)\mathrm{D}_X\mathcal{P}(\mathcal{P}+\mathcal{P}^\perp)\\ &= \mathcal{P}(\mathrm{D}_X\mathcal{P})\mathcal{P}^\perp + \mathcal{P}^\perp(\mathrm{D}_X\mathcal{P})\mathcal{P}. \end{aligned}\tag{43}$$

As we now show, these two terms relate to the second fundamental form and the Weingarten map.

The *second fundamental form* is a bilinear form which takes two tangent vectors to $\mathcal{M}$ at x and gives a normal vector, defined as follows: If X and Y are tangent vector fields, then

$$I\!I(X,Y) = \mathcal{P}^\perp\left(\mathrm{D}_X Y\right).$$

This depends only on the value of the (tangent) vector field Y at the point x: Since $Y = \mathcal{P}Y$ we have

$$I\!I(X,Y) = \mathcal{P}^\perp\left(\mathrm{D}_X(\mathcal{P}Y)\right) = \mathcal{P}^\perp\left((\mathrm{D}_X\mathcal{P})(Y) + \mathcal{P}(\mathrm{D}_X Y)\right) = \mathcal{P}^\perp\left((\mathrm{D}_X\mathcal{P})(Y)\right) \tag{44}$$

since $\mathcal{P}^\perp\mathcal{P} = 0$. Noting that $D\mathcal{P} = -D\mathcal{P}^\perp$, we can write $I\!I$ in terms of Φ using (42): Differentiating the expression for $\mathcal{P}^\perp$ gives zero when applied to a tangential vector Y, except where the derivative applies to the last term $D\Phi$. Thus we have (since $\mathcal{P}^\perp(D\Phi)^T = (D\Phi)^T$)

$$I\!I(X,Y) = -(D\Phi)^T(D\Phi D\Phi^T)^{-1}D^2\Phi(X,Y), \tag{45}$$

from which it follows that $I\!I$ is symmetric.

In the high codimension case the *Weingarten map* is defined to be the following operator which takes a tangent vector X and a normal vector V and produces a tangent vector:

$$A(X,V) = -\mathcal{P}\left(\mathrm{D}_X V\right).$$

We have

$$A(X,V) = -\mathcal{P}\left(\mathrm{D}_X(\mathcal{P}^\perp V)\right) = -(\mathcal{P}\mathrm{D}_X\mathcal{P}^\perp)(V) = \mathcal{P}(\mathrm{D}_X\mathcal{P}(V)). \tag{46}$$

The Weingarten map is related to the second fundamental form via the Weingarten relation, which is obtained by differentiating the equation $\langle Y,V\rangle = 0$ for any tangent vector field Y and normal vector field V:

$$0 = \mathrm{D}_X\langle Y,V\rangle = \langle(\mathrm{D}_X Y),V\rangle + \langle Y,\mathrm{D}_X V\rangle = \langle I\!I(X,Y),V\rangle - \langle Y,A(X,V)\rangle.$$

In summary, we have, for all tangent vector fields X, Y, and all normal vector fields V,

$$\begin{aligned} I\!I(X,Y) &= \mathcal{P}^\perp(\mathrm{D}_X\mathcal{P})Y,\\ A(X,V) &= \mathcal{P}(\mathrm{D}_X\mathcal{P})V. \end{aligned}$$

Hence, from (43), we have, for all vector fields W and tangent vector fields X on $\mathcal{M}$,

$$\begin{aligned}(\mathrm{D}_X\mathcal{P})W &= \mathcal{P}(\mathrm{D}_X\mathcal{P})\mathcal{P}^\perp W + \mathcal{P}^\perp(\mathrm{D}_X\mathcal{P})\mathcal{P}W\\ &= A(X,\mathcal{P}^\perp W) + I\!I(X,\mathcal{P}W).\end{aligned}$$

In these terms the intrinsic Hessian on $\mathcal{M}$ of a function f is given as follows: For any two tangent vectors X and Y, it follows from (41) that

$$\mathrm{Hess}\, f_{|\mathcal{M}}X = \mathcal{P}\nabla^2 f X + A(X,\mathcal{P}^\perp\nabla f)$$

and

$$\langle \mathrm{Hess}\, f_{|\mathcal{M}}X, Y\rangle = \mathrm{D}^2 f(X,Y) + \mathrm{D}f(I\!I(X,Y)).$$

6.2 Relation to curvature

We give some geometric interpretations of the second fundamental form.

For a fixed unit normal vector N at $x \in \mathcal{N}$, $\langle I\!I(.,.), N\rangle$ is the second fundamental form of the codimension one submanifold $\mathcal{M}_N$ given by projecting $\mathcal{M}$ orthogonally onto the $(d+1)$-dimensional subspace generated by T_xM and N. Thus for each unit tangent vector X, $\langle I\!I(X,X), N\rangle$ gives the curvature of the curve given by intersecting $\mathcal{M}_N$ with the plane generated by X and N.

Another useful interpretation is the following: For fixed X, the geodesic γ in $\mathcal{M}$ with $\gamma(0) = x$ and $\gamma'(0) = X$ has acceleration in $\mathbb{R}^n$ given by $\gamma''(0) = I\!I(X,X)$. (This follows directly from [O'N83, Cor. 4.9].)

As we saw in Section 5, the submanifold $\mathcal{M}$ can be written near x as the graph of a locally defined smooth function $u : \mathbb{R}^d \to \mathbb{R}^m$ over the tangent space $T_x\mathcal{M}$: Precisely, if we decompose $\mathbb{R}^n$ as $T_x\mathcal{M} \times N_x\mathcal{M}$ and choose the origin to be at x, the we can write $\mathcal{M}$ near x in the form

$$\{(z, u(z)) : \ z \in T_xM\}$$

for some smooth function $u : T_x\mathcal{M} \simeq \mathbb{R}^d \to N_x\mathcal{M} \simeq \mathbb{R}^m$ with $u(0) = 0$ and $\mathrm{D}u(0) = 0$. Then the second fundamental form of $\mathcal{M}$ at x is given by

$$I\!I(X,Y) = \mathrm{D}^2u(0)(X,Y) \in N_x\mathcal{M}$$

for each $X, Y \in T_x\mathcal{M}$. (This follows from [KN63, Example VII.3.3]. The result also comes directly from (45) and (30).)

Finally, we point out that the Riemannian curvature tensor and the sectional curvature can also be written in terms of the second fundamental form; see, e.g., [O'N83, §4].

7 Conclusion

In the context of equality-constrained optimization with a smooth feasible set, where the two worlds of constrained optimization and optimization on manifolds meet, we have presented connections between the feasibly-projected SQP, the Riemannian Newton method, and Newton's method in coordinates. We have also shown that the Riemannian Newton vector is preserved when the manifold is replaced by another manifold with which it agrees to second order at the current iterate, a situation that prevails when Maratos-like second-order corrections are applied. An interpretation of the Riemannian Hessian in terms of curvature of the feasible manifold has been given.

We point out that in the feasibly-projected SQP, it can be argued that setting the multipliers λ to their least-squares value (7) is a good choice since this ensures that the algorithm is intrinsic to the manifold; i.e., the algorithm is not affected by modifications of the objective function away from the feasible set.

The concepts discussed in this paper can be generalized to Newton trust region and Newton line search algorithms using the formulation from [ABG07, AMS08].

Acknowledgements

The authors would like to thank André Tits for pointing them to [WT04] and Jérôme Malick for introducing them to the literature on $\mathcal{U}$-Lagrangians.

References

[ABG06] P.-A. Absil, C. G. Baker, and K. A. Gallivan. A truncated-CG style method for symmetric generalized eigenvalue problems. *J. Comput. Appl. Math.*, 189(1–2):274–285, 2006.

[ABG07] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3):303–330, July 2007.

[ADM+02] Roy L. Adler, Jean-Pierre Dedieu, Joseph Y. Margulies, Marco Martens, and Mike Shub. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, 22(3):359–390, July 2002.

[AM09] P.-A. Absil and Jérôme Malick. Constructing retractions on matrix manifolds. in preparation, 2009.

[AMR88] R. Abraham, J. E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*, volume 75 of *Applied Mathematical Sciences*. Springer-Verlag, New York, second edition, 1988.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.

[AT07] P.-A. Absil and André L. Tits. Newton-KKT interior-point methods for indefinite quadratic programming. *Comput. Optim. Applic.*, 36(1):5–41, 2007.

[BAG08] C. G. Baker, P.-A. Absil, and K. A. Gallivan. An implicit trust-region method on Riemannian manifolds. *IMA J. Numer. Anal.*, 28(4):665–689, 2008.

[BGLS03] J. Frédéric Bonnans, J. Charles Gilbert, Claude Lemaréchal, and Claudia A. Sagastizábal. *Numerical Optimization*. Universitext. Springer-Verlag, Berlin, 2003. Theoretical and practical aspects. Translated and revised from the 1997 French original.

[Cha91] Geoffrey Chaucer. A treatise on the astrolabe, 1391. Prologue, II. 39-40.

[dC92] Manfredo Perdigão do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[DHM06] Aris Daniilidis, Warren Hare, and Jérôme Malick. Geometrical interpretation of the predictor-corrector type algorithms in structured optimization problems. *Optimization*, 55(5-6):481–503, 2006.

[dL75] Alain de Lille. Liber parabolarum, 591, 1175.

[EAS98] Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1998.

[Gab82] D. Gabay. Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.*, 37(2):177–219, 1982.

[GL76] Daniel Gabay and David G. Luenberger. Efficiently converging minimization methods based on the reduced gradient. *SIAM J. Control Optimization*, 14(1):42–61, 1976.

[GP07] Igor Grubišić and Raoul Pietersz. Efficient rank reduction of correlation matrices. *Linear Algebra Appl.*, 422(2-3):629–653, 2007.

[HHLM07] Uwe Helmke, Knut Hüper, Pei Yean Lee, and John B. Moore. Essential matrix estimation using Gauss-Newton iterations on a manifold. *Int. J. Computer Vision*, 74(2):117–136, 2007.

[HM94] Uwe Helmke and John B. Moore. *Optimization and Dynamical Systems*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1994. With a foreword by R. Brockett.

[HT04] Knut Hüper and Jochen Trumpf. Newton-like methods for numerical optimization on manifolds. In *Proceedings of the 38th IEEE Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 7–10, 2004*, 2004.

[JBAS08] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization for semidefinite convex problems. Technical Report arXiv:0807.4423, 2008.

[KN63] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of Differential Geometry*. Interscience Publishers, a division of John Wiley & Sons, New York-London, 1963. Volumes 1 and 2.

[Lew02] A. S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.*, 13(3):702–725 (electronic) (2003), 2002.

[LOS00] Claude Lemaréchal, François Oustry, and Claudia Sagastizábal. The $\mathcal{U}$-Lagrangian of a convex function. *Trans. Amer. Math. Soc.*, 352(2):711–729, 2000.

[MM05] Scott A. Miller and Jérôme Malick. Newton methods for nonsmooth convex minimization: connections among $\mathcal{U}$-Lagrangian, Riemannian Newton and SQP methods. *Math. Program.*, 104(2-3, Ser. B):609–633, 2005.

[MS00] Robert Mifflin and Claudia Sagastizábal. On $\mathcal{VU}$-theory for functions with primal-dual gradient structure. *SIAM J. Optim.*, 11(2):547–571 (electronic), 2000.

[NW06] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[O'N83] Barrett O'Neill. *Semi-Riemannian Geometry*, volume 103 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.

[Pol76] B. T. Polyak. Minimization methods with constraints. *J. Math. Sci.*, 5(1):97–128, 1976.

[Sak96] Takashi Sakai. *Riemannian Geometry*, volume 149 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1996. Translated from the 1992 Japanese original by the author.

[Shu86] Michael Shub. Some remarks on dynamical systems and numerical analysis. In L. Lara-Carrero and J. Lewowicz, editors, *Proc. VII ELAM.*, pages 69–92. Equinoccio, U. Simón Bolívar, Caracas, 1986.

[Smi94] Steven T. Smith. Optimization techniques on Riemannian manifolds. In Anthony Bloch, editor, *Hamiltonian and gradient flows, algorithms and control*, volume 3 of *Fields Inst. Commun.*, pages 113–136. Amer. Math. Soc., Providence, RI, 1994.

[Udr94] Constantin Udrişte. *Convex functions and optimization methods on Riemannian manifolds*, volume 297 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.

[WT04] Stephen J. Wright and Matthew J. Tenny. A feasible trust-region sequential quadratic programming algorithm. *SIAM J. Optim.*, 14(4):1074–1105 (electronic), 2004.