Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn
Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

P.-A. Absil. S. Hosseini

A collection of nonsmooth Riemannian optimization problems

INS Preprint No. 1713

September 2017

A Collection of Nonsmooth Riemannian Optimization Problems*

P.-A. Absil[†‡] S. Hosseini[§]

September 15, 2017

Abstract

Nonsmooth Riemannian optimization is a still scarcely explored subfield of optimization theory that concerns the general problem of minimizing (or maximizing), over a domain endowed with a manifold structure, a real-valued function that is not everywhere differentiable. The purpose of this paper is to illustrate, by means of nine concrete examples, that nonsmooth Riemannian optimization finds numerous applications in engineering and the sciences.

1 Introduction

Optimization on manifolds, also termed *Riemannian optimization*, concerns optimizing a real-valued function defined on a nonlinear search space endowed with a manifold structure. Manifolds that frequently arise in applications include the Stiefel manifold of orthonormal p-frames in $\mathbb{R}^n$, the Grassmann manifold of p-planes in $\mathbb{R}^n$, and the manifold of matrices of fixed rank and size. The area is motivated by numerous applications, notably in machine learning, but also in computer vision, imaging sciences, mechanics, physics, chemistry and genetics. A way to keep abreast of the rapidly evolving field of applications is to track papers that refer to general-purpose Riemannian optimization toolboxes such as Manopt [BMAS14], Pymanopt [TKW16], and ROPTLIB [HAGH16].

In considering an optimization problem on a Riemannian manifold, fundamental challenges occur due to the nonlinear nature of the search space. Indeed, any method based on the linear structure of the spaces—that is, virtually any method known in optimization—breaks down. A manifold is locally isomorphic to a linear space via chart maps, therefore one might ask whether it is possible to simply work in a chart domain and use classical linear algorithms to solve optimization problems on manifolds efficiently. Unfortunately, such an approach seldom leads to wieldy algorithms, because of the following reasons. First of all, symmetries of the underlying Riemannian manifold will in general not be respected by such algorithms. Second, while the existence of charts is inherent to the notion of manifold, they may not be readily available or they may be computationally impractical. Third, localizing to a chart inevitably

*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 Science Policy Office. This work was supported by the Belgian FNRS through grant FRFC 2.4585.12.

[†]Corresponding author

[‡]ICTEAM Institute, UCLouvain, B-1348 Louvain-la-Neuve, Belgium

[§]Hausdorff Center for Mathematics and Institute for Numerical Simulation, University of Bonn, 53115 Bonn, Germany

leads to distortions in the metric, which may cause slower convergence. Finally, for a chart-based algorithm to have access to the whole manifold, chart transition mechanisms need to be devised, and such mechanisms may prove to be intricate (see [SBCA13] for an example). For these reasons, together with the wealth of applications, Riemannian optimization has become a thriving area of research in the past few decades.

Early attempts to adapt standard optimization methods to smooth problems on *complete* Riemannian manifolds were presented by Gabay in 1982, who introduced the steepest descent, Newton and quasi-Newton algorithms on manifold settings. Furthermore, he provided the global and local convergence properties of the presented algorithms [Gab82]. Udriste also presented a steepest descent and a Newton algorithm for smooth objective functions on *complete* Riemannian manifolds and proved their (linear) convergence under the assumption of exact line search; see [Udr94]. Fairly recently, other versions of smooth optimization algorithms such as the steepest descent method, Newton's method, trust-region methods and conjugate gradient methods, have been extended to solving smooth optimization problems on *complete* Riemannian manifolds (see, e.g., [ADM$^+$02, AMS08, HGA15, Sat16]). These methods can be found in the general-purpose Riemannian optimization toolboxes mentioned above.

Regardless of the applications, a conceptual appeal of Riemannian optimization is that manifolds are arguably the natural generalization of $\mathbb{R}^n$ as the search space of *smooth* optimization problems: the notion of smoothness remains well defined for a real-valued function on an abstract manifold, and the notion of steepest descent direction follows from the Riemannian metric. In view of this, the concept of *nonsmooth* Riemannian optimization may seem paradoxical. However, it turns out that several important problems can be phrased as optimizing a *nonsmooth* function over a (smooth) Riemannian manifold.

The purpose of this paper is to give an overview of a few such problems that fit in the framework of nonsmooth Riemannian optimization (NRO). Our aim is to give just enough information for the reader to get a clear idea of what the problems are about and how they fit in the NRO framework, while referring to the original work for details.

The applications covered in this paper are not meant to form an exhaustive list. For example, Dirr et al. [DHL07, §3.4] present an application in grasping force optimization that we do not cover here, and our choice of applications is also complementary to [KGB16] where experimental results are shown for compressed modes, funcational correspondance, and robust Euclidean embedding.

Though NRO is a fairly recent and still scarcely explored area, a few general-purpose NRO methods are currently available. First, derivative-free techniques on manifolds (but not necessarily their convergence analysis) are readily applicable to NRO. A few derivative-free schemes such as Powell's method, direct local search methods and particle swarm optimization algorithms are available on complete Riemannian manifolds; see [BIA10, CSA15, Dre06]. Second, the NRO problems of interest have a locally Lipschitz—hence almost everywhere differentiable—objective function, and several methods exist that exploit this fact, such as gradient sampling [HU17, Hua13, §7.2], an ε-subdifferential algorithm [GH16b], a trust region algorithm [GH16a], nonsmooth BFGS methods [HHY16, Hua13, §7.3], and a manifold ADMM [KGB16]. In each section, we also give pointers to existing techniques to handle the specific problem considered.

2 Sparse PCA

Consider a data matrix $A = \begin{bmatrix} a_1 & \dots & a_j \end{bmatrix} \in \mathbb{R}^{m \times n}$, where m and n are the number of variables and observations, respectively. For example, in the case of gene expression data, A_{ij} may represent the expression level of gene i in experiment (or microarray) j.

Principal component analysis (PCA) relates to singular value decomposition (SVD) [GV96]. An SVD of A consists in expressing A as the matrix product $U\Sigma V^T$, where $p = \min\{m, n\}$, $U \in \mathbb{R}^{m \times p}$ is orthonormal ($U^T U = I$), $V \in \mathbb{R}^{n \times p}$ is also orthonormal, and $\Sigma \in \mathbb{R}^{p \times p}$ is a diagonal matrix with nonnegative diagonal entries sorted in decreasing order. The lth column of U, denoted by u_l, is then the *loading vector* of the lth principal component of A. In the gene expression application, the columns of U have been called *eigenarrays* and the columns of V, *eigengenes* [ABB00].

For simplicity, let us focus on the first loading vector, u_1, which has the remarkable property of giving the direction of the best fitting line in the least squares sense to the data points $a_1, \dots, a_n$, a concept that dates to Pearson [Pea01]. Equivalently, u_1 is solution of the optimization problem $\max_{u \in \mathbb{R}^m,\ \|u\|=1} u^T A A^T u$, which expresses that the columns of A are largest (in the mean square sense) along the direction of u_1. However, a possible drawback of u_1 is that its entries are typically nonzero. This may make the principal component—the projection onto the direction of u_1—unpleasantly heavy to compute, and hamper the interpretability of u_1 [Tre14].

Addressing this drawback leads to seek a new u that strikes a balance between the objectives of making $u^T A A^T u$ large and keeping small the cardinality $\|u\|_0$, i.e., the number of nonzero elements of u. Such a task can be approached in three ways: the Ivanov approach [Iva76] minimizes the data attachment term (here $u^T A A^T u$) under a constraint on the regularizer (here $\|u\|_0$); the Morozov approach [Mor84] minimizes the regularizer under a constraint on the data attachment term; the Tikhonov approach [TA77] mixes the two terms in the objective function. Let us focus on the last one, which yields the formulation

$$\begin{aligned} \max_{u \in \mathbb{R}^n} \quad & u^T A A^T u - \rho \|u\|_0 \\ \text{subject to} \quad & u^T u = 1, \end{aligned} \tag{1}$$

with the parameter $\rho \geq 0$. In problem (1), the objective function is not only nonsmooth; it is discontinuous, in view of the $\|u\|_0$ term. Optimizing a discontinuous function is an unpleasant task, hence the $\|u\|_0$ term is often relaxed to the sparsity-inducing (see, e.g., [Tib96, DT05]) ℓ_1 norm $\|u\|_1 := \sum_{i=1}^n |u_i|$, yielding the surrogate problem

$$\begin{aligned} \max_{u \in \mathbb{R}^n} \quad & u^T A A^T u - \rho \|u\|_1 \\ \text{subject to} \quad & u^T u = 1, \end{aligned} \tag{2}$$

a continuous but nonsmooth optimization problem on the unit sphere. Since the unit sphere is a submanifold of $\mathbb{R}^n$, (2) constitutes nonsmooth Riemannian optimization problem.

Techniques to handle this problem are discussed in [JTU03, ZHT06, dGJL07, JBAS10, JNRS10, GHT15] and references therein.

3 Secant-based dimensionality reduction

The purpose in this application is to pick the "easiest to invert" projection of a high-dimensional system. Let $\mathcal{S}$ be a subset of $\mathbb{R}^n$, which can be, e.g., the orbit of a dynamical

system, and let $\Sigma := \left\{ \frac{x-y}{\|x-y\|} : x, y \in \mathcal{S},\ x \neq y \right\}$ be its set of unit secants. We want to find a p-dimensional subspace $\mathcal{U}$ of $\mathbb{R}^n$ on which to orthogonally project the data. The projection is injective if Σ and the orthogonal complement of $\mathcal{U}$ have an empty intersection. In order to make the projection "as safely injective as possible", we look for the subspace $\mathcal{U}$ in the Grassmann manifold $\mathrm{Gr}(n, p)$ of p-planes in $\mathbb{R}^n$ that maximizes

$$f(\mathcal{U}) = \min_{s \in \Sigma} \|\pi_{\mathcal{U}} s\|,$$

where $\pi_{\mathcal{U}}$ denotes the orthogonal projection onto $\mathcal{U}$ and $\|\cdot\|$ denotes the Euclidean norm.

In practice, $\mathcal{S}$ comes to us as a finite set, e.g., the samples of a numerical integration of the dynamical system. The problem therefore consists in maximizing over the Grassmann manifold $\mathrm{Gr}(n, p)$ an objective function defined as the pointwise minimum of a finite collection of smooth functions. This is thus a nonsmooth optimization problem on a manifold.

This problem was addressed in [BK05] in the surrogate smooth form

$$\min_{\mathcal{U} \in \mathrm{Gr}(n,p)} \sum_{s \in \Sigma} \frac{1}{\|\pi_{\mathcal{U}} s\|}.$$

The only existing work that addresses the nonsmooth problem directly appears to be the very recent paper [Dre17] where a direct search method on the Grassmann manifold is advocated.

4 Economic load dispatch

The economic load dispatch problem consists in finding the most cost-effective repartition of power generation between production units in order to satisfy the demand, while accounting for transmission losses and keeping each unit within its allowed operating zone. Letting $p = \begin{bmatrix} p_1 & \dots & p_n \end{bmatrix}^T$ denote the power output of the n available units, a popular load dispatch model is

$$\min_{p \in \mathbb{R}^n} \quad f_{\mathrm{T}}(p) := \sum_{i=1}^{n} a_i p_i^2 + b_i p_i + c_i + \left| d_i \sin\left[e_i \left(p_i^{\min} - p_i \right) \right] \right| \tag{3a}$$

$$\text{subject to} \quad p_i^{\min} \leq p_i \leq p_i^{\max}, \quad i = 1, \dots, n, \tag{3b}$$

$$\sum_{i=1}^{n} p_i = p_{\mathrm{D}} + p_{\mathrm{L}}(p), \tag{3c}$$

where p_{D} is the power demand (MW) and $p_{\mathrm{L}}(p)$ stands for the power loss (MW) expressed as $p_{\mathrm{L}}(p) = p^T B p + p^T b^0 + b^{00}$. Coefficients B_{ij}, b_i^0, b^{00} are the transmission loss coefficients (B-coefficients) given by the elements of the square matrix B of size $n \times n$, the vector b^0 of length n, and the constant b^{00}, respectively. The matrix B is symmetric positive-definite, hence $p_{\mathrm{L}}(p)$ is a convex quadratic function of p, and thus the set of points that meet the power balance constraint (3c) form an ellipsoid, i.e., a submanifold of $\mathbb{R}^n$. The feasible set of (3) is thus the intersection of an ellipsoid and an axis-aligned box.

The objective function (3a) follows a frequently-encountered formulation where the contribution of each unit is made of a quadratic term and a rectified sine that models the so-called *valve-point effect* [WS93, WWS14]. The rectified term makes the objective function nonsmooth and turns (3) into a nonsmooth optimization problem on a Riemannian manifold with additional box constraints. The geometry of the problem was taken into account in [BSBA13] where a Riemannian subgradient approach is proposed.

5 Range-based independent component analysis

Independent component analysis (ICA) is the problem of extracting maximally independent linear combinations of given signals. An ICA algorithm typically consists of (i) a contrast function which measures the dependence between signals and (ii) an optimization method to minimize the contrast function. Since contrast functions are in principle invariant by scaling (the "measure of dependence" between signals should not be affected by scaling), it is common to break this indederminacy by imposing a normalization yielding, e.g., the *oblique manifold* $\mathcal{OB}(D)$ of $D \times D$ matrices with unit-norm columns [AG06, KS10, SGB15].

Letting matrix $M \in \mathbb{R}^{D\times N}$ contain the D signals of length N to be unmixed, a contrast function introduced in [VLV07] is

$$f(X; M) := \sum_{d=1}^{D} \log R(x_d^T M) - \log|\det X|, \tag{4}$$

where $X \in \mathcal{OB}(D)$ is the candidate unmixing matrix (yielding the candidate unmixed signals found in the rows of $X^T D$) and R returns the range of its argument. Vrins et al. [VLV07] propose to use a robust estimator of the range defined as

$$R([b_1, b_2, \ldots, b_N]) := \frac{1}{m}\sum_{r=1}^{m} R_r([b_1, b_2, \ldots, b_N])$$

where $R_r([b_1, b_2, \ldots, b_N]) := b_{(N-r+1)} - b_{(r)}$ and $b_{(i)}$ stands for the ith element of $[b_1, b_2, \ldots, b_N]$ by increasing values. We have thus obtained the problem

$$\begin{aligned} \min_{X} \quad & f(X; M) \\ \text{subject to} \quad & X \in \mathcal{OB}(D) \end{aligned} \tag{5}$$

which is a nonsmooth optimization problem over the oblique manifold. This problem was tackled by a derivative-free optimization method in [SBCA13] and by a nonsmooth quasi-Newton method in [SGB15]. The latter is a Riemannian generalization (see also [Hua13, Ch. 7]) of the nonsmooth BFGS method of Lewis and Overton [LO13].

6 Sphere packing on manifolds

Let $\mathcal{M}$ be an n-dimensional Riemannian manifold equipped with a Riemannian distance dist, and let $B(P, r)$ denote the ball with respect to this distance in $\mathcal{M}$ centered at P with radius r. The sphere packing problem aims at finding m points $P_1, \ldots, P_m$ in $\mathcal{M}$ such that

$$\max\{r : \quad B(P_i, r) \cap B(P_j, r) = \emptyset \quad \forall i < j\}, \tag{6}$$

is maximized. This problem (6) is equivalent to maximizing the following nonsmooth function,

$$F(P_1, \ldots, P_m) := \min_{1\le i<j\le m} \operatorname{dist}(P_i, P_j), \tag{7}$$

on $\mathcal{M} \times \ldots \times \mathcal{M}$. In [LH07] sphere packing on Grassmannians as an example of sphere packing on Riemannian manifolds is considered and a new formulation for the objective functions is

presented. Specifically, let the Grassmannian $\mathrm{Gr}(n,k)$ be the set of all k-dimensional linear subspaces of $\mathbb{R}^n$. It can be identified with

$$\{P \in \mathrm{S}(n): \quad P^2 = P \text{ and } \mathrm{trace}(P) = k\},$$

where $\mathrm{S}(n)$ denotes the set of $n \times n$ symmetric matrices. Moreover, assume that

$$\mathrm{dist}(P,Q) := \sqrt{\frac{1}{2}}\|P-Q\|_F,$$

where $\|.\|_F$ denotes the Frobenius norm. It is proved in [DHL07] that minimizing (7) is equivalent to minimizing the following nonsmooth function,

$$G(P_1,...,P_m) := \max_{1\le i<j\le m} \mathrm{trace}(P_iP_j),$$

on $\mathrm{Gr}(n,k) \times ... \times \mathrm{Gr}(n,k)$. Motivated by applications in multi-antenna channel communications [GD09, ZT02], techniques to solve this problem are discussed in [DHL07, LH07, GH16b, GH16a].

7 Robust low-rank matrix completion

The aim of low-rank matrix completion is to recover an unknown low-rank matrix by knowing only a small subset of its entries, which might be corrupted by noise or contain outliers. In the noiseless case, the problem could be stated as finding the lowest rank matrix X which matches the matrix M, which is the underlying matrix to be reconstructed, on some observed entries in a set Ω, i.e.,

$$X_{ij} = M_{ij} \quad (i,j) \in \Omega.$$

More explicitly,

$$\begin{aligned} \min_X \quad & \mathrm{rank}(X) \\ \text{subject to} \quad & X_{ij} = M_{ij} \quad (i,j) \in \Omega. \end{aligned} \tag{8}$$

If a matrix has rank r, then it has exactly r nonzero singular values, which means that the rank function in (8) is simply the number of nonzero singular values. Therefore, much as in section 2 where the cardinality was replaced by the ℓ_1 norm, one may consider a surrogate problem for (8), which minimizes the sum of the singular values, called nuclear norm, over the constraint set. As many of the existing applications for matrix completion involve very large data sets, it is therefore important to develop algorithms that can be applicable with such a large-scale setting, but, unfortunately, minimizing the nuclear norm by the existing methods for convex optimization scales very badly in the matrix dimension. This has motivated a considerable amount of algorithms that aim to minimize the nuclear norm relaxation by creating methods that use the low-rank structure of the solution, among which there exists an approach based on a direct optimization over the set of all fixed-rank matrices having assumed that the rank r of the target matrix is known in advance. Indeed, the problem can be stated as

$$\begin{aligned} \min_X \quad & \|P_\Omega(X-M)\|_F, \\ \text{subject to} \quad & X \in \mathcal{M}_r := \{X \in \mathbb{R}^{n\times m}: \quad \mathrm{rank}(X) = r\}, \end{aligned} \tag{9}$$

where P_Ω is the projector defined by $(P_\Omega(X))_{ij} = X_{ij}$ if $(i,j) \in \Omega$ and 0 otherwise, and $\|X\|_F$ denotes the Frobenius norm of X. It can be seen that this method is robust to Gaussian additive noise, in the sense that a small Gaussian additive noise still allows recovery of the underlying low-rank matrix with an error proportional to the noise level, however it is not well suited to recover the underlying low-rank matrix when some entries are corrupted by large errors, so-called outliers. To consider low-rank matrix completion in the presence of outliers, several methods have been proposed; see [CA16] and references therein. In [CA16] it has been shown that in this case the low-rank matrix completion can be solved by minimizing the ℓ_1 norm on the manifold of fixed rank matrices, i.e.,

$$\begin{aligned} \min_X \quad & \|P_\Omega(X - M)\|_{\ell_1}, \\ \text{subject to} \quad & X \in \mathcal{M}_r := \{X \in \mathbb{R}^{n\times m} : \quad \operatorname{rank}(X) = r\}, \end{aligned} \tag{10}$$

where $\|X\|_{\ell_1}$ denotes the sum of the absolute values of the entries of X. The adequacy of the framework of nonsmooth optimization on Riemannian manifolds here stems from the fact that the constraint set is a Riemannian manifold and the objective function is nonsmooth; see [CA16, HU17] for techniques solving this problem.

One of the most common applications of low-rank matrix completion is in recommender systems as in the Netflix Prize, for which the data is a big matrix with each entry recording the rating of customer i for movie j. The problem is how to predict the ratings that have not been made yet based on the current observation. A popular solution is to assume that the rating matrix is low-rank, an assumption that can be a priori motivated by the fact that it should be possible to accurately recover user profiles by linearly combining a few meta user profiles.

8 Finding the sparsest vector in a subspace

Suppose that a linear n-dimensional subspace W embedded in $\mathbb{R}^m$ contains a nonzero sparse vector. Given an arbitrary orthogonal basis of W, one aims to recover the sparsest vector in W. Letting $Q \in \mathbb{R}^{m\times n}$ denote a matrix whose columns form an orthonormal basis for W, the problem is as follows

$$\min \|Qx\|_0, \quad x \in \mathbb{S}, \tag{11}$$

where $\mathbb{S}$ is the Euclidean unit sphere in $\mathbb{R}^n$ and $\|Qx\|_0$ is the number of nonzero elements of Qx. As in most applications we only care about the solution up to scaling, it is natural to force x to live on the unit sphere. Since the objective function is not locally Lipschitz, we may replace it with the 1-norm as a surrogate; see [QSW16]. This leads to the problem

$$\min \|Qx\|_1, \quad x \in \mathbb{S}. \tag{12}$$

As the unit sphere is a submanifold of $\mathbb{R}^n$, this problem fits into the framework of nonsmooth optimization on Riemannian manifolds. Finding the sparsest vector in a subspace has several important applications for instance in sparse dictionary learning; [DH14]. Techniques to handle this problem are discussed in [QSW16, HU17] and references therein.

9 Restoring manifold-valued images

Consider an image $u : V \to \mathcal{M}$, where V is a set of pixels, which are usually on a two-dimensional grid, and $\mathcal{M}$ is a Riemannian manifold. In many applications, we have only

noisy measurements, i.e., $u^n = u + n$, where n denotes the noise. Furthermore, it sometimes happens that various pixel values are corrupted, which leads to the noisy measurements on a subset V_k of the pixel set V. The aim is to restore the image $u : V \to \mathcal{M}$, which can be considered as $u \in \mathcal{M}^V$, from partial and noisy data $u^n \in \mathcal{M}^{V_k}$. To this goal, one can consider the total variation (TV) of an image $u : V \to \mathcal{M}$ defined by

$$TV(u) := \sum_{(i,j)\in E} \operatorname{dist}(u_i, u_j),$$

where $E \subset V \times V$ is a given set of pairs of pixels that are close to each other and $\operatorname{dist} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ is the Riemannian distance on $\mathcal{M}$. The *TV denoising* approach consists in finding an image $u \in \mathcal{M}^V$ whose TV is relatively small while still being close to the noisy image $u^n \in \mathcal{M}^{V_k}$; see [ROF92]. This yields the problem

$$\min J(u) := \frac{1}{2} \sum_{i\in V_k} \operatorname{dist}^2(u_i, u_i^n) + \lambda TV(u), \quad u \in \mathcal{M}^V,$$

where $\lambda > 0$ is a positive number which balances the two parts of the functional. Since the distance function is not differentiable on the diagonal, the TV function is not differentiable. Therefore, this problem can be considered as a nonsmooth optimization problem on $\mathcal{M}^V$. We refer to [GS16, BBSW16, GH16a, GH16b] for techniques to solve this problem, which has applications, e.g., in in diffusion-tensor magnetic resonance imaging (where $\mathcal{M}$ is the set of positive-definite 3×3 matrices, aptly endowed with the so-called affine-invariant metric that turns it into a complete manifold) and in crystal lattice analysis (where $\mathcal{M}$ is the rotation group SO(3)).

10 Oriented bounding box

The oriented bounding box problem aims at finding a minimum volume parallelepipedic box containing a given collection of points in $\mathbb{R}^d$.

Suppose K points are given by a matrix $E \in \mathbb{R}^{d\times K}$, where the ith column contains the coordinates of the ith point. For a given orientation $Q \in \mathrm{SO}(3)$ of the box (i.e., the three columns of Q in the rotation group SO(3) give the direction of the three axes of the box), the minimal volume is readily found to be

$$f(Q) := \prod_{i=1}^{d} (e_{i,\max} - e_{i,\min}),$$

where $e_{i,\max}$ and $e_{i,\min}$ denote max and min entries, respectively, of the ith row of $Q^T E$. The oriented bounding box problem can thus be phrased as $\min_{Q\in\mathrm{SO}(3)} f(Q)$.

If there exists more than one entry at any row of $Q^T E$ reaching the maximum or minimum values, then the cost function f is not differentiable at Q. Therefore, the problem can be considered as a nonsmooth problem on a Riemannian manifold; see [BA10, HHY16] for techniques to solve this problem.

This problem arises in several applications in computer graphics [SE02], physical simulations [Eri04] and spatial data structures [GLM96], as well as other areas, such as computer-aided design [Cha03].

11 Conclusion

This paper has illustrated that NRO admits several important applications. NRO is moreover a challenging topic, as it calls for leveraging Riemannian optimization techniques to take advantage of the differentiable structure of the feasible set, while resorting to strategies to exploit the nonsmooth but almost everywhere smooth nature of the objective function. This mix of mathematical challenges and real-life applications makes NRO an exciting research area.

References

[ABB00] Orly Alter, Patrick O. Brown, and David Botstein. Singular value decomposition-expression data processingfor genome-wide and modeling. *PNAS*, 97(18):10101–10106, 2000. doi:10.1073/pnas.97.18.10101.

[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. doi:10.1093/imanum/22.3.359.

[AG06] P.-A. Absil and K. A. Gallivan. Joint diagonalization on the oblique manifold for independent component analysis. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 5, pages V–945–V–948, 2006.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008. URL: http://press.princeton.edu/titles/8586.html.

[BA10] Pierre B. Borckmans and P.-A. Absil. Fast oriented bounding box computation using particle swarm optimization. In *Proceedings of the 18th European Symposium on Artificial Neural Networks (ESANN)*, 2010.

[BBSW16] Miroslav Bačák, Ronny Bergmann, Gabriele Steidl, and Andreas Weinmann. A second order nonsmooth variational model for restoring manifold-valued images. *SIAM J. Sci. Comput.*, 38(1):A567–A597, 2016. doi:10.1137/15M101988X.

[BIA10] Pierre B. Borckmans, Mariya Ishteva, and P.-A. Absil. A modified particle swarm optimization algorithm for the best low multilinear rank approximation of higher-order tensors. In Marco Dorigo et al., editor, *Swarm Intelligence*, volume 6234 of *Lecture Notes in Computer Science*, pages 13–23. Springer-Verlag, 2010. doi:10.1007/978-3-642-15461-4_2.

[BK05] D. S. Broomhead and M. J. Kirby. Dimensionality reduction using secant-based projection methods: the induced dynamics in projected systems. *Nonlinear Dynam.*, 41(1-3):47–67, 2005. doi:10.1007/s11071-005-2792-1.

[BMAS14] Nicolas Boumal, Bamdev Mishra, P.-A. Absil, and Rodolphe Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning*

Research, 15:1455–1459, 2014. URL: http://jmlr.org/papers/v15/boumal14a.html.

[BSBA13] Pierre B. Borckmans, S. Easter Selvan, Nicolas Boumal, and P.-A. Absil. A Riemannian subgradient algorithm for economic dispatch with valve-point effect. *J. Comput. Applied. Math.*, 255:848–866, 2013. doi:10.1016/j.cam.2013.07.002.

[CA16] Léopold Cambier and P.-A. Absil. Robust low-rank matrix completion by Riemannian optimization. *SIAM Journal on Scientific Computing*, 38(5):S440–S460, 2016. doi:10.1137/15M1025153.

[Cha03] Chi Keung Chan. *Minimum Bounding Boxes and Volume Decomposition of Cad Models*. PhD thesis, University of Hong Kong (People's Republic of China), 2003.

[CSA15] A. Chattopadhyay, S. E. Selvan, and U. Amato. A derivative-free Riemannian Powell's method, minimizing Hartley-entropy-based ICA contrast. *IEEE Transactions on Neural Networks and Learning Systems*, PP(99):1–1, 2015. doi:10.1109/TNNLS.2015.2464157.

[dGJL07] Alexandre d'Aspremont, Laurent El Ghaoui, Michael I. Jordan, and Gert R. G. Lanckriet. A direct formulation for sparse pca using semidefinite programming. *SIAM Review*, 49(3):434–448, 2007. doi:10.1137/050645506.

[DH14] Laurent Demanet and Paul Hand. Scaling law for recovering the sparsest element in a subspace. *Information and Inference: A Journal of the IMA*, 3(4):295–309, 2014. doi:10.1093/imaiai/iau007.

[DHL07] G. Dirr, U. Helmke, and C. Lageman. Nonsmooth Riemannian optimization with applications to sphere packing and grasping. In *Lagrangian and Hamiltonian methods for nonlinear control 2006*, volume 366 of *Lecture Notes in Control and Inform. Sci.*, pages 29–45. Springer, Berlin, 2007.

[Dre06] David W. Dreisigmeyer. Direct search algorithms over Riemannian manifolds. Optimization Online 2007-08-1742, December 2006.

[Dre17] David W. Dreisigmeyer. A quasi-isometric embedding algorithm, 2017. arXiv:arXiv:1709.01972.

[DT05] David L. Donoho and Jared Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *PNAS*, 102(27):9446–9451, 2005. doi:10.1073/pnas.0502269102.

[Eri04] C. Ericson. *Real-Time Collision Detection*. CRC Press, Inc., Boca Raton, FL, USA, 2004.

[Gab82] D. Gabay. Minimizing a differentiable function over a differential manifold. *J. Optim. Theory Appl.*, 37(2):177–219, 1982. doi:10.1007/BF00934767.

[GD09] R. H. Gohary and T. N. Davidson. Noncoherent MIMO communication: Grassmannian constellations and efficient detection. *IEEE Trans. Inform. Theory*, 55(3):1176–1205, 2009. doi:10.1109/TIT.2008.2011512.

[GH16a] P. Grohs and S. Hosseini. Nonsmooth trust region algorithms for locally Lipschitz functions on Riemannian manifolds. *IMA J. Numer. Anal.*, 36(3):1167–1192, 2016. doi:10.1093/imanum/drv043.

[GH16b] P. Grohs and S. Hosseini. ε-subgradient algorithms for locally Lipschitz functions on Riemannian manifolds. *Adv. Comput. Math.*, 42(2):333–360, 2016. doi:10.1007/s10444-015-9426-z.

[GHT15] Matthieu Genicot, Wen Huang, and Nickolay T. Trendafilov. Weakly correlated sparse components with nearly orthonormal loadings. In *Geometric Science of Information*, pages 484–490, 2015. doi:10.1007/978-3-319-25040-3_52.

[GLM96] S. Gottschalk, M. C. Lin, and D. Manocha. Obbtree: A hierarchical structure for rapid interference detection. In *SIGGRAPH '96 Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 171–180. ACM, 1996.

[GS16] P. Grohs and M. Sprecher. Total variation regularization on Riemannian manifolds by iteratively reweighted minimization. *Inf. Inference*, 5(4):353–378, 2016. doi:10.1093/imaiai/iaw011.

[GV96] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[HAGH16] Wen Huang, P.-A. Absil, K. A. Gallivan, and Paul Hand. ROPTLIB: an object-oriented C++ library for optimization on Riemannian manifolds. Technical Report FSU16-14, Florida State University, 2016. URL: http://www.math.fsu.edu/~whuang2/papers/ROPTLIB.htm.

[HGA15] Wen Huang, K. A. Gallivan, and P.-A. Absil. A Broyden class of quasi-Newton methods for Riemannian optimization. *SIAM J. Optim.*, 25(3):1660–1685, 2015. doi:10.1137/140955483.

[HHY16] S. Hosseini, W. Huang, and R. Yousefpour. Line search algorithms for locally Lipschitz functions on Riemannian manifolds. November 2016. INS Preprint No. 1626.

[HU17] S. Hosseini and A. Uschmajew. A Riemannian gradient sampling algorithm for nonsmooth optimization on manifolds. *SIAM J. Optim.*, 27(1):173–189, 2017. doi:10.1137/16M1069298.

[Hua13] Wen Huang. *Optimization Algorithms on Riemannian Manifolds with Applications*. PhD thesis, Department of Mathematics, Florida State University, November 2013. URL: http://diginole.lib.fsu.edu/etd/8809/.

[Iva76] V. V. Ivanov. *The theory of approximate methods and their application to the numerical solution of singular integral equations*. Noordhoff International Publishing, Schuttersveld 9, P. O. Box 26, Leyden, The Netherlands, 1976.

[JBAS10] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization on the cone of positive semidefinite matrices. *SIAM J. Optim*, 20(5):2327–2351, 2010. doi:10.1137/080731359.

[JNRS10] Michel Journée, Yurii Nesterov, Peter Richtárik, and Rodolphe Sepulchre. Generalized power method for sparse principal component analysis. *J. Mach. Learn. Res.*, 11:517–553, 2010. URL: http://www.jmlr.org/papers/v11/journee10a.html.

[JTU03] Ian T Jolliffe, Nickolay T Trendafilov, and Mudassir Uddin. A modified principal component technique based on the lasso. *Journal of Computational and Graphical Statistics*, 12(3):531–547, 2003. doi:10.1198/1061860032148.

[KGB16] Artiom Kovnatsky, Klaus Glashoff, and Michael M. Bronstein. MADMM: A generic algorithm for non-smooth optimization on manifolds. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V*, pages 680–696. Springer International Publishing, Cham, 2016. doi:10.1007/978-3-319-46454-1_41.

[KS10] Martin Kleinsteuber and Hao Shen. Intrinsic Newton's method on oblique manifolds for overdetermined blind source separation. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems – MTNS*, pages 2139–2143, 2010.

[LH07] C. Lageman and U. Helmke. Optimization methods for the sphere packing problem on Grassmannians. In *Taming heterogeneity and complexity of embedded control*, pages 393–408. ISTE, London, 2007. doi:10.1002/9780470612217.ch22.

[LO13] Adrian S. Lewis and Michael L. Overton. Nonsmooth optimization via quasi-Newton methods. *Mathematical Programming*, 141(1):135–163, 2013. doi:10.1007/s10107-012-0514-2.

[Mor84] V. Morozov. *Methods for Solving Incorrectly Posed Problems*. Springer-Verlag, 1984.

[Pea01] Karl Pearson. LIII. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(11):559–572, 1901. doi:10.1080/14786440109462720.

[QSW16] Q. Qu, J. Sun, and J. Wright. Finding a sparse vector in a subspace: linear sparsity using alternating directions. *IEEE Trans. Inform. Theory*, 62(10):5855–5880, 2016. doi:10.1109/TIT.2016.2601599.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991). doi:10.1016/0167-2789(92)90242-F.

[Sat16] Hiroyuki Sato. A Dai-Yuan-type Riemannian conjugate gradient method with the weak wolfe conditions. *Computational Optimization and Applications*, 64(2):101–118, 2016. arXiv:1405.4371, doi:10.1007/s10589-015-9801-1.

[SBCA13] S. Easter Selvan, Pierre B. Borckmans, A. Chattopadhyay, and P.-A. Absil. Spherical mesh adaptive direct search for separating quasi-uncorrelated sources by range-based independent component analysis. *Neural Comput.*, 25(9):2486–2522, 2013. doi:10.1016/0167-2789(92)90242-F.

[SE02] P. J. Schneider and D. Eberly. *Geometric Tools for Computer Graphics*. Elsevier Science Inc., New York, NY, USA, 2002.

[SGB15] S. Easter Selvan, S. Thomas George, and R. Balakrishnan. Range-based ICA using a nonsmooth quasi-Newton optimizer for electroencephalographic source localization in focal epilepsy. *Neural Computation*, 27(3):628–671, 2015. doi:10.1162/NECO_a_00700.

[TA77] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill posed problems*. V. H. Winston and Sons, 1977.

[Tib96] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996. URL: http://www.jstor.org/stable/2346178.

[TKW16] James Townsend, Niklas Koep, and Sebastian Weichwald. Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation. *Journal of Machine Learning Research*, 17(137):1–5, 2016. URL: http://jmlr.org/papers/v17/16-177.html.

[Tre14] Nickolay T. Trendafilov. From simple structure to sparse components: a review. *Computational Statistics*, 29(3):431–454, Jun 2014. doi:10.1007/s00180-013-0434-5.

[Udr94] C. Udriste. *Convex functions and optimization methods on Riemannian manifolds*, volume 297. Kluwer Academic Publishers Group, Dordrecht, 1994.

[VLV07] F. Vrins, J. A. Lee, and M. Verleysen. A minimum-range approach to blind extraction of bounded sources. *IEEE Transactions on Neural Networks*, 18(3):809–822, May 2007. doi:10.1109/TNN.2006.889941.

[WS93] D.C. Walters and G.B. Sheble. Genetic algorithm solution of economic dispatch with valve point loading. *Power Systems, IEEE Transactions on*, 8(3):1325–1332, Aug 1993. doi:10.1109/59.260861.

[WWS14] Allen J. Wood, Bruce F. Wollenberg, and Gerald B. Sheblé. *Power Generation Operation and Control*. John Wiley & Sons, Hoboken, New Jersey, third edition, 2014.

[ZHT06] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, 2006. doi:10.1198/106186006X113430.

[ZT02] L. Zheng and D. N. C. Tse. Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. Inform. Theory*, 48(2):359–383, 2002. doi:10.1109/18.978730.