# Iterative Methods for Comparing Two Matrices by Means of Isometric Projections*

Thomas Cason, Pierre-Antoine Absil, and Paul Van Dooren

Université catholique de Louvain, Department of Mathematical Engineering
Bâtiment Euler, avenue Georges Lemaître 4, 1348 Louvain-la-Neuve, Belgium
http://www.inma.ucl.ac.be/{∼cason, ∼absil, ∼vandooren}

**Abstract.** In this paper, we go over a number of manifold-based iterative methods for solving optimization problems developed in order to compare arbitrary square matrices, possibly of different order.

## 1 Introduction

When comparing two square matrices, it is often natural to allow for a class of transformations acting on them. In [1] and [2], the authors propose to compare two matrices $A$ and $B$ possibly of different order (say $m$ and $n$) through their restriction on a lower dimensional subspace:

$$U^*AU \quad \text{and} \quad V^*BV \,,$$

with $U$ and $V$ belonging to $St(k,m)$ and $St(k,n)$ respectively, and where $St(k,m) = \left\{U \in \mathbb{C}^{m\times k} : U^*U = I_k\right\}$ denotes the *compact Stiefel manifold.*

Fraikin et *al.* [1, 3] propose in this context to maximize the inner product between the *isometric projections*, $U^*AU$ and $V^*BV$, namely:

$$\max_{\substack{U^*U=I_k\\ V^*V=I_k}} \langle U^*AU, V^*BV\rangle := \Re\,\mathrm{tr}\left((U^*AU)^*\,(V^*BV)\right) \,.$$

An other approach consists in looking at the minimization of some distance between the restrictions of $A$ and $B$ on a lower dimensional subspace. Cason et *al.* [2] considered the squared *Frobenius distance* defined by

$$\mathrm{dist}^2\,(M,N) = \|M-N\|_F^2 = \mathrm{tr}\,((M-N)^*(M-N)) \,,$$

and analyzed minimization problems that are essentially the counterparts of the inner product maximization defined above:

$$\min_{\substack{U^*U=I_k\\ V^*V=I_k}} \mathrm{dist}^2\,(U^*AU, V^*BV) \,,$$

* 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.

$$\min_{\substack{X = VU^* \\ U^*U = I_k \\ V^*V = I_k}} \operatorname{dist}^2(XA, BX) \ ,$$

and

$$\min_{\substack{X = VU^* \\ U^*U = I_k \\ V^*V = I_k}} \operatorname{dist}^2(X, BXA^*) \ .$$

Here, we further investigate and analyze some manifold-based algorithms to solve these problems.

## 2 Optimization on manifolds

All those problems are defined on feasible sets that have a *manifold* structure. Roughly speaking, this means that the feasible set is locally smoothly identified with $\mathbb{R}^d$, where $d$ is the dimension of the manifold (see [4] for details). Optimization on a manifold generalizes optimization in $\mathbb{R}^d$ while retaining the concept of smoothness. For this purpose it is often useful to consider the tangent space at a point of the manifold $\mathcal{M}$. Indeed, the tangent space is a linear space on which one can apply many results from the theory of optimization in $\mathbb{R}^d$. When the tangent spaces to a manifold $\mathcal{M}$ at $M$, denoted by $T_M\mathcal{M}$, are equiped with an inner product $g_M$ that varies smoothly with $M$, the manifold $\mathcal{M}$ becomes a Riemannian manifold. Let $M$ be an element of $\mathcal{M}$, a Riemannian submanifold of $\mathcal{E}$ and $\hat{f}$ a smooth real-valued function defined on a neighborhood of $M$ in the embedding space and $f$ the restriction of $\hat{f}$ on $\mathcal{M}$. The gradient of $f$ at $M$, denoted $\operatorname{grad} f(M)$, is then defined as the unique element of the tangent plane $T_M\mathcal{M} \subset \mathcal{E}$, that satisfies

$$\mathrm{D}\,\hat{f}(M) \cdot \xi_M = g_M(\operatorname{grad} f(M), \xi_M) \,, \qquad \forall \xi_M \in T_M\mathcal{M} \,,$$

where D is the directional derivative operator. The first-order necessary condition for a point $M_*$ to be optimal is $\operatorname{grad} f(M_*) = 0$. Finding a point where $\operatorname{grad} f$ vanishes can be done using iterative methods. These methods starts from a given point on the manifold and build a sequence of iterates that converge toward the stationary point.

A well known class of iterative methods are line-search methods. In $\mathbb{R}^d$, choose a starting point $x_0$ and proceed through the following iteration

$$x_{k+1} = x_k + t_k z_k \,,$$

where $z_k$ is a suitable *search direction* and $t_k$ a scalar called the *step size*. This iteration can be generalized as follows on manifolds :

$$M_{k+1} = R_{M_k}(t_k \xi_k) \ ,$$

where $\xi_k$ is a tangent vector and $R_{M_k}$ is a function called the *retraction* that "rolls" $t_k\xi_k$ on the manifold (see [4] for details). For minimization problems, one may choose the opposite of the gradient as search direction

$$\xi_k = -\mathrm{grad}\, f(M_k)\,.$$

This particular case is known as the *steepest descent* method. The step size can further be set using the so-called *Armijo Back-Tracking* scheme, for example.

Some iterative methods, like Newton's method, use higher-order derivatives of $f$. Solving $\mathrm{grad}\, f(M) = 0$ in $\mathbb{R}^d$ with Newton's method iteration scheme leads to

$$\mathrm{Hess}\, f\,(M_k) \cdot (M_{k+1} - M_k) = -\mathrm{grad}\, f\,(M_k)\ .$$

where $\mathrm{Hess}\, f(M) \cdot z$ (the Hessian of $f$ in the direction of $z$) is $\mathrm{D}\,\mathrm{grad}\, f(M) \cdot z$. On a manifold, the concept of directional derivative is generalized by the so-called affine connection. Let $\mathfrak{X}(\mathcal{M})$ denote the set of smooth vector fields on a manifold $\mathcal{M}$. An affine connection $\nabla$ is a bilinear mapping

$$\nabla : T_M\mathcal{M} \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M}) : \eta, \xi \mapsto \nabla_\eta \xi\,,$$

which satisfies Leibniz's law [4] and the Hessian of $f$ at $M \in \mathcal{M}$ in the direction $\xi \in T_M\mathcal{M}$ is defined as

$$\mathrm{Hess}\, f(M) \cdot \eta = \nabla_\eta\, \mathrm{grad}\, f(M)\,.$$

An iteration of the generalized Newton method becomes (see [5, 4])

1. Find $\eta_k \in T_{M_k}\mathcal{M}$ solution of the Newton equation

$$\mathrm{Hess}\, f\,(M_k) \cdot \eta_k = -\mathrm{grad}\, f\,(M_k)\ .$$

2. "Roll" $\eta_k$ on $\mathcal{M}$ in order to find the next iterate: $M_{k+1} = R_{M_k}\eta_k$.

Finally, we consider some *conjugate gradient* methods on manifolds [6, 4, 7]. These are line-search methods in which consecutive search directions are conjugate with respect to the Hessian of $f$ or some approximation of it. For some choices of the approximate conjugacy condition and the stepsize selection procedure, the method provably converges globally to stationary points of the cost function, with local $d$-step superlinear convergence.

## References


1. Fraikin, C., Nesterov, Y., Van Dooren, P.: Optimizing the coupling between two isometric projections of matrices. SIAM J. Matrix Anal. Appl. (2007)
2. Cason, T., Absil, P.-A., Van Dooren, P.: Comparing two matrices by means of isometric projections. *Summitted to* Numerical Linear Algebra in Signals, Systems and Control (2008)
3. Fraikin, C., Nesterov, Y., Van Dooren, P.: A gradient-type algorithm optimizing the coupling between matrices. Proceedings of the 13-th ILAS conference in Amsterdam (2006)

4. Absil, P.-A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, New Jersey (January 2008)
5. Adler, R.L., Dedieu, J.P., Margulies, J.Y., Martens, M., Shub, M.: Newton's method on Riemannian manifolds and a geometric model for the human spine. IMA J. Numer. Anal. **22**(3) (July 2002) 359–390
6. Smith, S.T.: Optimization techniques on Riemannian manifolds. In: Hamiltonian and gradient flows, algorithms and control. Volume 3 of Fields Inst. Commun. Amer. Math. Soc., Providence, RI (1994) 113–136
7. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2) (1998) 303–353