Optimization On Manifolds: Methods and Applications

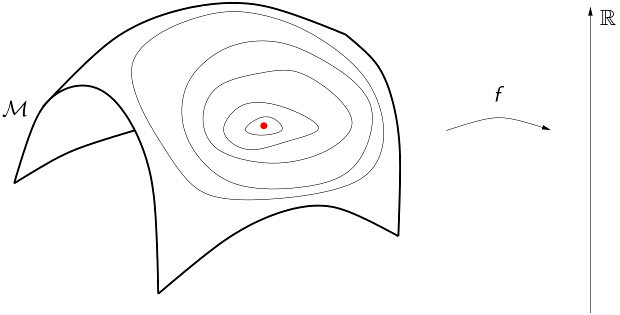
Pierre-Antoine Absil
(UCLouvain)

BFG'09, Leuven
18 Sep 2009

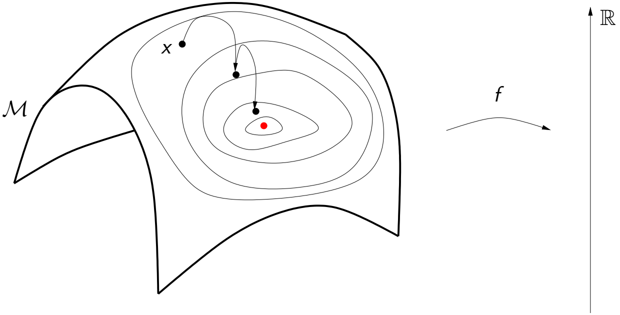
Joint work with:

- Chris Baker (Sandia)
- Kyle Gallivan (Florida State University)
- Eric Klassen (Florida State University)
- Damien Laurent (UCLouvain)
- Rob Mahony (Australian National University)
- Chafik Samir (U Clermont-Ferrand)
- Rodolphe Sepulchre (U of Liège)
- Anuj Srivastava (Florida State University)
- Paul Van Dooren (UCLouvain)
- ...

Optimization on Manifolds in one picture



Optimization on Manifolds in one picture

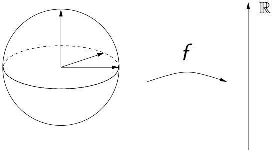


Why general manifolds? – Motivating examples

Given $A = A^T \in \mathbb{R}^{n \times n}$
and $N = \mathrm{diag}(1, \ldots, p)$,

$\min f(X) = \mathrm{trace}(X^T A X N)$

subj. to $X \in \mathbb{R}^{n \times p} : X^T X = I$



Domain: $\mathrm{St}(p, n)$

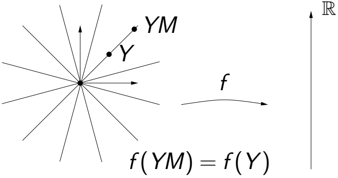
$= \{X \in \mathbb{R}^{n \times p} : X^T X = I\}$

Embedded submanifold

Given $A = A^T \in \mathbb{R}^{n \times n}$,

$\min f(Y) = \mathrm{trace}\left((Y^T Y)^{-1}(Y^T A Y)\right)$

subj. to $Y \in \mathbb{R}_*^{n \times p}$ (i.e., Y full rank)

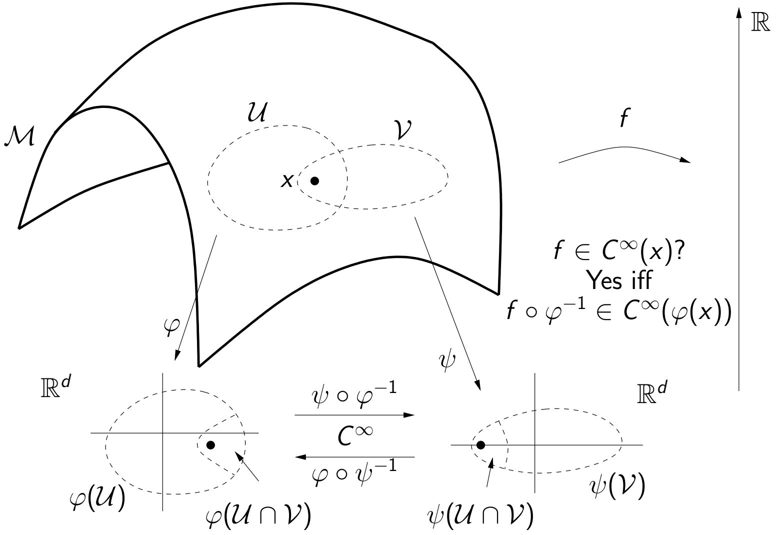


Domain: $\mathrm{Gr}(p, n)$

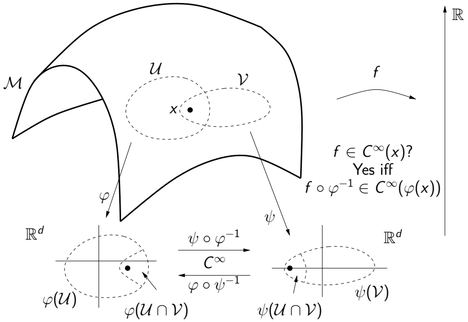
$= \left\{\{YM : M \in \mathbb{R}_*^{p \times p}\} : Y \in \mathrm{R}_*^{n \times p}\right\}$

Quotient manifold

Smooth optimization problems on general manifolds



Optimization on manifolds in its most abstract formulation

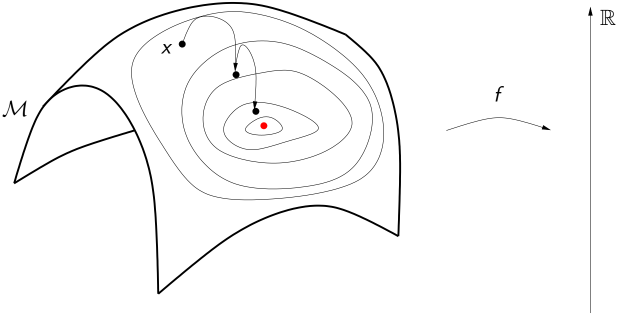


Given:

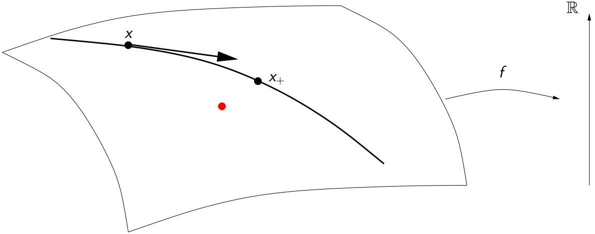
- A set $\mathcal{M}$ endowed (explicitly or implicitly) with a manifold structure (i.e., a collection of compatible charts).
- A function $f : \mathcal{M} \to \mathbb{R}$, smooth in the sense of the manifold structure.

Task: Compute a local minimizer of f.

Optimization on Manifolds in one picture



Some classics on Optimization On Manifolds (I)



Luenberger (1973), *Introduction to linear and nonlinear programming*. Luenberger mentions the idea of performing line search along geodesics, “which we would use if it were computationally feasible (which it definitely is not)”.

Some classics on Optimization On Manifolds (II)

Gabay (1982), *Minimizing a differentiable function over a differential manifold*. Stepest descent along geodesics; Newton's method along geodesics; Quasi-Newton methods along geodesics.

Smith (1994), *Optimization techniques on Riemannian manifolds*. Levi-Civita connection ∇; Riemannian exponential; parallel translation. But Remark 4.9: If Algorithm 4.7 (Newton's iteration on the sphere for the Rayleigh quotient) is simplified by replacing the exponential update with the update

$$x_{k+1} = \frac{x_k + \eta_k}{\|x_k + \eta_k\|}$$

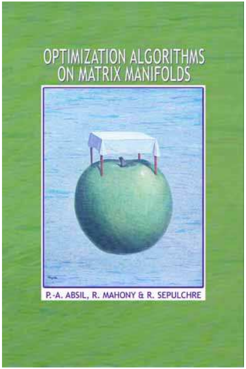
then we obtain the Rayleigh quotient iteration.

Some classics on Optimization On Manifolds (III)

Manton (2002), *Optimization algorithms exploiting unitary constraints* "The present paper breaks with tradition by not moving along geodesics". The geodesic update $\mathrm{Exp}_x\eta$ is replaced by a projective update $\pi(x+\eta)$, the *projection* of the point $x+\eta$ onto the manifold.

Adler, Dedieu, Shub, et al. (2002), *Newton's method on Riemannian manifolds and a geometric model for the human spine*. The exponential update is relaxed to the general notion of *retraction*. The geodesic can be replaced by any (smoothly prescribed) curve tangent to the search direction.

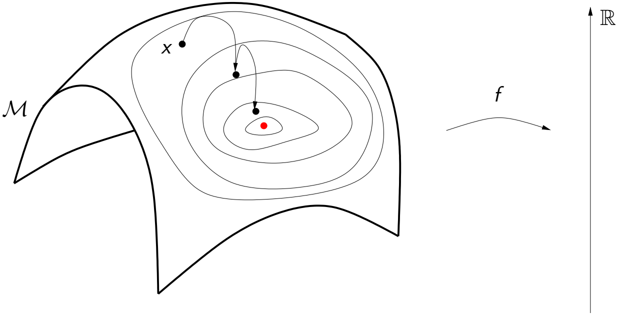
A recent book



Optimization Algorithms on Matrix Manifolds
P.-A. Absil, R. Mahony, R. Sepulchre
Princeton University Press, January 2008

1. Introduction
2. Motivation and applications
3. Matrix manifolds: first-order geometry
4. Line-search algorithms
5. Matrix manifolds: second-order geometry
6. Newton's method
7. Trust-region methods
8. A constellation of superlinear algorithms

Optimization on Manifolds in one picture



Specific manifolds, and where they appear

- Stiefel manifold $\mathrm{St}(p, n)$ and orthogonal group $O_p = \mathrm{St}(n, n)$

$$\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n\times p} : X^T X = I_p\}$$

  Applications: computer vision; principal component analysis; independent component analysis...
- Grassmann manifold $\mathrm{Gr}(p, n)$

  Set of all p-dimensional subspaces of $\mathbb{R}^n$

  Applications: various dimension reduction problems...
- Set of fixed-rank PSD matrices $\mathrm{S}_+(p, n)$. A quotient representation:

$$X \sim Y \Leftrightarrow \exists Q \in O_p : Y = XQ$$

  Applications: Low-rank approximation of symmetric matrices; low-rank approximation of tensors...

Specific manifolds, and where they appear

- Shape manifold $O_n \backslash \mathbb{R}_*^{n\times p}$

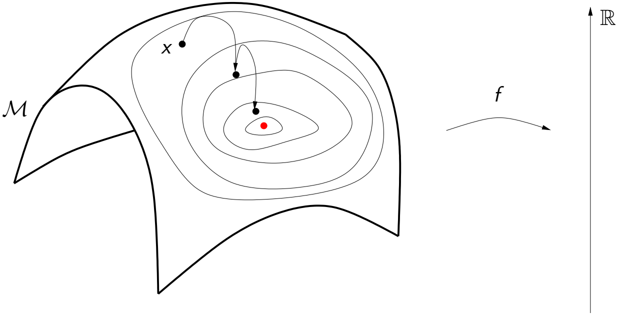
$$Y \sim X \Leftrightarrow \exists U \in O_n : Y = UX$$

  Applications: shape analysis
- Oblique manifold $\mathbb{R}_*^{n\times p}/\mathcal{S}_{\mathrm{diag}+}$

$$\mathbb{R}_*^{n\times p}/\mathcal{S}_{\mathrm{diag}+} \simeq \{Y \in \mathbb{R}_*^{n\times p} : \mathrm{diag}(Y^T Y) = I_p\}$$

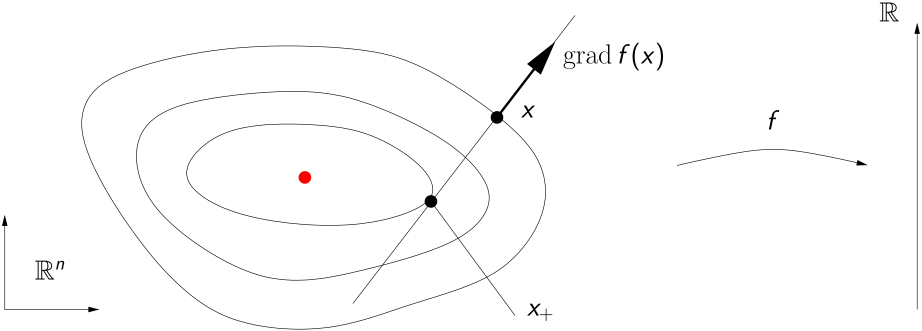
  Applications: blind source separation; factor analysis (oblique Procrustes problem)...
- Flag manifold $\mathbb{R}_*^{n\times p}/\mathcal{S}_{\mathrm{upp}_*}$
  Elements of the flag manifold can be viewed as a p-tuple of linear subspaces $(\mathcal{V}_1, \ldots, \mathcal{V}_p)$ such that $\dim(\mathcal{V}_i) = i$ and $\mathcal{V}_i \subset \mathcal{V}_{i+1}$.
  Applications: analysis of QR algorithm...

Optimization on Manifolds in one picture



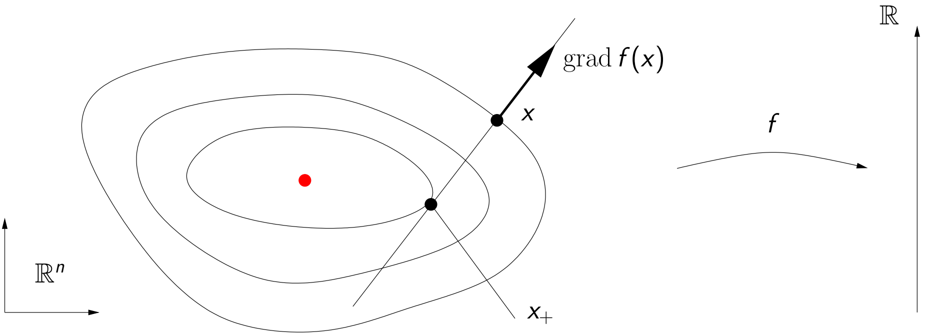
Steepest-descent methods on manifolds

Steepest-descent in $\mathbb{R}^n$



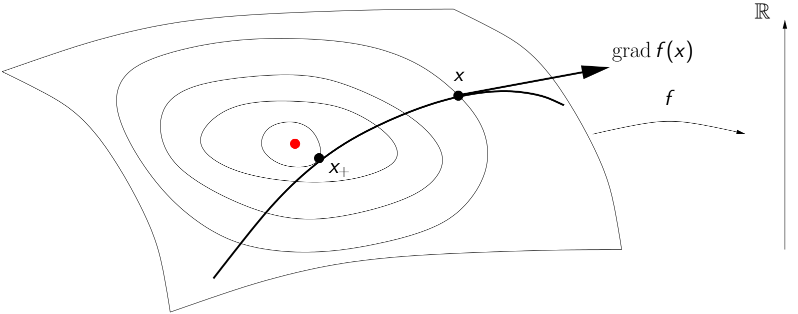
$$\operatorname{grad} f(x) = \begin{bmatrix} \partial_1 f(x) & \cdots & \partial_n f(x) \end{bmatrix}^T$$

Steepest-descent: from $\mathbb{R}^n$ to manifolds



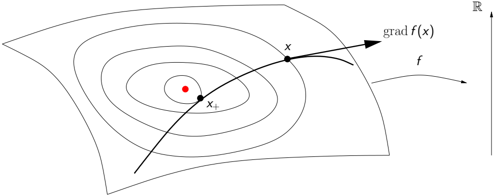
	$\mathbb{R}^n$	Manifold
Search direction	Vector at x	Tangent vector at x
Steepest-desc. dir.	$-\operatorname{grad} f(x)$	$-\operatorname{grad} f(x)$
Curve	$\gamma : t \mapsto x - t \operatorname{grad} f(x)$	γ s.t. $\gamma(0) = x$ and $\dot{\gamma}(0) = -\operatorname{grad} f(x)$

Steepest-descent: from $\mathbb{R}^n$ to manifolds



	$\mathbb{R}^n$	Manifold
Search direction	Vector at x	Tangent vector at x
Steepest-desc. dir.	$-\operatorname{grad} f(x)$	$-\operatorname{grad} f(x)$
Curve	$\gamma : t \mapsto x - t \operatorname{grad} f(x)$	γ s.t. $\gamma(0) = x$ and $\dot{\gamma}(0) = -\operatorname{grad} f(x)$

Steepest-descent: norm of tangent vectors



The steepest ascent direction at x is along

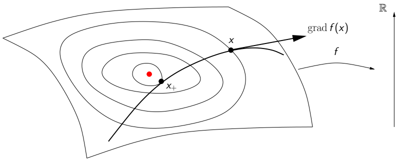
$$\underset{\substack{\xi \in T_x\mathcal{M} \\ \|\xi\|=1}}{\arg\max}\, \mathrm{D}f(x)[\xi].$$

To this end, we need a norm on the tangent space $T_x\mathcal{M}$.
For all $x \in \mathcal{M}$, let g_x denote an inner product in $T_x\mathcal{M}$, and define

$$\|\xi_x\| := \sqrt{g_x(\xi_x, \xi_x)}.$$

When g_x "smoothly" depends on x, we say that $(\mathcal{M}, g)$ is a *Riemannian manifold*.

Steepest-descent: gradient



There is a unique $\operatorname{grad} f(x)$, called the *gradient* of f at x, such that

$$\begin{cases} \operatorname{grad} f(x) \in T_x\mathcal{M} \\ g_x(\operatorname{grad} f(x), \xi_x) = \mathrm{D}f(x)[\xi_x], \quad \forall \xi_x \in T_x\mathcal{M}. \end{cases}$$

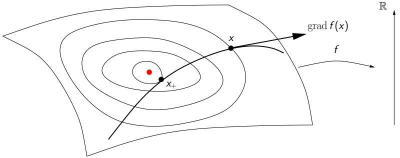
We have

$$\frac{\operatorname{grad} f(x)}{\|\operatorname{grad} f(x)\|} = \underset{\substack{\xi \in T_x\mathcal{M} \\ \|\xi\|=1}}{\arg\max}\, \mathrm{D}f(x)[\xi]$$

and

$$\|\operatorname{grad} f(x)\| = \mathrm{D}f(x)\left[\frac{\operatorname{grad} f(x)}{\|\operatorname{grad} f(x)\|}\right].$$

Steepest-descent: Riemannian submanifolds



Let $(\overline{\mathcal{M}}, \overline{g})$ be a Riemannian manifold and $\mathcal{M}$ be a submanifold of $\overline{\mathcal{M}}$. Then

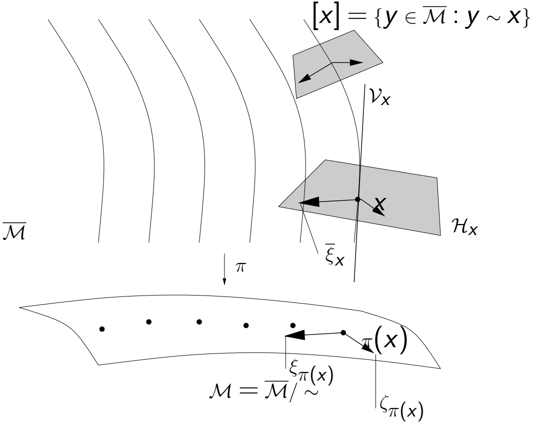
$$g_x(\xi_x, \zeta_x) := \overline{g}_x(\xi_x, \eta_x), \ \forall \xi_x, \zeta_x \in T_x\mathcal{M}$$

defines a Riemannian metric g on $\mathcal{M}$. With this Riemannian metric, $\mathcal{M}$ is a *Riemannian submanifold* of $\overline{\mathcal{M}}$.
Every $z \in T_x\overline{\mathcal{M}}$ admits a decomposition $z = \underbrace{\mathrm{P}_x z}_{\in T_x\mathcal{M}} + \underbrace{\mathrm{P}_x^{\perp} z}_{\in T_x^{\perp}\mathcal{M}}$.

If $\overline{f} : \overline{\mathcal{M}} \to \mathbb{R}$ and $f = \overline{f}|_{\mathcal{M}}$, then

$$\mathrm{grad}\, f(x) = \mathrm{P}_x \mathrm{grad}\, \overline{f}(x).$$

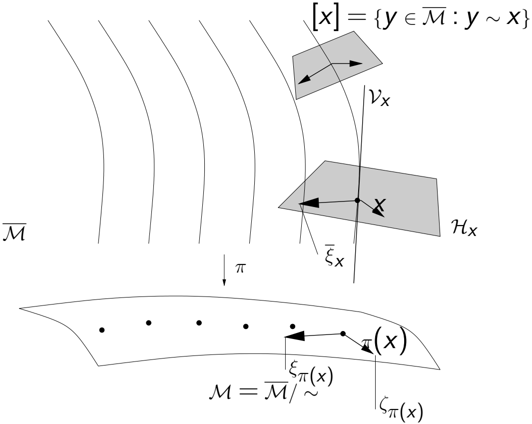
Steepest-descent: Riemannian quotient manifolds



Let $\tilde{g}$ be a Riemannian metric on $\overline{\mathcal{M}}$.
Suppose that, for all $\xi_{\pi(x)}$ and $\zeta_{\pi(x)}$ in $T_{\pi(x)}\overline{\mathcal{M}}/\sim$, and all $\tilde{x} \in \pi^{-1}(\pi(x))$, we have

$$\overline{g}_{\tilde{x}}(\overline{\xi}_{\tilde{x}}, \overline{\zeta}_{\tilde{x}}) = \overline{g}_x(\overline{\xi}_x, \overline{\zeta}_x).$$

Steepest-descent: Riemannian quotient manifolds

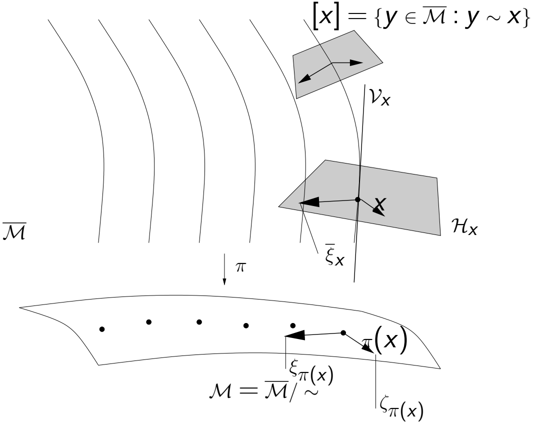


Then

$$g_{\pi(x)}(\xi_{\pi(x)}, \zeta_{\pi(x)}) := \overline{g}_x(\overline{\xi}_x, \overline{\zeta}_x).$$

defines a Riemannian metric on $\overline{\mathcal{M}}/\sim$. This turns $\overline{\mathcal{M}}/\sim$ into a Riemannian quotient manifold.

Steepest-descent: Riemannian quotient manifolds



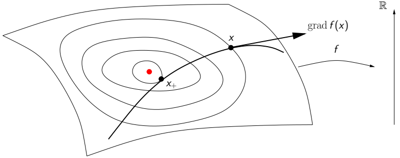
Let $f : \overline{\mathcal{M}}/\sim \to \mathbb{R}$. Let $\mathrm{P}_x^{h,\overline{g}}$ denote the orthogonal projection onto $\mathcal{H}_x$.

$$\overline{\mathrm{grad}\, f}_x = \mathrm{P}_x^{h,\overline{g}} \mathrm{grad}\,(f \circ \pi)(x).$$

If $\mathcal{H}_x$ is the orthogonal complement of $\mathcal{V}_x$ in the sense of $\overline{g}$ (π is a *Riemannian submersion*), then $\mathrm{grad}\,(f \circ \pi)(x)$ is already in $\mathcal{H}_x$, and thus

$$\overline{\mathrm{grad}\, f}_x = \mathrm{grad}\,(f \circ \pi)(x).$$

Steepest-descent: choosing the search curve



It remains to choose a curve γ such that

$$\begin{cases} \gamma(0) = x \\ \dot{\gamma}(0) = -\operatorname{grad} f(x). \end{cases} \tag{1}$$

Let $R : T\mathcal{M} \to \mathcal{M}$ be a *retraction* on $\mathcal{M}$, that is

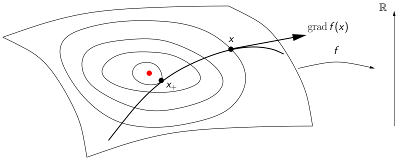
1. $R(0_x) = x$, where 0_x denotes the origin of $T_x\mathcal{M}$;
2. $\frac{\mathrm{d}}{\mathrm{d}t} R(t\xi_x)\big|_{t=0} = \xi_x$, for all $\xi_x \in T_x\mathcal{M}$.

Then the curve

$$\gamma : t \mapsto R(-t\operatorname{grad} f(x))$$

satisfies the requirements (1).

Steepest-descent: line-search procedure

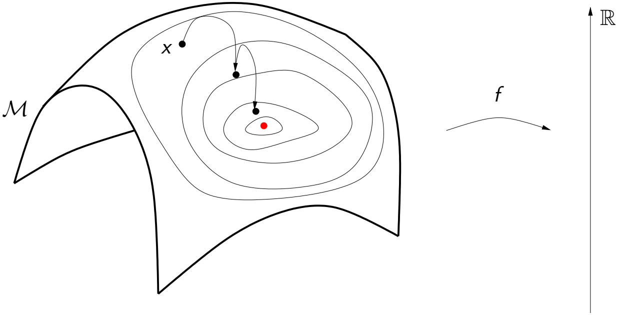


Find t such that $f(\gamma(t))$ is “sufficiently smaller” than $f(\gamma(0))$. Since

$$t \mapsto f(\gamma(t))$$

is just a function from $\mathbb{R}$ to $\mathbb{R}$, we can use the step selection techniques that are available for classical line-search methods.
For example: exact minimization, Armijo backtracking,...

Optimization on Manifolds in one picture



A simple steepest-descent method for Principal Component Analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the p dominant eigenvectors of A.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^T X = I_p} \mathrm{trace}(X^T AXN).$$

  The columns of X are the p dominant eigenvectors or A.
- A basic method: Steepest-ascent on $\mathrm{St}(p,n) := \{X \in \mathbb{R}^{n\times p} : X^T X = I\}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{trace}(X^T AXN)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X) = \mathcal{P}_{T_X \mathrm{St}(p,n)}(AXN) = AXN - X\mathrm{sym}(X^T AXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X)$.

A simple steepest-descent method for Principal Component Analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the p dominant eigenvectors of A.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^TX=I_p} \mathrm{trace}(X^TAXN).$$

  The columns of X are the p dominant eigenvectors or A.
- A basic method: Steepest-ascent on $\mathrm{St}(p,n) := \{X \in \mathbb{R}^{n\times p} : X^TX = I\}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{trace}(X^TAXN)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X) = \mathcal{P}_{T_X\mathrm{St}(p,n)}(AXN) = AXN - X\mathrm{sym}(X^TAXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X)$.

A simple steepest-descent method for Principal Component Analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the p dominant eigenvectors of A.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^TX=I_p} \mathrm{trace}(X^TAXN).$$

  The columns of X are the p dominant eigenvectors or A.
- A basic method: Steepest-ascent on $\mathrm{St}(p,n) := \{X \in \mathbb{R}^{n\times p} : X^TX = I\}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{trace}(X^TAXN)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X) = \mathcal{P}_{T_X\mathrm{St}(p,n)}(AXN) = AXN - X\mathrm{sym}(X^TAXN)$, where $\mathrm{sym}(Z) := (Z+Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X)$.

A simple steepest-descent method for Principal Component Analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the p dominant eigenvectors of A.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^TX=I_p} \mathrm{trace}(X^TAXN).$$

  The columns of X are the p dominant eigenvectors or A.
- A basic method: Steepest-ascent on $\mathrm{St}(p,n) := \{X \in \mathbb{R}^{n\times p} : X^TX = I\}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{trace}(X^TAXN)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X) = \mathcal{P}_{T_X\mathrm{St}(p,n)}(AXN) = AXN - X\mathrm{sym}(X^TAXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X)$.

A simple steepest-descent method for Principal Component Analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the p dominant eigenvectors of A.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

$$\max_{X^TX=I_p} \mathrm{trace}(X^TAXN).$$

  The columns of X are the p dominant eigenvectors or A.
- A basic method: Steepest-ascent on $\mathrm{St}(p,n) := \{X \in \mathbb{R}^{n\times p} : X^TX = I\}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{trace}(X^TAXN)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X) = \mathcal{P}_{T_X\mathrm{St}(p,n)}(AXN) = AXN - X\mathrm{sym}(X^TAXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X)$.

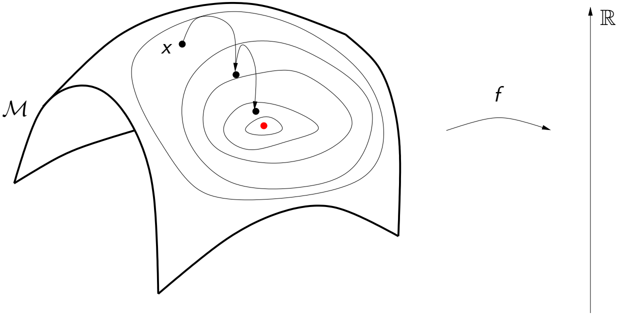
A simple steepest-descent method for Principal Component Analysis

- Let $A = A^T \in \mathbb{R}^{n\times n}$.
- Goal: Compute the p dominant eigenvectors of A.
- Principle: Let $N = \mathrm{diag}(p, p-1, \cdots, 1)$ and solve

  $$\max_{X^TX=I_p} \mathrm{trace}(X^TAXN).$$

  The columns of X are the p dominant eigenvectors or A.
- A basic method: Steepest-ascent on $\mathrm{St}(p,n) := \{X \in \mathbb{R}^{n\times p} : X^TX = I\}$.
- Let $f : \mathbb{R}^{n\times p} \to \mathbb{R} : X \mapsto \mathrm{trace}(X^TAXN)$.
- We have $\frac{1}{2}\mathrm{grad}\, f(X) = AXN$.
- Thus $\frac{1}{2}\mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X) = \mathcal{P}_{T_X\mathrm{St}(p,n)}(AXN) = AXN - X\mathrm{sym}(X^TAXN)$, where $\mathrm{sym}(Z) := (Z + Z^T)/2$.
- Basic algorithm: Follow $\dot{X} = \mathrm{grad}\, f|_{\mathrm{St}(p,n)}(X)$.

Optimization on Manifolds in one picture



Newton's method on manifolds

Newton on abstract manifolds

Required: Riemannian manifold $\mathcal{M}$; retraction R on $\mathcal{M}$; affine connection ∇ on $\mathcal{M}$; real-valued function f on $\mathcal{M}$.
Iteration $x_k \in \mathcal{M} \mapsto x_{k+1} \in \mathcal{M}$ defined by

1. Solve the Newton equation
$$\mathrm{Hess}\, f(x_k)\eta_k = -\mathrm{grad}\, f(x_k)$$
for the unknown $\eta_k \in T_{x_k}\mathcal{M}$, where
$$\mathrm{Hess}\, f(x_k)\eta_k := \nabla_{\eta_k}\mathrm{grad}\, f.$$
2. Set
$$x_{k+1} := R_{x_k}(\eta_k).$$

Newton on submanifolds of $\mathbb{R}^n$

Required: Riemannian submanifold $\mathcal{M}$ of $\mathbb{R}^n$; retraction R on $\mathcal{M}$; real-valued function f on $\mathcal{M}$.
Iteration $x_k \in \mathcal{M} \mapsto x_{k+1} \in \mathcal{M}$ defined by

1. Solve the Newton equation

$$\operatorname{Hess} f(x_k)\eta_k = -\operatorname{grad} f(x_k)$$

   for the unknown $\eta_k \in T_{x_k}\mathcal{M}$, where

$$\operatorname{Hess} f(x_k)\eta_k := \mathrm{P}_{T_{x_k}\mathcal{M}} \mathrm{D}\operatorname{grad} f(x_k)[\eta_k].$$

2. Set

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

Newton on the unit sphere S^{n-1}

Required: real-valued function f on S^{n-1}.
Iteration $x_k \in S^{n-1} \mapsto x_{k+1} \in S^{n-1}$ defined by

1. Solve the Newton equation

$$\begin{cases} \mathrm{P}_{x_k} \mathrm{D}(\operatorname{grad} f)(x_k)[\eta_k] = -\operatorname{grad} f(x_k) \\ x^T \eta_k = 0, \end{cases}$$

   for the unknown $\eta_k \in \mathbb{R}^n$, where

$$\mathrm{P}_{x_k} = (I - x_k x_k^T).$$

2. Set

$$x_{k+1} := \frac{x_k + \eta_k}{\|x_k + \eta_k\|}.$$

Newton for Rayleigh quotient optimization on unit sphere

Iteration $x_k \in S^{n-1} \mapsto x_{k+1} \in S^{n-1}$ defined by

1. Solve the Newton equation

$$\begin{cases} \mathrm{P}_{x_k} A \mathrm{P}_{x_k} \eta_k - \eta_k x_k^T A x_k = -\mathrm{P}_{x_k} A x_k, \\ x_k^T \eta_k = 0, \end{cases}$$

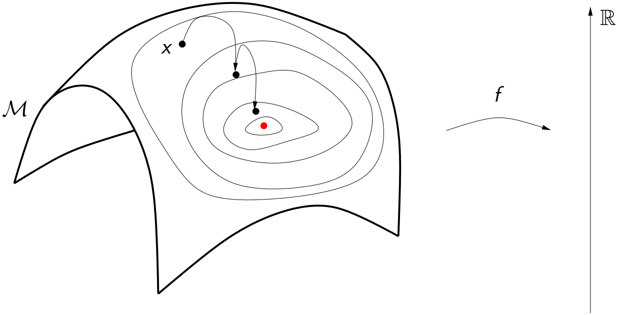
   for the unknown $\eta_k \in \mathbb{R}^n$, where

$$\mathrm{P}_{x_k} = (I - x_k x_k^T).$$

2. Set

$$x_{k+1} := \frac{x_k + \eta_k}{\|x_k + \eta_k\|}.$$

Optimization on Manifolds in one picture



An Application: Blind Source Separation

Joint diagonalization on the Stiefel manifold

- Measurements $X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_f) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_f) \end{bmatrix}$.
- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U\Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose Σ and U such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose V to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve

Joint diagonalization on the Stiefel manifold

- Measurements $X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_f) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_f) \end{bmatrix}$.
- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U \Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose Σ and U such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose V to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve

Joint diagonalization on the Stiefel manifold

- Measurements $X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \cdots & x_1(t_f) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \cdots & x_n(t_f) \end{bmatrix}$.
- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of

$$Y = W^T X$$

  look as statistically independent as possible.
- Decompose $W = U \Sigma V^T$. We have

$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$

- Whitening: Choose Σ and U such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose V to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve

Joint diagonalization on the Stiefel manifold

- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U \Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose Σ and U such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose V to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve
$$\max_{V^T V = I_p} \sum_{i=1}^{N} \|\mathrm{diag}(V^T C_i(\tilde{X}) V)\|_F^2.$$

Joint diagonalization on the Stiefel manifold

- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U \Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose Σ and U such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose V to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve
$$\max_{V^T V = I_p} \sum_{i=1}^{N} \|\mathrm{diag}(V^T C_i(\tilde{X}) V)\|_F^2.$$

Joint diagonalization on the Stiefel manifold

- Goal: Find a matrix $W \in \mathbb{R}^{n \times p}$ such that the rows of
$$Y = W^T X$$
look as statistically independent as possible.
- Decompose $W = U \Sigma V^T$. We have
$$Y = V^T \underbrace{\Sigma U^T X}_{=:\tilde{X}}.$$
- Whitening: Choose Σ and U such that $\tilde{X}\tilde{X}^T = I_n$. Then $YY^T = V^T \tilde{X}\tilde{X}^T V = V^T V = I_p$.
- Independence and dimension reduction: Consider a collection of covariance-like matrix functions $C_i(Y)$ such that $C_i(Y) = V^T C_i(\tilde{X}) V$. Choose V to make the $C_i(Y)$'s as diagonal as possible.
- Principle: Solve
$$\max_{V^T V = I_p} \sum_{i=1}^{N} \|\mathrm{diag}(V^T C_i(\tilde{X}) V)\|_F^2.$$

Joint diagonalization on the Stiefel manifold: application

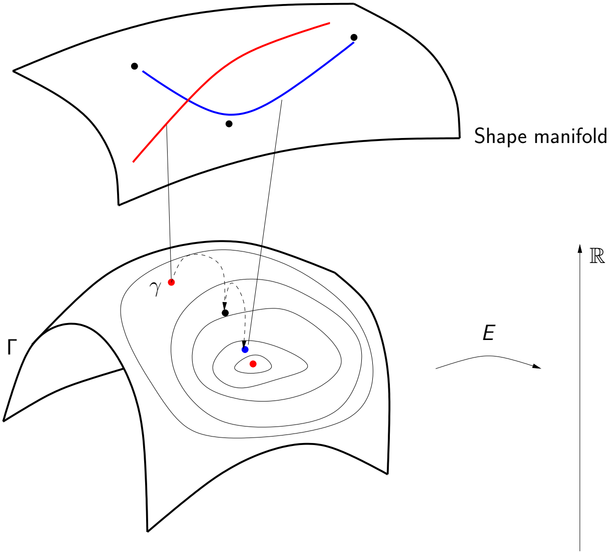
The application is blind source separation.
Two mixed pictures are given as input to a blind source separation algorithm based on a trust-region method on $\mathrm{St}(2, 2)$.

Joint diagonalization on the Stiefel manifold: application: input

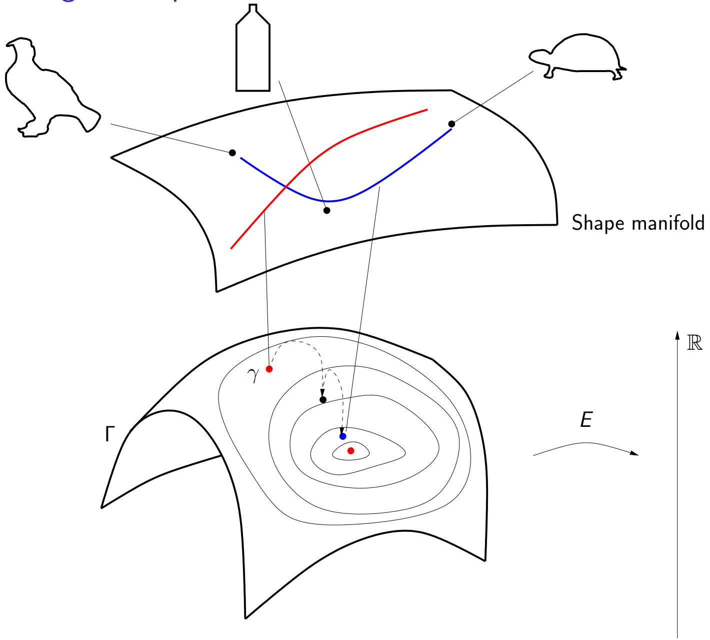
Joint diagonalization on the Stiefel manifold: application: output

Another Application: Curve Fitting

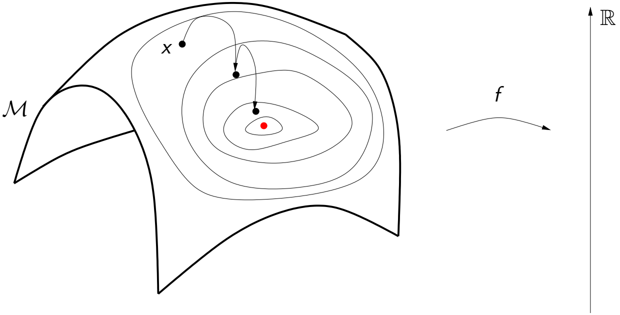
Curve fitting on manifolds



Curve fitting on shape manifold



Optimization on Manifolds in one picture



A few pointers I

- Optimization on manifolds in general: Luenberger [Lue73], Gabay [Gab82], Smith [Smi93, Smi94], Udriște [Udr94], Manton [Man02], Mahony and Manton [MM02], PAA *et al.* [ABG07]...
- Stiefel and Grassmann manifolds: Edelman *et al.* [EAS98], PAA *et al.* [AMS04]...
- Retractions: Shub [Shu86], Adler *et al.* [ADM$^+$02]...
- Eigenvalue problem: Chen and Amari [CA01], Lundström and Eldén [LE02], Simoncini and Eldén [SE02], Brandts [Bra03], Absil *et al.* [AMSV02, AMS04, ASVM04, ABGS05, ABG06] and Baker *et al.* [BAG06]

A few pointers II

- Independent component analysis: Amari *et al.* [ACC00], Douglas [Dou00], Rahbar and Reilly [RR00], Pham [Pha01], Joho and Mathis [JM02], Joho and Rahbar [JR02], Nikpour *et al.* [NMH02], Afsari and Krishnaprasad [AK04], Nishimori and Akaho [NA05], Plumbley [Plu05], PAA and Gallivan [AG06], Shen *et al.* [SHS06], Hüeper *et al.* [HSS06]...
- Pose estimation: Ma *et al.* [MKS01], Lee and Moore [LM04], Liu *et al.* [LSG04], Helmke *et al.* [HHLM07]
- Various matrix nearness problems: Trendafilov and Lippert [TL02], Grubisic and Pietersz [GP07]...

- 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** (2006), no. 1–2, 274–285.
- ______, *Trust-region methods on Riemannian manifolds*, Found. Comput. Math. **7** (2007), no. 3, 303–330.
- P.-A. Absil, C. G. Baker, K. A. Gallivan, and A. Sameh, *Adaptive model trust region methods for generalized eigenvalue problems*, International Conference on Computational Science (Vaidy S. Sunderam, Geert Dick van Albada, and Peter M. A. Sloot, eds.), Lecture Notes in Computer Science, vol. 3514, Springer-Verlag, 2005, pp. 33–41.
- Shun-ichi Amari, Tian-Ping Chen, and Andrzej Cichocki, *Nonholonomic orthogonal learning algorithms for blind source separation*, Neural Comput. **12** (2000), 1463–1484.

- 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** (2002), no. 3, 359–390.
- P.-A. Absil and K. A. Gallivan, *Joint diagonalization on the oblique manifold for independent component analysis*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, 2006, pp. V–945–V–948.
- Bijan Afsari and P. S. Krishnaprasad, *Some gradient based joint diagonalization methods for ICA*, Proceedings of the 5th International Conference on Independent Component Analysis and Blind Source Separation (Springer LCNS Series, ed.), 2004.
- P.-A. Absil, R. Mahony, and R. Sepulchre, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, Acta Appl. Math. **80** (2004), no. 2, 199–220.

P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren, *A Grassmann-Rayleigh quotient iteration for computing invariant subspaces*, SIAM Rev. **44** (2002), no. 1, 57–73. MR MR1924546 (2003h:65045)

P.-A. Absil, R. Sepulchre, P. Van Dooren, and R. Mahony, *Cubically convergent iterations for invariant subspace computation*, SIAM J. Matrix Anal. Appl. **26** (2004), no. 1, 70–96. MR MR2112852 (2005m:65074)

C. G. Baker, P.-A. Absil, and K. A. Gallivan, *An implicit Riemannian trust-region method for the symmetric generalized eigenproblem*, Computational Science—ICCS 2006 (Vassil N. Alexandrov, Geert Dick van Albada, Peter M.A. Sloot, and Jack Dongarra, eds.), LNCS, vol. 3991, Springer, New York, 2006, pp. 210–217.

- Jan Brandts, *The Riccati algorithm for eigenvalues and invariant subspaces of matrices with inexpensive action*, Linear Algebra Appl. **358** (2003), 335–365, Special issue on accurate solution of eigenvalue problems (Hagen, 2000). MR MR1942737 (2003m:65048)
- T. P. Chen and S. Amari, *Unified stabilization approach to principal and minor components extraction algorithms*, Neural Networks **14** (2001), no. 10, 1377–1387.
- Scott C. Douglas, *Self-stabilized gradient algorithms for blind source separation with orthogonality constraints*, IEEE Trans. Neural Networks **11** (2000), no. 6, 1490–1497.
- Alan Edelman, Tomás A. Arias, and Steven T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353. MR MR1646856 (99h:65075)

- D. Gabay, *Minimizing a differentiable function over a differential manifold*, J. Optim. Theory Appl. **37** (1982), no. 2, 177–219. MR MR663521 (84h:49071)
- Igor Grubišić and Raoul Pietersz, *Efficient rank reduction of correlation matrices*, Linear Algebra Appl. **422** (2007), no. 2-3, 629–653. MR MR2305146
- 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** (2007), no. 2, 117–136.
- Knut Hüper, Hao Shen, and Abd-Krim Seghouane, *Local convergence properties of FastICA and some generalisations*, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, 2006, pp. V–1009–V–1012.

Marcel Joho and Heinz Mathis, *Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation*, Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop SAM, 2002, pp. 273–277.

Marcel Joho and Kamran Rahbar, *Joint diagonalization of correlation matrices by using Newton methods with applications to blind signal separation*, Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop SAM, 2002, pp. 403–407.

Eva Lundström and Lars Eldén, *Adaptive eigenvalue computations using Newton's method on the Grassmann manifold*, SIAM J. Matrix Anal. Appl. **23** (2001/02), no. 3, 819–839. MR MR1896820 (2003d:65030)

Pei Yean Lee and John B. Moore, *Pose estimation via a Gauss-Newton-on-manifold approach*, Proceedings of the 16th International Symposium on Mathematical Theory of Network and System (MTNS), Leuven, 2004.

- Xiuwen Liu, Anuj Srivastava, and Kyle Gallivan, *Optimal linear representations of images for object recognition*, IEEE Pattern Anal. and Mach. Intell. **26** (2004), no. 5, 662–666.
- David G. Luenberger, *Introduction to linear and nonlinear programming*, Addison-Wesley, Reading, MA, 1973.
- Jonathan H. Manton, *Optimization algorithms exploiting unitary constraints*, IEEE Trans. Signal Process. **50** (2002), no. 3, 635–650. MR MR1895067 (2003i:90078)
- Yi Ma, Jana Kosecka, and Shankar S. Sastry, *Optimization criteria and geometric algorithms for motion and structure estimation*, Int. J. Computer Vision **44** (2001), no. 3, 219–249.
- Robert Mahony and Jonathan H. Manton, *The geometry of the Newton method on non-compact Lie groups*, J. Global Optim. **23** (2002), no. 3-4, 309–327, Nonconvex optimization in control. MR MR1923049 (2003g:90114)

- Yasunori Nishimori and Shotaro Akaho, *Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold*, Neurocomputing **67** (2005), 106–135.
- Maziar Nikpour, Jonathan H. Manton, and Gen Hori, *Algorithms on the Stiefel manifold for joint diagonalization*, Proc. ICASSP, 2002, pp. II–1481–1484.
- Dinh Tuan Pham, *Joint approximate diagonalization of positive definite Hermitian matrices*, SIAM J. Matrix Anal. Appl. **22** (2001), no. 4, 1136–1152. MR MR1824062 (2002e:65066)
- Mark D. Plumbley, *Geometrical methods for non-negative ICA: Manifolds, Lie groups and toral subalgebras*, Neurocomputing **67** (2005), 161–197.
- Kamran Rahbar and James P. Reilly, *Geometric optimization methods for blind source separation of signals*, International Conference on Independent Component Analysis ICA2000, Helsinki, Finland, June 2000.

Valeria Simoncini and Lars Eldén, *Inexact Rayleigh quotient-type methods for eigenvalue computations*, BIT **42** (2002), no. 1, 159–182. MR 2003d:65033

Hao Shen, Knut Hüper, and Alexander J. Smola, *Newton-like methods for nonparametric independent component analysis*, Neural Information Processing (Irwin King, Jun Wang, Laiwan Chan, and DeLiang Wang, eds.), LNCS, vol. 4232, Springer, 2006, pp. 1068–1077.

Michael Shub, *Some remarks on dynamical systems and numerical analysis*, Proc. VII ELAM. (L. Lara-Carrero and J. Lewowicz, eds.), Equinoccio, U. Simón Bolívar, Caracas, 1986, pp. 69–92.

Steven Thomas Smith, *Geometric optimization methods for adaptive filtering*, Ph.D. thesis, Division of Applied Sciences, Harvard University, Cambridge, MA, May 1993.

- Steven T. Smith, *Optimization techniques on Riemannian manifolds*, Hamiltonian and gradient flows, algorithms and control (Anthony Bloch, ed.), Fields Inst. Commun., vol. 3, Amer. Math. Soc., Providence, RI, 1994, pp. 113–136. MR MR1297990 (95g:58062)
- Nickolay T. Trendafilov and Ross A. Lippert, *The multimode Procrustes problem*, Linear Algebra Appl. **349** (2002), 245–264. MR MR1903736 (2003b:90071)
- Constantin Udrişte, *Convex functions and optimization methods on Riemannian manifolds*, Mathematics and its Applications, vol. 297, Kluwer Academic Publishers Group, Dordrecht, 1994. MR MR1326607 (97a:49038)