Optimization on manifolds

Pierre-Antoine Absil
(UCL/INMA)

`IAP Study Day`

`16 April 2007`

Collaboration

Chris Baker
(Florida State University)

Kyle Gallivan
(Florida State University)

Robert Mahony
(Australian National University)

Rodolphe Sepulchre
(Université de Liège)

Paul Van Dooren
(Université catholique de Louvain)

Optimization On Manifolds

What ?

Why ?

How ?

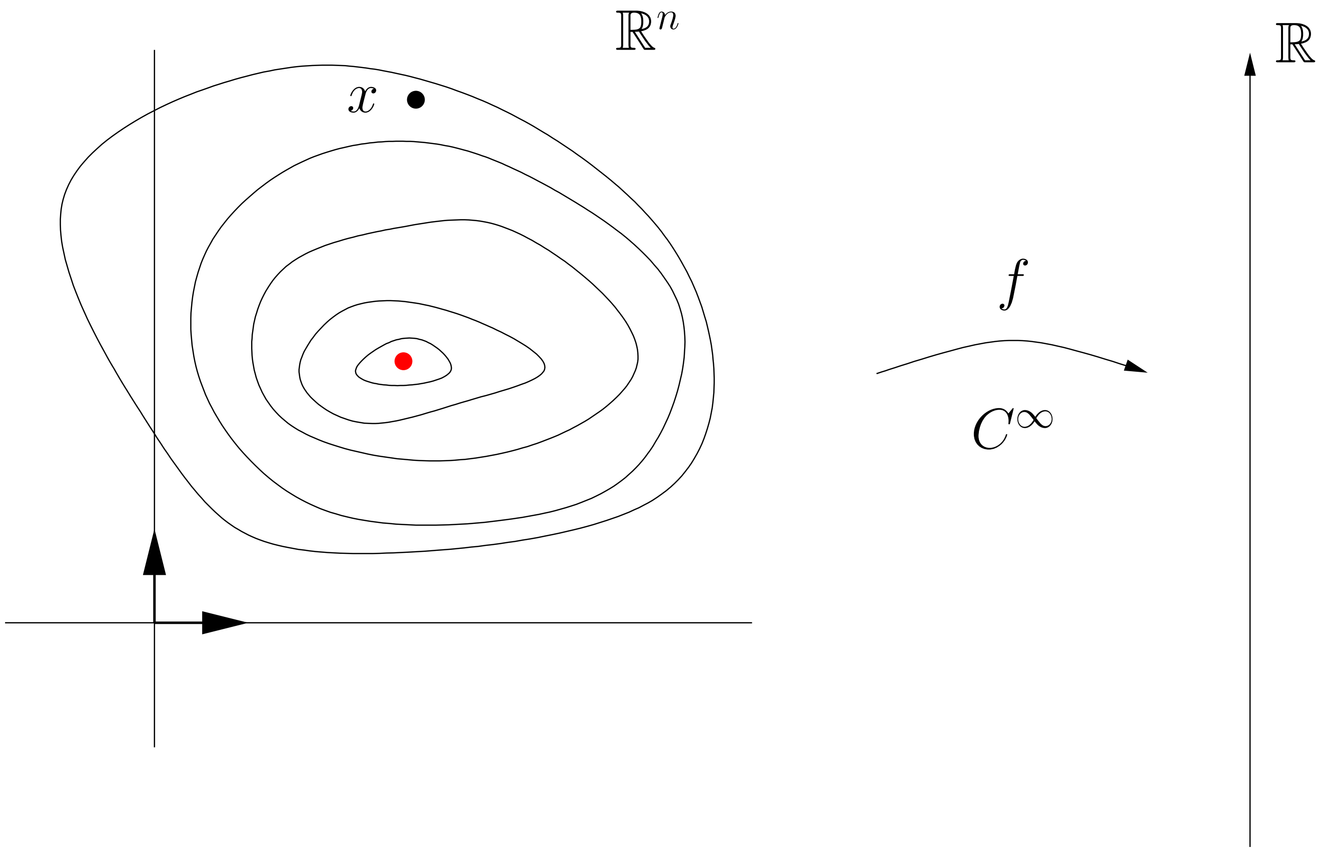
Optimization On Manifolds

What ?

Why ?

How ?

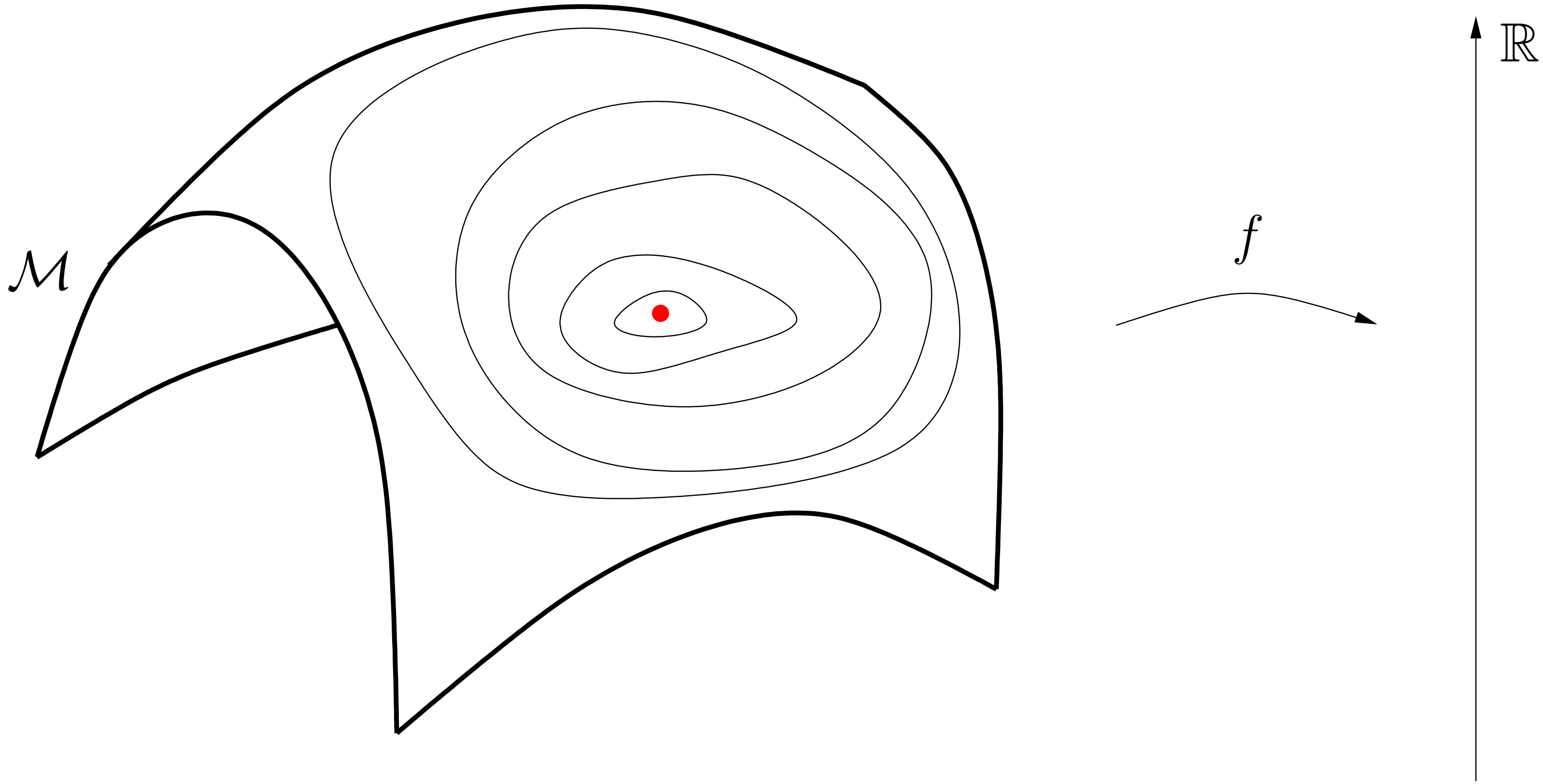
Smooth optimization in $\mathbb{R}^n$



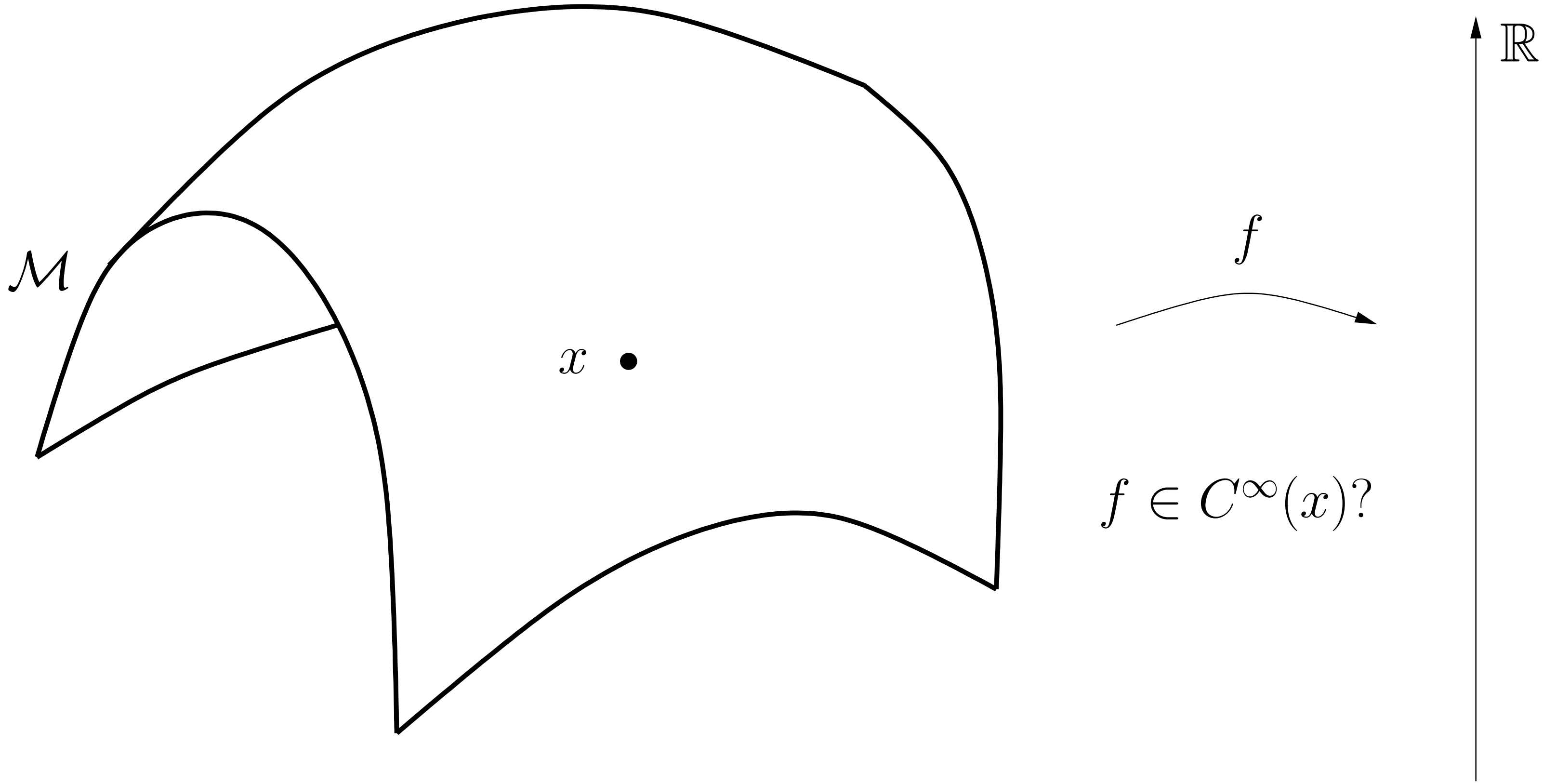
Optimization on "a set"

Differentiability? Generalization went too far!

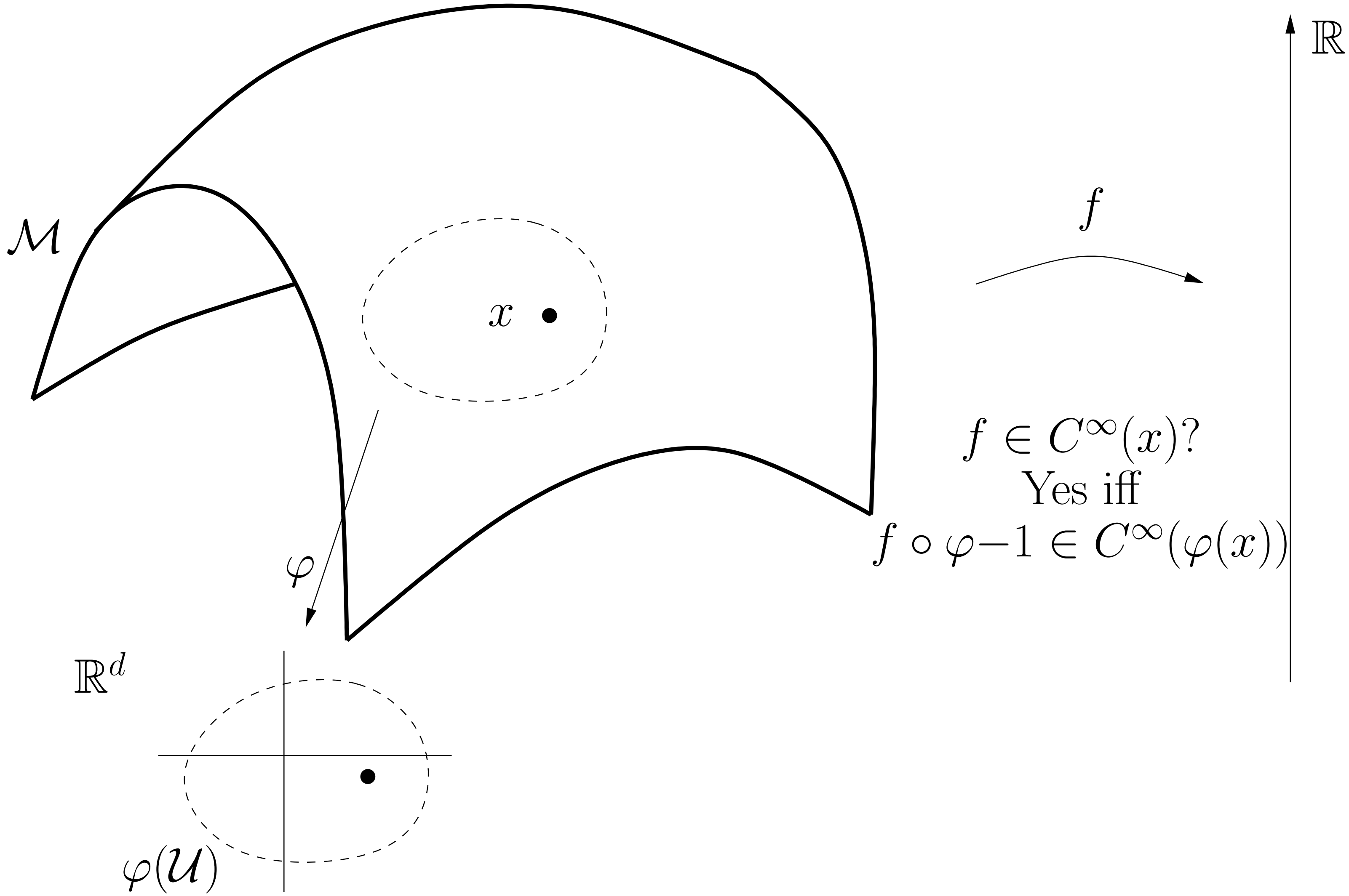
Smooth optimization on a manifold



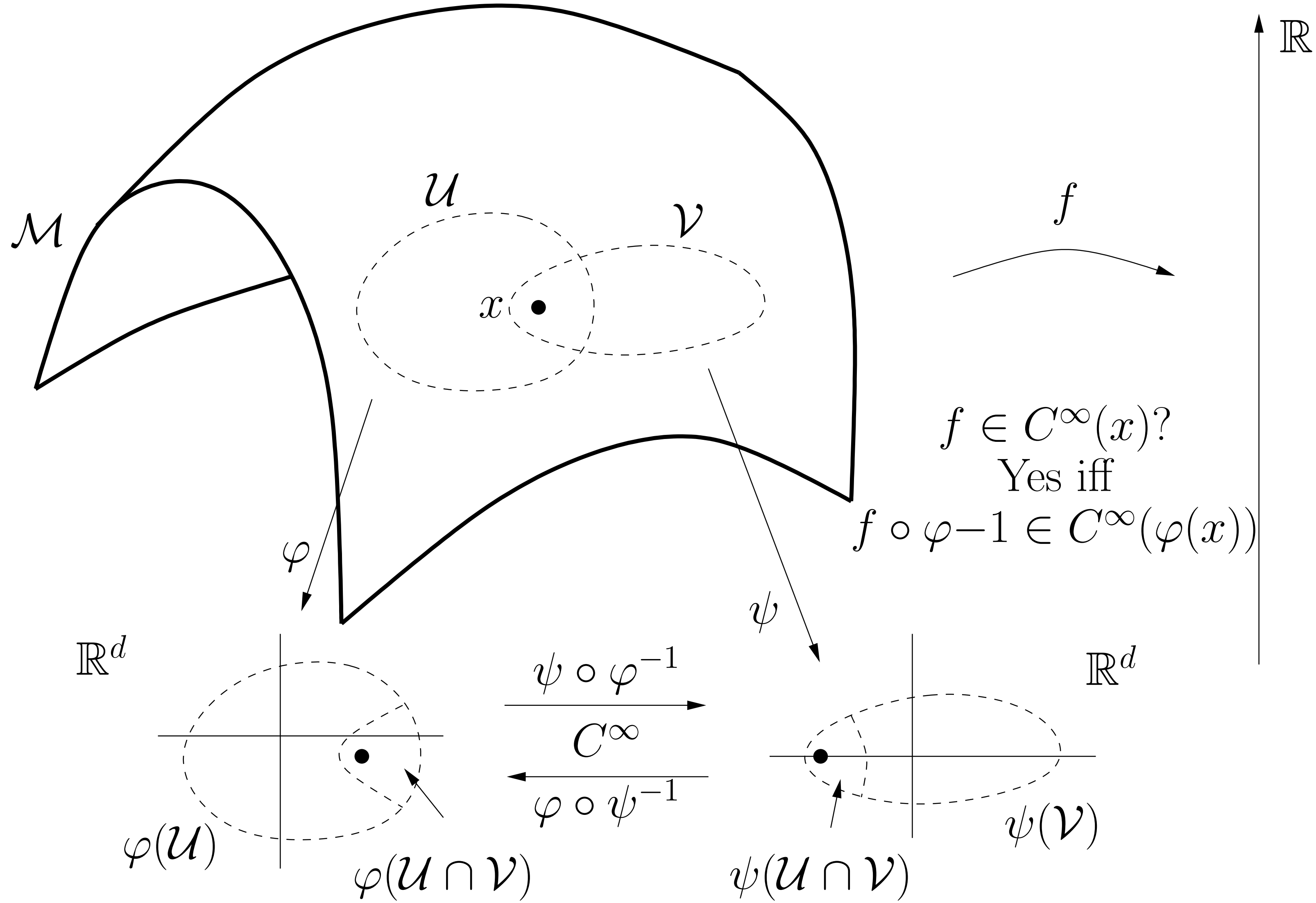
Smooth optimization on a manifold: what “smooth” means



Smooth optimization on a manifold: what "smooth" means



Smooth optimization on a manifold: what "smooth" means



Smooth optimization on a manifold: what "smooth" means

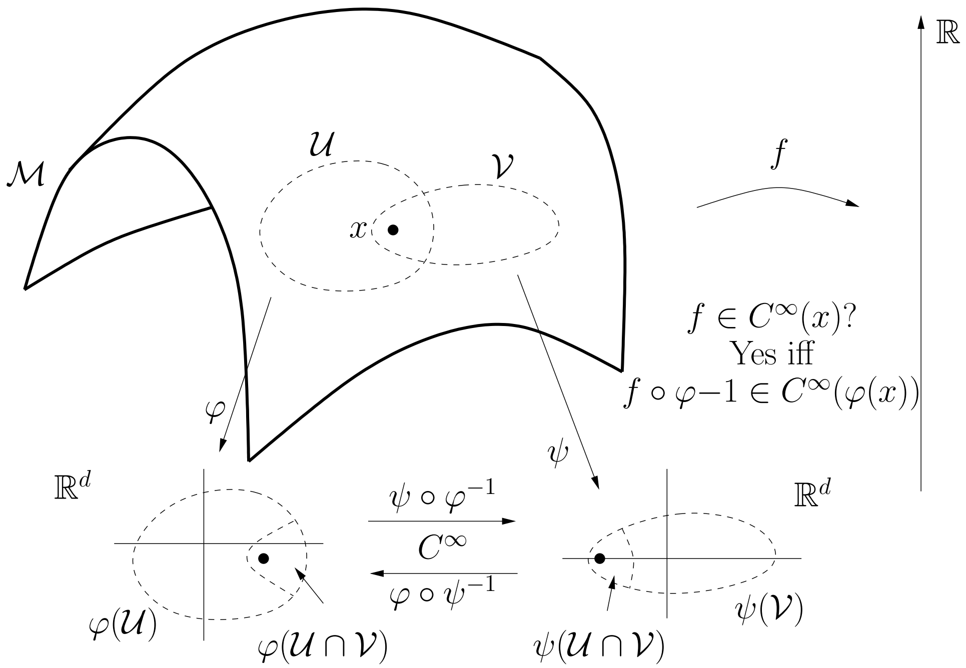
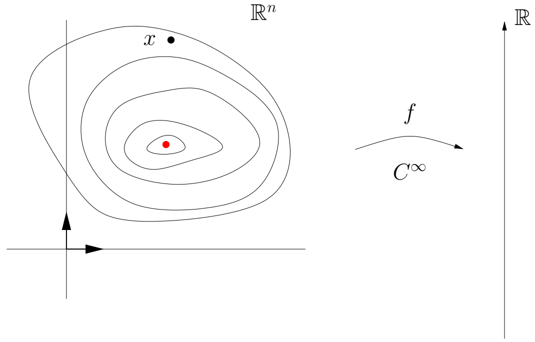


Chart: $\mathcal{U} \xrightarrow[\text{bij.}]{\varphi} \varphi(\mathcal{U})$

Atlas: Collection of "compatible charts" that cover $\mathcal{M}$

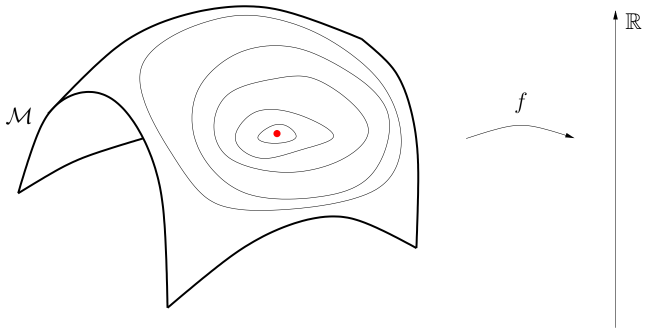
Manifold: Set with an atlas

(Highly Questionable) Summary



Optimization in $\mathbb{R}^n$: too easy

Optimization on arbitrary sets: too difficult



Optimization on manifolds: just right!

(Less Questionable) Summary

Smooth Optimization On Manifolds is a natural generalization of smooth optimization in $\mathbb{R}^n$.

Some important manifolds

- Stiefel manifold $\mathrm{St}(p, n)$: set of all orthonormal $n \times p$ matrices.
- Grassmann manifold $\mathrm{Grass}(p, n)$: set of all p-dimensional subspaces of $\mathbb{R}^n$
- Euclidean group $SE(3)$: set of all rotations-translations
- Flag manifold, shape manifold, oblique manifold...
- Several unnamed manifolds

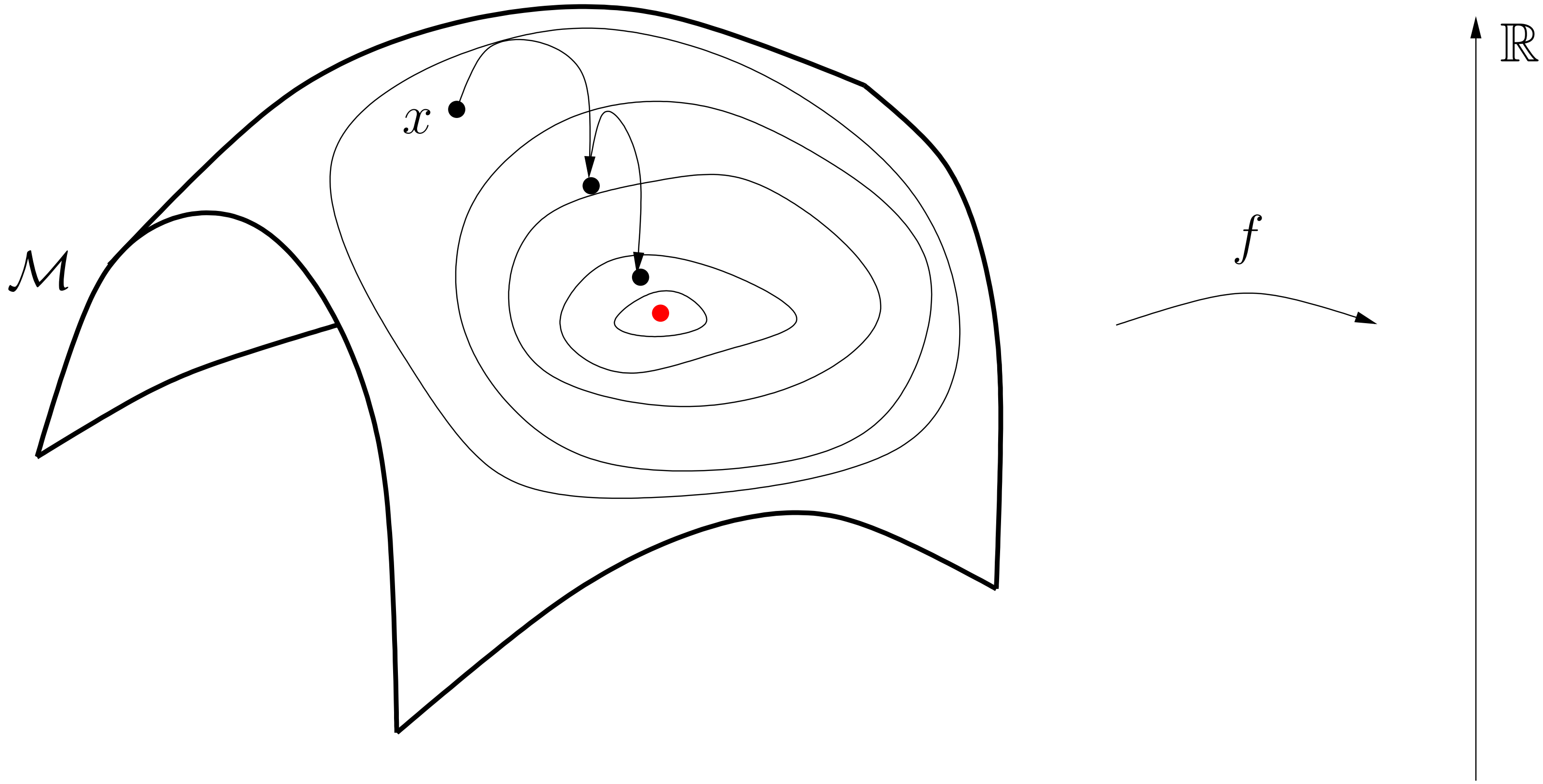
Optimization On Manifolds

What ?

Why ?

How ?

Optimization On Manifolds in one picture



Optimization On Manifolds

What ?

Why ?

How ?

Why?

Two examples of computational problems that can (should) be phrased as problems of Optimization On Manifolds:

- mechanical vibrations
- independent component analysis (ICA)

Mechanical vibrations

Stiffness matrix $A = A^T$, mass matrix $B = B^T \succ 0$.

Equation of vibrations (for undamped discretized linear structures):

$$Ax = \lambda Bx$$

where

- $\lambda = \omega^2$, ω angular frequency of vibration
- x is the corresponding mode of vibration.

Task: find lowest mode of vibration.

Generalized eigenvalue problem (GEP)

Given $n \times n$ matrices $A = A^T$ and $B = B^T \succ 0$, there exist $v_1, \ldots, v_n$ in $\mathbb{R}^n$ and $\lambda_1 \leq \ldots \leq \lambda_n$ in $\mathbb{R}$ such that

$$Av_i = \lambda_i B v_i$$
$$v_i^T B v_j = \delta_{ij}.$$

Task: find λ_1 and v_1.

We assume that $\lambda_1 < \lambda_2$ (simple eigenvalue).

GEP: optimization in $\mathbb{R}^n$

$$Av_i = \lambda_i B v_i$$

Cost function: Rayleigh quotient

$$\tilde{f} : \mathbb{R}^n_* \to \mathbb{R} : f(y) = \frac{y^T A y}{y^T B y}$$

Minimizers of $\tilde{f}$: αv_1, for all $\alpha \neq 0$.

The minimizers of $\tilde{f}$ yield the lowest mode of vibration.

The minimizers are not isolated.

GEP: optimization in $\mathbb{R}^n$

$$Av_i = \lambda_i B v_i$$

Cost function: Rayleigh quotient

$$\tilde{f} : \mathbb{R}^n_* \to \mathbb{R} : f(y) = \frac{y^T A y}{y^T B y}$$

Minimizers of $\tilde{f}$: αv_1, for all $\alpha \neq 0$.

The minimizers of $\tilde{f}$ yield the lowest mode of vibration.

The minimizers are not isolated.

Invariance property: $\tilde{f}(\alpha y) = \tilde{f}(y)$.

GEP: optimization in $\mathbb{R}^n$

$$Av_i = \lambda_i B v_i$$

Cost function: Rayleigh quotient

$$\tilde{f} : \mathbb{R}^n_* \to \mathbb{R} : f(y) = \frac{y^T A y}{y^T B y}$$

Minimizers of $\tilde{f}$: αv_1, for all $\alpha \neq 0$.

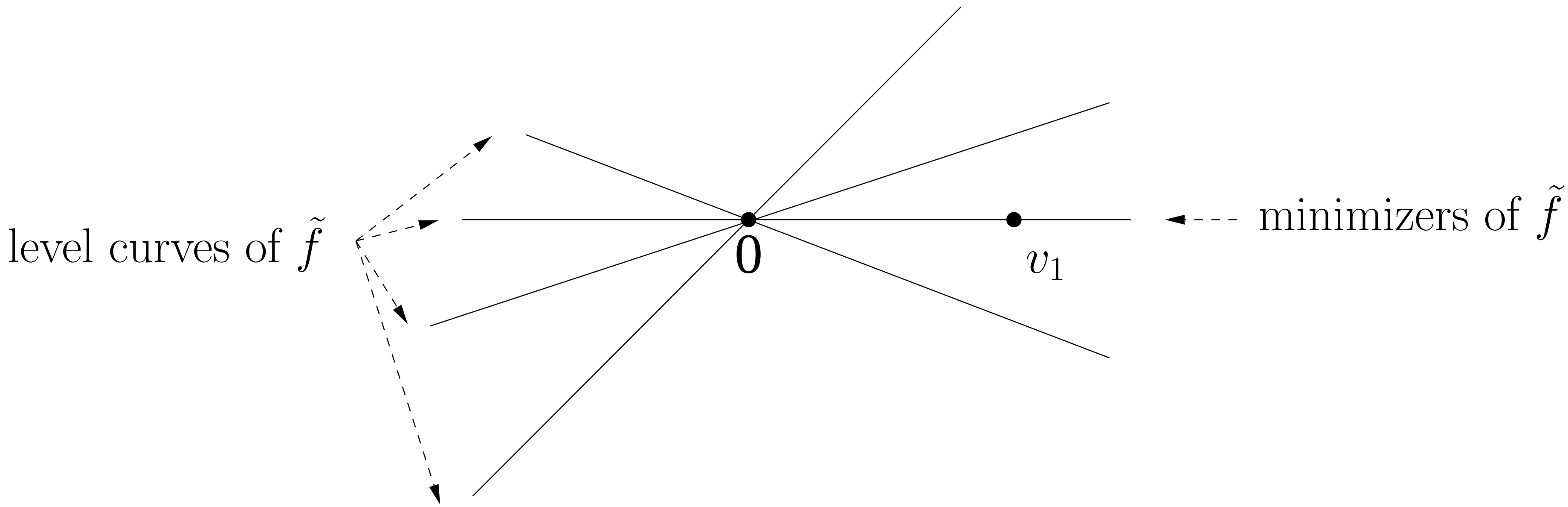
The minimizers of $\tilde{f}$ yield the lowest mode of vibration.

The minimizers are not isolated.

Invariance property: $\tilde{f}(\alpha y) = \tilde{f}(y)$. Idea: exploit the invariance property $\rightsquigarrow$ Optimization On Manifold.

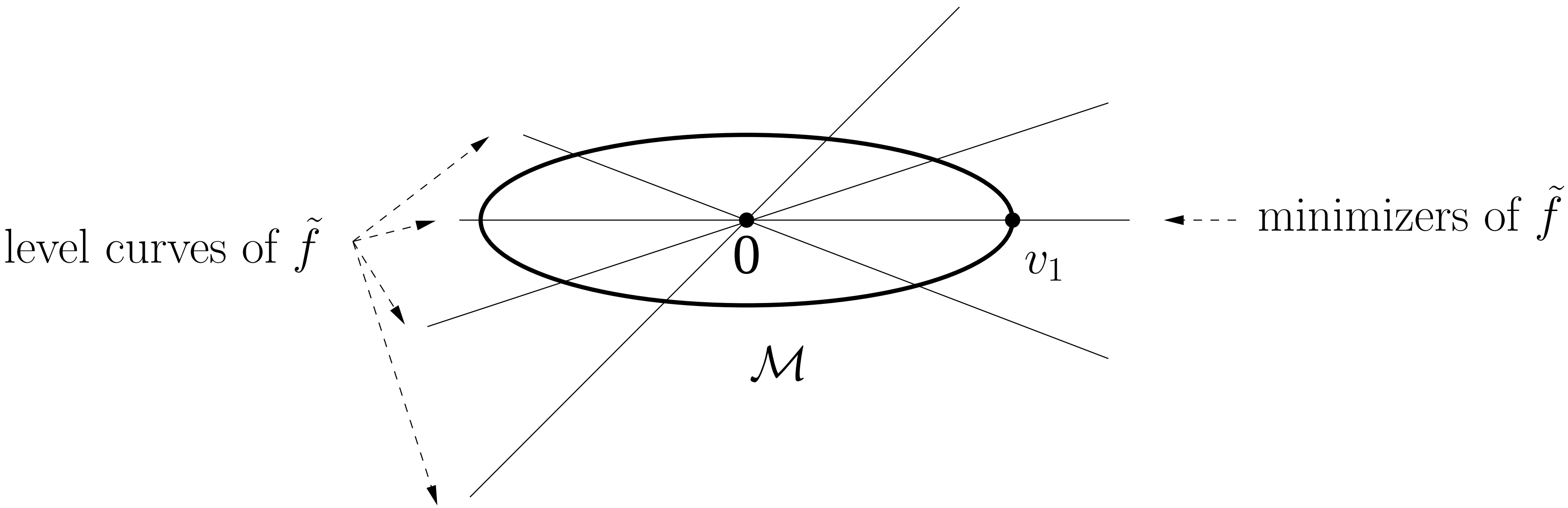
GEP: invariance by scaling

$$\tilde{f}(\alpha y) = \tilde{f}(y).$$



GEP: optimization on ellipsoid

$$\tilde{f}(\alpha y) = \tilde{f}(y).$$



GEP: optimization on ellipsoid

$$\tilde{f} : \mathbb{R}^n_* \to \mathbb{R} : f(y) = \frac{y^T A y}{y^T B y}$$

Invariance: $\tilde{f}(\alpha y) = \tilde{f}(y)$.

Remedy 1:

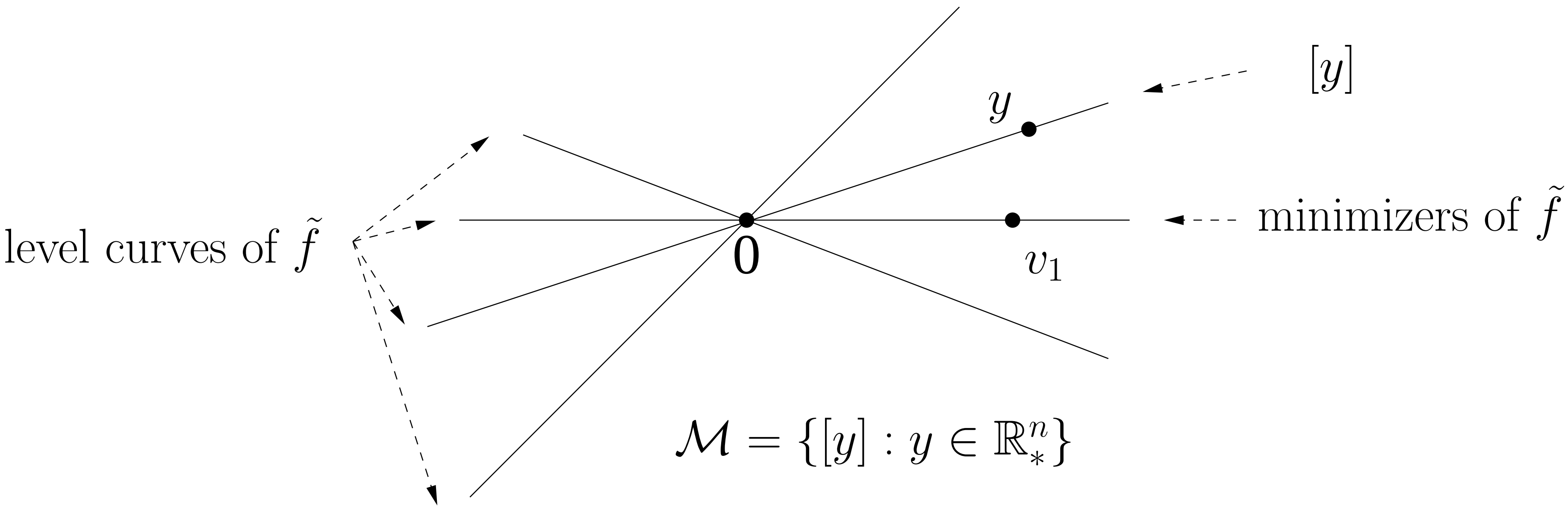
- $\mathcal{M} := \{y \in \mathbb{R}^n : y^T B y = 1\}$, *submanifold* of $\mathbb{R}^n$.
- $f : \mathcal{M} \to \mathbb{R} : f(y) = y^T A y$.

Stationary points of f: $\pm v_1, \ldots, \pm v_n$.

Minimizers of f: $\pm v_1$.

GEP: optimization on projective space

$$\tilde{f}(\alpha y) = \tilde{f}(y).$$



$$\mathcal{M} = \{[y] : y \in \mathbb{R}^n_*\}$$

GEP: optimization on projective space

$$\tilde{f} : \mathbb{R}^n_* \to \mathbb{R} : f(y) = \frac{y^T A y}{y^T B y}$$

Invariance: $\tilde{f}(\alpha y) = \tilde{f}(y)$.

Remedy 2:

- $[y] := y\mathbb{R} := \{y\alpha : \alpha \in \mathbb{R}\}$
- $\mathcal{M} := \mathbb{R}^n_* / \mathbb{R} = \{[y]\}$
- $f : \mathcal{M} \to \mathbb{R} : f([y]) := \tilde{f}(y)$

Stationary points of f: $[v_1], \ldots, [v_n]$.

Minimizer of f: $[v_1]$.

Block algorithm for GEP: optimization on Grassmann manifold

Goal: compute the p lowest modes simulateously.

$$\tilde{f} : \mathbb{R}_*^{n\times p} \to \mathbb{R} : \tilde{f}(Y) = \operatorname{trace}\left((Y^T BY)^{-1} Y^T AY\right)$$

Invariance: $\tilde{f}(YR) = \tilde{f}(Y)$ for all nonsing. $p \times p$ matrices R.

- $[Y] := \{YR : R \in \mathbb{R}_*^{p\times p}\}, \quad Y \in \mathbb{R}_*^{n\times p}$
- $\mathcal{M} := \mathrm{Grass}(p,n) := \{[Y]\}$
- $f : \mathcal{M} \to \mathbb{R} : f([Y]) := \tilde{f}(Y)$

Stationary points of f: $\operatorname{span}\{v_{i_1}, \ldots, v_{i_p}\}$.

Minimizer of f: $[Y] = \operatorname{span}\{v_1, \ldots, v_p\}$.

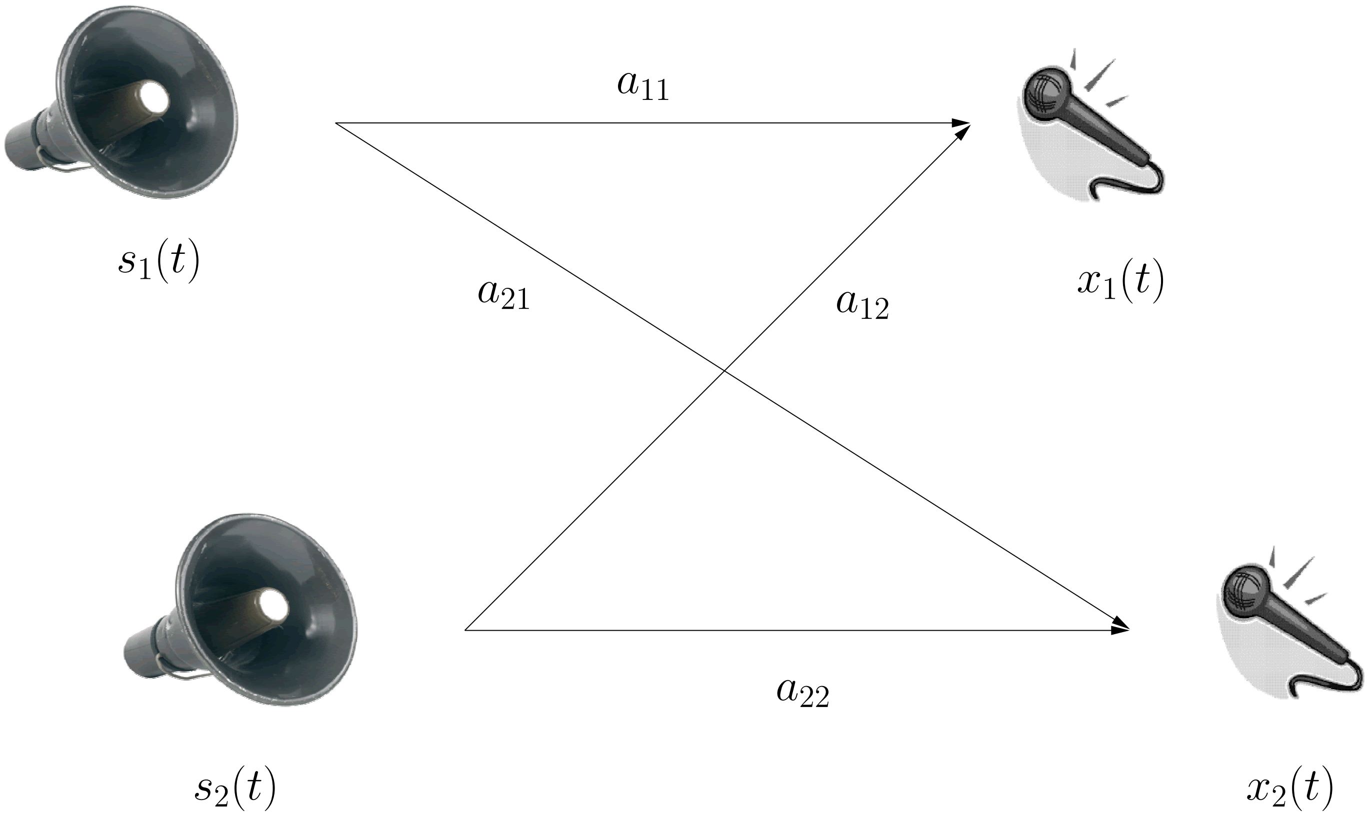
Why?

Two examples of computational problems that can (should) be phrased as problems of Optimization On Manifolds:

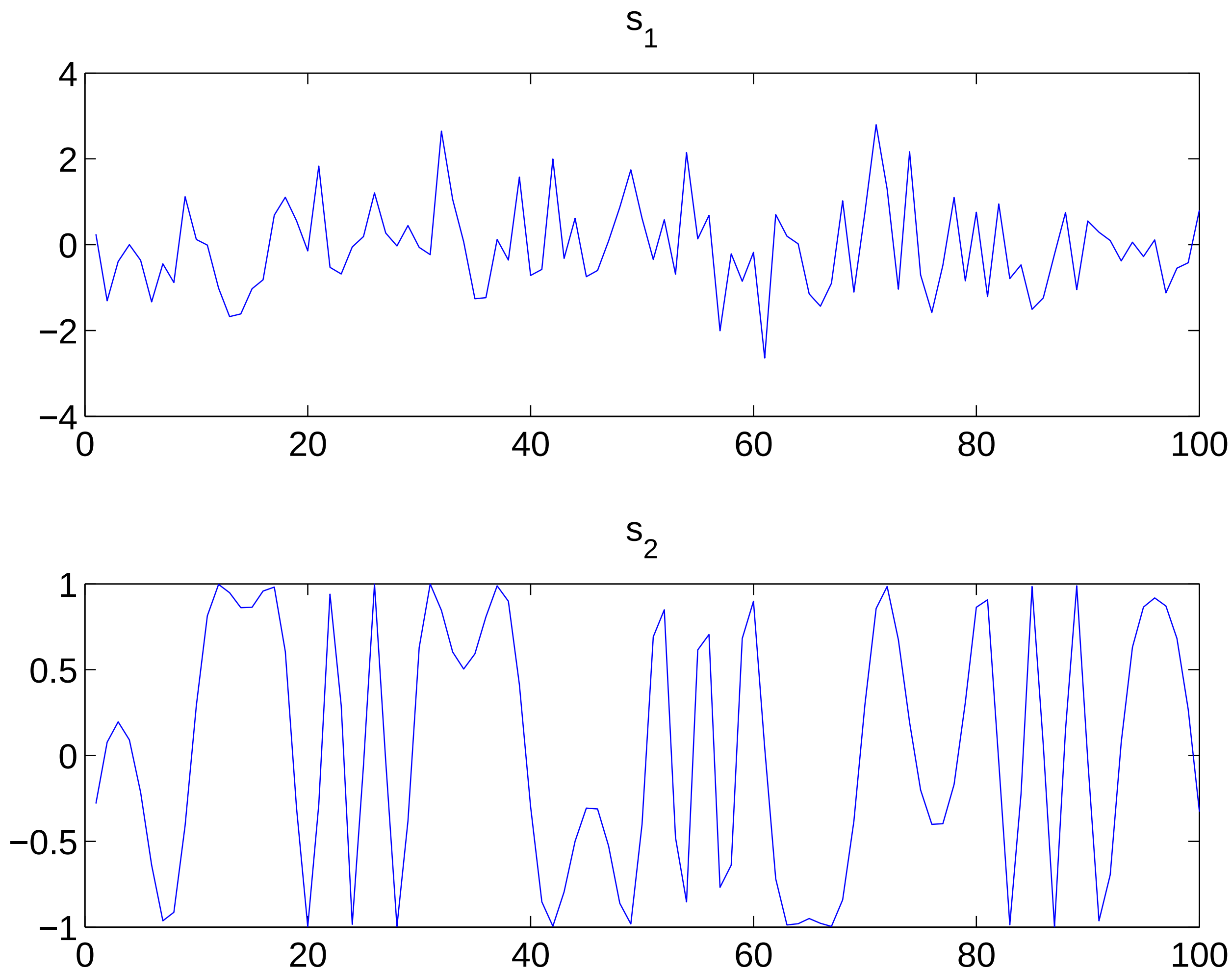
- mechanical vibrations
- independent component analysis (ICA)

Independent Component Analysis (ICA)

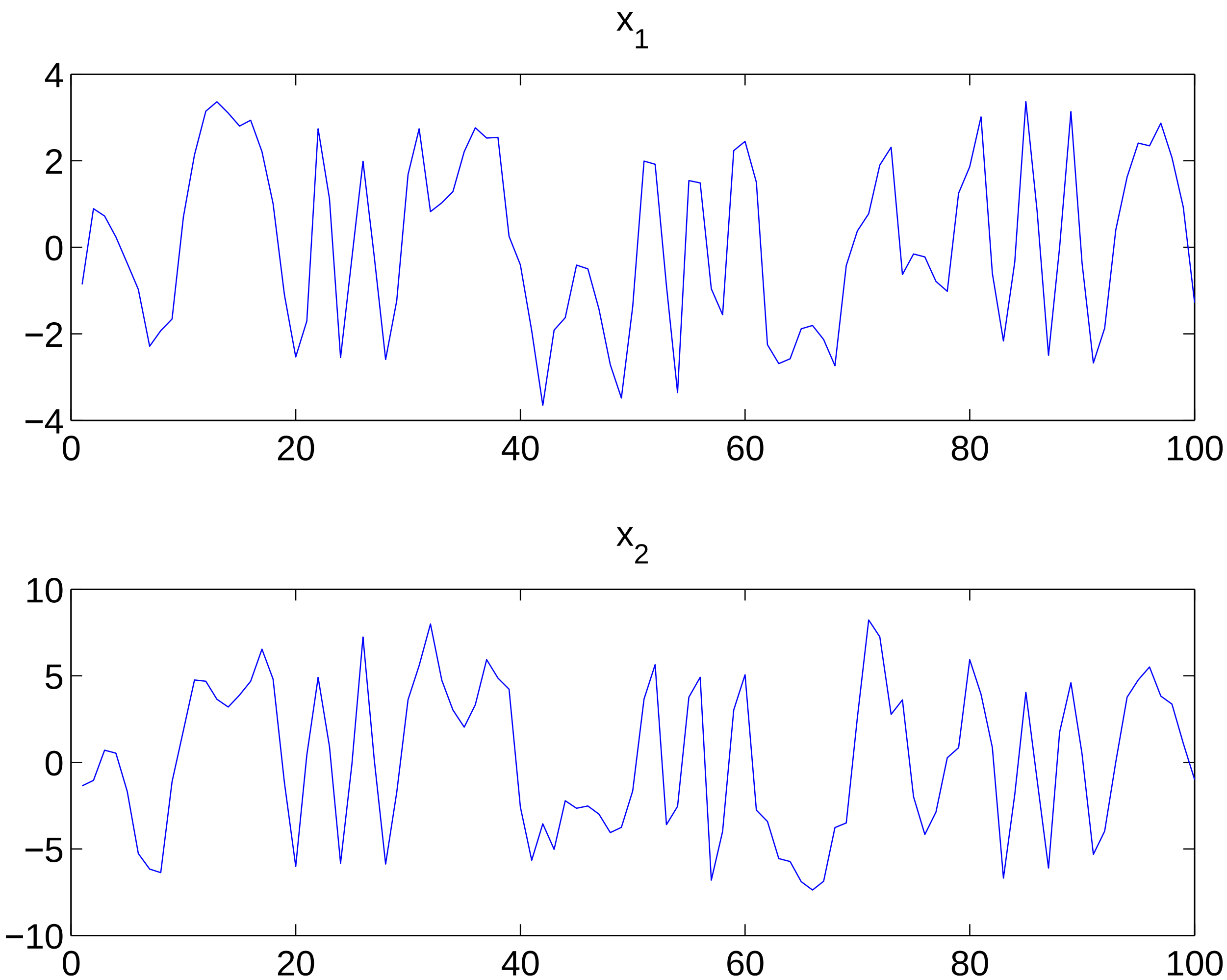
Cocktail party problem



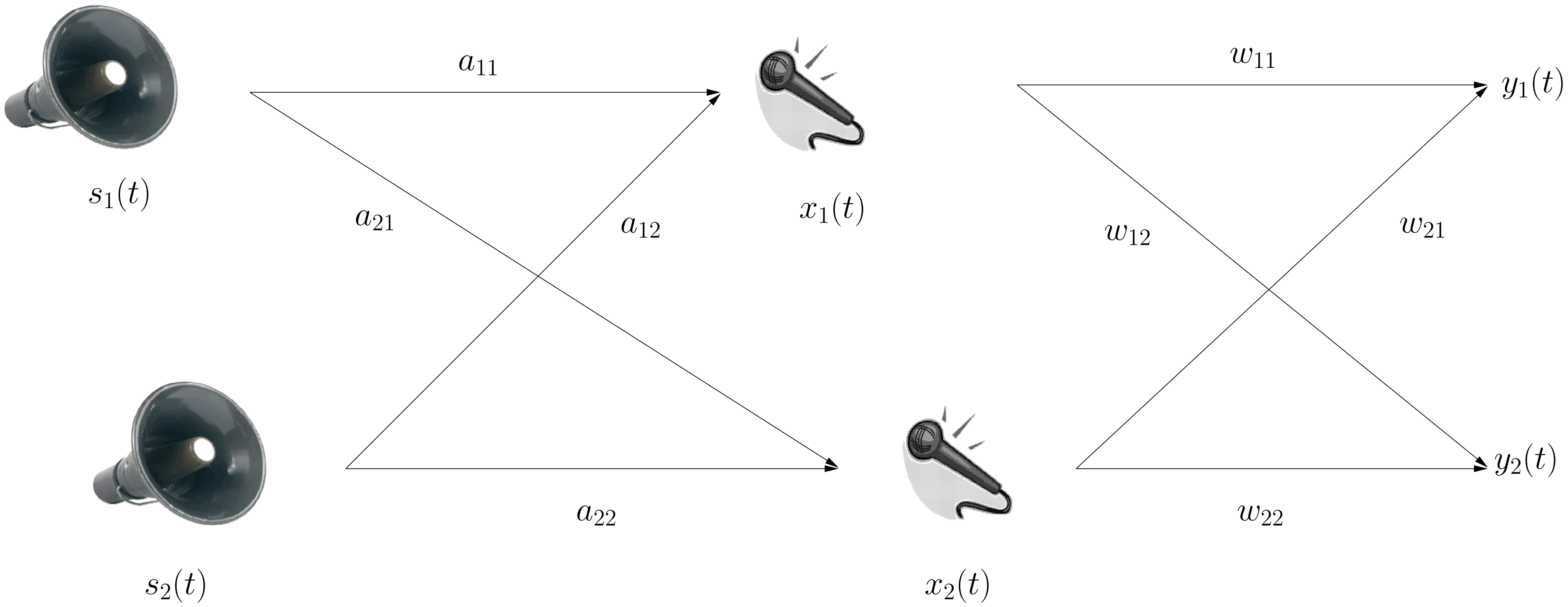
Independent Component Analysis (ICA)



Independent Component Analysis (ICA)



Independent Component Analysis (ICA)



ICA via Joint Diagonalization (JD)

$$y(t) = W^T x(t), \qquad x(t) = As(t)$$

Covariance matrices: $R_u(\tau) := E[u(t+\tau)u^T(t)]$.

Pick lags $\tau_1, \ldots, \tau_N$. It holds

$$R_y(\tau_1) = W^T R_x(\tau_1) W$$

$$\vdots$$

$$R_y(\tau_N) = W^T R_x(\tau_N) W.$$

Task: Select W to make $R_y(\tau_1), \ldots, R_y(\tau_N)$ "as diagonal as possible".

JD as optimization problem

Notation: $C_i := R_x(\tau_i)$.

Task: Make $W^T C_i W$, $i = 1, \ldots, N$, "as diagonal as possible".

Choose cost function to define the "best" joint diagonalization.

$$\tilde{f}(W) := \sum_{i=1}^{N} \left(\log\det \operatorname{ddiag}(W^T C_i W) - \log\det(W^T C_i W) \right).$$

Invariance property: $\tilde{f}(WD) = \tilde{f}(W)$ for all nonsingular diagonal matrix D.

Difficulty: The minimizers are not isolated.

JD as optimization on manifold

$$\tilde{f}(W) := \sum_{i=1}^{N} \left(\log \det \operatorname{ddiag}(W^T C_i W) - \log \det(W^T C_i W) \right).$$

Invariance $\tilde{f}(WD) = \tilde{f}(W)$, hence minimizers not isolated.

Two remedies:

1. Submanifold approach: restrict W to the *oblique manifold*

$$\mathcal{OB} := \{W \in \mathbb{R}^{n \times p} : \operatorname{ddiag}(W^T W) = I_p\}.$$

2. Quotient manifold approach: work on $\mathbb{R}^{n \times p}/\mathcal{D}$, the set of equivalence classes $[W] := W\mathcal{D} := \{WD : D \text{ diagonal}\}$.

Optimization On Manifolds

What ?

Why ?

How ?

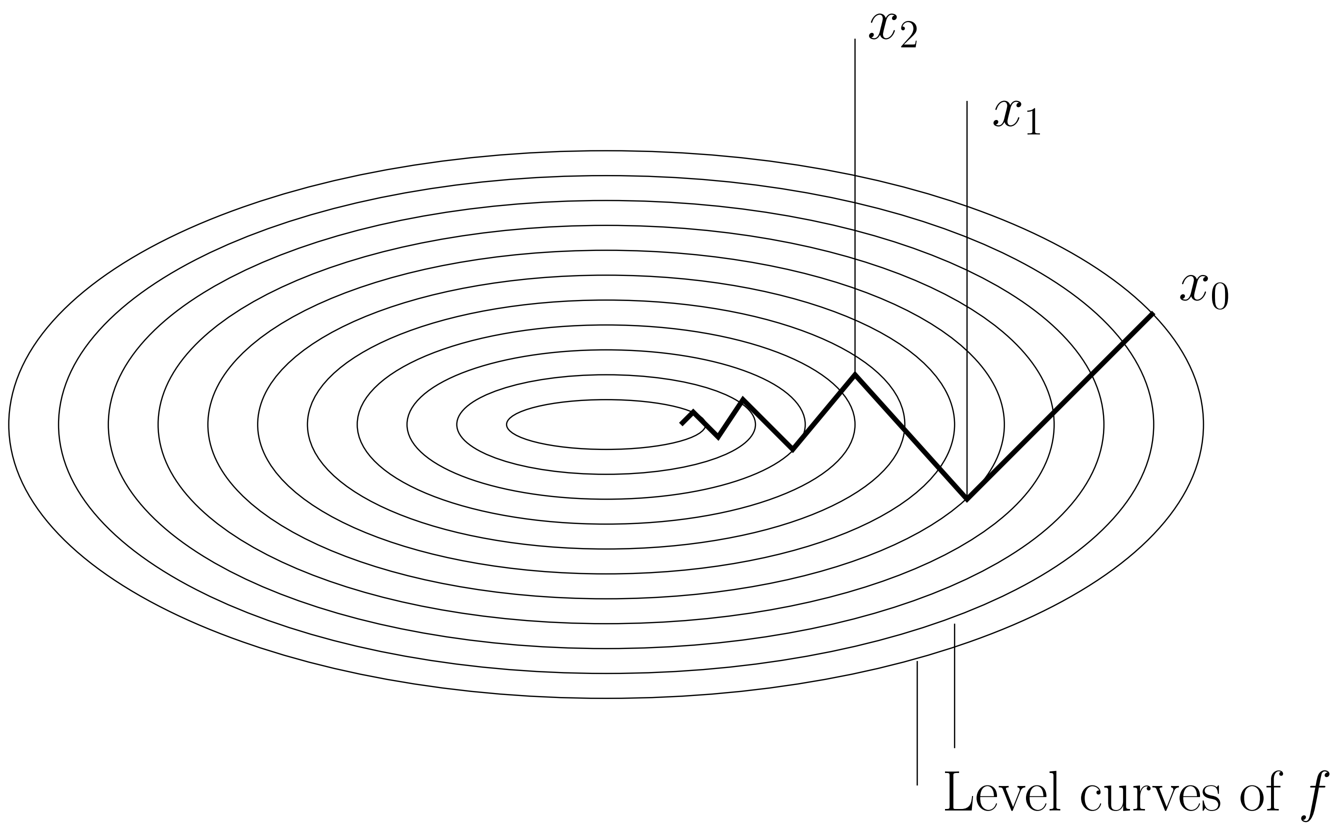
Optimization On Manifolds

What ?

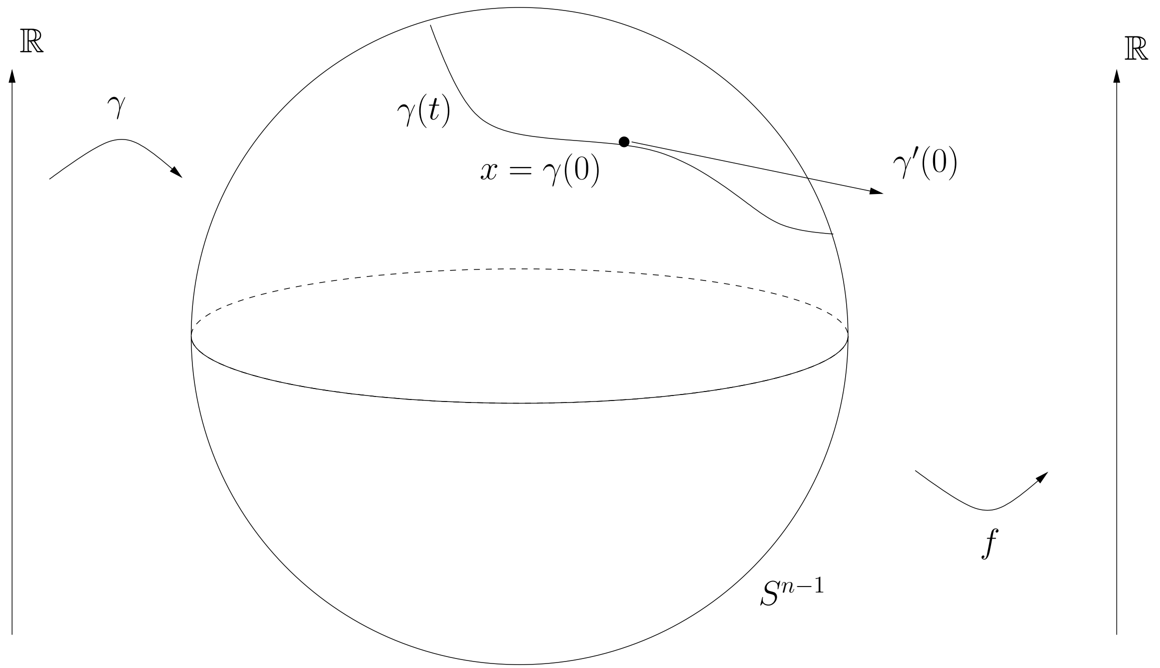
Why ?

How ?

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

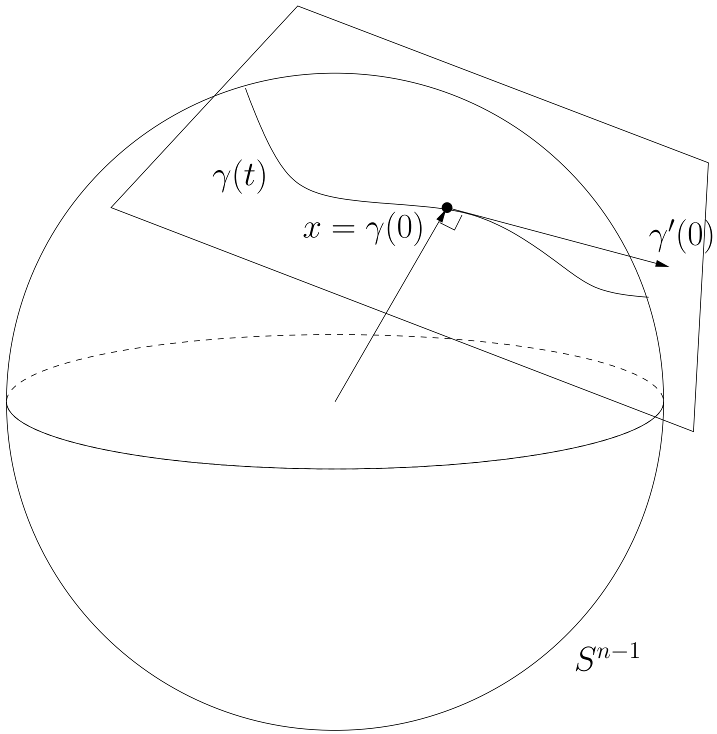


Steepest-descent on manifolds – Tangent vectors



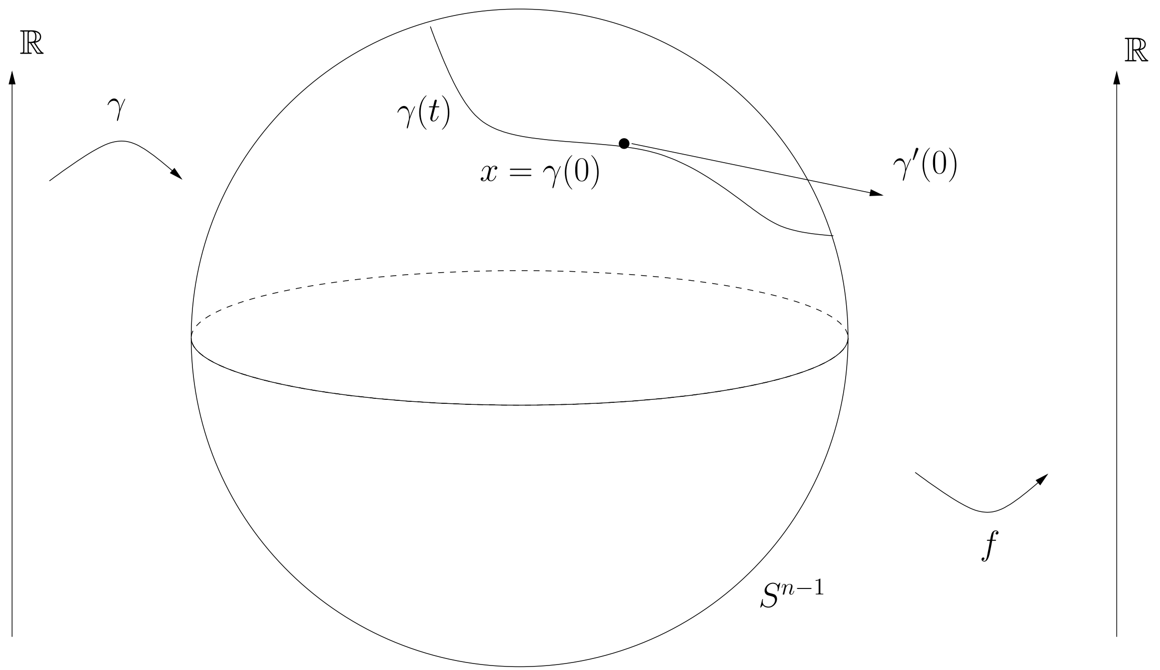
$$\gamma'(0): \quad f \in C^{\infty}(x) \quad \mapsto \quad \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\big|_{t=0} \in \mathbb{R}$$

Steepest-descent on manifolds – Tangent space



$$T_x\mathcal{M} = \{\gamma'(0) : \gamma \text{ curve in } \mathcal{M},\ \gamma(0) = x\}$$

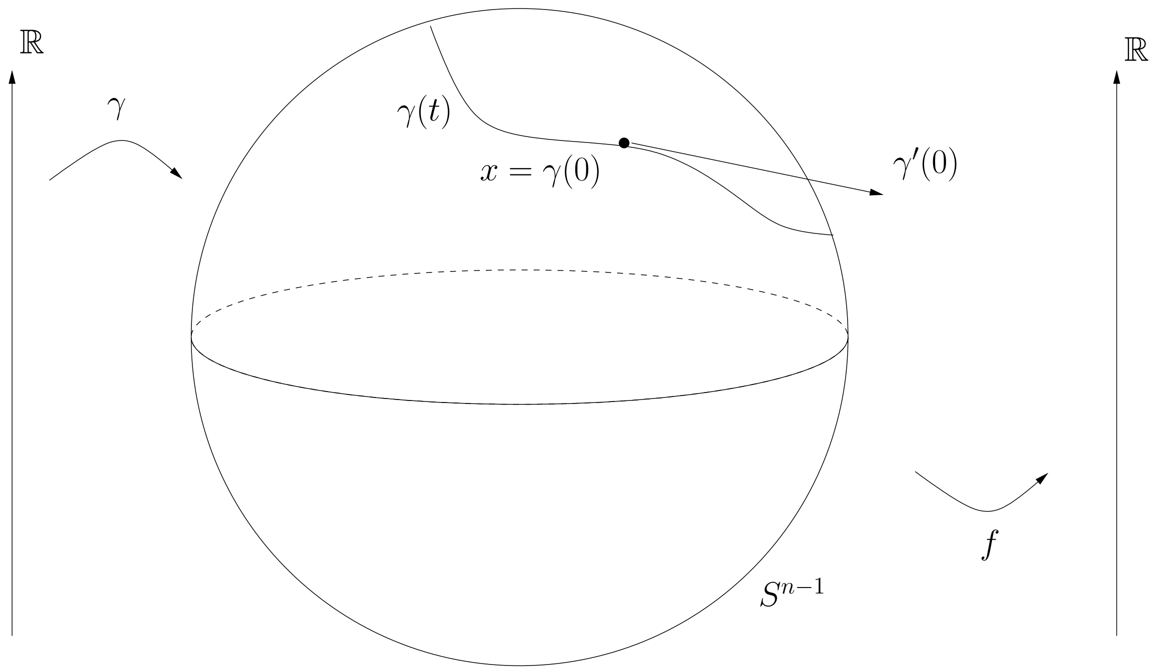
Steepest-descent on manifolds – Descent directions



$\gamma'(0)$ is a *descent direction* for f at x if

$$\gamma'(0)f := \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))|_{t=0} \quad < 0$$

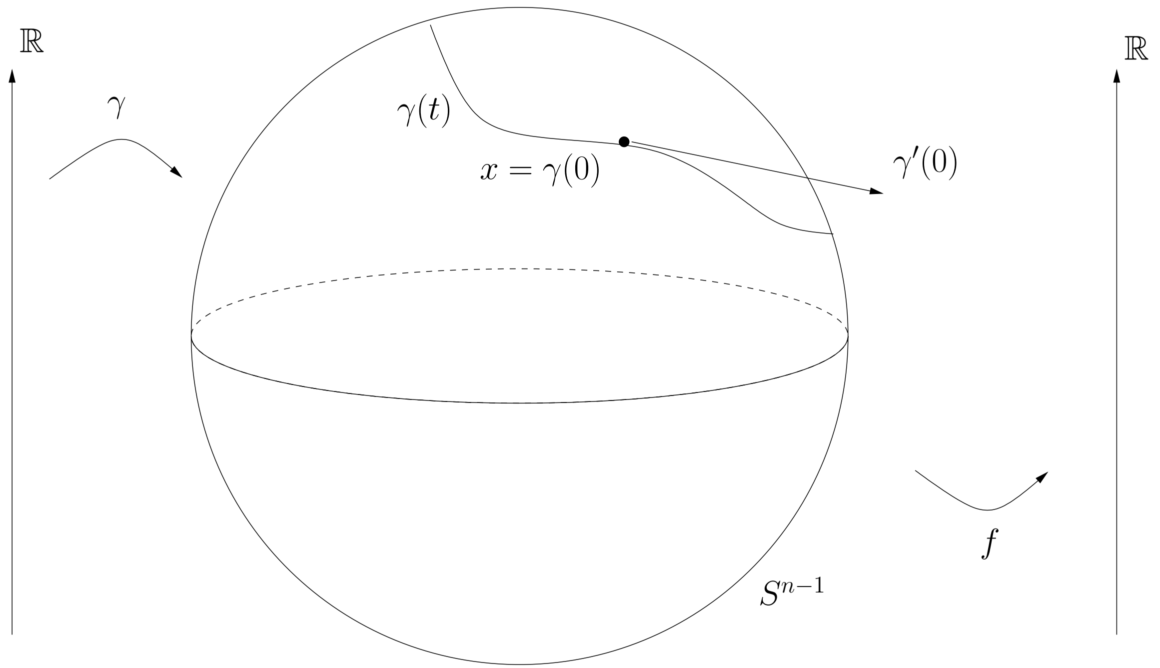
Steepest-descent on manifolds – Steepest descent direction



Define inner product $\langle \cdot, \cdot \rangle_x$ on the tangent space $T_x\mathcal{M}$. Then $\mathcal{M}$ is a *Riemannian manifold*.

Length of a tangent vector: $\|\gamma'(0)\|_x := \sqrt{\langle \gamma'(0), \gamma'(0) \rangle_x}$.

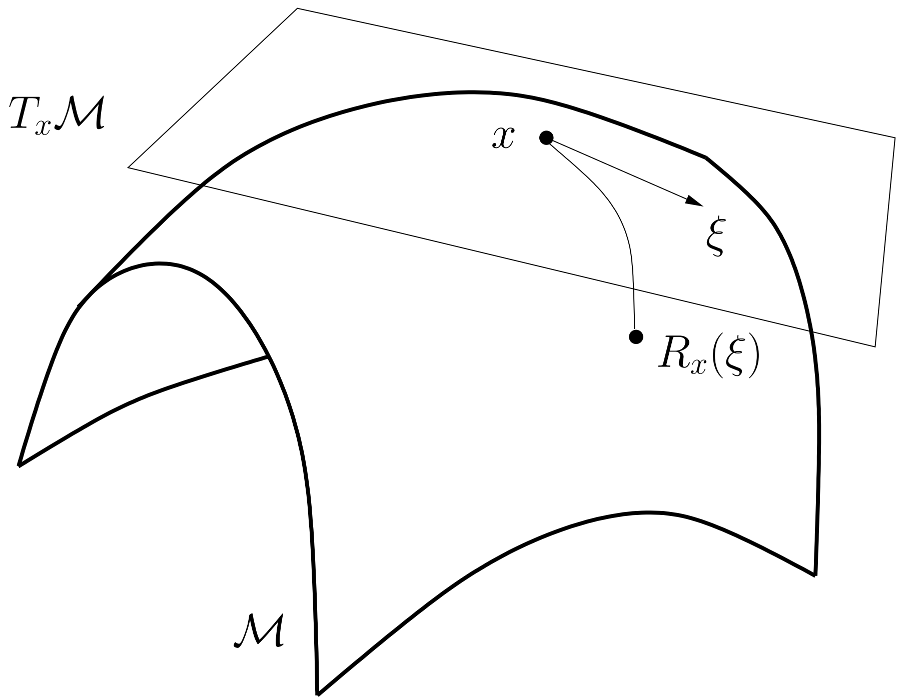
Steepest-descent on manifolds – Steepest descent direction



Steepest-descent direction along $\arg\min_{\xi \in T_x\mathcal{M},\ \|\xi\|_x = 1} \xi f$.

The steepest-descent direction is along the opposite of the *gradient* of f.

Steepest-descent on manifolds – Retraction



$$R_x(0_x) = x, \qquad \frac{\mathrm{d}}{\mathrm{d}t} R_x(t\xi)\Big|_{t=0} = \xi$$

Steepest-descent on manifolds – Summary

Let $\mathcal{M}$ be a Riemannian manifold with a retraction R. Let f be a cost function on $\mathcal{M}$. Let $x_0 \in \mathcal{M}$ be the initial iterate.

For $k = 0, 1, \ldots$:

1. Compute $\operatorname{grad} f(x_k)$.
2. Choose $x_{k+1} = R_{x_k}(-t \operatorname{grad} f(x_k))$ where $t > 0$ is chosen to satisfy a "sufficient decrease" condition.

Optimization On Manifolds

What ?

Why ?

How ?

A few pointers

- Optimization on manifolds in general: Luenberger [Lue73], Gabay [Gab82], Smith [Smi93, Smi94], Udrişte [Udr94], Manton [Man02], Mahony and Manton [MM02], PAA *et al.* [ABG06b]...
- 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], Simoncinin and Eldén [SE02], Brandts [Bra03], Absil *et al.* [AMSV02, AMS04, ASVM04, ABGS05, ABG06a] and Baker *et al.* [BAG06]

- 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 [GP05]...

Advertisement # 1: Graduate School

Course

"Optimization algorithms on matrix manifolds"

in the Graduate School on Systems, Optimization, Control and Networks (2007-2008)

Lecturers: PAA, Rodolphe Sepulchre

Advertisement # 2: forthcoming book

"Optimization algorithms on matrix manifolds"

by PAA, R. Mahony and R. Sepulchre, to appear (around December 2007)

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

References

[ABG06a] 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.

[ABG06b] ______, *Trust-region methods on Riemannian manifolds*, accepted for publication in Found. Comput. Math., doi:10.1007/s10208-005-0179-9, 2006.

[ABGS05] 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 et al. Peter M. A. Slot, eds.), Lecture Notes in Computer Science, vol. 3514, Springer-Verlag, 2005, pp. 33–41.

[ACC00] Shun-ichi Amari, Tian-Ping Chen, and Andrzej Cichocki, *Nonholonomic orthogonal learning algorithms for blind source separation*, Neural Computation **12** (2000), 1463–1484.

[ADM+02] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. 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.

[AG06] 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.

[AK04] 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.

[AMS04] 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.

[AMSV02] 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.

[ASVM04] 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.

[BAG06] 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, 2006, pp. 210–217.

[Bra03] J. Brandts, *The Riccati algorithm for eigenvalues and invariant subspaces of matrices with inexpensive action*, Linear Algebra Appl. **358** (2003), 335–365.

[CA01] T. P. Chen and S. Amari, *Unified stabilization approach to principal and minor components extraction algorithms*, Neural Networks **14** (2001), no. 10, 1377–1387.

[Dou00] Scott C. Douglas, *Self-stabilized gradient algorithms for blind source separation with orthogonality constraints*, IEEE Trans. Neural Networks **11** (2000), no. 6, 1490–1497.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353.

[Gab82] D. Gabay, *Minimizing a differentiable function over a differential manifold*, Journal of Optimization Theory and Applications **37** (1982), no. 2, 177–219.

[GP05] Igor Grubisic and Raoul Pietersz, *Efficient rank reduction of correlation matrices*, Tech. Report ERS-2005-009-F& A, Erasmus Research Institute of Management, January 2005, http://hdl.handle.net/1765/1933.

[HHLM07] Uwe Helmke, Knut Hüper, Pei Yean Lee, and John B. Moore, *Essential matrix estimation using Gauss-Newton iterations on a manifold*, International Journal of Computer Vision, to appear, 2007.

[HSS06] 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.

[JM02] Marcel Joho and Heinz Mathis, *Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation*, Proceedings of IEEE

Sensor Array and Multichannel Signal Processing Workshop SAM, 2002, pp. 273–277.

[JR02] Marcel Joho and Kamran Rahbar, *Joint diagonalization of correlation matrices by using Newton methods with applications to blind signal separation*, Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop SAM, 2002, pp. 403–407.

[LE02] E. Lundström and L. Eldén, *Adaptive eigenvalue computations using Newton's method on the Grassmann manifold*, SIAM J. Matrix Anal. Appl. **23** (2002), no. 3, 819–839.

[LM04] 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.

[LSG04] Xiuwen Liu, Anuj Srivastava, and Kyle Gallivan, *Optimal* *linear representations of images for object recognition*, IEEE Pattern Analysis and Machine Intelligence **26** (2004), no. 5, 662–666.

[Lue73] D. G. Luenberger, *Introduction to linear and nonlinear programming*, Addison-Wesley, 1973, Optimization techniques.

[Man02] J. H. Manton, *Optimization algorithms exploiting unitary constraints*, IEEE Trans. Signal Process. **50** (2002), no. 3, 635–650.

[MKS01] Yi Ma, Jana Kosecka, and Shankar S. Sastry, *Optimization criteria and geometric algorithms for motion and structure estimation*, International Journal of Computer Vision **44** (2001), no. 3, 219–249.

[MM02] R. Mahony and J. H. Manton, *The geometry of the Newton method on non-compact Lie groups*, J. Global Optim. **23** (2002), no. 3, 309–327.

[NA05] Yasunori Nishimori and Shotaro Akaho, *Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold*, Neurocomputing **67** (2005), 106–135.

[NMH02] Maziar Nikpour, Jonathan H. Manton, and Gen Hori, *Algorithms on the Stiefel manifold for joint diagonalization*, Proc. ICASSP, 2002, pp. II–1481–1484.

[Pha01] Dinh Tuan Pham, *Joint approximate diagonalization of positive definite Hermitian matrices*, SIAM J. Matrix Anal. Appl. **22** (2001), no. 4, 1136–1152.

[Plu05] M. D. Plumbley, *Geometrical methods for non-negative ICA: Manifolds, Lie groups and toral subalgebras*, Neurocomputing **67** (2005), 161–197.

[RR00] 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.

[SE02] Valeria Simoncini and Lars Eldén, *Inexact Rayleigh quotient-type methods for eigenvalue computations*, BIT **42** (2002), no. 1, 159–182.

[SHS06] Hao Shen, Knut Hüeper, 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.

[Shu86] M. 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.

[Smi93] S. T. Smith, *Geometric optimization methods for adaptive filtering*, Ph.D. thesis, Division of Applied Sciences, Harvard University, Cambridge, Massachusetts, 1993.

[Smi94] Steven T. Smith, *Optimization techniques on Riemannian* *manifolds*, Hamiltonian and gradient flows, algorithms and control, Fields Inst. Commun., vol. 3, Amer. Math. Soc., Providence, RI, 1994, pp. 113–136.

[TL02] N. T. Trendafilov and R. A. Lippert, *The multimode Procrustes problem*, Linear Algebra Appl. **349** (2002), 245–264.

[Udr94] C. Udrişte, *Convex functions and optimization methods on Riemannian manifolds*, Kluwer Academic Publishers, 1994.