Trust-region methods on Riemannian manifolds with applications in numerical linear algebra

Pierre-Antoine Absil

Christopher G. Baker

Kyle A. Gallivan

School of Computational Science and Information Technology,

Florida State University

Dynamical Systems and Computation Day

Thursday, July 15, 2004

CESAME, Université Catholique de Louvain, Belgium

Numerical Linear Algebra (NLA) problems

$$\equiv$$

Optimization on Manifold problems.

$\boxed{\text{Problem 1 : Minor Eigenvector Computation}}$

Given $n \times n$ matrix $A = A^T$ with (unknown) eigen-decomposition

$$A\,[v_1| \dots |v_n] = [v_1| \dots |v_n]\,\mathrm{diag}(\lambda_1, \dots, \lambda_n)$$

$$[v_1| \dots |v_n]^T\,[v_1| \dots |v_n] = I, \quad \lambda_1 < \lambda_2 \le \dots \le \lambda_n.$$

The problem is to compute the minor eigenvector $\pm v_1$.

Optimization problem on the unit sphere

Rayleigh quotient cost function:

$$f : S^{n-1} \to \mathbb{R} : y \mapsto y^T A y,$$

where S^{n-1} is the unit sphere $\{y \in \mathbb{R}^n : y^T y = 1\}$.

Useful properties:

- The stationary points of f are the eigenvectors of A.
- The local (and global) minima of f are $\pm v_1$.

Optimization problem on the real projective space

Rayleigh quotient cost function:

$$f : \mathbb{RP}^{n-1} \to \mathbb{R} : \mathrm{span}(y) \mapsto \frac{y^T A y}{y^T y},$$

where $\mathbb{RP}^{n-1}$ is the real projective space,
$\mathbb{RP}^{n-1} = \{\mathrm{span}(y) : y \in \mathbb{R}^n \setminus \{0\}\}$.

Useful properties:

- The stationary points of f are the "eigendirections" of A.
- The local (and global) minimum of f is $\mathrm{span}(v_1)$.

Problem 2 : Minor Eigenspace Computation

Given $n \times n$ matrix $A = A^T$ with (unknown) eigen-decomposition

$$A\,[v_1|\ldots|v_n] = [v_1|\ldots|v_n]\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

$$[v_1|\ldots|v_n]^T\,[v_1|\ldots|v_n] = I, \quad \lambda_1 \leq \ldots \leq \lambda_p < \lambda_{p+1} \leq \ldots \leq \lambda_n.$$

The problem is to compute the minor p-dimensional eigenspace $\mathrm{span}(v_1|\ldots|v_p)$.

Optimization problem on the Grassmann manifold

Rayleigh quotient cost function:

$$f : \mathrm{Grass}(p, n) \to \mathbb{R} : \mathrm{span}(Y) \mapsto \mathrm{trace}(Y^T A Y (Y^T Y)^{-1}),$$

where $\mathrm{Grass}(p, n)$ is the set of p-dimensional subspaces of $\mathbb{R}^n$.

Useful properties:

- The stationary points of f are the eigenspaces of A.
- The local (and global) minimum of f is $\mathrm{span}(v_1 | \ldots | v_p)$.

Problem 3 : Minor Eigenspace of Matrix Pencil

Given $n \times n$ matrix pencil (A, B), $A = A^T$, $B = B^T \succ 0$ with (unknown) eigen-decomposition

$$A\,[v_1|\ldots|v_n] = B\,[v_1|\ldots|v_n]\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

$$[v_1|\ldots|v_n]^T\,[v_1|\ldots|v_n] = I, \quad \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n.$$

Given integer p with $0 < p < n$.

The problem is to compute the minor p-dimensional eigenspace $\mathrm{span}(v_1|\ldots|v_p)$.

Optimization problem on the Grassmann manifold

Rayleigh quotient cost function:

$$f : \mathrm{Grass}(p, n) \to \mathbb{R} : \mathrm{span}(Y) \mapsto \mathrm{trace}(Y^T A Y (Y^T B Y)^{-1}),$$

where $\mathrm{Grass}(p, n)$ is the set of p-dimensional subspaces of $\mathbb{R}^n$.

Useful properties:

- The stationary points of f are the eigenspaces of (A, B).
- The local (and global) minimum of f is $\mathrm{span}(v_1|\ldots|v_p)$.

Numerical linear algebra problems

Several problems in numerical linear algebra can be expressed as finding a minimizer of a well-chosen cost function on a certain manifold. Examples:

- Full EVD. Full SVD.
- Full SVD.
- Balanced factorization.
- Nonlinear eigenvalue problem.
- Low rank approximation.
- ...

See Helmke and Moore [HM94], Lippert and Edelman [LE00] and references therein.

Manifolds

Roughly speaking, a manifold is a set that looks locally like $\mathbb{R}^n$. Local mappings from the manifold to $\mathbb{R}^n$ are called charts, and the inverse mappings are called parameterizations or systems of coordinates.

A Riemannian manifold is a manifold with a inner product on the tangent spaces, that varies in a smooth way.

References: do Carmo [dC92], Boothby [Boo75]...

Manifolds (cont'd)

The following manifolds are involved in the differential geometric approach to numerical linear algebra problems:

- Orthogonal group.
- Stiefel manifold: $n \times p$ orthonormal matrices.
- Grassmann manifold: p-dimensional subspaces in $\mathbb{R}^n$.
- Oblique manifold: matrices with normalized columns.
- Ellipsoids: $\{Y \in \mathbb{R}^{n\times p} : X^T R X = I\}$.
- Products of these manifolds.

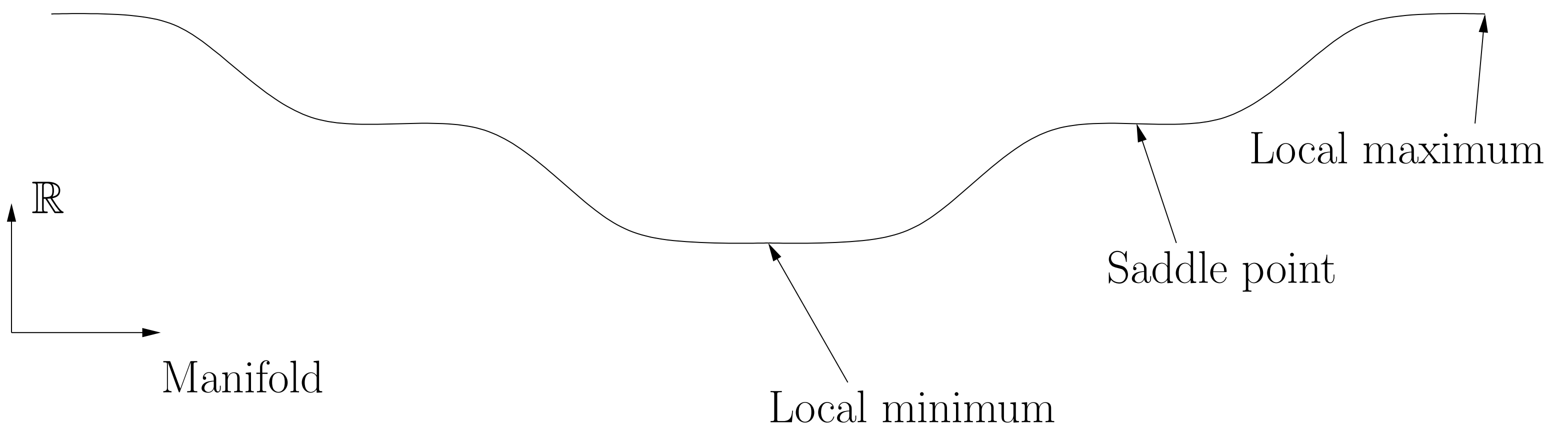
All these manifolds can be turned into Riemannian manifolds by smoothly defining an inner product on the tangent spaces.

Structure of the cost functions

Many cost functions related to linear algebra problems have:

- one or a few local minima which are also global minima,
- several other stationary points (i.e., critical points) that are either saddle points or local maxima.

We assume that only local minima are sought, although the other stationary points are sometimes interesting, too.



$\boxed{\text{Outline (Part 1)}}$

Numerical Linear Algebra (NLA) problems

$$\equiv$$

Optimization on Manifold problems.

- Motivating example: Minor Component Analysis.
- 'Conventional' methods: simple vector iterations.
- 'Unconventional' methods: optimization on manifold (here, the sphere).
- Need for a more efficient method with detailed convergence analysis.

Motivating problem : Minor Eigenvector Computation

Given $n \times n$ matrix $A = A^T$ with (unknown) eigen-decomposition

$$A\,[v_1|\ldots|v_n] = [v_1|\ldots|v_n]\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

$$[v_1|\ldots|v_n]^T\,[v_1|\ldots|v_n] = I, \quad 0 < \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n.$$

The problem is to compute the minor eigenvector $\pm v_1$.

$\boxed{\text{Outline (Part 1)}}$

Numerical Linear Algebra (NLA) problems

$$\equiv$$

Optimization on Manifold problems.

- Motivating example: Minor Component Analysis.
- 'Conventional' methods: simple vector iterations.
- 'Unconventional' methods: optimization on manifold (here, the sphere).
- Need for a more efficient method with detailed convergence analysis.

Simple vector iterations: Inverse Iteration

$$y_{k+1} = \frac{A^{-1} y_k}{\|A^{-1} y_k\|}$$

Properties:

- Global convergence to $\{\pm v_1, \ldots, \pm v_n\}$.
- Stable convergence to $\pm v_1$ only.
- Local linear convergence, with ratio $\frac{\lambda_1}{\lambda_2}$.

  Exemple: $n = 100$, $\lambda_i = i/n$ (evenly spaced eigenvalues on $(0, 1]$). Then $\frac{\lambda_1}{\lambda_2} = 0.5$.

  Possible evolution: error(1)=0.1, error(2)=0.05, error(3)=0.0025,...,error(27)$\simeq 1.4 \cdot 10^{-9}$.
- Computing a new iterate is expensive.

Simple vector iterations: Rayleigh Quotient Iteration (RQI)

$$\rho_k = \frac{y_k^T A y_k}{y_k^T y_k}$$

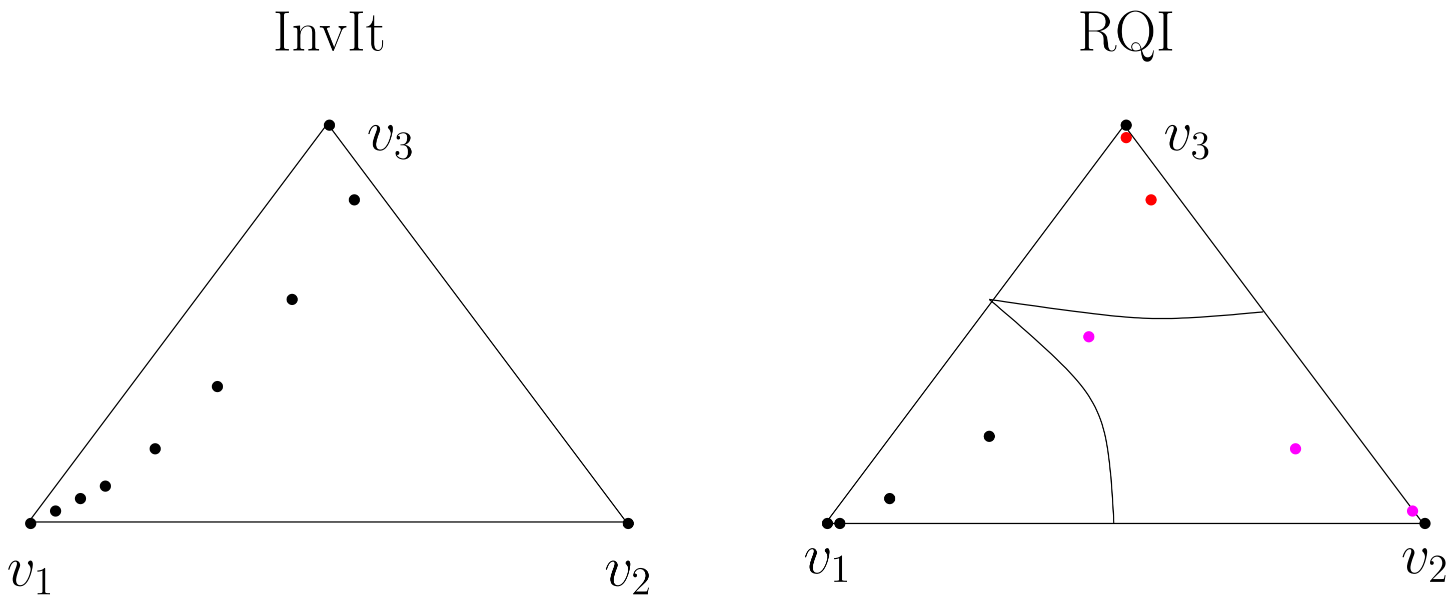
$$y_{k+1} = \frac{(A - \rho_k I)^{-1} y_k}{\|(A - \rho_k I)^{-1} y_k\|}$$

Properties:

- Converges to "nearest" eigenvector.
- Cubic local convergence.

  Possible evolution: error(1)=0.1, error(2)= 10^{-3}, error(3)= 10^{-9}.
- Computing a new iterate is expensive.

Simple vector iterations : illustrations



The ideal minor component algorithm

Given $A = A^T \succ 0$, with eigenvalues $0 < \lambda_1 \leq \ldots \lambda_n$ and associated eigenvectors $v_1, \ldots, v_n$.

1. Convergence to some eigenvector for all initial conditions.
2. Stable convergence to the minor eigenvector $\pm v_1$ only.
3. Superlinear (cubic) local convergence to $\pm v_1$.
4. No factorization of A.

   Matrix A only utilized as operator $x \mapsto Ax$.
5. Minimal storage space required.

$\boxed{\text{Outline (Part 1)}}$

Numerical Linear Algebra (NLA) problems

$$\equiv$$

Optimization on Manifold problems.

- Motivating example: Minor Component Analysis.
- 'Conventional' methods: simple vector iterations.
- 'Unconventional' methods: optimization on manifold (here, the sphere).
- Need for a more efficient method with detailed convergence analysis.

Optimization problem on the unit sphere

Rayleigh quotient cost function:

$$f : S^{n-1} \to \mathbb{R} : y \mapsto y^T A y,$$

where S^{n-1} is the unit sphere $\{y \in \mathbb{R}^n : y^T y = 1\}$.

Useful properties:

- The stationary points of f are the eigenvectors of A.
- The local (and global) minima of f are $\pm v_1$.

Available optimization methods on manifolds

A few references: Gabay [Gab82], Udrişte [Udr94], Smith [Smi94], Edelman *et al.* [EAS98], Manton [Man02].

It seems that all currently available methods on manifolds are either

- globally convergent but slow (linear), for example gradient descent methods; or
- fast but not (provably) globally convergent, for example the Newton method.

Moreover, exact Newton steps are expensive to compute.

Requirements on the optimization method

To achieve the "ideal minor component method", we need an optimization method with the following properties:

1. Global convergence to stationary points.
2. Stable convergence to local minima only.
3. Superlinear local convergence.
4. No factorization of the Hessian.
5. Minimal storage space needed.

Outline (Part 1)

Numerical Linear Algebra (NLA) problems

$$\equiv$$

Optimization on Manifold problems.

- Motivating example: Minor Component Analysis.
- ‘Conventional’ methods: simple vector iterations.
- ‘Unconventional’ methods: optimization on manifold (here, the sphere).
- Need for a more efficient method with detailed convergence analysis.

$\boxed{\text{In } \mathbb{R}^n\text{: Yes!}}$

TRUST-REGION METHOD

where the trust-region subproblems are solved with a

TRUNCATED CONJUGATE-GRADIENT

algorithm.

$\boxed{\text{Outline (Part 2)}}$

- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Minor Component Analysis.
  - Properties of the trust-region algorithm.
  - Extensions... Competitors...
  - Numerical experiments.

Principle of Trust-Region (TR) in $\mathbb{R}^n$

1. Consider a cost function f in $\mathbb{R}^n$. Let x_k be the current iterate.
2. Build a model $m_k(s)$ of f around x_k. The model should agree to f at x_k to the first order at least, and to the second order if superlinear convergence is sought.
3. Find (up to some precision) a minimizer s_k of the model within a "trust-region", i.e., a ball of radius Δ_k around x_k.
4. Compute the ratio

$$\rho = \frac{f(x_k) - f(x_k + s_k)}{m_k(0) - m_k(s_k)}$$

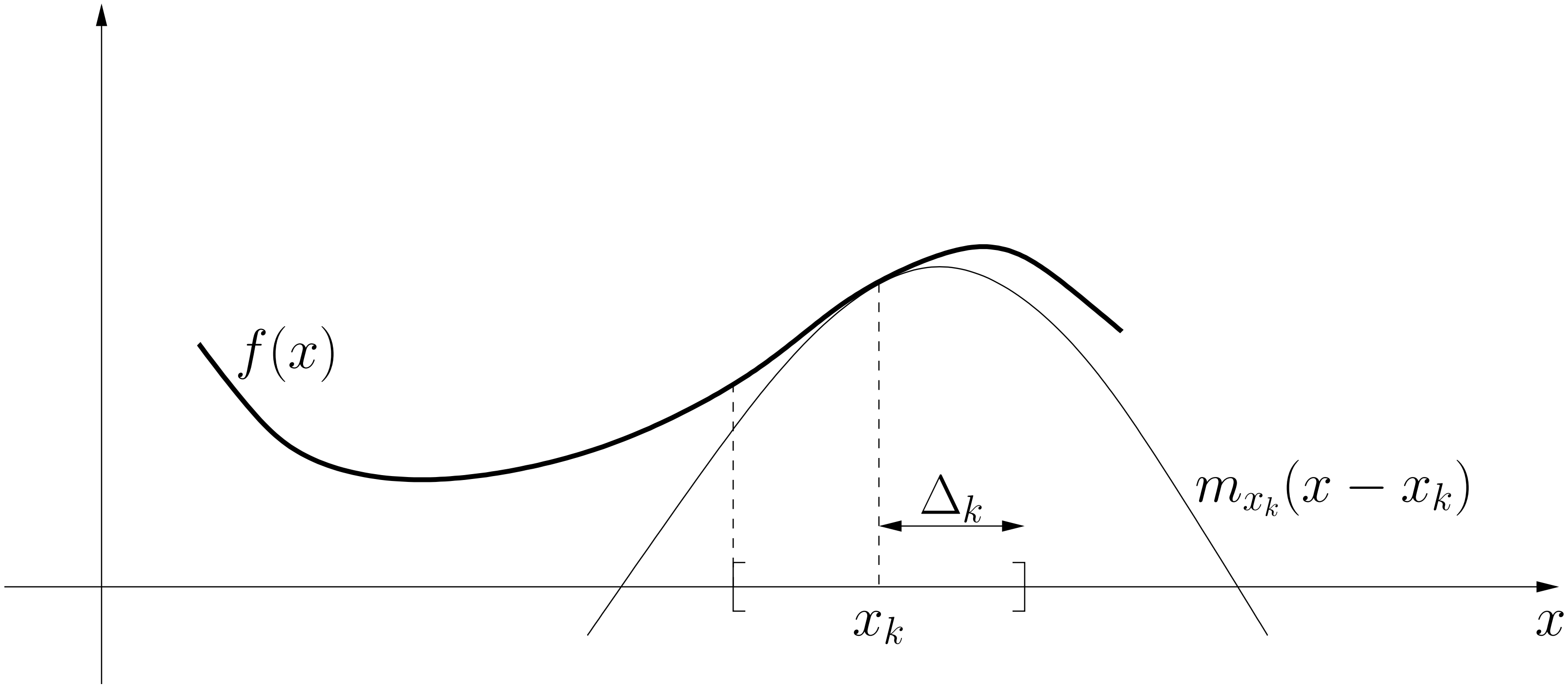
to compare the actual value of the cost function at the proposed new iterate with the value predicted by the model.

Principle of Trust-Region (TR) in $\mathbb{R}^n$ (cont'd)

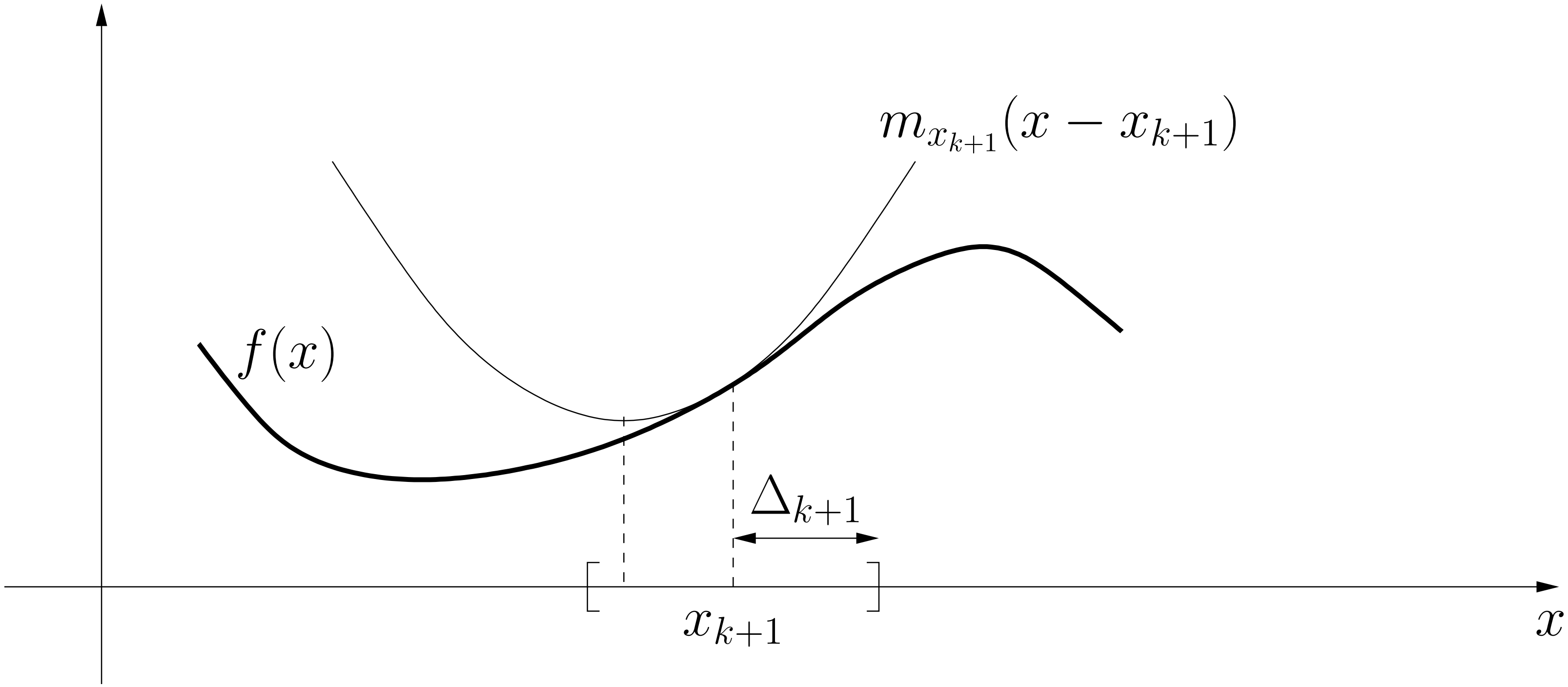
5. Shrink, enlarge or keep the trust-region radius according to the value of ρ.
6. Accept or reject the proposed new iterate $x_k + s_k$ according to the value of ρ.
7. Increment k and go to step 2.

For more detail, see e.g. [NW99, CGT00].

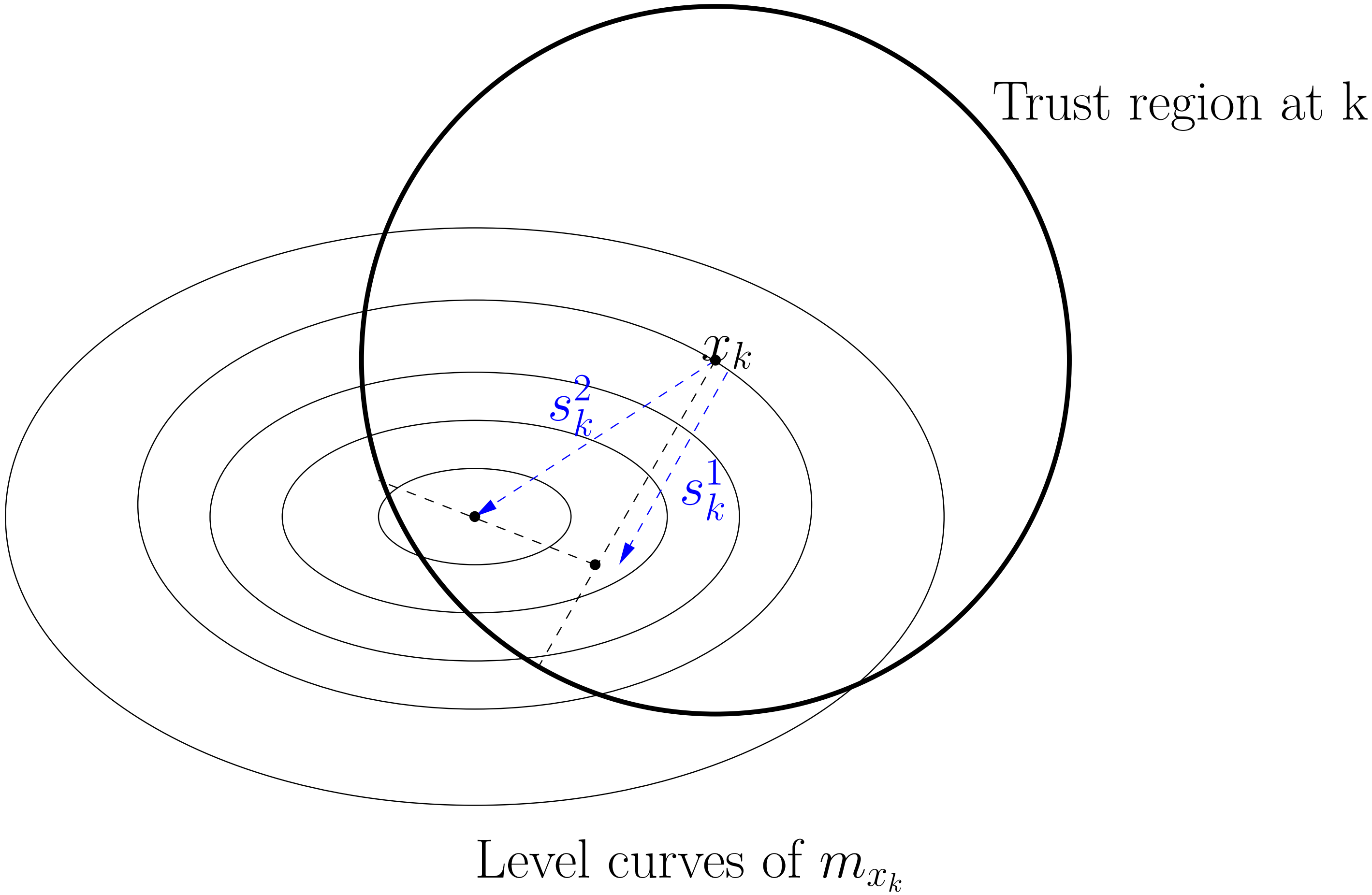
Principle of Trust-Region (TR) in $\mathbb{R}^n$



Principle of Trust-Region (TR) in $\mathbb{R}^n$



Principle of truncated CG (tCG)



Stopping criterion for tCG

Reasons for stopping tCG (inner iteration):

- The line-search algorithm hits the trust-region boundary. (This happens in particular when the model has a negative curvature along the current direction of search.)
- The norm of the residual has become sufficiently small. Criterion:

$$\|r_j\| \leq \|r_0\| \min(\|r_0\|^\theta, \kappa).$$

  Note that $r_n = 0$ in exact arithmetic (theory of linear CG).

$\longrightarrow$ Expected order of convergence:

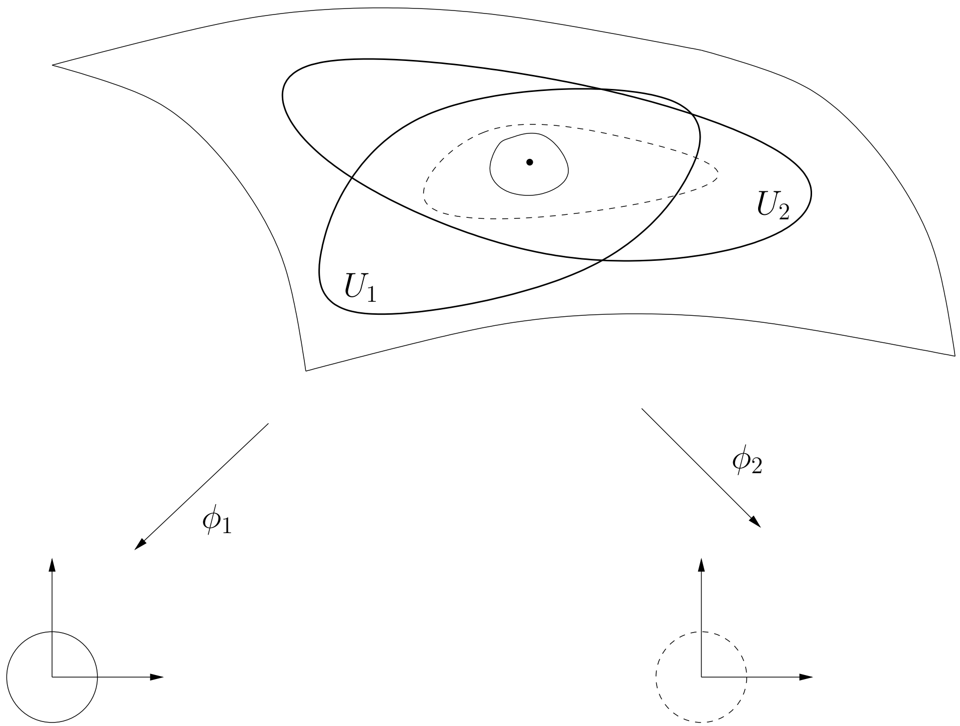
$$\min\{\theta + 1, 2\}.$$

Outline (Part 2)

- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Minor Component Analysis.
  - Properties of the trust-region algorithm.
  - Extensions... Competitors...
  - Numerical experiments.

Trust-region methods on Riemannian manifolds: difficulties

In general, coordinates systems can be scaled without restriction: If ϕ is a chart, then $\alpha\phi$ is still a chart, with $\alpha \in \mathbb{R}$.



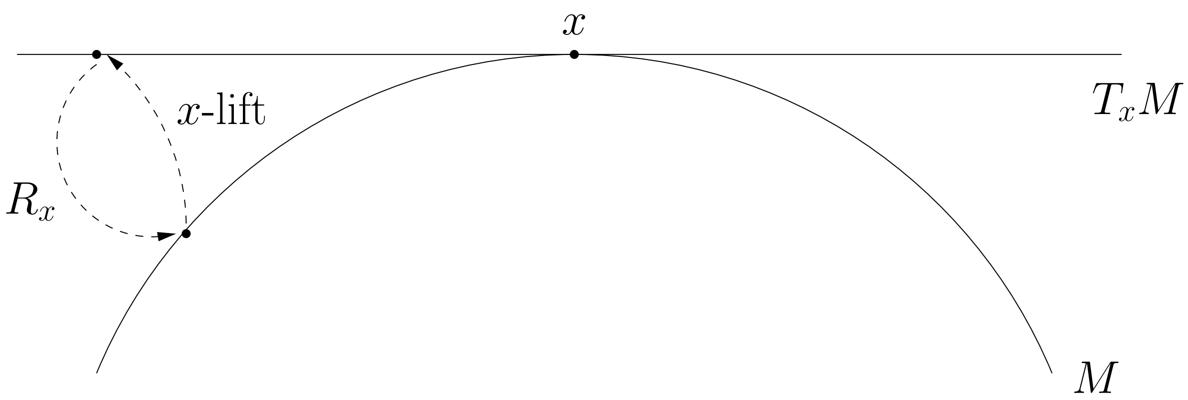
Trust-region methods on Riemannian manifolds: remedies

To define a notion of trust-region on Riemannian manifolds, one has to use charts with some "rigidity" property.

To assign a "locally rigid" chart to any point on a manifold M, we use the concept of *retraction* introduced (?) in Adler *et al.* [ADM$^+$02].

Trust-region methods on Riemannian manifolds: remedies (cont'd)

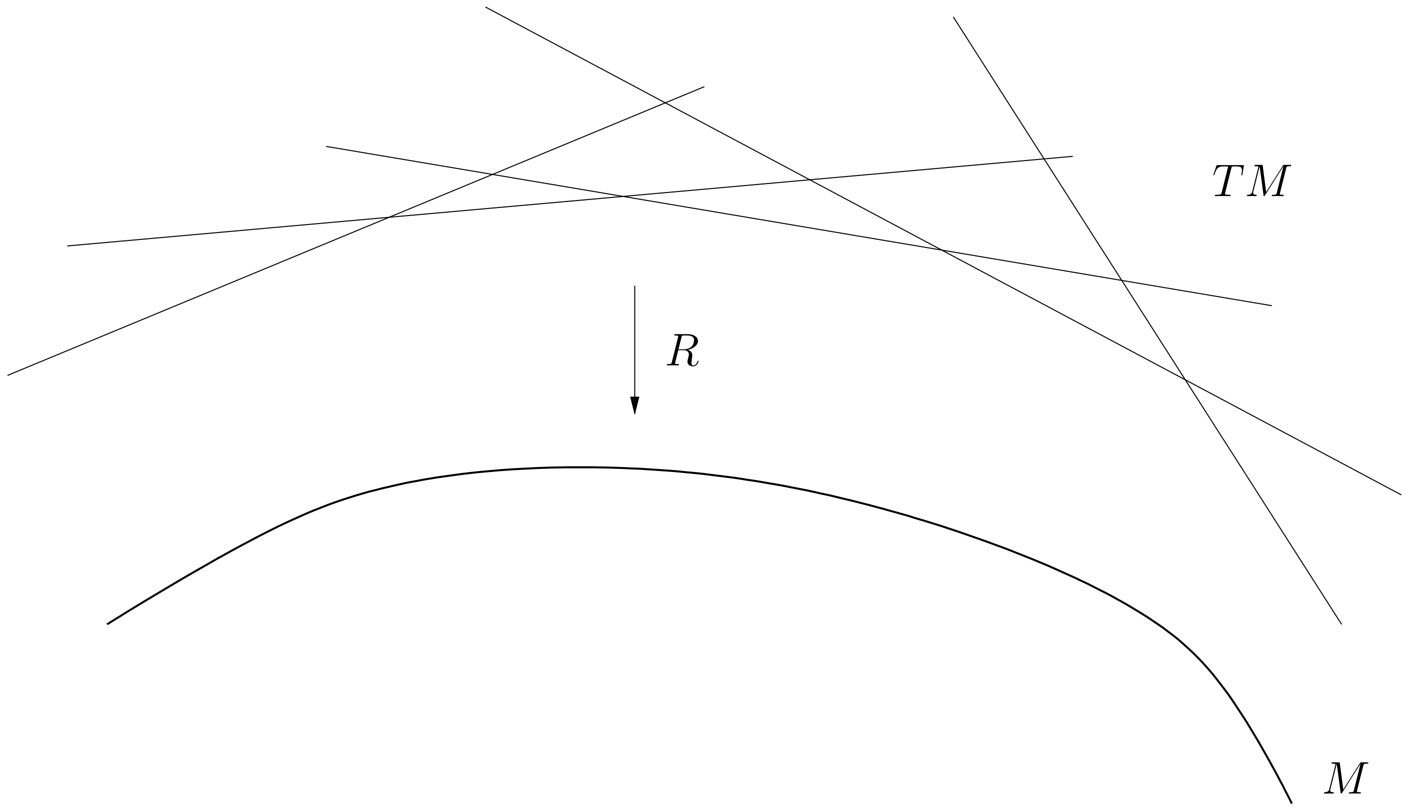
Concept of *retraction*:



1. R_x is defined and one-to-one in a neighbourhood of 0_x in T_xM.
2. $R_x(0_x) = x$.
3. $\mathrm{D}R_x(0_x) = \mathrm{id}_{T_xM}$, the identity mapping on T_xM, with the canonical identification $T_{0_x}T_xM \simeq T_xM$.

Trust-region methods on Riemannian manifolds: remedies (cont'd)

Retraction as a mapping from the tangent bundle TM to M.



Trust-region methods on Riemannian manifolds

1. Given: smooth manifold M; Riemannian metric g; smooth cost function f on M; retraction R from the tangent bundle TM to M; current iterate x_k.

1b. Lift up the cost function to the tangent space T_xM:

$$\hat{f}_x = f \circ R_x.$$

2. Build a model $m_k(s)$ of $\hat{f}_x$ around x_k.
3. Find (up to some precision) a minimizer s_k of the model within a "trust-region", i.e., a ball of radius Δ_k around x_k.

Trust-region methods on Riemannian manifolds (cont'd)

4. Compute the ratio

$$\rho = \frac{f(x_k) - f(R_{x_k} s_k)}{m_k(0) - m_k(s_k)}$$

   (note the presence of R_{x_k} !) to compare the actual value of the cost function at the proposed new iterate with the value predicted by the model.

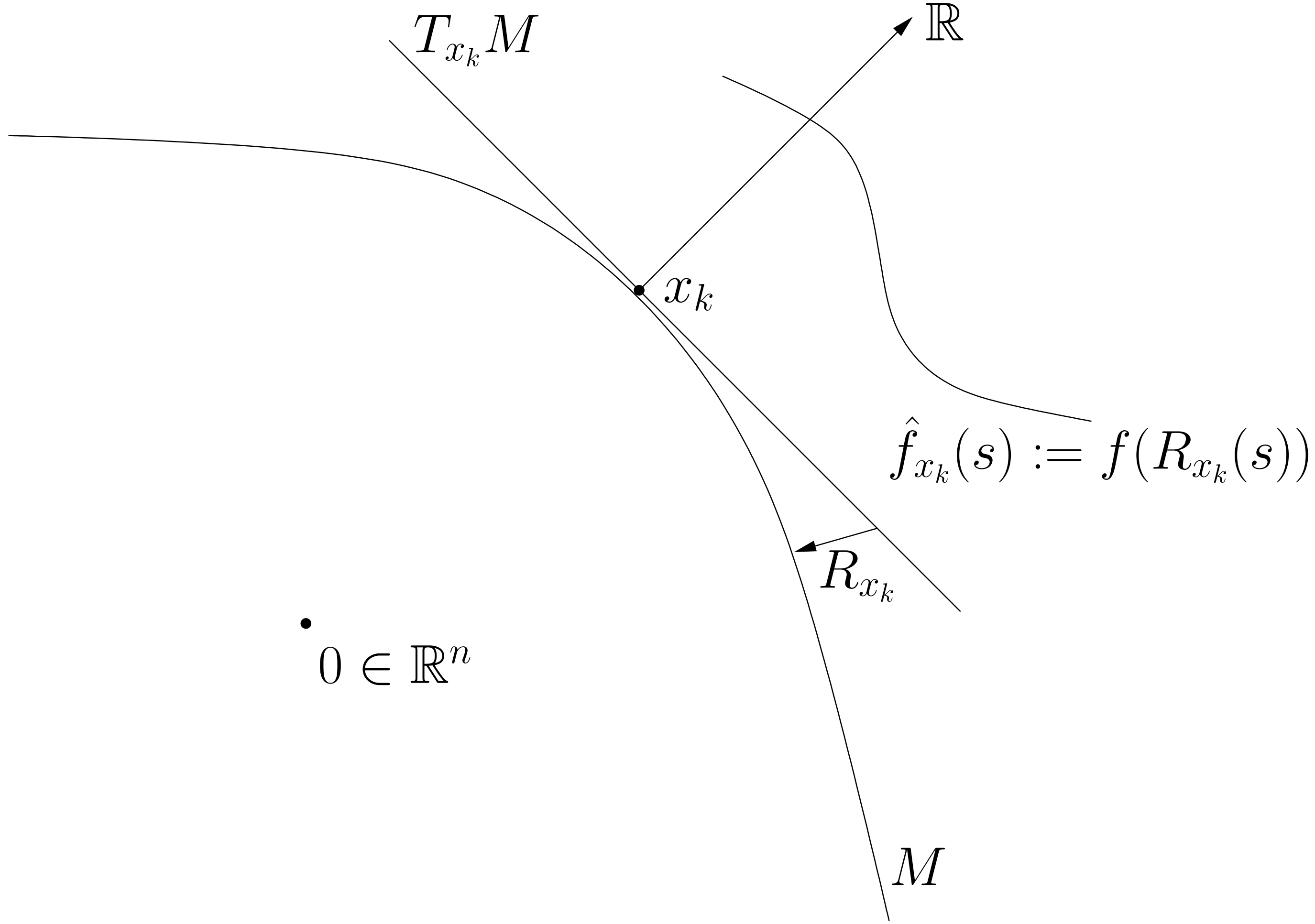
5. Shrink, enlarge or keep the trust-region radius according to the value of ρ.
6. Accept or reject the proposed new iterate $R_{x_k} s_k$ according to the value of ρ.
7. Increment k and go to step 2.

Solving the TR subproblem: truncated CG

- Start from the point $s^0 = 0$.
- Compute the first search direction $\delta^0 = -\mathrm{grad}\, f(x_k)$.
- Minimize the model $m_k(s)$ along δ_0 within the trust region. This yields s^1. If the boundary is reached, then stop.
- Compute the conjugate-gradient direction δ^1.
- Minimize the model along $s^1 + \alpha\delta^2$. If the boundary if reached, then stop.
- ... Repeat the procedure until some stopping criterion is satisfied, and return $s_k := s^j$.

Stopping criteria are based on the norm of the residual $\nabla m_k(s^j)$.

Principle of TR on Riemannian manifold



Required ingredients for Riemannian TR

- Manifold M, Riemannian metric g, and cost function f on M.
- Practical expression for $T_{x_k}M$.
- Retraction $R_{x_k} : T_{x_k}M \to M$.
- Function $\hat{f}_{x_k}(s) := f(R_{x_k}(s))$.
- Gradient $\operatorname{grad} \hat{f}_{x_k}(0)$.
- Hessian $\operatorname{Hess} \hat{f}_{x_k}(0)$.

Outline (Part 2)

- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Minor Component Analysis.
  - Properties of the trust-region algorithm.
  - Extensions... Competitors...
  - Numerical experiments.

Global convergence result

Let $\{x_k\}$ be a sequence of iterates generated by the RTR algorithm with $\rho' \in (0, \frac{1}{4})$. Suppose that f is C^2 and bounded below on the level set $\{x \in M : f(x) < f(x_0)\}$. Suppose that $\|\operatorname{grad} f(x)\| \leq \beta_g$ and $\|\operatorname{Hess} f(x)\| \leq \beta_H$ for some constants β_g, β_H, and all $x \in M$. Moreover suppose that

$$\|\tfrac{D}{dt}\tfrac{d}{dt} Rt\xi\| \leq \beta_D \tag{1}$$

for some constant β_D, for all $\xi \in TM$ with $\|\xi\| = 1$ and all $t < \delta_D$, where $\frac{D}{dt}$ denotes the covariant derivative along the curve $t \mapsto Rt\xi$. Further suppose that all approximate solutions s_k of the trust-region subproblems produce a decrease of the model that is at least a fixed fraction of the Cauchy decrease.

Global convergence result (cont'd)

It then follows that

$$\lim_{k\to\infty} \operatorname{grad} f(x_k) = 0.$$

And only the local minima are stable (the saddle points and local maxima are unstable).

Local convergence result

Consider the RTR-tCG algorithm. Suppose that f is a C^2 cost function on M and that

$$\|\mathcal{H}_k - \operatorname{Hess} \hat{f}_{x_k}(0_k)\| \leq \beta_{\mathcal{H}} \|\operatorname{grad} f(x_k)\|. \tag{2}$$

Let $v \in M$ be a nondegenerate local minimum of f, (i.e., $\operatorname{grad} f(v) = 0$ and $\operatorname{Hess} f(v)$ is positive definite). Further assume that $\operatorname{Hess} \hat{f}_{x_k}$ is Lipschitz-continuous at 0_x uniformly in x in a neighborhood of v, i.e., there exist $\beta_1 > 0$, $\delta_1 > 0$ and $\delta_2 > 0$ such that, for all $x \in B_{\delta_1}(v)$ and all $\xi \in B_{\delta_2}(0_x)$, it holds

$$\|\operatorname{Hess} \hat{f}_{x_k}(\xi) - \operatorname{Hess} \hat{f}_{x_k}(0_{x_k})\| \leq \beta_{L2} \|\xi\|. \tag{3}$$

Local convergence result (cont'd)

Then there exists $c > 0$ such that, for all sequences $\{x_k\}$ generated by the RTR-tCG algorithm converging to v, there exists $K > 0$ such that for all $k > K$,

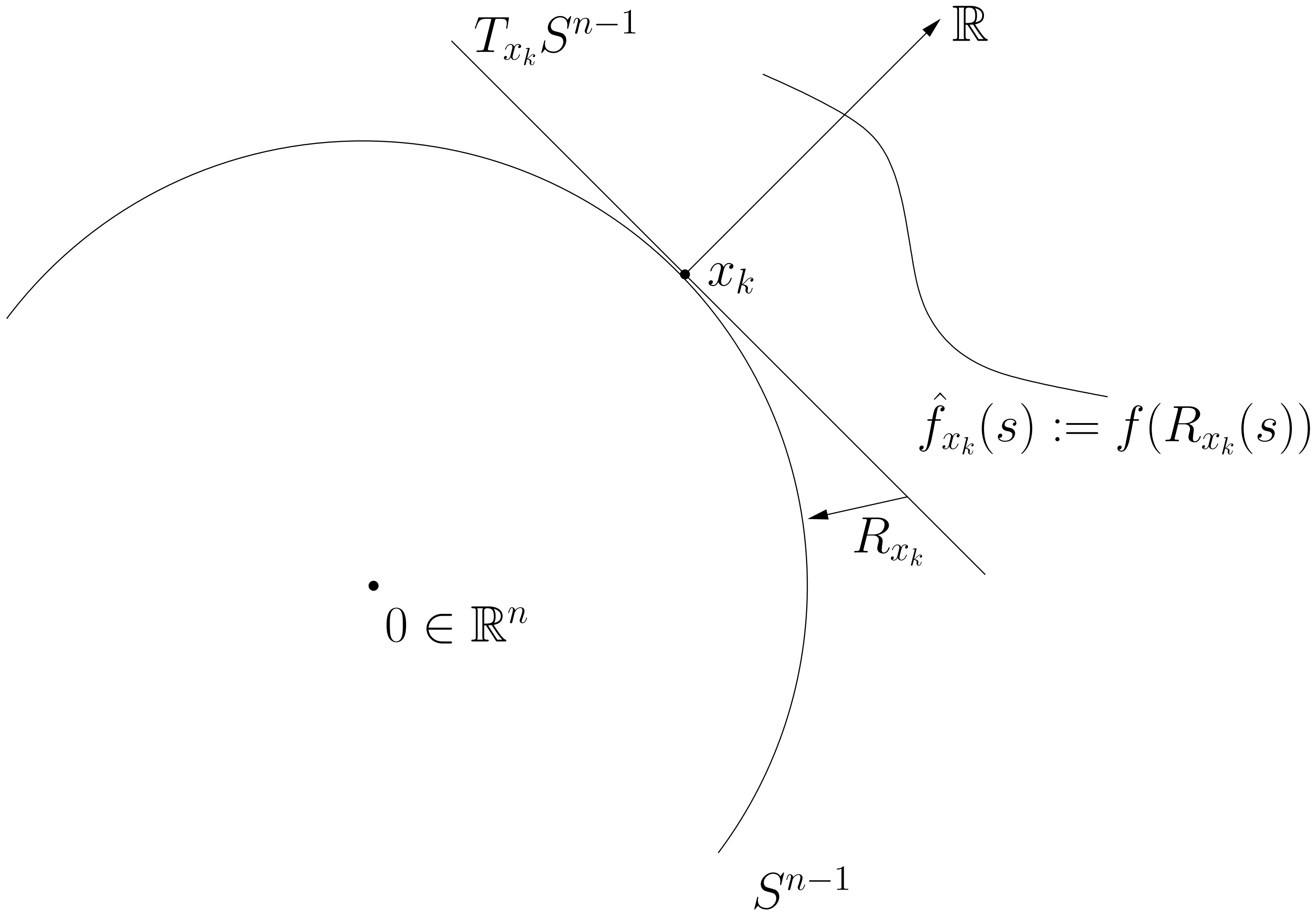
$$\mathrm{dist}(x_{k+1}, v) \leq c\,(\mathrm{dist}(x_k, v))^{\min\{\theta+1,2\}}, \tag{4}$$

where θ governs the stopping criterion of the tCG inner iteration.

Outline (Part 2)

- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Minor Component Analysis.
  - Properties of the trust-region algorithm.
  - Extensions... Competitors...
  - Numerical experiments.

Trust-region on the sphere



Properties of the algorithm

Algorithm: Riemannian Trust-Region method on the sphere with truncated-CG algorithm for minimizing the Rayleigh quotient.

Properties:

1. For all initial conditions, $\{y_k\}$ converges to an eigenvector.
2. Only the minor eigenvector $\pm v_1$ is stable.
3. Superlinear rate, with exponent $\min\{\theta + 1, 3\}$.
4. No factorization of A.
5. Minimal storage space needed (CG process).

Outline (Part 2)

- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Minor Component Analysis.
  - Properties of the trust-region algorithm.
  - Extensions... Competitors...
  - Numerical experiments.

Extensions of the algorithm

- $A = A^T \not\succ 0$. Then the algorithm computes the "leftmost eigenvector".
- Therefore, applied to $-A$, the algorithm computes the rightmost eigenvector.
- Algorithm for the symmetric/positive-definite generalized eigenvalue problem

$$Ax = \lambda Bx,$$

  using the Rayleigh quotient $y \mapsto (y^T Ay)/(y^T By)$.
- Block version. The iterates are $n \times p$ matrices Y. The cost function is $Y \mapsto \operatorname{trace}\left(Y^T AY(Y^T BY)^{-1}\right)$ and the relevant domain is the *Grassmann manifold.*

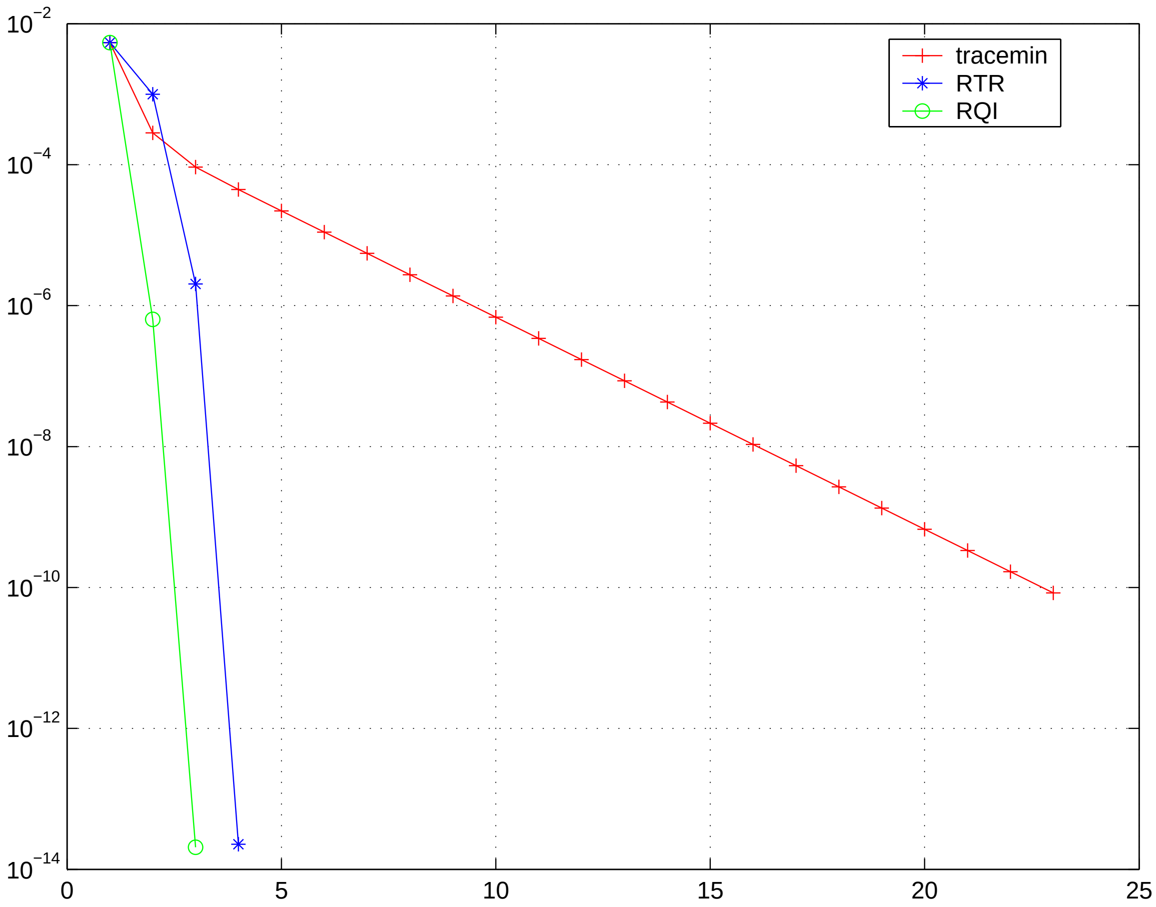
Competitors ?

- Scott [Sco81]. Restarted Lanczos method for the generalized eigenproblem, superlinear convergence, without matrix inversion.
  But the storage space becomes very large to ensure superlinear convergence. No proof of convergence.
- Golub and Ye [GY02]. Restarted Lanczos method for the generalized eigenproblem.
  But linear convergence (unless ideal preconditioning).
- SG algorithms of Lippert and Edelman [LE00]. Close precursors.
  But global convergence and generalized eigenproblem not considered.
- Nikpour *et al.* [NMMA04]...
  (See also Mongeau and Torki [MT99].)

Outline (Part 2)

- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Minor Component Analysis.
  - Properties of the trust-region algorithm.
  - Extensions... Competitors...
  - Numerical experiments.

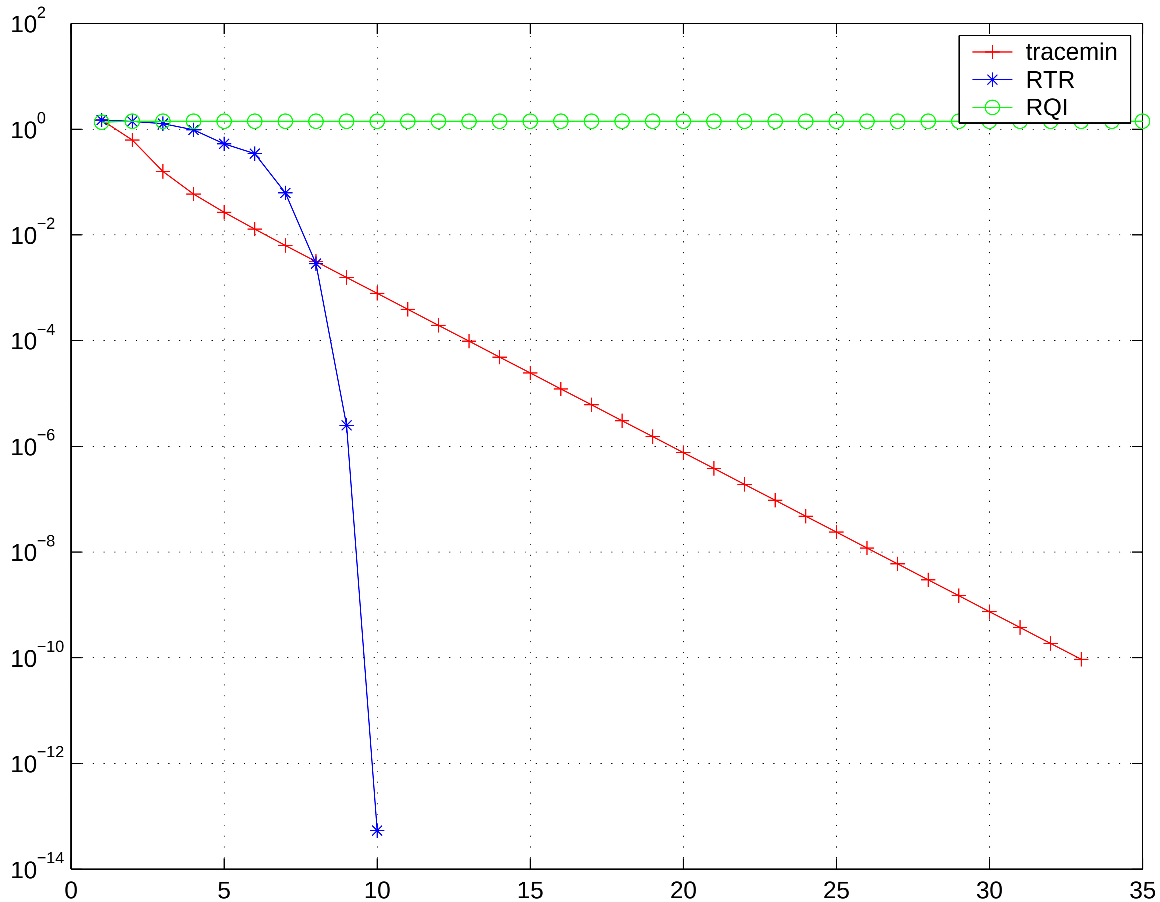
Numerical experiments



Distance to target versus number of outer iterations.

Simple symmetric positive-definite eigenvalue problem.

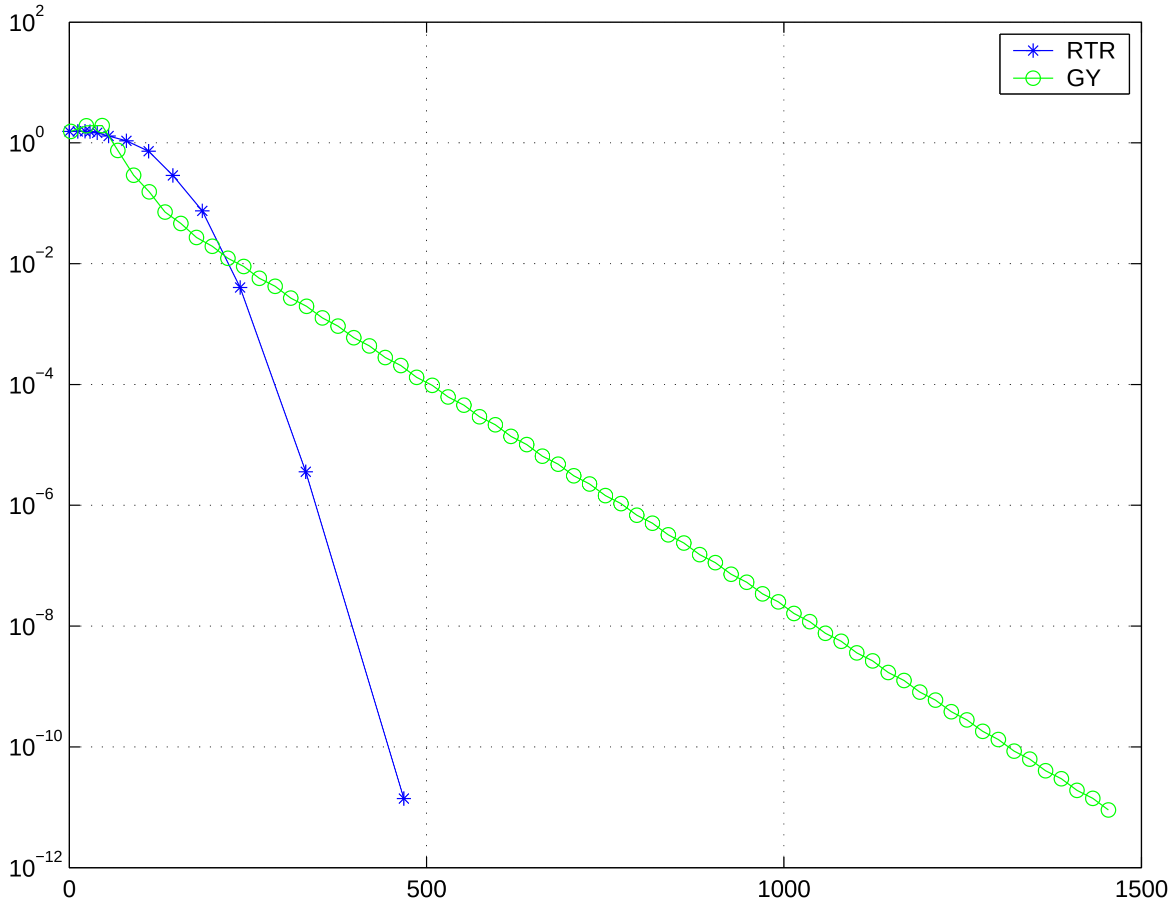
Numerical experiments



Distance to target versus number of outer iterations.

Simple symmetric positive-definite eigenvalue problem.

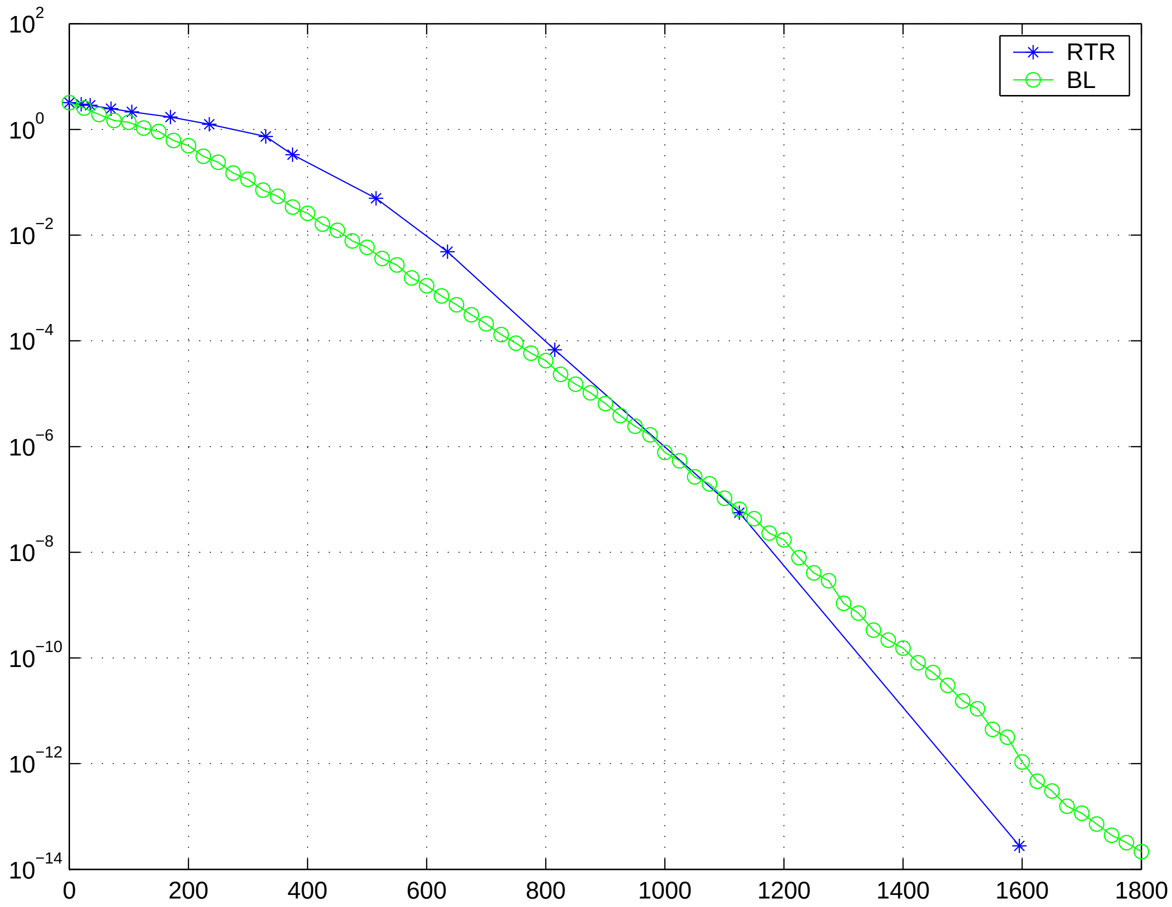
Numerical experiments



Distance to target versus matrix-vector multiplications.

Symmetric/positive-definite generalized eigenvalue problem.

Numerical experiments



Distance to target versus matrix-vector multiplications.

Block version, simple symmetric eigenvalue problem.

Conclusion (I)

Trust-region method on Riemannian manifolds.

1. Convergence to stationary points for **all** initial conditions.
2. Stable convergence to the nondegenerate local minima.
3. Superlinear local convergence to the nondegenerate local minima.
4. Approximate Hessian $\mathcal{H}$ only utilized as operator $s \mapsto \mathcal{H}s$.
5. Minimal storage space required.

Conclusion (II)

The "ideal" minor component algorithm

1. Convergence to some eigenvector for **all** initial conditions.
2. Stable convergence to the leftmost/rightmost eigenvector only.
3. Superlinear local convergence to $\pm v_1$.
4. Matrix A only utilized as operator $x \mapsto Ax$:
   - No exact system solve with matrix A.
   - No factorization of A.
5. Minimal storage space required.

Current work

Hybrid Lanczos-tCG method.

References

[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.

[Boo75] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, 1975.

[CGT00] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust-region methods*, MPS/SIAM Series on Optimization, Society for Industrial and Applied

Mathematics (SIAM), Philadelphia, PA, and Mathematical Programming Society (MPS), Philadelphia, PA, 2000.

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

[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.

[GY02] G. H. Golub and Q. Ye, *An inverse free preconditioned Krylov subspace method for symmetric generalized eigenvalue problems*, SIAM J. Sci. Comput. **24** (2002), no. 1, 312–334 (electronic).

[HM94] U. Helmke and J. B. Moore, *Optimization and dynamical systems*, Springer, 1994.

[LE00] R. Lippert and A. Edelman, *Nonlinear eigenvalue problems with orthogonality constraints (Section 9.4)*, Templates for the Solution of Algebraic Eigenvalue Problems (Zhaojun Bai, James Demmel, Jack Dongarra, Axel Ruhe, and Henk van der Vorst, eds.), SIAM, Philadelphia, 2000, pp. 290–314.

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

[MT99] M. Mongeau and M. Torki, *Computing eigenelements of real symmetric matrices via optimization*, Tech. Report MIP 99-54, Université Paul Sabatier, Toulouse, 1999, to appear in Computational Optimization and Applications.

[NMMA04] Maziar Nikpour, Jonathan H. Manton, Iven M. Y. Mareels, and Vadim Adamyan, *Algorithms for* *extreme eigenvalue problems*, Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004), Leuven, Belgium, 5-9 July 2004, 2004.

[NW99] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer Series in Operations Research, Springer-Verlag, New York, 1999.

[Sco81] David S. Scott, *Solving sparse symmetric generalized eigenvalue problems without factorization*, SIAM J. Numer. Anal. **18** (1981), no. 1, 102–110. MR 82d:65039

[Smi94] S. 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.

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

THE END