ITR Project Meeting, Jan 28, 2005

Kyle A. Gallivan

Christopher G. Baker

Pierre-Antoine Absil

School of Computational Science,

Florida State University

Collaboration with Paul Van Dooren and Ahmed Sameh

List of projects

- Symmetric Generalized Eigenvalue Problem (with A. Sameh)
- H_∞ norm computation (with P. Van Dooren)
- Methods for model reduction (with P. Van Dooren)
  - Gramian-based
  - (Tangential) interpolation
  - H_∞-based
  - Manifold-based

Trust-region methods on Riemannian manifolds for the symmetric generalized eigenproblem

Pierre-Antoine Absil

Christopher G. Baker

Kyle A. Gallivan

School of Computational Science,

Florida State University

Collaboration with Ahmed Sameh and Paul Van Dooren

`These slides and related documents are available at`

`http://www.csit.fsu.edu/∼absil/Publi/RTR.htm`

Why eigenproblem? Modal Analysis for Model Reduction

- Conceptually simple: project the system to the eigenspace corresponding to some eigenvalues.

- In structural mechanics, projection to the lower modes of vibration.
  $\rightsquigarrow$ Computation of the *leftmost* eigenpairs of stiffness-mass pencil (K, M).

- Useful as an initial step for very large sparse systems, to produce an intermediate transfer function of acceptable degree.

- Does not require a selection of input and output.

Modal Approximation of Structures (1)

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{b}u(t)$$

Assume proportional damping ($\mathbf{C}$ is a linear combination of $\mathbf{M}$ and $\mathbf{K}$). Assume that $\mathbf{M}$ is positive definite and $\mathbf{K}$ positive semi-definite. Then there exist a *modal basis* $\left(\mathbf{x}_{(s)}\right)_{s=1,\dots,n}$ such that

$$\mathbf{x}_{(r)}^T \mathbf{M} \mathbf{x}_{(s)} = \delta_{rs}, \ \mathbf{x}_{(r)}^T \mathbf{K} \mathbf{x}_{(s)} = \omega_r^2 \delta_{rs}, \ \mathbf{x}_{(r)}^T \mathbf{C} \mathbf{x}_{(s)} = 2\zeta_r \omega_r \delta_{rs},$$

and $0 \leq \omega_1 \leq \ldots \leq \omega_n$.

Modal Approximation of Structures (2)

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{b}u(t) \tag{1}$$

Decomposing the response in the modal basis,

$$\mathbf{q}(t) = \sum_{s=1}^{n} \mathbf{q}_{\mathbf{m}s}\mathbf{x}_{(s)},$$

and replacing in (1) yields the n decoupled equations

$$\ddot{\mathbf{q}}_{\mathbf{m}s} + 2\zeta_s\omega_s\dot{\mathbf{q}}_{\mathbf{m}s} + \omega_s^2\mathbf{q}_{\mathbf{m}s} = \mathbf{b}_{\mathbf{m}s}u(t), \quad s = 1, \ldots, n,$$

where

$$\mathbf{b}_{\mathbf{m}s} := \mathbf{x}_{(s)}^T\mathbf{b}.$$

Modal Approximation of Structures (3)

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{b}u(t)$$

$$\mathbf{q}(t) = \sum_{s=1}^{n} \quad \mathbf{x}_{(s)}\mathbf{x}_{(s)}^T\mathbf{b} \quad \frac{1}{\omega_{sd}} \int_0^t e^{-\zeta_s\omega_s(t-\tau)} \sin\left(\omega_{sd}(t-\tau)\right) u(\tau)d\tau, \tag{2}$$

where $\omega_{sd} = \omega_s\sqrt{1-\zeta_s^2}$.

Modal truncation consists in approximating $\mathbf{q}(t)$ by retaining only a few dominant terms in the development (2).

Modal Approximation of Structures (4)

Dominance of a mode s depends on two factors:

- A *spatial factor*

$$\mathbf{x}_{(s)}\ \mathbf{b}_{\mathbf{m}s} = \mathbf{x}_{(s)}\ \mathbf{x}_{(s)}^T \mathbf{b}$$

  that only depends on the spatial distribution $\mathbf{b}$ of the load. The factor $\mathbf{b}_{\mathbf{m}s}$is called *modal participation factor* for the considered mode; see [GR97, §2.5].

- A *temporal factor*

$$\theta_s(t) := \frac{1}{\omega_{sd}} \int_0^t e^{-\zeta_s \omega_s (t-\tau)} \sin\left(\omega_{sd}(t-\tau)\right) u(\tau) d\tau$$

  that only depends on $u(t)$.

Why model trust region? (1)

- Initial observation about single vector iterations for computing the leftmost eigenvector of a matrix $A = A^T \succ 0$:
  - Unshifted inverse iteration: global convergence, but only linear.
  - Rayleigh quotient iteration: cubic convergence, but no global convergence.
  - Hybrid method that retains the best of both??

Why model trust region? (2)

- Numerical Optimization:
  - For superlinear convergence, use Newton's method.
    At each step, compute the stationary point of the local quadratic model of the cost function.
  - For global convergence to local minima, introduce a trust-region constraint.
  - For numerical efficiency and low memory requirements, solve *approximately* the TR subproblems using truncated CG (Steihaug-Toint). Convergence properties are preserved!

- How does TR apply to the extreme symmetric generalized eigenproblem?

- Are we better off with a TR-based eigensolver?

Outline

- Extreme symmetric GEP as optimization on manifold.
- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - Numerical experiments.

The optimization problem

Given is $n \times n$ pencil (A, B), $A = A^T$, $B = B^T \succ 0$, with (unknown) eigensystem

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

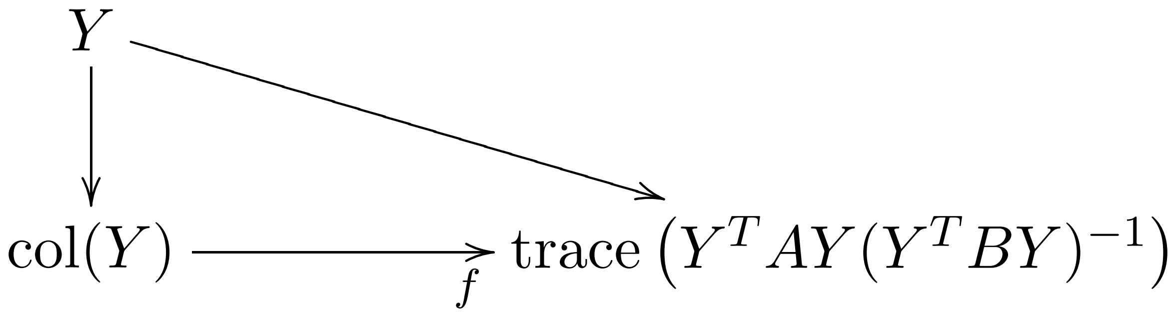
$$[v_1|\ldots|v_n]^T\,B\,[v_1|\ldots|v_n] = I, \quad \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_n.$$

The problem is to compute the “leftmost” eigenspace $\mathcal{V} := \mathrm{col}(v_1, \ldots, v_p)$.

Solution: $\mathcal{V} = \mathrm{col}(\arg\min_{Y \in \mathrm{ST}(p,n)} \mathrm{trace}\,\big(Y^T AY (Y^T BY)^{-1}\big))$.

Difficulty: continuum of minimizers Y.

Optimization problem on the Grassmann manifold



Then the leftmost p-dimensional eigenspace $\mathcal{V}$ of (A, B) satisfies

$$\mathcal{V} = \arg \min_{\mathcal{Y} \in \mathrm{Grass}(p,n)} f(\mathcal{Y})$$

where

$$f : \mathrm{Grass}(p, n) \to \mathbb{R} : \mathrm{col}(Y) \mapsto \mathrm{trace}(Y^TAY(Y^TBY)^{-1}).$$

Outline

- Extreme symmetric GEP as optimization on manifold.
- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - 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

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

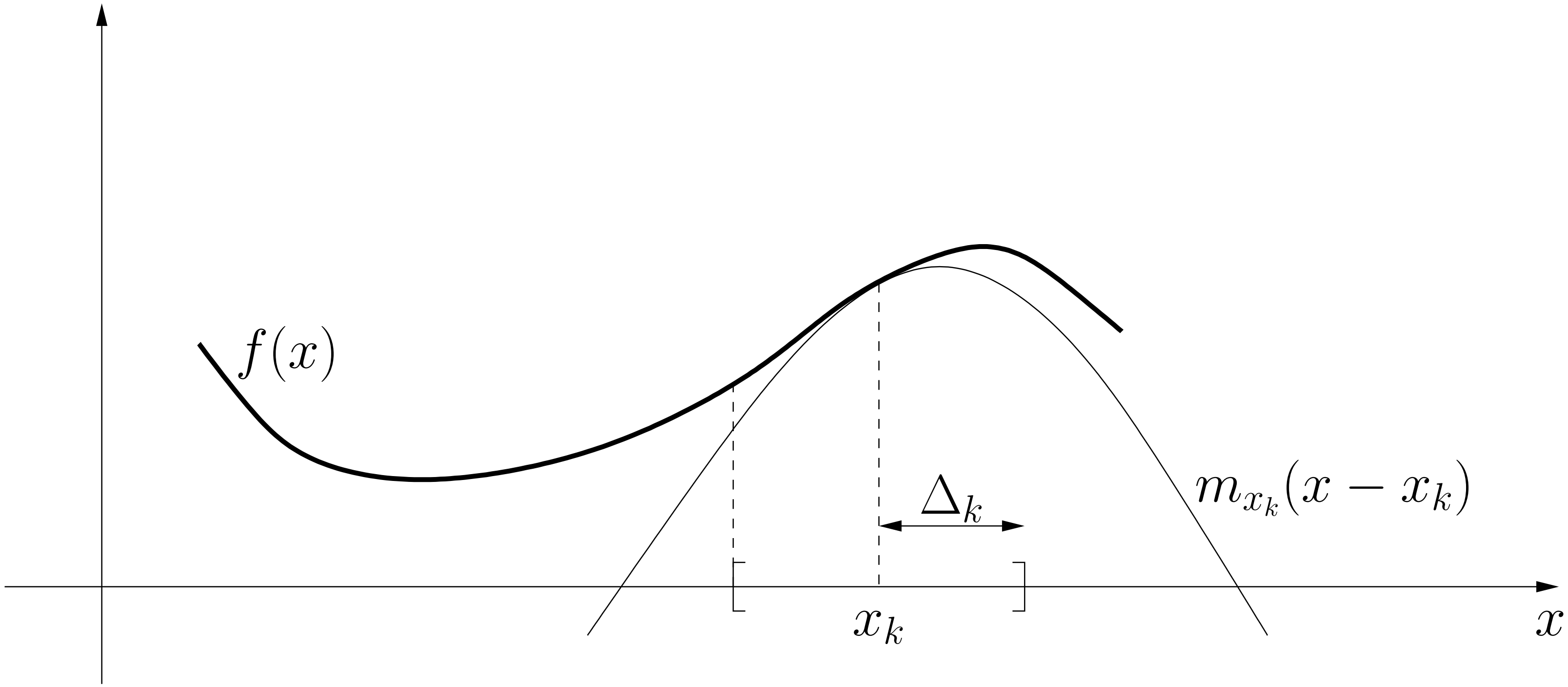
   to compare the actual value of the cost function at the proposed new iterate with the value predicted by the model.

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

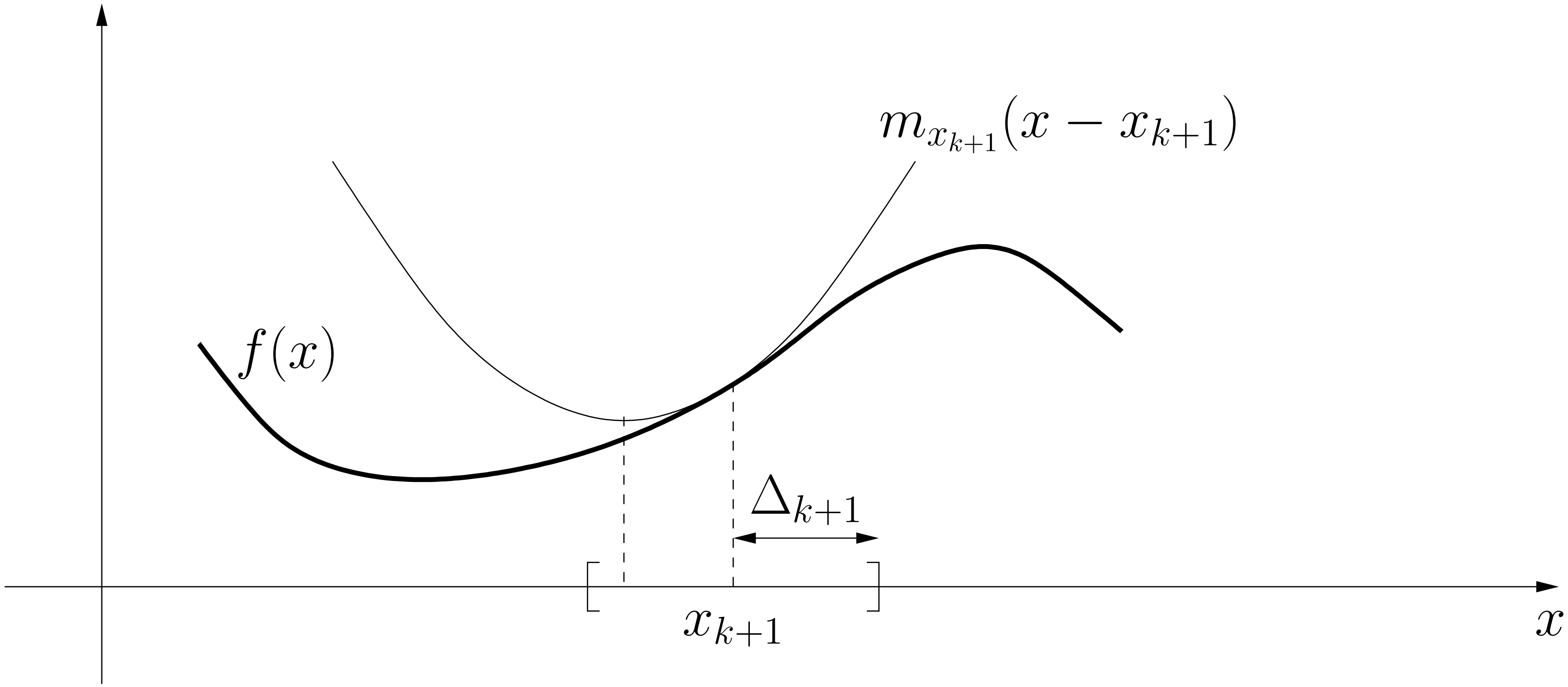
5. Shrink, enlarge or keep the trust-region radius according to the value of ρ.
6. Accept or reject the proposed new iterate $x_k + s_k$ according to the value of ρ.
7. Increment k and go to step 2.

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

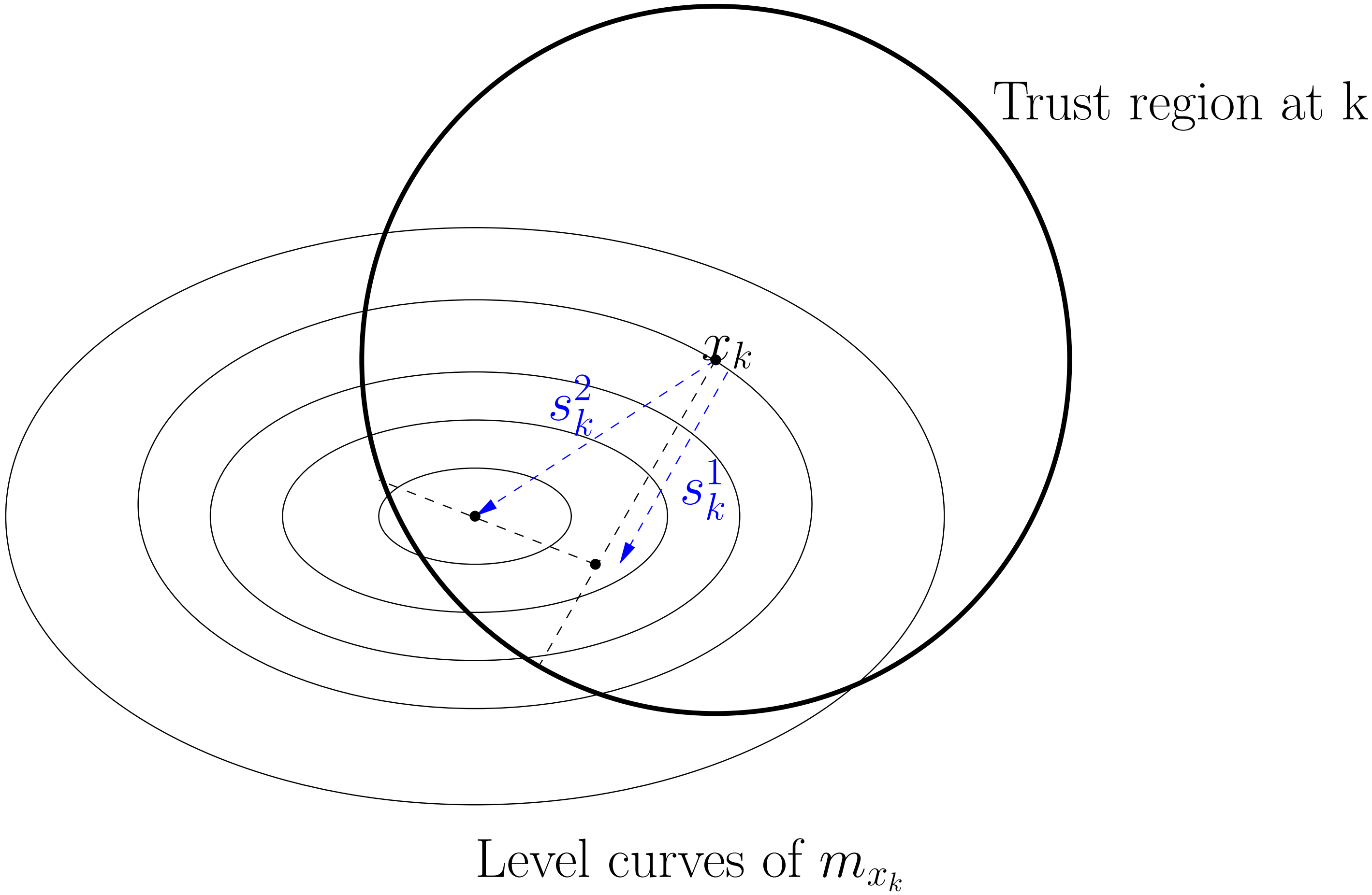
Principle of Trust-Region (TR) in $\mathbb{R}^n$



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

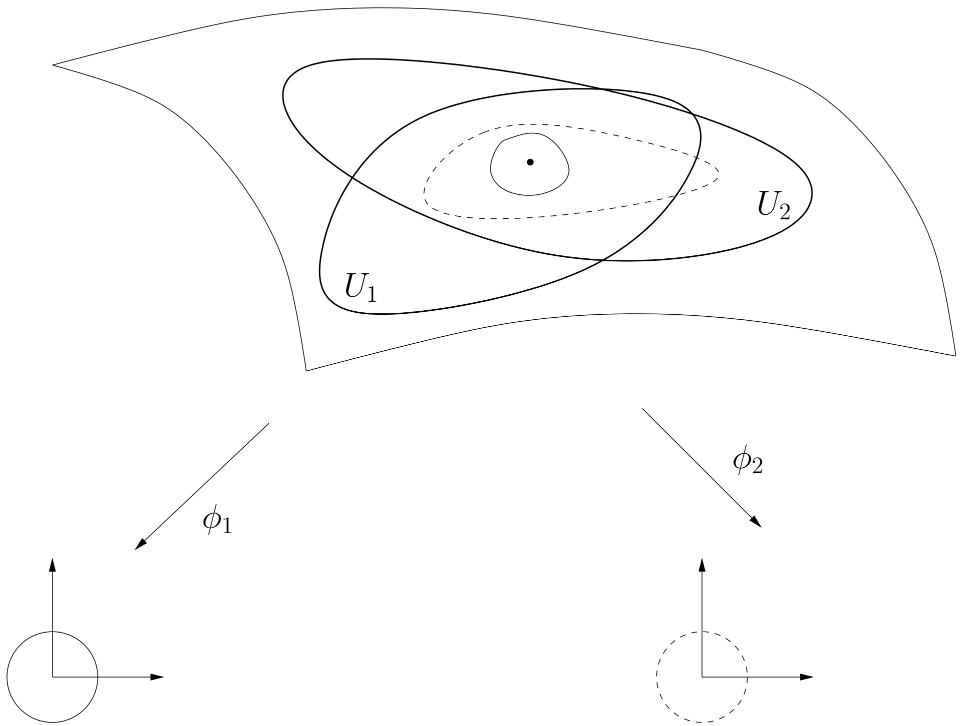
If cost fn symmetric around the limit point: $\min\{\theta + 1, 3\}$.

Outline

- Extreme symmetric GEP as optimization on manifold.
- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - 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 $\alpha \in \mathbb{R}$.



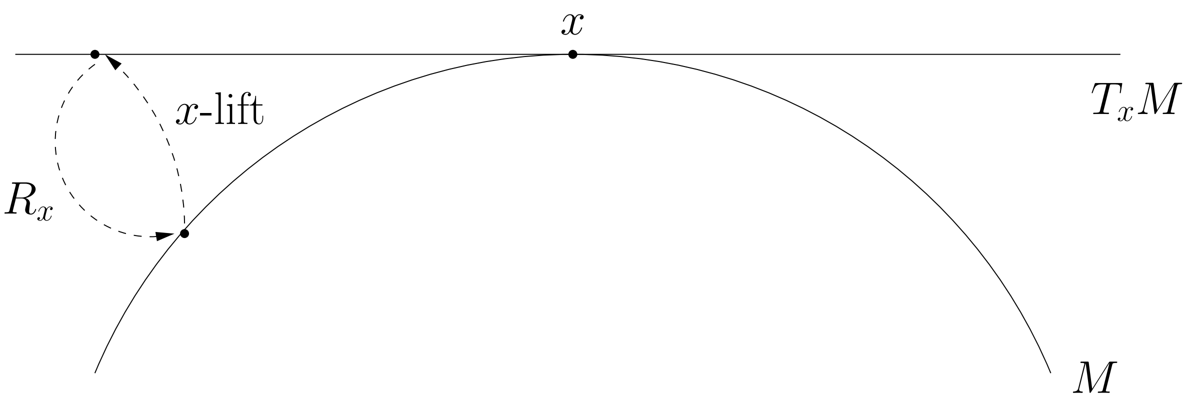
Trust-region methods on Riemannian manifolds: remedies

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)

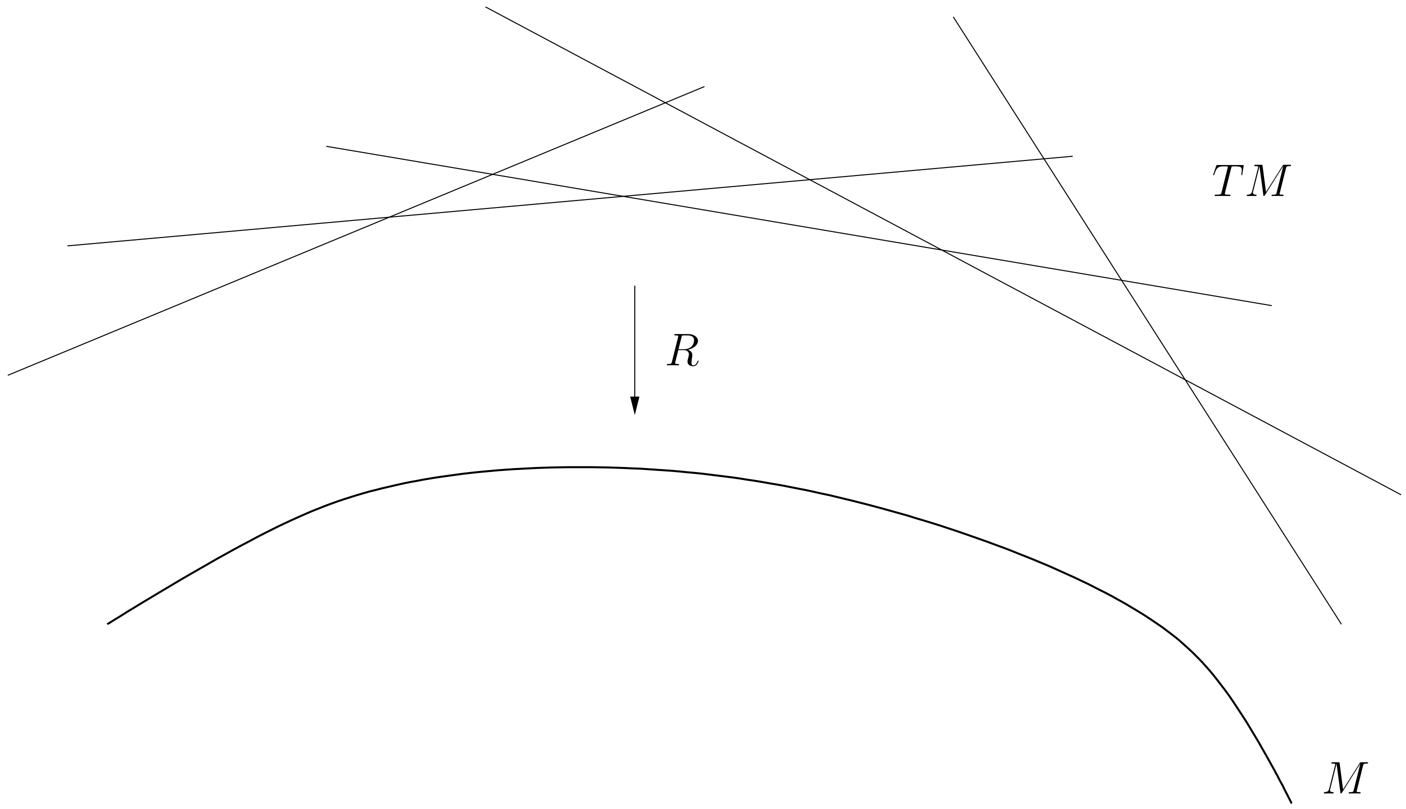
Concept of *retraction*:



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.

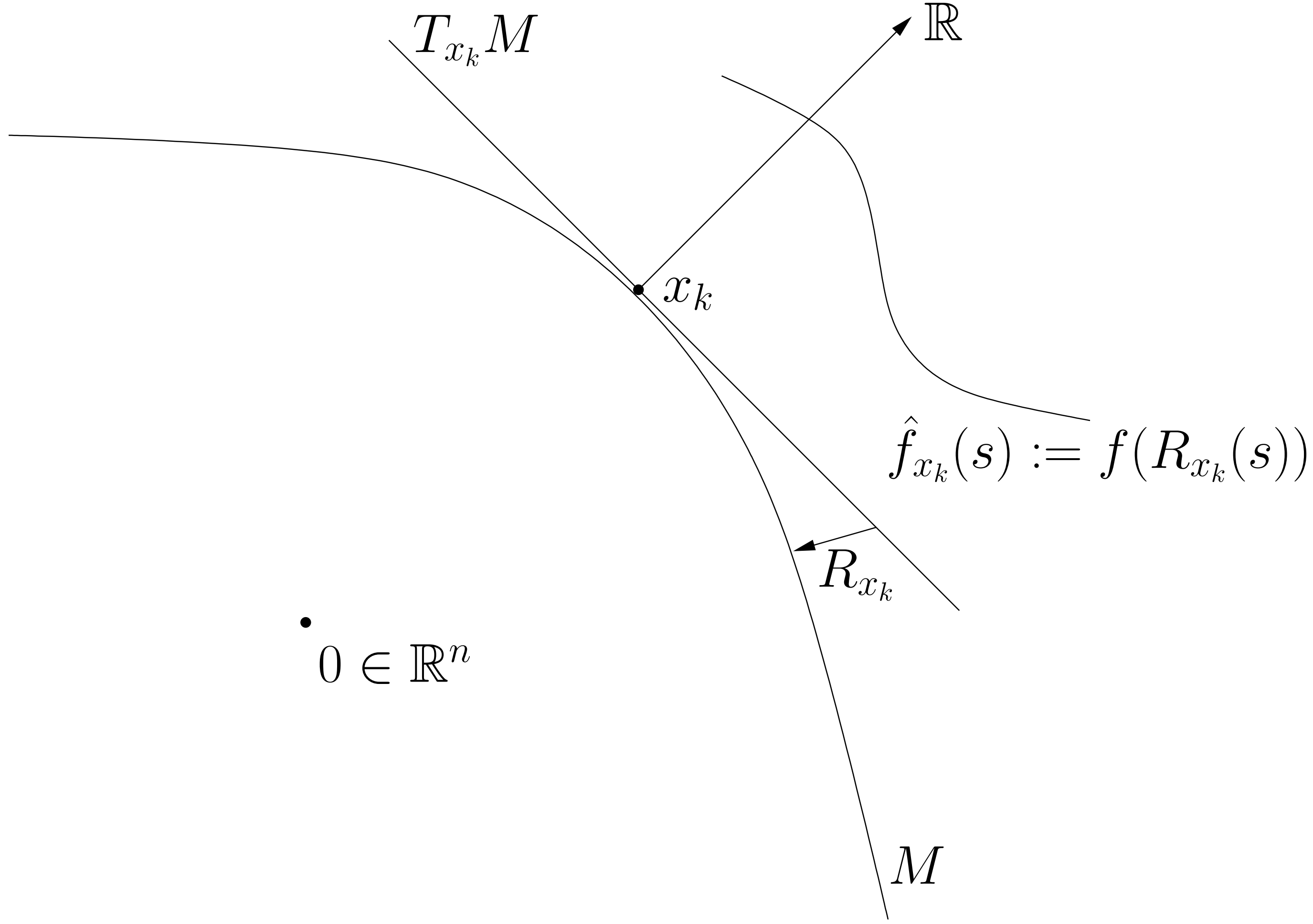
7. Increment k and go to step 2.

Solving the TR subproblem: truncated CG

- Start from the point $s^0 = 0$.
- Compute the first search direction $\delta^0 = -\text{grad}\, f(x_k)$.
- 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

- Extreme symmetric GEP as optimization on manifold.
- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - 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 $\|\mathrm{grad}\, f(x)\| \leq \beta_g$ and $\|\mathrm{Hess}\, f(x)\| \leq \beta_H$ for some constants β_g, β_H, and all $x \in M$. Moreover suppose that

$$\|\tfrac{D}{dt}\tfrac{d}{dt}Rt\xi\| \leq \beta_D \qquad (3)$$

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

$$\|\mathcal{H}_k - \operatorname{Hess} \hat{f}_{x_k}(0_k)\| \leq \beta_{\mathcal{H}} \|\operatorname{grad} f(x_k)\|. \tag{4}$$

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

$$\|\operatorname{Hess} \hat{f}_{x_k}(\xi) - \operatorname{Hess} \hat{f}_{x_k}(0_{x_k})\| \leq \beta_{L2} \|\xi\|. \tag{5}$$

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 $k > K$,

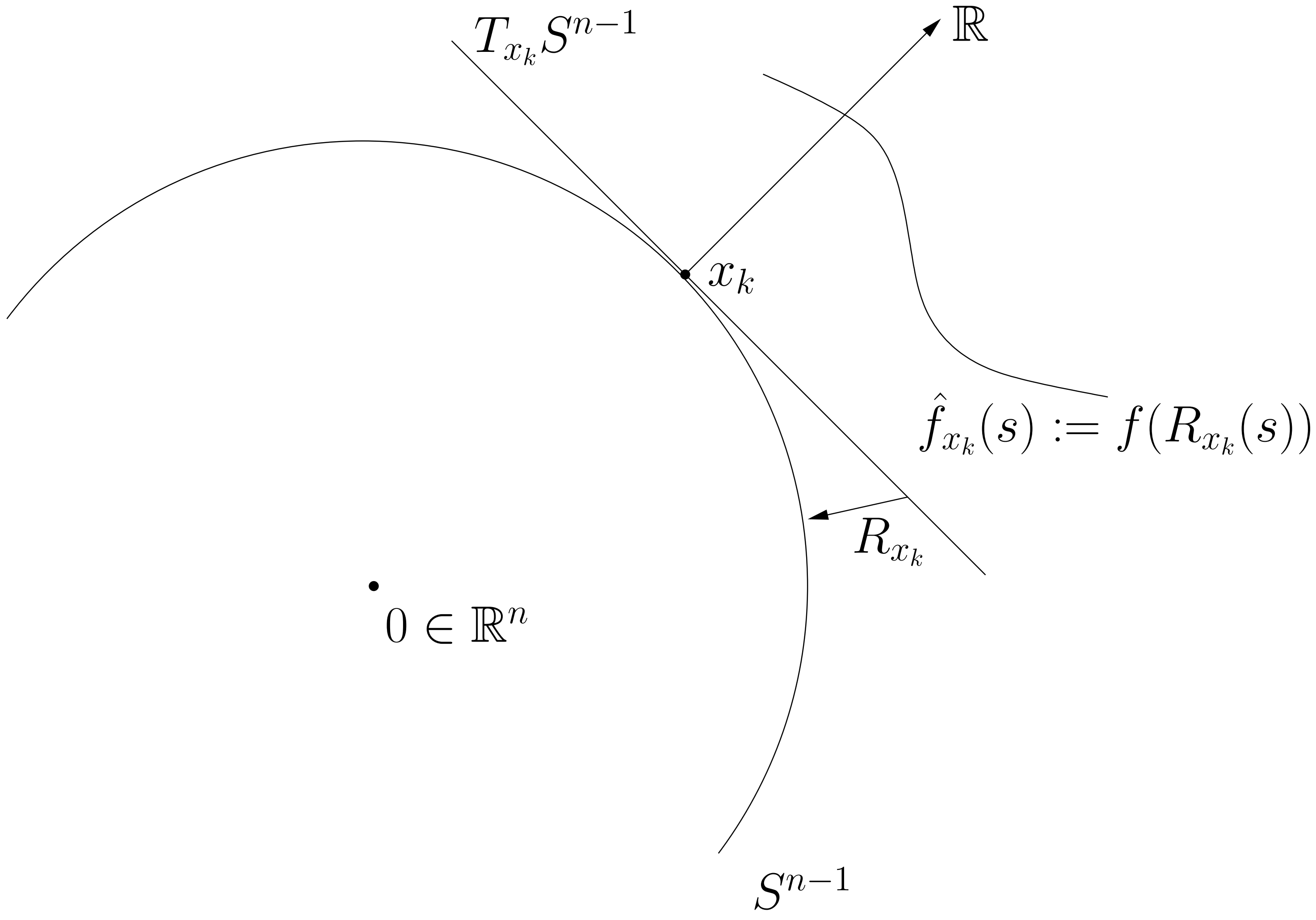
$$\operatorname{dist}(x_{k+1}, v) \leq c\,(\operatorname{dist}(x_k, v))^{\min\{\theta+1,2\}}, \tag{6}$$

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

Outline

- Extreme symmetric GEP as optimization on manifold.
- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - Numerical experiments.

Case $p = 1$: Trust-region on the sphere



Trust-region for extreme SGEVP: principles

Given: $n \times n$ symmetric matrices A and B, with $B \succeq 0$.
Problem: compute the 'leftmost' eigenvector v_1 of pencil (A, B).

Ingredients of the Riemannian trust-region method:

1. Manifold: $M = \{y \in \mathbb{R}^n : y^T B y = 1\} = \{y : \|y\|_B = 1\}$.
2. Tangent space: $T_y M = \{z : y^T B z = 0\}$.
3. Metric: $g_y(z_a, z_b) = z_a^T z_b$.
4. Retraction: $R_y(z) = (y + z)/\|y + z\|_B$.
5. Cost function: $f : \{y : \|y\|_B = 1\} \to \mathbb{R} : y \mapsto \frac{y^T A y}{y^T B y}$.

Underlying fact: $v_1 = \arg\min f(y)$.

Trust-region for extreme SGEVP: details

Lifted cost function:

$$\hat{f}_y(s) = f(R_y(s)) = f\left(\frac{y+s}{\|y+s\|_B}\right) = \frac{(y+s)^T A(y+s)}{(y+s)^T B(y+s)}, \quad y^T B s = 0.$$

Let $\langle u, v\rangle = u^T v$ denote the classical inner product on $\mathbb{R}^n$, and let P denote the orthogonal projector onto $\{s : y^T B s = 0\}$, that is

$$P = I - By(y^T B^2 y)^{-1} y^T B. \tag{7}$$

Trust-region for extreme SGEVP: details

One has:

$$
\begin{aligned}
\hat{f}_y(s) &= \frac{y^T A y}{y^T B y} + 2\frac{y^T A s}{y^T B y} \\
&\quad + \frac{1}{y^T B y}\left(s^T A s - \frac{y^T A y}{y^T B y} s^T B s \right) + O\left(\|s\|^3\right) \\
&= f(y) + 2\langle PAy, s\rangle \\
&\quad + \frac{1}{2}\langle 2P(A - f(y)B)Ps, s\rangle + O\left(\|s\|^3\right).
\end{aligned}
$$

Trust-region for extreme SGEVP: details

The second order approximation of $\hat{f}_y(s)$ is thus

$$m_y(s) = f(y) + 2\langle PAy, s\rangle + \frac{1}{2}\langle P(A - f(y)B)Ps, s\rangle, \quad y^T Bs = 0. \tag{8}$$

Exact trust-region method: compute
$s^* = \arg\min_{g_y(s,s)\leq\Delta^2} m_y(s) \qquad (y^T Bs = 0).$

Inexact trust-region: compute an approximate solution $\tilde{s}$ using truncated CG.

Update: $y_+ = R_y(\tilde{s}) = (y + \tilde{s})/\|y + \tilde{s}\|_B$.

Trust-region for BLOCK extreme SGEVP: principles

Given: $n \times n$ symmetric matrices A and B, with $B \succeq 0$.
Problem: compute the 'leftmost' eigenvectors $v_1, \ldots, v_p$ of pencil (A, B).

Ingredients of the Riemannian trust-region method:

1. Manifold: $M = \{p - \text{dimensional subspaces of } \mathbb{R}^n\}$ (Grassmann manifold).

2. Representations: $\mathcal{Y}$ represented by any $Y \in \mathbb{R}^{n \times p} : \text{col}(Y) = \mathcal{Y}$.

3. Tangent space: formally, $T_Y M = \{Z \in \mathbb{R}^{n \times p} : Y^T B Z = 0\}$.

4. Metric: formally, $g_Y(Z_a, Z_b) = \text{trace}\left((Y^T B Y)^{-1} Z_a^T Z_b\right)$.

5. Retraction: formally, $R_Y(Z) = (Y + Z)M$, where arbitrary M serves for normalization.

6. Cost function: formally,
   $f(Y) = \operatorname{trace}\left((Y^T BY)^{-1}(Y^T AY)\right)$.

Underlying fact: $[v_1|\ldots|v_p]M$ minimizes $f(Y)$ for all M invertible.

Trust-region for BLOCK extreme SGEVP: details

Lifted cost function:

$$
\begin{aligned}
\hat{f}_Y(Z) = f(R_Y(Z)) &= \operatorname{trace}\left(\left((Y+Z)^T B(Y+Z)\right)^{-1}\left((Y+Z)^T A(Y+Z)\right)\right) \\
&= \operatorname{trace}\left((Y^T BY)^{-1} Y^T AY\right) + 2\operatorname{trace}\left((Y^T BY)^{-1} Z^T AY\right) \\
&\quad + \operatorname{trace}\left((Y^T BY)^{-1} Z^T \left(AZ - BZ(Y^T AY)\right)\right) + HOT \\
&= \operatorname{trace}\left((Y^T BY)^{-1} Y^T AY\right) + 2\operatorname{trace}\left((Y^T BY)^{-1} Z^T P_{BY,BY} AY\right) \\
&\quad + \operatorname{trace}\left((Y^T BY)^{-1} Z^T P_{BY,BY} \left(AZ - BZ(Y^T AY)\right)\right) + HOT,
\end{aligned}
$$

where $P_{BY,BY} = I - BY(Y^T B^2 Y)^{-1} Y^T B.$

Trust-region for BLOCK extreme SGEVP: details

The second order approximation of $\hat{f}_Y(Z)$ is thus

$$
\begin{aligned}
m_Y(Z) &= f(Y) + g_Y(\operatorname{grad} f(Y), Z) + \frac{1}{2} g_Y(\mathcal{H}_Y Z, Z) \\
&= \operatorname{trace}\left((Y^T BY)^{-1} Y^T AY\right) + 2\operatorname{trace}\left((Y^T BY)^{-1} Z^T AY\right) \\
&\quad + \operatorname{trace}\left((Y^T BY)^{-1} Z^T \left(AZ - BZ(Y^T BY)^{-1} Y^T AY\right)\right).
\end{aligned}
$$

Exact trust-region method: compute
$Z^* = \arg\min_{g_Y(Z,Z) \leq \Delta^2} m_Y(Z) \qquad (Y^T BZ = 0).$

Inexact trust-region: compute an approximate solution $\tilde{Z}$ using truncated CG.

Update: $Y_+ = R_Y(\tilde{Z}) = (Y + \tilde{Z})M.$

Properties of the algorithm

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

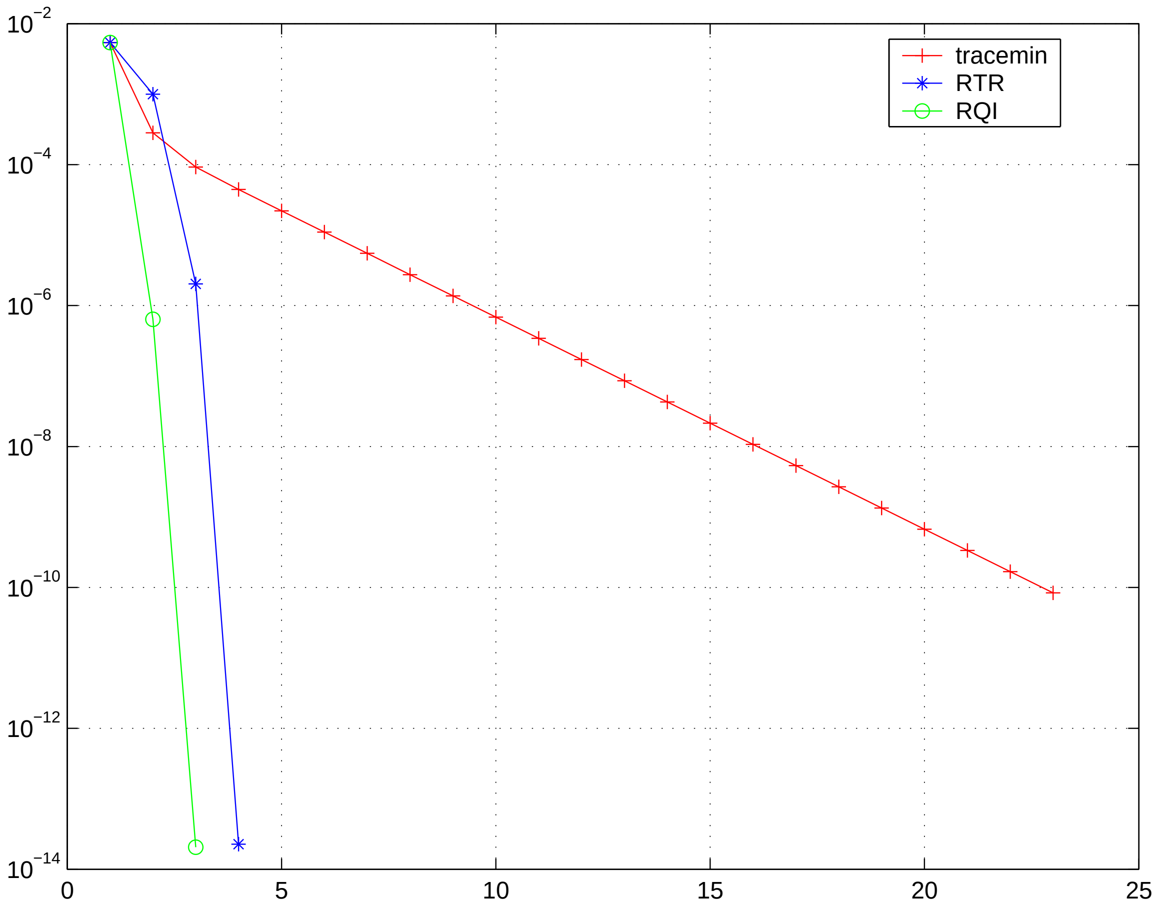
Properties:

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

- Extreme symmetric GEP as optimization on manifold.
- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - Numerical experiments.

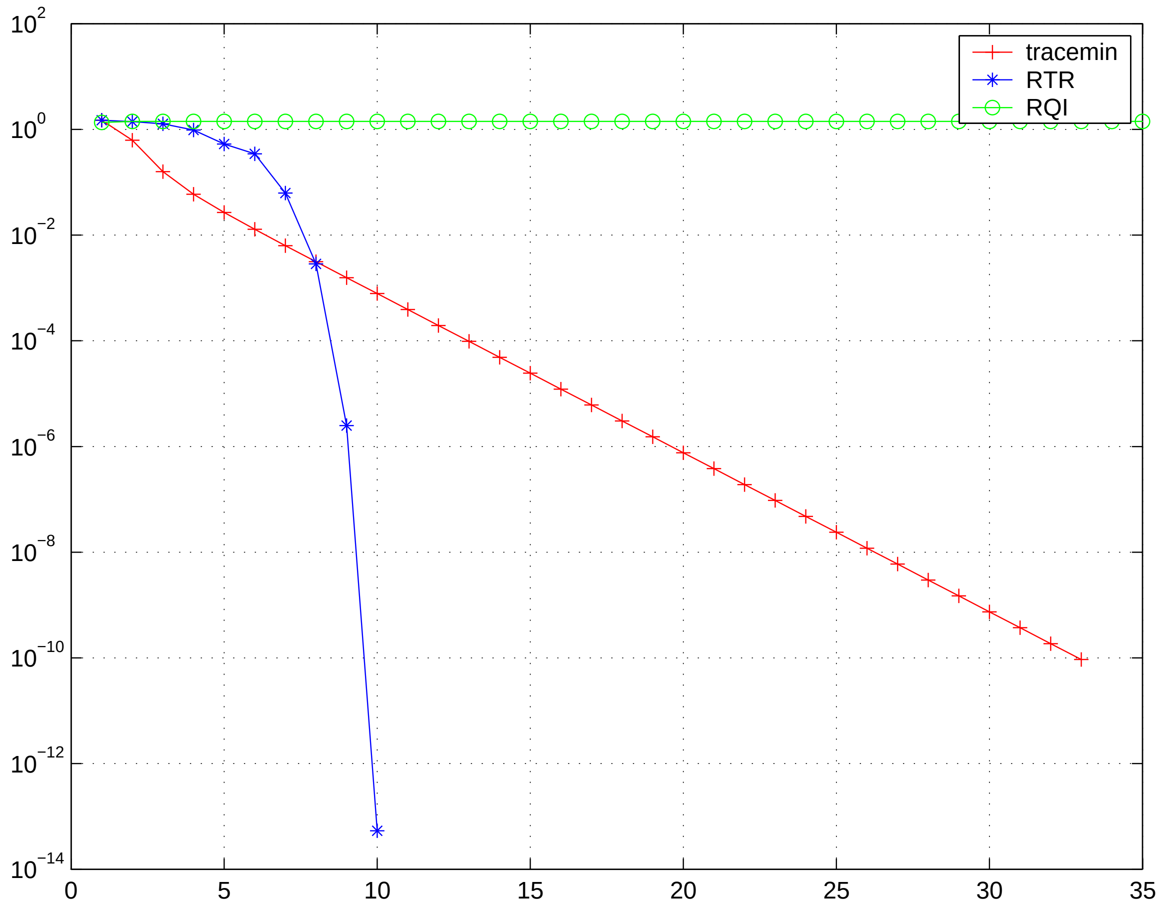
Numerical experiments: exact simple tracemin, RQI, RTR



Distance to target versus number of outer iterations.

Simple symmetric positive-definite eigenvalue problem.

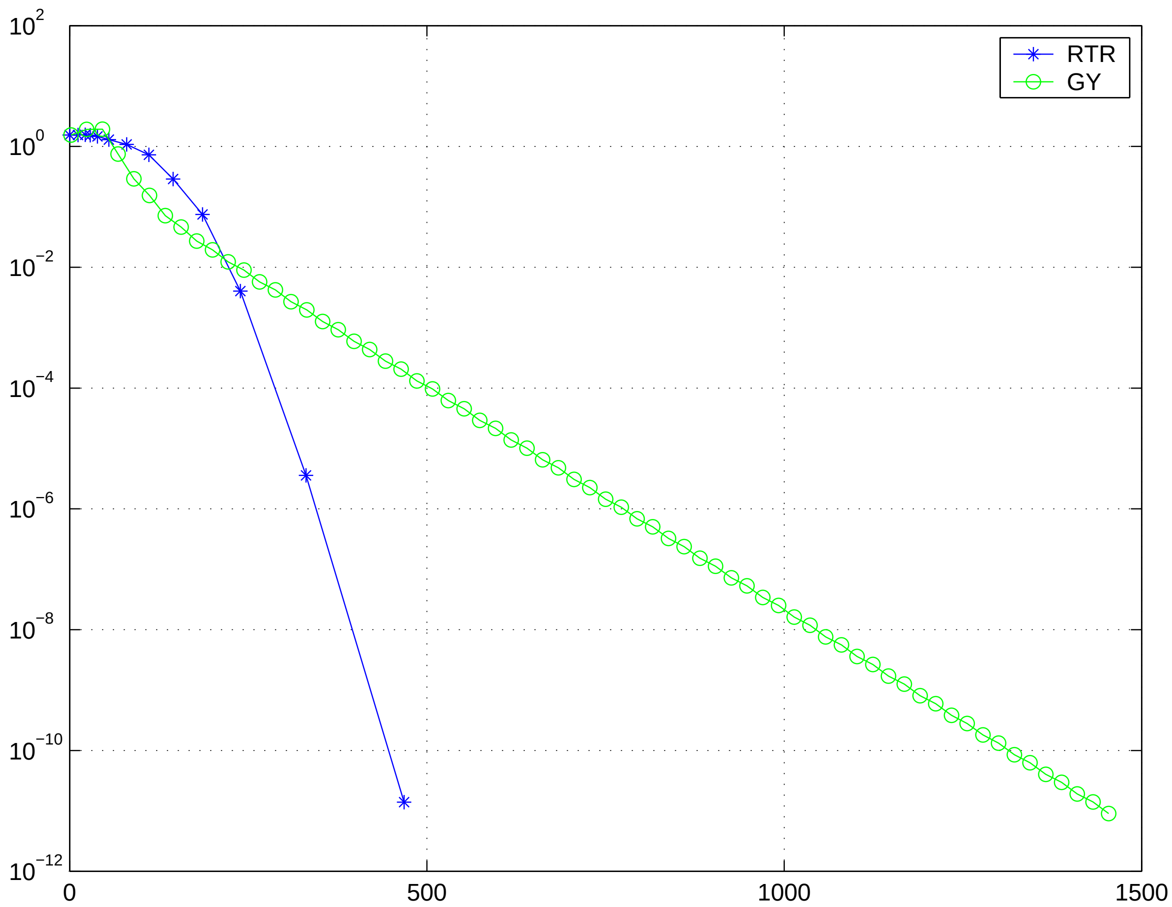
Numerical experiments: exact simple tracemin, RQI, RTR



Distance to target versus number of outer iterations.

Simple symmetric positive-definite eigenvalue problem.

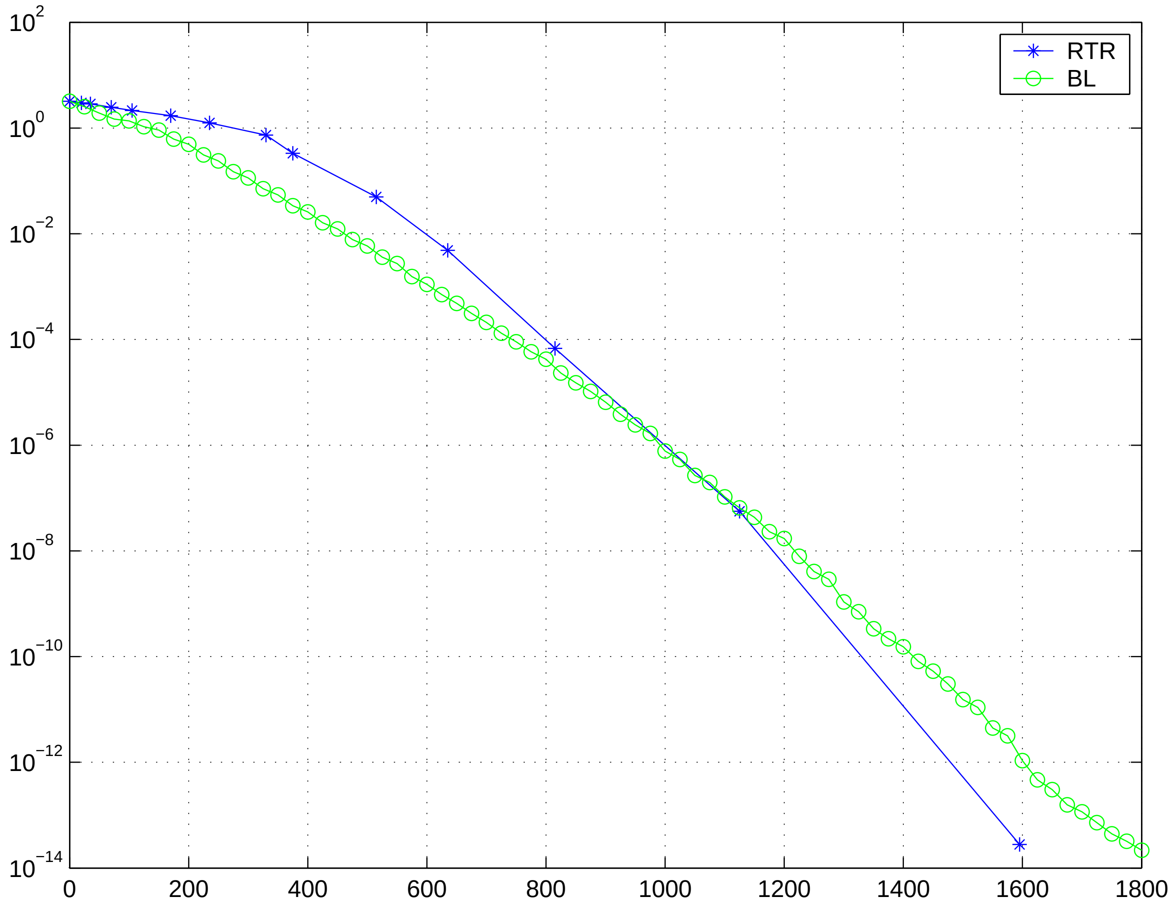
Numerical experiments: RTR vs Krylov [GY02]



Distance to target versus matrix-vector multiplications.

Symmetric/positive-definite generalized eigenvalue problem.

Numerical experiments: RTR vs Lanczos ($p > 1$)



Distance to target versus matrix-vector multiplications.

Block version, standard symmetric eigenvalue problem.

Outline

- Extreme symmetric GEP as optimization on manifold.
- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - Numerical experiments.

Link with Basic TraceMin

Basic Tracemin computes (assuming that $A \succ 0$, too)

$$Z^* = \arg\min \operatorname{trace}(Y+Z)^T A(Y+Z), \quad Y^T B Z = 0.$$

Notice that

$$\begin{aligned} & \operatorname{trace}(Y+Z)^T A(Y+Z) \\ = \; & \operatorname{trace}\left((Y^T B Y)^{-1} Y^T A Y\right) + 2\operatorname{trace}\left((Y^T B Y)^{-1} Z^T A Y\right) \\ & + \operatorname{trace}\left((Y^T B Y)^{-1} Z^T A Z\right). \end{aligned}$$

Useful property: with $Y_+ := (Y+Z)M$, one has

$$\operatorname{trace}\left((Y_+^T B Y_+)^{-1} Y_+^T A Y_+\right) \le \operatorname{trace}\left((Y^T B Y)^{-1} Y^T A Y\right).$$

But superlinear convergence is lost $\rightarrow$ dynamic shift strategy.

Link with "pure" Newton method

Remove the trust-region aspect and define the next iterate as

$$Y_+ = (Y + Z^\star)M$$

where $Z^\star$ solves the Newton equation

$$\mathrm{D}m_Y(Z^\star) = 0,$$

that is

$$P_{BY,BY}\left(AZ - BZ(Y^T AY)\right) = -P_{BY,BY}AY.$$

In the JD framework, this is called the *Jacobi equation*. Actually, it is just a (Grassmann-)Newton equation; see Edelman et al. [EAS98].

Global convergence to minor eigenspace is lost.

Subspace acceleration

Much like the pure Newton method and the Tracemin algorithm, the RTR-tCG approach lends itself to Davidson subspace acceleration enhancement. The subspace is appended with the RTR-tCG update vector $\tilde{Z}$.

Numerical experiments in progress.

Towards unification

The above-metioned (inexact-)Newton-like methods differ along the following lines:

1. Choice of the local model m ($\rightsquigarrow$ choice of shifts: Rayleigh shifts, no shifts...).
2. Stopping criterion for inner iteration.
3. Subspace acceleration enhancements.
4. Preconditioning.

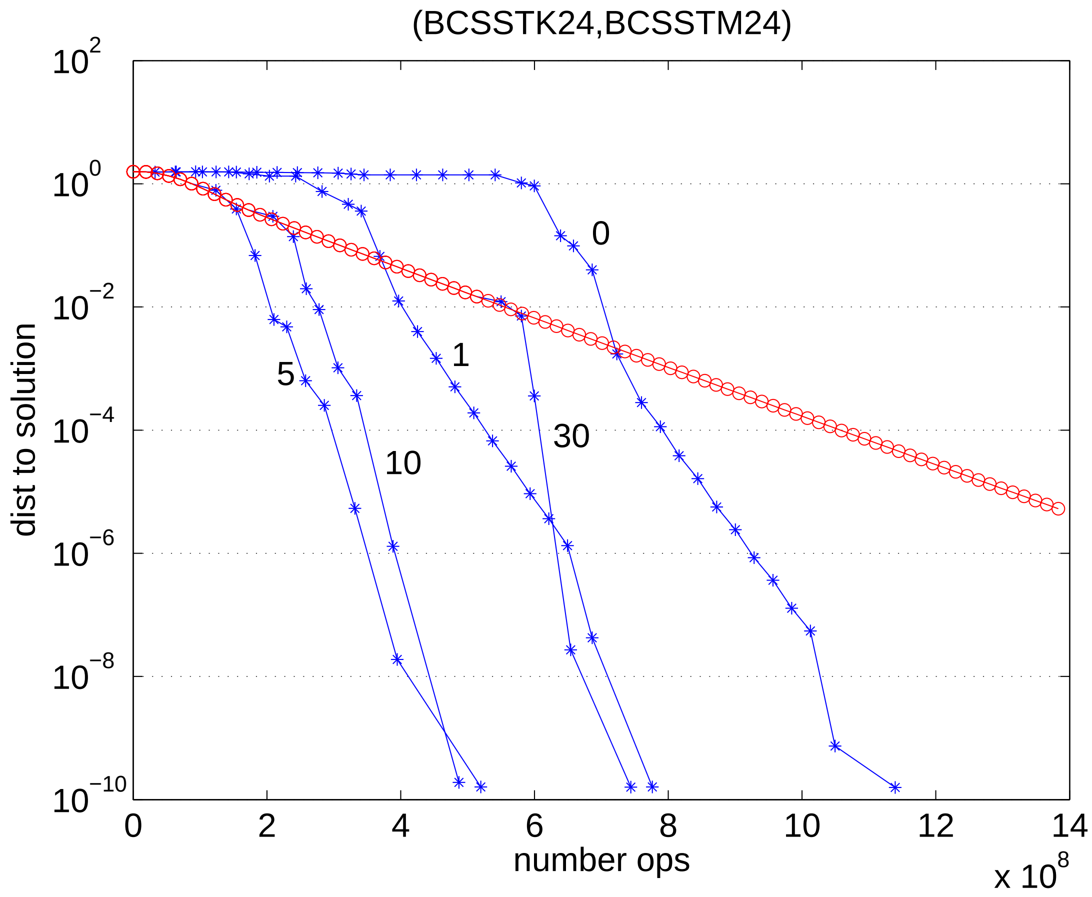
A hybrid Tracemin / TR method

Collaboration with Ahmed Sameh.

- Trust-region confinement may hamper efficient preconditioning far away from the solution.
  $\leadsto$ Use preconditioned Basic Tracemin in Phase I.

- Close to the solution, Basin Tracemin is linear.
  $\leadsto$ Use TR method in Phase II.

Hybrid Tracemin / TR



With exact preconditioner after `symamd`.

Are we better off with TR-based schemes?

- "Matrix-free" method: shift-and-inverse not needed.
- Superlinear convergence.
- Detailed global and local convergence analysis.
- Subspace acceleration enhancements; adaptive local models; preconditioning.

Future work

- Eigenvalue problem:
  - Metacode for the extreme symmetric GEP: TR method with adaptive local models and various subspace acceleration enhancements.
  - Case B positive semi-definite.
  - Compute interior eigenvalues.
  - Quadratic eigenvalue problem.
  - Nonsymmetric eigenvalue problem.
- Optimization-on-manifolds approach to model reduction (with Paul Van Dooren).

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.

[AL03] Peter Arbenz and Richard B. Lehoucq, *A comparison of algorithms for modal analysis in the absence of sparse direct methods*, Technical report SAND2003-1028J, submitted to Internat. J. Numer. Methods Engrg., 2003.

[BF66] W. W. Bradbury and R. Fletcher, *New iterative methods for solution of the eigenproblem*, Numer. Math. **9** (1966), 259–267.

[BGP97] Luca Bergamaschi, Giuseppe Gambolati, and Giorgio Pini, *Asymptotic convergence of conjugate gradient methods for the partial symmetric eigenproblem*, Numer. Linear Algebra

Appl. **4** (1997), no. 2, 69–84. MR 98b:65040

[CD74a] J. Cullum and W. E. Donath, *A block generalization of the symmetric s-step Lanczos algorithm*, Tech. Report RC 4845 (21570), IBM Thomas J. Watson Research Center, Yorktown Heights, New York, May 14 1974.

[CD74b] ______, *A block Lanczos algorithm for computing the q algebraically larges eigenvalues and a corresponding eigenspace of large, sparse, real symmetric matrices*, proceedings of the 1974 IEEE Conference on Decision and Control, Phoenix, Arizona, 1974, pp. 505–509.

[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.

[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.

[FSvdV98] Diederik R. Fokkema, Gerard L. G. Sleijpen, and Henk A. van der Vorst, *Jacobi-Davidson style QR and QZ algorithms for the reduction of matrix pencils*, SIAM J. Sci. Comput. **20** (1998), no. 1, 94–125 (electronic). MR 99e:65061

[GR97] M. Géradin and D. Rixen, *Mechanical vibrations : Theory and applications to structural dynamics*, John Wiley & Son Ltd, Chichester, 1997.

[GSF92] Giuseppe Gambolati, Flavio Sartoretto, and Paolo Florian, *An orthogonal accelerated deflation technique for large symmetric eigenproblems*, Comput. Methods Appl. Mech. Engrg. **94** (1992), no. 1, 13–23. MR 92i:65072

[GU77] G. H. Golub and R. Underwood, *The block Lanczos method for computing eigenvalues*, Mathematical software, III (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1977), Academic Press, New York, 1977, pp. 361–377. Publ. Math. Res. Center, No. 39.

[GY00] Gene H. Golub and Qiang Ye, *Inexact inverse iteration for generalized eigenvalue problems*, BIT **40** (2000), no. 4, 671–684. MR 2001j:65063

[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).

[KN03] Andrew V. Knyazev and Klaus Neymeyr, *A geometric theory for preconditioned inverse iteration. III. A short and sharp convergence estimate for generalized eigenvalue problems*, Linear Algebra Appl. **358** (2003), 95–114, Special issue on accurate solution of eigenvalue problems (Hagen, 2000). MR 2004c:65040

[Kny01] A. V. Knyazev, *Toward the optimal preconditioned* *eigensolver: locally optimal block preconditioned conjugate gradient method*, SIAM J. Sci. Comput. **23** (2001), no. 2, 517–541.

[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.

[LM80] D. E. Longsine and S. F. McCormick, *Simultaneous Rayleigh-quotient minimization methods for $Ax = \lambda Bx$*, Linear Algebra Appl. **34** (1980), 195–234.

[Not02] Y. Notay, *Combination of Jacobi-Davidson and conjugate gradients for the partial symmetric eigenproblem*, Numer. Linear Algebra Appl. **9** (2002), no. 1, 21–44.

[Not03] ———, *Convergence analysis of inexact Rayleigh quotient iteration*, SIAM J. Matrix Anal. Appl. **24** (2003), no. 1, 627–644.

[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

[SE02] Valeria Simoncini and Lars Eldén, *Inexact Rayleigh quotient-type methods for eigenvalue computations*, BIT **42** (2002), no. 1, 159–182. MR 2003d:65033

[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. MR MR1297990 (95g:58062)

[SP99] P. Smit and M. H. C. Paardekooper, *The effects of inexact solvers in algorithms for symmetric eigenvalue problems*, Linear Algebra Appl. **287** (1999), no. 1-3, 337–357, Special issue celebrating the 60th birthday of Ludwig Elsner. MR 2000a:65049

[ST00] A. Sameh and Z. Tong, *The trace minimization method for the symmetric generalized eigenvalue problem*, J. Comput. Appl. Math. **123** (2000), 155–175.

[SvdVM98] G. L. Sleijpen, H. A. van der Vorst, and E. Meijerink, *Efficient expansion of subspaces in the jacobi-davidson method for standard and generalized eigenproblems*, Electron. Trans. Numer. Anal. **7** (1998), 75–89.

[SW82] A. H. Sameh and J. A. Wisniewski, *A trace minimization algorithm for the generalized eigenvalue problem*, SIAM J. Numer. Anal. **19** (1982), no. 6, 1243–1259.

Future work

THE END

Review of Newton-like methods for extreme EVP

- Pure Newton: Ritz shifts, exact solve, preconditioning irrelevant.
- RTR-tCG: Ritz shifts, tCG inner stopping criterion.
- Dynamic tracemin: Ritz shifts pushed to the left, dynamic inner stopping criterion.
- JD: various shifts (usually Ritz values), various inner stopping criteria (usually a fixed number of inner iterations), Davidson acceleration.
- Lanczos (?): shifts irrelevant, only one step of inner solve (i.e., use RHS), subspace acceleration.

Tentative classification of methods for extreme EVP

The following classification is inspired from Arbenz and Lehoucq [AL03].

1. Inexact-Newton-based methods (optimize successive models of the Rayleigh quotient).
2. Nonlinear-CG-based methods for optimizing the Rayleigh quotient.
3. Lanczos-based methods (build Krylov subspaces and restart with best approximation from the subspace).

Apparently, most methods clearly fall within one category.

Classification: Newton methods

- 'Pure' Newton method on manifolds for the Rayleigh quotient: Smith [Smi94], Edelman, Arias and Smith [EAS98], Lundström and Eldén [LE02].
- Dynamic Tracemin of Sameh, Wisniewski and Tong [SW82, ST00]: Newton method with "shifted Ritz shifts".
- Jacobi-Davidson of Fokkema, Sleijpen, van der Vorst: see, e.g., [FSvdV98, SvdVM98].
- Vast and recent literature on inexact Newton and inverse iteration: [SP99, GY00, SE02, Not03, KN03]...
- Notay [Not02]: Newton, CG inner iteration, Davidson acceleration.

Classification: nonlinear CG

- Early work of Bradbury and Fletcher [BF66].
- Longsine and McCormick [LM80].
- Deflation-accelerated (nonlinear) CG (DACG) of Ganbolati, Pini and collaborators [GSF92, BGP97].
- Knyazev's Locally Optimal Block Preconditioned (nonlinear) CG (LOBPCG) [Kny01].

Classification: Lanczos methods

- Cullum and Donath [CD74a, CD74b], Golub and Underwood [GU77]: block Lanczos algorithms for the standard EVP.
- 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).
- ... (many other references)

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 and challenges

- Hybrid, “cross-classification” methods: Newton, nonlinear CG, Krylov.
- Go for interior eigenvalues.
- Nonsymmetric eigenvalue problem.
- Quadratic eigenvalue problem.

THE END