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

Slides compiled on December 15, 2004

These slides and related documents are available at
http://www.csit.fsu.edu/~absil/Publi/RTR.htm

Numerical Linear Algebra (NLA) problems

$\equiv$

Optimization on Manifold problems.

Problem 1 : Minor Eigenvector Computation

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

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

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

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

Problem 1 as 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 $\pm v_1$.

Problem 1 as optim problem on the 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 $\mathrm{span}(v_1|\ldots|v_p)$.

Problem 2 as optimization problem on Grassmann

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 the *Grassmann manifold* $\mathrm{Grass}(p,n)$ 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 $\mathrm{span}(v_1|\ldots|v_p)$.

Problem 3 as optimization problem on Grassmann

Rayleigh quotient cost function:

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

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 $\mathrm{span}(v_1|\dots|v_p)$.

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.
- Extreme partial EVD, SVD.
- Balanced factorization.
- Low rank approximation.
- Joint approximate diagonalization.
- ...

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

Manifolds: generalities

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:generalites (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: $\{Y \in \mathbb{R}^{n\times p} : Y^T RY = I\}$.
- Products of these manifolds.

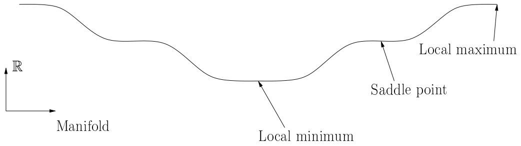
All these manifolds can be turned into Riemannian manifolds by smoothly defining an inner product on the tangent spaces.

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.



Optimization on Manifold
approach to
Principal/Minor Component Analysis
(PCA/MCA)

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

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

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

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

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)

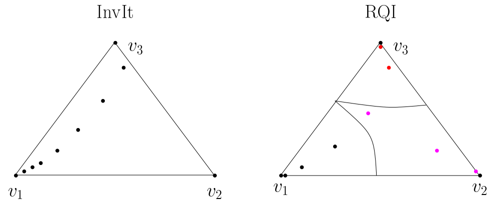
$$\rho_k = \frac{y_k^T A y_k}{y_k^T y_k}$$

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

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.

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.

MCA as 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.

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

TRUST-REGION METHOD

where the trust-region subproblems are solved with a

TRUNCATED CONJUGATE-GRADIENT

algorithm.

Riemannian Trust-Region (RTR) method
and application to
Principal/Minor Component Analysis
(PCA/MCA)

Outline (Part 2)

- 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 $\mathbb{R}^n$

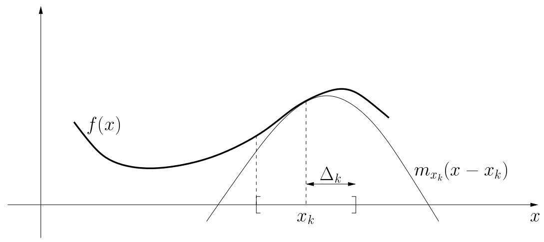
1. Consider a cost function f in $\mathbb{R}^n$. Let x_k be the current iterate.
2. Build a model $m_k(s)$ of f around x_k. The model should agree to f at x_k to the first order at least, and to the second order if superlinear convergence is sought.
3. Find (up to some precision) a minimizer s_k of the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

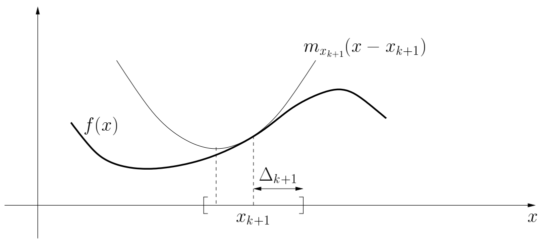
5. Shrink, enlarge or keep the trust-region radius according to the value of ρ.
6. Accept or reject the proposed new iterate $x_k + s_k$ according to the value of ρ.
7. Increment k and go to step 2.

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

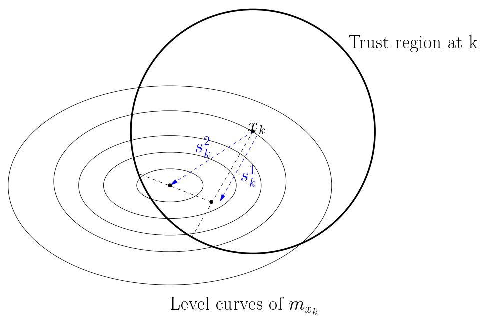
Principle of Trust-Region (TR) in $\mathbb{R}^n$



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:

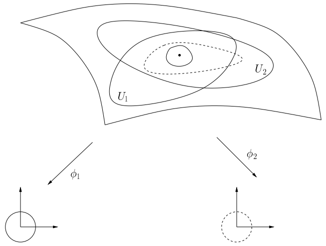
$$\min\{\theta + 1, 2\}.$$

Outline (Part 2)

- 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 $\alpha \in \mathbb{R}$.



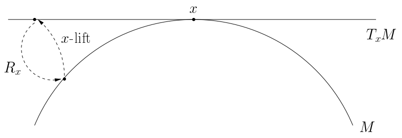
Trust-region methods on Riemannian manifolds: remedies

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)

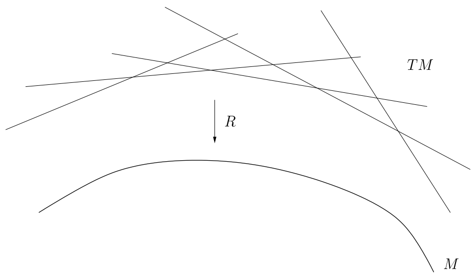
Concept of *retraction*:



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

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

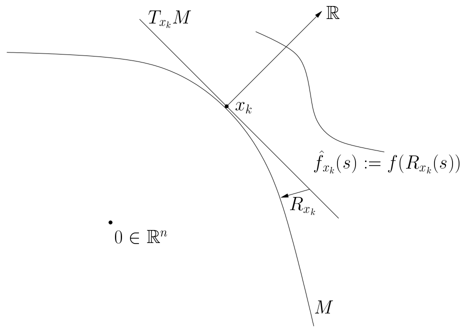
   (note the presence of R_{x_k} !) to compare the actual value of the cost function at the proposed new iterate with the value predicted by the model.
5. Shrink, enlarge or keep the trust-region radius according to the value of ρ.
6. Accept or reject the proposed new iterate $R_{x_k}s_k$ according to the 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 $\delta^0 = -\mathrm{grad}\, f(x_k)$.
- 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 grad $\hat{f}_{x_k}(0)$.
- Hessian Hess $\hat{f}_{x_k}(0)$.

Outline (Part 2)

- 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 $\|\text{grad}\, f(x)\| \leq \beta_g$ and $\|\text{Hess}\, f(x)\| \leq \beta_H$ 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

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

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

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

Let $v \in M$ be a nondegenerate local minimum of f, (i.e., grad $f(v) = 0$ and Hess $f(v)$ is positive definite). Further assume that 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

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

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 $k > K$,

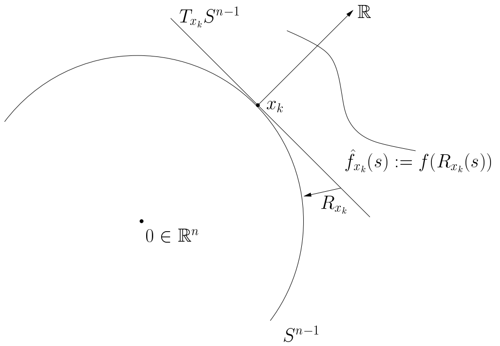
$$\text{dist}(x_{k+1}, v) \leq c\,(\text{dist}(x_k, v))^{\min\{\theta+1,2\}}, \tag{4}$$

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: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - Numerical experiments.

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

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 $\{s : y^T Bs = 0\}$, that is

$$P = I - By(y^T B^2 y)^{-1} y^T B. \tag{5}$$

Trust-region for extreme SGEVP: details

One has:

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

Trust-region for extreme SGEVP: details

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

$$m_y(s) = f(y) + 2\langle PAy, s\rangle + \frac{1}{2}\langle P(A - f(y)B)Ps, s\rangle, \quad y^T Bs = 0. \tag{6}$$

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

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 $Y \in \mathbb{R}^{n\times p} : \text{span}(Y) = \mathcal{Y}$.
3. Tangent space: formally, $T_Y M = \{Z \in \mathbb{R}^{n\times p} : Y^T BZ = 0\}$.
4. Metric: formally, $g_Y(Z_a, Z_b) = \text{trace}\left((Y^T BY)^{-1} Z_a^T Z_b\right)$.

5. Retraction: formally, $R_Y(Z) = (Y + Z)M$, where arbitrary M serves for normalization.
6. Cost function: formally, $f(Y) = \text{trace}\left((Y^T BY)^{-1}(Y^T AY)\right)$.

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:

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

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

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

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

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 (Part 2)

- 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

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

Notice that

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

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

$$\operatorname{trace}\left((Y_+^TBY_+)^{-1}Y_+^TAY_+\right) \leq \operatorname{trace}\left((Y^TBY)^{-1}Y^TAY\right).$$

But superlinear convergence is lost → 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

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

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.

Davidson 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 shifts (Ritz shifts in exact Newton equation).
2. Stopping criterion for inner iteration.
3. Davidson acceleration enhancement.
4. Preconditioning.

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].

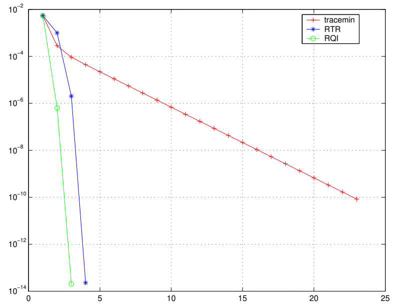
Classification: Lanczos methods

- 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)

Outline (Part 2)

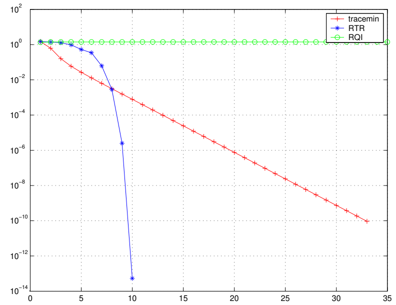
- Trust-region in $\mathbb{R}^n$.
- Trust-region on Riemannian manifolds.
  - Description.
  - Convergence analysis.
- Application: Extreme Component Analysis.
  - Algorithm details.
  - Links with other methods.
  - Numerical experiments.

Numerical experiments: exact simple tracemin, RQI, RTR



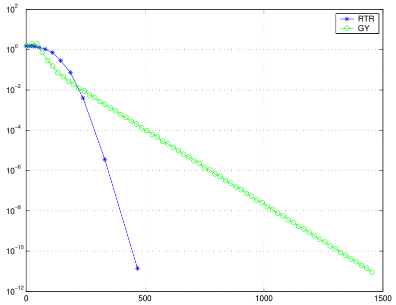
Distance to target versus number of outer iterations.
Simple symmetric positive-definite eigenvalue problem.

Numerical experiments: exact simple tracemin, RQI, RTR



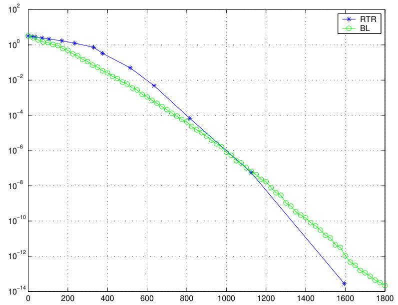
Distance to target versus number of outer iterations.
Simple symmetric positive-definite eigenvalue problem.

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.

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.

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

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

[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.

[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.

[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

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

[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).

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

systems, Springer, 1994.

[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.

[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.

[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.

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

[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 generlized 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.

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

THE END