

ITR Project Meeting, March 25, 2006

Kyle A. Gallivan

Christopher G. Baker

Pierre-Antoine Absil

School of Computational Science,
Florida State University

Collaboration with Ahmed Sameh and Danny Sorensen

List of projects

- Symmetric Generalized Eigenvalue Problem
 - Trust-region methods on Riemannian manifolds
- Low-rank Incremental SVD methods (with Danny Sorensen)
 - Multi-pass algorithms for increased accuracy and confidence

Implicit Riemannian Trust-Region Method for the symmetric generalized eigenproblem

Christopher G. Baker

Pierre-Antoine Absil

Kyle A. Gallivan

School of Computational Science, Florida State University

Département d'ingénierie mathématique, Université Catholique de Louvain

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

Outline

- Review: RTR method for Extreme SGEVP.
- Review: Adaptive Model RTR.
- Implicit Riemannian Trust-Region method.
 - Description.
 - Convergence results.
- IRTR for Extreme SGEVP.
 - Algorithm details.
 - Numerical experiments.

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 $T_x M$:

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

2. Build a model $m_k(s)$ of \hat{f}_{x_k} around 0.
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

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

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.

Required ingredients for Riemannian TR

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

ESGEV: The optimization problem

Given: $n \times n$ pencil (A, B) , $A = A^T$, $B = B^T \succ 0$,

$$Av_i = Bv_i\lambda_i, \quad i = 1, \dots, n$$

$$v_i^T Bv_j = \delta_{ij} \quad \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$$

Problem: compute the “leftmost” eigenspace

$$\mathcal{V} := \text{col}(v_1, \dots, v_p)$$

Solution: \mathcal{V} satisfies

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

$$f : \text{Grass}(p, n) \rightarrow \mathbb{R} : \text{col}(Y) \mapsto \text{tr} \left((Y^T B Y)^{-1} Y^T A Y \right).$$

Trust-region for Extreme SGEVP: principles

Ingredients of the RTR method for ESGEV [ABG06]:

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, $g_Y(Z_a, Z_b) = \text{tr}((Y^T B Y)^{-1} Z_a^T Z_b)$.

Trust-region for Extreme SGEVP: principles (2)

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

Underlying fact: $\begin{bmatrix} v_1 & \dots & v_p \end{bmatrix} M$ minimizes f for all M invertible.

Trust-region for Extreme SGEVP: details

Lifted cost function:

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

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

Trust-region for Extreme SGEVP: details (2)

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

$$\begin{aligned} m_Y(Z) &= f(Y) + g_Y(\text{grad } f(Y), Z) + \frac{1}{2}g_Y(\mathcal{H}_Y Z, Z) \\ &= \text{tr}((Y^T B Y)^{-1} Y^T A Y) + 2\text{tr}((Y^T B Y)^{-1} Z^T A Y) \\ &\quad + \text{tr}((Y^T B Y)^{-1} Z^T (A Z - B Z (Y^T B Y)^{-1} Y^T A Y)). \end{aligned}$$

Compute an approximate minimizer \tilde{Z} using truncated CG [CGT00]

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

Outline

- Review: RTR method for Extreme SGEVP.
- **Review: Adaptive Model RTR.**
- Implicit Riemannian Trust-Region method.
 - Description.
 - Convergence results.
- IRTR for Extreme SGEVP.
 - Algorithm details.
 - Numerical experiments.

A hybrid Tracemin / TR method

Collaboration with Ahmed Sameh.

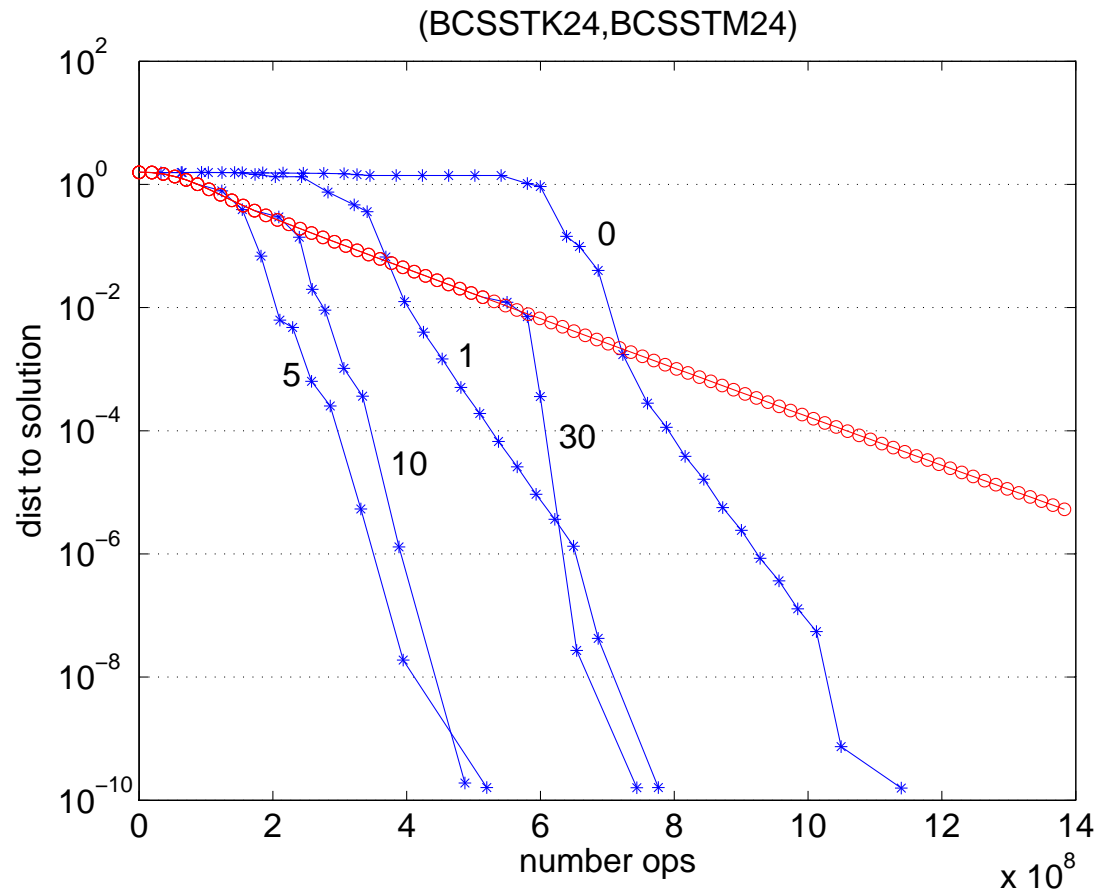
- **Problem:** Trust-region confinement may hamper efficient preconditioning far away from the solution.
 \rightsquigarrow Use preconditioned Basic Tracemin [SW82, ST00] in Phase I.
- **Problem:** Close to the solution, Basin Tracemin is linear.
 \rightsquigarrow Use TR method in Phase II.

Adaptive Model RTR

The method can be described in an Adaptive Model RTR framework [ABGS05].

- Phase I:
 - Use a model Hessian $P_{BX, BX} A P_{BX, BX}$
 - Set trust-region radius to infinity.
- Phase II:
 - Use model Hessian $\text{Hess} \hat{f}_X[S] = P_{BX, BX} (AS - BSX^T AX)$
 - Finite trust-region radius and $\rho' \in (0, 1)$.

EXP: Adaptive Model RTR



Calgary Saddledome, BCSST24; $n = 3562$; preconditioned with exact factorization of A after `symamd`; $p = 5$.

Adaptive Model RTR

- Method inherits the **global convergence** of constituent methods.
- The switching criterion affects efficiency only.
- Potential efficiency greater than constituents.
- **Take-home idea:** Framing the method as a model trust-region optimization allows us to choose best suited model at different points in the computation.

Outline

- Review: RTR method for Extreme SGEVP.
- Review: Adaptive Model RTR.
- **Implicit Riemannian Trust-Region method.**
 - Description.
 - Convergence results.
- IRTR for Extreme SGEVP.
 - Algorithm details.
 - Numerical experiments.

Complaints Against Trust-Region Mechanism

- Trust-region radius is heuristic.
 - Radius of current trust-region based on **performance of last model minimization**.
 - This may constrain current model minimization.
- Iterate may be rejected.
 - **Wasted time** spent computing potential iterate.
 - It can take a number of outer iterations to adjust trust-region radius.

Proposal for New Trust-Region

Suggestion: Based trust-region on the **current performance** of the surrogate model.

New trust-region is

$$\{s \in T_x M : \rho_x(s) \geq \rho'\}, \quad \rho' > 0.$$

ρ_x is as before:

$$\rho_x(s) = \frac{f(x) - f(R_x s)}{m_x(0) - m_x(s)}.$$

Implicit Riemannian Trust-Region (IRTR)

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 $T_x M$:

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

2. Build a model $m_k(s)$ of \hat{f}_{x_k} around 0.
3. Find (approximately) a minimizer s_k of the model within the new trust-region.
4. Accept $x_{k+1} = R_{x_k} s_k$.
5. Increment k and go to step 2.

Solving Model Minimization in IRTR

- Use truncated CG to solve model minimization.
- New trust-region definition requires some modifications.
- Boundary test:
 - Before: check $\|s^j\| \leq \Delta_k$
 - Now: check $\rho_{x_k}(s^j) \geq \rho'$
- If $\rho_{x_k}(s^j) < \rho'$:
 - Compute τ such that $\rho_{x_k}(s^{j-1} + \tau\delta_j) \geq \rho'$
 - This is potentially more difficult than finding $\|s^{j-1} + \tau\delta_j\| = \Delta_k$.

Required ingredients for Implicit RTR

- Manifold M , Riemannian metric g , and cost function f on M
- Practical expression for $T_{x_k}M$
- Retraction $R_{x_k} : T_{x_k}M \rightarrow M$
- Function $\hat{f}_{x_k}(s) := f(R_{x_k}(s))$
- Gradient $\text{grad } \hat{f}_{x_k}(0)$
- Hessian $\text{Hess } \hat{f}_{x_k}(0)$
- Trust-region test: $\rho_{x_k}(s)$
- Trust-region search: find τ s.t. $\rho_{x_k}(s + \tau\delta) = \rho'$

Convergence Results of IRTR

- The trust-region definition is **very strong**.
- As a result, standard TR global convergence results follow easily.
- Global Convergence of IRTR for ESGEV: Let $\{y_k\}$ be a sequence of iterates produced via IRTR-tCG with $\rho' \in (0, 1)$. Then

$$\lim_{k \rightarrow \infty} \|\text{grad} f(y_k)\|_2 = 0.$$

- The local convergence results have not yet been adapted from the RTR to the IRTR. Also, the global convergence results have yet to be adapted to a general IRTR.

Outline

- Review: RTR method for Extreme SGEVP.
- Review: Adaptive Model RTR.
- Implicit Riemannian Trust-Region method.
 - Description.
 - Convergence results.
- **IRTR for Extreme SGEVP.**
 - Algorithm details.
 - Numerical experiments.

IRTR for the Extreme SGEVP

As with RTR for Extreme SGEVP:

1. Manifold: Grassman, represented as $Y \in \mathbb{R}^{n \times p}$, $Y^T B Y = I$
2. Tangent space: $T_Y M = \{Z \in \mathbb{R}^{n \times p} : Y^T B Z = 0\}$.
3. Metric: $g_Y(Z_a, Z_b) = \text{tr}(Z_a^T Z_b)$.
4. Retraction: $R_Y(Z) = (Y + Z)M$, M for B -orthonormalization
5. Cost function: $f(Y) = \text{tr}((Y^T B Y)^{-1}(Y^T A Y))$.

IRTR for the Extreme SGEVP (2)

Therefore, cost function \hat{f}_Y is

$$\hat{f}_Y(S) = \text{tr} \left((I + S^T B S)^{-1} (Y + S)^T A (Y + S) \right)$$

$$\text{grad } \hat{f}_Y(0) = P_{BY, BY} A Y$$

$$\text{Hess } \hat{f}_Y(0)[S] = P_{BY} (A S - B S Y^T A Y)$$

Newton model of \hat{f}_Y is

$$m_Y(S) = \text{tr} (Y^T A Y) + 2 \text{tr} (S^T A Y) + \text{tr} (S^T (A S - B S Y^T A Y))$$

Case $p = 1$

If $p = 1$, then $\rho_y(s) = \frac{\hat{f}_y(0) - \hat{f}_y(s)}{m_y(0) - m_y(s)} = \frac{1}{1 + s^T B s}$.

- Checking trust-region inclusion requires checking $\|s\|_B$
- Solving ρ_y along a tangent vector has an **analytical solution**: τ s.t. $\rho_y(s + \tau\delta) = \rho'$ given by

$$\tau = \frac{-\delta^T B s + \sqrt{(\delta^T B s)^2 + \delta^T B \delta (\Delta_{\rho'}^2 - s^T B s)}}{\delta^T B \delta}$$

$$\Delta_{\rho'} = \sqrt{\frac{1}{\rho'} - 1}$$

- IRTR for $p = 1$ ESGEV is straightforward.

Implications for RTR

- Implications for RTR-ESGEV when $p = 1$, $B = I$, no preconditioning.
 - In this case, TR defined by $\|s\|_2 = \|s\|_B \leq \Delta_k$
- If $\frac{1}{\sqrt{3}} < \Delta_k < \sqrt{3}$, then
 - $\rho_{y_k}(s_k) > \frac{1}{4} \Rightarrow$ **iterate is acceptable!**
 - $\Delta_{k+1} = \Delta_k \Rightarrow$ **trust-region radius is maintained!**
- Properly chosen Δ_0 guarantees model performance.

Even without preconditioner or $B \neq I$, IRTR recommended.

Case $p > 1$

$$\rho_Y(S) = \frac{\text{tr} \left((I + S^T B S)^{-1} (S^T B S (Y^T A Y) - 2S^T A Y - S^T A S) \right)}{\text{tr} (S^T B S (Y^T A Y) - 2S^T A Y - S^T A S)}$$

Assume that $Y^T B Y = I$ and $Y^T A Y = \Sigma$. Then

$$\begin{aligned} m_Y(S) &= \text{tr} (Y^T A Y + 2S^T A Y + S^T (A S - B S Y^T A Y)) \\ &= \sum_{i=1}^p (\sigma_i + 2s_i^T A y_i + s_i^T (A s_i - B s_i \sigma_i)) \\ &= \sum_{i=1}^p m_{y_i}(s_i). \end{aligned}$$

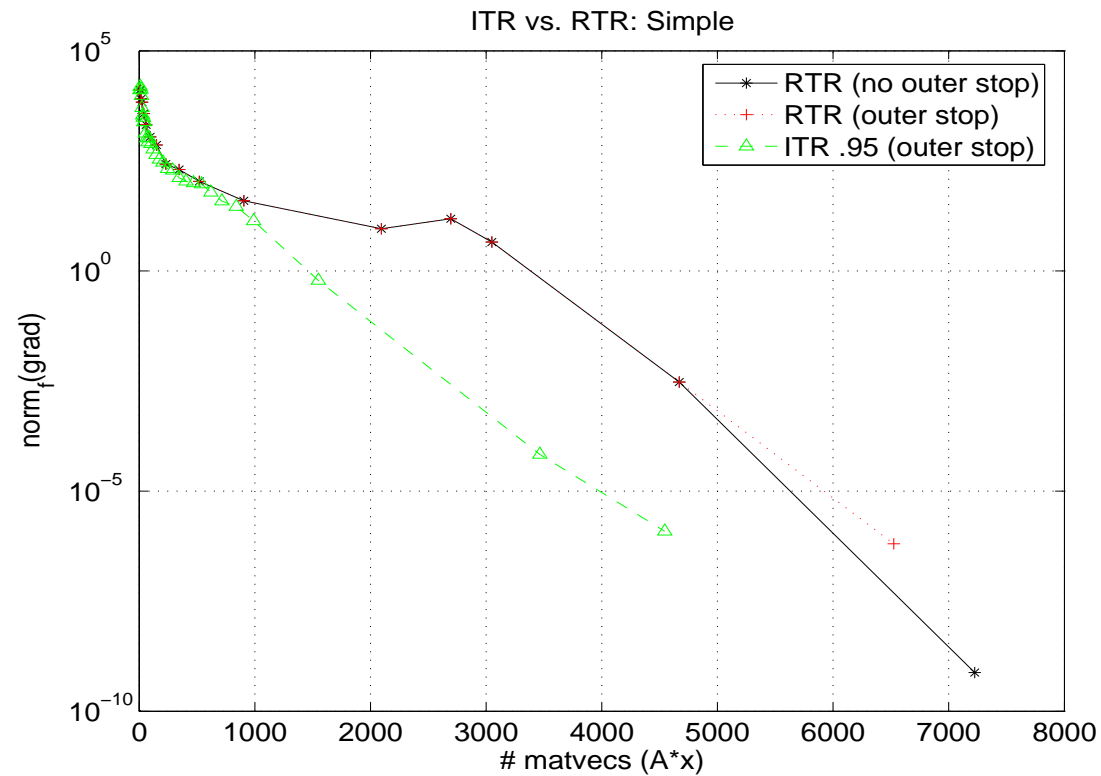
Case $p > 1$

- The $p > 1$ model $m_Y(S)$ can be decoupled into p “scalar” models, for which we have a formula for ρ .
- The block algorithm runs p **simultaneous** tCG algorithms.
- All processes are stopped if any satisfies a stopping criterion.
- Global convergence is still guaranteed.
- But $\rho_Y(S) \not\equiv \rho'$: **not a true IRTR!**

Outer Criterion Monitoring

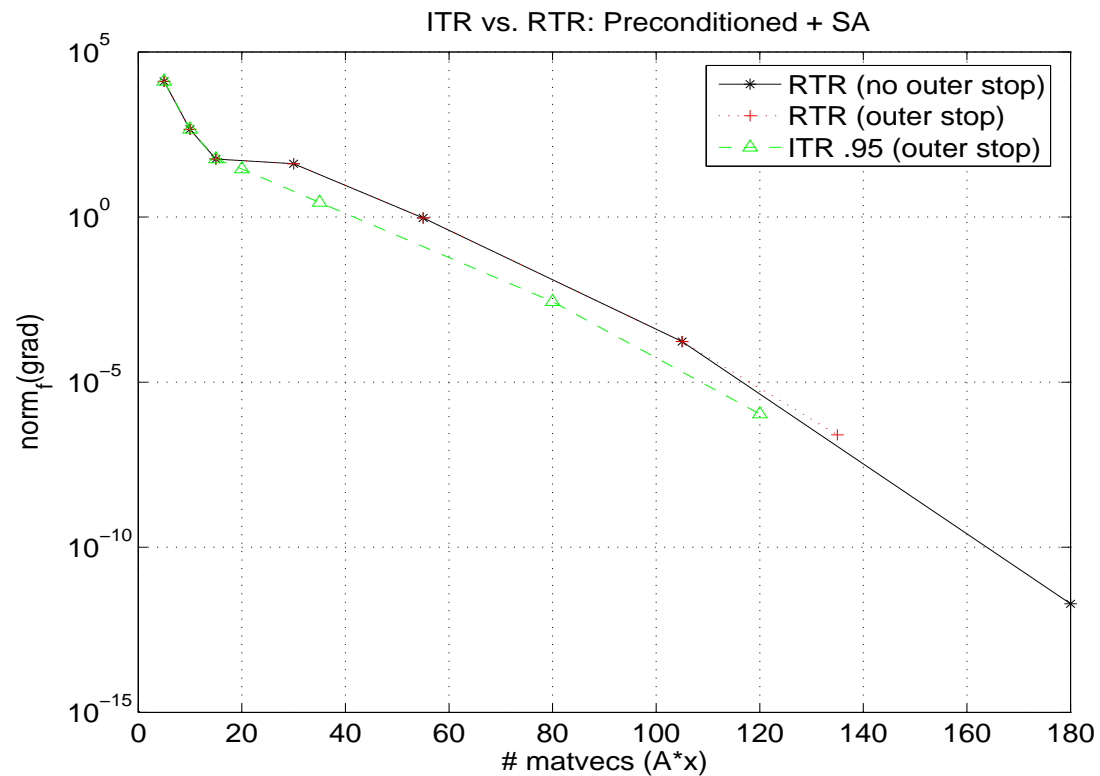
- Last call to tCG often performs more work than necessary to satisfy outer stopping criterion.
- Problem is typical for methods employing an inner iteration.
- Solution is (occasionally) compute outer residual in inner iteration, check stopping criterion.
- Similar to suggestion in [Not02], except we have no efficient formula for the residual norm.

EXP: Monitoring Outer Stopping Criterion (1)



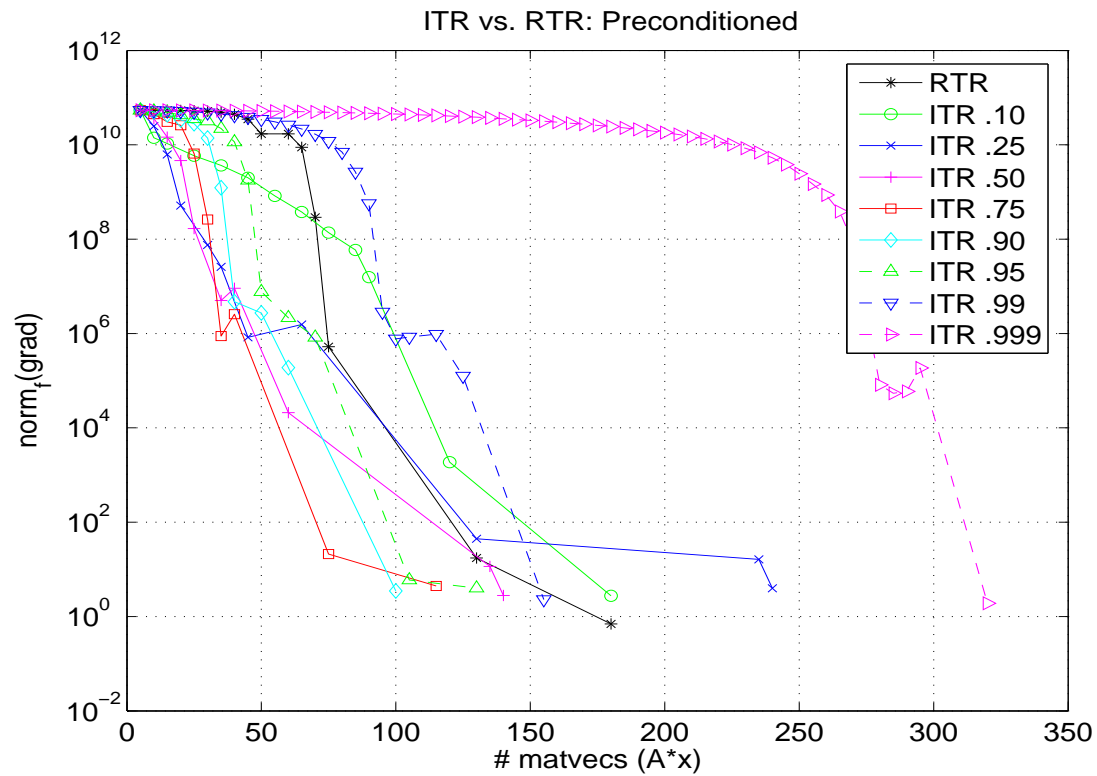
2-D Laplacian, $n = 10000$; no preconditioning; no subspace acceleration; $p = 5$

EXP: Monitoring Outer Stopping Criterion (2)



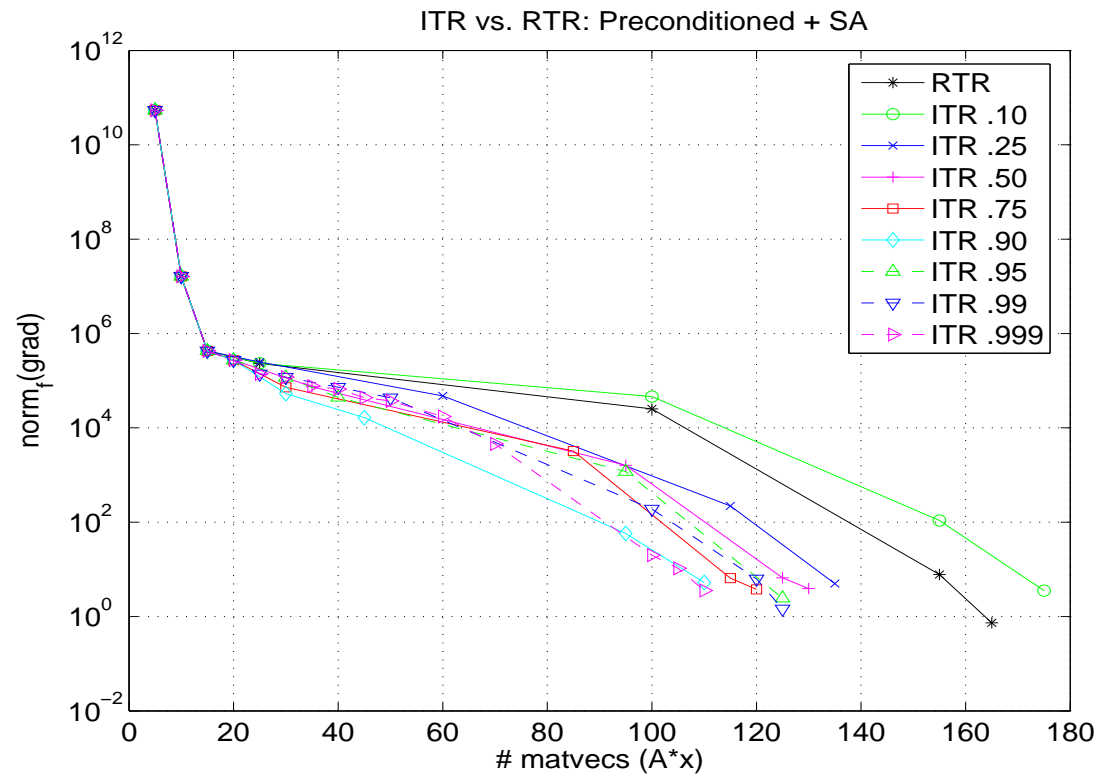
2-D Laplacian, $n = 10000$; preconditioned using exact factorization of A after symamd; 10-D subspace acceleration; $p = 5$

EXP: IRTR vs. RTR



BCSST24; preconditioned with exact factorization of A after `symamd`; no subspace acceleration; $p = 5$

EXP: IRTR vs. RTR



BCSST24; preconditioned with exact factorization of A after `symamd`; 10-D subspace acceleration; $p = 5$

Goal of IRTR

- Combined with SA, IRTR switches between optimizing on TM and M .
- **Take-home idea:** Break down the barrier between inner and outer iteration:
 - Outer criterion monitoring stops when iteration is ultimately satisfied; **always maintain awareness of outer error**
 - Base trust-region on the performance of surrogate model; **always maintain awareness of cost function**
- Both ideas touched on in [Not02].
- Reduce all TR parameters to one: ρ'

Future Work

- Formulation of a true block IRTR for Extreme SGEVP.
- Convergence results of IRTR for general (M, g, R, f)
- Application of IRTR to other NLA problems.
- Explore affect of ρ' parameter.
- Look at adaptive model mechanism for IRTR.

References

- [ABG06] P.-A. Absil, C. G. Baker, and K. A. Gallivan, *A truncated-CG style method for symmetric generalized eigenvalue problems*, J. Comput. Appl. Math. **189** (2006), no. 1–2, 274–285.
- [ABGS05] P.-A Absil, C. G. Baker, K. A. Gallivan, and A. Sameh, *Adaptive model trust region methods for generalized eigenvalue problems*, International Conference on Computational Science (Vaidy S. Sunderam, Geert Dick van Albada, and Peter M. A. Sloot, eds.), Lecture Notes in Computer Science, vol. 3514, Springer-Verlag, 2005, pp. 33–41.
- [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.

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

THE END

Algorithm 1 (Prec. Truncated CG for IRTR)

Set $s^0 = 0$, $r_0 = \text{grad} \hat{f}_y$, $z_0 = M^{-1}r_0$, $d^0 = -z_0$

for $j = 0, 1, 2, \dots$

Check inner stopping criterion

Check $\delta_j^T H_y[\delta_j]$

Compute $\tau \geq 0$ s.t. $s = s^j + \tau\delta_j$ satisfies $\rho_y(s) = \rho'$; return s

Set $\alpha^j = (z_j^T r_j) / (\delta_j^T H_y[\delta_j])$

Set $s^{j+1} = s^j + \alpha_j \delta_j$

if $\rho_y(s^{j+1}) < \rho'$

Compute $\tau \geq 0$ s.t. $s = s^j + \tau\delta_j$ satisfies $\rho_y(s) = \rho'$; return s

Check outer stopping criterion

Set $r_{j+1} = r_j + \alpha^j H_y[\delta_j]$

Set $z_{j+1} = M^{-1}r_{j+1}$

Set $\beta^{j+1} = (z_{j+1}^T r_{j+1}) / (z_j^T r_j)$

Set $\delta_{j+1} = -z_{j+1} + \beta^{j+1}\delta_j$

end.