Adaptive Model Trust-Region Methods for Generalized Eigenvalue Problems

Pierre-Antoine Absil [1]

Christopher G. Baker [1]

Kyle A. Gallivan [1]

Ahmed Sameh [2]

[1] *School of Computational Science, Florida State University*

[2] *Department of Computer Science, Purdue University*

`International Conference on Computational Science 2005`

Outline

Organization of this talk

- Introduction
- Model minimization framework
- Experimental Results
- Conclusion

The Problem: Leftmost Eigenspace of a Matrix

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

$$A \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} = B \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \operatorname{diag}(\lambda_1, \dots, \lambda_n)$$

$$\begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}^T B \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} = I$$

$$\lambda_1 < \lambda_2 \leq \dots \leq \lambda_n$$

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

Compute the minor p-dimensional eigenspace $\operatorname{col}(v_1, \dots, v_p)$.

Consider Inverse Iteration and Rayleigh Quotient Iteration

Preliminaries (1): Inverse Iteration (INVIT)

$$x_{k+1} = \frac{A^{-1}x_k}{\|A^{-1}x_k\|}$$

Properties:

- Global convergence to $\{\pm v_1, \ldots, \pm v_n\}$.
- Stable convergence to $\pm v_1$ only.
- Local linear convergence, with ratio $\frac{\lambda_1}{\lambda_2}$.
- Computing a new iterate is expensive.

Preliminaries (2): Rayleigh Quotient Iteration (RQI)

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

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

Properties:

- Cubic local convergence.
- Converges to "nearest" eigenvector.
- Computing a new iterate is expensive.

The Ideal Algorithm

1. Global convergence:
   - Convergence to some eigenvector for all initial conditions.
   - Stable convergence to the minor eigenvector $\pm v_1$ only.
2. Superlinear (cubic) local convergence to $\pm v_1$.
3. No factorization of A.
   Matrix A only utilized as operator $x \mapsto Ax$.
4. Low storage requirements.

A Hybrid Approach

Combine INVIT and RQI in two phases:

- Phase I: Use INVIT to reach domain of attraction of v_1
- Phase II: Use RQI to yield superlinear convergence

There are two problems with this approach:

- Need a practical and reliable switching criterion!
- Exact INVIT and RQI are expensive!

Proposed Remedy (1)

Phase I: Replace INVIT with Basic Tracemin (Sameh and Wisniewski [SW82], Sameh and Tong [ST00])

Basic Tracemin uses

- successive approximate minimization of...
- inexact local quadratic models...
- of the generalized Rayleigh quotient.

Similar to INVIT; equivalent when using exact preconditioner.

However, rate of convergence is only linear!

Proposed Remedy (2)

We desire a superlinear method for Phase II that is reliable and less expensive than RQI.

The recently proposed Riemannian Trust-Region (RTR) algorithm (Absil, Baker and Gallivan [ABG04]) is

- globally convergent
- superlinear convergence near the solution
- “matrix-free” and low-memory

Trust-region mechanism can prevent RTR from fully exploiting a good preconditioner.

Proposed hybrid method

Phase I: Tracemin

Phase II: RTR

Result: Efficient, globally convergent method with a superlinear rate of convergence near the solution.

We unify both methods in an adaptive model-based framework for minimizing the generalized Rayleigh quotient

Outline

Organization of this talk

- Introduction
- Model minimization framework
- Experimental Results
- Conclusion

Optimization Formulation of SGEP

Goal: compute the p leftmost eigenpairs of the generalized eigenproblem, where A and B are symmetric and positive definite:

$$(v_1, \lambda_1), \ldots, (v_p, \lambda_p)$$

$$\lambda_1 \leq \ldots \leq \lambda_p < \lambda_{p+1}$$

Subspace $\mathcal{V}_p = \text{col}(v_1, \ldots, v_p)$ is the column space of any minimizer of the Generalized Rayleigh cost function:

$$f : \mathbb{R}_*^{n \times p} \to \mathbb{R} : X \mapsto \text{trace}\left((X^T B X)^{-1}(X^T A X)\right)$$

Successive Model Minimization

Given $X \in \mathbb{R}_*^{n \times p}$, produce a correction S s.t. $f(X+S) < f(X)$

Choose S from $T_X := \{Z \in \mathbb{R}^{n \times p} : Z^T B X = 0\}$.

Successive minimizations attempt to minimize the function

$\hat{f}_X(S) := \mathrm{trace}\left(((Y+S)^T B(Y+S))^{-1}((Y+S)^T A(Y+S))\right)$,

for $S \in T_X$.

We will perform this minimization using a trust-region method on a quadratic model m_X of $\hat{f}_X$

Trust-region Strategy

Init: $X_0^T B X_0 = I_p$

for $k = 0, 1, 2, \ldots$

Obtain S_k as approximate solution of model minimization:

$$\min m_{X_k}(S) \quad \text{s.t. } \|S\| \leq \Delta_k, S \in T_X$$

$$m_{X_k}(S) := f(X_k) + \langle 2PAX_k, S\rangle + \frac{1}{2}\langle \mathcal{H}_{X_k}[S], S\rangle$$

Compute $\rho_k = \frac{\hat{f}(S_k) - \hat{f}(0)}{m(S_k) - m(0)}$

Update the trust-region radius and iterate:

$X_{k+1} = X_k$ or via B-orthonormalization of $(X_k + S_k)$

end for

$$P = I - BX(X^T B^2 X)^{-1} X^T B \qquad \langle X, Y\rangle = \text{trace}(X^T Y)$$

Solving the TR Subproblem

To solve the model minimization, we use a preconditioned truncated conjugate gradient algorithm (Steihaug [Ste83] and Toint [Toi81]).

Method has low-memory requirements and is matrix-free with respect to A and B.

For a preconditioner M, we use Olsen formula to solve $(PMP)\tilde{R} = R$, for $P\tilde{R} = \tilde{R}$ and $PR = R$:

$$\tilde{R} = M^{-1}R - M^{-1}BX(X^TBM^{-1}BX)^{-1}X^TBM^{-1}R$$

Truncated CG

Set $S_0 = 0$, $R_0 = 2PAX_k$, $\tilde{R}_0 = (PMP)^\dagger R_0$, $D_0 = -R_0$

for $j = 0, 1, 2, \ldots$ until an inner stopping criterion is satisfied

    if $\langle D_j, \mathcal{H}_{X_k}[D_j] \rangle \leq 0$

        Compute τ such that $S = S_j + \tau D_j$ minimizes $m(S)$

        and $\|S\|_M = \Delta$; **return** S;

    Set $\alpha_j = \langle R_j, \tilde{R}_j \rangle / \langle D_j, \mathcal{H}_{X_k}[D_j] \rangle$; Set $S_{j+1} = S_j + \alpha_j D_j$;

    if $\|S_{j+1}\|_M \geq \Delta$

        Compute $\tau \geq 0$ such that $S = S_j + \tau D_j$ satisfies

        $\|S\|_M = \Delta$; **return** S;

    Generate R_{j+1} and D_{j+1} using the standard CG recurrences

end for.

Choice of the model

$$m_X(S) := f(X) + \langle 2PAX, S\rangle + \frac{1}{2}\langle \mathcal{H}_X[S], S\rangle$$

$\mathcal{H}_X$ is allowed to be any symmetric operator:

- Tracemin-like:

$$\mathcal{H}_X[S] := 2PAPS$$

- Exact Hessian from quadratic expansion of $\hat{f}_X$:

$$\mathcal{H}_X[S] := 2P\left(AS - BS(X^TBX)^{-1}(X^TAX)\right)$$

Properties of Tracemin model

Recall that $A \succ 0$.

Tracemin-like model Hessian: $\mathcal{H}_X[S] := 2PAPS$.

- $\mathcal{H}_X \succ 0 \Rightarrow$ the stationary point of the model is a minimizer.
- $m_X(S) \leq m_X(0) \Rightarrow f(X+S) \leq f(X)$

Can take trust-region radius $\Delta = \infty$

A large S returned by a good preconditioner is always accepted!

Two-Phase Strategy

This dictates parameters for two phases of algorithm:

Phase I: Tracemin Phase

Use $\mathcal{H}_{X_k}[S] := PAPS$ and $\Delta := +\infty$.

Phase II: RTR Phase

Use exact Hessian and $\Delta_k < \infty$.

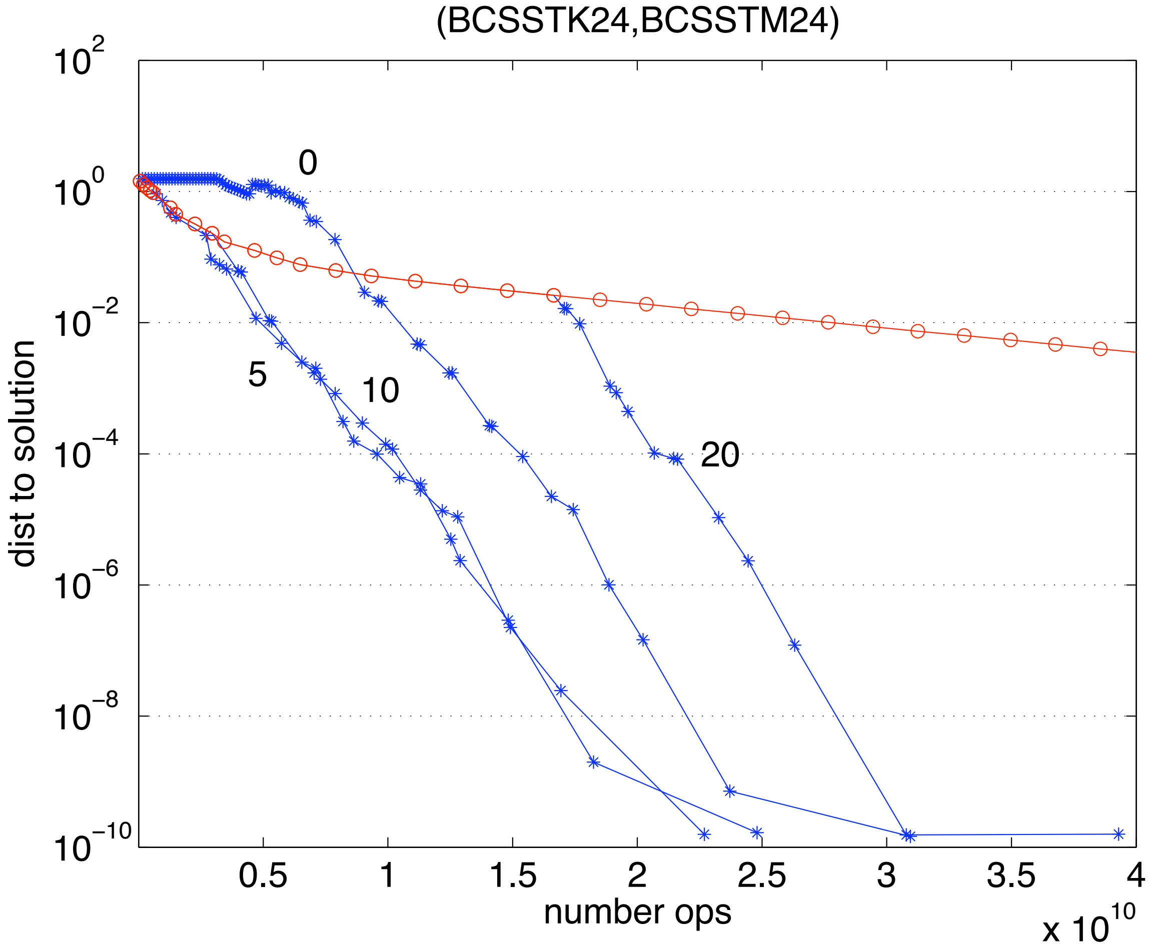
Regardless of switching criteria, algorithm always converges to leftmost eigenspace!

Outline

Organization of this talk

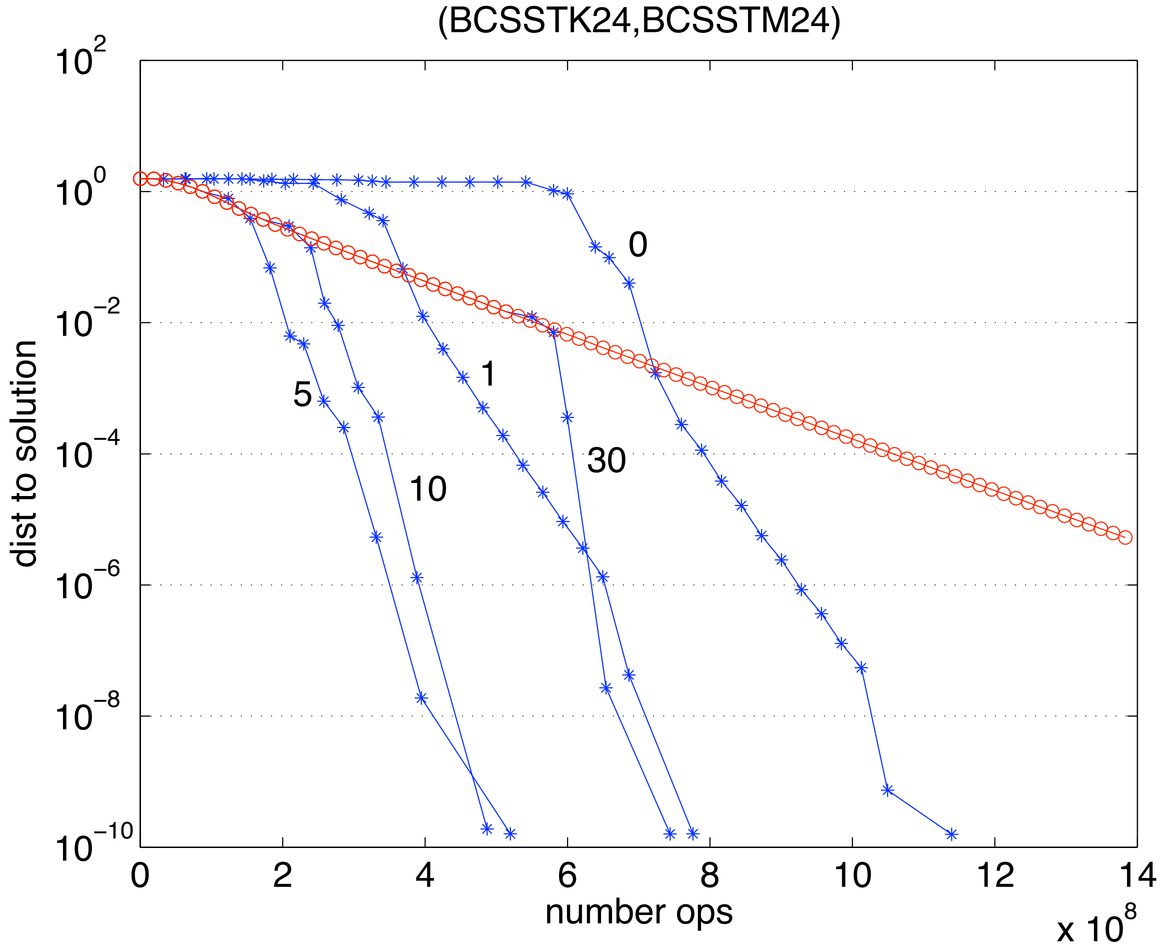
- Introduction
- Model minimization framework
- Experimental Results
- Conclusion

Results



Distance to target versus approximate number of operations. BCSST24 with Incomplete Cholesky preconditioner

Results



BCSST24 with exact (Cholesky) preconditioner after approximate minimum degree permutation.

Conclusion

- We presented a combination Basic Tracemin and Riemannian Trust Region for computing the leftmost eigenpairs of a symmetric positive definite pencil.
- The method is globally convergent with superlinear convergence.
- The appropriate combination is more efficient than the constituent methods.

Future research involves finding switching criteria to maximize the efficiency of the algorihtm.

References

[ABG04] P.-A. Absil, C. G. Baker, and K. A. Gallivan, *Trust-region methods on Riemannian manifolds*, Tech. Report FSU-CSIT-04-13, School of Computational Science, Florida State University, July 2004.

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

[Ste83] T. Steihaug, *The conjugate gradient method and trust regions in large scale optimization*, SIAM J. Numer. Anal. **20** (1983), 626–637.

[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.

[Toi81] Ph. L. Toint, *Towards an efficient sparsity exploiting Newton* *method for minimization*, Sparse Matrices and Their Uses (I. S. Duff, ed.), Academic Press, London, 1981, pp. 57–88.