Recent Work at the Intersection of Optimization and Linear Algebra

Christopher G. Baker*

*Computational Engineering & Energy Sciences (CEES)
Consortium for Advanced Simulation of LWRs (CASL)
Computer Science and Mathematics Division
Oak Ridge National Laboratory

ORNL/UTK Numerical Analysis Day

April 30th, 2011

Outline

Consortium for Advanced Simulation of LWRs

UT-BATTELLE

OAK RIDGE NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Collaborators

- Pierre-Antoine Absil, Université catholique de Louvain
- Kyle Gallivan, Florida State University
- Paul Van Dooren, Université catholique de Louvain

Eigenvalue Problem Background

Problem Definition

Generalized Eigenvalue Problem

Given $A, B \in \mathbb{R}^{n \times n}$, solve:

$$Av = Bv\lambda,$$

for eigenpair (λ, v). Specifically, when $A = A^T$, $B = B^T \succ 0$, we have n eigenpairs satisfying

$$(\lambda_i, v_i) \in \mathbb{R} \times \mathbb{R}^n \quad \text{and} \quad \langle v_i, v_j \rangle_B = \delta_{ij}$$

Application

- Many applications require only p extreme eigenpairs, $AV = V\Lambda$, corresponding to the largest or smallest eigenvalues.
- Examples include problems from structural dynamics, control, signal processing, informatics, etc.

Eigenvalue Problem Background

Solution Techniques

Matrix-free methods

Many applications result in matrices A, B with exploitable structure, cultivating our interest in matrix-free methods:

- Power/Krylov methods
  - power method, inverse iteration, subspace iteration
  - Arnoldi/Lanczos method
- Newton methods
  - Rayleigh quotient iteration
  - Jacobi-Davidson method
- Trace Minimization/Maximization methods
  - Generalized Davidson methods
  - Trace minimization (TRACEMIN) method
  - LOBPCG
  - RTR/IRTR

Optimizing Eigensolvers

Optimization Characterization

The optimization characterization of eigenvalue problems is well-known.

Generalized Eigenvalue Optimization Problem

For s.p.d. eigenproblem, we have that

$$\lambda_1 = \min_{x \neq 0} \frac{x^T A x}{x^T B x} \quad \text{and} \quad \lambda_n = \max_{x \neq 0} \frac{x^T A x}{x^T B x} \; .$$

For multiple eigenvalues,

$$V = \begin{bmatrix} v_1 & \dots & v_p \end{bmatrix}$$

is a minimizer of the generalized Rayleigh quotient:

$$\mathrm{GRQ}(X) = \mathrm{trace}\left(\left(X^T B X \right)^{-1} X^T A X \right)$$

Similarly, the rightmost eigenvectors maximize the GRQ.

Optimizing Eigensolvers

A Naïve First Stab

Newton's method for GRQ

Consider optimizing with Newton's method ($p = 1$ for simplicity):

$$\nabla \mathrm{GRQ}(x) = \frac{2}{x^T B x} (Ax - \rho B x)$$

$$\nabla^2 \mathrm{GRQ}(x) = \frac{2}{x^T B x} \left(I - \frac{2}{x^T B x} B x x^T \right) (A - \rho B) \left(I - \frac{2}{x^T B x} x x^T B \right)$$

for $\rho = \mathrm{GRQ}(x)$.

- Newton's method solves $\nabla^2 \,\mathrm{GRQ}(x) s = -\nabla \,\mathrm{GRQ}(x)$.
- If $\rho \neq \lambda_i$, solution is $s = x$, leading to the following iteration:
  - $x \mapsto 2x \mapsto 4x \mapsto \ldots$, and Newton's method fails!
- This is because $\mathrm{GRQ}(X) = \mathrm{GRQ}(XM)$ for non-singular M.
  - GRQ is invariant to basis, depends only on subspace.
- Failure not unique to GRQ; holds for functions homogenous of degree 0.

Optimizing Eigensolvers

Ties with Classical Methods

Addressing Invariance

- Jacobi-Davidson [SVdV96] and TRACEMIN [SW82,ST2000] methods explicitly normalize X and enforce orthogonality condition on step S.
- LOBPCG [Kny2001] does not specify basis for X; correction in [HL2006] adds basis selection to address other issues.
- Riemannian optimization approaches (RTR) [EAS98,ABG2006] recognize basis invariance, optimize GRQ over Grassmann manifold of subspaces.

Relationship to Classical Optimization Approaches

- J-D: Newton + subspace acceleration for better convergence
- TRACEMIN: Inexact/Quasi-Newton + subspace acceleration for faster convergence
- LOBPCG: CG iteration, using Rayleigh-Ritz for exact minimization
- RTR: GRQ on Riemannian manifold, solved via trust-region methods

Recent Eigensolver Approaches

GRQ and Riemannian Eigensolver Optimization

Riemannian setting

- GRQ is invariant to choice of basis, varies only with subspace.
- Consider the set of p-dimensional subspaces of $\mathbb{R}^n$.
  - This is the Grassmann manifold $\mathrm{Grass}(p, n, \mathbb{R})$
- $\mathrm{GRQ} : \mathrm{Grass}(p, n, \mathbb{R}) \to \mathbb{R} : \mathrm{span}(X) \mapsto \mathrm{trace}\left(\left(X^T B X\right)^{-1} X^T A X \right)$
- $\mathrm{span}(X)$ represented by any basis X.

How to solve this problem?

Previously mentioned algorithms equivalent/analogous to

- GRQ + Riemannian Newton $\Rightarrow$ Jacobi-Davidson
- GRQ + Riemannian Inexact-Newton $\Rightarrow$ TRACEMIN
- GRQ + Riemannian CG $\Rightarrow$ LOBPCG
- GRQ + Riemannian Trust-Region $\Rightarrow$ exciting new eigensolvers!

Recent Eigensolver Approaches

Riemannian Trust-Region Eigensolver

Trust-Region Idea

- Replace GRQ with (quadratic) model $m_X(S)$:

$$m_X(S) = \text{trace}\left(X^T AX\right) + 2\ \text{trace}\left(S^T AX\right) + \frac{1}{2}\text{trace}\left(S^T AS - S^T BSX^T AX\right)$$

- Limit step size to a "trust-region": $\min_{S^T BX=0,\ \|S\|\leq\Delta} m_X(S)$
- Actual vs. predicted performance dictates new trust-region size and whether iterate $X + S$ is accepted.

$$\rho_X(S) = \frac{\text{GRQ}(X+S) - \text{GRQ}(S)}{m_X(S) - m_X(0)}$$

- RTR developed in [ABG2007], eigensolver in [ABG2006]

Recent Eigensolver Approaches

Implicit Trust-Region Eigensolver

Inefficiencies in the trust-region mechanism

- TR too small leads to slow progress
- TR too large leads to rejected updates
- TR performs heuristic, based on previous performance

A New Trust-Region

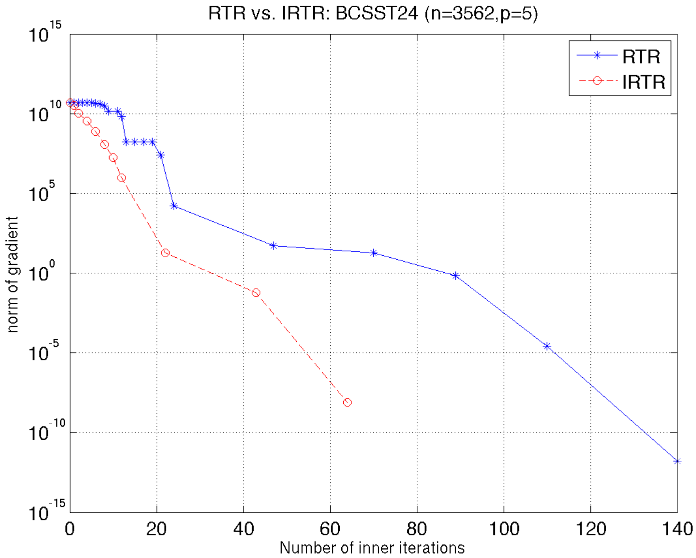
- Implicit RTR replaces trust-region definition. [BAG2008]
- New TR is $\{S \mid \rho_X(S) \geq c\}$; accept/reject can be discarded.
- In general, this formula is difficult to work with.
- However, GRQ with Newton model has nice structure ($p = 1$):

$$\rho_x(s) = \frac{1}{1 + s^T B s}$$

- Resulting method ensures that model is always high-fidelity.

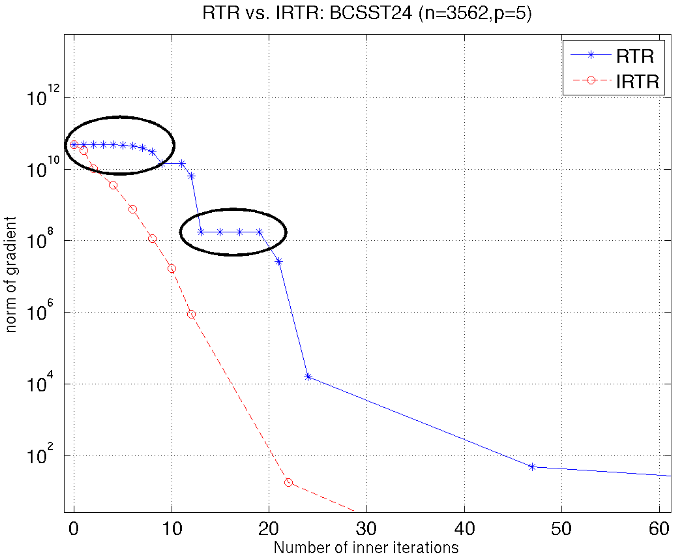
RTR vs. IRTR: A better trust-region

Problem: BCSST24 with Cholesky preconditioner



RTR vs. IRTR: A better trust-region

Better use of preconditioner, no stalling from rejections.



Lessons from Optimization

Propagating solver characteristics to the GRQ

Trust-Region vs. Newton

- TR algorithm has excellent convergence properties:
  - Global convergence, stable convergence to a local minimizer.
  - TR model minimization always well-posed (unlike Newton's linear solve)
  - Model minimization not require to be exact.
  - Both methods enjoy (at least) quadratic local convergence.
- Manifold setting directly addresses invariance problem of GRQ.

RTR vs. Jacobi-Davidson

- TR globalization less useful; provided for JD by Rayleigh-Ritz.
- JD implementations are concerned with shifting to positive definite; RTR eigensolvers enjoy indefiniteness.
- Inexact model minimization saves work in early iterations; in addition, IRTR solver tailored to efficiency of the iteration.
- Both methods can achieve cubic rate of local convergence.

Benchmark Timings: Trilinos/C++

Average speedup of IRTR: 1.33 over RTR, 3.46 over LOBPCG

Problem	Size	p	Prec	RTR	IRTR	LOBPCG
BCSST22	138	5	none	2.64	1.90	39.03
BCSST22	138	5	inexact	1.11	1.03	3.17
BCSST22	138	5	exact	0.29	0.24	0.45
BCSST20	485	5	inexact	49.04	34.40	*151.00
BCSST20	485	5	exact	0.11	0.08	0.14
BCSST13	2,003	25	exact	12.86	7.81	6.20
BCSST13	2,003	100	exact	79.41	56.95	56.12
BCSST23	3,134	25	exact	28.25	22.10	16.86
BCSST23	3,134	100	exact	168.76	129.06	180.40
BCSST24	3,562	25	exact	9.34	8.17	7.76
BCSST24	3,562	100	exact	98.23	69.83	108.20
BCSST25	15,439	25	exact	361.40	85.25	*3218.00

Extreme Singular Value Decomposition

Problem definition

Definition

The singular value decomposition of an $m \times n$ matrix A is

$$A = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^T = U_1 \Sigma V^T$$

with orthogonal U, V; Σ diagonal with non-decreasing, non-negative entries.

Extreme SVD

- Many application require only p extreme singular triplets (typ. largest).
- Compute the dominant/subordinate left and right singular bases for A.
- This is an optimization problem on orthogonal Stiefel manifolds.
- Optimize $f(U,V) = \text{trace}\left(U^T A V N\right)$
- This includes problems from structural dynamics, control, signal processing, and informatics (e.g., PCA, KLT, POD).

Extreme Singular Value Decomposition

Solution Techniques

Numerous characterizations with numerous solutions

- Compute the full SVD using dense methods and truncate.
- Transform to an eigenvalue problem:

$$B = \begin{bmatrix} & A^T \\ A & \end{bmatrix} \quad \text{or} \quad B = AA^T \quad \text{or} \quad B = A^T A$$

  Compute relevant eigenvectors via an iterative eigensolver, then back-transform.
- Use iterative SVD solver to compute just the desired singular triplets:
  - Non-linear equation → JD-SVD [Hochstenbach2000]
  - Riemannian optimization gives many approaches [ABG2007]
  - Low-rank incremental methods

Some of these are only amenable to computing the dominant singular triplets.

Extreme SVD Solvers

Low-Rank Incremental Methods

More efficient approach

The low-rank incremental SVD methods follow the example of the SVD updating methods, but track only a low-dimensional subspace.

History

Repeatedly and independently described in the literature:

- 1995: Manjunath et al.: "Eigenspace Update Algorithm"
- 2000: Levy, Lindenbaum: "Sequential Karhunen-Loeve"
- 2001: Chahlaoui, Gallivan, Van Dooren: "Recursive SVD"
- 2002: Brand: "Incremental SVD"
- 2004: Baker, Gallivan, Van Dooren (generalization, efficiency)
- 2012: Baker, Gallivan, Van Dooren (convergence, efficiency)

Extreme SVD Solvers

Low-Rank Incremental Methods

Kernel Step

Given a matrix A with factorization $A = U\Sigma V^T$, compute updated factorization of augmented matrix $\begin{bmatrix} A & A_+ \end{bmatrix}$:

$$U_+\Sigma_+V_+^T = \begin{bmatrix} A & A_+ \end{bmatrix} = \begin{bmatrix} U\Sigma V^T & A_+ \end{bmatrix}$$

IncSVD consumes all columns, making a single pass through the data matrix. Maintaining low-rank allows for high efficiency, at the expense of accuracy.

Related to an Optimizing Eigensolver

- algorithm can be restarted to take multiple passes through data
- multi-pass algorithms is globally convergent
- equivalent to a coordinate-ascent/descent eigensolver on A^TA
- gradient information can be injected to speed convergence

Extreme SVD Solvers

Optimizing Singular Value Solvers

Direct optimization approach

- Given $A \in \mathbb{R}^{m \times n}$, consider the objective function:

$$\begin{aligned} f &: \mathrm{St}(k,m,\mathbb{R}) \times \mathrm{St}(k,n,\mathbb{R}) \to \mathbb{R} \\ &: (U,V) \mapsto \operatorname{trace}\left(U^T A V N\right) \end{aligned}$$

- Compact Stiefel Manifold: $\mathrm{St}(k,m,\mathbb{R}) = \left\{U \in \mathbb{R}^{m \times k} \mid U^T U = I_k\right\}$
- Riemannian optimization characterization allows application of constellation of solvers over Riemannian manifolds.
- Can only compute dominant SVD, via maximization:
  - minimization of f yields $(-U_1, V_1)$, $f(-U_1, V_1) = -\max(U_1, V_1)$
  - minimization of f^2 yields (U_1, V_2), $f(U_1, V_2) = 0$
- Additional constraint needed to find subordinate singular triplets.
  - Incremental SVD natively addresses this.

Other Problems

Linear and Multi-Linear Riemannian Optimization Problems

Grassmannian/Subspace Optimizations

- Tensor Factorization/HO-SVD [Ishteva et al.][many many others]
  Compute optimal-rank tensor factorization of tensor A, via

$$f(U,V,W) = \|A \bullet_1 U^T \bullet_2 V^T \bullet_3 W^T\|^2$$

- H2-optimal reduced order models [Absil, Gallivan, Van Dooren]

$$f(\hat{H}) = \|\hat{H}(s) - H(s)\|^2_{\mathcal{H}_2}$$

- Interpolation of linear ROMs across parameter changes [Amsallem, Farhat, Lieu]
- Optimal linear subspace for face recognition [Liu, Srivastava, Gallivan]

Other Problems

Linear and Multi-Linear Riemannian Optimization Problems

Basis Optimizations (Stiefel/Oblique)

- ICA, blind-source separation, ("cocktail party problem") [Absil, Gallivan][many others]

$$f(Y) = \sum_{i=1}^{N} \mathrm{trace}\left(\mathrm{off}\left(Y^T C_i Y\right) Y^T C_i Y\right)$$

- Extreme singular triplets

Orthogonal Group Optimizations

- Computer vision problems over $SO(3) = O(2) \times \mathbb{R}^3$
  - Pose estimation
  - Motion recovery
- Full SVD over $O(M) \times O(N)$
- Full eigenvalue decomposition over $O(M)$

Conclusion

Optimization-derived Solvers

- Discussed links between well-understood optimization methods and (sometimes) less-understood eigenvalue and singular value solvers.
- Out-of-the-box optimization methods can produce fast linear algebra solvers, with robust convergence theory.
- Knowledge of the underlying linear algebra problem is still very useful in improving performance of these methods.
- Technology transfer between the domains critical for solver development, especially for non-traditional problems.