Using the Karush-Kuhn-Tucker Conditions to Analyze the Convergence Rate of Preconditioned Eigenvalue Solvers

Merico Argentati

University of Colorado Denver
Joint work with Andrew V. Knyazev, Klaus Neymeyr and Evgueni E. Ovtchinnikov

October 18, 2010

Outline of Presentation

- Introduction
- Preconditioning for linear systems and eigenvalue problems
- Background for our research and analysis
- Preconditioned gradient iteration with a fixed step size
- A convergence rate bound
- Apply Karush-Kuhn-Tucker (KKT) theory (non–linear programming)
- Conclusions

Introduction - Preconditioned Eigenvalue Solvers

- Preconditioned iterative solvers are designed for extremely large eigenvalue problems and popular in various application areas
- The convergence theory of such eigensolvers is still in its infancy with just a handful of convergence rate bounds derived.
- Only one of these known bounds, for the simplest preconditioned eigensolver–the preconditioned gradient iteration with a fixed step size—is sharp.
- We give a "text book" proof for this bound, that holds for the complex case, using novel ideas, and the Karush-Kuhn-Tucker conditions

Preconditioning for Linear Systems

- In linear algebra and numerical analysis, a preconditioner of a matrix A is a matrix such that $P^{-1}A$ has a smaller condition number than A
- Instead of solving the original linear system $Ax = b$, one may solve the left precondioned (or right) system $T(Ax - b) = 0$, where $T = P^{-1}$
- Generally neither $T = P^{-1}$ nor P are explicitly available and multiplication is done in a matrix free fashion
- Preconditioned iterative methods for $Ax - b = 0$ are, in most cases, mathematically equivalent to standard methods for $P^{-1}(Ax - b) = 0$
- Examples include the preconditioned Richardson iteration, the conjugate gradient method, the biconjugate gradient method, etc.
- Preconditoned Richardson iteration method for example implements

$$x_{n+1} = x_n - \gamma_n T(Ax_n - b),\ n \geq 0$$

See http://en.wikipedia.org/wiki/Preconditioner.

Preconditioning for Eigenvalue Problems

- Our goal is to find the eigenpair for $Ax = \lambda x$, or for the generalize eigenvalue problem $Bx = \lambda Ax$, often for the smallest or largest eigenvalue.
- Suppose that the targeted eigenvalue λ_* is known (approximately). Then one can compute the corresponding eigenvector for the linear system $T(Ax - \lambda_* I)x = 0$. Then solve using Richardson iteration

$$x_{n+1} = x_n - \gamma_n T(A - \lambda_* I)x_n,\ n \geq 0$$

- A popular choice for λ_* is the Rayleigh quotient

$$\mu(x_n) = x_n^T Bx_n / x_n^T Ax_n$$

which brings preconditioned optimization techniques to the scene.

Preconditioning for Eigenvalue Problems (Cont.)

- Due to the changing value λ_n, a comprehensive theoretical convergence analysis is much more difficult, compared to the linear systems case, even for the simplest methods, such as the Richardson iteration
- Preconditioned eigensolvers have been used in practice, e.g., for band structure calculations, thin elastic structures, electronic structure calculations, etc.
- Examples include preconditioned gradient iteration, steepest descent, power iteration, inverse iteration, Rayleigh quotient iteration, LOBPCG, etc.

Background for Our Research and Analysis

- The work to analyze the gradient iterative method with a fixed step size convergent rate has been in progress for several years by these collaborators
- The original proof [Knyazev and Neymeyr (2003)] has been recently significantly simplified and shortened using a gradient flow integration approach [Knyazev and Neymeyr (2009)]
- We use powerful tools, which are uncommon in numerical linear algebra, and provide *Preconditioned eigenvalue solver convergence theory in a nutshell*,
  [Argentati, Knyazev, Neymeyr, Ovtchinnikov - tbp]
- The problem is captivating and has an interesting geometrical interpretation involving norms, angles, closed balls, the Rayleigh quotient and optimization theory

Convergence Rate Bound

Problem Definition

- We consider a generalized eigenvalue problem $Bx = \mu(x)Ax$, with real symmetric positive definite matrices A and B with eigenvalues enumerated in decreasing order $\mu_1 \geq \ldots \geq \mu_n > 0$
- The objective is to approximate iteratively the largest eigenvalue μ_1 by maximizing the Rayleigh quotient $\mu(x) = x^T Bx / x^T Ax$. We iteratively correct a current iterate x along the preconditioned gradient of the Rayleigh quotient, i.e.

$$x' = x + \frac{1}{\mu(x)} T(Bx - \mu(x)Ax). \qquad (1)$$

- T is a real symmetric positive definite matrix called the preconditioner, where

$$(1-\gamma) z^T T^{-1} z \leq z^T A z \leq (1+\gamma) z^T T^{-1} z, \ \forall z, \ \text{for a given } \gamma \in [0,1)$$

Convergence Rate Bound

Main Theorem

Theorem

If $\mu_{i+1} < \mu(x) \leq \mu_i$ *then* $\mu(x') \geq \mu(x)$ *and either* $\mu_i < \mu(x')$ *or*

$$\frac{\mu_i - \mu(x')}{\mu(x') - \mu_{i+1}} \leq \sigma^2 \frac{\mu_i - \mu(x)}{\mu(x) - \mu_{i+1}}, \quad \sigma = \gamma + (1-\gamma)\frac{\mu_{i+1}}{\mu_i}. \tag{2}$$

Note that:

1. We have $\sigma < 1$
2. σ may be small if γ is small and there is a large gap between μ_{i+1} and μ_i

Simplify the Problem

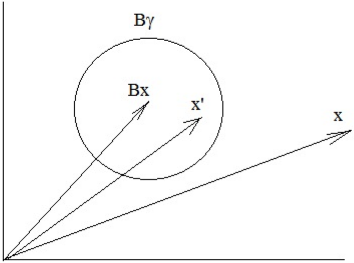
By changing the basis to the eigenvectors of $A^{-1}B$ we make B diagonal and $A = I$, and denote the new inner product by $(\cdot, \cdot)$, so

$$\mu(x)x' = Bx - (I - T)(Bx - \mu(x)x) \tag{3}$$

with $\|I - T\| \leq \gamma$ and $\mu(x) = (x, Bx)/(x, x)$

Geometry of Problem

- We denote $r = Bx - \mu(x)x$ and
$$\mathcal{B}_\gamma = \{y : \ \|Bx - y\| \leq \gamma\|r\|\},$$
which is a closed ball of radius $\gamma\|r\|$ centered at Bx
- Trivially, $\mu(x)x' = Bx - (I - T)r \in \mathcal{B}_\gamma$ since $\|(I - T)r\| \leq \gamma\|r\|$
- The $\mathrm{span}\{x\}$ does not intersect the ball $\mathcal{B}_\gamma = \mathcal{B}_\gamma(x)$



Motivation for Minimization

- The $\mathrm{span}\{x\}$ does not intersect the ball $\mathcal{B}_\gamma = \mathcal{B}_\gamma(x)$
- The latter gives $\cos(\angle\{x, Bx\}) < \cos(\angle\{y, Bx\}) < 1$ and thus $(y, Bx) \neq 0$ for any vector $y \in \mathcal{B}_\gamma$. By direct calculation,

  $$\frac{\cos(\angle\{y, Bx\})}{\cos(\angle\{x, Bx\})} = \frac{|(y, Bx)|\|x\|}{\|y\|(x, Bx)} = \sqrt{\frac{\mu(y)}{\mu(x)}}\frac{|(y, Bx)|}{\|y\|_B\|x\|_B},$$

  thus $\mu(y) > \mu(x)$
- This motivates us to vary $y \in \mathcal{B}_\gamma$ intending to minimize $\mu(y)$

The 2-D Case (Outline of Proof - 1 of 2)

This Proof Works for Complex Case

Proof.

For a minimizer y we can show $(Bx, y) = \|y\|^2$. Combining this with the assumption that $y \in \mathcal{B}_\gamma$, we have for normalize (unit) vectors $\hat{x}$ and $\hat{y}$

$$\|B\hat{x}\|^2 - |(B\hat{x}, \hat{y})|^2 \leq \gamma^2 \left(\|B\hat{x}\|^2 - \mu^2(\hat{x}) \right). \tag{4}$$

Form $\hat{x} = \alpha_1 x_1 + \alpha_2 x_2$, $|\alpha_1|^2 + |\alpha_2|^2 = 1$ and $\hat{y} = \beta_1 x_1 + \beta_2 x_2$, $|\beta_1|^2 + |\beta_2|^2 = 1$. This yields

$$|\alpha_1|^2 = \frac{\mu(x) - \mu_2}{\mu_1 - \mu_2}, \ |\alpha_2|^2 = \frac{\mu_1 - \mu(x)}{\mu_1 - \mu_2},$$

and

$$|\beta_1|^2 = \frac{\mu(y) - \mu_2}{\mu_1 - \mu_2}, \ |\beta_2|^2 = \frac{\mu_2 - \mu(y)}{\mu_1 - \mu_2}.$$

Then we obtain

$$\left| \frac{\beta_2}{\alpha_2} \frac{\alpha_1}{\beta_1} \right| \mu_1 - \mu_2 < \left| \frac{\beta_1}{\alpha_1} \right| \left| \frac{\beta_2}{\alpha_2} \frac{\alpha_1}{\beta_1} \mu_1 - \mu_2 \right| \leq \gamma(\mu_1 - \mu_2),$$

$\square$

The 2-D Case (Outline of Proof - 2 of 2)

This Proof Works for Complex Case

Proof.

Then we obtain

$$\left|\frac{\beta_2}{\alpha_2}\,\frac{\alpha_1}{\beta_1}\right|\mu_1-\mu_2<\left|\frac{\beta_1}{\alpha_1}\right|\left|\frac{\beta_2}{\alpha_2}\,\frac{\alpha_1}{\beta_1}\mu_1-\mu_2\right|\le\gamma(\mu_1-\mu_2).$$

Insertion of α_i, β_i we get

$$\frac{\mu_1-\mu(y)}{\mu(y)-\mu_2}\,\frac{\mu(x)-\mu_2}{\mu_1-\mu(x)}<\left(\gamma+(1-\gamma)\frac{\mu_2}{\mu_1}\right)^2=\sigma^2.$$

The 2-D Case (Outline of Proof - 2 of 2)

This Proof Works for Complex Case

Proof.

Then we obtain

$$\left|\frac{\beta_2}{\alpha_2}\ \frac{\alpha_1}{\beta_1}\right|\mu_1 - \mu_2 < \left|\frac{\beta_1}{\alpha_1}\right|\left|\frac{\beta_2}{\alpha_2}\ \frac{\alpha_1}{\beta_1}\mu_1 - \mu_2\right| \leq \gamma(\mu_1 - \mu_2).$$

Insertion of α_i, β_i we get

$$\frac{\mu_1 - \mu(y)}{\mu(y) - \mu_2}\ \frac{\mu(x) - \mu_2}{\mu_1 - \mu(x)} < \left(\gamma + (1-\gamma)\frac{\mu_2}{\mu_1}\right)^2 = \sigma^2.$$

$$\frac{\mu_1 - \mu(y)}{\mu(y) - \mu_2} \leq \sigma^2 \frac{\mu_1 - \mu(x)}{\mu(x) - \mu_2} \tag{5}$$

The 2-D Case (Outline of Proof - 2 of 2)

This Proof Works for Complex Case

Proof.

Then we obtain

$$\left|\frac{\beta_2}{\alpha_2}\frac{\alpha_1}{\beta_1}\right|\mu_1-\mu_2<\left|\frac{\beta_1}{\alpha_1}\right|\left|\frac{\beta_2}{\alpha_2}\frac{\alpha_1}{\beta_1}\mu_1-\mu_2\right|\le\gamma(\mu_1-\mu_2).$$

Insertion of α_i, β_i we get

$$\frac{\mu_1-\mu(y)}{\mu(y)-\mu_2}\frac{\mu(x)-\mu_2}{\mu_1-\mu(x)}<\left(\gamma+(1-\gamma)\frac{\mu_2}{\mu_1}\right)^2=\sigma^2.$$

$$\frac{\mu_1-\mu(y)}{\mu(y)-\mu_2}\le\sigma^2\frac{\mu_1-\mu(x)}{\mu(x)-\mu_2}\tag{5}$$

$$\frac{\mu_1-\mu(x')}{\mu(x')-\mu_2}\le\frac{\mu_1-\mu(y)}{\mu(y)-\mu_2}\le\sigma^2\frac{\mu_1-\mu(x)}{\mu(x)-\mu_2}\tag{6}$$

obtaining (2) ☐

Formulation of Problem Using Minimum $\mu(y)$

- Choose $\mu_{i+1} < \kappa < \mu_i$, and let $\kappa = \mu(x)$, then

$$\kappa = \mu(x) = \mu(x^*) < \mu(y^*) \leq \mu(w) \leq \mu(x')$$

where $\{x^*, y^*\}$ is a global minimizer for $\mu(y)$ and $w \in \mathcal{B}_\gamma(x)$ is a minimizer for $\mu(w)$

- Our goal is to prove that

$$\frac{\mu_i - \mu(y^*)}{\mu(y^*) - \mu_{i+1}} \leq \sigma^2 \frac{\mu_i - \mu(x)}{\mu(x) - \mu_{i+1}} \tag{7}$$

- In this case we have

$$\frac{\mu_i - \mu(x')}{\mu(x') - \mu_{i+1}} \leq \frac{\mu_i - \mu(y^*)}{\mu(y^*) - \mu_{i+1}} \leq \sigma^2 \frac{\mu_i - \mu(x)}{\mu(x) - \mu_{i+1}}, \tag{8}$$

since $\frac{\mu_i - d}{d - \mu_{i+1}}$ is a decreasing function of $d \in (\mu_{i+1}, \mu_i)$.

Setup for Non-Linear Programming

- Given $\mu_{i+1} < \mu(x) < \mu_i$, we want to find $y \in \mathcal{B}_\gamma$ such that $\mu(y)$ is a minimum. Now we will let x, y both vary. But we choose fixed κ such that $\mu_{i+1} < \kappa < \mu_i$ and constrain x such that $\mu(x) = \kappa$.
- We seek a pair of vectors $\{x^*, y^*\}$ that are a solution for the following constrained minimization problem:

$$\begin{aligned} \text{minimize} \quad & f(x,y) = \mu(y),\ x \neq 0 \\ \text{subject to} \quad & g(x,y) = \|Bx - y\|^2 - \gamma^2 \|Bx - \kappa x\|^2 \leq 0 \\ & h(x,y) = \kappa(x,x) - (x,Bx) = 0 \end{aligned}$$

Reminder:

$$\mathcal{B}_\gamma = \{y : \ \|Bx - y\| \leq \gamma \|r\|\} = \{y : \ \|Bx - y\| \leq \gamma \|Bx - \mu(x)x\|\}$$

Apply Karush-Kuhn-Tucker (KKT) Conditions

Real Case

- The KKT stationarity condition states that there exists constants a and b such that
$$\nabla f(x^*, y^*) + a\nabla g(x^*, y^*) + b\nabla h(x^*, y^*) = 0$$
at the critical point $\{x^*, y^*\}$.
- To simplify the notation, in the rest of the proof we drop the superscript * and substitute $\{x, y\}$ for $\{x^*, y^*\}$.
- Use $r = Bx - \kappa x$ and directly calculate separately the two components of the gradient

$$2a\left(B^2x - By - \gamma^2(B - \kappa I)r\right) - 2br = 0, \tag{9}$$

$$2\frac{By - \mu(y)y}{(y, y)} + 2a(y - Bx) = 0. \tag{10}$$

KKT Conditions - Remarks

- We notice $x \neq 0$ implies $y \neq 0$. Second, let us temporarily consider a stricter, compared to $x \neq 0$, constraint $\|x\| = 1$. Combined with the other constraints, this results in minimization of a smooth function $f(x, y)$ on a compact set, so there exists a solution $\{x^*, y^*\}$. Finally, let us remove the artificial constraint $\|x\| = 1$ and notice that the solution $\{x^*, y^*\}$ is scale-invariant.
- Thus we can consider the Karush-Kuhn-Tucker (KKT) conditions, e.g., [**?**, Theorem 9.1.1], in any neighborhood of $\{x^*, y^*\}$, which does not include the origin.
- The gradients of g and h are linearly independent simply since g depends only on x. Thus, the stationary point is regular, i.e., the KKT conditions are valid.

The Role of 2–D Invariant Subspaces

Lemma

For a fixed value $\kappa \in (\mu_n, \mu_1)$, let a pair of vectors $\{x^*, y^*\}$ denote a solution of the following constrained minimization problem:

$$\begin{aligned} \textit{minimize} \quad & f(x,y) = \mu(y),\ x \neq 0 \\ \textit{subject to} \quad & g(x,y) = \|Bx - y\|^2 - \gamma^2 \|Bx - \kappa x\|^2 \leq 0 \\ & h(x,y) = \kappa(x,x) - (x, Bx) = 0 \end{aligned}$$

Then either y^* is an eigenvector, or both x^* and y^* belong to a two-dimensional invariant subspace of B corresponding to two distinct eigenvalues.

2–D Invariant Subspaces (Outline of Proof - 1 of 2)

Real Case

Proof.

If y is an eigenvector we are done, otherwise introducing two new constants c and d, (10) implies $a \neq 0$, and (9) turns into

$$B(Bx - \gamma^2 r - y) = cr. \tag{11}$$

We rewrite (10) as

$$By - \mu(y)Bx = d(Bx - y). \tag{12}$$

After further manipulations we show that $cd \geq 0$ and we have

$$(1-\gamma^2)B^3x + \kappa\gamma^2(\kappa - c)B^2x - B^2y + c\kappa Bx = 0.$$

Eliminating x we obtain $p(B)y = (c_3B^3 + c_2B^2 + c_1B + c_0)y = 0$, where $p(\cdot)$ is a real polynomial with $c_3 = 1 - \gamma^2 > 0$ and $c_0 = cd\kappa \geq 0$. Since $c_3 > 0$ and $c_0 \geq 0$, the polynomial p_3 must have a non-positive root, and thus at most two positive roots. ☐

2–D Invariant Subspaces (Outline of Proof - 2 of 2)

Real Case

Proof.

Since B is diagonal we have

$$\begin{pmatrix} p(\mu_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p(\mu_n) \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = 0.$$

which implies that at most two components of y are nonzero. Solving for Bx in (10) we have

$$Bx = \left(\frac{(a||y||^2 - \mu(y))I - B}{a||y||^2} \right) y$$

Again since B is diagonal (and invertible), x and y are both in the same 2–D subspace. □

The Complex Case

- The 2–D proof works "as is" for the complex case
- For the KKT proof we use complex vectors $x = x_1 + x_2 i$ and $y = y_1 + y_2 i$. Then the formulation still involves real functions

$$\begin{aligned} \text{minimize} \quad & f(x, y) = \mu(y),\ x \neq 0 \\ \text{subject to} \quad & g(x, y) = \|Bx - y\|^2 - \gamma^2 \|Bx - \kappa x\|^2 \leq 0 \\ & h(x, y) = \kappa(x, x) - (x, Bx) = 0 \end{aligned}$$

- The gradients involve complex functions, but the proof and conclusions are exactly the same as for the real case.

Conclusions

- Use of the KKT approach provides a simple and elegant solution
- A detailed understanding of eigenvalue theory is not necessary
- KKT and non-linear programming should be of interest to a wider audience of mathematicians and graduate students
- This approach may be useful in analyzing other methods (e.g. LOBPCG)

- M. E. Argentati, A. V. Knyazev, K. Neymeyr and E. E. Ovtchinnikov, *Preconditioned eigenvalue solver convergence theory in a nutshell*, to be published.
- A. V. Knyazev, *Computation of eigenvalues and eigenvectors for mesh problems: algorithms and error estimates*, (In Russian), Dept. Num. Math., USSR Ac. Sci., Moscow, 1986.
- A. V. Knyazev, *Convergence rate estimates for iterative methods for a mesh symmetric eigenvalue problem*, Russian J. Numer. Anal. Math. Modelling, 2 (1987), pp. 371–396.
- A. V. Knyazev and K. Neymeyr, *A geometric theory for preconditioned inverse iteration. III: A short and sharp convergence estimate for generalized eigenvalue problems*, Linear Algebra Appl., 358 (2003), pp. 95–114.
- A. V. Knyazev and K. Neymeyr, *Gradient flow approach to geometric convergence analysis of preconditioned eigensolvers*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 621–628.
- E. E. Ovtchinnikov, *Sharp convergence estimates for the preconditioned steepest descent method for hermitian eigenvalue problems*, SIAM J. Numer. Anal., 43(6):2668–2689, 2006.