Gradient flow approach to geometric convergence analysis of preconditioned eigensolvers in A&M

Klaus Neymeyr

Institute of Mathematics, University of Rostock, Germany

Andrew Knyazev (speaker)

Department of Applied Mathematical and Statistical Sciences

Center for Computational Mathematics

University of Colorado Denver (Downtown Campus)

A conference honoring James H. Bramble: 50 years of scientific research.

Texas A&M University May 3, 2008

Supported by the National Science Foundation

Abstract

We first review ugly, but powerful, convergence rate bounds of [1] for symmetric generalized eigenvalue problems. Then a new elegant approach is presented, see arxiv.org/abs/0801.3099, that gives a geometric proof of a sharp convergence rate bound of a simple preconditioned eigensolver–the gradient iterative method with a fixed step size, where we use the gradient of the Rayleigh quotient as an optimization direction. The crucial step of the new proof is a reduction of the convergence analysis to a 2D subspace spanned by relevant eigenvectors, based on analyzing the gradient flow of the Rayleigh quotient. However, it is not currently known if this approach can be directly extended to the subspace iterations to improve and simplify the results of [1], which thus remain unbeatable.

[1] J.Bramble, J.Pasciak, A.Knyazev, A subspace preconditioning algorithm for eigenvector/eigenvalue computation, Adv Comp Math, 6 159–189 1996.

CONTENTS

Bramble-Pasciak-Knyazev (1996) results, p. 1/3

The block method

Matrix A is SPD. Eigenvalues $Av_i = \lambda_i v_i$ increasing and simple, we compute s smallest. A preconditioner B is SPD and $\|I - BA\| \leq \gamma < 1$ Given an approximation subspace V_s^n of dimension s, the eigenvectors and eigenvalues of A are approximated by computing the Ritz eigenvectors $\{v_i^n\} \subset V_s^n$ along with their corresponding eigenvalues λ_i^n. Then

$$\widehat{v}_i^{n+1} = v_i^n - B(Av_i^n - \lambda_i^n v_i^n), i = 1, \ldots, s,$$

and $V_s^{n+1} = \mathrm{Span}\,\{\widehat{v}_1^{n+1}, \ldots, \widehat{v}_s^{n+1}\}$.

Bramble-Pasciak-Knyazev (1996) results, p. 2/3

The convergence rate bound

Define $\theta_i^n \equiv \sin \angle_A \{v_i; V_s^n\}$ and $\Delta \equiv \max_{i=1,\ldots,s} \frac{\lambda_{i+1}+\lambda_i}{\lambda_{i+1}-\lambda_i}$. Let

$$\sum_{i=1}^{s} (\theta_i^0)^2 \leq \frac{(1-\gamma)^2(\lambda_1)^4(1-\lambda_s/\lambda_{s+1})^4}{1999\Delta^2(\lambda_s)^4}$$

Then the dimension of V_s^n for $n > 0$ is equal to s. Moreover, we have

$$0 \leq 1 - \lambda_i/\lambda_i^n \leq \frac{1.03}{1-\lambda_i/\lambda_{s+1}} \bar{\delta}_i^{2n} (\theta_i^0)^2$$

$$\delta_i = \gamma + (1-\gamma)\lambda_i/\lambda_{s+1} < 1, \bar{\delta}_i = \delta_i + \frac{(1-\delta_s)}{2}\left(\frac{\lambda_{s+1}-\lambda_s}{\lambda_{s+1}-\lambda_i}\right)^{1/2} \frac{\lambda_i}{\lambda_s} < 1.$$

Bramble-Pasciak-Knyazev (1996) results, p. 3/3

Possibilities for improvements

- **The bound is NOT SHARP even for the block size $s = 1$**
- **The assumption on the initial approximation is not realistic**
- **The proof is very technical and tedious**
- The bound is cluster robust, but the assumption is not
- The theory works only for this simplest method and thus does not explain the fast convergence, observed in practice, of more sophisticated methods, e.g., such as LOBPCG

Neymeyr (2001), Knyazev-Neymeyr (2003) results

Finally, a sharp bound under realistic assumptions, but

- **The proof is very technical and tedious**
- The bound is not cluster robust for $s > 1$ (but sharp!)
- The theory still works only for the simplest method

The new Knyazev-Neymeyr (2008) results: The same sharp bound and realistic assumptions with an ELEGANT reasonably SIMPLE proof

It gives us hope for progress in other directions!

Knyazev-Neymeyr (2008) results, p.1/6 Notations

We consider a generalized eigenproblem $B - \mu A$ with symmetric B and SPD A with eigenvalues μ_i enumerated in decreasing order. Let

$$x' = x + \frac{1}{\mu(x) - \mu_{\min}} T(Bx - \mu(x)Ax), \quad \mu(x) = \frac{(x, Bx)}{(x, Ax)}.$$

$B = I$ and $\mu_{\min} = 0$ recover BPK(1996) method with $s = 1$. The preconditioner T is SPD and $\|I - TA\| \leq \gamma < 1$

Knyazev-Neymeyr (2008) results, p.2/6 The main result

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

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

The convergence factor σ cannot be improved with the chosen terms and assumptions and the bound is asymptotically sharp as $\mu(x) \to \mu_i$.

When $B = I$ and $\mu_{\min} = 0$ substitutions $\lambda_i = 1/\mu_i$ and $\lambda(x) = 1/\mu(x)$ give

$$\frac{\lambda(x') - \lambda_i}{\lambda_{i+1} - \lambda(x')} \leq \left(\gamma + (1-\gamma)\frac{\lambda_i}{\lambda_{i+1}}\right)^2 \frac{\lambda(x) - \lambda_i}{\lambda_{i+1} - \lambda(x)},$$

which is a direct improvement of the BPK (1996) bound for $s = 1$.

The proof: reduction to the 2D case in $\operatorname{span}\{x_i, x_{i+1}\}$**!**

Knyazev-Neymeyr (2008) results, p.3/6 The proof

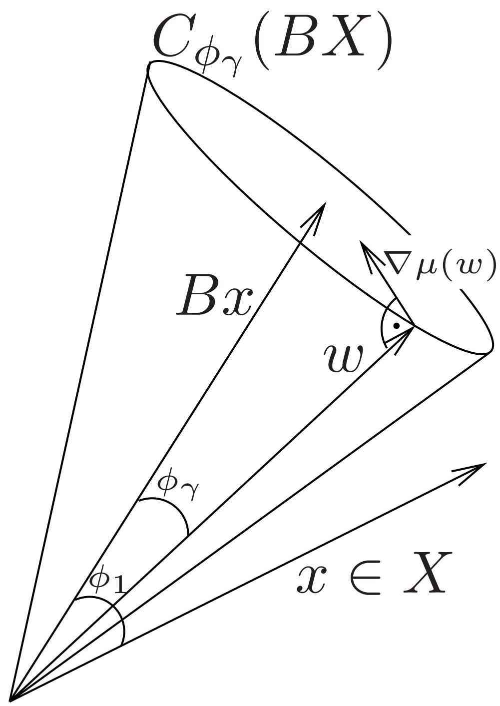
Without loss of generality, $A = I$ and $B > 0$, so $\|I - T\| \le \gamma < 1$ and

$$\mu(x)x' = Bx - (I - T)(Bx - \mu(x)x) \text{ or } X' = [B - (I - T)(B - \mu(X)I)]X.$$

A perturbed power method! Geometry comes to the rescue.

For a fixed one-dimensional subspace X the set of subspaces $\{X'\}$ generated by plugging all SPD preconditioners T with $\|I - T\| \le \gamma < 1$ belongs to (and makes in the real case) a circular cone $C_{\phi_\gamma(X)}(BX) := \{\cup Y : \dim Y = 1, \angle\{Y, BX\} \le \phi_\gamma(X)\}$ with the cone opening angle $\phi_\gamma(X) < \pi/2$ from $\sin \phi_\gamma(X) = \gamma \|Bx - \mu(x)x\| / \|Bx\| < 1$.

Knyazev-Neymeyr (2008) results, p.4/6 The proof



Sharp convergence rate bound = minimizing $\mu(\cdot)$ on this cone.

Knyazev-Neymeyr (2008) results, p.5/6 The proof

Let $\mu_{i+1} < \mu(X) < \mu_i$ and $0 < \gamma < 1$. Define the subspace $W \subseteq \arg\min \mu(C_{\phi_\gamma(X)}(BX))$, dim $W = 1$ and assume that $\mu(W) < \mu_i$, then $W \subset \partial C_{\phi_\gamma(X)}(BX)$ and $\exists\, \alpha = \alpha_\gamma(X) > -\mu_i$ such that $(B + \alpha I)W = BX$.

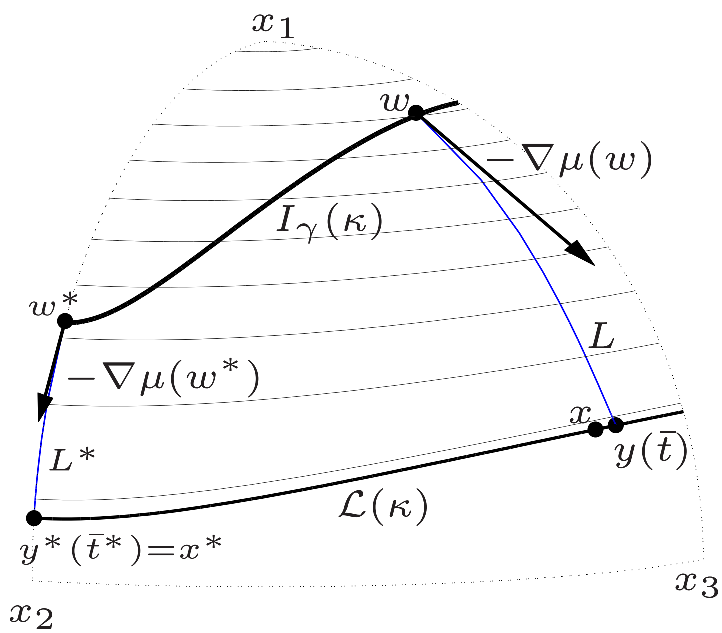
The shift α changes with X and is not known!

The inclusion $X \subset \mathrm{span}\{x_i, x_{i+1}\}$ implies $W \subset \mathrm{span}\{x_i, x_{i+1}\}$.

Let $\kappa \in (\mu_{i+1}, \mu_i)$ be fixed and the level set of the Rayleigh quotient be denoted by $\mathcal{L}(\kappa) := \{X : \dim X = 1, \mu(X) = \kappa\}$. Then $\|\nabla\mu(x)\|$, $x \in X, \|x\| = 1$ is the smallest on $X \in \mathcal{L}(\kappa)$ in $\mathrm{span}\{x_i, x_{i+1}\}$.

$I_\gamma(\kappa) := \{w : w \in \arg\min \mu(C_{\phi_\gamma(X)}(BX));\ X \in \mathcal{L}(\kappa)\}$—minimizers of the Rayleigh quotient. Then $\arg\min \mu(I_\gamma(\kappa)) \in \mathrm{span}\{x_i, x_{i+1}\}$.

Knyazev-Neymeyr (2008) results, p.6/6 The proof



Gradient flow : $y'(t) = -\frac{\nabla\mu(y(t))}{\|\nabla\mu(y(t))\|}$, $t \geq 0$, $y(0) = w \in I_\gamma(\kappa)$.

The total change $\mu(w) - \kappa = \mu(y(0)) - \mu(y(\bar{t})) = \int_0^{\bar{t}} \|\nabla\mu(y(t))\| dt > 0$.

Conclusions

- We present a novel geometric approach to the convergence analysis of a preconditioned fixed-step gradient eigensolver which reduces the derivation of the convergence rate bound to a two-dimensional case.
- The main novelty is in the use of a continuation method for the gradient flow of the Rayleigh quotient to locate the two-dimensional subspace corresponding to the smallest change in the Rayleigh quotient and thus to the slowest convergence of the gradient eigensolver.
- Such an approach may be helpful for analyzing advanced eigensolvers as well as in the theory of gradient iterative methods in smooth optimization connected to dynamical systems.
- The results of BPK (1996) so far remain unbeatable. however.