Steepest descent and conjugate gradient methods with variable preconditioning

Ilya Lashuk (the speaker) and Andrew Knyazev

Department of Mathematics and
Center for Computational Mathematics
University of Colorado at Denver

Supported by National Science Foundation
and Copper Mountain Conference 2006

Outline

1. Introduction
2. Setting the framework
3. The steepest descent convergence rate bound holds for the CG method with variable preconditioning
4. The steepest descent convergence rate bound in terms of the angle between the error and the preconditioned residual
5. What are the possible preconditioned residuals for a given error vector?
6. Main result – the SD convergence rate bound for the CG method with variable preconditioning cannot be improved
7. Conclusions

Introduction

- We study the conjugate gradient (CG) method for linear systems $Ax = b$ with variable preconditioning.
- The preconditioner is SPD on each step and the condition number of the preconditioned system matrix is bounded from above by a constant independent of the step number.
- How fast will this method converge? Can we obtain anything better than the steepest descent (SD) convergence rate?

Setting the framework

We use the following generalization of the CG method: given SPD matrices A, $\{B_k\}$, the right hand side b, the initial guess x_0 and integer parameters $\{m_k\}$, for $k = 0, 1, \ldots$

$$\begin{aligned} r_k &= b - Ax_k, \\ s_k &= B_k^{-1} r_k, \\ p_k &= s_k - \sum_{l=k-m_k}^{k-1} \frac{(As_k, p_l)}{(Ap_l, p_l)} p_l, \\ x_{k+1} &= x_k + \frac{(r_k, p_k)}{(Ap_k, p_k)} p_k. \end{aligned} \tag{1}$$

Integer parameters m_k should additionally satisfy

$$0 \le m_k \le k \text{ and } m_{k+1} \le m_k + 1. \tag{2}$$

Setting the framework (continued)

- Method (1) is known (see, e.g., Axelsson and Vassilevski [1991]) and sometimes referred to as *flexible conjugate gradient method.*
- Condition

  $$0 \leq m_k \leq k \text{ and } m_{k+1} \leq m_k + 1$$

  means that on iteration k we can't orthogonalize to search directions that we didn't orthogonalize to on previous iteration $k-1$ (except for the newly obtained direction p_{k-1}).

Setting the framework (continued)

Depending of the choice of m_k in method (1), we get (for example)

- steepest descent ($m_k = 0$)
- usual CG ($m_k = \min\{k, 1\}$)
- “full orthogonalization” ($m_k = k$)

There’s no need for full orthogonalization, when the preconditioner is SPD and fixed.

Steepest descent convergence rate bound holds for CG with variable preconditioning

The first question one might ask is whether the CG method with variable preconditioning converges at least as good as the SD method. The answer is YES (in our framework) – one can prove that for method (1) with (2) the A-norms of the errors $e_k = x - x_k$ on subsequent iterations satisfy

$$\frac{\|e_{k+1}\|_A}{\|e_k\|_A} \leq \frac{\kappa_{\max} - 1}{\kappa_{\max} + 1}, \tag{3}$$

where $\kappa_{\max}$ is the maximum condition number of $B_k^{-1}A$ (the ratio of the extreme eigenvalues)

SD convergence rate bound for the CG method with variable preconditioning – sketch of the proof

- For SD itself with variable preconditioning the proof is exactly the same as for the SD method with a constant preconditioner, since the bound and the proof are *local*, i.e. involve only the preconditioner on the *current* iteration.

- Now suppose for simplicity that we are doing "full orthogonalization" in (1), i.e., $m_k = k$. This implies that on the k-th iteration we minimize the A-norm of the error e_{k+1} over the plane

  $$e_{k+1} \in e_k + \operatorname{span}\left\{s_k, p_0, \ldots, p_{k-1}\right\},$$

  which obviously contains the line

  $$e_{k+1} \in e_k + \operatorname{span}\left\{s_k\right\}.$$

SD convergence rate is determined by the A-angle between the error vector and the preconditioned residual

- $\angle_A(x,y) = \arccos\left(\frac{(x,y)_A}{\|x\|_A \|y\|_A}\right) \in [0,\pi]$
- Since the “optimal” (minimizing the A-norm of e_{k+1}) step is taken in the direction of the preconditioned residual s_k, the convergence rate on the current iteration is determined by the angle between e_k and s_k. Namely, we have
$$\frac{\|e_{k+1}\|_A}{\|e_k\|_A} = \sin\angle_A(e_k, s_k)$$
- Therefore, because of the standard SD convergence rate bound,

$$\sin\angle_A(e_k, s_k) \le \frac{\kappa_{\max} - 1}{\kappa_{\max} + 1}$$

CG convergence rate on the current iteration

- For SD the A-angle between e_k and s_k is the only thing that matters.
- For CG the angles between s_k and the previous search directions p_l, $l < k$ are also important.
- If s_k is A-orthogonal to all previous search directions p_l, then the current step of CG is just a step of SD (the new search direction p_k is exactly s_k).
- If, in addition, $\angle_A (x, y) = \arcsin \left(\frac{\kappa_{\max} - 1}{\kappa_{\max} + 1} \right)$, then we have

$$\frac{\|e_{k+1}\|_A}{\|e_k\|_A} = \frac{\kappa_{\max} - 1}{\kappa_{\max} + 1} \quad \text{(equality!)}$$

Question: is this situation possible on every iteration?

What are the possible preconditioned residuals for a given error vector?

- We investigate how the preconditioned residual s_k changes, when we apply different preconditioners B_k to the residual r_k.
- We consider SPD preconditioners B_k satisfying $\kappa\left(B_k^{-1}A\right) \leq \kappa_{\max}$ for some $\kappa_{\max} > 1$.
- Since $s_k = B_k^{-1} r_k = B_k^{-1} A e_k$, we can reformulate this problem as follows: given an SPD matrix A, a vector e and a constant $\kappa_{\max} > 1$, describe the set of all vectors s such that there exists an SPD matrix B, satisfying $\kappa\left(B^{-1}A\right) \leq \kappa_{\max}$ and $s = B^{-1}Ae$.

What are the possible preconditioned residuals for a given error vector? (continued)

- We can generalize the problem of describing possible preconditioned residuals: suppose we are given some inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$, a vector $e \in \mathbb{R}^n$ and a constant $\kappa_{\max} > 1$.
- The problem now is to describe the set of all vectors s such that there exists an SPD, with respect to $\langle \cdot, \cdot \rangle$, matrix $C \in \mathbb{R}^{n \times n}$, satisfying $\kappa(C) \leq \kappa_{\max}$ and $Ce = s$.
- This covers the "possible residuals problem": take $\langle \cdot, \cdot \rangle$ to be an A-based inner product. If C is A-SPD, then $B = AC^{-1}$ will be SPD with respect to euclidean inner product – easy to check.

What are the possible preconditioned residuals for a given error vector? (resolution)

Theorem 1 *The set* $\{Cx\}$*, where* x *is a fixed nonzero real vector and* C *runs through all SPD matrices with condition number* $\kappa(C)$ *bounded from above by some* $\kappa_{\max}$*, is a pointed circular cone, specifically,*

$$\{Cx\} = \left\{ y : \angle(x, y) \leq \arcsin\left(\frac{\kappa_{\max} - 1}{\kappa_{\max} + 1}\right) \right\}$$

Main result – the SD convergence rate bound for the CG method with variable preconditioning cannot be improved

Theorem 2 *For any given SPD matrix A, vectors b and x_0, and $\kappa_{\max} > 1$, assuming a matrix size larger than the number of iterations, one can choose a sequence of SPD preconditioners B_k, satisfying $\kappa(B_k^{-1}A) \leq \kappa_{\max}$, such that the method (1) with (2) turns into the SD method, (1) with $m_k = 0$, and for each iteration*

$$\frac{\|e_{k+1}\|_A}{\|e_k\|_A} = \frac{\kappa_{\max} - 1}{\kappa_{\max} + 1} \tag{4}$$

Sketch of the proof – the strategy for choosing the preconditioners

On the 0-th iteration (which is always SD iteration, i.e., $p_0 = s_0$), we pick any preconditioner B_0 (SPD and satisfying condition number bound), such that

$$\sin \angle_A (e_0, s_0) = \frac{\kappa_{\max} - 1}{\kappa_{\max} + 1}.$$

Then we have

$$\frac{\|e_1\|_A}{\|e_0\|_A} = \frac{\kappa_{\max} - 1}{\kappa_{\max} + 1}$$

and $(e_1, s_0)_A = 0$.

Sketch of the proof (continued)

Now suppose that we were able to pick k preconditioners $B_0, \ldots, B_{k-1}$, such that the first $k-1$ iterations were actually SD iterations (i.e., $p_l = s_l$) with the worst possible convergence rate, and e_k is A-orthogonal to the previous search directions $p_l = s_l,\ l < k$. Then we can pick any vector q_k, such that $(q_k, p_l)_A = 0,\ l < k$ and q_k and e_k are linearly independent. We choose such B_k that $s_k \in \operatorname{span}\{e_1, q_1\}$ and

$$\sin \angle_A (e_k, s_k) = \frac{\kappa_{\max} - 1}{\kappa_{\max} + 1}.$$

Sketch of the proof (continued)

We then have $(s_k, p_l)_A = 0,\ l < k$, which implies $p_k = s_k$, so this iteration will again be an SD step. Finally, because of the A-angle between s_k and e_k, we have

$$\frac{\|e_1\|_A}{\|e_0\|_A} = \frac{\kappa_{\max} - 1}{\kappa_{\max} + 1}.$$

We also have $(e_{k+1}, p_l)_A = 0,\ l \leq k$ (this is important for choosing B_{k+1}). Indeed, $e_{k+1} \in \operatorname{span}\{e_k, p_k\} \perp_A p_l,\ l < k$. The orthogonality relation $(e_{k+1}, p_k)_A = 0$ holds because the "optimal" step is made in the direction $p_k = s_k$.

Conclusions

- Presented result basically shows that although the assumptions we have made look reasonable, they are insufficient.
- To prove better convergence properties, one needs to assert more about preconditioners B_k.
- This means preconditioners used on different iterations must be related to each other. One possibility is to assume that they all approximate some fixed preconditioner, see, e.g., Golub and Ye [1999/00] and Notay [2000].

References

O. Axelsson and P. S. Vassilevski. A black box generalized conjugate gradient solver with inner iterations and variable-step preconditioning. *SIAM J. Matrix Anal. Appl.*, 12(4):625–644, 1991. ISSN 0895-4798.

Gene H. Golub and Qiang Ye. Inexact preconditioned conjugate gradient method with inner-outer iteration. *SIAM J. Sci. Comput.*, 21(4):1305–1320 (electronic), 1999/00. ISSN 1064-8275.

Yvan Notay. Flexible conjugate gradients. *SIAM J. Sci. Comput.*, 22(4):1444–1460 (electronic), 2000. ISSN 1064-8275.