CHICAGO

5.4 CONJUGATE GRADIENT CONVERGENCE/ PRECONDITIONING

Consequences of using a Krylov space: matrix polynomial formulation

- Iteration in Krylov Space

$$\begin{aligned} x_{k+1} &= x_0 + \alpha_0 p_0 + \cdots + \alpha_k p_k \\ &= x_0 + \gamma_0 r_0 + \gamma_1 A r_0 + \cdots + \gamma_k A^k r_0, \end{aligned}$$

- Matrix Polynomial

$$P_k^*(A) = \gamma_0 I + \gamma_1 A + \cdots + \gamma_k A^k,$$

- Iteration as a matrix Polynomial

$$x_{k+1} = x_0 + P_k^*(A) r_0.$$

Error in A-space

- Error in A-norm

$$\|z\|_A^2 = z^T A z. \qquad \tfrac{1}{2}\|x - x^*\|_A^2 = \tfrac{1}{2}(x - x^*)^T A (x - x^*) = \phi(x) - \phi(x^*).$$

- So what is the conjugate gradient method computing?

- Another form of the error

$$x_{k+1} = x_0 + P_k^*(A) r_0. \qquad \min_{P_k} \|x_0 + P_k(A) r_0 - x^*\|_A.$$

$$x_{k+1} - x^* = x_0 + P_k^*(A) r_0 - x^* = [I + P_k^*(A) A](x_0 - x^*).$$

The calculation in eigenvalue space

Let $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of A, and let $v_1, v_2, \ldots, v_n$ be the corresponding orthonormal eigenvectors, so that

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^T .$$

$i=1$

$$\|x_{k+1} - x^*\|_A^2 = \min_{P_k} \sum_{i=1}^{n} \lambda_i [1 + \lambda_i P_k(\lambda_i)]^2 \xi_i^2 .$$

$$\begin{aligned} \|x_{k+1} - x^*\|_A^2 &\leq \min_{P_k} \max_{1 \leq i \leq n} [1 + \lambda_i P_k(\lambda_i)]^2 \left(\sum_{j=1}^{n} \lambda_j \xi_j^2 \right) \\ &= \min_{P_k} \max_{1 \leq i \leq n} [1 + \lambda_i P_k(\lambda_i)]^2 \|x_0 - x^*\|_A^2 , \end{aligned}$$

Consequences for Convergence

- Linear Convergence Rate Estimate: $\min_{P_k} \max_{1 \le i \le n} [1 + \lambda_i P_k(\lambda_i)]^2.$
- Consequences:

Theorem 5.4.

If A has only r distinct eigenvalues, then the CG iteration will terminate at the solution in at most r iterations.

Theorem 5.5.

If A has eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, we have that

$$\|x_{k+1} - x^*\|_A^2 \le \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)^2 \|x_0 - x^*\|_A^2.$$

- Note: finite termination in n steps.

5.2.2 PRECONDITIONGING

CG: PRACTICAL VERSION (MINIMAL STORAGE)

Algorithm 5.2 (CG).

Given x_0;

Set $r_0 \leftarrow Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$;

while $r_k \neq 0$

$$\alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T A p_k};$$
$$x_{k+1} \leftarrow x_k + \alpha_k p_k;$$
$$r_{k+1} \leftarrow r_k + \alpha_k A p_k;$$
$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};$$
$$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k;$$
$$k \leftarrow k + 1;$$

end (while)

NEEDS ONLY 1 MATRIX-VECTOR MULTIPLICATION PER STEP.
AX NEVER FORMED AS BEFORE.

Acceleration of Conjugate Gradient

- Rescaling of the problem

$$\hat{x} = Cx.$$

- The modified objective function

$$\hat{\phi}(\hat{x}) = \tfrac{1}{2}\hat{x}^T (C^{-T} A C^{-1})\hat{x} - (C^{-T} b)^T \hat{x}.$$

- Equivalent linear system.

$$(C^{-T} A C^{-1})\hat{x} = C^{-T} b,$$

Consequences for Convergence

- Linear Convergence Rate Estimate: $\min_{P_k} \max_{1 \le i \le n} [1 + \lambda_i P_k(\lambda_i)]^2.$
- Consequences:

Theorem 5.4.
If A has only r distinct eigenvalues, then the CG iteration will terminate at the solution in at most r iterations.

Theorem 5.5.
If A has eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, we have that

$$\|x_{k+1} - x^*\|_A^2 \le \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)^2 \|x_0 - x^*\|_A^2.$$

- Note: finite termination in n steps.

CHICAGO

How to find a preconditioner?

- Idea (from Theorem 5.5). Compute a C such that the eigenvalues are "clustered", then convergence is fast. For example

$$C^{-T} A C \approx I; \quad \text{or} \quad C \approx L^T; A = LL^T$$

- Preconditioners must be easy to factorize or invert.
- Example preconditioners:
  - Incomplete Cholesky (use sparsity pattern of A)
  - Symmetric Successive overrelaxation
  - Multigrid (Order (1) for PDEs)
- The Holy Grail: Condition number is O(1).

Preconditioned conjugate gradient

Algorithm 5.3 (Preconditioned CG).

Given x_0, preconditioner M;
Set $r_0 \leftarrow Ax_0 - b$;
Solve $My_0 = r_0$ for y_0;
Set $p_0 = -y_0$, $k \leftarrow 0$;
while $r_k \neq 0$

$$\alpha_k \leftarrow \frac{r_k^T y_k}{p_k^T A p_k};$$
$$x_{k+1} \leftarrow x_k + \alpha_k p_k;$$
$$r_{k+1} \leftarrow r_k + \alpha_k A p_k;$$
$$\text{Solve } M y_{k+1} = r_{k+1};$$
$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T y_{k+1}}{r_k^T y_k};$$
$$p_{k+1} \leftarrow -y_{k+1} + \beta_{k+1} p_k;$$
$$k \leftarrow k+1;$$

end (while)

Preconditioner action

$$M = C^T C$$

5.5 NONLINEAR CONJUGATE GRADIENT

The Fletcher-Reeves method

Algorithm 5.4 (FR).

Given x_0;

Evaluate $f_0 = f(x_0)$, $\nabla f_0 = \nabla f(x_0)$;

Set $p_0 \leftarrow -\nabla f_0$, $k \leftarrow 0$;

while $\nabla f_k \neq 0$

Compute α_k and set $x_{k+1} = x_k + \alpha_k p_k$;

Evaluate ∇f_{k+1};

$$\beta_{k+1}^{FR} \leftarrow \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k};$$

$$p_{k+1} \leftarrow -\nabla f_{k+1} + \beta_{k+1}^{FR} p_k;$$

$$k \leftarrow k+1;$$

end (while)

Actual line search

$$\alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T A p_k};$$

$$x_{k+1} \leftarrow x_k + \alpha_k p_k;$$

$$r_{k+1} \leftarrow r_k + \alpha_k A p_k;$$

$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};$$

$$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k;$$

$$k \leftarrow k+1;$$

$$r_k = \nabla f(x_k)$$

The line search method

- Use a search that satisfies the Wolfe Conditions.

$$f(x_k + \alpha_k p_k) \le f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$
$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \le -c_2 \nabla f_k^T p_k,$$

- Use the parameters

$$0 < c_1 < c_2 < \tfrac{1}{2}.$$

Convergence of Fletcher-Reeves

Lemma 5.6.

Suppose that Algorithm 5.4 is implemented with a step length α_k that satisfies the strong Wolfe conditions (5.43) with $0 < c_2 < \frac{1}{2}$. Then the method generates descent directions p_k that satisfy the following inequalities:

$$-\frac{1}{1-c_2} \le \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \le \frac{2c_2 - 1}{1-c_2}, \quad \text{for all } k = 0, 1, \ldots . \tag{5.53}$$

Fixed fraction of steepest descent

Assumptions 5.1.

(i) *The level set $\mathcal{L} := \{x \mid f(x) \le f(x_0)\}$ is bounded;*

(ii) *In some open neighborhood $\mathcal{N}$ of $\mathcal{L}$, the objective function f is Lipschitz continuously differentiable.*

Theorem 5.7 (Al-Baali [3]).

Suppose that Assumptions 5.1 hold, and that Algorithm 5.4 is implemented with a line search that satisfies the strong Wolfe conditions (5.43), with $0 < c_1 < c_2 < \frac{1}{2}$. Then

$$\liminf_{k \to \infty} \|\nabla f_k\| = 0. \tag{5.63}$$