Lecture 20
The Conjugate Gradients Algorithm II

MIT 18.335J / 6.337J

Introduction to Numerical Methods

Per-Olof Persson

November 21, 2006

Optimization by Conjugate Gradients

- We know that solving $Ax = b$ is equivalent to minimizing the quadratic function $\varphi(x) = \frac{1}{2}x^T Ax - x^T b$
- The minimization can be done by *line searches*, where $\varphi(x_n)$ is minimized along a *search direction* p_n
- The α_{n+1} that minimizes $\varphi(x_n + \alpha_{n+1} p_n)$ is

$$\alpha_{n+1} = \frac{p_n^T r_n}{p_n^T A p_n}$$

  with the residual $r_n = b - Ax_n$

  Proof. Black-board

- The residual is also minus the gradient of $\varphi(x_n)$:

$$\varphi'(x_n) = Ax_n - b = -r_n$$

The Method of Steepest Descent

- Very simple approach: Set search direction p_n to the negative gradient r_n
- Corresponds to moving in the direction $\varphi(x)$ changes the most

Algorithm: Steepest Descent

$x_0 = 0, r_0 = b$

for $k = 1, 2, 3, \ldots$

$\alpha_n = (r_{n-1}^T r_{n-1})/(r_{n-1}^T A r_{n-1})$	step length
$x_n = x_{n-1} + \alpha_n p_{n-1}$	approximate solution
$r_n = r_{n-1} - \alpha_n A p_{n-1}$	residual

- Poor convergence, tends to move along previous search directions

The Method of Conjugate Directions

- The optimization procedure can be improved by better search directions
- Let the search direction be *A-conjugate*, or $p_i^T A p_k = 0$
- Then the algorithm will converge in at most m steps, since the initial error can be decomposed along the p's:

$$e_0 = \sum_{n=0}^{m-1} \delta_n p_n, \quad \text{with} \quad \delta_n = \frac{p_n^T A e_0}{p_n^T A p_n}$$

But this is exactly the α we choose at step n:

$$\alpha_{n+1} = \frac{p_n^T r_n}{p_n^T A p_n} = \frac{p_n^T A e_n}{p_n^T A p_n} = \frac{p_n^T A e_0}{p_n^T A p_n}$$

since the error e_n is the initial e_0 plus a combination of $p_0, \ldots, p_{n-1}$, which are all A-conjugate to p_n. Each component δ_n is then subtracted out at step n, and the method converges after m steps.

Choosing A-conjugate Search Directions

- One method to choose p_n which is A-conjugate to previous search vectors is by Gram-Schmidt:

$$p_n = p_n^0 - \sum_{k=0}^{n-1} \beta_{nk} p_k, \quad \text{with} \quad \beta_{nk} = \frac{{p_n^0}^T A p_k}{p_k^T A p_k}$$

- The initial p_n^0 vectors should be linearly independent, for example column $n + 1$ of identity matrix
- Drawback: Must store all previous search vectors p_n
- Conjugate Gradients is simply Conjugate Directions with a particular initial vector in Gram-Schmidt: $p_n^0 = r_n$
- This gives orthogonal residuals $r_n^T r_j = 0$ for $j \neq n$, and $\beta_{nk} = 0$ for $n > k + 1$ (proof omitted)