MathGrp

L#11

Oct 4, 2011
Tuesday

The Conjugate Gradient Method for $Ax=b$

Steepest descent method suffers from one very unpleasant and potentially slowing down flaw: $r^{(k+1)} \perp r^{(k)}$, but often $r^{(k+1)}$ is the same as $r^{(k-1)}$, i.e. the search direction is repeated and often many times. Conjugate gradient is designed to avoid this. Here is how it does it.

Introduce the Krylov space

$$K_m(A,r) = \{r, Ar, A^2r, \dots, A^{m-1}r\}$$, r given (often $r = r^{(0)}$ - the initial residual)

We assume that the vectors in $K_m(A,r)$ are linearly independent. Therefore for $m=n$ $K_n(A,r) = \mathbb{R}^n$.

Then we minimize

$$(1) \qquad J(y) = \frac{1}{2}(A(x-y), x-y) - \frac{1}{2}(Ax, x)$$

over the Krylov subspaces. Namely, for given $x^{(0)} \in \mathbb{R}^n$ we define $r^{(0)} = b - Ax^{(0)}$ and $K_m = K_m(A, r^{(0)})$. The k-th CG iterate $x^{(m)}$ has the form

$$x^{(m)} = x^{(0)} + \theta^{(m)}$$, where $\theta^{(m)} \in K_m$ is

the minimizer or alltogether

$$J(x^{(m)}) = \min_{y \in \{x^{(0)}\} + K_m} J(y)$$

that is $\quad J(x^{(m)}) \le J(y)$ for any $y \in \{x^{0}\} + K_m$

i.e. $y = x^{(0)} + \sum_{\ell=0}^{m-1} \alpha_\ell A^\ell r \equiv x^{(0)} + \theta^{(m)}$

Then making the computations we get

$$\frac{1}{2}(A\underbrace{(x-x^{(m)})}_{-e^{(m)}}, x-x^{(m)}) - \frac{1}{2}(Ax,x) \le \frac{1}{2}(A(x-y),(x-y)) - \frac{1}{2}(Ax,x)$$

$\boxed{e^{(m)} = x - x^{(m)}}$ for any $y = x^{(0)} + \theta^{(m)}$

$$(Ae^{(m)}, e^{(m)}) \le (A(\underbrace{x-x^{(0)}}_{e^{0}} - \theta^{(m)}), \underbrace{x-x^{(0)}}_{e^{0}} - \theta^{(m)})$$

$$\boxed{(Ae^{(m)}, e^{(m)}) \le (A(e^{(0)} - \theta^{(m)}), e^{(0)} - \theta^{(m)}) \quad \text{for any } \theta^{m} \in K_m(A, r^{(0)})}$$

The main question now is how to find the minimizer $\theta^{(m)} \in K_m$.

Make one more hypothetical move: what about if we have an A-orthogonal basis of $K_m(A, r^{(0)})$?

say

$$\text{span}\{r^{(0)}, Ar^{(0)}, \dots, A^{m-1}r^{(0)}\} = \text{span}\{p^{(0)}, p^{(1)}, \dots, p^{(m-1)}\}$$

such that $(Ap^{(i)}, p^{(j)}) = 0 \quad i \ne j$

(we can think of as an application of Gram-Schmidt orthogonalization procedure applied to $r^{(0)}, Ar^{(0)}, \dots$

The idea of CG from a point of view minimization over a space spanned by the vectors

$$\{p^{(0)}, p^{(1)}, \dots, p^{(n-1)}\} \in \mathbb{R}^n$$

which are (by miracle) A-orthogonal, i.e.

$$(Ap^{(j)}, p^{(i)}) = 0 \quad i \neq j$$

Of course $(Ap^{(k)}, p^{(k)}) > 0$ since A is an SPD.

For an arbitrary given initial guess $x^{(0)}$ we can write the solution x of $Ax=b$ in the form

$$(*) \quad \boxed{x = x^{(0)} \pm \sum_{l=0}^{n-1} \alpha_l p^{(l)} = x^{(0)} + t^{n}}$$

this would be an exact representation of the solution

Therefore we have

$$Ax = Ax^{(0)} \pm \sum_{l=0}^{n-1} \alpha_l A p^{(l)} = b$$

Since $p^{(l)}$ are A-orthogonal we have

$$\pm \sum \alpha_l Ap^{(l)} = b - Ax^{(0)}$$

$$\pm \alpha_l (Ap^{(l)}, p^{(l)}) = (b - Ax^{(0)}, p^{(l)})$$

$$\alpha_l = \frac{(-Ax^{(0)}+b, p^{(l)})}{(Ap^{(l)}, p^{(l)})} \qquad l = 0,1,\dots,n-1$$

By construction we shall define $x^{(m)}$ as a vector spanned by the first k vectors $p^{(m)}$, that is

$$x^{(m)} = x^0 + \sum_{\ell=0}^{m-1} \alpha_\ell p^\ell \quad , \text{ i.e. } \theta^{(m)} = \sum_{\ell=0}^{m-1} \alpha_\ell p^\ell$$

By the same argument as before we have

$$x^{(m+1)} = x^{(m)} + \alpha_m p^{(m)} , \qquad m = 0, 1, \dots, n-1$$

where
$$\alpha_m = \frac{-(Ax^{(0)} - b, p^{(m)})}{(Ap^{(\kappa)}, p^{(\kappa)})}$$

$$\alpha_m = -\frac{(Ax^0 - \sum_{\ell=0}^{\kappa-1} \alpha_\ell A p^{(\ell)} - b, p^{(m)})}{(Ap^{(\kappa)}, p^{(\kappa)})}$$

Note $\sum_{\ell=0}^{m-1} \alpha_\ell \underbrace{(Ap^{(\ell)}, p^{(m)})}_{=0} = 0$

$$A \underbrace{(x^0 - \sum_{\ell=0}^{m-1} \alpha_\ell p^{(\ell)})}_{x^{(\kappa)}} \qquad \boxed{r^{(m)} = b - Ax^{(m)}}$$

$$\boxed{\alpha_\kappa = -\frac{(Ax^{(m)} - b, p^{(m)})}{(Ap^{(m)}, p^{(m)})} = \frac{(r^{(m)}, p^{(m)})}{(Ap^{(m)}, p^{(m)})}}$$

Also

$$r^{(m+1)} = b - Ax^{(m+1)} = b - Ax^{(m)} - \alpha_m A p^{(m)} = r^{(m)} - \alpha_m A p^{(m)}$$

$$\boxed{r^{(m+1)} = r^{(m)} - \alpha_m p^{(m)}}$$

The magic of CG is that we do not need all $p^{(\ell)}$ from the very beginning and they are generated in the process of computing $x^{(m)}$.

The main point of CG is that for an approximation of x we need much less than n-terms of the exact representation.

The CG method for $Ax=b$

$r^0 = b - Ax^{(0)}$ x^0 – given

$p^0 = r^0$

for $m = 0, 1, \ldots$ until convergence, compute

$$x^{m+1} = x^m + \alpha_m p^m \qquad \alpha_m = \frac{(r^m, p^m)}{(Ap^m, p^m)} \overset{1}{=} \frac{(r^m, r^m)}{(Ap^m, p^m)}$$

$$r^{m+1} = r^m - \alpha_m A p^m$$

$$p^{m+1} = r^{m+1} + \beta_m p^{(m)} \qquad \beta_m = -\frac{(Ap^m, r^{m+1})}{(Ap^m, p^m)} \overset{2}{=} \frac{(r^{m+1}, r^{m+1})}{(r^m, r^m)}$$

end m

The most important property of the vectors $p^0, p^1, \ldots, p^m$ is that they are A-orthogonal.
Indeed for $m+1$ & m we have

$$(Ap^{m+1}, p^m) = (Ar^{m+1}, p^m) + \beta_m (Ap^m, p^m) = 0$$ because of the choice of β_m.

It is valid for other pairs as well.

Steepest Descent Method

$x^{(0)}$ - given $r^0 = b - Ax^0$

$$x^{k+1} = x^k + \alpha_k r^k \qquad \alpha_k = \frac{(r^k, r^k)}{(Ar^k, r^k)}$$

$$r^{k+1} = r^k - \alpha_k A r^k$$

$$(r^{k+1}, r^k) = (r^k, r^k) - \alpha_k (Ar^k, r^k) = 0$$

$$r^{k+1} \perp r^k$$

but it could be parallel to r^{k-1}
so we have repetition of search directions

Gram-Schmidt Method

$$\{r^0, r^1, r^2, \ldots, r^{m-1}\} \to \{p^0, p^1, \ldots, p^{m-1}\}$$
A-orthogonal

$$p^0 = r^0$$

$$p^{k+1} = r^{k+1} - \sum_{\ell=0}^{k} \beta_{k\ell} p^\ell \qquad k = 0, 1, \ldots, m-2$$

where $\beta_{k\ell} = \dfrac{(Ar^{k+1}, p^\ell)}{(Ap^\ell, p^\ell)}$

$$(Ap^\ell, p^j) = 0 \qquad \ell \neq j$$

We get the vectors $\{p^0, p^{(1)}, \dots, p^{m-1}\}$
by Gram-Schmidt orthogonalization method
from

$$\text{span}\{r^0, Ar^0, \dots, A^{m-1}r^0\}$$
$$= \text{span}\{r^0, r^1, \dots, r^{m-1}\}$$

our w^j are r^j!

$$\boxed{\begin{aligned} & p^0 = r^0 \\ & p^{m+1} = r^{m+1} - \sum_{\ell=0}^{m} \beta_{m\ell}\, p^\ell \qquad \beta_{m\ell} = \frac{(A r^{(m+1)}, p^{(\ell)})}{(Ap^{(\ell)}, p^{(\ell)})} \\ & \text{Gram-Schmidt} \end{aligned}}$$

Obviously $x^{(1)} = x^{(0)} - \alpha_0 p^0 = r^0 - \alpha_0 r^0$

$$Ax^{(1)} - b = Ax^{(0)} - b - \alpha_0 Ar^0$$
$$-r^{(1)} = -r^{(0)} - \alpha Ar^{(0)}$$

$r^{(1)}$ in $\text{span}\{r^0, Ar^{(0)}\}$

$$r^{(2)} = x^{(1)} - \alpha_1 p^{(1)}$$

$$p^{(1)} \in \text{span}\{r_0^{(0)}, Ar^{(0)}\}$$

$$-r^{(2)} = -r^{(1)} - \alpha_1 Ap^{(1)}$$

$$r^{(1)} \in \text{span}\{r^0, Ar^{(0)}\}$$
$$Ap^{(1)} \in \text{span}\{Ar^0, A^2 r^{(0)}\}$$

$$r^{(2)} \in \text{span}\{r^0, Ar^{(0)}, A^2 r^{(0)}\}$$

Let us have a closer look at Gram-Schmidt and CG process defined above.

First, we prove that $r^{m+1} \perp p^{\ell} \quad \ell = 0, 1, \dots, m$

Indeed

$$(r^{m+1}, p^m) = (r^m, p^m) - \alpha_m (Ap^m, p^m)$$

$$= (r^m, p^m) - \frac{(r^m, p^m)}{(Ap^m, p^m)} (Ap^m, p^m) = 0$$

What about (r^{m+1}, p^{m-1}) ?

$$(r^{m+1}, p^{m-1}) = \underbrace{(r^m, p^{m-1})}_{\text{by the previous}} - \alpha_m \underbrace{(Ap^m, p^{m-1})}_{=0 \text{ A-orthogonality}}$$

Thus $(r^{m+1}, p^{\ell}) = 0 \quad \ell = 0, 1, \dots, m$

Let us look at

$$(r^m, p^m) = (r^m, r^m) + \beta_m \underbrace{(r^m, p^{m-1})}_{=0} = (r^m, r^m)$$

Therefore $\alpha_m = \dfrac{(r^m, r^m)}{(Ap^m, p^m)} > 0$

When $\alpha_m = 0$? $r^m = 0 \Rightarrow Ax^m - b = 0$

we have found the solution.

Final consideration

$$r^{\ell+1} = r^{\ell} - \alpha_\ell A p^{\ell}$$

$$A p^{\ell} = \frac{1}{\alpha_\ell}(r^{\ell} - r^{\ell+1}) \qquad \ell = 0, \dots, m$$

$$(r^{m+1}, Ap^{\ell}) = \frac{1}{\alpha_\ell}(r^{m+1}, r^{\ell} - r^{\ell+1})$$

$$(Ar^{m+1}, p^{\ell}) = \begin{cases} \dfrac{-1}{\alpha_m}(r^{m+1}, r^{m+1}) & \ell = m \\ 0 & \ell < m \end{cases}$$

But remember from Gram-Schmidt $\beta_{m\ell} = -\dfrac{(Ar^{m+1}, p^{\ell})}{(Ap^{\ell}, p^{\ell})} = \begin{cases} \beta_m & \ell = m \\ 0 \end{cases}$

$\Downarrow$

$$p^{m+1} = r^{m+1} - \beta_m p^m$$

Moreover

$$\beta_m = -\frac{(Ar^{m+1}, p^m)}{(Ap^m, p^m)} = \frac{1}{\underbrace{\alpha_m}_{\frac{(r^m, r^m)}{(Ap^m, p^m)}}} \frac{(r^{m+1}, r^{m+1})}{(Ap^m, p^m)} = \frac{(r^{m+1}, r^{m+1})}{(r^m, r^m)}$$