Conjugate Gradient

Thursday, March 12, 2020 7:46 AM

In all iterative algorithms so far we find $x^{(k)}$ in the span of $b, Ab, \ldots, A^{(k-1)}b$.

e.g. Richardson, Chebychev.

So what is the best "such" iteration?

$$x^{(k+1)} = \underset{x \in \text{Span}\{b, \ldots, A^{(k-1)}b\}}{\text{argmin}} \|x - x^*\|_A \qquad Ax^* = b.$$

$$\|x-x^*\|_A^2 = x^TAx + \underbrace{x^{*T}Ax^*}_{\text{fixed.}} - 2x^TAx^*$$

So same as $\min_x \frac{1}{2} x^TAx - b^Tx. = f(x).$

$\nabla f(x) = Ax - b$

$\nabla^2 f(x) = A \succcurlyeq 0.$

$$x^{(k)} = \underset{x \in \mathcal{K}_k}{\text{argmin}} \frac{1}{2} x^TAx - b^Tx.$$

Let $u^0, u^1, \ldots\ldots u^{(k)}$ be a basis of $\mathcal{K}_{k+1}$.

Let $u^0, u^1, \ldots \ldots u$

(not orthogonal)

$$x^{(k+1)} = \sum_{i=0}^{k} \alpha_i u^{(i)}$$

Then $f(x^{(k+1)}) = \frac{1}{2}\left(\sum_{i=0}^{k} \alpha_i u^i\right)^T A \left(\sum_{i=0}^{k} \alpha_i u^i\right) - b^T \left(\sum_{i=0}^{k} \alpha_i u^i\right)$

$$= \frac{1}{2} \sum \alpha_i^2 u^{i^T} A u^i - \sum \alpha_i b^T u^i + \frac{1}{2} \sum_{i \neq j} \alpha_i \alpha_j u^{i^T} A u^j.$$

We will choose u^i s.t. $u^{i^T} A u^j = 0 \quad \forall i \neq j$.

So to min $f(x^{(k+1)})$, we set $\nabla_\alpha f(x^{(k+1)}) = 0$

i.e. $\alpha_i = \dfrac{b^T u^i}{u^{i^T} A u^i}$.

$u^0 = b$

$u^1 = A u^0 - \dfrac{(A u^0)^T A u^0}{u^{0^T} A u^0} \cdot u^0$ s.t. $u^{1^T} A u^0 = 0$.

$$u^{(k+1)} = A u^k - \sum_{i=0}^{k} \frac{(A u^k)^T A u^i}{u^{i^T} A u^i} u^i$$

$$u - Au - \sum_{i=0} \frac{\cdots}{u^{iT} A u^i} u$$

Then

$$u^{(k+1)^T} A u^j = u^{k^T} A^2 u^j - \frac{u^{k^T} A^2 u^j}{u^{j^T} A u^j} u^{j^T} A u^j = 0.$$

$$\left(\forall j < k,\ u^{k^T} A u^j = 0\right)$$

using above again,

$$\boxed{u^{(k+1)} = Au^{(k)} - \frac{u^{k^T} A^2 u^k}{u^k A u^k} u^k - \frac{u^{k^T} A^2 u^{k-1}}{u^{k-1^T} A u^{k-1}} u^{k-1}}.$$

This is because

$$(Au^k)^T Au^i = 0 \quad \forall i < k-1$$

since $Au^i \in \operatorname{Span}\{u^0 \ldots u^{i+1}\} \perp u^k$.

$$x^{(k+1)} = \sum_{i=0}^{k} \frac{b^T u^i}{u^{i^T} A u^i} \cdot u^i$$

we need $O(k)$ Av operations $+ O(nk)$.

in fact, only need $Au^0, \ldots Au^k$

Recall that $\|x^{(k)} - x^*\| \le \varepsilon \|x^*\|_A$

Recall that $\|x^{(k)} - x^*\|_A \leq \varepsilon \|x^*\|_A$

after $k = O\left(\sqrt{K} \log \frac{1}{\varepsilon}\right)$ iterations since $\exists$ polynomial giving this approx (Chebychev)

and CG gets best approximation in A-norm.

Thm. # iterations $\leq$ # distinct eigenvalues of A.

Pf. $q(x) = \dfrac{\prod_{i=1}^{d} (\lambda_i - x)}{\prod \lambda_i}$ $\qquad q(0) = 1$

$\qquad q(\lambda_i) = 0$.

So after d. iterations $\|x - x^*\|_A = 0$.

This description avoids numerical issues.

u^i could blow up.

But computation can be rearranged to avoid this.
