Conjugate Gradient Iteration

Obervation 1. $Ax = b \Leftarrow \min_{x \in R^n} \underbrace{\frac{1}{2} x^T Ax - x^T b}_{\phi(x)}$

Observation 2. $\min_{x \in S \subset R^N} \phi(x) = \min_{x \in S \subset R^N} \left\| x - x^* \right\|_A$

Ideas:

1. Suppose we have found a search-direction P_{k+1}, at x^k

   we want to $\min_{\alpha} \phi(x^k + \alpha P_{k+1})$

$$\Rightarrow \frac{d}{d\alpha} \phi(x^k + \alpha P_{k+1}) = 0 \quad \Rightarrow \alpha_{k+1} = \frac{P_{k+1}^T r_k}{P_{k+1}^T A P_{k+1}}$$

2. Suppose $x^k \in \underbrace{\{r_0, Ar_0, A^2 r_0, \dots A^{k\text{-}1} r_0\}}_{K_k} + x_0$

   For any $J \in K_k$,

$$\frac{d}{dt} \phi(x_k + tJ) = \nabla \phi(x_k + tJ)^T \cdot J = 0 \quad at \ t = 0$$

$$\Rightarrow \nabla \phi(x_k)^T \cdot J = (b - Ax_k)^T \cdot J = -r_k^T \cdot J = 0$$

Since r_l, $l < k$, all in K_k, we have $r_l \cdot r_k = 0$, $l < k$

2. What search direction P_{k+1} should be looked for?

From $\begin{cases} r_{k+1} = b - Ax_{k+1} \\ r_k = b - Ax_k \end{cases} \Rightarrow A(x_{k+1} - x_k) = r_k - r_{k+1}$

$\Rightarrow$ Consider $x_{k+1} = x_k + \alpha_{k+1} P_{k+1}$

where $x_k \in K_k$ and $x_{k+1},\ P_{k+1} \in K_{k+1},$ we have

(*) $\alpha_k (JAP_{k+1}) = J(r_k - r_{k+1}) = 0 \ for\ all\ J \in K_k$

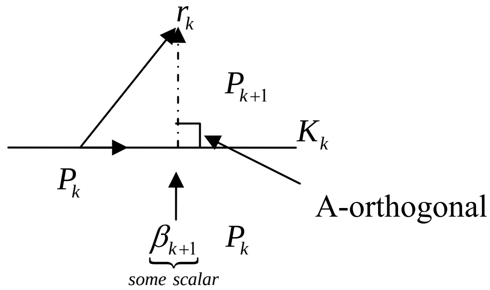
$\Rightarrow$ One should look for P_{k+1} in the subspace that is

$A\text{-}orthogonal\ to\ K_k$ (i.e $P_{k+1}^T AJ = 0$ for all $J \in K_k$)

4. Recall that $r_k \in K_{k+1}$.

To build an iterative updating algorithm for P_{k+1}, it is natural to consider (assuming we can find P_k that is A-orthogonal to K_{k-1})

(**) $P_{k+1} = r_k - \beta_{k+1} P_k$



To satisfy (*), we have

$$P_{k+1}^T A \underbrace{r_{l,0\le l\le k-1}}_{\in K_k} = 0 \quad \Rightarrow (r_k - \beta_{k+1} P_k)^T A r_l = 0$$

$$\left.\begin{array}{l} For\ l\ <\ k\text{-}1\ \Rightarrow Ar_l \in K_{k-1}\ \ \Rightarrow r_k^T A r_l = 0 \\ Moreover,\ by\ assumption,\ P_k^T A \underbrace{r_l}_{\in K_{\text{k-1}}} = 0 \end{array}\right\} \Rightarrow$$

$$P_{k+1}^T A r_l = 0,\ 0 \le l < k-1$$

So we only needs to determine β_{k+1} such that

$$r_k^T A r_{k-1} - \beta_{k+1} P_k^T A r_{k-1} = 0$$

$$\Rightarrow \beta_{k+1} = r_k^T A r_{k-1} / P_k^T A r_{k-1} \quad \text{provided } P_k^T A r_{k-1} \ne 0$$

$$\left(\begin{array}{l} Since\ r_k = b - Ax_k = b - A(x_{k-1} + \alpha_k P_k) = r_{k-1} - \alpha A P_k \\ \quad r_k^T \cdot r_{k-1} = \underbrace{\|r_{k-1}\|^2}_{\ne 0} - \alpha P_k^T A r_{k-1} = 0 \\ \Rightarrow P_k^T A r_{k-1} \ne 0\ and\ P_k^T A r_{k-1} = \dfrac{\|r_{k-1}\|^2}{\alpha}\ otherwise\ condradict \end{array}\right)$$

To tie the end, we only need to show

$\exists P_2$ that is A-orthogonal to $K_1 = r_0$

provided $\begin{cases} P_1 = r_0 \\ x_1 = x_0 + \alpha_1 \cdot P_1 \\ r_1 = b - Ax_1 \end{cases}$

Consider $P_2 = r_1 - (\underbrace{\frac{r_1^T A r_0}{P_1^T A r_0}}_{=P_2}) P_1$

one has $P_2^T A r_0 = (r_1^T - \frac{r_1^T A r_0}{P_1^T A r_0} \cdot r_0^T) A r_0$

$$= r_1^T A r_0 - r_1^T A r_0 = 0$$

Hence result.

5. Let's re-examine the quantities α_{k+1} *and* β_{k+1}

$$\alpha_{k+1} = \frac{P_{k+1}^T r_k}{P_{k+1}^T A P_{k+1}} = \frac{(r_k + \beta_{k+1} P_k)^T r_k}{P_{k+1}^T A P_{k+1}} \overset{(1)}{=} \frac{\|r_k\|^2}{P_{k+1}^T A P_{k+1}} (\because r_k \cdot r_l = 0 \ for \ l < k)$$

$$\beta_{k+1} = \frac{r_k^T A r_{k-1}}{P_k^T A r_{k-1}} \overset{(2),(3)}{=} \frac{r_k^T A P_k - \beta_k r_k^T A P_{k-1}}{\frac{1}{\alpha_k} \|r_{k-1}\|^2} \overset{(2),(3)}{=} \frac{r_k^T A P_k \alpha_k}{\|r_{k-1}\|^2} = \frac{\|r_k\|^2}{\|r_{k-1}\|^2}$$

(1) *by applying* (**) *recursively and* $r_k^T \cdot r_l = 0$ *for* $l < k$

(2) $r_{k\text{-}1} = P_k - \beta_k P_{k-1}$, $AP_{k\text{-}1} \in K_k$ and r_k is orthogonal to K_k

(3)*recall* $r_k = b - Ax_k = b - A(x_{k-1} + \alpha_k P_k) = r_{k-1} - \alpha_k A P_k$

$$\|r_k\|^2 = \alpha_k r_k^T A P_k$$

CG algorithm:

$$cg(x,\ b,\ A,\ \varepsilon,\ k_{\max})$$

1. $r = b - Ax,\ \rho_0 = \|r\|^2,\ k = 1$

2. $do\ while\ \sqrt{\rho_{k\text{-}1}} < \varepsilon \|b\|\ and\ k\ <\ k_{\max}$

   $(\rho_{k-1} = \|r_{k-1}\|^2\ relative\ residual\ less\ than\ \varepsilon)$

   $(a)\ if\ \ k = 1,\ then\ P = r\ \ (set\ P_1 = r_0)$

   $$else\ \beta = \frac{\rho_{k\text{-}1}}{\rho_{k\text{-}2}}\ \ and\ \ \underset{\rho_k}{P} = \underset{r_{k\text{-}1}}{r} + \underset{\beta_k \rho_{k\text{-}1}}{\beta P}$$

   $(b)\ \omega = A\rho$

   $$\alpha = \rho_{k\text{-}1} / P^T \omega$$

   $(c)\ \underset{(new)}{x} = \underset{(old)}{x} + \alpha P\ \ (update\ x)$

   $$(d)\ r = (b - Ax_{new})\ \ \ (update\ r)$$
   $$= b - A(x_{old} + \alpha P)$$
   $$= \underbrace{b - Ax_{old}}_{r_{old}} - \alpha \underbrace{AP}_{\omega} = r - \alpha\omega$$

   $(e)\ \rho_k = \|r\|^2$

   $(f)\ k = k + 1$

CG – convergence analysis:

Let x^* be the solution of $Ax = b$ and

x_k be the minimum of $\phi(x)$ over $x_0 + K_k$

consider $\omega \in \ x_0 + K_k \quad \Rightarrow \omega = \sum_{j=0}^{k\text{-}1} c_j A^j r_0 + x_0$

and $r_0 = \ b \text{ - } Ax_0 \ = \ A(x^* \text{ - } x_0)$

$$x^* - \omega = x^* - x_0 - \sum_{j=0}^{k-1} c_j A^{j+1}(x^* - x_0)$$

let $P(A) = 1 \text{ - } \sum_{j=0}^{k\text{-}1} c_j A^{j+1}$, we have $x^* - \omega = P(A)(x^* - x_0)$

Since $\left\| x^* - x_k \right\|_A \le \left\| x^* - \omega \right\|_A$ for all $\omega \in x_0 + K_k$, one has

$$\left\| x^* - x_k \right\|_A = \min_{\substack{P \in P_k \\ with\ P(0)=1}} \left\| P(A)(x^* - x_0) \right\|_A$$

Now suppose A is symmetric, positive definite and

diagonalizable. $U^T A U = \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

$P(A) = U^T P(\Lambda) U$

$$\left\| P(A)x \right\|_A = \left\| A^{1/2} P(A) x \right\|_2 \underset{A\ sym.}{\le} \left\| P(A) \right\|_2 \left\| A^{1/2} x \right\|_2 \le \left\| P(A) \right\|_2 \left\| x \right\|_A$$

$$\Rightarrow \left\| x_k - x^* \right\|_A \le \left\| x_0 - x^* \right\|_A \min_{\substack{P \in P_k \\ P(0)=1}} \ \max_{\substack{z \in \sigma(A) \\ = \{\lambda_1 \cdots \lambda_n\}}} \left| P(z) \right|$$

$$\Rightarrow \frac{\left\| x_k - x^* \right\|_A}{\left\| x_0 - x^* \right\|_A} \le \max_{z \in \sigma(A)} \left| P(z) \right|, \text{ for any } P \in P_k \text{ and } P(0) = 1$$

Theorem 2. Let A be symmetric and positive definite, then CG find the solution of AX=b within N iterations.

$$Ans: let\ \widetilde{P}(z) = \prod_{i=1}^{N} \frac{(\lambda_i - z)}{\lambda_i} \in P_N$$

$$clearly,\ \widetilde{P}(0) = 1\ and\ \max_{\substack{z\in\sigma(A)\\ =\{\lambda_1\cdots\lambda_n\}}} \widetilde{P}_N(z) = 0$$

$$\Rightarrow \frac{\left\|x_N - x^*\right\|_A}{\left\|x_0 - x^*\right\|_A} \le \max_{z\in\sigma(A)} \widetilde{P}_N(z) = 0 \quad \Rightarrow x_N = x^*$$

Theorem 3. Let A be symmetric and positive definite. Assume A has exactly k≤N distinct e-values of A. Then CG converges in at most k steps.

Observations:

$$Lemma: \frac{\left\|b - Ax_k\right\|_2}{\left\|b\right\|_2} \le \sqrt{K_2(A)} \cdot \frac{\left\|r_0\right\|_2}{\left\|b\right\|_2} \frac{\left\|x_k - x^*\right\|_A}{\left\|x_0 - x^*\right\|_A} \quad (Assume\ \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_N)$$

$$\frac{\left\|r_k\right\|_2}{\left\|r_0\right\|_2} = \underbrace{\frac{\left\|b - Ax_k\right\|_2}{\left\|b - Ax_0\right\|_2}} \le \frac{\left\|A(x^* - x_k)\right\|_2}{\left\|A(x^* - x_0)\right\|_2} \le \frac{\sqrt{\lambda_1}\left\|(x^* - x_k)\right\|_A}{\sqrt{\lambda_N}\left\|(x^* - x_0)\right\|_A}$$

$$\le \sqrt{Cond(A)} \cdot \frac{\left\|(x_k - x^*)\right\|_A}{\left\|(x_0 - x^*)\right\|_A}.$$

Here $K_2(A) = Cond(A)$ under the 2-norm.

J.W. Daniel The conjugate gradient method for linear and nonlinear operate equations SIAM J. Numer. Anal. 1983 p.296-314.

The sharpest estimations:

$$\left\|x_k - x^*\right\|_A \le 2\left\|x_0 - x^*\right\|_A \left[\frac{\sqrt{K_2(A)}-1}{\sqrt{K_2(A)}+1}\right]^k$$

$\Rightarrow$ *Observation*:

if $K_2(A) >> 1$, the convergence may be very slow.
On the other hand, if $K_2(A) \sim 1$, then the convergence is very fast.

How many iterations are needed for the relative error to be less than a given tolerance ε?

Exercise:

The number of iteration needed for the CG algorithm is about $O\left(\ln\left(\frac{\varepsilon}{2}\right)\frac{\sqrt{K_2(A)}}{2}\right)$.

To reduce the conditional number of A, we need preconditioning!

Solve $Ax=b \equiv \underbrace{M^{-1}Ax}_{\tilde{A}} = \underbrace{M^{-1}b}_{\tilde{b}}$ (M is called a precinditioner of A)

Candidates of preconditioner : (A=M-N)

1. Jacobi iteration: M=diag(A)
2. Gauss-Seidel iteration: M=lower triangular of A
3. incomplete Cholesky factorization $A = LL^T + E$ (E: a small perturbation)
4. multigrid iteration:

Preconditioned CG (PCG)

ALGORITHM : 9 ■ ***Preconditioned Conjugate Gradient***

1. Compute $r_0 := b - Ax_0$, $z_0 = M^{-1}r_0$, and $p_0 := z_0$
2. For $j = 0, 1, \ldots$, until convergence Do:
3. $\alpha_j := (r_j, z_j)/(Ap_j, p_j)$
4. $x_{j+1} := x_j + \alpha_j p_j$
5. $r_{j+1} := r_j - \alpha_j A p_j$
6. $z_{j+1} := M^{-1} r_{j+1}$
7. $\beta_j := (r_{j+1}, z_{j+1})/(r_j, z_j)$
8. $p_{j+1} := z_{j+1} + \beta_j p_j$
9. EndDo