Conjugate Gradient Methods

Useful for linear and nonlinear problems. Even for ~~the~~ linear problems the method is an iterative method (proceeds by successive approx.)

Linear CG.

$Ax = b$ $\quad$ A SPD $\quad$ (or HPD for complex probs)

$$Ax - b = 0 \longrightarrow \min_{x \in \mathbb{R}^n} \tfrac{1}{2} x^T A x - b^T x \quad (=: \phi(x))$$

Note that $\nabla \phi(x) = Ax - b =: r(x)$ (residual)

(often $r := b - Ax$). Let x_* be exact solution

For A SPD, we can define

$$\langle x, y \rangle_A = y^T A x \quad \text{and} \quad \|x\|_A = (x^T A x)^{1/2}$$

(see appendix A.1 elements of linear algebra)

Given $x_0 \rightarrow r_0 = Ax_0 - b$ and CG:

Step m: $z_m \in \text{span}\{r_0, Ar_0, \dots, A^{m-1} r_0\}$ s.t.

$x_m = x_0 + z_m$ satisfies $\|x_m - x_*\|_A$ minimum

At each step CG extends search space

$K_m(A, r_0) = \text{span}\{r_0, Ar_0, \dots, A^{m-1} r_0\}$ (Krylov (sub)space)

and computes best approx. solution over extended space.

CG build basis for $K_m(A, r_0)$ of search directions (vectors) $p_1, \cdots, p_m$ that are orthog in $\langle \cdot, \cdot \rangle_A$.

$$\operatorname{span}\{p_0, p_1, \cdots, p_m\} = K_{m+1}(A, r_0) = \operatorname{span}\{r_0, \cdots, A^m r_0\}$$

$$p_0 = -\nabla\phi(x_0) = -r_0$$

$$\|x_m - x_*\|_A^2 = (x_m - x_*)^T A (x_m - x_*) \quad \min_{x_m \in x_0 + K_m}$$

$$x_m = x_0 + z_m, \quad z_m \in K_m(A, r_0) \quad \text{s.t.}$$

<u>Theo</u> $\displaystyle z_m = \arg\min_{z \in K_m(A, r_0)} \|x_* - x_0 - z\|_A \implies$

$$(x_* - x_0) - z_m \perp_A K_m(A, r_0)$$

Proof: Assume $(x_* - x_0) - z_m \not\perp_A K_m(A, r_0)$.

~~Then~~ (and $\|x_* - x_0 - z_m\|$ min)

Then $\exists p \in K_m(A, r_0)$ s.t. $p^T A(x_* - x_0 - z_m) = \alpha > 0$ $(\neq 0)$

Consider $\|(x_* - x_0 - (z_m + t p)\|_A^2 =$

$$(x_* - x_0 - z_m - tp)^T A (x_* - x_0 - z_m - tp) =$$

$$\|x_* - x_0 - z_m\|_A^2 - 2t p^T A(x_* - x_0 - z_m) + t^2 p^T A p$$

Clearly $z + tp \in K_m(A, r_0)$. So by $\|x_* - x_0 - z_m\|_A$ min

~~$\|x_* - x_0 - z_m\|_A^2 - 2tp^TA(x_* - x_0 - z_m) + t^2p^TAp \geq$~~

~~for $t \geq 0$~~

$$\|x_* - x_0 - z_m\|_A^2 - 2tp^TA(x_* - x_0 - z_m) + t^2p^TAp \geq \|x_* - x_0 - z_m\|_A^2$$

for $t > 0$

$$\Leftrightarrow \quad t^2p^TAp - 2tp^TA(x_* - x_0 - z_m) \geq 0 \quad \text{for } t > 0$$

$$t\|p\|_A - 2\alpha \geq 0 \quad \text{for } t > 0$$ (and $\alpha > 0$ by assumption)

Contradiction ($0 < t < \frac{2\alpha}{\|p\|_A}$)

Hence such p must not exist for $\|x_* - x_0 - z_m\|$ min

Therefore $(x_* - x_0) - z_m \perp_A K_m(A, r_0)$

! (This is general result for norms and assoc. inner products for complete spaces $\rightarrow z_m$ guaranteed to always exist)

So, $z_m = \sum_{i=0}^{m-1} \alpha_i p_i$ s.t. $x_* - x_0 - \sum_{i=0}^{m-1} \alpha_i p_i \perp_A p_k$ for $k = 0..m-1$

$$p_k^T A (x_* - x_0 - \sum \alpha_i p_i) = 0 \quad \Leftrightarrow$$

$$p_k^T (b - Ax_0 - \sum \alpha_i A p_i) = 0 \quad \text{and}$$

$$p_k^T (b - Ax_0 - Az_m) = p_k^T(b - Ax_m) = -p_k^T r_m = 0$$

$$p_k^T \left(-r_0 - \sum_{i=0}^{m-1} \alpha_i A p_i\right) = 0 \qquad k = 0..m-1$$

$$-p_k^T r_0 - \sum_i p_k^T A p_i \alpha_i = 0 \qquad k = 0..m-1$$

$k=0 \quad \|r_0\|^2 - \alpha_0 p_0^T A p_0 - \sum_{i=1}^{m-1} \alpha_i p_0^T A p_i = 0$

We aim to construct p_i s.t. $p_i^T A p_j = 0 \quad i \neq j$

$$\rightarrow \quad \alpha_0 = \frac{r_0^T r_0}{p_0^T A p_0} \quad \text{(st. desc. with opt. LS)}$$

$$k=1 \quad -p_1^T r_0 - \sum_{i=0}^{m-1} \alpha_i p_1^T A p_i = 0 \iff$$

$$-p_1^T r_0 - \alpha_1 p_1^T A p_1 = 0 \iff$$

$$\alpha_1 = - \frac{p_1^T r_0}{p_1^T A p_1}$$

more gen: $-p_k^T r_0 - \alpha_k p_k^T A p_k = 0 \quad \alpha_k = \dfrac{-p_k^T r_0}{p_k^T A p_k}$

So, after computing p_m s.t. $\langle p_m, p_i \rangle_A = 0 \quad i = 0..m-1$

and $\text{span}\{p_0, \dots, p_m\} = K_{m+1}(A, r_0)$

(after prev. comp. $p_0 \cdots p_{m-1}$),

we compute α_m and x_m and r_m as ~~above~~ above.

Can we / How to compute p_m?

$p_0 = -r_0 \longrightarrow x_1 = x_0 + \alpha_0 p_0 = x_0 - \alpha_0 r_0$

$r_1 = A x_0 - \alpha_0 A r_0 - b = r_0 - \alpha_0 A r_0 \in K_2(A, r_0)$

We also showed that $r_1 \perp K_1(A, r_0) = s\{r_0\}$

Theo: $\text{span}\{p_0, \dots, p_m\} = \text{span}\{r_0, \dots, r_m\}$

Proof by induction $\quad | \quad = K_{m+1}(A, r_0)$

$p_1 \in K_2(A, r_0)$ and $p_1^T A p_0 = 0$

$p_1 = \gamma_0 p_0 + \gamma_1 r_1 \rightarrow p_0^T A(\gamma_0 p_0 + \gamma_1 r_1) = 0 \iff$

$\gamma_0 p_0^T A p_0 + \gamma_1 p_0^T A r_1 = 0$, to extend space $\gamma_1 = 0$ not allowed

In st. descent $p_1 = -r_1 \rightarrow \gamma_1 = -1$

$$\gamma_0 p_0^T A p_0 - p_0^T A r_1 = 0 \Rightarrow \gamma_0 = \frac{p_0^T A r_1}{p_0^T A p_0}$$

Proceed as follows $p_m = -r_m + \sum_{i=0}^{m-1} \gamma_i p_i \rightarrow$

$$p_m \in \text{span}\{r_m, p_{m-1}, \cdots, p_0\} = \text{span}\{r_m, r_{m-1}, \cdots, r_0\} = K_{m+1}(A, r_0)$$

$r_{m+1} \in K_{m+2}(A, r_0)$ and $r_{m+1} \perp K_{m+1}(A, r_0) = \text{span}\{r_0, \cdots, r_m\}$

Also $r_m = -p_m + \sum_{i=0}^{m-1} \gamma_i p_i \Rightarrow r_m \in \text{span}\{p_0, \cdots, p_m\}$

We can continue this process as long as $r_m \neq 0$

Note $r_m \perp \text{span}\{r_0, \cdots, r_{m-1}\}$ $m = 1, 2, 3, \cdots$

$r_i^T r_j = 0$ for $i \neq j$

Theo If $r_i \neq 0$ for $i = 0 .. n-1$, then $r_n = 0$ (CG actually finite termination method)

Proof ?

Note $r_i = 0 \iff A x_i - b = 0 \iff x_i = x_*$

Take $p_m = -r_m + \sum_{i=0}^{m-1} \gamma_i p_i$

$$p_j^T A p_m = -p_j^T A r_m + \sum_{i=0}^{m-1} \gamma_i p_j^T A p_i = 0 \iff$$

$$-p_j^T A r_m + \gamma_j p_j^T A p_j = 0 \implies \gamma_j = \frac{+p_j^T A r_m}{p_j^T A p_j}$$

However $p_j^T A r_m = r_m^T A p_j$

$$p_j \in K_{j+1}(A, r_0) = \text{span}\{r_0, r_1, \dots, r_j\}$$

$$A p_j \in K_{j+2}(A, r_0) = \text{span}\{r_0, r_1, \dots, r_{j+1}\}$$

$$r_m^T A p_j = 0 \quad \text{for } m > j+1 \iff j < m-1$$

For $j = 0, 1, \dots, m-2$ $\quad p_j^T A r_m = 0 \rightarrow \gamma_j = 0$

$$p_m = -r_m + \gamma_{m-1} p_{m-1} = -r_m + \frac{p_{m-1}^T A r_m}{p_{m-1}^T A p_{m-1}} p_{m-1}$$

$$x_m = x_0 + z_m = x_0 + \sum_{i=0}^{m-1} \alpha_i p_i = x_0 + \sum_{i=0}^{m-1} -\frac{p_i^T r_0}{p_i^T A p_i} p_i$$

Since α_i can be comp. as soon as p_i available

$$x_m = x_{m-1} - \frac{p_{m-1}^T r_0}{p_{m-1}^T A p_{m-1}} p_{m-1}$$

$$r_m = A x_{m-1} - \alpha_{m-1} A p_{m-1} - b = r_{m-1} - \alpha_{m-1} A p_{m-1}$$

Further simplification

$$r_m = r_{m-1} - \alpha_{m-1} A p_{m-1} = r_{m-2} - \alpha_{m-2} A p_{m-2} - \alpha_{m-1} A p_{m-1}$$

$$= r_0 - \alpha_0 A p_0 - \alpha_1 A p_1 - \dots - \alpha_{m-1} A p_{m-1}$$

$$r_0 = r_{m-1} + \alpha_0 A p_0 + \alpha_1 A p_1 + \cdots + \alpha_{m-2} A p_{m-2}$$

$$p_{m-1}^T r_0 = p_{m-1}^T r_{m-1} + \underbrace{\sum_{i=0}^{m-2} \alpha_i p_{m-1}^T A p_i}_{=0} = p_{m-1}^T r_{m-1}$$

But $p_{m-1} = -r_{m-1} + \sum_{i=0}^{m-2} \gamma_i p_i = -r_{m-1} + \dfrac{p_{m-2}^T A r_{m-1}}{p_{m-2}^T A p_{m-2}} p_{m-2}$

Since $r_{m-1}^T p_{m-2} = 0$ (why?)

$$p_{m-1}^T r_{m-1} = -r_{m-1}^T r_{m-1} \rightarrow \alpha_{m-1} = \frac{r_{m-1}^T r_{m-1}}{p_{m-1}^T A p_{m-1}}$$

$$p_{m-1}^T A r_m = r_m^T A p_{m-1} = -r_m^T r_m / \alpha_{m-1} = -\frac{r_m^T r_m}{r_{m-1}^T r_{m-1}} \cdot p_{m-1}^T A p_{m-1}$$

$$\gamma_{m-1} = \frac{p_{m-1}^T A r_m}{p_{m-1}^T A p_{m-1}} = -\frac{r_m^T r_m}{r_{m-1}^T r_{m-1}}$$

<u>CG</u> : $x_0 \rightarrow r_0 = A x_0 - b$, $p_0 = -r_0$, $m = 0$

while $\|r_m\| > \varepsilon$ {chosen tolerance} in practice add max. its.

$\quad \alpha_m = r_0^T r_0 / p_0^T A p_0$

$\quad x_{m+1} = x_m + \alpha_m p_m$

$\quad r_{m+1} = r_m + \alpha_m A p_m$

$\quad \beta_{m+1} = + \dfrac{r_{m+1}^T r_{m+1}}{r_m^T r_m}$

$\quad p_{m+1} = -r_{m+1} + \beta_{m+1} p_m$

$\quad m = m+1$

end

$$z_m \in K_m(A, r_0) = \gamma_0 r_0 + \gamma_1 A r_0 + \cdots + \gamma_{m-1} A^{m-1} r_0$$

$$= Z_{m-1}(A)\, r_0$$

$\hookrightarrow$ poly. degree $m-1$

$$r_m = A x_0 + A z_m - b = r_0 + A z_m = r_0 + A Z_{m-1}(A) r_0$$

$$= (I + A Z_{m-1}(A))\, r_0 = R_m(A)\, r_0$$

poly of degree m and $R_m(0) = 1$

$$R_m(A) r_0 = 1 \cdot r_0 + \gamma_0 A r_0 + \gamma_1 A^2 r_0 + \cdots + \gamma_{m-1} A^m r_0$$

Polynomial repr. can be used to prove many important properties.

$$e_m = x_m - x_* \Rightarrow A e_m = A x_m - b = r_m$$

$$e_m^T A e_m = \| x_m - x_* \|_A^2 \quad \text{(min. by CG)}$$

$$= r_m^T A^{-1} r_m = \| r_m \|_{A^{-1}}^2 \quad \text{(proper norm)}$$

$$\left(\frac{\| e_m \|_A}{\| e_0 \|_A} \right)^2 = \frac{e_m^T A e_m}{e_0^T A e_0} = \frac{e_m^T A A^{-1} A e_m}{e_0^T A A^{-1} A e_0} =$$

$$e_0 = \sum_i \eta_i v_i \quad (A = V \Lambda V^T)$$

$$\| e_0 \|_A^2 = \sum_i \eta_i^2 \lambda_i$$

$$\| e_m \|_A^2 = \| r_m \|_{A^{-1}}^2 = \sum_i R_m(\lambda_i)^2 \lambda_i v_i$$

$$e_m = A^{-1} r_m = A^{-1} R_m(A) r_0 = A^{-1} R_m(A) A e_0$$

$$= R_m(A)\, e_0$$

$$\| e_m \|_A^2 = \left(\sum_i R_m(\lambda_i) \eta_i v_i \right)^T A \left(\sum_i R_m(\lambda_i) \eta_i v_i \right)$$

$$= \sum_i \lambda_i R_m^2(\lambda_i) \eta_i^2 \quad \text{(optimal)}$$

~~So any other choice of poly~~

$$= \sum_i d_i \eta_i^2 \underbrace{\left(1 + d_i Z_m(d_i)\right)}_{R_m(d_i)}^2$$

Optimal, so any other choice of poly. $\tilde{Z}_m(d)$ leads to larger error (in A-norm)

$\rightarrow$ use for bounds on error.

Theo: If A has r distinct eigenvalues $d_1, \dots, d_r$ (>0)

CG conv. in r iterations

Proof: Take $R_r(d)$ s.t. $R_r(d_1) = \dots = R_r(d_r) = 0$

(since $d_i > 0$ we can still satisfy cond $R_m(0) = 1$)

(alt. consider expan. of r_0 in V and dim argument)

$$\|e_m\|_A = \|R_m(A)\, e_0\|_A \leq \|\tilde{R}_m(A)\, e_0\|_A \leq \|\tilde{R}_m(A)\|_2 \, \|e_0\|_A$$

$\hookrightarrow$ bounds by appropriate choices for $\tilde{R}_m(\cdot)$

$$\|\tilde{R}_m(A)\|_2 = \|V \tilde{R}_m(\Lambda) V^T\|_2 = \|\tilde{R}_m(\Lambda)\|_2 = \max_i |\tilde{R}_m(d_i)| = \max_i |1 - d_i \tilde{Z}_{m-1}(d_i)|$$

~~Taking $\tilde{R}_m$ s.t. $R_m(d_n) = R_m(d_{n-1}) = \dots = R_m(d_{n-m+2}) = 0$~~

~~and $R_m(\tfrac{1}{2}(d_1 + d_{n-m+1})) = 0$~~

Theo 5.5 : (After m+1 it.s)

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

Proof: Take $\tilde{R}_{m+1}$ s.t. $\tilde{R}_{m+1}(\lambda_n) = \dots = \tilde{R}_{m+1}(\lambda_{n-m+1}) = 0$

$$\tilde{R}_{m+1}\left(\tfrac{1}{2}(\lambda_1 + \lambda_{n-m})\right) = 0$$

work out max over remaining eigenvalues

Theo: $\| x_m - x_* \|_A \le 2 \left(\frac{\sqrt{\lambda_n/\lambda_1} - 1}{\sqrt{\lambda_n/\lambda_1} + 1} \right)^k \| x_0 - x_* \|_A$

where $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$

Proof: Take for $\tilde{R}_m$ Chebyshev poly taking min over interval $[\lambda_1, \lambda_n]$

Reducing λ_n/λ_1 and clustering eigenvalues yield better convergence $\rightarrow$ preconditioning

Common preconditioner is Incomplete Cholesky prec.

Consider prec. M (SPD) and $M \simeq A^{-1} \rightarrow$ fast convergence.

Define $\langle x, y \rangle_{M^{-1}} = y^T M^{-1} x$. Then

MA is "symm" (self-adjoint wrt $\langle \cdot, \cdot \rangle_{M^{-1}}$

$$\langle MAx, y\rangle_{M^{-1}} = y^T M^{-1} M A x = y^T A M M^{-1} x$$

$$= \langle x, MAy\rangle_{M^{-1}} \qquad = y^T A^T M^T M^{-1} x$$

$$\langle MAx, x\rangle_{M^{-1}} = x^T M^{-1} M A x = x^T A x > 0$$

for all $x \neq 0$

So, MA "symm." pos. def. wrt $\langle \cdot, \cdot \rangle_{M^{-1}}$

See alg. 5.3 (prec. CG)

Preconditioning is in general problem dependent.

Some common preconditioners for fairly general systems are those based on incomplete decompositions, like incomple Cholesky — a Cholesky decomp. in which (some or all) intermediate fill-in is ignored.

Nonlinear CG

How to go from linear CG $\to$ min $\phi(x) = \frac{1}{2}x^T A x - b^T x$ to more general $f(x)$.

(i) replace role of r_k by ∇f_k

(ii) Change choice of α_k by appropr. line search

Fletcher - Reeves

$x_0 \to f_0 = f(x_0)$, $\nabla f_0 = \nabla f(x_0)$
$p_0 = -\nabla f_0$, $k = 0$

while $\|\nabla f_0\| > \varepsilon$ and $k <$ maxits

$LS \to \alpha_k$; $x_{k+1} = x_k + \alpha_k p_k$

compute ∇f_{k+1}

$$\beta_{k+1} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$$

$$p_{k+1} = -\nabla f_{k+1} + \beta_{k+1} p_k$$

$k = k+1$

end

FR CG globally conv. but not so good in practice (read discussion book)

We guarantee p_k is descent direction by req. α_k satisfies strong Wolfe cond.s with $0 < c_1 < c_2 < \frac{1}{2}$ (stronger than in chap 3)

$$f(x_k + \alpha_k p_k) \le f(x_k) + c_1 \alpha_k \nabla f_k^T p_k$$

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

Variant(s):

Polak-Ribiere: $\beta_{k+1} = \dfrac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\|\nabla f_k\|^2}$

(otherwise like FR)

Very good in practice but may not converge. $\rightarrow$
modification: $k \geq 2$

$$\beta_k = \begin{cases} -\beta_k^{FR} & \text{if } \beta_k^{PR} < -\beta_k^{FR} \\ \beta_k^{PR} & \text{if } |\beta_k^{PR}| \le \beta_k^{FR} \\ \beta_k^{FR} & \text{if } \beta_k^{PR} > \beta_k^{FR} \end{cases}$$

β_k^{FRPR}

~~Kalter~~ Note $|\beta_k^{FR-PR}| \le \beta_k^{FR} \rightarrow$

show suff. for glob. conv.

Global Convergence

Assump.
(i) level set $\mathcal{L} = \{x \mid f(x) \le f(x_0)\}$ bounded

(ii) In some open nbhood N of $\mathcal{L}$, f is Lip. cont. differentiable

(i)+(ii) $\Rightarrow$ $\|\nabla f(x)\| \le \bar{\gamma}$ $\quad x \in \mathcal{L}$

Zoutendyks theorem (Theo 3.2) :

$$\sum_{k=0}^{\infty} \cos^2\theta_k \, \|\nabla f_k\|^2 < \infty$$

Assume at iterations $k_1, k_2, \ldots$ we restart CG

then $\sum_{k=k_1,k_2,\ldots} \|\nabla f_k\|^2 < \infty$ (*1)

If we restart at least every $\bar{n}$ it.s, $k_1, k_2, \ldots$ inf. seq. and (*1) implies $\lim_{k_j \to \infty} \|\nabla f_{k_j}\| = 0 \Leftrightarrow \liminf_{k \to \infty} \|\nabla f_k\| = 0$

Not surprising because steps k_j are steepest descent steps.

What about unrestarted CG (FR or FR-PR variant)?

Theo (variant of Theo 5.7 for PR)

Assume assumptions above and line search with strong W. cond.s with $0 < c_1 < c_2 < \frac{1}{2}$.

Then $\liminf_{k \to \infty} \|\nabla f_k\| = 0$

Proof (by contradiction)

Assume there exists $\gamma > 0$ s.t. $\|\nabla f_k\| \geq \gamma$ and $K > 0$ for all $k \geq K$

Fletcher-Reeves and variants

Lemma (var. lem 5.6)

FR with strong WC and $0 < c_2 < \frac{1}{2}$ yields

(*1) p_k with $-\frac{1}{1-c_2} \leq \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \leq \frac{2c_2-1}{1-c_2}$ $(k=0,1,2,\dots)$

holds for any choice $\tilde{\beta}_{k+1}$ s.t. $|\tilde{\beta}_{k+1}| \leq \beta_{k+1}$

Proof: Note $t(\xi) = \frac{2\xi-1}{1-\xi}$ strict monot. increasing on $[0, \frac{1}{2}]$

$t(0) = -1$, $t(\frac{1}{2}) = 0$. Therefore $-1 < \frac{2c_2-1}{1-c_2} < 0$

$k = 0 \rightarrow \frac{-\nabla f_0^T \nabla f_0}{\|\nabla f_0\|^2} = -1$ (*1) satisfied

(also fine for PR-FR)

Proceed by induction (true for some $k \geq 1$)

by (5.41b):
$$\frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} = \frac{-\nabla f_{k+1}^T \nabla f_{k+1} + \beta_{k+1} \nabla f_{k+1}^T p_k}{\|\nabla f_{k+1}\|^2} =$$

$$-1 + \beta_{k+1} \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_{k+1}\|^2} = -1 + \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_k\|^2}$$

for $\beta_{k+1}^{FR} = \frac{\|\nabla f_{k+1}\|^2}{\|\nabla f_k\|^2}$ (5.41a)

Now assume $\tilde{\beta}_{k+1}$ s.t. $|\tilde{\beta}_{k+1}| \leq \beta_{k+1}^{FR} \Rightarrow$

$$\frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} = -1 + \tilde{\beta}_{k+1} \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_{k+1}\|^2} = -1 + \frac{\tilde{\beta}_{k+1}}{\beta_{k+1}} \frac{\nabla f_{k+1}^T p_k}{\|\nabla f_k\|^2}$$

$$\left| \tilde{\beta}_{k+1} / \beta_{k+1} \right| \leq 1 \iff -1 \leq \frac{\tilde{\beta}_{k+1}}{\beta_{k+1}} \leq 1$$

From strong WC :

$$|\nabla f_{k+1}^T p_k| \le -c_2 \nabla f_k^T p_k \qquad \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2}$$

$$-1 + c_2 \frac{|\tilde{\beta}_{k+1}|}{\beta_{k+1}} \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \le \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \le -1 - c_2 \frac{|\tilde{\beta}_{k+1}|}{\beta_{k+1}}$$

For FR $\tilde{\beta}_{k+1} = \beta_{k+1} \overset{>0}{} \to \frac{|\tilde{\beta}_{k+1}|}{\beta_{k+1}} = 1$

From ind. hyp. $-\frac{1}{1-c_2} \le \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \le \frac{2c_2-1}{1-c_2} < 0$

$$-1 - \frac{c_2}{1-c_2} \frac{|\tilde{\beta}_{k+1}|}{\beta_{k+1}} \le \frac{\nabla f_{k+1}^T p_{k+1}}{\|\nabla f_{k+1}\|^2} \le -1 + \frac{c_2}{1-c_2} \frac{|\tilde{\beta}_{k+1}|}{\beta_{k+1}}$$

$$\underbrace{-1 - \frac{c_2}{1-c_2}} \le -1 - \frac{c_2}{1-c_2} \frac{|\tilde{\beta}_{k+1}|}{\beta_{k+1}} \le \cdots \le -1 + \frac{c_2}{1-c_2} \frac{|\tilde{\beta}_{k+1}|}{\beta_{k+1}} \le -1 + \frac{c_2}{1-c_2}$$

$$\frac{c_2-1}{1-c_2} - \frac{c_2}{1-c_2} = \frac{-1}{1-c_2} \qquad \frac{2c_2-1}{1-c_2} = \frac{c_2-1}{1-c_2} + \frac{c_2}{1-c_2}$$

$$-\frac{1}{1-c_2} \le \frac{\nabla f_k^T p_k}{\|\nabla f_k\|^2} \le \frac{2c_2-1}{1-c_2} \Rightarrow$$

$$-\frac{1}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|} \le \underbrace{\frac{\nabla f_k^T p_k}{\|\nabla f_k\| \|p_k\|}}_{-\cos\theta_k} \le \frac{2c_2-1}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|} \Rightarrow$$

$$-\frac{2c_2-1}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|} \le \cos\theta_k \le \frac{1}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|}$$

$$\| \quad \frac{1-2c_2}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|}$$

$\hookrightarrow$ bnd from 0 for $c_2 < \frac{1}{2}$

Global Conv.

Assumptions: (i) level set $\mathcal{L} := \{x \mid f(x) \le f(x_0)\}$ is bnd

(ii) In some open nbhood N of $\mathcal{L}$ obj. func.

f is Lip. cont. diff.

$\rightarrow \|\nabla f(x)\| \le \bar{\gamma}$ for all $x \in \mathcal{L}$

Theo Zoutendyk (Theo 3.2):

$$\sum_{k=0}^{\infty} \cos^2\theta_k \|\nabla f_k\|^2 < \infty$$

Global conv for FR and variants s.t.

$|\tilde{\beta}_{k+1}| \le \beta_{k+1}$ (like FR-PR CG)

Theo (var. Theo 5.7)

Under given assump.; assuming alg. 5.4 with β_{k+1}^{FR} or $|\tilde{\beta}_{k+1}| \le \beta_{k+1}^{FR}$, LS with strong WC and $0 < c_1 < c_2 < \frac{1}{2}$,

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

Proof (by contradiction)

Assume $\|\nabla f_k\| \ge \gamma$ for all k suff. large

(for suff. small γ we can take $\|\nabla f_k\| < \gamma$ as conv. crit. and assume $\|\nabla f_k\| \ge \gamma$ for all k)

Zoutendyk plus lower bnd $\cos\theta_k$:

$$\infty > \sum_{k=0}^{\infty} \cos^2\theta_k \|\nabla f_k\|^2 \geq \sum_{k=0}^{\infty} \left(\frac{1-2c_2}{1-c_2} \frac{\|\nabla f_k\|}{\|p_k\|}\right)^2 \|\nabla f_k\|^2$$

$$= \frac{1-2c_2}{1-c_2} \sum_{k=0}^{\infty} \frac{\|\nabla f_k\|^4}{\|p_k\|^2} \Rightarrow$$

$$\sum_{k=0}^{\infty} \frac{\|\nabla f_k\|^4}{\|p_k\|^2} < \infty$$

Since $\|\nabla f_k\| \geq \gamma$ ~~(~~ ~~→~~ ~~)~~

$$\sum_{k=0}^{\infty} \frac{\gamma^4}{\|p_k\|^2} = \gamma^4 \sum_k \frac{1}{\|p_k\|^2} < \infty$$

$$|\nabla f_k^T p_{k-1}| \leq -c_2 \nabla f_{k-1}^T p_{k-1} \leq \frac{c_2}{1-c_2} \|\nabla f_{k-1}\|^2$$

From (5.41b) with poss. $\tilde{\beta}_k$ i.s.o. $\beta_k = \beta_k^{FR}$.

$$\|p_{k+1}\|^2 = (-\nabla f_{k+1} + \tilde{\beta}_{k+1} p_k)^T(-\nabla f_{k+1} + \tilde{\beta}_{k+1} p_k)$$

$$\leq \|\nabla f_{k+1}\|^2 + 2|\tilde{\beta}_{k+1}| |\nabla f_{k+1}^T p_k| + \tilde{\beta}_{k+1}^2 \|p_k\|^2$$

$\downarrow$

(for $k \geq 1$)

$$\|p_k\|^2 \leq \|\nabla f_k\|^2 + 2|\tilde{\beta}_k| |\nabla f_k^T p_{k-1}| + \tilde{\beta}_k^2 \|p_{k-1}\|^2$$

$$\leq \|\nabla f_k\|^2 + \frac{2|\tilde{\beta}_k| c_2}{1-c_2} \|\nabla f_{k-1}\|^2 + \tilde{\beta}_k^2 \|p_{k-1}\|^2 =$$

$$\|\nabla f_k\|^2 + \underbrace{\frac{|\tilde{\beta}_k|}{\beta_k}}_{\leq 1} \cdot \underbrace{\frac{2c_2}{1-c_2}}_{>0} \underbrace{\beta_k \|\nabla f_{k-1}\|^2}_{\|\nabla f_k\|^2} + \underbrace{\frac{\tilde{\beta}_k^2}{\beta_k^2}}_{} \cdot \beta_k^2 \|p_{k-1}\|^2$$

$< \beta_k^2$

$$= \|\nabla f_k\|^2 \left(1 + \frac{|\tilde{\beta}_k|}{\beta_k} \frac{2c_2}{1-c_2}\right) + \tilde{\beta}_k^2 \|p_{k-1}\|$$

$$\leq \|\nabla f_k\|^2 \left(\frac{1+c_2}{1-c_2}\right) + \beta_k^2 \|p_{k-1}\|$$

let $c_3 = \frac{1+c_2}{1-c_2} > 1$:

$$\|p_k\|^2 \le c_3 \|\nabla f_k\|^2 + \beta_k^2 \|p_{k-1}\|^2$$

repeat

$$\le c_3 \|\nabla f_k\|^2 + \beta_k^2 c_3 \|\nabla f_{k-1}\|^2 + \beta_k^2 \beta_{k-1}^2 \|p_{k-2}\|^2$$

$$\le \cdots$$

$$\le c_3 \|\nabla f_k\|^2 + c_3 \beta_k^2 \|\nabla f_{k-1}\|^2 + c_3 \beta_k^2 \beta_{k-1}^2 \|\nabla f_{k-2}\|^2 + \cdots + c_3 \beta_k^2 \cdots \beta_2^2 \|\nabla f_1\|^2 + \beta_k^2 \cdots \beta_1^2 \underbrace{\|p_0\|^2}_{=\|\nabla f_0\|^2}$$

$$\le c_3 \|\nabla f_k\|^2 + c_3 \|\nabla f_k\|^4 \|\nabla f_{k-1}\|^{-2} + c_3 \|\nabla f_k\|^4 \|\nabla f_{k-2}\|^{-2} + \cdots + c_3 \|\nabla f_k\|^4 \|\nabla f_1\|^{-2} + c_3 \|\nabla f_k\|^4 \|\nabla f_0\|^{-2}$$

intro for convenienc $c_3 > 1$

$$= c_3 \|\nabla f_k\|^4 \left(\sum_{j=0}^{k} \|\nabla f_j\|^{-2} \right)$$

? $\|p_k\|^2 \le c_3 \bar{\gamma}^4 \cdot \sum_{j=0}^{k} \frac{1}{\gamma^2} = k\, c_3 \frac{\bar{\gamma}^4}{\gamma^2}$

$\|\nabla f_j\| \ge \gamma \iff \|\nabla f_j\|^{-1} \le \frac{1}{\gamma}$ holds for $k > K$ (for some $K > 0$) // problem with proof in book

for $k > K$

$$\|p_k\|^2 \le c_3 \|\nabla f_k\|^4 \sum_{j=0}^{k} \|\nabla f_j\|^{-2} \le c_3 \bar{\gamma}^4 \Big(\sum_{j=0}^{K} \|\nabla f_j\|^{-2} + \underbrace{\sum_{j=K+1}^{k} \|\nabla f_j\|^{-2}}_{\le (k-K)\gamma^{-2}} \Big)$$

Assume $\|\nabla f_k\| \le \gamma$ is convergence tol.

So, as long as "not converged" $\|\nabla f_k\| \geq \gamma$.

Then assume alg. does not converge:

$$\|\nabla f_k\| \geq \gamma \quad \text{for all } k = 0, \ldots$$

(rather than $\|\nabla f_k\| \geq \gamma$ for k suff. large.

Then $\|p_k\|^2 \le \frac{c_3 \bar{\gamma}^4}{\gamma^2}(k+1)$

Hence $\sum_{k=1}^{\infty} \frac{1}{\|p_k\|^2} \geq C \sum_{k=1}^{\infty} \frac{1}{k+1}$ diverges

However, earlier we derived $\sum_{k=0}^{\infty} \frac{1}{\|p_k\|^2} < \infty$

Hence, we derived a contradiction and the assumption $\|\nabla f_k\| \geq \gamma$ must be false.