Advanced Topics in Control. Exercise 5. Chao Ni
Gradient Methods.

1. (a). Exact line search.

$$f(x^{k+1}) = f(x^k+t^kd^k) = \frac{1}{2}(x^k+t^kd^k)^TA(x^k+t^kd^k) - b^T(x^k+t^kd^k)$$
$$= \frac{1}{2}(d^k)^TAd^k(t^k)^2 + \frac{1}{2}\left[(x^k)^TAd^k+(d^k)^TAx^k-2b^Td^k\right]t^k$$
$$+\frac{1}{2}(x^k)^TAx^k - b^Tx^k.$$

$$(t^k)^* = \underset{t^k}{\operatorname{argmin}}\, f(x^{k+1})$$
$$= -\frac{\frac{1}{2}\left[(x^k)^TAd^k+(d^k)^TAx^k-2b^Td^k\right]}{(d^k)^TAd^k}.$$

wlog. Assume $A=A^T$, otherwise. Let $\hat{A} = \frac{A+A^T}{2} + \frac{A-A^T}{2}$, we have $\frac{1}{2}x^TAx = \frac{1}{2}x^T\hat{A}x$.

$$\Rightarrow \quad t^* = \frac{(b-Ax^k)^Td^k}{(d^k)^TAd^k}.$$

(b) Steepest descent.

Taylor Expansion at x^k.

$$f(x^{k+1}) = f(x^k) + \nabla f^T(x^k)(t^kd^k) + o(\|t^kd^k\|_2^2).$$
$$= f(x^k) + t^k\|\nabla f(x^k)\|\cos\langle\nabla f(x^k), d^k\rangle.$$

minimize $f(x^{k+1}) \iff \cos\langle\nabla f(x^k), d^k\rangle = -1$

$$\Rightarrow d^k = -\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|_2}, \quad (\text{as } \|d^k\|=1).$$

(c). $x^{k+1}-x^k = t^k d^k = t^k\cdot\frac{-\nabla f(x^k)}{\|\cancel{\nabla f(x^k)}\|_2}$.

$x^{k+1} = x^k + t^kd^k \cancel{= x^k + \frac{(b-Ax^k)^Td^k\,d^k}{(d^k)^TAd^k}}$

$$x^{k+2}-x^{k+1} = t^{k+1}d^{k+1}$$
$$= t^{k+1}\cdot\frac{-\nabla f(x^k+t^kd^k)}{\|\cancel{\nabla f(x^k+t^kd^k)}\|_2}$$

$$\Rightarrow \langle x^{k+2}-x^{k+1}, x^{k+1}-x^k\rangle \propto \langle\nabla f(x^k), \nabla f(x^k+t^kd^k)\rangle$$
$$\propto \langle -d^k, \nabla f(x^k+t^kd^k)\rangle.$$

Assume $\nabla f(x^k+t^kd^k) = \alpha d^k + \beta(d^k)^\perp$. with $\alpha\neq 0$,

Otherwise. $\langle x^{k+2}-x^{k+1}, x^{k+1}-x^k\rangle = 0$. done.

Let $\tilde{t}^k = t^k - \alpha d^k$.

Then.

$$f(x^k+\tilde{t}^k d^k) = f(x^k+t^kd^k) + \nabla f(x^k+t^kd^k)^T(-\alpha d^k) + H.O.T$$

$$= f(x^k+t^kd^k) + (\alpha d^k + \beta (d^k)^\perp)^T(-\alpha d^k) + H.O.T$$

$$= f(x^k+t^kd^k) - \alpha^2\|d^k\|_2^2 + H.O.T.$$

Hence., Because $\alpha \neq 0$.

$\tilde{t}^k$ is a better step than t^k. which is contradictory to the fact that t^k is the result of exact line search.

$\Rightarrow \alpha = 0.$

$\Rightarrow \langle x^{k+2}-x^{k+1}, x^{k+1}-x^k \rangle = 0$. $\square$

(d) f is L-Lipschitz smooth $\Leftrightarrow$ f is differentiable and $\|\nabla f(x)-\nabla f(y)\|_2 \le L\|x-y\|_2$

For f is convex.

f is L-Lipschitz smooth $\Leftrightarrow$ $f(y) \le f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|_2^2$.

$$\Rightarrow f(x^k+t^kd^k_{SD}) \le f(x^k) + \nabla f(x^k)^T d^k_{SD} t^k + \frac{L}{2} d^{kT}_{SD} d^k_{SD} (t^k)^2$$

$d^k_{SD} = -\nabla f(x^k)$.

$$= f(x^k) - \|\nabla f(x^k)\|_2^2 t^k + \frac{L}{2}\|\nabla f(x^k)\|_2^2 (t^k)^2$$

$$= f(x^k) - t^k(1-\frac{Lt^k}{2})\|\nabla f(x^k)\|_2^2 \quad \square$$

Under the L-Smooth assumption, the largest value decrease can be achieved when $t^k(1-\frac{Lt^k}{2})\|\nabla f(x^k)\|_2^2$ is maximized.

$$\Rightarrow (t^k)^* = \frac{1}{L}.$$

(e) If $|x_2| \le x_1$.

$$\nabla f(x_1,x_2) = \begin{pmatrix} \frac{2x_1}{2\sqrt{x_1^2+\gamma x_2^2}} \\ \frac{2\gamma x_2}{2\sqrt{x_1^2+\gamma x_2^2}} \end{pmatrix} = \frac{1}{\sqrt{x_1^2+\gamma x_2^2}} \begin{pmatrix} x_1 \\ \gamma x_2 \end{pmatrix}.$$

Steepest descent has $d^k = -\nabla f(x_1,x_2)$

Hence, $x^{k+1} = x^k + t^kd^k = x^k - t^k \cdot \nabla f(x)$.

$\Rightarrow$ $x_1^{k+1} = x_1^k - \frac{t x_1^k}{\sqrt{x_1^2+\gamma x_2^2}}$. (skip the subscript k for simplicity)

$x_2^{k+1} = x_2^k - \frac{t \gamma x_2^k}{\sqrt{x_1^2+\gamma x_2^2}}$ in $\sqrt{x_1^2+\gamma x_2^2} \to \sqrt{(x_1^k)^2+\gamma(x_2^k)^2}$.

$\Rightarrow$ $f(x_1^{k+1}, x_2^{k+1}; t) = \sqrt{\left(1-\frac{t}{\sqrt{x_1^2+\gamma x_2^2}}\right)^2 (x_1^k)^2 + \gamma\left(1-\frac{t\gamma}{\sqrt{x_1^2+\gamma x_2^2}}\right)^2 (x_2^k)^2}$.

$t = \arg\min_t f(x_1^{k+1}, x_2^{k+1}; t)$

$= \arg\min_t \frac{x_1^2+\gamma^3 x_2^2}{x_1^2+\gamma x_2^2} t^2 - \frac{2(x_1^2+\gamma^2 x_2^2)}{\sqrt{x_1^2+\gamma x_2^2}}$

$= \frac{x_1^2+\gamma^2 x_2^2}{x_1^2+\gamma^3 x_2^2} \cdot \frac{x_1^2+\gamma x_2^2}{\sqrt{x_1^2+\gamma x_2^2}}$. (with subscription k).

$\Rightarrow$ $x_1^{k+1} = \left(1 - \frac{x_1^2+\gamma^2 x_2^2}{x_1^2+\gamma^3 x_2^2}\right) x_1^k = \frac{\gamma^2(\gamma-1)x_2^{k2}}{x_1^{k2}+\gamma^3 x_2^{k2}} x_1^k$ > 0 if $x_1^k > 0$ and $x_2^k \neq 0$.

$x_2^{k+1} = \left(1 - \gamma \cdot \frac{x_1^2+\gamma^2 x_2^2}{x_1^2+\gamma^3 x_2^2}\right) x_2^k = \frac{(1-\gamma)x_1^{k2}}{x_1^{k2}+\gamma^3 x_2^{k2}} x_2^k$.

Hence when $x_1^{k+1} \neq 0$, $\frac{|x_2^{k+1}|}{|x_1^{k+1}|} = \frac{|(1-\gamma)x_1^{k2}\cdot x_2^k|}{\gamma^2(\gamma-1)x_2^{k2}\cdot x_1^k} = \frac{1}{\gamma^2}\frac{x_1^k}{|x_2^k|}$

$\Rightarrow$ $\frac{|x_2^{k+1}|}{x_1^{k+1}} \cdot \frac{|x_2^k|}{x_1^k} = \frac{1}{\gamma^2}$.

Because $(x_1^0, x_2^0) = (\gamma, 1)$ and $\gamma > 1$,

$\Rightarrow$ $\frac{|x_2^1|}{x_1^1} = 1$, $\frac{|x_2^2|}{x_2^2} = \frac{1}{\gamma^2}$, $\frac{|x_2^3|}{x_1^3} = 1$ ... $\frac{|x_2^n|}{x_1^n} = \left(\frac{1}{\gamma^2}\right)^{(n+1) \bmod 2}$.

If ~~start~~ from $(x_1^0, x_2^0) = (\gamma, 1)$,
$|x_2| \leq x_1$ will always hold.

By induction, we will show the iterates are $x_1^k = \gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k$, $x_2^k = \left(-\frac{\gamma-1}{\gamma+1}\right)^k$.

when $k=0$, it holds as $(x_1^0, x_2^0) = (\gamma, 1)$.

Assume for k holds,

$k+1$, $x_1^{k+1} = \frac{\gamma^2(\gamma-1)\left(\frac{-(\gamma-1)}{\gamma+1}\right)^{2k} \cdot \gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k}{\left[\gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k\right]^2 + \gamma^3\left[\left(-\frac{\gamma-1}{\gamma+1}\right)^k\right]^2} = \cancel{\frac{\gamma^2\left(\frac{\gamma-1}{\gamma+1}\right)^{2k}\cdot\left(\frac{\gamma-1}{\gamma+1}\right)^k\cdot(\gamma-1)}{2\gamma^3\left(\frac{\gamma-1}{\gamma+1}\right)^{2k}}}$

$= \frac{\gamma^2(\gamma-1)\left(\frac{\gamma-1}{\gamma+1}\right)^{2k}\cdot\gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k}{\gamma^2(\gamma+1)\left(\frac{\gamma-1}{\gamma+1}\right)^{2k}} = \gamma\left(\frac{\gamma-1}{\gamma+1}\right)^{k+1}$

$$x_2^{k+1} = \frac{(1-\gamma)\gamma^2 \cdot (\frac{\gamma-1}{\gamma+1})^{2k}}{\gamma^2(\gamma+1) \cdot (\frac{\gamma-1}{\gamma+1})^{2k}} \cdot \left(-\frac{\gamma-1}{\gamma+1}\right)^k = \left(-\frac{\gamma-1}{\gamma+1}\right)^{k+1}$$

Hence, it holds for $k+1$.

$\Rightarrow$ Iterations hold for all $n \in \mathbb{N}$. $\square$

The sequence converges to $x^* = (0, 0)$, It's a local minimize in domina $|x_2| \leq x_1$.
The global minimizer is unbounded below. ($x_1 \to -\infty$.)

2. Subgradient method.

$$x^{k+1} = x^k - t^k g^k, \quad g^k \in \partial f(x^k)$$

(a). By definition of subgradient.

$\forall x, \forall y$. $\quad f(y) \geq f(x) + g^T(y-x)$.

Take y such that $g^T(y-x) \geq 0$.

$\Rightarrow 0 \leq g^T(y-x) \leq f(y) - f(x)$.

$\Rightarrow$ ~~$0 \leq \|g\|_2 \|y-x\|_2$~~

$0 \leq \|g^T(y-x)\|_2 = \|f(y) - f(x)\|_2 \leq L\|y-x\|_2$ holds for $\forall y$.

Take y such that $\|g^T(y-x)\|_2 = \|g\|_2 \|y-x\|_2$

Hence $\quad \|g\|_2 \|y-x\|_2 \leq L\|y-x\|_2$

$\Rightarrow \quad \|g\|_2 \leq L$. $\square$

(b) $\quad \|x^{i+1} - x\|_2^2 = \|x^i - t^i g^i - x\|_2^2$

$= \|x^i - x\|_2^2 - 2t^i (g^i)^T(x^i - x) + (t^i)^2 \|g^i\|_2^2$

$\leq \|x^i - x\|_2^2 - 2t^i(f(x^i) - f(x)) + (t^i)^2 \cdot L^2$ $\quad \cdots \left(f(x) \geq f(x^i) + (g^i)^T(x - x^i) \right)$

Telescopic summation.

$\Rightarrow 0 \leq \|x^{k+1} - x\|_2^2 \leq \|x^0 - x\|_2^2 - 2\sum_{i=0}^{k} t^i (f(x^i) - f(x)) + L^2 \sum_{i=0}^{k} (t^i)^2$. $\quad (*)$

$\Rightarrow$ Let $f_{Best} = \min_{0 \leq i \leq k} f(x^i)$. Assume $f_{Best} > f(x)$, otherwise $(*)$ is easier to hold

and take $\tilde{f}_{Best} = \min_i \{ f(x^i) > f(x) \} \geq f_{Best}$.

Also, if $f_{Best} < f(x)$, (5) is trivially true.

$\Rightarrow (f_{Best}^k - f(x)) \cdot \sum_{i=0}^{k} 2t^i \leq \sum_{i=0}^{k} 2t^i (f(x^i) - f(x)) \leq \|x^0 - x\|_2^2 + L^2 \sum_{i=0}^{k} (t^i)^2$

$\Rightarrow \quad f_{Best}^k \leq f(x) + \dfrac{L^2 \sum_{i=0}^{k} (t^i)^2 + \|x - x^0\|_2^2}{2\sum_{i=0}^{k} t^i}$ $\square$

(c)

Assume x^* is the global minimizer

$$\frac{\|x^{k+1}-x^*\|_2}{\|x^k-x^*\|_2} = \sqrt{1-\frac{2t^k(g^k)^T(x^k-x^*)+(t^k)^2\|g^k\|_2^2}{\|x^k-x^*\|_2^2}}$$

$$\overset{t^k=\beta/\|g^k\|_2}{=} \sqrt{1-\frac{2\beta(g^k)^T(x^k-x^*)/\|g^k\|_2+\beta^2}{\|x^k-x^*\|_2^2}}$$

$$\geq \sqrt{1+\frac{\beta^2}{\|x^k-x^*\|_2^2}-\frac{2\beta}{\|x^k-x^*\|_2}}$$

$$=\left|1-\frac{\beta}{\|x^k-x^*\|_2}\right|$$

If $\|x^k-x^*\|_2 \to 0$. then $\|x^{k+1}-x^*\|_2 \to +\infty$. It doesn't converge.

3. Conjugate gradient method.

(a) we show that ~~x^*~~ $\{p^k\}_{k=0}^{n-1}$ is a Basis in $\mathbb{R}^n$.

Let $\sum_{i=0}^{n-1}\alpha_i p^i = 0$.

$\{p^k\}_{k=0}^{n-1}$ is a basis $\Leftrightarrow$ $\alpha_i \neq 0$. $\forall i=0,\dots n-1$.

Assume $\exists j$. $\alpha_j \neq 0$.

$$p^j = -\frac{1}{\alpha_j}\cdot\left(\sum_{i\neq j}\alpha_i p^i\right).$$

$$\Rightarrow (p^j)^TAp^j = -\frac{1}{\alpha_j}\cdot(p^j)^T\cdot A\cdot\left(\sum_{i\neq j}\alpha_i p^i\right) = \sum_{i\neq j}\frac{-\alpha_i}{\alpha_j}(p^j)^TAp^i = 0.$$ by definition of Conjugate direction

$\Rightarrow p^j=0$. which is in contradiction with the fact.

Hence. $\alpha_i=0$. $\forall i\in\{0,\dots n-1\}$, $\{p^k\}_{k=0}^{n-1}$ is a Basis.

$\Rightarrow$ x^*-x^0 can be spanned uniquely by $\{p^k\}_{k=0}^{n-1}$

$$x^*-x^0 = \sum_{k=0}^{n-1}\beta^k p^k$$

Because t^k is the step size obtained by exact line search, by 1(a),

we have ~~$(x^{k+1})^Tp^k=0$~~. $t^k = \dfrac{(b-Ax^k)^Tp^k}{(p^k)^TAp^k}$

with these coefficients $\{t^k\}$

$x_0+\sum_{k=0}^{n-1}t^kp^k$ is uniquely defined in $\mathbb{R}^n$. ~~and $\langle x^n-x^0, p^j\rangle_A = 0 \ \forall j\in\{0,n\}$~~

$$\Rightarrow x^n = x^0+\sum_{k=0}^{n-1}t^kp^k = x^0+\sum_{k=0}^{n-1}\frac{(b-Ax^k)^Tp^k}{(p^k)^TAp^k} \quad (1)$$

We finish our proof by showing $x^* = x^n$.

Because $A \succ 0$. $\Rightarrow x^* = A^{-1} b \Leftrightarrow x^* - x^0 = A^{-1}(b - Ax_0)$.

Given $\{p^k\}_{k=0}^{n-1}$.

we construct $\tilde{p}^k = \frac{A^{1/2} p^k}{\sqrt{(p^k)^T A p^k}}$ (wlog. Assume $A = A^T$, Otherwise, Let $\hat{A} = \frac{A+A^T}{2} + \frac{A-A^T}{2}$).

$\Rightarrow (\tilde{p}^k)^T (\tilde{p}^i) = \delta_{ki}$.

Hence

$$
\begin{aligned}
I &= \tilde{p}^0 (\tilde{p}^0)^T + \cdots + \tilde{p}^{n-1} (\tilde{p}^{n-1})^T \\
&= \sum_{k=0}^{n-1} \frac{A^{1/2} p^k (p^k)^T A^{1/2}}{(p^k)^T A p^k} \\
&= A^{1/2} \cdot \sum_{k=0}^{n-1} \frac{p^k (p^k)^T}{(p^k)^T A p^k} \cdot A^{1/2}
\end{aligned}
$$

$$\Rightarrow \quad A^{-1} = \sum_{k=0}^{n-1} \frac{p^k (p^k)^T}{(p^k)^T A p^k}.$$

$$x^* = x^0 + A^{-1}(b - A x_0) = x_0 + \sum_{k=0}^{n-1} \frac{(p^k)^T (b - A x_0)}{(p^k)^T A p^k} \cdot p^k \qquad (2)$$

Compare (1) and (2). remain to show

$$
\begin{aligned}
& (b - A x_0)^T p^k = (b - A x_k)^T p^k \\
\Leftrightarrow \quad & x_0^T A p^k = x_k^T A p^k \\
\Leftrightarrow \quad & p_k^T A (x_k - x_0) = 0 \\
\Leftrightarrow \quad & p_k^T A \left(\sum_{i=0}^{k-1} t_i p^i \right) = 0. \\
\Leftarrow \quad & p_k^T A p^i = 0
\end{aligned}
$$

Hence $x^n = x^*$. $\square$

(b.) We will show that $(r^{k+1})^T p^i = 0$. $i \in \{0, \ldots k\}$.

$$
\begin{aligned}
r^{k+1} = A x^{k+1} - b &= A(x^k + t^k p^k) - b \\
&= A x^k - b + t^k A p^k \\
&= r^k + t^k A p^k.
\end{aligned}
$$

⑥

$$\Rightarrow (r^{k+1})^T p^k = (r^k)^T p^k + t^k (p^k)^T A p^k = (r^k)^T p^k - \frac{(r^k)^T p^k}{(p^k)^T A p^k} (p^k)^T A p^k = 0.$$

As

$$
\begin{aligned}
r^{k+i+1} &= r^{k+i} + t^{k+i} A p^{k+i} \\
&= r^{k+i} + \frac{-(r^{k+i})^T p^{k+i}}{(p^{k+i})^T A p^{k+i}} A p^{k+i}
\end{aligned}
$$

$$\Rightarrow (r^{k+i+1})^T p^k = (r^{k+i})^T p^k - \frac{(r^{k+i})^T p^{k+i}}{(p^{k+i})^T A p^{k+i}} \cdot (p^{k+i})^T A p^k = 0$$

$\Rightarrow (r^{k+1})^T p^i = 0 \quad i \in \{0, 1, \ldots k\}$.

$\Rightarrow (r^{k+1})^T (-r^i + \beta^i p^{i-1}) = 0$

$\Rightarrow (r^{k+1})^T r^i = 0$, $i \in \{0, \ldots k\}$.

$$\Rightarrow \quad t^k = \frac{(-r^k)^T p^k}{(p^k)^T A p^k} = \frac{(-r^k)^T \cdot (-r^k + \beta(k) p^{k-1})}{(p^k)^T U^T U p^k} \qquad (\text{Let } A = U^T U)$$

$$= \frac{(r^k)^T r^k}{(U p^k)^T U p^k}.$$

Hence. Let $z^k = r^k$, $y^k = U p^k$. then $t^k = \dfrac{(z^k)^T z^k}{(y^k)^T y^k}$.

~~Because~~ $\cancel{(p^k)^T r^{k+1} = 0.}$

$$\cancel{\Rightarrow \quad [-r^k + \beta(k) p^{k-1}]^T r^{k+1} = 0.}$$

$$\cancel{\Rightarrow \quad \beta^k = \frac{(r^k)^T r^{k+1}}{(p^{k-1})^T r^{k+1}} = \frac{(r^k)^T \cdot (r^k + t^k A p^k)}{(-r^{k-1} + \beta^{k-1} p^{k-2})^T r^{k+1}}} = \frac{(r^k)^T r^k}{\quad}$$

As $\{p^k\}$ is Conjugate directions

$$(p^k)^T A p^{k-1} = 0$$

$$\Rightarrow \quad (-r^k + \beta(k) p^{k-1})^T A p^{k-1} = 0$$

$$\Rightarrow \quad \beta(k) = \frac{r^{k^T} A p^{k-1}}{(p(k-1))^T A p^{k-1}} = \frac{(r^k)^T (r^k - r^{k-1}) \cdot \frac{1}{t^{k-1}}}{(-r^{k-1} + \beta^{k-1} p^{k-2})^T A p^{k-1}}$$

$$= \frac{(r^k)^T r^k}{(-r^{k-1})^T (r^k - r^{k-1})}$$

$$= \frac{(r^k)^T r^k}{(r^{k-1})^T r^{k-1}}$$

$$= \frac{(z^k)^T z^k}{(z^{k-1})^T (z^{k-1})} \qquad \square.$$

(c) The residual. $r_x^0 = A x^0 - b$.

$$r_y^0 = \cancel{\quad} \hat{A} y^0 - \hat{b} = A x^0 - b = r_x^0.$$

The Conjugate initial direction.

$$p_x^0 = -r_x^0, \quad p_y^0 = -r_y^0 = p_x^0.$$

Hence the initial residual and direction are ~~the~~ same. we finish the proof by showing at each step k. residual $r_x^k = r_y^k$, $p_x^k = p_x^y$.

As $r^{k+1} = r^k + t^k A p^k = r^k - \dfrac{(r^k)^T p^k}{(p^k)^T A p^k} \cdot A p^k = f(r^k, p^k)$

If $r_x^k = r_x^k$, $p_x^k = p_y^k$. then we have $r_x^{k+1} = r_y^{k+1}$,

And $p^k = -r^k + \beta(k) p^{k-1} = -r^k + \dfrac{(r^k)^T r^k}{(r^{k-1})^T (r^{k-1})} \cdot p^{k-1} = g(r^k, r^{k-1}, p^{k-1})$, Hence $p_x^{k+1} = p_y^{k+1}$ if $r_x^{k+1} = r_y^{k+1}$, $r_x^k = r_y^k$, $p_x^k = p_y^k$.

By induction. we know

$$p_x^k = p_y^k, \quad r_x^k = r_y^k \quad \text{holds for } \forall k \in N$$

Because the algorithm terminate when $r^k = 0$.

Hence these two problem are equivalent.

Algorithm for $f(y)$

Initialize $r_y^0 = Ay^0 - \hat{b} = Ax^0 - b$, $p_y^0 = -r_y^0 = b - Ax^0$, $k := 0$, $y^0 = 0$.

repeat

$$t^k = -\frac{(r^k)^T p^k}{(p^k)^T A p^k}$$

~~$x_{k+1} = x_k$~~

~~$y_{k+1} = y_k + t_k$~~ p^k.

$$y^{k+1} = y^k + t^k \cdot p^k.$$

$$r_y^{k+1} = r_y^k + t^k A p^k$$

If $|r_y^{k+1}| < \varepsilon$. then exit.

$$\beta^{k+1} = \frac{(r_y^{k+1})^T r_y^{k+1}}{(r_y^k)^T r_y^k}.$$

$$p_y^{k+1} = -r_y^{k+1} + \beta^{k+1} \cdot p_y^k.$$

$k := k+1$.

end repeat.

return y^{k+1}.

Each step of this algorithm is same as for initial problem. The only difference is the initial point. $x^0 \neq 0$ while $y^0 = 0$.

Hence $y^k - y^0 = x^k - x^0$

$$\Rightarrow \quad y^k = x^k - x^0$$