Exercise 4.

1. Show that if the nonzero vectors $p_0, p_1, \dots, p_\ell$ satisfy

$$p_i^T A p_j = 0 \quad \text{for all } i \neq j,$$

where A is symmetric and positive definite, then these vectors are linearly independent. (This result implies that A has at most n conjugate directions.)

Pf: Assume the converse. $\exists\, p_j = \sum_{i \neq j} c_i p_i$

c_i, $i \in [1:\ell]$, $i \neq j$ are not all $= 0$. w.l.g. assume $c_k \neq 0$.

$$p_k^T A p_j = p_k^T A (\sum c_i p_i) = \sum_{i \neq k,j} c_i p_k^T A p_i + c_k \cdot p_k^T A p_k = c_k \cdot p_k^T A p_k.$$

since $A \succ 0$. $p_k \neq 0$. $p_k^T A p_k > 0$. $c_k \neq 0$. $p_k^T A p_j \neq 0$. $(k \neq j)$. contradiction

2. Verify the formula for the step length in the conjugate gradient method (see page 5, LectureNotes5.pdf):

$$\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}.$$

Pf: $\phi(x) = \frac{1}{2} x^T A x - b^T x$ $\quad \nabla \phi(x) = Ax - b$. i.e. $\nabla \phi_k = r_k$.

find min ϕ among direction p_k. denote. $\psi(\alpha) = \phi(x_k + \alpha p_k) = \phi(x_k) + \alpha \nabla \phi_k^T p_k + \frac{1}{2} \alpha^2 p_k^T A p_k$

let $\psi'(\alpha) = 0 \Rightarrow \alpha = -\dfrac{\nabla \phi_k p_k}{p_k^T A p_k} = -\dfrac{r_k^T p_k}{p_k^T A p_k}$

3. Show that if $f(x)$ is a strictly convex quadratic, then the function

$$h(\sigma) \stackrel{\text{def}}{=} f(x_0 + \sigma_0 p_0 + \cdots + \sigma_{k-1} p_{k-1}),$$

is also a strictly convex quadratic in the variable $\sigma = (\sigma_0, \sigma_1, \dots, \sigma_{k-1})^T$.

$\rightarrow$ we defaultly regard $p_i \in \mathbb{R}^n$.

Pf: $\forall\, P = [p_0, p_1, \dots, p_{k-1}]^T$ $\quad h(\sigma) = f(x_0 + P^T \sigma)$.

set $f(x) = \frac{1}{2} x^T A x - b^T x + c$ $\quad (A \succ 0)$.

$$h(\sigma) = \frac{1}{2} (x_0 + P^T \sigma)^T A (x_0 + P^T \sigma) - b^T (x_0 + P^T \sigma) + c$$

$$= \frac{1}{2} [x_0^T A x_0 + \sigma^T P A x_0 + x_0^T A P^T \sigma + \sigma^T P A P^T \sigma] - b^T x_0 - b^T P^T \sigma + c.$$

$$= f(x_0) + x_0^T A P^T \sigma + \frac{1}{2} \sigma^T (P A P^T) \sigma - (Pb)^T \sigma = f(x_0) + [P(A x_0 - b)]^T \sigma + \frac{1}{2} \sigma^T (P A P^T) \sigma.$$

since $p_0, \dots, p_{k-1}$ are linear independent, P is non-singular $PAP^T \succ 0$.

thus. $h(\sigma) = \frac{1}{2} \sigma^T P A P^T \sigma + [P(Ax_0 - b)]^T \sigma + f(x)$ is strictly convex quadratic

4. Verify from the formulae (for a given x_0)

$$r_0 = Ax_0 - b, \quad p_0 \leftarrow -r_0, \quad \alpha_0 = -\frac{r_0^T p_0}{p_0^T A p_0},$$
$$x_1 = x_0 + \alpha_0 p_0, \quad r_1 = Ax_1 - b, \quad p_1 \leftarrow -r_1 + \frac{r_1^T A p_0}{p_0^T A p_0} p_0$$

that the following relations hold:

$$\mathbf{span}\{r_0, r_1\} = \mathbf{span}\{r_0, Ar_0\}$$
$$\mathbf{span}\{p_0, p_1\} = \mathbf{span}\{r_0, Ar_0\}.$$

Pf: i) $r_1 = Ax_1 - b = A(x_0 + \alpha_0 p_0) - b = Ax_0 + \alpha_0 A p_0 - b = r_0 - \alpha_0 A r_0 \Rightarrow Ar_0 = \frac{1}{\alpha_0}(r_0 - r_1)$.

thus $r_1 \in \text{span}\{r_0, Ar_0\}$ $\quad$ $\text{span}\{r_0, r_1\} \subseteq \text{span}\{r_0, Ar_0\}$

$Ar_0 \in \text{span}\{r_0, r_1\}$ $\quad$ $\text{span}\{r_0, r_1\} \supseteq \text{span}\{r_0, Ar_0\}$

$\Rightarrow \text{span}\{r_0, r_1\} = \text{span}\{r_0, Ar_0\}$.

ii) check $\text{span}\{p_0, p_1\} = \text{span}\{r_0, r_1\}$

$r_0 = -p_0$.

$p_1 = -r_1 + \frac{r_1^T A p_0}{p_0^T A p_0} \cdot p_0 = -r_1 - \frac{r_1^T A p_0}{p_0^T A p_0} r_0$ $\quad$ $p_1 \in \text{span}\{r_0, r_1\}$. $\Rightarrow$ "$\subseteq$" holds.

$r_1 = \frac{r_1^T A p_0}{p_0^T A p_0} \cdot p_0 - p_1$, "$\supseteq$" holds.

5. Let $\{(\lambda_i, v_i)\}_{i=1}^n$ be the eigenpairs of the symmetric matrix A. Show that the eigenvalues and eigenvectors of

$$[I + P_k(A)A]^T A [I + P_k(A)A]$$

are $\lambda_i[1 + \lambda_i P_k(\lambda_i)]^2$ and v_i, respectively.

Pf: $[I + P_k(A)A]\, v_i = v_i + P_k(A) A v_i = v_i + \lambda_i P_k(A) v_i = [1 + \lambda_i P_k(\lambda_i)] v_i$.

$[I + P_k(A)A] A [I + P_k(A)A]^T v_i = [I + P_k(A)A] (1 + \lambda_i P_k(\lambda_i)) A v_i = \lambda_i [1 + \lambda_i P_k(\lambda_i)]^2 v_i$.

that is $\lambda_i[1 + \lambda_i P_k(\lambda_i)]^2$ and v_i are eigenpairs of $[I + P_k(A)A] A [I + P_k(A)A]^T$

6. Consider the problem

$$\min_x x^2 \quad \text{s.t. } c(x) = 0,$$

where

$$c(x) = \begin{cases} x^6 \sin(1/x), & x \neq 0, \\ 0, & x = 0. \end{cases}$$

(a) Show that the constraint function is twice continuously differentiable at all x (including $x = 0$) and that the feasible points are $x = 0$ and $x = 1/(k\pi)$ for all nonzero integers k.

(b) Verify that each feasible point except $x = 0$ is an isolated local solution (see page 5, LectureNotes6.pdf).

(c) Verify that $x = 0$ is a global solution and a strict local solution, but not an isolated local solution.

Pf: (a). $c'(x) = 6x^6 \sin(1/x) - x^4 \cos(1/x)$. $\quad x \neq 0$.

$c''(x) = 30x^4 \sin(1/x) - 10x^3 \cos(1/x) + x^2 \sin(1/x)$.

$c'(0) = \lim\limits_{\Delta x \to 0} \frac{6\Delta x^6 \sin(1/\Delta x) - \Delta x^4 \cos(1/\Delta x)}{\Delta x} = \lim\limits_{\Delta x \to 0} \frac{\Delta x^4 (6\Delta x^2 \sin(1/\Delta x) - \cos(1/\Delta x))}{\Delta x} = 0$

$c''(0) = \lim\limits_{\Delta x \to 0} \frac{\Delta x^2 (\sin(1/\Delta x) \cdot 30\Delta x^2 - 10\Delta x \cos(1/\Delta x) + \sin(1/\Delta x))}{\Delta x} = 0$.

Let $c(x) = 0 \Rightarrow x^6 \sin(1/x) = 0 \Rightarrow x = 0$ or $\sin 1/x = 0 \Rightarrow x = 0, \frac{1}{k\pi}, \; k \in \mathbb{Z}\backslash\{0\}$.

(b). let $\varepsilon_k = \frac{1}{k\pi + \frac{\pi}{3}}$. $\quad x_k = \frac{1}{k\pi}$, in $\mathcal{N}_{\varepsilon_k}(x_k)$. there is no local solution other than x_k.

since in $\mathcal{N}_{\varepsilon_k}(x_k) \backslash \{x_k\}$. no other point s.t. $c(x) = 0$.

(c). let $f(x) = x^2$. $f'(x) = 2x$. $f(x)$ attains minimum at $x=0$ globally. and $x=0$ is feasible.

let $\varepsilon = 1$. $\mathcal{N}_\varepsilon(0)$: $\forall x \in \{\frac{1}{k\pi}\}$. $f(x) = \frac{1}{k^2\pi^2} > 0 = f(0)$. strictly

$\forall \varepsilon > 0$. $\exists k = [\frac{\pi}{\varepsilon}] + 1$ s.t. $x_k = \frac{1}{k\pi} \in \mathcal{N}_\varepsilon(0)$. and x_k is a local sol. (as shown in (b)). thus $x=0$ is not isolated.

7. Consider the constrained optimization problem

$$\min\ (x_2 + 100)^2 + 0.01x_1^2 \quad \textbf{s.t.}\ x_2 - \cos x_1 \geq 0.$$

Does this problem have a finite or infinite number of local solutions? Use the first-order optimality (KKT) conditions (see Theorem 21 (page 2, LectureNotes7.pdf))

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, && (1)\\
c_i(x^*) &= 0, \quad \text{for all } i \in \mathcal{E}, && (2)\\
c_i(x^*) &\geq 0, \quad \text{for all } i \in \mathcal{I}, && (3)\\
\lambda_i^* &\geq 0, \quad \text{for all } i \in \mathcal{I}, && (4)\\
\lambda_i^* c_i(x^*) &= 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}. && (5)
\end{aligned}$$

to justify your answer.

Sol: $\mathcal{L}(x, \lambda) = f(x) - \lambda c(x)$

$$\begin{cases}
\nabla \mathcal{L}_x = \begin{bmatrix} 0.02x_1 + \lambda \sin x_1 \\ 2(100 + x_2) - \lambda \end{bmatrix} = 0 \\
x_2 - \cos x_1 \geq 0 \qquad \lambda \geq 0 \\
\lambda(x_2 - \cos x_1) = 0.
\end{cases}$$

i) $\lambda = 0 \Rightarrow \begin{cases} x_1 = 0 \\ x_2 = -100 \end{cases}$

against $x_2 - \cos x_1 \geq 0$

ii) $\lambda \neq 0 \Rightarrow x_2 = \cos x_1$.

$$\begin{cases} 0.02x_1 + \lambda \sin x_1 = 0 \\ 200 + 2\cos x_1 - \lambda = 0. \end{cases} \Rightarrow 0.02x_1 + 2(\cos x_1 + 100)\sin x_1 = 0$$

Let $g(x) = 0.01x + (100 + \cos x)\sin x$. $|(100 + \cos x)\sin x| \leq 101$. thus. $x \in [10100, +\infty)$ $g > 0$

$x \in (-\infty, 10100]$ $g < 0$

$g(x)$ is smooth on $[-10100, 10100] \Rightarrow$ at most finite number of zero point.

thus, we have finite number $x^* = \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix}$. λ^* s.t. the KKT condition. ; i.e the finite number solution.

8. If f is convex and the feasible region

$$\Omega = \{x \mid c_i(x) = 0,\ i \in \mathcal{E};\ c_i(x) \geq 0,\ i \in \mathcal{I}\}$$

is convex, show that local solutions of

$$\min_{x \in \Omega} f(x)$$

are also global solutions. Moreover, show that the set of global solutions is convex.

Pf: (1). Assume that converse.

$x^* \in \Omega$ is local sol. but $\exists y^* \in \Omega$. s.t. $f(x^*) > f(y^*)$.

by local sol. $\exists \mathcal{N}_\varepsilon(x^*)$. s.t. $x \in \mathcal{N}_\varepsilon(x^*) \Rightarrow f(x) \geq f(x^*)$.

$\exists t$, let $z = tx^* + (1-t)y^* \in \mathcal{N}_\varepsilon(x^*)$. thus. $f(z_t) \leq tf(x^*) + (1-t)f(y^*)$

$f(z_t) = (1-t)(f(y^*) - f(x^*)) - f(x^*)$. the 1st summand < 0 thus $f(z_t) < f(x^*)$.

contradicts with x^* is local minimum.

(2) $\forall x_1, x_2$ be global solution.

$\forall t \in (0,1)$ $\quad f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2) = f(x_1) = f(x_2)$.

which means $tx_1 + (1-t)x_2$ also attains minimum. i.e. is a global sol.

9. **Let $v : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth vector function. Consider the (generally nonsmooth) unconstrained problems**

$$f(x) = \|v(x)\|_\infty = \max_{i=1,\dots,m} |v_i(x)|, \qquad \textbf{and} \qquad f(x) = \max_{i=1,\dots,m} v_i(x).$$

Reformulate these (generally nonsmooth) unconstrained optimization problems as smooth constrained optimization problems.

(1).

$$\min_x f(x) \iff \min_{x,t} f(x,t) = |t| \qquad t - v_i(x) \geq 0 \qquad t + v_i(x) \geq 0 \qquad i \in \{1, 2, \dots m\}.$$

(2)

$$\min_x f(x) \iff \min_{x,t.} f(x,t) = t \qquad t - v_i(x) \geq 0 \quad i \in \{1, 2, \dots, m\}.$$

10. **For Example 1 (page 7, LectureNotes6.pdf) show that the vector**

$$d = -\left(I - \frac{\nabla c_1(x)\nabla c_1(x)^T}{\|\nabla c_1(x)\|^2}\right)\nabla f(x) \Big/ \left\|\left(I - \frac{\nabla c_1(x)\nabla c_1(x)^T}{\|\nabla c_1(x)\|^2}\right)\nabla f(x)\right\|$$

satisfies

$$\nabla c_1(x)^T d = 0, \qquad \nabla f(x)^T d < 0,$$

whenever the first-order condition

$$\nabla f(x^*) = \lambda^* \nabla c_1(x^*)$$

is not satisfied.

Pf: If $\nabla f(x^*) = \lambda^* \nabla c_1(x^*)$ not satisfy. d is well-defined $\left(\|(I - \frac{\nabla c_1 \nabla c_1^T}{\|\nabla c_1\|})\nabla f\| \neq 0\right)$.

(1) $\nabla c_1(x)^T d = -\nabla c_1^T\left(I - \frac{\nabla c_1 \nabla c_1^T}{\|\nabla c_1\|^2}\right)\nabla f / \|\cdot\| = \left(-\nabla c_1^T + \frac{\nabla c_1^T \nabla c_1 \nabla c_1^T}{\|\nabla c_1\|^2}\right)\nabla f / \|\cdot\|$

$= (-\nabla c_1^T + \nabla c_1^T)\nabla f / \|\cdot\| = 0.$

(2). Consider $A = I - \frac{\nabla c_1 \nabla c_1^T}{\|\nabla c_1\|^2}$ $\quad$ we have $A^T = A$.

$A^2 = AA^T = \left(I - \frac{\nabla c_1 \nabla c_1^T}{\|\nabla c_1\|^2}\right)\left(I - \frac{\nabla c_1 \nabla c_1^T}{\|\nabla c_1^2\|}\right)$

$= I - 2\frac{\nabla c_1 \nabla c_1^T}{\|\nabla c_1\|^2} + \frac{\nabla c_1(\nabla c_1^T \nabla c_1)\nabla c_1^T}{\|\nabla c_1\|^4} = I - \frac{\nabla c_1 \nabla c_1^T}{\|\nabla c_1\|^2} = A.$

thus. $\forall v \in \mathbb{R}^n$. $\quad v^T A v = v^T A^T A v = (Av)^T (Av) = \|Av\|^2$

thus $\nabla f(x)^T d = -\frac{\nabla f^T A \nabla f}{\|A \nabla f\|} = -\frac{\|A\nabla f\|^2}{\|A \nabla f\|} = -\|A \nabla f\|^2 < 0.$