CS5321 Numerical Optimization Homework 2

Due Nov 27

1. (20%) Theorem 5.3 in textbook shows many good properties of CG, such as $\vec{p}_{k+1}^T A\vec{p}_i = 0$ and $\vec{r}_{k+1}^T\vec{r}_i = 0$ for $i = 0, 1, 2, ...$ In this problem, you only need to prove some special cases.

   (a) Show that $\vec{p}_2^T A\vec{p}_0 = 0$.

   (b) Show that $\vec{r}_2^T\vec{r}_0 = 0$.

(a)

we first derive the two property below

$$
\begin{aligned}
& r_2^T p_0 = (r_1 - \alpha_1 A p_1)^T p_0 \\
& = r_1^T p_0 - \alpha_1 p_1 A p_0 \\
& = (r_0 - \alpha_0 A p_0)^T p_0 \\
& = p_0^T p_0 - p_0^T p_0 \\
& = 0
\end{aligned}
$$

$$
\begin{aligned}
& p_k^T A p_0 = \frac{1}{\alpha_0} p_k^T p_0 \\
& A p_0 = \frac{1}{\alpha_0} p_0
\end{aligned}
$$

by textbook and the induction above we can have those useful property

$$
\begin{cases}
p_1^T A p_0 = 0 \\
r_0 = p_0 \\
A p_0 = \frac{1}{\alpha_0} p_0 \\
r_2^T p_0 = 0
\end{cases}
$$

$$
\begin{aligned}
& p_2^T A p_0 = (-r_2 + \beta_1 p_1)^T A p_0 \\
& = -r_2^T A p_0 + \beta_1 p_1^T A p_0 \\
& = -r_2^T A p_0 \\
& = -r_2^T \left(\frac{1}{\alpha_0} p_0 \right) \\
& = -\frac{1}{\alpha_0} r_2^T p_0 \\
& = 0
\end{aligned}
$$

(b)

since we know that $p_1^T A p_0 = 0, r_0 = p_0$

$r_2 = r_0 - \alpha_0 A p_0 - \alpha_1 A p_1$

$r_2^T r_0 = (r_0 - \alpha_0 A p_0 - \alpha_1 A p_1)^T r_0$

$= r_0^T r_0 - \alpha_0 p_0^T A r_0 - \alpha_1 p_1^T A r_0$

$= r_0^T r_0 - \alpha_0 p_0^T A r_0 - \alpha_1 p_1^T A p_0$

$= r_0^T r_0 - p_0^T r_0 - \alpha_1 p_1^T A p_0$

$= r_0^T r_0 - r_0^T r_0 - \alpha_1 p_1^T A p_0$

$= r_0^T r_0 - r_0^T r_0 - \alpha_1 p_1^T A p_0$

$= 0$

2. (10%) For BFGS, show that the inverse of B_{k+1}

$$B_{k+1} = B_k - \frac{B_k \vec{s}_k \vec{s}_k^T B_k}{\vec{s}_k^T B_k \vec{s}_k} + \frac{\vec{y}_k \vec{y}_k^T}{\vec{y}_k^T \vec{s}_k}$$

is

$$(I - \rho_k \vec{s}_k \vec{y}_k^T) B_k^{-1} (I - \rho_k \vec{y}_k \vec{s}_k^T) + \rho_k \vec{s}_k \vec{s}_k^T \text{ where } \rho_k = \frac{1}{\vec{y}_k^T \vec{s}_k}.$$

(Hint: only need to show their product is an identity matrix.)

here's some good tools can be used for the simplification

$$\begin{cases} y = Bs \\ s = B^{-1} y \\ y^T s = s^T y \\ s^T = y^T B^{-1} \end{cases}$$

$(I - \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{y}_k^T) B_k^{-1} (I - \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{y}_k \vec{s}_k^T) + \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{s}_k^T$

$= (B_k^{-1} - \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{y}_k^T B_k^{-1})(I - \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{y}_k \vec{s}_k^T) + \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{s}_k^T$

$= (B_k^{-1} - \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{y}_k^T B_k^{-1}) + (-B_k^{-1} \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{y}_k \vec{s}_k^T + \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{y}_k^T B_k^{-1} \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{y}_k \vec{s}_k^T) + \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{s}_k^T$

$= B_k^{-1} - \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{y}_k^T B_k^{-1} - \frac{1}{\vec{y}_k^T \vec{s}_k} B_k^{-1} \vec{y}_k \vec{s}_k^T + (\frac{1}{\vec{y}_k^T \vec{s}_k})^2 \vec{s}_k \vec{y}_k^T B_k^{-1} \vec{y}_k \vec{s}_k^T + \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{s}_k^T$

$\left(B_k - \frac{B_k \vec{s}_k \vec{s}_k^T B_k}{\vec{s}_k^T B_k \vec{s}_k} + \frac{\vec{y}_k \vec{y}_k^T}{\vec{y}_k^T \vec{s}_k} \right) \left((I - \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{y}_k^T) B_k^{-1} (I - \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{y}_k \vec{s}_k^T) + \frac{1}{\vec{y}_k^T \vec{s}_k} \vec{s}_k \vec{s}_k^T \right)$

$=$ here I only give the simplify result I got on pen and paper, it's too complicate to type in latex

$= I + (\frac{yy^T}{y^T s} B^{-1} - \frac{Bsy^T}{y^T s} B^{-1}) + (\frac{Bss^T}{y^T s} - \frac{BB^{-1} ys^T}{y^T s})$

$+ ((\frac{1}{y^T s})^2 Bsy^T B^{-1} ys^T - (\frac{1}{y^T s})^2 yy^T B^{-1} ys^T) + ((\frac{1}{y^T s})^2 yy^T ss^T - (\frac{1}{y^T s})^2 yy^T sy^T B^{-1})$

$+ ((\frac{1}{y^T s})^3 yy^T sy^T B^{-1} ys^T - (\frac{1}{y^T s})^2 \frac{1}{s^T Bs} Bss^T Bsy^T B^{-1} ys^T)$

$$
\begin{aligned}
&+ \left(\frac{1}{y^T s}\frac{1}{s^T Bs} Bss^T BB^{-1} ys^T - \frac{1}{y^T s}\frac{1}{s^T Bs} Bss^T Bss^T\right) \\
&+ \left(\frac{1}{y^T s}\frac{1}{s^T Bs} Bss^T Bsy^T B^{-1} - \frac{1}{s^T Bs} Bss^T BB^{-1}\right) \\
&= I
\end{aligned}
$$

3. (20%) Total least square problem

   (a) Show the distance of a point (x_i, y_i) to the line $ax + by + c = 0$ is $|ax_i + by_i + c|$ where $a^2 + b^2 = 1$.
   we can know the perpendicular vector is (a, b) suppose the length of the vector from (x_i, y_i) to $ax + by + c = 0$ is l we can deduce that

   $$
   \begin{aligned}
   a(x_i + la) + b(y_i + lb) + c &= 0 \\
   (a^2 + b^2)l &= -ax_i - by_i - c \\
   l &= -(ax_i + by_i + c)
   \end{aligned}
   $$

   since the length is positive:

   $$l = |ax_i + by_i + c|$$

   (b) Let $F(a, b, c) = \sum_{i=1}^{m}(ax_i + by_i + c)^2$. Show that $\partial F/\partial c = 0$ implies the point $(\bar{a}, \bar{b})$ is on the line $ax + by + c = 0$.

   $$
   \begin{aligned}
   &\frac{\partial F}{\partial c}\sum_{i=1}^{m}(ax_i + by_i + c)^2 \\
   &= \sum_{i=1}^{m} 2(ax_i + by_i + c) \\
   &= 2\left(a\left(\sum_{i=1}^{m} x_i\right) + b\left(\sum_{i=1}^{m} y_i\right) + mc\right) \\
   &= \frac{2}{m}\left(a\left(\bar{a}\right) + b\left(\bar{b}\right) + c\right) = 0
   \end{aligned}
   $$

   this means that

   $a\left(\bar{a}\right) + b\left(\bar{b}\right) + c = 0$

   implies point $(\bar{a}, \bar{b})$ is on the line $ax + by + c = 0$

4. (30%) Simplex method (the algorithm is shown in Figure 2): Consider the following linear program:

$$
\begin{array}{ll}
\max_{x_1,x_2} & z = 8x_1 + 5x_2 \\
\text{s.t.} & 2x_1 + x_2 \le 1000 \\
& 3x_1 + 4x_2 \le 2400 \\
& x_1 + x_2 \le 700 \\
& x_1 - x_2 \le 350 \\
& x_1, x_2 \ge 0
\end{array}
$$

(a) Transform it to the standard form.

$$-\min \ -c^T x$$
$$s.t. \ Ax = b$$
$$x \geq 0$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}, c = \begin{pmatrix} -8 \\ -5 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} A = \begin{pmatrix} 2 & 1 & 1 & 0 & 0 & 0 \\ 3 & 4 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \end{pmatrix}, b = \begin{pmatrix} 1000 \\ 2400 \\ 700 \\ 350 \end{pmatrix}$$

(b) Suppose the initial guess is $(0, 0)$. Use the simplex method to solve this problem. In each iteration, show

- Basic variables and non-basic variables, and their values.
- Pricing vector.
- Search direction.
- Ratio test result.

step 1:

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, N = \begin{pmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 1 \\ 1 & -1 \end{pmatrix} x_B = \begin{pmatrix} x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}, x_N = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} x = \begin{pmatrix} x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1000 \\ 2400 \\ 700 \\ 350 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{pricing vector: } c_N^T - c_B^T B^{-1} N = \begin{pmatrix} -8 \\ -5 \end{pmatrix}$$

since $-8 < 5$ we choose to decrease x_1 which can decease the value the most
Therefore we set $d_1 = 1$

$$\text{search direction: } d = \begin{pmatrix} d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} -2 \\ -3 \\ -1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$$

$$\text{ratio test: } min \left(\frac{1000}{2}, \frac{2400}{3}, \frac{700}{1}, \frac{350}{1} \right) = 350$$

step2:

$$B = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, N = \begin{pmatrix} 0 & 1 \\ 0 & 4 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} x_B = \begin{pmatrix} x_3 \\ x_4 \\ x_5 \\ x_1 \end{pmatrix}, x_N = \begin{pmatrix} x_6 \\ x_2 \end{pmatrix} x = \begin{pmatrix} x_3 \\ x_4 \\ x_5 \\ x_1 \\ x_6 \\ x_2 \end{pmatrix} = \begin{pmatrix} 300 \\ 1350 \\ 350 \\ 350 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{pricing vector: } c_N^T - c_B^T B^{-1} N = \begin{pmatrix} 8 \\ -13 \end{pmatrix}$$

$$\text{search direction: } d = \begin{pmatrix} d_3 \\ d_4 \\ d_5 \\ d_1 \\ d_6 \\ d_2 \end{pmatrix} = \begin{pmatrix} -3 \\ -7 \\ -2 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

$$\text{ratio test: } min\left(\frac{300}{3}, \frac{1350}{7}, \frac{350}{2}\right) = 100$$

step3:

$$B = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 4 & 1 & 0 & 3 \\ 1 & 0 & 1 & 1 \\ -1 & 0 & 0 & 1 \end{pmatrix}, N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{pmatrix} x_B = \begin{pmatrix} x_2 \\ x_4 \\ x_5 \\ x_1 \end{pmatrix}, x_N = \begin{pmatrix} x_6 \\ x_3 \end{pmatrix} x = \begin{pmatrix} x_2 \\ x_4 \\ x_5 \\ x_1 \\ x_6 \\ x_3 \end{pmatrix} = \begin{pmatrix} 100 \\ 650 \\ 150 \\ 450 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{pricing vector: } c_N^T - c_B^T B^{-1} N = \begin{pmatrix} -\frac{2}{3} \\ \frac{13}{3} \end{pmatrix}$$

$$\text{search direction: } d = \begin{pmatrix} d_2 \\ d_4 \\ d_5 \\ d_1 \\ d_6 \\ d_3 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{-5}{3} \\ \frac{-1}{3} \\ \frac{-1}{3} \\ 1 \\ 0 \end{pmatrix}$$

$$\text{ratio test: } min\left(\frac{650}{\frac{5}{3}}, \frac{150}{\frac{1}{3}}, \frac{150}{\frac{1}{3}}\right) = 390$$

step4:

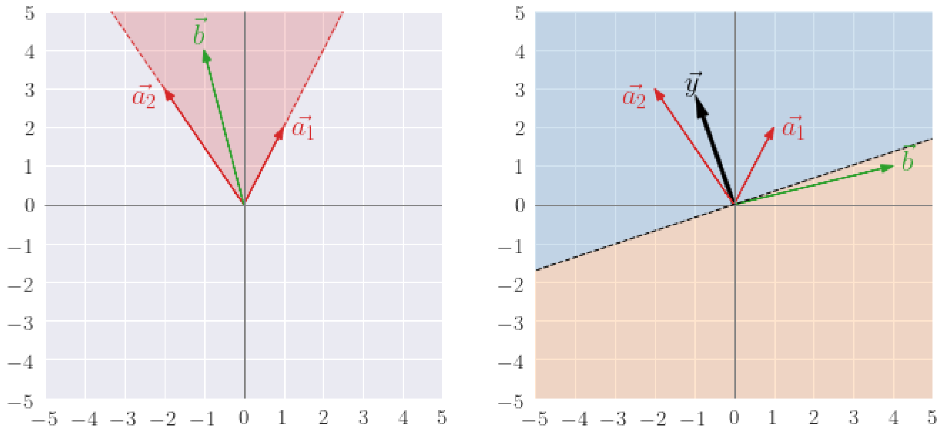
$$B = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 4 & 0 & 0 & 3 \\ 1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix}, N = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} x_B = \begin{pmatrix} x_2 \\ x_6 \\ x_5 \\ x_1 \end{pmatrix}, x_N = \begin{pmatrix} x_4 \\ x_3 \end{pmatrix} x = \begin{pmatrix} x_2 \\ x_6 \\ x_5 \\ x_1 \\ x_4 \\ x_3 \end{pmatrix} = \begin{pmatrix} 360 \\ 390 \\ 20 \\ 320 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{pricing vector: } c_N^T - c_B^T B^{-1} N = \begin{pmatrix} \frac{2}{5} \\ \frac{17}{5} \end{pmatrix}$$

pricing vector is all positive means we found optimal solution $(x_1, x_2) = (320, 360)$
answer $= 8320 + 5360 = 4360$

5. (20%) Farkas lemma: Let A be an $m \times n$ matrix and $\vec{b}$ be an m vector. Prove that exact one of the following two statements is true:

   (a) There exists a $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x} = \vec{b}$ and $\vec{x} \geq 0$.



case (a)

case (b)

Figure 1: Two cases of Farkas Lemma.

(b) There exists a $\vec{y} \in \mathbb{R}^m$ such that $A^T\vec{y} \geq 0$ and $\vec{b}^T\vec{y} < 0$.

(Hint: prove if (a) is true, then (b) cannot be true, and vice versa.)
if (a) is true then (b) is false:

$$Ax = b$$
$$x^TA^T = b^T$$
$$x^TA^Ty = b^Ty$$

since $x \geq 0$ therfore if $A^T\vec{y} \geq 0$ and $\vec{b}^T\vec{y} < 0$ the equation can't holds then we prove that if (a) is true, then (b) cannot be true.

if (a) is not true then (b) is true:
apply the separating hyperplane theorem:
we know that $\alpha^Tb > \beta$ and $\alpha^Ts < \beta$ for $\alpha \neq 0$, $\beta > 0$, s in the cone range of A.
We set $\alpha = -y$
We can get $y^Tb < 0$
And we can know $-y^T\lambda A < \beta$ for any $lambda > 0$ then we can set $\lambda = \infty$
we can get $y^TA > \beta/\lambda$, $A^Ty \geq 0$

Farkas's Lemma (1902) plays an important role in the proof of the KKT condition. The most critical part in the proof of the KKT condition is to show that the Lagrange multiplier $\vec{\lambda}^* \geq 0$ for inequality constraints. We can say if the LICQ condition is satisfied at $\vec{x}^*$, then any feasible direction $\vec{u}$ at $\vec{x}^*$ must have the following properties:

1. $\vec{u}^T \nabla f(\vec{x}^*) \geq 0$ since $\vec{x}^*$ is a local minimizer. (Otherwise, we find a feasible descent direction that decreases f.)
2. $\vec{u}^T \nabla c_i(\vec{x}^*) = 0$ for equality constraints, $c_i = 0$.
3. $\vec{u}^T \nabla c_i(\vec{x}^*) \geq 0$ for inequality constraints, $c_i \geq 0$.

Here is how Farkas Lemma enters the theme. Let $\vec{b}$ be $\nabla f(\vec{x}^*)$, $\vec{y}$ be $\vec{u}$ (any feasible direction at $\vec{x}^*$), the columns of A be $\nabla c_i(\vec{x}^*)$). Since no such $\vec{u}$ exists, according to the properties of $\vec{y}$, statement (a) must hold. The vector $\vec{x}$ in (a) corresponds to $\vec{\lambda}^*$, which just gives us the desired result of the KKT condition.

---

(1) Given a basic feasible point $\vec{x}_0$ and the corresponding index set $\mathcal{B}_0$ and $\mathcal{N}_0$.

(2) For $k = 0, 1, \ldots$

(3) Let $B_k = A(:, \mathcal{B}_k), N_k = A(:, \mathcal{N}_k)$, $\vec{x}_B = \vec{x}_k(\mathcal{B}_k), \vec{x}_N = \vec{x}_k(\mathcal{N}_k)$, and $\vec{c}_B = \vec{c}_k(\mathcal{B}_k), \vec{c}_N = \vec{c}_k(\mathcal{N}_k)$.

(4) Compute $\vec{s}_k = \vec{c}_N - N_k^T B_k^{-1} \vec{c}_B$ (pricing)

(5) If $\vec{s}_k \geq 0$, return the solution $\vec{x}_k$. (found optimal solution)

(6) Select $q_k \in \mathcal{N}_k$ such that $\vec{s}_k(i_q) < 0$, where i_q is the index of q_k in $\mathcal{N}_k$

(7) Compute $\vec{d}_k = B_k^{-1} A_k(:, q_k)$. (search direction)

(8) If $\vec{d}_k \leq 0$, return `unbounded`. (unbounded case)

(9) Compute $[\gamma_k, i_p] = \min\limits_{i, \vec{d}_k(i) > 0} \dfrac{\vec{x}_B(i)}{\vec{d}_k(i)}$ (ratio test)
(The first return value is the minimum ratio;
the second return value is the index of the minimum ratio.)

(10) $x_{k+1}\begin{pmatrix} \mathcal{B} \\ \mathcal{N} \end{pmatrix} = \begin{pmatrix} \vec{x}_B \\ \vec{x}_N \end{pmatrix} + \gamma_k \begin{pmatrix} -\vec{d}_k \\ \vec{e}_{i_q} \end{pmatrix}$
($\vec{e}_{i_q} = (0, \ldots, 1, \ldots, 0)^T$ is a unit vector with i_qth element 1.)

(11) Let the i_pth element in $\mathcal{B}$ be p_k. (pivoting)
$\mathcal{B}_{k+1} = (\mathcal{B}_k - \{p_k\}) \cup \{q_k\}$, $\mathcal{N}_{k+1} = (\mathcal{N}_k - \{q_k\}) \cup \{p_k\}$

Figure 2: The simplex method for solving (minimization) linear programming