CS5321 Numerical Optimization Homework 2

Due Dec 2

1. (20%) Check out the TRUST-REGION NEWTON- LANCZOS METHOD in Section 7.1 in the . What kind of problem it wants to solve? and how the Lanczos method solves it.

   考慮$\min\limits_{x \in R^n} f(x)$，其中$(x)$是定義在$R^n$的二階連續可微函數，
   定義當前點的鄰域(置信域)Ω_k

   $$\Omega_k = \{x \in R^n | \|x - x_k\| \leq \Delta_k\}$$

   這裡Δ_k稱為置信域半徑。
   假定在這個鄰域中，二次模型是目標函數f(x)的一個合適的近似，則在這個鄰域（稱為置信域）中極小化二次模型，得到近似極小點s_k，
   其中

   $$\|s_k\| \leq \Delta_k$$

   置信域方法的模型子問題是

   $$\begin{cases} min \quad q^{(k)}(s) = f(s_k) + g_k^T s + \frac{1}{2} s^T B_k s \\ s.t. \quad \|s_k\| \leq \Delta_k \end{cases}$$

   所以使用Lanczos method，
   假如矩陣不是正定義矩陣，會產生計算步長分母為0時，導致無限大的問題，而Lanczos method可以解決這個問題。
   Assume $A \in R^{nxn}$,is large, sparse and symmetric.
   There exists an orthogonal matrix Q,which transforms A to a tridiagonal matrix T.

   $$Q^T A Q = T \equiv tridiagonal$$

   T= $\begin{bmatrix} a_1 & b_1 & \dots & \dots & 0 \\ c_1 & a_2 & b_2 & \dots & \vdots \\ \vdots & c_2 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & b_{n-1} \\ 0 & 0 & 0 & c_{n-1} & a_n \end{bmatrix}$

   把上面模型子問題中s 替換成$Q_j\omega$
   $\Rightarrow min \quad f(s_k) + e_1^T Q_j (\nabla f_k) e_1^T \omega + \frac{1}{2} \omega^T T_j \omega$
   然後根據

   $$\|r_k\| \leq \eta_k \|\nabla f_k\|$$

   $$\Rightarrow rk = \nabla^2 f_k p_k + \nabla f_k$$

   來決定疊代何時停止。

2. (20%) Check out section 8.2 in the deep learning textbook. Give a summary about the major challenges in neural network optimization.

1. ill-conditioning:
非常普遍的問題，可能導致SGD卡住，即使step很小，也會導致cost function的增加。
Gradient descent step$\longrightarrow -\epsilon g$
add to the cost function $\longrightarrow -\epsilon g^T g + \frac{1}{2}\epsilon^2 g^T Hg$
當 $\frac{1}{2}\epsilon^2 g^T Hg > \epsilon g^T g$ ill-conditioning會是一個問題，
儘管有strong gradient，學習速度還是會很慢。
2. Local Minima 局部最小值:
由於神經網路的non-convexity性質，可能會有許多local minima。
但是這個不是主要問題。
3. Plateaus, Saddle Points etc
比起local minima/maxima更常見的問題是:
$\longrightarrow$zero gradient points: saddle points
• 在saddle時, Hessian 會有正值和負值
正值: cost greater than saddle point
負值: values have lower value
• 在低維度時: $\longrightarrow$Local minima are more common
• 在高維度時: $\longrightarrow$Local minima are rare, saddle points more common
主要的問題是saddle points，而不是multiple minima
大部分的訓練時間花在穿越flat valley of the Hessian matrix。
4. Cliffs and Exploding Gradients:
多重神經網路有類似cliffs的陡峭區域，因為multiplying several large weights
例如:RNN
這樣會導致參數的更新會跳到極遠，就像從懸崖跳下一樣。
5. Long-Term Dependencies:
像是RNN可能會有許多layers，這會讓computational graphs become extremely deep，這樣一直重複使用相同的參數會帶來困難。
6. Inexact Gradients:
在實際中，我們有noisy 或是biased estimate，因為深度學習演算法都依賴樣本抽樣的估計。
7. Poor Correspondence between Local and Global Structure:
可能很難走出一步，如果J(θ) is poorly conditioned當θ在不好的點，
例如: cliff或是saddle point。
就算解決上面問題，表現也可能不好，因為當改進最多的方向不是指向much lower cost區域。
8. Theoretical Limits of Optimization:
有些演算法只適合output discrete values，但是基本上大部分的神經網路output是smoothly increasing values。

3. (10%) Let J be an $m \times n$ matrix, $m \geq n$. Show that J has full column rank if and only if $J^T J$ is positive definite.

J 是m $\times$ n matrix ,m $\geq$ n
Since the fact that $J^T J$ is positive definite implies that it is nonsingular, we only need to prove two claims.
One of them is that the fact that J has full column rank implies $J^T J$ is positive definite; the other one is that the fact that $J^T J$ is nonsingular implies J has full column rank.
(1) If J has full column rank

$\Rightarrow Jx \neq 0, \quad \forall x \in R^n, x \neq 0$，所以
$\Rightarrow x^T(J^TJ)x > 0$
$\Rightarrow (Jx)^T Jx > 0, \forall x \neq 0$
$\Rightarrow J^TJ$is positive definite.
(2) 利用反證法Assume that J dosen't have full column rank.
存在$y \in R^n, y \neq 0$
$\Rightarrow s.t. Jy = 0$,Hence
$\Rightarrow J^TJy = 0$
$\Rightarrow J^TJy$ is singular 這是矛盾的
$\Rightarrow J^TJy$ is nonsingular $\Rightarrow$ J has full column rank

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

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

(a) Transform it to the standard form.

(a) 令$\vec{x}= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\vec{c}= \begin{bmatrix} 8 \\ 5 \end{bmatrix}$, A$= \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$, $\vec{b}= \begin{bmatrix} 1000 \\ 2400 \\ 700 \\ 350 \end{bmatrix}$
$\Rightarrow \underset{max}{\vec{x}} \quad z = \vec{c}^T\vec{x}$
$\Rightarrow s.t. \quad A\vec{x} \leq \vec{b}$
$\Rightarrow \vec{x} \geq 0$
轉換為standard form，將不等式變成等式

$$\Rightarrow \begin{cases} 2x_1 + x_2 + x_3 = 1000 \\ 3x_1 + 4x_2 + x_4 = 2400 \\ x_1 + x_2 + x_5 = 700 \\ x_1 - x_2 + _6 = 350 \end{cases}$$

x_3, x_4, x_5, x_6 are slack variables
所以

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

$$\Rightarrow \underset{-min}{\vec{x}} \quad z = \begin{bmatrix} -8 \\ -5 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}$$

$$\text{s.t.} \quad \begin{bmatrix} 2 & 1 & 1 & 0 & 0 & 0 \\ 3 & 4 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1000 \\ 2400 \\ 700 \\ 350 \end{bmatrix}, \vec{x} \geq 0$$

(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.

第1次iteration, k = 0:

$\beta_0 = \{3, 4, 5, 6\}$, $N_0 = \{1, 2\}$

$$B_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, N_0 = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\vec{x_B} \text{ ((Basic variables)} = \begin{bmatrix} x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = B_0^{-1}\vec{b} = \begin{bmatrix} 1000 \\ 2400 \\ 700 \\ 350 \end{bmatrix}$$

$$\vec{x_N} \text{(non-basic variables)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\vec{c_B} = \vec{c}(\beta_0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \vec{c_N} = \vec{c}(N_0) = \begin{bmatrix} -8 \\ -5 \end{bmatrix}$$

$$\Rightarrow \vec{s_0} \text{(Price vector)} = \vec{c_N} - (B_0^{-1}N_0)^T\vec{c_B} = \vec{c_N} - N_0^T(B_0^{-1})^T\vec{c_B}$$

$$= \begin{bmatrix} -8 \\ -5 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}^T \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \right)^T \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -8 \\ -5 \end{bmatrix}$$

$$\Rightarrow q_0 = 1, \quad i_q = 1$$

$$\Rightarrow s_0 < 0, \quad \vec{d_0} \text{(search direction)} = B_0^{-1}N(:, q_0) = B_0^{-1} \begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \end{bmatrix}$$

$[\gamma_0, \quad i_p]$(Ratio test)$= \underset{i, \vec{d_0}>0}{min} \ \frac{\vec{x_B(i)}}{\vec{d_0(i)}} = \underset{i, \vec{d_0}>0}{min} \begin{bmatrix} 500 \\ 800 \\ 700 \\ 350 \end{bmatrix}$

$\Rightarrow \gamma_0 = 350, \quad i_p = 6, \quad p_0 = 4$

so, $\vec{x_1} = \begin{bmatrix} B_0 \\ N_0 \end{bmatrix} = \begin{bmatrix} x_B \\ x_N \end{bmatrix} + \gamma_0 \begin{bmatrix} -\vec{d_0} \\ \vec{e_{i_q}} \end{bmatrix}$

$= \begin{bmatrix} 0 \\ 0 \\ 1000 \\ 2400 \\ 700 \\ 350 \end{bmatrix} +350 \begin{bmatrix} 1 \\ 0 \\ -2 \\ -3 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 350 \\ 0 \\ 300 \\ 1350 \\ 350 \\ 0 \end{bmatrix}$

第2次iteration, k = 1:
$\beta_1 = \{3, 4, 5, 6\}$, $N_1 = \{1, 2\}$

$B_1 = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, N_1 = \begin{bmatrix} 0 & 1 \\ 0 & 4 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}$

$\vec{x_B} = \begin{bmatrix} x_3 \\ x_4 \\ x_5 \\ x_1 \end{bmatrix} = \vec{x_1} \ (\beta_1) = \begin{bmatrix} 300 \\ 1350 \\ 350 \\ 350 \end{bmatrix} = B_1^{-1}\vec{b}$

$\vec{x_N} \ = \begin{bmatrix} x_6 \\ x_2 \end{bmatrix} = x_1 \ (\vec{N_1}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\vec{c_B} = \vec{c}(\beta_1) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -8 \end{bmatrix}, \ \vec{c_N} = \vec{c}(N_1) = \begin{bmatrix} 0 \\ -5 \end{bmatrix}$

$\Rightarrow \vec{s_1} = \vec{c_N} - N_1^T (B_1^{-1})^T \vec{c_B}$

$= \begin{bmatrix} 0 \\ -5 \end{bmatrix}$ - $\begin{bmatrix} 0 & 1 \\ 0 & 4 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}^T \left(\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \right)^T \begin{bmatrix} 0 \\ 0 \\ 0 \\ -8 \end{bmatrix} = \begin{bmatrix} 8 \\ -13 \end{bmatrix}$

$\Rightarrow q_1 = 2, \quad i_q = 2$

$\Rightarrow s_1 < 0, \quad \vec{d_1} = B_1^{-1} N(:, q_1) = B_1^{-1} \begin{bmatrix} 1 \\ 4 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \\ 2 \\ -1 \end{bmatrix}$

$$[\gamma_1, \quad i_p] = \min_{i, \vec{d_1}>0} \frac{\vec{x_B(i)}}{\vec{d_1(i)}} = \min_{i, \vec{d_1}>0} \begin{bmatrix} 100 \\ \frac{1350}{7} \\ 175 \\ -350 \end{bmatrix}$$

$$\Rightarrow \gamma_1 = 100, \quad i_p = 3, \quad p_1 = 1$$

so, $\vec{x_2} = \begin{bmatrix} B_1 \\ N_1 \end{bmatrix} = \begin{bmatrix} x_B \\ x_N \end{bmatrix} + \gamma_1 \begin{bmatrix} -\vec{d_1} \\ \vec{e_{i_q}} \end{bmatrix}$

$$= \begin{bmatrix} 350 \\ 0 \\ 300 \\ 1350 \\ 350 \\ 0 \end{bmatrix} + 100 \begin{bmatrix} 1 \\ 1 \\ -3 \\ -7 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 450 \\ 100 \\ 0 \\ 650 \\ 150 \\ 0 \end{bmatrix}$$

第3次iteration, k = 2:

$\beta_2 = \{2, 4, 5, 1\}$, $N_2 = \{6, 3\}$

$$B_2 = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 4 & 1 & 0 & 3 \\ 1 & 0 & 1 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \; N_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$\vec{x_B} = \begin{bmatrix} x_2 \\ x_4 \\ x_5 \\ x_1 \end{bmatrix} = \vec{x_2}\,(\beta_2) = \begin{bmatrix} 100 \\ 650 \\ 150 \\ 450 \end{bmatrix} = B_2^{-1}\vec{b}$$

$$\vec{x_N} = \begin{bmatrix} x_6 \\ x_3 \end{bmatrix} = x_2\,(\vec{N}_2) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\vec{c_B} = \vec{c}(\beta_2) = \begin{bmatrix} -5 \\ 0 \\ 0 \\ -8 \end{bmatrix}, \; \vec{c_N} = \vec{c}(N_2) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \vec{s_2} = \vec{c_N} - N_2^T (B_2^{-1})^T \vec{c_B}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}^T \left(\begin{bmatrix} 1 & 0 & 0 & 2 \\ 4 & 1 & 0 & 3 \\ 1 & 0 & 1 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix}^{-1} \right)^T \begin{bmatrix} -5 \\ 0 \\ 0 \\ -8 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} \\ \frac{13}{3} \end{bmatrix}$$

$$\Rightarrow q_2 = 1, \quad i_q = 6$$

$$\Rightarrow s_2 < 0, \quad \vec{d_2} = B_2^{-1} N(:, q_2) = B_2^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} \\ \frac{5}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$$

$$[\gamma_2, \quad i_p] = \min_{i, \vec{d_2} > 0} \frac{\vec{x_B(i)}}{\vec{d_2(i)}} = \min_{i, \vec{d_2} > 0} \begin{bmatrix} -150 \\ 390 \\ 450 \\ 1350 \end{bmatrix}$$

$\Rightarrow \gamma_2 = 390, \quad i_p = 4, \quad p_2 = 2$

so, $\vec{x_3} = \begin{bmatrix} B_2 \\ N_2 \end{bmatrix} = \begin{bmatrix} x_B \\ x_N \end{bmatrix} + \gamma_2 \begin{bmatrix} -\vec{d_2} \\ \vec{e_{i_q}} \end{bmatrix}$

$$= \begin{bmatrix} 450 \\ 100 \\ 0 \\ 650 \\ 150 \\ 0 \end{bmatrix} + 390 \begin{bmatrix} -\frac{1}{3} \\ \frac{2}{3} \\ 0 \\ -\frac{5}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 320 \\ 360 \\ 0 \\ 00 \\ 20 \\ 390 \end{bmatrix}$$

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 $\vec{x} \geq 0$.

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

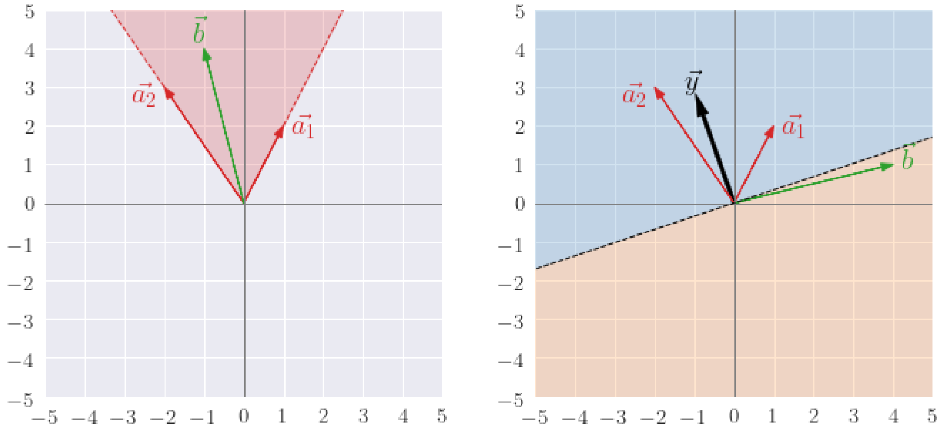
   (Hint: prove if (a) is true, then (b) cannot be true, and vice versa.)

1.
Assume(a) is true.
Assume Ax = b and x $\geq$ 0.
If $A^T y \geq 0$ then if x ≥ 0
$\Rightarrow$ we have that $(x^T A^T)y \geq 0 \Rightarrow b^T y \geq 0$.
Since $Ax = b$ this implies that $b^T y \geq 0$,
thus it cannot be that both $A^T y \geq 0$and $b^T y < 0$.
(b) is not true.

2.
Assume(a)is False.
Define

$$C = \{q \in R^m : \quad \exists x \geq 0, Ax = q\};$$

Notice that C is a convex set: for $q1, q2 \in C$
there exist $x_1, x_2 \quad s.t \quad q_1 = Ax_1 \quad and \quad q_2 = Aq_2$
and for any $\lambda \in [0, 1]$
we have that $\lambda q_2 + (1 - \lambda)q_2 = \lambda Aq_1 + (1 - \lambda)Aq_2 = A(\lambda q_1 + (1 - \lambda)q_2)$
and hence $\lambda q_2 + (1 - \lambda)q_2 \in C$.
Since (a) is false $b \notin C$.From the separating hyperplane theorem, we know there exists $y \in R^m$
$\Rightarrow s.t. \quad q^T y \geq 0 \quad and \quad b^T y < 0$, for all $q \in C$
Since $q = Ax$ that implies that $\forall x \geq 0$
we have that $x^T A^T y \geq 0$ and $b^T y < 0$,
$\Rightarrow A^T y \geq 0$ and $b^T y < 0$, as required.
(b) is True.



case (a) case (b)

Figure 1: Two cases of Farkas Lemma.

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