ACM 113 Introduction to Optimization - Problem Set 6

Ling Li, ling@cs.caltech.edu

June 5, 2001

6.1 *Linear programming.* Let the Lagrange multipliers $\lambda = \binom{y}{s}$, where y is for constraints $Ax - b = 0$ and s for $x \geq 0$. Then the Lagrangian function is

$$\mathcal{L}(x, \lambda) \equiv c^T x - y^T(Ax - b) - s^T x.$$

For optimal pair (x^*, λ^*), the 1st order KKT conditions are

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*) = c - A^T y^* - s^* &= 0, \\ s^* &\geq 0, \\ s_i^* x_i^* &= 0, \quad \forall i \in \mathcal{I}. \end{aligned}$$

For feasible point x^* satisfying above conditions, x^* must be an optimal solution since the above conditions, together with feasibility conditions, constitute the primal-dual LP optimality conditions: $A^T y + s = c$, $Ax = b$, $x \geq 0$, $s \geq 0$, and $x^T s = 0$.

Trust region subproblem. The problem is to minimize $m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p$ subject to $\|p\| \leq \Delta_k$, where B_k is symmetric but not necessarily positive definite. The constraint can be written as

$$c(p) = \Delta_k^2 - p^T p \geq 0.$$

Thus the Lagrangian function is

$$\mathcal{L}(p, \lambda) \equiv f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p - \lambda(\Delta_k^2 - p^T p).$$

The 1st order KKT conditions for optimal pair (p^*, λ^*) are

$$\begin{aligned} \nabla_p \mathcal{L}(p^*, \lambda^*) = \nabla f_k + B_k p^* + 2\lambda^* p^* &= 0, \\ \lambda^* &\geq 0, \\ \lambda^*(\Delta_k^2 - p^{*T} p^*) &= 0. \end{aligned}$$

If B_k is positive definite, and $\left\| B_k^{-1} \nabla f_k \right\| \leq \Delta_k$, then $p^* = -B_k^{-1} \nabla f_k$ and $\lambda^* = 0$ satisfy the above conditions.

6.2 Let x_1 and x_2 denote the length and width of the rectangle. The perimeter is $2(x_1 + x_2)$. The problem is formulated as

$$\begin{aligned} \min \quad & f(x) = -2(x_1 + x_2), \\ \text{subject to} \quad & x_1, x_2 \geq 0, \qquad (1) \\ & x_1^2 + x_2^2 - 16 = 0. \qquad (2) \end{aligned}$$

The Lagrangian function is

$$\mathcal{L}(x,\lambda) \equiv -2(x_1+x_2) - \lambda_1 x_1 - \lambda_2 x_2 - \lambda_3(x_1^2+x_2^2-16).$$

From the 1st order KKT conditions, the optimal (x^*, λ^*) satisfies

$$\begin{aligned} \nabla_x \mathcal{L}(x^*,\lambda^*) = \begin{pmatrix} -2-\lambda_1^*-2\lambda_3^* x_1^* \\ -2-\lambda_2^*-2\lambda_3^* x_2^* \end{pmatrix} &= 0, \\ \lambda_1^* x_1^* = \lambda_2^* x_2^* &= 0, \\ \lambda_1^*, \lambda_2^* &\geq 0. \end{aligned}$$

Thus $0 = (-2-\lambda_1^*-2\lambda_3^* x_1^*)\lambda_1^* = -\lambda_1^*(\lambda_1^*+2)$, i.e., $\lambda_1^* = 0$. Similarly, $\lambda_2^* = 0$. So $x_1^* = x_2^* = -\frac{1}{\lambda_3^*}$. From constraints (1) and (2), we have $x_1^* = x_2^* = 2\sqrt{2}$, and $\lambda_3^* = -\frac{1}{2\sqrt{2}}$. We can verify the optimality since

$$\nabla_{xx}\mathcal{L}(x^*,\lambda^*) = \begin{pmatrix} -2\lambda_3^* & 0 \\ 0 & -2\lambda_3^* \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$$

is positive definite.

Among constraints for the original problem, only (2) is active. Consider the perturbed problem

$$\begin{aligned} \min \quad & f(x) = -2(x_1+x_2), \\ \text{subject to} \quad & x_1^2 + x_2^2 - 16 = \epsilon. \end{aligned}$$

For $|\epsilon|$ small enough, the optimal solution is $x_1^*(\epsilon) = x_2^*(\epsilon) = \sqrt{\frac{16+\epsilon}{2}}$. The primal function is

$$p(\epsilon) = f(x^*(\epsilon)) = -4\sqrt{\frac{16+\epsilon}{2}}.$$

The derivative of $p(\epsilon)$ is

$$\nabla p(\epsilon) = -\sqrt{\frac{2}{16+\epsilon}}.$$

Thus $\nabla p(0) = -\frac{1}{2\sqrt{2}} = \lambda_3^*$.

6.3 Write the problem in the form of constrained nonlinear programming:

$$\begin{aligned} \min \quad & f(x) = -\pi h^2(r-h/3) \\ \text{subject to} \quad & c_1(x) = 2\pi r h - S = 0, \\ & c_2(x) = r \geq 0, \\ & c_3(x) = h \geq 0, \end{aligned}$$

where $x = (r,h)^T$, and $S > 0$ is the required spherical area. Let $\lambda = (\lambda_1, \lambda_2, \lambda_3)^T$ denote the Lagrange multipliers. The Lagrangian function is

$$\mathcal{L}(x,\lambda) \equiv f(x) - \lambda^T c(x) = -\pi h^2(r-h/3) - \lambda_1(2\pi r h - S) - \lambda_2 r - \lambda_3 h.$$

From the 1st order KKT conditions, the optimal pair (x^*, λ^*) satisfies

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*) = \begin{pmatrix} -\pi h^*(h^* + 2\lambda_1^*) - \lambda_2^* \\ \pi h^{*2} - 2\pi h^* r^* - 2\pi \lambda_1^* r^* - \lambda_3^* \end{pmatrix} &= 0, \\ \lambda_2^*, \lambda_3^* &\geq 0, \\ \lambda_2^* r^* = \lambda_3^* h^* &= 0, \end{aligned}$$

and $c_1(x^*) = 2\pi r^* h^* - S = 0$. Since $S > 0$, we get $r^* > 0$, $h^* > 0$. Thus $\lambda_2^* = \lambda_3^* = 0$. Only c_1 is active. Then we get $r^* = h^* = -2\lambda_1^* = \sqrt{\frac{S}{2\pi}}$.

We also have

$$\nabla_{xx} \mathcal{L}(x^*, \lambda^*) = \begin{pmatrix} 0 & -2\pi(h^* + \lambda_1^*) \\ -2\pi(h^* + \lambda_1^*) & 2\pi(h^* - r^*) \end{pmatrix} = \begin{pmatrix} 0 & -\pi h^* \\ -\pi h^* & 0 \end{pmatrix}.$$

For any 1st order feasible direction $d = (d_r, d_h)^T$ such that $d \neq 0$, and

$$\nabla c_1(x^*)^T d = 2\pi h^* d_r + 2\pi r^* d_h = 0,$$

we have $d_r = -d_h \neq 0$. Thus

$$d^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) d = -2\pi h^* d_r d_h = \sqrt{2\pi S} d_h^2 > 0.$$

Hence (from the sufficient conditions) $r^* = h^* = \sqrt{\frac{S}{2\pi}}$ is the only strict local minimizer, and thus maximizes the segment volume (the maximum is $\frac{2\pi}{3}\left(\sqrt{\frac{S}{2\pi}}\right)^3 = \frac{S^{3/2}}{3\sqrt{2\pi}}$).

6.4 The logarithmic barrier method solves the unconstrained problem

$$x(\mu) = \arg\min_x B(x, \mu) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - \mu \ln(x_1 + x_2 - 1), \tag{3}$$

and seeks convergence as $\mu \downarrow 0$. To solve (3), let

$$\nabla_x B(x, \mu) = \begin{pmatrix} x_1 - \frac{\mu}{x_1 + x_2 - 1} \\ x_2 - \frac{\mu}{x_1 + x_2 - 1} \end{pmatrix} = 0.$$

Since $x_1 + x_2 \geq 1$, we get $x_1 = x_2 = \frac{1 + \sqrt{1 + 8\mu}}{4}$. Thus when $\mu \downarrow 0$,

$$x(\mu) = \left(\frac{1 + \sqrt{1 + 8\mu}}{4}, \frac{1 + \sqrt{1 + 8\mu}}{4} \right)^T \to \left(\frac{1}{2}, \frac{1}{2} \right)^T.$$

The Lagrange multiplier estimate is defined as

$$\lambda(\mu) = \frac{\mu}{c(x(\mu))} = \frac{\mu}{\frac{1 + \sqrt{1 + 8\mu}}{2} - 1} = \frac{1 + \sqrt{1 + 8\mu}}{4}.$$

Thus $\lambda(\mu) \to \frac{1}{2}$ when $\mu \downarrow 0$. So we get $x^* = (\frac{1}{2}, \frac{1}{2})^T$ and $\lambda^* = \frac{1}{2}$.

The Hessian of $B(x, \mu)$ is

$$\nabla_{xx} B(x, \mu) = \begin{pmatrix} 1 + \frac{\mu}{(x_1 + x_2 - 1)^2} & \frac{\mu}{(x_1 + x_2 - 1)^2} \\ \frac{\mu}{(x_1 + x_2 - 1)^2} & 1 + \frac{\mu}{(x_1 + x_2 - 1)^2} \end{pmatrix}.$$

The condition number of the Hessian is

$$\kappa = 1 + \frac{2\mu}{(x_1 + x_2 - 1)^2}.$$

For the Hessian at $x(\mu)$, the condition number is

$$\kappa(\mu) = 1 + \frac{2\mu}{\left(\frac{1+\sqrt{1+8\mu}}{2} - 1\right)^2} = 1 + \frac{1}{2\mu}\left(\frac{1+\sqrt{1+8\mu}}{2}\right)^2.$$

When μ is very close to 0, $\kappa(\mu) \approx \frac{1}{2\mu}$ becomes very large, and $\kappa(\mu) \to +\infty$ when $\mu \downarrow 0$.

6.5 $f(x) = \frac{1}{2}x^TQx + \theta b^Tx$ and $c(x) = b^Tx = 0$.

- Since Q is positive definite on the subspace $b^Tx = 0$, the optimal solution is $x^* = 0$. Thus $\nabla f(x^*) = Qx^* + \theta b = \theta b$, $\nabla c(x^*) = b$. So $\lambda^* = \theta$.
- The augmented Lagrangian for this problem is ($a > 0$)

$$\mathcal{L}_a(x, \lambda) \equiv f(x) - \lambda^T c(x) + \frac{a}{2}\|c(x)\|^2 = \frac{1}{2}x^T(Q + abb^T)x + (\theta - \lambda)b^Tx. \tag{4}$$

For $\lambda = \lambda_k \neq \theta$, if (4) has a minimum, $Q + abb^T$ must be invertible.* Then from

$$(Q + abb^T)Q^{-1}b = b + abb^TQ^{-1}b = (1 + ab^TQ^{-1}b)b,$$

we have $Q^{-1}b = (1 + ab^TQ^{-1}b)(Q + abb^T)^{-1}b$. Also from $QQ^{-1}b = b \neq 0$, we know $Q^{-1}b \neq 0$. Thus $1 + ab^TQ^{-1}b \neq 0$, and

$$(Q + abb^T)^{-1}b = \frac{Q^{-1}b}{1 + ab^TQ^{-1}b}.$$

The 1st order condition $\nabla_x\mathcal{L}_a(x_k, \lambda_k) = (Q + abb^T)x_k + (\theta - \lambda_k)b = 0$ gives

$$x_k = (\lambda_k - \theta)(Q + abb^T)^{-1}b = \frac{(\lambda_k - \theta)Q^{-1}b}{1 + ab^TQ^{-1}b}. \tag{5}$$

So, if (4) has a minimum for $\lambda_k \neq \theta$, it is given by (5). (If $\lambda_k = \theta$, there may be many minimizers.)

- We would prove

$$x_k = -\frac{\theta Q^{-1}b}{(1 + ab^TQ^{-1}b)^k}, \quad \lambda_k = -\theta\left[\frac{1}{(1 + ab^TQ^{-1}b)^{k-1}} - 1\right], \tag{6}$$

by induction. For $k = 1$, by initialization, $\lambda_1 = 0 \neq \theta$, and by (5),

$$x_1 = \frac{(0 - \theta)Q^{-1}b}{1 + ab^TQ^{-1}b} = -\frac{\theta Q^{-1}b}{1 + ab^TQ^{-1}b}.$$

*Otherwise there exists $\hat{x} \neq 0$ such that $(Q + abb^T)\hat{x} = 0$. Since Q is positive definite on the subspace $b^Tx = 0$, if $b^T\hat{x} = 0$, $\hat{x}^T(Q + abb^T)\hat{x} = \hat{x}^TQ\hat{x} > 0$, contradicting $(Q + abb^T)\hat{x} = 0$. So $b^T\hat{x} \neq 0$. Thus $\mathcal{L}_a(k\hat{x}, \lambda_k) = k\left[(\theta - \lambda_k)b^T\hat{x}\right]$, where k is a scalar. Clearly $\mathcal{L}_a$ has no minimum.

So (6) holds for $k = 1$.

Suppose (6) holds for k. Then we estimate the multiplier as

$$\begin{aligned}\lambda_{k+1} &\equiv \lambda_k - a \cdot c(x_k) \\ &= -\theta \left[\frac{1 + ab^T Q^{-1} b}{(1 + ab^T Q^{-1} b)^k} - 1 \right] + ab^T \frac{\theta Q^{-1} b}{(1 + ab^T Q^{-1} b)^k} \\ &= -\theta \left[\frac{1}{(1 + ab^T Q^{-1} b)^k} - 1 \right].\end{aligned}$$

We still have $\lambda_{k+1} \neq \theta$. Then by (5),

$$x_{k+1} = \frac{\frac{-\theta}{(1+ab^T Q^{-1} b)^k} Q^{-1} b}{1 + ab^T Q^{-1} b} = -\frac{\theta Q^{-1} b}{(1 + ab^T Q^{-1} b)^{k+1}}.$$

So (6) also holds for $k + 1$.

- The convergences of x_k to x^* and λ_k to λ^* require $\left|1 + ab^T Q^{-1} b\right| > 1$. If $b^T Q^{-1} b > 0$, a can be any positive number; if $b^T Q^{-1} b < 0$, then $a > -\frac{2}{b^T Q^{-1} b}$; if $b^T Q^{-1} b = 0$, no a satisfies the convergence requirement. If x_k, λ_k converge, the larger a is, the faster they converge. However, since a is fixed, the convergence rate is always linear.

6.6 *Underneath idea*: Let $y^{(k)}$, $s^{(k)}$ be the feasible pair at step k for the dual problem. Let B, N represent basic and nonbasic variables for the primal problem. From $A^T y + s = c$,

$$s_N = c_N - (B^{-1} N)^T (c_B - s_B).$$

For step k, we have $s_N^{(k)} \geq 0$, $s_B^{(k)} = 0$, and $s_N^{(k)} = c_N - (B^{-1} N)^T c_B$. So

$$s_N = s_N^{(k)} + (B^{-1} N)^T s_B. \tag{7}$$

If x_p ($p \in B$) is a variable that doesn't satisfy the feasible condition for the primal problem, that is, $x_p < 0$, we need to find another variable x_q ($q \in N$) to replace x_p in the basis. Since

$$s_B^{(k+1)} = (0, \ldots, 0, s_p^{(k+1)}, 0, \ldots, 0), \tag{8}$$

denoting row p of $B^{-1} N$ by t, from (7) we get

$$s_N^{(k+1)} = s_N^{(k)} + t^T s_p^{(k+1)}. \tag{9}$$

If $t \geq 0$, then $s_p^{(k+1)}$ is unbounded, and the primal problem is infeasible. Otherwise, since we need to maintain $s^{(k+1)} \geq 0$, $s_p^{(k+1)}$ should satisfy

$$s_p^{(k+1)} = \min_{\substack{i \in N \\ t_i < 0}} \left\{ -\frac{s_i^{(k)}}{t_i} \right\}.$$

Algorithm: Suppose initially there exists an initial dual-feasible pair $y^{(0)}$, $s^{(0)}$. B and N represent the basic and nonbasic variables. $k = 0$.

(a) Calculate $x_B = B^{-1}b$. If $x_B \geq 0$, do optimal print and stop.

(b) Select $p \in B$ with $x_p < 0$. Calculate $t = (B^{-1}N)_{\text{row } p}$, or $t = (B^{-1})_{\text{row } p}N$.

(c) If $t \geq 0$, the primal problem is infeasible. Stop.

(d) Calculate

$$q = \arg\min_{\substack{i \in N \\ t_i < 0}} \left\{ -\frac{s_i^{(k)}}{t_i} \right\},$$

and $s_p^{(k+1)} = -\frac{s_q^{(k)}}{t_q}$. Update s using (8) and (9). Update B, N by removing p and adding q to B.

(e) $k \leftarrow k + 1$. Goto (a).

Example. For LP in the problem, add two surplus variables x_3 and x_4. The LP becomes

$$\begin{array}{rl} \min & z = 5x_1 + 4x_2 \\ \text{subject to} & 4x_1 + 3x_2 - x_3 = 10, \\ & 3x_1 - 5x_2 - x_4 = 12, \\ & x_1, x_2, x_3, x_4 \geq 0. \end{array}$$

Consider the initial basis $x_B = (x_3, x_4)^T$. Multiplying the constraints by -1, we obtain

basic	x_1	x_2	x_3	x_4	rhs	
$-z$	5	4	0	0	0	
x_3	-4	-3	1	0	-10	
x_4	$\boxed{-3}$	5	0	1	-12	$\Leftarrow$

$\rightarrow$

basic	x_1	x_2	x_3	x_4	rhs
$-z$	0	$\frac{37}{3}$	0	$\frac{5}{3}$	-20
x_3	0	$-\frac{29}{3}$	1	$-\frac{4}{3}$	6
x_1	1	$-\frac{5}{3}$	0	$-\frac{1}{3}$	4

Thus the optimal solution is $(x_1, x_2, x_3, x_4)^T = (4, 0, 6, 0)^T$ and the objective is 20.

6.7 The feasible set of the original integer linear program is shown in Figure 1(a). From the figure, the optimal continuous solution is $x = (1, \frac{3}{2})^T$, with objective $-\frac{3}{2}$; and the optimal integer solution is $x = (1, 1)^T$, with objective -1.

Introduce x_3 and x_4 as slack variables. The relaxation LP problem is

$$\begin{array}{rll} \min & z = -x_2 & \\ \text{subject to} & 3x_1 + 2x_2 + x_3 = 6, & (10) \\ & -3x_1 + 2x_2 + x_4 = 0, & (11) \\ & x \geq 0. & \end{array}$$

Start from basis $(x_3, x_4)^T$, we get

⇓

basic	x_1	x_2	x_3	x_4	rhs
$-z$	0	-1	0	0	0
x_3	3	2	1	0	6
x_4	-3	$\boxed{2}$	0	1	0

$\rightarrow$

⇓

basic	x_1	x_2	x_3	x_4	rhs
$-z$	$-\frac{3}{2}$	0	0	$\frac{1}{2}$	0
x_3	$\boxed{6}$	0	1	-1	6
x_2	$-\frac{3}{2}$	1	0	$\frac{1}{2}$	0

$\downarrow$

basic	x_1	x_2	x_3	x_4	rhs
$-z$	0	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{2}$
x_1	1	0	$\frac{1}{6}$	$-\frac{1}{6}$	1
x_2	0	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{2}$

The continuous solution is $(x_1, x_2)^T = (1, \frac{3}{2})^T$, where x_2 is not an integer. Consider row 2 of $B^{-1}N$ (the last row of the above table), we get $d_{23} = d_{24} = \frac{1}{4}$ and $d_{20} = \frac{3}{2}$. From $f_{2j} \equiv d_{2j} - \lfloor d_{2j} \rfloor$, the Gomory cut is

$$\frac{1}{4}x_3 + \frac{1}{4}x_4 \geq \frac{1}{2}. \tag{12}$$

Since (10) plus (11) gives $4x_2 + x_3 + x_4 = 6$, (12) is equivalent to

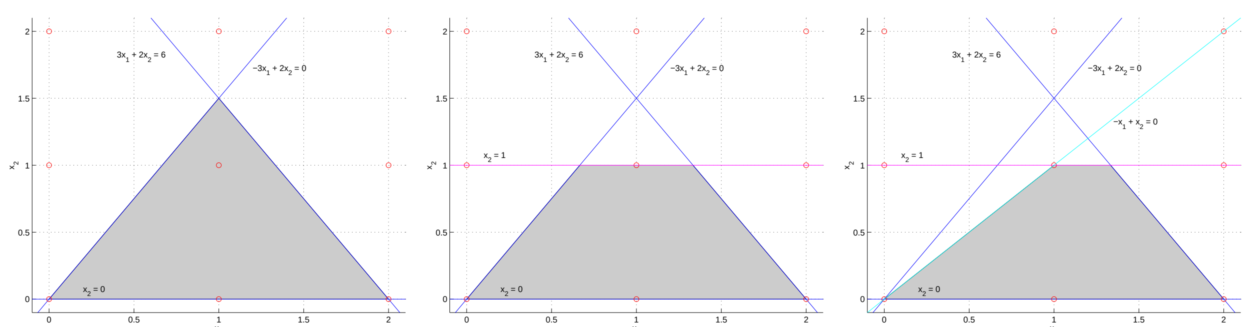
$$x_2 \leq 1. \tag{13}$$

The feasible set (on the x_1-x_2 plane) after this cut is in Figure 1(b).

Introduce slack variable x_5. Then (13) becomes another constraint:

$$x_2 + x_5 = 1. \tag{14}$$

Now the relaxation LP problem is solved as



(a) Original ILP (b) First cut (c) Second cut

Figure 1: Feasible sets (shadows) of the relaxation LP problems. Red circles are integer pointers.

		⇓				
basic	x_1	x_2	x_3	x_4	x_5	rhs
$-z$	0	−1	0	0	0	0
x_3	3	2	1	0	0	6
x_4	−3	$\boxed{2}$	0	1	0	0
x_5	0	1	0	0	1	1

→

	⇓					
basic	x_1	x_2	x_3	x_4	x_5	rhs
$-z$	$-\frac{3}{2}$	0	0	$\frac{1}{2}$	0	0
x_3	6	0	1	−1	0	6
x_2	$-\frac{3}{2}$	1	0	$\frac{1}{2}$	0	0
x_5	$\boxed{\frac{3}{2}}$	0	0	$-\frac{1}{2}$	1	1

↓

basic	x_1	x_2	x_3	x_4	x_5	rhs
$-z$	0	0	0	0	1	1
x_3	0	0	1	1	−4	2
x_2	0	1	0	0	1	1
x_1	1	0	0	$-\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$

$x_1 = \frac{2}{3}$ is not an integer. Consider row 1 of $B^{-1}N$ (the last row of the above table), we get $d_{14} = -\frac{1}{3}$, $d_{15} = \frac{2}{3}$, and $d_{10} = \frac{2}{3}$. Thus the Gomory cut is

$$\frac{2}{3}x_4 + \frac{2}{3}x_5 \geq \frac{2}{3}. \tag{15}$$

Since (11) plus (14) gives $-3x_1 + 3x_2 + x_4 + x_5 = 1$, (15) is equivalent to

$$-x_1 + x_2 \leq 0. \tag{16}$$

The feasible set (on the x_1-x_2 plane) after this second cut is in Figure 1(c).

Introduce slack variable x_6. Then (16) becomes constraint

$$-x_1 + x_2 + x_6 = 0. \tag{17}$$

With (17), the relaxation LP problem now is solved as

		⇓					
basic	x_1	x_2	x_3	x_4	x_5	x_6	rhs
$-z$	0	−1	0	0	0	0	0
x_3	3	2	1	0	0	0	6
x_4	−3	2	0	1	0	0	0
x_5	0	1	0	0	1	0	1
x_6	−1	$\boxed{1}$	0	0	0	1	0

→

	⇓						
basic	x_1	x_2	x_3	x_4	x_5	x_6	rhs
$-z$	−1	0	0	0	0	1	0
x_3	5	0	1	0	0	−2	6
x_4	−1	0	0	1	0	−2	0
x_5	$\boxed{1}$	0	0	0	1	−1	1
x_2	−1	1	0	0	0	1	0

↓

basic	x_1	x_2	x_3	x_4	x_5	x_6	rhs
$-z$	0	0	0	0	1	0	1
x_3	0	0	1	0	−5	3	1
x_4	0	0	0	1	1	−3	1
x_1	1	0	0	0	1	−1	1
x_2	0	1	0	0	1	0	1

Both x_1 and x_2 are integers. Thus we get the optimal integer solution $(x_1, x_2)^T = (1, 1)^T$.