First-Order Optimization Algorithms II

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/˜yyye

The SDM for Unconstrained Convex Lipschitz Optimization

Here we consider $f(\mathbf{x})$ being convex and differentiable everywhere and satisfying the (first-order) β-Lipschitz condition. Given the knowledge β, we again adopt the fixed step-size rule:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \frac{1}{\beta}\nabla f(\mathbf{x}^k). \tag{1}$$

Theorem 1 *For convex Lipschitz optimization the Steepest Descent Method generates a sequence of solutions such that*

$$f(\mathbf{x}^{k+1}) - f(\mathbf{x}^*) \le \frac{\beta}{k+2}\|\mathbf{x}^0 - \mathbf{x}^*\|^2 \text{ and } \min_{l=0,\ldots,k} \|\nabla f(\mathbf{x}^l)\|^2 = \frac{4\beta^2}{(k+1)(k+2)}\|\mathbf{x}^0 - \mathbf{x}^*\|^2,$$

where $\mathbf{x}^*$ is a minimizer of the problem.

Proof: For simplicity, we let $\delta^k = f(\mathbf{x}^k) - f(\mathbf{x}^*)(\ge 0)$, $\mathbf{g}^k = \nabla f(\mathbf{x}^k)$, and $\Delta^k = \mathbf{x}^k - \mathbf{x}^*$ in the rest of proof. As we have proved for general Lipschitz optimization

$$\delta^{k+1} - \delta^k \quad = f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \quad \le -\tfrac{1}{2\beta}\|\mathbf{g}^k\|^2, \quad \text{that is} \quad \delta^k - \delta^{k+1} \ge \tfrac{1}{2\beta}\|\mathbf{g}^k\|^2. \tag{2}$$

Furthermore, from the convexity,

$$-\delta^k = f(\mathbf{x}^*) - f(\mathbf{x}^k) \geq (\mathbf{g}^k)^T(\mathbf{x}^* - \mathbf{x}^k) = -(\mathbf{g}^k)^T\Delta^k, \quad \text{that is} \quad \delta^k \leq (\mathbf{g}^k)^T\Delta^k. \tag{3}$$

Thus, from (2) and (3)

$$\begin{aligned}
\delta^{k+1} &= \delta^{k+1} - \delta^k + \delta^k \\
&\leq -\tfrac{1}{2\beta}\|\mathbf{g}^k\|^2 + (\mathbf{g}^k)^T\Delta^k \\
&= -\tfrac{\beta}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - \beta(\mathbf{x}^{k+1} - \mathbf{x}^k)\Delta^k, \quad \text{(using (1))} \\
&= -\tfrac{\beta}{2}(\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + 2(\mathbf{x}^{k+1} - \mathbf{x}^k)^T\Delta^k) \\
&= -\tfrac{\beta}{2}(\|\Delta^{k+1} - \Delta^k\|^2 + 2(\Delta^{k+1} - \Delta^k)^T\Delta^k) \\
&= \tfrac{\beta}{2}(\|\Delta^k\|^2 - \|\Delta^{k+1}\|^2).
\end{aligned} \tag{4}$$

Sum up (4) from 1 to $k+1$, we have

$$\sum_{l=1}^{k+1} \delta^l \leq \frac{\beta}{2}(\|\Delta^0\|^2 - \|\Delta^{k+1}\|^2) \leq \frac{\beta}{2}\|\Delta^0\|^2.$$

From the proof of the Corollary 1 of last lecture, we have $\delta^0 \le \frac{\beta}{2}\|\Delta^0\|^2$. Thus,

$$\sum_{l=0}^{k+1} \delta^l \le \beta\|\Delta^0\|^2, \tag{5}$$

and

$$\begin{aligned}
\sum_{l=0}^{k+1} \delta^l &= \sum_{l=0}^{k+1} (l+1-l)\delta^l \\
&= \sum_{l=0}^{k+1} (l+1)\delta^l - \sum_{l=0}^{k+1} l\delta^l \\
&= \sum_{l=1}^{k+2} l\delta^{l-1} - \sum_{l=1}^{k+1} l\delta^l \\
&= (k+2)\delta^{k+1} + \sum_{l=1}^{k+1} l\delta^{l-1} - \sum_{l=1}^{k+1} l\delta^l \\
&= (k+2)\delta^{k+1} + \sum_{l=1}^{k+1} l(\delta^{l-1} - l\delta^l) \\
&\ge (k+2)\delta^{k+1} + \sum_{l=1}^{k+1} l\frac{1}{2\beta}\|\mathbf{g}^{l-1}\|^2,
\end{aligned}$$

where the first inequality comes from (2). Let $\|\mathbf{g}'\| = \min_{l=0,\ldots,k}\|\mathbf{g}^l\|$. Then we finally have

$$(k+2)\delta^{k+1} + \frac{(k+1)(k+2)/2}{2\beta}\|\mathbf{g}'\|^2 \le \beta\|\Delta^0\|^2, \tag{6}$$

which completes the proof.

The Accelerated Steepest Descent Method (ASDM)

There is an accelerated steepest descent method (Nesterov 83) that works as follows:

$$\lambda^0 = 0,\ \lambda^{k+1} = \frac{1+\sqrt{1+4(\lambda^k)^2}}{2},\ \alpha^k = \frac{1-\lambda^k}{\lambda^{k+1}}, \tag{7}$$

$$\tilde{\mathbf{x}}^{k+1} = \mathbf{x}^k - \frac{1}{\beta}\nabla f(\mathbf{x}^k),\ \mathbf{x}^{k+1} = (1-\alpha^k)\tilde{\mathbf{x}}^{k+1} + \alpha^k\tilde{\mathbf{x}}^k. \tag{8}$$

Note that $(\lambda^k)^2 = \lambda^{k+1}(\lambda^{k+1}-1)$, $\lambda^k > k/2$ and $\alpha^k \le 0$.

One can prove:

$$f(\tilde{\mathbf{x}}^{k+1}) - f(\mathbf{x}^*) \le \frac{2\beta}{k^2}\|\mathbf{x}^0 - \mathbf{x}^*\|^2,\ \forall k \ge 1.$$

Convergence Analysis of ASDM

Again for simplification, we let $\Delta^k = \lambda^k \mathbf{x}^k - (\lambda^k - 1)\tilde{\mathbf{x}}^k - \mathbf{x}^*$, $\mathbf{g}^k = \nabla f(\mathbf{x}^k)$ and $\delta^k = f(\tilde{\mathbf{x}}^k) - f(\mathbf{x}^*)(\geq 0)$ in the following.

Applying Lemma 1 for $\mathbf{x} = \tilde{\mathbf{x}}^{k+1}$ and $\mathbf{y} = \tilde{\mathbf{x}}^k$, convexity of f and (8) we have

$$\begin{aligned}
\delta^{k+1} - \delta^k &= f(\tilde{\mathbf{x}}^{k+1}) - f(\mathbf{x}^k) + f(\mathbf{x}^k) - f(\tilde{\mathbf{x}}^k) \\
&\leq -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + f(\mathbf{x}^k) - f(\tilde{\mathbf{x}}^k) \\
&\leq -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + (\mathbf{g}^k)^T(\mathbf{x}^k - \tilde{\mathbf{x}}^k) \\
&= -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 - \beta(\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k)^T(\mathbf{x}^k - \tilde{\mathbf{x}}^k).
\end{aligned} \tag{9}$$

Applying Lemma 1 for $\mathbf{x} = \tilde{\mathbf{x}}^{k+1}$ and $\mathbf{y} = \mathbf{x}^*$, convexity of f and (8) we have

$$\begin{aligned}
\delta^{k+1} &= f(\tilde{\mathbf{x}}^{k+1}) - f(\mathbf{x}^k) + f(\mathbf{x}^k) - f(\mathbf{x}^*) \\
&\leq -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + f(\mathbf{x}^k) - f(\mathbf{x}^*) \\
&\leq -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + (\mathbf{g}^k)^T(\mathbf{x}^k - \mathbf{x}^*) \\
&= -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 - \beta(\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k)^T(\mathbf{x}^k - \mathbf{x}^*).
\end{aligned} \tag{10}$$

Multiplying (9) by $\lambda^k(\lambda^k - 1)$ and (10) by λ^k respectively, and summing the two, we have

$$\begin{aligned}(\lambda^k)^2\delta^{k+1} - (\lambda^{k-1})^2\delta^k &\le -(\lambda^k)^2\tfrac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 - \lambda^k\beta(\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k)^T\Delta^k \\ &= -\tfrac{\beta}{2}((\lambda^k)^2\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + 2\lambda^k(\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k)^T\Delta^k) \\ &= -\tfrac{\beta}{2}(\|\lambda^k\tilde{\mathbf{x}}^{k+1} - (\lambda^k - 1)\tilde{\mathbf{x}}^k - \mathbf{x}^*\|^2 - \|\Delta^k\|^2) \\ &= \tfrac{\beta}{2}(\|\Delta^k\|^2 - \|\lambda^k\tilde{\mathbf{x}}^{k+1} - (\lambda^k - 1)\tilde{\mathbf{x}}^k - \mathbf{x}^*\|^2).\end{aligned}$$

Using (7) and (8) we can derive

$$\lambda^k\tilde{\mathbf{x}}^{k+1} - (\lambda^k - 1)\tilde{\mathbf{x}}^k = \lambda^{k+1}\mathbf{x}^{k+1} - (\lambda^{k+1} - 1)\tilde{\mathbf{x}}^{k+1}.$$

Thus,

$$(\lambda^k)^2\delta^{k+1} - (\lambda^{k-1})^2\delta^k \le \frac{\beta}{2}(\|\Delta^k\|^2 - \|\Delta^{k+1}\|^2.) \tag{11}$$

Sum up (11) from 1 to k we have

$$\delta^{k+1} \le \frac{\beta}{2(\lambda^k)^2}\|\Delta^1\|^2 \le \frac{2\beta}{k^2}\|\Delta^0\|^2$$

since $\lambda^k \ge k/2$ and $\|\Delta^1\| \le \|\Delta^0\|$.

First-Order Algorithms for Conic (Nonlinear) Optimization

$$\min\ f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in K.$$

- Nonnegative Linear Regression: given data $A \in R^{m\times n}$ and $\mathbf{b} \in R^m$

$$\min\ f(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2 \text{ s.t. } \mathbf{x} \geq \mathbf{0}; \quad \text{where } \nabla f(\mathbf{x}) = A^T(A\mathbf{x} - \mathbf{b}).$$

- Semidefinite Linear Regression: given data $A_i \in S^n$ for $i = 1, ..., m$ and $\mathbf{b} \in R^m$

$$\min\ f(X) = \frac{1}{2}\|\mathcal{A}X - \mathbf{b}\|^2 \text{ s.t. } X \succeq \mathbf{0}; \quad \text{where } \nabla f(X) = \mathcal{A}^T(\mathcal{A}X - \mathbf{b}).$$

$$\mathcal{A}X = \begin{pmatrix} A_1 \bullet X \\ \dots \\ A_m \bullet X \end{pmatrix} \quad \text{and} \quad \mathcal{A}^T\mathbf{y} = \sum_{i=1} y_i A_i.$$

SDM with Affine Scaling

Consider the nonnegative regression problem where the optimality conditions are

$$\mathbf{x} \geq \mathbf{0}, \quad \nabla f(\mathbf{x}) \geq \mathbf{0}, \quad x_j(\nabla f(\mathbf{x}))_j = 0 \; \forall j.$$

Let an iterate solution $\mathbf{x}^k > \mathbf{0}$ (for simplicity let each entry of $\mathbf{x}^k$ be bounded above by 1). Then, we can scale it to $\mathbf{e}$, the vector of all ones, by $\mathbf{x}' = (X^k)^{-1}\mathbf{x}$ where X^k is the diagonal matrix of vector $\mathbf{x}^k$.

$$f'(\mathbf{x}') = f(X^k\mathbf{x}') \quad \text{where} \quad \nabla f'(\mathbf{e}) = X^k \nabla f(\mathbf{x}^k),$$

the new SDM iterate would be

$$\mathbf{x}' = \mathbf{e} - \alpha_k X^k \nabla f(\mathbf{x}^k) \quad \text{and after scaling back} \quad \mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k (X^k)^2 \nabla f(\mathbf{x}^k).$$

If function f is β-Lipschitz, then so is f' with $\beta\|\mathbf{x}^k\|_\infty^2$, and we can fix $\alpha_k = \frac{1}{\beta\|\mathbf{x}^k\|_\infty^2}$ (or smaller to keep $\mathbf{x}^{k+1} > \mathbf{0}$) so that

$$f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \leq -\frac{1}{2\beta\|\mathbf{x}^k\|_\infty^2}\|X^k\nabla f(\mathbf{x}^k)\|^2),$$

so that the scaled gradient/complementarity vector $X^k\nabla f(\mathbf{x}^k)/\|\mathbf{x}^k\|_\infty$ converging to zero.

Affine Scaling for SDP Cone

Consider the semidefinite regression problem where the optimality conditions are

$$X \succeq \mathbf{0}, \quad \nabla f(X) \succeq \mathbf{0}, \quad X\nabla f(X) = \mathbf{0}, \text{ or } X^{1/2}\nabla f(X)X^{1/2} = \mathbf{0}.$$

Let an iterate solution $X^k \succ \mathbf{0}$. Then, one can scale it to I, the identity matrix, by $X' = (X^k)^{-1/2}X(X^k)^{-1/2}$. Now consider the objective function after scaling

$$f'(X') = f((X^k)^{1/2}X'(X^k)^{1/2}) \quad \text{where} \quad \nabla f'(I) = (X^k)^{1/2}\nabla f(X^k)(X^k)^{1/2},$$

the new SDM iterate would be

$$X' = I - \alpha_k (X^k)^{1/2}\nabla f(X^k)(X^k)^{1/2} \quad \text{and after scaling back} \quad X^{k+1} = X^k - \alpha_k X^k \nabla f(X^k) X^k.$$

If function f is β-Lipschitz, then so is f' with with $\beta\|X^k\|_\infty^2$, , and we can fix $\alpha_k = \frac{1}{\beta\|X^k\|_\infty^2}$ (or smaller to keep $X^{k+1} \succ \mathbf{0}$) so that

$$f(X^{k+1}) - f(X^k) \leq -\frac{1}{2\beta\|X^k\|_\infty^2}\|(X^k)^{1/2}\nabla f(X^k)(X^k)^{1/2}\|^2,$$

so that the scaled gradient/complementarity matrix converging to zero...

Conic Optimization with the Logarithmic Barrier

But one condition may be missing: $\nabla f(\mathbf{x}) \geq \mathbf{0}$ and $\nabla f(X) \succeq \mathbf{0}$, respectively...

One may add a barrier regularization:

$$\min \ \phi(\mathbf{x}) = f(\mathbf{x}) + \mu B(\mathbf{x}),$$

where $B(\mathbf{x}$ is a barrier function keeping $\mathbf{x}$ in the interior of cone K and μ is a positive barrier parameter.

- $K = R^n_+$:

$$B(\mathbf{x}) = -\sum_j \log(x_j), \quad \nabla B(\mathbf{x}) = \begin{pmatrix} \frac{-1}{x_1} \\ \dots \\ \frac{-1}{x_n} \end{pmatrix} \in R^n.$$

- $K = S^n_+$:

$$B(X) = -\log(\det(X)), \quad \nabla B(X) = -X^{-1} \in S^n.$$

Optimality Conditions for Optimization with Logarithmic Barrier

- Optimization in Nonnegative Cone ($\mathbf{x} > \mathbf{0}$):

$$\nabla\phi(\mathbf{x}) = \nabla f(\mathbf{x}) - \mu X^{-1}\mathbf{e} = \mathbf{0},\ X = \mathrm{diag}(\mathbf{x}),$$

  and the scaled gradient vector:

$$X\nabla\phi(\mathbf{x}) = X\nabla f(\mathbf{x}) - \mu\mathbf{e}$$

  and it converges to zero implies $\nabla f(\mathbf{x}) > \mathbf{0}$ for any positive μ. In practice, we set $\mu = \epsilon$ tolerance.

- Optimization in Semidefinite Cone ($X \succ \mathbf{0}$):

$$\nabla\phi(X) = \nabla f(X) - \mu X^{-1} = \mathbf{0},$$

  and the scaled gradient matrix:

$$X^{1/2}\nabla\phi(X)X^{1/2} = X^{1/2}\nabla f(X)X^{1/2} - \mu I$$

  and it converges to zero implies $\nabla f(X) \succ \mathbf{0}$ for any positive μ.

General Barrier and Penalty

We consider the general constrained optimization:

$$
\text{(GCO)} \qquad \begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & c_i(\mathbf{x}) & = 0, \; i \in \mathcal{E}, \\ & c_i(\mathbf{x}) & \geq 0, \; i \in \mathcal{I}. \end{array}
$$

We can convert it to an unconstrained problem:

$$
\min \quad f(\mathbf{x}) + \lambda \sum_{i \in \mathcal{E}} |c_i(\mathbf{x})| - \mu \sum_{i \in \mathcal{I}} \log(c_i(\mathbf{x}))
$$

where λ is sufficiently large and μ is sufficiently small.

Not robust if a high accuracy is desired...

A remedy strategy is to adjust λ is sufficiently large and μ dynamically, or use a projected gradient or reduced gradient first-order method, such as the Simplex Method of Dantzig...

Projected Gradient Method for Conic Optimization

Consider the nonnegative cone. At any iterate solution $\mathbf{x}^k \geq 0$, we project the gradient vector to the feasible direction space at $\mathbf{x}^k$:

$$g_j^k = \begin{cases} \nabla f(\mathbf{x}^k)_j & \text{if } x_j^k > 0 \text{ or } \nabla f(\mathbf{x}^k)_j < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then we apply the SDM with the projected gradient vector $\mathbf{g}^k$, that is, take a largest stepsize α such that

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha \mathbf{g}^k \geq \mathbf{0},\ 0 \leq \alpha \leq \frac{1}{\beta}.$$

Another Approach: we directly project the SDM iterate to the nonnegative cone in each step:

$$\mathbf{x}^{k+1} = \max\{0,\ \mathbf{x}^k - \frac{1}{\beta} \nabla f(\mathbf{x}^k)\}.$$

Does it converge? What's the difference of the two? What is the convergence speed? (Consider it a bonus question to Problem 6 of HW3.)

The Simplex Algorithm for LP: Reduced Gradient Method

$$\text{LP:} \quad \min\ \mathbf{c}^T\mathbf{x} \text{ s.t. } A\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0},$$

where $A \in R^{m \times n}$ has a full row rank m.

Theorem 2 *(The Fundamental Theorem of LP in Algebraic form) Given (LP) and (LD) where A has full row rank m,*

i) *if there is a feasible solution, there is a basic feasible solution (Carathéodory's theorem);*

ii) *if there is an optimal solution, there is an optimal basic solution.*

High-Level Idea:

1. Initialization Start at a BSF or corner point of the feasible polyhedron.
2. Test for Optimality. Compute the reduced gradient vector at the corner. If no descent and feasible direction can be found, stop and claim optimality at the current corner point; otherwise, select a new corner point and go to Step 2.

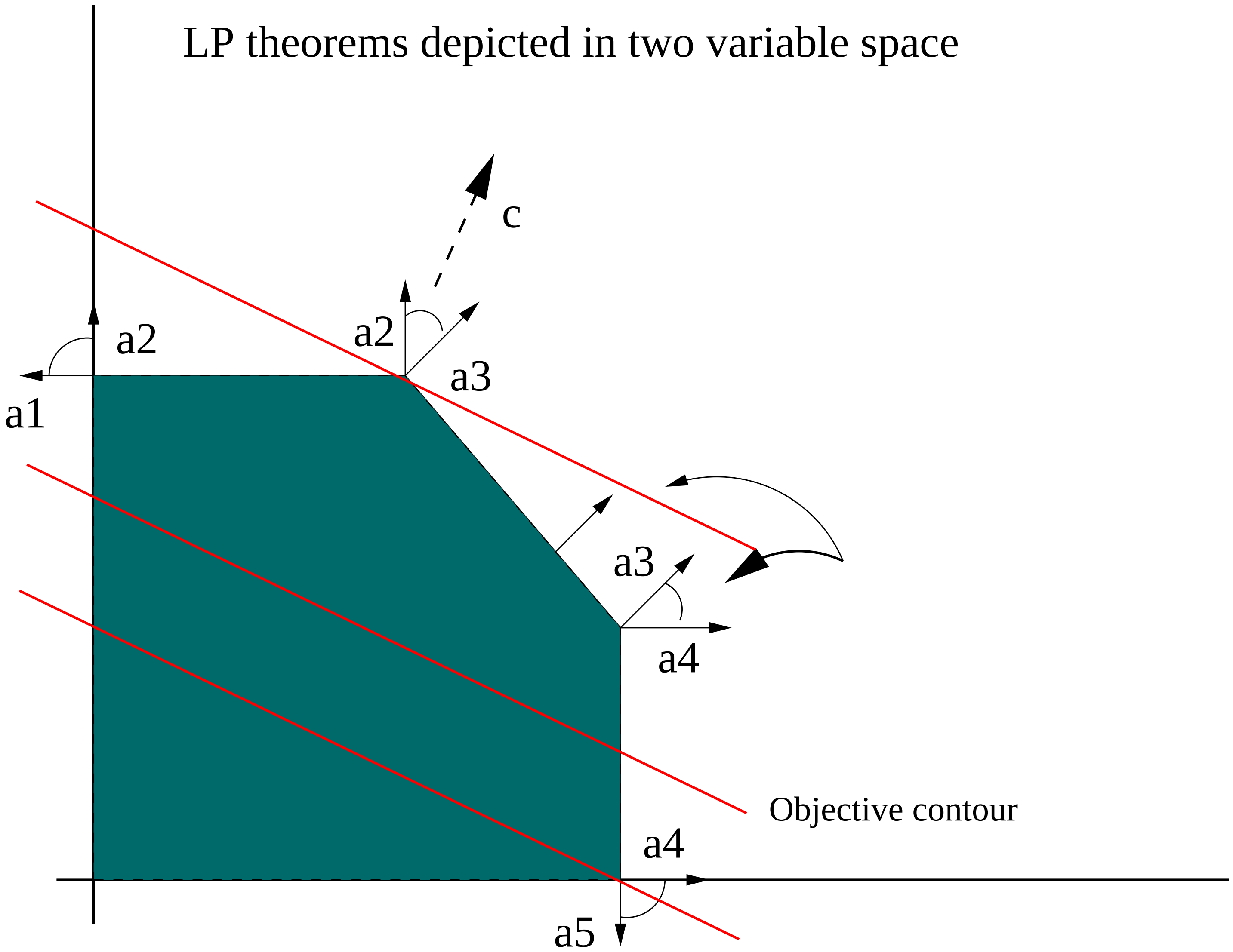

Figure 1: The LP Simplex Method

When a Basic Feasible Solution is Optimal

Suppose the basis of a basic feasible solution is A_B and the rest is A_N. One can transform the equality constraint to

$$A_B^{-1}A\mathbf{x} = A_B^{-1}\mathbf{b}, \quad \text{so that } \mathbf{x}_B = A_B^{-1}\mathbf{b} - A_B^{-1}A_N\mathbf{x}_N.$$

That is, we express $\mathbf{x}_B$ in terms of $\mathbf{x}_N$, the non-basic variables are are active for constraints $\mathbf{x} \geq \mathbf{0}$. Then the objective function equivalently becomes

$$\begin{aligned} \mathbf{c}^T\mathbf{x} = \mathbf{c}_B^T\mathbf{x}_B + \mathbf{c}_N^T\mathbf{x}_N &= \mathbf{c}_B^T A_B^{-1}\mathbf{b} - \mathbf{c}_B^T A_B^{-1}A_N\mathbf{x}_N + \mathbf{c}_N^T\mathbf{x}_N \\ &= \mathbf{c}_B^T A_B^{-1}\mathbf{b} + (\mathbf{c}_N^T - \mathbf{c}_B^T A_B^{-1}A_N)\mathbf{x}_N. \end{aligned}$$

Vector $\mathbf{r}^T = \mathbf{c}^T - \mathbf{c}_B^T A_B^{-1}A$ is called the Reduced Gradient/Cost Vector where $\mathbf{r}_B = \mathbf{0}$ always.

Theorem 3 *If Reduced Gradient Vector $\mathbf{r}^T = \mathbf{c}^T - \mathbf{c}_B^T A_B^{-1}A \geq \mathbf{0}$, then the BFS is optimal.*

Proof: Let $\mathbf{y}^T = \mathbf{c}_B^T A_B^{-1}$ (called Shadow Price Vector), then $\mathbf{y}$ is a dual feasible solution ($\mathbf{r} = \mathbf{c} - A^T\mathbf{y} \geq \mathbf{0}$) and $\mathbf{c}^T\mathbf{x} = \mathbf{c}_B^T\mathbf{x}_B = \mathbf{c}_B^T A_B^{-1}\mathbf{b} = \mathbf{y}^T\mathbf{b}$, that is, the duality gap is zero.

The Simplex Algorithm Procedures

0. Initialize Start a BFS with basic index set B and let N denote the complementary index set.

1. Test for Optimality: Compute the Reduced Gradient Vector $\mathbf{r}$ at the current BFS and let

$$r_e = \min_{j \in N}\{r_j\}.$$

   If $r_e \geq 0$, stop – the current BFS is optimal.

2. Determine the Replacement: Increase x_e while keep all other non-basic variables at the zero value (inactive) and maintain the equality constraints:

$$\mathbf{x}_B = A_B^{-1}\mathbf{b} - A_B^{-1}A_{.e}x_e \ (\geq \mathbf{0}).$$

   If x_e can be increased to ∞, stop – the problem is unbounded below. Otherwise, let the basic variable x_o be the one first becoming 0.

3. Update basis: update B with x_o being replaced by x_e, and return to Step 1.

A Toy Example

$$\begin{array}{llllllll} \text{minimize} & -x_1 & -2x_2 & & & & \\ \text{subject to} & x_1 & & +x_3 & & & = 1 \\ & & x_2 & & +x_4 & & = 1 \\ & x_1 & +x_2 & & & +x_5 & = 1.5. \end{array}$$

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1.5 \end{pmatrix}, \ \mathbf{c}^T = (-1 \ -2 \ 0 \ 0 \ 0).$$

Consider initial BFS with basic variables $B = \{3, 4, 5\}$ and $N = \{1, 2\}$.

Iteration 1:

1. $A_B = I$, $A_B^{-1} = I$, $\mathbf{y}^T = (0\,0\,0)$ and $\mathbf{r}_N = (-1 \ -2)$ – it's NOT optimal. Let $e = 2$.

2. Increase x_2 while

$$\mathbf{x}_B = A_B^{-1}\mathbf{b} - A_B^{-1}A_{.2}x_2 = \begin{pmatrix} 1 \\ 1 \\ 1.5 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} x_2.$$

   We see x_4 becomes 0 first.

3. The new basic variables are $B = \{3, 2, 5\}$ and $N = \{1, 4\}$.

Iteration 2:

1. $$A_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad A_B^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix},$$

   $\mathbf{y}^T = (0 \ \ -2 \ \ 0)$ and $\mathbf{r}_N = (-1 \ \ 2)$ – it's NOT optimal. Let $e = 1$.

2. Increase x_1 while

$$\mathbf{x}_B = A_B^{-1}\mathbf{b} - A_B^{-1}A_{.1}x_1 = \begin{pmatrix} 1 \\ 1 \\ 0.5 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} x_1 .$$

   We see x_5 becomes 0 first.

3. The new basic variables are $B = \{3, 2, 1\}$ and $N = \{4, 5\}$.

Iteration 3:

1.

$$A_B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad A_B^{-1} = \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix},$$

   $\mathbf{y}^T = (0 \ \ -1 \ \ -1)$ and $\mathbf{r}_N = (1 \ \ 1)$ – it's Optimal.

Is the Simplex Method always convergent to a minimizer? Which condition of the Global Convergence Theorem failed?