First-Order Methods for Unconstrained Optimization

Yinyu Ye

Department of Management Science and Engineering

Stanford University

Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

(LY: Chapters 8 and 14)

The Steepest Descent Method (SDM) with a Fixed Step Size

Here we consider the unconstrained convex optimization problem

$$\min \quad f(\mathbf{x})$$

where $f(\mathbf{x})$ is convex and differentiable every where, admits a minimizer $\mathbf{x}^*$, and satisfies the (first-order) β-Lipschitz condition, that is, for any two points $\mathbf{x}$ and $\mathbf{y}$

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq \beta\|\mathbf{x} - \mathbf{y}\|$$

for a positive real number β.

Starting from any point $\mathbf{x}^0$, the SDM is an iteration rule:

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

Does the sequence converge? How fast if it converges?

Convergence Analysis of the Method

Theorem 1 *The SDM generates a sequence of points $\mathbf{x}^k$, from any given initial point $\mathbf{x}^0$, such that*

$$\|\nabla f(\mathbf{x}^k)\|^2 \leq \frac{\beta^2\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{k+1}, \ \forall k \geq 1.$$

Proof: First, for any differentiable f, convex or nonconvex, we should have

$$f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})^T(\mathbf{x} - \mathbf{y}) \leq \frac{\beta}{2}\|\mathbf{x} - \mathbf{y}\|^2. \quad (2)$$

Now consider function $g_x(\mathbf{y}) = f(\mathbf{y}) - \nabla f(\mathbf{x})^T\mathbf{y}$ for any given $\mathbf{x}$. Note that g_x is also convex and satisfies the β-Lipschitz condition. Moreover, $\mathbf{x}$ is the minimizer of $g_x(\mathbf{y})$ and $\nabla g_x(\mathbf{y}) = \nabla f(\mathbf{y}) - \nabla f(\mathbf{x})$.

Applying (2) to g_x and noting the relations of g_x and $f(\mathbf{x})$, we have

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{x})^T(\mathbf{x}-\mathbf{y}) &= g_x(\mathbf{x}) - g_x(\mathbf{y}) \\ &\le g_x(\mathbf{y} - \tfrac{1}{\beta}\nabla g_x(\mathbf{y})) - g_x(\mathbf{y}) \\ &\le \nabla g_x(\mathbf{y})^T(-\tfrac{1}{\beta}\nabla g_x(\mathbf{y})) + \tfrac{\beta}{2}\tfrac{1}{\beta^2}\|\nabla g_x(\mathbf{y})\|^2 \\ &= -\tfrac{1}{2\beta}\|\nabla g_x(\mathbf{y})\|^2 \\ &= -\tfrac{1}{2\beta}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2. \end{aligned} \tag{3}$$

Similarly, we have

$$f(\mathbf{y}) - f(\mathbf{x}) - \nabla f(\mathbf{y})^T(\mathbf{y}-\mathbf{x}) \le -\frac{1}{2\beta}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2.$$

Adding the above two derived inequalities, we have another key inequality for any $\mathbf{x}$ and $\mathbf{y}$:

$$(\mathbf{x}-\mathbf{y})^T(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})) \ge \frac{1}{\beta}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|^2. \tag{4}$$

For simplification, in the following we let $\mathbf{d}^k = \mathbf{x}^k - \mathbf{x}^*$ and $\mathbf{g}^k = \nabla f(\mathbf{x}^k)$.

Let $\mathbf{x} = \mathbf{x}^k$ and $\mathbf{y} = \mathbf{x}^*$ in (4). Then, since $\nabla f(\mathbf{x}^*) = \mathbf{0}$,

$$(\mathbf{d}^k)^T \mathbf{g}^k \geq \frac{1}{\beta} \|\mathbf{g}^k\|^2;$$

so that

$$\begin{aligned} \|\mathbf{d}^{k+1}\|^2 &= \|\mathbf{x}^k - \tfrac{1}{\beta} \nabla f(\mathbf{x}^k) - \mathbf{x}^*\|^2 \\ &= \tfrac{1}{\beta^2} \|\mathbf{g}^k\|^2 - \tfrac{2}{\beta} (\mathbf{d}^k)^T \mathbf{g}^k + \|\mathbf{d}^k\|^2 \\ &\leq \tfrac{1}{\beta^2} \|\mathbf{g}^k\|^2 - \tfrac{2}{\beta^2} \|\mathbf{g}^k\|^2 + \|\mathbf{d}^k\|^2 \\ &= -\tfrac{1}{\beta^2} \|\mathbf{g}^k\|^2 + \|\mathbf{d}^k\|^2, \end{aligned}$$

that is,

$$\|\mathbf{d}^{k+1}\|^2 + \frac{1}{\beta^2} \|\mathbf{g}^k\|^2 \leq \|\mathbf{d}^k\|^2. \tag{5}$$

Inequality (5) implies that $\|\mathbf{d}^k\| = \|\mathbf{x}^k - \mathbf{x}^*\|$ is monotonically decreasing.

Now let $\mathbf{x} = \mathbf{x}^{k+1}$ and $\mathbf{y} = \mathbf{x}^k$ in (4). Then

$$\begin{aligned} -\tfrac{1}{\beta}(\mathbf{g}^k)^T(\mathbf{g}^{k+1} - \mathbf{g}^k) &= (\mathbf{x}^{k+1} - \mathbf{x}^k)^T(\mathbf{g}^{k+1} - \mathbf{g}^k) \\ &\geq \tfrac{1}{\beta}\|\mathbf{g}^{k+1} - \mathbf{g}^k\|^2, \end{aligned}$$

which leads to

$$\begin{array}{c} \|\mathbf{g}^{k+1}\|^2 \leq (\mathbf{g}^{k+1})^T\mathbf{g}^k \leq \|\mathbf{g}^{k+1}\|\|\mathbf{g}^k\|, \text{ or} \\ \|\mathbf{g}^{k+1}\| \leq \|\mathbf{g}^k\|. \end{array} \tag{6}$$

Inequality (6) implies that $\|\mathbf{g}^k\| = \|\nabla f(\mathbf{x}^k)\|$ is also monotonically decreasing.

Sum up (5) from 0 to k, we have

$$\|\mathbf{d}^{k+1}\|^2 + \frac{1}{\beta^2}\sum_{l=0}^{k}\|\mathbf{g}^l\|^2 \leq \|\mathbf{d}^0\|^2.$$

Then use (6), we have

$$\|\mathbf{d}^{k+1}\|^2 + \frac{k+1}{\beta^2}\|\mathbf{g}^k\|^2 \leq \|\mathbf{d}^0\|^2,$$

that is,

$$\|\nabla f(\mathbf{x}^k)\|^2 = \|\mathbf{g}^k\|^2 \leq \frac{\beta^2}{k+1}\|\mathbf{d}^0\|^2 = \frac{\beta^2}{k+1}\|\mathbf{x}^0 - \mathbf{x}^*\|^2,$$

which completes the proof.

Improved Convergence Analysis of the Method

We now improve the bound and prove:

Theorem 2 *The Steepest Descent Method of (1) generate a sequence of solutions such that*

$$\|\nabla f(\mathbf{x}^k)\|^2 = \frac{2\beta^2}{(k+1)(k+2)}\|\mathbf{x}^0 - \mathbf{x}^*\|^2.$$

Further for simplification, we let $\delta^k = f(\mathbf{x}^k) - f(\mathbf{x}^*)(\geq 0)$ in the rest of analyses.

Applying inequality (2) for $\mathbf{x} = \mathbf{x}^{k+1}$ and $\mathbf{y} = \mathbf{x}^k$ and noting (1) we have

$$\begin{aligned} \delta^{k+1} - \delta^k &= f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \\ &\leq (\mathbf{g}^k)^T(-\tfrac{1}{\beta}\mathbf{g}^k) + \tfrac{\beta}{2}\tfrac{1}{\beta^2}\|\mathbf{g}^k\|^2 \qquad (7) \\ &= -\tfrac{1}{2\beta}\|\mathbf{g}^k\|^2. \end{aligned}$$

This inequality implies that δ^k is monotonically decreasing.

Applying inequality (3) for $\mathbf{x} = \mathbf{x}^k$ and $\mathbf{y} = \mathbf{x}^*$ and noting $\mathbf{g}^* = \mathbf{0}$ we have

$$\begin{aligned}\delta^k &\le (\mathbf{g}^k)^T\mathbf{d}^k - \tfrac{1}{2\beta}\|\mathbf{g}^k\|^2 \\ &= -\beta(\mathbf{x}^{k+1} - \mathbf{x}^k)\mathbf{d}^k - \tfrac{\beta}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\ &= -\tfrac{\beta}{2}(\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + 2(\mathbf{x}^{k+1} - \mathbf{x}^k)^T\mathbf{d}^k) \\ &= -\tfrac{\beta}{2}(\|\mathbf{d}^{k+1} - \mathbf{d}^k\|^2 + 2(\mathbf{d}^{k+1} - \mathbf{d}^k)^T\mathbf{d}^k) \\ &= \tfrac{\beta}{2}(\|\mathbf{d}^k\|^2 - \|\mathbf{d}^{k+1}\|^2).\end{aligned} \qquad (8)$$

Sum up (8) from 0 to k, we have

$$\sum_{l=0}^{k} \delta^l \le \frac{\beta}{2}(\|\mathbf{d}^0\|^2 - \|\mathbf{d}^{k+1}\|^2) \le \frac{\beta}{2}\|\mathbf{d}^0\|^2. \qquad (9)$$

Repeatedly applying inequality (7), we have

$$
\begin{aligned}
\sum_{l=0}^{k} \delta^l \quad & \geq \delta^1 + \tfrac{1}{2\beta}\|\mathbf{g}^0\|^2 + \sum_{l=1}^{k} \delta^l \\
& = 2\delta^1 + \tfrac{1}{2\beta}\|\mathbf{g}^0\|^2 + \sum_{l=2}^{k} \delta^l \\
& \geq 2\delta^2 + \tfrac{2}{2\beta}\|\mathbf{g}^1\|^2 + \tfrac{1}{2\beta}\|\mathbf{g}^0\|^2 + \sum_{l=2}^{k} \delta^l \\
& = 3\delta^2 + \tfrac{2}{2\beta}\|\mathbf{g}^1\|^2 + \tfrac{1}{2\beta}\|\mathbf{g}^0\|^2 + \sum_{l=3}^{k} \delta^l \\
& \ldots \\
& \geq k\delta^k + \tfrac{k}{2\beta}\|\mathbf{g}^{k-1}\|^2 + \ldots + \tfrac{2}{2\beta}\|\mathbf{g}^1\|^2 + \tfrac{1}{2\beta}\|\mathbf{g}^0\|^2 + \sum_{l=k}^{k} \delta^l \\
& = (k+1)\delta^k + \tfrac{k}{2\beta}\|\mathbf{g}^{k-1}\|^2 + \ldots + \tfrac{2}{2\beta}\|\mathbf{g}^1\|^2 + \tfrac{1}{2\beta}\|\mathbf{g}^0\|^2 \\
& \geq (k+1)\delta^k + (\tfrac{k}{2\beta} + \ldots + \tfrac{2}{2\beta} + \tfrac{1}{2\beta})\|\mathbf{g}^{k-1}\|^2 \\
& = (k+1)\delta^k + \tfrac{k(k+1)/2}{2\beta}\|\mathbf{g}^{k-1}\|^2,
\end{aligned}
$$

where the last inequality comes from (6), that is, $\|\mathbf{g}^k\| = \|\nabla f(\mathbf{x}^k)\|$ is monotonically decreasing.

Using (9) we finally have

$$(k+1)\delta^k + \frac{k(k+1)/2}{2\beta}\|\mathbf{g}^{k-1}\|^2 \le \frac{\beta}{2}\|\mathbf{d}^0\|^2. \tag{10}$$

Inequality (10), since $\delta^k \ge 0$, $\mathbf{g}^k = \nabla f(\mathbf{x}^k)$ and $\mathbf{d}^0 = \mathbf{x}^0 - \mathbf{x}^*$, proves the desired bound:

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

which improves the early bound. It also implies that

$$\delta^k \le \frac{\beta}{2(k+1)}\|\mathbf{x}^0 - \mathbf{x}^*\|^2,$$

the standard convergence result of the SDM.

The Accelerated Steepest Descent Method (ASDM)

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

$$\lambda^0 = 0,\ \lambda^{k+1} = \frac{1+\sqrt{1+4(\lambda^k)^2}}{2},\ \alpha^k = \frac{1-\lambda^k}{\lambda^{k+1}}, \tag{11}$$

$$\tilde{\mathbf{x}}^{k+1} = \mathbf{x}^k - \frac{1}{\beta}\nabla f(\mathbf{x}^k),\ \mathbf{x}^{k+1} = (1-\alpha^k)\tilde{\mathbf{x}}^{k+1} + \alpha^k\tilde{\mathbf{x}}^k. \tag{12}$$

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

Applying inequality (2) for $\mathbf{x} = \tilde{\mathbf{x}}^{k+1}$ and $\mathbf{y} = \tilde{\mathbf{x}}^k$, convexity of f and (12) we have

$$\begin{aligned} \delta^{k+1} - \delta^k &= f(\tilde{\mathbf{x}}^{k+1}) - f(\mathbf{x}^k) + f(\mathbf{x}^k) - f(\tilde{\mathbf{x}}^k) \\ &\leq -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + f(\mathbf{x}^k) - f(\tilde{\mathbf{x}}^k) \\ &\leq -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + (\mathbf{g}^k)^T(\mathbf{x}^k - \tilde{\mathbf{x}}^k) \\ &= -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 - \beta(\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k)^T(\mathbf{x}^k - \tilde{\mathbf{x}}^k). \end{aligned} \tag{13}$$

Applying inequality (2) for $\mathbf{x} = \tilde{\mathbf{x}}^{k+1}$ and $\mathbf{y} = \mathbf{x}^*$, convexity of f and (12) we have

$$\begin{aligned}\delta^{k+1} &= f(\tilde{\mathbf{x}}^{k+1}) - f(\mathbf{x}^k) + f(\mathbf{x}^k) - f(\mathbf{x}^*) \\ &\leq -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + f(\mathbf{x}^k) - f(\mathbf{x}^*) \\ &\leq -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 + (\mathbf{g}^k)^T(\mathbf{x}^k - \mathbf{x}^*) \\ &= -\frac{\beta}{2}\|\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k\|^2 - \beta(\tilde{\mathbf{x}}^{k+1} - \mathbf{x}^k)^T(\mathbf{x}^k - \mathbf{x}^*).\end{aligned} \tag{14}$$

Multiplying (13) by $\lambda^k(\lambda^k - 1)$ and (14) by λ^k respectively, and summing the two, we have

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

Using (11) and (12) 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,

$$(\lambda^k)^2\delta^{k+1} - (\lambda^{k-1})^2\delta^k \le \frac{\beta}{2}(\|\mathbf{d}^k\|^2 - \|\mathbf{d}^{k+1}\|^2.) \qquad (15)$$

Sum up (15) from 1 to k we have

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

since $\lambda^k \ge k/2$ and $\|\mathbf{d}^1\| \le \|\mathbf{d}^0\|$.

The Barzilai and Borwein Method

Yet there is another two-point steepest descent method (Barzilai and Borwein 88) that works as follows:

$$\Delta_x^k = \mathbf{x}^k - \mathbf{x}^{k-1} \quad \text{and} \quad \Delta_g^k = \nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k-1}), \tag{16}$$

$$\alpha^k = \frac{(\Delta_x^k)^T \Delta_g^k}{(\Delta_g^k)^T \Delta_g^k} \quad \text{or} \quad \alpha^k = \frac{(\Delta_x^k)^T \Delta_x^k}{(\Delta_x^k)^T \Delta_g^k},$$

Then

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha^k \nabla f(\mathbf{x}^k). \tag{17}$$

An explanation why the BB method works

For convex quadratic minimization, let the distinct nonzero eigenvalues of the Hessian Q be λ_1, λ_2, ..., λ_K; and let the step size in the SDM be $\alpha^k = \frac{1}{\lambda_k}$, $k = 1, ..., K$. Then, the SDM terminates in K iterations.

In the BB method, α^k minimizes

$$\|\Delta_x^k - \alpha \Delta_g^k\| = \|\Delta_x^k - \alpha Q \Delta_x^k\|.$$

If the error becomes 0 plus $\|\Delta_x^k\| \neq 0$, $\frac{1}{\alpha^k}$ will be a nonzero eigenvalue of Q.

The Augmented Lagrangian Method

We consider

$$\min \quad f(\mathbf{x}) \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}.$$

The KKT conditions would be

$$\nabla f(\mathbf{x}) - A^T\mathbf{y} = \mathbf{0},$$

$$A\mathbf{x} - \mathbf{b} = \mathbf{0}.$$

Denote by $(\mathbf{x}^*, \mathbf{y}^*)$ an KKT pair, and consider the augmented Lagrangian function

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) - \mathbf{y}^T(A\mathbf{x} - \mathbf{b}) + \frac{\beta}{2}\|A\mathbf{x} - \mathbf{b}\|^2.$$

Then, for any given $(\mathbf{x}^k, \mathbf{y}^k)$, we compute a new iterate pair

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} L(\mathbf{x}, \mathbf{y}^k), \text{ and } \mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A\mathbf{x}^{k+1} - \mathbf{b}).$$

Analysis of the Augmented Lagrangian Method

Since $\mathbf{x}^{k+1}$ makes KKT condition:

$$\begin{aligned} \mathbf{0} &= \nabla f(\mathbf{x}^{k+1}) - A^T\mathbf{y}^k + \beta A^T(A\mathbf{x}^{k+1} - \mathbf{b}) \\ &= \nabla f(\mathbf{x}^{k+1}) - A^T(\mathbf{y}^k - \beta(A\mathbf{x}^{k+1} - \mathbf{b})) \\ &= \nabla f(\mathbf{x}^{k+1}) - A^T\mathbf{y}^{k+1}, \end{aligned}$$

we only need to consider whether or not $\|A\mathbf{x}^k - \mathbf{b}\|$ converges to zero and how fast it converges. From the convexity of $f(\mathbf{x})$, we have

$$\begin{aligned} \mathbf{0} &\le (\nabla f(\mathbf{x}^{k+1}) - \nabla f(\mathbf{x}^*))^T(\mathbf{x}^{k+1} - \mathbf{x}^*) \\ &= (A^T\mathbf{y}^{k+1} - A^T\mathbf{y}^*)^T(\mathbf{x}^{k+1} - \mathbf{x}^*) \\ &= (\mathbf{y}^{k+1} - \mathbf{y}^*)^T(A\mathbf{x}^{k+1} - A\mathbf{x}^*) = (\mathbf{y}^{k+1} - \mathbf{y}^*)^T(A\mathbf{x}^{k+1} - \mathbf{b}) \\ &= \tfrac{1}{\beta}(\mathbf{y}^{k+1} - \mathbf{y}^*)^T(\mathbf{y}^k - \mathbf{y}^{k+1}). \end{aligned}$$

Thus, we have

$$\begin{aligned}\|\mathbf{y}^k-\mathbf{y}^*\|^2 &= \|\mathbf{y}^k-\mathbf{y}^{k+1}+\mathbf{y}^{k+1}-\mathbf{y}^*\|^2\\ &\geq \|\mathbf{y}^k-\mathbf{y}^{k+1}\|^2+\|\mathbf{y}^{k+1}-\mathbf{y}^*\|^2\\ &= \beta\|A\mathbf{x}^{k+1}-\mathbf{b}\|^2+\|\mathbf{y}^{k+1}-\mathbf{y}^*\|^2.\end{aligned}$$

Again from the convexity,

$$\begin{aligned}\mathbf{0} &\leq (\nabla f(\mathbf{x}^{k+1})-\nabla f(\mathbf{x}^k))^T(\mathbf{x}^{k+1}-\mathbf{x}^k)\\ &= (A^T\mathbf{y}^{k+1}-A^T\mathbf{y}^k)^T(\mathbf{x}^{k+1}-\mathbf{x}^k)\\ &= (\mathbf{y}^{k+1}-\mathbf{y}^k)^T(A\mathbf{x}^{k+1}-A\mathbf{x}^k)\\ &= -\beta(A\mathbf{x}^{k+1}-\mathbf{b})((A\mathbf{x}^{k+1}-\mathbf{b}+\mathbf{b}-A\mathbf{x}^k),\end{aligned}$$

which implies that

$$\|A\mathbf{x}^{k+1}-\mathbf{b}\|^2 \leq \|A\mathbf{x}^k-\mathbf{b}\|^2.$$

Sum up from 0 to k of

$$\|\mathbf{y}^k-\mathbf{y}^*\|^2 \geq \beta\|A\mathbf{x}^{k+1}-\mathbf{b}\|^2+\|\mathbf{y}^{k+1}-\mathbf{y}^*\|^2,$$

we have

$$\begin{aligned}
\|\mathbf{y}^0-\mathbf{y}^*\|^2 &\geq \|\mathbf{y}^{k+1}-\mathbf{y}^*\|^2+\beta\sum_{l=0}^{k}\|A\mathbf{x}^{l+1}-\mathbf{b}\|^2 \\
&\geq \beta\sum_{l=0}^{k}\|A\mathbf{x}^{l+1}-\mathbf{b}\|^2 \\
&\geq (k+1)\beta\|A\mathbf{x}^{k+1}-\mathbf{b}\|^2,
\end{aligned}$$

which gives the complexity bound:

$$\|A\mathbf{x}^{k+1}-\mathbf{b}\|^2 \leq \frac{1}{(k+1)\beta}\|\mathbf{y}^0-\mathbf{y}^*\|^2.$$

The Alternating Direction Method with Multipliers

We consider

$$\min \quad f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \quad \text{s.t.} \quad A_1\mathbf{x}_1 + A_2\mathbf{x}_2 = \mathbf{b},\ \mathbf{x}_1 \in X_1,\ \mathbf{x}_2 \in X_2;$$

where X_1 and X_2 are (simple) convex sets.

Define its Augmented Lagrangian

$$L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) - \mathbf{y}^T(A_1\mathbf{x}_1 + A_2\mathbf{x}_2 - \mathbf{b}) + \frac{\beta}{2}\|A_1\mathbf{x}_1 + A_2\mathbf{x}_2 - \mathbf{b}\|^2.$$

Then, for any given $(\mathbf{x}_1^k, \mathbf{x}_2^k, \mathbf{y}^k)$, we compute a new iterate pair

$$\mathbf{x}_1^{k+1} = \arg\min_{\mathbf{x}_1 \in X_1} L(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{y}^k),\ \mathbf{x}_2^{k+1} = \arg\min_{\mathbf{x}_2 \in X_2} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{y}^k)$$

and

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A_1\mathbf{x}_1^{k+1} + A_2\mathbf{x}_2^{k+1} - \mathbf{b}).$$

Again, the iterates converge for any $\beta > 0$ with the same speed as the SDM.

The Splitting to Handle Inequalities

We consider

$$\min \quad f(\mathbf{x}_1) \quad \text{s.t.} \quad A\mathbf{x}_1 + \mathbf{x}_2 = \mathbf{b},\ \mathbf{x}_2 \geq \mathbf{0}.$$

$$L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = f(\mathbf{x}_1) - \mathbf{y}^T(A\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{b}) + \frac{\beta}{2}\|A\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{b}\|^2.$$

Then, for any given $(\mathbf{x}_1^k, \mathbf{x}_2^k, \mathbf{y}^k)$, we compute a new iterate pair

$$\mathbf{x}_1^{k+1} = \arg\min_{\mathbf{x}_1} L(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{y}^k)$$

$$\mathbf{x}_2^{k+1} = \arg\min_{\mathbf{x}_2 \geq \mathbf{0}} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{y}^k)$$

and

$$\mathbf{y}^{k+1} = \mathbf{y}^k - \beta(A\mathbf{x}_1^{k+1} + \mathbf{x}_2^{k+1} - \mathbf{b}).$$

Note that the solution of $\mathbf{x}_2$ can be computed in a close form!