Penalty and Barrier Method

LI-PING ZHANG

Department of Mathematical Sciences

Tsinghua University, Beijing 100084

Office: New Science Building #A302, Tel: 62798531

E-mail: lipingzhang@tsinghua.edu.cn

Outline

- Penalty Method   罚函数
- Barrier Method   障碍函数
- Augment Lagrangian Method   增广Lagrange算法
- The Alternating Direction Method with Multipliers

Main idea

Penalty and barrier methods are procedures for approximating constrained optimization problems by unconstrained problems. The approximation is accomplished in the case of penalty methods by adding to the objective function a term that prescribes a high cost for variation of the constraints and in the case of barrier methods by adding a term that favors points interior to the feasible region over these near the boundary. Associated with these methods is a parameter c that determines the severity of the penalty or barrier and consequently the degree to which the unconstrained problem approximates the original constrained problem.

把约束优化换成
无约束优化

Fundamental issues

The first has to do with how well the unconstrained problem approximates the constrained one. This is essential in examining whether, as the parameter c is increased toward infinity, the solution of the unconstrained problem converges to a solution of the constrained problem.

The other issue, most important from a practical viewpoint, is the question of how to solve a given unconstrained problem when its objective function contains a penalty or barrier term.

Penalty methods

Consider the problem

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ s.t. \quad & \mathbf{x} \in S, \end{aligned} \tag{1}$$

where f is a continuous function on $\mathcal{R}^n$ and S is a constraint set in $\mathcal{R}^n$. The idea of a penalty function method is to replace problem (1) by an unconstrained problem of the form

$$\min \quad q(\mathbf{x}, c) := f(\mathbf{x}) + cP(\mathbf{x}), \tag{2}$$

where c is a positive constant and P is a continuous function on $\mathcal{R}^n$ satisfying: (i) $P(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{R}^n$, and (ii) $P(\mathbf{x}) = 0$ if and only if $\mathbf{x} \in S$. The function P is called a penalty function.

Example 1. Suppose S is defined by

$$S = \{\mathbf{x} \in \mathcal{R}^n | \; g_i(\mathbf{x}) \geq 0, \;\; i \in \mathcal{I}, \quad h_i(\mathbf{x}) = 0, \;\; i \in \mathcal{E}\}.$$

A very useful penalty function in this case is

$$P(\mathbf{x}) = \sum_{i \in \mathcal{I}} (\max\{0, -g_i(\mathbf{x})\})^2 + \sum_{i \in \mathcal{E}} (h_i(\mathbf{x}))^2.$$

L2罚函数

一般来讲L2罚函数可微，但是L1的只连续

For large c it is clear that the minimum point of problem (2) will be in a region where P is small. It is expected that the region is S. Ideally then, as $c \to \infty$ the solution point of (2) will converge to a solution of (1).

The procedure of the penalty function method is: Let $\{c_k\}$ be a sequence tending to infinity such that for each k, $c_k > 0$ and $c_{k+1} > c_k$. For each k solve the problem

$$\min \quad q(\mathbf{x}, c_k),$$

obtaining a solution point $\mathbf{x}^k$. If $c_k P(\mathbf{x}^k)$ is sufficiently small, stop.

Note that we assume that, for each k, the penalty problem has a solution.

$\{x_k\}_{k=1}^{\infty}$ 的聚点一般来讲就是原问题最优解

Example 2. Consider the problem

$$\begin{array}{ll} \min & (x_1-1)^2 + x_2^2 \\ s.t. & x_2 \geq 1. \end{array}$$

Define the penalty problem

$$\min \quad q(\mathbf{x}, c) = (x_1-1)^2 + x_2^2 + c(\max\{0, 1-x_2\})^2.$$

Let

$$\nabla q(\mathbf{x}, c) = \begin{pmatrix} 2(x_1-1) \\ 2x_2 - 2c\max\{0, 1-x_2\} \end{pmatrix} = 0,$$

we obtain the stationary point of the penalty problem

$$\mathbf{x}_c^* = \left(1; \frac{c}{1+c}\right) \to (1;1) \quad \textbf{as} \quad c \to +\infty.$$

Hence $\mathbf{x}^* = (1;1)$ is the optimal solution of this problem.

Convergence of Penalty Method

Lemma 1 *Let* $\{\mathbf{x}^k\}$ *be the sequence generated by the penalty method. Then,*

(i) $q(\mathbf{x}^k, c_k) \le q(\mathbf{x}^{k+1}, c_{k+1})$,

(ii) $P(\mathbf{x}^k) \ge P(\mathbf{x}^{k+1})$,

(iii) $f(\mathbf{x}^k) \le f(\mathbf{x}^{k+1})$.

Proof. Since $c_{k+1} > c_k$ and P is a penalty function,

$$
\begin{aligned}
q(\mathbf{x}^{k+1}, c_{k+1}) &= f(\mathbf{x}^{k+1}) + c_{k+1}P(\mathbf{x}^{k+1}) \ge f(\mathbf{x}^{k+1}) + c_k P(\mathbf{x}^{k+1}) \\
&\underline{\ge} f(\mathbf{x}^k) + c_k P(\mathbf{x}^k) = q(\mathbf{x}^k, c_k),
\end{aligned}
$$

利用x^k最小性

which proves (i).

We also have

$$f(\mathbf{x}^{k+1}) + c_{k+1}P(\mathbf{x}^{k+1}) \underline{\leq} f(\mathbf{x}^k) + c_{k+1}P(\mathbf{x}^k)$$

利用 x^{k+1} 定义

and

$$f(\mathbf{x}^k) + c_k P(\mathbf{x}^k) \leq f(\mathbf{x}^{k+1}) + c_k P(\mathbf{x}^{k+1}).$$

Adding them yields

$$(c_{k+1} - c_k)P(\mathbf{x}^{k+1}) \leq (c_{k+1} - c_k)P(\mathbf{x}^k),$$

which proves (ii).

Also, $f(\mathbf{x}^{k+1}) + c_k P(\mathbf{x}^{k+1}) \geq f(\mathbf{x}^k) + c_k P(\mathbf{x}^k)$, and hence using (ii) we obtain (iii).

Lemma 2 *Let $\mathbf{x}^*$ be a solution of problem (1). Then for each k*

$$f(\mathbf{x}^*) \geq q(\mathbf{x}^k, c_k) \geq f(\mathbf{x}^k).$$

Proof.

$$f(\mathbf{x}^*) = f(\mathbf{x}^*) + c_k P(\mathbf{x}^*) \geq f(\mathbf{x}^k) + c_k P(\mathbf{x}^k) \geq f(\mathbf{x}^k).$$

The above two lemmas give global convergence of the penalty method.

Theorem 1 *Let $\{\mathbf{x}^k\}$ be the sequence generated by the penalty method. Then, any limit point of the sequence is a solution of problem (1).* 聚点即为最优解

Proof. Suppose the subsequence $\{\mathbf{x}^k\}_{k\in\mathcal{K}}$ is a convergent subsequence of $\{\mathbf{x}^k\}$ having limit $\bar{\mathbf{x}}$. Then by the continuity of f, we have

$$\lim_{k\to\infty,\ k\in\mathcal{K}} f(\mathbf{x}^k) = f(\bar{\mathbf{x}}). \tag{3}$$

Let f^* be the optimal value associated with problem (1). Then according Lemma 1 and 2, the sequence of values $q(\mathbf{x}^k, c_k)$ is nondecreasing and bounded above by f^*. Thus,

$$\lim_{k\to\infty,\ k\in\mathcal{K}} q(\mathbf{x}^k, c_k) = q^* \le f^*. \tag{4}$$

Subtracting (3) from (4) yields

$$\lim_{k\to\infty,\ k\in\mathcal{K}} c_k P(\mathbf{x}^k) = q^* - f(\bar{\mathbf{x}}). \tag{5}$$

Since $P(\mathbf{x}^k) \geq 0$ and $c_k \to \infty$, (5) implies

$$\lim_{k\to\infty,\ k\in\mathcal{K}} P(\mathbf{x}^k) = 0,$$

which, together with the continuity of P, implies $P(\bar{\mathbf{x}}) = 0$, and hence $\bar{\mathbf{x}} \in S$.

By Lemma 2, $f(\mathbf{x}^k) \leq f^*$, which together with (3) yields $f(\bar{\mathbf{x}}) \leq f^*$. Hence $\bar{\mathbf{x}}$ is optimal for (1).

1、不同的Ck下可能无解
2、Xk的收敛性
3、如果最终无解了，之前做的Xk基本没有用了

Remark

Note that we assume that, for any $c > 0$, the penalty problem has a solution. This is not true. The following problem

$$\begin{array}{ll} \min & -x_1^5 \\ s.t. & x_1^2 + x_2^2 = 1 \end{array}$$

has an unique optimal solution $\mathbf{x}^* = (1; 0)$. But, for any $c > 0$, the penalty problem

$$\min \quad q(\mathbf{x}, c) = -x_1^5 + c(x_1^2 + x_2^2 - 1)^2$$

has no minimal point in $\mathcal{R}^2$.

Consider the problem

$$\begin{array}{ll} \min & x_1^2 + x_2^2 \\ s.t. & x_1 + 1 = 0 \end{array}$$

The corresponding penalty problem

$$\min \ q(\mathbf{x}, c) = x_1^2 + x_2^2 + c(x_1 + 1)^2$$

The Hessian matrix of $q(\mathbf{x}, c)$

$$\nabla_x^2 q(\mathbf{x}, c) = \begin{pmatrix} 2 + 2c & 0 \\ 0 & 2 \end{pmatrix}$$

Clearly, it is ill-posed.

Barrier methods

Barrier methods are applicable to problems of the form

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ s.t. \quad & \mathbf{x} \in S, \end{aligned} \tag{6}$$

where the constraint set S is robust, namely, the set has an interior and it is possible to get to any boundary point by approaching it from the interior. This kind of set often takes the form

先使得刚开始在可行域里，然后在可行域边界设置很大的障碍

不等式约束

$$S = \{\mathbf{x} \in \mathcal{R}^n |\, g_i(\mathbf{x}) \geq 0, \;\; i = 1, 2, \ldots, m\}.$$

Barrier methods work by establishing a barrier on the boundary of the feasible region that prevents a search procedure from leaving the region.

A barrier function is a continuous function B defined on $int(S)$ such that (i) $B(\mathbf{x}) \geq 0$ and (ii) $B(\mathbf{x}) \to \infty$ as $\mathbf{x}$ approaches the boundary of S.

Example 3. Let $g_i,\ i = 1, 2, \ldots, m$ be continuous functions on $\mathcal{R}^n$. Suppose

$$S = \{\mathbf{x} \in \mathcal{R}^n |\, g_i(\mathbf{x}) \geq 0,\ \ i = 1, 2, \ldots, m\}$$

is robust, and suppose the interior of S is

$$int(S) = \{\mathbf{x} \in \mathcal{R}^n |\, g_i(\mathbf{x}) > 0,\ \ i = 1, 2, \ldots, m\}.$$

Then the following functions defined on $int(S)$

$$\boxed{B(\mathbf{x}) = \sum_{i=1}^{m} \frac{1}{g_i(\mathbf{x})}}$$

and

$$\boxed{B(\mathbf{x}) = -\sum_{i=1}^{m} \log \min\{1, g_i(\mathbf{x})\}}$$

are very useful barrier function in this case.

Corresponding to the problem (6), consider the approximate problem

$$\begin{array}{ll} \min & r(\mathbf{x}, c) := f(\mathbf{x}) + \frac{1}{c} B(x) \\ s.t. & \mathbf{x} \in int(S), \end{array} \tag{7}$$

where c is a positive constant. This is a constrained problem, and indeed the constraint is somewhat more complicated than in the original problem (6). The advantage of this problem, however, is that it can be solved by using an unconstrained search technique. Since the value of the objective function approaches infinity near the boundary of S, the search technique will automatically remain within the interior of S, and the constraint need not be accounted for explicitly. Thus, problem (7) is from a computational viewpoint unconstrained.

The barrier method is quite analogous to the penalty method. Let $\{c_k\}$ be a sequence tending to infinity such that for each k, $c_k > 0$ and $c_{k+1} > c_k$. For each k solve the problem

$$\min \quad r(\mathbf{x}, c_k),$$

obtaining a solution point $\mathbf{x}^k$. If $B(\mathbf{x}^k)/c_k$ is sufficiently small, stop.

Note that we assume that, for each k, the problem $\min\ r(\mathbf{x}, c_k)$ has a solution.

For the mixed constrained optimization problem as in Example 1, the combined penalty and barrier method is used:

$$\min\ F(\mathbf{x}, c) = f(x) - \frac{1}{c}\sum_{i\in\mathcal{I}} \log g_i(\mathbf{x}) + c\sum_{i\in\mathcal{E}} (h_i(\mathbf{x}))^2.$$

等式约束还是要用罚函数

Example 4. Consider the problem

$$\begin{aligned} \min \quad & x_1 + x_2 \\ s.t. \quad & -x_1^2 + x_2 \geq 0, \\ & x_1 \geq 0. \end{aligned}$$

Define the problem

$$\min \quad r(\mathbf{x}, c) = x_1 + x_2 - \frac{1}{c}[\ln(-x_1^2 + x_2) + \ln x_1].$$

Let

$$\nabla r(\mathbf{x}, c) = \begin{pmatrix} 1 - \frac{1}{c}\left(\frac{-2x_1}{-x_1^2+x_2} - \frac{1}{x_1}\right) \\ 1 - \frac{1}{c}\frac{1}{-x_1^2+x_2} \end{pmatrix} = 0,$$

we obtain the stationary point

$$\mathbf{x}_c^* = \left(\frac{1}{4}\left(-1 + \sqrt{1 + 8/c}\right); \frac{3}{2c} - \frac{1}{8}\left(-1 + \sqrt{1 + 8/c}\right)\right).$$

Hence as $c \to +\infty$, $\mathbf{x}_c^* \to \mathbf{x}^* = (0; 0)$ which is the optimal solution.

Convergence of Barrier Method

Lemma 3 *Let $\{\mathbf{x}^k\}$ be the sequence generated by the barrier method. Then,*

(i) $r(\mathbf{x}^k, c_k) \geq r(\mathbf{x}^{k+1}, c_{k+1})$,

(ii) $B(\mathbf{x}^k) \leq B(\mathbf{x}^{k+1})$,

(iii) $f(\mathbf{x}^k) \geq f(\mathbf{x}^{k+1})$.

Lemma 4 *Let $\mathbf{x}^*$ be a solution of problem (6). Then for each k*

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^k) \leq r(\mathbf{x}^k, c_k).$$

Theorem 2 *Any limit point of a sequence $\{\mathbf{x}^k\}$ generated by the barrier method is a solution to problem (6).*

Augmented Lagrange Penalty Method

Hestenes (1969) and Powell (1969) independently combined Lagrange function and penalty function to propose the augmented Lagrange penalty method to solve the equality-constrained optimization problems. Rockafeller (1973) extended the method to solve the inequality-constrained optimization problems. We first consider the equality-constrained optimization problem

对于等式约束的

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & h_i(\mathbf{x}) = 0, \; i \in \mathcal{E}. \end{array} \tag{8}$$

We wish to construct an unconstrained optimization problem such that $\mathbf{x}^*$ is optimal for this problem, where $\mathbf{x}^*$ is an optimal solution to (8).

Remark

Consider the problem

直接使用Lagrange函数失效的情况

$$\begin{array}{ll} \min & x_1^2 - 3x_2 - x_2^2 \\ s.t. & x_2 = 0 \end{array}$$

Clearly, it has an optimal solution $\mathbf{x}^* = (0; 0)$. However, the Lagrange function

$$L(\mathbf{x}, v) = x_1^2 - 3x_2 - x_2^2 - vx_2$$

has no minimum point w.r.t. $\mathbf{x}$ since the Hessian matrix

$$\nabla_x^2 L(\mathbf{x}, v) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

Procedure of Augment Lagrangian Method

增广Lagrange算法

For the equality-constraint problem (8), define the augment Lagrange function

$$\boxed{\varphi(\mathbf{x}, v, c) = f(\mathbf{x}) - \sum_{i\in\mathcal{E}} v_i h_i(\mathbf{x}) + \frac{c}{2}\sum_{i\in\mathcal{E}}(h_i(\mathbf{x}))^2.}$$

At the kth iteration, let $\mathbf{x}^k$ be the minimum point of $\varphi(\mathbf{x}, v^k, c)$ with the multiplier estimate v^k and the penalty parameter c. We have

$$\nabla_x \varphi(\mathbf{x}^k, v^k, c) = \nabla f(\mathbf{x}^k) - \sum_{i\in\mathcal{E}}(v_i^k - ch_i(\mathbf{x}^k))\nabla h_i(\mathbf{x}^k) = 0.$$

Hence, update

是解那么就是KKT点，于是

$$v_i^{k+1} = v_i^k - ch_i(\mathbf{x}^k), \quad i \in \mathcal{E}.$$

比Lagrange方法多了罚函数项，比直接罚函数法多了Lagrange乘式项

Step 1. Given starting point $\mathbf{x}^0$ and the multiplier estimate v^1, $c > 0, \varepsilon > 0, \alpha > 1$ and $\beta \in (0,1)$. Set $k := 1$.

Step 2. Compute the optimal solution $\mathbf{x}^k$ to the problem $\min \ \varphi(\mathbf{x}, v^k, c)$ from the starting point $\mathbf{x}^{k-1}$. If $\|h(\mathbf{x}^k)\| < \varepsilon$, stop. If

$$\frac{\|h(\mathbf{x}^k)\|}{\|h(\mathbf{x}^{k-1})\|} \geq \beta,$$

增加C的目的是保证罚函数要取等，因此如果h较小则不必增加C

set $c := \alpha c$ and go to Step 3. Otherwise, return Step 3.

Step 3. Update $v_i^{k+1} = v_i^k - ch_i(\mathbf{x}^k)$, $i \in \mathcal{E}$. Set $k := k+1$ and go to Step 2.

Example

$$\min \ \frac{1}{2}x_1^2 + \frac{1}{6}x_2^2 \quad \textbf{s.t. } x_1 + x_2 = 1$$

This problem has an optimal solution $x^* = (1/4; 3/4)$ with the Lagrange multiplier $v^* = 1/4$.

Define the augment Lagrange penalty function

$$\varphi(\mathbf{x}, v, c) = \frac{1}{2}x_1^2 + \frac{1}{6}x_2^2 - v(x_1 + x_2 - 1) + \frac{c}{2}(x_1 + x_2 - 1)^2.$$

The sequence generated by the augment Lagrange penalty method is

$$\mathbf{x}^k = \left(\frac{c_k + v^k}{1 + 4c_k}; \quad \frac{3c_k + 3v^k}{1 + 4c_k} \right), \quad v^{k+1} = \frac{v^k + c_k}{1 + 4c_k}.$$

The sequence v^k is convergent as c is large. Take $c = 10$, we have the limit point of v^k is $\bar{v} = \frac{1}{4}$, i.e., $\bar{v} = v^*$. Thus, $\mathbf{x}^k$ tends to x^*.

The sequence generated by the augment Lagrange penalty method is

$$\mathbf{x}^k = \left(\frac{c_k + v^k}{1 + 4c_k};\ \frac{3c_k + 3v^k}{1 + 4c_k} \right).$$

After 8 iterations, we obtain $(0.25; 0.75)$.

The sequence generated by the penalty method is

$$\mathbf{x}^k = \left(\frac{c_k}{1 + 4c_k};\ \frac{3c_k}{1 + 4c_k} \right).$$

After 20 iterations, we obtain $(0.25; 0.75)$.

Augment Lagrange function for inequality-constrained problem

We extend the Lagrange multiplier penalty method to solve the inequality-constrained optimization problem

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & g_i(\mathbf{x}) \geq 0, \ i \in \mathcal{I}. \end{array} \tag{9}$$

Reformulate problem (9) as the following equality-constrained problem by introducing variables y_i

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & g_i(\mathbf{x}) - y_i^2 = 0, \ i \in \mathcal{I}. \end{array}$$

The corresponding augment Lagrange function

$$\Psi(\mathbf{x}, \mathbf{y}, w, c) = f(\mathbf{x}) - \sum_{i \in \mathcal{I}} w_i(g_i(\mathbf{x}) - y_i^2) + \frac{c}{2} \sum_{i \in \mathcal{I}} (g_i(\mathbf{x}) - y_i^2)^2.$$

We first solve $\min_y \ \Psi(\mathbf{x}, \mathbf{y}, w, c)$ to obtain the optimal solution $\mathbf{y}^*(\mathbf{x}, w, c)$.

Since

$$\Psi(\mathbf{x}, \mathbf{y}, w, c) = f(\mathbf{x}) + \sum_{i \in \mathcal{I}} \left\{ \frac{c}{2} \left(y_i^2 - \frac{1}{c}(cg_i(\mathbf{x}) - w_i) \right)^2 - \frac{w_i^2}{2c} \right\},$$

the minimum point $\mathbf{y}^*(\mathbf{x}, w, c)$ is expressed as

$$(y_i^*)^2 = \begin{cases} \frac{1}{c}(cg_i(\mathbf{x}) - w_i), & \textbf{if } cg_i(\mathbf{x}) - w_i \geq 0, \\ 0, & \textbf{otherwise.} \end{cases}$$

Thus, the augment Lagrange function for (9) is expressed as

$$\varphi(\mathbf{x}, w, c) = f(\mathbf{x}) + \frac{1}{2c} \sum_{i \in \mathcal{I}} \left\{ [\max\{0, w_i - cg_i(\mathbf{x})\}]^2 - w_i^2 \right\}.$$

The Augmented Lagrangian Method for Convex Optimization

Consider the convex optimization case $h(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$. The augmented Lagrangian method (ALM) is:

Start from any $(\mathbf{x}^0, \mathbf{y}^0)$, we compute a new iterate pair

$$\mathbf{x}^{k+1} = arg\min L(\mathbf{x}, \mathbf{y}^k), \quad \mathbf{y}^{k+1} = \mathbf{y}^k + \beta h(\mathbf{x}^{k+1}),$$

where

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

Analysis of the Augmented Lagrangian Method

Since $\mathbf{x}^{k+1}$ 为L的最小值因此有梯度为0

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

we only need to concern about whether or not $\|A\mathbf{x}^k - \mathbf{b}\|$ converges to zero and how fast it converges.

First, from the convexity of $f(\mathbf{x})$, we have

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

which implies that

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

That is, the error in non-increasing.

Let $\mathbf{x}^*$ be the minimizer. Again, from the convexity of $f(\mathbf{x})$, we have

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

Thus, from the positivity of the cross product, we have

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

Sum up from 0 to k of the inequality, we have

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

Then, it gives the desired error bound:

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

The Alternating Direction Method with Multipliers

For the ADMM method, we consider structured problem

$$\begin{array}{ll} \min & f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) \\ s.t. & A_1\mathbf{x}_1 + A_2\mathbf{x}_2 = \mathbf{b}. \end{array}$$

Consider

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

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

$$\begin{aligned} \mathbf{x}_1^{k+1} &= arg\min_{\mathbf{x}_1} L(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{y}^k) \\ \mathbf{x}_2^{k+1} &= arg\min_{\mathbf{x}_2} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{y}^k) \\ \mathbf{y}^{k+1} &= \mathbf{y}^k + \beta(A_1\mathbf{x}_1^{k+1} + A_2\mathbf{x}_2^{k+1} - \mathbf{b}). \end{aligned}$$

Again, we can prove that the iterates converge with the same speed.

The ADMM method with three blocks

What about ADMM for

$$\min \quad f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_3(\mathbf{x}_3) \quad s.t. \quad A_1\mathbf{x}_1 + A_2\mathbf{x}_2 + A_3\mathbf{x}_3 = \mathbf{b},$$

with

$$\begin{aligned} L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{y}) \quad &= \quad f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_3(\mathbf{x}_3) + \mathbf{y}^T(A_1\mathbf{x}_1 + A_2\mathbf{x}_2 + A_3\mathbf{x}_3 - \mathbf{b}) \\ &\quad + \frac{\beta}{2}\|A_1\mathbf{x}_1 + A_2\mathbf{x}_2 + A_3\mathbf{x}_3 - \mathbf{b}\|^2. \end{aligned}$$

Then, for any given $(\mathbf{x}_1^k, \mathbf{x}_2^k, \mathbf{x}_3^k, \mathbf{y}^k)$, we compute a new iterate

$$\begin{aligned} \mathbf{x}_1^{k+1} \quad &= \quad arg \min\nolimits_{\mathbf{x}_1} L(\mathbf{x}_1, \mathbf{x}_2^k, \mathbf{x}_3^k, \mathbf{y}^k) \\ \mathbf{x}_2^{k+1} \quad &= \quad arg \min\nolimits_{\mathbf{x}_2} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2, \mathbf{x}_3^k, \mathbf{y}^k) \\ \mathbf{x}_3^{k+1} \quad &= \quad arg \min\nolimits_{\mathbf{x}_3} L(\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \mathbf{x}_3, \mathbf{y}^k) \\ \mathbf{y}^{k+1} \quad &= \quad \mathbf{y}^k + \beta(A_1\mathbf{x}_1^{k+1} + A_2\mathbf{x}_2^{k+1} + A_3\mathbf{x}_3^{k+1} - \mathbf{b}). \end{aligned}$$

Does it converges?

Consider the problem:

$$
\begin{array}{ll}
\min & 0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 \\
s.t. & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \mathbf{0}.
\end{array}
$$

The unique minimizer is $\mathbf{0}$.

Then, the ADMM with $\beta = 1$ would be a linear matrix mapping

$$\begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 6 & 0 & 0 & 0 & 0 \\ 5 & 7 & 9 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \\ \mathbf{y}^{k+1} \end{pmatrix} = \begin{pmatrix} 0 & -4 & -5 & 1 & 1 & 1 \\ 0 & 0 & -7 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1^{k} \\ x_2^{k} \\ x_3^{k} \\ \mathbf{y}^{k} \end{pmatrix},$$

which can be reduced to

$$\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \mathbf{y}^{k+1} \end{pmatrix} = M \begin{pmatrix} x_2^{k} \\ x_3^{k} \\ \mathbf{y}^{k} \end{pmatrix},$$

where

$$M = \frac{1}{162}\begin{pmatrix} 144 & -9 & -9 & -9 & 18 \\ 8 & 157 & -5 & 13 & -8 \\ 64 & 122 & 122 & -58 & -64 \\ 56 & -35 & -35 & 91 & -56 \\ -88 & -26 & -26 & -62 & 88 \end{pmatrix}.$$

But the spectral radius of the matrix is greater that 1, indicating the mapping is not a contraction.