ISyE8813
Augmented Lagrangian (AL) Method

Renato D.C. Monteiro

March 6, 2024

1 Problem description and assumptions

Consider the problem

$$\begin{aligned} &\min \ (f+h)(x) \\ &\text{s.t. } Ax = b \end{aligned}$$

where

- $f : \mathbb{R}^n \to \mathbb{R}$ is a L-smooth convex function on $\mathbb{R}^n$
- $h \in \overline{\text{Conv}}\,(\mathbb{R}^n)$
- $A : \mathbb{R}^n \to \mathbb{R}^m$ is a nonzero linear operator and $b \in \mathbb{R}^m$
- set of optimal solutions is nonempty

Optimality condition: $\bar{x}$ is an optimal solution if and only if there exists $\bar{p} \in \mathbb{R}^m$ such that $(x,p) = (\bar{x}, \bar{p})$ satisfies

$$\nabla f(x) + \partial h(x) + A^* p \ni 0, \quad b - Ax = 0$$

Hence, $\bar{z} = (\bar{x}, \bar{p})$ is a solution of the inclusion

$$0 \in T(z)$$

where

$$T(z) = T(x,p) := \begin{bmatrix} \nabla f(x) + A^* p \\ b - Ax \end{bmatrix} + \begin{bmatrix} \partial h(x) \\ 0 \end{bmatrix}$$

Observe that T is of the form

$$T(z) = F(z) + \partial g(z)$$

where, for every $z = (x, p) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$F(z) = F(x,p) := \begin{bmatrix} \nabla f(x) + A^* p \\ b - Ax \end{bmatrix}$$

and $g(z) = g(x,p) := h(x)$.

Since F is a continuous monotone operator and g is continuous, it follows that T is maximal monotone

2 Augmented Lagrangian Method

For a given $\beta > 0$, define

$$L_\beta(x;p) = (f+h)(x) + p^T(Ax-b) + \frac{\beta}{2}\|Ax-b\|^2$$

2.1 Exact version

Algorithm 1 (Exact Prox ALM)

0. Given $(x_0, p_0) \in \mathbb{R}^n \times \mathbb{R}^m$ and symmetric positive semidefinite matrix $Q \in \mathbb{R}^{n\times n}$
1. Compute
$$x \in \operatorname{Arg}\min_{x' \in \mathbb{R}^n} \; L_\beta(x';p_0) + \frac{1}{2}\|x'-x_0\|_Q^2. \tag{1}$$
2. Set
$$p = p_0 + \beta(Ax-b)$$
3. Set $(x_0, p_0) \leftarrow (x, p)$ and go to step 1

---

where

$$\|\cdot\|_Q = \sqrt{\langle \cdot, Q(\cdot)\rangle}$$

Optimality Condition for (1):

$$0 \in \nabla f(x) + \partial h(x) + A^*[p_0 + \beta(Ax-b)] + Q(x-x_0)$$

and hence

$$\begin{aligned} &0 \in \nabla f(x) + \partial h(x) + A^* p + Q(x-x_0) \\ &p = p_0 + \beta(Ax-b) \end{aligned} \tag{2}$$

or equivalently,

$$\begin{aligned} Q(x_0 - x) &\in \nabla f(x) + \partial h(x) + A^* p \\ \frac{p_0 - p}{\beta} &= b - Ax \end{aligned}$$

which is an exact prox iteration of the form

$$\frac{\nabla w(z_0) - \nabla w(z)}{\lambda} \in T(z)$$

where $\lambda = 1$, $z_0 = (x_0, p_0)$, $z = (x, p)$,

$$w(x,p) = \frac{1}{2}\|x\|_Q^2 + \frac{1}{2\beta}\|p\|^2$$

and

$$T(z) = T(x,p) := \begin{bmatrix} \nabla f(x) + A^* p \\ b - Ax \end{bmatrix} + \begin{bmatrix} \partial h(x) \\ 0 \end{bmatrix}$$

2.2 Inexact version

For a given $\beta > 0$, define

$$L_\beta(x;p) = (f+h)(x) + p^T(Ax-b) + \frac{\beta}{2}\|Ax-b\|^2$$

Algorithm 2 (Inexact Prox ALM)

0. Given $(x_0, p_0) \in \mathbb{R}^n \times \mathbb{R}^m$ and symmetric positive definite matrix $Q \in \mathbb{R}^{n\times n}$
1. Compute
$$\tilde{x} \approx \mathrm{Arg}\min_{x' \in \mathbb{R}^n} L_\beta(x';p_0) + \frac{1}{2}\|x' - x_0\|_Q^2. \tag{3}$$
2. Set
$$p = p_0 + \beta(A\tilde{x} - b) \tag{4}$$
3. Set $(x_0, p_0) \leftarrow (\tilde{x}, p)$ and go to step 1

For some $\sigma \in (0,1)$, assume that $\tilde{x} \in \operatorname{dom} h$ is an approximate solution of (3) in the sense that there exists $\tilde{v} \in \mathbb{R}^n$ such that

$$\begin{aligned}&\tilde{v} \in \partial\left(L_\beta(\cdot;p_0) + \frac{1}{2}\|\cdot - x_0\|_Q^2\right)(\tilde{x}),\\ &\|\tilde{v}\|_{Q^{-1}}^2 \le \sigma^2\left(\|\tilde{x} - x_0\|_Q^2 + \beta\|A\tilde{x} - b\|^2\right)\end{aligned} \tag{5}$$

The above inclusion implies that

$$\tilde{v} \in \nabla f(\tilde{x}) + \partial h(\tilde{x}) + A^*[p_0 + \beta(A\tilde{x} - b)] + Q(\tilde{x} - x_0)$$

and hence that

$$\tilde{v} \in \nabla f(\tilde{x}) + \partial h(\tilde{x}) + A^* p + Q(\tilde{x} - x_0) \tag{6}$$

Defining

$$\tilde{u} := Q^{-1}\tilde{v},$$

it follows that the previous inclusion is equivalent to

$$Q(x_0 + \tilde{u} - \tilde{x}) \in \nabla f(\tilde{x}) + \partial h(\tilde{x}) + A^* p$$

Letting

$$x := \tilde{x} - \tilde{u}$$

we have

$$\begin{aligned}Q(x_0 - x) &\in \nabla f(\tilde{x}) + \partial h(\tilde{x}) + A^* p\\ \frac{p_0 - p}{\beta} &= b - A\tilde{x}\end{aligned}$$

which is an inexact prox iteration of the form

$$\frac{\nabla w(z_0) - \nabla w(z)}{\lambda} \in T(\tilde{z})$$

where $\lambda = 1$, $z_0 = (x_0, p_0)$, $z = (x, p)$, $\tilde{z} = (\tilde{x}, p)$,

$$w(x,p) := \frac{1}{2}\|x\|_Q^2 + \frac{1}{2\beta}\|p\|^2,$$

and

$$T(z) = T(x,p) := \begin{bmatrix} \nabla f(x) + A^* p \\ b - Ax \end{bmatrix} + \begin{bmatrix} \partial h(x) \\ 0 \end{bmatrix}$$

Now

$$(dw)_{z_0}(\tilde{z}) = \frac{1}{2}\|x_0 - \tilde{x}\|_Q^2 + \frac{1}{2\beta}\|p - p_0\|^2 = \frac{1}{2}\|x_0 - \tilde{x}\|_Q^2 + \frac{\beta}{2}\|A\tilde{x} - b\|^2$$

and

$$(dw)_z(\tilde{z}) = \frac{1}{2}\|x - \tilde{x}\|_Q^2 = \frac{1}{2}\|\tilde{u}\|_Q^2 = \frac{1}{2}\|\tilde{v}\|_{Q^{-1}}^2$$

Thus, it follows from (5) and the above two identities that

$$(dw)_z(\tilde{z}) \le \sigma^2 (dw)_{z_0}(\tilde{z})$$

So, the Inexact Prox ALM is a special case of the generalized HPE framework.

According to the HPE convergence theory, the residual

$$r^k = \begin{pmatrix} r_1^k \\ r_2^k \end{pmatrix} := \begin{pmatrix} Q(x_{k-1} - x_k) \\ \beta^{-1}(p_{k-1} - p_k) \end{pmatrix}$$

satisfies

$$(\|r^k\|_*)^2 = \mathcal{O}(1/k)$$

Now consider the operator $P : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$

$$P(x,p) = \left(Qx, \frac{1}{\beta}p\right)$$

and note that

$$w(x,p) = \frac{1}{2}\|(x,p)\|_P^2$$

Then, P^{-1} is

$$P^{-1}(y,q) = \left(Q^{-1}y, \beta q\right)$$

Now, if $\|\cdot\| = \|\cdot\|_P$ for some positive definite operator P, then we know that

$$\|\cdot\|_* = \|\cdot\|_{P^{-1}}$$

Hence, it follows that

$$\|P^{-1/2} r^k\|^2 = (\|r^k\|_{P^{-1}})^2 = \|r_k\|_* = \mathcal{O}(1/k)$$

Thus

$$\|Q^{-1/2} r_1^k\|^2 + \beta \|r_2^k\|^2 = \mathcal{O}(1/k)$$

So

$$\|b - A\tilde{x}^k\| = \|r_2^k\| = \mathcal{O}\left(1/\sqrt{\beta k}\right)$$

and

$$r_1^k \in \nabla f(\tilde{x}_k) + \partial h(\tilde{x}_k) + A^* p_k, \quad \|Q^{-1/2} r_1^k\| = \mathcal{O}\left(1/\sqrt{k}\right)$$

so that

$$\|r_1^k\| = \mathcal{O}\left(\|Q\|^{1/2}/\sqrt{k}\right)$$