Chapter 15

ADMM

Underlying Spaces: In this chapter all the underlying spaces are Euclidean $\mathbb{R}^n$ spaces endowed with the dot product and the l_2-norm.

15.1 The Augmented Lagrangian Method

Consider the problem

$$H_{\text{opt}} = \min\{H(\mathbf{x},\mathbf{z}) \equiv h_1(\mathbf{x}) + h_2(\mathbf{z}) : \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}\}, \tag{15.1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{m\times p}$, and $\mathbf{c} \in \mathbb{R}^m$. For now, we will assume that h_1 and h_2 are proper closed and convex functions. Later on, we will specify exact conditions on the data $(h_1, h_2, \mathbf{A}, \mathbf{B}, \mathbf{c})$ that will guarantee the validity of some convergence results. To find a dual problem of (15.1), we begin by constructing a Lagrangian:

$$L(\mathbf{x},\mathbf{z};\mathbf{y}) = h_1(\mathbf{x}) + h_2(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\rangle.$$

The dual objective function is therefore given by

$$\begin{aligned} q(\mathbf{y}) &= \min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{z}\in\mathbb{R}^p} \{h_1(\mathbf{x}) + h_2(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\rangle\} \\ &= -h_1^*(-\mathbf{A}^T\mathbf{y}) - h_2^*(-\mathbf{B}^T\mathbf{y}) - \langle \mathbf{c}, \mathbf{y}\rangle, \end{aligned}$$

and the dual problem is given by

$$q_{\text{opt}} = \max_{\mathbf{y}\in\mathbb{R}^m} \{-h_1^*(-\mathbf{A}^T\mathbf{y}) - h_2^*(-\mathbf{B}^T\mathbf{y}) - \langle \mathbf{c}, \mathbf{y}\rangle\} \tag{15.2}$$

or, in minimization form, by

$$\min_{\mathbf{y}\in\mathbb{R}^m} \left\{h_1^*(-\mathbf{A}^T\mathbf{y}) + h_2^*(-\mathbf{B}^T\mathbf{y}) + \langle \mathbf{c}, \mathbf{y}\rangle\right\}. \tag{15.3}$$

The proximal point method was discussed in Section 10.5, where its convergence was established. The general update step of the proximal point method employed on problem (15.3) takes the form ($\rho > 0$ being a given constant)

$$\mathbf{y}^{k+1} = \operatorname{argmin}_{\mathbf{y}\in\mathbb{R}^m} \left\{h_1^*(-\mathbf{A}^T\mathbf{y}) + h_2^*(-\mathbf{B}^T\mathbf{y}) + \langle \mathbf{c}, \mathbf{y}\rangle + \frac{1}{2\rho}\|\mathbf{y} - \mathbf{y}^k\|^2\right\}. \tag{15.4}$$

Assuming that the sum and affine rules of subdifferential calculus (Theorems 3.40 and 3.43) hold for the relevant functions, we can conclude by Fermat's optimality condition (Theorem 3.63) that (15.4) holds if and only if

$$\mathbf{0} \in -\mathbf{A}\partial h_1^*(-\mathbf{A}^T\mathbf{y}^{k+1}) - \mathbf{B}\partial h_2^*(-\mathbf{B}^T\mathbf{y}^{k+1}) + \mathbf{c} + \frac{1}{\rho}(\mathbf{y}^{k+1} - \mathbf{y}^k). \tag{15.5}$$

Using the conjugate subgradient theorem (Corollary 4.21), we obtain that $\mathbf{y}^{k+1}$ satisfies (15.5) if and only if $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})$, where $\mathbf{x}^{k+1}$ and $\mathbf{z}^{k+1}$ satisfy

$$\begin{aligned}
\mathbf{x}^{k+1} &\in \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n}\{\langle \mathbf{A}^T\mathbf{y}^{k+1}, \mathbf{x}\rangle + h_1(\mathbf{x})\},\\
\mathbf{z}^{k+1} &\in \operatorname{argmin}_{\mathbf{z}\in\mathbb{R}^p}\{\langle \mathbf{B}^T\mathbf{y}^{k+1}, \mathbf{z}\rangle + h_2(\mathbf{z})\}.
\end{aligned}$$

Plugging the update equation for $\mathbf{y}^{k+1}$ into the above, we conclude that $\mathbf{y}^{k+1}$ satisfies (15.5) if and only if

$$\begin{aligned}
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}),\\
\mathbf{x}^{k+1} &\in \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n}\{\langle \mathbf{A}^T(\mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})), \mathbf{x}\rangle + h_1(\mathbf{x})\},\\
\mathbf{z}^{k+1} &\in \operatorname{argmin}_{\mathbf{z}\in\mathbb{R}^p}\{\langle \mathbf{B}^T(\mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})), \mathbf{z}\rangle + h_2(\mathbf{z})\},
\end{aligned}$$

meaning if and only if (using the properness and convexity of h_1 and h_2, as well as Fermat's optimality condition)

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}), \tag{15.6}$$
$$\mathbf{0} \in \mathbf{A}^T(\mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})) + \partial h_1(\mathbf{x}^{k+1}), \tag{15.7}$$
$$\mathbf{0} \in \mathbf{B}^T(\mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})) + \partial h_2(\mathbf{z}^{k+1}). \tag{15.8}$$

Conditions (15.7) and (15.8) are satisfied if and only if $(\mathbf{x}^{k+1}, \mathbf{z}^{k+1})$ is a coordinate-wise minimum (see Definition 14.2) of the function

$$\tilde{H}(\mathbf{x}, \mathbf{z}) \equiv h_1(\mathbf{x}) + h_2(\mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2.$$

By Lemma 14.7, coordinate-wise minima points of $\tilde{H}$ are exactly the minimizers of $\tilde{H}$, and therefore the system (15.6), (15.7), (15.8) leads us to the following primal representation of the dual proximal point method, known as the *augmented Lagrangian method*.

The Augmented Lagrangian Method

Initialization: $\mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$.
General step: for any $k = 0, 1, 2, \ldots$ execute the following steps:

$$(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}) \in \underset{\mathbf{x}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{ h_1(\mathbf{x}) + h_2(\mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2 \right\} \tag{15.9}$$
$$\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}). \tag{15.10}$$

Naturally, step (15.9) is called the *primal update step*, while (15.10) is the *dual update step*.

Remark 15.1 (augmented Lagrangian). *The augmented Lagrangian associated with the main problem* (15.1) *is defined to be*

$$L_\rho(\mathbf{x}, \mathbf{z}; \mathbf{y}) = h_1(\mathbf{x}) + h_2(\mathbf{z}) + \langle \mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c} \rangle + \frac{\rho}{2}\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}\|^2.$$

Obviously, $L_0 = L$ is the Lagrangian function, and L_ρ for $\rho > 0$ can be considered as a penalized version of the Lagrangian. The primal update step (15.9) *can be equivalently written as*

$$(\mathbf{x}^{k+1}, \mathbf{z}^{k+1}) \in \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n, \mathbf{z} \in \mathbb{R}^p} L_\rho(\mathbf{x}, \mathbf{z}; \mathbf{y}^k).$$

The above representation of the primal update step as the outcome of the minimization of the augmented Lagrangian function is the reason for the name of the method.

15.2 Alternating Direction Method of Multipliers (ADMM)

The augmented Lagrangian method is in general not an implementable method since the primal update step (15.9) can be as hard to solve as the original problem. One source of difficulty is the coupling term between the $\mathbf{x}$ and the $\mathbf{z}$ variables, which is of the form $\rho(\mathbf{x}^T\mathbf{A}^T\mathbf{B}\mathbf{z})$. The approach used in the *alternating direction method of multipliers* (ADMM) to tackle this difficulty is to replace the exact minimization in the primal update step (15.9) by one iteration of the alternating minimization method; that is, the objective function of (15.9) is first minimized w.r.t. $\mathbf{x}$, and then w.r.t. $\mathbf{z}$.

ADMM

Initialization: $\mathbf{x}^0 \in \mathbb{R}^n$, $\mathbf{z}^0 \in \mathbb{R}^p$, $\mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$.
General step: for any $k = 0, 1, \ldots$ execute the following:

(a) $\mathbf{x}^{k+1} \in \operatorname{argmin}_{\mathbf{x}} \left\{ h_1(\mathbf{x}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\|^2 \right\}$;

(b) $\mathbf{z}^{k+1} \in \operatorname{argmin}_{\mathbf{z}} \left\{ h_2(\mathbf{z}) + \frac{\rho}{2} \left\| \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right\|^2 \right\}$;

(c) $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})$.

15.2.1 Alternating Direction Proximal Method of Multipliers (AD-PMM)

We will actually analyze a more general method than ADMM in which a quadratic proximity term is added to the objective in the minimization problems of steps

(a) and (b). We will assume that we are given two positive semidefinite matrices $\mathbf{G} \in \mathbb{S}^n_+, \mathbf{Q} \in \mathbb{S}^p_+$, and recall that $\|\mathbf{x}\|^2_{\mathbf{G}} = \mathbf{x}^T\mathbf{G}\mathbf{x}, \|\mathbf{x}\|^2_{\mathbf{Q}} = \mathbf{x}^T\mathbf{Q}\mathbf{x}$.

AD-PMM

Initialization: $\mathbf{x}^0 \in \mathbb{R}^n$, $\mathbf{z}^0 \in \mathbb{R}^p$, $\mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$.
General step: for any $k = 0, 1, \ldots$ execute the following:

(a) $\mathbf{x}^{k+1} \in \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left\{h_1(\mathbf{x}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2 + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|^2_{\mathbf{G}}\right\}$;

(b) $\mathbf{z}^{k+1} \in \operatorname{argmin}_{\mathbf{z}\in\mathbb{R}^p} \left\{h_2(\mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z} - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\| + \frac{1}{2}\|\mathbf{z} - \mathbf{z}^k\|^2_{\mathbf{Q}}\right\}$;

(c) $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})$.

One important motivation for considering AD-PMM is that by using the proximity terms, the minimization problems in steps (a) and (b) of ADMM can be simplified considerably by choosing $\mathbf{G} = \alpha\mathbf{I} - \rho\mathbf{A}^T\mathbf{A}$ with $\alpha \geq \rho\lambda_{\max}(\mathbf{A}^T\mathbf{A})$ and $\mathbf{Q} = \beta\mathbf{I} - \rho\mathbf{B}^T\mathbf{B}$ with $\beta \geq \rho\lambda_{\max}(\mathbf{B}^T\mathbf{B})$. Then obviously $\mathbf{G}, \mathbf{Q} \in \mathbb{S}^n_+$, and the function that needs to be minimized in the $\mathbf{x}$-step can be simplified as follows:

$$\begin{aligned}
&h_1(\mathbf{x}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2 + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|^2_{\mathbf{G}}\\
&= h_1(\mathbf{x}) + \frac{\rho}{2}\left\|\mathbf{A}(\mathbf{x} - \mathbf{x}^k) + \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\|^2 + \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|^2_{\mathbf{G}}\\
&= h_1(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{A}(\mathbf{x} - \mathbf{x}^k)\|^2 + \left\langle \rho\mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\rangle\\
&\quad + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{x}^k\|^2 - \frac{\rho}{2}\|\mathbf{A}(\mathbf{x} - \mathbf{x}^k)\|^2 + \text{constant}\\
&= h_1(\mathbf{x}) + \rho\left\langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\rangle + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{x}^k\|^2 + \text{constant},
\end{aligned}$$

where by "constant" we mean a term that does not depend on $\mathbf{x}$. We can therefore conclude that step (a) of AD-PMM amounts to

$$\mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left\{h_1(\mathbf{x}) + \rho\left\langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\rangle + \frac{\alpha}{2}\|\mathbf{x} - \mathbf{x}^k\|^2\right\}, \tag{15.11}$$

and, similarly, step (b) of AD-PMM is the same as

$$\mathbf{z}^{k+1} = \operatorname{argmin}_{\mathbf{z}\in\mathbb{R}^p} \left\{h_2(\mathbf{z}) + \rho\left\langle \mathbf{B}\mathbf{z}, \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right\rangle + \frac{\beta}{2}\|\mathbf{z} - \mathbf{z}^k\|^2\right\}. \tag{15.12}$$

The functions minimized in the update formulas (15.11) and (15.12) are actually constructed from the functions minimized in steps (a) and (b) of ADMM by linearizing the quadratic term and adding a proximity term. This is the reason why the resulting method will be called the *alternating direction linearized proximal method of multipliers* (AD-LPMM). We can also write the update formulas (15.11) and

(15.12) in terms of proximal operators. Indeed, (15.11) can be rewritten equivalently as

$$\mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}} \left\{ \frac{1}{\alpha} h_1(\mathbf{x}) + \frac{1}{2} \left\| \mathbf{x} - \left(\mathbf{x}^k - \frac{\rho}{\alpha} \mathbf{A}^T \left(\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right) \right) \right\|^2 \right\}.$$

That is,

$$\mathbf{x}^{k+1} = \operatorname{prox}_{\frac{1}{\alpha}h_1} \left[\mathbf{x}^k - \frac{\rho}{\alpha} \mathbf{A}^T \left(\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right) \right].$$

Similarly, the $\mathbf{z}$-step can be rewritten as

$$\mathbf{z}^{k+1} = \operatorname{prox}_{\frac{1}{\beta}h_2} \left[\mathbf{z}^k - \frac{\rho}{\beta} \mathbf{B}^T \left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right) \right].$$

We can now summarize and write explicitly the AD-LPMM method.

AD-LPMM

Initialization: $\mathbf{x}^0 \in \mathbb{R}^n$, $\mathbf{z}^0 \in \mathbb{R}^p$, $\mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$, $\alpha \geq \rho\lambda_{\max}(\mathbf{A}^T\mathbf{A})$, $\beta \geq \rho\lambda_{\max}(\mathbf{B}^T\mathbf{B})$.
General step: for any $k = 0, 1, \ldots$ execute the following:

(a) $\mathbf{x}^{k+1} = \operatorname{prox}_{\frac{1}{\alpha}h_1} \left[\mathbf{x}^k - \frac{\rho}{\alpha} \mathbf{A}^T \left(\mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right) \right]$;

(b) $\mathbf{z}^{k+1} = \operatorname{prox}_{\frac{1}{\beta}h_2} \left[\mathbf{z}^k - \frac{\rho}{\beta} \mathbf{B}^T \left(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k \right) \right]$;

(c) $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c})$.

15.3 Convergence Analysis of AD-PMM

In this section we will develop a rate of convergence analysis of AD-PMM employed on problem (15.1). Note that both ADMM and AD-LPMM are special cases of AD-PMM. The following set of assumptions will be made.

Assumption 15.2.

(A) $h_1 : \mathbb{R}^n \to (-\infty, \infty]$ *and* $h_2 : \mathbb{R}^p \to (-\infty, \infty]$ *are proper closed convex functions.*

(B) $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{m\times p}$, $\mathbf{c} \in \mathbb{R}^m$, $\rho > 0$.

(C) $\mathbf{G} \in \mathbb{S}^n_+$, $\mathbf{Q} \in \mathbb{S}^p_+$.

(D) *For any* $\mathbf{a} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^p$ *the optimal sets of the problems*

$$\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ h_1(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}\|^2 + \frac{1}{2}\|\mathbf{x}\|^2_{\mathbf{G}} + \langle \mathbf{a}, \mathbf{x} \rangle \right\}$$

and

$$\min_{\mathbf{z}\in\mathbb{R}^p}\left\{h_2(\mathbf{z})+\frac{\rho}{2}\|\mathbf{B}\mathbf{z}\|^2+\frac{1}{2}\|\mathbf{z}\|_{\mathbf{Q}}^2+\langle\mathbf{b},\mathbf{z}\rangle\right\}$$

are nonempty.

(E) *There exists* $\hat{\mathbf{x}}\in\operatorname{ri}(\operatorname{dom}(h_1))$ *and* $\hat{\mathbf{z}}\in\operatorname{ri}(\operatorname{dom}(h_2))$ *for which* $\mathbf{A}\hat{\mathbf{x}}+\mathbf{B}\hat{\mathbf{z}}=\mathbf{c}$.

(F) *Problem* (15.1) *has a nonempty optimal set, denoted by* X^*, *and the corresponding optimal value is* $H_{\rm opt}$.

Property (D) guarantees that the AD-PMM method is actually a well-defined method.

By the strong duality theorem for convex problems (see Theorem A.1), under Assumption 15.2, it follows that strong duality holds for the pair of problems (15.1) and (15.2).

Theorem 15.3 (strong duality for the pair of problems (15.1) and (15.2)). *Suppose that Assumption* 15.2 *holds, and let* $H_{\rm opt}, q_{\rm opt}$ *be the optimal values of the primal and dual problems* (15.1) *and* (15.2), *respectively. Then* $H_{\rm opt}=q_{\rm opt}$, *and the dual problem* (15.2) *possesses an optimal solution.*

We will now prove an $O(1/k)$ rate of convergence result of the sequence generated by AD-PMM.

Theorem 15.4 ($O(1/k)$ rate of convergence of AD-PMM).[87] *Suppose that Assumption* 15.2 *holds. Let* $\{(\mathbf{x}^k,\mathbf{z}^k)\}_{k\geq 0}$ *be the sequence generated by AD-PMM for solving problem* (15.1). *Let* $(\mathbf{x}^*,\mathbf{z}^*)$ *be an optimal solution of problem* (15.1) *and* $\mathbf{y}^*$ *be an optimal solution of the dual problem* (15.2). *Suppose that* $\gamma>0$ *is any constant satisfying* $\gamma\geq 2\|\mathbf{y}^*\|$. *Then for all* $n\geq 0$,

$$H(\mathbf{x}^{(n)},\mathbf{z}^{(n)})-H_{\rm opt}\leq\frac{\|\mathbf{x}^*-\mathbf{x}^0\|_{\mathbf{G}}^2+\|\mathbf{z}^*-\mathbf{z}^0\|_{\mathbf{C}}^2+\frac{1}{\rho}(\gamma+\|\mathbf{y}^0\|)^2}{2(n+1)},\quad(15.13)$$

$$\|\mathbf{A}\mathbf{x}^{(n)}+\mathbf{B}\mathbf{z}^{(n)}-\mathbf{c}\|\leq\frac{\|\mathbf{x}^*-\mathbf{x}^0\|_{\mathbf{G}}^2+\|\mathbf{z}^*-\mathbf{z}^0\|_{\mathbf{C}}^2+\frac{1}{\rho}(\gamma+\|\mathbf{y}^0\|)^2}{\gamma(n+1)},\quad(15.14)$$

where $\mathbf{C}=\rho\mathbf{B}^T\mathbf{B}+\mathbf{Q}$ *and*

$$\mathbf{x}^{(n)}=\frac{1}{n+1}\sum_{k=0}^n\mathbf{x}^{k+1},\ \mathbf{z}^{(n)}=\frac{1}{n+1}\sum_{k=0}^n\mathbf{z}^{k+1}.$$

Proof. By Fermat's optimality condition (Theorem 3.63) and the update steps (a) and (b) of AD-PMM, it follows that $\mathbf{x}^{k+1}$ and $\mathbf{z}^{k+1}$ satisfy

$$-\rho\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^{k+1}+\mathbf{B}\mathbf{z}^k-\mathbf{c}+\frac{1}{\rho}\mathbf{y}^k\right)-\mathbf{G}(\mathbf{x}^{k+1}-\mathbf{x}^k)\in\partial h_1(\mathbf{x}^{k+1}),\quad(15.15)$$

$$-\rho\mathbf{B}^T\left(\mathbf{A}\mathbf{x}^{k+1}+\mathbf{B}\mathbf{z}^{k+1}-\mathbf{c}+\frac{1}{\rho}\mathbf{y}^k\right)-\mathbf{Q}(\mathbf{z}^{k+1}-\mathbf{z}^k)\in\partial h_2(\mathbf{z}^{k+1}).\quad(15.16)$$

[87] The proof of Theorem 15.4 on the rate of convergence of AD-PMM is based on a combination of the proof techniques of He and Yuan [65] and Gao and Zhang [58].

We will use the following notation:

$$\begin{aligned}
\tilde{\mathbf{x}}^k &= \mathbf{x}^{k+1},\\
\tilde{\mathbf{z}}^k &= \mathbf{z}^{k+1},\\
\tilde{\mathbf{y}}^k &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^k - \mathbf{c}).
\end{aligned}$$

Using (15.15), (15.16), the subgradient inequality, and the above notation, we obtain that for any $\mathbf{x} \in \mathrm{dom}(h_1)$ and $\mathbf{z} \in \mathrm{dom}(h_2)$,

$$h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \left\langle \rho\mathbf{A}^T\left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right) + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \right\rangle \geq 0,$$

$$h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \left\langle \rho\mathbf{B}^T\left(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c} + \frac{1}{\rho}\mathbf{y}^k\right) + \mathbf{Q}(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \right\rangle \geq 0.$$

Using the definition of $\tilde{\mathbf{y}}^k$, the above two inequalities can be rewritten as

$$\begin{aligned}
h_1(\mathbf{x}) - h_1(\tilde{\mathbf{x}}^k) + \left\langle \mathbf{A}^T\tilde{\mathbf{y}}^k + \mathbf{G}(\tilde{\mathbf{x}}^k - \mathbf{x}^k), \mathbf{x} - \tilde{\mathbf{x}}^k \right\rangle &\geq 0,\\
h_2(\mathbf{z}) - h_2(\tilde{\mathbf{z}}^k) + \left\langle \mathbf{B}^T\tilde{\mathbf{y}}^k + (\rho\mathbf{B}^T\mathbf{B} + \mathbf{Q})(\tilde{\mathbf{z}}^k - \mathbf{z}^k), \mathbf{z} - \tilde{\mathbf{z}}^k \right\rangle &\geq 0.
\end{aligned}$$

Adding the above two inequalities and using the identity

$$\mathbf{y}^{k+1} - \mathbf{y}^k = \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c}),$$

we can conclude that for any $\mathbf{x} \in \mathrm{dom}(h_1)$, $\mathbf{z} \in \mathrm{dom}(h_2)$, and $\mathbf{y} \in \mathbb{R}^m$,

$$H(\mathbf{x},\mathbf{z}) - H(\tilde{\mathbf{x}}^k, \tilde{\mathbf{z}}^k) + \left\langle \begin{pmatrix} \mathbf{x} - \tilde{\mathbf{x}}^k \\ \mathbf{z} - \tilde{\mathbf{z}}^k \\ \mathbf{y} - \tilde{\mathbf{y}}^k \end{pmatrix}, \begin{pmatrix} \mathbf{A}^T\tilde{\mathbf{y}}^k \\ \mathbf{B}^T\tilde{\mathbf{y}}^k \\ -\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{B}\tilde{\mathbf{z}}^k + \mathbf{c} \end{pmatrix} - \begin{pmatrix} \mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) \\ \mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) \\ \frac{1}{\rho}(\mathbf{y}^k - \mathbf{y}^{k+1}) \end{pmatrix} \right\rangle \geq 0, \tag{15.17}$$

where $\mathbf{C} = \rho\mathbf{B}^T\mathbf{B} + \mathbf{Q}$. We will use the following identity that holds for any positive semidefinite matrix $\mathbf{P}$:

$$(\mathbf{a} - \mathbf{b})^T\mathbf{P}(\mathbf{c} - \mathbf{d}) = \frac{1}{2}\left(\|\mathbf{a} - \mathbf{d}\|_{\mathbf{P}}^2 - \|\mathbf{a} - \mathbf{c}\|_{\mathbf{P}}^2 + \|\mathbf{b} - \mathbf{c}\|_{\mathbf{P}}^2 - \|\mathbf{b} - \mathbf{d}\|_{\mathbf{P}}^2\right).$$

Using the above identity, we can conclude that

$$\begin{aligned}
(\mathbf{x} - \tilde{\mathbf{x}}^k)^T\mathbf{G}(\mathbf{x}^k - \tilde{\mathbf{x}}^k) &= \frac{1}{2}\left(\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2 + \|\tilde{\mathbf{x}}^k - \mathbf{x}^k\|_{\mathbf{G}}^2\right)\\
&\geq \frac{1}{2}\|\mathbf{x} - \tilde{\mathbf{x}}^k\|_{\mathbf{G}}^2 - \frac{1}{2}\|\mathbf{x} - \mathbf{x}^k\|_{\mathbf{G}}^2, \qquad (15.18)
\end{aligned}$$

as well as

$$(\mathbf{z} - \tilde{\mathbf{z}}^k)^T\mathbf{C}(\mathbf{z}^k - \tilde{\mathbf{z}}^k) = \frac{1}{2}\|\mathbf{z} - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 - \frac{1}{2}\|\mathbf{z} - \mathbf{z}^k\|_{\mathbf{C}}^2 + \frac{1}{2}\|\mathbf{z}^k - \tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 \tag{15.19}$$

and

$$\begin{aligned}
&2(\mathbf{y} - \tilde{\mathbf{y}}^k)^T(\mathbf{y}^k - \mathbf{y}^{k+1})\\
&= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \|\tilde{\mathbf{y}}^k - \mathbf{y}^k\|^2 - \|\tilde{\mathbf{y}}^k - \mathbf{y}^{k+1}\|^2\\
&= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2\\
&-\|\mathbf{y}^k + \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}) - \mathbf{y}^k - \rho(\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\tilde{\mathbf{z}}^k - \mathbf{c})\|^2\\
&= \|\mathbf{y} - \mathbf{y}^{k+1}\|^2 - \|\mathbf{y} - \mathbf{y}^k\|^2 + \rho^2\|\mathbf{A}\tilde{\mathbf{x}}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}\|^2 - \rho^2\|\mathbf{B}(\mathbf{z}^k - \tilde{\mathbf{z}}^k)\|^2.
\end{aligned}$$

Therefore,

$$\frac{1}{\rho}(\mathbf{y}-\tilde{\mathbf{y}}^k)^T(\mathbf{y}^k-\mathbf{y}^{k+1}) \geq \frac{1}{2\rho}\left(\|\mathbf{y}-\mathbf{y}^{k+1}\|^2 - \|\mathbf{y}-\mathbf{y}^k\|^2\right) - \frac{\rho}{2}\|\mathbf{B}(\mathbf{z}^k-\tilde{\mathbf{z}}^k)\|^2. \quad (15.20)$$

Denoting

$$\mathbf{H} = \begin{pmatrix} \mathbf{G} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\rho}\mathbf{I} \end{pmatrix},$$

as well as

$$\mathbf{w} = \begin{pmatrix} \mathbf{x} \\ \mathbf{z} \\ \mathbf{y} \end{pmatrix}, \quad \mathbf{w}^k = \begin{pmatrix} \mathbf{x}^k \\ \mathbf{z}^k \\ \mathbf{y}^k \end{pmatrix}, \quad \tilde{\mathbf{w}}^k = \begin{pmatrix} \tilde{\mathbf{x}}^k \\ \tilde{\mathbf{z}}^k \\ \tilde{\mathbf{y}}^k \end{pmatrix},$$

we obtain by combining (15.18), (15.19), and (15.20) that

$$\begin{aligned} \left\langle \begin{pmatrix} \mathbf{x}-\tilde{\mathbf{x}}^k \\ \mathbf{z}-\tilde{\mathbf{z}}^k \\ \mathbf{y}-\tilde{\mathbf{y}}^k \end{pmatrix}, \begin{pmatrix} \mathbf{G}(\mathbf{x}^k-\tilde{\mathbf{x}}^k) \\ \mathbf{C}(\mathbf{z}^k-\tilde{\mathbf{z}}^k) \\ \frac{1}{\rho}(\mathbf{y}^k-\mathbf{y}^{k+1}) \end{pmatrix} \right\rangle &\geq \frac{1}{2}\|\mathbf{w}-\mathbf{w}^{k+1}\|_{\mathbf{H}}^2 - \frac{1}{2}\|\mathbf{w}-\mathbf{w}^k\|_{\mathbf{H}}^2 \\ &\quad + \frac{1}{2}\|\mathbf{z}^k-\tilde{\mathbf{z}}^k\|_{\mathbf{C}}^2 - \frac{\rho}{2}\|\mathbf{B}(\mathbf{z}^k-\tilde{\mathbf{z}}^k)\|^2 \\ &\geq \frac{1}{2}\|\mathbf{w}-\mathbf{w}^{k+1}\|_{\mathbf{H}}^2 - \frac{1}{2}\|\mathbf{w}-\mathbf{w}^k\|_{\mathbf{H}}^2. \end{aligned}$$

Combining the last inequality with (15.17), we obtain that for any $\mathbf{x} \in \mathrm{dom}(h_1)$, $\mathbf{z} \in \mathrm{dom}(h_2)$, and $\mathbf{y} \in \mathbb{R}^m$,

$$H(\mathbf{x},\mathbf{z}) - H(\tilde{\mathbf{x}}^k,\tilde{\mathbf{z}}^k) + \langle \mathbf{w}-\tilde{\mathbf{w}}^k, \mathbf{F}\tilde{\mathbf{w}}^k+\tilde{\mathbf{c}}\rangle \geq \frac{1}{2}\|\mathbf{w}-\mathbf{w}^{k+1}\|_{\mathbf{H}}^2 - \frac{1}{2}\|\mathbf{w}-\mathbf{w}^k\|_{\mathbf{H}}^2, \quad (15.21)$$

where

$$\mathbf{F} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{A}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{B}^T \\ -\mathbf{A} & -\mathbf{B} & \mathbf{0} \end{pmatrix}, \qquad \tilde{\mathbf{c}} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{c} \end{pmatrix}.$$

Note that

$$\begin{aligned} \langle \mathbf{w}-\tilde{\mathbf{w}}^k, \mathbf{F}\tilde{\mathbf{w}}^k+\tilde{\mathbf{c}}\rangle &= \langle \mathbf{w}-\tilde{\mathbf{w}}^k, \mathbf{F}(\tilde{\mathbf{w}}^k-\mathbf{w})+\mathbf{F}\mathbf{w}+\tilde{\mathbf{c}}\rangle \\ &= \langle \mathbf{w}-\tilde{\mathbf{w}}^k, \mathbf{F}\mathbf{w}+\tilde{\mathbf{c}}\rangle, \end{aligned}$$

where the second equality follows from the fact that $\mathbf{F}$ is skew symmetric (meaning $\mathbf{F}^T = -\mathbf{F}$). We can thus conclude that (15.21) can be rewritten as

$$H(\mathbf{x},\mathbf{z}) - H(\tilde{\mathbf{x}}^k,\tilde{\mathbf{z}}^k) + \langle \mathbf{w}-\tilde{\mathbf{w}}^k, \mathbf{F}\mathbf{w}+\tilde{\mathbf{c}}\rangle \geq \frac{1}{2}\|\mathbf{w}-\mathbf{w}^{k+1}\|_{\mathbf{H}}^2 - \frac{1}{2}\|\mathbf{w}-\mathbf{w}^k\|_{\mathbf{H}}^2.$$

Summing the above inequality over $k = 0, 1, \ldots, n$ yields the inequality

$$(n+1)H(\mathbf{x},\mathbf{z}) - \sum_{k=0}^{n} H(\tilde{\mathbf{x}}^k, \tilde{\mathbf{z}}^k) + \left\langle (n+1)\mathbf{w} - \sum_{k=0}^{n} \tilde{\mathbf{w}}^k, \mathbf{F}\mathbf{w} + \tilde{\mathbf{c}} \right\rangle \geq -\frac{1}{2}\|\mathbf{w} - \mathbf{w}^0\|_{\mathbf{H}}^2.$$

Defining

$$\mathbf{w}^{(n)} = \frac{1}{n+1}\sum_{k=0}^{n} \tilde{\mathbf{w}}^k, \mathbf{x}^{(n)} = \frac{1}{n+1}\sum_{k=0}^{n} \mathbf{x}^{k+1}, \mathbf{z}^{(n)} = \frac{1}{n+1}\sum_{k=0}^{n} \mathbf{z}^{k+1}$$

and using the convexity of H, we obtain that

$$H(\mathbf{x},\mathbf{z}) - H(\mathbf{x}^{(n)}, \mathbf{z}^{(n)}) + \langle \mathbf{w} - \mathbf{w}^{(n)}, \mathbf{F}\mathbf{w} + \tilde{\mathbf{c}} \rangle + \frac{1}{2(n+1)}\|\mathbf{w} - \mathbf{w}^0\|_{\mathbf{H}}^2 \geq 0.$$

Using (again) the skew-symmetry of $\mathbf{F}$, we can conclude that the above inequality is the same as

$$H(\mathbf{x},\mathbf{z}) - H(\mathbf{x}^{(n)}, \mathbf{z}^{(n)}) + \langle \mathbf{w} - \mathbf{w}^{(n)}, \mathbf{F}\mathbf{w}^{(n)} + \tilde{\mathbf{c}} \rangle + \frac{1}{2(n+1)}\|\mathbf{w} - \mathbf{w}^0\|_{\mathbf{H}}^2 \geq 0.$$

In other words, for any $\mathbf{x} \in \mathrm{dom}(h_1)$ and $\mathbf{z} \in \mathrm{dom}(h_1)$,

$$H(\mathbf{x}^{(n)}, \mathbf{z}^{(n)}) - H(\mathbf{x},\mathbf{z}) + \langle \mathbf{w}^{(n)} - \mathbf{w}, \mathbf{F}\mathbf{w}^{(n)} + \tilde{\mathbf{c}} \rangle \leq \frac{1}{2(n+1)}\|\mathbf{w} - \mathbf{w}^0\|_{\mathbf{H}}^2. \quad (15.22)$$

Let $(\mathbf{x}^*, \mathbf{z}^*)$ be an optimal solution of problem (15.1). Then $H(\mathbf{x}^*, \mathbf{z}^*) = H_{\mathrm{opt}}$ and $\mathbf{A}\mathbf{x}^* + \mathbf{B}\mathbf{z}^* = \mathbf{c}$. Plugging $\mathbf{x} = \mathbf{x}^*, \mathbf{z} = \mathbf{z}^*$, and the expressions for $\mathbf{w}^{(n)}, \mathbf{w}, \mathbf{w}^0, \mathbf{F}, \mathbf{H}, \tilde{\mathbf{c}}$ into (15.22), we obtain (denoting $\mathbf{y}^{(n)} = \frac{1}{n+1}\sum_{k=0}^{n} \tilde{\mathbf{y}}^k$)

$$\begin{aligned}
&H(\mathbf{x}^{(n)}, \mathbf{z}^{(n)}) - H_{\mathrm{opt}} + \langle \mathbf{x}^{(n)} - \mathbf{x}^*, \mathbf{A}^T\mathbf{y}^{(n)} \rangle + \langle \mathbf{z}^{(n)} - \mathbf{z}^*, \mathbf{B}^T\mathbf{y}^{(n)} \rangle \\
&+ \langle \mathbf{y}^{(n)} - \mathbf{y}, -\mathbf{A}\mathbf{x}^{(n)} - \mathbf{B}\mathbf{z}^{(n)} + \mathbf{c} \rangle \\
&\leq \frac{1}{2(n+1)}\left\{ \|\mathbf{x}^* - \mathbf{x}^0\|_{\mathbf{G}}^2 + \|\mathbf{z}^* - \mathbf{z}^0\|_{\mathbf{C}}^2 + \frac{1}{\rho}\|\mathbf{y} - \mathbf{y}^0\|^2 \right\}.
\end{aligned}$$

Cancelling terms and using the fact that $\mathbf{A}\mathbf{x}^* + \mathbf{B}\mathbf{z}^* = \mathbf{c}$, we obtain that the last inequality is the same as

$$H(\mathbf{x}^{(n)}, \mathbf{z}^{(n)}) - H_{\mathrm{opt}} + \langle \mathbf{y}, \mathbf{A}\mathbf{x}^{(n)} + \mathbf{B}\mathbf{z}^{(n)} - \mathbf{c} \rangle \leq \frac{\|\mathbf{x}^* - \mathbf{x}^0\|_{\mathbf{G}}^2 + \|\mathbf{z}^* - \mathbf{z}^0\|_{\mathbf{C}}^2 + \frac{1}{\rho}\|\mathbf{y} - \mathbf{y}^0\|^2}{2(n+1)}.$$

Since the above inequality holds for any $\mathbf{y} \in \mathbb{R}^m$, we can take the maximum of both sides over all $\mathbf{y} \in B[\mathbf{0}, \gamma]$ and obtain the inequality

$$H(\mathbf{x}^{(n)}, \mathbf{z}^{(n)}) - H_{\mathrm{opt}} + \gamma\|\mathbf{A}\mathbf{x}^{(n)} + \mathbf{B}\mathbf{z}^{(n)} - \mathbf{c}\| \leq \frac{\|\mathbf{x}^* - \mathbf{x}^0\|_{\mathbf{G}}^2 + \|\mathbf{z}^* - \mathbf{z}^0\|_{\mathbf{C}}^2 + \frac{1}{\rho}(\gamma + \|\mathbf{y}^0\|)^2}{2(n+1)}.$$

Since $\gamma \geq 2\|\mathbf{y}^*\|$ for some optimal dual solution $\mathbf{y}^*$ and strong duality holds (Theorem 15.3), it follows by Theorem 3.60 that the two inequalities (15.13) and (15.14) hold. $\square$

15.4 Minimizing $f_1(\mathbf{x}) + f_2(\mathbf{A}\mathbf{x})$

In this section we consider the model

$$\min_{\mathbf{x}\in\mathbb{R}^n} \{f_1(\mathbf{x}) + f_2(\mathbf{A}\mathbf{x})\}, \tag{15.23}$$

where f_1, f_2 are proper closed convex functions and $\mathbf{A} \in \mathbb{R}^{m\times n}$. As usual, $\rho > 0$ is a given constant. An implicit assumption will be that f_1 and f_2 are "proximable," which loosely speaking means that the prox operator of λf_1 and λf_2 can be efficiently computed for any $\lambda > 0$. This is obviously a "virtual" assumption, and its importance is only in the fact that it dictates the development of algorithms that rely on prox computations of λf_1 and λf_2.

Problem (15.23) can rewritten as

$$\min_{\mathbf{x}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^m} \{f_1(\mathbf{x}) + f_2(\mathbf{z}) : \mathbf{A}\mathbf{x} - \mathbf{z} = \mathbf{0}\}. \tag{15.24}$$

This fits the general model (15.1) with $h_1 = f_1, h_2 = f_2$, $\mathbf{B} = -\mathbf{I}$, and $\mathbf{c} = \mathbf{0}$. A direct implementation of ADMM leads to the following scheme ($\rho > 0$ is a given constant):

$$\begin{aligned}
\mathbf{x}^{k+1} &\in \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left[f_1(\mathbf{x}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right\|^2 \right], \qquad (15.25)\\
\mathbf{z}^{k+1} &= \operatorname{argmin}_{\mathbf{z}\in\mathbb{R}^m} \left[f_2(\mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z} + \frac{1}{\rho}\mathbf{y}^k\right\|^2 \right],\\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}).
\end{aligned}$$

The $\mathbf{z}$-step can be rewritten as a prox step, thus resulting in the following algorithm for solving problem (15.23).

Algorithm 1 [ADMM for solving (15.23)—version 1]

- **Initialization:** $\mathbf{x}^0 \in \mathbb{R}^n, \mathbf{z}^0, \mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$.
- **General step ($k \geq 0$):**
  (a) $\mathbf{x}^{k+1} \in \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n} \left[f_1(\mathbf{x}) + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right\|^2 \right]$;
  (b) $\mathbf{z}^{k+1} = \operatorname{prox}_{\frac{1}{\rho}f_2}\left(\mathbf{A}\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}^k\right)$;
  (c) $\mathbf{y}^{k+1} = \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1})$.

Step (a) of Algorithm 1 might be difficult to compute since the minimization in step (a) is more involved than a prox computation due to the quadratic term $\frac{\rho}{2}\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}$. We can actually employ ADMM in a different way that will refrain from the type of computation made in step (a). For that, we will rewrite problem (15.23) as

$$\min_{\mathbf{x},\mathbf{w}\in\mathbb{R}^n,\mathbf{z}\in\mathbb{R}^m} \{f_1(\mathbf{w}) + f_2(\mathbf{z}) : \mathbf{Ax} - \mathbf{z} = \mathbf{0}, \mathbf{x} - \mathbf{w} = \mathbf{0}\}.$$

The above problem fits model (15.1) with $h_1 \equiv 0, h_2(\mathbf{z},\mathbf{w}) = f_1(\mathbf{z}) + f_2(\mathbf{w})$, $\mathbf{B} = -\mathbf{I}$, and $\binom{\mathbf{A}}{\mathbf{I}}$ taking the place of $\mathbf{A}$. The dual vector $\mathbf{y} \in \mathbb{R}^{m+n}$ is of the form $\mathbf{y} = (\mathbf{y}_1^T, \mathbf{y}_2^T)^T$, where $\mathbf{y}_1 \in \mathbb{R}^m$ and $\mathbf{y}_2 \in \mathbb{R}^n$. In the above reformulation we have two blocks of vectors: $\mathbf{x}$ and $(\mathbf{z},\mathbf{w})$. The $\mathbf{x}$-step is given by

$$\begin{aligned}\mathbf{x}^{k+1} &= \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n}\left[\left\|\mathbf{Ax} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}_1^k\right\|^2 + \left\|\mathbf{x} - \mathbf{w}^k + \frac{1}{\rho}\mathbf{y}_2^k\right\|^2\right]\\ &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k - \frac{1}{\rho}\mathbf{y}_1^k\right] + \mathbf{w}^k - \frac{1}{\rho}\mathbf{y}_2^k\right).\end{aligned}$$

The $(\mathbf{z},\mathbf{w})$-step is

$$\begin{aligned}\mathbf{z}^{k+1} &= \operatorname{prox}_{\frac{1}{\rho}f_2}\left(\mathbf{Ax}^{k+1} + \frac{1}{\rho}\mathbf{y}_1^k\right),\\ \mathbf{w}^{k+1} &= \operatorname{prox}_{\frac{1}{\rho}f_1}\left(\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_2^k\right).\end{aligned}$$

The method is summarized in the following.

Algorithm 2 [ADMM for solving (15.23)—version 2]

- **Initialization:** $\mathbf{x}^0, \mathbf{w}^0, \mathbf{y}_2^0 \in \mathbb{R}^n, \mathbf{z}^0, \mathbf{y}_1^0 \in \mathbb{R}^m, \rho > 0$.
- **General step ($\boldsymbol{k \geq 0}$):**

$$\begin{aligned}\mathbf{x}^{k+1} &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k - \frac{1}{\rho}\mathbf{y}_1^k\right] + \mathbf{w}^k - \frac{1}{\rho}\mathbf{y}_2^k\right),\\ \mathbf{z}^{k+1} &= \operatorname{prox}_{\frac{1}{\rho}f_2}\left(\mathbf{Ax}^{k+1} + \frac{1}{\rho}\mathbf{y}_1^k\right),\\ \mathbf{w}^{k+1} &= \operatorname{prox}_{\frac{1}{\rho}f_1}\left(\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_2^k\right),\\ \mathbf{y}_1^{k+1} &= \mathbf{y}_1^k + \rho(\mathbf{Ax}^{k+1} - \mathbf{z}^{k+1}),\\ \mathbf{y}_2^{k+1} &= \mathbf{y}_2^k + \rho(\mathbf{x}^{k+1} - \mathbf{w}^{k+1}).\end{aligned}$$

Algorithm 2 might still be too computationally demanding since it involves the evaluation of the inverse of $\mathbf{I} + \mathbf{A}^T\mathbf{A}$ (or at least the evaluation of $\mathbf{A}^T\mathbf{A}$ and a solution of a linear system at each iteration), which might be a difficult task in large-scale problems. We can alternatively employ AD-LPMM on problem (15.24) and obtain the following scheme that does not involve any matrix inverse calculations.

Algorithm 3 [AD-LPMM for solving (15.23)]

- **Initialization:** $\mathbf{x}^0 \in \mathbb{R}^n, \mathbf{z}^0, \mathbf{y}^0 \in \mathbb{R}^m$, $\rho > 0$, $\alpha \geq \rho\lambda_{\max}(\mathbf{A}^T\mathbf{A}), \beta \geq \rho$.
- **General step ($k \geq 0$):**

$$\begin{aligned}
\mathbf{x}^{k+1} &= \operatorname{prox}_{\frac{1}{\alpha}f_1}\left[\mathbf{x}^k - \frac{\rho}{\alpha}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right)\right],\\
\mathbf{z}^{k+1} &= \operatorname{prox}_{\frac{1}{\beta}f_2}\left[\mathbf{z}^k + \frac{\rho}{\beta}\left(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right)\right],\\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}).
\end{aligned}$$

The above scheme has the advantage that it only requires simple linear algebra operations (no more than matrix/vector multiplications) and prox evaluations of λf_1 and λf_2 for different values of $\lambda > 0$.

Example 15.5 (l_1-regularized least squares). Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left\{\frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{x}\|_1\right\}, \tag{15.26}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$ and $\lambda > 0$. Problem (15.26) fits the composite model (15.23) with $f_1(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ and $f_2(\mathbf{y}) \equiv \frac{1}{2}\|\mathbf{y} - \mathbf{b}\|_2^2$. For any $\gamma > 0$, $\operatorname{prox}_{\gamma f_1} = \mathcal{T}_{\gamma\lambda}$ (by Example 6.8) and $\operatorname{prox}_{\gamma f_2}(\mathbf{y}) = \frac{\mathbf{y}+\gamma\mathbf{b}}{\gamma+1}$ (by Section 6.2.3). Step (a) of Algorithm 1 (first version of ADMM) has the form

$$\mathbf{x}^{k+1} \in \operatorname{argmin}_{\mathbf{x}\in\mathbb{R}^n}\left[\lambda\|\mathbf{x}\|_1 + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right\|^2\right],$$

which actually means that this version of ADMM is completely useless since it suggests to solve an l_1-regularized least squares problem by a sequence of l_1-regularized least squares problems.

Algorithm 2 (second version of ADMM) has the following form.

ADMM, version 2 (Algorithm 2):

$$\begin{aligned}
\mathbf{x}^{k+1} &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k - \frac{1}{\rho}\mathbf{y}_1^k\right] + \mathbf{w}^k - \frac{1}{\rho}\mathbf{y}_2^k\right),\\
\mathbf{z}^{k+1} &= \frac{\rho\mathbf{A}\mathbf{x}^{k+1} + \mathbf{y}_1^k + \mathbf{b}}{\rho + 1},\\
\mathbf{w}^{k+1} &= \mathcal{T}_{\frac{\lambda}{\rho}}\left(\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_2^k\right),\\
\mathbf{y}_1^{k+1} &= \mathbf{y}_1^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}),\\
\mathbf{y}_2^{k+1} &= \mathbf{y}_2^k + \rho(\mathbf{x}^{k+1} - \mathbf{w}^{k+1}).
\end{aligned}$$

An implementation of the above ADMM variant will require to compute the matrix $\mathbf{A}^T\mathbf{A}$ in a preprocess and to solve at each iteration an $n \times n$ linear system

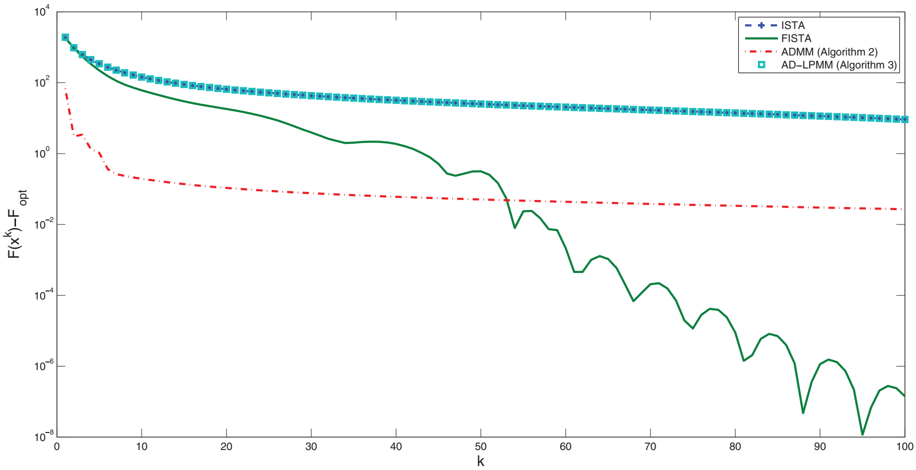

Figure 15.1. *Results of* 100 *iterations of ISTA, FISTA, ADMM (Algorithm* 2*) and AD-LPMM (Algorithm* 3*) on an* l_1*-regularized least squares problem.*

(or, alternatively, compute the inverse of $\mathbf{I} + \mathbf{A}^T\mathbf{A}$ in a preprocess). These operations might be difficult to execute in large-scale problems.

The general step of Algorithm 3 (which is essentially AD-LPMM) with $\alpha = \lambda_{\max}(\mathbf{A}^T\mathbf{A})\rho$ and $\beta = \rho$ takes the following form (denoting $L = \lambda_{\max}(\mathbf{A}^T\mathbf{A})$).

AD-LPMM (Algorithm 3):

$$\begin{aligned}
\mathbf{x}^{k+1} &= \mathcal{T}_{\frac{\lambda}{L\rho}}\left[\mathbf{x}^k - \frac{1}{L}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right)\right],\\
\mathbf{z}^{k+1} &= \frac{\rho\mathbf{A}\mathbf{x}^{k+1} + \mathbf{y}^k + \mathbf{b}}{\rho+1},\\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}).
\end{aligned}$$

The dominant computations in AD-LPMM are matrix/vector multiplications.

To illustrate the performance of the above two methods, we repeat the experiment described in Example 10.38 on the l_1-regularized least squares problem. We ran ADMM and AD-LPMM on the exact same instance, and the decay of the function values as a function of the iteration index k for the first 100 iterations is described in Figure 15.1. Clearly, ISTA and AD-LPMM exhibit the same performance, while ADMM seems to outperform both of them. This is actually not surprising since the computations carried out at each iteration of ADMM (solution of linear systems) are much heavier than the computations per iteration of AD-LPMM and ISTA (matrix/vector multiplications). In that respect, the comparison is in fact not fair and biased in favor of ADMM. What is definitely interesting is that FISTA significantly outperforms ADMM starting from approximately 50 iterations despite the fact that it is a simpler algorithm that requires substantially less computational effort per iteration. One possible reason is that FISTA is a method with a provably

$O(1/k^2)$ rate of convergence in function values, while ADMM is only guaranteed to converge at a rate of $O(1/k)$. ∎

Example 15.6 (robust regression). Consider the problem

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1, \tag{15.27}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Problem (15.27) fits the composite model (15.23) with $f_1 \equiv 0$ and $f_2(\mathbf{y}) = \|\mathbf{y} - \mathbf{b}\|_1$. Let $\rho > 0$. For any $\gamma > 0$, $\text{prox}_{\gamma f_1}(\mathbf{y}) = \mathbf{y}$ and $\text{prox}_{\gamma f_2}(\mathbf{y}) = \mathcal{T}_\gamma(\mathbf{y} - \mathbf{b}) + \mathbf{b}$ (by Example 6.8 and Theorem 6.11). Therefore, the general step of Algorithm 1 (first version of ADMM) takes the following form.

ADMM, version 1 (Algorithm 1):

$$\begin{aligned}
\mathbf{x}^{k+1} &= \text{argmin}_{\mathbf{x}\in\mathbb{R}^m} \left\| \mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k \right\|^2, \\
\mathbf{z}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left(\mathbf{A}\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}^k - \mathbf{b}\right) + \mathbf{b}, \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}).
\end{aligned}$$

The general step of Algorithm 2 (second version of ADMM) reads as follows:

$$\begin{aligned}
\mathbf{x}^{k+1} &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k - \frac{1}{\rho}\mathbf{y}_1^k\right] + \mathbf{w}^k - \frac{1}{\rho}\mathbf{y}_2^k\right), \\
\mathbf{z}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left(\mathbf{A}\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_1^k - \mathbf{b}\right) + \mathbf{b}, \\
\mathbf{w}^{k+1} &= \mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_2^k, \\
\mathbf{y}_1^{k+1} &= \mathbf{y}_1^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}), \\
\mathbf{y}_2^{k+1} &= \mathbf{y}_2^k + \rho(\mathbf{x}^{k+1} - \mathbf{w}^{k+1}).
\end{aligned}$$

Plugging the expression for $\mathbf{w}^{k+1}$ into the update formula of $\mathbf{y}_2^{k+1}$, we obtain that $\mathbf{y}_2^{k+1} = \mathbf{0}$. Thus, if we start with $\mathbf{y}_2^0 = 0$, then $\mathbf{y}_2^k = \mathbf{0}$ for all $k \geq 0$, and consequently $\mathbf{w}^k = \mathbf{x}^k$ for all k. The algorithm thus reduces to the following.

ADMM, version 2 (Algorithm 2):

$$\begin{aligned}
\mathbf{x}^{k+1} &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{z}^k - \frac{1}{\rho}\mathbf{y}_1^k\right] + \mathbf{x}^k\right), \\
\mathbf{z}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left(\mathbf{A}\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_1^k - \mathbf{b}\right) + \mathbf{b}, \\
\mathbf{y}_1^{k+1} &= \mathbf{y}_1^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}).
\end{aligned}$$

Algorithm 3 (which is essentially AD-LPMM) with $\alpha = \lambda_{\max}(\mathbf{A}^T\mathbf{A})\rho$ and $\beta = \rho$ takes the following form (denoting $L = \lambda_{\max}(\mathbf{A}^T\mathbf{A})$).

AD-LPMM (Algorithm 3):

$$\begin{aligned}
\mathbf{x}^{k+1} &= \mathbf{x}^k - \frac{1}{L}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right), \\
\mathbf{z}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left[\left(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b} + \frac{1}{\rho}\mathbf{y}^k\right)\right] + \mathbf{b}, \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{z}^{k+1}). \quad \blacksquare
\end{aligned}$$

Example 15.7 (basis pursuit). Consider the problem

$$\begin{aligned}
\min \quad & \|\mathbf{x}\|_1 \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b},
\end{aligned} \tag{15.28}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Problem (15.28) fits the composite model (15.23) with $f_1(\mathbf{x}) = \|\mathbf{x}\|_1$ and $f_2 = \delta_{\{\mathbf{b}\}}$. Let $\rho > 0$. For any $\gamma > 0$, $\text{prox}_{\gamma f_1} = \mathcal{T}_\gamma$ (by Example 6.8) and $\text{prox}_{\gamma_2 f_2} \equiv \mathbf{b}$. Algorithm 1 is actually not particularly implementable since its first update step is

$$\mathbf{x}^{k+1} \in \text{argmin}_{\mathbf{x}\in\mathbb{R}^n}\left\{\|\mathbf{x}\|_1 + \frac{\rho}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{z}^k + \frac{1}{\rho}\mathbf{y}^k\right\|^2\right\},$$

which does not seem to be simpler to solve than the original problem (15.28).

Algorithm 2 takes the following form (assuming that $\mathbf{z}^0 = \mathbf{b}$).

ADMM, version 2 (Algorithm 2):

$$\begin{aligned}
\mathbf{x}^{k+1} &= (\mathbf{I} + \mathbf{A}^T\mathbf{A})^{-1}\left(\mathbf{A}^T\left[\mathbf{b} - \frac{1}{\rho}\mathbf{y}_1^k\right] + \mathbf{w}^k - \frac{1}{\rho}\mathbf{y}_2^k\right), \\
\mathbf{w}^{k+1} &= \mathcal{T}_{\frac{1}{\rho}}\left(\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_2^k\right), \\
\mathbf{y}_1^{k+1} &= \mathbf{y}_1^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}), \\
\mathbf{y}_2^{k+1} &= \mathbf{y}_2^k + \rho(\mathbf{x}^{k+1} - \mathbf{w}^{k+1}).
\end{aligned}$$

Finally, assuming that $\mathbf{z}^0 = \mathbf{b}$, Algorithm 3 with $\alpha = \lambda_{\max}(\mathbf{A}^T\mathbf{A})\rho$ and $\beta = \rho$ reduces to the following simple update steps (denoting $L = \lambda_{\max}(\mathbf{A}^T\mathbf{A})$).

AD-LPMM (Algorithm 3):

$$\begin{aligned}
\mathbf{x}^{k+1} &= \mathcal{T}_{\frac{1}{\rho L}}\left[\mathbf{x}^k - \frac{1}{L}\mathbf{A}^T\left(\mathbf{A}\mathbf{x}^k - \mathbf{b} + \frac{1}{\rho}\mathbf{y}^k\right)\right], \\
\mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}). \quad \blacksquare
\end{aligned}$$

Example 15.8 (minimizing $\sum_{i=1}^p g_i(\mathbf{A}_i\mathbf{x})$). Consider now the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \sum_{i=1}^p g_i(\mathbf{A}_i\mathbf{x}), \tag{15.29}$$

where $g_1, g_2, \ldots, g_p$ are proper closed and convex functions and $\mathbf{A}_i \in \mathbb{R}^{m_i \times n}$ for all $i = 1, 2, \ldots, p$. Problem (15.29) fits the composite model (15.23) with

- $f_1 \equiv 0$;
- $f_2(\mathbf{y}) = \sum_{i=1}^p g_i(\mathbf{y}_i)$, where we assume that $\mathbf{y} \in \mathbb{R}^{m_1+m_2+\cdots+m_p}$ is of the form $\mathbf{y} = (\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_p^T)^T$, where $\mathbf{y}_i \in \mathbb{R}^{m_i}$;
- the matrix $\mathbf{A} \in \mathbb{R}^{(m_1+m_2+\ldots+m_p)\times n}$ given by $\mathbf{A} = (\mathbf{A}_1^T, \mathbf{A}_2^T, \ldots, \mathbf{A}_p^T)^T$.

For any $\gamma > 0$, $\mathrm{prox}_{\gamma f_1}(\mathbf{x}) = \mathbf{x}$ and $\mathrm{prox}_{\gamma f_2}(\mathbf{y})_i = \mathrm{prox}_{\gamma g_i}(\mathbf{y}_i)$, $i = 1, 2, \ldots, p$ (by Theorem 6.6). The general update step of the first version of ADMM (Algorithm 1) has the form

$$\mathbf{x}^{k+1} \in \mathrm{argmin}_{\mathbf{x}\in\mathbb{R}^n} \sum_{i=1}^p \left\| \mathbf{A}_i\mathbf{x} - \mathbf{z}_i^k + \frac{1}{\rho}\mathbf{y}_i^k \right\|^2, \tag{15.30}$$

$$\mathbf{z}_i^{k+1} = \mathrm{prox}_{\frac{1}{\rho}g_i}\left(\mathbf{A}_i\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_i^k\right), \quad i = 1, 2, \ldots, p,$$

$$\mathbf{y}_i^{k+1} = \mathbf{y}_i^k + \rho(\mathbf{A}_i\mathbf{x}^{k+1} - \mathbf{z}_i^{k+1}), \quad i = 1, 2, \ldots, p.$$

In the case where $\mathbf{A}$ has full column rank, step (15.30) can be written more explicitly, leading to the following representation.

ADMM, version 1 (Algorithm 1):

$$\mathbf{x}^{k+1} = \left(\sum_{i=1}^p \mathbf{A}_i^T\mathbf{A}_i\right)^{-1} \sum_{i=1}^p \mathbf{A}_i^T\left(\mathbf{z}_i^k - \frac{1}{\rho}\mathbf{y}_i^k\right),$$

$$\mathbf{z}_i^{k+1} = \mathrm{prox}_{\frac{1}{\rho}g_i}\left(\mathbf{A}_i\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_i^k\right), \quad i = 1, 2, \ldots, p,$$

$$\mathbf{y}_i^{k+1} = \mathbf{y}_i^k + \rho(\mathbf{A}_i\mathbf{x}^{k+1} - \mathbf{z}_i^{k+1}), \quad i = 1, 2, \ldots, p.$$

The second version of ADMM (Algorithm 2) is not simpler than the first version, and we will therefore not write it explicitly. AD-LPMM (Algorithm 3) invoked with the constants $\alpha = \lambda_{\max}(\sum_{i=1}^p \mathbf{A}_i^T\mathbf{A}_i)\rho$ and $\beta = \rho$ reads as follows (denoting $L = \lambda_{\max}(\sum_{i=1}^p \mathbf{A}_i^T\mathbf{A}_i)$).

AD-LPMM (Algorithm 3):

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \frac{1}{L}\sum_{i=1}^p \mathbf{A}_i^T\left(\mathbf{A}_i\mathbf{x}^k - \mathbf{z}_i^k + \frac{1}{\rho}\mathbf{y}_i^k\right),$$

$$\mathbf{z}_i^{k+1} = \mathrm{prox}_{\frac{1}{\rho}g_i}\left(\mathbf{A}_i\mathbf{x}^{k+1} + \frac{1}{\rho}\mathbf{y}_i^k\right), \quad i = 1, 2, \ldots, p,$$

$$\mathbf{y}_i^{k+1} = \mathbf{y}_i^k + \rho(\mathbf{A}_i\mathbf{x}^{k+1} - \mathbf{z}_i^{k+1}), \quad i = 1, 2, \ldots, p. \quad \blacksquare$$