Chapter 9

Error bounds and the restarting scheme for the primal-dual iteration

Theorems 8.2–8.4 reduce the problem of proving the main results (Theorems 3.4–3.12) to two tasks. The first involves bounding the error in the objective function, i.e., the term

$$\mathcal{G}(\tilde{c}) - \mathcal{G}(\mathcal{P}_h(c_\Lambda)),$$

where $\tilde{c}$ is either an exact minimizer or an approximate minimizer obtained via the primal dual iteration. The second involves the various approximation error terms depending on f and its polynomial coefficients.

In this chapter, we address the first task. We first provide an error bound for the (unrestarted) primal-dual iteration when applied to Hilbert-valued weighted SR-LASSO problem (7.2), and then use this to derive the specific restart scheme.

9.1 Error bounds for the primal-dual iteration

We now return to the general setting of the primal-dual iteration, where it is applied to the problem (4.9) and takes the form (4.12). The following result from [31, Theorem 5.1] establishes an important error bound for the Lagrangian difference.

Theorem 9.1. *Let $\tau, \sigma > 0$, initial points $(x^{(0)}, \xi^{(0)}) \in \mathcal{X} \times \mathcal{Y}$ and a bounded linear operator $A \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$, be such that $\|A\|^2_{\mathcal{B}(\mathcal{X},\mathcal{Y})} \leq (\tau\sigma)^{-1}$. Consider the sequence $\{(x^{(n)}, \xi^{(n)})\}_{n=1}^{\infty}$ generated by the primal-dual iteration (4.12). Then, for any $(x, \xi) \in \mathcal{X} \times \mathcal{Y}$,*

$$\mathcal{L}(\bar{x}^{(n)}, \xi) - \mathcal{L}(x, \bar{\xi}^{(n)}) \leq \frac{\tau^{-1}\|x - x^{(0)}\|^2_{2;\mathcal{V}} + \sigma^{-1}\|\xi - \xi^{(0)}\|^2_{2;\mathcal{V}}}{n}, \tag{9.1}$$

where

$$\bar{x}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} x^{(k)} \quad \text{and} \quad \bar{\xi}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} \xi^{(k)},$$

are the ergodic sequences and $\mathcal{L}$ is the Lagrangian (4.11).

The following lemma shows a decay rate of $1/n$ on the objective function in the case of the primal-dual iteration when applied to the problem (7.2). It is an extension of [13, Lemma 8.6] to the weighted and Hilbert-valued setting.

Lemma 9.2. *Let $A \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ and $\tau, \sigma > 0$ be such that $\|A\|^2_{\mathcal{B}(\mathcal{V}^N,\mathcal{V}^m)} \leq (\tau\sigma)^{-1}$. Consider the sequence $\{(\boldsymbol{x}^{(n)}, \boldsymbol{\xi}^{(n)})\}_{n=1}^{\infty}$ generated by the primal-dual iteration in (4.12) applied to (7.2) with $\boldsymbol{x}^{(0)} \in \mathcal{V}^N$ and $\boldsymbol{\xi}^{(0)} = \boldsymbol{0} \in \mathcal{V}^m$. Then, for any $\boldsymbol{x} \in \mathcal{V}^N$,*

$$\mathcal{G}(\bar{\boldsymbol{x}}^{(n)}) - \mathcal{G}(\boldsymbol{x}) \leq \frac{\tau^{-1}\|\boldsymbol{x} - \boldsymbol{x}_0\|^2_{2;\mathcal{V}} + \sigma^{-1}}{n}, \quad \bar{\boldsymbol{x}}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{x}^{(k)}. \tag{9.2}$$

Proof. Using (4.11) and (4.13), the left-hand side of (9.1) is given by

$$\begin{aligned}\mathcal{T}_n(\boldsymbol{x}, \boldsymbol{\xi}) := &\left(\lambda\|\bar{\boldsymbol{x}}^{(n)}\|_{1,\boldsymbol{w};\mathcal{V}} + \operatorname{Re}\langle A\bar{\boldsymbol{x}}^{(n)} - \boldsymbol{b}, \boldsymbol{\xi}\rangle_{2;\mathcal{V}} + \delta_B(\boldsymbol{\xi})\right)\\ &- \left(\lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} + \operatorname{Re}\langle A\boldsymbol{x} - \boldsymbol{b}, \bar{\boldsymbol{\xi}}^{(n)}\rangle_{2;\mathcal{V}} + \delta_B(\bar{\boldsymbol{\xi}}^{(n)})\right),\end{aligned}$$

where B is the unit ball in $\mathcal{V}^m$. Observe that the term $\boldsymbol{\xi}^{(n)}$ produced by this iteration satisfies $\|\boldsymbol{\xi}^{(n)}\|_{2;\mathcal{V}} \leq 1$. This follows from the observation shown in Section 4.4 that the proximal mapping

$$\operatorname{prox}_{\sigma h^*}(\boldsymbol{\xi}) = \operatorname{proj}_B(\boldsymbol{\xi} - \sigma\boldsymbol{b})$$

involves the projection onto the unit ball B. Hence, the ergodic sequence $\bar{\boldsymbol{\xi}}^{(n)}$ satisfies $\|\bar{\boldsymbol{\xi}}^{(n)}\|_{2;\mathcal{V}} \leq 1$ as well. Suppose now that $A\boldsymbol{x}^{(n)} - \boldsymbol{b} \neq \boldsymbol{0}$ and set

$$\boldsymbol{\xi} = \frac{A\boldsymbol{x}^{(n)} - \boldsymbol{b}}{\|A\boldsymbol{x}^{(n)} - \boldsymbol{b}\|_{2;\mathcal{V}}}.$$

Then $\delta_B(\boldsymbol{\xi}) = \delta_B(\bar{\boldsymbol{\xi}}^{(n)}) = 1$ and therefore

$$\begin{aligned}\mathcal{T}_n(\boldsymbol{x}, \boldsymbol{\xi}) &= \left(\lambda\|\bar{\boldsymbol{x}}^{(n)}\|_{1,\boldsymbol{w};\mathcal{V}} + \|A\bar{\boldsymbol{x}}^{(n)} - \boldsymbol{b}\|_{2;\mathcal{V}}\right) - \left(\lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} + \operatorname{Re}\langle A\boldsymbol{x} - \boldsymbol{b}, \bar{\boldsymbol{\xi}}^{(n)}\rangle_{2;\mathcal{V}}\right)\\ &\geq \left(\lambda\|\bar{\boldsymbol{x}}^{(n)}\|_{1,\boldsymbol{w};\mathcal{V}} + \|A\bar{\boldsymbol{x}}^{(n)} - \boldsymbol{b}\|_{2;\mathcal{V}}\right) - \left(\lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} + \|A\boldsymbol{x} - \boldsymbol{b}\|_{2;\mathcal{V}}\right).\end{aligned}$$

Clearly, the same bound also holds in the case $A\boldsymbol{x}^{(n)} - \boldsymbol{b} = \boldsymbol{0}$ where $\boldsymbol{\xi}$ is an arbitrary unit vector. Hence Theorem 9.1 and the fact that

$$\|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|_{2;\mathcal{V}} = \|\boldsymbol{\xi}\|_{2;\mathcal{V}} = 1$$

gives the result. ∎

9.2 The restarting scheme

For convenience, we now introduce new and slightly modify some existing notation. First, we redefine the objective function $\mathcal{G}$ of the Hilbert-valued weighted SR-LASSO problem (7.2) to make the dependence on the term $\boldsymbol{b}$ explicit: namely, we set

$$\mathcal{G}(\boldsymbol{x}, \boldsymbol{b}) = \lambda\|\boldsymbol{x}\|_{1,\boldsymbol{w};\mathcal{V}} + \|A\boldsymbol{x} - \boldsymbol{b}\|_{2;\mathcal{V}}, \quad \boldsymbol{x} \in \mathcal{V}^N, \ \boldsymbol{b} \in \mathcal{V}^m.$$

We then let

$$\mathcal{E}(z, x, b) = \mathcal{G}(z, b) - \mathcal{G}(x, b), \quad x, z \in \mathcal{V}^N,\ b \in \mathcal{V}^m.$$

Now consider the ergodic sequence $\bar{x}^{(n)}$ produced by n iterations of the primal-dual iteration (4.12) applied to (7.2) with parameters $\tau, \sigma > 0$, $x_0 \in \mathcal{V}^N$ and $\xi_0 = \mathbf{0} \in \mathcal{V}^m$. For reasons that will become clear in a moment, we now make the dependence on the vector b in (7.2), the number of iterations $\bar{x}^{(n)}$ and the initial vector x_0 explicit, by defining

$$\mathcal{P}(x_0, b, n) = \bar{x}^{(n)}.$$

With this in hand, we conclude this discussion by noting the following two scaling properties:

$$\mathcal{G}(ax, b) = a\mathcal{G}(x, b/a), \quad \mathcal{E}(az, x, b) = a\mathcal{E}(z, x/a, b/a). \tag{9.3}$$

These hold for any $a > 0$ and for any $x, z \in \mathcal{V}^N$ and $b \in \mathcal{V}^m$.

Lemma 9.3. *Suppose that $A \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ has the weighted rNSP over $\mathcal{V}$ of order (k, w) with constants $0 < \rho < 1$ and $\gamma > 0$. Consider the Hilbert-valued weighted SR-LASSO problem (7.2) with parameter $\lambda = c/\sqrt{k}$, where*

$$0 < c \leq \frac{(1+\rho)^2}{(3+\rho)\gamma}.$$

Let $\mathcal{E}$ and $\mathcal{P}$ be as defined above, τ, σ satisfy $\|A\|^2_{\mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)} \leq (\tau\sigma)^{-1}$ and $x, x_0 \in \mathcal{V}^N$, $b \in \mathcal{V}^m$, $a > 0$. Then

$$\mathcal{E}(a\mathcal{P}(x_0/a, b/a, n), x, b) \leq \frac{C^2}{a\tau n}(\mathcal{E}(x_0, x, b) + \xi)^2 + \frac{a}{\sigma n},$$

where

$$C = 2\max\{C_1'/c, C_2'\}, \tag{9.4}$$

C_1', C_2' are as in Lemma 7.3 and

$$\xi = \xi(x, b) = \frac{\sigma_k(x)_{1,w;\mathcal{V}}}{\sqrt{k}} + \|Ax - b\|_{2;\mathcal{V}}. \tag{9.5}$$

Proof. The scaling property (9.3) and Lemma 9.2 give

$$\begin{aligned}\mathcal{E}(a\mathcal{P}(x_0/a, b/a, n), x, b) &= a\mathcal{E}(\mathcal{P}(x_0/a, b/a, n), x/a, b/a)\\ &\leq a\left(\frac{\tau^{-1}\|x/a - x_0/a\|^2_{2;\mathcal{V}} + \sigma^{-1}}{n}\right)\\ &= \frac{\|x - x_0\|^2_{2;\mathcal{V}}}{a\tau n} + \frac{a}{\sigma n}.\end{aligned}$$

Now consider the term $\|\boldsymbol{x} - \boldsymbol{x}_0\|_{2;\mathcal{V}}$. Since $\boldsymbol{A}$ has the weighted rNSP and λ satisfies (7.3), we may use Lemma 7.4 to get

$$\begin{aligned}\|\boldsymbol{x}-\boldsymbol{x}_0\|_{2;\mathcal{V}} &\leq \frac{C_1'}{\sqrt{k}}\left(2\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}} + \frac{\mathcal{G}(\boldsymbol{x}_0,\boldsymbol{b})-\mathcal{G}(\boldsymbol{x},\boldsymbol{b})}{\lambda}\right) + \left(\frac{C_1'}{\sqrt{k}\lambda}+C_2'\right)\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_{2;\mathcal{V}} \\ &= \frac{C_1'}{\sqrt{k}\lambda}\mathcal{E}(\boldsymbol{x}_0,\boldsymbol{x},\boldsymbol{b}) + 2C_1'\frac{\sigma_k(\boldsymbol{x})_{1,\boldsymbol{w};\mathcal{V}}}{\sqrt{k}} + \left(\frac{C_1'}{\sqrt{k}\lambda} + C_2'\right)\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_{2;\mathcal{V}} \\ &\leq 2\max\{C_1'/c, C_2'\}(\mathcal{E}(\boldsymbol{x}_0,\boldsymbol{x},\boldsymbol{b}) + \xi).\end{aligned}$$

Substituting this into the previous expression now gives the result. ∎

This lemma gives the rationale behind the restarted scheme. It says the error in the objective function of the scaled output $a\mathcal{P}(\boldsymbol{x}_0/a, \boldsymbol{b}/a, n)$ of the primal-dual iteration with initial value $\boldsymbol{x}_0$ can be bounded in terms of the error in the objective function at the initial value, plus terms depending on the scaling parameter a, the number of iterations n and the compressed sensing error term ξ. By choosing these parameters suitably and iterating this procedure, we obtain the restarting scheme. We summarize this in the following theorem.

Theorem 9.4 (Restarting scheme). *Suppose that $\boldsymbol{A} \in \mathcal{B}(\mathcal{V}^N, \mathcal{V}^m)$ has the weighted rNSP over $\mathcal{V}$ of order $(k, \boldsymbol{w})$ with constants $0 < \rho < 1$ and $\gamma > 0$. Consider the Hilbert-valued weighted SR-LASSO problem (7.2) with parameter $\lambda = c/\sqrt{k}$, where $0 < c \leq \frac{(1+\rho)^2}{(3+\rho)\gamma}$. Let $\boldsymbol{x} \in \mathcal{V}^N$, $\boldsymbol{b} \in \mathcal{V}^m$, $\zeta' \geq \xi$, where ξ is as in (9.5), $0 < r < 1$ and define the sequence*

$$\varepsilon_0 = \|\boldsymbol{b}\|_{2;\mathcal{V}}, \quad \varepsilon_{k+1} = r(\varepsilon_k + \zeta'),\ k = 0, 1, 2, \ldots.$$

Let $\mathcal{E}$ and $\mathcal{P}$ be as defined above, τ, σ satisfy $\|\boldsymbol{A}\|^2_{\mathcal{B}(\mathcal{V}^N,\mathcal{V}^m)} \leq (\tau\sigma)^{-1}$ and set

$$n = \left\lceil \frac{2C}{r\sqrt{\sigma\tau}} \right\rceil, \quad a_k = \frac{1}{2}\sigma\varepsilon_{k+1}n,\ k = 0, 1, 2, \ldots,$$

where C is as in (9.4). Then the iteration $\tilde{\boldsymbol{x}}^{(0)}, \tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{x}}^{(2)}, \ldots$, defined by

$$\tilde{\boldsymbol{x}}^{(0)} = \boldsymbol{0}, \quad \tilde{\boldsymbol{x}}^{(k+1)} = a_k\mathcal{P}(\tilde{\boldsymbol{x}}^{(k)}/a_k, \boldsymbol{b}/a_k, n),\ k = 0, 1, 2, \ldots,$$

satisfies

$$\mathcal{E}(\boldsymbol{x}_k^\star, \boldsymbol{x}, \boldsymbol{b}) \leq \varepsilon_k \leq r^k\|\boldsymbol{b}\|_{2;\mathcal{V}} + \frac{r}{1-r}\zeta', \quad k = 0, 1, 2, \ldots.$$

Proof. We use induction on k. Suppose first that $k = 0$. Then, by definition,

$$\mathcal{E}(\tilde{\boldsymbol{x}}^{(k)}, \boldsymbol{x}, \boldsymbol{b}) = \mathcal{E}(\boldsymbol{0}, \boldsymbol{x}, \boldsymbol{b}) \leq \mathcal{G}(\boldsymbol{0}, \boldsymbol{b}) = \|\boldsymbol{b}\|_{2;\mathcal{V}} = \varepsilon_0.$$

Now suppose that the result holds for k. The previous lemma gives

$$\begin{aligned}\mathcal{E}(\tilde{\boldsymbol{x}}^{(k+1)}, \boldsymbol{x}, \boldsymbol{b}) &= \mathcal{E}(a_k \mathcal{P}(\tilde{\boldsymbol{x}}^{(k)}/a_k, \boldsymbol{b}/a_k, n), \boldsymbol{x}, \boldsymbol{b}) \\ &\leq \frac{C^2}{a_k \tau n}(\mathcal{E}(\tilde{\boldsymbol{x}}^{(k)}, \boldsymbol{x}, \boldsymbol{b}) + \zeta)^2 + \frac{a_k}{\sigma n} \\ &\leq \frac{C^2}{a_k \tau n}(\varepsilon_k + \zeta)^2 + \frac{a_k}{\sigma n}.\end{aligned}$$

We now substitute the values of n and a_k to obtain

$$\mathcal{E}(\tilde{\boldsymbol{x}}^{(k+1)}, \boldsymbol{x}, \boldsymbol{b}) = \frac{2C^2(\varepsilon_k + \zeta)}{r\sigma\tau n^2} + \frac{1}{2}r(\varepsilon_k + \zeta) \leq \frac{1}{2}r(\varepsilon_k + \zeta) + \frac{1}{2}r(\varepsilon_k + \zeta) = \varepsilon_{k+1}.$$

This completes the proof. ■

This theorem states that the restarted primal-dual iteration $\tilde{\boldsymbol{x}}^{(0)}, \tilde{\boldsymbol{x}}^{(1)}, \tilde{\boldsymbol{x}}^{(2)}, \ldots$ yields an objective function error $\mathcal{E}(\tilde{\boldsymbol{x}}^{(k)}, \boldsymbol{x}, \boldsymbol{b})$ that converges exponentially fast in the number of restarts k. Further, each (inner) primal-dual iteration involves a number of steps n that depends on the parameters C, r, σ and τ. In other words, n is a constant independent of k. Hence, the restarted scheme converges exponentially fast in the total number of primal-dual iterations as well.

As discussed in Section 5.1.1, it is typical to use this theorem to optimize the choice of r. Recall that this leads to the explicit choice $r = \mathrm{e}^{-1}$. We use this value in our algorithms – see Table 4.3.