MATHEMATICS OF COMPUTATION
Volume 66, Number 217, January 1997, Pages S1–S11
S 0025-5718(97)00777-1

Supplement to

A GLOBALLY CONVERGENT LAGRANGIAN BARRIER ALGORITHM FOR OPTIMIZATION WITH GENERAL INEQUALITY CONSTRAINTS AND SIMPLE BOUNDS

A. R. CONN, NICK GOULD, AND PH. L. TOINT

Appendix

A. Simplified statement of the Algorithm

In order to simplify the proofs given in the appendix, we give them in the case of a particular instance of Algorithm 3.1. This simplified algorithm corresponds to choosing parameters $\eta_s = \omega_s = \alpha_\omega = \beta_\omega = 1$, $\alpha_\eta = \frac{5}{6}, \beta_\eta = \frac{1}{6}$ and $\alpha_\lambda = \frac{1}{2}$.

For proofs that correspond to the more general statements the interested reader is referred to the technical report, Conn et al. [12]. However, the choice of parameters in practice appears to be significantly more critical than for augmented Lagrangian approaches. Thus the choice given here, which is just for convenience of exposition, should not be taken as an indication of suitable values.

Algorithm A.1 [Outer Iteration Algorithm]

step 0 : [Initialization] Choose the strictly positive constants

$$\omega_s, \eta_s, \mu_0 < 1, \tau < 1, \text{ and define } \omega_0 = \omega_s\mu_0 \text{ and } \eta_0 = \eta_s\mu_0^{\frac{5}{6}}. \tag{A.1}$$

An initial estimate of the solution, $x_{-1} \in B$, and vector of positive Lagrange multiplier estimates, λ_0, for which $c_i(x_{-1}) + \mu_0\sqrt{\lambda_{0,i}} > 0$ are specified. Set $k = 0$.

In addition set

$$\omega_* \ll 1 \text{ and } \eta_* \ll 1. \tag{A.2}$$

step 1 : [Inner iteration] Compute shifts

$$s_{k,i} = \mu_k\sqrt{\lambda_{k,i}}, \tag{A.3}$$

for $i = 1, ..., m$. Find $x_k \in B$ such that

$$\|P(x_k, \nabla\Psi_k)\| \le \omega_k \tag{A.4}$$

and

$$c_i(x_k) + s_{k,i} > 0 \quad \text{for} \quad i = 1, ..., m. \tag{A.5}$$

step 2 : [Test for convergence] If

$$\|P(x_k, \nabla\Psi_k)\| \le \omega_* \text{ and } \|[c_i(x_k)\bar{\lambda}_{k,i}]_{i=1}^m\| \le \eta_*, \tag{A.6}$$

stop. If

$$\left\| \left[\frac{c_i(x_k)\bar{\lambda}_{k,i}}{\sqrt{\lambda_{k,i}}} \right]_{i=1}^m \right\| \le \eta_k, \tag{A.7}$$

execute step 3. Otherwise, execute step 4.

step 3 : [Update Lagrange multiplier estimates] Set

$$\text{(A.8)}\qquad \begin{aligned} \lambda_{k+1} &= \bar{\lambda}_k, \\ \mu_{k+1} &= \mu_k, \\ \omega_{k+1} &= \mu_{k+1}\omega_k, \\ \eta_{k+1} &= \mu_{k+1}^{\frac{1}{6}}\eta_k. \end{aligned}$$

Increase k by one and go to step 1.

step 4 : [Reduce the penalty parameter] Set

$$\text{(A.9)}\qquad \begin{aligned} \lambda_{k+1} &= \lambda_k, \\ \mu_{k+1} &= \tau\mu_k, \\ \omega_{k+1} &= \omega_s\mu_{k+1}, \\ \eta_{k+1} &= \eta_s\mu_{k+1}^{\frac{5}{6}}. \end{aligned}$$

Increase k by one and go to step 1.

end of Algorithm A.1

B. Details of proofs from §4

B.1. An auxiliary lemma. We require the following lemma in the proof of global convergence of our algorithm. The lemma is the analog of Conn et al. [11, Lemma 4.1]. In essence, the result shows that the Lagrange multiplier estimates generated by the algorithm cannot behave too badly.

Lemma B.1 *Suppose that μ_k converges to zero as k increases when Algorithm* A.1 *is executed. Then the product $\mu_k(\lambda_{k,i})^{\frac{3}{2}}$ converges to zero for each $1 \le i \le m$.*

Proof. If μ_k converges to zero, step 4 of the algorithm must be executed infinitely often. Let $K = \{k_0, k_1, k_2, ...\}$ be the set of the indices of the iterations in which step 4 of the algorithm is executed and for which

$$\text{(B.1)}\qquad \mu_k \le (\tfrac{1}{2})^6.$$

We consider how the ith Lagrange multiplier estimate changes between two successive iterations indexed in the set K. First note that $\lambda_{k_p+1,i} = \lambda_{k_p,i}$. At iteration $k_p + j$, for $k_p + 1 < k_p + j \le k_{p+1}$, we have

$$\text{(B.2)}\qquad \lambda_{k_p+j,i} = \lambda_{k_p+j-1,i} - \left(\frac{c_{k_p+j-1,i}\lambda_{k_p+j,i}}{\sqrt{\lambda_{k_p+j-1,i}}}\right)\frac{1}{\mu_{k_p+j-1}},$$

from (2.5), (A.3) and (A.8) and

$$\text{(B.3)}\qquad \mu_{k_{p+1}} = \mu_{k_p+j} = \mu_{k_p+1} = \tau\mu_{k_p}.$$

Hence summing (B.2) and using the fact that $\lambda_{k_p+1,i} = \lambda_{k_p,i}$, we get

$$\text{(B.4)}\qquad \lambda_{k_p+j,i} = \lambda_{k_p,i} - \sum_{l=1}^{j-1}\left(\frac{c_{k_p+l,i}\lambda_{k_p+l+1,i}}{\sqrt{\lambda_{k_p+l,i}}}\right)\frac{1}{\mu_{k_p+1}}$$

where the summation in (B.4) is null if $j = 1$.

Now suppose that $j > 1$. Then for the set of iterations $k_p+l, 1 \le l < j$, step 3 of the algorithm must have been executed and hence, from (A.5), (B.3) and the recursive definition of η_k, we must also have

$$\text{(B.5)}\qquad \left\|\left[\frac{c_{k_p+l,i}\lambda_{k_p+l+1,i}}{\sqrt{\lambda_{k_p+l,i}}}\right]_{i=1}^m\right\| \le \eta_s\mu_{k_p+1}^{\frac{5}{6}+\frac{1}{6}(l-1)}.$$

Combining equations (B.1) to (B.5), we obtain the bound

$$\text{(B.6)}\qquad \begin{aligned} \|\lambda_{k_p+j}\| &\le \|\lambda_{k_p}\| + \textstyle\sum_{l=1}^{j-1}\left\|\left[\frac{c_{k_p+l,i}\lambda_{k_p+l+1,i}}{\sqrt{\lambda_{k_p+l,i}}}\right]_{i=1}^m\right\| \cdot \frac{1}{\mu_{k_p+1}} \\ &\le \|\lambda_{k_p}\| + 2\eta_s/\mu_{k_p+1}^{\frac{1}{6}}. \end{aligned}$$

Thus, multiplying (B.6) by $\mu_{k_p+j}^{\frac{2}{3}}$ and using (B.3), we obtain that

$$\text{(B.7)}\qquad \mu_{k_p+j}^{\frac{2}{3}}\|\lambda_{k_p+j}\| \le (\tau\mu_{k_p})^{\frac{2}{3}}\|\lambda_{k_p}\| + 2\eta_s\sqrt{\tau\mu_{k_p}}.$$

Equation (B.7) is also satisfied when $j = 1$, as equations (A.8) and (B.3) give

$$\text{(B.8)}\qquad \mu_{k_p+j}^{\frac{2}{3}}\|\lambda_{k_p+j}\| = (\tau\mu_{k_i})^{\frac{2}{3}}\|\lambda_{k_p}\|.$$

Hence from (B.7),

$$\text{(B.9)}\qquad \mu_{k_{p+1}}^{\frac{2}{3}}\|\lambda_{k_{p+1}}\| \le (\tau\mu_{k_p})^{\frac{2}{3}}\|\lambda_{k_p}\| + 2\eta_s\sqrt{\tau\mu_{k_p}}.$$

We now show that (B.9) implies that $\mu_{k_p}^{\frac{2}{3}}\|\lambda_{k_p}\|$ converges to zero as k increases. For, if we define

$$\text{(B.10)}\qquad \alpha_p \stackrel{\text{def}}{=} \mu_{k_p}^{\frac{2}{3}}\|\lambda_{k_p}\| \text{ and } \beta_p \stackrel{\text{def}}{=} 2\eta_s\sqrt{\mu_{k_p}},$$

equations (B.3), (B.9) and (B.10) give that

$$\text{(B.11)}\qquad \alpha_{p+1} \le \tau^{\frac{2}{3}}\alpha_p + \sqrt{\tau}\beta_p \text{ and } \beta_{p+1} = \sqrt{\tau}\beta_p$$

and hence that

$$\text{(B.12)}\qquad 0 \le \alpha_p \le \tau^{\frac{2}{3}p}\alpha_0 + \tau^{\frac{p}{2}}\sum_{l=0}^{p-1}\tau^{\frac{l}{6}}\beta_0.$$

It now follows that

$$\text{(B.13)}\qquad 0 \le \alpha_p \le \tau^{\frac{2}{3}p}\alpha_0 + \frac{\tau^{\frac{p}{2}}}{1-\tau^{\frac{l}{6}}}\beta_0.$$

But both α_0 and β_0 are finite. Thus, as p increases, α_p converges to zero; the second part of equation (B.11) implies that β_p converges to zero. Therefore, as the right-hand side of (B.7) converges to zero, so does $\mu_k^{\frac{2}{3}}\|\lambda_k\|$ for all k. The truth of the lemma is finally established by raising $\mu_k^{\frac{2}{3}}\|\lambda_{k,i}\|$ to the power $\frac{3}{2}$. □

We note that Lemma B.1 may be proved under much weaker conditions on the sequence $\{\eta_k\}$ than those imposed in Algorithm A.1. All that is needed is that, in the proof just given,

$$\sum_{l=1}^{j-1}\left\|\left[\frac{c_{k_p+l,i}\lambda_{k_p+l+1,i}}{\sqrt{\lambda_{k_p+l,i}}}\right]_{i=1}^{m}\right\|$$

in (B.6) should be bounded by some multiple of a positive power of $\mu_{k_{p+1}}$.

B.2. Proof of Theorem 4.2. In order to prove (i), (ii) and (iii), we consider each constraint in turn and distinguish two cases:

1. constraints for which $c_{*,i} \neq 0$; and
2. constraints for which $c_{*,i} = 0$.

For the first of these cases, we need to consider the possibility that

a. the penalty parameter μ_k is bounded away from zero; and

b. the penalty parameter μ_k converges to zero.

Case 1a. As μ_k is bounded away from zero, test (A.7) must be satisfied for all k sufficiently large and hence $|c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_k}|$ converges to zero. Thus, as $\{c_{k,i}\}$ converges to $c_{*,i} \neq 0$, for $k \in \mathcal{K}$, $\bar{\lambda}_{k,i}/\sqrt{\lambda_k}$ converges to zero. Hence, using (2.3) and (A.3), we have that

(B.14) $$\frac{\bar{\lambda}_{k,i}}{\sqrt{\lambda_{k,i}}} \equiv \frac{\mu\lambda_{k,i}}{c_{k,i}+\mu\sqrt{\lambda_{k,i}}} = \sqrt{\lambda_{k,i}}\frac{\mu\sqrt{\lambda_{k,i}}}{c_{k,i}+\mu\sqrt{\lambda_{k,i}}} \to 0.$$

We aim to show that $\bar{\lambda}_{k,i}$ converges to zero and that $c_{*,i} > 0$.

Suppose first that $\lambda_{k,i}$ does not converge to zero. It follows directly from (2.5) and (A.3) that

(B.15) $$c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}} = \mu_k(\lambda_{k,i} - \bar{\lambda}_{k,i}).$$

Then, as the left-hand side of (B.15) converges to zero and μ_k and $\lambda_{k,i}$ are bounded away from zero, we deduce that

(B.16) $$\bar{\lambda}_{k,i} = \lambda_{k,i}(1-\epsilon_{k,i}),$$

for some $\{\epsilon_{k,i}\}$, $k \in \mathcal{K}$, converging to zero. But then, by definition (2.3),

(B.17) $$\frac{\mu\sqrt{\lambda_{k,i}}}{c_{k,i}+\mu\sqrt{\lambda_{k,i}}} = 1-\epsilon_{k,i}.$$

However, as $\lambda_{k,i}$ is bounded away from zero, (B.17) contradicts (B.14). Thus $\lambda_{k,i}$ converges to zero, for $k \in \mathcal{K}$.

It now follows that, as $\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}}$ converges to zero, so does $\bar{\lambda}_{k,i}$. It also follows from (A.5) that $c_{k,i}+\mu_k\sqrt{\lambda_{k,i}} > 0$. As μ_k is bounded and $\lambda_{k,i}$ converges to zero, we have that $c_{*,i} \geq 0$. But as $c_{*,i} \neq 0$, we conclude that $c_{*,i} > 0$, $\bar{\lambda}_{k,i}$ converges to $\lambda_{*,i} = 0$, for $k \in \mathcal{K}$, and $c_{*,i}\lambda_{*,i} = 0$.

Case 1b. As μ_k converges to zero, Lemma B.1 shows that $\mu_k(\lambda_{k,i})^{\frac{3}{2}}$ and hence $\mu_k\lambda_{k,i}$ and $\mu_k\sqrt{\lambda_{k,i}}$ converges to zero. It follows immediately that the numerator of (2.3) converges to zero while the denominator converges to $c_{*,i}$ and hence that $\bar{\lambda}_{k,i}$ converges to zero for $k \in \mathcal{K}$. Furthermore, it follows from (A.5) that $c_{k,i} + \mu_k\sqrt{\lambda_{k,i}} > 0$: as $\mu_k\sqrt{\lambda_{k,i}}$ converges to zero, we have that $c_{*,i} \geq 0$. But as $c_{*,i}$ is, by assumption, nonzero, $c_{*,i} > 0$. Hence we may conclude that that $c_{*,i} > 0$, $\bar{\lambda}_{k,i}$ converges to $\lambda_{*,i} = 0$, for $k \in \mathcal{K}$, and $c_{*,i}\lambda_{*,i} = 0$.

We note from (2.15) that the set $\mathcal{I}^* \equiv \mathcal{I}(x_*)$ is precisely the set of constraints covered in Case 1. Having thus identified the constraints in $\mathcal{A}^* \equiv \mathcal{A}(x_*)$ as those in Case 2 above, we consider Case 2 in detail.

Case 2. By construction, at every iteration of the algorithm, $\bar{\lambda}_k > 0$. Moreover, from (2.6), (2.12), (A.4) and Case 1 above,

(B.18) $$\begin{aligned}&\|(g_k - A_{k,\mathcal{A}^*}^T\bar{\lambda}_{k,\mathcal{A}^*})_{\mathcal{F}_1}\|\\ \leq\ &\|(A_{k,\mathcal{I}^*}^T\bar{\lambda}_{k,\mathcal{I}^*})_{\mathcal{F}_1}\| + \|P(x_k,\nabla_x\Psi_k)_{\mathcal{F}_1}\|\\ \leq\ &\|(A_{k,\mathcal{I}^*}^T\bar{\lambda}_{k,\mathcal{I}^*})_{\mathcal{F}_1}\| + \omega_k \leq \bar{\omega}_k\end{aligned}$$

for some $\bar{\omega}_k$ converging to zero. Thus, in view of AS2 and Lemma 4.1, the Lagrange multiplier estimates $\bar{\lambda}_{k,\mathcal{A}^*}$ are bounded and, as $\mathcal{L}(x_k,\bar{\omega}_k;x_*,\mathcal{F}_1)$ is nonempty, these multipliers have at least one limit point. If $\lambda_{\mathcal{A}}$ is such a limit, AS1, (B.18) and the identity $c_{*,\mathcal{A}^*} = 0$ ensure that $(g_* - A_{*,\mathcal{A}^*}^T\lambda_{\mathcal{A}})_{\mathcal{F}_1} = 0$, $c_{*,\mathcal{A}^*}^T\lambda_{\mathcal{A}} = 0$ and $\lambda_{\mathcal{A}} \geq 0$.

Thus, from AS2, there is a subsequence $\mathcal{K}' \subseteq \mathcal{K}$ for which $\{x_k\}$ converges to x_* and $\{\bar{\lambda}_k\}$ converges to λ_* as $k \in \mathcal{K}'$ tends to infinity and hence, from (2.4), $\nabla_x\Psi_k$ converges to g_*^ℓ. We also have that

(B.19) $$c_*^T\lambda_* = 0$$

with both $c_{*,i}$ and $\lambda_{*,i}$ $(i = 1,\ldots,m)$ nonnegative and at least one of the pair equal to zero. We may now invoke Lemma 2.1, and the convergence of $\nabla_x\Psi_k$ to g_*^ℓ to see that

(B.20) $$g_{*,\mathcal{F}_1}^\ell = 0 \quad\text{and}\quad x_*^T g_*^\ell = 0.$$

The variables in the set $\mathcal{F}_1 \cap \mathcal{N}_b$ are, by definition, positive at x_*. The components of g_*^ℓ indexed by $\mathcal{D}_1$ are nonnegative from (2.10), as their corresponding variables are dominated. This then gives the conditions

(B.21) $$\begin{aligned} x_{*,i} > 0 \quad\text{and}\quad g_{*,i}^\ell &= 0 \quad\text{for}\quad i \in \mathcal{F}_1 \cap \mathcal{N}_b,\\ g_{*,i}^\ell &= 0 \quad\text{for}\quad i \in \mathcal{F}_1 \cap \mathcal{N}_f,\\ x_{*,i} = 0 \quad\text{and}\quad g_{*,i}^\ell &\geq 0 \quad\text{for}\quad i \in \mathcal{D}_1 \quad\text{and}\\ x_{*,i} = 0 \quad\text{and}\quad g_{*,i}^\ell &= 0 \quad\text{for}\quad i \in \mathcal{F}_4.\end{aligned}$$

Thus, we have shown that x_* is a Kuhn-Tucker point and hence we have established results (i), (ii) and (iii). It remains to prove (iv).

If μ_k is bounded away from zero, we have established in Case 1a above that $\lambda_{k,i}$ converges to zero. Hence, as μ_k is finite, $s_{k,i}$ also converges to zero. On the other hand, if μ_k converges to zero, we have established in Case 1b that $\mu_k\sqrt{\lambda_{k,i}}$ and hence, once again, $s_{k,i}$ converge to zero. But as $i \in \mathcal{I}^*$, $c_{k,i}$ is bounded away from zero for all $k \in \mathcal{K}$ sufficiently large, and therefore $\sigma_{k,i}$ converges to zero for all $i \in \mathcal{I}^*$ which establishes (iv).

B.3. Proof of Theorem 4.3. Assumption AS3 implies that there is at most one point in $\mathcal{L}(x_*, 0; x_*, \mathcal{F}_1)$ and thus AS2 holds. The conclusions of Theorem 4.2 then follow. The conclusion (v) of the current theorem is a direct consequence of AS3.

We have already identified the set of constraints for which $c_i(x_*) = 0$ with $\mathcal{A}^*$. Let

$$\sigma_{k,i} \stackrel{\text{def}}{=} \frac{s_{k,i}}{c_{k,i} + s_{k,i}}. \tag{B.22}$$

Then (2.3) shows that $\bar\lambda_{k,i} = \sigma_{k,i}\lambda_{k,i}$. We now prove that $\sigma_{k,i}$ converges to zero for all $i \in \mathcal{I}^*$ as $k \in \mathcal{K}$ tends to infinity.

To prove (vi), we let $\bar\Omega$ be any closed, bounded set containing the iterates x_k, $k \in \mathcal{K}$. We note that, as a consequence of AS1 and AS3, for $k \in \mathcal{K}$ sufficiently large, $A^+_{k,\mathcal{A}^*,\mathcal{F}_1}$ exists, is bounded and converges to $A^+_{*,\mathcal{A}^*,\mathcal{F}_1}$. Thus, we may write

$$\|A^+_{k,\mathcal{A}^*,\mathcal{F}_1}\| \le a_1 \tag{B.23}$$

for some constant $a_1 > 0$. As the variables in the set $\mathcal{F}_1$ are floating, equations (2.6), (2.7), (2.12) and the inner iteration termination criterion (A.4) give that

$$\|g_{k,\mathcal{F}_1} + A^T_{k,\mathcal{A}^*,\mathcal{F}_1}\bar\lambda_{k,\mathcal{A}^*} + A^T_{k,\mathcal{I}^*,\mathcal{F}_1}\bar\lambda_{k,\mathcal{I}^*}\| \le \omega_k. \tag{B.24}$$

By assumption, $\lambda(x)_{\mathcal{A}^*}$ is bounded for all x in a neighborhood of x_*. Thus, we may deduce from (4.2), (B.23) and (B.24) that

$$\begin{aligned} \|\bar\lambda_{k,\mathcal{A}^*} - \lambda_{k,\mathcal{A}^*}\| &= \|A^{+\,T}_{k,\mathcal{A}^*,\mathcal{F}_1} g_{k,\mathcal{F}_1} + \bar\lambda_{k,\mathcal{A}^*}\| \\ &= \|A^{+\,T}_{k,\mathcal{A}^*,\mathcal{F}_1}(g_{k,\mathcal{F}_1} + A^T_{k,\mathcal{A}^*,\mathcal{F}_1}\bar\lambda_{k,\mathcal{A}^*})\| \\ &\le \|A^{+\,T}_{k,\mathcal{A}^*,\mathcal{F}_1}\|(\omega_k + \|A^T_{k,\mathcal{I}^*,\mathcal{F}_1}\|\|\bar\lambda_{k,\mathcal{I}^*}\| \\ &\le a_1\omega_k + a_3\|\bar\lambda_{k,\mathcal{I}^*}\|, \end{aligned} \tag{B.25}$$

where $a_3 \stackrel{\text{def}}{=} a_1 \max_{x\in\bar\Omega}\|A(x)^T_{\mathcal{I}^*,\mathcal{F}_1}\|$. Moreover, from the integral mean-value theorem and the (local) differentiability of the least-squares Lagrange multiplier estimates (see, for example, Conn et al. [11, Lemma 2.2]) we have that

$$\lambda_{k,\mathcal{A}^*} - \lambda_{*,\mathcal{A}^*} = \left(\int_0^1 \nabla_x\lambda(x(t))_{\mathcal{A}^*}dt\right)(x_k - x_*), \tag{B.26}$$

where $\nabla_x\lambda(x)_{\mathcal{A}^*}$ is given by Conn et al. [11, equation 2.17], and where $x(t) = x_k + t(x_* - x_k)$. Now the terms within the integral sign are bounded for all x sufficiently close to x_* and hence (B.26) gives

$$\|\lambda_{\mathcal{A}^*,k} - \lambda_{\mathcal{A}^*,*}\| \le a_2\|x_k - x_*\| \tag{B.27}$$

for all $k \in \mathcal{K}$ sufficiently large, for some constant $a_2 > 0$, which is just the inequality (4.5). We then have that $\lambda_{\mathcal{A}^*,k}$ converges to $\lambda_{\mathcal{A}^*,*}$. Combining (4.1), (B.25) and (B.27), we obtain

$$\begin{aligned} \|\bar\lambda_{k,\mathcal{A}^*} - \lambda_{*,\mathcal{A}^*}\| &\le \|\bar\lambda_{k,\mathcal{A}^*} - \lambda_{k,\mathcal{A}^*}\| + \|\lambda_{k,\mathcal{A}^*} - \lambda_{*,\mathcal{A}^*}\| \\ &\le a_1\omega_k + a_2\|x_k - x_*\| + a_3\|\bar\lambda_{k,\mathcal{I}^*}\| \\ &\le a_1\omega_k + a_2\|x_k - x_*\| + a_3\sigma_k\|\lambda_{k,\mathcal{I}^*}\|, \end{aligned} \tag{B.28}$$

the required inequality (4.4). It remains to establish (4.6) and (4.7).

The relationships (2.5) and (A.3) imply that

$$c_{k,i} = \mu_k(\sqrt{\lambda_{k,i}}/\bar\lambda_{k,i})(\lambda_{k,i} - \bar\lambda_{k,i}) \tag{B.29}$$

and

$$c_{k,i}\bar\lambda_{k,i}/\sqrt{\lambda_{k,i}} = \mu_k(\lambda_{k,i} - \bar\lambda_{k,i}) \tag{B.30}$$

for $1 \le i \le m$. Bounding (B.29) and using the triangle inequality and the inclusion $\mathcal{A} \subseteq \mathcal{A}^*$, we obtain

$$\begin{aligned} \|c_{k,\mathcal{A}}\| &\le \mu_k\left\|[\sqrt{\lambda_{k,i}}/\bar\lambda_{k,i}]_{i\in\mathcal{A}}\right\|\|(\bar\lambda_k - \lambda_k)_{\mathcal{A}}\| \\ &\le \mu_k\left\|[\sqrt{\lambda_{k,i}}/\bar\lambda_{k,i}]_{i\in\mathcal{A}}\right\|[\|(\bar\lambda_k - \lambda_*)_{\mathcal{A}}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}}\|] \\ &\le \mu_k\left\|[\sqrt{\lambda_{k,i}}/\bar\lambda_{k,i}]_{i\in\mathcal{A}}\right\|[\|(\bar\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}}\|]. \end{aligned} \tag{B.31}$$

But then, combining (B.28) and (B.31), we see that (4.7) holds for all $k \in \mathcal{K}$ sufficiently large. Furthermore, the triangle inequality, the relationships (4.1), (4.4) and

$$\lambda_{*,\mathcal{I}^*} = 0 \tag{B.32}$$

yield the bound

$$\begin{aligned} \|\bar\lambda_k - \lambda_k\| &\le \|\bar\lambda_k - \lambda_*\| + \|\lambda_k - \lambda_*\| \\ &\le \|(\bar\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + \|\bar\lambda_{k,\mathcal{I}^*}\| + \|\lambda_{k,\mathcal{I}^*}\| \\ &\le a_1\omega_k + a_2\|x_k - x_*\| + (1 + (1 + a_3)\sigma_k)\|\lambda_{k,\mathcal{I}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\|. \end{aligned} \tag{B.33}$$

Hence, taking norms of (B.30) and using (B.33), we see that (4.6) holds for all $k \in \mathcal{K}$ sufficiently large.

C. Details of proofs from §5

C.1. Proof of Lemma 5.1. We first need to make some observations concerning the status of the variables as the limit point is approached. We pick k sufficiently large that the sets $\mathcal{F}_1$ and $\mathcal{D}_1$, defined in (2.13), have been determined. Then, for $k \in \mathcal{K}$, the remaining variables either float (variables in $\mathcal{F}_2$) or oscillate between floating and being dominated (variables in $\mathcal{F}_3$). Now recall the definition (2.14) of $\mathcal{F}_4$ and pick an infinite subsequence, $\tilde{\mathcal{K}}$, of $\mathcal{K}$ such that:

(i) $\mathcal{F}_4 = \mathcal{F}_5 \cup \mathcal{D}_2$ with $\mathcal{F}_5 \cap \mathcal{D}_2 = \emptyset$;
(ii) variables in $\mathcal{F}_5$ are floating for all $k \in \tilde{\mathcal{K}}$; and
(iii) variables in $\mathcal{D}_2$ are dominated for all $k \in \tilde{\mathcal{K}}$.

Notice that the set $\mathcal{F}_2$ of (2.13) is contained within $\mathcal{F}_5$. Note, also, that there are only a finite number ($\le 2^{|\mathcal{F}_4|}$) of such subsequences $\tilde{\mathcal{K}}$ and that for k sufficiently large, each $k \in \mathcal{K}$ is in one such subsequence. It is thus sufficient to prove the lemma for $k \in \tilde{\mathcal{K}}$.

Now, for $k \in \tilde{\mathcal{K}}$, define

$$\mathcal{F} \stackrel{\text{def}}{=} \mathcal{F}_1 \cup \mathcal{F}_5 \quad\text{and}\quad \mathcal{D} \stackrel{\text{def}}{=} \mathcal{D}_1 \cup \mathcal{D}_2. \tag{C.1}$$

So, the variables in $\mathcal{F}$ are floating while those in $\mathcal{D}$ are dominated.

We also need to consider the status of the constraints in $\mathcal{A}_2^*$. We choose a χ satisfying (5.5) and pick an infinite subsequence, $\bar{\mathcal{K}}$, of $\tilde{\mathcal{K}}$ such that

(a) $\mathcal{A}_2^* = \mathcal{A}_s^* \cup \mathcal{A}_b^*$ with $\mathcal{A}_s^* \cap \mathcal{A}_b^* = \emptyset$, where $\mathcal{A}_s^*$ and $\mathcal{A}_b^*$ are defined below;

(b) the Lagrange multiplier estimates satisfy

$$\bar{\lambda}_{k,i} \le \mu_k^{1-\chi}\sqrt{\lambda_{k,i}} \tag{C.2}$$

for all constraints $i \in \mathcal{A}_s^*$ and all $k \in \bar{\mathcal{K}}$; and

(c) the Lagrange multiplier estimates satisfy

$$\bar{\lambda}_{k,i} > \mu_k^{1-\chi}\sqrt{\lambda_{k,i}} \tag{C.3}$$

for all constraints $i \in \mathcal{A}_b^*$ and all $k \in \bar{\mathcal{K}}$.

We note that there are only a finite number ($\le 2^{|\mathcal{A}_2^*|}$) of such subsequences $\bar{\mathcal{K}}$ and that for k sufficiently large, each $k \in \mathcal{K}$ is in one such subsequence. It is thus sufficient to prove the lemma for $k \in \bar{\mathcal{K}}$.

We define

$$\mathcal{A} = \mathcal{A}_1^* \cup \mathcal{A}_b^* \tag{C.4}$$

and note that this set is consistent with the set $\mathcal{A}$ described by AS5. It then follows from (5.1) and (C.4) that

$$\mathcal{A}^* = \mathcal{A} \cup \mathcal{A}_s^* \quad \text{with} \quad \mathcal{A} \cap \mathcal{A}_s^* = \emptyset. \tag{C.5}$$

We note that, if $i \in \mathcal{A}_b^*$, (C.3) gives

$$\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i} < \mu_k^{\chi-1} \tag{C.6}$$

for all $k \in \bar{\mathcal{K}}$. Moreover, inequalities (5.4) and (5.5) imply

$$\left\| \left[\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i}\right]_{i \in \mathcal{A}_1^*} \right\| \le a_4 \mu_k^{\zeta-1} \le a_4 \mu_k^{\chi-1}. \tag{C.7}$$

It then follows directly from (C.6) and (C.7) that

$$\left\| \left[\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i}\right]_{i \in \mathcal{A}} \right\| \le a_{14} \mu_k^{\chi-1} \tag{C.8}$$

for some positive constants χ, satisfying (5.5), and a_{14} and for all $k \in \mathcal{K}$. Furthermore

$$\lambda_{*,\mathcal{A}_s^*} = 0, \tag{C.9}$$

as $\mathcal{A}_s^* \subseteq \mathcal{A}_2^*$. Finally, the same inclusion and (C.2) imply that

$$\|\bar{\lambda}_{k,\mathcal{A}_s^*}\| \le \mu_k^{1-\chi} \left\| \left[\sqrt{\lambda_{k,i}}\right]_{i \in \mathcal{A}_s^*} \right\| \le \mu_k^{1-\chi} \left\| \left[\sqrt{\lambda_{k,i}}\right]_{i \in \mathcal{A}_2^*} \right\| \tag{C.10}$$

for all $k \in \bar{\mathcal{K}}$.

We may now invoke Theorem 4.3, part (vi), the bound (C.8) and the inclusion $\mathcal{A} \subseteq \mathcal{A}^*$ to obtain the inequalities

$$\|(\bar{\lambda}(x_k, \lambda_k, s_k) - \lambda_*)_{\mathcal{A}}\| \le a_1 \omega_k + a_2 \|x_k - x_*\| + a_3 \sigma_k \|\lambda_{k,\mathcal{I}^*}\| \tag{C.11}$$

and

$$\|c_{k,\mathcal{A}}\| \le \; a_{14}\mu_k^{\chi}[a_1\omega_k + a_2\|x_k - x_*\| + a_3\sigma_k\|\lambda_{k,\mathcal{I}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\|] \tag{C.12}$$

for all sufficiently large $k \in \bar{\mathcal{K}}$. Moreover, $\bar{\lambda}_k$ converges to λ_* and hence (2.4) implies that $\nabla_x \Psi_k$ converges to g_*^ℓ. Therefore, from Lemma 2.1,

$$x_{*,i} = 0 \quad \text{for all} \quad i \in \mathcal{D} \quad \text{and} \quad g_{*,i}^\ell = 0 \quad \text{for all} \quad i \in \mathcal{F}. \tag{C.13}$$

Using Taylor's theorem and the identities (B.32), (C.5) and (C.9), we have

$$\begin{aligned} \nabla_x \Psi_k &= g_k + A_k^T \bar{\lambda}_k \\ &= g_* + H_*(x_k - x_*) + A_*^T \bar{\lambda}_k + \\ &\quad \textstyle\sum_{j=1}^m \bar{\lambda}_{k,j} H_{*,j}(x_k - x_*) + r_1(x_k, x_*, \bar{\lambda}_k) \\ &= g_*^\ell + H_*^\ell(x_k - x_*) + A_{*,\mathcal{A}}^T(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}} + A_{*,\mathcal{A}_s^*}^T \bar{\lambda}_{k,\mathcal{A}_s^*} + \\ &\quad A_{*,\mathcal{I}^*}^T \bar{\lambda}_{k,\mathcal{I}^*} + r_1(x_k, x_*, \bar{\lambda}_k) + r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*), \end{aligned} \tag{C.14}$$

where

$$r_1(x_k, x_*, \bar{\lambda}_k) = \int_0^1 (H^\ell(x_k + t(x_* - x_k), \bar{\lambda}_k) - H^\ell(x_*, \bar{\lambda}_k))(x_k - x_*)dt \tag{C.15}$$

and

$$r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*) = \sum_{j=1}^m (\bar{\lambda}_{k,j} - \lambda_{j,*}) H_j(x_*)(x_k - x_*). \tag{C.16}$$

The boundedness and Lipschitz continuity of the Hessian matrices of f and the c_i in a neighborhood of x_* along with the convergence of $\bar{\lambda}_k$ to λ_* for which the relationships (4.8) and (B.32) hold then give that

$$\|r_1(x_k, x_*, \bar{\lambda}_k)\| \le a_{15}\|x_k - x_*\|^2 \tag{C.17}$$

and

$$\begin{aligned} \|r_2(x_k, x_*, \bar{\lambda}_k, \lambda_*)\| &\le a_{16}\|x_k - x_*\| \|\bar{\lambda}_k - \lambda_*\| \\ &\le a_{16}\|x_k - x_*\|(\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}^*}\| + \sigma_k\|\lambda_{k,\mathcal{I}^*}\|) \end{aligned} \tag{C.18}$$

for some positive constants a_{15} and a_{16}, using (4.1). In addition, again using Taylor's theorem and that $c_{*,\mathcal{A}} = 0$, we have

$$c_{k,\mathcal{A}} = A_{*,\mathcal{A}}(x_k - x_*) + r_3(x_k, x_*)_{\mathcal{A}}, \tag{C.19}$$

where

$$(r_3(x_k, x_*))_i = \int_0^1 t_2 \int_0^1 (x_k - x_*)^T H_i(x_* + t_1 t_2(x_k - x_*))(x_k - x_*) dt_1 dt_2 \tag{C.20}$$

for $i \in \mathcal{A}$ (see Gruver and Sachs [26, p. 11]). The boundedness of the Hessian matrices of the c_i in a neighborhood of x_* then gives that

$$\|r_3(x_k, x_*)_{\mathcal{A}}\| \le a_{17}\|x_k - x_*\|^2 \tag{C.21}$$

for some constant $a_{17} > 0$. Combining (C.14) and (C.19), we obtain

$$\begin{aligned} &\begin{pmatrix} H_*^\ell & A_{*,\mathcal{A}}^T \\ A_{*,\mathcal{A}} & 0 \end{pmatrix} \begin{pmatrix} x_k - x_* \\ (\bar{\lambda}_k - \lambda_*)_{\mathcal{A}} \end{pmatrix} \\ &= \begin{pmatrix} \nabla_x \Psi_k - g_*^\ell - A_{*,\mathcal{I}^*}^T \bar{\lambda}_{k,\mathcal{I}^*} - A_{*,\mathcal{A}_s^*}^T \bar{\lambda}_{k,\mathcal{A}_s^*} \\ c_{k,\mathcal{A}} \end{pmatrix} - \begin{pmatrix} r_1 + r_2 \\ (r_3)_{\mathcal{A}} \end{pmatrix}, \end{aligned} \tag{C.22}$$

where we have suppressed the arguments of r_1, r_2 and r_3 for brevity. We may then use (C.13) to rewrite (C.22) as

$$\begin{pmatrix} H^\ell_{*,\mathcal{F},\mathcal{F}} & H^\ell_{*,\mathcal{F},\mathcal{D}} & A^T_{*,\mathcal{A},\mathcal{F}} \\ H^\ell_{*,\mathcal{D},\mathcal{F}} & H^\ell_{*,\mathcal{D},\mathcal{D}} & A^T_{*,\mathcal{A},\mathcal{D}} \\ A_{*,\mathcal{A},\mathcal{F}} & A_{*,\mathcal{A},\mathcal{D}} & 0 \end{pmatrix} \begin{pmatrix} (x_k - x_*)_\mathcal{F} \\ x_{k,\mathcal{D}} \\ (\bar\lambda_k - \lambda_*)_\mathcal{A} \end{pmatrix}$$
$$= \begin{pmatrix} \nabla_x \Psi_{k,\mathcal{F}} - A^T_{*,\mathcal{A}_2^*,\mathcal{F}} \bar\lambda_{k,\mathcal{A}_2^*} - A^T_{*,\mathcal{I}^*,\mathcal{F}} \bar\lambda_{k,\mathcal{I}^*} \\ \nabla_x \Psi_{k,\mathcal{F}} - g^\ell_{*,\mathcal{D}} - A^T_{*,\mathcal{A}_2^*,\mathcal{D}} \bar\lambda_{k,\mathcal{A}_2^*} - A^T_{*,\mathcal{I}^*,\mathcal{D}} \bar\lambda_{k,\mathcal{I}^*} \\ c_{k,\mathcal{A}} \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_\mathcal{F} \\ (r_1 + r_2)_\mathcal{D} \\ (r_3)_\mathcal{A} \end{pmatrix}. \quad \text{(C.23)}$$

Then, rearranging (C.23) and removing the middle horizontal block, we obtain

$$\begin{pmatrix} H^\ell_{*,\mathcal{F},\mathcal{F}} & A^T_{*,\mathcal{A},\mathcal{F}} \\ A_{*,\mathcal{A},\mathcal{F}} & 0 \end{pmatrix} \begin{pmatrix} (x_k - x_*)_\mathcal{F} \\ (\bar\lambda_k - \lambda_*)_\mathcal{A} \end{pmatrix} =$$
$$\begin{pmatrix} \nabla_x \Psi_{k,\mathcal{F}} - H^\ell_{*,\mathcal{F},\mathcal{D}} x_{k,\mathcal{D}} - A^T_{*,\mathcal{A}_2^*,\mathcal{F}} \bar\lambda_{k,\mathcal{A}_2^*} - A^T_{*,\mathcal{I}^*,\mathcal{F}} \bar\lambda_{k,\mathcal{I}^*} \\ c_{k,\mathcal{A}} - A_{*,\mathcal{A},\mathcal{D}} x_{k,\mathcal{D}} \end{pmatrix} - \begin{pmatrix} (r_1 + r_2)_\mathcal{F} \\ (r_3)_\mathcal{A} \end{pmatrix}. \quad \text{(C.24)}$$

Roughly, the rest of the proof proceeds by showing that the right-hand side of (C.24) is $O(\omega_k) + O(\sigma_k \|\lambda_{k,\mathcal{I}^*}\|) + O(\mu_k \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\|)$. This will then ensure that the vector on the left-hand side is of the same size, which is the result we require. First observe that

$$\|x_{k,\mathcal{D}}\| \le \omega_k, \quad \text{(C.25)}$$

from (2.11) and (A.4), and

$$\|\nabla_x \Psi_{k,\mathcal{F}}\| \le \omega_k, \quad \text{(C.26)}$$

from (2.12). Consequently, using (C.13) and (C.25), we have

$$\|x_k - x_*\| \le \|(x_k - x_*)_\mathcal{F}\| + \omega_k. \quad \text{(C.27)}$$

Let $\Delta x_k = \|(x_k - x_*)_\mathcal{F}\|$. Combining (4.8), (B.32), (C.11) and (C.27), we obtain

$$\|(\bar\lambda_k - \lambda_*)_\mathcal{A}\| \le a_{18}\omega_k + a_2 \Delta x_k + a_3 \sigma_k \|\lambda_{k,\mathcal{I}^*}\|, \quad \text{(C.28)}$$

where $a_{18} \stackrel{\text{def}}{=} a_1 + a_2$. Furthermore, from (C.17), (C.18), (C.21), (C.27) and (C.28),

$$\left\| \begin{pmatrix} (r_1 + r_2)_\mathcal{F} \\ (r_3)_\mathcal{A} \end{pmatrix} \right\| \le \begin{array}{l} a_{19}(\Delta x_k)^2 + a_{20}\Delta x_k \omega_k + a_{21}\omega_k^2 + \\ \quad a_{22}\sigma_k \|\lambda_{k,\mathcal{I}^*}\| (\omega_k + \Delta x_k), \end{array} \quad \text{(C.29)}$$

where $a_{19} \stackrel{\text{def}}{=} a_{15} + a_{17} + a_{16}a_2$, $a_{20} \stackrel{\text{def}}{=} 2(a_{15} + a_{17}) + a_{16}(a_{18} + a_2)$, $a_{21} \stackrel{\text{def}}{=} a_{15} + a_{17} + a_{16}a_{18}$ and $a_{22} \stackrel{\text{def}}{=} a_{16}(1 + a_3)$. Moreover, from (C.10), (C.12), (C.25), (C.26) and (C.27),

$$\left\| \begin{pmatrix} \nabla_x \Psi_{k,\mathcal{F}} - H^\ell_{*,\mathcal{F},\mathcal{D}} x_{k,\mathcal{D}} - A^T_{*,\mathcal{A}_2^*,\mathcal{F}} \bar\lambda_{k,\mathcal{A}_2^*} - A^T_{*,\mathcal{I}^*,\mathcal{F}} \bar\lambda_{k,\mathcal{I}^*} \\ c_{k,\mathcal{A}} - A_{*,\mathcal{A},\mathcal{D}} x_{k,\mathcal{D}} \end{pmatrix} \right\| \le$$
$$a_{23}\omega_k + a_{24}\sigma_k \|\lambda_{k,\mathcal{I}^*}\| + a_{25}\mu_k^{1-\chi} \left\| \left[\sqrt{\lambda_{k,i}} \right]_{i \in \mathcal{A}_2^*} \right\| +$$
$$a_{14}\mu_k^\chi \left[a_{18}\omega_k + a_2\Delta x_k + a_3\sigma_k \|\lambda_{k,\mathcal{I}^*}\| + \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| \right], \quad \text{(C.30)}$$

where

$$a_{23} \stackrel{\text{def}}{=} 1 + \left\| \begin{pmatrix} H^\ell_{*,\mathcal{F},\mathcal{D}} \\ A_{*,\mathcal{A},\mathcal{D}} \end{pmatrix} \right\|, \quad a_{24} \stackrel{\text{def}}{=} \|A^T_{*,\mathcal{I}^*,\mathcal{F}}\| \quad \text{and} \quad a_{25} \stackrel{\text{def}}{=} \|A^T_{*,\mathcal{A}_2^*,\mathcal{F}}\|. \quad \text{(C.31)}$$

By assumption AS5, the coefficient matrix on the left-hand side of (C.24) is nonsingular. Let its inverse have norm M. Multiplying both sides of the equation by this inverse and taking norms, we obtain

$$\begin{array}{rl} \left\| \begin{pmatrix} (x_k - x_*)_\mathcal{F} \\ (\bar\lambda_k - \lambda_*)_\mathcal{A} \end{pmatrix} \right\| \le & M[a_{19}(\Delta x_k)^2 + a_{20}\Delta x_k \omega_k + a_{21}\omega_k^2 + \\ & a_{22}\sigma_k \|\lambda_{k,\mathcal{I}^*}\| (\omega_k + \Delta x_k) + a_{23}\omega_k + a_{24}\sigma_k \|\lambda_{k,\mathcal{I}^*}\| + \\ & a_{25}\mu_k^{1-\chi} \left\| \left[\sqrt{\lambda_{k,i}} \right]_{i \in \mathcal{A}_2^*} \right\| + a_{14}\mu_k^\chi (a_{18}\omega_k + a_2 \Delta x_k + \\ & \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + a_3\sigma_k \|\lambda_{k,\mathcal{I}^*}\|)] \\ = & (M a_{19}\Delta x_k + M a_{20}\omega_k + M a_2 a_{14}\mu_k^\chi)\Delta x_k + \\ & (M a_{21}\omega_k + M a_{14}a_{18}\mu_k^\chi + M a_{23})\omega_k + \\ & M a_{14}\mu_k^\chi \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + M a_{25}\mu_k^{1-\chi} \left\| \left[\sqrt{\lambda_{k,i}} \right]_{i \in \mathcal{A}_2^*} \right\| + \\ & (M a_{24} + M a_{22}(\omega_k + \Delta x_k) + M a_3 a_{14}\mu_k^\chi)\sigma_k \|\lambda_{k,\mathcal{I}^*}\|. \end{array} \quad \text{(C.32)}$$

The mechanisms of Algorithm A.1 ensure that ω_k converges to zero. Moreover, Theorem 4.2 guarantees that Δx_k also converges to zero for $k \in \bar{\mathcal{K}}$. Thus, there is a k_0 for which

$$\omega_k \le \min(1, 1/(4M a_{20})) \quad \text{(C.33)}$$

and

$$\Delta x_k \le \min(1, 1/(4M a_{19})) \quad \text{(C.34)}$$

for all $k \ge k_0$ $(k \in \bar{\mathcal{K}})$. Furthermore, let

$$\mu_{\max} \equiv \min(1, 1/(4M a_2 a_{14})^{1/\chi}). \quad \text{(C.35)}$$

Then, if $\mu_k \le \mu_{\max}$, (C.32), (C.33), (C.35) and (C.34) give

$$\begin{array}{rl} \Delta x_k \le & \frac{3}{4}\Delta x_k + M(a_{21} + a_{14}a_{18} + a_{23})\omega_k + \\ & M a_{14}\mu_k^\chi \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + M a_{25}\mu_k^{1-\chi} \left\| \left[\sqrt{\lambda_{k,i}} \right]_{i \in \mathcal{A}_2^*} \right\| + \\ & M(a_{24} + 2a_{22} + a_3 a_{14})\sigma_k \|\lambda_{k,\mathcal{I}^*}\|. \end{array} \quad \text{(C.36)}$$

Cancelling the Δx_k terms in (C.36), multiplying the resulting inequality by four and substituting into (C.27), we obtain the desired inequality (5.6), where $a_5 \stackrel{\text{def}}{=} 1 + 4M(a_{21} + a_{14}a_{18} + a_{23})$, $a_6 \stackrel{\text{def}}{=} 4M a_{14}$, $a_7 \stackrel{\text{def}}{=} 4M a_{25}$ and $a_8 \stackrel{\text{def}}{=} 4M(a_{24} + 2a_{22} + a_3 a_{14})$. The remaining inequalities (5.7) and (5.8) follow directly by substituting (5.6) into (4.4) and (4.6), the required constants being $a_9 \stackrel{\text{def}}{=} a_1 + a_2 a_5$, $a_{10} \stackrel{\text{def}}{=} a_2 a_6$, $a_{11} \stackrel{\text{def}}{=} a_2 a_7$, $a_{12} \stackrel{\text{def}}{=} a_3 + a_2 a_8$ and $a_{13} \stackrel{\text{def}}{=} 1 + a_2 a_6$.

C.2. Proof of Lemma 5.2. We have, from Theorem 4.3 and AS6, that the complete sequence of Lagrange multiplier estimates $\{\bar\lambda_k\}$ generated by Algorithm A.1 converges to λ_*. We now consider the sequence $\{\lambda_k\}$.

There are three possibilities. First, μ_k may be bounded away from zero. In this case, step 3 of Algorithm A.1 must be executed for all k sufficiently large, which ensures that $\{\lambda_k\}$ and $\{\bar\lambda_{k-1}\}$ are identical for all large k. As the latter sequence converges to λ_*, so does the former.

Secondly, μ_k may converge to zero but nonetheless there may be an infinite number of iterates for which (A.7) is satisfied. In this case, the only time adjacent members of the sequence $\{\lambda_k\}$ differ, $\lambda_k = \bar{\lambda}_{k-1}$, and we have already observed that the latter sequence $\{\bar{\lambda}_{k-1}\}$ converges to λ_*.

Finally, if the test (A.7) were to fail for all $k > k_1$, $\|\lambda_{k,\mathcal{I}^*}\|$ and $\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\|$ will remain fixed for all $k \geq k_1$, as step 4 would then be executed for all subsequent iterations. But then (4.6) implies that

$$\left\| \left[c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}} \right]_{i=1}^{m} \right\| \leq a_{26}\mu_k \tag{C.37}$$

for some constant a_{26} for all $k \geq k_2 \geq k_1$. As μ_k converges to zero as k increases, we have

$$a_{26}\mu_k \leq \eta_s \mu_k^{\frac{5}{6}} = \eta_k \tag{C.38}$$

for all k sufficiently large. But then inequality (A.7) must be satisfied for some $k \geq k_1$, contradicting the supposition. Hence, this latter possibility proves to be impossible. Thus, $\{\lambda_k\}$ converges to λ_*.

Inequality (5.9) then follows immediately for $i \in \mathcal{I}^*$ by considering the definitions (A.3), (4.8) and (B.22) and using the convergence of $\lambda_{k,\mathcal{I}^*}$ to $\lambda_{*,\mathcal{I}^*} = 0$; a suitable representation of θ_k would be

$$\theta_k = \max_{i \in \mathcal{I}^*} \left(\frac{\sqrt{\lambda_{k,i}}}{c_{k,i} + \mu_k \sqrt{\lambda_{k,i}}} \right). \tag{C.39}$$

Now $\bar{\lambda}_{k,i}$ converges to $\lambda_{*,i} > 0$ and is thus bounded away from zero for all k, for each $i \in \mathcal{A}_1^*$. But this and the convergence of $\{\bar{\lambda}_k\}$ to λ_* implies that $\sqrt{\lambda_{k,i}}/\bar{\lambda}_{k,i}$ is bounded and hence inequality (5.4), with $\zeta = 1$, holds for all k. The remaining results follow directly from Lemma 5.1 on substituting $\zeta = 1$ into (5.5).

C.3. Proof of Theorem 5.3. The appropriate version of Theorem 5.3 for the simplified Algorithm 2 is now stated:

Theorem C.1 *Suppose that the iterates $\{x_k\}$ generated by Algorithm* 3.1 *satisfy* AS6 *and that* AS4 *and* AS5 *hold. Furthermore, suppose that* AS7 *holds. Then there is a constant $\mu_{\min} > 0$ such that $\mu_k \geq \mu_{\min}$ for all k.*

Proof. Suppose, otherwise, that μ_k tends to zero. Then, step 4 of the algorithm must be executed infinitely often. We aim to obtain a contradiction to this statement by showing that step 3 is always executed for k sufficiently large. We note that our assumptions are sufficient for the conclusions of Theorem 4.3 to hold.

Lemma 5.2, part (i), ensures that $\{\lambda_k\}$ converges to λ_*. We note that, by definition,

$$\mu_k < 1. \tag{C.40}$$

Consider the convergence tolerance ω_k as generated by the algorithm. By construction

$$\omega_k \leq \omega_s \mu_k \tag{C.41}$$

for all k. (This follows by definition if step 4 of the algorithm occurs and because the penalty parameter is unchanged while ω_k is reduced when step 3 occurs.) As Lemma 5.2, part (iii), ensures that (5.4) is satisfied for all k, we may apply Lemma 5.1 to the iterates generated by the algorithm. We identify the set $\mathcal{K}$ with the complete set of integers. As we are currently assuming that μ_k converges to zero, we can ensure that μ_k is sufficiently small so that Lemma 5.1 applies to Algorithm A.1 and thus that there is an integer k_1 and constants $a_9, \ldots, a_{13}$ so that (5.7) and (5.8) hold for all $k \geq k_1$. In particular, if we choose

$$\chi = \chi_0 \stackrel{\text{def}}{=} 1, \tag{C.42}$$

we obtain the bounds

$$\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}^*}\| \leq a_9\omega_k + (a_{10} + a_{11})\mu_k \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + a_{12}\sigma_k \|\lambda_{k,\mathcal{I}^*}\| \tag{C.43}$$

and

$$\left\| \left[c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}} \right]_{i=1}^{m} \right\| \leq \mu_k \left[a_9\omega_k + (a_{11} + a_{13})\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + (1 + (1 + a_{12})\sigma_k)\|\lambda_{k,\mathcal{I}^*}\| \right] \tag{C.44}$$

for all $k \geq k_1$, from (C.40) and the inclusion $\mathcal{A}_2^* \subseteq \mathcal{A}^*$. Moreover, as Lemma 5.2, part (ii), ensures that θ_k converges to zero, there is an integer k_2 for which

$$\sigma_k \leq \mu_k \tag{C.45}$$

for all $k \geq k_2$. Thus, combining (C.40), (C.43), (C.44) and (C.45), we have that

$$\|(\bar{\lambda}_k - \lambda_*)_{\mathcal{A}^*}\| \leq a_9\omega_k + a_{27}\mu_k \|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + + a_{12}\mu_k \|\lambda_{k,\mathcal{I}^*}\| \tag{C.46}$$

and

$$\left\| \left[c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}} \right]_{i=1}^{m} \right\| \leq \mu_k \left[a_9\omega_k + a_{28}\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| + a_{29}\|\lambda_{k,\mathcal{I}^*}\| \right] \tag{C.47}$$

for all $k \geq \max(k_1, k_2)$, where $a_{27} \stackrel{\text{def}}{=} a_{10} + a_{11}$, $a_{28} \stackrel{\text{def}}{=} a_{11} + a_{13}$ and $a_{29} \stackrel{\text{def}}{=} 2 + a_{12}$.

Now, let k_3 be the smallest integer such that

$$\mu_k^{\frac{1}{6}} \leq \frac{\eta_s}{\omega_s a_{30}}, \tag{C.48}$$

$$\mu_k^{\frac{5}{6}} \leq \min\left(1, \frac{1}{a_{31}}\right), \tag{C.49}$$

$$\mu_k \leq \frac{\eta_s}{2\omega_s a_9} \tag{C.50}$$

and

$$\mu_k \leq \frac{\eta_s}{2\omega_s(a_{29} + a_{28}a_{31})} \tag{C.51}$$

for all $k \geq k_3$, where $a_{30} \stackrel{\text{def}}{=} a_9 + a_{28} + a_{29}$ and $a_{31} \stackrel{\text{def}}{=} a_9 + a_{12} + a_{27}$. Furthermore, let k_4 be such that

$$\|(\lambda_k - \lambda_*)_{\mathcal{A}^*}\| \leq \omega_s \quad \text{and} \quad \|\lambda_{k,\mathcal{I}^*}\| \leq \omega_s \tag{C.52}$$

for all $k \geq k_4$.

Finally, define $k_5 = \max(k_1, k_2, k_3, k_4)$, let Γ be the set $\{k|$ step 4 is executed at iteration $k-1$ and $k \geq k_5\}$, and let k_0 be the smallest element of Γ. By assumption, Γ has an infinite number of elements.

For iteration k_0, $\omega_{k_0} = \omega_s \mu_{k_0}$ and $\eta_{k_0} = \eta_s \mu_{k_0}$. Then (C.47) gives

$$\begin{aligned} &\| [c_{k_0,i}\bar{\lambda}_{k_0,i}/\sqrt{\lambda_{k_0,i}}]_{i=1}^m \| \\ &\quad \leq \mu_{k_0}\left[a_9\omega_{k_0} + a_{28}\|(\lambda_{k_0}-\lambda_*)_{\mathcal{A}^*}\| + a_{29}\|\lambda_{k_0,\mathcal{I}^*}\|\right] \\ &\quad \leq \omega_s(a_9 + a_{28} + a_{29})\mu_{k_0} = \omega_s a_{30}\mu_{k_0} \quad \text{[from (C.52)]} \\ &\quad \leq \eta_s \mu_{k_0}^{\frac{5}{6}} = \eta_{k_0} \quad \text{[from (C.48)]}. \end{aligned} \tag{C.53}$$

Thus, from (C.53), step 3 of Algorithm A.1 will be executed with $\lambda_{k_0+1} = \bar{\lambda}(x_{k_0}, \lambda_{k_0}, s_{k_0})$. Inequality (C.46), in conjunction with (C.41) and (C.52) guarantees that

$$\begin{aligned} \|(\lambda_{k_0+1}-\lambda_*)_{\mathcal{A}^*}\| &\leq a_9\omega_{k_0} + a_{27}\mu_{k_0}\|(\lambda_{k_0}-\lambda_*)_{\mathcal{A}^*}\| + a_{12}\mu_{k_0}\|\lambda_{k_0,\mathcal{I}^*}\| \\ &\leq a_9\omega_s\mu_{k_0} + a_{27}\omega_s\mu_{k_0} + a_{12}\omega_s\mu_{k_0} \\ &\leq \omega_s a_{31}\mu_{k_0}. \end{aligned} \tag{C.54}$$

Furthermore, inequality (4.1), in conjunction with (4.8), (C.45), and (C.52), ensures that

$$\|\lambda_{k_0+1,\mathcal{I}^*}\| \leq \sigma_{k_0}\|\lambda_{k_0,\mathcal{I}^*}\| \leq \omega_s\mu_{k_0}. \tag{C.55}$$

We shall now assume that step 3 is executed for iterations $k_0 + i$ $(0 \leq i \leq j)$ and show that

$$\|(\lambda_{k_0+i+1}-\lambda_*)_{\mathcal{A}^*}\| \leq \omega_s a_{31}\mu_{k_0}^{1+\frac{1}{6}i} \tag{C.56}$$

and

$$\|\lambda_{k_0+i+1,\mathcal{I}^*}\| \leq \omega_s\mu_{k_0}^{1+\frac{1}{6}i}. \tag{C.57}$$

Inequalities (C.54) and (C.55) show that this is true for $j = 0$. We aim to show that the same is true for $i = j+1$. Under our supposition, we have, for iteration $k_0 + j + 1$, that $\mu_{k_0+j+1} = \mu_{k_0}$, $\omega_{k_0+j+1} = \omega_s\mu_{k_0}^{j+2}$ and $\eta_{k_0+j+1} = \eta_s\mu_{k_0}^{1+\frac{1}{6}(j+1)}$. Then (C.47) gives

$$\begin{aligned} &\| [c_{k_0+j+1,i}\bar{\lambda}_{k_0+j+1,i}/\sqrt{\lambda_{k_0+j+1,i}}]_{i=1}^m \| \\ &\quad \leq \mu_{k_0}\left[a_9\omega_s\mu_{k_0}^{j+2} + a_{28}\|(\lambda_{k_0+j+1}-\lambda_*)_{\mathcal{A}^*}\| + a_{29}\|\lambda_{k_0+j+1,\mathcal{I}^*}\|\right] \\ &\quad \leq \mu_{k_0}\left[a_9\omega_s\mu_{k_0}^{j+2} + a_{28}a_{31}\omega_s\mu_{k_0}^{1+\frac{1}{6}j} + a_{29}\omega_s\mu_{k_0}^{1+\frac{1}{6}j}\right] \\ &\qquad \text{[from (C.56)–(C.57)]} \\ &\quad \leq \omega_s(a_9 + (a_{29} + a_{28}a_{31})\mu_{k_0}^{2+\frac{1}{6}j} \\ &\quad \leq \eta_s\mu_{k_0}^{1+\frac{1}{6}j} = \eta_{k_0+j+1} \quad \text{[from (C.50)–(C.51)]}. \end{aligned} \tag{C.58}$$

Thus, from (C.58), step 3 of Algorithm A.1 will be executed with $\lambda_{k_0+j+2} = \bar{\lambda}(x_{k_0+j+1}, \lambda_{k_0+j+1}, s_{k_0+j+1})$. Inequality (C.46) then guarantees that

$$\begin{aligned} &\|(\lambda_{k_0+j+2}-\lambda_*)_{\mathcal{A}^*}\| \\ &\quad \leq a_9\omega_{k_0+j+1} + a_{27}\mu_{k_0+j+1}\|(\lambda_{k_0+j+1}-\lambda_*)_{\mathcal{A}^*}\| + a_{12}\mu_{k_0+j+1}\|\lambda_{k_0+j+1,\mathcal{I}^*}\| \\ &\quad \leq a_9\omega_s\mu_{k_0}^{j+2} + \omega_s(a_{27}a_{31}\mu_{k_0}^{\frac{5}{6}} + a_{12})\mu_{k_0}^{1+\frac{1}{6}(j+1)} \quad \text{[from (C.56)–(C.57)]} \\ &\quad \leq \omega_s a_{31}\mu_{k_0}^{1+\frac{1}{6}(j+1)} \quad \text{[from (C.49)]}, \end{aligned} \tag{C.59}$$

which establishes (C.56) for $i = j+1$.

Furthermore, inequalities (4.1) and (C.45) ensure that

$$\begin{aligned} \|\lambda_{k_0+j+2,\mathcal{I}^*}\| &\leq \sigma_{k_0+j+1}\|\lambda_{k_0+j+1,\mathcal{I}^*}\| \leq \mu_{k_0+j+1}\|\lambda_{k_0+j+1,\mathcal{I}^*}\| \quad \text{[from (4.8)]} \\ &\leq \omega_s\mu_{k_0}^{1+\frac{1}{6}(j+1))} \quad \text{[from (C.57)]}, \end{aligned} \tag{C.60}$$

which establishes (C.57) for $i = j+1$. Hence, step 3 of the algorithm is executed for all iterations $k \geq k_0$. But this implies that Γ is finite, which contradicts the assumption that step 4 is executed infinitely often. Hence the theorem is proved. □

C.4. Proof of Theorem 5.4. We proceed by considering an example which has more than one Kuhn-Tucker point and for which the optimal Lagrange multipliers are distinct. We consider a sequence of iterates which is converging satisfactorily to a single Kuhn-Tucker point $(x_{*,1}, \lambda_{*,1})$ (and thus the penalty parameter has settled down to a single value). We now introduce an "extra" iterate x_k near to a different Kuhn-Tucker point $(x_{*,2}, \lambda_{*,2})$. We make use of the identity

$$c_{k,i}\bar{\lambda}_{k,i}/\sqrt{\lambda_{k,i}} = \mu_k(\lambda_{k,i} - \bar{\lambda}_{k,i}), \tag{C.61}$$

derived from (2.5) and (A.3), to show that if the Lagrange multiplier estimate $\bar{\lambda}_{k,i}$ calculated at x_k is a sufficiently "accurate" approximation of $\lambda_{*,2}$ (while $\lambda_{k,i}$ is an "accurate" representation of $\lambda_{*,1}$), the acceptance test (A.7) will fail and the penalty parameter will be reduced. Moreover, we show that this behavior can be repeated indefinitely.

To be specific, we consider the following problem:

$$\underset{x \in \mathbb{R}}{\text{minimize}} \quad \epsilon(x-1)^2 \quad \text{such that} \quad c(x) = x^2 - 4 \geq 0, \tag{C.62}$$

where ϵ is a (yet to be specified) positive constant. It is straightforward to show that the problem has two local solutions, which occur at the Kuhn-Tucker points

$$(x_{*,1}, \lambda_{*,1}) = \left(-2, \frac{3\epsilon}{2}\right) \quad \text{and} \quad (x_{*,2}, \lambda_{*,2}) = \left(2, \frac{\epsilon}{2}\right), \tag{C.63}$$

and that the constraint is active at both local solutions. Moreover, there are no specific bounds on the variable in the problem, and hence $P(x, \nabla_x\Psi(x,\lambda,s)) = \nabla_x\Psi(x,\lambda,s)$ for all x.

We intend to construct a cycle of iterates x_{k+i}, $i = 0, \ldots, j$, for some integer j, which are allowed by Algorithm A.1. The penalty parameter remains fixed throughout the cycle until it is reduced at the end of the final iteration. We start with $\lambda_0 = \lambda_{*,1}$. We also pick ϵ so that

$$\epsilon \leq \min\left(\frac{2}{3}, \frac{\omega_s}{(6 + \frac{1}{1-\mu_0})}, \frac{2\eta_s}{3\mu_0^{\frac{1}{6}}}\right). \tag{C.64}$$

We define j to be the smallest integer for which

$$\mu_0^{\frac{j-1}{6}} < \tfrac{1}{2}\epsilon/\eta_s. \tag{C.65}$$

We let μ denote the value of the penalty parameter at the start of the cycle.

i = 0: We have $\omega_k = \omega_s\mu$ and $\eta_k = \eta_s\mu^{\frac{5}{6}}$. We are given $\lambda_k = \lambda_{*,1}$. We pick x_k near $x_{*,1}$ so that $\bar\lambda_k = (1-\mu)\lambda_{*,1}$. We show that such a choice guarantees that the convergence and acceptance tests (A.5) and (A.7) are satisfied, and thus step 3 of the algorithm is executed.

i = 1, ... , j - 2: We have $\omega_{k+i} = \omega_s\mu^{1+i}$ and $\eta_{k+i} = \eta_s\mu^{\frac{5+i}{6}}$. We have $\lambda_{k+i} = (1-\mu_i)\lambda_{*,1}$. We pick x_{k+i} near $x_{*,1}$ so that $\bar\lambda_{k+i} = (1-\mu_{i+1})\lambda_{*,1}$. We again show that such a choice guarantees that the convergence and acceptance tests (A.5) and (A.7) are satisfied, and thus step 3 of the algorithm is executed.

i = j - 1: We have $\omega_{k+i} = \omega_s\mu^{1+i}$ and $\eta_{k+i} = \eta_s\mu^{\frac{5+i}{6}}$. We have $\lambda_{k+i} = (1-\mu_i)\lambda_{*,1}$. We pick x_{k+i} near $x_{*,1}$ so that $\bar\lambda_{k+i} = \lambda_{*,1}$. Once again, we show that such a choice guarantees that the convergence and acceptance tests (A.5) and (A.7) are satisfied, and thus step 3 of the algorithm is executed.

i = j: We have $\omega_{k+j} = \omega_s\mu^{1+j}$ and $\eta_{k+j} = \eta_s\mu^{\frac{5+j}{6}}$. We have $\lambda_{k+j} = \lambda_{*,1}$. We pick x_{k+i} as the local minimizer of the Lagrangian barrier function which is larger than $x_{*,2}$, which trivially ensures that the convergence test (A.5) is satisfied. We also show that the acceptance test (A.7) is violated at this point, so that step 4 of the algorithm will be executed and the penalty parameter reduced.

It is clear that if an infinite sequence of such cycles occur, the penalty parameter μ_k will converge to zero. We now show that this is possible.

If a is a real number, we will make extensive use of the trivial inequalities

(C.66) $$1 \le \sqrt{1+a} \le 1+a \quad \text{whenever} \quad a \ge 0$$

and

(C.67) $$1-a \le \sqrt{1-a} \le 1-\tfrac{1}{2}a \quad \text{whenever} \quad 0 \le a \le 1.$$

We also remind the reader that

(C.68) $$\mu \le \mu_0 < 1.$$

1. Let

(C.69) $$x_k = -2\sqrt{1+\tfrac{1}{4}\mu s_k/(1-\mu)},$$

where the shift $s_k = \mu\sqrt{\frac{3}{2}}\epsilon$. Then it is easy to verify that $\bar\lambda_k = (1-\mu)\lambda_{*,1}$. Moreover,

(C.70) $$\begin{aligned}\nabla_x\Psi(x_k,\lambda_k,s_k) &= 2\epsilon(x_k-1) - 3\epsilon(1-\mu)x_k = -\epsilon(2+(1-3\mu)x_k)\\ &= -2\epsilon\left(1-(1-3\mu)\sqrt{1+\mu s_k/(4(1-\mu))}\right).\end{aligned}$$

Taking norms of (C.70) and using (C.66) yields

(C.71) $$\|P(x_k,\nabla_x\Psi(x_k,\lambda_k,s_k))\| \le \begin{cases} 6\epsilon\mu & \text{if } \mu \le \frac{1}{3},\\ 2\epsilon\mu\left(3+\dfrac{(3\mu-1)s_k}{4(1-\mu)}\right) & \text{otherwise.}\end{cases}$$

Now (C.64) implies that $s_k \le \mu < 1$, and thus we obtain the overall bound

(C.72) $$\|P(x_k,\nabla_x\Psi(x_k,\lambda_k,s_k))\| \le \epsilon\left(6+\frac{1}{1-\mu_0}\right)\mu$$

from (C.68) and (C.71). But then (C.72) and (C.64) give

(C.73) $$\|P(x_k,\nabla_x\Psi(x_k,\lambda_k,s_k))\| \le \omega_s\mu = \omega_k,$$

as $1 \le \omega_s/(6+1/(1-\mu_0))\epsilon$. Furthermore, from (C.61) and (C.64),

(C.74) $$\|c(x_k)\bar\lambda_k/\sqrt{\lambda_k}\| = \mu\|\lambda_{k,i}-\bar\lambda_{k,i}\| = \tfrac{3}{2}\mu^2\epsilon \le \eta_s\mu^{\frac{5}{6}} = \eta_k,$$

as $\mu^{\frac{7}{6}} \le \mu_0^{\frac{1}{6}} \le 1 \le 2\eta_s/3\epsilon$. Thus, x_k satisfies (A.5) and (A.7), and hence step 3 of the algorithm will be executed. Therefore, in particular, $\omega_{k+1} = \omega_s\mu^2$, $\eta_{k+1} = \eta_s\mu$ and $\lambda_{k+1} = (1-\mu)\lambda_{*,1}$.

2. For $i = 1,\ldots j-2$, let

(C.75) $$x_{k+i} = -2\sqrt{1-\tfrac{1}{4}\mu^i(1-\mu)s_{k+i}/(1-\mu^{i+1})},$$

where the shift $s_{k+i} = \mu\sqrt{\frac{3}{2}(1-\mu^i)\epsilon}$. Note that (C.75) is well defined, as the second term within the square root is less than $\frac{1}{4}$ in magnitude because (C.64) and (C.68) imply that $s_k < \mu$ and $\mu^i(1-\mu)/(1-\mu^{i+1}) < 1$. It is then easy to verify that $\bar\lambda_{k+i} = (1-\mu^{i+1})\lambda_{*,1}$. Moreover,

(C.76) $$\begin{aligned}\nabla_x\Psi(x_{k+i},\lambda_{k+i},s_{k+i}) &= 2\epsilon(x_{k+i}-1) - 3\epsilon(1-\mu^{i+1})x_{k+i}\\ &= -\epsilon(2+(1-3\mu^{i+1})x_{k+i})\\ &= -2\epsilon\left(1-(1-3\mu^{i+1})\sqrt{1-\tfrac{\mu^i(1-\mu)s_{k+i}}{4(1-\mu^{i+1})}}\right)\end{aligned}$$

Now suppose $\mu^{i+1} \le \frac{1}{3}$. Then (C.76), (C.67), (C.68) and $s_k \le \mu$ yield

(C.77) $$\begin{aligned}\|P(x_{k+i},\nabla_x\Psi(x_{k+i},\lambda_{k+i},s_{k+i}))\| &\le 2\epsilon\left(1-(1-3\mu^{i+1})(1-\tfrac{\mu^i(1-\mu)s_{k+i}}{8(1-\mu^{i+1})})\right)\\ &= 2\epsilon\left(3\mu^{i+1}+\tfrac{\mu^i(1-\mu)(1-3\mu^{i+1})s_{k+i}}{8(1-\mu^{i+1})}\right)\\ &\le 2\epsilon\mu^{i+1}\left(3+\tfrac{(1-\mu)(1-3\mu^{i+1})}{8(1-\mu^{i+1})}\right)\\ &\le 2\epsilon\mu^{i+1}\left(3+\tfrac{1}{8(1-\mu_0)}\right).\end{aligned}$$

If, on the other hand, $\mu^{i+1} > \frac{1}{3}$, the same relationships give

(C.78) $$\begin{aligned}\|P(x_{k+i},\nabla_x\Psi(x_{k+i},\lambda_{k+i},s_{k+i}))\| &\le 2\epsilon\left(1-(1-3\mu^{i+1})(1-\tfrac{\mu^i(1-\mu)s_{k+i}}{4(1-\mu^{i+1})})\right)\\ &= 2\epsilon\left(3\mu^{i+1}+\tfrac{\mu^i(1-\mu)(1-3\mu^{i+1})s_{k+i}}{4(1-\mu^{i+1})}\right)\\ &\le 6\epsilon\mu^{i+1}.\end{aligned}$$

Thus, combining (C.77) and (C.78), we certainly have that

(C.79) $$\|P(x_{k+i},\nabla_x\Psi(x_{k+i},\lambda_{k+i},s_{k+i}))\| \le \epsilon\left(6+\frac{1}{1-\mu_0}\right)\mu^{i+1}.$$

But then (C.79) and (C.64) give

(C.80) $$\|P(x_{k+i},\nabla_x\Psi(x_{k+i},\lambda_{k+i},s_{k+i}))\| \le \omega_s\mu^{1+i} = \omega_{k+i},$$

as $1 \leq \omega_s/((6+1/(1-\mu_0))\epsilon)$. Furthermore, from (C.61) and (C.64),

$$\text{(C.81)}\quad \begin{aligned} \|c(x_{k+i})\bar\lambda_{k+i}/\sqrt{\lambda_{k+i}}\| &= \mu\|\lambda_{k+i,i}-\bar\lambda_{k+i,i}\| = \tfrac{3}{2}\mu^{i+1}(1-\mu)\epsilon \\ &\leq \tfrac{3}{2}\mu^{i+1}\epsilon \leq \eta_s\mu^{\frac{5+i}{6}} = \eta_{k+i}, \end{aligned}$$

as $\mu^{\frac{1+5i}{6}} \leq \frac{2}{3}\eta_s/\epsilon$. Thus, x_{k+i} satisfies (A.5) and (A.7), and hence step 3 of the algorithm will be executed. Therefore, in particular, $\omega_{k+i+1} = \omega_s\mu^{i+2}$, $\eta_{k+i+1} = \eta_s\mu^{1+\frac{i}{6}}$ and $\lambda_{k+i+1} = (1-\mu^{i+1})\lambda_{*,1}$.

3. Let

$$\text{(C.82)}\quad x_{k+j-1} = -2\sqrt{1-\tfrac{1}{4}\mu^{j-1}s_{k+j-1}},$$

where the shift $s_{k+j-1} = \mu\sqrt{\frac{3}{2}(1-\mu^{j-1})\epsilon}$. Once again, (C.64) and (C.68) imply that $s_{k+j-1} \leq \mu$, and thus (C.82) is well defined. Furthermore, it is easy to verify that $\bar\lambda_{k+j-1} = \lambda_{*,1}$. Moreover

$$\text{(C.83)}\quad \begin{aligned} \nabla_x\Psi(x_{k+j-1},\lambda_{k+j-1},s_{k+j-1}) &= 2\epsilon(x_{k+j-1}-1) - 3\epsilon x_{k+j-1} \\ &= -\epsilon(2+x_{k+j-1}) \\ &= -2\epsilon\left(1-\sqrt{1-\tfrac{1}{4}\mu^{j-1}s_{k+j-1}}\right). \end{aligned}$$

But then (C.67), (C.83) and the inequality $s_{k+j-1} \leq \mu$ yield

$$\text{(C.84)}\quad \|P(x_{k+j-1},\nabla_x\Psi(x_{k+j-1},\lambda_{k+j-1},s_{k+j-1}))\| \leq \tfrac{1}{2}\epsilon\mu^{j-1}s_{k+j-1} \leq \tfrac{1}{2}\epsilon\mu^{j}.$$

Thus, (C.84) and (C.64) give

$$\text{(C.85)}\quad \|P(x_{k+j-1},\nabla_x\Psi(x_{k+j-1},\lambda_{k+j-1},s_{k+j-1}))\| \leq \omega_s\mu^{j} = \omega_{k+j-1},$$

as $1 \leq \omega_s/((6+1/(1-\mu_0))\epsilon) < 2\omega_s/\epsilon$. Furthermore, from (C.61) and (C.64),

$$\text{(C.86)}\quad \begin{aligned} \|c(x_{k+j-1})\bar\lambda_{k+j-1}/\sqrt{\lambda_{k+j-1}}\| &= \mu\|\lambda_{k+j-1,i}-\bar\lambda_{k+j-1,i}\| = \tfrac{3}{2}\mu^{j}\epsilon \\ &\leq \eta_s\mu^{\frac{4+j}{6}} = \eta_{k+j-1}, \end{aligned}$$

as $\mu^{\frac{5j-4}{6}} \leq \frac{2}{3}\eta_s/\epsilon$. Thus, x_{k+j-1} satisfies (A.5) and (A.7), and hence step 3 of the algorithm will be executed. Therefore, in particular, $\omega_{k+j} = \omega_s\mu^{1+j}$, $\eta_{k+j} = \eta_s\mu^{\frac{5+j}{6}}$ and $\lambda_{k+j} =)\lambda_{*,1}$.

4. We pick x_{k+j} as the largest root of the nonlinear equation

$$\text{(C.87)}\quad \phi(x) \equiv 2(x-1) - \frac{3xs_{k+j}}{x^2-4+s_{k+j}} = 0,$$

where $s_{k+j} = \mu\sqrt{\frac{3}{2}\epsilon}$. Equation (C.87) defines the stationary points of the Lagrangian barrier function for the problem (C.63). This choice ensures that (A.5) is trivially satisfied. As $\phi(2) = -4$ and $\phi(x)$ increases without bound as x tends to infinity, the largest root of (C.87) is greater than 2. The function $\bar\lambda$ given by (2.4) is a decreasing function of x as x grows beyond 2. Now let $\hat x = \sqrt{4+\frac{1}{2}s_{k+j}}$. It is easy to show that $\bar\lambda(\hat x,\lambda_{*,1},s_{k+j}) = \epsilon$. Moreover, we get $\phi(\hat x) = 2(\hat x-1) - 2\hat x = -2$. Therefore, $x_{k+j} > \hat x$, and thus $\bar\lambda_{k+j} < \epsilon$. But then, using (C.61), we have that

$$\text{(C.88)}\quad \begin{aligned} \|c(x_{k+j})\bar\lambda_{k+j}/\sqrt{\lambda_{k+j}}\| &= \mu(\lambda_{k+j}-\bar\lambda_{k+j}) \geq \mu(\tfrac{3}{2}\epsilon-\epsilon) = \tfrac{1}{2}\epsilon\mu \\ &> \eta_s\mu^{\frac{5+j}{6}} = \eta_{k+j} \end{aligned}$$

from (C.65). Thus, the test (A.7) is violated and the penalty parameter subsequently reduced. This ensures that $\omega_{k+j+1} = \omega_s\mu$, $\eta_{k+j+1} = \eta_s\mu$ and $\lambda_{k+j+1} = \lambda_{*,1}$.

Hence, a cycle as described at the start of this section is possible and we conclude that, in the absence of AS6, the penalty parameter generated by Algorithm A.1 may indeed converge to zero. □

C.5. Proof of Theorem 5.5. First, as (A.7) holds for all $k \geq k_0$, the penalty parameter μ_k remains fixed at some value μ_k, say, the convergence tolerances satisfy

$$\text{(C.89)}\quad \omega_{k+1} = \omega_k\mu_k \quad\text{and}\quad \eta_{k+1} = \eta_k\mu_k^{\frac{1}{6}},$$

and $\lambda_{k+1} = \bar\lambda_k$ for all $k > k_0$.

The Q-superlinear convergence of the Lagrange multiplier estimates for inactive constraints follows directly from Theorem 4.2, part (iv). Lemma 5.2, part (ii), the convergence of θ_k to zero and the relationships (4.1) and (4.8) then give that

$$\text{(C.90)}\quad \|\lambda_{k+1,\mathcal{I}^*}\| \leq \mu_k\|\lambda_{k,\mathcal{I}^*}\|$$

for all k sufficiently large.

The identities (2.5), (A.3) and the assumption that (A.7) holds for all $k \geq k_0$ gives

$$\text{(C.91)}\quad \begin{aligned} \|(\lambda_{k+1}-\lambda_k)_{\mathcal{A}^*}\| &= \mu_k^{-1}\left\|[c_{k,i}\bar\lambda_{k,i}/\sqrt{\lambda_{k,i}}]_{i\in\mathcal{A}^*}\right\| \\ &\leq \mu_k^{-1}\left\|[c_{k,i}\bar\lambda_{k,i}/\sqrt{\lambda_{k,i}}]_{i=1}^{m}\right\| \leq \mu_k^{-1}\eta_k \end{aligned}$$

for all such k. But then the triangle inequality and (C.91) imply that

$$\text{(C.92)}\quad \begin{aligned} \|(\lambda_{k+j}-\lambda_*)_{\mathcal{A}^*}\| &\leq \|(\lambda_{k+j+1}-\lambda_*)_{\mathcal{A}^*}\| + \|(\lambda_{k+j+1}-\lambda_{k+j})_{\mathcal{A}^*}\| \\ &\leq \|(\lambda_{k+j+1}-\lambda_*)_{\mathcal{A}^*}\| + \mu_k^{-1}\eta_{k+j} \end{aligned}$$

for all $k \geq k_0$. Thus, summing (C.92) from $j = 0$ to $j_{\max}-1$ and using the relationship (C.89) yields

$$\text{(C.93)}\quad \begin{aligned} \|(\lambda_k-\lambda_*)_{\mathcal{A}^*}\| &\leq \|(\lambda_{k+j_{\max}}-\lambda_*)_{\mathcal{A}^*}\| + \mu_k^{-1}\textstyle\sum_{i=0}^{j_{\max}-1}\eta_{k+j} \\ &\leq \|(\lambda_{k+j_{\max}}-\lambda_*)_{\mathcal{A}^*}\| + \mu_k^{-1}\eta_k(1-\mu_k^{\frac{1}{6}j_{\max}})/(1-\mu_k). \end{aligned}$$

Hence, letting $j_{\max}$ tend to infinity and recalling that λ_k converges to λ_*, we see that (C.93) gives

$$\text{(C.94)}\quad \|(\lambda_k-\lambda_*)_{\mathcal{A}^*}\| \leq \frac{\mu_k^{-1}\eta_k}{1-\mu_k}$$

for all $k \geq k_0$. As η_k converges to zero R-linearly, with R-factor $\mu_k^{\frac{1}{6}}$, (C.94) gives the required result (ii).

The remainder of the proof parallels that of Lemma 5.1. As (A.7) holds for all sufficiently large k, the definition (C.4) of $\mathcal{A}$ and the bound (C.8) ensure that

$$\text{(C.95)}\quad \begin{aligned} \|c_{k,\mathcal{A}}\| &\leq \left\|[\sqrt{\lambda_{k,i}}/\bar\lambda_{k,i}]_{i\in\mathcal{A}}\right\| \left\|[c_{k,i}\bar\lambda_{k,i}/\sqrt{\lambda_{k,i}}]_{i\in\mathcal{A}}\right\| \\ &\leq a_{14}\mu_k^{\chi-1}\left\|[c_{k,i}\bar\lambda_{k,i}/\sqrt{\lambda_{k,i}}]_{i=1}^{m}\right\| \leq a_{14}\mu_k^{\chi-1}\eta_k. \end{aligned}$$

Thus, combining (C.25) and (C.26), (C.90) and replacing (C.12) by (C.95), we may replace the bound on the right-hand side of (C.30) by $a_{23}\omega_k + a_{24}\sigma_k\|\lambda_{k,\mathcal{I}^\bullet}\| + a_{25}\mu_k^{1-\chi}\left\|[\sqrt{\lambda_{k,i}}]_{i\in\mathcal{A}_2^\bullet}\right\| + a_{14}\mu_k^{\chi-1}\eta_k$, and consequently (C.32) by

$$
\begin{aligned}
\Delta x_k \le\ & M[a_{19}(\Delta x_k)^2 + a_{20}\Delta x_k\omega_k + a_{21}\omega_k^2 + \\
& a_{22}\sigma_k\|\lambda_{k,\mathcal{I}^\bullet}\|(\omega_k + \Delta x_k) + a_{23}\omega_k + a_{24}\sigma_k\|\lambda_{k,\mathcal{I}^\bullet}\| + \\
& a_{25}\mu_k^{1-\chi}\left\|[\sqrt{\lambda_{k,i}}]_{i\in\mathcal{A}_2^\bullet}\right\| + a_{14}\mu_k^{\chi-1}\eta_k] \\
=\ & (Ma_{19}\Delta x_k + Ma_{20}\omega_k)\Delta x_k + (Ma_{21}\omega_k + Ma_{23})\omega_k + \\
& (Ma_{24} + Ma_{22}(\omega_k + \Delta x_k))\sigma_k\|\lambda_{k,\mathcal{I}^\bullet}\| \\
& Ma_{25}\mu_k^{1-\chi}\left\|[\sqrt{\lambda_{k,i}}]_{i\in\mathcal{A}_2^\bullet}\right\| + Ma_{14}\mu_k^{\chi-1}\eta_k.
\end{aligned} \tag{C.96}
$$

Hence, if k is sufficiently large that

$$\Delta x_k \le 1/(4Ma_{19}), \quad \omega_k \le \min(1, 1/(4Ma_{20})) \quad \text{and} \quad \sigma_k \le 1, \tag{C.97}$$

(C.96) and (C.97) can be rearranged to give

$$
\begin{aligned}
\Delta x_k \le\ & 2M[(a_{21} + a_{23})\omega_k + (a_{24} + 2a_{22})\|\lambda_{k,\mathcal{I}^\bullet}\| + \\
& a_{25}\mu_k^{1-\chi}\left\|[\sqrt{\lambda_{k,i}}]_{i\in\mathcal{A}_2^\bullet}\right\| + a_{14}\mu_k^{\chi-1}\eta_k].
\end{aligned} \tag{C.98}
$$

But then (C.27) and (C.98) give

$$
\begin{aligned}
\|x_k - x_*\| \le\ & a_{32}\omega_k + a_{33}\|\lambda_{k,\mathcal{I}^\bullet}\| \\
& + a_{34}\mu_k^{\chi-1}\eta_k + a_{35}\mu_k^{1-\chi}\left\|[\sqrt{\lambda_{k,i}}]_{i\in\mathcal{A}_2^\bullet}\right\|,
\end{aligned} \tag{C.99}
$$

where $a_{32} \stackrel{\text{def}}{=} 1 + 2M(a_{21} + a_{23})$, $a_{33} \stackrel{\text{def}}{=} 2M(a_{24} + 2a_{22})$, $a_{34} \stackrel{\text{def}}{=} 2Ma_{14}$ and $a_{35} \stackrel{\text{def}}{=} 2Ma_{25}$. Each term on the right-hand-side of (C.99) converges at least R-linearly to zero; the R-factors (in order) being no larger than μ_k, μ_k, $\mu_k^{\frac{1}{6}}$ and $\mu_k^{\frac{1}{12}}$, respectively, following (C.89), (C.90) and (C.94). Hence, (C.99) shows that x_k converges at least R-linearly with R-factor at most $\mu_k^{\frac{1}{12}}$.

C.6. Proof of Corollary 5.6. This follows directly from Theorem C.1 (§ C.3), as this ensures that (A.7) is satisfied for all k sufficiently large.