Appendix 4a Proofs for Part 4a

Proof of Lemma 4.a.1:  Note that if $\delta_n$ stays away from zero, $\|\Theta - f(\beta^{**})\| = O_p(1)$. But

\[
\|\Theta - f(\beta^{**})\| = \|\hat{g} - (\hat{g} - (\Theta - f(\beta^{**}))\|
\]

\[
\leq \|\hat{g}\| + \|\hat{g} - (\Theta - f(\beta^{**}))\|
\]

\[
= \|\hat{g}\| + O_p(\gamma_n)
\]

by assumption A2, so that

\[
\|\hat{g}\| = O_p(1)
\]

If $\delta_n = 0$ then by assumption A2 $\|\hat{g}\|$

\[
= \|\hat{g} - (\Theta - f(\beta^{**}))\| = O_p(\gamma_n)
\]

as desired.//.

Proof of Lemma 4.a.2:  First observe that $\lambda^*$ (given by 4.A.1) becomes

\[
\frac{<\hat{g}, \Theta - f> + <\hat{g}, f - \hat{f}>}{\|\hat{g}\|^2}
\]

\[
\leq \frac{\|\hat{g}\| \|\Theta - f\| + \|\hat{g}\| \|f - \hat{f}\|}{\|\hat{g}\|^2}
\]

(by the Cauchy-Schwarz Inequality)

\[
= \frac{\|\Theta - f\| + \|f - \hat{f}\|}{\|\hat{g}\|}
\]

(A4.A.1)

(We will often use $f$ to designate $f(\beta^{**})$). If $\delta_n = 0$, then by Lemma 4.a.1, and 4.1, A4.A.1

\[
= \frac{0 + O_p(n^{-5})}{O_p(\gamma_n)} = O_p(n^{-5}\gamma_n^{-1}).
\]
If $\delta_n$ stays away from zero we want to see how quickly $\lambda^*$ approaches 1. To that end

$$1 - \lambda^* = \frac{\|g\|^2 - \langle \hat{g}, \Theta - \hat{f} \rangle}{\|\hat{g}\|^2}$$

$$= \frac{\langle \hat{g}, \hat{g} \rangle - \langle \hat{g}, \Theta - \hat{f} \rangle}{\|\hat{g}\|^2}$$

$$= \frac{\langle \hat{g}, \hat{g} - \Theta + \hat{f} \rangle}{\|\hat{g}\|^2}$$

$$\leq \frac{\|\hat{g}\| \|\hat{g} - (\Theta - f)\| + \|\hat{g}\| \|\hat{f} - f\|}{\|\hat{g}\|^2}$$

(by the Triangle and Cauchy-Schwarz Inequalities)

$$= \frac{\|\hat{g} - (\Theta - f)\| + \|\hat{f} - f\|}{\|\hat{g}\|}$$

(A4.A.2)

So by Lemma 4.1.a, equation 4.1, and assumption A2, A4.A.2

$$= \frac{O_p(Y_n) + O_p(n^{-5})}{O_p(1)}$$

$$= O_p(Y_n),$$

as desired.//.

**Proof of Lemma 4.1.a:** Consider $\hat{\lambda}^* - \lambda^*$

$$= \frac{\langle \hat{g}, Y - \hat{f} \rangle}{\|\hat{g}\|^2} - \frac{\langle \hat{g}, \Theta - \hat{f} \rangle}{\|\hat{g}\|^2} = \frac{\langle \hat{g}, \epsilon \rangle}{\|\hat{g}\|^2}$$

Note that $\langle \hat{g}, \epsilon \rangle = n^{-1} \sum_{i=1}^{n} \hat{g}(x_i) \epsilon_i = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} W_2(x_i, x_j)(Y_j - \hat{f}_j) \epsilon_i$

$$= n^{-2} \sum_{i=1}^{n} \epsilon_i \sum_{j=1}^{n} W_2(x_i, x_j)(\Theta_j + \epsilon_j - \hat{f}_j)$$

74
\[
\begin{align*}
&= n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \epsilon_i W_2(x_i, x_j)(\Theta_j - f_j) + n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} W_2(x_i, x_j) \epsilon_i \epsilon_j - n^{-2} \sum_{i=1}^{n} \epsilon_i \sum_{j=1}^{n} W_2(x_i, x_j)(\hat{f}_j - f_j) \\
&= T_1 + T_2 - T_3 \text{ (say).}
\end{align*}
\]

Note that \( E(T_1) = 0 \). Hence, we turn our attention to \( V(T_1) = E(T_1^2) \)

\[
\leq n^{-4} \sum_{i=1}^{n} c(\sum_{j=1}^{n} W_2(x_i, x_j)(\Theta_j - f_j))^2.
\]

(for some real positive constant \( c \), by Whittle’s Inequality (Whittle (1960)))

\[
\leq c_2 n^{-2} \delta_n^2 \tau_n
\]

for some positive constant \( c_2 \), by Assumption A4, the definition of \( \tilde{\delta}_n \), and the Cauchy-Schwarz Inequality. So \( V(T_1) \)

\[
= O_p \left( \gamma_n^2 n^{-1} \delta_n^2 \right)
\]

by Burman and Chaudhuri (1992) results 6.14, and 6.16. So that \( T_1 \)

\[
= O_p(n^{-5} \gamma \delta)
\]

Next observe that

\[
T_2 = n^{-2} \sum_{i \neq j} W_2(x_i, x_j) \epsilon_i \epsilon_j + n^{-2} \sum_{i} W_2(x_i, x_i) \epsilon_i^2
\]

so that \( E(T_2) \) approaches 0 asymptotically. Note that

\[
V(T_2) = O_p (n^{-3} \gamma_n^2)
\]

by Whittle’s Inequality (Whittle (1960)), Assumptions A3 and A4, and the definition of \( \epsilon_i \). Then

\[
T_2 = O_p (n^{-5} \gamma)
\]

(note also that this result is confirmed by Burman and Chaudhuri (1992) result 6.14).
\[ T_3 = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \sum_{k=1}^{n} W_2(x_i, x_j) (\hat{f}_j - f_j) \]

\[ = n^{-2} \sum_{i=1}^{n} \varepsilon_i \sum_{j=1}^{n} W_2(x_i, x_j) (n^{-1} \sum_{k=1}^{n} W_1(x_j, x_k) \varepsilon_k + O_p(n^{-1})) \]

(by Assumption A1)

\[ = n^{-2} \sum_{i=1}^{n} \varepsilon_i (\sum_{j=1}^{n} W_2(x_i, x_j) (n^{-1} \sum_{k=1}^{n} W_1(x_j, x_k) \varepsilon_k) + \sum_{j=1}^{n} W_2(x_i, x_j) (O_p(n^{-1}))) \]

\[ = n^{-2} \sum_{i=1}^{n} \varepsilon_i \left( \sum_{j=1}^{n} \sum_{k=1}^{n} W_2(x_i, x_j) W_1(x_j, x_k) \varepsilon_k \right) + n^{-2} \sum_{i=1}^{n} \varepsilon_i (\sum_{j=1}^{n} W_2(x_i, x_j) (O_p(n^{-1}))) \]

\[ = T_{31} + O_p(n^{-5/2}) \text{ (say)} \]

by noting the convergence rate of $T_1$. Next rewrite $T_{31}$ as

\[ n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \varepsilon_i \varepsilon_j \left( n^{-1} \sum_{k=1}^{n} W_2(x_i, x_j) W_1(x_j, x_k) \right) \]

and observe that by Whittle’s Inequality (Whittle (1960)), $V(T_{31})$

\[ \leq n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} c (n^{-1} \sum_{k=1}^{n} W_2(x_i, x_j) W_1(x_j, x_k))^2 \]

(for some real positive constant $c$)

\[ \leq n^{-2} c_2 \tau_n \]

(for some positive real constant $c_2$) by assumptions A1 and A4. So that $T_{31}$ (and $T_3$)

\[ = O_p(n^{-5/2}) \]


Then $T_1 + T_2 - T_3 = \langle \hat{g}, \varepsilon \rangle = O_p(n^{-5/2})$

so that

\[ \hat{\lambda}^* - \lambda^* = \frac{O_p(n^{-5/2})}{\| \hat{g} \|^2} \]

(A4.A.3)
\[
O_p(n^{-5} \gamma_n), \quad \text{if } \lim_{n \to \infty} \delta_n \neq 0
\]
\[
O_p(n^{-5} \gamma_n^{-1}), \quad \text{if } \delta_n = 0
\]

by Lemma 4.a.1, as desired.\\

**Proof of Lemma 4.a.4:** We will prove parts a, b and c in order. First note that
\[
\| \hat{g} \| = \| \hat{g} + (\Theta - f) - (\Theta - f) \|
\]
\[
\leq \| \hat{g} - (\Theta - f) \| + \| (\Theta - f) \|
\]
\[
= O_p(\gamma_n) + \delta_n
\]
by the Triangle Inequality and Assumption A2. Hence, \( \| \hat{g} \| \) takes the larger of \( O_p(\gamma_n) \) or \( O_p(\delta_n) \) as \( n \) approaches \( \infty \). Next observe that \( \lambda^* = \frac{< \hat{g}, \Theta - \hat{f} >}{\| \hat{g} \|^2} \)
\[
= \frac{< \hat{g}, \Theta - f >}{\| \hat{g} \|^2} + \frac{< \hat{g}, f - \hat{f} >}{\| \hat{g} \|^2}
\]
\[
\leq \frac{\| \Theta - f \|}{\| \hat{g} \|} + \frac{\| f - \hat{f} \|}{\| \hat{g} \|}
\]
\[
= \frac{\delta_n}{\| \hat{g} \|} + \frac{O_p(n^{-5})}{\| \hat{g} \|} \quad (A4.A.4)
\]
by the Cauchy-Schwartz inequality, Assumption A1, and 4.1. By simply observing the form of \( \lambda^* \) above it is clear from Assumption A2 that \( \lambda^* \) approaches 1 at the rate of \( O_p(\gamma_n) \). Using Lemma 4.a.4 (part a), we see that term A4.A.4
\[
\begin{cases}
  O_p(\delta_n \gamma_n^{-1}), \quad \text{if } \frac{n^{-5}}{\gamma_n} < \frac{\delta_n}{\gamma_n} < 1 \\
  O_p(n^{-5} \gamma_n^{-1}), \quad \text{if } \frac{\delta_n}{\gamma_n} < \frac{n^{-5}}{\gamma_n} < 1
\end{cases}
\]
Finally, combining A4.A.3 with Lemma 4.a.4 (part a) it is clear that $\hat{\lambda}^* - \lambda^* =$

$$
\begin{cases}
O_p(n^{-5} \gamma_n \delta_n^{-2}), & \text{if } \frac{\delta_n}{\gamma_n} > 1 \\
O_p(n^{-5} \gamma_n^{-1}), & \text{if } \frac{\delta_n}{\gamma_n} < 1
\end{cases}
$$

as desired.//.

**Proof of Lemma 4.a.5**: Note that Lemmas 4.a.1, 4.a.3 and 4.a.4 together imply that

$$
|\hat{\lambda}^* - \lambda^*| \|\hat{g}\| = O_p(n^{-5})
$$

in not only either of the cases in which $\delta_n = 0$, or $\lim_{n \to \infty} \delta_n \neq 0$ but in any of the situations in which $\lim_{n \to \infty} \delta_n = 0$ as well.//.

**Proof of Theorem 4.A.1**: Observe that

$$
\|\hat{\lambda}^* + \hat{f} - \Theta\| = \|\hat{\lambda}^* - (\Theta - f) + (\hat{f} - f)\|
$$

$$
= \|\hat{\lambda}^* (\hat{g} - (\Theta - f)) + (1 - \lambda^*)(\Theta - f) + (\hat{f} - f)\|
$$

$$
\leq \|\hat{\lambda}^*\| \|\hat{g} - (\Theta - f)\| + |1 - \lambda^*| \|\Theta - f\| + \|\hat{f} - f\| \quad (A4.A.5)
$$

by the Triangle inequality. We may easily assert by Assumptions A1 and A2, Lemma 4.a.2 and equation 4.1, that A4.A.5

$$
= O_p(1) O_p(\gamma_n) + O_p(\gamma_n) O_p(1) + O_p(n^{-5})
$$

$$
= O_p(\gamma_n)
$$

if $\lim_{n \to \infty} \delta_n \neq 0$, and A4.A.5

$$
= O_p(n^{-5} \gamma_n^{-1}) O_p(\gamma_n) + 0 + O_p(n^{-5})
$$

$$
= O_p(n^{-5}),
$$

if $\delta_n = 0$, as desired.//.
Proof of Theorem 4.A.2: As in Part 1a, observe that
\[ \left\| \hat{\lambda}^* \hat{g} + \hat{f} - \Theta \right\|^2 - \left\| \lambda^* \hat{g} + \hat{f} - \Theta \right\|^2 \]
\[ = n^{-1} \sum_{i=1}^{n} \left( (\hat{\lambda}^* \hat{g} + \hat{f} - \Theta)^2 - (\lambda^* \hat{g} + \hat{f} - \Theta)^2 \right) \]
\[ = n^{-1} \sum_{i=1}^{n} \left( (\hat{\lambda}^* - \lambda^*) \hat{g}^2 + 2(\lambda^* - \lambda^*) (\hat{g} \hat{f} - g \Theta) \right) \]
\[ = n^{-1} \sum_{i=1}^{n} \left( (\hat{\lambda}^* - \lambda^*) \hat{g} \left( (\hat{\lambda}^* + \lambda^*) \hat{g} + 2(\hat{f} - \Theta) \right) \right) \]
\[ = n^{-1} \sum_{i=1}^{n} \left( (\hat{\lambda}^* - \lambda^*) \hat{g} \left( (\hat{\lambda}^* \hat{g} + \hat{f} - \Theta) + (\lambda^* \hat{g} + \hat{f} - \Theta) \right) \right) \]
\[ \leq \left\| \hat{\lambda}^* - \lambda^* \right\| \left\| \hat{g} \right\| \left\| \hat{\lambda}^* \hat{g} + \hat{f} - \Theta \right\| + \left\| \hat{\lambda}^* - \lambda^* \right\| \left\| \lambda^* \hat{g} + \hat{f} - \Theta \right\| \]  
(A4.6)

by the Cauchy-Schwartz inequality. Next rewrite A4.6 as
\[ T3^2 - T5^2 \leq T4*T3 + T4*T5 \] (say).

Then
\[ T3^2 \leq T4*T3 + T4*T5 + T5^2 \]  
(A4.7)

Lemma 4.a.5 indicates that \( T4 = \left\| \hat{\lambda}^* - \lambda^* \right\| \left\| \hat{g} \right\| \) converges at a rate of \( O_p(n^{-5}) \), which is at least as fast as \( \left\| \lambda^* \hat{g} + \hat{f} - \Theta \right\| \) in the two cases considered here. We will show that
\[ T5 = \left\| \hat{\lambda}^* \hat{g} + \hat{f} - \Theta \right\| \) and \( T3 = \left\| \hat{\lambda}^* \hat{g} + \hat{f} - \Theta \right\| \) converge at exactly the same rate.

First, suppose that \( T3 \) converges at a slower rate than \( T5 \). By Lemma 4.a.5, A4.7 becomes
\[ T3^2 \leq k*T4*T3 + k*T5^2 \]
(where \( k \in \mathbb{R} \)). Then
\[ T3 \leq k^{\frac{1}{2}}*(T4 + T5) \]
which implies that \( T3 \) converges at a rate at least as fast as \( T5 \). This is a contradiction (#).
Next, suppose that \( T5 \) converges at a slower rate than \( T3 \). This is a somewhat absurd assumption, nevertheless we will proceed as planned. Notice that we may rewrite A4.A.6 as

\[
T5^2 - T3^2 \leq T4*T3 + T4*T5
\]

so that

\[
T5^2 \leq T4*T3 + T4*T5 + T3^2
\]

and

\[
T5 \leq T4^5*(T3^5 + T5^5) + T3
\]

\[
\leq T4^5*(k*T5^5) + T3
\]

again for some \( k \in \mathbb{R} \), by our assumption. Then \( T5 \) converges at a rate at least as fast as \((T4*T5)^5 \) (which is a \# based on Lemma 4.a.5, assumption A5, and Theorem 4.A.1.a), or \( T5 \) converges at a rate at least as fast as \( T3 \) (which is a \# of the assumption of this part). \#. So \( T5 = \|\lambda^\star \hat{g} + \hat{f} - \Theta\| \) and \( T3 = \|\lambda^\star \hat{g} + \hat{f} - \Theta\| \) converge at exactly the same rate for the cases \( \delta_n = 0 \) or \( \lim_{n \to \infty} \delta_n \neq 0 \), as desired.\/. 

**Proof of Theorem 4.A.3**: Recall that A4.A.5 had that \( \|\lambda^\star \hat{g} + \hat{f} - \Theta\| \)

\[
\leq |\lambda^\star|\|\hat{g} - (\Theta - f)\| + |1 - \lambda^\star|\|\Theta - f\| + \|\hat{f} - f\|. 
\]

The result follows as a consequence of combining this result with Lemma 4.a.4, as desired.\/. 

**Proof of Theorem 4.A.4**: Note that in the proof of Theorem 4.A.2 we proved that

asymptotically \( \|\lambda^\star \hat{g} + \hat{f} - \Theta\| = \|\lambda^\star \hat{g} + \hat{f} - \Theta\| \). The proof of that theorem was dependent upon the fact that \( \|\hat{g}\| \) converges at a rate of \( O_p(n^{-5}) \); at least as fast as

\( \|\lambda^\star \hat{g} + \hat{f} - \Theta\| \). Observe that the same holds true when \( \lim_{n \to \infty} \delta_n = 0 \), by Theorem 4.A.3.

80
Then Theorem 4.A.4 is proven identically with the exception that the second false supposition will use Lemma 4.a.5, assumption A5 and Theorem 4.A.3.a) for the first #. So, again

\[ \| \hat{\lambda} \ast \hat{g} + \hat{f} - \Theta \| = \| \lambda \ast \hat{g} + \hat{f} - \Theta \| \]

asymptotically when \( \lim_{n \to \infty} \delta_n = 0 \), as desired.///.