Online Appendix to: Decoupling Noise and Features via Weighted ℓ_1-Analysis Compressed Sensing

RUIMIN WANG, ZHOUWANG YANG, LIGANG LIU, JIANSONG DENG, and FALAI CHEN
University of Science and Technology of China

In this appendix, we will prove the convergence rate and asymptotic optimality for the DLRS estimator, based on the asymptotic behavior of eigenvalues of matrix M and statistical theory.

Let Ω be a bounded 2D manifold (the domain) and $\mathscr{H}^2(\Omega)$ the space of C^2-continuous functions defined on Ω. A semi-norm in $\mathscr{H}^2(\Omega)$ is defined by

$$|f|_{\Omega,2}^2 = \int_\Omega (\Delta_\Omega f)^2.$$

With a set of sampling points $\Pi = \{X_i\}_{i=1}^n$ in the domain Ω, we can also give a discrete version of the aforementioned semi-norm as

$$|f|_{\Pi,2}^2 = \sum_{i=1}^n |\Delta_\Omega f(X_i)|^2.$$

Specially, $|f|_{\Omega,0}^2 = \int_\Omega f^2$ and $|f|_{\Pi,0}^2 = \frac{1}{n}\sum_{i=1}^n f(X_i)^2$.

We now have a few assumptions as follows.

(A.1) The input Ω is a bounded Lipschitz domain satisfying the uniform cone conditions. See Utreras [1988] for detailed definition.

(A.2) The set of sampling points $\Pi = \{X_i\}_{i=1}^n$ in domain Ω satisfies the following quasi-uniform assumption: there exists a constant $\xi_0 > 0$ such that

$$\frac{\delta_{\max}}{\delta_{\min}} \le \xi_0,$$

where $\delta_{\max} = \sup_{X\in\Omega}\inf_{X_i\in\Pi}\|X - X_i\|$ and $\delta_{\min} = \min_{j\neq i}\|X_j - X_i\|$.

(A.3) Given $\Pi = \{X_i\}_{i=1}^n \subset \Omega$, there exist constants $\underline{\xi}$ and $\overline{\xi}$ (depending on Π) such that

$$\underline{\xi}|f|_{\Omega,2}^2 \le |f|_{\Pi,2}^2 \le \overline{\xi}|f|_{\Omega,2}^2$$

for any function $f \in \mathscr{H}^2(\Omega)$.

Remark 1. Suppose $\Pi = \{X_i\}_{i=1}^n$ is an equidistributed sequence in the region Ω. From the law of large numbers, we have

$$\lim_{n\to\infty} |f|_{\Pi,2}^2 = \frac{1}{\text{Area}(\Omega)}|f|_{\Omega,2}^2.$$

Since Ω is bounded, Area(Ω) is also bounded. Thus (A.3) is satisfied with probability one as the sample size goes to infinity.

DOI: http://dx.doi.org/10.1145/2557449

1. PROOF OF THEOREM 1

Before we prove Theorem 1, we have some propositions.

PROPOSITION 1.1. *For any $f \in \mathscr{H}^2(\Omega)$, there exists a matrix $M_{\Pi,2}$ (depending on Π) such that*

$$|f|_{\Pi,2}^2 = \min_{\substack{\phi\in\mathscr{H}^2(\Omega)\\ \phi(X_i)=f_i,\, i=1,\dots,n}} \frac{1}{n}\mathbf{f}^T M_{\Pi,2}\mathbf{f}, \qquad (1)$$

where $\mathbf{f} = (f_1,\dots,f_n)^T = (f(X_1),\dots,f(X_n))^T$ is the vector of function values at $\Pi = \{X_i\}_{i=1}^n$.

The proof of the preceding proposition can be found in textbook Halmos [1982] using the Riesz representation theorem and thus the details are omitted.

PROPOSITION 1.2. *If Ω is a bounded 2D manifold and μ_n is the largest eigenvalue of matrix $M_{\Pi,2}$, then $n\delta_{\max}^2$ and $\delta_{\max}^4\mu_n$ are both bounded from above.*

PROOF. Suppose that V_{unit} is the area of unit geodestic disk on Ω. So we have

$$nV_{\text{unit}}\delta_{\min}^2 \le \text{Area}(\Omega),$$

and then get

$$\delta_{\max}^2 \le n^{-1}\frac{\text{Area}(\Omega)}{V_{\text{unit}}}\frac{\delta_{\max}^2}{\delta_{\min}^2} = n^{-1}\frac{\text{Area}(\Omega)}{V_{\text{unit}}}\xi_0^2 = O(n^{-1}). \qquad (2)$$

So $n\delta_{\max}^2$ is bounded from above.

Let u be the function such that

$$\frac{1}{n}\mathbf{u}^T M_{\Pi,2}\mathbf{u} = |u|_{\Pi,2}^2 = \min_{\substack{\phi\in\mathscr{H}^2(\Omega)\\ \phi(X_i)=u_i,\, i=1,\dots,n}} |\phi|_{\Pi,2}^2,$$

where $\mathbf{u} = (u_1,\dots,u_n)^T$ is the eigenvector of $M_{\Pi,2}$ corresponding to the largest eigenvalue, that is, $M_{\Pi,2}\mathbf{u} = \mu_n\mathbf{u}$. We define a compactly supported radial basis function

$$w(s) = \begin{cases} e^{-\|\mathbf{s}\|/(1-\|\mathbf{s}\|)}, & 0 \le \|\mathbf{s}\| \le 1 \\ 0, & \|\mathbf{s}\| > 1 \end{cases}$$

and specify an interpolant $\phi(X) = \sum_{i=1}^n u_i w_i(X)$, where $w_i(X) = w(\frac{X-X_i}{\delta_{\min}})$. By the definition of $\delta_{\min}$, it is easy to see that $\phi(X_i) = u_i, i = 1,\dots,n$. Moreover, we have for $\beta \in \mathbb{Z}_+^3$

$$D^\beta w_i(X_j) = 0, \quad \forall i \neq j$$

and with $|\beta| = 2$

$$D^\beta w_j(X_j) = \delta_{\min}^{-2} D^\beta w(\mathbf{0}).$$

Hence, we have

$$\begin{aligned}|u|_{\Pi,2}^2 &\le |\phi|_{\Pi,2}^2\\ &= \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{|\beta|=2}\frac{2!}{\beta!}|D^\beta\phi(X_j)|^2\right)\\ &= \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{|\beta|=2}\frac{2}{\beta!}\left|\sum_{i=1}^{n}u_i D^\beta w_i(X_j)\right|^2\right)\\ &= \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{|\beta|=2}\frac{2}{\beta!}u_j^2|D^\beta w_j(X_j)|^2\right)\\ &= \frac{1}{n}\sum_{j=1}^{n}u_j^2\left(\sum_{|\beta|=2}\frac{2}{\beta!}|D^\beta w(\mathbf{0})|^2\right)\delta_{\min}^{-4},\end{aligned}$$

which implies that $\mu_n \le c(w)\delta_{\min}^{-4}$ by denoting the constant $c(w) = \sum_{|\beta|=2}\frac{2}{\beta!}|D^\beta w(\mathbf{0})|^2$. Finally we get

$$\delta_{\max}^4\mu_n \le c(w)\frac{\delta_{\max}^4}{\delta_{\min}^4} = c(w)\xi_0^4$$

and prove that $\delta_{\max}^4\mu_n$ is bounded from above. □

PROPOSITION 1.3. *Suppose that $\xi_1 j^m \le \mu_j \le \xi_2 j^m$ for $m > 0$ and $j = 1, 2, \ldots$, where $\xi_1, \xi_2 > 0$ are constants. Then we have for $n > 0, \lambda > 0$,*

$$\sum_{j=1}^{n}\frac{1}{(1+\lambda\mu_j)^2} = O(\lambda^{-1/m}).$$

PROOF. First of all we have

$$\sum_{j=1}^{n}\frac{1}{(1+\lambda\xi_2 j^m)^2} \le \sum_{j=1}^{n}\frac{1}{(1+\lambda\mu_j)^2} \le \sum_{j=1}^{n}\frac{1}{(1+\lambda\xi_1 j^m)^2}.$$

For $i = 1, 2$, we have

$$\begin{aligned}\sum_{j=1}^{n}\frac{1}{(1+\lambda\xi_i j^m)^2} &\ge \int_1^{n+1}\frac{1}{(1+\lambda\xi_i x^m)^2}dx\\ &= \frac{1}{m}\int_{\lambda\xi_i}^{\lambda\xi_i(n+1)^m}\frac{y^{-\frac{m-1}{m}}}{(1+y)^2}dy\cdot(\lambda\xi_i)^{-1/m}\\ &\to m^{-1}\left(\int_{\lambda\xi_i}^{\infty}\frac{y^{-(m-1)/m}}{(1+y)^2}dy\right)\xi_i^{-1/m}\cdot\lambda^{-1/m}\\ &= O(\lambda^{-1/m}),\end{aligned}$$

where the second equation reflects the change of variable ($y = \lambda\xi_i x^m$), and "$\to$" corresponds to "$n \to \infty$." Similarly, with the same change of variable, we also have

$$\begin{aligned}\sum_{j=1}^{n}\frac{1}{(1+\lambda\xi_i j^m)^2} &\le \int_0^{n}\frac{dx}{(1+\lambda\xi_i x^m)^2}\\ &= \frac{1}{m}\int_{0}^{\lambda\xi_i n^m}\frac{y^{-\frac{m-1}{m}}}{(1+y)^2}dy\cdot(\lambda\xi_i)^{-1/m}\\ &\to m^{-1}\left(\int_{\lambda\xi_i}^{\infty}\frac{y^{-(m-1)/m}}{(1+y)^2}dy\right)\xi_i^{-1/m}\cdot\lambda^{-1/m}\\ &= O(\lambda^{-1/m}).\quad\square\end{aligned}$$

We are now ready to exhibit the Rayleigh quotient inequalities connecting the semi-norms in $\mathscr{H}^2(\Omega)$ and their discretized version.

LEMMA 1.4. *Let Ω satisfy (A.1) and $f \ne 0$ satisfy (A.3). Then there exists constant $\gamma_1 > 0$ (depending only on $\Omega, \xi_0, \underline{\xi}$) and $\delta_0 > 0$, such that if $\delta_{\max} \le \delta_0$ we have*

$$\frac{|f|_{\Pi,2}^2}{|f|_{\Pi,0}^2} \ge \frac{|f|_{\Omega,2}^2}{\gamma_1\left(|f|_{\Omega,0}^2 + \delta_{\max}^4|f|_{\Omega,2}^2\right)},$$

for any $|f|_{\Pi,0}^2 \ne 0$.

PROOF. According to Theorem 3.3 in Utreras [1988], there exists constant $c(\Omega, \xi_0) > 0$ and $\delta_0 > 0$ such that for $\delta_{\max} \le \delta_0$,

$$|f|_{\Pi,0}^2 \le C(d, m, \Omega, \xi_0)\left(|f|_{\Omega,0}^2 + \delta_{\max}^4|f|_{\Omega,2}^2\right).$$

Since $|f|_{\Pi,2}^2 \ge \underline{\xi}|f|_{\Omega,2}^2$, we have

$$\begin{aligned}\frac{|f|_{\Pi,2}^2}{|f|_{\Pi,0}^2} &\ge \frac{\underline{\xi}|f|_{\Omega,2}^2}{c(\Omega,\xi_0)\left(|f|_{\Omega,0}^2 + \delta_{\max}^4|f|_{\Omega,2}^2\right)}\\ &\ge \frac{|f|_{\Omega,2}^2}{\gamma_1\left(|f|_{\Omega,0}^2 + \delta_{\max}^4|f|_{\Omega,2}^2\right)},\end{aligned}$$

where $\gamma_1 = c(\Omega, \xi_0)/\underline{\xi}$. □

LEMMA 1.5. *Assume the same conditions as in Lemma 1. Then there exists constant $\gamma_2 > 0$ (depending only on $\Omega, \xi_0, \underline{\xi}, \overline{\xi}$) and $\delta_0 > 0$, such that if $\delta_{\max} \le \delta_0$ we have*

$$\frac{|f|_{\Omega,2}^2}{|f|_{\Omega,0}^2} \ge \frac{|f|_{\Pi,2}^2}{\gamma_2\left(|f|_{\Pi,0}^2 + \delta_{\max}^4|f|_{\Pi,2}^2\right)}, \tag{3}$$

for any $0 \ne f$.

PROOF. According to Theorem 3.4 in Utreras [1988], there exists constant $c'(\Omega, \xi_0) > 0$ and $\delta_0 > 0$ such that for $\delta_{\max} \le \delta_0$,

$$|f|_{\Omega,0}^2 \le c'(\Omega, \xi_0)\left(|f|_{\Pi,0}^2 + \delta_{\max}^4|f|_{\Omega,2}^2\right).$$

Since $\underline{\xi}|f|_{\Omega,2}^2 \le |f|_{\Pi,2}^2 \le \overline{\xi}|f|_{\Omega,2}^2$, we have

$$\begin{aligned}\frac{|f|_{\Omega,2}^2}{|f|_{\Omega,0}^2} &\ge \frac{|f|_{\Pi,2}^2/\overline{\xi}}{c'(\Omega,\xi_0)\left(|f|_{\Omega,0}^2 + \delta_{\max}^4|f|_{\Pi,2}^2/\underline{\xi}\right)}\\ &\ge \frac{|f|_{\Pi,2}^2}{\gamma_2\left(|f|_{\Pi,0}^2 + \delta_{\max}^4|f|_{\Pi,2}^2\right)},\end{aligned}$$

where $\gamma_2 = c'(\Omega, \xi_0)\overline{\xi}\max(1, 1/\underline{\xi})$. □

Lemma 1.4 and Lemma 1.5 build a connection between the continuous semi-norms and discrete semi-norms. This enables us to study the behavior of the eigenvalues of $M_{\Pi,2}$ through studying the variational eigenvalue problem. Let $\mu_1 \le \cdots \le \mu_n$ be the eigenvalues of $M_{\Pi,2}$ in ascending order. Clearly $\{\mu_j\}$ are non-negative real numbers since the matrix $M_{\Pi,2}$ is semi-positive define. Next we study the behavior of these eigenvalues and show that they can be bounded by the discrete spectrum of the differential operator $(-\Delta_\Omega)^2$, where Δ_Ω is the Laplacian-Beltrami operator on Ω.

LEMMA 1.6. *Let Ω satisfy (A.1) and $\Pi = \{X_j\}_{j=1}^n$ satisfy (A.2). Then there exist constants $c_1, c_2 > 0$ such that*

$$c_1\rho_j \le \mu_j \le c_2\rho_j,$$

where $\rho_1 \le \rho_2 \le \cdots \le \rho_n$ are the first n eigenvalues of the variational eigenvalue problem

$$\int_\Omega \phi \Delta_\Omega^2 \psi = \rho \int_\Omega \phi\psi, \quad \forall\, \psi \in \mathscr{H}^2(\Omega).$$

PROOF. From Lemma 1.4 we get

$$\frac{|\phi|^2_{\Pi,2}}{|\phi|^2_{\Pi,0}} \ge \frac{|\phi|^2_{\Omega,2}}{\gamma_1\left(|\phi|^2_{\Omega,0} + \delta^4_{\max}|\phi|^2_{\Omega,2}\right)}$$

for any $\phi \in \mathscr{H}^2(\Omega)$ with $|\phi|^2_{\Pi,0} \neq 0$. Thus

$$\mu_j \ge \frac{1}{\gamma_1}\vartheta_j,$$

where $\vartheta_1 \le \cdots \le \vartheta_n$ are the first n eigenvalues of the variational eigenvalue problem

$$|\phi|^2_{\Omega,2} = \vartheta \cdot \left(|\phi|^2_{\Omega,0} + \delta^4_{\max}|\phi|^2_{\Omega,2}\right),$$

which implies

$$\vartheta_j = \frac{\rho_j}{1+\delta^4_{\max}\rho_j},\ j = 1,\ldots,n.$$

Note that $\delta^4_{\max}\rho_j$ is bounded from above, since $\rho_j \sim j^2$ according to Theorem 14.6 in Agmon [1965] and the fact $\delta^4_{\max} = O(n^{-2})$ from Eq. (2). So there exists $c_1 > 0$ such that $\frac{1}{\gamma_1(1+\delta^4_{\max}\rho_j)} \ge c_1$, then we have

$$\mu_j \ge c_1\rho_j.$$

On the other hand, using Lemma 1.5

$$\frac{|\phi|^2_{\Omega,2}}{|\phi|^2_{\Omega,0}} \ge \frac{|\phi|^2_{\Pi,2}}{\gamma_1\left(|\phi|^2_{\Pi,0} + \delta^4_{\max}|\phi|^2_{\Pi,2}\right)}$$

we have

$$\rho_j \ge \frac{1}{\gamma_2}\nu_j,$$

where $\nu_1 \le \cdots \le \nu_n$ are the first n eigenvalues of the variational eigenvalue problem

$$|\phi|^2_{\Pi,2} = \nu \cdot \left(|\phi|^2_{\Pi,0} + \delta^4_{\max}|\phi|^2_{\Pi,2}\right),$$

which gives

$$\nu_j = \frac{\mu_j}{1+\delta^4_{\max}\mu_j},\ j = 1,\ldots,n.$$

So there exists $c_2 > 0$ such that

$$\mu_j \le \gamma_2\left(1+\delta^4_{\max}\mu_j\right)\rho_j \le \gamma_2\left(1+\delta^4_{\max}\mu_n\right)\rho_j \le c_2\rho_j,$$

since $\delta^4_{\max}\mu_n$ is bounded according to Proposition 1.2. □

LEMMA 1.7. *Suppose Ω satisfy (A.1). Let $\{\mu_1 \le \cdots \le \mu_n\}$ be the eigenvalues of $M_{\Pi,2}$ in ascending order. Then there exist constants $c_3, c_4 > 0$ such that for $2 < j \le n$ we have*

$$c_3 j^2 \le \mu_j \le c_4 j^2. \tag{4}$$

PROOF. According to Lemma 1.6, it suffices to prove that the eigenvalues $\rho_1 \le \rho_2 \le \cdots$ satisfy the type of relationship in Eq. (4).

By using integration by parts, we observe that $\rho_1 \le \rho_2 \le \cdots$ are the eigenvalues of the differential operator $(-\Delta_\Omega)^2$ which has discrete spectrum contained in the non-negative real axis. We can then apply Theorem 14.6 in Agmon [965] to get

$$\rho_j \sim j^2, \quad j > 2.$$

This concludes the proof. □

THEOREM 1.8. *Let f be an element of $\mathscr{H}^2(\Omega)$ and the samples satisfy*

$$y_i = f(X_i) + \varepsilon_i,\ i = 1,\ldots,n, \tag{5}$$

where $y_1,\ldots,y_n$ are the observed functional values at $\Pi = \{X_i\}_{i=1}^n \subset \Omega$, and $\varepsilon_1,\ldots,\varepsilon_n$ are i.i.d random variables with zero mean and finite variance $\sigma^2 > 0$. Suppose (A.1) and (A.2) are fulfilled. Let $\hat{\mathbf{f}}_n(\lambda) = A_n(\lambda)\mathbf{y} = (I_n + \lambda M_{\Pi,2})^{-1}\mathbf{y}$ be the estimator from the DLRS model. Denote $r_n(\lambda) = n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^2$. As $n \to \infty$ and $\lambda \sim n^{-2/3}$ is chosen, then

$$\mathbf{E}[r_n(\lambda)] = O(n^{-\frac{2}{3}}).$$

PROOF. By using the bounds of eigenvalues $\mu_j = O(j^2)$ obtained in Lemma 1.7, we have

$$\begin{aligned}\mathbf{E}[r_n(\lambda)] &= \mathbf{E}[n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^2] \\ &= n^{-1}\left(\mathbf{f}^T\,(A_n(\lambda) - I_n)^2\,\mathbf{f} + \sigma^2\mathrm{tr}[A_n(\lambda)^2]\right) \\ &\le \frac{\lambda}{4n}\mathbf{f}^T M\mathbf{f} + \frac{\sigma^2}{n}\sum_{j=1}^n \frac{1}{(1+\lambda\mu_j)^2} \\ &= O(\lambda) + O\left(n^{-1}\lambda^{-\frac{1}{2}}\right),\end{aligned} \tag{6}$$

where the last equation is based on the result of Proposition 1.3 with $m = 2$. In particular, if the smoothing parameter is chosen to satisfy $\lambda \sim n^{-2/3}$, then we achieve the convergence rate $\mathbf{E}[r_n(\lambda)] = O(n^{-\frac{2}{3}})$. According to Stone [1982], $-\frac{2}{3}$ is the optimal for multivariate function estimation with the order 2 in the 2D domain Ω with some standard assumptions. Since the assumption (A.3) is satisfied with probability one as $n \to \infty$, we know the DLRS estimator achieves the optimal convergence rate with probability one. □

Using Theorem 1.8, we can easily prove Theorem 1 in the submission. Specifically, in the DLRS model we let the unknown function f be a C^2-smooth surface $\mathcal{S}$ itself and the observations $\mathbf{y} = (y_1,\ldots,y_n)^T$ be the noisy samples of surface position $P = (\mathbf{p}_1,\ldots,\mathbf{p}_n)^T$. Therefore we come to the conclusion of Theorem 1 in the submission.

2. PROOF OF THEOREM 2

We will show that the DLRS estimator satisfies some general conditions and then prove the asymptotic optimality of GCV under our proposed framework.

Let $\hat{\mathbf{f}}_n(\lambda) = A_n(\lambda)\mathbf{y} = (I_n + \lambda M)^{-1}\mathbf{y}$ be the estimator of our DLRS model and denote $r_n(\lambda) = n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^2$. The asymptotic optimality of GCV is defined as

$$\frac{r_n(\hat{\lambda}_G)}{\inf_{\lambda\in\mathbb{R}_+} r_n(\lambda)} \to_p 1, \tag{7}$$

which verifies the closeness between the values of risk function given by the GCV choice $\hat{\lambda}_G$ and the theoretically optimal choice $\lambda^* = \arg\inf_{\lambda\in\mathbb{R}_+} r_n(\lambda)$.

The main result here is to show that our estimator satisfies the following three conditions.

- (C.1) $\inf_{\lambda\in\mathbb{R}_+} n\mathbf{E}[r_n(\lambda)] \to \infty$.
- (C.2) There exists a sequence $\{\lambda_n\}$ such that $r_n(\lambda_n) \to_p 0$ (the convergence in probability).
- (C.3) Let $0 \le \kappa_1 \le \cdots \le \kappa_n$ be the eigenvalues of $K(\lambda) = \lambda M$. For any ℓ such that $\frac{\ell}{n} \to 0$, then $\frac{(n^{-1}\sum_{i=\ell+1}^n \kappa_i^{-1})^2}{n^{-1}\sum_{i=\ell+1}^n \kappa_i^{-2}} \to 0$ as $n \to \infty$.

The condition (C.1) states that the convergence rate of the risk function to zero should be lower than $O(n^{-1})$. Otherwise, the estimates may possess unattainably small risk.

Denote null(Δ_Ω) the null space of Laplacian operator Δ_Ω. Actually from the behavior of eigenvalues as shown in Lemma 1.7, it is not difficult to verify that our proposed model meets the condition (C.1) except for $f \in \text{null}(\Delta_\Omega)$.

LEMMA 2.1. *If $f \notin \text{null}(\Delta_\Omega)$, the estimator $\hat{\mathbf{f}}_n(\lambda)$ from our DLRS model holds*

$$\inf_{\lambda\in\mathbb{R}_+} n\mathbf{E}[r_n(\lambda)] \to \infty.$$

This verifies the condition (C.1).

PROOF. Let $0 \le \mu_1 \le \cdots \le \mu_n$ be the eigenvalues of design matrix M, and $\mathbf{u}_j$ the unit eigenvector corresponding to μ_j, $j = 1, \ldots, n$. So we have

$$\begin{aligned} n\mathbf{E}[r_n(\lambda)] &= n\mathbf{E}[n^{-1}\|\hat{\mathbf{f}}_n(\lambda) - \mathbf{f}\|^2] \\ &= \mathbf{E}[(\hat{\mathbf{f}}_n(\lambda) - \mathbf{f})^T(\hat{\mathbf{f}}_n(\lambda) - \mathbf{f})] \\ &= \mathbf{f}^T\,(A_n(\lambda) - I)^2\,\mathbf{f} + \sigma^2\text{tr}[A_n(\lambda)^2] \\ &= \sum_{j=1}^{n} \frac{\lambda^2\mu_j^2}{(1+\lambda\mu_j)^2} e_j^2 + \sigma^2 \sum_{j=1}^{n} \frac{1}{(1+\lambda\mu_j)^2}, \end{aligned} \tag{8}$$

where $e_j = \mathbf{u}_j^T\mathbf{f}$.

If $\lambda \sim O(1)$ or $\lambda \to \infty$ (corresponds to $n \to \infty$), since $\mu_j \sim j^2$ there exists j^* such that $\frac{j^*}{n} \to 0$ and $\frac{\lambda\mu_j}{1+\lambda\mu_j} \ge \frac{1}{2}$ for $j > j^*$, then

$$\begin{aligned} n\mathbf{E}[r_n(\lambda)] &\ge \sum_{j=1}^{n} \frac{\lambda^2\mu_j^2}{(1+\lambda\mu_j)^2} e_j^2 \\ &\ge \frac{1}{4}\sum_{j>j^*} e_j^2 \\ &\ge \frac{n}{4}|f|^2_{\Pi,0} - \frac{1}{4} j^* \max\left\{e_1^2, \ldots, e_{j^*}^2\right\} \\ &= O(n) \to \infty. \end{aligned}$$

On the other hand, if $\lambda \to 0$ corresponds to $n \to \infty$, we have

$$\begin{aligned} n\mathbf{E}[r_n(\lambda)] &\ge \sigma^2 \sum_{j=1}^{n} \frac{1}{(1+\lambda\mu_j)^2} \\ &= O(\lambda^{-\frac{1}{2}}) \\ &\to \infty, \end{aligned}$$

where the second equation is also based on Proposition 1.3. □

LEMMA 2.2. *Under condition (C.1), we have in probability*

$$\sup_{\lambda>0} \left| \frac{r_n(\lambda)}{\mathbf{E}[r_n(\lambda)]} - 1 \right| \to 0. \tag{9}$$

PROOF. To get Eq. (9), it suffices to show in probability

$$\sup_{\lambda>0} \frac{n^{-1}\left|\mathbf{f}^T(A_n(\lambda) - I_n)A_n(\lambda)\varepsilon\right|}{\mathbf{E}[r_n(\lambda)]} \to 0 \tag{10}$$

and

$$\sup_{\lambda>0} \frac{n^{-1}\left|\|A_n(\lambda)\varepsilon\|^2 - \sigma^2\text{tr}[A_n(\lambda)^2]\right|}{\mathbf{E}[r_n(\lambda)]} \to 0. \tag{11}$$

According to the Chebyshev inequality, we have for any given $\delta > 0$

$$\begin{aligned} &\Pr\left\{\frac{n^{-1}\left|\mathbf{f}^T(A_n(\lambda) - I_n)A_n(\lambda)\varepsilon\right|}{\mathbf{E}[r_n(\lambda)]} > \delta\right\} \\ &\le \delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-2}\mathbf{E}\left[(\mathbf{f}^T(A_n(\lambda) - I_n)A_n(\lambda)\varepsilon)^2\right] \\ &= \delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-2}\sigma^2 \\ &\quad \text{tr}\left[A_n(\lambda)(A_n(\lambda) - I_n)\mathbf{f}\,\mathbf{f}^T(A_n(\lambda) - I_n)A_n(\lambda)\right] \\ &= \delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-2}\sigma^2\|A_n(\lambda)(A_n(\lambda) - I_n)\mathbf{f}\|^2 \\ &\le \delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-1}\sigma^2\frac{\|(A_n(\lambda) - I_n)\mathbf{f}\|^2}{n\mathbf{E}[r_n(\lambda)]} \\ &\le \delta^{-2}\sigma^2(n\mathbf{E}[r_n(\lambda)])^{-1} \to 0, \end{aligned}$$

since $n\mathbf{E}[r_n(\lambda)] \ge \|(A_n(\lambda) - I_n)\mathbf{f}\|^2$. Thus Eq. (10) holds in probability.

Again using the Chebyshev inequality, we have for any given $\delta > 0$

$$\begin{aligned} &\Pr\left\{\frac{n^{-1}\left|\|A_n(\lambda)\varepsilon\|^2 - \sigma^2\text{tr}[A_n(\lambda)^2]\right|}{\mathbf{E}[r_n(\lambda)]} > \delta\right\} \\ &\le \delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-2}\mathbf{E}\left[(\|A_n(\lambda)\varepsilon\|^2 - \sigma^2\text{tr}[A_n(\lambda)^2])^2\right] \\ &= \delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-1}\frac{\mathbf{E}\left[\|A_n(\lambda)\varepsilon\|^4\right] - \left(\sigma^2\text{tr}[A_n(\lambda)^2]\right)^2}{n\mathbf{E}[r_n(\lambda)]}. \end{aligned}$$

Since $n\mathbf{E}[r_n(\lambda)] \ge \sigma^2\text{tr}[A_n(\lambda)^2]$, we only need to show

$$\frac{\mathbf{E}\left[\|A_n(\lambda)\varepsilon\|^4\right] - \left(\sigma^2\text{tr}[A_n(\lambda)^2]\right)^2}{\sigma^2\text{tr}[A_n(\lambda)^2]} < \text{Constant.} \tag{12}$$

Denote $B = A_n(\lambda)^2 = (B_{ij})_{n\times n}$, then we have

$$\begin{aligned} \mathbf{E}\left[\|A_n(\lambda)\varepsilon\|^4\right] &= \mathbf{E}\left[(\varepsilon^T B\varepsilon)^2\right] \\ &= \mathbf{E}\left[\left(\sum_{i,j} B_{ij}\varepsilon_i\varepsilon_j\right)\left(\sum_{i',j'} B_{i'j'}\varepsilon_{i'}\varepsilon_{j'}\right)\right] \\ &= \mathbf{E}\left[\left(\sum_{i} B_{ii}\varepsilon_i^2\right)\left(\sum_{i'} B_{i'i'}\varepsilon_{i'}^2\right)\right] \\ &\quad + \mathbf{E}\left[\left(\sum_{i\neq j} B_{ij}\varepsilon_i\varepsilon_j\right)\left(\sum_{i'\neq j'} B_{i'j'}\varepsilon_{i'}\varepsilon_{j'}\right)\right] \\ &\le \left(\sum_{i=1}^{n} B_{ii}\sigma^2\right)^2 + \sum_{i=1}^{n} B_{ii}^2\mathbf{E}\left[\varepsilon_i^4\right] + \sum_{i\neq j} B_{ij}^2\sigma^4. \end{aligned}$$

There exists a constant c such that $\mathbf{E}[\varepsilon_i^4] \le c\sigma^2$ and $\sigma^4 \le c\sigma^2$, so we get

$$\begin{aligned} \mathbf{E}\left[\|A_n(\lambda)\varepsilon\|^4\right] &\le \left(\sum_{i=1}^{n} B_{ii}\sigma^2\right)^2 + c\sum_{i=1}^{n} B_{ii}^2\sigma^2 + c\sum_{i\neq j} B_{ij}^2\sigma^2 \\ &= \left(\sum_{i=1}^{n} B_{ii}\sigma^2\right)^2 + c\sum_{i,j} B_{ij}^2\sigma^2 \\ &= \left(\sigma^2\text{tr}[A_n(\lambda)^2]\right)^2 + c\sigma^2\text{tr}[A_n(\lambda)^4] \\ &\le \left(\sigma^2\text{tr}[A_n(\lambda)^2]\right)^2 + c\sigma^2\text{tr}[A_n(\lambda)^2], \end{aligned}$$

which implies Eq. (12), and immediately leads to (11) in probability. □

The condition (C.2) shows that the risk function $r_n(\lambda_n)$ converges to zero in probability with appropriate sequence $\{\lambda_n\}$. Obviously, the conclusion of condition (C.2) can be easily derived from

Theorem 1.8 and Lemma 2.2. Therefore, the condition (C.2) holds true.

The condition (C.3) gives a ratio

$$\frac{\left(n^{-1}\sum_{i=\ell+1}^{n}\kappa_i^{-1}\right)^2}{n^{-1}\sum_{i=\ell+1}^{n}\kappa_i^{-2}}, \tag{13}$$

which is defined on the eigenvalues of $K(\lambda)=\lambda M$ and often plays an important role in the asymptotic analysis.

LEMMA 2.3. *In our model, for any ℓ such that $\frac{\ell}{n}\to 0$ and $\kappa_{\ell+1}>0$, then the ratio of Eq.* (13) *converges to zero as n (the sample size) goes to infinity. This verifies the condition (C.3).*

PROOF. From Lemma 1.7, namely, $\mu_i=O(i^2)$, we get

$$\begin{aligned}\lim_{n\to\infty}\frac{\left(n^{-1}\sum_{i=\ell+1}^{n}\kappa_i^{-1}\right)^2}{n^{-1}\sum_{i=\ell+1}^{n}\kappa_i^{-2}} &= \lim_{n\to\infty}\frac{\left(\sum_{i=\ell+1}^{n}\mu_i^{-1}\right)^2}{n\sum_{i=\ell+1}^{n}\mu_i^{-2}}\\ &= \lim_{n\to\infty}\frac{\left(\int_{\ell+1}^{n}\mu^{-2m/d}\mathrm{d}\mu\right)^2}{n\int_{\ell+1}^{n}\mu^{-4m/d}\mathrm{d}\mu}\\ &= \lim_{n\to\infty}\frac{(4m-d)d}{(2m-d)^2}\cdot\frac{\left((\ell+1)^{1-\frac{2m}{d}}-n^{1-\frac{2m}{d}}\right)^2}{n\left((\ell+1)^{1-\frac{4m}{d}}-n^{1-\frac{4m}{d}}\right)}\\ &= \lim_{n\to\infty}\frac{(4m-d)d}{(2m-d)^2}\cdot\frac{\ell+1}{n}\cdot\frac{\left(1-(\frac{\ell+1}{n})^{\frac{2m}{d}-1}\right)^2}{\left(1-(\frac{\ell+1}{n})^{\frac{4m}{d}-1}\right)}\\ &= 0. \quad\square\end{aligned}$$

By conclusion, we have verified that the three conditions (C.1), (C.2), and (C.3) hold true for our model. Then we will prove the asymptotic optimality of GCV under these three conditions.

LEMMA 2.4. *Under the condition (C.2), we have*

$$n^{-1}\mathrm{tr}[I_n-A_n(\lambda_n)]\to 1, \tag{14}$$

and

$$n^{-1}\|(I_n-A_n(\lambda_n))\mathbf{y}\|^2\to\sigma^2. \tag{15}$$

PROOF. From the fact that

$$\sigma^2(n^{-1}\mathrm{tr}[A_n(\lambda_n)])^2\le\sigma^2 n^{-1}\mathrm{tr}[A_n(\lambda_n)^2]\le\mathbf{E}[r_n(\lambda_n)]\to 0,$$

we have $n^{-1}\mathrm{tr}[A_n(\lambda_n)]\to 0$ and then get

$$n^{-1}\mathrm{tr}[I_n-A_n(\lambda_n)]\to 1.$$

By the fact $n^{-1}\|\varepsilon\|^2\to\sigma^2$ and the Cauchy-Schwartz inequality, we have

$$\begin{aligned}n^{-1}\|(I_n-A_n(\lambda_n))\mathbf{y}\|^2 &= n^{-1}\|\varepsilon\|^2\\ &+n^{-1}\|\mathbf{f}-\hat{\mathbf{f}}_n(\lambda_n)\|^2+\frac{2}{n}|(\mathbf{f}-\hat{\mathbf{f}}_n(\lambda_n))^T\varepsilon|\to\sigma^2. \quad\square\end{aligned}$$

LEMMA 2.5. *Under the condition (C.3), for λ_n such that $r_n(\lambda_n)\to 0$, we have*

$$\frac{\left(n^{-1}\mathrm{tr}[A_n(\lambda_n)]\right)^2}{n^{-1}\mathrm{tr}[A_n(\lambda_n)^2]}\to 0. \tag{16}$$

PROOF. Recall $A_n(\lambda_n)=(I_n+\lambda_n M)^{-1}=(I_n+K_n(\lambda_n))^{-1}$. We get

$$\frac{\left(n^{-1}\mathrm{tr}[A_n(\lambda_n)]\right)^2}{n^{-1}\mathrm{tr}[A_n(\lambda_n)^2]}=\frac{\left(n^{-1}\sum_{i=1}^{n}(1+\kappa_i)^{-1}\right)^2}{n^{-1}\sum_{i=1}^{n}(1+\kappa_i)^{-2}}, \tag{17}$$

where $0\le\kappa_1\le\cdots\le\kappa_n$ are the eigenvalues of $K_n(\lambda_n)$. Let ℓ be the number holding $\kappa_\ell\le 1<\kappa_{\ell+1}$, then we have

$$\sum_{i=1}^{n}(1+\kappa_i)^{-1}\le\ell+\sum_{i=\ell+1}^{n}\kappa_i^{-1}, \tag{18}$$

and

$$\sum_{i=1}^{n}(1+\kappa_i)^{-2}\ge\frac{1}{4}\left(\ell+\sum_{i=\ell+1}^{n}\kappa_i^{-2}\right). \tag{19}$$

To reach Eq. (16), it suffices to show

$$\frac{\left(\frac{\ell}{n}+\frac{1}{n}\sum_{i=\ell+1}^{n}\kappa_i^{-1}\right)^2}{\frac{1}{4}\left(\frac{\ell}{n}+\frac{1}{n}\sum_{i=\ell+1}^{n}\kappa_i^{-2}\right)}\to 0. \tag{20}$$

On the other hand, $\mathbf{E}[r_n(\lambda_n)]\to 0$ since $r_n(\lambda_n)$ is non-negative, thus we get $n^{-1}\mathrm{tr}[A_n(\lambda_n)^2]\to 0$ and have $\frac{\ell}{n}\to 0$ due to Eq. (19). So it is not hard to see that (20) holds under the condition (C.3). $\square$

LEMMA 2.6. *For any $\hat{\lambda}$ such that $r_n(\hat{\lambda})\to 0$ and*

$$\frac{\left(n^{-1}\mathrm{tr}[A_n(\hat{\lambda})]\right)^2}{n^{-1}\mathrm{tr}[A_n(\hat{\lambda})^2]}\to 0, \tag{21}$$

under the condition (C.1) we have

$$\frac{\left|\mathrm{SURE}_n(\hat{\lambda})-\tilde{r}_n(\hat{\lambda})-n^{-1}\|\varepsilon\|^2+\sigma^2\right|}{r_n(\hat{\lambda})}\longrightarrow_p 0, \tag{22}$$

and

$$\frac{n^{-1}\|\tilde{\mathbf{f}}_n(\hat{\lambda})-\hat{\mathbf{f}}_n(\hat{\lambda})\|^2}{r_n(\hat{\lambda})}\longrightarrow_p 0, \tag{23}$$

where $\mathrm{SURE}_n(\lambda)=\sigma^2-\sigma^4\frac{\left(n^{-1}\mathrm{tr}[I_n-A_n(\lambda)]\right)^2}{n^{-1}\|(I_n-A_n(\lambda))\mathbf{y}\|^2}$, $\tilde{\mathbf{f}}_n(\lambda)=\mathbf{y}-\sigma^2\frac{\mathrm{tr}[I_n-A_n(\lambda)]}{\|(I_n-A_n(\lambda))\mathbf{y}\|^2}(I_n-A_n(\lambda))\mathbf{y}$ *and* $\tilde{r}_n(\lambda)=n^{-1}\|\tilde{\mathbf{f}}_n(\lambda)-\mathbf{f}\|^2$.

Proof of the Lemma 2.6 is left in the Appendix.

LEMMA 2.7. *Under conditions (C.2) and (C.3), $\hat{\mathbf{f}}_n(\hat{\lambda}_G)$ is consistent, that is, $r_n(\hat{\lambda}_G)\to 0$, where $\hat{\lambda}_G$ is chosen by GCV.*

PROOF. According to the proof of Lemma 5.2 in Li [1985] and similarly as in Girard [1991], the preceding lemma can be established. $\square$

2.1 Asymptotic Optimality Theorem

THEOREM 2.8. *Under conditions (C.1), (C.2), and (C.3), $\hat{\mathbf{f}}_n(\hat{\lambda}_G)$ is asymptotically optimal, where $\hat{\lambda}_G$ is the GCV choice.*

PROOF. From the condition (C.2), for λ_n^* that is the minimizer of $r_n(\lambda)$, we have $r_n(\lambda_n^*)\to 0$. According to Lemma 2.5, we have

$$\frac{\left(n^{-1}\mathrm{tr}[A_n(\lambda_n^*)]\right)^2}{n^{-1}\mathrm{tr}[A_n(\lambda_n^*)^2]}\to 0. \tag{24}$$

Hence from Lemma 2.6, we have $\mathrm{SURE}_n(\lambda_n^*)-n^{-1}\|\varepsilon_n\|^2+\sigma^2=r_n(\lambda_n^*)(1+o_p(1))$.

On the other hand, from Lemma 2.7 this also holds for $\hat{\lambda}=\hat{\lambda}_G$. Therefore we have

$$\mathrm{SURE}_n(\hat{\lambda}_G)-n^{-1}\|\varepsilon_n\|^2+\sigma^2=r_n(\hat{\lambda}_G)(1+o_p(1)). \tag{25}$$

Since $\mathrm{SURE}_n(\hat{\lambda}_G)\le\mathrm{SURE}_n(\lambda_n^*)$ and $r_n(\lambda_n^*)\le r_n(\hat{\lambda}_G)$, we have $r_n(\hat{\lambda}_G)/r_n(\lambda_n^*)\to 1$ in probability. $\square$

Proof of Lemma 2.6

PROOF. We first prove Eq. (22), which can be rewritten as

$$\frac{2\left|\frac{\sigma^2\mathrm{tr}[I_n-A_n(\lambda)]\mathbf{y}^T(I_n-A_n(\lambda))\varepsilon}{n\|(I_n-A_n(\lambda))\mathbf{y}\|^2}-\frac{\sigma^4(\mathrm{tr}[I_n-A_n(\lambda)])^2}{n\|(I_n-A_n(\lambda))\mathbf{y}\|^2}-n^{-1}\|\varepsilon\|^2+\sigma^2\right|}{r_n(\lambda)}$$
$$\le 2\frac{\sigma^2\mathrm{tr}[I_n-A_n(\lambda)]}{\|(I_n-A_n(\lambda))\mathbf{y}\|^2}\cdot\frac{n^{-1}\left|\mathbf{f}^T(I_n-A_n(\lambda))\varepsilon\right|}{r_n(\lambda)}$$
$$+2\frac{\sigma^2\mathrm{tr}[I_n-A_n(\lambda)]}{\|(I_n-A_n(\lambda))\mathbf{y}\|^2}\cdot\frac{n^{-1}\left|\varepsilon^TA_n(\lambda)\varepsilon-\sigma^2\mathrm{tr}[A_n(\lambda)]\right|}{r_n(\lambda)}$$
$$+2\frac{\left|\left(\frac{\sigma^2\mathrm{tr}[I_n-A_n(\lambda)]}{\|(I_n-A_n(\lambda))\mathbf{y}\|^2}-1\right)\left(\sigma^2-n^{-1}\|\varepsilon\|^2\right)\right|}{r_n(\lambda)}. \quad (26)$$

Note that $n^{-1}\mathrm{tr}[I_n-A_n(\lambda_n)]\to 1$, $n^{-1}\|(I_n-A_n(\lambda_n))\mathbf{y}\|^2\to\sigma^2$ from Lemma 2.4, and $\sup_{\lambda>0}|\frac{r_n(\lambda)}{\mathbf{E}[r_n(\lambda)]}-1|\to 0$ by Lemma 2.2. Thus it suffices for us to show the following three equations

$$\sup_{\lambda>0}\frac{n^{-1}\left|\mathbf{f}^T(I_n-A_n(\lambda))\varepsilon\right|}{\mathbf{E}[r_n(\lambda)]}\to 0, \quad (27)$$

$$\sup_{\lambda>0}\frac{n^{-1}\left|\varepsilon^TA_n(\lambda)\varepsilon-\sigma^2\mathrm{tr}[A_n(\lambda)]\right|}{\mathbf{E}[r_n(\lambda)]}\to 0, \quad (28)$$

$$\sup_{\lambda>0}\frac{|(\sigma^2n^{-1}\mathrm{tr}[I_n-A_n(\lambda)]-n^{-1}\|(I_n-A_n(\lambda))\mathbf{y}\|^2)(\sigma^2-n^{-1}\|\varepsilon\|^2)|}{\mathbf{E}[r_n(\lambda)]}\to 0. \quad (29)$$

For Eq. (27), according to the Chebyshev inequality, we have for any given $\delta>0$

$$\begin{aligned}
&\Pr\left\{\frac{n^{-1}|\mathbf{f}^T(I_n-A_n(\lambda))\varepsilon|}{\mathbf{E}[r_n(\lambda)]}>\delta\right\}\\
&\le\delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-2}\mathbf{E}\left[(\mathbf{f}^T(I_n-A_n(\lambda))\varepsilon)^2\right]\\
&=\delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-2}\sigma^2\mathrm{tr}\left[(I_n-A_n(\lambda))\mathbf{f}\,\mathbf{f}^T(I_n-A_n(\lambda))\right]\\
&=\delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-2}\sigma^2\|(I_n-A_n(\lambda))\mathbf{f}\|^2\\
&=\delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-1}\sigma^2\tfrac{\|(I_n-A_n(\lambda))\mathbf{f}\|^2}{n\mathbf{E}[r_n(\lambda)]}\\
&\le\delta^{-2}\sigma^2(n\mathbf{E}[r_n(\lambda)])^{-1}\to 0,
\end{aligned}$$

since $n\mathbf{E}[r_n(\lambda)]\ge\|(I_n-A_n(\lambda))\mathbf{f}\|^2$.

For Eq. (28), again using the Chebyshev inequality, we have for any given $\delta>0$

$$\begin{aligned}
&\Pr\left\{\frac{n^{-1}\left|\varepsilon^TA_n(\lambda)\varepsilon-\sigma^2\mathrm{tr}[A_n(\lambda)]\right|}{\mathbf{E}[r_n(\lambda)]}>\delta\right\}\\
&\le\delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-2}\mathbf{E}\Big[\left(\varepsilon^TA_n(\lambda)\varepsilon-\sigma^2\mathrm{tr}[A_n(\lambda)]\right)^2\Big]\\
&=\delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-1}\tfrac{\mathbf{E}\left[(\varepsilon^TA_n(\lambda)\varepsilon)^2\right]-\left(\sigma^2\mathrm{tr}[A_n(\lambda)]\right)^2}{n\mathbf{E}[r_n(\lambda)]}.
\end{aligned}$$

Since $n\mathbf{E}[r_n(\lambda)]\ge\sigma^2\mathrm{tr}[A_n(\lambda)^2]$, we only need to show

$$\frac{\mathbf{E}\left[(\varepsilon^TA_n(\lambda)\varepsilon)^2\right]-\left(\sigma^2\mathrm{tr}[A_n(\lambda)]\right)^2}{\sigma^2\mathrm{tr}[A_n(\lambda)^2]}<\text{Constant}. \quad (30)$$

Denote $A_n(\lambda)=(A_{ij})_{n\times n}$, then we have

$$\begin{aligned}
\mathbf{E}\left[(\varepsilon^TA_n(\lambda)\varepsilon)^2\right]&=\mathbf{E}\Big[\Big(\sum_{i,j}A_{ij}\varepsilon_i\varepsilon_j\Big)\Big(\sum_{i',j'}A_{i'j'}\varepsilon_{i'}\varepsilon_{j'}\Big)\Big]\\
&=\mathbf{E}\Big[\Big(\sum_{i}A_{ii}\varepsilon_i^2\Big)\Big(\sum_{i'}A_{i'i'}\varepsilon_{i'}^2\Big)\Big]\\
&\quad+\mathbf{E}\Big[\Big(\sum_{i\ne j}A_{ij}\varepsilon_i\varepsilon_j\Big)\Big(\sum_{i'\ne j'}A_{i'j'}\varepsilon_{i'}\varepsilon_{j'}\Big)\Big]\\
&\le\left(\sum_{i=1}^nA_{ii}\sigma^2\right)^2+\sum_{i=1}^nA_{ii}^2\mathbf{E}\left[\varepsilon_i^4\right]+\sum_{i\ne j}A_{ij}^2\sigma^4.
\end{aligned}$$

There exists a constant c such that $\mathbf{E}[\varepsilon_i^4]\le c\sigma^2$ and $\sigma^4\le c\sigma^2$, so we get

$$\begin{aligned}
\mathbf{E}\left[(\varepsilon^TA_n(\lambda)\varepsilon)^2\right]&\le\left(\sum_{i=1}^nA_{ii}\sigma^2\right)^2+c\sum_{i=1}^nA_{ii}^2\sigma^2+c\sum_{i\ne j}A_{ij}^2\sigma^2\\
&=\left(\sum_{i=1}^nA_{ii}\sigma^2\right)^2+c\sum_{i,j}A_{ij}^2\sigma^2\\
&=\left(\sigma^2\mathrm{tr}[A_n(\lambda)]\right)^2+c\sigma^2\mathrm{tr}[A_n(\lambda)^2],
\end{aligned}$$

which implies Eq. (30), and immediately leads to (28).

For Eq. (29), using the proved (27), (28), and $\sigma^2(n^{-1}\mathrm{tr}[A_n(\lambda)])^2\le\sigma^2n^{-1}\mathrm{tr}[A_n(\lambda)^2]\le\mathbf{E}[r_n(\lambda)]$, we only need to show

$$\sup_{\lambda>0}\frac{\left|\sigma^2-n^{-1}\|\varepsilon\|^2\right|}{(\mathbf{E}[r_n(\lambda)])^{1/2}}\to 0, \quad (31)$$

since the fact that

$$\begin{aligned}
&\left|\sigma^2n^{-1}\mathrm{tr}[I_n-A_n(\lambda)]-n^{-1}\|(I_n-A_n(\lambda))\|^2\right|\\
&=\left|\sigma^2-\sigma^2n^{-1}\mathrm{tr}[A_n(\lambda)]-n^{-1}\|\varepsilon+\mathbf{f}-\hat{\mathbf{f}}_n(\lambda)\|^2\right|\\
&=|\sigma^2-\sigma^2n^{-1}\mathrm{tr}[A_n(\lambda)]-n^{-1}\|\varepsilon\|^2-r_n(\lambda)-2n^{-1}(\mathbf{f}-\hat{\mathbf{f}}_n(\lambda))^T\varepsilon|\\
&=|\sigma^2-n^{-1}\|\varepsilon\|^2-\sigma^2n^{-1}\mathrm{tr}[A_n(\lambda)]\\
&\quad-r_n(\lambda)-2n^{-1}\mathbf{f}^T(I_n-A_n(\lambda))\varepsilon+2n^{-1}\varepsilon^TA_n(\lambda)\varepsilon|\\
&\le\left|\sigma^2-n^{-1}\|\varepsilon\|^2\right|+r_n(\lambda)+2n^{-1}\left|\mathbf{f}^T(I_n-A_n(\lambda))\varepsilon\right|\\
&\quad+2n^{-1}\left|\varepsilon^TA_n(\lambda)\varepsilon-\sigma^2\mathrm{tr}[A_n(\lambda)]\right|+\sigma^2n^{-1}\mathrm{tr}[A_n(\lambda)].
\end{aligned}$$

By the Chebyshev inequality, we have for any given $\delta>0$

$$\begin{aligned}
&\Pr\left\{\frac{\left|\sigma^2-n^{-1}\|\varepsilon\|^2\right|}{(\mathbf{E}[r_n(\lambda)])^{1/2}}>\delta\right\}\\
&\le\delta^{-2}(\mathbf{E}[r_n(\lambda)])^{-1}\mathbf{E}\Big[\left(\sigma^2-n^{-1}\|\varepsilon\|^2\right)^2\Big]\\
&=\delta^{-2}(\mathbf{E}[r_n(\lambda)])^{-1}\left(n^{-2}\mathbf{E}[\|\varepsilon\|^4]-\sigma^4\right)\\
&\le\delta^{-2}(\mathbf{E}[r_n(\lambda)])^{-1}\left(n^{-2}\left(n^2\sigma^4+n\mathbf{E}\left[\varepsilon_i^4\right]\right)-\sigma^4\right)\\
&=\delta^{-2}(n\mathbf{E}[r_n(\lambda)])^{-1}\mathbf{E}\left[\varepsilon_i^4\right]\to 0,
\end{aligned}$$

which implies (31).

Now it remains to prove Eq. (23), the numerator of which can be rearranged as

$$\begin{aligned}
&n^{-1}\|\tilde{\mathbf{f}}_n(\hat\lambda)-\hat{\mathbf{f}}_n(\hat\lambda)\|^2\\
&=\left(\frac{\sigma^2n^{-1}\mathrm{tr}[I_n-A_n(\lambda)]}{n^{-1}\|(I_n-A_n(\lambda))\mathbf{y}\|^2}-1\right)^2n^{-1}\|(I_n-A_n(\lambda))\mathbf{y}\|^2\\
&=\tfrac{\left((\sigma^2-n^{-1}\|\varepsilon\|^2)-r_n(\lambda)-2n^{-1}\mathbf{f}^T(I_n-A_n(\lambda))\varepsilon+2n^{-1}(\varepsilon^TA_n(\lambda)\varepsilon-\sigma^2\mathrm{tr}[A_n(\lambda)])+\sigma^2n^{-1}\mathrm{tr}[A_n(\lambda)]\right)^2}{n^{-1}\|(I_n-A_n(\lambda))\mathbf{y}\|^2}.
\end{aligned}$$

To get (23), since $n^{-1}\|(I_n - A_n(\lambda))\mathbf{y}\|^2 \to \sigma^2$, it suffices to show the following

$$\frac{\left(\sigma^2 - n^{-1}\|\varepsilon\|^2\right)^2}{r_n(\lambda)} \to 0, \tag{32}$$

$$\frac{\left(n^{-1}\mathbf{f}^T(I_n - A_n(\lambda))\varepsilon\right)^2}{r_n(\lambda)} \to 0, \tag{33}$$

$$\frac{\left(n^{-1}(\varepsilon^T A_n(\lambda)\varepsilon - \sigma^2 \mathrm{tr}[A_n(\lambda)])\right)^2}{r_n(\lambda)} \to 0, \tag{34}$$

$$\frac{\left(n^{-1}\mathrm{tr}[A_n(\lambda)]\right)^2}{r_n(\lambda)} \to 0. \tag{35}$$

Note that $\sup_{\lambda>0}|\frac{r_n(\lambda)}{\mathbf{E}[r_n(\lambda)]} - 1| \to 0$, then Eqs. (32), (33), and (34) can be easily proved from (31), (27), and (28) respectively. The last equation (35) follows from $\sigma^2 n^{-1}\mathrm{tr}[A_n(\lambda)^2] \leq \mathbf{E}[r_n(\lambda)]$ and (21).

Hence, we complete the proof of Lemma 2.6. □

REFERENCES

S. Agmon. 1965. *Lectures on Elliptic Boundary Value Problems*. D. Van Norstrand, Princeton, NJ.

D. A. Girard. 1991. Asymptotic optimality of the fast randomized versions of GCV and cl in ridge regression and regularization. *Ann. Statist.* 19, 4, 1950–1963.

P. R. Halmos. 1982. *A Hilbert Space Problem Book*. Vol. 19 of Graduate Texts in Mathematics, 2nd Ed. Springer.

K.-C. Li. 1985. From Stein's unbiased risk estimates to the method of generalized cross validation. *Ann. Statist.* 13, 4, 147 1352–1377.

C. J. Stone. 1982. Optimal global rates of convergence for nonparametric regression. *The Ann. Statist.* 10, 4, 1040–1053.

F. I. Utreras. 1988. Convergence rates for multivariate smoothing spline functions. *J. Approx. Theory* 52, 1, 152 1–27.

Received December 2012; revised December 2013; accepted December 2013