Supplementary Materials of "The Reconstruction Approach: From Interpolation to Regression"

Shifeng Xiong

NCMIS, KLSC, Academy of Mathematics and Systems Science

Chinese Academy of Sciences, Beijing 100190

xiong@amss.ac.cn

Abstract All proofs are given in the Supplementary Materials.

Appendix: Proofs

Proof of Theorem 1. We have

$$\mathrm{MSE}\left(\hat{f}(\mathbf{x}_0)\right) = E\left(\hat{f}(\mathbf{x}_0) - f(\mathbf{x}_0)\right)^2 = E\left(\sum_{j=1}^{m} b_j(\mathbf{x}_0)\hat{\gamma}_j - f(\mathbf{x}_0)\right)^2$$

$$\leqslant E\left(\sum_{j=1}^{m}\left[\hat{\gamma}_j - f(\mathbf{a}_j)\right] b_j(\mathbf{x}_0)\right)^2 /2 + \left(\sum_{j=1}^{m} b_j(\mathbf{x}_0) f(\mathbf{a}_j) - f(\mathbf{x}_0)\right)^2 /2$$

$$\leqslant E \sup_{j=1,\ldots,m} |\hat{\gamma}_j - f(\mathbf{a}_j)|^2 \left(\sum_{j=1}^{m} |b_j(\mathbf{x}_0)|\right)^2 /2 + \delta_m^2/2 = O\left(\delta_m^2 + \epsilon_n\right),$$

which completes the proof. $\square$

Proof of (11). Let $b_j(x) = \prod_{1\leqslant k\leqslant m,\ k\neq j}(x-a_k)/(a_j-a_k)$. We have

$$E\hat{f}(x_0) - f(x_0) = \sum_{j=1}^{m} b_j(x_0) E\hat{\gamma}_j - f(x_0) = \sum_{j=1}^{m} b_j(x_0) f(a_j) - f(x_0);$$

$$Var\left(\hat{f}(x_0)\right) = \sum_{j=1}^{m} b_j^2(x_0) Var(\hat{\gamma}_j) = \sigma^2 \|\mathbf{b}(x_0)\|^2/l.$$

By the theoretical results on the Chebyshev nodes-based polynomial interpolation (Stewart

1996), the convergence rate $\delta_m = \max_{t\in[0,1]} |f^{(m)}(t)|/2^{2m-1}m!$ and $\|\mathbf{b}(x_0)\| \leqslant 1$. Therefore,

$$\text{MSE}\left(\hat{f}(x_0)\right) \leqslant \delta_m^2 + m\sigma^2/n = \left[\frac{1}{2^{2m-1}m!}\max_{t\in[0,1]}|f^{(m)}(t)|\right]^2 + m\sigma^2/n.$$

(i) Assume that $\max_{t\in[0,1]}|f^{(m)}(t)| = O(m!)$. Take $m = C\log\left(n/\log(n)\right)$ with constant $C \in (0, 1/2]$. We have

$$\frac{1}{2^{2m-1}m!}\max_{t\in[0,1]}|f^{(m)}(t)| = O\left(\frac{1}{4^m}\right) = o\left(\frac{1}{e^m}\right) = o\left(\sqrt{\frac{\log(n)}{n}}\right),$$

which implies

$$\text{MSE}\left(\hat{f}(x_0)\right) = o\left(\sqrt{\frac{\log(n)}{n}}\right) + m\sigma^2/n = O\left(\frac{\log(n)}{n}\right).$$

(ii) Assume that $\max_{t\in[0,1]}|f^{(m)}(t)| = O\left((4/e)^m m\right)$ and $m\sqrt{\log(m)} = C\log(n)$ with constant $C > 0$. By Stirling's formula, we have

$$\frac{1}{2^{2m-1}m!}\max_{t\in[0,1]}|f^{(m)}(t)| \cdot \sqrt{\frac{n}{\log(n)}} = O\left(\frac{m\sqrt{n}}{\sqrt{m}m^m\sqrt{\log(n)}}\right)$$
$$= O\left(\frac{\exp(m\sqrt{\log(m)}/C)}{\exp(m\log(m))(\log(m))^{1/4}}\right) = o(1),$$

which implies

$$\text{MSE}\left(\hat{f}(x_0)\right) = o\left(\sqrt{\frac{\log(n)}{n}}\right) + m\sigma^2/n = o\left(\frac{\log(n)}{n}\right). \quad \square$$

Proof of (13). Assume $f \in C^4([0,1])$. Since the convergence rate of the cubic spline interpolator is $1/m^4$ (Stewart 1996), similar to the proof of (11), we have

$$\text{MSE}\left(\hat{f}(x_0)\right) \leqslant (m^{-4})^2 + O(m/n).$$

With $m \sim n^{1/9}$,

$$\text{MSE}\left(\hat{f}(x_0)\right) = O\left(n^{-8/9}\right). \quad \square$$

Let $\tilde{f}$ denote the solution to the general KRR problem (26) in Section 4.2. Recall that $\hat{f}$ is defined in (25). We now prove $\tilde{f} = \hat{f}$.

Proof. For simplicity, let $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{r}(\mathbf{x})$ denote $\mathbf{R}_{\mathcal{X}}$, $\mathbf{G}_{\mathcal{X}}$, and $\mathbf{r}_{\mathcal{X}}(\mathbf{x})$, respectively.

By (25),

$$\hat{f}(\mathbf{x}) = \mathbf{y}'(\mathbf{I}_n + n\lambda\mathbf{VRV}')^{-1}\mathbf{b}(\mathbf{x}) = \mathbf{y}'(\mathbf{I}_n + n\lambda\mathbf{VRV}')^{-1}\left\{\mathbf{Ug}(\mathbf{x}) + \mathbf{Vr}(\mathbf{x})\right\}.$$

By (26) and the representer theorem (Schölkopf, Herbrich, and Smola 2001),

$$\tilde{f}(\mathbf{x}) = \mathbf{y}'(\mathbf{R}\ \mathbf{G})\begin{pmatrix} \mathbf{R}^2 + n\lambda\mathbf{R} & \mathbf{RG} \\ \mathbf{G}'\mathbf{R} & \mathbf{G}'\mathbf{G} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{r}(\mathbf{x}) \\ \mathbf{g}(\mathbf{x}) \end{pmatrix}.$$

It suffices to prove

$$(\mathbf{R}\ \mathbf{G})\begin{pmatrix} \mathbf{R}^2 + n\lambda\mathbf{R} & \mathbf{RG} \\ \mathbf{G}'\mathbf{R} & \mathbf{G}'\mathbf{G} \end{pmatrix}^{-1} = (\mathbf{I}_n + n\lambda\mathbf{VRV}')^{-1}(\mathbf{V}\ \mathbf{U}),$$

which is implied by

$$(\mathbf{V}\ \mathbf{U})\begin{pmatrix} \mathbf{R}^2 + n\lambda\mathbf{R} & \mathbf{RG} \\ \mathbf{G}'\mathbf{R} & \mathbf{G}'\mathbf{G} \end{pmatrix} = (\mathbf{I}_n + n\lambda\mathbf{VRV}')(\mathbf{R}\ \mathbf{G}). \tag{0.1}$$

Recall $\mathbf{U} = \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}$ and $\mathbf{V} = \left[\mathbf{I}_n - \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\right]\mathbf{R}^{-1}$. We have

$$\begin{aligned}
&\mathbf{VRV}'\mathbf{R} = \left[\mathbf{I}_n - \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\right]\mathbf{R}^{-1}\left[\mathbf{R} - \mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\right] \\
&= \mathbf{I}_n - \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}' = \mathbf{VR}.
\end{aligned}$$

Therefore,

$$\left[\mathbf{V}(\mathbf{R}^2 + n\lambda\mathbf{R}) + \mathbf{U}\mathbf{G}'\mathbf{R}\right] - [\mathbf{R} + n\lambda\mathbf{V}\mathbf{R}\mathbf{V}'\mathbf{R}] = \mathbf{V}\mathbf{R}^2 + \mathbf{U}\mathbf{G}'\mathbf{R} - \mathbf{R} + n\lambda(\mathbf{V}\mathbf{R} - \mathbf{V}\mathbf{R}\mathbf{V}'\mathbf{R})$$
$$= \mathbf{R} - \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{R} + \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{R} - \mathbf{R} = \mathbf{0}. \tag{0.2}$$

On the other hand, we have

$$\mathbf{V}'\mathbf{G} = \mathbf{R}^{-1}\left[\mathbf{I} - \mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{R}^{-1}\right]\mathbf{G} = \mathbf{0}.$$

Therefore,

$$[\mathbf{V}\mathbf{R}\mathbf{G} + \mathbf{U}\mathbf{G}'\mathbf{G}] - [\mathbf{G} + n\lambda\mathbf{V}\mathbf{R}\mathbf{V}'\mathbf{G}] = \mathbf{V}\mathbf{R}\mathbf{G} + \mathbf{U}\mathbf{G}'\mathbf{G} - \mathbf{G}$$
$$= \left[\mathbf{I}_n - \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\right]\mathbf{G} + \mathbf{R}^{-1}\mathbf{G}(\mathbf{G}'\mathbf{R}^{-1}\mathbf{G})^{-1}\mathbf{G}'\mathbf{G} - \mathbf{G} = \mathbf{0}. \tag{0.3}$$

Obviously (0.1) follows from (0.2) and (0.3). $\square$

References

Stewart, G. W. (1996). *Afternotes on Numerical Analysis*, Society for Industrial and Applied Mathematics.

Schölkopf, B., Herbrich, R., and Smola, A. J. (2001). A generalized representer theorem. In *International Conference on Computational Learning Theory*, 416–426, Springer.