Data Driven Adaptive Spline Smoothing

Ziyue Liu and Wensheng Guo

University of Pennsylvania

Supplementary Material

S1 Filtering and smoothing algorithms

Filtering Steps. Define $a_{j+1} = \mathrm{E}\{x(t_{j+1})\,|y_1, \cdots, y_j\}$, $P_{j+1} = \mathrm{var}\{x(t_{j+1})\,|y_1, \cdots, y_j\}$, for $j = 1, \cdots, n$, the filtering equations are

$$
\begin{aligned}
v_j &= y_j - Z a_j, \\
V_j &= Z P_j Z' + \sigma^2, \\
K_j &= H_{j,j-1} P_j Z' V_j^{-1}, \\
L_j &= H_{j,j-1} - K_j Z, \\
a_{j+1} &= H_{j,j-1} a_j + K_j v_j, \\
P_{j+1} &= H_{j,j-1} P_j L_j' + \Omega_{j,j-1}.
\end{aligned}
$$

The log-likelihood can be calculated through the filtering step as

$$
\begin{aligned}
l(\theta \mid y) &= p(y_1, \cdots, y_n) = \sum_{j=1}^{n} \log p(y_j \mid y_0, \cdots, y_{j-1}) \\
&= -\frac{n}{2} \log 2\pi - \frac{1}{2} \sum_{j=1}^{n} \left(\log \mid V_j \mid + v_j' V_j^{-1} v_j \right).
\end{aligned}
$$

Smoothing Steps. Define $s_j = \mathrm{E}\{x(t_j)\,|y_1, \cdots, y_n\}$, $W_j = \mathrm{var}\{x(t_j)\,|y_1, \cdots, y_n\}$, for $j = n, \cdots, 1$, initialized with $r_n = 0$ and $N_n = 0$, the smoothing equations are

$$
\begin{aligned}
r_{j-1} &= Z' V_j^{-1} v_j + L_j' r_j, \\
N_{j-1} &= Z' V_j^{-1} Z + L_j' N_j L_j, \\
s_j &= a_j + P_j r_{j-1}, \\
W_j &= P_j - P_j N_{j-1} P_j.
\end{aligned}
$$

S2 Proofs

Proof of the Lemma. Observe

$$\begin{aligned}\Sigma_\lambda(i,j) &= \lambda_{min}^{-1}\int_0^1 G_m(t_i,u)\,G_m(t_j,u)\,du \\ &\quad -\int_0^1 \{\lambda_{min}^{-1}-\lambda^{-1}(u)\}G_m(t_i,u)\,G_m(t_j,u)\,du \\ &= \lambda_{min}^{-1}\Sigma(i,j)-\Sigma_\lambda^\Delta(i,j),\end{aligned}$$

where Σ_λ^Δ is nonnegative definite. Recall a result in linear algebra (e.g. Fulton(2000)): let A and B be two real symmetric matrices, let $C=A+B$, denote the eigenvalues of A by

$$\alpha:\ \alpha_1\geq\cdots\geq\alpha_n,$$

and similarly β for B and γ for C, then the i^{th} largest eigenvalue of C satisfies the following inequality

$$\max_{j+k=n+i}\alpha_j+\beta_k\leq\gamma_i\leq\min_{j+k=i+1}\alpha_j+\beta_k.$$

Let $j=i$, apply the inequality, we have

$$\delta_{in}\leq\lambda_{min}^{-1}\delta_{in}^{*}.$$

Similarly by factoring λ_{max} out we get the other part of the inequality.

Proof of the Theorem As in classical smoothing spline (Eubank (1988), Wahba (1990)), we can decompose IR into the bias part and the variance part.

$$\begin{aligned}\mathrm{IR}_n(\lambda) &= \int_0^1 [f(t)-\mathrm{E}\{f_\lambda(t)\}]^2\,p(t)\ dt\ +\ \int_0^1 \mathrm{var}\{f_\lambda(t)\}p(t)\ dt \\ &= \mathrm{B}_n^2(\lambda)+\mathrm{V}_n(\lambda).\end{aligned}$$

According to the weighted calculus theory (Grossman, Grossman and Katz (2006)), the design points $t_1,\cdots,t_n$ form a weighted arithmetic partition of $[0,1]$, which means

$$\int_{t_j}^{t_{j+1}} p(t)\,dt=\frac{1}{n}.$$

Therefore, as $n\to\infty$,

$$\begin{aligned}\mathrm{B}_n^2(\lambda) &= \int_0^1 [f(t)-\mathrm{E}\{f_\lambda(t)\}]^2\,p(t)\,dt \\ &= \lim_{n\to\infty}\left[\frac{1}{n}\sum_{j=1}^{n}\{f(t)-\mathrm{E}(f_\lambda(t))\}^2\right]\end{aligned}$$

$$
\begin{aligned}
&\leq \lim_{n\to\infty}\left[\frac{1}{n}\sum_{j=1}^{n}\{f(t)-(g(t))\}^2+\int_0^1\lambda(t)\{g^{(m)}(t)\}^2dt\right]\\
&\leq \int_0^1\lambda(t)\{g_*^{(m)}(t)\}^2dt\\
&\leq \lambda_{max}\int_0^1\{g_*^{(m)}(t)\}^2dt\\
&= O(\lambda_{max}).
\end{aligned}
$$

The third line follows because $\mathrm{E}\{f_\lambda(t)\}$ minimizes the r.h.s. of the third line. The fourth line follows by choosing $g_*(t)$ that interpolates $f(t)$.

Similarly, as $n\to\infty$,

$$
\begin{aligned}
\mathrm{V}_n(\lambda) &= \int_0^1 \mathrm{var}\{f_\lambda(t)\}p(t)\ dt\\
&= \lim_{n\to\infty}\left[\frac{\sigma^2}{n}\mathrm{trace}\{A^2(\lambda(t))\}\right],
\end{aligned}
$$

an increasing function of the individual eigenvalues of $Q_2'\Sigma_\lambda Q_2$ (Wahba 1990, page 55-56), which in turn can be approximated by the eigenvalues of Σ_λ because of Cauchy interlacing theorem. Applying the lemma, the variance is less than or equal to the corresponding variance of classical smoothing spline with smoothing parameter λ_{min}, which means as $n\to\infty$

$$
\mathrm{V}_n(\lambda)\leq O\left(\lambda_{min}^{-1/2m}n^{-1}\right).
$$

Combine the bias and the variance part, let λ_{min} and λ_{max} as $O\left(n^{-2m/(2m+1)}\right)$, then IR decays at the same rate of $n^{-2m/(2m+1)}$ as classical smoothing splines.

S3 References

Eubank, R. L. (1988), *Spline smoothing and nonparametric regression.* Marcel Dekker, Inc.

Fulton, W. (2000), Eigenvalues, invariant factors, highest weights, and Schubert calculus, *Bulletin of the American Mathematical Society* **37**, 209-249.

Grossman, J., Grossman M. and Katz, R (2006). *The first systems of weighted differential and integral calculus.* BookSurge Pulishing.

Wahba, G. (1990). *Spline models for observational data.* CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.