Sobolev penalty:univariate

The Sobolev penalty is

$$P(h) = \int (h^{(r)}(x))^2 dx$$

Suppose $\sup_g \|g\|_\infty \le G$.

We shall suppose for simplicity that the domain is $[0, 1]$.

Suppose we have the function class (so no additional ridge penalty)

$$\hat{\mathcal{G}}(T) = \left\{ \hat{g}(\cdot|\lambda) = \arg\min_{g \in \mathcal{G}} \frac{1}{2}\|y - g\|_T^2 + \lambda P(g) : \lambda \in \Lambda \right\}$$

Using the logic in Example 9.3.2 in Vandegeer, we can express any function in $\mathcal{G}$ as

$$f + g$$

where

$$g = \sum_{k=1}^{r} \alpha_k \psi_k, f = \int_0^1 \beta_u \tilde{\phi}_u$$

where $\langle \psi_k, \tilde{\phi}_u \rangle_T = 0$ and $P(\psi_k) = 0$.

Suppose the observations were drawn from $y = f^* + g^* + \epsilon$.

Now we have the function class

$$\hat{\mathcal{G}}(T) = \left\{ \hat{g}(\cdot|\lambda), \hat{f}(\cdot|\lambda) = \arg\min_{g \in \mathcal{G}} \frac{1}{2}\|y - (f+g)\|_T^2 + \lambda P(f) : \lambda \in \Lambda, g = \sum_{k=1}^{r} \alpha_k \psi_k, f = \int_0^1 \beta_u \tilde{\phi}_u \right\}$$

We will show that

$$\left\| \left(\hat{g}(\cdot|\lambda^{(1)}) + \hat{f}(\cdot|\lambda^{(1)}) \right) - \left(\hat{g}(\cdot|\lambda^{(2)}) + \hat{f}(\cdot|\lambda^{(2)}) \right) \right\|_\infty \le |\lambda^{(1)} - \lambda^{(2)}| n^{\tau_{min}} \sqrt{\frac{n^{\tau_{min}}}{2} \|\epsilon\|_T^2 + P(f^*)} G$$

Proof

First by Vandegeer Example 9.3.2, we know that

$$\hat{g}(\cdot|\lambda) = \arg \min_{g = \sum \alpha_k \psi_k} -2\langle \epsilon, g - g^{*\perp} \rangle_T + \|g - g^*\|_T^2$$

$$\hat{f}(\cdot|\lambda) = \arg \min_{f = \int_0^1 \beta_u \tilde{\phi}_u} -2\langle \epsilon, f - f^* \rangle_T + \|f - f^*\|_T^2 + \lambda P(f)$$

So $\hat{g}(\cdot|\lambda)$ is actually independent of λ and is therefore constant. We will just denote it $\hat{g}$ from now on.

Now consider

$$h = c\left(\hat{f}(\cdot|\lambda^{(1)}) - \hat{f}(\cdot|\lambda^{(2)}) \right)$$

where c is some constant s.t. $P(h) = 1$.

We can assume that $P(h) \neq 0$. Otherwise, if

$$P\left(\hat{f}(\cdot|\lambda^{(1)}) - \hat{f}(\cdot|\lambda^{(2)}) \right) = 0$$

then we know that

$$\hat{f}(\cdot|\lambda^{(1)}) - \hat{f}(\cdot|\lambda^{(2)}) \in span\left\{\psi_k\right\}_{k=1}^r$$

This is true if and only if $\hat{f}(\cdot|\lambda^{(1)}) \equiv \hat{f}(\cdot|\lambda^{(2)})$ (by the fact that the function spaces are orthogonal).
Consider the optimization problem

$$\hat{m}_h(\lambda) = \arg\min_m \frac{1}{2}\|y - (\hat{g} + \hat{f}(\cdot|\lambda^{(1)}) + mh)\|_T^2 + \lambda P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right)$$

By implicit differentiation of the KKT conditions, we get

$$\frac{\partial}{\partial\lambda}\hat{m}_h(\lambda) \quad = \quad -\left(\|h\|_T^2 + \lambda\frac{\partial^2}{\partial m^2}P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right)\right)^{-1} \frac{\partial}{\partial m}P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right)\Bigg|_{m=\hat{m}_\lambda(\lambda_0)}$$

Then the first multiplicand is bounded by

$$\begin{aligned}
\left|\|h\|_T^2 + \lambda\frac{\partial^2}{\partial m^2}P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right)\right|^{-1} &\leq n^{\tau_{min}}\frac{\partial^2}{\partial m^2}P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right)^{-1} \\
&= \frac{n^{\tau_{min}}}{2P(h)}
\end{aligned}$$

The equality follows from the Lemma Sobolev Facts (see below).
From the Lemma Sobolev Facts and by the fact that $P(h) = 1$, we have

$$\begin{aligned}
\left|\frac{\partial}{\partial\lambda}\hat{m}_h(\lambda)\right| &\leq \frac{n^{\tau_{min}}}{P(h)}\sqrt{P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right)P(h)} \\
&= n^{\tau_{min}}\sqrt{P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right)}
\end{aligned}$$

We know that

$$\begin{aligned}
\lambda P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right) &\leq \frac{1}{2}\|y - (\hat{g} + \hat{f}(\cdot|\lambda^{(1)}))\|_T^2 + \lambda P\left(\hat{f}(\cdot|\lambda^{(1)})\right) \\
&\leq \frac{1}{2}\|y - (g^* + f^*)\|_T^2 + \lambda^{(1)}P(f^*) + \left(\lambda - \lambda^{(1)}\right)P\left(\hat{f}(\cdot|\lambda^{(1)})\right)
\end{aligned}$$

and

$$P\left(\hat{f}(\cdot|\lambda^{(1)})\right) \leq \frac{1}{2\lambda^{(1)}}\|y - (g^* + f^*)\|_T^2 + P(f^*)$$

So

$$P\left(\hat{f}(\cdot|\lambda^{(1)}) + mh\right) \leq \frac{n^{\tau_{min}}}{2}\|\epsilon\|_T^2 + P(f^*)$$

Then by the MVT, we have

$$\begin{aligned}
\|\hat{f}(\cdot|\lambda^{(1)}) - \hat{f}(\cdot|\lambda^{(2)})\|_\infty &= \|m_h(\lambda)h\|_\infty \\
&\leq |\lambda^{(1)} - \lambda^{(2)}|\left|\frac{\partial}{\partial\lambda}\hat{m}_h(\lambda)\right|G \\
&\leq |\lambda^{(1)} - \lambda^{(2)}|n^{\tau_{min}}\sqrt{\frac{n^{\tau_{min}}}{2}\|\epsilon\|_T^2 + P(f^*)}G
\end{aligned}$$

Sobolev penalty: multivariate

The function class of interest

$$\hat{\mathcal{G}}(T) = \left\{\left\{\hat{g}_j(\cdot|\lambda), \hat{f}_j(\cdot|\lambda)\right\} = \arg\min_{g\in\mathcal{G}} \frac{1}{2}\|y - \sum_{j=1}^J g_j(x_j)\|_T^2 + \sum_{j=1}^J \lambda_j P(g_j) : \lambda \in \Lambda\right\}$$

We conclude the same thing

Proof

First by Vandegeer Example 9.3.2, we know that

$$\{\hat{g}_j(\cdot|\lambda)\} = \arg \min_{g_j=\sum \alpha_k \psi_k} -2\langle \epsilon, \sum_{j=1}^J g_j - g_j^* \rangle_T + \| \sum_{j=1}^J g_j - g_j^* \|_T^2$$

$$\left\{ \hat{f}_j(\cdot|\lambda) \right\} = \arg \min_{f_j=\int_0^1 \beta_u \tilde{\phi}_u} -2\langle \epsilon, \sum_{j=1}^J f_j - f_j^* \rangle_T + \| \sum_{j=1}^J f_j - f_j^* \|_T^2 + \sum_{j=1}^J \lambda_j P(f_j)$$

So $\hat{g}(\cdot|\lambda)$ is actually independent of λ and is therefore constant. We will just denote it $\hat{g}_j$ from now on.

Now consider

$$h_j = c\left(\hat{f}_j(\cdot|\lambda^{(1)}) - \hat{f}_j(\cdot|\lambda^{(2)}) \right)$$

where c is some constant s.t. $P(h_j) = 1$.

We can assume that $P(h_j) \neq 0$. Otherwise, if

$$P\left(\hat{f}_j(\cdot|\lambda^{(1)}) - \hat{f}_j(\cdot|\lambda^{(2)}) \right) = 0$$

then we know that

$$\hat{f}_j(\cdot|\lambda^{(1)}) - \hat{f}_j(\cdot|\lambda^{(2)}) \in span\,\{\psi_k\}_{k=1}^r$$

This is true if and only if $\hat{f}_j(\cdot|\lambda^{(1)}) \equiv \hat{f}_j(\cdot|\lambda^{(2)})$ (by the fact that the function spaces are orthogonal).

Now consider the optimization problem

$$\{\hat{m}_j(\lambda, h)\} = \arg \min_{m_j} \frac{1}{2} \|y - \sum_{j=1}^J (\hat{g}_j + \hat{f}_j(\cdot|\lambda^{(1)}) + m_j h_j)\|_T^2 + \lambda P\left(\hat{f}_j(\cdot|\lambda^{(1)}) + m_j h_j \right)$$

(If $h_j \equiv 0$, then set $m_j = 0$ as a constant.) For simplicity, we will assume $h_j \neq 0$.

By implicit differentiation of the KKT conditions, we get for all $\ell = 1 : J$

$$\frac{\partial}{\partial \lambda_\ell} \hat{m}_\ell(\lambda, h) = -\left(\|h\|_T^2 + \sum_{j=1}^J \lambda_j \frac{\partial^2}{\partial m_j^2} P\left(\hat{f}_j(\cdot|\lambda^{(1)}) + m_j h_j \right) \right)^{-1} \frac{\partial}{\partial m_\ell} P\left(\hat{f}_\ell(\cdot|\lambda^{(1)}) + m_\ell h \right) \Bigg|_{m=\hat{m}(\lambda,h)}$$

and

$$\frac{\partial}{\partial \lambda_k} \hat{m}_\ell(\lambda, h) = 0 \text{ if } \ell \neq k$$

From the Lemma Sobolev Facts, we have

$$\begin{aligned}
\left| \frac{\partial}{\partial \lambda} \hat{m}_h(\lambda) \right| &\leq \frac{n^{\tau_{min}}}{P(h_\ell)} \sqrt{P\left(\hat{f}_\ell(\cdot|\lambda^{(1)}) + m_\ell h_\ell \right) P(h_\ell)} \\
&= n^{\tau_{min}} \sqrt{P\left(\hat{f}_\ell(\cdot|\lambda^{(1)}) + m_\ell h_\ell \right)}
\end{aligned}$$

We know that

$$\begin{aligned}
\lambda_\ell P\left(\hat{f}_\ell(\cdot|\lambda^{(1)}) + m_\ell h_\ell \right) &\leq \frac{1}{2} \|y - (\hat{g} + \hat{f}(\cdot|\lambda^{(1)}))\|_T^2 + \sum_{j=1}^J \lambda_j P\left(\hat{f}(\cdot|\lambda^{(1)}) \right) \\
&\leq \frac{1}{2} \|y - (g^* + f^*)\|_T^2 + \sum_{j=1}^J \lambda_j^{(1)} P(f^*) + \sum_{j=1}^J \left(\lambda_j - \lambda_j^{(1)} \right) P\left(\hat{f}(\cdot|\lambda^{(1)}) \right)
\end{aligned}$$

and

$$P\left(\hat{f}_\ell(\cdot|\lambda^{(1)})\right) \le \frac{1}{2\lambda_\ell^{(1)}} \|y - (g^* + f^*)\|_T^2 + \sum_{j=1}^J \lambda_j^{(1)} P\left(f^*\right)$$

So

$$P\left(\hat{f}_\ell(\cdot|\lambda^{(1)}) + m_\ell h_\ell\right) \le \frac{n^{\tau_{min}}}{2} \|\epsilon\|_T^2 + n^{t_{\max}+t_{\min}} \sum_{j=1}^J P\left(f^*\right)$$

Then by the MVT, we have

$$\begin{aligned}
\|\hat{f}_\ell(\cdot|\lambda^{(1)}) - \hat{f}_\ell(\cdot|\lambda^{(2)})\|_\infty &= \|\hat{m}_\ell(\lambda, h) h_\ell\|_\infty \\
&\le \left\|\lambda^{(1)} - \lambda^{(2)}\right\| \|\nabla_\lambda \hat{m}_\ell(\lambda, h)\| \, G \\
&\le \left\|\lambda^{(1)} - \lambda^{(2)}\right\| n^{\tau_{min}} \sqrt{\frac{n^{\tau_{min}}}{2} \|\epsilon\|_T^2 + n^{t_{\max}+t_{\min}} \sum_{j=1}^J P\left(f^*\right) G}
\end{aligned}$$

Lemma: Sobolev Facts

For any function h, we have

$$\begin{aligned}
\left|\frac{\partial}{\partial m} P(g + mh)\right| &= \left|2 \int (g^{(r)}(x) + m h^{(r)}(x)) h^{(r)}(x) dx\right| \\
&\le 2\sqrt{P(g + mh) P(h)}
\end{aligned}$$

and

$$\frac{\partial^2}{\partial m^2} P(g + mh) = 2 \int (h^{(r)}(x))^2 dx = 2P(h)$$