0.1 Another cute lemma?

Lemma 1

Suppose $I : \mathcal{G} \mapsto [0, \infty)$ is a pseudonorm. So $I(g) + I(f) \geq I(g + f)$ and $I(cg) = cI(g) \forall c > 0$. Then

$$\left| \frac{\partial}{\partial t} I(g + th) \right| \leq I(h)$$

Proof:
By the triangle inequality for I, we have

$$|I(g + th) - I(g)| \leq I(th)$$

then diving by t and taking the limit, we have

$$\frac{\partial}{\partial t} I(g + th) = \lim_{t \to 0} \frac{|I(g + th) - I(g)|}{t} \leq \lim_{t \to 0} \frac{I(th)}{t} = I(h)$$

Lemma 2

Suppose the conditions in lemma 1, as well as

$$I^v(g) \leq M \|g\|_n^2 + M_0$$

For any $\lambda < \tilde{\lambda}$,

$$\left\| \hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda \right\|_n^2 = O_p(1) v \tilde{\lambda}^2 I^v(\hat{g}_\lambda)$$

Proof:
Since $\hat{g}_\lambda$ is the minimizer of the penalized criterion, then for h s.t. $\hat{g}_\lambda + th \in \mathcal{G} \forall t > 0$

$$\frac{\partial}{\partial t} \left(\|y - (\hat{g}_\lambda + th)\|_n^2 + \lambda^2 I^v(\hat{g}_\lambda + th) \right) \Big|_{t=0} = 0$$

But we also know that

$$\frac{\partial}{\partial t} \left(\|y - (\hat{g}_\lambda + th)\|_n^2 + \lambda^2 I(\hat{g}_\lambda + th) \right) \Big|_{t=0} = -2 \left(y - \hat{g}_\lambda, h \right)_T + \lambda^2 v I^{v-1}(\hat{g}_\lambda) \frac{\partial}{\partial t} I(\hat{g}_\lambda + th)$$

Then for $\lambda < \tilde{\lambda}$, we have

$$-2 \left(y - \hat{g}_\lambda, h \right) + \lambda^2 v I^{v-1}(\hat{g}_\lambda) \frac{\partial}{\partial t} I(\hat{g}_\lambda + th) = -2 \left(y - \hat{g}_{\tilde{\lambda}}, h \right)_T + \tilde{\lambda}^2 v I^{v-1}(\hat{g}_{\tilde{\lambda}}) \frac{\partial}{\partial t} I(\hat{g}_{\tilde{\lambda}} + th)$$

Rearranging, we get

$$\begin{aligned} 0 &= 2 \left(\hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda, h \right)_T + v \left(\tilde{\lambda}^2 I^{v-1}(\hat{g}_{\tilde{\lambda}}) \frac{\partial}{\partial t} I(\hat{g}_{\tilde{\lambda}} + th) - \lambda^2 I^{v-1}(\hat{g}_\lambda) \frac{\partial}{\partial t} I(\hat{g}_\lambda + th) \right) \\ &\leq 2 \left(\hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda, h \right)_T + v I^{v-1}(\hat{g}_\lambda) I(h) \tilde{\lambda}^2 \end{aligned}$$

where the first inequality follows from Lemma 1. Setting $h = \hat{g}_\lambda - \hat{g}_{\tilde{\lambda}}$, we get

$$\left\| \hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda \right\|_n^2 = v \tilde{\lambda}^2 I^{v-1}(\hat{g}_\lambda) I \left(\hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda \right)$$

Since

$$I \left(\hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda \right) \leq I(\hat{g}_{\tilde{\lambda}}) + I(\hat{g}_\lambda)$$

then

$$\left\| \hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda \right\|_n^2 = O_p(1) v \tilde{\lambda}^2 I^v(\hat{g}_\lambda)$$

(Note that if we assume $I^v \left(\hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda \right) \leq O_P(1) \left\| \hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda \right\|_n^2 + C$, then $\left\| \hat{g}_{\tilde{\lambda}} - \hat{g}_\lambda \right\|_n^{2-2/v} = O_p(1) v I^{v-1}(\hat{g}_\lambda) \tilde{\lambda}^2 + C$.)

Lemma 3

Suppose

$$H\left(\delta, \{g : I(g) \le 1\}, \|\cdot\|_n\right) \le A\delta^{-\alpha}$$

and

$$\frac{\|g\|_n}{I(g)} \le K < \infty$$

Suppose for some $\hat{\lambda} \le \tilde{\lambda}$, $\left\|\hat{g}_{\tilde{\lambda}} - \hat{g}_{\hat{\lambda}}\right\|_n$ is bounded (doesn't grow with n).
Suppose

$$I^v(\hat{g}_{\hat{\lambda}}) \le M\|\hat{g}_{\hat{\lambda}}\|_n^2 + M_0$$

(Note: this is probably not necessary, but I'm sleepy)
Then

$$H\left(\delta, \left\{\hat{g}_\lambda : \lambda \ge \hat{\lambda}\right\}, \|\cdot\|_n\right) \le A\delta^{-\alpha}$$

Proof:
Let $I(\hat{g}_{\hat{\lambda}}) = R$. We'll suppose $R \ge 1$. Otherwise, we'll be done.
By the assumptions, R is bounded since

$$I^v(\hat{g}_{\hat{\lambda}}) \le M\|\hat{g}_{\hat{\lambda}}\|_n^2 + M_0 \le M\left(\|\hat{g}_{\hat{\lambda}} - \hat{g}_{\tilde{\lambda}}\|_n + \|\hat{g}_{\tilde{\lambda}}\|_n\right)^2 + M_0$$

Note that $\left\{\hat{g}_\lambda : \lambda \ge \hat{\lambda}\right\} \subseteq \{g : I(g) \le R\} = \{Rg : I(g) \le 1\}$. Note that if h is the closest function in the δ-cover for $\{g : I(g) \le 1\}$, then

$$\|g - h\|_n \le \delta \implies \|Rg - \delta\lfloor\frac{R}{\delta}\rfloor h\|_n \le \|Rg - Rh\|_n + \delta\|h\|_n \le \delta(1 + R + K)$$

Then for some constant $\tilde{A}$ dependent on A, R, K,

$$H\left(\delta, \left\{\hat{g}_\lambda : \lambda \ge \hat{\lambda}\right\}, \|\cdot\|_n\right) \le H\left(\delta, \{Rg : I(g) \le 1\}, \|\cdot\|_n\right) \le \tilde{A}\delta^{-\alpha}$$

Lemma 4

Suppose

$$H\left(\delta, \{\hat{g}_\lambda : \lambda \in \Lambda\}, \|\cdot\|_n\right) \le A\delta^{-\alpha}$$

Suppose $\hat{\lambda}$ is the CV-fitted lambda and $\tilde{\lambda}$ is the oracle lambda given in Vandegeer.
Suppose

$$\left\|g^* - \hat{g}_{\tilde{\lambda}}\right\|_V = O_p(1)\left\|g^* - \hat{g}_{\tilde{\lambda}}\right\|_T$$

Then

$$\left\|\hat{g}_{\tilde{\lambda}} - \hat{g}_{\hat{\lambda}}\right\|_V = O_p(n^{-1/(2+\alpha)})$$

Proof:
The basic inequality gives us

$$\left\|\hat{g}_{\tilde{\lambda}} - \hat{g}_{\hat{\lambda}}\right\|_V^2 \le 2\left|\left(\epsilon, \hat{g}_{\tilde{\lambda}} - \hat{g}_{\hat{\lambda}}\right)_V\right| + 2\left|\left(g^* - \hat{g}_{\tilde{\lambda}}, \hat{g}_{\tilde{\lambda}} - \hat{g}_{\hat{\lambda}}\right)_V\right|$$

If $\left|\left(\epsilon, \hat{g}_{\tilde{\lambda}} - \hat{g}_{\hat{\lambda}}\right)_T\right|$ is the bigger term, then

$$\left\|\hat{g}_{\tilde{\lambda}} - \hat{g}_{\hat{\lambda}}\right\|_V^2 \le O_P(1)\left|\left(\epsilon, \hat{g}_{\tilde{\lambda}} - \hat{g}_{\hat{\lambda}}\right)_V\right|$$

Use the same arguments as Thrm 9.1 to show that $\left\|\hat{g}_{\tilde{\lambda}}-\hat{g}_{\hat{\lambda}}\right\|_V = O_p(n^{-1/2})$.

If $\left|\left(g^*-\hat{g}_{\tilde{\lambda}}, \hat{g}_{\tilde{\lambda}}-\hat{g}_{\hat{\lambda}}\right)_T\right|$ is the bigger term, then by Cauchy Schwarz

$$\begin{aligned}\left\|\hat{g}_{\tilde{\lambda}}-\hat{g}_{\hat{\lambda}}\right\|_V^2 &\leq O_p(1)\left|\left(g^*-\hat{g}_{\tilde{\lambda}}, \hat{g}_{\tilde{\lambda}}-\hat{g}_{\hat{\lambda}}\right)_V\right| \\ &\leq O_p(1)\left\|g^*-\hat{g}_{\tilde{\lambda}}\right\|_V\left\|\hat{g}_{\tilde{\lambda}}-\hat{g}_{\hat{\lambda}}\right\|_V\end{aligned}$$

Hence

$$\left\|\hat{g}_{\tilde{\lambda}}-\hat{g}_{\hat{\lambda}}\right\|_V \leq O_p(n^{-1/(2+\alpha)})$$

Need to show:

If we show that $I(\hat{g}_{\hat{\lambda}}) = O_p(n^{\frac{2}{(2+\alpha)v}}) \leq O_p(n^{1/\alpha})$, then by Lemma 2, 3, 4, we have that $\left\|\hat{g}_{\tilde{\lambda}}-\hat{g}_{\hat{\lambda}}\right\|_V \leq O_p(n^{-1/(2+\alpha)})$.

It might be easier to show that for many penalties, $I(\hat{g}_{\lambda=0}) = O_p(n^{\frac{2}{(2+\alpha)v}})$???

OLD

Lemma 3:

Suppose the function class $\mathcal{F}$ is bounded s.t. $\sup_{f\in\mathcal{F}} \|f\|_n \leq R < \infty$. Let

$$\tilde{\mathcal{F}} = \{\gamma f \ : \ f \in \mathcal{F}, \gamma \in (0,1]\}$$

$$H\left(\delta(1+R+\delta), \tilde{\mathcal{F}}, \|\cdot\|_n\right) \leq \log(1 + \lfloor\frac{1}{\delta}\rfloor) + H\left(\delta, \mathcal{F}, \|\cdot\|_n\right)$$

Proof: Let $\{h_i\}_{i=1}^N$ be the δ-cover for $\mathcal{F}$. Consider any $f \in \mathcal{F}$ and let $h_{(f)}$ be the closest function in δ-cover for $\mathcal{F}$. Choose $j \in \mathbb{Z}^+$ such that $|\gamma - \delta j| < \delta$.

$$\begin{aligned}
\|\gamma f - \delta j h_{(f)}\|_n &\leq \|\gamma f - \gamma h_{(f)}\|_n + \|\gamma h_{(f)} - \delta j h_{(f)}\|_n \\
&\leq \gamma\|f - h_{(f)}\|_n + |\gamma - \delta j|\|h_{(f)}\|_n \\
&\leq \gamma\delta + \delta\left(\|f - h_{(f)}\|_n + \|f\|_n\right) \\
&\leq \gamma\delta + \delta\left(\delta + R\right) \\
&\leq \delta\left(1 + R + \delta\right)
\end{aligned}$$

Hence we have found that the following $N(1 + \lfloor\frac{1}{\delta}\rfloor)$ functions form a $\delta(1+R+\delta)$-cover for $\tilde{\mathcal{F}}$:

$$\{h_i\}_{i=1}^N \cup \left\{ j\delta h_i \ : \ j \in 1 : \lfloor\frac{1}{\delta}\rfloor, i \in 1 : N \right\}$$

Lemma 4:

Define function classes $\{\mathcal{F}_j\}_{j=1}^J$ and

$$\tilde{\mathcal{F}} = \left\{ \sum_{j=1}^J f_j \ : \ f_j \in \mathcal{F}_j \right\}$$

Then

$$H\left(J\delta, \tilde{\mathcal{F}}, \|\cdot\|_n\right) \leq \sum_{j=1}^J H\left(\delta, \mathcal{F}_j, \|\cdot\|_n\right)$$

Proof: For every $j = 1 : J$, consider any $f_j \in \mathcal{F}_j$ and let $h_{(j)}$ be the closest function in the δ-cover for $\mathcal{F}_j$.

$$\|\sum_{j=1}^J f_j - \sum_{j=1}^J h_{(j)}\| \quad \leq \quad \sum_{j=1}^J \|f_j - h_{(j)}\| \leq J\delta$$

Hence $\exp\left(\sum_{j=1}^J H\left(\delta, \mathcal{F}_j, \|\cdot\|_n\right)\right)$ functions form a $J\delta$-cover for $\tilde{\mathcal{F}}$.

Lemma 5:

Suppose for all $j = 1, ..., J$, there is some $\alpha_j > 0$ and $A_j > 0$ s.t. the following entropy bound holds for all $\delta > 0$

$$H\left(\delta, \left\{\frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|_T\right) \leq A\delta^{-\alpha_j}$$

Then for sufficiently small$\delta > 0$, we have

$$H\left(\delta, \left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sup_{j\in 1:J}\left(I(g_j) + I(g_j^*)\right)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|_T\right) \leq 2JA\left(\frac{\delta}{2J\left(1+R\right)}\right)^{-\alpha_{max}}$$

where $\alpha_{max} = \max_{j\in 1:J} \alpha_j$.

Proof: By Lemma 3,

$$H\left(\delta(1+R+\delta), \left\{\gamma\frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} \ : \ g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0, \gamma \in (0,1]\right\}, \|\cdot\|_T\right) \leq \log(1+\lfloor\frac{1}{\delta}\rfloor)+A\delta^{-\alpha_j}$$

Note that

$$\frac{\sum_{j=1}^J g_j - g_j^*}{\sup_{j\in 1:J}\left(I(g_j) + I(g_j^*)\right)} = \sum_{j=1}^J \left(\frac{I(g_j) + I(g_j^*)}{\sup_{\ell\in 1:J} I(g_\ell) + I(g_\ell^*)}\right)\frac{g_j - g_j^*}{I(g_j) + I(g_j^*)}$$

By Lemma 4,

$$H\left(J\delta(1+R+\delta), \left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sup_{j\in 1:J}\left(I(g_j) + I(g_j^*)\right)} \ : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|_T\right) \leq J\log(1+\lfloor\frac{1}{\delta}\rfloor)+JA\delta^{-\alpha_j}$$

Hence for sufficiently small δ,

$$H\left(J\delta(1+R+\delta), \left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sup_{j\in 1:J}\left(I(g_j) + I(g_j^*)\right)} \ : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|_T\right) \leq 2JA\delta^{-\alpha_{max}}$$

Rearranging, we get

$$\begin{aligned} H\left(\delta, \left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sup_{j\in 1:J}\left(I(g_j) + I(g_j^*)\right)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|_T\right) &\leq 2AJ\left(\sqrt{\left(\frac{1+R}{2}\right)^2 + \frac{\delta}{J}} - \frac{1+R}{2}\right)^{-\alpha_{max}} \\ &\leq 2AJ\left(\frac{\delta}{2J\left(1+R\right)}\right)^{-\alpha_{max}} \end{aligned}$$

(Used the fact that for $b > 0$ small enough, $\sqrt{a^2+b} - a \geq \sqrt{(a+\frac{b}{4a})^2} - a = \frac{b}{4a}$)

Lemma 5b:

Suppose for all $j = 1, ..., J$, there is some $\alpha_j > 0$ and $A_j > 0$ s.t. the following entropy bound holds for all $\delta > 0$

$$H\left(\delta, \left\{\frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|\right) \leq A\delta^{-\alpha_j}$$

Then for sufficiently small$\delta > 0$, we have

$$H\left(\delta, \left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sum_{j=1}^J I(g_j) + I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|\right) \leq 2JA\,(STUFF)^{-\alpha_{max}}$$

where $\alpha_{max} = \max_{j\in 1:J} \alpha_j$.

Proof: Note that

$$\frac{\sum_{j=1}^J g_j - g_j^*}{\sum_{j=1}^J I(g_j) + I(g_j^*)} = \sum_{j=1}^J \left(\frac{I(g_j) + I(g_j^*)}{\sum_{j=1}^J I(g_\ell) + I(g_\ell^*)}\right) \frac{g_j - g_j^*}{I(g_j) + I(g_j^*)}$$

So we can express

$$\left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sum_{j=1}^J I(g_j) + I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\} \subseteq \left\{\sum_{j=1}^J \gamma_j \frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0, \sum_{j=1}^J \gamma_j = 1\right\}$$

Let $\mathcal{H}_j$ be the set of functions that form a δ-cover for $\left\{\frac{g_j - g_j^*}{I(g_j)+I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}$. Consider the set of functions

$$\left\{\sum_{j=1}^J \delta k_j h_j \ : \ h_j \in \mathcal{H}_j, 1 - \frac{1}{\delta} \leq \delta {\textstyle\sum_{j=1}^J} k_j \leq 1, k_j \in 1 : \lfloor\frac{1}{\delta}\rfloor\right\}$$

Let $|\delta k_j - \gamma_i| < \delta/2$. Then

$$\begin{aligned}
\left\|\sum_{j=1}^J \gamma_j \frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} - \sum_{j=1}^J \delta k_j h_j\right\| &\leq \sum_{j=1}^J \left\|\gamma_j \frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} - \delta k_j h_j\right\| \\
&\leq \sum_{j=1}^J \left\|\gamma_j \frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} - \gamma_i h_j\right\| + |\delta k_j - \gamma_i|\|h_j\| \\
&\leq \sum_{j=1}^J \left(\gamma_j \delta + \frac{\delta}{2}\left(\left\|\frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} - h_j\right\| + \left\|\frac{g_j - g_j^*}{I(g_j) + I(g_j^*)}\right\|\right)\right) \\
&\leq \delta(1 + JR + J\delta)
\end{aligned}$$

Hence these $\left(\Pi_{j=1}^J N_j\right)\binom{\lfloor\frac{1}{\delta}\rfloor + J - 1}{J-1}$ functions form a $\delta(1 + JR + J\delta)$ cover. Hence the entropy is

$$H\left(\delta(1 + JR + J\delta), \left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sum_{j=1}^J I(g_j) + I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|\right) \leq (J-1)\log(1+J+\lfloor\frac{1}{\delta}\rfloor) + A\sum_{j=1}^J \delta^{-\alpha_j}$$

Note:

$$\binom{\lfloor\frac{1}{\delta}\rfloor + J - 1}{J - 1} \leq \left(\lfloor\frac{1}{\delta}\rfloor + J - 1\right)^{J-1}$$

Hence for sufficiently small δ,

$$H\left(\delta(1+JR+J\delta), \left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sum_{j=1}^J I(g_j)+I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j)+I(g_j^*)>0\right\}, \|\cdot\|\right) \le 2JA\delta^{-\alpha_{max}}$$

Rearranging, we get

$$\begin{aligned} H\left(\delta, \left\{\frac{\sum_{j=1}^J g_j - g_j^*}{\sum_{j=1}^J I(g_j)+I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j)+I(g_j^*)>0\right\}, \|\cdot\|\right) &\le 2AJ\left(\frac{-JR+1+\sqrt{(JR+1)^2+4\delta J}}{2J}\right)^{-\alpha_{max}} \\ &\le 2AJ\left(\frac{\sqrt{2}\delta J^{3/2}}{1+JR}\right)^{-\alpha_{max}} \end{aligned}$$

(Used the fact that for $b>0$ small enough, $\sqrt{a^2+b}-a \ge \sqrt{(a+\frac{b}{4a})^2}-a = \frac{b}{4a}$)

Lemma 6:

Suppose ϵ_i are sub-gaussian errors and for the function class $\mathcal{F}$, we have that for some $0<\alpha<2$, $A'>0$, and $J>0$

$$H(\delta, \mathcal{F}, \|\cdot\|_T) \le A' J^\tau \delta^{-\alpha} \ \forall \delta > 0$$

Then for $T = 2C_1 C A'^{1/2} J^{\tau/2} 2^{1-\alpha/2}$

$$Pr\left(\sup_{f\in\mathcal{F}} \frac{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i f(z_i)\right|}{\|f\|_n^{1-\alpha/2}} \ge T\right) \le c\exp(-T^2/c^2)$$

Proof: Follow proof for Lemma 8.4 in Vandegeer, but with $A = A'J^{-\alpha}$. Note that we then have $A_0 = A'^{1/2}J^{\tau/2}$. We then get

$$Pr\left(\sup_{f\in\mathcal{F}} \frac{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i f(z_i)\right|}{\|f\|_n^{1-\alpha/2}} \ge 2C_1 C A'^{1/2} J^{\tau/2} 2^{1-\alpha/2}\right) \le c\exp(-T^2/c^2)$$

Note that we can write via shorthand that

$$\sup_{f\in\mathcal{F}} \frac{\left|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(z_i)\right|}{\|f\|_n^{1-\alpha/2}} = O_p(J^{\tau/2} n^{-1/2})$$

Result 3: Single λ, Multiple Penalties, Optimal $\tilde{\lambda}_T$ over X_T

Consider function classes $\mathcal{G}_j$ that are cones. Also, suppose we have an additive model:

$$y = \sum_{j=1}^J g_j^* + \epsilon$$

where $g_j^* \in \mathcal{G}_j$.
We fit the model by least squares with separate penalties for each function g_j:

$$\{\hat{g}_j\}_{j=1}^J = \arg\min_{g_j\in\mathcal{G}_j} \|y - \sum_{j=1}^J g_j\|_T^2 + \lambda^2 \sum_{j=1}^J I_j^{v_j}(g_j)$$

Suppose $v_j \geq 1$ for all j. (This requirement on v_j stricter than Vandegeer Thrm 10.2)
Suppose for all j, there is some $0 < \alpha < 2$ s.t. for all $\delta > 0$,

$$H\left(\delta, \{g_j \in \mathcal{G}_j : I(g_j) \leq 1\}, \|\cdot\|_T\right) \leq A\delta^{-\alpha}$$

and that for all j

$$\sup_{g_j \in \mathcal{G}_j} \frac{\|g_j\|_T}{I(g_j)} \leq R < \infty$$

If we choose λ s.t.

$$\tilde{\lambda}_T^{-1} = O_p\left(n^{1/(2+\alpha)}\right)\left(J + \sum_{j=1}^{J} I_j^{v_j}(g_j^*)\right)^{(2-\alpha)/2(2+\alpha)}$$

then

$$\|\sum_{j=1}^{J} g_j - g_j^*\|_T = O_p\left(\tilde{\lambda}_T\right) J \left(\sum_{j=1}^{J} I_j^{v_j}(g_j^*)\right)^{1/2}$$

and

$$\sum_{j=1}^{J} I_j(\hat{g}_j) \leq O_p(J)\left(J + \sum_{j=1}^{J} I_j^{v_j}(g_j^*)\right)$$

Proof:

The basic inequality gives us:

$$\|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T^2 + \lambda^2 \sum_{j=1}^{J} I_j^{v_j}(\hat{g}_j) \leq 2\left|\left(\epsilon_T, \sum_{j=1}^{J} \hat{g}_j - g_j^*\right)\right| + \lambda^2 \sum_{j=1}^{J} I_j^{v_j}\left(g_j^*\right)$$

Case 1: $\left|\left(\epsilon_T, \sum_{j=1}^{J} \hat{g}_j - g_j^*\right)\right| \leq \lambda^2 \sum_{j=1}^{J} I_j^{v_j}\left(g_j^*\right)$

$$\|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T \leq O_p(\lambda)\left(\sum_{j=1}^{J} I_j^{v_j}\left(g_j^*\right)\right)^{1/2}$$

Case 2: $\left|\left(\epsilon_T, \sum_{j=1}^{J} \hat{g}_j - g_j^*\right)\right| \geq \lambda^2 \sum_{j=1}^{J} I_j^{v_j}\left(g_j^*\right)$
By Lemma 3,

$$H\left(\delta, \left\{\frac{\sum_{j=1}^{J} g_j - g_j^*}{\sum_{j=1}^{J} I(g_j) + I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|_T\right) \leq \tilde{A} J^{1-\alpha}\delta^{-\alpha}$$

Hence by (10.6) in Vandegeer,

$$\sup_{g_j \in \mathcal{G}_j} \frac{\left|\left(\epsilon_T, \sum_{j=1}^{J} g_j - g_j^*\right)\right|}{\|\sum_{j=1}^{J} g_j - g_j^*\|^{1-\alpha/2}\left(\sum_{j=1}^{J} I(g_j) + I(g_j^*)\right)^{\alpha/2}} = O_p\left(n^{-1/2}\right) J^{1-\alpha}$$

and the basic inequality becomes

$$\|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T^2 + \lambda^2 \sum_{j=1}^{J} I_j^{v_j}(\hat{g}_j) \leq O_p\left(n^{-1/2}\right) J^{1-\alpha} \|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T^{1-\alpha/2}\left(\sum_{j=1}^{J} I(\hat{g}_j) + I(g_j^*)\right)^{\alpha/2}$$

Case 2a: Suppose $\sum_{j=1}^{J} I(\hat{g}_j) \leq \sum I(g_j^*)$.
Then

$$\|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T \quad \leq \quad O_p\left(n^{-1/(2+\alpha)}\right) J^{\frac{2(1-\alpha)}{\alpha+2}} \left(\sum_{j=1}^{J} I(g_j^*)\right)^{\alpha/(2+\alpha)}$$

Case 2b: Suppose $\sum_{j=1}^{J} I(\hat{g}_j) \geq \sum I(g_j^*)$.
First note that by assuming $v_j \geq 1$, we must have $I_j(\hat{g}_j) \leq I_j^{v_j}(\hat{g}_j) + 1$. So

$$\begin{aligned}
\sum_{j=1}^{J} I_j(\hat{g}_j) &\leq J + \sum_{j=1}^{J} I_j^{v_j}(\hat{g}_j) \\
&\leq J + O_p\left(n^{-1/2}\right) J^{1-\alpha} \lambda^{-2} \|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T^{1-\alpha/2} \left(\sum_{j=1}^{J} I(\hat{g}_j)\right)^{\alpha/2}
\end{aligned}$$

Case 2ba: If the second term on the RHS in the inequality above is bigger, then

$$\sum_{j=1}^{J} I_j(\hat{g}_j) \leq O_p\left(n^{-1/(2-\alpha)}\right) J^{2(1-\alpha)/(2-\alpha)} \lambda^{-4/(2-\alpha)} \|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T$$

which implies

$$\|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T \leq O_p\left(n^{-1/(2-\alpha)}\right) J^{\alpha(1-\alpha)/(2-\alpha)} \lambda^{-2\alpha/(2-\alpha)}$$

and

$$\sum_{j=1}^{J} I_j(\hat{g}_j) \leq J^{(\alpha+2)(1-\alpha)/(2-\alpha)} \left(J + \sum_{j=1}^{J} I_j^{v_j}(g_j^*)\right)$$

Case 2bb: If the first term on the RHS in the inequality above is bigger, then

$$\sum_{j=1}^{J} I_j(\hat{g}_j) \leq J \implies \|\sum_{j=1}^{J} \hat{g}_j - g_j^*\|_T \leq O_p\left(n^{-1/(2+\alpha)}\right) J^{(2-\alpha)/(2+\alpha)}$$

Result 5: Multiple λ, Multiple Penalties, Optimal λ on X_T

Consider an additive model:

$$y = \sum_{j=1}^{J} g_j^* + \epsilon$$

We fit the model by least squares with separate penalties and separate λ for each function g_j:

$$\{\hat{g}_j\}_{j=1}^{J} = \arg\min_{g_j \in \mathcal{G}_j} \|y - \sum_{j=1}^{J} g_j\|_T^2 + \frac{1}{J} \sum_{j=1}^{J} \lambda_j^2 I_j^{v_j}(g_j)$$

Suppose $v_j > \frac{2\alpha_j}{2+\alpha_j}$ for all j.

Suppose for all j, there is some $0 < \alpha_j < 2$ s.t. for all $\delta > 0$,

$$H\left(\delta, \left\{\frac{g_j - g_j^*}{I(g_j) + I(g_j^*)} : g_j \in \mathcal{G}_j, I(g_j) + I(g_j^*) > 0\right\}, \|\cdot\|_T\right) \leq \frac{A}{J}\delta^{-\alpha_j}$$

and for all j,

$$\sup_{g_j \in \mathcal{G}_j} \frac{\|g_j - g_j^*\|_T}{I(g_j) + I(g_j^*)} \leq R < \infty$$

If we choose λ s.t.

$$\tilde{\lambda}_j^{-1} = ???$$

then

$$\|\sum_{j=1}^{J} g_j - g_j^*\|_T = ???$$

and

$$\sum_{j=1}^{J} I^{v_j}(\hat{g}_{\lambda,j}) = ???$$

Example 1: Sobelov norm (NOT DONE)

Consider the functions

$$\mathcal{G} = \left\{ g : [0,1] \mapsto \mathbb{R} : \int_0^1 g^{(m)}(z)^2 dz < \infty \right\}$$

Suppose x_i are all unique. Then the Sobelov norm for the class $\{\hat{g}_\lambda \in \mathcal{G} : \lambda \in \Lambda\}$ is bounded above by its $L_2(P_n)$ norm.

$$I^2(\hat{g}_\lambda) = \int_0^1 \left(\hat{g}_\lambda^{(m)}(z) \right)^2 dz \leq 2\|\hat{g}_\lambda\|_n^2 + 4I^2(\tilde{g}) + 4\|y\|_n^2 \ \forall \lambda \in \Lambda$$

PROBLEM: as defined, it is possible that $I^2(\tilde{g})$ grows with n, which is not okay!

Proof:

Let $\tilde{g}$ satisfy $\tilde{g}(x_i) = y_i$ and have the smallest value for $\int_0^1 \left(\tilde{g}^{(m)}(z) \right)^2 dz$. This function $\tilde{g}$ should always exist.

Case 1: $\lambda \leq 1/2$

By definition of $\hat{g}_\lambda$

$$\|y - \hat{g}_\lambda\|_n^2 + \lambda^2 I^2(\hat{g}_\lambda) \leq \|y - (\tilde{g} - \lambda \hat{g}_\lambda)\|_n^2 + \lambda^2 I^2(\tilde{g} - \lambda \hat{g}_\lambda)$$

Note that

$$\begin{aligned} I^2(\tilde{g} - \lambda \hat{g}_\lambda) &= \int_0^1 \left(\tilde{g}^{(m)} - \lambda \hat{g}_\lambda^{(m)} \right)^2 dz \\ &= 2 \int_0^1 \max \left(\left| \tilde{g}^{(m)} \right|^2, \left| \lambda \hat{g}_\lambda^{(m)} \right|^2 \right) dz \\ &= 2 \left(\int_0^1 \left| \tilde{g}^{(m)} \right|^2 dz + \int_0^1 \left| \lambda \hat{g}_\lambda^{(m)} \right|^2 dz \right) \end{aligned}$$

Hence

$$\lambda^2 I^2(\hat{g}_\lambda) \leq \lambda^2 \|\hat{g}_\lambda\|_n^2 + 2\lambda^2 I^2(\tilde{g}) + 2\lambda^4 I^2(\hat{g}_\lambda)$$

The following ineq follows, where the RHS is maximized when $\lambda = 1/2$

$$I^2(\hat{g}_\lambda) \leq \frac{\lambda^2}{\lambda^2 - 2\lambda^4} \left(\|\hat{g}_\lambda\|_n^2 + 2I^2(\tilde{g}) \right) \leq 2\|\hat{g}_\lambda\|_n^2 + 4I^2(\tilde{g})$$

Case 2: $\lambda > 1/2$

By definition of $\hat{g}_\lambda$

$$\|y - \hat{g}_\lambda\|_n^2 + \lambda^2 I^2(\hat{g}_\lambda) \leq \|y\|_n^2$$

The RHS is maximized when $\lambda = 1/2$, so

$$I^2(\hat{g}_\lambda) \leq 4\|y\|_n^2$$

Hence we have an upper bound for the Sobelov norm

$$I^2(\hat{g}_\lambda) \leq 2\|\hat{g}_\lambda\|_n^2 + 4I^2(\tilde{g}) + 4\|y\|_n^2$$

Appendix

A cute lemma I found but never used: Supposing that $I^v(\hat{g}_\lambda)$ is continuous in λ, then given training data T,

$$\frac{\partial}{\partial \lambda} L_T(\hat{g}_\lambda, \lambda) \quad = \quad 2\lambda I^v(\hat{g}_\lambda)$$

Also, L_T is convex in λ.
Proof:
By definition,

$$L_T(\hat{g}_\lambda, \lambda) = \|y - \hat{g}_\lambda\|_T^2 + \lambda^2 I^v(\hat{g}_\lambda) \leq \|y - \hat{g}_{\lambda'}\|_T^2 + \lambda^2 I^v(\hat{g}_{\lambda'}) = L_T(\hat{g}_{\lambda'}, \lambda)$$

Then we can provide upper and lower bounds for $L_T(\hat{g}_{\lambda_2}, \lambda_2) - L_T(\hat{g}_{\lambda_1}, \lambda_1)$:

$$\begin{aligned}
L_T(\hat{g}_{\lambda_2}, \lambda_2) - L_T(\hat{g}_{\lambda_1}, \lambda_1) &\leq L_T(\hat{g}_{\lambda_1}, \lambda_2) - L_T(\hat{g}_{\lambda_1}, \lambda_1) \\
&= \|y - \hat{g}_{\lambda_1}\|_T^2 + \lambda_2^2 I^v(\hat{g}_{\lambda_1}) - \|y - \hat{g}_{\lambda_1}\|_T^2 - \lambda_1^2 I^v(\hat{g}_{\lambda_1}) \\
&= (\lambda_2^2 - \lambda_1^2) I^v(\hat{g}_{\lambda_1})
\end{aligned}$$

$$\begin{aligned}
L_T(\hat{g}_{\lambda_2}, \lambda_2) - L_T(\hat{g}_{\lambda_1}, \lambda_1) &\geq L_T(\hat{g}_{\lambda_2}, \lambda_2) - L_T(\hat{g}_{\lambda_2}, \lambda_1) \\
&= \|y - \hat{g}_{\lambda_2}\|_T^2 + \lambda_2^2 I^v(\hat{g}_{\lambda_2}) - \|y - \hat{g}_{\lambda_2}\|_T^2 - \lambda_1^2 I^v(\hat{g}_{\lambda_2}) \\
&= (\lambda_2^2 - \lambda_1^2) I^v(\hat{g}_{\lambda_2})
\end{aligned}$$

So suppose WLOG $\lambda_2 > \lambda_1$:

$$(\lambda_2 + \lambda_1) I^v(\hat{g}_{\lambda_2}) \leq \frac{L_T(\hat{g}_{\lambda_2}, \lambda_2) - L_T(\hat{g}_{\lambda_1}, \lambda_1)}{\lambda_2 - \lambda_1} \leq (\lambda_2 + \lambda_1) I^v(\hat{g}_{\lambda_1})$$

So as $\lambda_1 \to \lambda_2 = \lambda$, we have by the sandwich theorem,

$$\frac{\partial}{\partial \lambda} L_T(\hat{g}_\lambda, \lambda) = 2\lambda I^v(\hat{g}_\lambda)$$

Furthermore, given training data T

$$\frac{\partial}{\partial \lambda} L_T(\hat{g}_\lambda, \lambda) = \frac{\partial}{\partial \lambda} \|y - \hat{g}_\lambda\|_T^2 + 2\lambda I^v(\hat{g}_\lambda) + \lambda^2 \frac{\partial}{\partial \lambda} I^v(\hat{g}_\lambda)$$

then, combining this with the lemma, we have that

$$\frac{\partial}{\partial \lambda} \|y - \hat{g}_\lambda\|_T^2 = -\lambda^2 \frac{\partial}{\partial \lambda} I^v(\hat{g}_\lambda)$$

Finally, to see that L_T is convex in λ, note that

$$\frac{\partial^2}{\partial \lambda^2} L_T(\hat{g}_\lambda, \lambda) = 2 I^v(\hat{g}_\lambda) + 2\lambda v I^{v-1}(\hat{g}_\lambda) \frac{\partial}{\partial \lambda} I(\hat{g}_\lambda) > 0$$

since $\frac{\partial}{\partial \lambda} I(\hat{g}_\lambda) > 0$.