# Supplemental Appendix to "Compactness of Infinite Dimensional Parameter Spaces"

Joachim Freyberger* Matthew A. Masten†

December 23, 2015

**Abstract**

This supplemental appendix provides proofs for all results not already proven in the appendix of the main paper. We also provide several additional results discussed in the main paper.

## A Some useful lemmas: Proofs

*Proof of lemma 1.* Let $A \subseteq X$ be $\|\cdot\|_X$-bounded. Then it is contained in a $\|\cdot\|_X$-ball. That ball is $\|\cdot\|_Y$-relatively compact by assumption. So $A$ is a subset of a $\|\cdot\|_Y$-relatively compact set. Containment is preserved by taking closures of both sets, and hence the $\|\cdot\|_Y$-closure of $A$ is a subset of a $\|\cdot\|_Y$-compact set, and is also $\|\cdot\|_Y$-compact since it is a closed subset of a compact set. □

*Proof of lemma 2.* Let $\{a_n\}$ be a sequence in $A$. Since $A$ is $\|\cdot\|_X$-closed, any element $a$ such that $\|a_n - a\|_X \to 0$ must be in $A$. Let $a$ be such that $\|a_n - a\|_Y \to 0$. Then $\|a_n - a\|_X \to 0$ by our norm inequality. Hence $a \in A$. □

*Proof of corollary 1.* Follows by repeatedly applying lemma 2. □

*Proof of lemma 3.* This proof is given in lemma A.1 of Santos (2012) and we therefore omit it. □

*Proof of lemma 4.* Since $(X, \|\cdot\|_X)$ is embedded in $(Z, \|\cdot\|_Z)$, there exists a constant $M_1 > 0$ such that

$$\|\cdot\|_Z \leq M_1 \|\cdot\|_X.$$

Likewise, by assumption 2, there is a constant constant $M_2 > 0$ such that $\|\cdot\|_Y \leq M_2\|\cdot\|_Z$. Hence

$$\|\cdot\|_Y \leq M_1 M_2 \|\cdot\|_X.$$

Thus $(X, \|\cdot\|_X)$ is embedded in $(Y, \|\cdot\|_Y)$. Next we need to show that this embedding is compact. Let $A \subseteq X$ be $\|\cdot\|_X$-bounded. Let $\{a_n\}$ be a sequence in $A$. By assumption 1 there is a subsequence

*Department of Economics, University of Wisconsin-Madison, `jfreyberger@ssc.wisc.edu`
†Department of Economics, Duke University, `matt.masten@duke.edu`

$\{a_{n_k}\}$ that $\|\cdot\|_Z$-converges. But by assumption 2, $\|\cdot\|_Z$ is a stronger norm than $\|\cdot\|_Y$ and hence this subsequence $\|\cdot\|_Y$-converges. Thus every sequence in $A$ has a $\|\cdot\|_Y$-convergent subsequence and so $A$ is $\|\cdot\|_Y$-compact. $\square$

# B Norm inequality lemmas: Proofs

In the proof of lemma 5 and other lemmas, we use the following: The product rule tells us how to differentiate functions like $h(x)g(x)$. The generalization of this rule is called *Leibniz's formula* or the *General Leibniz rule*. For functions $u$ and $v$ that are $|\alpha|$ times continuously differentiable near $x$, it is

$$[\nabla^\alpha(uv)](x) = \sum_{\{\beta:\beta\leq\alpha\}} \begin{bmatrix}\alpha\\ \beta\end{bmatrix} \nabla^\beta u(x)\nabla^{\alpha-\beta}v(x).$$

Here $\beta \leq \alpha$ is interpreted as being component-wise: $\beta \leq \alpha$ if $\beta_j \leq \alpha_j$ for $1 \leq j \leq d_x$, where $d_x$ is the number of components in the multi-indices $\beta$ and $\alpha$, and is also equal to the dimension of the argument $x$ of the functions $u$ and $v$. Also,

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix} = \prod_{j=1}^{d_x} \binom{\alpha_j}{\beta_j}$$

where

$$\binom{\alpha_j}{\beta_j} = \frac{\alpha_j!}{\beta_j!(\alpha_j-\beta_j)!}$$

is the binomial coefficient. For a reference on this formula, see Adams and Fournier (2003), page 2.

*Proof of lemma 5.* Applying Leibniz's formula to the function $\mu(x)^{1/2}f(x)$ we have

$$\nabla^\lambda(\mu^{1/2}f) = \sum_{\{\beta:\beta\leq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} (\nabla^\beta f)(\nabla^{\lambda-\beta}\mu^{1/2}),$$

for $|\lambda| \leq m + m_0$. By the triangle inequality, this implies

$$\|\nabla^\lambda(\mu^{1/2}f)\|_{0,2,\mathbb{1}_\mathcal{C}} \leq \sum_{\{\beta:\beta\leq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} \|\nabla^{\lambda-\beta}\mu^{1/2}\nabla^\beta f\|_{0,2,\mathbb{1}_\mathcal{C}}.$$

Using the bound on the derivatives of $\mu^{1/2}$ we have

$$
\begin{aligned}
\|\nabla^{\lambda-\beta}\mu^{1/2}\nabla^{\beta}f\|_{0,2,\mathbb{1}_{\mathcal{C}}} &= \left(\int_{\mathcal{C}}[\nabla^{\lambda-\beta}\mu^{1/2}(x)\nabla^{\beta}f(x)]^2\,dx\right)^{1/2} \\
&= \left(\int_{\mathcal{C}}|\nabla^{\lambda-\beta}\mu^{1/2}(x)|^2[\nabla^{\beta}f(x)]^2\,dx\right)^{1/2} \\
&\leq \left(\int_{\mathcal{C}}|K_{\mathcal{C}}\mu^{1/2}(x)|^2[\nabla^{\beta}f(x)]^2\,dx\right)^{1/2} \\
&= K_{\mathcal{C}}^2\left(\int_{\mathcal{C}}[\nabla^{\beta}f(x)]^2\mu(x)\,dx\right)^{1/2} \\
&= K_{\mathcal{C}}^2\|\nabla^{\beta}f\|_{0,2,\mu\mathbb{1}_{\mathcal{C}}} \\
&\leq K_{\mathcal{C}}^2\|f\|_{m+m_0,2,\mu\mathbb{1}_{\mathcal{C}}},
\end{aligned}
$$

where the last line follows since $m+m_0\geq 0$. Thus, for $|\lambda|\leq m+m_0$,

$$
\|\nabla^{\lambda}(\mu^{1/2}f)\|_{0,2,\mathbb{1}_{\mathcal{C}}} \leq \left(\sum_{\{\beta:\beta\leq\lambda\}}\begin{bmatrix}\lambda\\ \beta\end{bmatrix}\right)K_{\mathcal{C}}^2\|f\|_{m+m_0,2,\mu\mathbb{1}_{\mathcal{C}}}.
$$

Next,

$$
\begin{aligned}
\|\mu^{1/2}f\|^2_{m+m_0,2,\mathbb{1}_{\mathcal{C}}} &= \sum_{0\leq|\lambda|\leq m+m_0}\|\nabla^{\lambda}(\mu^{1/2}f)\|^2_{0,2,\mathbb{1}_{\mathcal{C}}} \\
&\leq \sum_{0\leq|\lambda|\leq m+m_0}\left[\left(\sum_{\{\beta:\beta\leq\lambda\}}\begin{bmatrix}\lambda\\ \beta\end{bmatrix}\right)K_{\mathcal{C}}^2\|f\|_{m+m_0,2,\mu\mathbb{1}_{\mathcal{C}}}\right]^2 \\
&= \|f\|^2_{m+m_0,2,\mu\mathbb{1}_{\mathcal{C}}}\left[K_{\mathcal{C}}^2\sum_{0\leq|\lambda|\leq m+m_0}\left(\sum_{\{\beta:\beta\leq\lambda\}}\begin{bmatrix}\lambda\\ \beta\end{bmatrix}\right)\right]^2 \\
&\equiv \|f\|^2_{m+m_0,2,\mu\mathbb{1}_{\mathcal{C}}}M_{\mathcal{C}}^2
\end{aligned}
$$

and hence

$$
\|\mu^{1/2}f\|_{m+m_0,2,\mathbb{1}_{\mathcal{C}}} \leq M_{\mathcal{C}}\|f\|_{m+m_0,2,\mu\mathbb{1}_{\mathcal{C}}}
$$

as desired. When assumption 3 holds, the same proof above applies, but the constants now hold over all $\mathcal{D}$. □

*Proof of lemma 6.* We use induction. The inequality holds for $m = 0$ with $M_{\mathcal{C}} = 1$ since

$$\begin{aligned}\|f\|_{0,\infty,\mu^{1/2}\mathbb{1}_{\mathcal{C}}} &= \sup_{x\in\mathcal{D}} |f(x)|\mu^{1/2}(x)\mathbb{1}_{\mathcal{C}}(x) \\ &= \sup_{x\in\mathcal{D}} |\mu^{1/2}(x)f(x)|\mathbb{1}_{\mathcal{C}}(x) \\ &= \|\mu^{1/2}f\|_{0,\infty,\mathbb{1}_{\mathcal{C}}}.\end{aligned}$$

Suppose the inequality holds for $m$ and let $0 < |\lambda| \leq m+1$. By Leibniz's formula,

$$\nabla^\lambda(\mu^{1/2}f) = (\nabla^\lambda f)\mu^{1/2} + \sum_{\{\beta:\beta\leq\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} (\nabla^{\lambda-\beta}\mu^{1/2})(\nabla^\beta f),$$

which implies that

$$\begin{aligned}|(\nabla^\lambda f)\mu^{1/2}| &\leq |\nabla^\lambda(\mu^{1/2}f)| + \left|\sum_{\{\beta:\beta\leq\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} (\nabla^{\lambda-\beta}\mu^{1/2})(\nabla^\beta f)\right| \\ &\leq |\nabla^\lambda(\mu^{1/2}f)| + K_{\mathcal{C}} \sum_{\{\beta:\beta\leq\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} \mu^{1/2}|\nabla^\beta f|.\end{aligned}$$

The second line follows by assumption 4, assuming we only evaluate this inequality at $x \in \mathcal{C}$. Taking the supremum over $x$ in $\mathcal{C}$ and the maximum over $|\lambda| \leq m+1$ gives

$$\|f\|_{m+1,\infty,\mu^{1/2}\mathbb{1}_{\mathcal{C}}} \leq \|\mu^{1/2}f\|_{m+1,\infty,\mathbb{1}_{\mathcal{C}}} + K'_{\mathcal{C}}\|f\|_{m,\infty,\mu^{1/2}\mathbb{1}_{\mathcal{C}}},$$

by the definition of the norms, and since $\lambda$ isn't included in the sum we get only $m$ derivatives in this last term on the right hand side. Moreover, we picked up an extra $\leq$ since we moved the max and supremum inside the summation in the second term, and then were left with the constant

$$K'_{\mathcal{C}} \equiv K_{\mathcal{C}} \sum_{|\lambda|\leq m} \sum_{\{\beta:\beta\leq\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} < \infty.$$

By the induction hypothesis there is an $M'_{\mathcal{C}} < \infty$ such that

$$\|f\|_{m,\infty,\mu^{1/2}\mathbb{1}_{\mathcal{C}}} \leq M'_{\mathcal{C}}\|\mu^{1/2}f\|_{m,\infty,\mathbb{1}_{\mathcal{C}}}.$$

Moreover,

$$\|\mu^{1/2}f\|_{m,\infty,\mathbb{1}_{\mathcal{C}}} \leq \|\mu^{1/2}f\|_{m+1,\infty,\mathbb{1}_{\mathcal{C}}}.$$

Thus

$$\|f\|_{m,\infty,\mu^{1/2}\mathbb{1}_{\mathcal{C}}} \leq M'_{\mathcal{C}}\|\mu^{1/2}f\|_{m+1,\infty,\mathbb{1}_{\mathcal{C}}}.$$

Plugging this into our expression from earlier yields

$$\begin{aligned}\|f\|_{m+1,\infty,\mu^{1/2}\mathbb{1}_{\mathcal{C}}} &\leq \|\mu^{1/2}f\|_{m+1,\infty,\mathbb{1}_{\mathcal{C}}} + K'_{\mathcal{C}}\|f\|_{m,\infty,\mu^{1/2}\mathbb{1}_{\mathcal{C}}} \\ &\leq \|\mu^{1/2}f\|_{m+1,\infty,\mathbb{1}_{\mathcal{C}}} + K'_{\mathcal{C}}M'_{\mathcal{C}}\|\mu^{1/2}f\|_{m+1,\infty,\mathbb{1}_{\mathcal{C}}} \\ &= (1+K'_{\mathcal{C}}M'_{\mathcal{C}})\|\mu^{1/2}f\|_{m+1,\infty,\mathbb{1}_{\mathcal{C}}} \\ &\equiv M_{\mathcal{C}}\|\mu^{1/2}f\|_{m+1,\infty,\mathbb{1}_{\mathcal{C}}}.\end{aligned}$$

When assumption 3 holds, the same proof above applies, but the constants now hold over all $\mathcal{D}$. □

*Proof of lemma 7.* We will modify the proof of lemma 6 as appropriate. As there, we use proof by induction. For the base case, set $m = 0$. Then

$$\begin{aligned}\|f\|_{0,2,\mu\mathbb{1}_{\mathcal{C}}} &= \left(\int_{\mathcal{C}} [f(x)]^2\mu(x)\ dx\right)^{1/2} \\ &= \left(\int_{\mathcal{C}} [\mu^{1/2}(x)f(x)]^2\ dx\right)^{1/2} \\ &= \|\mu^{1/2}f\|_{0,2,\mathbb{1}_{\mathcal{C}}}.\end{aligned}$$

Thus the result holds for $m = 0$. Now suppose it holds for $m$. Let $|\lambda|$ be such that $0 < |\lambda| \leq m+1$. Then, as in the proof of lemma 6, we have

$$\nabla^\lambda(\mu^{1/2}f) = (\nabla^\lambda f)\mu^{1/2} + \sum_{\{\beta:\beta\leq\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda \\ \beta\end{bmatrix} (\nabla^{\lambda-\beta}\mu^{1/2})(\nabla^\beta f)$$

by Leibniz's formula. As in that proof, applying our bound on the derivative of the weight function, we get

$$|\nabla^\lambda f|\mu^{1/2} \leq |\nabla^\lambda(\mu^{1/2}f)| + K_{\mathcal{C}} \sum_{\{\beta:\beta\leq\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda \\ \beta\end{bmatrix} |\nabla^\beta f|\mu^{1/2}.$$

Now we square both sides and integrate over $\mathcal{C}$ to obtain

$$\begin{aligned}
\int_{\mathcal{C}} |\nabla^\lambda f(x)|^2 \mu(x)\, dx &\le \int_{\mathcal{C}} |[\nabla^\lambda(\mu^{1/2} f)](x)|^2\, dx \\
&\quad + \int_{\mathcal{C}} K_{\mathcal{C}}^2 \sum_{\{\tilde{\beta}:\tilde{\beta}\le\lambda,\tilde{\beta}\neq\lambda\}} \sum_{\{\beta:\beta\le\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \tilde{\beta}\end{bmatrix} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} |\nabla^{\tilde{\beta}} f(x)| \cdot |\nabla^\beta f(x)| \mu(x)\, dx \\
&\quad + \int_{\mathcal{C}} 2|[\nabla^\lambda(\mu^{1/2} f)](x)| K_{\mathcal{C}} \sum_{\{\beta:\beta\le\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} |\nabla^\beta f(x)| \mu^{1/2}(x)\, dx \\
&= \int_{\mathcal{C}} |[\nabla^\lambda(\mu^{1/2} f)](x)|^2\, dx \\
&\quad + K_{\mathcal{C}}^2 \sum_{\{\tilde{\beta}:\tilde{\beta}\le\lambda,\tilde{\beta}\neq\lambda\}} \sum_{\{\beta:\beta\le\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \tilde{\beta}\end{bmatrix} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} \int_{\mathcal{C}} |\nabla^{\tilde{\beta}} f(x)| \cdot |\nabla^\beta f(x)| \mu(x)\, dx \\
&\quad + 2K_{\mathcal{C}} \sum_{\{\beta:\beta\le\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} \int_{\mathcal{C}} |[\nabla^\lambda(\mu^{1/2} f)](x)| \cdot |\nabla^\beta f(x)| \mu^{1/2}(x)\, dx \\
&\equiv (1) + (2) + (3).
\end{aligned}$$

In the third term, we can apply Leibniz's formula again,

$$|\nabla^\beta f| \mu^{1/2} \le |\nabla^\beta(\mu^{1/2} f)| + K_{\mathcal{C}} \sum_{\{\eta:\eta\le\beta,\eta\neq\beta\}} \begin{bmatrix}\beta\\ \eta\end{bmatrix} |\nabla^\eta f| \mu^{1/2}$$

to get

$$\begin{aligned}
(3) &\equiv 2K_{\mathcal{C}} \sum_{\{\beta:\beta\le\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} \int_{\mathcal{C}} |[\nabla^\lambda(\mu^{1/2} f)](x)| \cdot |\nabla^\beta f(x)| \mu^{1/2}(x)\, dx \\
&\le 2K_{\mathcal{C}} \sum_{\{\beta:\beta\le\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda\\ \beta\end{bmatrix} \Bigg( \int_{\mathcal{C}} |[\nabla^\lambda(\mu^{1/2} f)](x)| \cdot |[\nabla^\beta(\mu^{1/2} f)](x)|\, dx \\
&\qquad + K_{\mathcal{C}} \sum_{\{\eta:\eta\le\beta,\eta\neq\beta\}} \begin{bmatrix}\beta\\ \eta\end{bmatrix} \int_{\mathcal{C}} |[\nabla^\lambda(\mu^{1/2} f)](x)| \cdot |\nabla^\eta f(x)| \mu^{1/2}(x)\, dx \Bigg).
\end{aligned}$$

We can apply Leibniz's formula again to eliminate the $|\nabla^\eta f(x)|\mu^{1/2}(x)$ term. Continuing in this manner, we get a sum solely of integrals of the form

$$\int_{\mathcal{C}} |[\nabla^\lambda(\mu^{1/2} f)](x)| \cdot |[\nabla^\beta(\mu^{1/2} f)](x)|\, dx.$$

Now replace one of the two absolute value terms in the integrand with whichever one is largest.

Suppose its the $\lambda$ piece. This yields

$$\int_{\mathcal{C}} |[\nabla^{\lambda}(\mu^{1/2}f)](x)| \cdot |[\nabla^{\beta}(\mu^{1/2}f)](x)| \; dx \leq \int_{\mathcal{C}} |[\nabla^{\lambda}(\mu^{1/2}f)](x)|^2 \; dx.$$

Thus the third piece is now a sum of terms like this one, where the multi-index in the differential operator can go as high as $|\lambda|$. Summing (3) over $|\lambda|$ with $0 \leq |\lambda| \leq m+1$ we obtain a sum of many unweighted integrals over $\mathcal{C}$ with integrands of the form $|[\nabla^{\lambda}(\mu^{1/2}f)](x)|^2$. Now all we have to do is group all these integrals such that our entire expression (3) is a multiple of

$$\sum_{0 \leq |\lambda| \leq m+1} \int_{\mathcal{C}} |[\nabla^{\lambda}(\mu^{1/2}f)](x)|^2 \; dx = \|\mu^{1/2}f\|^2_{m+1,2,\mathbb{1}_{\mathcal{C}}}.$$

If there are any 'missing' integrals, we can just add on the missing ones (which will give us another inequality, but that's ok since we only need an upper bound). Thus we see that, after summing over $0 \leq |\lambda| \leq m+1$, the term (3) is bounded above by

$$C_{3,\mathcal{C}} \|\mu^{1/2}f\|^2_{m+1,2,\mathbb{1}_{\mathcal{C}}}$$

for some constant $C_{3,\mathcal{C}} > 0$.

Consider now the second piece. It is a sum of integrals of the form

$$\int_{\mathcal{C}} |\nabla^{\tilde{\beta}} f(x)| \cdot |\nabla^{\beta} f(x)| \mu(x) \; dx.$$

Basically the same argument from third piece applies. We can replace one of the absolute values here with whichever is the largest, thus obtaining an integral of the form

$$\int_{\mathcal{C}} |\nabla^{\beta} f(x)|^2 \mu(x) \; dx.$$

Now summing these terms over $0 \leq |\lambda| \leq m+1$ we see that after grouping all the integrals and adding any missing terms, the entire expression (2) is a multiple of

$$\sum_{0 \leq |\lambda| \leq m} \int_{\mathcal{C}} |\nabla^{\lambda} f(x)|^2 \; dx = \|f\|^2_{m,2,\mathbb{1}_{\mathcal{C}}}.$$

It is important here that the sum only goes up to $m$, not $m+1$. This is because, in the term (2), the $\beta$ and $\tilde{\beta}$ pieces are always strictly smaller than $\lambda$, and $\lambda$ itself can only go up to $m+1$. Hence $\beta$ and $\tilde{\beta}$ can only go up to $m$. Thus we see that the term (2) is bounded above by

$$C_{2,\mathcal{C}} \|f\|^2_{m,2,\mathbb{1}_{\mathcal{C}}}$$

for some constant $C_{2,\mathcal{C}} > 0$. Finally, consider the term (1). This term is easy because when we sum over $0 \leq |\lambda| \leq m+1$ this term exactly equals

$$\|\mu^{1/2} f\|^2_{m+1,2,\mathbb{1}_\mathcal{C}}$$

without having to add any extra terms or mess with the integrands. Combining all these results, we see (by also summing over the left hand side of our original inequality) that

$$\|f\|^2_{m+1,2,\mu\mathbb{1}_\mathcal{C}} \leq (1 + C_{3,\mathcal{C}})\|\mu^{1/2} f\|^2_{m+1,2,\mathbb{1}_\mathcal{C}} + C_{2,\mathcal{C}}\|f\|^2_{m,2,\mathbb{1}_\mathcal{C}}.$$

Now apply the induction hypothesis to the last term to get

$$\begin{aligned}\|f\|^2_{m+1,2,\mu\mathbb{1}_\mathcal{C}} &\leq (1 + C_{3,\mathcal{C}})\|\mu^{1/2} f\|^2_{m+1,2,\mathbb{1}_\mathcal{C}} + C_{2,\mathcal{C}}\|\mu^{1/2} f\|^2_{m+1,2,\mathbb{1}_\mathcal{C}} \\ &= (1 + C_{3,\mathcal{C}} + C_{2,\mathcal{C}})\|\mu^{1/2} f\|^2_{m+1,2,\mathbb{1}_\mathcal{C}}.\end{aligned}$$

Finally, take the square root of both sides to get

$$\|f\|_{m+1,2,\mu\mathbb{1}_\mathcal{C}} \leq (1 + C_{3,\mathcal{C}} + C_{2,\mathcal{C}})^{1/2}\|\mu^{1/2} f\|_{m+1,2,\mathbb{1}_\mathcal{C}}$$

as desired. When assumption 3 holds, the same proof above applies, but the constants now hold over all $\mathcal{D}$. □

*Proof of lemma 8.* As in the proof of lemma 6, we have

$$\nabla^\lambda(\mu^{1/2} f) = (\nabla^\lambda f)\mu^{1/2} + \sum_{\{\beta:\beta\leq\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda \\ \beta\end{bmatrix} (\nabla^{\lambda-\beta}\mu^{1/2})(\nabla^\beta f).$$

Hence

$$|\nabla^\lambda(\mu^{1/2} f)| \leq |(\nabla^\lambda f)\mu^{1/2}| + \sum_{\{\beta:\beta\leq\lambda,\beta\neq\lambda\}} \begin{bmatrix}\lambda \\ \beta\end{bmatrix} |(\nabla^\beta f)\mu^{1/2}|.$$

Take the sup over $x$ and the max over $|\lambda| \leq m+1$ to get

$$\|\mu^{1/2} f\|_{m+1,\infty} \leq \|f\|_{m+1,\infty,\mu^{1/2}} + K'\|f\|_{m,\infty,\mu^{1/2}}.$$

Since $\|f\|_{m,\infty,\mu^{1/2}} \leq \|f\|_{m+1,\infty,\mu^{1/2}}$ we get

$$\|\mu^{1/2} f\|_{m+1,\infty} \leq (1 + K')\|f\|_{m+1,\infty,\mu^{1/2}}.$$

The result follows by evaluating this inequality with the weight $\mu^2$. □

## C Proofs of the compact embedding theorems 5 and 7

*Proof of theorem 5 (Compact embedding for unbounded domains with product weighting).* For cases 1–3, we apply lemma S1 below, which allows us to convert our previous compact embedding and closedness results for equal weighting to results for product weighting. For case 4, we do not have such a prior result because it's not clear how to define equal weighted Hölder norms, as discussed in the main paper. Hence for this case we instead modify the proof of the previous compact embedding and closedness results.

**Cases 1–3**: Theorem 3 (case 1: part 1 with the $s$ weight equal to the constant 1 and the $c$ weight equal to $\tilde{\mu}^2$) (case 2: part 2 with the $s$ weight equal to 1 and the $c$ weight equal to $\tilde{\mu}$) (case 3: part 3, with weights chosen as in case 2) implies that (cases 1 and 2: $\mathscr{W}_{m+m_0,2,\mathbb{1}}$) (case 3: $\mathscr{C}_{m+m_0,\infty,\mathbb{1}}$) is compactly embedded in (cases 1 and 3: $\mathscr{C}_{m,\infty,\tilde{\mu}}$) (case 2: $\mathscr{W}_{m,2,\tilde{\mu}}$). Note that both the constant weight function, $\tilde{\mu}$, and $\tilde{\mu}^2$ satisfy the local integrability assumptions $6'$ and $6''$ as well as assumption 3.

By proposition 6, (cases 1 and 3: $\|\cdot\|_{m,\infty,\tilde{\mu}}$) (case 2: $\|\cdot\|_{m,2,\tilde{\mu}}$) and (cases 1 and 3: $\|\cdot\|_{m,\infty,\tilde{\mu},\text{ALT}}$) (case 2: $\|\cdot\|_{m,2,\tilde{\mu},\text{ALT}}$) are equivalent norms. Therefore (cases 1 and 2: $\mathscr{W}_{m+m_0,2,\mathbb{1}} = \mathscr{W}_{m+m_0,2,\mathbb{1},\text{ALT}}$) (case 3: $\mathscr{C}_{m+m_0,\infty,\mathbb{1},\text{ALT}}$) is compactly embedded in (cases 1 and 3: $\mathscr{C}_{m,\infty,\tilde{\mu},\text{ALT}}$) (case 2: $\mathscr{W}_{m,2,\tilde{\mu},\text{ALT}}$). Lemma S1 part 1 now implies that (cases 1 and 2: $\mathscr{W}_{m+m_0,2,\mu_s,\text{ALT}}$) (case 3: $\mathscr{C}_{m+m_0,\infty,\mu_s,\text{ALT}}$) is compactly embedded in (cases 1 and 3: $\mathscr{C}_{m,\infty,\mu_c,\text{ALT}}$) (case 2: $\mathscr{W}_{m,2,\mu_c,\text{ALT}}$).

**Case 4**: The proof is similar to the proof of theorem 3. Since we have already given a detailed proof of that theorem, here we only comment on the nontrivial modifications to that proof. The numbers here refer to the steps in that proof.

1. $\Theta = \{f \in \mathscr{C}_{m+m_0,\infty,\mu_s,\nu} : \|\mu_s f\|_{m+m_0,\infty,\mathbb{1},\nu} \leq B\}$.

2. Completeness of the function spaces under product weighting follows by completeness of the unweighted spaces.

4. This step is not necessary since, by definition of the product weighted norms, $f_n \in \Theta$ for all $n$ implies

   $\{\mu_s f_n\}$ is $\|\cdot\|_{m+m_0,\infty,\mathbb{1},\nu}$-bounded. In particular, this implies it is $\|\cdot\|_{m+m_0,\infty,\mathbb{1}_{\Omega_J},\nu}$-bounded for each $J$, where here

$$\|g\|_{m+m_0,\infty,\mathbb{1}_{\Omega_J},\nu} = \|g\|_{m+m_0,\infty,\mathbb{1}_{\Omega_J}} + \max_{|\lambda|=m+m_0} \sup_{x,y\in\Omega_J, x\neq y} \frac{|\nabla^\lambda f(x) - \nabla^\lambda f(y)|}{\|x-y\|_e^\nu}.$$

   Generally, in this proof indicators in the weight function placeholder denote the set over which integration or suprema are taken.

5. Apply theorem 1 part 5. Since $\{\mu_s f_n\}$ is $\|\cdot\|_{m+m_0,\infty,\mathbb{1}_{\Omega_J},\nu}$-bounded, it is $\|\cdot\|_{m,\infty,\mathbb{1}_{\Omega_J}}$-relatively compact.

9. By identical calculations as before, we have

$$\|f_j - f_k\|_{m,\infty,\mu_c,\text{ALT}} \leq \|f_j - f_k\|_{m,\infty,\mu_c \mathbb{1}_{\Omega_J},\text{ALT}} + \|f_j - f_k\|_{m,\infty,\mu_c \mathbb{1}_{\Omega_J^c},\text{ALT}}.$$

10. For $f_j \in \Theta$ we have

$$\begin{aligned}
\|f_j\|_{m,\infty,\mu_c \mathbb{1}_{\Omega_J^c},\text{ALT}} &= \|\mu_c f_j\|_{m,\infty,\mathbb{1}_{\Omega_J^c}} \\
&= \|\mu_s \tilde{\mu} f_j\|_{m,\infty,\mathbb{1}_{\Omega_J^c}} \\
&\leq M \|\mu_s f_j\|_{m,\infty,\tilde{\mu}\mathbb{1}_{\Omega_J^c}} \\
&= M \max_{0 \leq |\lambda| \leq m} \sup_{x \in \Omega_J^c} |\nabla^\lambda (\mu_s(x) f_j(x))| \tilde{\mu}(x) \\
&\leq M \max_{0 \leq |\lambda| \leq m} \sup_{x \in \mathbb{R}^{d_x}} |\nabla^\lambda (\mu_s(x) f_j(x))| \sup_{x \in \Omega_J^c} \tilde{\mu}(x) \\
&\leq M \|\mu_s f_j\|_{m+m_0,\infty,\mathbb{1},\nu} \sup_{x \in \Omega_J^c} \tilde{\mu}(x) \\
&\leq MB \sup_{x \in \Omega_J^c} \tilde{\mu}(x).
\end{aligned}$$

The third line follows by lemma 8. The last line follows since $f_j \in \Theta$. Now since $\tilde{\mu}(x) = (1 + x'x)^{-\delta}$, $\delta > 0$, converges to zero in the tails, we can choose $J$ large enough such that

$$\sup_{x \in \Omega_J^c} \tilde{\mu}(x) < \frac{\varepsilon}{4MB}.$$

Hence, by the triangle inequality,

$$\|f_j - f_k\|_{m,\infty,\mu_c \mathbb{1}_{\Omega_J^c}} < \frac{\varepsilon}{2}.$$

11. Since $\{\mu_s f_j^{(J)}\}$ converges in the norm $\|\cdot\|_{m,\infty,\mathbb{1}_{\Omega_J}}$ it is also Cauchy in that norm. Thus there is some $K$ large enough (take $K > J$) such that

$$\|\mu_s (f_j - f_k)\|_{m,\infty,\mathbb{1}_{\Omega_J}} < \frac{\varepsilon}{2M}$$

for all $k, j > K$, where $M$ is a constant given below. Hence

$$\begin{aligned}
\|f_j - f_k\|_{m,\infty,\mu_c \mathbb{1}_{\Omega_J},\text{ALT}} &= \|\mu_c(f_j - f_k)\|_{m,\infty,\mathbb{1}_{\Omega_J}} \\
&= \|\mu_s \tilde{\mu}(f_j - f_k)\|_{m,\infty,\mathbb{1}_{\Omega_J}} \\
&\leq M \|\mu_s(f_j - f_k)\|_{m,\infty,\tilde{\mu}\mathbb{1}_{\Omega_J}} \\
&\leq M \|\mu_s(f_j - f_k)\|_{m,\infty,\mathbb{1}_{\Omega_J}} \\
&< M \frac{\varepsilon}{2M} \\
&= \frac{\varepsilon}{2}.
\end{aligned}$$

The third line follows by lemma 8. The fourth line follows since $\tilde{\mu}(x) = (1 + x'x)^{-\delta} \leq 1$ for all $x$.

□

**Lemma S1.** Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces where $\|f\|_X < \infty$ for all $f \in X$ and $\|f\|_Y < \infty$ for all $f \in Y$. Moreover, suppose that for all $f \in X$

$$\|f\|_X = \|f\|_s$$

and for all $f \in Y$

$$\|f\|_Y = \|f\tilde{\mu}\|_c$$

where $\|\cdot\|_s$ and $\|\cdot\|_c$ are norms and $\tilde{\mu}$ is a weight function. Let $(\tilde{X}, \|\cdot\|_{\tilde{X}})$ and $(\tilde{Y}, \|\cdot\|_{\tilde{Y}})$ be Banach spaces where $\|f\|_{\tilde{X}} < \infty$ for all $f \in \tilde{X}$ and $\|f\|_{\tilde{Y}} < \infty$ for all $f \in \tilde{Y}$. Moreover, suppose that for all $f \in \tilde{X}$

$$\|f\|_{\tilde{X}} = \|f\mu_s\|_s$$

and for all $f \in \tilde{Y}$

$$\|f\|_{\tilde{Y}} = \|f\mu_s\tilde{\mu}\|_c$$

for some weight function $\mu_s$.

1. (Compact embedding) Suppose $(X, \|\cdot\|_X)$ is compactly embedded in $(Y, \|\cdot\|_Y)$. Then $(\tilde{X}, \|\cdot\|_{\tilde{X}})$ is compactly embedded in $(\tilde{Y}, \|\cdot\|_{\tilde{Y}})$.

2. (Closedness) Suppose

   $$\Omega = \{f \in X : \|f\|_X \leq B\}$$

   is $\|\cdot\|_Y$-closed. Then

   $$\tilde{\Omega} = \{f \in \tilde{X} : \|f\|_{\tilde{X}} \leq B\}$$

   is $\|\cdot\|_{\tilde{Y}}$-closed.

*Proof of lemma S1.*

1. Let $f \in \tilde{X}$. By definition, $\|f\|_{\tilde{X}} = \|f\mu_s\|_s < \infty$. Define $h = f\mu_s$ and notice that $h \in X$. Since $(X, \|\cdot\|_X)$ is compactly embedded in $(Y, \|\cdot\|_Y)$, $X \subseteq Y$ and there exists a constant $C$ such that $\|h\|_Y \leq C\|h\|_X$. First note that $h \in X$ implies $\|h\|_Y < \infty$ and hence $\|h\tilde{\mu}\|_c = \|f\mu_s\tilde{\mu}\|_c < \infty$. So $f \in \tilde{Y}$ and thus $\tilde{X} \subseteq \tilde{Y}$. Next, note that

$$\begin{aligned}\|h\|_Y \leq C\|h\|_X &\Leftrightarrow \|h\tilde{\mu}\|_c \leq C\|h\|_s \\ &\Leftrightarrow \|f\mu_s\tilde{\mu}\|_c \leq C\|f\mu_s\|_s \\ &\Leftrightarrow \|f\|_{\tilde{Y}} \leq C\|f\|_{\tilde{X}}.\end{aligned}$$

Next let $\{f_n\}$ be a sequence in the $\|\cdot\|_{\tilde{Y}}$-closure of

$$\tilde{\Omega} = \{f \in \tilde{X} : \|f\|_{\tilde{X}} \leq B\} = \{f \in \tilde{X} : \|f\mu_s\|_s \leq B\}.$$

Let $h_n = f_n\mu_s$. Then by definition of the norms, $h_n$ is a sequence in the $\|\cdot\|_Y$-closure of

$$\Omega = \{h \in X : \|h\|_X \leq B\}.$$

Since $(X, \|\cdot\|_X)$ is compactly embedded in $(Y, \|\cdot\|_Y)$, there exists a subsequence $h_{n_j} = f_{n_j}\mu_s$, which is $\|\cdot\|_Y$-Cauchy. That is, for any $\varepsilon > 0$, there exists an $N$ such that $\|h_{n_j} - h_{n_k}\|_Y \leq \varepsilon$ for all $j, k > N$. But

$$\|h_{n_j} - h_{n_k}\|_Y = \|(h_{n_j} - h_{n_k})\tilde{\mu}\|_c = \|(f_{n_j} - f_{n_k})\mu_s\tilde{\mu}\|_c = \|f_{n_j} - f_{n_k}\|_{\tilde{Y}}.$$

Therefore, $f_{n_j}$ is a subsequence of $f_n$ which is $\|\cdot\|_{\tilde{Y}}$-Cauchy. Since $(\tilde{Y}, \|\cdot\|_{\tilde{Y}})$ is Banach, $f_j$ converges to a point in $\tilde{Y}$. Hence $(\tilde{X}, \|\cdot\|_{\tilde{X}})$ is compactly embedded in $(\tilde{Y}, \|\cdot\|_{\tilde{X}})$.

2. Let $f_n$ be a sequence in $\tilde{\Omega}$ such that for some $f \in \tilde{X}$, $\|f_n - f\|_{\tilde{Y}} \to 0$ as $n \to \infty$. Since $f_n \in \tilde{\Omega}$ we have $\|f_n\mu_s\|_s = \|f\|_{\tilde{X}} \leq B$. Let $h_n = f_n\mu_s$ and $h = f\mu_s$. Since

$$\|h_n\|_X = \|h_n\|_s = \|f_n\mu_s\|_s = \|f\|_{\tilde{X}} \leq B$$

we have $h_n \in \Omega$. Moreover,

$$\|h_n - h\|_Y = \|(h_n - h)\tilde{\mu}\|_c = \|f_n - f\|_{\tilde{Y}} \to 0.$$

Since $\Omega = \{f \in X : \|f\|_X \leq B\}$ is $\|\cdot\|_Y$-closed, $h \in \Omega$. That is, $f\mu_s \in \Omega$, which implies that

$$\|f\|_{\tilde{X}} = \|f\mu_s\|_X \leq B.$$

Hence $f \in \tilde{\Omega}$. So $\tilde{\Omega}$ is $\|\cdot\|_{\tilde{Y}}$-closed.

□

*Proof of theorem 7 (Compact embedding for weighted norms on bounded domains).* The proof is similar to the proof of theorem 3. Since we have already given a detailed proof of that theorem, here we only comment on the nontrivial modifications to that proof. The numbers here refer to the steps in that proof.

2. For case 1, $\Omega_1 = \cdots = \Omega_k = \mathcal{D}$ and $\Omega^{(0)} = \mathcal{D}$ when applying Rodríguez, Álvarez, Romera, and Pestana (2004).

3. We use the following more general domain truncation: Let $\{\Omega_J\}$ be a sequence of open subsets of $\mathcal{D}$ such that

   (a) $\Omega_J \subseteq \Omega_{J+1}$ for any $J$,

   (b) $\bigcup_{J=1}^{\infty} \Omega_J = \mathcal{D}$, and

   (c) The closure of $\Omega_J$ does not contain the boundary of the closure of $\mathcal{D}$ for any $J$. That is, $\text{Boundary}(\overline{\mathcal{D}}) \cap \overline{\Omega}_J = \emptyset$ for all $J$.

   Roughly speaking, the sets $\Omega_J$ are converging to $\mathcal{D}$ from the inside. They do this in such a way that for any $J$, the boundary points of $\overline{\mathcal{D}}$ are well separated from $\Omega_J$.

The rest of the steps go through with very minor modifications. □

# D Proofs of closedness theorems

*Proof of theorem 2 (Closedness for bounded domains).* For this proof we let $d_x = 1$ to simplify the notation. All arguments generalize to $d_x > 1$.

1. We want to show that the $\|\cdot\|_s = \|\cdot\|_{m+m_0,2}$-ball $\Theta$ is $\|\cdot\|_c = \|\cdot\|_{m,\infty}$-closed. $(\mathscr{W}_{m+m_0,2}, \|\cdot\|_{m+m_0,2})$ is compactly embedded in $(\mathscr{W}_{m,2}, \|\cdot\|_{m,2})$ by part 2 of theorem 1, which applies since we assumed $\mathcal{D}$ satisfies the cone condition and $m_0 > d_x/2$. Lemma A.1 in Santos (2012) (reproduced in the main paper's appendix on page 38 for convenience) then implies that that the $\|\cdot\|_{m+m_0,2}$-ball $\Theta$ is $\|\cdot\|_{m,2}$-closed, because the Sobolev $L_2$ spaces are separable Hilbert spaces (theorem 3.6 of Adams and Fournier 2003). Finally, since $\|\cdot\|_{m,2} \leq \|\cdot\|_{m,\infty}$ corollary 1 implies that $\Theta$ is $\|\cdot\|_{m,\infty}$-closed.

2. We want to show that the $\|\cdot\|_s = \|\cdot\|_{m+m_0,2}$-ball $\Theta$ is $\|\cdot\|_c = \|\cdot\|_{m,2}$-closed. We already showed this in the proof of part 1.

3. We want to show that the $\|\cdot\|_s = \|\cdot\|_{m+m_0,\infty}$-ball $\Theta$ is not $\|\cdot\|_c = \|\cdot\|_{m,\infty}$-closed. Consider the case $m = 0$ and $m_0 = 1$, so that $\Theta$ is the set of continuously differentiable functions whose levels and first derivatives are uniformly bounded by $B$. We will show that this set is not closed in the ordinary sup-norm $\|\cdot\|_{0,\infty}$.

   Suppose $\mathcal{D} = (-1, 1)$. Define
   $$g_k(x) = \sqrt{x^2 + 1/k}.$$

for integers $k \geq 1$. These are smooth approximations to the absolute value function: For each $x \in \mathcal{D}$, $g_k(x) \to \sqrt{x^2} = |x|$ as $k \to \infty$. $g_k$ is continuous and differentiable, with first derivative

$$\begin{aligned} g_k'(x) &= \frac{1}{2}(x^2 + 1/k)^{-1/2} \cdot 2x \\ &= \frac{x}{\sqrt{x^2 + 1/k}}. \end{aligned}$$

So

$$|g_k'(x)| \leq \frac{|x|}{\sqrt{x^2 + 1/k}} \leq \frac{|x|}{\sqrt{x^2}} = 1$$

for all $k$. Also,

$$|g_k(x)| = \sqrt{x^2 + 1/k} \leq \sqrt{1 + 1/k} \leq \sqrt{1+1} = \sqrt{2}$$

for all $k$. Hence $g_k \in \Theta = \{f \in \mathscr{C}_1(\mathcal{D}) : \|f\|_{1,\infty} \leq B\}$ for each $k$, where $B = 1 + \sqrt{2}$. But, letting $f(x) = |x|$,

$$\|g_k - f\|_{0,\infty} = \sup_{x \in \mathcal{D}} |g_k(x) - f(x)| \to 0$$

as $k \to \infty$. Since $f$ is not differentiable at 0, $f \notin \Theta$. This implies that $\Theta$ is not closed under $\|\cdot\|_{0,\infty}$.

4. We want to show that the $\|\cdot\|_s = \|\cdot\|_{m+m_0,\infty}$-ball $\Theta$ is not $\|\cdot\|_c = \|\cdot\|_{m,2}$-closed. The same counterexample from part 4 applies here as well. Letting $m = 0$ and $m_0 = 1$, we will show that the $\|\cdot\|_{1,\infty}$-ball $\Theta$ is not closed in the ordinary $L_2$ norm $\|\cdot\|_{0,2}$. From part 4, we constructed a sequence $g_k$ in $\Theta$ such that

$$\|g_k - f\|_{0,\infty} \to 0$$

as $k \to \infty$, for $f \notin \Theta$. Convergence in $\|\cdot\|_{0,\infty}$ implies convergence in $\|\cdot\|_{0,2}$ and hence

$$\|g_k - f\|_{0,2} \to 0$$

as $k \to \infty$. Therefore $\Theta$ is not closed under $\|f\|_{0,2}$.

5. We want to show that $\|\cdot\|_{m+m_0,\infty,\mathbb{1},\nu}$-balls are $\|\cdot\|_{m,\infty}$-closed, where $m_0 \geq 0$. Since $\|\cdot\|_{0,\infty} \leq \|\cdot\|_{m,\infty}$, corollary 1 shows that it is sufficient to prove the result for $m = 0$. That is, it is sufficient to prove that the $\|\cdot\|_{m_0,\infty,\mathbb{1},\nu}$-ball

$$\Theta_{m_0} \equiv \{f \in \mathscr{C}_{m_0,\infty,\mathbb{1},\nu} : \|f\|_{m_0,\infty,\mathbb{1},\nu} \leq B\}$$

is $\|\cdot\|_{0,\infty}$-closed, for all $m_0 \geq 0$. We proceed by induction on $m_0$.

**Step 1 (Base Case):** Let $m_0 = 0$. We want to show that $\Theta_0$ is $\|\cdot\|_{0,\infty}$-closed, so we will show that its complement $\Theta_0^c = \mathscr{C}_{0,\infty} \setminus \Theta_0$ is $\|\cdot\|_{0,\infty}$-open. That is, for any $f \in \Theta_0^c$ there exists an $\varepsilon > 0$ such that

$$\{g \in \mathscr{C}_{0,\infty} : \|f - g\|_{0,\infty} \leq \varepsilon\} \subseteq \Theta_0^c.$$

So take an arbitrary $f \in \Theta_0^c$. Since $f$ is outside the Hölder ball $\Theta_0$, its Hölder norm is larger than $B$,

$$\sup_{x \in \mathcal{D}} |f(x)| + \sup_{x_1, x_2 \in \mathcal{D}, x_1 \neq x_2} \frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|^\nu} > B.$$

Hence there exist points $\bar{x}, \bar{x}_1, \bar{x}_2$ in the Euclidean closure of $\mathcal{D}$ with $\bar{x}_1 \neq \bar{x}_2$ such that

$$|f(\bar{x})| + \frac{|f(\bar{x}_1) - f(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} > B.$$

Define

$$\delta = |f(\bar{x})| + \frac{|f(\bar{x}_1) - f(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} - B > 0.$$

Our goal is find a $\|\cdot\|_{0,\infty}$-ball around $f$ with some positive radius $\varepsilon$ such that all functions $g$ in that ball are also not in the Hölder ball $\Theta_0$. So we need these functions $g$ to have a large Hölder norm (larger than $B$). Let's examine that. For all $g \in \mathscr{C}_{0,\infty}$,

$$\begin{aligned}
\|g\|_{0,\infty,\mathbb{1},\nu} &= \sup_{x \in \mathcal{D}} |g(x)| + \sup_{x_1, x_2 \in \mathcal{D}, x_1 \neq x_2} \frac{|g(x_1) - g(x_2)|}{|x_1 - x_2|^\nu} \\
&\geq |g(\bar{x})| + \frac{|g(\bar{x}_1) - g(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} \\
&\geq |f(\bar{x})| - |f(\bar{x}) - g(\bar{x})| + \frac{|g(\bar{x}_1) - g(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} \\
&= |f(\bar{x})| - |f(\bar{x}) - g(\bar{x})| \\
&\qquad + \frac{|f(\bar{x}_1) - f(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} - \frac{|f(\bar{x}_1) - f(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} + \frac{|g(\bar{x}_1) - g(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} \\
&\geq |f(\bar{x})| - |f(\bar{x}) - g(\bar{x})| \\
&\qquad + \frac{|f(\bar{x}_1) - f(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} - \frac{|\,(f(\bar{x}_1) - g(\bar{x}_1)) - (f(\bar{x}_2) - g(\bar{x}_2))\,|}{|\bar{x}_1 - \bar{x}_2|^\nu} \\
&= B + \delta - \left( |f(\bar{x}) - g(\bar{x})| + \frac{|\,(f(\bar{x}_1) - g(\bar{x}_1)) - (f(\bar{x}_2) - g(\bar{x}_2))\,|}{|\bar{x}_1 - \bar{x}_2|^\nu} \right).
\end{aligned}$$

The third and fifth lines follow by the reverse triangle inequality. The last line follows by the definition of $\delta$. If we can make this last piece in parentheses small enough, we'll be done. For any $\varepsilon > 0$,

$$g \in \{g \in \mathscr{C}_{0,\infty} : \|f - g\|_{0,\infty} \leq \varepsilon\}$$

implies

$$|f(\bar{x}) - g(\bar{x})| + \frac{|(f(\bar{x}_1) - g(\bar{x}_1)) - (f(\bar{x}_2) - g(\bar{x}_2))|}{|\bar{x}_1 - \bar{x}_2|^\nu} \leq \varepsilon + \frac{2\varepsilon}{|\bar{x}_1 - \bar{x}_2|^\nu}$$

by the triangle inequality. So suppose we choose $\varepsilon$ so that

$$\varepsilon + \frac{2\varepsilon}{|\bar{x}_1 - \bar{x}_2|^\nu} \leq \frac{\delta}{2}.$$

Note that this choice of $\varepsilon$ depends on the particular $f \in \Theta_0^c$ chosen at the beginning, via $\delta$ and $\bar{x}_1$ and $\bar{x}_2$. Then for all $g \in \mathscr{C}_{0,\infty}$ with $\|f - g\|_{0,\infty} \leq \varepsilon$ we have

$$\begin{aligned} \|g\|_{0,\infty,\mathbb{1},\nu} &\geq B + \delta - \frac{\delta}{2} \\ &= B + \frac{\delta}{2} \\ &> B. \end{aligned}$$

Hence $g \in \Theta_0^c$ for all such $g$. Thus $\Theta_0^c$ is $\|\cdot\|_{0,\infty}$-open and hence $\Theta_0$ is $\|\cdot\|_{0,\infty}$-closed.

**Step 2 (Induction Step):** Next we suppose that $\Theta_{m_0}$ is $\|\cdot\|_{0,\infty}$-closed for some integer $m_0 \geq 0$. We will show that this implies $\Theta_{m_0+1}$ is $\|\cdot\|_{0,\infty}$-closed.

Since $\Theta_{m_0}$ is $\|\cdot\|_{0,\infty}$-closed, we have that for all $f$ in $\Theta_{m_0}^c = \mathscr{C}_{0,\infty} \setminus \Theta_{m_0}$ there exists an $\varepsilon > 0$ such that for all $g \in \mathscr{C}_{0,\infty}$ with

$$\|f - g\|_{0,\infty} \leq \varepsilon,$$

it holds that $g \in \Theta_{m_0}^c$. As in the base case, we will show that $\Theta_{m_0+1}^c$ is $\|\cdot\|_{0,\infty}$-open. So take an arbitrary $f \in \Theta_{m_0+1}^c$. We will show that there exists an $\varepsilon > 0$ such that for all $g \in \mathscr{C}_{0,\infty}$ with $\|f - g\|_{0,\infty} \leq \varepsilon$ we have $g \in \Theta_{m_0+1}^c$. We have to consider several cases, depending on the properties of the $f$ we're given. First, $\Theta_{m_0+1} \subsetneq \Theta_{m_0}$ implies

$$\Theta_{m_0}^c \subsetneq \Theta_{m_0+1}^c.$$

So it might be the case that $f \in \Theta_{m_0}^c$. This is case (a) below. Moreover, it is possible that $f \in \Theta_{m_0+1}^c$ but $f \notin \Theta_{m_0}^c$. This case could occur for several reasons. It might be that $f \in \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$, so $\|f\|_{m_0+1,\infty,\mathbb{1},\nu} \leq D$ for some constant $D < \infty$, but that this norm, while finite, is still too big:

$$\|f\|_{m_0+1,\infty,\mathbb{1},\nu} > B.$$

This is case (b) below. Another possibility is that $f \notin \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$. But $f \notin \Theta_{m_0}^c$, $f \in \Theta_{m_0}$ and hence its $m_0$'th derivative exists and is Hölder continuous. So there are three reasons why $f \notin \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$ could occur: Either the $(m_0+1)$'th derivative does not exist (case (c) below), the $(m_0+1)$'th derivative exists but is not $\|\cdot\|_{0,\infty}$-bounded (i.e., the first piece of the Hölder norm $\|f\|_{m_0+1,\infty,\mathbb{1},\nu}$ is infinite) (case (d) below), or the $(m_0+1)$'th derivative exists and is $\|\cdot\|_{0,\infty}$-bounded, but is not Hölder continuous (i.e., the first piece of the Hölder norm $\|f\|_{m_0+1,\infty,\mathbb{1},\nu}$ is finite, but the second piece is infinite) (case (e) below).

(a) Suppose $f \in \Theta_{m_0}^c$. But we already know from the induction assumption that $\Theta_{m_0}^c$ is open. Hence there exists an $\varepsilon > 0$ such that for all $g \in \mathscr{C}_{0,\infty}$ with $\|f - g\|_{0,\infty} \leq \varepsilon$ it holds that $g \in \Theta^c_{m_0} \subsetneq \Theta^c_{m_0+1}$.

(b) Suppose $f \notin \Theta^c_{m_0}$ and $f \in \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$ with

$$B < \|f\|_{m_0+1,\infty,\mathbb{1},\nu} \leq D$$

for some constant $D < \infty$. Since $f \notin \Theta^c_{m_0}$, $f \in \Theta_{m_0}$ and hence

$$\|f\|_{m_0,\infty,\mathbb{1},\nu} \leq B.$$

Let $g \in \mathscr{C}_{0,\infty}$ be such that $\|f - g\|_{0,\infty} \leq \varepsilon$. Remember that our goal is to find an $\varepsilon > 0$ such that all of these $g$ are in $\Theta^c_{m_0+1}$. Regardless of the value of $\varepsilon$, if $g \notin \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$ (in which case $g \notin \Theta_{m_0+1}$ and so $g \in \Theta^c_{m_0+1}$) or if $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \geq C$ for some finite constant $C > B$, then $g \in \Theta^c_{m_0+1}$. So suppose that $g \in \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$ and

$$\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C.$$

We will show that although this norm is smaller than $C$, it is still larger than $B$. For each $x \in \mathcal{D}$ and $\delta > 0$ with $x + \delta \in \mathcal{D}$,[1] the mean value theorem implies that there exists an $x_g \in [x, x + \delta]$ such that

$$g'(x_g) = \frac{g(x+\delta) - g(x)}{\delta}$$

and hence

$$\begin{aligned} g'(x) &= g'(x_g) + (g'(x) - g'(x_g)) \\ &= \frac{g(x+\delta) - g(x)}{\delta} + (g'(x) - g'(x_g)). \end{aligned}$$

Note that $g$ is differentiable because $g \in \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$. Likewise, there exists an $x_f \in [x, x + \delta]$ such that

$$f'(x) = \frac{f(x+\delta) - f(x)}{\delta} + (f'(x) - f'(x_f)).$$

[1] The cone condition implies that there exists a single $\delta > 0$ such that, for all $x \in \mathcal{D}$, at least one of $x + \delta \in \mathcal{D}$ or $x - \delta \in \mathcal{D}$ holds.

It follows that

$$
\begin{aligned}
&\|f' - g'\|_{0,\infty} \\
&= \sup_{x\in\mathcal{D}} |f'(x) - g'(x)| \\
&= \sup_{x\in\mathcal{D}} \left| \left( \frac{f(x+\delta) - f(x)}{\delta} + (f'(x) - f'(x_f)) \right) - \left( \frac{g(x+\delta) - g(x)}{\delta} + (g'(x) - g'(x_g)) \right) \right| \\
&= \sup_{x\in\mathcal{D}} \left| \frac{f(x+\delta) - g(x+\delta)}{\delta} - \frac{f(x) - g(x)}{\delta} + (f'(x) - f'(x_f)) + (g'(x) - g'(x_g)) \right| \\
&\leq \sup_{x\in\mathcal{D}} \left( \frac{|f(x+\delta) - g(x+\delta)|}{\delta} + \frac{|f(x) - g(x)|}{\delta} + |f'(x) - f'(x_f)| + |g'(x) - g'(x_g)| \right) \\
&\leq \frac{2\varepsilon}{\delta} + D\delta^\nu + C\delta^\nu
\end{aligned}
$$

The fourth line follows by the triangle inequality. The last line by $\|f - g\|_{0,\infty} \leq \varepsilon$, $x_f \in [x, x+\delta]$, $x_g \in [x, x+\delta]$, and since $f'$ and $g'$ are both Hölder continuous with Hölder constants $D$ and $C$, respectively (which follows because $\|f\|_{m_0+1,\infty,\mathbb{1},\nu} \leq D$ and $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C$).
Let $\varepsilon_1 > 0$ be arbitrary. Choose $\delta > 0$ such that $D\delta^\nu \leq \varepsilon_1/3$ and $C\delta^\nu \leq \varepsilon_1/3$. After choosing $\delta$, choose $\varepsilon$ such that $2\varepsilon/\delta \leq \varepsilon_1/3$. Thus

$$\|f' - g'\|_{0,\infty} \leq \varepsilon_1.$$

We have shown that if the first derivatives of $f$ and $g$ are Hölder continuous, we can make the derivatives for all $g$ with $\|f - g\|_{0,\infty} \leq \varepsilon$ arbitrarily close to the derivative of $f$ by choosing $\varepsilon$ small enough. An analogous argument shows that if $\|f' - g'\|_{0,\infty} \leq \varepsilon_1$ and if the second derivatives are Hölder continuous, then we can make the second derivatives arbitrarily close. Applying this argument recursively to higher order derivative shows that for any $\varepsilon_{m_0+1} > 0$, we can pick an $\varepsilon > 0$ such that for all $g$ with $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C$ and $\|f - g\|_{0,\infty} \leq \varepsilon$,

$$\|\nabla^{m_0+1} f - \nabla^{m_0+1} g\|_{0,\infty} \leq \varepsilon_{m_0+1}.$$

Our argument from the base case (step 1) now implies that if $\varepsilon_{m_0+1}$ is small enough, then $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} > B$ for all $g \in \mathscr{C}_{0,\infty}$ with $\|f - g\|_{0,\infty} \leq \varepsilon$. Hence $g \in \Theta^c_{m_0+1}$. Note that we use $\|f\|_{m_0+1,\infty,\mathbb{1},\nu} > B$ when applying the base case argument.

(c) Suppose that for some $\bar{x} \in \mathcal{D}$, $\nabla^{m_0+1} f(\bar{x})$ does not exist. Then $f \notin \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$. But since $f \notin \Theta^c_{m_0}$, we know that the $m_0$'th derivative of $f$ exists and is Hölder continuous. As in case (b), take $g \in \mathscr{C}_{0,\infty}$ such that $\|f - g\|_{0,\infty} \leq \varepsilon$ and suppose that $g \in \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$ $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C$ for $C > B$ (remember from part (b) that otherwise we know $g \in \Theta^c_{m_0+1}$ already). Since the $m_0$'th derivative of $f$ exists and is Hölder continuous, we know that the only way for the derivative $\nabla^{m_0+1} f(\bar{x})$ to not exist is if it has a kink—its right hand side derivative does not exist, its left hand side derivative does not exist, or both exist but are not equal. So we consider each of these three cases separately.

i. Suppose the right hand side derivative of $\nabla^{m_0} f$ at $\bar{x}$ does not exist. That is,

$$\lim_{h \searrow 0} \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h}$$

does not exist. Then there exists a $\delta > 0$ such that for any $\eta > 0$ we can find an $h$ with $0 < h < \eta$ and

$$\left| \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h} - \nabla^{m_0+1} g(\bar{x}) \right| > \delta.$$

If such a $\delta$ did not exists, then

$$\lim_{h \searrow 0} \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h} = \nabla^{m_0+1} g(\bar{x})$$

by definition of the limit. For such a fixed $h$, we have

$$\begin{aligned}
\delta &< \left| \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h} - \nabla^{m_0+1} g(\bar{x}) \right| \\
&\leq \left| \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} g(\bar{x}+h) + \nabla^{m_0} g(\bar{x}) - \nabla^{m_0} f(\bar{x})}{h} \right| \\
&\quad + \left| \frac{\nabla^{m_0} g(\bar{x}+h) - \nabla^{m_0} g(\bar{x})}{h} - \nabla^{m_0+1} g(\bar{x}) \right| \\
&\leq \left| \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} g(\bar{x}+h) + \nabla^{m_0} g(\bar{x}) - \nabla^{m_0} f(\bar{x})}{h} \right| \\
&\quad + \left| \nabla^{m_0+1} g(\tilde{x}) - \nabla^{m_0+1} g(\bar{x}) \right| \\
&\leq \left| \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} g(\bar{x}+h) + \nabla^{m_0} g(\bar{x}) - \nabla^{m_0} f(\bar{x})}{h} \right| + Ch^{\nu}.
\end{aligned}$$

The second line follows by the triangle inequality. The third line by the mean value theorem, since $\nabla^{m_0} g$ is differentiable, and here $\tilde{x} \in [\bar{x}, \bar{x}+h]$. The fourth line follows since $\nabla^{m_0+1} g$ is Hölder continuous with constant $C$, and since $\tilde{x} \in [\bar{x}, \bar{x}+h]$ so that $\|\tilde{x} - \bar{x}\| \leq h$. Now choose $h$ small enough such that $Ch^{\nu} \leq \delta/2$. For this fixed $h$, pick $\varepsilon$ small enough such that

$$\|\nabla^{m_0} f - \nabla^{m_0} g\|_{0,\infty} \leq \frac{\delta h}{4}.$$

Then

$$\delta < \left| \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h} - \nabla^{m_0+1} g(\bar{x}) \right| \leq \delta,$$

a contraction.

ii. Suppose the left hand side derivative of $\nabla^{m_0} f$ at $\bar{x}$ does not exist. That is,

$$\lim_{h\searrow 0} \frac{\nabla^{m_0} f(\bar{x}) - \nabla^{m_0} f(\bar{x}-h)}{h}$$

does not exist. This case proceeds analogously to the previous case.

iii. Both the left hand and right hand side derivatives of $\nabla^{m_0} f$ at $\bar{x}$ exist, but they are not equal:

$$\lim_{h\searrow 0} \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h} \neq \lim_{h\searrow 0} \frac{\nabla^{m_0} f(\bar{x}) - \nabla^{m_0} f(\bar{x}-h)}{h}.$$

Considering the distance between the right hand side and left hand side secant lines, for any $h > 0$ such that $[\bar{x}-h, \bar{x}+h] \subseteq \mathcal{D}$, we obtain

$$\begin{aligned}
&\left|\left(\frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h}\right) - \left(\frac{\nabla^{m_0} f(\bar{x}) - \nabla^{m_0} f(\bar{x}-h)}{h}\right)\right| \\
&\quad \leq 4\frac{\varepsilon_{m_0}}{h} + \left|\left(\frac{\nabla^{m_0} g(\bar{x}+h) - \nabla^{m_0} g(\bar{x})}{h}\right) - \left(\frac{\nabla^{m_0} g(\bar{x}) - \nabla^{m_0} g(\bar{x}-h)}{h}\right)\right| \\
&\quad = 4\frac{\varepsilon_{m_0}}{h} + \left|\left(\nabla^{m_0+1} g(\tilde{x}_1) - \nabla^{m_0+1} g(\tilde{x}_2)\right)\right| \\
&\quad \leq 4\frac{\varepsilon_{m_0}}{h} + C(2h)^{\nu}.
\end{aligned}$$

For the first line, we used the triangle inequality plus the fact that for any $\varepsilon_{m_0} > 0$, there exists an $\varepsilon > 0$ not depending on $g$ such that $\|f-g\|_{0,\infty} \leq \varepsilon$ implies

$$\|\nabla^{m_0} f - \nabla^{m_0} g\|_{0,\infty} \leq \varepsilon_{m_0}.$$

This follows from our argument in part (b), since $\nabla^{m_0} f$ and $\nabla^{m_0} g$ are Hölder continuous.

In the second line, we used the mean value theorem, since $g \in \mathscr{C}_{m_0+1,\infty,\mathbb{1},\nu}$, where $\tilde{x}_1 \in [\bar{x}, \bar{x}+h]$ and $\tilde{x}_2 \in [\bar{x}-h, \bar{x}]$. In the third line we used Hölder continuity of $\nabla^{m_0+1} g$ since $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C$, plus the fact that $|\tilde{x}_1 - \tilde{x}_2| \leq 2h$.

Since

$$\lim_{h\searrow 0} \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h} \neq \lim_{h\searrow 0} \frac{\nabla^{m_0} f(\bar{x}) - \nabla^{m_0} f(\bar{x}-h)}{h}$$

there exists a $\delta > 0$ such that for an arbitrarily small $h$

$$\left|\left(\frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h}\right) - \left(\frac{\nabla^{m_0} f(\bar{x}) - \nabla^{m_0} f(\bar{x}-h)}{h}\right)\right| > \delta.$$

Choose $h$ such that $C(2h)^{\nu} \leq \delta/2$. Then for this fixed $h$, pick $\varepsilon$ small enough such that $4\varepsilon_{m_0}/h \leq \delta/2$. Then

$$\delta < \left| \left( \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h} \right) - \left( \frac{\nabla^{m_0} f(\bar{x}) - \nabla^{m_0} f(\bar{x}-h)}{h} \right) \right| \leq \delta,$$

a contraction.

In all three cases where $\nabla^{m_0+1} f(\bar{x})$ does not exist, we have derived a contradiction. Hence there does not exist a $g \in \mathscr{C}_{0,\infty}$ with $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C$ and $\|f-g\|_{0,\infty} \leq \varepsilon$. This implies that for all $g \in \mathscr{C}_{0,\infty}$ with $\|f-g\|_{0,\infty} \leq \varepsilon$ it holds that $g \in \Theta^c_{m_0+1}$.

(d) Suppose $\nabla^{m_0+1} f(x)$ exists for all $x \in \mathcal{D}$ but

$$\sup_{x \in \mathcal{D}} |\nabla^{m_0+1} f(x)| = \infty.$$

For example, this happens with $f(x) = \sqrt{x}$ when $\mathcal{D} = (0,1)$ and $m_0 = 0$. Then there exists a $\bar{x} \in \mathcal{D}$ such that

$$C < |\nabla^{m_0+1} f(\bar{x})| < \infty$$

for some constant $C > B$. Thus, for all $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C$,

$$\begin{aligned}
|\nabla^{m_0+1} g(\bar{x})| &\geq |\nabla^{m_0+1} f(\bar{x})| - \left|\nabla^{m_0+1} g(\bar{x}) - \nabla^{m_0+1} f(\bar{x})\right| \\
&= |\nabla^{m_0+1} f(\bar{x})| - \left| \lim_{h \to 0} \frac{\nabla^{m_0} g(\bar{x}+h) - \nabla^{m_0} g(\bar{x})}{h} - \lim_{h \to 0} \frac{\nabla^{m_0} f(\bar{x}+h) - \nabla^{m_0} f(\bar{x})}{h} \right| \\
&= |\nabla^{m_0+1} f(\bar{x})| - \lim_{h \to 0} \left| \frac{\nabla^{m_0} g(\bar{x}+h) - \nabla^{m_0} f(\bar{x}+h)}{h} - \frac{\nabla^{m_0} g(\bar{x}) - \nabla^{m_0} f(\bar{x})}{h} \right|.
\end{aligned}$$

The first line follows by the reverse triangle inequality. Since the limit in the last line exists and is finite, for any $\delta > 0$, we can find an $\bar{h} > 0$ with $[\bar{x}, \bar{x}+\bar{h}] \subseteq \mathcal{D}$ such that the difference between the limit and the term we're taking the limit of evaluated at $\bar{h}$ is smaller than $\delta$. Hence

$$\begin{aligned}
|\nabla^{m_0+1} g(\bar{x})| &\geq |\nabla^{m_0+1} f(\bar{x})| - \left| \frac{\nabla^{m_0} g(\bar{x}+\bar{h}) - \nabla^{m_0} f(\bar{x}+\bar{h})}{\bar{h}} - \frac{\nabla^{m_0} g(\bar{x}) - \nabla^{m_0} f(\bar{x})}{\bar{h}} \right| - \delta \\
&\geq C - \delta - \left| \frac{\nabla^{m_0} g(\bar{x}+\bar{h}) - \nabla^{m_0} f(\bar{x}+\bar{h})}{\bar{h}} - \frac{\nabla^{m_0} g(\bar{x}) - \nabla^{m_0} f(\bar{x})}{\bar{h}} \right|.
\end{aligned}$$

As in part (b), for any $\varepsilon_{m_0} > 0$, there is an $\varepsilon > 0$ such that $\|f-g\|_{0,\infty} \leq \varepsilon$ implies

$$\|\nabla^{m_0} f - \nabla^{m_0} g\|_{0,\infty} \leq \varepsilon_{m_0}.$$

Let $\varepsilon_{m_0}$ such that

$$\left| \frac{\nabla^{m_0} g(\bar{x}+\bar{h}) - \nabla^{m_0} f(\bar{x}+\bar{h})}{\bar{h}} - \frac{\nabla^{m_0} g(\bar{x}) - \nabla^{m_0} f(\bar{x})}{\bar{h}} \right| \leq \delta.$$

Then

$$\begin{aligned}|\nabla^{m_0+1}g(\bar{x})| &\geq C - 2\delta \\ &> B\end{aligned}$$

where the last line follows if we choose $\delta > 0$ such that $C-2\delta > B$, that is, $\delta < (C-B)/2$, which is possible since $C > B$. We have shown that the first piece of the Hölder norm $\|g\|_{m_0+1,\infty,\mathbb{1},\nu}$ is larger than $B$, and so the entire norm is larger than $B$ and hence $g \in \Theta^c_{m_0+1}$.

(e) Finally, suppose

$$\sup_{x\in\mathcal{D}} |\nabla^{m_0+1}f(x)| \leq D < \infty$$

but $\nabla^{m_0+1}f$ is not Hölder continuous:

$$\sup_{x_1,x_2\in\mathcal{D},x_1\neq x_2} \frac{|\nabla^{m_0+1}f(x_1) - \nabla^{m_0+1}f(x_2)|}{|x_1-x_2|^\nu} = \infty.$$

Again take $g \in \mathscr{C}_{0,\infty}$ such that $\|f-g\|_{0,\infty} \leq \varepsilon$ and suppose that $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C$ for $C > B$. Since $\nabla^{m_0+1}f$ is not Hölder continuous, there exist $x_1$ and $x_2$ in $\mathcal{D}$, $x_1 \neq x_2$, such that

$$\left|\frac{\nabla^{m_0+1}f(x_1) - \nabla^{m_0+1}f(x_2)}{|x_1-x_2|^\nu}\right| > B + C.$$

Moreover, by the triangle inequality,

$$\begin{aligned}&\left|\frac{\nabla^{m_0+1}f(x_1) - \nabla^{m_0+1}f(x_2)}{|x_1-x_2|^\nu}\right| \\ &\leq \left|\frac{\nabla^{m_0+1}g(x_1) - \nabla^{m_0+1}g(x_2)}{|x_1-x_2|^\nu}\right| + \\ &+ \lim_{h\to 0}\left|\frac{(\nabla^{m_0}g(x_1+h) - \nabla^{m_0}g(x_1)) - (\nabla^{m_0}f(x_1+h) - \nabla^{m_0}f(x_1))}{h}\Big/|x_1-x_2|^\nu\right| \\ &+ \lim_{h\to 0}\left|\frac{(\nabla^{m_0}g(x_2+h) - \nabla^{m_0}g(x_2)) - (\nabla^{m_0}f(x_2+h) - \nabla^{m_0}f(x_2))}{h}\Big/|x_1-x_2|^\nu\right|.\end{aligned}$$

As in part (b), for any $\varepsilon_{m_0} > 0$, there is an $\varepsilon > 0$ such that $\|f-g\|_{0,\infty} \leq \varepsilon$ implies

$$\|\nabla^{m_0}f - \nabla^{m_0}g\|_{0,\infty} \leq \varepsilon_{m_0}.$$

Returning to our previous inequality, we see that since the limits on the right hand side are finite and since $\nabla^{m_0+1}g$ is Hölder continuous, for any $\delta > 0$ there is an $\bar{h} > 0$ which does not depend on $g$ such that

$$
\begin{aligned}
&\left|\frac{\nabla^{m_0+1} f(x_1) - \nabla^{m_0+1} f(x_2)}{|x_1 - x_2|^\nu}\right| \\
&\leq \left|\frac{\nabla^{m_0+1} g(x_1) - \nabla^{m_0+1} g(x_2)}{|x_1 - x_2|^\nu}\right| \\
&+ \left|\frac{\big(\nabla^{m_0} g(x_1 + \bar{h}) - \nabla^{m_0} g(x_1)\big) - \big(\nabla^{m_0} f(x_1 + \bar{h}) - \nabla^{m_0} f(x_1)\big)}{\bar{h}} \Bigg/ |x_1 - x_2|^\nu\right| \\
&+ \left|\frac{\big(\nabla^{m_0} g(x_2 + \bar{h}) - \nabla^{m_0} g(x_2)\big) - \big(\nabla^{m_0} f(x_2 + \bar{h}) - \nabla^{m_0} f(x_2)\big)}{\bar{h}} \Bigg/ |x_1 - x_2|^\nu\right| + \delta \\
&\leq C + \frac{4\varepsilon_{m_0}}{\bar{h}|x_1 - x_2|^\nu} + \delta.
\end{aligned}
$$

This is the same argument we used in part (d). In the last line we used $\|g\|_{m_0+1,\infty,\mathbb{1},\nu} \leq C$, the triangle inequality, and $\|\nabla^{m_0} f - \nabla^{m_0} g\|_{0,\infty} \leq \varepsilon_{m_0}$. Choose $\delta = B/2$. Then choose $\varepsilon_{m_0}$ small enough so that

$$
\frac{4\varepsilon_0}{\bar{h}|x_1 - x_2|^\nu} < \frac{B}{2}.
$$

Combining our results, we have shown

$$
C + B < \left|\frac{\nabla^{m_0+1} f(x_1) - \nabla^{m_0+1} f(x_2)}{|x_1 - x_2|^\nu}\right| \leq C + B,
$$

a contradiction.

□

*Proof of theorem 4 (Closedness under equal weightings).*

1. We want to show that the $\|\cdot\|_s = \|\cdot\|_{m+m_0,2,\mu_s}$-ball $\Theta$ is $\|\cdot\|_c = \|\cdot\|_{m,\infty,\mu_c^{1/2}}$-closed. Part 1 of our compact embedding result theorem 3 says that $\mathscr{W}_{m+m_0,2,\mu_s}$ is compactly embedded in $\mathscr{C}_{m,\infty,\mu_c^{1/2}}$. Now consider the space $(\mathscr{W}_{m,2,\mu_a}, \|\cdot\|_{m,2,\mu_a})$ where $\mu_a$ is such that

$$
\int_{\mathbb{R}^{d_x}} \frac{\mu_a(x)}{\mu_c(x)}\, dx \leq C_1.
$$

Then for any $f \in \mathscr{C}_{m,\infty,\mu_c^{1/2}}$,

$$\begin{aligned}
\|f\|^2_{m,2,\mu_a} &= \sum_{0 \leq |\lambda| \leq m} \int_{\mathbb{R}^{d_x}} |\nabla^\lambda f(x)|^2 \mu_a(x)\, dx \\
&= \sum_{0 \leq |\lambda| \leq m} \int_{\mathbb{R}^{d_x}} |\nabla^\lambda f(x)|^2 \mu_c(x) \frac{\mu_a(x)}{\mu_c(x)}\, dx \\
&\leq C\|f\|^2_{m,\infty,\mu_c^{1/2}} \int_{\mathbb{R}^{d_x}} \frac{\mu_a(x)}{\mu_c(x)}\, dx \\
&\leq CC_1 \|f\|_{m,\infty,\mu_c^{1/2}}.
\end{aligned}$$

Hence

$$\mathscr{C}_{m,\infty,\mu_c^{1/2}} \subseteq \mathscr{W}_{m,2,\mu_a}.$$

But we also know that $\mathscr{W}_{m+m_0,2,\mu_s}$ is compactly embedding in $\mathscr{C}_{m,\infty,\mu_c^{1/2}}$. Therefore, by lemma 4, $\mathscr{W}_{m+m_0,2,\mu_s}$ is compactly embedded in $\mathscr{W}_{m,2,\mu_a}$. Both of these are separable Hilbert spaces by arguments as in the proof of theorem 3.6 in Kufner (1980), which is analogous to Adams and Fournier (2003) theorem 3.6. Hence lemma A.1 of Santos (2012) implies that $\Theta$ is $\|\cdot\|_{m,2,\mu_a}$-closed. But now lemma 2 and the inequality $\|\cdot\|_{m,2,\mu_a} \leq (CC_1)^{1/2}\|\cdot\|_{m,\infty,\mu_c^{1/2}}$ imply that $\Theta$ is $\|\cdot\|_{m,\infty,\mu_c^{1/2}}$-closed.

2. We want to show that the $\|\cdot\|_s = \|\cdot\|_{m+m_0,2,\mu_s}$-ball $\Theta$ is $\|\cdot\|_c = \|\cdot\|_{m,2,\mu_c}$-closed. Part 2 of our compact embedding result theorem 3 says that $\mathscr{W}_{m+m_0,2,\mu_s}$ is compactly embedded in $\mathscr{W}_{m,2,\mu_c}$. Both of these are separable Hilbert spaces, as discussed in the previous part. Hence lemma A.1. of Santos (2012) implies that $\Theta$ is $\|\cdot\|_{m,2,\mu_c}$-closed.

3. We want to show that the $\|\cdot\|_s = \|\cdot\|_{m+m_0,\infty,\mu_s}$-ball $\Theta$ is not $\|\cdot\|_c = \|\cdot\|_{m,\infty,\mu_c}$-closed. The same counterexample from the proof of part 3 of theorem 2 can be adapted here as well, by smoothly extending its domain definition to $\mathcal{D} = \mathbb{R}$.

4. We want to show that the $\|\cdot\|_s = \|\cdot\|_{m+m_0,\infty,\mu_s}$-ball $\Theta$ is not $\|\cdot\|_c = \|\cdot\|_{m,2,\mu_c}$-closed. As in the previous part, this can be shown by extending the same counterexample from theorem 2.

□

*Proof of theorem 6 (Closedness under product weightings).* **Cases 1 and 2**. This follows exactly as in the proof of theorem 5, except we apply theorem 4 and then lemma S1 part 2

**Case 3**. As in theorem 4, we can adapt the counterexample from theorem 2 by smoothly extending its domain to $\mathcal{D} = \mathbb{R}$.

**Case 4**. Assume $d_x = 1$ for simplicity. This proof is a close modification to the corresponding proof of theorem 2 for bounded domains. As in that proof, it suffices to prove the result for $m = 0$. For any $g \in \mathscr{C}_{m_0,\infty,\mu_s,\nu}$ define $g_s(x) = \mu_s(x)g(x)$ and $g_c(x) = \mu_c(x)g(x)$. We want to prove that

$$\Theta_{m_0} \equiv \{g \in \mathscr{C}_{m_0,\infty,\mu_s,\nu} : \|g\|_{m_0,\infty,\mu_s,\nu} \leq B\}$$

is $\|\cdot\|_{m_0,\infty,\mu_c}$-closed, for all $m_0 \geq 0$. We proceed by induction on $m_0$.

**Step 1 (Base Case):** Let $m_0 = 0$. We want to show that $\Theta_0$ is $\|\cdot\|_{0,\infty,\mu_c}$-closed, so we will show that its complement $\Theta_0^c = \mathscr{C}_{0,\infty,\mu_c} \setminus \Theta_0$ is $\|\cdot\|_{0,\infty,\mu_c}$-open. So take an arbitrary $f \in \Theta_0^c$. We will show that there exists an $\varepsilon > 0$ such that

$$\{g \in \mathscr{C}_{0,\infty,\mu_c} : \|f - g\|_{0,\infty,\mu_c} \leq \varepsilon\} \subseteq \Theta_0^c.$$

Since $f$ is outside the weighted Hölder ball $\Theta_0$, its weighted Hölder norm is larger than $B$,

$$\sup_{x\in\mathbb{R}} |f_s(x)| + \sup_{x_1,x_2\in\mathbb{R}} \frac{|f_s(x_1) - f_s(x_2)|}{|x_1 - x_2|^\nu} > B.$$

Hence there exist points $\bar{x}, \bar{x}_1, \bar{x}_2 \in \mathbb{R}$ with $\bar{x}_1 \neq \bar{x}_2$ such that

$$|f_s(\bar{x})| + \frac{|f_s(\bar{x}_1) - f_s(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} > B.$$

Define

$$\delta = |f_s(\bar{x})| + \frac{|f_s(\bar{x}_1) - f_s(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} - B > 0.$$

Next, for all $g \in \mathscr{C}_{0,\infty,\mu_c}$,

$$\begin{aligned}
\|g\|_{0,\infty,\mu_s,\nu} &\geq |g_s(\bar{x})| + \frac{|g_s(\bar{x}_1) - g_s(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} \\
&\geq |f_s(\bar{x})| - |f_s(\bar{x}) - g_s(\bar{x})| \\
&\qquad + \frac{|f_s(\bar{x}_1) - f_s(\bar{x}_2)|}{|\bar{x}_1 - \bar{x}_2|^\nu} - \frac{|(f_s(\bar{x}_1) - g_s(\bar{x}_1)) - (f_s(\bar{x}_2) - g_s(\bar{x}_2))|}{|\bar{x}_1 - \bar{x}_2|^\nu} \\
&= B + \delta - \left( |f_s(\bar{x}) - g_s(\bar{x})| + \frac{|(f_s(\bar{x}_1) - g_s(\bar{x}_1)) - (f_s(\bar{x}_2) - g_s(\bar{x}_2))|}{|\bar{x}_1 - \bar{x}_2|^\nu} \right) \\
&= B + \delta - \left( |f_c(\bar{x}) - g_c(\bar{x})| \frac{\mu_s(\bar{x})}{\mu_c(\bar{x})} + \frac{\left|(f_c(\bar{x}_1) - g_c(\bar{x}_1)) \frac{\mu_s(\bar{x}_1)}{\mu_c(\bar{x}_1)} - (f_c(\bar{x}_2) - g_c(\bar{x}_2)) \frac{\mu_s(\bar{x}_2)}{\mu_c(\bar{x}_2)}\right|}{|\bar{x}_1 - \bar{x}_2|^\nu} \right).
\end{aligned}$$

For all $g \in \mathscr{C}_{0,\infty,\mu_c}$ with

$$\|f - g\|_{0,\infty,\mu_c} = \|f_c - g_c\|_{0,\infty} \leq \varepsilon$$

we have

$$|f_c(\bar{x}) - g_c(\bar{x})| \frac{\mu_s(\bar{x})}{\mu_c(\bar{x})} + \frac{\left|(f_c(\bar{x}_1) - g_c(\bar{x}_1)) \frac{\mu_s(\bar{x}_1)}{\mu_c(\bar{x}_1)} - (f_c(\bar{x}_2) - g_c(\bar{x}_2)) \frac{\mu_s(\bar{x}_2)}{\mu_c(\bar{x}_2)}\right|}{|\bar{x}_1 - \bar{x}_2|^\nu} \leq \varepsilon \frac{\mu_s(\bar{x})}{\mu_c(\bar{x})} + \frac{\varepsilon \frac{\mu_s(\bar{x}_1)}{\mu_c(\bar{x}_1)} + \varepsilon \frac{\mu_s(\bar{x}_2)}{\mu_c(\bar{x}_2)}}{|\bar{x}_1 - \bar{x}_2|^\nu}$$

by the triangle inequality. So suppose we choose $\varepsilon$ small enough that the right hand side is $\leq \delta/2$.

Then for all $g \in \mathscr{C}_{0,\infty,\mu_c}$ with $\|f - g\|_{0,\infty,\mu_c} \leq \varepsilon$ we have

$$\begin{aligned}\|g\|_{0,\infty,\mu_s,\nu} &\geq B + \delta - \frac{\delta}{2} \\ &> B.\end{aligned}$$

Hence $g \in \Theta_0^c$ for all such $g$. Thus $\Theta_0^c$ is $\|\cdot\|_{0,m,\mu_c}$-open and hence $\Theta_0$ is $\|\cdot\|_{0,m,\mu_c}$-closed.

**Step 2 (Induction Step):** This step follows the same arguments as those with bounded support. As in step 1, the main idea is simply to replace $g$ with either $g_c$ or $g_s$, as appropriate. □

*Proof of theorem 8 (Closedness for weighted norms on bounded domains).* This proof is identical to the proof of theorem 6, except that now we use the compact embedding results of theorem 7 when necessary. □

# E Proofs of propositions from section 4

*Proof of proposition 1.* This proof is straightforward and we therefore omit it. □

*Proof of proposition 2.* This proof is straightforward and we therefore omit it. □

*Proof of proposition 3.* This proof is given in Gallant and Nychka (1987) as lemma A.2, and hence we omit it. □

*Proof of proposition 4.* This proof is similar to the proof of proposition 3, which was shown in lemma A.2 of Gallant and Nychka (1987). Let $\mathcal{C} \subseteq \mathcal{D}$ be compact. We prove the proposition by induction on $m$ (letting $m_0 = 0$, since it is irrelevant for the present result). For the base case, $m = 0$, the result holds trivially by letting $K_{\mathcal{C}} = 1$. Next suppose it holds for $m - 1$. Choose $\lambda$ such that $|\lambda| = m$ and let $\nabla^\lambda = \nabla^\beta \nabla^\alpha$ where $|\alpha| = 1$ and $|\beta| = m - 1$. The result holds trivially if $\delta_s = 0$, so let $\delta_s \neq 0$. Then

$$\begin{aligned}\nabla^\lambda[\mu_s^{1/2}(x)] &= \nabla^\lambda \left[\exp\left(\frac{\delta_s}{2}(x'x)\right)\right] \\ &= \nabla^\beta \left(\nabla^\alpha \left[\exp\left(\frac{\delta_s}{2}(x'x)\right)\right]\right) \\ &= \nabla^\beta \left(\frac{\delta_s}{2} \exp\left(\frac{\delta_s}{2}(x'x)\right) \cdot \nabla^\alpha (x'x)\right) \\ &= \frac{\delta_s}{2} \sum_{\gamma \leq \beta} \begin{bmatrix}\beta \\ \gamma\end{bmatrix} \nabla^\gamma \left[\exp\left(\frac{\delta_s}{2}(x'x)\right)\right] \nabla^{\alpha+\beta-\gamma}(x'x) \\ &= \frac{\delta_s}{2} \sum_{\gamma \leq \beta} \begin{bmatrix}\beta \\ \gamma\end{bmatrix} [\nabla^\gamma \mu_s^{1/2}(x)] \nabla^{\alpha+\beta-\gamma}(x'x).\end{aligned}$$

In the fourth line we used Leibniz's formula. Next,

$$\begin{aligned}|\nabla^{\alpha+\beta-\gamma}(x'x)| &\leq \sum_{i=1}^{d_x}(x_i^2+2|x_i|+2)\\ &\leq 4(1+x'x).\end{aligned}$$

Hence

$$\begin{aligned}|\nabla^\lambda[\mu_s^{1/2}(x)]| &\leq \frac{|\delta_s|}{2}\sum_{\gamma\leq\beta}\begin{bmatrix}\beta\\ \gamma\end{bmatrix}|\nabla^\gamma\mu_s^{1/2}(x)|\cdot|4(1+x'x)|\\ &\leq 2|\delta_s|\sum_{\gamma\leq\beta}\begin{bmatrix}\beta\\ \gamma\end{bmatrix}K_{\mathcal{C},m-1}\mu_s^{1/2}(x)\cdot|1+x'x|\\ &\leq 2|\delta_s|\sum_{\gamma\leq\beta}\begin{bmatrix}\beta\\ \gamma\end{bmatrix}K_{\mathcal{C},m-1}\mu_s^{1/2}(x)\cdot M_{\mathcal{C}}\\ &= \mu_s^{1/2}(x)\left(2|\delta_s|\sum_{\gamma\leq\beta}\begin{bmatrix}\beta\\ \gamma\end{bmatrix}K_{\mathcal{C},m-1}\cdot M_{\mathcal{C}}\right).\end{aligned}$$

Here $M_{\mathcal{C}}=\sup_{x\in\mathcal{C}}|1+x'x|$, which is finite since $\mathcal{C}$ is compact. The second line follows by the induction hypothesis. ☐

*Proof of proposition 5.* Pick $g(x)=1+x'x$. Notice that $g(x)\to\infty$ as $\|x\|_e\to\infty$. We prove the result by showing that for any $0\leq|\lambda|\leq m_0$,

$$\nabla^\lambda\tilde{\mu}_c^{1/2}(x)=\exp\left[\frac{\delta_c}{2}(x'x)\right]\cdot p_\lambda(x) \tag{$*$}$$

for some polynomial $p_\lambda(x)$. Consequently, dividing by $\mu_s^{1/2}(x)$ yields

$$\frac{\nabla^\lambda\tilde{\mu}_c^{1/2}(x)}{\mu_s^{1/2}(x)}=\exp\left[\frac{\delta_c-\delta_s}{2}(x'x)\right]\cdot p_\lambda(x).$$

Since $\delta_c<\delta_s$,

$$\left|\frac{\nabla^\lambda\tilde{\mu}_c^{1/2}(x)}{\mu_s^{1/2}(x)}\right|$$

converges to zero as $\|x\|_e\to\infty$. This implies there is a $J$ such that for all $x$ with $\|x\|_e>J$, this ratio is smaller than $M_1$. For all $x$ with $\|x\|_e\leq J$, this ratio is a continuous function (the product of an exponential and a polynomial) on a compact set, and hence achieves a maximum $M_2$. Let $M=\max\{M_1,M_2\}$. Thus the ratio is bounded by $M$ for all $x\in\mathbb{R}^{d_x}$.

So it suffices to show equation ($*$). We proceed by induction. For the base case, $|\lambda| = 0$,

$$\begin{aligned} \nabla^0 \tilde{\mu}_c^{1/2}(x) &= \exp[\delta_c(x'x)/2] \cdot g(x) \\ &= \exp[\delta_c(x'x)/2] \cdot (1 + x^2). \end{aligned}$$

So the base case holds with $p_0(x) = g(x) = 1 + x^2$. Next, suppose it holds for $|\lambda| = m - 1$. Choose $\lambda$ such that $|\lambda| = m$ and let $\nabla^\lambda = \nabla^\beta \nabla^\alpha$ where $|\alpha| = 1$ and $|\beta| = m - 1$. Then

$$\begin{aligned} \nabla^\lambda[\tilde{\mu}_c^{1/2}(x)] &= \nabla^\alpha[\nabla^\beta \tilde{\mu}_c^{1/2}(x)] \\ &= \nabla^\alpha[\exp[\delta_c(x'x)/2] \cdot p_\beta(x)] \\ &= \exp[\delta_c(x'x/2)](\delta_c/2) p_\beta(x) \nabla^\alpha(x'x) + \exp[\delta_c(x'x)/2] \nabla^\alpha p_\beta(x) \\ &= \exp[\delta_c(x'x)/2] \left( (\delta_c/2) p_\beta(x) \nabla^\alpha(x'x) + \nabla^\alpha p_\beta(x) \right). \end{aligned}$$

Since the derivative of a polynomial is a polynomial, we're done. □

*Proof of proposition 6.*

1. This follows immediately from lemmas 5 and 7:

$$\|\mu^{1/2} f\|_{m,2} \leq M_1 \|f\|_{m,2,\mu} \leq M_1 M \|\mu^{1/2} f\|_{m,2}.$$

2. This follows immediately from lemmas 6 and 8.

□

## F Proofs of propositions from section 5

*Proof of proposition 7.* Suppose such a function $\mu$ existed. Define $g : (0,1) \to \mathbb{R}$ by $g(x) = \log \mu(x)$. Then (1) implies that $g(x) \to -\infty$ as $x \to 0$. (2) implies that

$$g'(x) = \frac{1}{\mu(x)} \mu'(x) \leq K.$$

Hence $|g'(x)| \leq K$ for all $x \in (0,1)$. This is a contradiction to $g(x) \to -\infty$ as $x \to 0$. □

*Proof of proposition 8.* First consider the polynomial weight case, $\mu_s(x) = [x(1-x)]^{\delta_s}$. The proof is similar to the proof of propositions 3. We proceed by induction. For the base case $m = 0$, the result holds trivially by letting $K_{\mathcal{C}} = 1$. Next suppose it holds for $m - 1$. If $\delta_s = 0$ the result holds trivially, so let $\delta_s \neq 0$. We have

$$\begin{aligned}
\nabla^m[\mu_s^{1/2}(x)] &= \nabla^m \left([x(1-x)]^{\delta_s/2}\right) \\
&= \nabla^{m-1}\nabla^1 \left([x(1-x)]^{\delta_s/2}\right) \\
&= \nabla^{m-1} \left(\frac{\delta_s}{2}[x(1-x)]^{\delta_s/2-1}\nabla^1[x(1-x)]\right) \\
&= \frac{\delta_s}{2} \sum_{\gamma \leq m-1} \begin{bmatrix} m-1 \\ \gamma \end{bmatrix} \nabla^\gamma \left([x(1-x)]^{\delta_s/2-1}\right) \nabla^{1+(m-1)-\gamma}[x(1-x)] \\
&= \frac{\delta_s}{2} \sum_{\gamma \leq m-1} \begin{bmatrix} m-1 \\ \gamma \end{bmatrix} \nabla^\gamma \left(\mu_{s,\tilde{\delta}}^{1/2}(x)\right) \nabla^{m-\gamma}[x(1-x)].
\end{aligned}$$

Here $\tilde{\delta} = \delta_s - 1/2$. $\nabla^n[x(1-x)]$ is either $x - x^2$ for $n = 0$, $1 - 2x$ for $n = 1$, $-2$ for $n = 2$, and 0 for $n > 2$. Hence

$$\begin{aligned}
M_\mathcal{C} &\equiv \sup_{x \in \mathcal{C}} |\nabla^{m-\gamma}[x(1-x)]| \\
&< \infty
\end{aligned}$$

since $\mathcal{D}$ is bounded. So for all $x \in \mathcal{C}$,

$$\begin{aligned}
|\nabla^m[\mu_s^{1/2}(x)]| &\leq \frac{|\delta_s|}{2} \sum_{\gamma \leq m-1} \begin{bmatrix} m-1 \\ \gamma \end{bmatrix} |\nabla^\gamma[\mu_{s,\tilde{\delta}}^{1/2}(x)]| \cdot |\nabla^{m-\gamma}[x(1-x)]| \\
&\leq \frac{|\delta_s|}{2} \sum_{\gamma \leq m-1} \begin{bmatrix} m-1 \\ \gamma \end{bmatrix} K_{\mathcal{C},m-1}\mu_{s,\tilde{\delta}}^{1/2}(x) \cdot M_\mathcal{C} \\
&= \mu_{s,\tilde{\delta}}^{1/2}(x) \left(\frac{|\delta_s|}{2} \sum_{\gamma \leq m-1} \begin{bmatrix} m-1 \\ \gamma \end{bmatrix} K_{\mathcal{C},m-1}M_\mathcal{C}\right) \\
&= [x(1-x)]^{\delta_s/2-1} \left(\frac{|\delta_s|}{2} \sum_{\gamma \leq m-1} \begin{bmatrix} m-1 \\ \gamma \end{bmatrix} K_{\mathcal{C},m-1}M_\mathcal{C}\right) \\
&= \mu_s^{1/2}(x)\frac{1}{x(1-x)} \left(\frac{|\delta_s|}{2} \sum_{\gamma \leq m-1} \begin{bmatrix} m-1 \\ \gamma \end{bmatrix} K_{\mathcal{C},m-1}M_\mathcal{C}\right) \\
&\leq \mu_s^{1/2}(x)M'_\mathcal{C} \left(\frac{|\delta_s|}{2} \sum_{\gamma \leq m-1} \begin{bmatrix} m-1 \\ \gamma \end{bmatrix} K_{\mathcal{C},m-1}M_\mathcal{C}\right).
\end{aligned}$$

The second line follows by our $M_\mathcal{C}$ bound from above, and by the induction hypothesis with constant $K_{\mathcal{C},m-1}$. The last line follows since $\mathcal{C} \subseteq (0,1)$ is compact, and hence $x$ is bounded away from zero and one. So

$$M'_\mathcal{C} \equiv \sup_{x \in \mathcal{C}} \frac{1}{x(1-x)} < \infty.$$

Next consider the exponential weight case, $\mu_s(x) = \exp[\delta_s x^{-1}(1-x)^{-1}]$. The proof for this case is similar to the proofs of propositions 3 and 4. Let $\mathcal{C} \subseteq \mathcal{D}$ be compact. We prove the proposition by induction on $m$ (letting $m_0 = 0$, since it is irrelevant for the present result). For the base case, $m = 0$, the result holds trivially by letting $K_{\mathcal{C}} = 1$. Next suppose it holds for $m-1$. The result holds trivially if $\delta_s = 0$, so let $\delta_s \neq 0$. Then

$$\begin{aligned}
\nabla^m[\mu_s^{1/2}(x)] &= \nabla^m \left[\exp\left(\frac{\delta_s}{2}\frac{1}{x(1-x)}\right)\right] \\
&= \nabla^{m-1}\left(\nabla^1\left[\exp\left(\frac{\delta_s}{2}\frac{1}{x(1-x)}\right)\right]\right) \\
&= \nabla^{m-1}\left(\frac{\delta_s}{2}\exp\left(\frac{\delta_s}{2}\frac{1}{x(1-x)}\right)\cdot\nabla^1\left(\frac{1}{x(1-x)}\right)\right) \\
&= \frac{\delta_s}{2}\sum_{\gamma\leq m-1}\begin{bmatrix} m-1 \\ \gamma \end{bmatrix}\nabla^\gamma\left[\exp\left(\frac{\delta_s}{2}\frac{1}{x(1-x)}\right)\right]\nabla^{1+(m-1)-\gamma}\left(\frac{1}{x(1-x)}\right) \\
&= \frac{\delta_s}{2}\sum_{\gamma\leq m-1}\begin{bmatrix} m-1 \\ \gamma \end{bmatrix}[\nabla^\gamma\mu_s^{1/2}(x)]\nabla^{m-\gamma}\left(\frac{1}{x(1-x)}\right).
\end{aligned}$$

In the fourth line we used Leibniz's formula. Next, for any natural number $n$,

$$\nabla^n\left(\frac{1}{x(1-x)}\right) = n!\sum_{j=0}^{n}\frac{(-1)^{n-j}}{(1-x)^{j+1}x^{n+1-j}}.$$

Hence

$$\begin{aligned}
|\nabla^m[\mu_s^{1/2}(x)]| &\leq \frac{|\delta_s|}{2}\sum_{\gamma\leq m-1}\begin{bmatrix} m-1 \\ \gamma \end{bmatrix}|\nabla^\gamma\mu_s^{1/2}(x)|\cdot\left|n!\sum_{j=0}^{n}\frac{(-1)^{n-j}}{(1-x)^{j+1}x^{n+1-j}}\right| \\
&\leq \frac{|\delta_s|}{2}\sum_{\gamma\leq m-1}\begin{bmatrix} m-1 \\ \gamma \end{bmatrix}|\nabla^\gamma\mu_s^{1/2}(x)|\cdot M_{\mathcal{C}} \\
&\leq \frac{|\delta_s|}{2}\sum_{\gamma\leq m-1}\begin{bmatrix} m-1 \\ \gamma \end{bmatrix}K_{\mathcal{C},m-1}\mu_s^{1/2}(x)\cdot M_{\mathcal{C}} \\
&= \mu_s^{1/2}(x)\left(\frac{|\delta_s|}{2}\sum_{\gamma\leq m-1}\begin{bmatrix} m-1 \\ \gamma \end{bmatrix}K_{\mathcal{C},m-1}\cdot M_{\mathcal{C}}\right).
\end{aligned}$$

Here

$$M_{\mathcal{C}} = \sup_{x\in\mathcal{C}}\left|n!\sum_{j=0}^{n}\frac{(-1)^{n-j}}{(1-x)^{j+1}x^{n+1-j}}\right|,$$

which is finite since $\mathcal{C} \subseteq (0,1)$ is compact, and hence $x$ is bounded away from zero and one. The third line follows by the induction hypothesis. □

*Proof of proposition 9.* Let $g(x) = x^{-1}(1-x)^{-1}$. Then $g(x) \to \infty$ as $x \to 0$ or $x \to 1$. Note that

$\mathrm{Bd}(\overline{\mathcal{D}}) = \{0, 1\}$. The rest of the proof is similar to that of proposition 5. It suffices to show that for any $0 \leq |\lambda| \leq m_0$,

$$\nabla^\lambda \tilde{\mu}_c^{1/2}(x) = \mu_c(x) \cdot r_\lambda(x) \tag{$*$}$$

for some rational function $r_\lambda$. Dividing $(*)$ by $\mu_s^{1/2}(x)$ yields

$$\frac{\nabla^\lambda \tilde{\mu}_c^{1/2}(x)}{\mu_s^{1/2}(x)} = \exp[(\delta_c - \delta_s)g(x)] \cdot r_\lambda(x).$$

Since $\delta_c < \delta_s$, the absolute value of this expression converges to zero as $x \to 0$ or 1. This proves part 2 of assumption 5. The proof of equation $(*)$ is as in the proof of 5: The base case holds immediately with $r_0(x) = g(x)$. The induction step follows since the derivative of a rational function is still rational. $\square$

# G Discussion of assumption 5

To get some intuition for assumption 5, consider the one dimensional case $d_x = 1$. In this case, we can usually take $m_0 = 1$, since $m_0 > d_x/2$ is then satisfied (see theorem 3 below). Then

$$\begin{aligned}
\frac{|\nabla^0 \tilde{\mu}_c^{1/2}(x)|}{\mu_s^{1/2}(x)} &= \left| \frac{\nabla^0[\mu_c^{1/2}(x) g(x)]}{\mu_s^{1/2}(x)} \right| \\
&\leq \left( \frac{\mu_c(x)}{\mu_s(x)} \right)^{1/2} |g(x)|
\end{aligned}$$

and

$$\begin{aligned}
\frac{|\nabla^1 \tilde{\mu}_c^{1/2}(x)|}{\mu_s^{1/2}(x)} &= \left| \frac{\nabla^1[\mu_c^{1/2}(x) g(x)]}{\mu_s^{1/2}(x)} \right| \\
&= \left| \frac{\nabla^1 \mu_c^{1/2}(x)}{\mu_s^{1/2}(x)} g(x) + \frac{\mu_c^{1/2}(x)}{\mu_s^{1/2}(x)} \nabla^1 g(x) \right| \\
&\leq \frac{|\nabla^1 \mu_c^{1/2}(x)|}{\mu_s^{1/2}(x)} |g(x)| + \left( \frac{\mu_c(x)}{\mu_s(x)} \right)^{1/2} |\nabla^1 g(x)|.
\end{aligned}$$

So when $d_x = 1$ with $m_0 = 1$, a sufficient condition for 5 is that there is a function $g$ that diverges to infinity in the tails, but whose levels diverge slow enough that

$$|g(x)| = o\left( \left[ \frac{\mu_c(x)}{\mu_s(x)} \right]^{-1/2} \right) \qquad \text{and} \qquad |g(x)| = o\left( \left[ \frac{|\nabla^1 \mu_c^{1/2}(x)|}{\mu_s^{1/2}(x)} \right]^{-1} \right)$$

and whose first derivative also satisfies

$$|\nabla^1 g(x)| = o\left(\left[\frac{\mu_c(x)}{\mu_s(x)}\right]^{-1/2}\right).$$

For further intuition, suppose assumption 3 held for $\mu_c$. Then for all $x \in \mathbb{R}^{d_x}$ and any $0 \leq |\lambda| \leq m_0$,

$$\begin{aligned}|\nabla^\lambda \mu_c^{1/2}(x)| &\leq K\mu_c^{1/2}(x)\\ &= K\left(\frac{\mu_c(x)}{\mu_s(x)}\right)^{1/2}\mu_s^{1/2}(x)\end{aligned}$$

and hence

$$\frac{|\nabla^\lambda \mu_c^{1/2}(x)|}{\mu_s^{1/2}(x)} \leq K\left(\frac{\mu_c(x)}{\mu_s(x)}\right)^{1/2}$$

Now suppose assumption 1 holds. Then the right hand side converges to zero as $\|x\|_e \to \infty$. Thus, in this special case, a sufficient condition for assumption 5 is that $|g(x)|$ and its derivative $|\nabla^1 g(x)|$ do not diverge faster than $\sqrt{\mu_c(x)/\mu_s(x)}$ converges to zero.

# H Closure of differentiable functions

The following lemma shows that the Sobolev sup-norm closure of a Sobolev sup-norm (with more derivatives) ball is a Hölder space with exponent 1. We assume $d_x = 1$ for notational simplicity, but the result can be extended to $d_x > 1$.

**Lemma S2.** Let $\mathcal{D}$ be a convex open subset of $\mathbb{R}$. Let $m, m_0 \geq 0$ be integers. Define

$$\Theta_D = \{f \in \mathscr{C}_{m+m_0+1}(\mathcal{D}) : \|f\|_{m+m_0+1,\infty} \leq B\}$$

and

$$\Theta_L = \{f \in \mathscr{C}_{m+m_0}(\mathcal{D}) : \|f\|_{m+m_0,\infty,\mathbb{1},1} \leq B\}.$$

Let $\bar{\Theta}_D$ be the $\|\cdot\|_{m,\infty}$-closure of $\Theta_D$. Then $\bar{\Theta}_D = \Theta_L$.

*Proof.* We prove equality by showing that $\bar{\Theta}_D \subseteq \Theta_L$ and $\Theta_L \subseteq \bar{\Theta}_D$.

1. ($\bar{\Theta}_D \subseteq \Theta_L$). Let $f \in \bar{\Theta}_D$. We will show that $f \in \Theta_L$. By the definition of the $\|\cdot\|_{m,\infty}$-closure, there exists a sequence $f_n \in \Theta_D$ such that

$$\|f_n - f\|_{m,\infty} \to 0.$$

   Since $f_n \in \Theta_D$,

$$\|f_n\|_{m+m_0+1,\infty} = \max_{0\leq|\lambda|\leq m+m_0+1} \sup_{x\in\mathcal{D}} |\nabla^\lambda f_n(x)| \leq B.$$

Also notice that for all $x, y \in \mathcal{D}$,

$$\frac{|\nabla^{m+m_0} f_n(x) - \nabla^{m+m_0} f_n(y)|}{|x-y|} \leq |\nabla^{m+m_0+1} f_n(\tilde{x})| \leq \sup_{x \in \mathcal{D}} |\nabla^{m+m_0+1} f_n(x)|$$

where $\tilde{x}$ is between $x$ and $y$, by the mean value theorem and convexity of $\mathcal{D}$. It follows that

$$\max_{|\lambda| \leq m+m_0} \sup_{x \in \mathcal{D}} |\nabla^\lambda f_n(x)| + \max_{|\lambda|=m+m_0} \sup_{x,y \in \mathcal{D}, x \neq y} \frac{|\nabla^\lambda f_n(x) - \nabla^\lambda f_n(y)|}{|x-y|} \leq \|f_n\|_{m+m_0+1,\infty} \leq B$$

and therefore $f_n \in \Theta_L$. But from part 5 of Theorem 2 we know that $\Theta_L$ is $\|\cdot\|_{m,\infty}$-closed and since $\|f_n - f\|_{m,\infty} \to 0$ it follows that $f \in \Theta_L$.

2. ($\Theta_L \subseteq \bar{\Theta}_D$) Let $f \in \Theta_L$. We will show that $f \in \bar{\Theta}_D$. Specifically, we will show how to $\|\cdot\|_{m,\infty}$-approximate $f$ by a sequence of functions $\tilde{f}_n$ in $\Theta_D$. Define

$$M_1 = \max_{|\lambda| \leq m+m_0} \sup_{x,y \in \mathcal{D}, x \neq y} \frac{|\nabla^\lambda f(x) - \nabla^\lambda f(y)|}{|x-y|} < \infty$$

and

$$M_2 = \sup_{x,y \in \mathcal{D}, x \neq y} \frac{|\nabla^{m+m_0} f(x) - \nabla^{m+m_0} f(y)|}{|x-y|} < \infty.$$

If $\mathcal{D} \neq \mathbb{R}$, then since $\nabla^{m+m_0} f$ is Lipschitz, the Kirszbraun theorem (e.g., theorem 6.1.1 on page 189 of Dudley 2002) allows us to extend $\nabla^{m+m_0} f$ to a function "$\nabla^{m+m_0} F$" on $\mathbb{R}$ with the same Lipschitz constant. Define $F$ to be the $m+m_0$ times antiderivative of $\nabla^{m+m_0} F$. Then $F$ is $(m+m_0)$-times differentiable, $\nabla^{m+m_0} F$ is Lipschitz with constant $M_2$, and $F|_{\mathcal{D}} = f$. In particular, for this extension $F$,

$$\max_{|\lambda| \leq m+m_0} \sup_{x,y \in \mathbb{R}, x \neq y} \frac{|\nabla^\lambda F(x) - \nabla^\lambda F(y)|}{|x-y|} = M_1$$

and

$$\sup_{x,y \in \mathbb{R}, x \neq y} \frac{|\nabla^{m+m_0} F(x) - \nabla^{m+m_0} F(y)|}{|x-y|} = M_2.$$

From here on we let $f(x) = F(x)$ denote the value of this extension of $f$ if $x \notin \mathcal{D}$. The main issue is that $f$ is only $(m+m_0)$-times differentiable, but we want to approximate it by functions that are just a little bit smoother—functions that are $(m+m_0+1)$-times differentiable. To do this, we convolve $f$ with a smoother function:

$$f_n(x) = [f * \psi_{\varepsilon_n}](x) = \int_{\mathbb{R}} f(x + \varepsilon_n y)\psi(y)\, dy.$$

Here $*$ denotes convolution. $\varepsilon_n$ is a sequence with $\varepsilon_n \to 0$ as $n \to \infty$. $\psi_{\varepsilon_n}$ is an approximation to the identity: a function $\psi_{\varepsilon_n}(u) = \psi(u/\varepsilon_n)/\varepsilon_n$ where $\psi : \mathbb{R} \to \mathbb{R}$ is a $(m+m_0+1)$-times continuously differentiable function such that $\psi(y) \geq 0$ for all $y \in \mathbb{R}$, $\psi(y) = 0$ if $|y| \geq 1$, and

$\int_{-1}^{1} \psi(y)\, dy = 1$. For example,

$$\psi(y) = B_k (1 - y^2)^k \mathbb{1}(|y| \leq 1).$$

where $k > m + m_0 + 1$ and $B_k$ is such that the function integrates to 1. Note that $f_n$ is $(m + m_0 + 1)$-times differentiable.

For all $\lambda \leq m + m_0$,

$$\begin{aligned}
[\nabla^\lambda f_n](x) &= [\nabla^\lambda f * \psi_{\varepsilon_n}](x) \\
&= \int_{\mathbb{R}} [\nabla^\lambda f](x - z) \frac{1}{\varepsilon_n} \psi\left(\frac{z}{\varepsilon_n}\right) dz \\
&= \int_{-1}^{1} [\nabla^\lambda f](x - \varepsilon_n y) \psi(y)\, dy.
\end{aligned}$$

The last line follows by a change of variables and since $\psi$ is zero outside $[-1, 1]$. Hence

$$\begin{aligned}
|\nabla^\lambda f_n(x) - \nabla^\lambda f(x)| &\leq \int_{-1}^{1} |\nabla^\lambda f(x - \varepsilon_n y) - \nabla^\lambda f(x)| \psi(y)\, dy \\
&\leq \int_{-1}^{1} |M_1 \varepsilon_n y| \psi(y)\, dy \\
&= \varepsilon_n M_1 \int_{-1}^{1} |y| \psi(y)\, dy \\
&\equiv \delta_n
\end{aligned}$$

for all $\lambda \leq m + m_0$. The first line follows since $\psi$ integrates to 1. Since $\delta_n \to 0$, it follows that

$$\|f_n - f\|_{m+m_0,\infty} \to 0.$$

Moreover,

$$\begin{aligned}
|\nabla^{m+m_0} f_n(x_1) - \nabla^{m+m_0} f_n(x_2)| &\leq \int |\nabla^{m+m_0} f(x_1 - \varepsilon_n y) - \nabla^{m+m_0} f(x_2 - \varepsilon_n y)| \psi(y)\, dy \\
&\leq M_2 |x_1 - x_2|.
\end{aligned}$$

Since $f_n$ is $(m + m_0 + 1)$-times continuously differentiable,

$$|\nabla^{m+m_0+1} f_n(x)| = \lim_{h \to 0} \frac{|\nabla^{m+m_0} f_n(x + h) - \nabla^{m+m_0} f_n(x)|}{h} \leq M_2$$

for each $x \in \mathbb{R}$. Recall that

$$M_2 = \sup_{x, y \in \mathcal{D}, x \neq y} \frac{|\nabla^{m+m_0} f(x) - \nabla^{m+m_0} f(y)|}{|x - y|}.$$

This implies that

$$
\begin{aligned}
\|f_n\|_{m+m_0+1,\infty} &\le \|f_n\|_{m+m_0,\infty} + \sup_{x\in\mathcal{D}} |\nabla^{m+m_0+1} f_n(x)| \\
&\le \|f\|_{m+m_0,\infty} + \|f_n - f\|_{m+m_0,\infty} + \sup_{x\in\mathcal{D}} |\nabla^{m+m_0+1} f_n(x)| \\
&\le \|f\|_{m+m_0,\infty} + \delta_n + \sup_{x\in\mathcal{D}} |\nabla^{m+m_0+1} f_n(x)| \\
&\le \left( \|f\|_{m+m_0,\infty} + \sup_{x,y\in\mathcal{D}, x\neq y} \frac{|\nabla^{m+m_0} f(x) - \nabla^{m+m_0} f(y)|}{|x-y|} \right) + \delta_n \\
&\le B + \delta_n.
\end{aligned}
$$

The last line follows since $f \in \Theta_L$. Thus $f_n$ is *almost* in $\Theta_D$, but not quite. But we can just rescale $f_n$ to put it inside $\Theta_D$: Let

$$
\tilde{f}_n(x) = \frac{B}{B+\delta_n} f_n(x).
$$

Then $\|\tilde{f}_n\|_{m+m_0+1,\infty} \le B$ and so $\tilde{f}_n \in \Theta_D$. Moreover,

$$
\begin{aligned}
\|\tilde{f}_n - f\|_{m,\infty} &\le \|\tilde{f}_n - f\|_{m+m_0,\infty} \\
&\le \|\tilde{f}_n - f_n\|_{m+m_0,\infty} + \|f_n - f\|_{m+m_0,\infty} \\
&= \max_{0\le|\lambda|\le m+m_0} \sup_{x\in\mathcal{D}} \left| \nabla^\lambda \left( \frac{B}{B+\delta_n} f_n(x) \right) - \nabla^\lambda f_n(x) \right| + \|f_n - f\|_{m+m_0,\infty} \\
&= \left| \frac{B}{B+\delta_n} - 1 \right| \|f_n\|_{m+m_0,\infty} + \|f_n - f\|_{m+m_0,\infty} \\
&= \frac{\delta_n}{B+\delta_n} \|f_n\|_{m+m_0,\infty} + \|f_n - f\|_{m+m_0,\infty}.
\end{aligned}
$$

Since $\|f_n\|_{m+m_0,\infty} \le \|f_n\|_{m+m_0+1,\infty} \le B + \delta_n$,

$$
\frac{\delta_n}{B+\delta_n} \|f_n\|_{m+m_0,\infty} \to 0.
$$

We also know that $\|f_n - f\|_{m+m_0,\infty} \to 0$. It follows that

$$
\|\tilde{f}_n - f\|_{m,\infty} \to 0.
$$

But remember that $\tilde{f}_n \in \Theta_D$. So, by definition of the $\|\cdot\|_{m,\infty}$-closure, $f \in \bar{\Theta}_D$.

$\square$

# I Sup-norm convergence over closed domains $\mathcal{D}$

Throughout the paper we have focused on functions with open domains $\mathcal{D}$. In practice we may also be interested in functions with closed domains $\mathcal{D}$. First, note that convergence of a sequence of functions in a Sobolev $L_p$ norm where the integral is taken over the interior of $\mathcal{D}$ implies convergence in the Sobolev $L_p$ norm where the integral is taken over the entire $\mathcal{D}$. This follows since $\mathcal{D}$ is a subset of $\mathbb{R}^{d_x}$ and hence its boundary has measure zero. So the value of the integral is not affected by its values on the boundary. For Sobolev sup-norms, however, convergence over the interior of $\mathcal{D}$ does not automatically imply convergence over all of $\mathcal{D}$. In the following lemma, we illustrate how to do this extension for sequences from a Hölder ball which are known to converge in the ordinary sup-norm over the interior. Similar results can be obtained with different parameter spaces and for convergence in general Sobolev sup-norms.

**Lemma S3.** Let $\mathcal{D} \subseteq \mathbb{R}^{d_x}$ be closed and convex. Let $f_n : \mathcal{D} \to \mathbb{R}$ be a sequence of functions in

$$\Theta = \{f \in \mathscr{C}_0(\mathcal{D}) : \|f\|_{0,\infty,\mathbb{1},\nu} \leq B\}.$$

Suppose

$$\sup_{x \in \text{int}\mathcal{D}} |f_n(x) - f(x)| \to 0.$$

for some function $f$. Suppose $f$ is continuous at each boundary point in $\mathcal{D}$. Then

$$\sup_{x \in \mathcal{D}} |f_n(x) - f(x)| \to 0.$$

*Proof of lemma S3.* We want to show that for any $\varepsilon > 0$, there is an $N$ such that

$$|f_n(x) - f(x)| \leq \varepsilon$$

for all $n \geq N$, for all $x \in \mathcal{D}$. For each $x \in \mathcal{D}$, choose an element $z_x \in \text{int}\mathcal{D}$ such that $\|x - z_x\|_e^\nu \leq \varepsilon/(3B)$ and

$$|f(x) - f(z_x)| \leq \frac{\varepsilon}{3}.$$

This is possible since $f$ is continuous on all of $\mathcal{D}$, including at boundary points, and by convexity of $\mathcal{D}$. By the triangle inequality,

$$\begin{aligned} |f_n(x) - f(x)| &= |f_n(x) - f(x) - f_n(z_x) + f_n(z_x) - f(z_x) + f(z)| \\ &\leq |f_n(x) - f_n(z_x)| + |f(x) - f(z_x)| + |f_n(z_x) - f(z_x)|. \end{aligned}$$

By the definition of this parameter space we have

$$\sup_{x \in \mathcal{D}} |f_n(x) - f_n(z_x)| \leq B\|x - z_x\|_e^\nu \leq \frac{\varepsilon}{3}.$$

By uniform convergence of $f_n$ to $f$ on the interior of $\mathcal{D}$, there is an $N$ such that

$$|f_n(z_x) - f(z_x)| \leq \frac{\varepsilon}{3}$$

for all $n \geq N$. Thus we're done. $\square$

# J Proofs for section 6

*Proof of proposition 10.* We omit this proof because it is almost identical to the proof of lemma A1 in Newey and Powell (2003). $\square$

*Proof of proposition 11.* We verify the conditions of proposition 10.

1. The parameter space is $\|\cdot\|_c$-compact by part 1 of theorems 3 and 4. Since the sieve space is a $\|\cdot\|_c$-closed subset of the $\|\cdot\|_c$-compact set $\Theta$, it is also $\|\cdot\|_c$-compact.

2. Define $Q(g) = -\mathbb{E}((Y - g(X))^2)$. Then for $g_1, g_2 \in \Theta$,

$$\begin{aligned}
&|Q(g_1) - Q(g_2)| \\
&= \left|\mathbb{E}(g_2(X)^2 - g_1(X)^2) + \mathbb{E}(2Y(g_1(X) - g_2(X)))\right| \\
&\leq \left|\mathbb{E}(g_2(X)^2 - g_1(X)^2)\right| + |\mathbb{E}(2Y(g_1(X) - g_2(X)))| \\
&= |\mathbb{E}(g_2(X) - g_1(X))(g_2(X) + g_1(X))| + 2\,|\mathbb{E}(Y(g_1(X) - g_2(X)))| \\
&\leq \sqrt{\mathbb{E}\left((g_2(X) - g_1(X))^2\right)\mathbb{E}\left((g_2(X) + g_1(X))^2\right)} + 2\sqrt{\mathbb{E}\left(Y^2\right)\mathbb{E}\left((g_1(X) - g_2(X))^2\right)} \\
&\leq \sqrt{\mathbb{E}\left((g_2(X) - g_1(X))^2\right)\mathbb{E}\left(2g_2(X)^2 + 2g_1(X)^2\right)} + 2\sqrt{\mathbb{E}\left(Y^2\right)\mathbb{E}\left((g_1(X) - g_2(X))^2\right)}.
\end{aligned}$$

   The fourth line follows from the Cauchy-Schwarz inequality and the last line from $(a+b)^2 \leq 2a^2 + 2b^2$ for any $a, b, \in \mathbb{R}$. Next,

$$\mathbb{E}((g_1(X) - g_2(X))^2) \leq \left(\sup_{x \in \mathbb{R}} |g_1(x) - g_2(x)|\mu_c(x)\right)^2 \mathbb{E}(\mu_c(X)^{-2}) = \|g_1 - g_2\|_c^2 \cdot \mathbb{E}(\mu_c(X)^{-2}).$$

   Moreover, for all $g \in \Theta$,

$$\begin{aligned}
\mathbb{E}(g(X)^2) &= \mathbb{E}(g(X)^2\mu_c(X)^2\mu_c(X)^{-2}) \\
&\leq \|g\|_c^2 \cdot \mathbb{E}(\mu_c(X)^{-2}) \\
&\leq C^2\|g\|_s^2 \cdot \mathbb{E}(\mu_c(X)^{-2}) \\
&\leq C^2B^2\mathbb{E}(\mu_c(X)^{-2}).
\end{aligned}$$

   The third line follows by the compact embedding result, part 1 of theorem 3. Therefore

$$|Q(g_1) - Q(g_2)| \leq 2\left(BC\mathbb{E}(\mu_c(X)^{-2}) + \sqrt{\mathbb{E}\left(Y^2\right)\mathbb{E}\left(\mu_c(X)^{-2}\right)}\right)\|g_1 - g_2\|_c.$$

Since $\mathbb{E}(Y^2) < \infty$ and $\mathbb{E}(\mu_c(X)^{-2}) < \infty$, $Q$ is $\|\cdot\|_c$-continuous. Similarly, let $\widehat{Q}_n(g) = -\frac{1}{n}\sum_{i=1}^n (Y_i - g(X_i))^2$. Identical arguments imply that

$$|\widehat{Q}_n(g_1) - \widehat{Q}_n(g_2)| \leq 2\left(BC\frac{1}{n}\sum_{i=1}^n \mu_c(X_i)^{-2} + \sqrt{\left(\frac{1}{n}\sum_{i=1}^n Y_i^2\right)\left(\frac{1}{n}\sum_{i=1}^n \mu_c(X_i)^{-2}\right)}\right)\|g_1 - g_2\|_c.$$

Hence $\widehat{Q}$ is $\|\cdot\|_c$-continuous.

3. Suppose $Q(g) = Q(g_0)$. Then $\mathbb{E}((Y - g(X))^2) = \mathbb{E}((Y - g_0(X))^2)$, which implies that $g(X) = g_0(X)$ almost everywhere. If $g(\bar{x}) \neq g_0(\bar{x})$ for some $\bar{x}$, then $g(\bar{x}) \neq g_0(\bar{x})$ in a neighborhood of $\bar{x}$ by continuity of $g_0$, a contradiction. Hence $g(x) = g_0(x)$ for all $x \in \mathbb{R}$. Thus $\|g - g_0\|_c = \sup_{x\in\mathbb{R}} |g(x) - g_0(x)|\mu_c(x) = 0$. Moreover,

$$Q(g_0) = -\mathbb{E}((Y - g_0(X))^2) > -\mathbb{E}(2Y^2 + 2g_0(X)^2) > -\infty.$$

4. For any $g_k \in \Theta_k$

$$\|g_k - g_0\|_c \leq \sup_{|x|\leq M} |g_k(x) - g(x)| \sup_{|x|\leq M} \mu_c(x) + \sup_{|x|\geq M} |(g_k(x) - g(x))\mu_s(x)| \sup_{|x|\geq M} \frac{\mu_c(x)}{\mu_s(x)}.$$

Let $\varepsilon > 0$. Since $g_k$ and $g_0$ are in $\Theta$,

$$\sup_{|x|\geq M} |(g_k(x) - g(x))\mu_s(x)| \leq \|g_k - g\|_s \leq 2B.$$

Thus, since $\mu_c$ and $\mu_s$ satisfy assumption 1, we can choose $M$ such that

$$\sup_{|x|\geq M} |(g_k(x) - g(x))\mu_s(x)| \sup_{|x|\geq M} \frac{\mu_c(x)}{\mu_s(x)} \leq \frac{\varepsilon}{2}.$$

By assumption, for a fixed $M$, we can pick $k$ large enough to make $\sup_{|x|\leq M} |g_k(x) - g(x)|$ arbitrarily small. By $\mu_c^2$ satisfying the integrability assumption $6'$ and continuity of $\mu_c$, $\sup_{|x|\leq M} \mu_c(x) < \infty$. Hence we can pick $k$ large enough so that

$$\sup_{|x|\leq M} |g_k(x) - g(x)| \sup_{|x|\leq M} \mu_c(x) \leq \frac{\varepsilon}{2}.$$

Thus $\|g_k - g_0\|_c \leq \varepsilon$. Hence we have shown that $\|g_k - g_0\|_c \to 0$ as $k \to 0$.

5. For all $g \in \Theta_{k_n} \subseteq \Theta$,

$$(Y - g(X))^2 \leq 2Y^2 + g(X)^2 \leq 2Y^2 + 2B^2C^2\mu_c(X)^{-2}.$$

Since $\mathbb{E}(Y^2) < \infty$ and $\mathbb{E}(\mu_c(X)^{-2}) < \infty$ we have

$$\mathbb{E}\left(\sup_{g\in\Theta}(Y - g(X))^2\right) < \infty.$$

Hence Jennrich's uniform law of large numbers implies that

$$\sup_{g\in\Theta_{k_n}} |\widehat{Q}_n(g) - Q(g)| \xrightarrow{p} 0.$$

□

*Proof of proposition 12.* The proof is similar to the one of proposition 11 and verifies the conditions of proposition 10.

1. This step is identical to the corresponding step in the proof of proposition 11.

2. Define $Q(g) = -\mathbb{E}((Y - g(X))^2\mu_c(X)^2)$. Then for $g_1, g_2 \in \Theta$,

$$\begin{aligned}|Q(g_1) - Q(g_2)| &= \left|\mathbb{E}\left((g_2(X)^2 - g_1(X)^2)\mu_c(X)^2\right) + \mathbb{E}\left(2Y(g_1(X) - g_2(X))\mu_c(X)^2\right)\right| \\ &\le \sqrt{\mathbb{E}\left((g_2(X) - g_1(X))^2\mu_c(X)^2\right)\mathbb{E}\left((g_2(X) + g_1(X))^2\mu_c(X)^2\right)} \\ &\quad + 2\sqrt{\mathbb{E}\left(Y^2\mu_c(X)^2\right)\mathbb{E}\left((g_1(X) - g_2(X))^2\mu_c(X)^2\right)} \\ &\le \sqrt{\mathbb{E}\left((g_2(X) - g_1(X))^2\mu_c(X)^2\right)\mathbb{E}\left(2g_2(X)^2\mu_c(X)^2 + 2g_1(X)^2\mu_c(X)^2\right)} \\ &\quad + 2\sqrt{\mathbb{E}\left(Y^2\mu_c(X)^2\right)\mathbb{E}\left((g_1(X) - g_2(X))^2\mu_c(X)^2\right)}.\end{aligned}$$

   Next,

$$\mathbb{E}\left((g_1(X) - g_2(X))^2\mu_c(X)^2\right) \le \|g_1 - g_2\|_c^2.$$

   Moreover, for all $g \in \Theta$,

$$\mathbb{E}\left(g(X)^2\mu_c(X)^2\right) \le B^2M_5^2.$$

   Therefore

$$|Q(g_1) - Q(g_2)| \le 2\left(BM_5 + \sqrt{\mathbb{E}\left(Y^2\mu_c(X)^2\right)}\right)\|g_1 - g_2\|_c.$$

   Since $\mathbb{E}\left(Y^2\mu_c(X)^2\right) < \infty$, $Q$ is continuous. Similarly, let $\widehat{Q}_n(g) = -\frac{1}{n}\sum_{i=1}^n(Y_i - g(X_i))^2\mu_c(X_i)^2$. Identical arguments imply that

$$|\widehat{Q}_n(g_1) - \widehat{Q}_n(g_2)| \le 2\left(BM_5 + \sqrt{\frac{1}{n}\sum_{i=1}^n Y_i^2\mu_c(X_i)^2}\right)\|g_1 - g_2\|_c.$$

   Hence $\widehat{Q}$ is continuous.

3. As before, $\mathbb{E}((Y - g(X))^2\mu_c(X)^2) = \mathbb{E}((Y - g_0(X))^2\mu_c(X)^2)$ implies $g(X)\mu_c(X) = g_0(X)\mu_c(X)$ almost everywhere. If $g(\bar{x}) \ne g_0(\bar{x})$ for some $\bar{x}$, then $g(\bar{x}) \ne g_0(\bar{x})$ in a neighborhood of $\bar{x}$

by continuity of $g_0$. Moreover if $\mu_c(\bar{x}) > 0$, then $\mu_c(x) > 0$ with positive probability in a neighborhood of $\bar{x}$, which contradicts that $g(X)\mu_c(X) = g_0(X)\mu_c(X)$ almost everywhere. Thus, $g(\bar{x}) \neq g_0(\bar{x})$ implies $\mu_c(\bar{x}) = 0$. Therefore $\|g - g_0\|_c = 0$. Moreover,

$$Q(g_0) = -\mathbb{E}((Y - g_0(X))^2 \mu_c(X)^2) > -\mathbb{E}(2Y^2\mu_c(X)^2 + 2g_0(X)^2\mu_c(X)^2) > -\infty.$$

4. This step is identical to the corresponding step in the proof of proposition 11.

5. For all $g \in \Theta_{k_n} \subseteq \Theta$,

$$(Y - g(X))^2 \mu_c(X)^2 \leq 2Y^2\mu_c(X)^2 + 2g(X)^2\mu_c(X)^2 \leq 2Y^2\mu_c(X)^2 + 2B^2M_5^2.$$

This combined with $\mathbb{E}(Y^2\mu_c(X)^2) < \infty$ let us apply Jennrich's uniform law of large numbers, which gives

$$\sup_{\theta \in \Theta_{k_n}} |\widehat{Q}_n(\theta) - Q(\theta)| \xrightarrow{p} 0.$$

□

*Proof of proposition 13.* Let $g_{k_n} \in \widetilde{\Theta}_{k_n}$ such that $\|g_{k_n} - g_0\|_c \to 0$. Then $\|g_{k_n}\|_c \leq \|g_0\|_c + 1$ for $n$ large enough. Moreover, $\|g_0\|_c \leq C\|g_0\|_s < \infty$. From the proof of proposition 12 we know that

$$|Q(g_{k_n}) - Q(g_0)| \leq 2\left(M_5(\|g_0\|_c + 1) + \sqrt{\mathbb{E}\left(Y^2\mu_c(X)^2\right)}\right)\|g_{k_n} - g_0\|_c$$

and

$$|\widehat{Q}_n(g_{k_n}) - \widehat{Q}_n(g_0)| \leq 2\left(M_5(\|g_0\|_c + 1) + \sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2\mu_c(X_i)^2}\right)\|g_{k_n} - g_0\|_c.$$

Now write

$$\widehat{Q}_n(g_{k_n}) - Q(g_{k_n}) = \left(\widehat{Q}_n(g_{k_n}) - \widehat{Q}_n(g_0)\right) + \left(\widehat{Q}_n(g_0) - Q(g_0)\right) + \left(Q(g_0) - Q(g_{k_n})\right).$$

$\widehat{Q}_n(g_0) - Q(g_0) = O_p(1/\sqrt{n})$ by the central limit theorem, which applies since $\mathbb{E}((Y - g_0(X))^4) < \infty$ and $\mu_c$ is uniformly bounded above. Thus,

$$\widehat{Q}_n(g_{k_n}) - Q(g_{k_n}) = O_p(\|g_{k_n} - g_0\|_c + 1/\sqrt{n}).$$

Since $\max\{1/\sqrt{n}, \|g_{k_n} - g_0\|_c\} = O(\lambda_n)$, lemma A.3 in Chen and Pouzo (2012) implies that for some $M_0 > 0$ it holds that $\|g_0\|_s \leq M_0$ and

$$\tilde{g}_w \in \{g \in \mathscr{W}_{1,2,\mu_s} : \|g\|_{1,2,\mu_s} \leq M_0\}$$

with probability arbitrarily close to 1 for all large $n$. Hence it suffices to prove that $\|\bar{g}_w - g_0\|_c \xrightarrow{p} 0$, where

$$\bar{g}_w(x) = \underset{g \in \tilde{\Theta}_{k_n}^{M_0}}{\operatorname{argmax}} \; -\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))^2\mu_c(X_i)^2 + \lambda_n\|g\|_s\right)$$

and $\tilde{\Theta}_{k_n}^{M_0} = \{g \in \widetilde{\Theta}_{k_n} : \|g\|_s \leq M_0\}$.

Consistency now follows from proposition 12 under two additional arguments:

1. First, $\sup_{g\in\tilde{\Theta}_{k_n}^{M_0}} \lambda_n\|g\|_s \leq \lambda_n M_0 \to 0$ and therefore the sample objective function (including the penalty) still converges to $Q$ uniformly over $g \in \tilde{\Theta}_{k_n}^{M_0}$.
2. Second, since $\tilde{\Theta}_{k_n}^{M_0}$ is finite dimensional, for any $g_1, g_2 \in \tilde{\Theta}_{k_n}^{M_0}$ there exists $D > 0$ such that $|\|g_1\|_s - \|g_2\|_s| \leq D|\|g_1\|_c - \|g_2\|_c| \leq D\|g_1 - g_2\|_c$. Hence the sample objective function (including the penalty) is still continuous on $\tilde{\Theta}_{k_n}^{M_0}$.

All other assumptions of proposition 10 hold using the same arguments as those in the proof of proposition 12. Thus $\|\bar{g}_w - g_0\|_c \xrightarrow{p} 0$ and hence $\|\tilde{g}_w - g_0\|_c \xrightarrow{p} 0$. □

*Proof of proposition 14.* The proof is adapted from the proof of theorem 4.3 in Newey and Powell (2003). Again we verify the conditions of proposition 10.

1. This step is identical to the corresponding step in the proof of proposition 11.

2a. Define $Q(g) = -\mathbb{E}(\mathbb{E}(Y - g(X) \mid Z)^2)$. For $g_1, g_2 \in \Theta$,

$$\begin{aligned}
&|\mathbb{E}(Y - g_1(X) \mid Z)^2 - \mathbb{E}(Y - g_2(X) \mid Z)^2| \\
&= |\mathbb{E}(2Y \mid Z)\mathbb{E}(g_2(X) - g_1(X) \mid Z) + \mathbb{E}(g_2(X) - g_1(X) \mid Z)\mathbb{E}(g_2(X) + g_1(X) \mid Z)| \\
&\leq |\mathbb{E}(2Y + g_2(X) + g_1(X) \mid Z)| \cdot |\mathbb{E}(g_2(X) - g_1(X) \mid Z)| \\
&= |\mathbb{E}((2g_0(X) + g_2(X) + g_1(X))\mu_c(X)\mu_c(X)^{-1} \mid Z)| \cdot |\mathbb{E}((g_2(X) - g_1(X))\mu_c(X)\mu_c(X)^{-1} \mid Z)| \\
&\leq 4BM_5|\mathbb{E}(\mu_c(X)^{-1} \mid Z)| \cdot M_5\|g_1 - g_2\|_c \cdot |\mathbb{E}(\mu_c(X)^{-1} \mid Z)| \\
&= 4BM_5^2\mathbb{E}(\mu_c(X)^{-1} \mid Z)^2\|g_1 - g_2\|_c \\
&\leq 4BM_5^2\mathbb{E}(\mu_c(X)^{-2} \mid Z)\|g_1 - g_2\|_c.
\end{aligned}$$

The fourth line uses $\mathbb{E}(U \mid Z) = 0$ and the last uses Jensen's inequality. Therefore

$$\begin{aligned}
|Q(g_1) - Q(g_2)| &\leq \mathbb{E}\left(|\mathbb{E}(Y - g_1(X) \mid Z)^2 - \mathbb{E}(Y - g_2(X) \mid Z)^2|\right) \\
&\leq 4BM_5^2\mathbb{E}(\mu_c(X)^{-2})\|g_1 - g_2\|_c.
\end{aligned}$$

Hence, $Q$ is continuous.

2b. Let

$$\Theta_{k_n} = \left\{ g \in \Theta : g = \sum_{j=1}^{k_n} b_j p_j(x) \text{ for some } b_1, \ldots, b_{k_n} \in \mathbb{R} \right\}.$$

Define $P_Z$ as the $n \times k_n$ matrix with $(i,j)$th element $p_j(X_i)$. Let $Q_Z = P_Z(P_Z' P_Z)^- P_Z'$ where $(P_Z'P_Z)^-$ denotes the Moore-Penrose generalized inverse of $(P_Z'P_Z)$. Let $Y$ and $g(X)$ be the $n \times 1$ vectors with elements $Y_i$ and $g(X_i)$, respectively. Define $\widehat{Q}_n(g) = -\frac{1}{n}\|Q_Z(Y - g(X))\|^2$. Then for $g_1, g_2 \in \Theta$,

$$
\begin{aligned}
&|\widehat{Q}_n(g_1) - \widehat{Q}_n(g_2)| \\
&= \left| \frac{1}{n}\|Q_Z(Y - g_1(X))\|^2 - \frac{1}{n}\|Q_Z(Y - g_2(X))\|^2 \right| \\
&\leq \frac{1}{n}\|Q_Z(g_1(X) - g_2(X))\| \cdot \|Q_Z(2Y - g_1(X) - g_2(X))\| \\
&\leq \frac{1}{n}\|g_1(X) - g_2(X)\| \cdot \|2Y - g_1(X) - g_2(X)\| \\
&= \sqrt{\frac{1}{n}\sum_{i=1}^n (g_1(X_i) - g_2(X_i))^2 \mu_c(X_i)^2 \mu_c(X_i)^{-2}} \sqrt{\frac{1}{n}\sum_{i=1}^n (2Y_i - g_1(X_i) - g_2(X_i))^2} \\
&\leq \left( \sqrt{\frac{1}{n}\sum_{i=1}^n \mu_c(X_i)^{-2}} \sqrt{\frac{1}{n}\sum_{i=1}^n 4Y_i^2 + 4B^2 M_5^2 \mu_c(X_i)^{-2}} \right) \|g_1 - g_2\|_c.
\end{aligned}
$$

The second line follows because, by the Cauchy-Schwarz inequality,

$$
|(a'a) - (b'b)| = |(a-b)'(a+b)| \leq \sqrt{(a-b)'(a-b)}\sqrt{(a+b)'(a+b)}
$$

for all $a, b \in \mathbb{R}^n$. The third line follows because $Q_Z$ is idempotent and thus $\|Q_Z b\| \leq \|b\|$ for all $b \in \mathbb{R}^n$. Hence $\widehat{Q}_n$ is continuous.

3. By completeness, $Q(g) = -\mathbb{E}(\mathbb{E}(Y - g(X) \mid Z)^2) = 0$ implies that $g(x) = g_0(x)$ almost everywhere. Identical arguments as those in the proof of proposition 11 then imply that $\|g - g_0\|_c = 0$, by continuity of $g_0$. Moreover,

$$
Q(g_0) = -\mathbb{E}(\mathbb{E}(U \mid Z)^2) = 0 > -\infty.
$$

4. Assumption 4 of proposition 10 holds using identical arguments as those in the proof of proposition 11.

5. Assumption 5 of proposition 10 requires convergence of $\widehat{Q}_n$ to $Q$ uniformly over the sieve spaces. We show this by applying corollary 2.2 in Newey (1991). $\Theta$ is $\|\cdot\|_c$-compact, which is Newey's assumption 1. $Q$ is $\|\cdot\|_c$-continuous, which is Newey's equicontinuity assumption. Next, define

$$
B_n = \left( \sqrt{\frac{1}{n}\sum_{i=1}^n \mu_c(X_i)^{-2}} \sqrt{\frac{1}{n}\sum_{i=1}^n 4Y_i^2 + 4B^2 M_5^2 \mu_c(X_i)^{-2}} \right)
$$

and recall that

$$
|\widehat{Q}_n(g_1) - \widehat{Q}_n(g_2)| \leq B_n \|g_1 - g_2\|_c.
$$

By Kolmogorov's strong law of large numbers and the existence of the relevant moments, $B_n = O_p(1)$. Hence Newey's assumption 3A holds. All that remains is to show Newey's assumption 2, pointwise convergence: $|\widehat{Q}(g) - Q(g)| = o_p(1)$ for all $g \in \Theta$. First write

$$\begin{aligned}|\widehat{Q}(g) - Q(g)| &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y - g(X) \mid Z = Z_i)^2 - \mathbb{E}\left(\mathbb{E}(Y - g(X) \mid Z)^2\right) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\mathbb{E}}(Y - g(X) \mid Z = Z_i)^2 - \mathbb{E}(Y - g(X) \mid Z = Z_i)^2\right),\end{aligned}$$

where $\widehat{\mathbb{E}}(Y - g(X) \mid Z = Z_i)$ is the series estimator of the conditional expectation evaluated at $Z_i$. For the first part notice that $\mathbb{E}(Y - g(X) \mid Z = Z_i)^2$ is iid and

$$\begin{aligned}\mathbb{E}\left(\mathbb{E}(Y - g(X) \mid Z)^2\right) &\le \mathbb{E}\left(\mathbb{E}((Y - g(X))^2 \mid Z)\right) \\ &\le \mathbb{E}(2Y^2 + 2g(X)^2) \\ &\le 2\mathbb{E}(Y^2) + 2\mathbb{E}(\mu_c(X)^{-1})\|g\|_c^2 \\ &< \infty.\end{aligned}$$

It follows from Kolmogorov's strong law of large numbers that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y - g(X) \mid Z = Z_i)^2 - \mathbb{E}\left(\mathbb{E}(Y - g(X) \mid Z)^2\right) \xrightarrow{p} 0.$$

Next, following Newey (1991), define $\rho$ as the $n \times 1$ vector containing $Y_i - g(X_i)$ and $h$ as the $n \times 1$ vector containing $\mathbb{E}(Y - g(X) \mid Z = Z_i)$. Then

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{E}(Y - g(X) \mid Z = Z_i)^2 - \mathbb{E}(Y - g(X) \mid Z = Z_i)^2\right)\right| = \left|\|Q_Z\rho\|^2 - \|h\|^2\right|/n.$$

Since for all $a, b \in \mathbb{R}^n$ it holds that $a'a - b'b = (a-b)'(a-b) + 2b'(a-b)$,

$$\left|\|Q_Z\rho\|^2 - \|h\|^2\right|/n \le \left(\|Q_Z\rho - h\|^2 + 2\|h\| \cdot \|Q_Z\rho - h\|\right)/n.$$

Since

$$\|h\|^2/n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y - g(X) \mid Z = Z_i)^2,$$

the previous arguments imply that $\|h\|^2/n = O_p(1)$. It therefore suffices to prove that $\|Q_Z\rho - h\|^2/n = o_p(1)$, which by Markov's inequality is implied by

$$\mathbb{E}\left(\|Q_Z\rho - h\|^2\right)/n \to 0.$$

as $n \to 0$. Newey (1991) shows

$$\mathbb{E}\left(\|Q_Z\rho - h\|^2\right)/n \le \mathbb{E}\left(\mathrm{trace}(Q_Z \,\mathrm{var}(h \mid Z))\right)/n + o(1).$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left(\|Q_Z\rho - h\|^2\right)/n &\le \mathbb{E}\left(\sum_{i=1}^n (Q_Z)_{ii}\,\mathrm{var}(Y_i - g(X_i) \mid Z_i)\right)/n + o(1) \\
&\le \mathbb{E}\left(\sqrt{\frac{1}{n}\sum_{i=1}^n (Q_Z)_{ii}^2 \frac{1}{n}\sum_{i=1}^n \mathrm{var}(Y_i - g(X_i) \mid Z_i)^2}\right) + o(1) \\
&\le \mathbb{E}\left(\sqrt{\frac{1}{n}\mathrm{trace}(Q_Z' Q_Z)\frac{1}{n}\sum_{i=1}^n \mathrm{var}(Y_i - g(X_i) \mid Z_i)^2}\right) + o(1) \\
&= \mathbb{E}\left(\sqrt{\frac{1}{n}\mathrm{trace}(Q_Z)\frac{1}{n}\sum_{i=1}^n \mathrm{var}(Y_i - g(X_i) \mid Z_i)^2}\right) + o(1) \\
&\le \sqrt{\frac{k_n}{n}}\mathbb{E}\left(\sqrt{\frac{1}{n}\sum_{i=1}^n \mathrm{var}(Y_i - g(X_i) \mid Z_i)^2}\right) + o(1) \\
&\le \sqrt{\frac{k_n}{n}}\sqrt{\mathbb{E}\left(\mathrm{var}(Y_i - g(X_i) \mid Z_i)^2\right)} + o(1).
\end{aligned}$$

The second line follows from the Cauchy-Schwarz inequality. The third line from the definition of the trace. The fourth line because $Q_Z$ is idempotent. The fifth line because $\mathrm{trace}(Q_Z) \le k_n$. The last line by Jensen's inequality. Since $\mathbb{E}\left((\mathrm{var}(Y_i - g(X_i) \mid Z_i))^2\right) < \infty$ and $k_n/n \to 0$, it follows that

$$\mathbb{E}\left(\|Q_Z\rho - h\|^2\right)/n \to 0.$$

□

# References


Adams, R. A. and J. J. Fournier (2003): *Sobolev spaces*, vol. 140, Academic press, 2nd ed.

Chen, X. and D. Pouzo (2012): "Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals," *Econometrica*, 80, 277–321.

Dudley, R. M. (2002): *Real analysis and probability*, vol. 74, Cambridge University Press.

Gallant, A. and D. Nychka (1987): "Semi-nonparametric maximum likelihood estimation," *Econometrica*, 55, 363–390.

Kufner, A. (1980): *Weighted Sobolev spaces*, BSB B. G. Teubner Verlagsgesellschaft.

Newey, W. K. (1991): "Uniform convergence in probability and stochastic equicontinuity," *Econometrica*, 1161–1167.

Newey, W. K. and J. L. Powell (2003): "Instrumental variable estimation of nonparametric models," *Econometrica*, 71, 1565–1578.

Rodríguez, J. M., V. Álvarez, E. Romera, and D. Pestana (2004): "Generalized weighted Sobolev spaces and applications to Sobolev orthogonal polynomials I," *Acta Applicandae Mathematicae*, 80, 273–308.

Santos, A. (2012): "Inference in nonparametric instrumental variables with partial identification," *Econometrica*, 80, 213–275.