1. The authors are grateful to Weigutian Ou for pointing out an issue affecting Lemma II. 6 and Lemma A. 8. Specifically, in Lemma II. 6 and Lemma A. 8, $\mathcal{B}(\Psi) = \max_i\{|a_i|\mathcal{B}(\Phi_i)\}$ should be replaced by

$$\mathcal{B}(\Psi) \leq \max\left\{\sum_{i=1}^{n} |a_i|\mathcal{B}(\Phi_i), \max_{i\in\{1,\dots,n\}} \mathcal{B}(\Phi_i), \max_{i\in\{1,\dots,n\}} |a_i|\mathcal{B}(\Phi_i)\right\}. \quad (1)$$

In Lemma II. 6 this bound is obtained as follows. Given networks $\Phi_i \in \mathcal{N}_{d_i,d'}$ with associated matrix-vector pairs (A^i_ℓ, b^i_ℓ), $i \in \{1,\dots,n\}$, $\ell \in \{1,\dots,L\}$, the first $L-1$ matrix-vector pairs of the network $\Psi \in \mathcal{N}_{\sum_{i=1}^n d_i,d'}$ are obtained as

$$A_\ell = \operatorname{diag}(A^1_\ell, A^2_\ell, \dots, A^n_\ell), \quad b_\ell = (b^1_\ell, b^2_\ell, \dots, b^n_\ell).$$

Hence, we have, for all $\ell \in \{1,\dots,L-1\}$, that $\|A_\ell\|_\infty \leq \max_i \mathcal{B}(\Phi_i)$ and $\|b_\ell\|_\infty \leq \max_i \mathcal{B}(\Phi_i)$. In the last layer, i.e., for $\ell = L$, we have

$$A_L = (a_1 A^1_L, a_2 A^2_L, \dots, a_n A^n_L), \quad b_L = \sum_{i=1}^{n} a_i b^i_L,$$

and therefore $\|A_L\|_\infty \leq \max_i\{|a_i|\mathcal{B}(\Phi_i)\}$ and $\|b_L\|_\infty \leq \sum_{i=1}^n |a_i|\mathcal{B}(\Phi_i)$. In the case of Lemma A. 8, (1) follows upon noting that the network $\Psi \in \mathcal{N}_{d,d'}$ is specified by the same matrix-vector pairs as in the modified Lemma II. 6, except for the matrix A_1, which is instead given by

$$A_1 = \begin{pmatrix} A^1_1 \\ \vdots \\ A^n_1 \end{pmatrix}.$$

Note that in the special case $a_i \geq 1$, $b^i_L = 0$, both for all $i \in \{1,\dots,n\}$, the original claim $\mathcal{B}(\Psi) = \max_i\{|a_i|\mathcal{B}(\Phi_i)\}$ remains correct.

These corrections do not affect the instances where Lemma II. 6 and Lemma A. 8 are applied. Specifically, in the proof of Theorem VII. 2,

$$\mathcal{B}(\Psi_{f,M}) \leq \pi_3(M), \quad \text{for all } f \in \mathcal{C}, M \in \mathbb{N},$$

still holds for some (other) polynomial π_3 due to $I_{f,M} \subseteq \{1,\dots,\pi_1(M)\}$ and $\max_{i\in I_{f,M}} |c_i| \leq \pi_1(M)$.

In the proof of Theorem VIII. 10, the modification resulting from the correction in Lemma II. 6 yields an upper bound on the weight magnitude of the networks approximating ψ_m which continues to be independent of ε and, as such, does not lead to changes in the subsequent arguments.

Finally, in the proof of Lemma IX. 5, we consider a linear combination (with $a_i = 1$, for all $i \in \{1,\dots,d\}$) of squaring networks in order to obtain an approximation of $f(x) = \sum_{i=1}^{d} x_i^2$. As $b_{m+1} = 0$ in the proof of Proposition III. 2, the bound $\mathcal{B}(\Psi_{d,D,\varepsilon}) \leq 1$ in (84) remains valid thanks to the comment at the end of the first paragraph in this document.

2. In the third paragraph on p. 32, the formulation "let $\mathcal{D} = \{\varphi_i\}_{i\in\mathbb{N}}$ be an ordered orthonormal basis for $\mathcal{C}$" is meant to say that $\mathcal{D} = \{\varphi_i\}_{i\in\mathbb{N}}$ is an ordered orthonormal basis for a space that contains $\mathcal{C}$.

3. In the paragraph following Definition IV. 3, "for a given $x \in \mathcal{X}$" should read "for a given $x \in \mathcal{C}$" and "the resulting error satisfies $\|D(E(x)) - x\| \leq \varepsilon$" should read "the resulting error satisfies $\rho(D(E(x)), x) \leq \varepsilon$".

4. In Definition IV. 4, "$M(\varepsilon; \mathcal{X}, \rho)$" should read "$M(\varepsilon; \mathcal{C}, \rho)$".

5. In the caption of Table 1 it should say "bounded Lipschitz domain" instead of "Lipschitz domain".

6. On p. 29, the second and third display should read as follows:

   Now, note that

$$\begin{aligned}\Bigg\| \sum_{i\in\tilde{I}_{f,\widetilde{M}}} \tilde{c}_i \tilde{\varphi}_i \Bigg\|_{L^2(\Omega)} &= \Bigg\| f - (f - \sum_{i\in\tilde{I}_{f,\widetilde{M}}} \tilde{c}_i \tilde{\varphi}_i) \Bigg\|_{L^2(\Omega)} \\ &\leq \|f\|_{L^2(\Omega)} + \Bigg\| f - \sum_{i\in I_{f,M}} c_i \varphi_i \Bigg\|_{L^2(\Omega)} .\end{aligned}$$

   Making use of the orthonormality of the $\tilde{\varphi}_i$, we can conclude that

$$\sum_{i\in\tilde{I}_{f,\widetilde{M}}} |\tilde{c}_i|^2 \leq \left(\sup_{f\in\mathcal{C}} \|f\|_{L^2(\Omega)} + CM^{-\gamma} \right)^2 .$$

7. The authors are grateful to Erwin Riegler for pointing out the following issues in Appendix B. The first sentence should read "We consider, for $m \in \mathbb{R}_+$ and $p, q \in [1, \infty]$, the Besov space ..." and the first paragraph on p. 75 should be replaced by the following:

   In order to find an upper bound on $\|(\|A_n(f)\|_{\ell^2})_{n=N+1}^{\infty}\|_{\ell^2}$, we first note that

$$\|\cdot\|_{\ell^\alpha} \leq \|\cdot\|_{\ell^\beta}, \qquad \leq \beta \leq \alpha \leq \infty, \tag{2}$$

holds for sequences as well as vectors. For $x \in \mathbb{R}^d$ and $\alpha \in [2, \infty]$, we have, in addition, that

$$\|x\|_{\ell^2} \leq d^{\frac{1}{2} - \frac{1}{\alpha}} \|x\|_{\ell^\alpha}, \tag{3}$$

which is a direct consequence of Hölder's inequality applied to the vectors $(x_1^2, \ldots, x_d^2)$ and $(1, \ldots, 1)$ with corresponding Hölder exponents $\frac{\alpha}{2}$ and $\frac{\alpha}{\alpha-2}$, respectively. Upon noting that the $A_n(f)$ are vectors of length $|\mathcal{D}_n| = 2^n$, application of (3) combined with (2) yields, for $p \in [1, \infty]$, that

$$\|A_n(f)\|_{\ell^2} \leq 2^{n(\frac{1}{2} - \frac{1}{p})_+} \|A_n(f)\|_{\ell^p}, \tag{4}$$

where $x_+ := \max\{0, x\}$. Consequently, we get, for all $f \in \mathcal{U}(B^m_{p,q}([0,1]))$, $p \in [1, \infty]$, $q \in [1, 2]$, that

$$\begin{aligned}
&\|(\|A_n(f)\|_{\ell^2})_{n=N+1}^\infty\|_{\ell^2} \\
&\quad \overset{(a)}{\leq} \|(2^{n(\frac{1}{2} - \frac{1}{p})_+} \|A_n(f)\|_{\ell^p})_{n=N+1}^\infty\|_{\ell^q} \\
&\quad \overset{(b)}{\leq} 2^{-(N+1)(m - (\frac{1}{p} - \frac{1}{2})_+)} \|(2^{n(m + \frac{1}{2} - \frac{1}{p})} \|A_n(f)\|_{\ell^p})_{n=N+1}^\infty\|_{\ell^q} \\
&\quad \overset{(c)}{\leq} 2^{-(N+1)(m - (\frac{1}{p} - \frac{1}{2})_+)} \|f\|_{m,p,q} \\
&\quad \overset{(d)}{\leq} (2^{N+1})^{-(m - (\frac{1}{p} - \frac{1}{2})_+)},
\end{aligned}$$

which establishes (94) with $C = 1$ and $\beta = m - (\frac{1}{p} - \frac{1}{2})_+$. Note that for (a) we used (4) for the inner norm and (2) for the outer norm. For (b) we employed the fact that the sequence starts with index $n = N + 1$ and we used that $(a - b)_+ = (b - a)_+ + a - b$, for $a, b \in \mathbb{R}$, which yields

$$\begin{aligned}
2^{n(\frac{1}{2} - \frac{1}{p})_+} &= 2^{n((\frac{1}{p} - \frac{1}{2})_+ + \frac{1}{2} - \frac{1}{p})} \\
&= 2^{-n(m - (\frac{1}{p} - \frac{1}{2})_+)} \, 2^{n(m + \frac{1}{2} - \frac{1}{p})},
\end{aligned}$$

for $n, m \in \mathbb{R}$ and $p \in [1, \infty]$, with the convention $\frac{1}{\infty} = 0$. Finally, (c) and (d) follow by application of the definitions in (92) and (93), respectively.