CORRECTION TO "NONPARAMETRIC REGRESSION USING DEEP NEURAL NETWORKS WITH RELU ACTIVATION FUNCTION"

BY JOHANNES SCHMIDT-HIEBER AND DON VU

University of Twente and Vrije Universiteit Amsterdam

**Correction:** Condition (ii) in Theorem 1 should be changed to

$$(ii') : \sum_{i=0}^{q} \frac{\beta_i + t_i}{2 \beta_i^* + t_i} \log_2(4t_i \vee 4\beta_i) \log_2(n) \leq L \lesssim n \phi_n.$$  

Moreover, the constants $C, C'$ in Theorem 1 also depend on the implicit constants that appear in the Conditions (ii) - (iv). There are large regimes where the new condition (ii') is weaker than (ii).

**Explanation:** Rather than choosing $m$ and $N$ in the proof of Theorem 1 globally, one should instead apply Theorem 5 individually to each $i$ with

$$m_i := \left\lceil \frac{\beta_i + t_i}{2 \beta_i^* + t_i} \log_2(n) \right\rceil \quad \text{and} \quad N_i := \left\lceil cn_t/(2 \beta_i^* + t_i) \right\rceil,$$

where $0 < c \leq 1/2$ is a sufficiently small constant. As mentioned at the beginning of the proof of Theorem 1, it is sufficient to prove the result for sufficiently large $n$. Therefore, we can assume that $m_i \geq 1$ for all $i = 0, \ldots, q$ and $N_i \leq n t_i/(2 \beta_i^* + t_i)$. The latter implies

$$N_i 2^{-m_i} \leq N_i \left(n^{-\frac{t_i}{2 \beta_i^* + t_i}}\right)^{\beta_i + t_i} \leq N_i^{-\frac{\beta_i}{t_i}}.$$  

(1)

If we now define

$$L_i' := 8 + (m_i + 5)(1 + \lceil \log_2(t_i \vee \beta_i) \rceil),$$

then there exists a network $\tilde{h}_{ij} \in F(L_i', (t_i, 6(t_i + \lceil \beta_i \rceil)N_i, \ldots, 6(t_i + \lceil \beta_i \rceil)N_i, 1), s_i)$ with $s_i \leq 141(t_i + \beta_i + 1)^{3+t_i}N_i(m_i + 6)$, such that using (1),

$$\|\tilde{h}_{ij} - h_{ij}\|_{L^\infty([0,1]^d)} \leq (2Q_i + 1)(1 + t_i^2 + \beta_i^2)6t_iN_i 2^{-m_i} + Q_i 3^{\beta_i} N_i^{-\frac{\beta_i}{t_i}}$$

(3)

$$\leq \left((2Q_i + 1)(1 + t_i^2 + \beta_i^2)6t_i + Q_i 3^{\beta_i}\right)N_i^{-\frac{\beta_i}{t_i}}.$$
where $Q_i$ is any upper bound of the Hölder norms of $h_{ij}$, $j = 1, \ldots, d_i+1$. We can now argue as in the original proof to show that the composite network $f^*$ is in the class $\mathcal{F}(E, (d, 6r_i \max_i N_i, \ldots, 6r_i \max_i N_i, 1), \sum_{i=0}^q d_{i+1}(s_i+4))$, with $E := 3q + \sum_{i=0}^q L_i$. Using the definition of $L_i$ in (2) it can be shown as in the original proof that $E \leq \sum_{i=0}^q \frac{\beta_i + 1}{2^\beta_i + 1} (\log_2(4 \log_2(t_i \vee t_1)) \log_2(n)$ for all sufficiently large $n$. All remaining steps are the same as in the original proof of Theorem 1. The constant $c$ in the definition of $N_i$ will also depend on the implicit constant in the conditions $L \lesssim n\phi n$, $n\phi n \lesssim \min_{i=1,\ldots,L} p_i$ and $s \asymp n\phi n \log n$.

Further comments:

- Lemma 1 requires that the constant $K$ is large enough such that Theorem 3 is applicable.
- First display on p.1886: The value $t_2$ is $N$ not $Nd$.
- Equation (18) also requires that the inputs are non-negative.
- In Lemma 3, the $L^\infty$-norms should be replaced by the supremum, that is, $\|f\|_{L^\infty(A)}$ should be changed to $\sup_{x \in A} |f(x)|$.
- In (22) and two lines after (22), $R(\hat{f}, f)$ should be $R(\hat{f}_n, f)$.
- In the proof of Theorem 1, $r_i$ does not depend on $i$ and should be named $r$. Three lines after Equation (26), $C$ should be replaced by $C'$.
- In the proof of Theorem 3, $\beta^*$ in the first line on p.1893 should be $\beta^{**}$. It is sufficient to check that the Hölder constant of $\phi_w$ is bounded by $(\beta^* + 1)^t(t^* + 1)$ as all later arguments of the proof carry over. Moreover $g_i(x) = (x_1, \ldots, x_{d_i})^\top$ should be $g_i(x) = (x_1, \ldots, x_{d_i+1})^\top$ if $d_i \geq d_{i+1}$ and $g_i(x) = (x_1, \ldots, x_{d_i}, 0, \ldots, 0)^\top$ if $d_i < d_{i+1}$. Finally $\|\psi_u\|_2^2$ should be replaced by $\|\psi_u^r\|_2^2$.
- In the proof of Lemma 2, one can simply take the constant function $h_{j,0} = K$ if $\mu_0 \neq 0$. This immediately gives $d_{j,k} = K\mu_0^{d_2}j_{-j_2}^2$ for all wavelet coefficients $d_{j,k}$. For the case $\mu_0 = 0$, one should replace the binomial coefficient $\binom{dr}{r}$ by the multinomial coefficient $\binom{dr}{r_1,\ldots,r} = (dr)!/(r!)^d$.
- To verify the last inequality in (B.8) of the Supplementary Material, one can replace $1/M$ by $< 1/M$.
- In the proof of Lemma 4, some $F$ are missing. In particular it should be $r_j := F\sqrt{n^{-1} \log N_n} \vee E^{1/2}((f_j(X) - f_0(X))^2]$. Also on p.12 of the Supplementary Material it should be

$$P(T \geq t) \leq 1 \wedge 2N_n \max_j \exp\left( -\frac{t^2}{8t F/(3r_j) + 16n}\right).$$

Since $r_j \geq F\sqrt{n^{-1} \log N_n}$ we can argue as before and obtain $P(T \geq t)$...
$t) \leq 2\mathcal{N}_n \exp(-3t \sqrt{\log \mathcal{N}_n/(16\sqrt{n})})$ for all $t \geq 6\sqrt{n \log \mathcal{N}_n}$. The conclusion of (I) is still valid.

- In the proof of Lemma 5, we always work with the $\| \cdot \|_\infty$-norm for vectors. The grid size of an individual parameter should be taken as $\delta/((L + 1)V)$. 