Random Vector Functional Link Neural Networks as Universal Approximators

Palina Salanevich
joint with D. Needell, A. Nelson, and R. Saab

University of California, Los Angeles
Email: psalanevich@math.ucla.edu

UCLA

September 23, 2019

Introduction: a neuron model

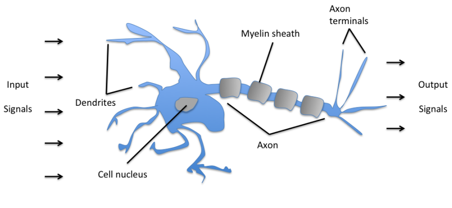


Figure: Anatomy (left) and a mathematical model (right) of a neuron.

Introduction: a neuron model

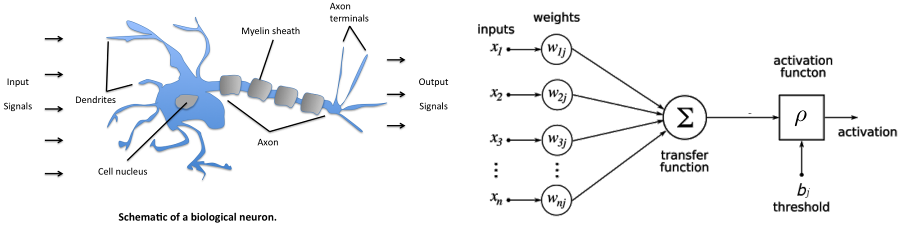
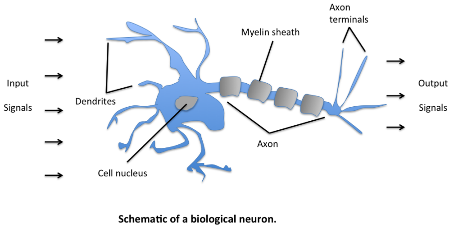


Figure: Anatomy (left) and a mathematical model (right) of a neuron.

Introduction: a neuron model



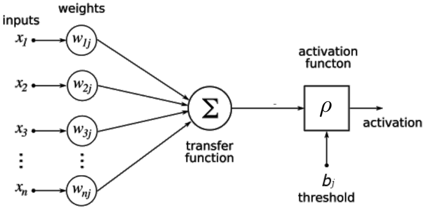


Figure: Anatomy (left) and a mathematical model (right) of a neuron.

$$x = (x_1, \ldots, x_m) \mapsto$$

- $x = (x_1, \ldots, x_m)$ is an input signal

Introduction: a neuron model

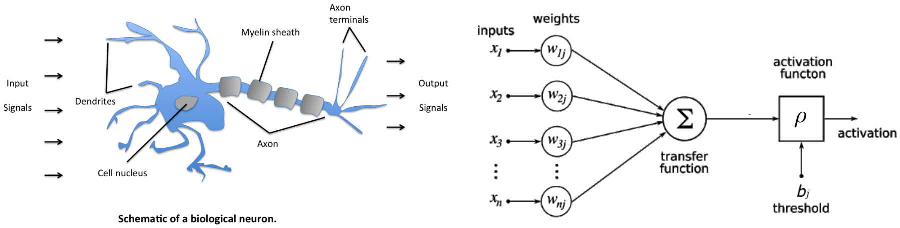


Figure: Anatomy (left) and a mathematical model (right) of a neuron.

$$x = (x_1, \ldots, x_m) \mapsto \langle x, \omega \rangle$$

- $x = (x_1, \ldots, x_m)$ is an input signal
- $\omega = (w_1, \ldots, w_m)$ is the vector of input weights

Introduction: a neuron model

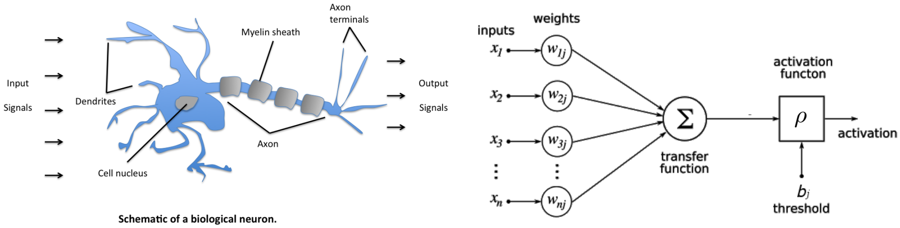
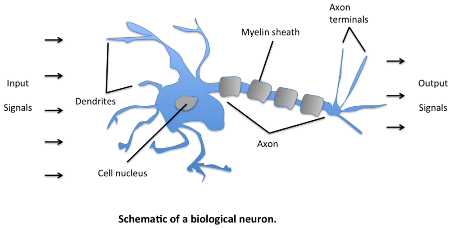


Figure: Anatomy (left) and a mathematical model (right) of a neuron.

$$x = (x_1, \dots, x_m) \mapsto \langle x, \omega \rangle + b$$

- $x = (x_1, \dots, x_m)$ is an input signal
- $\omega = (w_1, \dots, w_m)$ is the vector of input weights
- b is a threshold

Introduction: a neuron model



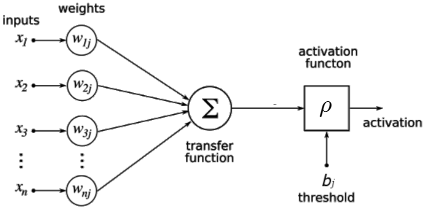


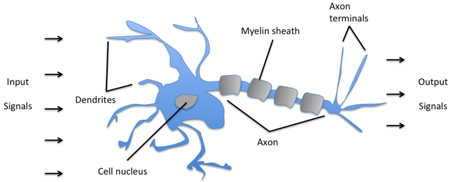
Figure: Anatomy (left) and a mathematical model (right) of a neuron.

$$x = (x_1, \dots, x_m) \mapsto \rho\big(\langle x, \omega \rangle + b\big)$$

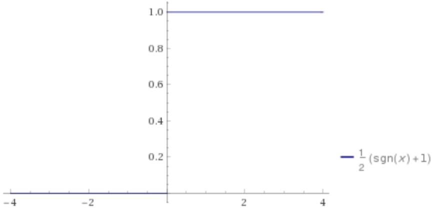
- $x = (x_1, \dots, x_m)$ is an input signal
- $\omega = (w_1, \dots, w_m)$ is the vector of input weights
- b is a threshold
- ρ is an activation function

Introduction: activation function

$$x = (x_1, \ldots, x_m) \mapsto \rho(\langle x, \omega \rangle + b)$$

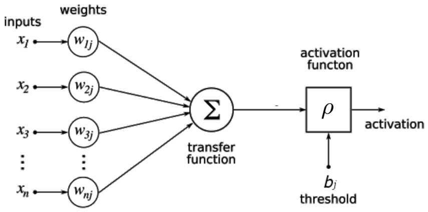


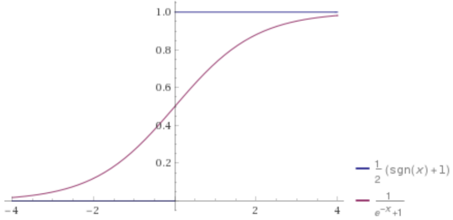
Schematic of a biological neuron.



Introduction: activation function

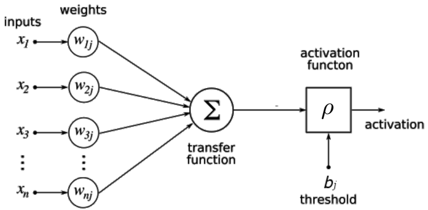
$$x = (x_1, \ldots, x_m) \mapsto \rho\big(\langle x, \omega \rangle + b\big)$$





Introduction: activation function

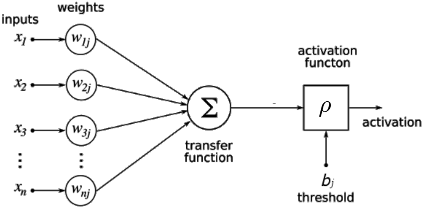
$$x = (x_1, \ldots, x_m) \mapsto \rho\big(\langle x, \omega \rangle + b\big)$$



Sigmoid	$\rho(z) = \frac{1}{1+\exp(-z)}$
ReLU	$\rho(z) = \max\{0, z\}$
Sine	$\rho(z) = \sin(z)$
Hardlim	$\rho(z) = \frac{1}{2}(1 + \sin(z))$
Radbas	$\rho(z) = \exp(-z^2)$

Introduction: activation function

$$x = (x_1, \ldots, x_m) \mapsto \rho\big(\langle x, \omega\rangle + b\big)$$



Sigmoid	$\rho(z) = \frac{1}{1+\exp(-z)}$
ReLU	$\rho(z) = \max\{0, z\}$
Sine	$\rho(z) = \sin(z)$
Hardlim	$\rho(z) = \frac{1}{2}(1 + \sin(z))$
Radbas	$\rho(z) = \exp(-z^2)$

Our assumptions: $g \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ (plus some piecewise continuity)

OR g differentiable with $g' \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$

Introduction: neural nets

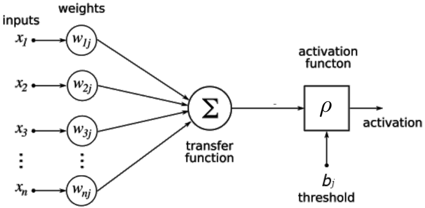


Figure: Single neuron: $\mathcal{F}_1 = \Big\{ f_1(\cdot) = \rho(\langle \cdot, \omega \rangle + b) : b \in \mathbb{R}, \omega \in \mathbb{R}^m \Big\}$.

Introduction: neural nets

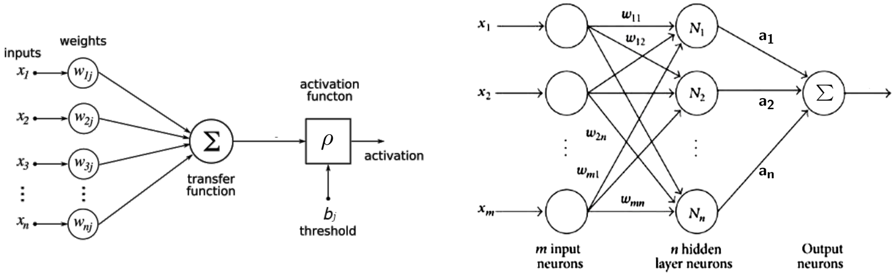
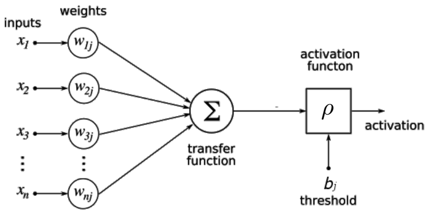


Figure: Single neuron: $\mathcal{F}_1 = \left\{ f_1(\cdot) = \rho(\langle \cdot, \omega \rangle + b) : b \in \mathbb{R}, \omega \in \mathbb{R}^m \right\}$.
Single layer neural net: $\mathcal{F}_n = \left\{ f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j) : a_j, b_j \in \mathbb{R}, \omega_j \in \mathbb{R}^m \right\}$.

Introduction: neural nets



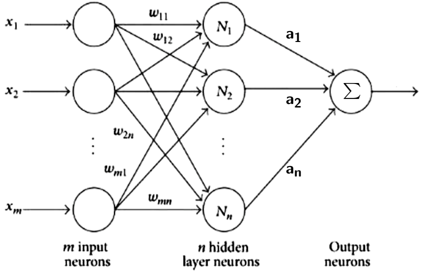


Figure: Single neuron: $\mathcal{F}_1 = \left\{ f_1(\cdot) = \rho(\langle \cdot, \omega \rangle + b) : b \in \mathbb{R}, \omega \in \mathbb{R}^m \right\}$.
Single layer neural net: $\mathcal{F}_n = \left\{ f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j) : a_j, b_j \in \mathbb{R}, \omega_j \in \mathbb{R}^m \right\}$.

Question

What class $\mathcal{F}$ of functions can be approximated by $\mathcal{F}_n$ so that $\forall f \in \mathcal{F}$ there exists $(f_n)_{n \in \mathbb{N}}$, $f_n \in \mathcal{F}_n$, s.t. $\lim_{n \to \infty} ||f - f_n|| = 0$?

Neural nets as universal approximators

Theorem (Barron, 1993)

Single layer neural net is a universal approximator for $\mathcal{F} = C_c(\mathbb{R})$. More precisely, $\forall f \in C_c(\mathbb{R})$ there exists $(f_n)_{n\in\mathbb{N}}$, $f_n \in \mathcal{F}_n$, s.t. $||f - f_n||_2 = O\left(\frac{1}{\sqrt{n}}\right)$.

Neural nets as universal approximators

Theorem (Barron, 1993)

Single layer neural net is a universal approximator for $\mathcal{F} = C_c(\mathbb{R})$. More precisely, $\forall f \in C_c(\mathbb{R})$ there exists $(f_n)_{n\in\mathbb{N}}$, $f_n \in \mathcal{F}_n$, s.t. $||f - f_n||_2 = O\left(\frac{1}{\sqrt{n}}\right)$.

But: How do we find the parameters $a_j, b_j \in \mathbb{R}$, $\omega_j \in \mathbb{R}^m$, $j \in \{1, \ldots n\}$?

given	$T = \{(x_i, f(x_i))\}_{i=1}^N$
find	$f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j)$
subject to	$\|\|f - f_n\|\|_2 < \varepsilon,$

Neural nets as universal approximators

Theorem (Barron, 1993)

Single layer neural net is a universal approximator for $\mathcal{F} = C_c(\mathbb{R})$. More precisely, $\forall f \in C_c(\mathbb{R})$ there exists $(f_n)_{n\in\mathbb{N}}$, $f_n \in \mathcal{F}_n$, s.t. $||f - f_n||_2 = O\left(\frac{1}{\sqrt{n}}\right)$.

But: How do we find the parameters $a_j, b_j \in \mathbb{R}$, $\omega_j \in \mathbb{R}^m$, $j \in \{1, \dots n\}$?

given	$T = \{(x_i, f(x_i))\}_{i=1}^{N}$
find	$f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j)$
subject to	$\|\|f - f_n\|\|_2 < \varepsilon,$

- slow convergence, algorithms get stuck in local minima
- highly sensitive to training data

Neural nets as universal approximators

Idea: optimize only a part of parameters, while keeping the others fixed.

given	$b_j \in \mathbb{R}, \omega_j \in \mathbb{R}^m$
find	$f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j)$
subject to	$\|\|f - f_n\|\|_2 < \varepsilon,$

Theorem (Barron, 1993)

For $n \in \mathbb{N}$, let us fix $b_j \in \mathbb{R}$, $\omega_j \in \mathbb{R}^m$, $j \in \{1, \dots, n\}$. Then, $\forall f \in C_c(\mathbb{R})$ there exists $(a_j)_{j=1}^n$, such that for $f_n(\cdot) = \sum_{j=1}^n a_j \rho(\langle \cdot, \omega_j \rangle + b_j)$,

$$||f - f_n||_2 = O\left(\frac{1}{n^{\frac{2}{m}}}\right).$$

Note: As $\lim_{m \to \infty} n^{\frac{2}{m}} = 1$, the bound is not useful for large dimensions.

Random vector functional link network (RVFL)

Better idea: optimize only $(a_j)_{j=1}^n$, choose $b_j \in \mathbb{R}$, $\omega_j \in \mathbb{R}^m$ at random.

$$\begin{array}{|ll|} \text{find} & \text{distribution for } b_j \in \mathbb{R}, \omega_j \in \mathbb{R}^m \\ & \text{and } f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j) \\ \text{subject to} & ||f - f_n||_2 < \varepsilon, \end{array}$$

Random vector functional link network (RVFL)

Better idea: optimize only $(a_j)_{j=1}^n$, choose $b_j \in \mathbb{R}$, $\omega_j \in \mathbb{R}^m$ at random.

$$\begin{array}{|ll|} \text{find} & \text{distribution for } b_j \in \mathbb{R}, \omega_j \in \mathbb{R}^m \\ & \text{and } f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j) \\ \text{subject to} & ||f - f_n||_2 < \varepsilon, \end{array}$$

- used in time-series data prediction, handwritten word recognition, visual tracking, and other signal classification and regression problems

Random vector functional link network (RVFL)

Better idea: optimize only $(a_j)_{j=1}^n$, choose $b_j \in \mathbb{R}$, $\omega_j \in \mathbb{R}^m$ at random.

$$\begin{array}{|ll|}\hline \text{find} & \text{distribution for } b_j \in \mathbb{R}, \omega_j \in \mathbb{R}^m \\ & \text{and } f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j) \\ \text{subject to} & ||f - f_n||_2 < \varepsilon, \\ \hline \end{array}$$

- used in time-series data prediction, handwritten word recognition, visual tracking, and other signal classification and regression problems
- show similar performance to the classical SLFN (with all parameters learned), but with much faster and more efficient learning process

Random vector functional link network (RVFL)

Better idea: optimize only $(a_j)_{j=1}^n$, choose $b_j \in \mathbb{R}$, $\omega_j \in \mathbb{R}^m$ at random.

$$\begin{array}{ll} \text{find} & \text{distribution for } b_j \in \mathbb{R}, \omega_j \in \mathbb{R}^m \\ & \text{and } f_n(\cdot) = \sum_{j=1}^{n} a_j \rho(\langle \cdot, \omega_j \rangle + b_j) \\ \text{subject to} & ||f - f_n||_2 < \varepsilon, \end{array}$$

- used in time-series data prediction, handwritten word recognition, visual tracking, and other signal classification and regression problems
- show similar performance to the classical SLFN (with all parameters learned), but with much faster and more efficient learning process
- to date, luck of theoretical analysis

RVFL as a uniform approximator (on average)

Theorem (Igelnik and Pao, 1995)

Let $f \in C_c(\mathbb{R}^m)$. There exist distributions for parameters b_j, ω_j, $j \in \{1, \ldots, n\}$ and weights $\{a_j\}_{k=1}^n$ such that the sequence $\left\{f_n(\cdot) = \sum_{j=1}^n a_j \rho(\langle \cdot, \omega_j \rangle + b_j)\right\}_{n=1}^\infty$ of RVFL networks satisfies

$$\lim_{n\to\infty} \mathbb{E} \int_{\mathrm{supp}(f)} |f(x) - f_n(x)|^2 dx = 0,$$

with convergence rate $O(1/n)$.

RVFL as a uniform approximator (on average)

Theorem (Igelnik and Pao, 1995)

Let $f \in C_c(\mathbb{R}^m)$. There exist distributions for parameters b_j, ω_j, $j \in \{1, \ldots, n\}$ and weights $\{a_j\}_{k=1}^n$ such that the sequence $\left\{f_n(\cdot) = \sum_{j=1}^n a_j \rho(\langle \cdot, \omega_j \rangle + b_j)\right\}_{n=1}^\infty$ of RVFL networks satisfies

$$\lim_{n\to\infty} \mathbb{E} \int_{\mathrm{supp}(f)} |f(x) - f_n(x)|^2 dx = 0,$$

with convergence rate $O(1/n)$.

RVFL as a uniform approximator (on average)

Theorem (Igelnik and Pao, 1995)

Let $f \in C_c(\mathbb{R}^m)$. For any $\varepsilon > 0$, there exist distributions for parameters b_j, ω_j, $j \in \{1, \dots, n\}$ and weights $\{a_j\}_{k=1}^n$ such that the sequence of RVFL networks $\left\{f_n(\cdot) = \sum_{j=1}^n a_j \rho(\langle \cdot, \omega_j \rangle + b_j)\right\}_{n=1}^\infty$ satisfies

$$\lim_{n\to\infty} \mathbb{E} \int_{\mathrm{supp}(f)} |f(x) - f_n(x)|^2 dx < \varepsilon,$$

with convergence rate $< O(1/n)$.

Distribution: there exist constants $\alpha(\varepsilon)$, $\Omega(\varepsilon)$ large enough, so that

$$\begin{aligned}
\omega_j &\sim U([-\alpha\Omega, \alpha\Omega])^m;\\
y_j &\sim U(\mathrm{supp}(f));\\
u_j &\sim U([-\tfrac{\pi}{2}(2L+1), \tfrac{\pi}{2}(2L+1)]), \quad \text{where } L := \lceil \tfrac{2m}{\pi}\mathrm{rad}(K)\Omega - \tfrac{1}{2} \rceil;\\
b_j &= -\langle \omega_j, y_j \rangle - \alpha u_j,
\end{aligned}$$

RVFL as a uniform approximator (on average)

Theorem (Needell, Nelson, Saab, S.)

Let $f \in C_c(\mathbb{R}^m)$. There exist distributions for parameters b_j, ω_j, $j \in \{1, \dots, n\}$ and weights $\{a_j\}_{k=1}^n$ such that the sequence $\left\{f_n(\cdot) = \sum_{j=1}^n a_j \rho(\langle \cdot, \omega_j \rangle + b_j)\right\}_{n=1}^\infty$ of RVFL networks satisfies

$$\lim_{n\to\infty} \mathbb{E} \int_{\mathrm{supp}(f)} |f(x) - f_n(x)|^2 dx = 0,$$

(but with no convergence rate guarantees).

RVFL as a uniform approximator (on average)

Theorem (Needell, Nelson, Saab, S.)

Let $f \in C_c(\mathbb{R}^m)$. There exist distributions for parameters b_j, ω_j, $j \in \{1, \dots, n\}$ and weights $\{a_j\}_{k=1}^n$ such that the sequence $\left\{f_n(\cdot) = \sum_{j=1}^n a_j \rho(\langle \cdot, \omega_j \rangle + b_j)\right\}_{n=1}^\infty$ of RVFL networks satisfies

$$\lim_{n\to\infty} \mathbb{E} \int_{\mathrm{supp}(f)} |f(x) - f_n(x)|^2 dx = 0,$$

(but with no convergence rate guarantees).

Distribution: there exist $\alpha_n \to \infty$, $\Omega_n \to \infty$ as $n \to \infty$, so that for f_n

$$\begin{aligned}
\omega_j &\sim U([-\alpha_n\Omega_n, \alpha_n\Omega_n])^m;\\
y_j &\sim U(\mathrm{supp}(f));\\
u_j &\sim U([-\tfrac{\pi}{2}(2L_n+1), \tfrac{\pi}{2}(2L_n+1)]), \quad \text{where } L_n := \lceil \tfrac{2m}{\pi}\mathrm{rad}(K)\Omega_n - \tfrac{1}{2}\rceil;\\
b_j &= -\langle \omega_j, y_j\rangle - \alpha u_j,
\end{aligned}$$

RVFL: non-asymptotic probabilistic bounds

Theorem (Needell, Nelson, Saab, S.)

Let $f \in C_c(\mathbb{R}^m)$ with $K = \mathrm{supp}(f)$. For any $\varepsilon > 0$ and $\eta \in (0,1)$, there exist distributions (as above) for parameters b_j, ω_j, $j \in \{1,\ldots,n\}$ and weights $\{a_j\}_{k=1}^n$ such that if

$$n \gtrsim \frac{\alpha m^2 \Omega^{m+1} \mathrm{rad}(K)^{(3m+2)/2} ||f||_\infty ||\rho||_\infty \log(\eta^{-m}\delta^{-1} m^{1/2} \mathrm{rad}(K))}{\varepsilon \log(1 + \frac{\varepsilon ||\rho||_\infty}{\alpha m \Omega^{m+1} \mathrm{rad}(K)^{(m+2)/2} ||f||_\infty ||\rho||_2^2})},$$

then the RVFL network $f_n(\cdot) = \sum_{j=1}^n a_j \rho(\langle \cdot, \omega_j \rangle + b_j)$ satisfies

$$\mathbb{P}\left(\int_K |f(x) - f_n(x)|^2 dx < \varepsilon \right) \geq 1 - \eta.$$

RVFL: non-asymptotic probabilistic bounds

Theorem (Needell, Nelson, Saab, S.)

Let $f \in C_c(\mathbb{R}^m)$ with $K = \mathrm{supp}(f)$. For any $\varepsilon > 0$ and $\eta \in (0,1)$, there exist distributions (as above) for parameters b_j, ω_j, $j \in \{1,\dots,n\}$ and weights $\{a_j\}_{k=1}^n$ such that if

$$n \gtrsim \frac{\alpha m^2 \Omega^{m+1} \mathrm{rad}(K)^{(3m+2)/2} ||f||_\infty ||\rho||_\infty \log(\eta^{-m}\delta^{-1} m^{1/2} \mathrm{rad}(K))}{\varepsilon \log(1 + \frac{\varepsilon ||\rho||_\infty}{\alpha m \Omega^{m+1} \mathrm{rad}(K)^{(m+2)/2} ||f||_\infty ||\rho||_2^2})},$$

then the RVFL network $f_n(\cdot) = \sum_{j=1}^n a_j \rho(\langle \cdot, \omega_j\rangle + b_j)$ satisfies

$$\mathbb{P}\left(\int_K |f(x) - f_n(x)|^2 dx < \varepsilon\right) \geq 1 - \eta.$$

Note: for small $\varepsilon > 0$, the requirement on the number of nodes behaves like

$$n \gtrsim \varepsilon^{-2} \log(\eta^{-1} \mathcal{N}(\delta, K)).$$

Idea of the proof

Goal: for a function $f \in C_c(\mathbb{R}^m)$, construct a random approximation

$$f(x) \approx f_n(x) = \sum_{j=1}^{n} a_j \rho \left(\sum_{i=1}^{m} x_i \omega_{ji} + b_j \right)$$

Idea of the proof

Goal: for a function $f \in C_c(\mathbb{R}^m)$, construct a random approximation

$$f(x) \approx f_n(x) = \sum_{j=1}^{n} a_j \rho \left(\sum_{i=1}^{m} x_i \omega_{ji} + b_j \right)$$

- **Step 1:** Construct a limit-integral representation of f using *convolution identity*.

Idea of the proof

Goal: for a function $f \in C_c(\mathbb{R}^m)$, construct a random approximation

$$f(x) \approx f_n(x) = \sum_{j=1}^{n} a_j \rho \left(\sum_{i=1}^{m} x_i \omega_{ji} + b_j \right)$$

- **Step 1:** Construct a limit-integral representation of f using *convolution identity*.
- **Step 2:** Use the *Monte Carlo method* to approximate the integral in the constructed limit-integral representation of f using a linear combination of random realizations of the activation function ρ.

Idea of the proof

Goal: for a function $f \in C_c(\mathbb{R}^m)$, construct a random approximation

$$f(x) \approx f_n(x) = \sum_{j=1}^{n} a_j \rho \left(\sum_{i=1}^{m} x_i \omega_{ji} + b_j \right)$$

- **Step 1:** Construct a limit-integral representation of f using *convolution identity*.
- **Step 2:** Use the *Monte Carlo method* to approximate the integral in the constructed limit-integral representation of f using a linear combination of random realizations of the activation function ρ.

Note: As a byproduct, we obtain an *explicit formula* for parameters $\{a_j\}_{j=1}^n$ in terms of function f and random parameters ω_j, b_j.

Limit-integral representation

Assume wlog that $\int_{\mathbb{R}} g(x)dx = 1$ and consider approximate δ-functions

$$h_w(y) = \prod_{j=1}^{m} w(j)\rho\big(w(j)y(j)\big) \quad y, w \in \mathbb{R}^m.$$

Limit-integral representation

Assume wlog that $\int_{\mathbb{R}} g(x)dx = 1$ and consider approximate δ-functions

$$h_w(y) = \prod_{j=1}^{m} w(j)\rho\big(w(j)y(j)\big) \quad y, w \in \mathbb{R}^m.$$

Lemma

Let $f \in C_0(\mathbb{R}^m)$. Then for all $x \in \mathbb{R}^m$ we have

$$f(x) = \lim_{\Omega\to\infty} \frac{1}{\Omega^m} \int_{[0,\Omega]^m} (f * h_w)(x)dw$$

$$= \lim_{\Omega\to\infty} \frac{1}{\Omega^m} \int_{[0,\Omega]^m} \int_{\mathbb{R}^m} f(y)\left(\prod_{j=1}^{m} w(j)\rho\Big(w(j)\big(x(j) - y(j)\big)\Big)\right)dydw.$$

Limit-integral representation

Assume wlog that $\int_{\mathbb{R}} g(x)dx = 1$ and consider approximate δ-functions

$$h_w(y) = \prod_{j=1}^{m} w(j)\rho\big(w(j)y(j)\big) \quad y, w \in \mathbb{R}^m.$$

Lemma

Let $f \in C_0(\mathbb{R}^m)$. Then for all $x \in \mathbb{R}^m$ we have

$$f(x) = \lim_{\Omega\to\infty} \frac{1}{\Omega^m} \int_{[0,\Omega]^m} (f * h_w)(x)dw$$

$$= \lim_{\Omega\to\infty} \frac{1}{\Omega^m} \int_{[0,\Omega]^m} \int_{\mathbb{R}^m} f(y)\left(\prod_{j=1}^{m} w(j)\rho\Big(w(j)\big(x(j) - y(j)\big)\Big)\right)dydw.$$

Problem: Need to replace the product with a sum.

Limit-integral representation

Assume wlog that $\int_{\mathbb{R}} g(x)dx = 1$ and consider approximate δ-functions

$$h_w(y) = \prod_{j=1}^{m} w(j)\rho\big(w(j)y(j)\big) \quad y, w \in \mathbb{R}^m.$$

Lemma

Let $f \in C_0(\mathbb{R}^m)$. Then for all $x \in \mathbb{R}^m$ we have

$$f(x) = \lim_{\Omega\to\infty} \frac{1}{\Omega^m} \int_{[0,\Omega]^m} (f * h_w)(x)dw$$

$$= \lim_{\Omega\to\infty} \frac{1}{\Omega^m} \int_{[0,\Omega]^m} \int_{\mathbb{R}^m} f(y)\Bigg(\prod_{j=1}^{m} w(j)\rho\Big(w(j)\big(x(j) - y(j)\big)\Big)\Bigg)dydw.$$

Problem: Need to replace the product with a sum.

Idea: Use $2\cos(a)\cos(b) = \cos(a-b) + \cos(a+b)$ iteratively to obtain

$$\prod_{j=1}^{m} \cos\big(w(j)z(j)\big) = \frac{1}{2^m} \sum_{\pm} \cos\big(\pm w(1)z(1)) \pm \cdots \pm w(m)z(m)\big)$$

Limit-integral representation

Let $L = \lceil \frac{2m}{\pi}\mathrm{rad}(K)\Omega - \frac{1}{2} \rceil$ and define

$$\cos_\Omega(x) := \begin{cases} \cos(x), & x \in [-\frac{1}{2}(2L+1)\pi, \frac{1}{2}(2L+1)\pi], \\ 0, & \text{otherwise.} \end{cases}$$

Limit-integral representation

Let $L = \lceil \frac{2m}{\pi}\mathrm{rad}(K)\Omega - \frac{1}{2} \rceil$ and define

$$\cos_\Omega(x) := \begin{cases} \cos(x), & x \in [-\frac{1}{2}(2L+1)\pi, \frac{1}{2}(2L+1)\pi], \\ 0, & \text{otherwise.} \end{cases}$$

Then $f(x) = \lim_{\Omega\to\infty} \frac{1}{(2\Omega)^m} \int_{K\times[-\Omega,\Omega]^m} f(y) \cos_\Omega\left(\langle w, x-y\rangle\right) \left| \prod_{j=1}^m w(j) \right| dydw$.

Limit-integral representation

Let $L = \lceil \frac{2m}{\pi}\mathrm{rad}(K)\Omega - \frac{1}{2} \rceil$ and define

$$\cos_\Omega(x) := \begin{cases} \cos(x), & x \in [-\frac{1}{2}(2L+1)\pi, \frac{1}{2}(2L+1)\pi], \\ 0, & \text{otherwise.} \end{cases}$$

Then $f(x) = \lim_{\Omega\to\infty} \frac{1}{(2\Omega)^m} \int_{K\times[-\Omega,\Omega]^m} f(y)\cos_\Omega\left(\langle w, x-y\rangle\right)\left|\prod_{j=1}^m w(j)\right| dydw$.
We have $\cos_\Omega(z) = \lim_{\alpha\to\infty}(\cos_\Omega * h_\alpha)(z)$, where $h_\alpha(y) = \alpha\rho(\alpha y)$.

Limit-integral representation

Let $L = \lceil \frac{2m}{\pi}\mathrm{rad}(K)\Omega - \frac{1}{2} \rceil$ and define

$$\cos_\Omega(x) := \begin{cases} \cos(x), & x \in [-\frac{1}{2}(2L+1)\pi, \frac{1}{2}(2L+1)\pi], \\ 0, & \text{otherwise.} \end{cases}$$

Then $f(x) = \lim_{\Omega\to\infty} \frac{1}{(2\Omega)^m} \int_{K\times[-\Omega,\Omega]^m} f(y) \cos_\Omega\left(\langle w, x-y\rangle\right) \left|\prod_{j=1}^m w(j)\right| dydw.$

We have $\cos_\Omega(z) = \lim_{\alpha\to\infty} (\cos_\Omega * h_\alpha)(z)$, where $h_\alpha(y) = \alpha\rho(\alpha y)$.

Lemma

Let $f \in C_c(\mathbb{R}^m)$ with $K := \mathrm{supp}(f)$. For all $\Omega \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$, define

$$F_{\alpha,\Omega}(y, w, u) := \frac{\alpha}{(2\Omega)^m} \left| \prod_{j=1}^m w(j) \right| f(y) \cos_\Omega(u),$$

$$b_\alpha(y, w, u) := -\alpha(\langle w, y\rangle + u)$$

Then, for any $x \in K$ and $K(\Omega) := K \times [-\Omega, \Omega]^m \times [-\frac{\pi}{2}(2L+1), \frac{\pi}{2}(2L+1)]$, we have

$$f(x) = \lim_{\Omega\to\infty} \lim_{\alpha\to\infty} \int_{K(\Omega)} F_{\alpha,\Omega}(y, w, u) \rho\big(\alpha\langle w, x\rangle + b_\alpha(y, w, u)\big) dydwdu.$$

Monte-Carlo approximation

$$\omega_j \sim U([-\alpha\Omega, \alpha\Omega])^m;$$
$$y_j \sim U(\mathrm{supp}(f));$$
$$u_j \sim U([-\tfrac{\pi}{2}(2L+1), \tfrac{\pi}{2}(2L+1)]), \quad \text{where } L := \lceil \tfrac{2m}{\pi}\mathrm{rad}(K)\Omega - \tfrac{1}{2} \rceil;$$
$$b_j = -\langle \omega_j, y_j \rangle - \alpha u_j,$$

Monte-Carlo approximation

$$\begin{aligned}
\omega_j &\sim U([-\alpha\Omega, \alpha\Omega])^m; \\
y_j &\sim U(\mathrm{supp}(f)); \\
u_j &\sim U([-\tfrac{\pi}{2}(2L+1), \tfrac{\pi}{2}(2L+1)]), \quad \text{where } L := \lceil \tfrac{2m}{\pi}\mathrm{rad}(K)\Omega - \tfrac{1}{2} \rceil; \\
b_j &= -\langle \omega_j, y_j \rangle - \alpha u_j,
\end{aligned}$$

Lemma

For $f \in C_c(\mathbb{R}^m)$ define $f_n(x) = \sum_{j=1}^n a_j \rho\big(\langle w_j, x\rangle + b_j\big)$, where

$$a_j = \frac{\mathrm{vol}(K(\Omega))}{n} F_{\alpha,\Omega}\big(y_j, \frac{w_j}{\alpha^m}, u_j\big), \quad j \in \{1, \ldots, n\}.$$

Then we have, for $C_{f,\rho,\alpha,\Omega,m} := \alpha^2 \|f\|_\infty^2 \Omega^{2m} \pi^2 (2L+1)^2 \mathrm{vol}(K)^2 \|\rho\|_2^2$,

$$\lim_{n\to\infty} \mathbb{E} \int_K \left| \int_{K(\Omega)} F_{\alpha,\Omega}(y, w, u) \rho\big(\alpha\langle w, x\rangle + b_\alpha(y, w, u)\big)\, dydwdu - f_n(x) \right|^2 dx \le \frac{C_{f,\rho,\alpha,\Omega,m}}{n}.$$

Monte-Carlo approximation

$$\omega_j \sim U([-\alpha\Omega, \alpha\Omega])^m;$$
$$y_j \sim U(\mathrm{supp}(f));$$
$$u_j \sim U([-\tfrac{\pi}{2}(2L+1), \tfrac{\pi}{2}(2L+1)]), \quad \text{where } L := \lceil \tfrac{2m}{\pi}\mathrm{rad}(K)\Omega - \tfrac{1}{2} \rceil;$$
$$b_j = -\langle \omega_j, y_j \rangle - \alpha u_j,$$

Lemma

For $f \in C_c(\mathbb{R}^m)$ *define* $f_n(x) = \sum_{j=1}^{n} a_j \rho\big(\langle w_j, x\rangle + b_j\big)$*, where*

$$a_j = \frac{\mathrm{vol}(K(\Omega))}{n} F_{\alpha,\Omega}\Big(y_j, \frac{w_j}{\alpha^m}, u_j\Big), \quad j \in \{1, \dots, n\}.$$

Then we have, for $C_{f,\rho,\alpha,\Omega,m} := \alpha^2 \|f\|_\infty^2 \Omega^{2m} \pi^2 (2L+1)^2 \mathrm{vol}(K)^2 \|\rho\|_2^2$*,*

$$\lim_{n\to\infty} \mathbb{E} \int_K \left| \int_{K(\Omega)} F_{\alpha,\Omega}(y,w,u) \rho\big(\alpha\langle w, x\rangle + b_\alpha(y,w,u)\big)\, dydwdu - f_n(x) \right|^2 dx \le \frac{C_{f,\rho,\alpha,\Omega,m}}{n}.$$

As $I(x) = \int_{K(\Omega)} F_{\alpha,\Omega}(y,w,u)\rho\big(\alpha\langle w, x\rangle + b_\alpha(y,w,u)\big)\, dydwdu \to f(x)$ as $\alpha, \Omega \to \infty$, can choose $\alpha, \Omega \to \infty$ large enough, so that $|I(x) - f(x)| < \varepsilon'$. Then

$$\big|f(x) - f_n(x)\big| < \varepsilon' + \big|I(x) - f_n(x)\big|$$

RVFL on Manifolds

How does the bound depend on the ambient dimension m?

$$\lim_{n\to\infty} \mathbb{E} \int_K |I(x) - f_n(x)|^2 \, dx \leq \frac{\alpha^2 \|f\|_\infty^2 \Omega^{2m} \pi^2 (2L+1)^2 \mathrm{vol}(K)^2 \|\rho\|_2^2}{n}.$$

RVFL on Manifolds

How does the bound depend on the ambient dimension m?

$$\lim_{n\to\infty} \mathbb{E} \int_K \left| I(x) - f_n(x) \right|^2 dx \leq \frac{\alpha^2 \|f\|_\infty^2 \Omega^{2m} \pi^2 (2L+1)^2 \mathrm{vol}(K)^2 \|\rho\|_2^2}{n}.$$

- The constant $C_{f,\rho,\alpha,\Omega,m}$ (and, hence, the number n of hidden nodes) scales with $\mathrm{vol}(K)^2$.

RVFL on Manifolds

How does the bound depend on the ambient dimension m?

$$\lim_{n\to\infty} \mathbb{E} \int_K |I(x) - f_n(x)|^2 \, dx \leq \frac{\alpha^2 \|f\|_\infty^2 \Omega^{2m} \pi^2 (2L+1)^2 \mathrm{vol}(K)^2 \|\rho\|_2^2}{n}.$$

- The constant $C_{f,\rho,\alpha,\Omega,m}$ (and, hence, the number n of hidden nodes) scales with $\mathrm{vol}(K)^2$.
- If $K = \mathrm{supp}(f)$ is full-dimensional in $\mathbb{R}^m$, $C_{f,\rho,\alpha,\Omega,m}$ (and, hence, n) is exponential in m.

RVFL on Manifolds

How does the bound depend on the ambient dimension m?

$$\lim_{n\to\infty} \mathbb{E} \int_K |I(x) - f_n(x)|^2 \, dx \leq \frac{\alpha^2 \|f\|_\infty^2 \Omega^{2m} \pi^2 (2L+1)^2 \mathrm{vol}(K)^2 \|\rho\|_2^2}{n}.$$

- The constant $C_{f,\rho,\alpha,\Omega,m}$ (and, hence, the number n of hidden nodes) scales with $\mathrm{vol}(K)^2$.
- If $K = \mathrm{supp}(f)$ is full-dimensional in $\mathbb{R}^m$, $C_{f,\rho,\alpha,\Omega,m}$ (and, hence, n) is exponential in m.

Question

Can we improve $C_{f,\rho,\alpha,\Omega,m}$ and the lower bound on n if $K = \mathrm{supp}(f)$ has a lower dimensional structure, e.g., lies on a d-dimensional manifold $\mathcal{M} \subset \mathbb{R}^m$?

Detour - Smooth, Compact Manifolds

Let $\mathcal{M} \subset \mathbb{R}^m$ be a smooth, compact, d-dimensional manifold with

- atlas $\{U_j, \phi_j\}_{j \in \mathcal{A}}$
- partition of unity $\{\eta_j\}_{j \in \mathcal{A}}$ s.t. $\sum_{j \in \mathcal{A}} \eta_j(x) = 1$ and $\mathrm{supp}(\eta_j) \subset U_j$.

Detour - Smooth, Compact Manifolds

Let $\mathcal{M} \subset \mathbb{R}^m$ be a smooth, compact, d-dimensional manifold with

- atlas $\{U_j, \phi_j\}_{j\in\mathcal{A}}$
- partition of unity $\{\eta_j\}_{j\in\mathcal{A}}$ s.t. $\sum_{j\in\mathcal{A}} \eta_j(x) = 1$ and $\mathrm{supp}(\eta_j) \subset U_j$.

Theorem

Any function $f : \mathcal{M} \to \mathbb{R}$ may be represented by a (compactly supported) partition of unity:

$$f(x) = \sum_{\{j\in\mathcal{A}:\ x\in U_j\}} (\hat{f}_j \circ \phi_j)(x)$$

$$\hat{f}_j(z) := \begin{cases} f(\phi_j^{-1}(z))\,\eta_j(\phi_j^{-1}(z)) & z \in \phi_j(U_j) \\ 0 & \textit{otherwise}, \end{cases}$$

so that $\hat{f}_j$ are supported on compact subsets $\phi_j(\mathrm{supp}(\eta_j))$ of $U_j \subset \mathbb{R}^d$.

RVFL on manifolds

To approximate $f : \mathcal{M} \to \mathbb{R}$ by lower dimensional RVFL:

Step 1: Approximate $\hat{f}_j$ by a RVFL on $\phi_j(\mathrm{supp}(\eta_j)) \subset \mathbb{R}^d$:

$$\hat{f}_j(z) \approx \hat{f}_{n_j}(z) = \sum_{k=1}^{n_j} v_k \rho(\langle w_k, z \rangle + b_k)$$

Step 2: Approximate f by summing RVFLs over $\mathcal{M}$:

$$f(x) \approx \sum_{\{j \in \mathcal{A} : \, x \in U_j\}} (\hat{f}_{n_j} \circ \phi_j)(x)$$

RVFL on manifolds

To approximate $f\colon \mathcal{M} \to \mathbb{R}$ by lower dimensional RVFL:

Step 1: Approximate $\hat{f}_j$ by a RVFL on $\phi_j(\mathrm{supp}(\eta_j)) \subset \mathbb{R}^d$:

$$\hat{f}_j(z) \approx \hat{f}_{n_j}(z) = \sum_{k=1}^{n_j} v_k \rho(\langle w_k, z\rangle + b_k)$$

Step 2: Approximate f by summing RVFLs over $\mathcal{M}$:

$$f(x) \approx \sum_{\{j\in\mathcal{A}\colon\, x\in U_j\}} (\hat{f}_{n_j} \circ \phi_j)(x)$$

Theorem (Needell, Nelson, Saab, S.)

Let $\varepsilon > 0$. For each $j \in J$ there exist a sequence of RVFL networks $\left\{\tilde{f}_{n_j}(\cdot) = \sum_{k=1}^{n_j} v_k^{(j)} \rho\left(\langle w_k^{(j)}, \cdot\rangle + b_k^{(j)}\right)\right\}_{n=1}^{\infty}$ such that

$$\lim_{\{n_j\}_{j\in J}\to\infty} \mathbb{E}\int_{\mathcal{M}} \left| f(x) \quad - \sum_{\{j\in J\colon\, x\in U_j\}} (\tilde{f}_{n_j} \circ \phi_j)(x)\right|^2 dx < \varepsilon.$$

RVFL on manifolds: non-asymptotic result

Theorem (Needell, Nelson, Saab, S.)

Let $\mathcal{M} \subset \mathbb{R}^N$ be a smooth, compact, d-dimensional manifold with atlas $\{U_j, \phi_j\}_{j \in J}$, $f \in C_c(\mathcal{M})$, $\varepsilon > 0$, and $\eta \in (0,1)$. There exists $\{\delta_j\}_{j \in J}$ such that if

$$n \gtrsim \frac{|J|\sqrt{\mathrm{vol}(\mathcal{M})}\log(|J|\eta^{-1}\mathcal{N}(\delta_j, \phi_j(U_j)))}{\varepsilon \log(1 + \frac{\varepsilon}{c_j^d |J| \sqrt{\mathrm{vol}(\mathcal{M})}\mathrm{vol}(\phi_j(U_j))^2})},$$

then for each $j \in J$ there exist RVFL networks $\tilde{f}_{n_j}(\cdot) = \sum_{k=1}^{n_j} v_k^{(j)} \rho(\langle w_k^{(j)}, \cdot \rangle + b_k^{(j)}$ such that, with probability at least $1 - \eta$,

$$\int_{\mathcal{M}} \left| f(x) - \sum_{\{j \in J :\ x \in U_j\}} (\tilde{f}_{n_j} \circ \phi_j)(x) \right|^2 dx < \varepsilon.$$

Note: (Shaham et. al , 2018) can choose $|J| \lesssim 2^d d \log(d) \mathrm{vol}(\mathcal{M}) \delta^{-d}$. Then the total number n of the hidden layer nodes has exponential dependence on d (instead of the m).

Numerical results

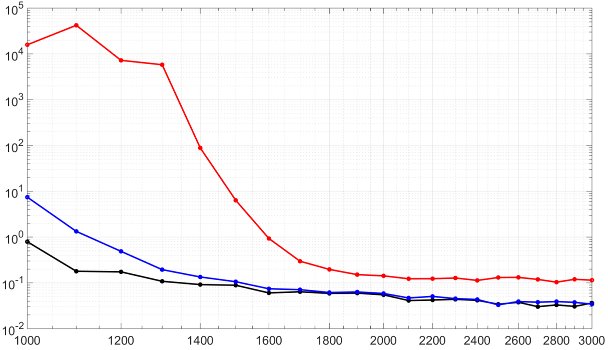


Figure: Log-scale plot of average relative RVFL error as a function of the number of nodes n in each RVFL. Geometric multiresolution analysis manifold approximations with resolution levels $\mathbf{j} = \mathbf{12}$, $j = 9$, and $j = 6$. For each j, reconstruction error decays as a function of n until reaching a floor due to error in the GMRA approximation of $\mathcal{M}$.

Numerical results

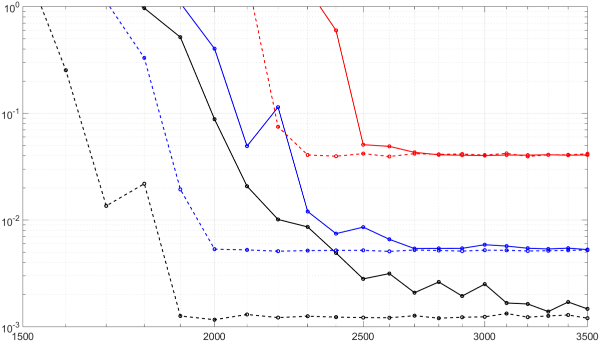


Figure: Log-scale plot of average relative RVFL error as a function of the number of nodes n in each RVFL. GMRA manifold approximations with resolution levels $\mathbf{j = 12}$, $j = 9$, and $j = 6$. For each j, we fix $\alpha = 2$ and vary $w = 10, 15$ (solid and dashed lines, resp.). Reconstruction error decays as a function of n until reaching a floor due to error in the GMRA approximation of $\mathcal{M}$.

Questions?

Thank you for your attention!