Controlled Differential Equations on Long Sequences via Non-standard Wavelets

Sourav Pal, Zhanpeng Zeng, Sathya N. Ravi, Vikas Singh
https://github.com/sourav-roni/BCR-DE

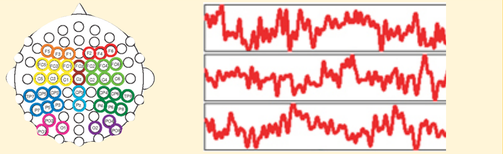
ICML 40 Years International Conference On Machine Learning

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON
UIC UNIVERSITY OF ILLINOIS CHICAGO

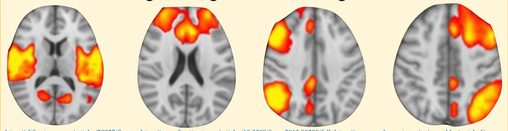
Overview

- We present an **efficient** way to model controlled differential equations via the lens of an **integral transform (operator)** and **multi-resolution analysis (MRA)**
- Proposed method enables the handling of **coupled differential equations** and yields competitive results on standard benchmarks.

Applications



Sequential measurements over long time intervals or continuous monitoring: blood pressure, EEG, fMRI, heart rate etc. Goal: Representation learning from such data for various downstream tasks including prediction, regression, generative modeling.



https://elifesciences.org/articles/28927/figures, https://www.frontiersin.org/articles/10.3389/fneur.2013.00200/full, https://consumer.huawei.com/en/wearables/watch-d/

NCDE & CZ Operators

- **Neural controlled differential equations (NCDE)**s are modeled using the **Riemann-Stieltjes** integral

$$z_t = z_0 + \int_0^t f_\theta(z_s)\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}s}(s)\mathrm{d}s = z_0 + \int_0^t f_\theta(z_s)\,\mathbf{X}'(s)\mathrm{d}s$$

where $\mathbf{X}$ is a continuous function of bounded variation, serving as the "**control**" and is usually assumed to be differentiable.

- Integral transform, $\mathscr{A}$ with associated kernel $a(t,s)$ acting on function f is denoted as:

$$\mathscr{A}(f)(t) = \int a(t,s)f(s)\mathrm{d}s$$

- **Calderon-Zygmund (CZ)** operators, involve a kernel which is smooth away from the diagonal, enabling their compression via **non-standard wavelet decomposition**.

$$|a(t,s)| \le \frac{1}{|t-s|}, \quad |\partial_t^M a(t,s)| + |\partial_s^M a(t,s)| \le \frac{C_0}{|t-s|^{1+M}}$$

- By thinking about transformation between function spaces, we can cast the NCDE as an **integral transform**

$$(h_\theta(z_s))^T \times X'(s) := v_s$$

$$u_t = \int_0^t a_\theta(t,s)v(s)\mathrm{d}s$$

where, $u_t = z_t - z_0$

BCR & Simplifications

- Solving the integral transform via **unrolling** along time dimension is equivalent to performing a matrix-vector product $\mathbf{u} = \mathbf{A}\mathbf{v}$.
- $\mathbf{u} \in \mathbb{R}^T, \mathbf{v} \in \mathbb{R}^T, \mathbf{A} \in \mathbb{R}^{T\times T}$, where T is sequence length, this has complexity $O(T^2)$, which is infeasible for long sequences.

- **Beylkin, Coifman and Rokhlin (BCR)** exploited non-standard wavelet decomposition to speed-up dense matrix-vector product.
- For an operator with kernel $a(t,s)$ and the non-standard 2D MRA involves:

$$\alpha_{km}^l = \iint \psi_k^l(t)a(t,s)\psi_m^l(s)\mathrm{d}t\mathrm{d}s \qquad \beta_{km}^l = \iint \psi_k^l(t)a(t,s)\phi_m^l(s)\mathrm{d}t\mathrm{d}s$$

$$\gamma_{km}^l = \iint \phi_k^l(t)a(t,s)\psi_m^l(s)\mathrm{d}t\mathrm{d}s \qquad A_{km}^l = \iint \phi_k^l(t)a(t,s)\phi_m^l(s)\mathrm{d}t\mathrm{d}s$$

where, ψ's and ϕ's are the scaled and translated copies of the scaling function and wavelet function at level l of decomposition respectively.

- Considering the forward and inverse wavelet transform to be denoted by $\mathscr{W}^T$ and $\mathscr{W}$ respectively, the 2D MRA of operator A can be written as

$$\mathscr{W}^T A^l \mathscr{W} = \begin{bmatrix} \alpha^{l+1} & \beta^{l+1} \\ \gamma^{l+1} & A^{l+1} \end{bmatrix}; A^l = \mathscr{W}\begin{bmatrix} \alpha^{l+1} & \beta^{l+1} \\ \gamma^{l+1} & A^{l+1} \end{bmatrix}\mathscr{W}^T$$

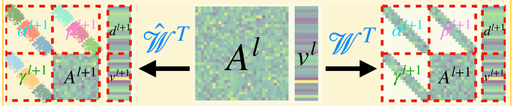
where, $\alpha^{l+1}, \beta^{l+1}, \gamma^{l+1}$ and A^{l+1} are all matrices.

- For certain class of operators, e.g. CZ, the non-standard decomposition leads to a diagonally banded structure.

$$|\alpha_{km}^l| + |\beta_{km}^l| + |\gamma_{km}^l| \le \frac{C_M}{1+|k-m|^{1+M}}$$

- Hence, the dense product is only at the coarsest level, via the recurrence

$$u^l = A^l v^l = \mathscr{W}\left(\begin{bmatrix} \alpha^{l+1} & \beta^{l+1} \\ \gamma^{l+1} & 0 \end{bmatrix}\begin{bmatrix} d^{l+1} \\ v^{l+1} \end{bmatrix} + \begin{bmatrix} 0 \\ u^{l+1} \end{bmatrix}\right)$$



- We propose **Partially Un-shared Convolution (PUC)** to parameterize the above, which helps in regularization. This layer performs convolution using filters which are shared for some length of the sequence before being unshared.

Evaluations

Physiological Measurements (Regression)

Model	RMSE			Time (hrs)		
	RR	HR	SpO_2	RR	HR	SpO_2
ODE-RNN (s512)	1.66 ± 0.06	6.75 ± 0.9	1.98 ± 0.31	0.0	0.1	0.1
NCDE (s1)	2.79 ± 0.04	9.82 ± 0.34	2.83 ± 0.27	23.8	22.1	28.1
NCDE (s512)	2.53 ± 0.03	12.22 ± 0.11	2.98 ± 0.04	0.1	0.0	0.1
NRDE (d3s8)	2.42 ± 0.19	7.67 ± 0.40	2.55 ± 0.13	2.9	3.2	3.1
NRDE (d3s128)	$\mathbf{1.51 \pm 0.08}$	$\mathbf{2.97 \pm 0.45}$	1.37 ± 0.22	0.5	1.7	1.7
NRDE (d3s512)	$\mathbf{1.49 \pm 0.08}$	3.46 ± 0.13	$\mathbf{1.29 \pm 0.15}$	0.3	0.4	0.4
BCR-DE	$\mathbf{1.53 \pm 0.09}$	$\mathbf{3.27 \pm 0.16}$	$\mathbf{1.18 \pm 0.15}$	0.4	0.5	0.9

- Performance of BCR-DE compared to baselines for regression to Respiratory Rate (RR), Heart Rate (HR) and oxygen saturation (SpO2) from PPG and ECG data (BIDMC32 dataset). BCR-DE achieves comparable performance relative to NRDE and improves on test set RMSE over alternatives significantly.

Eigenworms (Classification)

Model	Accuracy (%)	Time (hrs)
ODE-RNN (s128)	47.9 ± 5.3	0.01
NCDE (s4)	66.7 ± 11.8	5.5
NCDE (s128)	48.7 ± 2.6	0.1
NRDE (d2s4)	$\mathbf{83.8 \pm 3.0}$	2.4
NRDE (d3s128)	68.4 ± 8.2	0.1
BCR-DE	77.8 ± 1.2	0.01
BCR-DE (Noise)	78.7 ± 2.4	0.01

- BCR-DE achieves comparable performance to the best baseline of NRDE, with a runtime that is two orders of magnitude faster.
- The noise variant of BCR-DE where detail coefficients are perturbed achieves a better classification accuracy

Auto-encoding (Medium Length sequence)

Task	Dataset	NCDE		NRDE		BCR-DE	
		MSE	Time (hrs)	MSE	Time (hrs)	MSE	Time (hrs)
AE	PPG	**6.05e-5**	3.63	0.014	0.67	0.012	0.2
	ECG	**6.06e-5**	3.03	0.014	0.57	0.024	0.19
DAE	PPG	**0.008**	4.1	0.023	0.92	**0.009**	0.18
	ECG	**0.008**	3.04	0.023	0.73	0.02	0.18
MAE	PPG	0.28	2.23	0.106	5.47	**0.024**	0.22
	ECG	0.29	1.5	0.106	3.76	**0.097**	0.23

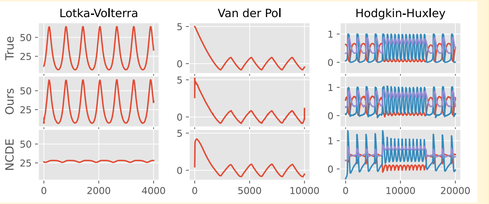
- For auto-encoding (reconstruction) (AE), denoising auto-encoding (DAE) and masked auto-encoding (MAE) on PPG and ECG data of sequence length 4K, BCR-DE achieves comparable or better performance (MSE), but needs a much lower runtime.

Coupled Differential Equations

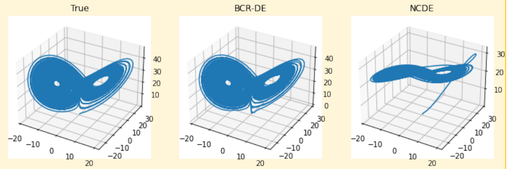
- Common dynamical systems often feature multiple dependent variables and one independent variable (e.g. time), often such systems are referred to as **coupled differential equations.**
- Since BCR-DE models the operator of an integral transform, it is able to model the systems described by coupled differential equations.

Setting (Seq Len)	NCDE		NRDE		BCR_DE	
	MSE	Time (hrs)	MSE	Time (hrs)	MSE	Time (hrs)
Toy Coupled DE (4k)	1e-4	0.62	**6e-5**	0.02	3e-4	0.009
Lotka-Volterra (4k)	377.9	43.74	365.4	3.04	**0.19**	0.134
Van der Pol (10k)	0.023	43.6	0.94	7.43	**1e-3**	0.34
Chaotic Lorenz (10k)	66.15	42.9	133.3	3.7	**0.05**	0.35
Hodgkin-Huxley (20k)	0.02	45.35	1.24	4.28	**4e-4**	0.35
Benzene Conc. (240)	250.3	3.64	725.4	1.17	**212.9**	0.046

- Table above shows the MSE on test set over the entire sequence length and training time. In almost all cases BCR-DE achieves the best performance in significantly less time.



- Comparing predicted trajectories for different coupled systems with simulated ground truth. BCR-DE matches the true trajectory almost perfectly.

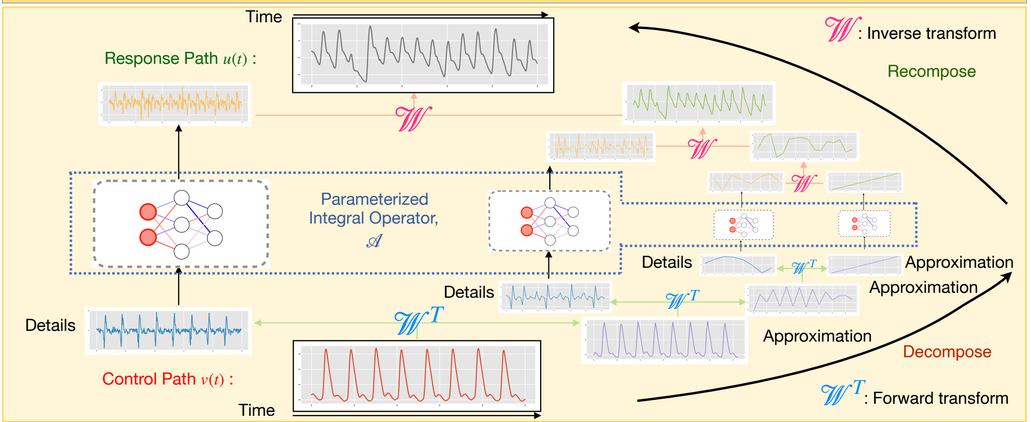


- Comparison of trajectories for chaotic Lorenz system. BCR-DE matches the ground truth nearly exactly.

Conclusions

- BCR-DE provides an **efficient** strategy to model **long** but *fixed* length sequences, by **unrolling** the dynamics along the length of the sequence
- Computational benefits are best achieved when sequence is long and number of levels of decomposition is large enough to have a small but effective coarse representation of the signal.

BCR-DE Framework



International Conference on Machine Learning (ICML), 2023

Research supported by Vilas Board of Trustees, NIH grant RF1 AG059312, UIC startup fund