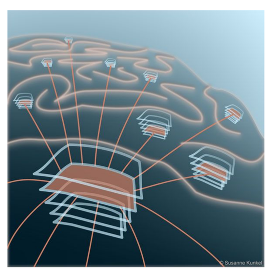
Unfolding recurrence by Green's functions for optimized reservoir computing



Sandra Nestler[1,2], Christian Keup[1,2], David Dahmen[1], Wiebke Bartolomaeus[2], Youness Boutaib[2], Matthieu Gilson[1,3], Holger Rauhut[2], Moritz Helias[1,2]

[1] Institute of Neuroscience and Medicine (INM-6), Forschungszentrum Jülich, Germany
[2] RWTH Aachen University, Germany
[3] Universitat Pompeu Fabra, Barcelona, Spain

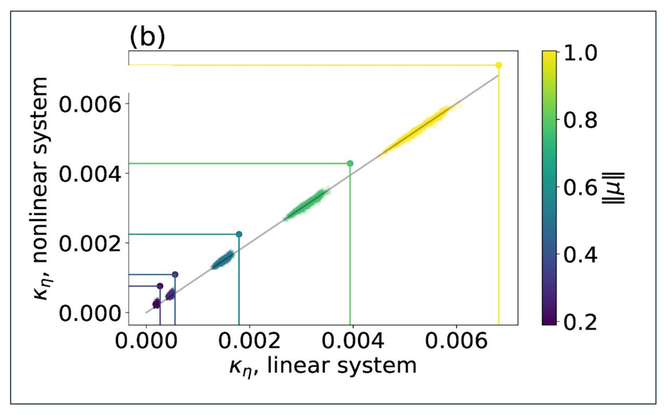
Published in: Advances in Neural Information Processing Systems 33 (2020), Advances in Continuous and Discrete Models 2022.1 (2022)

Setup

- Reservoir Computing as computationally efficient machine learning system[1,2]
- Task: Binary classification of one-dimensional, time-dependent stimuli
- Dependence of the performance on reservoir properties has already been studied[3,4]

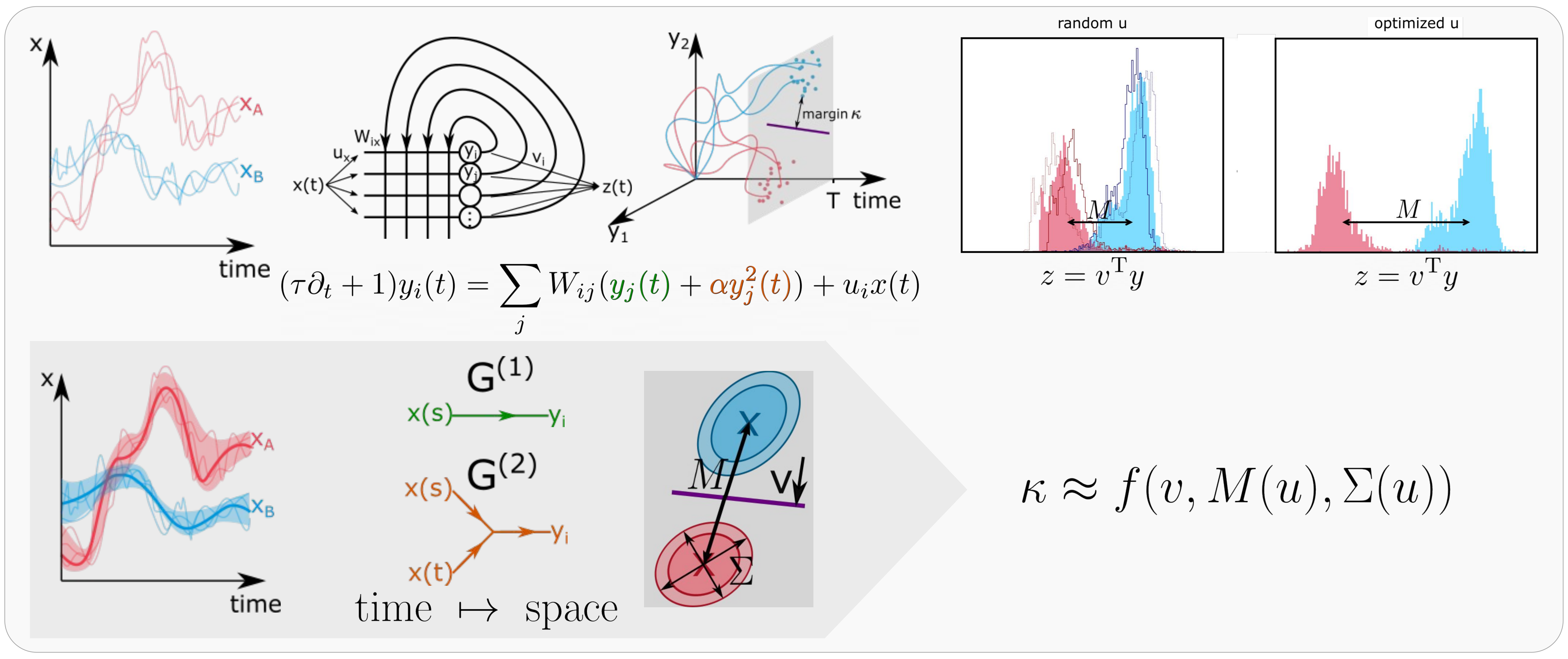
Non-linear vs. linear system

- Clear benefit compared to random u
- Gain from non-linearity varies with linear separability of stimuli
- Significant performance increase for low linear separabilities



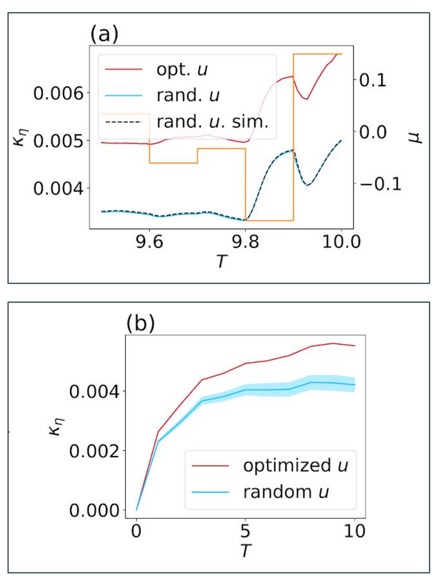
The soft margin

- Joint optimization of input and readout projections
- Classification quality measure: margin
  $\kappa(u,v) = \min_\nu(\zeta_\nu v^\mathrm{T} y^{u,\nu})$
- Differentiable and less sensitive to exact realizations of stimuli: soft margin
  $\kappa_\eta(u,v) = -\frac{1}{\eta}\ln\left[\sum_\nu \exp(-\eta\zeta_\nu v^\mathrm{T} y^{u,\nu})\right]$
- For large set of sample data: κ_η becomes cumulant generating function
  $\kappa_\eta(u,v) \approx v^\mathrm{T} M^u - \frac{1}{2}\eta v^\mathrm{T}\Sigma^u v$



Dynamics and optimization

- Non-linear dynamics can be approximated as perturbation series for small α
- M becomes sensitive to second order stimulus statistics, Σ becomes sensitive to fourth order
- For fixed reservoir, stimulus and readout time: considerable increase in classification performance



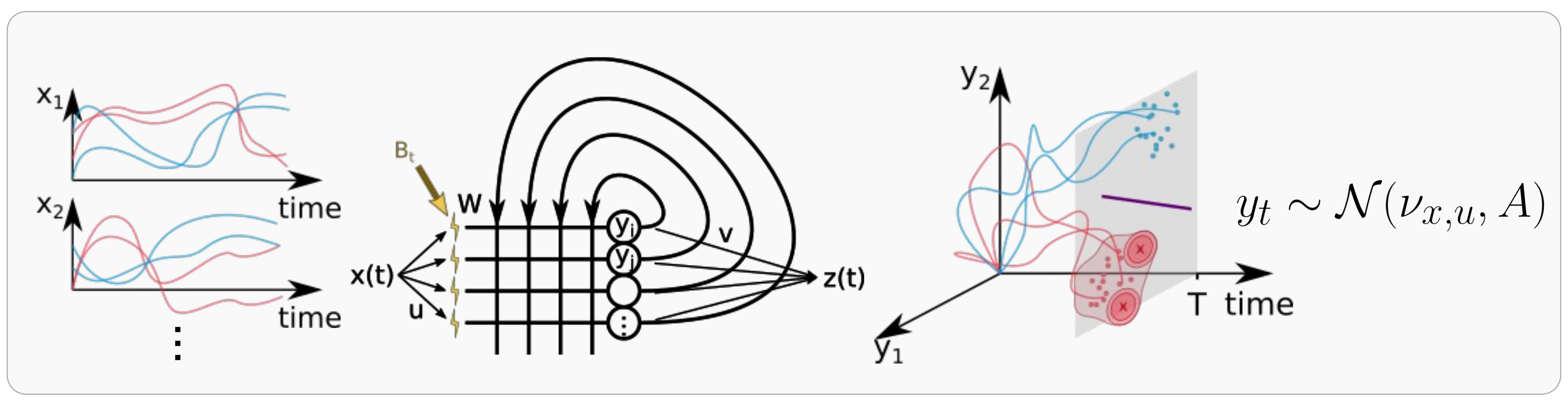
Conclusion

- Unfold recurrent dynamics via Green's functions
- Soft margin yields closed-form expressions for optimization
- Trade-off between separation and variability in readout direction
- Significant gain from non-linearity for weakly linear separable data
- Clear absolute performance gain also in linearly well separable ECG5000 dataset

Classification by stochastic linear RNN

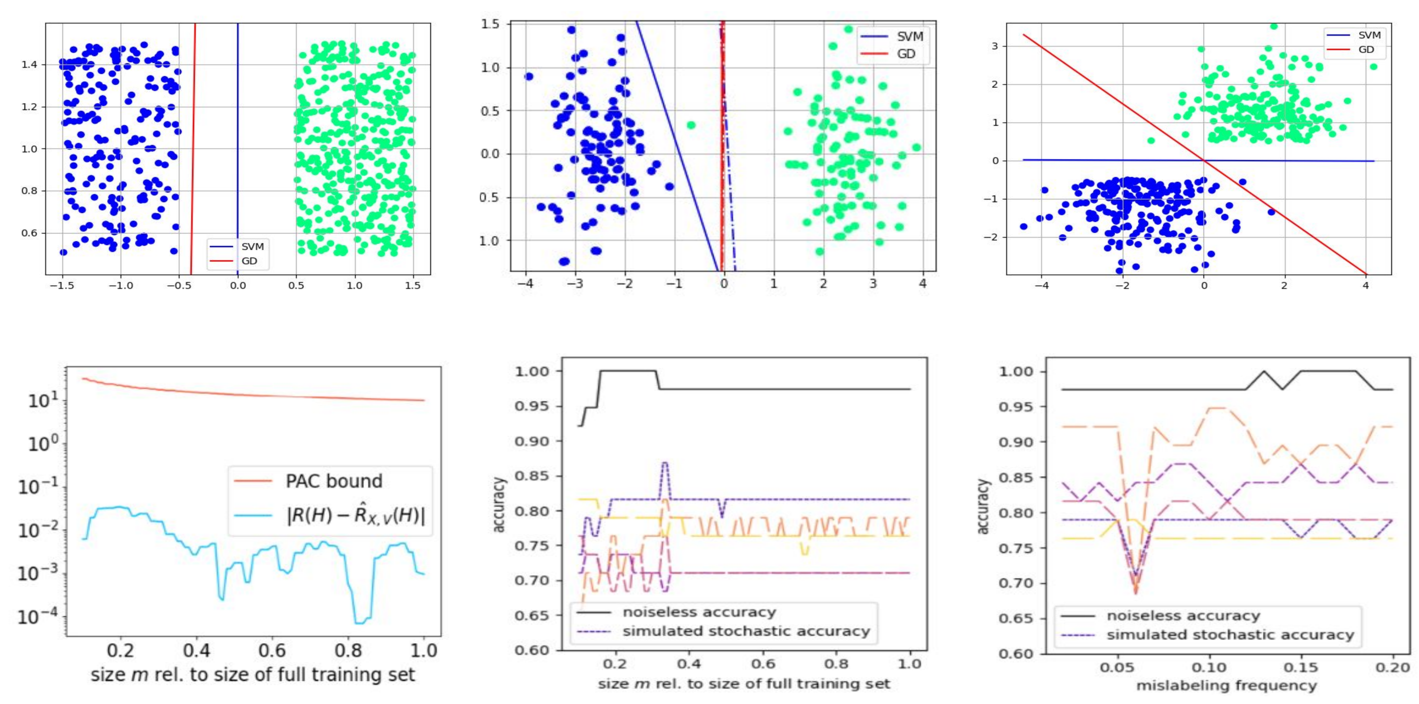
- Introduce noise to the dynamics to leverage sensitivity on exact realizations
- Minimize empirical risk of misclassification based on probability density

$$\min_{u,\omega,b} \widehat{R}_{(x,v)}(H) = \min_{u,\omega,b} \frac{1}{m}\sum_{i=1}^{m} \underbrace{\mathbb{P}_B(H(x_i) \neq v_i)}_{\mathbb{P}(\mathrm{sign}(\langle y_t, \omega\rangle + b) \neq v)} = \min_{u,\omega,b} \frac{1}{m}\sum_{i=1}^{m}\Phi\left(-v_i \cdot \frac{\langle \nu_{x_i,u}, \omega\rangle + b}{\sqrt{\omega^T A \omega}}\right)$$



Optimization strategy

- Use Rademacher complexity to guarantee empirical risk convergence
- Contrast empirical risk for noisy and noise-free dynamics
- The optimal classifier is robust to outliers and perturbations



Analytically unfolding recurrent dynamics into Green's functions is a versatile approach that may be used as a general purpose scheme to analyze recurrent networks.

References

[1] Jäger, H. (2001) Tech. Rep. GMD Report 148
[2] Maass, H., Natschläger, T. (2002) Neural Computation, 14.11, 2531-2560.
[3] Bertschinger N, Natschläger T, Legenstein R. (2004) NeurIPS 17
[4] Toyoizumi T, Abbott L. (2004) Phys. Rev. E., 84, 051908
[5] Chen, Yanping, et al. (2015) The ucr time series classification archive.

Links to the papers: