Optimized Reservoir Computing with Stochastic Recurrent Networks

Sandra Nestler[1,2], Christian Keup[1], David Dahmen[1], Moritz Helias[1,3]

[1] Institute of Neuroscience and Medicine (INM-6), Jülich Research Centre, Jülich, Germany
[2] Mathematics of Information Processing, RWTH Aachen University, Aachen, Germany
[3] Department of Physics, RWTH Aachen University, Germany

Contact: s.nestler@fz-juelich.de

Binary Classification Task

- Reservoir Computing as computationally efficient machine learning system [1, 2]
- The dependence of the performance on reservoir properties has been studied [3, 4]
- We aim at a joint optimization of input and output projection

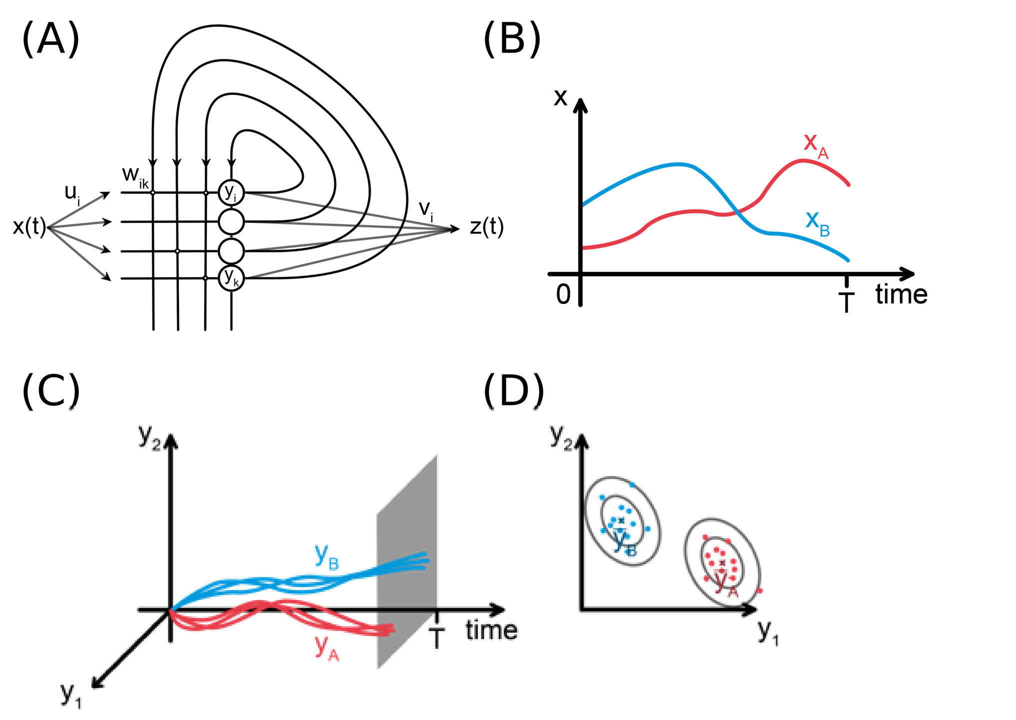


Figure 1 Binary classification with Reservoir Computing. (A) The Reservoir Computing scheme comprises the projection of inputs onto recurrently connected neurons and a linear readout. Left: Projection of time-dependent input $x(t)$ via afferent connections u. Middle: Neural network with random recurrent connectivity W (reservoir). Right: One-dimensional linear readout of high-dimensional network activity via efferent connections v. (B) Time-dependent stimuli represent classes. (C) The trajectories of the same class are similar. (D) At some readout time, the states of each class are approximately Gaussian distributed.

Linear Dynamics

- Linear dynamics with neuron-internal noise

$$(\tau\partial_t + 1)y_i(t) = \sum_j W_{ij}y_j + u_i x(t) + \xi_i(t)$$

- Analytically solvable by decomposition into eigenmodes r^α, l^α
- Green's function acts as propagator from input to output
- Covariance matrix depends only on noise amplitude

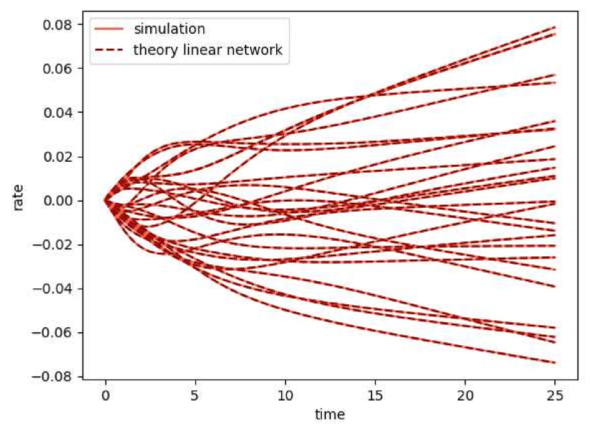


Figure 2 Examplary neural response. The analytical solution of the network dynamics is compared with a numerical simulation using NEST[5] for some arbitrary input.

$$\langle y_i(t)\rangle = \sum_j \int_0^t \mathrm{d}t'\, G^{(1)}_{ij}(t-t')\, u_j x(t')$$

$$G^{(1)}_{ij}(t-t') = \frac{1}{\tau}\sum_\alpha \exp\Big(-\lambda_\alpha \frac{t-t'}{\tau}\Big)\, r^\alpha_i l^\alpha_j \qquad j,t' \xrightarrow{G^{(1)}} i,t$$

$$\Sigma_{ij}(t) = \sum_{kl}\int_0^t \mathrm{d}t' \int_0^t \mathrm{d}t''\, G^{(1)}_{ik}(t-t')G^{(1)}_{jl}(s-t'')\langle \xi_k(t')\xi_l(t'')\rangle$$

- Analytical expressions enable systematic optimizations
- Approximate distributions of states as Gaussian

The Surprise

- States of classes form clouds
- For classification, consider classes as Gaussian distributed
- The optimal readout vector then follows as

$$v = \Sigma^{-1}d$$

- Readout is robust against outliers
- Measure quality of the classification: surprise

$$S = \frac{1}{2}d^{\mathrm{T}}\Sigma^{-1}d$$

- Monotonically connected to misclassification probability

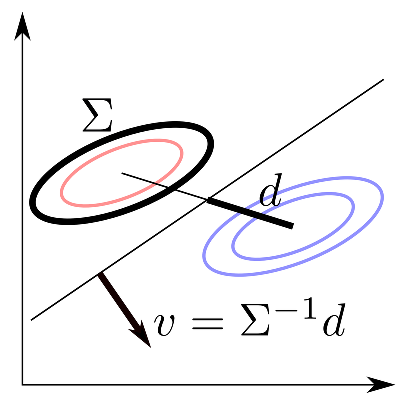


Figure 3 Finding the optimal readout vector. The covariance matrix Σ and the distance vector d characterize the separability of classes. The normal vector of the decision plane is then $v = \Sigma^{-1}d$.

Optimization in the Linear System

- Analytical results for d and Σ allow optimization of S
- Surprise assumes optimal readout, corresponds to joint optimization of u and v
- S can be expressed as quadratic form in the input projection u

$$S = u^{\mathrm{T}}\mathcal{M}u$$

- Maximized by eigenvector to largest eigenvalue of $\mathcal{M}$
- For fixed reservoir, stimulus and readout time, a large increase in classification performance can be achieved
- The surprise exhibits a sensitive angle dependence on the input projection

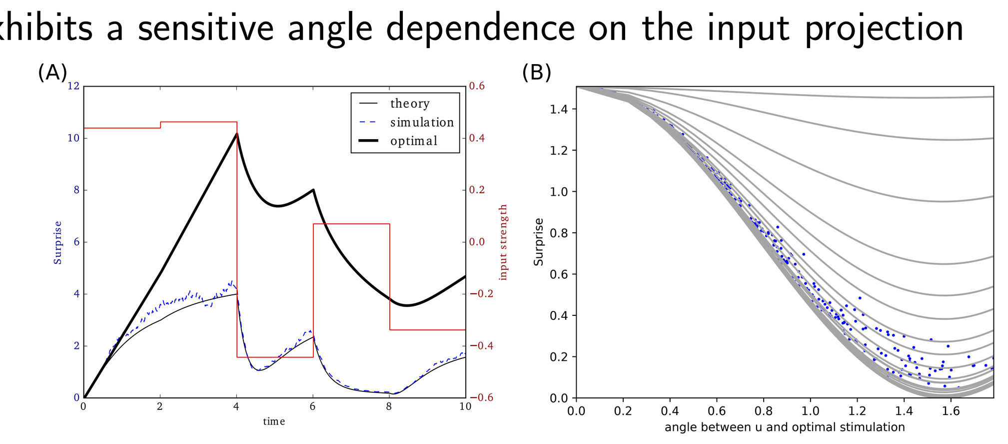


Figure 4 Optimal surprise in response to a stepwise constant stimulus. (A) The surprise reachable with optimal input projection for each readout time is compared to an arbitrary projection. (B) The input projection is chosen from variation of the angle between the optimal input direction and a random direction. The variability of the surprise follows from the distribution of eigenvalues of $\mathcal{M}$. Although separabilities similar to the optimal one can be reached, it is unlikely to achieve a similar classification performance using random input projections.

Readout in the Nonlinear System

- Introduce a small quadratic nonlinearity to mimic transfer function of realistic neurons
- Treat as small perturbation of linear dynamics
- Potential for improved separation by nonlinear dynamics
- Higher order cumulants of stimuli influence the mean and covariance of network states
- The effect of perturbations on the neural trajectory becomes dependent on the state in phase space
- Covariances become class-dependent for neuron-internal noise
- Consider nonlinear system with deterministic units and class-specific noise

$$(\tau\partial_t + 1)y_i(t) = \sum_j W_{ij}(y_j(t) + \frac{\alpha}{2}y_j^2(t)) + u_i x(t)$$

- Analytical connection between stimuli and network states required for optimzation of the input projection
- An approximation of the dynamics can in general be achieved from field theoretical methods [6]
- First order correction given by diagrams:

0th order: $x \longrightarrow y$ $\quad$ $G^{(1)}x$

$+$

1st order: $x, x \longrightarrow y$ $\quad$ $xG^{(2)}x$

$= y$

The Soft Margin

- Surprise not applicable for class-dependent covariance
- Optimize distances to decision plane instead of misclassification probability

$$\kappa = \min_\nu(\zeta_\nu(v^T y^\nu - \theta))$$

- Soft margin replaces margin by differentiable approximation

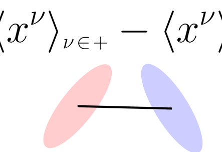
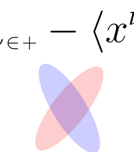
$$O_\eta(v,\theta) = -\frac{1}{\eta}\ln\sum_\nu \exp(-\eta\zeta_\nu(v^T y^\nu - \theta))$$

- The dependence on the threshold θ can be addressed with appropriate shift of coordinates
- The gradient can be calculated to desired degree of complexity of the network state distribution using a cumulant expansion

$$\frac{\partial}{\partial v_i}O_\eta(v) = \langle \zeta_\nu y^\nu\rangle - \eta\sum_j \Sigma_{ij}v_j$$

- Σ denotes the second order cumulant of $\zeta_\nu y^\nu$ and can be understood as the covariance of all points, irrespective of the class
- In Gaussian approximation, a stable fix point can be found for invertible Σ that is equivalent to the surprise
- Separation of classes is composed of linear and nonlinear separability of the input

$$\langle \zeta_\nu y^\nu\rangle = G^{(1)}(\langle x^\nu\rangle_{\nu\in+} - \langle x^\nu\rangle_{\nu\in-}) + G^{(2)}(\langle x^\nu x^{\nu\mathrm{T}}\rangle_{\nu\in+} - \langle x^\nu x^{\nu\mathrm{T}}\rangle_{\nu\in-})$$

- Optimizaton of the soft margin nearly maximizes the margin
- Analytic connection between input and classification performance as in the linear system

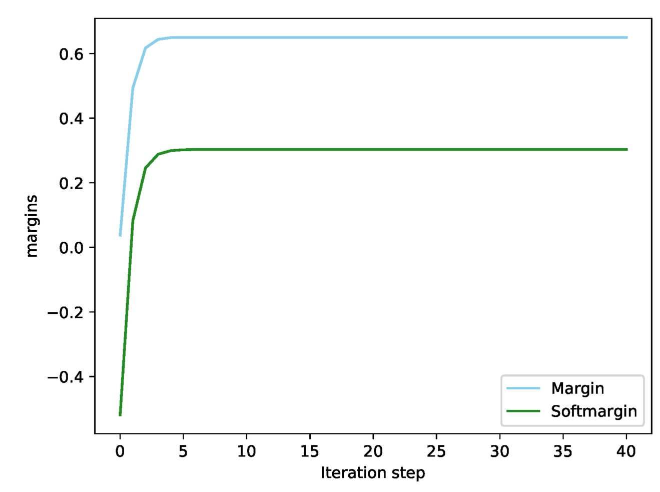


Figure 6 Gradient descent of the soft margin. The response of a nonlinear reservoir to a delta peak stimulus is classified using the gradient descent on the soft margin. The margin for the readout vector found in every optimization step resembles the soft margin, shifted by an offset.

Discussion

- Classification performance increases in linear system by joint optimization of input and output
- Readout specified by direct calculation rather than training algorithm
- Readout method robust against outliers
- Analogous method for nonlinear system
- Benefit from optimization of the input projection still unanswered in the nonlinear system

References

[1] Jäger, H. (2001) The echo state approach to analysing and training recurrent neural networks. Tech. Rep. GMD Report 148

[2] Maass, H., Natschläger, T. (2002) Real-time computing without stable states: A new framework for neural computation based on perturbations. Neural Computation, 14.11, 2531-2560.

[3] Bertschinger N, Natschläger T, Legenstein R. (2004) At the Edge of Chaos: Real-time Computations and Self-Organized Criticality in Recurrent Neural Networks. Conference: Advances in Neural Information Processing Systems 17 NIPS

[4] Toyoizumi T, Abbott L. (2004) Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime. Phys. Rev. E., 84, 051908

[5] Peyser, Alexander et al. (2017). NEST 2.14.0. Zenodo. 10.5281/zenodo.882971

[6] Helias M, Dahmen D. (2019) Statistical field theory for neural networks, arXiv:1901.10416

Acknowledgments: Partially supported by HGF young investigator's group VH-NG-1028 and European Union Horizon 2020 grant 785907 (Human Brain Project SGA2).