Learning Spectral Regularizations for Linear Inverse Problems

Hartmut Bauermeister,
Martin Burger,
Michael Moeller

UNIVERSITÄT SIEGEN
FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

NEURAL INFORMATION PROCESSING SYSTEMS

Inverse Problem setting

We consider a compact linear operator between infinite-dimensional Hilbert spaces and try to reconstruct data from noisy observations.

- Compact linear operator:

$$A\colon X \to Y$$

- Noisy observation:

$$y^\delta = Ax + n^\delta$$

- Singular value decomposition:

$$Ax = \sum_{n=1}^{\infty} \sigma_n \langle x, u_n \rangle v_n$$

- Pseudoinverse:

$$A^\dagger y = \sum_{n=1}^{\infty} \frac{1}{\sigma_n} \langle y, v_n \rangle u_n$$

Problem: 0 is an accumulation point of the singular values and the pseudoinverse is hence discontinuous!

Spectral regularization

To circumvent the problem of discontinuity, the pseudoinverse is usually replaced by a continuous regularization operator:

$$R_\alpha y = \sum_{n=1}^{\infty} g_\alpha(\sigma_n) \langle y, v_n \rangle u_n$$

For the noise level going to 0, we want the surrogate to resemble the exact pseudoinverse:

$$\|R_\alpha y^\delta - A^\dagger y\| \to 0, \text{ as } \delta \to 0$$

In this work, we investigate a parametrization of a learnable regularization of the singular values:

$$g_\alpha(\sigma) = \mathcal{N}(\sigma, \delta, y^\delta; \theta)$$

Advantage: For proper choices of the network architecture, we can maintain guarantees for converging to the pseudoinverse and at the same time exploit the problem specific data prior modeled by the neural network.

Choices of learnable models with guarantees

A-priori approach

For guaranteeing the convergence to the pseudoinverse for the noise level going to zero, we incorporate a network as part of classical, provable regularization approaches. In our work we consider models inspired by Lavrentiev and Tikhonov regularization:

$$\mathcal{N}_{\text{Lav}}(\sigma, \delta; \theta) = \frac{1}{\sigma + \delta^p \tilde{N}(\sigma, \delta; \theta)},$$

$$\mathcal{N}_{\text{Tik}}(\sigma, \delta; \theta) = \frac{\sigma}{\sigma^2 + \delta^q \tilde{N}(\sigma, \delta; \theta)}$$

In order to obtain provable guarantees, we restrict the network architecture as follows:

$$\tilde{N}(\sigma, \delta; \theta) = \theta_{\text{scale}} \cdot \text{sigmoid}(\text{FCN}(\sigma, \delta; \theta))$$

FCN in our experiments is chosen to be a 2-layer fully-connected network.

A-posteriori approach

Instead of casting a learnable model directly into a classical regularization approach, we also consider taking the output of an unrestricted network and adapting the corresponding singular values afterwards. The network G tries to find the data directly from the corresponding noisy observation itself. For a network G we can calculate the resulting 'singular values' of the corresponding regularization operator in the following sense:

$$g_\alpha(\sigma_n) = \frac{\langle u_n, \mathcal{G}(y^\delta; \theta) \rangle}{\langle v_n, y^\delta \rangle}$$

Harnessing technical convergence conditions, a provable regularization scheme is obtained by performing a projection of the singular values:

$$\mathcal{N}(\sigma_n, \delta, y^\delta; \theta) = \text{proj}_{[a,b]}\left(\frac{\langle u_n, \mathcal{G}(y^\delta; \theta) \rangle}{\langle v_n, y^\delta \rangle} \right)$$

In particular, the projection interval is chosen as

$$[a, b] = \left[(1 - \sqrt{\delta}\theta_l) \frac{\sigma}{\sigma^2 + \alpha_l \delta}, (1 + \sqrt{\delta}\theta_u) \frac{\sigma}{\sigma^2 + \alpha_u \delta} \right].$$

Numerical experiments

Investigated problems

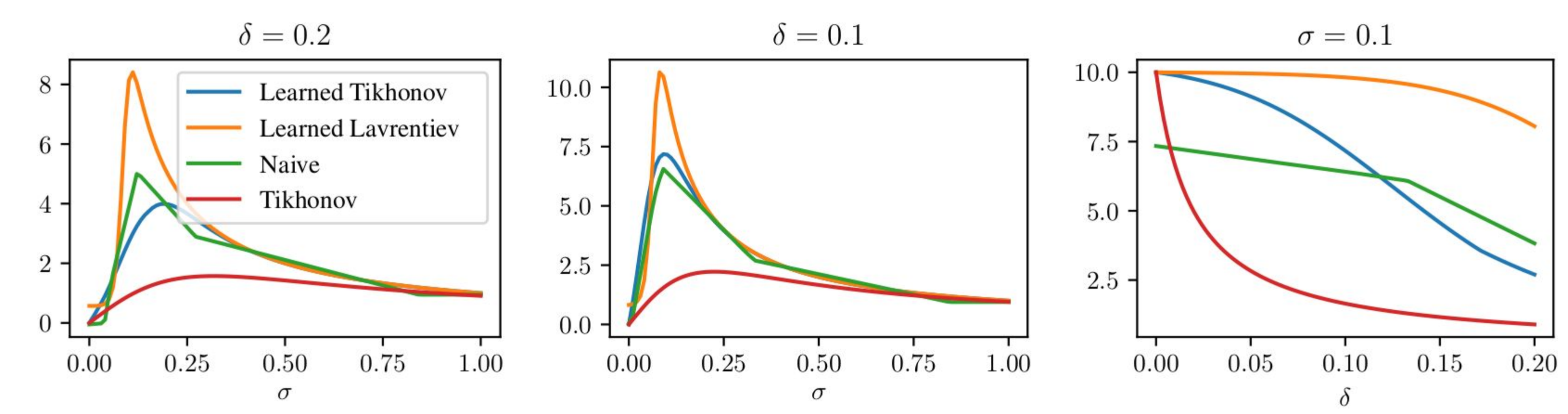
We consider the two inverse problems of differentiating and deblurring a function:

$$A_{\text{int}} x(t) = \int_0^t x(s)\, ds,$$

$$A_{\text{blur}} x(t) = \int_0^1 g(s - t) x(s)\, ds$$

Observations

The figure below shows the behaviour of g for the different investigated methods. The naive approach simply learns a mapping of the singular values without provable guarantees. Note how all learnable approaches resemble a Tikhonov type regularization, but more freedom of the curves shape. The right plot shows for a fixed singular value the behaviour for the noise level converging to zero. While the proposed approaches provably converge to the pseudoinverse behaviour, the naive approach fails to do so.



Exemplifying the results of the a-priori parameter choice rules. Left and middle: singular value regularization g as a function of the the singular values for fixed noise levels. Regularization for a fixed singular value as a function of the noise level

The table below shows the overall performance of all investigated approaches on both tasks. The proposed learnable regularizations clearly outperform the classical methods while remaining provably convergent for the noise level approaching zero.

		Naive	Lav.	Tik.	Learned Lav.	Learned Tik.	A-Post.
Deblur	training	29.58	20.05	27.17	28.59	29.73	31.89
	test	29.56	19.98	27.00	28.55	29.67	31.51
Diff.	training	28.93	21.35	26.44	28.94	29.27	31.63
	test	29.00	21.30	26.45	28.99	29.31	30.74

PSNR values during training and testing for deblurring and differentiation for various different regularization strategies.

Conclusion

We studied learnable regularization methods and showed that they can outperform their classical counterparts while still featuring provable convergence guarantees. An extension of this analysis to more general notions of distances, e.g. Bregman distances is a promising outlook for future work.