Learning parametrised regularisation functions via quotient minimisation

Martin Benning[1]*, **Guy Gilboa**[2], and **Carola-Bibiane Schönlieb**[1]

[1] Department of Applied Mathematics & Theoretical Physics (DAMTP), Centre for Mathematical Sciences (CMS), Wilberforce Road, Cambridge, CB3 0WA, United Kingdom

[2] Electrical Engineering Department, Technion - IIT, Technion City, Haifa 32000, Israel

We propose a novel strategy for the computation of adaptive regularisation functions. The general strategy consists of minimising the ratio of a parametrised regularisation function; the numerator contains the regulariser with a desirable training signal as its argument, whereas the denominator contains the same regulariser but with its argument being a training signal one wants to avoid. The rationale behind this is to adapt parametric regularisations to given training data that contain both wanted and unwanted outcomes. We discuss the numerical implementation of this minimisation problem for a specific parametrisation, and present preliminary numerical results which demonstrate that this approach is able to recover total variation as well as second-order total variation regularisation from suitable training data.

1 Introduction

Variational regularisation methods are powerful tools for the (approximate) solution of ill-posed inverse problems. However, even many of the popular methods, like total variation regularisation [19] or the (weighted) one-norm of wavelet coefficients [9], are not tailored to specific applications. They represent rather generic approaches that exploit certain structures of the desired signals, such as sparsity of the edge-sets or compressibility with respect to a certain basis.

Optimising parametrised regularisation functions represents a systematic approach for creating adaptive regularisations tailored to specific applications. The works of [6–8, 13, 15, 18] have gained considerable attention and brought several achievements to the field of parameter learning, amongst numerous other publications.

In this work we are going to look into a novel idea of constructing adaptive regularisers by minimising the ratio of parametrised regularisation functions. This idea is inspired by recent work on the concept of generalised singular vectors of convex regularisation functions [3] and the connected concept of the spectral total variation transform [5, 10, 16]. We are going to explain the general idea of the quotient minimisation and its motivation in Section 2. In Section 3 we are going to discuss a numerical implementation of the model for a given parametrisation of the regularisation function. Finally, we will present numerical results in Section 4 and conclude this work in Section 5.

2 Minimising quotients of regularisers

We propose to minimise the quotient of the same regularisation function J for different inputs $u^+ \in \mathbb{R}^m$ and $u^- \in \mathbb{R}^m$ with respect to a parametrisation h of the regularisation functional J, i.e.

$$\hat{h} \in \underset{\|h\|_2=1}{\arg\min} \frac{J(u^+; h)}{J(u^-; h)} \, . \tag{1}$$

The rationale behind (1) is that the regularisation function J with optimal parametrisation $\hat{h}$ will be very small for signals similar to u^+, whilst very large for signals u^-. In the context of variational regularisation - like Tikhonov-type regularisation with quadratic data fidelities and regularisation functions J - this implies that u^+ will be a favourable solution compared to u^-.

In order to compute a minimiser of (1) we need to pick a specific parametrisation first. Throughout this work we choose J to be absolutely one-homogeneous, and focus on models of the form

$$J(u; h) := \|u * h\|_1 \, , \tag{2}$$

with $*$ denoting the discrete convolution operator with Dirichlet boundary conditions, $\| \cdot \|_1$ being the discrete one norm, and $h \in \mathbb{R}^n$ the discrete convolution kernel. In this setup we aim at learning a convolution kernel that makes u^+ sparse, but u^- dense. Note that we can (and will) restrict the support of h in order to reconstruct fewer parameters than given data variables.

* Corresponding author: e-mail mb941@cam.ac.uk, phone +44 1223 760449

3 Numerical computation

With the specific parametrisation of J in (2), solving (1) requires the numerical solution of a generalised Eigenfunction problem. Based on the inverse power method by Hein and Bühler in [14], we propose the following variant:

$$h^{k+\frac{1}{2}} = \underset{\sum_{j=1}^{n} h_j = 0}{\arg\min} \left\{ J(u^+; h) - \mu^k \langle h, p^k \rangle \right\} , \tag{3a}$$

$$h^{k+1} = \frac{h^{k+\frac{1}{2}}}{\|h^{k+\frac{1}{2}}\|_2} , \tag{3b}$$

$$p^{k+1} \in \partial J(u^-, h^{k+1}) , \tag{3c}$$

$$\mu^{k+1} = \frac{J(u^+, h^{k+1})}{J(u^-, h^{k+1})} . \tag{3d}$$

Here ∂J denotes the subdifferential of J. We want to emphasise that the subdifferential is usually multi-valued, so that (3c) does not have a unique solution. However, if we choose u^- such that $u^- * h^k \neq 0$ for all its entries and all k, (3c) will have a single-valued explicit solution. This is typically true for signals u^- that represent merely noise.

Note that in contrast to the original formulation in [14] (and similar to [4]), we include an additional zero-mean constraint for the filter in (3a), in order to reconstruct filters that do not change the mean of the underlying signals u^+ and u^-. In addition, we do not incorporate the normalisation into (3a) but rather perform it as a separate step (3b). This simplifies the convergence analysis and seems more stable from an experimental point of view. Starting with the update for $h^{k+\frac{1}{2}}$, we observe

$$J(u^+; h^{k+\frac{1}{2}}) - \mu^k \langle h^{k+\frac{1}{2}}, p^k \rangle \leq J(u^+; h^k) - \mu^k \langle h^k, p^k \rangle = J(u^+; h^k) - \mu^k J(u^-, h^k) = 0 .$$

Furthermore, if $J(u^+; h^{k+\frac{1}{2}}) - \mu^k \langle h^{k+\frac{1}{2}}, p^k \rangle = 0$, we can already conclude $h^{k+\frac{1}{2}} = h^k$. Hence, we have $J(u^+; h^{k+\frac{1}{2}}) < \mu^k \langle h^{k+\frac{1}{2}}, p^k \rangle$ for $h^{k+\frac{1}{2}} \neq h^k$. From the one-homogeneity we further observe $J(u^+; h^{k+\frac{1}{2}}) < \mu^k \langle h^{k+\frac{1}{2}}, p^k \rangle \leq \mu^k J(u^-; h^{k+\frac{1}{2}})$, which yields

$$\mu^{k+1} = \frac{J(u^+; h^{k+1})}{J(u^-; h^{k+1})} = \frac{J(u^+; h^{k+\frac{1}{2}})}{J(u^-; h^{k+\frac{1}{2}})} < \frac{J(u^+; h^k)}{J(u^-; h^k)} = \mu^k .$$

This is basically the descent result of [14, Lemma 3.1]. Moreover can it be shown that the sequence converges globally to a local solution of (1), according to [14, Theorem 3.1].

4 Results

In the following, we are going to show one- and a two-dimensional numerical examples to demonstrate the overall ability to compute adaptive filters numerically via the proposed framework. We further demonstrate that we can rediscover known regularisation methods (like the total variation regularisation) for suitable choices of u^+ and u^-.

4.1 The one-dimensional case

For the first result we consider the function u^+ to be a singular vector of the one-dimensional total variation (see [3]), which is displayed in Figure 1 a). For u^- we simply choose a vector of Gaußian-distributed random variables with standard deviation $\sigma \approx 3{\times}10^{-1}$, shown in Figure 1 b). Now we compute a one-dimensional convolution filter $\hat{h} \in \mathbb{R}^{n \times 1}$, for $n = 5$, via algorithm (3). We initialise the algorithm with a random initialisation h^0 that is normalised and has mean zero. We then iterate (3) until the iterated Eigenvalues violate $|\mu^{k+1} - \mu^k| < \varepsilon \mu^k$, for $\varepsilon = 10^{-5}$. Due to the non-convexity of the problem, we repeat the procedure 100 times and pick $\hat{h}$ with smallest ratio $J(u^+; \hat{h})/J(u^-; \hat{h})$. In order to solve (3a), we use CVX [11, 12] with Mosek [1] as a solver. The result $\hat{h}$ of this procedure is visualised in Figure 1 c). Despite having set $n = 5$ we observe that the computed filter is a two-point finite difference approximation of the derivative, verifying that the solution of (1) has led to a meaningful convolution kernel $\hat{h}$. We apply the same procedure with u^+ being a singular vector of the second-order total variation (see [2] for an example) as visualised in Figure 1 d). We choose u^- to be pure noise again (Figure 1 e)) and proceed as in the previous example. It can be seen in Figure 1 f) that the reconstructed kernel $\hat{h}$ resembles a central finite difference approximation of the second derivative, such that (2) mimics the second-order total variation in this case.

In order to verify the behaviour of the reconstructed filters, we compute a Morozov regularised solution [17] of a noisy version f of u^+ via

$$\hat{u} = \underset{u \in \mathbb{R}^m}{\arg\min}\, J(u; \hat{h}) \qquad \text{subject to} \qquad \|u - f\|_2 \leq \eta \sigma \sqrt{m} , \tag{4}$$

with m denoting the number of elements of u^+ and f, respectively. Numerically, we also compute (4) via CVX, with the same settings as before. Figure 2 a) displays the comparison between u^+, f and $\hat{u}$ in case of $\hat{h}$ given as in Figure 1 c), for $\mu = 1.2$. Figure 2 b) shows the same result for $\hat{h}$ as given in Figure 1 f).

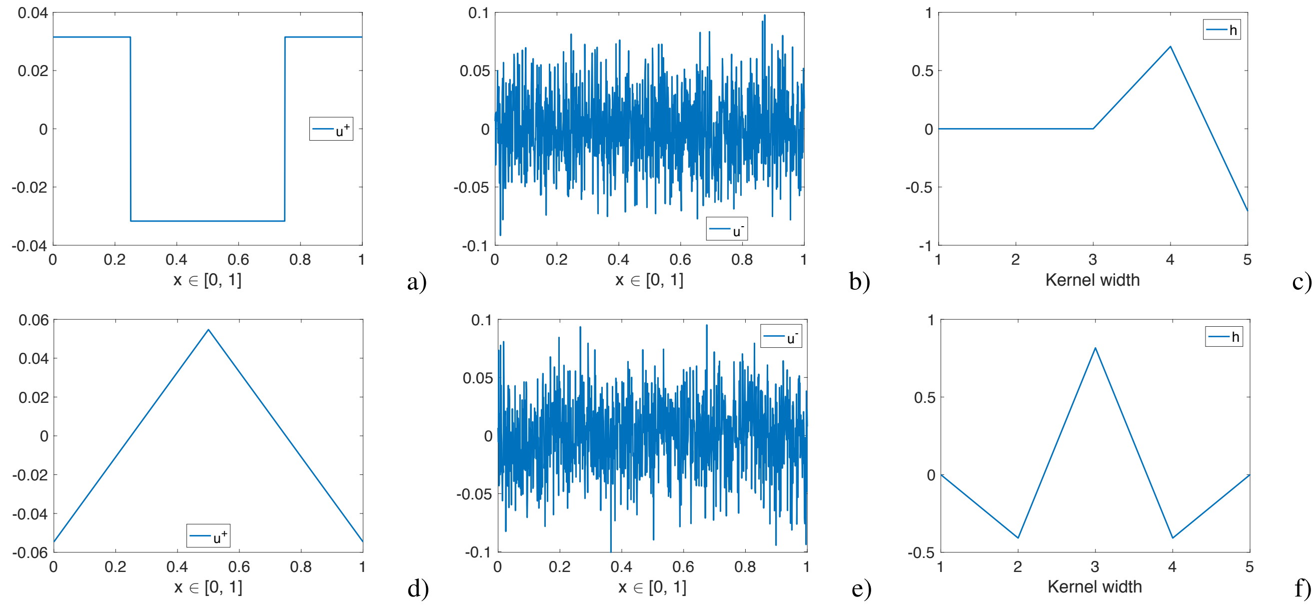

Fig. 1: Figure 1 a) shows u^+, a one-dimensional singular vector of the total variation. Figure 1 b) shows u^-, being a vector containing pure noise. Figure 1 c) shows the corresponding result of (1) with J given as in (2). Figure 1 d)-f) show the same as Figure 1 a)-c), but for u^+ being a singular vector of the second-order total variation.

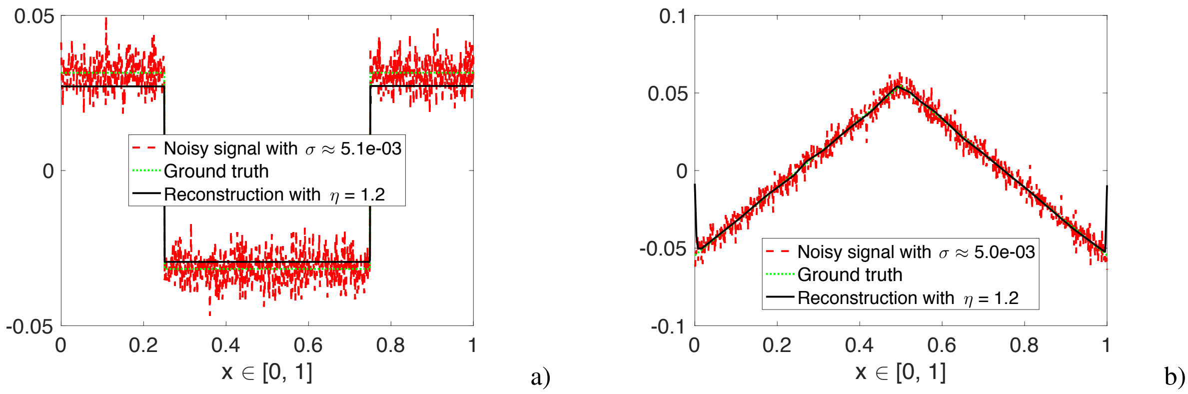

Fig. 2: Figure 2 a) compares the signal u^+ and its noisy version f with the Morozov regularised solution $\hat{u}$ of (4). The latter is computed with the filter $\hat{h}$ from Figure 2 c) and $\mu = 1.2$. Figure 2 b) is similar to Figure 2 a), but here the filter $\hat{h}$ from Figure 1 f) is being used for the Morozov regularisation.

4.2 The two-dimensional case

For the two-dimensional case we consider the Shepp-Logan phantom as our desirable signal u^+, and produce a noisy version u^- of it, corrupted by Gaußian noise with standard deviation $\sigma = 3.21 \times 10^{-2}$. The image u^- is visualised in Figure 3 a). We proceed in the exact same manner as in the one-dimensional case, with the only difference being the use of the two-dimensional discrete convolution. The filter $\hat{h} \in \mathbb{R}^{n\times n}$ is parametrised with $n = 5$. Figure 3 c) shows the reconstruction of the filter kernel $\hat{h}$, given a random, mean-zero initialisation h^0. As in the previous example, we apply the newly-computed filter via the Morozov regularisation (4) for denoising. Figure 3 b) visualises the outcome, for $\mu = 1.2$. We observe that $\hat{h}$ resembles finite differences in a diagonal direction, again mimicking the behaviour of total variation regularisation (though only in one dimension). We want to emphasise that for different initialisations other directions may be approximated instead.

5 Conclusions & Outlook

We have presented a novel method for the computation of adaptive regularisation functions based on the minimisation of a quotient of parametrised regularisation functions. We have considered a numerical algorithm based on [14] that is guaranteed

a) b) c)

Fig. 3: A noisy version of the Shepp-Logan phantom (Figure 3 a)). The noise is Gaußian with standard deviation $\sigma = 3.21 \times 10^{-2}$. Figure 3 b) shows the solution of (4) for f given as another noisy version of u^{+}, and $\hat{h}$ as in Figure 3 c). The latter has been computed via (1).

to converge globally to a local solution of this problem, and have presented preliminary numerical results that support the idea of quotient minimisation in the context of parameter learning.

Future work will include more sophisticated parametrisations of regularisation functions, as well as the incorporation of more diverse training data that consist of more than two signals. An extended work will also include a detailed numerical and theoretical analysis of the proposed method, which is beyond the scope of this paper.

6 Acknowledgements

MB and CBS acknowledge support from the EPSRC grant EP/M00483X/1 and the Leverhulme grant 'Breaking the non-convexity barrier'. GG acknowledges support from the Israel Science Foundation (grant No. 718/15) and by the Magnet program of the OCS, Israel Ministry of Economy, in the framework of Omek Consortium.

References

[1] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 7.1 (Revision 28).*, 2015.
[2] Martin Benning, Christoph Brune, Martin Burger, and Jahn Müller. Higher-order tv methods—enhancement via bregman iteration. *Journal of Scientific Computing*, 54(2-3):269–310, 2013.
[3] Martin Benning and Martin Burger. Ground states and singular vectors of convex variational regularization methods. *Methods and Applications of Analysis*, 20(4):295–334, 2013.
[4] Xavier Bresson, Thomas Laurent, David Uminsky, and James V Brecht. Convergence and energy landscape for cheeger cut clustering. In *Advances in Neural Information Processing Systems*, pages 1385–1393, 2012.
[5] Martin Burger, Lina Eckardt, Guy Gilboa, and Michael Moeller. Spectral representations of one-homogeneous functionals. In *Scale Space and Variational Methods in Computer Vision*, pages 16–27. Springer International Publishing, 2015.
[6] Yunjin Chen, Thomas Pock, and Horst Bischof. Learning ℓ_1-based analysis and synthesis sparsity priors using bi-level optimization. In *Workshop on Analysis Operator Learning vs. Dictionary Learning, NIPS 2012*, 2012.
[7] Julianne Chung, Matthias Chung, and Dianne P O'Leary. Designing optimal spectral filters for inverse problems. 33(6):3132–3152, 2011.
[8] Juan Carlos De los Reyes and Carola-Bibiane Schönlieb. Image denoising: Learning the noise model via nonsmooth PDE-constrained optimization. 7(4), 2013.
[9] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
[10] Guy Gilboa. A total variation spectral framework for scale and texture analysis. *SIAM Journal on Imaging Sciences*, 7(4):1937–1961, 2014.
[11] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. http://stanford.edu/~boyd/graph_dcp.html.
[12] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.
[13] Eldad Haber and Luis Tenorio. Learning regularization functionals – a supervised training approach. 19(3):611, 2003.
[14] Matthias Hein and Thomas Bühler. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse pca. In *Advances in Neural Information Processing Systems*, pages 847–855, 2010.
[15] Karl Kunisch and Thomas Pock. A bilevel optimization approach for parameter learning in variational models. 6(2):938–983, 2013.
[16] Michael Moeller, Julia Diebold, Guy Gilboa, and Daniel Cremers. Learning nonlinear spectral filters for color image reconstruction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 289–297, 2015.
[17] Vladimir Alekseevich Morozov. On the solution of functional equations by the method of regularization. In *Soviet Math. Dokl*, volume 7, pages 414–417, 1966.
[18] Stefan Roth and Michael J Black. Fields of experts: A framework for learning image priors. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 860–867. IEEE, 2005.
[19] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.