Boosting Weakly Convex Ridge Regularizers with Spatial Adaptivity

Sebastian Neumayer, Mehrsa Pourya, Alexis Goujon, Michael Unser
Department of Electrical Engineering
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland
{sebastian.neumayer, mehrsa.pourya, alexis.goujon, michael.unser}@epfl.ch

Abstract

We propose to enhance 1-weakly convex ridge regularizers for image reconstruction by incorporating spatial adaptivity. To this end, we resort to a neural network that generates a weighting mask from an initial reconstruction, which is obtained with the baseline regularizer. Empirically, the learned mask can capture long-range dependencies and leads to a smaller penalization of inherent image structures. Our experiments show that spatial adaptivity improves the performance of image denoising and MRI reconstruction.

1 Introduction

A popular approach for solving linear inverse problems in imaging [22] is to compute a reconstruction $\hat{\mathbf{x}}$ based on the variational problem

$$\hat{\mathbf{x}} = \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^d} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda R(\mathbf{x}), \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{m\times d}$ is the measurement operator, $\mathbf{y} \in \mathbb{R}^m$ is the observed data, and $R\colon \mathbb{R}^d \to \mathbb{R}^+$ is some regularizer with weight $\lambda \in \mathbb{R}^+$ that encodes prior information about the desirable solution. Classic regularizers R, such as the Tikhonov [32] or total-variation (TV) [30] ones, are interpretable and lead to guarantees on consistency and stability [6]. However, they are not state-of-the-art anymore and are (almost) systematically outperformed by data-driven approaches [2]. The latter, though, rarely offer such guarantees [24], and sometimes even hallucinate or remove meaningful structures [15]. Hence, their usage for critical applications such as diagnostic imaging remains controversial.

Recently, the authors of [8, 9] proposed a framework to learn ridge-based regularizers (WCRR), see also [29, 3], which offer theoretical guarantees similar to those of the classic regularizers. For a vectorized image $\mathbf{x} \in \mathbb{R}^d$, the proposed convolutional regularizers read

$$R\colon \mathbf{x} \mapsto \sum_{j=1}^{N_\mathrm{C}} \big\langle \mathbf{1}_d, \psi_j(\mathbf{W}_j\mathbf{x}) \big\rangle, \tag{2}$$

where the differentiable potentials $\psi_j\colon \mathbb{R} \to \mathbb{R}^+$ are applied pixelwise, the $\mathbf{W}_j \in \mathbb{R}^{d\times d}$ are convolution matrices, and j indexes along the N_C channels. Using the shorthand $\mathbf{W} = [\mathbf{W}_1^T \cdots \mathbf{W}_{N_\mathrm{C}}^T]^T \in \mathbb{R}^{dN_\mathrm{C}\times d}$, we get a multichannel filtered version $\mathbf{W}\mathbf{x}$ of the image $\mathbf{x}$, which is penalized via the channel-specific ψ_j. Recall that R is called ρ-weakly convex if $R + \frac{\rho}{2}\|\cdot\|^2$ is convex. By enforcing constraints on the ψ_j and $\mathbf{W}$, it is possible to obtain (1-weakly) convex regularizers, which allows for an efficient minimization of (1). In particular, for image denoising ($\mathbf{H} = \mathbf{I}$) with $\lambda \leq 1$, the objective (1) is convex and an efficient global minimization is possible. Moreover, the data-to-reconstruction map is Lipschitz-continuous in $\mathbf{y}$ [10].

Most recent works based on the variational framework (1) also rely on a data-driven R. However, R is usually chosen as a deep CNN, which makes an interpretation more challenging than for the shallow

NeurIPS 2023Workshop on Deep Learning and Inverse Problems.

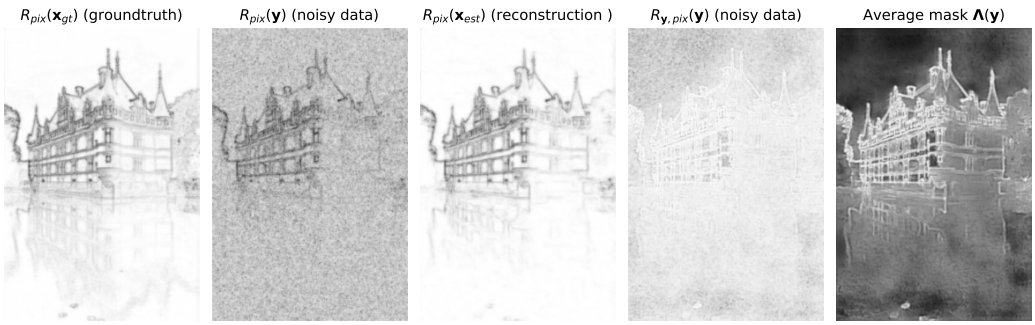

Figure 1: The first three images correspond to the pixel-wise cost $R_{\text{pix}}(\mathbf{x}) = \sum_j \psi_j(\mathbf{W}_j\mathbf{x})$ for $\mathbf{x} \in \{\mathbf{x}_{\text{gt}}, \mathbf{y}, \mathbf{x}_{\text{est}}\}$, respectively, where black corresponds to high values. The last two depict the pixel-wise adapted cost $R_{\mathbf{y},\text{pix}}(\mathbf{x}) = \sum_j \mathbf{\Lambda}_j(\mathbf{y}) \odot \psi_j(\mathbf{W}_j\mathbf{x})$, and the pixel-wise average of the mask.

R in (2). Examples in the convex setting include adversarial convex regularization [23] and specific instances of variational networks [17]. The literature devoted to the non-convex setting is much richer. Examples include [3], which is based on (2) without constraints on the ψ_j; total deep variation [16]; or the network Tikhonov (NETT) approach [20], where the authors also study the robustness of the data-to-reconstruction map. A slightly different approach is taken in [13], where the authors propose to directly learn $\operatorname{prox}_R(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^d} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + R(\mathbf{x})$ as a Gaussian denoiser. To do so, they explicitly parameterize prox_R as a gradient-step denoiser [4, 12] and promote the necessary 1-weak convexity of the underlying R via regularization during training.

Starting with total-variation regularization [11, 35], it has been observed in several works (see for example [7, 36]) that the use of a spatially adaptive regularization mask (as opposed to a constant weight) can significantly boost the performance for the reconstruction model (1). More precisely, assuming a regularizer of the form (2), one modifies it as

$$R_{\mathbf{y}} \colon \mathbf{x} \mapsto \sum_{j=1}^{N_{\text{C}}} \big\langle \mathbf{\Lambda}_j(\mathbf{y}), \psi_j(\mathbf{W}_j\mathbf{x}) \big\rangle, \tag{3}$$

where the regularization mask $\mathbf{\Lambda} \colon \mathbb{R}^m \to (\mathbb{R}_+^d)^{N_{\text{C}}}$ depends on $\mathbf{y}$ according to some suitable heuristic. In (3), the mask can be justified in two different ways.

i) The noise might not be uniform throughout the data, such that a spatially (noise-)adapted regularizer is better suited than a standard one.
ii) Inspecting the pixel-wise regularization cost, see Figure 1, we observe that both the noise and the structure in the ground truth $\mathbf{x}_{\text{gt}}$ lead to high cost. Consequently, to avoid oversmoothing, the values of $\mathbf{\Lambda}$ should be smaller at structures. Further, as structure might respond to some $\mathbf{W}_j$ and not to others, it is reasonable to use different masks across the channels.

Recently, spatial adaptivity has also found its way into data-driven regularization [19, 18, 25], where $\mathbf{\Lambda}$ is computed with a neural newtwork (NN) based on the data $\mathbf{y}$. For an overview, we refer to [28].

Outline and contributions We propose to learn a mask-generating NN $G \colon \mathbb{R}^d \to [0,1]^{dN_{\text{C}}}$ for the construction of a spatially adaptive ridge regularizer (SARR) $R_{\mathbf{y}}$ of the form (3), which takes an initial reconstruction $\mathbf{x}_{\text{est}} \in \mathbb{R}^d$ as input. We implicitly assume that $\mathbf{x}_{\text{est}}$ is good enough, so that the mask-generating NN G in $\mathbf{\Lambda}(\mathbf{y}) = G(\mathbf{x}_{\text{est}})$ is mostly independent of the operator $\mathbf{H}$. Thus, we can learn G with the same Gaussian denoising task as in [9], which allows for efficient multi-noise level training. Note that $\mathbf{\Lambda}(\mathbf{y})$ can be interpreted as a conditioning of R on the data $\mathbf{y}$.

The architecture of a SARR and the training details are given in Section 2. In particular, it holds that the denoising task (1) is convex in $\mathbf{x}$, which allows for robust training. Based on the learned $\mathbf{\Lambda}$, we provide an experimental evaluation in Section 3, both for denoising and magnetic resonance imaging (MRI) reconstruction. We find that the conditioning of R on $\mathbf{y}$ increases the performance for both setups. Additionally, we provide comparisons with other reconstruction methods based on (1). Our implementation builds on [9][1], and pre-trained models are available upon request. Finally, conclusions are drawn in Section 4.

[1] `https://github.com/axgoujon/weakly_convex_ridge_regularizer`

2 Architecture and training

The backbone of our scheme is the parameterization of R proposed in [9]. It involves 80 learnt multi-convolutions for which the corresponding operator $\mathbf{W}$ satisfies $\|\mathbf{W}\| = 1$. The associated potentials ψ_j depend on the standard deviation σ of the underlying noise as $\psi_{j,\sigma} = \alpha_j(\sigma)^{-2}\psi(\alpha_j(\sigma)\cdot)$ with $\alpha_j(\sigma) = \mathrm{e}^{s_{\alpha_j}(\sigma)}/(\sigma + 10^{-5})$, where s_{α_j} are learned linear splines, and $\psi = \mu\psi_+ - \psi_-$ with the two quadratic splines ψ_+, ψ_- satisfying $\psi_+''(x), \psi_-''(x) \in [0, 1]$ for almost every $x \in \mathbb{R}$. The $\alpha_j(\sigma)$ have 11 equally distant knots in $[0, 30/255]$, and the ψ_+, ψ_- have 101 equally spaced knots with spacing $\Delta = 0.002$. It is shown in [9, Prop. 3.1] that the resulting R is 1-weakly convex. Throughout this paper, the convolution filters in $\mathbf{W}$ and the potential ψ of the baseline regularizer remain fixed, and we only consider the model calibration parameters, namely the weight μ and the scalings $\alpha_{j,\sigma}$, as learnable. Note that a modification of these will not change the estimate of the weak-convexity modulus of R [9, Rem. 3.3.].

For the mask generation in the SARR, we first compute an unconditional reconstruction $\mathbf{x}_{\text{est}}$ based on (1) using accelerated gradient descent (AGD) [26] with a restart technique [27]. For denoising, we use the pre-trained R from [9] with the true σ and, for inverse problems, we tune its hyperparameters as described in [9], knowing that AGD converges either to a global minimum (denoising) or a critical point (inverse problems). We observed that the reconstruction is relatively independent of the initialization, which is important for a stable mask generation. The obtained $\mathbf{x}_{\text{est}}$ is then plugged into the NN G, which is chosen as a so-called RFDN [21] with one input channel, 40 features, 80 output channels, and superresolution factor one (the same choice as in [19]), which results in $\sim 3 \cdot 10^5$ parameters. Further, we apply the sigmoid function to its output to ensure that the values are in $[0, 1]$. Due to this constraint, we have $\mathbf{\Lambda}_{j,k}(\mathbf{y})\psi_{j,\sigma}'' \geq -1$ almost everywhere. Hence, $R_{\mathbf{y}}$ remains 1-weakly convex [9, Prop. 3.1], and we can use the same stepsize $1/(1 + \max(1, \mu))$ as with R for minimizing (1) with AGD. Other architectures for G have been successfully deployed, for example, in [18].

Now, let $\boldsymbol{\theta}(\sigma)$ represent the aggregated set of learnable parameters of the conditional (i.e. spatially adaptive) $R_{\mathbf{y}}$, which we write as $R_{\mathbf{y},\boldsymbol{\theta}(\sigma)}$ for an explicit reference. Our goal is to learn $\boldsymbol{\theta}(\sigma)$ such that $R_{\mathbf{y},\boldsymbol{\theta}}$ generalizes well to a variety of inverse problems. Specifically, we optimize $\boldsymbol{\theta}(\sigma)$ (and, in particular, G) such that the conditional multi-noise-level denoiser

$$D_{\boldsymbol{\theta}(\sigma)}(\mathbf{y}) = \operatorname{prox}_{R_{\mathbf{y},\boldsymbol{\theta}(\sigma)}}(\mathbf{y}) = \operatorname*{arg\,min}_{\mathbf{x}\in\mathbb{R}^d} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + R_{\mathbf{y},\boldsymbol{\theta}(\sigma)}(\mathbf{x}) \tag{4}$$

is a good Gaussian denoiser across a variety of standard deviations σ. To this end, we extract $M = 238\,400$ grayscale patches $\{\mathbf{x}^m\}_{m=1}^M$ of size (40×40) from 400 images of the BSD500 data set [1]. Each patch $\mathbf{x}^m$ is corrupted as $\mathbf{y}^m = \mathbf{x}^m + \sigma^m\mathbf{n}^m$ with Gaussian noise $\mathbf{n}^m \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ scaled by $\sigma^m \sim \mathcal{U}[0, 30/255]$. Then, we define the multi-noise-level training problem

$$\operatorname*{arg\,min}_{\boldsymbol{\theta}} \sum_{m=1}^{M} \mathbb{E}_{(\mathbf{n}^m,\sigma^m)}\big(\|D_{\boldsymbol{\theta}(\sigma^m)}(\mathbf{y}^m) - \mathbf{x}^m\|_1\big). \tag{5}$$

There, $D_{\boldsymbol{\theta}(\sigma)}(\mathbf{y})$ is computed using AGD. Its gradient with respect to $\boldsymbol{\theta}(\sigma)$ is computed with the same implicit differentiation techniques as in [9]. Hence, we can solve the training problem (5) using the methods know as ADAM [14]. Our model is trained for 25 epochs with batches of size 100, which takes around 18 hours on a Tesla V100 GPU. The learning rates are initially set to $5 \cdot 10^{-2}$ for μ, $5 \cdot 10^{-3}$ for s_{α_j}, and to 10^{-4} for G. Then, they are decayed by 0.75 every 250 batches.

3 Experimental evaluation

Denoising The evaluation of the proposed SARR (3) and several other methods on the BSD68 test set is provided in Table 1. The task is not blind, in the sense that the standard deviation σ is used directly as an input for all methods. The first important observation is that the conditioning of WCRRs on $\mathbf{y}$ allows one to significantly improve the denoising performance. If we inspect the response of clean images to $\mathbf{W}$ after application of the mask, as in Figure 1, we observe that the response of the structure is damped. This mechanism may explain the quality improvement. Moreover, we almost reach the performance of the deep proximal denoiser Prox-DRUNet [13], which is currently one of the best-performing methods for convergent plug-and-play image-reconstruction algorithms. However, for Prox-DRUNet the conditions to be a proximal operator (as stated in [10]) are only

Table 1: Denoising performance (measured by PSNR) on the BSD68 test set.

Method	BM3D [5]	WCRR[1] [9]	Proposed SARR[1]	Prox-DRUNet[2] [13]
$\sigma = 5/255$	37.54	37.68	<u>37.84</u>	**37.98**
$\sigma = 15/255$	31.11	31.22	<u>31.54</u>	**31.70**
$\sigma = 25/255$	28.60	28.69	<u>29.07</u>	**29.18**

[1]Minimization of a convex functional. [2]Minimization of an *approximately* convex functional.

Table 2: PSNR (first columns) and SSIM (second columns) values for the MRI experiment.

	4-fold single coil				8-fold multi-coil			
	PD	PDFS	PD	PDFS	PD	PDFS	PD	PDFS
Zero-fill ($\mathbf{H}^T\mathbf{y}$)	27.40	29.68	0.729	0.745	23.80	27.19	0.648	0.681
TV	32.44	32.67	0.833	0.781	32.77	33.38	0.850	0.824
CRR-NN [8]	33.99	33.75	0.880	0.831	34.29	34.50	0.881	0.852
PnP-DnCNN [31]	35.24	34.63	0.884	<u>0.840</u>	35.11	35.14	0.881	**0.858**
WCRR [9]	35.78	34.63	0.899	0.838	35.57	<u>35.16</u>	<u>0.894</u>	0.856
Proposed SARR	**36.25**	<u>34.77</u>	**0.904**	0.839	**35.98**	**35.26**	**0.901**	**0.858**
Prox-DRUNet [13]	<u>36.20</u>	**35.05**	<u>0.901</u>	**0.847**	<u>35.78</u>	35.12	<u>0.894</u>	0.857

enforced via regularization. Given that the metrics found in Table 1 are quite close, the conceptual advantage of a SARR is that we have an interpretation as analysis prior. A visual example is provided in Appendix A. If σ is actually unknown, we can use an estimator or tune it on a calibration set.

Inverse problems The SARR trained to denoise on the BSD500 dataset is now deployed to solve the single- and 15-coil MRI-reconstruction problems given in [8]. The ground-truth images are generated from fully sampled k-space measurements and consist of proton-density-weighted knee MR images from the fastMRI dataset [15] with fat suppression (PDFS) and without fat suppression (PD). The hyperparameters λ and σ in (1) for the unconditional R are tuned with a coarse-to-fine grid search [8] on a 10-image calibration set, independently for PD and PDFS images. The tests are then performed on 50 images. In the single-coil setup, we simulate the data by masking the Fourier transform of the ground-truth image. In the 15-coil setup, we simulate the data by subsampling the Fourier transforms of the multiplication of the ground-truth images with sensitivity maps computed with the BART [34] implementation of the ESPIRiT algorithm [33] from the raw k-space data. The subsampling of the measurements for both setups is determined by the acceleration factor M_{acc}. More precisely, a fraction of $0.32/M_{\text{acc}}$ columns in the center of the k-space (low frequencies) is kept, and columns in the other region of the k-space are uniformly sampled so that the overall expected proportion of selected columns is $1/M_{\text{acc}}$. Gaussian noise with $\sigma_{\mathbf{n}} = 2 \cdot 10^{-3}$ is then added to the real and imaginary parts of the measurements. Both the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) values reported in Table 2 are computed on a centered (320×320) patch. The full implementation details for CRR-NN and PnP-DnCNN reconstructions are given in [8] for the same setup. The Prox-DRUNet reconstructions are obtained with the DRS-PnP algorithm given in [13], for which the hyperparameters are tuned as for the other methods. The proximal operator $\text{prox}_{\tau\|\mathbf{H}\cdot-\mathbf{y}\|^2}$ required by the DRS-PnP is computed through the conjugate-gradient method, which is iterated to numerical precision.

As in the denoising scenario, we observe that the conditioning of WCRRs boosts the performance in all test settings. Moreover, we outperform PnP-DnCNN (a method that does not even provide convergence guarantees for inverse problems) and are overall comparable to Prox-DRUNet. A visual reconstruction example for each setting is provided in Appendix A.

4 Conclusion

We have provided a proof of concept for the benefit of conditioning in ridge-based regularizers. Although a good $\mathbf{x}_{\text{est}}$ is essential for computing the mask $\mathbf{\Lambda}$, we observed that our approach yields a SARR $R_{\mathbf{y}}$ that generalizes well to new tasks. In the future, we want to investigate the theoretical properties and to search for simpler mask-generation networks.

Acknowledgments and Disclosure of Funding

The research leading to this publication was supported by the European Research Council (ERC) under European Union's Horizon 2020 (H2020), Grant Agreement - Project No 101020573 FunLearn, and by the Swiss National Science Foundation, Grant 200020 184646/1. The authors have no conflicts of interest to declare.

References

[1] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011.

[2] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019.

[3] Y. Chen, R. Ranftl, and T. Pock. Insights into analysis operator learning: From patch-based sparse models to higher order MRFs. *IEEE Transactions on Image Processing*, 23(3):1060–72, 2014.

[4] R. Cohen, Y. Blau, D. Freedman, and E. Rivlin. It has potential: Gradient-driven denoisers for convergent solutions to inverse problems. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007.

[6] P. del Aguila Pla, S. Neumayer, and M. Unser. Stability of image-reconstruction algorithms. *IEEE Transactions on Computational Imaging*, 9:1–12, 2023.

[7] Y. Dong and C.-B. Schönlieb. Tomographic reconstruction with spatially varying parameter selection. *Inverse Problems*, 36(5):054002, 2020.

[8] A. Goujon, S. Neumayer, P. Bohra, S. Ducotterd, and M. Unser. A neural-network-based convex regularizer for inverse problems. *IEEE Transactions on Computational Imaging*, 9:781–795, 2023.

[9] A. Goujon, S. Neumayer, and M. Unser. Learning weakly convex regularizers for convergent image-reconstruction algorithms. *ArXiv preprint #2308.10542*, 2023.

[10] R. Gribonval and M. Nikolova. A characterization of proximity operators. *Journal of Mathematical Imaging and Vision*, 62(6-7):773–789, 2020.

[11] M. Hintermüller, K. Papafitsoros, and C. N. Rautenberg. Analytical aspects of spatially adapted total variation regularisation. *Journal of Mathematical Analysis and Applications*, 454(2):891–935, 2017.

[12] S. Hurault, A. Leclaire, and N. Papadakis. Gradient step denoiser for convergent Plug-and-Play. In *10th International Conference on Learning Representations*, 2022.

[13] S. Hurault, A. Leclaire, and N. Papadakis. Proximal denoiser for convergent Plug-and-Play optimization with nonconvex regularization. In *39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 9483–9505, 2022.

[14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*, 2015.

[15] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdzalv, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui. fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiology: Artificial Intelligence*, 2(1):e190007, 2020.

[16] E. Kobler, A. Effland, K. Kunisch, and T. Pock. Total deep variation for linear inverse problems. In *2020 Conference on Computer Vision and Pattern Recognition*, pages 7549–7558, 2020.

[17] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock. Variational networks: Connecting variational methods and deep learning. In *Pattern Recognition*, pages 281–293. Springer, 2017.

[18] A. Kofler, F. Altekrüger, F. A. Ba, C. Kolbitsch, E. Papoutsellis, D. Schote, C. Sirotenko, F. F. Zimmermann, and K. Papafitsoros. Learning regularization parameter-maps for variational image reconstruction using deep neural networks and algorithm unrolling. *ArXiv preprint #2301.05888*, 2023.

[19] S. Lefkimmiatis and I. S. Koshelev. Learning sparse and low-rank priors for image recovery via iterative reweighted least squares minimization. In *11th International Conference on Learning Representations*, 2023.

[20] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier. NETT: Solving inverse problems with deep neural networks. *Inverse Problems*, 36(6):065005, 2020.

[21] J. Liu, J. Tang, and G. Wu. Residual feature distillation network for lightweight image super-resolution. In *Computer Vision–ECCV 2020 Workshops*, pages 41–55. Springer, 2020.

[22] M. T. McCann and M. Unser. Biomedical image reconstruction: From the foundations to deep neural networks. *Foundations and Trends® in Signal Processing*, 13(3):283–359, 2019.

[23] S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C.-B. Schönlieb. Learned convex regularizers for inverse problems. *arXiv:2008.02839*, 2021.

[24] S. Mukherjee, A. Hauptmann, O. Öktem, M. Pereyra, and C.-B. Schönlieb. Learned reconstruction methods with convergence guarantees: A survey of concepts and applications. *IEEE Signal Processing Magazine*, 40(1):164–182, 2023.

[25] R. R. D. Nekhili, X. Descombes, and L. Calatroni. A hybrid approach combining CNNs and variational modelling for blind image denoising. *HAL preprint #03596605*, 2022.

[26] Y. E. Nesterov. A method of solving a convex programming problem with convergence rate $\mathrm{o}(1/k^2)$. *Doklady Akademii Nauk*, 269(3):543–547, 1983.

[27] B. O'Donoghue and E. Candès. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15(3):715–732, 2015.

[28] M. Pragliola, L. Calatroni, A. Lanza, and F. Sgallari. On and beyond total variation regularization in imaging: The role of space variance. *SIAM Review*, 65(3):601–685, 2023.

[29] S. Roth and M. J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205–229, 2009.

[30] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[31] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin. Plug-and-Play methods provably converge with properly trained denoisers. In *36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5546–5557, 2019.

[32] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics*, 4:1035–1038, 1963.

[33] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. ESPIRiT-An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, 2014.

[34] M. Uecker, P. Virtue, F. Ong, M. J. Murphy, M. T. Alley, S. S. Vasanawala, and M. Lustig. Software toolbox and programming library for compressed sensing and parallel imaging. In *ISMRM Workshop on Data Sampling and Image Reconstruction*, page 41, 2013.

[35] C. Van Chung, J. C. De los Reyes, and C. Schönlieb. Learning optimal spatially-dependent regularization parameters in total variation image denoising. *Inverse Problems*, 33(7):074005, 2017.

[36] Q. Zhong, R. W. Liu, and Y. Duan. Spatially adapted first and second order regularization for image reconstruction: From an image surface perspective. *Journal of Scientific Computing*, 92(2):Paper No. 33, 2022.

A Denoising and reconstruction examples

Figure 2 contains a visual comparison of the denoising capability for the methods of Table 1 applied to a natural image. Significant differences appear, for example in the sky or in the spire. Since the top of the spire is already quite smoothed out in the WCRR reconstruction, one has no hope of recovering it via the proposed two-step conditioning and a direct approach like ProxDRUNet works better.

Figures 3 and 4 contain MRI reconstructions (single-coil setup). Here, the improvement is not consistent across images. If we inspect the computed masks (Figure 5), we observe that these are quite conservative for the images without much improvement over the baseline WCRR. In particular, they do not aggressively smooth the background (as done for the images with larger improvement), and they have less variation throughout the structure. This indicates that there is indeed room for improvement with the mask-generation process. However, we want to stress that the mask generation in this task is challenging since boundary effects related to the data occur in the reconstructions (which is why the metrics for this challenge are actually only computed on a centered patch). To complement these results, we provide reconstructions for the 15-coil setup in Figure 6.

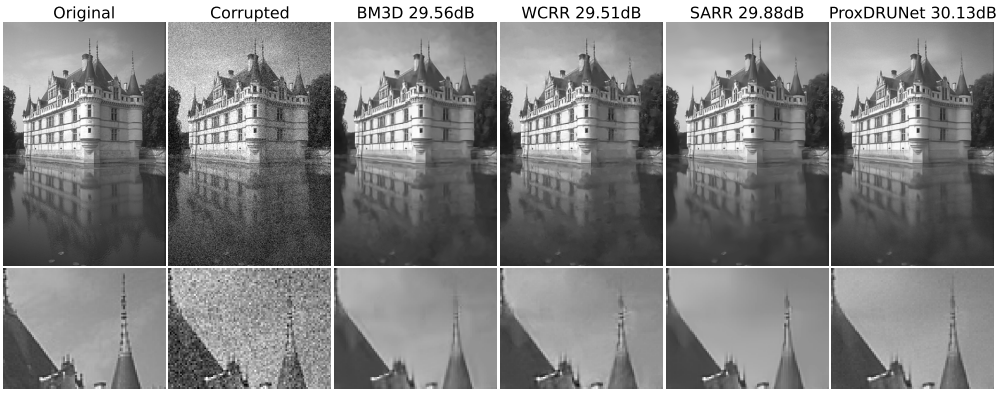

Figure 2: Denoising of the castle image using several different methods for $\sigma = 25/255$.

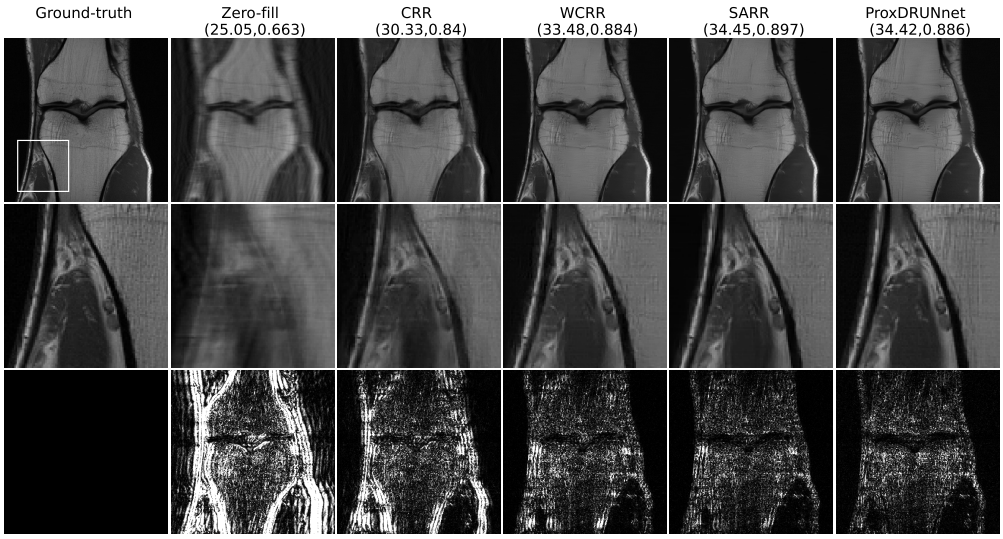

Figure 3: Reconstructions for single-coil MRI (PD). The reported metric is (PSNR, SSIM) and the last row shows the squared value of the residuals.

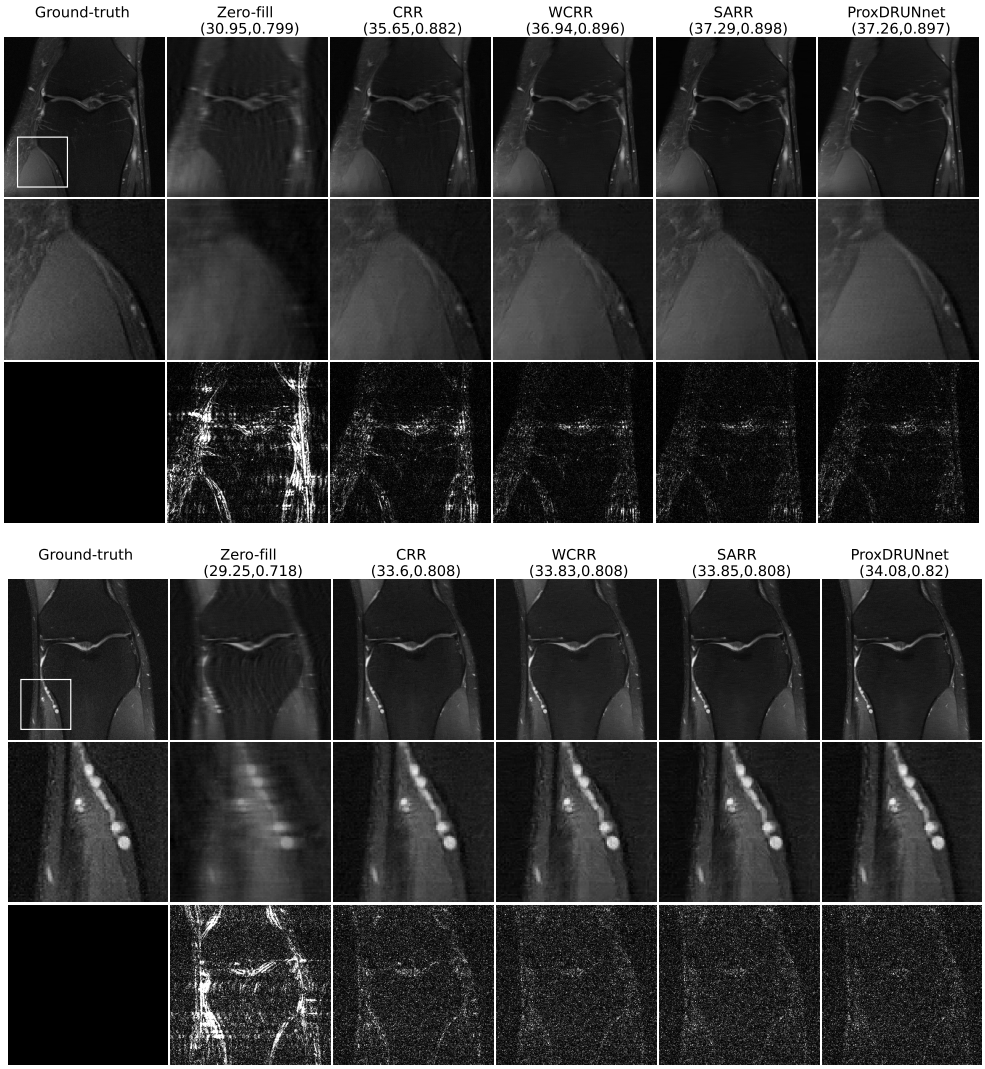

Figure 4: Reconstructions for single-coil MRI (PDFS). The reported metric is (PSNR, SSIM) and the last row shows the squared value of the residuals.

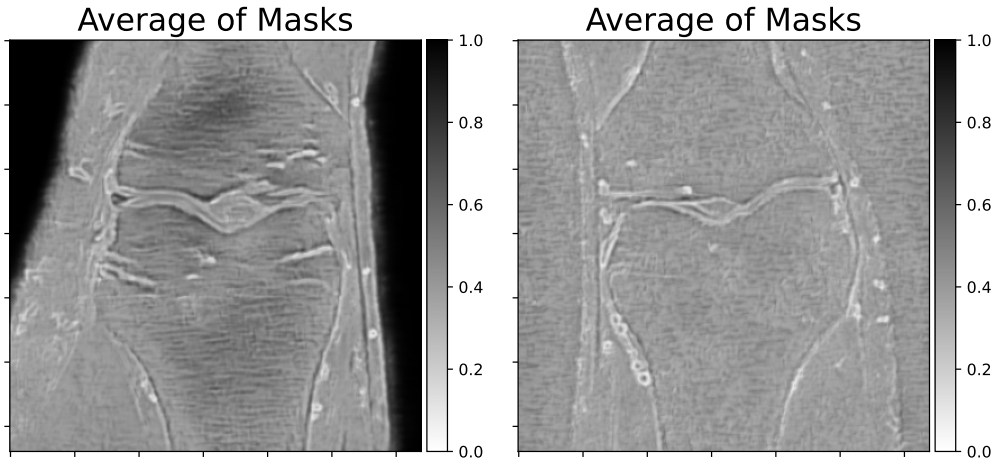

Figure 5: Pixel-wise average of the masks $\mathbf{\Lambda}$ for the images in Figure 4. The left mask corresponds to a reconstruction with improvement over WCRR, whereas the right one did not improve.

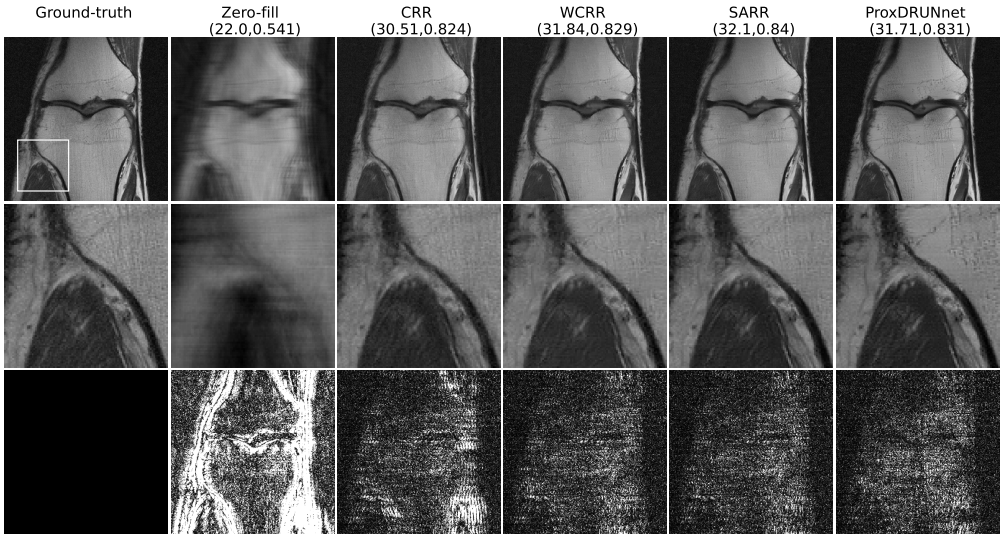

Figure 6: Reconstructions for multi-coil MRI (PD). The reported metric is (PSNR, SSIM) and the last row shows the squared value of the residuals.