Provably Convergent Data-Driven Convex-Nonconvex Regularization

Zakhar Shumaylov
University of Cambridge
zs334@cam.ac.uk

Jeremy Budd
California Institute of Technology
jmbudd@caltech.edu

Subhadip Mukherjee
IIT Kharagpur
subhadipju@gmail.com

Carola-Bibiane Schönlieb
University of Cambridge
cbs31@cam.ac.uk

Abstract

An emerging new paradigm for solving inverse problems is via the use of deep learning to learn a regularizer from data. This leads to high-quality results, but often at the cost of provable guarantees. In this work, we show how well-posedness and convergent regularization arises within the convex-nonconvex (CNC) framework for inverse problems. We introduce a novel input weakly convex neural network (IWCNN) construction to adapt the method of learned adversarial regularization to the CNC framework. Empirically we show that our method overcomes numerical issues of previous adversarial methods.

1 Introduction

Inverse problems appear as the main mathematical formulation of a number of scientific applications, including numerous problems in medical imaging. In inverse problems, one seeks to estimate an unknown parameter $x^* \in \mathcal{X}$ from a transformed and noisy measurement

$$y^\delta = \mathcal{A}x^* + e \in \mathcal{Y}\,. \tag{1}$$

Here, $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ is the *forward operator*, assumed to be linear and bounded, e.g. representing imaging physics, and $e \in \mathcal{Y}$, with $\|e\|_{\mathcal{Y}} \leq \delta$, describes measurement noise. However, (1) is frequently *ill-posed*, i.e. the inverse may not exist, be unique, or be continuous in the measurement y^δ.

Traditionally, variational approaches were used to overcome this ill-posedness by hand-crafting a *regularizer* that aims to incorporate prior information about x^* and they are built on top of a rigorous function-analytic foundation. A number of approaches have appeared in recent years which propose to learn a (deep) regularizer directly from data, see e.g. [3, 7, 12, 15, 16, 20, 21, 26, 28, 31, 32, 34, 36, 37] and see [6, 9] for overviews. These techniques are often able to achieve high-quality reconstructions, but (unlike traditional variational approaches) often lack provable properties.

The methods described above typically only approach regularization in the sense of model space regularization. This work will explicitly utilize both data and model space regularization (see [10]). We will employ the technique of learning *adversarial regularizers* (ARs). This was first proposed in [21] and has since seen extensions including multi-step regularizers [23], latent optimized AR [38], and in [25] deep *input convex neural networks* (ICNNs) (see [4]) were used to learn an adversarial *convex* regularizer (ACR), which allowed for a number of desirable theoretical guarantees to be proved. However, the imposition of convexity is quite restrictive, and it has been observed in the literature that nonconvex regularizers often have better performance; see [19, 24, 29, 33]. Of particular interest is the *convex-nonconvex (CNC) framework*, wherein the regularizer is kept nonconvex in a

NeurIPS 2023 Workshop on Deep Learning and Inverse Problems

structured enough way to guarantee convexity of the overall objective, for a review of such methods see [18]. In [11] for example this is achieved by using a ridge regularizer and enforcing the overall function to be 1-weakly convex, turning the denoising objective convex. In a similar manner, a proximal operator corresponding to a weakly convex regularizer can be learned, as in [13].

In this work, we adapt the AR method to the CNC framework. In particular, we introduce the *input weakly convex neural network* (IWCNN) construction, generalizing ICNNs, to impose weak convexity on regularizers. This is motivated by a desire to retain the provable guarantees and desirable optimization properties of ACRs, but also to exploit the advantages of nonconvex regularization. In this formulation we are able to show well-posedness and convergent regularization in the sense of stationary points of the regularizer, using existing results from the literature. We then use our IWCNNs to learn an *adversarial convex-nonconvex regularizer* (ACNCR) and show that this is more versatile than the AR or ACR in computed tomography (CT) experiments.

2 Background and problem formulation

In the function-analytic formulation of inverse problems, the unknown parameter x^* is modeled as deterministic and one approximates it from y^δ by solving a variational reconstruction problem:

$$x_\alpha(y^\delta) \in \operatorname*{arg\,min}_{x\in\mathcal{X}} J_\alpha\left(x; y^\delta\right) := \mathcal{L}_\mathcal{Y}\left(y^\delta, \mathcal{A}\,x\right) + \alpha\,\mathcal{R}(x). \tag{2}$$

Here, $\mathcal{X}$ and $\mathcal{Y}$ are normed vector spaces, $\mathcal{L}_\mathcal{Y} : \mathcal{Y}\times\mathcal{Y} \to \mathbb{R}^+$ measures *data fidelity*, and the *regularizer* $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ penalizes undesirable images. The parameter $\alpha > 0$ trades off data fidelity with regularization and is chosen depending on the noise strength δ. Henceforth, we will choose $\mathcal{L}_\mathcal{Y}(y_1, y_2) := \|y_1 - y_2\|_\mathcal{Y}^2$ for convenience.

In practice, (2) is solved using iterative optimization schemes, and the quality of reconstructions depends heavily on the choice of $\mathcal{R}$. While convexity of $\mathcal{R}$ may be desirable analytically to provide efficient optimization schemes [27] with various guarantees, in practice reconstruction quality is much better for nonconvex regularizers, as discussed in Section 1. This, however, comes at a cost: finding global minima becomes impossible in general and sometimes even finding stationary points can not be guaranteed. The *convex-nonconvex (CNC) framework* addresses this problem, wherein the regularizer is kept nonconvex in a structured way to guarantee convexity of the overall objective $J_\alpha(\cdot, y)$. With this goal, we choose the regularizer as a combination of a weakly convex function over the data space and convex over the model/parameter space, i.e. $\mathcal{R}(x) := \mathcal{R}^{cnc}(x, \mathcal{A}\,x)$ where $\mathcal{R}^{cnc}(x, y) := \mathcal{R}^c(x) + \mathcal{R}^{wc}(y)$, where $\mathcal{R}^c$ is convex and $\mathcal{R}^{wc}$ is weakly convex.

Definition 2.1 (ρ-weak convexity)**.** Let $\rho > 0$. For U a nonempty convex subset of $\mathcal{X}$, a function $f : U \to \overline{\mathbb{R}}$ is said to be *ρ-weakly convex*, if for all $x_1, x_2 \in U$ and $\lambda \in [0, 1]$,

$$f\left(\lambda x_1 + (1-\lambda)x_2\right) \leqslant \lambda f\left(x_1\right) + (1-\lambda) f\left(x_2\right) + \rho\lambda(1-\lambda)\left\|x_1 - x_2\right\|^2. \tag{3}$$

The class of weakly convex functions is quite broad and it includes all convex functions and all smooth functions with Lipschitz continuous gradients.

3 Theoretical results

Under the CNC parametrization of the regularizer, we now provide convergence guarantees in terms of stability, convergent regularization, and fixed point convergence [27].

Definition 3.1. Given $y^0 \in \mathcal{Y}$, we say that $x^\dagger$ is an *$\mathcal{R}$-minimizing solution* if

$$x^\dagger \in \operatorname*{arg\,min}_{x\in\mathcal{X}} \mathcal{R}(x) \text{ subject to } \mathcal{A}\,x = y^0. \tag{4}$$

Note that an $\mathcal{R}$-minimizing solution can be written as $x^\dagger \in \operatorname*{arg\,min}_{x\in\mathcal{X}} \mathcal{R}^{cnc}(x, y^0)$ s.t. $\mathcal{A}\,x = y^0$ due to the constraint, implying uniqueness under e.g. strict convexity of $\mathcal{R}^{cnc}$ in the first argument.

Theorem 3.1. For $\mathcal{R}^{cnc}(x, y) := \mathcal{R}^c(x) + \mathcal{R}^{wc}(y)$ proper, lower semi-continuous, μ-strongly convex in the first argument, and ρ-weakly convex and bounded in the second argument, we have:

1. **Weak convexity:** For $\alpha\rho > 1$: $J_\alpha(\cdot, y)$ is $-\alpha\mu + (\alpha\rho - 1)\| \mathcal{A} \|^2$-weakly convex; For $\alpha\rho \leq 1$: $J_\alpha(\cdot, y)$ is $\alpha\mu$ - strongly convex.

2. **Existence:** $J_\alpha(x; y)$ has a minimizer $x_\alpha(y)$ for every y and $\alpha > 0$. Furthermore for $\alpha\rho \leq 1$, $x_\alpha(y)$ is unique.

3. **Stability:** Sequences of minimizers of $J_\alpha(\cdot, y^\delta)$ are stable with respect to the data y^δ, i.e. if $\lim_{k\to\infty} \|y^\delta - y_k\| = 0$, then every sequence $x_k \in \arg\min J_\alpha(\cdot, y_k)$ has a subsequence weakly convergent to some minimizer $x_\alpha\left(y^\delta\right)$.

Typically, finding global minima of nonconvex objectives is infeasible; we will recover stationary points instead. For $\widetilde{x}_\alpha(y) \in \{x \in \mathcal{X} \mid 0 \in \partial_x J_\alpha(x; y)\}$, for $\partial_x J_\alpha(x; y)$ the Clarke subdifferential:

4. **Locally convergent regularization:** For $\|y^\delta - y^0\| \leq \delta$, $\delta \to 0$, $\alpha(\delta) \to 0$, and $\frac{\delta}{\alpha(\delta)} \to 0$, the reconstruction $\widetilde{x}_\alpha\left(y^\delta\right)$ converges to the unique $\mathcal{R}$-minimizing solution $x^\dagger$ in (4). In words: in the vanishing noise limit, there exists a regularization parameter selection strategy under which reconstructions converge to the solution of the noiseless operator equation.

5. **Convergence of sub-gradient updates:** Given the subgradient descent method $x_{k+1} = x_k - \eta_k v_k$, with $v_k \in \partial J_\alpha(\cdot, y)(x_k)$, if $\alpha\rho \leq 1$: there exists a choice of η_k^* such that x_k converge to the minimizer $x_\alpha(y)$ with respect to the norm on $\mathcal{X}$.

Proof. Items (1-5) directly follow from [25, 30]. □

4 Adversarial convex-nonconvex regularization

4.1 Adversarial regularization and the ACR

Within the adversarial regularization approach, $\mathcal{R}_\theta$ is chosen as a neural network. Assume that we have a dataset of samples $(x_i) \in \mathcal{X}$ and $(y_i) \in \mathcal{Y}$ i.i.d. from the distributions of ground truth images $\mathbb{P}_r$ and measurements $\mathbb{P}_Y$, respectively. We are in the setting of weakly supervised learning, i.e. these are *not* samples of measurement-ground truth pairs. In order to 'compare' the two distributions, we map $\mathbb{P}_Y$ from the space of measurements $\mathcal{Y}$ to the original space $\mathcal{X}$ using some pseudo-inverse $\mathcal{A}^\dagger$. We denote the projected distribution by $\mathbb{P}_n := (\mathcal{A}^\dagger)_\# \mathbb{P}_Y$, where $\#$ denotes the push-forward of measures. Then $\mathbb{P}_n$ will correspond to the distribution of images with reconstruction artifacts.

Now, $\mathcal{R}_\theta$ is meant to penalize artificial images and promote real images, so we want $\mathcal{R}_\theta$ to be large on $\mathbb{P}_n$ and small on $\mathbb{P}_r$. Therefore, [21] chose the following loss functional to minimize:

$$\mathbb{E}_{X\sim\mathbb{P}_r}\left[\mathcal{R}_\theta(X)\right] - \mathbb{E}_{X\sim\mathbb{P}_n}\left[\mathcal{R}_\theta(X)\right] + \lambda \cdot \mathbb{E}\left[\left(\|\nabla_x \mathcal{R}_\theta(X)\| - 1\right)_+^2\right]. \tag{5}$$

Here, the last term is used to enforce the neural network to be 1-Lipschitz with respect to the input.[1]

The ACR is then defined in [25] to be of the form $\mathcal{R}_\theta(x) = \mathcal{R}_\theta^{\text{ICNN}}(x) + \mu\|x\|^2$ where $\mathcal{R}_\theta^{\text{ICNN}}$ is an ICNN [4]. This is again trained by minimizing (5).

4.2 Input weakly convex neural network (IWCNN)

Based on the discussion in Sections 2 and 3, we wish to construct a neural network parameterization that is nonconvex, but weakly convex, with respect to the input, to be used for adversarial regularization. A natural question to ask is whether the original AR is nonconvex and weakly convex. Unfortunately, the answer is negative due to the following result, which we prove in Appendix C.

Theorem 4.1. A piecewise linear continuous function $f : \mathbb{R}^d \to \mathbb{R}$ with a finite number of pieces (e.g., a `ReLU` or `leakyReLU` neural net) is weakly convex if and only if it is convex.

Thus, we need to construct a network which is nonconvex but guaranteed to be weakly convex. For this, we make use of the following fact (for a generalisation to Banach spaces, see Appendix D).

[1] This is done in similarity with the Wasserstein GAN loss (see [5]), with the expected value in the last term taken over all lines connecting samples in $\mathbb{P}_n$ and $\mathbb{P}_r$.

Theorem 4.2 (Weak convexity of compositions [8])**.** Let $F(x) = h(c(x))$ for $h : \mathbb{R}^m \to \mathbb{R}$ convex and L-Lipschitz, and $c : \mathbb{R}^d \to \mathbb{R}^m$ a C^1-smooth map with β-Lipschitz gradient. Then F is $L\beta$-weakly convex. Furthermore, by the chain rule, $\partial F(x) = \nabla c\,(x)^\top \partial h\,(c\,(x))$.

Definition 4.1 (IWCNN)**.** By Theorem 4.2, we can therefore define an IWCNN by

$$f_\theta^{\text{IWCNN}} = g_{\theta_1}^{\text{ICNN}} \circ g_{\theta_2}^{\text{sm}},$$

where $g_{\theta_1}^{\text{ICNN}}$ is an ICNN and $g_{\theta_2}^{\text{sm}}$ is a neural network with smooth activations.

We note that the IWCNN construction is only necessary because of non-smooth activation functions (like `ReLU` or `leakyReLU`), which are used in both the AR and ACR. A neural network with all smooth activation functions, e.g. `sigmoid`, is automatically weakly convex. However, usage of smooth non-linearities has been shown to worsen performance in machine learning models [17] and furthermore using smooth activations in the original AR formulation turns out to harm performance.

4.3 Adversarial convex-nonconvex regularizer (ACNCR)

With the IWCNN in hand, we can now parametrize the learned regularizer as $\mathcal{R}_\theta^{cnc}$ defined as $\mathcal{R}_\theta^{cnc}(x, y) := \mathcal{R}_{\theta_1}^{c}(x) + \mathcal{R}_{\theta_2}^{wc}(y)$, for $\mathcal{R}_{\theta_1}^{c}$ parameterized in the same way as the ACR, while $\mathcal{R}_{\theta_2}^{wc}$ is parameterized using an IWCNN. Thus, we can view this approach as learning to denoise in both the data and observation domain. Note that this ACNCR is strictly more expressive than the ACR. Thus, denoting $\mathbb{P}_{Y_r} := (\mathcal{A})_{\#}\,\mathbb{P}_r$, we train both of the networks in a decoupled way by minimizing:

$$\begin{aligned}
&\mathbb{E}_{X\sim\mathbb{P}_r}\left[\mathcal{R}_{\theta_1}^{c}(X)\right] - \mathbb{E}_{X\sim\mathbb{P}_n}\left[\mathcal{R}_{\theta_1}^{c}(X)\right] \quad + \lambda\cdot\mathbb{E}\left[\left(\left\|\nabla_x\,\mathcal{R}_{\theta_1}^{c}(X)\right\| - 1\right)_+^2\right] \quad + \\
&\mathbb{E}_{Y\sim\mathbb{P}_{Y_r}}\left[\mathcal{R}_{\theta_2}^{wc}(Y)\right] - \mathbb{E}_{Y\sim\mathbb{P}_Y}\left[\mathcal{R}_{\theta_2}^{wc}(Y)\right] \quad + \lambda\cdot\mathbb{E}\left[\left(\left\|\nabla_y\,\mathcal{R}_{\theta_2}^{wc}(Y)\right\| - 1\right)_+^2\right].
\end{aligned}$$

In choosing this objective, we normalized the operator $\mathcal{A}$ to have norm 1. This is also done in experiments to further aid stable training of the networks.

5 Computed Tomography (CT) numerical experiments

To evaluate our ACNCR, we consider two applications: CT reconstruction with (i) sparse-view and (ii) limited-angle projection. For details on the experimental set-up, see Appendix B. Main results are shown in Table 1, with further visual examples in Appendix A.

Sparse view CT As in [21] performance of AR during reconstruction deteriorates if the network is over-trained and early stopping is not employed. For ACR this does not occur due to reduced expressivity, at a price of reduced performance as seen on Table 1. Akin to ACR, ACNCR overcomes this limitation and performs on par with AR, without over-training and without early stopping thanks to better expressivity.

Limited view CT Reconstruction from limited-angle projection data, with no measurement in a specific angular region, is an inverse problem with a severely ill-posed forward operator where the reconstruction performance depends critically on the image prior. One of the main benefits of imposing convexity on the regularizer is the improved performance in the limited-angle setting as compared to AR, wherein, even with early stopping, artifacts arise in the reconstructions. ACNCR overcomes this issue without having to employ early stopping, while also performing on par with ACR and outperforming both model-based approaches and AR as seen from Table 1.

6 Conclusion

In this work, we have shown that a CNC regularizer defined as the sum of a weakly convex function over the data space and a convex function over the model/parameter space exhibits existence of solutions, stability, convergent regularization, and convergence of subgradient descent. Furthermore, through our novel IWCNN construction, we have shown how to learn such a regularizer adversarially. In CT experiments, this ACNCR is observed to better adapt to both sparse and limited-angle settings, showing that it is more versatile with respect to ill-posedness than the AR and ACR approaches.

A Visualization of experimental results

Table 1: Average PSNR and SSIM over test data in CT experiments.

	Limited			Sparse		
Methods	PSNR (dB)	SSIM	# param.	PSNR (dB)	SSIM	# param.
FBP	17.1949	0.1852	1	21.0157	0.1877	1
TV	**25.6778**	**0.7934**	1	**31.7619**	**0.8883**	1
LPD	28.9480	**0.8394**	127 370	**37.4868**	0.9217	700 180
FBP + U-Net	**29.1103**	0.8067	14 787 777	37.1075	**0.9265**	14 787 777
AR	23.6475	0.6257	133 792	36.4079	**0.9101**	33 952 481
ACR	26.4459	**0.8184**	34 897	34.5844	0.8765	9 448
ACNCR	**26.5420**	0.8161	1 085 448	35.6476	0.9094	1 085 448

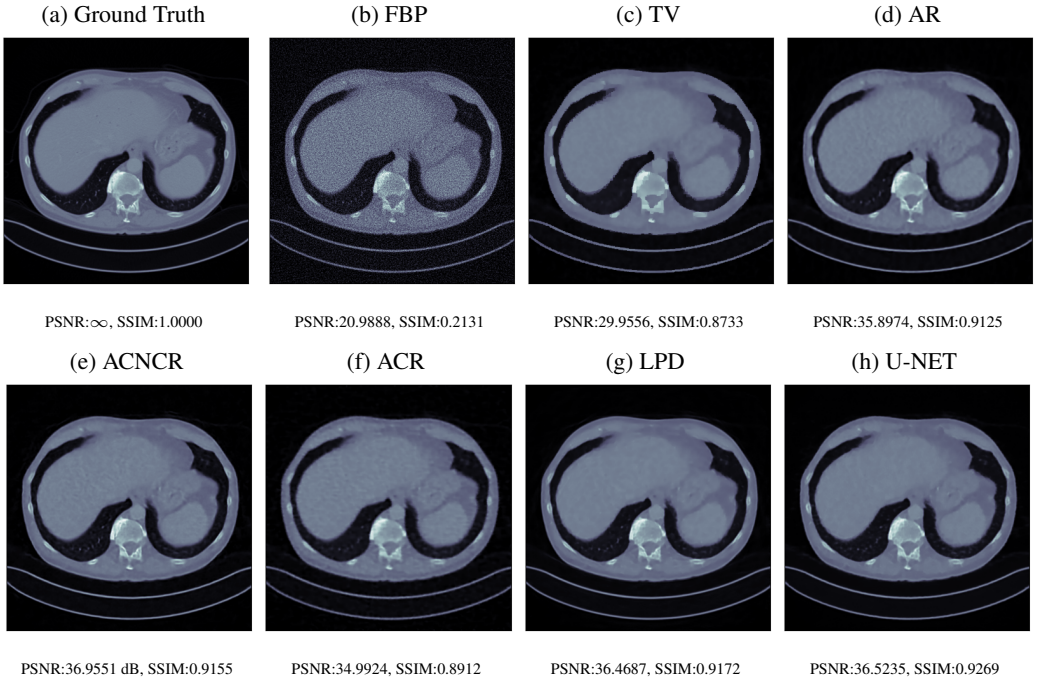

Figure 1: Reconstructed images obtained using different methods, along with the associated PSNR and SSIM, for sparse view CT. In this case, the ACNCR is shown to outperform the AR in terms of SSIM and PSNR.

B Experimental set-up

For experiments, human abdominal CT scans for 10 patients in the Mayo Clinic low-dose CT grand challenge dataset [22] are used. We simulate the projection data in ODL [1]. Our training dataset for the CT experiments consists of 2250 2D slices of size 512×512 corresponding to nine patients, and the slices extracted from the remaining one patient are used for evaluation.

Our ACNCR method is compared with two model-based techniques: filtered back-projection (FBP) and total variation (TV) regularization; two supervised data-driven methods: the learned primal-dual (LPD) method [2] and UNet-based post-processing of FBP [14]; and two adversarial regularization approaches: the AR and the ACR. In comparing with these we illustrate the trade-off in levels of constraints versus stability and performance. The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [39] are used as quality metrics. For fairness, we report the highest PSNR achieved by all methods during reconstruction. Results are displayed in Appendix A.

LPD is trained on pairs of target images and projection data, whereas the U-net post-processor is trained on pairs of true images and the corresponding FBP. AR, ACR and AWCR, in contrast, require ground-truth and FBP images drawn from their marginal distributions (and hence not necessarily paired). The hyperparameters λ and ρ_0 are chosen in the same as done in [25].

For both CT experiments, projection data is simulated using a parallel-beam acquisition geometry with 350 angles and 700 rays/angle, using additive Gaussian noise with $\sigma = 3.2$. The pseudoinverse reconstruction is taken to be images obtained using FBP. For limited angle experiments, data is simulated with a missing angular wedge of 60°. The native `odl` power method is used to approximate the norm of the operator. For all three adversarial regularisation methods a fixed number of steps of accelerated gradient descent is performed and for consistency best PSNR value images are reported in Table 1.

For the IWCNN Section 4.2 architecture, the usual ICNN [4] architecture using `leakyReLU` activations is used for $g^{\text{ICNN}}_{\theta_1}$, while the smooth network $g^{\text{sm}}_{\theta_2}$ is parameterised a deep convolutional network with `SiLU` activations and 5 layers. The *RMSprop* optimizer (following [21]) with a learning rate of 1×10^{-4} is used for training.

C Proof of Theorem 4.1

Proof. If f is convex, then by definition it is weakly convex.

Suppose that f is weakly convex. As f is convex if and only if $f|_\ell$ is convex for all lines $\ell \subseteq \mathbb{R}^d$, and if f is weakly convex then so is $f|_\ell$ for all lines $\ell \subseteq \mathbb{R}^d$, it suffices to prove the theorem for $d = 1$.

Since $f : \mathbb{R} \to \mathbb{R}$ is piecewise linear with finitely many pieces, there exist $(a_i)_{i=1}^{n+1}$, $(b_i)_{i=1}^{n+1}$, and $(x_i)_{i=1}^{n}$ such that the x_i are monotonically increasing and, defining $x_0 = -\infty$ and $x_{n+1} = \infty$,

$$f(x) = \{a_i x + b_i, \quad x \in (x_{i-1}, x_i],\ i \in \{1, 2, ..., n+1\}.$$

Furthermore, by continuity, for all $i \in \{1, ..., n\}$, $a_i x_i + b_i = a_{i+1} x_i + b_{i+1}$.

Since f is weakly convex, there exists $\rho > 0$ such that, by (3), for all $\varepsilon > 0$ and $i \in \{1, ..., n\}$,

$$f(x_i) = f\left(\frac{1}{2}(x_i - \varepsilon) + \frac{1}{2}(x_i + \varepsilon)\right) \leq \frac{1}{2} f(x_i - \varepsilon) + \frac{1}{2} f(x_i + \varepsilon) + \frac{1}{2} \times \frac{1}{2} \rho (2\varepsilon)^2.$$

For $\varepsilon < \min\{x_i - x_{i-1}, x_{i+1} - x_i\}$, this becomes

$$a_i x_i + b_i \leq \frac{1}{2}(a_i(x_i - \varepsilon) + b_i) + \frac{1}{2}(a_{i+1}(x_i + \varepsilon) + b_{i+1}) + \rho\varepsilon^2$$

which simplifies to (using that $a_i x_i + b_i = a_{i+1} x_i + b_{i+1}$)

$$0 \leq \frac{1}{2}\varepsilon(a_{i+1} - a_i) + \rho\varepsilon^2, \qquad \text{and therefore} \qquad a_{i+1} - a_i \geq -2\rho\varepsilon,$$

for all sufficiently small $\varepsilon > 0$. Hence, $a_{i+1} \geq a_i$ for all $i \in \{1, .., n\}$.

We make the following claim:

$$f(x) = \max_{j \in \{1, ..., n+1\}} a_j x + b_j =: f_j(x). \tag{6}$$

To prove (6), let $x \in (x_{i-1}, x_i]$, and hence $f(x) = f_i(x)$. Note that for all j, $f_j(x_j) = f_{j+1}(x_j)$ and that for all j and y, $f_j'(y) = a_j \leq a_{j+1} = f_{j+1}'(y)$. Hence:

- For all j and $y \geq x_{j-1}$, $f_j(y) \geq f_{j-1}(y)$.
- For all j and $y \leq x_j$, $f_j(y) \geq f_{j+1}(y)$.

It follows that:

- For all $j < i$ and $y \geq x_{i-1}$, $f_i(y) \geq f_j(y)$.
- For all $j > i$ and $y \leq x_i$, $f_i(y) \geq f_j(y)$.

Since $y = x$ satisfies both conditions, we have that for all $j \neq i$, $f_i(x) \geq f_j(x)$, as desired. Finally, from (6) it immediately follows that f is convex, as it is the pointwise maximum of affine (and therefore convex) functions. □

D Banach composition weak convexity

Theorem D.1 (Special case of [35] Proposition 10.21(c))**.** Let U open and V be nonempty convex sets of two normed spaces X and Y respectively, let $f : V \to \bar{\mathbb{R}}$ be finite, convex and K_f-Lipschitz continuous on V. Let $F : U \to Y$ be a differentiable mapping with $F(U) \subset V$ and such that DF is uniformly continuous on U with a linear modulus of continuity (i.e. Lipschitz continuous) with coefficient K_F on U, then $f \circ F$ is $K_f K_F$-weakly convex.

References

[1] J. Adler, H. Kohr, and Ozan Öktem. Operator discretization library (ODL). *GitHub repository*, 2017. URL https://github.com/odlgroup/odl.

[2] Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE transactions on medical imaging*, 37(6):1322–1332, 2018.

[3] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[4] Brandon Amos, Lei Xu, and J Zico Kolter. Input convex neural networks. In *International Conference on Machine Learning*, pages 146–155, 2017.

[5] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/arjovsky17a.html.

[6] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019.

[7] Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98, 2016.

[8] Damek Davis, Dmitriy Drusvyatskiy, Kellie J. MacPhee, and Courtney Paquette. Subgradient methods for sharp weakly convex functions. *Journal of Optimization Theory and Applications*, 179(3):962–982, Dec 2018. ISSN 1573-2878. doi: 10.1007/s10957-018-1372-8.

[9] Alexandros G. Dimakis. *Deep Generative Models and Inverse Problems*, page 400–421. Cambridge University Press, 2022. doi: 10.1017/9781009025096.010.

[10] Sergey Fomel. On model-space and data-space regularization: A tutorial. *SEP-94: Stanford Exploration Project*, pages 141–164, 1997.

[11] Alexis Goujon, Sebastian Neumayer, and Michael Unser. Learning weakly convex regularizers for convergent image-reconstruction algorithms, 2023.

[12] Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Gradient step denoiser for convergent plug-and-play. In *International Conference on Learning Representations*, 2022.

[13] Samuel Hurault, Arthur Leclaire, and Nicolas Papadakis. Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 9483–9505. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/hurault22a.html.

[14] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[15] Ulugbek S Kamilov, Charles A Bouman, Gregery T Buzzard, and Brendt Wohlberg. Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 40(1):85–97, 2023.

[16] Erich Kobler, Alexander Effland, Karl Kunisch, and Thomas Pock. Total deep variation for linear inverse problems. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7549–7558, 2020.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[18] Alessandro Lanza, Serena Morigi, Ivan W Selesnick, and Fiorella Sgallari. Convex non-convex variational models. In *Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision*, pages 1–57. Springer, 2022.

[19] Oscar Leong, Eliza O'Reilly, Yong Sheng Soh, and Venkat Chandrasekaran. Optimal convex and nonconvex regularizers for a data source, 2022.

[20] Housen Li, Johannes Schwab, Stephan Antholzer, and Markus Haltmeier. NETT: solving inverse problems with deep neural networks. *Inverse Problems*, 36(6):065005, 2020. doi: 10.1088/1361-6420/ab6d57.

[21] Sebastian Lunz, Ozan Öktem, and Carola-Bibiane Schönlieb. Adversarial regularizers in inverse problems. In *Advances in Neural Information Processing Systems*, pages 8507–8516, 2018.

[22] C. McCollough. TU-FG-207A-04: Overview of the Low Dose CT Grand Challenge. *Medical Physics*, 43:3759–3760, 2016. doi: 10.1118/1.4957556.

[23] Tristan Milne, Étienne Bilocq, and Adrian Nachman. A new method for determining Wasserstein 1 optimal transport maps from Kantorovich potentials, with deep learning applications, 2022.

[24] Hosein Mohimani, Massoud Babaie-Zadeh, and Christian Jutten. A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm. *IEEE Transactions on Signal Processing*, 57(1):289–301, 2009. doi: 10.1109/TSP.2008.2007606.

[25] S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C.-B. Schönlieb. Learned convex regularizers for inverse problems. *arXiv preprint arXiv:2008.02839v2*, 2021.

[26] Subhadip Mukherjee, Marcello Carioni, Ozan Öktem, and Carola-Bibiane Schönlieb. End-to-end reconstruction meets data-driven regularization for inverse problems. In *Advances in Neural Information Processing Systems*, volume 34, pages 21413–21425, 2021.

[27] Subhadip Mukherjee, Andreas Hauptmann, Ozan Öktem, Marcelo Pereyra, and Carola-Bibiane Schönlieb. Learned reconstruction methods with convergence guarantees: A survey of concepts and applications. *IEEE Signal Processing Magazine*, 40(1):164–182, 2023. doi: 10.1109/MSP.2022.3207451. URL https://ieeexplore.ieee.org/abstract/document/10004773.

[28] Pei Peng, Shirin Jalali, and Xin Yuan. Auto-encoders for compressed sensing. In *NeurIPS 2019 Workshop on Solving Inverse Problems with Deep Networks*, 2019.

[29] Konstantin Pieper and Armenak Petrosyan. Nonconvex regularization for sparse neural networks. *Applied and Computational Harmonic Analysis*, 61:25–56, 2022. ISSN 1063-5203. doi: 10.1016/j.acha.2022.05.003. URL https://www.sciencedirect.com/science/article/pii/S1063520322000434.

[30] C. Pöschl. An overview on convergence rates for tikhonov regularization methods for non-linear operators. *Journal of Inverse and Ill-posed Problems*, 17(1):77–83, 2009. doi: 10.1515/JIIP.2009.009.

[31] E. T. Reehorst and P. Schniter. Regularization by denoising: clarifications and new interpretations. *IEEE Transactions on Computational Imaging*, 5(1):52–67, 2019.

[32] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

[33] Stefan Roth and Michael J Black. Fields of experts. *International Journal of Computer Vision*, 82:205–229, 2009.

[34] Ernest Ryu, Jialin Liu, Sicheng Wang, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. Plug-and-play methods provably converge with properly trained denoisers. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 5546–5557. PMLR, 09–15 Jun 2019.

[35] Lionel Thibault. *Unilateral variational analysis in Banach spaces*. World Scientific, 2021.

[36] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[37] Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948, 2013. doi: 10.1109/GlobalSIP.2013.6737048.

[38] Huayu Wang, Chen Luo, Taofeng Xie, Qiyu Jin, Guoqing Chen, Zhuo-Xu Cui, and Dong Liang. Convex latent-optimized adversarial regularizers for imaging inverse problems, 2023.

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.