REGULARIZATION USING DENOISING: EXACT AND ROBUST SIGNAL RECOVERY

Ruturaj G. Gavaskar and Kunal N. Chaudhury

Department of Electrical Engineering, Indian Institute of Science, Bengaluru, India

ABSTRACT

We consider the problem of signal reconstruction from linearly corrupted data using plug-and-play (PnP) regularization. As opposed to traditional sparsity-promoting regularizers, PnP uses an off-the-shelf denoiser within a proximal algorithm such as ISTA or ADMM for image reconstruction. Although PnP has become popular in the imaging community, its regularization capacity is not fully understood. For example, it is not known if PnP can in theory recover a signal from few noiseless measurements as in classical compressed sensing and if the recovery is robust. We explore these questions in this work and present some theoretical and experimental results. In particular, we prove that if the denoiser in question has low rank and if the ground-truth lies in the range of the denoiser, then it can be recovered exactly from noiseless measurements. To the best of knowledge, this is first such result. Furthermore, we show using numerical simulations that even if the aforementioned conditions are violated, PnP recovery is robust in practice. We formulate a theorem regarding the recovery error based on these observations.

***Index Terms*—** Image regularization, plug-and-play, robust recovery, compressed sensing, inpainting.

1. INTRODUCTION

In variational signal reconstruction, an unknown signal $\boldsymbol{x}_0 \in \mathbb{R}^n$ is estimated from partial noisy measurements $\boldsymbol{y} \in \mathbb{R}^m$. This is done by minimizing the sum of a data-fidelity function $f(\boldsymbol{x})$ and a regularizer $\Phi(\boldsymbol{x})$. As is well-known, this can be viewed as a maximum a-posteriori estimation [1], where $f(\boldsymbol{x})$ is the negative log-likelihood and $\Phi(\boldsymbol{x})$ is the negative log of some prior on $\boldsymbol{x}_0$. We focus on linear inverse problems in this work, where the forward model is given by

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x}_0 + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a measurement operator and $\boldsymbol{\eta} \in \mathbb{R}^m$ is Gaussian noise. The data fidelity in this case is $f(\boldsymbol{x}) = \|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2$.

For computational tractability, Φ is taken to be convex. Moreover, for popular choices of Φ such as total-variation and its variants [2], $f + \Phi$ can be efficiently minimized using iterative algorithms such as ADMM and ISTA [3]. These algorithms crucially rely on the fact that the *proximal map* of Φ,

$$\operatorname{prox}_{\Phi}(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{u}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{x}\|_2^2 + \Phi(\boldsymbol{u}), \tag{2}$$

can be computed efficiently in closed-form or numerically. From a signal processing perspective, $\operatorname{prox}_{\Phi}(\boldsymbol{x})$ is the denoising of $\boldsymbol{x}$ using regularizer Φ [1]. Motivated by this observation, a novel form of regularization, called *plug-and-play* (PnP), was proposed in [4,5]. In this framework, the proximal map within ISTA or ADMM is replaced by a Gaussian denoiser such as BM3D [6] or DnCNN [7], which are much more powerful than the proximal maps of commonly used convex regularizers. More precisely, (2) is replaced by $D(\boldsymbol{x})$ within a proximal algorithm, where $D : \mathbb{R}^n \to \mathbb{R}^n$ is the functional representation of the denoiser. The idea is to directly deploy the denoiser instead of having to specify Φ and go through its proximal map—indeed, it is not known if complex denoisers such as BM3D are the proximal map of some Φ. While PnP has been shown to work well for several imaging problems [8,9], this comes with a price, namely, it is difficult to interpret PnP as a regularization mechanism—minimization of $f(\boldsymbol{x}) + \Phi(\boldsymbol{x})$—and this makes it difficult to understand its restoration capacity.

In a series of recent works [5, 10–13], it was shown that an explicit regularizer Φ can be associated with specific linear denoisers. By linear denoiser, we mean that the denoiser can be expressed as $D(\boldsymbol{x}) = \mathbf{W}\boldsymbol{x}$ for some matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$. The essence of the results in the aforementioned papers is that if $\mathbf{W}$ satisfies a specific set of properties, then we can construct a convex function $\Phi_{\mathbf{W}}$ (described in Section 2) such that $D(\boldsymbol{x}) = \operatorname{prox}_{\Phi_{\mathbf{W}}}(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$. In this case, the reconstruction $\boldsymbol{x}^*$ obtained using PnP regularization is a minimizer of $f(\boldsymbol{x}) + \lambda\Phi_{\mathbf{W}}(\boldsymbol{x})$, where λ is a penalty parameter. In contrast, it is not known whether any nonlinear denoiser can be expressed as a proximal map.

While the ad-hoc idea of replacing the proximal map by a denoiser is intuitive, it is however surprising that PnP should work in the first place. Understanding why PnP works so well in practice is a difficult question and seems beyond reach at this point. Instead, as a first step, we look at simple questions that are mathematically tractable. For example, can we come up with conditions on the ground-truth $\boldsymbol{x}_0$, the forward operator $\mathbf{A}$ and the denoiser $\mathbf{W}$ that can guarantee exact recovery? Can we establish stable recovery in the presence of noise? By stable, we mean that the reconstruction error is proportional to the noise in the model or measurements [14].

To address these questions, we restrict ourselves to linear inverse problems. Moreover, to leverage the explicit formula of the associated regularizer $\Phi_{\mathbf{W}}$, we work with linear denoisers. In this setting, the PnP reconstruction is given by

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} \|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda\Phi_{\mathbf{W}}(\boldsymbol{x}). \tag{3}$$

For the analysis, we work with the constrained formulation:

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} \Phi_{\mathbf{W}}(\boldsymbol{x}) \quad \text{subject to} \quad \|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|_2 \leqslant \delta, \tag{4}$$

where δ is set proportional to the noise level of $\boldsymbol{\eta}$ in (1). In particular, for noiseless measurements ($\boldsymbol{\eta} = \mathbf{0}$), we set $\delta = 0$ and consider the problem:

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} \Phi_{\mathbf{W}}(\boldsymbol{x}) \quad \text{subject to} \quad \mathbf{A}\boldsymbol{x} = \boldsymbol{y}. \tag{5}$$

As opposed to the soft constraint imposed by the penalty function in (3), the hard constraint in (4) makes it easier to analyze. This switch

Correspondence: ruturajg@iisc.ac.in, kunal@iisc.ac.in. K. N. Chaudhury was supported by Core Research Grant CRG/2020/000527 and SERB-STAR Award STR/2021/000011 from Department of Science and Technology, Government of India.

from the penalty to the constrained formulation is used in compressed sensing literature [14,15]. In fact, problems (3) and (4) are equivalent for appropriate choices of δ and λ [16].

The main contributions of this work are as follows:

1. We formulate sufficient conditions on the ground-truth, denoiser, and forward operator under which we can guarantee that $\boldsymbol{x}^* = \boldsymbol{x}_0$ (exact recovery), where $\boldsymbol{x}^*$ is obtained from (5). We specialize this result to the compressed sensing problem, which results in a probabilistic guarantee on exact recovery.
2. We numerically validate that the sufficient conditions indeed result in exact signal reconstruction. Moreover, we present numerical evidence that even if the measurements are noisy ($\boldsymbol{\eta} \neq \mathbf{0}$) and even if the conditions are violated, the recovery error $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2$ is proportional to the noise floor if we use (4) to obtain $\boldsymbol{x}^*$ (robust recovery). We formalize this as a theorem.

2. EXACT RECOVERY

Notation. *We denote the null space and range space of a matrix* $\mathbf{X}$ *by* $\mathcal{N}(\mathbf{X})$ *and* $\mathcal{R}(\mathbf{X})$. *The pseudoinverse of* $\mathbf{X}$ *is denoted by* $\mathbf{X}^\dagger$.

The observation that a linear denoiser $\mathbf{W}$ is (under some conditions) the proximal map of a convex function $\Phi_{\mathbf{W}}$ was first made in [5]. An explicit formula for $\Phi_{\mathbf{W}}$ was provided in [10, 11] for symmetric $\mathbf{W}$; this was later generalized to non-symmetric denoisers in [12, 13]. To avoid certain technicalities specific to non-symmetric denoisers, we will work with symmetric denoisers. In this case, the formula for the extended-valued regularizer $\Phi_{\mathbf{W}} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is as follows.

Theorem 1. *Let* $\mathbf{W} \in \mathbb{R}^{n \times n}$ *be symmetric positive semidefinite with eigenvalues in* $[0, 1]$. *Define*

$$\Phi_{\mathbf{W}}(\boldsymbol{x}) = \begin{cases} \frac{1}{2}\boldsymbol{x}^\top(\mathbf{I} - \mathbf{W})\mathbf{W}^\dagger\boldsymbol{x}, & \text{if } \boldsymbol{x} \in \mathcal{R}(\mathbf{W}), \\ \infty, & \text{otherwise.} \end{cases} \tag{6}$$

Then $\mathbf{W}\boldsymbol{x} = \operatorname{prox}_{\Phi_{\mathbf{W}}}(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$.

We note that the expression for $\Phi_{\mathbf{W}}$ is given differently in [10,11], but they are in fact equivalent to the compact expression in (6). It can be shown that $(\mathbf{I} - \mathbf{W})\mathbf{W}^\dagger$ is symmetric positive semidefinite if $\mathbf{W}$ satisfies the conditions in Theorem 1, so that $\Phi_{\mathbf{W}}$ is convex. There are several linear denoisers that satisfy the conditions in Theorem 1. This includes kernel filters such as nonlocal means (NLM) [6] and the bilateral filter [17]. Although $\mathbf{W}$ is generally non-symmetric for these filters, they can be well-approximated by a symmetric positive semidefinite matrix with eigenvalues in $[0, 1]$; e.g., see [18]. Additionally, the symmetric approximation $\mathbf{W}$ is doubly stochastic.

We consider the question of exact recovery: If $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}_0$ (noiseless measurements), can we guarantee that $\boldsymbol{x}_0$ is a unique minimizer of (5)? We first consider the case where $\mathbf{W}$ is invertible. It was shown in [5, 12] that if $\mathbf{W}$ corresponds to a kernel filter (such as nonlocal means), then $\mathbf{W}$ is typically invertible and irreducible. In this case, we have $\mathcal{R}(\mathbf{W}) = \mathbb{R}^n$ and $\mathbf{W}^\dagger = \mathbf{W}^{-1}$, so that the formula in (6) becomes $\Phi_{\mathbf{W}}(\boldsymbol{x}) = 0.5\boldsymbol{x}^\top(\mathbf{W}^{-1} - \mathbf{I})\boldsymbol{x}$. In this case, we have the following impossibility result. (The proof is technical and is deferred to a future paper [19].)

Proposition 2. *Let* $\mathbf{W}$ *be a kernel filter that is invertible and irreducible. Then* $\boldsymbol{x}_0$ *is a minimizer of* (5) *if and only if the vector* $(\mathbf{W}^{-1} - \mathbf{I})\boldsymbol{x}_0$ *is orthogonal to* $\mathcal{N}(\mathbf{A})$. *In particular, if* $\mathbf{A}$ *is a random matrix with i.i.d. Gaussian entries (as in compressed sensing [14]) and* $\boldsymbol{x}_0$ *is a non-constant signal, then the probability that* $\boldsymbol{x}_0$ *is a minimizer of* (5) *is* 0.

In other words, except for the uninteresting case where $\boldsymbol{x}_0$ is a constant signal, we cannot recover the original signal from compressive measurements if $\mathbf{W}$ has full rank. Is exact recovery possible for non-trivial signals if $\mathbf{W}$ is constrained to be low rank? In case of symmetric kernel filters, it was shown in [18] that if the input image (which is used to compute the kernel matrix) does not contain repetitive textures, then the eigenvalues of $\mathbf{W}$ decay rapidly to zero. Therefore, we expect that $\mathbf{W}$ can be well-approximated in practice by a low-rank matrix. It turns out that (under certain conditions) we can exactly recover $\boldsymbol{x}_0$ using a low-rank denoiser.

Theorem 3. *Let* $\mathbf{W}$ *be such that* $\operatorname{rank}(\mathbf{W}) \leqslant m$. *Suppose that*

1. $\boldsymbol{x}_0 \in \mathcal{R}(\mathbf{W})$.
2. *The restriction of* $\mathbf{A}$ *to* $\mathcal{R}(\mathbf{W})$ *is one-to-one.*

Then $\boldsymbol{x}_0$ *is the unique solution of* (5).

Proof. From the definition of $\Phi_{\mathbf{W}}$, we can write (5) as

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \boldsymbol{x}^\top(\mathbf{I} - \mathbf{W})\mathbf{W}^\dagger\boldsymbol{x} \quad \text{s.t.} \quad \boldsymbol{x} \in \mathcal{R}(\mathbf{W}),\ \mathbf{A}\boldsymbol{x} = \mathbf{A}\boldsymbol{x}_0. \tag{7}$$

We claim that $\boldsymbol{x}_0$ is the only feasible solution and hence the unique minimizer of (7). Indeed, if $\boldsymbol{x}$ is feasible, then it is clear that $\boldsymbol{x} - \boldsymbol{x}_0 \in \mathcal{R}(\mathbf{W}) \cap \mathcal{N}(\mathbf{A})$. However, note that the second assumption is equivalent to $\mathcal{R}(\mathbf{W}) \cap \mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$, so that $\boldsymbol{x} = \boldsymbol{x}_0$. □

As far as the assumptions in Theorem 3 are concerned, we have the following remarks:

1. The condition $\boldsymbol{x}_0 \in \mathcal{R}(\mathbf{W})$ cannot be relaxed since this is necessary for exact recovery. Indeed, since the feasible set in (7) is a subset of $\mathcal{R}(\mathbf{W})$, $\boldsymbol{x}_0$ cannot be a solution if $\boldsymbol{x}_0 \notin \mathcal{R}(\mathbf{W})$.
2. In most applications, we have $m < n$, so that the dimension of $\mathcal{N}(\mathbf{A})$ is $\geqslant n - m > 0$. It follows that if the dimension of $\mathcal{R}(\mathbf{W})$ is $> m$, then $\mathcal{R}(\mathbf{W}) \cap \mathcal{N}(\mathbf{A}) \neq \{\mathbf{0}\}$. Therefore, the condition $\operatorname{rank}(\mathbf{W}) \leqslant m$ is in fact necessary for the restriction of $\mathbf{A}$ to $\mathcal{R}(\mathbf{W})$ to be one-to-one.

As a corollary of Theorem 3, we have the following exact recovery result for compressed sensing where $\mathbf{A}$ is a random matrix.

Proposition 4. *Let* $\operatorname{rank}(\mathbf{W}) \leqslant m$ *and* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *be a random matrix with i.i.d. Gaussian entries. Then the restriction of* $\mathbf{A}$ *to* $\mathcal{R}(\mathbf{W})$ *is one-to-one with probability* 1. *Consequently, if* $\boldsymbol{x}_0 \in \mathcal{R}(\mathbf{W})$, *then* $\boldsymbol{x}_0$ *can be exactly recovered with probability* 1 *by solving* (5).

We skip the formal proof of this result but explain the main idea. Since the entries of $\mathbf{A}$ are random, $\mathbf{A}$ has full row rank with probability 1. Thus, $\mathcal{N}(\mathbf{A})$ is almost surely a subspace with dimension $n - m$. On the other hand, since $\dim \mathcal{R}(\mathbf{W}) \leqslant m$, we have

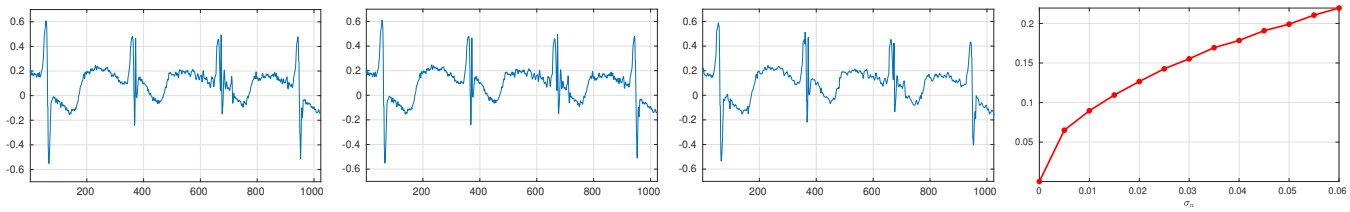
$$\dim \mathcal{R}(\mathbf{W}) + \dim \mathcal{N}(\mathbf{A}) \leqslant n$$

almost surely. Now, it can be shown that if the sum of the dimensions of two subspaces of $\mathbb{R}^n$ is $\leqslant n$, then their intersection is almost surely $\{\mathbf{0}\}$. Hence, $\mathcal{R}(\mathbf{W}) \cap \mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$ almost surely, from which the result follows.

Note that there are two sources of error that can violate the hypothesis of Theorem 3 in practice:

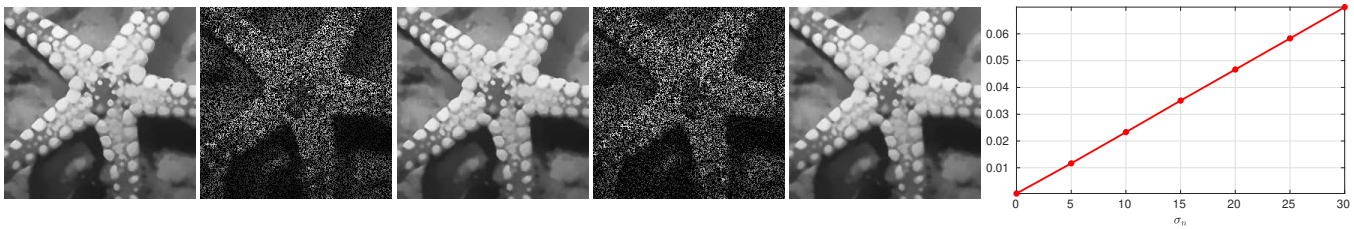
- **Model noise:** The condition $\boldsymbol{x}_0 \in \mathcal{R}(\mathbf{W})$ may not hold. We refer to this as *model noise*.
- **Measurement noise:** We refer to the additive Gaussian noise $\boldsymbol{\eta}$ as *measurement noise*.

Theorem 3 is applicable only when both model noise and measurement noise are absent. However, in the next section, we present numerical results which suggest that PnP regularization can recover the signal robustly in the presence of noise. For our experiments, we



(a) Ground-truth, $\boldsymbol{x}_0$. (b) $\boldsymbol{x}^*$ (noiseless). (c) $\boldsymbol{x}^*$ ($\sigma_n = 0.01$). (d) Relative error.

Fig. 1. CS reconstruction of a signal of length 1024 from 512 random Gaussian measurements. The signal $\boldsymbol{x}_0$ and denoiser $\mathbf{W}$ are chosen such that the conditions in Theorem 3 are satisfied. The relative error $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2 \,/\, \|\boldsymbol{x}_0\|_2$ is as follows: (b) 1×10^{-7} and (c) 0.088.



(a) Ground-truth, $\boldsymbol{x}_0$. (b) $\boldsymbol{y}$ (noiseless). (c) $\boldsymbol{x}^*$ (noiseless). (d) $\boldsymbol{y}$ ($\sigma_n = 30$). (e) $\boldsymbol{x}^*$ ($\sigma_n = 30$). (f) Relative error.

Fig. 2. Image inpainting from about 60% missing pixels. The image size is 256×256. The signal $\boldsymbol{x}_0$ and denoiser $\mathbf{W}$ are chosen such that the two conditions in Theorem 3 are satisfied. The relative error is as follows: (c) 2×10^{-8} and (e) 0.021.

use the DSG-NLM filter [5] (a symmetrized variant of NLM), the GLIDE filter [20], and a graph signal denoiser proposed in [21]. All these denoisers satisfy the condition in Theorem 1, so that $\Phi_{\mathbf{W}}$ is well-defined and PnP regularization amounts to solving (4).

3. EXPERIMENTS

3.1. Reconstruction in the absence of model noise

In this section, we show results for the case where the only source of noise is the measurement noise $\boldsymbol{\eta}$. In particular, we ensure that both conditions in Theorem 3 are satisfied.

Compressed Sensing. We take $\mathbf{A}$ to be a random matrix with i.i.d. Gaussian entries having mean 0 and variance $1/m$. This corresponds to a widely used setting in compressed sensing (CS) [14, 15]. We consider the case where $\boldsymbol{x}_0$ is an ECG signal of length $n = 1024$ [22]. We set $m = 512$, corresponding to 50% compression.

The denoiser $\mathbf{W}$ is constructed as follows. First, we construct the matrix $\overline{\mathbf{W}} \in \mathbb{R}^{n \times n}$ for the DSG-NLM filter [5], which is symmetric positive semidefinite with eigenvalues in $[0, 1]$. (DSG-NLM was originally proposed for images, but it can be easily adapted to one-dimensional signals.) We then take $\mathbf{W}$ to be the best low-rank approximation of $\overline{\mathbf{W}}$ with rank m, which is found using the singular value decomposition (SVD) of $\overline{\mathbf{W}}$. In order to ensure that the restriction of $\mathbf{A}$ to $\mathcal{R}(\mathbf{W})$ is one-to-one, we randomly generate $\mathbf{A}$ until it satisfies the property that the smallest singular value of $\mathbf{AU}$ is strictly positive, where $\mathbf{U} \in \mathbb{R}^{n \times m}$ is a matrix whose columns form a basis of $\mathcal{R}(\mathbf{W})$. Finally, to ensure that $\boldsymbol{x}_0 \in \mathcal{R}(\mathbf{W})$, we generate $\boldsymbol{x}_0$ by applying $\mathbf{W}$ to a real-world ECG signal taken from the MIT-BIH Arrhythmia Database [23]; see Fig. 1(a). Thus, both the conditions in Theorem 3 are satisfied.

For the case $\boldsymbol{\eta} = \mathbf{0}$, Fig. 1(b) shows the recovered signal $\boldsymbol{x}^*$ obtained by solving (5). This is exactly $\boldsymbol{x}_0$, which validates Theorem 3. Next, we add Gaussian noise $\boldsymbol{\eta}$ with standard deviation $\sigma_n = 0.01$ to the measurements, and solve (4) to obtain $\boldsymbol{x}^*$ (with $\delta = \|\boldsymbol{\eta}\|_2$). The reconstruction $\boldsymbol{x}^*$ for this case is shown in Fig. 1(c). Although $\boldsymbol{x}^*$ is not equal to $\boldsymbol{x}_0$, the error $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2$ is small relative to $\|\boldsymbol{x}_0\|_2$. We repeat this experiment for $\sigma_n = \{0.005, 0.01, \ldots, 0.06\}$, and plot the values of the relative error (averaged over 100 random trials) in Fig. 1(d). We use CVX [24] to solve (4) and (5).

Image Inpainting. We now consider the image inpainting problem where $\mathbf{A}$ is a selection of m random rows of the $n \times n$ identity matrix. We consider an image of size 256×256, so that $n = 256^2$. We set $m/n \approx 0.4$ (60% compression). We synthesize $\boldsymbol{x}_0$ as in the previous experiment. The only difference is in the choice of denoiser. Instead of DSG-NLM, we choose the GLIDE filter [20] as our denoiser $\mathbf{W}$. GLIDE is low-rank by construction (so we do not separately need to take a low-rank approximation); its rank is user-configurable and we set it equal to 200.

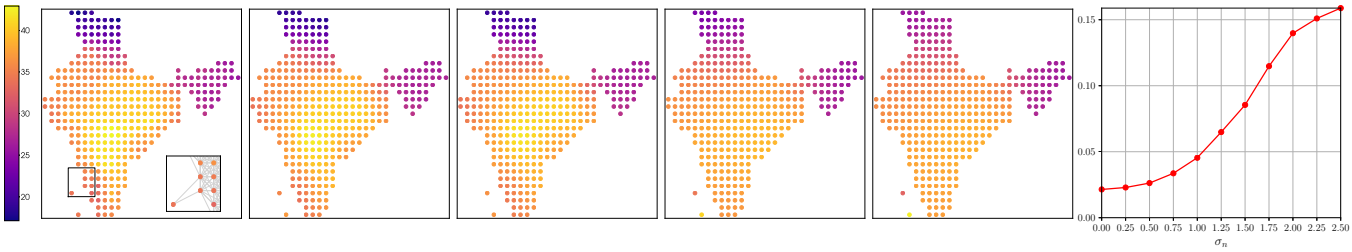
In Fig. 2, we show the results separately for $\sigma_n = 0$ (no measurement noise) and $\sigma_n = 30$. (The pixel intensities are in the range $[0, 255]$.) As predicted by Theorem 3, for $\sigma_n = 0$ we have $\boldsymbol{x}^* = \boldsymbol{x}_0$. However, the error $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2$ is small even for $\sigma_n = 30$. As before, we show a plot of $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2$ (relative to $\|\boldsymbol{x}_0\|_2$) as a function of σ_n in Fig. 2(f). We used the C-SALSA algorithm [25] to solve problem (4) in this experiment. Note that the error in Fig. 1(d) and Fig. 2(f) is approximately affine in σ_n. This seems to be the pattern in general with other experiments (not reported here).

3.2. Reconstruction in the presence of model noise

We now show experiments with model noise, i.e. $\boldsymbol{x}_0 \notin \mathcal{R}(\mathbf{W})$, in addition to measurement noise. This is achieved simply by taking some real-world signal $\boldsymbol{x}_0$, instead of synthesizing one via the application of $\mathbf{W}$ as done in the previous section.

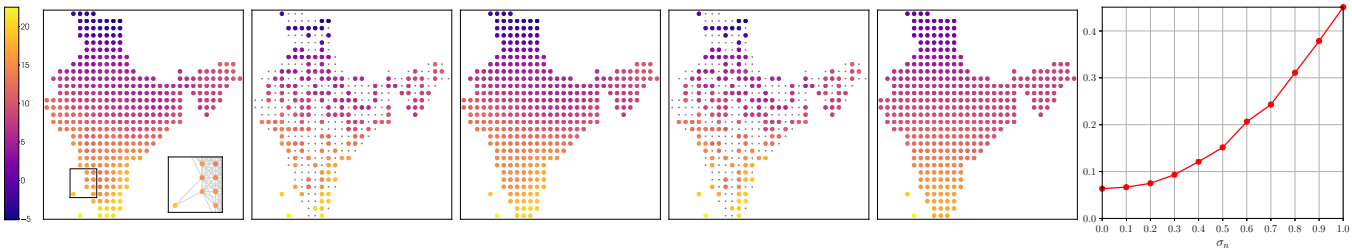
Compressed Sensing of Graph Signals. A graph is formed by a set of vertices $\mathcal{V}$ and a set of edges $\mathcal{E}$. Its characteristics are compactly captured by its adjacency matrix $\mathbf{E} \in \mathbb{R}^{n \times n}$, where n is the number of vertices. For all $i, j = 1, \ldots, n$, the value of $\mathbf{E}(i, j)$ is the weight of the path from the ith to the jth vertex (according to some arbitrary but fixed ordering of the vertices). A graph signal $\boldsymbol{x}_0$ is a function $\boldsymbol{x}_0 : \mathcal{V} \to \mathbb{R}$, which can be treated as a vector in $\mathbb{R}^n$ [26].

Compressed sensing of graph signals has been explored in [27]. In order to construct a real-world graph signal, we use a dataset of temperature values published by the Indian Meteorological Department [28]. The data we use consists of the maximum temperature



(a) Ground-truth, $\boldsymbol{x}_0$. (b) $\boldsymbol{x}^*$ (noiseless). (c) $\boldsymbol{x}^*$ ($\sigma_n = 0.5$). (d) $\boldsymbol{x}^*$ ($\sigma_n = 1.0$). (e) $\boldsymbol{x}^*$ ($\sigma_n = 1.5$). (f) Relative error.

Fig. 3. CS reconstruction of a temperature signal (values in $^\circ$C) from 140 random Gaussian projections ($\approx 40\%$). The signal resides on a graph having 355 vertices. The relative error $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2 \,/\, \|\boldsymbol{x}_0\|_2$ is as follows: (b) 0.021, (c) 0.026, (d) 0.062 and (e) 0.078.



(a) Ground-truth, $\boldsymbol{x}_0$. (b) $\boldsymbol{y}$ (noiseless). (c) $\boldsymbol{x}^*$ (noiseless). (d) $\boldsymbol{y}$ ($\sigma_n = 0.5$). (e) $\boldsymbol{x}^*$ ($\sigma_n = 0.5$). (f) Relative error.

Fig. 4. Interpolation of a temperature profile (values in $^\circ$C) from 175 missing measurements ($\approx 50\%$). The graph size is 355. The small black dots indicate missing values. The relative error is as follows: (c) 0.064 and (e) 0.185.

values (in $^\circ$C) on the 111th day of the year 2018, evaluated at discrete points on a grid of $1^\circ \times 1^\circ$ (latitude $\times$ longitude) spread over the Indian region. Each grid point is a vertex of our graph, and the number of vertices is $n = 355$. We construct $\mathbf{E}$ based on the method in [21]:

$$\mathbf{E}(i,j) = \begin{cases} \dfrac{1}{Z_i} \exp\left(-d_{ij}^2/\beta^2\right), & \text{if } d_{ij} \leqslant C, \\ 0, & \text{otherwise,} \end{cases}$$

where d_{ij} is the geodesic distance between the locations corresponding to vertices i and j, C is a cutoff distance, $\beta > 0$ is a constant, and Z_i is a normalization factor. We set $C = 320$ and $\beta = C/3$ (all values in kilometers). The forward model is given by (1), where $\boldsymbol{x}_0$ is the temperature signal and $m < n$. We set $m = 140$ ($\approx 60\%$ compression), and randomly generate $\mathbf{A}$ with i.i.d. Gaussian entries having mean 0 and variance $1/m$.

To obtain a symmetric denoiser satisfying Theorem 1, we first construct a linear graph signal denoiser proposed in [21], which is defined as

$$\overline{\mathbf{W}} = \Big(\mathbf{I} + \alpha(\mathbf{I} - \mathbf{E})^\top (\mathbf{I} - \mathbf{E})\Big)^{-1},$$

where $\alpha > 0$ controls the degree of smoothing. Note that $\overline{\mathbf{W}}$ is symmetric and positive semidefinite, and its eigenvalues lie in $[0, 1]$. We take the SVD-based low-rank approximation of $\overline{\mathbf{W}}$, with rank equal to 120, as our denoiser $\mathbf{W}$. We verify that the second condition in Theorem 3 is satisfied, and generate $\boldsymbol{y}$ as per (1).

The recovered signal $\boldsymbol{x}^*$ obtained by solving (4) is shown in Fig. 3 for different levels σ_n of the measurement noise. In Fig. 3(f), we show a plot of the error (averaged over 100 random trials). Note that since model noise is present, the error is non-zero even when $\boldsymbol{\eta} = \mathbf{0}$. We use the `cvxpy` package in Python to solve (4).

Graph Signal Interpolation. Graph signal interpolation is the counterpart of inpainting for graph signals [29]. As before, $\mathbf{A}$ is a random selection of m rows of the $n \times n$ identity matrix. We consider the same graph and denoiser as in the previous experiment, but the ground-truth signal $\boldsymbol{x}_0$ now consists of the minimum temperature values on the 6th day of the year 2018. We fix $m = 175$, corresponding to about 50% missing samples. We generate $\mathbf{A}$ so as to satisfy the first condition in Theorem 3, and generate $\boldsymbol{y}$ according to (1). The reconstruction $\boldsymbol{x}^*$ as well as a plot of the errors (averaged over 100 random trials) are shown in Fig. 4 for different values of σ_n.

From Figs. 3 and 4, we see that the error $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2$ increases smoothly with σ_n. Moreover, we expect the minimum error (corresponding to the case where $\boldsymbol{\eta} = \mathbf{0}$), which arises purely due to model noise, to depend on the distance of $\boldsymbol{x}_0$ from the subspace $\mathcal{R}(\mathbf{W})$. It turns out that these observations can be formalized as stated in the theorem below. We leave the proof to a future paper [19].

Theorem 5. *Suppose that the restriction of* $\mathbf{A}$ *to* $\mathcal{R}(\mathbf{W})$ *is one-to-one. Let* $\boldsymbol{x}^*$ *be a solution of problem* (4)*, where* $\delta \geqslant \|\boldsymbol{\eta}\|_2$.

- *If* $\boldsymbol{x}_0 \in \mathcal{R}(\mathbf{W})$*, then* $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2 \leqslant a\|\boldsymbol{\eta}\|_2 + c$ *for some* $a, c > 0$.
- *If* $\boldsymbol{x}_0 \notin \mathcal{R}(\mathbf{W})$*, then* $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2 \leqslant a\|\boldsymbol{\eta}\|_2 + b \cdot \mathrm{dist}\big(\boldsymbol{x}_0, \mathcal{R}(\mathbf{W})\big) + c$ *for some* $a, b, c > 0$*, where* $\mathrm{dist}\big(\boldsymbol{x}_0, \mathcal{R}(\mathbf{W})\big)$ *is the distance of* $\boldsymbol{x}_0$ *from* $\mathcal{R}(\mathbf{W})$.

4. CONCLUSION

We formally established that a signal can be exactly recovered from m noiseless measurements using plug-and-play regularization if the rank of the linear denoiser is at most m and if the signal is in the range of the denoiser. This should be of interest since many linear denoisers used in practice are (near) low-rank [18]. For compressed sensing using random Gaussian measurements, we established that exact recovery can be achieved with probability 1. Using one-dimensional signals, images and graph signals, we showed that the recovery is robust even if the condition regarding the range is violated and the measurements are noisy. Based on these observations, we put forth a theorem that the recovery error $\|\boldsymbol{x}^* - \boldsymbol{x}_0\|_2$ is bounded by the noise level provided certain conditions are met. To the best of our knowledge, this is the first work to address the question of exact/stable recovery for plug-and-play regularization.

5. REFERENCES

[1] B. R. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. Comput.*, vol. C-26, no. 3, pp. 219–229, 1977.

[2] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Elsevier, Burlington, MA, USA, 2009.

[3] A. Beck, *First-Order Methods in Optimization*, SIAM, Philadelphia, PA, USA, 2017.

[4] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," *Proc. IEEE Global Conf. Signal Inform. Process.*, pp. 945–948, 2013.

[5] S. Sreehari, S. V. Venkatakrishnan, B. Wohlberg, G. T. Buzzard, L. F. Drummy, J. P. Simmons, and C. A. Bouman, "Plug-and-play priors for bright field electron tomography and sparse interpolation," *IEEE Trans. Comput. Imaging*, vol. 2, no. 4, pp. 408–423, 2016.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[7] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[8] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. Chan, S. Liu, E. T. Reehorst, and P. Schniter, "Plug-and-play methods for magnetic resonance imaging: Using denoisers for image recovery," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 105–116, 2020.

[9] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021.

[10] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "A convergent image fusion algorithm using scene-adapted Gaussian-mixture-based denoising," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 451–463, 2019.

[11] S. H. Chan, "Performance analysis of plug-and-play ADMM: A graph signal processing perspective," *IEEE Trans. Comput. Imaging*, vol. 5, no. 2, pp. 274–286, 2019.

[12] P. Nair, R. G. Gavaskar, and K. N. Chaudhury, "Fixed-point and objective convergence of plug-and-play algorithms," *IEEE Trans. Comput. Imaging*, vol. 7, pp. 337–348, 2021.

[13] R. G. Gavaskar, C. D. Athalye, and K. N. Chaudhury, "On plug-and-play regularization using linear denoisers," *IEEE Trans. Image Process.*, vol. 30, pp. 4802–4813, 2021.

[14] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[15] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012.

[16] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Top. Signal Process.*, vol. 1, no. 4, pp. 586–597, 2007.

[17] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," *Proc. IEEE Int. Conf. Comput. Vision*, pp. 839–846, 1998.

[18] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, 2013.

[19] R. G. Gavaskar, C. D. Athalye, and K. N. Chaudhury, "Exact and robust compressed sensing using plug-and-play regularization," In preparation.

[20] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 755–768, 2013.

[21] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovacevic, "Signal denoising on graphs via graph filtering," *Proc. IEEE Global Conf. Signal Inform. Process.*, pp. 872–876, 2014.

[22] Y. Liu, M. de Vos, I. Gligorijevic, V. Matic, Y. Li, and S. van Huffel, "Multi-structural signal recovery for biomedical compressive sensing," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2794–2805, 2013.

[23] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, 2001.

[24] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.2," `http://cvxr.com/cvx`, January 2020, Accessed: June 2021.

[25] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, 2010.

[26] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, 2013.

[27] H. Zheng, F. Yang, X. Tian, X. Gan, X. Wang, and S. Xiao, "Data gathering with compressive sensing in wireless sensor networks: A random walk based approach," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 1, pp. 35–44, 2014.

[28] A. K. Srivastava, M. Rajeevan, and S. R. Kshirsagar, "Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region," *Atmos. Sci. Letters*, vol. 10, no. 4, pp. 249–254, 2009.

[29] Y. Yazaki, Y. Tanaka, and S. H. Chan, "Interpolation and denoising of graph signals using plug-and-play ADMM," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 5431–5435, 2019.