Sparsity-Promoting Multiplicative Denoising via Block Coordinate Descent

DARTMOUTH
Guarini School of Graduate and Advanced Studies

Jonathan Lindbloom, Anne Gelb

Dartmouth College, Department of Mathematics

Abstract

Images produced by coherent imaging techniques, such as synthetic aperture radar (SAR), ultrasound imaging, and positron emission tomography (PET), are frequently contaminated by multiplicative noise. Unfortunately, numerical methods for additive (Gaussian) denoising cannot easily be modified to handle multiplicative noise.

We propose two new block coordinate descent methods for sparsity-promoting multiplicative denoising assuming a Gamma multiplicative noise model. The first method is applicable when the unknown has first been obscured by a linear measurement operator $\boldsymbol{A} \in \mathbb{R}_+^{m \times n}$, whereas the second method is carried out in the log domain but requires $\boldsymbol{A} = \boldsymbol{I}_n$ (the identity). Both methods accommodate general sparsifying transformations $\boldsymbol{R} \in \mathbb{R}^{k \times n}$. Convexity analysis and convergence guarantees are presented.

To accelerate the methods, we employ a Newton-Krylov method for the solution of their most computationally-intensive subproblems. Finally, we develop an efficient ***priorconditioned Newton-Krylov method*** for a very general class of nonlinear optimization problems, which can be viewed as a generalization of the popular priorconditioning technique of the statistical inverse problem community to non-Gaussian likelihoods. We then apply this method to further accelerate the proposed multiplicative denoising methods.

General model

For an unknown $\boldsymbol{u} \in \mathbb{R}^n_{++}$, we assume that the corrupted observation $\boldsymbol{f} \in \mathbb{R}^n_{++}$ arises from the observational model

$$\boldsymbol{f} = (\boldsymbol{A}\boldsymbol{u}) \odot \boldsymbol{\eta}.$$

Here $\boldsymbol{A} \in \mathbb{R}_+^{m \times n}$ is a nonnegative linear measurement operator, and $\boldsymbol{\eta} \in \mathbb{R}^m_{++}$ having entries independently and identically distributed according to a gamma distribution $[\boldsymbol{\eta}]_i \sim \mathrm{Gamma}(L, L)$ for some parameter $L > 0$.

For the reconstruction of $\boldsymbol{u}$, we propose to model $\boldsymbol{u}$ via the hierarchical generative model

$$[\boldsymbol{f}]_i \,|\, \boldsymbol{u} \overset{\text{ind}}{\sim} \mathrm{Gamma}\left(L, \frac{L}{[\boldsymbol{A}\boldsymbol{u}]_i}\right), \quad i = 1, \dots, m,$$
$$\boldsymbol{u} \,|\, \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{0}, \left(\boldsymbol{R}^T \boldsymbol{D}_{\boldsymbol{\theta}}^{-1} \boldsymbol{R}\right)^{-1}\right),$$
$$[\boldsymbol{\theta}]_i \overset{\text{iid}}{\sim} \mathcal{GG}(r, \beta, \vartheta), \quad i = 1, \dots, k.$$

where we employ a sparsity-promoting, conditionally-Gaussian prior for the signal vector $\boldsymbol{u}$. The degree of sparsity is controlled by the three hyperparameters r, β, ϑ.

Applying Bayes' theorem, the negative log posterior density is of the form

$$\mathcal{E}(\boldsymbol{u}, \boldsymbol{\theta}) = \sum_{i=1}^m L \log\left([\boldsymbol{A}\boldsymbol{u}]_i\right) + L \frac{[\boldsymbol{f}]_i}{[\boldsymbol{A}\boldsymbol{u}]_i} + \frac{1}{2} \|\boldsymbol{D}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{R}\boldsymbol{u}\|_2^2 - \log \pi(\boldsymbol{\theta}),$$

with $\pi(\boldsymbol{\theta})$ collecting hyperprior terms. We then seek the MAP estimate of the parameters $(\boldsymbol{u}^{\text{MAP}}, \boldsymbol{\theta}^{\text{MAP}}) = \arg\min_{(\boldsymbol{u}, \boldsymbol{\theta})} \mathcal{E}(\boldsymbol{u}, \boldsymbol{\theta})$.

Log model

If $\boldsymbol{A} = \boldsymbol{I}_n$, then one can formulate a similar model in the log-domain. Defining $\boldsymbol{y} \coloneqq \log(\boldsymbol{f})$, $\boldsymbol{x} \coloneqq \log(\boldsymbol{u})$, $\boldsymbol{\epsilon} \coloneqq \log \boldsymbol{\eta}$, we obtain the additive model $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\epsilon}$. Formulating a similar hierarchical model for $\boldsymbol{x}$, the negative log posterior density we consider has the form

$$\mathcal{H}(\boldsymbol{x}, \boldsymbol{\theta}) = \sum_{i=1}^n L([\boldsymbol{x}]_i - [\boldsymbol{y}]_i) + L \exp\{[\boldsymbol{y}]_i - [\boldsymbol{x}]_i\}$$
$$+ \frac{1}{2} \|\boldsymbol{D}_{\boldsymbol{\theta}}^{-1/2} \boldsymbol{R}\boldsymbol{x}\|_2^2 - \log \pi(\boldsymbol{\theta}).$$

A benefit of the log model is that the likelihood is log-concave, whereas the likelihood in the general model is not. However, reconstructions via the log model can be biased towards lower intensity.

Block coordinate descent methods

To solve the MAP problem, we employ the block coordinate descent method

$$\boldsymbol{\theta}^{(k+1)} = \arg\min_{\boldsymbol{\theta}} \ \mathcal{E}(\boldsymbol{u}^{(k)}, \boldsymbol{\theta}),$$
$$\boldsymbol{u}^{(k+1)} = \arg\min_{\boldsymbol{u}} \ \mathcal{E}(\boldsymbol{u}, \boldsymbol{\theta}^{(k+1)})$$

for the general model, and similarly for the log model.

The subproblem for updating $\boldsymbol{\theta}$ can be solved via an ODE method or by an analytic formula depending on the hyperparameters r and β, and in either case is relatively cheap to compute.

The subproblem for updating $\boldsymbol{u}$ amounts to minimizing the negative log likelihood regularized by a quadratic penalty. With $\mathcal{F}(\boldsymbol{u})$ the objective, we consider solving this subproblem using a either first-order gradient, or by a second-order Newton-Krylov method of the form

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} - \alpha_k \left(D_{\boldsymbol{u},\boldsymbol{u}} \mathcal{F}(\boldsymbol{u}^{(k)})\right)^{-1} D_{\boldsymbol{u}} \mathcal{F}(\boldsymbol{u}^{(k)}).$$

The step sizes $\{\alpha_k\}$ chosen by an Armijo line search. For the general model the hessian may be nonsingular, but can be made singular with a simple modification. For the log model, the hessian is always nonsingular.

Analysis

Lemma. *The objective functions $\mathcal{E}(\boldsymbol{u}, \boldsymbol{\theta})$ and $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\theta})$ are proper, lower semicontinuous, closed, have closed sublevel sets, and are coercive. They also achieve minimum values.*

Theorem. *The objective function $\mathcal{E}(\boldsymbol{u}, \boldsymbol{\theta})$ is generally nonconvex.* The objective function $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\theta})$ is strictly convex if $r \geq 1$ and $\eta = r\beta - \frac{3}{2} > 0$, but is generally nonconvex otherwise.

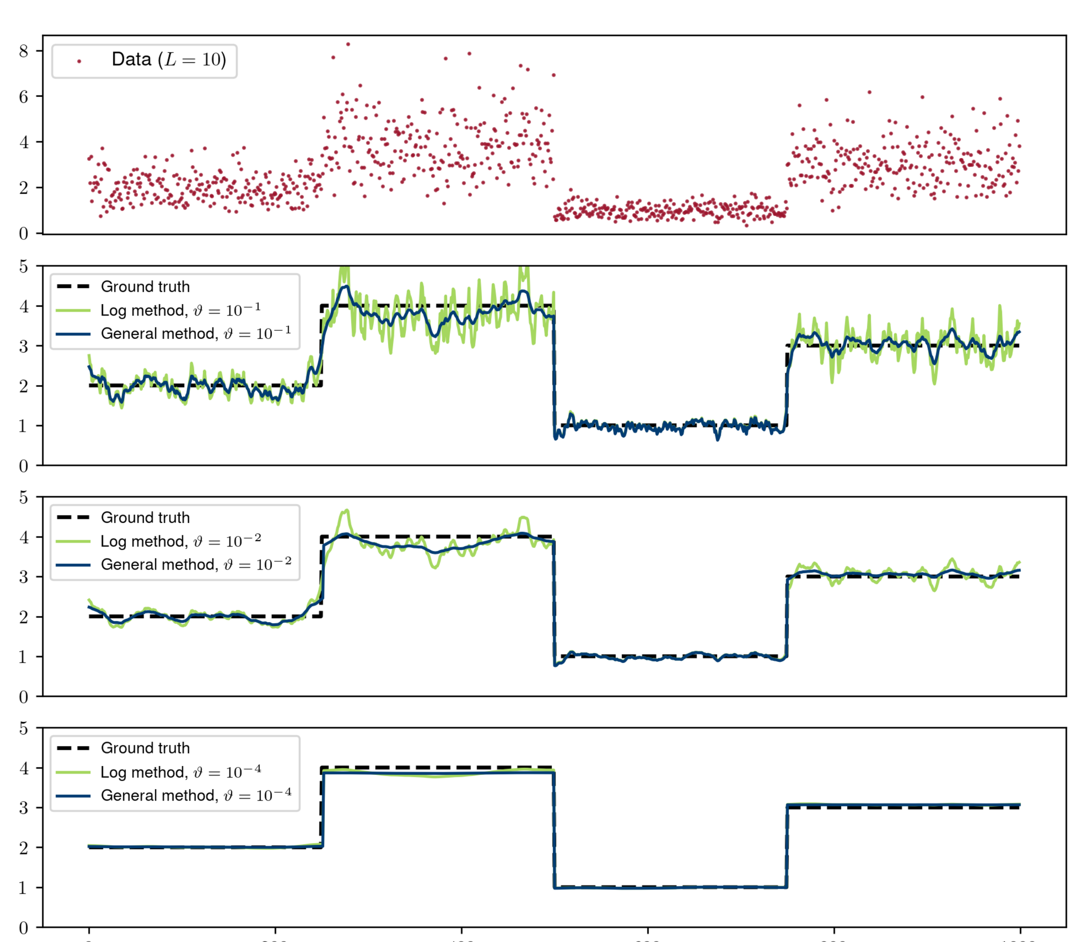
Convergence

Theorem. *Let $\{\boldsymbol{x}^{(k)}, \boldsymbol{\theta}^{(k)}\}$ denote the iterates of the block coordinate descent applied the log model objective $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\theta})$. Then any limit point of $\{\boldsymbol{x}^{(k)}, \boldsymbol{\theta}^{(k)}\}$ is a stationary point of $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\theta})$. If $r \geq 1$ and $\eta = r\beta - 3/2 > 0$, then any limit point of $\{\boldsymbol{x}^{(k)}, \boldsymbol{\theta}^{(k)}\}$ is a unique minimizer of $\mathcal{H}(\boldsymbol{x}, \boldsymbol{\theta})$.*

Conjecture. *Let $\{\boldsymbol{u}^{(k)}, \boldsymbol{\theta}^{(k)}\}$ denote the iterates of the block coordinate descent applied the general model objective $\mathcal{E}(\boldsymbol{u}, \boldsymbol{\theta})$. Then any limit point of $\{\boldsymbol{u}^{(k)}, \boldsymbol{\theta}^{(k)}\}$ is a stationary point of $\mathcal{E}(\boldsymbol{u}, \boldsymbol{\theta})$.*

One-dimensional denoising

Here we consider the denoising task ($\boldsymbol{A} = \boldsymbol{I}_n$) with sparsity in the discrete gradient. We fix $r = -1, \beta = 1$, and show results for varying ϑ.



Priorconditioned Newton-Krylov method

Utilizing a Newton-Krylov method for the $\boldsymbol{u}/\boldsymbol{x}$ subproblem of the coordinate descent requires the repeated solution of (possibly large) linear systems with a Krylov method such as the Conjugate Gradient (CG) method. Hence it is advantageous to precondition these system solves.

A variety of preconditioning strategies may be employed. One popular approach is known as *priorconditioning*, which for models with additive Gaussian noise turns the MAP problem of solving

$$\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{R} \end{bmatrix} \boldsymbol{x}^\star = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0}_k \end{bmatrix} \text{ in least-squares} \iff \boldsymbol{x}^\star = \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \frac{1}{2}\|\boldsymbol{R}\boldsymbol{x}\|_2^2$$

into the related problem of solving

$$\begin{bmatrix} \boldsymbol{A}\boldsymbol{R}^\# \\ \boldsymbol{I}_k \end{bmatrix} \boldsymbol{z}^\star = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0}_k \end{bmatrix} \text{ in least-squares} \iff \boldsymbol{z}^\star = \arg\min_{\boldsymbol{z} \in \mathbb{R}^k} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{R}^\#\boldsymbol{z} - \boldsymbol{y}\|_2^2 + \frac{1}{2}\|\boldsymbol{z}\|_2^2$$

where $\boldsymbol{R}^\#$ is an oblique ($\boldsymbol{A}$-weighted) pseudoinverse. This has an interpretation as a standard form transformation for regularized least-squares problems.

We propose a novel extension of the priorconditioning technique to the non-Gaussian setting. Specifically, we consider the general class of problems

$$\arg\min_{\boldsymbol{x} \in \mathbb{R}^n} \sum_{i=1}^m \phi_i([\boldsymbol{A}\boldsymbol{x}]_i\,;\,[\boldsymbol{y}]_i) + \frac{1}{2}\|\boldsymbol{R}\boldsymbol{x}\|_2^2.$$

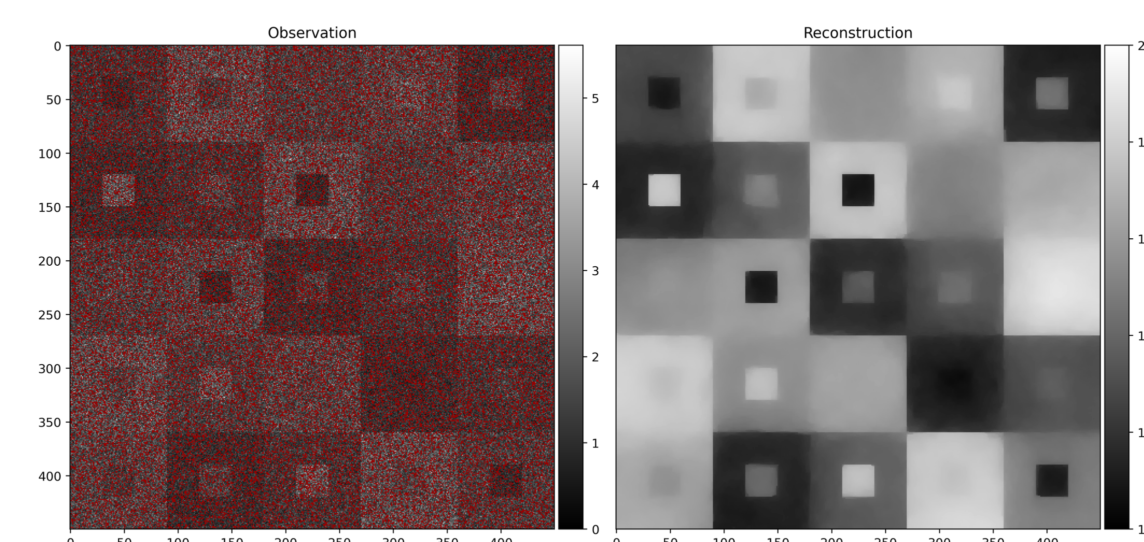
of which the $\boldsymbol{u}/\boldsymbol{x}$ subproblems in our multiplicative denoising methods are special cases. Letting $\mathcal{W}(\boldsymbol{x})$ be this objective, we show how for this class of problems the computation of the Newton direction

$$\boldsymbol{d}^{(k)} = -(D_{\boldsymbol{x},\boldsymbol{x}} \mathcal{W}(\boldsymbol{x}^{(k)}))^{-1} (D_{\boldsymbol{x}} \mathcal{W}(\boldsymbol{x}^{(k)}))$$

can be cast as the solution to a least-squares problem, to which the priorconditoning technique can be employed as in the Gaussian setting.

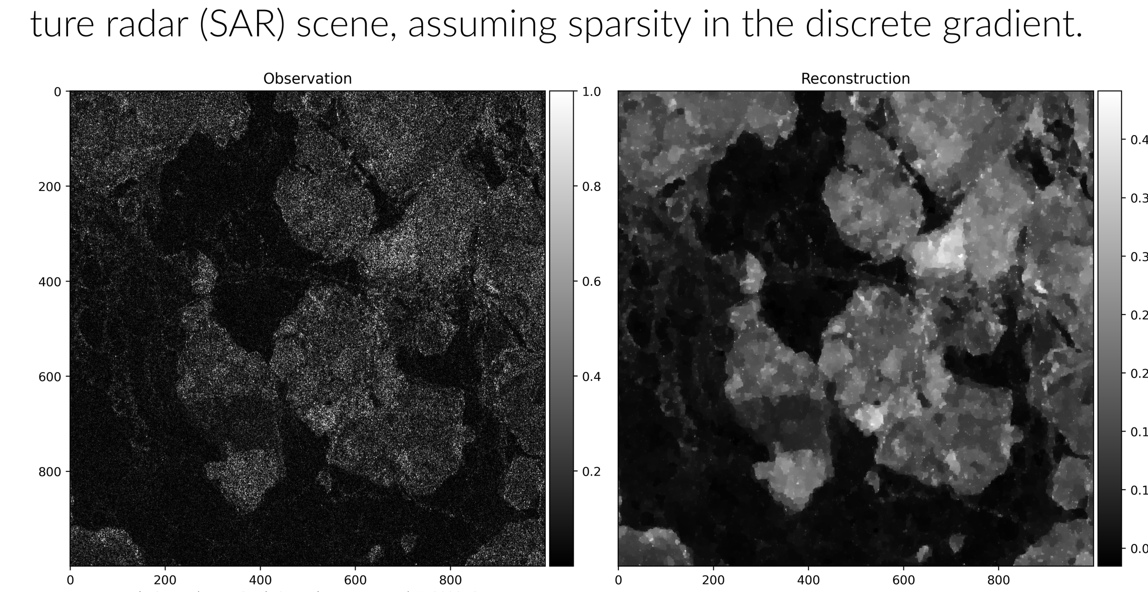
Joint denoising and inpainting

Here we use the general model algorithm for a joint denoising ($L = 10$) and inpainting task, assuming sparsity in the discrete gradient. Red pixels are unobserved, and we pick $r = 1$, $\eta = 10^{-2}$, $\vartheta = \frac{1}{2}$.



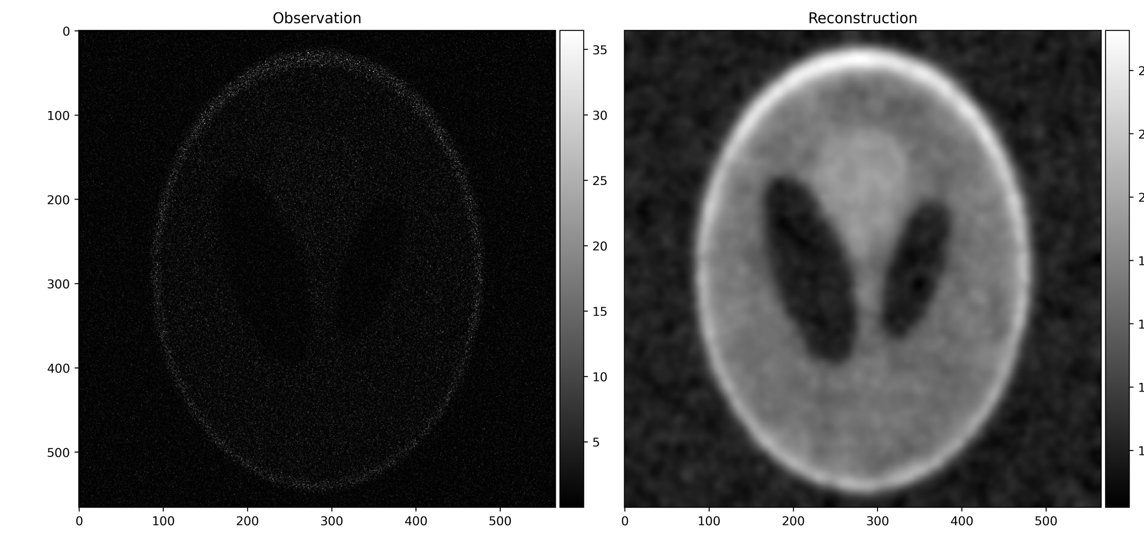
SAR despeckling

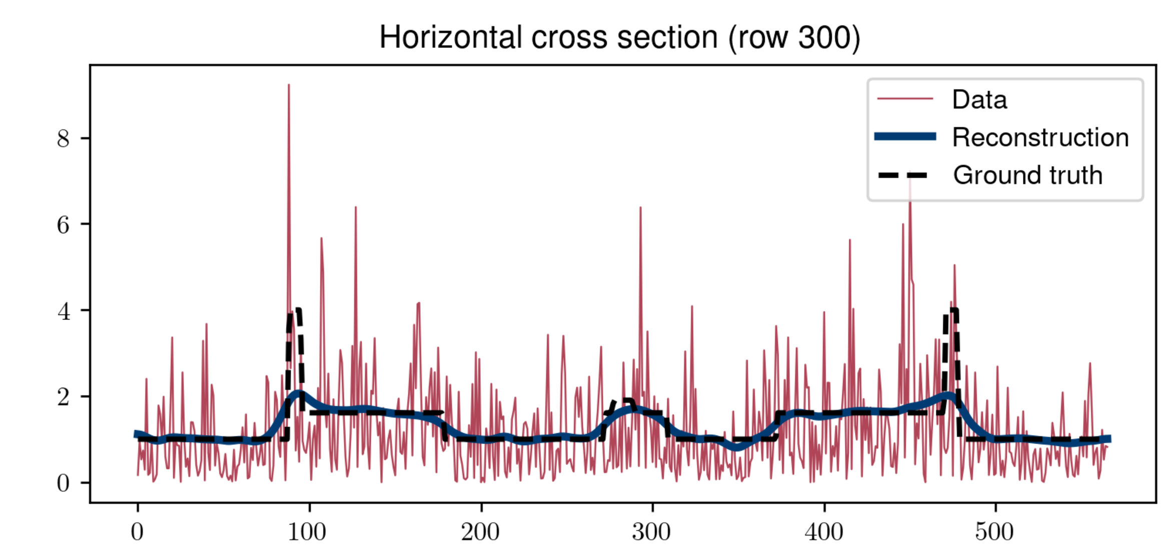
Here we use the log model algorithm for despeckling a synthetic aperture radar (SAR) scene, assuming sparsity in the discrete gradient.



Joint denoising and deblurring

Here we use the general model algorithm for a joint denoising and deblurring task, assuming sparsity in the discrete gradient. This is the most challenging example shown, since here $L = 1$. We pick $r = -1, \beta = 1, \vartheta = -1$.





Future directions

- **Automated regularization parameter selection:** in our methods the parameter ϑ plays the role of a regularization strength parameter which must be specified or hand-tuned by the user.
- **Nonconvexity:** since we work with generalized Gamma hyperpriors, one might consider hybrid or path-following approaches to obtain better results with strongly promoting priors.
- **Uncertainty quantification:** here we seek only the MAP estimate of the unknown. To obtain UQ information one might consider sampling the posterior distribution via MCMC, or other approaches to UQ.
- **Other sparsifying transformations:** our examples assume sparsity in the discrete gradient. However, our approach allows for a general sparsifying transformation, so one could easily consider sparsity in our bases, e.g., wavelet coefficients.

References

Aubert, G., Aujol, J.F. (2008). A Variational Approach to Removing Multiplicative Noise. *SIAM Journal on Applied Mathematics, 68(4), 925-946.*

Beck, A. (2017). First-Order Methods in Optimization. SIAM.

Calvetti, D., Pragliola, M., Somersalo, E., Strang, A. (2020). Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors. *Inverse Problems, 36(2), 025010.*

Hansen, P. (2013). Oblique projections and standard-form transformations for discrete inverse problems. *Numerical Linear Algebra with Applications, 20(2), 250-258.*

Acknowledgements

This work is supported by AFOSR grant #F9550-22-1-0411, DOE ASCR #DE-ACO5-000R22725, and ONR MURI grant #N00014-20-1-2595.

jonathan.t.lindbloom.gr@dartmouth.edu