Generalized hybrid solvers for large-scale edge-preserving inversion

Jonathan Lindbloom* **Jan Glaubitz** **Anne Gelb**

Nov. 1, 2022

MURI Annual Meeting

This work is supported by ONR MURI #N00014-20-1-2595.

Narrative of this talk

- Reconstructing environmental variables of sea ice state from observations is typically ill-posed due to lack of data and noise
- We need to introduce regularization in the form of a *prior* to obtain good reconstructions
- Not all priors are created equal; some priors recover certain features (such as floe edges) better than others
- To use better priors, the price we pay is that we lose convexity and algorithmic convergence guarantees
- **The goal of this work is to obtain good reconstructions with sparsity priors in spite of this loss of convexity, for large-scale problems**

Outline

1. **Bayesian inverse problems**
2. **Sparsity-promoting priors**
3. **Numerical examples**

Bayesian inverse problems

Inverse problems

- The typical case: we have some data $\boldsymbol{y} \in \mathbb{R}^m$ that we assume comes from the generative model

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}), \tag{1}$$

  for some known linear measurement operator $\boldsymbol{F} \in \mathbb{R}^{m \times n}$, unknown ground truth $\boldsymbol{x} \in \mathbb{R}^n$, and noise $\boldsymbol{\epsilon}$.

- Even if $\boldsymbol{F}^{-1}$ exists, inverting the observation with $\boldsymbol{F}^{-1}$ yields poor reconstructions due to ill-posedness of the reconstruction problem

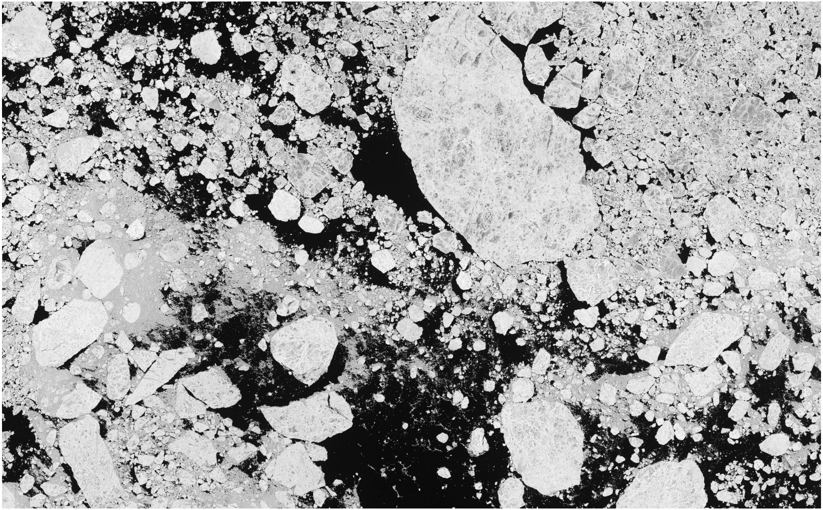

Figure 1: Optical imagery (ground truth).

Figure 2: $\boldsymbol{F}$ representing a blurring/averaging.

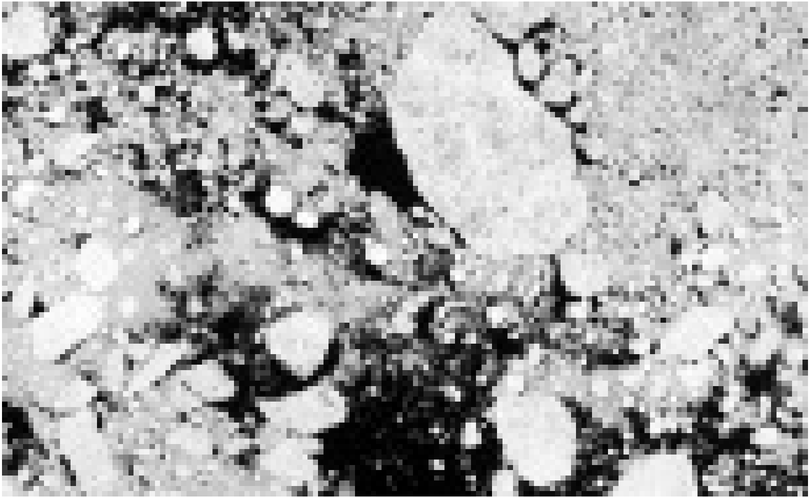

Figure 3: $\boldsymbol{F}$ representing an up-sampling, used in super-resolution, combining observations of disparate resolutions, pan-sharpening.

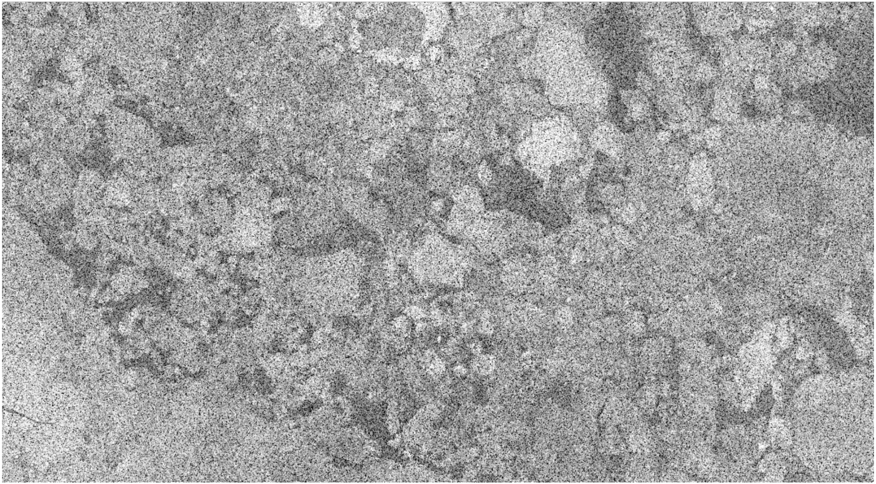

Figure 4: $\boldsymbol{F}$ is just the identity, as in the product model for synthetic aperture radar (SAR) de-despeckling.

ICEYE Strip SAR Example, Arctic Ocean (~ 12 km x 27 km)

Bayesian formulation of the reconstruction task

- Goal: characterize the posterior probability density

$$\underbrace{\pi(\boldsymbol{x} \mid \boldsymbol{y})}_{\text{(posterior)}} \propto \overbrace{\pi(\boldsymbol{y} \mid \boldsymbol{x})}^{\text{likelihood (data fidelity)}} \times \underbrace{\pi(\boldsymbol{x})}_{\text{prior (regularization)}} \tag{2}$$

- In our work, we focus on finding the *maximum a posteriori* (MAP) point estimate

$$\boldsymbol{x}_{\text{MAP}} = \arg\min_{\boldsymbol{x}} \{-\log \pi(\boldsymbol{x} \mid \boldsymbol{y})\}\,. \tag{3}$$

- Future work will consider characterizing the full posterior density, which then permits uncertainty quantification for the reconstruction.

Common (convex) priors

1. ℓ_p-norm priors, $1 \leq p \leq 2$

$$\pi(\boldsymbol{x}) \propto \exp\left\{-\lambda \|\boldsymbol{R}\boldsymbol{x}\|_p^p\right\} \tag{4}$$

$$\Rightarrow \boldsymbol{x}_\lambda = \arg\min_{\boldsymbol{x}} \left\{\|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{R}\boldsymbol{x}\|_p^p\right\} \tag{5}$$

2. Total variation (TV) prior

$$\pi(\boldsymbol{x}) \propto \exp\left\{-\lambda \|\boldsymbol{R}\boldsymbol{x}\|_1\right\} \tag{6}$$

$$\Rightarrow \boldsymbol{x}_\lambda = \arg\min_{\boldsymbol{x}} \left\{\|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \|\nabla \boldsymbol{x}\|_1\right\} \tag{7}$$

The selection of the "best" *regularization parameter* λ is a field in and of itself.

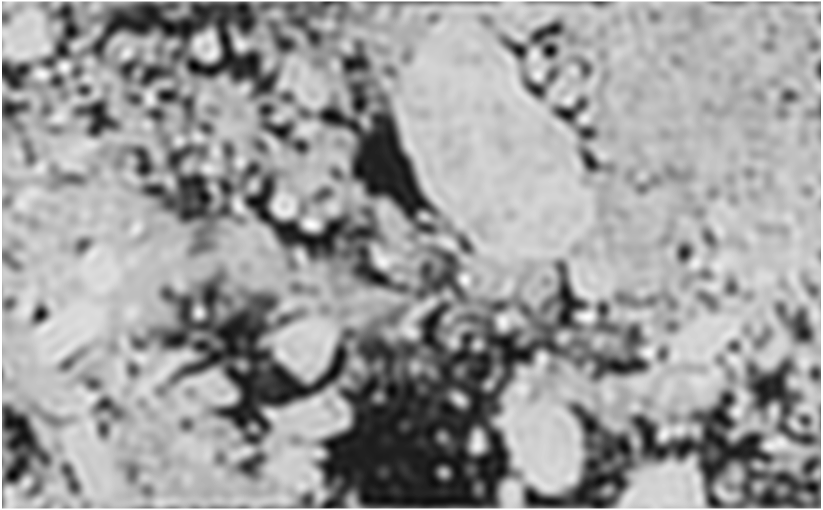

Figure 5: Tikhonov reconstruction for de-blurring problem with $\boldsymbol{R} = \nabla$ (image gradient), $\pi(\boldsymbol{x}) \propto \exp\left\{-\lambda\|\nabla \boldsymbol{x}\|_2^2\right\}$.

Sparsity-promoting priors

Motivation

- Hypothetically the prior is a free parameter (choose whatever regularization penalty you want)

- However, convexity of the MAP estimate depends on the prior

- Non-convex priors can yield superior reconstructions in sparse signal recovery, edge-preserving inversion, etc. 👍

- With non-convex problems you may lose convergence guarantees, convex algorithms may no longer produce good solutions, generally harder to approach 👎

The idea

- Represent non-convex priors marginally as scale-mixtures of Gaussians

$$\pi(\boldsymbol{x}) = \int \pi(\boldsymbol{x} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \, d\boldsymbol{\theta}$$

  where the joint density $\pi(\boldsymbol{x}, \boldsymbol{\theta})$ is *conditionally Gaussian* when conditioned on $\boldsymbol{\theta}$. Here $\pi(\boldsymbol{\theta})$ is a *hyper-prior* for the hyper-parameters $\boldsymbol{\theta}$, related to the prior we want to work with.

- We now infer a posterior over both the unknown source, as well as the parameters $\boldsymbol{\theta}$. We can also infer unknown noise levels in this framework, which are vital to the quality of the reconstruction.

$$\pi(\boldsymbol{x}, \boldsymbol{\theta} \,|\, \boldsymbol{y}) \propto \pi(\boldsymbol{y} \,|\, , \boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x} \,|\, \boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

The new problem

$$\underset{(\boldsymbol{x},\alpha,\boldsymbol{\beta})}{\arg\min}\ \underbrace{\frac{1}{2\alpha^2}\|\boldsymbol{F}\boldsymbol{x}-\boldsymbol{y}\|_2^2}_{\text{data fidelity}}+\overbrace{\|\boldsymbol{R}\boldsymbol{x}\|^2_{\boldsymbol{D}_{\boldsymbol{\beta}}^{-1}}}^{\text{prior}}+\underbrace{\frac{m}{2}\log\alpha+\log\det\big(\boldsymbol{D}_{\boldsymbol{\beta}}\big)-\log\pi(\alpha,\boldsymbol{\beta})}_{\text{hyper-parameters}}$$

- Bayesian coordinate descent (BCD) algorithm for obtaining the posterior mean[1]
- Iterative alternating sequential (IAS) algorithm for obtaining the MAP estimate, **with the global hybrid variant to tackle non-convexity**[2]

---

[1] Glaubitz, Gelb, and Song, *Generalized sparse Bayesian learning and application to image reconstruction.*

[2] Calvetti, Pragliola, and Somersalo, "Sparsity Promoting Hybrid Solvers for Hierarchical Bayesian Inverse Problems".

Our novel contributions

1. We have proven convexity conditions for the hierarchical model that apply in more general settings than has been shown before; (1) for non-invertible regularization $\boldsymbol{R}$, (2) unknown noise variance, and (3) multiple observations with unknown noise variance (data fusion)

2. We have applied auxiliary variable techniques to develop a new variant of IAS, *partially-collapsed Gibbs* IAS (PCG-IAS), which can exploit diagonalizations of operators and is computationally feasible for large-scale inversion

Numerical examples

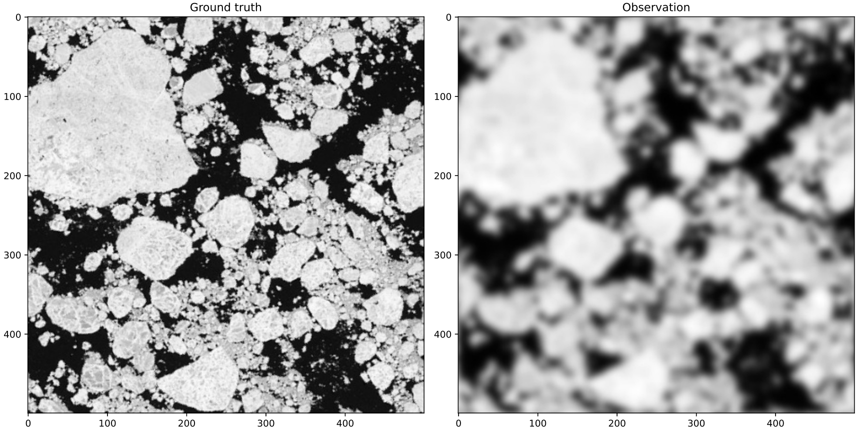

Figure 6: Example data.

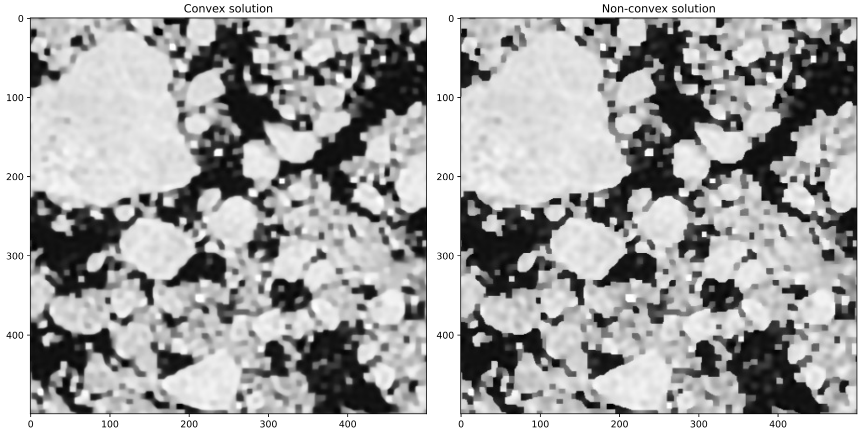

Figure 7: Two reconstructions via PCG-IAS; convex solution (left) used to initialize sparse gradient, non-convex solution (right)

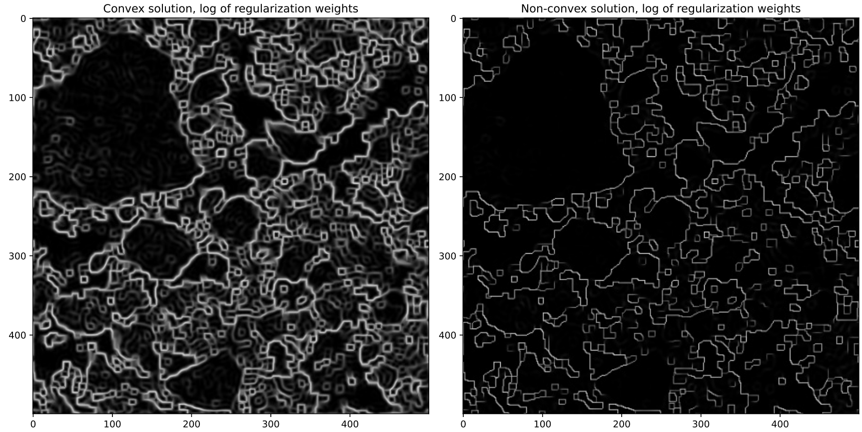

Figure 8: Inferred (gradient) regularization weights for the convex solution and non-convex solution.

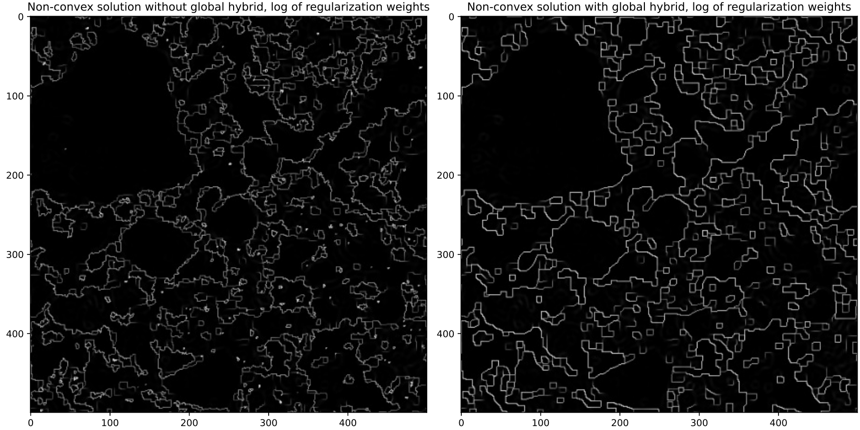

Figure 9: Inferred (gradient) regularization weights for the non-convex problem, with and without global hybrid initialization. Both solutions are local minima of the same problem.

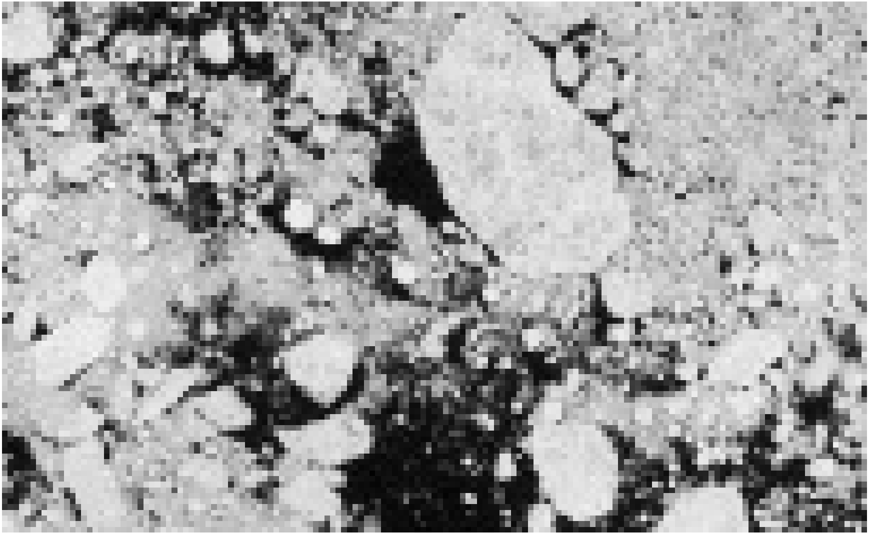

Figure 10: Low-resolution data (96 x 126).

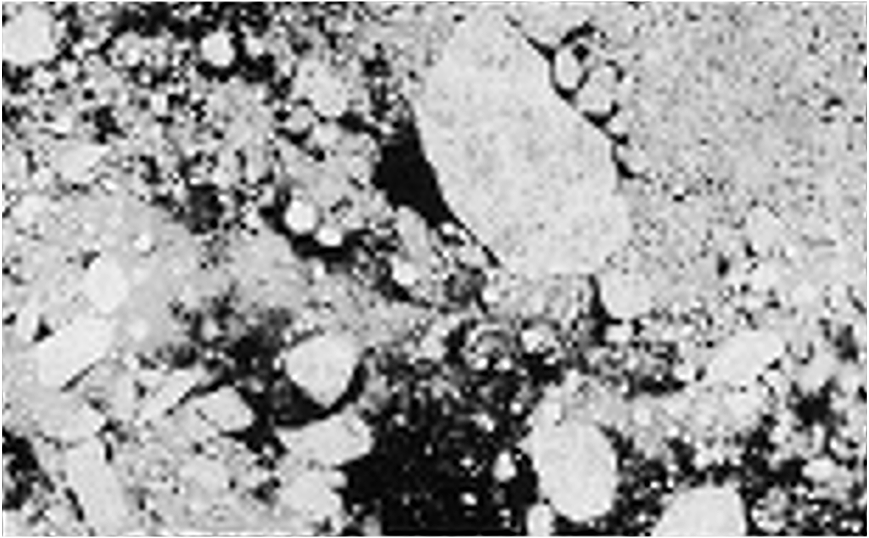

Figure 11: Super-resolution reconstruction (2000 x 3240).

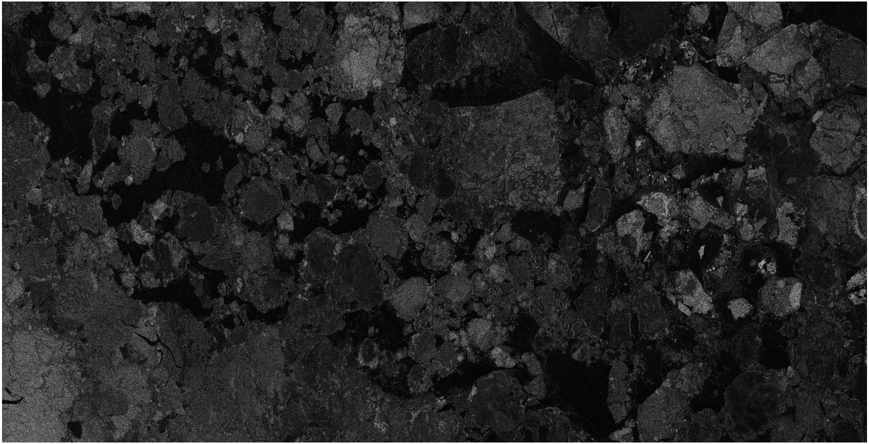

Figure 12: Real synthetic aperture radar image (contaminated by speckle noise).

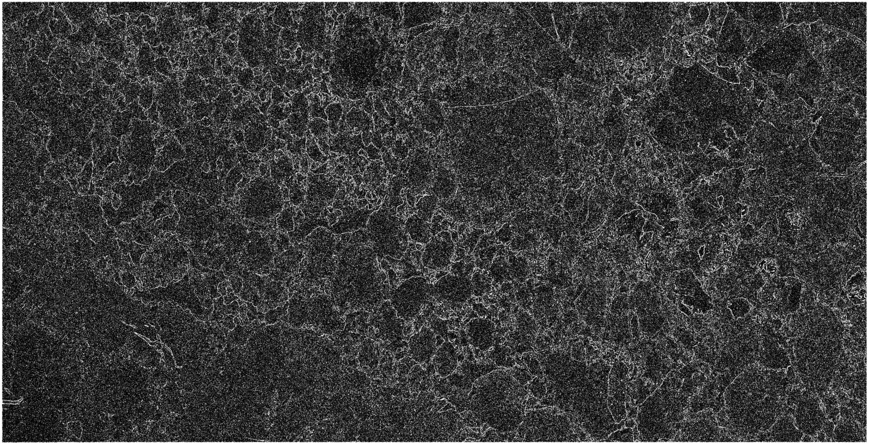

Figure 13: Variance image in reconstruction, according to reconstruction with a total variation prior.

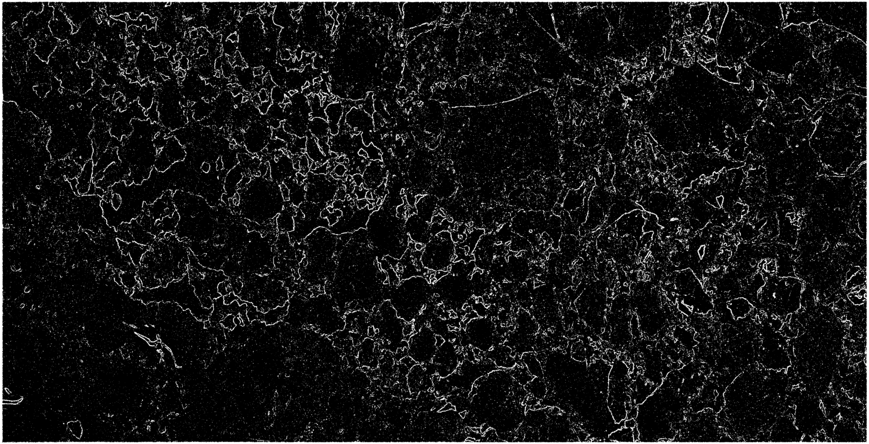

Figure 14: Variance image in reconstruction, according to reconstruction with a Cauchy difference prior.

Future directions

- Testing with MURI challenge problem 1
- Development of toolkit for easy access to solvers (to be released)
- Applying our solvers to de-speckling
- Uncertainty quantification
- Data fusion
- Application to pan-sharpening