DARTMOUTH
Department of
Mathematics

DARTMOUTH
Guarini School of Graduate
and Advanced Studies

Computational Strategies for Bayesian Inversion with Conditionally Gaussian Sparsity Priors

SIAM CSE23

3rd March 2023

Jonathan Lindbloom†, Jan Glaubitz, and Anne Gelb

Outline

1. Introduction
2. Hierarchical model
3. Data fusion
4. Despeckling
5. Data augmentation

Problem setup

We would like to reconstruct an unknown signal $\boldsymbol{x} \in \mathbb{R}^n$ given an indirect noisy observation

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{noise}}).$$

Here $\boldsymbol{y}, \boldsymbol{\epsilon} \in \mathbb{R}^m$ with linear forward operator $\boldsymbol{F} \in \mathbb{R}^{m \times n}$ and SPD noise covariance $\boldsymbol{\Sigma}_{\text{noise}} \in \mathbb{R}^{m \times m}$.

Problem setup

We would like to reconstruct an unknown signal $\boldsymbol{x} \in \mathbb{R}^n$ given an indirect noisy observation

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\text{noise}}).$$

Here $\boldsymbol{y}, \boldsymbol{\epsilon} \in \mathbb{R}^m$ with linear forward operator $\boldsymbol{F} \in \mathbb{R}^{m \times n}$ and SPD noise covariance $\boldsymbol{\Sigma}_{\text{noise}} \in \mathbb{R}^{m \times m}$.

Assumptions:

1. $\boldsymbol{x}$ is (relatively) sparse under some linear transformation $\boldsymbol{R} \in \mathbb{R}^{k \times n}$
2. We permit $\ker(\boldsymbol{R}) \neq \{\boldsymbol{0}\}$, but require $\ker(\boldsymbol{F}) \cap \ker(\boldsymbol{R}) = \{\boldsymbol{0}\}$
3. We do not know $\boldsymbol{\Sigma}_{\text{noise}}$ a priori but believe it has a simple parametric form

Conditionally Gaussian sparsity priors

We could opt to work with a sparsity prior directly.

Alternatively, many sparsity priors admit a representation as a conditionally Gaussian hierarchical prior, in which case of the form

$$\begin{aligned} \beta_j &\overset{iid}{\sim} \text{Hyper-prior} \\ \boldsymbol{x} \,|\, \boldsymbol{\beta} &\sim \mathcal{N}\left(\mathbf{0}, \left(\boldsymbol{R}^T \boldsymbol{D}_{\boldsymbol{\beta}}^{-1} \boldsymbol{R}\right)^{-1}\right) \\ \boldsymbol{y} \,|\, \boldsymbol{x} &\sim \mathcal{N}\left(\boldsymbol{F}\boldsymbol{x}, \boldsymbol{\Sigma}_{\text{noise}}\right) \end{aligned}$$

Conditionally Gaussian sparsity priors

ℓ_1-regularization:

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \ \|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|^2_{\boldsymbol{\Sigma}_{\text{noise}}} + \lambda\|\boldsymbol{R}\boldsymbol{x}\|_1$$

via conditionally Gaussian representation [1] :

$$(\boldsymbol{x}^\star, \boldsymbol{\beta}^\star) \overset{\eta\to 0}{=} \arg\min_{\boldsymbol{x},\boldsymbol{\beta}} \ \frac{1}{2}\|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|^2_{\boldsymbol{\Sigma}_{\text{noise}}} + \frac{1}{2}\|\boldsymbol{R}\boldsymbol{x}\|^2_{\boldsymbol{D}_{\boldsymbol{\beta}}^{-1}} + \frac{1}{2}\sum_{j=1}^{k} \log\beta_j - \log\pi(\boldsymbol{\beta})$$

[1] Calvetti, Somersalo and Strang, "Hierachical Bayesian models and sparsity: ℓ-2 -magic", 2019.

Negative log posterior

$$\pi_{\mathrm{GG}}(\theta \mid r, s, \vartheta) \propto \theta^{rs-1} \exp\left\{ -\left(\frac{\theta}{\vartheta}\right)^{r} \right\} \mathbf{1}_{>0}(\theta)$$

$$E(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|^2_{\boldsymbol{\Sigma}^{-1}_{\mathrm{noise}}} + \frac{1}{2}\|\boldsymbol{R}\boldsymbol{x}\|^2_{\boldsymbol{D}^{-1}_{\boldsymbol{\beta}}} - \eta \sum_{j=1}^{n} \log \beta_j + \sum_{j=1}^{n} \left(\frac{\beta_j}{\vartheta}\right)^{r}$$

Theorem [2]

Let $\eta := rs - \frac{3}{2}$ and $\ker(\boldsymbol{R}) = \{\mathbf{0}\}$. Then $E(\boldsymbol{x}, \boldsymbol{\beta})$ is globally convex if $r \geq 1$ and $\eta > 0$, and is locally convex at $(\boldsymbol{x}, \boldsymbol{\beta})$ if either (i) $0 < r < 1$ and $\eta > 0$ or (ii) $r < 0$ and $s > 0$, and

$$\beta_j < \vartheta \left(\eta/(r|r-1|)\right)^{1/r}, \quad j = 1, \ldots, k.$$

[2] Calvetti et al., "Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors", 2020.

MAP estimation via block coordinate descent (IAS)

IAS Algorithm

Choose initial $\boldsymbol{x}_0, \boldsymbol{\beta}_0$.
for $k = 1, \ldots, n_{\text{maxits}}$ do
    $\boldsymbol{\beta}_k \leftarrow \arg\max_{\boldsymbol{\beta}} \pi(\boldsymbol{x}_{k-1}, \boldsymbol{\beta} \mid \boldsymbol{y})$
    $\boldsymbol{x}_k \leftarrow \arg\max_{\boldsymbol{x}} \pi(\boldsymbol{x}, \boldsymbol{\beta}_k \mid \boldsymbol{y})$
end for

$$\arg\max_{\boldsymbol{x}} \pi(\boldsymbol{x}, \boldsymbol{\beta} \mid \boldsymbol{y}) = \left(\boldsymbol{F}^T \boldsymbol{\Sigma}_{\text{noise}}^{-1} \boldsymbol{F} + \boldsymbol{R}^T \boldsymbol{D}_{\boldsymbol{\beta}}^{-1} \boldsymbol{R} \right)^{-1} \boldsymbol{F}^T \boldsymbol{\Sigma}_{\text{noise}}^{-1} \boldsymbol{y}$$

Hierarchical model

Consider $\boldsymbol{\Sigma}_{\text{noise}} = \alpha \boldsymbol{I}_m$. Our first modification is to consider the model

$$
\begin{aligned}
\alpha &\sim \text{GG}(r_1, s_1, \vartheta_1), \\
\beta_j &\overset{iid}{\sim} \text{GG}(r_2, s_2, \vartheta_2), \quad j = 1, \ldots, k, \\
\boldsymbol{x} \,|\, \boldsymbol{\beta} &\sim \mathcal{N}\left(\boldsymbol{0}, \left(\boldsymbol{R}^T \boldsymbol{D}_{\boldsymbol{\beta}}^{-1} \boldsymbol{R}\right)^{-1}\right), \\
\boldsymbol{y} \,|\, \boldsymbol{x}, \alpha &\sim \mathcal{N}\left(\boldsymbol{F}\boldsymbol{x}, \alpha \boldsymbol{I}_m\right).
\end{aligned}
$$

Negative log posterior

The negative log posterior is given by $E(\boldsymbol{x}, \alpha, \boldsymbol{\beta}) = E_1(\boldsymbol{x}, \alpha) + E_2(\boldsymbol{x}, \boldsymbol{\beta})$, where

$$E_1(\boldsymbol{x}, \alpha) = \frac{1}{2\alpha}\|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \eta_1 \log \frac{\alpha}{\vartheta_1} + \left(\frac{\alpha}{\vartheta_1}\right)^{r_1}$$

$$E_2(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{R}\boldsymbol{x}\|^2_{\boldsymbol{D}_{\boldsymbol{\beta}}^{-1}} - \eta_2 \sum_{j=1}^{k} \log \frac{\beta_j}{\vartheta_2} + \sum_{j=1}^{k} \left(\frac{\beta_j}{\vartheta_2}\right)^{r_2}$$

Here $\eta_2 \coloneqq r_2 s_2 - \frac{3}{2}$, and $\eta_1 \coloneqq r_1 s_1 - \frac{m+2}{2}$.

Convexity in the variance parameterization

Convexity result for $E_1(\boldsymbol{x}, \alpha)$

Same result in terms of (r, s, η), but η depends on m (size of measurement vector).

Convexity result for $E_2(\boldsymbol{x}, \boldsymbol{\beta})$

Same result, but we have shown explicitly that it still holds for $\ker(\boldsymbol{R}) \neq \{\boldsymbol{0}\}$.

Main observation:

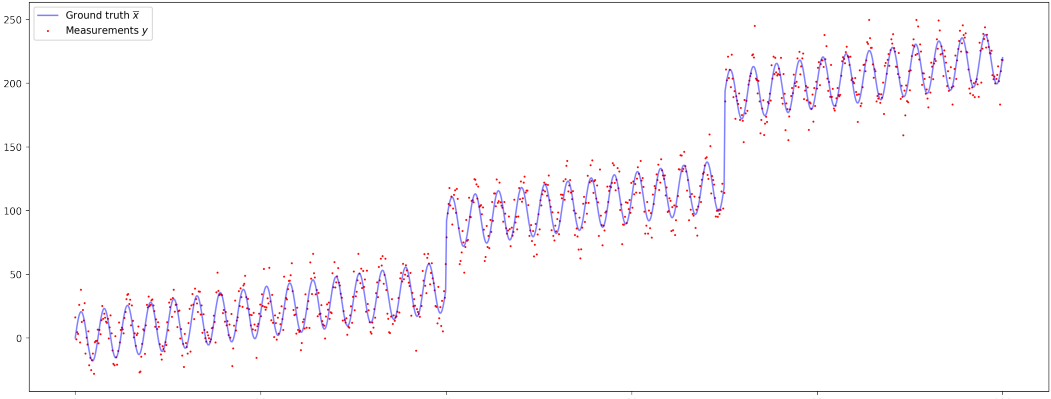
$$\nabla_{\boldsymbol{x}} \nabla_{\boldsymbol{\beta}} \; E_2(\boldsymbol{x}, \boldsymbol{\beta}) = -\boldsymbol{R}^T \boldsymbol{D}_{\boldsymbol{\beta}}^{-1} \operatorname{diag}\left(\boldsymbol{D}_{\boldsymbol{\beta}}^{-1} \boldsymbol{R}\boldsymbol{x}\right)$$

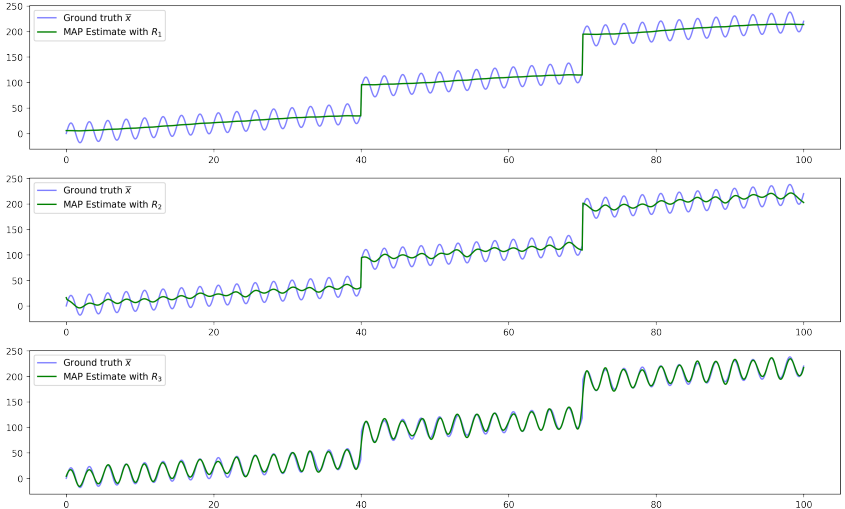
Uninformative hyper-priors

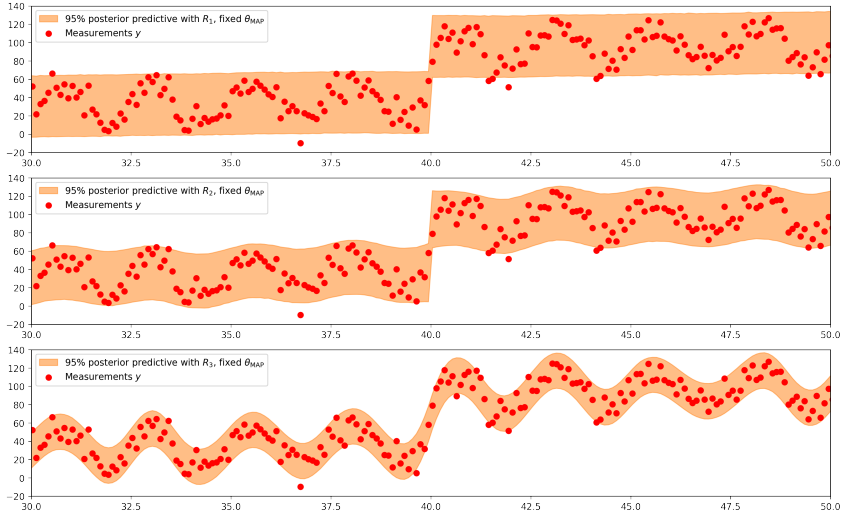
- The choice $r_1 = -1$, $s_1 = 1$, $\vartheta_1 = 10^{-4}$ corresponds to an "uninformative" inverse Gamma variance hyper-prior
- In literature and in practice, accurately recovers true noise variance
- Stationary points satisfy discrepancy principle w.r.t. inferred noise variance
- If a stationary point $(\boldsymbol{x}^\star, \alpha^\star)$ satisfies the discrepancy principle $\|\boldsymbol{F}\boldsymbol{x}^\star - \boldsymbol{y}\|_2^2 \approx m\alpha_{\text{true}}$, then $\alpha^\star \approx \alpha_{\text{true}}$
- However, our sufficient condition for local convexity in this case is that

$$\alpha \leq \frac{4 \times 10^{-4}}{m + 4}$$

- Condition is not useful?


Ground truth $\bar{x}$
Measurements y
250
200
150
100
50
0
0
20
40
60
80
100


Ground truth $\overline{x}$
MAP Estimate with R_1
Ground truth $\overline{x}$
MAP Estimate with R_2
Ground truth $\overline{x}$
MAP Estimate with R_3
0
50
100
150
200
250
0
20
40
60
80
100


95% posterior predictive with R_1, fixed θ_{MAP}
Measurements y
95% posterior predictive with R_2, fixed θ_{MAP}
Measurements y
95% posterior predictive with R_3, fixed θ_{MAP}
Measurements y

Precision parameterization

What changes if α and $\boldsymbol{\beta}$ denote inverse variances (precisions)?

Lemma

$\theta \sim \mathrm{GG}(r, s, \vartheta) \Leftrightarrow \frac{1}{\theta} \sim \mathrm{GG}(-r, s, \vartheta^{-1})$, so we can easily convert between the two parameterizations.

But in general the convexity and the MAP estimate will depend on the parameterization.

Negative log posterior

The negative log posterior is given by $E(\boldsymbol{x}, \alpha, \boldsymbol{\beta}) = E_3(\boldsymbol{x}, \alpha) + E_4(\boldsymbol{x}, \boldsymbol{\beta})$, where

$$E_3(\boldsymbol{x}, \alpha) = \frac{\alpha}{2}\|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|_2^2 - \eta_3 \log \frac{\alpha}{\vartheta_3} + \left(\frac{\alpha}{\vartheta_3}\right)^{r_3},$$

$$E_4(\boldsymbol{x}, \boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{R}\boldsymbol{x}\|_{\boldsymbol{D}_{\boldsymbol{\beta}}}^2 - \eta_4 \sum_{j=1}^{k} \log \frac{\beta_j}{\vartheta_4} + \sum_{j=1}^{k} \left(\frac{\beta_j}{\vartheta_4}\right)^{r_4}.$$

Here $\eta_3 = r_3 s_3 + \frac{m-2}{2}$, $\eta_4 = r_4 s_4 - \frac{1}{2}$.

Convexity in the precision parameterization

Convexity result for $E_3(\boldsymbol{x}, \alpha)$

$$\eta_3 \alpha^{-1} + \frac{r_3(r_3-1)}{\vartheta_3^{r_3}} \alpha^{r_3-1} - \|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|_2^2 \geq 0$$

Convexity result for $E_4(\boldsymbol{x}, \boldsymbol{\beta})$

$$\eta_4 \beta_j^{-1} + \frac{r_4(r_4-1)}{\vartheta_4^{r_4}} \beta_j^{r_4-1} - [\boldsymbol{R}\boldsymbol{x}]_j^2 \geq 0, \quad j = 1, \ldots, k$$

Uninformative hyper-priors

- The choice $r_3 = 1, s_3 = 1, \vartheta_3 = 10^4$ corresponds to an "uninformative" Gamma precision hyper-prior
- From our convexity condition, local convexity only guaranteed in this case if

$$\frac{m}{2} \geq \alpha \|\boldsymbol{F}\boldsymbol{x} - \boldsymbol{y}\|_2^2$$

- If $\boldsymbol{x}^\star$ approximately satisfies the discrepancy principle $\|\boldsymbol{F}\boldsymbol{x}^\star - \boldsymbol{y}\|_2^2 \approx \frac{m}{\alpha_{\text{true}}}$, then this becomes

$$\frac{1}{\alpha} \geq \frac{2}{\alpha_{\text{true}}}$$

IAS in the precision parameterization

- The $\boldsymbol{x}$-update is still given by the solution to a linear system
- What about the α and $\boldsymbol{\beta}$ updates?

Lemma

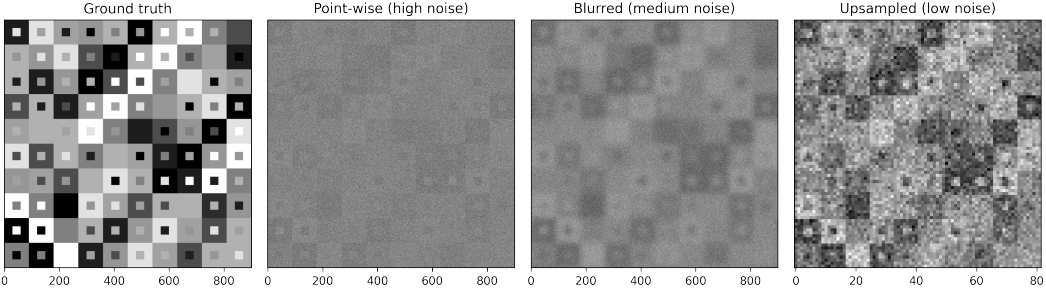
If $r < 0$ and $\eta < 0$, or if $r > 0$ and $\eta > 0$, for $z \geq 0$ and $\xi > 0$ the unique solution to

$$\frac{1}{2}z^2 - \eta\xi^{-1} + r\xi^{r-1} = 0$$

is given by the solution to the initial value problem

$$\psi'(t) = -\frac{2t\psi(t)}{2r^2\psi(t)^{r-1} + t^2}, \quad \psi(0) = (\eta/r)^{1/r}$$

evaluated at $t = |z|$.


Ground truth
Point-wise (high noise)
Blurred (medium noise)
Upsampled (low noise)
0
200
400
600
800
0
200
400
600
800
0
200
400
600
800
0
20
40
60
80

Conditioned on any hyper-parameters $\boldsymbol{\theta}$, the posterior density looks like

$$\pi(\boldsymbol{x} \mid \boldsymbol{y}^{1:p}, \boldsymbol{\theta}) \propto \pi_1(\boldsymbol{x}) \times \cdots \times \pi_K(\boldsymbol{x}), \quad \pi_j(\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}\|\boldsymbol{F}_j\boldsymbol{x} - \boldsymbol{y}_j\|^2_{\boldsymbol{Q}_j}\right\}.$$

Conditioned on any hyper-parameters $\boldsymbol{\theta}$, the posterior density looks like

$$\pi(\boldsymbol{x} \mid \boldsymbol{y}^{1:p}, \boldsymbol{\theta}) \propto \pi_1(\boldsymbol{x}) \times \cdots \times \pi_K(\boldsymbol{x}), \quad \pi_j(\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}\|\boldsymbol{F}_j\boldsymbol{x} - \boldsymbol{y}_j\|^2_{\boldsymbol{Q}_j}\right\}.$$

Then $\pi(\boldsymbol{x} \mid \boldsymbol{y}^{1:p}, \boldsymbol{\theta})$ is the Gaussian $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ with

$$\boldsymbol{Q} = \sum_{j=1}^{K} \boldsymbol{F}_j^T \boldsymbol{Q}_j \boldsymbol{F}_j,$$

$$\boldsymbol{\mu} = \boldsymbol{Q}^{-1}\left(\sum_{j=1}^{K} \boldsymbol{F}_j^T \boldsymbol{Q}_j \boldsymbol{y}_j\right).$$

IAS algorithm for data fusion

Input Data $\{\boldsymbol{y}^{(j)}\}_{j=1}^{p}$, forward operators $\boldsymbol{F}^{(j)}$, sparsifying transformation $\boldsymbol{R}$, hyper-prior parameters, initialization $(\boldsymbol{x}_0, \{\alpha_0\}_{j=1}^{p}, \boldsymbol{\beta}_0)$, number of iterations N.
Output Approximation to the MAP estimate.

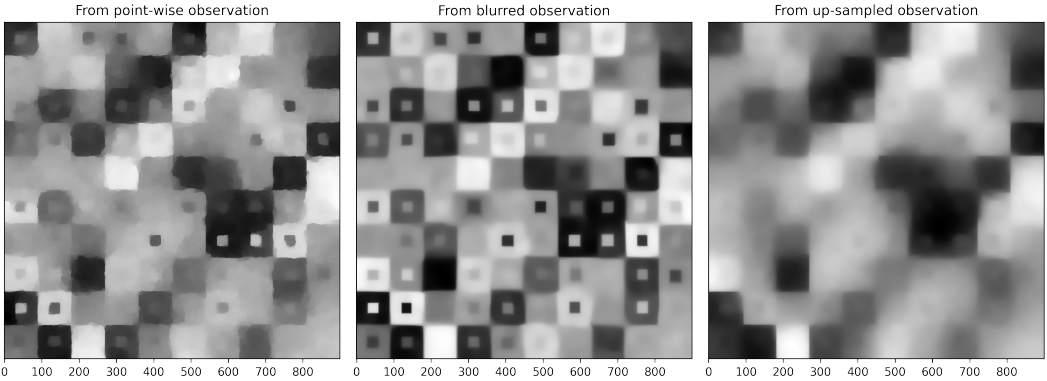
1 For $k = 1, \ldots, N$ do

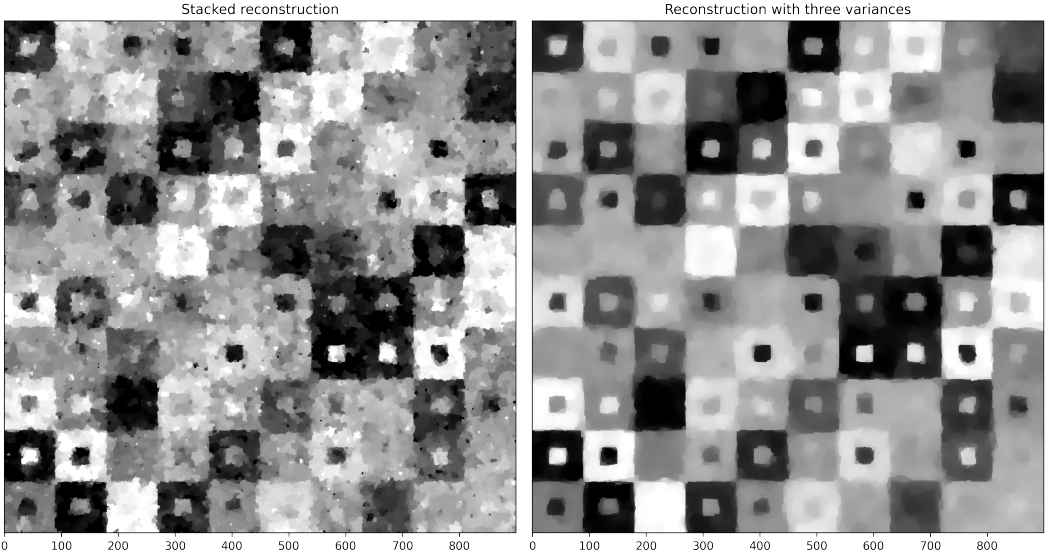
i. Set $\alpha_k^{(j)} = \arg\max_{\alpha^{(j)}} \pi(\boldsymbol{x}_{k-1}, \alpha_{k-1}^{(1)}, \ldots, \alpha^{(j)}, \ldots, \alpha_{k-1}^{(p)}, \boldsymbol{\beta}_{k-1} \mid \boldsymbol{y})$ for each $j = 1, \ldots, p$.

ii. Set $\boldsymbol{\beta}_k = \arg\max_{\boldsymbol{\beta}} \pi(\boldsymbol{x}_{k-1}, \alpha_k^{(1)}, \ldots, \alpha_k^{(p)}, \boldsymbol{\beta} \mid \boldsymbol{y})$.

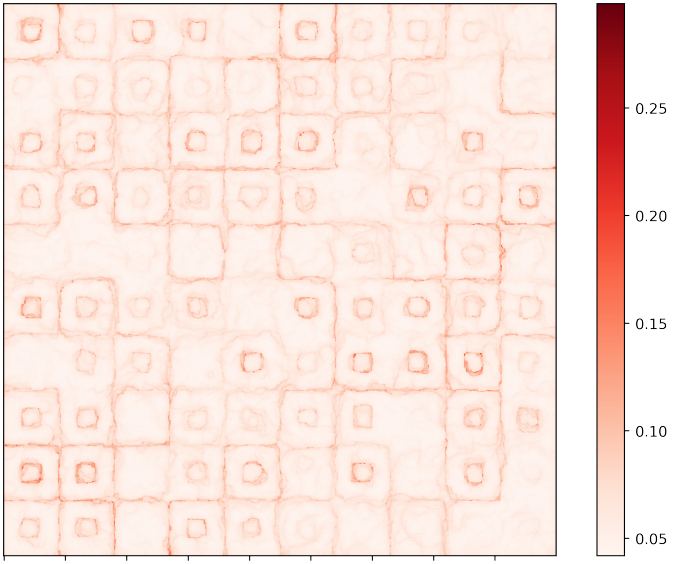
iii. Set $\boldsymbol{x}_k = \arg\max_{\boldsymbol{x}} \pi(\boldsymbol{x}, \alpha_k^{(1)}, \ldots, \alpha_k^{(p)}, \boldsymbol{\beta}_k \mid \boldsymbol{y})$.

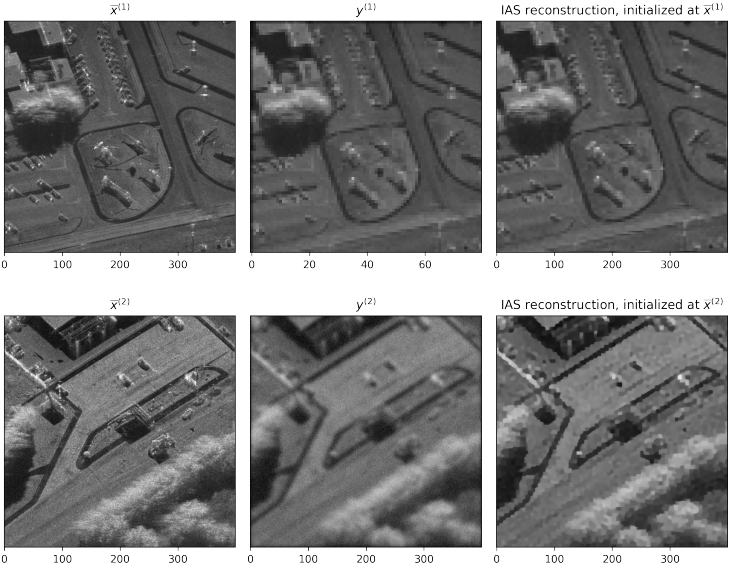
2 Return $(\boldsymbol{x}_N, \{\alpha_N^{(j)}\}_{j=1}^{p}, \boldsymbol{\beta}_N)$.


From point-wise observation
From blurred observation
From up-sampled observation
0 100 200 300 400 500 600 700 800
0 100 200 300 400 500 600 700 800
0 100 200 300 400 500 600 700 800


Stacked reconstruction
Reconstruction with three variances
0
100
200
300
400
500
600
700
800
0
100
200
300
400
500
600
700
800

Reconstruction	MSE	SSIM
From point-wise	0.0223	0.7518
From blurred	0.0113	0.7763
From up-sampled	0.0478	0.7101
Stacked	0.0157	0.6339
With three learned variances	0.0073*	0.8162*


Marginal pixel-wise standard deviation, fixed θ_{MAP}
0
100
200
300
400
500
600
700
800
0.25
0.20
0.15
0.10
0.05


$\bar{x}^{(1)}$
$y^{(1)}$
IAS reconstruction, initialized at $\bar{x}^{(1)}$
$\bar{x}^{(2)}$
$y^{(2)}$
IAS reconstruction, initialized at $x^{(2)}$
0
100
200
300
20
40
60

Marginal penalty

In the variance parameterization, it is known that we can pick (r, s, ϑ) such that the local minimizers agree with those corresponding to ℓ_p-norm $(0 < p < 2)$ and Student-t $(\nu = 2)$ regularization.

In the precision parameterization, we have shown we can do the same to achieve Cauchy regularization.

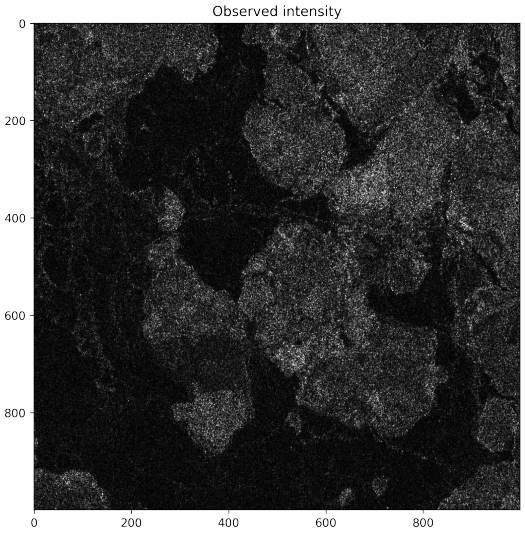
Despeckling

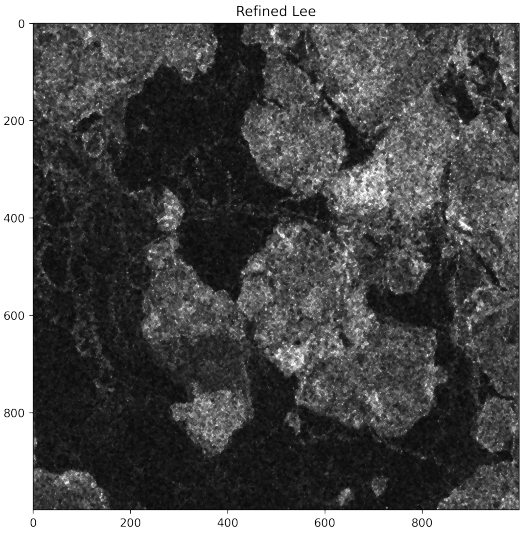
- Multiplicative gamma-distributed noise
- Logarithmic transformation yields a model with log-concave Fisher-Tippett likelihood with

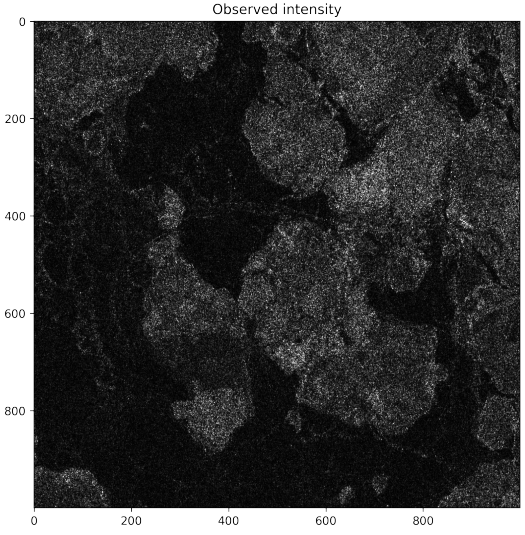
$$-\log \pi(\boldsymbol{y} \mid \boldsymbol{x}) = L \left(\sum_{i=1}^{n} \exp\{\boldsymbol{y}_i - \boldsymbol{x}_i\} + \boldsymbol{x}_i - \boldsymbol{y}_i \right)$$

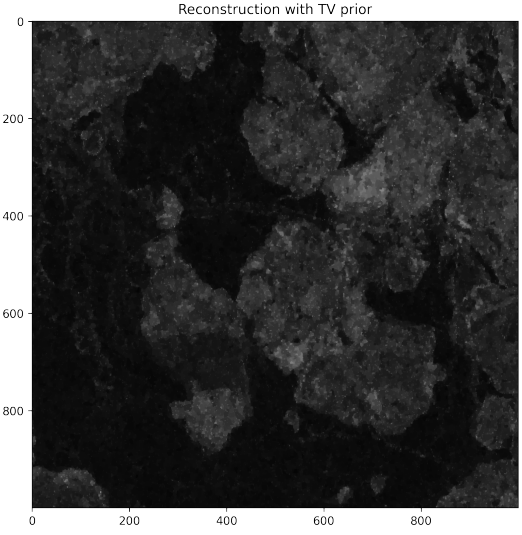
- Coupled with hierarchical sparsity prior, only modification to IAS is solution to nonlinear equation via Newton's method or other

Sentinel-1, courtesy of ESA


Observed intensity
0
200
400
600
800
0
200
400
600
800


Refined Lee
0
200
400
600
800
0
200
400
600
800


Observed intensity
0
200
400
600
800
0
200
400
600
800


Reconstruction with TV prior
0
200
400
600
800
0
200
400
600
800

Data augmentation

- When using a Gaussian prior $\pi(\boldsymbol{x}) \propto \exp\{-\frac{\lambda}{2}\|\boldsymbol{R}\boldsymbol{x}\|_2^2\}$, we are able to make use of a common diagonalization of $\boldsymbol{F}^T\boldsymbol{F}$ and $\boldsymbol{R}^T\boldsymbol{R}$ to directly compute an $\boldsymbol{x}$-update, e.g., if both block-circulant with circulant blocks (BCCB) then

$$\boldsymbol{F}^T\boldsymbol{F} = \boldsymbol{\mathcal{F}}^H\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{\mathcal{F}}, \quad \boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{\mathcal{F}}^H\boldsymbol{\Lambda}_{\boldsymbol{R}}\boldsymbol{\mathcal{F}}$$

[3] Marnissi et al., "A Data Augmentation Approach for Sampling Gaussian Models in High Dimension", 2019.

Data augmentation

- When using a Gaussian prior $\pi(\boldsymbol{x}) \propto \exp\{-\frac{\lambda}{2}\|\boldsymbol{R}\boldsymbol{x}\|_2^2\}$, we are able to make use of a common diagonalization of $\boldsymbol{F}^T\boldsymbol{F}$ and $\boldsymbol{R}^T\boldsymbol{R}$ to directly compute an $\boldsymbol{x}$-update, e.g., if both block-circulant with circulant blocks (BCCB) then

$$\boldsymbol{F}^T\boldsymbol{F} = \boldsymbol{\mathcal{F}}^H\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{\mathcal{F}}, \quad \boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{\mathcal{F}}^H\boldsymbol{\Lambda}_{\boldsymbol{R}}\boldsymbol{\mathcal{F}}$$

- When applying IAS with a sparse prior, $\boldsymbol{D}_{\boldsymbol{\beta}}^{-1}$ "gets-in-the-way" since

$$\boldsymbol{R}^T\boldsymbol{D}_{\boldsymbol{\beta}}^{-1}\boldsymbol{R} = \textbf{?}$$

[3] Marnissi et al., "A Data Augmentation Approach for Sampling Gaussian Models in High Dimension", 2019.

Data augmentation

- When using a Gaussian prior $\pi(\boldsymbol{x}) \propto \exp\{-\frac{\lambda}{2}\|\boldsymbol{R}\boldsymbol{x}\|_2^2\}$, we are able to make use of a common diagonalization of $\boldsymbol{F}^T\boldsymbol{F}$ and $\boldsymbol{R}^T\boldsymbol{R}$ to directly compute an $\boldsymbol{x}$-update, e.g., if both block-circulant with circulant blocks (BCCB) then

$$\boldsymbol{F}^T\boldsymbol{F} = \boldsymbol{\mathcal{F}}^H\boldsymbol{\Lambda}_{\boldsymbol{F}}\boldsymbol{\mathcal{F}}, \quad \boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{\mathcal{F}}^H\boldsymbol{\Lambda}_{\boldsymbol{R}}\boldsymbol{\mathcal{F}}$$

- When applying IAS with a sparse prior, $\boldsymbol{D}_{\boldsymbol{\beta}}^{-1}$ "gets-in-the-way" since

$$\boldsymbol{R}^T\boldsymbol{D}_{\boldsymbol{\beta}}^{-1}\boldsymbol{R} = \textbf{?}$$

- We have made use of recent data augmentation techniques[3] to get around this

[3] Marnissi et al., "A Data Augmentation Approach for Sampling Gaussian Models in High Dimension", 2019.

$$\boldsymbol{v} \,|\, \boldsymbol{x}, \boldsymbol{\beta} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{v}}, \boldsymbol{Q}_{\boldsymbol{v}}^{-1}\right), \quad \boldsymbol{Q}_{\boldsymbol{v}}^{-1} = \frac{1}{\lambda(\boldsymbol{\beta})}\boldsymbol{I} - \boldsymbol{D}_{\boldsymbol{\beta}}^{-1}, \quad \boldsymbol{\mu}_{\boldsymbol{v}} = \boldsymbol{Q}^{-1}\boldsymbol{R}\boldsymbol{x},$$

$$\Rightarrow \boldsymbol{x} \,|\, \boldsymbol{v}, \alpha, \boldsymbol{\beta}, \boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{Q}^{-1}\right), \quad \boldsymbol{Q} = \frac{1}{\alpha}\boldsymbol{F}^T\boldsymbol{F} + \frac{1}{\lambda}\boldsymbol{R}^T\boldsymbol{R}, \quad \boldsymbol{\mu} = \boldsymbol{Q}^{-1}\left(\frac{1}{\alpha}\boldsymbol{F}^T\boldsymbol{y} + \boldsymbol{R}^T\boldsymbol{v}\right).$$

PCG-IAS Algorithm

Choose initial $\boldsymbol{x}_0, \alpha_0, \boldsymbol{\beta}_0, \boldsymbol{v}_0$.
for $k = 1, \ldots, n_{\text{maxits}}$ do
  $\alpha_k \leftarrow \arg\max_{\alpha} \pi(\alpha \mid \boldsymbol{x}_{k-1}, \boldsymbol{\beta}_{k-1}, \boldsymbol{y})$.
  $\boldsymbol{\beta}_k \leftarrow \arg\max_{\boldsymbol{\beta}} \pi(\boldsymbol{\beta} \mid \alpha_k, \boldsymbol{x}_{k-1}, \boldsymbol{y})$.
  $\boldsymbol{v}_k \leftarrow \arg\max_{\boldsymbol{v}} \pi(\boldsymbol{v} \mid \alpha_k, \boldsymbol{\beta}_k, \boldsymbol{x}_{k-1}, \boldsymbol{y})$
  $\boldsymbol{x}_k \leftarrow \arg\max_{\boldsymbol{x}} \pi(\boldsymbol{x} \mid \alpha_k, \boldsymbol{\beta}_k, \boldsymbol{v}_k, \boldsymbol{y})$.
end for