Sparsity promoting hybrid solvers for hierarchical bayesian inverse problems

Hybrid solver for hierarchical Bayesian inverse problems

Daniela Calvetti[a], Monica Pragliola[b], Erkki Somersalo[a]

[a] *Department of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University*
[b] *Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, Bologna, IT*

Abstract

The recovery of sparse generative models from few noisy measurements is an important and challenging problem. Many deterministic algorithms rely on some form of ℓ_1-ℓ_2 minimization to combine the computational convenience of the ℓ_2 penalty and the sparsity promotion of the ℓ_1. It was recently shown within the Bayesian framework that sparsity promotion and computational efficiency can be attained with hierarchical models with conditionally Gaussian priors and gamma hyperpriors. The related Gibbs energy function is a convex functional and its minimizer, which is the MAP estimate of the posterior, can be computed efficiently with the globally convergent Iterated Alternating Sequential (IAS) algorithm [5]. Generalization of the hyperpriors for these sparsity promoting hierarchical models to generalized gamma family yield either globally convex Gibbs energy functionals, or can exhibit local convexity for some choices for the hyperparameters. [6]. The main problem in computing the MAP solution for greedy hyperpriors that strongly promote sparsity is the presence of local minima. To overcome the premature stopping at a spurious local minimizer, we propose two hybrid algorithms that first exploit the global convergence associated with gamma hyperpriors to arrive in a neighborhood of the unique minimizer, then adopt a generalized gamma hyperprior that promote sparsity more strongly. The performance of the two algorithms is illustrated with computed examples.

1 Introduction

The recovery of a sparse vector from noisy indirect observations continues to be an active research topic. After the groundbreaking work on compressed sensing and its connections to sparsity-promoting regularization methods [2, 12, 13, 16, 17] and the ℓ_1-penalty in particular, the interest in sparse recovery has been revived by dictionary learning methods in data science, where the goal is to match an observed vector with few dictionary entries in a huge data base [20, 24, 25]. The connections between regularization methods and penalty functionals on one hand, and Bayesian inference techniques on the other, have been thoroughly investigated [3, 19, 4], and families of priors that promote sparsity have been identified in the Bayesian framework.

Sparsity is a qualitative rather than quantitative trait because in general the size of the support and its location cannot be specified in advance. While there is a wealth of different priors that promote sparsity, the results may differ significantly depending on the cost for non-vanishing entries. In the classical regularization setting, this is well illustrated by the different ℓ_p-penalties, with $p \leq 1$. Penalty functionals with $0 \leq p < 1$ tend to promote sparsity more strongly than $p = 1$. The convexity of the objective function for the latter, and the results on the exact recovery of sparse generative models under suitable conditions have contributed to the popularity of ℓ_1 regularization for sparse problems, while the presence of local minima has damped the enthusiasm for penalties with $p < 1$.

Analogous consideration hold in the Bayesian framework for sparsity promoting hierarchical prior models, with generalized gamma hyperpriors. Recently [5, 6], it has been shown that the Maximum A Posteriori (MAP) iterative sparse reconstruction algorithm is particularly well suited for heavily underdetermined but large scale problems (see, e.g., [8] for an application). The Iterative Alternating Sequential algorithm (IAS) is based on hierarchical Bayesian models, and uses sparsity promoting hyperpriors selected from a family of generalized gamma distributions. As pointed out in [6], some choices of the hyperparameters yield algorithms that are closely related, e.g., to the ℓ_p-penalization methods.

Moreover, the convexity properties of the objective function also depend on the parameter choice, as does the convergence rate of algorithms for computing the MAP estimate. Our aim is to combine the properties of generalized gamma hyperpriors to design robust and computationally efficient methods for sparse recovery from few noisy observations. More specifically, we propose hybrid algorithms for the MAP computation: a gamma hyperprior guides the approximate solution towards the unique minimizer of the objective functions at the beginning, and subsequently a greedier hyperprior is employed to promote sparsity more strongly. We focus on two such hybrid algorithms, that we refer to as local and global because of the different strategy to switch hyperpriors. In the local version, the hyperprior is changed componentwise, guaranteeing local convexity, while in the global version, the hyperprior is changed for all components. In addition to analyzing the convergence properties of each approach, we provide a criterion for ensuring that the a priori beliefs are consistent with the two different hyperpriors. The performance of the algorithms is assessed in the light of computed examples.

2 Hierarchical Bayesian models

In this section we introduce the Iterative Alternating Sequential (IAS) algorithm for the MAP computation, and review some of its key properties: additional details can be found in [8, 5, 6].

Consider the linear observation model with additive Gaussian noise,

$$b = \mathsf{A}x + e\,, \quad e \sim \mathcal{N}(0, \Sigma),$$

where $\mathsf{A} \in \mathbb{R}^{m\times n}$, with $m < n$, is a known ill-conditioned matrix describing the forward model, $x \in \mathbb{R}^n$ is the unknown of interest, $\Sigma \in \mathbb{R}^{n\times n}$ is the symmetric positive definite covariance matrix of the noise. We remark that by letting $\mathsf{A}' = \mathsf{S}\mathsf{A}$ and $b' = \mathsf{S}b$, where S is the Cholesky factor of the precision matrix, $\Sigma^{-1} = \mathsf{S}^\mathsf{T}\mathsf{S}$, we can assume the noise to be white, i.e. $\Sigma = \mathsf{I}$, hence, the likelihood probability density function (pdf) of b with given x takes the form

$$\pi_{b|x}(b \mid x) \propto \exp\Big(-\frac{1}{2}\|\mathsf{A}x - b\|^2\Big).$$

We are interested in estimating x from the observed measurements in b under the a priori assumption that x is sparse, that is, $\|x\|_0 = \mathrm{card}(\mathrm{supp}(x)) \ll n$. In general, the approach can be generalized to cases where the unknown of interest itself is not sparse, but admits a sparse representation in some dictionary, by making the coefficients of the representation the unknown of primary interest.

To encode the sparsity belief in the prior model, we begin by considering a componentwise Gaussian prior model,

$$x_j \sim \mathcal{N}(0, \theta_j)\,, \quad \theta_j > 0\,, \quad 1 \le j \le n,$$

or equivalently,

$$x \sim \mathcal{N}(0, \mathsf{D}_\theta)\,, \mathsf{D}_\theta = \mathrm{diag}(\theta_1, \ldots, \theta_n) \in \mathbb{R}^{n\times n},$$

where the variances of the individual components are not known. The conditional prior density of x given θ is of the form

$$\pi_{x|\theta}(x \mid \theta) \propto \frac{1}{\prod_{j=1}^n \sqrt{\theta_j}} \exp\Big(-\frac{1}{2}\|\mathsf{D}_\theta^{-1/2}x\|^2\Big) = \exp\Big(-\frac{1}{2}\|\mathsf{D}_\theta^{-1/2}x\|^2 - \frac{1}{2}\sum_{j=1}^n \log\theta_j\Big),$$

and following the Bayesian paradigm that all unknowns are modeled as random variables, the a priori belief about θ is encoded in a *hyperprior* pdf $\pi_\Theta(\theta)$. The price to pay for this hierarchical prior model is that we need to estimate not only x but also θ based on data in terms of the joint posterior distribution of (x, θ) conditioned on b,

$$\pi_{x,\theta|b}(x, \theta \mid b) \propto \underbrace{\pi_{x|\theta}(x \mid \theta)\,\pi_\theta(\theta)}_{\pi_{x,\theta}(x,\theta)}\,\pi_{b|x}(b \mid x). \tag{1}$$

A way to promote sparse solutions is to choose a hyperprior π_θ that favors small values of θ but allows occasional large outliers. A family with these properties, thoroughly investigated in [6], is that of the generalized gamma distributions,

$$\pi_\theta(\theta) = \pi_\theta(\theta \mid r, \beta, \vartheta) = \frac{|r|^n}{\Gamma(\beta)^n}\prod_{j=1}^n \frac{1}{\vartheta_j}\left(\frac{\theta_j}{\vartheta_j}\right)^{r\beta-1}\exp\left(-\left(\frac{\theta_j}{\vartheta_j}\right)^r\right),$$

where $r \in \mathbb{R} \setminus \{0\}$, $\beta > 0$, $\vartheta_j > 0$.

The MAP estimate of the posterior pdf model (1) is also the minimizer of the negative logarithm of the posterior pdf,

$$(x^*, \theta^*) = \arg\min_{x,\theta} \left\{ -\log \pi_{x,\theta|b}(x, \theta \mid b) =: \mathcal{F}(x, \theta) \right\}. \tag{2}$$

The objective function $\mathcal{F}(x, \theta)$ can be written as

$$\begin{aligned} \mathcal{F}(x, \theta) &= \mathcal{F}(x, \theta \mid r, \vartheta, \beta) \\ &= \overbrace{\frac{1}{2}\|b - \mathsf{A}x\|^2}^{(a)} + \underbrace{\frac{1}{2}\sum_{j=1}^{n} \frac{x_j^2}{\theta_j} - \left(r\beta - \frac{3}{2}\right)\sum_{j=1}^{n} \log \frac{\theta_j}{\vartheta_j} + \sum_{j=1}^{n} \left(\frac{\theta_j}{\vartheta_j}\right)^r}_{(b)}, \end{aligned} \tag{3}$$

to emphasize that only the terms in (a) depend on x, and only those in (b) depend on θ. These observations play a key role for the design of a computationally efficient algorithm for computing the MAP estimate. We start by recalling the IAS algorithm for the solution of problem (2); see [7, 8, 5] for further details and for a comprehensive study of the effect of the choice of hyperparameters (r, β, ϑ) on the promotion of sparsity and the properties of the objective function.

2.1 IAS algorithm

Given the initial value θ^0, each step of the IAS for problem (2) consists of the two updates,

$$\theta^t \to x^{t+1} \to \theta^{t+1}, \quad t \geq 0,$$

where

$$x^{t+1} = \arg\min_{x} \left\{ \mathcal{F}(x, \theta^t) \right\}, \quad \theta^{t+1} = \arg\min_{\theta} \left\{ \mathcal{F}(x^{t+1}, \theta) \right\}.$$

Due to the particular form of the objective function (3), each step comprises first the computation of the minimizer of (a) with respect to x keeping θ fixed, then the minimizer of (b) with respect to θ with the updated value of x fixed. While this procedure is remarkably similar to the Alternating Direction Method of Multipliers (ADMM) [1], there are some fundamental differences. In fact, while in ADMM, the alternating structure is achieved via an artificial partial decoupling of the fidelity term and the penalty term by introducing auxiliary variables, in IAS the partial decoupling is automatic, with the common term of (a) and (b) being the link between the two minimization tasks. Moreover, both minimization tasks are relatively simple with an exact condition for the minimizer. For some choices of hyperparameters, IAS has been shown to be globally at least linearly convergent [7, 5]. In the following, we review some of the computational details of the IAS algorithm that are particularly relevant for the proposed hybrid schemes.

Update of x The update of x given θ by minimizing part (a) in (3) reduces to solution of a quadratic minimization problem, i.e.,

$$x^{t+1} = \arg\min_{x} \left\{ \|\mathsf{A}x - b\|^2 + \|\mathsf{D}_\theta^{-1/2} x\|^2 \right\}, \ \theta = \theta^t,$$

thus x^{t+1} is the least squares solution of the linear system

$$\begin{bmatrix} \mathsf{A} \\ \mathsf{D}_\theta^{-1/2} \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix}. \tag{4}$$

When the dimensions of the problem and the computing resources make it unfeasible to solve the least squares problem (4) exactly, a suitable approximate solution can be obtained by solving a reduced problem via the Conjugate Gradient for Least Squares (CGLS) algorithm [9], often without any real loss of information [6]. To define the reduced problem, introduce the change of variables,

$$\mathsf{D}_\theta^{-1/2} x = w,$$

which corresponds to a whitening of the conditional prior, and reformulate the linear system (4) in terms of w as

$$\begin{bmatrix} \mathsf{A}_\theta \\ \mathsf{I} \end{bmatrix} w = \begin{bmatrix} b \\ 0 \end{bmatrix}, \ \mathsf{A}_\theta = \mathsf{A}\mathsf{D}_\theta^{1/2}. \tag{5}$$

The least squares solution of the linear system after the change of variables is the Tikhonov regularized solution of

$$\mathsf{A}_\theta w = b \tag{6}$$

with regularization operator equal to the identity and regularization parameter 1. As pointed out in [9], the Tikhonov regularized solution solving (5) is remarkably similar to the solution of (6) computed by the CGLS iteration with an early stopping criterion [6] based on Morozov discrepancy principle and the coupling between the two least squares problems. More precisely, denote the kth Krylov subspace corresponding to the above system by

$$\mathscr{K}_k = \mathscr{K}_k(\mathsf{A}_\theta^\mathsf{T} b, \mathsf{A}_\theta^\mathsf{T}\mathsf{A}) = \operatorname{span}\left\{ \left(\mathsf{A}_\theta^\mathsf{T}\mathsf{A}_\theta\right)^\ell \mathsf{A}_\theta b \mid 0 \le \ell \le k-1 \right\}.$$

Define the Reduced Krylov Subspace (RKS) solution as

$$w^{(k)} = \operatorname{argmin}\left\{ \|b - \mathsf{A}_\theta w\| \mid w \in \mathscr{K}_k \right\},$$

where k is the first index satisfying

$$\|b - \mathsf{A}_\theta w^{(k+1)}\| \le \sqrt{m}, \quad G(w^{(k+1)}) > \tau G(w^{(k)}),$$

where $\tau > 1$, $\varepsilon = \tau - 1 > 0$ is a small safeguard parameter, and the functional G, given by

$$G(w) = \|b - \mathsf{A}_\theta w\|^2 + \|w\|^2,$$

is the objective function approximately minimized by the surrogate reduced model. To update x, we set

$$x^{t+1} = \mathsf{D}_\theta^{1/2} w^{(k)}.$$

The purpose of the early termination of the CGLS iteration is to obtain a good approximation of the solution to (5) with the surrogate reduced model, not to introduce additional regularization. In our setting, the information about the type of solutions to favor is included in the matrix A_θ via the multiplication by the matrix $\mathsf{D}_\theta^{1/2}$.

Update of θ It follows from the independence of the components that the first order optimality condition that needs to be satisfied by the updated θ can be imposed componentwise. Setting the partial derivatives of (b) in (3) with respect to θ_j equal to zero, we find that θ_j must satisfy

$$-\frac{1}{2}\frac{x_j^2}{\theta_j^2} - \left(r\beta - \frac{3}{2}\right)\frac{1}{\theta_j} + r\frac{\theta_j^{r-1}}{\vartheta_j^r} = 0, \quad x = x^{t+1}. \tag{7}$$

While for some values of r, notably $r = \pm 1$, (7) admits an analytic solution, in general we need to solve it numerically. It was shown in [6] that after the changes of variables $\theta_j = \vartheta_j \xi_j$, $x_j = \sqrt{\vartheta_j} z_j$, we may write $\xi_j = \varphi(|z_j|)$, and via implicit differentiation, the function φ satisfies the initial value problem

$$\varphi'(z) = \frac{2z\varphi(z)}{2r^2\varphi(z)^{r+1} + z^2}, \quad \varphi(0) = \left(\frac{\eta}{r}\right)^{1/r}, \tag{8}$$

therefore the updated θ_j can be computed by a numerical time integrator.

We conclude this section with the main results on selecting the model parameters (r, β, ϑ). The values of the parameters r and β affect how strongly the sparsity of the solution is promoted and determine the convexity of the objective function, while the value of ϑ_j can be related to the sensitivity of the data to x_j. Recall that for a linear model $b = \mathsf{A}x + \varepsilon$, a classical measure of the sensitivity of the data b to the component x_j is $\|\mathsf{A}e_j\|$, where $e_j \in \mathbb{R}^n$ is the canonical jth Cartesian unit vector. It was proven recently [8, 5, 6] that under rather natural conditions, a judicious choice of the parameter ϑ is

$$\vartheta_j = \frac{C}{\|\mathsf{A}e_j\|^2},$$

where the constant $C > 0$ is related to the expected sparsity of of the solution and to an estimate of the signal-to-noise ratio (SNR). Due to the connection with sensitivity, we this choice of ϑ is referred to as sensitivity scaling.

3 Hybrid IAS algorithms

In this section, we will propose a hybrid version of the IAS algorithm in which the hypermodel in the generalized gamma family is updated componentwise as the iteration proceeds. The following theorem, see [6] for details, summarizes how the values of the hyperparameters r and β affect the convexity properties of the functional $\mathcal{F}$.

Theorem 1. *Let $\beta > 0$ and $r \neq 0$, and let $\mathcal{F}(x, \theta)$ be the objective function for the minimization problem in (2).*

(a) If $r \geq 1$ and $\eta = r\beta - 3/2 > 0$, the function $\mathcal{F}(x, \theta)$ is globally convex.

(b) If $0 < r < 1$ and $\eta = r\beta - 3/2 > 0$, or, if $r < 0$ and $\beta > 0$, the function $\mathcal{F}(x, \theta)$ is convex provided that

$$\theta_j < \overline{\theta} = \vartheta_j \left(\frac{\eta}{r|r-1|} \right)^{1/r}.$$

As far as the computation of the MAP estimate is concerned, the global convexity of the objective function when $r \geq 1$ is very convenient, although there are several reasons for considering other choices of r that yield hierarchical priors that promote sparsity more strongly. It has been observed that, by and large, the further the objective function is from being globally convex, the stronger the sparsity of the minimizer is promoted. We review below some recent results, see [10, 5], relating generalized gamma hyperpriors and classical sparsity-promoting priors.

Let

$$\begin{aligned} \mathcal{P}(x, \theta \mid r, \beta, \vartheta) &= \frac{1}{2} \sum_{j=1}^{n} \frac{x_j^2}{\theta_j} - \left(r\beta - \frac{3}{2} \right) \sum_{j=1}^{n} \log \frac{\theta_j}{\vartheta_j} - \sum_{j=1}^{n} \left(\frac{\theta_j}{\vartheta_j} \right)^r \\ &= \sum_{j=1}^{n} p(x_j, \theta_j \mid r, \beta, \vartheta_j) \end{aligned}$$

denote the penalty term (b) in the objective function, and express the IAS updating formula (8) for θ_j as a function of x_j

$$g_j(x_j) = \theta_j = \vartheta_j \varphi \left(\frac{|x_j|}{\sqrt{\vartheta_j}} \right).$$

It has been shown in [5] that for $r = 1$, as $\eta \to 0+$ the penalty function $\mathcal{P}(x, \theta, 1, 3/2 + \eta, \vartheta)$ approaches a weighted ℓ_1-penalty in the sense that

$$\lim_{\eta \to 0+} \mathcal{P}(x, g(x) \mid 1, \frac{3}{2} + \eta, \vartheta) = \sqrt{2} \sum_{j=1}^{n} \frac{|x_j|}{\sqrt{\vartheta_j}}$$

and, moreover, the corresponding minimizer x^* found by the IAS algorithm converges to scaled ℓ_1 regularized solution.

More generally as shown in [6], by choosing $r\beta = 3/2$, the penalty function coincides with the weighted ℓ_p-norm, with $p = 2r/(r+1)$,

$$\mathcal{P}\big(x, g(x) \mid r, \frac{3}{2r}, \vartheta\big) = C_r \sum_{j=1}^{n} \frac{|x_j|^p}{\sqrt{\vartheta_j^{\,p}}}, \quad C_r = \frac{r+1}{(2r)^{r/(r+1)}}.$$

While this result holds in general, for $0 < r < 1$ and $\beta = 3/2r$, the model corresponds to ℓ_p penalties with $0 < p < 1$, which are known to promote strongly the sparsity of the solution.

For the inverse gamma hypermodel, corresponding to $r = -1$, the penalty term approaches the Student distribution, a prominently fat tailed distribution favoring large outliers, and leading to a greedy algorithm that strongly promotes sparsity [6]. The main problem with the lack of global convexity of the objective function is that optimization-based algorithms for the MAP computation may stop at a spurious local minimizer.

In this work, we propose two modifications to the IAS algorithm that take advantage of the global convexity of the objective function corresponding to the gamma hyperprior ($r = 1$), and of the stronger sparsity promotion of hierarchical models with $r < 1$ whose associated objective functions are only

locally convex. In both proposed algorithms, the gamma hyperprior is used initially to drive the IAS iterates towards the unique minimizer of the globally convex objective function, then switched to greedier hypermodel. The two different algorithms are referred to as *local* and *global* hybrid models. In the local hybrid, the hyperprior is changed componentwise as soon as the corresponding variance falls inside the convexity region of the second model, while in the global model, the hyperprior is changed for all components after a given number of IAS steps. Next we present the details relative to the two hybrid schemes.

3.1 Local hybrid IAS

We write the objective function $\mathcal{F}(x, \theta \mid r, \vartheta, \beta)$ with the given model parameters (r, β, ϑ) as

$$\mathcal{F}(x, \theta \mid r, \vartheta, \beta) = \|b - \mathsf{A}x\|^2 + \sum_{j=1}^{n} p(x_j, \theta_j \mid r, \vartheta_j, \beta),$$

where

$$p(x_j, \theta_j \mid r, \vartheta_j, \beta) = \frac{1}{2}\frac{x_j^2}{\theta_j} - \left(r\beta - \frac{3}{2}\right)\log\frac{\theta_j}{\vartheta_j} + \left(\frac{\theta_j}{\vartheta_j}\right)^r.$$

Unlike in the standard IAS algorithm, where the parameters r, β and ϑ are kept fixed, the local hybrid algorithm updates the parameters for those component pairs (x_j, θ_j) that satisfies the convexity criterion in Theorem 1 for the second hypermodel.

More precisely, consider two hypermodels with parameters $(r^{(1)}, \beta^{(1)}, \vartheta^{(1)})$ and $(r^{(2)}, \beta^{(2)}, \vartheta^{(2)})$, with $r^{(2)} < 1 \leq r^{(1)}$, $r^{(2)} \neq 0$, referred to as $\mathscr{M}_1$ and $\mathscr{M}_2$, respectively, and start the IAS algorithm with the model $\mathscr{M}_1$.

Let $(x, \theta) = (x^t, \theta^t)$ denote the IAS iterate after t steps. For each component x_j of x, we compute the θ_j update corresponding to model $\mathscr{M}_2$,

$$\theta_j^{(2)} = g(x_j \mid r^{(2)}, \beta^{(j)}, \vartheta_j^{(2)}) = g^{(2)}(x_j).$$

If

$$\theta_j^{(2)} < \overline{\theta}_j = \vartheta_j^{(2)} \left(\frac{\eta^{(2)}}{r^{(2)}|r^{(2)} - 1|}\right)^{1/r^{(2)}}, \tag{9}$$

we update θ_j switching to $\mathscr{M}_2$, otherwise we continue with $\mathscr{M}_1$. Observe that since the function $g^{(2)}$ is strictly increasing for $x_j > 0$, we may write the above condition in terms of x_j,

$$|x_j| < \left[g^{(2)}\right]^{-1}(\overline{\theta}_j) = \overline{x}_j.$$

Let $I \subset \{1, 2, \ldots, n\}$ denote an index set such that

$$j \in I \text{ if and only if } |x_j| < \overline{x}_j,$$

and by I^c its complement. Define the local hybrid objective function,

$$\begin{aligned}\mathcal{F}(x, \theta \mid I) \quad = \quad & \|b - \mathsf{A}x\|^2 + \sum_{j \in I^c} p(x_j, \theta_j \mid r^{(1)}, \vartheta_j^{(1)}, \beta^{(1)}) + \\ & \sum_{j \in I} p(x_j, \theta_j \mid r^{(2)}, \vartheta_j^{(2)}, \beta^{(2)}).\end{aligned}$$

whose convexity can be guaranteed by a bound constraint

$$|x_j| < \overline{x}_j \text{ for } j \in I. \tag{10}$$

It was shown in [6] that to add a bound constraint to the IAS algorithm it suffices to project the updated vector x onto the feasible set. The selection of the hyperparameter $\vartheta^{(j)}$, $j = 1, 2$, deserves some attention. For $\mathscr{M}_1$, the value of $\vartheta^{(1)}$ can be decided by taking sensitivity analysis into consideration, as suggested in [6]. We assign the value of $\vartheta^{(2)}$ based on the following consideration: *If $x_j = 0$, the corresponding variance θ_j should be the same regardless of the choice of the hypermodel, and should reflect the expected*

variance of a background signal. We recall that if $x_j = 0$, the updating of θ_j in the IAS algorithm according to (7) yields

$$g(0 \mid r, \beta, \vartheta_j) = \vartheta_j \left(\frac{\eta}{r}\right)^{1/r}, \quad \eta = r\beta - 3/2,$$

and in order for the two models to agree, it suffices to set

$$\vartheta_j^{(2)} = \left(\frac{\eta^{(1)}}{r^{(1)}}\right)^{1/r^{(1)}} \left(\frac{r^{(2)}}{\eta^{(2)}}\right)^{1/r^{(2)}} \vartheta_j^{(1)}.$$

We are now ready to summarize the proposed local hybrid IAS scheme in algorithmic form. Here we assume that $x \in \mathbb{R}^n$ itself is sparse; suitable adjustments need to be made when the sparsity assumption concerns the increments.

Algorithm 1 Local Hybrid IAS

inputs: Noisy data $b \in \mathbb{R}^m$,
linear forward operator $\mathsf{A} \in \mathbb{R}^{m\times n}$, noise covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{m\times m}$
hyperparameters $(r^{(1)}, \beta^{(1)}, \vartheta^{(1)})$, $(r^{(2)}, \beta^{(2)}, \vartheta^{(2)})$

output: estimated signal and variance $x^*, \theta^* \in \mathbb{R}^n$

1. **initialize:** set $t = 0$, $\theta^t = \vartheta^{(1)}$, $I = \emptyset$
2. **for** *t = 0, 1, 2, ... until convergence* **do**:
3. update x^{t+1} by solving (6)
4. project components x_j^{t+1}, $j \in I$, to $[-\overline{x}, \overline{x}]$
5. **for** *j = 1, ... , n*
6. **if** $\theta_j \geq \overline{\theta}$
7. update $\theta_j^{t+1} = g(x_j^{t+1} \mid r^{(1)}, \beta^{(1)}, \vartheta_j^{(1)})$
8. **else**
9. update $\theta_j^{t+1} = g(x_j^{t+1} \mid r^{(2)}, \beta^{(2)}, \vartheta_j^{(2)})$
10. update $I = I \cup \{j\}$
11. **endif**
12. **end for**
13. $x^* = x^{t+1}$, $\theta^* = \theta^{t+1}$

Before discussing a modification of the above algorithm, a comment on the projection on convexity interval (step 4) is of order. The projection step is included in the algorithm to ensure that the index set I of components being updated using the hypermodel $\mathscr{M}_2$ is monotonically increasing, which, in general, may not be automatically guaranteed. However, the numerical experiments show that the projection step in practice may not be necessary, and the bound constraint $|x_j| < \overline{x}$ is not active.

In [6], the stability of the convexity condition was briefly discussed in terms of the scaled (dimensionless) variables, $z_j = x_j/\sqrt{\vartheta_j^{(2)}}$, $\xi_j = \theta_j/\vartheta_j^{(2)}$. It was shown (see Lemma 4.2 in [6]) that if $\xi_j^t < \overline{\xi}$, then

$$|z_j^{t+1}| \leq M\xi_j^t = M\varphi(|z_j^t|),$$

where φ is the IAS updating function of the scaled variable ξ_j given the current z_j, and M depends on the matrix A and the data b. For example, in the case $r = 1/2$, the convexity bound is $\overline{\xi} = 16\eta^2$. A natural question is whether, if $|z_j^t| < \varepsilon < \overline{z}$, where $\overline{z} = \varphi^{-1}(\overline{\xi})$ is the convexity bound for the scaled variable z_j, it can be guaranteed that $|z_j^{t+1}| < \overline{z}$, or, equivalently,

$$\varphi(|z_j^{t+1}|) < \overline{\xi} = 16\eta^2,$$

where $\eta = r^{(2)}\beta^{(2)} - 3/2$. In [6] it was shown that

$$\varphi(s) = 4\eta^2 + \frac{1}{\eta}s^2 + \mathcal{O}(s^4),$$

therefore,

$$|z_j^{t+1}| \leq M(4\eta^2 + \frac{1}{\eta}\varepsilon^2).$$

To check if (9) is satisfied up to fourth order terms, it suffices to have

$$4\eta^2 + \frac{1}{\eta}(M(4\eta^2 + \frac{1}{\eta}\varepsilon^2))^2 < 16\eta^2,$$

or

$$M^2(4\eta^2 + \frac{1}{\eta}\varepsilon^2)^2 < 12\eta^3,$$

that is,

$$\varepsilon^2 < \frac{\sqrt{12}}{M}\eta^{3/2} - 4\eta^2.$$

The positivity of the right side can be guaranteed by choosing η sufficiently small. While the above estimate is only approximate and qualitative, it conveys the idea that the stability of the convexity condition may depend on the forward model as well as on hyperparameter selection.

3.2 Global hybrid IAS

As the numerical experiments confirm, the gain from switching to the model $\mathscr{M}_2$ for components that are already in the convexity region of that model is not so much in enhancing, e.g., sudden discontinuities in the solution, but more on cleaning the background. An alternative approach is to relax the convexity requirement and use the global convexity of the first model to find a good starting point for the optimization with the second model, taking the risk of minimizing a non-convex objective function from an initial guess sufficiently close to the global minimum of the first model.

More specifically, we run first the IAS algorithm for a fixed number $\bar{t}$ of iterations with model $\mathscr{M}_1$, whose conservative parameter choice guarantees convergence towards a global minimizer, and then switch to the less conservative hypermodel $\mathscr{M}_2$, trading the global convexity for stronger sparsity promotion. We refer to this scheme as *global* hybrid IAS, since the change of hyperprior is carried out at once for all the variances θ_j. unlike in the local version where only selected components followed the model $\mathscr{M}_2$. The computational details are summarized in Algorithm 2.

Algorithm 2 Global Hybrid IAS

inputs: Noisy data $b \in \mathbb{R}^m$,
linear forward operator $\mathsf{A} \in \mathbb{R}^{m\times n}$, noise covariance matrix $\mathsf{\Sigma} \in \mathbb{R}^{m\times m}$
hyperparameters $(r^{(1)}, \beta^{(1)}, \vartheta^{(1)})$, $(r^{(2)}, \beta^{(2)}, \vartheta^{(2)})$
integer $\bar{t} > 0$ defining the switch point

output: estimated signal and variance $x^*, \theta^* \in \mathbb{R}^n$

initialize: set $\theta^0 = \vartheta^{(1)}$
for *t = 0, 1, 2, ... until convergence* **do**:
  update x^{t+1} by solving (6)
  for $j = 1, \ldots, n$
    if $t < \bar{t}$
      update $\theta_j^{t+1} = g(x_j^{t+1} \mid r^{(1)}, \beta^{(1)}, \vartheta_j^{(1)})$
    else
      update $\theta_j^{t+1} = g(x_j^{t+1} \mid r^{(2)}, \beta^{(2)}, \vartheta_j^{(2)})$
    endif
end for
$x^* = x^{(t+1)}$, $\theta^* = \theta^{(t+1)}$

In the description of the Algorithm 2, the value $\bar{t}$ is given as input. Alternatively, one could run the model $\mathscr{M}_1$ until the variances θ stop changing significantly. Since in general we have little information

of the nature of the minima of the objective function when $r < 1$, a definitive automatic switching rule is not easy to justify. We illustrate the performance of the algorithm on a few test cases in the section on computed examples.

4 IAS for sparse increments

The IAS algorithm, and the hybrid versions of it, assume that the unknown has a sparse representation, and x is the vector of coefficients in this representation. In the case where the a priori sparsity belief is not about the signal x but its increments, the IAS algorithms needs to be suitably adapted. In the one dimensional case the changes are rather straightforward, while the treatment in the higher dimensional cases is more delicate.

In the one-dimensional case assume that the unknown is a piecewise constant signal in $\mathbb{R}$ characterized by few discontinuities. If $f(s)$ is the signal, $0 \leq s \leq 1$ and $x_j = f(jh)$, where $h = 1/n$ is the discretization parameter, we may express x in terms of the increments,

$$x_j = x_0 + \sum_{k=1}^{j}(x_k - x_{k-1}), 1 \leq j \leq n,$$

or, letting $y_j = x_j - x_{j-1}$, as

$$x = x_0 e_1 + \mathsf{L}^{-1} y,$$

where

$$\mathsf{L} = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}, \quad e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Assuming, for simplicity that $x_0 = 0$, it follows from the invertibility of L that we may reformulate the problem as estimating y from the observation model

$$b = \mathsf{A}\mathsf{L}^{-1} y + \varepsilon,$$

with the a priori belief that y is sparse. We update x in the IAS algorithm by computing

$$y^{t+1} = \operatorname{argmin}\left\{ \frac{1}{2}\|b - \mathsf{A}\mathsf{L}^{-1}y\|^2 + \frac{1}{2}\sum_{j=1}^{n}\frac{y_j^2}{\theta_j^t} \right\}, \quad x^{t+1} = \mathsf{L}^{-1}y^{t+1},$$

where the sparsity of y is playing a role in the update of the θ

$$\theta^{t+1} = \operatorname{argmin}\left\{ \frac{1}{2}\sum_{j=1}^{n}\frac{(y_j^{t+1})^2}{\theta_j} + \left(r\beta - \frac{3}{2}\right)\sum_{j=1}^{n}\log\left(\frac{\theta_j}{\vartheta_j}\right) - \sum_{j=1}^{n}\left(\frac{\theta_j}{\vartheta_j}\right)^r \right\}.$$

The passage from the signal to its increments is more challenging in dimensions $d \geq 2$ where the one-to-one correspondence no longer holds.

To illustrate how to proceed, consider a quadrilateral graph, the nodes representing, e.g., the pixels in an image that we want to estimate, with adjacent pixels connected by an edge, see Figure 1. Assume for simplicity that the values of the image vanish at the boundary nodes, that we refer to as bound nodes, thus we are only interested in estimating the values at the remaining nodes, referred to as free nodes. Let n_v be the number of the free nodes and n_e the number of edges with at least one free node as an end point, referred to as free edges. Let $\mathsf{L} \in \mathbb{R}^{n_e \times n_v}$ denote the mapping from the free nodal values collected in the vector x to the increments along free edges in the vector y,

$$y = \mathsf{L}x. \tag{11}$$

Since the nodal values at the bound nodes, not included in the vector x, are equal to zero, the matrix L has a trivial null space, i.e., $\mathscr{N}(\mathsf{L}) = \{0\}$.

Let n_ℓ denote the number of all loops T_j in the graph, see Figure 1, including those defined by the edges between bound nodes, and let $\mathsf{M} \in \mathbb{R}^{n_e \times n_\ell}$ be the matrix computing the circulation around each loop by

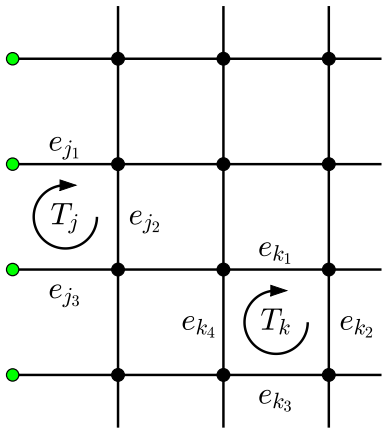

Figure 1: Schematics of the circulation condition $\mathsf{M}z = 0$. In this figure, the black dots indicate the free nodes and the green dots are bound nodes in which the grid function is assumed to vanish. Between free nodes, no edge is defined, thus corresponding to a zero contribution to the circulation. If z is a jump vector corresponding to a grid function x, the sums around edges of each loop must vanish.

summing the increments over edges in clockwise order. If the increments along the edges correspond the nodal values, then the circulation in each element must vanish, i.e., $\mathsf{M}y = 0$ or, equivalently, $y \in \mathscr{N}(\mathsf{M})$ for every $y \in \mathscr{R}(\mathsf{L})$. Since the edge increments associated with the nodal values are computed via the matrix L. Consequently, the matrices L and M define a short exact sequence,

$$\{0\} \longrightarrow \mathbb{R}^{n_v} \xrightarrow{\mathsf{L}} \mathbb{R}^{n_e} \xrightarrow{\mathsf{M}} \mathbb{R}^{n_\ell} \longrightarrow \{0\}.$$

To define a prior promoting sparse increments, we consider a conditionally Gaussian prior model in terms of the increments y_j along the free edges, written concisely as

$$\pi_{y,\theta}(y,\theta) = \pi_{y|\theta}(y \mid \theta)\pi_\theta(\theta) \propto \exp\left(-\frac{1}{2}\sum_{j=1}^{n_e}\frac{y_j^2}{\theta_j} + \phi(\theta)\right), \tag{12}$$

where the function $\phi(\theta) = \phi_{r,\beta}(\theta)$ does not depend on y. It follows from the definition of the increments in terms of the nodal values (11) that $y \in \mathscr{R}(\mathsf{L}) = \mathscr{N}(\mathsf{M})$, therefore the support of the prior is restricted to $\mathscr{N}(\mathsf{M})$. Introducing the auxiliary variable

$$\beta = \mathsf{D}_\theta^{-1/2} y,$$

where $\mathsf{D}_\theta \in \mathbb{R}^{n_e \times n_e}$ is the diagonal matrix with entries θ_j, the compatibility condition on y can be written in terms of β as

$$\beta \in \mathscr{R}(\mathsf{L}_\theta), \quad \mathsf{L}_\theta = \mathsf{D}_\theta^{-1/2}\mathsf{L}.$$

Consider the QR factorization of L_θ,

$$\mathsf{L}_\theta = \mathsf{QR} = \begin{bmatrix} \mathsf{Q}_1 & \mathsf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathsf{R}_1 \\ \mathsf{O} \end{bmatrix},$$

where the orthogonal matrix Q is partitioned in the two blocks, $\mathsf{Q}_1 \in \mathbb{R}^{n_e \times n_v}$, $\mathsf{Q}_2 \in \mathbb{R}^{n_e \times (n_e - n_v)}$, and $\mathsf{R}_1 \in \mathbb{R}^{n_v \times n_v}$ is upper triangular and nonsingular because L is of full rank, and O is the zero matrix of size$(n_e - n_v) \times n_v$. For any $\beta \in \mathscr{R}(\mathsf{L}_\theta)$, there exists $x \in \mathbb{R}^{n_v}$ such that

$$\beta = \mathsf{L}_\theta x = \mathsf{QR}x, \tag{13}$$

hence, by multiplying both sides of (13) by the transpose of Q, we get

$$Q^\mathsf{T}\beta = \begin{bmatrix} \mathsf{Q}_1^\mathsf{T}\beta \\ \mathsf{Q}_2^\mathsf{T}\beta \end{bmatrix} = \begin{bmatrix} \mathsf{R}_1 x \\ 0 \end{bmatrix},$$

or, equivalently,

$$\mathsf{R}_1^{-1}\mathsf{Q}_1^\mathsf{T}\beta = x, \quad \mathsf{Q}_2^\mathsf{T}\beta = 0.$$

Therefore we can express the compatibility condition in terms of the auxiliary variable as

$$\beta \in \mathscr{N}(\mathsf{Q}_2^{\mathsf{T}}) = \mathscr{H}.$$

The posterior density for the prior (12) on the increments,

$$\widetilde{\pi}_{\beta|b,\theta}(\beta \mid b,\theta) \propto \exp\left(-\frac{1}{2}\|b - \mathsf{A}\mathsf{R}_1^{-1}\mathsf{Q}_1^{\mathsf{T}}\beta\|^2 - \frac{1}{2}\|\beta\|^2 + \phi(\theta)\right) \tag{14}$$

if we neglect the compatibility conditions, when restricted to the subspace $\mathscr{H}$ becomes

$$\begin{aligned} \pi_{\beta|b,\theta}(\beta \mid b,\theta) &= \widetilde{\pi}_{\text{post}}(\beta \mid b,\theta) \otimes \delta_{\mathscr{H}}(\beta) \\ &\propto \exp\left(-\frac{1}{2}\|b - \mathsf{A}\mathsf{R}_1^{-1}\mathsf{Q}_1^{\mathsf{T}}\beta\|^2 - \frac{1}{2}\|\mathsf{Q}_1^{\mathsf{T}}\beta\|^2 + \phi(\theta)\right)\Big|_{\mathsf{Q}_2^{\mathsf{T}}\beta=0}, \end{aligned} \tag{15}$$

where $\delta_{\mathscr{H}}$ is the singular measure concentrated on $\mathscr{H}$. The following theorem shows that it is possible to carry out the iterations of the IAS algorithm for the posterior (15) working directly with (14).

Theorem 2. *The vector β_* that maximizes* (14) *satisfies $\mathsf{Q}_2^{\mathsf{T}}\beta_* = 0$, therefore also maximizes* (15)*. Moreover, β_* can be found by minimizing the expression*

$$F(\beta) = \frac{1}{2}\|b - \mathsf{A}\mathsf{L}_\theta^{\dagger}\beta\|^2 + \frac{1}{2}\|\beta\|^2,$$

where $\mathsf{L}_\theta^{\dagger}$ is the pseudoniverse of L_θ.

Proof. From the observation that

$$\|\beta\|^2 = \|\mathsf{Q}^{\mathsf{T}}\beta\|^2 = \|\mathsf{Q}_1^{\mathsf{T}}\beta\|^2 + \|\mathsf{Q}_2^{\mathsf{T}}\beta\|^2,$$

it follows that

$$\begin{aligned} &\frac{1}{2}\|b - \mathsf{A}\mathsf{R}_1^{-1}\mathsf{Q}_1^{\mathsf{T}}\beta\|^2 + \frac{1}{2}\|\beta\|^2 - \phi(\theta) \\ &\qquad = \frac{1}{2}\|b - \mathsf{A}\mathsf{R}_1^{-1}\mathsf{Q}_1^{\mathsf{T}}\beta\|^2 + \frac{1}{2}\|\mathsf{Q}_1^{\mathsf{T}}\beta\|^2 + \frac{1}{2}\|\mathsf{Q}_2^{\mathsf{T}}\beta\|^2 - \phi(\theta). \end{aligned}$$

and its minimizer, for fixed θ, is

$$\beta_* = \operatorname{argmin}\left\{\frac{1}{2}\|b - \mathsf{A}\mathsf{R}_1^{-1}\mathsf{Q}_1^{\mathsf{T}}\beta\|^2 + \frac{1}{2}\|\mathsf{Q}_1^{\mathsf{T}}\beta\|^2\right\}, \quad \mathsf{Q}_2^{\mathsf{T}}\beta^* = 0,$$

which also maximizes (15). Moreover, for β^* such that $\mathsf{Q}_2^{\mathsf{T}}\beta^* = 0$,

$$\mathsf{R}_1^{-1}\mathsf{Q}_1^{\mathsf{T}}\beta_* = \mathsf{L}_\theta^{\dagger}\beta_*,$$

which completes the proof. □

The previous theorem shows that to find the MAP estimate, is not necessary to form the matrix M or to compute the QR factorization of the matrix L_θ. Instead, it suffices to solve the linear system

$$\begin{bmatrix} \mathsf{A}\mathsf{L}_\theta^{\dagger} \\ \mathsf{I} \end{bmatrix} \beta = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

in the least squares sense, because its solution automatically satisfies $\mathsf{Q}_2^{\mathsf{T}}\beta = 0$, thus guaranteeing the existence of a vector x such that (13) holds. The vector $y = \mathsf{D}_\theta^{1/2}\beta$ satisfies the compatibility condition $\mathsf{M}y = 0$, representing feasible and consistent increments along the edges. When resorting to the RKS approximation of the update of the signal inside the IAS iteration, we need to have a procedure to multiply a vector β by the matrix $\mathsf{A}\mathsf{L}_\theta^{\dagger}$ and its transpose. The matrix-vector product of β with $\mathsf{A}\mathsf{L}_\theta^{\dagger}$ can be computed by first solving $\mathsf{L}_\theta\alpha = \beta$ for α in the least squares sense, then multiplying α by A. To evaluate the product of a vector z we observe that the transpose of $\mathsf{A}\mathsf{L}_\theta^{\dagger}$ is

$$\left(\mathsf{L}_\theta^{\dagger}\right)^{\mathsf{T}}\mathsf{A}^{\mathsf{T}} = \mathsf{L}_\theta\left(\mathsf{L}_\theta^{\mathsf{T}}\mathsf{L}_\theta\right)^{-1}\mathsf{A}^{\mathsf{T}},$$

where, fortunately, in our case $\mathsf{L}_\theta^{\mathsf{T}}\mathsf{L}_\theta$ is very sparse. Therefore we solve $\left(\mathsf{L}_\theta^{\mathsf{T}}\mathsf{L}_\theta\right)w = \mathsf{A}^{\mathsf{T}}z$, then multiply the solution w by L_θ.

5 Computed examples

In our evaluation of the performance of the local and global hybrid IAS algorithms, we focus on the following questions:

Stability of the convexity condition in local IAS: To monitor how the components behave with respect to the local convexity region, we run the local hybrid IAS and monitor the behavior of the index set I in Algorithm 1. In particular, we track the indices $j \in I$, pointing to components x_j that enter the local convexity region, satisfying (10), and check whether or not they remain in I without forcing the bound constraint for x_j.

Identification of the support: It is of particular interest to see whether the local hybrid method identifies correctly the support of a generative signal, avoiding stopping at a local minimum that may miss some of the components in the support, as sometimes happens when the non-convex prior models are used. Likewise, with the global hybrid algorithm, we monitor the indices corresponding to the variances in the convexity region at the switching iteration $\bar{t}$ and at the final iteration of Algorithm 2.

In our examples, the hyperpriors for the hybrid schemes are the gamma ($r^{(1)} = 1$) and the inverse gamma ($r^{(2)} = -1$). The performance of local and global hybrid IAS algorithms are also compared with the plain IAS with either the gamma or the inverse gamma hyperprior. In the global hybrid IAS algorithm, the switch to the non-convex model occurs at iteration $\bar{t} = 10$.

Example 1 The first test case is a one-dimensional deconvolution problem. The generative model is a piecewise constant signal $f : [0, 1] \to \mathbb{R}$, $f(0) = 0$, and the data consist of a few discrete noisy observations,

$$b_j = \int_0^1 A(s_j - s) f(s) ds + \varepsilon_j, \quad 1 \le j \le m, \quad A(s) = \left(\frac{J_1(\kappa|s|)}{\kappa|s|} \right)^2,$$

where J_1 is the Bessel function of the first kind and κ is a scalar controlling the width of the kernel that we set to $\kappa = 40$, yielding a kernel with FWHM = 0.08. We discretize the integral as

$$\int_0^1 A(s_j - s) f(s) ds \approx \sum_{j=1}^{n} w_k A(s_j - t_k) f(t_k), \quad 1 \le k \le n,$$

where $t_k = (k-1)/(n-1)$ and the w_k's are the trapezoidal quadrature weights. We generate the data with a dense discretization with $n = n_{\text{dense}} = 1253$, while in the forward model used for solving the inverse problem, we set $n = 500$. The observation points are given by $s_j = (4 + j)/100$, $1 \le j \le m = 91$, and the additive noise is assumed to be scaled white noise, with standard deviation σ set to 2% of the maximum of the noiseless generated signal. We denote $x_j = f(t_j)$. Figure 2 shows the generative signal and the data.

While the generative signal, a piecewise constant function, is not sparse, it admits a sparse representation in terms of its increments $z_j = x_j - x_{j-1}$ over the interval of definition. Assuming $x_0 = 0$, then

$$z = \mathsf{L}x\,, \quad \mathsf{L} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -1 & 1 & \dots & 0 \\ & & \ddots & \\ 0 & \dots & -1 & 1 \end{bmatrix} \in \mathbb{R}^{n \times n},$$

hence

$$x = \mathsf{C}z \quad \text{with} \quad \mathsf{C} = \mathsf{L}^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & & \ddots & \\ 1 & \dots & 1 & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Therefore our inverse problem is to estimate the vector z, assumed to be sparse, from the data vector b, given the forward model

$$b = \mathsf{A}\mathsf{C}z + e, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 \mathsf{I}), \quad \mathsf{A}_{jk} = w_k A(s_j - t_k).$$

The reconstruction results, together with the final variance vector θ and the number of CGLS steps per IAS iteration, are shown in Figure 3. The locations of the first two increments in the generative signal are not easy to detect from the data, see Figure 2, and they are not sharply restored by the IAS algorithm

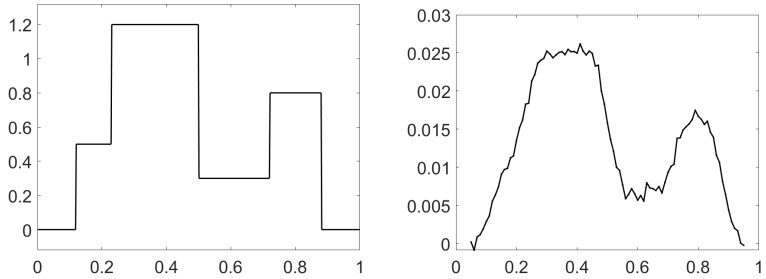

Figure 2: Left: The generative model. Right: the blurred and noisy data vector $b \in \mathbb{R}^{91}$.

with gamma hyperprior (see the first row of Figure 3), while the IAS with inverse gamma (second row) hyperprior lumps the increments, stopping at a local minimizer that corresponds to a simpler profile.

The reconstruction with the local hybrid algorithm is shown in the first panel of the third row of Figure 3. The middle panel of the same row shows in dashed blue the components of θ that follow the inverse gamma distribution at the last iteration of IAS, and in solid red those that never switch from the gamma distribution. The effect of changing to the inverse gamma is a cleaner background. The global hybrid hyperprior returns a sharp restoration of the signal, as shown in the first panel of the fourth row of Figure 3, with the five jumps accurately identified in the correct positions. In both cases, after a few steps, the number of CGLS steps per IAS iteration equals the number of increments detected, indicating that both hybrid IAS algorithms can determine very accurately the cardinality of the support: see also [6].

To address the stability of the convexity condition, we follow iteration by iteration the convexity condition, classifying each index in the sets I (convexity condition satisfied) and its complement I^c (condition not satisfied). The left panel of Figure 4, where the indices in I are marked in green, and those in I^c in yellow, indicate that the set I is monotonously increasing, that is, once a component enters the convexity region, it does not leave it, thus effectively removing the need for imposing the bound constraint (10).

The middle panel of Figure 4 shows the variances θ_j in the global hybrid algorithm at the end of the iteration $\bar{t} - 1 = 9$, prior to switching to the inverse gamma model. The components θ_j satisfying the convexity bound, in dashed blue, are those for which the switch to the inverse gamma distribution does not compromise the convexity of the objective function. The panel on the right indicates for each component at each global hybrid IAS iteration whether it satisfies (green) or not (yellow) the convexity bound. Although in this case, unlike for the local hybrid IAS algorithm, the index set I is not monotonically increasing, eventually the support is correctly detected to high accuracy.

Example 2 The second test case is an image restoration problem. Let Ω be a square compact region in $\mathbb{R}^2$ and x be the generative image defined over Ω. The discrete and noisy data consists of observation at points $q_j \in \Omega$ of a convolved version of the image,

$$b_j = \int_\Omega A(q_j, p')x(p')dp' + \varepsilon_j.$$

with a Gaussian convolution kernel

$$A(p, p') = \frac{1}{2\pi w^2} \exp\left(-\frac{\|p - p'\|_2^2}{2w^2}\right), \quad \text{with} \quad w = 0.015. \tag{16}$$

The integral is discretized over an $n \times n$ pixel grid with $n = 136$, whereas the number of observation points is $m = 68 \times 68$. The noiseless signal is corrupted by additive scaled white noise with standard deviation approximately 2% of the maximum noiseless signal. We assume a priori sparsity of the horizontal and vertical increments of the discrete image x, and implement the sparsity prior in the IAS algorithm according to the procedure detailed in Section 4. The original image, the observed data, vertical and horizontal increments of the original image are shown in Figure 5.

The IAS is performed by constraining the values x_j in the interval $[0, 1]$, $1 \le j \le n^2$. More details on constrained IAS are given in [6].

The restored images computed by the IAS algorithm with gamma, inverse gamma, and by the local and global IAS algorithms are shown in Figure 6. Figure 7 shows the logarithmic plot of variances θ_j,

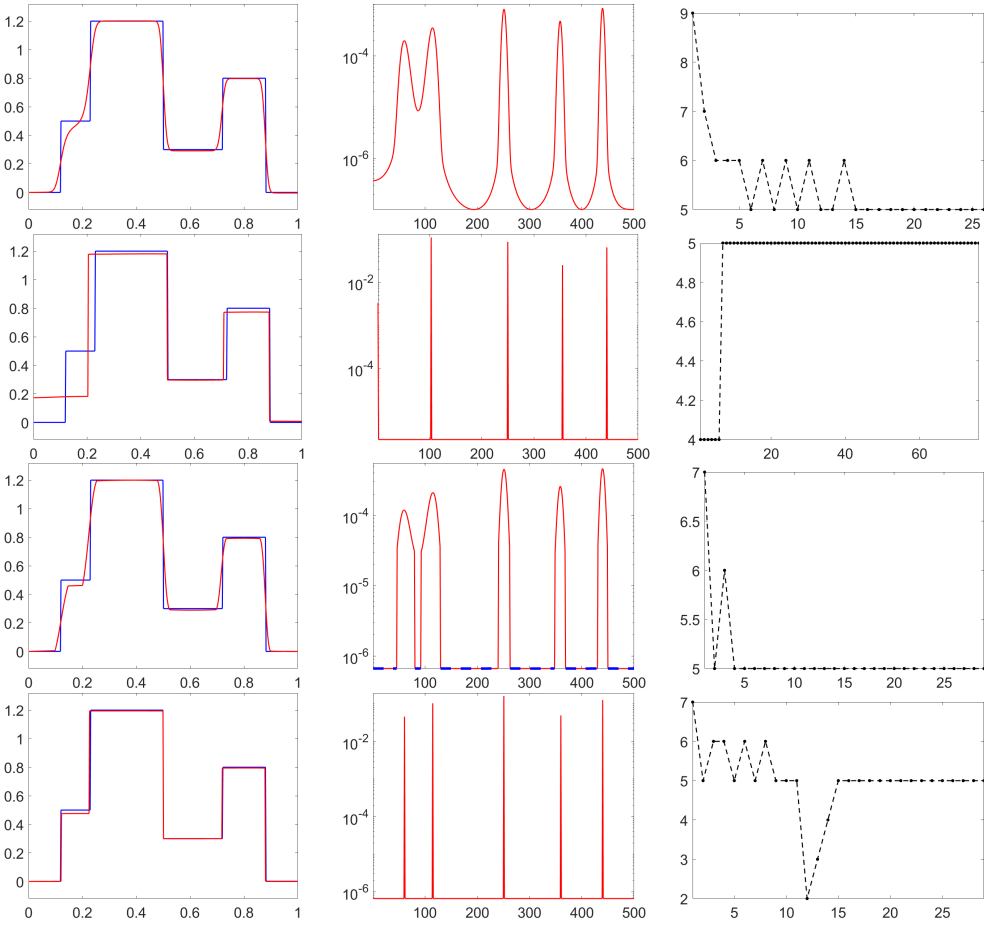

Figure 3: Reconstruction of the signal via gamma, inverse gamma, local hybrid and global hybrid hyperprior (left), the hyperparameter θ (center) and the CGLS iterations per each IAS iteration (right). For the gamma hyperprior in the top row the parameter values are $\eta = 10^{-2}$ and $\vartheta = 10^{-5}$, for the inverse gamma hypeprior in the second row $\eta = -4.5$ and $\vartheta = 10^{-5}$. The hybrid hyperpriors in the bottom rows inherit the parameters from the generative hyperpriors.

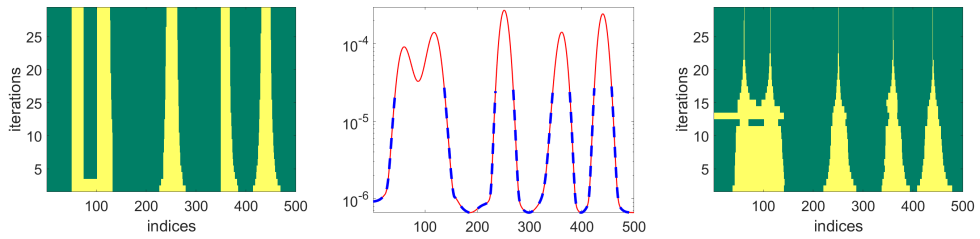

Figure 4: Left: Pseudocolor image of the indices iteration by iteration of the local hybrid algorithm, green indicating the indices of those components that are in the convexity domain (index set I), and yellow those that are outside of it (index set I^c). Observe that when moving up, the yellow intervals shrink and the green ones increase, indicating stable convexity without the need to force the bound constraint (10). Center: Variables θ_j in the global hybrid algorithm at the iteration $\bar{t} - 1 = 9$, before the switch to inverse gamma model. The values in the convexity region are indicated in blue, the rest in red. Right: Pseudocolor image of the indices iteration by iteration in the global hybrid algorithm, green indicating the indices with components in the stability region. Observe that while the algorithm converges, correctly identifying the support, the index sets are not monotonous. In particular, after the switch ($\bar{t} = 10$), the discontinuities close to the left end of the interval create some confusion.

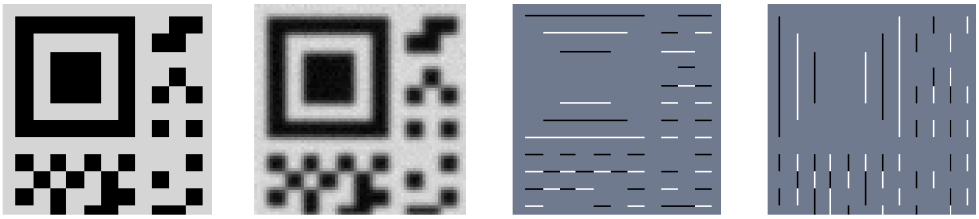

Figure 5: From left to right: original test image $x \in \mathbb{R}^{136\times136}$; observed data $b \in \mathbb{R}^{68\times68}$ corrupted by Gaussian blur and additive Gaussian noise; vector of horizonal increments of the image; and vertical increments.

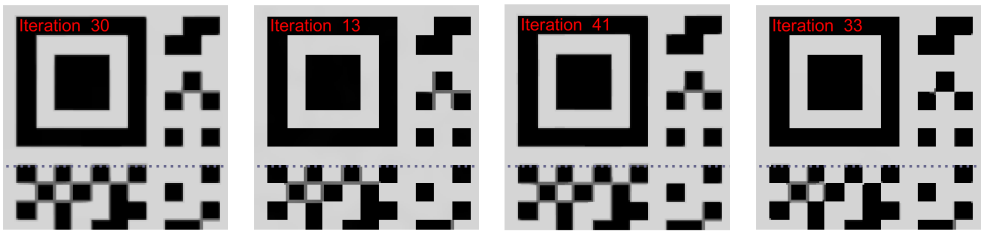

Figure 6: From left to right: Restored images by IAS algorithm based on gamma and inverse gamma hyperpriors, and by the local and global hybrid IAS algorithm using the combination of gamma and inverse gamma models. In the gamma hyperprior, the parameter values are $\eta = 10^{-4}$ and $\vartheta = 10^{-3}$, and in the inverse gamma hypeprior, $\eta = -6.5$ and $\vartheta = 10^{-4}$. The hybrid hyperpriors inherit these parameters from the generative hyperpriors. The dotted horizontal line indicates a cut across the reconstruction defining the profiles shwn in Figure 7.

the profile of the restorations along the dotted horizontal cut lines indicated in the reconstructions of Figure 6, and the profile of the original image. Not surprisingly, the restoration using gamma hyperprior shows slightly rounded corners, while the algorithm based on inverse gamma hyperprior produces some staircasing artifacts along the edges. Both effects are mitigated in the restorations computed with the hybrid IAS algorithms. The reconstruction of the global hybrid IAS algorithm is of remarkably high quality.

The number of CGLS steps in each IAS iteration for the four models is reported in Figure 8.

The left panels of Figure 9 display pseudocolor images of the indices of the variances θ_j of the horizontal and vertical increments at the last iteration of local hybrid IAS, with green corresponding to increments that satisfy the local convexity condition for the inverse gamma, and yellow to the complement. The remaining panels of Figure 9 show the corresponding pseudocolor images for the global hybrid algorithm at the switching iteration $\bar{t}$ (center), and at the last iteration of global hybrid IAS (right), respectively.

Example 3 In the third example, we consider the problem of estimating a nearly black two-dimensional object. The generative model is a starry night impulse image, defined as a distribution on $\Omega = [0,1]\times[0,1]$,

$$d\mu(p) = \sum_{k=1}^{J} a_k \delta(p - p_k)dp, \quad p_k \sim \text{Uniform}(\Omega), \quad a_k \sim \text{Uniform}([1.5, 2])\,,$$

is observed through a Gaussian convolution kernel - see (16), with the discrete and noisy data at observation points $q_j \in \Omega$ given by

$$b_j = \int_\Omega A(q_j, p')d\mu(p') + \varepsilon_j = \sum_{k=1}^{K} a_k A(q_j, p_k) + \varepsilon_j.$$

To solve the inverse problem, we subdivide the image Ω into $n = 128 \times 128 = 16\,384$ pixels, denoted by Ω_ℓ, and let A be the matrix representing the discretized kernel,

$$\int_\Omega A(q_j, p)d\mu(p) \approx \sum_{\ell=1}^{n} \underbrace{|\Omega_\ell| A(q_j, q'_\ell)}_{=\mathsf{A}_{j\ell}} x_\ell, \quad x_\ell = \frac{1}{|\Omega_\ell|}\int_{\Omega_\ell} d\mu(p),$$

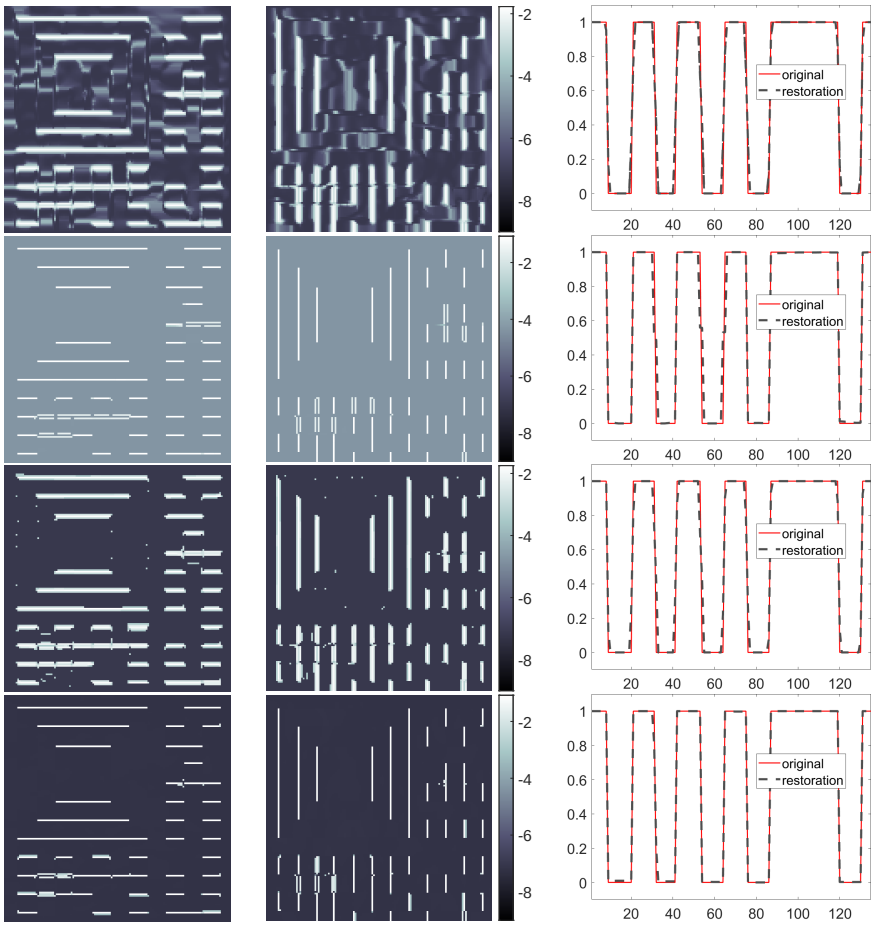

Figure 7: From top to bottom: Logarithmic plots of variances corresponding to vertical and horizontal increments, and on the right, one-dimensional profiles extracted from the restorations in Figure 6 for the gamma, inverse gamma, local hybrid and global hybrid hyperprior.

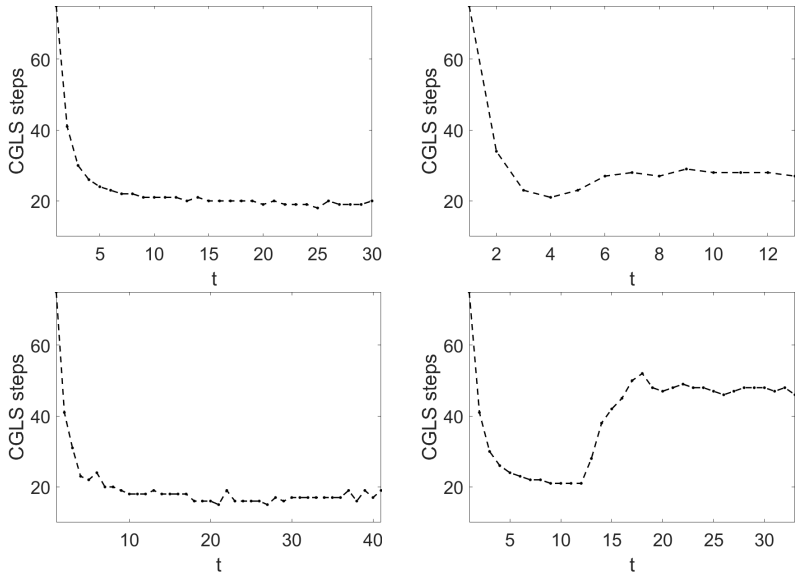

Figure 8: Number of CGLS steps per outer iteration, from top to bottom and from left to right, for gamma, inverse gamma, local hybrid and global hybrid hyperprior.

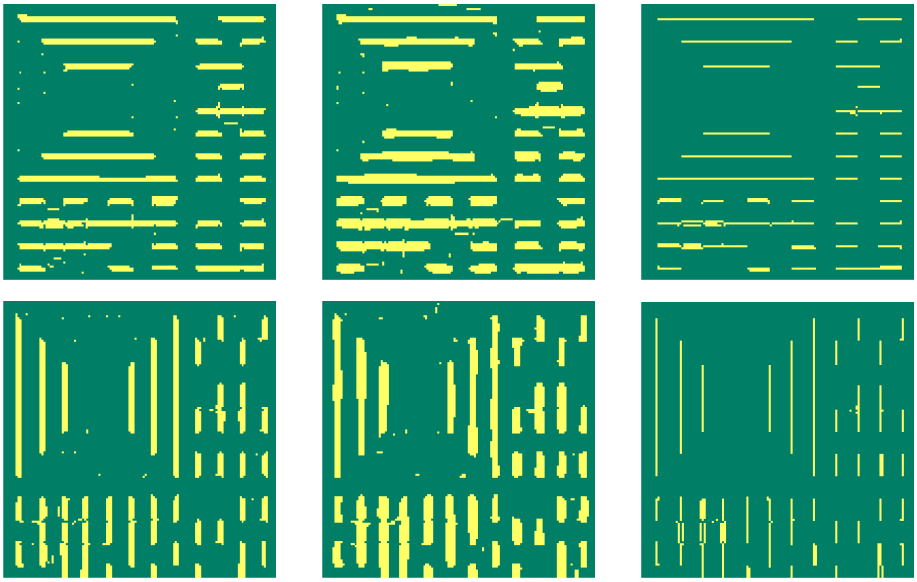

Figure 9: Pseudocolor image of the indices to the variances θ_j for vertical (top) and horizontal increments (bottom) with color coding indicating whether $\theta_j < \bar{\theta}$ (green) or $\theta_j \geq \bar{\theta}$ (yellow). The right panels represent the final iteration of the local hybrid algorithm, middle panels the iteration $\bar{t} - 1$ right before the switch of the global hybrid algorithm, and right panels the final iteration of global hybrid algorithm.

where q'_ℓ denotes the center point of the pixel Ω_ℓ and $|\Omega_\ell|$ is its area. The number of observation points is $m = 64 \times 64 = 4\,096$ and the noiseless signal is corrupted by scaled white noise with standard deviation approximately 1.8% of the maximum noiseless signal. In this case, since the signal itself is sparse, no change of variable is needed. Figure 10 shows the original impulse image characterized by $k = 80$ non-zero points, and the noisy blurred image with kernel width $w = 0.015$.

The restored images and the variances θ represented as pixel images obtained with the IAS algorithm with gamma and inverse gamma hyperpriors, and the local and global hybrid IAS algorithms are shown in Figure 11. The differences in the four algorithms are clearly visible from the estimates of the variance θ and the one-dimensional profiles along the dotted horizontal cut line in the image. The changes in the image of the variances estimated by the IAS with gamma hyperprior (top row, middle) is less pronounced than for the estimates obtained with the other three algorithms, and the intensity of the second star along the cut line (top row, right panel) is significantly lower than in the original image. On the other hand, while the reconstruction from the IAS with inverse gamma hyperprior (second row) is very sharp, the algorithm is too greedy and misses the second star on the horizontal cut line (right panel). Both hybrid reconstructions (third and fourth row) are sharper than that obtained with the gamma hyperprior

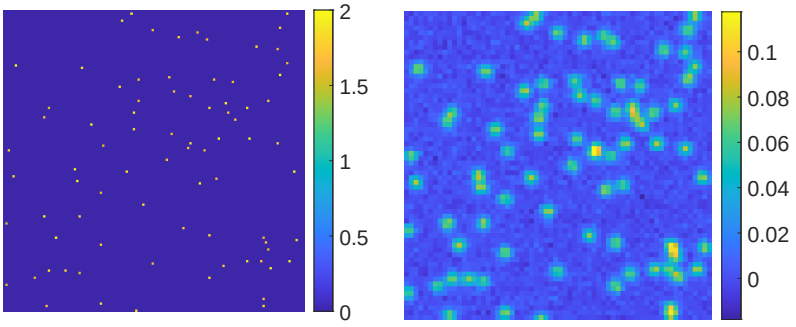

Figure 10: Left: Original generative impulse image plotted on a 128×128 grid as a pixel image. This is the discretization used in the inverse solver, so the pixel image shown here represents the best reconstruction that the algorithm could produce. The reconstructions are compared with this image. Right panel: The 64×64 blurred and noisy observation, degraded by Gaussian blur and additive while Gaussian noise, scaled so as to achieve SNR ≈ 25.

image, and reproduce the original profile with higher fidelity than the IAS algorithm with inverse gamma hyperprior.

Finally, in Figure 12 the behavior of the variances in terms of distribution is shown as a pseudocolor map in the local hybrid case at the final IAS iteration (left panel) and for the global hybrid case at the switching (middle panel) and final (right panel) IAS iteration.

6 Conclusions

In the present work, we discuss the minimization of conditionally Gaussian hypermodels under the adoption of generalized gamma hyperpriors. Based on the results derived in [6], the two proposed hybrid algorithms, namely the local and global hybrid IAS, exploit the global convexity ensured by gamma hyperpriors ($r = 1$) and the stronger sparsity promotion of the generalized gamma hyperpriors with $r < 1$. The local hybrid hypermodel preserves the global convexity characterizing the gamma hyperpriors and, as confirmed by numerical examples, is particularly effective in cleaning the background, while not ensuring a sharp recovery of sudden discontinuities in the signals. On the other hand, the global hybrid hypermodel, which relies on the detection of a suitable initial guess for the minimization of the locally convex hypermodel $\mathscr{M}_2$, returns high quality restorations at the expense of global convexity.

Acknowlwdgements

The work of DC was partly supported by the NSF grant DMS-1522334, and of ES by the NSF grant DMS-1714617. Part of this work was done while the authors DC, MP and ES were visiting Institut Henri Poincaré during the worshop "The Mathematics of Imaging" in March - April 2019 and the Institue for Mathematics and its Applications during the workshop "Computational Imaging" in October 2019. The hospitality and support of IHP and IMA are gratefully acknowledged.

References

[1] Boyd S, Parikh N, Chu E, Peleato B and Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine learning, 3: 1-122.

[2] Bruckstein AM, Donoho DL and Elad M (2009) From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Review, 51(1): 43-81.

[3] Calvetti D and Somersalo E (2007) An introduction to Bayesian scientific computing: ten lectures on subjective computing. Springer New York.

[4] Calvetti D and Somersalo E (2018) Inverse problems: From regularization to Bayesian inference. Wiley Interdisciplinary Reviews: Computational Statistics, 10(3), p.e1427.

[5] Calvetti D, Somersalo E and Strang A (2019) Hierarchical Bayesian models and sparsity: ℓ_2-magic. Inverse Problems, 35: 035003 (26pp).

[6] Calvetti D, Pragliola M, Somersalo E and Strang A (2019) Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors. Inverse Problems, 36: 025010 (29pp).

[7] Calvetti D, Pascarella A, Pitolli F, Somersalo E and Vantaggi B (2015) A hierarchical Krylov-Bayes iterative inverse solver for MEG with physiological preconditioning. Inverse Problems, 31: 125005 (23pp).

[8] Calvetti D, Pitolli F, Prezioso J, Somersalo E and Vantaggi B (2017) Priorconditioned CGLS-based quasi-MAP estimate, statistical stopping rule, and ranking of priors. SIAM Journal on Scientific Computing, 39: 477-500.

[9] Calvetti D, Pitolli F, Somersalo E and Vantaggi B (2018) Bayes meets Krylov: Statistically inspired preconditioners for CGLS. SIAM Review, 60: 429-461.

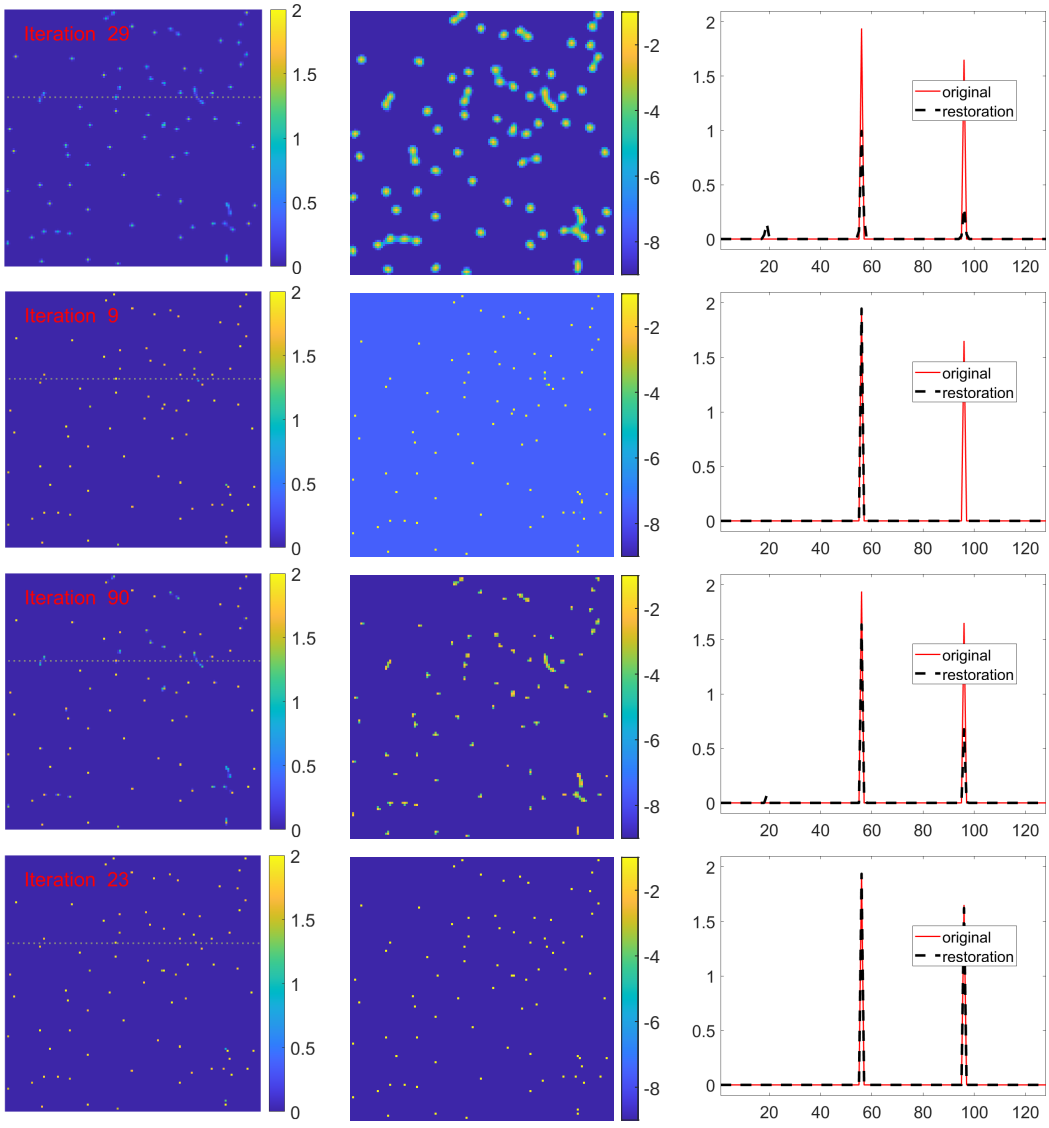

Figure 11: Reconstructions, from top to bottom, using gamma and inverse gamma hyperpriors, and the local and global hybrid models (first column). The second column shows the corresponding variances, and the last column shows the reconstructed profiles along the horizontal dotted line across the image. The profiles are compared to the corresponding profile of the generative model represented as a pixel image in the same grid. For the gamma hyperprior in the top row the parameter values are $\eta = 10^{-5}$ and $\vartheta = 10^{-4}$, for the inverse gamma hypeprior in the second row $\eta = -4.5$ and $\vartheta = 10^{-6}$. The hybrid hyperpriors in the bottom rows inherit the parameters from the generative hyperpriors.

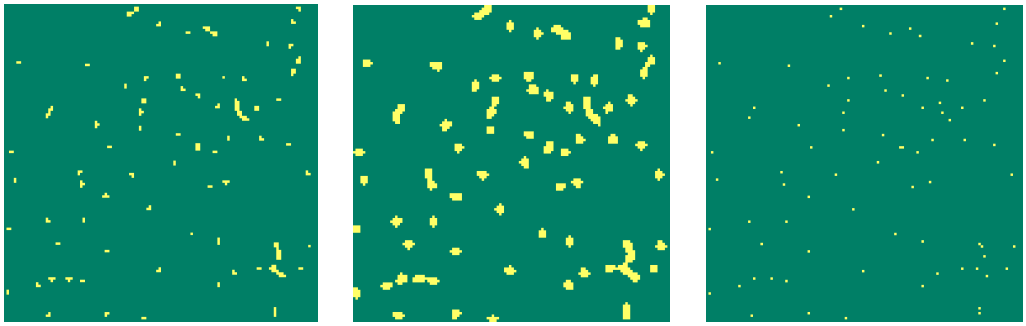

Figure 12: Image of the variances θ_j with color coding indicating if $\theta_j < \bar{\theta}$ (green) or $\theta_j \geq \bar{\theta}$ (yellow). Left panel represents the final iteration of the local hybrid algorithm, middle panel the iteration $\bar{t} - 1$, right before the switch of the global hybrid algorithm, and right panel the final iteration of global hybrid algorithm.

[10] Calvetti D, Hakula H, Pursiainen S Somersalo E (2009) Conditionally Gaussian Hypermodels for Cerebral Source Localization. SIAM on Imaging Science, 2: 879-909.

[11] Calvetti D, S Somersalo E (2007) A Gaussian hypermodel to recover blocky objects. Inverse Problems, 23: 733-754.

[12] Candes E, Romberg J and Tao T (2006) Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, LIX: 1207-1223.

[13] Candes EJ and Tao T (2005) Decoding by linear programming. IEEE Transactions on Information Theory, 51(12): 4203-4215.

[14] Daubechies I, Devore R, Fornasier M and Güntürk CS (2010) Iteratively reweighted least squares minimization for sparse recovery. Communications on Pure and Applied Mathematics, LXIII: 1-38.

[15] DeVore RA. Jawerth B and Lucier B (1992) Image compression through wavelet transform coding. IEEE Transactions on Information Theory, 38(2): 719-746.

[16] Donoho DL (2006) For most large undetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. Communications on Pure and Applied Mathematics, LIX: 907-934

[17] Donoho, DL, Elad M and Temlyakov V. (2006) Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Transactions on Information Theory, 52(1): 6-18.

[18] Gorodnitsky IF and Rao BD (1997) Sparse signal reconstruction from limited data using FOCUSS, a re-weighted minimum norm algorithm. IEEE Transactions on Signal Processing, 45(3): 600-616.

[19] Kaipio J and Somersalo E (2004) *Statistical and Computational Inverse Problems.* Springer Verlag, New York.

[20] Kreutz-Delgado K, Murray JF, Rao BD, Engan K, Lee TW and Sejnowski TJ (2003) Dictionary learning algorithms for sparse representation. Neural computation, 15(2): 349-396.

[21] Lin FH, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW and Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage, 31(1): 160-171.

[22] Li Y and Oldenburg DW (1996) 3-D inversion of magnetic data. Geophysics, 61(2): 394-408.

[23] Li Y and Oldenburg DW (1998) 3-D inversion of gravity data. Geophysics, 63(1): 109-119.

[24] Mairal J, Bach F, Ponce J and Sapiro G (2009) Online dictionary learning for sparse coding. In Proceedings of the 26th annual international conference on machine learning: 689-696.

[25] Tariyal S, Majumdar A, Singh R and Vatsa M (2016) Deep dictionary learning. IEEE Access, 4: 10096-10109.