# Analysis of a prior model for a blocky inversion of seismic amplitude versus offset data

Short title: **Blocky inversion of seismic AVO data**

**Ulrich Theune***

*StatoilHydro Research Center, 7005 Trondheim, NORWAY,*

*uth@statoilhydro.com*

**Ingrid Østgård Jensås**

*StatoilHydro, 7501 Stjørdal, NORWAY,*

*ioj@statoilhydro.com*

**Jo Eidsvik**

*Department of Mathematical Sciences, NTNU, 7491 Trondheim, NORWAY,*

*joeid@math.ntnu.no*

**Corresponding author: uth@statoilhydro.com*

Manuscript submitted: May 18, 2009

Revised manuscript submitted:

## ABSTRACT

Resolving thinner layers and better focussing layer boundaries in inverted seismic sections is an important challenge in exploration and production seismology, as this will help to better identify a potential drilling target. It is generally accepted that the limitation of resolution in seismic inversion is due to the physical processes involved (namely the propagation of a relative low-frequency wavelet) and the way we often invert seismic data. Many seismic inversion methods are based on a least-square optimization approach that can intrinsically lead to unfocussed transitions between adjacent layers. It is therefore necessary to include constraints in the inversion that enhance the contrasts at layer boundaries. In this article, we describe a Bayesian seismic amplitude versus offset inversion algorithm that enforces sharper boundaries between layers by enforcing sparseness in the vertical gradients of the inversion results. The underlying principle idea is similar to high resolution processing algorithms and has been adapted from digital image sharpening algorithms. In particular, we investigate the potential of two statistical distribution, the Cauchy and the Laplace functions, for their potential to improve the contrasts between layers. However, including these constraints into a seismic inversion algorithm results in a large, sparse and non-linear system of equations. We describe the statistical background, develop a numerical algorithm for such inversion problems, and suggest statistical routines to determine the relevant distribution parameters that control the suggested seismic inversion. We then demonstrate the feasibility and potential of the

method with synthetic and field data. These examples demonstrate that the inversion algorithm is indeed capable of detecting and enhancing layer boundaries in seismic inversion. The analysis of the synthetic data suggests that the Laplace constraint performs more reliably, while the Cauchy constraint may not find the optimum solution by converging to a local minimum of the cost function.

## INTRODUCTION

The inversion of seismic data is an important contribution in reservoir characterization, as it allows us to estimate physical properties of the subsurface from observations. One major challenge in inverting seismic data is to achieve a sufficient resolution in order to distinguish thin layers, which may, for example, contain a hydrocarbon reservoir of interest. In general, two major causes may limit the seismic resolution in seismic imaging. Firstly, it is generally accepted that layers thinner than one quatre of a characteristic wavelength cannot be resolved in seismic data (e.g., Sheriff, 2002). A second cause is related to the seismic inversion process itself, that often leaves unfocused boundaries between adjacent layers. It is this second limitation of vertical resolution that we aim to address in this paper.

Inverse problems are usually solved by minimizing the differences between the observed data and synthetic data that may be modeled based on some knowledge about the physics processes involved and an initial model of the subsurface properties. Since seismic inversion is an ill-posed problem, regularization is necessary to obtain a unique and stable solution (e.g., Jackson,

1979). A well chosen regularization term will improve the chances of finding an optimum solution that contains the best possible information about the properties of the earth given the data and prior knowledge. It can also impose certain desirable characteristics onto the solution. A common assumption is a normally distributed misfit (or "noise") between the observations and the synthetic data. The prior model is often assumed to be normally distributed (e.g., Buland and Omre, 2003, among others). However, other prior assumptions (e.g., Amundsen, 1991) and data misfit descriptions (e.g., Farquharson and Oldenburg, 1998) are possible as well.

In the inversion scheme presented in this paper, the misfit and model values are also assumed to be normally distributed, but another constraint is added that ensures a more "blocky" structure of the inversion result. Here, a blocky structure represents an inversion profile in time that exhibits sharp contrasts between adjacent layers, while variations within a layer are suppressed. A similar problem exists in image processing algorithms, in particular in image sharpening applications, where contrasts between adjacent objects are essential for a sharp image. In fact, the choice of inversion constraint presented here is motivated by the work of Geman and Young (1995) and Charbonnier et al. (1997) on image sharpening algorithms.

In this paper, we present work on practicability tests for such algorithms for seismic inversion. Our main focus is on investigating the potentials of the Cauchy and the differentiable Laplace constraints to improve the vertical resolution in seismic inversion. The Cauchy norm is often used in seismic signal processing in order to achieve sparse data representation. For example, Sac-

chi and Ulrych (1995) demonstrated its usefulness for a better representation of seismic data in the Radon domain. Amundsen (1991) used the Cauchy norm in a seismic inversion scheme and found this inversion approach more robust in the presence of noise. However, the inversion algorithm presented here applies the Cauchy or Laplace norm to the gradients of the model data, as proposed by Charbonnier et al. (1997). We note that similar constraints have also been used in geophysical imaging studies (Oldenburg et al., 1983; Portniaguine and Zhdanov, 1999; Youzwishen and Sacchi, 2006) in order to better focus migration and inversion. The problem presented here with seismic prestack inversion needs to be posed in a slightly different way: the aim is to achieve a better vertical resolution by imposing a blocky structure on the seismic inversion result. However, in the horizontal direction, the values should be laterally continuous to reflect the nature of sedimentary layers in the earth.

In the following, we describe the proposed inversion scheme based on a modified least-square inversion technique that enforces sharp layer boundaries in the vertical direction. We start with describing the formulation of the blockyness model. Then, a mathematical description of the inversion approach and discussion related to numerical issues follows, before the applicability is demonstrated with synthetic test data and an example from the Norwegian Sea.

## BLOCKYNESS MODEL FOR SEISMIC INVERSION

The inversion algorithm presented here can be derived from Bayes' theorem,

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{m})} \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}), \tag{1}$$

which relates the posterior distribution $p(\mathbf{m}|\mathbf{d})$ to any prior knowledge about the model, $p(\mathbf{m})$, and the likelihood of the data, $p(\mathbf{d}|\mathbf{m})$. A common model for the likelihood is used, and is therefore only shortly described. The main part of this section focuses on the blockyness constraint, which we include in the prior model $p(\mathbf{m})$.

Inversion of seismic prestack data $\mathbf{d}$, which are represented for several traveltimes ($j$), seismic traces ($i$), and angles ($k$) is the topic of interest. Thus, $d_{ijk} = d_{ij}(\theta_k)$ is the recorded seismic reflection data at incidence angle $\theta_k$, $k = 1, \ldots, N_\theta$, for seismic trace $i$, and at traveltime $j$. The lateral coordinates $i = 1, \ldots, N_1$, could represent $N_1$ inlines / crosslines, organized on a grid, or simply a 2D seismic line, and the discrete traveltimes are indicated by $j = 1, \ldots, N_2$. Further, the collection of data along trace $i$ is denoted $\mathbf{d}_i$. Similarly, the parameter model at trace $i$ is $\mathbf{m}_i$, and the model at traveltime $j$ is $\mathbf{m}_j$. At each sample $ij$, the model consists of the logarithms of the three seismic parameters, which are P- and S-wave velocity and density. Later, $m_{ijl}$ is also needed, where $l = 1, 2, 3$ denotes these three elastic parameters at trace $i$ and traveltime $j$.

For each seismic trace, the data are assumed to be linearly related to the elastic properties of the earth at this trace,

$$\mathbf{d}_i = \mathbf{G}_i\mathbf{m}_i + \mathbf{n}_i = \mathbf{W}_i\mathbf{A}_i\mathbf{D}_i\mathbf{m}_i + \mathbf{n}_i,\ i = 1, \ldots, N_1. \tag{2}$$

In this expression, $\mathbf{G}_i$ represents the modeling operator and $\mathbf{n}_i$ stands for the omnipresent noise, or more precisely, the misfit between the observed data $\mathbf{d}_i$ and the synthetic data $\mathbf{G}_i\mathbf{m}_i$. This misfit consists of random noise and of signal in the data that is not described by the modeling operator $\mathbf{G}_i$. This noise term is assumed Gaussian with zero mean and covariance matrix $\mathbf{\Sigma}_{d,i}$. As shown, the modeling operator consists of three parts: $\mathbf{W}_i$ is a convolution matrix containing aptly chosen wavelets and $\mathbf{A}_i$ is the reflectivity matrix. In principle these two can vary laterally, thus the index $i$. In this work, the weak contrast approximation to the Zoeppritz equations (Aki and Richards, 2002) is assumed valid. Finally, $\mathbf{D}_i$ is a first order vertical differential operator, identical for all traces. A more detailed description of these operators is found in the literature (e.g., Buland and Omre, 2003).

**Data likelihood**

The different data traces are treated as independent so that the likelihood of the data $\mathbf{d}$ for all traces can be summarized by

$$p(\mathbf{d}|\mathbf{m}) = \prod_{i=1}^{N_1} p(\mathbf{d}_i|\mathbf{m}_i) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{N_1}(\mathbf{d}_i - \mathbf{G}_i\mathbf{m}_i)'\mathbf{\Sigma}_{d,i}^{-1}(\mathbf{d}_i - \mathbf{G}_i\mathbf{m}_i)\right). \quad (3)$$

In short, all $\mathbf{G}_i$ and $\mathbf{\Sigma}_{d,i}$ can be combined into two large block diagonal matrices leading to

$$p(\mathbf{d}|\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{G}\mathbf{m})'\mathbf{\Sigma}_d^{-1}(\mathbf{d} - \mathbf{G}\mathbf{m})\right). \quad (4)$$

For convenience, the negative log-likelihood $l_d(\mathbf{m})$ is used later, which is defined as

$$l_d(\mathbf{m}) = const + \frac{1}{2}(\mathbf{d} - \mathbf{G}\mathbf{m})'\mathbf{\Sigma}_d^{-1}(\mathbf{d} - \mathbf{G}\mathbf{m}), \quad (5)$$

where *const* is a constant term independent on $\mathbf{m}$. In the applications described here, we assume that the data noise can be described by uncorrelated normal distributions, such that the noise covariance matrix is given by $\mathbf{\Sigma}_d = \sigma^2\mathbf{I}$. Remember that the noise terms $\mathbf{n_i}$ in equation (2) consist of random noise and data signal that is not described by the modeled data $\mathbf{Gm}$. The latter contribution could in theory be laterally or vertically correlated. However, estimating and describing such correlated noise is challenging and beyond the scope of the work reported here. Therefore, we assume that the simplification of using uncorrelated noise is accurate enough for our purpose.

**Prior model**

The prior model is composed of two parts, which are first a Gaussian term with the expected level and lateral correlation, and secondly a blockyness term with heavy tails for the vertical gradients of the model. Prior knowledge about the mean level of the model is summarized in the vector $\boldsymbol{\mu}$, and the lateral covariance structure at each traveltime $j$ is represented by $\mathbf{\Sigma}_{m,j}$. The mean $\boldsymbol{\mu}$ takes three values at each traveltime and trace; one for each elastic parameter. In practice, the mean level typically varies in traveltime, while it might be constant laterally. In this Gaussian part the model at different traveltimes are assumed independent, i.e.

$$\begin{aligned} p_{lat}(\mathbf{m}) &\propto \exp\Big[-\frac{1}{2}(\mathbf{m}-\boldsymbol{\mu})'\mathbf{\Sigma}_m^{-1}(\mathbf{m}-\boldsymbol{\mu})\Big] \\ &= \exp\Big[-\sum_{j=1}^{N_2}\frac{1}{2}(\mathbf{m}_j-\boldsymbol{\mu}_j)'\mathbf{\Sigma}_{m,j}^{-1}(\mathbf{m}_j-\boldsymbol{\mu}_j)\Big]. \end{aligned} \tag{6}$$

The compact notation for the covariance matrix is used such that $\mathbf{\Sigma}_m$ is a

large block diagonal matrix composed of $\mathbf{\Sigma}_{m,j}$, $j = 1, \ldots, N_2$. This lateral covariance matrix is specified further below. The blockyness constraint is added to enforce sparseness in the vertical direction, such that the negative log-prior model is defined by $l_m(\mathbf{m}) = const + l_{lat}(\mathbf{m}) + l_{vert}(\mathbf{m})$, and now becomes

$$l_m(\mathbf{m}) = const + \frac{1}{2}(\mathbf{m} - \boldsymbol{\mu})' \mathbf{\Sigma}_m^{-1} (\mathbf{m} - \boldsymbol{\mu}) + \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \sum_{l=1}^{3} C\left( \frac{[\mathbf{D}_i(\mathbf{m}_i - \boldsymbol{\mu}_i)]_{jl}}{\kappa_l^2} \right). \tag{7}$$

In the last term, the subscript notation $jl$ refers to a vector element $l$ at traveltime $j$ in trace $i$ and the $\kappa_l$ parameters, $l = 1, 2, 3$ in the denominator are scaling parameters, which can be different for the three elastic parameters. The kernel of the additional vertical constraint $C(x)$ is a regularization term that enforces blockyness in the inversion, and is defined as

$$C(x) = \begin{cases} \frac{1}{2}x^2 & \text{if Gaussian constraint} \\ \log(1 + x^2) & \text{if Cauchy constraint} \\ \sqrt{1 + x^2} - 1 & \text{if differentiable Laplace.} \end{cases} \tag{8}$$

The differentiable Laplace constraint in (8) is similar to the Laplace distribution, but it is differentiable at $x = 0$, which is an essential criterion for the inversion problem to be developed below. The differentiable Laplace distribution defines a convex function, i.e. $C(y) > C(x) + C'(x) \cdot (y - x)$, for $y > x > 0$, and similarly for negative arguments, since $C$ is an even function. The Cauchy constraint, on the other hand, defines a non-convex function which increases very slowly, allowing some large values with small probabil-

ities far out in the tail, while most values concentrate near 0. In fact, the tails of the Cauchy distribution are so heavy that the expected value is not defined since $\int_{-\infty}^{\infty} x^{\alpha} \exp[-C(x)]dx = \infty$, $\alpha \geq 1$.

Let us discuss the principal idea with a conceptual example. Figure 1 shows schematically an idealized inversion problem consisting of two layers with different seismic properties (solid line). A potential result of a standard minimum norm inversion is illustrated by the dotted line. It shows many features that often make the interpretation difficult: it exhibits strong side lobes and the boundary between the two layers is rather smeared out. On the contrary, the schematic blocky inversion solution is rather constant within the layers and follows the interface closer. One way of interpreting these artifacts in the minimum norm solution is that the inversion intrinsically allows for a rather wide distribution of vertical gradients in the inverted property. In contrast, the Cauchy distribution, when applied to the gradients of the model parameters, encourages only a few large changes in the vertical direction, while most small variations with depth will be small. Similar statements apply to the differentiable Laplace distribution. Figure 2 (left) illustrates the Cauchy and differentiable Laplace constraint functions $C(x)$ along with a Gaussian function, while Figure 2(right) gives the corresponding normalized density functions. In Figure 2(right), the plots are scaled so that 90% of the probability mass lies between $x = -1$ and $x = 1$. This means that the scale parameter $\kappa$ is set to 0.16 for Cauchy, 0.37 for differentiable Laplace and 0.6 for Gaussian.

The covariance structure for the lateral part remains to be specified. For

every traveltime, $\mathbf{\Sigma}_{m,j}$ is composed of two components: i) a $3 \times 3$ matrix with the variance / covariance for the three elastic parameters at this traveltime $j$ and trace location $i$, and ii) a $N_2 \times N_2$ correlation matrix for the lateral correlation between elastic parameters in different traces. The $3 \times 3$ covariance matrix $\mathbf{\Sigma}_{0,j}$ is defined as

$$\mathbf{\Sigma}_{0,j} = \begin{bmatrix} \tau^2_{v_p(j)} & \eta_{v_p(j),v_s(j)} & \eta_{v_p(j),\rho(j)} \\ \eta_{v_p(j),v_s(j)} & \tau^2_{v_s(j)} & \eta_{v_s(j),\rho(j)} \\ \eta_{v_p(j),\rho(j)} & \eta_{v_s(j),\rho(j)} & \tau^2_{\rho(j)} \end{bmatrix}, \tag{9}$$

where $\tau$'s are the variances, while $\eta$'s are the covariances between the elastic parameters. The lateral correlation part is modeled by the inverse covariance matrix, which is sparse under Markov assumptions. In our specific 2D case below, this implies that the conditional inversion result at trace $i$ is only dependent on the two neighboring traces $i-1$ and $i+1$, and within each trace only the elastic parameters at the same traveltimes are coupled. Such a choice of local correlation corresponds to the model of autoregressive time series (Brockwell and Davis, 1991). By construction, the inverse $N_2 \times N_2$ covariance matrix is defined by

$$\mathbf{\Sigma}^{-1}_{m,1} = \begin{bmatrix} 1 & -\phi & & & & & \\ -\phi & 1+\phi^2 & -\phi & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & -\phi & 1+\phi^2 & -\phi \\ & & & & & -\phi & 1 \end{bmatrix}. \tag{10}$$

Here, $-1 < \phi < 1$ is the interaction parameter, which in the 2D model is the autocorrelation along the line. The covariance matrix for traveltime $j$ is then defined by the Kronecker product $\mathbf{\Sigma}_{m,j} = \mathbf{\Sigma}_{m,1} \otimes \mathbf{\Sigma}_{0,j}$. All these parameters in the prior model can in principle vary both in time and laterally, if one can make justified guesses. However, in our examples below, we consider these parameters as invariant.

When including the different contribution to the laterally correlated prior model, the $l_{lat}$ term becomes

$$\begin{aligned} l_{lat}(\mathbf{m}) = const + \frac{1}{2} \sum_{i=1}^{N_1} (\mathbf{m}_i - \boldsymbol{\mu}_i)' \mathbf{\Sigma}_{m,i}^{-1} (\mathbf{m}_i - \boldsymbol{\mu}_i) \\ - \phi \sum_{i=1}^{N_1 - 1} (\mathbf{m}_i - \boldsymbol{\mu}_i)' \mathbf{\Sigma}_{m,i}^{-1} (\mathbf{m}_{i+1} - \boldsymbol{\mu}_{i+1}), \end{aligned} \tag{11}$$

where the covariance matrix $\mathbf{\Sigma}_{m,i}$ is block diagonal with the $3 \times 3$ matrices $\mathbf{\Sigma}_{0,j}$ along the diagonal blocks $i$. In the examples below, we will regard only 2D applications. However, some considerations on extending the algorithm to 3D applications are in order: in 3D applications the sparsity of $\mathbf{\Sigma}_{m,1}^{-1}$ is maintained, but now this matrix includes 9 non-zero elements, if only nearest and diagonal neighbors are used in a Markovian setting. Assembling the laterally correlated covariance matrix remains straight forward; however, as we will show in the next section, the solution needs to be found numerically. Then, the numerical handling of the inversion problem becomes more complicated, as discussed in the next section.

## METHODOLOGY

With the definitions of the terms occurring in Bayes' theorem (equation 1) in the previous section, the negative log-posterior is given as

$$\begin{aligned} l_{\mathrm{m|d}}(\mathbf{m}) = const + \Bigg\{ & \frac{1}{2}(\mathbf{d}-\mathbf{G}\mathbf{m})^T \boldsymbol{\Sigma}_d^{-1}(\mathbf{d}-\mathbf{G}\mathbf{m}) \\ & + \frac{1}{2}(\mathbf{m}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}_m^{-1}(\mathbf{m}-\boldsymbol{\mu}) \\ & + \sum_{j=1}^{N_1}\sum_{i=1}^{N_2}\sum_{l=1}^{3} C\left(\frac{[\mathbf{D}_i(\boldsymbol{m}_i-\boldsymbol{\mu}_i)]_{jl}}{\kappa_l^2}\right)\Bigg\}. \end{aligned} \tag{12}$$

The likelihood and lateral prior parts are Gaussian, but the blockyness constraint applied to the derivative along the vertical direction makes the problem non-Gaussian.

We are now in a position to determine the optimum blocky solution of the seismic inversion problem. This solution is found by differentiating the above negative log-likelihood $l_{\mathrm{m|d}}(\mathbf{m})$, where the argument is the model parameter $\mathbf{m}$. By introducing the blockyness term, the problem becomes non-linear and needs to be solved iteratively. Our optimization routine follows a procedure proposed by Charbonnier et al. (1997), where the last part of equation (12) is linearized. A least-squares problem is obtained and subsequently solved iteratively. The main idea can be summarized as follows: equation (12) is differentiated with respect to $\mathbf{m}$ and set to zero

$$\frac{dl_{\mathrm{m|d}}(\mathbf{m})}{d\mathbf{m}} = -\mathbf{G}^T\boldsymbol{\Sigma}_d^{-1}(\mathbf{d}-\mathbf{G}\mathbf{m}) + \boldsymbol{\Sigma}_m^{-1}(\mathbf{m}-\boldsymbol{\mu}) + \mathbf{D}^T\mathbf{B}\mathbf{D}(\mathbf{m}-\boldsymbol{\mu}) = \mathbf{0}. \tag{13}$$

Here, $\mathbf{B}$ is a diagonal $3N_1N_2 \times 3N_1N_2$ matrix with the diagonal elements corresponding to trace $i$, traveltime $j$ that are defined for each elastic parameter

$l$ as

$$B_{kk}(\mathbf{m}) = \begin{cases} \frac{1}{\kappa^2}, & \text{if Gauss norm} \\ \frac{2}{\kappa_l^2 + [\mathbf{D}_i(\mathbf{m}_i - \boldsymbol{\mu}_i)]_{jl}^2}, & \text{if Cauchy norm} \\ \frac{1}{\sqrt{\kappa_l^4 + \kappa_l^2 [\mathbf{D}_i(\mathbf{m}_i - \boldsymbol{\mu}_i)]_{jl}^2}}, & \text{if differentiable Laplace norm} \end{cases}, \quad (14)$$

where the index $k$ is related to the three basic indices $i, j, k$ by $k = (l-1)N_1N_2 + (i-1)N_1 + j$, and $[\,]_{jl}$ means that the $(l-1)N_2 + j^{th}$ element of the $i^{th}$ trace is used.

This expression is exact, using no approximation, but it cannot be solved analytically for $\mathbf{m}$, as the matrix $\mathbf{B}$ depends on $\mathbf{m}$. Instead, an iterative optimization scheme solves $dl_{\mathrm{m|d}}(\mathbf{m})/d\mathbf{m} = 0$ assuming that $\mathbf{B}$ is independent of $\mathbf{m}$ at each iteration. Charbonnier et al. (1997) showed that this assumption is justified because the dependency on $\mathbf{m}$ in $\mathbf{B}$ is weak. The solution of equation (13) is

$$\begin{aligned} \widehat{\mathbf{m}}_{\mathrm{MAP}} = &\left[\mathbf{G}^T\boldsymbol{\Sigma}_d^{-1}\mathbf{G} + \boldsymbol{\Sigma}_m^{-1} + \mathbf{D}^T\mathbf{B}\mathbf{D}\right]^{-1} \times \\ &\left[\mathbf{G}\boldsymbol{\Sigma}_d^{-1}\mathbf{d} + \boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu} + \mathbf{D}^T\mathbf{B}\mathbf{D}\boldsymbol{\mu}\right]. \end{aligned} \quad (15)$$

In order to obtain the final solution, an iteratively reweighed algorithm as suggested by Scales et al. (1988) applies (see also Figure 3): the current $\widehat{\mathbf{m}}_{\mathrm{MAP}}$ is used to compute a new $\mathbf{B}$, which is then kept fixed to obtain a new $\widehat{\mathbf{m}}_{\mathrm{MAP}}$ in the subsequent iteration step. At each iteration, we recompute the scaling matrix $\mathbf{B}$, and then solve a linear equation to obtain the new inversion result. When the inversion result does no longer change substantially in the iteration, convergence it achieved. Our tests showed that convergence is fast, and the algorithm typically reaches machine precision within $3-5$ steps.

The mathematical properties of the solution $\widehat{\mathbf{m}}_{\mathrm{MAP}}$ are discussed next. As noted by Charbonnier et al. (1997), the algorithm will converge to a minimum of $l_{\mathrm{m|d}}(\mathbf{m})$, corresponding to a maximum for the posterior density of $\mathbf{m}$ given $\mathbf{d}$. When the blockyness constraint is a convex function, this maximum is unique. This entails that a unique maximum for the posterior can be detected when using a differential Laplace as blockyness constraint. When the blockyness constraint is non-convex, the algorithm converges to a local minimum. The edges are possibly sharper in this case, but the solution is not global and may include some artifacts caused by the non-uniqueness. In the proof of this statements by Charbonnier et al. (1997), a full rank of the model matrix (in this case $\mathbf{G}$) is required, but in the formulation presented here, a global maximum is still reached since the Gaussian part of the prior imposes full rank. Recall that the Gaussian part of the prior tie the model parameters in the lateral direction, with independence in the vertical.

The laterally correlated blocky inversion methods requires a large sparse matrix to be inverted. For our examples, we chose to facilitate a Cholesky factorization procedure, which we describe in Appendix A.

## EXAMPLES

The methodology developed in the previous section aims at enhancing the contrasts between layers in seismic inversion. In this section, two examples are presented to illustrate the feasibility and potential of the blocky inversion method. In the first example, a synthetic representation of a blocky elastic

model is generated, and from this data are simulated. These data are then inverted with four different inversion constraints. First, we will determine a standard minimum norm solution, which results from equation (12) by omitting the constraint term containing the gradients. The other three inversions include the blockyness constraint for the three different functions in equation (8). This example is used for demonstration of the methodology by comparing the inversion results to a known data set. The second example deals with seismic AVA data from the Norwegian Sea. In this example, assessment of the parameters become more important, for which we will suggest a statistical procedure.

**Synthetic example**

We construct a model that consists of 25 traces and that span over $1000ms$ two way travel time. This model contains six major layers with boundaries at 200, 240, 400, 480, 680, and $720ms$ travel time, respectively. In addition, each layer contains further minor vertical and lateral variations of the P-wave and S-wave velocity and density, which are laterally and vertically correlated. The resulting synthetic model is displayed in Figure 4. Synthetic seismic AVA data are simulated conditional on this elastic model for four angles ($10^o$, $20^o$, $30^o$ and $40^o$). We used a linearized Aki and Richards forward model and a $30Hz$ Ricker wavelet to create synthetic AVA gathers for 10, 20, 30 and 40 degree for the angle of incidence in equation (2). These data were subsequently perturbed with independent random noise of variance $\sigma^2 = 0.01^2$.

For any inversion strategy, we need to specify model parameters in order to compute the inversion. In this synthetic case, we set the prior covariance in the Gaussian prior part from the synthetic elastic model, and the lateral correlation parameter to $\phi = 0.9$. The mean level $\boldsymbol{\mu}$ in the Gaussian part is constructed by lowpass filtering the synthetic elastic model for each of the three elastic parameters. Further, the blockyness parameter is estimated from the true synthetic vertical gradients, and becomes $\kappa = 0.012$ (Cauchy), $\kappa = 0.015$ (Laplace), and $\kappa = 0.03$ (Gaussian). These $\kappa$ are used for all three elastic parameters in each case.

First, consider blocky inversion with Laplace constraints for the vertical gradients. The pseudocode in Figure 3 is initiated with prior mean $\boldsymbol{\mu}$, and run for five iterations. The decline of the posterior objective function $l_{\mathrm{m|d}}(\mathbf{m})$ is monitored at each iteration. After the second iteration, the value of the objective function decreases to 0.05% of its original value, and eventually to $6 \times 10^{-7}$ after the fifth iteration. This type of decline indicates that an optimum a posteriori solution is reached, and that convergence occurs within a few iterations. The blocky inversion result obtained at the last iteration is displayed in Figure 5. This result should be compared with the synthetic elastic model in Figure 4. We note that the inversion result has the same trends as the synthetic model, with a blocky appearance in the vertical direction at the synthetic layer boundaries, in particular for the P-wave velocity.

The next example employed the Cauchy constraints for the vertical gradients. The pseudocode in Figure 3 is initiated and run as for Laplace described

above, and the blocky inversion result is displayed in Figure 6. These results based on Cauchy prior show similar trends as with Laplace prior, but the vertical jumps are larger and the extremes of the true model are better captured with the Cauchy prior. On the other hand, the Cauchy inversion results contain some artifacts, which appear as small fluctuations in Figure 6. These artifacts indicate that the inversion using Cauchy constraints for the vertical gradients converges only to a local optimum of the posterior. For the Laplace solution in Figure 5, we can also see some fluctuations, but less pronounced. These smaller artifacts in the Laplace solution may be attributed to numerical noise, while those for Cauchy are mathematical and most likely caused by the non-convex objective function.

In order to illustrate the improvement of the different blocky inversion results over a standard minimum norm inversion, we compare the inversion results for the P-wave velocity using various priors at the central trace. All inversion results seem to represent the true synthetic P-wave velocity reasonably well. As a measure for the proximity of the different inversion results to the true model, we calculate the sum of squared differences $\beta$ in vertical gradients between true ($\mathbf{m}$) and estimated P-wave velocity ($\widehat{\mathbf{m}}$) for each trace,

$$\beta = ||\mathbf{D}\widehat{\mathbf{m}}_i - \mathbf{D}\mathbf{m}_i||^2, \tag{16}$$

where a smaller value of $\beta$ indicates a better correspondence to the true model. For the P-wave velocity at the center trace, we calculate the following values: $\beta = 58.0$ for Cauchy, $\beta = 58.5$ for Laplace, $\beta = 60.2$ for blocky

Gaussian, and $\beta = 153.2$ for the minimum norm inversion. This means that the blocky models capture the true vertical gradients significantly better than the minimum norm Gaussian solution, which seems to contain considerably more side-lobe energy.

These imperfections of the minimum norm solution are in particular evident in the detail around $200ms$ in Figure 7(right). The three blocky models follow the velocities in the layers above and below the interface fairly well, whereas the minimum norm solution contains some considerable side-lobe energy. The change of the velocity at the interface in the solution for the blockyness constraint with the Gaussian function (blue) is an improvement over the minimum norm solution (magenta). This example also shows that both the Cauchy (red) and Laplace (green) functions provide even more constant values within the layers and also sharper layer boundaries.

As a last test for this synthetic example, we investigate the confidence that we can have in the automated detection of layer boundaries using the blocky inversion constraints. In order to do so, we apply a Markov chain Monte Carlo (MCMC) sampling algorithm to assess the uncertainty of the posterior solution for the model with Laplace constraints. The MCMC algorithm we use here is a Langevin–Metropolis method, whose proposal $m'$ is normally distributed with prior mean $\mathbf{m}+\frac{h^2}{2}$ and variance $h^2\mathbf{I}$ (Roberts and Rosenthal, 2001). Here, $h$ is a tuning parameter in the MCMC algorithm. The optimal value for $h$ is found by trial and error such that the acceptance rate in the MCMC algorithm is about 57% (Roberts and Rosenthal, 2001). To test the accuracy of the blocky inversion method, we run the MCMC algorithm for

100, 000 iterations. After each iteration, we determine the occurrence of the ten largest vertical gradients in the P-wave velocity for the central trace. We count the frequency of such occurrence at each sample, which gives an estimate of the probability of each travel time to contain one of the ten largest gradients. Figure 8 shows these estimates for the Laplace model for vertical gradients (left) and for the minimum norm model (right). When comparing the Laplace results and the minimum norm results, we see that the largest gradients are more clearly identified in the Laplace case. Consider for instance the layer boundary at $200ms$ travel time: the analysis for the inversion data using the Laplace constraint contains a clear spike indicating that in approximately 85% of the simulations, the P-wave velocity contains a large change at this travel time. On the contrary, the analysis for the minimum norm solution shows that this is only the case for roughly 15% of the samples. In general, the data for the minimum norm solution barely indicate a clear detection of a layer boundary, as its distribution is rather uniform over the entire trace. In comparison, the Laplace result detects the boundaries at 200, 240, 400, 480, 680, and $720ms$ travel time quite well. However, especially in the window between 480 and $600ms$, this analysis shows the occurrence of spurious large gradients in the simulated Laplace models, which may due to noise in the synthetic realizations of the model and seismic data.

**Field data from the Norwegian Sea**

This section contains a description of an inversion of 2D seismic data

from the Norwegian sea using the blocky inversion method. We will discuss three different inversions in this example. The first inversion is based on the minimum norm inversion, which we will use again as a reference to illustrate the impact of the blocky inversion on the vertical structure of the calculated models. The other two inversions are the blocky methodology with the Cauchy and Laplace constraints.

We chose this line because of a borehole in the vicinity that contains a full suite of sonic measurements that allows for developing a detailed prior model. This prior model consists of three contributions: a) a low frequency background model, which forms the a priori expectation of the model values $\boldsymbol{\mu}$; b) an assessment of the covariance parameters; and c) the distribution parameters $\kappa$. In order to determine the different parts of the prior model, the log data are first converted from depth to time and subsequently resampled to the seismic sampling interval. The low frequency background model was created by first copying the log data laterally, where interpreted horizons were used as guides, and then smoothed using a suitable $f-k$ filter. The prior covariance matrix results from statistical analyzing the log data for the variances and correlation coefficients; these properties are laterally invariant. The last parameter to be determined is the distribution parameter $\kappa$ for the two blocky inversions, which results from log data by using a maximum likelihood estimator. In order to do so, we first define a likelihood function $L$ given by

$$L(\mathbf{x};\kappa) = \prod_{t=1}^{N_2-1} C(x_t;\kappa), \tag{17}$$

where $C(\mathbf{x})$ is defined in equation (8), $x_t$ are the vertical gradients of the log data, and stating explicitly $\kappa$ in this expression is to highlight that $\kappa$ is a varying parameter in this estimation process. As before, we consider the log-likelihood function $l = \log(L)$ and find the optimum distribution parameter $\kappa^*$ by optimizing $l$ with respect to $\kappa$, which results in the following condition for $\kappa^*$

$$\frac{\partial l}{\partial \kappa} = -\frac{N_2}{\kappa} + \frac{2}{\kappa^3} \sum_{t=1}^{N_2} \frac{\partial C(x_t)}{\partial \kappa} = 0 \text{ for } \kappa = \kappa^*. \tag{18}$$

This optimum distribution parameter needs to be found numerically; table 1 summarizes the obtained values for the Cauchy and Laplace distribution parameters when analyzing the log data.

In order to obtain a suitable interaction parameter $\phi$ for the lateral correlation (equation 10), we performed tests on a small subset of the seismic section and selected an optimum parameter of $\phi = 0.4$ by visual inspection.

Furthermore, wavelets were extracted for each common angle stack at the borehole location and assumed to be invariant throughout the entire data set. Lastly, data noise was assessed by statistically analyzing the operator mismatch $\mathbf{d} - \mathbf{Gm}$ at the borehole location with the assumption that this assessment applies to the entire data set.

As in the synthetic example, the inversion followed the scheme of Figure 3. In this case, the algorithm converges after one update, as the cost function does not decrease further for additional iterations. Although the inversion calculates models for both seismic velocities and density, we will focus only on the inverted acoustic impedance. Figure 9 displays the inverted acous-

tic impedance model from the minimum norm inversion (top), the blocky inversion using a Cauchy constraint (middle) and Laplace constraint (bottom). Note that we display the inverted acoustic impedance variations with respect to the prior expectation to better illustrate the contribution of the inversion. All three inversions deliver similar models but we can indeed see minor differences. Careful inspection of these images reveals that the minimum norm inversion contains several side lobes and layer boundaries appear more blurry than in the other two models. This is highlighted in Figure 10, which compares a single trace from each inversion model; the location of this trace is indicated by the dotted line in Figure 9. While the results of the blocky inversion differ not significantly from each other, the inversion result from the minimum norm algorithm shows significantly less detail. This is in particular apparent for the event at $0.1s$, which appears less focussed for the minimum norm inversion than for the blocky inversions. In addition, there seems also to be more details present in the models stemming from the blocky inversions between $0.4s$ and $0.5s$ than for the minimum norm model.

## DISCUSSION

The two examples in the previous section show that the modified Bayesian inversion methodology is capable of delivering inverted models that exhibit improved layer contrasts over the minimum norm solution. The main motivation for the synthetic example was to analyze the potential and behavior of the Cauchy and Laplace functions as a constraint in seismic inversion. Both functions deliver a blocky structure in the inversion result, however, a major

difference became apparent. While the inversion result using the Laplace function as a constraint for the gradients of the models in Figure 5 are well behaved in the sense that the inverted models are more or less constant within the layers and exhibit only large changes at the layer boundaries, the results for the Cauchy constraint are different (see Figure 6). These models exhibit severe numerical artifacts that might be caused by the inversion algorithm converging to a local minimum. Remember that the Cauchy function is non-convex and the posterior distribution $p(\mathbf{m}|\mathbf{d})$ is multi-modal. In the local optimum, the small oscillations are the most likely, while a global optimum may not show these artifacts. However, although the Cauchy solution has some fluctuation artifacts it might capture the largest jumps more realistically than the solution using Laplace constraints. Like in the case with Laplace constraints, the side lobes typically shown in inversion methods suppressed in the inversion as a result of using the blockyness constraints.

The field example, on the other hand, did not exhibit these problems with the Cauchy function. Either inversion results shown in Figures 9 show reasonable data ranges for models of the P-wave velocity, and the detail in Figure 10 also shows that the blocky inversion can indeed deliver an improved layer contrast than the minimum norm inversion that we used as a reference case. The challenges we wanted to address in this example was the estimation of the relevant parameters that control the inversion, in particular the distribution coefficient $\kappa$ and the interaction parameter $\phi$. In order to determine the distribution parameter, we employed a statistical algorithm. The example showed that the distribution parameters in such a way are well suited to

enable a reasonable inversion with well developed layer boundaries. On the other hand, the determination of the interaction parameter was performed by trying different values for $\phi$ and choosing that one that gave visually the best looking inverted models. As this is often difficult to decide and may vary from interpreter to interpreter, the development of objective mathematical routines for a proper estimation of $\phi$ would be a useful extension of the presented work so far.

## CONCLUSION

In this paper, we discussed an inversion constraint that aims at enhancing the contrasts between layers in seismic prestack inversion. The major result of the work presented here is the introduction and feasibility testing of a blocky seismic inversion methodology. We introduced an additional term in the optimization term that utilizes Cauchy and Laplace distributions, which we applied to the gradients of the model parameters. The resulting inversion method enforces sparsity onto these gradients and thereby achieves a better focusing of layer boundaries. An attractive feature of this approach is that it is not necessary to a priori specify the location of layer boundaries in the seismic data. Instead, the algorithm analyzes a first inversion attempt and determines an optimum constraint that will enforce sharper contrasts between adjacent layers in a subsequent inversion. As such, the enhancement of layer boundaries is purely data driven. In addition, we also included a lateral correlation of neighboring traces in order to stabilize the inversion. When developing this term, we were particularly interested in maintaining

the possibility of laterally varying correlation, which could be an important feature when geological structures such as faults are present in the area of interest. This results in a very large sparse system of equations. The use of Cholesky factorization as a computational tool for inverting a large sparse matrix was very appropriate for this purpose. It was relatively fast, and the size of the matrices to invert was $3N_2 \times 3N_2$ instead of $3N_1N_2 \times 3N_1N_2$ as it would have been for the direct solution. For 3D data sets, Cholesky factorization may not be feasible any longer and other numerical methods such as conjugate gradients may be a better alternative.

By introducing this additional constraint, the optimum solution becomes non-linear and can therefore not be calculated directly. Instead, iterative solution algorithms need to be applied, which increases the computational costs of the inversion by at least a factor of two. In our experience, performing one additional iteration with the blockyness constraint $\mathbf{B}$ is sufficient for realistic data in order to obtain the optimum solution, as further iteration steps do not improve the results significantly.

The synthetic example suggests that the Laplace function is a more suitable for enforcing a blocky inversion result because of its convex form. Applying the Cauchy norm to the same data resulted in inversion results that contain substantial artifacts that can possibly be explained by the inversion not finding the optimum global minimum but converging to a local minimum. However, when analyzing the inversion results from the Norwegian Sea, such problems did not occur, indicating that this problem is possibly not as severe in real data.

Determining the necessary parameters for sparse inversion schemes or high resolution processing algorithms has always been a challenge. We addressed this issue for the blocky inversion scheme by developing a statistical analysis workflow using log data to find the optimum distribution parameters $\kappa$. At this stage, however, a careful visual inspection and supervision of the dependence of the inversion on the select distribution parameters $\kappa$ is recommended. Also, the presented algorithm can only determine one distribution parameter for the entire time interval of interest. The estimation of a time-variant distribution parameter $\kappa$ is the topic of ongoing investigation. Another short-coming of the current algorithm is that it determines the gradient and statistical parameters, $\kappa$ and $\mathbf{\Sigma}_0$, separately. Since these parameters are most likely linked to each other, alternative methods for parameter tuning, such as cross validation, might be of interest.


## ACKNOWLEDGEMENTS

We like to thank StatoilHydro ASA for the permission to publish this paper.


## Appendix A
## Algorithm considerations

The blocky inversion algorithm summarized in equation (15) must be solved for $\widehat{\mathbf{m}}_{\mathrm{MAP}}$. This requires an inversion of a large sparse matrix. This section describes the Cholesky factorization of the large sparse matrix that we employed in order to express the matrix to be inverted as $[\mathbf{G}^T\mathbf{\Sigma}_d^{-1}\mathbf{G} + \mathbf{\Sigma}_m^{-1} +$

$\mathbf{D}^T\mathbf{B}\mathbf{D}]^{-1} = \mathbf{R} = \mathbf{L}\mathbf{L}'$ (see, e.g., Davis and Hager, 2005; Rue and Held, 2005). Using Cholesky factorization, the problem is given by

$$\mathbf{L}\mathbf{L}'\widehat{\mathbf{m}}_{\text{MAP}} = \mathbf{a}, \tag{A-1}$$

where $\mathbf{a} = \mathbf{G}\boldsymbol{\Sigma}_d^{-1}\mathbf{d} + \boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu} + \mathbf{D}^T\mathbf{B}\mathbf{D}\boldsymbol{\mu}$. The solution to equation (A-1) is given in two steps:

- Solve 'forward' for $\widehat{\mathbf{m}}$ in $\mathbf{L}\widehat{\mathbf{m}} = \mathbf{a}$.
- Solve 'backward' for $\widehat{\mathbf{m}}_{\text{MAP}}$ in $\mathbf{L}'\widehat{\mathbf{m}}_{\text{MAP}} = \widehat{\mathbf{m}}$.

The forward and backward steps are both easy to compute since $\mathbf{L}$ inherits the sparse structure of $\boldsymbol{R}$ and hence $\boldsymbol{\Sigma}_m^{-1}$. The only non-zero elements of $\mathbf{L}$ are on the $3N_2 \times 3N_2$ block diagonal and on the $3N_2 \times 3N_2$ block next to the diagonal, caused by the lateral correlation in $\boldsymbol{\Sigma}_{m,j}$. From the model formulation, the sparse matrix $\mathbf{R} = \mathbf{L}\mathbf{L}'$ equals

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_{1,2} & & & & & & \\ \mathbf{R}_{1,2} & \mathbf{R}_2 & \mathbf{R}_{2,3} & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & & \mathbf{R}_{N_1-1,N_1-2} & \mathbf{R}_{N_1-1} & \mathbf{R}_{N_1-1,N_1} \\ & & & & & \mathbf{R}_{N_1,N_1-1} & \mathbf{R}_{N_1} \end{bmatrix}, \tag{A-2}$$

where $\mathbf{R}_i = \mathbf{G}_i^T\boldsymbol{\Sigma}_{d,i}^{-1}\mathbf{G}_i + \boldsymbol{\Sigma}_{m,i}^{-1} + \mathbf{D}_i^T\mathbf{B}_i\mathbf{D}_i$, and $\mathbf{R}_{i,i+1} = \mathbf{R}_{i+1,i} = -\phi\boldsymbol{\Sigma}_{m,i}^{-1}$. The equations for $\mathbf{L}$, which inherits exactly the same structure as $\mathbf{R}$, are then given by

$$\mathbf{L}_1\mathbf{L}_1' = \mathbf{R}_1, \quad \mathbf{L}_i\mathbf{L}_i' + \mathbf{L}_{i,i-1}\mathbf{L}_{i,i-1}' = \mathbf{R}_i \tag{A-3}$$

for the diagonal terms, while the one off diagonal is

$$\mathbf{L}_{i-1}\mathbf{L}'_{i,i-1} = \mathbf{R}_{i-1,i}. \tag{A-4}$$

The order of computations to solve equation (A-1) is $O((3N_2)^3 2N_1)$, where the order $N_2^3$ is caused by matrix inversion for each trace, and the linear order $2N_1$ is caused by looping through the Cholesky factorization forward and backward through all traces in the data set. In 3D, the factor $2N_1$ would increase to $bN_1$, where $b$ is the bandwidth given by the ordering of cells and the Markov neighborhood on a two dimensional lattice of inlines / crosslines. There exists fast numerical schemes for minimizing this bandwidth and thus the computation time (Davis and Hager, 2005; Rue and Held, 2005).

# References


Aki, K., and P. Richards, 2002, Quantitative seismology, 2 ed.: University Science Books.

Amundsen, L., 1991, Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion: Geophysics, **56**, 2027–2035.

Brockwell, P., and R. Davis, 1991, Time series: theory and methods: Springer New York.

Buland, A., and H. Omre, 2003, Bayesian linearized avo inversion: Geophysics, **68**, 185–198.

Charbonnier, P., L. Blanc-Feraud, G. Aubert, and B. M., 1997, Deterministic

Edge-Preserving Regularization in Computed Imaging: IEEE Transactions on image processing, **6**, 298–311.

Davis, T. A., and W. W. Hager, 2005, Row modifications of a sparse cholesky factorization: SIAM Journal on Matrix Analysis and Applications, **26**, 621–639.

Farquharson, C. G., and D. W. Oldenburg, 1998, Non-linear inversion using general measures of data misfit and model structure: Geophys. J. Int., **134**, 213–227.

Geman, D., and C. Young, 1995, Nonlinear image recovery with half-quadratic regularization: IEEE Trans. Image Processing, **IP-4**, 932–946.

Jackson, D., 1979, The use of a priori data to resolve non-uniqueness in linear inversion: Geophys. J. Int., **57**, 137–157.

Oldenburg, D., T. Scheuer, and S. Levy, 1983, Recovery of the acoustic impedance from reflection seismograms: Geophysics, **48**, 1318–1337.

Portniaguine, O., and M. S. Zhdanov, 1999, Focusing geophysical inversion images: Geophysics, **64**, 874–887.

Roberts, G., and J. Rosenthal, 2001, Optimal scaling for various metropolis-hestings algorithms: Statistical science, **16**, 351–367.

Rue, H., and L. Held, 2005, Gaussian markov random fields, theory and applications, 1 ed.: Chapman & Hall/CRC.

Sacchi, M., and T. Ulrych, 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, 1169–1177.

Scales, J., A. Gersztenkorn, and S. Treitel, 1988, Fast $\ell_p$ solution of large, sparse, linear systems: Application to seismic travel time tomography: J.

Comput. Phys., **75**, 314–333.

Sheriff, R., 2002, Encyclopedic dictionary of applied geophysics, 4 ed.: Society of Exploration Geophysicists.

Youzwishen, C. F., and M. D. Sacchi, 2006, Edge Preserving Imaging: Journal of Seismic Exploration, **15**, 45–58.

## TABLES

Table 1: Optimum distribution parameters for blocky inversion.

| | $\kappa^*_{\text{Cauchy}}$ | $\kappa^*_{\text{Laplace}}$ |
|---|---|---|
| $\ln(v_P)$ | 0.0099 | 0.0166 |
| $\ln(v_S)$ | 0.0159 | 0.0298 |
| $\ln(\rho)$ | 0.0040 | 0.0089 |

## FIGURES

- Figure 1: Schematic sketch of the blocky inversion idea. Solid line: idealized boundary between two layers; dotted line: inversion result exhibiting side lobes and unfocussed layer boundary; dashed line: blocky inversion with reduced side lobes and sharp layer contrasts.

- Figure 2: Left) Potential function $C(x)$ for Cauchy (solid), differentiable Laplace (dashed) and Gaussian (dotted). Right) Normalized density functions for Cauchy (solid), differentiable Laplace(dashed) and Gaussian (dotted).

- Figure 3: Illustration of the algorithmic workflow of the iterative blocky inversion scheme.

- Figure 4: Properties of the synthetic example. Left: P-wave velocity, center: S-wave velocity; right: density.

- Figure 5: Inversion result with Laplace constraints on vertical gradients for the synthetic example. Left: P-wave velocity; center: S-wave velocity; right: density.

- Figure 6: Inversion result with Cauchy constraints on vertical gradients. Left: P-wave velocity; center: S-wave velocity; right: density.

- Figure 7: Left: comparison of the true synthetic P-wave velocity with the minimum norm and blocky inversion results at the center trace.

Right: comparison of the of the different inversion results at the center trace around the layer boundary at $0.2s$ (black line: true model, magenta: minimum norm solution, blue: Gaussian norm, red: Cauchy norm, green: Laplace norm).

- Figure 8: Proportions of realizations among the ten largest gradients. Left) Laplace blocky model. Right) Minimum norm Gaussian model. The dotted lines indicate the locations of the six major layer boundaries in the model.

- Figure 9: Comparison of the inverted acoustic impedance from a minimum norm inversion (a), a blocky inversion using the Cauchy constraint (b) and a Laplace constraint (c). Note that we show the inverted velocity variations. The dotted line indicate the position of the trace used for the comparison shown in Figure 10. The data are shown with respect to an arbitrary reference point.

- Figure 10: Comparison of the inverted acoustic impedance from a minimum norm inversion (left), a blocky inversion using the Cauchy constraint (center) and a Laplace constraint (right).

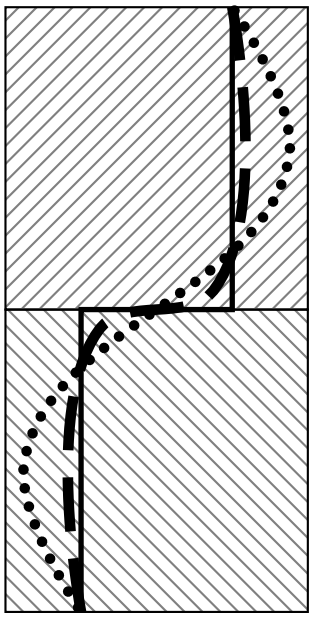

Figure 1:

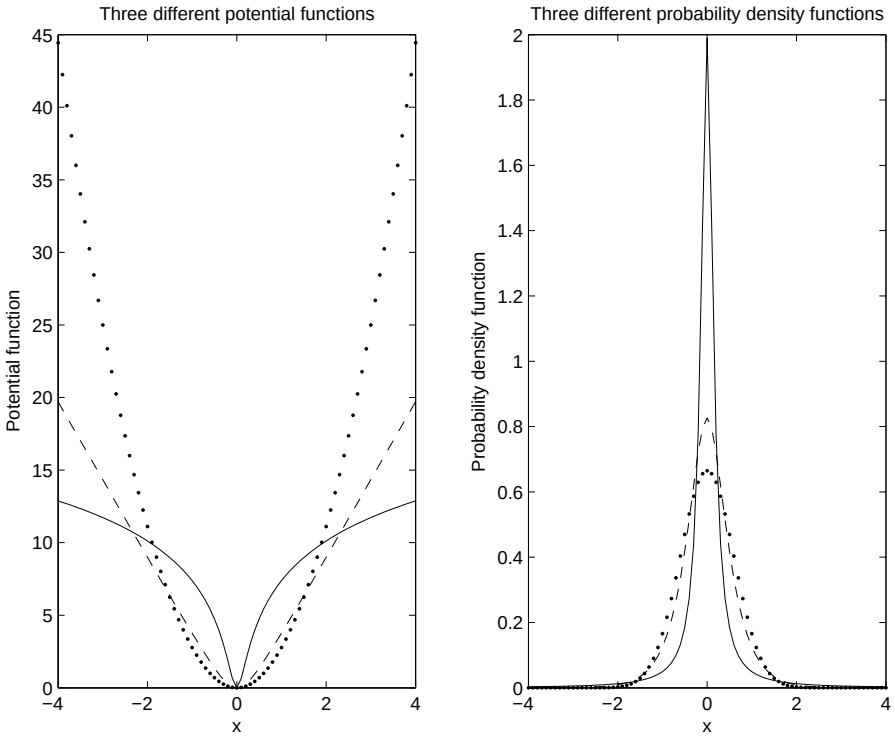
Three different potential functions
Potential function
x
Three different probability density functions
Probability density function
x

Figure 2:

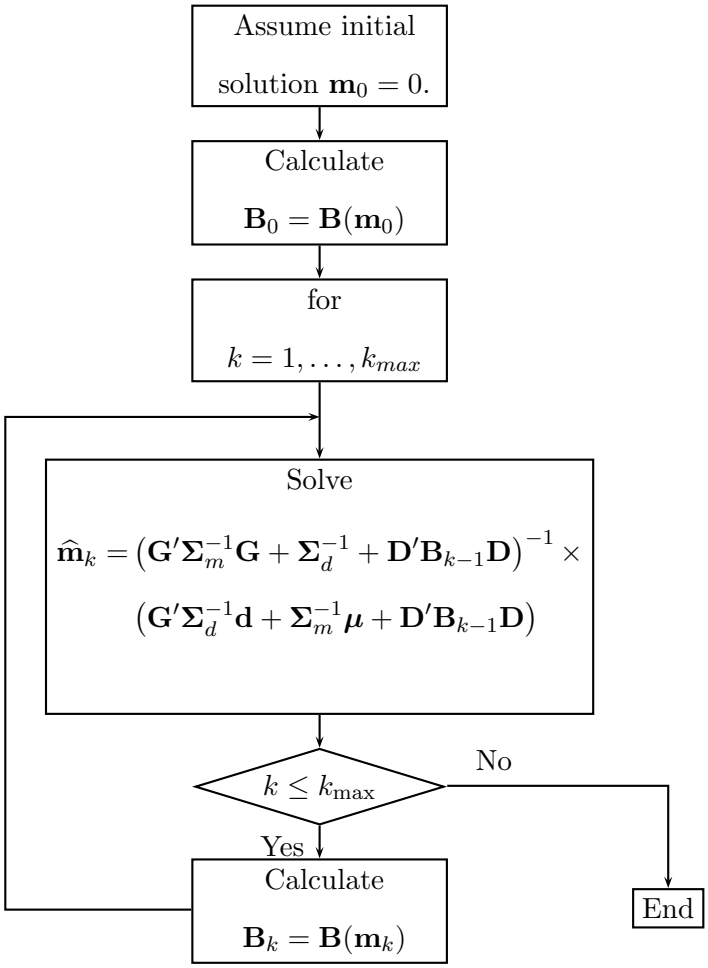
Assume initial
solution $\mathbf{m}_0 = 0$.
Calculate
$\mathbf{B}_0 = \mathbf{B}(\mathbf{m}_0)$
for
$k = 1, \ldots, k_{max}$
Solve
$\widehat{\mathbf{m}}_k = \left(\mathbf{G}'\boldsymbol{\Sigma}_m^{-1}\mathbf{G} + \boldsymbol{\Sigma}_d^{-1} + \mathbf{D}'\mathbf{B}_{k-1}\mathbf{D}\right)^{-1} \times$
$\left(\mathbf{G}'\boldsymbol{\Sigma}_d^{-1}\mathbf{d} + \boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu} + \mathbf{D}'\mathbf{B}_{k-1}\mathbf{D}\right)$
$k \leq k_{\max}$
No
Yes
Calculate
$\mathbf{B}_k = \mathbf{B}(\mathbf{m}_k)$
End

Figure 3:

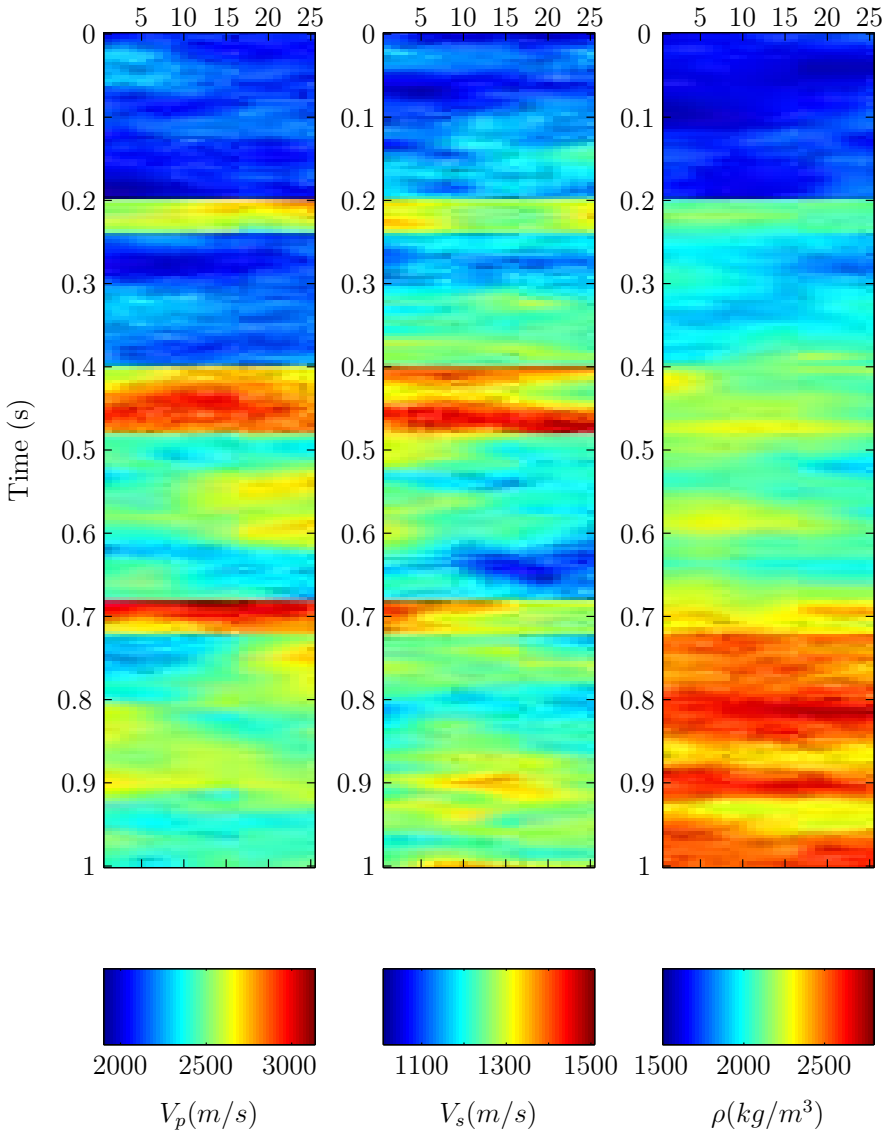
5 10 15 20 25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Time (s)
2000 2500 3000
$V_p(m/s)$
1100 1300 1500
$V_s(m/s)$
1500 2000 2500
$\rho(kg/m^3)$

Figure 4:

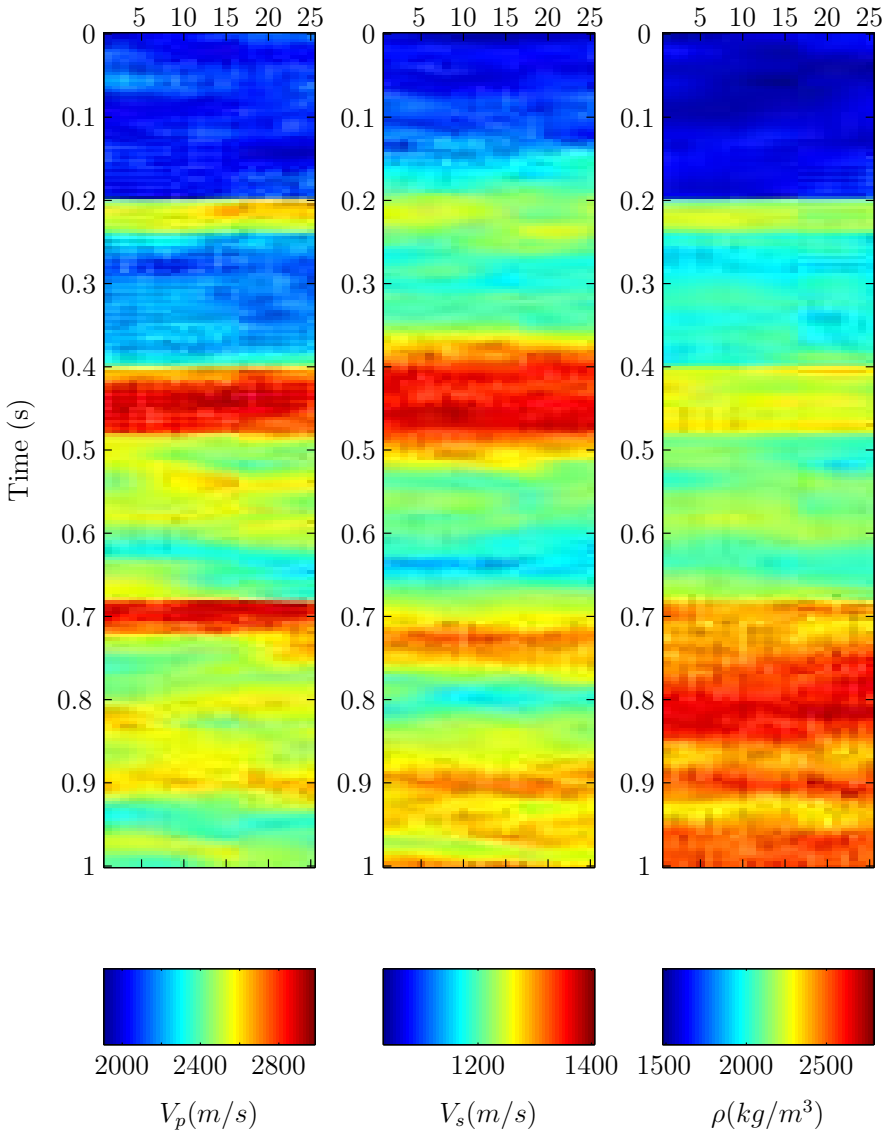
5 10 15 20 25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Time (s)
2000 2400 2800
$V_p(m/s)$
1200 1400
$V_s(m/s)$
1500 2000 2500
$\rho(kg/m^3)$

Figure 5:

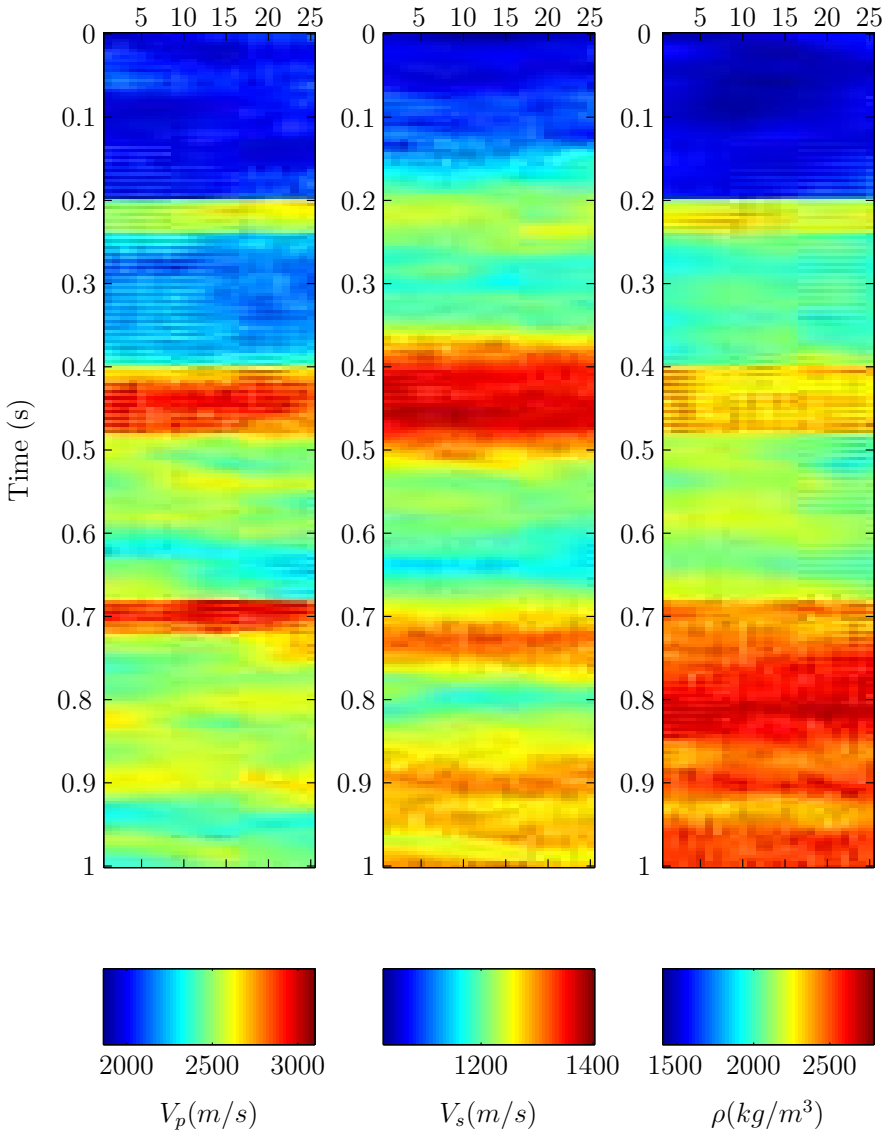
5 10 15 20 25
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Time (s)
2000 2500 3000
$V_p(m/s)$
1200 1400
$V_s(m/s)$
1500 2000 2500
$\rho(kg/m^3)$

Figure 6:

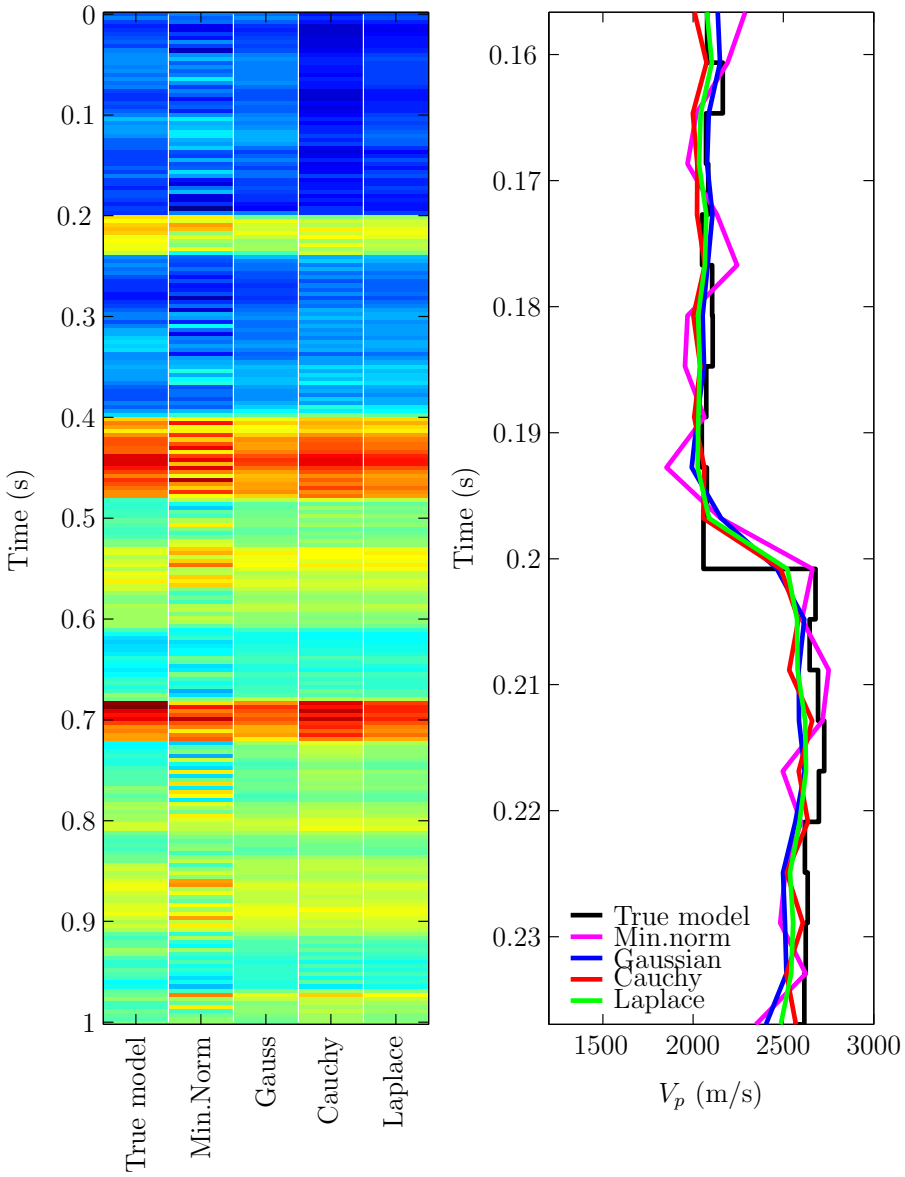
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Time (s)
True model
Min.Norm
Gauss
Cauchy
Laplace
0.16
0.17
0.18
0.19
0.2
0.21
0.22
0.23
Time (s)
True model
Min.norm
Gaussian
Cauchy
Laplace
1500
2000
2500
3000
$V_p$ (m/s)

Figure 7:

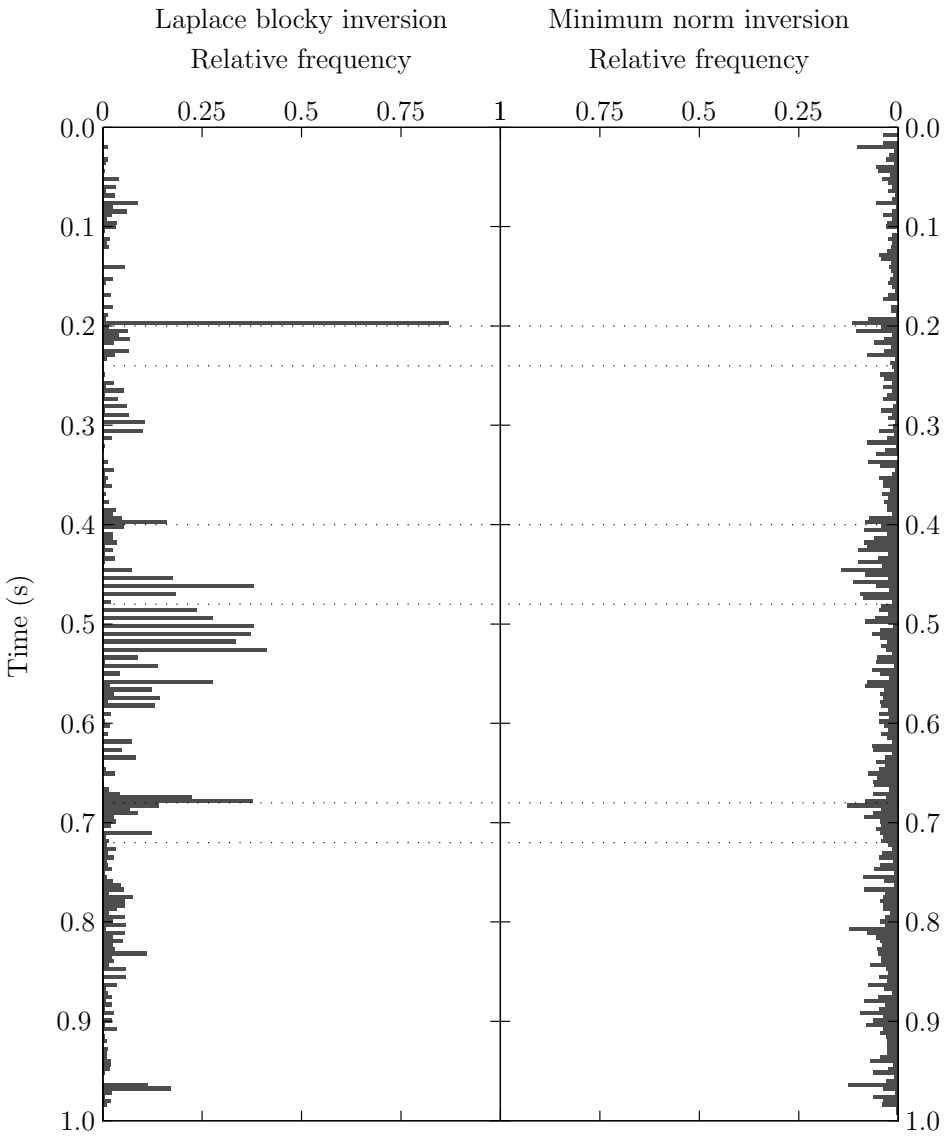
Laplace blocky inversion
Relative frequency
Minimum norm inversion
Relative frequency
Time (s)
0
0.25
0.5
0.75
1
0.75
0.5
0.25
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Figure 8:

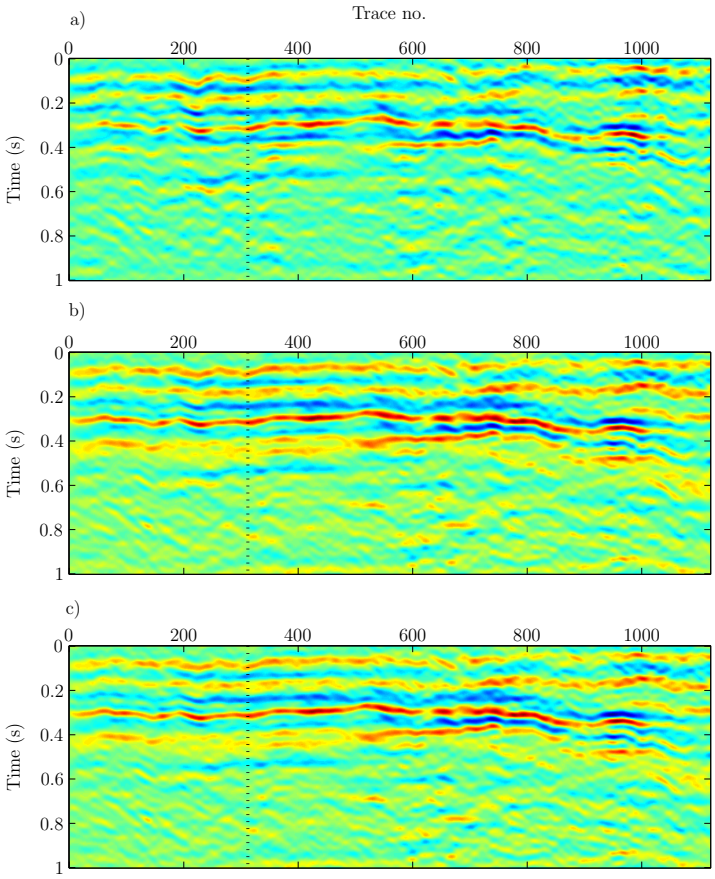
a)
Trace no.
0
200
400
600
800
1000
0
0.2
0.4
0.6
0.8
1
Time (s)
b)
0
200
400
600
800
1000
0
0.2
0.4
0.6
0.8
1
Time (s)
c)
0
200
400
600
800
1000
0
0.2
0.4
0.6
0.8
1
Time (s)

Figure 9:

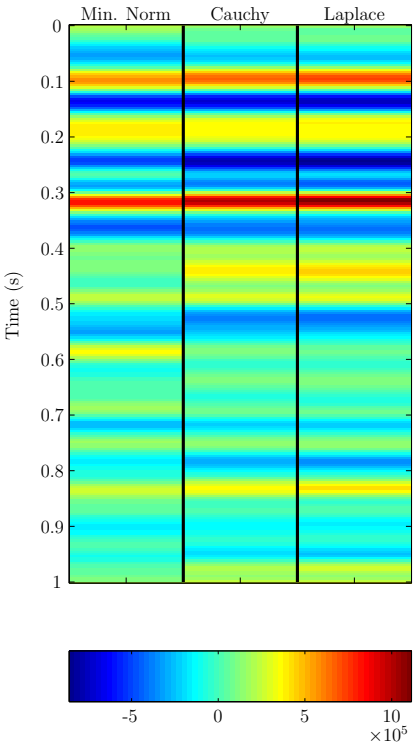
Min. Norm
Cauchy
Laplace
Time (s)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
-5
0
5
10
×10^5

Figure 10: