Total-variation minimization with bound constraints

Musa Maharramov and Stewart A. Levin

ABSTRACT

We present a powerful and easy-to-implement algorithm for solving constrained optimization problems that involve L_1/total-variation regularization terms, and both equality and inequality constraints. We discuss the relationship of our method to earlier works of Goldstein and Osher (2009) and Chartrand and Wohlberg (2010), and demonstrate that our approach is a combination of the augmented Lagrangian method with splitting and model projection. We test the method on a geomechanical problem and invert highly compartmentalized pressure change from noisy surface uplift observations. We conclude the paper with a discussion of possible extension to a wide class of regularized optimization problems with bound and equality constraints.

INTRODUCTION

The primary focus of this work is a class of least-squares fitting problems with a total-variation (TV) regularization and bound model constraints:

$$\begin{aligned} &\||\nabla \mathbf{m}|\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 \;\rightarrow\; \min, \\ &\mathbf{m}_1 \;\leq\; \mathbf{m} \;\leq\; \mathbf{m}_2. \end{aligned} \tag{1}$$

In (1) we seek a model vector $\mathbf{m}$ such that forward-modeled data $\mathbf{F}(\mathbf{m})$ match observed data $\mathbf{d}$ in the least squares sense, while imposing blockiness-promoting total-variation (TV) regularization (Rudin et al., 1992) and lower ($\mathbf{m}_1$) and upper ($\mathbf{m}_2$) model bounds. Rather than using a regularization parameter as a coefficient of the regularization term, we use a data-fitting weight α. TV regularization (also know as the Rudin-Osher-Fatemi, or ROF, model) acts as a form of model styling that helps to preserve sharp contrasts and boundaries in the model, even when spectral content of input data has limited resolution. Examples of successful geophysical application of unconstrained TV-regularized optimization are included in this report (Maharramov and Biondi, 2015; Maharramov et al., 2015; Ma et al., 2015a,b). The regularization provided by bounded total-variation sometimes produces sufficient smoothing side-effect on the inverted model that obviates explicit bound constraints. However, many applications still require the imposition of additional constraints regardless of the regularization. For example, reservoir pore-pressure inversion problems often come

with *a priori* bounds on the estimated pore pressure change, such as the pore pressure change being non-negative as a result of fluid injection (lower bound) or never exceeding a hydraulic fracturing pressure (upper bound). TV regularization is a key tool in imaging and de-noising applications (Rudin et al., 1992; Chambolle and Lions, 1997; Goldstein and Osher, 2009; Chartrand and Wohlberg, 2010) and require an efficient mechanism for including *a priori* model constraints that can significantly reduce model space (Chartrand and Wohlberg, 2010). While barrier or penalty function methods, such as nonlinear interior-point methods (Nocedal and Wright, 2006), can be used to tackle the general constrained formulation (1), the presence of a non-differentiable L_1-norm total-variation term and non-quadratic penalty terms pose considerable challenges to practical implementation. A log-barrier function such as

$$-\text{const} \times \sum_{i=1}^{n} \log \frac{m_2^i - m^i}{\delta} + \log \frac{m^i - m_1^i}{\delta}, \tag{2}$$

where n is the model space dimension, can be added to the right-hand side of the objective function to keep solution iterates away from the rectangular bounds. However, this adds a non-quadratic term to the objective function. For large-scale inversion problems with $n > 10^5$ (such as typical in geophysical applications) often only iterative gradient-based solution techniques like the nonlinear conjugate gradients (Nocedal and Wright, 2006) are available, and adding non-quadratic terms may significantly affect convergence properties. Note that this is in addition to the challenges associated with handling the non-differentiable TV-regularization term.

Chartrand and Wohlberg (2010) used a splitting approach to decouple the TV-regularized problem from enforcing the constraints. In our approach, we perform three-way splitting of problem (1) into a smooth optimization, gradient thresholding and projection steps using the Alternating Direction Method of Multipliers (ADMM) (Boyd et al., 2010). For unconstrained TV-regularized problems this approach is equivalent to the split-Bregman method of Goldstein and Osher (2009). However, we integrate the projection step associated with enforcing the bound constraints into the TV-minimization loop and avoid unnecessary iterations in the minimization of a proximal term (Parikh and Boyd, 2013) associated with the projection.

METHOD

First, we recast the TV-regularization part of (1) as a constrained optimization problem following the approach of Goldstein and Osher (2009). We introduce an auxiliary variable $\mathbf{x}$ and operator $\mathbf{\Phi} : \mathbf{m} \rightarrow \mathbf{x}$ such that for isotropic TV regularization we have a vector of the model-space dimension

$$\mathbf{\Phi}(\mathbf{m}) \;=\; \sqrt{(\nabla_x \mathbf{m})^2 + (\nabla_y \mathbf{m})^2}, \tag{3}$$

and for anisotropic regularization a vector twice the model-space dimension

$$\mathbf{\Phi}(\mathbf{m}) \;=\; \begin{bmatrix} \nabla_x \mathbf{m} \\ \nabla_y \mathbf{m} \end{bmatrix}. \tag{4}$$

Problem (1) can now be reformulated with an additional equality constraint:

$$\begin{aligned} &\|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 \;\rightarrow\; \min, \\ &\mathbf{x} \;=\; \mathbf{\Phi}(\mathbf{m}), \\ &\mathbf{m}_1 \;\leq\; \mathbf{m} \;\leq\; \mathbf{m}_2. \end{aligned} \tag{5}$$

Problem (5) is still a bound-constrained problem. Introducing the projection operator

$$\mathbf{\Pi}(\mathbf{m}) \;=\; \max\{\min\{\mathbf{m},\mathbf{m}_2\},\mathbf{m}_1\}, \tag{6}$$

where min and max are applied component-wise, we reduce (5) to a fully equality-constrained formulation:

$$\begin{aligned} &\|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 \;\rightarrow\; \min, \\ &\mathbf{x} \;=\; \mathbf{\Phi}(\mathbf{m}), \\ &\mathbf{m} \;=\; \mathbf{y}, \\ &\mathbf{y} = \mathbf{\Pi}(\mathbf{m}). \end{aligned} \tag{7}$$

Following the augmented Lagrangian recipe for (7) while assuming the last equality constraint still enforced explicitly, we obtain a sequence of problems (Nocedal and Wright, 2006)

$$\begin{aligned} &(\mathbf{x}^{k+1},\mathbf{m}^{k+1}) \;=\; \operatorname{argmin}\; \|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 \;+ \\ &\frac{\lambda}{2}\|\mathbf{x} \;-\; \mathbf{\Phi}(\mathbf{m})\|_2^2 - \boldsymbol{\mu}_k{}^T\left(\mathbf{x} \;-\; \mathbf{\Phi}(\mathbf{m})\right) + \frac{\delta}{2}\|\mathbf{m}-\mathbf{y}\|_2^2 - \boldsymbol{\nu}_k{}^T\left(\mathbf{m}-\mathbf{y}\right) \;\rightarrow\; \min, \\ &\boldsymbol{\mu}_{k+1} \;=\; \boldsymbol{\mu}_k - \lambda\left[\mathbf{x}^{k+1} - \mathbf{\Phi}(\mathbf{m}^{k+1})\right], \\ &\boldsymbol{\nu}_{k+1} \;=\; \boldsymbol{\nu}_k - \delta\left[\mathbf{m}^{k+1} - \mathbf{y}\right],\; k=0,1,2,\ldots \end{aligned} \tag{8}$$

Coefficients λ and δ are any positive constants above certain problem-specific "threshold" values (Nocedal and Wright, 2006), and can be selected experimentally. Vectors $\boldsymbol{\mu}_k$ and $\boldsymbol{\nu}_k$ are vectors of multipliers that converge to the set of Lagrange multipliers for the first two equality constraints of problem (7). At each step, (8) solves an L_1-regularized problem with respect to the combined model vector $(\mathbf{x},\mathbf{m})$. Introducing new scaled multiplier vectors

$$\mathbf{b}^k \;=\; \frac{\boldsymbol{\mu}_k}{\lambda},\; \mathbf{c}^k \;=\; \frac{\boldsymbol{\nu}_k}{\delta},\; k=0,1,2,\ldots \tag{9}$$

a little algebra shows that (8) is equivalent to

$$\begin{aligned} &(\mathbf{x}^{k+1},\mathbf{m}^{k+1}) \;=\; \operatorname{argmin}\; \|\mathbf{x}\|_1 \;+\; \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 \;+ \\ &\frac{\lambda}{2}\|\mathbf{x} \;-\; \mathbf{\Phi}(\mathbf{m}) - \mathbf{b}^k\|_2^2 + \frac{\delta}{2}\|\mathbf{m} \;-\; \mathbf{y} - \mathbf{c}^k\|_2^2 \;\rightarrow\; \min, \\ &\mathbf{b}^{k+1} \;=\; \mathbf{b}^k + \Phi(\mathbf{m}^{k+1}) - \mathbf{x}^{k+1}, \\ &\mathbf{c}^{k+1} \;=\; \mathbf{c}^k + \mathbf{y} - \mathbf{m}^{k+1},\; k=0,1,2,\ldots \end{aligned} \tag{10}$$

Here we used the fact that adding a constant term $\lambda/2\|\mathbf{b}^k\|_2^2 + \delta/2\|\mathbf{c}^k\|_2^2$ to the objective function obviously does not change the minimizing solution.

Problem (7) can be solved by iteratively projecting the current model vector $\mathbf{m}$ onto $\mathbf{y}$, then conducting the iterations (10) to convergence, then repeating the process. However, presence of the proximal term $\delta/2\|\mathbf{m}-\mathbf{y}\|_2^2$ in (8) due to the constraint $\mathbf{m}=\mathbf{y}$ means that a very accurate solution of (10) at early iterations is wasteful and unnecessary. We instead carry out a single iteration of (10) followed by the model projection:

$$\begin{aligned}
(\mathbf{x}^{k+1},\mathbf{m}^{k+1}) &= \text{argmin } \|\mathbf{x}\|_1 + \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 + \\
&\frac{\lambda}{2}\|\mathbf{x} - \mathbf{\Phi}(\mathbf{m}) - \mathbf{b}^k\|_2^2 + \frac{\delta}{2}\|\mathbf{m} - \mathbf{y}^k - \mathbf{c}^k\|_2^2 \rightarrow \text{min}, \\
&\mathbf{b}^{k+1} = \mathbf{b}^k + \Phi(\mathbf{m}^{k+1}) - \mathbf{x}^{k+1}, \\
&\mathbf{c}^{k+1} = \mathbf{c}^k + \mathbf{y}^k - \mathbf{m}^{k+1}, \\
&\mathbf{y}^{k+1} = \mathbf{\Pi}(\mathbf{m}^{k+1}) = \max\{\min\{\mathbf{m}^{k+1},\mathbf{m}_2\},\mathbf{m}_1\},\ k=0,1,2,\ldots
\end{aligned} \tag{11}$$

The iterative process (11) still requires soling an L_1-regularized problem. However, the L_1-norm term now involves only the vector $\mathbf{x}$. Therefore, we can apply Douglas-Rachford splitting, minimizing

$$\|\mathbf{x}\|_1 + \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_2^2 + \frac{\lambda}{2}\|\mathbf{x} - \mathbf{\Phi}(\mathbf{m}) - \mathbf{b}^k\|_2^2 + \frac{\delta}{2}\|\mathbf{m} - \mathbf{y}^k - \mathbf{c}^k\|_2^2 \tag{12}$$

alternately with respect to $\mathbf{m}$ and $\mathbf{x}$ in an inner loop of $N_1 \geq 1$ cycles. Because the proximal constraint $\mathbf{m}=\mathbf{y}$ renders good fitting accuracy at early stages unnecessary, N_1 can be small. Further we note that the minimization of (12) with respect to $\mathbf{x}$ (in a splitting step with $\mathbf{m}$ fixed) is given trivially by the "shrinkage" operator (Goldstein and Osher, 2009):

$$\mathbf{x}^{k+1} = \text{shrink}\left\{\mathbf{\Phi}(\mathbf{m}) + \mathbf{b}^k, \frac{1}{\lambda}\right\}, \tag{13}$$

where

$$\text{shrink}\left\{\mathbf{x},\gamma\right\} = \frac{\mathbf{x}}{|\mathbf{x}|}\max\left(|\mathbf{x}|-\gamma,0\right), \tag{14}$$

and is effectively thresholding the model gradient. Our algorithm can be described by the following 5 steps:

1 Initialization

$$\begin{aligned}
\mathbf{m}^0 &= \text{starting guess}, \\
\mathbf{x}^0 &= \mathbf{0}, \\
\mathbf{y}^0 &= \max\{\min\{\mathbf{m}^0,\mathbf{m}_2\},\mathbf{m}_1\}, \\
\mathbf{b}^0 &= \mathbf{0}, \\
\boldsymbol{c}^0 &= \mathbf{0},
\end{aligned} \tag{15}$$

2 **Outer loop**. Repeat steps 3-5 for $k=0,1,2,\ldots$

3 **Inner loop**. Iterate (16) $N_1 \geq 1$ times.

$$\begin{aligned} \mathbf{m}^{k+1} &= \text{argmin } \frac{\lambda}{2}\|\mathbf{x}^k - \mathbf{\Phi}(\mathbf{m}) - \mathbf{b}^k\|_2^2 + \frac{\alpha}{2}\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_2^2 + \\ &\quad \frac{\delta}{2}\|\mathbf{m} - \mathbf{y}^k - \boldsymbol{c}^k\|_2^2, \\ \mathbf{x}^{k+1} &= \text{shrink}\left\{\mathbf{\Phi}(\mathbf{m}^{k+1}) + \mathbf{b}^k, \frac{1}{\lambda}\right\}, \ \mathbf{x}^k \ = \ \mathbf{x}^{k+1}, \end{aligned} \tag{16}$$

4 Update the multipliers and project the model onto the bounding rectangle:

$$\begin{aligned} \mathbf{b}^{k+1} &= \mathbf{b}^k + \mathbf{\Phi}(\mathbf{m}^{k+1}) - \mathbf{x}^{k+1}, \\ \boldsymbol{c}^{k+1} &= \boldsymbol{c}^k + \mathbf{y}^k - \mathbf{m}^{k+1}, \\ \mathbf{y}^{k+1} &= \max\{\min\{\mathbf{m}^{k+1}, \mathbf{m}_2\}, \mathbf{m}_1\}. \end{aligned} \tag{17}$$

5 Terminate if the target accuracy is reached

$$\frac{\|\mathbf{m}^{k+1} - \mathbf{m}^k\|_2}{\|\mathbf{m}^k\|} \ \leq \ \text{target accuracy}. \tag{18}$$

or go back to step 2 otherwise.

Optimizing (16) with respect to $\mathbf{m}$ is in itself a large-scale optimization problem, nonlinear for a nonlinear modeling operator $\mathbf{F}$. Solving the optimization problem (16) exactly is unnecessary because for small k (i.e., at early stages of the inversion) vector $\mathbf{y}^k$ is not the true model, vector $\mathbf{x}^k$ is far from the true model gradient, and the multipliers $\mathbf{b}^k, \mathbf{x}^k$ could be far from scaled Lagrange multipliers, s. Therefore, for large-scale problems only a few steps of an iterative method like conjugate gradients need be carried out. As the solution converges to the true solution and critical sharp contrasts in the model are identified, an iterative solver can begin to take advantage of the objective function curvature information collected at previous iterations of the *outer* loop, potentially leading to a significantly faster convergence. Optimal strategies for spanning iterations of nonlinear conjugate gradients across iterations of the *outer* loop of our algorithm are the subject of an upcoming report.

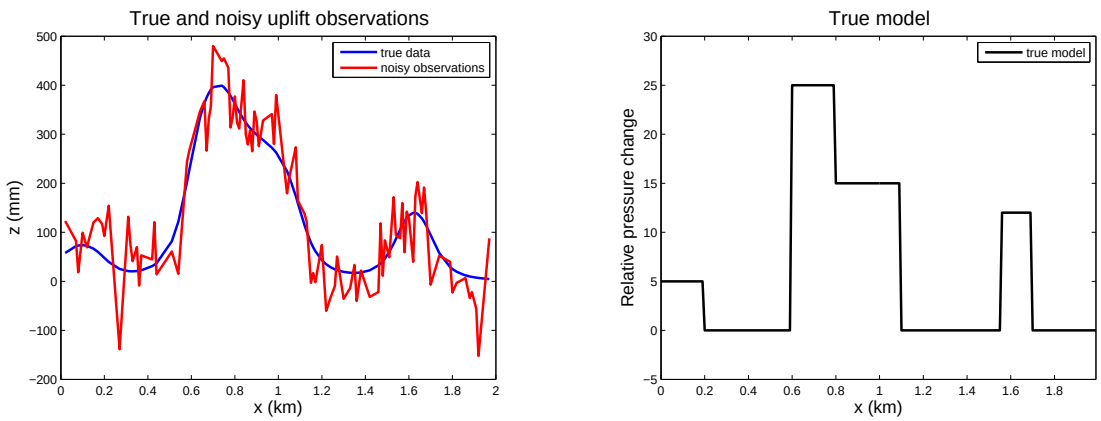
RESULTS

We demonstrate our method with a test problem that simulates vertical surface uplift in response to distributed dilatational sources, mathematically equivalent to surface deformation due to pore pressure change (Segall, 2010). Our modeling operator is defined as

$$\mathbf{F}(\mathbf{m}) \ = \ u(x), \ u(x) \ = \ \int_0^A \frac{D^3 m(\xi) d\xi}{(D^2 + (x-\xi)^2)^{3/2}}, \tag{19}$$

where we assume that $\mathbf{m} = m(\xi), \xi \in [0, A]$ is a relative pore pressure change along a linear segment $[0, A]$ of a reservoir at a constant depth D, and $\mathbf{u} = u(x), x \in$

$[0, A]$, within a proportionality factor determined by poroelastic medium properties (Maharramov and Zoback, 2015), is the induced vertical uplift on the surface. For demonstration purposes we consider a one-dimensional model but the results trivially extend to realistic reservoir and surface geometries. Operator (19) is a smoothing operator, and recovering sharp pressure contrasts e.g. due to permeability barriers requires model "styling" or regularization such as blockiness-promoting ROF model. As a true model we used a highly compartmentalized pressure model of Figure 1(b). In our experiments, we set $D = 100$m $A = 2$km, and discretized both the model and data space using a 200-point uniform grid. Random Gaussian noise with $\sigma = 15\%$ of the maximum clean data amplitude was added to the clean forward-modeled data to produce the noisy observations shown in Figure 1(a).



(a) (b)

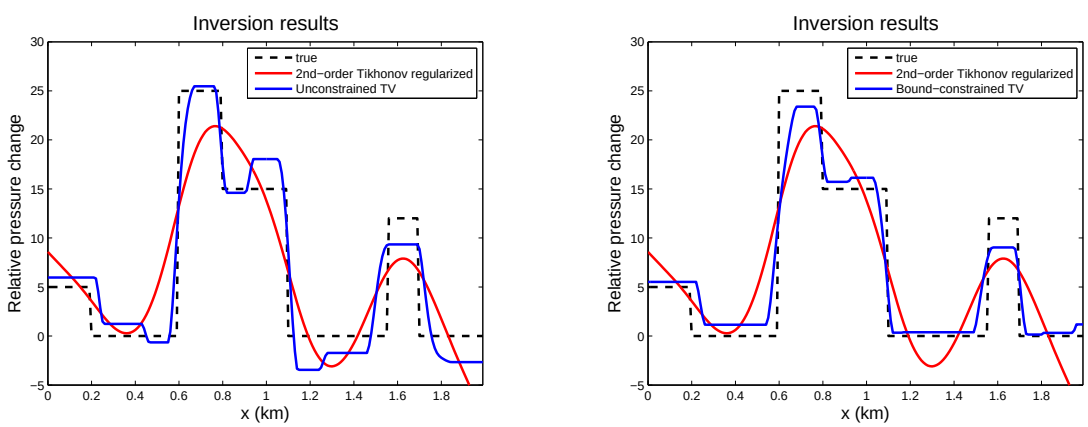
Figure 1: (a) True and noisy uplift observations. Random Gaussian noise with $\sigma = 15\%$ of maximum clean data amplitude was added to the clean data. (b) True model exhibibits a highly compartmentalized "blocky" behavior. [**CR**]

The result of a TV-regularized *unconstrained* inversion is shown in Figure 2(a) against the true model and a Tikhonov-regularized inversion. This result was obtained using the above algorithm by setting $\delta = 0$ (no bound constraints) and using the values of $\alpha = 1$ and $\lambda = 2$. The TV-regularized result captures the compartmentalized picture of pressure distribution better than the highly smoothed Tikhonov regularization result. However, due to absence of bound constraints, lower pressure bounds are not honored, resulting in negative pressure areas that are not present in the true model. The result of running our bound-constrained TV-regularization algorithm is shown in Figure 2(b). The imposition of bound constraints not only removed the negative relative pressure areas, but also removed the pressure spike at about $x \approx 1$km in the unconstrained inversion of Figure 2(a) that apparently had resulted from compensating negative pressures. In both the constrained and the unconstrained runs we conducted 1000 outer loop iterations with 2 inner loops cycles. However, the algorithm converged quickly, with only a few initial iterates outside a tight neighborhood of the final curve, as shown in Figure 3(b). Finally, we note that many practical

implementations of bound constraints often resort to a simplistic way of enforcing the constraints: the inverted model is projected onto the bounding rectangle either once after applying a direct unconstrained solver, or at each iteration of an unconstrained solver. In this case variable $\mathbf{y}$ and the associated quadratic regularization term are not introduced into the objective function. This may result in a violation of the KKT optimality conditions where the bound constraints are active (Nocedal and Wright, 2006), and is demonstrated by the blue plot in Figure 3(a). While the bound constraints are honored, the solution is both qualitatively and quantitatively far from optimal.

CONCLUSIONS AND PERSPECTIVES

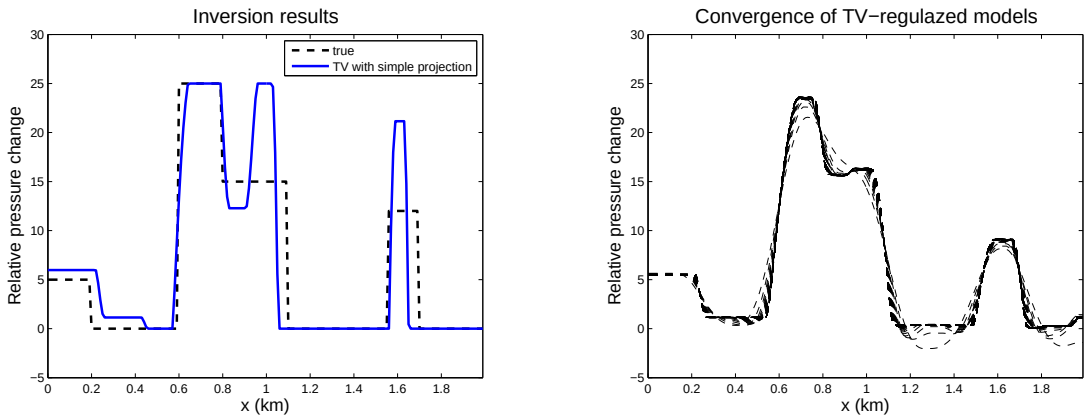
Our algorithm combines the advantages of the blockiness-promoting and edge-preserving ROF model with the ability to impose bound constraints. The splitting mechanism used for enforcing the bound constraints is naturally integrated into the split-Bregman iterations and results in no extra computational cost. The method was able to resolve compartmentalized subsurface pressure changes from noisy surface uplift observations despite the highly diffusive nature of the underlying deformation process. The



(a) (b)

Figure 2: (a) Unconstrained TV-regularized inversion. The algorithm tries to fit the data by allowing negative relative pressure changes. (b) Bound constrained TV-regularized result. Note that enforcing lower bounds resulted in a more accurate shape matching of the true model. [**CR**]

method can be implemented around any large-scale nonlinear solver such as conjugate gradients or quasi-Newton methods. Additional equality and inequality constraints can be incorporated into the algorithm using the general ADMM framework.



(a) (b)

Figure 3: (a) Direct imposition of the bound constraints at each iteration of the unconstrained solver resulted in a qualitatively and quantitatively wrong inversion. (b) Convergence of TV-regularized inverted models with bound constraints. The method quickly resolves both sharp contrasts and active bounds as only a few initial curves out of 1000 iterates lie outside a small neighborhood of the final curve. [**CR**]

REFERENCES

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, 2010, Distributed optimization and statistical learning via the alternating direction method of multipliers: Foundations and Trends in Machine Learning, **3**, 1–122.

Chambolle, A. and P. L. Lions, 1997, Image recovery via total variational minimization and related problems: Numerische Mathematik, **76**, 167–188.

Chartrand, R. and B. Wohlberg, 2010, Total-variation regularization with bound constraints: Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference, 766–769.

Goldstein, T. and S. Osher, 2009, The split bregman method for L1-regularized problems: SIAM Journal on Imaging Sciences, **2**, 323–343.

Ma, Y., M. Maharramov, R. Clapp, and B. Biondi, 2015a, Illumination compensation by L1 regularization and steering filters: SEP Report, **158**.

———, 2015b, Time lapse seismic imaging with L1 regularization and steering filters: SEP Report, **158**.

Maharramov, M. and B. Biondi, 2015, Resolving the effects of production-induced overburden dilation using simultaneous TV-regularized time-lapse FWI: SEP Report, **158**.

Maharramov, M., Y. Ma, and B. Biondi, 2015, Double-difference time-lapse FWI with a total-variation regularization: SEP Report, **158**.

Maharramov, M. and M. Zoback, 2015, Reservoir monitoring by inverting pore pressure change from surface deformation: SEP Report, **158**.

Nocedal, J. and S. J. Wright, 2006, Numerical optimization: Springer.

Parikh, N. and S. Boyd, 2013, Proximal algorithms: Foundations and Trends in Optimization, **1**, 123–231.

Rudin, L. I., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: Physica D: Nonlinear Phenomena, **60**, 259–268.

Segall, P., 2010, Earth and volcano deformation: Princeton University Press.