Introduction

Large-scale non-linear inverse problems frequently arise in different fields of geosciences. Such problems are inherently ill-posed and thus require regularization techniques to be included in iterative local optimization (gradient-based) methods. Popular regularizations such as those promoting sparsity and blockiness are non-smooth, which makes their implementation in non-linear problems challenging. In this abstract, we propose two proximal Newton-type algorithms to solve non-linear problems with non-smooth regularization. Similar to classical Newton-type methods, the nonlinear misfit function is replaced by a locally-quadratic majorizing function, while a proximal algorithm is used to determined the step direction in a particular way to involve the gradients/subgradients of the non-differentiable regularization function. We propose two distinct algorithms to determine the step direction, which are referred to as regularized least-squares minimization. The first algorithm relies on operator splitting and updates the step direction iteratively in an inner loop through a proximal gradient method and a generalized iterative shrinkage-thresholding algorithm (ISTA) (Daubechies et al., 2004), leading to Newton-ISTA (NISTA). A key property of this method is that we only need Hessian-vector products to determine the step direction. The second algorithm relies on the alternating direction method of multipliers (ADMM), where auxiliary variables are introduced to decouple the quadratic term from the non-differentiable one and minimize them in alternating mode at each iteration (Boyd et al., 2010; Aghamiry et al., 2019b,a), leading to Newton-ADMM (NADMM). This algorithm requires to solve a Newton system at each iteration, with approximate or exact methods. We implement the proposed methods in full-waveform inversion (FWI), an ill-posed nonlinear optimization problem, which aims to estimate subsurface parameters $m$ by minimizing the distance between observed and calculated data (Tarantola, 1984; Pratt et al., 1998):

$$\mathcal{M}(m) = \frac{1}{2} ||d - F(m)||^2_2,$$

(1)

where the observed data are denoted by $d$, the calculated data by $F(m) = PA^{-1}(m)b$, with $A$ the wave-equation operator, $b$ the source and the detection operator $P$ samples the wavefield $A^{-1}(m)b$ at the receiver position. Numerical tests performed show outstanding performance of the proposed algorithms for regularization of FWI.

Method

Full-waveform inversion with a general form of regularization can be written as

$$\min_m \mathcal{M}(m) + \lambda \mathcal{R}(m),$$

(2)

where $\mathcal{R}(m)$ is a possibly non-differentiable regularization, which encodes the prior knowledge about the model parameters and prevents data overfitting, and $\lambda$ is the trade-off parameter that balances the misfit and regularization terms. A Newton-type method approximates the misfit function $\mathcal{M}$ with a local quadratic function given by

$$\tilde{\mathcal{M}}_k(m) = \mathcal{M}(m_k) + \nabla \mathcal{M}(m_k)^T (m - m_k) + \frac{1}{2} (m - m_k)^T H_k (m - m_k),$$

(3)

where $k$ is the iteration number, $m_k$ is the $k^{th}$ model-parameter iterate, $\nabla \mathcal{M}(m_k)$ is the gradient vector, and $H_k$ is the Hessian matrix $\nabla^2 \mathcal{M}(m_k)$ or an approximation of it (if quasi-Newton methods are desired). Using this approximation, proximal Newton-type methods (Lee et al., 2014) solve eq. 2 iteratively as

$$m_{k+1} = m_k + c_k \Delta m_k,$$

(4)

where $c_k$ is the step length, which can be determined by a line search method, and

$$\Delta m_k = \arg \min_{\Delta m} \tilde{\mathcal{M}}_k(m_k + \Delta m) + \lambda \mathcal{R}(m_k + \Delta m)$$

(5)

is the step direction. Computation of the step direction $\Delta m_k$ with Newton methods is the most computationally expensive part of this algorithm because it involves the Hessian matrix with the regularization term $\mathcal{R}$. However, various approximations of the Hessian can be employed leading to proximal quasi-Newton methods or proximal gradient method if $H_k$ reduces to a scaled version of the identity matrix. In the following we propose two methods to solve eq. 5.

Newton-ISTA (NISTA)

Generalized ISTA solves eq. 5 iteratively as (Attouch and Peypouquet, 2016)

$$\begin{cases} \Delta m^{\ell+1}_k = \text{prox}_{\lambda c_k \mathcal{R}} \left( m_k + \Delta m^{\ell} - c_k (H_k \Delta p^{\ell} + \nabla \mathcal{M}(m_k)) \right) - m_k, \\ \Delta p^{\ell+1} = \Delta m^{\ell+1} + \left( \frac{1}{c_k^2} \Delta m^{\ell+1}_k - \Delta m^{\ell}_k \right), \end{cases}$$

(6)

where $\ell$ denote the inner iteration count, $c_k \in (0, 1/\|H_k\|^2)$, and

$$\text{prox}_{\lambda c_k \mathcal{R}}(m) = \arg \min_x \lambda \mathcal{R}(x) + \frac{1}{2c_k} ||x - m||^2_2.$$ 

(7)
The NISTA method is summarized in Algorithm 1.

**Algorithm 1:** Regularization by NISTA.

Require: starting point \( m_0 \)
set \( \Delta p^0 = 0 \)
repeat
  Compute the Hessian \( H_k \) or an approximation to it.
  Compute the step direction:
  for \( i = 1 \) to \( N - 1 \) do
    \( \Delta m_i^{k+1} \leftarrow \text{prox}_{\lambda/m_i} (m_i + \Delta p - c_i (H_i \Delta p + \nabla \mathcal{A}(m_i))) - m_i \)
    \( \Delta p \leftarrow \Delta m_i^{k+1} + \frac{1}{2} (\Delta m_i^{k+1} - \Delta m_i^k) \)
  end for
  Select step length \( \alpha_k \) with a backtracking line search.
Update: \( m_{k+1} \leftarrow m_k + \alpha_k \Delta m_k \)
until stopping conditions are satisfied.

Newton-ADMM (NADMM)

NADMM is obtained by solving (5) via ADMM (Boyd et al., 2010). By introducing the auxiliary variable \( p = m_k + \Delta m \), we recast the minimization in eq. 5 as the following constrained problem:

\[
\min_{\Delta m, p} J_k(m_k + \Delta m) + \lambda \mathcal{R}(p) \quad \text{subject to} \quad m_k + \Delta m = p. \tag{8}
\]

Solving (8) with an augmented Lagrangian method leads to the following saddle point problem

\[
\min_{\Delta m, p} \max_{q} \left( J_k(m_k + \Delta m) + \lambda \mathcal{R}(p) + \langle q, m_k + \Delta m - p \rangle + \frac{1}{2c_k} \|m_k + \Delta m - p - q\|^2 \right), \tag{9}
\]

where \( q \) is the Lagrange multiplier. Applying the scaled form of ADMM to eq. 9, when combined with using eq. 4, gives the iteration

\[
\begin{align*}
\Delta m_k &= \arg \min_{\Delta m} J_k(m_k + \Delta m) + \frac{1}{2c_k} \|m_k + \Delta m - p_k - q_k\|^2 \\
m_{k+1} &= m_k + \alpha_k \Delta m_k \\
p_{k+1} &= \arg \min_p \lambda \mathcal{R}(p) + \frac{1}{2c_k} \|m_{k+1} - p - q_k\|^2 \\
q_{k+1} &= q_k + p_{k+1} - m_{k+1}.
\end{align*} \tag{10}
\]

The first subproblem in eq. 10 is a generalized gradient step because it implicitly includes the gradient information of the possibly non-differentiable regularizer. The regularization effect appears as adding a constant to the diagonal of the Hessian. It is quadratic and admit a closed form solution. The second subproblem is a proximal mapping and, in many cases, easy to solve. This iteration described by eq. 10 appears as the outer loop labeled by subscript \( k \) in the NADMM algorithm. One may view to iteratively refine the solution of each subproblem independently by interlacing one inner loop per subproblem in the outer loop 10. However, it has been shown numerically that a single inner iteration leads to the fastest convergence. Accordingly, the proposed NADMM is summarized in Algorithm 2.

**Algorithm 2:** Regularization by NADMM.

Require: starting point \( m_0 \)
set \( p = q = 0 \)
repeat
  Compute the Hessian \( H_k \) or an approximation to it.
  Compute the step direction:
  \( \Delta m_k \leftarrow (c_k H_k + I)^{-1} (c_k \nabla \mathcal{A}(m_k) + (p + q - m_k)) \).
  Select step length \( \alpha_k \) with a backtracking line search.
Update: \( m_{k+1} \leftarrow m_k + \alpha_k \Delta m_k \)
Update: \( p \leftarrow \text{prox}_{\lambda c_k} (m_{k+1} - q) \).
Update: \( q \leftarrow q + p - m_{k+1} \)
until stopping conditions are satisfied.

Results

We first assess NISTA and NADMM against a simple two-variable nonlinear optimization problem.

\[
\min_{m_1, m_2} J(m_1, m_2) + \lambda \mathcal{R}(m_1, m_2), \tag{11}
\]

where \( J(m_1, m_2) = 75(m_2 - m_1^2)^2 + (1 - m_1)^2 \) is the Rosenbrock function and \( \mathcal{R}(m_1, m_2) = |m_1| + |m_2| \) is the \( l_1 \)-norm. The Rosenbrock function is continuously differentiable and has a global minimum at \((1,1)\).

Adding the sparsity-promoting regularization term to this function however moves this global minimum toward zero along a specific path. For \( 0 \leq \lambda \leq 3/2 \), the global minimum occurs at \((m_1^*, m_2^*)\) where \( m_1^* = (2 - \lambda)/(2 + 2\lambda) \) and \( m_2^* = (m_1^*)^2 - \lambda / 150 \). For \( 3/2 < \lambda \leq 2 \), it occurs at \((m_1^*,0)\) where \( m_1^* \) satisfies \( 300(m_1^*)^3 + 2m_1^* + \lambda - 2 = 0 \), and for \( \lambda > 2 \) the function reaches its global minimum at \((0,0)\). We applied both NISTA (with 50 inner iterations) and NADMM to minimize this nonlinear and non-differentiable function for \( \lambda = 3/2 \) (having the global minimum at \((0,1.0)\), Fig. 1). We also used the full Hessian and BFGS quasi-Newton Hessian in each algorithm. Both methods successfully converged to the desired global minimum with a faster convergence of NADMM (Fig. 1).

We assess now our methods for FWI, eq. 1. The gradient and Hessian are given by (Pratt et al., 1998)

\[
\nabla J(m) = -J^T \Delta d, \quad \text{and} \quad \nabla^2 J(m) = J^T J + \frac{\partial^2 F[J(m)]}{\partial m^2} \Delta d, \tag{12}
\]

where \( \Delta d = d - F(m) \) and \( J \) is the sensitivity or the Fréchet derivative matrix. In order to show the effectiveness of NISTA and NADMM, we build a challenging velocity model shown in Fig. 2a. The velocity
model contains four inclusions of different structure (piecewise constant, piecewise linear, smooth), in a 2 km × 2 km homogeneous background model. The data are generated with five sources at the surface (with 400 m spacing) and 50 m equally spaced receivers placed on all the boundaries except the surface. The forward modeling is performed with a 9-point stencil finite-difference method implemented with anti-lumped mass and PML absorbing boundary conditions to solve the Helmholtz equation. The source signature is a Ricker wavelet with a 10 Hz dominant frequency. We start the inversion from the homogeneous background model (Fig. 2b) and invert simultaneously four frequency components (5, 7, 10, and 12.5 Hz) with noiseless and noisy data (SNR=5db). We use the L-BFGS quasi-Newton method with line search for these results. We first performed the FWI without regularization (for λ = 0). The estimated models are shown in Fig. 2c and 2d for noiseless and noisy data, respectively. The estimated models contain severe artifacts and the inclusions are badly reconstructed, which prompts us to add regularization. Generally, it is very difficult to image these different structures simultaneously with classical regularizers. Here, we use a more sophisticated regularizer developed by the image denoising community to address this challenge. Obviously, the proximal operator in eq. 7 is a denoiser, which has not a functional form. This means that the proximity operator in NISTA or NADMM (Algorithm 1 or 2) can be applied without prior knowledge of the medium to be imaged and, hence can be used as a black-box regularizer (Kamilov et al., 2017). Here, we employ BM3D (Dabov et al., 2007) as the state-of-the-art denoising algorithm to perform regularization of FWI. Fig. 2e and 2g show the estimated models obtained by NISTA and NADMM for noiseless data, while Fig. 2f and 2h show the same results for noisy data. A direct comparison between the true model, the initial model, and the final models without/with regularization along with two vertical logs at horizontal distances 0.65 km, 1.80 km, and two horizontal logs at vertical depths 0.65 km and 1.9 km are also shown in Fig. 3. These results show that regularization allows for the successful reconstruction of the inclusions. The number of inner iterations for NISTA was 100, although it can be tuned adaptively based on the method of Lee et al. (2014). This test also confirms the faster convergence of NADMM relative to NISTA (Fig. 4). Since we assess both methods with the same L-BFGS algorithm, we would recommend NADMM at the expense of NISTA.

Conclusions

We proposed NISTA and NADMM for solving a general non-linear inverse problem with non-smooth regularization. They showed successful performance for minimization of the Rosenbrock function with one-norm regularization. We also successfully used them to solve a FWI problem with BM3D regular-

Figure 1: Minimization of the sparsity-promoting regularized Rosenbrock function with AISTA and NADMM using full Hessian matrix (red stars) and BFGS quasi-Newton Hessian (green circles).

Figure 2: Inclusion test: (a) True model, (b) Initial model. (c-d) Results of FWI without regularization for (c) noiseless and (d) noisy data. (e-h) Results of BM3D-regularized FWI with NISTA (e-f) and NADMM (g-h) for (e,g) noiseless and (f,h) noisy data.
Figure 3: Inclusion test: Direct comparison between FWI results of Fig. 2 along logs. For noiseless data, (a-b) vertical logs at $x=0.65$ km and $x=1.80$ km, (c-d) horizontal logs at $z=0.65$ km and $z=1.90$ km. (e-h) Same as (a-d) for noisy data.

Figure 4: Convergence history of NISTA and NADMM for the FWI example of Fig. 2.

ization, an adaptive regularizer suitable to build complicated subsurface models. Our future works aim at comparing NISTA and NADMM for more realistic FWI problems.

Acknowledgements
This work was granted access to the HPC resources of CINES under the allocation 2020 0596 made by GENCI. This study was partially funded by the WIND consortium (https://www.geoazur.fr/WIND), sponsored by Chevron, Shell and Total.

References