Fast "online" migration with Compressive Sensing

Felix J. Herrmann, Ning Tu, and Ernie Esser University of British Columbia, Earth and Ocean Sciences

Abstract

We present a novel adaptation of a recently developed relatively simple iterative algorithm to solve large-scale sparsity-promoting optimization problems. Our algorithm is particularly suitable to large-scale geophysical inversion problems, such as sparse least-squares reverse-time migration or Kirchoff migration since it allows for a tradeoff between parallel computations, memory allocation, and turnaround times, by working on subsets of the data with different sizes. Comparison of the proposed method for sparse least-squares imaging shows a performance that rivals and even exceeds the performance of state-of-the art one-norm solvers that are able to carry out least-squares migration at the cost of a single migration with all data.



Motivation

Wave-equation based migration corresponds to computing

$$\delta \mathbf{m} = \sum_{i=1}^{K} \nabla \mathbf{F}_{i}^{*}(\mathbf{m}_{0}, \mathbf{q}_{i}) \delta \mathbf{d}_{i}, \qquad (1)$$

where the sum runs over the number of experiments, indexed by $i = 1 \cdots K$. For seismic applications, this number *K* is typically very large. These experiments—with sources \mathbf{q}_i and corresponding data $\delta \mathbf{d}_i$ —involve either individual traces for different source/receiver pairs as in Kirchoff migration or (monochromatic) shot records as in (time-harmonic) reverse-time migration. In either case, the computational costs, including data-storage and movement, grows with *K* and the size of the survey area. To make matters worse, migration often relies on increasing the fold, to stack out unwanted artifacts in the image space, or on carrying out least-squares migration. In either case, the aforementioned increased costs strain even the largest computational resources stifling the adaptation of sparsity-promoting inversion techniques.

To overcome these costs, which are exacerbated by rapidly increasing costs of RTM as the frequency increases, we proposed randomized algorithms based on certain subsampling techniques that work on randomized subsets of the sources and corresponding data. These techniques have either lead to so-called batching techniques (van Leeuwen and Herrmann, 2013a), where the above sum is approximated by carrying out the sum over subsets of randomly selected sources (or simultaneous sources), or to techniques based on Compressive Sensing (Herrmann and Li, 2012). Contrary to batching techniques, where the error is controlled by increasing the size of the subsets, sampling related errors during Compressive Imaging are controlled by promoting curvelet-domain sparsity. By incorporating rerandomization techniques from stochastic optimization into sparsity-promoting imaging, we arrived at an imaging technology that is able to carry out sparsity-promoting least-squares migration working on small subsets at the cumulative costs of roughly one RTM carried out over all data (Herrmann and Li, 2012; Tu and Herrmann, 2014).

While our Compressive Imaging technology is certainly promising, it relies on the combination of a sophisticated ℓ_1 -norm solver (SPG11, van den Berg and Friedlander, 2008), rerandomizations, and warm starts. This makes the resulting algorithm complicated, and therefore difficult to implement on industry-scale problems. In addition, there is, aside from empirical observations, no formal proof what optimization problem this heuristic randomized approach actually solves. Below, we present a new highly flexible sparse inversion methodology that derives from a recently proposed sparse Kaczmarz solver for online problems in Compressive Sensing using linearized Bregman iterations (Lorenz et al., 2014). The resulting algorithm is relatively simple and suitable for online problems where measurements—read application of rows of the linearized modelling operator and therefore terms in the sum of Equation 1—take time and parallel resources to evaluate.

We organize our work as follows. First, we briefly state a simplified version of the algorithm undergirding our Compressive Imaging technology and juxtapose it with the proposed algorithm. Next, we compare the performance of the proposed algorithm on seismic imaging problems, followed by conclusions.

"Online" sparse solvers

We are interested in solving the following (partly) separable—i.e., we have access to individual rows as in trace-based migration or blocks of rows as in shot migration—linear inverse problems, where we aim to estimate the image **x** from measurements $\mathbf{b} \in \mathbb{R}^M$ given by $\mathbf{b} = \mathbf{A}\mathbf{x}$, with **x** sparse or compressible, and $\mathbf{A} \in \mathbb{R}^{M \times N}$ tall—i.e., $M \gg N$. This overdetermined problem is different from Compressive Sensing where $M \ll N$ and could in principle by solved easily by inverting **A** with least-squares or by simply applying the adjoint. Unfortunately, since we are dealing with problems where there are many rows of **A** that are expensive to evaluate, we can not follow this approach and we have to resort to working on randomized pairs $\{\mathbf{A}_i, \mathbf{b}_i\}$, made of randomly selected subsets of rows only—i.e., we only have access to $\mathbf{A}_i \in \mathbb{R}^{m \times N}$ and $\mathbf{b}_i \in \mathbb{R}^m$ with $m \ll N \ll M$. Our aim is now to come up with an algorithm where the total



amount of work, read the total number of rows participating in the inversion, is roughly the same as the total number of rows of the total system, or terms in the sum of Equation (1)—i.e., $L = m \times n_k \sim M$ with n_k the number of iterations of the solver.

Solution by Randomized Iterative Shrinkage-Thresholding Algorithm (RISTA)

While many algorithms have been developed to solve the convex Basis Pursuit problem for underdetermined **A**'s with $m \ll N$

BP: minimize $\|\mathbf{x}\|_1$ subject to $\mathbf{A}\mathbf{x} = \mathbf{b}$,

its solution often requires too many iterations, and therefore evaluations of **A** and its adjoint **A**^{*}, to make this a viable option for industrial problems, where the matrices are either very large or expensive to evaluate or both. To overcome this problem, the authors proposed a method to accelerate this type of algorithms using ideas from approximate message passing. The key idea of that approach was to "break" buildup of correlations between the model iterates, \mathbf{x}_k , and the matrix **A** by drawing randomized pairs $\{\mathbf{A}_k, \mathbf{b}_k\}$ for each iteration *k*. This results in Algorithm 1, a randomized version of the Iterative Shrinkage-Thresholding Algorithm (ISTA, Yin et al. (2008);Beck and Teboulle (2009)). For a fixed pair $\{\mathbf{A}_k, \mathbf{b}_k\}$ and appropriately chosen step length t_k and $\lambda \to 0^+$ (read as λ goes to 0 from above), Algorithm 1 provably solves BP. While RISTA type algorithms—where techniques from stochastic gradients are combined with sparsity promotion—lead to fast and therefore practical algorithms, their performance hinges on choices for the threshold. Not only do small λ 's lead to slow convergence, there is also no general proof that Algorithm 1 solves BP when choosing different pairs $\{\mathbf{A}_k, \mathbf{b}_k\}$ for each iteration. So the question is whether there exists an alternative iterative solver that makes few assumptions on **A** and that is capable of working with small randomized subsets of rows and that is relatively simple to implement.

Algorithm 1 Randomized Iterative Shrinkage-Thresholding Algorithm (RISTA) for appropriately chosen step lengths t_k 's and threshold λ for the soft thresholding with $S_{\lambda}(x) = \max(|x| - \lambda, 0)$.

Result: Estimate for the model x^k

1. **for** $k = 0, 1, \cdots$ **do** 2. Draw **A**_k and **b**_k

- 3. $\mathbf{z}_{k+1} = \mathbf{x}_k t_k \mathbf{A}_k^* (\mathbf{A}_k \mathbf{x}_k \mathbf{b}_k)$
- 4. $\mathbf{x}_{k+1} = S_{\lambda}(\mathbf{z}_{k+1})$ 5. end for

Solution by Randomized Shrinkage-Thresholding Kaczmarz Algorithm with linearized Bregman (RISKA)

Consider the following slight modification—i.e. we replace line 3 of Algorithm 1 by $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^* (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$. We call this Solution by Randomized Shrinkage-Thresholding Kaczmarz Algorithm with linearized Bregman (RISKA,Yin et al. (2008)). According to recent work by Lorenz et al. (2014), this modified algorithm solves for step lengths, $t_k = \frac{\|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|^2}{\|\mathbf{A}_k^* (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|^2}$,

KB:
$$\min_{\mathbf{x}} \operatorname{minimize} \lambda \|\mathbf{x}\|_1 + \frac{1}{2} \|\mathbf{x}\|_2^2$$
 subject to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

For large enough λ (van Leeuwen and Herrmann, 2013b), RISKA provably converges to the solution of BP. While the modification w.r.t. RISTA is at first sight relatively minor, its implications are profound and manifold. First, compared to RISTA the inversion involves two iterates, namely \mathbf{z}_{k+1} , which seeks to minimize the ℓ_2 -norm of the residual, and sparse model iterates \mathbf{x}_k obtained by soft thresholding \mathbf{z}_{k+1} with the threshold λ . Contrary to RISTA, updates for the \mathbf{z}_k 's are calculated by subtracting gradients computed with respect to the sparse model iterate \mathbf{x}_k . Third, the role of the threshold λ fundamentally changes. Within RISTA, the λ determines the projection of the model iterates on a λ dependent ℓ_1 ball. For a fixed λ the iterations without drawing new pairs { $\mathbf{A}_k, \mathbf{b}_k$ } solve the unconstrained problem min \mathbf{x} $\frac{1}{2} || \mathbf{A}\mathbf{x} - \mathbf{b} || + \lambda || \mathbf{x} ||_1$. By taking the limit $\lambda \to 0^+$ these unconstrained problems solve BP. Unfortunately, small λ lead to slow convergence (Hennenfent et al., 2008). For large-scale problems



this makes it necessary to devise ingenious cooling techniques where the ℓ_1 -norm constraint is gradually relaxed, e.g. by decreasing the λ 's as the algorithm progresses. This process allows components to enter into the solution slowly while bringing down the residual. Thresholding within RISKA, on the other hand, plays a completely different role since it balances ℓ_1 - and ℓ_2 -norms on the model. This means in a way that RISKA is "auto tuning"—i.e., when we choose λ too high, the \mathbf{x}_k will be zeros but not the \mathbf{z}_k 's. As k increases, the large entries—these are the first candidates to enter the solution—in \mathbf{z}_k will grow while the noisy crosstalk will become smaller if we chose to work with randomized $\{A_k, b_k\}$'s. After sufficient iterations, the number of which can be reduced by using intelligent step lengths, components will enter the solution. A too large λ for RISTA, on the other hand, would lead to zeros and would require a lowering of λ . Finally, RISKA solves a constrained problem, i.e., the data is fitted exactly (this can be relaxed to different norms error balls (Lorenz et al., 2013)). This has the advantage that step lengths can be larger leading to a very fast but above all extremely simple algorithm that lends itself well to (i) work on (randomized) subsets of the data. Depending on the available computational resources, this gives us the freedom to choose, for a fixed number of passes through the rows L, an optimal batch size m by trading of numbers of iterations and numbers of available nodes partaking in the inversion. For RTM this means that we can control the number of PDE solves while for trace-based migration this would limit the total number of source-receiver pairs touched by migration; (ii) implement RISKA in existing industrial settings as long as one has access to randomized subsets of data \mathbf{b}_k , the Born modelling \mathbf{A}_k , and its adjoint migration A_{k}^{*} . For RTM this means randomly selected shots or simultaneous shots while for Kirchhoff migration one should have access to randomized traces; (iii) select data and rows adaptively, e.g. according to the norm of the rows of A or condition number of block of rows as described in the literature (Strohmer and Vershynin, 2009; Needell and Tropp, 2014).

Examples

To illustrate the performance of RISKA, we compare sparsity-promoting linearized imaging results for primaries only and data including primaries and surface-related multiples obtained with our randomized SPG11 solver and with the above described algorithm. In all examples, we fix the number of passes through the data to one (L = M), which are equivalent to the cost of a single RTM with all data. The results included in Figure 1 show improvements, especially below the Salt (juxtapose Figure 1a and 1b). These results are remarkable because we are able to obtain high-fidelity inversion results at the cost of a single RTM. An advantage of RISKA is that the underlying algorithm is simple to implement and relatively robust w.r.t. choices of λ .

For the same budget (cost of 1 RTM), the proposed algorithm also produces improved results for an imaging example that includes surface-related multiples. In this case, we invert the linearized Born scattering operator with an areal source wavefield that contains the upgoing wavefield (Tu and Herrmann, 2014). Again, the results in Figure 2 are excellent despite the relative simplicity of the algorithm.

Conclusions

We presented a simplified randomized sparsity-promoting solver that provably converges to a mixed onetwo-norm penalized solution while working on arbitrarily small subsets of data. This latter ability, in combination with its simplicity and "auto-tuning, makes this algorithm particular suitable for large-scale and parallel industrial applications.

References

- Beck, A. and Teboulle, M. [2009] A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Img. Sci.*, **2**(1), 183–202, ISSN 1936-4954, doi:10.1137/080716542.
- Hennenfent, G., van den Berg, E., Friedlander, M.P. and Herrmann, F.J. [2008] New insights into one-norm solvers from the Pareto curve. *Geophysics*, **73**(4), A23–A26, doi:10.1190/1.2944169.
- Herrmann, F.J. and Li, X. [2012] Efficient least-squares imaging with sparsity promotion and compressive sensing. *Geophysical Prospecting*, **60**(4), 696–712, doi:10.1111/j.1365-2478.2011.01041.x.
- Lorenz, D.A., Schöpfer, F. and Wenger, S. [2013] The Linearized Bregman Method via Split Feasibility Problems: Analysis and Generalizations. *ArXiv e-prints*.
- Lorenz, D.A., Wenger, S., Schöpfer, F. and Magnor, M. [2014] A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing. *ArXiv e-prints*.





Figure 1: Comparison of randomized sparse inversions on primary only data using SPGL1 (a) and RISKA (b). It is clear that for the same computational budget, the inversion result for RISKA is less noisy and recovers the reflectors below the Salt better.



Figure 2: Comparison of randomized sparse inversions on data with surface-related multiples using SPGL1 (a) and RISKA (b). Again, the much simpler RISKA does an excellent job compared to SPGL1.

- Needell, D. and Tropp, J.A. [2014] Paved with good intentions: Analysis of a randomized block kaczmarz method. *Linear Algebra and its Applications*, **441**(0), 199 221, ISSN 0024-3795, doi: http://dx.doi.org/10.1016/j.laa.2012.12.022, special Issue on Sparse Approximate Solution of Linear Systems.
- Strohmer, T. and Vershynin, R. [2009] A randomized kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications*, 15(2), 262–278, ISSN 1069-5869, doi:10.1007/s00041-008-9030-4.
- Tu, N. and Herrmann, F.J. [2014] Fast imaging with surface-related multiples by sparse inversion, accepted for Geophysical Journal International.
- van den Berg, E. and Friedlander, M.P. [2008] Probing the pareto frontier for basis pursuit solutions. *SIAM Journal* on Scientific Computing, **31**(2), 890–912, doi:10.1137/080714488.
- van Leeuwen, T. and Herrmann, F.J. [2013a] Fast waveform inversion without source encoding. *Geophysical Prospecting*, **61**, 10–19, doi:10.1111/j.1365-2478.2012.01096.x, article first published online: 10 JULY 2012.
- van Leeuwen, T. and Herrmann, F.J. [2013b] Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, **195**, 661–667, doi:10.1093/gji/ggt258.
- Yin, W., Osher, S., Goldfarb, D. and Darbon, J. [2008] Bregman iterative algorithms for ℓ₁-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, **1**(1), 143–168, doi:10.1137/070703983.