Pass on the message: recent insights in large-scale sparse recovery

Felix J. Herrmann

January 19, 2012

Abstract

Data collection, data processing, and imaging in exploration seismology increasingly hinge on large-scale sparsity promoting solvers to remove artifacts caused by efforts to reduce costs. We show how the inclusion of a "message term" in the calculation of the residuals improves the convergence of these iterative solvers by breaking correlations that develop between the model iterate and the linear system that needs to be inverted. We compare this message-passing scheme to state-of-the-art solvers for problems in missing-trace interpolation and in dimensionality-reduced imaging with phase encoding.

Introduction

Modern-day exploration seismology increasingly relies on the solution of large-scale optimization problems that require multiple passes through the data. This leads to major challenges because seismic imaging increasingly depends on long-offset and full-azimuth sampling, which leads to exponential growing costs for data collection, storage, and processing. We discuss how recent insights from statistical physics and compressive sensing can be used to reduce processing costs by minimizing the number of required passes through the data. More specifically, we look at the method of approximate message passing (AMP, see e.g. Donoho et al. (2009)), which leads to significant improvements in the convergence of large-scale sparsity-promoting solvers for specific problems. The outline is as follows. First, we briefly introduce ideas behind randomized dimensionality reduction techniques that include compressive sensing for seismic acquisition, processing, and imaging. Second, we discuss conventional solution strategies for large-scale sparsity-promoting optimization problems. Third, we introduce a new solution method based on approximate message passing. We conclude by applying this specialized framework to missing-trace interpolation and imaging.

Randomized dimensionality reduction

By working on smaller randomized subproblems, while exploiting structure within signals, challenges related to the so-called 'curse of dimensionality' are being addressed in many different research areas. This curse leads to exponential growth in data-collection and processing costs as the surface area and desired resolution increase. Recent developments in randomized acquisition and efficient imaging with source encoding are examples that address these challenges. However, in both examples cost reductions go at the expense of creating artifacts, such as aliasing or source crosstalk. Therefore, the challenge is to render these interferences incoherent so they can be removed and amplitudes can be restored.

Compressive sensing

To remove, or at least surpress, subsampling-related artifacts transform-domain sparsity promotion is often employed. During sparsity promotion incoherent artifacts, such as spectral leakage or source crosstalk, are mapped back to coherent signal. This is done via a iterative procedure during which we transform the data to separate the significant transform coefficients from interference "noise". Mathematically, this approach corresponds to solving the sparsity-promoting program Basis Pursuit:

$$\mathrm{BP}: \qquad \underset{\mathbf{x}}{\text{minimize}}\, \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{1}$$

During recent years, a tremendous body of work has been developed in support of this convex-optimization program where sparse vectors $\mathbf{x} \in \mathbb{C}^N$ are recovered from incomplete measurements $\mathbf{b} = \mathbf{A}\mathbf{x}_0$ with $\mathbf{x}_0$ sparse and $\mathbf{b} \in \mathbb{C}^n$ with $n \ll N$. So far, focus has been on (i) deriving conditions that guarantee recovery of k-sparse vectors with sparsity levels $\rho = k/n$ from measurements by sampling matrices $\mathbf{A}$ with aspect ratios $\delta = n/N$; (ii) designing sampling schemes that favour recovery, e.g., the design of randomized acquisition or phase encoding; (iii) implementing fast ($\mathscr{O}(N \log N)$) sparsifying transforms that exploit structure, e.g. curvelets in seismic exploration; (iv) developing algorithms that are frugal in the number of matrix-vector multiplies they require to solve BP.

Solutions by cooling

Even though the framework of compressive sensing has let to major breakthroughs in many research areas including MRI, the application of its principles to exploration seismology has been challenging because of the large scale of problems that easily involve vectors with billions of unknowns. In addition, BP corresponds to solving the limiting case ($\lambda \downarrow 0$) of the quadratic problem: $\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|x\|_1$, which is known to converge slowly as a function of the number of matvec's as the trade-off parameter $\lambda \downarrow 0$. To overcome this problem, 'optimizers' use continuation methods that employ properties of the Pareto tradeoff curve that traces the one-norm of the solution against the two-norm of the residual. This leads to approaches were series of subproblems are solved that allow components to enter the solution contollably by slowly increasing the one-norm of the solution. Each subproblem involves gradient

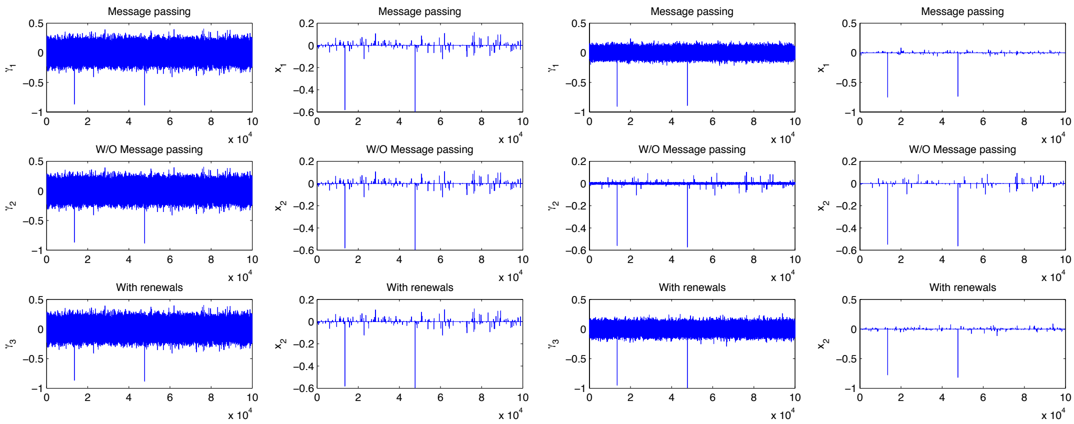

Figure 1 *Model iterates after one (left) and and two (right) iterations before and after soft thresholding. Notice the spurious spiky artifacts that develop after the first iteration in the second row. These correlations are removed by the message term (top row) or by drawing a new sampling matrix and data for each iteration (bottom row).*

updates, $\mathbf{x}^{t+1} = \mathbf{x}^t + \mathbf{A}^H(\mathbf{b} - \mathbf{A}\mathbf{x}^t)$ with t the iteration number, in combination with a nonlinear projection promoting sparsity. Not withstanding the success of these continuation methods, which undergird state-of-the-art versatile solvers such as SPGℓ_1 (van den Berg and Friedlander, 2008), convergence for large systems remains challenging in particular when given a small budget ($\mathscr{O}(50)$) of matvecs.

Solutions by approximate message passing

To understand the performance of solvers for BP, consider the second row of Fig. 1 that includes the first two model iterates before and after soft thresholding for $\mathbf{A}$ a Gaussian 10×10^5 matrix and $\mathbf{x}$ a vector with two ones at random locations. (Remember that soft thresholding is an elementwise operation that jointly minimizes the two-norm on the residual and the absolute value of the output and is given by $\eta(x, \tau) = \mathrm{sgn}(x)\max(|x| - \tau, 0)$, with $\tau > 0$ the threshold value.) We use Gaussian matrices because they lead to Gaussian interferences, i.e., $\mathbf{A}^H\mathbf{A}\mathbf{x}_0 = \mathbf{x}_0 + \mathbf{w}$ with $\mathbf{w}$ zero-mean Gaussian noise with variance $n^{-1}\|\mathbf{x}_0\|_2^2$ (Montanari, 2010). This property, which is common amongst compressive-sensing matrices that favor recovery by BP, reduces the recovery problem to a simple "denoising" problem, which can be solved by soft thresholding with the proper threshold level. Unfortunately, this property no longer holds after the first iteration because dependencies emerge between the model iterate $\mathbf{x}^t$ and the matrix $\mathbf{A}$ for iterations $t > 1$. These dependencies cause spurious artifacts and this leads to slow convergence.

Suppose now that we select a new matrix $\mathbf{A}$ and corresponding measurement vector $\mathbf{b}$ at each iteration as suggested by Montanari (2010). In that case, correlations can no longer develop between the model iterate and the matrix rendering soft thresholding effective for each iteration (juxtapose the second and third row of Fig. 1). While this idea has been used successfully in situations where data is abundant, e.g., in seismic imaging where different independent randomized subsets of fully-sampled data volumes are drawn to speed up convergence (Herrmann and Li, 2011), this solution is unworkable in data-scarce situations, such as during the recovery from incomplete field data.

Using arguments from statistical physics, Donoho et al. (2009) address this issue by including an additional 'message-passing' term in iterative soft thresholding schemes that solve BP. With this additional term, we iterate the following approximate message-passing scheme (AMP):

$$\begin{aligned} \mathbf{x}^{t+1} &= \eta_t\left(\mathbf{A}^H\mathbf{r}^t + \mathbf{x}^t\right) \\ \mathbf{r}^t &= \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{I_t}{n}\mathbf{r}^{t-1}, \end{aligned} \tag{2}$$

with $\mathbf{x}^0 = 0$, $\mathbf{r}^0 = \mathbf{b}$. Here, $\eta_t(\cdot)$ is a iteration-dependent soft thresholding, and I_t is the number of entries that survived the threshold of the previous iteration. As we can see from the first row in Fig. 1, inclusion of this extra term in Eq. 2 cancels the correlations and corresponds to adding the 'residual' of

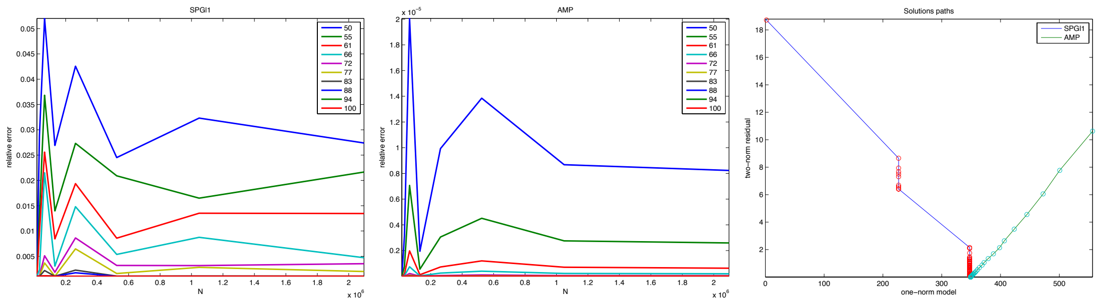

Figure 2 *Performance comparison for SPGℓ_1 and AMP. Relative recovery errors for increasing problem size for and increasing number of matvecs for SPGℓ_1 (left) and AMP (middle). Solution path is plotted on the right.*

the previous iteration scaled by the ratio of the number of elements that survived the threshold and the number of measurements. This extra term clearly breaks the spurious interferences and guarantees that each iteration becomes a simple denoising problem ideally suited for soft thresholding.

Large dimensionality: a blessing in disguise

The theory explaining the improved performance by adding the message term in Eq. 2 is involved but can be summarized as follows. First, the message term leads to asymptotic cancellation of the damaging correlations and is valid for system sizes going to infinity ($N \to \infty$) for recovery problems with Gaussian matrices. In that limit, Montanari (2010) argues that the linear system of equations decouples such that the recovery for each entry of the model iterate boils down to estimating x_i^t using the property $\left(\mathbf{x}^t + \mathbf{A}^H \mathbf{r}^t\right)_i = (\mathbf{x} + \tilde{\mathbf{w}})_i$ for $i = 1 \cdots N$. This can be done effectively by soft thresholding using the fact that the $\{\tilde{w}_i\}_{i=1 \cdots N}$ are asymptotically Gaussian in the large-scale limit. Second, and this is somewhat of a hand waiving argument, large matrices tend to behave as Gaussian matrices even if their entries are not Gaussian or if they are obtained by some other randomization such as random phase encoding or random restriction. For example, the randomly restricted Fourier matrix behaves approximately as a Gaussian matrix. Third, the iterative procedure with the message-passing term solves BP and therefore recovers k-sparse vectors with high probability if certain conditions are met on the aspect ratio of $\mathbf{A}$ and the sparsity level of the vector $\mathbf{x}_0$ (see e.g., Donoho et al. (2009) for a detailed overview on phase diagrams that predict the transition from recoverable to non-recoverable combinations of (ρ, δ)). Now, if these large-scale limit asymptotic arguments indeed hold then the message-passing algorithm is truly remarkable because we may be able to improve the performance of highly sophisticated one-norm solvers for exceedingly large problem sizes. Evidently, this promise comes with the caveat that message-passing algorithms are specifically designed to solve sparse-recovery problems for Gaussian matrices while methods such as SPGℓ_1 are versatile and solve BP for any $\mathbf{A}$ and $\mathbf{b}$ as long one is willing to spend a sufficient number of iterations to bring the residual down.

Discussion and examples

SPGℓ_1 versus AMP as $N \to \infty$: To illustrate the performance of AMP for a limited number of matvec's, we compare recoveries for $N = 2^j$, $j = 14 \cdots 22$ with fixed aspect ratio $\delta = 1/300$, sparsity level $\rho = 1/20$, and increasing number of matvecs. For small numbers of matvec's, the relative errors for AMP is significantly smaller than for SPGℓ_1 as shown by Fig. 2. These results are consistent with theoretical predictions. Also notice that the solution paths are very different (Fig. 2(right)).

Recovery from missing traces: While AMP is strictly speaking designed for Gaussian matrices and strictly sparse vectors only, we examine its performance for the recovery of a seismic line from 50% randomly missing sources using roughly 100 matvec's. We use the restricted 3D curvelet transform for the recovery. Output shot records for SPGℓ_1 and AMP are plotted in Fig. 3 and show clear improvements for AMP despite violations of the underlying assumptions. (The restricted 3D curvelet transform matrix is not a Gaussian matrix and the vector is not strictly sparse but compressible.) The SNRs, 7.75dB for SPGℓ_1 and 9.75dB for AMP, confirm this observation. Aside from this remarkable improvement, the

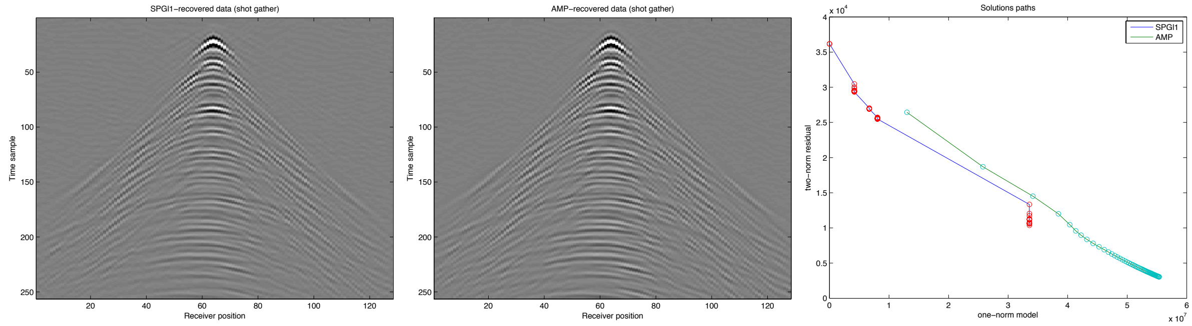

Figure 3 *Recovery from* 50% *missing shots with SPG*ℓ_1*(left) and AMP (middle). Solution paths (right).*

residue is significantly smaller after 50 iterations and the solution paths for AMP points to the ℓ_1-norm of the curvelet coefficient vector obtained by solving SPGℓ_1 for 300 matvecs on the complete data. This is remarkable and highly encouraging because this is a very large-scale problem ($N = 1.12 \times 10^9$).

Efficient imaging: Aside from improvements in the recovery of missing data, message-passing ideas also provide an explanation for efficient sparsity-promoting imaging where new subsets of simultaneous shots are drawn each time one of the subproblems of SPGℓ_1 is solved (Herrmann and Li, 2011). As can be seen from Fig. 4, these renewals have a similar positive effect as the message term in Eq. 2.

Acknowledgements: Thanks to the sponsors of the SINBAD consortium, NSERC, and the authors of CurveLab and Ulugbek Kamilov for making the AMP software available.

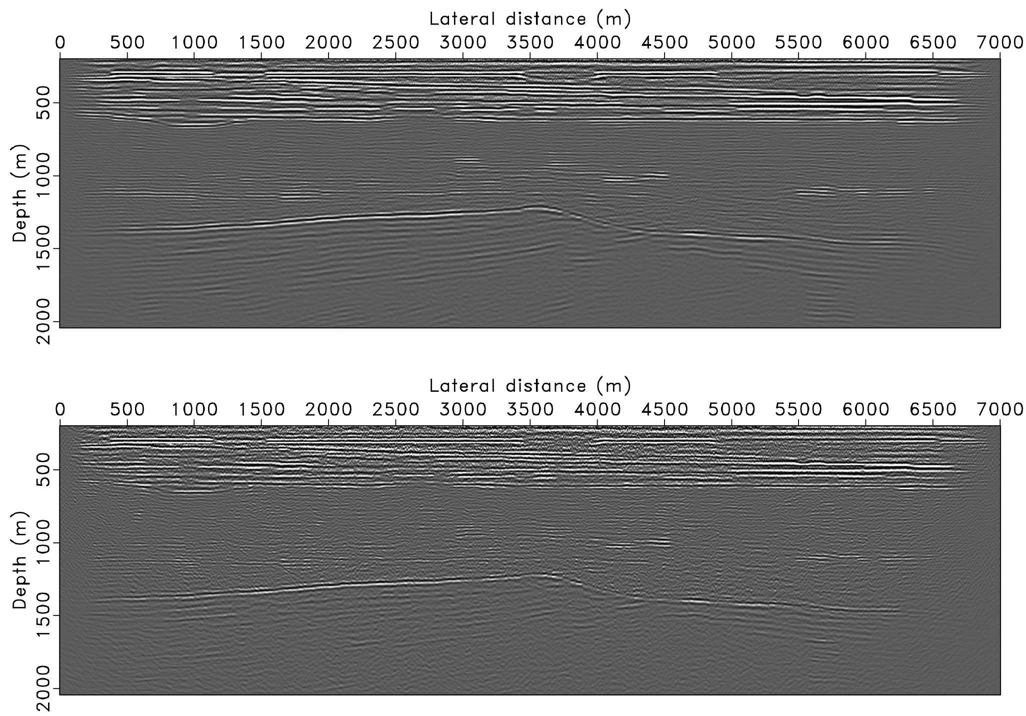

Figure 4 *Imaging with and without 'messaging' (adapted from Herrmann and Li, 2011).*

References

Donoho, D., Maleki, A. and Montanari, A. [2009] Message passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, **106**(45), 18914–18919.

Herrmann, F.J. and Li, X. [2011] Efficient least-squares imaging with sparsity promotion and compressive sensing, to appear in Geophysical Prospecting.

Montanari, A. [2010] Graphical models concepts in compressed sensing. *CoRR*, **abs/1011.4328**.

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.