Controlling linearization errors in ℓ_1 regularized inversion by rerandomization

Ning Tu, Xiang Li and Felix J. Herrmann, Earth and Ocean Sciences, University of British Columbia

SUMMARY

Linearized inversion is a data fitting procedure that tries to match the observed seismic data with data predicted by linearized modelling. In practice, the observed data is not necessarily in the column space of the linearized modelling operator. This can be caused by lack of an accurate background velocity model or by coherent noises not explained by linearized modelling. Through carefully designed experiments, we observe that a moderate data mismatch does not pose an issue if we can use *all* the data in the inversion. However, artifacts do arise from the mismatch when randomized dimensionality reduction techniques are adopted to speed up the inversion. To stabilize the inversion for dimensionality reduction with randomized source aggregates, we propose to *rerandomize* by drawing independent simultaneous sources occasionally during the inversion. The effect of this rerandomization is remarkable because it results in virtually artifact-free images at a cost comparable to a single reverse-time migration. Implications of our method are profound because we are now able to resolve fine-scale steep subsalt features in a computationally feasible manner.

INTRODUCTION

Linearized inversion aims to obtain the perturbation of the subsurface medium parameterized by a model vector $\mathbf{m}$, by fitting the predicted data to the observed seismic data. The predicted data is identified as a linear function of the model perturbation $\delta\mathbf{m}$ over a background model $\mathbf{m}_0$ where $\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}$. The underlying optimization problem can be written as:

$$\mathbf{LS}: \quad \underset{\delta\mathbf{m}}{\text{minimize}} \quad \|\nabla\mathbf{F}[\mathbf{m}_0,\mathbf{q}]\delta\mathbf{m} - \mathbf{d}\|_2,$$

where $\nabla\mathbf{F}$ is the linearized Born scattering operator, $\mathbf{q}$ is the vectorized source wavefield for all sources, which is either assumed to be known, or can be estimated by variable projection (Aravkin et al., 2012; Tu et al., 2013), and $\mathbf{d}$ is the observed data, usually after removal of surface-related multiples.

Despite all its advantages, such as true-amplitude preservation and robustness to missing data (Tu et al., 2013; Nemeth et al., 1999), linearized inversion is not widely adopted because of its high computational cost. Solving problem **LS** usually requires iterative evaluations of the Born scattering operator and its adjoint, while each evaluation is already expensive as it solves four PDEs for each monochromatic source experiment. Since the simulation cost is mainly determined by the number of sources and frequencies, dimensionality reduction techniques—such as source encoding and using randomized frequency subsets—are proposed to reduce these costs (Dai et al., 2011; Herrmann and Li, 2012). To overcome the negative effects these subsampling techniques have (e.g., source cross-talk) on the seismic image, Herrmann and Li (2012) proposed to promote Curvelet sparsity in the image space by solving the following ℓ_1 minimization problem (known as the Basis Pursuit De-Noise (BPDN) problem, Chen et al., 2001; van den Berg and Friedlander, 2008):

$$\mathbf{BPDN}: \quad \begin{aligned} &\text{minimize} && \|\mathbf{x}\|_1 \\ &\text{subject to} && \|\nabla\mathbf{F}[\mathbf{m}_0,\underline{\mathbf{q}}]\mathbf{C}^*\mathbf{x} - \underline{\mathbf{d}}\|_2 \le \sigma, \end{aligned}$$

where underlined variables means subsampled quantities, i.e., $\underline{\mathbf{d}} = \mathbf{RMd}$, and $\mathbf{RM}$ is the subsampling matrix (see Herrmann and Li, 2012, equation (3) for more details). Matrix $\mathbf{C}^*$ is the Curvelet synthesis operator (Candès et al., 2006), and σ is adjusted to allow for data mismatch due to noise and modelling errors. As the objective function of problem **BPDN** is non-differentiable, **SPG**ℓ_1 implicitly solves a series of LASSO subproblems with gradually relaxed τ's (Tibshirani, 1996; van den Berg and Friedlander, 2008):

$$\mathbf{LASSO}: \quad \begin{aligned} &\text{minimize} && \|\nabla\mathbf{F}[\mathbf{m}_0,\underline{\mathbf{q}}]\mathbf{C}^*\mathbf{x} - \underline{\mathbf{d}}\|_2 \\ &\text{subject to} && \|\mathbf{x}\|_1 \le \tau. \end{aligned}$$

In practice, the input data $\mathbf{d}$ does not necessarily lie in the column space of the linearized modelling operator $\nabla\mathbf{F}$. This can be caused by lack of an accurate background model, or by coherent noises in the data that the linearized modelling fails to explain, such as internal multiples. Therefore, there will be a mismatch between the input data and the predicted data. However, it is rarely discussed in the literature how to choose the "right" σ in this case. Moreover, how the mismatch is going to affect the image quality, especially when we subsample the data, is not well understood. In this article, we carefully design several experiments to examine these issues. We also propose possible solutions if the image quality is compromised under any circumstance. We start with the case where hypothetically we can use all the data (i.e., **RM** is simply a Dirac operator in this case), and then delve into the case where we randomly subsample the data to speed up the inversion. In the end, we verify our observations with a case study using a realistic synthetic model.

THE FULL-DATA CASE

We use a simple two-layer model to examine how the choice of σ in problem **BPDN** affects the inversion. The model is 400m wide and 400m deep, with a five meter grid spacing. The true model $\mathbf{m}$ has two layers. The top layer ranges from the surface to 145m in depth, and has a velocity of 1500m/s. The bottom layer ranges from 145m to 400m in depth, and has a velocity of 2500m/s. We smooth the true model to get the background model $\mathbf{m}_0$, and obtain the perturbation $\delta\mathbf{m}$ by subtracting the true model and the background model. There are 21 sources put at a depth of five meters with 20m lateral spacing, and 81 receivers put at the same depth as the sources with five meter lateral spacing.

We generate two sets of data here. We make the first dataset by linearized modelling (i.e., $\nabla\mathbf{F}[\mathbf{m}_0,\mathbf{q}]\delta\mathbf{m}$), which we refer to as

the *linearized* data. We make the second dataset by first modelling the entire wavefield with the true model (i.e., $\mathbf{F}[\mathbf{m},\mathbf{q}]$ where $\mathbf{F}$ is the forward modelling operator), and then subtracting the direct waves modelled with the smooth background model (i.e., $\mathbf{F}[\mathbf{m}_0,\mathbf{q}]$). We refer to this dataset as the *forward modelling* data. For both datasets, we use a Ricker wavelet of 25Hz peak frequency as the source. From an inversion point of view, the major difference between the two datasets is that the linearized data is perfectly in the column space of the Born scattering operator, while the forward modelling data is not. Traces from the two datasets are plotted in Figure (1(a)). The visual difference between the two datasets is mainly a phase shift, most probably caused by the slightly different kinematics between the true and the background model.

We compare three scenarios here. In the first scenario, we obtain a baseline image by inverting the linearized data. As the data is completely noise-free, we choose $\sigma = 0$ and run for 100 iterations using the **SPG**ℓ_1 solver (van den Berg and Friedlander, 2008). In the second scenario, we use the forward modelling data as the input, and use the *true* σ defined as the ℓ_2 norm of the difference (by subtraction) between the linearized data and the forward modelling data. The third scenario is otherwise the same as the second scenario, except that we simply choose $\sigma = 0$ and let the solver run for 100 iterations.

The inversion results are shown in Figure (1(b)) to (1(d)). We can see that the results of the first and the third scenarios are comparable. The failure of the second scenario indicates that a "wrong" σ was chosen. As we can see from Figure (1(a)), although the phase shift between the two datasets is minimal, the difference (in red) is almost as large as the input data itself. Therefore using the true σ will terminiate the optimization prematurely. As a result, the second scenario finished with only four iterations. In fact, the linearized modelling in the third scenario well fits the forward modelling data (Figure (1(e))). Figure (1(f))) shows that the phase shift between the two datasets (Figure (1(a))) are explained mainly by a spatial shift of the model perturbation. Although the spatial shift is not desirable, it is definitely more favorable than any interfering artifact.

THE SUBSAMPLING CASE

With the same setup as the previous section, we examine how the data mismatch affects the image quality when we subsample the data to speed up the inversion. Instead of using all 21 sequential sources and 61 frequencies, we use 5 simultaneous sources by randomized source superposition, and 15 randomly selected frequencies (see Herrmann and Li, 2012, for details). Based on our observations in the previous section, we choose $\sigma = 0$ and run for 100 iterations.

We first compare two scenarios here. Again, we use linearized data as the input in the first scenario and the forward modelling data in the second scenario. The results are shown in Figure (1(g)) and (1(h)). We can see that the mismatch does result in interfering artifacts when we subsample the data. In other words, the subsampled **BPDN** problem fails to converge to the same solution as the full problem. This is closely related to the calibration problem in compressive sensing (Herman and Strohmer, 2010; Gribonval et al., 2012).

In our previous work, we noticed that introducing *rerandomization* greatly improves the convergence in terms of model error decrease (Tu and Herrmann, 2012). By rerandomization, we mean to draw an independent random subsampling matrix **RM** for each **LASSO** subproblem, by virtue of **SPG**ℓ_1's warm start feature. To maximize the benefit of rerandomization, we redraw both simultaneous sources and randomized frequency subsets (Tu and Herrmann, 2012).

Expecting rerandomization would help to remove the artifacts caused by data mismatch, we setup the third scenario, which is otherwise the same as the second scenario but with rerandomization. The result is shown in Figure (1(i)), which is virtually artifacts free. Although not shown here, we observe that the middle traces of Figure (1(f)) and Figure (1(i)) highly resemble each other. Therefore we argue that rerandomization enables the *subsampled* **BPDN** problem to converge to virtually the same solution as the *full* problem. This also explains the improved convergence observed by Tu and Herrmann (2012).

CASE STUDY

In this section, we verify our previous observations using a more challenging model modified from a 2D slice of the SEG /EAGE salt model. We pad 10 grid points to the water layer of the model for easier removal of direct waves. The model is 3.9km deep and 15.7km wide. We again smooth the true model to obtain the background model. As salt makes a high velocity contrast with the surrounding medium, smoothing a salt structure significantly degrades the accuracy of the background model. We then obtain the true perturbation by subtracting the true and the background model, displayed in Figure (2(b)). We put 323 co-located sources and receivers with 48.8m lateral spacing at a depth of 24.4m. We again generate two sets of data: the *linearized* data and the *forward modelling* data. For both datasets, we use a Ricker wavelet of 5Hz peak frequency as the source and record for 8 seconds. Traces from the two datasets, as well as their difference, are plotted in Figure (2(a)). Different from the two-layer model, the data mismatch here is mainly caused by smoothing the salt structure.

We compare four scenarios here. In the first scenario, we run an RTM with all the data to obtain a baseline image. In the second scenario, we invert from the subsampled linearized data. In the third scenario, we use the subsampled forward modelling data as the input. In the fourth scenario, we again invert from the subsampled forward modelling data, but with rerandomization. In all scenarios, we use 15 frequencies and 15 simultaneous sources. We set the misfit allowance $\sigma = 0$ and run for 100 iterations. With such a setup, the simulation cost is about 1.45X of a single RTM with all the data. We also mute the water layer in all *inversion* scenarios, as the artifacts imaged inside the water layer can adversely affect data fitting.

The results are shown in Figure (2(c)) to (2(f)). The target zone that we are most interested in is below the salt structure. Figure (2(d)) and (2(e)) show that the data mismatch does significantly degrade the image quality, especially in the target zone, when subsampling is adopted. However, by rerandomization ,

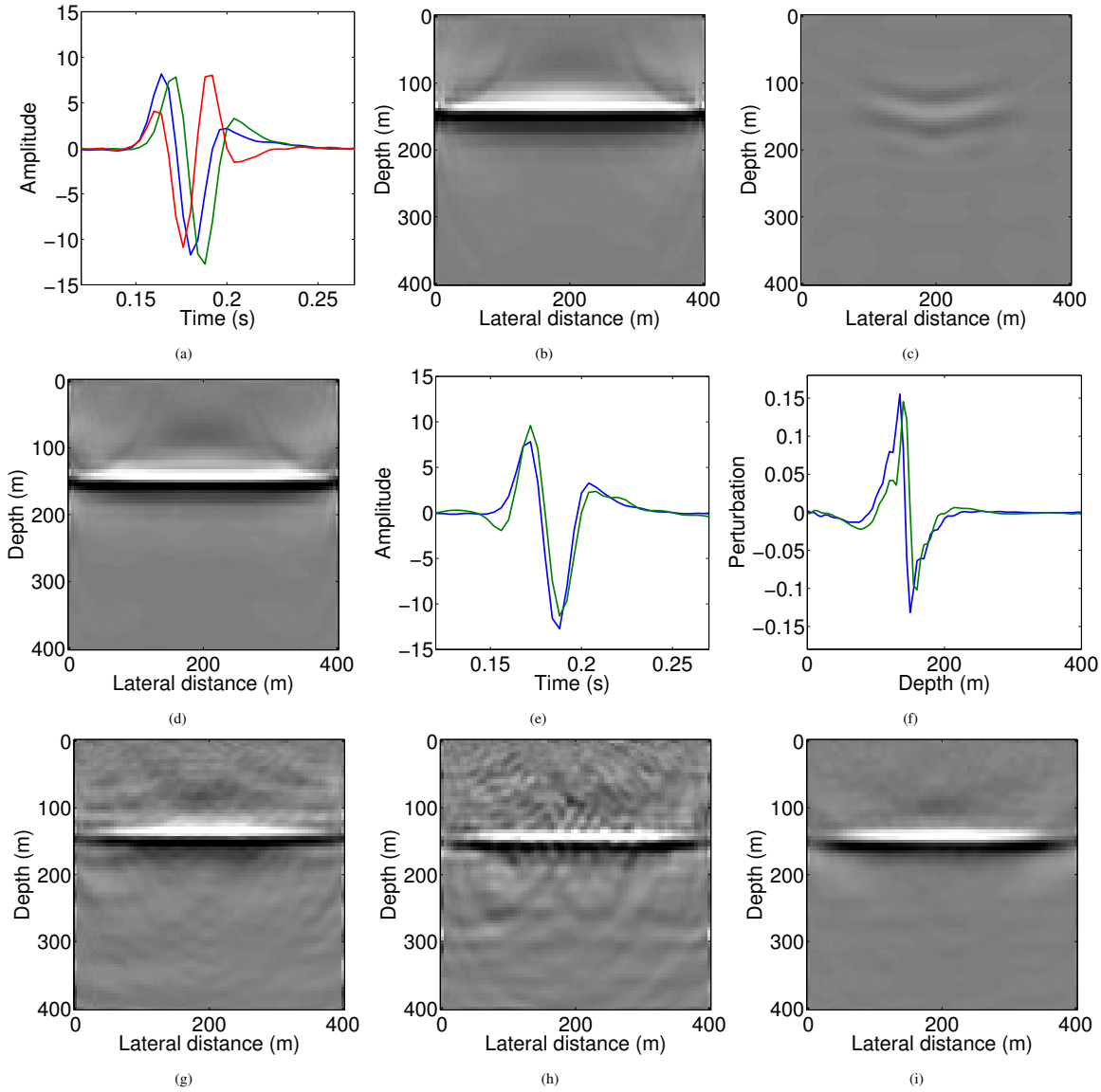

Figure 1: Examples with the two-layer model. The images are plotted with the same color scale. **(a)**. Traces of the linearized data (blue), the forward modelling data (green), and their difference (red). **(b)**. Inversion of the full linearized data. **(c)**. Inversion of the full forward modelling data with the true σ. **(d)**. Inversion of the full forward modelling data with $\sigma = 0$. **(e)**. A comparison of the input forward modelling data (blue) and the linearized modelling data with the inversion results of **(d)** (green). **(f)**. The middle traces of **(b)** and **(d)**. **(g)**. Inversion of the subsampled linearized data, $\sigma = 0$. **(h)**. Inversion of the subsampled forward modelling data, $\sigma = 0$. **(i)**. Otherwise the same as **(h)** but with rerandomization, $\sigma = 0$.

the image quality is greatly improved (Figure (2(f))). We make linearized modelling data with the inversion results both before (Figure (2(e))) and after (Figure (2(f))) taking these measures, and compare them with the input forward modelling data. Traces from these datasets are plotted in Figure (2(g)). We can see that the improvements in the image quality and data fitting are consistent and can validate each other.

We have yet another interesting observation from this example. Compared with linearized inversion, RTM is usually considered to be more robust to a data mismatch. However, this example shows that with linearized inversion, we can better resolve the subsalt structures despite the mismatch. The target zones from Figure (2(c)) and (2(f)) are zoomed and compared in Figure (2(h)).

CONCLUSION

In this abstract, we analyzed the effects of mismatches on ℓ_1-norm regularized inversion related to errors in the linearization. This can lead to major artifacts especially when randomized sources are used to make the ℓ_1-norm regularized inversion computationally feasible. Allowing for a tolerance de-

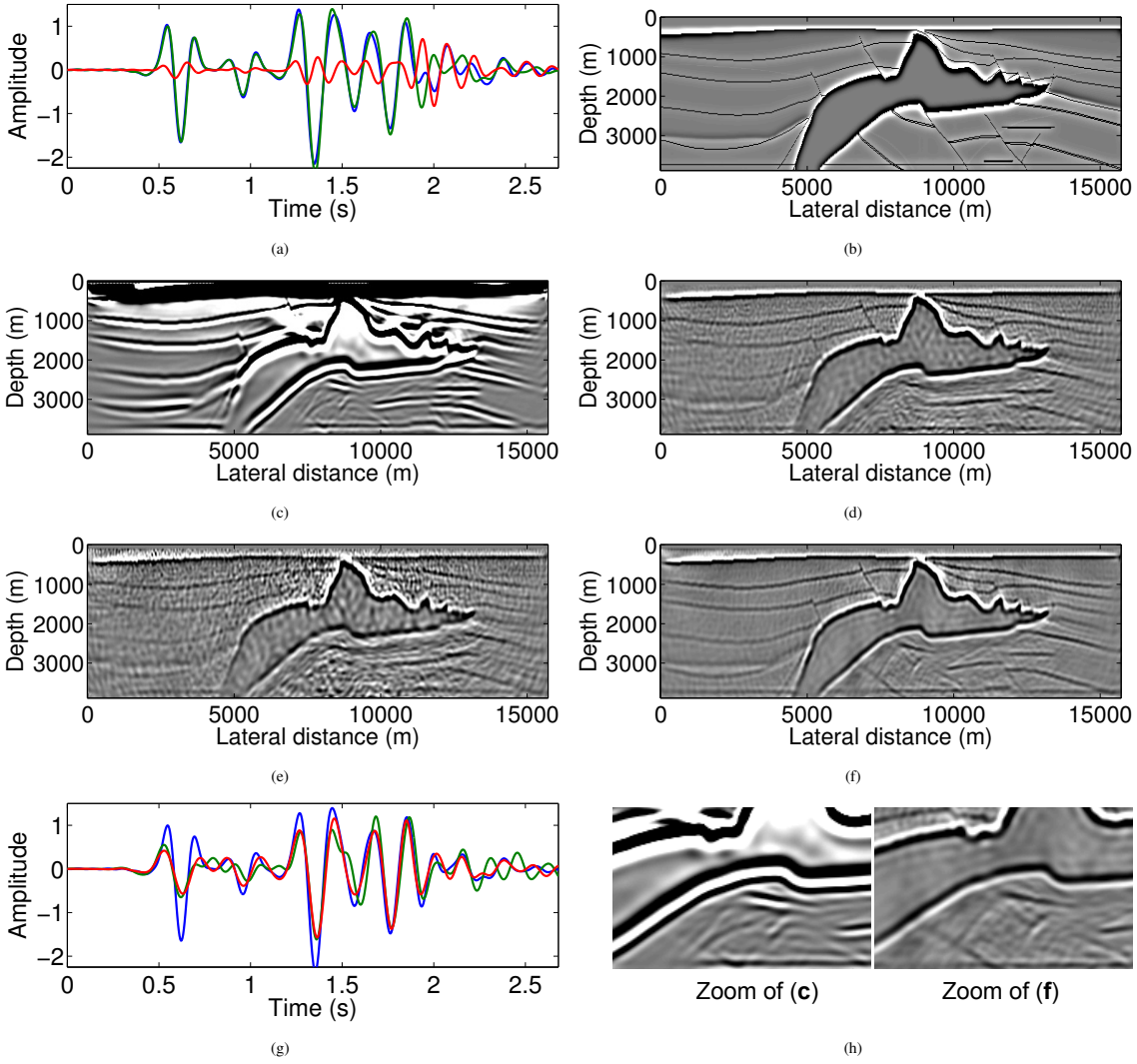

Figure 2: Case study with the synthetic SEG/EAGE 2D salt model. All images are plotted with the same color scale *except* the RTM image. **(a)**. Traces of the linearized data (blue), the forward modelling data (green), and their difference (red). **(b)**. True model perturbation. **(c)**. The RTM image of the forward modelling data. **(d)**. Inversion of the subsampled linearized data. **(e)**. Inversion of the subsampled forward modelling data. **(f)**. Otherwise the same as **(e)** but with rerandomization. **(g)**. Traces of the forward modelling data (blue) and the linearized modelling data with the inversion results of **(e)** (green) and **(f)** (red). **(h)**. Zoomed subsalt areas of **(c)** (left) and **(f)** (right).

pending on the linearization errors makes the situation only worse because the inversion stops prematurely in that case. We overcome this detrimental weakness of ℓ_1-norm regularized inversion by choosing independent subsets of randomized sources during the inversion. Not only does this rerandomization lead to faster convergence but it also removes many of the artifacts related to errors in the linearization. We tested this algorithm on stylized and realistic synthetic examples. Our method produced virtually artifact-free high-resolution images at low computational cost despite errors in the linearized forward model. Contrary to relying on the fold to stack out imaging related errors during reverse-time migration, our method uses redundancy in the data to stabilize the inversion. This allows us to create high-resolution images at a computational cost comparable to the cost of a single reverse-time migration with all data.

ACKNOWLEDGEMENT

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco. Ning Tu and Xiang Li also receive scholarships from the China Scholarship Council.

REFERENCES

Aravkin, A. Y., T. van Leeuwen, and N. Tu, 2012, Sparse seismic imaging using variable projection.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, no. 3, 861–899.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129.

Dai, W., X. Wang, and G. T. Schuster, 2011, Least-squares migration of multisource data with a deblurring filter: Geophysics, **76**, R135–R146.

Gribonval, R., G. Chardon, and L. Daudet, 2012, Blind calibration for compressed sensing by convex optimization: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2713–2716.

Herman, M., and T. Strohmer, 2010, General deviants: An analysis of perturbations in compressed sensing: Selected Topics in Signal Processing, IEEE Journal of, **4**, 342–349.

Herrmann, F. J., and X. Li, 2012, Efficient least-squares imaging with sparsity promotion and compressive sensing: Geophysical Prospecting, **60**, 696–712.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

Tibshirani, R., 1996, Regression shrinkage and selection via the lasso: Journal of the Royal Statistical Society Series B Methodological, **58**, 267–288.

Tu, N., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2013, Fast least-squares migration with multiples and source estimation.

Tu, N., and F. J. Herrmann, 2012, Imaging with multiples accelerated by message passing: Presented at the SEG Technical Program Expanded Abstracts, SEG.

van den Berg, E., and M. P. Friedlander, 2008, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, **31**, 890–912.