

Compressive imaging by wavefield inversion with group sparsity

Felix J. Herrmann*, *UBC-Seismic Laboratory for Imaging and Modeling*

SUMMARY

Migration relies on multi-dimensional correlations between source- and residual wavefields. These multi-dimensional correlations are computationally expensive because they involve operations with explicit and full matrices that contain both wavefields. By leveraging recent insights from compressive sampling, we present an alternative method where linear correlation-based imaging is replaced by imaging via multidimensional deconvolutions of compressibly sampled wavefields. Even though this approach goes at the expense of having to solve a sparsity-promotion recovery program for the image, our wavefield inversion approach has the advantage of reducing the system size in accordance to transform-domain sparsity of the image. Because seismic images also exhibit a focusing of the energy towards zero offset, the compressive-wavefield inversion itself is carried out using a recent extension of one-norm solver technology towards matrix-valued problems. These so-called hybrid (1, 2)-norm solvers allow us to penalize pre-stack energy away from zero offset while exploiting joint sparsity amongst near-offset images. Contrary to earlier work to reduce modeling and imaging costs through random phase-encoded sources, our method compressively samples wavefields in model space. This approach has several advantages amongst which improved system-size reduction, and more flexibility during subsequent inversions for subsurface properties.

INTRODUCTION

With the recent resurgence of full-waveform inversion—i.e., adjoint-state methods applied to solve PDE-constrained optimization problems—the computational cost of computing the model updates has become one of the major impediments withstanding successful application of this technology to industry-size data volumes. To overcome this impediment, we argue that further improvements will depend on a problem formulation with a computational complexity that is no longer strictly determined by the *size* of the discretization but by transform-domain *sparsity* of its solution. In this new paradigm, we bring computational costs in par with our ability to compress solutions of certain PDEs. This premise is related to two recent developments. First, there is the new field of compressive sensing (CS in short throughout the paper, Candès et al., 2006; Donoho, 2006)—where the argument is made, and rigorously proven—that compressible signals can be recovered from severely sub-Nyquist sampling by solving a sparsity promoting program. Second, there is in the seismic community the recent resurgence of simultaneous-source acquisition (Beasley, 2008; Krohn and Neelamani, 2008; Berkhout, 2008; Neelamani et al., 2008; Herrmann et al., 2009), and continuing efforts to reduce the cost of seismic modeling, imaging, and inversion through phase encoding of simultaneous sources (Morton and Ober, 1998; Romero et al., 2000; Krohn and Neelamani, 2008; Neelamani et al., 2008; Herrmann et al., 2009), and the removal of subsets of angular frequencies (Sirgue and Pratt, 2004; Mul-

der and Plessix, 2004; Lin et al., 2008; Herrmann et al., 2009) or plane waves (Vigh and Starr, 2008). All these approaches correspond to instances of CS. By using CS principles, we have been able to remove the associated sub-sampling interferences through a combination of exploiting transform-domain sparsity, properties of certain sub-sampling schemes, and the existence of sparsity promoting solvers. In this paper, we will extend these approaches by applying these principles towards compressively sampling in the model space. This means that we are intervening at the heart of the imaging process. First, we briefly introduce the objectives of full-waveform inversion. Next, we introduce a imaging procedure based on multi-dimensional deconvolutions, which combined with hybrid (1, 2)-norm solvers, leads to a formulation for compressive imaging. We conclude by presenting a proof-of-principle example, followed by a discussion and conclusions.

FULL-WAVEFORM INVERSION

Full-waveform inversion entails solving PDE-constrained optimization problems of the following type:

$$\min_{\mathbf{U}, \mathbf{m}} \frac{1}{2} \|\mathbf{P} - \mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U} = \mathbf{F}, \quad (1)$$

where \mathbf{P} and \mathbf{U} are the observed data volumes and the solution of the multi-source (in its columns)-frequency Helmholtz equation over the domain of interest. \mathbf{D} represents the detection operator that extracts simulated data from time-harmonic solutions at the receiver locations (its adjoint inserts the residue into the co-located sources), \mathbf{H} is a matrix with the discretized multi-frequency Helmholtz equation, and \mathbf{F} is a matrix with the frequency-transformed source distributions in its columns. In the above optimization problem (from which—after casting Eq. 1 in its unconstrained form—most quasi-Newton type full-waveform inversion schemes derive), solutions for the unknown velocity model, \mathbf{m} , and for the wave equation, \mathbf{U} , are pursued that minimize the energy mismatch. Because Eq. 1 is nonlinear in the model variables collected in the vector \mathbf{m} , solutions of Eq. 1 require multiple solves of the (implicit) Helmholtz equation. At the j^{th} depth-level, these solves yield the following updates for the model:

$$\delta \mathbf{m}_j = \sum_{\omega} \omega^2 \Re \left(\text{diag} \left(\mathbf{U}_j \mathbf{V}_j^* \right)_{\omega} \right), \quad (2)$$

where the monochromatic forward- and inversely propagated source-receiver wavefields, \mathbf{U}_j , \mathbf{V}_j are solved from the Helmholtz, and adjoint (denoted by the symbol $*$) Helmholtz systems,

$$\mathbf{H}[\mathbf{m}]\mathbf{U} = \mathbf{F}, \quad \text{and} \quad \mathbf{H}[\mathbf{m}]^* \mathbf{V} = \Delta \mathbf{R}, \quad (3)$$

with $\Delta \mathbf{R} := \mathbf{D}^* (\mathbf{P} - \mathbf{D}\mathbf{U})$ the multi-shot residual wavefield. For later extension to pre-stack imaging, we deliberately formulated imaging in Eq. 2 in terms of matrix multiplications amongst the source- and residual wavefields at depth level j . In that case, the zero-offset image resides on the diagonal and is equivalent to summing the Hadamard products (elementwise products)

Imaging by wavefield inversion

between \mathbf{U} , and the complex conjugate of \mathbf{V} over all shots and frequencies (Plessix, 2006). (To simplify notation, I will omit frequency and depth dependence.)

At this point, the following observations are in order. First, the updates in Equation 2 correspond, up to some constants, to 'post-stack' migration (see e.g. Plessix, 2006), where the zero-offset image is extracted from the diagonal after matrix-matrix multiplication. Second, the above migration procedure can be generalized to prestack migration by dropping the zero-offset diagonal extraction. This allows for the incorporation of a focusing procedure during which the energy in the off-diagonals—i.e., the non-zero-offset imaged wavefields—is minimized through a differential-semblance-like optimization procedure (see e.g. Shen and Symes, 2008). Third, the creation of the image suffers from amplitude deterioration and requires, when implemented naively, explicit storage of both wavefields. By turning above correlation-based imaging into a focused wavefield inversion procedure for compressively-sampled data (read image-space blended data), we address these computational and amplitude problems. First, we introduce focused prestack imaging through wavefield inversion, followed by compression according to compressive sensing. We conclude by introducing a method based on joint sparsity promotion.

IMAGING VIA FOCUSED-WAVEFIELD INVERSION

Following earlier work by Claerbout (1971); Berkhout (1982); Shen and Symes (2008); Misra and Sacchi (2008), and many others, Equation 2 can for each angular frequency be recast into the following 'pre-stack' form:

$$\mathbf{I} = \mathbf{T}_{(x_s, x_r) \rightarrow (m, h)}^{\Delta h} (\mathbf{U}\mathbf{V}^*), \quad (4)$$

with $\mathbf{T}_{(x_s, x_r) \rightarrow (m, h)}^{\Delta h}$ a linear operator that maps source-receiver coordinates to midpoint ($m = \frac{1}{2}(x_s + x_r)$)/half-offset ($h = \frac{1}{2}(x_s - x_r)$) coordinates for offsets on the interval $h \in [-\Delta h, +\Delta h]$. For now, we assume that the source-receiver positions lie in the same horizontal plane for each depth level.

Instead of calculating the image through multi-dimensional correlation, we propose to image by inverting the adjoint of the wavefield \mathbf{U} with respect to the wavefield \mathbf{V}^* . This type of wavefield inversion has been applied successfully before, a finding reported widely in the literature with applications that range from interferometric deconvolution, to missing-trace interpolation with the focal transform Herrmann et al. (2008), curvelet-based multiple and primary prediction. (For the latter, refer to another contribution by the first author to the proceedings of this conference.) As shown by Herrmann and Wang (2008), these wavefield inversions benefit greatly from transform-domain sparsity promotion. Transform-domain sparsity, however, is not the only aspect that we can exploit towards our advantage. The fact that imaging is based on focusing can also be used by penalizing energy residing in the non-zero offsets, followed by extracting the zero-offset trace—i.e., $\delta \mathbf{m} = \mathbf{I}(\cdot, h = 0)$ —after the inversion is completed.

We now arrive at the mathematical formulation of the imaging problem through focused wavefield inversion. This formulation

involves the inversion of the following system:

$$\mathbf{U}^* \mathbf{S}^* \mathbf{X} \approx \mathbf{V}^* \quad (5)$$

$$\mathbf{H}\mathbf{X} \approx \mathbf{0}, \quad (6)$$

or $\mathbf{A}\mathbf{X} \approx \mathbf{B}$. Here, \mathbf{H} is a linear operator designed to penalize the non-zero offsets—, i.e. $\mathbf{I}(\cdot, h)$ for $h \neq 0$. The synthesis operator \mathbf{S}^* brings data back from the sparsifying domain to the near offsets, including a conversion back to source-receiver coordinates (with \mathbf{T}^*). Finally, the matrix $\mathbf{0}$ contains zeros, and the symbol \approx indicates that the above equations hold in the least-squares sense. Focusing is accomplished by multiplying the images with a function that increases (linearly) with distance with respect to the zero-offset coordinate (Shen and Symes, 2008). (Notice this argument also applies to focusing towards zero time.) With the linearized 'forward' model, we create a focused image by inverting the adjoint of the wavefield \mathbf{U} , compounded with the synthesis matrix \mathbf{S}^* . Focusing is accomplished through minimization of the energy functional $\|\mathbf{H}\mathbf{I}\|_2$.

Recovery from incomplete compressively sampled data involves the solution of the following sparsity-promotion program (Berg and Friedlander, 2008)

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \sigma \quad (7)$$

Here, compressively-sampled data is given by $\mathbf{b} = \mathbf{A}\mathbf{x}_0$ with \mathbf{x}_0 sparse (e.g., a transform-domain vector with a limited number of large entries). The matrix $\mathbf{A} = \mathbf{R}\mathbf{M}\mathbf{S}^*$ is made out of matrices that, respectively, subsample, phase encode, and sparsify (Herrmann et al., 2009). The above optimization problems seeks a vector with the smallest ℓ_1 -norm provided data is fitted to within an energy σ . This constant depends on the noise level or on the level to which the forward model explains data, which in our case represents the residual wavefield.

The key contribution of CS lies in the fact that the above sparsity promoting program is able to recover sparse vectors \mathbf{x}_0 from data sub sampled at rates that are proportional to the sparsity-level of \mathbf{x}_0 . Even though this framework has proven to be powerful, the recovery until recently was not able to benefit from situations where multiple compressive measurements are taken of a series of objects that have similar transform-domain supports. In that case, transform coefficients are likely to be large at the same entries. This situation occurs when images align for the near offsets. Following recent work by Berg and Friedlander (2008), recovery of objects from multiple measurement vectors (collected in the matrix \mathbf{B}), can be formulated as follows

$$\tilde{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{X}\|_{1,2} \quad \text{subject to} \quad \|\mathbf{A}\mathbf{X} - \mathbf{B}\|_{2,2} \leq \sigma, \quad (8)$$

with the mixed (1,2)-norm defined as

$$\|\mathbf{X}\|_{1,2} := \sum_{i \in \text{rows}(\mathbf{X})} \|\text{row}_i(\mathbf{X})\|_2. \quad (9)$$

The (2,2)-norm for the residue is defined as

$$\|\mathbf{X}\|_{2,2} := \left(\sum_{i \in \text{rows}(\mathbf{X})} \|\text{row}_i(\mathbf{X})\|_2^2 \right)^{\frac{1}{2}}. \quad (10)$$

Imaging by wavefield inversion

The sparsity-promoting program in Eq. 8 differs from Eq. 7 because it seeks to find multiple sparse vectors from multiple measurements while asserting minimal energy amongst the rows. This penalty assures 'lateral' continuity amongst the recovered sparse vectors. We will use this joint-sparsity promotion for our focused-imaging procedure.

COMPRESSIVE IMAGING

Solutions of Eq. 1 require multiple solves of the (implicit) Helmholtz system. This may prove computationally prohibitive, even after preconditioning (with a complexity of $\mathcal{O}(n^4)$ in 2-D (Erlangga and Nabben, 2007; Erlangga and Herrmann, 2008)). Storage of the wavefields is also problematic. We address these issues by using CS (Romberg, 2008; Herrmann et al., 2009) to reduce the size of these wavefields. We accomplish this by extending earlier work (Herrmann et al., 2009)—where sources and angular frequencies in the data space are compressively sampled—towards a compressive sampling of the source- and residual wavefields in the model space:

$$\mathbf{RM} := \overbrace{[\mathbf{R}^\sigma \otimes \mathbf{R}^\rho \otimes \mathbf{R}^\zeta]}^{\text{sub sampler}} \overbrace{(\mathbf{F}_3^* \text{diag}(e^{i\theta})) \mathbf{F}_3}^{\text{random phase encoder}},$$

with \mathbf{F}_3 the 3-D Fourier transforms, and $\theta = \text{Uniform}([0, 2\pi])$ a random phase rotation. The matrices \mathbf{R}^σ , \mathbf{R}^ρ , and \mathbf{R}^ζ represent CS-subsampling matrices acting along the depth, receiver, and source coordinates. The subsampling operators reduce the system size, i.e., $n_\sigma \times n_\rho \times n_\zeta \ll n_s \times n_r \times n_z$ with n_s , n_r , n_z the number of sources, receivers, and depthlevels, respectively. As shown by Herrmann et al. (2009), application of this type of CS-sampling matrix, \mathbf{RM} , is in this case equivalent to a spatial blending along the source-receiver coordinates, and along the depth coordinate. To maximize randomization during the CS sampling, we choose a different \mathbf{RM} for each frequency.

Aside from proper CS sampling, imaging from compressively-sampled wavefields depends on a sparsifying transform that compresses seismic images, is fast, and reasonably incoherent with the CS sampling matrix. We accomplish this by defining the sparsity transform as the Kronecker product between the 2-D discrete curvelet transform (Candès et al., 2006) along the midpoint-depth coordinates, and the Dirac basis along the offset coordinate—i.e., $\mathbf{S} := \text{vec}^{-1}((\mathbf{Id} \otimes \mathbf{C}) \mathbf{T}_0) \text{vec}(\cdot)$ with \mathbf{C} , \mathbf{Id} the curvelet synthesis and Dirac matrices, respectively. \mathbf{T}_0 represents the mapping defined in Eq. 4, supplemented with the imaging condition at $t = 0$. The functions vec and vec^{-1} reorganize a matrix into a long vector, and vice versa. To further simplify notation, I drop these from the equations.

We are now in the position to image from compressively-sampled wavefields. This is done by applying the CS-sub-sample matrix to Eq. 5, yielding a reduced system: $\mathbf{RMU}^* \mathbf{S}^* \mathbf{X} \approx \mathbf{RMV}^*$. For the remainder of the paper, we overload the definition for the original system, $\mathbf{AX} \approx \mathbf{B}$, with its compressively sampled counterpart. By solving the joint sparsity-promoting program of Eq. 8, followed by extraction of the diagonal from $\delta \tilde{\mathbf{m}} = \Re(\text{diag}(\mathbf{S}^* \tilde{\mathbf{X}}))$, we obtain an estimate for the image. We solve this problem with SPGL₁ (Berg and Friedlander, 2008), a projected-gradient algorithm based on root finding.

EXAMPLE

We consider an example with $n = n_t = n_s = n_r = n_z = 128$ for 8 frequencies selected randomly from the band 20 – 100Hz, and the first 8 near offsets. The background model \mathbf{m}_0 and perturbation $\delta \mathbf{m} := \mathbf{m} - \mathbf{m}_0$ are plotted in Fig. 1(a)-1(b). The wavefields are computed with Eq. 3. Next, *linearized* data is computed (Eq. 5), followed by a compressive sampling for each frequency separately, yielding a system of 1/64 its original size. Migrated images and selected common-image gathers for the migrated (by multi-dimensional correlation) and focused wavefield inversion are plotted in Fig.'s 1(c)-1(f). Even though we used sub-optimal wavelets instead of curvelets as the sparsity transform (to save memory usage in this prototype implemented using SPARCO (van den Berg et al., 2007).), we were able to get a good recovery from a substantial degree of subsampling.

DISCUSSION & CONCLUSIONS

The key contributions are: (i) Focused image recovery through mixed (1, 2)-norm minimization during which transform-domain sparsity is promoted for each offset in conjunction with a ℓ_2 penalty term that promotes joint sparsity amongst these images. Consequently, energy amongst the larger offsets is penalized, yielding a focusing of the energy. (ii) A reduction in modeling costs with some computational overhead related to the solution of the sparsity promoting program (see Herrmann et al., 2009, for details). Because, the concept of compressive sampling is carried a step further compared to earlier work (Herrmann et al., 2009), we anticipate additional performance improvements by combining focusing with model-space sparsity promotion. (iii) A reduction in storage requirements. The more compressible the image, the more these wavefields can be sub sampled. This opens the possibility of storing these matrices explicitly. (iv) A reduction in the wavefield-inversion cost by virtue of the reduced system size. (v) Finally, compressive sampling in the model space has the additional advantage that it offers more flexibility during the subsequent imaging/inversion step. For instance, extremely high degrees of CS subsampling can be envisaged for velocity-model estimation where the unknown velocity model permits an extremely low-dimensional parameterization by splines. Conversely, high-resolution migration will require more compressive samples.

In summary, by combining sparsity promotion and focusing towards the near offsets, a new framework for pre-stack migration and inversion is presented. In this framework, model updates are obtained by multi-dimensional deconvolutions of the source wavefield with respect to the residual wavefield. This approach has the advantage that it allows for a recovery from compressively sampled wavefields. This leads to system-size reduction with sub-sampling rates, and hence modeling, and storage costs, that can be chosen in par with the sparsity of the object of interest. Finally, we envisage an integration of these ideas into our formulation for full-waveform inversion based on our Gauss-Newton-Krylov method reported elsewhere in the proceedings of this conference.

ACKNOWLEDGMENTS

FJH. would like to thank Martin Lacasse for insightful discussions and TU Delft for hosting part of FJH's sabbatical. This work was in part supported by NSERC Discovery Grant (22R81254) and CRD Grant DNOISE (334810-05).

Imaging by wavefield inversion

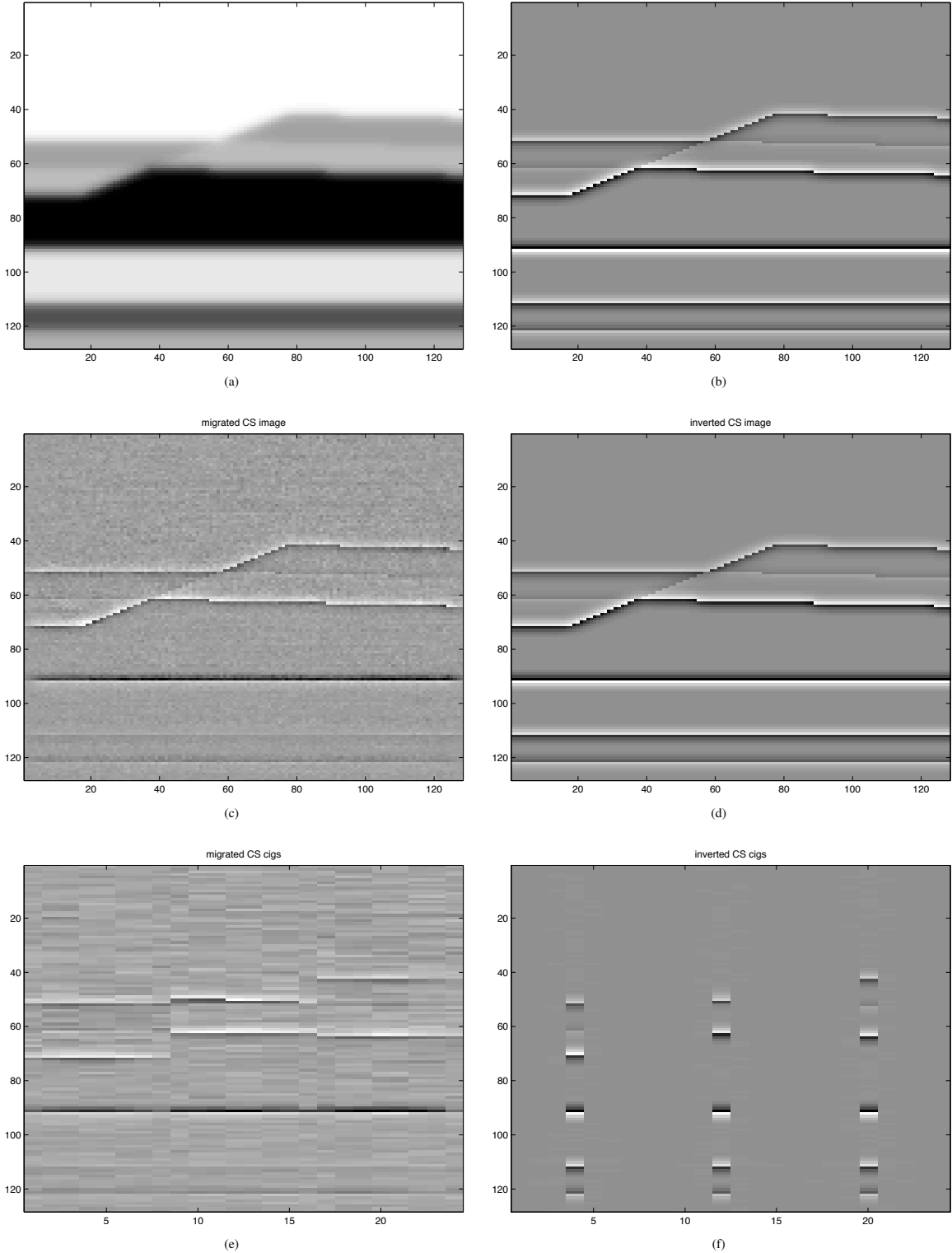


Figure 1: Proof of principle for the recovery of the image from compressively sampled wavefields by joint sparsity promotion. **(a)** The background velocity model, \mathbf{m}_0 . **(b)** The perturbation $\delta\mathbf{m} := \mathbf{m} - \mathbf{m}_0$. **(c)** “Noisy” migrated image according to Eq. 2 for compressively-sampled *linearized* data (cf. Eq. 5). **(d)** Image obtained by wavefield inversion via joint sparsity promotion—i.e., $\delta\tilde{\mathbf{m}} = \Re\left(\text{diag}\left(\mathbf{S}^*\tilde{\mathbf{X}}\right)\right)$. **(e)** Three selected common-image gathers for the migrated image as a function depth and the subsurface half offset. **(f)** The same but after wavefield inversion. These results clearly show that subsampling according to the principles of CS leads to noisy interference artifacts. Our formulation based on wavefield inversion removes these incoherent artifacts and leads to a substantial improvement in the recovery. This improvement can be attributed to sparsity promotion combined with focusing. This focusing can clearly be observed in the common-image gathers for the result obtained by wavefield inversion.

Imaging by wavefield inversion

REFERENCES

- Beasley, C. J., 2008, A new look at marine simultaneous sources: *The Leading Edge*, **27**, 914–917.
- Berg, E. v. and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: Technical Report 2, Department of Computer Science, University of British Columbia.
- Berkhout, A. J., 1982, *Seismic migration. Imaging of acoustic energy by wave field extrapolation*: Elsevier, Amsterdam.
- , 2008, Changing the mindset in seismic data acquisition: *The Leading Edge*, **27**, 924–938.
- Candès, E., J. Romberg, and T. Tao, 2006, Stable signal recovery from incomplete and inaccurate measurements: *Communications on Pure and Applied Mathematics*, **59**, 1207–1223.
- Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: *Multiscale Modeling and Simulation*, **5**, 861–899.
- Clairbout, J. F., 1971, Toward a unified theory of reflector mapping: *Geophysics*, **36**, 467–481.
- Donoho, D. L., 2006, Compressed sensing: *IEEE Transactions on Information Theory*, **52**, 1289–1306.
- Erlangga, Y. A. and F. J. Herrmann, 2008, An iterative multilevel method for computing wavefields in frequency-domain seismic inversion: *SEG Technical Program Expanded Abstracts*, 1957–1960, SEG.
- Erlangga, Y. A. and R. Nabben, 2007, On multilevel projection Krylov method for the preconditioned Helmholtz system. Submitted for publication.
- Herrmann, F. J., Y. A. Erlangga, and T. T. Lin, 2009, Compressive simultaneous full-waveform simulation. TR-2008-09, to appear as a letter in *geophysics*, august, 2009.
- Herrmann, F. J. and D. Wang, 2008, Seismic wavefield inversion with curvelet-domain sparsity promotion: *SEG Technical Program Expanded Abstracts*, **27**, 2497–2501.
- Herrmann, F. J., D. Wang, G. Hennenfent, and P. P. Moghaddam, 2008, Curvelet-based seismic data processing: a multiscale and nonlinear approach: *Geophysics*, **73**, no. 1, A1–A5.
- Krohn, C. and R. Neelamani, 2008, Simultaneous sourcing without compromise: Rome 2008, 70th EAGE Conference & Exhibition, B008.
- Lin, T., E. Lebed, Y. A. Erlangga, and F. J. Herrmann, 2008, Interpolating solutions of the helmholtz equation with compressed sensing: *SEG Technical Program Expanded Abstracts*, 2122–2126, SEG.
- Misra, S. and M. D. Sacchi, 2008, Global optimization with model-space preconditioning: Application to avo inversion: *Geophysics*, **73**, R71–R82.
- Morton, S. A. and C. C. Ober, 1998, Faster shot-record depth migrations using phase encoding: *SEG Technical Program Expanded Abstracts*, 1131–1134, SEG.
- Mulder, W. and R. Plessix, 2004, How to choose a subset of frequencies in frequency-domain finite-difference migration: *Geophysical Journal International*, **158**, 801–812.
- Neelamani, N., C. Krohn, J. Krebs, M. Deffenbaugh, and J. Romberg, 2008, Efficient seismic forward modeling using simultaneous random sources and sparsity: *SEG International Exposition and 78th Annual Meeting*, 2107–2110.
- Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503.
- Romberg, J., 2008, Compressive sensing by random convolution: submitted. Preprint available at http://users.ece.gatech.edu/~justin/Publications_files/RandomConvolution.pdf.
- Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: *Geophysics*, **65**, no. 2, 426–436.
- Shen, P. and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: *Geophysics*, **73**, VE49–VE59.
- Sirgue, L. and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248.
- van den Berg, E., M. P. Friedlander, G. Hennenfent, F. J. Herrmann, R. Saab, and O. Yilmaz, 2007, Sparco: a testing framework for sparse reconstruction: Technical Report TR-2007-20, UBC Computer Science Department.
- Vigh, D. and E. W. Starr, 2008, 3d prestack plane-wave, full-waveform inversion: *Geophysics*, **73**, VE135–VE144.