Introduction

With the recent resurgence of full-waveform inversion, the computational cost of solving forward modeling problems has become—aside from issues with non-uniqueness—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 in this area 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 seismic data and models. This premise is based on two recent developments, namely 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 samplings by solving a sparsity promoting program, and the recent resurgence of simultaneous sources (Beasley, 2008; Berkhout, 2008; Krebs et al., 2009; Herrmann et al., 2009; Herrmann, 2009). In this abstract, we focuss on how simultaneous sources can be used to reduce the computational complexity of migration. The presented technique follows in the footsteps of attempts towards cost reductions for the computation of gradients, which are central to imaging and inversion, through phase encoded simultaneous sources (Romero et al., 2000; Herrmann et al., 2009), possibly in combination with the removal of subsets of angular frequencies (Sirgue and Pratt, 2004; Mulder and Plessix, 2004; Lin et al., 2008; Herrmann et al., 2009). Our contribution to earlier work is twofold. First, we show that randomly phase-encoded sources harness interferences by turning them into incoherent noise. We confirm predictions stating that the larger the degree of under sampling (read the smaller the number of frequency-restricted simultaneous source experiments), the more energy leaks towards noisy interferences. Second, we show that we can remove these noise artifacts and restore the amplitudes by solving a sparsity promoting program. The solution of this program corresponds to one-norm regularized linearized inversion from frequency-restricted simultaneous data.

Dimensionality reduction by compressive sampling

Full-waveform inversion (FWI) involves the solution of the following multi-experiment unconstrained optimization problem: $\min_{\mathbf{m}} \frac{1}{2}\|\mathbf{P} - \boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{Q}]\|_{2,2}^2$, where each column of $\mathbf{P}$ contains the observed data for one shot and all frequencies. The nonlinear operator $\boldsymbol{\mathcal{F}}[\mathbf{m},\mathbf{Q}] = \mathbf{D}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{Q}$ simulates data by solving the Helmholtz system $\mathbf{H}$ for all sources in the columns of $\mathbf{Q}$. We restrict the simulated data to the receiver positions with the detection operator $\mathbf{D}$ to obtain observed data. For simplicity, we assume that the sources are known and co-located with the receivers. We also neglect surface-related multiples by by using an absorbing boundary condition at the surface. Unfortunately, the solution of the above minimization problem with (quasi) Newton techniques is extremely costly because each iteration requires the solution of the forward and time-reversed (adjoint) Helmholtz system for each of the n_f frequencies and for each of the n_s sources. Moreover, improvements in convergence of these schemes require expensive inversions of the reduced Hessian (see e.g. Erlangga and Herrmann, 2009, and the references therein). Here, we address these problems by combining dimensionality-reduction strategies with recovery based on sparsity promotion. We reduce the data volume and hence the size of the Helmholtz system by applying the following randomized block-diagonal sampling operator (Herrmann et al., 2009):

$$\mathbf{RM} = \mathrm{vec}^{-1}\mathrm{blockdiag}\left[(\mathbf{RM})_1 \cdots (\mathbf{RM})_{n_s'}\right]\mathrm{vec}, \tag{1}$$

with $(\mathbf{RM})_{1\cdots n_s'} := \left(\mathbf{R}^{\boldsymbol{\Sigma}}{}_{1\cdots n_s'}\mathbf{F}_{\boldsymbol{\Sigma}}^{*}\mathrm{diag}\left(e^{\hat{i}\boldsymbol{\theta}_{1\cdots n_s'}}\right)\mathbf{F}_{\boldsymbol{\Sigma}} \otimes \mathbf{I} \otimes \mathbf{R}^{\boldsymbol{\Omega}}{}_{1\cdots n_s'}\mathbf{F}_{\boldsymbol{\Omega}}\right)$, with $\mathbf{R}^{\boldsymbol{\Sigma},\boldsymbol{\Omega}}{}_{1\cdots n_s'}$ the randomized restriction operators that for each source experiment randomly sample along the source (Σ) and frequency (Ω) axes. The $\mathbf{F}_{\boldsymbol{\Sigma},\boldsymbol{\Omega}}$ represent the Fourier transforms with respect to these two axes. To allow for a formulation with Kronecker products, we sandwiched the above expression between the operator vec that maps a matrix columnwise to a vector (matlab's colon operator), and its inverse vec^{-1}. (For notational convenience, we will drop this detail in the remaining equations.) The block diagonals correspond to $n_s' \ll n_s$ different simultaneous source experiments that are phase encoded by the random phases $\boldsymbol{\theta}_{1\cdots n_s'} = \mathrm{Uniform}([0,2\pi])$ and sampled by different realizations for $\mathbf{R}^{\Sigma}_{1\cdots n_s'}$. The fre-

quencies in each experiment are also randomly selected by the restriction matrix $\mathbf{R}^{\Omega}_{1\cdots n'_s}$. After applying this sampling operator, we obtain $\min_{\mathbf{m}} \frac{1}{2}\|\underline{\mathbf{P}} - \underline{\boldsymbol{\mathcal{F}}}[\mathbf{m}, \underline{\mathbf{Q}}]\|^2_{2,2}$, where the underbarred quantities refer to compressively sampled wavefields—i.e., $\mathbf{RM}\left(\mathbf{P} - \boldsymbol{\mathcal{F}}[\mathbf{m}, \mathbf{Q}]\right) = \underline{\mathbf{P}} - \underline{\boldsymbol{\mathcal{F}}}[\mathbf{m}, \underline{\mathbf{Q}}]$. After this subsampling, FWI involves solutions of the dimensionality reduced Helmholtz system over $n'_s \ll n_s$ sources with $n'_f \ll n_f$ frequencies.

Gradient updates from simultaneous sources

FWI typically involves model updates $\mathbf{m} \leftarrow \mathbf{m} + \delta\widetilde{\mathbf{m}}$ with $\delta\widetilde{\mathbf{m}} = \Re\left(\sum_\omega \omega^2 \sum_s (\bar{\mathbf{u}} \odot \mathbf{v})_{s,\omega}\right) = \mathbf{K}^*[\mathbf{m}, \mathbf{Q}]\delta\mathbf{P}$, where $\mathbf{u}, \mathbf{v}$ are the forward modeled source and reverse-time modeled residual wavefields. $\delta\mathbf{P} = (\mathbf{P} - \boldsymbol{\mathcal{F}}[\mathbf{m}, \mathbf{Q}])$ is the residual wavefield at the surface. Grosso modo, these gradient updates correspond to migrations given by the adjoint of the linearized multi-experiment Born scattering operator $\mathbf{K}[\mathbf{m}, \mathbf{Q}]$. (Again, similar expressions hold for the reduced system where impulsive sources are replaced by simultaneous sources and where the two sums are only evaluated over random subsets of the source and frequency experiments.) To illustrate how simultaneous sources may affect migrated images ($\delta\widetilde{\mathbf{m}}$), we simulate for all frequencies linearized data for a single impulsive shot and for a single simultaneous shot. This is an instance of one block of $\mathbf{RM}$. From Fig 1 (a-b), we can clearly see that the migrated image from the simultaneous-source experiment is better resolved then the image of a single sequential shot. However, there is clearly leakage from the image towards incoherent noise-like artifacts.

Recovery of gradient updates by sparsity promotion

Aside from interference artifacts, gradient updates do not resolve the amplitudes correctly and this leads to slow convergence. Instead of shifting towards a formulation that includes updates that involve the inversion of the reduced Hessian (as reported in Erlangga and Herrmann (2009)), we choose a formulation where we invert the adjoint of the migration operator—i.e., the linearized Born scattering operator. As common in approaches with reduced Hessians, we ignore contributions from internal multiples in the updates. Because the problem is now linear, we can use the techniques from compressive sensing to recover the updates by carrying out one-norm regularized linearized inversions—i.e., $\delta\widetilde{\mathbf{m}} = \mathbf{S}^H\delta\widetilde{\mathbf{x}}$ with $\delta\widetilde{\mathbf{x}} = \arg\min_{\delta\mathbf{x}} \|\delta\mathbf{x}\|_{\ell_1}$ subject to $\delta\underline{\mathbf{P}} = \mathbf{A}\delta\mathbf{x}$, where $\mathbf{A} = \mathbf{RMKS}^* = \underline{\mathbf{K}}\mathbf{S}^*$, $\delta\underline{\mathbf{P}} = \mathbf{RM}\delta\mathbf{P}$, and $\underline{\mathbf{K}} = \underline{\mathbf{K}}[\mathbf{m}, \underline{\mathbf{Q}}]$ and where $\mathbf{S}^H$ is the adjoint of some sparsifying transform (e.g. curvelets or wavelets). By solving these sparsity-promoting programs for each model update (see Alg. 1, where we solve FWI upto an error ϵ), we end up with an algorithm that will be competitive with reduced Hessian updates and that offers sufficient flexibility to implement frequency continuation by choosing the $\mathbf{R}^{\Omega}$'s appropriately. Each subproblem is solved with SPGℓ_1 van den Berg and Friedlander (2008).

Example

To illustrate the perfomance of our algorithms, we study the behavior of a single gradient by conducting a series of experiments where we vary the subsampling ratio—i.e., the aspect ratio of RM—and the frequency-to-shot subsampling ratio. All simulations are carried with 256 collocated shot and receiver positions sampled at a 29 m interval. The time sample interval is $0.016\,\mathrm{s}$. Comparison between the gradient updates and the updates by sparsity promotion shows remarkable high-fidelity results even for increasing subsampling ratios. As expected, the numbers in Table 1 confirm increasing recovery errors for increasing subsampling ratios. For fixed subsampling ratios, however, we observe improved results for decreasing frequency-to-shot ratios, which suggests that simultaneous shots contribute more to the solution. These trends also hold for migrated images—i.e., standard gradient updates. Because the speedups are proportional to subsampling ratios, we can conclude that the dimensionality reductions offset costs of the ℓ_1 recovery approximately.

Conclusions

Dimensionality-reduction strategies will be instrumental for the success of FWI. These strategies are build on the premise that whenever models exhibit structure (read transform-domain sparsity), these models can be reconstructed from randomized subsampling that are proportional to transform-domain

sparsity. Because these approaches remove the "curse of dimensionality", which is one of the major impediments of FWI, this contribution may lead to a paradigm shift where Newton-type updates, which otherwise would have to be computed over all frequencies and all sequential sources, can now be computed by inverting the dimensionality-reduced linearized Born scattering operator. This result will help bring FWI into the main stream of seismic data processing.

Acknowledgments

FJH. would like to thank Yogi Erlangga for insightful discussions. This work was financed through the SINBAD II project with support from BG, BP, Chevron, Petrobras, and WesternGeco and was partly supported by a NSERC Discovery Grant (22R81254) and CRD Grant DNOISE (334810-05).

Algorithm 1: FWI by repeated one-norm minimization.

Result: Estimate for the model $\widetilde{\mathbf{m}}$

$\mathbf{m} \longleftarrow \mathbf{m}_0$; // initial model

while $\|\underline{\mathbf{P}} - \underline{\boldsymbol{\mathcal{F}}}[\mathbf{m}, \underline{\mathbf{Q}}]\|_{2,2}^2 \geq \epsilon$ **do**

$\mathbf{A} \longleftarrow \underline{\mathbf{K}}[\mathbf{m}, \underline{\mathbf{Q}}]\mathbf{S}^H$; // compute linearized system

$\delta\widetilde{\mathbf{x}} \longleftarrow \arg\min_{\delta\mathbf{x}} \|\delta\mathbf{x}\|_{\ell_1}$ subject to $\delta\underline{\mathbf{P}} = \mathbf{A}\delta\mathbf{x}$; // do the recovery

$\mathbf{m} \longleftarrow \mathbf{m} + \mathbf{S}^H t\delta\widetilde{\mathbf{x}}$; // compute model updates

end

Subsample ratio	0.015	0.006	0.002
n'_f/n'_s	recovery error (dB)		
5	**17.44** (1.32)	**11.66** (0.78)	**6.83** (-0.14)
1	**17.53** (1.59)	**11.89** (1.05)	**7.19** (0.15)
0.2	**18.22** (1.68)	**12.11** (1.32)	**7.46** (0.27)
Speed up ($\times$)	66	166	500

Table 1 *Signal-to-noise ratios, SNR* $= -20\log_{10}(\frac{\|\delta\mathbf{m}-\delta\widetilde{\mathbf{m}}\|_2}{\|\delta\mathbf{m}\|_2})$ *for reconstructions with the wavelet sparsity transform for different subsample and frequency-to-shot ratios. SNRs for* ℓ_1 *are in bold and SNRS for migration are in parentheses.*

References

Beasley, C.J. [2008] A new look at marine simultaneous sources. *The Leading Edge*, **27**(7), 914–917, doi:10.1190/1.2954033.

Berkhout, A.J. [2008] Changing the mindset in seismic data acquisition. *The Leading Edge*, **27**(7), 924–938, ISSN 7, doi: 10.1190/1.2954035.

Candès, E., Romberg, J. and Tao, T. [2006] Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, **59**(8), 1207–1223.

Donoho, D.L. [2006] Compressed sensing. *IEEE Transactions on Information Theory*, **52**(4), 1289–1306.

Erlangga, Y.A. and Herrmann, F.J. [2009] Seismic waveform inversion with Gauss-Newton-Krylov method. *SEG Technical Program Expanded Abstracts*, SEG, vol. 28, 2357–2361.

Herrmann, F.J. [2009] Compressive imaging by wavefield inversion with group sparsity. *SEG Technical Program Expanded Abstracts*, SEG, SEG, vol. 28, 2337–2341.

Herrmann, F.J., Erlangga, Y.A. and Lin, T. [2009] Compressive simultaneous full-waveform simulation. *Geophysics*, **74**, A35.

Krebs, J.R. et al. [2009] Fast full wave seismic inversion using source encoding. SEG, vol. 28, 2273–2277, doi:10.1190/1.3255314.

Lin, T., Lebed, E., Erlangga, Y.A. and Herrmann, F.J. [2008] Interpolating solutions of the helmholtz equation with compressed sensing. *SEG Technical Program Expanded Abstracts*, SEG, vol. 27, 2122–2126.

Mulder, W. and Plessix, R. [2004] How to choose a subset of frequencies in frequency-domain finite-difference migration. **158**, 801–812.

Romero, L.A., Ghiglia, D.C., Ober, C.C. and Morton, S.A. [2000] Phase encoding of shot records in prestack migration. *Geophysics*, **65**(2), 426–436, doi:10.1190/1.1444737.

Sirgue, L. and Pratt, R.G. [2004] Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, **69**(1), 231–248, doi:10.1190/1.1649391.

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.

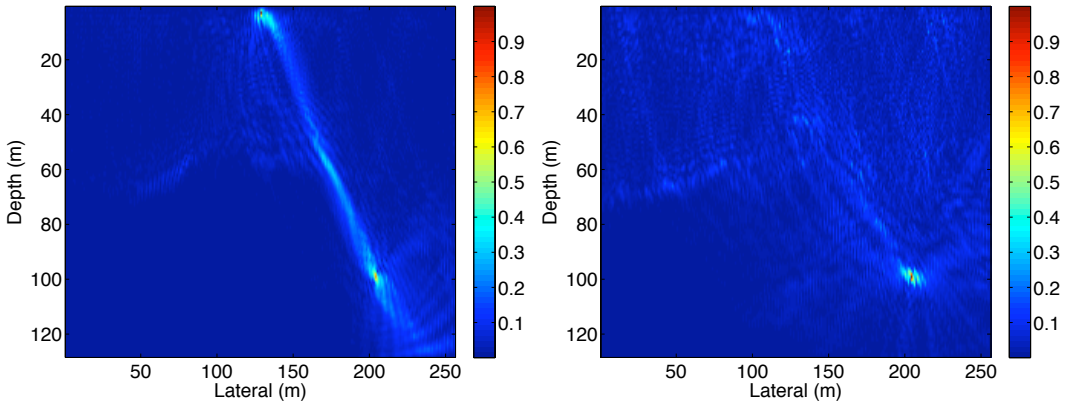

Figure 1 *Sequential-shot versus simultaneous-source imaging of a single point diffractor in the target zone.* ***(a)*** *Migrated image for a single sequential shot for all frequencies ([0-31.25 Hz])* ***(b)*** *The same as* ***(c)*** *but for a single simultaneous shot. Notice the improved focusing for the simultaneous shot.*

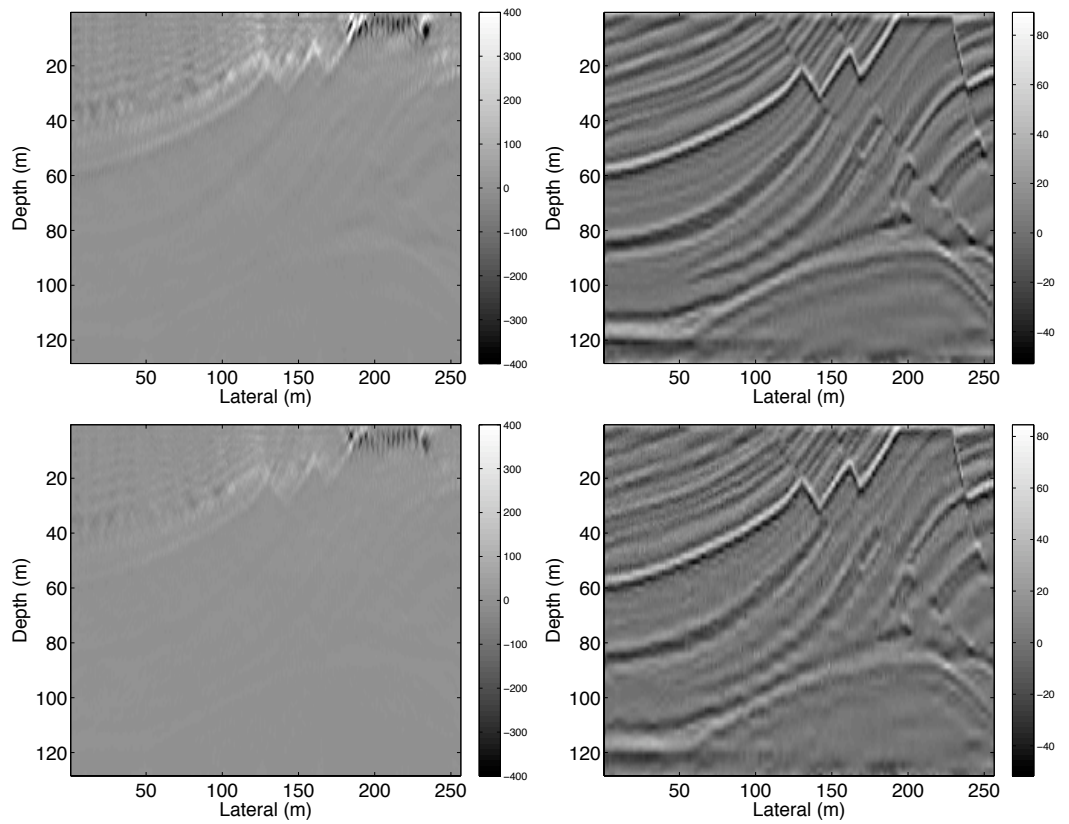

Figure 2 *Comparison between standard dimensionality-reduced gradient updates, and updates obtained by sparsity promotion.* ***(a)*** *Migrated image for simultaneous data for* $n'_f = 10$ *and* $n'_s = 50$. ***(b)*** *The same as* ***(a)*** *but now with sparsity-promoting recovery, yielding an improvement in SNR from* $1.68\,\mathrm{dB}$ *to* $18.22\,\mathrm{dB}$. ***(c-d)*** *are the same as* ***(a-b)*** *but for* $n'_f = 8$ *and* $n'_s = 25$*, yielding an improvement in SNR from* $1.32\,\mathrm{dB}$ *to* $12.11\,\mathrm{dB}$.