Tackling the “data deluge”: a dimensionality-reduction approach

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

Current-day imaging and inversion technology increasingly relies on faithful samplings and simulations of seismic wavefields. This reliance on full sampling and high-fidelity wavefield simulations strains our acquisition and processing systems and overcoming this impediment is becoming one of the main challenges faced by our industry.

By using randomized dimensionality-reduction techniques, we propose a new strategy where acquisition and computational costs are no longer dictated by the sampling grid but by transform-domain compressibility of the image. To arrive at this result, we combine recent findings from machine learning / stochastic optimization—where (nonlinear) inversions are carried out on random subsets of data—and compressive sensing—where data that permit compressible representations are deliberately subsampled.

The key idea of the stochastic approximation is to reduce computational costs by computing each gradient update on a different randomly selected subset of data. In seismic exploration, this corresponds to carrying out migrations with one or a few incoherent supershots made of superpositions of random source-encoded experiments. While this approach introduces source cross talk, it has been applied successfully to reduce the cost of least-squares migration and full-waveform inversion because it reduces the number of wave-equation solves. However, the method is sensitive to noise and relies on relatively large numbers of supershots and wave simulations to get reasonable results.

Compressive sensing also relies on random superpositions but it differs by deriving rigorous recovery guarantees expressed in terms of the compressibility of the solution—read the percentage of curvelet coefficients required to approximate the image accurately—and the degree of subsampling—read the number of simultaneous-source experiments. In this case, data is recovered by transform-domain sparsity promotion and this leads to cost reductions as long the cost of sparsity promotion is smaller than the gains obtained by reducing the number of source experiments.

Motivated by our successful application of this strategy to wavefield simulations with the Helmholtz equation, we extend and adapt this method to seismic imaging and full-waveform inversion, which are different because the number of unknowns of the image is typically much smaller than the size of the data. Remember that compressive sensing is designed for the opposite case where the number of unknowns is much smaller than the number of observations.

We leverage the efficiency of the stochastic approximation and the robustness of compressive sensing by casting imaging and full-waveform inversion into a series of compressive-sampling experiments made of simultaneous-source experiments. The data for each experiment is “collected” with a new randomization. Aside from reducing the number of wave simulations, the proposed method also has the advantage that it exploits the “natural” sparsity of the image domain.

During this talk, we demonstrate that this approach can lead to significant speed and quality improvements in least-squares migration and full-waveform inversion. Because the method consists of randomized subselections of data, we envisage enticing perspectives of this technology when combined with future seismic-data acquisition practices.