Shot encoding with random projections

Ludwig Schmidt and Piotr Indyk, MIT, Chaohui Chen, Amik St.-Cyr and Detlef Hohl, Shell International E&P, Inc.

SUMMARY

In order to image complex geological structures, seismic surveys acquire an increasingly large amount of data. While the resulting data sets enable higher-resolution images of the subsurface, they also contain redundant information and require large computational resources for processing. One approach for mitigating this trend is blended imaging, which combines the original shot records into a smaller number of blended shots at the expense of crosstalk in the final image. Since the cost of imaging is roughly proportional to the number of shots, blended imaging directly leads to a faster imaging process. In contrast to the existing shot encoding schemes, we establish a novel connection between blended imaging and dimensionality reduction using the Johnson-Lindenstrauss lemma. We introduce three new shot encoding schemes based on random projections and evaluate their performance. Our experiments on three data sets show that our random shot encoding schemes are competitive with existing shot encoding schemes and outperform decimated shot encoding for small numbers of shots.

INTRODUCTION

With the increasing amount of data collected in seismic surveys, imaging the resulting data sets becomes a considerable computational challenge. For many migration methods such as reverse-time migration (RTM), the computational cost scales linearly with the number of shots. Hence, one promising approach to decreasing the cost of imaging is to decrease the number of shots used for migration. However, simply discarding a significant fraction of the available shots also leads to a loss of information and consequently image quality. Therefore, *blended imaging* combines several original shot records into a smaller number of blended shot records before migration. The goal is to achieve a small number of blended shots while minimizing the resulting crosstalk.

In the past, several shot encoding schemes have been proposed for blended imaging. Overall, shot encoding schemes try to minimize the crosstalk between the original shot records which have been combined into a single, encoded shot. Since this crosstalk is the main source of unwanted noise in blended imaging, minimizing the crosstalk leads to a final image that is close to the result of migrating each original shot individually. In order to achieve this, shot encoding schemes vary various properties of the shot records such as amplitude, phase or time delay. Examples for shot encoding schemes are phase-encoding (Romero et al., 2000) and modulated-shot migration (Soubaras, 2006). A recent survey contains an overview of many shot encoding schemes (Godwin and Sava, 2011).

The main goal of our work is to study the performance of *random projections* as shot encoding schemes. Random projections are a powerful tool for reducing the dimensionality of a data set while approximately preserving relevant properties of the data, such as lengths of vectors or inner products. Random projections are widely used in the design of algorithms, machine learning and compressive sensing. We use three different random projection schemes in order to reduce the dimensionality of a seismic data set along the source axis, which corresponds to blending the original shot records.

BACKGROUND

Here, we briefly review the mathematical formulation of blended imaging, which we then use to formally define our random shot-encoding schemes in the following section. Moreover, we introduce an important concept from the field of dimensionality reduction as a justification for our choice of shot-encoding schemes.

Conventional imaging

Let $W_S(x,t,s)$ and $W_R(x,t,s)$ be the source and receiver wavefields for location x, time t and shot s. Using the conventional cross-correlation imaging condition, the final image $R(x)$ is given by

$$R(x) = \sum_s \sum_t W_S(x,t,s) W_R(x,t,s). \tag{1}$$

Since the image values at different locations are independent, we drop the location parameter and focus on a single location. Moreover, let W_t^S and W_t^R be the column vectors containing the source and receiver wavefields from the different shots at time t. Then we can write the image value as a sum of inner products:

$$R = \sum_t (W_t^S)^H W_t^R, \tag{2}$$

where $(W_t^S)^H$ denotes the conjugate transpose. In this formulation, blended imaging can be seen as a way of reducing the dimensionality of W_t^S and W_t^R while preserving their inner products.

Blended imaging

For N_s original shots and N_e blended shots, a linear shot encoding scheme can be described by a $N_e \times N_s$ encoding matrix E. Each row of E corresponds to one blended shot and the entry E_{ij} contains the coefficient of original shot j in blended shot i. The main computational advantage of blended imaging comes from the following observation: for linear migration schemes (e.g. RTM), we can compute the blended wavefield

$$B_t^S = E W_t^S \tag{3}$$

with only N_e migrations by first blending the source data according to E and then migrating the blended data. By using the same approach for the receiver wavefields, we get an overall speedup of N_s/N_e compared to migrating all N_s original shots individually. Since migration is usually the computationally

DOI http://dx.doi.org/10.1190/segam2013-1260.1

most expensive part of imaging, this speedup can reduce the overall cost significantly.

Using the conventional imaging condition on the blended wavefields gives the folllowing image:

$$R_e = \sum_t (B_t^S)^H B_t^R \tag{4}$$

$$= \sum_t W_t^S E^H E W_t^R \tag{5}$$

$$= R + \sum_t W_t^S (E^H E - I) W_t^R, \tag{6}$$

where I is the identity matrix. Note that for $E^H E = I$, blended imaging gives the same result as imaging with individual shots. The off-diagonal entries of the crosstalk matrix $C = E^H E$ correspond to unwanted noise introduced by correlating wavefields from different shots. Therefore, a good shot encoding scheme should have the following two properties: N_e is small and $E^H E \approx I$.

Johnson-Lindenstrauss transforms

One of the main results in the theory of dimensionality reduction is the Johnson-Lindenstrauss (JL) lemma (Johnson and Lindenstrauss, 1984). Informally, the lemma states that there are mappings that approximately preserve the distances between a set of points while significantly reducing the dimensionality of the point set.

Theorem 1. *Let $x_1, \ldots, x_n \in \mathbb{R}^d$ be a set of points and let $0 < \varepsilon < 1/2$. Then there exists a mapping $f : \mathbb{R}^d \to \mathbb{R}^k$ with $k = O\left(\frac{\log n}{\varepsilon^2}\right)$ such that for all i, j*

$$(1-\varepsilon)\|x_i - x_j\|_2 \le \|f(x_i) - f(x_j)\|_2 \le (1+\varepsilon)\|x_i - x_j\|_2. \tag{7}$$

We call such a mapping f a JL-transform. It can be shown that functions of the form $f(x) = Mx$ are JL-transforms for several types of random matrices M. The JL-lemma and JL-transforms have many applications in the design of algorithms, machine learning and signal processing, e.g. as measurement matrices in compressive sensing (Baraniuk et al., 2008).

For our use case, it is worth noting that JL-transforms also preserve inner products. Note that

$$\|x+y\|_2^2 - \|x-y\|_2^2 = 4x^H y. \tag{8}$$

So approximately preserving the pointwise distances $\|x+y\|_2$ and $\|x-y\|_2$ means that we also approximately preserve the inner product $x^H y$.

RANDOM SHOT-ENCODING SCHEMES

We now introduce three shot-encoding schemes based on JL transforms. In order to complement the survey of Godwin and Sava (2011), we also analyze the distributions of the corresponding crosstalk matrices.

Gaussian ensemble

Each entry in the encoding matrix E is given by a standard normal random variable, normalized by the number of encoded shots:

$$E_{ij} \sim N(0, 1/N_e). \tag{9}$$

The diagonal entries of the crosstalk matrix are distributed according to a χ^2-distribution with N_e degrees of freedom:

$$C_{ii} = \sum_{j=1}^{N_e} E_{ij}^2$$

$$\sim \chi^2_{N_e}.$$

So $\mathbb{E}[C_{ii}] = 1$ and $\mathrm{Var}(C_{ii}) = 2/N_e$.

The off-diagonal entries are sums of product-normal distributions. Hence for $i \neq j$, $\mathbb{E}[C_{ij}] = 0$ and $\mathrm{Var}(C_{ij}) = 1/N_e$. This shows that $\mathbb{E}[E^H E] = I$.

Moreover, Gaussian ensembles give JL transforms (Indyk and Motwani, 1998; Dasgupta and Gupta, 2003).

Rademacher matrices

Each entry in the encoding matrix E is given by a ± 1 fair coin flip with normalized outcome:

$$E_{ij} = \begin{cases} +1/\sqrt{N_e} & \text{with } p = 1/2 \\ -1/\sqrt{N_e} & \text{with } p = 1/2 \end{cases}. \tag{10}$$

Since $C_{ii} = \sum_{j=1}^{N_e} E_{ij}^2 = 1$, the diagonal entries of the crosstalk matrix are always 1.

Note that $E_{ij} \sim (2\mathrm{Ber}(1/2) - 1)/\sqrt{N_e}$, where $\mathrm{Ber}(p)$ is the Bernoulli distribution with success probability p. For $i \neq j$, E_{ik} and E_{jk} are independent and we get

$$C_{ij} = \sum_{k=1}^{N_e} E_{ik} E_{jk} \tag{11}$$

$$\sim (2\mathrm{Bin}(N_e, 1/2) - N_e)/N_e, \tag{12}$$

where $\mathrm{Bin}(n, p)$ is the Binomial distribution with n trials and success probability p. So for the off-diagonal entries, we have $\mathbb{E}[C_{ij}] = 0$ and $\mathrm{Var}(C_{ij}) = 1/N_e$.

Finally, Rademacher matrices are also JL transforms (Achlioptas, 2003).

Sparse random matrices

We consider sparse random matrices with an entrywise distribution similar to Rademacher matrices but with an additional sparsity parameter q:

$$E_{ij} = \begin{cases} +1/\sqrt{N_e q} & \text{with } p = q/2 \\ 0 & \text{with } p = 1-q \\ -1/\sqrt{N_e q} & \text{with } p = q/2 \end{cases}. \tag{13}$$

Due to the similarity with Rademacher matrices, we omit the calculations for the distribution of C_{ij}. The mean and variance of C_{ij} are the same as for Rademacher matrices.

For q down to $1/3$, this type of sparse random matrix is known to give JL transforms (Achlioptas, 2003).

DOI http://dx.doi.org/10.1190/segam2013-1260.1

RESULTS

In order to evaluate our random shot-encoding schemes, we study the trade-off between the number of blended shots N_e and the final image quality on two small synthetic models and the Sigsbee2A data set.

Evaluation baseline

We compare our shot-encoding schemes with the images computed from N_e equidistant original shots. This selection of original shots is also called *decimated* shot encoding. We use this baseline for two reasons:

- The computational cost of migrating N_e original shots is the same as that of migrating N_e blended shots.
- Decimated shot encoding performs better than most other shot-encoding schemes reported in the literature (Godwin and Sava, 2011).

Hence, our random shot encoding schemes would ideally produce a better image than the result of decimated shot encoding with the same number of shots.

Image quality comparison

In order to assess the quality of an image computed with a certain shot encoding scheme, we compare the it to a corresponding reference image. For all our models, we use the result of imaging with a large number of individual shots as reference image. The rationale behind this comparison method is that we use shot-encoding in order to approximate the result of imaging with individual shots.

The first measure we consider is the normalized ℓ_2-error:

$$\mathrm{Err}_2(R_e,R) = \frac{\|R-R_e\|_2}{\|R\|_2}. \qquad (14)$$

Moreover, we consider the structural similarity index (SSIM), which is a standard measure in the evaluation of image and video codecs (Wang et al., 2004). The SSIM is designed to quantify similarity as perceived by the human visual system. It is defined on windows (e.g. 8×8 blocks) of the two images:

$$\mathrm{SSIM}(a,b) = \frac{(2\mu_a\mu_b+c_1)(2\sigma_{ab}+c_2)}{(\mu_a^2+\mu_b^2+c_1)(\sigma_a^2+\sigma_b^2+c_2)}, \qquad (15)$$

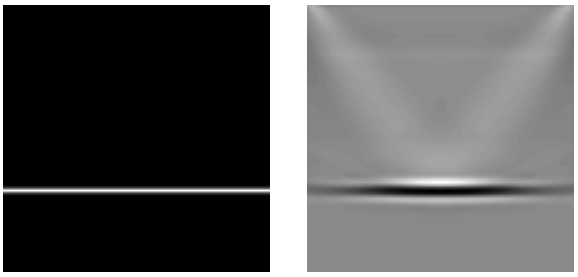
where

- μ_x is the average of window x
- σ_x is the variance of window x
- σ_{xy} is the covariance of windows x and y
- c_1 and c_2 are stabilization constants.

The final SSIM value is the average over all windows. The SSIM of two identical images is 1.

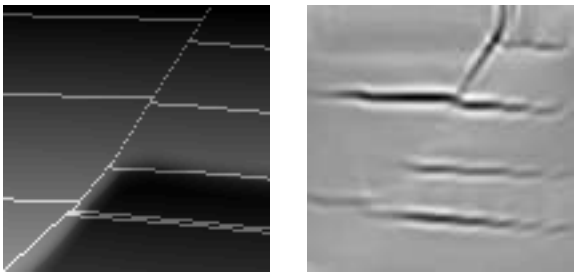
Small synthetic models

We now apply the evaluation methods outlined above on two small synthetic models. Both models consist of 100×100 grid points and the reference images are computed from 100 individual shots. Figures 2 and 1 show the velocity models and reference images. Figures 5 to 4 contain the corresponding results. For the random encoding schemes, each point is the average of 20 runs.



(a) Velocity model (b) Reference image

Figure 1: The horizontal reflector model.



(a) Velocity model (b) Reference image

Figure 2: The fault model.

While decimated shot encoding gives better results for larger number of shots (ten or more), random shot encoding schemes perform better for small number of shots.

Sigsbee2A

For the Sigsbee2A model (figure 7(a)), we use a reference image computed from all 500 shots. In order to focus the quality comparison on the relevant features below the salt structure, we compute our quality measures only on the lower-left triangular part of the image. Figure 7(b) contains the relevant part of the final reference image.

Figures 8 and 9 show the results of the Sigsbee2A experiments. Due to the size of the data set, each data point of the random schemes now correponds to only a single run. As before, decimated shot encoding performs better for larger number of shots while random shot encoding has a small advantage for a small number of shots.

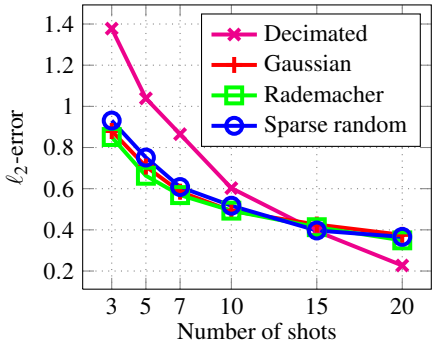


Figure 3: ℓ_2-error results for the horizontal reflector model.

DOI http://dx.doi.org/10.1190/segam2013-1260.1

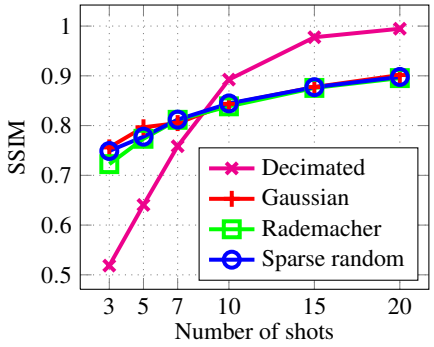


Figure 4: SSIM results for the horizontal reflector model.

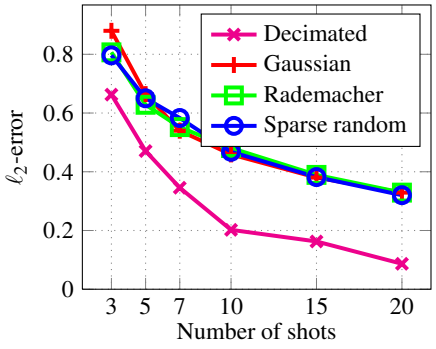


Figure 5: ℓ_2-error results for the fault model.

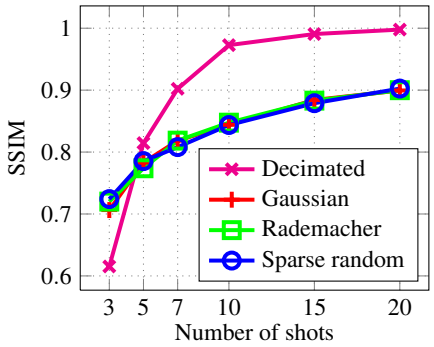
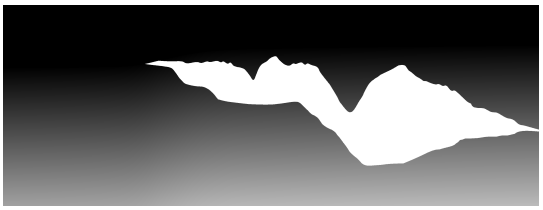
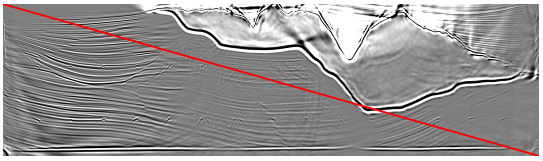


Figure 6: SSIM results for the fault model.



(a) Velocity model



(b) Reference image. We use only the part below the red line for comparing final images.

Figure 7: The Sigsbee2A model.

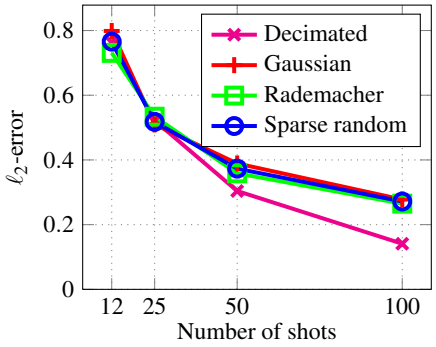


Figure 8: ℓ_2-error results for the Sigsbee2A model.

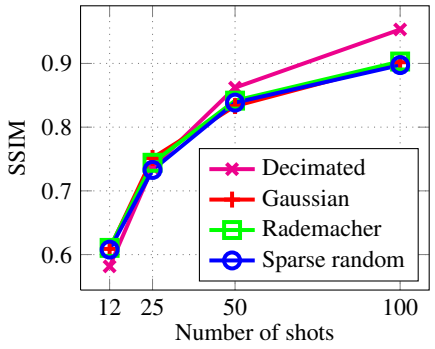


Figure 9: SSIM results for the Sigsbee2A model.

CONCLUSIONS

We have established a novel connection between blended imaging and dimensionality reduction with the JL lemma. Following this approach, we have proposed three shot encoding schemes based on random projections. Our experiments on three data sets demonstrate that our random encoding schemes are competitive with decimated shot encoding (and hence also with other shot encoding schemes). For small numbers of shots, random shot encoding performs at least as good as decimated shot encoding and significantly better on some velocity models. Since the cost of blending shot records is usually much smaller than the cost of migration, the computational overhead incurred by random shot encoding is negligible. Moreover, random shot encoding has the potential to preserve more information of the blended shots, which can enable more sophisticated approaches to recovering the original wavefields from blended data. This direction is the focus of our current work.

ACKNOWLEDGMENTS

The authors thank Shell International E&P for financial support, computational resources and permission to publish this work. Some experiments in this paper rely on test data from SMAART and software from the Madagascar project. The authors also thank Jonathan Kane, Chinmay Hegde and Paul Sava for helpful discussions.

http://dx.doi.org/10.1190/segam2013-1260.1

EDITED REFERENCES

Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2013 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Achlioptas, D., 2003, Database-friendly random projections: Johnson-Lindenstrauss with binary coins: Journal of Computer and System Sciences, **66**, no. 4, 671–687, http://dx.doi.org/10.1016/S0022-0000(03)00025-4.

Baraniuk, R., M. Davenport, R. DeVore, and M. Wakin, 2008, A simple proof of the restricted isometry property for ran- dom matrices: Constructive Approximation, **28**, no. 3, 253–263, http://dx.doi.org/10.1007/s00365-007-9003-x.

Dasgupta, S., and A. Gupta, 2003, An elementary proof of a theorem of Johnson and Lindenstrauss: Random Structures and Algorithms, **22**, no. 1, 60–65, http://dx.doi.org/10.1002/rsa.10073.

Godwin, J., 2011, Blended source imaging by amplitude encoding: Master's thesis, Colorado School of Mines.

Godwin, J., and P. Sava, 2011, A comparison of shot-encoding schemes for wave-equation migration: 81st Annual International Meeting, SEG, Expanded Abstracts, 32–36.

Indyk, P., and R. Motwani, 1998, Approximate nearest neighbors: Towards removing the curse of dimensionality: Proceedings of the 30th Annual ACM Symposium on Theory of Computing, 604–613.

Johnson, W. B., and J. Lindenstrauss, 1984, Extensions of Lipschitz mappings into a Hilbert space: Contemporary Mathematics, **26**, 189–206.

Romero, L., D. Ghiglia, C. Ober, and S. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436, http://dx.doi.org/10.1190/1.1444737.

Soubaras, R., 2006, Modulated-shot migration: 76th Annual International Meeting, SEG, Expanded Abstracts, 2430–2434.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, 2004, Image quality assessment: from error visibility to structural similarity: IEEE Transactions on Image Processing, **13**, no. 4, 600–612, http://dx.doi.org/10.1109/TIP.2003.819861.