

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Análise de Erro do *Compressive Sensing* para Dados Sísmicos

Éberton da Silva Marinho¹

Programa de Pós-Graduação em Ciências e Engenharia do Petróleo, UFRN, Natal, Brasil

Francisco Iranildo Ferreira do Nascimento Gomes²

Programa de Pós-Graduação em Ciências e Engenharia do Petróleo, UFRN, Natal, Brasil

Tiago Cavalcanti da Rocha³

Programa de Pós-Graduação em Ciências e Engenharia do Petróleo, UFRN, Natal, Brasil

Gilberto Corso⁴

Departamento de Biofísica e Farmacologia, Centro de Biociências, UFRN, Natal, Brasil

Liacir dos Santos Lucena⁵

Departamento de Física Teórica e Experimental, UFRN, Natal, Brasil

International Center for Complex Systems, UFRN, Natal, Brasil

Programa de Pós-Graduação em Ciência e Engenharia do Petróleo, UFRN, Natal, Brasil

Resumo. O *Bayesian Compressive Sensing* é aplicado a um sismograma real usando várias amostragens e duas bases *Wavelets*: *Daubechies 6* e *Symlet 9*. A distribuição de erro do sinal recuperado pelo *Compressive Sensing* tende a ser normal para baixas amostragens e não normal, muito assimétrica e leptocúrtica, para altas amostragens.

Palavras-chave. *Compressive Sensing*, *Bayesian Compressive Sensing*, *Wavelets*, *Sismogramas*, *Análise de erro*.

1 Introdução

A aquisição e processamento de sinais sísmicos têm uma grande importância na indústria do petróleo. Tais atividades são primordiais para um melhor entendimento e

¹ ebertonm@gmail.com, eberton.marinho@ifrn.edu.br

² ironrdc@gmail.com

³ supremotiago@gmail.com

⁴ gfcorso@gmail.com

⁵ liacir.lucena@gmail.com

predição de camadas e estruturas formadoras do subsolo. Elas comporão o estudo de viabilidade de exploração de uma determinada região juntamente com outras análises [8, 9].

A técnica do *Compressive Sensing* (CS) [1, 2, 3] tem se mostrado útil na tarefa de reduzir o volume de dados na aquisição de um sinal. O CS contorna um dos maiores limitadores da amostragem que é o teorema de Shannon-Nyquist. Este teorema afirma que para se restaurar um sinal com precisão, faz-se necessária uma taxa de amostragem que tenha pelo menos o dobro da frequência mais alta dos dados.

Apesar do sucesso da CS, sua utilização na área do petróleo ainda é recente. Neste trabalho realizamos, sucintamente, um estudo do erro na recuperação de sinais sísmicos usando o CS para várias taxas de amostragem do sinal e com o emprego de duas *Wavelets* [4, 5, 7] distintas como bases. De fato, fizemos uma exploração heurística do comportamento de várias *Wavelets*, sendo que a *Daubechies 6* e a *Symlet 9* se mostraram as mais promissoras para nossas análises. Além disto, estudamos o algoritmo *Bayesian Compressive Sensing* (BCS) [6, 10] e o empregamos neste trabalho. Para analisar o erro da reconstrução do sinal estudamos sua distribuição. Em especial, testamos a normalidade, a simetria e verificamos o quanto a distribuição pode ser mesocúrtica.

2 Metodologia

2.1 Dados sísmicos

A reflexão sísmica é uma das técnicas mais utilizadas no levantamento geológico na área de exploração de petróleo. Cada traço sísmico é o resultado da captação das informações refletidas pelo sistema de camadas terrestres, sendo que, a junção de vários traços sísmicos dá origem a um sismograma. Os dados sísmicos aqui utilizados são compostos por 33 traços que foram retirados do Capítulo 4 de [8] e encontram-se disponíveis em (<http://www.morganclaypool.com/doi/pdfplus/10.2200/S00384ED1V01Y201109SPR010>). Todos os *scripts* utilizados foram criados no ambiente de programação MATLAB (<http://www.mathworks.com/>).

2.2 Compressive Sensing

Considere um sinal $x[i]$, como por exemplo, uma série temporal com N elementos. Em seguida usamos uma base do tipo *Wavelet*, $\Psi = [\psi_1 | \psi_2 | \dots | \psi_N]$, $N \times N$, para ψ_i um vetor coluna. Desta forma, o sinal original x pode ser expresso como

$$x = \sum_{i=1}^N s_i \psi_i \quad \text{ou} \quad x = \Psi s \quad (1)$$

onde s é um vetor coluna de coeficientes reescrito como segue: $s_i = \langle x, \psi_i \rangle = \psi_i^T x$. Pode-se ver que s e x são representações equivalentes de um mesmo sinal, porém x se encontra no domínio do tempo, enquanto s_i está no domínio de Ψ . O sinal x será do tipo K -esparso se ele puder ser escrito como uma combinação linear de apenas K vetores, isto é, $K \ll N$.

Para simularmos o processo de aquisição comprimida do sinal, vamos supor um processo de medição linear geral $\Phi = \{\phi_j\}_{j=1}^M$. Este processo computa $M < N$ produtos internos entre o sinal original x e uma base de vetores, como em $y_j = \langle x, \phi_j \rangle$, onde $M \sim K$. Com auxílio de (1) pode-se escrever

$$y = \Phi x + e = \Phi \Psi s + e = \Theta s + e \tag{2}$$

onde $\Theta = \Phi \Psi$ é uma matriz $M \times N$. Ademais, a quantidade e aqui será o erro. Destacamos que no processo do CS a medição é não adaptativa, ou seja, a matriz Θ não depende de x [1, 2, 4].

2.3 Bayesian Compressive Sensing

Uma abordagem Bayesiana [6, 10] pode ser aplicada ao problema de recuperação, onde o sinal s desconhecido é associado a uma probabilidade $p(s | \gamma)$, chamada de distribuição a priori, que modela nosso conhecimento sobre a natureza de s . O y é modelado como um conjunto de observações resultantes de um processo aleatório com distribuição condicional $p(y | s, \beta)$, onde $\beta = 1/\sigma^2$ é o inverso da variância do ruído. Estas distribuições dependem do modelo de parâmetros γ e β , que são hiperparâmetros, e distribuições a priori adicionais, chamadas hiperprioris.

A modelagem proposta constitui uma forma hierárquica de três estágios. Os primeiros dois estágios desta priori hierárquica resultam em uma distribuição de Laplace $p(s | \lambda)$, e o último estágio é incorporado para calcular λ . Na Figura 1, em (a), temos o sinal original (em azul) do traço 17 e, em (b), a comparação com a recuperação pelo método BCS para 30% do sinal original utilizando a *Wavelet Daubechies 6* (sinal recuperado em vermelho).

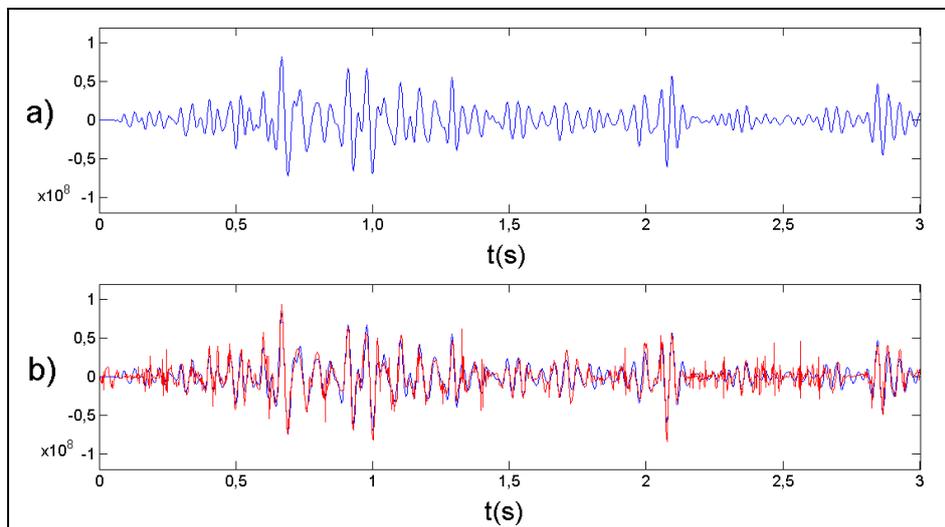


Figura 1: Exemplo de um traço sísmico. Em a), a série original de dados do traço número 17 com 1501 pontos. Em b), a superposição do traço sísmico original (em azul) e o recuperado (em vermelho) com amostragem de 30%, utilizando a *Wavelet Daubechies 6*.

3 Análises dos Histogramas do Erro

Após procedermos a amostragem e a recuperação para 1000 execuções de cada traço utilizando o método BCS com vários parâmetros (cinco amostragens e duas *Wavelets*), calculamos o erro relativo. A matriz de sensoriamento que usamos neste trabalho foi aleatória gaussiana com média nula e desvio padrão inversamente proporcional ao número de amostras. Em seguida, para cada traço, criamos histogramas do erro relativo, como os apresentados na Figura 2. Para a execução do algoritmo BCS foi usada a função *BCS_fast_rvm* disponível na biblioteca SparseLab versão 2.1 que se encontram no site <https://sparselab.stanford.edu/>.

Com os dados dos histogramas, calculamos as médias e desvios padrões da recuperação do sinais com cada uma das 5 amostras utilizando as *Wavelets Daubechies 6* e *Symlet 9*, apresentados na Tabela 1. Em seguida, para cada traço, computamos os testes de normalidade (pelo método *Kolmogorov-Smirnov*), de simetria e de curtose, apresentados na Tabela 2.

As Tabelas 1 e 2 indicam, de uma forma resumida, três tendências: (i) o erro relativo diminui com o aumento da amostragem; (ii) amostragens muito altas (50%) levam a distribuições não normais, ainda mais assimétricas à direita e leptocúrticas; (iii) parece haver uma transição na resposta do *Compressive Sensing* quando a taxa de aquisição se situa entre 40 e 50%. Ressaltamos também que o desvio da curtose observado foi para tendência leptocúrtica (os erros se tornam mais concentrados junto a média) quando a distribuição deixa de ser normal. Observou-se também que o desvio da assimetria sempre foi para a direita, ou seja, a recuperação do *Compressive Sensing* tem maior probabilidade de produzir sinais recuperados com grande erro relativo. Acrescenta-se que qualitativamente os resultados se mostraram semelhantes para as duas *Wavelets* estudadas.

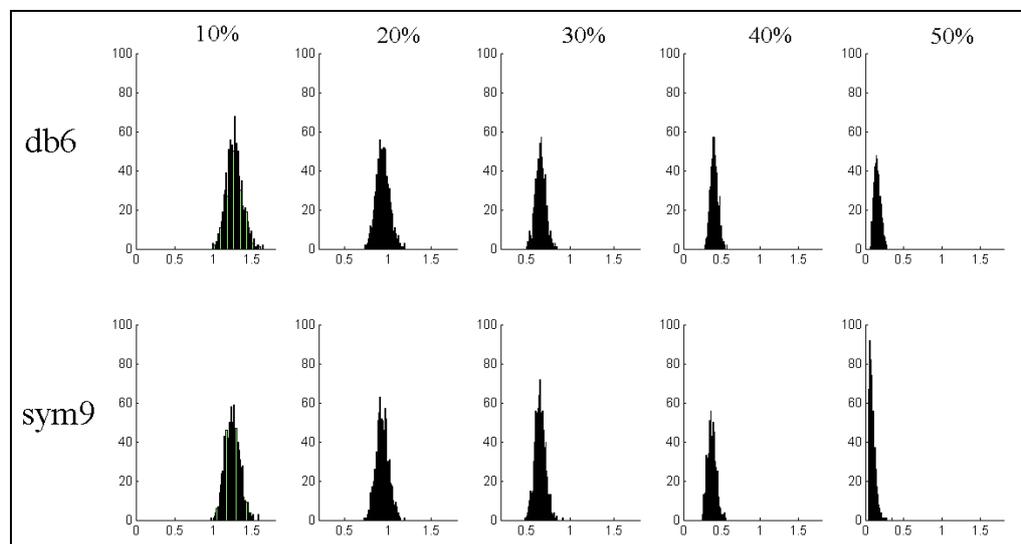


Figura 2: Histogramas do erro relativo da recuperação do traço de número 17 através do BCS para várias amostragens. O eixo da abscissa representa o erro relativo e o eixo da ordenada a frequência absoluta.

Tabela 1: Medições das médias (μ) e desvio padrões (σ) para o conjunto de todos os traços. Nesta tabela são indicados os valores para as taxas de amostragem e duas *Wavelets*.

$\mu \pm \sigma$	10%	20%	30%	40%	50%
db6	$1,28 \pm 0,12$	$0,96 \pm 0,13$	$0,67 \pm 0,12$	$0,37 \pm 0,10$	$0,09 \pm 0,05$
sym9	$1,33 \pm 0,15$	$0,94 \pm 0,13$	$0,65 \pm 0,12$	$0,36 \pm 0,09$	$0,07 \pm 0,05$

Tabela 2: Tabela quantitativa para os testes de normalidade, simetria e curtose da recuperação do sismograma pelo método BCS.

	Testes	10%	20%	30%	40%	50%
db6	Normal por <i>Kolmogorov-Smirnov</i>	32	29	33	28	0
	Simetria	8	9	5	15	0
	Inclinação à Direita	25	24	28	18	33
	Curtose (Mesocúrtico)	27	20	30	27	1
	Leptocúrtico	6	13	3	6	32
sym9	Normal por <i>Kolmogorov-Smirnov</i>	28	27	33	29	1
	Simetria	10	6	6	14	1
	Inclinação à Direita	23	27	27	18	32
	Curtose (Mesocúrtico)	28	26	30	26	2
	Leptocúrtico	5	7	3	7	31

4 Conclusões e Trabalhos Futuros

Aplicamos a técnica do *Comprehensive Sensing* sobre um sismograma real usando duas bases *Wavelets*: *Daubechies 6* e *Symlet 9*. O estudo da distribuição dos erros é importante na indústria do petróleo, porque ajuda a identificar falhas em imagens sísmicas que podem levar a problemas de interpretação geofísicas. Uma distribuição de erro que se desvia da normalidade é potencialmente perigosa em engenharia, pois a presença de assimetria marcante no sinal gera caudas longas e a possibilidade de falsos resultados oriundos de dados muito distantes da média.

Uma análise sobre o erro do sinal recuperado mostra que, como esperado, o erro diminui com o aumento da amostragem do sinal, tanto do ponto de vista da média do erro, quanto da sua flutuação em torno da média. Porém, amostragens altas (50%) levam a distribuições de erro não normais, ainda mais assimétricas e não mesocúrticas. Este resultado sugere um ponto frágil do *Comprehensive Sensing* que deve ser explorado em mais detalhes. Por fim uma outra questão também se mostrou de interesse. Os resultados indicam uma transição na distribuição de erro quando se varia a taxa de aquisição neste sistema. Existe um ponto entre 40 e 50% onde parece haver uma descontinuidade na resposta do *Comprehensive Sensing*. Uma análise em torno desta transição é um dos objetos de futuras pesquisas, bem como a realização de testes semelhantes com outras *Wavelets* e outros métodos de recuperação, com ou sem filtros.

Referências

- [1] R. G. Baraniuk, Compressive sensing, *IEEE Signal Processing Magazine*, vol. 24, 118-121, (2007), DOI: 10.1109/MPS.2007.4286571.
- [2] E. Candès, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 489–509, (2006), DOI: 10.1109/TIT.2005.862083.
- [3] D. L. Donoho, Compressed sensing, *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, (2006), DOI: 10.1109/TIT.2006.871582.
- [4] M. Duarte, M. Wakin and R. Baraniuk, Wavelet-domain Compressive Signal Reconstruction Using a Hidden Markov Tree Model, *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, Nevada, (2008), DOI: 10.1109/ICASSP.2008.4518815.
- [5] L. He and L. Carin, Exploiting Structure in Wavelet-Based Bayesian Compressive Sensing, *IEEE Trans. Signal Processing*, vol. 57, no. 9, pp.3488-3497, (2009), DOI: 10.1109/TSP.2009.2022003.
- [6] S. Ji, Y. Xue and L. Carin, Bayesian compressive sensing, *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346–2356, (2008), DOI: 10.1109/TSP.2009.2022003.
- [7] S. Mallat, *A wavelet tour of signal processing*, 2nd ed., Academic Press, (1998), DOI: 10.1109%2F34.192463.
- [8] W. A. Mousa and A. A. Al-Shuhail, *Processing of Seismic Reflection Data Using MATLAB*, Synthesis Lectures on Signal Processing, Morgan & Claypool Publishers, (2011), DOI: 10.2200/S00384ED1V01Y201109SPR010.
- [9] E. Onajite, *Seismic Data Analysis Techniques in Hydrocarbon Exploration*, 1st ed., Elsevier (2013), ISBN-10: 0124200230
- [10] M. W. Seeger and H. Nickisch, Compressed sensing and Bayesian experimental design, in *ICML*, pp. 912–919. (2008), DOI: 10.1145/1390156.1390271.