



Relatório - MS877

DETECÇÃO DE COMPRESSÃO COM A NORMA 1

ALUNO: MATHEUS QUEIROZ ZABIN

ORIENTADOR: PROF. DR. LÚCIO TUNES DOS SANTOS

Campinas, junho de 2017

Sumário

1	Resumo	2
2	Introdução ao problema	2
3	Unicidade da solução	3
3.1	Análise inicial	3
3.2	Propriedade Isométrica Restrita	4
3.3	Propriedade Isométrica Restrita em matrizes aleatórias normais	5
4	Encontrando a solução esparsa de $Ax=b$ com a norma-1	5
5	Matriz de detecção	6
6	Aplicações	7
6.1	Compressão de dados	7
6.2	Reconstrução de dados	11
7	Considerações finais	13
8	Referências Bibliográficas	14

1 Resumo

No mundo atual, o fluxo de dados (sinais, imagens, etc), está crescendo cada dia mais e de forma cada vez mais rápida. Porém, o número de características realmente significativas em uma quantidade relativamente grande de dados é, de forma geral, muito menor do que o número de coeficientes em uma representação de dados, ou seja, os dados são compressíveis. Em processamento de dados, o padrão é medir os dados brutos, e em seguida, comprimir estas medidas antes do armazenamento. Este processo de aquisição de dados pode ser nomeado como “detecção completa mais compressão”.

A técnica nomeada “Detecção de Compressão” (Compressive Sensing - CS), é uma metodologia cada vez mais utilizada no âmbito de processamento digitais de sinais, atraindo cada vez mais o olhar dos pesquisadores da área. O objetivo é reduzir o número de medições durante a fase de aquisição, de modo que nenhuma compressão adicional seja necessária. A teoria básica de CS, se fundamenta no fato de que um sinal qualquer de comprimento N , pode ser recuperado a partir de um número muito menor que N medições, através de otimização. O preço a ser pago é a necessidade de procedimentos de recuperação mais sofisticados.

2 Introdução ao problema

Considere um sinal discreto como um vetor $\alpha \in \mathbb{R}^n$ e seja $b \in \mathbb{R}^n$ um vetor de medições tal que $b = Ax$, onde $A \in \mathbb{R}^{m \times n}$. O processo de obtenção de x a partir de um sinal desconhecido b é denominado codificação e o processo de recuperação de x a partir do vetor de medição b é denominado decodificação.

No problema

$$Ax = b, \tag{1}$$

existem situações onde $m = n$ e neste caso para realizar o processo de decodificação, basta resolver o sistema linear.

Porém, em muitas aplicações $m \ll n$ e com isso o problema (1) será um sistema possível indeterminado, i.e., existirão infinitas soluções para (1) e não seria possível realizar a recuperação do sinal. Nestes casos, para realizar a recuperação de x de forma eficaz, é razoável adicionar restrições ao problema (1) de forma a garantir a unicidade da solução.

Em geral, em situações práticas, b é obtida a partir de um sinal altamente esparso, ou seja, grande parte das componentes de x são nulas, e essa informação pode ser suficiente para se definir a restrição almejada.

Definindo $\|x\|_0$ como o número de elementos não nulos do vetor x , o problema (1) pode ser reformulado para a seguinte forma,

$$\min_x \{ \|x\|_0 \mid Ax = b \}, \tag{2}$$

ou seja, o sinal que será recuperado deverá satisfazer $Ax = b$ e ao mesmo tempo ser o mais esparso possível. Porém, encontrar a solução com maior número de componentes nulas é computacionalmente intratável para sistemas muito grandes, e portanto uma alternativa mostra-se necessária.

A alternativa que será utilizada para tornar o problema computacionalmente viável será minimizar a norma- ℓ_1 de x em vez de minimizar $\|x\|_0$. Com isso, podemos formalizar um novo modelo,

$$\min_x \{ \|x\|_1 \mid Ax = b \}, \tag{3}$$

que pode ser tratado computacionalmente como um problema de programação linear.

A escolha levanta algumas questões, como: Por que a norma- ℓ_1 recupera soluções esparsas? Sob quais condições a unicidade dessa solução é garantida?

Essas questões serão respondidas nas próximas seções, assim como a escolha da norma- ℓ_1 justificada.

3 Unicidade da solução

3.1 Análise inicial

Esta seção buscará apresentar condições sob as quais o problema (3) irá possuir solução única.

Para o desenvolvimento de tal argumentação, suponha que a solução buscada é k -esparsa, i.e., o vetor solução x possui no máximo k componentes não-nulas, sendo $k \ll m$.

Defina também Σ_k como sendo o conjunto de vetores k -esparsos em \mathbb{R}^n .

Vamos agora criar algumas impressões à respeito do núcleo de A . Seja x^* satisfazendo $Ax^* = b$. Sabemos que todas as outras soluções de $Ax = b$ são da forma $x^* + \eta$ sendo $\eta \in \mathcal{N}(A)$. Suponhamos também que $x^* \in \Sigma_k$ e vamos procurar condições para que x^* seja a única solução k -esparsa de $Ax = b$.

Supondo por contradição que existe um outro $x^{**} \in \Sigma_k$ tal que $Ax^{**} = b$, então temos que $A(x^* - x^{**}) = 0$, i.e., $x^* - x^{**} \in \mathcal{N}(A)$. Sabemos também que como $x^*, x^{**} \in \Sigma_k$, então a combinação linear $x^* - x^{**} \in \Sigma_{2k}$ (que também é não-nula).

Com isso concluímos que se $Ax = b$ possui mais do que um vetor k -esparsos como solução, então $\mathcal{N}(A)$ possui um vetor $2k$ -esparsos não-nulo.

Utilizando a contra-positiva desta afirmação nos fornece o seguinte lema.

Lema 1 *Suponha que $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$, i.e., todos os elementos não-nulos no núcleo de A possuem ao menos $2k + 1$ componentes não-nulos. Então qualquer solução k -esparsa de $Ax = b$ é única.*

Uma variação da condição $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$ é dada pelo seguinte lema.

Lema 2 *A condição $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$ é válida se e somente se cada subconjunto de $2k$ colunas de A são linearmente independentes.*

Prova [1] Suponha que $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$. Defina A_j como sendo a j -ésima coluna de A e $T \subseteq \{1, \dots, n\}$ seja qualquer subconjunto de índices com cardinalidade $|T| = 2k$. Considere o subconjunto $\{A_j : j \in T\}$ de $2k$ colunas de A . Se $\sum_{j \in T} x_j A_j = 0$ para alguns escalares x_j , então $Ax = 0$ para o vetor x com componentes x_j com $j \in T$ e $x_j = 0$ caso $j \notin T$. Assim $x \in \Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$, que implica em $x = 0$. Portanto $x_j = 0$ para todo j e então $\{A_j : j \in T\}$ é um conjunto linearmente independente.

Para a volta, suponha cada subconjunto $\{A_j : j \in T\}$ de $2k$ colunas de A é linearmente independente. Considere um vetor $x \in \Sigma_{2k} \cap \mathcal{N}(A)$ e seja $T \subseteq \{1, \dots, n\}$ um conjunto qualquer de $2k$ índices tais que se $x_j \neq 0$, então $j \in T$ (e $x_j = 0$ se $j \notin T$). Nós usamos isso para formar o subconjunto $\{A_j : j \in T\}$ de $2k$ colunas de A que satisfaça

$$\sum_{j \in T} x_j A_j = Ax = 0,$$

porque $x \in \mathcal{N}(A)$. Mas $\{A_j : j \in T\}$ é linearmente independente por hipótese, então $x_j = 0$ para cada $j = 1, \dots, n$, i.e., $x = 0$. Portanto $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$. \square

Porém, mesmo com ambos os lemas em mãos é computacionalmente impraticável verificar a condição $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$, portanto precisamos de alguma outra estratégia.

3.2 Propriedade Isométrica Restrita

A condição $\Sigma_{2k} \cap \mathcal{N}(A)$ diz que não existe vetor não-nulo $2k$ -esparso que satisfaça $Ax = 0$. Não perdemos generalidade se supormos o mesmo para vetores unitários (com respeito à norma-2), já que se $x \neq 0$, então $Ax = 0$ se e somente se $Au = 0$ com $u = x/\|x\|_2$. Então iremos procurar uma condição para que não exista vetor unitário não-nulo $2k$ -esparso que satisfaça $Au = 0$.

É razoável assumir que neste caso, existe uma constante $c_1 > 0$ tal que todos vetores unitários $2k$ -esparsos satisfaçam $\|Au\|_2^2 > c_1$. Pois $Au = 0$ se e somente se $\|x\|_2^2 = 0$.

Sendo assim, podemos enunciar o seguinte lema.

Lema 3 *Se existe uma constante positiva c_1 tal que $c_1 \leq \|Au\|_2^2$ para todos os vetores unitários $2k$ -esparsos, então $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$.*

Uma variação do Lema 3 que será útil a posteriori, quando formos analisar a efetividade da norma-1 para recuperar soluções esparsas pode ser construída considerando o fato de que se $c_1 \leq \|Au\|_2^2$ então sempre existirá c_2 tal que $\|Au\|_2^2 \leq c_2$ para qualquer vetor unitário $u \in \Sigma_{2k}$. Dessa forma, temos a seguinte desigualdade,

$$c_1 \leq \|Au\|_2^2 \leq c_2. \quad (4)$$

É razoável considerar que reescalando a matriz A por uma constante, a condição $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$ continuará se verificando. Sendo assim, podemos multiplicar convenientemente a desigualdade (4) por $2/(c_1 + c_2)$, reescalar a matriz A pelo coeficiente $\sqrt{2/(c_1 + c_2)}$ e definir $\delta = (c_2 - c_1)/(c_2 + c_1)$, tendo assim, a desigualdade,

$$1 - \delta \leq \|Au\|_2^2 \leq 1 + \delta. \quad (5)$$

Note que como $c_2, c_1 > 0$, temos que $0 < \delta < 1$.

Com isso, podemos definir a propriedade isométrica restrita.

Definição 1 *Uma matriz $A \in \mathbb{R}^{m \times n}$ satisfaz a propriedade isométrica restrita de ordem k se existe alguma constante $\delta_K \in (0, 1)$ tal que*

$$1 - \delta_k \leq \|Au\|_2^2 \leq 1 + \delta_k,$$

para todos os vetores unitários u k -esparsos.

Com esta definição, podemos criar um novo lema a partir do Lema 3.

Lema 4 *Se uma matriz A satisfaz a propriedade isométrica restrita de ordem $2k$ para algum $k \geq 1$, então $\Sigma_{2k} \cap \mathcal{N}(A) = \{0\}$ e qualquer vetor k -esparso é único.*

3.3 Propriedade Isométrica Restrita em matrizes aleatórias normais

Apesar de a propriedade isométrica restrita ser de difícil verificação, podemos desenvolver algo mais com a hipótese da matriz A possuir componentes aleatórios de distribuição normal.

Para tal situação, existe o seguinte teorema [1].

Teorema 1 *Seja A uma matriz $m \times n$ com $m = qn$ para algum $q \in (0, 1)$ e suas componentes sendo variáveis aleatórias independentes de uma distribuição normal de média 0 e variância $1/m$. Para qualquer $\epsilon \in (0, 1)$ fixado, $\delta \in (0, 1)$, e $m \geq 12/\delta$, a matriz A irá satisfazer a propriedade isométrica restrita de ordem k com constante $\delta_k = \delta$ com probabilidade de pelo menos $1 - \epsilon$ se n for escolhido grande o suficiente para satisfazer a desigualdade,*

$$-c_0qn + (k + 1/2) \ln(n) + c_1 \leq \ln(\epsilon),$$

onde $c_0 = \delta^2/144 - \delta^3/1296 > 0$ e $c_1 = k \ln\left(\frac{36\epsilon}{k\delta}\right) + \frac{1}{2} \ln(q/\pi)$.

Sendo assim, com o Teorema 1 podemos verificar que quando estamos lidando com matrizes de tamanho consideravelmente grande, teremos uma probabilidade quase certa de tal matriz possuir a propriedade isométrica restrita da ordem necessária para que a solução mais esparsa do problema $Ax = b$ seja única.

A demonstração do Teorema 1 é melhor abordada em [1].

4 Encontrando a solução esparsa de $Ax=b$ com a norma-1

O teorema que garante a suficiência de minimização da norma-1 para recuperação de sinais k -esparsos é o seguinte. [1]

Teorema 2 *Seja A uma matriz $m \times n$ que satisfaz a propriedade isométrica restrita de ordem $3k$ com constante $\delta_{3k} < \frac{1}{3}$. Então, a solução de,*

$$\min \|x\|_1 \quad \text{s.a.} \quad Ax = b,$$

recupera vetores k -esparsos.

(A demonstração do teorema é abordada em [1])

E pelo Lema 4, como A possui propriedade isométrica restrita de ordem $3k$, sabemos que tal vetor k -esparso é único.

Um fator a ser levado em conta ao tomar conhecimento de tal teorema, é de que tais condições garantem a convergência, porém ela ainda existe amplamente quando as hipóteses não são verificadas, garantindo assim um bom uso prático da detecção de compressão. Tomemos como exemplo o método de Newton, que possui hipóteses de convergência bem restritivas, porém na prática tais hipóteses não são sempre verificadas e o método converge mesmo assim.

Para resolver computacionalmente a minimização, é realizada uma reformulação para torna-la um problema clássico de programação linear. Para tal é tomado $x = u - v$ com $u, v \in \mathbb{R}^n$ e $z = [u^T | v^T]^T$, ambos vetores não-negativos, de forma que o problema se resumirá em,

$$\min \mathbf{1}_{2n}^T z \quad \text{sujeito à} \quad [A | -A]z = b.$$

5 Matriz de detecção

Muito é dito à respeito da matriz de detecção A associada à restrição do problema de minimização, porém, como ela seria em um exemplo numérico?

O teorema apresentado na Seção 3.3 é muito sugestivo quanto à essa pergunta, bastando tomar uma matriz $A \in \mathbb{R}^{m \times n}$ com elementos pertencentes à uma distribuição normal de média 0 e variância $1/m$ e a questão se resolveria. Porém, se fosse possível obter uma matriz com uma estrutura mais simples, de forma a simplificar as operações e as restrições do problema e que ainda obtivesse bons resultados na recuperação do sinal esparsos, seria de grande proveito.

Para a construção de tal matriz primeiro será definida a matriz B dada por,

$$B = 1^{n \times n} - 2I = \begin{bmatrix} -1 & 1 & 1 & \dots & 1 \\ 1 & -1 & 1 & \dots & 1 \\ 1 & 1 & -1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & -1 \end{bmatrix}, \quad B \in \mathbb{R}^{n \times n}.$$

A matriz $A \in \mathbb{R}^{m \times n}$ que será efetivamente utilizada nos problemas de detecção de compressão será dada por um subconjunto de m linhas da matriz B .

As vantagens da utilização da mesma em vez da matriz randômica normal estão, computacionalmente falando, na possibilidade de não guardar efetivamente a matriz A na memória (o que é de imensa ajuda já que a matriz A costuma atingir tamanhos muito grandes mesmo em aplicações simples), dada sua estrutura trivial, além de que, na multiplicação entre matrizes não é necessário realizar multiplicação termo-a-termo já que todos os elementos de A são unitários, fazendo com que a complexidade computacional do problema decaia significativamente.

Também é importante levantar que quando é realizada a recuperação do sinal através da matriz randômica normal o sinal acaba possuindo muito ruído, como pode ser observado na comparação feita pela Figura 1, enquanto a recuperação através da matriz construída acaba sendo bem mais "suave".

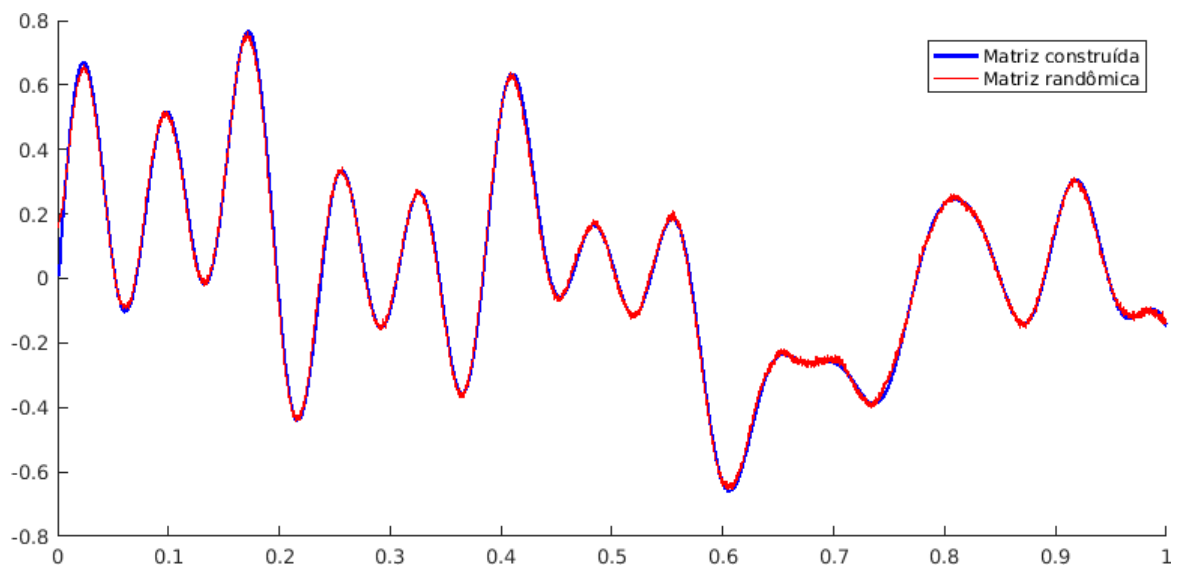


Figura 1: Comparação entre a recuperação de um sinal utilizando a matriz construída e a matriz randômica normal.

6 Aplicações

6.1 Compressão de dados

Uma das aplicações da detecção de compressão é a compressão de dados utilizando os conceitos estudados. Para fins de exemplificação tome a curva,

$$f(x) = \frac{1}{5}(\sin 49x + \sin 83x + \sin 79x + \sin 7x + \sin 17x), \quad (6)$$

e defina o sinal original como $x \in \mathbb{R}^{5000}$, de 5000 pontos da curva (6), igualmente espaçados no intervalo $[0, 1]$. O sinal está explicitado na Figura 2.

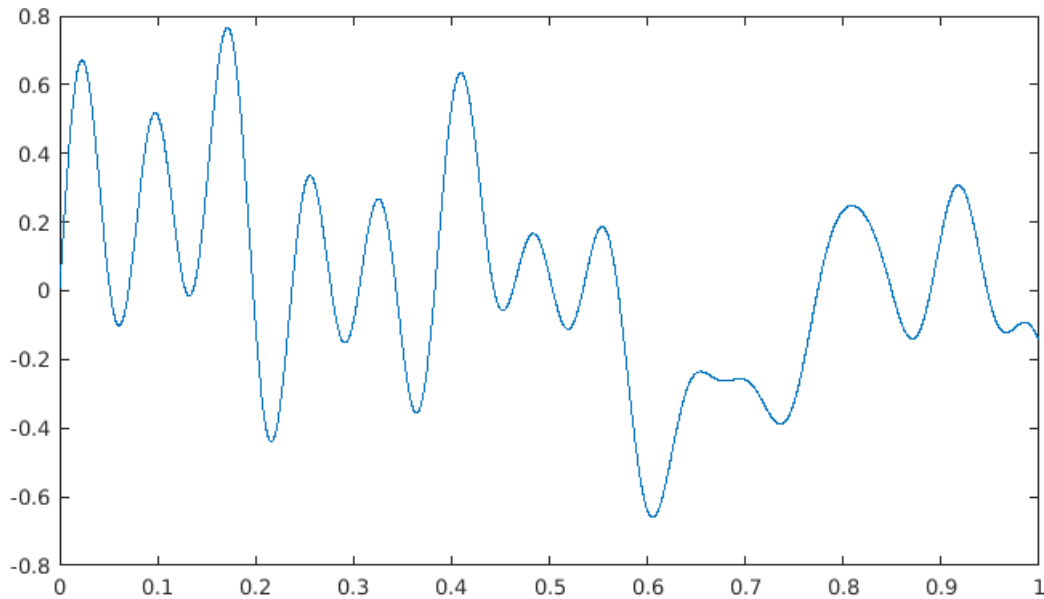


Figura 2: Sinal original retirado da curva (6).

Como facilmente constatado, o sinal x não é esparso e isso é uma condição necessária para que o mesmo possa ser recuperado via minimização com a norma- ℓ_1 após a compressão. Para resolver esse impasse se faz necessário algum tipo de transformação em x que possa ser revertida a posteriori e que torne x um vetor esparso. A transformação que será utilizada no exemplo, muito comumente usada neste tipo de problema, é a transformação discreta de cosseno (que denotaremos por \mathcal{C}).

A Figura 3 ilustra a aplicação da transformada \mathcal{C} ao sinal x .

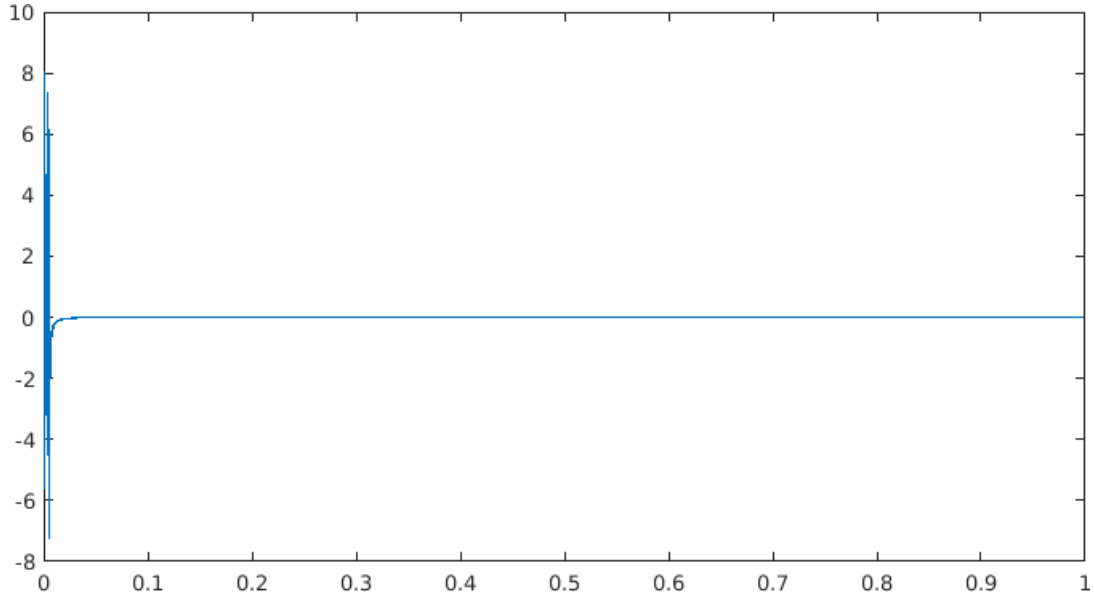


Figura 3: Sinal $\mathcal{C}(x)$, notoriamente esparso.

Para fins de generalização será utilizada a notação matricial da transformada \mathcal{C} , tomando $\Phi = \mathcal{C}(I)$ sendo $I \in \mathbb{R}^{n \times n}$ a matriz identidade. Dessa forma $\mathcal{C}(x)$ será denotado por Φx (e $\mathcal{C}^{-1}(x)$ por $\Phi^{-1}x$), onde Φ em um caso mais geral poderá ser qualquer transformação conveniente que torne x um sinal esparso.

A compressão do sinal x será dada por,

$$b = A\Phi x,$$

onde $A \in \mathbb{R}^{m \times n}$ é a matriz de detecção já abordada na Seção 5. Acontecerá uma redução de dimensão do sinal x de n para m (lembrando que $m \ll n$). Neste exemplo, será utilizado $m = 500$ (i.e. $m = n/10$), ou seja, a compressão será de 90%.

Dado que o sinal $s = \Phi x$ é esparso, para sua descompressão serão utilizados os conceitos estudados de detecção de compressão, tendo assim o problema,

$$\min \|s\|_1 \quad \text{s.a.} \quad As = b, \quad (7)$$

que, como visto na Seção 4, tem como problema de programação linear associado,

$$\min 1_{2n}^T z \quad \text{s.a.} \quad [A, -A]z = b,$$

onde $s(i) = z(i) - z(n+i)$ para todo $i \in \{1, \dots, n\}$.

Resolvido o problema (7) e obtido s basta tomar $x = \Phi^{-1}s$, onde Φ^{-1} será calculado com a transformada discreta de cosseno inversa \mathcal{C}^{-1} (sendo mais barato do que efetivamente inverter uma matriz quadrada com os métodos habituais).

Também é importante levantar que ao resolver o PL, poderia ser utilizado $[A\Phi^{-1}, -A\Phi^{-1}]$ em vez de $[A, -A]$, com fim de já obter a solução final (o sinal x) ao fim da minimização, porém tal substituição pode muitas vezes tirar a "trivialidade" da minimização com a matriz A original, como já visto na Seção 5.

Como resultado da descompressão de x , seguindo as rotinas explicitadas acima, a Figura 4, que visivelmente recupera perfeitamente o sinal original.

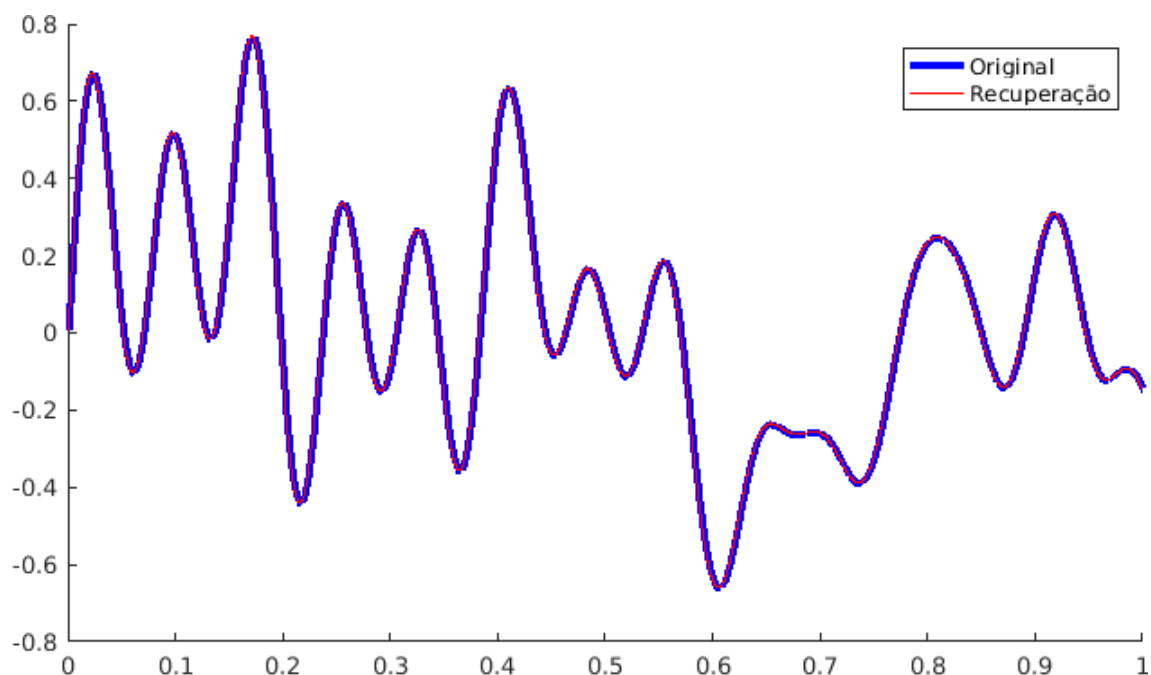


Figura 4: Descompressão do sinal original através da detecção de compressão.

É importante salientar que o tempo de processamento com a matriz construída é aumentado drasticamente quando utiliza-se a transformação $\Phi^{-1}A$ na formulação do problema (como já comentado), já que a matriz perde sua estrutura "simples".

O mesmo procedimento pode, por exemplo, ser realizado em imagens digitais. Considerando uma imagem digital como uma matriz de valores, onde cada elemento da matriz define a posição e a intensidade de cada pixel, pode ser construído um sinal associado à imagem concatenando todas as colunas de pixel em um único vetor x .

Tomando como exemplo a região explicitada da Figura 5,

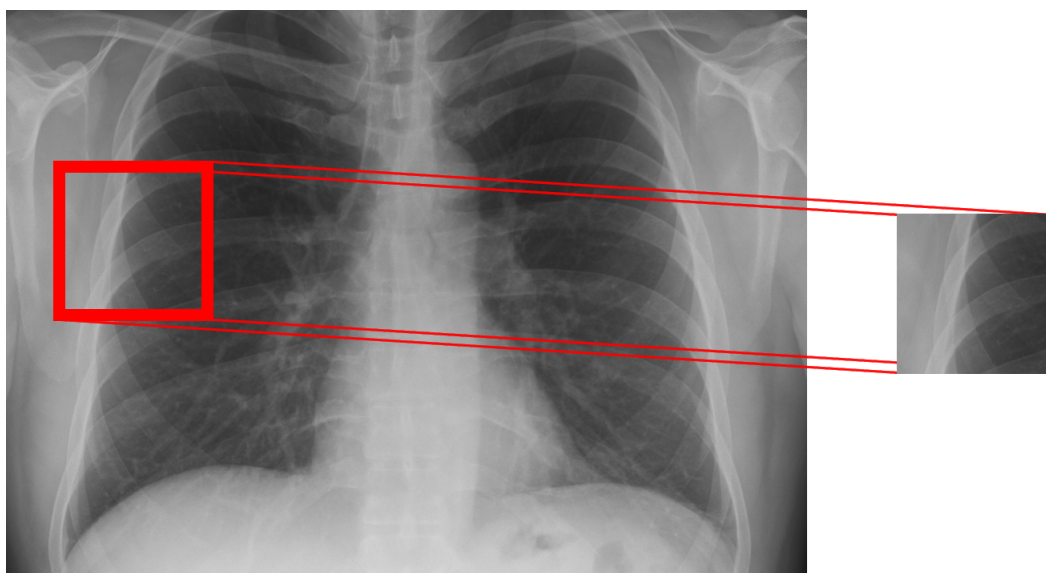


Figura 5: Figura de um exame raio X de uma costela. (Retirada de https://simple.wikipedia.org/wiki/Chest_x-ray, com direitos de uso não comercial com modificação.)

com 200×200 pixels, pode ser construído um sinal de dimensão 40000, que é ilustrado pela Figura 6 bem como o resultado de sua descompressão após a compressão com taxa de 90%.

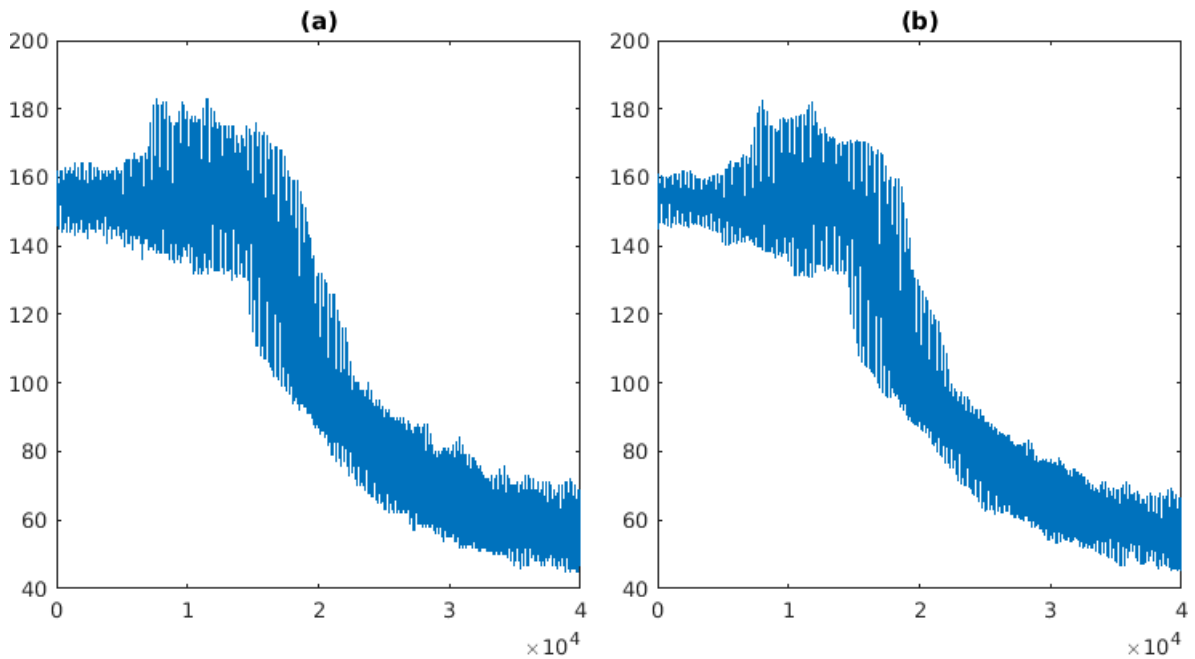


Figura 6: **(a)** Sinal original associado à imagem da Figura 5 **(b)** Sinal descomprimido (após compressão) associado à imagem da Figura 5.

Após reescalonar os sinais ilustrados, temos o resultado gráfico explicitado na Figura 7.

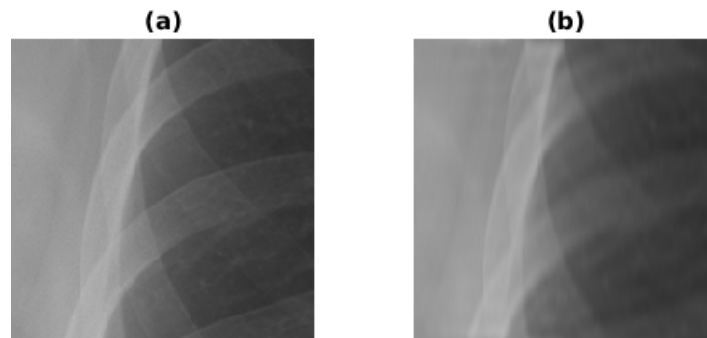


Figura 7: **(a)** Imagem original. **(b)** Imagem comprimida com taxa de 90%.

É explícito que a imagem foi reconstruída de forma a manter sua estrutura original, apesar de ter sofrido uma leve suavização visual.

É importante salientar que o impacto da detecção de compressão no campo da compressão de dados não é dado pela sua taxa de diminuição de informação armazenada e sim pelo fato de que a compressão pode ser feita simultaneamente com a detecção dos dados, dado que cada píxel pode ser adicionado à compressão de forma paralela e contínua, fazendo com que em nenhum momento se faça necessário o armazenamento da imagem em estado "cru".

6.2 Reconstrução de dados

É de interesse das aplicações práticas a adição de perturbações ao sinal original. Aplicando ruído randômico de intensidade $\pm 0,05$ ao sinal x , obtem-se o sinal x^* ilustrado pela Figura 8.

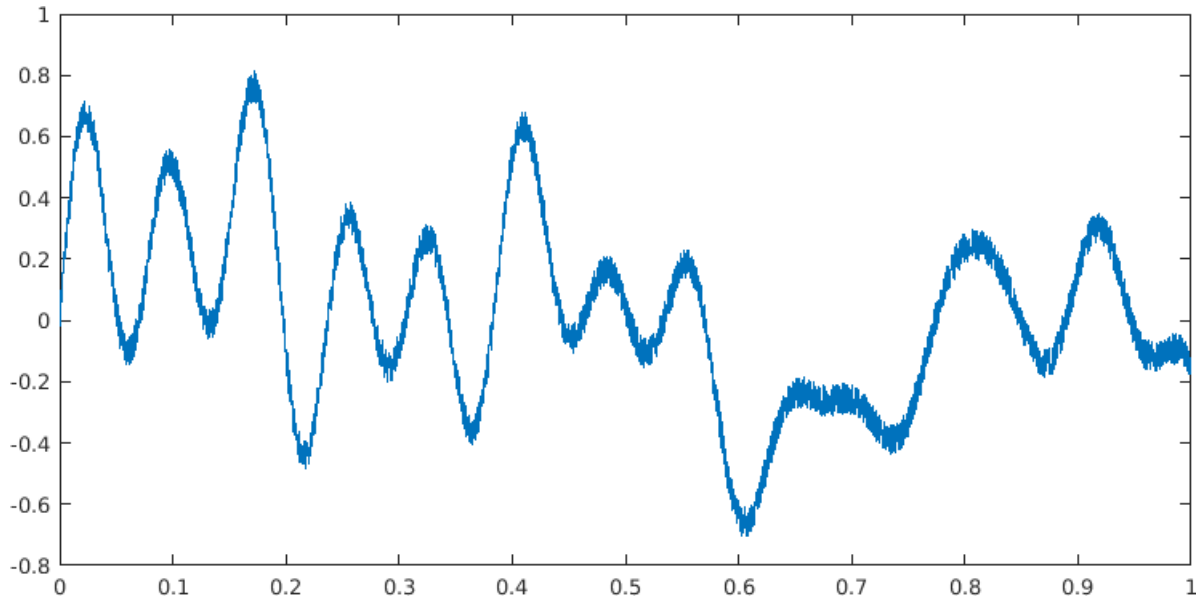


Figura 8: Sinal x com perturbações randômicas, que será denotado por x^* .

A transformação Φx^* é ilustrada pela Figura 9 e como facilmente constatável, não é um sinal esparso, possuindo várias entradas muito próximas de 0, porém nunca nulas.

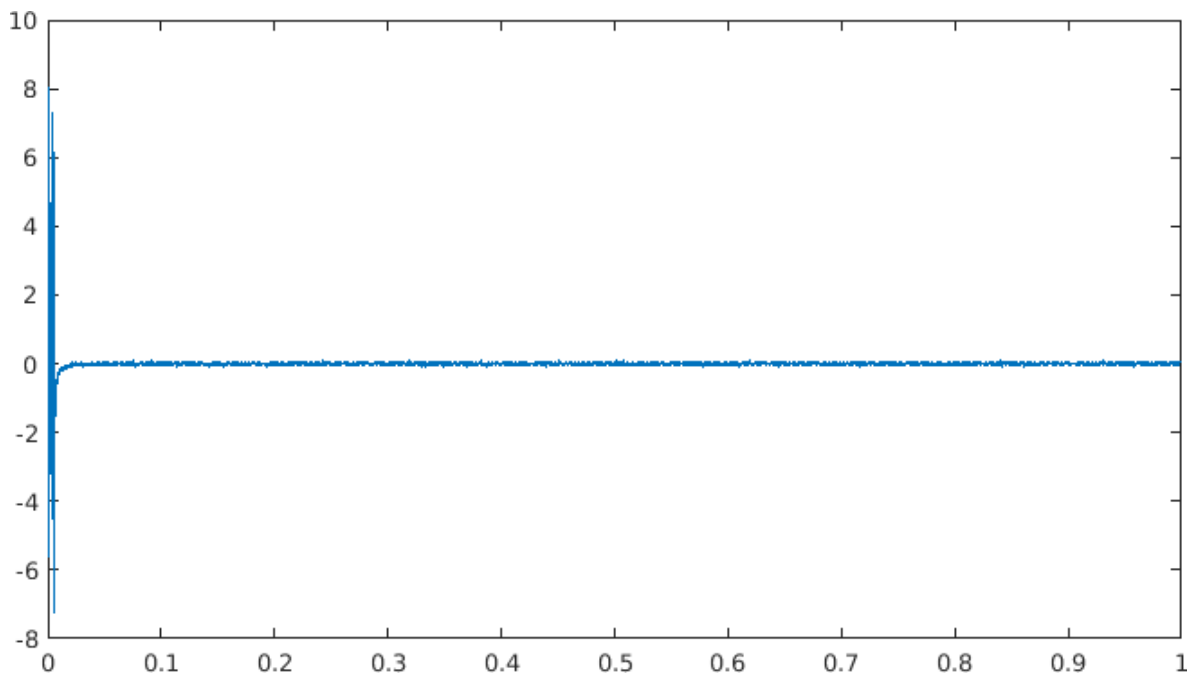


Figura 9: Sinal Φx^* , não esparso, com muitos elementos quase nulos.

Após a compressão, a recuperação do sinal Φx^* irá retornar um sinal esparso, como indica a teoria estudada. Usualmente a esparsidade surge das entradas que anteriormente eram próximas de zero, fazendo com que o sinal ilustrado pela Figura 10 seja muito semelhante ao do caso sem ruído (Figura 3).

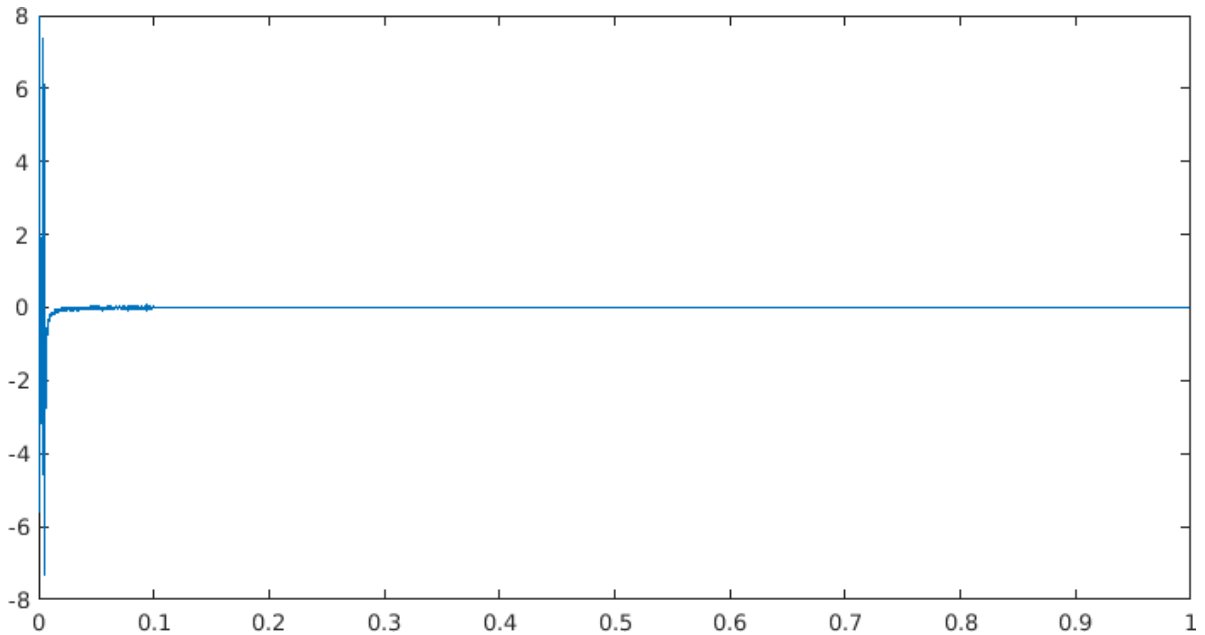


Figura 10: Sinal Φx^* após a restauração utilizando detecção de compressão.

Com tal devida semelhança, é razoável esperar que a transformação inversa $\Phi^{-1}\Phi x^*$ seja muito parecida com o sinal original sem ruído (2). A comparação é ilustrada pela Figura 11, indicando que de fato houve uma "restauração" do sinal original a partir de um sinal com perturbações.

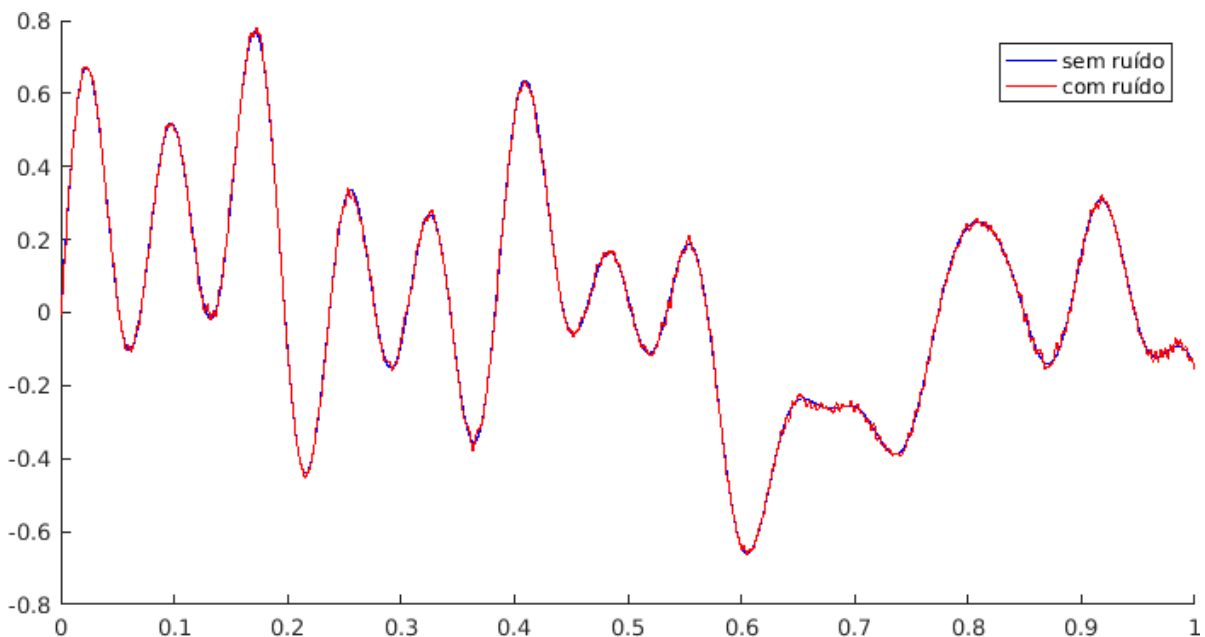


Figura 11: Comparação entre a recuperação do sinal com ruído e sem ruído.

7 Considerações finais

A segunda parte do projeto, a qual busca aplicações reais dos conceitos teóricos estudados, ainda se estenderá buscando estudar aspectos práticos da reconstrução de dados em situações reais, assunto que não foi completamente abordado nesta versão do relatório.

Todos as situações até então estudadas e aplicadas retornaram resultados coerentes com a bibliografia estudada, justificando e contextualizando de forma satisfatória as aplicações da detecção de compressão.

8 Referências Bibliográficas

- [1] K. Bryan & T. Leise, Making Do with Less: An Introduction to Compressed Sensing, **SIAM Review**, 55 (2013), 547-566.
- [2] E.J. Candès, J.K. Romberg & T. Tao, Stable Signal Recovery from Incomplete and Inaccurate Measurements, **Communications on Pure and Applied Mathematics**, 59 (2006), 1207-1223.
- [3] Y.C. Eldar & G. Kutyniok, **Compressed Sensing- Theory and Applications**, Cambridge, 2012.
- [4] S. Foucart & H. Rauhut, **A Mathematical Introduction to Compressive Sensing**, Birkhäuser, 2010.
- [5] Y. Zhang, Theory of Compressive Sensing via ℓ_1 -Minimization: a Non-RIP Analysis and Extensions, **Journal of the Operations Research Society of China**, 1 (2013), 79-105.
- [6] M. Elad, **Sparse and Redundant Representations**, Springer, 2010.
- [7] D. Donoho, For most large underdetermined systems of linear equations, the minimal ℓ_1 -norm solution is also the sparsest solution, **Comm. Pure Appl. Math.** 59 (2006), pp. 797-829.
- [8] K. Hayashi, M. Nagahara & T. Tanaka, A User's Guide to Compressed Sensing for Communications Systems, **IEICE Trans. Commun.** Vol E96-B, No. 3, 2013, pp. 685-712