Ricardo Antonio Zanetti

Separação de Eventos Sísmicos por Métodos de Decomposição de Sinais

CAMPINAS
2013

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação

Ricardo Antonio Zanetti

Separação de Eventos Sísmicos por Métodos de Decomposição de Sinais

Dissertação de Mestrado apresentada ao programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas para obtenção do título de Mestre em Engenharia Elétrica, na área de Telecomunicações e Telemática.

Orientador: Prof. Dr. João Marcos Travassos Romano

Co-orientador: Prof. Dr. Leonardo Tomazeli Duarte

Este exemplar corresponde à versão final da dissertação defendida pelo aluno Ricardo Antonio Zanetti e orientada pelo Prof. Dr. João Marcos Travassos Romano.

CAMPINAS
2013

Ficha Catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

	Zanetti, Ricardo Antonio, 1978-
Z16s	Separação de eventos sísmicos por métodos de decomposição de sinais / Ricardo Antonio Zanetti. – Campinas, SP : [s.n.], 2013.
	Orientador: João Marcos Travassos Romano. Coorientador: Leonardo Tomazeli Duarte. Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.
	1. Método sísmico de reflexão - Processamento de dados. 2. Decomposição (Matemática). 3. Processamento de sinais - Técnicas digitais. 4. Processamento de sinais - Métodos estatísticos. 5. Ondas sísmicas - Processamento de dados. I. Romano, João Marcos Travassos,1960-. II. Duarte, Leonardo Tomazeli. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Seismic events separation by means of signal decomposition
Palavras-chave em inglês:
Seismic reflection method - Data processing
Mathematical decomposition
Signal processing - Digital techniques
Signal processing - Statistical methods
Seismic waves - Data processing
Área de concentração: Telecomunicações e Telemática
Titulação: Mestre em Engenharia Elétrica
Banca examinadora:
João Marcos Travassos Romano [Orientador]
Allan Kardec Duailibe Barros Filho
Renato da Rocha Lopes
Data de defesa: 05-08-2013
Programa de Pós-Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Ricardo Antonio Zanetti

Data da Defesa: 5 de agosto de 2013

Título da Tese: "Separação de Eventos Sísmicos por Métodos de Decomposição de Sinais"

Prof. Dr. João Marcos Travassos Romano (Presidente): ______________________

Prof. Dr. Allan Kardec Duailibe Barros Filho: ______________________

Prof. Dr. Renato da Rocha Lopes: ______________________

Dedico esta dissertação aos meus pais, pelo suporte, incondicional, em todas as etapas da minha vida.

Agradecimentos

A conclusão deste trabalho só foi possível graças à ajuda de diversas pessoas. Algumas que já me conheciam há um bom tempo e outras que tive o prazer de conhecer durante esta jornada. Todas elas foram importantes, cada uma de uma maneira. Portanto, tentarei ser sucinto apesar de merecerem homenagens mais apropriadas.

Durante todo esse tempo eu pude sentir a intervenção divina. As dificuldades foram grandes no início e sei que Deus, em sua infinita sabedoria, deu-me as oportunidades e colocou as pessoas certas em meu caminho. Sendo assim, primeiramente agradeço a Ele e à Sua providência divina. "*Isto é uma ordem: sê firme e corajoso. Não te atemorizes, não tenhas medo, porque o Senhor está contigo em qualquer parte para onde fores.*", (Js 1,9).

Agradeço ao professor João Marcos Travassos Romano pela confiança em mim depositada para este trabalho, além da paciência e compreensão nos momentos difíceis. Foi uma honra ter o senhor como orientador.

Meus agradecimentos aos meus amigos da turma de graduação em Engenharia Elétrica da Unicamp de 2000: a Everton Zaccaria Nadalin e André Kazuo Takahata, por me convencerem a participar dessa incrível família que é o laboratório DSPCom e por sempre estarem dispostos a me ajudar com minhas dúvidas (como sempre...); a Leonardo Tomazeli Duarte, por ter a coragem de me coorientar e por tê-lo feito com a mesma precisão com que me dava assistências nas quadras de basquete; e a Diogo Coutinho Soriano e Filipe Ieda Fazanaro pelo apoio e por contribuírem para que eu me sentisse de novo na graduação.

Agradeço aos amigos do DSPCom: a Kenji Nose Filho, Rafael Ferrari, Rafael Krummenauer, Tiago Tavares Leite Barros e Marcos Ricardo Covre pelas sugestões e por toda ajuda durante o meu trabalho; aos professores Renato da Rocha Lopes e Romis Ribeiro de Faissol Attux pelos conselhos e por sempre estarem dispostos a nos ajudar com nossas dúvidas; e também a todos os outros amigos do DSPCom pelo companheirismo e pela rica troca de ideias sobre os mais diversos assuntos. Considerem-se todos homenageados!

Finalmente, eu gostaria de agradecer à minha família: meu pai, José Ticiano Zanetti; minha mãe, Nílcia Iamondi Zanetti; e minha irmã, Tais Helena Zanetti. Obrigado por tudo!

"*Do, or do not. There is no try.*"

Yoda, *Star Wars: The Empire Strikes Back*

Resumo

A geofísica de exploração reúne diversas técnicas de análise e processamento de dados sísmicos, com o objetivo específico de encontrar depósitos de elementos de interesse econômico ou científico no subsolo. Muitas destas técnicas são também utilizadas em processamento de sinais digitais. Da mesma forma, técnicas recentemente apresentadas para processamento digitais de sinais podem ser aplicadas no processamento de dados sísmicos com o objetivo de remoção de ruídos e separação de eventos nas imagens que serão interpretadas por um geólogo. Neste caso específico, apresentamos a técnica da RPCA (Análise de Componentes Principais Robusta) para extração de ruído de ondas Rayleigh, também conhecido como *ground roll*.
Neste trabalho, comparamos o desempenho da técnica de decomposição de sinais RPCA a outras técnicas normalmente utilizadas nesta tarefa específica, como transformada Radon, transformada *f-k*, SVD (Decomposição em Valores Singulares) e SVD-ICA (Análise de Componentes Independentes). Em um primeiro momento, as análises foram feitas aplicando as diferentes técnicas em dados sintéticos que simulam eventos sísmicos, *ground roll* e outros ruídos. Em seguida, de forma análoga, foram utilizados dados reais que passaram por todo o pré-processamento necessário. Para os diferentes cenários, foram obtidos resultados que se apresentaram equivalentes ou até mesmo melhores em relação às outras técnicas de decomposição utilizadas, mostrando que a RPCA pode ser uma ferramenta considerável em certos casos.

Palavras-chave: Dado Sísmico. Separação de Sinais. Decomposição em Valores Singulares. Análise de Componentes Principais Robusta.

Abstract

The exploration geophysics combines several analysis and processing techniques of seismic data, with the specific aim of finding underground reservoir of economic and scientific concernment. Many of these techniques are also used in digital signal processing. In the same way, recently introduced techniques for digital signal processing can be applied in seismic data processing, seeking noise removal and event separation on images that will be interpreted by a geologist. In this specific case, we introduce the RPCA (Robust Principal Component Analysis) techinic for Rayleigh waves extraction, also known as ground roll.
In this work, we compare the performance of RPCA decomposition technique to other techniques generaly used in this specific task, like Radon transform, f-k transform, SVD (Singular Value Decomposition) and SVD-ICA (Independent Component Analysis). At first, the analyses were done applying the different techniques to synthetic data that simulate seismic events, ground roll and other noises. Then, similarly, real data were used after the necessary pre-processing. For the different scenarios, we obtained results that showed them similar or even better compared to the other used decomposition techniques, showing that RPCA may be a considerable tool in specific cases.

Key-words: Seismic Data. Signal Separation. Singular Value Decomposition. Robust Principal Component Analysis.

Lista de Figuras

Conteúdo

Introdução

O imageamento de subsuperfície é de suma importância em diversos problemas práticos. Em um contexto de extração de hidrocarbonetos, é imperativo extrair o conjunto de informações *a priori* mais acurado possível sobre a área em estudo. Os custos envolvidos no processo de prospecção são muito elevados; portanto, a etapa de processamento dos dados sísmicos digitalizados ganha um papel fundamental nessa tarefa. Em *geofísica de exploração* temos diversos métodos de aquisição que podem ser aplicados de acordo com os objetivos do estudo. A sísmica de reflexão reúne o conjunto de métodos mais adequado quando buscamos informações que podem levar à descoberta de depósitos de hidrocarbonetos, seja na forma de gás ou óleo.

Dentre estes métodos está a geração de um pulso sísmico no solo e a captação das reflexões deste pulso nas diferentes camadas do subsolo por meio de sensores instalados na superfície. Estes dados digitalizados contêm a informação de interesse misturada a diversas categorias de ruídos coerentes e incoerentes. As técnicas de processamento de dados sísmicos buscam separar os sinais de interesse do restante do ruído. Em específico, para o caso do ruído causado por ondas de *ground roll* (um ruído coerente de alta magnitude), existem algumas técnicas específicas baseadas em filtragem de espectro e outras baseadas em métodos de decomposição de sinais.

Na presente dissertação, é abordado um conjunto de técnicas de decomposição de sinais que exploram, dentre outras propriedades de sinal, a coerência presente em dados sísmicos ruidosos, digitalizados na forma de sismogramas. Em especial, dar-se-á atenção a um novo paradigma de filtragem conhecido como Análise de Componentes Principais Robusta (*Robust Principal Component Analysis* - RPCA). Esta abordagem pode ser vista como uma extensão da clássica decomposição estatística Análise de Componentes Principais (*Principal Component Analysis* - PCA), porém menos suscetível a sinais gravemente corrompidos. Também estudamos o uso da Decomposição em Valores Singulares, já usada para este mesmo propósito em outros trabalhos (Porsani, Silva, Melo & Ursin 2010) e sua associação ao paradigma de separação cega de fontes Análise de Componentes Independentes (Independent Component Analysis - ICA) (Vrabie, Mars & Lacoume 2004).

Com o intuito de situar o desempenho das técnicas analisadas, foram conduzidas simulações em diferentes situações. Primeiramente, emulamos características encontradas em seções sísmicas em dados artificiais, de modo a observar isoladamente o desempenho das técnicas aplicadas em um ambiente controlado. Em seguida, aplicamos as técnicas em um conjunto de dados reais corrompido por ruído de *ground roll*, além de realizar comparações com o método de filtragem por Transformada *f-k* — a ferramenta mais comumente utilizada na extração deste tipo de ruído.

Os resultados obtidos permitem uma interessante análise dos métodos de decomposição aplicados a este problema específico, e fornecem outras diversas possibilidades de aplicação das mesmas no contexto de geofísica de exploração.

Organização deste Trabalho

Capítulo 1: Neste capítulo introdutório são apresentados alguns conceitos utilizados em processamento digital de sinais sísmicos e que serão citados no decorrer do nosso trabalho. Temos também uma visão geral das etapas necessárias a se obter uma visualização de uma seção sísmica, algo muito próximo à realidade das camadas inferiores do subsolo. É introduzido também o principal problema abordado neste trabalho: a extração do ruído provocado por ondas de rolamento superficial (*ground roll*). Alguns dos métodos mais utilizados na indústria de geofísica de exploração para a extração deste tipo de ruído também são abordados de forma introdutória, de maneira a preparar o leitor para as novas propostas deste trabalho nas seções posteriores.

Capítulo 2: Contém a apresentação e uma curta introdução aos métodos de decomposição de sinais sísmicos utilizados neste estudo. A já tradicional ferramenta de decomposição SVD é apresentada em primeiro lugar. A seguir, uma seção introdutória aos métodos de separação cega de fontes e uma breve descrição com exemplos das decomposições baseadas nas estatísticas dos sinais envolvidos, PCA e ICA. Encerrando o capítulo temos a apresentação de um técnica híbrida, também conhecida por aplicações em processamentos de sinais sísmicos, a SVD-ICA.

Capítulo 3: A nossa proposta de aplicação da decomposição de sinais por RPCA é apresentada neste capítulo. A técnica é introduzida junto de uma descrição das aplicações em que já é empregada na literatura. Também discutimos os dois principais algoritmos utilizados neste estudo: o GoDec e o SSGoDec.

Capítulo 4: Os principais resultados obtidos durante todo o trabalho de pesquisa são apresentados neste capítulo. Primeiramente, são utilizados dados sintéticos para a simulação de situações encontradas em processamento digital de sinais sísmicos. Cruzamento de eventos (ondas refletidas) e difrações são separados via os métodos de decomposição SVD, a SVD-ICA e por fim, a RPCA. Em seguida, as técnicas são aplicadas em dados reais pré-processados. Por fim, comparamos os resultados destas decomposições com o método clássico utilizado para a extração do ruído de *ground roll*, a filtragem por transformada *f-k*.

Conclusão: Nesta etapa, buscamos comentar os resultados de forma a mostrar as possibilidades de aplicação dos métodos de decomposição de sinais (e no caso específico, a RPCA) para a solução de problemas específicos de extração de ruídos em dados sísmicos. Discutimos também possíveis trabalhos futuros e questões que não foram tratadas neste trabalho.

Capítulo 1

O Problema da Separação de Eventos Sísmicos

Neste início de trabalho, são apresentados os conceitos básicos de processamento de sinais sísmicos, tais como os traços sísmicos, a organização dos mesmos em famílias e algumas das etapas de processamento às quais os dados são submetidos. Em seguida temos a descrição do problema de extração de ruído coerente e as principais técnicas utilizadas nesta tarefa.

1.1 Introdução à Sísmica de Reflexão

Um levantamento geofísico trata da obtenção das características geológicas de uma determinada área em estudo. A maneira como é feito o levantamento depende de diversos fatores, como por exemplo o meio-ambiente da área em questão, o objeto de busca ou suas características de interesse, se o método de prospecção envolve os campos naturais da Terra ou fontes artificiais, os aspectos econômicos e científicos envolvidos, dentre outros.

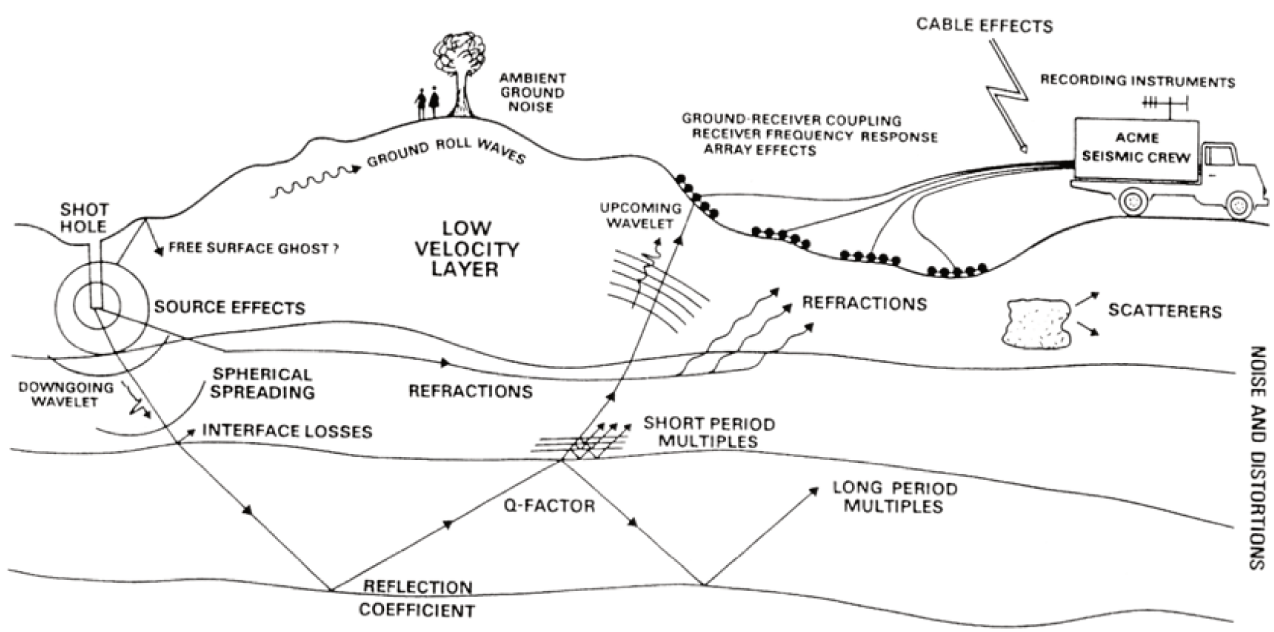

Figura 1.1: Aquisição de dados sísmicos para exploração (reimpresso de (Reynolds 1997)).

Em sísmica de reflexão, a geologia de uma determinada região é analisada usando-se ondas sísmicas, produzidas por uma fonte artificial. No caso de uma aquisição de dados terrestre, esta fonte pode ser a explosão de uma carga de dinamite enterrada próxima à superfície. Uma instalação de prospecção, bem como diversos fenômenos e elementos que envolvem a aquisição de dados sísmicos, é representada na Figura 1.1. As ondas sísmicas geradas na detonação penetram no subsolo e suas reflexões, devido às diferentes camadas geológicas, são captadas por sensores (chamados *geofones*, no caso de levantamentos terrestres ou *hidrofones*, no caso de levantamentos marítimos) dispostos em arranjos na superfície, como na Figura 1.2.

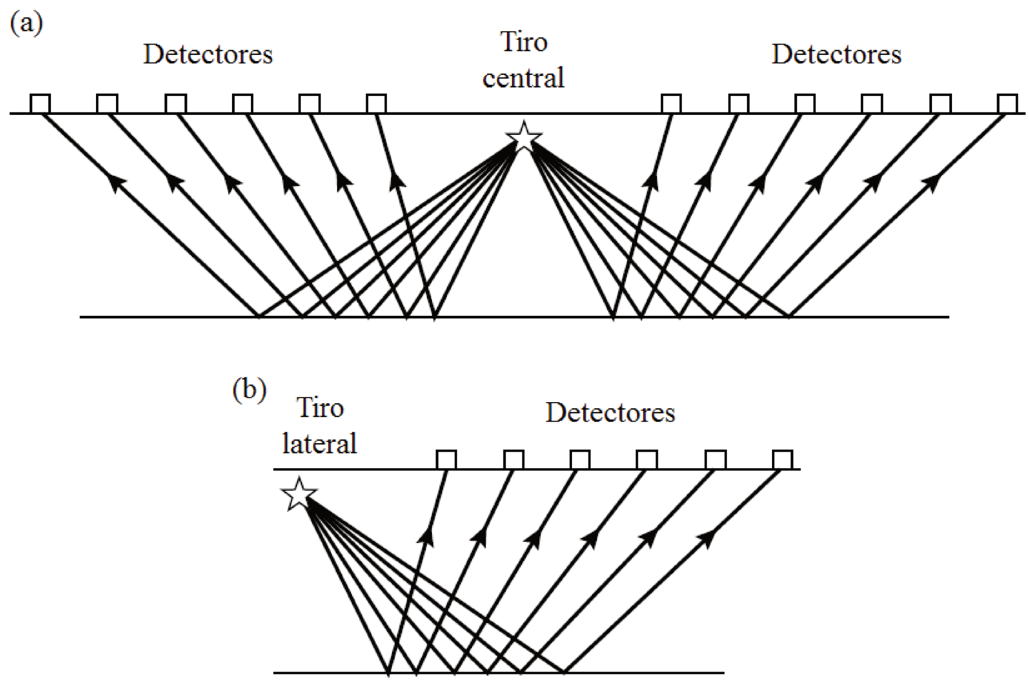

Figura 1.2: Arranjos de geofones para captação de reflexões: (a) Tiro central; (b) Tiro lateral (reimpresso de (Kearey et al. 2002)).

O *traço sísmico*, que contém as informações sobre as características das camadas geológicas, é o sinal obtido das reflexões de parte da energia incidente do pulso sísmico em cada um dos sensores. Estas reflexões são captadas em instantes diferentes pelos receptores, em virtude dos diferentes coeficientes de propagação e reflexão das camadas inferiores do subsolo. Assumindo que o formato do pulso permanece inalterado durante a penetração do mesmo pelas diferentes camadas, o traço sísmico pode ser visto como uma convolução do pulso de entrada com uma série conhecida como função refletividade (Kearey et al. 2002). A relação entre a forma de onda captada e a estratificação das camadas geológicas pode ser vista na Figura 1.3. A este modelo convolutivo poderíamos somar um termo de ruído que é constituído de todos os sinais indesejados que também são captados pelos sensores. Esta seria a descrição de um traço sísmico real.

Os traços sísmicos podem ser arranjados de diferentes formas, de acordo com a necessidade de visualização ou do procedimento de processamento sísmico. Isto é possível graças aos dados contidos no *header* (cabeçalho) de cada traço sísmico. Podemos visualizá-los agrupados por tiro comum (*Common Shot* - CS), por exemplo, que facilita uma visão primária dos eventos em relação ao deslocamento na linha, ou por ponto comum em profundidade (*Common-depth Point* - CDP) (Yilmaz 2001). A Figura 1.4 ilustra estas duas configurações de organização. Em um outro exemplo, a correção dos dados por sobretempo normal (*Normal-moveout* - NMO) é

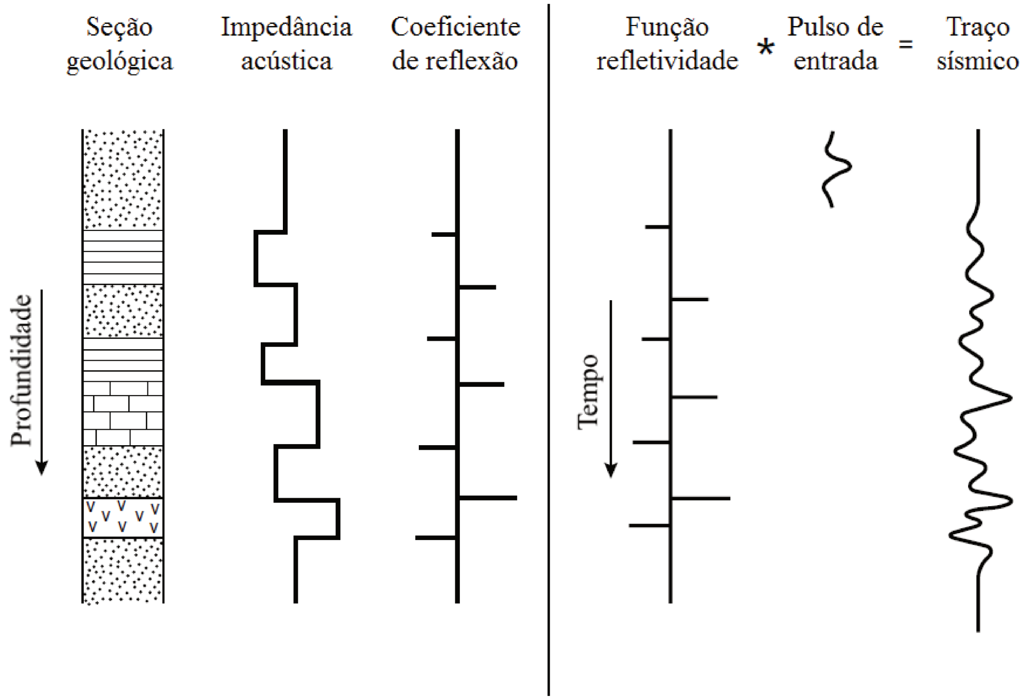

Figura 1.3: Modelo convolutivo para a composição do traço sísmico (reimpresso de (Kearey et al. 2002)).

aplicada aos dados agrupados for CDP e é uma etapa preparatória para para o empilhamento (stacking) das seções, para posterior visualização dos dados em profundidade e não mais em tempo (Figura 1.5) (Yilmaz 2001).

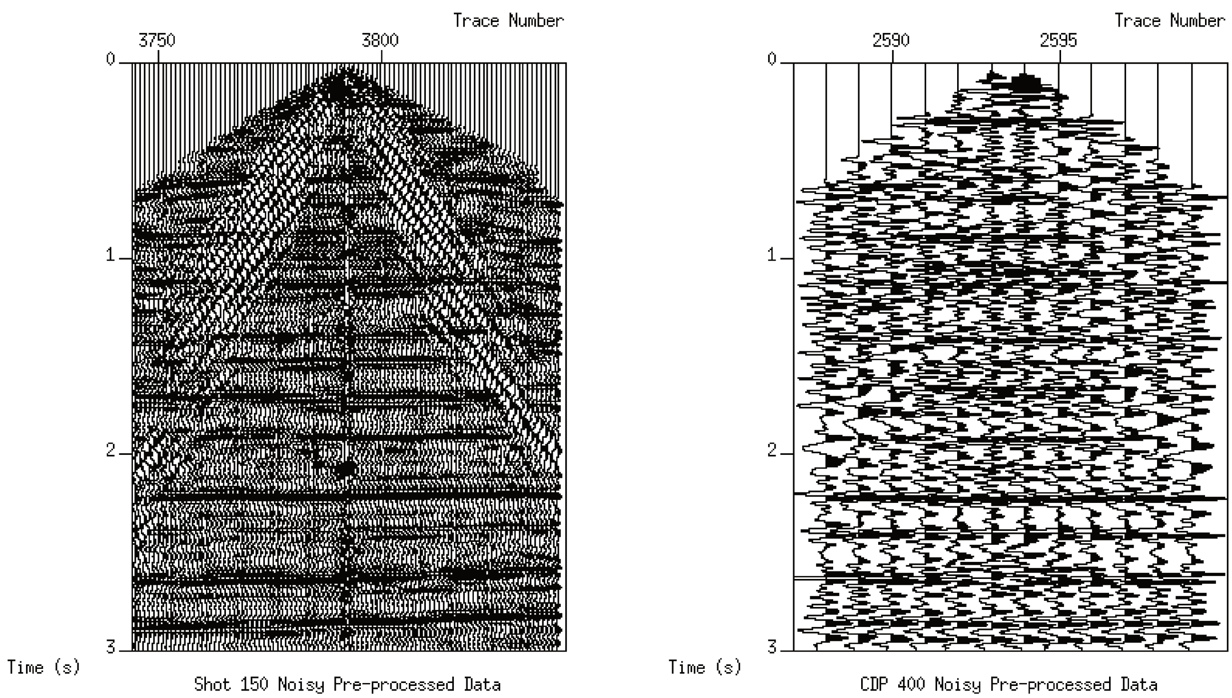

Figura 1.4: Traços sísmicos ordenados por CS à esquerda, e por CDP à direita.

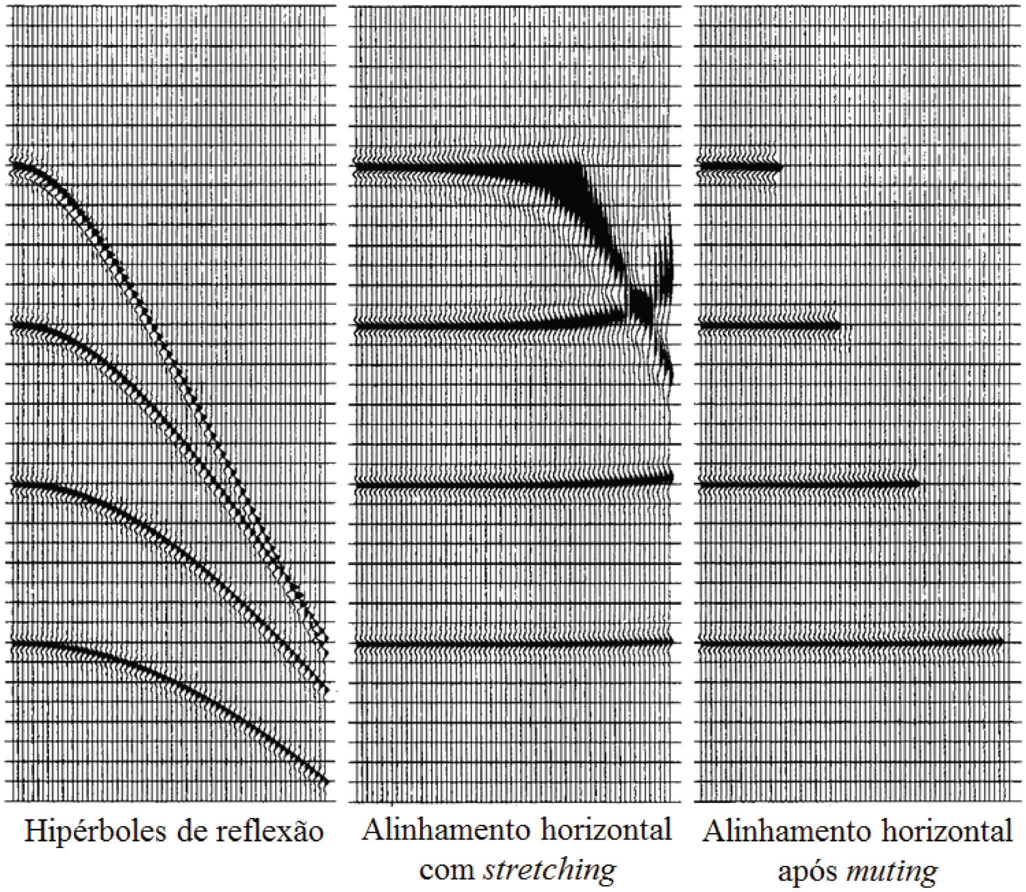

Figura 1.5: Correção NMO de uma seção CDP (adaptado de (Yilmaz 2001)).

A exploração de hidrocarbonetos se dá em camadas geológicas a profundidades que variam de 100 m a 5 km (Van der Kruk 2003). Os *eventos sísmicos* (refletores, interfaces entre camadas, falhas geológicas, entre outras estruturas) podem oferecer indícios de depósitos desses elementos de alto valor econômico e científico. Dada a importância da interpretação das seções sísmicas, é necessário que os dados digitais recolhidos passem por diversas etapas de processamento de maneira a eliminar ruídos, excluir traços comprometidos por sensores com defeito, efetuar correções relativas ao perfil do terreno e outras análises e procedimentos que permitam, no final, obter uma imagem que ilustre uma seção ou uma área representando a estrutura geológica da região, tal como na Figura 1.6. Esta é uma imagem obtida após todo o processamento de eliminação de ruídos, correções de amplitude, empilhamento das seções organizadas em famílias CDP ou de ponto médio comum (*Common-midpoint* - CMP) e migração dos dados.

Devido às inúmeras variáveis impostas pelo ambiente em estudo e também, é claro, às limitações de equipamentos de aquisição de informação, é impossível obter dados livres de ruídos. Existem inúmeras técnicas para recuperação de dados com ruído descorrelacionado, com baixa ou alta relação sinal-ruído. Entretanto, há situações onde os dados são corrompidos por ruídos coerentes, com amplitudes iguais ou muito superiores ao sinal desejado. É o caso, por exemplos dos diversos tipos de múltiplas de reflexão, difrações e ondas de superfície. Nessas situações é necessário recorrer a técnicas específicas de filtragem ou separação de sinais.

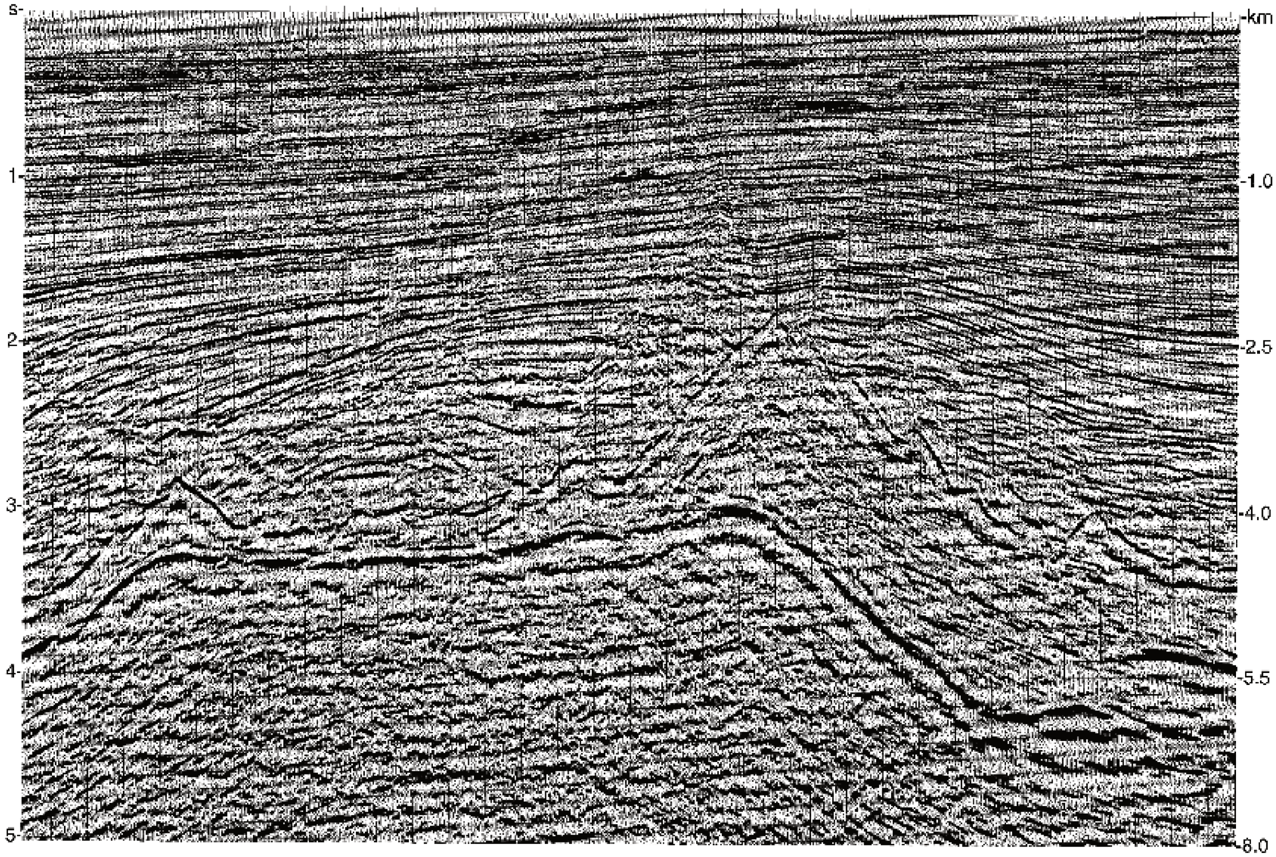

Figura 1.6: Seção sísmica do Golfo do México, usada para exploração de hidrocarbonetos (reimpresso de (Yilmaz 2001)).

1.2 Ondas de Rolamento Superficial (*Ground Roll*)

O ruído causado por ondas de rolamento superficial (*ground roll*) é justamente um caso onde o sinal indesejado apresenta amplitude geralmente muito elevada em relação à informação de interesse. Além disso, este tipo de ruído apresenta um padrão bem definido, sendo facilmente reconhecido no dado — às vezes, eclipsando praticamente toda a informação, sendo necessária a aplicação de um ajuste de amplitude para a visualização dos eventos. Na Figura 1.4 ele pode ser melhor observado na ilustração à esquerda, com os traços sísmicos organizados por CS, representado pelas linhas inclinadas e aparentemente paralelas, que se originam na fonte e se propagam até aproximadamente 2 segundos.

As ondas de *ground roll* seguem os padrões das ondas Rayleigh (Rayleigh *waves*). Estas ondas se propagam ao longo de uma superfície livre ou ao longo da interface entre dois meios sólidos não similares, sendo que as partículas afetadas descrevem uma órbita elíptica retrógrada num plano perpendicular à superfície (Kearey et al. 2002), de acordo com a Figura 1.7. A amplitude das ondas de *ground roll* decresce exponencialmente com a distância abaixo da superfície. Elas têm uma velocidade de propagação mais baixa que a das ondas principais de reflexão e, num meio-espaço homogêneo, deveriam ser não dispersivas (Kearey et al. 2002).

Dentre as técnicas utilizadas para a remoção dessas ondas, destacam-se: a *Transformada Radon*, que consiste em mapear os dados no domínio tempo-distância (t-x) para o domínio *tau-p* (τ-p) onde τ representa o tempo de intercepção e p a vagarosidade (*slowness*) (Yilmaz 2001).

Esta técnica, porém, é mais apropriada à extração de *múltiplas de reflexão*; a *Decomposição em Valores Singulares* (*Singular Value Decomposition* - SVD), uma técnica conhecida para a melhoria da relação sinal-ruído em dados sísmicos (Bekara & Van der Baan 2007), também pode ser usada para, especificamente, extrair as ondas de *ground roll* (Porsani et al. 2010) como veremos mais adiante; e aquela que talvez seja a técnica mais comumente utilizada na indústria para a remoção do *ground roll*, a *Filtragem por Transformada f-k*.

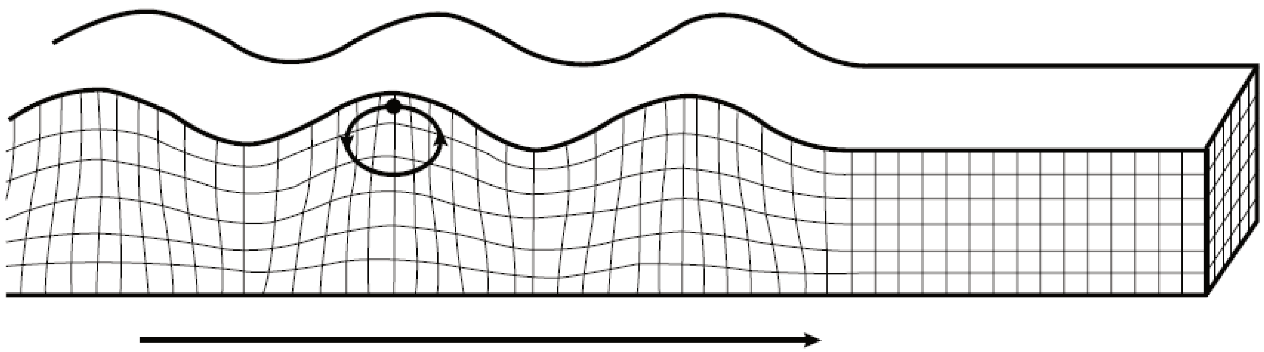

Figura 1.7: Representação da propagação de uma onda de *ground roll* em um meio sólido (reimpresso de (Kearey et al. 2002)).

1.3 Separação de Eventos por Transformada *f-k*

Transformada *f-k* é o nome normalmente utilizado em sísmica para se referir à *Transformada Bidimensional de Fourier*, que mapeia os dados no domínio tempo-distância (t-x) para o domínio frequência-número de onda (f-k), também conhecido como *Espectro f-k*. A frequência f indica o número de oscilações por segundo, enquanto que o número de onda k indica o número de comprimentos de onda por metro ao longo do eixo horizontal (alguns autores definem k como o número de comprimentos de onda ao longo do eixo horizontal multiplicado por 2π) (Van der Kruk 2003). Portanto, dada uma função $r(t,x)$, sua transformada *f-k* pode então ser descrita como:

$$\mathcal{F}_{2D}\{r(t,x)\} = \mathcal{F}_x\big\{\mathcal{F}_t\{r(t,x)\}\big\} = R(\upsilon_t, \upsilon_x), \qquad (1.1)$$

onde $\mathcal{F}\{\cdot\}$ denota a transformada de Fourier unilateral, υ_t é a frequência temporal e υ_x é a frequência espacial. Com isto, temos que $\omega = \upsilon_t$, que fornece a pulsação e $k = \upsilon_x$, que fornece o número de onda (Van der Kruk 2003).

Na Figura 1.8 podemos observar uma representação do espectro *f-k* para uma típica família de tiro, contendo elementos de reflexão e de ruído. Percebe-se que o espectro referente às reflexões aparecem separados do restante do ruído, tornando possível assim a eliminação dos mesmos por meio de uma filtragem em envelope ou “fatia de pizza”. Para isso devemos determinar as inclinações (*slopes*) do filtro no espectro *f-k* (estimando as velocidades), excluir o sinal indesejado e retornar a porção de sinal desejado novamente para o domínio tempo-distância (Van der Kruk 2003).

Uma questão levantada neste trabalho e que será abordada mais adiante é a de uma situação onde o espectro *f-k* do *ground roll* se sobrepõe ao espectro das reflexões, fazendo com que a aplicação da filtragem por transformada *f-k* elimine informação de interesse juntamente com o

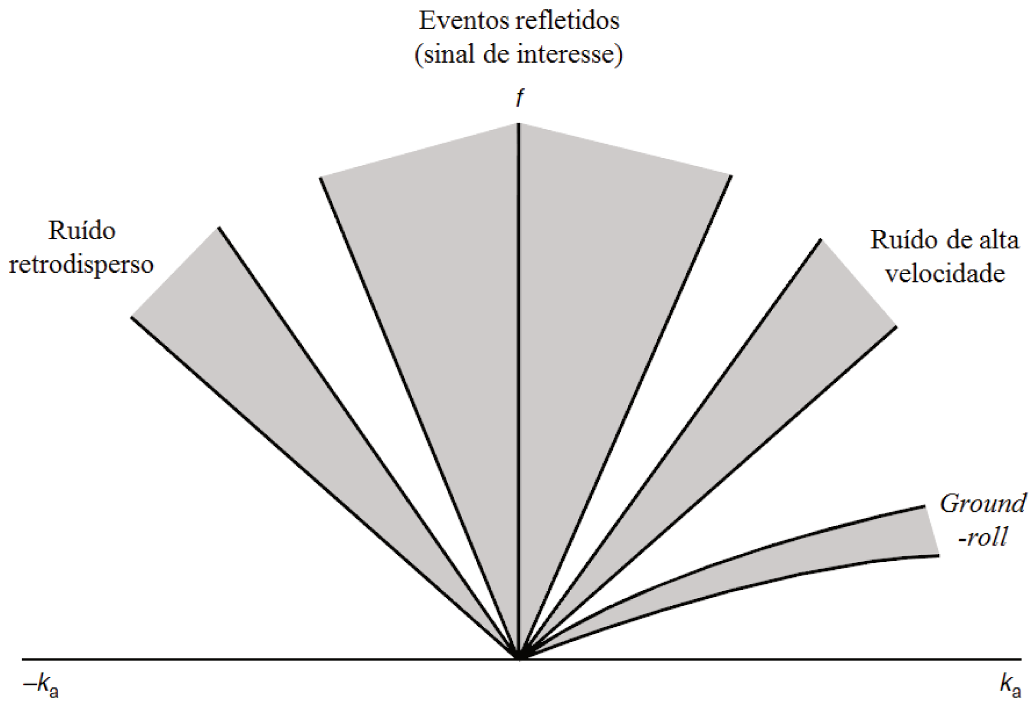

Figura 1.8: Representação do espectro *f*-*k* e os elementos de sinal e ruído que o compõem (readaptado de (Kearey et al. 2002)).

ruído indesejado. No capítulo seguinte serão abordadas algumas técnicas de decomposição de sinais que já são bem conhecidas da literatura e são aplicadas pela indústria em diversos lugares na remoção de ruído coerente.

Capítulo 2

Métodos de Decomposição de Sinais

Este capítulo apresenta alguns métodos de decomposição e separação de sinais que também foram levados em consideração durante os trabalhos. Primeiramente, é apresentada a Decomposição em Valores Singulares (*Singular Value Decomposition* - SVD), já utilizada no contexto da remoção do *ground roll* em (Porsani et al. 2010). Em seguida, é introduzido o conceito de Separação Cega de Fontes (*Blind Source Separation* - BSS), a amplamente utilizada Análise de Componentes Principais (*Principal Component Analysis* - PCA) e a Análise de Componentes Independentes (*Independent Component Analysis* - ICA). Por fim, apresentamos um método híbrido, a separação de sinais por SVD-ICA. Todas as técnicas apresentadas neste capítulo são acompanhadas de exemplos de aplicação em dados sintéticos.

2.1 Decomposição em Valores Singulares

A Decomposição em Valores Singulares é uma técnica largamente utilizada em aplicações que buscam o aprimoramento de um determinado sinal. Por exemplo, pode ser utilizada na melhoria da relação sinal-ruído e na restauração de imagens corrompidas. Em processamento de dados sísmicos, ela pode ser empregada na atenuação de múltiplas, correção de estática residual e, de modo geral, na separação de eventos sísmicos (Yilmaz 2001).

Esta decomposição se assemelha bastante à decomposição em autovalores e autovetores (*eigendecomposition*). Ambas são técnicas de processamento de matrizes que buscam fatorar uma matriz de dados em vetores ponderados por uma matriz diagonal (ou pseudodiagonal, no caso da SVD). Elas se diferenciam basicamente no fato de que a SVD pode ser aplicada em matrizes retangulares e não apenas quadradas, podendo assim ser considerada uma extensão da decomposição em autovalores e autovetores (Diamantaras & Kung 1996).

A SVD consiste na fatoração de uma matriz de dados $\mathbf{X} \in \mathbb{R}^{m \times n}$ como se segue:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}, \tag{2.1}$$

onde $\mathbf{U} \in \mathbb{R}^{m \times m}$ e $\mathbf{V} \in \mathbb{R}^{n \times n}$ são matrizes quadradas ortogonais e $\mathbf{D} \in \mathbb{R}^{m \times n}$ é uma matriz pseudo-diagonal retangular. Novamente, associando a SVD à decomposição em autovalores e autovetores, podemos dizer que as colunas de $\mathbf{U}$ são autovetores de $\mathbf{X}\mathbf{X}^{\mathrm{T}}$, as colunas de $\mathbf{V}$ são autovetores de $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ e os valores singulares contidos em $\mathbf{D}$ são raízes quadradas positivas dos autovalores das matrizes $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ e $\mathbf{X}^{\mathrm{T}}\mathbf{X}$.

Conforme ilustrado na Figura 2.1, a mesma fatoração na equação 2.1 também pode ser representada em forma de somatório:

$$\mathbf{UDV}^{\mathrm{T}} = \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}, \tag{2.2}$$

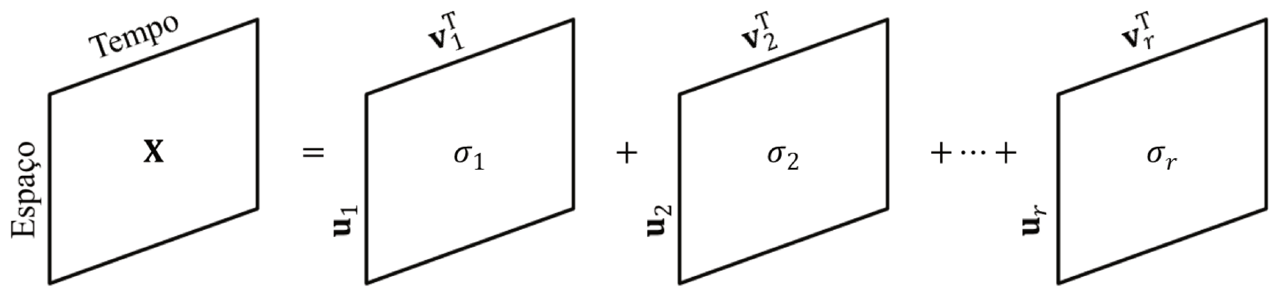

Figura 2.1: Fatoração da matriz $\mathbf{X}$ em uma soma de r autoimagens ponderadas. Adaptado de (Kirlin & Done 1999).

onde r é o valor mínimo entre m e n. Os vetores $\mathbf{u}_k \in \mathbb{R}^m$ e $\mathbf{v}_k^{\mathrm{T}} \in \mathbb{R}^n$ correspondem às k-ésimas colunas das matrizes $\mathbf{U}$ e $\mathbf{V}^{\mathrm{T}}$, respectivamente. Além disso, a matriz de posto unitário $\mathbf{u}_k\mathbf{v}_k^{\mathrm{T}}$ é chamada k-ésima autoimagem da matriz de dados $\mathbf{X}$ (Kirlin & Done 1999), e influencia a decomposição ponderada por σ_k, ou seja, o k-ésimo valor singular de $\mathbf{X}$). Sejam $\mathbf{u}_k = [u_{k1}, \ldots, u_{ki}, \ldots, u_{km}]^T$ e $\mathbf{v}_k = [v_{k1}, \ldots, v_{kj}, \ldots, v_{kn}]$ os k-ésimos vetores singulares esquerdo e direito respectivamente. Na k-ésima autoimagem de $\mathbf{X}$, a amostra de dado x_{ij}^k no tempo j no sensor i é espressa por $x_{ij}^k = u_{ki}v_{kj}$ (Vrabie et al. 2004). Portanto, $\mathbf{v}_k$ é chamado de *wavelet normalizada*, sendo que $v_{kj}(1 \leq j \leq n)$ fornece a dependência no tempo da componente associada à k-ésima autoimagem. Assim, $\mathbf{u}_k$ é chamado *vetor de propagação em distância*, já que $u_{ki}(1 \leq i \leq m)$ fornece a amplitude da *wavelet* normalizada no i-ésimo sensor (Vrabie et al. 2004, Glangeaud & Mari 1994).

Os valores singulares são ordenados por ordem decrescente de magnitude, concentrando assim a maior parte da energia do sinal nas primeiras autoimagens. Esta característica pode ser observada na Figura 2.2, onde percebe-se a grande diferença de amplitude entre o primeiro e o quinto valores singulares — energia correspondente à maioria dos eventos horizontais — e um grande número de valores singulares com energia bem menos expressiva. De fato, a matriz $\mathbf{X}$ pode também ser representada como se segue:

$$\mathbf{X} = \mathbf{S} + \mathbf{N}, \tag{2.3}$$

onde $\mathbf{S} \in \mathbb{R}^{m \times n}$ é uma matriz correspondente ao sinal desejado e $\mathbf{N} \in \mathbb{R}^{m \times n}$ é a matriz que contém o ruído a ser extraído pela SVD. Dessa forma, após a aplicação da decomposição, teremos dois conjuntos de dados: o *subespaço de sinal*, composto da soma das autoimagens referentes aos mais elevados valores singulares e, portanto, a maior parcela do sinal desejado; e o *subespaço*

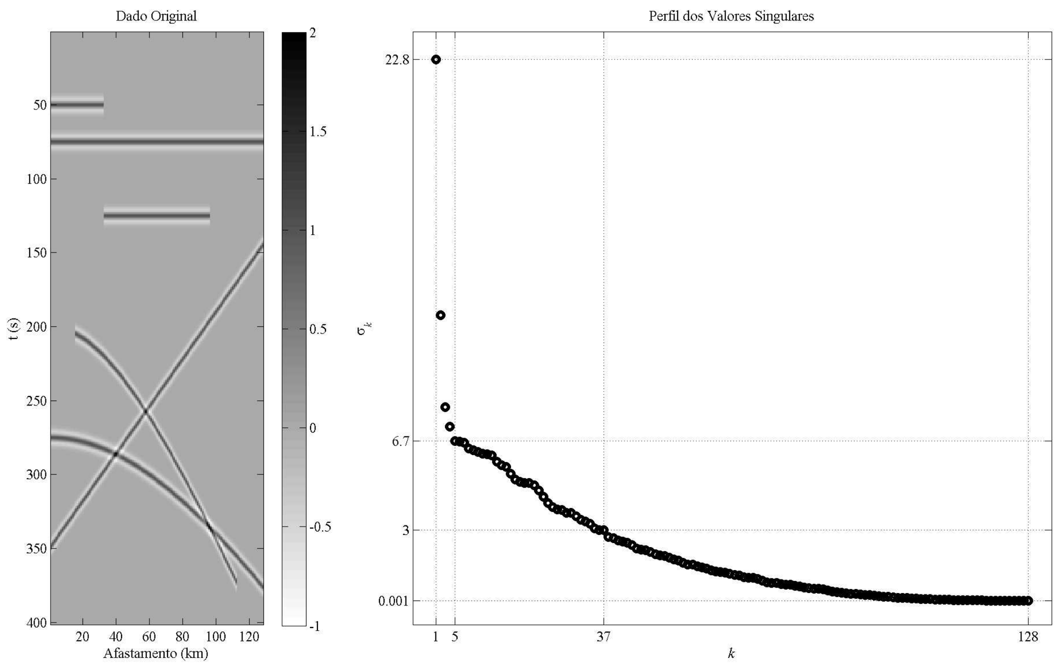

Figura 2.2: À esquerda, imagem originalmente submetida à SVD. À direita, perfil dos valores singulares calculados para a mesma imagem.

de ruído, composto da soma das autoimagens restantes, ou seja, o ruído de $\mathbf{X}$. Logo:

$$\mathbf{X} = \sum_{k=1}^{p} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}} + \sum_{k=p+1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}, \tag{2.4}$$

onde p é igual ao número de autoimagens pertencentes ao subespaço de sinal.

Em processamento de dados sísmicos, a SVD é designada a extrair os eventos caracterizados por um alto grau de correlação traço-a-traço (Kirlin & Done 1999). Tais eventos tendem a estar concentrados nas autoimagens associadas aos mais altos valores singulares, ou seja, dentro do subespaço de sinal, enquanto que o restante das autoimagens estão contidas dentro do subespaço de ruído. Portanto, se considerarmos apenas estas autoimagens quando reconstruirmos o dado original, será possível estimar, por exemplo, eventos horizontais (vide Figura 2.3). A SVD pode também ser usada para separar eventos lineares que não estão horizontalmente alinhados. Isto pode ser feito executando-se uma correção de tempo *a priori* na seção de dado sísmico, de forma a alinhar o evento desejado (Kirlin & Done 1999). Eventos não lineares também podem ser separados pela SVD, desde que uma correção de tempo adequada seja feita.

O número de autoimagens p escolhido para representar o subespaço de sinal depende da análise de perfil dos valores singulares. No caso da Figura 2.3, foram utilizadas duas autoimagens enquanto que o restante foi incluído no subespaço de ruído. Entretanto, é possível observar nas Figuras 2.2 (especificamente o perfil de valores singulares) e 2.3 que o subespaço de ruído não é composto apenas de ruído e sim de eventos com baixa correlação traço-a-traço. De fato, como sugerido em (Vrabie et al. 2004), a matriz de dados $\mathbf{X}$ pode ser projetada em três subespaços

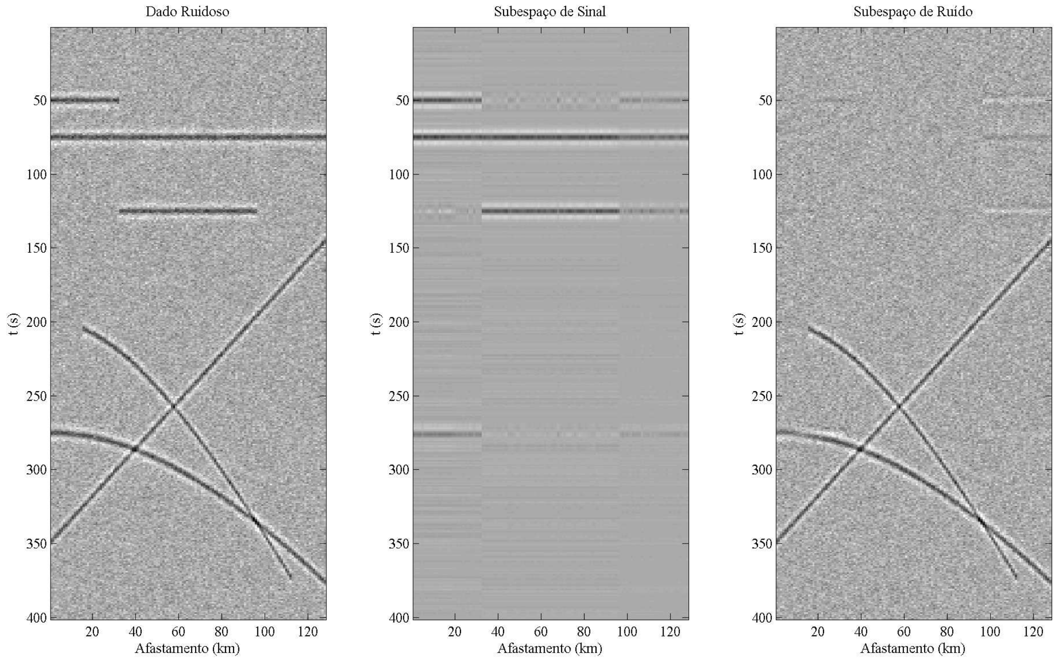

Figura 2.3: Aplicação da SVD em um dado ruidoso e sua separação entre subespaço de sinal e subespaço de ruído.

ortogonais:

$$\begin{aligned}\mathbf{X} &= \mathbf{X}_{LP} + \mathbf{X}_{BP} + \mathbf{X}_{HP} \\ &= \sum_{k=1}^{p} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}} + \sum_{k=p+1}^{q} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}} + \sum_{k=q+1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}\end{aligned} \tag{2.5}$$

- Subespaço *Passa-Baixas* (*Low-Pass* - LP), associado ao subespaço de sinal, contendo os eventos com maior correlação traço-a-traço;
- Subespaço *Passa-Faixa* (*Band-Pass* - BP), associado aos $q - p$ valores residuais do subespaço de sinal, composto por elementos de baixa correlação traço-a-traço;
- Subespaço *Passa-Altas* (*High-Pass* - HP), contendo apenas o ruído;

Em geofísica, temos que os eventos caracterizados por refletores horizontais estão contidos em $\mathbf{X}_{LP} \in \mathbb{R}^{m \times n}$, refletores inclinados ou que apresentam certa curvatura compõem $\mathbf{X}_{BP} \in \mathbb{R}^{m \times n}$ enquanto os ruídos provenientes do ambiente ou equipamentos de medição ou interferências diversas estão contidos em $\mathbf{X}_{HP} \in \mathbb{R}^{m \times n}$.

Voltando à representação gráfica do subespaço de sinal na Figura 2.3, nota-se o aparecimento de segmentos de sinal que não existem no dado original. Este cenário é mais acentuado em conjuntos de dados como o utilizado neste exemplo, onde o perfil dos valores singulares possui grande quantidade de energia concentrada na porção $\mathbf{X}_{BP}$ da decomposição. Estes segmentos são chamados de *artefatos* e, neste caso, estão ligados à condição de ortogonalidade (que não é justificada fisicamente) imposta pela SVD aos vetores de propagação, forçando o vetor de *wavelets* normalizadas a ser uma mistura de fontes (Vrabie et al. 2004).

2.2 Separação Cega de Fontes

O conjunto de métodos conhecidos como Separação Cega de Fontes é usado para separar dados em componentes de informação subjacentes (Stone 2004). Esses dados são compostos por misturas de sinais provenientes de diferentes fontes geradoras, das quais se conhece pouca ou nenhuma informação *a priori* — daí o significado da expressão *separação cega*. Um exemplo clássico deste problema é o de dois grupos de pessoas conversando separadamente em uma sala, onde é possível ouvir parcialmente a conversa do grupo externo ao mesmo tempo em que se ouve o que é falado dentro do próprio grupo. Os métodos de BSS buscam separar a mistura de vozes ouvida em dois discursos que correspondem ao que foi realmente falado em cada grupo com o mínimo possível de perdas.

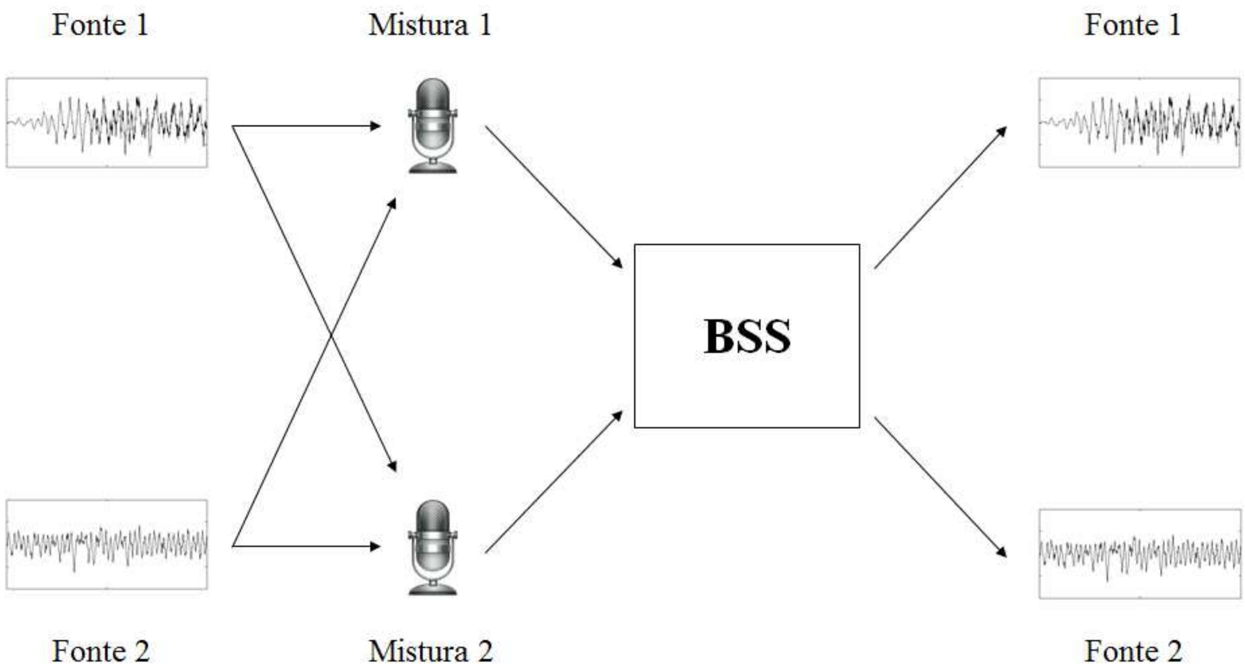

Figura 2.4: Exemplo de um problema de separação de fontes.

Este exemplo pode ser ilustrado pela Figura 2.4, com a diferença de que as duas misturas de sinais são captadas pelos microfones.

As técnicas clássicas consideram as seguintes propriedades no processo de separação dessas misturas (Stone 2004):

Independência: Se os sinais emitidos pelas fontes são independentes, as misturas dos mesmos não o serão. Isto ocorre pois cada mistura carrega as características dos sinais que as compõe e isso garante a dependência entre ambas. Sinais independentes e misturas dependentes são requisitos básicos para diversos métodos de separação de fontes.

Normalidade: Faz referência ao formato dos histogramas em relação à distribuição *normal*, também chamada de *gaussiana*. As misturas serão mais facilmente separáveis quando suas fontes apresentarem menor gaussianidade, embora o histograma das misturas apresentem um formato mais gaussiano do que qualquer uma de suas fontes componentes.

Complexidade: A complexidade temporal (estrutura ao longo do tempo) de uma mistura é sempre maior ou igual à complexidade da menos complexa de suas fontes. Dessa forma, é possível recuperar uma mistura extraindo-se o sinal da fonte de estrutura mais simples. A complexidade neste caso é explorada a partir de matrizes descorrelacionadas.

Busca-se, portanto, informações comuns nas misturas através de análises de histogramas, formas de onda e outras características de forma a suprir os métodos de BSS com a maior quantidade possível de informações sobre as misturas e fontes para aumentar a possibilidade de separação dos sinais.

É necessário também observar o número de misturas e fontes. Os métodos de BSS funcionam melhor quando o número de fontes é igual ou menor do que o número de misturas — o que ocorre mais frequentemente na prática. A separação também é possível quando há mais fontes do que misturas, porém esses casos são, de certo modo, mais desafiadores dado que há um aumento do número de variáveis desconhecidas. Justamente, por particularidades de cada problema, a seleção do método de BSS mais apropriado é outro aspecto muito importante, pois pode determinar a falha ou o sucesso na separação das misturas (Comon & Jutten 2010).

Basicamente, os métodos de BSS tentam extrair as fontes das misturas por meio do produto interno entre as colunas de uma matriz de ponderação e as misturas destes sinais, o que produz uma projeção ortogonal das mesmas (Stone 2004). Por exemplo, no caso de métodos baseados na recuperação da não gaussianidade, esta matriz de ponderação, também chamada *matriz de separação*, é obtida de forma que a *função densidade de probabilidade* (*Probability Density Function* - PDF) do sinal extraído seja o mais diferente possível de uma gaussiana. Geralmente, os métodos de separação são precedidos pela aplicação da Análise de Componentes Principais, discutida na sequência.

2.3 Análise de Componentes Principais

Uma técnica muito utilizada na redução da quantidade de ruído em um sinal é a Análise de Componentes Principais (Stone 2004). A proposta da PCA é identificar a estrutura de dependência por trás de uma observação estocástica multivariada de forma a obter uma descrição compacta da mesma (Diamantaras & Kung 1996). No caso de um dado de dimensionalidade m com correlação diferente de zero entre as variáveis observadas, é possível obter uma descrição deste usando apenas as n variáveis independentes, onde $m > n$ em qualquer situação. Quanto mais correlacionadas forem as variáveis observadas, menor será o número de variáveis independentes que podem descrever os dados.

A PCA, que se baseia numa formulação estatística, utiliza funções matemáticas para descrever a estrutura de dependência entre as variáveis. Transformações lineares são as mais funcionais formas para representação de funções matemáticas e a PCA tradicional está relacionada a estas transformações (Diamantaras & Kung 1996). Basicamente, o objetivo é impor uma transformação linear ortogonal que converte os dados em um novo sistema de coordenadas, tal que a projeção de maior variância determina o eixo chamado de primeiro componente principal, a projeção de segunda maior variância determina o segundo componente principal (ortogonal ao primeiro) e assim por diante.

Considerando um vetor aleatório $\mathbf{x} = [x_1, \ldots, x_m]^\mathrm{T}$ com média $E\{\mathbf{x}\} = 0$ e matriz de covariância $\mathbf{R_x} = E\{\mathbf{x}\mathbf{x}^\mathrm{T}\} \in \mathbb{R}^{m \times m}$. O vetor de realizações $\mathbf{y}$, determinado pela PCA, é dado pela seguinte transformação:

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \tag{2.6}$$

onde as colunas da matriz $\mathbf{W}$ formam uma base ortonormal de seu próprio espaço vetorial e $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{I}$. Além disso, suas linhas são vetores de coeficientes e cada um deles tem comprimento unitário ($||\mathbf{w}_k|| = 1$) e extrai exatamente um sinal y_k, ou componente principal, da mistura $\mathbf{x}$. A variância de cada componente principal é conhecida como autovalor (ou λ_k). Por definição, um vetor $\mathbf{w}_k$ é um autovetor de $\mathbf{R}_{\mathbf{x}}$ se:

$$\lambda_k \mathbf{w}_k^{\mathrm{T}} = \mathbf{R}_{\mathbf{x}} \mathbf{w}_k, \tag{2.7}$$

ou seja, se $\mathbf{w}_k$ é um autovetor da matriz $\mathbf{R}_{\mathbf{x}}$, então a magnitude de $||\mathbf{w}_k||$ é alterada por um fator λ_k (Stone 2004).

A PCA busca a maximização da *variância da projeção* (J_v) de $\mathbf{x}$ no espaço vetorial de $\mathbf{R}_{\mathbf{x}}$, ou seja, a reconstrução de $\mathbf{x}$ a partir de $\mathbf{y}$, cujas realizações $y_1, y_2, \ldots, y_m$ são chamadas de *componentes principais* de $\mathbf{x}$. Isto é o equivalente a buscar a minimização do *erro quadrático médio da reconstrução* de $\mathbf{x}$ (J_e) (Diamantaras & Kung 1996). Seja o vetor de autovalores de $\mathbf{R}_{\mathbf{x}}$ ordenado de forma decrescente ($\lambda_1 > \lambda_2 > \ldots > \lambda_m$), temos que o erro mínimo de reconstrução será dado por:

$$\min J_e = \sum_{k=n+1}^{m} \lambda_k, \tag{2.8}$$

enquanto que a máxima variância será:

$$\min J_v = \sum_{k=1}^{n} \lambda_k \tag{2.9}$$

Cada componente principal é uma combinação linear das variáveis observadas:

$$y_k = w_{k1}x_k + \cdots + w_{kn}x_n = w_k^{\mathrm{T}} x, \tag{2.10}$$

tal como suas respectivas variâncias:

$$E\{y_k^2\} = w_k^{\mathrm{T}} E\{xx^{\mathrm{T}}\} w_k = w_k^{\mathrm{T}} \mathbf{R}_{\mathbf{x}} w_k \tag{2.11}$$

O primeiro componente principal tem a máxima variância sob a condição de que o vetor de coeficientes é normalizado $\|w_1\| = 1$. O segundo componente principal é ortogonal ao primeiro e da mesma forma os componentes principais em sequência (como ilustrado na Figura 2.5), sendo que o resultado final levará em conta apenas os n primeiros componentes principais, necessários à melhor representação em termos de informação e compactação.

A Figura 2.6 mostra, de forma análoga ao que acontece no caso da SVD (Figura 2.2), a grande contribuição dos primeiros componentes principais na formação da imagem, observada no perfil dos autovalores calculados usando-se a PCA. A simulação na qual aplicamos a SVD ao dado de exemplo, referente à Figura 2.3, foi também executada aplicando-se a PCA e seu resultado pode ser visto na Figura 2.7 a seguir. Para efeito de comparação, no caso da SVD foram utilizadas as duas primeiras autoimagens, enquanto que no caso da PCA, levamos em conta os dois primeiros componentes principais na reconstrução do dado para o subespaço de sinal.

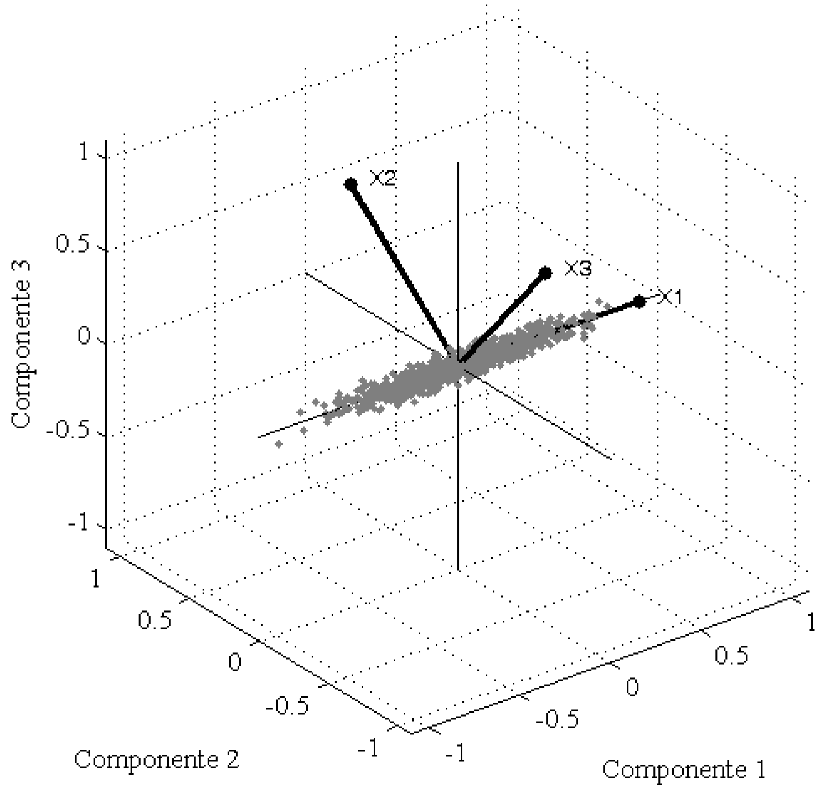

Figura 2.5: Representação da PCA para um conjunto de pontos e seus eixos principais.

Existem várias semelhanças entre as decomposições efetuadas pela PCA e pela SVD: Ambas produzem uma redução de dimensionalidade, são baseadas em decomposições vetoriais e utilizam de estatísticas de segunda ordem em seus algoritmos. Frequentemente são estudadas formas de combinar diferentes técnicas de decomposição de forma a usar os pontos fortes de cada uma, ao mesmo tempo em que se procura minimizar os efeitos de seus pontos fracos. Mais adiante, veremos uma dessas formas híbridas de decomposição de sinais, a SVD-ICA (Vrabie et al. 2004). Mas antes apresentaremos a Análise de Componentes Independentes.

2.4 Análise de Componentes Independentes

A Análise de Componentes Independentes faz parte do conjunto de métodos de BSS. É baseada no princípio fisicamente realista de que sinais diferentes, provenientes de processos físicos diferentes, são estatisticamente independentes. Portanto, de forma reversa, considera que sinais independentes extraídos de misturas, pertencem a processos distintos (Stone 2004).

Apesar de estar relacionada aos métodos de decomposição vistos anteriormente, a ICA utiliza uma propriedade estatística mais forte do que a descorrelação (ortogonalidade) na extração de fontes: A *independência estatística*. Em comparação, uma dada mistura, composta por vozes captadas por microfones, quando processada via PCA seria extraída em fontes ortogonais entre si, porém elas ainda seriam misturas de vozes. Enquanto que as misturas, quando processadas via ICA, seriam decompostas em fontes independentes estatisticamente, correspondendo a cada

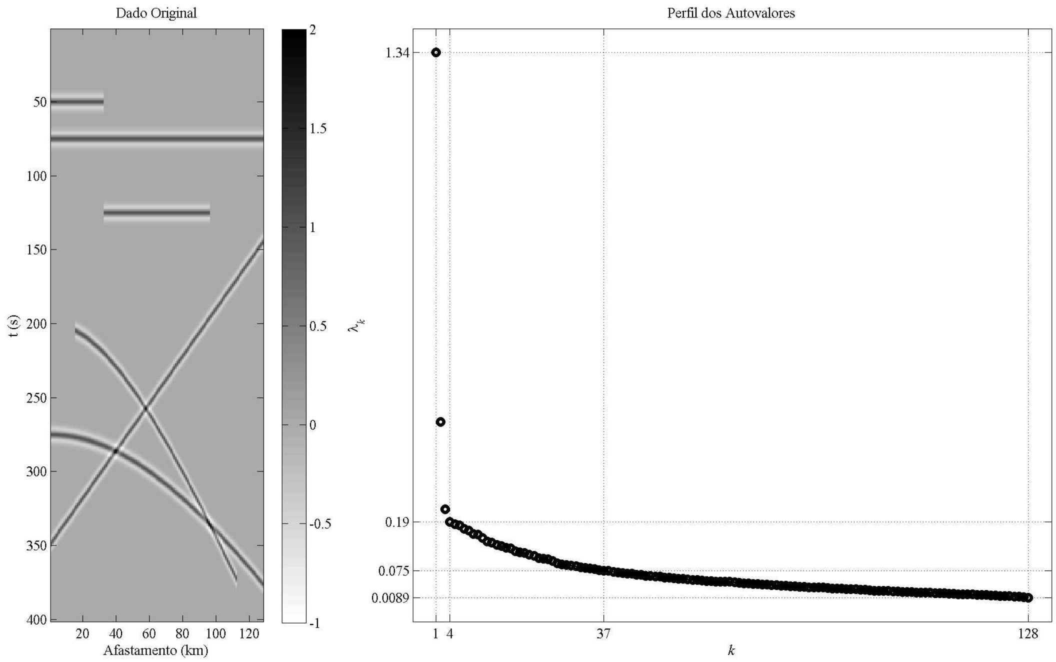

Figura 2.6: Dado original utilizado e o respectivo perfil de autovalores obtidos aplicando a PCA.

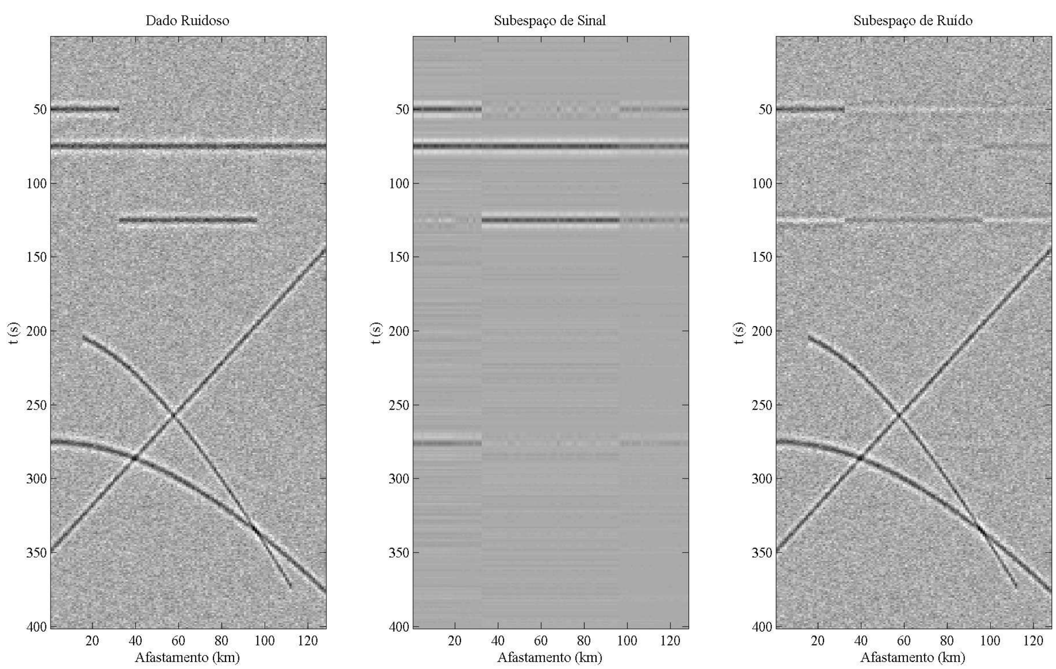

Figura 2.7: Separação do dado ruidoso em subespaços de sinal e de ruído, usando-se PCA.

uma das vozes. Este resultado, que é válido sob a condição de independência das fontes, foi demonstrado em (Comon 1994).

Com o intuito de introduzir a ICA, vamos considerar um modelo de mistura linear e instantâneo. Assim sendo, considere uma matriz $\mathbf{X} \in \mathbb{R}^{m\times n}$ composta por n amostras de m sinais, representados por:

$$\mathbf{X} = \mathbf{AS}, \tag{2.12}$$

onde $\mathbf{A} \in \mathbb{R}^{m\times m}$ é chamada *matriz de mistura*, e $\mathbf{S} \in \mathbb{R}^{m\times n}$ é uma matriz que contém os sinais fontes, também desconhecidos. É fundamental que os coeficientes de mistura em $\mathbf{A}$ sejam suficientemente diferentes, de forma a tornar inversível a matriz por eles composta (Hyvärinen, Karhunen & Oja 2001).

Com o intuito de estimar $\mathbf{S}$, é necessário encontrar uma matriz de separação $\mathbf{W} \in \mathbb{R}^{m\times m}$ que forneça a seguinte estimativa dos dados:

$$\mathbf{Y} = \mathbf{WX}. \tag{2.13}$$

Idealmente, $\mathbf{W}$ deve ser a matriz inversa de $\mathbf{A}$. Porém, dado que $\mathbf{A}$ é desconhecida, torna-se necessário utilizar alguma propriedade das fontes, conforme descrito anteriormente.

A técnica de ICA considera que os sinais fontes são estatisticamente independentes e, assim sendo, busca ajustar a matriz $\mathbf{W}$ de modo que os sinais representados por $\mathbf{Y} \in \mathbb{R}^{m\times n}$ satisfaçam esta condição. Assim, temos que $\mathbf{Y}$ será igual a $\mathbf{S}$ (talvez, multiplicada por uma constante, porém com peso insignificante) (Hyvärinen et al. 2001). De fato, a característica de independência entre os sinais e não gaussianidade são determinantes na total separação das misturas usando ICA.

Figura 2.8: Recuperação de dois sinais uniformes *i.i.d.* após sua mistura.

Diferentemente das fontes não gaussianas, conforme visto na Figura 2.8, as fontes gaussianas tem sua estimação inviabilizada a partir de suas misturas. Como a PDF conjunta resultante é simétrica, ela não contém nenhuma informação sobre a matriz de mistura, impossibilitando sua estimação. A distribuição das fontes gaussianas não é afetada por qualquer transformação ortogonal.

Os métodos para BSS que usam a ICA como principal ferramenta necessitam de estatísticas de ordem superior para solucionar os problemas de fontes independentes. Como no caso de fontes gaussianas estas estatísticas são zero, a separação de fontes, neste caso, é impossível.

Em comparação aos métodos de separação anteriormente listados, que utilizam estatísticas de segunda ordem em seus algoritmos, a ICA utiliza estatísticas de ordem superior como a curtose, que é muito importante na determinação da não-gaussianidade dos sinais. Porém, devido ao fato da curtose ser bastante sensível a *outliers* (valores atípicos), pode-se usar outras características do sinal, como a negentropia para a determinação da não-gaussianidade nos algoritmos de ICA que utilizam esta propriedade (Stone 2004).

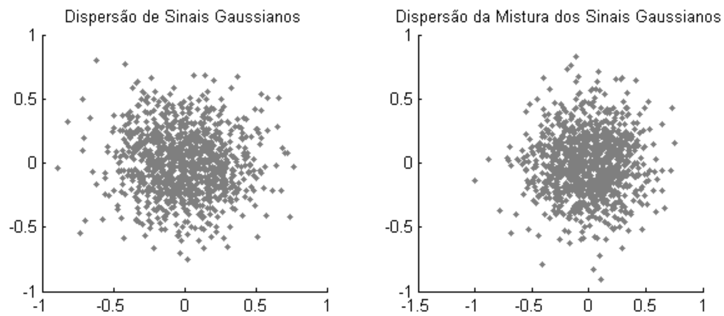

Figura 2.9: Resultado da mistura de dois sinais gaussianos independentes.

A ICA também possui variantes e extensões, como a que adiciona um termo de ruído ao modelo:

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}, \tag{2.14}$$

onde assume-se que $\mathbf{N} \in \mathbb{R}^{m \times n}$ representa um ruído branco, gaussiano e estatisticamente independente dos outros sinais de interesse que compõem a mistura.

Dentre os diversos algoritmos para resolução da ICA, destacamos:

FastICA: É um algoritmo de ponto-fixo que possui três variantes: usando curtose com deflação; usando ortogonalização simétrica; e usando a não linearidade da tangente hiperbólica com ortogonalização simétrica (Hyvarinen 1999).

Algoritmos de gradiente: Usados para estimação de máxima verossimilhança usando uma não linearidade fixa dada pela tangente hiperbólica: *steepest descent* ou algoritmo Bell-Sejnowski; o algoritmo do gradiente natural; gradiente natural usando não linearidade adaptativa (neste caso, o sinal da curtose), ou algoritmo Bell-Sejnowski extendido.

NPCA-RLS: (*Nonlinear PCA - Recursive Least-Squares*) É o algoritmo de Mínimos Quadrados Recursivo para o critério de PCA Não Linear.

EASI: (*Equivariant Adaptive Separation via Independence*) Separação Adaptativa Equivariante via Independência. É um algoritmo de aprendizagem que utiliza a técnica do gradiante relativo.

JADE: Método de Diagonalização Aproximada Conjunta de Matrizes (*Joint Approximate Diagonalization of Eigenmatrices*) (Cardoso & Souloumiac 1993).

Este último é o mesmo utilizado na técnica híbrida SVD-ICA que será apresentada na sequência.

2.5 Separação de Eventos por SVD-ICA

A SVD cumpre um papel importante na separação de sinais, diminuição de dimensionalidade e aumento da relação sinal-ruído (*Signal-to-Noise Ratio* - SNR). Porém, esta decomposição possui limitações, principalmente relacionadas à imposição de ortogonalidade (que não possui justificativa física) aos vetores de propagação que compõem a matriz $\mathbf{U} \in \mathbb{R}^{m \times m}$ em 2.1. Uma maneira de relaxar esta restrição é encontrar uma nova matriz de *wavelets* normalizadas $\tilde{\mathbf{V}}_R \in \mathbb{R}^{n\times R}$ para as quais estas ondas serão tão estatisticamente independentes quanto possível. Isto pode ser obtido através do uso da ICA (Vrabie et al. 2004).

Esta técnica híbrida consiste em, primeiramente, reduzir a dimensionalidade dos dados em análise com a aplicação da SVD. É definido um número de autoimagens arbitrariamente, após a análise do perfil de valores singulares, de forma a excluir grande parte do ruído ($\mathbf{X}_{HP} \in \mathbb{R}^{m\times n}$ em 2.5) e reduzir o custo computacional da ICA a ser aplicada. Seja R o número de autoimagens desejado, temos:

$$\mathbf{X}_R = \mathbf{U}_R\mathbf{D}_R\mathbf{V}_R^{\mathrm{T}} = \sum_{k=1}^{R} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}, \tag{2.15}$$

onde $\mathbf{X}_R \in \mathbb{R}^{m\times n}$, $\mathbf{U}_R \in \mathbb{R}^{m\times R}$, $\mathbf{D}_R \in \mathbb{R}^{R\times R}$ e $\mathbf{V}_R \in \mathbb{R}^{n\times R}$. Em seguida, é necessário encontrar uma matriz ortogonal de rotação $\mathbf{B} \in \mathbb{R}^{R\times R}$ para garantir que as *wavelets* normalizadas de $\tilde{\mathbf{V}} \in \mathbb{R}^{n\times R}$ sejam estatisticamente independentes (Vrabie et al. 2004). Para isso é usado o algoritmo JADE, já citado anteriormente, e assim obtemos:

$$\tilde{\mathbf{V}}_R = \mathbf{V}_R\mathbf{B}. \tag{2.16}$$

Assim, o subespaço $\mathbf{X}_R$ fica descrito como:

$$\mathbf{X}_R = \sum_{k=1}^{R} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}} = \mathbf{U}_R\mathbf{D}_R\mathbf{V}_R^{\mathrm{T}} = \mathbf{C}_R\tilde{\mathbf{V}}_R^{\mathrm{T}}, \tag{2.17}$$

onde $\mathbf{C}_R \in \mathbb{R}^{m\times R}$, que pode ser representada como $\mathbf{C}_R = \mathbf{U}_R\mathbf{D}_R\mathbf{B}$. A partir deste ponto, obtemos outras duas matrizes: $\tilde{\mathbf{U}}_R \in \mathbb{R}^{m\times R}$, composta dos novos vetores de propagação normalizados e $\tilde{\mathbf{D}}_R \in \mathbb{R}^{R\times R}$, que é diagonal e composta pelos valores singulares modificados (Vrabie et al. 2004).

Finalmente, é escolhido um novo número de autoimagens que representarão os novos subespaços de sinal e de ruído:

$$\mathbf{X}_{sinal} = \sum_{k=1}^{Q} \tilde{\sigma}_k \tilde{\mathbf{u}}_k \tilde{\mathbf{v}}_k^{\mathrm{T}} \quad \text{e} \quad \mathbf{X}_{rudo} = \sum_{k=Q+1}^{R} \tilde{\sigma}_k \tilde{\mathbf{u}}_k \tilde{\mathbf{v}}_k^{\mathrm{T}} \tag{2.18}$$

O resultado pode ser observado na Figura 2.10 a seguir:

Tendo apresentado neste capítulo os métodos mais consagrados de separação, discutiremos a seguir uma nova abordagem, mais diretamente ligada à motivação inicial deste trabalho e aos resultados aqui obtidos: a Análise de Componentes Principais Robusta.

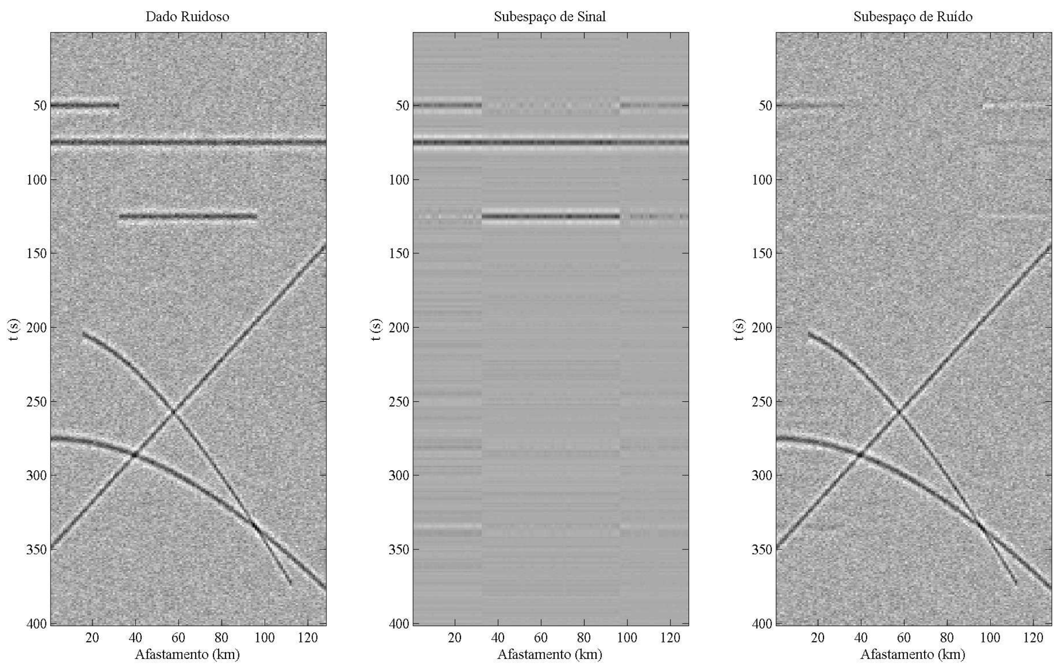

Figura 2.10: Decomposição do dado ruidoso em subespaços de sinal e de ruído usando-se SVD-ICA.

Capítulo 3

Análise de Componentes Principais Robusta

Neste capítulo é introduzida a técnica de Análise de Componentes Principais Robusta (*Robust Principal Component Analysis* - RPCA), que consiste em decompor um sinal em uma componente de posto reduzido e uma componente esparsa. Em seguida, a técnica da RPCA é descrita como método de separação. Enfim, encerramos o capítulo com um sumário a respeito do algoritmo utilizado nas simulações e um exemplo do seu funcionamento em um dado sísmico sintético.

3.1 Descrição do Problema/Modelo

Nas seções anteriores foram apresentados diversos métodos para separação de sinais. Cada um destes métodos apresenta um desempenho superior aos outros em determinadas situações e o oposto também é válido. Entretanto, há uma característica comum a todos: separar, com certa precisão, os elementos de posto baixo do restante dos dados (que são tratados como um subespaço de elementos indesejados ou simplesmente ruído). Esta também é uma característica da Análise de Componentes Principais Robusta, porém é usado o conceito de *esparsidade* no tratamento do restante dos dados separados (que não podem ser simplesmente rotulados como ruído, por exemplo).

No que se refere ao desempenho destes métodos mencionados até aqui, pode-se dizer que a natureza do sinal interferente ou ruído tem influência determinante sobre o resultado final. Até então, vimos exemplos onde os dados foram submetidos a um ruído branco gaussiano e a relação sinal-ruído era razoavelmente baixa, com 14 dB. Porém, há diversas aplicações reais onde o ruído corrompe a informação de forma muito grosseira e sua amplitude pode até mesmo superar a do sinal desejado. Este ruído também pode apresentar um certo nível de coerência, o que pode levar alguns algoritmos de separação a uma convergência errônea, incluindo vestígios do ruído entre as informações de interesse.

Em sísmica de reflexão, o *ground roll* é um dos casos de ruído que se adapta a essa descrição. Devido à diferença de velocidade das ondas em propagação, o *ground roll* tende a aparecer separado dos eventos de interesse, quando os dados são observados no domínio *f-k*. Porém, há casos onde as ondas se sobrepõem às de interesse e, neste caso, as filtragens convencionalmente utilizadas podem retirar, junto com o ruído, uma parcela de sinais desejados.

Existem várias outras aplicações importantes para a RPCA, onde os dados sob estudo podem ser modelados como um componente de posto baixo somado a outro componente esparso. Por exemplo (Candès, Li, Ma & Wright 2011):

Vigilância por Vídeo: O problema consiste na separação das ações que se pretende monitorar de elementos estáticos ao fundo (tais como mobílias e estruturas) em uma dada sequência de *frames* de vídeo de vigilância. Se empilharmos os *frames* como colunas de uma matriz $\mathbf{M}$, então o componente de posto baixo $\mathbf{L_0}$ corresponderá aos elementos estáticos das imagens enquanto que o componente esparso $\mathbf{S_0}$ será composto pelos objetos e pessoas em movimento.

Reconhecimento Facial: Esta aplicação é baseada no princípio em que imagens de uma superfície convexa Lambertiana, sob iluminação variante, produzem um subespaço de dimensões reduzidas (Basri & Jacobs 2003). Particularmente, imagens de faces humanas podem ser bem aproximadas a um subespaço de dimensões reduzidas e ser capaz de recuperar corretamente este subespaço é de suma importância no reconhecimento facial e alinhamento. As imagens normalmente contém outros elementos como sombras ou saturações em brilho. Às vezes, por exemplo, é necessário submeter a análise uma fotografia de revista, com caracteres e outros elementos se sobrepondo à face, ou então a pessoa está usando óculos escuros ou chapéu e todos esses fatores podem comprometer as técnicas de reconhecimento facial.

Indexação Semântica Latente: Mecanismos de busca *web*, frequentemente, precisam analisar e indexar o conteúdo de um imenso conjunto de documentos. Um método popular para para executar esta tarefa é o LSI (*Latent Semantic Indexing*) (Deerwester, Dumais, Furnas, Landauer & Harshman 1990, Papadimitriou, Tamaki, Raghavan & Vempala 1998). Basicamente, ele consiste em coletar uma matriz $\mathbf{M}$ de “documento-versus-termo”, cujos registros codificam a relevância de um termo (ou uma palavra) para um documento, tal como a frequência com que ele aparece no documento. As técnicas já descritas PCA ou SVD são tradicionalmente utilizadas para decompor a matriz em uma parte de posto baixo e outra residual, que não é necessariamente esparsa. Porém, se fosse possível garantir que esta parte residual fosse esparsa, então $\mathbf{L_0}$ seria composta por palavras comuns usada em todo o texto enquanto que $\mathbf{S_0}$ conteria as poucas palavras-chave que melhor distinguiriam cada documento.

Classificação e Filtragem Colaborativa: Antecipar a preferência ou o gosto de usuários é o problema que vem ganhando importância no comércio e nos anúncios *online*. Empresas coletam as informações de classificação de usuários para os mais diversos produtos como filmes, livros, jogos, software, etc. O Netflix Prize para classificação de filmes *online* é um dos mais famosos exemplos (Netflix Inc. 2009). O problema consiste em usar classificações incompletas de usuários para preencher lacunas e poder predizer com precisão as preferências de um dado usuário com relação a filmes ou produtos. É um problema tipicamente classificado como de preenchimento de matrizes. Porém, muitas vezes algumas entradas estão corrompidas por ruídos ou alguma classificação totalmente incoerente feita por um usuário. Assim, o problema se torna mais desafiador, precisando completar e ao mesmo tempo corrigir as informações. Isto significa, que é necessário inferir uma matriz de posto baixo $\mathbf{L_0}$ a partir de um conjunto de dados incompletos e corrompidos.

A seguir, será apresentada a RPCA e suas características de execução.

3.2 O Método

A PCA é, indiscutivelmente, a ferramenta estatística de análise de dados e redução de dimensionalidade mais utilizada atualmente (Candès et al. 2011). Entretanto, a busca pela projeção de maior variância (principal característica do algoritmo da PCA) também torna esta técnica vulnerável a valores muito discrepantes (*outliers*). Isto é um problema quando a PCA é aplicada a sinais ou dados corrompidos grosseiramente por ruídos de amplitude mais elevada, pois a busca pelos componentes principais analisando as projeções de maior variância não converge adequadamente. A PCA modela uma matriz $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ de dados da seguinte forma:

$$\mathbf{M} = \mathbf{L}_0 + \mathbf{N}_0, \tag{3.1}$$

onde $\mathbf{L}_0 \in \mathbb{R}^{n_1 \times n_2}$ é uma matriz de posto baixo e $\mathbf{N}_0 \in \mathbb{R}^{n_1 \times n_2}$ é uma matriz de perturbação baixa.

A PCA clássica (Hotelling 1933, Eckart & Young 1936, Jolliffe 1986) busca a melhor estimativa (em termos de ℓ_2) de posto k para $\mathbf{L}_0$, resolvendo o seguinte problema de otimização (Candès et al. 2011):

$$\begin{aligned} \underset{\mathbf{L}}{\text{minimizar}} \quad & ||\mathbf{M} - \mathbf{L}||_2 \\ \text{sujeito a} \quad & \text{posto}(\mathbf{L}) \leq k. \end{aligned} \tag{3.2}$$

Por outro lado, a RPCA propõe uma forma de recuperar uma matriz de posto baixo $\mathbf{L}_0$ de um conjunto de medidas grosseiramente corrompidas:

$$\mathbf{M} = \mathbf{L}_0 + \mathbf{S}_0. \tag{3.3}$$

Diferentemente do caso da PCA clássica onde há termo de ruído reduzido $\mathbf{N}_0$, os componentes da matriz $\mathbf{S}_0 \in \mathbb{R}^{n_1 \times n_2}$ podem ter amplitude arbitrariamente elevada e supostamente esparsa, porém a localização das informações é desconhecida (Candès et al. 2011). Tal característica é ilustrada na Figura 3.1, onde uma dada matriz $\mathbf{M}$ é decomposta em um termo de posto reduzido e um termo esparso.

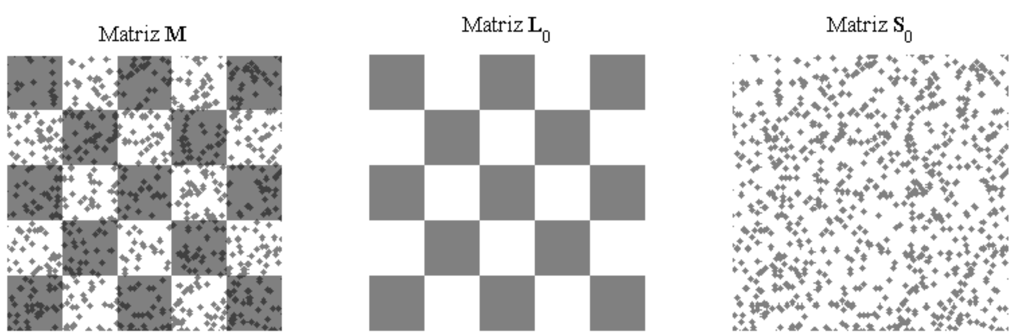

Figura 3.1: Matriz $\mathbf{M}$ composta da soma das componentes $\mathbf{L_0}$ e $\mathbf{S_0}$

O problema parece ser extremamente desafiante, dadas as condições de ruído e o pouco que se sabe a respeito dos dados em estudo. Entretanto, trata-se de um caso que pode ser resolvido via otimização convexa tratável. Primeiramente, façamos algumas definições: seja $||\mathbf{M}||_* = \sum_i \sigma_i(\mathbf{M})$ a representação da norma nuclear da matriz $\mathbf{M}$, que é a soma dos valores singulares desta matriz; e seja $||\mathbf{M}||_1 = \sum_i j|\mathbf{M}_{ij}|$ a representação da norma ℓ_1 de $\mathbf{M}$ vista como um grande vetor em $\mathbb{R}^{n_1 \times n_2}$. Então, mostra-se que sob hipóteses bem fracas, a estimação por busca de componentes principais (*Principal Component Pursuit* - PCP) recupera exatamente os termos de posto baixo $\mathbf{L}_0$ e esparso $\mathbf{S}_0$ resolvendo o seguinte problema de otimização (Candès et al. 2011):

$$\begin{array}{ll} \underset{\mathbf{L},\mathbf{S}}{\text{minimizar}} & ||\mathbf{L}||_* + \lambda ||\mathbf{S}||_1 \\ \text{sujeito a} & \mathbf{L} + \mathbf{S} = \mathbf{M}. \end{array} \tag{3.4}$$

Em (Candès et al. 2011) é garantido que este método funcione mesmo se o posto de $\mathbf{L}_0$ crescer quase linearmente com as dimensões da matriz, e os erros em $\mathbf{S}_0$ sejam até uma fração constante de todos os registros. Este problema pode ser resolvido por algoritmos eficientes e escaláveis a um custo computacional não muito maior do que na PCA clássica.

Existe entretanto um problema quanto à classificação dos elementos que compõem a matriz de dados. Em determinadas situações o termo de posto baixo pode também ser classificado como esparso e vice-versa, por exemplo, em um caso onde a matriz $\mathbf{M} = \mathbf{e}_1 \mathbf{e}_1^*$, ou seja, possui valor um na posição superior esquerda e zeros em todas as outras posições. Se $\mathbf{M}$ é tanto esparsa quanto de posto baixo, a decisão sobre a sua classificação pode ser impossível. Para que o problema apresente um significado, é necessário impor que a componente de posto baixo não seja esparso.

É nesse ponto em que o problema da RPCA se relaciona com o problema de preenchimento de matrizes e faz uso da *noção geral de incoerência*, introduzida em (Candès & Recht 2009). Trata-se de uma hipótese que leva em consideração a SVD da componente esparsa $\mathbf{L}_0 \in \mathbb{R}^{n_1 \times n_2}$ como:

$$\mathbf{L}_0 = \mathbf{U}\mathbf{D}\mathbf{V}^* = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*, \tag{3.5}$$

onde r é o posto da matriz, $\mathbf{D} \in \mathbb{R}^{n_1 \times n_2}$ é diagonal e contém os valores singulares positivos e $\mathbf{U} \in \mathbb{R}^{n_1 \times n_2}$ e $\mathbf{V} \in \mathbb{R}^{n_2 \times n_2}$ são os vetores singulares esquerdo e direito, respectivamente. Então, o fator de incoerência μ atesta que:

$$\max_i \; ||\mathbf{U}^* \mathbf{e}_i||^2 \leq \frac{\mu r}{m}, \quad \max_i \; ||\mathbf{V}^* \mathbf{e}_i||^2 \leq \frac{\mu r}{n}, \tag{3.6}$$

e

$$||\mathbf{U}\mathbf{V}^*||_\infty \leq \sqrt{\frac{\mu r}{mn}}. \tag{3.7}$$

Aqui, $||\mathbf{M}||_\infty = \max_{i,j} \; |\mathbf{M}_{ij}|$, que é a norma ℓ_∞ de $\mathbf{M}$ vista como se fosse um longo vetor. Como discutido em (Candès & Recht 2009), (Candès & Tao 2010) e (Gross 2011), a condição de incoerência afirma que para pequenos valores de μ, os vetores singulares são razoavelmente espalhados — ou seja, não esparsos.

Ainda resta a questão sobre o fato da matriz esparsa possuir posto baixo, o que ocorreria caso todos os registros diferentes de zero ocorressem em uma mesma coluna ou poucas colunas. Por exemplo, se a primeira coluna de $\mathbf{S}_0$ fosse o oposto da respectiva coluna em $\mathbf{L}_0$ e todas as outras colunas de $\mathbf{S}_0$ contivessem apenas zeros. Neste caso não seria possível recuperar $\mathbf{L}_0$ e

$\mathbf{S}_0$ por quaisquer métodos visto que $\mathbf{M} = \mathbf{L}_0 + \mathbf{S}_0$ teria um espaço de colunas igual ou contido naquele de $\mathbf{L}_0$. Para evitar tais situações sem sentido, assumiremos que o padrão de esparsidade do componente esparso é uniformemente aleatório (Candès et al. 2011).

Portanto, com apenas uma suposição de aleatoriedade (a respeito da esparsidade de $\mathbf{S}_0$) e considerando que os vetores singulares — ou componentes principais — de $\mathbf{L}_0$ são razoavelmente espalhados, a solução da PCP para o problema em 3.4 pode recuperar, com probabilidade próxima a 1, as componentes de posto baixo e esparsa de $\mathbf{M}$. Formalmente, de (Candès et al. 2011), temos:

Teorema 3.1. *Suponha que $\mathbf{L}_0 \in \mathbb{R}^{n\times n}$, obedeça 3.6 e 3.7. Estabeleça qualquer matriz $\mathbf{D} \in \mathbb{R}^{n\times n}$ de sinais. Suponha que o conjunto de suporte Ω de $\mathbf{S}_0 \in \mathbb{R}^{n\times n}$ é uniformemente distribuído entre todos os conjuntos de cardinalidade m, e que $\operatorname{sgn}([\mathbf{S}_0]_{ij}) = \mathbf{D}_{ij}$ para todo $(i,j) \in \Omega$. Então, existe uma constante numérica c tal qual com probabilidade ao menos $1 - cn^{-10}$ (sobre a escolha de suporte de $\mathbf{S}_0$), de que a PCP 3.4 com $\lambda = 1/\sqrt{n}$ seja exata, isto é, $\hat{\mathbf{L}} = \mathbf{L}_0$ e $\hat{\mathbf{S}} = \mathbf{S}_0$, dado que*

$$\operatorname{posto}(\mathbf{L}_0) \leq \rho_r n \mu^{-1} (\log n)^{-2} \quad \text{e} \quad m \leq \rho_s n^2. \tag{3.8}$$

Nesta equação, ρ_r e ρ_s são constantes numéricas positivas. No caso geral retangular, onde $\mathbf{L}_0 \in \mathbb{R}^{n_1\times n_2}$, a PCP com $\lambda = 1/\sqrt{n_{(1)}}$ obtém sucesso com probabilidade de pelo menos $1 - cn_{(1)}^{-10}$, dado que $\operatorname{posto}(\mathbf{L}_0) \leq \rho_r n_{(2)} \mu^{-1} (\log n_{(1)})^{-2}$ e $m \leq \rho_s n_1 n_2$

De fato, isto funciona para altos valores de posto e é capaz de recuperar $\mathbf{L}_0$ com probabilidade na ordem de $n/(\log n)^2$ — quando μ é muito alto — e elementos diferentes de zero em $\mathbf{S}_0$, com probabilidade n^2. Para evitar ambiguidades, determinamos que o nosso modelo para $\mathbf{S}_0$ é obtido como se segue: tome uma matriz arbitrária $\mathbf{S}$ e iguale a zero seus registros no conjunto aleatório Ω^c; isto retornará $\mathbf{S}_0$ (Candès et al. 2011). Ainda, sob a suposição do teorema, a minimização de

$$||\mathbf{L}||_* + \frac{1}{\sqrt{n_{(1)}}} ||\mathbf{S}||_1, \quad n_{(1)} = \max(n_1, n_2), \tag{3.9}$$

sempre retorna a resposta correta. Este é um fato curioso pois poderia-se esperar que o escalar correto λ devesse ser escolhido para balancear os termos em $||\mathbf{L}||_* + \lambda ||\mathbf{S}||_1$ de forma apropriada. Porém, este não é o caso e a escolha $\lambda = 1/\sqrt{n_{(1)}}$ parece ser universal, embora não esteja claro o porquê deste valor ser uma escolha apropriada, não importando quais sejam $\mathbf{L}_0$ e $\mathbf{S}_0$. A análise matemática revelou a eficácia deste valor e, apesar da prova do teorema fornecer uma faixa inteira de valores corretos, foi escolhido um que fosse suficientemente simples dentro da mesma faixa (Candès et al. 2011).

A RPCA trabalha com muitos dos conceitos de preenchimento de matrizes, que consiste na recuperação de uma matriz de posto baixo a partir de uma pequena fração de seus registros. Suas principais referências neste trabalho são (Candès & Recht 2009), (Candès & Tao 2010), (Candes & Plan 2010), (Keshavan, Montanari & Oh 2010), (Gross, Liu, Flammia, Becker & Eisert 2010) e (Gross 2011). Apesar destas similaridades entre os trabalhos, os resultados aqui são de uma natureza diferente, mesmo nosso problema de separação podendo ser visto, de certa forma, como um problema de preenchimento de matrizes. A principal diferença é que ao invés de termos uma fração de dados e querermos obter os outros faltantes, nós buscamos a informação que foi perdida e em seu lugar há apenas ruído. O problema de encontrar a informação perdida é menos complexo do que o problema de recuperá-la a partir de um dado corrompido. A RPCA, então, pode ser vista como uma extensão do problema de preenchimento de matrizes.

3.3 O Algoritmo

Existem várias abordagens para a solução do problema de separação de uma matriz $\mathbf{M}$ em uma componente de posto baixo $\mathbf{L}$ e uma componente esparsa $\mathbf{S}$ (em termos de otimização convexa), como pode ser visto em (Perception and Decision Lab 2012). Algumas delas sugerem até mesmo uma extensão do problema, incluindo um termo de ruído à expressão:

$$\mathbf{M} = \mathbf{L} + \mathbf{S} + \mathbf{N}, \tag{3.10}$$

onde $\mathbf{N} \in \mathbb{R}^{n_1 \times n_2}$, de mesmas dimensões das demais matrizes da equação, é composta por ruído. O termo também pode ser considerado como um elemento que não possui posto baixo e nem é esparso. Este é o caso do algoritmo GoDec (*Go Decomposition*) e sua variante SSGoDec (*Semi-Soft GoDec*) (Zhou & Tao 2011), usados neste trabalho.

Diferentemente da RPCA pura que modela os dados de acordo com a decomposição em 3.3, o GoDec considera um modelo similar a 3.10:

$$\mathbf{M} = \mathbf{L} + \mathbf{S} + \mathbf{N}, \quad \text{posto}\,(\mathbf{L}) \leq r, \ \text{card}\,(\mathbf{S}) \leq k, \tag{3.11}$$

ou seja, tanto o posto de $\mathbf{L}$ quanto a cardinalidade de $\mathbf{S}$ são estipulados antes de se iniciar os cálculos. O GoDec atribui a aproximação de posto r de $\mathbf{M} - \mathbf{S}$ a $\mathbf{L}$ e atribui a aproximação esparsa de cardinalidade k de $\mathbf{M} - \mathbf{L}$ a $\mathbf{S}$ (Zhou & Tao 2011). Este é um dos fatores que contribuem para a redução do custo computacional do algoritmo e o resultado, é claro, acaba sendo aproximado, e não exato como proposto na seção anterior para a RPCA.

Apesar disso, o algoritmo GoDec se apresentou uma excelente ferramenta para a decomposição dos dados em um componente de posto baixo e outro esparso (nosso objetivo ao estudarmos a RPCA). O algoritmo resolve o problema de otimização minimizando o erro de decomposição:

$$\begin{aligned} \underset{\mathbf{L},\mathbf{S}}{\text{minimizar}} \quad & ||\mathbf{N}||_F^2 = ||\mathbf{M} - \mathbf{L} - \mathbf{S}||_F^2 \\ \text{sujeito a} \quad & \text{posto}\,(\mathbf{L}) \leq j \\ & \text{card}\,(\mathbf{S}) \leq k, \end{aligned} \tag{3.12}$$

onde $||\mathbf{N}||_F^2$ denota a norma ℓ_F de Frobenius, isto é, a raiz quadrada da soma dos quadrados absolutos de seus elementos.

Essencialmente, o algoritmo GoDec utiliza o projeções aleatórias bilaterais (*Bilateral Random Projections* - BRP) para uma rápida aproximação de posto baixo para $\mathbf{L}$. Este é o principal elemento de aumento do desempenho do algoritmo. O custo computacional é inferior ao da SVD e o erro da aproximação BRP fica próximo ao erro da aproximação por SVD, sob suaves condições (Zhou & Tao 2011). Basicamente, dadas r BRPs:

$$\mathbf{L} = \mathbf{Y}_1(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1)^{-1}\mathbf{Y}_2^{\mathrm{T}}, \tag{3.13}$$

onde $\mathbf{Y}_1 = \mathbf{M}\mathbf{A}_1$ e $\mathbf{Y}_2 = \mathbf{M}^{\mathrm{T}}\mathbf{A}_2$, nas quais $\mathbf{A}_1 \in \mathbb{R}^{n_2 \times r}$ e $\mathbf{A}_2 \in \mathbb{R}^{n_1 \times r}$ são matrizes aleatórias.

Quando os valores singulares de $\mathbf{M}$ decaem vagarosamente, o desempenho de 3.13 fica comprometida. Para resolver o problema, foi projetada uma modificação para esta situação baseada em um esquema de performance (*power scheme*) (Roweis 1998). Na modificação por esquema de performance, é calculado a BRP de uma matriz $\tilde{\mathbf{M}} = (\mathbf{M}\mathbf{M}^{\mathrm{T}})^q\mathbf{M}$, cujos valores singulares decaem mais rapidamente do que em $\mathbf{M}$. Em particular, $\sigma_i(\tilde{\mathbf{M}}) = \sigma_i(\tilde{\mathbf{M}})^{2q+1}$. Ambas $\mathbf{M}$ e $\tilde{\mathbf{M}}$ compartilham os mesmos valores singulares (Zhou & Tao 2011). A BRP de $\tilde{\mathbf{M}}$ é:

$$\mathbf{Y}_1 = \tilde{\mathbf{M}}\mathbf{A}_1, \quad \mathbf{Y}_2 = \tilde{\mathbf{M}}^{\mathrm{T}}\mathbf{A}_2. \tag{3.14}$$

De acordo com a equação 3.13, a aproximação de posto r baseada em BRP fica:

$$\tilde{\mathbf{L}} = \mathbf{Y}_1(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1)^{-1}\mathbf{Y}_2^{\mathrm{T}}. \tag{3.15}$$

De forma a obter a aproximação de $\mathbf{L}$ com posto r, calculamos a decomposição QR de $\mathbf{Y}_1$ e $\mathbf{Y}_1$:

$$\mathbf{Y}_1 = \mathbf{Q}_1\mathbf{R}_1, \quad \mathbf{Y}_2 = \mathbf{Q}_2\mathbf{R}_2. \tag{3.16}$$

Finalmente, a aproximação de posto baixo de $\mathbf{L}$ é dada por:

$$\mathbf{L} = (\tilde{\mathbf{L}})^{\frac{1}{2q+1}} = \mathbf{Q}_1\left[\mathbf{R}_1(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1)^{-1}\mathbf{R}_2^{\mathrm{T}}\right]^{1/(2q+1)}\mathbf{Q}_2^{\mathrm{T}}. \tag{3.17}$$

A seguir, temos um sumário do algoritmo GoDec, reproduzido de (Zhou & Tao 2011):

Algoritmo 3.1 GoDec

Entrada: $\mathbf{M}$, r, k, ϵ, q
Saída: $\mathbf{L}$, $\mathbf{S}$
  $\mathbf{L}_0 \leftarrow \mathbf{M}$, $\mathbf{S}_0 \leftarrow \mathbf{0}$, $t \leftarrow 0$
  enquanto $||\mathbf{M} - \mathbf{L}_t - \mathbf{S}_t||_F^2/||\mathbf{M}||_F^2 > \epsilon$ **faça**
    $t \leftarrow t + 1$;
    $\tilde{\mathbf{L}} \leftarrow \left[(\mathbf{M} - \mathbf{S}_{t-1})(\mathbf{M} - \mathbf{S}_{t-1})^{\mathrm{T}}\right]^q(\mathbf{M} - \mathbf{S}_{t-1})$;
    $\mathbf{Y}_1 \leftarrow \tilde{\mathbf{L}}\mathbf{A}_1,\ \mathbf{A}_2 \leftarrow \mathbf{Y}_1$;
    $\mathbf{Y}_1 \leftarrow \tilde{\mathbf{L}}^{\mathrm{T}}\mathbf{Y}_1 = \mathbf{Q}_2\mathbf{R}_2,\ \mathbf{Y}_1 \leftarrow \tilde{\mathbf{L}}^{\mathrm{T}}\mathbf{Y}_2 = \mathbf{Q}_1\mathbf{R}_1$;
    se posto $(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1) < r$ **então**
      $r \leftarrow$ posto $(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1)$, volte ao primeiro passo;
    fim se
    $\mathbf{L}_t \leftarrow \mathbf{Q}_1\left[\mathbf{R}_1(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1)^{-1}\mathbf{R}_2^{\mathrm{T}}\right]^{1/(2q+1)}\mathbf{Q}_2^{\mathrm{T}}$;
    $\mathbf{S}_t \leftarrow \mathcal{P}_\Omega(\mathbf{M}-\mathbf{L}_t)$, onde Ω é o subconjunto de não zeros dos k maiores registros de $|\mathbf{X}-\mathbf{L}_t|$;
  fim enquanto

De acordo com (Zhou & Tao 2011), foram feitos experimentos onde compararam o desempenho do GoDec (disponível em (Zhou 2013)) com o desempenho da RPCA (`inexact_alm_rpca`, em (Perception and Decision Lab 2012)). Os resultados apontaram que ambos obtiveram sucesso na estimação das decomposições em posto baixo e esparsa, porém com erros relativos menores (menos de 10^{-6}) com o GoDec e vários segundos de vantagem para o mesmo em relação à RPCA. A melhoria na acurácia é devida ao modelo de decomposição do GoDec em 3.11, que adiciona um termo de ruído sendo mais generalista do que o modelo da RPCA em 3.3. Já a melhoria na velocidade se deve à aproximação de posto baixo baseada em BRP, que significativamente poupa esforço computacional (Zhou & Tao 2011).

Diferentemente do GoDec, que impõe uma rigorosa condição de limiar para os valores singulares da componente de posto baixo $\mathbf{L}$ e os registros da componente esparsa $\mathbf{S}$, a variante do GoDec, o SSGoDec (*Semi-soft Go Decomposition*) adota um limiar suave para os registros de $\mathbf{S}$. Esta alteração traz duas vantagens principais: a primeira é que o parâmetro k, restrito em 3.12, agora pode ser determinado automaticamente por um limite suavizado por uma regularização de norma ℓ_1. Dessa forma, evita-se a situação onde o parâmetro k escolhido é muito grande e uma parte do ruído $\mathbf{N}$ vaza para $\mathbf{S}$; a segunda é que o custo computacional, e por consequência o tempo de processamento, é substancialmente menor do que usando o GoDec (Zhou 2013).

Para o SSGoDec, o problema de otimização é então alterado:

$$\begin{aligned} \underset{\mathbf{L},\mathbf{S}}{\text{minimizar}} \quad & ||\mathbf{M}-\mathbf{L}-\mathbf{S}||_F^2 + \lambda||\mathbf{S}||_1 \\ \text{sujeito a} \quad & \text{posto}\,(\mathbf{L}) \leq j, \end{aligned} \tag{3.18}$$

onde λ é o limiar de suavização, muito mais fácil de se encontrar do que k em 3.12. Isto se deve ao fato do erro resultante de decomposição ser mais robusto às alterações de λ (Zhou & Tao 2013). As outras alterações podem ser vistas no algoritmo do SSGoDec:

Algoritmo 3.2 Semi-soft GoDec

Entrada: $\mathbf{M}$, r, λ, ϵ, q
Saída: $\mathbf{L}$, $\mathbf{S}$
  $\mathbf{L}_0 \leftarrow \mathbf{M}$, $\mathbf{S}_0 \leftarrow \mathbf{0}$, $t \leftarrow 0$
  enquanto $||\mathbf{M}-\mathbf{L}_t-\mathbf{S}_t||_F^2/||\mathbf{M}||_F^2 > \epsilon$ **faça**
    $t \leftarrow t+1;$
    $\tilde{\mathbf{L}} \leftarrow \left[(\mathbf{M}-\mathbf{S}_{t-1})(\mathbf{M}-\mathbf{S}_{t-1})^{\mathrm{T}}\right]^q(\mathbf{M}-\mathbf{S}_{t-1});$
    $\mathbf{Y}_1 \leftarrow \tilde{\mathbf{L}}\mathbf{A}_1,\ \mathbf{A}_2 \leftarrow \mathbf{Y}_1;$
    $\mathbf{Y}_1 \leftarrow \tilde{\mathbf{L}}^{\mathrm{T}}\mathbf{Y}_1 = \mathbf{Q}_2\mathbf{R}_2,\ \mathbf{Y}_1 \leftarrow \tilde{\mathbf{L}}^{\mathrm{T}}\mathbf{Y}_2 = \mathbf{Q}_1\mathbf{R}_1;$
    se $\text{posto}\,(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1) < r$ **então**
      $r \leftarrow \text{posto}\,(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1)$, volte ao primeiro passo;
    fim se
    $\mathbf{L}_t \leftarrow \mathbf{Q}_1\left[\mathbf{R}_1(\mathbf{A}_2^{\mathrm{T}}\mathbf{Y}_1)^{-1}\mathbf{R}_2^{\mathrm{T}}\right]^{1/(2q+1)}\mathbf{Q}_2^{\mathrm{T}};$
    $\mathbf{S}_t \leftarrow \mathcal{P}_\lambda(\mathbf{M}-\mathbf{L}_t)$, onde $\mathcal{P}_\lambda(x) = \text{sgn}\,(x)\max\,(|x|-\lambda,0);$
  fim enquanto

Foram também testados outros algoritmos para a solução da RPCA:

- Multiplicador de Lagrange Aumentado (*Augmented Lagrange Multiplier* - ALM) em (Perception and Decision Lab 2012);
- Gradiente Próximo Acelerado (*Accelerated Proximal Gradient* - APG) em (Perception and Decision Lab 2012);
- Método Dual (*Dual Method*) em (Perception and Decision Lab 2012);
- Limiarização por Valor Singular (*Singular Value Thresholding* - SVT) em (Perception and Decision Lab 2012);
- Método de Direção Alternada (*Alternating Direction Method* - ADM) em (Perception and Decision Lab 2012);
- PCA Robusto Bayesiano (*Bayesian robust PCA*) em (Carin 2013);

Dentre toda esta variedade de soluções, selecionamos os métodos do GoDec e do SSGoDec para estudarmos mais a fundo. Finalmente, selecionamos o algoritmo SSGoDec como a nossa solução para o problema da RPCA por seus melhores resultados, apesar de sua alta velocidade (o que apenas enfatiza sua eficiência).

Para o mesmo exemplo de dado sísmico sintético usado nos exemplos de decomposições anteriores, o resultado obtido pode ser visto na Figura 3.2. A decomposição em 3.11, que adiciona um termo extra de ruído, é bastante conveniente para o nosso tipo de problema. Afinal, uma classificação de eventos sísmicos mais precisa, em termos de decomposição de sinais, seria a soma de um componente de posto baixo, um componente esparso (ou tudo aquilo o que não tem posto baixo e também não é ruído) e o ruído propriamente dito. A variação dos parâmetros de entrada no SSGoDec (posto, limiar de suavização e fator de esquema de performance), permitiu visualizar como podemos trocar elementos de lugar entre os subespaços, por exemplo, incluindo mais ou menos elementos do subespaço esparso no subespaço de ruído.

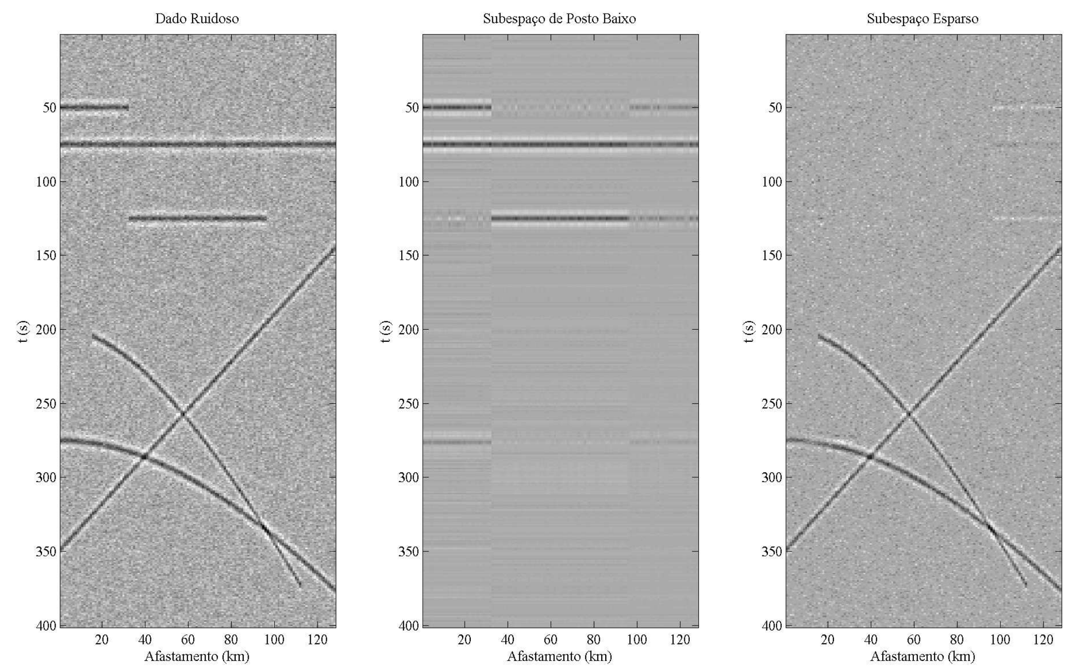

Figura 3.2: Decomposição de $\mathbf{M}$ em subespaço de posto baixo e subespaço esparso, usando SSGoDec (RPCA).

O método de decomposição aqui apresentado e o algoritmo a ele associado, constituem a nossa contribuição teórica. No Capítulo 4, vamos apresentar os resultados práticos obtidos, comparando-os com a filtragem por Transformada *f*-*k*.

Capítulo 4

Resultados

Iniciamos este capítulo com a apresentação do resultado das técnicas de decomposição apresentadas nos Capítulos 3 e 4 em dados sintéticos, que simulam problemas de separação de eventos cruzados e separação entre reflexões e difrações. Introduzimos também um modelo sintético de ruído de *ground roll*. Em seguida são apresentados os resultados obtidos em dados reais pré-processados.

4.1 Dados Sintéticos

A principal motivação do trabalho é a aplicação da técnica da RPCA aos dados sísmicos para a separação dos eventos de interesse do ruído *ground roll*. Entretanto, esta aplicação depende de testes prévios em um ambiente controlado. Para isso, usamos dados sintéticos que simulam os eventos sísmicos, as ondas de *ground roll* e o ruído aleatório de modo a imitar ao máximo os dados reais, porém com a capacidade de comparar os resultados obtidos com um modelo conhecido em sua totalidade.

Dentre os possíveis cenários que podem ser encontrados em estudos de dados sísmicos, estão os *eventos cruzados* e *difrações*. Estes dois exemplos foram a motivação para os trabalhos apresentados na Conferência da Associação Europeia de de Geocientistas e Engenheiros (*European Association of Geoscientists & Engineers* - EAGE) e na Conferência Europeia de Processamento de Sinais (*European Signal Processing Conference* - EUSIPCO), ambos em 2012. A seguir, serão reproduzidos tais experimentos.

O processamento e as simulações deste capítulo foram executados usando o MathWorks MATLAB R2012b (The MathWorks, Inc. 2012) e o CWP Seismic Un*x 43R3 (Center for Wave Phenomena 2012).

4.1.1 Separação de Eventos Cruzados

Para simular o cruzamento entre dois refletores (eventos) foi utilizado o dado da Figura 4.1, onde também podemos ver o perfil dos dez primeiros valores singulares, referentes às dez primeiras autoimagens. Este número foi escolhido por duas razões: a primeira é que é necessária uma redução de dimensionalidade para a aplicação da SVD-ICA, reduzindo o custo computacional. Logo, para isto temos $R = 10$ para a equação 2.15; a segunda razão é a baixa amplitude dos valores singulares seguintes, tornando-se muito pouco representativos para os exemplos utilizados. Estes dois argumentos também valem para todos os outros resultados apresentados mais adiante.

Este dado foi corrompido com ruído gaussiano branco aditivo com SNR de 14 dB. O número de traços é $m = 40$ e o número de amostras é $n = 70$. Percebe-se, através do perfil dos valores singulares, que a primeira autoimagem tem um peso consideravelmente mais alto do que as outras na composição do dado. A diferença entre o segundo valor singular e o décimo é muito menor do que a diferença daquele para o primeiro valor singular. Logo, o número de autoimagens escolhido para representar o subespaço de sinal no cálculo da SVD é $p = 1$, na equação 2.4. É usado o mesmo valor como parâmetro de número de imagens extraídas pela ICA no cálculo da SVD-ICA ($Q = 1$ em 2.18), representando igualmente o subespaço de sinal desta decomposição. No caso do cálculo da RPCA, isto é o equivalente a selecionar posto $j = 1$ como parâmetro de entrada do algoritmo.

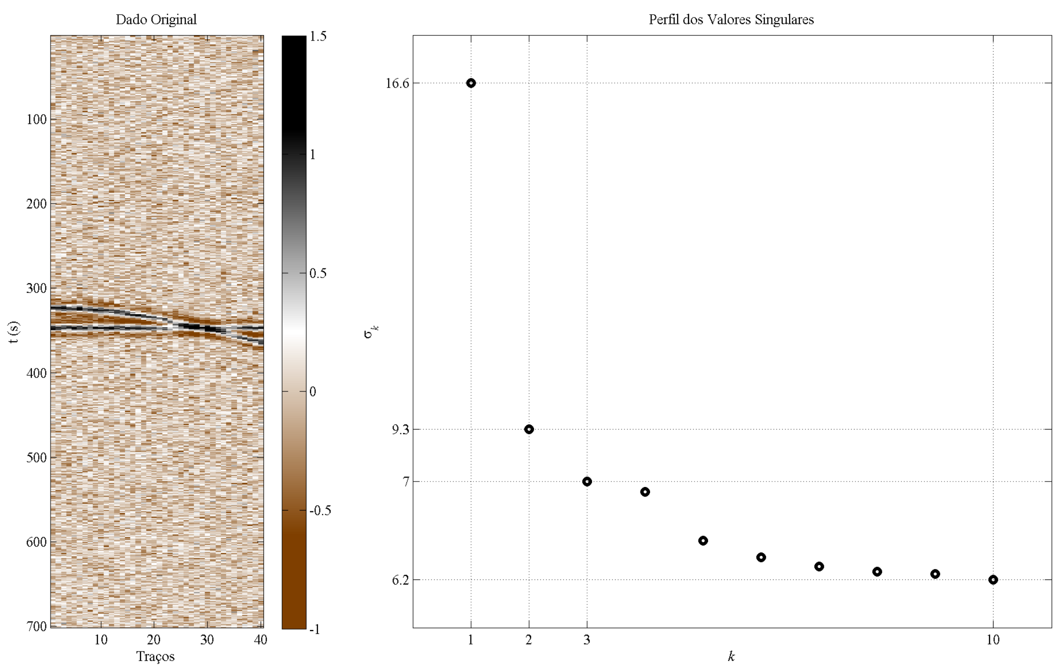

Figura 4.1: Dado utilizado no experimento de separação de eventos cruzados.

Os resultados a seguir seguiram as características supracitadas e replicam os resultados apresentados em (Duarte, Nadalin, Nose Filho, Zanetti, Romano & Tygel 2012a) (74th EAGE *Conference & Exhibition incorporating* SPE EUROPEC 2012) e (Duarte, Nadalin, Filho, Zanetti, Romano & Tygel 2012b) (20th *European Signal Processing Conference* EUSIPCO 2012).

Como pode ser observado na Figura 4.2, a decomposição SVD consegue separar quase que totalmente os dois eventos, incluindo o evento curvilíneo no subespaço de ruído. Porém, a decomposição cria alguns artefatos e descontinuidades em ambos os subespaços, como uma linha branca acima do evento retilíneo no subespaço de sinal e uma interrupção no evento curvilíneo no subespaço de ruído. O tempo de processamento da decomposição por SVD foi de 0,0288 segundo.

A decomposição híbrida SVD-ICA, que aplica a ICA em $R = 10$ autoimagens extraídas usando SVD, como dito anteriormente, apresenta a primeira autoimagem pós-ICA como subespaço de sinal na Figura 4.3. Podemos observar que a separação dos eventos é melhor executada

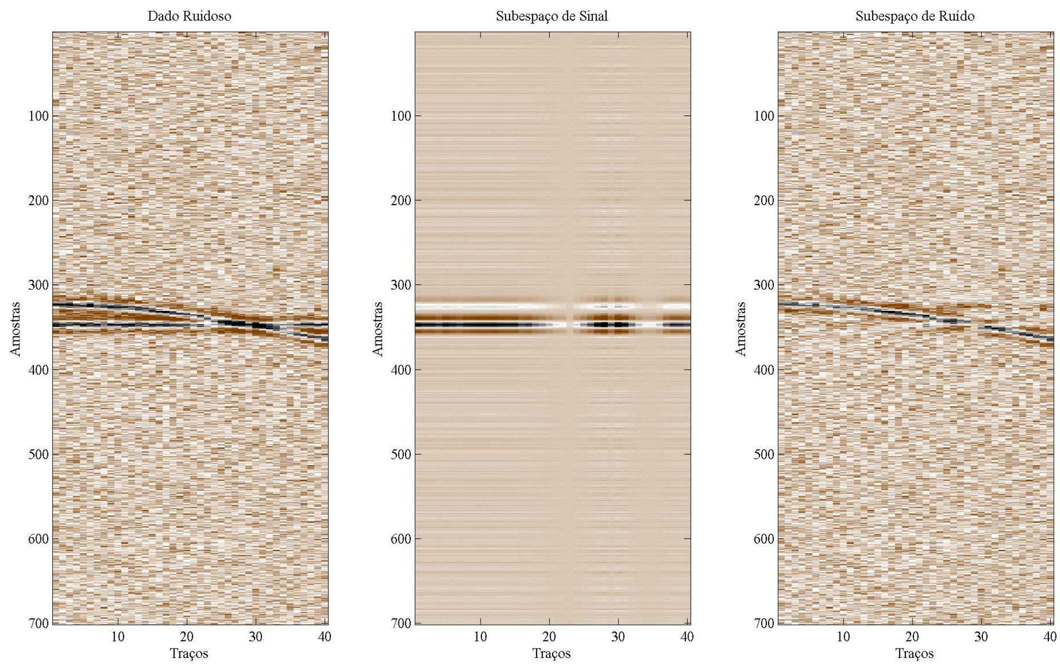

Figura 4.2: Resultado da decomposição SVD para o caso de eventos cruzados.

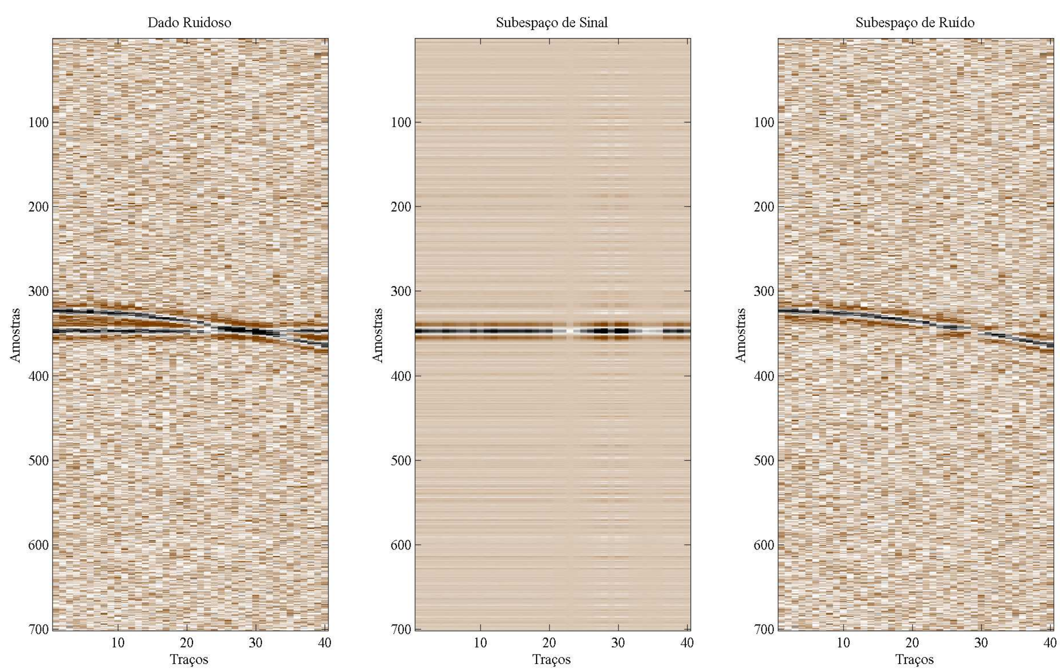

Figura 4.3: Resultado da decomposição SVD-ICA para o caso de eventos cruzados.

neste caso, com o evento retilíneo bem definido no subespaço de sinal e o evento curvilíneo com apenas uma pequena interrupção junto do ruído, no subespaço de ruído. Ainda existem alguns artefatos nas imagens, mas a melhoria em relação à aplicação da SVD pura é evidente. Porém, em termos de tempo de execução do algoritmo, a decomposição SVD-ICA levou 2,5029 segundos para terminar. Um tempo consideravelmente mais alto do que o necessário pela SVD.

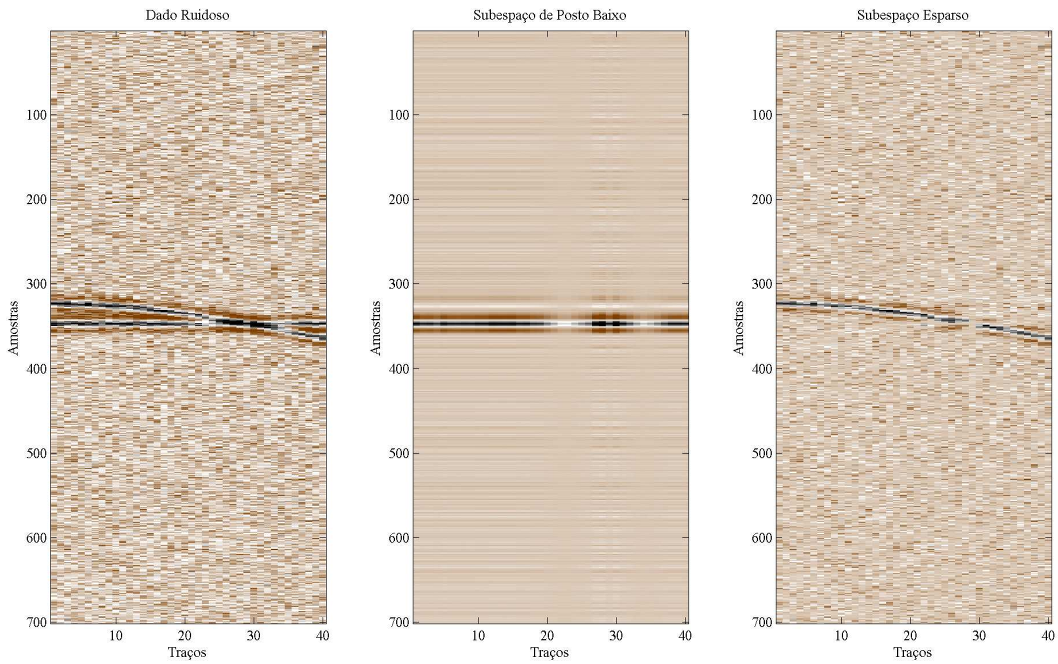

Figura 4.4: Resultado da decomposição RPCA para o caso de eventos cruzados.

Já no caso da decomposição por RPCA, cujo resultado pode ser observado na Figura 4.4, a tendência é classificar a separação em um meio-termo entre a separação via SVD e a separação via SVD-ICA. No subespaço de posto baixo (o equivalente ao subespaço de sinal nas outras decomposições), a RPCA foi capaz de recuperar muito bem o evento retilíneo, porém com resquícios do mesmo artefato deixado pela decomposição SVD. No subespaço esparso (análogo ao subespaço de ruído para as outras decomposições), o evento curvilíneo é quase tão bem recuperado como no caso da SVD-ICA. Entretanto, a grande diferença da RPCA para as outras duas decomposições é que esta conseguiu eliminar grande parte do ruído que estaria no subespaço esparso. Esta é uma característica do algoritmo SSGoDec ajustada pelo parâmetro de cardinalidade (no caso, $k = 0{,}1$ em 3.12), que é ajustada de forma a diferenciar o ruído dos registros esparsos. O tempo necessário para os cálculos foi de 0,0859 segundo, ainda significativamente menor do que a SVD-ICA.

4.1.2 Separação de Difrações e Reflexões

Outro modelo sintético foi utilizado para a separação de difrações de reflexões, desta vez com $m = 300$ traços e $n = 300$ amostras e SNR de 20 dB. Uma característica que torna este modelo um tanto quanto desafiador, é que as difrações, que possuem formato hiperbólico, possuem amplitudes menores do que os refletores. Além disso, as amplitudes não são constantes ao longo das hipérboles, como pode ser visto na Figura 4.5.

Nesta mesma figura, observamos o perfil dos valores singulares para o dado em questão e identificamos que as duas primeiras autoimagens contêm quase que toda a informação de interesse. Sendo assim, o subespaço de sinal será representado agora pelas duas primeiras

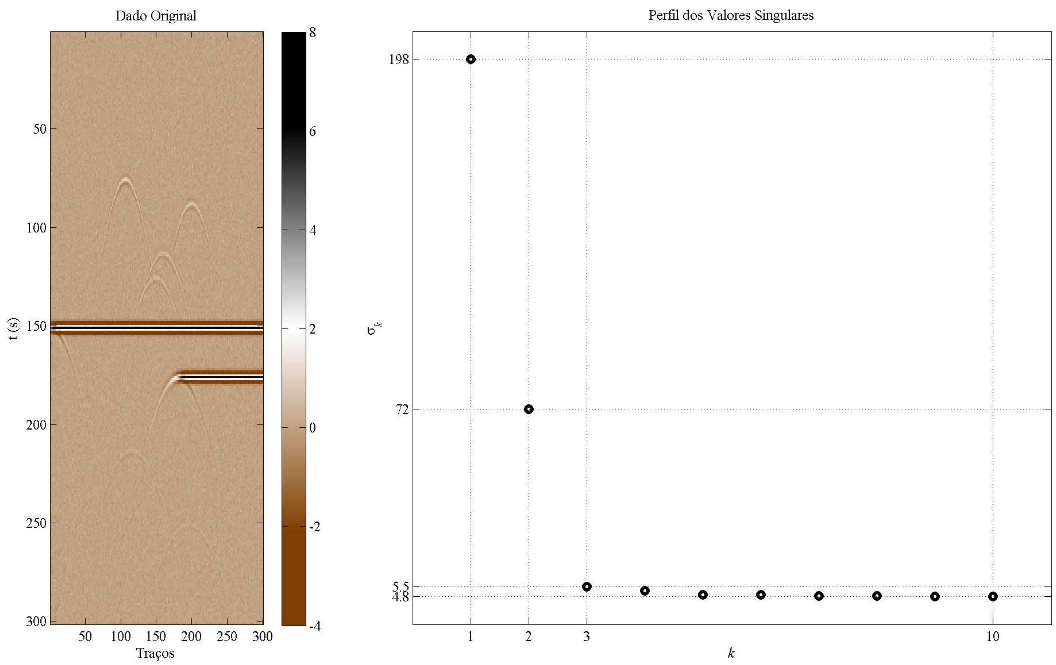

Figura 4.5: Dado utilizado no experimento de separação de difrações e reflexões.

autoimagens da SVD (o equivalente a duas autoimagens da ICA em SVD-ICA e posto dois como entrada para o cálculo da RPCA).

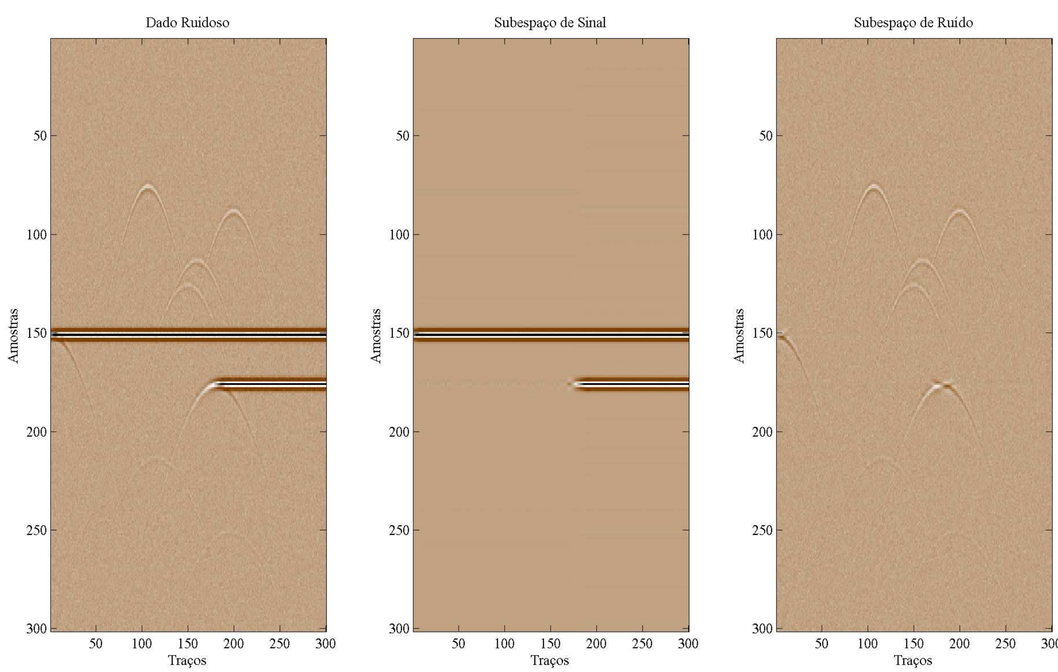

Figura 4.6: Resultado da decomposição SVD para a separação de difrações e reflexões.

O resultado da decomposição via SVD na Figura 4.6 mostra que esta decomposição consegue separar muito bem as reflexões das difrações. O ruído do sinal ficou praticamente todo no subespaço de ruído, o que confirma que foi observado anteriormente no perfil de valores singulares na Figura 4.5. O tempo decorrido nesta separação foi de 0,0442 segundo.

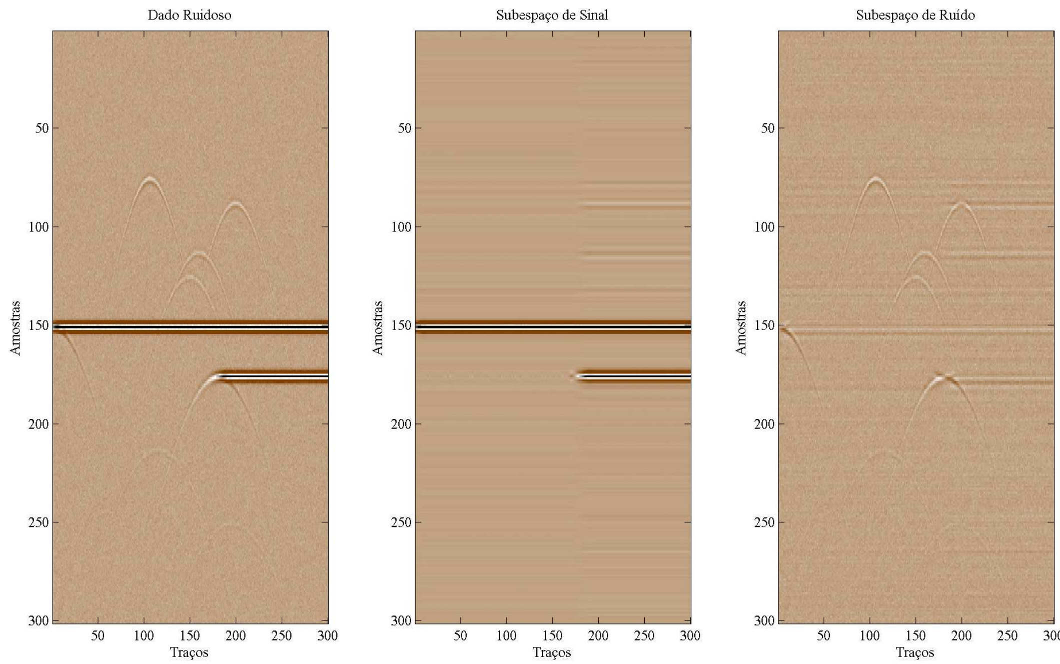

Figura 4.7: Resultado da decomposição SVD-ICA para a separação de difrações e reflexões.

Diferentemente do caso anterior, com eventos cruzados, onde a SVD-ICA obteve o melhor resultado dentre as três decomposições, neste caso podemos observar o inverso. Percebe-se claramente o surgimento de artefatos em ambos os subespaços e resquícios dos refletores no subespaço de ruído na Figura 4.7, o que significa que a separação não foi completa. Assim como na seção anterior, o tempo de execução do algoritmo SVD-ICA foi o maior dentre os três, levando agora 2,378 segundos para completar a separação.

De acordo com a Figura 4.8, a decomposição RPCA obteve um resultado ligeiramente melhor do que a decomposição SVD quando comparamos o subespaço de sinal desta com o subespaço de posto baixo da RPCA. Já no subespaço esparso, as difrações estão um pouco melhor definidas e, assim como no caso de eventos cruzados, o ruído foi consideravelmente eliminado, resultando em uma imagem mais limpa. Em termos de desempenho computacional, a RPCA executou a separação em 0,253 segundo.

4.1.3 Eliminação de *Ground Roll*

A utilização de um dado sintético que simule as características das ondas de *ground roll* ocorreu diversas vezes durante os trabalhos. Porém, para este caso, tentamos simular um dado onde o espectro *f-k* dos eventos fosse sobreposto pelo espectro das ondas de *ground roll* e a comparação fosse também feita com a filtragem *f-k*, que é o método mais comumente utilizado em processamento de dados sísmicos.

O que torna este cenário mais desafiador é esta sobreposição de espectros. Normalmente, devido à diferença de velocidade das ondas correspondentes aos eventos e as ondas de *ground*

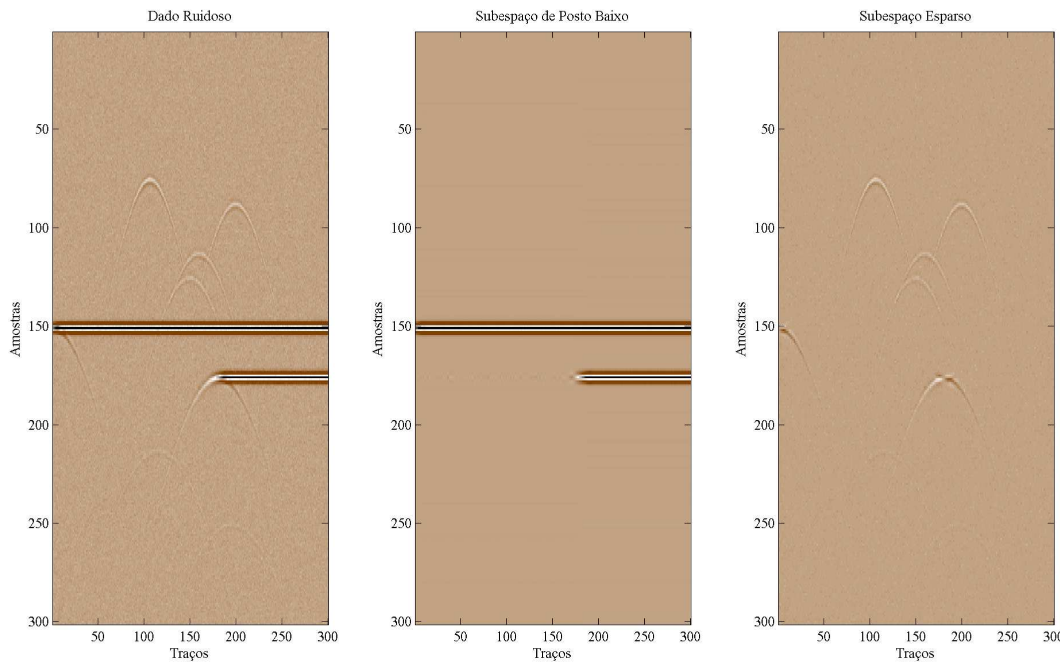

Figura 4.8: Resultado da decomposição RPCA para a separação de difrações e reflexões.

roll, seus espectros permanecem separados, o que torna a tarefa fácil para uma filtragem que utilize uma *slope* de acordo com a localização do *ground roll* no espectro *f-k* do dado sísmico (Yilmaz 2001). Por outro lado, quando os espectros estão sobrepostos, é mais difícil saber a localização exata do *ground roll* no dado, além do fato de que a filtragem *f-k* elimina informação de interesse juntamente com o ruído.

O intuito deste experimento é verificar se a RPCA — e sua capacidade de recuperação de registros corrompidos — é capaz de separar o *ground roll* dos eventos de interesse em um cenário de sobreposição de espectros *f-k* sem perdas. Sendo assim, utilizamos o dado sintético da Figura 4.9, que simula uma seção CMP. Como pode ser observado, o dado segue o modelo em 2.3. Na Figura 4.10 podemos observar o espectro *f-k* de cada componente separadamente e ambos sobrepostos.

Como pode ser observado no perfil de valores singulares na Figura 4.11, a informação de interesse está espalhada em várias autoimagens. Isto se deve ao fato dos eventos não estarem alinhados como nos casos analisados até agora. Para podermos aplicar as técnicas de decomposição de sinais estudadas, é necessário um pré-processamento dos dados, de forma a horizontalizar os eventos.

Sendo assim, as decomposições de sinais serão aplicadas em seções pós-correção NMO. Como os dados são sintéticos e sabemos a velocidade dos eventos de interesse, os mesmos serão horizontalizados enquanto que o *ground roll* sofrerá distorção. O lado positivo é que, desta forma, os métodos separarão com maior facilidade os eventos do ruído. Porém a correção NMO provoca um *stretching* (alargamento) das extremidades do dado, obrigando-nos a eliminar estes segmentos por *muting* (silenciamento), levando, como consequência, à perda de informação de interesse. A Figura 4.12 mostra os dados após a correção NMO e exclusão das áreas com *stretching*, juntamente do novo perfil de valores singulares. A partir da análise deste perfil, definimos $p = 2$ para a SVD, $Q = 2$ para a SVD-ICA e finalmente $j = 2$ para a RPCA.

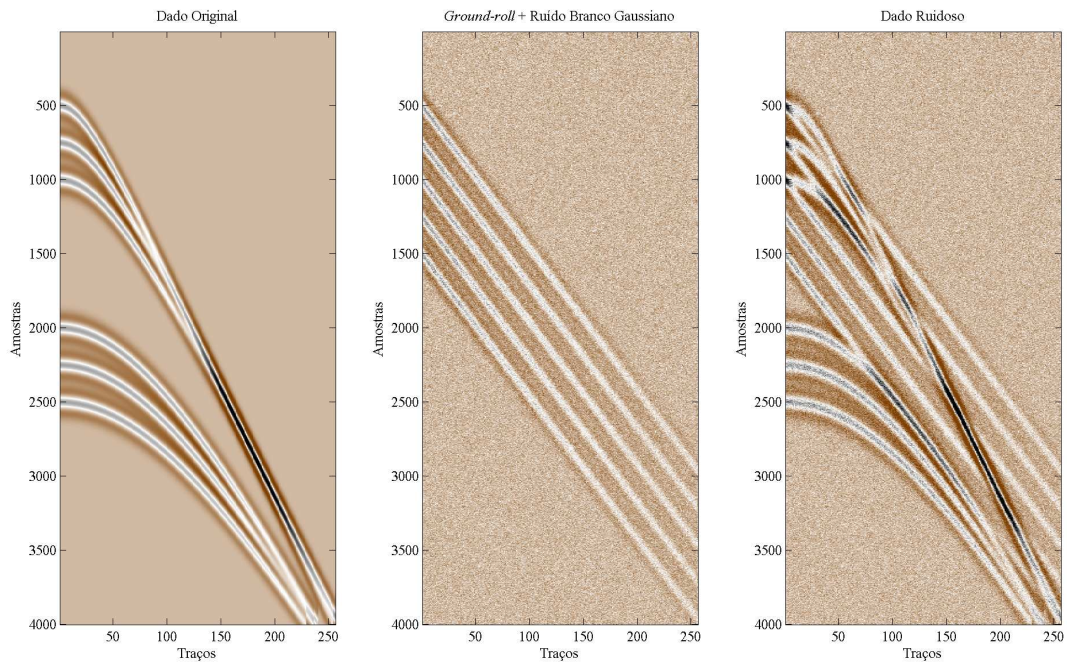

Figura 4.9: Composição do dado sintético para eliminação de *ground roll*.

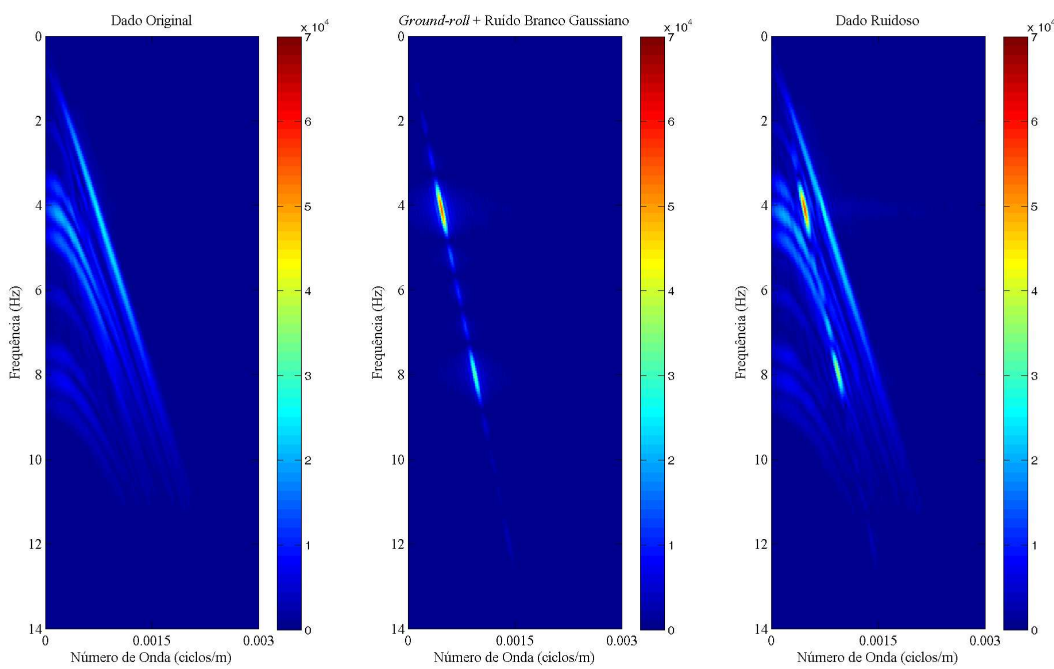

Figura 4.10: Espectro *f-k* do dado sintético para eliminação de *ground roll*.

A filtragem *f-k* é normalmente aplicada antes do processamento NMO e seu resultado pode ser observado na Figura 4.13 e seu respectivo espectro *f-k* na Figura 4.14. Percebemos claramente que a filtragem elimina uma parcela considerável de informação de interesse, juntamente

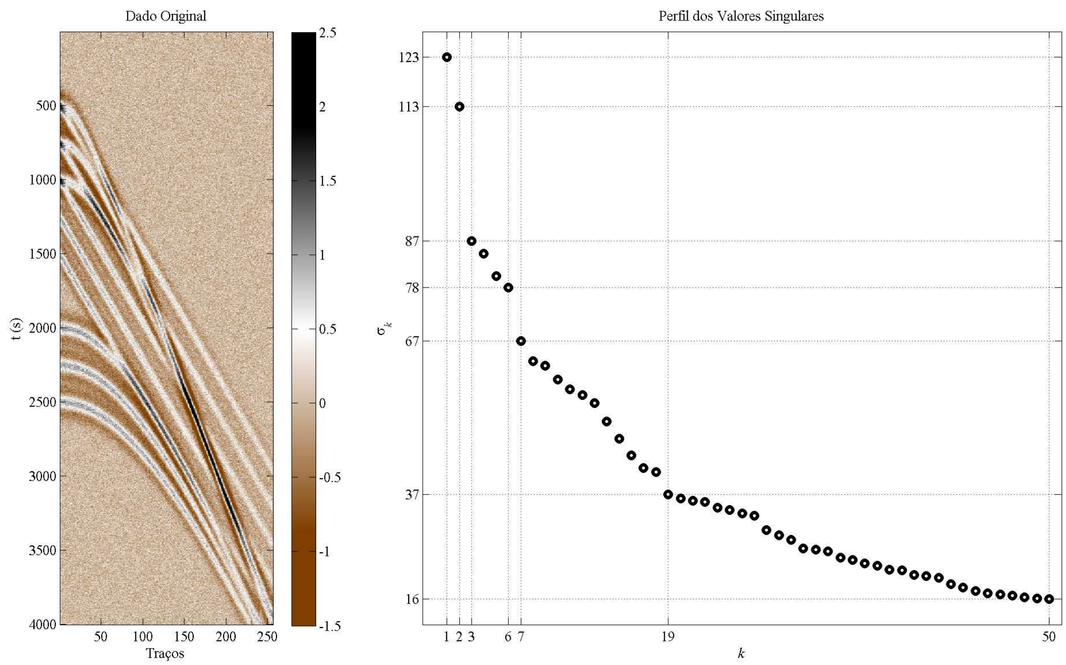

Figura 4.11: Dado original sem pré-processamento e o perfil de valores singulares.

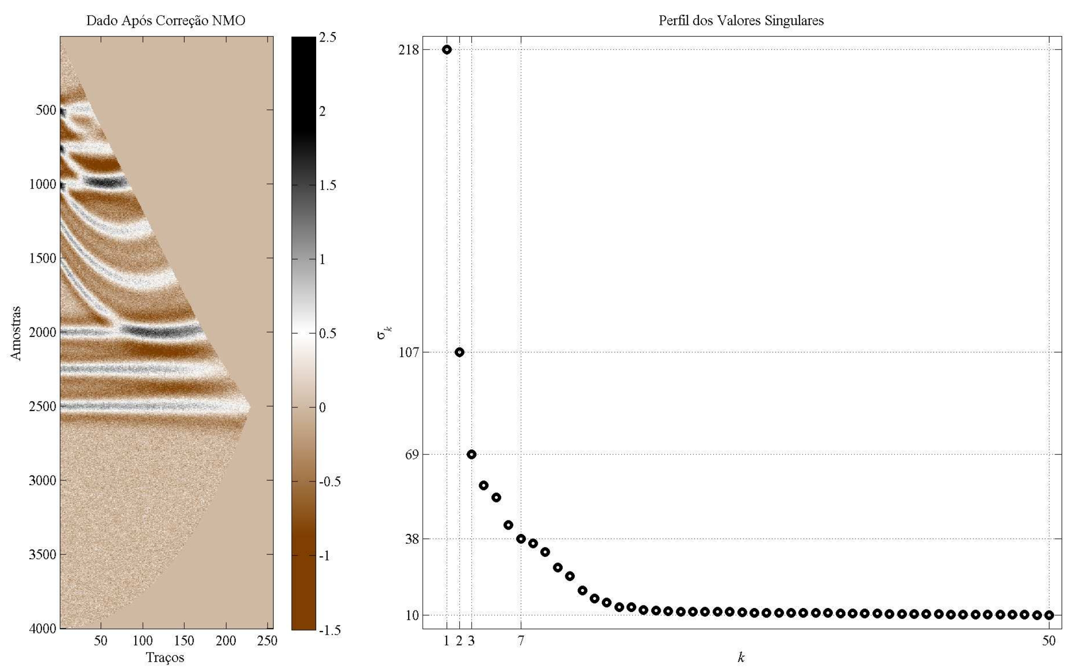

Figura 4.12: Dado pós-correção NMO e o novo perfil de valores singulares.

com a maior parte do ruído de *ground roll*. Também é possível observar alguns artefatos próximos às bordas no subespaço de sinal da Figura 4.13, que normalmente são eliminados em etapas posteriores de processamento por meio de *muting*.

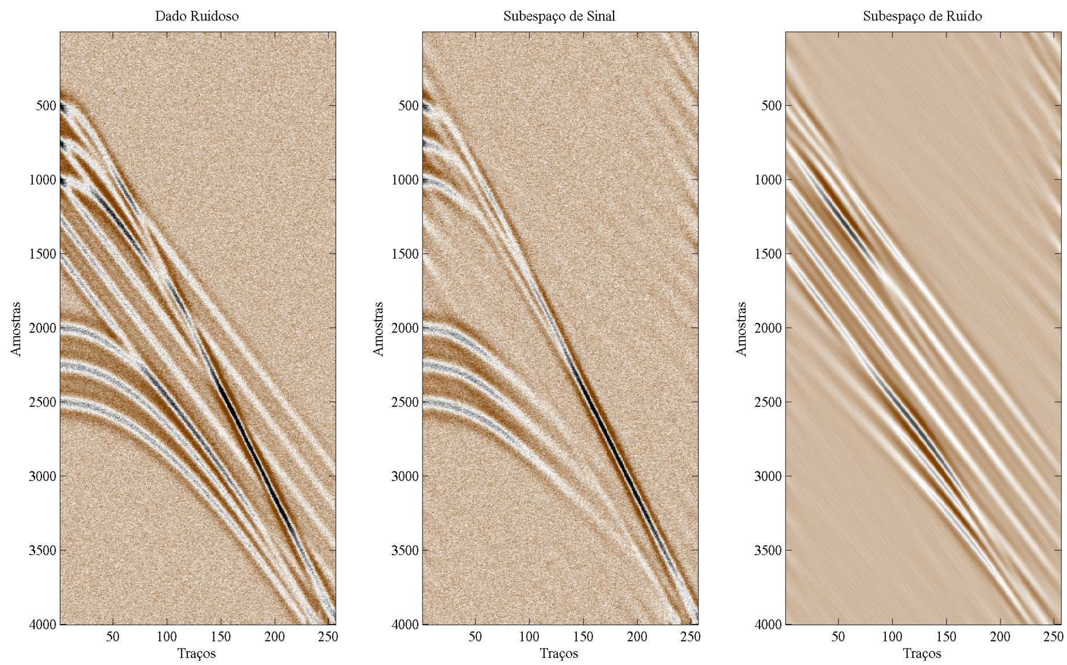

Figura 4.13: Dado original pós-filtragem *f-k*.

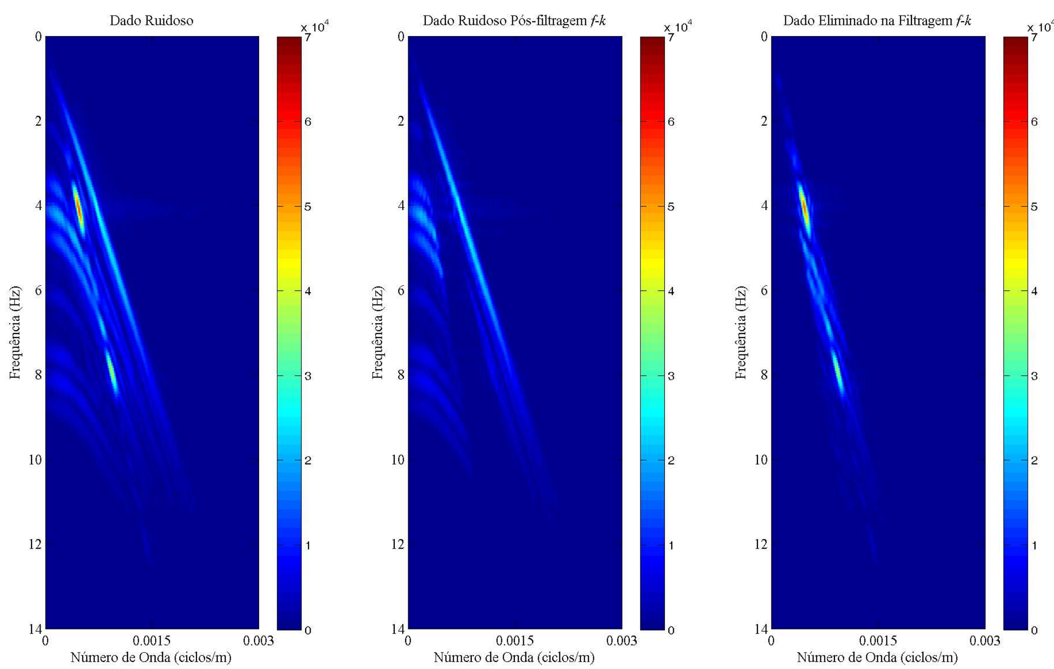

Figura 4.14: Espectro do dado original pós-filtragem *f-k*.

Como dito anteriormente, o dado pós-correção NMO perde uma certa quantidade de informação de interesse devido ao *muting* de *stretching*. Logo, fazer uma comparação dos dados obtidos na Figura 4.13 com os resultados das decomposições a seguir pode não parecer justo. Sendo assim, mais adiante, apresentamos também uma situação na qual a filtragem *f-k* é executada nos dados pós-processamento NMO.

Na sequência, exibiremos as ilustrações das decomposições corrigidas por NMO seguidas dos respectivos resultados finais, após aplicarmos a NMO inversa.

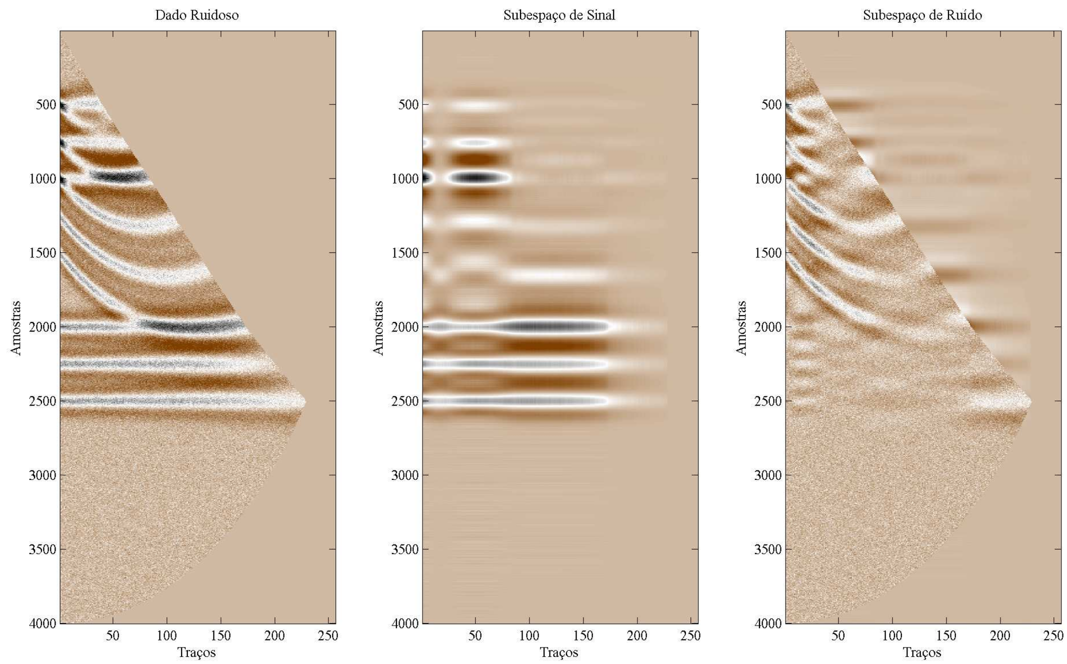

Figura 4.15: Resultado da decomposição SVD para o dado corrigido via NMO.

Os resultados observados nas Figuras 4.15 e 4.16 mostram que a SVD conseguiu separar a maior parte do *ground roll* dos dados, junto da maioria do ruído aleatório. Entretanto surgiram refletores falsos no resultado final, pois a inversa da NMO fez com que os artefatos deixados pela SVD fossem "corrigidos", visto que eram horizontais antes da transformação. Como esperávamos, parte dos dados foi perdida devido à extração de *stretching* (fato que se repetirá em todos os casos seguintes).

A SVD-ICA, como pode ser visto nas Figuras 4.17 e 4.18, recuperou parcialmente apenas os três últimos refletores, criando diversos artefatos acima deles. Por isso, o subespaço de sinal corrigido pela inversa NMO apresentou todos estes artefatos como tênues falsos refletores. O *ground roll* foi praticamente eliminado com o restante do ruído e uma fração dos refletores inferiores, como pode ser observado no subespaço de ruído.

A RPCA, dentre as três decomposições, foi a que melhor recuperou os refletores, apesar de restar muito pouco do refletor mais ao topo. Como pode ser visto nas Figuras 4.19 e 4.20, de forma similar à SVD, os dois artefatos deixados se tornaram refletores falsos. O ground roll também foi removido, junto de todo o ruído restante.

A Figura 4.21 demonstra o desempenho da filtragem *f-k* na remoção do *ground roll* no dado recuperado após a inversa da NMO. Resta muito pouco do *ground roll* no subespaço de sinal, porém os eventos também perdem amplitude e há a criação de diversos artefatos. O ruído aleatório permanece no subespaço de sinal, ao contrário das decomposições vistas anteriormente.

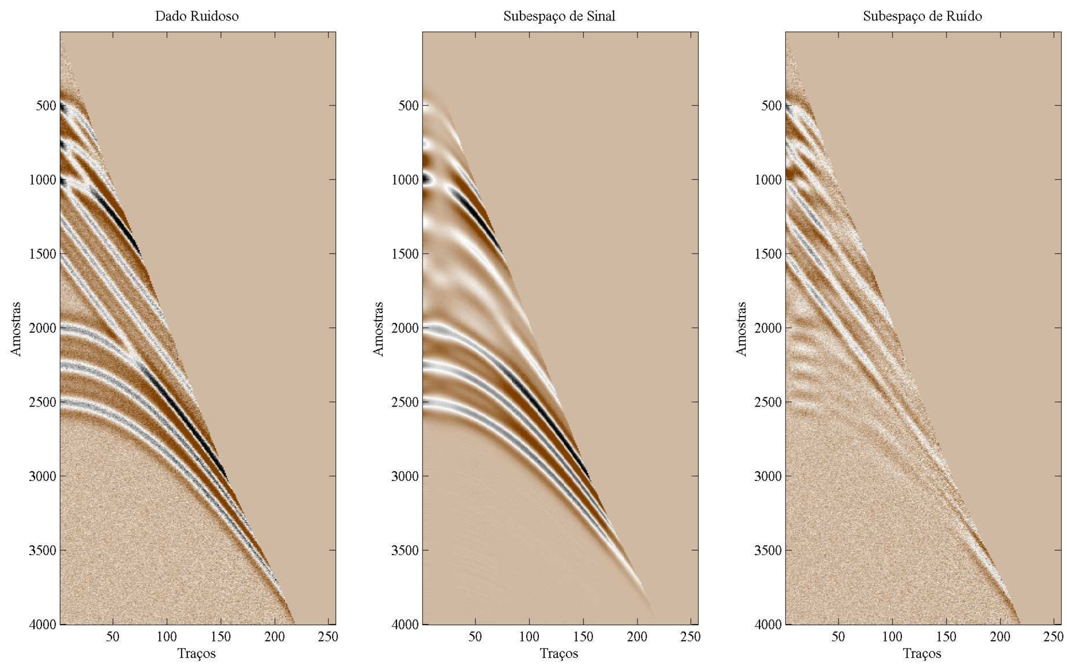

Figura 4.16: Resultado final para a decomposição SVD após a NMO inversa.

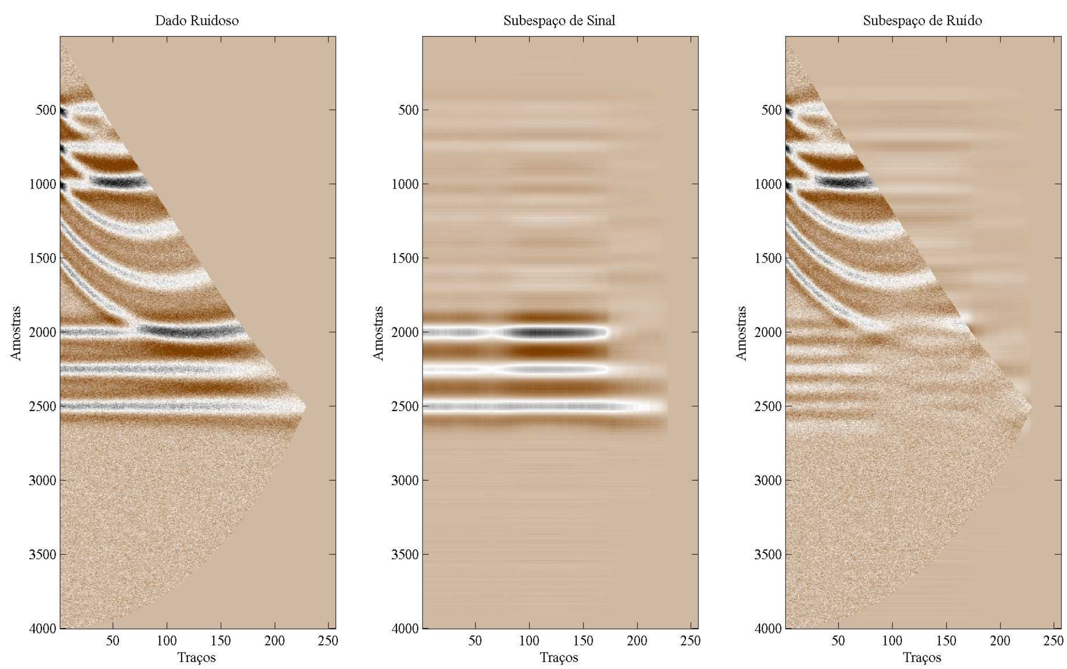

Figura 4.17: Resultado da decomposição SVD-ICA para o dado corrigido via NMO.

4.2 Dados Reais

A aplicação da decomposição RPCA em um dado real era o maior objetivo deste trabalho. Após as diversas análises executadas em dados sintéticos, a experiência adquirida permitiu

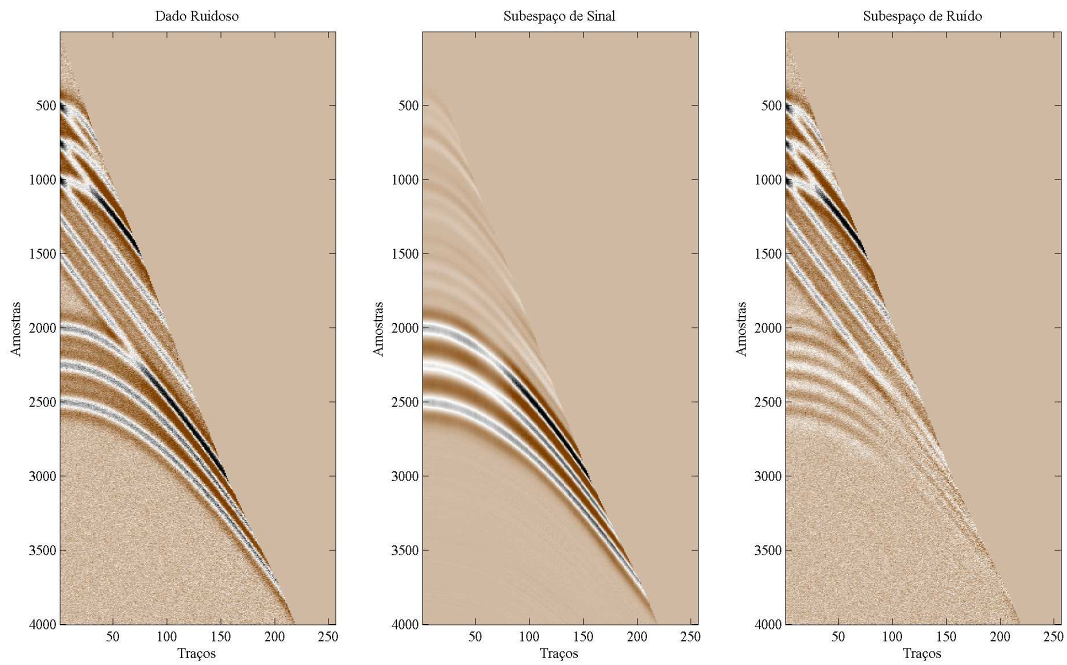

Figura 4.18: Resultado final para a decomposição SVD-ICA após a NMO inversa.

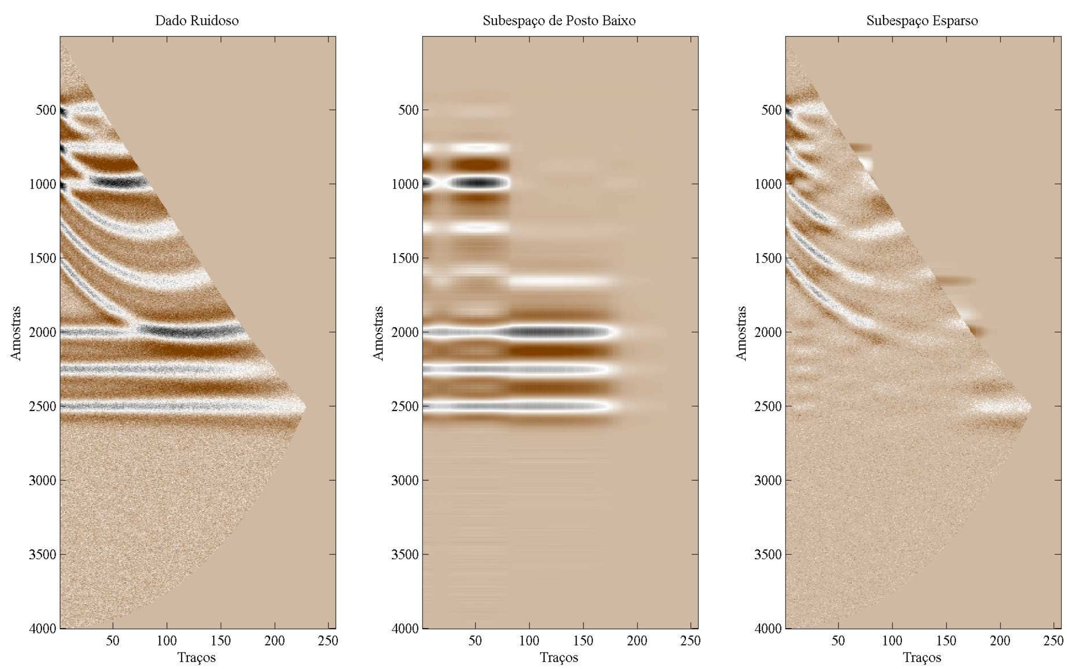

Figura 4.19: Resultado da decomposição RPCA para o dado corrigido via NMO.

prosseguir com esta etapa que era evidentemente natural.

Foram feitos alguns ajustes nos algoritmos de decomposição de forma a permitir a análise de um dado de dimensões e complexidade maiores. A aplicação da decomposição SVD local

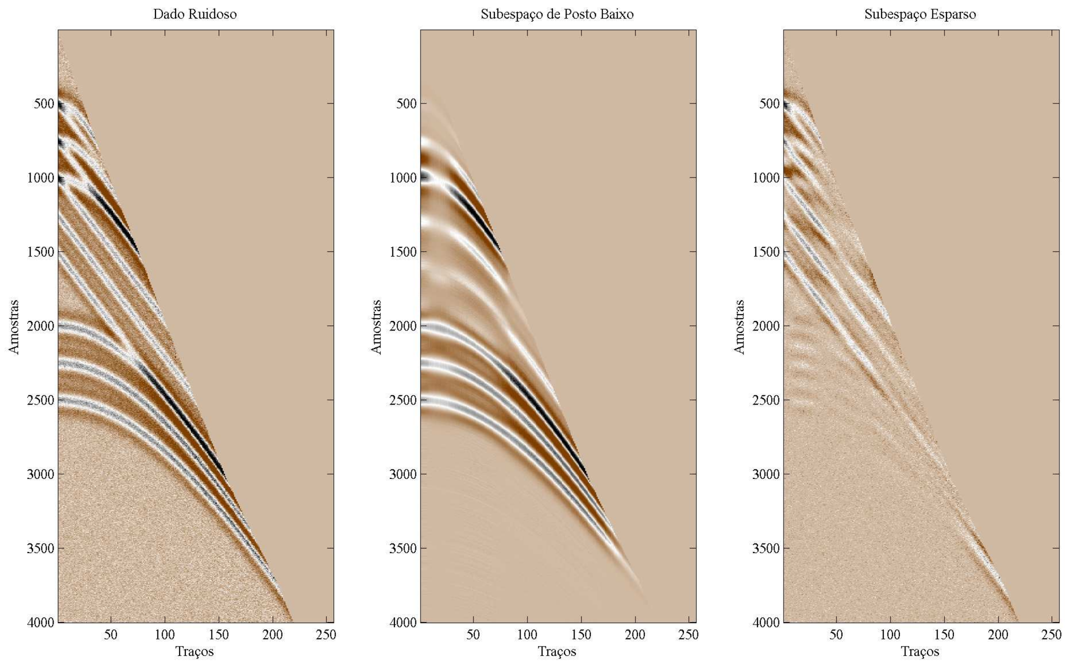

Figura 4.20: Resultado final para a decomposição RPCA após a NMO inversa.

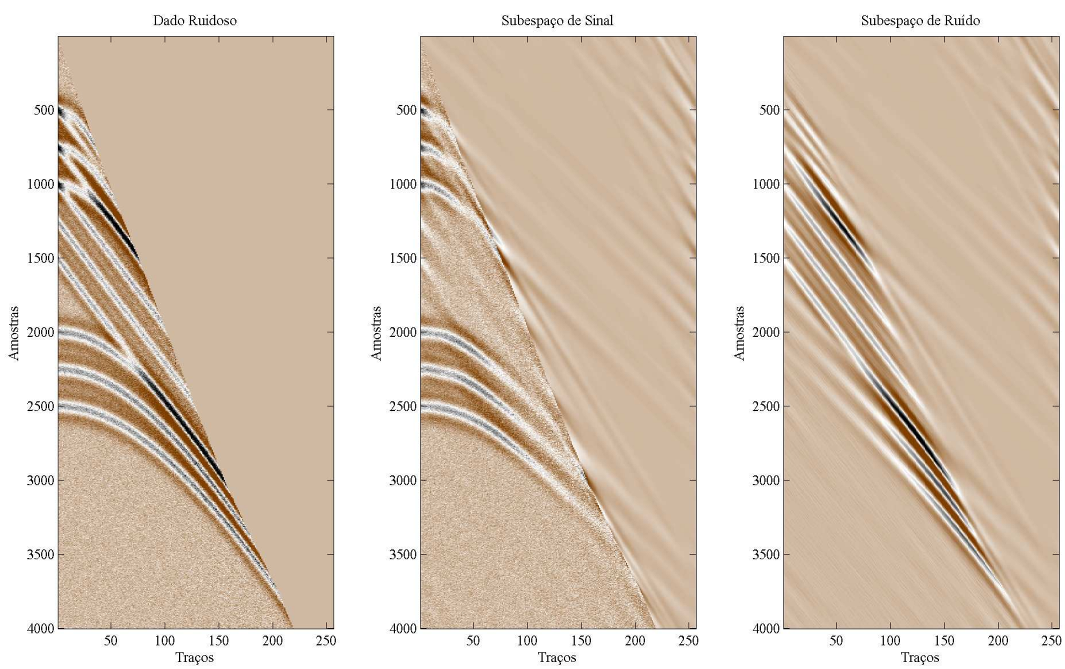

Figura 4.21: Resultado final para a filtragem *f-k* após a NMO inversa.

foi proposta em (Bekara & Van der Baan 2007) com o objetivo de elevar a relação sinal-ruído do dado em estudo. Já em (Porsani et al. 2010) a aplicação da SVD local por meio de uma janela fixa deslizante foi utilizada com o objetivo de eliminar o *ground roll*. Foi justamente este trabalho que nos motivou e nele foi baseada nossa aplicação da RPCA local para eliminação de *ground roll*.

Esta janela, que se desloca um traço a cada iteração, possui dimensão $2n+1$ e é neste intervalo de traços onde ocorre o cálculo local da SVD. Conforme a janela se movimenta, o cálculo é sobreposto sobre os $2n$ traços anteriores e assim prossegue até que a janela alcance o último traço. Dessa forma, o cálculo da SVD local pode ser descrito como (Golub & Van Loan 1996, Porsani et al. 2010):

$$\mathbf{X} = \sum_{k=1}^{2n+1} \sigma_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}, \tag{4.1}$$

compartilhando as mesmas características da SVD descritas no capítulo 2. De forma análoga, aplicamos esta técnica de janelamento e cálculo local também para a RPCA. A análise não foi feita para a SVD-ICA por esta decomposição ter apresentado os piores resultados processando dados sintéticos mais complexos.

Excepcionalmente, no caso da decomposição RPCA que usa o algoritmo SSGoDec, devido às projeções aleatórias bilaterais (em 3.13) não é possível repetir exatamente o mesmo resultado de uma execução do algoritmo duas vezes. De forma a contornar esta condição, que pode ser indesejável em uma aplicação que requer certa precisão ou que necessite futuras revisões ou reprocessamentos, nós utilizamos o método de Monte Carlo (Weisstein 2013) de forma a aproximar o resultado das simulações a um padrão determinístico. Para tanto, foi calculada a média de cem simulações.

O dado escolhido corresponde a uma prospecção em 2D feita na costa norte do Alasca, mais precisamente na Reserva Nacional de Petróleo (*National Petroleum Reserve* - NPRA), entre os anos de 1974 e 1981. Foi selecionada para este estudo a linha 31-81, por ser curta e por ter sido coletada e processada na última temporada (Schleicher 2012). Algumas características importantes para o processamento: 96 traços, multiplicidade (*fold*) 12, 56 tiros (100 – 155) e 535 CDPs (101 – 636).

Primeiramente, foi realizado o que chamamos de *pré-processamento* do dado (Yilmaz 2001), a etapa anterior à aplicação dos métodos de decomposição. Este pré-processamento, por sua vez, possui diversas etapas: primeiramente o dado é ordenado por tiros (*shots*) e recebe os ajustes de geometria, correção de divergência esférica e *muting* eliminando traços e setores extremamente corrompidos; em seguida, ele é novamente reordenado, desta vez por CDPs para em seguida ser aplicada a correção NMO. O dado então foi novamente reordenado por *shots* e recebe uma correção de amplitude por controle automático de ganho (*Automatic Gain Control* - AGC) de 25%.

Poderíamos aplicar as decomposições após a correção NMO e a correção de amplitude, antes do reordenamento por *shots* pois os dados ficam melhor alinhados após a correção NMO, no domínio do CDP. Porém, como o dado utilizado é antigo e o *fold* de apenas 12 (e pode ser considerado pequeno para o valor de $n = 2$ utilizado), optamos por aplicar as decomposições no domínio dos *shots*. Isto também foi possível dado o fato do dado em estudo apresentar os eventos muito bem alinhados neste mesmo domínio.

Esclarecidos os procedimentos utilizados em nosso pré-processamento, vamos à apresentação dos resultados com dados reais. O dado real parcialmente completo (3 de 6 segundos de aquisição), pré-processado e empilhado, está ilustrado na Figura 4.22. É possível observar uma grande quantidade de ruído, além do próprio *ground roll* que se estende aproximadamente até os 3 segundos. A porção superior esquerda é a parte mais corrompida por ruído, portanto focamos nossa análise do dado final processado neste setor, compreendido entre 0 e 3 segundos no tempo e entre os CDPs 111 e 221.

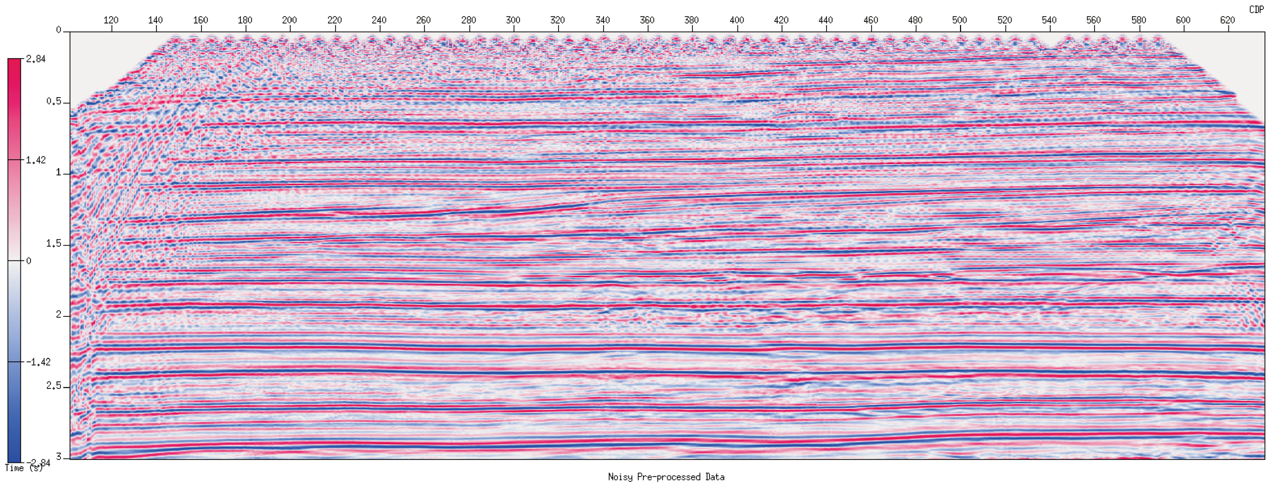

Figura 4.22: Dado real pré-processado e parcialmente completo.

Escolhemos o tiro número 150 arbitrariamente para ilustrar a atuação das decomposições, assim como foi feito anteriormente com os dados sintéticos. Assim, seguindo o mesmo procedimento da seção anterior, iniciamos pela decomposição SVD. Buscamos portanto reproduzir o procedimento em (Porsani et al. 2010) e obter assim um resultado para comparação com a RPCA mais adiante. A Figura 4.23 faz uma comparação entre o tiro 150 ruidoso e os respectivos subespaços de sinal e ruído.

Percebe-se que a decomposição SVD foi capaz de retirar grande parte do *ground roll* e de outros tipos de ruídos, assim como demonstrado em (Porsani et al. 2010). Também, é possível observar que o processo introduziu alguns artefatos e lacunas. Como explicado no Capítulo 2, estes elementos, que não existem no dado original, estão ligados à condição de ortogonalidade imposta pela SVD (Vrabie et al. 2004).

A Figura 4.24 ilustra o resultado da decomposição RPCA para o mesmo tiro. É possível observar que os resultados foram sensivelmente melhores do que os obtidos usando a SVD. As lacunas e artefatos desapareceram e foi possível recuperar alguma informação que estava encoberta pelo *ground roll*, apesar deste não ter sido removido por completo. Outra observação interessante é que o subespaço de posto baixo ficou livre de ruído descorrelacionado praticamente na mesma proporção que em seu equivalente na SVD (subespaço de sinal), porém é notória a diferença entre o subespaço esparso da RPCA e o subespaço de ruído da SVD. Isto ocorre pois o modelo de decomposição, graças ao algoritmo GoDec/SSGoDec, segue a mesma formulação em 3.10, e não como em 2.3 para o caso da SVD. A diferença entre os subespaços foi decomposta pela RPCA em um termo de ruído e separada.

A comparação com a filtragem *f-k* para o tiro 150 é exibida na Figura 4.25. O desempenho na eliminação de *ground roll* foi superior à SVD e à RPCA. Em contrapartida, também é visível que outras porções de dado foram extraídas durante a filtragem, que em parte constituem informação de interesse. Este resultado já era esperado após a análise do espectro *f-k* para determinarmos as *slopes* de filtragem. De fato, isto ocorre pois o *ground roll* está situado em uma porção ligeiramente separada dos eventos no referido espectro. As Figuras 4.26 e 4.27, quando comparadas, ilustram a situação mostrando o espectro da decomposição RPCA e o espectro da filtragem *f-k* realizadas.

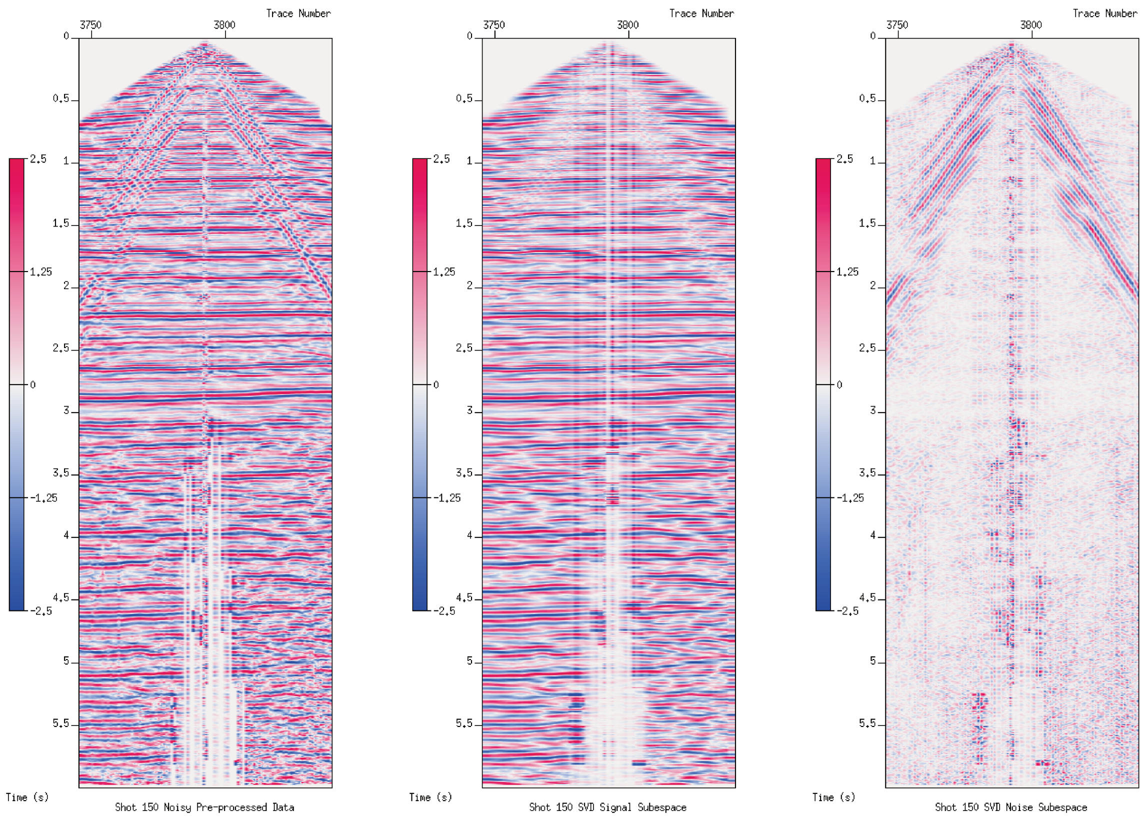

Figura 4.23: Resultado da decomposição SVD para o tiro 150.

Após a aplicação das decomposições e da filtragem, os dado é reordenado novamente por CDP para que seja feito o empilhamento. Com isso, obtemos a ilustração mais aproximada do que seria o perfil do subsolo. O restante do processamento sísmico que incluiria, por exemplo, a migração (que forneceria uma imagem em profundidade e não em tempo, como estivemos trabalhando até agora), não será executado pois nosso objetivo está restrito à extração do *ground roll* e seus efeitos diretos.

Sendo assim, após o processamento das decomposições SVD e RPCA e da filtragem *f*-*k*, podemos comparar lado a lado a região de interesse do dado empilhado na Figura 4.28. Este é o principal resultado do trabalho pois mostra o peso das vantagens e desvantagens de cada método em uma etapa avançada de todo o processamento sísmico.

No quadrante superior esquerdo da Figura 4.28, temos a seção de interesse do dado pré-processado ilustrado na Figura 4.22, ainda sem a aplicação de nenhuma técnica de eliminação de ruídos. Ao lado, temos o resultado do empilhamento do subespaço de sinal da decomposição SVD. É possível reparar que grande parte do ruído é excluída, mas em alguns locais há um tipo de *stretching* no sentido horizontal. Este efeito é devido à técnica de janelamento, que sobrepõe até cinco cálculos da decomposição, e notamos por meio da variação do comprimento da janela durante as simulações que a melhor escolha é $2n+1=5$, ou seja, $n=2$. Para $n=1$ a SVD ou a RPCA tem menos traços para diferenciar eventos de maior correlação horizontal e o resultado não é satisfatório. Para $n=3$ ou valores mais altos, este efeito se intensifica, causando excessiva deformação nos dados. Entretanto, é um efeito que se manifesta principalmente nas extremidades dos eventos (o setor em questão, por se tratar de uma extremidade do dado, contém diversas extremidades de eventos) e não afeta significativamente outras regiões do dado processado.

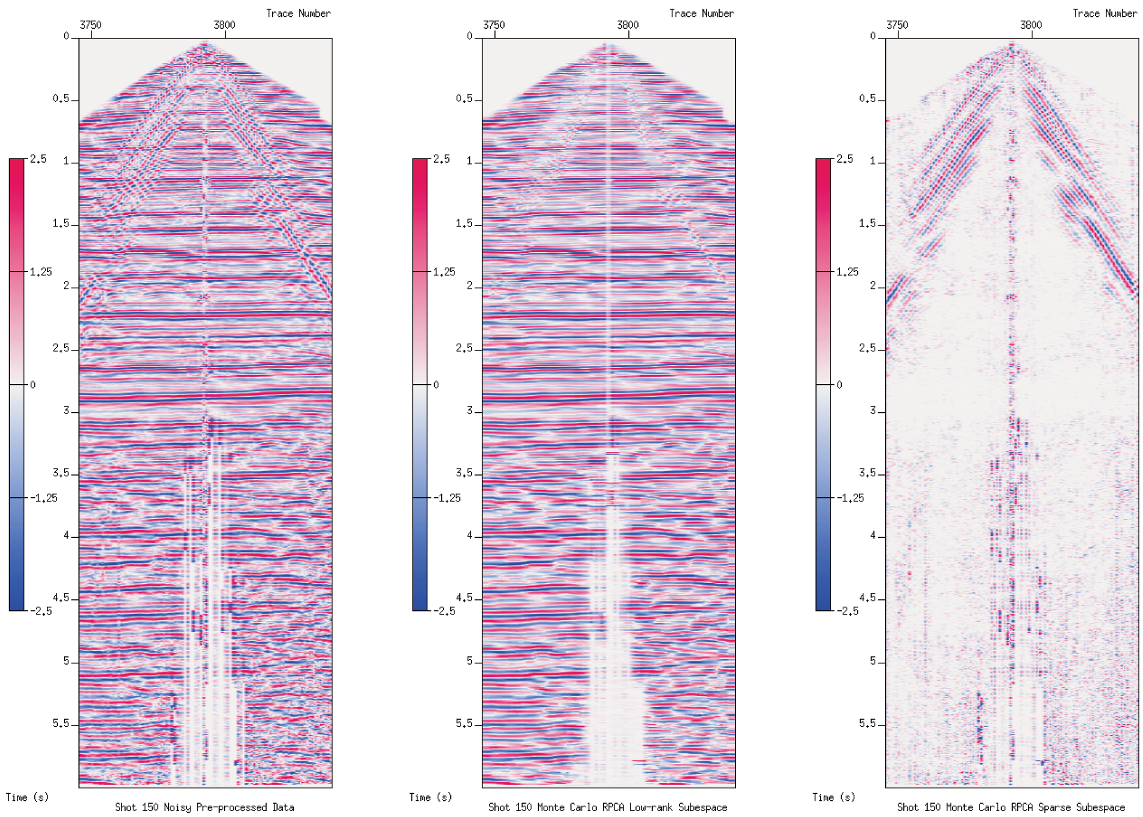

Figura 4.24: Resultado da decomposição RPCA para o tiro 150.

No canto inferior esquerdo da mesma figura, temos o resultado do empilhamento do dado processado via RPCA. O efeito de *stretching* ainda está presente, porém em proporção bem menor, o que permite observar detalhes menores obtidos a menos de 0,5 segundo. O mesmo serve para eventos entre 1 e 2 segundos, que acabam encobertos pelo efeito de *stretching* no caso da SVD. Na Figura 4.29, onde é mostrado o dado completo, é possível ver o que a RPCA é capaz de fazer em relação ao ruído descorrelacionado, além da remoção do *ground roll*.

Por fim, no canto inferior direito da Figura 4.28, temos o resultado da filtragem *f-k*. A eliminação do *ground roll* é quase total, superando os outros métodos como já foi dito em testes anteriores. Entretanto também é possivel observar a perda de informação em diversos segmentos da imagem, além de um efeito de distorção na borda esquerda do dado. Em uma comparação entre as três seções empilhadas, é a que apresenta maior quantidade de ruído descorrelacionado, já que a função da filtragem *f-k* é apenas de eliminar os dados dentro de seu envelope de filtragem, diferentemente das outras técnicas de separação de sinais utilizadas, que contribuem para uma melhora da relação sinal-ruído. Os artefatos surgidos no canto superior esquerdo do dado ilustram um pouco este cenário, já que ele é resultante de uma subtração em uma porção onde o valor da informação já era zero. Nas etapas seguintes de processamento pode ser executado um *muting* que elimina esses efeitos (dessa forma, também poderiam ser eliminados os efeitos de *stretching* nas extremidades nos casos da SVD e da RPCA).

Resta portanto, exibirmos o resultado considerando os dados empilhados para as três soluções na Figura 4.29. De cima para baixo temos o resultado do empilhamento para a SVD, para a RPCA e para a filtragem *f-k*.

Infelizmente, as dimensões no papel são pequenas para que possamos observar moelhor os detalhes, mas a primeira impressão é de que o dado processado via SVD é o menos ruidoso. Mais

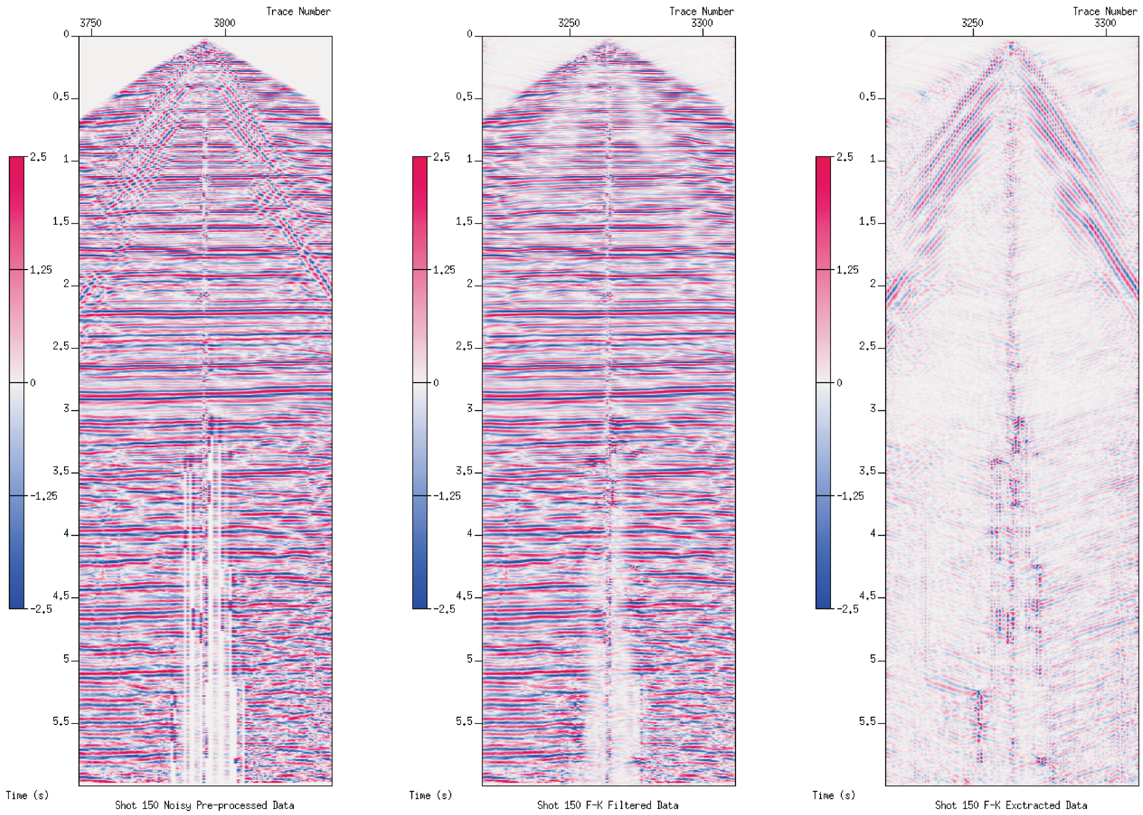

Figura 4.25: Resultado da filtragem *f-k* para o tiro 150.

de perto é possível observar que os eventos até 1 segundo são bastante afetados pelo *stretching* mencionado há pouco.

A filtragem *f-k*, ao pé da Figura 4.29 fez aquilo o que se esperava dela: eliminou a quase totalidade de ruído *ground roll*. O dado ainda permanece com o ruído não coerente, os artefatos são visíveis principalmente nas bordas superiores e informação de interesse também é perdida no processo.

Por último, vemos que a RPCA nos fornece uma imagem consideravelmente limpa de *ground roll* e sensível recuperação de informação antes eclipsada por este ruído. Os detalhes dos refletores são melhor detalhados que na SVD e a extração de ruído não coerente é melhor executada, permitindo afirmar que o resultado, em termos de imageamento, é o melhor dentre os três métodos.

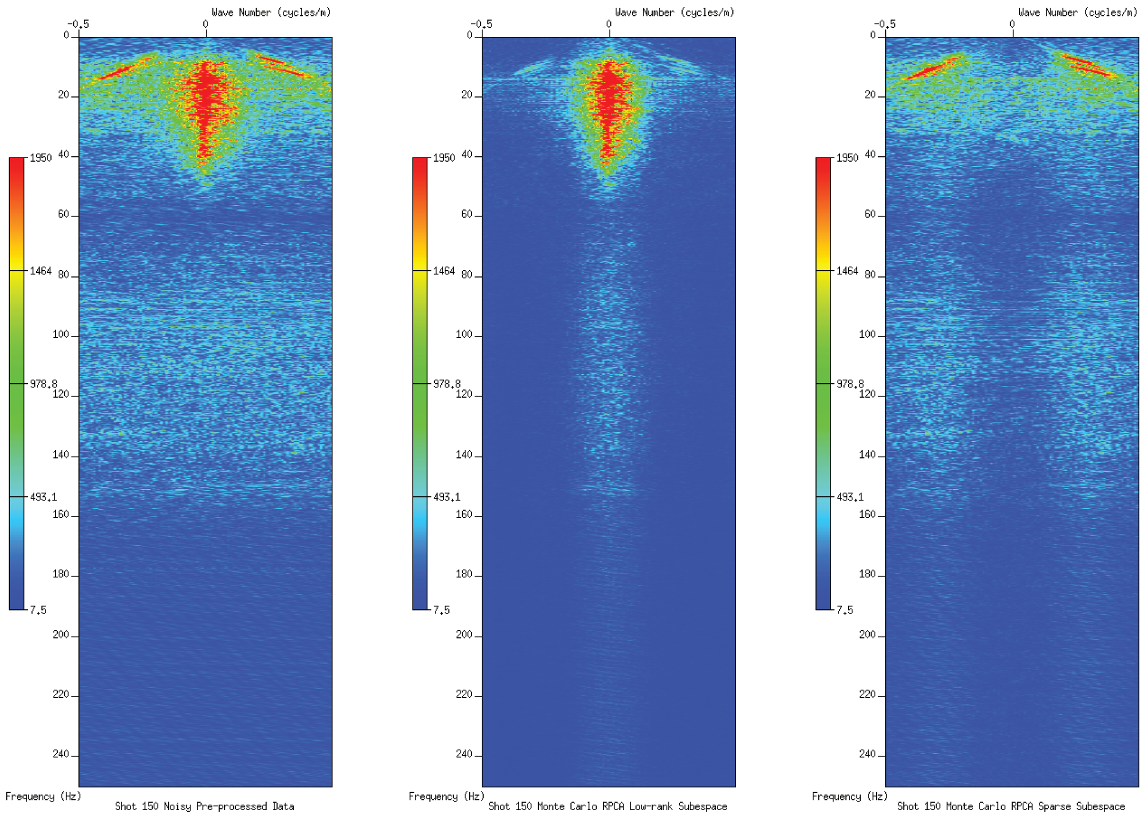

Figura 4.26: Espectro *f-k* para a decomposição RPCA para o tiro 150.

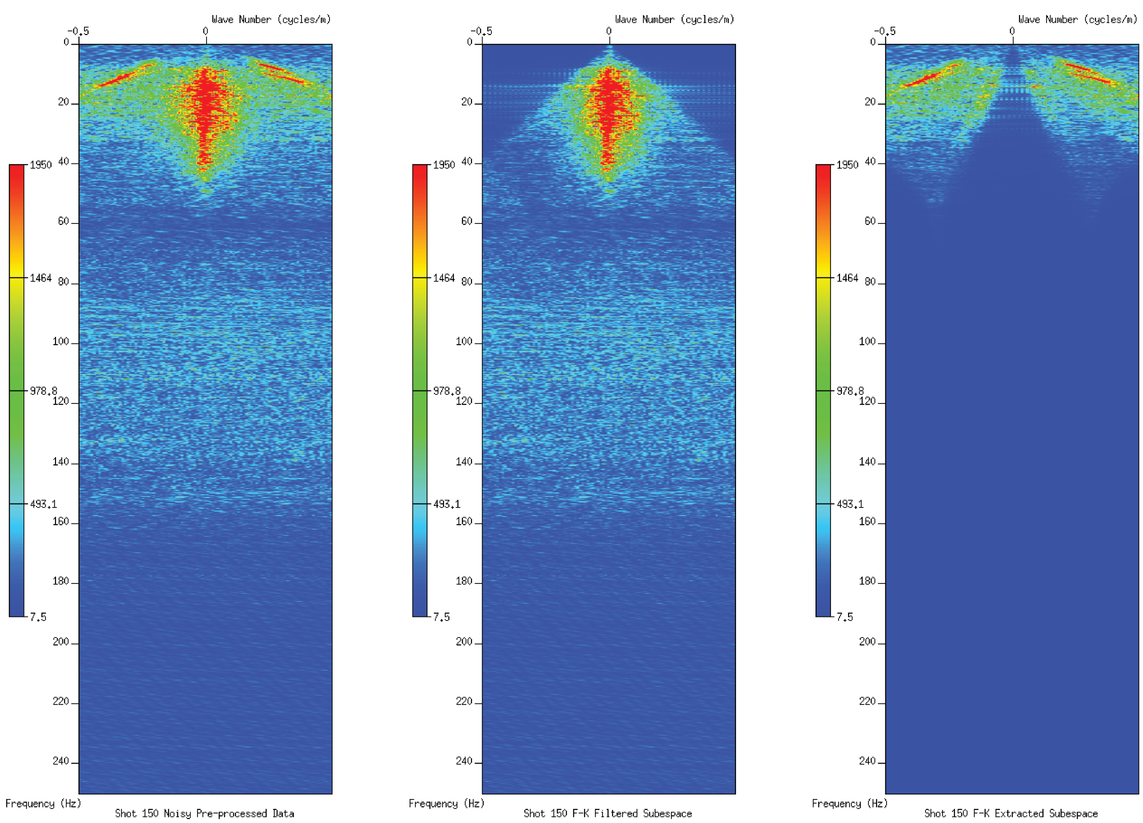

Figura 4.27: Espectro *f-k* para da filtragem realizada para o tiro 150.

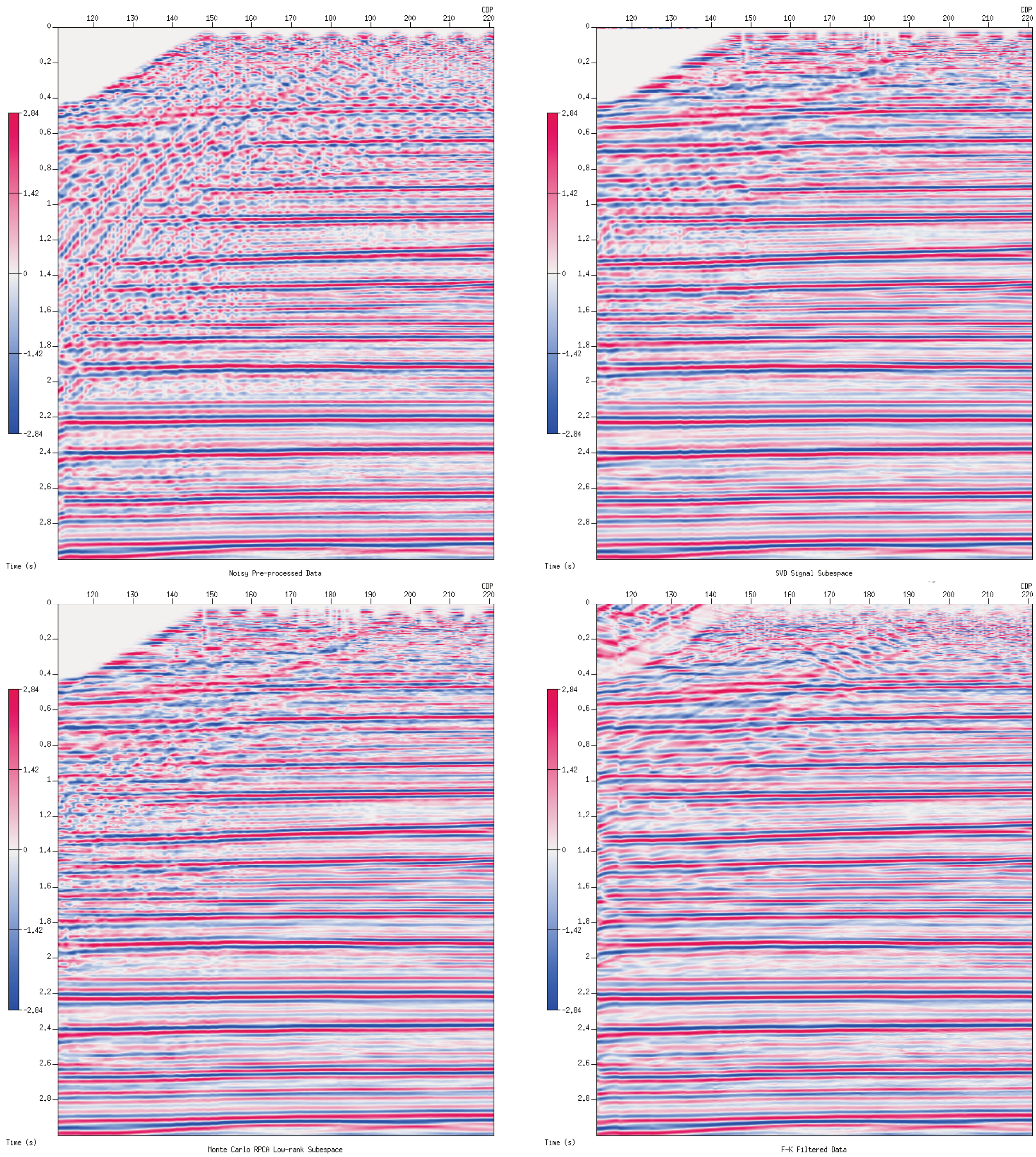

Figura 4.28: Região de interesse do dado empilhado para os quatro cenários.

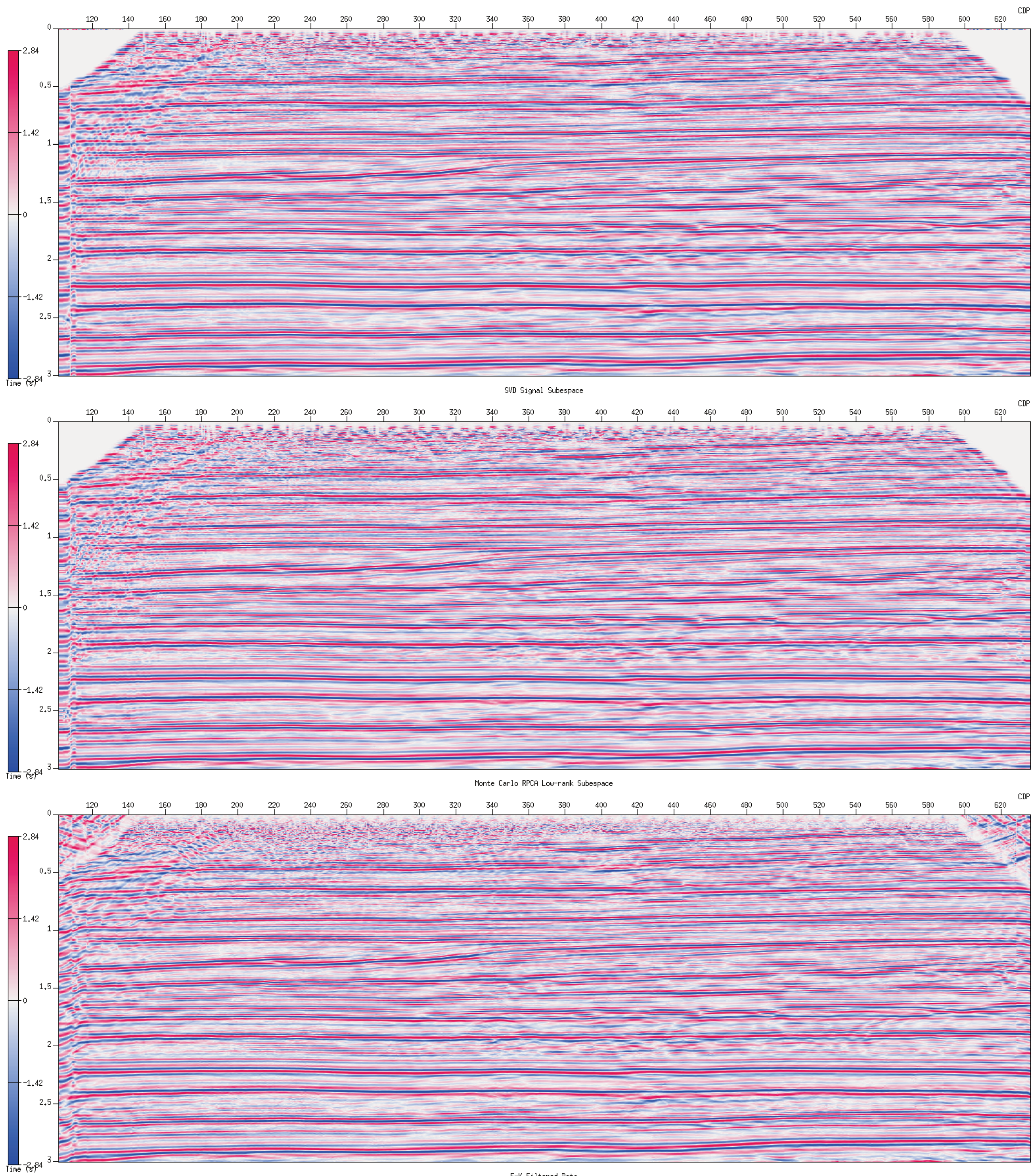

Figura 4.29: Dado completo empilhado para os três cenários de processamento.

Conclusão

Neste trabalho foi possível reproduzir os procedimentos executados em (Porsani et al. 2010) para a remoção de ruído *ground roll* via SVD, utilizar os resultados obtidos e adaptar o mesmo procedimento para a recém-introduzida técnica da RPCA e, com isso, realizar um estudo comparativo entre estes métodos e a tradicional ferramenta utilizada em processamento de dados sísmicos, a filtragem *f-k*.

Foram apresentados os resultados publicados nas conferências EAGE 2012 (Duarte et al. 2012a) e EUSIPCO 2012 (Duarte et al. 2012b), que aplicavam as técnicas de decomposição de sinais SVD, SVD-ICA e RPCA em um modelo para separação de eventos cruzados e em outro para separação de difrações de reflexões. Repetimos os mesmos parâmetros, servindo de introdução para uma nova simulação em dados sintéticos, desta vez para um cenário semelhante ao corrompido por ruído de *ground roll*, ilustrado na Figura 4.9. Além disso, conduzimos testes em uma situação desafiadora, mas dada a dificuldade em emular um cenário deste tipo, os resultados não foram os esperados.

A necessidade da correção NMO eliminou parte considerável do modelo construído e foram observadas criações de elementos não existentes no dado originalmente concebido para os diferentes métodos de processamento. Acreditamos que isto seja um indicador de que existem casos específicos onde as técnicas de decomposição utilizadas podem ser melhor aproveitadas. A análise dos perfis de valores singulares nas Figuras 4.1, 4.5 e 4.12, e o resultado das respectivas decomposições, indicam que os métodos de decomposição de sinais se adaptam melhor a situações onde há poucas autoimagens com valores singulares elevados em amplitudes muito superiores ao restante delas. Por esta razão, os métodos utilizados em (Porsani et al. 2010) e (Vrabie et al. 2004) utilizam poucas autoimagens para definirem o subespaço de interesse. Contudo, foi possível observar as propriedades e também os problemas de cada técnica utilizada, servindo assim como uma interessante introdução para a análise mais importante por vir: o processamento em dados sísmicos reais.

Desde a apresentação dos processamentos do tiro 150, foi possível observar nas Figuras 4.23, 4.24 e 4.25, que a filtragem *f-k* se mostraria a melhor ferramenta para a remoção de ruído de *ground roll* para este cenário, onde o espectro *f-k* mostra que este ruído está ligeiramente separado da maior parte da informação de interesse, evitando que a perda de informação provocada pelo envelope de filtragem fosse maior. A SVD e a RPCA tiveram desempenho semelhante neste sentido, ao passo que eliminaram grandes quantidades de ruído não coerente enquanto que isso nem poderia ser esperado da filtragem *f-k*.

Como explicamos no início da Seção 4.2, o ideal seria aplicarmos a SVD ao dado ordenado por CDPs dada a maior horizontalidade dos eventos neste domínio. Os efeitos de *stretching* seriam minimizados e seria possível eliminar uma maior porção de *ground roll*. O mesmo vale para a decomposição RPCA, que também separa melhor eventos com alta correlação traço-a-traço. Infelizmente, devido às limitações do dado real utilizado (*fold* de apenas 12) por ser muito antigo, tivemos que operar no domínio dos tiros, que felizmente apresentavam uma boa horizontalidade em seus eventos.

Outro fator que também pode influenciar os resultados das decomposições é o pré-processamento dos dados sísmicos. Antes de aplicarmos o AGC de 25%, não era possível enxergar nada além do *ground roll*, dada a enorme diferença de amplitude entre este ruído e os eventos refletidos. Neste caso, também podemos afirmar que um pré-processamento diferente talvez permitisse obter melhores resultados mesmo com este conjunto de dados.

Por fim, a análise das seções empilhadas permitiu-nos observar que as técnicas de decomposição de sinais vão além da simples extração do *ground roll*. A eliminação de ruído não coerente e a recuperação de informação antes encoberta por ruído demonstrou ser de bastante utilidade na obtenção de uma imagem mais limpa e os estudos efetuados demonstraram que ainda há outras investigações a fazer.

Perspectivas

No início dos testes, buscávamos uma maneira de horizontalizar os eventos e facilitar assim a atuação de técnicas como a SVD e a RPCA. Criamos um procedimento que, por meio de um *picking* manual sobre o *ground roll* obtinha sua inclinação e, com base nela era efetuado um *shifting* dos traços verticalmente, de forma a tornar horizontal o *ground roll* (o que inverteria a convenção adotada neste trabalho sobre subespaços de sinal e de ruído), e assim eliminar com precisão o ruído.

Entretanto, com o surgimento de artefatos quando realizávamos o *shift back*, havia o surgimento de artefatos diferentes dos que costumam surgir como resquícios de refletores ou extremidades dos mesmos. Após a análise do espectro de Fourier 2D dos dados processados, a melhor explicação para esse problema seria uma espécie de *aliasing*, provocado pelo *shifting* dos traços. Algumas sugestões para a solução do problema surgiram, como por exemplo a aplicação de *oversampling*. Por acharmos que a complexidade do problema poderia ser muito maior, deixamos de lado esta linha de análise. Porém, se descoberta uma solução em trabalhos futuros, é possível contornar os problemas causados pela correção NMO, evitando a perda de informações e a criação de artefatos por técnicas de decomposição de sinais.

Uma outra ideia é aplicar a RPCA na extração de múltiplas de reflexão, dada a possibilidade de horizontalizar estes elementos preservando os reais refletores dos dados.

Visto que as técnicas de decomposição de sinais apresentaram um bom desempenho na eliminação de ruído não coerente em seções cujos eventos apresentam alto grau de correlação traço-a-traço, ao invés de tentarmos substituir a filtragem *f-k*, em dados cujo espectro *f-k* do *ground roll* não apresenta sobreposição a informações de interesse, seria interessante combinar os métodos, aproveitando assim as qualidades de cada um e minimizando suas falhas.

Enfim, podemos concluir que ainda há várias aplicações onde os métodos de decomposição de sinais podem ser aplicadas em processamento de dados sísmicos.

Bibliografia

Basri, R. & Jacobs, D. (2003). Lambertian reflectance and linear subspaces, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **25**(2): 218–233.

Bekara, M. & Van der Baan, M. (2007). Local singular value decomposition for signal enhancement of seismic data, *GEOPHYSICS* **72**(2): V59–V65.

Candès, E. J., Li, X., Ma, Y. & Wright, J. (2011). Robust principal component analysis?, *Journal of the ACM* **58**(3): 1–37.

Candès, E. J. & Recht, B. (2009). Exact matrix completion via convex optimization, *Foundations of Computational Mathematics* **9**(6): 717–772.

Candes, E. & Plan, Y. (2010). Matrix completion with noise, *Proceedings of the IEEE* **98**(6): 925–936.

Candès, E. & Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion, *Information Theory, IEEE Transactions on* **56**(5): 2053–2080.

Cardoso, J. F. & Souloumiac, A. (1993). Blind beamforming for non-gaussian signals, *IEE Proceedings F (Radar and Signal Processing)* **140**: 362–370.

Carin, L. (2013). Bayesian Compressive Sensing.
URL: `http://people.ee.duke.edu/~lcarin/BCS.html`

Center for Wave Phenomena (2012). CWP/SU: Seismic Un*x.
URL: `http://www.cwp.mines.edu/cwpcodes`

Comon, P. (1994). Independent component analysis, a new concept ?, *Signal Processing* **36**(3): 287–314.

Comon, P. & Jutten, C. (2010). *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, 1st edn, Academic Press.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R. (1990). Indexing by latent semantic analysis, *Journal of the American Society for Information Science* **41**: 391–407.

Diamantaras, K. I. & Kung, S. Y. (1996). *Principal Component Neural Networks: Theory and Applications*, Wiley-Interscience.

Duarte, L., Nadalin, E., Filho, K., Zanetti, R. A., Romano, J. M. T. & Tygel, M. (2012b). Seismic wave separation by means of robust principal component analysis, *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pp. 1494–1498.

Duarte, L. T., Nadalin, E. Z., Nose Filho, K., Zanetti, R. A., Romano, J. M. T. & Tygel, M. (2012a). Application of robust principal component analysis to seismic data processing, *74th EAGE Conference & Exhibition*.

Eckart, C. & Young, G. (1936). The approximation of one matrix by another of lower rank, *Psychometrika* **1**(3): 211–218.

Glangeaud, F. & Mari, J. L. (1994). *Wave Separation*, Editions Technip.

Golub, G. H. & Van Loan, C. F. (1996). *Matrix Computations*, 3rd edn, Johns Hopkins University Press.

Gross, D. (2011). Recovering low-rank matrices from few coefficients in any basis, *Information Theory, IEEE Transactions on* **57**(3): 1548–1566.

Gross, D., Liu, Y.-K., Flammia, S. T., Becker, S. & Eisert, J. (2010). Quantum state tomography via compressed sensing, *Physical Review Letters* **105**: 150401.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology* **24**(6): 417–441.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis, *Neural Networks, IEEE Transactions on* **10**(3): 626–634.

Hyvärinen, A., Karhunen, J. & Oja, E. (2001). *Independent Component Analysis*, Wiley-Interscience.

Jolliffe, I. T. (1986). *Principal Component Analysis*, Springer.

Kearey, P., Brooks, M. & Hill, I. (2002). *An Introduction to Geophysical Exploration*, 3rd edn, Blackwell Science.

Keshavan, R., Montanari, A. & Oh, S. (2010). Matrix completion from a few entries, *Information Theory, IEEE Transactions on* **56**(6): 2980–2998.

Kirlin, R. L. & Done, W. J. (1999). *Covariance Analysis for Seismic Signal Processing*, SEG Books.

Netflix Inc. (2009). The Netflix Prize.
URL: `http://www.netflixprize.com`

Papadimitriou, C. H., Tamaki, H., Raghavan, P. & Vempala, S. (1998). Latent semantic indexing: A probabilistic analysis, *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, PODS '98, ACM, New York, NY, USA, pp. 159–168.

Perception and Decision Lab (2012). Low-rank matrix recovery and completion via convex optimization.
URL: `http://perception.csl.illinois.edu/matrix-rank/sample_code.html`

Porsani, M. J., Silva, Silva, M. G., Melo, P. E. M. & Ursin, B. r. (2010). Svd filtering applied to ground-roll attenuation, *Journal of Geophysics and Engineering* **7**(3): 284–289.

Reynolds, J. M. (1997). *An Introduction to Applied and Enviromental Geophysics*, Wiley.

Roweis, S. (1998). Em algorithms for pca and spca, *Advances in Neural Information Processing Systems* pp. 626–632.

Schleicher, K. (2012). Open dataopen source: Seismic unix scripts to process a 2d land line.

Stone, J. V. (2004). *Independent Component Analysis: A tutorial Introduction*, A Bradford Book.

The MathWorks, Inc. (2012). MATLAB The Language of Technical Computing.
URL: `http://www.mathworks.com/products/matlab`

Van der Kruk, J. (2003). *Reflection Seismic 1*, ETH Zürich, Institut für Geophysik.

Vrabie, V. D., Mars, J. I. & Lacoume, J.-L. (2004). Modified singular value decomposition by means of independent component analysis, *Signal Processing* **84**(3): 645–652.

Weisstein, E. W. (2013). Monte Carlo Method — MathWorld – A Wolfram Web Resource.
URL: `http://mathworld.wolfram.com/MonteCarloMethod.html`

Yilmaz, Özdoğan. (2001). *Seismic Data Analysis*, 2nd edn, Society of Exploration Geophysicists.

Zhou, T. (2013). Go Decomposition.
URL: `https://sites.google.com/site/godecomposition`

Zhou, T. & Tao, D. (2011). Godec: Randomized low-rank & sparse matrix decomposition in noisy case, *in* L. Getoor & T. Scheffer (eds), *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, ICML '11, ACM, New York, NY, USA, pp. 33–40.

Zhou, T. & Tao, D. (2013). Shifted subspaces tracking on sparse outlier for motion segmentation, *23rd International Joint Conference on Artificial Intelligence - IJCAI 2013*. (Aceito para apresentação).