



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"

Campus Presidente Prudente

LARISSA VITÓRIA RIBEIRO DE ANDRADE

## **EIGENFACE, EIGENBEEF PARTE 2**

PRESIDENTE PRUDENTE

2025

**LARISSA VITÓRIA RIBEIRO DE ANDRADE**

eigenface, eigenbeef parte 2

Tarefa desenvolvido na disciplina de Álgebra Linear para Ciências de Dados no Programa de Pós-Graduação em Matemática Aplicada e Computacional da Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP-FCT.

Professor: Cássio Machiaveli Oishi

**PRESIDENTE PRUDENTE**

2025

# Sumário

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Metodologia</b>	<b>2</b>
2.1	Pré-processamento dos Dados . . . . .	2
2.2	Decomposição RPCA . . . . .	3
2.3	Análise dos Componentes . . . . .	3
2.4	Detecção de Anomalias . . . . .	3
2.5	Ferramentas Utilizadas . . . . .	3
<b>3</b>	<b>Resultados</b>	<b>3</b>
3.1	Exercício . . . . .	3
3.2	Resolução . . . . .	4
3.2.1	Caso das imagens de peças de corte bovino . . . . .	4

# 1 Introdução

A análise de componentes principais (PCA) é uma técnica fundamental em álgebra linear e ciência de dados, amplamente utilizada para redução de dimensionalidade, compressão de dados e reconhecimento de padrões. No entanto, sua sensibilidade a *outliers* e dados corrompidos limita sua aplicabilidade em cenários do mundo real, onde ruídos e anomalias são frequentes. Para superar essa limitação, Candès et al. (2011) propuseram a **Análise de Componentes Principais Robusta (RPCA)**, que decompõe uma matriz de dados  $\mathbf{X}$  em dois componentes: uma matriz de baixo posto  $\mathbf{L}$ , que contém a estrutura principal dos dados, e uma matriz esparsa  $\mathbf{S}$ , que captura outliers, ruídos e anomalias:

$$\mathbf{X} = \mathbf{L} + \mathbf{S}.$$

Esta decomposição tem sido aplicada com sucesso em diversas áreas, como processamento de vídeo (separação entre fundo e movimento), reconhecimento facial (extração de eigenfaces livres de oclusões) e, mais recentemente, em análise de imagens biomédicas e de alimentos. No contexto de inspeção de qualidade de carnes, a RPCA pode ser utilizada para identificar regiões anômalas, como manchas, texturas irregulares ou sinais de deterioração, de forma automática e robusta.

Este trabalho tem como objetivo principal reproduzir e adaptar a metodologia de RPCA para o conjunto de imagens de *eigenfaces* (base de dados Yale B) e, em seguida, aplicá-la a um conjunto de imagens de cortes bovinos (*eigenbeef*), a fim de detectar e analisar anomalias visuais associadas à qualidade da carne.

## 2 Metodologia

A implementação da RPCA foi baseada no algoritmo de *Principal Component Pursuit (PCP)* via *Alternating Directions Method (ADM)*, conforme descrito por Candès et al. e reproduzido nos códigos em MATLAB. A metodologia adotada segue as etapas abaixo:

### 2.1 Pré-processamento dos Dados

- **Eigenfaces:** Utilizou-se a base de dados Yale B, composta por imagens de rostos humanos com variações de iluminação e expressões. As imagens foram convertidas para escala de cinza, redimensionadas e organizadas em uma matriz  $\mathbf{X}$ , onde cada coluna representa uma imagem vetorizada.
- **Eigenbeef:** Um conjunto de imagens de cortes bovinos foi preparado, contendo amostras das classes *fresh* (fresca) e *rotten* (deteriorada). As imagens foram padronizadas em tamanho e intensidade, e posteriormente dispostas em uma matriz de dados de forma análoga ao caso das faces.

## 2.2 Decomposição RPCA

Para cada conjunto de dados, aplicou-se o algoritmo de RPCA implementado em MATLAB (Código 3.5), que resolve o problema convexo:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{sujeito a} \quad \mathbf{L} + \mathbf{S} = \mathbf{X},$$

onde  $\|\cdot\|_*$  é a norma nuclear (soma dos valores singulares) e  $\lambda = 1/\sqrt{\max(n, m)}$ , com  $n$  e  $m$  sendo as dimensões de  $\mathbf{X}$ .

## 2.3 Análise dos Componentes

- **Componente de Baixo Posto (L):** Representa a estrutura global e redundante das imagens. Foi analisado por meio da decomposição em valores singulares (SVD) e do gráfico de energia acumulada.
- **Componente Esparso (S):** Contém variações locais, ruídos e possíveis anomalias. A norma  $\|\mathbf{S}\|_2$  foi utilizada como métrica de anomalia por imagem.

## 2.4 Detecção de Anomalias

Para o caso das imagens de carne, as normas de  $\mathbf{S}$  foram comparadas entre as classes *fresh* e *rotten* por meio de histogramas e boxplots. Imagens com maiores valores de norma foram identificadas como as mais anômalas, e seus mapas de anomalias foram visualizados para interpretação qualitativa.

## 2.5 Ferramentas Utilizadas

- MATLAB R2023a para implementação do algoritmo RPCA.
- Funções customizadas *shrink* e *SVT* para operadores de limiaramento.
- Visualização dos resultados com gráficos de valores singulares, energia acumulada e distribuição de normas.

# 3 Resultados

## 3.1 Exercício

Estude o RPCA para o dataset de imagens de faces (eigenfaces), como no exemplo dado pelo Fig 3.20 do livro. Depois disso, repita o processo para o seu estudo no caso das imagens de peças de corte bovino.

## 3.2 Resolução

### 3.2.1 Caso das imagens de peças de corte bovino

Nossa reprodução foi fiel a seção do livro onde retrata os eigenface, implementados no MATLAB, aqui estão os resultados e análises após a implementação.

#### Decomposição RPCA das Imagens

A Figura 1 apresenta a decomposição das imagens de carne obtida por meio da RPCA, mostrando, para cada amostra selecionada, a imagem original ( $X$ ), o componente de baixo posto ( $L$ ) e o componente esparsos ( $S$ ).

Observa-se que o componente  $L$  preserva a estrutura global das imagens, como o formato geral da carne e padrões de intensidade de baixa frequência, enquanto variações locais e ruídos são significativamente atenuados. Esse comportamento confirma que a maior parte da informação visual relevante pode ser representada em um subespaço de baixa dimensionalidade.

Por outro lado, o componente esparsos  $S$  evidencia regiões pontuais de alta variabilidade, destacando texturas irregulares, sombras, reflexos e possíveis sinais de deterioração. A separação clara entre estrutura e anomalias indica que a RPCA foi eficaz na modelagem robusta das imagens, mesmo na presença de perturbações locais.

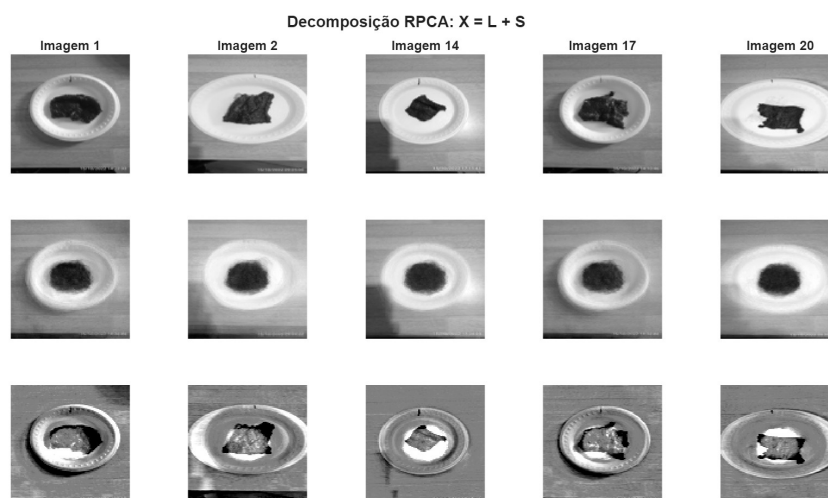


Figura 1: Decomposição RPCA das imagens de carne: imagem original ( $X$ ), componente de baixo posto ( $L$ ) e componente esparsos ( $S$ ).

#### Análise Espectral e Energia Acumulada

A Figura 2 mostra a análise espectral comparativa entre a matriz de dados original  $X$  e o componente de baixo posto  $L$ . No gráfico de valores singulares normalizados, observa-se um decaimento mais acentuado para  $L$ , caracterizando sua natureza de baixo posto.

Esse comportamento indica que a RPCA removeu componentes associados a ruído e variações esparsas, concentrando a energia do sinal em um número reduzido de modos principais. O

gráfico de energia acumulada reforça essa interpretação, mostrando que aproximadamente 95% da energia de  $L$  é explicada por um número significativamente menor de componentes quando comparado à matriz original.

Esse resultado justifica o uso do componente  $L$  como uma representação compacta e robusta das imagens para tarefas subsequentes, como redução de dimensionalidade e classificação.

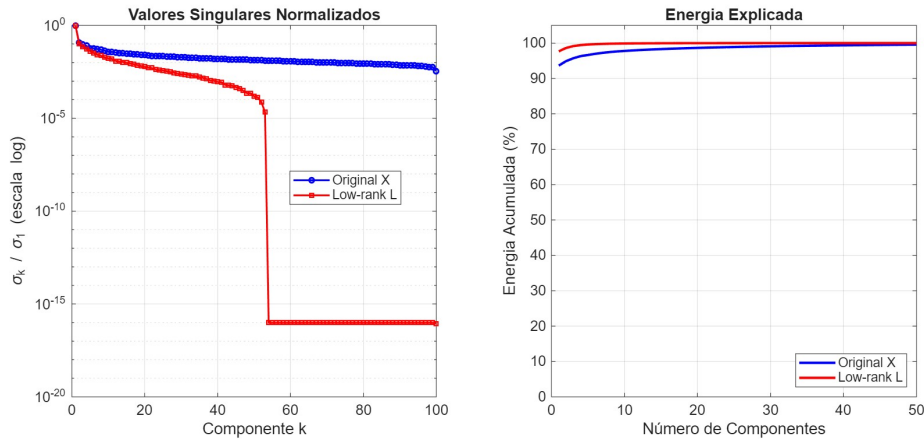


Figura 2: Análise espectral comparativa entre a matriz original  $X$  e o componente de baixo posto  $L$ .

### Distribuição do Componente Esparso e Detecção de Anomalias

A Figura 3 apresenta a distribuição das normas do componente esparso  $S$  para as classes *fresh* e *rotten*. A norma do componente esparso foi utilizada como uma medida quantitativa de anomalia por imagem.

Nota-se que as imagens da classe *rotten* apresentam, em média, valores mais elevados da norma de  $S$ , indicando maior presença de variações locais e irregularidades estruturais. Esse comportamento sugere que carnes deterioradas tendem a possuir padrões visuais mais complexos e menos homogêneos do que carnes frescas.

Além disso, a separação entre as distribuições, evidenciada tanto pelo histograma quanto pelo boxplot, indica que o componente esparso contém informação discriminativa relevante para a diferenciação entre as classes.

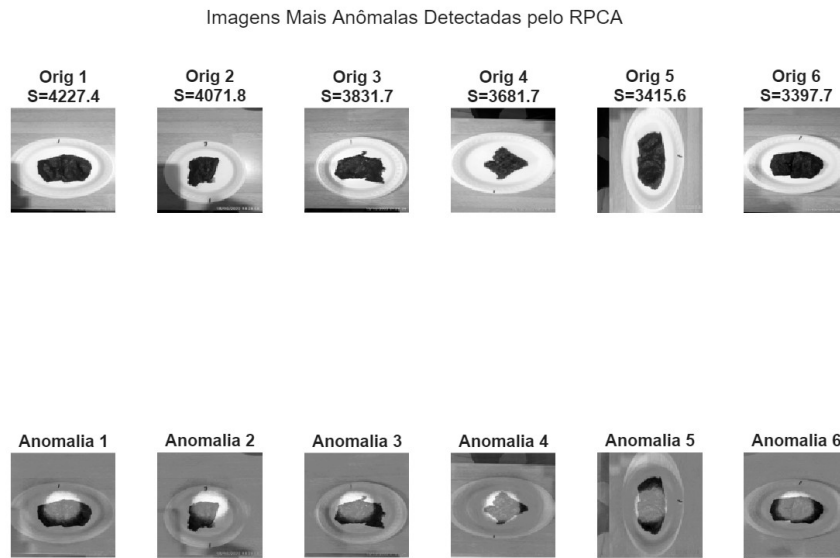


Figura 3: Distribuição das normas do componente esparsos  $S$  para imagens *fresh* e *rotten*.

### Imagens Mais Anômalas Identificadas pela RPCA

A Figura 4 apresenta as imagens identificadas como mais anômalas com base na norma do componente esparsos  $S$ . Para cada caso, são mostradas a imagem original e o respectivo mapa de anomalias.

Observa-se que o componente  $S$  destaca regiões específicas da imagem associadas a manchas, descontinuidades de textura e alterações abruptas de intensidade. Essas regiões são pouco representadas no componente de baixo posto e, portanto, interpretadas como desvios em relação ao padrão estrutural dominante.

Esse resultado demonstra a capacidade da RPCA de atuar como uma ferramenta eficaz para detecção automática de regiões suspeitas, oferecendo suporte visual e quantitativo para aplicações de inspeção e controle de qualidade.



Figura 4: Imagens mais anômalas detectadas pela RPCA: comparação entre imagem original e componente esparsos  $S$ .