

# NMF

## Introdução

A Non-negative matrix factorization (NMF) é um grupo de algoritmos utilizado para a análise de problemas multivariável e de álgebra linear, onde uma matriz  $V$  é fatorada em duas matrizes  $W$  e  $H$  que possuem a propriedade de serem três matrizes sem elementos negativos. No entanto, essa fatoração não possui solução exata, sendo necessária uma aproximação numérica.

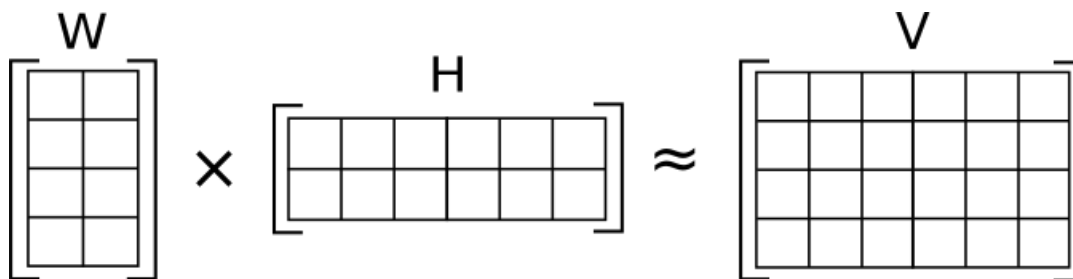


Figura 1: Ilustração da aproximação da matriz  $V$  representada por duas matrizes menores  $W$  e  $H$ , usando non-negative matrix factorization (Wikipedia).

No âmbito de dimensões das matriciais, considere um conjunto multivariável de  $m$  vetores de dimensão  $n$  e suponha que esses vetores sejam colocados como colunas de uma matriz  $V$ , onde  $m$  é o número de exemplos de uma base de dados. Essa matriz pode então ser aproximadamente fatorada pelo NMF em uma matriz  $W$   $n \times r$  e uma matriz  $H$   $r \times m$ . Geralmente, o valor de  $r$  é escolhido pelo usuário, sendo frequentemente menor que  $n$  ou  $m$ , de forma que  $W$  e  $H$  sejam menores que  $V$ , resultando numa versão comprimida da matriz original.

A partir disso, cada coluna da matriz  $V$  pode ser escrita da seguinte forma:  $v \approx Wh$ , onde  $v$  e  $h$  são as colunas correspondentes de  $V$  e  $H$ . Em outras palavras, cada vetor  $v$  pode ser aproximado por uma combinação linear das colunas de  $W$  ponderadas pelos componentes de  $h$ . Dessa forma,  $W$  pode ser considerada uma matriz que contém uma base que é otimizada por uma aproximação linear dos dados de  $V$ .

## Aplicações

O principal diferencial dessa fatoração é a sua positividade, pois as matrizes resultantes positivas são mais fáceis de interpretar e gerar análises do que matrizes com valores negativos. Dessa forma, esse algoritmo é utilizado em diversas aplicações que necessitam dessa propriedade, como problemas de astronomia, visão computacional, clusterização de documentos, processamento de sinais de áudio, sistemas de recomendação e bioinformática.

Para exemplificar sua utilidade, considere o problema em que temos  $m$  imagens vetorizadas de  $n$  pixels com valores não negativos — apenas cores entre preto e branco —, sendo cada imagem a representação de um rosto de uma pessoa. Caso esses  $m$  vetores sejam colocados numa matriz  $V$ , essa matriz pode ser aproximada em  $W$  e  $H$ , tal que  $V \approx WH$ , de modo que cada rosto de  $V$  é

representada pela combinação de múltiplas imagens bases em  $W$  e a combinação dessas imagens é feita utilizando apenas adições descritas em  $H$ , pois todas as matrizes possuem apenas elementos positivos. Nesse sentido, o NMF oferece uma noção mais intuitiva de combinação de partes — de características (nariz, olho, boca, etc) — que formam um “todo”.

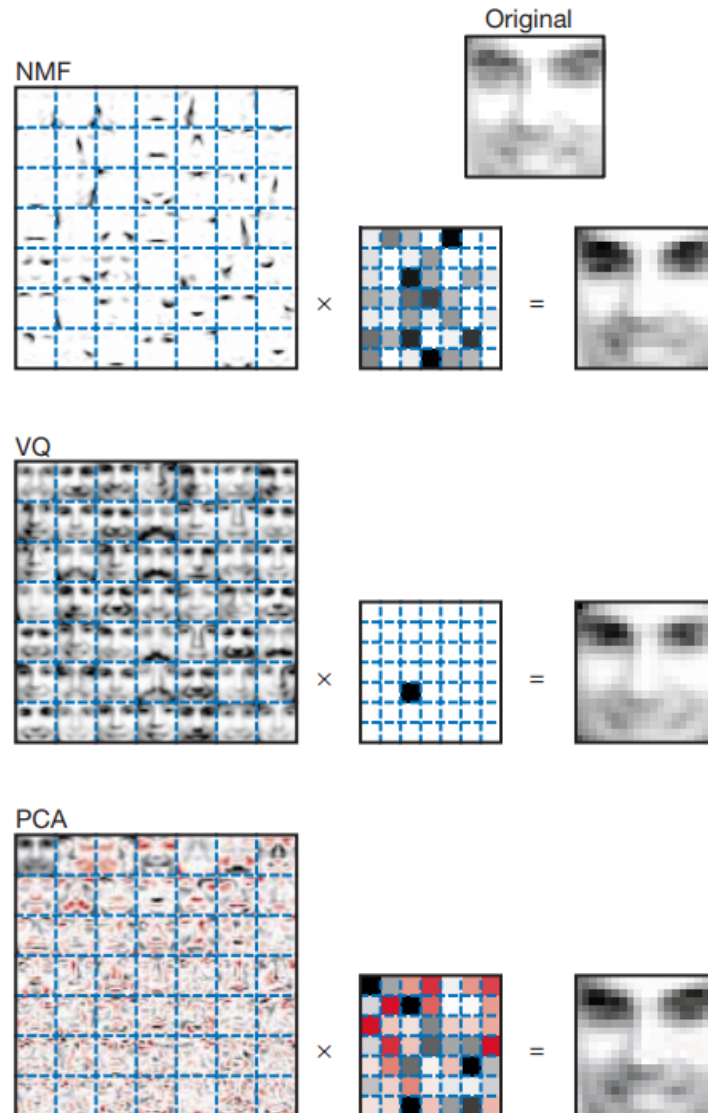


Figura 2: Representação de faces usando os algoritmos de aprendizado non-negative matrix factorization (NMF), principal component analysis (PCA) e vector quantization (VQ). A base de dados utilizada tinha 2429 imagens de 19 X 19 pixels e foram fatoradas em 49 imagens base (Nature 1999).

Por outro lado, se for utilizado o SVD nessa mesma base de dados, cada imagem da matriz  $V$  será aproximada por combinações lineares que envolvem junções complexas entre números positivos e negativos. Diante disso, torna-se difícil extrair qualquer significado intuitivo dos rostos dos indivíduos representados na base de dados.

Outro exemplo interessante de analisar consiste na análise de documentos textuais, onde cada documento é representado por um vetor de frequências de cada palavra que aparece no texto. Nesse contexto, com NMF é possível extrair características semânticas de cada documento, por meio da observação de padrões de combinação das bases de  $W$  fornecidas por  $H$ , revelando documentos com assuntos e estruturas semelhantes.

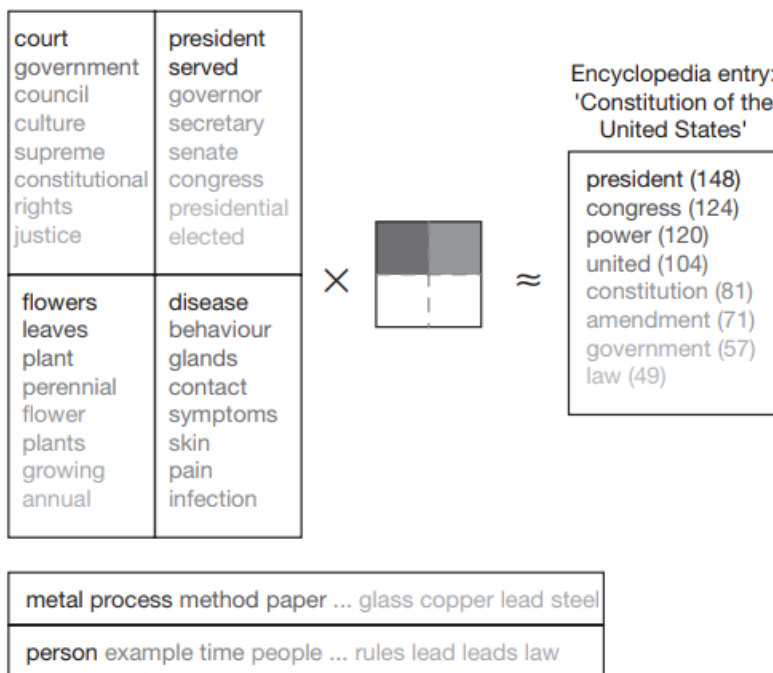


Figura 3: Utilização do non-negative matrix factorization (NMF) para descobrir características semânticas de 30991 artigos da enciclopédia Grolier, considerando um vocabulário de 15276 palavras para cada documento. Foram extraídas 200 características semânticas desses documentos, sendo exibidas no canto esquerdo superior algumas delas.

Por fim, como podemos observar nos exemplos acima, há diversos quadrados brancos que retratam a não-utilização de uma base de  $W$ . A partir disso, é possível notar uma "esparcividade" característica do NMF, que se trata da ocorrência de muitos valores 0 nas matrizes  $W$  e  $H$ . Essa particularidade do NMF advém do fato de muitas aproximações não precisarem utilizar todas essas "partes" — todos os vetores de  $W$  — para aproximarem um vetor de  $V$ .

## Especificação

Non-negative Matrix Factorization (NMF) é uma decomposição que retorna matrizes não-negativas. Algumas notações para não-negatividade:

$$\begin{aligned}
 V_{ij} &\geq 0, \forall i, j \\
 V &\in \mathbb{R}_+^{m \times n} \\
 V &\geq 0
 \end{aligned}$$

## Problema exato do NMF

Dado  $V$  não-negativo, encontre  $W$  e  $H$  não-negativos, tal que  $V = WH$

Problema NP-completo!

## Problema aproximado do NMF

Dado  $V$  não-negativo, encontre  $W$  e  $H$  não-negativos, tal que  $V \approx WH$

i.e, dado  $V$  não-negativo, encontre  $W$  e  $H$  não-negativos que realizam a minimização abaixo:

$$[WH] = \arg \min_{W \geq 0, H \geq 0} f(W, H) = \frac{1}{2} \|V - WH\|_F^2$$

Essa minimização de uma função de 2 variáveis gera um problema de otimização não-convexo, de forma que é difícil encontrar seu mínimo global.

## Algoritmo

O problema da minimização é resolvido usando Gradiente Descendente Alternado (GD).

Para minimizar uma função de uma variável  $f(x)$ , o Gradiente Descendente começa com um ponto inicial  $x_0$  e executa as iterações no formato abaixo:

$$x^{(k+1)} = x^{(k)} - \alpha^{(k)} \nabla_x f(x^{(k)})$$

- $k \in \mathbb{N}$  o número da iteração
- $x^{(k)}$  é a variável na iteração  $k$
- $x^{(k+1)}$  é a variável na iteração  $k + 1$
- $\alpha^{(k)} \geq 0$  é o tamanho do passo na iteração  $k$
- $\nabla_x f(x^{(k)})$  é o gradiente da função  $f$  com respeito à  $x$

Nesse caso, ao invés de utilizar uma função de uma variável, será considerada a função objetivo de erro  $f(W, H)$  do NMF.

A partir disso, temos duas opções para execução do Gradiente Descendente:

1. Atualizar  $W$  e  $H$  ao mesmo tempo
2. Atualizar  $W$  e  $H$  separadamente

Independente do estratégia escolhida, o objetivo é produzir uma sequência de  $\{W^k, H^k\}_{k \in \mathbb{N}}$  tal que o custo da função está sempre decrescendo, de forma que:

$$f(W^{(k+1)}, H^{(k+1)}) \leq f(W^{(k)}, H^{(k)})$$

Geralmente, o segundo método é utilizado, no qual primeiro  $W$  é atualizado com  $H$  fixo, e então  $H$  é atualizado com  $W$  fixo.

$$W^{(k+1)} = W^{(k)} - \alpha_W^{(k)} \nabla_W f(W^{(k)}, H^{(k)})$$

$$H^{(k+1)} = H^{(k)} - \alpha_H^{(k)} \nabla_H f(W^{(k+1)}, H^{(k)})$$

Como já são conhecidos  $W^{(k)}$  e  $H^{(k)}$ , só é necessário descobrir  $\nabla_W f, \nabla_H f, \alpha_W^{(k)}, \alpha_H^{(k)}$

## Propriedades do traço

Para prosseguimento do texto, considere as seguintes propriedades:

$$\|X\|_F^2 = \text{tr}(X^T X) \quad (1)$$

$$\text{tr}(X) = \text{tr}(X^T) \quad (2)$$

$$\nabla_X \text{tr}(AXB) = A^T B^T \quad (3)$$

$$\nabla_X \text{tr}(X^T AX) = (A + A^T)X \quad (4)$$

$$\nabla_X \text{tr}(B^T X^T XB) = 2XBB^T \quad (5)$$

que são variações das propriedades abaixo sobre gradiente aplicado em traços

$$\nabla_X \text{tr}(AX) = A^T$$

$$\nabla_X \text{tr}(X^T A) = A$$

$$\nabla_X \text{tr}(X^T AX) = (A + A^T)X$$

$$\nabla_X \text{tr}(XAX^T) = X(A + A^T)$$

Para exemplificar o funcionamento dessas propriedades, considere o seguinte exemplo  $2 \times 2$ :

$$A = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix}, X = \begin{bmatrix} | & | \\ x_1 & x_2 \\ | & | \end{bmatrix}, X^T = \begin{bmatrix} -x_1^T - \\ -x_2^T - \end{bmatrix}$$

sendo  $a_i$  e  $x_i$  a  $i$ -ésima coluna de  $A$  e  $X$ , respectivamente

$$X^T A = \begin{bmatrix} x_1^T a_1 & x_1^T a_2 \\ x_2^T a_1 & x_2^T a_2 \end{bmatrix}$$

$$\text{tr}(X^T A) = x_1^T a_1 + x_2^T a_2$$

$$\left[ \nabla_X \text{tr}(X^T A) \right] = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix} = A$$

## Desenvolvimento da fórmula

Desenvolvendo a função objetivo

$$\begin{aligned}
\|V - WH\|_F^2 &\stackrel{(1)}{=} \text{tr} \left\{ (V - WH)^T (V - WH) \right\} \\
&= \text{tr} \left\{ V^T V - V^T WH - (WH)^T V + (WH)^T WH \right\} \\
&\stackrel{(2)}{=} \text{tr} \left\{ V^T V - 2W^T H^T V + H^T W^T WH \right\} \\
&\stackrel{(1)}{=} \|V\|_F^2 - 2\text{tr}(W^T H^T V) + \text{tr}(H^T W^T WH)
\end{aligned}$$

Portanto, temos a seguinte igualdade:

$$f(W, H) = \frac{1}{2} \|V - WH\|_F^2 = \|V\|_F^2 - \text{tr}(W^T H^T V) + \frac{1}{2} \text{tr}(H^T W^T WH)$$

Derivando essa função, temos que

$$\begin{aligned}
\nabla_W f &\stackrel{(3,5)}{=} (WH - V)H^T \\
\nabla_H f &\stackrel{(3,4)}{=} W^T(WH - V)
\end{aligned}$$

Substituindo na fórmula de atualização

$$\begin{aligned}
W^{(k+1)} &= W^{(k)} - \alpha_W^{(k)} (WH - V)H^T \\
H^{(k+1)} &= H^{(k)} - \alpha_H^{(k)} W^T(WH - V)
\end{aligned}$$

Nesse momento, só falta escolher os tamanhos dos passos.

Há diversas abordagens para esse problema, porém por enquanto considere

$$[\alpha_W]_{ij} = \frac{W_{ij}}{[WHH^T]_{ij}}, [\alpha_H]_{ij} = \frac{H_{ij}}{[W^T WH]_{ij}}$$

onde  $A_{ij}$  denota o elemento pertencente à linha  $i$  e coluna  $j$  de  $A$ .

Os tamanhos dos passos são calculados para cada elemento da matriz ao invés de serem estruturados para uma matriz toda, de forma que  $\alpha$  é um número e não uma matriz. Essa construção é feita pois são os tamanhos dos passos que garantem a não-negatividade e seus cálculos envolvem divisões — operação que matrizes não contemplam.

Substituindo o tamanho do passos nas equações, temos que

$$\begin{aligned}
W_{ij}^{(k+1)} &= W_{ij} - [\alpha_W]_{ij} [(WH - V)H^T]_{ij} \\
&= W_{ij} - \frac{W_{ij}}{[WHH^T]_{ij}} [(WH - V)H^T]_{ij} \\
&= W_{ij} - \frac{[W(WH - V)H^T]_{ij}}{[WHH^T]_{ij}} \\
&= W_{ij} \left( 1 - \frac{[(WH - V)H^T]_{ij}}{[WHH^T]_{ij}} \right) \\
&= W_{ij} \left( \frac{[WHH^T]_{ij}}{[WHH^T]_{ij}} - \frac{[(WH - V)H^T]_{ij}}{[WHH^T]_{ij}} \right) \\
&= W_{ij} \frac{[VH^T]_{ij}}{[WHH^T]_{ij}}
\end{aligned}$$

$$\begin{aligned}
H_{ij}^{(k+1)} &= H_{ij} - [\alpha_H]_{ij} [W^T(WH - V)]_{ij} \\
&= H_{ij} - \frac{H_{ij}}{[W^TWH]_{ij}} [W^T(WH - V)]_{ij} \\
&= H_{ij} \left( 1 - \frac{[W^T(WH - V)]_{ij}}{[W^TWH]_{ij}} \right) \\
&= H_{ij} \left( \frac{[W^TWH]_{ij}}{[W^TWH]_{ij}} - \frac{[W^T(WH - V)]_{ij}}{[W^TWH]_{ij}} \right) \\
&= H_{ij} \frac{[W^TV]_{ij}}{[W^TWH]_{ij}}
\end{aligned}$$

Repare que o valor de cada elemento nas matrizes da próxima iteração é gerado a partir de operações de produtos e divisões de elementos das matrizes  $V$ ,  $W$  e  $H$  que possuem apenas elementos não-negativos. Portanto, a cada iteração a não-negatividade dos elementos é mantida por esse tamanho de passo sugerido.

Usando a notação para multiplicação elemento-elemento da matriz  $\circ$ , temos o seguinte resultado para cada iteração do algoritmo:

$$\begin{aligned}
W^{(k+1)} &= W^{(k)} \circ \frac{VH^{(k)T}}{WH^{(k)}H^{(k)T}} \\
H^{(k+1)} &= H^{(k)} \circ \frac{W^{(k)T}V}{W^{(k)T}W^{(k)}H^{(k)}}
\end{aligned}$$

### Características

- A cada iteração a função objetivo não aumenta seu valor (Prova de convergência abaixo)
- O algoritmo não garante a convergência para um mínimo global, apenas para um mínimo local
- Para manter a não-negatividade, o tamanho do passo realiza operações elemento a elemento com valores não-negativos

# Convergência

## Teorema (Lee e Seung 2001)

A norma Euclidiana  $\frac{1}{2} \|V - WH\|_F$  não tem seu valor aumentado após as seguintes atualizações:

$$W^{(k+1)} = W^{(k)} \circ \frac{VH^{(k)T}}{WH^{(k)}H^{(k)T}} \quad (6)$$

$$H^{(k+1)} = H^{(k)} \circ \frac{W^{(k)T}V}{W^{(k)T}W^{(k)}H^{(k)}} \quad (7)$$

Por simplicidade, a prova será feita atualizando a função objetivo com foco numa coluna de  $H$ , denotada como  $h$ , para  $W$  fixo. Nesse caso, considere  $f(h)$  expressada como:

$$f(h) = \frac{1}{2} \|Wh - v\|_2^2$$

Na  $k$ -ésima iteração do algoritmo, utilizando a expansão de Taylor de segunda ordem em torno de  $h$  em  $f$  temos

$$F(h) = f(h^{(k)}) + \nabla f(h^{(k)})^T (h - h^{(k)}) + \frac{1}{2} (h - h^{(k)})^T \nabla^2 f(h^{(k)}) (h - h^{(k)}) \quad (8)$$

onde  $h^{(k)}$  é uma constante da obtida na iteração anterior.

Essa função  $F(h)$  foi definida usando  $f(h)$ , pois ao encontrarmos valores menores para  $F$  também serão encontrados valores menores para  $f$ .

## Técnica de otimização

A prova utiliza a técnica de otimização chamada Majorization Minimization, onde uma função auxiliar  $G$  com imagem maior é minimizada, com intuito de minimizar uma função  $F$  de menor imagem; de forma que,  $G$  é um limite superior de  $F$ .

**Definição 1:**  $G(h|h^k)$  é uma função auxiliar para  $F(h)$  se as seguintes condições

$$G(h|h^{(k)}) \geq F(h), G(h|h) = F(h) \quad (9)$$

são satisfeitas.

### Lema 1

Se  $G$  é uma função auxiliar, então o valor de  $F$  não cresce quando a seguinte atualização for feita:

$$h^{(k+1)} = \arg \min_h G(h|h^{(k)})$$

**Prova:**  $F(h^{(k)}) = G(h^{(k)}|h^{(k)}) \geq G(h^{(k+1)}|h^{(k)}) \geq F(h^{(k+1)})$



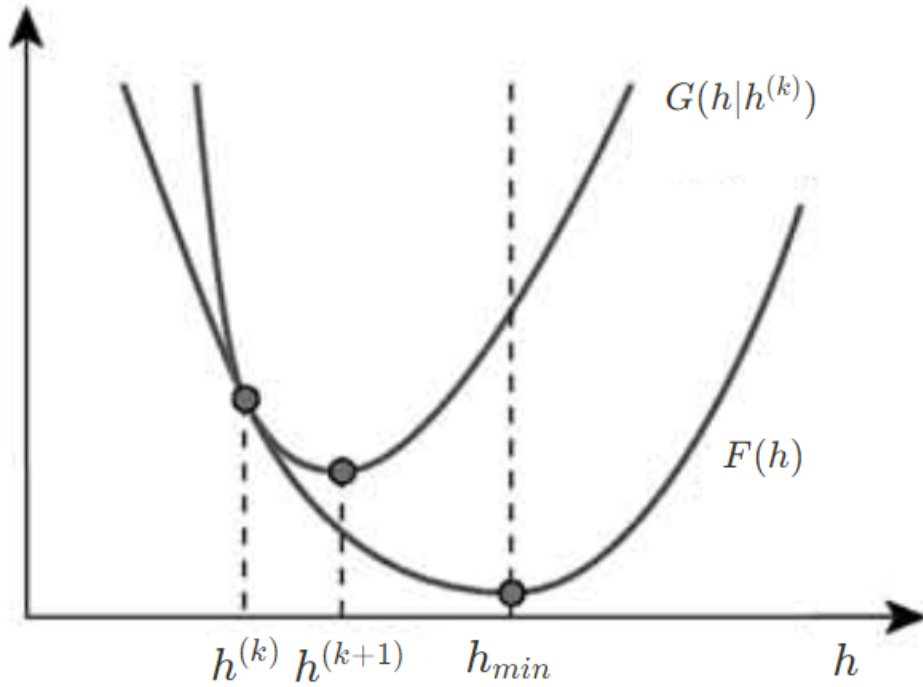


Figura 4: Minimização da função auxiliar  $G(h|h^{(k)}) \geq F(h)$  garante que  $F(h^{(k)}) \geq F(h^{(k+1)})$  para  $h^{(k+1)} = \arg \min_h G(h, h^{(k)})$ .

Para esse problema, considere a seguinte função auxiliar  $G$ :

$$G(h|h^{(k)}) = f(h^{(k)}) + \nabla f(h^{(k)})^T (h - h^{(k)}) + \frac{1}{2} (h - h^{(k)})^T K(h^{(k)}) (h - h^{(k)})$$

Onde propriedade  $G(h|h) = F(h)$  é garantida nos termos  $(h - h^{(k)})$  presentes em  $F$  e  $G$ , e com a condição de  $K(h^{(k)}) \geq \nabla^2 f(h^{(k)})$  para  $G$  ser o limite superior de  $F$ .

## Desenvolvimento de $\nabla^2 f(h^{(k)})$

Para construir  $K$ , é preciso saber mais sobre  $\nabla f$ :

Dado que  $f(h) = \frac{1}{2} \|Wh - v\|_2^2$

$$\begin{aligned} \frac{1}{2} \|Wh - v\|_2^2 &= \frac{1}{2} (Wh - v)^T (Wh - v) \\ &= \frac{1}{2} (h^T W^T - v^T) (Wh - v) \\ &= \frac{1}{2} (h^T W^T Wh - h^T W^T v - v^T Wh - v^T v) \\ &= \frac{1}{2} (h^T W^T Wh - 2h^T W^T v + v^T v) \end{aligned}$$

Derivando uma vez a expressão acima

$$\nabla_h f(h) = W^T W h - W^T v$$

Derivando novamente

$$\nabla_h^2 f(h) = W^T W$$

- $W^T W$  é simétrica
- $W$  é não-negativa  $\rightarrow W^T W$  também é não-negativa

## Matriz $K$ e sua convergência

Com base na condição  $K(h^{(k)}) \geq \nabla^2 f(h^{(k)})$ , Lee e Seung sugerem a seguinte estrutura para  $K(h^{(k)})$ :

$$K_{ab}(h^{(k)}) = \text{Diag}\left(\frac{[W^T W h^{(k)}]_a}{[h^{(k)}]_a}\right)_{ab}$$

Onde a matriz  $\text{Diag}$  para uma matriz  $A$  e um vetor  $x$  é dado por:

$$\begin{aligned} \text{Diag}\left(\frac{[Ax]_i}{[x]_i}\right) &= \text{Diag}\left(\frac{[y]_i}{[x]_i}\right) \\ &= \frac{\text{Diag}(y)}{\text{Diag}(x)} \\ &= D_x^{-1} D_y \\ &= \begin{bmatrix} \frac{y_1}{x_1} & 0 & \dots & 0 \\ 0 & \frac{y_1}{x_1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{y_d}{x_d} \end{bmatrix} \end{aligned}$$

Reescrevendo as funções  $F$  e  $G$

$$\begin{aligned} F(h) &= f(h^{(k)}) + \nabla f(h^{(k)})^T (h - h^{(k)}) + \frac{1}{2} (h - h^{(k)})^T W^T W (h - h^{(k)}) \\ G(h|h^{(k)}) &= f(h^{(k)}) + \dots + \frac{1}{2} (h - h^{(k)})^T \text{Diag}\left(\frac{[W^T W h^{(k)}]_i}{[h^{(k)}]_i}\right) (h - h^{(k)}) \end{aligned}$$

Desenvolvendo  $G(h|h^{(k)}) \geq F(h)$  para atender à condição (9) da função auxiliar

$$(h - h^{(k)})^T \left( \text{Diag}\left(\frac{[W^T W h^{(k)}]_i}{[h^{(k)}]_i}\right) - W^T W \right) (h - h^{(k)}) \geq 0 \quad (10)$$

Para atender essa condição (10), a matriz  $K - W^T W$  deve ser positiva semidefinida, pois toda matriz positiva semidefinida  $A$  possui a condição necessária e suficiente:

$$x^T A x \geq 0 \quad \forall x \neq 0$$

Dessa forma, considere a matriz:

$$L_{ab}(h^{(k)}) = h_a^{(k)} \left( K(h^{(k)}) - W^T W \right)_{ab} h_b^{(k)}$$

que se trata apenas da matriz  $K - W^T W$  com seus componentes escalonados. Então,  $K - W^T W$  é positiva semidefinida se e somente se  $L$  for, e

$$\begin{aligned} v^T L v &= \sum_{ab} v_a L_{ab} v_b \\ &= \sum_{ab} v_a h_a^{(k)} \left( \text{Diag} \left( \frac{[W^T W h^{(k)}]_a}{[h^{(k)}]_a} \right)_{ab} - W^T W_{ab} \right) h_b^{(k)} v_b \\ &= \sum_{ab} h_a^{(k)} (W^T W)_{ab} h_b^{(k)} v_a^2 - v_a h_a^{(k)} (W^T W)_{ab} h_b^{(k)} v_b \\ &= \sum_{ab} (W^T W)_{ab} h_a^{(k)} h_b^{(k)} \left[ \frac{1}{2} v_a^2 + \frac{1}{2} v_b^2 - v_a v_b \right] \\ &= \frac{1}{2} \sum_{ab} (W^T W)_{ab} h_a^{(k)} h_b^{(k)} (v_a - v_b)^2 \\ &\geq 0 \end{aligned}$$

O que prova o teorema de convergência para a variável  $h$  — a prova para  $w$  é similar.

Logo, a cada iteração do algoritmo  $f(W^{(k+1)}, H^{(k+1)}) \leq f(W^{(k)}, H^{(k)})$ .

## Obtenção do tamanho do passo

Para a iteração  $h^{(k+1)}$ , o tamanho do passo pode ser obtido resolvendo o seguinte problema:

$$\frac{\partial G(h|h^{(k)})}{\partial h} = 0$$

$$\frac{\partial G(h|h^{(k)})}{\partial h} = \nabla f(h^{(k)}) + K(h - h^{(k)})$$

$$h = h^{(k)} - K^{-1} \nabla f(h^{(k)})$$

$$\text{Como } K = \text{Diag} \left( \frac{[W^T W h^{(k)}]_i}{[h^{(k)}]_i} \right) \implies K^{-1} = \text{Diag} \left( \frac{[h^{(k)}]_i}{[W^T W h^{(k)}]_i} \right) \text{ e } \nabla f(h^{(k)}) = W^T (W h^{(k)} - v)$$

Portanto, temos

$$\begin{aligned}
h &= h^{(k)} - K^{-1} \nabla f(h^{(k)}) \\
&= h^{(k)} - \text{Diag} \left( \frac{[h^{(k)}]_i}{[W^T W h^{(k)}]_i} \right) W^T (W h^{(k)} - v) \\
&= h^{(k)} - h^{(k)} \otimes \frac{[W^T (W h^{(k)} - v)]_{ij}}{[W^T W h^{(k)}]_{ij}} \\
&= \frac{h^{(k)} W^T W h^{(k)}}{W^T W h^{(k)}} - \frac{h^{(k)} (W^T W h^{(k)} + W^T v)}{W^T W h^{(k)}} \\
&= h^{(k)} \otimes \frac{[W^T v]_{ij}}{[W^T W h^{(k)}]_{ij}}
\end{aligned}$$

que é a fórmula de atualização do teorema.

## Referências

ANG, A. Convergence analysis of nmf algorithm. Mathematique et recherche operationnelle, UMONS, Belgium, 2018. First draft: June 6, 2017; Last update: February 27, 2018. Disponível em:

[https://angms.science/doc/NMF/nmf\\_1.pdf](https://angms.science/doc/NMF/nmf_1.pdf).

ANG, A. Non-negative matrix factorization and multiplicative update. Mathematique et recherche operationnelle, UMONS, Belgium, 2019. First draft: June 6, 2017; Last update: February 25, 2019.

Disponível em: [https://angms.science/doc/NMF/nmf\\_0.pdf](https://angms.science/doc/NMF/nmf_0.pdf).

BURRED, J. J. Detailed derivation of multiplicative update rules for nmf. Paris, France, 2014.

CONTRIBUTORS, W. Non-negative Matrix Factorization. 2024. [https://en.wikipedia.org/wiki/Non-negativematrix\\_factorization](https://en.wikipedia.org/wiki/Non-negativematrix_factorization). Accessed: 2024-07-13.

DUCHI, J. Properties of the trace and matrix derivatives. Available electronically at [https://web.stanford.edu/~jduchi/projects/matrix\\_prop.pdf](https://web.stanford.edu/~jduchi/projects/matrix_prop.pdf), 2007.

LEE, D.; SEUNG, H. S. Algorithms for non-negative matrix factorization. Advances in neural information processing systems, v. 13, 2000.

LEE, D. D.; SEUNG, H. S. Learning the parts of objects by non-negative matrix factorization. nature, Nature Publishing Group UK London, v. 401, n. 6755, p. 788–791, 1999.