Dual Simplex Volume Maximization for Simplex-Structured Matrix Factorization

Maryam Abdolali [1] Giovanni Barbarino [2] Nicolas Gillis [2]

UNIVERSITÀ DEGLI STUDI DELL'AQUILA

DISIM
Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica

03 December 2024

[1]K.N.Toosi University, Tehran, Iran

[2]Université de Mons, Belgium

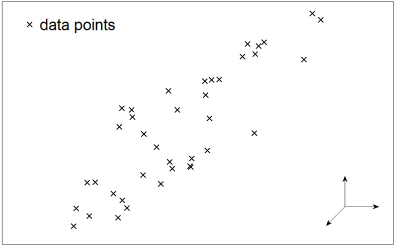
Low-Rank Nonnegative Matrix Factorization

Dimensionality reduction and Constrained Low-Rank Matrix Approximations

- Given n data points m_j $(j = 1, 2, \ldots, n)$, we would like to understand the underlying structure of this data through linear dimensionality reduction: find a set of r basis vectors u_k $(1 \leq k \leq r)$ so that for some weights v_{kj}
- This is **equivalent** to the low-rank approximation of matrix M:

$$M = [m_1\, m_2\, \ldots\, m_n] \;\approx\; [u_1\, u_2\, \ldots\, u_r]\,[v_1\, v_2\, \ldots\, v_n] \;=\; UV$$

m_j



- How to measure the **error** $||M - UV||$?

  Ex. PCA/truncated SVD use $||X||$ or $||X||_F^2$.

- What **constraints** should the factors $U \in \Omega_U$ and $V \in \Omega_V$ satisfy?

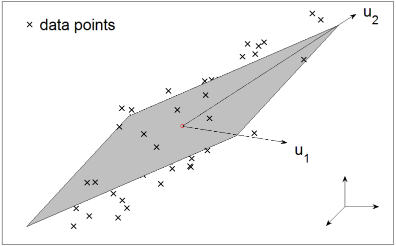
  Ex. PCA has no constraints, k-means a single '1' per column of V.

Dimensionality reduction and Constrained Low-Rank Matrix Approximations

- Given n data points m_j ($j = 1, 2, \dots, n$), we would like to understand the underlying structure of this data through linear dimensionality reduction: find a set of r basis vectors u_k ($1 \leq k \leq r$) so that for some weights v_{kj}
- This is **equivalent** to the low-rank approximation of matrix M:

$$M = [m_1\, m_2\, \dots\, m_n] \;\approx\; [u_1\, u_2\, \dots\, u_r]\,[v_1\, v_2\, \dots\, v_n] \;=\; UV$$

$$m_j \;\approx\; \sum_{k=1}^{r} u_k\, v_{kj}$$



- How to measure the **error** $||M - UV||$?

  Ex. PCA/truncated SVD use $||X||$ or $||X||_F^2$.

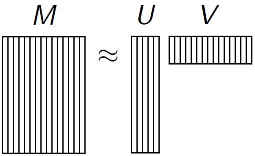
- What **constraints** should the factors $U \in \Omega_U$ and $V \in \Omega_V$ satisfy?

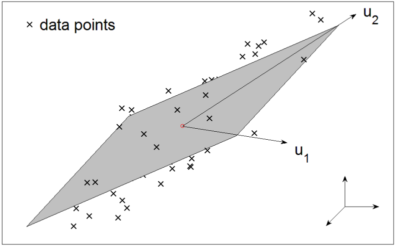
  Ex. PCA has no constraints, k-means a single '1' per column of V.

Dimensionality reduction and Constrained Low-Rank Matrix Approximations

- Given n data points m_j ($j = 1, 2, \dots, n$), we would like to understand the underlying structure of this data through linear dimensionality reduction: find a set of r basis vectors u_k ($1 \leq k \leq r$) so that for some weights v_{kj}
- This is **equivalent** to the low-rank approximation of matrix M:

$$M = [m_1\ m_2\ \dots\ m_n] \approx [u_1\ u_2\ \dots\ u_r]\,[v_1\ v_2\ \dots\ v_n] = UV$$





- How to measure the **error** $||M - UV||$?

  Ex. PCA/truncated SVD use $||X||$ or $||X||_F^2$.

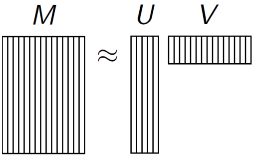
- What **constraints** should the factors $U \in \Omega_U$ and $V \in \Omega_V$ satisfy?

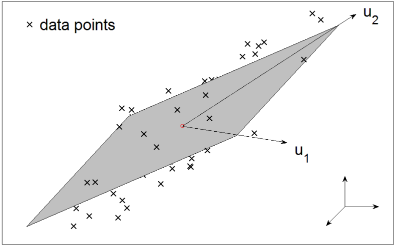
  Ex. PCA has no constraints, k-means a single '1' per column of V.

Dimensionality reduction and Constrained Low-Rank Matrix Approximations

- Given n data points m_j $(j = 1, 2, \dots, n)$, we would like to understand the underlying structure of this data through linear dimensionality reduction: find a set of r basis vectors u_k $(1 \leq k \leq r)$ so that for some weights v_{kj}
- This is **equivalent** to the low-rank approximation of matrix M:

$$M = [m_1\, m_2\, \dots\, m_n] \;\approx\; [u_1\, u_2\, \dots\, u_r]\,[v_1\, v_2\, \dots\, v_n] \;=\; UV$$





- How to measure the **error** $||M - UV||$?

  Ex. PCA/truncated SVD use $||X||$ or $||X||_F^2$.

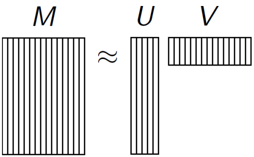
- What **constraints** should the factors $U \in \Omega_U$ and $V \in \Omega_V$ satisfy?

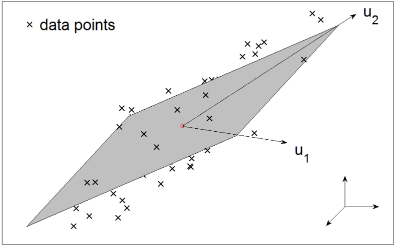
  Ex. PCA has no constraints, k-means a single '1' per column of V.

Dimensionality reduction and Constrained Low-Rank Matrix Approximations

- Given n data points m_j ($j = 1, 2, \ldots, n$), we would like to understand the underlying structure of this data through linear dimensionality reduction: find a set of r basis vectors u_k ($1 \leq k \leq r$) so that for some weights v_{kj}
- This is **equivalent** to the low-rank approximation of matrix M:

$$M = [m_1\, m_2\, \ldots\, m_n] \approx [u_1\, u_2\, \ldots\, u_r]\,[v_1\, v_2\, \ldots\, v_n] = UV$$





- How to measure the **error** $||M - UV||$?

  Ex. PCA/truncated SVD use $||X||$ or $||X||_F^2$.

- What **constraints** should the factors $U \in \Omega_U$ and $V \in \Omega_V$ satisfy?

  Ex. PCA has no constraints, k-means a single '1' per column of V.

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}_+^{p \times n}$ and a factorization rank $r \ll \min(p, n)$, find $U \in \mathbb{R}_+^{p \times r}$ and $V \in \mathbb{R}_+^{r \times n}$ such that

$$\min_{U \geq 0, V \geq 0} ||M - UV||_F^2 = \sum_{i,j} (M - UV)_{ij}^2 \qquad \text{(NMF)}$$

NMF is a linear dimensionality reduction technique for nonnegative data :

$$\underbrace{M(:,i)}_{\geq 0} \approx \sum_{k=1}^{r} \underbrace{U(:,k)}_{\geq 0} \underbrace{V(k,i)}_{\geq 0} \quad \text{for all } i$$

Why nonnegativity?

→ **Interpretability**: Nonnegativity constraints lead to easily interpretable factors (and a sparse and part-based representation)
→ **Many applications.** image processing, text mining, audio source separation, recommender systems, hyperspectral unmixing, community detection, clustering, etc.

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}_{+}^{p \times n}$ and a factorization rank $r \ll \min(p, n)$, find $U \in \mathbb{R}_{+}^{p \times r}$ and $V \in \mathbb{R}_{+}^{r \times n}$ such that

$$\min_{U \geq 0, V \geq 0} ||M - UV||_F^2 = \sum_{i,j} (M - UV)_{ij}^2 \qquad \text{(NMF)}$$

NMF is a linear dimensionality reduction technique for nonnegative data :

$$\underbrace{M(:, i)}_{\geq 0} \approx \sum_{k=1}^{r} \underbrace{U(:, k)}_{\geq 0} \underbrace{V(k, i)}_{\geq 0} \qquad \text{for all } i$$

Why nonnegativity?

→ **Interpretability**: Nonnegativity constraints lead to easily interpretable factors (and a sparse and part-based representation)

→ **Many applications.** image processing, text mining, audio source separation, recommender systems, hyperspectral unmixing, community detection, clustering, etc.

Nonnegative Matrix Factorization (NMF)

Given a matrix $M \in \mathbb{R}_+^{p \times n}$ and a factorization rank $r \ll \min(p, n)$, find $U \in \mathbb{R}_+^{p \times r}$ and $V \in \mathbb{R}_+^{r \times n}$ such that

$$\min_{U \geq 0, V \geq 0} ||M - UV||_F^2 = \sum_{i,j} (M - UV)_{ij}^2 \qquad \text{(NMF)}$$

NMF is a linear dimensionality reduction technique for nonnegative data :

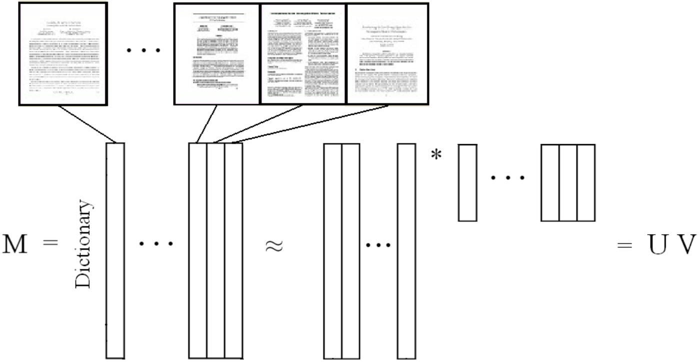
$$\underbrace{M(:, i)}_{\geq 0} \approx \sum_{k=1}^{r} \underbrace{U(:, k)}_{\geq 0} \underbrace{V(k, i)}_{\geq 0} \qquad \text{for all } i$$

Why nonnegativity?

→ **Interpretability**: Nonnegativity constraints lead to easily interpretable factors (and a sparse and part-based representation)

→ **Many applications.** image processing, text mining, audio source separation, recommender systems, hyperspectral unmixing, community detection, clustering, etc.

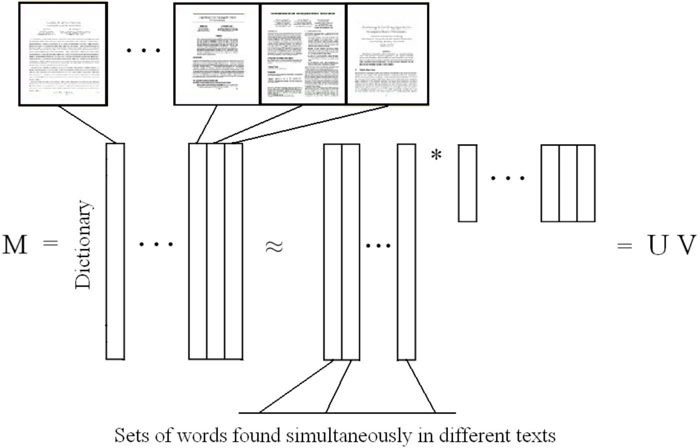
Application 1: topic recovery and document classification



- $M_{i,j}$ are the frequencies of word i in document j
- The columns $U_{:,k}$ represent the topics in the documents
- Weights in $V_{:,j}$ allow to assign each document j to its corresponding topics

Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999)

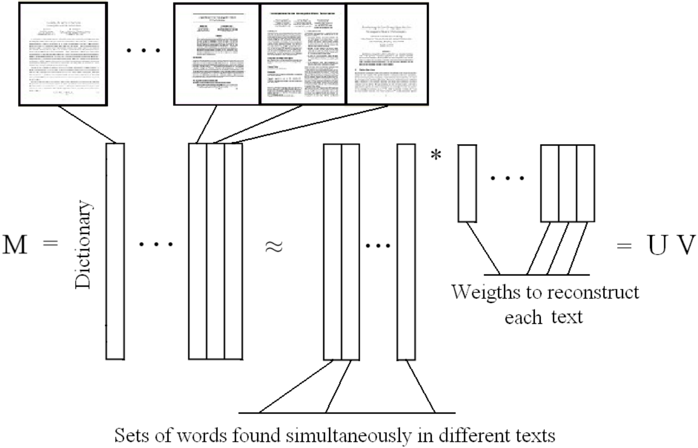
Application 1: topic recovery and document classification



- $M_{i,j}$ are the frequencies of word i in document j
- The columns $U_{:,k}$ represent the topics in the documents
- Weights in $V_{:,j}$ allow to assign each document j to its corresponding topics

Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999)

Application 1: topic recovery and document classification



- $M_{i,j}$ are the frequencies of word i in document j
- The columns $U_{:,k}$ represent the topics in the documents
- Weights in $V_{:,j}$ allow to assign each document j to its corresponding topics

Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999)

Application 2: recommender systems

In some cases, some entries are missing/unknown

For example, we would like to predict how much someone is going to like a movie based on its movie preferences (e.g., 1 to 5 stars) :

$$\text{Movies} \quad \overset{\text{Users}}{\begin{bmatrix} 2 & 3 & 2 & ? & ? \\ ? & 1 & ? & 3 & 2 \\ 1 & ? & 4 & 1 & ? \\ 5 & 4 & ? & 3 & 2 \\ ? & 1 & 2 & ? & 4 \\ 1 & ? & 3 & 4 & 3 \end{bmatrix}}$$

Movies ratings are modeled as linear combinations of 'feature' movies (related to the genres - child oriented, serious vs. escapist, thriller, romantic, etc.)

$$\underbrace{M(i,:)}_{\text{movie } i} \approx \sum_{k=1}^{r} \underbrace{U(i,k)}_{\text{weights}} \underbrace{V(k,:)}_{\text{genre } k}$$

Application 2: recommender systems

In some cases, some entries are missing/unknown

For example, we would like to predict how much someone is going to like a movie based on its movie preferences (e.g., 1 to 5 stars) :

$$\text{Movies} \quad \overset{\text{Users}}{\begin{bmatrix} 2 & 3 & 2 & ? & ? \\ ? & 1 & ? & 3 & 2 \\ 1 & ? & 4 & 1 & ? \\ 5 & 4 & ? & 3 & 2 \\ ? & 1 & 2 & ? & 4 \\ 1 & ? & 3 & 4 & 3 \end{bmatrix}}$$

Movies ratings are modeled as linear combinations of 'feature' movies (related to the genres - child oriented, serious vs. escapist, thriller, romantic, etc.)

$$\underbrace{M(i,:)}_{\text{movie } i} \approx \sum_{k=1}^{r} \underbrace{U(i,k)}_{\text{weights}} \underbrace{V(k,:)}_{\text{genre } k}$$

For example, using a rank-2 factorization on the Netflix dataset, female vs. male and serious vs. escapist behaviors were extracted

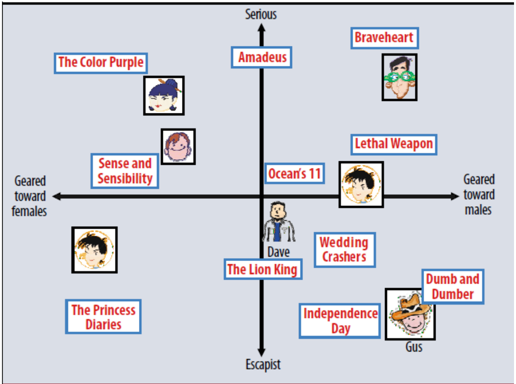


Figure 2. A simplified illustration of the latent factor approach, which characterizes both users and movies using two axes—male versus female and serious versus escapist.

Koren, Bell, Volinsky, *Matrix Factorization Techniques for Recommender Systems, 2009*
Winners of the Netflix prize 1,000,000$

Simplex-Structured Matrix Factorization

Geometric interpretation of exact NMF

Given $M = UV$, one can scale M and U such that they become column stochastic implying that V is column stochastic:

$$M = UV \iff M' = MD_M = (UD_U)(D_U^{-1}VD_M) = U'V'$$

The columns of M' are convex combinations of the columns of U':

$$M'_{:j} = \sum_{i=1}^{k} U'_{:i} V'_{ij} \quad \text{with} \quad \sum_{i=1}^{k} V'_{ij} = 1 \ \forall j, \ V'_{ij} \geq 0 \, \forall ij$$

In other terms

$$\text{conv}(M') \quad \subseteq \quad \text{conv}(U') \quad \subseteq \quad \Delta^n,$$

where $\text{conv}(X)$ is the convex hull of the columns of X, and $\Delta^n = \{x \in \mathbb{R}^n \,|x \geq 0, \sum_{i=1}^n x_i = 1 \}$ is the unit simplex

Exact NMF $\equiv$ Find r points whose convex hull is nested between two given polytopes (Nested Polytope Problem)

Vavasis, S.A.: On the complexity of nonnegative matrix factorization. SIAM Journal on Optimization 20(3), 1364–1377 (2009)

Geometric interpretation of exact NMF

Given $M = UV$, one can scale M and U such that they become column stochastic implying that V is column stochastic:

$$M = UV \iff M' = MD_M = (UD_U)(D_U^{-1} V D_M) = U'V'$$

The columns of M' are convex combinations of the columns of U':

$$M'_{:j} = \sum_{i=1}^{k} U'_{:i} V'_{ij} \quad \text{with} \quad \sum_{i=1}^{k} V'_{ij} = 1 \ \forall j, \ V'_{ij} \geq 0 \, \forall ij$$

In other terms

$$\text{conv}(M') \quad \subseteq \quad \text{conv}(U') \quad \subseteq \quad \Delta^n,$$

where $\text{conv}(X)$ is the convex hull of the columns of X, and $\Delta^n = \{x \in \mathbb{R}^n \ | x \geq 0, \sum_{i=1}^n x_i = 1 \ \}$ is the unit simplex

Exact NMF $\equiv$ Find r points whose convex hull is nested between two given polytopes (Nested Polytope Problem)

Vavasis, S.A.: On the complexity of nonnegative matrix factorization. SIAM Journal on Optimization 20(3), 1364–1377 (2009)

Geometric interpretation of exact NMF

Given $M = UV$, one can scale M and U such that they become column stochastic implying that V is column stochastic:

$$M = UV \iff M' = MD_M = (UD_U)(D_U^{-1}VD_M) = U'V'$$

The columns of M' are convex combinations of the columns of U':

$$M'_{:j} = \sum_{i=1}^{k} U'_{:i} V'_{ij} \quad \text{with} \quad \sum_{i=1}^{k} V'_{ij} = 1 \;\; \forall j, \;\; V'_{ij} \geq 0 \, \forall ij$$

In other terms

$$\text{conv}(M') \quad \subseteq \quad \text{conv}(U') \quad \subseteq \quad \Delta^n,$$

where conv(X) is the convex hull of the columns of X, and $\Delta^n = \{x \in \mathbb{R}^n \; | x \geq 0, \sum_{i=1}^n x_i = 1 \; \}$ is the unit simplex

Exact NMF $\equiv$ Find r points whose convex hull is nested between two given polytopes (Nested Polytope Problem)

Vavasis, S.A.: On the complexity of nonnegative matrix factorization. SIAM Journal on Optimization 20(3), 1364–1377 (2009)

Geometric interpretation of exact NMF

Given $M = UV$, one can scale M and U such that they become column stochastic implying that V is column stochastic:

$$M = UV \iff M' = MD_M = (UD_U)(D_U^{-1}VD_M) = U'V'$$

The columns of M' are convex combinations of the columns of U':

$$M'_{:j} = \sum_{i=1}^{k} U'_{:i} V'_{ij} \quad \text{with} \quad \sum_{i=1}^{k} V'_{ij} = 1 \;\; \forall j, \;\; V'_{ij} \geq 0 \, \forall ij$$

In other terms

$$\text{conv}(M') \quad \subseteq \quad \text{conv}(U') \quad \subseteq \quad \Delta^n,$$

where conv(X) is the convex hull of the columns of X, and
$\Delta^n = \{x \in \mathbb{R}^n \; | x \geq 0, \sum_{i=1}^{n} x_i = 1 \; \}$ is the unit simplex

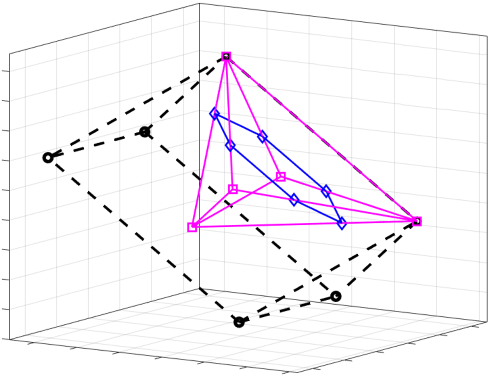
Exact NMF $\equiv$ Find r points whose convex hull is nested between two given polytopes (Nested Polytope Problem)

Vavasis, S.A.: On the complexity of nonnegative matrix factorization. SIAM Journal on Optimization 20(3), 1364–1377 (2009)

Geometric interpretation of exact NMF

$$\text{(NMF)} \qquad M = UV \iff \text{conv}(M') \subseteq \text{conv}(U') \subseteq \Delta^n \qquad (NPP)$$

Exact NMF $\equiv$ Find r points whose convex hull is nested between two given polytopes (Nested Polytope Problem)



Vavasis, S.A.: On the complexity of nonnegative matrix factorization. SIAM Journal on Optimization 20(3), 1364–1377 (2009)

from NMF to SSMF and back

Simplex-Structured Matrix Factorization requires V nonnegative and column stochastic: the columns of M are approximated as convex combinations of the basis vectors in U

$$\min_{U,V\geq 0} ||M - UV||_F^2 \quad : \quad V(:,j) \in \Delta := \{x \geq 0 : e^T x = 1\} \quad \forall j \qquad \text{(SSMF)}$$

Notice that we do not require M, U nonnegative or stochastic

Equivalently

$$\text{conv}(M) \quad \tilde{\subseteq} \quad \text{conv}(U)$$

from NMF to SSMF and back

Simplex-Structured Matrix Factorization requires V nonnegative and column stochastic: the columns of M are approximated as convex combinations of the basis vectors in U

$$\min_{U,V\geq 0} ||M - UV||_F^2 \quad : \quad V(:,j) \in \Delta := \{x \geq 0 : e^T x = 1\} \quad \forall j \qquad \text{(SSMF)}$$

Notice that we do not require M, U nonnegative or stochastic

Equivalently

$$\text{conv}(M) \quad \widetilde{\subseteq} \quad \text{conv}(U)$$

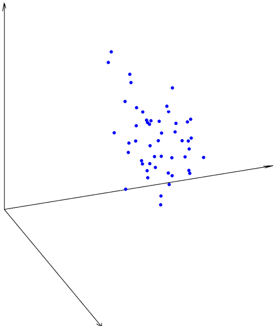
Simplex-Structured Matrix Factorization requires V nonnegative and column stochastic: the columns of M are approximated as convex combinations of the basis vectors in U

$$\min_{U,V\geq 0} ||M - UV||_F^2 \quad : \quad V(:,j) \in \Delta := \{x \geq 0 : e^T x = 1\} \quad \forall j \qquad \text{(SSMF)}$$

Notice that we do not require M, U nonnegative or stochastic

Equivalently

$$\text{conv}(M) \quad \widetilde{\subseteq} \quad \text{conv}(U)$$



from NMF to SSMF and back

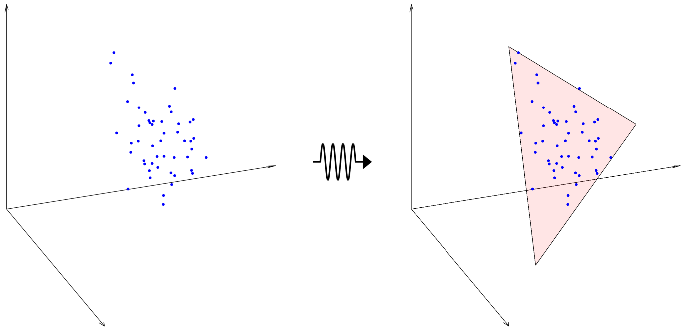
Simplex-Structured Matrix Factorization requires V nonnegative and column stochastic: the columns of M are approximated as convex combinations of the basis vectors in U

$$\min_{U,V\geq 0} ||M - UV||_F^2 \quad : \quad V(:,j) \in \Delta := \{x \geq 0 : e^T x = 1\} \quad \forall j \qquad \text{(SSMF)}$$

Notice that we do not require M, U nonnegative or stochastic

Equivalently

$$\text{conv}(M) \quad \tilde{\subseteq} \quad \text{conv}(U)$$

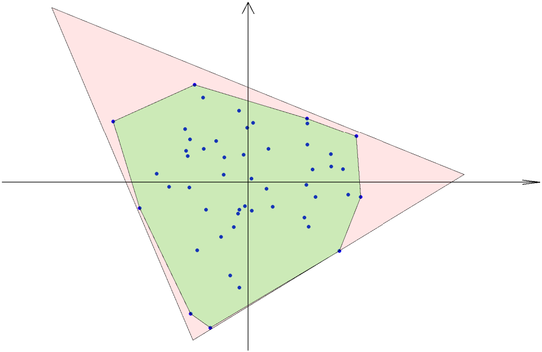


Solution to SSMF

$$Conv(M) \subseteq Conv(U) \qquad U \in \mathbb{R}^{m\times r}$$

Exists? **Yes** for $r \geq \text{dimaff}(M) + 1$ ...

but it is far from being ***Unique***

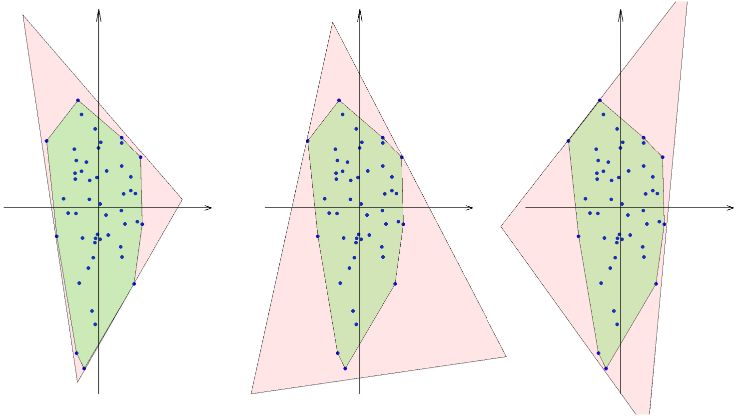


Solution to SSMF

$$Conv(M) \subseteq Conv(U) \qquad U \in \mathbb{R}^{m \times r}$$

Exists? **Yes** for $r \geq \text{dimaff}(M) + 1$ ...

but it is far from being ***Unique***

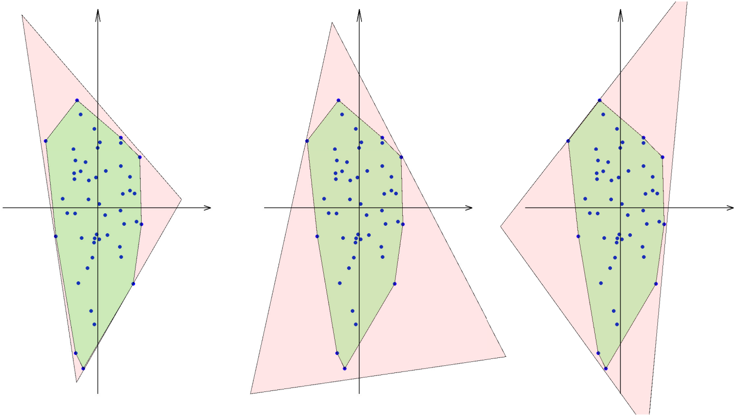


Solution to SSMF

$$Conv(M) \subseteq Conv(U) \qquad U \in \mathbb{R}^{m \times r}$$

Exists? **Yes** for $r \geq \text{dimaff}(M) + 1$ ...

but it is far from being ***Unique***



This is a problem for the **Interpretability** of the solution and the **Stability** of the algorithms

Application: Blind hyperspectral unmixing

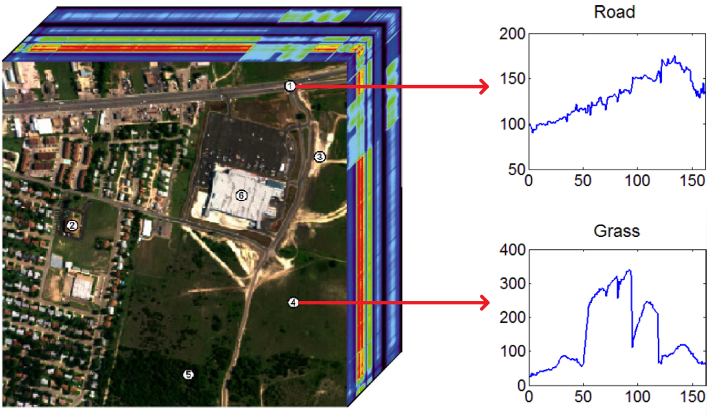


Figure 1: Urban hyperspectral image, 162 spectral bands and 307-by-307 pixels.

Problem. Identify the materials and classify the pixels

Application: Blind hyperspectral unmixing

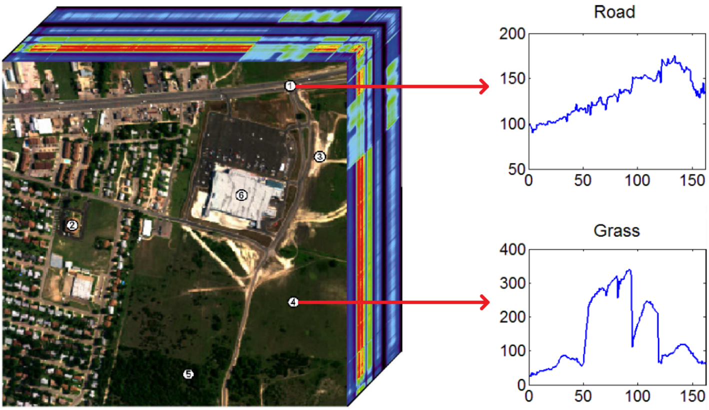
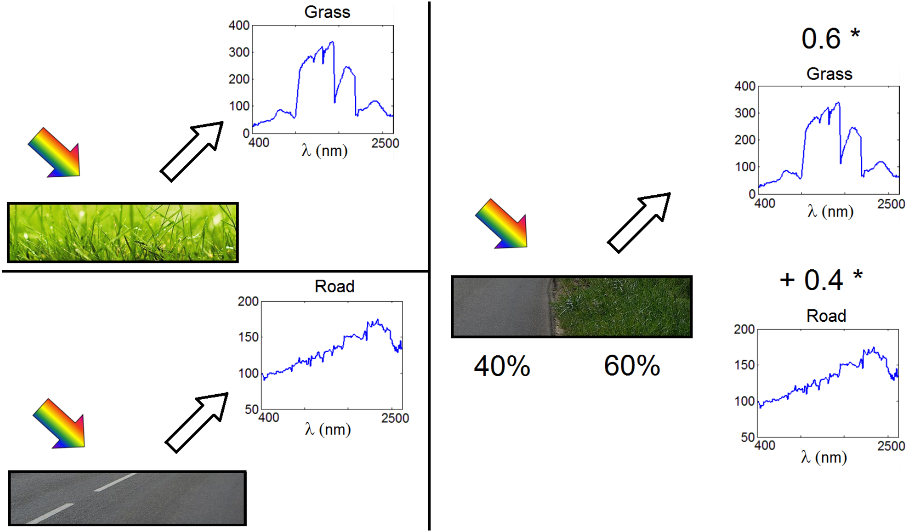


Figure 1: Urban hyperspectral image, 162 spectral bands and 307-by-307 pixels.

Problem. Identify the materials and classify the pixels

Application: Blind hyperspectral unmixing



Urban hyperspectral image

$$\underbrace{\mathrm{M}(:,j)}_{\substack{\text{spectral signature}\\ \text{of } j\text{th pixel}}} \approx \sum_{k=1} \underbrace{\mathrm{U}(:,k)} \qquad \underbrace{\mathrm{V}(k,j)} \quad .$$

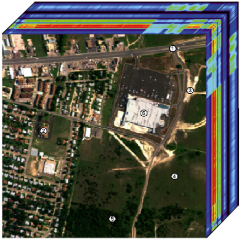
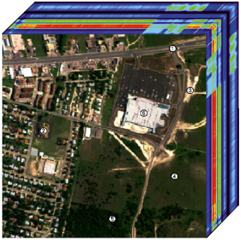


Figure 1: Decomposition of the Urban dataset

Zhu, F.: Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey. (2017) Arxiv

Urban hyperspectral image

$$\underbrace{\mathrm{M}(:,j)}_{\substack{\text{spectral signature}\\ \text{of } j\text{th pixel}}} \approx \sum_{k=1} \underbrace{\mathrm{U}(:,k)}_{\substack{\text{spectral signature}\\ \text{of } k\text{th endmember}}} \underbrace{\mathrm{V}(k,j)} .$$



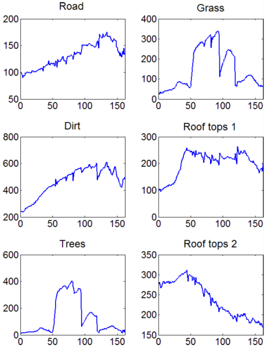


Figure 1: Decomposition of the Urban dataset

Zhu, F.: Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey. (2017) Arxiv

Urban hyperspectral image

$$\underbrace{\mathrm{M}(:,j)}_{\substack{\text{spectral signature}\\ \text{of } j\text{th pixel}}} \approx \sum_{k=1} \underbrace{\mathrm{U}(:,k)}_{\substack{\text{spectral signature}\\ \text{of } k\text{th endmember}}} \underbrace{\mathrm{V}(k,j)}_{\substack{\text{abundance of } k\text{th endmember}\\ \text{in } j\text{th pixel}}} .$$

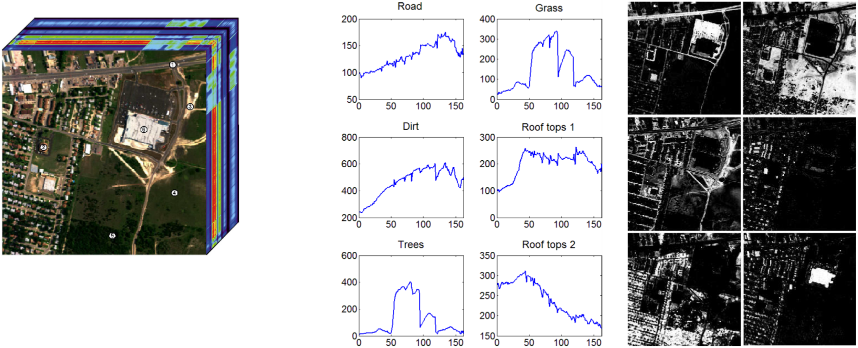
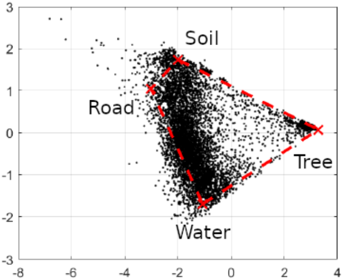


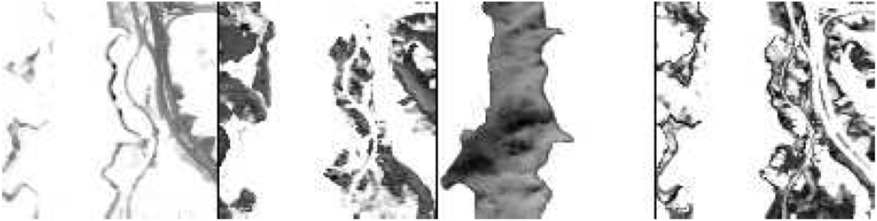
Figure 1: Decomposition of the Urban dataset

Zhu, F.: Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey. (2017) Arxiv

Urban hyperspectral image

Jasper Ridge Data set



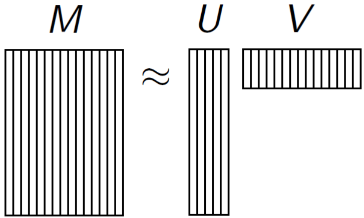


Zhu, F.: Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey. (2017) Arxiv

Separability and Successive Projections Algorithm

Separability Assumption

Separability of M: for an r-index set $\mathcal{K}$ and a V column stochastic, $M = \underbrace{M(:,\mathcal{K})}_{U} V$



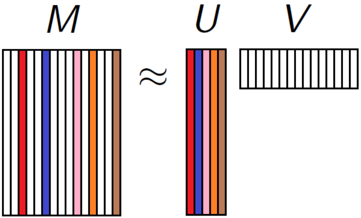
Arora, Ge, Kannan, Moitra, *Computing a Nonnegative Matrix Factorization – Provably*, STOC 2012

If U is full rank, then the separability of $M = UV$ can be expressed as either

- U is a subset of the columns of M
- $V \in \mathbb{R}_+^{r \times n}$ has I_r as submatrix (up to permutation)
- $\text{conv}(U) = \text{conv}(M)$, so
  $M = \widetilde{U}\widetilde{V} \implies \text{conv}(U) \subseteq \text{conv}(\widetilde{U})$

Separability Assumption

Separability of M: for an r-index set $\mathcal{K}$ and a V column stochastic, $M = \underbrace{M(:,\mathcal{K})}_{U} V$



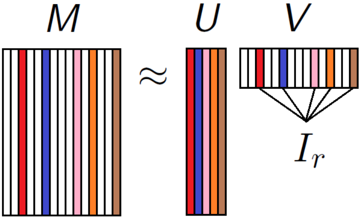
Arora, Ge, Kannan, Moitra, *Computing a Nonnegative Matrix Factorization – Provably*, STOC 2012

If U is full rank, then the separability of $M = UV$ can be expressed as either

- U is a subset of the columns of M
- $V \in \mathbb{R}_+^{r \times n}$ has I_r as submatrix (up to permutation)
- $\text{conv}(U) = \text{conv}(M)$, so $M = \widetilde{U}\widetilde{V} \implies \text{conv}(U) \subseteq \text{conv}(\widetilde{U})$

Separability Assumption

Separability of M: for an r-index set $\mathcal{K}$ and a V column stochastic, $M = \underbrace{M(:,\mathcal{K})}_{U} V$



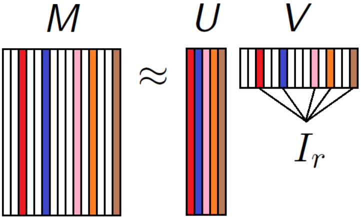
Arora, Ge, Kannan, Moitra, *Computing a Nonnegative Matrix Factorization – Provably*, STOC 2012

If U is full rank, then the separability of $M = UV$ can be expressed as either

- U is a subset of the columns of M
- $V \in \mathbb{R}_+^{r \times n}$ has I_r as submatrix (up to permutation)
- $\text{conv}(U) = \text{conv}(M)$, so
  $M = \widetilde{U}\widetilde{V} \implies \text{conv}(U) \subseteq \text{conv}(\widetilde{U})$

Separability Assumption

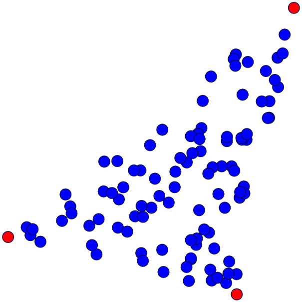
Separability of M: for an r-index set $\mathcal{K}$ and a V column stochastic, $M = \underbrace{M(:,\mathcal{K})}_{U}\, V$



Arora, Ge, Kannan, Moitra, *Computing a Nonnegative Matrix Factorization – Provably*, STOC 2012

If U is full rank, then the separability of $M = UV$ can be expressed as either

- U is a subset of the columns of M
- $V \in \mathbb{R}_+^{r \times n}$ has I_r as submatrix (up to permutation)
- conv(U) = conv(M), so
  $M = \widetilde{U}\widetilde{V} \implies \text{conv}(U) \subseteq \text{conv}(\widetilde{U})$



Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$

The solution will be $U = M(:,\mathcal{K})$ and $V = U^\dagger M$

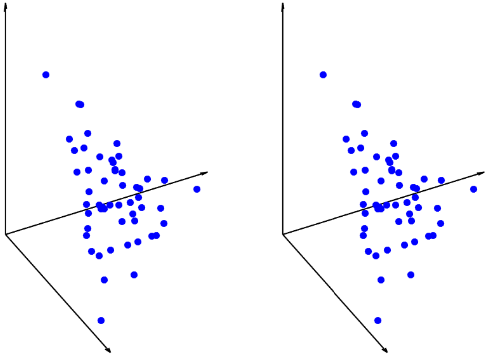
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^{\dagger} M$

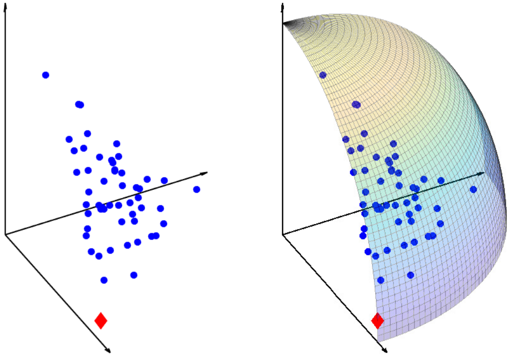
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^{\dagger} M$

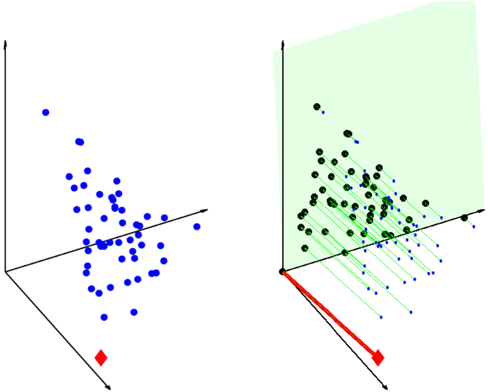
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^\dagger M$

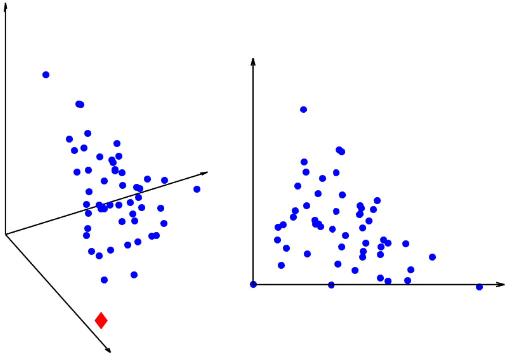
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^{\dagger} M$

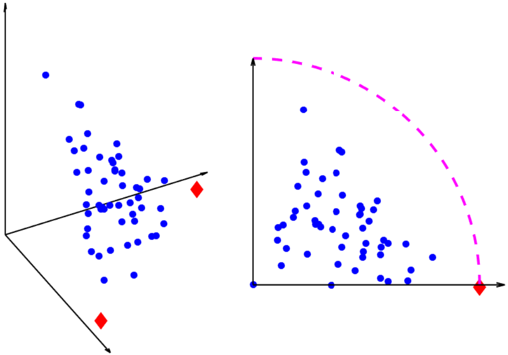
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^\dagger M$

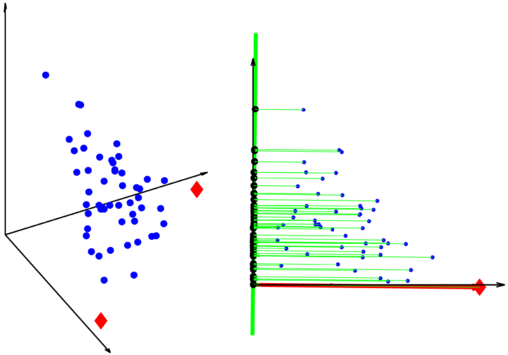
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^\dagger M$

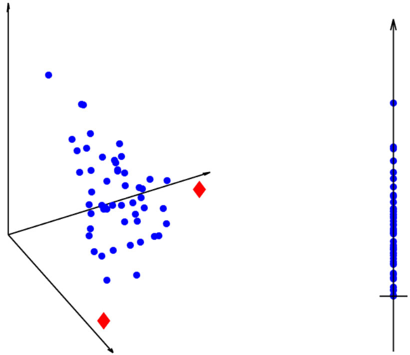
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^{\dagger} M$

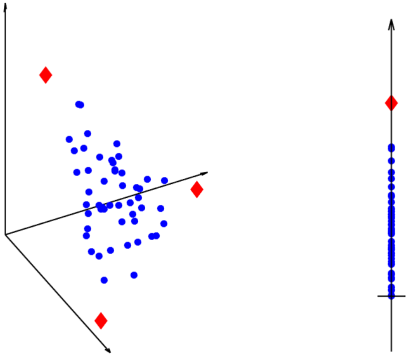
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^\dagger M$

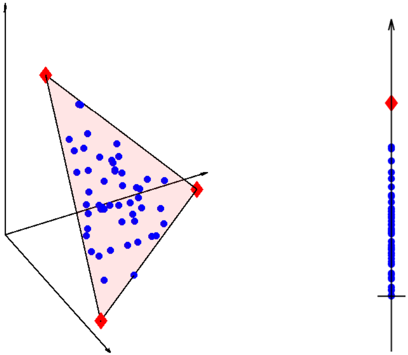
Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Given $M = UV$ separable, U full column rank equal to r, repeat for r times:

1: Find $j^* = \text{argmax}_j ||M(:,j)||$ and add it to $\mathcal{K}$

2: $M \leftarrow \left(I - uu^T\right) M$ where $u = M(:,j^*)/||M(:,j^*)||$



The solution will be $U = M(:,\mathcal{K})$ and $V = U^\dagger M$

Araújo, U., Saldanha, B., Galvão, R., Yoneyama, T., Chame, H., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometrics and Intelligent Laboratory Systems 57 (2001)

Successive Projection Algorithm (SPA)

Perturbation robustness: suppose $M = UV + N$ with UV separable, U full rank and each column of N with norm at most ε

- If $\varepsilon \leq \mathcal{O}(\sigma_r(U)/\mathcal{K}(U)^2)$ then SPA extract a matrix $\widetilde{U}$ such that

$$\max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)^2\right) \quad \text{(sharp for } r \geq 3\text{)}$$

Gillis, N., Vavasis, S.A.: Fast and robust recursive algorithms for separable nonnegative matrix factorization. IEEE Transactions on Pattern Analysis and Machine Intelligence 36(4), 698–714 (2013)

Barbarino G, Gillis N.: On the Robustness of the Successive Projection Algorithm, (2024) Arxiv

Variants to improve robustness to perturbations and outliers:

- **SPA^2:** Apply SPA to M to obtain U_1 and apply SPA to $U_1^\dagger M = (U_1^\dagger U)V + U_1^\dagger N$

$$\varepsilon \leq \mathcal{O}(\sigma_r(U)/\mathcal{K}(U)^2) \implies \max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)\right) \quad \text{(sharp)}$$

- **MVE:** Minimum Volume Ellipsoid $A \succ 0$ s.t. $m_j^\top A m_j \leq 1 \ \forall j$ and use $(A^{-1/2})^\dagger$

$$\varepsilon \leq \mathcal{O}(\sigma_r(U)/(r\sqrt{r})) \implies \max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)\right) \quad \text{(sharp)}$$

- Randomized/Smoothed variants (RandNMF, VCA): instead of looking for

Successive Projection Algorithm (SPA)

Perturbation robustness: suppose $M = UV + N$ with UV separable, U full rank and each column of N with norm at most ε

- If $\varepsilon \leq \mathcal{O}(\sigma_r(U)/\mathcal{K}(U)^2)$ then SPA extract a matrix $\widetilde{U}$ such that

$$\max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)^2\right) \quad \text{(sharp for } r \geq 3\text{)}$$

Variants to improve robustness to perturbations and outliers:

- **SPA^2:** Apply SPA to M to obtain U_1 and apply SPA to $U_1^\dagger M = (U_1^\dagger U)V + U_1^\dagger N$

$$\varepsilon \leq \mathcal{O}(\sigma_r(U)/\mathcal{K}(U)^2) \implies \max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)\right) \quad \text{(sharp)}$$

Gillis, N., Ma, W.K.: Enhancing pure-pixel identification performance via preconditioning. SIAM Journal on Imaging Sciences 8(2), 1161–1186 (2015)

- **MVE:** Minimum Volume Ellipsoid $A \succ 0$ s.t. $m_j^\top A m_j \leq 1 \; \forall j$ and use $(A^{-1/2})^\dagger$

$$\varepsilon \leq \mathcal{O}(\sigma_r(U)/(r\sqrt{r})) \implies \max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)\right) \quad \text{(sharp)}$$

- Randomized/Smoothed variants (RandNMF, VCA): instead of looking for $\text{argmax}_j \|M(:,j)\|$ take a random $u \in \mathbb{R}^n$ and choose as vertex the average of the p

Successive Projection Algorithm (SPA)

Perturbation robustness: suppose $M = UV + N$ with UV separable, U full rank and each column of N with norm at most ε

- If $\varepsilon \leq \mathcal{O}(\sigma_r(U)/\mathcal{K}(U)^2)$ then SPA extract a matrix $\widetilde{U}$ such that

$$\max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)^2\right) \quad \text{(sharp for } r \geq 3\text{)}$$

Variants to improve robustness to perturbations and outliers:

- **SPA^2:** Apply SPA to M to obtain U_1 and apply SPA to $U_1^\dagger M = (U_1^\dagger U)V + U_1^\dagger N$

$$\varepsilon \leq \mathcal{O}(\sigma_r(U)/\mathcal{K}(U)^2) \implies \max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)\right) \quad \text{(sharp)}$$

- **MVE:** Minimum Volume Ellipsoid $A \succ 0$ s.t. $m_j^\top A m_j \leq 1 \; \forall j$ and use $(A^{-1/2})^\dagger$

$$\varepsilon \leq \mathcal{O}(\sigma_r(U)/(r\sqrt{r})) \implies \max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)\right) \quad \text{(sharp)}$$

Gillis, N., Vavasis, S.A.: Semidefinite programming based preconditioning for more robust near-separable nonnegative matrix factorization. SIAM Journal on Optimization 25(1), 677–698 (2015)

- Randomized/Smoothed variants (RandNMF, VCA): instead of looking for $\mathrm{argmax}_j ||M(:,j)||$ take a random $u \in \mathbb{R}^n$ and choose as vertex the average of the p columns of M corresponding to the p greatest entries of $u^\top M$

Successive Projection Algorithm (SPA)

Perturbation robustness: suppose $M = UV + N$ with UV separable, U full rank and each column of N with norm at most ε

- If $\varepsilon \leq \mathcal{O}(\sigma_r(U)/\mathcal{K}(U)^2)$ then SPA extract a matrix $\widetilde{U}$ such that

$$\max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)^2\right) \quad \text{(sharp for } r \geq 3\text{)}$$

Variants to improve robustness to perturbations and outliers:

- **SPA²:** Apply SPA to M to obtain U_1 and apply SPA to $U_1^\dagger M = (U_1^\dagger U)V + U_1^\dagger N$

$$\varepsilon \leq \mathcal{O}(\sigma_r(U)/\mathcal{K}(U)^2) \implies \max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)\right) \quad \text{(sharp)}$$

- **MVE:** Minimum Volume Ellipsoid $A \succ 0$ s.t. $m_j^\top A m_j \leq 1 \ \forall j$ and use $(A^{-1/2})^\dagger$

$$\varepsilon \leq \mathcal{O}(\sigma_r(U)/(r\sqrt{r})) \implies \max_{1 \leq k \leq r} \|U(:,k) - \widetilde{U}(:,k)\| \leq \mathcal{O}\left(\varepsilon \mathcal{K}(U)\right) \quad \text{(sharp)}$$

- Randomized/Smoothed variants (RandNMF, VCA): instead of looking for $\text{argmax}_j ||M(:,j)||$ take a random $u \in \mathbb{R}^n$ and choose as vertex the average of the p columns of M corresponding to the p greatest entries of $u^\top M$

Vu Thanh, O., Nadisic, N., Gillis, N.: Randomized successive projection algorithm, GRETSI (2022).

Nadisic, N., Gillis, N., Kervazo, C.: Smoothed separable nonnegative matrix factorization. Linear Algebra and its Applications 676, 174–204 (2023).

Rank Deficient case: SNPA

What if $M = UV$ is separable but U is rank deficient?

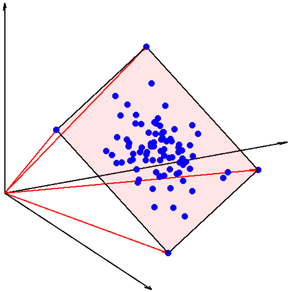
SNPA: Successive Nonnegative Projection Algorithm

Modify the projection step as

1: Project the original M on $\text{conv}(M(:,\mathcal{K}))$ to obtain M_p

2: Find $j^* = \text{argmax}_j ||M(:,j) - M_p(:,j)||$ and add it to $\mathcal{K}$

When $M_p = 0$, return $U = M(:,\mathcal{K})$



- ✓ Can handle the deficient rank case $\text{rk}(U) < r$
- × The bound on the error is $\mathcal{O}(\varepsilon \widetilde{\mathcal{K}}(U)^3)$
- ✓ If U is full rank, the error is the same as SPA and empirically it is more robust

Gillis, N.: Successive nonnegative projection algorithm for robust nonnegative blind source separation. SIAM Journal on Imaging Sciences 7(2), 1420–1450 (2014)

Rank Deficient case: SNPA

What if $M = UV$ is separable but U is rank deficient?

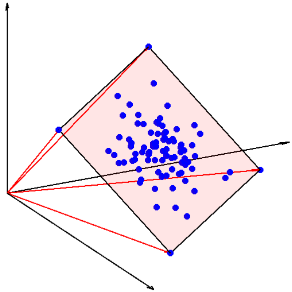
SNPA: Successive Nonnegative Projection Algorithm

Modify the projection step as

1: Project the original M on $\text{conv}(M(:,\mathcal{K}))$ to obtain M_p

2: Find $j^* = \text{argmax}_j ||M(:,j) - M_p(:,j)||$ and add it to $\mathcal{K}$

When $M_p = 0$, return $U = M(:,\mathcal{K})$



- ✓ Can handle the deficient rank case $\text{rk}(U) < r$
- × The bound on the error is $\mathcal{O}(\varepsilon \widetilde{\mathcal{K}}(U)^3)$
- ✓ If U is full rank, the error is the same as SPA and empirically it is more robust

Gillis, N.: Successive nonnegative projection algorithm for robust nonnegative blind source separation. SIAM Journal on Imaging Sciences 7(2), 1420–1450 (2014)

Rank Deficient case: SNPA

What if $M = UV$ is separable but U is rank deficient?

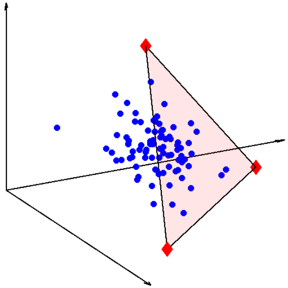
SNPA: Successive Nonnegative Projection Algorithm

Modify the projection step as

1: Project the original M on $\text{conv}(M(:,\mathcal{K}))$ to obtain M_p

2: Find $j^* = \text{argmax}_j ||M(:,j) - M_p(:,j)||$ and add it to $\mathcal{K}$

When $M_p = 0$, return $U = M(:,\mathcal{K})$



✓ Can handle the deficient rank case $\text{rk}(U) < r$

× The bound on the error is $\mathcal{O}(\varepsilon \widetilde{\mathcal{K}}(U)^3)$

✓ If U is full rank, the error is the same as SPA and empirically it is more robust

Gillis, N.: Successive nonnegative projection algorithm for robust nonnegative blind source separation. SIAM Journal on Imaging Sciences 7(2), 1420–1450 (2014)

Rank Deficient case: SNPA

What if $M = UV$ is separable but U is rank deficient?

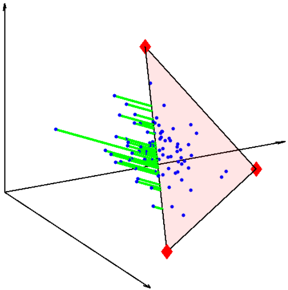
SNPA: Successive Nonnegative Projection Algorithm

Modify the projection step as

1: Project the original M on $\text{conv}(M(:,\mathcal{K}))$ to obtain M_p

2: Find $j^* = \text{argmax}_j ||M(:,j) - M_p(:,j)||$ and add it to $\mathcal{K}$

When $M_p = 0$, return $U = M(:,\mathcal{K})$



- ✓ Can handle the deficient rank case $\text{rk}(U) < r$
- × The bound on the error is $\mathcal{O}(\varepsilon \widetilde{\mathcal{K}}(U)^3)$
- ✓ If U is full rank, the error is the same as SPA and empirically it is more robust

Gillis, N.: Successive nonnegative projection algorithm for robust nonnegative blind source separation. SIAM Journal on Imaging Sciences 7(2), 1420–1450 (2014)

Rank Deficient case: SNPA

What if $M = UV$ is separable but U is rank deficient?

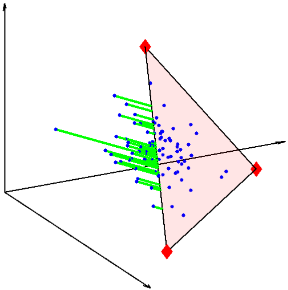
SNPA: Successive Nonnegative Projection Algorithm

Modify the projection step as

1: Project the original M on $\text{conv}(M(:,\mathcal{K}))$ to obtain M_p

2: Find $j^* = \text{argmax}_j ||M(:,j) - M_p(:,j)||$ and add it to $\mathcal{K}$

When $M_p = 0$, return $U = M(:,\mathcal{K})$



- ✓ Can handle the deficient rank case $\text{rk}(U) < r$
- × The bound on the error is $\mathcal{O}(\varepsilon \widetilde{\mathcal{K}}(U)^3)$
- ✓ If U is full rank, the error is the same as SPA and empirically it is more robust

Gillis, N.: Successive nonnegative projection algorithm for robust nonnegative blind source separation. SIAM Journal on Imaging Sciences 7(2), 1420–1450 (2014)

SSC and Minimum Volume

Sufficiently Scattered Condition

Separability leads to fast and robust algorithms, but it is a strong assumption

A column stochastic matrix V is sufficiently scattered if

SSC1: $\mathcal{C} := \{x \mid 1 = e^\top x \geq \sqrt{r-1}\|x\|\} \subseteq \text{conv}(V)$

SSC2: if Q is orthogonal and $\text{conv}(V) \subseteq \text{conv}(Q)$ then Q is a permutation matrix

Tl;dr: $\mathcal{C} \subseteq \text{conv}(V)$

Notice: Separability $\implies$ V contains I as submatrix $\implies$ $\mathcal{C} \subseteq \Delta = \text{conv}(V)$ $\implies$ SSC

Theorem

If $M = UV$ with V SSC, U full rank exists, then it is the **unique solution** to

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) : Conv(M) \subseteq Conv(U)$$

Notice2: SSC1 ensures the minimality, SSC2 ensures the uniqueness

Fu, X., Ma, W.K., Huang, K., Sidiropoulos, N.D.: Blind separation of quasi-stationary sources: exploiting convex geometry in covariance domain. IEEE Transactions on Signal Processing 63(9), 2306–2320 (2015)

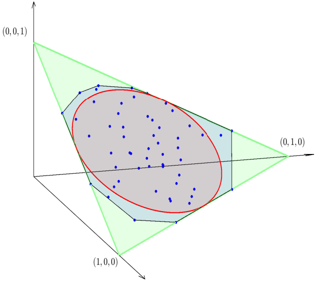
Sufficiently Scattered Condition

Separability leads to fast and robust algorithms, but it is a strong assumption

A column stochastic matrix V is sufficiently scattered if

SSC1: $\mathcal{C} := \{x \mid 1 = e^\top x \geq \sqrt{r-1}\|x\|\} \subseteq \text{conv}(V)$

SSC2: if Q is orthogonal and $\text{conv}(V) \subseteq \text{conv}(Q)$ then Q is a permutation matrix



Tl;dr: $\mathcal{C} \subseteq \text{conv}(V)$

Notice: Separability $\implies$ V contains I as submatrix $\implies$ $\mathcal{C} \subseteq \Delta = \text{conv}(V)$ $\implies$ SSC

Theorem

If $M = UV$ with V SSC, U full rank exists, then it is the **unique solution** to

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) : Conv(M) \subseteq Conv(U)$$

Notice2: SSC1 ensures the minimality, SSC2 ensures the uniqueness

Fu, X., Ma, W.K., Huang, K., Sidiropoulos, N.D.: Blind separation of quasi-stationary sources: exploiting convex geometry in covariance domain. IEEE Transactions on Signal Processing 63(9), 2306–2320 (2015)

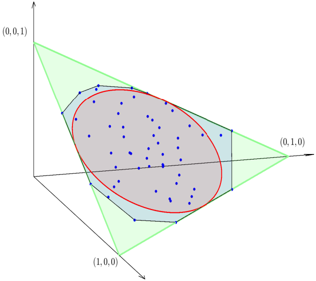
Sufficiently Scattered Condition

Separability leads to fast and robust algorithms, but it is a strong assumption

A column stochastic matrix V is sufficiently scattered if

SSC1: $\mathcal{C} := \{x \mid 1 = e^\top x \geq \sqrt{r-1}\|x\|\} \subseteq \text{conv}(V)$

SSC2: if Q is orthogonal and $\text{conv}(V) \subseteq \text{conv}(Q)$ then Q is a permutation matrix



Tl;dr: $\mathcal{C} \subseteq \text{conv}(V)$

Notice: Separability $\implies$ V contains I as submatrix $\implies$ $\mathcal{C} \subseteq \Delta = \text{conv}(V)$ $\implies$ SSC

Theorem

If $M = UV$ with V SSC, U full rank exists, then it is the **unique solution** to

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) : Conv(M) \subseteq Conv(U)$$

Notice2: SSC1 ensures the minimality, SSC2 ensures the uniqueness

Fu, X., Ma, W.K., Huang, K., Sidiropoulos, N.D.: Blind separation of quasi-stationary sources: exploiting convex geometry in covariance domain. IEEE Transactions on Signal Processing 63(9), 2306–2320 (2015)

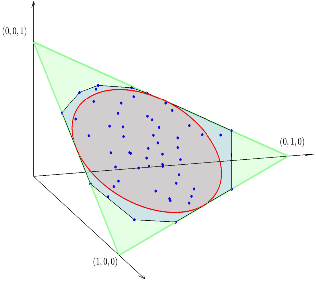
Sufficiently Scattered Condition

Separability leads to fast and robust algorithms, but it is a strong assumption

A column stochastic matrix V is sufficiently scattered if

SSC1: $\mathcal{C} := \{x \mid 1 = e^\top x \geq \sqrt{r-1}\|x\|\} \subseteq \mathrm{conv}(V)$

SSC2: if Q is orthogonal and $\mathrm{conv}(V) \subseteq \mathrm{conv}(Q)$ then Q is a permutation matrix



Tl;dr: $\mathcal{C} \subseteq \mathrm{conv}(V)$

Notice: Separability $\implies$ V contains I as submatrix $\implies$ $\mathcal{C} \subseteq \Delta = \mathrm{conv}(V)$ $\implies$ SSC

Theorem

If $M = UV$ with V SSC, U full rank exists, then it is the **unique solution** to

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) : Conv(M) \subseteq Conv(U)$$

Notice2: SSC1 ensures the minimality, SSC2 ensures the uniqueness

Fu, X., Ma, W.K., Huang, K., Sidiropoulos, N.D.: Blind separation of quasi-stationary sources: exploiting convex geometry in covariance domain. IEEE Transactions on Signal Processing 63(9), 2306–2320 (2015)

Sufficiently Scattered Condition

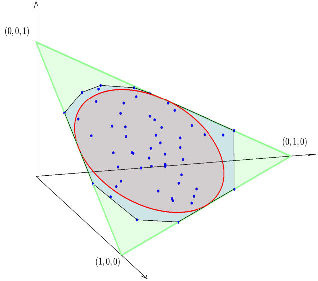
Separability leads to fast and robust algorithms, but it is a strong assumption

A column stochastic matrix V is sufficiently scattered if

SSC1: $\mathcal{C} := \{x \mid 1 = e^\top x \geq \sqrt{r-1}\|x\|\} \subseteq \text{conv}(V)$

SSC2: if Q is orthogonal and $\text{conv}(V) \subseteq \text{conv}(Q)$ then Q is a permutation matrix

Tl;dr: $\mathcal{C} \subseteq \text{conv}(V)$



Notice: Separability $\Longrightarrow$ V contains I as submatrix $\Longrightarrow$ $\mathcal{C} \subseteq \Delta = \text{conv}(V) \Longrightarrow$ SSC

Theorem

If $M = UV$ with V SSC, U full rank exists, then it is the **unique solution** to

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) : Conv(M) \subseteq Conv(U)$$

Notice2: SSC1 ensures the minimality, SSC2 ensures the uniqueness

Fu, X., Ma, W.K., Huang, K., Sidiropoulos, N.D.: Blind separation of quasi-stationary sources: exploiting convex geometry in covariance domain. IEEE Transactions on Signal Processing 63(9), 2306–2320 (2015)

Sufficiently Scattered Condition

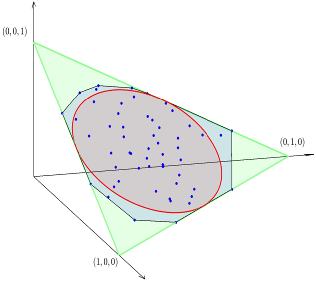
Separability leads to fast and robust algorithms, but it is a strong assumption

A column stochastic matrix V is sufficiently scattered if

SSC1: $\mathcal{C} := \{x \mid 1 = e^\top x \geq \sqrt{r-1}\|x\|\} \subseteq \text{conv}(V)$

SSC2: if Q is orthogonal and $\text{conv}(V) \subseteq \text{conv}(Q)$ then Q is a permutation matrix

Tl;dr: $\mathcal{C} \subseteq \text{conv}(V)$



Notice: Separability $\Longrightarrow$ V contains I as submatrix $\Longrightarrow$ $\mathcal{C} \subseteq \Delta = \text{conv}(V)$ $\Longrightarrow$ SSC

Theorem

If $M = UV$ with V SSC, U full rank exists, then it is the **unique solution** to

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) : Conv(M) \subseteq Conv(U)$$

Notice2: SSC1 ensures the minimality, SSC2 ensures the uniqueness

Fu, X., Ma, W.K., Huang, K., Sidiropoulos, N.D.: Blind separation of quasi-stationary sources: exploiting convex geometry in covariance domain. IEEE Transactions on Signal Processing 63(9), 2306–2320 (2015)

Simplex Volume Minimization

Exact Case:

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) \;:\; Conv(M) \subseteq Conv(U)$$

Inexact Case:

$$\min_{U,V} \|M - UV\|_F^2 + \lambda \log\det(U^\top U) \;:\; V \text{ column stochastic}$$

Alternating Method: Given $(\widetilde{U}, \widetilde{V})$ initial approximation,

Update of U

$$\log\det(A) \leq \langle B^{-1}, A \rangle + \log\det(B) - r$$

with $=$ iff $B = A \succ 0$

$$\|M - U\widetilde{V}\|_F^2 + \lambda \log\det(U^\top U) \leq \langle UU^\top, E \rangle - \langle U, C \rangle + b$$

where $E = \lambda(\widetilde{U}^\top \widetilde{U})^{-1} + \widetilde{V}\widetilde{V}^\top$, $C = 2M\widetilde{V}^T$ and b do not depend on U

$$\min_U \sum_i u_i^\top E u_i - c_i^\top u_i$$

are m quadratic and strongly convex optimization problems on the rows of U

Update of V

$$\|M - \widetilde{U}V\|_F^2 + \lambda \log\det(\widetilde{U}^\top \widetilde{U}) = \langle VV^\top, E \rangle - \langle V, C \rangle + b$$

where $E = \widetilde{W}^\top \widetilde{W}$, $C = 2\widetilde{U}^T M$ and b do not depend on V

$$\min_V \sum_i v_i^\top E v_i - c_i^\top v_i \;:\; V \text{ col. stoc.}$$

are n quadratic and strongly convex optimization problems on the columns of V, over convex domains (unit simplices)

Simplex Volume Minimization

Exact Case:

$$\min_{U\in\mathbb{R}^{m\times r}} Vol(U) \;:\; Conv(M) \subseteq Conv(U)$$

Inexact Case:

$$\min_{U,V} \|M - UV\|_F^2 + \lambda \log\det(U^\top U) \;:\; V \text{ column stochastic}$$

Leplat, V., Ang, A.M., Gillis, N.: Minimum-volume rank-deficient nonnegative matrix factoriza- tions. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3402–3406 (2019)

Alternating Method: Given $(\widetilde{U}, \widetilde{V})$ initial approximation,

Update of U

$$\log\det(A) \leq \langle B^{-1}, A\rangle + \log\det(B) - r$$

with $=$ iff $B = A \succ 0$

$$\|M - U\widetilde{V}\|_F^2 + \lambda \log\det(U^\top U) \leq \langle UU^\top, E\rangle - \langle U, C\rangle + b$$

where $E = \lambda(\widetilde{U}^\top \widetilde{U})^{-1} + \widetilde{V}\widetilde{V}^\top$, $C = 2M\widetilde{V}^T$ and b do not depend on U

$$\min_U \sum_i u_i^\top E u_i - c_i^\top u_i$$

are m quadratic and strongly convex

Update of V

$$\|M - \widetilde{U}V\|_F^2 + \lambda \log\det(\widetilde{U}^\top \widetilde{U}) = \langle VV^\top, E\rangle - \langle V, C\rangle + b$$

where $E = \widetilde{W}^\top \widetilde{W}$, $C = 2\widetilde{U}^T M$ and b do not depend on V

$$\min_V \sum_i v_i^\top E v_i - c_i^\top v_i \;:\; V \text{ col. stoc.}$$

are n quadratic and strongly convex optimization problems on the columns of V,

Simplex Volume Minimization

Exact Case:

$$\min_{U\in\mathbb{R}^{m\times r}} Vol(U) \;:\; Conv(M) \subseteq Conv(U)$$

Inexact Case:

$$\min_{U,V} \|M - UV\|_F^2 + \lambda \log\det(U^\top U) \;:\; V \text{ column stochastic}$$

Alternating Method: Given $(\widetilde{U}, \widetilde{V})$ initial approximation,

Update of U

$$\log\det(A) \le \langle B^{-1}, A\rangle + \log\det(B) - r$$

with $=$ iff $B = A \succ 0$

$$\|M - U\widetilde{V}\|_F^2 + \lambda \log\det(U^\top U) \le \langle UU^\top, E\rangle - \langle U, C\rangle + b$$

where $E = \lambda(\widetilde{U}^\top \widetilde{U})^{-1} + \widetilde{V}\widetilde{V}^\top$, $C = 2M\widetilde{V}^T$ and b do not depend on U

$$\min_U \sum_i u_i^\top E u_i - c_i^\top u_i$$

are m quadratic and strongly convex optimization problems on the rows of U

Update of V

$$\|M - \widetilde{U}V\|_F^2 + \lambda \log\det(\widetilde{U}^\top \widetilde{U}) = \langle VV^\top, E\rangle - \langle V, C\rangle + b$$

where $E = \widetilde{W}^\top \widetilde{W}$, $C = 2\widetilde{U}^T M$ and b do not depend on V

$$\min_V \sum_i v_i^\top E v_i - c_i^\top v_i \;:\; V \text{ col. stoc.}$$

are n quadratic and strongly convex optimization problems on the columns of V, over convex domains (unit simplices)

Simplex Volume Minimization

Exact Case:

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) \ : \ Conv(M) \subseteq Conv(U)$$

Inexact Case:

$$\min_{U,V} \|M - UV\|_F^2 + \lambda \log\det(U^\top U) \ : \ V \text{ column stochastic}$$

Alternating Method: Given $(\widetilde{U}, \widetilde{V})$ initial approximation,

Update of U

$$\log\det(A) \leq \langle B^{-1}, A \rangle + \log\det(B) - r$$

with $=$ iff $B = A \succ 0$

$$\|M - U\widetilde{V}\|_F^2 + \lambda \log\det(U^\top U) \leq \langle UU^\top, E \rangle - \langle U, C \rangle + b$$

where $E = \lambda(\widetilde{U}^\top \widetilde{U})^{-1} + \widetilde{V}\widetilde{V}^\top$, $C = 2M\widetilde{V}^T$ and b do not depend on U

$$\min_U \sum_i u_i^\top E u_i - c_i^\top u_i$$

are m quadratic and strongly convex optimization problems on the rows of U

Update of V

$$\|M - \widetilde{U}V\|_F^2 + \lambda \log\det(\widetilde{U}^\top \widetilde{U}) = \langle VV^\top, E \rangle - \langle V, C \rangle + b$$

where $E = \widetilde{W}^\top \widetilde{W}$, $C = 2\widetilde{U}^T M$ and b do not depend on V

$$\min_V \sum_i v_i^\top E v_i - c_i^\top v_i \ : \ V \text{ col. stoc.}$$

are n quadratic and strongly convex optimization problems on the columns of V, over convex domains (unit simplices)

Simplex Volume Minimization

Exact Case:

$$\min_{U\in\mathbb{R}^{m\times r}} Vol(U) \;:\; Conv(M) \subseteq Conv(U)$$

Inexact Case:

$$\min_{U,V} \|M - UV\|_F^2 + \lambda \log\det(U^\top U) \;:\; V \text{ column stochastic}$$

Alternating Method: Given $(\widetilde{U}, \widetilde{V})$ initial approximation,

Update of U

$$\log\det(A) \leq \langle B^{-1}, A\rangle + \log\det(B) - r$$

with $=$ iff $B = A \succ 0$

$$\|M - U\widetilde{V}\|_F^2 + \lambda \log\det(U^\top U) \leq \langle UU^\top, E\rangle - \langle U, C\rangle + b$$

where $E = \lambda(\widetilde{U}^\top \widetilde{U})^{-1} + \widetilde{V}\widetilde{V}^\top$, $C = 2M\widetilde{V}^T$ and b do not depend on U

$$\min_U \sum_i u_i^\top E u_i - c_i^\top u_i$$

are m quadratic and strongly convex optimization problems on the rows of U

Update of V

$$\|M - \widetilde{U}V\|_F^2 + \lambda \log\det(\widetilde{U}^\top \widetilde{U}) = \langle VV^\top, E\rangle - \langle V, C\rangle + b$$

where $E = \widetilde{W}^\top \widetilde{W}$, $C = 2\widetilde{U}^T M$ and b do not depend on V

$$\min_V \sum_i v_i^\top E v_i - c_i^\top v_i \;:\; V \text{ col. stoc.}$$

are n quadratic and strongly convex optimization problems on the columns of V, over convex domains (unit simplices)

Simplex Volume Minimization

Exact Case:

$$\min_{U \in \mathbb{R}^{m \times r}} Vol(U) \ : \ Conv(M) \subseteq Conv(U)$$

Inexact Case:

$$\min_{U,V} \|M - UV\|_F^2 + \lambda \log\det(U^\top U) \ : \ V \text{ column stochastic}$$

Alternating Method: Given $(\widetilde{U}, \widetilde{V})$ initial approximation,

Update of U

$$\log\det(A) \leq \langle B^{-1}, A \rangle + \log\det(B) - r$$

with $=$ iff $B = A \succ 0$

$$\|M - U\widetilde{V}\|_F^2 + \lambda \log\det(U^\top U) \leq \langle UU^\top, E \rangle - \langle U, C \rangle + b$$

where $E = \lambda(\widetilde{U}^\top \widetilde{U})^{-1} + \widetilde{V}\widetilde{V}^\top$, $C = 2M\widetilde{V}^T$ and b do not depend on U

$$\min_U \sum_i u_i^\top E u_i - c_i^\top u_i$$

are m quadratic and strongly convex optimization problems on the rows of U

Update of V

$$\|M - \widetilde{U}V\|_F^2 + \lambda \log\det(\widetilde{U}^\top \widetilde{U}) = \langle VV^\top, E \rangle - \langle V, C \rangle + b$$

where $E = \widetilde{W}^\top \widetilde{W}$, $C = 2\widetilde{U}^T M$ and b do not depend on V

$$\min_V \sum_i v_i^\top E v_i - c_i^\top v_i \ : \ V \text{ col. stoc.}$$

are n quadratic and strongly convex optimization problems on the columns of V, over convex domains (unit simplices)

Simplex Volume Minimization

Exact Case:

$$\min_{U\in\mathbb{R}^{m\times r}} Vol(U) \;:\; Conv(M) \subseteq Conv(U)$$

Inexact Case:

$$\min_{U,V} \|M - UV\|_F^2 + \lambda \log\det(U^\top U) \;:\; V \text{ column stochastic}$$

Alternating Method: Given $(\widetilde{U}, \widetilde{V})$ initial approximation,

Update of U

$$\log\det(A) \leq \langle B^{-1}, A\rangle + \log\det(B) - r$$

with $=$ iff $B = A \succ 0$

$$\|M - U\widetilde{V}\|_F^2 + \lambda \log\det(U^\top U) \leq \langle UU^\top, E\rangle - \langle U, C\rangle + b$$

where $E = \lambda(\widetilde{U}^\top \widetilde{U})^{-1} + \widetilde{V}\widetilde{V}^\top$, $C = 2M\widetilde{V}^T$ and b do not depend on U

$$\min_U \sum_i u_i^\top E u_i - c_i^\top u_i$$

are m quadratic and strongly convex optimization problems on the rows of U

Update of V

$$\|M - \widetilde{U}V\|_F^2 + \lambda \log\det(\widetilde{U}^\top \widetilde{U}) = \langle VV^\top, E\rangle - \langle V, C\rangle + b$$

where $E = \widetilde{W}^\top \widetilde{W}$, $C = 2\widetilde{U}^T M$ and b do not depend on V

$$\min_V \sum_i v_i^\top E v_i - c_i^\top v_i \;:\; V \text{ col. stoc.}$$

are n quadratic and strongly convex optimization problems on the columns of V, over convex domains (unit simplices)

Facet Identification

PCA Preprocessing: Given $\widetilde{M} = \widetilde{U}V \in \mathbb{R}^{m\times n}$ and U of rank r we can always reduce to

$$Q(\widetilde{M} - ze^\top) = Q(\widetilde{U} - ze^\top)V \in \mathbb{R}^{(r-1)\times n}$$

In other words, we have to find a simplex: r vertices in $r-1$ dimensions

$$Conv(U) = \cap_{i=1}^{r}\mathcal{S}_i \quad \text{where} \quad \mathcal{S}_i := \{x : \theta_i^T x \leq 1\}$$

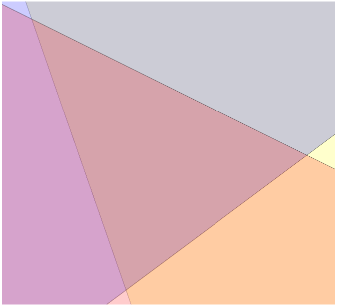
$$Conv(M) \subseteq Conv(U) \qquad \Longleftrightarrow \qquad \Theta = \begin{pmatrix}\theta_1 & \dots & \theta_r\end{pmatrix} \qquad \Theta^\top M \leq 1$$

MVIE *Maximum Volume Inscribed Ellipsoid*

Enumerates the facets of $Conv(M)$, very expensive

GFPI *Greedy Facet-based Polytope Identification*

Mixed integer programming, also expensive



In order to deal with facets GFPI works in the **Polar Space**

Facet Based Algorithms

PCA Preprocessing: Given $\widetilde{M} = \widetilde{U}V \in \mathbb{R}^{m \times n}$ and U of rank r we can always reduce to

$$Q(\widetilde{M} - ze^\top) = Q(\widetilde{U} - ze^\top)V \in \mathbb{R}^{(r-1)\times n}$$

In other words, we have to find a simplex: r vertices in $r-1$ dimensions

$$Conv(U) = \cap_{i=1}^{r} \mathcal{S}_i \quad \text{where} \quad \mathcal{S}_i := \{x : \theta_i^T x \leq 1\}$$

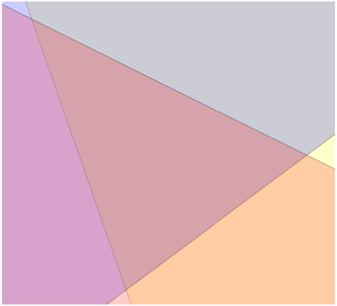
$$Conv(M) \subseteq Conv(U) \quad \Longleftrightarrow \quad \Theta = \begin{pmatrix} \theta_1 & \dots & \theta_r \end{pmatrix} \quad \Theta^\top M \leq 1$$

MVIE *Maximum Volume Inscribed Ellipsoid*

Enumerates the facets of $Conv(M)$, very expensive

GFPI *Greedy Facet-based Polytope Identification*

Mixed integer programming, also expensive



In order to deal with facets GFPI works in the **Polar Space**

Facet Based Algorithms

PCA Preprocessing: Given $\widetilde{M} = \widetilde{U}V \in \mathbb{R}^{m \times n}$ and U of rank r we can always reduce to

$$Q(\widetilde{M} - ze^\top) = Q(\widetilde{U} - ze^\top)V \in \mathbb{R}^{(r-1)\times n}$$

In other words, we have to find a simplex: r vertices in $r-1$ dimensions

$$Conv(U) = \cap_{i=1}^{r} \mathcal{S}_i \quad \text{where} \quad \mathcal{S}_i := \{x : \theta_i^T x \le 1\}$$

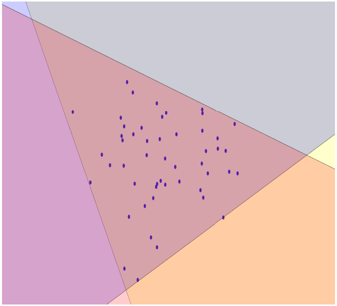
$$Conv(M) \subseteq Conv(U) \quad \Longleftrightarrow \quad \Theta = \begin{pmatrix} \theta_1 & \dots & \theta_r \end{pmatrix} \quad \Theta^\top M \le 1$$

MVIE *Maximum Volume Inscribed Ellipsoid*

Enumerates the facets of $Conv(M)$, very expensive

GFPI *Greedy Facet-based Polytope Identification*

Mixed integer programming, also expensive



In order to deal with facets GFPI works in the **Polar Space**

Facet Based Algorithms

PCA Preprocessing: Given $\widetilde{M} = \widetilde{U}V \in \mathbb{R}^{m \times n}$ and U of rank r we can always reduce to

$$Q(\widetilde{M} - ze^\top) = Q(\widetilde{U} - ze^\top)V \in \mathbb{R}^{(r-1)\times n}$$

In other words, we have to find a simplex: r vertices in $r-1$ dimensions

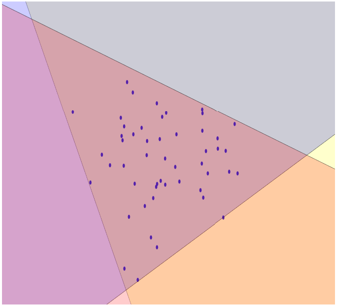
$$Conv(U) = \cap_{i=1}^{r} \mathcal{S}_i \quad \text{where} \quad \mathcal{S}_i := \{x : \theta_i^T x \le 1\}$$

$$Conv(M) \subseteq Conv(U) \qquad \iff \qquad \Theta = \begin{pmatrix}\theta_1 & \dots & \theta_r\end{pmatrix} \qquad \Theta^\top M \le 1$$

MVIE *Maximum Volume Inscribed Ellipsoid*

Enumerates the facets of $Conv(M)$, very expensive

Lin, C.H., Wu, R., Ma, W.K., Chi, C.Y., Wang, Y.: Maximum volume inscribed ellipsoid: A new simplex- structured matrix factorization framework via facet enumeration and convex optimization. SIAM Journal on Imaging Sciences 11 (2018)



GFPI *Greedy Facet-based Polytope Identification*

Mixed integer programming, also expensive

In order to deal with facets GFPI works in the **Polar Space**

Facet Based Algorithms

PCA Preprocessing: Given $\widetilde{M} = \widetilde{U}V \in \mathbb{R}^{m \times n}$ and U of rank r we can always reduce to

$$Q(\widetilde{M} - ze^\top) = Q(\widetilde{U} - ze^\top)V \in \mathbb{R}^{(r-1)\times n}$$

In other words, we have to find a simplex: r vertices in $r-1$ dimensions

$$Conv(U) = \cap_{i=1}^{r} \mathcal{S}_i \quad \text{where} \quad \mathcal{S}_i := \{x : \theta_i^T x \leq 1\}$$

$$Conv(M) \subseteq Conv(U) \qquad \Longleftrightarrow \qquad \Theta = \begin{pmatrix} \theta_1 & \dots & \theta_r \end{pmatrix} \qquad \Theta^\top M \leq 1$$

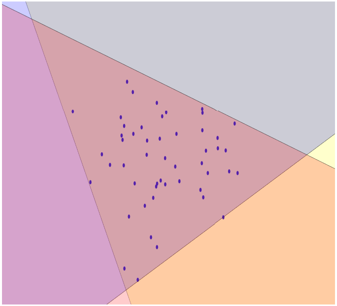
MVIE *Maximum Volume Inscribed Ellipsoid*

Enumerates the facets of $Conv(M)$, very expensive

GFPI *Greedy Facet-based Polytope Identification*

Mixed integer programming, also expensive

Abdolali, M., Gillis, N.: Simplex-structured matrix factorization: Sparsity-based identifiability and provably correct algorithms. SIAM Journal on Mathematics of Data Science 3(2), 593–623 (2021)



Facet Based Algorithms

PCA Preprocessing: Given $\widetilde{M} = \widetilde{U}V \in \mathbb{R}^{m \times n}$ and U of rank r we can always reduce to

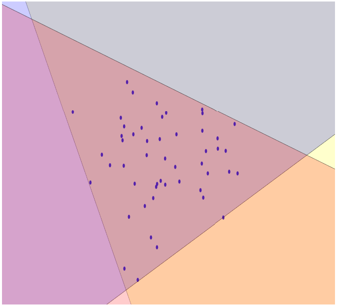
$$Q(\widetilde{M} - ze^\top) = Q(\widetilde{U} - ze^\top)V \in \mathbb{R}^{(r-1)\times n}$$

In other words, we have to find a simplex: r vertices in $r-1$ dimensions

$$Conv(U) = \cap_{i=1}^r \mathcal{S}_i \quad \text{where} \quad \mathcal{S}_i := \{x : \theta_i^T x \le 1\}$$

$$Conv(M) \subseteq Conv(U) \qquad \Longleftrightarrow \qquad \Theta = \begin{pmatrix}\theta_1 & \dots & \theta_r\end{pmatrix} \qquad \Theta^\top M \le 1$$

MVIE *Maximum Volume Inscribed Ellipsoid*
Enumerates the facets of $Conv(M)$, very expensive



GFPI *Greedy Facet-based Polytope Identification*
Mixed integer programming, also expensive

In order to deal with facets GFPI works in the **Polar Space**

Polarity

$$\mathcal{S} \subseteq \mathbb{R}^{r-1} \qquad \mathcal{S}^* := \{\theta : \theta^T x \leq 1 \; \forall x \in \mathcal{S}\}$$

- Swaps points and hyperplanes

$$\{x : \theta^T x = 1\} \rightsquigarrow \theta$$

- Sends simplexes into simplexes
- It is an involution for convex sets
- **Reverses Containments**

$$Conv(M) \subseteq Conv(U) \iff Conv(U)^* \subseteq Conv(M)^*$$

$$\iff \Theta^T M \leq 1 \quad \text{where} \quad Conv(U)^* = Conv(\Theta)$$

We can thus seek the simplex Θ with **maximum volume** inside $Conv(M)^*$ as in

$$\max_{\theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T M \leq 1 \qquad (MaxVol)$$

Polarity

$$\mathcal{S} \subseteq \mathbb{R}^{r-1} \qquad \mathcal{S}^* := \{\theta : \theta^T x \leq 1 \ \forall x \in \mathcal{S}\}$$

- Swaps points and hyperplanes

$$\{x : \theta^T x = 1\} \rightsquigarrow \theta$$

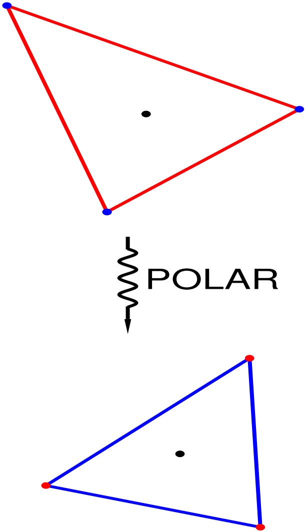
- Sends simplexes into simplexes
- It is an involution for convex sets
- **Reverses Containments**

$$Conv(M) \subseteq Conv(U) \iff Conv(U)^* \subseteq Conv(M)^*$$

$$\iff \Theta^T M \leq 1 \quad \text{where} \quad Conv(U)^* = Conv(\Theta)$$

We can thus seek the simplex Θ with **maximum volume** inside $Conv(M)^*$ as in

$$\max_{\theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T M \leq 1 \qquad (MaxVol)$$



Polarity

$$\mathcal{S} \subseteq \mathbb{R}^{r-1} \qquad \mathcal{S}^* := \{\theta : \theta^T x \leq 1 \ \forall x \in \mathcal{S}\}$$

- Swaps points and hyperplanes

$$\{x : \theta^T x = 1\} \rightsquigarrow \theta$$

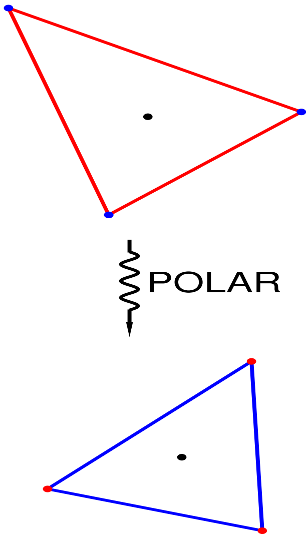
- Sends simplexes into simplexes
- It is an involution for convex sets
- **Reverses Containments**

$$Conv(M) \subseteq Conv(U) \iff Conv(U)^* \subseteq Conv(M)^*$$

$$\iff \Theta^T M \leq 1 \quad \text{where} \quad Conv(U)^* = Conv(\Theta)$$

We can thus seek the simplex Θ with **maximum volume** inside $Conv(M)^*$ as in

$$\max_{\theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T M \leq 1 \qquad (MaxVol)$$



Polarity

$$\mathcal{S} \subseteq \mathbb{R}^{r-1} \qquad \mathcal{S}^* := \{\theta : \theta^T x \leq 1 \ \forall x \in \mathcal{S}\}$$

- Swaps points and hyperplanes

$$\{x : \theta^T x = 1\} \rightsquigarrow \theta$$

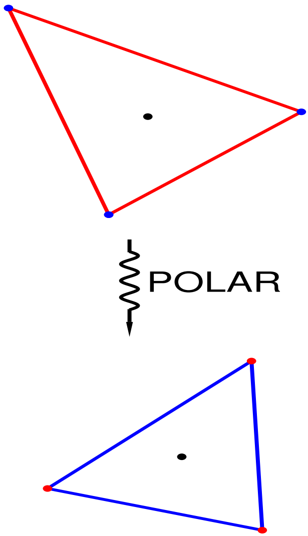
- Sends simplexes into simplexes
- It is an involution for convex sets
- **Reverses Containments**

$$Conv(M) \subseteq Conv(U) \iff Conv(U)^* \subseteq Conv(M)^*$$

$$\iff \Theta^T M \leq 1 \quad \text{where} \quad Conv(U)^* = Conv(\Theta)$$

We can thus seek the simplex Θ with **maximum volume** inside $Conv(M)^*$ as in

$$\max_{\theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T M \leq 1 \qquad (MaxVol)$$



Polarity

$$\mathcal{S} \subseteq \mathbb{R}^{r-1} \qquad \mathcal{S}^* := \{\theta : \theta^T x \leq 1 \; \forall x \in \mathcal{S}\}$$

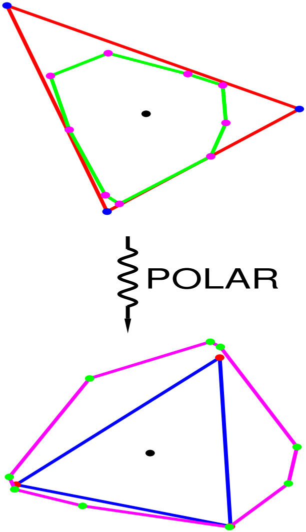
- Swaps points and hyperplanes

$$\{x : \theta^T x = 1\} \rightsquigarrow \theta$$

- Sends simplexes into simplexes
- It is an involution for convex sets
- **Reverses Containments**

$$Conv(M) \subseteq Conv(U) \iff Conv(U)^* \subseteq Conv(M)^*$$

$$\iff \Theta^\top M \leq 1 \quad \text{where} \quad Conv(U)^* = Conv(\Theta)$$



We can thus seek the simplex Θ with **maximum volume** inside $Conv(M)^*$ as in

$$\max_{\theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T M \leq 1 \qquad (MaxVol)$$

Polarity

$$\mathcal{S} \subseteq \mathbb{R}^{r-1} \qquad \mathcal{S}^* := \{\theta : \theta^T x \leq 1 \ \forall x \in \mathcal{S}\}$$

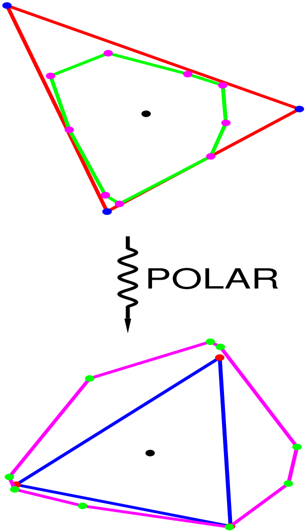
- Swaps points and hyperplanes

$$\{x : \theta^T x = 1\} \rightsquigarrow \theta$$

- Sends simplexes into simplexes
- It is an involution for convex sets
- **Reverses Containments**

$$Conv(M) \subseteq Conv(U) \iff Conv(U)^* \subseteq Conv(M)^*$$

$$\iff \Theta^\top M \leq 1 \quad \text{where} \quad Conv(U)^* = Conv(\Theta)$$



We can thus seek the simplex Θ with **maximum volume** inside $Conv(M)^*$ as in

$$\max_{\theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T M \leq 1 \qquad (MaxVol)$$

Identifiability and η-Expansion

Theorem (M.A., G.B., N.G., 2023)

Let $M = UV \in \mathbb{R}^{r-\mathbf{1} \times n}$ SSC and for any $u \in \mathbb{R}^{r-\mathbf{1}}$ define

$$\mathcal{V}(u) := \max_{\Theta \in \mathbb{R}^{r-\mathbf{1} \times r}} Vol(\Theta) \quad : \quad \Theta^T(M - ue^T) \leq 1$$

Then $\mathcal{V}(u)$ is convex in u with unique minimum for $u = Ue/r$ and Θ polar of U

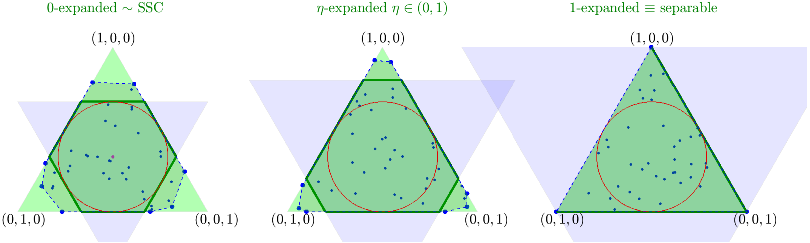
Identifiability and η-Expansion

Theorem (M.A., G.B., N.G., 2023)

Let $M = UV \in \mathbb{R}^{r-1 \times n}$ SSC and for any $u \in \mathbb{R}^{r-1}$ define

$$\mathcal{V}(u) := \max_{\Theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T (M - ue^T) \leq 1$$

Then $\mathcal{V}(u)$ is convex in u with unique minimum for $u = Ue/r$ and Θ polar of U



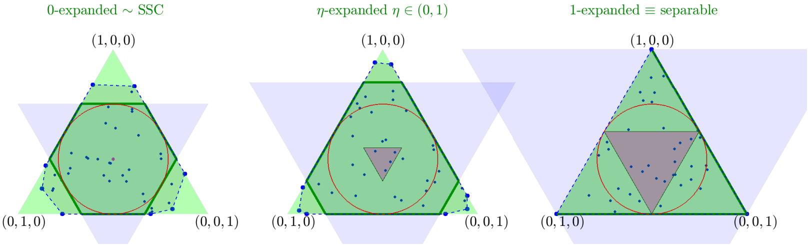
Identifiability and η-Expansion

Theorem (M.A., G.B., N.G., 2023)

Let $M = UV \in \mathbb{R}^{r-1 \times n}$ SSC and for any $u \in \mathbb{R}^{r-1}$ define

$$\mathcal{V}(u) := \max_{\Theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T(M - ue^T) \leq 1$$

Then $\mathcal{V}(u)$ is convex in u with unique minimum for $u = Ue/r$ and Θ polar of U



Theorem (M.A., G.B., N.G., 2023)

Let $M = UV \in \mathbb{R}^{r-1 \times n}$ be η-expanded and suppose $u = Uv$, $v \in \blacktriangledown$. Then

$$\max_{\Theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T(X - ue^T) \leq 1$$

is solved uniquely by Θ polar of U

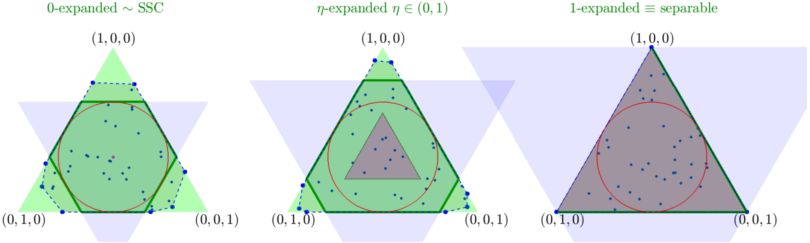
Identifiability and η-Expansion

Theorem (M.A., G.B., N.G., 2023)

Let $M = UV \in \mathbb{R}^{r-1 \times n}$ SSC and for any $u \in \mathbb{R}^{r-1}$ define

$$\mathcal{V}(u) := \max_{\Theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T(M - ue^T) \leq 1$$

Then $\mathcal{V}(u)$ is convex in u with unique minimum for $u = Ue/r$ and Θ polar of U



Conjecture (M.A., G.B., N.G., 2023)

Let $M = UV \in \mathbb{R}^{r-1 \times n}$ be η-expanded and suppose $u = Uv$, $v \in \blacktriangle$. Then

$$\max_{\Theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) \quad : \quad \Theta^T(M - ue^T) \leq 1$$

is solved uniquely by Θ polar of U

Maximum Volume in Dual

Algorithm Maximum Volume in the Dual (MV-Dual)

Input: Data matrix $\widetilde{M} \in \mathbb{R}^{m \times n}$ and a factorization rank r

Output: A matrix $\widetilde{U} \in \mathbb{R}^{m \times r}$ and a vector z such that $\widetilde{M} \approx z + \widetilde{U}V$ where V is column stochastic

1: Use PCA to reduce $\widetilde{M} = z + QM$ with $M \in \mathbb{R}^{r-\mathbf{1} \times n}$

2: Initialize $u_1 = Me/n$, $p = 1$ and $\Theta \in \mathcal{N}(0,1)^{r-\mathbf{1} \times r}$

3: **while** not converged: $p = 1$ or $\frac{\|u_p - u_{p-\mathbf{1}}\|_\mathbf{2}}{\|u_{p-\mathbf{1}}\|_\mathbf{2}} > 0.01$ **do**

4: Solve

$$\arg \max_{\Theta \in \mathbb{R}^{r-\mathbf{1} \times r}} Vol(\Theta) : \Theta^T (X - v_p e^T) \leq 1$$

via alternating optimization on the columns of Θ

5: Recover U by computing the polar of $Conv\,(\Theta)$

6: Let $u_{p+1} \leftarrow Ue/r$, and $p = p + 1$

7: **end while**

8: Compute $\widetilde{U} = QU$

Cost : PCA $\mathcal{O}(mnr)$ plus Maximization problem solver for a single column $\mathcal{O}(nr^2)$ times the number of iterations

Maximum Volume in Dual

Algorithm Maximum Volume in the Dual (MV-Dual)

Input: Data matrix $\widetilde{M} \in \mathbb{R}^{m \times n}$ and a factorization rank r

Output: A matrix $\widetilde{U} \in \mathbb{R}^{m \times r}$ and a vector z such that $\widetilde{M} \approx z + \widetilde{U}V$ where V is column stochastic

1: Use PCA to reduce $\widetilde{M} = z + QM$ with $M \in \mathbb{R}^{r-1 \times n}$

2: Initialize $u_1 = Me/n$, $p = 1$ and $\Theta \in \mathcal{N}(0,1)^{r-1 \times r}$

3: **while** not converged: $p = 1$ or $\frac{\|u_p - u_{p-1}\|_2}{\|u_{p-1}\|_2} > 0.01$ **do**

4: Solve

$$\arg \max_{\Theta \in \mathbb{R}^{r-1 \times r}} Vol(\Theta) : \Theta^T (X - v_p e^T) \leq 1$$

via alternating optimization on the columns of Θ

5: Recover U by computing the polar of $Conv(\Theta)$

6: Let $u_{p+1} \leftarrow Ue/r$, and $p = p + 1$

7: **end while**

8: Compute $\widetilde{U} = QU$

Cost : PCA $\mathcal{O}(mnr)$ plus Maximization problem solver for a single column $\mathcal{O}(nr^2)$ times the number of iterations

Maximum Volume in Dual

Algorithm Maximum Volume in the Dual (MV-Dual)

Input: Data matrix $\widetilde{M} \in \mathbb{R}^{m \times n}$ and a factorization rank r

Output: A matrix $\widetilde{U} \in \mathbb{R}^{m \times r}$ and a vector z such that $\widetilde{M} \approx z + \widetilde{U}V$ where V is column stochastic

Use PCA to reduce $\widetilde{M} = z + QM$ with $M \in \mathbb{R}^{r-1 \times n}$

Initialize $u_1 = Me/n$, $p = 1$ and $\Theta \in \mathcal{N}(0,1)^{r-1 \times r}$

while not converged: $p = 1$ or $\frac{\|u_p - u_{p-1}\|_2}{\|u_{p-1}\|_2} > 0.01$ **do**

    Solve

$$\arg \max_{\Theta \in \mathbb{R}^{r-1 \times r}, \Delta \in \mathbb{R}^{r \times n}} Vol(\Theta)^2 - \lambda \|\Delta\|_F^2 : \Theta^T (X - v_p e^T) \leq 1 + \Delta^T$$

    via alternating optimization on the columns of Θ, Δ

    Recover U by computing the polar of $Conv(\Theta)$

    Let $u_{p+1} \leftarrow Ue/r$, and $p = p + 1$

end while

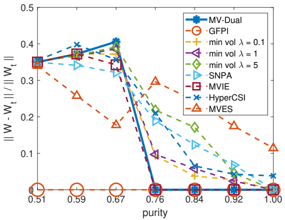
Compute $\widetilde{U} = QU$

Cost : PCA $\mathcal{O}(mnr)$ plus Maximization problem solver for a single column $\mathcal{O}(nr^2)$ times the number of iterations

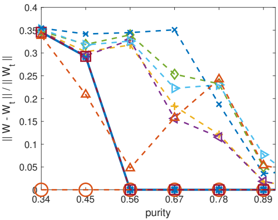
Experiments

Exact Case

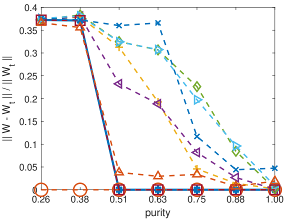
U^*, V^* ground truth $\quad ERR = \frac{\|U^* - U\|_F}{\|U^*\|_F} \quad$ purity $p = \max_{i,j} |V^*_{i,j}| = \eta + (1-\eta)\frac{2}{r}$



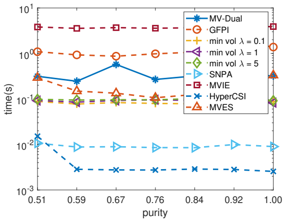
ERR for $r = 3,\ n = 30r$



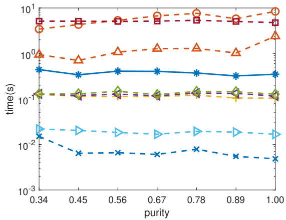
ERR for $r = 4,\ n = 30r$



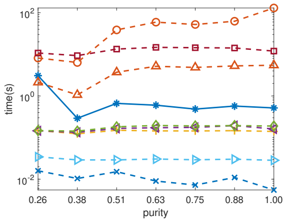
ERR for $r = 5,\ n = 30r$



Time for $r = 3,\ n = 30r$



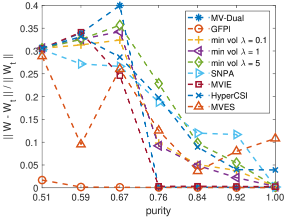
Time for $r = 4,\ n = 30r$



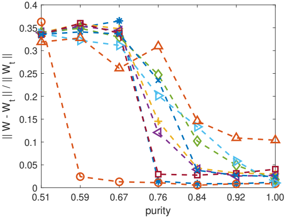
Time for $r = 5,\ n = 30r$

Noisy Case

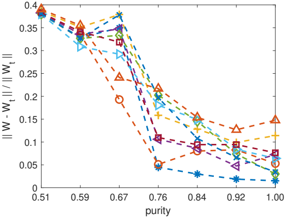
U^*, V^* ground truth $\quad ERR = \frac{\|U^* - U\|_F}{\|U^*\|_F} \quad$ purity $p = \max_{i,j} |V^*_{i,j}| = \eta + (1-\eta)\frac{2}{r}$



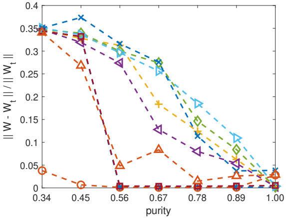
ERR for $r = 3$, SNR $= 60$



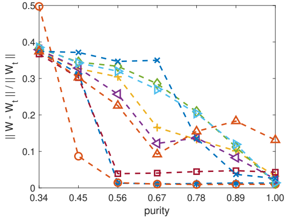
ERR for $r = 3$, SNR $= 40$



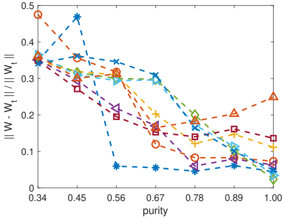
ERR for $r = 3$, SNR $= 30$



ERR for $r = 4$, SNR $= 60$



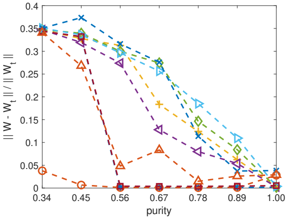
ERR for $r = 4$, SNR $= 40$



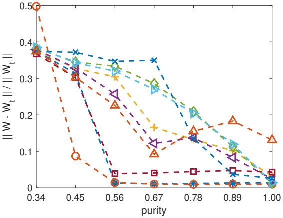
ERR for $r = 4$, SNR $= 30$

Noisy Case

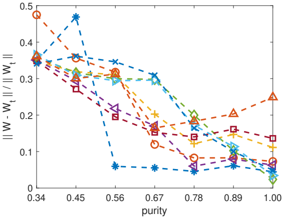
U^*, V^* ground truth $\quad ERR = \frac{\| U^* - U \|_F}{\| U^* \|_F} \quad$ purity $p = \max_{i,j} |V^*_{i,j}| = \eta + (1-\eta)\frac{2}{r}$



ERR for $r = 4$, SNR $= 60$



ERR for $r = 4$, SNR $= 40$

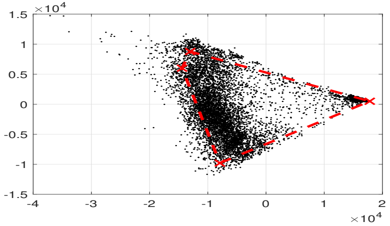


ERR for $r = 4$, SNR $= 30$

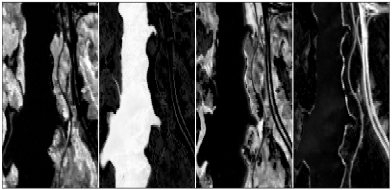
SNR	MVDual	GFPI	min vol $\lambda = 0.1$	min vol $\lambda = 1$	min vol $\lambda = 5$	SNPA	MVIE	HyperCSI	MVES
30	0.56±0.11	7.76±3.51	0.12±0.01	0.13±0.01	0.14±0.02	0.01±0.001	5.28±0.23	0.01±0.004	0.30±0.04
40	0.45±0.06	4.18±1.12	0.10±0.01	0.11±0.01	0.13±0.01	0.01±0.00	4.96±0.12	0.005±0.004	0.30±0.05
60	0.42±0.06	1.47±0.45	0.07±0.01	0.08±0.01	0.09±0.01	0.01±0.00	3.78±0.12	0.001±0.00	0.26±0.07

Unmixing Hyperspectral Imaging

$$\mathrm{MRSA}(x, y) = \frac{100}{\pi} \cos^{-1}\left(\frac{(x-\bar{x}e)^\top (y-\bar{y}e)}{\|x-\bar{x}e\|_2 \|y-\bar{y}e\|_2}\right) \qquad ERR = \sum_k \mathrm{MRSA}(U_k^*, U_k)$$



Projection of data points
and the symplex computed by MV-Dual



Abundance maps estimated by MV-Dual
From left to right: road, tree, soil, water

	SNPA	Min-Vol	HyperCSI	GFPI	MV-Dual
MRSA	22.27	6.03	17.04	4.82	3.74
Time (s)	0.60	1.45	0.88	100*	43.51

Comparing the performances of MV-Dual with the state-of-the-art SSMF algorithms on Jasper-Ridge data set. Numbers marked with * indicate that the corresponding algorithms did not converge within 100 seconds.

Thank You!

Abdolali M., Barbarino G., and Gillis N. **Dual simplex volume maximization for simplex-structured matrix factorization.** *SIAM Journal of Scientific Imaging*, 2024.

Nicolas Gillis. ***Nonnegative matrix factorization.*** SIAM, Philadelphia, 2020.