Completely Positive Factorization via Orthogonality Constrained Problem

Lai Zhijian; Akiko Yoshise

University of Tsukuba

March 2, 2021

Background: Basic Concepts

Definition 1.1 c.f. [Abraham and Naomi, 2003]

1. A matrix $A \in \mathbb{S}_n$ is called completely positive if there exists an entrywise nonnegative matrix $B \in \mathbb{R}^{n \times r}$ such that $A = BB^T$. Such B is called a CP factorization of A.

Background: Basic Concepts

Definition 1.1 c.f. [Abraham and Naomi, 2003]

1. A matrix $A \in \mathbb{S}_n$ is called completely positive if there exists an entrywise nonnegative matrix $B \in \mathbb{R}^{n \times r}$ such that $A = BB^T$. Such B is called a CP factorization of A.
2. $\mathcal{CP}_n := \left\{ BB^\top \in \mathbb{S}_n \mid B \text{ is a nonnegative matrix} \right\}$ denotes the completely positive cone.

Background: Basic Concepts

Definition 1.1 c.f. [Abraham and Naomi, 2003]

1. A matrix $A \in \mathbb{S}_n$ is called completely positive if there exists an entrywise nonnegative matrix $B \in \mathbb{R}^{n \times r}$ such that $A = BB^T$. Such B is called a CP factorization of A.
2. $\mathcal{CP}_n := \left\{ BB^\top \in \mathbb{S}_n \mid B \text{ is a nonnegative matrix} \right\}$ denotes the completely positive cone.

Consider the matrix $A \in \mathcal{CP}_3$ where $A = B_1 B_1^T = B_2 B_2^T = B_3 B_3^T$.

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}.$$

Generally, one can have many CP factorizations, even those numbers of columns differ.

$$B_1 = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 3 & 3 & 0 \\ 3 & 0 & 3 \\ 0 & 3 & 3 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 3 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 3 & 0 & 0 & 3 \end{pmatrix}.$$

Background: Basic Concepts

Definition 1.1 c.f. [Abraham and Naomi, 2003]

1. A matrix $A \in \mathbb{S}_n$ is called completely positive if there exists an entrywise nonnegative matrix $B \in \mathbb{R}^{n\times r}$ such that $A = BB^T$. Such B is called a CP factorization of A.
2. $\mathcal{CP}_n := \left\{BB^\top \in \mathbb{S}_n \mid B \text{ is a nonnegative matrix}\right\}$ denotes the completely positive cone.

Consider the matrix $A \in \mathcal{CP}_3$ where $A = B_1B_1^T = B_2B_2^T = B_3B_3^T$.

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}.$$

Generally, one can have many CP factorizations, even those numbers of columns differ.

$$B_1 = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 3 & 3 & 0 \\ 3 & 0 & 3 \\ 0 & 3 & 3 \end{pmatrix}, \quad B_3 = \begin{pmatrix} 3 & 3 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 3 & 0 & 0 & 3 \end{pmatrix}.$$

Definition 1.2 c.f. [Abraham and Naomi, 2003]

The minimum of the number of columns among CP factorizations of $A \in \mathcal{CP}_n$ is called cp-rank of A, written as $\mathrm{cp}(A)$.

Background: Application of $\mathcal{CP}_n$

Many nonconvex NP-hard quadratic and combinatorial optimizations have a linear program over completely positive cone, $\mathcal{CP}_n$.

Background: Application of $\mathcal{CP}_n$

Many nonconvex NP-hard quadratic and combinatorial optimizations have a linear program over completely positive cone, $\mathcal{CP}_n$. For example, standard quadratic optimization:

$$\min \left\{ x^T M x \mid e^T x = 1, x \in \mathbb{R}^n_+ \right\},$$

can equivalently be written as

Standard quadratic optimization via $\mathcal{CP}_n$ [Bomze et al., 2000]

$$\min \left\{ \langle M, X \rangle \mid \left\langle ee^T, X \right\rangle = 1, X \in \mathcal{CP}_n \right\},$$

where $M \in \mathbb{S}_n$ possibly indefinite, and e is the all ones vector.

Background: Application of $\mathcal{CP}_n$

Many nonconvex NP-hard quadratic and combinatorial optimizations have a linear program over completely positive cone, $\mathcal{CP}_n$. For example, standard quadratic optimization:

$$\min\left\{x^T M x \mid e^T x = 1, x \in \mathbb{R}^n_+\right\},$$

can equivalently be written as

Standard quadratic optimization via $\mathcal{CP}_n$ [Bomze et al., 2000]

$$\min\left\{\langle M, X\rangle \mid \left\langle ee^T, X\right\rangle = 1, X \in \mathcal{CP}_n\right\},$$

where $M \in \mathbb{S}_n$ possibly indefinite, and e is the all ones vector.

An application of above is

Independence number α of a graph G [De Klerk and Pasechnik, 2002]

$$\alpha(G) = \max\left\{\langle ee^T, X\rangle \mid \langle A + I, X\rangle = 1, X \in \mathcal{CP}_n\right\},$$

where A is the adjacency matrix of G.

Background: Open Problem — Finding a CP factorization

Open Problem of $\mathcal{CP}_n$

There are many fundamental open problems in completely positive cone, $\mathcal{CP}_n$.

Background: Open Problem — Finding a CP factorization

Open Problem of $\mathcal{CP}_n$

There are many fundamental open problems in completely positive cone, $\mathcal{CP}_n$.

Here is a list of these open problems [Berman et al., 2015]:

1. Checking membership in $\mathcal{CP}_n$.
2. Determining geometry of $\mathcal{CP}_n$.
3. **Finding a factorization of a matrix in $\mathcal{CP}_n$. (Our goal)**
   $\longrightarrow$ Given $A \in \mathcal{CP}_n$, find a nonnegative B such that $A = BB^T$.
4. Computing the cp-rank.
5. Finding cutting planes for completely positive optimization problems.

CP factorization as a Feasibility Problem: Lemma 2.1

Lemma 2.1

Suppose that $A \in \mathbb{S}_n$, $r \in \mathbb{N}$. Then

$$r \geq \mathrm{cp}(A) \Longleftrightarrow A \text{ has a CP factorization } B \text{ with } r \text{ columns.}$$

CP factorization as a Feasibility Problem: Lemma 2.1

Lemma 2.1

Suppose that $A \in \mathbb{S}_n$, $r \in \mathbb{N}$. Then

$$r \geq \mathrm{cp}(A) \Longleftrightarrow A \text{ has a CP factorization } B \text{ with } r \text{ columns.}$$

Consider the matrix $A \in \mathcal{CP}_3$,

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}.$$

Given $A = B_1 B_1^T$, we can easily construct $\widehat{B}_1$ such that $\widehat{B}_1 \widehat{B}_1^\top = A$.

$$B_1 = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix} \geq 0, \quad \longrightarrow \quad \widehat{B}_1 := \begin{pmatrix} 4 & 1 & 1 & 0 \\ 1 & 4 & 1 & 0 \\ 1 & 1 & 4 & 0 \end{pmatrix} \geq 0, \text{or} \begin{pmatrix} 4 & 1 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 \\ 1 & 1 & 4 & 0 & 0 \end{pmatrix} \geq 0.$$

CP factorization as a Feasibility Problem: Lemma 2.1

Lemma 2.1

Suppose that $A \in \mathbb{S}_n$, $r \in \mathbb{N}$. Then

$$r \geq \mathrm{cp}(A) \iff A \text{ has a CP factorization } B \text{ with } r \text{ columns.}$$

Consider the matrix $A \in \mathcal{CP}_3$,

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}.$$

Given $A = B_1 B_1^T$, we can easily construct $\widehat{B}_1$ such that $\widehat{B}_1 \widehat{B}_1^\top = A$.

$$B_1 = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix} \geq 0, \quad \longrightarrow \quad \widehat{B}_1 := \begin{pmatrix} 4 & 1 & 1 & 0 \\ 1 & 4 & 1 & 0 \\ 1 & 1 & 4 & 0 \end{pmatrix} \geq 0, \text{or} \begin{pmatrix} 4 & 1 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 & 0 \\ 1 & 1 & 4 & 0 & 0 \end{pmatrix} \geq 0.$$

Thus, if we have had a CP-factorization B with r columns, then we can easily get another CP-factorization $\widehat{B}$ with r' columns for every positive integer $r' \geq r$.

Lemma 2.2 [Xu, 2004, Lemma 1.]

Let $\mathcal{O}_r$ denote the set of $r \times r$ orthogonal matrices. Suppose that $B, C \in \mathbb{R}^{n \times r}$. Then

$$BB^T = CC^T \iff \exists X \in \mathcal{O}_r \text{ such that } BX = C.$$

CP factorization as a Feasibility Problem: Lemma 2.2

Lemma 2.2 [Xu, 2004, Lemma 1.]

Let $\mathcal{O}_r$ denote the set of $r \times r$ orthogonal matrices. Suppose that $B, C \in \mathbb{R}^{n \times r}$. Then

$$BB^T = CC^T \Longleftrightarrow \exists X \in \mathcal{O}_r \text{ such that } BX = C.$$

We have known that $A = B_1 B_1^T = B_2 B_2^T$.

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}, \text{ and } B_1 = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 3 & 3 & 0 \\ 3 & 0 & 3 \\ 0 & 3 & 3 \end{pmatrix}.$$

CP factorization as a Feasibility Problem: Lemma 2.2

Lemma 2.2 [Xu, 2004, Lemma 1.]

Let $\mathcal{O}_r$ denote the set of $r \times r$ orthogonal matrices. Suppose that $B, C \in \mathbb{R}^{n \times r}$. Then

$$BB^T = CC^T \iff \exists X \in \mathcal{O}_r \text{ such that } BX = C.$$

We have known that $A = B_1 B_1^T = B_2 B_2^T$.

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}, \text{ and } B_1 = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 3 & 3 & 0 \\ 3 & 0 & 3 \\ 0 & 3 & 3 \end{pmatrix}.$$

In fact, there is an orthogonal matrix X such that $B_1 X = B_2$.

$$X = \frac{1}{3}\begin{pmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{pmatrix} \in \mathcal{O}_3.$$

CP factorization as a Feasibility Problem

1. From a "bad" factorization $B_0 \ngeq 0$. Using spectral decomposition $A = VDV^T$, we define $B_0 := V\sqrt{D}$, then $A = B_0 B_0^T$.

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}, B_0 = \begin{bmatrix} \frac{3}{\sqrt{2}} & \frac{\sqrt{6}}{2} & 2\sqrt{3} \\ -\frac{3}{\sqrt{2}} & \frac{\sqrt{6}}{2} & 2\sqrt{3} \\ 0 & -\sqrt{6} & 2\sqrt{3} \end{bmatrix} \ngeq 0.$$

CP factorization as a Feasibility Problem

1. From a "bad" factorization $B_0 \not\geq 0$. Using spectral decomposition $A = VDV^T$, we define $B_0 := V\sqrt{D}$, then $A = B_0 B_0^T$.

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}, B_0 = \begin{bmatrix} \frac{3}{\sqrt{2}} & \frac{\sqrt{6}}{2} & 2\sqrt{3} \\ -\frac{3}{\sqrt{2}} & \frac{\sqrt{6}}{2} & 2\sqrt{3} \\ 0 & -\sqrt{6} & 2\sqrt{3} \end{bmatrix} \not\geq 0.$$

2. To a "good" factorization $B_0 X \geq 0$. If we find a suitable orthogonal matrix X, e.g.,

$$X = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix} \in \mathcal{O}_3, B_0 X = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix} = B_1 \geq 0.$$

CP factorization as a Feasibility Problem

1. From a "bad" factorization $B_0 \not\geq 0$. Using spectral decomposition $A = VDV^T$, we define $B_0 := V\sqrt{D}$, then $A = B_0 B_0^T$.

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}, B_0 = \begin{bmatrix} \frac{3}{\sqrt{2}} & \frac{\sqrt{6}}{2} & 2\sqrt{3} \\ -\frac{3}{\sqrt{2}} & \frac{\sqrt{6}}{2} & 2\sqrt{3} \\ 0 & -\sqrt{6} & 2\sqrt{3} \end{bmatrix} \not\geq 0.$$

2. To a "good" factorization $B_0 X \geq 0$. If we find a suitable orthogonal matrix X, e.g.,

$$X = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix} \in \mathcal{O}_3, B_0 X = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix} = B_1 \geq 0.$$

CP factorization as a feasibility problem [Groetzner and Dür, 2020]

$$\begin{array}{ll} \text{find} & X \\ \text{s.t.} & BX \geq 0 \\ & X \in \mathcal{O}_r \end{array} \tag{1}$$

where $r \geq \text{cp}(A)$, $B \in \mathbb{R}^{n \times r}$ is an arbitrary initial factorization $A = BB^T$ (need not nonnegative).

CP factorization as a Feasibility Problem

1. From a "bad" factorization $B_0 \ngeq 0$. Using spectral decomposition $A = VDV^T$, we define $B_0 := V\sqrt{D}$, then $A = B_0 B_0^T$.

$$A = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}, B_0 = \begin{bmatrix} \frac{3}{\sqrt{2}} & \frac{\sqrt{6}}{2} & 2\sqrt{3} \\ -\frac{3}{\sqrt{2}} & \frac{\sqrt{6}}{2} & 2\sqrt{3} \\ 0 & -\sqrt{6} & 2\sqrt{3} \end{bmatrix} \ngeq 0.$$

2. To a "good" factorization $B_0 X \geq 0$. If we find a suitable orthogonal matrix X, e.g.,

$$X = \begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{3} \\ \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix} \in \mathcal{O}_3, B_0 X = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix} = B_1 \geq 0.$$

CP factorization as a feasibility problem [Groetzner and Dür, 2020]

$$\begin{aligned} \text{find} \quad & X \\ \text{s.t.} \quad & BX \geq 0 \\ & X \in \mathcal{O}_r \end{aligned} \tag{1}$$

where $r \geq \text{cp}(A)$, $B \in \mathbb{R}^{n \times r}$ is an arbitrary initial factorization $A = BB^T$ (need not nonnegative). **From Lemma 2.1 and 2.2, we have $A \in \mathcal{CP}_n \iff$ (1) is feasible.**

Our Approach to feasibility problem

1. We first establish the connection between (1) and (2):

$$\begin{array}{ll} \text{find} & X \\ \text{s.t.} & BX \geq 0 \\ & X \in \mathcal{O}_r \end{array} \quad (1) \quad \longrightarrow \quad \begin{array}{ll} \max & \min\,(BX)_{ij} \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \quad \longrightarrow \quad \begin{array}{ll} \min & \max\,(-BX)_{ij} \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \quad (2)$$

Our Approach to feasibility problem

1. We first establish the connection between (1) and (2):

$$
\begin{array}{ll} \text{find} & X \\ \text{s.t.} & BX \geq 0 \\ & X \in \mathcal{O}_r \end{array} \quad (1) \quad \longrightarrow \quad \begin{array}{ll} \max & \min\,(BX)_{ij} \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \quad \longrightarrow \quad \begin{array}{ll} \min & \max\,(-BX)_{ij} \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \quad (2)
$$

2. Introducing a differentiable approximation for the sake of adapting the method below:

$$
\max\,(-BX)_{ij} \stackrel{\text{approximate}}{\longrightarrow} LSE_{\mu}\,(-BX)\,.
$$

Our Approach to feasibility problem

1. We first establish the connection between (1) and (2):

$$\begin{array}{ll} \text{find} & X \\ \text{s.t.} & BX \geq 0 \\ & X \in \mathcal{O}_r \end{array} \quad (1) \quad \longrightarrow \quad \begin{array}{ll} \max & \min (BX)_{ij} \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \quad \longrightarrow \quad \begin{array}{ll} \min & \max (-BX)_{ij} \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \quad (2)$$

2. Introducing a differentiable approximation for the sake of adapting the method below:

$$\max (-BX)_{ij} \overset{\text{approximate}}{\longrightarrow} LSE_{\mu} (-BX) .$$

3. Adopting a state-of-the-art curvilinear search method, which aims to solve the general optimization with orthogonality constrains:

$$\min_{X \in \mathbb{R}^{n \times p}} \mathcal{F}(X), \text{ s.t. } X^T X = I,$$

where $\mathcal{F}(X) : \mathbb{R}^{n \times p} \to \mathbb{R}$ is C^1.

CP Factorization via Orthogonality Constrained Problem

LogSumExp: Smooth Approximation to Max Function

The LogSumExp (LSE) function is given by $LSE_\mu(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$,

$$LSE_\mu(\mathbf{x}) = \mu \log \left(\sum_{i=1}^{n} \exp(x_i/\mu) \right).$$

LogSumExp: Smooth Approximation to Max Function

The LogSumExp (LSE) function is given by $LSE_\mu(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$,

$$LSE_\mu(\mathbf{x}) = \mu \log \left(\sum_{i=1}^n \exp(x_i/\mu) \right).$$

The gradient of LSE function is the well-known softmax function, which is given by $\sigma : \mathbb{R}^n \to \mathbb{R}^n$,

$$\sigma_\mu(x) := \frac{1}{\sum_{j=1}^n \exp\left(x_j/\mu\right)} \begin{bmatrix} \exp\left(x_1/\mu\right) \\ \vdots \\ \exp\left(x_n/\mu\right) \end{bmatrix}.$$

LogSumExp: Smooth Approximation to Max Function

The LogSumExp (LSE) function is given by $LSE_\mu(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$,

$$LSE_\mu(\mathbf{x}) = \mu \log \left(\sum_{i=1}^n \exp(x_i/\mu)\right).$$

The gradient of LSE function is the well-known softmax function, which is given by $\sigma : \mathbb{R}^n \to \mathbb{R}^n$,

$$\sigma_\mu(x) := \frac{1}{\sum_{j=1}^n \exp(x_j/\mu)} \begin{bmatrix} \exp(x_1/\mu) \\ \vdots \\ \exp(x_n/\mu) \end{bmatrix}.$$

Let $x = (2, 5, -1, 3)$.

$n = 4$	$\mu = 1$	$\mu = 1/2$	$\mu = 1/4$	$\mu = 1/8$
$LSE_\mu(\mathbf{x})$	5.1719	5.0103	5.0001	5.0000

Table: Example of approximation effect with different parameters μ.

LogSumExp: Smooth Approximation to Max Function

The LogSumExp (LSE) function is given by $LSE_\mu(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$,

$$LSE_\mu(\mathbf{x}) = \mu \log \left(\sum_{i=1}^n \exp(x_i/\mu) \right).$$

LogSumExp: Smooth Approximation to Max Function

The LogSumExp (LSE) function is given by $LSE_\mu(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$,

$$LSE_\mu(\mathbf{x}) = \mu \log \left(\sum_{i=1}^{n} \exp(x_i/\mu) \right).$$

Theorem 3.1 — Approximation theorem of LogSumExp [Lai, 2020]

Suppose that $\mu > 0$, and $\max x_i$ denotes the maximum entry of $\mathbf{x}$. For all $\mathbf{x} \in \mathbb{R}^n$, we have

1. $\max x_i < LSE_\mu(\mathbf{x}) \leq \max x_i + \mu \log(n)$, hence $|\max x_i - LSE_\mu(\mathbf{x})| \leq \mu \log(n)$.
2. if $0 < \mu_2 < \mu_1$, then $LSE_{\mu_2}(\mathbf{x}) < LSE_{\mu_1}(\mathbf{x})$.

LogSumExp: Smooth Approximation to Max Function

The LogSumExp (LSE) function is given by $LSE_\mu(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$,

$$LSE_\mu(\mathbf{x}) = \mu \log \left(\sum_{i=1}^n \exp(x_i/\mu)\right).$$

Theorem 3.1 — Approximation theorem of LogSumExp [Lai, 2020]

Suppose that $\mu > 0$, and $\max x_i$ denotes the maximum entry of $\mathbf{x}$. For all $\mathbf{x} \in \mathbb{R}^n$, we have

1. $\max x_i < LSE_\mu(\mathbf{x}) \le \max x_i + \mu \log(n)$, hence $|\max x_i - LSE_\mu(\mathbf{x})| \le \mu \log(n)$.
2. if $0 < \mu_2 < \mu_1$, then $LSE_{\mu_2}(\mathbf{x}) < LSE_{\mu_1}(\mathbf{x})$.

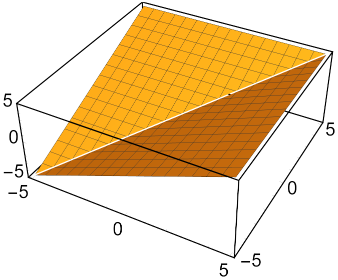

Figure: Graph of $\max(x, y)$.

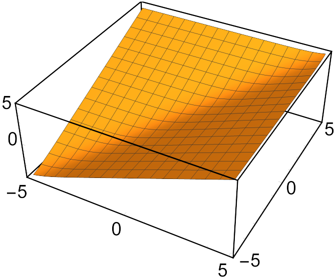

Figure: Graph of $\log(e^x + e^y)$.

LogSumExp: Smooth Approximation to Max Function

The LogSumExp (LSE) function is given by $LSE_\mu(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$,

$$LSE_\mu(\mathbf{x}) = \mu \log \left(\sum_{i=1}^n \exp(x_i/\mu) \right).$$

Theorem 3.1 — Approximation theorem of LogSumExp [Lai, 2020]

Suppose that $\mu > 0$, and $\max x_i$ denotes the maximum entry of $\mathbf{x}$. For all $\mathbf{x} \in \mathbb{R}^n$, we have

1. $\max x_i < LSE_\mu(\mathbf{x}) \leq \max x_i + \mu \log(n)$, hence $|\max x_i - LSE_\mu(\mathbf{x})| \leq \mu \log(n)$.
2. if $0 < \mu_2 < \mu_1$, then $LSE_{\mu_2}(\mathbf{x}) < LSE_{\mu_1}(\mathbf{x})$.

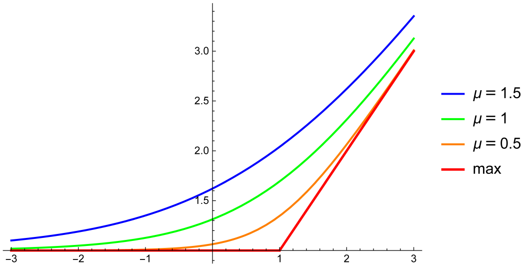

Figure: Slices through the line $y = 1$ of $\max(x, y)$ and some $\mu \log(e^{x/\mu} + e^{y/\mu})$ on $\mathbb{R}^2$.

LogSumExp: Smooth Approximation to Max Function

Approximate the problem (2) by problem (3):

$$\begin{aligned} \min \quad & \max\,(-BX)_{ij} \\ \text{s.t.} \quad & X \in \mathcal{O}_r \end{aligned} \qquad (2)$$

$$\begin{aligned} \min \quad & LSE_\mu\,(-BX) \\ \text{s.t.} \quad & X \in \mathcal{O}_r \end{aligned} \qquad (3)$$

LogSumExp: Smooth Approximation to Max Function

Approximate the problem (2) by problem (3):

$$\begin{array}{ll} \min & \max\,(-BX)_{ij} \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \qquad (2) \qquad\qquad \begin{array}{ll} \min & LSE_{\mu}\,(-BX) \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \qquad (3)$$

Proposition 3.2 — Approximation between two minimum [Lai, 2020]

Let t resp. t_μ denote global minimum of problem (2) resp. (3). We have

$$0 < t_\mu - t \leq \mu \log(nr).$$

LogSumExp: Smooth Approximation to Max Function

Approximate the problem (2) by problem (3):

$$\begin{array}{ll} \min & \max\,(-BX)_{ij} \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \qquad (2)$$

$$\begin{array}{ll} \min & LSE_{\mu}\,(-BX) \\ \text{s.t.} & X \in \mathcal{O}_r \end{array} \qquad (3)$$

Proposition 3.2 — Approximation between two minimum [Lai, 2020]

Let t resp. t_μ denote global minimum of problem (2) resp. (3). We have

$$0 < t_\mu - t \leq \mu \log(nr).$$

Proposition 3.3 [Lai, 2020]

If problem (3) has a feasible solution X such that $LSE_\mu\,(-BX) \leq 0$, then we find a CP factorization $A = (BX)(BX)^T$ with $BX > 0$.

A Curvilinear Search Method [Wen and Yin, 2013]

Consider

$$\min_{X \in \mathcal{M}_n^p} \mathcal{F}(X), \tag{StOp}$$

where $\mathcal{F}(X) : \mathbb{R}^{n\times p} \to \mathbb{R}$ is C^1, and $\mathcal{M}_n^p := \left\{X \in \mathbb{R}^{n\times p} : X^T X = I\right\}$ is called Stiefel manifold.

A Curvilinear Search Method [Wen and Yin, 2013]

Consider

$$\min_{X \in \mathcal{M}_n^p} \mathcal{F}(X), \qquad \text{(StOp)}$$

where $\mathcal{F}(X) : \mathbb{R}^{n\times p} \to \mathbb{R}$ is C^1, and $\mathcal{M}_n^p := \left\{X \in \mathbb{R}^{n\times p} : X^T X = I\right\}$ is called Stiefel manifold.

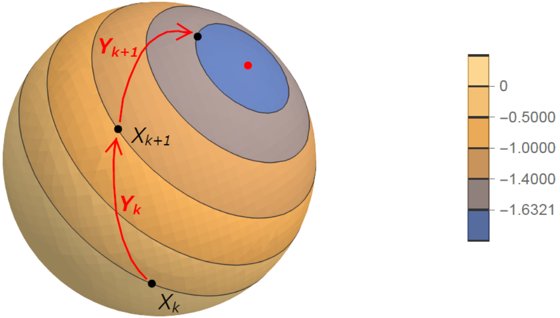

Figure: Illustration of unit sphere $\mathcal{M}_3^1$.

Gradient Descent (Linear Search) Method on $\mathbb{R}^n$.

$\xrightarrow{\text{extend}}$ Curvilinear Search Method on Manifold.

1. At a point X on $\mathcal{M}_n^p$, we construct a smooth curve $Y(\tau) : \mathbb{R} \to \mathcal{M}_n^p$ such that $Y(0) = X$.

A Curvilinear Search Method [Wen and Yin, 2013]

Consider

$$\min_{X \in \mathcal{M}_n^p} \mathcal{F}(X), \qquad \text{(StOp)}$$

where $\mathcal{F}(X) : \mathbb{R}^{n\times p} \to \mathbb{R}$ is C^1, and $\mathcal{M}_n^p := \left\{X \in \mathbb{R}^{n\times p} : X^T X = I\right\}$ is called Stiefel manifold.

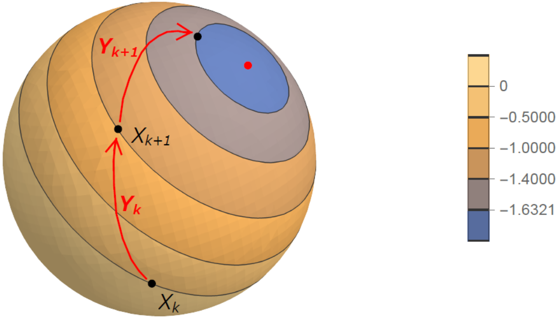

Figure: Illustration of unit sphere $\mathcal{M}_3^1$.

Gradient Descent (Linear Search) Method on $\mathbb{R}^n$.
$\xrightarrow{\text{extend}}$ Curvilinear Search Method on Manifold.

1. At a point X on $\mathcal{M}_n^p$, we construct a smooth curve $Y(\tau) : \mathbb{R} \to \mathcal{M}_n^p$ such that $Y(0) = X$.
2. If X is not a local minimizer of (StOp), then $\exists \bar{\tau} \in \mathbb{R}, \mathcal{F}\left(Y\left(\bar{\tau}\right)\right) < \mathcal{F}\left(Y(0)\right)$. It is true if $\left.\frac{d\mathcal{F}(Y(\tau))}{d\tau}\right|_{\tau=0} < 0$ for $\mathcal{F}\left(Y\left(\tau\right)\right) : \mathbb{R} \to \mathbb{R}$.

Consider

$$\min_{X \in \mathcal{M}_n^p} \mathcal{F}(X), \qquad \text{(StOp)}$$

where $\mathcal{F}(X) : \mathbb{R}^{n\times p} \to \mathbb{R}$ is C^1, and $\mathcal{M}_n^p := \left\{X \in \mathbb{R}^{n\times p} : X^T X = I\right\}$ is called Stiefel manifold.

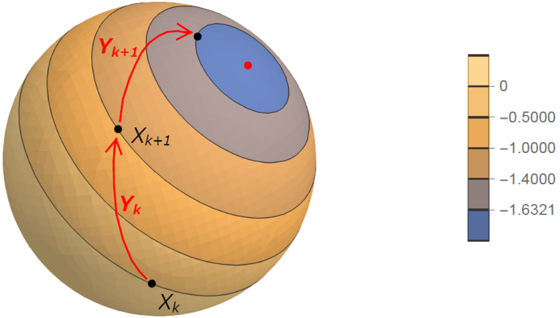

Figure: Illustration of unit sphere $\mathcal{M}_3^1$.

Gradient Descent (Linear Search) Method on $\mathbb{R}^n$.
$\xrightarrow{\text{extend}}$ Curvilinear Search Method on Manifold.

1. At a point X on $\mathcal{M}_n^p$, we construct a smooth curve $Y(\tau) : \mathbb{R} \to \mathcal{M}_n^p$ such that $Y(0) = X$.
2. If X is not a local minimizer of (StOp), then $\exists \bar{\tau} \in \mathbb{R}, \mathcal{F}\left(Y\left(\bar{\tau}\right)\right) < \mathcal{F}\left(Y(0)\right)$. It is true if $\left.\frac{d\mathcal{F}(Y(\tau))}{d\tau}\right|_{\tau=0} < 0$ for $\mathcal{F}\left(Y\left(\tau\right)\right) : \mathbb{R} \to \mathbb{R}$.
3. Get a better point $X := Y(\bar{\tau})$. Return to the first step until it is close to local minimizer.

A Curvilinear Search Method [Wen and Yin, 2013]

Suppose $X \in \mathcal{M}_n^p$. Let gradient $G := \mathcal{DF}(X)$. Define

$$\nabla \mathcal{F}(X) := G - XG^T X \text{ and } A := GX^T - XG^T.$$

Then $\nabla \mathcal{F}(X) = AX$. Note that $\nabla \mathcal{F}(X) = 0$ if and only if $A = 0$.

Lemma 3.3 — First-order Optimality Condition [Wen and Yin, 2013, Lemma 1.]

If a feasible point X is a local minimizer of (StOp). Then X satisfies $\nabla \mathcal{F}(X) = 0$.

Lemma 3.4 — Update Scheme [Wen and Yin, 2013, Lemma 3.]

1. X is a feasible point. Given any skew-symmetric matrix $W \in \mathbb{R}^{n \times n}$ (i.e. $W^T = -W$), the matrix

$$Y(\tau) := \left(I + \frac{\tau}{2} W\right)^{-1} \left(I - \frac{\tau}{2} W\right) X$$

is a smooth curve $Y(\tau) : \mathbb{R} \to \mathcal{M}_n^p$ such that $Y(0) = X$.

2. If set $W = A := GX^T - XG^T$. Then $\left.\frac{d\mathcal{F}(Y(\tau))}{d\tau}\right|_{\tau=0} = -\frac{1}{2}\|A\|_F^2$.

A Curvilinear Search Method

Algorithm 1: A New Method for Completely Positive Matrix Factorization

Data: Given $A \in \mathcal{CP}_n, r \geq \text{cp}(A)$.
Result: An $n \times r$ CP factorization of A.
Initialization: Choose an initial decomposition $B \in \mathbb{R}^{n\times r}$ and starting point $X_0 \in \mathcal{O}_r$. Set $0 < \theta_1 < \theta_2 < 1, \mu < 0, \epsilon > 0, k \leftarrow 0$;
while $\|\nabla \mathcal{F}(X_k)\| > \epsilon$ **do**

Generate $G_k \leftarrow -B^T \frac{\partial LSE_\mu(BX_k)}{\partial(BX_k)}, A_k \leftarrow G_k X_k^T - X_k G_k^T, W_k \leftarrow A_k$;
Find a step size $\tau_k > 0$ that satisfies the Armijo-Wolfe conditions:

$$\mathcal{F}\left(Y_k\left(\tau_k\right)\right) \leq \mathcal{F}\left(Y_k(0)\right) + \theta_1 \tau_k \mathcal{F}_\tau'\left(Y_k(0)\right)$$
$$\mathcal{F}_\tau'\left(Y_k\left(\tau_k\right)\right) \geq \theta_2 \mathcal{F}_\tau'\left(Y_k(0)\right);$$

Set $X_{k+1} \leftarrow Y_k\left(\tau_k\right)$, $k \leftarrow k+1$;

end

Global convergence of local minimizer of (3) [Wen and Yin, 2013, Theorem 2]

The global convergence of local minimizer is guaranteed, that is $\lim_{k\to\infty} \|\nabla \mathcal{F}\left(X_k\right)\|_F = 0$.

Numerical Results

Let $A := HH^T$, where $H \in \mathbb{R}^{n\times n}$ with entries randomly generated in $\{1, 2, \ldots, 10\}$. We use Algorithm 1 and Groetzner's method to factorize A for finding other CP factorization of A.

- Intel Core i7-4770 3.40 GHz and 16GB Ram, and MatlabR2020a.
- We take $n \in \{10, 15, 20, 25, 30\}$ and $r = t * n$ for $t \in \{1, 1.5, 2, 3\}$.
- For each pair of n and r, we generated 100 instances to examine.

Numerical Results

Let $A := HH^T$, where $H \in \mathbb{R}^{n \times n}$ with entries randomly generated in $\{1, 2, \ldots, 10\}$. We use Algorithm 1 and Groetzner's method to factorize A for finding other CP factorization of A.

- Intel Core i7-4770 3.40 GHz and 16GB Ram, and MatlabR2020a.
- We take $n \in \{10, 15, 20, 25, 30\}$ and $r = t * n$ for $t \in \{1, 1.5, 2, 3\}$.
- For each pair of n and r, we generated 100 instances to examine.
- For each instance, both methods have up to 100 initial point opportunities. If it is not successful after 5000 iterations, then the next initial point will be tried.
- As long as one succeeds, no other initial points are tried. The time for a single instance is calculated from the first initial point to the last successful one.

Numerical Results

Table: A direct comparison of Algorithm 1 and Groetzner's method.

$N = 100$		Algorithm 1		Groetzner's method	
n	$r = n$	no. of successful cases	av. time (sec.) for successful cases	no. of successful cases	av. time (sec.) for successful cases
10	10	100	0.0010	95	3.17
15	15	100	0.0016	77	13.98
20	20	100	0.0025	2	26.76
25	25	100	0.0043	0	-
30	30	100	0.0056	0	-
n	$r \approx 1.5n$	no. of successful cases	av. time (sec.) for successful cases	no. of successful cases	av. time (sec.) for successful cases
10	15	100	0.0014	100	0.39
15	23	100	0.0027	100	3.78
20	30	100	0.0044	93	18.33
25	38	100	0.0089	65	55.10
30	45	100	0.0121	36	122.37
n	$r = 2n$	no. of successful cases	av. time (sec.) for successful cases	no. of successful cases	av. time (sec.) for successful cases
10	20	100	0.0031	100	0.41
15	30	100	0.0041	100	1.94
20	40	100	0.0076	99	9.56
25	50	100	0.0127	82	40.82
30	60	100	0.0166	50	64.87

Summary: Why so fast is our approach?

Table: Dominant computation of some CP factorization algorithms.

CP factorization Methods	Dominant Computation at Each Iteration
Our approach	matrix inverse (only!)
Groetzner's method	Second Order Cone Problem
[Jarre and Schmallowsky, 2009]	Second Order Cone Problem
[Nie, 2014]	Semidefinite Optimization Problem
[Sponsel and Dür, 2014]	Second Order Cone problem

Why so fast is our approach?

Only the inversion of an $n \times n$ matrix dominates the computation at each iteration, not other subproblems like SOCP. Although solving an SOCP can be done in polynomial time, it is still very costly overall. Clearly, if subproblems have to be solved in each iteration, algorithm will always be slow.

Thank you for listening.

Reference I

[Abraham and Naomi, 2003] Abraham, B. and Naomi, S.-m. (2003).
Completely Positive Matrices.
World Scientific.

[Berman et al., 2015] Berman, A., Dür, M., and Shaked-Monderer, N. (2015).
Open problems in the theory of completely positive and copositive matrices.
Electronic Journal of Linear Algebra, 29(1):46–58.

[Bomze et al., 2015] Bomze, I. M., Dickinson, P. J., and Still, G. (2015).
The structure of completely positive matrices according to their cp-rank and cp-plus-rank.
Linear algebra and its applications, 482:191–206.

[Bomze et al., 2000] Bomze, I. M., Dür, M., De Klerk, E., Roos, C., Quist, A. J., and Terlaky, T. (2000).
On copositive programming and standard quadratic optimization problems.
Journal of Global Optimization, 18(4):301–320.

[De Klerk and Pasechnik, 2002] De Klerk, E. and Pasechnik, D. V. (2002).
Approximation of the stability number of a graph via copositive programming.
SIAM Journal on Optimization, 12(4):875–892.

Reference II

[Dickinson, 2010] Dickinson, P. J. (2010).
An improved characterisation of the interior of the completely positive cone.
Electronic Journal of Linear Algebra, 20(1):723–729.

[Dickinson and Dür, 2012] Dickinson, P. J. and Dür, M. (2012).
Linear-time complete positivity detection and decomposition of sparse matrices.
SIAM Journal on Matrix Analysis and Applications, 33(3):701–720.

[Groetzner and Dür, 2020] Groetzner, P. and Dür, M. (2020).
A factorization method for completely positive matrices.
Linear Algebra and its Applications, 591:1–24.

[Jarre and Schmallowsky, 2009] Jarre, F. and Schmallowsky, K. (2009).
On the computation of C^* certificates.
Journal of Global Optimization, 45(2):281.

[Nie, 2014] Nie, J. (2014).
The $\mathcal{A}$-truncated K-moment problem.
Foundations of Computational Mathematics, 14(6):1243–1276.

Reference III

[Sikirić et al., 2020] Sikirić, M. D., Schürmann, A., and Vallentin, F. (2020).
A simplex algorithm for rational cp-factorization.
Mathematical Programming, pages 1–21.

[Sponsel and Dür, 2014] Sponsel, J. and Dür, M. (2014).
Factorization and cutting planes for completely positive matrices by copositive projection.
Mathematical Programming, 143(1-2):211–229.

[Wen and Yin, 2013] Wen, Z. and Yin, W. (2013).
A feasible method for optimization with orthogonality constraints.
Mathematical Programming, 142(1-2):397–434.

[Xu, 2004] Xu, C. (2004).
Completely positive matrices.
Linear algebra and its applications, 379:319–327.

Appendix: Copositive program (primal and dual problem)

Consider the so-called copositive program (primal problem)

$$\min\left\{\langle C, X\rangle \mid \langle A_i, X\rangle = b_i \ (i = 1, \ldots, m), X \in \mathcal{COP}_n\right\}, \tag{4}$$

where

$$\mathcal{COP}_n \triangleq \left\{A \in \mathcal{S}_n | x^T A x \geq 0 \text{ for all } x \in \mathbb{R}^n_+\right\}$$

is the cone of so-called copositive matrices. The dual problem of (4) is

$$\max\left\{\sum_{i=1}^m b_i y_i \mid C - \sum_{i=1}^m y_i A_i \in \mathcal{CP}_n,\ y_i \in \mathbb{R}\right\}, \tag{5}$$

where $\mathcal{CP}_n$ denotes the set of $n \times n$ completely positive matrices.

Dual cones, cf. [Abraham and Naomi, 2003]

$\mathcal{CP}_n$ and $\mathcal{COP}_n$ are mutually dual to each other.

Appendix: Relationship with other matrix cones

1. The cone of so-called copositive matrices: $\mathcal{COP}_n := \left\{A \in \mathcal{S}_n | x^T A x \geq 0 \text{ for all } x \in \mathbb{R}^n_+\right\}$.
2. The cone of symmetric entrywise nonnegative matrices: $\mathcal{N}_n := \{A \in \mathcal{S}_n \mid A_{ij} \geq 0 \text{ for all } i, j = 1, \dots, n\}$.
3. The cone of symmetric positive semidefinite matrices: $\mathcal{S}_n^+ := \left\{A \in \mathcal{S}_n \mid x^T A x \geq 0 \text{ for all } x \in \mathbb{R}^n\right\}$.
4. The cone of doubly nonnegative matrices: $\mathcal{DNN}_n := \mathcal{S}_n^+ \cap \mathcal{N}_n$.
5. The Minkowski sum of $\mathcal{S}_n^+$ and $\mathcal{N}_n$: $\mathcal{S}_n^+ + \mathcal{N}_n := \{A + B \in \mathcal{S}_n \mid A \in \mathcal{S}_n^+, B \in \mathcal{N}_n\}$.

Relationship with other matrix cones

$$\mathcal{CP}_n \subseteq \mathcal{S}_n^+ \cap \mathcal{N}_n \subseteq \mathcal{S}_n^+ \subseteq \mathcal{S}_n^+ + \mathcal{N}_n \subseteq \mathcal{COP}_n \tag{6}$$

Appendix: Existing methods

Many different methods to CP-factorization problem have studied before.

1. Some work well for the matrices with specific property.
   —*special sparse matrices [Dickinson and Dür, 2012], rational CP-factorization [Sikirić et al., 2020].*
2. Some work for all matrices but are numerically expensive.
   —*[Nie, 2014], [Jarre and Schmallowsky, 2009], [Sponsel and Dür, 2014].*

Appendix: Upper bound of cp-rank

Lemma-Upper bound of cp-rank [Bomze et al., 2015, Theorem 4.1]

For all $A \in \mathcal{CP}_n$, we have $\text{cp}(A) \leq \mathrm{cp}_n := \begin{cases} n & \text{for } n \in \{2,3,4\} \\ \frac{1}{2}n(n+1) - 4 & \text{for } n \geq 5 \end{cases}$

Appendix: cp-plus-rank and $\text{int}\,(\mathcal{CP}^n)$

The cp-rank and the cp-plus-rank

We define cp-rank of $A \in \mathcal{S}_n$ as the minimum of the number of columns for all CP-factor of A, that is

$$\text{cp}(A) := \min_B \left\{ r \in \mathbb{N} | \exists B \in \mathbb{R}^{n \times r}, B \geq 0, A = BB^T \right\}.$$

Notice that $\text{cp}(A) := \infty$ if $A \notin \mathcal{CP}_n$. We also define cp-plus-rank of $A \in \mathcal{S}_n$ as

$$\text{cp}^+(A) := \min_B \left\{ r \in \mathbb{N} | \exists B \in \mathbb{R}^{n \times r}, B > 0, A = BB^T \right\}.$$

Theorem 2.2 Interior of $\mathcal{CP}^n$ [Dickinson, 2010, Theorem 3.8]

$$\text{int}\,(\mathcal{CP}^n) = \{A \in \mathbb{S}_n \mid \text{cp}^+(A) < \infty \text{ and } \text{rank}(A) = n\}$$

Appendix: Approximation Lemma for LSE of matrix form

Lemma 3.3 — Approximation Lemma for LSE of matrix form [L, 2020]

Suppose that $\mu > 0, B \in \mathbb{R}^{n \times r}$, and $\max(-BX)_{ij}$ denotes the minimum entry of $-BX$. For all $X \in \mathbb{R}^{r \times r}$, we have

1. $\max(-BX)_{ij} < LSE_{\mu}(-BX) \leq \max(-BX)_{ij} + \mu \log(nr)$.
2. if $0 < \mu_2 < \mu_1$, then $LSE_{\mu_2}(-BX) < LSE_{\mu_1}(-BX)$.