CP-factorization via Orthogonality Constrained Problem

Zhijian Lai, Akiko Yoshise

University of Tsukuba

August 24, 2020

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*.
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 $A_1 \in \mathcal{CP}_3$ where $A_1 = BB^T$.

$$A_1 = \begin{pmatrix} 18 & 9 & 9 \\ 9 & 18 & 9 \\ 9 & 9 & 18 \end{pmatrix}, B = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{pmatrix}.$$

A counterexample

$$A_2 = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 3 \end{pmatrix} \in \mathcal{S}_5^+ \cap \mathcal{N}_5 \backslash \mathcal{C}P_5.$$

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

Many nonconvex NP-hard quadratic and combinatorial optimizations have a linear program over $\mathcal{CP}_n$. 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 by $\mathcal{CP}_n$ [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 by $\mathcal{CP}_n$ [De Klerk and Pasechnik, 2002]

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

where A is the adjacency matrix of G and E is the all-ones matrix.

Background: Open Problem — Finding a CP-factorization

Checking membership in $\mathcal{CP}_n$ is NP-hard [Dickinson and Gijben, 2014].

RECALL: 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*.

Our Goal

Given $A \in \mathcal{CP}_n$, find a CP-factorization of A.

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].*

Existing Results on $\mathcal{CP}_n$

Example 1.2 [Groetzner and Dür, 2020, Example 2.1.]

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.

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

Existing Results on $\mathcal{CP}_n$: cp-(plus)-rank and $\mathrm{int}\,(\mathcal{CP}^n)$

Definition 2.1 c.f. [Abraham and Naomi, 2003]

We define *cp-rank* of $A \in \mathbb{S}_n$ as minimum of the number of columns for CP-factorization B, that is

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

Notice that $\mathrm{cp}(A) = \infty$ if $A \notin \mathcal{CP}_n$. Also define *cp-plus-rank* of $A \in \mathbb{S}_n$, that is

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

Sometimes, we need to distinguish the completely positive matrices in either interior or boundary of $\mathcal{CP}^n$.

Theorem 2.2 [Dickinson, 2010, Theorem 3.8]

For $A \in \mathbb{S}_n$, we have $A \in \mathrm{int}\,(\mathcal{CP}^n) \iff \mathrm{cp}^+(A) < \infty$ and $\mathrm{rank}(A) = n$.

Existing Results on $\mathcal{CP}_n$: Essential Lemma

Observation: If we have had a CP-factorization B of A with r columns, then we can easily get another CP-factorization $\widehat{B}$ of A with r' columns for every positive integer $r' \geq r$. For instance, let $k := r' - r$, then

$$\widehat{B} := [B, 0_{n\times k}] \geq 0, \text{ and } \widehat{B}\widehat{B}^\top = BB^\top = A.$$

Lemma 2.3

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

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

Essential Lemma 2.4 [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 via Orthogonality Constrained Problem

1. Background
2. Existing Results on $\mathcal{CP}_n$
3. Our Approach: CP-factorization via Orthogonality Constrained Problem
   - CP-factorization as a nonconvex feasibility problem — Groetzner and Dür
   - LogSumExp: Smooth Approximation to Min Function
   - A curvilinear search method — Wen and Yin
4. Numerical Results

CP-factorization as a nonconvex feasibility problem

RECALL:

Lemma 2.3 $r \geq \mathrm{cp}(A) \iff A$ has a CP-factorization with r columns.

Lemma 2.4 $BB^T = CC^T \iff \exists X \in \mathcal{O}_r$ such that $BX = C$.

1. a "bad" factorization $B \ngeq 0$
   $A = BB^T$, but B not nonnegative. (spectral decomposition and zero appending to r)
   $\downarrow$ a suitable orthogonal matrix $X \in \mathcal{O}_r$
2. a "good" factorization $BX \geq 0$
   $A = (BX)(BX)^T$, meanwhile BX nonnegative.

CP-factorization as a nonconvex 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 $r \geq \mathrm{cp}(A)$, $B \in \mathbb{R}^{n \times r}$ is an arbitrary initial factorization $A = BB^T$ not nonnegative.
From Lemma 2.3 and 2.4 we have $A \in \mathcal{CP}_n \iff (1)$ is feasible.

Our Approach

[Groetzner and Dür, 2020] applied the so-called alternating projection method to (1).

Our Approach to nonconvex feasibility problem (1) [L, 2020]

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

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

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

$$\min\,(BX)_{ij} \xrightarrow{\text{approximate}} LSE_\rho\,(BX)$$

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 a differentiable function.

CP-factorization via Orthogonality Constrained Problem

LogSumExp: Smooth Approximation to Min Function

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

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

Lemma 3.1 — Basic properties of LSE

Suppose that $\rho < 0$, $\mathbf{x} \in \mathbb{R}^n$. The following results indicate that $LSE_\rho(\mathbf{x})$ nicely approximates the minimum function from below.

1. For all $\mathbf{x} \in \mathbb{R}^n$, we have $\lim_{\rho \to -\infty} LSE_\rho(\mathbf{x}) = \min x_i$.
2. For all $\mathbf{x} \in \mathbb{R}^n$, we have

$$\min x_i + \frac{1}{\rho} \log(n) \leq LSE_\rho(\mathbf{x}) < \min x_i,$$

i.e., the error item $\epsilon_{\mathbf{x}} := \min x_i - LSE_\rho(\mathbf{x})$ is in the interval $\left(0, -\frac{1}{\rho}\log(n)\right]$.

LogSumExp: Smooth Approximation to Min Function

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

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

Example 3.2: Values of $LSE_\rho\left(\mathbf{x}\right)$

n=4	$\rho = -1$	$\rho = -2$	$\rho = -3$	$\rho = -10$
$\mathbf{x}_1 = [5, 2, 6, 3]^t$	1.6381	1.9353	1.9838	1.9999
$\mathbf{x}_2 = [2, 2, 2, 2]^t$	0.6137	1.3068	1.5379	1.8613
$\epsilon_\rho := -1/\rho log(n)$	1.3863	0.6931	0.4621	0.1387

We find that the less ρ is, the better approximation effects.

LogSumExp: Smooth Approximation to Min Function

Lemma 3.3 — Basic properties of matrix LSE [L, 2020]

Suppose that $\rho < 0, B \in \mathbb{R}^{n\times r}$, matrix varible $X \in \mathbb{R}^{r\times r}$. The following results indicate that $LSE_\rho(BX) : \mathbb{R}^{r\times r} \to \mathbb{R}$ nicely approximates the $\min{(BX)_{ij}}$ from below.

1. For all $X \in \mathbb{R}^{r\times r}$, we have $\lim\limits_{\rho\to-\infty} LSE_\rho(BX) = \min(BX)_{ij}$.
2. For all $X \in \mathbb{R}^{r\times r}$, we have

$$\min(BX)_{ij} + \frac{1}{\rho}\log(nr) \le LSE_\rho(BX) < \min(BX)_{ij},$$

   i.e., the error item $\epsilon_X := \min(BX)_{ij} - LSE_\rho(BX)$ is in the interval $(0, -\frac{1}{\rho}\log(nr)]$.

LogSumExp: Smooth Approximation to Min Function

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

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

Proposition 3.4 [L, 2020]

Let t resp. t_ρ denote global maximum of problem (2) resp. (3), and $\epsilon_\rho := -\frac{1}{\rho}\log(nr) > 0$ is maximum error of $LSE_\rho\,(BX)$. Then,

$$0 < t - t_\rho \le \epsilon_\rho.$$

Proposition 3.5 [L, 2020]

If problem (3) has a feasible solution X such that $LSE_\rho\,(BX) \ge 0$, then

1. for the same X, $min\,(BX)_{ij} > 0$ for problem (2).
2. hence, we find a CP factorization $(BX)(BX)^T = A$ with $BX > 0$.

A curvilinear search method

Attempting to optimize (3), we apply the famous curvilinear search method, proposed by Wen and Yin [Wen and Yin, 2013], for general optimization with orthogonality constrains:

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

Let gradient $G := \mathcal{D}\mathcal{F}(X) = \left(\frac{\partial \mathcal{F}(X)}{\partial X_{i,j}}\right)$. The Lagrangian function is $\mathcal{L}(X, \Lambda) = \mathcal{F}(X) - \frac{1}{2} \operatorname{tr}\left(\Lambda\left(X^T X - I\right)\right)$ where $\Lambda \in \mathcal{S}^n$ is lagrangian multiplier.

Lemma 3.5 — First-order optimality conditions [Wen and Yin, 2013, Lemma 1.]

- Suppose that X is a local minimizer of problem above. Then X satisfies the first-order optimality conditions $\mathcal{D}_X \mathcal{L}(X, \Lambda) = G - XG^T X = 0$ and $X^T X = I$ with the associated lagrangian multiplier $\Lambda = G^T X$.
- Define

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

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

A curvilinear search method

Lemma 3.6 — 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 $Y(\tau)$ defined below satisfies $Y(\tau)^T Y(\tau) = X^T X$ and $Y(0) = X$,

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

   $Y(\tau) : \mathbb{R} \to \mathbb{R}^{n \times p}$ is a smooth curve.

2. If set $W = A := GX^T - XG^T$. Then $Y(\tau)$ is a descent curve at $\tau = 0$, that is

$$\mathcal{F}'_\tau(Y(0)) := \left.\frac{\partial \mathcal{F}(Y(\tau))}{\partial \tau}\right|_{\tau=0} = -\frac{1}{2}\|A\|_F^2.$$

   Since $\nabla \mathcal{F}(X) = 0$ if and only if $A = 0$, as long as X is yet to be local minimizer, then $\mathcal{F}'_\tau(Y(0)) < 0$.

Algorithm 1: CP-factorization via Orthogonality Constrained Problem

Data: Given $A \in \mathcal{CP}_n, r \geq \mathrm{cp}(A)$.

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 $0 < \theta_1 < \theta_2 < 1, \rho < 0, \epsilon > 0, k \leftarrow 0$;

while $\|\nabla \mathcal{F}(X_k)\| > \epsilon$ **do**

Generate $G_k \leftarrow -B^T \frac{\partial LSE_\rho(BX_k)}{\partial(BX_k)}, A_k \leftarrow G_k X_k^T - X_k G_k^T, W_k \leftarrow A_k$;

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

The global convergence of local minimizer is guaranteed in [Wen and Yin, 2013, Theorem 2], that is $\lim_{k\to\infty} \|\nabla \mathcal{F}\left(X_k\right)\|_F = 0$.

CP-factorization via Orthogonality Constrained Problem

Numerical Results

We investigate random instances were generated like [Groetzner and Dür, 2020, Section 7.7][1].

Table: **For every value of n and r, our approach has 100% success rate and is significantly faster than [Groetzner and Dür, 2020] even if our computer is very poor.**

		Our Approach[2] *Intel Core i7-4770 3.40 GHz and 16GB Ram Very poor environment!*		[Groetzner and Dür, 2020] *88 Intel Xenon ES-2699 cores (2.2 GHz each) and a total of 0.792 TB Ram*	
n	r	av. success rate (%)	av. time (sec.)	av. success rate (%)	av. time (sec.)
50	51	100	0.06	1.1	504
50	151	100	0.27	100	0.9
100	151	100	0.42	100	9.8
100	301	100	1.41	100	4
150	201	100	0.89	83.3	42
200	301	100	2.31	90.9	47

[1] Compute C by setting $C_{ij} := |B_{ij}|$ for all i, j, where B is a random $n \times k$ matrix based on Matlab command randn and finally we take $A = CC^T$ to be factorized. For randomly generated matrices A with several n and $k = 2n$. For $n \leq 50$, we used 100 starting points, for $n > 50$, we used 10 starting points. In each case, we used a maximum of 5000 iterations per starting point. The numbers in columns $3 - 6$ represent the average of 100 randomly generated instances.

[2] Some improvements in practice: Instead of Armijo-Wolfe conditions, we use well-known Barzilai-Borwein step. And a heuristic extension — decreasing of parameter ρ. Also we decide $\min\,(BX_{k+1})_{ij} \geq 0$ as stopping condition.

Reference I

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

[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.

[Dickinson, 2010] Dickinson, P. J. (2010).
An improved characterisation of the interior of the completely positive cone.
Electronic Journal of Linear Algebra, 20(1):51.

[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.

Reference II

[Dickinson and Gijben, 2014] Dickinson, P. J. and Gijben, L. (2014).
On the computational complexity of membership problems for the completely positive cone and its dual.
Computational optimization and applications, 57(2):403–415.

[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 a-truncated k-moment problem.
Foundations of Computational Mathematics, 14(6):1243–1276.

[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.

Reference III

[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.

Improvement: BB step and Heuristic extension

In practice, parameter ρ is not as small as possible. Although a smaller ρ gives a much tighter approximation, the speed of convergence, however, often slows.

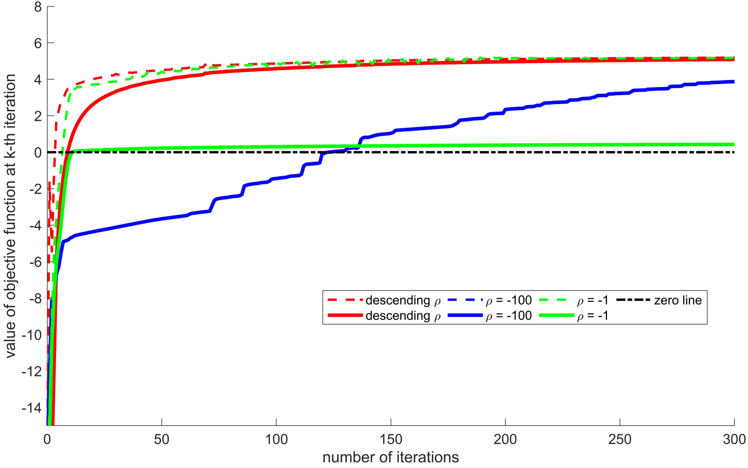

Figure: Value of objective function

Lemma [Bomze et al., 2015, Theorem 4.1]

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

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

$$\begin{aligned} & \text{find} && X \\ 1 \quad & \text{s.t.} && BX \ge 0 \\ & && X \in \mathcal{O}_r \end{aligned} \tag{4}$$

where $r \ge \mathrm{cp}(A)$, $B \in \mathbb{R}^{n \times r}$ is an arbitrary initial factorization $A = BB^T$ not nonnegative.
From Lemma 2.3 and 2.4 we have $A \in \mathcal{CP}_n \Longleftrightarrow (1)$ is feasible.

Proof of Lemma 3.1.

1. We first show an auxiliary result that is also applied for avoiding numerical overflow on computing value $LSE_\rho(\mathbf{x})$ in Algorithm 1. Given $\mathbf{x} \in \mathbb{R}^n$, we have

$$LSE_\rho(\mathbf{x}) = \frac{1}{\rho} \log \left(\sum_{i=1}^n \exp\left(\rho(x_i - c)\right)\right) + c$$

for all $c \in \mathbb{R}$. It follows from

$$\begin{aligned}
\frac{1}{\rho} \log \left(\sum_{i=1}^n \exp\left(\rho(x_i - c)\right)\right) + c &= \frac{1}{\rho} \log \left(\exp(-\rho c) \sum_{i=1}^n \exp\left(\rho x_i\right)\right) + c \\
&= \frac{1}{\rho} \log \left(\exp(-\rho c)\right) + \frac{1}{\rho} \log \left(\sum_{i=1}^n \exp\left(\rho x_i\right)\right) + c \\
&= \frac{1}{\rho} \log \left(\sum_{i=1}^n \exp\left(\rho x_i\right)\right).
\end{aligned}$$

In particular, if we let $c := \min x_i$, say x_j, then

$$\begin{aligned} LSE_\rho(\mathbf{x}) &= \frac{1}{\rho} \log \left(\sum_{i=1}^n \exp\left(\rho(x_i - x_j)\right)\right) + x_j \\ &= \frac{1}{\rho} \log \left(1 + \sum_{i \neq j}^n \exp\left(\rho(x_i - x_j)\right)\right) + x_j. \end{aligned} \tag{5}$$

Since for any $i \neq j$, $\rho(x_i - x_j) \leq 0$ implies $1 < 1 + \sum_{i \neq j}^n \exp\left(\rho(x_i - x_j)\right) \leq n$, then the term $\log\left(1 + \sum_{i \neq j}^n \exp\left(\rho(x_i - x_j)\right)\right)$ is bounded, thus $\lim_{\rho \to -\infty} LSE_\rho(\mathbf{x}) = x_j$.

2. Note that x_j denotes $\min x_i$. Due to the equation (5) and

$$\frac{1}{\rho} \log(n) \leq \frac{1}{\rho} \log\left(1 + \sum_{i \neq j}^n \exp\left(\rho(x_i - x_j)\right)\right) < 0,$$

we have $\frac{1}{\rho} \log(n) \leq LSE_\rho(\mathbf{x}) - x_j < 0$.

3. It is easily seen that the gradient of $LSE_\rho(\mathbf{x})$ is the so-called famous "softmax function", and the individual components of it is

$$\left(\frac{\partial LSE_\rho(\mathbf{x})}{\partial x_i}\right) = \frac{\exp(\rho x_i)}{\sum_{j=1}^{n} \exp(\rho x_j)} = \exp\{\rho(x_i - LSE_\rho(\mathbf{x}))\}. \tag{6}$$

Notice that we do not use the form $\exp(-\rho LSE_\rho(\mathbf{x}))\exp(\rho x_i)$ due to numerical underflow and overflow in Algorithm 1. We only prove the next property to complete the last proof. We declare that for any fixed $\mathbf{x} \in \mathbb{R}^n$, if we regard $LSE_\rho(\mathbf{x})$ as a map of variable $\rho \in (-\infty, 0)$, denoted by $LSE_{\mathbf{x}}(\rho)$, then

$$\frac{\partial LSE_{\mathbf{x}}(\rho)}{\partial \rho} < 0$$

for all $\rho \in (-\infty, 0)$. For convenience, we replace $LSE_{\mathbf{x}}(\rho)$ or $LSE_{\rho}(\mathbf{x})$ sometimes with LSE at next procedure. We have

$$\begin{aligned}\frac{\partial LSE_{\mathbf{x}}(\rho)}{\partial \rho} &= -\frac{1}{\rho^2}\log\left(\sum_{i=1}^n \exp\left(\rho x_i\right)\right) + \frac{1}{\rho}\frac{1}{\sum_{i=1}^n \exp\left(\rho x_i\right)}\left(\sum_{i=1}^n x_i \exp\left(\rho x_i\right)\right)\\ &= \frac{1}{\rho}\{-LSE + \exp(-\rho LSE)(\sum_{i=1}^n x_i \exp(\rho x_i))\}\\ &= \frac{1}{\rho}\{\sum_{i=1}^n x_i \exp\{\rho(x_i - LSE)\} - LSE\}\\ &= \frac{1}{\rho}\{\mathbf{x}^T \frac{\partial LSE}{\partial \mathbf{x}} - LSE\} < 0.\end{aligned}$$

For the last inequality, we observe from (6) that $\sum_{i=1}^n \frac{\partial LSE}{\partial x_i} = 1$ and every entry $\frac{\partial LSE}{\partial x_i} > 0$, hence the term $\mathbf{x}^T \frac{\partial LSE}{\partial \mathbf{x}}$ is a convex combination of all entries of $\mathbf{x}$, which implies that $\mathbf{x}^T \frac{\partial LSE}{\partial \mathbf{x}} \geq \min x_i > LSE$.

Proof of Proposition 3.4.

By definition of global maximum and Lemma 3.3 we have

$$LSE_\rho(BX^*) \leq LSE_\rho(BX^*_\rho) = t_\rho < \min\,(BX^*_\rho)_{ij} \leq \min\,(BX^*)_{ij} = t,$$

thus $0 < t - t_\rho$. And
$t - t_\rho = \min\,(BX^*)_{ij} - LSE_\rho(BX^*_\rho) \leq \min\,(BX^*)_{ij} - LSE_\rho\,(BX^*) \leq \epsilon_\rho.$

Consider the so-called copositive program (primal problem)

$$\min \{\langle C, X\rangle \mid \langle A_i, X\rangle = b_i \ (i = 1, \dots, m), X \in \mathcal{COP}_n\}, \tag{7}$$

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. Here $\mathcal{S}_n$ is the set of real symmetric $n \times n$ matrices, and the inner product of two matrices $\langle A, B\rangle := \mathrm{trace}(A^T B)$ as usual. The dual problem of (7) is

$$\max \{\textstyle\sum_{i=1}^m b_i y_i \mid C - \sum_{i=1}^m y_i A_i \in \mathcal{CP}_n,\ y_i \in \mathbb{R}\}, \tag{8}$$

where $\mathcal{CP}_n$ denotes the set of $n \times n$ completely positive matrices, which is a proper cone (i.e., closed, convex, pointed, and full dimensional) and also is the dual cone of $\mathcal{COP}_n$, cf. [Abraham and Naomi, 2003].