Primal and dual optimization problems and induced factorizations

Radu Balan

University of Maryland
Department of Mathematics and the Norbert Wiener Center
College Park, Maryland *rvbalan@umd.edu*

August 10, 2024

Santaló School 2024 , "Data Science, Signal Processing and Harmonic Analysis"
Buenos Aires, Argentina, August 5-9, 2024.

Down the memory lane ...

When did I first meet Carlos ?

Down the memory lane ...

When did I first meet Carlos ?
One day in Spring of 2005:

Down the memory lane ...

When did I first meet Carlos ?
One day in Spring of 2005:

Down the memory lane ...

When did I first meet Carlos ?
One day in Spring of 2005:

Carlos C., Ursula M., John Benedetto, Alex Powell, Joe Lakey , Mark Lammers

CIMPA 2013, Mar Del Plata

CIMPA 2013, Mar Del Plata

Happy Birthday Carlos!

Acknowledgments

This material is based upon work partially supported by the National Science Foundation under grant no. DMS-2108900 and by Simons Foundation. "Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

Collaborators:
Fushuai (Black) Jiang (UMD), Kasso Okoudjou (Tufts), Anirudha Poria(Bar-Ilan U.), Michael Rawson (UMD/PNNL), Yang Wang (HKUST), Rui Zhang (HKUST), Felix Krahmer (TUM).

Works:

1. R. Balan, K. Okoudjou, A. Poria, *On a Feichtinger Problem*, Operators and Matrices vol. 12(3), 881-891 (2018) http://dx.doi.org/10.7153/oam-2018-12-53
2. R. Balan, K.A. Okoudjou, M. Rawson, Y. Wang, R. Zhang, *Optimal l1 Rank One Matrix Decomposition*, in "Harmonic Analysis and Applications", Rassias M., Ed. Springer (2021)

Warm-Up Exercise

Let $A \in \mathbb{R}^{n \times n}$ be a matrix and consider the following optimization problem:

$$\gamma(A) := \inf_{A=\sum_{k\geq 1} x_k y_k^T} \sum_k \|x_k\|_1 \|y_k\|_1$$

Warm-Up Exercise

Let $A \in \mathbb{R}^{n\times n}$ be a matrix and consider the following optimization problem:

$$\gamma(A) := \inf_{A=\sum_{k\geq 1} x_k y_k^T} \sum_k \|x_k\|_1 \|y_k\|_1$$

Note:

$$A = [A_{i,j}]_{i,j\in[n]} = \begin{bmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{n1} \end{bmatrix} \cdot [1, 0, \cdots, 0] + \cdots + \begin{bmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{nn} \end{bmatrix} \cdot [0, 0, \cdots, 1]$$

From where: $\gamma(A) \leq \sum_{i,j} |A_{i,j}| =: \|A\|_1$.

Warm-Up Exercise

Let $A \in \mathbb{R}^{n\times n}$ be a matrix and consider the following optimization problem:

$$\gamma(A) := \inf_{A=\sum_{k\geq 1} x_k y_k^T} \sum_k \|x_k\|_1 \|y_k\|_1$$

Note:

$$A = [A_{i,j}]_{i,j\in[n]} = \begin{bmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{n1} \end{bmatrix} \cdot [1, 0, \cdots, 0] + \cdots + \begin{bmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{nn} \end{bmatrix} \cdot [0, 0, \cdots, 1]$$

From where: $\gamma(A) \leq \sum_{i,j} |A_{i,j}| =: \|A\|_1$.
For converse: Let $A = \sum_k x_k y_k^T$ be the optimal decomposition. Then:

$$\|A\|_1 = \|\sum_k x_k y_k^T\|_1 \leq \sum_k \|x_k y_k^T\|_1 = \sum_k \|x_k\|_1 \|y_k\|_1 = \gamma(A).$$

Projective norm

We obtained:

$$\gamma(A) := \inf_{A=\sum_{k\geq 1} x_k y_k^T} \sum_k \|x_k\|_1 \|y_k\|_1 = \sum_{i,j} |A_{i,j}| =: \|A\|_1$$

folowing Grothendieck, the last norm is sometime referred to as *projective norm*, $\|A\|_\wedge$.

Projective norm

We obtained:

$$\gamma(A) := \inf_{A=\sum_{k\geq 1} x_k y_k^T} \sum_k \|x_k\|_1 \|y_k\|_1 = \sum_{i,j} |A_{i,j}| =: \|A\|_1$$

folowing Grothendieck, the last norm is sometime referred to as *projective norm*, $\|A\|_\wedge$.

Assume now that $A = A^T$. Considered a more constrained optimization problem:

$$\gamma_0(A) := \inf_{\substack{A=\sum_{k\geq 1} \varepsilon_k x_k x_k^T \\ \varepsilon_k \in \{+1,-1\}}} \sum_k \|x_k\|_1^2$$

Projective norm

We obtained:

$$\gamma(A) := \inf_{A=\sum_{k\geq 1} x_k y_k^T} \sum_k \|x_k\|_1 \|y_k\|_1 = \sum_{i,j} |A_{i,j}| =: \|A\|_1$$

folowing Grothendieck, the last norm is sometime referred to as *projective norm*, $\|A\|_\wedge$.
Assume now that $A = A^T$. Considered a more constrained optimization problem:

$$\gamma_0(A) := \inf_{\substack{A=\sum_{k\geq 1} \varepsilon_k x_k x_k^T \\ \varepsilon_k \in \{+1,-1\}}} \sum_k \|x_k\|_1^2$$

It is not hard to show that γ_0 is a norm on $Sym(\mathbb{R}^n)$ (or $Sym(\mathbb{C}^n)$), and $\|A\|_1 \leq \gamma_0(A)$.

Projective norm

We obtained:

$$\gamma(A) := \inf_{A=\sum_{k\geq 1} x_k y_k^T} \sum_k \|x_k\|_1 \|y_k\|_1 = \sum_{i,j} |A_{i,j}| =: \|A\|_1$$

folowing Grothendieck, the last norm is sometime referred to as *projective norm*, $\|A\|_\wedge$.

Assume now that $A = A^T$. Considered a more constrained optimization problem:

$$\gamma_0(A) := \inf_{\substack{A = \sum_{k\geq 1} \varepsilon_k x_k x_k^T \\ \varepsilon_k \in \{+1,-1\}}} \sum_k \|x_k\|_1^2$$

It is not hard to show that γ_0 is a norm on $Sym(\mathbb{R}^n)$ (or $Sym(\mathbb{C}^n)$), and $\|A\|_1 \leq \gamma_0(A)$. Leveraging the fact that $\frac{1}{2}(xy^T + yx^T) = \frac{1}{4}((x+y)(x+y)^T - (x-y)(x-y)^T)$ one obtains:

$$\|A\|_1 \leq \gamma_0(A) := \inf_{\substack{A = \sum_{k\geq 1} \varepsilon_k x_k x_k^T \\ \varepsilon_k \in \{+1,-1\}}} \sum_k \|x_k\|_1^2 \leq 2\|A\|_1$$

Problem Formulation

Let $Sym^+(\mathbb{C}^n) = \{A \in \mathbb{C}^{n\times n} \ , \ A^* = A \geq 0\}$. For $A \in Sym^+(\mathbb{C}^n)$, denote

$$\gamma_+(A) := \inf_{A=\sum_{k\geq 1} x_x x_k^*} \sum_k \|x_k\|_1^2$$

It is obvious that $\|A\|_1 \leq \gamma_0(A) \leq \gamma_+(A)$.

Problem Formulation

Let $Sym^+(\mathbb{C}^n) = \{A \in \mathbb{C}^{n \times n}\ ,\ A^* = A \geq 0\}$. For $A \in Sym^+(\mathbb{C}^n)$, denote

$$\gamma_+(A) := \inf_{A=\sum_{k\geq 1} x_x x_k^*} \sum_k \|x_k\|_1^2$$

It is obvious that $\|A\|_1 \leq \gamma_0(A) \leq \gamma_+(A)$.
The *matrix problem*: For every $n \geq 1$ find the best constant C_n such that, for every $A \in Sym^+(\mathbb{C}^n)$,

$$\gamma_+(A) \leq C_n \|A\|_1 := C_n \sum_{k,l=1}^{n} |A_{k,l}|$$

That is, we are interested in finding:

$$C_n = \sup_{A \geq 0} \frac{\gamma_+(A)}{\|A\|_1}$$

Properties of $\gamma_+(A)$

The infimum is achieved:

$$\gamma_+(A) := \inf_{A=\sum_{k\geq 1} x_x x_k^*} \sum_k \|x_k\|_1^2 = \min_{A=\sum_{k=1}^{n^2} x_x x_k^*} \sum_k \|x_k\|_1^2.$$

Upper bounds:

$$\gamma_+(A) \leq n\, trace(A) \leq n\|A\|_1 = n \sum_{k,j} |A_{k,j}|$$

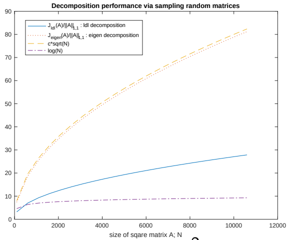
$$\gamma_+(A) \leq n\, trace(A) \leq n^2 \|A\|_{Op}$$

Lower bounds:

$$\|A\|_1 = \min_{A=\sum_{k\geq 1} x_x y_k^*} \sum_k \|x_k\|_1 \|y_k\|_1 \leq \gamma_+(A)$$

Convexity: for $A, B \in Sym^+(\mathbb{C}^n)$ and $t \geq 0$,

$$\gamma_+(A+B) \leq \gamma_+(A) + \gamma_+(B) \quad , \quad \gamma_+(tA) = t\gamma_+(A)$$



Maximum of $\sum_k \|x_k\|_1^2 / \|A\|_1$ over 30 random noise realizations, where $x_k's$ are obtained from the eigendecomposition, or the LDL factorization.

Properties of $\gamma_+(A)$

Lower bound is achieved, $\gamma_+(A) = \|A\|_1$ in the following cases:

1. If $A = xx^*$ is of rank one.
2. If $A \geq 0$ is a diagonally dominant matrix, $A_{ii} \geq \sum_{k \neq i} |A_{i,k}|$.
3. If $A \geq 0$ admits a Non Negative Matrix Factorization (NNMF), $A = BB^T$ with $B_{ij} \geq 0$.

Continuity, Lipschitz and linear program reformulation:

1. $\gamma_+ : Sym^+(\mathbb{C}^n) \to \mathbb{R}$ is continuous.
2. If $A, B \geq \delta I$ and $trace(A), trace(B) \leq 1$ then

$$|\gamma_+(A) - \gamma_+(B)| \leq \left(\frac{n}{\delta^2} + n^2\right) \|A - B\|_{Op}.$$

3. Let $S_1 = \{x \in \mathbb{C}^n \ , \ \|x\|_1 = 1\}$ denote the compact unit sphere with respect to the l^1 norm, and let $\mathcal{B}(S_1)$ denote the set of Borel measures over S_1. Then:

$$\gamma_+(A) = \inf_{\mu \in \mathcal{B}(S_1): \int_{S_1} xx^* d\mu(x) = A} \mu(S_1) \ , \ \mu^*(x) = \sum_{k=1}^{m} \lambda_k \delta(x - g_k)$$

where $\gamma_+(A) = \sum_{k=1}^m \lambda_k$ and $A = \sum_{k=1}^m \lambda_k g_k g_k^*$ is the optimal factorization.

Primal and dual problems for γ_+

The linear program is a convex optimization problem (which is great), but it is defined on an infinite dimensional space (not so great!).
Its dual problem enjoys strong duality (this may not be obvious due to infinite dimensional technical issues):

Theorem

Assume $A \geq 0$. Its associated primal (min) & dual (max) problems are:

$$\max_{T=T^*:\langle Tx,x\rangle\leq 1\ ,\ \forall \|x\|_1\leq 1} trace(TA) = \min_{\mu\in\mathcal{B}(S_1):\int_{S_1} xx^*\,d\mu(x)=A} \mu(S_1) = \gamma_+(A)$$

Primal and dual problems for γ_+

The linear program is a convex optimization problem (which is great), but it is defined on an infinite dimensional space (not so great!).
Its dual problem enjoys strong duality (this may not be obvious due to infinite dimensional technical issues):

Theorem

Assume $A \geq 0$. Its associated primal (min) & dual (max) problems are:

$$\max_{T=T^*:\langle Tx,x\rangle\leq 1\ ,\ \forall \|x\|_1\leq 1} trace(TA) = \min_{\mu\in\mathcal{B}(S_1):\int_{S_1} xx^*\,d\mu(x)=A} \mu(S_1) = \gamma_+(A)$$

Note the quantity:

$$\rho_1(T) = \max_{x:\|x\|_1\leq 1} \langle Tx, x\rangle$$

The dual problem and C_n turn into:

$$\max_{T=T^*:\rho_1(T)\leq 1} trace(TA)$$

$$C_n = \max_{A\geq 0:\|A\|_1\leq 1} \gamma_+(A) = \max_{\substack{A\geq 0:\\ \|A\|_1\leq 1}} \max_{\substack{T=T^*:\\ \rho_1(T)\leq 1}} trace(TA) = \max_{\substack{A\geq 0:\\ A\neq 0}} \max_{\substack{T=T^*:\\ \rho_1(T)>0}} \frac{trace(TA)}{\|A\|_1\,\rho_1(T)}$$

The bound ρ_1

Recall, for $T = T^*$:

$$\rho_1(T) = \max_{x:\|x\|_1 \leq 1} \langle Tx, x \rangle$$

How to compute it?
Easy cases:

1. If $T \leq 0$ then $\rho_1(T) = 0$
2. If $T \geq 0$ then

$$\rho_1(T) = \max_k T_{k,k} = \max_{i,j} |T_{i,j}| =: \|T\|_\infty$$

This resembles the *numerical radius* of a matrix, $r(T) = \max_{\|x\|_2=1} |\langle Tx, x \rangle|$, which for hermitian matrices equals the largest singular value (operator norm).
Note differences: (i) $\|\cdot\|_2 \to \|\cdot\|_1$; (ii) no absolute value $|.|$.

The bound ρ_1 (2)

Assume $\lambda_{max}(T) > 0$, i.e. T is NOT negative semi-definite. Then:

$$\rho_1(T) = \max_{x:\|x\|_1=1} \langle Tx, x\rangle = \max_{\substack{A \geq 0 : \\ rank(A)=1 \\ \|A\|_1 = 1}} trace(TA) = \max_{\substack{A \geq 0 : \\ rank(A)=1 \\ \|A\|_1 \leq 1}} trace(TA)$$

The bound ρ_1 (2)

Assume $\lambda_{max}(T) > 0$, i.e. T is NOT negative semi-definite. Then:

$$\rho_1(T) = \max_{x:\|x\|_1=1} \langle Tx, x\rangle = \max_{\substack{A \geq 0 : \\ rank(A)=1 \\ \|A\|_1 = 1}} trace(TA) = \max_{\substack{A \geq 0 : \\ rank(A)=1 \\ \|A\|_1 \leq 1}} trace(TA)$$

Convex relaxation:

$$\pi_+(T) := \max_{\substack{A \geq 0 : \\ \|A\|_1 \leq 1}} trace(TA)$$

which is a semi-definite program (SDP). Thus:

$$\rho_1(T) \leq \pi_+(T).$$

Primal and dual problems for ρ_1

The SDP enjoys strong duality:

Theorem

Assume $T = T^*$. The primal-dual programs have strong duality:

$$\pi_+(T) = \max_{\substack{A \geq 0 : \\ \|A\|_1 \leq 1}} trace(TA) = \min_{Y \geq 0} \|T + Y\|_\infty$$

where $\|Z\|_\infty = \max_{i,j} |Z_{i,j}|$.

The proof of this theorem is based on the Von Neumann's min-max theorem:

$$\min_{Y \geq 0} \|T + Y\|_\infty = \min_{Y \geq 0} \max_{A:\|A\|_1 \leq 1} trace((T + Y)A) \stackrel{vN}{=} \max_{A:\|A\|_1 \leq 1} \min_{Y \geq 0} trace((T + Y)A) =$$

$$= \max_{A:\|A\|_1 \leq 1} \left(trace(TA) + \min_{Y \geq 0} trace(YA) \right) = \max_{A \geq 0:\|A\|_1 \leq 1} \left(trace(TA) + \min_{Y \geq 0} trace(YA) \right) =$$

$$= \max_{A \geq 0:\|A\|_1 \leq 1} trace(TA) = \pi_+(T)$$

Closing the loop

The final result: the connection between $\gamma_+(A)$ and C_n on one hand, and $\rho_1(T)$ and $\pi_+(T)$ on the other hand:

Theorem

$$C_n := \max_{\substack{A \geq 0 \\ A \neq 0}} \frac{\gamma_+(A)}{\|A\|_1} = \max_{\substack{T = T^* \\ \rho_1(T) \neq 0}} \frac{\pi_+(T)}{\rho_1(T)}$$

The proof is based on an earlier derivation:

$$C_n = \max_{\substack{A \geq 0 : \\ A \neq 0}} \max_{\substack{T = T^* : \\ \rho_1(T) > 0}} \frac{trace(TA)}{\|A\|_1 \rho_1(T)} = \max_{\substack{T = T^* : \\ \rho_1(T) > 0}} \max_{\substack{A \geq 0 : \\ A \neq 0}} \frac{trace(TA)}{\|A\|_1 \rho_1(T)} =$$

$$= \max_{\substack{T = T^* : \\ \rho_1(T) > 0}} \frac{1}{\rho_1(T)} \max_{\substack{A \geq 0 : \\ \|A\|_1 = 1}} trace(TA) = \max_{\substack{T = T^* : \\ \rho_1(T) > 0}} \frac{\pi_+(T)}{\rho_1(T)}$$

Latest Result

Theorem (Afonso Bandeira, Dustin Mixon, Stefan Steinerberger - Oberwolfach 2024; ACHA 2024)

There are $\alpha > 0$, $N_0 > 1$ so that for any $n \geq N_0$,

$$C_n \geq \alpha\sqrt{n}$$

Happy Birthday Carlos!

Happy Birthday Carlos!

Thank you for listening! ... QUESTIONS?

Motivation

A Feichtinger Problem

At a 2004 Oberwolfach meeting, Hans Feichtinger asked the following question:
(Q1) Given a positive semi-definite trace-class operator $T : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$, $Tf(x) = \int K(x,y)f(y)dy$, with $K \in M^1(\mathbb{R}^d \times \mathbb{R}^d)$, and its spectral factorization, $T = \sum_k \langle \cdot, h_k \rangle h_k$, must it be $\sum_k \|h_k\|_{M^1}^2 < \infty$?

A modified version of the question is:
(Q2) Given T as before, i.e., $T = T^* \geq 0$, $K \in M^1(\mathbb{R}^d \times \mathbb{R}^d)$, is there a factorization $T = \sum_k \langle \cdot, g_k \rangle g_k$ such that $\sum_k \|g_k\|_{M^1}^2 < \infty$?

Problem Reformulation

Matrix Language

Consider an infinite matrix $A = (A_{m,n})_{m,n\geq 0}$ so that $\|A\|_\wedge := \|A\|_1 := \sum_{m,n\geq 0} |A_{m,n}| < \infty$.
This implies that A acts on $l^2(\mathbb{N})$ as a trace-class compact operator.
Assume additionally $A = A^* \geq 0$ as a quadratic form.
Let $(e_k)_{k\geq 0}$ denote an orthogonal set of eigenvectors normalized so that $A = \sum_{k\geq 0} e_k e_k^*$. It is easy to check that $e_k \in l^1(\mathbb{N})$, for each k.
Equivalent reformulations of the two problems (Heil, Larson '08):

Problem Reformulation

Matrix Language

Consider an infinite matrix $A = (A_{m,n})_{m,n\geq 0}$ so that $\|A\|_\wedge := \|A\|_1 := \sum_{m,n\geq 0} |A_{m,n}| < \infty$.
This implies that A acts on $l^2(\mathbb{N})$ as a trace-class compact operator.
Assume additionally $A = A^* \geq 0$ as a quadratic form.
Let $(e_k)_{k\geq 0}$ denote an orthogonal set of eigenvectors normalized so that $A = \sum_{k\geq 0} e_k e_k^*$. It is easy to check that $e_k \in l^1(\mathbb{N})$, for each k.
Equivalent reformulations of the two problems (Heil, Larson '08):
Q1: Does it hold $\sum_{k\geq 0} \|e_k\|_1^2 < \infty$?

Problem Reformulation

Matrix Language

Consider an infinite matrix $A = (A_{m,n})_{m,n\geq 0}$ so that $\|A\|_{\wedge} := \|A\|_1 := \sum_{m,n\geq 0} |A_{m,n}| < \infty$.
This implies that A acts on $l^2(\mathbb{N})$ as a trace-class compact operator.
Assume additionally $A = A^* \geq 0$ as a quadratic form.
Let $(e_k)_{k\geq 0}$ denote an orthogonal set of eigenvectors normalized so that $A = \sum_{k\geq 0} e_k e_k^*$. It is easy to check that $e_k \in l^1(\mathbb{N})$, for each k.
Equivalent reformulations of the two problems (Heil, Larson '08):
Q1: Does it hold $\sum_{k\geq 0} \|e_k\|_1^2 < \infty$? Answer: Negative in general! (see [1])

Problem Reformulation

Matrix Language

Consider an infinite matrix $A = (A_{m,n})_{m,n\geq 0}$ so that $\|A\|_\wedge := \|A\|_1 := \sum_{m,n\geq 0} |A_{m,n}| < \infty$.
This implies that A acts on $l^2(\mathbb{N})$ as a trace-class compact operator.
Assume additionally $A = A^* \geq 0$ as a quadratic form.
Let $(e_k)_{k\geq 0}$ denote an orthogonal set of eigenvectors normalized so that $A = \sum_{k\geq 0} e_k e_k^*$. It is easy to check that $e_k \in l^1(\mathbb{N})$, for each k.
Equivalent reformulations of the two problems (Heil, Larson '08):

Q1: Does it hold $\sum_{k\geq 0} \|e_k\|_1^2 < \infty$? Answer: Negative in general! (see [1])

Q2: Is there a factorization $A = \sum_{k\geq 0} f_k f_k^*$ so that $\sum_{k\geq 0} \|f_k\|_1^2 < \infty$?

Problem Reformulation

Matrix Language

Consider an infinite matrix $A = (A_{m,n})_{m,n\geq 0}$ so that $\|A\|_\wedge := \|A\|_1 := \sum_{m,n\geq 0} |A_{m,n}| < \infty$.
This implies that A acts on $l^2(\mathbb{N})$ as a trace-class compact operator.
Assume additionally $A = A^* \geq 0$ as a quadratic form.
Let $(e_k)_{k\geq 0}$ denote an orthogonal set of eigenvectors normalized so that $A = \sum_{k\geq 0} e_k e_k^*$. It is easy to check that $e_k \in l^1(\mathbb{N})$, for each k.
Equivalent reformulations of the two problems (Heil, Larson '08):
Q1: Does it hold $\sum_{k\geq 0} \|e_k\|_1^2 < \infty$? Answer: Negative in general! (see [1])

Q2: Is there a factorization $A = \sum_{k\geq 0} f_k f_k^*$ so that $\sum_{k\geq 0} \|f_k\|_1^2 < \infty$?

Using previous equivalence and some functional analysis arguments:

Proposition

If (Q2) is answered affirmatively, then the matrix conjecture must be true.

Linear program result

Optimal Factorization from a Measure Theory Perspective

Let $S_1 = \{x \in \mathbb{C}^n \ , \ \|x\|_1 = 1\}$ denote the compact unit sphere with respect to the l^1 norm, and let $\mathcal{B}(S_1)$ denote the set of Borel measures over S_1. For $A \in \mathrm{Sym}(\mathbb{C}^n)^+(\mathbb{C}^n)$ consider the optimization problem:

$$(p^*, \mu^*) = inf_{\mu \in \mathcal{B}(S_1): \int_{S_1} xx^* d\mu(x) = A} \ \mu(S_1) \quad (M)$$

Theorem (Optimal Measure)

For any $A \in Sym^+(\mathbb{C}^n)$ the optimization problem (M) is convex and its global optimum (minimum) is achieved by

$$p^* = \gamma_+(A) \ \ , \ \ \mu^*(x) = \sum_{k=1}^m \lambda_k \delta(x - g_k)$$

where $A = \sum_{k=1}^m (\sqrt{\lambda_k} g_k)(\sqrt{\lambda_k} g_k)^*$ is an optimal decomposition that achieves $\gamma_+(A) = \sum_{k=1}^m \lambda_k$.

Super-resolution and Convex Optimizations

$$\gamma_+(A) = \min_{x_1,\ldots,x_m \,:\, A=\sum_k x_k x_k^*} \sum_{k=1}^{m} \|x_k\|_1^2 \;,\;\; m = n^2 \quad (P)$$

$$p^* = \inf_{\mu \in \mathcal{B}(S_1) \,:\, A=\int_{S_1} xx^* d\mu(x)} \int_{S_1} d\mu(x) \quad (M)$$

Remarks

1. *The optimization problem (P) is non-convex, but finite-dimensional. The optimization problem (M) is convex, but infinite-dimensional.*
2. *If $g_1, ..., g_m \in S_1$ in the support of μ^* are known so that $\mu^* = \sum_{k=1}^{m} \lambda_k \delta(x - g_k)$, then the optimal $\lambda_1, ..., \lambda_m \geq 0$ are determined by a linear program. More general, (M) is an infinite-dimensional linear program.*
3. *Finding the support of μ^* is an example of a super-resolution problem. One possible approach is to choose a redundant dictionary (frame) that includes the support of μ^*, and then solve the induced linear program.*

Proof of the Optimal Measure Result

Recall: we want to show the following problems admit same solution:

$$\gamma_+(A) = \min_{x_1,\ldots,x_m \,:\, A=\sum_k x_k x_k^*} \sum_{k=1}^{m} \|x_k\|_1^2 \;,\;\; m = n^2 \quad (P)$$

$$p^* = \inf_{\mu\in\mathcal{B}(S_1)\,:\,A=\int_{S_1} xx^*\,d\mu(x)} \int_{S_1} d\mu(x) \quad (M)$$

a. Assume $A=\sum_{k=1}^m x_k x_k^*$ is a global minimum for (P). Then $\mu(x)=\sum_{k=1}^m \|x_k\|_1^2\delta(x-\frac{x_k}{\|x_k\|_1})$ is a feasible solution for (M). This shows $p^*\leq\gamma_+(A)$.

b. For reverse: Let μ^* be an optimal measure in (M). Fix $\varepsilon>0$. Construct a disjoint partition $(U_l)_{1\leq l\leq L}$ of S_1 so that each U_l is included in some ball $B_\varepsilon(z_l)$ of radius ε with $\|z_l\|_1=1$. Thus $U_l\subset B_\varepsilon(z_l)\cap S_1$.

For each l, compute $x_l=\frac{1}{\mu^*(U_l)}\int_{U_l} x\,d\mu^*(x)\in B_\varepsilon(z_l)$. Let $g_l=\sqrt{\mu^*(U_l)}x_l$.

Proof: The Optimal Measure Result (cont)

Key inequality:

$$0 \leq R_l := \int_{U_l} (x - x_l)(x - x_l)^* \, d\mu^*(x) = \int_{U_l} xx^* \, d\mu^*(x) \; - \mu^*(U_l) x_l x_l^*$$

Sum over l and with $R = \sum_{l=1}^{L} R_l$ get

$$A = \sum_{l=1}^{L} \int_{U_l} xx^* \, d\mu^*(x) \leq \sum_{l=1}^{L} g_l g_l^* + R$$

By sub-additivity and homogeneity:

$$\gamma_+(A) \leq \sum_{l=1}^{L} \|g_l\|_1^2 + \gamma_+(R) \leq \sum_{l=1}^{L} \mu^*(U_l) \|x_l\|_1^2 + n \, trace(R)$$

But $\|x_l - z_l\|_1 \leq \varepsilon$ and $\|x - x_l\|_1 \leq 2\varepsilon$ for every $x \in U_l$. Hence $\|x_l\|_1 \leq 1 + \varepsilon$ and $trace(R_l) \leq 4\mu^*(U_l)\varepsilon^2$. (In fact, $\|x_l\|_1 \leq 1$ by triangle inequality)

Proof: The Optimal Measure Result (end)

Thus:

$$\gamma_+(A) \leq \mu^*(S_1) + (2\varepsilon + \varepsilon^2 + 4n\varepsilon^2)\mu^*(S_1)$$

Since $\varepsilon > 0$ is arbitrary, it follows

$$\gamma_+(A) \leq \mu^*(S_1) = p^*$$

This ends the proof of the measure result. $\square$

Second New Result: The Continuity Property

Theorem (The Continuity Property)

The map $\gamma_+ : (Sym^+(\mathbb{C}^n), \|\cdot\|) \to \mathbb{R}$ *is continuous.*

Remarks

1. This statement extends the continuity result from $Sym^{++}(\mathbb{C}^n) = \{A = A^* > 0\}$ to $Sym^+(\mathbb{C}^n) = \{A = A^* \geq 0\}$.
2. Proof is based on a (new?) comparison result between non-negative operators.
3. Global Lipschitz is still open.

The Continuity Property

The proof is based on the following two lemmas:

Lemma (L1)

Let $A \in Sym^+(\mathbb{C}^n)$ of rank $r > 0$. Let $\lambda_r > 0$ denote the r^{th} eigenvalue of A, and let $P_{A,r}$ denote the orthogonal projection onto the range of A. For any $0 < \varepsilon < 1$ and $B \in Sym^+(\mathbb{C}^n)$ such that $\|A - B\|_{Op} \leq \frac{\varepsilon\lambda_r}{1-\varepsilon}$, the following holds true:

$$A - (1-\varepsilon)P_{A,r}BP_{A,r} \geq 0 \qquad (1)$$

Lemma (L2)

Let $A \in Sym^+(\mathbb{C}^n)$ of rank $r > 0$. Let $\lambda_r > 0$ denote the r^{th} eigenvalue of A. For any $0 < \varepsilon < \frac{1}{2}$ and $B \in Sym^+(\mathbb{C}^n)$ such that $\|A - B\|_{Op} \leq \varepsilon\lambda_r$, the following holds true:

$$B - (1-\varepsilon)P_{B,r}AP_{B,r} \geq 0 \qquad (2)$$

where $P_{B,r}$ denotes the orthogonal projection onto the top r eigenspace of B.

Proof of Continuity of γ_+

Fix $A \in Sym^+(\mathbb{C}^n)$. Let $(B_j)_{j\geq 1}$, $B_j \in Sym^+(\mathbb{C}^n)$, be a convergent sequence to A. We need to show $\gamma_+(B_j) \to \gamma_+(A)$.

Let $A = \sum_{k=1}^{n^2} x_k x_k^*$ be the optimal decomposition of A such that $\gamma_+(A) = \sum_{k=1}^{n^2} \|x_k\|_1^2$.

If $A = 0$ then $\gamma_+(A) = 0$ and

$$0 \leq \gamma_+(B_j) \leq n\, trace(B_j) \leq n^2 \|B_j\|_{Op}.$$

Hence $\lim_j \gamma_+(B_j) = 0$.

Assume $rank(A) = r > 0$ and let $\lambda_r > 0$ denote the smallest strictly positive eigenvalue of A. Let $\varepsilon \in (0, \frac{1}{2})$ be arbitrary. Let $J = J(\varepsilon)$ be so that $\|A - B_j\|_{Op} < \varepsilon \lambda_r$ for all $j > J$. Let $B_j = \sum_{k=1}^{n^2} y_{j,k} y_{j,k}^*$ be the optimal decomposition of B_j such that $\gamma_+(B_j) = \sum_{k=1}^{n^2} \|y_{j,k}\|_1^2$.

Let $\Delta_j = A - (1-\varepsilon) P_{A,r} B_j P_{A,r}$. By Lemma L1, for any $j > J$,

$$\gamma_+(A) \leq (1-\varepsilon)\gamma_+(P_{A,r} B_j P_{A,r}) + \gamma_+(\Delta_j) \leq (1-\varepsilon) \sum_{k=1}^{n^2} \|P_{A,r} y_{j,k}\|_1^2 + n\, trace(\Delta_j)$$

Proof of Continuity of γ_+ (cont)

Pass to a subsequence j' of j so that $y_{j',k} \to y_k$, for every $k \in [n^2]$, and $\gamma_+(B_{j'}) \to \liminf_j \gamma_+(B_j)$. Then $\lim_{j'} P_{A,r} y_{j',k} = P_{A,r} y_k = y_k$ and

$$\lim_{j'} \sum_{k=1}^{n^2} \|P_{A,r} y_{j',k}\|_1^2 = \lim_{j'} \sum_{k=1}^{n^2} \|y_{j',k}\|_1^2 = \liminf_j \gamma_+(B_j)$$

On the other hand, $\lim_j trace(\Delta_j) = \varepsilon\, trace(A)$. Hence:

$$\gamma_+(A) \leq (1-\varepsilon) \liminf_j \gamma_+(B_j) + \varepsilon\, trace(A)$$

Since $\varepsilon > 0$ is arbitrary, it follows $\gamma_+(A) \leq \liminf_j \gamma_+(B_j)$.
The inequality $\limsup_j \gamma_+(B_j) \leq \gamma_+(A)$ follows from Lemma L2 similarly: with $\Delta_j = B_j - (1-\varepsilon) P_{B_j,r} A P_{B_j,r}$ and $A = \sum_{k=1}^{n^2} x_k x_k^*$ optimal,

$$\gamma_+(B_j) \leq (1-\varepsilon)\gamma_+(P_{B_j,r} A P_{B_j,r}) + n\, trace(\Delta_j) = (1-\varepsilon) \sum_{k=1}^{n^2} \|P_{B_j,r} x_k\|_1^2 + n\, trace(\Delta_j).$$

Next take limsup of lhs by noticing $P_{B_j,r} \to P_{A,r}$ and $\limsup_j \|\Delta_j\|_{Op} = \varepsilon \|A\|_{Op}$:
$\limsup_j \gamma_+(B_j) \leq (1-\varepsilon)\gamma_+(A) + n^2 \varepsilon \|A\|_{Op}$. Take $\varepsilon -> 0$ and result follows. □

Proof of Lemmas

Proof of Lemma L1

Let $P = P_{A,r}$. and $\Delta = A - (1-\varepsilon)P_{A,r}BP_{A,r}$. For any $x \in \mathbb{C}^n$:

$$\langle \Delta x, x\rangle = \langle APx, Px\rangle - (1-\varepsilon)\langle BPx, Px\rangle = \langle (A-(1-\varepsilon)B)Px, Px\rangle =$$

$$= \varepsilon\langle APx, Px\rangle + (1-\varepsilon)\langle (A-B)Px, Px\rangle \geq \varepsilon\lambda_r\|Px\|^2 - (1-\varepsilon)\|A-B\|_{Op}\|Px\|^2 \geq 0$$

because $\|A-B\|_{Op} \leq \frac{\varepsilon\lambda_r}{1-\varepsilon}$.

Proof of Lemma L2

Let $P = P_{B,r}$ and $\Delta = B - (1-\varepsilon)P_{B,r}AP_{B,r}$. Let $C = B - P_{B,r}BP_{B,r} \geq 0$. Let μ_r be the r^{th} eigenvalue of B. Note $|\mu_r - \lambda_r| \leq \|A-B\|_{Op} \leq \varepsilon\lambda_r$. Thus $\mu_r \geq (1-\varepsilon)\lambda_r$. For any $x \in \mathbb{C}^n$:

$$\langle \Delta x, x\rangle = \langle Cx, x\rangle + \langle BPx, Px\rangle - (1-\varepsilon)\langle APx, Px\rangle = \langle Cx, x\rangle + \varepsilon\langle BPx, Px\rangle +$$

$$+(1-\varepsilon)\langle (B-A)Px, Px\rangle \geq \langle Cx, x\rangle + (\varepsilon\mu_r - (1-\varepsilon)\|A-B\|_{Op})\|Px\|^2 \geq 0$$

because $\|A-B\|_{Op} \leq \varepsilon\lambda_r \leq \frac{\varepsilon\mu_r}{1-\varepsilon}$.

Third new result: Strong duality for γ_+

Theorem

For every $A \geq 0$,

$$\max_{\substack{T = T^* \\ \langle Tx, x\rangle \leq 1 \ , \ \forall \|x\|_1 \leq 1}} trace(TA) = \min_{\substack{\mu \in \mathcal{B}(S_1) \\ \int_{S_1} xx^* d\mu(x) = A}} \mu(S_1) = \gamma_+(A)$$

Proof [Fushuai "Black" Jiang]

The second equality was established earlier as a "super-resolution" result.

For the first equality:

1. Let $A = \sum_{k=1}^m x_k x_k^*$ be its optimal decomposition such that $\gamma_+(A) = \sum_{k=1}^m \|x_k\|_1^2$, and let $T = T^*$ be a generic matrix so that $\langle Ty, y\rangle \leq 1$ for all $\|y\|_1 \leq 1$. Denote $y_k = \frac{x_k}{\|x_k\|_1}$. Then

$$trace(TA) = \sum_{k=1}^m \langle Tx_k, x_k\rangle = \sum_{k=1}^m \|x_k\|_1^2 \langle Ty_k, y_k\rangle \leq \sum_{k=1}^m \|y_k\|_1^2 = \gamma_+(A)$$

Proof of strong duality for γ_+ (2)

2. For the reverse inequality, let $H \subset Sym^+(\mathbb{C}^n) \times \mathbb{R}$ denote the set

$$H = \left\{ (\int_{S_1} zz^* \, d\mu(z), r + \int_{S_1} d\mu) \ \ , \ \ \mu \in \mathcal{B}(S_1) \ , \ r \geq 0 \right\}$$

Claim 1: H is closed.
Use Banach-Alaoglou theorem that the set of unit Borel measures is weak-* compact.
Claim 2: H is convex. – immediate
Let $q = max_{T=T^*} \ trace(TA)$ subject to $\langle Tx, x \rangle \leq 1$ for all $\|x\|_1 \leq 1$.
Claim 3: $(A, q) \in H$, which establishes the theorem.
Assume the contrary: $(A, q) \notin H$. Then it is separated by a hyperplane from H:

$$trace\left(R \int_{S_1} xx^* \, d\mu(z)\right) + a(r + \int_{S_1} d\mu) \geq c_0 > trace(AR) + aq \ , \ \forall \mu \in \mathcal{B}(S_1), r \geq 0$$

Deduce: $a \geq 0$, $c_0 \leq 0$. If $a = 0$ then contradiction for $\mu = \mu^*$. Rescale by dividing through a. Denote $T_0 = -R/a$.

Proof of strong duality for γ_+ (3)

We obtained:

$$\int_{S_1} (1 - \langle T_0 x, x\rangle) d\mu \geq c_0 > q - trace(AT_0)$$

for every Borel measure $\mu \in \mathcal{B}(S_1)$. This means $\langle T_0 x, x\rangle \leq 1$ for all $\|x\| = 1$. This also implies $\langle T_0 x, x\rangle \leq 1$ for all $\|x\|_1 \leq 1$. On the other hand $q < trace(AT_0) + c_0 \leq trace(AT_0)$ which contradicts the optimality of q. Q.E.D.