# General Probabilistic Theories, tensor products, and projective transformations

Stanislaw Szarek

Case Western Reserve U. & Sorbonne U.

G. Aubrun, L. Lami, C. Palazuelos, S. Szarek, A. Winter, Commun. Math. Phys. 375 (2020), 679-724
G. Aubrun, L. Lami, C. Palazuelos, M. Plavala, Geom. Funct. Anal. 31 (2021), 181-205

# Abstract

Generalized Probabilistic Theories (GPTs) are theories of nature that have random features. A GPT must specify the set of states purporting to represent the physical reality, the allowable measurements, the rules for outcome statistics of the latter, and the composition rules for merging subsystems and creating a larger system. Examples include classical probability and quantum theory.

The composition rules alluded to above usually involve tensor products, including tensor products of normed spaces, convex sets and of cones. Among tensor products that have operational meaning in the GPT context, the projective and the injective product are the extreme ones, which leads to the natural question "How much do they differ?" considered already by Grothendieck and Pisier, resp. in the 1950s and 1980s.

We report on quantitative results concerning projective/injective discrepancy for finite-dimensional normed spaces. Some of the results are essentially optimal, but others can be likely improved. The methods involve a wide range of techniques from geometry of Banach spaces and random matrices. We also report on parallel results in the context of cones. Finally, we will encourage a more systematic study of convex bodies with the allowed morphisms being projective transformations.

# Outline

• a few words about GPTs

• projective and injective tensor products and norms: definitions, notation

• historical background; the infinite dimensional case; qualitative vs. quantitative

• tensor products of cones vs. tensor products of normed spaces and convex bodies

• a selection of results and examples of tools from geometric functional analysis and random matrices

*Buzzwords* : CHSH inequality; Dvoretzky-Milman's theorem; $p$-summing norms; Chevet-Gordon's inequality; Grothendieck's inequalities; $K$-convexity & the $MM^*$-estimate

# Generalized Probabilistic Theories or GPTs

A general probabilistic theory is a triple $(V, C, u)$, where: (i) $V$ is a finite-dimensional real vector space; (ii) $C \subset V$ is a closed, convex, salient and generating cone; and (iii) $u$, called the order unit or the unit effect, is a functional in the interior of the dual cone

$$C^* := \{x^* \in V^* : x^*(x) \geqslant 0 \; \forall \, x \in C\}.$$

$(V, C)$ can be thought of as an ordered vector space: $x \leqslant y \Leftrightarrow y - x \in C$.

# Generalized Probabilistic Theories or GPTs

A general probabilistic theory is a triple $(V, C, u)$, where: (i) $V$ is a finite-dimensional real vector space; (ii) $C \subset V$ is a closed, convex, salient and generating cone; and (iii) $u$, called the order unit or the unit effect, is a functional in the interior of the dual cone

$$C^* := \{x^* \in V^* : x^*(x) \geqslant 0 \; \forall \, x \in C\}.$$

$(V, C)$ can be thought of as an ordered vector space: $x \leqslant y \Leftrightarrow y - x \in C$. The dual space $V^*$ can be thought of as ordered by the dual cone $C^*$.

# Generalized Probabilistic Theories or GPTs

A general probabilistic theory is a triple $(V, C, u)$, where: (i) $V$ is a finite-dimensional real vector space; (ii) $C \subset V$ is a closed, convex, salient and generating cone; and (iii) $u$, called the order unit or the unit effect, is a functional in the interior of the dual cone

$$C^* := \{x^* \in V^* : x^*(x) \geqslant 0 \ \forall x \in C\}.$$

$(V, C)$ can be thought of as an ordered vector space: $x \leqslant y \Leftrightarrow y - x \in C$. The dual space $V^*$ can be thought of as ordered by the dual cone $C^*$. The states of the physical system modeled by $(V, C, u)$ are represented by vectors in $C \cap u^{-1}(1) =: \Omega$.

# Generalized Probabilistic Theories or GPTs

A general probabilistic theory is a triple $(V, C, u)$, where: (i) $V$ is a finite-dimensional real vector space; (ii) $C \subset V$ is a closed, convex, salient and generating cone; and (iii) $u$, called the order unit or the unit effect, is a functional in the interior of the dual cone

$$C^* := \{x^* \in V^* : x^*(x) \geqslant 0 \ \forall x \in C\}.$$

$(V, C)$ can be thought of as an ordered vector space: $x \leqslant y \Leftrightarrow y - x \in C$.
The dual space $V^*$ can be thought of as ordered by the dual cone $C^*$.
The states of the physical system modeled by $(V, C, u)$ are represented by vectors in $C \cap u^{-1}(1) =: \Omega$.
The associated normed spaces have unit balls $B_V := \mathrm{conv}(-\Omega \cup \Omega)$ and $B_{V^*} := \{x^* : -u \leqslant x^* \leqslant u\}$.

# Generalized Probabilistic Theories or GPTs

A general probabilistic theory is a triple $(V, C, u)$, where: (i) $V$ is a finite-dimensional real vector space; (ii) $C \subset V$ is a closed, convex, salient and generating cone; and (iii) $u$, called the order unit or the unit effect, is a functional in the interior of the dual cone

$$C^* := \{x^* \in V^* : x^*(x) \geqslant 0 \; \forall \, x \in C\}.$$

$(V, C)$ can be thought of as an ordered vector space: $x \leqslant y \Leftrightarrow y - x \in C$.
The dual space $V^*$ can be thought of as ordered by the dual cone $C^*$.
The states of the physical system modeled by $(V, C, u)$ are represented by vectors in $C \cap u^{-1}(1) =: \Omega$.
The associated normed spaces have unit balls $B_V := \mathrm{conv}(-\Omega \cup \Omega)$ and $B_{V^*} := \{x^* : -u \leqslant x^* \leqslant u\}$.
If $(V_j, C_j, u_j)$, $j = 1, 2$ represent subsystems, the global system is modeled by $V = V_1 \otimes V_2$, with the choice of cone $C$ depending on further assumptions

# Generalized Probabilistic Theories or GPTs

A general probabilistic theory is a triple $(V, C, u)$, where: (i) $V$ is a finite-dimensional real vector space; (ii) $C \subset V$ is a closed, convex, salient and generating cone; and (iii) $u$, called the order unit or the unit effect, is a functional in the interior of the dual cone

$$C^* := \{x^* \in V^* : x^*(x) \geqslant 0 \;\forall\, x \in C\}.$$

$(V, C)$ can be thought of as an ordered vector space: $x \leqslant y \Leftrightarrow y - x \in C$.
The dual space $V^*$ can be thought of as ordered by the dual cone $C^*$.
The states of the physical system modeled by $(V, C, u)$ are represented by vectors in $C \cap u^{-1}(1) =: \Omega$.
The associated normed spaces have unit balls $B_V := \mathrm{conv}(-\Omega \cup \Omega)$ and $B_{V^*} := \{x^* : -u \leqslant x^* \leqslant u\}$.
If $(V_j, C_j, u_j)$, $j = 1, 2$ represent subsystems, the global system is modeled by $V = V_1 \otimes V_2$, with the choice of cone $C$ depending on further assumptions, e.g., global vs. local measurements being allowed.

# Classical and quantum GPTs

In the classical case, $\Omega \subset \mathbb{R}^n$ is the set of (discrete) probability densities.

$$\Omega = \{(p_j)_{j=1}^n : \sum_j p_j = 1\}$$

and the cone is $\mathbb{R}_+^n$.

# Classical and quantum GPTs

In the classical case, $\Omega \subset \mathbb{R}^n$ is the set of (discrete) probability densities.

$$\Omega = \{(p_j)_{j=1}^n : \sum_j p_j = 1\}$$

and the cone is $\mathbb{R}^n_+$.

In the quantum case, $\Omega$ is the set $n \times n$ matrices of trace 1 which further belong to the positive semi-definite cone (density matrices).

# Classical and quantum GPTs

In the classical case, $\Omega \subset \mathbb{R}^n$ is the set of (discrete) probability densities.

$$\Omega = \{(p_j)_{j=1}^n : \sum_j p_j = 1\}$$

and the cone is $\mathbb{R}_+^n$.

In the quantum case, $\Omega$ is the set $n \times n$ matrices of trace 1 which further belong to the positive semi-definite cone (density matrices).

The norms are the $\ell_1^n$-norm for the base space and the $\ell_\infty^n$-norm for the dual in the classical case, and the trace class norm and the "spectral" norm (as an operator on $\ell_2^n$) in the quantum case.

# Classical and quantum GPTs

In the classical case, $\Omega \subset \mathbb{R}^n$ is the set of (discrete) probability densities.

$$\Omega = \{(p_j)_{j=1}^n : \sum_j p_j = 1\}$$

and the cone is $\mathbb{R}_+^n$.

In the quantum case, $\Omega$ is the set $n \times n$ matrices of trace 1 which further belong to the positive semi-definite cone (density matrices).

The norms are the $\ell_1^n$-norm for the base space and the $\ell_\infty^n$-norm for the dual in the classical case, and the trace class norm and the "spectral" norm (as an operator on $\ell_2^n$) in the quantum case.

Roughly speaking, measurements of a state give probabilistic outcomes according to the weights $(p_j)$ and similarly in the quantum case.
This is measurement "in a basis", other schemes are also allowed.

# Definitions and notation : the projective norm

If $X, Y$ are real Banach spaces, we will consider norms on $X \otimes Y$ (the algebraic tensor product) verifying

$$\|x \otimes y\| = \|x\| \cdot \|y\|. \tag{1}$$

# Definitions and notation : the projective norm

If $X, Y$ are real Banach spaces, we will consider norms on $X \otimes Y$ (the algebraic tensor product) verifying

$$\|x \otimes y\| = \|x\| \cdot \|y\|. \tag{1}$$

By the triangle inequality, every such norm must satisfy

$$\|z\| \leqslant \min \left\{ \sum_i \|x_i\| \cdot \|y_i\| : \ z = \sum_i x_i \otimes y_i \right\} \tag{2}$$

# Definitions and notation : the projective norm

If $X, Y$ are real Banach spaces, we will consider norms on $X \otimes Y$ (the algebraic tensor product) verifying

$$\|x \otimes y\| = \|x\| \cdot \|y\|. \tag{1}$$

By the triangle inequality, every such norm must satisfy

$$\|z\| \leqslant \min \left\{ \sum_i \|x_i\| \cdot \|y_i\| : \ z = \sum_i x_i \otimes y_i \right\} \tag{2}$$

and replacing "$\leqslant$" by "$:=$" in (2) we get the definition of the projective tensor norm $\|z\|_{X \otimes_\pi Y}$, the largest norm on $X \otimes Y$ verifying (1)

# Definitions and notation : the projective norm

If $X, Y$ are real Banach spaces, we will consider norms on $X \otimes Y$ (the algebraic tensor product) verifying

$$\|x \otimes y\| = \|x\| \cdot \|y\|. \tag{1}$$

By the triangle inequality, every such norm must satisfy

$$\|z\| \leqslant \min \left\{ \sum_i \|x_i\| \cdot \|y_i\| : \; z = \sum_i x_i \otimes y_i \right\} \tag{2}$$

and replacing "$\leqslant$" by "$:=$" in (2) we get the definition of the projective tensor norm $\|z\|_{X \otimes_\pi Y}$, the largest norm on $X \otimes Y$ verifying (1), also denoted sometimes by $\|z\|_{X \widehat{\otimes} Y}$. We will usually write simply $\|z\|_\pi$.

# Definitions and notation : duality and the injective norm

For the smallest "reasonable" norm on $X \otimes Y$ it is most convenient to appeal to duality: if $x^* \in X^*, y^* \in Y^*$, we want $x^* \otimes y^*$ to induce a functional on $X \otimes Y$ whose norm is $\|x^*\| \cdot \|y^*\|$, which implies

$$\|z\| \geqslant \max\left\{(x^* \otimes y^*)(z) : \|x^*\| \leqslant 1,\ \|y^*\| \leqslant 1\right\}. \tag{3}$$

Again, replacing "$\geqslant$" by "$:=$" in (3) we get the definition of injective tensor norm $\|z\|_{X \otimes_\varepsilon Y}$ (or simply $\|z\|_\varepsilon$), denoted sometimes by $\|z\|_{X \check{\otimes} Y}$. Equivalently, $\|z\|_\varepsilon$ is the norm of $z$ as a bilinear form on $X^* \times Y^*$.

# Definitions and notation : duality and the injective norm

For the smallest "reasonable" norm on $X \otimes Y$ it is most convenient to appeal to duality: if $x^* \in X^*, y^* \in Y^*$, we want $x^* \otimes y^*$ to induce a functional on $X \otimes Y$ whose norm is $\|x^*\| \cdot \|y^*\|$, which implies

$$\|z\| \geqslant \max\left\{(x^* \otimes y^*)(z) : \|x^*\| \leqslant 1,\ \|y^*\| \leqslant 1\right\}. \tag{3}$$

Again, replacing "$\geqslant$" by "$:=$" in (3) we get the definition of injective tensor norm $\|z\|_{X \otimes_\varepsilon Y}$ (or simply $\|z\|_\varepsilon$), denoted sometimes by $\|z\|_{X \check{\otimes} Y}$. Equivalently, $\|z\|_\varepsilon$ is the norm of $z$ as a bilinear form on $X^* \times Y^*$.

A compact way to relate these two notions (at least in the finite dimensional case) is

$$X \otimes_\varepsilon Y = (X^* \otimes_\pi Y^*)^*.$$

# Definitions and notation : duality and the injective norm

For the smallest "reasonable" norm on $X \otimes Y$ it is most convenient to appeal to duality: if $x^* \in X^*, y^* \in Y^*$, we want $x^* \otimes y^*$ to induce a functional on $X \otimes Y$ whose norm is $\|x^*\| \cdot \|y^*\|$, which implies

$$\|z\| \geqslant \max\left\{(x^* \otimes y^*)(z) : \|x^*\| \leqslant 1,\ \|y^*\| \leqslant 1\right\}. \tag{3}$$

Again, replacing "$\geqslant$" by "$:=$" in (3) we get the definition of injective tensor norm $\|z\|_{X \otimes_\varepsilon Y}$ (or simply $\|z\|_\varepsilon$), denoted sometimes by $\|z\|_{X \check{\otimes} Y}$. Equivalently, $\|z\|_\varepsilon$ is the norm of $z$ as a bilinear form on $X^* \times Y^*$.

A compact way to relate these two notions (at least in the finite dimensional case) is

$$X \otimes_\varepsilon Y = (X^* \otimes_\pi Y^*)^*.$$

If the spaces are infinite dimensional, completions are required and there are reflexivity issues, but we will largely ignore this side of the story and – unless explicitly stated otherwise – will assume that $\dim X, \dim Y < \infty$.

# An equivalent language: tensor products of convex sets

In geometric functional analysis, we often identify norms on a finite dimensional vector space $V$ with symmetric convex bodies:

$$X = (V, \|\cdot\|) \quad \to \quad B_X := \{x : \|x\| \leqslant 1\} \quad = \quad \text{the unit ball of } X$$

$$V \supset K \to \|x\|_K := \inf\{t \geqslant 0 : x \in tK\} = \text{ the Minkowski functional of } K$$

# An equivalent language: tensor products of convex sets

In geometric functional analysis, we often identify norms on a finite dimensional vector space $V$ with symmetric convex bodies:

$$X = (V, \|\cdot\|) \quad \to \quad B_X := \{x : \|x\| \leqslant 1\} \quad = \quad \text{the unit ball of } X$$

$$V \supset K \to \|x\|_K := \inf\{t \geqslant 0 : x \in tK\} = \text{ the Minkowski functional of } K$$

In this setting we define the projective tensor product as

$$K \otimes_\pi L := \operatorname{conv}\{x \otimes y : x \in K, y \in L\}$$

# An equivalent language: tensor products of convex sets

In geometric functional analysis, we often identify norms on a finite dimensional vector space $V$ with symmetric convex bodies:

$$X = (V, \|\cdot\|) \quad \to \quad B_X := \{x : \|x\| \leqslant 1\} \quad = \quad \text{the unit ball of } X$$

$$V \supset K \to \|x\|_K := \inf\{t \geqslant 0 : x \in tK\} = \text{ the Minkowski functional of } K$$

In this setting we define the projective tensor product as

$$K \otimes_\pi L := \operatorname{conv}\{x \otimes y : x \in K, y \in L\}$$

and the previous definitions can be restated as

$$B_{X \otimes_\pi Y} := B_X \otimes_\pi B_Y \quad \text{and} \quad B_{X \otimes_\varepsilon Y} := \left(B_{X^*} \otimes_\pi B_{Y^*}\right)^\circ,$$

where $K^\circ := \{x \in V^* : \forall y \in K \ \langle y, x\rangle \leqslant 1\}$ is the polar of $K$.

# An equivalent language: tensor products of convex sets

In geometric functional analysis, we often identify norms on a finite dimensional vector space $V$ with symmetric convex bodies:

$$X = (V, \|\cdot\|) \quad \to \quad B_X := \{x : \|x\| \leqslant 1\} \quad = \quad \text{the unit ball of } X$$

$$V \supset K \to \|x\|_K := \inf\{t \geqslant 0 : x \in tK\} = \text{ the Minkowski functional of } K$$

In this setting we define the projective tensor product as

$$K \otimes_\pi L := \operatorname{conv}\{x \otimes y : x \in K, y \in L\}$$

and the previous definitions can be restated as

$$B_{X \otimes_\pi Y} := B_X \otimes_\pi B_Y \quad \text{and} \quad B_{X \otimes_\varepsilon Y} := (B_{X^*} \otimes_\pi B_{Y^*})^\circ ,$$

where $K^\circ := \{x \in V^* : \forall y \in K \ \langle y, x\rangle \leqslant 1\}$ is the polar of $K$.

One can likewise define projective and injective tensor products for not-necessarily-symmetric convex sets, most notably for cones.

# Considering operators rather than tensors

Since $X^* \otimes Y$ is canonically isomorphic to $\mathcal{L}(X, Y)$, it is also possible to avoid talking about tensors and rephrase all questions in terms of operators. In that setting, if $z = \sum_i |y_i\rangle\langle x_i^*|$, then

$$\|z\|_\varepsilon = \|z : X \to Y\|,$$

the operator norm, while $\|z\|_\pi = \min \sum_i \|y_i\| \cdot \|x_i^*\|$ (the minimum over all representations) is the nuclear norm.

# Considering operators rather than tensors

Since $X^* \otimes Y$ is canonically isomorphic to $\mathcal{L}(X, Y)$, it is also possible to avoid talking about tensors and rephrase all questions in terms of operators. In that setting, if $z = \sum_i |y_i\rangle\langle x_i^*|$, then

$$\|z\|_\varepsilon = \|z : X \to Y\|,$$

the operator norm, while $\|z\|_\pi = \min \sum_i \|y_i\| \cdot \|x_i^*\|$ (the minimum over all representations) is the nuclear norm. Moreover, appealing to duality gives

$$\|z\|_\pi = \max_{\|w : Y \to X\| \leqslant 1} \operatorname{tr} wz.$$

This allows to analyze both concepts in terms of operator norms, which are arguably conceptually simpler.

# Considering operators rather than tensors

Since $X^* \otimes Y$ is canonically isomorphic to $\mathcal{L}(X, Y)$, it is also possible to avoid talking about tensors and rephrase all questions in terms of operators. In that setting, if $z = \sum_i |y_i\rangle\langle x_i^*|$, then

$$\|z\|_\varepsilon = \|z : X \to Y\|,$$

the operator norm, while $\|z\|_\pi = \min \sum_i \|y_i\| \cdot \|x_i^*\|$ (the minimum over all representations) is the nuclear norm. Moreover, appealing to duality gives

$$\|z\|_\pi = \max_{\|w: Y \to X\| \leqslant 1} \operatorname{tr} wz.$$

This allows to analyze both concepts in terms of operator norms, which are arguably conceptually simpler. In particular

$$\rho(X, Y) := \max_{z \in X \otimes Y, z \neq 0} \frac{\|z\|_\pi}{\|z\|_\varepsilon} = \max_{\|w: Y \to X^*\| \leqslant 1, \|z: X^* \to Y\| \leqslant 1} \operatorname{tr} wz.$$

The quantity $\rho(X, Y)$ quantifies discrepancy between $X \otimes_\pi Y$ and $X \otimes_\varepsilon Y$ and is arguably the most important concept of this presentation.

# Grothendieck and Pisier

Tensor products of normed spaces were studied in detail by Grothendieck in 1950s. In particular, he proposed and studied 14 "natural tensor norms" and posed a number of open questions, one of which was whether the norms $\|\cdot\|_{X\otimes_\pi Y}$ and $\|\cdot\|_{X\otimes_\varepsilon Y}$ can be equivalent when when $\dim X = \dim Y = \infty$.

# Grothendieck and Pisier

Tensor products of normed spaces were studied in detail by Grothendieck in 1950s. In particular, he proposed and studied 14 "natural tensor norms" and posed a number of open questions, one of which was whether the norms $\|\cdot\|_{X\otimes_\pi Y}$ and $\|\cdot\|_{X\otimes_\varepsilon Y}$ can be equivalent when when $\dim X = \dim Y = \infty$.

It was a surprise when in 1980s Pisier answered this question in the positive, even more so because he showed earlier that if $\dim X \to \infty$ and $\dim Y \to \infty$, then

$$\rho(X, Y) \quad \to \quad \infty.$$

# Grothendieck and Pisier

Tensor products of normed spaces were studied in detail by Grothendieck in 1950s. In particular, he proposed and studied 14 "natural tensor norms" and posed a number of open questions, one of which was whether the norms $\|\cdot\|_{X\otimes_\pi Y}$ and $\|\cdot\|_{X\otimes_\varepsilon Y}$ can be equivalent when when $\dim X = \dim Y = \infty$.

It was a surprise when in 1980s Pisier answered this question in the positive, even more so because he showed earlier that if $\dim X \to \infty$ and $\dim Y \to \infty$, then

$$\rho(X, Y) \quad \to \quad \infty.$$

Also surprisingly, no quantitative analysis of the finite high-dimensional case was made until very recently. Such analysis is the main topic of this presentation. We also show that $\rho(X, Y) > 1$ in all nontrivial cases (i.e., $\min\{\dim W, \dim Y\} \geqslant 2$).

# The analogous problem for cones

*Given two non-degenerate cones $C \subset \mathbb{R}^n$, $C' \subset \mathbb{R}^m$, when do we have $C \otimes_\varepsilon C' = C \otimes_\pi C'$?*

# The analogous problem for cones

*Given two non-degenerate cones $C \subset \mathbb{R}^n$, $C' \subset \mathbb{R}^m$, when do we have $C \otimes_\varepsilon C' = C \otimes_\pi C'$?*

**Answer**: *If and only if one of the cones $C, C'$ is classical*

# The analogous problem for cones

*Given two non-degenerate cones* $C \subset \mathbb{R}^n$, $C' \subset \mathbb{R}^m$, *when do we have* $C \otimes_\varepsilon C' = C \otimes_\pi C'$*?*

**Answer**: *If and only if one of the cones* $C, C'$ *is classical (i.e., isomorphic to the standard classical cone).*

# The analogous problem for cones

*Given two non-degenerate cones $C \subset \mathbb{R}^n$, $C' \subset \mathbb{R}^m$, when do we have $C \otimes_\varepsilon C' = C \otimes_\pi C'$?*

**Answer**: *If and only if one of the cones $C, C'$ is classical (i.e., isomorphic to the standard classical cone).*

(G. Aubrun, L. Lami, C. Palazuelos, M. Plavala, 2021)

# The equality case for cones

Consider now any cone $C \subset \mathbb{R}^n$ and the classical cone $C' = \mathbb{R}^m_+ \subset \mathbb{R}^m$. We want to show that

$$C \otimes_\varepsilon \mathbb{R}^m_+ = C \otimes_\pi \mathbb{R}^m_+.$$

# The equality case for cones

Consider now any cone $C \subset \mathbb{R}^n$ and the classical cone $C' = \mathbb{R}_+^m \subset \mathbb{R}^m$. We want to show that

$$C \otimes_\varepsilon \mathbb{R}_+^m = C \otimes_\pi \mathbb{R}_+^m.$$

We have

$C \otimes_\pi \mathbb{R}_+^m = \text{conv}\{x \otimes y : x \in C, y \in \mathbb{R}_+^m\}$

# The equality case for cones

Consider now any cone $C \subset \mathbb{R}^n$ and the classical cone $C' = \mathbb{R}^m_+ \subset \mathbb{R}^m$.
We want to show that

$$C \otimes_\varepsilon \mathbb{R}^m_+ = C \otimes_\pi \mathbb{R}^m_+.$$

We have

$C \otimes_\pi \mathbb{R}^m_+ = \mathrm{conv}\{x \otimes y : x \in C, y \in \mathbb{R}^m_+\} = \mathrm{conv}\{\sum_{k=1}^m x_k \otimes e_k : x_k \in C\}$
because each $y \in \mathbb{R}^m_+$ is a positive linear combination of $e_k$'s

# The equality case for cones

Consider now any cone $C \subset \mathbb{R}^n$ and the classical cone $C' = \mathbb{R}^m_+ \subset \mathbb{R}^m$. We want to show that

$$C \otimes_\varepsilon \mathbb{R}^m_+ = C \otimes_\pi \mathbb{R}^m_+.$$

We have

$C \otimes_\pi \mathbb{R}^m_+ = \text{conv}\{x \otimes y : x \in C, y \in \mathbb{R}^m_+\} = \text{conv}\{\sum_{k=1}^m x_k \otimes e_k : x_k \in C\}$ because each $y \in \mathbb{R}^m_+$ is a positive linear combination of $e_k$'s, and likewise

$$C^* \otimes_\pi \mathbb{R}^m_+ = \text{conv}\{\sum_{k=1}^m y_k \otimes e_k : y_k \in C^*\}.$$

# The equality case for cones

Consider now any cone $C \subset \mathbb{R}^n$ and the classical cone $C' = \mathbb{R}^m_+ \subset \mathbb{R}^m$.
We want to show that

$$C \otimes_\varepsilon \mathbb{R}^m_+ = C \otimes_\pi \mathbb{R}^m_+.$$

We have

$C \otimes_\pi \mathbb{R}^m_+ = \text{conv}\{x \otimes y : x \in C, y \in \mathbb{R}^m_+\} = \text{conv}\{\sum_{k=1}^m x_k \otimes e_k : x_k \in C\}$
because each $y \in \mathbb{R}^m_+$ is a positive linear combination of $e_k$'s, and likewise

$$C^* \otimes_\pi \mathbb{R}^m_+ = \text{conv}\{\textstyle\sum_{k=1}^m y_k \otimes e_k : y_k \in C^*\}.$$

By definition, $\tau \in \mathbb{R}^n \otimes \mathbb{R}^m$ belongs to $C \otimes_\varepsilon \mathbb{R}^m_+$ iff

$$\langle \tau, \sigma \rangle \geqslant 0 \text{ for all } \sigma \in C^* \otimes_\pi \mathbb{R}^m_+.$$

# The equality case for cones

Consider now any cone $C \subset \mathbb{R}^n$ and the classical cone $C' = \mathbb{R}^m_+ \subset \mathbb{R}^m$.
We want to show that

$$C \otimes_\varepsilon \mathbb{R}^m_+ = C \otimes_\pi \mathbb{R}^m_+.$$

We have

$C \otimes_\pi \mathbb{R}^m_+ = \mathrm{conv}\{x \otimes y : x \in C, y \in \mathbb{R}^m_+\} = \mathrm{conv}\{\sum_{k=1}^m x_k \otimes e_k : x_k \in C\}$
because each $y \in \mathbb{R}^m_+$ is a positive linear combination of $e_k$'s, and likewise

$$C^* \otimes_\pi \mathbb{R}^m_+ = \mathrm{conv}\{\textstyle\sum_{k=1}^m y_k \otimes e_k : y_k \in C^*\}.$$

By definition, $\tau \in \mathbb{R}^n \otimes \mathbb{R}^m$ belongs to $C \otimes_\varepsilon \mathbb{R}^m_+$ iff

$$\langle \tau, \sigma \rangle \geqslant 0 \text{ for all } \sigma \in C^* \otimes_\pi \mathbb{R}^m_+.$$

Writing $\tau = \sum_{k=1}^m z_j \otimes e_j$ and testing it against $\sigma = y \otimes e_k$, $y \in C^*$ we get

$$\langle \tau, \sigma \rangle = \textstyle\sum_{k=1}^m \langle z_j, y \rangle \langle e_j, e_k \rangle = \langle z_k, y \rangle \geqslant 0$$

# The equality case for cones

Consider now any cone $C \subset \mathbb{R}^n$ and the classical cone $C' = \mathbb{R}^m_+ \subset \mathbb{R}^m$. We want to show that

$$C \otimes_\varepsilon \mathbb{R}^m_+ = C \otimes_\pi \mathbb{R}^m_+.$$

We have

$C \otimes_\pi \mathbb{R}^m_+ = \mathrm{conv}\{x \otimes y : x \in C, y \in \mathbb{R}^m_+\} = \mathrm{conv}\{\sum_{k=1}^m x_k \otimes e_k : x_k \in C\}$ because each $y \in \mathbb{R}^m_+$ is a positive linear combination of $e_k$'s, and likewise

$$C^* \otimes_\pi \mathbb{R}^m_+ = \mathrm{conv}\{\textstyle\sum_{k=1}^m y_k \otimes e_k : y_k \in C^*\}.$$

By definition, $\tau \in \mathbb{R}^n \otimes \mathbb{R}^m$ belongs to $C \otimes_\varepsilon \mathbb{R}^m_+$ iff

$$\langle \tau, \sigma \rangle \geqslant 0 \text{ for all } \sigma \in C^* \otimes_\pi \mathbb{R}^m_+.$$

Writing $\tau = \sum_{k=1}^m z_j \otimes e_j$ and testing it against $\sigma = y \otimes e_k$, $y \in C^*$ we get

$$\langle \tau, \sigma \rangle = \textstyle\sum_{k=1}^m \langle z_j, y \rangle \langle e_j, e_k \rangle = \langle z_k, y \rangle \geqslant 0 \text{ for all } y \in C^*,$$

which means that each $z_k \in C$ (by the bipolar theorem for cones).

# The equality case for cones

Consider now any cone $C \subset \mathbb{R}^n$ and the classical cone $C' = \mathbb{R}^m_+ \subset \mathbb{R}^m$. We want to show that

$$C \otimes_\varepsilon \mathbb{R}^m_+ = C \otimes_\pi \mathbb{R}^m_+.$$

We have

$C \otimes_\pi \mathbb{R}^m_+ = \mathrm{conv}\{x \otimes y : x \in C, y \in \mathbb{R}^m_+\} = \mathrm{conv}\{\sum_{k=1}^m x_k \otimes e_k : x_k \in C\}$ because each $y \in \mathbb{R}^m_+$ is a positive linear combination of $e_k$'s, and likewise

$$C^* \otimes_\pi \mathbb{R}^m_+ = \mathrm{conv}\{\textstyle\sum_{k=1}^m y_k \otimes e_k : y_k \in C^*\}.$$

By definition, $\tau \in \mathbb{R}^n \otimes \mathbb{R}^m$ belongs to $C \otimes_\varepsilon \mathbb{R}^m_+$ iff

$$\langle \tau, \sigma \rangle \geqslant 0 \text{ for all } \sigma \in C^* \otimes_\pi \mathbb{R}^m_+.$$

Writing $\tau = \sum_{k=1}^m z_j \otimes e_j$ and testing it against $\sigma = y \otimes e_k$, $y \in C^*$ we get

$$\langle \tau, \sigma \rangle = \textstyle\sum_{k=1}^m \langle z_j, y \rangle \langle e_j, e_k \rangle = \langle z_k, y \rangle \geqslant 0 \text{ for all } y \in C^*,$$

which means that each $z_k \in C$ (by the bipolar theorem for cones). That is, every $\tau \in C \otimes_\varepsilon \mathbb{R}^m_+$ belongs to $C \otimes_\pi \mathbb{R}^m_+$, so the two cones are equal.

The remarkable fact is that this is effectively the only case of equality.

# Back to normed spaces: some special cases

If $\mathcal{H}, \mathcal{K}$ are Hilbert (inner product) spaces, the situation is very simple: $\| \cdot \|_\varepsilon$ is the operator (spectral) norm, while $\| \cdot \|_\pi$ is the trace class norm and so

$$\rho(\mathcal{H}, \mathcal{K}) = \min\{\dim \mathcal{H}, \dim \mathcal{K}\}.$$

# Back to normed spaces: some special cases

If $\mathcal{H}, \mathcal{K}$ are Hilbert (inner product) spaces, the situation is very simple: $\| \cdot \|_\varepsilon$ is the operator (spectral) norm, while $\| \cdot \|_\pi$ is the trace class norm and so

$$\rho(\mathcal{H}, \mathcal{K}) = \min\{\dim \mathcal{H}, \dim \mathcal{K}\}.$$

(This in particular saturates the easy general upper bound for $\rho(X, Y)$.)

# Back to normed spaces: some special cases

If $\mathcal{H}, \mathcal{K}$ are Hilbert (inner product) spaces, the situation is very simple: $\| \cdot \|_\varepsilon$ is the operator (spectral) norm, while $\| \cdot \|_\pi$ is the trace class norm and so

$$\rho(\mathcal{H}, \mathcal{K}) = \min\{\dim \mathcal{H}, \dim \mathcal{K}\}.$$

(This in particular saturates the easy general upper bound for $\rho(X, Y)$.)
For a general lower bound, a naive attempt is to appeal to the John's theorem, which says that if $\dim X = n = \dim \mathcal{H}$, where $\mathcal{H}$ is a Hilbert space, then $d(X, \mathcal{H}) \leqslant n^{1/2}$, where

$$d(E, F) = \min\{\|v : E \to F\| \cdot \|v^{-1} : F \to E\|\}$$

is the Banach-Mazur distance.

# Back to normed spaces: some special cases

If $\mathcal{H}, \mathcal{K}$ are Hilbert (inner product) spaces, the situation is very simple: $\|\cdot\|_\varepsilon$ is the operator (spectral) norm, while $\|\cdot\|_\pi$ is the trace class norm and so

$$\rho(\mathcal{H}, \mathcal{K}) = \min\{\dim \mathcal{H}, \dim \mathcal{K}\}.$$

(This in particular saturates the easy general upper bound for $\rho(X, Y)$.) For a general lower bound, a naive attempt is to appeal to the John's theorem, which says that if $\dim X = n = \dim \mathcal{H}$, where $\mathcal{H}$ is a Hilbert space, then $d(X, \mathcal{H}) \leqslant n^{1/2}$, where

$$d(E, F) = \min\{\|v : E \to F\| \cdot \|v^{-1} : F \to E\|\}$$

is the Banach-Mazur distance. This allows to obtain some nontrivial information; for example using $v, v^{-1}$ certifying $d(X, \mathcal{H}) \leqslant n^{1/2}$ as $w, z$ in

$$\rho(X, \mathcal{H}) = \max_{\|w:\mathcal{H}\to X^*\|\leqslant 1,\ \|z:X^*\to\mathcal{H}\|\leqslant 1} \operatorname{tr} wz$$

we obtain $\rho(X, \mathcal{H}) \geqslant n^{1/2}$.

# Back to normed spaces: some special cases

If $\mathcal{H}, \mathcal{K}$ are Hilbert (inner product) spaces, the situation is very simple: $\|\cdot\|_\varepsilon$ is the operator (spectral) norm, while $\|\cdot\|_\pi$ is the trace class norm and so

$$\rho(\mathcal{H}, \mathcal{K}) = \min\{\dim \mathcal{H}, \dim \mathcal{K}\}.$$

(This in particular saturates the easy general upper bound for $\rho(X, Y)$.)
For a general lower bound, a naive attempt is to appeal to the John's theorem, which says that if $\dim X = n = \dim \mathcal{H}$, where $\mathcal{H}$ is a Hilbert space, then $d(X, \mathcal{H}) \leqslant n^{1/2}$, where

$$d(E, F) = \min\{\|v : E \to F\| \cdot \|v^{-1} : F \to E\|\}$$

is the Banach-Mazur distance. This allows to obtain some nontrivial information; for example using $v, v^{-1}$ certifying $d(X, \mathcal{H}) \leqslant n^{1/2}$ as $w, z$ in

$$\rho(X, \mathcal{H}) = \max_{\|w:\mathcal{H}\to X^*\|\leqslant 1,\ \|z:X^*\to\mathcal{H}\|\leqslant 1} \operatorname{tr} wz$$

we obtain $\rho(X, \mathcal{H}) \geqslant n^{1/2}$. The same circle of ideas allows to handle the case of different dimensions: $\rho(X, \mathcal{H}) \geqslant \min\{\dim X, \dim \mathcal{H}\}^{1/2}$.

# Some special cases, cont'd

The same argument proves a cute equality $\rho(X, X^*) = \dim X$, but it doesn't help in the general case: by a 1981 result of Gluskin $\max\{d(E, F) : \dim E = \dim F = n\} = \Theta(n)$ and no nontrivial lower bound can be directly inferred.

# Some special cases, cont'd

The same argument proves a cute equality $\rho(X, X^*) = \dim X$, but it doesn't help in the general case: by a 1981 result of Gluskin $\max\{d(E, F) : \dim E = \dim F = n\} = \Theta(n)$ and no nontrivial lower bound can be directly inferred.

Here are other interesting special cases that can be handled. If (say) $\dim X \geqslant n$, then

$$\rho(X, \ell_1^n) \geqslant (n/2)^{1/2} \quad \text{and} \quad \rho(X, \ell_\infty^n) \geqslant (n/2)^{1/2}.$$

# Some special cases, cont'd

The same argument proves a cute equality $\rho(X, X^*) = \dim X$, but it doesn't help in the general case: by a 1981 result of Gluskin $\max\{d(E, F) : \dim E = \dim F = n\} = \Theta(n)$ and no nontrivial lower bound can be directly inferred.

Here are other interesting special cases that can be handled. If (say) $\dim X \geqslant n$, then

$$\rho(X, \ell_1^n) \geqslant (n/2)^{1/2} \quad \text{and} \quad \rho(X, \ell_\infty^n) \geqslant (n/2)^{1/2}.$$

The first inequality follows by relating $\rho(X, \ell_1^n)$ to the so-called $p$-summing norms of the identity on $X$; these concepts were fashionable in 1970s and 1980s.

# Some special cases, cont'd

The same argument proves a cute equality $\rho(X, X^*) = \dim X$, but it doesn't help in the general case: by a 1981 result of Gluskin $\max\{d(E, F) : \dim E = \dim F = n\} = \Theta(n)$ and no nontrivial lower bound can be directly inferred.

Here are other interesting special cases that can be handled. If (say) $\dim X \geqslant n$, then

$$\rho(X, \ell_1^n) \geqslant (n/2)^{1/2} \text{ and } \rho(X, \ell_\infty^n) \geqslant (n/2)^{1/2}.$$

The first inequality follows by relating $\rho(X, \ell_1^n)$ to the so-called $p$-summing norms of the identity on $X$; these concepts were fashionable in 1970s and 1980s. The second one is then a consequence of (generally true) $\rho(X, Y) = \rho(X^*, Y^*)$.

# Some special cases, cont'd

The same argument proves a cute equality $\rho(X, X^*) = \dim X$, but it doesn't help in the general case: by a 1981 result of Gluskin $\max\{d(E, F) : \dim E = \dim F = n\} = \Theta(n)$ and no nontrivial lower bound can be directly inferred.

Here are other interesting special cases that can be handled. If (say) $\dim X \geqslant n$, then

$$\rho(X, \ell_1^n) \geqslant (n/2)^{1/2} \text{ and } \rho(X, \ell_\infty^n) \geqslant (n/2)^{1/2}.$$

The first inequality follows by relating $\rho(X, \ell_1^n)$ to the so-called $p$-summing norms of the identity on $X$; these concepts were fashionable in 1970s and 1980s. The second one is then a consequence of (generally true) $\rho(X, Y) = \rho(X^*, Y^*)$. No substantial improvement is possible since $\rho(\mathcal{H}, \ell_1^n) = n^{1/2}$ (easy), but we do not know whether $(n/2)^{1/2}$ can be replaced by $n^{1/2}$ in general.

# The quantum case

The case that is of relevance to quantum theory is when $X, Y$ are spaces of Hermitian matrices endowed with the trace class norm. We have then

$$\rho(X, Y) = \Theta(\min\{\dim X, \dim Y\}^{3/4}).$$

# The quantum case

The case that is of relevance to quantum theory is when $X, Y$ are spaces of Hermitian matrices endowed with the trace class norm. We have then

$$\rho(X, Y) = \Theta(\min\{\dim X, \dim Y\}^{3/4}).$$

For simplicity, consider $X = Y$ to be spaces of $d \times d$ matrices, so $n = \dim X = \dim Y = d^2$.

# The quantum case

The case that is of relevance to quantum theory is when $X, Y$ are spaces of Hermitian matrices endowed with the trace class norm. We have then

$$\rho(X, Y) = \Theta(\min\{\dim X, \dim Y\}^{3/4}).$$

For simplicity, consider $X = Y$ to be spaces of $d \times d$ matrices, so $n = \dim X = \dim Y = d^2$. The lower bound uses random tensors. Let

$$z = \sum_{j=1}^{d^2} \sum_{k=1}^{d^2} g_{jk} x_j \otimes y_k$$

where $(x_j)$ and $(y_k)$ are Hilbert–Schmidt orthonormal bases of $X$ and $Y$ and $g_{jk}$ are i.i.d. Gaussian random variables.

# The quantum case

The case that is of relevance to quantum theory is when $X, Y$ are spaces of Hermitian matrices endowed with the trace class norm. We have then

$$\rho(X, Y) = \Theta(\min\{\dim X, \dim Y\}^{3/4}).$$

For simplicity, consider $X = Y$ to be spaces of $d \times d$ matrices, so $n = \dim X = \dim Y = d^2$. The lower bound uses random tensors. Let

$$z = \sum_{j=1}^{d^2} \sum_{k=1}^{d^2} g_{jk} x_j \otimes y_k$$

where $(x_j)$ and $(y_k)$ are Hilbert–Schmidt orthonormal bases of $X$ and $Y$ and $g_{jk}$ are i.i.d. Gaussian random variables. We are going to use

$$\rho(X, Y) = \max\{\operatorname{tr} wz : \|w : Y \to X^*\| \leqslant 1, \|z : X^* \to Y\| \leqslant 1\}$$

with $w = z^T$.

# The quantum case

The case that is of relevance to quantum theory is when $X, Y$ are spaces of Hermitian matrices endowed with the trace class norm. We have then

$$\rho(X, Y) = \Theta(\min\{\dim X, \dim Y\}^{3/4}).$$

For simplicity, consider $X = Y$ to be spaces of $d \times d$ matrices, so $n = \dim X = \dim Y = d^2$. The lower bound uses random tensors. Let

$$z = \sum_{j=1}^{d^2} \sum_{k=1}^{d^2} g_{jk} x_j \otimes y_k$$

where $(x_j)$ and $(y_k)$ are Hilbert–Schmidt orthonormal bases of $X$ and $Y$ and $g_{jk}$ are i.i.d. Gaussian random variables. We are going to use

$$\rho(X, Y) = \max\{\operatorname{tr} wz : \|w : Y \to X^*\| \leqslant 1, \ \|z : X^* \to Y\| \leqslant 1\}$$

with $w = z^T$. It is clear what $\operatorname{tr} z^T z$ is

# The quantum case

The case that is of relevance to quantum theory is when $X, Y$ are spaces of Hermitian matrices endowed with the trace class norm. We have then

$$\rho(X, Y) = \Theta(\min\{\dim X, \dim Y\}^{3/4}).$$

For simplicity, consider $X = Y$ to be spaces of $d \times d$ matrices, so $n = \dim X = \dim Y = d^2$. The lower bound uses random tensors. Let

$$z = \sum_{j=1}^{d^2} \sum_{k=1}^{d^2} g_{jk} x_j \otimes y_k$$

where $(x_j)$ and $(y_k)$ are Hilbert–Schmidt orthonormal bases of $X$ and $Y$ and $g_{jk}$ are i.i.d. Gaussian random variables. We are going to use

$$\rho(X, Y) = \max\{\operatorname{tr} wz : \|w : Y \to X^*\| \leqslant 1, \|z : X^* \to Y\| \leqslant 1\}$$

with $w = z^T$. It is clear what $\operatorname{tr} z^T z$ is, and the norms $\|z^T : Y \to X^*\|$, $\|z : X^* \to Y\|$ are controlled (ultimately) via the Chevet-Gordon inequality.

# The quantum case, continued

And here is the idea behind the $O(\cdot)$ argument (the upper bound).

# The quantum case, continued

And here is the idea behind the $O(\cdot)$ argument (the upper bound).

We note first that $z \in X \otimes_\varepsilon Y$ can be thought of as a bilinear form on $X^* \times Y^*$ and that $X^* = Y^*$ is (the self-adjoint part of) the $C^*$-algebra $\mathcal{A}$ of $k \times k$ matrices with the usual operator norm. Thus we are in the realm of the Haagerup-Pisier non-commutative Grothendieck inequality, which says that for such bilinear form there are states $\varphi, \psi$ on $\mathcal{A}$ such that

$$|z(a,b)| \leqslant 2\|z\|_\varepsilon \, \varphi(a^2)^{1/2} \, \psi(b^2)^{1/2} \quad \text{for all} \quad a, b \in \mathrm{Re}\,\mathcal{A}.$$

With this information, we need to upper-bound $\|z\|_\pi$, or the nuclear norm of $z : \mathcal{A} \to \mathcal{A}^*$.

# The quantum case, conclusion

Again, we need to upper-bound the nuclear norm of $z : \mathcal{A} \to \mathcal{A}^*$, using

$$|z(a, b)| \leqslant 2\|z\|_\varepsilon \, \varphi(a^2)^{1/2} \, \psi(b^2)^{1/2} \quad \text{for all} \quad a, b \in \mathrm{Re}\, \mathcal{A}.$$

Here is a calculation which is not quite right, but supplies the gist of the trick.

# The quantum case, conclusion

Again, we need to upper-bound the nuclear norm of $z : \mathcal{A} \to \mathcal{A}^*$, using

$$|z(a,b)| \leqslant 2\|z\|_\varepsilon\, \varphi(a^2)^{1/2}\, \psi(b^2)^{1/2} \quad \text{for all} \quad a, b \in \operatorname{Re}\mathcal{A}.$$

Here is a calculation which is not quite right, but supplies the gist of the trick. Let $\varphi = \sum_{i=1}^d \lambda_i |u_i\rangle\langle u_i|$ be the spectral decomposition. We will estimate the nuclear norm of $z : \mathcal{A} \to \mathcal{A}^*$ (say, with $\|z\|_\varepsilon \leqslant 1$) by writing

$$z(a) = \sum_{i,j} \operatorname{tr}(aE_{ji}) z(E_{ij}), \quad \text{or} \quad z = \sum_{i,j} |z(E_{ij})\rangle\langle E_{ji}|$$

where $E_{ij} = |u_i\rangle\langle u_j|$. For a single term, we have

$$\|z(E_{ij})\|_{\mathcal{A}^*} = \max_{\|b\|_{\mathcal{A}} \leqslant 1} |z(E_{ij}, b)| \leqslant 2\varphi(|E_{ij}|^2)^{1/2} \leqslant 2\lambda_i^{1/2}$$

(note that $\psi(b^2) \leqslant 1$ if $\|b\|_{\mathcal{A}} \leqslant 1$)

# The quantum case, conclusion

Again, we need to upper-bound the nuclear norm of $z : \mathcal{A} \to \mathcal{A}^*$, using

$$|z(a,b)| \leqslant 2\|z\|_\varepsilon \, \varphi(a^2)^{1/2} \, \psi(b^2)^{1/2} \quad \text{for all} \quad a, b \in \operatorname{Re}\mathcal{A}.$$

Here is a calculation which is not quite right, but supplies the gist of the trick. Let $\varphi = \sum_{i=1}^d \lambda_i |u_i\rangle\langle u_i|$ be the spectral decomposition. We will estimate the nuclear norm of $z : \mathcal{A} \to \mathcal{A}^*$ (say, with $\|z\|_\varepsilon \leqslant 1$) by writing

$$z(a) = \sum_{i,j} \operatorname{tr}(aE_{ji}) z(E_{ij}), \quad \text{or} \quad z = \sum_{i,j} |z(E_{ij})\rangle\langle E_{ji}|$$

where $E_{ij} = |u_i\rangle\langle u_j|$. For a single term, we have

$$\|z(E_{ij})\|_{\mathcal{A}^*} = \max_{\|b\|_{\mathcal{A}} \leqslant 1} |z(E_{ij}, b)| \leqslant 2\varphi(|E_{ij}|^2)^{1/2} \leqslant 2\lambda_i^{1/2}$$

(note that $\psi(b^2) \leqslant 1$ if $\|b\|_{\mathcal{A}} \leqslant 1$) and summing over $1 \leqslant i, j \leqslant d$ gives $2d \sum_i \lambda_i^{1/2} \leqslant 2d^{3/2} = 2n^{3/4}$ as a bound on $\|z\|_\pi$ ($4n^{3/4}$ if we don't cheat).

# The general case, or cases

Modulo logarithmic factors (indicated by $^*$ in the $\Omega$ notation), we have:

- $X = Y, \dim X = n$: $\rho(X, X) = \Omega^*(n^{1/2})$ (almost optimal, see $X = \ell_1^n$)
- $\dim X = \dim Y = n$: $\rho(X, Y) = \Omega^*(n^{1/6})$
- $\dim X = n \leqslant \dim Y$: $\rho(X, Y) = \Omega^*(n^{1/8})$

# The general case, or cases

Modulo logarithmic factors (indicated by $^*$ in the $\Omega$ notation), we have:

- $X = Y, \dim X = n$: $\rho(X, X) = \Omega^*(n^{1/2})$ (almost optimal, see $X = \ell_1^n$)
- $\dim X = \dim Y = n$: $\rho(X, Y) = \Omega^*(n^{1/6})$
- $\dim X = n \leqslant \dim Y$: $\rho(X, Y) = \Omega^*(n^{1/8})$

The bounds can not be better than, respectively, $(2n)^{1/2}$, $n^{1/2}$, and again $n^{1/2}$. We know that $(2n)^{1/2}$ is not sharp, but we do not know whether the factor 2 can be removed. It is conceivable that all these quantities are actually $\Omega^*(n^{1/2})$ or even $\Omega(n^{1/2})$.

# The general case, or cases

Modulo logarithmic factors (indicated by $^*$ in the $\Omega$ notation), we have:

- $X = Y, \dim X = n$: $\rho(X, X) = \Omega^*(n^{1/2})$ (almost optimal, see $X = \ell_1^n$)
- $\dim X = \dim Y = n$: $\rho(X, Y) = \Omega^*(n^{1/6})$
- $\dim X = n \leqslant \dim Y$: $\rho(X, Y) = \Omega^*(n^{1/8})$

The bounds can not be better than, respectively, $(2n)^{1/2}$, $n^{1/2}$, and again $n^{1/2}$. We know that $(2n)^{1/2}$ is not sharp, but we do not know whether the factor 2 can be removed. It is conceivable that all these quantities are actually $\Omega^*(n^{1/2})$ or even $\Omega(n^{1/2})$.

While the above results are asymptotic, it is worthwhile noting that

- $\dim X, \dim Y \geqslant 2 \implies \rho(X, Y) \geqslant \frac{19}{18}$.

# An Auerbach-type lemma

Just for fun, here is a simple lemma that is fundamental in proving this 2-dimensional lower bound and which is susceptible to an intuitive geometrical interpretation:

*Let $X$ be a Banach space of dimension at least 2. Then there exist vectors $e_1, e_2 \in X$, $e_1^*, e_2^* \in X^*$ such that for any $i, j \in \{1, 2\}$ we have $\|e_i\|_X = \|e_j^*\|_{X^*} = 1$, $e_j^*(e_i) = \delta_{ij}$, and moreover*

$$\|e_1 + e_2\|_X \leqslant 3/2.$$

Once established, this allows to mimic the entanglement-detecting CHSH-inequality to obtain a non-trivial 19/18 lower bound for $\rho(X, Y)$. This is the bound shown in our paper; we have a more complicated argument that gives $\frac{8}{7}$, and we suspect that the right number is $\sqrt{2}$.

# Affine vs. projective transformations

One of the reasons for the difference between the setting of cones and the setting of normed spaces is that two different sets of morphisms are involved.

# Affine vs. projective transformations

One of the reasons for the difference between the setting of cones and the setting of normed spaces is that two different sets of morphisms are involved.

A cone $C$ is uniquely determined by its base

$$B_C = C \cap u^{-1}(1)$$

corresponding to an interior element $u$ of the dual cone $C^*$.

# Affine vs. projective transformations

One of the reasons for the difference between the setting of cones and the setting of normed spaces is that two different sets of morphisms are involved.

A cone $C$ is uniquely determined by its base

$$B_C = C \cap u^{-1}(1)$$

corresponding to an interior element $u$ of the dual cone $C^*$.

While the bases of $C$ and $C^*$ are related by the standard polarity relation for convex bodies (not entirely trivial, but true)

# Affine vs. projective transformations

One of the reasons for the difference between the setting of cones and the setting of normed spaces is that two different sets of morphisms are involved.

A cone $C$ is uniquely determined by its base

$$B_C = C \cap u^{-1}(1)$$

corresponding to an interior element $u$ of the dual cone $C^*$.

While the bases of $C$ and $C^*$ are related by the standard polarity relation for convex bodies (not entirely trivial, but true), the set of morphisms in the category of cones (linear maps of the underlying real vector spaces) corresponds to *projective* maps between the bases and not to *affine* maps.

# Affine vs. projective transformations

One of the reasons for the difference between the setting of cones and the setting of normed spaces is that two different sets of morphisms are involved.

A cone $C$ is uniquely determined by its base

$$B_C = C \cap u^{-1}(1)$$

corresponding to an interior element $u$ of the dual cone $C^*$.

While the bases of $C$ and $C^*$ are related by the standard polarity relation for convex bodies (not entirely trivial, but true), the set of morphisms in the category of cones (linear maps of the underlying real vector spaces) corresponds to *projective* maps between the bases and not to *affine* maps.

One illustration of this difference is the fact that all convex quadrangles are projectively equivalent, while obviously they are not all affinely equivalent. This suggests investigating the class of $n$-dimensional (not necessarily-symmetric) convex bodies by identifying bodies that are projectively equivalent, and with some projective version of the Banach-Mazur distance quantifying their similarity.

# The toolbox for the general case

This is again based on $\rho(X, Y) = \max_{\|w:Y\to X^*\|\leqslant 1,\, \|z:X^*\to Y\|\leqslant 1} \operatorname{tr} wz$ and an appropriate relaxation of the choices $w = v, z = v^{-1}$. First, we define the factorization constant of $X$ through $Y$ as

$$\mathrm{f}(X, Y) := \inf_{u,v} \{\|u : X \to Y\| \cdot \|v : Y \to X\| \,:\, vu = \mathrm{Id}_X\},$$

which allows $\dim X \neq \dim Y$ and means that a subspace "well-isomorphic" to $X$ is "well-complemented" in $Y$.

# The toolbox for the general case

This is again based on $\rho(X,Y) = \max_{\|w:Y\to X^*\|\leqslant 1,\, \|z:X^*\to Y\|\leqslant 1} \operatorname{tr} wz$ and an appropriate relaxation of the choices $w = v, z = v^{-1}$. First, we define the factorization constant of $X$ through $Y$ as

$$\mathrm{f}(X,Y) := \inf_{u,v} \left\{ \|u : X \to Y\| \cdot \|v : Y \to X\| \,:\, vu = \mathrm{Id}_X \right\},$$

which allows $\dim X \neq \dim Y$ and means that a subspace "well-isomorphic" to $X$ is "well-complemented" in $Y$. Next, the weak factorization constant is

$$\mathrm{wf}(X,Y) := \inf_{u,v} \left\{ \mathbf{E}\left[ \|u : X \to Y\| \cdot \|v : Y \to X\| \right] \,:\, \mathbf{E}\left[vu\right] = \mathrm{Id}_X \right\},$$

where $u, v$ are now operator-valued random variables.

# The toolbox for the general case

This is again based on $\rho(X, Y) = \max_{\|w:Y\to X^*\|\leqslant 1, \|z:X^*\to Y\|\leqslant 1} \operatorname{tr} wz$ and an appropriate relaxation of the choices $w = v, z = v^{-1}$. First, we define the factorization constant of $X$ through $Y$ as

$$\mathsf{f}(X, Y) := \inf_{u,v} \{\|u : X \to Y\| \cdot \|v : Y \to X\| \, : \, vu = \mathrm{Id}_X\},$$

which allows $\dim X \neq \dim Y$ and means that a subspace "well-isomorphic" to $X$ is "well-complemented" in $Y$. Next, the weak factorization constant is

$$\mathsf{wf}(X, Y) := \inf_{u,v} \{\mathbf{E}\,[\|u : X \to Y\| \cdot \|v : Y \to X\|] \, : \, \mathbf{E}\,[vu] = \mathrm{Id}_X\},$$

where $u, v$ are now operator-valued random variables.

Clearly $\mathsf{wf}(X, Y) \leqslant \mathsf{f}(X, Y) \leqslant d(X, Y)$ and one easily checks that

$$\rho(X', Y) \leqslant \mathsf{wf}(X', X)\rho(X, Y) \quad \text{and} \quad \rho(X, Y) \geqslant \frac{\dim X}{\mathsf{wf}(X, Y^*)}.$$

# The toolbox for the general case, cont'd

If $X = Y$, we select $u, v$ "at random."

# The toolbox for the general case, cont'd

If $X = Y$, we select $u, v$ "at random." At first, we choose a representation of $X$ of $\mathbb{R}^n$ and let $u = n^{-1/2} G^T$, $v = n^{-1/2} G$, where $G$ is a GOE matrix.

# The toolbox for the general case, cont'd

If $X = Y$, we select $u, v$ "at random." At first, we choose a representation of $X$ of $\mathbb{R}^n$ and let $u = n^{-1/2}G^T$, $v = n^{-1/2}G$, where $G$ is a GOE matrix. The tool which allows to estimate $\mathbf{E}\|G : Y^* \to X\|$ is the Chevet-Gordon inequality, which upper-bounds it by $n^{-1/2}$ times

$$\|\mathrm{Id} : \ell_2^n \to X\| \cdot \mathbf{E}\|g\|_Y + \|\mathrm{Id} : \ell_2^n \to Y\| \cdot \mathbf{E}\|g\|_X,$$

where $g$ is the standard Gaussian vector on $\mathbb{R}^n$.

# The toolbox for the general case, cont'd

If $X = Y$, we select $u, v$ "at random." At first, we choose a representation of $X$ of $\mathbb{R}^n$ and let $u = n^{-1/2} G^T$, $v = n^{-1/2} G$, where $G$ is a GOE matrix. The tool which allows to estimate $\mathbf{E}\|G : Y^* \to X\|$ is the Chevet-Gordon inequality, which upper-bounds it by $n^{-1/2}$ times

$$\|\mathrm{Id} : \ell_2^n \to X\| \cdot \mathbf{E}\|g\|_Y + \|\mathrm{Id} : \ell_2^n \to Y\| \cdot \mathbf{E}\|g\|_X,$$

where $g$ is the standard Gaussian vector on $\mathbb{R}^n$. If $X = Y$, the two terms coincide and – bounding similarly $\mathbf{E}\|G^T : X \to Y^*\|$ and noting that $\|\mathrm{Id} : \ell_2^n \to X^*\| = \|\mathrm{Id} : X \to \ell_2^n\|$ – we see that we need to control

$$\|\mathrm{Id} : \ell_2^n \to X\| \cdot \|\mathrm{Id} : X \to \ell_2^n\| \cdot \mathbf{E}\|n^{-1/2} g\|_X \cdot \mathbf{E}\|n^{-1/2} g\|_{X^*}.$$

# The toolbox for the general case, cont'd

If $X = Y$, we select $u, v$ "at random." At first, we choose a representation of $X$ of $\mathbb{R}^n$ and let $u = n^{-1/2}G^T$, $v = n^{-1/2}G$, where $G$ is a GOE matrix. The tool which allows to estimate $\mathbf{E}\|G : Y^* \to X\|$ is the Chevet-Gordon inequality, which upper-bounds it by $n^{-1/2}$ times

$$\|\mathrm{Id} : \ell_2^n \to X\| \cdot \mathbf{E}\|g\|_Y + \|\mathrm{Id} : \ell_2^n \to Y\| \cdot \mathbf{E}\|g\|_X,$$

where $g$ is the standard Gaussian vector on $\mathbb{R}^n$. If $X = Y$, the two terms coincide and – bounding similarly $\mathbf{E}\|G^T : X \to Y^*\|$ and noting that $\|\mathrm{Id} : \ell_2^n \to X^*\| = \|\mathrm{Id} : X \to \ell_2^n\|$ – we see that we need to control

$$\|\mathrm{Id} : \ell_2^n \to X\| \cdot \|\mathrm{Id} : X \to \ell_2^n\| \cdot \mathbf{E}\|n^{-1/2}g\|_X \cdot \mathbf{E}\|n^{-1/2}g\|_{X^*}.$$

For an appropriate representation of $X$ of $\mathbb{R}^n$, the first two factors give $d(X, \ell_2^n) \leqslant n^{1/2}$.

# The toolbox for the general case, cont'd

If $X = Y$, we select $u, v$ "at random." At first, we choose a representation of $X$ of $\mathbb{R}^n$ and let $u = n^{-1/2}G^T$, $v = n^{-1/2}G$, where $G$ is a GOE matrix. The tool which allows to estimate $\mathbf{E}\|G : Y^* \to X\|$ is the Chevet-Gordon inequality, which upper-bounds it by $n^{-1/2}$ times

$$\|\mathrm{Id} : \ell_2^n \to X\| \cdot \mathbf{E}\|g\|_Y + \|\mathrm{Id} : \ell_2^n \to Y\| \cdot \mathbf{E}\|g\|_X,$$

where $g$ is the standard Gaussian vector on $\mathbb{R}^n$. If $X = Y$, the two terms coincide and – bounding similarly $\mathbf{E}\|G^T : X \to Y^*\|$ and noting that $\|\mathrm{Id} : \ell_2^n \to X^*\| = \|\mathrm{Id} : X \to \ell_2^n\|$ – we see that we need to control

$$\|\mathrm{Id} : \ell_2^n \to X\| \cdot \|\mathrm{Id} : X \to \ell_2^n\| \cdot \mathbf{E}\|n^{-1/2}g\|_X \cdot \mathbf{E}\|n^{-1/2}g\|_{X^*}.$$

For an appropriate representation of $X$ of $\mathbb{R}^n$, the first two factors give $d(X, \ell_2^n) \leqslant n^{1/2}$. The last two factors are essentially the same as spherical means, or mean widths, which can be controlled by the $MM^*$-estimate.

# Mean (half-)width of $K \subset \mathbb{R}^n$ and the $MM^*$-estimate

If $|u| = 1$ and $w(K, u) := \sup_{x \in K} \langle u, x \rangle = \|u\|_{K^\circ}$, then $w(K, u) + w(K, -u)$ is the width of $K$ in the direction of $u$. The average over $u$ is the mean half-width of $K$.

# Mean (half-)width of $K \subset \mathbb{R}^n$ and the $MM^*$-estimate

If $|u| = 1$ and $w(K, u) := \sup_{x \in K} \langle u, x \rangle = \|u\|_{K^\circ}$, then $w(K, u) + w(K, -u)$ is the width of $K$ in the direction of $u$. The average over $u$ is the mean half-width of $K$.

The $MM^*$-estimate says that, for some well-balanced linear image $\tilde{K}$ of any centrally symmetric convex body $K \subset \mathbb{R}^n$ we can achieve $w(\tilde{K}) \cdot w(\tilde{K}^\circ) = O(\log n)$.

# Mean (half-)width of $K \subset \mathbb{R}^n$ and the $MM^*$-estimate

If $|u| = 1$ and $w(K, u) := \sup_{x \in K} \langle u, x \rangle = \|u\|_{K^\circ}$, then $w(K, u) + w(K, -u)$ is the width of $K$ in the direction of $u$. The average over $u$ is the mean half-width of $K$.

The $MM^*$-estimate says that, for some well-balanced linear image $\tilde{K}$ of any centrally symmetric convex body $K \subset \mathbb{R}^n$ we can achieve $w(\tilde{K}) \cdot w(\tilde{K}^\circ) = O(\log n)$. Thus, for the corresponding representation of $X$ on $\mathbb{R}^n$, we obtain the same bound for $\mathbf{E}\|n^{-1/2}g\|_X \cdot \mathbf{E}\|n^{-1/2}g\|_{X^*}$.

# Mean (half-)width of $K \subset \mathbb{R}^n$ and the $MM^*$-estimate

If $|u| = 1$ and $w(K,u) := \sup_{x \in K} \langle u, x \rangle = \|u\|_{K^\circ}$, then $w(K,u) + w(K,-u)$ is the width of $K$ in the direction of $u$. The average over $u$ is the mean half-width of $K$.

The $MM^*$-estimate says that, for some well-balanced linear image $\tilde{K}$ of any centrally symmetric convex body $K \subset \mathbb{R}^n$ we can achieve $w(\tilde{K}) \cdot w(\tilde{K}^\circ) = O(\log n)$. Thus, for the corresponding representation of $X$ on $\mathbb{R}^n$, we obtain the same bound for $\mathbf{E}\|n^{-1/2} g\|_X \cdot \mathbf{E}\|n^{-1/2} g\|_{X^*}$.

Some additional tweaking is needed since we need to reconcile two requirements for the representation of $X$ of $\mathbb{R}^n$, the one witnessing $d(X, \ell_2^n)$ and the other consistent with the $MM^*$-estimate, but ultimately gathering all bounds we get

$$\rho(X,X) = \Omega\left(\frac{\dim X}{d(X,\ell_2^n)\log^3 n}\right) \geqslant \Omega\left(\frac{n^{1/2}}{\log^3 n}\right),$$

as needed.

# The case $X \neq Y$

There are too many “balancing requirements” to be simultaneously achievable, so instead the argument is based on the following trichotomy.

# The case $X \neq Y$

There are too many "balancing requirements" to be simultaneously achievable, so instead the argument is based on the following trichotomy.

*Let $X$ be a normed space of dimension $n$. Then for every $1 \leqslant A \leqslant n^{1/2}$ at least one of the following holds*

1. *$X$ contains a subspace $E$ of dimension $d = \Omega(n^{1/2})$ such that $d(E, \ell_\infty^d) = O(A\sqrt{\log n})$.*
2. *$X^*$ contains a subspace $F$ of dimension $d = \Omega(n^{1/2})$ such that $d(F, \ell_\infty^d) = O(A\sqrt{\log n})$.*
3. *$X$ contains a subspace $H$ of dimension $d = \Omega(A^2/\log n)$ such that $d(H, \ell_2^d) \leqslant 4$ and, additionally, $H$ is $O(\log n)$-complemented in $X$.*

# The case $X \neq Y$

There are too many "balancing requirements" to be simultaneously achievable, so instead the argument is based on the following trichotomy.

*Let $X$ be a normed space of dimension $n$. Then for every $1 \leqslant A \leqslant n^{1/2}$ at least one of the following holds*

1. *$X$ contains a subspace $E$ of dimension $d = \Omega(n^{1/2})$ such that $d(E, \ell_\infty^d) = O(A\sqrt{\log n})$.*
2. *$X^*$ contains a subspace $F$ of dimension $d = \Omega(n^{1/2})$ such that $d(F, \ell_\infty^d) = O(A\sqrt{\log n})$.*
3. *$X$ contains a subspace $H$ of dimension $d = \Omega(A^2/\log n)$ such that $d(H, \ell_2^d) \leqslant 4$ and, additionally, $H$ is $O(\log n)$-complemented in $X$.*

Since subspaces $\lambda$-isomorphic to $\ell_\infty^d$ are automatically $\lambda$-complemented, each of the conditions above leads to an upper bound on $\mathrm{wf}(\ell_p^d, X)$ for the appropriate $p \in \{1, 2, \infty\}$. Given that $\rho(\ell_p^d, \ell_{p'}^{d'})$ are known, every combination of these conditions for $X$ and $Y$ leads to a lower bound on $\rho(X, Y)$, and the final step is optimizing over $A$.

THANK YOU!