Math. Program., Ser. A
DOI 10.1007/s10107-015-0899-9

FULL LENGTH PAPER

Two pairs of families of polyhedral norms versus ℓ_p-norms: proximity and applications in optimization

Jun-ya Gotoh[1] · **Stan Uryasev**[2]

Received: 27 October 2013 / Accepted: 13 March 2015
© Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society 2015

Abstract This paper studies four families of polyhedral norms parametrized by a single parameter. The first two families consist of the CVaR norm (which is equivalent to the D-norm, or the largest-k norm) and its dual norm, while the second two families consist of the convex combination of the ℓ_1- and ℓ_∞-norms, referred to as the deltoidal norm, and its dual norm. These families contain the ℓ_1- and ℓ_∞-norms as special limiting cases. These norms can be represented using linear programming (LP) and the size of LP formulations is independent of the norm parameters. The purpose of this paper is to establish a relation of the considered LP-representable norms to the ℓ_p-norm and to demonstrate their potential in optimization. On the basis of the ratio of the tight lower and upper bounds of the ratio of two norms, we show that in each dual pair, the primal and dual norms can equivalently well approximate the ℓ_p- and ℓ_q-norms, respectively, for $p, q \in [1, \infty]$ satisfying $1/p + 1/q = 1$. In addition, the deltoidal norm and its dual norm are shown to have better proximity to the ℓ_p-norm than the CVaR norm and its dual. Numerical examples demonstrate that LP solutions with optimized parameters attain better approximation of the ℓ_2-norm than the ℓ_1- and ℓ_∞-norms do.

Keywords ℓ_p-norm · CVaR norm · Deltoidal norm · Linear programming (LP) · pth order cone programming (pOCP)

✉ Stan Uryasev
uryasev@ufl.edu

Jun-ya Gotoh
jgoto@indsys.chuo-u.ac.jp

1 Department of Industrial and Systems Engineering, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

2 Risk Management and Financial Engineering Lab, Department of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, Gainesville, FL 32611, USA

Published online: 08 April 2015

Mathematics Subject Classification 47A30 · 51M16 · 51M20 · 52B12 · 52B15 · 90C05 · 90C20 · 90C90

1 Introduction

Norms are used in many fields of mathematics for gauging the size of vectors. The ℓ_2-norm (or the Euclidean norm), $\|\cdot\|_2$, has been treated preferentially among many candidates of norms, despite the seemingly counterintuitive square-operation in its definition.[1] It is frequently used also in optimization because its minimization can be reduced to a quadratic optimization, including the second order cone program (SOCP), which can be efficiently solved via an interior point method.

As an alternative to the ℓ_2-norm, the ℓ_1-norm (or the Manhattan or taxicab norm), $\|\cdot\|_1$, has gathered attention in computer science, statistics, operations research and so on.[2] Since it is easily interpretable and minimization of the ℓ_1-norm results in a linear program (LP), substituting the ℓ_1-norm for the ℓ_2-norm has been a fundamental technique and employed in many applications.[3]

We think that the ℓ_∞-norm (or the supremum or uniform norm), $\|\cdot\|_\infty$, deserves the third place, regarding its popularity. Like the ℓ_1-norm, its minimization results in an LP, but its advantage can be found in the dependence on a single extremal element of the vector. Such a worst-case-based feature is advantageous in reflecting a 'conservativeness' in the measurement.[4]

The three norms mentioned above belong to a parametrized family of the ℓ_p*-norm*. For a vector $\boldsymbol{x} \in \mathbb{R}^n$, it is defined by

$$\|\boldsymbol{x}\|_p := \begin{cases} (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}, & p \in [1, \infty), \\ \lim_{r\to\infty} \|\boldsymbol{x}\|_r \equiv \max\{|x_1|, |x_2|, \ldots, |x_n|\}, & p = \infty. \end{cases}$$

Although some applications have employed the ℓ_p-norm with $p \neq 1, 2$, or ∞ (e.g., [16,18,26]), unfortunately, it is hard to find advantages in its computational tractability and interpretability over the aforementioned three cases, $p = 1, 2$, and ∞. Indeed, we can easily make an intractable numerical example. For example, to evaluate the ℓ_p-norm of a vector $\boldsymbol{x} = (100, -101, 102)$ with $p = 500$, we have to compute

[1] In many practical situation, the distance between two points is measured by the length of the line segment connecting them. In statistics, random variables are usually evaluated via the ℓ_2-norm (e.g., variance, χ^2 value, and the sum of squared errors).

[2] For example, in machine learning (e.g., the lasso of [24] and the elastic net of [27]) and compressed sensing (e.g., [7]), it is employed because its minimization often brings a sparse solution, i.e., a solution having many zero-elements, which is in contrast to the ℓ_2-norm minimization usually resulting in a dense solution.

[3] For example, with a slight modification, minimizing the absolute deviation from the expected value in place of the variance is considered as an application of the approximation of the ℓ_2-norm by the ℓ_1-norm and often employed in various contexts (e.g., [1] for the statistical regression; [15] for agricultural planning; [17] for financial optimization).

[4] Also, its counterparts are used in various context (e.g., in gauging the distance of two cumulative distribution functions for the Kolmogorov-Smirnov test).

$(100^{500}+101^{500}+102^{500})^{\frac{1}{500}}$. The exponentiation may result in overflow of a standard computer or require some sophisticated numerical method, which contrasts to the cases with $p = 1, 2$, and ∞.[5]

While the p-th order cone program (pOCP) or the ℓ_p-norm minimization can be efficiently solved with sophisticated solver software packages, the advantages of the LP-representability of the ℓ_1- and ℓ_∞-norms are still significant. Actually, LP is a subclass of pOCP and LP solvers are more accessible and stable than pOCP solvers. In addition, in recent papers such as [16], the parameter p of the ℓ_p-norm is supposed to be calibrated to given data, which would require an optimization with respect to the norm parameter p. Such a parametric optimization of pOCP is, however, still developing. In contrast, with a parametrized family of polyhedral norms, we might be able to employ the well-established parametric LP methods.[6]

Given these circumstances, this paper explores relations between the family of ℓ_p-norms and parametrized families of polyhedral norms which do not require significant change of the complexity of computation compared to the ℓ_1- or ℓ_∞-norm and is more interpretable than ℓ_p-norms with $p \neq 1$ or ∞. Specifically, we consider two (primal and dual) pairs of families of LP-representable norms containing both the ℓ_1- and ℓ_∞-norms as special cases.

Let us denote by $x_{(i)}$ the element whose absolute value is the i-th largest among the n elements of a vector $\boldsymbol{x} \in \mathbb{R}^n$, i.e., $|x_{(1)}| \geq |x_{(2)}| \geq \cdots \geq |x_{(n)}|$. For any integer $k \in \{1, \ldots, n\}$, the function of the form

$$|x_{(1)}| + |x_{(2)}| + \cdots + |x_{(k)}|$$

is also a norm in $\mathbb{R}^n$ (see, e.g., [6,8] for its convexity). This notion is known as the *largest-k norm* and has been further generalized for $k \in [1, n]$ in [4] or $k \in (0, n]$ in [20]. Following [20], we call it the *CVaR norm*.

Since the size of related LPs is independent of the parameter k, as will be presented, the CVaR norm and its dual generalize the ℓ_1- and ℓ_∞-norms without significant increase of computational complexity. Now we can formulate the research question as follows.

Q.1.(a). For what value of $k \in [1, n]$ does the CVaR norm (or its dual) give the best approximation of the ℓ_p-norm, and in which sense is it the best?

Another candidate of LP-representable norm is the convex combination of the ℓ_1- and ℓ_∞-norms, i.e., $(1 - \lambda)\|\cdot\|_1 + \lambda\|\cdot\|_\infty$ for some $\lambda \in [0, 1]$. Due to a reason explained later, we call this the *deltoidal norm*. For this norm, the question Q.1.(a) is modified as below.

Q.1.(b). For what value of $\lambda \in [0, 1]$ does the deltoidal norm (or its dual) give the best approximation of the ℓ_p-norm, and in which sense is it the best?

[5] MATLAB and R on a 64-bit PC output '`Inf`' (i.e., ∞), which is clearly a wrong answer. On the other hand, algebraic computation softwares such as SageMathCloud output a reasonable value.

[6] Another promising example is in a recent paper of the authors. Indeed, [12] reports that by applying the cross-validation method for calibrating the norm parameter, a support vector machine (SVM) with the CVaR norm regularizer attained better out-of-sample performance than the standard ℓ_2 regularizer SVM within a comparable amount of computation time.

Also, we address the following question.

Q.2. Which norm better approximates the series of ℓ_p-norms at the best parameters, the CVaR norm or the deltoidal norm (or one of their duals)?

Here we would like to emphasize that our primal concern is not in the approximation of pOCP itself.[7] Rather, we would like to bring the attention to the usability of the families of LP-representable norms.[8]

The contribution of the paper lies in several aspects. First, to answer the questions above, we derive the tight bounds of the ratio of each parametrized family of the polyhedral norms and the family of ℓ_p-norms, and define a measure of proximity between them. The derivation enjoys techniques common in the mathematical optimization literature. Second, as a byproduct, a new characterization of polytopes is derived. Also, the paper provides a review of the representative norms in $\mathbb{R}^n$.

The structure of the paper is as follows. In Sect. 2, following a brief summary of known properties of the CVaR norm, we examine the proximity between the CVaR norm and the ℓ_p-norm and derive the parameter of the CVaR norm which approximates the ℓ_p-norm in the best way. Sect. 3 analyzes the deltoidal norms, similar to the preceding section. Section 4 presents applications of the proximity results to optimization problems, such as a norm minimization, a cone-constrained optimization, and a refinement of the probabilistic bound of the robust optimization studied in [4]. Some numerical examples are given in Sect. 5, demonstrating how well the worst-case-based theoretical bounds work in numerical applications, including the projection of a point on a polyhedron, the mean-variance portfolio selection, and a robust LP problem.

2 Approximation by CVaR Norm and Its Dual

In order to make this paper self-contained, we begin with a brief review of the CVaR norm [20] and relate it to the D-norm [4].

[7] It is known that pOCP can be efficiently approximated by a single LP via the polyhedral approximations by [3,25]. Also, we cannot assert that the replacement of pOCP with an LP-representable norm approximation always improves the computational efficiency. Indeed, there are cases which need cautious treatment. For example, let us consider the least square regression via variable selection. It is known to be formulated as a 0-1 mixed integer quadratic programming (MIQP) problem, and that some researches propose an alternative of a 0-1 mixed integer LP (MILP) by replacing the absolute error to the squared error. However, as long as we compared these two formulations for the UCI repository data sets by using IBM ILOG CPLEX 12.5, we could not find any advantage of the MILP approach. In fact, the MILP approach spent much more time than the MIQP one. This may be because the additional increase of the number of constraints excessively relaxed the subproblems generated in the branch-and-bound scheme.

[8] Indeed, several recent software packages, such as Portfolio Safeguard [21] and CVX [13], include the ℓ_1- and ℓ_∞-norms as built-in functions. These examples show a potential usability of the parametrized families of norms. Those packages can easily incorporate the user-defined functions on the basis of simple operations such as '+' and '`max`.' The two norms are easily readily available like the ℓ_2-norm since the dual CVaR and the deltoidal norm can be represented using the ℓ_1 and ℓ_∞ norms and those operations, as will be described later.

2.1 CVaR Norm and Its Dual

Definition 1 (*CVaR norm* [20]) For $\alpha = \frac{j}{n}$, $j = 0, 1, \ldots, n-1$, the CVaR norm, denoted by $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha$, of a vector $\boldsymbol{x} \in \mathbb{R}^n$ with confidence level α is defined by the formula:

$$\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha := \sum_{i=1}^{n-j} |x_{(i)}|.$$

For $\alpha \in (\frac{j}{n}, \frac{j+1}{n})$, $j = 0, \ldots, n-2$, it is defined by the weighted average of $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\frac{j}{n}}$ and $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\frac{j+1}{n}}$:

$$\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha := (j + 1 - n\alpha)\langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\frac{j}{n}} + (n\alpha - j)\langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\frac{j+1}{n}}.$$

For $\alpha \in (\frac{n-1}{n}, 1)$, it is $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha := n(1-\alpha)\langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\frac{n-1}{n}}$.

For example, the CVaR norm of a vector $\boldsymbol{x} = (2, -5, \sqrt{7})$ with $\alpha = \frac{1}{3}$ is calculated by sorting the vector as $(5, \sqrt{7}, 2)$ and summing elements corresponding to α: $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\frac{1}{3}} = 5 + \sqrt{7}$. When $\alpha = 0$, the CVaR norm is equal to the ℓ_1-norm, whereas when $\alpha = \frac{n-1}{n}$, it is equal to the ℓ_∞-norm. Obviously, the CVaR norm with $\alpha = \frac{j}{n}$, $j = 0, 1, \ldots, n-1$, corresponds to the largest-$(n-j)$ norm (e.g., [8]). Roughly speaking, the CVaR norm with confidence level α generalizes the sum of elements whose absolute values are among the top $100(1-\alpha)$ percent of the n absolute values.

[20] shows that for $\alpha \in [0, 1)$, the CVaR norm of $\boldsymbol{x} \in \mathbb{R}^n$ is given by

$$\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha \equiv \min_c \left\{ n(1-\alpha)c + \sum_{i=1}^{n} [|x_i| - c]^+ \right\}, \tag{1}$$

where $[x]^+ := \max\{x, 0\}$. This representation is originally given in [22] for a risk measure known as the *conditional value-at-risk (CVaR)* in the context of financial risk minimization. More specifically, (1) corresponds to the CVaR in which the absolute value of a component of the vector $\boldsymbol{x}$ is defined as a loss (see [23] for the details). This is the reason why we call $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ the CVaR norm.

An advantage of the CVaR norm is its LP-representability. Indeed, the formula (1) is also represented by an LP, whose size is independent of the parameter α:

$$\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha \equiv \left| \begin{array}{l} \underset{c,z}{\text{minimize}} \ n(1-\alpha)c + \sum_{i=1}^{n} z_i \\ \text{subject to } z_i \geq x_i - c, \ z_i \geq -x_i - c, \ z_i \geq 0, \quad i = 1, \ldots, n. \end{array} \right. \tag{2}$$

Note that by using the (strong) duality theory of LP, we have another LP formulation of the CVaR norm of $\boldsymbol{x} \in \mathbb{R}^n$:

$$\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha \equiv \max \left\{ \sum_{i=1}^{n} |x_i| q_i : \sum_{i=1}^{n} q_i = n(1-\alpha), 0 \le q_i \le 1, \quad i = 1, \ldots, n \right\},$$

which is an example of the so-called *continuous (or fractional) knapsack problem*.[9] An optimal solution to the above LP can be calculated by a greedy algorithm,[10] and the optimal objective value is given by

$$\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha \equiv \sum_{i=1}^{\lfloor n(1-\alpha) \rfloor} \left| x_{(i)} \right| + \{ n(1-\alpha) - \lfloor n(1-\alpha) \rfloor \} \left| x_{(\lfloor n(1-\alpha) \rfloor + 1)} \right|, \tag{3}$$

where $\lfloor x \rfloor$ denotes the largest integer no greater than x. The formula (3) provides another perspective of Definition 1.

A mathematically equivalent notion in combinatorial terms is earlier given by [4] and it is called the *D-norm* therein.

Definition 2 (*D-norm* [4]) For $\kappa \in [1, n]$, the D-norm of $\boldsymbol{x} \in \mathbb{R}^n$ is defined by

$$|\!|\!| \boldsymbol{x} |\!|\!|_\kappa := \max_{S,t} \left\{ \sum_{j \in S} |x_j| + (\kappa - \lfloor \kappa \rfloor) |x_t| : S \subset N, |S| \le \lfloor \kappa \rfloor, t \in N \backslash S \right\}, \tag{4}$$

where $N := \{1, \ldots, n\}$, and note that the maximization in the formula (4) is carried out with respect to $S \subset N$ and $t \in N$.

Unfortunately, however, paper [2] has recently started using the name, D-norm, for a different norm in a distant context. In order to avoid an unnecessary mix-up and to emphasize the relation with the optimization-based formulation (1), we call this norm the CVaR norm for the remainder of this paper.

For $\alpha \in [0, \frac{n-1}{n}]$, the CVaR norm is equivalent to the D-norm with parameter κ defined by

$$\kappa = n(1-\alpha), \tag{5}$$

(see [20][11]). In the rest part of this paper, we limit $\alpha \in [0, \frac{n-1}{n}]$ so that the ℓ_1- and ℓ_∞-norms would be the extreme cases.

The following relation provides a key to the analysis below about the proximity between the ℓ_p-norm and the CVaR norm.

[9] In this sense, $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ may be called the *knapsack norm*.

[10] The complexity of calculating the CVaR norm itself is defined by the complexity of a selection algorithm (and summation), i.e., $O(n)$.

[11] In [20], 'p' is used in place of κ. In order to avoid a mix-up with the parameter 'p' for the ℓ_p-norm, we employ κ instead. Also, this replacement may be convenient to remind the readers of the fact that for $\alpha = \frac{j}{n}$, $j = 0, 1, \ldots, n-1$, the CVaR norm is equivalent to the largest-'k' norm. Indeed, the usage of κ implies that it can take any non-integer values, while k can often be linked only to integers.

Proposition 1 *For any $p \in (1, \infty)$, $\alpha \in [0, \frac{n-1}{n}]$, and $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, it is valid*

$$\min\left\{1, n^{1-\frac{1}{p}}(1-\alpha)\right\} \leq \frac{\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha}{\|\boldsymbol{x}\|_p} \leq \left\{\lfloor \kappa \rfloor + (\kappa - \lfloor \kappa \rfloor)^{\frac{p}{p-1}}\right\}^{\frac{p-1}{p}}, \tag{6}$$

where $\kappa = n(1-\alpha)$. For $p = 1$, the bounds (6) are specified as $1-\alpha \leq \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha / \|\boldsymbol{x}\|_1 \leq 1$, whereas for $p = \infty$, they are $1 \leq \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha / \|\boldsymbol{x}\|_\infty \leq n(1-\alpha)$. All the above bounds are tight.

Proposition 1 is an extension of [4], which establishes the tight bounds of the ratio of the CVaR norm (D-norm) to the ℓ_2-norm. To our best knowledge, Proposition 1 is new for $p \neq 2$. See "Proof of Proposition 1" in Appendix for the proof.

Let us emphasize that each of the two inequalities in (6) is tight, i.e., there is some $\boldsymbol{x}$ satisfying equality. Indeed, $\boldsymbol{x} = (1, 0, \ldots, 0)$ or $\boldsymbol{x} = (1, 1, \ldots, 1)$ attains the lower bound, while $\boldsymbol{x}$ with $x_1 = \cdots = x_{\lfloor \kappa \rfloor} = 1$, $x_{\lfloor \kappa \rfloor + 1} = (\kappa - \lfloor \kappa \rfloor)^{\frac{1}{p-1}}$ and $x_{\lfloor \kappa \rfloor + 2} = \cdots = x_n = 0$ attains the upper bound.

Before exploring how to attain the best approximation, we introduce the dual norm. For a norm, denoted by $\|\cdot\|$,[12] in $\mathbb{R}^n$, its dual norm of $\boldsymbol{x} \in \mathbb{R}^n$ is defined by[13]

$$\|\boldsymbol{x}\|^* := \max\{\boldsymbol{x}^\top \boldsymbol{y} : \|\boldsymbol{y}\| \leq 1\}.$$

Paper [4] shows that for $\boldsymbol{x} \in \mathbb{R}^n$, the dual norm of the CVaR norm is simply given by

$$\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* = \max\left\{\frac{1}{n(1-\alpha)}\|\boldsymbol{x}\|_1, \|\boldsymbol{x}\|_\infty\right\}.$$

We below refer to $\langle\!\langle \cdot \rangle\!\rangle_\alpha^*$ as the *dual CVaR norm*.[14]

Similarly to the CVaR norm, the dual CVaR norm attains the following tight bounds of the associated ratio.

Proposition 2 *For any $p \in (1, \infty)$, $\alpha \in [0, \frac{n-1}{n}]$, and $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, we have*

$$\min\left\{1, n^{\frac{1}{p}}(1-\alpha)\right\} \leq \frac{\|\boldsymbol{x}\|_p}{\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^*} \leq \{\lfloor \kappa \rfloor + (\kappa - \lfloor \kappa \rfloor)^p\}^{\frac{1}{p}}, \tag{7}$$

with $\kappa = n(1-\alpha)$. For $p = 1$, the bounds (7) are specified as $1 \leq \|\boldsymbol{x}\|_1 / \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* \leq n(1-\alpha)$, whereas for $p = \infty$, they are $1-\alpha \leq \|\boldsymbol{x}\|_\infty / \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* \leq 1$. All the above bounds are tight.

See "Proof of proposition 2" in Appendix for the proof.

Consequently, we now see that both the CVaR norm and its dual have a symmetric pair of lower and upper bounds on the ratios to the ℓ_p- and $\ell_{\frac{p}{p-1}}$-norms, respectively:

[12] Throughout the paper the notation $\|\cdot\|$ (or $\|\cdot\|'$) is reserved for any norm, while $\|\cdot\|_p$, $\langle\!\langle \cdot \rangle\!\rangle_\alpha$, $(\!(\cdot)\!)_\lambda$ and $|\!|\!|\cdot|\!|\!|_\kappa$ are for the named norms.

[13] We write vectors in $\mathbb{R}^n$ as column vectors in the inner products.

[14] Note that the dual CVaR norm formula implies the duality between the ℓ_1-norm and the ℓ_∞-norm, i.e., $\|\cdot\|_1^* = \|\cdot\|_\infty$ and $\|\cdot\|_\infty^* = \|\cdot\|_1$.

$$\min\left\{1, n^{1-\frac{1}{p}}(1-\alpha)\right\} \le \left\{\begin{array}{l} \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha / \|\boldsymbol{x}\|_p \\ \|\boldsymbol{x}\|_{\frac{p}{p-1}} / \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* \end{array}\right\} \le \left\{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^{\frac{p}{p-1}}\right\}^{1-\frac{1}{p}}, \quad (8)$$

where $\kappa = n(1-\alpha)$. Especially when $p = 2$, the tight bounds (8) are given in a more symmetric manner by

$$\min\{1, \sqrt{n}(1-\alpha)\} \le \left\{\begin{array}{l} \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha / \|\boldsymbol{x}\|_2 \\ \|\boldsymbol{x}\|_2 / \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* \end{array}\right\} \le \sqrt{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^2}. \quad (9)$$

Remark 1 The upper and lower bounds in (9) are derived in [4] for the CVaR norm. The upper and lower bounds associated with the dual CVaR norm derived in [4] are as follows:

$$\min\{1, \sqrt{n}(1-\alpha)\} \le \frac{\|\boldsymbol{x}\|_2}{\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^*} \le \max\{n(1-\alpha), \sqrt{n}\}.$$

However, the upper bound is not tight.[15] Moreover, the proof for the lower bound in [4] does not guarantee tightness. Proposition 2 thus upgrades Proposition 3 of [4]. As will be shown in Sect. 4.3, Theorem 3 (b) of [4] can be strengthened based on the modification above.

2.2 The Best Approximation of ℓ_p-Norm by CVaR Norm and Its Dual

Based on the two pairs of bounds (8), we can find the best value of α so that $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ and $\langle\!\langle \cdot \rangle\!\rangle_\alpha^*$ approximate $\|\cdot\|_p$ and $\|\cdot\|_{\frac{p}{p-1}}$, respectively, in the best way. Due to the symmetry of the pairs of the bounds (8), we can analyze the best values of the CVaR norm and the dual CVaR norm in one shot.

To find parameters providing the best approximation, let us define the function showing how well a norm $\|\cdot\|$ approximates another norm $\|\cdot\|'$. Our suggestion is that the goodness of the approximation of $\|\cdot\|'$ by $\|\cdot\|$ be defined on the basis of the tight bounds of the ratio between the two norms. Specifically, supposing that a pair of tight bounds of the form:

$$\forall \boldsymbol{x} \in \mathbb{R}\backslash\{\boldsymbol{0}\}, \quad L \le \frac{\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|'} \le U, \quad \left(\text{or equivalently, } \frac{1}{U} \le \frac{\|\boldsymbol{x}\|'}{\|\boldsymbol{x}\|} \le \frac{1}{L}\right) \quad (10)$$

is given with some positive constants L and U, then, the ratio U/L can be a measure of the proximity, i.e., the goodness of approximation.[16] For example, since for any

[15] Indeed, it can be shown that if we have $L \le \|\boldsymbol{x}\|/\|\boldsymbol{x}\|_2 \le U$ for a norm $\|\cdot\|$, then we also have $L \le \|\boldsymbol{x}\|/\|\boldsymbol{x}\|_2 \le \|\boldsymbol{x}\|_2/\|\boldsymbol{x}\|^* \le U$.

[16] Note that this ratio is independent of the locations of the norms in the ratio. Indeed, $L \le \|\boldsymbol{x}\|/\|\boldsymbol{x}\|' \le U$ is equivalently presented by $1/U \le \|\boldsymbol{x}\|'/\|\boldsymbol{x}\| \le 1/L$, but the ratio remains to be equal to U/L. Note also that $U/L \ge 1$ for any pair of norms $\|\cdot\|, \|\cdot\|'$, and $U/L = 1$ implies $\|\cdot\| = \|\cdot\|'$. Further, the value $\log(U/L)$ becomes a metric (or distance) of norms of $\mathbb{R}^n$. From the approximation viewpoint, the closer U/L is to 1 (or the closer $\log(U/L)$ is to 0), the better.

$\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, we have $1 \leq \|\boldsymbol{x}\|_1 / \|\boldsymbol{x}\|_2 \leq \sqrt{n}$, and the ratio U/L between the ℓ_1- and ℓ_2-norms is $\sqrt{n}$. As for the ratio between the ℓ_∞- and ℓ_2-norms, it is again equal to $\sqrt{n}$ since we have the tight bounds: $1/\sqrt{n} \leq \|\boldsymbol{x}\|_\infty / \|\boldsymbol{x}\|_2 \leq 1$. With this measure, we can say that the proximities of the ℓ_1- and ℓ_∞-norms to the ℓ_2-norm are the same.[17]

Based on the bounds (8), let us define the function which evaluates the proximity for the pair of the families of the CVaR norms, i.e., the ratio U/L of the bounds on $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha / \|\boldsymbol{x}\|_p$ and $\|\boldsymbol{x}\|_{\frac{p}{p-1}} / \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^*$. Given a positive integer n and $p \in (1, \infty)$, let us denote it by

$$f_{n,p}(\kappa) := \frac{\left\{ \lfloor \kappa \rfloor + (\kappa - \lfloor \kappa \rfloor)^{\frac{p}{p-1}} \right\}^{\frac{p-1}{p}}}{\min\{1, n^{-\frac{1}{p}} \kappa\}}, \tag{11}$$

where $\kappa = n(1-\alpha)$.[18]

Lemma 1 *$f_{n,p}(\kappa)$ is continuous at any $\kappa \in (1, n)$, and differentiable at any non-integer except at $\kappa = n^{\frac{1}{p}}$, i.e., $\kappa \notin \{1, \ldots, n\} \cup \{n^{\frac{1}{p}}\}$.*

See "Proof of Lemma 1" in Appendix for the proof of the continuity of $f_{n,p}(\kappa)$.

Proposition 3 *The function $f_{n,p}(\kappa)$ is decreasing for $\kappa \leq n^{\frac{1}{p}}$. The function $f_{n,p}(\kappa)$ is increasing for $\kappa \geq n^{\frac{1}{p}}$. Accordingly, $f_{n,p}(\kappa)$ uniquely attains its minimum value, $\{\lfloor n^{\frac{1}{p}} \rfloor + (n^{\frac{1}{p}} - \lfloor n^{\frac{1}{p}} \rfloor)^{\frac{p}{p-1}}\}^{\frac{p-1}{p}}$, at $\kappa = n^{\frac{1}{p}}$.*

See "Proof of Proposition 3" in Appendix for the proof.

Figure 1 shows an example of $y = f_{n,2}(\kappa)$ with $n = 1, 2, \ldots, 12$, and the minimum values are attained at $\kappa = 1, \sqrt{2}, \ldots, \sqrt{12}$, respectively.

Substituting (5) into $\kappa = n^{\frac{1}{p}}$, we reach a conclusion that the value of α with which the CVaR norm approximates the ℓ_p-norm in the best way is given by

$$\alpha^\star = 1 - n^{\frac{1}{p} - 1}. \tag{12}$$

Using the relation $1/p + 1/q = 1$, we also attain the value of α with which the dual CVaR norm approximates the ℓ_p-norm in the best way:

$$\alpha^\star = 1 - n^{-\frac{1}{p}}, \tag{13}$$

which corresponds to $\kappa = n^{1 - \frac{1}{p}}$.

Example 1 Paper [20] uses formula $p = 1/(1-\alpha)^2$ to illustrate how the CVaR norm of a vector behaves in comparison to the ℓ_p-norm. Figure 2 demonstrates the formulae

[17] By construction, the ratio U/L evaluates the proximity in a worst-case sense.

[18] The function (11) is well-defined and satisfies $f_{n,p}(\kappa) \geq 1$ for any $\kappa \in [1, n]$.

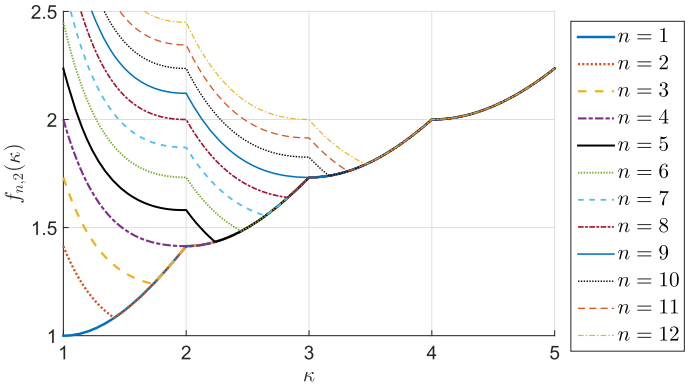
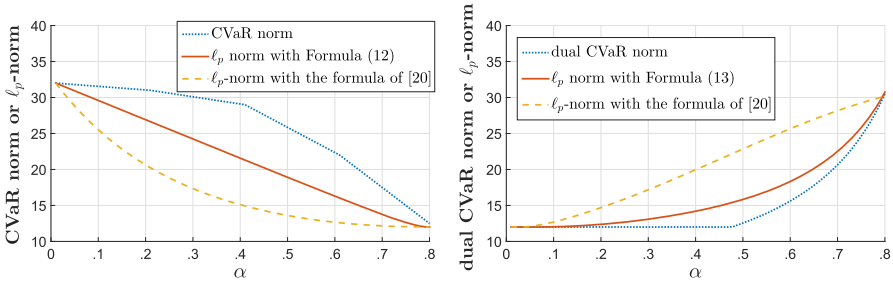


Fig. 1 Shapes of functions $y = f_{n,2}(\kappa)$ defined in (11) with $n = 1, 2, \ldots, 12$. The function $f_{n,p}(\kappa)$ yields the ratio between the tight upper and lower bounds of the ratios $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha / \|\boldsymbol{x}\|_p$ and $\|\boldsymbol{x}\|_{\frac{p}{p-1}} / \langle\!\langle \boldsymbol{x} \rangle\!\rangle^*_\alpha$. It is nonconvex on $[1, n]$ and nonsmooth at several points, but monomodal (i.e., quasi-convex). Also, it is decreasing for $\kappa \in [1, n^{1/p}]$, while increasing for $\kappa \in [n^{1/p}, n]$, and accordingly, its minimum is attained at $\kappa = n^{1/p}$



(a) CVaR norm **(b)** dual CVaR norm

Fig. 2 CVaR norm and dual CVaR norm of the vector $\boldsymbol{x} = (2, 1, 7, 10, -12)$. **a** Compares the CVaR norm versus the ℓ_p-norms with different p formulae, whereas **b** compares the dual CVaR norm versus the ℓ_q-norms with different q formulae, where p in **a** and q in **b** satisfy $1/p + 1/q = 1$. The ℓ_p- or ℓ_q-norm is computed with two different formulae for p (or q): (i) p corresponding to the best α formula (12) in **a** or (13) in **b**; (ii) $p = 1/(1-\alpha)^2$ in [20]. The former is closer to the CVaR norm (or dual CVaR norm) at every α

(12) and (13) for the same example from [20]. We see that the ℓ_p-norms based on the suggested formulae (12) and (13) give closer values to the CVaR and dual CVaR norms, respectively, than the formula of [20] does. □

Corollary 1 *With $\alpha = \alpha^\star$, defined by (12) or (13), the tight bounds (8) are given by*

$$1 \le \left\{ \begin{array}{c} \langle\!\langle \boldsymbol{x} \rangle\!\rangle_{1-n^{(1/p-1)}} / \|\boldsymbol{x}\|_p \\ \|\boldsymbol{x}\|_{\frac{p}{p-1}} / \langle\!\langle \boldsymbol{x} \rangle\!\rangle^*_{1-n^{(-1/p)}} \end{array} \right\} \le \left\{ \lfloor n^{\frac{1}{p}} \rfloor + (n^{\frac{1}{p}} - \lfloor n^{\frac{1}{p}} \rfloor)^{\frac{p}{p-1}} \right\}^{\frac{p-1}{p}}. \tag{14}$$

Example 2 Let us consider the vector $\boldsymbol{x} = (100, -101, 102)$ with $p = 500$ given in the introduction. Noting that (14) implies

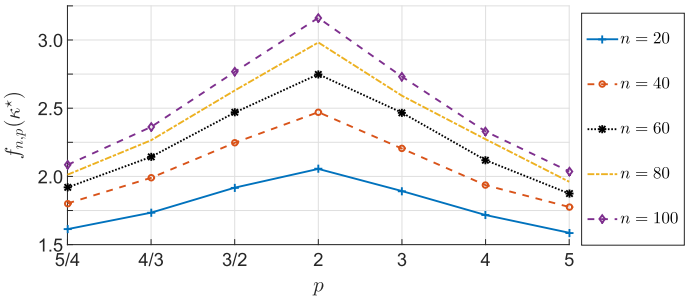


Fig. 3 Relation between $f_{n,p}(\kappa^\star)$ and p for $n = 20, 40, 60, 80, 100$. The relation between p and the best ratio $f_{n,p}(\kappa^\star)$ where $\kappa^\star$ minimizes the ratio of the upper and lower bounds of the ratio between the CVaR norm (or the dual CVaR norm) and the ℓ_p-norm (or, respectively, $\ell_{p/(p-1)}$-norm). The ratio $f_{n,p}(\kappa^\star)$ becomes the largest around $p = 2$, which indicates that the case of $p = 2$ is the hardest to approximate for the CVaR norm and its dual norm

$$\left\{\lfloor n^{\frac{1}{p}}\rfloor + (n^{\frac{1}{p}} - \lfloor n^{\frac{1}{p}}\rfloor)^{\frac{p}{p-1}}\right\}^{-\frac{p-1}{p}} \langle\!\langle \boldsymbol{x}\rangle\!\rangle_{1-n^{\frac{1}{p}-1}} \le \|\boldsymbol{x}\|_p$$
$$\le \left\{\lfloor n^{1-\frac{1}{p}}\rfloor + \left(n^{1-\frac{1}{p}} - \lfloor n^{1-\frac{1}{p}}\rfloor\right)^p\right\}^{\frac{1}{p}} \langle\!\langle \boldsymbol{x}\rangle\!\rangle^*_{1-n^{-\frac{1}{p}}},$$

we have $102.0010 \le \|\boldsymbol{x}\|_{500} \le 102.1452$.[19] □

Figure 3 shows the relation between the value of p and the best-approximating ratio $f_{n,p}(\kappa^\star)$. The ratio $f_{n,p}(\kappa^\star)$ becomes the largest around $p = 2$, which indicates that the case of $p = 2$ is the hardest to approximate for the CVaR norm and its dual norms. Indeed, (by ignoring the term $(n^{\frac{1}{p}} - \lfloor n^{\frac{1}{p}}\rfloor)^{\frac{p}{p-1}}$ in (14),) the best ratio of the upper and lower bounds ranges over $[1, n^{\frac{p-1}{p^2}}]$. Since the power of n in the right-end of the range is $-\left(\frac{1}{p} - \frac{1}{2}\right)^2 + \frac{1}{4}$, which attains the maximum at $p = 2$, the ℓ_2-norm can be regarded as the worst case among $p \in [1, \infty]$. Indeed, when $p = 2$, (12) and (13) coincide and we have

$$\alpha^\star = 1 - \frac{1}{\sqrt{n}}. \tag{15}$$

Corollary 2 *For $p = 2$, the tight bounds (14) at the best value of α, (15), is given by*

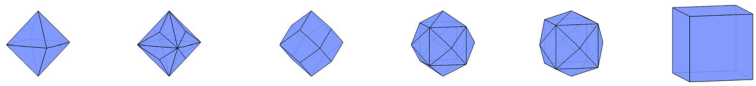
$$1 \le \left\{\begin{array}{l} \langle\!\langle \boldsymbol{x}\rangle\!\rangle_{\alpha^\star}/\|\boldsymbol{x}\|_2 \\ \|\boldsymbol{x}\|_2/\langle\!\langle \boldsymbol{x}\rangle\!\rangle^*_{\alpha^\star} \end{array}\right\} \le \sqrt{\lfloor\sqrt{n}\rfloor + (\sqrt{n} - \lfloor\sqrt{n}\rfloor)^2}. \tag{16}$$

Roughly speaking, (by ignoring the squared term in the outer square root,) the best ratio of the upper and lower bounds ranges over $[1, n^{\frac{1}{4}}]$. Noting that

$$\forall \boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}, \quad 1 \le \left\{\begin{array}{l} \|\boldsymbol{x}\|_1/\|\boldsymbol{x}\|_2 \\ \|\boldsymbol{x}\|_2/\|\boldsymbol{x}\|_\infty \end{array}\right\} \le \sqrt{n},$$

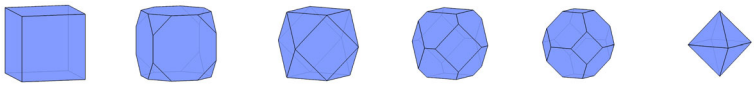
[19] These values are computed by MATLAB.

α	0	$\frac{2-\sqrt{2}}{3}$	$\frac{1}{3}$	$1-\frac{1}{\sqrt{3}}$	$\frac{1}{2}$	$\frac{2}{3}$
$\kappa = 3(1-\alpha)$	3	$\sqrt{2}+1$	2	$\sqrt{3}$	$\frac{3}{2}$	1



name	regular octahedron	triakis octahedron	rhombic dodecahedron		tetrakis hexahedron	cube
no.vertices	6	14	14	14	14	8
no.faces	8	24	12	24	24	6

(a) Unit CVaR balls $Z_\alpha := \{\boldsymbol{x} \in \mathbb{R}^3 : \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha \leq 1\}$



name	cube	truncated cube	cuboctahedron		truncated octahedron	regular octahedron
no.vertices	8	24	12	24	24	6
no.faces	6	14	14	14	14	8

(b) Unit dual CVaR balls $Z^*_\alpha := \{\boldsymbol{x} \in \mathbb{R}^3 : \langle\!\langle \boldsymbol{x} \rangle\!\rangle^*_\alpha \leq 1\}$

Fig. 4 CVaR balls and dual CVaR balls in $\mathbb{R}^3$ and correspondence to famous solids. The upper row, named **a**, shows six three-dimensional polytopes which are defined with the CVaR norm $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ with $\alpha = 0$, $(2-\sqrt{2})/3$, $1/3$, $1-1/\sqrt{3}$, $1/2$, and $2/3$. The left-most one is known as the *regular octahedron* and corresponds to the ℓ_1-norm, whereas the right-most one is known as the *cube* (aka *regular hexahedron*) and corresponds to the ℓ_∞-norm. They are two of the five *Platonic solids*, which are convex regular polytopes in $\mathbb{R}^3$. These two appear also in the lower row, named **b**, again, but in mirror-reversed positions. The other polytopes between the two regular polytopes are nontrivial examples of the CVaR norm or its dual. The dual CVaR norms $\langle\!\langle \cdot \rangle\!\rangle^*_\alpha$ with $\alpha = (2-\sqrt{2})/3$, $1/3$, $1/2$, respectively, correspond to *truncated cube*, *cuboctahedron*, *truncated octahedron*, which are three of the thirteen Archimedean solids. The CVaR norms with the same αs define *triakis octahedron*, *rhombic dodecahedron*, *tetrakis hexahedron* (aka *tetrahexahedron*), which are Catalan solids, i.e., dual solids to Archimedean ones. The polytope defined with $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ under the best α, denoted $\alpha^\star = 1-1/\sqrt{3}$, and its dual polytope, which is defined with $\langle\!\langle \cdot \rangle\!\rangle^*_\alpha$ under $\alpha^\star$, are also given above. Although this pair can be characterized as the most spherical in a sense, to our best knowledge, they seem to be unfamiliar in the literature

(16) implies that the CVaR norm or its dual can improve on the approximation of the ℓ_2-norm from $n^{\frac{1}{2}}$ to $n^{\frac{1}{4}}$. This roughly corresponds to the bounds of the ratio with the ℓ_4- or $\ell_{\frac{4}{3}}$-norm.[20]

Note also that we have $\langle\!\langle \boldsymbol{x} \rangle\!\rangle^*_{\alpha^\star} \leq \|\boldsymbol{x}\|_2 \leq \langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\alpha^\star}$ for all $\boldsymbol{x} \in \mathbb{R}^n$. This indicates that at $\alpha = \alpha^\star$, the dual pairs of the CVaR norm sandwich the ℓ_2-norm. Besides, the inequalities are tight. Indeed, $\langle\!\langle \boldsymbol{x} \rangle\!\rangle^*_{\alpha^\star} = \|\boldsymbol{x}\|_2 = \langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\alpha^\star}$ at $\boldsymbol{x} = \boldsymbol{e}_i$ for any $i \in N$, where $\boldsymbol{e}_i$ is the vector that has 1 at i-th element and 0 otherwise.

2.3 Geometric Correspondence of $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ and $\langle\!\langle \cdot \rangle\!\rangle^*_\alpha$ in $\mathbb{R}^3$

In order to enrich our understanding about the relation among the CVaR norm, its dual, and the parameter α, we provide its geometric interpretation.

Figure 4 illustrates the three-dimensional unit balls corresponding to the CVaR norm and its dual norm, i.e., $Z_\alpha := \{\boldsymbol{x} \in \mathbb{R}^3 : \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha \leq 1\}$ and $Z^*_\alpha := \{\boldsymbol{x} \in \mathbb{R}^3 : \langle\!\langle \boldsymbol{x} \rangle\!\rangle^*_\alpha \leq 1\}$. As shown in the figures, with $\alpha = \frac{2-\sqrt{2}}{3}$, $\frac{1}{3}$, and $\frac{1}{2}$, the balls of $\langle\!\langle \cdot \rangle\!\rangle_\alpha$ and $\langle\!\langle \cdot \rangle\!\rangle^*_\alpha$ in

[20] With the ℓ_p-norm, we have $1 \leq \|\boldsymbol{x}\|_2/\|\boldsymbol{x}\|_p \leq n^{\frac{1}{2}-\frac{1}{p}}$ for $p \geq 2$; $1 \leq \|\boldsymbol{x}\|_p/\|\boldsymbol{x}\|_2 \leq n^{\frac{1}{p}-\frac{1}{2}}$ for $p \leq 2$, which can be shown by using the Hölder's inequality.

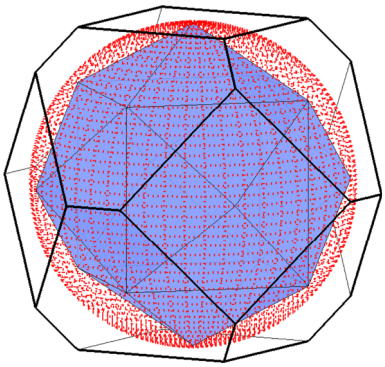


Fig. 5 Tight sandwich of Euclidean ball by two dual polytopes induced by CVaR norm and its dual norm with $\alpha^\star$. This figure illustrates the tight inclusion among the three solids in $\mathbb{R}^3$: the Euclidean unit ball, $B := \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\|_2 \leq 1\}$, and two unit balls, $Z_{\alpha^\star}$, $Z^*_{\alpha^\star}$, defined with the CVaR norm and its dual under the best parameter $\alpha^\star$, i.e., $Z_{\alpha^\star} := \{\boldsymbol{x} \in \mathbb{R}^3 : \langle\!\langle \boldsymbol{x} \rangle\!\rangle_{\alpha^\star} \leq 1\}$ and $Z^*_{\alpha^\star} := \{\boldsymbol{x} \in \mathbb{R}^3 : \langle\!\langle \boldsymbol{x} \rangle\!\rangle^*_{\alpha^\star} \leq 1\}$, respectively. Namely, $Z_{\alpha^\star} \subset B \subset Z^*_{\alpha^\star}$. These two inclusions are tight in two ways: 1) in terms of $\boldsymbol{x}$, i.e., there exists an $\boldsymbol{x}$ which attains the equality for a given α; 2) in terms of α, i.e., there is no α greater than $\alpha^\star$ which keep the including relation

$\mathbb{R}^3$ form famous solids, known as the *Catalan solids* and the *Archimedean solids*, respectively.[21] Surprisingly, to the best of our knowledge, no particular names are given to the polytopes corresponding to $\alpha^\star = 1 - \frac{1}{\sqrt{3}}$. In the sense based on the minimization of the ratio $f_{n,p}$, the pair of polytopes, $Z_{\alpha^\star}$ and $Z^*_{\alpha^\star}$, can be characterized as the most spherical polytope among the series of polytopes parametrized with α. Furthermore, the series of the solids defined by $\langle\!\langle \cdot \rangle\!\rangle^*_\alpha$ (Fig. 4 (b)) corresponds to a geometrical operation known as the *Wythoffian truncation*, which cuts off the vertices of the cube ($Z_{\frac{2}{3}} = Z^*_0$) or the regular octahedron ($Z_0 = Z^*_{\frac{2}{3}}$).

For $\alpha \in [0, \alpha^\star]$, the CVaR ball, Z_α, inscribes the Euclidean unit ball, B, and the dual CVaR ball, Z^*_α, circumscribes B, whereas for $\alpha \in (\alpha^\star, \frac{2}{3})$, Z_α and Z^*_α intersect B. In this sense, the CVaR norm and the dual CVaR norm tightly sandwich the ℓ_2-norm at $\alpha^\star = 1 - \frac{1}{\sqrt{3}}$ in $\mathbb{R}^3$ (Fig. 5). On the other hand, at $\alpha = \frac{2}{3}$, the sandwich relation again holds with the inside-out manner, i.e., $Z_{\frac{2}{3}}$ (actually, the ℓ_∞-norm ball) circumscribes B and $Z^*_{\frac{2}{3}}$ (actually, the ℓ_1-norm ball) inscribes B.

3 Approximation by a Convex Combination of ℓ_1- and ℓ_∞-Norms and Its Dual

In this section, we consider the approximation of the ℓ_p-norm by a convex combination of the ℓ_1- and ℓ_∞-norms and its dual norm.

[21] An Archimedean solid is a highly symmetric, semi-regular, three-dimensional convex polytope where two or more types of regular polygons meet at each vertex, while a Catalan solid is the polytope dual to an Archimedean solid and faces therein are not regular polygons.

3.1 Convex Combination of ℓ_1- and ℓ_∞-Norms and Its Dual

For $\lambda \in [0, 1]$, let us introduce the convex combination of the ℓ_1- and ℓ_∞-norms:

$$(\!(\boldsymbol{x})\!)_\lambda := (1-\lambda)\|\boldsymbol{x}\|_1 + \lambda\|\boldsymbol{x}\|_\infty.$$

We call this the *deltoidal norm* (of $\boldsymbol{x}$) for the reason to be described later. Obviously, for $\lambda = 0$, $(\!(\cdot)\!)_\lambda$ is equal to $\|\cdot\|_1$, while for $\lambda = 1$ it is equal to $\|\cdot\|_\infty$, and it is LP-representable for $\lambda \in [0, 1]$.

Further, we explore the dual norm of $(\!(\cdot)\!)_\lambda$, which is called the *dual deltoidal norm* in this paper. Apparently, the dual, $(\!(\cdot)\!)^*_\lambda$, is also LP-representable[22]; by using the theory of LP, we obtain a simple formula of the dual norm.

Proposition 4 *For $\lambda \in [0, 1]$, the dual deltoidal norm of $\boldsymbol{x} \in \mathbb{R}^n$ is given by*

$$(\!(\boldsymbol{x})\!)^*_\lambda = \max\left\{|x_{(1)}|, \frac{|x_{(1)}| + |x_{(2)}|}{2-\lambda}, \ldots, \frac{|x_{(1)}| + \cdots + |x_{(n)}|}{n-(n-1)\lambda}\right\}. \tag{17}$$

See "Proof of Proposition 4" in Appendix for the proof.

For example, the dual norm of a vector $\boldsymbol{x} = (2, -5, \sqrt{7})$ with $\lambda = 2 - \sqrt{2}$ is calculated by the formula as $(\!(\boldsymbol{x})\!)^*_{2-\sqrt{2}} = \max\left\{|-5|, \frac{|-5|+|\sqrt{7}|}{\sqrt{2}}, \frac{|-5|+|\sqrt{7}|+|2|}{2\sqrt{2}-1}\right\} = \frac{|-5|+|\sqrt{7}|}{\sqrt{2}}$. Similarly to the CVaR norm, $\langle\!\langle\cdot\rangle\!\rangle_\alpha$, the above representation is convenient for computing the norm, but its LP-representation is useful in the minimization of the dual deltoidal norm.[23] Along the lines of the CVaR norm and its dual, $\langle\!\langle\cdot\rangle\!\rangle_\alpha$ and $\langle\!\langle\cdot\rangle\!\rangle^*_\alpha$, as shown in the previous section, we derive symmetric pairs of tight bounds of the ratio between $\|\cdot\|_p$ and $(\!(\cdot)\!)_\lambda$ or $(\!(\cdot)\!)^*_\lambda$.

Proposition 5 *For any $p \in (1, \infty)$, $\lambda \in [0, 1]$, and $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, it is valid*

$$g_{n,p}(\lambda) \le \left\{\begin{array}{l} (\!(\boldsymbol{x})\!)_\lambda / \|\boldsymbol{x}\|_p \\ \|\boldsymbol{x}\|_q / (\!(\boldsymbol{x})\!)^*_\lambda \end{array}\right\} \le \left\{(n-1)(1-\lambda)^{\frac{p}{p-1}} + 1\right\}^{\frac{p-1}{p}}, \tag{18}$$

[22] The LP representation of the dual deltoidal norm of $\boldsymbol{x} \in \mathbb{R}^n$ is given by the dual of (43):

$$(\!(\boldsymbol{x})\!)^*_\lambda \equiv \left|\begin{array}{l} \underset{z}{\text{minimize}}\ \ \boldsymbol{1}_n^\top z \\ \text{subject to}\ \left\{\lambda \boldsymbol{I}_n + (1-\lambda)\boldsymbol{1}_n\boldsymbol{1}_n^\top\right\} z \ge |\boldsymbol{x}|,\ z \ge \boldsymbol{0}. \end{array}\right.$$

[23] The deltoidal norm and its dual can be written by using the CVaR-norm. Indeed, we have $(\!(\cdot)\!)_\lambda \equiv (1-\lambda)\langle\!\langle\cdot\rangle\!\rangle_0 + \lambda\langle\!\langle\cdot\rangle\!\rangle_{\frac{n-1}{n}}$, $(\!(\cdot)\!)^*_\lambda \equiv \max\left\{\langle\!\langle\cdot\rangle\!\rangle_{\frac{n-1}{n}}, \frac{1}{2-\lambda}\langle\!\langle\cdot\rangle\!\rangle_{\frac{n-2}{n}}, \ldots, \frac{1}{n-(n-1)\lambda}\langle\!\langle\cdot\rangle\!\rangle_0\right\}$. This representation can be useful if the CVaR operation is readily available as a built-in function.

where $\frac{1}{p}+\frac{1}{q}=1$ *and*

$$g_{n,p}(\lambda) := \begin{cases} 1, & 0 \le \lambda < \frac{p-1}{p}, \\ \min\left\{\frac{(1-\lambda)c+\lambda}{c^{\frac{1}{p}}}, \frac{(1-\lambda)c+1}{(c+1)^{\frac{1}{p}}}\right\}, & \frac{p-1}{p} \le \lambda < \frac{(p-1)n}{(p-1)n+1}, \\ \frac{n-(n-1)\lambda}{n^{\frac{1}{p}}}, & \frac{(p-1)n}{(p-1)n+1} \le \lambda \le 1, \end{cases} \tag{19}$$

where $c := \left\lfloor \frac{\lambda}{(p-1)(1-\lambda)} \right\rfloor$. *For* $p=1$, *the bounds (18) are specified as*

$$\frac{n-(n-1)\lambda}{n} \le \left\{ \begin{array}{c} ((\boldsymbol{x}))_\lambda / \|\boldsymbol{x}\|_1 \\ \|\boldsymbol{x}\|_\infty / ((\boldsymbol{x}))_\lambda^* \end{array} \right\} \le 1,$$

whereas for $p=\infty$, *they are*

$$1 \le \left\{ \begin{array}{c} ((\boldsymbol{x}))_\lambda / \|\boldsymbol{x}\|_\infty \\ \|\boldsymbol{x}\|_1 / ((\boldsymbol{x}))_\lambda^* \end{array} \right\} \le n-(n-1)\lambda.$$

All the above bounds are tight.

See "Proof of Proposition 5" in Appendix for the proof.

Especially, when $p=2$, bounds are completely symmetric.

Corollary 3 *For any* $\lambda \in [0,1]$ *and* $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, *it is valid*

$$g_{n,2}(\lambda) \le \left\{ \begin{array}{c} ((\boldsymbol{x}))_\lambda / \|\boldsymbol{x}\|_2 \\ \|\boldsymbol{x}\|_2 / ((\boldsymbol{x}))_\lambda^* \end{array} \right\} \le \sqrt{(n-1)(1-\lambda)^2+1}. \tag{20}$$

It is obvious that for $n=1$, both sides of (18) are equal to 1, independently of λ, i.e., we have $((x))_\lambda / \|x\|_p = 1$ for any $x \in \mathbb{R}$. Therefore, in the following, we consider only the case of $n \ge 2$.

Given n, let us consider the following segmentation of the range $[\frac{p-1}{p}, \frac{(p-1)n}{(p-1)n+1})$:

$$\left[\frac{p-1}{p}, \frac{2p-2}{2p-1}\right), \left[\frac{2p-2}{2p-1}, \frac{3p-3}{3p-2}\right), \ldots, \left[\frac{(p-1)(n-1)}{(p-1)(n-1)+1}, \frac{(p-1)n}{(p-1)n+1}\right).$$

Noting that $\left\lfloor \frac{\lambda}{(p-1)(1-\lambda)} \right\rfloor = k$ holds in the subrange $[\frac{(p-1)k}{(p-1)k+1}, \frac{(p-1)(k+1)}{(p-1)(k+1)+1})$, we see that on the subrange, the function $g_{n,p}(\lambda)$ is a piece-wise linear concave function consisting of two linear functions. Note also that the right-piece function in the k-th subrange, i.e., $\frac{(1-\lambda)k+1}{(k+1)^{1/p}}$, is equal to the left-piece function of the $k+1$-th subrange, i.e., $\frac{(1-\lambda)(k+1)+\lambda}{(k+1)^{1/p}}$. Consequently, letting $\hat{\lambda}_k$ be the value of λ intersecting the two linear lines within the k-th subrange, the function $g_{n,p}(\lambda)$ can be rewritten by

$$g_{n,p}(\lambda) = \begin{cases} \frac{(k-1)(1-\lambda)+1}{k^p}, & \lambda \in [\hat{\lambda}_{k-1}, \hat{\lambda}_k], \\ \frac{k(1-\lambda)+1}{(k+1)^p}, & \lambda \in [\hat{\lambda}_k, \hat{\lambda}_{k+1}), \end{cases} \quad k = 1, 2, \ldots, n-1,$$

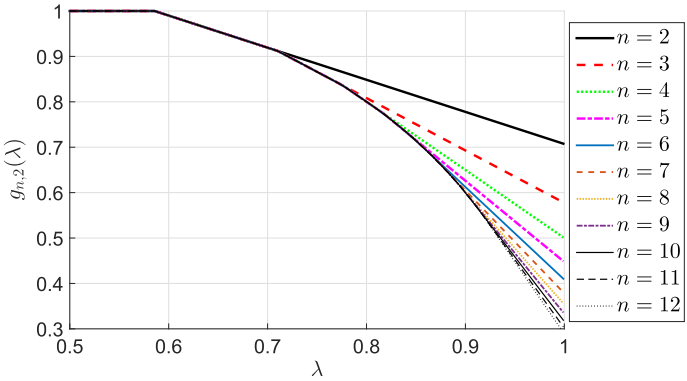


Fig. 6 Shapes of functions $y = g_{n,2}(\lambda)$ defined in (19) for $n = 2, 3, \ldots, 12$. The function $g_{n,2}(\lambda)$ yields the tight lower bound of the ratios $(\!(\boldsymbol{x})\!)_\lambda / \|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_2 / (\!(\boldsymbol{x})\!)^*_\lambda$. It is piecewise linear, nonincreasing and concave on [0, 1]. For $\lambda \geq 2 - \sqrt{2}$, it is decreasing

where

$$\hat{\lambda}_k := \frac{(k+1)k^{\frac{1}{p}} - k(k+1)^{\frac{1}{p}}}{k^{\frac{p+1}{p}} - (k-1)(k+1)^{\frac{1}{p}}}, \quad k = 0, 1, \ldots, n-1,$$

and $\hat{\lambda}_n = 1$. Note that $\hat{\lambda}_k \in (\frac{(p-1)k}{(p-1)k+1}, \frac{(p-1)(k+1)}{(p-1)(k+1)+1})$ for $k = 1, \ldots, n-1$.

Accordingly, $g_{n,p}(\lambda)$ is a piecewise-linear nonincreasing concave function over the whole range [0, 1] (Fig. 6). In particular, for $\lambda > \hat{\lambda}_1$, $g_{n,p}(\lambda)$ is decreasing (whereas for $\lambda \in [0, \hat{\lambda}_1]$ it is equal to one).

3.2 The Best Approximation of Euclidean Norm by $(\!(\cdot)\!)_\lambda$ and $(\!(\cdot)\!)^*_\lambda$

Similarly to the case of the CVaR norm and its dual, we can discuss the best values for $(\!(\cdot)\!)_\lambda$ and $(\!(\cdot)\!)^*_\lambda$ in one shot again due to the symmetry of the bounds (18). Let us introduce a function which evaluates the discrepancy between the upper bound and lower bound of (18). Given a positive integer n and $p \in (0, 1)$, let us denote it by

$$h_{n,p}(\lambda) := \frac{\left\{(n-1)(1-\lambda)^{\frac{p}{p-1}} + 1\right\}^{\frac{p-1}{p}}}{g_{n,p}(\lambda)}, \tag{21}$$

for $\lambda \in [0, 1]$.[24]

Lemma 2 *$h_{n,p}(\lambda)$ is continuous at any $\lambda \in (0, 1)$, and differentiable at any λ except at $\hat{\lambda}_k$, $k = 1, \ldots, n-1$. In particular, it is quasiconvex on each interval $[\hat{\lambda}_{k-1}, \hat{\lambda}_k]$,*

[24] The function (21) is well-defined and satisfies $h_{n,p}(\lambda) \geq 1$ for any $\lambda \in [0, 1]$.

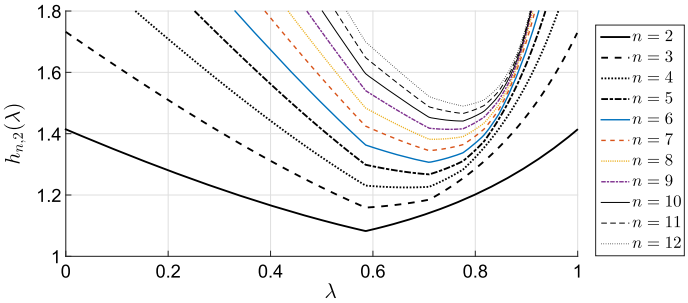


Fig. 7 Shapes of functions $y = h_{n,2}(\lambda)$ defined in (21) for $n = 2, 3, \ldots, 12$. The function $h_{n,2}(\lambda)$ yields the ratio between the tight upper and lower bounds of the ratios $(\!(\boldsymbol{x})\!)_\lambda / \|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_2 / (\!(\boldsymbol{x})\!)^*_\lambda$. It has unique minimizer for each $n \geq 2$

$k = 1, \ldots, n-1$. *Furthermore,* $h_{n,p}(\lambda)$ *is quasiconvex on the whole range* $\lambda \in [0, 1]$ *and has a unique minimum.*

See "Proof of Lemma 2" in Appendix for the proof and Fig. 7 for the graphs of $h_{n,2}$.

Indeed, λ which minimizes $h_{n,p}(\lambda)$ can take either of the forms $1 - \left(\frac{k-1}{n-1}\right)^{p-1}$ and $\hat{\lambda}_k$ for an integer $k \in \{1, \ldots, n-1\}$. Since $1 - \left(\frac{k-1}{n-1}\right)^{p-1}$ is decreasing in k while $\hat{\lambda}_k$ is increasing in k, we have to consider their relative locations on $[0, 1]$. Based on these facts, we have the following statements.

Proposition 6 *For integer* $n \geq 2$, *let* $\lambda^\star$ *denote the minimizer of* $h_{n,p}(\lambda)$.

[*Case 1:* n *satisfies* $\hat{\lambda}_{k-1} \leq 1 - \left(\frac{k-1}{n-1}\right)^{p-1} \leq \hat{\lambda}_k$ *with some* $k \in \{1, \ldots, n-1\}$]
$h_{n,p}(\lambda)$ *uniquely attains its minimum value* $k^{\frac{1}{p}} \left\{(n-1)\left(\frac{k-1}{n-1}\right)^p + 1\right\}^{-\frac{1}{p}}$ *at* $\lambda^\star = 1 - \left(\frac{k-1}{n-1}\right)^{p-1}$.

[*Case 2:* n *satisfies* $1 - \left(\frac{k}{n-1}\right)^{p-1} \leq \hat{\lambda}_k \leq 1 - \left(\frac{k-1}{n-1}\right)^{p-1}$ *with some* $k \in \{1, \ldots, n-1\}$]
$h_{n,p}(\lambda)$ *uniquely attains its minimum value*

$$\left[(n-1)\left\{(k+1)k^{\frac{1}{p}} - k(k+1)^{\frac{1}{p}}\right\}^{\frac{p}{p-1}} + \left\{k \cdot k^{\frac{1}{p}} - (k-1)(k+1)^{\frac{1}{p}}\right\}^{\frac{p}{p-1}}\right]^{\frac{p-1}{p}}$$

at $\lambda = \hat{\lambda}_k$.

Especially, when $p = 2$ and n is a square number, i.e., there is a positive integer k' such that $n = k'^2$, we can obtain a simple analytical formula. Indeed, the condition is fulfilled with $p = 2$ and $n = k^2$, and $h_{n,2}(\lambda)$ then attains its minimum value $\sqrt{\frac{\sqrt{n}+1}{2}}$ at $\lambda = \frac{\sqrt{n}}{\sqrt{n}+1}$ uniquely.

Figure 8 shows the relation between p and the best-approximating ratio $h_{n,p}(\lambda^\star)$. The ratio $h_{n,p}(\lambda^\star)$ becomes the largest around $p = 2$, which indicates that the case

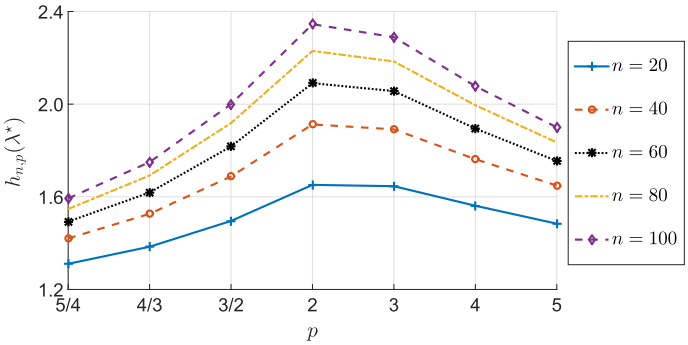


Fig. 8 Relation between $h_{n,p}(\lambda^\star)$ and p for $n = 20, 40, 60, 80, 100$. The relation between p and the best ratio $h_{n,p}(\lambda^\star)$ where $\lambda^\star$ minimizes the ratio of the upper and lower bounds of the ratio between the deltoidal norm (or the dual deltoidal norm) and the ℓ_p-norm (or, respectively, $\ell_{p/(p-1)}$-norm). The ratio $h_{n,p}(\lambda^\star)$ becomes the largest around $p = 2$, which indicates that the case of $p = 2$ is the hardest to approximate for the deltoidal norm and its dual norms. Comparing with Fig. 3, we see that $h_{n,p}(\lambda^\star)$ is smaller than $f_{n,p}(n(1-\alpha^\star))$, i.e., the deltoidal norm and its dual better approximate the ℓ_p-norm than the CVaR norm and its dual

Table 1 Comparison of the best proximities of CVaR norms and deltoidal norms

n	$\langle\!\langle\cdot\rangle\!\rangle_{\alpha^\star}$ and $\langle\!\langle\cdot\rangle\!\rangle^*_{\alpha^\star}$					$(\!(\cdot)\!)_{\lambda^\star}$ and $(\!(\cdot)\!)^*_{\lambda^\star}$			
	$\alpha^\star$	$\frac{n\alpha^\star}{n-1}$	$f^\star_{n,2}$	L	U	$\lambda^\star$	$h^\star_{n,2}$	L	U
2	0.2929	0.5858	1.0824	1.0000	1.0824	0.5858	1.0824	1.0000	1.0824
3	0.4226	0.6340	1.2393	1.0000	1.2393	0.5858	1.1589	1.0000	1.1589
4	0.5000	0.6667	1.4142	1.0000	1.4142	0.6667	1.2247	0.9428	1.1547
10	0.6838	0.7597	1.7396	1.0000	1.7396	0.7760	1.4412	0.8360	1.2048
100	0.9000	0.9091	3.1623	1.0000	3.1623	0.9091	2.3452	0.5750	1.3484
1000	0.9684	0.9693	5.6025	1.0000	5.6025	0.9692	4.0386	0.3454	1.3950

"$\alpha^\star$" in the second column shows the best α, given by the formula (15), with which the CVaR norm, $\langle\!\langle\cdot\rangle\!\rangle_{\alpha^\star}$, and its dual, $\langle\!\langle\cdot\rangle\!\rangle^*_{\alpha^\star}$, approximate the ℓ_2-norm, $\|\cdot\|_2$, in the best way. "$\lambda^\star$" in the seventh column shows the best λ, given in Proposition 6, with which the deltoidal norm, $(\!(\cdot)\!)_{\lambda^\star}$, and its dual, $(\!(\cdot)\!)^*_{\lambda^\star}$, approximate the ℓ_2-norm in the best way. The third column, "$\frac{n\alpha^\star}{n-1}$," shows $\alpha^\star$ multiplied by $n/(n-1)$, for comparison to $\lambda^\star$. The fourth column, "$f^\star_{n,2}$," shows the minimum of the ratio $f_{n,2}$, as given in Proposition 3, whereas the eighth column, "$h^\star_{n,2}$," shows the minimum of the ratio $h_{n,2}$, as given in Proposition 6. "L" in the fifth column stands for the lower bound of the ratios $\langle\!\langle \boldsymbol{x}\rangle\!\rangle_{\alpha^\star}/\|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_2/\langle\!\langle \boldsymbol{x}\rangle\!\rangle^*_{\alpha^\star}$, which is explicitly given in (8), while "L" in the ninth column stands for the lower bound of the ratios $(\!(\boldsymbol{x})\!)_{\lambda^\star}/\|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_2/(\!(\boldsymbol{x})\!)^*_{\lambda^\star}$, which is explicitly given in (20). Likewise, "U" stands for the upper bounds $\sqrt{\lfloor\sqrt{n}\rfloor + (\sqrt{n} - \lfloor\sqrt{n}\rfloor)^2}$ for the CVaR norm and its dual (in the sixth column) and $\sqrt{(n-1)(1-\lambda^\star)^2 + 1}$ for the deltoidal norm and its dual (in the rightmost column)

of $p = 2$ is the hardest to approximate for the deltoidal norm and its dual norms. Comparing with Fig. 3, we see that $h_{n,p}(\lambda^\star)$ is smaller than $f_{n,p}(n(1-\alpha^\star))$, which indicates that the deltoidal norm and its dual better approximate the ℓ_p-norm than the CVaR norm and its dual.

Table 1 demonstrates several values of $f_{n,2}$ and $h_{n,2}$ under the best parameters $\alpha^\star$ and $\lambda^\star$, respectively. The smallest value of $h_{n,2}$ is greater than $f_{n,2}$ for $n \geq 3$. This

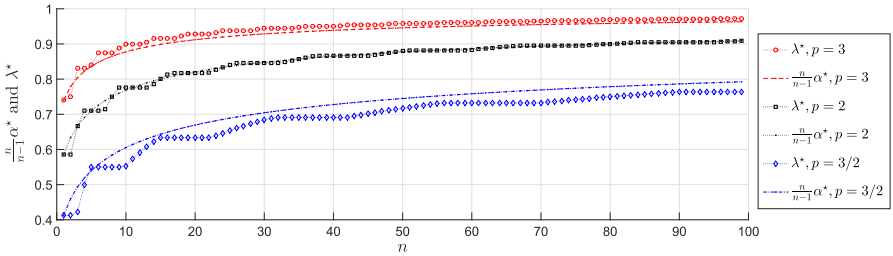


Fig. 9 $\lambda^\star$ and $\frac{n}{n-1}\alpha^\star$ when $p = 3, 2, 3/2$. The relation between the best values of α and λ, i.e., $\alpha^\star$ and $\lambda^\star$, when $p = 3, 2$, and $3/2$. Both $\alpha^\star$ and $\lambda^\star$ increase as the dimension n grows. More precisely, $\alpha^\star$ is multiplied by a constant $n/(n-1)$. After the multiplication, two sequences coincide at periodical n for $p = 2$

implies that the primal and dual deltoidal norms better approximate the ℓ_2-norm than the primal and dual CVaR norms.[25]

Besides, we see that both of $\alpha^\star$ and $\lambda^\star$ are nondecreasing in n. More specifically, there is a certain correspondence between $\alpha^\star$ and $\lambda^\star$ when $p = 2$. Figure 9 plots the relation of $\frac{n}{n-1}\alpha^\star$ and $\lambda^\star$ versus n for $p = 3, 2, \frac{3}{2}$.[26] When $p = 2$, at periodical n, such as $2, 4, 6, 9, 12, 16, 20, 25, 30, 36, 42, 49, 56, 64, 72, 81, 90, 100, \ldots$, the two sequences coincide. In this sense, we can obtain the following rule of thumb for the optimal λ when $p = 2$: $\lambda^\star \approx \frac{n}{n-1}\alpha^\star = \frac{\sqrt{n}}{\sqrt{n+1}}$.

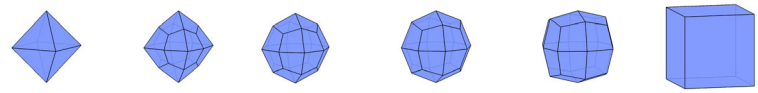
3.3 Geometric Correspondence of $(\!(\cdot)\!)_\alpha$ and $(\!(\cdot)\!)^*_\alpha$ in $\mathbb{R}^3$

As in the case of $\langle\!\langle\cdot\rangle\!\rangle_\alpha$ and $\langle\!\langle\cdot\rangle\!\rangle^*_\alpha$, we here give the geometric interpretation of $(\!(\cdot)\!)_\lambda$ and $(\!(\cdot)\!)^*_\lambda$. Figure 10 summarizes unit balls of $(\!(\cdot)\!)_\lambda$ and $(\!(\cdot)\!)^*_\lambda$ in $\mathbb{R}^3$ and the correspondence to famous solids. Similarly to the case of $\langle\!\langle\cdot\rangle\!\rangle_\alpha$ and $\langle\!\langle\cdot\rangle\!\rangle^*_\alpha$, a Catalan solid is defined by the primal norm $(\!(\cdot)\!)_\lambda$ with $\lambda^\star$ while its dual polytope (i.e., an Archimedean solid) is by the dual norm $(\!(\cdot)\!)^*_\lambda$ with $\lambda^\star$. Solids corresponding to $(\!(\cdot)\!)_{\lambda^\star}$ and $(\!(\cdot)\!)^*_{\lambda^\star}$ are known as the *deltoidal icositetrahedron*, which has 24 trapezial faces, and the *rhombicuboctahedron*, which has 26 faces consisting of 8 triangles and 18 squares, respectively, and this is the reason why we name $(\!(\cdot)\!)_\lambda$ the deltoidal norm. In contrast to $\langle\!\langle\cdot\rangle\!\rangle_\alpha$ and $\langle\!\langle\cdot\rangle\!\rangle^*_\alpha$, the shapes of the polytopes continuously transform as λ changes, i.e., the numbers of vertices and faces remain the same for any $\lambda \in (0, 1)$, and except for $\lambda = 0, 2-\sqrt{2}, 1$, no particular name is given. Besides, the transformation of the dual norm ball with respect to the parameter $\lambda \in (0, 1)$ is known as the *Wythoffian cantellation*, which cuts both vertices and edges of the cube or the regular octahedron.

Similarly to $\langle\!\langle\cdot\rangle\!\rangle^*_{\alpha^\star}$ and $\langle\!\langle\cdot\rangle\!\rangle_{\alpha^\star}$, the tight inclusion among the Euclidean ball and the two balls of $(\!(\cdot)\!)_{\lambda^\star}$ and $(\!(\cdot)\!)^*_{\lambda^\star}$ holds in $\mathbb{R}^3$ (Fig. 11). However, different from $\langle\!\langle\cdot\rangle\!\rangle^*_{\alpha^\star}$ and $\langle\!\langle\cdot\rangle\!\rangle_{\alpha^\star}$, this does not hold at $\lambda^\star$ except for $n = 2$ or 3 (see the columns 'L' and 'U'

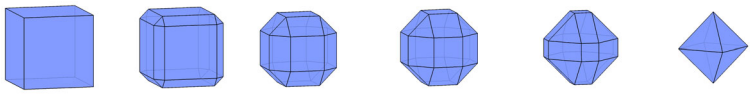
[25] In contrast, when $n = 2$, the CVaR norm and the deltoidal norm coincide and the smallest value of $h_{2,2}$ is equal to that of $f_{2,2}$.

[26] The $\alpha^\star$ is scaled by $\frac{n}{n-1}$ so that the range $[0, \frac{n-1}{n}]$ of α covers $[0, 1]$ as λ does.



λ	0	$\frac{1}{4}$	$\frac{1}{2}$	$2-\sqrt{2}$	$\frac{3}{4}$	1
name	regular octahedron			deltoidal icositetrahedron		cube
no.vertices	6	26	26	26	26	8
no.faces	8	24	24	24	24	6

(a) Unit deltoidal ball $S_\lambda := \{\boldsymbol{x} \in \mathbb{R}^3 : ((\boldsymbol{x}))_\lambda \leq 1\}$



name	cube			rhombicuboctahedron		regular octahedron
no.vertices	8	24	24	24	24	6
no.faces	6	26	26	26	26	8

(b) Unit dual deltoidal ball $S^*_\lambda := \{\boldsymbol{x} \in \mathbb{R}^3 : ((\boldsymbol{x}))^*_\lambda \leq 1\}$

Fig. 10 Deltoidal balls and dual deltoidal balls in $\mathbb{R}^3$ and correspondence to famous solids. The upper row, named **a**, shows six three-dimensional polytopes defined with the deltoidal norm with $\lambda = 0, 1/4, 1/2, 2-\sqrt{2}, 3/4$, and 1. The left-most one is known as the *regular octahedron* and corresponds to ℓ_1-norm, whereas the right-most one is known as the *cube* and corresponds to ℓ_∞-norm. They are two of the five *Platonic solids*, which are convex regular polytopes in $\mathbb{R}^3$. These two appear also in the lower row, named **b**, but in mirror-reversed positions. The other polytopes between the two regular polytopes are nontrivial examples of the deltoidal norm or the dual deltoidal norm. The deltoidal norm with the best $\lambda = 2-\sqrt{2}$ corresponds to the polytope known as the *deltoidal icositetrahedron*, and this is the reason why we name the convex combination of the ℓ_1- and ℓ_∞-norms the *deltoidal norm*. This polytope is the dual polytope to the *rhombicuboctahedron*, which corresponds to the dual deltoidal norm with the same λ. These two polytopes can be characterized as the most spherical polytope within the series of polytopes parametrized by λ. To our best knowledge, no names are assigned to polytopes located between the four named polytopes. The series of polytopes in (**b**), however, corresponds to a geometric operation called the *Wythoffian cantellation*, which cuts both vertices and edges of the cube or the regular octahedron

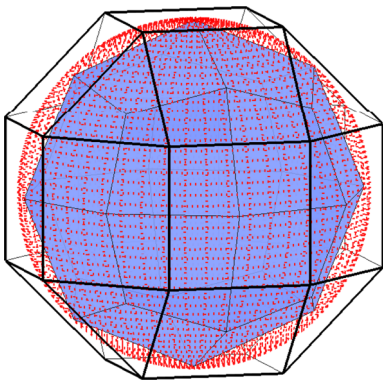


Fig. 11 Tight sandwich of Euclidean ball by two dual polytopes induced by deltoidal norm and its dual norm with $\lambda^\star$. This figure illustrates the tight inclusion among the three solids in $\mathbb{R}^3$: the Euclidean unit ball, $B := \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\|_2 \leq 1\}$, and two unit balls, $S_{\lambda^\star}$, $S^*_{\lambda^\star}$, defined with the deltoidal norm and its dual under the best parameter $\lambda^\star$, i.e., $S_{\lambda^\star} := \{\boldsymbol{x} \in \mathbb{R}^3 : ((\boldsymbol{x}))_{\lambda^\star} \leq 1\}$ and $S^*_{\lambda^\star} := \{\boldsymbol{x} \in \mathbb{R}^3 : ((\boldsymbol{x}))^*_{\lambda^\star} \leq 1\}$, respectively. Namely, $S_{\lambda^\star} \subset B \subset S^*_{\lambda^\star}$. These two inclusions are tight in two ways: 1) in terms of $\boldsymbol{x}$, i.e., there exists an $\boldsymbol{x}$ which attains the equality for a given λ; 2) in terms of λ, i.e., there is no λ greater than $\lambda^\star$ which keeps the including relation

for the deltoidal norm in Table 1). Also, comparing Figs. 5 and 11, the sandwich by the pair of deltoidal norm balls looks more stuffed than that of the CVaR norm balls, which corresponds to the difference in U/L reported as the values of $f^\star_{n,2}$ and $h^\star_{n,2}$ in the row of $n = 3$ of Table 1.

4 Applications in Optimization

This section demonstrates how the tight bounds of two norms work in the context of optimization.

4.1 Norm Minimization Problems

We start with an optimization problem of the following form:

$$\phi^\star := \left| \begin{array}{l} \underset{\boldsymbol{x},\boldsymbol{y}}{\text{minimize}} \ \ \|\boldsymbol{y}\|' \\ \text{subject to } \boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}, \ \boldsymbol{x} \in X, \end{array} \right. \tag{22}$$

where $\|\cdot\|'$ is a norm in $\mathbb{R}^r$, $\boldsymbol{R} \in \mathbb{R}^{r\times n}$, and X is a set in $\mathbb{R}^n$. For example, a quadratic program (QP):

$$(\phi^\star)^2 := \left| \begin{array}{l} \underset{\boldsymbol{x}}{\text{minimize}} \ \ \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ \text{subject to } \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \end{array} \right. \tag{23}$$

can be written in the form (22) where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is a symmetric positive semidefinite matrix, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$, and $\boldsymbol{c} \in \mathbb{R}^n$.[27] Indeed, employing the ℓ_2-norm $\|\cdot\|_2$ as $\|\cdot\|'$ and letting $\boldsymbol{Q}^{1/2} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]\mathrm{diag}(\sqrt{\theta_1}, \ldots, \sqrt{\theta_n})[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]^\top$ with θ_i eigenvalues of $\boldsymbol{Q}$ and $\boldsymbol{u}_i$ their corresponding eigenvectors, $i = 1, \ldots, n$, and setting $\boldsymbol{R} = \boldsymbol{Q}^{1/2}$ and $X = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}\}$, the QP (23) can be written in (22). Just replacing the norm $\|\cdot\|'$ by a norm $\|\cdot\|$, an alternative to (23) can be posed:

$$\hat{\phi} := \left| \begin{array}{l} \underset{\boldsymbol{x},\boldsymbol{y}}{\text{minimize}} \ \ \|\boldsymbol{y}\| \\ \text{subject to } \boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}, \ \boldsymbol{x} \in X. \end{array} \right. \tag{24}$$

Proposition 7 *Let L and U be positive constants satisfying (10). Besides, let $\phi^\star$ be the optimal value of (22),*[28] *$\hat{\phi}$ be the optimal value of (24), and $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$ be an optimal solution to (24). Then we have*

[27] Note that minimization of (seemingly more general) quadratic function of the form $\frac{1}{2}\boldsymbol{\xi}^\top \boldsymbol{D}\boldsymbol{\xi} + \boldsymbol{d}^\top \boldsymbol{\xi}$ subject to $\boldsymbol{H}\boldsymbol{\xi} \le \boldsymbol{h}$ with a positive definite matrix $\boldsymbol{D}$ can be reduced to the above form by replacing the matrix and variables by

$$\boldsymbol{Q} = \frac{1}{2}\begin{pmatrix} \boldsymbol{D} & \boldsymbol{d} \\ \boldsymbol{d}^\top & d \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} \boldsymbol{\xi} \\ \xi' \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{H} & \boldsymbol{0} \\ \boldsymbol{0}^\top & 1 \\ \boldsymbol{0}^\top & -1 \end{pmatrix}, \quad \boldsymbol{d} = \begin{pmatrix} \boldsymbol{h} \\ 1 \\ -1 \end{pmatrix},$$

for a constant $d > 0$ such that $\boldsymbol{D} - \boldsymbol{d}\boldsymbol{d}^\top/d$ is positive semidefinite.

[28] Note that $\phi^\star \equiv \|\boldsymbol{y}^\star\|'$ with an optimal solution $(\boldsymbol{x}^\star, \boldsymbol{y}^\star)$ to (22).

$$\frac{1}{U}\hat{\phi} \le \phi^\star \le \|\hat{\boldsymbol{y}}\|' \le \frac{1}{L}\hat{\phi}. \tag{25}$$

Proposition 7 indicates that the pair of tight bounds also guarantees a proximity of the approximating formulation (24) to the original formulation (22). Note that we have from (25)

$$L \le \frac{\hat{\phi}}{\phi^\star} \le U \tag{26}$$

and

$$1 \le \frac{\|\hat{\boldsymbol{y}}\|'}{\phi^\star} \le \frac{U}{L}. \tag{27}$$

Inequalities (26) indicate that the lower and upper bounds, L and U, guarantee the proximity of the optimal values of the problem (24) and the original problem (22), while (27) indicates that the ratio U/L guarantees the proximity of solutions to (24) and the solutions to (22).[29]

A possible merit of (24) is that when the condition $\boldsymbol{x} \in X$ is given by a system of linear inequalities (or equalities) it is further rewritten by a single LP. Specifically, denoting the linear constraints by $\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}$, LP-formulations of (24) with $\langle\!\langle\cdot\rangle\!\rangle_\alpha$, $\langle\!\langle\cdot\rangle\!\rangle_\alpha^*$, $(\!(\cdot)\!)_\lambda$ and $(\!(\cdot)\!)_\lambda^*$ are given by, respectively,

$$\hat{\phi}_{\langle\!\langle\cdot\rangle\!\rangle} := \left|\begin{array}{l} \underset{\boldsymbol{x},z,c}{\text{minimize}}\ n(1-\alpha)c + \boldsymbol{1}_n^\top z \\ \text{subject to } \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b},\ z \ge \boldsymbol{R}\boldsymbol{x} - c\boldsymbol{1}_n,\ z \ge -\boldsymbol{R}\boldsymbol{x} - c\boldsymbol{1}_n,\ z \ge \boldsymbol{0}, \end{array}\right.$$

$$\hat{\phi}_{\langle\!\langle\cdot\rangle\!\rangle^*} := \left|\begin{array}{l} \underset{\boldsymbol{x},\boldsymbol{y},s}{\text{minimize}}\ s \\ \text{subject to } \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b},\ n(1-\alpha)s \ge \boldsymbol{1}_n^\top \boldsymbol{y}, \\ \qquad\qquad \boldsymbol{y} \ge \boldsymbol{R}\boldsymbol{x},\ \boldsymbol{y} \ge -\boldsymbol{R}\boldsymbol{x},\ s\boldsymbol{1}_n \ge \boldsymbol{R}\boldsymbol{x},\ s\boldsymbol{1}_n \ge -\boldsymbol{R}\boldsymbol{x}, \end{array}\right.$$

$$\hat{\phi}_{(\!(\cdot)\!)} := \left|\begin{array}{l} \underset{\boldsymbol{x},\boldsymbol{y},s}{\text{minimize}}\ (1-\lambda)\boldsymbol{1}_n^\top \boldsymbol{y} + \lambda s \\ \text{subject to } \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b},\ \boldsymbol{y} \ge \boldsymbol{R}\boldsymbol{x},\ \boldsymbol{y} \ge -\boldsymbol{R}\boldsymbol{x},\ s\boldsymbol{1}_n \ge \boldsymbol{R}\boldsymbol{x}, s\boldsymbol{1}_n \ge -\boldsymbol{R}\boldsymbol{x}, \end{array}\right.$$

$$\hat{\phi}_{(\!(\cdot)\!)^*} := \left|\begin{array}{l} \underset{\boldsymbol{x},z}{\text{minimize}}\ \boldsymbol{1}_n^\top z \\ \text{subject to } \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b},\ z \ge \boldsymbol{0},\ \{\lambda \boldsymbol{I}_n + (1-\lambda)\boldsymbol{1}_n\boldsymbol{1}_n^\top\}z \ge \boldsymbol{R}\boldsymbol{x}, \\ \qquad\qquad \{\lambda \boldsymbol{I}_n + (1-\lambda)\boldsymbol{1}_n\boldsymbol{1}_n^\top\}z \ge -\boldsymbol{R}\boldsymbol{x}, \end{array}\right.$$

where $\boldsymbol{1}_n$ the all-one column vector in $\mathbb{R}^n$, and $\boldsymbol{I}_n$ the $n \times n$ unit matrix.

Let us consider the case where the ℓ_p-norm is employed as the norm in the objective of (22). For the norms studied in this paper, the bounds (25) are specified.

Corollary 4 *Let $\phi^\star$ be the optimal value of (22) with the ℓ_p-norm, i.e., $\|\cdot\|' = \|\cdot\|_p$, and let $\hat{\phi}_{\langle\!\langle\cdot\rangle\!\rangle}, \hat{\phi}_{\langle\!\langle\cdot\rangle\!\rangle^*}, \hat{\phi}_{(\!(\cdot)\!)}$, and $\hat{\phi}_{(\!(\cdot)\!)^*}$ be the optimal values of (24) and $(\boldsymbol{x}_{\langle\!\langle\cdot\rangle\!\rangle}, \boldsymbol{y}_{\langle\!\langle\cdot\rangle\!\rangle})$, $(\boldsymbol{x}_{\langle\!\langle\cdot\rangle\!\rangle^*}, \boldsymbol{y}_{\langle\!\langle\cdot\rangle\!\rangle^*})$, $(\boldsymbol{x}_{(\!(\cdot)\!)}, \boldsymbol{y}_{(\!(\cdot)\!)})$, and $(\boldsymbol{x}_{(\!(\cdot)\!)^*}, \boldsymbol{y}_{(\!(\cdot)\!)^*})$ be the corresponding optimal solutions with norms $\langle\!\langle\cdot\rangle\!\rangle_\alpha$, $\langle\!\langle\cdot\rangle\!\rangle_\alpha^*$, $(\!(\cdot)\!)_\lambda$, and $(\!(\cdot)\!)_\lambda^*$, respectively. Then,*

[29] (27) explains that minimizing the ratio U/L makes sense for better approximation.

$$\left\{\lfloor\kappa\rfloor+(\kappa-\lfloor\kappa\rfloor)^{\frac{p}{p-1}}\right\}^{\frac{1-p}{p}}\hat{\phi}_{\langle\langle\cdot\rangle\rangle}\le\phi^\star\le\|\boldsymbol{y}_{\langle\langle\cdot\rangle\rangle}\|_p\le\max\left\{1,n^{\frac{1-p}{p}}(1-\alpha)^{-1}\right\}\cdot\hat{\phi}_{\langle\langle\cdot\rangle\rangle},$$
$$\min\left\{1,n^{\frac{1}{p}}(1-\alpha)\right\}\cdot\hat{\phi}_{\langle\langle\cdot\rangle\rangle^*}\le\phi^\star\le\|\boldsymbol{y}_{\langle\langle\cdot\rangle\rangle^*}\|_p\le\{\lfloor\kappa\rfloor+(\kappa-\lfloor\kappa\rfloor)^p\}^{\frac{1}{p}}\hat{\phi}_{\langle\langle\cdot\rangle\rangle^*},$$
$$\left\{(n-1)(1-\lambda)^{\frac{p}{p-1}}+1\right\}^{\frac{1-p}{p}}\hat{\phi}_{(\!(\cdot)\!)}\le\phi^\star\le\|\boldsymbol{y}_{(\!(\cdot)\!)}\|_p\le\{g_{n,p}(\lambda)\}^{-1}\hat{\phi}_{(\!(\cdot)\!)},$$
$$g_{n,p/(p-1)}(\lambda)\cdot\hat{\phi}_{(\!(\cdot)\!)^*}\le\phi^\star\le\|\boldsymbol{y}_{(\!(\cdot)\!)^*}\|_p\le\{(n-1)(1-\lambda)^p+1\}^{\frac{1}{p}}\hat{\phi}_{(\!(\cdot)\!)^*},$$

where $\kappa=n(1-\alpha)$.

In addition, the use of the best values, $\alpha^\star$ and $\lambda^\star$, guarantees the best approximation of the solutions in the sense of (27).

4.2 Cone-constrained Optimization Problems

We further apply the norms to an optimization problem with conic constraints.

$$\phi^\star:=\left|\begin{array}{l}\underset{\boldsymbol{x}\in X}{\text{minimize}}\ \boldsymbol{c}_0^\top\boldsymbol{x}\\ \text{subject to }\|\boldsymbol{A}_i\boldsymbol{x}-\boldsymbol{b}_i\|'\le\boldsymbol{c}_i^\top\boldsymbol{x}-d_i,\quad i=1,\dots,\ell,\end{array}\right.\tag{28}$$

where $\boldsymbol{c}_0,\boldsymbol{c}_i\in\mathbb{R}^n$, $\boldsymbol{A}_i\in\mathbb{R}^{m\times n}$, $\boldsymbol{b}_i\in\mathbb{R}^m$, $d_i\in\mathbb{R}$, $i=1,\dots,\ell$. Note that (28) can be considered as a generalized version of (22).

By replacing the norm $\|\cdot\|'$ in (28) with another norms, denoted $\|\cdot\|$, we can consider approximated problems.

$$\hat{\phi}:=\left|\begin{array}{l}\underset{\boldsymbol{x}\in X}{\text{minimize}}\ \boldsymbol{c}_0^\top\boldsymbol{x}\\ \text{subject to }\|\boldsymbol{A}_i\boldsymbol{x}-\boldsymbol{b}_i\|\le\boldsymbol{c}_i^\top\boldsymbol{x}-d_i,\quad i=1,\dots,\ell,\end{array}\right.\tag{29}$$

$$\phi_\nu:=\left|\begin{array}{l}\underset{\boldsymbol{x}\in X}{\text{minimize}}\ \boldsymbol{c}_0^\top\boldsymbol{x}\\ \text{subject to }\nu\|\boldsymbol{A}_i\boldsymbol{x}-\boldsymbol{b}_i\|\le\boldsymbol{c}_i^\top\boldsymbol{x}-d_i,\quad i=1,\dots,\ell,\end{array}\right.\tag{30}$$

where ν is a parameter for indicating the constant in front of the norm $\|\cdot\|$ in the constraints. We prove the following fact.

Proposition 8 *Let L and U be positive constants satisfying (10). Besides, let $\phi^\star$, $\hat{\phi}$, and ϕ_ν be the optimal values of (28), (29), and (30), respectively. Then,*

$$\phi_{\frac{L}{U}}\le\hat{\phi}\le\phi_{\frac{U}{L}},\quad\phi_{\frac{1}{U}}\le\phi^\star\le\phi_{\frac{1}{L}}.\tag{31}$$

Especially, when $L=1$, we have $\phi_{\frac{1}{U}}\le\phi^\star\le\hat{\phi}\le\phi_U$. On the other hand, when $U=1$, we have $\phi_L\le\hat{\phi}\le\phi^\star\le\phi_{\frac{1}{L}}$.

The bounds (31) are slightly less informative than (25). This can be the price of the generality. However, as in the case of the CVaR norm or the dual CVaR norm to the ℓ_p-norm under $\alpha^\star$, if $L=1$ holds, we can achieve the similar bounds to the case where only the objective includes the norm as in (25).

Another interesting view on (29) can be found in a relation to the conic programming (e.g., [6]). For simplicity, let us consider the case where $X = \mathbb{R}^n$. Then, the dual problem to (29) is

$$\hat{\hat{\phi}} := \left| \begin{array}{l} \underset{\{z_i\},\boldsymbol{w}}{\text{maximize}} \ \sum_{i=1}^{\ell} \left(\boldsymbol{b}_i^\top z_i + d_i w_i \right) \\ \text{subject to} \ \sum_{i=1}^{\ell} \left(\boldsymbol{A}_i^\top z_i + \boldsymbol{c}_i w_i \right) = \boldsymbol{c}_0, \quad \|z_i\|^* \le w_i, \quad i = 1, \ldots, \ell. \end{array} \right. \tag{32}$$

If $\|\cdot\| = \|\cdot\|_p$, this pair is pOCP and the strong duality, i.e., $\hat{\phi} = \hat{\hat{\phi}}$, holds under the strict feasibility condition. On the other hand, if we employ an LP-representable norm, the strong duality holds without the strict feasibility condition because (29) and (32) can be equivalently rewritten by LPs.

4.3 Probabilistic Guarantee for Robust Optimization

The tight bounds of the ratio between two norms can be utilized also in robust optimization. As shown in [4], the CVaR norm (D-norm in their term) leads to a tractable robust counterpart with a probabilistic guarantee. Here we simply apply the methods developed to extend the results from [4].

Let us consider an uncertain LP:

$$\left| \begin{array}{l} \underset{\boldsymbol{x} \in X}{\text{minimize}} \ \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \ \tilde{\boldsymbol{A}} \boldsymbol{x} \le \boldsymbol{b}, \ \text{for all} \ \tilde{\boldsymbol{A}} \subset \mathcal{U}, \end{array} \right.$$

where X is a set consisting of constraints involving only certain coefficients and $\mathcal{U}$ is the uncertainty set given by $\mathcal{U} := \{\tilde{\boldsymbol{A}} \in \mathbb{R}^{m \times n} : \|\boldsymbol{M}(\mathrm{vec}(\tilde{\boldsymbol{A}}) - \mathrm{vec}(\bar{\boldsymbol{A}}))\| \le \delta\}$ with $\boldsymbol{M}$ an invertible matrix, $\delta \ge 0$, and $\mathrm{vec}(\boldsymbol{A}) \in \mathbb{R}^{mn}$ is the vector obtained by stacking its rows on top of each other. This can be reduced to

$$\left| \begin{array}{l} \underset{\boldsymbol{x} \in X}{\text{minimize}} \ \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \ \bar{\boldsymbol{a}}_i^\top \boldsymbol{x} + \delta \|(\boldsymbol{M}^\top)^{-1} \boldsymbol{x}_i\|^* \le \boldsymbol{b}_i, \quad i = 1, \ldots, m, \end{array} \right. \tag{33}$$

where $\boldsymbol{x}_i \in \mathbb{R}^{mn}$ is a vector that contains $\boldsymbol{x} \in \mathbb{R}^n$ in entries $(i-1) \cdot n + 1$ through $i \cdot n$, and 0 everywhere else (see [4] for the details).

In addition, we assume $\mathrm{vec}(\tilde{\boldsymbol{A}})$ follows an unknown distribution, but it is known to have the expected value $\mathrm{vec}(\bar{\boldsymbol{A}}) \in \mathbb{R}^{mn}$ and the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{mn \times mn}$.

Proposition 9 ([4]) *Let $\boldsymbol{x}^\star$ be an optimal solution to (33). The probability that $\boldsymbol{x}^\star$ satisfies constraint i for any realization of the uncertainty matrix $\tilde{\boldsymbol{A}}$ is*

$$\mathbb{P}\left\{ \tilde{\boldsymbol{a}}_i^\top \boldsymbol{x}^\star \le b_i \right\} = \mathbb{P}\left\{ (\mathrm{vec}(\tilde{\boldsymbol{A}}))^\top \boldsymbol{x}_i^\star \le b_i) \right\} \ge 1 - \frac{1}{1 + \delta^2 (\|\boldsymbol{\Sigma}^{1/2} \boldsymbol{x}_i^\star\|^* / \|\boldsymbol{\Sigma}^{1/2} \boldsymbol{x}_i^\star\|_2)^2}.$$

By applying the ratio of the tight bounds, we obtain the bounds depending on any norm.

Corollary 5 *For a norm $\|\cdot\|$, let L and U be constants such that for any $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, it is valid $L \leq \|\boldsymbol{x}\|/\|\boldsymbol{x}\|_2 \leq U$. If the norm used in $\mathcal{U}$ is $\|\cdot\|$, then the probability bound is given explicitly by*

$$\mathbb{P}\left\{\tilde{\boldsymbol{a}}_i^\top \boldsymbol{x}^\star \leq b_i\right\} \geq 1 - \frac{1}{1 + (\delta/U)^2}.$$

If the norm used in $\mathcal{U}$ is $\|\cdot\|^*$, then the probability bound is given explicitly by

$$\mathbb{P}\left\{\tilde{\boldsymbol{a}}_i^\top \boldsymbol{x}^\star \leq b_i\right\} \geq 1 - \frac{1}{1 + (\delta L)^2}.$$

Using this result, Theorem 3 (b) of [4] can be sharpened:

Corollary 6 *If the norm used in $\mathcal{U}$ is $\langle\!\langle\cdot\rangle\!\rangle_\alpha$ (or $|||\cdot|||_\kappa$), then*

$$\mathbb{P}\left\{\tilde{\boldsymbol{a}}_i^\top \boldsymbol{x}^\star \leq b_i\right\} \geq 1 - \frac{1}{1 + \delta^2/\{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^2\}},$$

where $\kappa = n(1-\alpha)$.

Also, the bounds become weaker as the bound L or $1/U$ is further below 1, which can be a suggestion for determination of the norm describing the uncertainty set. For example, if we employ $\alpha^\star$, the dual CVaR norm is preferable to the CVaR norm for describing the uncertain set.

5 Case Studies

To examine the proximity based on a worst-case analysis of the ratio between two norms, we provide three case studies: 1) Projection of a point on a polyhedron; 2) Mean-variance portfolio selection; 3) Robust LP problem.

5.1 Projection of a Point on a Polyhedron

The projection of a point $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ on a polyhedron is formulated as

$$\phi^\star := \left| \begin{array}{l} \underset{\boldsymbol{x}}{\text{minimize}} \ \|\boldsymbol{x} - \bar{\boldsymbol{x}}\|_2 \\ \text{subject to } \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}, \end{array} \right.$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$.[30] We compare the objective value and solution of this quadratic optimization problem to those of LPs defined via the six LP-representable

[30] Instances were generated as follows. We set $\bar{\boldsymbol{x}} = \boldsymbol{0}$ and $\boldsymbol{b} = \boldsymbol{1}_m$. The matrices $\boldsymbol{A}$ were randomly generated so that the reciprocal of each element followed a uniform distribution over (0, 1), i.e., $1/a_{ij} \sim \mathrm{U}(0, 1)$.

norms, $\|\cdot\|_1$, $\|\cdot\|_\infty$, $\langle\!\langle\cdot\rangle\!\rangle_\alpha$, $\langle\!\langle\cdot\rangle\!\rangle^*_\alpha$, $(\!(\cdot)\!)_\lambda$, and $(\!(\cdot)\!)^*_\lambda$, under the optimal parameter $\alpha^\star$ or $\lambda^\star$.[31] Below we denote by $\hat{\boldsymbol{x}}$ and $\hat{\phi}$ an optimal solution and the optimal value, respectively, of the optimization with an LP-representable norm. If some clarification is required, we denote them by indicating what norm corresponds to. For example, if a solution is obtained by the ℓ_1-norm, we denote the solution by $\boldsymbol{x}_1$ and its optimal value by $\|\boldsymbol{x}_1\|_1$; if a solution is obtained by the CVaR norm, we denote the solution by $\boldsymbol{x}_{\langle\!\langle\cdot\rangle\!\rangle}$ and its optimal value by $\langle\!\langle\boldsymbol{x}_{\langle\!\langle\cdot\rangle\!\rangle}\rangle\!\rangle_{\alpha^\star}$.

Proximity of optimal solutions. Table 2 shows the average and standard deviation of the ratio, $\hat{\phi}/\phi^\star$ or $\phi^\star/\hat{\phi}$, over 1000 instances.[32] Obviously, solutions to six LPs with the LP-representable norms under the best parameters are located within the range as theoretically guaranteed by (26). The average values are stably in the same order as the theoretical bounds tell. Due to that, comparing to the ℓ_1- and ℓ_∞-norms, the increase of the ratio for the generalized four norms is rather moderate as n grows. In addition, the ℓ_1- and ℓ_∞-norms show larger standard deviation than the pair of the CVaR norms and the pair of the deltoidal norms. It is noteworthy that the average value is much smaller than the theoretical upper bounds.

Figure 12 shows the percentiles of the ratio. First, we see that at every level of percentage, the ordering is stable. Second, especially the deltoidal norm and its dual are the closest to the value of one, which implies that their optimal values very well approximate the minimum Euclidean distance between a point to a polyhedron. On the other hand, the optimal values of the ℓ_1- and ℓ_∞-norm minimizations widely fluctuate and they can be away from the minimum Euclidean distance. This explains the reason for the large standard deviation in Table 2.

Solution accuracy. Next let us evaluate the approximation accuracy in terms of the quadratic objective, i.e., (27). Table 3 and Fig. 13 report statistics of the ratio of the ℓ_2-norm of the obtained LP solutions, i.e., $\|\hat{\boldsymbol{x}}\|_2$, and the true optimal value of the quadratic optimization, i.e., $\phi^\star = \|\boldsymbol{x}_2\|_2$. We see that, as already observed in Table 2 and Fig. 12, the obtained solutions with the deltoidal norm and its dual achieve better solution accuracy than the others, followed by the CVaR norms and its dual. In particular, from Fig. 13 we see that the performance of the ℓ_1- and ℓ_∞-norm solutions deteriorate for a small portion of instances ranked within the bottom 5%. This indicates that the performance of the ℓ_1- or ℓ_∞-approximation is susceptible to the given positions of a point and a polytope.

Footnote 30 continued
The computation ran over 1000 instances for each combination of $(n, m) \in \{(2, 1), (5, 2), (10, 3), (20, 4), (50, 7), (100, 10)\}$, where m was chosen equal to $\lfloor\sqrt{n}\rfloor$ in order for optimal solutions not to be crowded in a certain area. In the following, only the results of $n = 2, 20, 100$ are presented for reducing the number of pages. See the discussion paper version [11] for the results of $n = 5, 10, 50$.

[31] Computations in Sections 5.1 and 5.2 were done in MATLAB on a laptop PC, Windows 7 OS. We used IBM ILOG CPLEX 12.5 for solving the QPs (ℓ_2-norm) and LPs (six LP-representable norms).

[32] To be comparable within each dual pair (e.g., the ℓ_1-norm minimizer and the ℓ_∞-norm minimizer), the positions of numerators and denominators are arranged so that the ratio is greater than 1.

Table 2 Average and standard deviation of ratio $\hat{\phi}/\phi^\star$ (or $\phi^\star/\hat{\phi}$) for the projection problem

n	ℓ_1- and ℓ_∞-norms				CVaR norm and its dual				deltoidal norm and its dual			
	L	$\frac{\lVert \boldsymbol{x}_1 \rVert_1}{\phi^\star}$	$\frac{\phi^\star}{\lVert \boldsymbol{x}_\infty \rVert_\infty}$	U	L	$\frac{\langle\!\langle \boldsymbol{x}_{\langle\langle\rangle\rangle} \rangle\!\rangle_{\alpha^\star}}{\phi^\star}$	$\frac{\phi^\star}{\langle\!\langle \boldsymbol{x}_{\langle\langle\rangle\rangle^*} \rangle\!\rangle^*_{\alpha^\star}}$	U	L	$\frac{(\!(\boldsymbol{x}_{(\!(\rangle\!)})\!)_{\lambda^\star}}{\phi^\star}$	$\frac{\phi^\star}{(\!(\boldsymbol{x}_{(\!(\rangle\!)^*})\!)^*_{\lambda^\star}}$	U
2	1	1.1531	1.2998	1.4142	1	1.0244	1.0524	1.0824	1	1.0244	1.0524	1.0824
		(0.1270)	(0.1210)			(0.0231)	(0.0256)			(0.0231)	(0.0256)	
20	1	2.2309	3.4319	4.4721	1	1.2839	1.6329	2.0550	0.7741	0.9532	1.0893	1.2785
		(0.3450)	(0.3951)			(0.1114)	(0.1005)			(0.0433)	(0.0554)	
100	1	3.8893	6.7812	10.0000	1	1.4539	2.1868	3.1623	0.5750	0.7770	0.9902	1.3484
		(0.4426)	(0.6202)			(0.1081)	(0.1238)			(0.0302)	(0.0389)	

The average and standard deviation (in the parentheses) of the ratio of the optimal value of each LP, $\hat{\phi}$, and the minimum Euclidean norm, $\phi^\star$, among the 1000 randomly generated instances. For comparison, the lower and upper bounds, denoted by L and U, respectively, are also reported on the both sides of the ratios. The values of L and U for the CVaR norm and its dual are given by the left-hand side and right-hand side of (16), respectively, while those for the deltoidal norm and its dual are given by $g_{n,p}(\lambda^\star)$, which is defined in (19), and the right-hand side of (20) with $\lambda = \lambda^\star$, respectively

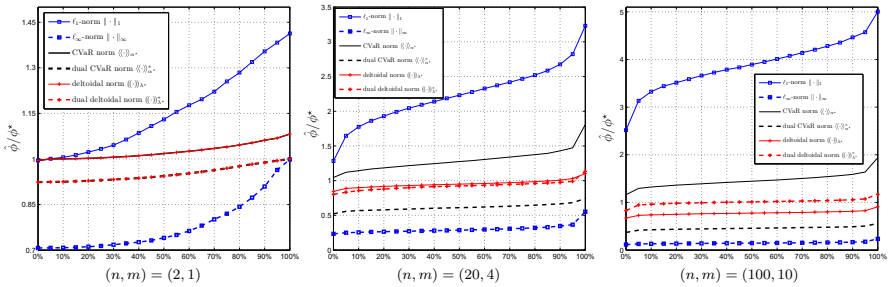


Fig. 12 Percentiles of ratio $\hat{\phi}/\phi^\star$ for the projection problem (x-axis: %; y-axis: $\hat{\phi}/\phi^\star$). The percentiles of the ratio of $\hat{\phi}$ and $\phi^\star$ among the 1000 randomly generated instances, where $\hat{\phi}$ denotes the optimal values of the problem (24) with an LP-representable norm while $\phi^\star$ denotes the optimal value of the original problem (22)

Table 4 shows the distance of the LP solutions to the true QP solutions. We see that the dual deltoidal norm solution is closer to the ℓ_2-norm solution, as indicated by the ratios in Tables 2 and 3.

Large-scale case study with Portfolio Safeguard [21]. One of the merits of using LP-representable norms is the solvability of large-scale problems. For example, an instance of the projection problem with $(n, m) = (1000, 900000)$ can be solved in 4654.56 seconds with Portfolio Safeguard, a state-of-the-art solver software, on a single personal computer. Portfolio Safeguard solves very efficiently problems with CVaR, ℓ_1-, ℓ_2-, ℓ_∞- and deltoidal norm because these norms are included in the library of precoded nonlinear functions. Interested readers can be referred to the webpage with the case study.[33]

5.2 Markowitz's Mean-Variance Model for Portfolio Selection

Given n investable assets, the so-called Markowitz's mean-variance model is formulated in the following form:

$$(\phi^\star)^2 := \left| \begin{array}{l} \underset{\boldsymbol{x}}{\text{minimize}} \ \ \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x} \\ \text{subject to } \boldsymbol{r}^\top \boldsymbol{x} = d, \ \mathbf{1}_n^\top \boldsymbol{x} = 1, \ (\boldsymbol{x} \geq \mathbf{0}), \end{array} \right. \tag{34}$$

where d is a parameter representing a desired level of mean return, $\boldsymbol{r} \in \mathbb{R}^n$ is the mean return vector of n assets, and $\boldsymbol{\Sigma}$ is an $n \times n$ positive semidefinite symmetric matrix representing the covariance matrix of the n asset returns. As described in Sect. 4.1, the QP (34) is rewritten by

$$\phi^\star := \left| \begin{array}{l} \underset{\boldsymbol{x}, \boldsymbol{y}}{\text{minimize}} \ \ \|\boldsymbol{y}\|_2 \\ \text{subject to } \boldsymbol{r}^\top \boldsymbol{x} = d, \ \mathbf{1}_n^\top \boldsymbol{x} = 1, \ (\boldsymbol{x} \geq \mathbf{0}), \ \boldsymbol{y} = \boldsymbol{M}\boldsymbol{x}, \end{array} \right. \tag{35}$$

[33] The link to the large-scale case study on projection with different norms: http://www.ise.ufl.edu/uryasev/research/testproblems/advanced-statistics/case-study-projection-on-polyhedron-with-cvar-absolute-norm-4/.

Table 3 Average and standard deviation of ratio $\|\hat{\boldsymbol{x}}\|_2/\phi^\star$ for the projection problem

n	ℓ_1- and ℓ_∞-norms			CVaR norm and its dual			deltoidal norm and its dual		
	$\frac{\|\boldsymbol{x}_1\|_2}{\phi^\star}$	$\frac{\|\boldsymbol{x}_\infty\|_2}{\phi^\star}$	$\frac{U}{L}$	$\frac{\|\boldsymbol{x}_{\langle\langle\cdot\rangle\rangle}\|_2}{\phi^\star}$	$\frac{\|\boldsymbol{x}_{\langle\langle\cdot\rangle\rangle^*}\|_2}{\phi^\star}$	$\frac{U}{L}$	$\frac{\|\boldsymbol{x}_{(\!(\cdot)\!)}\|_2}{\phi^\star}$	$\frac{\|\boldsymbol{x}_{(\!(\cdot)\!)^*}\|_2}{\phi^\star}$	$\frac{U}{L}$
2	1.1430	1.1050	1.4142	1.0241	1.0277	1.0824	1.0241	1.0277	1.0824
	(0.1279)	(0.1175)		(0.0334)	(0.0332)		(0.0334)	(0.0332)	
20	1.5309	1.3228	4.4721	1.2138	1.2236	2.0550	1.1686	1.1018	1.6515
	(0.2766)	(0.1753)		(0.0952)	(0.0785)		(0.0543)	(0.0494)	
100	1.7568	1.4880	10.0000	1.3829	1.3431	3.1623	1.2527	1.1736	2.3452
	(0.2642)	(0.1480)		(0.0984)	(0.0770)		(0.0489)	(0.0574)	

The average and standard deviation (in the parentheses) of the ratio of $\|\hat{\boldsymbol{x}}\|_2$ and $\phi^\star$ among the 1000 randomly generated instances, where $\hat{\boldsymbol{x}}$ denotes an obtained solution to the problem of the form (24) with an LP-representable norm while $\phi^\star$ denotes the optimal value of the original problem of the form (22). The value of U/L for the CVaR norm and its dual is given by the right-hand side of (16), while that for the deltoidal norm and its dual is given by $h_n(\lambda^\star)$ in Proposition 6

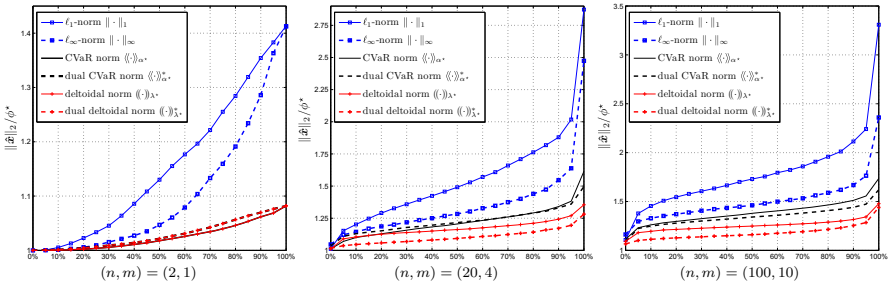


Fig. 13 Percentiles of ratio $\|\hat{\boldsymbol{x}}\|_2/\phi^\star$ for the projection problem (x-axis: %; y-axis: $\|\hat{\boldsymbol{x}}\|_2/\phi^\star$). The percentiles of the ratio of $\|\hat{\boldsymbol{x}}\|_2$ and $\phi^\star$ among the 1000 randomly generated instances, where $\hat{\boldsymbol{x}}$ denotes an obtained solution to the problem (24) with an LP-representable norm while $\phi^\star$ denotes the optimal value of the original problem (22)

Table 4 Average and standard deviation of distance $\|\hat{\boldsymbol{x}}-\boldsymbol{x}_2\|_2$ from QP solution $\boldsymbol{x}_2$ for the projection problem

n	$\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2$	$\|\boldsymbol{x}_\infty-\boldsymbol{x}_2\|_2$	$\|\boldsymbol{x}_{\langle\langle\cdot\rangle\rangle}-\boldsymbol{x}_2\|_2$	$\|\boldsymbol{x}_{\langle\langle\cdot\rangle\rangle^*}-\boldsymbol{x}_2\|_2$	$\|\boldsymbol{x}_{(\!(\cdot)\!)}-\boldsymbol{x}_2\|_2$	$\|\boldsymbol{x}_{(\!(\cdot)\!)^*}-\boldsymbol{x}_2\|_2$
2	0.1703	0.0672	0.0503	0.0576	0.0503	0.0576
	(0.1633)	(0.0415)	(0.0403)	(0.0625)	(0.0403)	(0.0625)
20	0.0718	0.0429	0.0368	0.0428	0.0360	0.0266
	(0.0427)	(0.0093)	(0.0121)	(0.0205)	(0.0146)	(0.0127)
100	0.0263	0.0171	0.0154	0.0162	0.0135	0.0107
	(0.0097)	(0.0024)	(0.0029)	(0.0046)	(0.0032)	(0.0030)

The average and standard deviation (in the parentheses) of the Euclidean distance between the ℓ_2-norm minimizer, $\boldsymbol{x}_2$, and each LP-based solution, i.e., $\boldsymbol{x}_1$, $\boldsymbol{x}_\infty$, $\boldsymbol{x}_{\langle\langle\cdot\rangle\rangle}$, $\boldsymbol{x}_{\langle\langle\cdot\rangle\rangle^*}$, $\boldsymbol{x}_{(\!(\cdot)\!)}$, or $\boldsymbol{x}_{(\!(\cdot)\!)^*}$, for the 1000 randomly generated instances

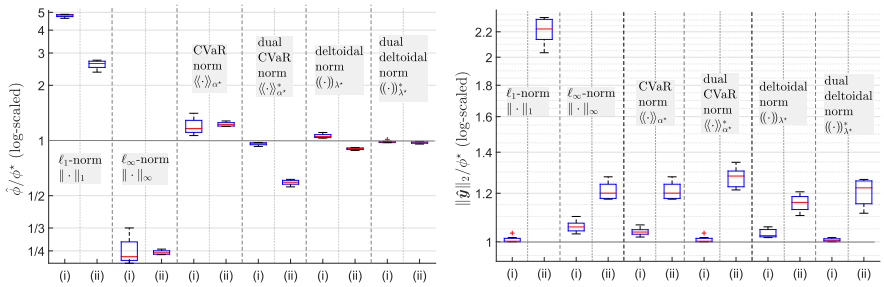
for some matrix $\boldsymbol{M}\in\mathbb{R}^{m\times n}$.[34]

[17] reformulates the problem by employing the ℓ_1-norm in place of the ℓ_2-norm. This criterion is known as the MAD (mean-absolute deviation) model. In contrast, we below approach the Markowitz's portfolio by solving

$$\hat{\phi} := \left| \begin{array}{l} \underset{\boldsymbol{x},\boldsymbol{y}}{\text{minimize}} \ \ \|\boldsymbol{y}\| \\ \text{subject to } \boldsymbol{r}^\top\boldsymbol{x}=d,\ \mathbf{1}_n^\top\boldsymbol{x}=1,\ (\boldsymbol{x}\geq\mathbf{0}),\ \ \boldsymbol{y}=\boldsymbol{M}\boldsymbol{x}, \end{array} \right. \tag{36}$$

with the norms, $\langle\langle\cdot\rangle\rangle_{\alpha^\star}$, $\langle\langle\cdot\rangle\rangle^*_{\alpha^\star}$, $(\!(\cdot)\!)_{\lambda^\star}$, and $(\!(\cdot)\!)^*_{\lambda^\star}$, in place of the norm $\|\cdot\|$ in the objective, and compare with the MAD model, $\|\cdot\|_1$, in the approximation accuracy to the Markowitz portfolio.

[34] With m historical returns $\boldsymbol{R}_1,\ldots,\boldsymbol{R}_m\in\mathbb{R}^n$, $\boldsymbol{r}$ is often estimated by $(\boldsymbol{R}_1,\ldots,\boldsymbol{R}_m)\mathbf{1}_m/m$ and $\boldsymbol{\Sigma}$ is estimated by $\boldsymbol{M}^\top\boldsymbol{M}$ with $\boldsymbol{M}=(\boldsymbol{R}_1-\boldsymbol{r},\ldots,\boldsymbol{R}_m-\boldsymbol{r})^\top/\sqrt{m}$. Note that the number of elements in $\boldsymbol{y}$ is equal to m, which is not necessarily equal to n.



(a) $\hat{\phi}/\phi^\star$: Ratio of two different objectives **(b)** $\|\hat{y}\|_2/\phi^\star$: Ratio of quadratic objectives

Fig. 14 Two ratios of LP approaches (36) for Markowitz's problem (35): (i) With $\boldsymbol{x} \geq \boldsymbol{0}$ versus (ii) Without $\boldsymbol{x} \geq \boldsymbol{0}$. **a** $\hat{\phi}/\phi^\star$: Ratio of two different objectives and **b** $\|\hat{\boldsymbol{y}}\|_2/\phi^\star$: Ratio of quadratic objectives. **a** shows the ratio between the optimal value of the approximating LP-representable problem (36) and the optimal value of the ℓ_2-norm-based problem (35), i.e., $\hat{\phi}/\phi^\star$, whereas **b** shows the ratio between the standard deviation of the resulting portfolio and the minimum standard deviation, i.e., $\|\hat{\boldsymbol{y}}\|_2/\phi^\star$. Case (i) shows results with short-sale constraint, i.e., $\boldsymbol{x} \geq \boldsymbol{0}$, while Case (ii) shows results without the constraint

In the experiment, we solved the problem with $n = 25$, changing nine values of d.[35] Fig. 14 shows the ratio and the solution accuracy of the six LP-representable norms in boxplots. If short sales are not allowed, i.e., the nonnegativity $\boldsymbol{x} \geq \boldsymbol{0}$ is imposed, the MAD model well approximates the Markowitz's portfolio in terms of $\|\hat{\boldsymbol{y}}\|_2/\phi^\star$, shown in Fig. (b) despite the poor proximity shown in Fig. (a). However, if short sales are allowed, i.e., the constraint is relaxed, the solution to the MAD model is furthest away from the Markowitz's portfolio. On the other hand, the other norms, especially the dual deltoidal norm, achieved better approximating solutions.

5.3 A Robust LP Problem

Finally, let us consider an SOCP of the form (28), which is viewed as a robust optimization problem:[36]

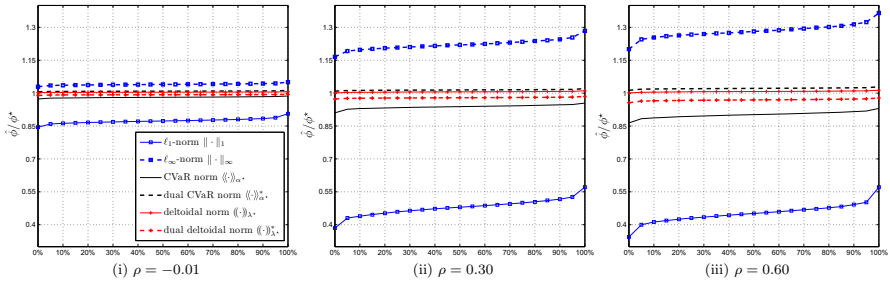
$$\phi^\star := \left| \begin{array}{l} \underset{\boldsymbol{x}}{\text{maximize}}\ \boldsymbol{1}_n^\top \boldsymbol{x} \\ \text{subject to}\ \boldsymbol{c}^\top \boldsymbol{x} + \delta \|\boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{x}\|_2 \leq C,\ \boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{1}_n, \end{array} \right. \tag{37}$$

[35] The covariance matrix $\boldsymbol{\Sigma}$ and the expected return vector $\boldsymbol{r}$ were estimated based on the "25 Portfolios Formed on Size and Book-to-Market Daily" data downloaded from [9]. The matrix $\boldsymbol{M}$ was obtained by the eigenvalue decomposition of the estimated covariance matrix $\boldsymbol{\Sigma}$. We solved nine instances, each corresponding to $d = \underline{d} + (\overline{d} - \underline{d})i/10$, $i = 1, \ldots, 9$, where $\underline{d} := \min\{r_1, \ldots, r_{25}\}$ and $\overline{d} := \max\{r_1, \ldots, r_{25}\}$.

[36] Formulation (37) is motivated by a robust optimization of an uncertain continuous knapsack problem:

$$\left| \begin{array}{l} \underset{\boldsymbol{x}}{\text{maximize}}\ \boldsymbol{1}_n^\top \boldsymbol{x} \\ \text{subject to}\ \tilde{\boldsymbol{c}}^\top \boldsymbol{x} \leq C,\ \boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{1}_n, \end{array} \right.$$

where C represents a capacity of a knapsack and $\tilde{\boldsymbol{c}}$ represents volumes of n (liquid) items. In this formulation, $\tilde{\boldsymbol{c}}$ is assumed to be under uncertainty, $\mathcal{U}$, of the form: $\tilde{\boldsymbol{c}} \in \mathcal{U} := \{\boldsymbol{u} \in \mathbb{R}^n : \sqrt{(\boldsymbol{u} - \boldsymbol{c})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{u} - \boldsymbol{c})} \leq \delta\}$, where $\boldsymbol{\Sigma}^{-1}$ is a positive definite matrix determining the shape of $\mathcal{U}$ and δ defines the size of the set $\mathcal{U}$.



(i) $\rho = -0.01$ (ii) $\rho = 0.30$ (iii) $\rho = 0.60$

Fig. 15 Percentiles of ratio $\hat{\phi}/\phi^\star$ for the problem (37) (x-axis: %; y-axis: $\hat{\phi}/\phi^\star$). The percentiles of the ratio between the approximating objective value, $\hat{\phi}$, and the true one, $\phi^\star$, among the 1000 solutions to the randomly generated instances of (37). We see that the difference among the methods becomes larger as the underlying correlation coefficient ρ departs from 0. Especially, the dual CVaR norm and the deltoidal norm show nice mimicking performance stably both in percentage terms and in ρ

where $\boldsymbol{c} \in \mathbb{R}^n$, $C > 0$, $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix, and δ is a positive constant.[37] What we would like to compare is

$$\hat{\phi} := \left| \begin{array}{l} \underset{\boldsymbol{x}}{\text{maximize}}\ \mathbf{1}_n^\top \boldsymbol{x} \\ \text{subject to } \boldsymbol{c}^\top \boldsymbol{x} + \delta \|\boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{x}\| \le C,\ \mathbf{0} \le \boldsymbol{x} \le \mathbf{1}_n, \end{array} \right. \tag{38}$$

where $\|\cdot\|$ is $\|\cdot\|_1$, $\|\cdot\|_\infty$, $\langle\langle\cdot\rangle\rangle_{\alpha^\star}$, $\langle\langle\cdot\rangle\rangle^*_{\alpha^\star}$, $(\!(\cdot)\!)_{\lambda^\star}$, or $(\!(\cdot)\!)^*_{\lambda^\star}$.

Fig. 15 compares the ratio of the optimal value of the approximating formulations (38) and that of the original formulation (37). As the correlation, ρ, increases, the ℓ_1- or ℓ_∞-norm approximation differs from the original version. The ℓ_1-solution is far below the value of 1, showing an excessive conservativeness relative to the Euclidean case, as pointed out in the literature. The CVaR norm also showed a conservative solution. In contrast, the use of the dual CVaR norm and the deltoidal norm well mimicked the Euclidean case.

6 Concluding Remarks

This paper studied two pairs of families of polyhedral norms. Each pair has a conservativeness parameter and includes the ℓ_1- and ℓ_∞-norms as special cases. As shown, these norms are advantageous over the ℓ_p-norm in their computer-friendly evaluations and interpretability with respect to the involved parameters. With some specific parameters, these families define some known polytopes.

[37] We randomly generated 1000 instances of (37) of size $n = 100$ each, as follows: Generating $c_i \sim \mathrm{U}(0, 1)$, $i = 1, \ldots, n$, we set $\sigma_i = 0.1 c_i$ and $C = \lfloor \boldsymbol{c}^\top \mathbf{1}_n / 2 \rfloor$. (This indicates that the nominal volume, c_i, of each item ranges over (0, 1) and its standard deviation is 1/10 of the nominal volume, and the capacity is about a half of the aggregate total, $\boldsymbol{c}^\top \mathbf{1}_n$. We also set the covariance matrix as $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma})\{(1-\rho)\boldsymbol{I}_n + \rho \mathbf{1}_n \mathbf{1}_n^\top\}\mathrm{diag}(\boldsymbol{\sigma})$ so that the each pair of two assets has correlation coefficient of ρ. (Note that $\boldsymbol{\Sigma}$ is positive definite as long as $\rho > -1/(n-1)$.) Also, we set $\delta = 10.8857$. To solve the norm constrained problems, we used CVX, a package for specifying and solving convex programs [13,14].

Regarding the relation to the ℓ_p-norm, the answers to the questions posed in the introduction are listed below:

Q.1 *In which sense do the norms give the best approximation?*

A. In order to quantify the proximity of two norms, we utilize the ratio of the tight upper and lower bounds of the ratio between the two norms. The bounds are symmetric between the primal and dual pair of norms, and the analysis after the derivation can be done in one shot. Also, both of the ratios for the two pairs of families of norms analyzed in this paper are quasiconvex functions with respect to the involved parameters, and therefore, we can uniquely attain the optimal parameters.

Q.1.(a) *With what value of α do the CVaR norm and its dual give the best approximation?*

A. The best value of α for the CVaR norm approximating the ℓ_p-norm is given by $\alpha^\star = 1 - n^{-1+\frac{1}{p}}$, whereas that for the dual norm is given by $\alpha^\star = 1 - n^{-\frac{1}{p}}$. Accordingly, in the sense of the largest-k norm, the best k approximating the ℓ_p-norm is $n^{\frac{1}{p}}$, whereas for the dual norm it is $n^{1-\frac{1}{p}}$.

Q.1.(b) *With what value of λ do the deltoidal norm and its dual give the best approximation?*

A. The best values of λ for the deltoidal norm and its dual for approximating the ℓ_p- and $\ell_{\frac{p}{p-1}}$-norms, respectively, are similarly given in a symmetric manner (as in the case of the pair of the CVaR norms). When $p = 2$ and n is a square number, the best parameter for the deltoidal norm is approximately given by $\frac{n}{n-1}\alpha^\star$.

Q.2 *Which norms better approximate the ℓ_p-norm, the CVaR norm or the deltoidal norm?*

A. Comparing the two pairs of the norms, the deltoidal pair approximates the ℓ_p-norm in a better way.

Needless to say, there are infinitely many LP-representable norms, and we expect developments of more appealing parametrized families. However, the four norms studied in this paper look so simple that they can provide a benchmark for future studies.

On the other hand, non-convex functions formally-defined as $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p$ with $p \in [0, 1)$ have recently gathered attention since they are related to the number of non-zero elements of a vector $\boldsymbol{x}$ (see, e.g., [10]). A nonconvex extension of the CVaR norms to such a direction can be found in a recent report [19].

Acknowledgments This research was done during the visit of the first author of the paper to the University of Florida, USA. The first author appreciates the financial support by Chuo university and the environmental support by the Department of Industrial and Systems Engineering of the University of Florida. The research of the first author is supported in part by a MEXT Grant-in-Aid for Young Scientists (B) 23710176. Research of the second author of the paper was partially supported by the AFOSR grant FA9550-11-1-0258, "New Developments in Uncertainty: Linking Risk Management, Reliability, Statistics and Stochastic Optimization." The authors also thank Matt Norton for his help with proofreading.

Appendix: Proof of Propositions

Proof of Proposition 1

To show the lower bound of (6), consider the following maximization problem:

$$\left|\begin{array}{l}\text{maximize } \|\boldsymbol{x}\|_p^p \equiv |x_1|^p + \cdots + |x_n|^p \\ \text{subject to } \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha = 1.\end{array}\right.$$

By symmetry, the optimal value is equal to

$$\left|\begin{array}{l}\text{maximize } x_1^p + \cdots + x_n^p \\ \text{subject to } x_1 + \cdots + x_{\lfloor\kappa\rfloor} + (\kappa - \lfloor\kappa\rfloor)x_{\lfloor\kappa\rfloor+1} = 1, \\ \qquad\qquad x_j \geq x_{\lfloor\kappa\rfloor+1},\ j=1,\ldots,\lfloor\kappa\rfloor;\ \ x_{\lfloor\kappa\rfloor+1} \geq 0;\ \ x_{\lfloor\kappa\rfloor+1} = x_j,\ j=\lfloor\kappa\rfloor+2,\ldots,n.\end{array}\right.$$

Since this is a convex maximization over a polytope, an optimal solution is attained at some vertex of the polytope. The constraints consist of one equality and $\lfloor\kappa\rfloor + 1$ inequalities while the dimension is $\lfloor\kappa\rfloor + 1$, and therefore, at most one inequality can be unbinding. Besides, due to the first (equality) constraint, it is impossible that all inequalities hold with equalities.

[Case 1: $x_{\lfloor\kappa\rfloor+1} > 0$] The corresponding vertex satisfies $x_1 = \cdots = x_{\lfloor\kappa\rfloor+1} = \frac{1}{\kappa}$, and its objective value is $\frac{n}{\kappa^p}$.

[Case 2: $x_j = x_{\lfloor\kappa\rfloor+1} = 0$ for $j \neq \hat{j}$ and $x_{\hat{j}} > 0$ for some $\hat{j} \in \{1,\ldots,\lfloor\kappa\rfloor\}$] In this case, $x_{\hat{j}} = 1$ and the vertices attain the objective value of 1.

Therefore, we have $\|\boldsymbol{x}\|_p / \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha \leq \max\{1, n^{\frac{1}{p}}\kappa^{-1}\}$, and $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha / \|\boldsymbol{x}\|_p \leq \min\{1, \kappa n^{-\frac{1}{p}}\}$.

To show the upper bound of (6), consider the following minimization problem:

$$\left|\begin{array}{l}\text{minimize } \|\boldsymbol{x}\|_p^p := |x_1|^p + \cdots + |x_n|^p \\ \text{subject to } \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha = 1.\end{array}\right.$$

By using the symmetry, the optimal value is equal to

$$\left|\begin{array}{l}\text{minimize } x_1^p + \cdots + x_n^p \\ \text{subject to } x_1 + \cdots + x_{\lfloor\kappa\rfloor} + (\kappa - \lfloor\kappa\rfloor)x_{\lfloor\kappa\rfloor+1} = 1, \\ \qquad\qquad x_j \geq x_{\lfloor\kappa\rfloor+1},\ j=1,\ldots,\lfloor\kappa\rfloor;\ \ x_{\lfloor\kappa\rfloor+1} \geq 0;\ \ x_{\lfloor\kappa\rfloor+2} = \cdots = x_n = 0,\end{array}\right.$$

and we can consider the reduced problem having $\lfloor\kappa\rfloor + 1$ variables. A solution with $x_j = \{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^{\frac{p}{p-1}}\}^{-1}$, $j = 1,\ldots,\lfloor\kappa\rfloor$, and $x_{\lfloor\kappa\rfloor+1} = (\kappa - \lfloor\kappa\rfloor)^{\frac{1}{p-1}}\{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^{\frac{p}{p-1}}\}^{-1}$ (and $x_j = 0$, $j = \lfloor\kappa\rfloor + 2,\ldots,n$) satisfies the Karush-Kuhn-Tucker (KKT) condition. Since the problem is a convex minimization, this solution is optimal and the optimal value is $\{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^{\frac{p}{p-1}}\}^{1-p}$. Consequently, $\|\boldsymbol{x}\|_p / \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha \geq \{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^{\frac{p}{p-1}}\}^{-\frac{p-1}{p}}$ and $\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha / \|\boldsymbol{x}\|_p \leq \{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^{\frac{p}{p-1}}\}^{\frac{p-1}{p}}$. □

Proof of Proposition 2

Similarly to the proof of Proposition 1, to prove the upper bound of (7), consider the following maximization problem:

$$\left|\begin{array}{l}\text{maximize } \|\boldsymbol{x}\|_p^p \equiv |x_1|^p + \cdots + |x_n|^p \\ \text{subject to } \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* \equiv \max\{\|\boldsymbol{x}\|_1/n(1-\alpha), \|\boldsymbol{x}\|_\infty\} \le 1.\end{array}\right. \tag{39}$$

Note that due to the symmetry in the n elements of $\boldsymbol{x}$, the optimal value of (39) can be obtained by the following convex maximization program:

$$\left|\begin{array}{l}\text{maximize } x_1^p + \cdots + x_n^p \\ \text{subject to } x_1 + \cdots + x_n = \kappa,\ 1 = x_1 \ge x_2 \ge \cdots \ge x_n \ge 0,\end{array}\right.$$

where $\kappa = n(1-\alpha)$. Obviously, an optimal solution can be obtained by setting $x_1 = \cdots = x_{\lfloor\kappa\rfloor} = 1$, $x_{\lfloor\kappa\rfloor+1} = \kappa - \lfloor\kappa\rfloor$, and $x_{\lfloor\kappa\rfloor+2} = \cdots = x_n = 0$, and its optimal value is $\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^p$. Consequently, $\|\boldsymbol{x}\|_p/\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* \le \{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^p\}^{\frac{1}{p}}$.

Next, we prove the lower bound of (7). Consider

$$\left|\begin{array}{l}\text{minimize } \|\boldsymbol{x}\|_p^p \\ \text{subject to } \langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* \equiv \max\{\|\boldsymbol{x}\|_1/n(1-\alpha), \|\boldsymbol{x}\|_\infty\} = 1,\end{array}\right. \tag{40}$$

which is equivalent to the minimum of the following two convex optimization problems:

$$\left|\begin{array}{l}\text{minimize } x_1^p + \cdots + x_n^p \\ \text{subject to } x_1 + \cdots + x_n = \kappa,\ 1 \ge x_1 \ge x_2 \ge \cdots \ge x_n \ge 0;\end{array}\right. \tag{41}$$

$$\left|\begin{array}{l}\text{minimize } x_1^p + \cdots + x_n^p \\ \text{subject to } x_1 + \cdots + x_n \le \kappa,\ 1 = x_1 \ge x_2 \ge \cdots \ge x_n \ge 0.\end{array}\right. \tag{42}$$

It is easy to see that the solution $x_1 = \cdots = x_n = \kappa/n$ is a KKT solution to (41) and the optimal value is $n(\kappa/n)^p$. On the other hand, $x_1 = 1, x_2 = \cdots = x_n = 0$ is a KKT solution to (42) and the optimal value is 1. Consequently, the optimal value of (40) is given by $\min\{\kappa n^{-1+\frac{1}{p}}, 1\}$, and we have $\|\boldsymbol{x}\|_p/\langle\!\langle \boldsymbol{x} \rangle\!\rangle_\alpha^* \ge \min\{n^{\frac{1}{p}}(1-\alpha), 1\}$. □

Proof of Lemma 1

We here prove the continuity of the function at $\kappa \in \mathbb{Z}$ since the continuity and differentiability at $\kappa \notin \mathbb{Z}$ is evident.

[Case: $\kappa \ge n^{\frac{1}{p}}$] We have $f_{n,p}(\kappa) = \{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^{\frac{p}{p-1}}\}^{\frac{p-1}{p}}$ and $\lim_{\varepsilon\to 0+} f_{n,p}(\kappa - \varepsilon) = \lim_{\varepsilon\to 0+} [\kappa - 1 + \{\kappa - \varepsilon - (\kappa - 1)\}^{\frac{p}{p-1}}]^{\frac{p-1}{p}} = \kappa^{\frac{p-1}{p}}$. On the other hand, $\lim_{\varepsilon\to 0+} f_{n,p}(\kappa + \varepsilon) = \lim_{\varepsilon\to 0+} \{\kappa + (\kappa - \varepsilon - \kappa)^{\frac{p}{p-1}}\}^{\frac{p-1}{p}} = \kappa^{\frac{p-1}{p}}$.

[Case: $\kappa \leq n^{\frac{1}{p}}$] We have $f_{n,p}(\kappa) = \frac{n^{\frac{1}{p}}}{\kappa}\{\lfloor\kappa\rfloor + (\kappa - \lfloor\kappa\rfloor)^{\frac{p}{p-1}}\}^{\frac{p-1}{p}}$ and $\lim_{\varepsilon\to 0+} f_{n,p}(\kappa - \varepsilon) = \lim_{\varepsilon\to 0+} \frac{n^{\frac{1}{p}}[\kappa-1+\{\kappa-\varepsilon-(\kappa-1)\}^{\frac{p}{p-1}}]^{\frac{p-1}{p}}}{\kappa-\varepsilon} = \left(\frac{n}{\kappa}\right)^{\frac{1}{p}}$. On the other hand, $\lim_{\varepsilon\to 0+} f_{n,p}(\kappa + \varepsilon) = \lim_{\varepsilon\to 0+} \frac{n^{\frac{1}{p}}\{\kappa+(\kappa-\varepsilon-\kappa)^{\frac{p}{p-1}}\}^{\frac{p-1}{p}}}{\kappa+\varepsilon} = \left(\frac{n}{\kappa}\right)^{\frac{1}{p}}$.

Proof of Proposition 3

[Case: $\kappa \geq n^{\frac{1}{p}}$] For any $\kappa \notin \mathbb{Z}$, it is differentiable and $f'_{n,p}(\kappa) = \frac{(\kappa-C)^{\frac{1}{p-1}}}{\{C+(\kappa-C)^{\frac{p}{p-1}}\}^{\frac{1}{p}}}$. Since $f_{n,p}(\kappa)$ is continuous at any $\kappa \in \mathbb{Z}$, it is increasing for $\kappa \geq n^{\frac{1}{p}}$.

[Case: $\kappa \leq n^{\frac{1}{p}}$] For any $\kappa \notin \mathbb{Z}$, it is differentiable and $f'_{n,p}(\kappa) = \frac{n^{\frac{1}{p}}C\{(\kappa-C)^{\frac{p}{p-1}}-(\kappa-C)\}}{\kappa^2(\kappa-C)\{(\kappa-C)^{\frac{p}{p-1}}+C\}^{\frac{1}{p}}} < 0$. Since $f_{n,p}(\kappa)$ is continuous at any $\kappa \in \mathbb{Z}$, it is decreasing for $\kappa \leq n^{\frac{1}{p}}$. □

Proof of Proposition 4

Note that the dual norm to $(\!(\boldsymbol{x})\!)_\lambda$ is written by an LP:

$$(\!(\boldsymbol{x})\!)^*_\lambda \equiv \left| \begin{array}{l} \underset{\boldsymbol{y}}{\text{maximize}}\ |\boldsymbol{x}|^\top \boldsymbol{y} \\ \text{subject to } \{\lambda \boldsymbol{I}_n + (1-\lambda)\mathbf{1}_n\mathbf{1}_n^\top\}\boldsymbol{y} \leq \mathbf{1}_n,\ \ \boldsymbol{y} \geq \mathbf{0}, \end{array} \right. \tag{43}$$

where $|\boldsymbol{x}| := (|x_1|, \ldots, |x_n|)^\top$. It is easy to see that any extreme point of the feasible region of the above LP is described as a solution to a system of n equalities of the form

$$\begin{cases} \{\lambda \boldsymbol{I}_{|S|} + (1-\lambda)\mathbf{1}_{|S|}\mathbf{1}_{|S|}^\top\}\boldsymbol{y}_S = \mathbf{1}_{|S|}, \\ y_i = 0,\ i \notin S, \end{cases}$$

where $\boldsymbol{y}_S$ denotes the vector $(y_i)_{i\in S}$ with $S \subset \{1, \ldots, n\}$. With the Sherman-Morrison-Woodbury formula, the explicit expression of the solution is given by $\boldsymbol{y}_S = \frac{1}{|S|-(|S|-1)\lambda}\mathbf{1}_{|S|}$ and $y_i = 0, i \notin S$, and its objective value is $\frac{1}{|S|-(|S|-1)\lambda}\sum_{i\in S}|x_i|$. Note that this solution is feasible to (43). It is easy to see that given $|S|$, the maximum objective value is $\frac{1}{|S|-(|S|-1)\lambda}\sum_{i=1}^{|S|}|x_{(i)}|$. Consequently, the optimal value is given by the formula which is to be proved. □

Proof of Proposition 5

In order to obtain the lower bound for the deltoidal norm, we consider the following optimization problem.

$$\left|\begin{array}{l}\text{maximize } \|\boldsymbol{x}\|_p^p \\ \text{subject to } (\!(\boldsymbol{x})\!)_\lambda \le 1.\end{array}\right.$$

Taking into account the symmetry of n variables, it is sufficient to solve the following optimization problem:

$$\left|\begin{array}{l}\text{maximize } x_1^p + \cdots + x_n^p \\ \text{subject to } (1-\lambda)(x_1 + \cdots + x_n) + \lambda x_1 = 1, \quad x_1 \ge x_2 \ge \cdots \ge x_n \ge 0.\end{array}\right. \tag{44}$$

Since this is a convex maximization over a polytope, we can find an optimal solution at an extreme point of the feasible region. Noting that the feasible region of (44) consists of the one equality and n inequalities, no less than n inequalities hold with equality at any extreme point. It is not hard to see that the extreme solution is defined by $x_1 = \cdots = x_k = \frac{1}{(1-\lambda)k+\lambda} > 0$, $x_{k+1} = \cdots = x_n = 0$, for some $k \in \{1, 2, \ldots, n\}$, and the objective value is $f_p(k) := \frac{k}{\{(1-\lambda)k+\lambda\}^p}$. The derivative of f_p is $f_p'(k) = \frac{\{(1-\lambda)k+\lambda\}^{p-1}\{(1-p)(1-\lambda)k+\lambda\}}{\{(1-\lambda)k+\lambda\}^{2p}}$, which is increasing for $k < \frac{\lambda}{(p-1)(1-\lambda)} =: \hat{k}$, whereas decreasing for $k > \hat{k}$.

[Case: $\lambda \in [0, 1-\frac{1}{p})$] Note that this condition is equivalent to $\lfloor \hat{k} \rfloor + 1 \le 1$ (or $\hat{k} < 1$). In this case, the maximum of f_p over $k \in \{1, \ldots, n\}$ is attained at $k = 1$ and its value is 1.

[Case: $\lambda \in [1-\frac{1}{p}, \frac{(p-1)n}{(p-1)n+1})$] This condition is equivalent to $1 < \lfloor \hat{k} \rfloor + 1 \le n$ (or $1 \le \hat{k} < n$).

The maximum is attained at $k = \lfloor \hat{k} \rfloor$ or $k = \lfloor \hat{k} \rfloor + 1$, and the objective value is

$$\max\left\{\frac{\lfloor \hat{k} \rfloor}{\{(1-\lambda)\lfloor \hat{k} \rfloor + \lambda\}^p}, \frac{\lfloor \hat{k} \rfloor + 1}{\{(1-\lambda)(\lfloor \hat{k} \rfloor + 1) + \lambda\}^p}\right\}.$$

[Case: $\lambda \in [\frac{(p-1)n}{(p-1)n+1}, 1]$] This condition is equivalent to $n \le \lfloor \hat{k} \rfloor$ (or $\hat{k} \ge n$). The maximum is attained at $k = n$ and its objective is $\frac{n}{\{(1-\lambda)n+\lambda\}^p}$.

By taking the p-th root of the result above and taking the reciprocal, we obtain the lower bound, $g_{n,p}(\lambda)$, of $(\!(\boldsymbol{x})\!)_\lambda / \|\boldsymbol{x}\|_p$.

In order to obtain the upper bound, we consider the following optimization problem:

$$\left|\begin{array}{l}\text{minimize } \|\boldsymbol{x}\|_p^p \\ \text{subject to } (\!(\boldsymbol{x})\!)_\lambda = 1.\end{array}\right.$$

Taking into account the symmetry of n variables, it is sufficient to solve the following optimization problem:

$$\left|\begin{array}{l}\text{minimize } x_1^p + \cdots + x_n^p \\ \text{subject to } (1-\lambda)(x_1 + \cdots + x_n) + \lambda x_1 = 1, \; x_1 \ge x_2 \ge \cdots \ge x_n \ge 0.\end{array}\right. \tag{45}$$

It is easy to see that the solution $x_1 = \frac{1}{(n-1)(1-\lambda)^2+1}$, $x_i = \frac{1-\lambda}{(n-1)(1-\lambda)^2+1}$, $i = 2, \ldots, n$, satisfies the KKT condition. Since (45) is a convex program, this is an optimal solution, and the objective value is $\frac{(n-1)(1-\lambda)^p+1}{\{(n-1)(1-\lambda)^2+1\}^p}$. Therefore, we get $\|\boldsymbol{x}\|_p/(\!(\boldsymbol{x})\!)_\lambda \geq \{(n-1)(1-\lambda)^p+1\}^{\frac{1}{p}}/\{(n-1)(1-\lambda)^2+1\}$. By taking the reciprocal to the result above, we obtain the upper bound of $(\!(\boldsymbol{x})\!)_\lambda/\|\boldsymbol{x}\|_p$.

Next, we prove the bounds of the dual norm. Similarly to the proof of the dual CVaR norm case, we consider the maximization of $\|\boldsymbol{x}\|_p$ over a constraint $(\!(\boldsymbol{x})\!)^* \leq 1$ in order to show the upper bound.

$$\left|\begin{array}{l}\text{maximize } \|\boldsymbol{x}\|_p^p \\ \text{subject to } (\!(\boldsymbol{x})\!)_\lambda^* \leq 1.\end{array}\right.$$

Based on the explicit formula (17) of $(\!(\boldsymbol{x})\!)_\lambda^*$, this is equivalent to

$$\left|\begin{array}{l}\text{maximize } x_1^p + \cdots + x_n^p \\ \text{subject to } \max\left\{x_1, \frac{x_1+x_2}{2-\lambda}, \ldots, \frac{x_1+\cdots+x_n}{n-(n-1)\lambda}\right\} = 1, \quad x_1 \geq x_2 \geq \cdots \geq x_n \geq 0,\end{array}\right.$$

which is reduced to a convex maximization over a polytope

$$\left|\begin{array}{l}\text{maximize } x_1^p + \cdots + x_n^p \\ \text{subject to } x_1 + \cdots + x_j \leq j - (j-1)\lambda, \ j = 1, \ldots, n; \quad x_1 \geq x_2 \geq \cdots \geq x_n \geq 0.\end{array}\right.$$

Let us observe that any optimal solution satisfies the first n inequality constraints with equalities. (Suppose on the contrary that there exists the smallest k such that $x_1 + \cdots + x_k < k - (k-1)\lambda$ at an optimal solution $\boldsymbol{x}$. Noting that $x_1 = 1, x_2 = \cdots = x_{k-1} = 1-\lambda$ holds, this assumption implies that $x_k < 1-\lambda$ and the objective value can increase by setting $x_k = 1-\lambda$.) Therefore, the solution of the system of the n equalities, $x_1 = 1, x_2 = \cdots = x_n = 1-\lambda$, is the unique solution, and the optimal value is then given by $1 + (n-1)(1-\lambda)^p$. Consequently, $\|\boldsymbol{x}\|_p/(\!(\boldsymbol{x})\!) \leq \{1 + (n-1)(1-\lambda)^p\}^{\frac{1}{p}}$.

In order to obtain the lower bound for the dual deltoidal norm, we consider the following optimization problem.

$$\left|\begin{array}{l}\text{minimize } \|\boldsymbol{x}\|_p^p \\ \text{subject to } (\!(\boldsymbol{x})\!)_\lambda^* = 1.\end{array}\right. \tag{46}$$

Based on the explicit formula (17) of $(\!(\boldsymbol{x})\!)_\lambda^*$, this is equivalent to

$$\left|\begin{array}{l}\text{minimize } x_1^p + \cdots + x_n^p \\ \text{subject to } \max\left\{x_1, \frac{x_1+x_2}{2-\lambda}, \ldots, \frac{x_1+\cdots+x_n}{n-(n-1)\lambda}\right\} = 1, \quad x_1 \geq x_2 \geq \cdots \geq x_n \geq 0,\end{array}\right.$$

whose optimal value is equal to $\min\{F_p(1), \ldots, F_p(n)\}$ where $F_p(k)$ can be computed by the following LP:

$$F_p(k) := \left| \begin{array}{l} \text{minimize } x_1^p + \cdots + x_n^p \\ \text{subject to } x_1 + \cdots + x_k = k - (k-1)\lambda,\ x_1 \geq x_2 \geq \cdots \geq x_n \geq 0, \\ \qquad\qquad x_1 + \cdots + x_{k'} \leq k' - (k'-1)\lambda,\ \text{for } k' \in \{1, \ldots, n\} \backslash \{k\}. \end{array} \right. \tag{47}$$

In place of (47), consider the following (relaxed) LP for some $k \in \{1, \ldots, n\}$:

$$\hat{F}(k) := \left| \begin{array}{l} \text{minimize } x_1^p + \cdots + x_n^p \\ \text{subject to } x_1 + \cdots + x_k = k - (k-1)\lambda,\ x_1 \geq x_2 \geq \cdots \geq x_n \geq 0. \end{array} \right. \tag{48}$$

It is easy to observe that the solution such that $x_1 = \cdots = x_k = \frac{k-(k-1)\lambda}{k}$ and $x_{k+1} = \cdots = x_n = 0$ satisfies the KKT condition. Since (48) is a convex program, such a KKT solution is optimal to that. Also, note that this KKT solution satisfies the inequalities $x_1 + \cdots + x_{k'} \leq k' - (k'-1)\lambda$ for any $k' \neq k$, which implies that this KKT solution is also optimal to (47), i.e., $F(k) = \hat{F}(k)$. Therefore, the optimal value of (46) can be obtained by $\min\{\hat{F}(1), \ldots, \hat{F}(n)\}$, where $\hat{F}_p(k) = \frac{\{(1-\lambda)k+\lambda\}^p}{k^{p-1}}$. Since the first order derivative of $\hat{F}_p(k)$ is given by $\hat{F}_p'(k) = \frac{\{(k-1)(1-\lambda)+1\}^{p-1}}{k^p} \times \{(1-\lambda)k + \lambda(1-p)\}$, we see that $\hat{F}_p(k)$ is increasing (decreasing) if $k > \frac{\lambda(p-1)}{1-\lambda}$ ($k < \frac{\lambda(p-1)}{1-\lambda}$) for $\lambda \in (0, 1)$ and $p > 1$, and therefore, attained its minimum at $\tilde{k} := \frac{\lambda(p-1)}{1-\lambda}$.

[Case: $\lambda \in [0, \frac{1}{p})$] Note that this condition is equivalent to $\lfloor \tilde{k} \rfloor + 1 \leq 1$ (or $\tilde{k} < 1$). The minimum of F_p over $k \in \{1, \ldots, n\}$ is attained at $k = 1$ and its value is 1.

[Case: $\lambda \in [\frac{1}{p}, \frac{n}{n+p-1})$] This condition is equivalent to $1 < \lfloor \tilde{k} \rfloor + 1 \leq n$ (or $1 \leq \tilde{k} < n$). The maximum is attained at $k = \lfloor \tilde{k} \rfloor$ or $k = \lfloor \tilde{k} \rfloor + 1$, and the objective value is then

$$\min\{F(\lfloor \tilde{k} \rfloor), F(\lfloor \tilde{k} \rfloor + 1)\} = \min\left\{ \frac{\{(1-\lambda)\lfloor \tilde{k} \rfloor + \lambda\}^p}{\lfloor \tilde{k} \rfloor^{p-1}}, \frac{\{(1-\lambda)(\lfloor \tilde{k} \rfloor + 1) + \lambda\}^p}{(\lfloor \tilde{k} \rfloor + 1)^{p-1}} \right\}.$$

[Case: $\lambda \in [\frac{n}{n+p-1}, 1]$] This condition is equivalent to $n \leq \lfloor \tilde{k} \rfloor$ (or $\tilde{k} \geq n$). The maximum is attained at $k = n$ and its objective is $\frac{\{(1-\lambda)n+\lambda\}^p}{n^{p-1}}$.

By taking the p-th root, we finalize the proof of the lower bound.

Proof of Lemma 2

Note that on the interval $(\hat{\lambda}_{k-1}, \hat{\lambda}_k)$, the function $h_{n,p}(\lambda)$ is explicitly written by

$$h_{n,p,k}(\lambda) := \frac{k^{\frac{1}{p}} \left\{ (n-1)(1-\lambda)^{\frac{p}{p-1}} + 1 \right\}}{(k-1)(1-\lambda) + 1},$$

and its first derivative is

$$h'_{n,p,k}(\lambda) = \frac{k^{\frac{1}{p}}\left\{k-1-(n-1)(1-\lambda)^{\frac{1}{p-1}}\right\}}{\left\{(n-1)(1-\lambda)^{\frac{p}{p-1}}+1\right\}^{\frac{1}{p}}\{(k-1)(1-\lambda)+1\}^2}.$$

Noting that the sign of $h'_{n,p,k}(\lambda)$ depends only on $k-1-(n-1)(1-\lambda)^{\frac{1}{p-1}}$, we see that for each $k \in \{1,\ldots,n-1\}$, $h_{n,p,k}(\lambda)$ uniquely attains the minimum at $1-\left(\frac{k-1}{n-1}\right)^{p-1}$ over $[0,1]$. Besides, we have

$$\lim_{\lambda\to\hat{\lambda}_k+0} h'_{n,p}(\lambda) - \lim_{\lambda\to\hat{\lambda}_k-0} h'_{n,p}(\lambda) = \lim_{\lambda\to\hat{\lambda}_k} h'_{n,p,k+1}(\lambda) - h'_{n,p,k}(\lambda)$$

$$= \frac{\left\{k\cdot k^{\frac{1}{p}}-(k-1)(k+1)^{\frac{1}{p}}\right\}^3\left[\left\{\frac{(k+1)^{\frac{1}{p}}-k^{\frac{1}{p}}}{k\cdot k^{\frac{1}{p}}-(k-1)(k+1)^{\frac{1}{p}}}\right\}^{\frac{p}{p-1}}(n-1)+1\right]^{\frac{p-1}{p}}}{k^{\frac{1}{p}}(k+1)^{\frac{1}{p}}} > 0.$$

From these facts, we see that $h_{n,p}(\lambda)$ is quasiconvex and has the unique minimum. □

References

1. Arthanari, T.S., Dodge, Y.: Mathematical Programming in Statistics. Wiley, New York (1981)
2. Aulbach, S., Falk, M., Hofmann, M.: On max-stable processes and the functional D-norm. Extremes **16**, 255–283 (2013)
3. Ben-Tal, A., Nemirovski, A.: On polyhedral approximations of the second-order cone. Math. Oper. Res. **26**, 193–205 (2001)
4. Bertsimas, D., Pachamanova, D., Sim, M.: Robust linear optimization under general norms. Oper. Res. Lett. **32**, 510–516 (2004)
5. Bertsimas, D., Sim, M.: The price of robustness. Oper. Res. **52**, 35–53 (2004)
6. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
7. Candés, E.J., Wakin, M.B.: An introduction to compressive sampling. IEEE Signal Proc. Mag. **25**, 21–30 (2008)
8. Dattorro, J.: Convex Optimization & Euclidean Distance Geometry. Meboo Publishing, USA (2005)
9. French, K.: Data Library, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html Accessed 14 May 2013
10. Ge, D., Jiang, X., Ye, Y.: A note on the complexity of Lp minimization. Math. Prog. **129**, 285–299 (2011)
11. Gotoh, J., Uryasev, S.: Two pairs of families of polyhedral norms versus ℓ_p-Norms: proximity and applications in optimization, research report 2013–3, Department of Industrial Systems and Engineering, University of Florida, Gainesville, FL, http://www.ise.ufl.edu/uryasev/files/2013/10/TwoPairsOfPolyhedralNormsVersusLpNorms_20131025_UFISEPD.pdf (2013)
12. Gotoh, J., Uryasev, S.: Support vector machines based on convex risk functionals and general norms, research report 2013–6, Department of Industrial Systems and Engineering, University of Florida, Gainesville, FL
13. Grant, M., Boyd, S.: CVX: MATLAB software for disciplined convex programming, version 2.0 beta, http://cvxr.com/cvx (2012)
14. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs, in Recent Advances in Learning and Control, V. Blondel, S. Boyd and H. Kimura, eds., Springer. 2008, pp. 95–110. http://stanford.edu/boyd/graph_dcp.html

15. Hazell, P.B.R.: A linear alternative to quadratic and semivariance programming for farm planning under uncertainty. Am. J. Agr. Econ. **53**, 53–62 (1971)
16. Kloft, M., Brefeld, U., Sonnenburg, S., Zien, A.: ℓ_p-norm multiple kernel learning. J. Mach. Learn. Res. **12**, 953–997 (2011)
17. Konno, H., Yamazaki, H.: Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market. Manage. Sci. **37**, 519–531 (1991)
18. Krokhmal, P.: Higher moment risk measures. Quant. Financ. **7**, 373–387 (2007)
19. Mafusalov, A., Uryasev, S.: Conditional value-at-risk (CVaR) norm: stochastic case, research report 2013–5, Department of Industrial Systems and Engineering, University of Florida, Gainesville, FL, http://www.ise.ufl.edu/uryasev/files/2014/01/CVaR_norm_stochastic_case.pdf (2013)
20. Pavlikov, K., Uryasev, S.: CVaR norm and applications in optimization. Optim. Lett. **8**, 1–22 (2014)
21. Portfolio Safeguard: http://www.aorda.com/aod/psg.action
22. Rockafellar, T.R., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–41 (2000)
23. Rockafellar, T.R., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Financ. **26**, 1443–1471 (2002)
24. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. B Met. **58**, 267–288 (1996)
25. Vinel, A., Krokhmal, P.: Polyhedral approximations in p-order cone programming. Optim. Method Softw. **29**, 1210–1237 (2014)
26. Xue, G., Ye, Y.: An efficient algorithm for minimizing a sum of P-norms. SIAM J.Optim. **10**, 551–579 (1997)
27. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. Roy. Stat. Soc. B Met. **67**, 301–320 (2005)