# Asymptotically tight Lagrangian dual of smooth nonconvex problems via improved error bound of Shapley-Folkman Lemma

Santanu S. Dey[1] and Jingye Xu[1*]

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 755 Ferst Dr NW, Atlanta, 30332, GA, USA.

*Corresponding author(s). E-mail(s): jxu673@gatech.edu;
Contributing authors: sdey30@gatech.edu;

**Abstract**

In convex geometry, the Shapley–Folkman Lemma asserts that the nonconvexity of a Minkowski sum of $\boldsymbol{n}$-dimensional bounded nonconvex sets does not accumulate once the number of summands exceeds the dimension $\boldsymbol{n}$, and thus the sum becomes approximately convex. Originally published by Starr in the context of quasi-equilibrium in nonconvex market models in economics, the lemma has since found widespread use in optimization, particularly for estimating the duality gap of the Lagrangian dual of separable nonconvex problems. Given its foundational nature, we pose the following geometric question: *Is it possible for the nonconvexity of the Minkowski sum of $\boldsymbol{n}$-dimensional nonconvex sets to even diminish instead of just not accumulating as the number of summands increases, under some general conditions?* We answer this affirmatively. First, we provide an elementary geometric proof of the Shapley–Folkman Lemma based on the facial structure of the convex hull of each set. This leads to refinement of the classical error bound derived from the lemma. Building on this new geometric perspective, we further show that when most of the sets satisfy a certain local smoothness condition which naturally arises in the epigraphs of smooth functions, their Minkowski sum converges directly to a convex set, with a vanishing nonconvexity measure. In optimization, this implies that the Lagrangian dual of block-structured smooth nonconvex problems—with potentially additional sparsity constraints—is asymptotically tight under mild assumptions, which contracts non-vanishing duality gap obtained via classical Shapley-Folkman Lemma.

## 1 Introduction

Consider a block-structured nonconvex program:

$$\mathrm{OPT}(\mathbf{b}) := \inf \sum_{i=1}^{k} \left\langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \right\rangle \tag{1a}$$

$$\text{s.t. } \sum_{i=1}^{k} B^{(i)} \mathbf{x}^{(i)} \leqslant \mathbf{b}, \tag{1b}$$

$$\mathbf{x}^{(i)} \in \mathcal{X}^{(i)} \subseteq \mathbb{R}^n, \forall i \in [k], \tag{1c}$$

where $\mathcal{X}^{(i)}$ is a closed possibly nonconvex set of $\mathbb{R}^n$, $B^{(i)} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Although the objective and the coupling constraints in (1) appear linear, this formulation is general enough to capture nonlinear problems. Indeed, all nonlinearities can be embedded in the feasible sets $\mathcal{X}^{(i)}$ by introducing suitable auxiliary variables.

Another important program associated with (1) is the Lagrangian dual (also called Lagrangian relaxation) of (1) by dualizing the coupling constraints (1b):

$$\begin{aligned} L(\lambda) := \inf_{\mathbf{x}} & \left( \sum_{i=1}^{k} \left\langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \right\rangle \right) + \left\langle \lambda, B^{(i)} \mathbf{x}^{(i)} - \mathbf{b} \right\rangle \\ \text{s.t. } & \mathbf{x}^{(i)} \in \mathcal{X}^{(i)} \subseteq \mathbb{R}^n, \forall i \in [k], \end{aligned} \tag{2}$$

and

$$\mathrm{DUAL}(\mathbf{b}) := \sup_{\lambda \geqslant 0} L(\lambda). \tag{3}$$

The Lagrangian dual is computationally promising because it is always a (non-smooth) convex program and provides a valid dual bound of (1) regardless of convexity of (1): $\mathrm{OPT}(\mathbf{b}) \geqslant \mathrm{DUAL}(\mathbf{b})$. On the other hand, the Lagrangian dual exploits the block-structure of (1) so that its subgradient can be obtained by solving $L(\lambda)$, which is fully decomposed into $k$ independent problems. That is, for each fixed $\lambda \geqslant 0$,

$$B^{(i)} \mathbf{x}_*^{(i)} - \mathbf{b} \in \partial L(\lambda), \text{ where } \mathbf{x}_* := \operatorname{arginf}_{\mathbf{x}} L(\lambda).$$

Therefore, given an oracle that minimizing a linear function over each $\mathcal{X}^{(i)}$, one can compute subgradient of $L(\lambda)$ and solve (3) via non-smooth optimization methods [1–3]. This allows distributed algorithm to solve block-structure problems (1) and it is also known as column generation or Dantzig–Wolfe decomposition [4, 5]. Aside from computational advantages, Lagrangian dual of nonconvex problems also finds other applications: sensitivity analysis of nonconvex problems [6], pricing in nonconvex markets [7, 8], market equilibrium [9] and online algorithm [10, 11].

Many textbooks and notes state that the strong duality, that is $\mathrm{OPT}(\mathbf{b}) = \mathrm{DUAL}(\mathbf{b})$, holds under certain regularity condition if all $\mathcal{X}^{(i)}$s are convex. This brings an impression that such level of convexity is needed for the existence of strong duality. However, the strong duality actually only requires the convexity of linear image of (1) onto the space of objective and constraints. More specifically, Let

$$\mathcal{P}^{(i)} := \left\{(t, \mathbf{d}) \;\middle|\; \exists \mathbf{x} \in \mathcal{X}^{(i)}, t = \left\langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \right\rangle, \mathbf{d} = B^{(i)}\mathbf{x}^{(i)} \right\},$$
$$\mathcal{P} := \sum_{i=1}^{k} \mathcal{P}^{(i)}$$

where we overload standard sum for vectors with Minkowski sum for sets and $\mathcal{P}$ is the Minkowski sum of all $\mathcal{P}^{(i)}$s. Under suitable regularity conditions, strong duality can be guaranteed whenever the image set $\mathcal{P}$ is convex [12]. In particular, when each set $\mathcal{X}^{(i)}$ is convex, the overall set $\mathcal{P}$ remains convex because both linear mappings and Minkowski sums preserve convexity. Therefore, the condition frequently shown in the textbooks and notes can be viewed as a sufficient condition. Although this distinction may appear subtle, it is in fact crucial—it underpins several nontrivial results concerning the strong duality in Lagrangian relaxations of nonconvex programs [13, 14]. A notable example is the celebrated $S$-Lemma [13], which establishes that the Lagrangian dual of a quadratic program with a single quadratic constraint is tight. This result can be interpreted as a direct consequence of Dines Lemma [15] that the image of two homogeneous quadratic mappings is always convex.

In this paper, we adopt this perspective and investigate the convexity of $\mathcal{P}$ to ensure the strong duality of the Lagrangian dual of (1). While each individual set $\mathcal{P}^{(i)}$ may be nonconvex, it is possible that their Minkowski sum $\mathcal{P}$ closely approximates its convex hull. This phenomenon is formally characterized by the *Shapley–Folkman Lemma* [9]. In the setting where $k \geqslant n$, the Shapley-Folkman Lemma states that

$$\forall \mathbf{x} \in \mathbf{conv}(\mathcal{P}),\ \exists I \in \binom{[k]}{n} \text{ such that } \mathbf{x} \in \left(\sum_{i \in [k] \setminus I} \mathcal{P}^{(i)}\right) + \left(\sum_{i \in I} \mathbf{conv}(\mathcal{P}^{(i)})\right).$$

Intuitively, the Shapley-Folkman Lemma implies that every point in $\mathbf{conv}(\mathcal{P})$ is close to some point in $\mathcal{P}$, up to at most $n$ nonconvex summands. If all $\mathcal{P}^{(i)}$ are contained within a ball of radius $R$, a quantitative measure of this approximate convexity can be obtained. In particular, under the Hausdorff distance, it has been shown [16] that

$$d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P})) \leqslant \sqrt{n}R.$$

This quantitative bound can further be employed to derive an a priori estimate of the duality gap $\Delta := \mathrm{OPT}(\mathbf{b}) - \mathrm{DUAL}(\mathbf{b})$ associated with problem (1) (see Theorem 3).

The Shapley–Folkman lemma was originally introduced by Starr in the study of quasi-equilibria in nonconvex market models in economics. In this context, it can be interpreted as an approximate form of strong duality for Lagrangian relaxations. Its connection to optimization was later developed in works such as [17–19], which used the lemma to quantify the duality gap of problem (1). More recently, several works have focused on refining the quantitative bounds provided by the Shapley–Folkman lemma, particularly on improving estimates of the Hausdorff distance $d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P}))$

or similar notion of nonconvexity; see, for example, [20–22]. An analogue of the Shapley–Folkman lemma for discrete convex sets has also been developed in [23]

In this work, we seek to refine the quantitative characterization of the Hausdorff distance $d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P}))$, given its central role in assessing the tightness of Lagrangian duals. We begin by presenting a new and elementary geometric proof of the Shapley-Folkman Lemma, which leverages the facial structure of each $\mathrm{conv}(\mathcal{P}^{(i)})$. This perspective enables us to refine the classical error bound. Building on this geometric insight, we further demonstrate that the Minkowski sum of the sets $\mathcal{P}^{(i)}$ can directly converges to its convex hull as the number of summands grows, i.e., $d_H(\mathcal{P}, \mathrm{conv}(\mathcal{P})) \to 0$ as $k \to \infty$, provided enough sets satisfy some mild "local smoothness" condition (see Theorem 4). Such smoothness naturally arises in the epigraphs of smooth (possibly nonconvex) functions, allowing us to establish that the following problem admits an asymptotically tight Lagrangian dual as $k \to \infty$:

$$\begin{aligned} \inf_{\mathbf{x}} \quad & \sum_{i=1}^{k} f^{(i)}(\mathbf{x}^{(i)}) \\ \text{s.t.} \quad & \sum_{i=1}^{k} B^{(i)}\mathbf{x}^{(i)} \leqslant \mathbf{b}, \\ & \|\mathbf{x}^{(i)}\|_0 \leqslant s^{(i)}, \quad \forall i \in [k], \end{aligned} \tag{4}$$

where each $f^{(i)}$ is a smooth (possibly nonconvex) function, and $B^{(i)}$ and $s^{(i)}$ satisfy some mild conditions (see Theorem 5). Note that we only requires that the sparsity level $s^{(i)}$ is not too small, i.e., $s^{(i)} \geqslant m+1, \forall i \in [k]$. To model problems without sparsity constraints, one may simply let $s^{(i)} = n, \forall i \in [k]$. To the best of our knowledge, this is the first result establishing an asymptotically vanishing nonconvexity measures for Minkowski sums of nonconvex sets within the framework of the Shapley–Folkman lemma.

This work focuses on Minkowski sum of $\mathcal{P}^{(i)}$ without averaging since its convexity directly reflects the duality gap of problem (1). In the classical applications of the Shapley–Folkman Lemma, the analysis is typically performed on the Minkowski average of the sets $\mathcal{P}^{(i)}$. This choice arises because the the error bound $d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P}))$ does not vanish in the classic analysis of Shapley-Folkman Lemma but $d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P}))$ is both homogeneous and independent of $k$, which immediately yields

$$d_H\left(\tfrac{1}{k}\mathcal{P}, \tfrac{1}{k}\mathbf{conv}(\mathcal{P})\right) \leqslant \tfrac{\sqrt{n}R}{k},$$

implying a linear convergence of the averaged sum toward its convex hull as $k$ increases. In contrast, our analysis applies directly to the unnormalized Minkowski sum. When averaging is applied, our result demonstrates a superlinear convergence rate, thereby strengthening the classical result.

## 2 Preliminary

In this section, we introduce several useful notation and results to study Shapley-Folkman Lemma and the Lagrangian relaxation of (1).

## 2.1 Notation

For a set $A \subset \mathbb{R}^d$, we define its *radius* as $\mathrm{R}(A) := \inf_{x \in \mathbb{R}^d} \{r \geqslant 0 : A \subseteq \mathcal{B}(x,r)\}$, where $\mathcal{B}(x,r) := \{y \in \mathbb{R}^d : \|y-x\| \leqslant r\}$ denotes the closed Euclidean ball of radius $r$ centered at $x$. Given two sets $A, B \subset \mathbb{R}^d$, their *Minkowski sum* is defined as $A+B := \{a+b : a \in A,\ b \in B\}$. For any scalar $\lambda \geqslant 0$, the *Minkowski scaling* of $A$ is $\lambda A := \{\lambda a : a \in A\}$.We denote by $\mathbf{conv}(A) := \{\sum_{i=1}^m \lambda_i a_i \mid a_i \in A,\ \lambda_i \geqslant 0,\ \sum_{i=1}^m \lambda_i = 1\}$ the convex hull of $A$. The *closed convex hull* is defined as $\overline{\mathbf{conv}}(A) := \mathrm{cl}(\mathbf{conv}(A))$, where $\mathrm{cl}(A)$ denotes the topological closure of $A$.

## 2.2 Nonconvexity measure and convex hull operator

Given a set $A \subseteq \mathbb{R}^n$, the exists a rich body of literature on how to measure non-convexity of $A$ due to its wide-ranging applications in nonconvex optimization. Throughout this paper, we will focus on two measures:

1. (Hausdorff distance from the convex hull) Given a closed convex set $K$ with $\mathbf{0}$ in its interior, we define

$$\Phi^K(A) := \inf \{r \geqslant 0 : \mathbf{conv}(A) \subseteq A + rK\}.$$

When $K$ is omitted, we refer it as the Hausdorff distance from the convex hull under $l_2$ norm:

$$\Phi(A) := \inf \{r \geqslant 0 : \mathbf{conv}(A) \subseteq A + r\mathcal{B}(\mathbf{0},1)\}.$$

2. (The inner radius of a nonconvex set [9])

$$\Xi(A) := \sup_{\mathbf{x} \in \mathbf{conv}(A)} \{\inf \mathrm{R}(T) : T \subseteq A, \mathbf{x} \in \mathbf{conv}(T)\}$$


**Fig. 1**: $A$ and $\mathbf{conv}(A)$

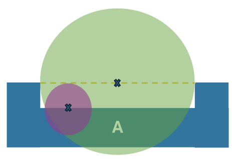


**Fig. 2**: two nonconvexity measures where pink ball corresponds to $\Phi(A)$ and green ball corresponds to $\Xi(A)$.

In the literature on the Shapley–Folkman Lemma, several measures of nonconvexity have been proposed. The measure $\Xi(\cdot)$, introduced in the seminal work by Starr [9], is used to establish the notion of quasi-equilibrium in nonconvex market models in economics. In this paper, we focus primarily on $\Phi(\cdot)$, as it is more directly related to the duality gap in Lagrangian dual (see Theorem 3). Other nonconvexity measures, exhibiting distinct behaviors and intended for different analytical purposes, are not listed in this paper; for a comprehensive overview, we refer readers to the survey [16].

**Lemma 1** [16, 24] Let $A$ be a subset of $\mathbb{R}^n$, then

$$\Phi(A) \leqslant \Xi(A) \leqslant \mathrm{R}(A).$$

**Claim 1** [16] Given a closed bounded convex set $K$, $\Phi^K(\cdot)$ are subadditive and positive homogeneous. That is for any $A, B \subseteq \mathbb{R}^n$ and $\kappa > 0$, it follows that

$$\Phi^K(A+B) \leqslant \Phi^K(A) + \Phi^K(B);$$
$$\Phi^K(\kappa A) = \kappa \Phi^K(A).$$

Similar results hold for $\Xi^2(\cdot)$.

The analysis in this paper relies heavily on the convex hull operator $\mathbf{conv}(\cdot)$. Two structural properties of convex hulls are used repeatedly: commutativity with Minkowski sums and commutativity with linear projections. We restate these well-known facts below for convenience.

**Claim 2** [16] (**conv** commutes with Minkowski sum) Let $A_1, \ldots, A_k$ be nonempty sets of $\mathbb{R}^n$. Then it follows that

$$\mathbf{conv}\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} \mathbf{conv}\left(A_i\right)$$

**Claim 3** [25] (**conv** commutes with linear projection)

$$\mathbf{conv}(\mathcal{P}^{(i)}) = \left\{(t, \mathbf{d}) \,\middle|\, \exists \mathbf{x} \in \mathbf{conv}\left(\mathcal{X}^{(i)}\right), t = \left\langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \right\rangle, \mathbf{d} = B^{(i)}\mathbf{x}^{(i)}\right\}.$$

# 3 Shapley-Folkman Lemma and duality gap of Lagrangian relaxation

In this section, we state Shapley-Folkman Lemma formally and show how it is related to duality gap of Lagrangian relaxation.

## 3.1 Shapley-Folkman Lemma

**Theorem 1** *(Shapley-Folkman Lemma) Let $A_1, \ldots, A_k$ be nonempty closed sets of $\mathbb{R}^n$ with $k \geqslant n+1$. Let $\mathbf{x} \in \mathbf{conv}(\sum_{i\in[k]} A_i) = \sum_{i\in[k]} \mathbf{conv}(A_i)$. Then there exists $\mathcal{I} \subseteq [k]$ with cardinality at most $n$ such that*

$$\mathbf{x} \in \sum_{i\in\mathcal{I}} \mathbf{conv}(A_i) + \sum_{i\in[k]\setminus\mathcal{I}} A_i.$$

Shapley-Folkman Lemma is a direct consequence of Carathéodory's theorem and there are many different ways to prove it [20, 21] and we defer our geometric proof in Lemma 2. Shapley-Folkman Lemma can be interpreted that the nonconvexity of the Minkowski sum of nonconvex sets, once reaching certain cap depending on the dimension, will not accumulate as the number of nonconvex sets increase. This phenomenon can be quantified by the following corollary. The proof of a slightly improved version is provided later in Lemma 4 under the pointedness assumption. In general, estimating the nonconvexity measures $\Phi(\cdot)$ and $\Xi(\cdot)$ is challenging. Many applications of the Shapley–Folkman Lemma therefore consider bounded nonconvex sets using the second part of Corollary 1.

**Corollary 1** Let Let $A_1, \ldots, A_k$ be nonempty sets of $\mathbb{R}^n$ with $k \geqslant n$ such that $\Xi(A_i) \leqslant \beta, \forall i \in [k]$. Then it follows that

$$\Phi\left(\sum_{i\in[k]} A_i\right) \leqslant \Xi\left(\sum_{i\in[k]} A_i\right) \leqslant \sqrt{n}\beta \text{ and therefore } \Phi\left(\frac{1}{k}\sum_{i\in[k]} A_i\right) \leqslant \frac{n}{k}\beta.$$

If we further assumes $\mathrm{R}(A_i) \leqslant \gamma, \forall i \in [k]$, then it follows that

$$\Phi\left(\sum_{i\in[k]} A_i\right) \leqslant \Xi\left(\sum_{i\in[k]} A_i\right) \leqslant \sqrt{n}\gamma \text{ and therefore } \Phi\left(\frac{1}{k}\sum_{i\in[k]} A_i\right) \leqslant \frac{n}{k}\gamma.$$

## 3.2 Duality gap of Lagrangian relaxation

In optimization theory, a key quantity associated with the Lagrangian dual that reflects the tightness of the relaxation is the duality gap:

$$\Delta := \mathrm{OPT}(\mathbf{b}) - \mathrm{DUAL}(\mathbf{b}).$$

In this subsection, we wish to obtain a priori bound on $\Delta$ based on $\Phi(\mathcal{P})$. To achieve this, we first use the following primal characterization [26, 27] of Lagrangian relaxation (bi-conjugacy in convex analysis [28]) to reduce $\mathrm{DUAL}(\mathbf{b})$ from min-max optimization to a minimization problem in the original space. Consider

$$\begin{aligned} \mathrm{OPT}_{\mathbf{L}}(\mathbf{b}) := \min & \sum_{i=1}^{k} \left\langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \right\rangle \\ \text{s.t. } & \sum_{i=1}^{k} B^{(i)}\mathbf{x}^{(i)} = \mathbf{b}, \\ & \mathbf{x}^{(i)} \in \overline{\mathbf{conv}}(\mathcal{X}^{(i)}) \subseteq \mathbb{R}^n, \forall i \in [k]. \end{aligned} \tag{5}$$

**Theorem 2** *[27, 28] Under proper regularity condition,* $\mathrm{OPT}_{\mathbf{L}}(\mathbf{b}) = \mathrm{DUAL}(\mathbf{b})$.

Common regularity conditions required for Theorem 2 include the existence of a Slater point or the assumption that each $\mathbf{conv}(\mathcal{X}^{(i)})$ is a polyhedron [29, 30]. The absence of such regularity may result in cases where $\mathrm{OPT}_{\mathbf{L}}(\mathbf{b}) > \mathrm{DUAL}(\mathbf{b})$ [30].

Throughout this paper, we exclude these pathological instances and directly assume that the following equivalence holds:

**Assumption 1** $\mathrm{OPT}_{\mathbf{L}}(\mathbf{b}) = \mathrm{DUAL}(\mathbf{b})$.

We finally requires two additional assumption on the geometry of $\mathcal{X}^{(i)}$.

**Assumption 2** $\overline{\mathbf{conv}}\{\mathcal{X}^{(i)}\} = \mathbf{conv}(\mathcal{X}^{(i)}), \forall i \in [k]$ .

**Assumption 3** $\mathbf{conv}(\mathcal{X}^{(i)})$ is pointed, $\forall i \in [k]$. That is, each $\mathbf{conv}(X^{(i)})$ contains no line.

We note that Assumption 2 and Assumption 3 are quite general. They are automatically satisfied when $\mathcal{X}^{(i)}$ is compact or $\mathcal{X}^{(i)}$ is an epigraph of a closed 1-coercive function. Even if each set $\mathcal{X}^{(i)}$ is closed, it is still possible that its convex hull $\mathbf{conv}(\mathcal{X}^{(i)})$ is not closed. Therefore, the closed convex hull operator is required in (5) in general. Since our goal is to compare (5) with (1), additional structures or Assumption 2 are expected. On the other hand, Assumption 3 guarantees that each $\mathbf{conv}(\mathcal{X}^{(i)})$ has at least one extreme point, and moreover any such extreme point must lie in the original set $\mathcal{X}^{(i)}$. The pointedness assumption is subsequently used in Lemma 2 to describe how the Minkowski sum of pointed closed convex sets decomposes in terms of their facial structure. This decomposition provides a geometric proof of the Shapley–Folkman Lemma. We remark that many results in this paper continue to hold—after modest modifications—even without Lemma 2 and Assumption 3. Nevertheless, we include Lemma 2 because it gives a clean geometric characterization of Minkowski sums of pointed closed convex sets and conveys the central theme of this paper: the local geometry around extreme points impacts the global behavior of Minkowski sums.

As we mentioned earlier, the tightness of Lagrangian dual of (1) directly relies on the convexity of $\mathcal{P} = \sum_{i=1}^{k} \mathcal{P}^{(i)}$ and we close this section by quantifying $\Delta$ in terms of $\Phi(\mathcal{P})$.

**Theorem 3** *Let $\mathcal{E} := \Phi(\mathcal{P})$. Under Assumption 1 and Assumption 2, it then follows that*

$$\mathrm{OPT}(\mathbf{b} + \mathcal{E}\mathbf{1}) - \mathcal{E} \leqslant \mathrm{DUAL}(\mathbf{b}) \leqslant \mathrm{OPT}(\mathbf{b}),$$

*and therefore*

$$\Delta \leqslant \mathcal{E} + \mathrm{OPT}(\mathbf{b}) - \mathrm{OPT}(\mathbf{b} + \mathcal{E}\mathbf{1}).$$

*Proof* Under the Assumption 1, we know that $\mathrm{DUAL}(\mathbf{b}) = \mathrm{OPT}_{\mathbf{L}}(\mathbf{b})$.

Let $\mathbf{x}_{\mathbf{L}}$ be the optimal solution of (5). By weak duality, we know that

$$\mathrm{OPT}_{\mathbf{L}}(\mathbf{b}) = \sum_{i=1}^{k} \left\langle \mathbf{c}^{(i)}, \mathbf{x}_{\mathbf{L}}^{(i)} \right\rangle = \mathrm{DUAL}(\mathbf{b}) \leqslant \mathrm{OPT}(\mathbf{b}). \tag{6}$$

Let $\mathbf{p}_{\mathbf{L}}^{(i)} := \begin{bmatrix} \left\langle \mathbf{c}^{(i)}, \mathbf{x}_{\mathbf{L}}^{(i)} \right\rangle \\ B^{(i)} \mathbf{x}_{\mathbf{L}}^{(i)} \end{bmatrix}$. By Claim 3 and Assumption 2, $\mathbf{p}^{(i)} \in \mathbf{conv}(\mathcal{P}^{(i)})$. By Claim 2, $\sum_{i=1}^{k} \mathbf{p}^{(i)} \in \sum_{i=1}^{k} \mathbf{conv}(\mathcal{P}^{(i)}) = \mathbf{conv}(\mathcal{P})$. By the definition of $\Phi(\mathcal{P})$, there exists some $\mathbf{x}_{*} \in \mathcal{X}^{(i)}$ and $\mathbf{p}_{*}^{(i)} := \begin{bmatrix} \left\langle \mathbf{c}^{(i)}, \mathbf{x}_{*}^{(i)} \right\rangle \\ B^{(i)} \mathbf{x}_{*}^{(i)} \end{bmatrix}$ such that

$$\left\| \left( \sum_{i=1}^{k} \mathbf{p}_{\mathbf{L}}^{(i)} \right) - \left( \sum_{i=1}^{k} \mathbf{p}_{*}^{(i)} \right) \right\|_2 \leqslant \mathcal{E}.$$

This means that $\mathbf{x}^{*}$ is an feasible solution of $\mathrm{OPT}(\mathbf{b} + \mathcal{E}\mathbf{1})$ and therefore

$$\mathrm{OPT}(\mathbf{b} + \mathcal{E}\mathbf{1}) \leqslant \sum_{i=1}^{k} \left\langle \mathbf{c}^{(i)}, \mathbf{x}_{*}^{(i)} \right\rangle \leqslant \left( \sum_{i=1}^{k} \left\langle \mathbf{c}^{(i)}, \mathbf{x}_{\mathbf{L}}^{(i)} \right\rangle \right) + \mathcal{E}.$$

Combining with (6), this yields that

$$\Delta = \mathrm{OPT}(\mathbf{b}) - \mathrm{OPT}_{\mathbf{L}}(\mathbf{b}) \leqslant \mathcal{E} + \mathrm{OPT}(\mathbf{b}) - \mathrm{OPT}(\mathbf{b} + \mathcal{E}\mathbf{1}).$$

□

# 4 Error bounds via Shapley-Folkman Lemma

## 4.1 Geometric proof of Shapley-Folkman Lemma

In this section, we give a proof of Shapley-Folkman Lemma that utilized the facial structure of the convex hull of each nonconvex sets.

**Definition 1** Given a convex set $C$, we call a close convex set $F \subseteq C$ is a **face** of $C$ if for any line segment $[a, b] \subseteq C$ such that $(a, b) \cap F \neq \varnothing$, we have $[a, b] \in F$. If there exists a supporting hyperplane $H$ of $C$ such that $F = C \cap H$, then we call $F$ an **exposed face**. The dimension of face $F$ denoted by $\dim(F)$ is the dimension of its affine hull.

**Lemma 2** Let $A_1, \ldots, A_k$ be nonempty sets of $\mathbb{R}^n$ with $k \geqslant n + 1$ such that $\mathbf{conv}(A_i)$ is closed and pointed. Then for every $\mathbf{y} \in \mathbf{conv}\left(\sum_{i=1}^{k} A_i\right)$, there exist faces $F_i$ of $\mathbf{conv}(A_i)$ such that

$$\mathbf{y} \in \sum_{i=1}^{k} F_i \text{ and } \sum_{i=1}^{k} \dim(F_i) \leqslant n.$$

*Proof* For any $\mathbf{y} \in \mathbf{conv}\left(\sum_{i=1}^{k} A_i\right)$, consider $\mathbf{y} = \sum_{i=1}^{k} \mathbf{y}_i$ where each $\mathbf{y}_i$ lies in the relative interior some face $F_i$ of $\mathbf{conv}(A_i)$. If $\sum_{i=1}^{k} \dim(F_i) > n$, we show how to find a different $\mathbf{y}_i'$ from $\mathbf{conv}(A_i)$ so that the total dimension $\sum_{i=1}^{k} \dim(F_i)$ is decrease. For each $F_i$, we can find a set $V_{F_i}$ of $\dim(F_i)$ linear independent vectors so that $\mathbf{y}_i \pm \epsilon \mathbf{v} \in F_i$ for all $\mathbf{v} \in V_{F_i}$ and small enough $\epsilon > 0$. Now extending $V_{F_i}$ in the space of $A_1 \times A_2 \times \cdots A_k$, we construct

$$V = \left\{ \tilde{\mathbf{v}} := \begin{bmatrix} \mathbf{0} \\ \mathbf{v} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(i-1)n} \times \mathbb{R}^{n} \times \mathbb{R}^{(k-i)n}, \forall \mathbf{v} \in V_{F_i}, \forall i \in [k] \right\}.$$

Clearly, $V$ includes $\sum_{i=1}^{k} \dim(F_i)$ linearly independent vectors. If $\sum_{i=1}^{k} \dim(F_i) > n$, there exists $\mathbf{d} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_k] \in \text{span}\{V\}$ that

$$\mathbf{d}_1 + \mathbf{d}_2 + \cdots + \mathbf{d}_k = \mathbf{0}$$

Let $[\mathbf{y}'_1, \mathbf{y}'_2, \cdots, \mathbf{y}'_k] = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_k] + \lambda \mathbf{d}$. Since each $\mathbf{conv}(A_i)$ is pointed, there exists some choice of $\lambda, \epsilon$ such that at least one $\mathbf{y}'_i$ reaches the boundary of $F_i$ while keeping other $\mathbf{y}'_j$ in $F_j$. In this case, the total dimension $\sum_{i=1}^{k} \dim(F_i)$ is decreased and we still ensure that $\mathbf{y} = \sum_{i=1}^{k} \mathbf{y}'_i$ for some $\mathbf{y}'_i \in \mathbf{conv}(A_i)$. □

Since $\dim(F_i) \in \{0, 1, \ldots, n\}$ and $\sum_{i=1}^{k} \dim(F_i) \leqslant n$ implies that at least $k - n$ faces have dimension 0 and are extreme points which lies in $A_i$. Let $\mathcal{I} = \{i \in [k] : \dim(F^{(i)}) > 0\}$. It is straightforward to check that

$$\mathbf{y} \in \sum_{i \in \mathcal{I}} \mathbf{conv}(A_i) + \sum_{i \in [k] \setminus \mathcal{I}} A_i.$$

and therefore Lemma 2 recovers the class Shapley-Folkman Lemma (Theorem 1) under the pointedness condition.

Moreover, one can further use Lemma 2 to strength the refined Shapley-Folkman Lemma recently proposed in [20] where the notion of $r$-th convex hull in [20] corresponds to dimension of face $F_i$ in Lemma 2 but we additionally requires that the points used to construct $r$-th convex hull lie in a $r$-dimension face $F_i$.

## 4.2 Error bounds via randomized rounding

Let $A_1, A_2, \ldots, A_k$ be nonempty set of $\mathbb{R}^n$ such that $\mathbf{conv}(A_i)$ is pointed. In this subsection, we adopt the geometric perspective introduced in Lemma 2 to refine the classical bound of $\Phi\left(\sum_{i=1}^{k} A_i\right)$ stated in Corollary 1. For any $\mathbf{y} \in \mathbf{conv}\left(\sum_{i=1}^{k} A_i\right)$, Lemma 2 implies the existence of faces $F_i$ of $\mathbf{conv}(A_i)$ such that $\mathbf{y} = \sum_{i=1}^{k} \mathbf{y}_i$where $\mathbf{y}_i \in F_i, \ \forall i \in [k]$. By definition, $\Phi\left(\sum_{i=1}^{k} A_i\right)$ quantifies how close $\mathbf{y} \in \sum_{i=1}^{k} \mathbf{conv}(A_i)$ is to some point $\mathbf{y}' \in \sum_{i=1}^{k} A_i$ in the Euclidean norm, that is, $\|\mathbf{y} - \mathbf{y}'\|_2 \leqslant \Phi(\sum_{i=1}^{k} A_i)$. We observe that a good $\mathbf{y}'$ can be constructed via randomly sampling some points from $F_i$. We first establish some results that is used to obtain such sampling.

**Claim 4** Let $C$ be a convex set and $F$ be its face. For any $\mathbf{x} \in F$, if $\mathbf{x} = \sum_i \lambda_i \mathbf{x}_i$ for some $\lambda_i > 0, \mathbf{x}_i \in C, \forall i$ and $\sum_{i=1} \lambda_i = 1$. Then $\mathbf{x}_i \in F, \forall i$.

*Proof* Fix some index $j$. Write

$$x = (1 - \lambda_j) \underbrace{\left(\tfrac{1}{1-\lambda_j} \sum_{i \neq j} \lambda_i x_i\right)}_{=:y} + \lambda_j x_j.$$

By convexity of $C$, we have $y \in C$ and $x_j \in C$. Since $0 < \lambda_j < 1$ and $x \in F$, the point $x$ is a strict convex combination of $y$ and $x_j$. By the definition of a face, we conclude $y \in F$ and $x_j \in F$. As this holds for each $j$, all $x_i \in F$. ☐

**Claim 5** Let $A$ be a nonempty set of $\mathbb{R}^n$. If $\Xi(A) < \infty$, then for every face $F$ of $\mathbf{conv}(A)$, $\Xi(A \cap F) \leqslant \Xi(A)$.

*Proof* Consider any $\mathbf{x} \in F \subseteq \mathbf{conv}(A)$, by the definition of $\Xi(A)$, there exists $T \subseteq A$ such that $\mathrm{R}(T) \leqslant \Xi(A)$ and $\mathbf{x}$ is a strict convex combination of some points $\mathbf{x}_i$ from $T$. Since $F$ is the face of $\mathbf{conv}(A)$, Claim 4 implies that $\mathbf{x}_i \subseteq F$. Since $\mathbf{x}$ is an arbitrary point in $F$ and $\mathbf{conv}(A \cap F) \subseteq F$, this further implies that $\Xi(A \cap F) \leqslant \Xi(A)$. ☐

**Lemma 3** Let $A$ be a set of $\mathbb{R}^n$. Then for every $\mathbf{x} \in \mathbf{conv}(A)$, there exists a finite distribution $\mathcal{D}$ such that $\mathcal{D}$ is supported on $A$, $\mathbb{E}_{Z \sim \mathcal{D}}(Z) = \mathbf{x}$ and $\mathbb{E}(\|Z - \mathbb{E}(Z)\|_2^2) \leqslant \Xi^2(A)$ .

*Proof* For any $\mathbf{x} \in \mathbf{conv}(A)$, by the definition of $\Xi(A)$, there exist some $T \subseteq A$ such that $\mathbf{x} \in \mathbf{conv}(T)$ and $\mathrm{R}(T) \leqslant \Xi(A)$. Since the statement of the Lemma is invariant under translation, we may translate $A$ so that $\|\mathbf{y}\|_2 \leqslant \Xi(A)$ for all $\mathbf{y} \in T$. By Carathéodory's theorem, $\mathbf{x}$ can be written as

$$\mathbf{x} = \sum_{i \in [n+1]} \lambda_i \mathbf{p}_i \text{ where } \lambda \in \Delta^{n+1} \text{ and } \mathbf{p}_i \in T.$$

We define random variable $Z \sim \mathcal{D}$ that $Z = \mathbf{p}_i$ with probability $\lambda_i$. It is clear that $\mathbb{E}_{Z \sim D}(Z) = \mathbf{x}$ and

$$\mathbb{E}(\|Z - \mathbb{E}(Z)\|_2^2) = \mathbb{E}(\|Z\|_2^2) - \|\mathbb{E}Z\|_2^2 \leqslant \mathbb{E}(\|Z\|_2^2) \leqslant \Xi^2(A).$$

where the last line uses that fact that $\|\mathbf{p}\|_2 \leqslant \Xi(A)$ ☐

**Lemma 4** Let $A_1, A_2, \ldots, A_k$ be nonempty subsets of $\mathbb{R}^n$. Let $F_i$ be a face of $\mathbf{conv}(A_i), \forall i \in [k]$ and let $\mathcal{I} := \{i \in [k] : \dim(F_i) > 0\}$. Then it follows that

$$\Phi\left(\sum_{i=1}^{k} A_i \cap F_i\right) \leqslant \sqrt{\sum_{i \in \mathcal{I}} \Xi^2(F_i)} \leqslant \sqrt{\sum_{i \in \mathcal{I}} \Xi^2(A_i)}.$$

*Proof* For every $\mathbf{x} \in \mathbf{conv}\left(\sum_{i=1}^k A_i \cap F_i\right) = \sum_{i=1}^k \mathbf{conv}(A_i \cap F_i)$, there exits some $\mathbf{x}_i$ such that $\mathbf{x} = \sum_{i=1}^k \mathbf{x}_i$ and $\mathbf{x}_i \in \mathbf{conv}(A_i \cap F_i)$. We aim to construct $\mathbf{y}_i$ such that $\mathbf{y}_i \in A_i$ and $\left\|\left(\sum_{i=1}^k \mathbf{x}_i\right) - \left(\sum_{i=1}^k \mathbf{y}_i\right)\right\|_2 \leqslant \sqrt{\sum_{i \in \mathcal{I}} \Xi^2(F_i)}$. For any $i \notin \mathcal{I}$, $\dim(F_i) = 0$ and therefore it is a extreme point so that $\mathbf{x}_i \in A_i \cap F_i$. For $i \in I$, by Lemma 3, there exists a distribution $Z_i \sim \mathcal{D}_i$ such that $\mathbb{E}_{Z_i \sim \mathcal{D}_i}(Z_i) = \mathbf{x}_i$ and $\mathcal{D}_i$ is supported on $F_i \cap A_i$ and $\mathbb{E}(\|Z_i - \mathbf{y}_i\|_2^2) \leqslant \Xi^2(A_i \cap F_i) \leqslant \beta^2$. Then we (randomly) construct $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 + \ldots \mathbf{y}_k$ in the following way:

$$\mathbf{y}_i = \begin{cases} \mathbf{x}_i & \text{if } i \in [k] \backslash \mathcal{I}, \\ Z_i \sim \mathcal{D}_i & \text{if } i \in \mathcal{I}, \end{cases}$$

Therefore, it follows that

$$\mathbb{E}\left(\|\mathbf{x} - \mathbf{y}\|_2^2\right) = \mathbb{E}\left(\left\|\sum_{i \in \mathcal{I}} (\mathbf{x}_i - Z_i)\right\|_2^2\right)$$

$$= \sum_{i\in\mathcal{I}} \mathbb{E}\left(\|(\mathbf{x}_i - Z_i)\|_2^2\right)$$
$$\leqslant \sum_{i\in\mathcal{I}} \Xi^2(F_i).$$

Therefore there exists a choice of $\mathbf{y}$ such that

$$\left\|\left(\sum_{i=1}^k \mathbf{x}_i\right) - \left(\sum_{i=1}^k \mathbf{y}_i\right)\right\|_2 \leqslant \sqrt{\sum_{i\in\mathcal{I}} \Xi^2(F_i)}.$$

Since $\mathbf{x}$ is arbitrary and by Claim 5, this implies that $\Phi\left(\sum_{i=1}^k A_i \cap F_i\right) \leqslant \sqrt{\sum_{i\in\mathcal{I}} \Xi^2(F_i)} \leqslant \sqrt{\sum_{i\in\mathcal{I}} \Xi^2(A_i)}$. □

## 4.3 Improved error bounds via local geometry

In this subsection, we seek sharper bounds on $\Phi\left(\sum_{i=1}^k A_i\right)$ by leveraging the geometric structure of the summands $A_i$. Our goal is to identify sufficient conditions under which

$$\Phi\left(\sum_{i=1}^k A_i\right) \longrightarrow 0 \quad \text{as } k \to \infty,$$

that is, the Minkowski sum becomes asymptotically convex.

For each $\mathbf{y} = \sum_{i=1}^k \mathbf{y}_i \in \mathbf{conv}\left(\sum_{i=1}^k A_i\right)$, Lemma 2 implies that each $\mathbf{y}_i$ lies in the Minkowski sum of certain faces $F_i$ of $\mathbf{conv}(A_i)$. The error bound presented in Lemma 4 can be interpreted as follows: we attempt to round $\mathbf{y}$ as a whole to a point in $\sum_{i=1}^k (F_i \cap A_i)$. This perspective naturally suggests a potential further refinement of Lemma 4, wherein we aim to round $\mathbf{y}$ directly to a point in $\sum_{i=1}^k A_i$. However, this ambitious idea encounters a significant challenge, as it requires full characterization of the nonconvex sets $A_i$, which is typically unavailable in practice. To address this limitation, we instead focus on exploiting only the local geometric properties of each $A_i$.

**Definition 2** Let $A$ be an nonempty set. For any $Q \subseteq \mathbf{conv}(A)$, we call $H \subseteq A$ a *hidden convex component* associated with $Q$ if

1. $H \subseteq A$ is a closed convex set;
2. $Q \subseteq H$;

The definition above is illustrated in figure 3. Such hidden convex component may not be defined for general $Q$. On the other hand, for each extreme point $\mathbf{v}_i$ of $\mathbf{conv}(A_i)$, there always exists a trivial hidden convex component—namely the singleton set $H_i = \{\mathbf{v}_i\}$. It is immediate that the existence of such trivial hidden convex components does not drive $\Phi\left(\sum_{i=1}^k A_i\right)$ to zero. For instance, take $A_i = \{0,1\}^n$ for all $i \in [k]$. Their Minkowski sum $\sum_{i=1}^k A_i$ is contained in the integer lattice $\{0,1,\ldots,k\}^n$,

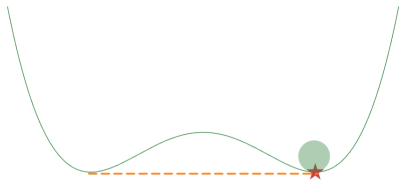

**Fig. 3**: Epigraph of $y = x^4 - x^2$ and the star point associates a $l_2$ ball as a hidden convex component.

and clearly $\Phi\left(\sum_{i=1}^k A_i\right)$ does not vanish. A natural next attempt is to require each hidden convex component $H_i$ to have nonempty interior. However, as illustrated in Figure 4, the mere existence of hidden convex components with non-vanishing interior still does not guarantee any improvement in the convexification of the Minkowski sum. In fact, diminishing the nonconvexity measure $\Phi\left(\sum_{i=1}^k A_i\right)$ requires smoothness condition of $H_i$.

**Definition 3** A differentiable function $f(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is called $L$-smooth if $f(\mathbf{y}) \leqslant f(\mathbf{x}) + \langle \mathbf{y} - \mathbf{x}, \nabla f(\mathbf{x})\rangle + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2, \forall \mathbf{y} \in \mathbb{R}^n$.

**Definition 4** Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set with 0 with its interior. The *gauge function* associated with $C$ is

$$\|\mathbf{x}\|_C := \inf\{\lambda > 0 : x \in \lambda C\}, \qquad x \in \mathbb{R}^n.$$

For a hidden convex component $H$ with $\mathbf{0}$ in its interior, we quantify its smoothness by the smoothness of the square gauge function $\|\cdot\|_H^2$, which is related to the concept of $(2, D)$-smooth in functional analysis [31].

To utilize the local geometry, we first overload the definition of $\Phi^K(\cdot)$: given a closed convex set $K$ with $\mathbf{0}$ in its interior, we let

$$\Phi^K(A, D) := \sup_{\mathbf{x} \in A} \inf_{\mathbf{y} \in D} \|\mathbf{x} - \mathbf{y}\|_K .$$

It is therefore straightforward to see that $\Phi^K(A) = \Phi^K(\mathbf{conv}(A), A)$. Moreover, it is worth mentioning that both $\Phi(A)$ and $\Xi(A)$ are invariant under translating $A$.

**Lemma 5** Let $A_1, A_2, \ldots, A_k$ be nonempty subsets of $\mathbb{R}^n$ and $\Xi(A_i) \leqslant \beta, \forall i \in [k]$. Let $F_i$ be a face of $\mathbf{conv}(A_i), \forall i \in [k]$ and $\mathcal{I} := \{i \in [k] : \dim(F_i) > 0\}$. Let $H_i$ be a hidden convex component associated with $F_i$ for all $i \in [k] \backslash \mathcal{I}$ and $H = \sum_{i \in [k] \backslash \mathcal{I}} H_i$. Suppose $H$ has nonempty interior, let $\mathcal{H}$ be a translation of $H$ that contains $\mathbf{0}$ in its interior and if we have $\|\cdot\|_{\mathcal{H}}^2$ is

$L$-smooth, then

$$\Phi^{\mathcal{H}}\left(\sum_{i=1}^{k} F_i, \sum_{i=1}^{k} A_i\right) \leqslant \sqrt{1 + \frac{L}{2}\sum_{i\in\mathcal{I}}^{n} \Xi^2(F_i)} - 1.$$

*Proof* $\Phi^{\mathcal{H}}\left(\sum_{i=1}^{k} F_i, \sum_{i=1}^{k} A_i\right)$ is equivalent to the maximum possible distance (under $\|\cdot\|_{\mathcal{H}}$) from points in $\sum_{i=1}^{k} F_i$ to $\sum_{i=1}^{k} A_i$. If $\sum_{i=1}^{k} F_i \subseteq \mathcal{F}$ for some $\mathcal{F} \subseteq \mathbb{R}^n$, it is clear that $\Phi^{\mathcal{H}}\left(\sum_{i=1}^{k} F_i, \sum_{i=1}^{k} A_i\right) \leqslant \Phi^{\mathcal{H}}\left(\mathcal{F}, \sum_{i=1}^{k} A_i\right)$. We derive an upper bound of this quantity by choosing $\mathcal{F} := \left(\sum_{i\in\mathcal{I}} F_i\right) + \left(\sum_{i\in[k]\setminus\mathcal{I}} H_i\right)$. We try to show that any point in $\left(\sum_{i\in\mathcal{I}} F_i\right) + \left(\sum_{i\in[k]\setminus\mathcal{I}} H_i\right)$ can be rounded to some point in $\left(\sum_{i\in\mathcal{I}} F_i \cap A_i\right) + \left(\sum_{i\in[k]\setminus\mathcal{I}} H_i\right) \subseteq \sum_{i=1}^{k} A_i$ with a desired distance.

Since the target statement is invariant under translation, we first translate $A_i$ and $H_i$ so that $\sum_{i\in[k]\setminus\mathcal{I}} H_i = \mathcal{H}$. Therefore, any point $\mathbf{p}$ in $\left(\sum_{i\in\mathcal{I}} F_i\right) + \left(\sum_{i\in[k]\setminus\mathcal{I}} H_i\right)$, can be written as $\mathbf{p} = \left(\sum_{i\in\mathcal{I}} \mathbf{p}_i\right) + \mathbf{v}$ where $\mathbf{p}_i \in F_i$ and $\mathbf{v} \in \mathcal{H}$. By Lemma 3, we can equip each $F_i$ with a distribution $Z_i \sim D_i$ with support on $A_i \cap F_i$ such that $\mathbb{E}_{Z_i\sim\mathcal{D}_i}(Z_i) = \mathbf{p}_i$ and $\mathbb{E}(\|Z_i - \mathbb{E}(Z_i)\|_2^2) \leqslant \Xi^2(F_i)$. Now we try to bound

$$\mathbb{E}\left(\left\|\left(\sum_{i\in\mathcal{I}} Z_i\right) - \mathbf{p}\right\|_{\mathcal{H}}^2\right) = \mathbb{E}\left(\left\|\left(\sum_{i\in\mathcal{I}} Z_i - \mathbf{p}_i\right) - \mathbf{v}\right\|_{\mathcal{H}}^2\right)$$

Let $\mathbf{g} := \nabla \|\mathbf{v}\|_{\mathcal{H}}^2$ and applying the definition of $L$-smoothness, this further yields that

$$\begin{aligned}
\mathbb{E}\left(\left\|\left(\sum_{i\in\mathcal{I}} Z_i - \mathbf{p}_i\right) - \mathbf{v}\right\|_{\mathcal{H}}^2\right) &\leqslant \mathbb{E}\left(\|\mathbf{v}\|_{\mathcal{H}}^2\right) + \mathbb{E}\left(\left\langle \sum_{i\in\mathcal{I}} (Z_i - \mathbf{p}_i), \mathbf{g}\right\rangle\right) + \frac{L}{2}\mathbb{E}\left(\left\|\left(\sum_{i\in\mathcal{I}} Z_i - \mathbf{p}_i\right)\right\|_2^2\right)\\
&= \mathbb{E}\left(\|\mathbf{v}\|_{\mathcal{H}}^2\right) + \frac{L}{2}\mathbb{E}\left(\left\|\left(\sum_{i\in\mathcal{I}} Z_i - \mathbf{p}_i\right)\right\|_2^2\right)\\
&= \mathbb{E}\left(\|\mathbf{v}\|_{\mathcal{H}}^2\right) + \frac{L}{2}\sum_{i\in\mathcal{I}} \mathbb{E}(\|Z_i - \mathbf{p}_i\|_2^2)\\
&\leqslant 1 + \frac{L}{2}\sum_{i\in\mathcal{I}} \Xi^2(F_i).
\end{aligned}$$

This implies there exists some point in $\sum_{i\in\mathcal{I}} A_i \cap F_i$ whose distance defined by $\|\cdot\|_{\mathcal{H}}$ to $\mathbf{p}$ is at most $\sqrt{1 + \frac{L}{2}\sum_{i\in\mathcal{I}} \Xi^2(F_i)}$. Therefore, by the positive homogeneousness of $\|\cdot\|_{\mathcal{H}}$, the $\|\cdot\|_{\mathcal{H}}$ distance of $\mathbf{p}$ to $\left(\sum_{i\in\mathcal{I}} A_i \cap F_i\right) + \mathcal{H}$ is at most $\sqrt{1 + \frac{L}{2}\sum_{i\in\mathcal{I}} \Xi^2(F_i)} - 1$. □

**Theorem 4** *Let $A_1, A_2, \ldots, A_k$ be nonempty subsets of $\mathbb{R}^n$ and $\Xi(A_i) \leqslant \beta, \forall i \in [k]$ such that $\mathbf{conv}(A_i)$ is pointed and closed. If every extreme point of $A_i$ has a $l_2$ ball with radius*

*$r_i$ ($r_i$ can be zero) as hidden convex component, let $r_* = \inf_{\mathcal{I}\subseteq[k]:|\mathcal{I}|=n} \sum_{i\in[k]\setminus\mathcal{I}} r_i$. Then it follows that*

$$\Phi\left(\sum_{i=1}^{k} A_i\right) \leqslant \sqrt{r_*^2 + \frac{1}{2}n\beta^2} - r_*.$$

*If there exists some $r > 0$ such that $r_i \geqslant r$ for all $i \in [k]$, then*

$$\Phi\left(\sum_{i=1}^{k} A_i\right) \leqslant \sqrt{(k-n)^2 r^2 + \frac{1}{2}n\beta^2} - (k-n)r.$$

*As $k \to \infty$, it is clear the above $\Phi\left(\sum_{i=1}^{k} A_i\right) \to 0$.*

*Proof* To bound $\Phi\left(\sum_{i=1}^{k} A_i\right)$, it suffices to show that for any $\mathbf{p} \in \sum_{i=1}^{k} \mathbf{conv}(A_i)$, we have $\Phi\left(\mathbf{p}, \sum_{i=1}^{k} A_i\right) \leqslant \sqrt{r_*^2 + \frac{1}{2}n\beta^2} - r_*$. By Lemma 2, there exists face $F_i$ of $\mathbf{conv}(A_i)$ such that $\mathbf{p} \subseteq \sum_{i=1}^{k} F_i$ and $\sum_{i=1}^{k} \dim(F_i) = n$. By the hypothesis, we can find $H_i = \mathcal{B}(\mathbf{0}, r_i)$ as hidden convex component associated with $F_i$ for all $i \in [k]\setminus\mathcal{I}$ such that $\mathcal{H} = \sum_{i\in[k]\setminus\mathcal{I}} H_i = \mathcal{B}(\mathbf{0}, r_*)$. Then $\|\mathbf{x}\|_{\mathcal{H}} = \frac{1}{r_*}\|\mathbf{x}\|$ and $\|\cdot\|_{\mathcal{H}}^2$ is $\frac{1}{r_*^2}$-smooth. Lemma 5 implies that

$$\Phi^{\mathcal{H}}\left(\sum_{i=1}^{k} F_i, \sum_{i=1}^{k} A_i\right) \leqslant \sqrt{1 + \frac{1}{2r_*^2}\sum_{i\in\mathcal{I}}^{n} \Xi^2(F_i)} - 1.$$

This implies that

$$\begin{aligned}\Phi\left(\sum_{i=1}^{k} F_i, \sum_{i=1}^{k} A_i\right) &\leqslant \sqrt{r_*^2 + \frac{1}{2}\sum_{i\in\mathcal{I}}^{n} \Xi^2(F_i)} - r_* \\ &\leqslant \sqrt{r_*^2 + \frac{1}{2}\sum_{i\in\mathcal{I}}^{n} \beta^2} - r_*\end{aligned}$$

where the last inequality uses Claim 5. Since $\mathbf{p}$ is arbitrary, this implies that $\Phi\left(\sum_{i=1}^{k} A_i\right) \leqslant \sqrt{r_*^2 + \frac{1}{2}n\beta^2} - r_*$.. The rest of statement directly follows that $r_* \leqslant (k-n)r$. □

We end the section with serval examples and remarks.

**Remark 1** Lemma 5 can be generalized by quantifying the smoothness of the hidden convex components through the smoothness (possibly with respect to a different norm) of the $p$-th power of a gauge function $\|\cdot\|_H^p$ for some $p > 1$, following the same probabilistic argument. We do not pursue this extension here, as it naturally requires alternative notions of nonconvexity beyond those considered in this work. For simplicity, we therefore focus on the classical nonconvexity measures $\Phi(\cdot)$ and $\Xi(\cdot)$ that appear in the literature on the Shapley–Folkman lemma, due to their direct relevance in estimating the duality gap $\Delta$.

**Remark 2** In general, the smoothness of hidden convex component is required for establishing the vanishing of nonconvexity. To illustrate this phenomenon, consider the simple example $A_0 + \sum_{i=1}^{k} \mathcal{B}_p(\mathbf{0}, 1) = A_0 + \mathcal{B}_p(\mathbf{0}, k)$ where $A_0 := \{(0,0), (0,1)\} \subseteq \mathbb{R}^2$ is a nonconvex set and $\mathcal{B}_p(\mathbf{0}, 1)$ denotes the unit $p$-norm ball. Figures 4, 5 and 6 depict the resulting Minkowski sums for $k = 1, 3$, and 6, respectively. We observe that the "nonconvex hollow" in the set gradually vanishes as the number of summands increases when $p = 1.5, 2$. In contrast, this hollow persists for all $k$ in the case $p = 1$. The underlying reason is that $\|\cdot\|_{\mathcal{B}_1(\mathbf{0},1)}^{f}$ is not smooth for any choice of $f > 1$.

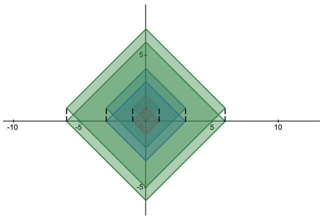

**Fig. 4**: $p = 1$

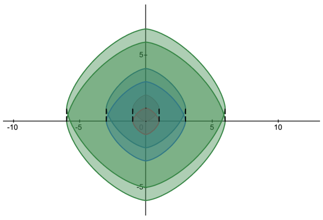

**Fig. 5**: $p = 1.5$

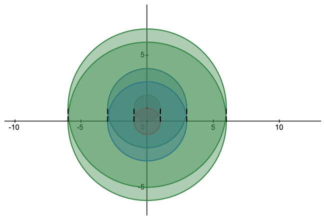

**Fig. 6**: $p = 2$

**Remark 3** It requires the existence of smooth hidden convex components only around the extreme points of each nonconvex set in order to ensure that their Minkowski sums converge to a convex set. Such smooth hidden components need not exist around non-extreme points. To illustrate this, consider the $\mathbb{R}^2$ examples $T_1$ and $T_2$ shown in Figures 7 and 9. Both sets contain some points on which no smooth hidden convex components are associated with. Nevertheless, one can verify that $\sum_{i=1}^{k} T_1$ converges to a convex set as $k \to \infty$, whereas $\sum_{i=1}^{k} T_2$ does not. The difference arises from the behavior at extreme points: every extreme point of $\mathbf{conv}(T_1)$ admits a smooth hidden convex component (e.g., a ball contained in $T_1$ in a neighborhood of that point), while $\mathbf{conv}(T_2)$ possesses five extreme points for which no such smooth hidden convex component exists.

**Fig. 7**: $T_1$

**Fig. 8**: $\mathbf{conv}(T_1)$

**Fig. 9**: $T_2$

**Fig. 10**: $\mathbf{conv}(T_2)$

## 5 Asymptotically tight Lagrangian dual of smooth nonconvex problems

In this section, we aim to leverage Theorem 4 and Lemma 5 to establish the asymptotic tightness of problem (1). The key insight is that the smooth hidden convex components naturally emerge at the extreme points of the convex hull of the epigraphs of smooth functions. Under mild assumptions on the constraint matrices $B^{(i)}$, the smoothness of these hidden convex components is approximately preserved under projection, thereby implying that $\Phi(\mathcal{P})$ vanishes as $k$ increases. We rewrite (4):

$$\begin{aligned} \mathrm{OPT}_{\mathbf{s}}(\mathbf{b}) := \inf & \sum_{i=1}^{k} t_i \\ \text{s.t. } & \sum_{i=1}^{k} B^{(i)}\mathbf{x}^{(i)} \leqslant \mathbf{b}, \\ & \left\|\mathbf{x}^{(i)}\right\|_0 \leqslant s_i, t_i \geqslant f^{(i)}(\mathbf{x}^{(i)}), \forall i \in [k]. \end{aligned} \tag{7}$$

In this case, we have

$$\mathcal{X}_{\mathbf{s}}^{(i)} := \left\{(t, \mathbf{x}) \in \mathbb{R}^{m+1} \mid \left\|\mathbf{x}^{(i)}\right\|_0 \leqslant s_i, t^{(i)} \geqslant f^{(i)}(\mathbf{x}^{(i)})\right\}$$

and the projected set $\mathcal{P}^{(i)}$ for (7) is

$$\begin{aligned} \mathcal{P}_{\mathbf{s}}^{(i)} &:= \left\{(t, \mathbf{d}) \in \mathbb{R}^{m+1} \;\middle|\; \begin{array}{l} \exists (t^{(i)}, \mathbf{x}^{(i)}) \text{ that } \left\|\mathbf{x}^{(i)}\right\|_0 \leqslant s_i, t^{(i)} \geqslant f^{(i)}(\mathbf{x}^{(i)}) \\ t = t^{(i)}, \mathbf{d} = B^{(i)}\mathbf{x}^{(i)} \end{array}\right\} \\ &= \left\{(t, \mathbf{d}) \in \mathbb{R}^{m+1} \;\middle|\; \begin{array}{l} \exists \mathbf{x}^{(i)} \text{ that } \left\|\mathbf{x}^{(i)}\right\|_0 \leqslant s_i, \\ t \geqslant f^{(i)}(\mathbf{x}^{(i)}), \mathbf{d} = B^{(i)}\mathbf{x}^{(i)} \end{array}\right\}. \end{aligned}$$

Note that we can simply choose $s_i = n$ in (7) to model the block-structured smooth nonconvex problem without sparsity constrained. On the other hand, $f^{(i)}$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$ and therefore is proper.

**Definition 5** (1-coercive function) A function $f : \mathbb{R}^n \to (-\infty, +\infty]$ is called *1-coercive* if

$$\frac{f(x)}{\|x\|} \longrightarrow +\infty \qquad \text{as } \|x\| \to \infty.$$

Equivalently, for every $M > 0$ there exists $R > 0$ such that

$$\|x\|_2 > R \implies f(x) > M\|x\|_2.$$

1-coercive function is commonly used in analysis and optimization. Standard example includes strictly convex quadratic functions and univariate polynomial with a positive leading coefficient. The existence of asymptotically tight Lagrangian dual requires two additional assumption. Assumption 4 requires the nonconvexity of each block in (4) is bounded. Assumption 5 implies that $\mathbf{conv}(\mathcal{X}_{\mathbf{s}}^{(i)})$ is closed and pointed in Lemma 6, which ensures Assumption 2 and Assumption 3.

**Assumption 4** There exists some $\beta$ such that $\Xi(\mathcal{P}_{\mathbf{s}}^{(i)}) \leqslant \beta, \forall i \in [k]$.

**Assumption 5** Each $f^{(i)}(\cdot)$ is 1-coercive and closed.

**Lemma 6** Under assumption 5, $\mathbf{conv}(\mathcal{X}_{\mathbf{s}}^{(i)})$ is closed and pointed.

*Proof* Let $\chi_{s_i}(\mathbf{x}) := \begin{cases} \infty & \text{if } \left\|\mathbf{x}^{(i)}\right\|_0 \geqslant s_i \\ 0 & \text{if otherwise} \end{cases}$. Let $\tilde{f}^{(i)}(\mathbf{x}) := f^{(i)}(\mathbf{x}) + \chi_{s_i}(\mathbf{x})$. By the definition of 1-coercive, it is straightforward that $\tilde{f}^{(i)}(\mathbf{x})$ is 1-coercive and closed. Moreover, it is clear that $\operatorname{epi} \tilde{f}^{(i)}(\mathbf{x}) = \mathcal{X}_{\mathbf{s}}^{(i)}$. It is well known that the convex hull of epigraph of proper, 1-coercive and closed function is closed and therefore $\mathbf{conv}\left(\mathcal{X}_{\mathbf{s}}^{(i)}\right)$ is closed [25]. Suppose that $\mathbf{conv}\left(\mathcal{X}_{\mathbf{s}}^{(i)}\right)$ is not pointed, we seek contradiction. We first observe that the definition of 1-coercive implies that there exists some $lb$ such that $lb \leqslant \tilde{f}^{(i)}(\mathbf{x})$. This implies that $t \geqslant lb, \forall (t, \mathbf{x}) \in \mathbf{conv}(\mathcal{X}_{\mathbf{s}}^{(i)})$. Therefore, any line contained in $\mathbf{conv}(\mathcal{X}_{\mathbf{s}}^{(i)})$ must take form of $l(t) := (t_0, \mathbf{x}_0) + t(0, \mathbf{r}) \forall t \in \mathbb{R}^n$ for some nonzero $\mathbf{r}$. Since $\tilde{f}^{(i)}(\mathbf{x})$ is 1-coercive, there exists some $R > 1$ such that $\|\mathbf{x}\|_2 \geqslant R \implies \tilde{f}^{(i)}(\mathbf{x}) \geqslant (|lb| + |t_0| + 1)\|\mathbf{x}\|$. Choose $t$ sufficiently large so that $\|\mathbf{x}_0 + t\mathbf{r}\|_2 \geqslant (n+2)R$, since $(t_0, \mathbf{x}_0 + t\mathbf{r}) \in \mathbf{conv}(\mathcal{X}_{\mathbf{s}}^{(i)})$, this implies that $\mathbf{x}_0 + t\mathbf{r} = \sum_{i \in \mathcal{K}} \lambda_i (t_i, \mathbf{x}_i)$ for some $(t_i, \mathbf{x}_i) \in \mathcal{X}_{\mathbf{s}}^{(i)}$ and $\lambda$ from a simplex of dimension $|\mathcal{K}|$. Since $\|\cdot\|_2$ is subadditive and positive homogeneous, this implies that $\|\mathbf{x}_0 + t\mathbf{r}\| \leqslant \sum_{i \in \mathcal{K}} \lambda_i \|\mathbf{x}_i\|$. By Carathéodory's Theorem, we have $|\mathcal{K}| \leqslant (n+2)$. Since $\|\mathbf{x}_0 + t\mathbf{r}\| \geqslant (n+2)R$, a simple pigeonhole argument implies that there exists some $j \in \mathcal{K}$, such that $\lambda_j \left\|\mathbf{x}_j\right\| \geqslant R$. Finally, this show that

$$\left\|\mathbf{x}_j\right\| \geqslant R \implies \lambda_j t_j \geqslant \frac{R}{\|\mathbf{x}\|} t_j \geqslant R(|lb| + |t_0| + 1) \geqslant |lb| + |t_0| + 1.$$

This implies that

$$\sum_{i \in \mathcal{K}} \lambda_i t_i = \lambda_j t_j + \sum_{i \in \mathcal{K} \setminus \{j\}} \lambda_i t_i \geqslant |lb| + |t_0| + 1 + lb > t_0,$$

which leads to contraction. □

The last ingredient for applying Lemma 5 is to show that there exists a smooth hidden convex component associated with each extreme point of $\mathbf{conv}(\mathcal{P}_{\mathbf{s}}^{(i)})$. We obtain this from the following observation. The set $\mathcal{P}_{\mathbf{s}}^{(i)}$ can be viewed as a linear projection of a union of epigraphs of certain smooth functions. By the definition of smoothness, for any point in such an epigraph we can find a ball that lies entirely in the epigraph and contains this point by Claim 6. Under the linear projection induced by the coupling matrix $B^{(i)}$, this ball is mapped to an ellipsoid in $\mathcal{P}_{\mathbf{s}}^{(i)}$, which serves as the smooth hidden convex component we are looking for. The maximal distortion under this linear projection is quantified by $\mathcal{L}_s(B^{(i)})$, defined below.

**Definition 6** Given a matrix $B \in \mathbb{R}^{m \times n}$ and let $\tilde{B} \in \mathbb{R}^{(m+1) \times n} := \begin{bmatrix} \mathbf{1}^\top \\ B \end{bmatrix}$ be a matrix with appending a row of all ones. We define the *projection factor with respect to sparsity level s* as

$$\mathcal{L}_s(B) := \inf_{S \subseteq [n] : |S| = s} \sigma_{\inf}\left(\tilde{B}_S (\tilde{B}_S)^\top\right)$$

where $\sigma_{\inf}(\cdot)$ is the minimal singular value and $\tilde{B}_S \in \mathbb{R}^{(m+1)\times|S|}$ is the submatrix of $\tilde{B}$ that consists columns of $B$ with indices in $S$.

**Remark 4** To obtain a nondegenerate hidden convex component, we require $\mathcal{L}_s(B^{(i)}) > 0$. This condition is equivalent to the existence of $m+1$ affinely independent vectors among every collection of $s^{(i)}$ columns of $B^{(i)}$. A necessary condition for this is $s^{(i)} \geqslant m+1$. Apart from this simple dimensional requirement, the condition $\mathcal{L}_s(B^{(i)}) > 0$ is not very restrictive: it holds generically and can be ensured almost surely by adding an arbitrarily small smooth random perturbation to $B^{(i)}$ whenever $s^{(i)} \geqslant m+1$.

**Claim 6** For any $L$-smooth function convex $f(\cdot)$ and any $\mathbf{x} \in \mathbb{R}^n$, there exists a $l_2$-ball $B$ with radius $\frac{1}{L}$ such that $B \subseteq \operatorname{epi} f$ and $(\mathbf{x}, f(\mathbf{x})) \subseteq B$.

*Proof* For any $\mathbf{x} \in \mathbb{R}^n$, let $\mathbf{g} = \nabla f(\mathbf{x})$. We construct the center of the ball $B$ with radius $r := \frac{1}{L}$ by

$$(\mathbf{x}_c, t_c) = (\mathbf{x}, f(\mathbf{x}) + \frac{1}{L}\frac{1}{\sqrt{\|\mathbf{g}\|_2^2 + 1}}(-\mathbf{g}, 1)\,.$$

Clearly, $(\mathbf{x}, f(\mathbf{x})) \in B$ and it remains to prove that $B \subseteq \operatorname{epi} f$. For every point $(\mathbf{y}, t) \subseteq B$, the lowest possible $t$ is

$$t_{\inf}(\mathbf{y}) := t_c - \sqrt{r^2 - \|\mathbf{y} - \mathbf{x}_c)\|_2^2}$$

Therefore, it suffices to prove $f(\mathbf{y}) \leqslant t_{\inf}(\mathbf{y})$ for all $\mathbf{y}$ such that $(\mathbf{y} - \mathbf{x}_c)^2 \leqslant r^2$. Note that $f(\mathbf{y}) \leqslant t_{\inf}(\mathbf{y})$ can be further simplified to prove that

$$r^2 \leqslant (f(\mathbf{y}) - t_c)^2 + \|\mathbf{y} - \mathbf{x}_c\|_2^2\,.$$

Since $f(\cdot)$ is a smooth convex function, it admits a quadratic upper bound

$$f(\mathbf{y}) \leqslant f(\mathbf{x}) + \langle \mathbf{y} - \mathbf{x}, \mathbf{g}\rangle + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2\,, \forall \mathbf{y} \in \mathbb{R}^n. \tag{8}$$

Let $\alpha := \dfrac{1}{\sqrt{\|\mathbf{g}\|_2^2 + 1}}$, $\Delta\mathbf{x} := \mathbf{y} - \mathbf{x}$ and $s := f(\mathbf{y}) - f(\mathbf{x}) - \langle \mathbf{g}, \Delta\mathbf{x}\rangle$. In this case, it follows that

$$\begin{aligned}
(f(\mathbf{y}) - t_c)^2 + \|\mathbf{y} - \mathbf{x}_c\|_2^2 &= (f(\mathbf{y}) - f(\mathbf{x}) - r\alpha)^2 + \|\mathbf{y} - \mathbf{x} + r\alpha\mathbf{g}\|_2^2 \\
&= (s + \langle \mathbf{g}, \Delta\mathbf{x}\rangle - r\alpha)^2 + \|\Delta\mathbf{x} + r\alpha\mathbf{g}\|_2^2 \\
&= (s + \langle \mathbf{g}, \Delta\mathbf{x}\rangle)^2 + r^2\alpha^2 - 2r\alpha(s + \langle \mathbf{g}, \Delta\mathbf{x}\rangle) + \|\Delta\mathbf{x}\|_2^2 + \|r\alpha\mathbf{g}\|_2^2 + 2\langle r\alpha\mathbf{g}, \Delta\mathbf{x}\rangle \\
&= (s + \langle \mathbf{g}, \Delta\mathbf{x}\rangle)^2 + r^2\alpha^2 - 2r\alpha s + \|\Delta\mathbf{x}\|_2^2 + \|r\alpha\mathbf{g}\|_2^2 \\
&= (s + \langle \mathbf{g}, \Delta\mathbf{x}\rangle)^2 - 2r\alpha s + \|\Delta\mathbf{x}\|_2^2 + r^2 \\
&\geqslant -2r\alpha s + \|\Delta\mathbf{x}\|_2^2 + r^2 \\
&\geqslant (-\alpha + 1)\|\Delta\mathbf{x}\|_2^2 + r^2 \\
&\geqslant r^2
\end{aligned}$$

where the first inequality just drops $(s + \langle \mathbf{g}, \Delta\mathbf{x}\rangle)^2$ term; the second inequality uses (8); the third inequality uses the fact that $\alpha \leqslant 1$.

□

**Claim 7** For every extreme point $\mathbf{v}$ (0-dimensional face) of $\mathbf{conv}(\mathcal{P}^{(i)})$, there exists a ball $l_2$-ball $B$ with radius $\frac{\mathcal{L}_s(B^{(i)})}{L}$ such that $\mathbf{v} \in B \subseteq \mathcal{P}^{(i)}$.

*Proof* Fixed any support set $S \subseteq [n]$, we define

$$\mathcal{P}_S^{(i)} := \left\{ (t, \mathbf{d}) \in \mathbb{R}^{m+1} \;\middle|\; \begin{array}{l} \exists \mathbf{x}, x_i = 0, \forall i \notin S; \\ \mathbf{d} = B^{(i)}\mathbf{x}; \\ t \geqslant f(\mathbf{x}); \end{array} \right\}, \tag{9}$$

It is clear that

$$\mathcal{P}^{(i)} = \bigcup_{S \in \binom{[n]}{s_i}} \mathcal{P}_S^{(i)}.$$

Let $\mathbf{v} := (t_\mathbf{v}, \mathbf{d}_\mathbf{v})$ be any extreme point of $\mathbf{conv}(\mathcal{P}^{(i)})$. In this case, there must exist some $S \subseteq \binom{[n]}{s_i}$ such that $\mathbf{v} \in P_S^{(i)}$. By construction of $P_S^{(i)}$, there exists some $\mathbf{x}_\mathbf{v} \in \mathbb{R}^n$ such that

$$\begin{aligned} &\mathbf{d}_\mathbf{v} = B^{(i)}\mathbf{x}_\mathbf{v}, t_\mathbf{v} \geqslant f^{(i)}(\mathbf{x}_\mathbf{v}), \\ &(\mathbf{x}_\mathbf{v})_i = 0, \forall i \notin S. \end{aligned}$$

Since $\mathbf{v}$ is a extreme point, we can further assume that $t_\mathbf{v} = f^{(i)}(\mathbf{x})$. Consider the following function $f_S^{(i)}(\mathbf{y}) : \mathbb{R}^S \to \mathbb{R}$ which is constructed from $f^{(i)}(\mathbf{x})$ by fixing the variables outside of $S$ to be zero. Clearly, $f_S^{(i)}$ is a $L$-smooth convex function. Applying Claim 6, there is a ball $B_S$ with radius $\frac{1}{L}$ with dimension $|S| + 1$ that $((\mathbf{x}_\mathbf{v})_S, f(\mathbf{x}_\mathbf{v})) \subseteq B_S \subseteq \text{epi}\, f_S^{(i)}$. Now consider the linear map:

$$(\mathbf{x}_S, t) \to \begin{bmatrix} \mathbf{1}^\top \\ B_S^{(i)} \end{bmatrix} (\mathbf{x}_S, t)$$

Let $Q := \begin{bmatrix} \mathbf{1}^\top \\ B_S^{(i)} \end{bmatrix}$. It is well known that a linear map transforms a ball into an ellipse. In particular, $B_S$ is mapped to an ellipse $E$ (up to translation) with form: $\left\{ \mathbf{y} : \mathbf{y}^\top (QQ^\top)^{-1}\mathbf{y} \leqslant (\frac{1}{L})^2 \right\}$. This ellipse includes a ball with radius $\frac{\sigma_{\inf}(QQ^\top)}{L}$ and by the definition of projection factor, this ellipse includes a ball with radius $\frac{\mathcal{L}_s(B^{(i)})}{L}$. Since $E \subseteq \mathcal{P}_S^{(i)} \subseteq \mathcal{P}^{(i)}$, this implies that there exists a ball with the desired statement. □

**Theorem 5** *Consider problem (4), assume that each $f^{(i)}(\cdot)$ is $L$-smooth and under the Assumption 1, Assumption 4 and Assumption 5 that there exists some $\beta$ such that $\Xi(\mathcal{P}^{(i)}) \leqslant \beta$. Then it follow that*

$$\begin{aligned} \mathcal{E} := \Phi\left(\sum_{i=1}^{k} \mathcal{P}^{(i)}\right) &\leqslant \sqrt{\mathcal{L}^2 \frac{1}{L^2} + \frac{1}{2}(m+1)\beta^2} - \mathcal{L}^2 \frac{1}{L^2}, \\ \text{OPT}_s(\mathbf{b} + \mathcal{E}\mathbf{1}) - \mathcal{E} &\leqslant \text{DUAL}_s(\mathbf{b}) \leqslant \text{OPT}_s(\mathbf{b}), \end{aligned}$$

*where $\mathcal{L} = \inf_{\mathcal{I} \subseteq [k] : |\mathcal{I}| = m+1} \sum_{i \notin \mathcal{I}} \mathcal{L}(B^{(i)})$. If there exists some $\omega$ such that $\mathcal{L}(B^{(i)}) \geqslant \omega, \forall i \in [k]$, then it follows that*

$$\begin{aligned} \mathcal{L} &\to \infty \text{ as } k \to \infty, \\ \mathcal{E} &\to 0 \text{ as } k \to \infty. \end{aligned}$$

*Proof* Using Claim 7 and Theorem 4, we obtain

$$\mathcal{E} = \Phi\left(\sum_{i=1}^{k} \mathcal{P}^{(i)}\right) \leqslant \sqrt{\mathcal{L}^2 \frac{1}{L^2} + \frac{1}{2}(m+1)\beta^2} - \mathcal{L}^2 \frac{1}{L^2}.$$

On the other hand, if there exists some $\omega$ such that $\mathcal{L}(B^{(i)}) \geqslant \omega, \forall i \in [k]$, then $\mathcal{L} \geqslant (k-m-1)\omega \to \infty$ as $k \to \infty$ and therefore $\mathcal{E} \to 0$. □

**Remark 5** The smoothness assumption in Theorem 5 can be relaxed. Instead of requiring that $f$ be globally smooth—i.e., globally upper bounded by a convex quadratic function—it suffices to assume that $f$ is locally upper bounded by a convex quadratic function at each point $\mathbf{x}$. In particular, as long as this local smoothness condition guarantees the existence of an $l_2$ -ball with nontrivial radius, the conclusion of Theorem 5 remains valid.

# Acknowledgments

We would also like to thank Diego Cifuentes and Greg Blekherman for the various wonderful discussions.

## References


[1] Bagirov, A., Karmitsa, N., Mäkelä, M.M.: Introduction to Nonsmooth Optimization: Theory, Practice and Software vol. 12. Springer, New York, USA (2014)

[2] Frangioni, A.: Standard bundle methods: Untrusted models and duality. Numerical nonsmooth optimization: state of the art algorithms, 61–116 (2020)

[3] Van Ackooij, W., Frangioni, A.: Incremental bundle methods using upper models. SIAM Journal on Optimization **28**(1), 379–410 (2018)

[4] Lübbecke, M.E.: Column generation. Wiley encyclopedia of operations research and management science **17**, 18–19 (2010)

[5] Desaulniers, G., Desrosiers, J., Solomon, M.M.: Column Generation vol. 5. Springer, New York, USA (2006)

[6] Guo, L., Ye, J.J., Zhang, J.: Sensitivity analysis of the maximal value function with applications in nonconvex minimax programs. Mathematics of Operations Research **49**(1), 536–556 (2024)

[7] Disfani, V.R., Fan, L., Piyasinghe, L., Miao, Z.: Multi-agent control of community and utility using lagrangian relaxation based dual decomposition. Electric Power Systems Research **110**, 45–54 (2014)

[8] Yang, Z., Zheng, T., Yu, J., Xie, K.: A unified approach to pricing under nonconvexity. IEEE Transactions on Power Systems **34**(5), 3417–3427 (2019)

[9] Starr, R.M.: Quasi-equilibria in markets with non-convex preferences. Econometrica: journal of the Econometric Society, 25–38 (1969)

[10] Balseiro, S., Lu, H., Mirrokni, V.: Dual mirror descent for online allocation problems. In: International Conference on Machine Learning, pp. 613–628 (2020). PMLR

[11] Argue, C., Gupta, A., Molinaro, M., Singla, S.: Robust secretary and prophet algorithms for packing integer programs. In: Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1273–1297 (2022). SIAM

[12] Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. SIAM, Philadelphia, USA (2001)

[13] Pólik, I., Terlaky, T.: A survey of the s-lemma. SIAM review **49**(3), 371–418 (2007)

[14] Chandrasekaran, V., Duff, T., Rodriguez, J.I., Shu, K.: Lagrangian dual sections:

A topological perspective on hidden convexity. arXiv preprint arXiv:2510.06112 (2025)

[15] Dines, L.L.: On the mapping of quadratic forms. Bulletin of the American Mathematical Society **47**(6), 494–498 (1941)

[16] Fradelizi, M., Madiman, M., Marsiglietti, A., Zvavitch, A.: The convexification effect of minkowski summation. EMS Surveys in Mathematical Sciences **5**(1), 1–64 (2018)

[17] Bertsekas, D., Lauer, G., Sandell, N., Posbergh, T.: Optimal short-term scheduling of large-scale power systems. IEEE Transactions on Automatic Control **28**(1), 1–11 (1983)

[18] Aubin, J.-P., Ekeland, I.: Estimates of the duality gap in nonconvex optimization. Mathematics of Operations Research **1**(3), 225–245 (1976)

[19] Udell, M., Boyd, S.: Bounding duality gap for separable problems with linear constraints. Computational Optimization and Applications **64**(2), 355–378 (2016)

[20] Bi, Y., Tang, A.: Duality gap estimation via a refined shapley–folkman lemma. SIAM Journal on Optimization **30**(2), 1094–1118 (2020)

[21] Kerdreux, T., Colin, I., d'Aspremont, A.: Stable bounds on the duality gap of separable nonconvex optimization problems. Mathematics of Operations Research **48**(2), 1044–1065 (2023)

[22] Nguyen, T., Vohra, R.: (near) substitute preferences and equilibria with indivisibilities," working paper (2022)

[23] Murota, K., Tamura, A.: Shapley–folkman-type theorem for integrally convex sets. Discrete Applied Mathematics **360**, 42–50 (2025)

[24] Wegmann, R.: Einige maßzahlen für nichtkonvexe mengen. Archiv der Mathematik **34**(1), 69–74 (1980)

[25] Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex analysis and minimization algorithms ii. Grundlehren der mathematischen Wissenschaften (1993)

[26] Boland, N., Christiansen, J., Dandurand, B., Eberhard, A., Linderoth, J., Luedtke, J., Oliveira, F.: Combining progressive hedging with a frank–wolfe method to compute lagrangian dual bounds in stochastic mixed-integer programming. SIAM Journal on Optimization **28**(2), 1312–1336 (2018)

[27] Geoffrion, A.M.: Lagrangean relaxation for integer programming. In: Approaches to Integer Programming, pp. 82–114. Springer, Amsterdam, Netherlands (2009)

[28] Hiriart-Urruty, J.-B., Lemaréchal, C.: Fundamentals of Convex Analysis. Springer, New York, USA (2004)

[29] Dey, S.S., Meunier, F., Ramirez, D.M.: Geoffrion's theorem beyond finiteness and rationality. arXiv preprint arXiv:2510.10966 (2025)

[30] Lemaréchal, C., Renaud, A.: A geometric study of duality gaps, with applications. Mathematical Programming **90**(3), 399–427 (2001)

[31] Molinaro, M.: Strong convexity of feasible sets in off-line and online optimization. Mathematics of Operations Research **48**(2), 865–884 (2023)