

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

Santanu S. Dey¹ and Jingye Xu^{1*}

¹H. 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

The Shapley–Folkman Lemma, a foundational result in convex geometry, asserts that the nonconvexity of a Minkowski sum of n -dimensional bounded nonconvex sets does not accumulate once the number of summands exceeds the dimension n , and thus the sum becomes approximately convex. This Lemma, originally published by Starr is used to show quasi-equilibrium in nonconvex market models in economics and equivalently used to estimate 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 n -dimensional nonconvex sets to vanish asymptotically, rather than merely converge to a constant, as the number of summands increases? We answer this question 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 under a pointedness assumption. This leads to a small refinement on the classical error bound derived from the Lemma in the pointed case. Building on this new geometric perspective, we next show that if the sets satisfy a certain local smoothness condition, which naturally arises, for example, in the epigraphs of smooth functions, then their Minkowski sums converge directly to a convex set as the nonconvexity measure vanishes asymptotically. This result has direct implications for optimization and economic theory. In particular, we show that the Lagrangian dual of block-structured smooth nonconvex problems, possibly subject to additional sparsity constraints, is asymptotically tight under mild assumptions. This sharply improves upon classical Shapley–Folkman based bounds, which generally predict a nonvanishing duality gap. Equivalently, from an economic perspective, this result implies that, asymptotically as the size of the

economy increases, competitive equilibrium exists with smooth utility functions, notwithstanding underlying nonconvexities.

Keywords: Shapley–Folkman Lemma, Strong Lagrangian Dual, Competitive Equilibrium

Mathematics Subject Classification: 90C11 · 90C46

1 Introduction

Consider a block-structured nonconvex program:

$$\text{OPT}(\mathbf{b}) := \inf \sum_{i=1}^k \langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \rangle \quad (1a)$$

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

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

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) are linear, this formulation is general enough to capture many nonlinearities, by embedding these into the feasible sets $\mathcal{X}^{(i)}$, potentially through the introduction of suitable auxiliary variables.

An important relaxation of (1) is the Lagrangian dual of (1), obtained by dualizing the coupling constraints (1b) across various blocks:

$$L(\lambda) := \inf_{\mathbf{x}} \left(\sum_{i=1}^k \langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \rangle + \langle \lambda, B^{(i)} \mathbf{x}^{(i)} \rangle \right) - \langle \lambda, \mathbf{b} \rangle \quad (2)$$

$$\text{s.t. } \mathbf{x}^{(i)} \in \mathcal{X}^{(i)} \subseteq \mathbb{R}^n, \forall i \in [k],$$

and

$$\text{DUAL}(\mathbf{b}) := \sup_{\lambda \geq 0} L(\lambda). \quad (3)$$

The Lagrangian dual of the block-structured problem (1) plays a central role in both optimization and economic theory. From an optimization perspective, dualizing the coupling constraints yields a convex (possibly nonsmooth) program (3) regardless of the nonconvexity of the feasible sets $\mathcal{X}^{(i)}$, thereby providing a tractable and globally valid lower bound on $\text{OPT}(\mathbf{b})$. Moreover, the dual formulation exploits the separable block structure of the primal problem: for any fixed multiplier $\lambda \geq 0$, the Lagrangian $L(\lambda)$ decomposes into independent subproblems across blocks, enabling scalable algorithms based on decomposition, subgradient methods, and price-based coordination [1–3]. From an economic viewpoint, the dual variables λ admit a natural interpretation as equilibrium prices associated with the shared resource constraint $\sum_{i=1}^k B^{(i)} x^{(i)} \leq \mathbf{b}$

[4]. In this interpretation, each block represents an individual agent or firm that optimizes its own objective given prices λ , while the dual objective aggregates these decentralized decisions and enforces market clearing. Consequently, the duality gap between $\text{OPT}(\mathbf{b})$ and $\text{DUAL}(\mathbf{b})$ quantifies the extent to which a competitive equilibrium exists or fails to exist in the presence of nonconvexities, making the Lagrangian framework a unifying bridge between large-scale optimization and equilibrium analysis in economic theory.

One of the most desirable properties one aims to establish for the Lagrangian dual, in both optimization and economic theory, is *strong duality*. In optimization, strong duality certifies that the optimal value of a (possibly relaxed) dual problem coincides with that of the primal, thereby justifying convex relaxations, decomposition methods, and price-based algorithms. In economic theory, strong duality corresponds to the existence of supporting prices that decentralize a socially optimal allocation, that is, establishing the existence of a competitive equilibrium.

Many textbooks state that strong duality, that is $\text{OPT}(\mathbf{b}) = \text{DUAL}(\mathbf{b})$, holds under certain regularity conditions when all $\mathcal{X}^{(i)}$'s are convex. However, strong duality actually only requires the convexity of linear image of (1) onto the space of objective and constraints [5]. More specifically, let

$$\begin{aligned}\mathcal{P}^{(i)} &:= \{(t, \mathbf{d}) \mid \exists \mathbf{x} \in \mathcal{X}^{(i)}, t = \langle \mathbf{c}^{(i)}, \mathbf{x}^{(i)} \rangle, \mathbf{d} = B^{(i)}\mathbf{x}^{(i)}\}, \\ \mathcal{P} &:= \sum_{i=1}^k \mathcal{P}^{(i)},\end{aligned}$$

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 [5]. 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 textbooks 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 [6, 7]. A notable example is the celebrated *S-Lemma* [6], 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 [8] 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* [4]. In the setting where $k \geq n$, the Shapley–Folkman Lemma states that

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

Here $[k] = \{1, \dots, k\}$. 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 $\mathcal{P}^{(i)}$ s are contained within a ball of radius R , then a quantitative measure of this approximate convexity can be obtained. In particular, under the Hausdorff distance with l_2 norm, it has been shown [9] that

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

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

The Shapley–Folkman lemma was originally introduced by Starr [4] 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 duals. Its connection to optimization was later developed in works such as [10–12], 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, [13–15]. An analog of the Shapley–Folkman lemma for discrete convex sets has also been developed in [16].

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 $\mathbf{conv}(\mathcal{P}^{(i)})$. This perspective enables us to refine the classical error bound under a pointedness assumption (see Theorem 4 in Section 4). Building on this geometric insight, we further demonstrate that the Minkowski sum of the sets $\mathcal{P}^{(i)}$ can directly converge to its convex hull as the number of summands grows, i.e., $d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P})) \rightarrow 0$ as $k \rightarrow \infty$, provided enough sets satisfy some mild ‘local smoothness’ condition (see Theorem 5 in Section 4). To the best of our knowledge, this is the first result establishing an asymptotically vanishing nonconvexity measure for Minkowski sums of nonconvex sets within the framework of the Shapley–Folkman Lemma.

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 $k \rightarrow \infty$:

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

where each $f^{(i)}$ is a smooth (possibly nonconvex) function, $B^{(i)}$'s satisfy some mild conditions, and $s^{(i)}$ is not too small, i.e., $s^{(i)} \geq m, \forall i \in [k]$ where $B^{(i)}$ has m rows (see Theorem 6 in Section 5). To model problems without sparsity constraints, one may simply drop these constraints, that is let $s^{(i)} = \text{number of variables in } \mathbf{x}^i$, for all $i \in [k]$.

Finally, we note here that this work focuses on Minkowski sum of $\mathcal{P}^{(i)}$'s 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 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

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

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.

The remainder of the paper is organized as follows. In Section 2, we present standard notation and preliminary results. Section 3 establishes a connection between $d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P}))$ and $\Delta := \text{OPT}(\mathbf{b}) - \text{DUAL}(\mathbf{b})$ in Theorem 3. Section 4 presents a new proof of Shapley–Folkman Lemma under the assumption of pointedness of the convex hull of the non-convex sets (see Lemma 4 and discussion following it) which allows us to establish a slightly better bound on $d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P}))$, in Theorem 4. Our main result of showing that $d_H(\mathcal{P}, \mathbf{conv}(\mathcal{P}))$ goes to zero as k goes to ∞ under the presence of a local smoothness property, is presented in this section in Theorem 5. Finally, in Section 5, we establish the strong duality result stated informally above, formalized as Theorem 6.

2 Preliminary Results and Notation

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

2.1 Notation

For a set $A \subset \mathbb{R}^d$, we define its *radius* as $R(A) := \inf_{x \in \mathbb{R}^d} \{r \geq 0 : A \subseteq \mathcal{B}(x, r)\}$, where $\mathcal{B}(x, r) := \{y \in \mathbb{R}^d : \|y - x\|_2 \leq r\}$ 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 $\lambda \geq 0$, the *Minkowski scaling* of A is $\lambda A := \{\lambda a : a \in A\}$. We denote by $\mathbf{conv}(A) := \{\sum_{i=1}^m \lambda_i a_i \mid a_i \in A, \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1\}$ the convex hull of A . The *closed convex hull* is defined as $\overline{\mathbf{conv}}(A) := \text{cl}(\mathbf{conv}(A))$, where $\text{cl}(A)$ denotes the topological closure of A . The relative boundary and the recession cone of a closed convex set C are denoted using $\text{relbd}(C)$ and $\text{rec}(C)$. The kernel of a linear map \mathcal{A} is denoted by $\ker(\mathcal{A})$. We use $\text{epi}(f)$ to denote the epigraph of a function f .

2.2 Nonconvexity measure and convex hull operator

For a given set $A \subseteq \mathbb{R}^n$, there 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

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

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

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

2. The inner radius of a nonconvex set [4]:

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

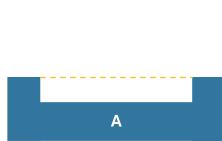


Fig. 1: A and $\mathbf{conv}(A)$

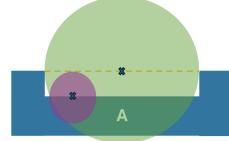


Fig. 2: two nonconvexity measures where pink ball corresponds to $\Phi(A)$ and green ball corresponds to $\Xi(A)$.

The measure $\Xi(\cdot)$, introduced in the seminal work by Starr [4], 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). See Figure 1 and 2 for illustration of the $\Phi(\cdot)$ and $\Xi(\cdot)$ nonconvexity measure. 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 [9]. We present next a standard result on these non-convexity measures.

Lemma 1 [9, 17] Let A be a subset of \mathbb{R}^n , then

$$\Phi(A) \leq \Xi(A) \leq \mathbf{R}(A).$$

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.

Lemma 2 [9] (**conv** commutes with Minkowski sum) Let A_1, \dots, A_k be nonempty sets of \mathbb{R}^n . Then it follows that

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

Lemma 3 [18] (**conv** commutes with linear projection)

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

2.3 Shapley-Folkman Lemma

Theorem 1 (*Shapley-Folkman Lemma*) Let A_1, \dots, A_k be nonempty closed sets of \mathbb{R}^n with $k \geq n + 1$. 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 [13, 14] and we defer our geometric proof in Lemma 4, presented in Section 4.

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. 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 A_1, \dots, A_k be nonempty sets of \mathbb{R}^n with $k \geq n$ such that $\Xi(A_i) \leq \beta, \forall i \in [k]$. Then it follows that

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

If we further assume $R(A_i) \leq \gamma, \forall i \in [k]$, then it follows that

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

3 Shapley-Folkman Lemma and duality gap of Lagrangian relaxation

A key quantity associated with the Lagrangian dual that reflects the tightness of the relaxation is the duality gap (also called the additive gap) which is defined as:

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

In this section, we present a result showing how to upper bound Δ based on upper bounds on $\Phi(\mathcal{P})$.

To achieve this, we first use the following primal characterization [19, 20] of Lagrangian dual (bi-conjugacy in convex analysis [21]) to reduce $\text{DUAL}(\mathbf{b})$ from min-max optimization to a minimization problem in the original space. Consider

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

Theorem 2 [20, 21] *Under proper regularity condition, $\text{OPT}_{\mathbf{L}}(\mathbf{b}) = \text{DUAL}(\mathbf{b})$.*

Common regularity conditions required for Theorem 2 include the existence of a Slater point or the assumption that each $\text{conv}(\mathcal{X}^{(i)})$ is a polyhedron [22, 23]. The absence of such regularity may result in cases where $\text{OPT}_{\mathbf{L}}(\mathbf{b}) > \text{DUAL}(\mathbf{b})$ [23]. Throughout this paper, we exclude these pathological instances and directly assume that the following equivalence holds:

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

We finally present the main result of this Section showing that the tightness of Lagrangian dual of (1) directly relies on the convexity of $\mathcal{P} = \sum_{i=1}^k \mathcal{P}^{(i)}$ by quantifying Δ in terms of $\Phi(\mathcal{P})$. This theorem is similar to the gap formula in [24].

Theorem 3 *Let $\mathcal{E} := \Phi(\mathcal{P})$. Under Assumption 1 and $\text{conv}(\mathcal{X}^{(i)})$ being closed for all $i \in [k]$, it follows that*

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

and therefore

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

Proof Under Assumption 1, we know that $\text{DUAL}(\mathbf{b}) = \text{OPT}_{\mathbf{L}}(\mathbf{b})$. Let $\mathbf{x}_{\mathbf{L}}$ be the ϵ -optimal solution of (5). By weak duality, we know that

$$\text{OPT}_{\mathbf{L}}(\mathbf{b}) + \epsilon = \sum_{i=1}^k \left\langle \mathbf{c}^{(i)}, \mathbf{x}_{\mathbf{L}}^{(i)} \right\rangle = \text{DUAL}(\mathbf{b}) + \epsilon \leq \text{OPT}(\mathbf{b}) + \epsilon. \quad (6)$$

Let $\mathbf{p}_{\mathbf{L}}^{(i)} := \begin{bmatrix} \left\langle \mathbf{c}^{(i)}, \mathbf{x}_{\mathbf{L}}^{(i)} \right\rangle \\ B^{(i)} \mathbf{x}_{\mathbf{L}}^{(i)} \end{bmatrix} \in \mathbf{conv}(\mathcal{P}^{(i)})$ where the last containment is due to the fact that $\mathbf{conv}(\mathcal{X}^{(i)})$ is closed (which implies $\mathbf{x}_{\mathbf{L}}^{(i)} \in \mathbf{conv}(\mathcal{X}^{(i)})$) and using Lemma 3. By Lemma 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

$$\left\| \left(\sum_{i=1}^k \mathbf{p}_{\mathbf{L}}^{(i)} \right) - \left(\sum_{i=1}^k \mathbf{p}_*^{(i)} \right) \right\|_2 \leq \mathcal{E}. \quad (7)$$

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

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

where the last inequality follows from (7). Now, using (6) with the above inequality, we obtain that

$$\begin{aligned} \text{OPT}(\mathbf{b} + \mathcal{E}\mathbf{1}) - \mathcal{E} &\leq \text{DUAL}(\mathbf{b}) + \epsilon \\ \Rightarrow \Delta &\leq \mathcal{E} + \text{OPT}(\mathbf{b}) + \epsilon - \text{OPT}(\mathbf{b} + \mathcal{E}\mathbf{1}), \end{aligned}$$

where the second inequality is obtained by subtracting $\text{OPT}(\mathbf{b})$ on both sides. Letting $\epsilon \rightarrow 0$ yields the desired statement. \square

4 Error bounds via Shapley–Folkman Lemma

In this section (and for the rest of the paper), we make an additional assumption:

Assumption 2 Let $A_i \subseteq \mathbb{R}^n$, $i \in [k]$ be the sets that are the summands in the Shapley–Folkman lemma. For all $i \in [k]$, we assume that $\mathbf{conv}(A_i)$ is closed and pointed.

Assumption 2 guarantees that each $\mathbf{conv}(A_i)$ has at least one extreme point. Moreover, since $\mathbf{conv}(A_i)$ is closed, any such extreme point must lie in the original set A_i . The pointedness assumption is subsequently used in Lemma 4 in the next section 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.

4.1 Geometric proof of Shapley-Folkman Lemma

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

For completeness, we begin by presenting the standard definition of a face of a convex set and related notions.

Definition 1 Given a closed convex set C , we call a closed convex set $F \subseteq C$ is a **face** of C if for any line segment $[a, b] \subseteq C$ such that $(a, b) \cap F \neq \emptyset$, we have $[a, b] \subseteq F$. The dimension of face F denoted by $\dim(F)$ is the dimension of its affine hull.

Note that a convex set C is a face of itself. Moreover a face of a face of C is also a face of C . Finally, faces of closed convex sets are closed convex sets [21].

We begin some simple observations from linear algebra and convex analysis. The first claim directly follows from a rank-nullity argument.

Claim 1 Let $L_i \in \mathbb{R}^n$, $i \in [k]$ be linear subspaces, such that $\sum_{i=1}^k \dim(L_i) > n$. Then there exists $d_i \in L_i$ for $i \in [k]$, with not all d_i s zero, such that $\sum_{i=1}^k d_i = 0$.

Claim 2 Let $C \subseteq \mathbb{R}^n$ be a closed convex set, and let F be a nonempty face of C . If $x \in \text{relbd}(F)$, then x is contained in a face of C of strictly smaller dimension than F .

Proof Since $x \in \text{relbd}(F)$, there exists a hyperplane H in the affine hull of F such that (i) F is contained in one of the half-spaces defined by H , (ii) $x \in H$, and (iii) $F \not\subseteq H$. Therefore, by definition of a face, $F \cap H$ is a face of F , and thus a face of C . Moreover, (iii) implies that the dimension of F is at least 1 more than the dimension of $F \cap H$. \square

Lemma 4 Let A_1, \dots, A_k be nonempty sets of \mathbb{R}^n with $k \geq n + 1$ such that $\text{conv}(A_i)$ is closed and pointed. Then for every $\mathbf{y} \in \text{conv}\left(\sum_{i=1}^k A_i\right)$, there exist faces F_i of $\text{conv}(A_i)$ such that

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

Proof For any $\mathbf{y} \in \text{conv}\left(\sum_{i=1}^k A_i\right)$, 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 $\text{conv}(A_i)$. If $\sum_{i=1}^k \dim(F_i) > n$, we show how to find a different \mathbf{y}'_i from $\text{conv}(A_i)$ so that the total dimension $\sum_{i=1}^k \dim(F'_i)$, where F'_i is a face of $\text{conv}(A_i)$ containing \mathbf{y}'_i , decreases while ensuring $\mathbf{y} = \sum \mathbf{y}'_i$. Let G_i be the linear subspace obtained by translating the affine hull of F_i to the origin. If $\sum_{i=1}^k \dim(G_i) = \sum_{i=1}^k \dim(F_i) > n$, applying Claim 1, we can find $d_i \in G_i$, not all zero, such that $\mathbf{d}_1 + \mathbf{d}_2 + \dots + \mathbf{d}_k = \mathbf{0}$. Let $\mathbf{y}'_i(\lambda) := \mathbf{y}_i + \lambda d_i$ for $i \in [k]$. Since each $\text{conv}(A_i)$ is pointed, there exists some choice of λ such that for at least one $i^* \in [k]$ we have that $\mathbf{y}'_{i^*}(\lambda)$ belongs to the relative boundary of F_{i^*} while keeping other $\mathbf{y}'_j(\lambda)$ in F_j for all $j \in [k]$. By Claim 2, $\mathbf{y}'_{i^*}(\lambda) \in F'_{i^*}$, a face of

$\mathbf{conv}(A_i)$ such that $\dim(F'_{i*}) < \dim(F_{i*})$. Letting $F'_j = F_j$ for all $j \neq i*$, the total dimension $\sum_{i=1}^k \dim(F'_i)$ decreases and we still ensure that $\mathbf{y} = \sum_{i=1}^k \mathbf{y}'_i$ for some $\mathbf{y}'_i \in \mathbf{conv}(A_i)$. \square

Proof of Shapley–Folkman Lemma using Lemma 4 For every $\mathbf{y} \in \mathbf{conv}\left(\sum_{i=1}^k A_i\right)$, Lemma 4 implies that there exist faces F_i of $\mathbf{conv}(A_i)$ for $i \in [k]$, such that $\mathbf{y} \in \sum_{i=1}^k F_i$ and $\sum_{i=1}^k \dim(F_i) \leq n$. Since $\dim(F_i) \in \{0, 1, \dots, n\}$ and $\sum_{i=1}^k \dim(F_i) \leq n$, it implies that at least $k-n$ faces have dimension 0. These dimension-0 faces are extreme points of corresponding $\mathbf{conv}(A_i)$ s, therefore must lie in A_i . Let $\mathcal{I} = \{i \in [k] : \dim(F_i) > 0\}$ where from the above we have that $|\mathcal{I}| \leq n$. Moreover, we have,

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

\square

We note here that Lemma 4 can be used to strengthen the refined Shapley–Folkman Lemma recently proposed in [13].

4.2 Error bounds via randomized rounding

Let A_1, A_2, \dots, A_k be nonempty set of \mathbb{R}^n such that $\mathbf{conv}(A_i)$ is closed and pointed. For any $\mathbf{y} \in \mathbf{conv}\left(\sum_{i=1}^k A_i\right)$, Lemma 4 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, $\|\mathbf{y} - \mathbf{y}'\|_2 \leq \Phi\left(\sum_{i=1}^k A_i\right)$.

The central idea of this section is that a suitable \mathbf{y}' can be constructed by randomly sampling points from each F_i . We first establish some preliminary results that facilitate this sampling approach.

Claim 3 Let C be a closed convex set and let F be a face of C . For any $\mathbf{x} \in F$, suppose \mathbf{x} is a strict convex combination of \mathbf{x}_i s, that is, $\mathbf{x} = \sum_i \lambda_i \mathbf{x}_i$, with $\lambda_i > 0$ and $\sum_i \lambda_i = 1$. If $\mathbf{x}_i \in C$ for all i , then $\mathbf{x}_i \in F$ for all i .

Proof Directly follows from the definition of a face. \square

Claim 4 Let A be a nonempty set of \mathbb{R}^n . If $\Xi(A) < \infty$, then for every face F of $\mathbf{conv}(A)$, $\Xi(A \cap F) \leq \Xi(A)$.

Proof Consider any $\mathbf{x} \in F \subseteq \mathbf{conv}(A)$, by the definition of $\Xi(A)$, there exists $T \subseteq A$ such that $R(T) \leq \Xi(A)$ 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 3 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 implies that $\Xi(A \cap F) \leq \Xi(A)$. \square

The above claim is interesting because, in general $\Xi(A \cap B)$ may not be smaller than $\Xi(A)$ or $\Xi(B)$.

Lemma 5 Let $A \subseteq \mathbb{R}^n$. 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 $\mathbb{E}(\|Z - \mathbb{E}(Z)\|_2^2) \leq \Xi^2(A)$.

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 $R(T) \leq \Xi(A)$. Since the statement of the Lemma is invariant under translation, we may translate A so that $\|\mathbf{y}\|_2 \leq \Xi(A)$ for all $\mathbf{y} \in T$. By Carathéodory's theorem, \mathbf{x} can be written a convex combination of \mathbf{p}_i and each $\mathbf{p}_i \in T$, that is $\mathbf{x} = \sum_{i=1}^{n+1} \lambda_i \mathbf{p}_i$ where $\mathbf{p}_i \in T$ and λ belongs to the $n+1$ dimensional standard simplex. We define random variable $Z \sim \mathcal{D}$ that $Z = \mathbf{p}_i$ with probability λ_i . It is clear that $\mathbb{E}_{Z \sim \mathcal{D}}(Z) = \mathbf{x}$ and

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

where the last line uses that fact that Z only has support over T and $\|\mathbf{y}\|_2 \leq \Xi(A)$ for all $\mathbf{y} \in T$. \square

Lemma 6 Let A_1, A_2, \dots, A_k be nonempty subsets of \mathbb{R}^n . Let F_i be a face of $\mathbf{conv}(A_i)$ for all $i \in [k]$ and let $\mathcal{I} := \{i \in [k] : \dim(F_i) > 0\}$. Then it follows that

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

Proof For every $\mathbf{x} \in \mathbf{conv}\left(\sum_{i=1}^k A_i \cap F_i\right) = \sum_{i=1}^k \mathbf{conv}(A_i \cap F_i)$ (the equality follows from Lemma 2), there exists 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 $\mathbf{y}_i \in A_i \cap F_i$ and $\left\|\left(\sum_{i=1}^k \mathbf{x}_i\right) - \left(\sum_{i=1}^k \mathbf{y}_i\right)\right\|_2 \leq \sqrt{\sum_{i \in \mathcal{I}} \Xi^2(A_i \cap F_i)}$.

For any $i \notin \mathcal{I}$, $\dim(F_i) = 0$ and therefore \mathbf{x}_i is an extreme point of $\mathbf{conv}(A_i)$, so $\mathbf{x}_i \in A_i \cap F_i$. For $i \in \mathcal{I}$, by Lemma 5, 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 $A_i \cap F_i$ and $\mathbb{E}(\|Z_i - \mathbf{y}_i\|_2^2) \leq \Xi^2(A_i \cap F_i)$. Then we (randomly) construct $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 + \dots + \mathbf{y}_k$ in the following way:

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

Therefore, it follows that

$$\begin{aligned} \mathbb{E}(\|\mathbf{x} - \mathbf{y}\|_2^2) &= \mathbb{E}\left(\left\|\sum_{i \in \mathcal{I}} (\mathbf{x}_i - Z_i)\right\|_2^2\right) \\ &= \sum_{i \in \mathcal{I}} \mathbb{E}(\|(\mathbf{x}_i - Z_i)\|_2^2) \\ &\leq \sum_{i \in \mathcal{I}} \Xi^2(A_i \cap F_i). \end{aligned}$$

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

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

Since \mathbf{x} is arbitrary and by Claim 4, this implies that $\Phi\left(\sum_{i=1}^k A_i \cap F_i\right) \leq \sqrt{\sum_{i \in \mathcal{I}} \Xi^2(A_i \cap F_i)} \leq \sqrt{\sum_{i \in \mathcal{I}} \Xi^2(A_i)}$. \square

Applying Lemma 4 and Lemma 6 directly yields the following theorem.

Theorem 4 *Under conditions that each $\mathbf{conv}(A_i)$ is closed and pointed for all $i \in [k]$, we have that,*

$$\Phi\left(\sum_{i=1}^k A_i\right) \leq \sup \left\{ \sqrt{\sum_{i=1}^k \Xi^2(A_i \cap F_i)} : F_i \text{ is a face of } A_i \text{ and } \sum_{i=1}^k \dim(F_i) = n \right\}.$$

Theorem 4 is a refinement on the classical error bounds, due to the fact that $\Xi(A_i \cap F_i)$ can be significantly smaller than $\Xi(A_i)$ unless F_i is a high-dimensional face. However, due to the global dimension-sum constraints, the number of such high-dimensional faces is limited.

A more concrete illustration is as follows. Consider any full-dimensional convex body $C \in \mathbb{R}^n$ with $R(C) = R$, and define $A_i = C \setminus \text{int}(C)$ for all $i \in [k]$. In this case, every face F_i of $\mathbf{conv}(A_i)$ that is not $\mathbf{conv}(A_i)$ itself, satisfies $\Xi(A_i \cap F_i) = 0$ because each $F_i \cap A_i$ is a convex set. Therefore, $\Xi(A_i \cap F_i) \leq R$, where $\Xi(A_i \cap F_i) \neq 0$, only when $F_i = \mathbf{conv}(A_i)$ and in this case $\dim(F_i) = n$. Now applying Theorem 4 implies that $\Phi\left(\sum_{i=1}^k A_i\right) \leq \Xi(A_1) \leq R$, which is significantly smaller than the $\sqrt{n}R$ bound predicted by the classical error estimates derived from the Shapley–Folkman lemma.

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

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

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 4 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 6 can be interpreted as follows: we attempt to round \mathbf{y} as a whole to a point in $\sum_{i=1}^k (A_i \cap F_i)$. This perspective naturally suggests a potential further refinement of Lemma 6, wherein we aim to round \mathbf{y} directly to a point in $\sum_{i=1}^k A_i$. However, this appealing idea encounters a significant challenge, as it requires full characterization of the nonconvex sets A_i , which is typically unavailable in practice.

The key insight of our main result is the following: We do not need to understand the global structure of the nonconvex sets A_i . It is sufficient to instead focus on exploiting only the local geometric properties of each A_i .

We next introduce a definition that plays a key role in identifying the required local geometry.

Definition 2 Let A be a 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. $Q \subseteq H$;

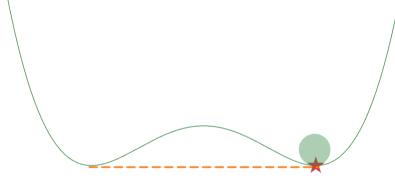


Fig. 3: A is the epigraph of the function $y = x^4 - x^2$. The star point is the set Q . The l_2 ball contained in the epigraph is a hidden convex component.

The definition above is illustrated via an example in Figure 3.

Such hidden convex component may not exist 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 $\{0, 1, \dots, k\}^n$, 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 discussed in Remark 2 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 an appropriate smoothness condition of the hidden convex component H_i s.

Definition 3 [5] A differentiable function $f(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ is called L -smooth if $f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2, \forall \mathbf{y} \in \mathbb{R}^n$, where $\nabla f(\cdot)$ is the gradient.

Definition 4 [21] Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set with 0 with its interior. The *gauge function* associated with C is

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

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 [25].

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, we note that $\Phi^K(A, D)$ is invariant under translation of A and D by the same vector.

Lemma 7 Let A_1, A_2, \dots, A_k be nonempty subsets of \mathbb{R}^n and $\Xi(A_i) \leq \beta, \forall i \in [k]$. Let F_i be a face of $\mathbf{conv}(A_i)$, for all $i \in [k]$ and let $\mathcal{I} := \{i \in [k] : \dim(F_i) > 0\}$. Let H_i be a hidden convex component associated with F_i for all $i \in [k] \setminus \mathcal{I}$ and $H = \sum_{i \in [k] \setminus \mathcal{I}} H_i$. Suppose H has

nonempty interior, let \mathcal{H} be a translation of H that contains $\mathbf{0}$ in its interior. If we have $\|\cdot\|_{\mathcal{H}}^2$ is L -smooth, then

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

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 $\Phi^{\mathcal{H}}\left(\sum_{i=1}^k F_i, \sum_{i=1}^k A_i\right) \leq \Phi^{\mathcal{H}}\left(\mathcal{F}, \sum_{i=1}^k A_i\right)$. 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 will show that for any point in $\left(\sum_{i \in \mathcal{I}} F_i\right) + \left(\sum_{i \in [k] \setminus \mathcal{I}} H_i\right)$ there exists a point in $\left(\sum_{i \in \mathcal{I}} A_i \cap F_i\right) + \left(\sum_{i \in [k] \setminus \mathcal{I}} H_i\right) \subseteq \sum_{i=1}^k A_i$ within the desired distance.

Since the target statement is invariant under equal translation of $\sum_{i=1}^k F_i$ and $\sum_{i=1}^k A_i$, we first translate A_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 5, we can equip each F_i with a distribution $Z_i \sim D_i$ with support on $A_i \cap F_i$ such that $\mathbb{E}_{Z_i \sim D_i}(Z_i) = \mathbf{p}_i$ and $\mathbb{E}(\|Z_i - \mathbb{E}(Z_i)\|_2^2) \leq \Xi^2(A_i \cap F_i)$. 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

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

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

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 $\sqrt{1 + \frac{L}{2} \sum_{i \in \mathcal{I}} \Xi^2(A_i \cap F_i)}$. Therefore, by the positive homogeneousness of $\|\cdot\|_{\mathcal{H}}$, the $\|\cdot\|_{\mathcal{H}}$ distance of \mathbf{p} to $(\sum_{i \in \mathcal{I}} A_i \cap F_i) + \mathcal{H}$ is at most $\sqrt{1 + \frac{L}{2} \sum_{i \in \mathcal{I}} \Xi^2(A_i \cap F_i)} - 1$. \square

Theorem 5 *Let A_1, A_2, \dots, A_k be nonempty subsets of \mathbb{R}^n and $\Xi(A_i) \leq \beta$, for all $i \in [k]$ such that $\mathbf{conv}(A_i)$ is closed and pointed. Suppose every extreme point of A_i has an associated l_2 -ball with radius r_i (r_i can be zero) as a hidden convex component. Let $r_* = \min_{\mathcal{I} \subseteq [k]: |\mathcal{I}|=n} \sum_{i \in [k] \setminus \mathcal{I}} r_i$. Then it follows that*

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

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

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

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

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 $\Phi \left(\mathbf{p}, \sum_{i=1}^k A_i \right) \leq \sqrt{r_*^2 + \frac{1}{2} n \beta^2} - r_*$. By Lemma 4, there exists face F_i of $\mathbf{conv}(A_i)$ such that $\mathbf{p} \in \sum_{i=1}^k F_i$ and $\sum_{i=1}^k \dim(F_i) \leq n$.

By the hypothesis, after suitable translation of all the sets, 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 7 implies that

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

This implies that

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

where the last inequality uses Claim 4. Since \mathbf{p} is arbitrary, this implies that $\Phi \left(\sum_{i=1}^k A_i \right) \leq \sqrt{r_*^2 + \frac{1}{2} n \beta^2} - r_*$. The rest of statement directly follows that $r_* \geq (k-n)r$. \square

We end the section with several examples and remarks.

Remark 1 Lemma 7 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 Δ .

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 $f > 1$.

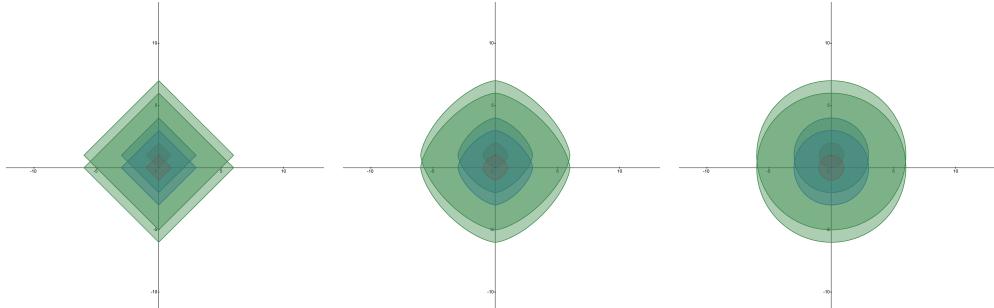


Fig. 4: $p = 1$

Fig. 5: $p = 1.5$

Fig. 6: $p = 2$

Remark 3 To ensure that the Minkowski sums converge to a convex set, it is sufficient that smooth, hidden convex components exist only in neighborhoods of the extreme points of convex hull of each nonconvex set. Such components need not exist around non-extreme points of the convex hulls. To illustrate this, consider the \mathbb{R}^2 examples T_1 and T_2 shown in Figures 7, 8, 9 and 10. 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 $k \rightarrow \infty$, whereas $\sum_{i=1}^k T_2$ does not. The difference arises from the behavior at extreme points: every extreme point of $\text{conv}(T_1)$ admits a smooth hidden convex component (e.g., a ball contained in T_1 in a neighborhood of that point), while $\text{conv}(T_2)$ possesses five extreme points for which no such smooth hidden convex component exists.



Fig. 7: T_1



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



Fig. 9: T_2



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

5 Asymptotically tight Lagrangian dual of smooth nonconvex problems

In this section, we aim to leverage Theorem 4 and Lemma 7 to establish the asymptotic tightness of problem (4). 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) as:

$$\begin{aligned} \text{OPT}_{\mathbf{s}}(\mathbf{b}) := \inf & \sum_{i=1}^k t_i \\ \text{s.t. } & \sum_{i=1}^k B^{(i)} \mathbf{x}^{(i)} \leq \mathbf{b}, \\ & \left\| \mathbf{x}^{(i)} \right\|_0 \leq s_i, t_i \geq f^{(i)}(\mathbf{x}^{(i)}), \forall i \in [k]. \end{aligned} \tag{8}$$

In this case, we have

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

and the projected set $\mathcal{P}^{(i)}$ for (8) is

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

Note that we can simply choose $s_i = n$ in (8) to model the block-structured smooth nonconvex problem without sparsity constraints.

Definition 5 (1-coercive function) [21] A function $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ is called *1-coercive* if

$$\frac{f(x)}{\|x\|_2} \longrightarrow +\infty \quad \text{as } \|x\|_2 \rightarrow \infty.$$

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

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

The notion of 1-coercive functions is commonly used in analysis and optimization [24]. Some example includes strictly convex quadratic functions and univariate polynomials of even degree with a positive leading coefficient.

The existence of an asymptotically tight Lagrangian dual requires two additional assumptions. Assumption 3 requires the nonconvexity of each block in (4) is bounded. Assumption 4 implies that $\mathbf{conv}(\mathcal{P}_s^{(i)})$ is closed and shown in Lemma 10.

Assumption 3 There exists some β such that $\Xi(\mathcal{P}_s^{(i)}) \leq \beta, \forall i \in [k]$.

A function $f(\mathbf{x}) : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ is called proper if there exists some \mathbf{x}_0 such that $f(\mathbf{x}_0) < \infty$.

Assumption 4 Each $f^{(i)}(\cdot)$ is 1-coercive, closed and proper and $f^{(i)}(\mathbf{0}) < \infty$.

Lemma 8 Under Assumption 4, $\mathbf{conv}(\mathcal{X}_s^{(i)})$ is closed and pointed. Moreover, $\text{rec}(\mathbf{conv}(\mathcal{X}_s^{(i)})) = \{(t, 0) : t \geq 0\}$.

Proof Let $\chi_{s_i}(\mathbf{x}) := \begin{cases} \infty & \text{if } \|\mathbf{x}^{(i)}\|_0 \geq s_i + 1 \\ 0 & \text{if otherwise} \end{cases}$. 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. Since $f^{(i)}(\mathbf{0}) < \infty$, $\tilde{f}^{(i)}(\mathbf{x})$ is also proper. Moreover, it is clear that $\text{epi } \tilde{f}^{(i)}(\mathbf{x}) = \mathcal{X}_s^{(i)}$. It is well known that the convex hull of the epigraph of a proper, 1-coercive and closed function is closed and therefore $\mathbf{conv}(\mathcal{X}_s^{(i)})$ is closed [18]. Suppose, for the sake of contradiction, that $\mathbf{conv}(\mathcal{X}_s^{(i)})$ is not pointed. We first observe that the definition of 1-coercive-ness implies that there exists some lb such that $lb \leq \tilde{f}^{(i)}(\mathbf{x})$. This implies that $t \geq lb$, for all $(t, \mathbf{x}) \in \mathbf{conv}(\mathcal{X}_s^{(i)})$. Therefore, any line contained in $\mathbf{conv}(\mathcal{X}_s^{(i)})$ must take the form of $l(\gamma) := (t_0, \mathbf{x}_0) + \gamma(0, \mathbf{r})$, $\gamma \in \mathbb{R}^n$ for some nonzero \mathbf{r} . To also establish the properties of $\text{rec}(\mathbf{conv}(\mathcal{X}_s^{(i)}))$, we show the stronger statement that $\mathbf{conv}(\mathcal{X}_s^{(i)})$ does not contain any ray taking the form of $l^+(\gamma) := (t_0, \mathbf{x}_0) + \gamma(0, \mathbf{r})$, $\gamma \in \mathbb{R}_+$ for some nonzero \mathbf{r} . This implies that $\mathbf{conv}(\mathcal{X}_s^{(i)})$ is pointed. Combined with the fact that the function is also lower bounded by some constant, this also implies that $\text{rec}(\mathbf{conv}(\mathcal{X}_s^{(i)})) = \{(t, 0) : t \geq 0\}$. The argument is as follows.

Suppose such ray $l^+(\gamma)$ exists, since $\tilde{f}^{(i)}(\mathbf{x})$ is 1-coercive, there exists some $R > 1$ such that $\|\mathbf{x}\|_2 \geq R \implies \tilde{f}^{(i)}(\mathbf{x}) \geq (|lb| + |t_0| + 1) \|\mathbf{x}\|_2$. Choose $|\gamma|$ sufficiently large so that $\|\mathbf{x}_0 + \gamma\mathbf{r}\|_2 \geq (n+2)R$, since $(t_0, \mathbf{x}_0 + \gamma\mathbf{r}) \in \mathbf{conv}(\mathcal{X}_s^{(i)})$, this implies that $\mathbf{x}_0 + \gamma\mathbf{r} = \sum_{j \in [n+2]} \lambda_j(t_j, \mathbf{x}_j)$ for some $(t_j, \mathbf{x}_j) \in \mathcal{X}_s^{(i)}$ and λ belongs to a simplex of dimension $n+2$. Since $\|\cdot\|_2$ is subadditive and positively homogeneous, this implies that $\|\mathbf{x}_0 + \gamma\mathbf{r}\|_2 \leq \sum_{i \in [n+2]} \lambda_i \|\mathbf{x}_j\|_2$. Since $\|\mathbf{x}_0 + \gamma\mathbf{r}\|_2 \geq (n+2)R$, a simple pigeonhole argument implies that there exists some $j_* \in [n]$, such that $\lambda_{j_*} \|\mathbf{x}_{j_*}\|_2 \geq R$. Since $\lambda_{j_*} \in [0, 1]$, this further shows

that $\|\mathbf{x}_{j_*}\|_2 \geq R$ and by the choice of R , this implies that $t_{j_*} \geq (|lb| + |t_0| + 1) \|\mathbf{x}_{j_*}\|_2$. This shows that

$$\lambda_{j_*} \|\mathbf{x}_{j_*}\|_2 \geq R \implies \lambda_{j_*} t_{j_*} \geq \frac{R}{\|\mathbf{x}_{j_*}\|_2} t_{j_*} \geq R(|lb| + |t_0| + 1) \geq |lb| + |t_0| + 1.$$

This implies that

$$\sum_{j \in [n+2]} \lambda_j t_j = \lambda_{j_*} t_{j_*} + \sum_{j \in [n+2] \setminus \{j_*\}} \lambda_j t_j \geq |lb| + |t_0| + 1 + (1 - \lambda_{j_*})lb > t_0,$$

which leads to contraction. \square

Lemma 9 (Theorem 9.1 in [26]) Let C be a closed pointed convex set of \mathbb{R}^n . For any linear map $\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, if $\text{rec}(C) \cap \ker(\mathcal{A}) = \{0\}$, then $\mathcal{A}(C)$ is closed and pointed.

Lemma 10 Under assumption 4, $\text{conv}(\mathcal{P}_s^{(i)})$ is closed and pointed.

Proof By Claim 3, $\text{conv}(\mathcal{P}_s^{(i)})$ can be viewed as the image of $\text{conv}(\mathcal{X}_s^{(i)})$ under the following linear map $\mathcal{A} : (t, \mathbf{x}) \rightarrow \begin{bmatrix} 1 \\ B \end{bmatrix} (t, \mathbf{x})$. In this case, we have $\text{rec}(\text{conv}(\mathcal{X}_s^{(i)})) \cap \ker(\mathcal{A}) = \{\mathbf{0}\}$ because any non-zero $\mathbf{v} \in \text{rec}(\text{conv}(\mathcal{X}_s^{(i)}))$ has form of $(r, \mathbf{0})$ for some $r > 0$ by Lemma 8 and it is clear that $\mathcal{A}\mathbf{v} \neq \mathbf{0}$ in this case. Applying Lemma 9 and Lemma 8 yields the desired statement. \square

The last ingredient for applying Theorem 5 is to show that there exists a smooth hidden convex component (l_2 -ball) associated with each extreme point of $\text{conv}(\mathcal{P}_s^{(i)})$. We obtain this from the following observation. The set $\mathcal{P}_s^{(i)}$ 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, as established in Claim 5 below. Under the linear projection induced by the coupling matrix $B^{(i)}$, this ball is mapped to an ellipsoid in $\mathcal{P}_s^{(i)}$. This ellipsoid provides a desired smooth hidden convex component. We then extract an l_2 -ball contained within this ellipsoid, which ultimately serves as the hidden convex component used in our proof. The maximal distortion under this linear projection and the choice l_2 -ball contained in this ellipsoid is quantified by $\mathcal{L}_s(B^{(i)})$, defined below.

Definition 6 Given a matrix $B \in \mathbb{R}^{m \times n}$ and let $\tilde{B}_S \in \mathbb{R}^{(m+1) \times (|S|+1)} := \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & B_S \end{bmatrix}$ be a matrix where B_S is the submatrix of B that consists of columns of B with indices in S . We define the *projection factor with respect to sparsity level s* as

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

where $\sigma_\diamond(A) := \begin{cases} 0 & \text{if } A \text{ is not positive definite;} \\ \frac{\sqrt{\sigma_{\min}(A^{-1/2})}}{\sigma_{\max}(A^{-1/2})} & \text{otherwise;} \end{cases}$ and $\sigma_{\min}(\cdot), \sigma_{\max}(\cdot)$ are the minimal singular value and maximal singular value, respectively. $A^{1/2}$ denotes the principal square root of a positive definite matrix A .

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 linear independent columns among every possible collection of $s^{(i)}$ columns of $B^{(i)}$. A necessary condition for this is $s^{(i)} \geq m$. 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 $s^{(i)} \geq m$.

Claim 5 For any L -smooth function $f(\cdot)$ and any $\mathbf{x} \in \mathbb{R}^n$, there exists a l_2 -ball \mathcal{B} with radius $\frac{1}{L}$ such that $\mathcal{B} \subseteq \text{epi } f$ and $(\mathbf{x}, f(\mathbf{x})) \subseteq \mathcal{B}$.

Proof For any $\mathbf{x} \in \mathbb{R}^n$, let $\mathbf{g} = \nabla f(\mathbf{x})$. We construct the center of the ball \mathcal{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, $(\mathbf{x}, f(\mathbf{x})) \in \mathcal{B}$ and it remains to prove that $\mathcal{B} \subseteq \text{epi } f$. For every point $(\mathbf{y}, t) \subseteq \mathcal{B}$, the lowest possible value of t is

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

Therefore, it suffices to prove $f(\mathbf{y}) \leq t_{\inf}(\mathbf{y})$ for all \mathbf{y} such that $\|\mathbf{y} - \mathbf{x}_c\|_2^2 \leq r^2$. Note that $f(\mathbf{y}) \leq t_{\inf}(\mathbf{y})$ can be further simplified to prove that

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

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

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \mathbf{y} - \mathbf{x}, \mathbf{g} \rangle + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2, \forall \mathbf{y} \in \mathbb{R}^n. \quad (9)$$

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

that

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

where the first inequality just drops $(s + \langle \mathbf{g}, \Delta \mathbf{x} \rangle)^2$ term; the second inequality uses (9); the third inequality uses the fact that $\alpha \leq 1$.

□

Lemma 11 [27] Let $E := \{\mathbf{x} : \|Ax - \mathbf{k}\|_2 \leq 1\}$ be an ellipsoid, where A is an invertible matrix. Then, for every $\mathbf{p} \in E$, there exists a l_2 -norm ball containing \mathbf{p} that is subset of E and has a radius of $\frac{\sqrt{\sigma_{\min}(A)}}{\sigma_{\max}(A)}$.

Claim 6 For every extreme point \mathbf{v} (0-dimensional face) of $\text{conv}(\mathcal{P}^{(i)})$, there exists a l_2 -ball B with radius $\frac{\mathcal{L}_s(B^{(i)})}{L}$ such that $\mathbf{v} \in B \subseteq \mathcal{P}^{(i)}$.

Proof For a fixed support set $S \subseteq [n]$, we define

$$\mathcal{P}_S^{(i)} := \left\{ (t, \mathbf{d}) \in \mathbb{R}^{m+1} \left| \begin{array}{l} \exists \mathbf{x}, x_i = 0, \forall i \notin S; \\ \mathbf{d} = B^{(i)} \mathbf{x}; \\ t \geq f(\mathbf{x}); \end{array} \right. \right\}, \quad (10)$$

It is clear that

$$\mathcal{P}^{(i)} = \bigcup_{S \subseteq [n], |S|=s_i} \mathcal{P}_S^{(i)}.$$

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

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

Since \mathbf{v} is a extreme point, we can further assume that $t_{\mathbf{v}} = f^{(i)}(\mathbf{x})$. Consider the following function $f_S^{(i)}(\mathbf{y}) : \mathbb{R}^S \rightarrow \mathbb{R}$ 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 5, there is a ball \mathcal{B}_S with radius $\frac{1}{L}$ with dimension $|S| + 1$ that $((\mathbf{x}_{\mathbf{v}})_S, f(\mathbf{x}_{\mathbf{v}})) \subseteq \mathcal{B}_S \subseteq \text{epi } f_S^{(i)}$. Now consider the linear map:

$$(\mathbf{x}_S, t) \rightarrow \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & B_S^{(i)} \end{bmatrix} (\mathbf{x}_S, t)$$

Let $Q := \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & B_S^{(i)} \end{bmatrix}$. If QQ^\top is not invertible, we have $\mathcal{L}_s(B^{(i)}) = 0$ and the lemma directly holds. Therefore, we assume that QQ^\top is invertible. It is well known that a linear map transforms a ball into an ellipse. In particular, \mathcal{B}_S is mapped to an ellipse E (up to translation) with form: $\left\{ \mathbf{y} : \mathbf{y}^\top (QQ^\top)^{-1} \mathbf{y} \leq \left(\frac{1}{L}\right)^2 \right\} = \left\{ \mathbf{y} : \|U\mathbf{y}\| \leq \left(\frac{1}{L}\right) \right\}$ where $U = (QQ^\top)^{-1/2}$. By Lemma 11 and the definition of $\mathcal{L}_s(B^{(i)})$, we can find a l_2 -ball with radius $\frac{\mathcal{L}_s(B^{(i)})}{L}$ that both contains \mathbf{v} and is included in E and therefore included in $\mathcal{P}^{(i)}$.

□

Theorem 6 Consider problem (4), assume that each $f^{(i)}(\cdot)$ is L -smooth and under the Assumption 1, Assumption 3 and Assumption 4 that there exists some β such that $\Xi(\mathcal{P}^{(i)}) \leq \beta$. Then it follow that

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

$$\text{OPT}_s(\mathbf{b} + \mathcal{E} \mathbf{1}) - \mathcal{E} \leq \text{DUAL}_s(\mathbf{b}) \leq \text{OPT}_s(\mathbf{b}),$$

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 ω such that $\mathcal{L}(B^{(i)}) \geq \omega, \forall i \in [k]$, then it follows that

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

Proof Using Claim 6 and Theorem 5, we obtain

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

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

Remark 5 The smoothness assumption in Theorem 6 can be relaxed. Instead of requiring that f be globally smooth, that is, 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 6 remains valid.

Acknowledgments

We would also like to thank Diego Cifuentes, Greg Blekherman, Santosh Vempala, and Kazuo Murota for the various wonderful discussions.

References

- [1] Lübbecke, M.E.: Column generation. Wiley encyclopedia of operations research and management science **17**, 18–19 (2010)
- [2] Desaulniers, G., Desrosiers, J., Solomon, M.M.: Column Generation vol. 5. Springer, New York, USA (2006)
- [3] Cifuentes, D., Dey, S.S., Xu, J.: Lagrangian dual for integer optimization with zero duality gap that admits decomposition. In: International Conference on Integer Programming and Combinatorial Optimization, pp. 184–198 (2025). Springer
- [4] Starr, R.M.: Quasi-equilibria in markets with non-convex preferences. Econometrica: journal of the Econometric Society, 25–38 (1969)
- [5] Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. SIAM, Philadelphia, USA (2001)
- [6] Pólik, I., Terlaky, T.: A survey of the S-lemma. SIAM review **49**(3), 371–418 (2007)

- [7] Chandrasekaran, V., Duff, T., Rodriguez, J.I., Shu, K.: Lagrangian dual sections: A topological perspective on hidden convexity. arXiv preprint arXiv:2510.06112 (2025)
- [8] Dines, L.L.: On the mapping of quadratic forms. *Bulletin of the American Mathematical Society* **47**(6), 494–498 (1941)
- [9] Fradelizi, M., Madiman, M., Marsiglietti, A., Zvavitch, A.: The convexification effect of minkowski summation. *EMS Surveys in Mathematical Sciences* **5**(1), 1–64 (2018)
- [10] 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)
- [11] Aubin, J.-P., Ekeland, I.: Estimates of the duality gap in nonconvex optimization. *Mathematics of Operations Research* **1**(3), 225–245 (1976)
- [12] Udell, M., Boyd, S.: Bounding duality gap for separable problems with linear constraints. *Computational Optimization and Applications* **64**(2), 355–378 (2016)
- [13] Bi, Y., Tang, A.: Duality gap estimation via a refined shapley–folkman lemma. *SIAM Journal on Optimization* **30**(2), 1094–1118 (2020)
- [14] 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)
- [15] Nguyen, T., Vohra, R.: (near-) substitute preferences and equilibria with indivisibilities. *Journal of Political Economy* **132**(12), 4122–4154 (2024)
- [16] Murota, K., Tamura, A.: Shapley–Folkman-type theorem for integrally convex sets. *Discrete Applied Mathematics* **360**, 42–50 (2025)
- [17] Wegmann, R.: Einige maßzahlen für nichtkonvexe mengen. *Archiv der Mathematik* **34**(1), 69–74 (1980)
- [18] Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex analysis and minimization algorithms ii. *Grundlehren der mathematischen Wissenschaften* (1993)
- [19] 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)
- [20] Geoffrion, A.M.: Lagrangean relaxation for integer programming. In: *Approaches to Integer Programming*, pp. 82–114. Springer, Amsterdam, Netherlands (2009)

- [21] Hiriart-Urruty, J.-B., Lemaréchal, C.: *Fundamentals of Convex Analysis*. Springer, New York, USA (2004)
- [22] Dey, S.S., Meunier, F., Ramirez, D.M.: Geoffrion's theorem beyond finiteness and rationality. arXiv preprint arXiv:2510.10966 (2025)
- [23] Lemaréchal, C., Renaud, A.: A geometric study of duality gaps, with applications. *Mathematical Programming* **90**(3), 399–427 (2001)
- [24] Dubois-Taine, B., d'Aspremont, A.: Frank-wolfe meets shapley-folkman: a systematic approach for solving nonconvex separable problems with linear constraints: B. dubois-taine, a. d'aspremont. *Mathematical Programming*, 1–51 (2025)
- [25] Molinaro, M.: Strong convexity of feasible sets in off-line and online optimization. *Mathematics of Operations Research* **48**(2), 865–884 (2023)
- [26] Rockafellar, R.T.: *Convex analysis*:(pms-28) (2015)
- [27] Liu, N., Grimmer, B.: Gauges and accelerated optimization over smooth and/or strongly convex sets. arXiv preprint arXiv:2303.05037 (2023)