Received: 28 November 2023 | Revised: 5 April 2024 | Accepted: 13 May 2024

DOI: 10.1112/blms.13116

RESEARCH ARTICLE

Bulletin of the London Mathematical Society

Graph convex hull bounds as generalized Jensen inequalities

Ilja Klebanov

Mathematical Institute, Freie Universität Berlin, Berlin, Germany

Correspondence
Ilja Klebanov, Mathematical Institute, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany.
Email: iljusch@gmail.com

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: EXC-2046/1

Abstract

Jensen's inequality is ubiquitous in measure and probability theory, statistics, machine learning, information theory and many other areas of mathematics and data science. It states that, for any convex function $f : K \to \mathbb{R}$ defined on a convex domain $K \subseteq \mathbb{R}^d$ and any random variable X taking values in K, $\mathbb{E}[f(X)] \geqslant f(\mathbb{E}[X])$. In this paper, sharp upper and lower bounds on $\mathbb{E}[f(X)]$, termed 'graph convex hull bounds', are derived for arbitrary functions f on arbitrary domains K, thereby extensively generalizing Jensen's inequality. The derivation of these bounds necessitates the investigation of the convex hull of the graph of f, which can be challenging for complex functions. On the other hand, once these inequalities are established, they hold, just like Jensen's inequality, for *any* K-valued random variable X. Therefore, these bounds are of particular interest in cases where f is relatively simple and X is complicated or unknown. Both finite- and infinite-dimensional domains and codomains of f are covered as well as analogous bounds for conditional expectations and Markov operators.

MSC 2020
26D15, 28B05, 46A55, 52A40, 60B11 (primary), 37A30 (secondary)

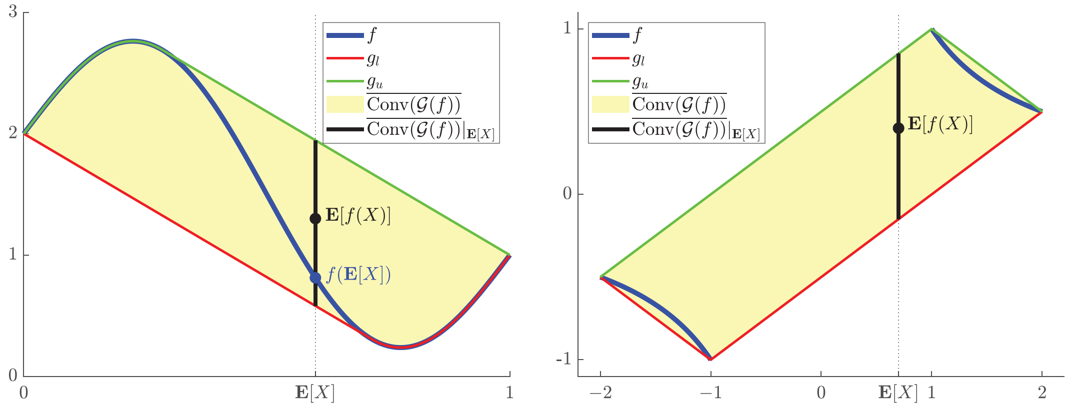

FIGURE 1 Illustration of Examples 3.10 and 3.11. Since the mean $\mathbb{E}[(X, f(X))] = (\mathbb{E}[X], \mathbb{E}[f(X)])$ (black dot) of the pair $(X, f(X))$ lies in the closure $\overline{\mathrm{Conv}(\mathcal{G}(f))}$ of the convex hull of the graph of f, the value $\mathbb{E}[f(X)]$ is restricted to the thick black line $\overline{\mathrm{Conv}}_{\mathbb{E}[X]}$, providing the bounds in Theorem 3.4. Note that, in the second example, the domain $K = [-2, -1] \cup [1, 2]$ of f is disconnected and $f(\mathbb{E}[X])$ is not even defined whenever $\mathbb{E}[X] \in (-1, 1)$, which does not affect Theorem 3.4.

1 | INTRODUCTION

In many theoretical and practical derivations, it is challenging to exactly compute the mean $\mathbb{E}[f(X)]$ of a function $f : K \to \mathbb{R}$, where $K \subseteq \mathbb{R}^d$ and $d \in \mathbb{N}$, applied to a K-valued random variable X. Thus, one has to rely on bounds on $\mathbb{E}[f(X)]$. This is particularly the case when X is known only in terms of its mean $\mathbb{E}[X]$, requiring dependence solely on the knowledge of $\mathbb{E}[X]$ and f, leading to the following general problem addressed in this paper:

Problem 1.1. Establish (sharp) bounds on the mean $\mathbb{E}[f(X)]$ given $\mathbb{E}[X]$ and f.

In the finite-dimensional setup described above, Jensen's inequality [9] offers partial answers for convex functions f on convex domains K, asserting that

$$\mathbb{E}[f(X)] \geqslant f(\mathbb{E}[X]) \tag{1.1}$$

(without specifying an upper bound). While Jensen's inequality is a foundational results in measure theory and is widely utilized in probability theory, statistics, information theory, statistical physics and many other research areas, it exhibits several crucial limitations and restrictions. The aim of this paper is to derive new bounds on $\mathbb{E}[f(X)]$, termed *graph convex hull bounds*, that address these shortcomings, in particular:

- Jensen's inequality says nothing about functions f that are neither convex nor concave, while the graph convex hull bounds hold for arbitrary functions.
- A significant limitation of Jensen's inequality is its requirement for the function f to be defined over a *convex* domain K. Our graph convex hull bounds, on the other hand, are free from such constraints on the domain of f, allowing for even disconnected domains as illustrated in Example 3.11 and Figure 1, enhancing their versatility.

- Jensen's inequality does not offer any form of 'reverse' bounds of the forms $\mathbb{E}[f(X)] \leqslant s + f(\mathbb{E}[X])$ or $\mathbb{E}[f(X)] \leqslant cf(\mathbb{E}[X])$ for some $s = s(X) \geqslant 0$, $c = c(X) \geqslant 1$, a feature that could be invaluable in various mathematical and applied contexts and can certainly be obtained for specific functions or function classes [4, 6, 12, 20]. The minimal value $s = s(X) \geqslant 0$ which satisfies above property is called the *Jensen gap* [1, 11, 19]. The graph convex hull bounds provide sharp bounds on $\mathbb{E}[f(X)]$ both from above and below. However, unlike the Jensen gap, our bounds depend on X only through $\mathbb{E}[X]$ (cf. Problem 1.1), hence they are more generally applicable but can only be sharp in a weaker sense.
- For domains K within infinite-dimensional topological vector spaces $\mathcal{X}$, Jensen's inequality requires f to meet certain continuity conditions [14], a stipulation not needed by our graph convex hull bounds.

The graph convex hull bounds are obtained by exploiting the basic fact that the mean of the pair $(X, f(X))$ lies in the closure $\overline{\mathrm{Conv}(\mathcal{G}(f))}$ of the convex hull of the graph $\mathcal{G}(f)$ of f; cf. Corollary 3.3 and Figure 1. This can be shown by a simple application of the Hahn–Banach separation theorem — characterize $\overline{\mathrm{Conv}(\mathcal{G}(f))}$ as an intersection of half-spaces and note that the expected value $\mathbb{E}$ as an operator 'respects' each half-space, since it is linear, positive and $\mathbb{E}[Z] = Z$ for constant Z. While the graph convex hull bounds in concrete applications require some investigation of $\overline{\mathrm{Conv}(\mathcal{G}(f))}$ (note that Jensen's inequality also requires some analysis of f, namely the verification of its convexity), all inequalities are derived solely from the properties of f and hold, just like Jensen's inequality, for *any* random variable X no matter its probability distribution. This property is crucial because, in practice, the function f is often sufficiently simple to analyze, at least numerically, whereas the distribution of X might be complex, unknown, or one requires a priori bounds that hold *uniformly* for all random variables taking values in K.

Note that the three properties of $\mathbb{E}$ mentioned above for the derivation of $(\mathbb{E}[X], \mathbb{E}[f(X)]) \in \overline{\mathrm{Conv}(\mathcal{G}(f))}$ are of algebraic nature and have little to do with measure or integration theory. Hence, the graph convex hull bounds can be extended to a broader class of linear operators beyond expected values, so-called Markov operators, as well as to conditional expectations, which are well-known facts in the case of Jensen's inequality, see, for example, [2, Equation (1.2.1); 7, Theorem 10.2.7]. Once these three versions of the graph convex hull bounds are proven, the corresponding versions of Jensen's inequality follow as simple corollaries, providing a novel and considerably simpler proof of this famous result (cf. Remark 3.7). By working in a more general setup than is typical for Markov operators (Remark 4.2), we also strongly generalize the corresponding version of Jensen's inequality.

The paper is structured as follows. After laying out the general setup and notation in Section 2, the graph convex hull bounds are derived for expected values (Section 3), Markov operators (Section 4) and conditional expectations (Section 5). The corresponding versions of Jensen's inequality follow as Corollaries 3.6, 4.11 and 5.7 in their respective sections.

2 | PRELIMINARIES

In order to formulate the results within the most general setting possible, we assume that $\mathcal{X}$ and $\mathcal{Y}$ are both real, Hausdorff and locally convex topological vector spaces and $f : K \to \mathcal{Y}$, where $K \subseteq \mathcal{X}$ is an arbitrary domain. However, readers are encouraged to conceptualize $\mathcal{X} = \mathbb{R}^d$ and $\mathcal{Y} = \mathbb{R}$ for simplicity. Furthermore, $(\Omega, \Sigma, \mathbb{P})$ will denote a probability space, and $X : \Omega \to \mathcal{X}$ will represent a function such that $(X, f(X))$ is weakly integrable.[†] Therefore, all expected values are interpreted in the weak sense, which implies that X does not necessarily need to be measurable, that is, a random variable. Recall that a map $X : \Omega \to \mathcal{X}$ is called *weakly integrable* if there exists $e \in \mathcal{X}$ such that, for any $\ell \in \mathcal{X}'$, $\ell(X) \in L^1(\mathbb{P})$ and $\ell(e) = \int_\Omega \ell(X)\,\mathrm{d}\mathbb{P}$, in which case the expected value of X is defined by $\mathbb{E}[X] := \int_\Omega X\,\mathrm{d}\mathbb{P} := e$ [17, Definition 3.26]. Here, $\mathcal{X}'$ denotes the continuous dual of $\mathcal{X}$. In particular, each Pettis or Bochner integrable map is weakly integrable, and readers may consider Bochner integrable random variables X for simplicity. When working with Markov operators in Section 4, we will not impose any measurability or integrability assumptions on $(X, f(X))$, while for conditional expectations in Section 5 we will require $\mathcal{X}$ and $\mathcal{Y}$ to be Banach spaces and $(X, f(X))$ to be Bochner integrable. Recall that a random variable $X : (\Omega, \Sigma, \mathbb{P}) \to \mathcal{X}$ is Bochner or strongly integrable if and only if X is measurable and $\int_\mathcal{X} \|X\|_\mathcal{X}\,\mathrm{d}\mathbb{P} < \infty$ [5, Theorem II.2.2], in which case we write $X \in L^1(\mathbb{P}; \mathcal{X})$. The relevant assumptions will be detailed separately in each section, namely in Assumptions 3.1, 4.5 and 5.1.

Throughout the paper, we use the following notation. We denote the closure of a subset A of a real topological vector space by $\overline{A}$ and the convex hull of A by

$$\operatorname{Conv}(A) := \bigcap \{C \supset A \mid C \text{ convex}\} = \left\{ \textstyle\sum_{j=1}^J w_j a_j \,\middle|\, J \in \mathbb{N},\ a_j \in A,\ w_j \geqslant 0,\ \sum_{j=1}^J w_j = 1 \right\}. \tag{2.1}$$

Further, $\mathcal{G}(f) = \{(x, f(x)) \mid x \in K\}$ denotes the graph of a function $f : K \to \mathcal{Y}$ and, for $M \subseteq \mathcal{X} \times \mathcal{Y}$ and $x \in \mathcal{X}$,

$$M|_x := \{y \in \mathcal{Y} \mid (x, y) \in M\}$$

(note that $\overline{M}|_x$ and $\overline{M|_x}$ might not coincide). Finally, while we clearly have in mind inequalities with respect to some total or partial order or preorder on $\mathcal{Y}$, most statements are presented for arbitrary binary relations on $\mathcal{Y}$, which we continue to denote by $\leq$ for readability reasons. However, in the case where $\mathcal{Y} = \mathbb{R}$, we consistently assume the canonical total order. In certain situations, we will extend the space $\mathcal{Y}$ by adding the elements $\pm\infty$, which are presumed to satisfy $-\infty \leqslant y \leqslant \infty$ for each $y \in \mathcal{Y}$, and denote $\overline{\mathcal{Y}} := \mathcal{Y} \cup \{\pm\infty\}$, as is typical for the extended real number line $\overline{\mathbb{R}} = [-\infty, \infty] = \mathbb{R} \cup \{\pm\infty\}$. Closed intervals in $\overline{\mathcal{Y}}$ are defined by $[a, b] := \{y \in \overline{\mathcal{Y}} \mid a \leqslant y \leqslant b\}$, $a, b \in \overline{\mathcal{Y}}$. We now state a simple preliminary lemma that we will use to derive Jensen's inequality from the graph convex hull bounds:

Lemma 2.1. *Let $f : K \to \mathbb{R}$ be a convex function on a convex domain $K \subseteq \mathbb{R}^d$. Then its epigraph* $\operatorname{epi}(f) := \{(x, y) \in K \times \mathbb{R} \mid y \geqslant f(x)\} \supseteq \mathcal{G}(f)$ *is convex and, consequently,* $\operatorname{Conv}(\mathcal{G}(f))|_x \subseteq [f(x), \infty)$ *for each $x \in K$. If $K = \mathbb{R}^d$, then* $\overline{\operatorname{Conv}(\mathcal{G}(f))}|_x \subseteq [f(x), \infty)$ *for each $x \in K$.*

Proof. The convexity of $\operatorname{epi}(f)$ is a well-known result [16, Theorem 4.1]. It follows that $\operatorname{Conv}(\mathcal{G}(f)) \subseteq \operatorname{epi}(f)$ by (2.1) and $\operatorname{Conv}(\mathcal{G}(f))|_x \subseteq \operatorname{epi}(f)|_x = [f(x), \infty)$ for each $x \in K$. If $K = \mathbb{R}^d$, then f is continuous [3, Theorem 4.1.3] and $\operatorname{epi}(f)$ is closed. Hence, $\overline{\operatorname{Conv}(\mathcal{G}(f))}|_x \subseteq \overline{\operatorname{epi}(f)}|_x = \operatorname{epi}(f)|_x = [f(x), \infty)$ for each $x \in K$. □

[†] As is typical for random variables X, we write $f(X)$ for the composition $f \circ X$; and extend this notation to maps X that are not measurable.

3 | GRAPH CONVEX HULL BOUNDS FOR EXPECTED VALUES

Throughout this section, we make the following general assumptions:

Assumption 3.1. $(\Omega, \Sigma, \mathbb{P})$ is a probability space, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are real, Hausdorff and locally convex topological vector spaces (equipped with their Borel σ-algebras). Furthermore, $\leq$ represents any binary relation on $\mathcal{Y}$, with the stipulation that it defaults to the canonical total order in the case $\mathcal{Y} = \mathbb{R}$ (hence the notation). Finally, $K \subseteq \mathcal{X}$, $f : K \to \mathcal{Y}$ and $X : \Omega \to K$ such that the pair $(X, f(X))$ is weakly integrable.

The following theorem presents a well-known and intuitive result that the mean $\mathbb{E}[Z]$ of a random variable Z taking values in a closed convex subset C of $\mathcal{Z}$ lies within this set[†]: $\mathbb{E}[Z] \in C$. If $\mathcal{Z} = \mathbb{R}^d$, the closedness assumption is not required. Its proof utilizes the Hahn–Banach separation theorem.

Theorem 3.2 (mean/barycenter of convex set lies within its closure). *Let Assumption 3.1 hold and let $Z : \Omega \to A$ be a weakly integrable map taking values in a subset $A \subseteq \mathcal{Z}$. Then $\mathbb{E}[Z] \in \mathrm{Conv}(A)$ if $\mathcal{Z} = \mathbb{R}^d$, $d \in \mathbb{N}$ and $\mathbb{E}[Z] \in \overline{\mathrm{Conv}(A)}$ in general.*[‡] *Further, for any $z \in \mathrm{Conv}(A)$ there exist a probability space $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}})$ and a weakly integrable map $\tilde{Z} : \tilde{\Omega} \to A$ such that $\mathbb{E}[\tilde{Z}] = z$.*

Proof. See [14, Theorem 3.1] for the general statement $\mathbb{E}[Z] \in \overline{\mathrm{Conv}(A)}$ and [7, Theorem 10.2.6.] for the proof of $\mathbb{E}[Z] \in \mathrm{Conv}(A)$ in the case $\mathcal{Z} = \mathbb{R}^d$ (though our setup is slightly more general, the proof goes similarly). Now let $z \in \mathrm{Conv}(A)$. Then it can be written as a convex combination of elements of A, that is, $z = \sum_{j=1}^{J} w_j a_j$, for some $J \in \mathbb{N}$, $a_j \in A$, $w_j \geqslant 0$, $j = 1, \dots, J$, with $\sum_{j=1}^{J} w_j = 1$. Choose $\tilde{\Omega} = \{1, \dots, J\}$ with distribution $\tilde{\mathbb{P}}$ given by the probability vector $(w_1, \dots, w_J)$ and let $\tilde{Z}(j) = a_j$. Then $\mathbb{E}[\tilde{Z}] = \sum_{j=1}^{J} w_j a_j = z$. □

An application of above theorem to the convex hull of the graph of a function f yields:

Corollary 3.3. *Under Assumption 3.1, $\mathbb{E}[(X, f(X))] \in \mathrm{Conv}(\mathcal{G}(f))$ if $\mathcal{X} = \mathbb{R}^m$, $\mathcal{Y} = \mathbb{R}^n$, $m, n \in \mathbb{N}$ and $\mathbb{E}[(X, f(X))] \in \overline{\mathrm{Conv}(\mathcal{G}(f))}$ in general. Further, for any $(x, y) \in \mathrm{Conv}(\mathcal{G}(f))$ there exist a probability space $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}})$ and $\tilde{X} : \tilde{\Omega} \to K$ such that the pair $(\tilde{X}, f(\tilde{X}))$ is weakly integrable and $\mathbb{E}[(\tilde{X}, f(\tilde{X}))] = (x, y)$.*

Proof. This follows from Theorem 3.2 for $Z := (X, f(X))$ and $A := \mathcal{G}(f) \subseteq \mathcal{X} \times \mathcal{Y} =: \mathcal{Z}$ (note that any map $\tilde{Z}$ taking values in $\mathcal{G}(f)$ is of the form $(\tilde{X}, f(\tilde{X}))$). □

The following result is basically a restatement of Corollary 3.3. Since it is the main result of this section, we state it as a theorem:

[†] Note that this addresses a distinct issue from that explored by Choquet theory, and the Krein–Milman theorem in particular, as discussed in [15], where a bounded, closed convex set is analyzed in relation to the convex hull of its *extreme points*. Extreme points will be of no relevance to this paper.

[‡] For an example of a random variable Z taking values in a convex subset C of an infinite-dimensional space which satisfies $\mathbb{E}[Z] \in \overline{C} \setminus C$, see [14, Remark 3.2].

Theorem 3.4 (Graph convex hull bounds on $\mathbb{E}[f(X)]$). *Under Assumption 3.1, let the convex hull of the graph of f be 'enclosed' by two 'envelope' functions $g_l, g_u \colon \overline{\operatorname{Conv}(K)} \to \overline{\mathcal{Y}}$ in the following way:*

$$
\begin{aligned}
&\overline{\operatorname{Conv}(\mathcal{G}(f))} \subseteq \{(x,y) \in \overline{\operatorname{Conv}(K)} \times \mathcal{Y} \mid g_l(x) \leqslant y \leqslant g_u(x)\}, \\
\text{that is,} \qquad &\overline{\operatorname{Conv}(\mathcal{G}(f))}|_x \subseteq [g_l(x), g_u(x)] \qquad \text{for each } x \in \overline{\operatorname{Conv}(K)}.
\end{aligned}
\tag{3.1}
$$

Then

$$
g_l(\mathbb{E}[X]) \leqslant \mathbb{E}[f(X)] \leqslant g_u(\mathbb{E}[X]). \tag{3.2}
$$

If $\mathcal{X} = \mathbb{R}^m$, $\mathcal{Y} = \mathbb{R}^n$, $m, n \in \mathbb{N}$, then $\overline{\operatorname{Conv}(K)}$ and $\overline{\operatorname{Conv}(\mathcal{G}(f))}$ can be replaced by $\operatorname{Conv}(K)$ *and* $\operatorname{Conv}(\mathcal{G}(f))$*, respectively. Further, for $\mathcal{Y} = \mathbb{R}$, (3.2) is sharp in the following sense:*

- *If $\mathcal{X} = \mathbb{R}^m$ and $g_l(x) = \inf(\operatorname{Conv}(\mathcal{G}(f))|_x)$ and $g_u(x) = \sup(\operatorname{Conv}(\mathcal{G}(f))|_x)$ for each $x \in \operatorname{Conv}(K)$, then, for every $x \in \operatorname{Conv}(K)$ and $y \in (g_l(x), g_u(x))$, there exist a probability space $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}})$ and $\tilde{X} \colon \tilde{\Omega} \to K$ such that the pair $(\tilde{X}, f(\tilde{X}))$ is weakly integrable and $(\mathbb{E}[\tilde{X}], \mathbb{E}[f(\tilde{X})]) = (x, y)$.*
- *In general, if $g_l(x) = \inf(\overline{\operatorname{Conv}(\mathcal{G}(f))}|_x)$ and $g_u(x) = \sup(\overline{\operatorname{Conv}(\mathcal{G}(f))}|_x)$ for each $x \in \overline{\operatorname{Conv}(K)}$, then, for every $x \in \overline{\operatorname{Conv}(K)}$ and $y \in (g_l(x), g_u(x))$ and for any neighborhood U of (x, y), there exist a probability space $(\tilde{\Omega}, \tilde{\Sigma}, \tilde{\mathbb{P}})$ and $\tilde{X} \colon \tilde{\Omega} \to K$ such that the pair $(\tilde{X}, f(\tilde{X}))$ is weakly integrable and $(\mathbb{E}[\tilde{X}], \mathbb{E}[f(\tilde{X})]) \in U$.*

Proof. Since $\mathbb{E}[X] \in \overline{\operatorname{Conv}(K)}$ and $(\mathbb{E}[X], \mathbb{E}[f(X)]) \in \overline{\operatorname{Conv}(\mathcal{G}(f))}$ by Theorem 3.2, and Corollary 3.3,

$$
\mathbb{E}[f(X)] \in \overline{\operatorname{Conv}(\mathcal{G}(f))}|_{\mathbb{E}[X]} \subseteq [g_l(\mathbb{E}[X]), g_u(\mathbb{E}[X])];
$$

cf. Figure 1. In the finite-dimensional case $\mathcal{X} = \mathbb{R}^m$, $\mathcal{Y} = \mathbb{R}^n$, we can omit all the closures by Theorem 3.2 and Corollary 3.3. The claim on the sharpness of the bounds also follows directly from Corollary 3.3. □

Remark 3.5. We emphasize again that, just like Jensen's inequality, the bounds g_l and g_u are determined solely by the function f and hold for *any* K-valued map X that meets the minimal integrability assumptions.

A direct application of the graph convex hull bounds from Theorem 3.4 yields Jensen's inequality:

Corollary 3.6 (Jensen's inequality). *Let Assumption 3.1 hold and $f \colon K \to \mathbb{R}$ be a convex function on a convex domain $K \subseteq \mathbb{R}^d$. Then $\mathbb{E}[f(X)] \geqslant f(\mathbb{E}[X])$.*

Proof. This follows directly from Lemma 2.1 and Theorem 3.4 with $g_l := f$. □

Remark 3.7. It is worth noting that this novel proof strategy of Jensen's inequality is significantly simpler than traditional approaches. In the proof of, for example, [7, Theorem 10.2.6.], the author outlines:

- a first argument showing $\mathbb{E}[X] \in K$, akin to Theorem 3.2, ensuring $f(\mathbb{E}[X])$ is well defined;
- a second, straightforward argument similar to Lemma 2.1; and
- a third, more complex argument to establish the inequality itself.

Yet, this latter argument can be entirely bypassed by applying the first argument a second time, this time to the pair $(X, f(X))$, resulting in $(\mathbb{E}[X], \mathbb{E}[f(X)]) \in \operatorname{Conv}(\mathcal{G}(f))$ and thus directly proving Jensen's inequality. This elegant approach appears to have been previously overlooked, as well as its application to non-convex functions and domains as demonstrated in Theorem 3.4.

Note that the lower bound $f(\mathbb{E}[X])$ on $\mathbb{E}[f(X)]$ in Jensen's inequality (1.1) is superfluous in our setup and may not even be well defined (since $\mathbb{E}[X]$ is not guaranteed to lie in K; cf. Example 3.11). However, if comparing $\mathbb{E}[f(X)]$ and $f(\mathbb{E}[X])$ is crucial, the stated result leads to the following consequences whenever $f(\mathbb{E}[X])$ exists:

Corollary 3.8. *Let the assumptions of Theorem 3.4 hold and let $\mathcal{Y} = \mathbb{R}$. Then*

(a) $c_l(\mathbb{E}[X])\, f(\mathbb{E}[X]) \leqslant \mathbb{E}[f(X)] \leqslant c_u(\mathbb{E}[X])\, f(\mathbb{E}[X])$;
(b) $\max\left(\inf_{x\in K} f(x),\, f(\mathbb{E}[X]) - s_u\right) \leqslant \mathbb{E}[f(X)] \leqslant \min\left(\sup_{x\in K} f(x),\, f(\mathbb{E}[X]) + s_l\right)$;
(c) $\hat{c}_l\, f(\mathbb{E}[X]) \leqslant \mathbb{E}[f(X)] \leqslant \hat{c}_u\, f(\mathbb{E}[X])$

whenever the quantities

$$c_l(x) := \frac{g_l(x)}{f(x)}, \qquad s_l := \|f - g_l\|_\infty, \qquad \hat{c}_l := \inf_{x\in\overline{\operatorname{Conv}(K)}} \frac{g_l(x)}{f(x)},$$

$$c_u(x) := \frac{g_u(x)}{f(x)}, \qquad s_u := \|f - g_u\|_\infty, \qquad \hat{c}_u := \sup_{x\in\overline{\operatorname{Conv}(K)}} \frac{g_u(x)}{f(x)},$$

as well as $f(\mathbb{E}[X])$ are well defined. If $\mathcal{X} = \mathbb{R}^m$, $\mathcal{Y} = \mathbb{R}^n$, $m, n \in \mathbb{N}$, $\overline{\operatorname{Conv}(K)}$ can be replaced by $\operatorname{Conv}(K)$ *in the definition of $\hat{c}_l$ and $\hat{c}_u$.*

Proof. (a) and (b) follow directly from Theorem 3.4, while (c) follows from (a). □

Remark 3.9. While the bounds in Theorem 3.4 and Corollary 3.8(a) require the knowledge of $\mathbb{E}[X]$ (cf. Problem 1.1), statements (b) and (c) compare $\mathbb{E}[f(X)]$ with $f(\mathbb{E}[X])$ without relying on this information. In particular, (b) and (c) offer a priori bounds applicable uniformly across all random variables X with values in K. Consequently, in transitioning from (a) to (c), the lower bound $c_l(\mathbb{E}[X])$ must be substituted with its most conservative estimate, namely, the infimum of $c_l(x)$ across all potential values of x (and the supremum in the case of the upper bound). However, these bounds are still sharper than the obvious inequalities (whenever well defined)

$$\inf_{x\in K} f(x) \leqslant \mathbb{E}[f(X)] \leqslant \sup_{x\in K} f(x), \qquad f(\mathbb{E}[X]) \inf_{x,y\in\overline{K}} \frac{f(x)}{f(y)} \leqslant \mathbb{E}[f(X)] \leqslant f(\mathbb{E}[X]) \sup_{x,y\in\overline{K}} \frac{f(x)}{f(y)}.$$

Example 3.10. Let $K = [0, 1]$ and $f : K \to \mathbb{R}$, $f(x) = 2 - x + \sin(2\pi x)$, visualized in Figure 1 (left), which is neither convex nor concave. Then the 'envelope' functions of $\operatorname{Conv}(\mathcal{G}(f))$ are given by

$$g_l(x) = \begin{cases} 2 - ax & \text{if } x \leqslant x_*, \\ f(x) & \text{otherwise,} \end{cases} \qquad g_u(x) = \begin{cases} f(x) & \text{if } x \leqslant 1 - x_*, \\ 1 + a - ax & \text{otherwise,} \end{cases}$$

where $x_* \approx 0.715$ is the solution of $2\pi \cos(2\pi x)\, x = \sin(2\pi x)$ and $a = 1 - 2\pi \cos(2\pi x_*)$. Note that the constants $\hat{c}_l \approx 0.5$ and $\hat{c}_u \approx 6.5$ from Corollary 3.8 compare favorably with the obvious bounds $\inf_{x,y\in K} \frac{f(x)}{f(y)} \approx 0.09$ and $\sup_{x,y\in K} \frac{f(x)}{f(y)} \approx 11.6.$ from Remark 3.5 (it is easy to see how the function f can be modified to make the improvements over these bounds as large as desired).

Example 3.11. Let $K = [-2, -1] \cup [1, 2]$ be the *disconnected* domain of the function $f : K \to \mathbb{R}$, $f(x) = x^{-1}$, visualized in Figure 1 (right), which again is neither convex nor concave. Again, the 'envelope' functions of $\mathrm{Conv}(\mathcal{G}(f))$ from Theorem 3.4, given by the piecewise linear functions

$$g_l(x) = \begin{cases} (-3 - x)/2 & \text{if } x \in [-2, -1], \\ (x - 1)/2 & \text{if } x \in [-1, 2], \end{cases} \qquad g_u(x) = \begin{cases} (x + 1)/2 & \text{if } x \in [-2, 1], \\ (3 - x)/2 & \text{if } x \in [1, 2], \end{cases}$$

provide sharp bounds on $\mathbb{E}[f(X)]$. In this case, it might be impossible to compare $\mathbb{E}[f(X)]$ with $f(\mathbb{E}[X])$, since the latter is not defined whenever $\mathbb{E}[X] \in (-1, 1)$.

Remark 3.12. While in the above examples the envelope functions g_l and g_u were determined analytically (up to the approximation of x_*), they usually can be computed numerically with high precision and little computational effort even for a complicated function f, as long as its domain K is bounded.

4 | GRAPH CONVEX HULL BOUNDS FOR MARKOV OPERATORS

Note that the only properties of the expected value $\mathbb{E}$ used to prove $\mathbb{E}[Z] \in \overline{\mathrm{Conv}(A)}$ in Theorem 3.2 are its linearity, the positivity of $\ell(Z)$ for each $\ell \in \mathcal{Z}'$ (meaning that $\ell(Z) \geqslant 0$ implies $\mathbb{E}[\ell(Z)] \geqslant 0$) and the property that $\mathbb{E}[Z] = Z$ for each constant random variable $Z \in \mathcal{Z}$ (and the same is true for Jensen's inequality). Therefore, Jensen's inequality and the graph convex hull bounds can be interpreted as algebraic results rather than as measure theoretic ones, and can be extended to operators that do not necessarily involve integration, but merely satisfy the three properties above. This generalization is the purpose of this section and will also enable us to cover conditional expectations in the subsequent section. Such operators are known as *Markov operators* and play a crucial role in the analysis of time evolution phenomena and dynamical systems [2]. They are well known to satisfy Jensen's inequality [2, Equation (1.2.1)], yet we significantly broaden the setting in which it applies; cf. Remark 4.2.

Definition 4.1. Let Ω_1, Ω_2 be sets and $\mathcal{S}_1, \mathcal{S}_2$ be vector spaces of functions on Ω_1 and Ω_2 that contain the unit constant functions $\mathbb{1}_{\Omega_1} : \Omega_1 \to \mathbb{R}, t \mapsto 1$ and $\mathbb{1}_{\Omega_2}$, respectively, for example, $\mathcal{S}_1 = L^1(\mathbb{P}_1)$, if $(\Omega_1, \Sigma_1, \mathbb{P}_1)$ is a probability space. A linear operator $\mathcal{E} : \mathcal{S}_1 \to \mathcal{S}_2$ is called a *Markov operator* if

(i) $\mathcal{E}(\mathbb{1}_{\Omega_1}) = \mathbb{1}_{\Omega_2}$;

(ii) $\mathcal{E}(\varphi) \geqslant 0$ for each $\varphi \in \mathcal{S}_1$ with $\varphi \geqslant 0$.

Further, let $\mathcal{Z}$ be a real topological vector space and

$$\mathcal{S}_1^{\mathcal{Z}} \subseteq \{Z_1 : \Omega_1 \to \mathcal{Z} \mid \ell \in \mathcal{Z}' : \ell(Z_1) \in \mathcal{S}_1\}, \quad \mathcal{S}_2^{\mathcal{Z}} \subseteq \{Z_2 : \Omega_2 \to \mathcal{Z} \mid \ell \in \mathcal{Z}' : \ell(Z_2) \in \mathcal{S}_2\},$$

for example, $\mathcal{S}_1^{\mathcal{Z}} = L^1(\mathbb{P}_1; \mathcal{Z})$ in the case $\mathcal{S}_1 = L^1(\mathbb{P}_1)$. We call $\mathcal{E}^{\mathcal{Z}} : \mathcal{S}_1^{\mathcal{Z}} \to \mathcal{S}_2^{\mathcal{Z}}$ a *$\mathcal{Z}$-valued Markov operator* associated with the Markov operator $\mathcal{E}$ if $\ell(\mathcal{E}^{\mathcal{Z}}(Z_1)) = \mathcal{E}(\ell(Z_1))$ for each $\ell \in \mathcal{Z}'$ and $Z_1 \in \mathcal{S}_1^{\mathcal{Z}}$, that is, if the following diagram commutes for each $\ell \in \mathcal{Z}'$.

$$\begin{array}{ccc} \mathcal{S}_1^{\mathcal{Z}} & \xrightarrow{\mathcal{E}^{\mathcal{Z}}} & \mathcal{S}_2^{\mathcal{Z}} \\ \big\downarrow{\scriptstyle \ell\circ} & & \big\downarrow{\scriptstyle \ell\circ} \\ \mathcal{S}_1 & \xrightarrow{\mathcal{E}} & \mathcal{S}_2 \end{array}$$

Remark 4.2. Typically, Ω_1, Ω_2 are assumed to be measurable spaces or measure spaces, and the elements in $\mathcal{S}_1, \mathcal{S}_2$ to be measurable or to lie in L^1 with respect to certain measures [2, 8]. Further, some authors require Markov operators to be endomorphisms, that is, $\mathcal{S}_1 = \mathcal{S}_2$ [2], to be L^1-contractions [13, 18] or to preserve L^1-norm [8]. By working under the minimal requirements in Definition 4.1 and by introducing the entirely new notion of a $\mathcal{Z}$-valued Markov operator, we thus significantly expand the scope of the currently known version of Jensen's inequality for Markov operators [2, Equation (1.2.1)].

Example 4.3 (cf. Section 3). Let $(\Omega_1, \Sigma_1, \mathbb{P}_1) = (\Omega_2, \Sigma_2, \mathbb{P}_2)$ be a probability space and $\mathcal{S}_1 = L^1(\mathbb{P}_1)$ and $\mathcal{S}_2 = \operatorname{span}(\mathbb{1}_{\Omega_2})$. Further, let $\mathcal{Z}$ be a real, Hausdorff, locally convex topological vector space and $\mathcal{S}_1^{\mathcal{Z}} = \{Z_1 : \Omega_1 \to \mathcal{Z} \mid Z_1 \text{ weakly integrable}\}$, $\mathcal{S}_2^{\mathcal{Z}} = \{Z_2 : \Omega_2 \to \mathcal{Z} \mid Z_2 \text{ constant}\}$. Then the expected value $\mathbb{E} : \mathcal{S}_1^{\mathcal{Z}} \to \mathcal{S}_2^{\mathcal{Z}}$ is a $\mathcal{Z}$-valued Markov operator associated with the Markov operator $\mathbb{E} : \mathcal{S}_1 \to \mathcal{S}_2$, where $\mathcal{S}_2^{\mathcal{Z}}$ is identified with $\mathcal{Z}$ and $\mathcal{S}_2$ with $\mathbb{R}$. Note that the same notation $\mathbb{E}$ is used for both operators as is the standard in probability theory.

Example 4.4 (cf. Section 5). Let $(\Omega_1, \Sigma_1, \mathbb{P}_1) = (\Omega_2, \Sigma_2, \mathbb{P}_2)$ be a probability space and $\mathcal{F} \subseteq \Sigma_1$ be a sub-σ-algebra of Σ_1. Let $\mathcal{S}_1 = L^1(\Sigma_1, \mathbb{P}_1)$ and $\mathcal{S}_2 = L^1(\mathcal{F}, \mathbb{P}_1|_{\mathcal{F}})$. Further, let $\mathcal{Z}$ be a Banach space, $\mathcal{S}_1^{\mathcal{Z}} = L^1(\Sigma_1, \mathbb{P}_1; \mathcal{Z})$ and $\mathcal{S}_2^{\mathcal{Z}} = L^1(\mathcal{F}, \mathbb{P}|_{\mathcal{F}}; \mathcal{Z})$. Then the conditional expectation $\mathbb{E}[\,\cdot\,|\mathcal{F}] : \mathcal{S}_1^{\mathcal{Z}} \to \mathcal{S}_2^{\mathcal{Z}}$ is a $\mathcal{Z}$-valued Markov operator associated with the Markov operator $\mathbb{E}[\,\cdot\,|\mathcal{F}] : \mathcal{S}_1 \to \mathcal{S}_2$ (this follows easily from, for example, [5, Theorem II.2.6]). Again, the same notation $\mathbb{E}[\,\cdot\,|\mathcal{F}]$ is used for both operators as is common practice.

To extend the findings in Section 3 to Markov operators, we establish the subsequent assumptions for this section:

Assumption 4.5. $\mathcal{X}, \mathcal{Y}$ are two real, Hausdorff and locally convex topological vector spaces and Ω_1, Ω_2, $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_1^{\mathcal{X}}$, $\mathcal{S}_2^{\mathcal{X}}$, $\mathcal{S}_1^{\mathcal{Y}}$, $\mathcal{S}_2^{\mathcal{Y}}$ are as in Definition 4.1, with $\mathcal{Z}$ replaced by $\mathcal{X}$ and $\mathcal{Y}$, respectively. Further, $\mathcal{E}^{\mathcal{X}} : \mathcal{S}_1^{\mathcal{X}} \to \mathcal{S}_2^{\mathcal{X}}$ and $\mathcal{E}^{\mathcal{Y}} : \mathcal{S}_1^{\mathcal{Y}} \to \mathcal{S}_2^{\mathcal{Y}}$ are $\mathcal{X}$-valued and $\mathcal{Y}$-valued Markov operators, respectively, associated with the *same* Markov operator $\mathcal{E} : \mathcal{S}_1 \to \mathcal{S}_2$. $\leq$ is any binary relation on $\mathcal{Y}$, which we assume to be the canonical total order in the case $\mathcal{Y} = \mathbb{R}$. Finally $K \subseteq \mathcal{X}$, $f : K \to \mathcal{Y}$ and $X : \Omega_1 \to K$ such that the pair $(X, f(X)) \in \mathcal{S}_1^{\mathcal{X}} \times \mathcal{S}_1^{\mathcal{Y}}$.

Lemma 4.6. *Under Assumption 4.5, the operator* $\mathcal{E}^{\mathcal{X}\times\mathcal{Y}} \colon \mathcal{S}_1^{\mathcal{X}} \times \mathcal{S}_1^{\mathcal{Y}} \to \mathcal{S}_2^{\mathcal{X}} \times \mathcal{S}_2^{\mathcal{Y}}$, $(\tilde{X}, \tilde{Y}) \mapsto (\mathcal{E}^{\mathcal{X}}(\tilde{X}), \mathcal{E}^{\mathcal{Y}}(\tilde{Y}))$ *is an* $(\mathcal{X}\times\mathcal{Y})$*-valued Markov operator associated with the same Markov operator* $\mathcal{E}$.

Proof. Let $\tilde{Z} = (\tilde{X}, \tilde{Y}) \in \mathcal{S}_1^{\mathcal{X}} \times \mathcal{S}_1^{\mathcal{Y}}$ and $\ell \in (\mathcal{X}\times\mathcal{Y})'$. Then ℓ is of the form $\ell(x, y) = \ell^{\mathcal{X}}(x) + \ell^{\mathcal{Y}}(y)$ for some $\ell^{\mathcal{X}} \in \mathcal{X}'$, $\ell^{\mathcal{Y}} \in \mathcal{Y}'$. Hence, $\ell(\tilde{Z}) = \ell^{\mathcal{X}}(\tilde{X}) + \ell^{\mathcal{Y}}(\tilde{Y}) \in \mathcal{S}$ (since $\mathcal{S}$ is a vector space), as required (and similarly for $\mathcal{S}_2^{\mathcal{X}} \times \mathcal{S}_2^{\mathcal{Y}}$). Further,

$$\ell(\mathcal{E}^{\mathcal{X}\times\mathcal{Y}}(\tilde{Z})) = \ell^{\mathcal{X}}(\mathcal{E}^{\mathcal{X}}(\tilde{X}) + \ell^{\mathcal{Y}}(\mathcal{E}^{\mathcal{Y}}(\tilde{Y}) = \mathcal{E}(\ell^{\mathcal{X}}(\tilde{X})) + \mathcal{E}(\ell^{\mathcal{Y}}(\tilde{Y})) = \mathcal{E}(\ell(\tilde{Z})).$$ □

Let us now formulate the analogue of Theorem 3.2 for Markov operators:

Theorem 4.7 (Markov operators respect closed convex set constraints). *Let* Ω_1, Ω_2, $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{E}$, $\mathcal{Z}$, $\mathcal{S}_1^{\mathcal{Z}}$, $\mathcal{S}_2^{\mathcal{Z}}$, $\mathcal{E}^{\mathcal{Z}}$ *be as in Definition 4.1 and assume that* $\mathcal{Z}$ *is Hausdorff and locally convex. Further, let* $Z \in \mathcal{S}_1$ *with* $Z(\Omega_1) \subseteq A$ *for some subset* $A \subseteq \mathcal{Z}$. *Then* $(\mathcal{E}^{\mathcal{Z}}(Z))(\Omega_2) \subseteq \overline{\mathrm{Conv}(A)}$.

Proof. By the Hahn–Banach separation theorem, since $\overline{\mathrm{Conv}(A)}$ is convex, it coincides with the intersection of all closed half-spaces $\mathcal{H}_{\ell,s} := \{z \in \mathcal{Z} \mid \ell(z) \geqslant s\} \supseteq \overline{\mathrm{Conv}(A)}$ including it, where $\ell \in \mathcal{Z}' \setminus \{0\}$ and $s \in \mathbb{R}$. Since, for $\omega \in \Omega$,

$$\begin{aligned} Z(\Omega_1) \subseteq \mathcal{H}_{\ell,s} &\implies \ell(Z) \geqslant s \\ &\implies \mathcal{E}(\ell(Z)) \geqslant s \\ &\implies \ell(\mathcal{E}^{\mathcal{Z}}(Z) \geqslant s \\ &\implies \mathcal{E}^{\mathcal{Z}}(Z)(\Omega_2) \subseteq \mathcal{H}_{\ell,s}, \end{aligned}$$

$(\mathcal{E}^{\mathcal{Z}}(Z))(\Omega_2)$ lies in each half-space $\mathcal{H}_{\ell,s}$ including $\overline{\mathrm{Conv}(A)}$, proving the claim. □

An application of above theorem to the convex hull of the graph of a function f yields:

Corollary 4.8. *Under Assumption 5.1,* $(\mathcal{E}^{\mathcal{X}}(X), \mathcal{E}^{\mathcal{Y}}(f(X)))(\Omega_2) \subseteq \overline{\mathrm{Conv}(\mathcal{G}(f))}$.

Proof. Using Lemma 4.6, this follows from Theorem 4.7 for $Z := (X, f(X))$ and $A := \mathcal{G}(f) \subseteq \mathcal{X}\times\mathcal{Y} =: \mathcal{Z}$. □

As in Section 3, a reformulation of above corollary results in graph convex hull bounds for Markov operators in place of expected values:

Theorem 4.9 (Graph convex hull bounds on $\mathcal{E}^{\mathcal{Y}}(f(X))$). *Under Assumption 5.1, let the convex hull of the graph of* f *be 'enclosed' by two 'envelope' functions* $g_l, g_u \colon \overline{\mathrm{Conv}(K)} \to \overline{\mathcal{Y}}$ *satisfying* (3.1). *Then*

$$g_l(\mathcal{E}^{\mathcal{X}}(X)) \leqslant \mathcal{E}^{\mathcal{Y}}(f(X)) \leqslant g_u(\mathcal{E}^{\mathcal{X}}(X)). \tag{4.1}$$

Note that, in contrast to Theorem 3.4, all objects in (4.1) *are functions on* Ω_2.

Proof. Since $(\mathcal{E}^{\mathcal{X}}(X))(\Omega_2) \subseteq \overline{\mathrm{Conv}(K)}$ and $(\mathcal{E}^{\mathcal{X}}(X), \mathcal{E}^{\mathcal{Y}}(f(X)))(\Omega_2) \subseteq \overline{\mathrm{Conv}(\mathcal{G}(f))}$ by Theorem 4.7 and Corollary 4.8,

$$(\mathcal{E}^{\mathcal{Y}}(f(X)))(\Omega_2) \subseteq \overline{\mathrm{Conv}(\mathcal{G}(f))}|_{\mathcal{E}^{\mathcal{X}}(X)} \subseteq [g_l(\mathcal{E}^{\mathcal{X}}(X)), g_u(\mathcal{E}^{\mathcal{X}}(X))].$$

□

Corollary 4.10. *Let the assumptions of Theorem 4.9 hold and let $\mathcal{Y} = \mathbb{R}$. Then, using the notation from Corollary 3.8,*

(a) $c_l(\mathcal{E}^{\mathcal{X}}(X))\, f(\mathcal{E}^{\mathcal{X}}(X)) \leqslant \mathcal{E}^{\mathcal{Y}}(f(X)) \leqslant c_u(\mathcal{E}^{\mathcal{X}}(X))\, f(\mathcal{E}^{\mathcal{X}}(X))$,
(b) $\max\left(\inf_{x\in K} f(x),\, f(\mathcal{E}^{\mathcal{X}}(X)) - s_u\right) \leqslant \mathcal{E}^{\mathcal{Y}}(f(X)) \leqslant \min\left(\sup_{x\in K} f(x),\, f(\mathcal{E}^{\mathcal{X}}(X)) + s_l\right)$,
(c) $\hat{c}_l\, f(\mathcal{E}^{\mathcal{X}}(X)) \leqslant \mathcal{E}^{\mathcal{Y}}(f(X)) \leqslant \hat{c}_u\, f(\mathcal{E}^{\mathcal{X}}(X))$,

whenever these quantities are well defined.

Proof. (a) and (b) follow directly from Theorem 4.9, while (c) follows from (a). □

We conclude this section by deriving Jensen's inequality as a straightforward corollary of Theorem 4.9. To the best of the author's knowledge, Jensen's inequality for Markov operators has not been established with this level of generality, cf. Remark 4.2.

Corollary 4.11 (Jensen's inequality for Markov operators). *Let Assumption 4.5 hold and $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then $\mathcal{E}^{\mathcal{Y}}(f(X)) \geqslant f(\mathcal{E}^{\mathcal{X}}(X))$.*

Proof. This follows directly from Theorem 4.9 and Lemma 2.1. □

5 | GRAPH CONVEX HULL BOUNDS FOR CONDITIONAL EXPECTATIONS

It is well known that Jensen's inequality also holds for *conditional expectations* [7, Theorem 10.2.7]. In this section, we establish graph convex hull bounds for conditional expectations, representing a specific instance of the broader framework outlined in Section 4 (cf. Example 4.4) and deriving directly from the results therein. Since conditional expectations are commonly defined over Banach spaces and for Bochner integrable random variables [5, Section V.1], this will require slightly stronger assumptions:

Assumption 5.1. $(\Omega, \Sigma, \mathbb{P})$ is a probability space, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are real Banach spaces and $K \subseteq \mathcal{X}$ a subset of $\mathcal{X}$. Further, $\leq$ is any binary relation on $\mathcal{Y}$, which we assume to be the canonical total order in the case $\mathcal{Y} = \mathbb{R}$ and $\mathcal{F} \subseteq \Sigma$ is a sub-σ-algebra with $\mathbb{E}[\,\cdot\,|\mathcal{F}]$ denoting the corresponding conditional expectation. Finally, $K \subseteq \mathcal{X}$, $f : K \to \mathbb{R}$ and $X : \Omega \to K$ is a random variable such that the pair $(X, f(X))$ is Bochner integrable.

Theorem 5.2 (Conditional expectations respect closed convex set constraints). *Let $Z : (\Omega, \Sigma, \mathbb{P}) \to A$ be a Bochner integrable random variable taking values in a subset $A \subseteq \mathcal{Z}$ of a real Banach space $\mathcal{Z}$ and let $\mathcal{F} \subseteq \Sigma$ be a sub-σ-algebra. Then $\mathbb{E}[Z|\mathcal{F}] \in \overline{\mathrm{Conv}(A)}$ $\mathbb{P}$-almost surely.*

Proof. Note that, as is typical for conditional expectations, no difference is made in the notation between $\mathbb{E}[\cdot|\mathcal{F}]\colon L^1(\Sigma,\mathbb{P})\to L^1(\mathcal{F},\mathbb{P}|_{\mathcal{F}})$ and $\mathbb{E}[\cdot|\mathcal{F}]\colon L^1(\Sigma,\mathbb{P};\mathcal{Z})\to L^1(\mathcal{F},\mathbb{P}|_{\mathcal{F}};\mathcal{Z})$ (corresponding to $\mathcal{E}$ and $\mathcal{E}_{\mathcal{Z}}$ in Definition 4.1). Since $\mathbb{E}[\ell(Z)|\mathcal{F}]=\ell(\mathbb{E}[Z|\mathcal{F}])$ for each $\ell\in\mathcal{Z}'$ (this follows easily from, for example, [5, Theorem II.2.6]), and that $W\geqslant s$ implies $\mathbb{E}[W|\mathcal{F}]\geqslant s$ $\mathbb{P}$-almost surely for each real-valued random variable $W\colon (\Omega,\Sigma,\mathbb{P})\to\mathbb{R}$ by [10, Theorem 8.1(ii)], the claim follows directly from Theorem 4.7. □

Remark 5.3. In contrast to Section 4, all statements in L^1 can only hold $\mathbb{P}$-almost surely. Hence, when drawing the final conclusion from the proof of Theorem 4.7 to the one of Theorem 5.2, we must exercise caution when considering the intersection over all (possibly uncountably many) half-spaces that contain $\overline{\operatorname{Conv}(A)}$. However, following the construction of conditional expectations [5, Theorem V.1.4], there exists a representative of $\mathbb{E}[Z|\mathcal{F}]$ which lies in all these half-spaces simultaneously.

This theorem's application to the convex hull of the graph of a function f yields:

Corollary 5.4. *Under Assumption 5.1,* $\mathbb{E}[(X,f(X))|\mathcal{F}]\in\overline{\operatorname{Conv}(\mathcal{G}(f))}$ $\mathbb{P}$*-almost surely.*

Proof. This follows from Theorem 5.2 for $Z:=(X,f(X))$ and $A:=\mathcal{G}(f)\subseteq\mathcal{X}\times\mathcal{Y}=:\mathcal{Z}$. □

Theorem 5.5 (Graph convex hull bounds on $\mathbb{E}[f(X)|\mathcal{F}]$). *Under Assumption 5.1, let the convex hull of the graph of f be 'enclosed' by two 'envelope' functions* $g_l, g_u\colon \overline{\operatorname{Conv}(K)}\to\overline{\mathcal{Y}}$ *satisfying* (3.1)*. Then*

$$g_l(\mathbb{E}[X|\mathcal{F}])\leqslant\mathbb{E}[f(X)|\mathcal{F}]\leqslant g_u(\mathbb{E}[X|\mathcal{F}])\ \mathbb{P}\text{-almost surely.}\tag{5.1}$$

Note that, in contrast to Theorem 3.4, all objects in (5.1) *are random variables.*

Proof. Since $\mathbb{E}[X|\mathcal{F}]\in\overline{\operatorname{Conv}(K)}$ $\mathbb{P}$-almost surely and $(\mathbb{E}[X|\mathcal{F}],\mathbb{E}[f(X)|\mathcal{F}])\in\overline{\operatorname{Conv}(\mathcal{G}(f))}$ $\mathbb{P}$-almost surely by Theorem 5.2 and Corollary 5.4,

$$\mathbb{E}[f(X)|\mathcal{F}]\subseteq\overline{\operatorname{Conv}(\mathcal{G}(f))}|_{\mathbb{E}[X|\mathcal{F}]}\subseteq[g_l(\mathbb{E}[X|\mathcal{F}]),g_u(\mathbb{E}[X|\mathcal{F}])]\ \mathbb{P}\text{-almost surely.}$$

□

Corollary 5.6. *Let the assumptions of Theorem 5.5 hold and let* $\mathcal{Y}=\mathbb{R}$*. Then, using the notation from Corollary 3.8,*

(a) $c_l(\mathbb{E}[X|\mathcal{F}])\,f(\mathbb{E}[X|\mathcal{F}])\leqslant\mathbb{E}[f(X)]\leqslant c_u(\mathbb{E}[X|\mathcal{F}])\,f(\mathbb{E}[X|\mathcal{F}])$,
(b) $\max\big(\inf_{x\in K}f(x),\ f(\mathbb{E}[X|\mathcal{F}])-s_u\big)\leqslant\mathbb{E}[f(X)]\leqslant\min\big(\sup_{x\in K}f(x),\ f(\mathbb{E}[X|\mathcal{F}])+s_l\big)$,
(c) $\hat{c}_l\,f(\mathbb{E}[X|\mathcal{F}])\leqslant\mathbb{E}[f(X)]\leqslant\hat{c}_u\,f(\mathbb{E}[X|\mathcal{F}])$,

$\mathbb{P}$*-almost surely, whenever these quantities are well defined.*

Proof. (a) and (b) follow directly from Theorem 5.5, while (c) follows from (a). □

Again, conditional Jensen's inequality follows almost directly from Theorem 5.5:

Corollary 5.7 (Conditional Jensen's inequality). *Let Assumption 5.1 hold and $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Then $\mathbb{E}[f(X)|\mathcal{F}] \geqslant f(\mathbb{E}[X|\mathcal{F}])$ $\mathbb{P}$-almost surely.*

Proof. This follows directly from Theorem 5.5 and Lemma 2.1. □

ACKNOWLEDGMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC-2046/1, project 390685689) through the project EF1-10 of the Berlin Mathematics Research Center MATH+. The author thanks Tim Sullivan for carefully proofreading this manuscript and for helpful suggestions.

Open access funding enabled and organized by Projekt DEAL.

JOURNAL INFORMATION

The *Bulletin of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

ORCID

Ilja Klebanov https://orcid.org/0000-0002-1911-4281

REFERENCES

1. S. Abramovich and L.-E. Persson, *Some new estimates of the 'Jensen gap'*, J. Inequal. Appl. (2016), **9**. doi: 10.1186/s13660-016-0985-4.
2. D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014. doi: 10.1007/978-3-319-00227-9.
3. J. M. Borwein and A. S. Lewis, *Convex analysis and nonlinear optimization: theory and examples*, 2nd ed., CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, vol. 3, Springer, New York, 2006. doi: 10.1007/978-0-387-31256-9.
4. I. Budimir, S. S. Dragomir, and J. Pečarić, *Further reverse results for Jensen's discrete inequality and applications in information theory*, JIPAM. J. Inequal. Pure Appl. Math. **2** (2001), no. 1, 14.
5. J. Diestel and J. J. Uhl, *Vector measures*, Mathematical Surveys, vol. 15, American Mathematical Society, Providence, RI, 1977. doi: 10.1090/surv/015.
6. S. S. Dragomir, *Some reverses of the Jensen inequality with applications*, Bull. Aust. Math. Soc. **87** (2013), no. 2, 177–194. doi: 10.1017/S0004972712001098.
7. R. M. Dudley, *Real analysis and probability*, Cambridge Studies in Advanced Mathematics, vol. 74, Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original. doi: 10.1017/CBO9780511755347.
8. T. Eisner, B. Farkas, M. Haase, and R. Nagel, *Operator theoretic aspects of ergodic theory*, Graduate Texts in Mathematics, vol. 272, Springer, Cham, 2015. doi: 10.1007/978-3-319-16898-2.
9. J. L. W. V. Jensen, *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*, Acta Math. **30** (1906), no. 1, 175–193. doi: 10.1007/BF02418571.
10. O. Kallenberg, *Foundations of modern probability*, 3rd ed., Probability Theory and Stochastic Modelling, vol. 99, Springer, Cham, 2021. doi: 10.1007/978-3-030-61871-1.
11. S. Khan, M. Adil Khan, S. I. Butt, and Y.-M. Chu, *A new bound for the Jensen gap pertaining twice differentiable functions with applications*, Adv. Difference Equ. **2020** (2020), 11. doi: 10.1186/s13662-020-02794-8.
12. S. Khan, M. Adil Khan, and Y.-M. Chu, *Converses of the Jensen inequality derived from the Green functions with applications in information theory*, Math. Methods Appl. Sci. **43** (2020), no. 5, 2577–2587. doi: 10.1002/mma.6066.

13. M. Lin, *Weak mixing for semigroups of Markov operators without finite invariant measures*, Ergodic theory (Proc. Conf., Math. Forschungsinst., Oberwolfach, 1978), Lecture Notes in Math., vol. 729, Springer, Berlin, 1979, pp. 89–92. doi: 10.1007/BFb0063286.
14. M. D. Perlman, *Jensen's inequality for a convex vector-valued function on an infinite-dimensional space*, J. Multivariate Anal. **4** (1974), 52–65. doi: 10.1016/0047-259X(74)90005-0.
15. R. R. Phelps, *Lectures on Choquet's theorem*, 2nd ed., Lecture Notes in Mathematics, vol. 1757, Springer, Berlin, 2001. doi: 10.1007/b76887.
16. R. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics and Physics, Princeton University Press, Princeton, NJ, 1997.
17. W. Rudin, *Functional analysis*, 2nd ed., International Series in Pure and Applied Mathematics, McGraw-Hill, Inc., New York, 1991.
18. R. Rudnicki, *Markov operators: applications to diffusion processes and population dynamics*, vol. 27, 2000, pp 67–79. Dedicated to the memory of Wiesław Szlenk. doi: 10.4064/am-27-1-67-79.
19. H. Ullah, M. Adil Khan, and T. Saeed, *Determination of bounds for the Jensen gap and its applications*, Mathematics **9** (2021), no. 23. doi: 10.3390/math9233132.
20. G. Wunder, B. Groß, R. Fritschek, and R. F. Schaefer, *A reverse Jensen inequality result with application to mutual information estimation*, 2021 IEEE Information Theory Workshop (ITW), 2021, pp. 1–6. doi: 10.1109/ITW48936.2021.9611449.