Asymptotic confidence sets for random linear programs

Shuyu Liu SL7695@NYU.EDU
Courant Institute of Mathematical Sciences
New York University, New York, NY, 10012

Florentina Bunea FB238@CORNELL.EDU
Department of Statistics and Data Science
Cornell University, Ithaca, NY, 14853

Jonathan Niles-Weed JNW@CIMS.NYU.EDU
Courant Institute of Mathematical Sciences and Center for Data Science
New York University, New York, NY, 10012

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

Motivated by the statistical analysis of the discrete optimal transport problem, we prove distributional limits for the solutions of linear programs with random constraints. Such limits were first obtained by Klatt, Munk, & Zemel (2022), but their expressions for the limits involve a computationally intractable decomposition of $\mathbb{R}^m$ into a possibly exponential number of convex cones. We give a new expression for the limit in terms of auxiliary linear programs, which can be solved in polynomial time. We also leverage tools from random convex geometry to give distributional limits for the entire set of random optimal solutions, when the optimum is not unique. Finally, we describe a simple, data-driven method to construct asymptotically valid confidence sets in polynomial time.

Keywords: Linear programming, distributional inference, confidence sets

1. Introduction

Linear programming is one of the core techniques in convex optimization, capturing many canonical problems such as maximum flow, shortest path, bipartite matching, and optimal transport. Linear programs (LPs) are notable for their versatility, their rich combinatorial theory, and their algorithmic tractability: the pioneering work of Hačijan (1979) showed that LPs can be solved in polynomial time, and the last 70 years of research in theoretical computer science and scientific computing have made solving linear programs a "mature technology" in practice (Boyd and Vandenberghe, 2004).

We consider throughout a standard form LP, given by

$$\min_{\mathbf{x}\in\mathbb{R}^m} \langle \mathbf{c}, \mathbf{x}\rangle, \qquad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{k\times m}$, $\mathbf{b} \in \mathbb{R}^k$ and $\mathbf{c} \in \mathbb{R}^m$. The goal of this paper is to understand the distributional behavior of solutions to Eq. (1) when $\mathbf{b}$ is replaced by a random vector $\mathbf{b}_n$. We assume the existence of a random variable $\mathbb{G}$ such that

$$r_n(\mathbf{b}_n - \mathbf{b}) \xrightarrow{D} \mathbb{G} \tag{2}$$

for some rate $r_n \to \infty$, and we will seek a corresponding limit law for the solutions to Eq. (1). This setting is motivated by applications of linear programming in statistics and machine learning, where the "right-hand side" vector $\mathbf{b}$ corresponds to random capacities, demands, or prices. An important example, which motivates many of the developments of this paper, is the linear programming formulation of the optimal transportation problem between discrete distributions, where the vector $\mathbf{b}$ corresponds to the probability mass function of the two measures. The statistician who only has access to these measures via samples can compute a solution to an *empirical optimal transport* problem by replacing $\mathbf{b}$ with an estimator $\mathbf{b}_n$. Quantifying the uncertainty in the resulting solution requires constructing an asymptotic confidence set for this random linear program.

Obtaining distributional limit results for solutions to random optimization problems is, of course, a well studied subject both in scientific computing and in statistics (Dupačová and Wets, 1988; King and Rockafellar, 1993; Linderoth et al., 2006; Polyak and Juditsky, 1992; Shapiro, 1991), but the LP lacks the regularity conditions necessary to apply classical results: neither smoothness nor strong convexity holds for Eq. (1) in general, solutions are generally not unique, and optimal solutions to Eq. (1) always lie on the boundary of the feasible set. By contrast, standard distributional limit results, for instance in the analysis of M-estimators, require local strong convexity, uniqueness, and that the solution to the population-level problem lies in the relative interior of the feasible set (see, e.g., Vaart, 1998). The challenges met in circumventing these classical conditions are well known (Aitchison and Silvey, 1958; Andrews, 2002; Chernoff, 1954). Statistically, the lack of regularity in Eq. (1) is the source of several pathologies: even when the solution to Eq. (1) is unique, the limiting distribution will in general not be Gaussian, and if there are multiple solutions to Eq. (1) it is not even clear how to formulate the desired distributional limit results. The typical path forward, not taken in this work, is to impose extra conditions to guarantee that uniqueness holds and to focus on settings where there is sufficient regularity to ensure a Gaussian limit.

Let us give a very simple example which illustrates some of the difficulties of this problem. Consider a 2×2 optimal transport problem:

$$\min_{\pi \in \mathbb{R}^{2\times 2}} \pi_{12} + \pi_{21}\,, \qquad \text{s.t. } \pi\mathbf{1} = \mathbf{q}, \pi^\top \mathbf{1} = \mathbf{s}, \pi \geq \mathbf{0}\,,$$

where $\mathbf{q} = \mathbf{s} = (1/2, 1/2)$. In this case, the target solution $\pi^* = (1/2, 0; 0, 1/2)$ is unique. If we suppose that $\mathbf{q}$ is replaced by random vector in the probability simplex $\mathbf{q}_n = (q_n^{(1)}, q_n^{(2)})$, then the optimal solution to the perturbed program is

$$\hat{\pi}_n = (1/2, q_n^{(1)} - 1/2; 0, q_n^{(2)})\mathbb{1}_{\{q_n^{(1)} > q_n^{(2)}\}} + (q_n^{(1)}, 0; q_n^{(2)} - 1/2, 1/2)\mathbb{1}_{\{q_n^{(1)} \leq q_n^{(2)}\}}\,,$$

and if we assume $\sqrt{n}(\mathbf{q}_n - \mathbf{q})$ converges in distribution to a centered Gaussian vector, the rescaled solution $\sqrt{n}(\hat{\pi}_n - \pi^*)$ converges to a mixture distribution with two non-Gaussian components. A 3×3 version of the same problem, with the same objective function and $\mathbf{q} = \mathbf{s} = (1/3, 1/3, 1/3)$, has multiple optimal solutions, and *a priori* it is not clear how to quantify the uncertainty of a solution obtained when $\mathbf{r}$ is replaced by a random counterpart.

The challenges in obtaining distributional limits for LPs were first tackled by the pioneering work of Klatt et al. (2022), who derived distributional limits for (1) in a very general setting. Their results are expressed in terms of a partition of $\mathbb{R}^m$ into closed convex cones; the restriction of the limiting distribution on each cone is a linear function of the limit of the sequence $r_n(\mathbf{b}_n - \mathbf{b})$. To handle the fact that solutions to (1) may not be unique, Klatt et al. (2022) adopt a framework of an algorithmic flavor: they assume, informally speaking, that there exists a consistent, possibly randomized, selection procedure to specify a solution within the optimal set. This strategy allows them to prove a distributional limit for the particular optimal solution selected by this procedure, without having to assume that the optimal solution is unique.

Despite the completeness and sophistication of their approach, Klatt et al. (2022) leave open several fundamental questions. First, it is not clear whether it is possible to sample from their limit laws in polynomial time: all of their limits are expressed in terms of a decomposition of $\mathbb{R}^m$ into a possibly exponential number of closed convex cones. Even evaluating the functions involved in their limiting expressions therefore appears to be computationally intractable. Second, their approach to non-unique solutions cleverly sidesteps the need to assume that the optimal solution is unique; however, the resulting limit law does not give insight into the overall geometry of the random solution set. Third, even ignoring issues of computational feasibility, their results do not yield a method to obtain asymptotically valid confidence sets from data, because the limiting distributions they obtain depend on the (typically unknown) optimal solutions to the original LP.

In this work, we propose solutions to these three questions. First, in the case the solution to the original LP is unique, we give a new representation of the limit that can be sampled from in polynomial time; in fact, we show that the limit can be generated by solving an auxiliary random linear program. Second, in the general (non-unique) case, we define and prove a distributional limit for the optimal solutions *in the space of convex sets*—the resulting limit captures the random geometry of the entire solution set. Finally, we develop a practical and computationally cheap data-driven method for constructing asymptotically valid confidence sets.

2. Preliminaries on linear programming

In this section, we recall some facts about the structure of linear programs. We point the reader to standard reference works (Bertsimas and Tsitsiklis, 1997; Boyd and Vandenberghe, 2004; Bradley et al., 1977; Nocedal and Wright, 2006) for additional background information.

We denote the set of optimal solutions to (1) by

$$\mathbf{x}^*(\mathbf{b}) := \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^m} \langle \mathbf{c}, \mathbf{x}\rangle, \qquad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}. \tag{3}$$

The notation $\mathbf{x}^*(\mathbf{b})$ emphasizes that this optimal set depends on the right-hand side $\mathbf{b}$. In general, LPs do not possess unique solutions, so that typically $|\mathbf{x}^*(\mathbf{b})| \neq 1$. However, if the solution is unique, by slight abuse of notation we write $\mathbf{x}^*(\mathbf{b})$ for both the (single-element) set of optimal solutions and for the optimal solution itself. We sometimes refer to $\mathbf{x}^*(\mathbf{b})$ as the set of "target solutions," to contrast it with the random solution set $\mathbf{x}^*(\mathbf{b}_n)$ obtained by replacing $\mathbf{b}$ by its random counterpart. We denote the optimal objective value of (1) by $f(\mathbf{b})$.

Throughout, we make the following assumptions on (1).

Assumption 1 *The constraint matrix $\mathbf{A}$ has full rank, the optimal solution set $\mathbf{x}^*(\mathbf{b})$ is nonempty and bounded, and (1) satisfies the Slater condition (Boyd and Vandenberghe, 2004, Section 5.2.3), i.e., $\exists \mathbf{x}_0 \in \mathbb{R}^m$, such that $\mathbf{A}\mathbf{x}_0 = \mathbf{b},\ \mathbf{x}_0 > \mathbf{0}$.*

The assumption that $\mathbf{A}$ is full rank is without loss of generality, as redundant constraints in the matrix can always be removed. The assumption that $\mathbf{x}^*(\mathbf{b})$ is nonempty and bounded is also made by Klatt et al. (2022) and holds for many LPs of interest, including optimal transport problems. Finally, the Slater condition is a standard assumption in convex programming and is only a minor strengthening of Assumption (B2) of Klatt et al. (2022, see Lemma 5.4).

2.1. Bases

For any subset $\mathrm{I} \subseteq \{1, \ldots, m\}$, we denote by $\mathbf{A}_\mathrm{I}$ the $k \times |\mathrm{I}|$ submatrix of $\mathbf{A}$ formed by taking the columns of $\mathbf{A}$ corresponding to the elements of I. Analogously, for $\mathbf{x} \in \mathbb{R}^m$, we write $\mathbf{x}_\mathrm{I}$ for the vector of length $|\mathrm{I}|$ consisting of the coordinates of $\mathbf{x}$ corresponding to I.

Definition 1 *A set* $\mathrm{I} \subseteq [m]$ *is a* basis *if*

$$|\mathrm{I}| = k, \qquad \mathrm{rank}(\mathbf{A}_\mathrm{I}) = k \tag{4}$$

Given a basis I, we can define the *basic solution* $\mathbf{x}(\mathrm{I}; \mathbf{b})$ to be the vector $\mathbf{x}$ satisfying

$$\begin{aligned} \mathbf{x}_\mathrm{I} &= \mathbf{A}_\mathrm{I}^{-1}\mathbf{b} \\ \mathbf{x}_{\mathrm{I}^C} &= \mathbf{0}\,. \end{aligned} \tag{5}$$

Explicitly, $\mathbf{x}(\mathrm{I}, \mathbf{b})$ is defined by setting the coordinates not in I to zero and inverting the matrix $\mathbf{A}_\mathrm{I}$ to obtain the values on the coordinates corresponding to I. This vector is a feasible solution to (1) if and only if the vector $\mathbf{A}_\mathrm{I}^{-1}\mathbf{b}$ is nonnegative; if it is, we say that $\mathbf{x}(\mathrm{I}; \mathbf{b})$ is a *basic feasible solution*. By construction, basic feasible solutions have at most k non-zero entries: if we denote the support (i.e., the set of non-zero entries) of a vector $\mathbf{x}$ by $S(\mathbf{x})$, then

$$S(\mathbf{x}(\mathrm{I}; \mathbf{b})) \subseteq \mathrm{I}\,.$$

This inclusion can be strict if the vector $\mathbf{A}_\mathrm{I}^{-1}\mathbf{b}$ has zero coordinates. When the inclusion is strict, the solution is called *degenerate*. If $\mathbf{x}$ is a degenerate basic feasible solution, then any basis I such that $S(\mathbf{x}) \subseteq \mathrm{I}$ satisfies $\mathbf{x} = \mathbf{x}(\mathrm{I}; \mathbf{b})$; in particular, several different bases may give rise to the same (degenerate) basic feasible solution.

Geometrically, basic feasible solutions are precisely extreme points (vertices) of the feasible set of (1) (Bertsimas and Tsitsiklis, 1997, Theorem 2.3); we will therefore use the terms basic feasible solution and vertex interchangeably in what follows. Our justification for focusing on basic feasible solutions is the "fundamental theorem of linear programming" (Bertsimas and Tsitsiklis, 1997, Theorem 2.7), which ensures that if any optimal solution to (1) exists, then there exists an optimum which is a basic feasible solution.

We denote by $\mathcal{I}(\mathbf{b})$ the set of all bases I for which $\mathbf{x}(\mathrm{I}; \mathbf{b})$ is a basic feasible solution, and by $\mathcal{I}^*(\mathbf{b})$ the set of all bases I for which $\mathbf{x}(\mathrm{I}; \mathbf{b})$ is an optimal solution. Concretely, $\mathcal{I}(\mathbf{b})$ or $\mathcal{I}^*(\mathbf{b})$ is a set of sets: each of its elements is one basis. The set of optimal vertices of (1) is defined by

$$\mathbf{V}^*(\mathbf{b}) \coloneqq \{\mathbf{x}(\mathrm{I}; \mathbf{b}) : \mathrm{I} \in \mathcal{I}^*(\mathbf{b})\}\,. \tag{6}$$

The general theory of polyhedral geometry implies that since $\mathbf{x}^*(\mathbf{b})$ is bounded, we may write $\mathbf{x}^*(\mathbf{b}) = \mathrm{conv}(\mathbf{V}^*(\mathbf{b}))$, the convex hull of $\mathbf{V}^*(\mathbf{b})$. Moreover, the assumption that $\mathbf{x}^*(\mathbf{b})$ is bounded implies that $\mathbf{x}^*(\mathbf{b}')$ is bounded for all perturbations $\mathbf{b}'$.[1] We therefore also have $\mathbf{x}^*(\mathbf{b}') = \mathrm{conv}(\mathbf{V}^*(\mathbf{b}'))$.

We summarize the main notation used in this paper in Table 1.

1. This follows from the fact that $\mathbf{x}^*(\mathbf{b})$ and $\mathbf{x}^*(\mathbf{b}')$ are polyhedra with the same recession cone, which must equal $\{\mathbf{0}\}$ since $\mathbf{x}^*(\mathbf{b})$ is bounded.

Symbol	Meaning
$f(\mathbf{b})$	Optimal objective value of (1)
$\mathbf{b}$, $\mathbf{b}_n$	True and random right-hand side constraints, (2)
$S(\mathbf{x})$	Set of nonzero coordinates of $\mathbf{x}$
$\mathbf{x}^*(\mathbf{b})$	Set of optimal solutions, (3)
$\mathbf{x}(\mathrm{I}; \mathbf{b})$	Basic feasible solution, (5)
$\mathcal{I}(\mathbf{b}), \mathcal{I}^*(\mathbf{b})$	Set of feasible and optimal bases
$\mathbf{V}^*(\mathbf{b})$	Extreme points of $\mathbf{x}^*(\mathbf{b})$, (6)
h_K	Support function of K, (9)

Table 1: Important notation

3. Vertex and base stability

This section presents two stability results that are central to our analysis. Though simple and likely well known, we present them explicitly here to highlight the important role they play in our theorems.

The first is a Lipschitzian property of polytopes due to Walkup and Wets (1969), which shows that the set of optimal solutions of Eq. (1) is Lipschitz with respect to the Hausdorff distance.

Proposition 2 *Under Assumption 1, there exists a constant* $C = C(\mathbf{A}, \mathbf{c}) > 0$ *such that if* $\mathbf{b}_1, \mathbf{b}_2 \in \mathbb{R}^k$ *are such that* $\mathbf{x}^*(\mathbf{b}_1)$ *and* $\mathbf{x}^*(\mathbf{b}_2)$ *are nonempty, then* $\rho_H(\mathbf{x}^*(\mathbf{b}_1), \mathbf{x}^*(\mathbf{b}_2)) \leq C\|\mathbf{b}_1 - \mathbf{b}_2\|$.

The second proposition shows that optimal bases for $\mathbf{b}'$ are also optimal for $\mathbf{b}$.

Proposition 3 *Under Assumption 1, there exists* $\delta = \delta(\mathbf{A}, \mathbf{b}) > 0$ *such that if* $\|\mathbf{b}' - \mathbf{b}\| \leq \delta$*, then* $\mathcal{I}^*(\mathbf{b}')$ *is nonempty and* $\mathcal{I}^*(\mathbf{b}') \subseteq \mathcal{I}^*(\mathbf{b})$.

4. A tractable limiting distribution when the target solution is unique

In this section, we first consider the simplified setting where the target solution $\mathbf{x}^*(\mathbf{b})$ is unique. Even under this simplification, however, the limiting distribution obtained by Klatt et al. (2022) does not have a tractable form. In particular, it is not even clear whether it is possible to generate samples from this distribution in polynomial time. The goal of this section is to obtain an expression for the limiting distribution that can be computed efficiently.

Stating this result requires defining a notion of distributional convergence suitable for a random set. Even when $|\mathbf{x}^*(\mathbf{b})| = 1$, it is possible that $|\mathbf{x}^*(\mathbf{b}_n)| > 1$. This situation can arise when $\mathbf{x}^*(\mathbf{b})$ is *unique but degenerate*, i.e., when there exist multiple optimal bases in $\mathcal{I}^*(\mathbf{b})$. Even if these bases all give rise to the same solution $\mathbf{x}^*(\mathbf{b})$ in the original program, they can correspond to *different* optimal solutions when $\mathbf{b}$ is replaced by $\mathbf{b}_n$. In this situation, $|\mathbf{x}^*(\mathbf{b}_n)| > 1$, and it is not possible to formulate a distributional limit for $r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b}))$ viewed as the difference of two vectors in $\mathbb{R}^m$. However, when $|\mathbf{x}^*(\mathbf{b})| = 1$, we can consider the set defined by translating the elements of $\mathbf{x}^*(\mathbf{b}_n)$ by $\mathbf{x}^*(\mathbf{b})$ and rescaling them by r_n:

$$r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b})) := \{r_n(\mathbf{x} - \mathbf{x}^*(\mathbf{b})) : \mathbf{x} \in \mathbf{x}^*(\mathbf{b}_n)\} \subseteq \mathbb{R}^m .$$

Our first main result is that this random set enjoys a *set-valued* distributional limit, with limit equal to the distribution of the optimal set of a random auxiliary linear program.[2]

Theorem 4 *Suppose that Eq. (1) satisfies Assumption 1. If $\mathbf{b}_n$ satisfies the distributional limit Eq. (2) and $|\mathbf{x}^*(\mathbf{b})| = 1$, then*

$$r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b})) \xrightarrow{D} \mathbf{p}^*_{\mathbf{b}}(\mathbb{G})\,, \tag{7}$$

where $\mathbf{p}^*_{\mathbf{b}}(\mathbb{G})$ is the set of optimal solutions to the following linear program:

$$\min\langle \mathbf{c}, \mathbf{p}\rangle : \mathbf{A}\mathbf{p} = \mathbb{G}, \quad \mathbf{p}_i \geq 0 \quad \forall i \notin S(\mathbf{x}^*(\mathbf{b}))\,. \tag{8}$$

The linear program Eq. (8) is a random optimization problem, with a random feasible set. For each ω in the background probability space Ω, or equivalently, for each realization of $\mathbb{G}$, Eq. (8) is a *bona fide* linear program.

The continuous mapping theorem implies that continuous functionals of the set $r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b}))$ also enjoy weak convergence. To give a concrete example of the statistical implications of this fact, consider the problem of obtaining a confidence set for $\mathbf{x}^*(\mathbf{b})$. Doing so requires knowing how far $\mathbf{x}^*(\mathbf{b})$ typically is from $\mathbf{x}^*(\mathbf{b}_n)$. If we let $d(S, \mathbf{x}) = \inf_{\mathbf{y}\in S} \|\mathbf{y} - \mathbf{x}\|$, then the following corollary shows that we can obtain a distributional limit for $d(\mathbf{x}^*(\mathbf{b}_n), \mathbf{x}^*(\mathbf{b}))$.

Corollary 5 $r_n d(\mathbf{x}^*(\mathbf{b}_n), \mathbf{x}^*(\mathbf{b})) \xrightarrow{D} d(\mathbf{p}^*_{\mathbf{b}}(\mathbb{G}), \mathbf{0})$.

In words, the rescaled distance of the target solution $\mathbf{x}^*(\mathbf{b})$ to the set of optimal solutions of the random program converges in distribution to the distance of zero to the optimal set of the random auxiliary LP. Importantly, this is a convex program, whose solution can be found in polynomial time.

Let us compare Corollary 5 with what would be obtained by a more standard approach. If one finds estimators by solving an optimization problem that can yield multiple optima, a standard path to inference consists in first identifying a subset of them that are close to one another, and then deriving the limiting distribution of any one of them, relative to the unique target. By contrast, Corollary 5 gives information about the distance of $\mathbf{x}^*(\mathbf{b})$ to the whole set of optima for the random program.

We stress our limit law is equivalent to the one obtained by Klatt et al. (2022, Theorem 3.5). The benefit of Theorem 4 is that $\mathbf{p}^*_{\mathbf{b}}(\mathbb{G})$ is given explicitly: though this set can be large, it is algorithmically accessible since it possesses an explicit polyhedral representation in terms of separating hyperplanes. This implies, for instance, that it is possible to solve convex optimization problems involving $\mathbf{p}^*_{\mathbf{b}}(\mathbb{G})$ in polynomial time via the ellipsoid method. We can therefore generate samples from the limiting distribution of (functionals of) the solution set in polynomial time, assuming $\mathbf{b}$ and the limiting distribution of $r_n(\mathbf{b}_n - \mathbf{b})$ are known. Specifically, we can obtain such a sample by 1) generating a sample $\mathbb{G}$ from the known limiting distribution of $r_n(\mathbf{b}_n - \mathbf{b})$ (in many applications, this will be a Gaussian distribution) and 2) solving the linear program Eq. (8), with that particular sample $\mathbb{G}$ used to form the constraint $\mathbf{A}\mathbf{p} = \mathbb{G}$ (this can be done by a standard efficient linear programming algorithm). Given an oracle to generate samples $\mathbb{G}$ and a linear programming algorithm,

2. To define weak convergence in this setting, we view these random sets as random elements in the metric space of compact subsets of $\mathbb{R}^m$ equipped with the Hausdorff distance, and weak convergence means, as usual, the convergence of expectations of bounded, continuous functions in this topology (King, 1989; Molchanov, 2005).

therefore, we obtain a polynomial-time oracle to generate samples from the limiting distribution of $r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b}))$. Such samples can be used, for instance, to obtain Monte Carlo estimates for calibrating asymptotic hypothesis tests and confidence sets.

On the other hand, Klatt et al. prove the same result but where the expression on the right side is a sum over a decomposition of $\mathbb{R}^m$ into a possibly exponential number of pieces. Even assuming $\mathbf{b}$ is known and $\mathbb{G}$ can efficiently be sampled from, such a decomposition typically cannot be evaluated in polynomial time. We emphasize that by formulating the limit as a linear program itself, we leverage the algorithmic theory of linear programming to sidestep the inefficient enumeration over the elements of this decomposition.

When the unique optimal solution $\mathbf{x}^*(\mathbf{b})$ is also non-degenerate, then Proposition 3 implies that for $\mathbf{b}_n$ sufficiently close to $\mathbf{b}$, the perturbed linear program also possesses a unique solution. In this situation, Theorem 4 is a *standard* distributional limit: asymptotically almost surely, the set $\mathbf{x}^*(\mathbf{b}_n)$ reduces to a singleton, and Theorem 4 shows that the distributional limit of the vector $r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b}))$ is the (unique) solution to Eq. (8), which is just $\mathbf{x}(\mathrm{I}^*; \mathbb{G})$ for the unique $\mathrm{I}^* \in \mathcal{I}^*(\mathbf{b})$. This recovers the limit for this simplified setting mentioned by Klatt et al. (2022, discussion after Remark 3.2).

5. Distributional convergence in the space of convex sets

When $\mathbf{x}^*(\mathbf{b})$ is not unique, the approach to defining a set-valued distributional limit taken in Theorem 4 no longer succeeds. Indeed, if $\mathbf{x}^*(\mathbf{b}_n)$ and $\mathbf{x}^*(\mathbf{b})$ are general closed sets, then even if $\mathbf{x}^*(\mathbf{b}_n) \to \mathbf{x}^*(\mathbf{b})$ in Hausdorff distance, the set

$$\mathbf{x}^*(\mathbf{b}_n) \ominus \mathbf{x}^*(\mathbf{b}) \coloneqq \{\mathbf{x} - \mathbf{x}' : \mathbf{x} \in \mathbf{x}^*(\mathbf{b}_n), \mathbf{x}' \in \mathbf{x}^*(\mathbf{b})\}$$

will not converge to $\{\mathbf{0}\}$ in general, so that no meaningful limit of $r_n(\mathbf{x}^*(\mathbf{b}_n) \ominus \mathbf{x}^*(\mathbf{b}))$ exists. In the non-unique case, Klatt et al. (2022) therefore define a distributional limit under the additional assumption that there exists a consistent scheme for selecting a single element of $\mathbf{x}^*(\mathbf{b}_n)$ and $\mathbf{x}^*(\mathbf{b})$; they then show that this selection satisfies a distributional limit in the classical sense. This ingenious approach captures the behavior of practical algorithms for solving LPs, since reasonable LP solvers give rise to such selection schemes (see Klatt et al., 2022, Lemma 5.5). However, as in the case where the target solution is unique, their limiting distribution is expressed as a sum over a decomposition of $\mathbb{R}^m$ into a possibly exponential number of pieces. Moreover, their techniques do not give insight into the overall fluctuations of the random set $\mathbf{x}^*(\mathbf{b}_n)$. By contrast, in the unique case, Theorem 4 shows that it is possible to obtain simultaneous control over the whole random set.

In this section, we leverage techniques from random convex geometry to obtain similar results for the non-unique case. Unlike Theorem 4, Theorem 6 goes beyond the setting analyzed by Klatt et al. (2022). Like Theorem 4, we state our convergence results in terms of the optimal solutions to a random auxiliary LP, implying that evaluating the limits we obtain can be computationally tractable in applications.

To formulate our distributional limit, we adopt a strategy developed by Artstein and Vitale (1975), Weil (1982), and independently by Lyashenko (1983) to prove central limit theorems for random compact sets. To any compact, convex set $K \subseteq \mathbb{R}^m$, we associate its *support function* $\mathsf{h}_K : \mathbb{S}^{m-1} \to \mathbb{R}$ defined by

$$\mathsf{h}_K(\alpha) \coloneqq \sup_{\mathbf{x} \in K} \langle \alpha, \mathbf{x} \rangle . \tag{9}$$

The mapping $K \mapsto \mathsf{h}_K$ provides an *isometric embedding* of the metric space of convex, compact sets equipped with the Hausdorff metric into the Banach space $\mathcal{C}(\mathbb{S}^{m-1})$ of continuous functions on the sphere equipped with the uniform norm (see Molchanov, 2005, section 3.1.2). Explicitly, given two compact, convex sets K_1 and K_2, we have

$$\rho_H(K_1, K_2) = \sup_{\alpha \in \mathbb{S}^{m-1}} |\mathsf{h}_{K_1}(\alpha) - \mathsf{h}_{K_2}(\alpha)| . \tag{10}$$

In particular, the map from a convex set to its support function is injective; K can be recovered from h_K by taking its Legendre transform. This embedding has two profound implications. First, the geometry of convex sets is entirely captured by their support functions. In particular, we may associate to a random convex set its support function, viewed as a random element of $\mathcal{C}(\mathbb{S}^{m-1})$, and study its distribution instead.[3] Second, since $\mathcal{C}(\mathbb{S}^{m-1})$ is a Banach space, we may leverage the theory of probability in Banach spaces to prove limit theorems for support functions.

Our main result of this section is a distributional limit for the set $\mathbf{x}^*(\mathbf{b}_n)$. We will focus on the support functions of $\mathbf{x}^*(\mathbf{b}_n)$ as defined in Eq. (9). Once again, it is stated in terms of the solutions to an auxiliary linear program.

Theorem 6 *Let h_n and h be the support functions of $\mathbf{x}^*(\mathbf{b}_n)$ and $\mathbf{x}^*(\mathbf{b})$, respectively. Suppose that* (1) *satisfies Assumption 1. If $\mathbf{b}_n$ satisfies the distributional limit* (2)*, then*

$$r_n(\mathsf{h}_n - \mathsf{h}) \xrightarrow{D} \mathsf{g}_{\mathbb{G}} , \tag{11}$$

where $\mathsf{g}_{\mathbb{G}}$ is the random element of $\mathcal{C}(\mathbb{S}^{m-1})$ defined by

$$\mathsf{g}_{\mathbb{G}}(\alpha) = \sup_{\mathbf{x} \in \mathbf{q}^*_\alpha(\mathbb{G})} \langle \alpha, \mathbf{x} \rangle ,$$

and $\mathbf{q}^*_\alpha(\mathbb{G})$ is the set of optimal vertex solutions to the following linear program:[4]

$$\min \langle \mathbf{c}, \mathbf{q} \rangle : \mathbf{A}\mathbf{q} = \mathbb{G}, \quad \mathbf{q}_i \geq 0 \quad \forall i \notin S(\nabla \mathsf{h}(\alpha)) . \tag{12}$$

Informally, Theorem 6 shows that when n is large, $\mathsf{h}_n \overset{d}{\approx} \mathsf{h} + r_n^{-1}\mathsf{g}_{\mathbb{G}}$. By the isometry described in Eq. (10), this translates into a statement about the fluctuations of the random set $\mathbf{x}^*(\mathbf{b}_n)$ around $\mathbf{x}^*(\mathbf{b})$. The proof of Theorem 6 is based on establishing the directional Hadamard differentiability of the mapping $\mathbf{b} \mapsto \mathsf{h}_{\mathbf{x}^*(\mathbf{b})}$ viewed as a function from $\mathbb{R}^k$ to $\mathcal{C}(\mathbb{S}^{m-1})$, and then applying a functional delta method due to Römisch (2006).

Like Theorem 4, Theorem 6 has statistical implications for the problem of obtaining a confidence set for $\mathbf{x}^*(\mathbf{b})$. The isometry (10) implies the following analogue of Corollary 5.

Corollary 7 $r_n \rho_H(\mathbf{x}^*(\mathbf{b}_n), \mathbf{x}^*(\mathbf{b})) \xrightarrow{D} \sup_{\alpha \in \mathbb{S}^{m-1}} |\mathsf{g}_{\mathbb{G}}(\alpha)|$

3. We omit a detailed discussion of measurability here, but it can be shown that if the space of convex, compact subsets of $\mathbb{R}^m$ is equipped with an appropriate σ-algebra (known as the Effros σ-algebra), then for a random set K the support function h_K is indeed a random variable in $\mathcal{C}(\mathbb{S}^{m-1})$ (see Molchanov, 2005, Proposition 2.5).
4. The function h is only differentiable almost everywhere, but since $\mathsf{g}_{\mathbb{G}}(\alpha)$ is almost surely continuous it suffices to specify its values on a dense subset.

In other words, the rescaled Hausdorff distance between the solution sets converges in distribution to the supremum of a random continuous function on the sphere. Corollary 7 can be compared to (Klatt et al., 2022, Proposition 3.7), which shows that $\rho_H(\mathbf{x}^*(\mathbf{b}_n), \mathbf{x}^*(\mathbf{b})) = O_P(r_n^{-1})$. Our result gives finer control over the behavior of the rescaled distance in terms of the solutions to auxiliary linear programs. However, unlike Corollary 5, we are not aware of an algorithm that can compute the supremum on the right side of Corollary 7 in polynomial time. Finding a computationally tractable expression for this limit is an attractive open problem.

6. Data-driven confidence sets

Theorems 4 and 6 give explicit distributional limits for $\mathbf{x}^*(\mathbf{b}_n)$ in terms of auxiliary linear programs. Though evaluating these limits is computationally tractable, they fail to be suitable for concrete inference tasks because the limiting distributions depend on properties of the true optimal solution set $\mathbf{x}^*(\mathbf{b})$. Since this set is almost always unknown in practice, Theorems 4 and 6 do not provide a data-driven way to obtain asymptotically valid confidence sets.

In principle, the fact that Theorem 6 is proven by directional Hadamard differentiability arguments implies that the m-out-of-n bootstrap is consistent (Dümbgen, 1993). However, using the bootstrap for inference raises other practical difficulties: it is an open question how to choose m for good performance, and convergence is slow. Therefore, even though Theorems 4 and 6 provide a complete answer to the theoretical question of obtaining a valid distributional limit, they are a poor way to construct confidence sets in practice.

In this section, we give a simple procedure to obtain such sets. Specifically, we suppose that that statistician has solved the perturbed linear program and obtained a random solution $\hat{\mathbf{x}}_n \in \mathbf{V}^*(\mathbf{b}_n)$ along with a corresponding basis $\mathrm{I}_n \in \mathcal{I}^*(\mathbf{b}_n)$.[5] We will construct a confidence set based on $\hat{\mathbf{x}}_n$ that is guaranteed to contain at least one element of $\mathbf{x}^*(\mathbf{b})$ with high probability. Specifically, let us consider the basic solution $\mathbf{x}(\mathrm{I}_n; \mathbf{b})$ defined by the random basis I_n. This solution may not be feasible for (1), much less optimal, so we define the projection

$$\bar{\mathbf{x}}_n^* \coloneqq \operatorname*{argmin}_{\mathbf{x} \in \mathbf{x}^*(\mathbf{b})} \|\mathbf{x}(\mathrm{I}_n; \mathbf{b}) - \mathbf{x}\| . \tag{13}$$

The following result shows that we can construct a set containing this point with high probability.

Theorem 8 *Suppose that (1) satisfies Assumption 1 and $\mathbf{b}_n$ satisfies the distributional limit (2). Let G_α be an open set such that $\mathbb{P}\{\mathbb{G} \in G_\alpha\} \geq 1 - \alpha$. Then*

$$\liminf_{n \to \infty} \mathbb{P}\left(r_n(\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_n^*) \in \mathbf{x}(\mathrm{I}_n; G_\alpha)\right) \geq 1 - \alpha\,, \tag{14}$$

where $\mathbf{x}(\mathrm{I}_n; G_\alpha) \coloneqq \{\mathbf{x}(\mathrm{I}_n; \mathbf{G}) : \mathbf{G} \in G_\alpha\}$.

Corollary 9 (Confidence set for an optimal solution) *In the setting of Theorem 8, the set $C_n \coloneqq \{\hat{\mathbf{x}}_n - r_n^{-1}\mathbf{x} : \mathbf{x} \in \mathbf{x}(\mathrm{I}_n; G_\alpha)\}$ contains an element of $\mathbf{x}^*(\mathbf{b})$ with asymptotic probability at least $1 - \alpha$.*

5. Algorithms such as the simplex method always return an optimal vertex when one exists, along with a corresponding basis (Bertsimas and Tsitsiklis, 1997, Theorem 3.3).

Theorem 8 and Corollary 9 are weaker than Theorems 4 and 6: they do not give any information about the whole set of optimal solutions $\mathbf{x}^*(\mathbf{b})$. Instead, Corollary 9 only guarantees that C_n contains *an* optimal solution with high probability. As our simulations in Section 7 show, when the optimal solution is non-unique, the confidence sets constructed by this procedure sometimes cover one solution, sometimes another. Nevertheless, Theorem 9 does offer the practitioner an asymptotic guarantee that *some* optimal solution is in a neighborhood of the estimator.

On the other hand, unlike Theorems 4 and 6, Corollary 9 is eminently practical. It requires only the outputs $\hat{\mathbf{x}}_n$ and I_n from a standard linear programming algorithm, and the set $\mathbf{x}(\mathrm{I}_n : G_\alpha)$ is easy to compute, since the mapping $\mathbf{G} \mapsto \mathbf{x}(\mathrm{I}_n : \mathbf{G})$ is an explicit linear transformation. For instance, if G_α is an ellipsoid of the form $\{\mathbf{y} \in \mathbb{R}^k : \mathbf{y}^\top \Sigma^{-1} \mathbf{y} < 1\}$, then recalling definition (5) in Section 2 we have

$$\mathbf{x}(\mathrm{I}_n; G_\alpha) = \{\mathbf{x} \in \mathbb{R}^m : \mathbf{x}_{\mathrm{I}_n}^\top \mathbf{M}_n \mathbf{x}_{\mathrm{I}_n}^\top < 1,\, \mathbf{x}_{\mathrm{I}_n^c} = \mathbf{0}\}\,, \tag{15}$$

where $\mathbf{M}_n := \mathbf{A}_{\mathrm{I}_n}^\top \Sigma^{-1} \mathbf{A}_{\mathrm{I}_n} \in \mathbb{R}^{k \times k}$.

7. Examples

We will provide two examples in this section to show the effectiveness of the method described in Theorem 8 and Corollary 9 for generating a confidence set for solutions to LPs.

We first return to the simple discrete optimal transport problem described in the introduction, which is a linear program with a unique degenerate optimal vertex. We then treat a more complicated example arising from a min-cost flow problem (see Bradley et al., 1977, section 8.1). In this example, there are two optimal vertex solutions at the population level. In both cases, our simulations confirm that the method gives confidence sets which cover an optimal solution with high probability.

7.1. Empirical Optimal Transport

We consider the optimal transport example given in the introduction, where we suppose that $n\mathbf{q}_n \sim \mathrm{Mult}(n, (1/2, 1/2))$. This corresponds to the situation where we aim to estimate the solution to an optimal transport problem involving an unknown distribution $\mathbf{q} = (1/2, 1/2)$ on the basis of n i.i.d. samples from $\mathbf{q}$. In this setting, the classical central limit theorem implies $\sqrt{n}(\mathbf{q}_n - \mathbf{q}) \overset{D}{\to} (Z, -Z)$, where $Z \sim \mathcal{N}(0, 1/4)$. We therefore choose $G_\alpha = \{(x, -x) : x \in [-z_{0.025}/2, z_{0.025}/2]\}$, where $[-z_{0.025}/2, z_{0.025}/2]$ is a 95% confidence interval for an $\mathcal{N}(0, 1/4)$ random variable, and use Corollary 9 to construct a confidence set for the entries of π.

Figure 1 shows examples of the confidence intervals produced by our method. We plot one realization for each of the labeled values of n. Note that for each realization, the confidence intervals for two (random) entries of π are singletons: for example, when $n = 20$, the solution we obtained to the LP was $\hat{\pi}_n = (0.5, 0.05; 0, 0.45)$ and the confidence intervals given by Corollary 9 were $\pi_{11} = 0.5$, $\pi_{12} \in [-0.169, 0.269]$, $\pi_{21} = 0$ and $\pi_{22} \in [0.23, 0.67]$. Even though the confidence intervals for π_{11} and π_{21} have zero width, this set does in fact contain the optimal solution $(\frac{1}{2}, 0; 0, \frac{1}{2})$. The somewhat counterintuitive fact that a confidence set with empty interior covers the true parameter with probability approaching 95% is a consequence of the fact that the distribution of $\hat{\pi}_n$ is *not* absolutely continuous with respect to the Lebesgue measure.

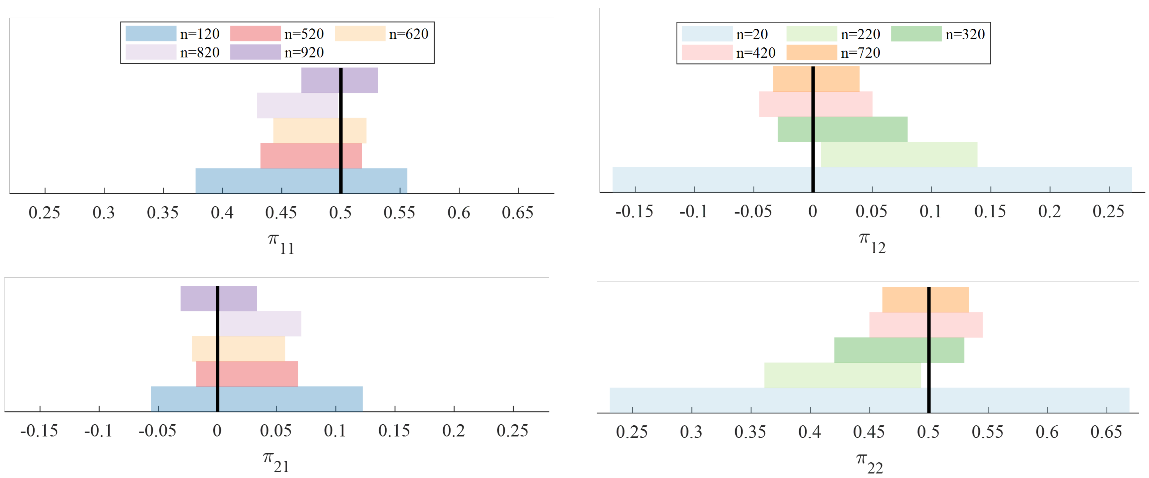

Figure 1: Example confidence intervals for $\pi = (\frac{1}{2}, 0; 0, \frac{1}{2})$ computed with different values of n (one replicate each). For the values of n appearing in the box on the left, the confidence intervals for π_{12} and π_{22} were singletons at 0 and $1/2$, respectively; for the values of n appearing on the right, the confidence intervals for π_{11} and π_{12} were singletons.

We also estimate the observed coverage probabilities for finite n. For each n, we generate 1000 independent replicates, calculate the 95% confidence intervals and count the replicates that successfully capture a true solution.

n	1	3	5	10	50	100	500	10000
Coverage Probability	0.480	0.892	0.941	0.981	0.935	0.922	0.947	0.950

7.2. Minimal Cost Flow Problem

We adapt an example from Bradley et al. (1977, Section 8.1) arising in operations research. Consider the problem of moving goods from origins to destinations along routes with certain volume constraints and costs. We model an instance of this problem as the directed graph depicted in Fig. 2, with 5 nodes and 9 arcs. Each arc is unidirectional, labeled with its capacity and transportation cost (the pair of numbers in the parentheses adjacent to the arc). Each node is labeled with its supply or demand. For example, the supply of node 1 is 20. The arc x_{12} transports products from node 1 to node 2 with the maximum capacity of 15 units of product and the cost \$4 per unit of product. Assuming that the total demand matches the total supply, the goal is to fulfill all the demands in the network at a minimum cost.

This minimal-cost flow problem can be written in a linear program form:

$$\min \sum_{i,j} c_{ij} x_{ij} : \sum_j x_{ij} - \sum_k x_{ki} = b_i \ (i = 1, 2, ..., 5), \ 0 \le x_{ij} \le u_{ij}, \tag{16}$$

where b_i is the supply of each node, u_{ij} is the capacity of each arc, and c_{ij} is the transportation cost of each arc. A standard linear program in the form of Eq. (1) can be obtained for this problem by introducing the auxiliary variable y_{ij}, which satisfies $y_{ij} + x_{ij} = u_{ij}$ and $y_{ij} \ge 0$. The auxiliary

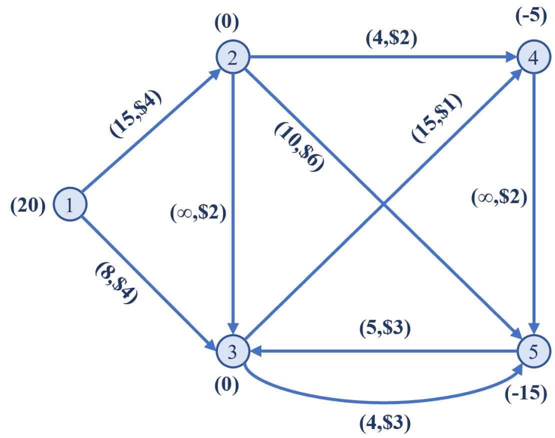

Figure 2: Minimal-cost flow problem. Each arc is labeled with its *capacity* (the total amount of flow it can carry) and the cost of moving a single unit of flow across it. Vertices are labeled with supplies (positive quantities) or demands (negative quantities) for goods at each location.

variable y_{ij} represents the remaining capacity for each arc. The standard form for Eq. (16) is

$$\min \sum_{i,j} c_{ij}x_{ij} : \sum_j x_{ij} - \sum_k x_{ki} = b_i, \quad y_{ij} + x_{ij} = u_{ij},\ x_{ij} \geq 0,\ y_{ij} \geq 0. \tag{17}$$

Note that the equality constraints $\sum_j x_{ij} - \sum_k x_{ki} = b_i$ are redundant due to the flow balance condition of the network, and deleting any one of them will not change the program. Suppose the flow balance constraint on the third node is deleted and we have the modified supply vector $\tilde{\mathbf{b}} = (b_1, b_2, b_4, b_5) = (20, 0, -5, -15)$.

The program in Fig. 2 has two optimal vertex solutions:

	x_{12}	x_{13}	x_{23}	x_{24}	x_{25}	x_{34}	x_{35}	x_{45}	x_{53}
solution 1	12	8	8	4	0	15	1	14	0
solution 2	12	8	8	4	0	12	4	11	0

In applications, the true supply and demand at each node may not be known precisely, but rather must be estimated by an empirical supply vector $\tilde{\mathbf{b}}_n$ obtained by averaging the observed supplies and demands over n days. Suppose that we know $\sqrt{n}(\tilde{\mathbf{b}}_n - \tilde{\mathbf{b}}) \xrightarrow{D} \mathbb{G}_0$, where $\mathbb{G}_0 \sim \mathcal{N}(0, \mathrm{diag}(4, 1, 1, 3))$. We calculate a min-cost flow $\hat{\mathbf{x}}_n$ using the estimated supply vector $\tilde{\mathbf{b}}_n$, and employ Corollary 9 to build a confidence set.

To visualize the confidence set for $\hat{\mathbf{x}}_n$ for various n, we show the projection of 4 dimensional confidence sets to lower dimensional spaces. As an example, we plot the confidence interval for the x_{45} coordinate (Fig. 3, Fig. 4) and the confidence set for the 2 dimensional arc pair (x_{23}, x_{45}) (Fig. 5). In Fig. 3, we show several examples of the confidence sets we obtain for x_{45}. We plot a single realization for each value of n. Figure 4 shows many replicates for the $n = 50$ case to illustrate the sampling variability of the sets we construct, and Fig. 5 depicts the same procedure for the two-dimensional confidence set for (x_{23}, x_{45}). We can see that for each replicate, the given confidence sets capture *one* of the solutions very well—which solution is covered depends on the random fluctuations in each replicate.

In short, Corollary 9 gives a practical means of obtaining asymptotically valid confidence sets for the solution to a linear program. To our knowledge, this is the first procedure satisfying these requirements.

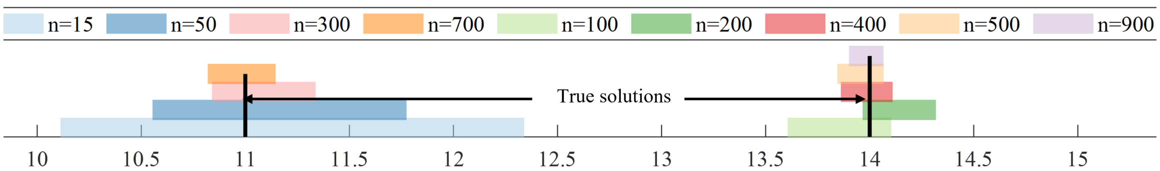

Figure 3: Example confidence intervals for flow through arc x_{45} computed with different values of n.

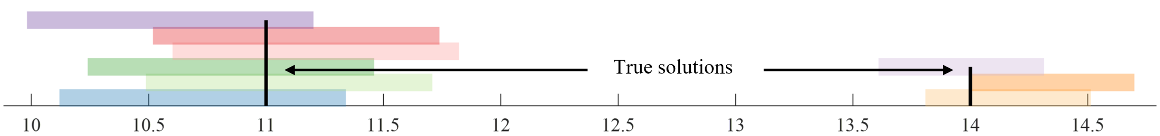

Figure 4: Confidence intervals for flow through arc x_{45} when $n = 50$ (many replicates).

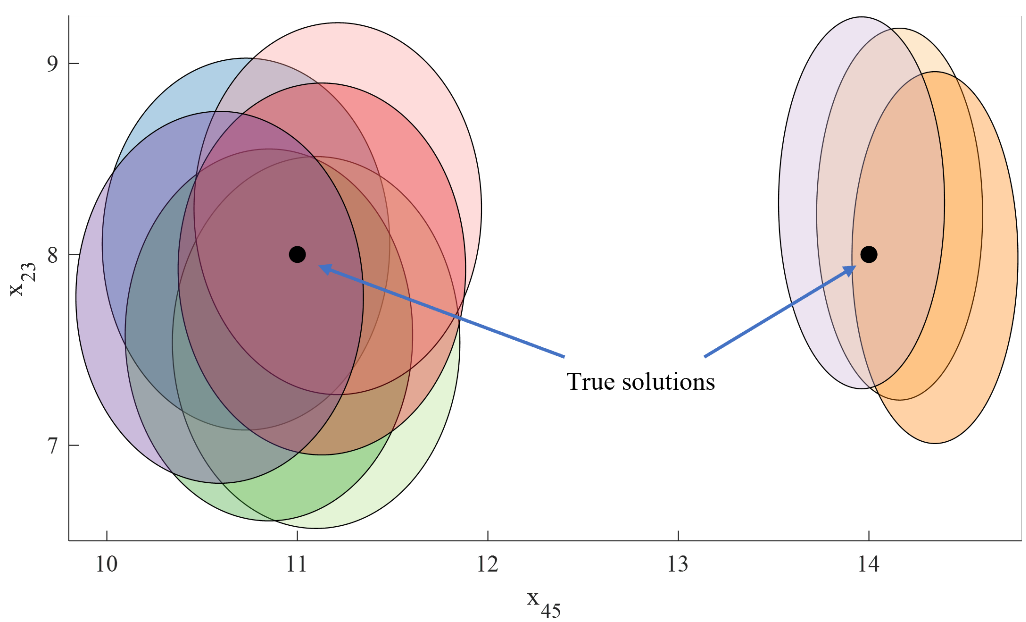

Figure 5: Confidence sets for flow through the arc pair (x_{23}, x_{45}) when $n = 50$ (many replicates).

Acknowledgments

Bunea was supported in part by NSF grant DMS-2210563, and Niles-Weed was supported in part by NSF grant DMS-2210583 and a Sloan Research Fellowship.

References

J. Aitchison and S. D. Silvey. Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.*, 29:813–828, 1958. ISSN 0003-4851. doi: 10.1214/aoms/1177706538. URL https://doi.org/10.1214/aoms/1177706538.

Donald W. K. Andrews. Generalized method of moments estimation when a parameter is on a boundary. volume 20, pages 530–544. 2002. doi: 10.1198/073500102288618667. URL https://doi.org/10.1198/073500102288618667. Twentieth anniversary GMM issue.

Zvi Artstein and Richard A Vitale. A strong law of large numbers for random compact sets. *The Annals of Probability*, pages 879–882, 1975.

Dimitris Bertsimas and John Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1st edition, 1997. ISBN 1886529191.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004. ISBN 0521833787.

Stephen P Bradley, Arnoldo C Hax, and Thomas L Magnanti. *Applied mathematical programming*. Addison-Wesley, 1977.

Herman Chernoff. On the distribution of the likelihood ratio. *Ann. Math. Statistics*, 25:573–578, 1954. ISSN 0003-4851. doi: 10.1214/aoms/1177728725. URL https://doi.org/10.1214/aoms/1177728725.

Lutz Dümbgen. On nondifferentiable functions and the bootstrap. *Probab. Theory Related Fields*, 95(1):125–140, 1993. ISSN 0178-8051. doi: 10.1007/BF01197342. URL https://doi.org/10.1007/BF01197342.

Jitka Dupačová and Roger Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.*, 16(4):1517–1549, 1988. ISSN 0090-5364. doi: 10.1214/aos/1176351052. URL https://doi.org/10.1214/aos/1176351052.

L. G. Hačijan. A polynomial algorithm in linear programming. *Dokl. Akad. Nauk SSSR*, 244(5): 1093–1096, 1979. ISSN 0002-3264.

Alan J. King. Generalized delta theorems for multivalued mappings and measurable selections. *Math. Oper. Res.*, 14(4):720–736, 1989. ISSN 0364-765X. doi: 10.1287/moor.14.4.720. URL https://doi.org/10.1287/moor.14.4.720.

Alan J. King and R. Tyrrell Rockafellar. Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.*, 18(1):148–162, 1993. ISSN 0364-765X. doi: 10.1287/moor.18.1.148. URL https://doi.org/10.1287/moor.18.1.148.

Marcel Klatt, Axel Munk, and Yoav Zemel. Limit laws for empirical optimal solutions in random linear programs. *Ann. Oper. Res.*, 315(1):251–278, 2022. ISSN 0254-5330. doi: 10.1007/s10479-022-04698-0. URL https://doi.org/10.1007/s10479-022-04698-0.

Jeff Linderoth, Alexander Shapiro, and Stephen Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142(1):215–241, 2006.

NN Lyashenko. Statistics of random compacts in euclidean space. *Journal of Soviet Mathematics*, 21:76–92, 1983.

Ilya Molchanov. *Theory of random sets*. Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2005. ISBN 978-185223-892-3; 1-85233-892-X.

Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, 2e edition, 2006.

B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30(4):838–855, 1992. ISSN 0363-0129. doi: 10.1137/0330046. URL https://doi.org/10.1137/0330046.

Werner Römisch. *Delta Method, Infinite Dimensional*. John Wiley & Sons, Ltd, 2006. ISBN 9780471667193. doi: https://doi.org/10.1002/0471667196.ess3139. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/0471667196.ess3139.

Alexander Shapiro. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30:169–186, 1991.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.

David W Walkup and Roger J-B Wets. A lipschitzian characterization of convex polyhedra. *Proceedings of the American Mathematical Society*, 23(1):167–173, 1969.

Wolfgang Weil. An application of the central limit theorem for banach-space-valued random variables to the theory of random sets. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 60(2):203–208, 1982.

Appendix A. Proofs of propositions

We first establish a few basic lemmas. In the proofs, we utilize the optimal conditions of the linear program Eq. (1) and its dual program:

$$\max_{\lambda \in \mathbb{R}^k} \langle \mathbf{b}, \lambda \rangle, \qquad \text{s.t. } \mathbf{c} - \mathbf{A^T}\lambda \geq 0. \tag{18}$$

The linear program Eq. (1) and the dual program Eq. (18) achieve their optima at $(\mathbf{x}^*(\mathbf{b}),\ \lambda^*(\mathbf{b}))$ if and only if $\exists\, \mathbf{s} \in \mathbb{R}^m$ such that:

$$\mathbf{A}^T\lambda^*(\mathbf{b}) + \mathbf{s} = \mathbf{c},\ \mathbf{A}\mathbf{x}^*(\mathbf{b}) = \mathbf{b},\ \mathbf{x}^*(\mathbf{b}) \geq 0,\ \mathbf{s} \geq 0,\ \mathbf{x}^*(\mathbf{b})^T\mathbf{s} = 0. \tag{19}$$

The last condition is called *complementary slackness*, and is equivalent to the condition that $\mathbf{x}^*(\mathbf{b})_i > 0 \implies \mathbf{s}_i = 0$ for all $i \in [m]$.

Lemma 10 *Under Assumption 1, there exists* $\delta = \delta(\mathbf{A}, \mathbf{b}) > 0$*,* $C_1 = C(\mathbf{A})$*, and* $C_2 = C(\mathbf{A}, \mathbf{c})$ *such that the following properties hold:*

1. *If* $\|\mathbf{b}' - \mathbf{b}\| \leq \delta$*, then* $\mathcal{I}(\mathbf{b}') \subseteq \mathcal{I}(\mathbf{b})$
2. *If* $\|\mathbf{b}' - \mathbf{b}\| \leq \delta$*, then* $\mathbf{x}^*(\mathbf{b}') \neq \emptyset$
3. $\|\mathbf{x}(I; \mathbf{b}') - \mathbf{x}(I; \mathbf{b})\| \leq C_1 \|\mathbf{b}' - \mathbf{b}\|$*, for all* $I \in \mathcal{I}(\mathbf{b}')$.
4. *If* $f(\mathbf{b}')$ *is finite, then* $|f(\mathbf{b}') - f(\mathbf{b})| \leq C_2 \|\mathbf{b}' - \mathbf{b}\|$,

Proof The perturbed LP with linear constraint $\mathbf{A}\mathbf{x} = \mathbf{b}'$ reads:

$$\min_{\mathbf{x} \in \mathbb{R}^m} \langle \mathbf{c}, \mathbf{x} \rangle, \qquad \text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}', \ \mathbf{x} \geq \mathbf{0}, \tag{20}$$

1. Inclusion of feasible bases: $\mathcal{I}(\mathbf{b}') \subseteq \mathcal{I}(\mathbf{b})$ If $\exists \mathrm{I}_0 \in \mathcal{I}(\mathbf{b}') \backslash \mathcal{I}(\mathbf{b})$, there exist $1 \leq p \leq k$ such that $(\mathbf{A}_{\mathrm{I}_0}^{-1} \mathbf{b}')_p \geq 0$ and $(\mathbf{A}_{\mathrm{I}_0}^{-1} \mathbf{b})_p < 0$. However, when $\|\mathbf{b}' - \mathbf{b}\| < \frac{|(\mathbf{A}_{I_0}^{-1}\mathbf{b})_p|}{\|\mathbf{A}_{I_0}^{-1}\|}$,

$$|(\mathbf{A}_{I_0}^{-1} \mathbf{b}')_p - (\mathbf{A}_{I_0}^{-1} \mathbf{b})_p| \leq \|\mathbf{A}_{I_0}^{-1} \mathbf{b}' - \mathbf{A}_{I_0}^{-1} \mathbf{b}\| < |(\mathbf{A}_{I_0}^{-1} \mathbf{b})_p|.$$

Therefore, $(\mathbf{A}_{I_0}^{-1} \mathbf{b}')_p < 0$. Take $\delta_{b_0} = \min_{\{I_0 | \mathbf{A}_{I_0} \text{ invertible}\}} \min_{p : (\mathbf{A}_{I_0}^{-1}\mathbf{b})_p < 0} \frac{|(\mathbf{A}_{I_0}^{-1}\mathbf{b})_p|}{\|\mathbf{A}_{I_0}^{-1}\|}$. When $\|\mathbf{b}' - \mathbf{b}\| < \delta_{b_0}$, there is no such basis I_0 and $\mathcal{I}(\mathbf{b}') \subseteq \mathcal{I}(\mathbf{b})$.

2. Existence of optimal solution $\mathbf{x}^*(\mathbf{b}')$**.** We first show that the perturbed LP is feasible.

By Assumption 1, there exists $\mathbf{x}_0$ satisfying

$$\mathbf{A}\mathbf{x}_0 = \mathbf{b}, \qquad \mathbf{x}_0 > 0\,.$$

Let $s_{\mathbf{x}_0}$ be the smallest entry of $\mathbf{x}_0$ and let I_0 be an arbitrary element of $\mathcal{I}(\mathbf{b})$. When $\|\mathbf{b}' - \mathbf{b}\| < \frac{s_{\mathbf{x}_0}}{\|\mathbf{A}_{I_0}^{-1}\|} := \delta_{b_1}$, we have

$$\|\mathbf{x}(I_0; \mathbf{b}' - \mathbf{b})\| = \|\mathbf{A}_{I_0}^{-1}(\mathbf{b}' - \mathbf{b})\| < s_{\mathbf{x}_0}.$$

Then $\mathbf{x}_0' := \mathbf{x}_0 + \mathbf{x}(I_0; \mathbf{b}' - \mathbf{b})$ satisfies $\mathbf{A}\mathbf{x}_0' = \mathbf{b}'$ and $\mathbf{x}_0' > 0$, which indicates that $\mathbf{x}_0'$ lies in the feasible region of Eq. (20).

The dual problem of Eq. (20) is

$$\max_{\lambda \in \mathbb{R}^k} \langle \mathbf{b}', \lambda \rangle, \qquad \text{s.t. } \mathbf{c} - \mathbf{A}^{\mathbf{T}} \lambda \geq 0. \tag{21}$$

The fact that $\mathbf{x}^*(\mathbf{b})$ is nonempty implies that Eq. (21) is feasible, since the feasible set of Eq. (21) does not depend on $\mathbf{b}'$. Hence the value of Eq. (20) is bounded and there exist optimal solutions.

3. Lipschitz continuity of basic feasible solutions. We argue as in Part 1. For any $\mathrm{I} \in \mathcal{I}(\mathbf{b}')$, $\mathbf{x}(\mathrm{I}; \mathbf{b}')_{\mathrm{I}^C} = \mathbf{x}(\mathrm{I}; \mathbf{b})_{\mathrm{I}^C} = \mathbf{0}$, and

$$\|(\mathbf{A}_{\mathrm{I}}^{-1} \mathbf{b}') - (\mathbf{A}_{\mathrm{I}}^{-1} \mathbf{b})\| = \|\mathbf{A}_{\mathrm{I}}^{-1} \mathbf{b}' - \mathbf{A}_{\mathrm{I}}^{-1} \mathbf{b}\| \leq \|\mathbf{A}_{\mathrm{I}}^{-1}\| \|\mathbf{b}' - \mathbf{b}\|\,.$$

Taking $C_1 = \max_{\mathrm{I} : \mathbf{A}_{\mathrm{I}} \text{ invertible}} \|\mathbf{A}_{\mathrm{I}}^{-1}\|$ yields the bound.

4. Local Lipschitz continuity of optimal value The fact that target and perturbed primal problems have finite values indicates that there exist optimal solutions to the target and perturbed dual problems. Denote optimal vertex solutions to the dual programs by $\lambda^*(\mathbf{b})$ and $\lambda^*(\mathbf{b}')$, respectively Strong duality implies

$$\langle \mathbf{b}', \lambda^*(\mathbf{b}') \rangle = f(\mathbf{b}') = \langle \mathbf{c}, \mathbf{x}^*(\mathbf{b}') \rangle.$$

Therefore, we have

$$\langle \mathbf{b}', \lambda^*(\mathbf{b}') \rangle - \langle \mathbf{b}, \lambda^*(\mathbf{b}) \rangle = \langle \mathbf{c}, \mathbf{x}^*(\mathbf{b}') - \mathbf{x}^*(\mathbf{b}) \rangle = f(\mathbf{b}') - f(\mathbf{b}).$$

Since $\lambda^*(\mathbf{b})$ and $\lambda^*(\mathbf{b}')$ are optimal vertices of Eq. (18) and Eq. (21), respectively, we obtain

$$\langle \mathbf{b}' - \mathbf{b}, \lambda^*(\mathbf{b}) \rangle \leq (f(\mathbf{b}') - f(\mathbf{b})) \leq \langle \mathbf{b}' - \mathbf{b}, \lambda^*(\mathbf{b}') \rangle.$$

Therefore

$$|f(\mathbf{b}') - f(\mathbf{b})| \leq \|\mathbf{b}' - \mathbf{b}\| \max_{\lambda \in \mathbf{\Lambda}} \|\lambda\|,$$

where $\mathbf{\Lambda}$ is the set of all vertices of the polytope $\mathbf{A}^\top \lambda \geq \mathbf{c}$.

■

We now turn to proofs of the propositions.

Proof [Proof of Proposition 2] We follow the same argument as is given in the proof of Proposition 3.7 in Klatt et al. (2022). If $\mathbf{x}^*(\mathbf{b}_1)$ and $\mathbf{x}^*(\mathbf{b}_2)$ are both nonempty, then Lemma 10, part 4, implies that $\|f(\mathbf{b}_1) - f(\mathbf{b}_2)\| \leq C_2 \|\mathbf{b}_1 - \mathbf{b}_2\|$. We then apply the main theorem of Walkup and Wets (1969) with K being the positive orthant and $\tau(\mathbf{x}) = (\mathbf{A}\mathbf{x}, \langle \mathbf{c}, \mathbf{x} \rangle)$. ■

Proof [Proof of Proposition 3] Let δ be small enough that Lemma 10 holds. Parts 1 and 2 of that lemma imply that $\emptyset \neq \mathcal{I}^*(\mathbf{b}') \subseteq \mathcal{I}(\mathbf{b}') \subseteq \mathcal{I}(\mathbf{b})$. It therefore suffices to show that $\mathrm{I}_0 \in \mathcal{I}^*(\mathbf{b})$ for all $\mathrm{I}_0 \in \mathcal{I}^*(\mathbf{b}')$.

Assume that $\mathbf{x}(\mathrm{I}_0; \mathbf{b}') \in \mathbf{x}^*(\mathbf{b}')$. Denote by λ_{I_0} an optimal dual solution to Eq. (21), which satisfies

$$\mathbf{A}^T \lambda_{I_0} + \mathbf{s} = \mathbf{c},\ \mathbf{A}\mathbf{x}(I_0; \mathbf{b}') = \mathbf{b}',\ \mathbf{x}(I_0; \mathbf{b}') \geq 0,\ \mathbf{s} \geq 0,\ \mathbf{x}(I_0; \mathbf{b}')^T \mathbf{s} = 0$$

for some $\mathbf{s} \in \mathbb{R}^m$. We will now show that $(\mathbf{x}(\mathrm{I}_0; \mathbf{b}), \lambda_{\mathrm{I}_0})$ is also an optimal primal-dual pair for the unperturbed program when δ is small enough.

The first four conditions still hold for $(\mathbf{x}(I_0; \mathbf{b}), \lambda_{I_0})$:

$$\mathbf{A}^T \lambda_{I_0} + \mathbf{s} = \mathbf{c},\ \mathbf{A}\mathbf{x}(I_0; \mathbf{b}) = \mathbf{b},\ \mathbf{x}(I_0; \mathbf{b}) \geq 0,\ \mathbf{s} \geq 0.$$

To show the complementary slackness condition holds, we use Part 3 of Lemma 10, since $S(\mathbf{x}(I_0; \mathbf{b})) \subseteq S(\mathbf{x}(I_0; \mathbf{b}'))$ as long as $\|\mathbf{x}(I_0; \mathbf{b}) - \mathbf{x}(I_0; \mathbf{b}')\| < \tau(\mathbf{A}, \mathbf{b})$, where

$$\tau(\mathbf{A}, \mathbf{b}) := \max_{\mathrm{I}: \mathrm{I} \in \mathcal{I}(\mathbf{b})} \min_{i \in S(\mathbf{x}(\mathrm{I}; \mathbf{b}))} \mathbf{x}(\mathrm{I}; \mathbf{b})_i > 0.$$

By Part 3 of Lemma 10, we can choose $\delta'(\mathbf{A}, \mathbf{b}) > 0$ small enough that $\|\mathbf{x}(I_0; \mathbf{b}) - \mathbf{x}(I_0; \mathbf{b}')\| < \tau(\mathbf{A}, \mathbf{b})$ whenever $\|\mathbf{b} - \mathbf{b}'\| \leq \delta'$.

We obtain that if $\|\mathbf{b}' - \mathbf{b}'\| \leq \delta^*(\mathbf{A}, \mathbf{b}) =: \delta \wedge \delta'$, then $\mathcal{I}^*(\mathbf{b}') \subseteq \mathcal{I}^*(\mathbf{b})$, as desired.

■

Appendix B. Proofs of main theorems

This section contains the proofs of our main results. We first show how to derive Theorem 6 and Corollary 7 (Appendix B.1). We then obtain Theorem 4 and Corollary 5 as easy consequences (Appendix B.2). Finally, we give the elementary proofs of Theorem 8 and Corollary 9 in Appendix B.3.

B.1. Proofs for Section 5

Our proof is based on the Hadamard differentiability properties of the mapping $H : \mathbb{R}^k \to \mathcal{C}(\mathbb{S}^{m-1})$ which sends a vector $\mathbf{b}$ to the support function $\mathsf{h}_{\mathbf{x}^*(\mathbf{b})}$. Specifically, we will show the following:

Theorem 11 *The mapping $H : \mathbb{R}^k \to \mathcal{C}(\mathbb{S}^{m-1})$ is directionally Hadamard differentiable, with derivative $\mathsf{g}_\cdot$, where g is as in the statement of Theorem 6. That is,*

$$\lim_{t_n \searrow 0, \xi_n \to \xi} \frac{H(\mathbf{b} + t_n \xi_n) - H(\mathbf{b})}{t_n} = \mathsf{g}_\xi \,, \tag{22}$$

in $\mathcal{C}(\mathbb{S}^{m-1})$.

Theorem 6 then follows directly from Römisch (2006).

Proof [Proof of Theorem 11] First, Proposition 2 and Eq. (10) imply that if t_n is sufficiently small and ξ_n is sufficiently close to ξ, then

$$\|\mathsf{h}_{\mathbf{x}^*(\mathbf{b}+t_n\xi_n)} - \mathsf{h}_{\mathbf{x}^*(\mathbf{b}+t_n\xi)}\|_{L^\infty} = \rho_H(\mathbf{x}^*(\mathbf{b}+t_n\xi_n), \mathbf{x}^*(\mathbf{b}+t_n\xi)) \lesssim t_n \|\xi_n - \xi\| \,.$$

Therefore

$$\lim_{t_n \to 0, \xi_n \to \xi} \frac{\|H(\mathbf{b} + t_n\xi_n) - H(\mathbf{b} + t_n\xi)\|_{L^\infty}}{t_n} \lesssim \lim_{\xi_n \to \xi} \|\xi_n - \xi\| = 0 \,, \tag{23}$$

so that

$$\lim_{t_n \searrow 0, \xi_n \to \xi} \frac{H(\mathbf{b} + t_n\xi_n) - H(\mathbf{b} + t_n\xi)}{t_n} = 0 \tag{24}$$

in $\mathcal{C}(\mathbb{S}^{m-1})$.

It therefore suffices to show that

$$\lim_{t_n \searrow 0} \frac{H(\mathbf{b} + t_n\xi) - H(\mathbf{b})}{t_n} = \mathsf{g}_\xi \,. \tag{25}$$

The function $\mathsf{h}(\alpha)$ is differentiable whenever $\sup_{\mathbf{x} \in \mathbf{V}^*(\mathbf{b})} \langle \alpha, \mathbf{x} \rangle$ is uniquely achieved, and the gradient is precisely the vertex giving the supremum. For a vertex $\mathbf{v} \in \mathbf{V}^*(\mathbf{b})$ we write $K_\mathbf{v}$ for the subset of $\mathbb{S}^{m-1}$ consisting of all α for which h is differentiable at α, with derivative $\mathbf{v}$. The collection $\{K_\mathbf{v} : \mathbf{v} \in \mathbf{V}^*(\mathbf{b})\}$ forms a finite disjoint partition of a sphere up to a measure zero set. We shall show that $H(\mathbf{b} + t_n\xi)$ converges uniformly to $H(\mathbf{b})$ on each element of this partition, which establishes almost everywhere uniform convergence and the desired limit.

In what follows, we therefore fix a $\mathbf{v} \in \mathbf{V}^*(\mathbf{b})$ and consider the functions $H(\mathbf{b} + t_n\xi)$ and $H(\mathbf{b})$ on $K_\mathbf{v}$. By assumption, $\sup_{\mathbf{x} \in \mathbf{V}^*(\mathbf{b})} \langle \alpha, \mathbf{x} \rangle$ is uniquely attained at $\mathbf{v}$ for all α in this set. We will now show that for all $\alpha \in K_\mathbf{v}$ and t_n smaller than a constant that depends on $\mathbf{v}$ but not on α, we may restrict the supremum in $\sup_{\mathbf{x} \in \mathbf{x}^*(\mathbf{b}+t_n\xi)} \langle \alpha, \mathbf{x} \rangle$ to vectors of the form $\mathbf{x}(\mathrm{I}; \mathbf{b} + t_n\xi)$ where $\mathrm{I} \in \mathcal{I}^*(\mathbf{b} + t_n\xi)$ and $\mathbf{x}(\mathrm{I}; \mathbf{b}) = \mathbf{v}$.

The optimal set $\mathbf{x} \in \mathbf{x}^*(\mathbf{b} + t_n\xi)$ is the set of nonnegative vectors in $\mathbb{R}^m$ that satisfy the linear constraint $\mathbf{A}\mathbf{x} = \mathbf{b} + t_n\xi$ and that achieve the value $\langle \mathbf{c}, \mathbf{x}\rangle = f(\mathbf{b} + t_n\xi)$. Therefore $\sup_{\mathbf{x}\in\mathbf{x}^*(\mathbf{b}+t_n\xi)}\langle \alpha, \mathbf{x}\rangle$ is equivalent to the linear program

$$\max\langle \alpha, \mathbf{x}\rangle : \mathbf{A}\mathbf{x} = \mathbf{b} + t_n\xi, \langle \mathbf{c}, \mathbf{x}\rangle = f(\mathbf{b} + t_n\xi), \mathbf{x} \geq 0\,. \tag{26}$$

Analogously, we have by assumption that $\mathbf{v}$ is the unique solution to

$$\max\langle \alpha, \mathbf{x}\rangle : \mathbf{A}\mathbf{x} = \mathbf{b}, \langle \mathbf{c}, \mathbf{x}\rangle = f(\mathbf{b}), \mathbf{x} \geq 0\,. \tag{27}$$

Since $\mathbf{x}^*(\mathbf{b} + t_n\xi)$ is compact, Eq. (26) has an optimal solution, and therefore so does its dual problem:

$$\min\langle \lambda, \mathbf{b} + t_n\xi\rangle + \mu f(\mathbf{b} + t_n\xi) : \mathbf{A}^\top\lambda + \mu\mathbf{c} \geq \alpha\,, \tag{28}$$

where $\lambda \in \mathbb{R}^k$ and $\mu \in \mathbb{R}$. Denote by λ^* and μ^* arbitrary optimal solutions to this problem. Complementary slackness implies that any optimal solution $\mathbf{x}_n^*$ to Eq. (26) satisfies

$$i \in S(\mathbf{x}_n^*) \implies (\mathbf{A}^\top\lambda^*)_i + \mu^*\mathbf{c}_i = \alpha_i\,. \tag{29}$$

We can always assume that $\sup_{\mathbf{x}\in\mathbf{x}^*(\mathbf{b}+t_n\xi)}\langle \alpha, \mathbf{x}\rangle$ is achieved at an extreme point, and so is given by some basic feasible solution $\mathbf{x}(\mathrm{I}; \mathbf{b} + t_n\xi)$ for $\mathrm{I} \in \mathcal{I}^*(\mathbf{b} + t_n\xi)$. So it suffices to show that if such an I gives rise to an optimal solution to Eq. (26), then $\mathbf{x}(\mathrm{I}; \mathbf{b}) = \mathbf{v}$. By Eq. (29),

$$(\mathbf{A}^\top\lambda^*)_i + \mu^*\mathbf{c}_i = \alpha_i \quad \forall i \in S(\mathbf{x}(\mathrm{I}; \mathbf{b} + t_n\xi))\,. \tag{30}$$

By Proposition 3, for t_n small enough (independent of α), the fact that $\mathrm{I} \in \mathcal{I}^*(\mathbf{b} + t_n\xi)$ implies $\mathrm{I} \in \mathcal{I}^*(\mathbf{b})$ and $S(\mathbf{x}(\mathrm{I}; \mathbf{b})) \subseteq S(\mathbf{x}(\mathrm{I}; \mathbf{b} + t_n\xi))$. Combining this fact with Eq. (30) gives that

$$\langle \mathbf{A}^\top\lambda^* + \mu^*\mathbf{c} - \alpha, \mathbf{x}(\mathrm{I}; \mathbf{b})\rangle = 0\,. \tag{31}$$

But this implies that $\mathbf{x}(\mathrm{I}; \mathbf{b})$ must be optimal for Eq. (27) by weak duality. To see this explicitly, we first observe that $\mathrm{I} \in \mathcal{I}^*(\mathbf{b})$ shows that $\mathbf{x}(\mathrm{I}; \mathbf{b})$ is feasible in Eq. (27). Second, λ^* and μ^* are feasible for Eq. (28). Therefore, if $\mathbf{x}$ is any feasible point for Eq. (28), we have

$$\begin{aligned}\langle \alpha, \mathbf{x}(\mathrm{I}; \mathbf{b}) - \mathbf{x}\rangle &= \langle \mathbf{A}^\top\lambda + \mu\mathbf{c}, \mathbf{x}(\mathrm{I}; \mathbf{b}) - \mathbf{x}\rangle + \langle \mathbf{A}^\top\lambda + \mu\mathbf{c} - \alpha, \mathbf{x} - \mathbf{x}(\mathrm{I}; \mathbf{b})\rangle \\ &\geq 0\,,\end{aligned}$$

where we have used that $\langle \mathbf{A}^\top\lambda + \mu\mathbf{c}, \mathbf{x}(\mathrm{I}; \mathbf{b}) - \mathbf{x}\rangle = 0$ since $\mathbf{A}\mathbf{x}(\mathrm{I}; \mathbf{b}) = \mathbf{A}\mathbf{x}$ and $\langle \mathbf{c}, \mathbf{x}(\mathrm{I}; \mathbf{b})\rangle = \langle \mathbf{c}, \mathbf{x}\rangle$ and the second term is nonnegative in light of Eq. (31) and the fact that $\mathbf{A}^\top\lambda + \mu\mathbf{c} - \alpha$ and $\mathbf{x}$ are both nonnegative. Therefore $\mathbf{x}(\mathrm{I}; \mathbf{b})$ is optimal for Eq. (27), so we must have $\mathbf{x}(\mathrm{I}; \mathbf{b}) = \mathbf{v}$, which was what we wanted to show.

For $\alpha \in K_{\mathbf{v}}$, we therefore have that for t_n sufficiently small, depending only on $\mathbf{v}$,

$$\sup_{\mathbf{x}\in\mathbf{V}^*(\mathbf{b}+t_n\xi)}\langle \alpha, \mathbf{x}\rangle - \sup_{\mathbf{x}\in\mathbf{V}^*(\mathbf{b})}\langle \alpha, \mathbf{x}\rangle = \max_{\mathrm{I}\in\mathcal{I}^*(\mathbf{b}+t_n\xi):\mathbf{x}(\mathrm{I};\mathbf{b})=\mathbf{v}}\langle \alpha, \mathbf{x}(\mathrm{I}; \mathbf{b} + t_n\xi) - \mathbf{v}\rangle\,. \tag{32}$$

Consider now the linear program appearing in the definition of $\mathbf{q}_\alpha^*(\xi)$:

$$\min\langle \mathbf{c}, \mathbf{q}\rangle : \mathbf{A}\mathbf{q} = \xi, \mathbf{q}_i \geq 0 \quad \forall i \notin S(\mathbf{v})\,. \tag{33}$$

Note that this program does not depend on α, only on $\mathbf{v}$. We wish to show that for t_n small enough, the basic feasible solutions of this program are exactly the vectors of the form $t_n^{-1}(\mathbf{x}(\mathrm{I};\mathbf{b}+t_n\xi)-\mathbf{v})$ for $\mathrm{I} \in \mathcal{I}(\mathbf{b}+t_n\xi)$ such that $\mathbf{x}(\mathrm{I};\mathbf{b}) = \mathbf{v}$. A basic feasible solution corresponds to a selection of m linearly independent constraints: k that arise from the equality constraints, and $m-k$ tight inequality constraints selected from the set $S(\mathbf{v})^C$.

Fix a basis for this linear program, denote by $\bar{J}$ the set of equality constraints selected from the set $S(\mathbf{v})^C$, and let $J = \bar{J}^C$. The fact that $\bar{J} \subseteq S(\mathbf{v})^C$ implies $S(\mathbf{v}) \subseteq J$. Since these give rise to a basis, the system of equations given by $\mathbf{A}\mathbf{q} = \xi$ and $\mathbf{q}_j = 0$ for $j \notin J$ has a unique solution. Equivalently, the set J satisfies that $\mathbf{A}_J\mathbf{q}_J = \xi$ has a unique solution, so that $\mathbf{A}_J$ is full rank. If this basis gives rise to a basic feasible solution of Eq. (33), then $(\mathbf{A}_J^{-1}\xi)_i \geq 0$ for all $i \notin S(\mathbf{v})$. To conclude, basic feasible solutions to Eq. (33) are of the form $\mathbf{q}_J = \mathbf{A}_J^{-1}\xi$ and $\mathbf{q}_{J^C} = \mathbf{0}$, where $J \supseteq S(\mathbf{v})$ satisfies

$$|J| = k, \mathrm{rank}(\mathbf{A}_J) = k, (\mathbf{A}_J^{-1}\xi)_i \geq 0 \quad \forall i \notin S(\mathbf{v})\,. \tag{34}$$

Conversely, every set J satisfying these requirements gives rise to a basic feasible solution of Eq. (33).

On the other hand, if $\mathbf{y} = t_n^{-1}(\mathbf{x}(\mathrm{I};\mathbf{b}+t_n\xi)-\mathbf{v})$ for some $\mathrm{I} \in \mathcal{I}(\mathbf{b}+t_n\xi)$ such that $\mathbf{x}(\mathrm{I};\mathbf{b}) = \mathbf{v}$, then $\mathbf{y}_\mathrm{I} = t_n^{-1}\mathbf{A}_\mathrm{I}^{-1}(\mathbf{b}+t_n\xi-\mathbf{b}) = \mathbf{A}_\mathrm{I}^{-1}\xi$ and $\mathbf{y}_{\mathrm{I}^C} = \mathbf{0}$. Moreover, the requirement that $\mathrm{I} \in \mathcal{I}(\mathbf{b}+t_n\xi)$ implies that $\mathbf{y}_\mathrm{I} = \mathbf{A}_\mathrm{I}^{-1}\xi \geq -t_n^{-1}\mathbf{v}_\mathrm{I}$, since this is equivalent to I being feasible for $\mathbf{b}+t_n\xi$, and the requirement that $\mathbf{x}(\mathrm{I};\mathbf{b}) = \mathbf{v}$ implies $S(\mathbf{v}) \subseteq I$. To conclude, $\mathbf{y}$ is a vector such that $\mathbf{y}_\mathrm{I} = \mathbf{A}_\mathrm{I}^{-1}\xi$ and $\mathbf{y}_{\mathrm{I}^C} = \mathbf{0}$, where $\mathrm{I} \supseteq S(\mathbf{v})$ satisfies

$$|\mathrm{I}| = k, \mathrm{rank}(\mathbf{A}_\mathrm{I}) = k, \mathbf{A}_\mathrm{I}^{-1}\xi \geq -t_n^{-1}\mathbf{v}_\mathrm{I}\,. \tag{35}$$

Conversely, any set I satisfying these properties gives rise to a vector $\mathbf{y}$ of the form $t_n^{-1}(\mathbf{x}(\mathrm{I};\mathbf{b}+t_n\xi)-\mathbf{v})$ for some $\mathrm{I} \in \mathcal{I}(\mathbf{b}+t_n\xi)$ such that $\mathbf{x}(\mathrm{I};\mathbf{b}) = \mathbf{v}$.

We now notice that Eq. (34) and Eq. (35) are nearly the same. Clearly, all sets $I \supseteq S(\mathbf{v})$ satisfying Eq. (35) also satisfy $(\mathbf{A}_I^{-1}\xi)_i \geq 0 \quad \forall i \notin S(\mathbf{v})$, since if $i \notin S(\mathbf{v})$ this is equivalent to the requirement that $(\mathbf{A}_\mathrm{I}^{-1}\xi)_i \geq -t_n^{-1}\mathbf{v}_i = 0$ in Eq. (35). Conversely, for t_n sufficiently small, every set $J \supseteq S(\mathbf{v})$ satisfying Eq. (34) also satisfies $\mathbf{A}_J^{-1}\xi \geq -t_n^{-1}\mathbf{v}_J$ This is because every coordinate of the vector $\mathbf{A}_J^{-1}\xi$ is bounded, uniformly in J. So for t_n small enough, if $\mathbf{v}_i > 0$, we will have $(\mathbf{A}_J^{-1}\xi)_i \geq -t_n^{-1}\mathbf{v}_i$. Therefore, for t_n small enough, the allowable subsets in Eq. (34) and Eq. (35) agree. In other words, the basic feasible solutions to Eq. (33) are precisely the vectors of the form $t_n^{-1}(\mathbf{x}(\mathrm{I};\mathbf{b}+t_n\xi)-\mathbf{v})$ for some $\mathrm{I} \in \mathcal{I}(\mathbf{b}+t_n\xi)$ such that $\mathbf{x}(\mathrm{I};\mathbf{b}) = \mathbf{v}$. Moreover, it is now easy to see that *optimal* vertices in Eq. (33) correspond to vectors of the form $t_n^{-1}(\mathbf{x}(\mathrm{I};\mathbf{b}+t_n\xi)-\mathbf{v})$ for some $\mathrm{I} \in \mathcal{I}^*(\mathbf{b}+t_n\xi)$ (i.e., the set of *optimal* bases) for which $\mathbf{x}(\mathrm{I};\mathbf{b}) = \mathbf{v}$. Indeed, in each case we simply need to select the subset of vertices that minimize the inner product with $\mathbf{c}$: that is obviously true in the case of solutions to Eq. (33), and by linearity a basic feasible solution $\mathbf{x}(\mathrm{I};\mathbf{b}+t_n\xi)$ minimizes the inner product with $\mathbf{c}$ if and only if $t_n^{-1}(\mathbf{x}(\mathrm{I};\mathbf{b}+t_n\xi)-\mathbf{v})$ minimizes the inner product with $\mathbf{c}$.

We conclude that for t_n small enough (depending only on $\mathbf{v}$ and not on α), for all $\alpha \in K_\mathbf{v}$,

$$\max_{\mathrm{I}\in\mathcal{I}^*(\mathbf{b}+t_n\xi):\mathbf{x}(\mathrm{I};\mathbf{b})=\mathbf{v}} \langle \alpha, \mathbf{x}(\mathrm{I};\mathbf{b}+t_n\xi)-\mathbf{v}\rangle = \max_{\mathbf{x}\in\mathbf{q}_\alpha^*(\xi)} \langle \alpha, t_n\mathbf{x}\rangle = t_n\mathrm{g}_\xi(\alpha)\,. \tag{36}$$

Therefore

$$\lim_{t_n\searrow 0} \frac{H(\mathbf{b}+t_n\xi)(\alpha) - H(\mathbf{b})(\alpha)}{t_n} = \mathrm{g}_\xi(\alpha) \tag{37}$$

uniformly on $K_{\mathbf{v}}$, as claimed. ■

Proof [Proof of Corollary 7] The functional $f \mapsto \sup_{\alpha \in \mathbb{S}^{m-1}} f(\alpha)$ is clearly a continuous map from $\mathcal{C}(\mathbb{S}^{m-1})$ to $\mathbb{R}$, so the continuous mapping theorem combined with Theorem 6 implies

$$r_n \sup_{\alpha \in \mathbb{S}^{m-1}} |\mathsf{h}_n(\alpha) - \mathsf{h}(\alpha)| \xrightarrow{D} \sup_{\alpha \in \mathbb{S}^{m-1}} |\mathsf{g}_{\mathbb{G}}(\alpha)| .$$

Combined with Eq. (10), this implies

$$r_n \rho_H(\mathbf{x}^*(\mathbf{b}_n), \mathbf{x}^*(\mathbf{b})) \xrightarrow{D} \sup_{\alpha \in \mathbb{S}^{m-1}} |\mathsf{g}_{\mathbb{G}}(\alpha)| .$$

■

B.2. Proofs for Section 4

The results of this section will follow from specializing the results of Section 5 to the case where the target solution $\mathbf{x}^*(\mathbf{b})$ is unique.

Proof [Proof of Theorem 4] We will apply Theorem 6. We first need to verify that the sense of convergence is the same, and then that the expressions for the limit agree. The random solution set $\mathbf{x}^*(\mathbf{b}_n)$ is nonempty with probability approaching 1 as $n \to \infty$ by Lemma 10. The sets on the left side of the limit in Theorem 4 are therefore (on an event of probability approaching one) non-empty, convex, compact sets. If $K_1, \dots,$ is a sequence of such random sets, Molchanov (2005, Theorem 6.13) implies that it converges weakly to a random set K if and only if for any $N \in \mathbb{N}$, $\alpha_1, \dots, \alpha_N \in \mathbb{S}^{m-1}$, the vector $(\mathsf{h}_{K_n}(\alpha_1), \dots, \mathsf{h}_{K_n}(\alpha_N))$ converges to $(\mathsf{h}_K(\alpha_1), \dots, \mathsf{h}_K(\alpha_N))$, and the sets are tight, in the sense that $\lim_{c \to \infty} \sup_n \mathbb{P}\{\|K_n\| \geq c\} \to 0$.

To compute the support function of $r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b}))$, we use the fact that $\mathbf{x}^*(\mathbf{b})$ is a singleton to write

$$\begin{aligned}
\sup_{\mathbf{x} \in r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b}))} \langle \alpha, \mathbf{x} \rangle &= \sup_{\mathbf{x}' \in \mathbf{x}^*(\mathbf{b}_n)} \langle \alpha, r_n(\mathbf{x}' - \mathbf{x}^*(\mathbf{b})) \\
&= r_n \sup_{\mathbf{x}' \in \mathbf{x}^*(\mathbf{b}_n)} \langle \alpha, \mathbf{x}' \rangle - r_n \langle \alpha, \mathbf{x}^*(\mathbf{b}) \rangle \\
&= r_n(\mathsf{h}_n(\alpha) - \mathsf{h}(\alpha)) ,
\end{aligned}$$

where h_n and h are as in Theorem 6. Theorem 6 shows that the support function of $r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b}))$ converges to $\mathsf{g}_{\mathbb{G}}$. The tightness condition is therefore trivially satisfied, so we will be done as long as we can show that the function $\mathsf{g}_{\mathbb{G}}$ is the support function of the set $\mathbf{p}^*_{\mathbf{b}}(\mathbb{G})$. Since $\mathsf{h}(\alpha) = \langle \alpha, \mathbf{x}^*(\mathbf{b}) \rangle$, the gradient $\nabla h(\alpha)$ is identically equal to $\mathbf{x}^*(\mathbf{b})$, so that the linear program Eq. (12) reduces to Eq. (8). Since $\mathsf{g}_{\mathbb{G}}$ is defined as the supremum of a linear functional, we may replace the set $\mathbf{q}^*(\mathbb{G})$ of optimal vertices by its convex hull $\operatorname{conv}(\mathbf{q}^*(\mathbb{G}))$, and we just need to show that this set agrees with $\mathbf{p}^*_{\mathbf{b}}(\mathbb{G})$ to show that $\mathsf{g}_{\mathbb{G}}$ is its support function. To do so, we use the fact that the recession cone of $\mathbf{p}^*_{\mathbf{b}}(\mathbb{G})$ is $\{\mathbf{0}\}$ when $\mathbf{x}^*(\mathbf{b})$ is unique. To see this, first observe that a vector in the recession cone must satisfy $\langle \mathbf{c}, \mathbf{d} \rangle = 0, \mathbf{A}\mathbf{d} = 0, \mathbf{d}_i \geq 0$ for all $i \notin S(\mathbf{x}^*(\mathbf{b}))$. If $\mathbf{d}$ is such a vector, then for $\epsilon > 0$ small enough the vector $\mathbf{x}^*(\mathbf{b}) + \epsilon \mathbf{d}$ is also optimal for Eq. (1). Indeed this vector satisfies the linear constraints and has the same objective value, and for ϵ sufficiently small no coordinates of $\mathbf{x}^*(\mathbf{b}) + \epsilon\mathbf{d}$ will be negative. Since we have assumed that $\mathbf{x}^*(\mathbf{b})$ is a singleton, we must have that $\mathbf{d} = \mathbf{0}$, so that the recession cone of the optimal set in this LP is $\{\mathbf{0}\}$. Therefore $\mathbf{p}^*_{\mathbf{b}}(\mathbb{G}) = \operatorname{conv}(\mathbf{q}^*(\mathbb{G}))$, and therefore $\mathrm{g}_{\mathbb{G}}$ is the support function of $\mathbf{p}^*_{\mathbf{b}}(\mathbb{G})$, proving the claim. ■

Proof [Proof of Corollary 5] For any vector $\mathbf{x}$, the functional $S \mapsto d(S, \mathbf{x})$ is continuous with respect to the Hausdorff distance. The continuous mapping theorem therefore implies that

$$d(r_n(\mathbf{x}^*(\mathbf{b}_n) - \mathbf{x}^*(\mathbf{b})), \mathbf{0}) \xrightarrow{D} d(\mathbf{p}^*_{\mathbf{b}}(\mathbb{G}), \mathbf{0})\,.$$

It then suffices to note that the quantity on the left is equal to $r_n d(\mathbf{x}^*(\mathbf{b}_n), \mathbf{x}^*(\mathbf{b}))$. ■

B.3. Proofs for Section 6

Proof [Proof of Theorem 8] Define $\mathbf{A}_n := (\mathbf{A}; \mathbf{e}_{\mathrm{I}_n^c}) \in \mathbb{R}^{m\times m}$ to be the matrix whose first k rows are $\mathbf{A}$ and whose remaining $m-k$ rows consist of the elementary basis vectors $\mathbf{e}_i$ for $i \notin \mathrm{I}_n$. Since I_n is a basis, $\mathbf{A}_n$ has full rank. Moreover, the fact that the basis I_n corresponds to $\hat{\mathbf{x}}_n$ implies that $S(\hat{\mathbf{x}}_n) \subseteq \mathrm{I}_n$. Therefore $\mathbf{A}_n\hat{\mathbf{x}}_n = \mathbf{b}_n^0$, where $\mathbf{b}_n^0 := (\mathbf{b}_n, \mathbf{0}) \in \mathbb{R}^m$ is the augmented vector whose first k coordinates are $\mathbf{b}_n$ and whose remaining $m-k$ coordinates are zero. Similarly, $\mathbf{A}_n\mathbf{x}(\mathrm{I}_n; \mathbf{b}) = \mathbf{b}^0$, where $\mathbf{b}^0 \in \mathbb{R}^m$ is defined in an analogous way.

We obtain

$$r_n\mathbf{A}_n(\hat{\mathbf{x}}_n - \mathbf{x}(\mathrm{I}_n; \mathbf{b})) = r_n(\mathbf{b}_n^0 - \mathbf{b}^0) \xrightarrow{D} \mathbb{G}^0, \tag{38}$$

where as above $\mathbb{G}^0$ is the random variable obtained by appending $m-k$ zeroes to $\mathbb{G}$.

We will now show that $r_n(\bar{\mathbf{x}}_n^* - \mathbf{x}(\mathrm{I}_n; \mathbf{b})) \xrightarrow{p} \mathbf{0}$, so that we can replace $\mathbf{x}(\mathrm{I}_n; \mathbf{b})$ by $\bar{\mathbf{x}}_n^*$ in the limit. By Theorem 3, there exists a constant $\delta > 0$ such that if $\|\mathbf{b}_n - \mathbf{b}\| \le \delta$, then $\mathcal{I}^*(\mathbf{b}_n) \subseteq \mathcal{I}^*(\mathbf{b})$. Since $\mathrm{I}_n \in \mathcal{I}^*(\mathbf{b}_n)$ by assumption, this fact implies that if $\|\mathbf{b}_n - \mathbf{b}\| \le \delta$, then I_n is an optimal basis for the target problem, i.e., $\mathbf{x}(\mathrm{I}_n; \mathbf{b}) \in \mathbf{x}^*(\mathbf{b})$. In particular, on the event that $\|\mathbf{b}_n - \mathbf{b}\| \le \delta$, we have $\bar{\mathbf{x}}_n^* = \mathbf{x}(\mathrm{I}_n; \mathbf{b})$. The distributional convergence assumption Eq. (2) implies $\mathbf{b}_n \xrightarrow{p} \mathbf{b}$. We therefore have that $\mathbb{P}\{r_n\|\bar{\mathbf{x}}_n^* - \mathbf{x}(\mathrm{I}_n; \mathbf{b})\| > 0\} \le \mathbb{P}\{\|\mathbf{b}_n - \mathbf{b}\| > \delta\} \to 0$ as $n \to \infty$, so that $r_n(\bar{\mathbf{x}}_n^* - \mathbf{x}(\mathrm{I}_n; \mathbf{b})) \xrightarrow{p} \mathbf{0}$. Combining this fact with Eq. (38) yields

$$r_n\mathbf{A}_n(\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_n^*) \xrightarrow{D} \mathbb{G}^0, \tag{39}$$

If we define G_α as in the theorem, we therefore obtain that

$$\liminf_{n\to\infty} \mathbb{P}\left\{r_n(\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_n^*) \in \mathbf{A}_n^{-1}G_\alpha^0\right\} = \liminf_{n\to\infty} \mathbb{P}\left\{r_n\mathbf{A}_n(\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_n^*) \in G_\alpha^0\right\} \ge 1-\alpha \tag{40}$$

where $G_\alpha^0 \in \mathbb{R}^m$ is obtained from $G_\alpha \in \mathbb{R}^k$ by padding each vector with zeros. To conclude, we note that $\mathbf{A}_n^{-1}G_\alpha^0 = \mathbf{x}(\mathrm{I}_n; G_\alpha)$. Indeed, for any $\mathbf{G} \in G_\alpha$, the definition of $\mathbf{A}_n$ implies that the first k coordinates of $\mathbf{A}_n\mathbf{x}(\mathrm{I}_n; \mathbf{G})$ are $\mathbf{A}\mathbf{x}(\mathrm{I}_n; \mathbf{G}) = \mathbf{G}$ and the last $m-k$ coordinates are zero. Therefore $G_\alpha^0 = \mathbf{A}_n\mathbf{x}(\mathrm{I}_n; G_\alpha)$, and since $\mathbf{A}_n$ is invertible this proves the claim. ■

Corollary 9 is an immediate consequence.

Proof [Proof of Corollary 9] If $r_n(\hat{\mathbf{x}}_n - \bar{\mathbf{x}}_n^*) \in \mathbf{x}(\mathrm{I}_n; G_\alpha)$, then there exists an $\mathbf{x} \in \mathbf{x}(\mathrm{I}_n; G_\alpha)$ such that $\bar{\mathbf{x}}_n^* = \hat{\mathbf{x}}_n - r_n^{-1}\mathbf{x}$. Since $\bar{\mathbf{x}}_n^* \in \mathbf{x}^*(\mathbf{b})$, the result follows from Theorem 8. ■