A SIMPLEX METHOD FOR COUNTABLY INFINITE LINEAR PROGRAMS*

ARCHIS GHATE†, CHRISTOPHER THOMAS RYAN‡, AND ROBERT L. SMITH§

Abstract. We introduce a simplex method for general countably infinite linear programs (CILPs). Previous literature has focused on special cases, such as infinite network flow problems or Markov decision processes. A novel aspect of our approach is the placing of data and decision variables in a Hilbert space that elegantly encodes a "discounted" weighting to ensure the continuity of infinite sums. Under some assumptions, including that all basic feasible solutions are nondegenerate with strictly positive support, and the set of bases is closed in an appropriate topology, we show convergence to the optimal value for our proposed simplex algorithm. We show that existing applications naturally fit this more general framework.

Key words. countably infinite linear programs, infinite-dimensional optimization, simplex method

AMS subject classifications. 90C49, 65K05

1. Introduction.

Infinite-dimensional linear programming plays an important role in the theory of stochastic, robust, and dynamic optimization [4, 19, 23, 26], bearing fruit in applications to inventory management [2], revenue management [1], production planning [18], workforce planning [22], and equipment replacement [5], among others.

The special case of countably infinite linear programs (CILPs) has received increasing attention [14, 16, 32, 36]. In a CILP, the decision-maker has countably many decisions and faces countably many linear constraints. Although a comprehensive theory of duality for CILPs has been proposed in [14], a general theory of simplex methods for CILPs is still missing. To date, efforts have primarily focused on devising algorithms for special cases, including nonstationary and countable-state Markov decision processes [19, 26], and networks with countably infinite nodes and arcs [32, 36]. A goal of this paper is to extract analytical insight from these cases in the literature, discover what they have in common, and connect this to a deeper understanding of the topological structure of (at least partially) "tractable" countably infinite linear programs.

In addition to tackling as yet intractable problems from the above applications, a general simplex theory could provide insights into and a foundation for future solution approaches to a larger class of problems where CILPs and their extensions arise. These include computing the stationary distributions, occupation measures, and exit distributions of Markov chains [24]; nonstationary stochastic optimization including multi-armed bandit problems with time-varying rewards [8]; countably infinite monotropic programs [9, 15] and convex cost flow problems on countably infinite networks [30]; optimization problems with infinite sums [27]; fluid approximations of decomposable Markov decision processes [6]; search problems in robotics [13]; infinite

*Submitted on December 2, 2019, revision submitted on XXX

Funding: Funded in part by National Science Foundation grants CMMI 1561918 and 1333260 and the Natural Sciences and Engineering Research Council of Canada Discovery Grant RGPIN-2020-06488.

†Industrial and Systems Engineering, University of Washington (archis@uw.edu).

‡UBC Sauder School of Business, University of British Columbia (chris.ryan@sauder.ubc.ca)

§Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI (rlsmith@umich.edu)

horizon stochastic programs [20]; and games with partial information [11]. Unfortunately, the lack of such a theory has prevented the broader optimization community from fully utilizing CILPs in their work. This paper attempts to partially overcome this hurdle.

One reason for the focus thus far on special cases is that infinite-dimensional linear programming involves complex topological considerations in general. Indeed, selecting the topological space to embed the data is an important modeling choice [4]. Depending on the topology, it can be more or less easy to state the dual, more or less easy to prove weak and strong duality, and more or less easy to build the components of a simplex method. By examining a special case, the choice of dual and the elements of a simplex algorithm often become easier to identify. To deal with greater generality, this paper proposes a novel topology for CILPs (inspired by earlier work in [35]) that frames the problem in a Hilbert space setting.

Before discussing further implications of this modeling choice, we clarify what we mean by a "simplex method". The geometric essence of the simplex method is the traversing of edges (called "pivoting") between extreme points of a polyhedron in search of an optimal solution. In the finite case, since the objective function is linear (and hence both convex and concave) and the linear constraints describe a convex feasible region, the existence of an extreme point optimal solution is guaranteed and determined by "local" considerations – if there are no improving directions along edges from a given extreme point then it is a global optimum.

The computational realization of this geometric view of the simplex method involves the algebraic notions of basic feasible solutions, basic directions, and reduced costs. These are in direct correspondence to the geometric notions of extreme points, edges, and improving directions, respectively. The success of the simplex method crucially depends on this tight connection between algebra and geometry.

A core difficulty in designing a simplex method for CILPs, even at the abstract level, is that both the geometric view and the relationship between algebra and geometry are more tenuous. Indeed, one can easily write down an innocent-looking infinite-dimensional linear program that is bounded and feasible but has no optimal solution. Consider, for example, a minimum cost flow problem with two nodes with supply and demand one, joined by a countably infinite number of arcs with costs $(1/2)^k, k = 1, 2, \ldots$. The infimum over all feasible costs is zero but is not attained. Even when optimal solutions are known to exist, the feasible region may have no extreme points (p. 61 of [4]). Without extreme points, the geometric essence of the simplex method has no grounding. Even when extreme points do exist, there are cases where there do not exist edges on which to "pivot" between them. Consider, for example, the feasible region of the closed unit disk centered around the origin in $\mathbb{R}^2$ and represented by the intersection of its countably many supporting half-spaces along the rational points of its boundary. The boundary of the disk constitutes its extreme points while it has no edges to pivot along. Indeed, the cone of improving directions from a given extreme point may lack extreme rays (p. 28 of [4]).

Other desirable properties we take for granted in the finite simplex method — beyond mere clarity about the objects and steps involved — may also fail in the infinite-dimensional setting. Ideally, a simplex method would satisfy the following:

- (P1) The iterates have monotone non-increasing objective values.
- (P2) The objective values of the iterates converge to the optimal value of the problem (optimal value convergence).
- (P3) Each iteration of the algorithm can be performed in finite time and with a finite amount of data.

(P4) The iterates converge to an optimal solution of the problem.

Property (P1) is helpful since algorithms are terminated after finitely many iterations in practice. Property (P1) ensures that the last iterate of the algorithm is always the best among the sequence of iterates (keeping track of the *incumbent* iterate, which is a common practice in non-monotonic algorithms, is difficult in infinite-dimensional problems, where calculating objective values already requires infinite time and space).

It is well documented (see, for instance, [16]) that properties (P1)–(P4) need not hold in general. Designing algorithms that meet some or all of these properties for special cases have been the focus of a stream of papers in recent years [19, 26, 32, 36].

In this paper, we provide a set of sufficient conditions (captured as assumptions (A1)–(A8) below) that ensure our proposed simplex-method satisfies (P1) and (P2) for a broad class of problems. This is the main result of the paper, captured as Theorem 8.3. The result is nontrivial, and the set of sufficient conditions critically depend on the problem's embedding in the Hilbert space discussed above. The closest result in the literature is the "shadow simplex method" in [16]. There, an algorithm is provided that satisfies (P2) and (P3) under a set of conditions that does not guarantee (P1). It is a simplex method in the sense that it pivots among extreme points of finite-dimensional projections (or "shadows") of the feasible region (that may not correspond to adjacent pivots on the original feasible region). A general approach to resolving (P3) is beyond the scope of this paper, however, the examples we discuss in Section 9 does have a finite implementation. As for (P4), our main result on optimal value convergence (Theorem 8.3) establishes the existence of a *subsequence* of iterates that converges to an optimal solution. To establish convergence of the entire sequence of iterates involves careful selection arguments in the spirit of [34], which is not the focus of the current paper. However, we do show in Theorem 8.4 that the set of iterates of the simplex method become arbitrarily close to the set of optimal solutions and, by implication, if there is a unique optimal solution, (P4) holds.

The reader may notice that we have not included among our desiderata (P1)–(P4) a statement about the rate of convergence of the simplex algorithm in question. Although in finite-dimensional optimization this type of analysis is commonplace, in the infinite-dimensional setting we know of only a few cases where convergence rates have been posited (for instance, [29, 33]). These papers leverage compactness and continuity properties of continuous linear programs that fail to hold in our setting.

The dearth of convergence rates results in the literature is not a surprise. The finite-dimensional simplex algorithm itself, arguably the most impactful optimization algorithm ever developed, evaded complexity analysis for decades and remains an open area of research until the present day. Klee and Minty showed worst-case performance can be exponential, and recent results show that this worst-case performance holds under numerous pivot rules. Indeed, a celebrated result is a recent subexponential (although not polynomial) worst case for a particularly successful pivot rule [21].

We organize the remainder of the paper as follows. We start in Section 2 with a few preliminaries and provide an overview of the Hilbert space structure leveraged throughout the paper. In Section 3, we state our general CILP problem. In Section 4, we define the concept of a basic feasible solution and show that the extreme points are basic feasible solutions. Section 5 describes the mechanics of pivoting between extreme points. In Section 6, we introduce the concept of reduced costs to provide an optimality condition analogous to the finite-dimensional simplex method. In Section 7, we construct our simplex method based on choosing pivots of "steepest descent"; i.e., reduce the objective value by the greatest possible rate. This guarantees property (P1) but also proves crucial in establishing (P2). In Section 8, we show that this simplex method converges to optimal value. Section 9 provides a concrete example that satisfies our assumptions.

2. Preliminaries. This section contains basic notation and definitions. Most importantly, it defines a type of topology on the space of real sequences that is used throughout the rest of the paper.

Let $\mathbb{R}$ and $\mathbb{N}$ denote the set of real and natural numbers, respectively. The **vector space of all real sequences** is denoted $\mathbb{R}^{\mathbb{N}}$. We denote an **element** x of $\mathbb{R}^{\mathbb{N}}$ by $(x_j)_{j=1}^{\infty}$ (or more simply (x_j)) where x_j is called the jth **component** of x. The vector space **ordering** on $\mathbb{R}^{\mathbb{N}}$ is denoted $\geq$ where $x \geq 0$ if $x_i \geq 0$ for $i = 1, 2, \ldots$. A matrix $A = (a_{ij})_{i,j=1}^{\infty}$ (or more simply $A = (a_{ij})$) where a_{ij} is a real number for all i and j is called a **doubly infinite matrix**. The jth column of A is denoted $a_{\cdot j}$ and the ith row is denoted $a_{i\cdot}$. The columns and rows of A can be viewed as sequences in $\mathbb{R}^{\mathbb{N}}$. We let Ax denote the vector $(\sum_{j=1}^{\infty} a_{ij}x_j : i = 1, 2, \ldots)$. Let u and v be two sequences in $\mathbb{R}^{\mathbb{N}}$. For brevity, we sometimes let $u^\top v$ denote the infinite sum $\sum_{j=1}^{\infty} u_j v_j$.

For any countable set B of vectors in $\mathbb{R}^{\mathbb{N}}$, let $\operatorname{cspan}(B)$ denote their **countable span**; that is, for $B = \{B^1, B^2, \ldots\}$ let $\operatorname{cspan}(B) = \{\sum_{j=1}^{\infty} \alpha_j B^j : \alpha \in \mathbb{R}^{\mathbb{N}}$ where $\sum_{j=1}^{\infty} \alpha_j B^j$ converges $\}$ where $\sum_{j=1}^{\infty} \alpha_j B^j = \lim_{N\to\infty} \sum_{j=1}^{N} \alpha_j B^j$ denotes component-wise convergence of partial sums.[1] We abuse notation and let A denote both a doubly-infinite matrix as well as the set of columns in A. This notation will save a lot of tedious distinctions throughout the paper. Accordingly, we may write $\operatorname{cspan}(A)$ as the countable span of the set of columns of A (recall each column is a vector in $\mathbb{R}^{\mathbb{N}}$).

For any $x \in \mathbb{R}^{\mathbb{N}}$, the **support set** $\mathcal{S}(x)$ of x is the set of indices j where x_j is nonzero; that is, $\mathcal{S}(x) := \{j : x_j \neq 0\}$. Let $\mathcal{S}^c(x)$ denote the **complement** of the support set of x; that is, $\mathcal{S}^c(x) := \{j : x_j = 0\}$. Let F be a subset of $\mathbb{R}^{\mathbb{N}}$. A vector $x \in F$ is an **extreme point** of F if it *cannot* be expressed as $x = \lambda x^1 + (1-\lambda)x^2$ where $\lambda \in (0, 1)$ and $x^1, x^2 \in F$ with $x^1 \neq x^2$. The **set of all extreme points** of F is denoted $\operatorname{ext} F$.

We define a particular class of Hilbert topologies on the space of real sequences. Earlier work using a similar topology can be found in [35]. Define $\mathbb{R}^{\infty} = \prod_{j=1}^{\infty} H_j$ where $H_j = \mathbb{R}$ (as a set, but with a different topology defined below) for all $j = 1, 2, \ldots$. The standard inner product and norm on $\mathbb{R}$ are denoted $\langle\cdot,\cdot\rangle$ and $|\cdot|$, respectively. That is, for $x, y \in \mathbb{R}$, $\langle x, y\rangle = xy$ and $|x|$ is the absolute value of x. We endow each H_j with a slightly modified topology. Fix a $\delta_j \in (0, 1)$ and define the inner product and norm on H_j as $\langle\cdot,\cdot\rangle_j = \delta_j^2\langle\cdot,\cdot\rangle$ and $|\cdot|_j = \delta_j|\cdot|$. That is, if $x, y \in H_j$ then $\langle x, y\rangle_j = \delta_j^2 xy$ and $|x|_j = \delta_j|x_j|$. Under these operations, it is straightforward to show that H_j is a Hilbert space with an appropriately defined norm topology associated with $|\cdot|_j$, which agrees with the usual Euclidean topology on $\mathbb{R}$.

The Hilbert sum $H = \{(x_j) \in \prod_{j=1}^{\infty} H_j : \sum_{j=1}^{\infty} |x_j|_j^2 = \sum_{j=1}^{\infty} \delta_j^2|x_j|^2 < \infty\}$ of the spaces H_j is endowed with inner product $(x|y) = \sum_{j=1}^{\infty} |x_j y_j|_j = \sum_{j=1}^{\infty} \delta_j^2\langle x_j, y_j\rangle$ and norm

$$||x|| = \Big(\sum_{j=1}^{\infty} |x_j|_j^2\Big)^{1/2} = \Big(\sum_{j=1}^{\infty} \delta_j^2|x_j|^2\Big)^{1/2}. \tag{2.1}$$

and is a Hilbert space (see Section I.6 in [10]). Using this notation, another way to define H is the set of sequences in $\prod_{j=1}^{\infty} H_j$ with finite $||\cdot||$ norm. Note that every

[1] When B is a finite set of vectors, the sums defining $\operatorname{cspan}(B)$ are finite.

choice of the sequence (δ_j) may give rise to a different Hilbert space H.

For every index j, define a compact set $V_j \subseteq H_j$ where $|v_j| \leq r_j$ for every $v_j \in V_j$. Let $V = \prod_{j=1}^{\infty} V_j$. By Tychonoff's theorem, V is compact in the product norm topology on H consisting of the product of the norm topologies associated with $|\cdot|_j$ for every j, no matter the choice of (δ_j). However, we would like to describe when V is compact in the norm topology (of $||\cdot||$) on H. This is achieved only under certain conditions, as stated in the following lemma.

Lemma 2.1. *Let $V_j \subseteq H_j$ where $|v_j| \leq r_j$ for every $v_j \in V_j$ for some sequence (r_j) and $V = \prod_{j=1}^{\infty} V_j$. If the sequence (δ_j) is such that $\sum_{j=1}^{\infty} \delta_j^2 r_j^2 < \infty$ then the norm topology (of $||\cdot||$) and the product norm topology on V are equivalent.*

Proof. See pages 120 and 153 of [25]. ∎

Along with this characterization of compactness of V in the norm topology, it is critical to understand the notion of continuity of linear functionals in the same topology. By the Riesz-Fréchet Theorem, continuous linear functionals over H are precisely of the form $\varphi(x) = (z|x)$ for $x \in H$, where z is another element of H. Consider the linear function $\varphi(x) = \sum_{j=1}^{\infty} a_j x_j$ where (a_j) is an arbitrary real sequence (not necessarily in H). The function φ is well-defined and continuous in the norm topology if there exists a sequence $(\tilde{a}_j) \in H$ such that $\sum_{j=1}^{\infty} a_j x_j = (\tilde{a}|x) = \sum_{j=1}^{\infty} \delta_j^2 \tilde{a}_j x_j$ for all $x \in H$. The above equation holds if $\tilde{a}_j = a_j/\delta_j^2$ where $||\tilde{a}||^2 = \sum_{j=1}^{\infty} \delta_j^2 \left|a_j/\delta_j^2\right|^2 = \sum_{j=1}^{\infty} |a_j|^2/\delta_j^2 < \infty$. We summarize this in the following lemma.

Lemma 2.2 (Continuity of linear functionals). *Given a real sequence (a_j), the linear functional $\varphi(x) = \sum_{j=1} a_j x_j$ over $x \in H$ is continuous in the norm topology if $\sum_{j=1}^{\infty} |a_j|^2/\delta_j^2 < \infty$.*

A sufficient condition for Lemma 2.2 is that there exists a $\rho \in (0, 1)$, scalar $\bar{a} < \infty$, and real sequence (α_j) such that $|a_j| \leq \bar{a}\alpha_j$ and $0 < \alpha_j < \delta_j$ with $0 < \alpha_j^2/\delta_j^2 < \rho^j$ for all j. Indeed, in this case

$$\sum_{j=1}^{\infty} \tfrac{1}{\delta_j^2}|a_j|^2 \leq \sum_{j=1}^{\infty} \tfrac{1}{\delta_j^2}\bar{a}^2\alpha_j^2 = \bar{a}^2 \sum_{j=1}^{\infty} \tfrac{\alpha_j^2}{\delta_j^2} < \bar{a}^2 \sum_{j=1}^{\infty} \rho^j = \bar{a}^2 \tfrac{\rho}{1-\rho} < \infty.$$

A particular choice that achieves this is to set δ_j to δ^j for some $\delta \in (0, 1)$ and α_j to α^j for some $\alpha \in (0, 1)$ where $\alpha/\delta < \rho$ for some $\rho \in (0, 1)$.

3. Countably infinite linear programs.

The problem under study in this paper is the countably infinite linear program (CILP):

$$\text{(P.1)} \qquad f^* := \inf_{x \in \mathbb{R}^{\mathbb{N}}} \sum_{j=1}^{\infty} c_j x_j,$$

$$\text{(P.2)} \qquad \text{(P)} \qquad \text{subject to } \sum_{j=1}^{\infty} a_{ij} x_j = b_i \text{ for } i = 1, 2, \ldots$$

$$\text{(P.3)} \qquad x \geq 0$$

where c_j, a_{ij}, and b_i are real numbers for all $i, j = 1, 2, \ldots$. Let c denote the sequence (c_j), b denote the sequence (b_i), and A denotes the doubly infinite matrix (a_{ij}).

The first task is to set conditions on the data so that an optimal extreme point solution of (P) is guaranteed to exist. The literature has imposed a variety of conditions on (P) to ensure an extreme point optimal solution exists (see [16] for a discussion). Our approach is different and leverages the Hilbert topology defined in Section 2.

First, we assume:

(A1) the set $\mathcal{F}$ of all feasible solutions to (P) is non-empty, and

(A2) there exists a nonnegative sequence $r = (r_j) \in \mathbb{R}^{\mathbb{N}}$ such that $|x_j| \leq r_j$ for every sequence $x = (x_j) \in \mathcal{F}$. We also assume that there is a $0 < \delta < 1$ such that $\sum_{j=1}^{\infty} \delta^j r_j < \infty$.

(A3) there exists an $\alpha \in (0, \delta)$ and an $\bar{a} < \infty$ such that
  (i) $|a_{ij}| \leq \bar{a}\alpha^j$ for all $i, j = 1, 2, \ldots$ and
  (ii) $|a_{ij}| \leq \bar{a}\alpha^i$ for all $i, j = 1, 2, \ldots$.

Let $X_j = [0, r_j]$ and set $X = \prod_{j=1}^{\infty} X_j$. Define the Hilbert space H with norm $||\cdot||_H$ as defined in (2.1) with $\delta_j = \delta^j$, where δ is defined in (A2). By Lemma 2.1 and Tychonoff's theorem, X is compact in the norm topology on H. It remains to discuss the continuity properties of the linear functions defining (P). A preliminary result is as follows.

Lemma 3.1. *Suppose (A2) and (A3) hold. The infinite series $\sum_{j=1}^{\infty} a_{ij}x_j$ is absolutely convergent for $i = 1, 2, \ldots$ and all $x \in H$ if $\alpha < \delta$.*

Proof. For all $i, j = 1, 2, \ldots$ we have the basic property that $|a_{ij}x_j| \leq |a_{ij}||x_j|$. This means that

$$\sum_{j=1}^{\infty} |a_{ij}x_j| \leq \sum_{j=1}^{\infty} |a_{ij}||x_j| = \sum_{j=1}^{\infty} \delta^{2j} \left(\tfrac{|a_{ij}|}{\delta^{2j}}\right) |x_j|$$
$$= ((|a_{ij}|/\delta^{2j}) \mid (|x_j|)) \leq ||(|a_{ij}|/\delta^{2j})||_H ||(x_j)||_H$$

where the second equality follows by multiplying and dividing term j in the sum by δ^{2j}, the third equality observes that this is the inner product of the vectors $(|a_{ij}|/\delta^{2j})$ and (x_j) in the Hilbert space H, and the final inequality is the Cauchy-Schwartz inequality. It thus remains to show that $||(|a_{ij}|/\delta^{2j})||_H ||(x_j)||_H < \infty$. We have assumed that $x \in H$ and so $||(x_j)||_H < \infty$, so it remains to show that $||(|a_{ij}|/\delta^{2j})||_H < \infty$. Observe that

$$||(|a_{ij}|/\delta^{2j})||_H = \sqrt{\sum_{j=1}^{\infty} \delta^{2j} \left(|a_{ij}|/\delta^{2j}\right)^2} = \sqrt{\sum_{j=1}^{\infty} |a_{ij}|^2/\delta^{2j}}$$
$$\leq \sqrt{\sum_{j=1}^{\infty} \bar{a}^2\alpha^{2j}/\delta^{2j}} = \tfrac{\bar{a}\alpha/\delta}{\sqrt{1-(\alpha/\delta)^2}} < \infty,$$

where the first inequality follows from (A3) and the second (strict) inequality follows under the assumption that $\alpha < \delta$. ☐

The last of our basic assumptions on the data ensures that the objective function is continuous in the same topology:

(A4) The sequence (c_j) is such that $\sum_{j=1}^{\infty} |c_j|^2/\delta_j^2 < \infty$.

Theorem 3.2 (Existence of optimal extreme point). *If (A1)–(A4) hold then (P) has an optimal extreme point solution.*

Proof. This follows from Bauer Maximum Principle (Theorem 7.69 in [3]) in the Hilbert norm topology. First, (A1) tells us the feasible region $\mathcal{F}$ is nonempty. As argued above, the set $X = \prod_{j=1}^{\infty} X_j = \prod_{j=1}^{\infty}[0, r_j]$ is compact, using (A2). Thus, it suffices to show that the sets $\{x \in H : \sum_{j=1}^{\infty} a_{ij}x_j = b_i\}$ are closed for $i = 1, 2, \ldots$, since then $\mathcal{F}$ is the intersection of X and these sets. The closedness of $\{x \in H : \sum_{j=1}^{\infty} a_{ij}x_j = b_i\}$ follows if $\sum_{j=1}^{\infty} a_{ij}x_j$ is a continuous function. It is straightforward to see (A3)(i) implies that $\sum_{j=1}^{\infty} |a_{ij}|^2/\delta_j^2 < \infty$ holds and so, by Lemma 2.2, the constraint functions in (P.2) are continuous. Hence, $\mathcal{F}$ is compact in the Hilbert norm topology. It is left to show that the objective function of (P) is well-defined, convex and continuous. Convexity follows from linearity, while well-definedness and continuity follow by (A4) and Lemma 2.2. ∎

Later, we will need to leverage structure on the range of the doubly infinite matrix A; that is, the space containing b. For now, we will assume that range space is another Hilbert space Y in $\mathbb{R}^{\mathbb{N}}$ defined by a norm as in Section 2 but now taking $\delta_j = \beta^j$ for some $\beta \in (0, 1)$. That is, for $y \in Y$ we have $||y||_Y^2 = \sum_{i=1}^{\infty} \beta^{2i} y_i^2$. The next result shows that when the linear map defined by A maps feasible solutions into Y.

LEMMA 3.3. *Suppose (A2) and (A3) hold. Then* $\operatorname{cspan}(A)$ *is a subspace of* Y *if* $0 < \beta < 1$ *and* $0 < \alpha < \delta < 1$.

Proof. Let $x \in H$ and set $y = Ax \in \mathbb{R}^{\mathbb{N}}$ by Lemma 3.1. This means $y_i = \sum_{j=1}^{\infty} a_{ij} x_j$ and $|y_i| \leq \sum_{j=1}^{\infty} |a_{ij} x_j| \leq \bar{a}(\alpha/\delta)/\sqrt{1-(\alpha/\delta)^2}||x||_H$ from the proof of Lemma 3.1. This then implies

$$||y||_Y = \sqrt{\sum_{i=1}^{\infty} \beta^{2i} |y_i|^2} \leq \sqrt{\sum_{i=1}^{\infty} \beta^{2i} \bar{a}^2 \tfrac{(\alpha/\delta)^2}{(1-(\alpha/\delta)^2)^2} ||x||_H^2} \tag{3.1}$$

$$= \bar{a} \tfrac{\alpha/\delta}{1-(\alpha/\delta)^2} ||x||_H \sqrt{\sum_{i=1}^{\infty} \beta^{2i}} = \bar{a} \tfrac{\alpha/\delta}{1-(\alpha/\delta)^2} ||x||_H \tfrac{\beta}{\sqrt{1-\beta^2}} < \infty \tag{3.2}$$

for $0 < \beta < 1$ since $||x||_H < \infty$ for all $x \in H$. This implies $y \in Y$. ∎

We now show that A defines a continuous linear operator. Recall (see, for instance, Chapter IV of [37]) that the **operator norm** $||L||$ of linear operator L is equal to $\sup_{x:||x||_H \leq 1} ||L(x)||_Y$. We say the linear map L is **continuous** (or equivalently **bounded**) if $||L|| < \infty$. This result is critical for establishing optimal policy convergence of the simplex algorithm we define below. The proof involves establishing an isometric isomorphism between H and ℓ^2 and using the Schur Test for boundedness of operators mapping ℓ^2 into ℓ^2 (see page 260 of [12]). Due to its technical nature, we place the proof in the appendix.

LEMMA 3.4 (Continuity of constraint operator). *Suppose (A2) and (A3) hold. The doubly infinite matrix A defines a continuous linear operator from H into Y if* $0 < \beta < 1$ *and* $0 < \alpha < \delta < 1$.

4. Extreme points and basic feasible solutions. As with finite-dimensional versions of the simplex method, our algorithm works with the algebraic characterization of extreme points as basic feasible solutions. Defining basic solutions is more delicate in the infinite-dimensional setting than in the finite setting (for an extended discussion see [4]). We make the following preliminary definitions.

DEFINITION 4.1. *We call* $B(x) \triangleq \{a_{\cdot j} : j \in \mathcal{S}(x)\}$ *the* ***active set*** *of columns of A associated with a feasible x.*

The name 'active set' comes from the fact that Ax is a linear combination of the columns in $B(x)$. That is, only the columns in $B(x)$ are 'active' in the product Ax. Informally, we may think of $B(x)$ as the 'support of columns of A' associated with x, whereas $\mathcal{S}(x)$ is the 'support of indices' of x.

DEFINITION 4.2. *A subset B of columns of A is a* ***basis*** *if*

(B1) $\{z : Az = 0, B(z) \subseteq B\} = \{0\}$.

(B2) $\operatorname{cspan}(B) = \operatorname{cspan}(A)$.

*We say B is a **basis of feasible solution** x if, additionally,*

(B3) $B(x) \subseteq B$

This condition is analogous to the familiar definition of a basis of an extreme point solution from finite-dimensional linear programming (see, for instance, Chapter 3 in [7]). Conditions (B1) and (B2) correspond to the fact that a basis forms a column basis of the constraint matrix, with (B1) yielding linear independence and (B2) a spanning condition. Condition (B3) captures the fact that nonbasic variables are set to zero. Strict containment in (B3) allows the possibility of basic variables taking a value of zero.

If B is a basis of A, then it determines a linear operator from H_B into Y where $H_B = \{x \in H \mid x_j = 0 \text{ for } j \notin \mathcal{S}(B)\}$ with $\mathcal{S}(B)$ denoting the set of indices of columns of A that are in B. We abuse notation and also let B denote this linear operator. We need another assumption on the structure of the constraint matrix A that yields the invertibility of our basis matrices.

(A5) The doubly infinite matrix A and scalar β are such that $A : H \to Y$ is an onto map. That is, $\operatorname{cspan} A = Y$.

LEMMA 4.3 (Continuity of bases in operator norm). *Suppose (A2), (A3) and (A5) hold, $0 < \beta < 1$ and $0 < \alpha < \delta < 1$. Let B be a basis of A. Then, the doubly infinite matrix B defines a continuous linear operator with an inverse B^{-1} that is also a continuous linear operator.*

Proof. The proof that B defines a continuous linear operator is nearly identical to that of Lemma 3.4 since B is a submatrix of A. See the appendix. The fact that B^{-1} exists comes from the definition of a basis. Indeed, property (B1) implies that B is one-to-one. Let w^1 and w^2 be such that $Bw^1 = Bw^2$. Note that w^1 and w^2 can be extended (by appending zeros) to vectors z^1 and z^2 such that $Az^1 = Az^2$ where $B(z^i) \subseteq B$ for $i = 1, 2$. Thus, according to (B1), $A(z^1 - z^2) = 0$, which implies $z^1 - z^2 = 0$ and so $z^1 = z^2$. This, in turn, implies $w^1 = w^2$ and B is a one-to-one mapping. The fact that B is onto follows from (B2) and (A5). Finally, by the Banach Inverse Theorem (see Theorem 1 on page 149 of [28]), B^{-1} is a continuous map from Y to H. □

DEFINITION 4.4. *A vector $x \in H$ is a **basic solution** if its admits a basis B (as defined in (B1)–(B3)). If a basic solution is feasible it is called a **basic feasible solution (bfs)**. If $B(x)$ is a basis of x then x is called a **nondegenerate bfs**.*

Given a basis B, one can construct an associated basic feasible solution. Recall that B is a subset of columns in A. Let x_B denote the elements of x that correspond to the columns in B, we call the elements of x_B **basic variables**. Let N denote the columns in A that are not in B. The elements in x_N are called **nonbasic variables**. Then, the basic solution associated with B satisfies $Bx_B = b$ and $x_N = 0$. Since B is invertible, we know $x_B = B^{-1}b$. The solution (x_B, x_N) is a basic feasible solution if and only if $B^{-1}b \geq 0$. We summarize this in the following result.

LEMMA 4.5. *If B be a basis then the solution $x = (x_B, x_N)$ with $x_B = B^{-1}b$ and $x_N = 0$ is a basic solution.*

Observe that if x is a nondegenerate bfs then $B(x)$ is its unique basis. In general, there is not a one-to-one correspondence between basic feasible solutions and extreme points (for a thorough discussion see [4], and in the specific context of CILPs see [17]). The following concepts help to resolve this challenge.

DEFINITION 4.6. *For any non-negative $x \in H$, let $\sigma(x)$ denote the infimal positive value of a component of x; that is, $\sigma(x) \triangleq \inf_{j \in \mathcal{S}(x)} x_j$. We say that x has* ***strictly positive support (SPS)*** *if $\sigma(x) > 0$.*

The concept of SPS first appeared in [31] and was later generalized to CILPs in [17]. Observe that a real sequence x can have all positive entries and yet fail to have SPS. Indeed, consider the vector (x_j) where $x_j = 1/j$ for $j = 1, 2, \ldots$. The following two assumptions align the algebraic and geometric notions of extreme points, and as we shall see in Remark 5.7 below, also insures that pivots move from an extreme point to a different extreme point:

(A6) every bfs of (P) is a nondegenerate bfs,

(A7) $\sigma \triangleq \inf_{x \in \text{ext} F} \sigma(x) > 0$. In particular, every extreme point of $\mathcal{F}$ has SPS.

In Section 9, we will see an example of a problem where these conditions hold. It is also straightforward to see that they do not hold in general. Failure of (A6) is common even in finite dimensional linear programming. As for assumption (A7), the binary tree in Figure 1 of [16] provides an example with a bfs that fails the SPS condition.

THEOREM 4.7 (Extreme points are basic feasible solutions). *Suppose (A6) and (A7) and the conditions of Theorem 3.2 hold. Then a feasible solution is extreme point if and only if it is a nondegenerate bfs. In particular, problem (P) has an optimal nondegenerate bfs.*

Proof. The 'if and only if' follows from Proposition 2.6 and Corollary 2.12 in [17]. The 'in particular' is then immediate from Theorem 3.2. ☐

5. Pivoting. The key step in any simplex method is pivoting – moving systematically from one bfs to another in a way that monotonically improves the objective value of the optimization problem.

Before exploring pivoting in the infinite-dimensional setting, we refresh the mechanics of a pivot in the finite-dimensional setting at a high level. This may help the reader visualize some of our development. We describe the finite setting only for the most well-behaved case where the problem is bounded and the basic feasible solutions involved are nondegenerate.

Pivoting involves selecting an appropriate nonbasic variable (called an **entering variable**) to add to B and selecting an appropriate basic variable (called a **leaving variable**) to remove from B. This results in a new basis of vectors B' that can be associated with a new bfs x'. In general, there is some choice over both the entering and leaving variables.

Geometrically, a pivot entails a movement from one extreme point of the feasible region to another along an edge. When an entering variable is chosen, it determines which edge is traversed by defining a **basic direction** d that takes a value of 1 in the component of the entering variable, zero on all other nonbasic variables, and otherwise satisfies the constraint $Ax = b$ to determine the values of d on the components of the basic variables. The new bfs x' equals the sum $x + \lambda d$ for some $\lambda \geq 0$. The value of λ is increased as the basic direction is traversed until the value of one of the basic variables hits zero (this is unique by nondegeneracy). The basic variable whose value first hits zero in $x' = x + \lambda d$ is the leaving variable.

Finally, which nonbasic variable to choose as an entering variable depends on its *reduced cost*. The reduced cost of a nonbasic variable is the change in objective value associated with its basic direction d; that is, $\sum_j c_j d_j$ where c is the objective vector of the linear program. Thus, an entering variable must be chosen among those nonbasic variables where $\sum_j c_j d_j$ improves the value of the objective. In the case of a minimization problem, this is precisely when $\sum_j c_j d_j < 0$. A key result in the finite-dimensional setting is that a basic feasible solution is optimal if it has no nonbasic variables with an improving reduced cost (Theorem 3.1 in [7]). This is the termination condition of the finite-dimensional simplex method.

We turn now to detail the infinite-dimensional setting. We highlight important differences with the finite-dimensional case as we proceed. We assume (A1)–(A7) throughout this discussion. By Theorem 3.2, a feasible extreme point solution x exists. By Theorem 4.7, x is a nondegenerate bfs.

DEFINITION 5.1. *Let x be a nondegenerate bfs and $k \in \mathcal{S}^c(x)$ the index of a nonbasic variable. The kth* ***basic direction*** *$d(x;k)$ with respect to x (or simply kth basic direction when the context is clear) is the* unique *vector $d \in H$ such that*

(BD1) $d_k = 1$,

(BD2) $d_j = 0$ for all $j \in \mathcal{S}^c(x)$ not equal to k,

(BD3) $Ad = 0$.

It is important to note that the basic direction depends on the current basis. That is captured directly in the notation $d(x;k)$.

The above definition asserts that there is a unique vector in H that satisfies (BD1)–(BD3). To see this, for (BD3) to hold, we must have for every constraint $i = 1, 2, \ldots$:

$$\sum_{j=1}^{\infty} a_{ij}d_j = \sum_{j \in \mathcal{S}(x)} a_{ij}d_j + \sum_{j \in \mathcal{S}^c(x)} a_{ij}d_j = \sum_{j \in \mathcal{S}(x)} a_{ij}d_j + a_{ik}d_k + \sum_{k \neq j \in \mathcal{S}^c(x)} a_{ij}d_j = 0 \tag{5.1}$$

using $d_k = 1$ by (BD1). This is equivalent to

$$\sum_{j \in \mathcal{S}(x)} a_{ij}d_j = -a_{ik}, \text{ for } i = 1, 2, \ldots \tag{5.2}$$

since $d_j = 0$ for all $j \in \mathcal{S}^c(x)$ not equal to k by (BD2). Our attention turns to analyzing (5.2).

Now, given a basic feasible solution, the set $B(x)$ is a basis. As shown in Lemma 4.3, this implies that $B(x)$ is an invertible linear operator with inverse $B(x)^{-1}$. We may write $d(x;k)$ into two components $(d_{B(x)}, d_{N(x)})$ where $N(x)$ consists of the columns of A not in $B(x)$. Then (5.2) is equivalent to writing $B(x)d_{B(x)} = -a_{\cdot k}$ where $a_{\cdot k}$ is the kth column of A: $d_{B(x)} = -B(x)^{-1}a_{\cdot k}$. Also, (BD1) implies $d_k = 1$ and $d_j = 0$ for $j \in S^c(x) \setminus \{k\}$. That is, $d_{N(x)} = e^k$ where e^k is the vector with a one in entry k and zero otherwise on $N(x)$. Putting this together we have

$$d(x;k) = (-B(x)^{-1}a_{\cdot k}, e_k) \tag{5.3}$$

The existence and uniqueness of d is thus a consequence of the properties of the matrix B and its inverse.

Condition (BD3) ensures that $x + \lambda d(x;k)$ satisfies constraint (P.2) for all $\lambda \in \mathbb{R}$ since $A(x + \lambda d(x;k)) = Ax + \lambda Ad(x;k) = b + 0 = b$, where $Ax = b$ since x is a feasible solution of (P). We next characterize the set of λ such that $x + \lambda d(x;k) \geq 0$; that is, (P.3) holds. If every component $d_j(x;k)$ of $d(x;k)$ is nonnegative then λ can be taken arbitrarily large and (P.3) continues to hold. The next result shows that, under our assumptions, this cannot happen.

LEMMA 5.2. *Suppose (A2) holds. Let x be a nondegenerate bfs and k be the index of a nonbasic variable at x. The set $\{j \in \mathcal{S}(x) : d_j(x;k) < 0\}$ is nonempty.*

Proof. Suppose not. Then, $d_j(x;k) \geq 0$ for all $j \in \mathcal{S}(x)$. Also recall that $d_k(x;k) = 1$ and $d_j(x;k) = 0$ for all $j \in \mathcal{S}^c(x)$ not equal to k. This implies that

$x + \lambda d(x;k) \geq 0$ and, in particular, $x + \lambda d(x;k) \in \mathcal{F}$ for all $\lambda \geq 0$ since both (P.2) and (P.3) are satisfied. This violates the boundedness assumption (A2). ∎

Given this lemma, we may look for the leaving variable associated with the basic direction k. Informally, the leaving variable is the basic variable that first reaches a value of zero along the basic direction. We need a few lemmas to make this precise.

The object of interest here is the **infimum ratio**

$$\lambda(x;k) \triangleq \inf\left\{\frac{x_j}{-d_j(x;k)} : j \in \mathcal{S}(x) \text{ such that } d_j(x;k) < 0\right\}. \tag{5.4}$$

Below (in Theorem 5.6) we show λ is well-defined and that there always exists a unique j that attains the infimum in (5.4).

Next, we show that $\lambda(x;k)$ behaves as expected, in the sense that it defines how far the feasible region extends in the basic direction $d(x;k)$.

LEMMA 5.3. *Let x be a nondegenerate bfs and k be the index of a nonbasic variable. Then $x + \lambda d(x;k) \geq 0$ for all $\lambda \in [0, \lambda(x;k)]$. Moreover, $x + \lambda d(x;k) \not\geq 0$ for $\lambda \notin [0, \lambda(x;k)]$.*

Proof. For the first part, consider any $0 \leq \lambda \leq \lambda(x;k)$. We only need to consider $j \in \mathcal{S}(x)$ for which $d_j(x;k) < 0$ (because $d_j(x;k) \geq 0$ for all other j and hence $x_j + \lambda d_j(x;k) \geq 0$ for those j). For any such j, we have, $x_j + \lambda d_j(x;k) \geq x_j + \lambda(x;k) d_j(x;k) \geq x_j + \frac{x_j}{-d_j(x;k)} d_j(x;k) = 0$ as claimed.

For the second part, first consider any $\lambda > \lambda(x;k)$. We need to show that there is a $j \in \mathcal{S}(x)$ such that $x_j + \lambda d_j(x;k) < 0$. Any such j must be such that $d_j(x;k) < 0$. There are two possibilities. The first one is that the infimum ratio is attained for some j, say j^*. Then, $x_{j^*} + \lambda d_{j^*}(x;k) < x_{j^*} + \lambda(x;k) d_{j^*}(x;k) = x_{j^*} + \frac{x_{j^*}}{-d_{j^*}(x;k)} d_{j^*}(x;k) = 0$. The second one is that the infimum ratio is not attained. Suppose $\lambda = \lambda(x;k) + \epsilon$ for some $\epsilon > 0$. Now, by definition of the infimum, there exists a j^* such that $\frac{x_{j^*}}{-d_{j^*}(x;k)} < \lambda(x;k) + \epsilon$, and for this j^*, we have, $x_{j^*} + \lambda d_{j^*}(x;k) = x_{j^*} + (\lambda(x;k) + \epsilon) d_{j^*}(x;k) < 0$. Finally, if $\lambda < 0$, then $x_k + \lambda d_k(x;k) = 0 + \lambda < 0$. ∎

It remains to define the leaving variable. Any x_j such that j achieves the infimum in the definition of $\lambda(x;k)$ in (5.4) is a candidate (by nondegeneracy there exists at most one such index). However, it is not clear whether or not this infimum is attained. Indeed, in the CILP setting, a leaving variable may not exist in general.

Under our assumptions, however, we show that a leaving variable always exists in every basic direction. Our proof of this requires geometric reasoning. We first show that $x' \triangleq x + \lambda d(x;k)$ from the previous lemma is an extreme point (see Proposition 5.5 below). In the process, we show that each basic direction goes along an 'edge' of the feasible region (a precise definition of 'edge' is given). This conforms with our intuition from the finite-dimensional setting that pivots occur along edge directions.

Having established x' is an extreme point, we will use Theorem 4.7 to conclude that x' is a nondegenerate bfs. This algebraic property of x' rules out the possibility that the infimum in (5.4) is not attained. Details of this argument are in Theorem 5.6.

We start with a formal definition of extremality that captures the notion of extreme points as a special case. For (P.3) extended discussion of extremality in general infinite-dimensional vector spaces, see Section 7.12 in [3].

DEFINITION 5.4. **(Extreme subset)** *Let S be a non-empty subset of $\mathbb{R}^{\mathbb{N}}$. A non-empty subset $E \subset S$ is called **S-extreme** if it has the following property: if $x, y \in S$ and if there exists a t, $0 < t < 1$ such that $tx + (1-t)y \in E$, then x, y necessarily belong to E. A 0-dimensional extreme subset is a called an **extreme point** of S. A 1-dimensional extreme subset of is called an **edge** of S.*

PROPOSITION 5.5. *Suppose (A1)–(A7) hold, x is a nondegenerate bfs, and k is the index of a nonbasic variable. Then,*

(i) the set $\mathcal{Z}(x;k) \triangleq \{z \in H : z = x + \lambda d(x;k),\ \lambda \in [0, \lambda(x;k)]\}$ is an edge of $\mathcal{F}$, and

(ii) $x + \lambda(x;k)d(x;k)$ is an extreme point of $\mathcal{F}$.

Proof. See appendix. ∎

THEOREM 5.6 (Existence and uniqueness of leaving variable). *Suppose the condition of Theorem 4.7 hold and let x be a nondegenerate bfs and k be the index of a nonbasic variable. There exists a unique leaving basic variable; that is, there exists a unique $j^* \in \mathcal{S}(x)$ with $d_j(x;k) < 0$ that attains the infimum ratio in (5.4). Thus, $x' \triangleq x + \lambda(x;k)d(x;k)$ is a nondegenerate bfs with basis $B(x') = B(x) \cup \{a_{\cdot k}\} \setminus \{a_{\cdot j^*}\}$.*

Proof. By Proposition 5.5, x' is an extreme point of $\mathcal{F}$ and thus by Theorem 4.7, x' is a nondegenerate bfs. Suppose by way of the contradiction that there is no leaving basic variable when pivoting in nonbasic variable x_k to form x'. We will contradict property (B1) of the basis $B(x')$ of x'.

Since there is no leaving basic variable, this means that $\mathcal{S}(x') = \mathcal{S}(x) \cup \{k\}$. Indeed, by the definition of $d(x;k)$ we have $x'_k > 0$, $x'_j = 0$ for $j \in \mathcal{S}^c(x)$ and since the infimum is not attained for any $j \in \mathcal{S}(x)$, we must also have $x'_j > 0$.

Let $z \triangleq x' - x$. Note that $B(x) \subseteq B(x')$ since, as we have just argued, $\mathcal{S}(x) \subseteq \mathcal{S}(x')$. For all $i = 1, 2, \ldots$

$$\sum_{j=1}^{\infty} a_{ij} z_j = \sum_{j \in \mathcal{S}(x')} a_{ij} z_j = \sum_{j \in \mathcal{S}(x')} a_{ij} x'_j - \sum_{j \in \mathcal{S}(x')} a_{ij} x_j$$

$$= \sum_{j \in \mathcal{S}(x')} a_{ij} x'_j - \sum_{j \in \mathcal{S}(x)} a_{ij} x_j = b_i - b_i = 0$$

and thus $Az = 0$. Since $z \neq 0$ this contradicts property (B1) of the basis $B(x')$ of nondegenerate bfs x'. Clearly, $B(x') = B(x) \cup \{a_{\cdot k}\} \setminus \{a_{\cdot j^*}\}$. ∎

This result shows that, under our assumptions, every basic direction admits a unique leaving variable (uniqueness invokes nondegeneracy).

Remark 5.7. By (BD1) in Definition 5.1, the value of the entering variable in the new basic feasible solution x' is $\lambda(x;k)$, since $x' = x + \lambda(x;k)d(x;k)$. Thus, if we assume (A6) and (A7), we must have $\lambda(x;k) > \sigma$. That is, every pivot operation "moves" to a different bfs.

6. Reduced costs and optimality conditions.

In this section, we explore the properties of entering nonbasic variables. This discussion leads to establishing an optimality condition for CILPs based on pivoting, which serves as the condition for optimal termination of our simplex method.

DEFINITION 6.1. *Let x be a nondegenerate bfs and k the index of a nonbasic variable. The **reduced cost** $r(x;k)$ of nonbasic variable k at basis x is the sum $\sum_{j=1}^{\infty} c_j d_j(x;k)$. Using the structure of $d(x;k)$ detailed in (BD1)–(BD3), the reduced cost is typically expressed as $r(x;k) \triangleq c_k + \sum_{j \in \mathcal{S}(x)} c_j d_j(x;k)$.*

An alternate way of writing reduced cost is using matrix notation. Recalling our expression for $d(x;k)$ in (5.3), we may write the reduced cost as $r(x;k) = c_k - \sum_{j \in \mathcal{S}(x)} c_j (B(x)^{-1} a_{\cdot k})_j$ or as a **reduced cost vector** $r(x) = c - c_{B(x)}^{\top} B(x)^{-1} A$ with entries $r(x;k)$ and where $c_{B(x)}^{\top} B(x)^{-1} A$ denotes the sum $\sum_{j \in S(x)} c_j (B(x))^{-1} A)_j$.

Note that here $r(x;k) = 0$ for any basic variable $k \in S(x)$. Moreover,

$$r(x; N(x)) \triangleq (r(x;k) : k \notin S(x)) = c_{N(x)} - c_{B(x)}^\top B(x)^{-1} N(x). \tag{6.1}$$

By our assumptions on c and d, the reduced cost vector is well-defined. Moreover, it is critical to note that the reduced cost of a nonbasic variable depends on the basis of the current bfs.[2] This is reflected in our choice of notation $r(x;k)$ and $r(x)$.

The reduced cost allows us to succinctly capture the change in objective value when pivoting from x to $x' \triangleq x + \lambda(x;k)d(x;k)$, which is equal to

$$\sum_{j=1}^{\infty} c_j x'_j - \sum_{j=1}^{\infty} c_j x_j = \lambda(x;k) \sum_{j=1}^{\infty} c_j d_j(x;k) = \lambda(x;k) r(x;k) \tag{6.2}$$

and so pivoting in a nonbasic variable with negative reduced cost will strictly improve the objective value over the current feasible solution of (P) (recall that when (A6) and (A7) hold, $\lambda(x;k) > 0$, as discussed in Remark 5.7).

The set $\mathcal{T}(x) \triangleq \{k \in \mathcal{S}^c(x) : r(x;k) < 0\}$ of nonbasic variables at x with negative reduced costs are the candidate choices for entering variables in a pivot. The main result of this section is to show, under certain conditions, that if $\mathcal{T}(x) = \emptyset$ then we can conclude that x is an optimal solution. This implies that the basic directions are a sufficient set of improving directions.

THEOREM 6.2 (Optimality condition). *Suppose (A4) and the conditions of Lemma 3.3 hold. If x is a bfs and $r(x) \geq 0$ then x is an optimal solution.*

Proof. Suppose $r(x) \geq 0$ for some bfs x. For notational simplicity let B denote the basis $B(x)$ of x and let N denote $N(x)$.

Let y be any feasible solution and let $z \triangleq y - x$. Since x and y are both feasible and thus $Ax = Ay = b$, we have $Az = 0$ since A is a linear operator. As above, we write z as $z = (z_B, z_N)$ so that $0 = Az = Bz_B + Nz_D$. Since B is invertible, multiplying both sides by B^{-1} yields $0 = B^{-1}Bz_B + B^{-1}Nz_N$ and so $z_B = -B^{-1}Nz_N$. Hence, we have

$$c^\top z = (c_N - c_B^\top B^{-1} N) z_N \qquad \text{(more details below)} \tag{6.3}$$

$$= r(x;N)^\top z_N. \qquad \text{(using (6.1))} \tag{6.4}$$

We give some more details on (6.3). In finite dimensions, this step is trivial, here it requires some additional reasoning.

Let $c_N = (\nu_1, \nu_2, \ldots)$, $c_B^\top B^{-1} N = (\mu_1, \mu_2, \ldots)$, and $z_N = (\eta_1, \eta_2, \ldots)$. The goal is to show that (to yield (6.3)): $\sum_{k=1}^{\infty} \nu_k \eta_k - \sum_{k=1}^{\infty} \mu_k \eta_k = \sum_{k=1}^{\infty} (\nu_k - \mu_k)\eta_k$, and this holds as long as each sum on the left-hand side is finite. We first argue that the sum $\sum_{k=1}^{\infty} \nu_k \eta_k$ is finite. Note that $z_N \in H$ since $z \in H$ and c_N satisfies the condition $\sum_{k=1}^{\infty} |\nu_k|^2/\delta_k^2 < \infty$ since c satisfies (A4). By Lemma 2.2, the sum $c^\top z$ is finite, which implies $c_N^\top z_N = \sum_{k=1}^{\infty} \nu_k \eta_k$ is also finite. Next, recall that $\sum_{k=1}^{\infty} \mu_k \eta_k = c_B^\top B^{-1} N z_N$ where the right-hand side is finite for the following reasons. We know $z_N \in H$ and so $Nz_N \in Y$ by Lemma 3.3. Thus, $B^{-1}Nz_N$ is again in H since B^{-1} maps Y to H. By similar reasoning as for the previous sum, we can thus conclude that $c_B^\top(B^{-1}Nz_N)$ is finite. This allows us to conclude (6.3).

Now, observe that $x_N = 0$ by definition of a basic variable, and so $z_N = y_N - x_N = y_N \geq 0$ since y is feasible and thus satisfies (P.3). Moreover, by hypothesis, $r(x;N) \geq 0$. This implies that $r(x;N)^\top z_N \geq 0$ and so from (6.4), $c^\top z \geq 0$ and thus $c^\top y \geq c^\top x$ for all feasible y. This implies that x is an optimal solution. ∎

[2] When degeneracy is allowed, different bases for the same basic feasible solution may yield different reduced costs for nonbasic variables. Under (A6), a single basis exists and so there is a unique reduced cost for a nonbasic variable at any bfs.

7. An (abstract) simplex method. Given our description of pivoting in Section 5 and optimality condition in Theorem 6.2, we are now ready to state our simplex method. We should note that we do not claim the finite implementability of this method, merely that each operation is well-defined and the termination condition is valid. For this reason, we call our simplex method "abstract" — additional structure or assumptions are needed to implement it in general. Issues of finite implementability have been discussed for special cases in the literature [19, 26, 36].

Since we have assumed that every basic solution is nondegenerate in (A6), any choice of entering variable suffices because there is no chance of cycling (that is, returning to a previously visited basic feasible solution). Indeed, as long as there is an entering variable k with negative reduced cost $r(x;k) < 0$, Remark 5.7 shows that $\lambda(x;k) > \sigma$ and so by (6.2) the objective value strictly drops with each pivot. Hence, cycling is not possible. Thus, property (P1) holds for our simplex method. The next results structures the possible reduced costs.

LEMMA 7.1. *Suppose (A4) and the conditions of Lemma 3.3 hold. For every bfs x, let $\mathcal{T}(x) = \{k_1, k_2, \dots\}$ be the set of indices on nonbasic variables, taking $k_1 \leq k_2 \leq \cdots$ without loss. Then either $\mathcal{T}(x)$ is finite (possibly empty) or $\lim_{\ell\to\infty} r(x;k_\ell) = 0$.*

Proof. It suffices to show that if $\mathcal{T}(x)$ is not finite then $\lim_{\ell\to\infty} r(x;k_\ell) = 0$. From the definition of reduced cost, we have $r(x;k) = c_k - c_{B(x)}^\top B(x)^{-1} a_{\cdot k}$ for any $k \in \mathcal{T}(x)$. Note that $a_{\cdot k} \in Y$ since $a_{\cdot k} \in \operatorname{cspan}(A) \subseteq Y$ by Lemma 3.3. Hence $|r(x;k)| \leq |c_k| + |c_{B(x)}((B(x))^{-1} a_{\cdot k})|$. Now,

$$|c_{B(x)}^\top B(x)^{-1} a_{\cdot k})| \leq \|c_{B(x)}\|_H \|B(x)^{-1} a_{\cdot k})\|_H \leq ||c_{B(x)}||_H ||B(x)^{-1}||_L ||a_{\cdot k}||_Y \tag{7.1}$$

where $||\cdot||_L$ is the operator norm for the space $L(H,Y)$ of continuous linear operators mapping H into Y. Hence, $|r(x;k)| \leq |c_k| + ||c_{B(x)}||_H ||B(x)^{-1}||_L ||a_{\cdot k}||_Y$. From the proof of Lemma 3.3 we can conclude $||a_{\cdot k}||_Y \to 0$ as $k \to \infty$. Indeed, since $a_{\cdot k} = Ae^k$, where e^k is the unit vector with $e_k^k = 1$ and $e_j^k = 0$ otherwise, we have from (3.2) that

$$||a_{\cdot k}|| \leq \bar{a} \tfrac{\alpha/\delta}{1-(\alpha/\delta)^2} \tfrac{\beta}{\sqrt{1-\beta^2}} ||e^k||_H = \bar{a} \tfrac{\alpha/\delta}{1-(\alpha/\delta)^2} \tfrac{\beta}{\sqrt{1-\beta^2}} \delta^k$$

that converges to 0 as $k \to \infty$. Also $\|c_{B(x)}\|_H < \infty$ and $\|B(x)^{-1}\|_L < \infty$ since they are bounded linear functionals and operators respectively, and $|c_k| \to 0$ as $k \to \infty$ by (A4). Taken together, we can use this to conclude that $\lim_{\ell\to\infty} r(x;k_\ell) = 0$. ◻

LEMMA 7.2 (Most negative reduced cost). *Let x be a bfs. If $\mathcal{T}(x)$ is nonempty, then the most negative reduced cost $r_* \triangleq \inf_{k\in\mathcal{T}(x)} r(x;k)$ is* attained *by some nonbasic variable $k_* \in \mathcal{T}(x)$.*

Proof. Let $\epsilon = r(x;k_1) < 0$. By Lemma 7.1, there exists an index $\bar{\ell}$ such that $r(x;k_\ell) > \epsilon$ for all $\ell > \bar{\ell}$. Thus, $\inf_{k\in\mathcal{T}(x)} r(x;k) = \min\{r(x;k_\ell) : \ell = 0, 1, \dots, \bar{\ell}\}$. The latter is a finite set and so the minimum is clearly attained by some $k^* \in \{0, 1, \dots, \bar{\ell}\}$. ◻

We now have all of the ingredients to state our simplex method.

SIMPLEX METHOD

1. *(Initialization)* Let x^1 denote an initial bfs of (P). Set an iteration counter m to 1.
2. *(Compute reduced costs)* Compute reduced costs $r(x^m; k)$ for all nonbasic variables $x \in \mathcal{S}^c(x^m)$.
3. *(Optimality test and termination)* If $r(x^m; k) \geq 0$ for all $k \in \mathcal{S}^c(x^m)$, return x^m as an optimal solution and terminate.
4. *(Determine entering variable)* Otherwise, select as entering variable $x_{k_*^m}$, a variable with the most negative reduced cost (as defined in Lemma 7.2).
5. *(Pivot)* Determine a new bfs $x' \triangleq x^m + \lambda(x^m; k_*^m) d(x^m; k_*^m)$.
6. *(Update bfs)* Set $x^m \leftarrow x'$ and $m \leftarrow m + 1$. Continue at Step 2.

We briefly justify the steps of the algorithm. The optimality test in Step 3 suffices to conclude optimality by Theorem 6.2. The pivoting step (Step 5) is discussed in detail in Section 5, where the objects $\lambda(x^m; k_*^m)$ and $d(x^m; k_*^m)$ are discussed. The fact that x' is again a bfs was established in Theorem 5.6.

LEMMA 7.3 (Reduced costs converge to zero). *Suppose (A6) and (A7) and the conditions of Theorem 5.6 and Lemma 7.2 hold. The most negative reduced cost r_*^m at iteration m converges to zero as $m \to \infty$. That is, for any $\epsilon > 0$, there exists an iteration counter M_ϵ such that $-\epsilon < r_*^m \leq 0$ for all iterations $m \geq M_\epsilon$.*

Proof. Suppose not. There exists a subsequence of iterations m_n in which $r_{m_n}^* \leq -\epsilon$ (note that $r_{m_n}^*$ exists for each m_n by Lemma 7.2 and Theorem 5.6). Since the value of the entering basic variable at the end of iteration m_n is $\lambda(x^{m_n}; k_n)$, Remark 5.7 implies that $\lambda(x^{m_n}; k_n) \geq \sigma$ since (A6) and (A7) hold. Therefore, the objective function is reduced by at least $\sigma\epsilon$ in each one of these iterations, since the entering variable in Step 4 of the simplex method has reduced cost $r_{m_n}^* \leq -\epsilon$. But this is impossible since the sequence of function values $c^\top x^{m_n}$ is bounded below by f^*. ∎

We do not discuss how to determine an initial basic feasible solution. This remains an open challenge for many papers on CILP (see, for instance, [16, 32, 36]). In certain contexts (like those we discuss in Section 9), a starting basic feasible solution can be determined by inspection. More generally, a Big M approach seems appropriate.

8. Convergence to optimality. We now show that our simplex algorithm satisfies property (P2). More precisely, we will say our algorithm has **optimal value convergence** if the values of the sequence of iterates x^m converge to the optimal value f^* of (P). More formally, let $f^m \triangleq c^\top x^m$. Our goal is to show that $f^m \to f^*$ as $m \to \infty$. Of course, if the algorithm terminates, the optimal value f^* is attained. The interesting case is when the algorithm never terminates.

To show optimal convergence we need one final assumption. To state it we define a topology for the subsets of columns of A that allows us to talk about convergence of bases. Let B be a subset of columns of A. Then, the sequence $j^B = (j_1^B, j_2^B, \dots)$ where $j_i^B \in \{0, 1\}$ for all i encodes a subset of columns in A where $j_i^B = 1$ if column $a_i \in B$ and 0 otherwise. We encode convergence of bases "column by column" via convergence in this space of sequences. Let I be the set of all $\{0, 1\}$ sequences and define the product discrete topology on I where j^{B^m} converges to j^{B^*} if for every i there exists an m_i such that $j^{B^m} = j^{B^*}$ for all $m \geq m_i$. In other words, convergence corresponds to "lock in" in every element. We say a sequence $\{B^m\}$ of subsets of columns of A **converges** to another subset B^* of columns of A if and only if j^{B^m} converges to j^{B^*} in the above product discrete topology on I. It is straightforward to see that the resulting topology on subsets of columns of A is a homeomorphism for the product discrete topology on I. We say a collection of subsets of columns of A is **closed** if the limit of every convergent sequence taken from this collection is also contained in the collection.

(A8) The set $\mathcal{B} \triangleq \{B(x) : x \text{ is a bfs of (P)}\}$ is closed.[3]

The next section explores an example where (A8) holds. It is worth noting that there are very natural settings where this assumption fails. Consider the min-cost flow setting of [32] but now relax the condition that the graph G contains no infinite directed cycles. Indeed, consider the graph that consists of a single infinite directed cycle. Removing a single edge from this cycle yields a bfs corresponding to a spanning tree. Consider the sequence of bfs's that arise by successively removing edges along the outward directed portion of the infinite directed cycle. This sequence of bfs's converges in the product discrete topology to the entire infinite directed cycle, which is clearly not a bfs.

LEMMA 8.1 (Bases converge in product discrete topology). *Suppose assumption (A8) holds. Let $(B^m : m = 1, 2, \dots)$ be a sequence of bases. Then there exists a subsequence B^{m_n} and a basis B^* such that B^{m_n} converges to B^* in the product discrete topology.*

Proof. To prove the lemma it suffices to show that the set $\mathcal{B}$ of bases is sequentially compact in the product discrete topology. Since closed subsets of sequentially compact spaces are sequentially compact, by assumption (A8), it suffices to show that the set of all columns of A is a sequentially compact space under the product discrete topology described above. Indeed, the product discrete topology on A is metrizable and compact by Theorems 2.61 and 3.36 in [3]. Compact subspaces of metric spaces are sequentially compact (Theorem 3.28 in [3]) and thus the product discrete topology on A is sequentially compact. ∎

Convergence in the product discrete topology is not a standard notion of convergence of linear operators. Accordingly, some work needs to be done to leverage this condition.

First, we show that convergence in the product discrete topology implies the more common notion of convergence in operator norm. The difficulty here is that, as an operator, we think of each B defining an invertible operator on a different space. That is, the basis B defines the invertible operator $B : H_B \to Y$ where H_B is defined above Lemma 4.3. It is important in the arguments that follow to redefine B over a common domain. Let B be the basis of A that consists of columns of A indexed by j_k for $k = 1, 2, \dots$. Let T_B denote the mapping from ℓ^2 into H_B with $T_B(x) = x'$ where

$$x'_j = \begin{cases} x_k/\delta^{j_k} & \text{if } j = j_k \text{ for } k = 1, 2, \dots \\ 0 & \text{otherwise} \end{cases} \tag{8.1}$$

Thus, we can define $\tilde{B} := BT_B$, which remains an invertible and continuous linear operator from ℓ^2 into Y since both B (by Lemma 4.3) and T_B (trivially) are invertible and continuous linear operators.

Suppose L_m (for $m = 1, 2, \dots$) and L are bounded linear maps between ℓ^2 and Y. Then we say that the L_m converge to L in **operator norm** if $||L_m - L|| \to 0$ as $m \to \infty$ (where here, $||\cdot||$ denotes the operator norm). This is equivalent to the statement that $||L_m x - Lx||_Y \to 0$ uniformly for all $x \in \ell^2$ such that $||x||_{\ell^2} \le 1$.

Consider the linear operators $\tilde{B}^m$ and $\tilde{B}^*$, where B^m and B^* are defined as above.

[3]The fact that $\mathcal{B}$ is the collection of *all* bases relies on the assumption that all basic feasible solutions are nondegenerate (B2) and thus every basis is of the form $B(x)$ for some bfs x.

The following result shows that convergence of B^{m_n} to B^* in the product discrete topology implies that $\tilde{B}^{m_n} \to \tilde{B}^*$ in the operator norm.

LEMMA 8.2 (Bases converge in operator norm). *Suppose (A3), the conditions of Lemma 8.1 hold, and $0 < \alpha < \delta < 1$. Then the subsequence of linear operators $\tilde{B}^{m_n}$ converges to $\tilde{B}^*$ in the operator norm (where B^{m_n} and B^* are defined in Lemma 8.1).*

Proof. By Lemma 8.1, the B^{m_n} converges to B^* in the product discrete topology. To simplify notation, we let $\tilde{B}^n$ denote the linear operator $\tilde{B}^{m_n}$ from ℓ^2 to Y defined by $\tilde{B}^{m_n} = B^{m_n} T_{B^{m_n}}$ where $T_{B^{m_n}}$ is defined in (8.1). To show $\tilde{B}^n \to \tilde{B}^*$ in the operator norm we must show $||\tilde{B}_n x - \tilde{B}^* x||_Y \to 0$ uniformly for all x with $||x||_{\ell^2} \leq 1$. Let $x \in \ell^2$ be such that $||x||_{\ell^2} \leq 1$. Using the above constructs, we have $\tilde{B}x = B(T_B x) = Bx' = B(x_k/\delta^{j_k}) = (a_{\cdot j_1}/\delta^{j_1}, a_{\cdot j_2}/\delta^{j_2}, \dots)x$. Hence, we have $\tilde{B}^n x = \sum_{k=1}^{\infty} \delta^{-j_k^n} x_k a_{\cdot j_k^n}$ and $\tilde{B}^* x = \sum_{k=1}^{\infty} \delta^{-j_k^*} x_k a_{\cdot j_k^*}$ (where we use the shorthand $j_k^{m_n}$ to denote $j_k^{B^{m_n}}$ and j_k^* to denote $j_k^{B^*}$) so that

$$\tilde{B}^n x - \tilde{B}^* x = \sum_{k=k_n+1}^{\infty} (\delta^{-j_k^n} x_k a_{\cdot j_k^n} - \delta^{-j_k^*} x_k a_{\cdot j_k^*}) = \sum_{k=k_n+1}^{\infty} (\delta^{-j_k^n} a_{\cdot j_k^n} - \delta^{-j_k^*} a_{\cdot j_k^*}) x_k$$

since $j_k^n = j_k^*$ for $k \leq k_n$ for some k_n for each n where $k_n \to \infty$ as $n \to \infty$. This follows from the fact B^n converges to B^* in the product discrete topology. Thus, we have

$$||\tilde{B}^n x - \tilde{B}^* x||_Y \leq \sum_{k=k_n+1}^{\infty} ||(\delta^{-j_k^n} a_{\cdot j_n^k} - \delta^{-j_k^*} a_{\cdot j_k^*}) x_k||_Y$$

$$= \sum_{k=k_n+1}^{\infty} \sqrt{\sum_{i=1}^{\infty} \beta^{2i} |\delta^{-j_k^n} a_{ij_k^n} - \delta^{-j_k^*} a_{ij_k^*}|^2 |x_k|^2} \tag{8.2}$$

By (A3), we have $a_{ij_k^n} \leq \bar{a}\alpha^{j_k^n}$ and $a_{ij_n^*} \leq \bar{a}\alpha^{j_k^*}$. The significance of this bound is that we can unravel much of the dependency of the square root terms in (8.2) on the index i, yielding:

$$||\tilde{B}^n x - \tilde{B}^* x||_Y \leq \sum_{k=k_n+1}^{\infty} \sqrt{\sum_{i=1}^{\infty} \beta^{2i} \bar{a}^2 |\delta^{-j_k^n} \alpha^{j_k^n} - \delta^{-j_k^*} \alpha^{j_k^*}|^2 |x_k|^2}$$

$$= \bar{a} \sum_{k=k_n+1}^{\infty} |(\tfrac{\alpha}{\delta})^{j_n^k} - (\tfrac{\alpha}{\delta})^{j_n^*}| |x_k| \sqrt{\sum_{i=1}^{\infty} \beta^{2i}}$$

$$= \tfrac{\bar{a}\beta}{\sqrt{1-\beta^2}} \sum_{k=k_n+1}^{\infty} |(\tfrac{\alpha}{\delta})^{j_n^k} - (\tfrac{\alpha}{\delta})^{j_n^*}| |x_k|$$

$$\leq \tfrac{\bar{a}\beta}{\sqrt{1-\beta^2}} \gamma^{k_n} \sum_{k=1}^{\infty} |\gamma^k + \gamma^k| |x_{k+k_n}| \tag{8.3}$$

where, in the last step, $\gamma = \alpha/\delta$ and since $j_k^n \geq k$ and $j_k^* \geq k$. Finally we can develop the remaining sum in (8.3) as follows:

$$\sum_{k=1}^{\infty} |\gamma^k - \gamma^k| |x_{k+k_n}| = 2 \sum_{k=1}^{\infty} \gamma^k |x_{k+k_n}| \leq 2 \sum_{k=1}^{\infty} \gamma^k = 2 \tfrac{\gamma}{1-\gamma}$$

where the inequality follows since $||x||_{\ell^2} \leq 1$. Returning to (8.3), we have

$$||\tilde{B}^n x - \tilde{B}^* x||_Y \leq \tfrac{2\bar{a}\alpha\beta\gamma}{\sqrt{1-\beta^2}(1-\gamma)} \gamma^{k_n}$$

Since $\gamma < 1$ and $k_n \to \infty$ as $n \to \infty$, and the fact that right-hand side of the above equation does not depend on x for any $x \in \ell^2$, we have $\tilde{B}^n \to \tilde{B}^*$ in operator norm, completing the proof. ∎

We can now state and prove the main result of the paper.

THEOREM 8.3 (Optimal value convergence). *Suppose (A1)–(A8) hold with $0 < \beta < 1$ and $0 < \alpha < \delta < 1$ and the* SIMPLEX METHOD *does not terminate. Let $f^m \triangleq \sum_{j=1} c_j x_j^m$ be the sequence of values of iterates x^m of the* SIMPLEX METHOD. *Then $f^m \to f^*$. Moreover, there exists a subsequence of the x^m that converge to an optimal solution x^*.*

Proof. By Lemmas 8.1 and 8.2, there exists a subsequence of bases B^{m_n} of that converges to a basis B^* in the product discrete topology and associated maps $\tilde{B}^{m_n}$ that converge to $\tilde{B}^*$ in the operator norm. As noted below equation (8.1), each of the $\tilde{B}^{m_n}$ are continuous and invertible maps from ℓ^2 to Y. Let Φ denote the mapping that sends invertible operators to their inverse; that is, $\Phi(\tilde{B}) = \tilde{B}^{-1}$. By Theorem IV.1.5 in [37],[4] the mapping Φ is continuous. This implies that $(\tilde{B}^{m_n})^{-1}$ converges to $(\tilde{B}^*)^{-1}$ in the operator norm.

Let $x^{m_n} = (B^{m_n})^{-1}b$ and $x^* = (B^*)^{-1}b$. Accordingly, $x^{m_n} = T^{-1}_{B^{m_n}}(\tilde{B}^{m_n})^{-1}b$ and $x^* = T^{-1}_{B^*}(\tilde{B}^*)^{-1}b$. It is straightforward to see that since B^{m_n} converges to B^* in the product discrete topology, we have $T_{B^{m_n}} \to T_{B^*}$ and thus $T^{-1}_{B^{m_m}} \to T^{-1}_{B^*}$ again by appealing to Theorem IV.1.5 in [37]. Hence, we have $x^{m_n} = T^{-1}_{B^{m_n}}(\tilde{B}^{m_n})^{-1}b \to T^{-1}_{B^*}(\tilde{B}^*)^{-1}b = x^*$ since $T^{-1}_{B^{m_n}} \to T^{-1}_{B^*}$ and $(\tilde{B}^{m_n})^{-1} \to (\tilde{B}^*)^{-1}$, both in the operator norm. That is, there exists a subsequence of the x^m that converge to a basic solution x^* in the norm topology of H. Moreover, since $(B^{m_n})^{-1}b \geq 0$, because each of the x^{m_n} is a basic feasible solution, we can conclude that $(B^*)^{-1}b \geq 0$ by continuity. This implies that x^* is a basic feasible solution.

Finally, we claim that x^* is an optimal solution. To do so, we use Theorem 6.2 and show that the reduced costs $r(x^*; k) \geq 0$ for all $k \in S^c(x^*)$. Recall the definition of reduced cost has $r(x^*; k) = c_k + \sum_{j \in S^*} c_j (B^*)^{-1} a_{\cdot k}$, where S^* is the support of x^* and $k \notin S^*$. Similarly, let S^{m_n} denote the support of x^{m_n}.[5] We will show that $r(x^{m_n}; k) \to r(x^*; k)$ as $n \to \infty$ for all $k \notin S^*$. Indeed,

$$
\begin{aligned}
|r(x^{m_n}; k) - r(x^*; k)| &= |\sum_{j \in S^{m_n}} c_j((B^{m_n})^{-1} a_{\cdot k})_j - \sum_{j \in S^*} c_j((B^*)^{-1} a_{\cdot k})_j| \\
&= |\sum_{j \in S^{m_n} \cap S^*} c_j(((B^{m_n})^{-1} - (B^*)^{-1}) a_{\cdot k})_j + \sum_{j \in S^{m_n} \setminus S^*} c_j((B^{m_n})^{-1} a_{\cdot k})_j \\
&\qquad - \sum_{j \in S^* \setminus S^{m_n}} c_j((B^*)^{-1} a_{\cdot k})_j| \\
&\leq \sum_{j \in S^{m_n} \cap S^*} |c_j(((B^{m_n})^{-1} - (B^*)^{-1}) a_{\cdot k})_j| + \sum_{j \in S^{m_n} \setminus S^*} |c_j((B^{m_n})^{-1} a_{\cdot k})_j| \\
&\qquad + \sum_{j \in S^* \setminus S^{m_n}} |c_j((B^*)^{-1} a_{\cdot k})_j.|
\end{aligned}
$$

The first time on the right-hand side converges to zero since $(\tilde{B}^{m_n})^{-1}$ converges to $(\tilde{B}^*)^{-1}$ in the operator norm. Moreover, the sets $S^{m_n} \setminus S^*$ and $S^* \setminus S^{m_n}$ vanish in

[4] Note that Theorem IV.1.5 is stated for settings where $B : X \to X$ is a linear operator for some given Banach space X. However, the paragraph following the proof of the theorem (see page 193 of [37]) shows that it applies to linear operators $B : X \to Y$, where X and Y are (potentially different) Banach spaces under conditions satisfied in our setting. Here we take $X = \ell^2$.

[5] We make these changes in notation in order for the displayed equation below to be less crowded.

the limit (by Lemma 8.1) and so the second two sums also converge to 0. These observations involve an exchange of an infinite sum with a limit (as $n \to \infty$). This exchange is legitimate under the dominated convergence theorem since for any subset S of $\{1, 2, \dots\}$, $\sum_{j \in S} |c_j((B^{m_n})^{-1} a_{\cdot k})_j| \leq \sum_{j=1}^{\infty} |c_j x_j^{m_n}| < \infty$ since x^{m_n} is a basic feasible solution and all feasible solutions have finite cost (and also when replacing B^{m_n} and x^{m_n} with B^* and x^*, respectively).

It remains to argue that $r(x^*; k) \geq 0$ for all $k \notin S^*$. Suppose otherwise, that $r(x^*; k) = -\epsilon < 0$ for some $k \notin S^*$ and $\epsilon > 0$. Since $r(x^{m_n}; k) \to r(x^*; k)$ this implies that for sufficiently large n, $r(x^{m_n}; k) = -\epsilon < 0$. This contradicts Lemma 7.3. Hence, we can conclude that the reduced costs of all non-basic variables at x^* are nonnegative. Hence, by Theorem 6.2, x^* is an optimal solution.

By construction, the iterates of the simplex method have nondecreasing objective value. Thus, since we have just argued that x^* is optimal, we know $f^{m_n} \to f^*$ and since objective values are nondecreasing, the implies $f^m \to f^*$. ∎

A brief comment on how the various assumptions are used in our main Theorem 8.3. Assumptions (A1)–(A4) are invoked in the call to Theorem 6.2, the call to Lemma 7.3 additionally uses (A6) and (A7), and finally the call to Lemma 8.2 additionally uses (A8).

Although Theorem 8.3 does not furnish that the optimal solution convergence desired in (P4), the next result shows that the iterates of the simplex method become "arbitrarily close" to the set of optimal solutions. The Hilbert topology has an associated metric d where $d(x, y) = ||x - y||_H$. The distance from a point y to a set S is denoted $d(y, S) := \inf \{d(y, s) : s \in S\}$. We say a sequence y^n gets *arbitrarily close* to S if $d(y^n, S) \to 0$ as $n \to \infty$.

THEOREM 8.4. *The sequence of simplex iterates gets arbitrarily close to the set of optimal solutions to* (P)*. In particular, if there is a unique optimal solution then the full sequence of iterates converges to an optimal solution.*

Proof. Let F^* denote the set of optimal solutions of (P). Suppose there exists a subsequence x^{m_n} of simplex iterates and an $\epsilon > 0$ such that $d(x^{m_n}, F^*) > \epsilon$ for all n sufficiently large. By the compactness argument in the proof of the previous theorem, there exists a convergent sub-subsequence of x^{m_n} that converges to an optimal feasible solution $x^* \in F^*$. However, this contradicts the supposition that $d(x^{m_n}, F^*) > \epsilon$ for all n sufficiently large. ∎

9. Examples. In this section, we look at a class of CILPs that satisfy (A1)–(A8) and thus, by Theorem 8.3, our simplex method converges to optimal value. A goal of this paper was to extract analytical insight from this example to build the topological structure of "tractable" countably infinite linear programs. This was achieved in the previous sections. In this section, we will reflect this theory back on this special case to ground our contributions.

The following set up of minimum cost flow problems on *pure supply networks* is due to [32]. We show that these flow problems satisfy (A1)–(A8), under the observation that (A6)–(A8) can actually be weakened. Instead of applying to *all* basic feasible solutions (and extreme points), it suffices for (A6)–(A8) for all basic feasible solutions *encountered in a run of the simplex method.*

Let $G = (\mathcal{N}, \mathcal{A})$ be a directed graph with countably many nodes $\mathcal{N} = \{1, 2, \dots\}$ and arcs $\mathcal{A} \subseteq \mathcal{N} \times \mathcal{N}$. Each arc (i, j) has cost c_{ij}, and each node has supply b_i (with $b_i < 0$ corresponding to a demand). The goal of the *countably infinite network flow*

(CINF) problem is to solve:

$$\inf_x \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} \tag{9.1a}$$

$$\text{s.t.} \sum_{j:(i,j)\in\mathcal{A}} x_{ij} - \sum_{j:(j,i)\in\mathcal{A}} x_{ji} = b_i \text{ for } i \in \mathcal{N} \tag{9.1b}$$

$$x_{ij} \geq 0 \text{ for } (i,j) \in \mathcal{A}. \tag{9.1c}$$

A graph is *locally finite* if every node has finite in- and out-degree. Two nodes i and j are *finitely connected* in G if there exists a finite path P_{ij} between i and j. The graph G is *finitely connected* if all pairs of nodes in G are finitely connected. A *path to infinity* is a sequence of distinct nodes $i_1, i_2, \ldots$ where $(i_k, i_{k+1}) \in \mathcal{A}$ or $(i_{k+1}, i_k) \in \mathcal{A}$ for $k = 1, 2, \ldots$. An *infinite cycle* consists of two paths to infinity from some node i, $(i, i_1, i_2 \ldots)$ and $(i, j_1, j_2, \ldots)$, where all intermediate nodes i_k and j_ℓ are distinct. A *spanning tree* is a subgraph of G that contains no finite or infinite cycles and is incident to all nodes. A basic feasible flow in G is a feasible solution of (9.1) such that the subgraph induced by the arcs with positive flow is contained in a spanning tree of the graph. When the set of arcs of a flow x with positive flow themselves form a spanning tree, we call x a nondegenerate basic feasible flow. Of particular importance to the analysis in [32] is the following special class of spanning trees. A *spanning in-tree S rooted at infinity* is a spanning tree where for each node $i \in N$ there is a unique path from i to infinity in S that contains only forward arcs directed to "infinity". [32] also make the following additional assumptions:

- (NF1) G is locally finite,
- (NF2) G is finitely connected,
- (NF3) G contains no finite or infinite directed cycles,
- (NF4) b_i is integer for all $i \in \mathcal{N}$,
- (NF5) $b \in \ell_\infty(\mathcal{N})$, i.e., there exists a uniform upper bound $\bar{b}$ on absolute values of all node supplies.
- (NF6) G has finitely many nodes with in-degree 0,
- (NF7) $b_i \geq 0$ for all $i \in \mathcal{N}$ (all nodes are either transshipment nodes or supply nodes).

Assumptions (NF6) and (NF7) ensure that graph G permits *stages*, defined as follows. Stage 0 is the finite set of all nodes with in-degree 0. Stage 1 consists of all nodes with in-degree 0 in the modified graph that results from removing all stage 0 nodes and their adjacent arcs. Thus, all stage 1 nodes are adjacent to stage 0 nodes in the graph. We construct the subsequent stages by repeating this procedure.

In [32], the following additional assumption is made on the structure of stages:

- (NF8) There exist $\beta \in (0,1)$ and $\gamma \in (0,+\infty)$ such that for every $(i,j) \in \mathcal{A}$, $|c_{ij}| \leq \gamma\beta^{s(i)}$, where β can be interpreted as a discount factor (discounted arc costs) and $s(i)$ is the stage of node i,
- (NF9) There exists a sub-exponential function $g(k)$ where $|S_k| \leq g(k)$ for all k.

We refer to problems satisfying (NF1)–(NF9) as *pure supply problems.* Clearly, (9.1) is in the form (P), so it remains to check that (A1)–(A8) hold when (NF1)–(NF9) are taken.

Before checking these, it will be convenient to reformulate (9.1) by augmenting supply on certain nodes (for reasons that will become apparent once we check (A6)). Let $N' = (\mathcal{N}, \mathcal{A}, b', c)$ denote the network with the same graph and arc costs, but with supply $b'_i = b_i$ if $b_i > 0$ and $b'_i = 1$ if $b_i = 0$. Observe that if N is a pure supply network, then so is N'.

The key property of network N' is given in Lemma 4.8 of [32], which we recall as follows. Let T denote a spanning tree in N. Any arc (i,j) not in T has a reduced cost that corresponds to the cost of the cycle that it is formed in T when arc (i,j) is added to T (where the costs of arcs are weighted with 1 or -1 according to whether they are in the same direction as (i,j) in the cycle or not; for a formal definition see the discussion preceding Lemma 3.3 in [32]). The key property of Lemma 4.8 is that the reduced cost of arc (i,j) with respect to spanning tree T in the augmented network N' is the same as the reduced cost of arc (i,j) with respect to T in the original network N. Moreover, flows in N' can easily be converted to flows in N. Indeed, an optimal solution for the augmented problem yields an optimal solution for the original problem if we remove all flows originating from augmented supplies. Hence, it suffices to run a simplex algorithm on N' to recover a simplex method on N. It only remains to verify (A1)–(A8) hold for N'.

Not every instance of (9.1) is feasible, but we will only discuss feasible instances and so we may assume that (A1) holds. If an instance of (9.1) is feasible, then taking a single outgoing arc from every node forms an initial spanning tree T_0 and corresponds to a basic feasible flow (Lemma 4.4 in [32]). Lemma 4.2 in [32] shows that trees constructed in this way are always spanning in-trees rooted at infinity.

Although there are no explicit bounding constraints in (9.1), Lemma 2.6 in [32] shows that there is an implied bound on the flow on every arc. This is implicit from the uniform boundedness of supplies (NF5) and finiteness of the stages. Condition (A4) is a direct implication of (NF8) when δ is taken sufficiently large. The argument here is similar in spirit to the proof of Lemma 2.4, details are omitted. For (A3), we can rescale the constraints (9.1b) to satisfy the necessary conditions. The finite support of both rows and columns of the constraint matrix makes such a rescaling possible. This finiteness of rows and columns in a consequence of the fact that graph G is finitely-connected (NF2). Condition (A4) follows easily from (NF8) and (NF9).

Establishing (A5) requires more effort. In fact, we will show that every basis defines an onto map into Y, thus establishing the result for A since we have $\mathrm{cspan}(A) = \mathrm{cspan}(B)$ for every basis B. In [32], a basis B corresponds to the arcs of a spanning in-tree rooted at infinity. It suffices to argue that $B : H_B \to Y$ is an onto map for $\beta > \delta$, where H_B is defined before Lemma 4.3. We already know that $B : H_B \to Y$ by Lemma 3.3. Let $y \in Y$ and we will show that there exists an $x \in H_B$ such that $Bx = y$. We have $||y||_Y^2 = \sum_{i=1}^{\infty} \beta^{2i}|y_i|^2 < \infty$ since $y \in Y$. Let $\tilde{y}_i = \max\{1, |y_i|\}$ for $i = 1, 2, \ldots$ and note that $\sum_{i=1}^{\infty} \beta^{2i}|\tilde{y}_i|^2 < \infty$. Let the nodes in the tree $T(B)$ be numbered so that arc $(i,j) \in T(B)$ only if $i < j$. We have that there is a unique directed path to infinity out of each node i in $T(B)$. Let $P(i)$ be the finite set of all nodes k such that the unique path to infinity out of node k passes through node i. This set is finite by Lemma 4.1 in [32]. The flow constraints $Bx = b$ then gives $x_{ij} = \sum_{k \in P(i)} y_k$ where (i,j) is the unique arc leaving node i in $T(B)$ (the uniqueness of this arc is also guaranteed by Lemma 4.1 in [32]). It remains to show that $||x||_H < \infty$ for such an x. We have $|x_{ij}| \le \sum_{k \in P(i)} |y_k| \le \sum_{k=1}^{i} |y_k| \le \sum_{k=1}^{i} |\tilde{y}_k|$ so that $|x_{ij}|^2 \le (\sum_{k=1}^{i} |\tilde{y}_k|)^2$ since $\sum_{k=1}^{i} |\tilde{y}_k| \ge 1$. Hence,

$$||x||_H^2 = \sum_{(i,j) \in T(B)} \delta^{2i}|x_{ij}|^2 \le \sum_{i=1}^{\infty} \delta^{2i}(\sum_{k=1}^{i} |\tilde{y}_k|)^2 \tag{9.2}$$

since $x_{ij} = 0$ for $(i,j) \notin T(B)$. It thus remains to argue that $\sum_{i=1}^{\infty} \delta^{2i}(\sum_{k=1}^{i} |\tilde{y}_k|)^2 < \infty$, which will complete the proof. First, observe that there exists an I and a $\bar{y} > 1$ such that $|\tilde{y}_i| < \bar{y}/\beta^i$ for all $i \ge I$. Indeed, suppose otherwise that $|\tilde{y}_i| \ge \bar{y}/\beta^i$ for some subsequence $i = i_1, i_2, \ldots,$ in which case

$$\sum_{i=1}^{\infty} \beta^{2i} |\tilde{y}_i|^2 \geq \sum_{k=1}^{\infty} |\tilde{y}_{i_k}|^2 \geq \sum_{k=1}^{\infty} \beta^{2i_k} (\bar{y}/\beta^{i_k})^2 = \sum_{k=1}^{\infty} \bar{y} = \infty,$$

which contradicts the fact that $y \in Y$ and thus $\sum_{i=1}^{\infty} \beta^{2i} |\tilde{y}_i|^2 < \infty$. Thus, we may develop the second sum in the right-hand side of (9.2) as $\sum_{k=1}^{i} |\tilde{y}_k| \leq \sum_{k=1}^{i} (\bar{y}(I) + \bar{y}/\beta^i)$ where $\bar{y}(I) = \max_{k \leq I} |\tilde{y}_k|$. Hence, $\sum_{k=1}^{i} |\tilde{y}_k| \leq i\bar{y}(I) + i\bar{y}/\beta_i$. Thus, returning to (9.2), we have:

$$||x||_H^2 \leq \sum_{i=1}^{\infty} \delta^{2i} (\sum_{k=1}^{i} |\tilde{y}_k|)^2 \leq \sum_{i=1}^{\infty} \delta^{2i} (i\bar{y}(I) + i\bar{y}/\beta^i)^2$$

$$= \tilde{y}(I) \sum_{i=1}^{\infty} \delta^{2i} i^2 + 2\bar{y}(I)\bar{y} \sum_{i=1}^{\infty} (\delta^2/\beta)^i i^2 + \bar{y}^2 \sum_{i=1}^{\infty} (\delta/\beta)^{2i} i^2 < \infty$$

whenever $0 < \delta < \beta < 1$. Hence, $x \in H_B$ and we conclude that A is an onto map, establishing (A5).

In general, problem (9.1) need not be nondegenerate and so (A6) may not hold. However, under the transformation to N', all basic feasible solutions are nondegenerate. It is easy to see that every spanning tree in N' is a spanning in-tree rooted at infinity. Moreover, in the augmented N', a spanning in-tree rooted at infinity S corresponds to a nondegenerate basic feasible flow x^S, since every node has positive supply and a single outgoing arc. Accordingly, every arc carries positive flow and thus x^S is nondegenerate. In other words, there is a way to pivot from a nondegenerate basic feasible flow to a nondegenerate basic feasible flow for every choice of entering variable back in the original problem using the augmented network N'. Undertaking only such pivots in the simplex method defined in Section 7, we see that only nondegenerate basic feasible flows can be encountered by the simplex method.

Condition (A7) on the supports of extreme points follows from Theorem 3.2 in [32]. That result shows that every basic feasible flow is integer valued when the data is integer and, consequently, $\sigma \geq 1$.

When we showed (A6) above, we remarked on how the simplex method can be made to pivot from spanning in-trees rooted at infinity to spanning in-trees rooted at infinity. Corollary 4.15 in [32] shows that any convergent subsequence of such a sequence of iterate trees converges to yet another spanning in-tree rooted as infinity in the product discrete topology. The verifies (A8) and completes our verification the pure supply CINFs fit the setting of current paper and can be solved via the simplex method proposed in Section 7.

10. Conclusion. In this conclusion, we will provide a high-level summary of some of the insights our framework provides – particularly, in its novel topological underpinning – for solving CILPs via a simplex method. First, (A6) is critical. This assumption guarantees that we are able to "move", at least a little bit, at every pivot. The SPS assumption (A7) means that there is a lower bound on this "little bit" that is moved. Taken together, these properties guarantee that progress towards optimality is achieved as the simplex method runs.

However, "positive progress" towards optimality does not guarantee convergence. A key ingredient is (A8). The SPS condition (A7) guarantees that extreme points have an algebraic characterization as basic feasible solutions, which gives rise to the mechanics of tracking how the simplex method iterates from bfs to bfs through exploring successive bases. The closure of the set of bases implies a convergence of a subsequence of these bfs iterates, and hence in their objective values. The property that reduced costs converge to zero (Lemma 7.1), along with the optimality condition in Theorem 6.2, ensure convergence to optimality (Lemma 7.2).

In future work, it would be interesting to find settings where some of our assumptions fail, and yet a simplex method can be constructed that converges in value to optimality. Of course, this paper has only examined general conditions to ensure properties (P1) and (P2) discussed in the introduction. Exploration of what general conditions ensure (P3) and (P4) is a promising future direction. Some of the examples in the previous section have these properties, giving the interested reader a foothold on that journey.

REFERENCES

[1] D. Adelman, *Dynamic bid prices in revenue management*, Oper. Res., 55 (2007), pp. 647–661.

[2] D. Adelman and D. Klabjan, *Duality and existence of optimal policies in generalized joint replenishment*, Math. of Oper. Res., 30 (2005), pp. 28–50.

[3] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, 3rd ed., 2006.

[4] E. J. Anderson and P. Nash, *Linear Programming in Infinite-Dimensional Spaces: Theory and Applications*, Wiley, 1987.

[5] J. C. Bean, J. R. Lohmann, and R. L. Smith, *A dynamic infinite horizon replacement economy decision model*, The Engineering Economist, 30 (1985), pp. 99–120.

[6] D. Bertsimas and V. Misic, *Decomposable markov decision processes: A fluid optimization approach*, Operations Research, 64 (2016), pp. 1537–1555.

[7] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, Athena, 1997.

[8] O. Besbes, Y. Gur, and A. Zeevi, *Nonstationary stochastic optimization*, Operations Research, 63 (2015), pp. 1227–1244.

[9] R. S. Burachik and S. N. Majeed, *Strong duality for generalized monotropic programming in infinite dimensions*, Journal of Mathematical Analysis and Applications, 400 (2013), pp. 541–557.

[10] J. B. Conway, *A course in functional analysis*, vol. 96, Springer, 2019.

[11] W. D. Cook, C. A. Field, and M. J. L. Kirby, *Infinite linear programs in games with partial information*, Operations Research, 23 (1975), pp. 996–1010.

[12] R. G. Cooke, *Infinite matrices and sequence spaces*, Courier Corporation, 2014.

[13] E. D. Demaine, S. P. Fekete, and S. Gal, *Online searching with turn cost*, Theoretical Computer Science, 361 (2006), pp. 342–355.

[14] A. Ghate, *Circumventing the Slater conundrum in countably infinite linear programs*, European Journal of Oper. Res., 246 (2015), pp. 708–720.

[15] A. Ghate, *Duality in countably infinite monotropic programs*, SIAM J. on Opt., 27 (2017), pp. 2010–2033.

[16] A. Ghate, D. Sharma, and R. L. Smith, *A shadow simplex method for infinite linear programs*, Oper. Res., 58 (2010), pp. 865–877.

[17] A. Ghate and R. L. Smith, *Characterizing extreme points as basic feasible solutions in infinite linear programs*, Oper. Res. Lett., 37 (2009), pp. 7–10.

[18] A. Ghate and R. L. Smith, *Optimal backlogging over an infinite horizon under time-varying convex production and inventory costs*, Manufacutring and Service Operations Management, 11 (2009), pp. 362–368.

[19] A. Ghate and R. L. Smith, *A linear programming approach to non stationary infinite-horizon Markov decision processes*, Oper. Res., 61 (2013), pp. 413–425.

[20] R. C. Grinold, *Infinite horizon stochastic programs*, SIAM Journal on Control and Optimization, 24 (1986), pp. 1246–1260.

[21] T. D. Hansen and U. Zwick, *An improved version of the random-facet pivoting rule for the simplex algorithm*, in ACM-STOC, ACM, 2015, pp. 209–218.

[22] W. Hu, M. S. Lavieri, A. Toriello, and X. Liu, *Strategic health workforce planning*, IIE Trans., 48 (2016), pp. 1127–1138.

[23] D. Klabjan and D. Adelman, *An infinite-dimensional linear programming algorithm for deterministic semi-Markov decision processes on Borel spaces*, Math. of Oper. Res., 32 (2007), pp. 528–550.

[24] J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona, *Approximations of countably-infinite*

linear programs over bounded measure spaces. https://arxiv.org/abs/1810.03658.

[25] K. Kuratowski, *Introduction to Set Theory and Topology*, Elsevier, 1972.

[26] I. Lee, M. A. Epelman, H. E. Romeijn, and R. L. Smith, *Simplex algorithm for countable-state discounted Markov decision processes*, Oper. Res., 65 (2017), pp. 1029–1042.

[27] D. T. Luc and M. Volle, *Duality for optimization problems with infinite sums*, SIAM Journal on Optimization, 29 (2019), pp. 1819–1843.

[28] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, 1997.

[29] P. Mohajerin Esfahani, T. Sutter, D. Kuhn, and J. Lygeros, *From infinite to finite programs: explicit error bounds with applications to approximate dynamic programming*, SIAM J. on Opt., 28 (2018), pp. 1968–1998.

[30] S. Nourollahi and A. Ghate, *Duality in convex minimum cost flow problems on infinite networks and hypernetworks*, Networks, 70 (2017), pp. 98–115.

[31] H. E. Romeijn, D. Sharma, and R. L. Smith, *Extreme point characterizations for infinite network flow problems*, Networks, 48 (2006), pp. 209–22.

[32] C. T. Ryan, R. L. Smith, and M. A. Epelman, *A simplex method for uncapacitated pure-supply infinite network flow problems*, SIAM J. on Opt., 28 (2018), pp. 2022–2048.

[33] N. Saldi, S. Yüksel, and T. Linder, *On the asymptotic optimality of finite approximations to Markov decision processes with Borel spaces*, Math. of Oper. Res., 42 (2017), pp. 945–978.

[34] I. E. Schochetman and R. L. Smith, *Convergence of selections with applications in optimization*, Journal of Mathematical Analysis and Applications, 155 (1991), pp. 278–292.

[35] I. E. Schochetman and R. L. Smith, *Finite dimensional approximation in infinite dimensional mathematical programming*, Math. Prog., 54 (1992), pp. 307–333.

[36] T. C. Sharkey and H. E. Romeijn, *A simplex algorithm for minimum-cost network-flow problems in infinite networks*, Networks, 52 (2008), pp. 14–31.

[37] A. E. Taylor and D. C. Lay, *Introduction to Functional Analysis*, Wiley New York, 1958.

Appendix A. Proofs of Lemmas 3.3 and 4.3.

The first step is to establish an isometric isomorphism between H and ℓ^2, the space of square-summable sequences. Consider the transformation T_δ from H into $\mathbb{R}^{\mathbb{N}}$ defined by $T_\delta(x) = (\delta^j x_j)$. Let $x(\delta)$ denote the image of x under T_δ for notational convenience.

Claim A.1. The spaces H and ℓ^2 are isometrically isomorphic under mapping T_δ.

Proof. First, we claim that T_δ is an isometry. Indeed, $||x||_H = \sqrt{\sum_{j=1}^{\infty} \delta^{2j}|x_j|^2} = \sqrt{\sum_{j=1}^{\infty} |\delta^j x_j|^2} = ||T_\delta(x)||_{\ell^2}$. Next, observe that $T_\delta : H \to \ell^2$. Indeed, for $x \in H$ note that $||x(\delta)||_2^2 = ||x||_H^2 < \infty$ and so $x(\delta) \in \ell^2$. Second, we claim that $T_\delta : H \to \ell^2$ is onto. Let $y \in \ell^2$ and set $x_j = (y_j/\delta^j)$ for $j = 1, 2, \ldots$. Observe that $T_\delta(x) = (\delta^j(y_j/\delta^j)) = (y_j) = y$. Thus, it suffices to argue that $x \in H$. This follows since $||x||_H = \sum_{j=1}^{\infty} \delta^{2j}|x_j|^2 = \sum_{j=1}^{\infty} \delta^{2j}|y_j/\delta^j|^2 = \sum_{j=1}^{\infty} \delta^{2j}|y_j|^2/\delta^{2j} = \sum_{j=1}^{\infty} |y_j| < \infty$, since $y \in \ell^2$. Third, we claim that $T_\delta : H \to \ell^2$ is one-to-one. Indeed, if $x \neq x'$ in H then since T_δ is a linear map, $||T_\delta(x) - T_\delta(x')||_{\ell^2} = ||x - x'||_H \neq 0$. Hence, $T_\delta(x) \neq T_\delta(x')$ and T_δ is one-to-one. ∎

Consider now the transformation $T_{\beta,A} : \text{cspan}(A) \to \ell^2$ where $\text{cspan}(A)$ is the column span of the infinite matrix A over H and $T_{\beta,A}(y) = (\beta^i y_i)$. By an identical argument as above, $T_{\beta,A}$ is an isometric isomorphism between $\text{cspan}(A)$ and ℓ^2. Using T_δ and $T_{\beta,A}$ we construct a "pullback" linear operator $A' := T_{\beta,A} A T_\delta^{-1}$ from ℓ^2 to ℓ^2 from the operator from H to Y defined by A.

Claim A.2. The linear operator A is continuous if and only if A' is continuous.

Proof. It is straightforward to see that T_δ^{-1} and $T_{\beta,A}$ are bounded linear operators with an operator norm equal 1 since both are isometries and so (for instance)

$$||T_{\beta,A}|| = \sup_{y \in \text{cspan}(A)} \frac{||T_{\beta,A}(y)||_{\ell^2}}{||y||_Y} = \sup_{y \in \text{cspan}(A)} \frac{||y||_Y}{||y||_Y} = 1 < \infty$$

Now, since $A' = T_{\beta,A} A T_\delta^{-1}$ we have $||A'|| \le ||T_{\beta,A}|| ||A|| ||T_\delta^{-1}|| = ||A||$ so A' is a bounded linear operator whenever A is. Multiplying the equation defining A' the above equation on the left by $T_{\beta,A}^{-1}$ and on the right by T_δ we get $A = T_{\beta,A}^{-1} A' T_\delta$ and so A is bounded whenever A' is. In fact, $||A|| = ||A'||$. ∎

Thus, we have reduced showing the continuity of A to establishing the continuity of A'. Since A' is a linear operator from ℓ^2 to ℓ^2, we can leverage from the following lemma.

Lemma A.3 (Schur test, page 260 in [12]). *If a doubly infinite matrix $M = (m_{ij})$ satisfies (i) $\sum_{j=1}^\infty |m_{ij}| \le B_1$ for every i, and (ii) $\sum_{i=1}^\infty |m_{ij}| \le B_2$ for every j, then the operator M is bounded and $||M|| \le \sqrt{B_1 B_2}$.*

We now apply the Schur test to A'. It a straightforward exercise to show that $A' = (m_{ij})$ has $m_{ij} = \beta^i/\delta^j a_{ij}$. To check (i) in the Schur test holds, observe that

$$\sum_{j=1}^\infty \tfrac{\beta^i}{\delta^j}|a_{ij}| = \beta^i \sum_{j=1}^\infty \tfrac{1}{\delta^j}|a_{ij}| \le \beta^i \sum_{j=1}^\infty \tfrac{1}{\delta^j}\bar{a}\alpha^j = \beta^i \bar{a} \sum_{j=1}^\infty \left(\tfrac{\alpha}{\delta}\right)^j \le \bar{a}\tfrac{\alpha/\delta}{1-\alpha/\delta} = B_1,$$

where the first inequality holds by (A3) and the fact $0 < \beta < 1$ and $0 < \alpha < \delta < 1$. Similarly,

$$\sum_{i=1}^\infty \tfrac{\beta^i}{\delta^j}|a_{ij}| = \tfrac{1}{\delta^j}\sum_{i=1}^\infty \beta^i |a_{ij}| \le \tfrac{1}{\delta^j}\sum_{i=1}^\infty \beta^i \bar{a}\alpha^j = \tfrac{1}{\delta^j}\bar{a}\sum_{i=1}^\infty (\alpha\beta)^i \le \bar{a}\tfrac{\alpha\beta}{1-\alpha\beta} = B_2.$$

Proof of Lemma 3.4. Under the assumptions, A' is a continuous map from ℓ^2 to ℓ^2 by the Schur Test (Lemma A.3). Then by Claim A.2, we have A is a continuous mapping from H to Y. This completes the proof. ∎

Proof of Lemma 4.3. It remains to prove the B is a continuous operator. Recall that the basis B defines an operator $B : H_B \to Y$. Under the assumptions, B is a bounded linear operator. Indeed, $||B|| = \sup_{x\in H_B} \frac{||Bx||_Y}{||x||_H} = \sup_{x\in H_B} \frac{||Ax||_Y}{||x_H||} \le \sup_{x\in H} \frac{||Ax||_Y}{||x||_H} = ||A|| < \infty$, where the second equality follows since $B(x) = A(x)$ for $x \in H_B$ and the last (strict) inequality follows from Lemma 3.4. ∎

Appendix B. Proof of Proposition 5.5.

Lemma B.1. *Let E be an extreme subset of S, a non-empty subset of $\mathbb{R}^{\mathbb{N}}$. Given another non-empty subset T of $\mathbb{R}^{\mathbb{N}}$: (i) if $E \subseteq T \subseteq S$ then E is an extreme subset of T and (ii) $E \cap T$ is an extreme subset of $S \cap T$.*

Definition B.2. *Let x be a nondegenerate bfs. The* ***cone of feasible directions*** *(from x) is $\mathcal{C}(x) \triangleq \{z \in H : x + \lambda z \in \mathcal{F} \text{ for some } \lambda > 0\}$. Define also the translation $\bar{\mathcal{C}}(x)$ of $\mathcal{C}(x)$ by x. That is, $\bar{\mathcal{C}}(x) \triangleq x + \mathcal{C}(x) = \{y \in H : y = x + z, z \in \mathcal{C}(x)\}$.*

Observe that $\mathcal{F}$ itself is a subset of $\bar{\mathcal{C}}(x)$ since $y - x \in \mathcal{C}(x)$ for every $y \in \mathcal{F}$ (simply take $\lambda = 1$). In light of Lemma B.1(ii), we may focus attention on understanding extreme subsets E of $\bar{\mathcal{C}}(x)$ (which turns out to be an easier task) since $E \cap \mathcal{F}$ is an extreme subset of $\mathcal{F} = \bar{\mathcal{C}}(x) \cap \mathcal{F}$.

Following the above logic, we will examine an extreme subset of the translated cone $\mathcal{C}(x)$. First, consider the set $\mathcal{E}(x;k) \triangleq \{\xi \in H : \xi = \mu d(x;k),\ \mu \ge 0\}$. We show this is an extreme subset (in fact, an edge) of the cone of feasible directions.

Claim B.3. $\mathcal{E}(x;k)$ is $\mathcal{C}(x)$-extreme.

Proof of Claim B.3: First notice that $\mathcal{E}(x;k) \subseteq \mathcal{C}(x)$. To see this, consider a $\xi = \mu d(x;k)$ for some $\mu > 0$ (we omit the trivial case of $\mu = 0$). Thus, $\xi \in \mathcal{E}(x;k)$. In order to show that $\mathcal{E}(x;k) \subseteq \mathcal{C}(x)$, we must show that $\xi \in \mathcal{C}(x)$, that is, that there exists a $\lambda > 0$ such that $x + \lambda\mu d(x;k) \in \mathcal{F}$. Note that setting $\lambda = \lambda(x;k)/\mu$ works. Now to prove our claim, let $\eta, \chi \in \mathcal{C}(x)$ and $0 < t < 1$ be such that $t\eta + (1-t)\chi \in \mathcal{E}(x;k)$. We need to prove that $\eta, \chi \in \mathcal{E}(x;k)$. Since $\eta, \chi \in \mathcal{C}(x)$, there exists $\lambda_\eta > 0$ and $\lambda_\chi > 0$ such that $x + \lambda_\eta \eta \in \mathcal{F}$ and $x + \lambda_\chi \chi \in \mathcal{F}$. That is, $x + \lambda_\eta \eta \geq 0$, $\sum_{j=1}^{\infty} a_{ij}\eta_j = 0$, $i = 1, 2, \ldots$ and $x + \lambda_\chi \xi \geq 0$, $\sum_{j=1}^{\infty} a_{ij}\xi_j = 0$, $i = 1, 2, \ldots$. Moreover, since $t\eta + (1-t)\chi \in \mathcal{E}(x;k)$, there exists a $\mu \geq 0$ such that $\mu d(x;k) = t\eta + (1-t)\chi$. To establish that $\eta, \chi \in \mathcal{E}(x;k)$, we need to construct $\mu_1 \geq 0$ and $\mu_2 \geq 0$ such that $\eta = \mu_1 d(x;k)$ and $\chi = \mu_2 d(x;k)$. To achieve this, we consider three types of components of η and χ. The first type is components $j \in \mathcal{S}^c(x)$ such that $j \neq k$. For these components, $x_j = 0$ and hence we know that $\eta_j \geq 0$, $\chi_j \geq 0$. In addition, $d_j(x;k) = 0$. Thus, $\mu d_j(x;k) = t\eta_j + (1-t)\chi_j$ implies that $\eta_j = 0$ and $\chi_j = 0$. Our second type of components in fact only includes component k. For this component, $d_k(x;k) = 1$. In addition, $x_k = 0$ implies that $\eta_k \geq 0$ and $\chi_k \geq 0$. As a result, $\mu = t\eta_k + (1-t)\chi_k$ implies $\chi_k = \frac{\mu - t\eta_k}{1-t}$.

The third type of components is $j \in \mathcal{S}(x)$. For these components, we have,

$$\sum_{j \in \mathcal{S}(x)} a_{ij}\eta_j = -\eta_k a_{ik},\ i = 1, 2, \ldots, \text{ and} \tag{B.1}$$

$$\sum_{j \in \mathcal{S}(x)} a_{ij}\chi_j = -\chi_k a_{ik} = -\tfrac{\mu - t\eta_k}{1-t} a_{ik},\ i = 1, 2, \ldots. \tag{B.2}$$

But since the basic direction $d(x;k)$ is unique, the system of equations (B.1) implies that $\eta_j = \eta_k d_j(x;k)$ for all $j \in \mathcal{S}(x)$. It is clear that this is a solution to (B.1). To see that this is the only solution, we proceed by contradiction. So, suppose there is an alternate solution ζ_j, for $j \in \mathcal{S}(x)$, to (B.1). This implies that $\sum_{j \in \mathcal{S}(x)} a_{ij}(\eta_j - \zeta_j) = 0$ for $i = 1, 2, \ldots$ with $\eta_j \neq \zeta_j$ for at least one $j \in \mathcal{S}(x)$. But this contradicts the fact that x is a basic solution. Similarly, the system of equations (B.2) implies that $\chi_j = \frac{\mu - t\eta_k}{1-t} d_j(x;k)$ for all $j \in \mathcal{S}(x)$. In summary, we have shown that, by choosing $\mu_1 = \eta_k$ and $\mu_2 = \frac{\mu - t\eta_k}{1-t}$, we ensure $\eta = \mu_1 d(x;k)$ and $\chi = \mu_2 d(x;k)$ as required. This completes our proof of Claim B.3. This result is a precursor to showing that the translated set $\bar{\mathcal{E}}(x;k) \triangleq \{z \in H : z = x + \xi,\ \xi \in \mathcal{E}(x;k)\}$ is an edge $\bar{\mathcal{C}}(x)$.

Claim B.4. $\bar{\mathcal{E}}(x;k)$ is $\bar{\mathcal{C}}(x)$-extreme.

Proof of Claim B.4: Consider any $z^1, z^2 \in \bar{\mathcal{C}}(x)$. That is, there are $\xi^1, \xi^2 \in \mathcal{C}(x)$ such that $z^1 = x + \xi^1$ and $z^2 = x + \xi^2$. Consider any $0 < t < 1$ such that $tz^1 + (1-t)z^2 \in \bar{\mathcal{E}}(x;k)$. That is, there is some $\xi^0 \in \mathcal{E}(x;k)$ such that $tz^1 + (1-t)z^2 = x + \xi^0$. We need to establish that $z^1, z^2 \in \bar{\mathcal{E}}(x;k)$. In other words, we need to establish that $\xi^1, \xi^2 \in \mathcal{E}(x;k)$. To see that this holds, note that $tz^1 + (1-t)z^2 = t(x + \xi^1) + (1-t)(x + \xi^2) = x + t\xi^1 + (1-t)\xi^2$. But since this must equal $x + \xi^0$, we have, $t\xi^1 + (1-t)\xi^2 = \xi^0$. Since $\mathcal{E}(x;k)$ is $\mathcal{C}(x)$-extreme, this implies that $\xi^1, \xi^2 \in \mathcal{E}(x;k)$ as required. This completes the proof of Claim B.4. Claim B.4 implies that $\bar{\mathcal{E}}(x;k) \cap \mathcal{F}$ is $(\bar{\mathcal{C}}(x) \cap \mathcal{F})$-extreme. Observe that the set $\mathcal{Z}(x;k) = \bar{\mathcal{E}}(x;k) \cap \mathcal{F}$ in view of Lemma 5.3. Thus, since $\mathcal{F} \subseteq \bar{\mathcal{C}}(x)$ (as was observed before the statement of the result) $\mathcal{Z}(x;k)$ is $\mathcal{F}$-extreme, using Lemma B.1(ii). It is straightforward to see that $x + \lambda(x;k)d(x;k)$ is an extreme point of the set $\mathcal{Z}(x;k)$. Thus, by Lemma B.1(i), $x + \lambda(x;k)d(x;k)$ is an extreme point of $\mathcal{F}$.