DUALITY IN COUNTABLY INFINITE MONOTROPIC PROGRAMS

ARCHIS GHATE

Abstract. Finite-dimensional monotropic programs form a class of convex optimization problems that includes linear programs, convex minimum cost flow problems on networks and hypernetworks, and separable convex programs with linear constraints. Countably infinite monotropic programs arise, for example, in infinite-horizon sequential decision problems and in robust optimization. Their applications encompass (i) countably infinite linear programs such as the shortest path formulations of infinite-horizon non-stationary as well as countable-state Markov decision processes; and (ii) convex minimum cost flow problems on countably infinite networks and hypernetworks. Duality results for finite-dimensional monotropic programs are as powerful as those available for linear programs. On the contrary, applicable duality results for countably infinite monotropic programs are currently non-existent owing to several mathematical pathologies in infinite dimensional sequence spaces. This paper overcomes this hurdle by first embedding the dual variables in a sequence space where the Lagrangian function is well-defined and finite. Weak duality and complementary slackness are derived using finite-dimensional proof techniques. Conditions under which zero duality gaps and strong duality between a sequence of finite-dimensional primal-dual projections of the infinite dimensional problem are preserved in the limit are established. Essentially all known duality results about countably infinite mathematical programs are recovered as special cases.

Key words. convex optimization, infinite dimensional optimization, duality

AMS subject classifications. 90C25, 90C46, 90C48, 90C35

1. Introduction. Finite-dimensional monotropic programs form a class of convex optimization problems wherein duality results are as powerful as those available for linear programs. Research on monotropic programs can be traced back to the seminal works of Minty [16] and Rockafellar [19, 20, 21, 22]. Indeed, in the preface to his classic book [22] on the subject, Rockafellar commented that duality in monotropic programs is “as important a tool in computation as it is in theory and interpretation.”

A finite-dimensional monotropic program is written as

\[ \min \sum_{j=1}^{n} c_j(x_j) \]
\[ x \in S, \]
\[ x_j \in X_j, \quad j = 1, 2, \ldots, n, \]

where \( n \) is a positive integer; \( x \) is a vector in \( \mathbb{R}^n \) with real-valued components \( x_1, x_2, \ldots, x_n \); \( X_j \) is a nonempty interval of \( \mathbb{R} \) for each \( j \); \( c_j(\cdot) \) is a convex function for each \( j \); and \( S \) is a subspace of \( \mathbb{R}^n \) (see Section 9.7 in [5]). This class of monotropic programs includes linear programs, minimum cost network flow problems, and separable convex programs with linear constraints. Rockafellar [21] was the first to prove a zero duality gap result for this class of monotropic programs. He used a variant of the \( \epsilon \)-descent method for this purpose. More recently, Bertsekas [6] generalized this result to the so-called extended monotropic programs where components \( x_j \) themselves could be finite-dimensional vectors.

There has been a surge of interest in deriving duality results for countably infinite mathematical programs over the last two decades. Three of these studies [9, 24, 25]...
focused on countably infinite linear programs (CILPs) — these are linear optimization problems with a countably infinite number of variables and a countably infinite number of constraints. One study focused on linear minimum cost flow problems on countably infinite networks [29], which are a special case of CILPs. The most recent ones considered convex minimum cost flow problems on countably infinite networks and hypernetworks [17]. These duality results have been applied mainly to infinite-horizon planning problems such as Markov decision processes (MDPs) [10, 11, 14] but also to robust optimization [9]. This paper extends these duality results to countably infinite monotropic programs.

There is only one existing strong duality result for countably infinite monotropic programs [7]. There, the primal variables were embedded in a countable product of locally convex spaces and the dual variables resided in the topological dual of this primal space. A strong duality result was then derived under a constraint qualification that called for a certain set to be closed (see their Theorem 3.5 and Corollary 3.1). Unfortunately, such closedness conditions, although theoretically elegant, have long been thought to be hard to apply in practice (see the discussions in [2, 24, 25] for the special case of CILPs and [15] for the more general case of convex programs). Moreover, in most applications in Operations Research and Economics (including in CILPs, in infinite dimensional convex network flow problems, and in convex infinite-horizon planning problems), this choice of variable spaces would amount to selecting the space of all real sequences for the primal variables and the space of all real sequences with finitely many nonzero entries for the dual variables (this dual variable space is called the generalized finite sequence space (GFSS) [2]). This choice is unsatisfactory. For instance, in countable state MDPs, where each dual variable corresponds to the value of an MDP state, choosing GFSS as the dual variable space would imply that only a finite number of states have nonzero values. Thus, an alternative approach is needed to accommodate most applications of interest to the Operations Research and Economics communities.

Since countably infinite monotropic problems are a special case of infinite dimensional convex programs, it is perhaps tempting to use Slater’s constraint qualification to derive a zero duality gap result (see Theorem 3.11.2 in Ponstein [18], for instance). This would require that the positive cone of the primal constraint space has a non-empty interior. For many problems of interest in Operations Research and Economics, one natural choice for the constraint space is again the space of all real sequences. Unfortunately, the interior of the positive cone of this space is empty in its natural product topology, rendering Slater’s constraint qualification inapplicable. A possible remedy for some problems is to instead use the space of all bounded sequences with its usual supremum norm topology because the positive cone of this constraint space does have a non-empty interior. Unfortunately, however, the corresponding dual problem is notoriously difficult to characterize owing to the existence of the so-called singular functionals. For instance, in CILPs, this dual problem cannot, in general, be written using the ordinary transpose of the primal constraint matrix and then the dual variables cannot be interpreted as shadow prices. These hurdles in establishing a zero duality gap result in infinite dimensional convex programs were recently formalized by Martin et al. [15] via what they called the Slater conundrum. Specifically, they made a remarkable observation by employing the notion of a core point. A core point is an algebraic counterpart of the topological concept of an interior point. Informally, they commented, “on the one hand, existence of a core point ensures a zero duality gap (a desirable property), but on the other hand, existence of a core point implies the existence of singular dual functionals (an undesirable property).” All in all, such
negative results have historically posed significant stumbling blocks in establishing concrete and applicable duality results for countably infinite mathematical programs using closedness or interior point-type sufficient conditions. As in [9, 17], these challenges are circumvented in this paper by applying a hybrid approach that (i) provides a way to choose a space for dual variables such that weak duality and complementary slackness hold, and (ii) switches to the planning horizon method to derive conditions under which zero duality gaps and strong duality in finite-dimensional approximations are preserved in the limit.

The paper is organized as follows. A countably infinite monotropic program is formulated in the next section. Hypothesis (H1) on the associated convex functions is then introduced. Under H1, the objective function in the monotropic program is well-defined and finite. Two other hypotheses (H2, H3) are then stated. These define a linear subspace (Lemma 2) where the dual variables reside. H1, H2, and H3 enable a proof that the Lagrangian function is well-defined and finite (Lemma 3). This problem setup from Section 2 facilitates the use of finite-dimensional proof techniques to establish weak duality (Proposition 4) and complementary slackness (Proposition 10) in Section 3. A planning horizon approach is then devised to establish a zero duality gap result (Proposition 19) in Section 4. To achieve this, a sequence of finite-dimensional projections of the infinite dimensional monotropic program is constructed. These finite-dimensional problems are monotropic programs themselves. Optimal values of these finite-dimensional problems are shown to converge to the optimal value of the infinite dimensional problem (Proposition 18). The proof is similar to that of Berge’s maximum principle and employs the concept of Kuratowski set convergence (Definition 14). The dual of the finite-dimensional monotropic program is then written, and it is shown that the infinite-dimensional dual optimal value is bounded below by the finite-dimensional dual optimal value when costs are nonnegative (Lemma 20). This, when combined with (i) a zero duality gap result in finite-dimensional monotropic programs (see Proposition 9.19 in [5]), (ii) value convergence, and finally (iii) weak duality, yields the no duality gap result. Dual value convergence is then also derived. Finally, strong duality is established in Section 5 (see Theorem 22) under the assumption that optimal dual variables in the finite-dimensional dual problems are uniformly bounded. This result does not require nonnegativity of costs. The proof derives a convergent subsequence from a sequence of pairs of optimal solutions to the finite-dimensional monotropic program and its dual, and shows that its accumulation point is feasible and complementary for the infinite dimensional monotropic program and its dual. This implies, by the complementary slackness result in Section 3, that this accumulation point pair is optimal to the infinite dimensional monotropic program and its dual, and that the two problems have identical optimal values. Although, from a high-level viewpoint, the approach and hence the sequence of results in Sections 3 - 5 in this paper are identical to the author’s previous work in [17], a more sophisticated analysis is needed to ensure that the methodology works here. For instance, the dual problems in the author’s previous work did not utilize the notion of an orthogonal subspace, but this is needed to properly characterize the dual problem here. Moreover, it is harder to define finite-dimensional problems, and the idea of Kuratowski set convergence is employed in the planning horizon approach in this paper. Sections 6 and 7 show that known duality results from the author’s recent work on CILPs [9] and on convex minimum cost flow problems on infinite networks and hypernetworks [17] can be recovered as special cases of the monotropic programming results here. In the finite-dimensional case, it is easy to show that linear programs and convex minimum cost flow problems, as well as the corresponding duality results, are special cases of
monotropic programs. As we shall see, showing this in the countably infinite context
requires considerable work. Moreover, a nontrivial argument is needed to characterize
the orthogonal subspaces that define the dual problems in each case. Sections 6 and
7 therefore include some of the main contributions of this paper although this might
not be evident at first glance. New duality results for countably infinite separable
convex programs with linear constraints are derived in Section 8 as a special case of
the monotropic programming results.

2. Problem formulation. Throughout this paper, the symbol $\mathbb{R}$ is employed
to define new notations. The symbol $\mathbb{N}$ will denote the set $\{1, 2, 3, \ldots\}$ of natural
numbers and $\#$ will be reserved for set cardinalities. The linear vector space of all
real-valued sequences $x \triangleq (x_1, x_2, x_3, \ldots)$ is denoted by $\mathbb{R}^\mathbb{N}$. The usual metrizable
product topology of pointwise convergence on $\mathbb{R}^\mathbb{N}$ is employed throughout.

Let $S$ be any subspace of $\mathbb{R}^\mathbb{N}$. For each $j \in \mathbb{N}$, let $-\infty < a_j \leq b_j < +\infty$ be any
two real numbers, and let $u_j \triangleq \max\{|a_j|, |b_j|\}$. Suppose that, for each $j \in \mathbb{N}$, $c_j(\cdot)$ is
a real-valued, convex, and continuous function over the closed and bounded interval
$X_j \triangleq [a_j, b_j]$. Now consider the infinite dimensional optimization problem

\[
(P) \quad V \triangleq \inf_{x \in S} C(x) \triangleq \sum_{j \in \mathbb{N}} c_j(x_j)
\]

Problem $(P)$ will be called a countably infinite monotropic program. In the sequel,
$X \triangleq \prod_{j \in \mathbb{N}} X_j$ will denote the Cartesian product of intervals $X_j$ across $j \in \mathbb{N}$. The
feasible region of $(P)$ is denoted by $F \triangleq S \cap X$.

Since $|c_j(\cdot)|$ is continuous over the compact interval $X_j$, it attains its maximum
on this interval, and let $c_j^* \triangleq \max_{x_j \in X_j} |c_j(x_j)|$. Assume that

**H1.** the series $\sum_{j \in \mathbb{N}} c_j^*$ of nonnegative terms is finite.

This hypothesis was inspired by a similar assumption that the author used in his
recent work on duality in CILPs [9] and in convex minimum cost flow problems in
countably infinite networks and hypernetworks [17]. Such hypotheses are standard in
the literature on infinite dimensional optimization [10, 27, 29] as they ensure that the
primal objective function is well-defined and finite.

**Lemma 1.** Suppose $S$ is closed. If $(P)$ has a feasible solution, then it has an
optimal solution.

**Proof.** The series of functions $\sum_{j \in \mathbb{N}} c_j(x_j)$ is easily seen to be uniformly convergent
over $F$ by applying the Weierstrass test [3] using H1. This implies that the objective
function $C(x)$ in $(P)$ is continuous over $F$ (see Theorem 9.7 in [3]). The product $X$
of compact intervals is compact by Tychonoff’s theorem (see Theorem 2.61 in [1]).
Hence the feasible region $F$ is an intersection of a closed set $S$ and a compact set
$X$; $F$ is thus compact (see Theorem 2.35 and its corollary in [26]). The result then
follows by Corollary 2.35 in [1], which states that a continuous function achieves its
minimum on a compact set.

The assumption that $S$ is closed is not vacuous because, unlike finite-dimensional
Euclidean spaces, not all subspaces of $\mathbb{R}^\mathbb{N}$ are closed in the product topology (consider
the subspace GFSS, for instance, with the sequence of vectors whose first $n$ components equal 1 and all tail components equal 0; this sequence converges to the vector of all ones, which is not in GFSS). We will show in Sections 6 and 7 that $S$ is closed in many applications of interest.

As in finite-dimensional monotropic programs, use a new variable $y \triangleq (y_1, y_2, \ldots)$ to first rewrite $(P)$ in the equivalent format

\[
(P) \quad V \triangleq \inf_{x \in X} C(x) \triangleq \sum_{j \in \mathbb{N}} c_j(x_j)
\]

\[
(8) \quad x_j = y_j, \ j \in \mathbb{N},
\]

\[
(9) \quad y \in S,
\]

\[
(10) \quad x_j \in X_j, \ j \in \mathbb{N}.
\]

In order to write the dual problem, a dual variable $\lambda \triangleq (\lambda_1, \lambda_2, \lambda_3, \ldots)$ is attached to the equality constraint (8) for each $j \in \mathbb{N}$. The resulting sequence $\lambda \triangleq (\lambda_1, \lambda_2, \lambda_3, \ldots)$ needs to be embedded into an appropriate subset $\Lambda$ of $\mathbb{R}^\mathbb{N}$. Specifically, let $\Lambda$ be the subset of all sequences $\lambda$ such that

H2. the series $\sum_{j \in \mathbb{N}} |\lambda_j|u_j$ of nonnegative terms is finite; and

H3. the series $\sum_{j \in \mathbb{N}} |\lambda_j y_j|$ of nonnegative terms is finite for each $y \in S$.

These two hypotheses were motivated by similar assumptions in the author’s work on CILPs [9] and on convex minimum cost flow problems in countably infinite networks and hypernetworks [17]. H2 and H3 help ensure that two series that appear in our Lagrangian function below converge.

**Lemma 2.** The subset $\Lambda$ is a subspace of $\mathbb{R}^\mathbb{N}$.

**Proof.** Straightforward, hence omitted. \[
\]

Now define the Lagrangian function as

\[
L(x, y; \lambda) \triangleq \sum_{j \in \mathbb{N}} c_j(x_j) + \sum_{j \in \mathbb{N}} \lambda_j(y_j - x_j), \ x \in X, \ y \in S, \ \lambda \in \Lambda.
\]

**Lemma 3.** The Lagrangian is well-defined and finite for every $x \in X$, $y \in S$, and $\lambda \in \Lambda$.

**Proof.** The first series $\sum_{j \in \mathbb{N}} c_j(x_j)$ converges (absolutely) over $X$ by H1. Also, the series $\sum_{j \in \mathbb{N}} \lambda_j x_j$ converges (absolutely) over $X$ by H2. Similarly, the series $\sum_{j \in \mathbb{N}} \lambda_j y_j$ converges (absolutely) by H3. Thus, the series $\sum_{j \in \mathbb{N}} (\lambda_j y_j - \lambda_j x_j)$ converges by Theorem 8.8 in [3]. This means that the second series $\sum_{j \in \mathbb{N}} \lambda_j(y_j - x_j)$ in the Lagrangian function also converges. This proves the claim. \[
\]

Now define

\[
\phi(\lambda) \triangleq \inf_{x \in X, \ y \in S} L(x, y; \lambda), \ \lambda \in \Lambda,
\]

and write the dual of $(P)$ as

\[
(D) \quad W = \sup_{\lambda \in \Lambda} \phi(\lambda).
\]
3. Weak duality and complementary slackness. Since \( \sum_{j \in \mathbb{N}} c_j(x_j), \sum_{j \in \mathbb{N}} \lambda_j x_j \) and \( \sum_{j \in \mathbb{N}} \lambda_j y_j \) converge by H1, H2, and H3, the Lagrangian function in (11) can be rewritten equivalently as

\[
L(x, y; \lambda) = \sum_{j \in \mathbb{N}} \left( c_j(x_j) - \lambda_j x_j \right) + \sum_{j \in \mathbb{N}} \lambda_j y_j, \quad x \in X, \ y \in S, \ \lambda \in \Lambda.
\]

Consequently, the function \( \phi(\cdot) \) from (12) can also be rewritten as

\[
\phi(\lambda) = \inf_{x \in X, \ y \in S} L(x, y; \lambda) = \inf_{x \in X, \ y \in S} \left[ \sum_{j \in \mathbb{N}} \left( c_j(x_j) - \lambda_j x_j \right) + \sum_{j \in \mathbb{N}} \lambda_j y_j \right]
\]

\[
= \inf_{y \in S} \sum_{j \in \mathbb{N}} \lambda_j y_j + \inf_{x \in X} \left[ \sum_{j \in \mathbb{N}} \left( c_j(x_j) - \lambda_j x_j \right) \right]
\]

\[
= \inf_{y \in S} \sum_{j \in \mathbb{N}} \lambda_j y_j + \sum_{j \in \mathbb{N}} \inf_{x \in X_j} \left[ \left( c_j(x_j) - \lambda_j x_j \right) \right].
\]

The function \( c_j(x_j) - \lambda_j x_j \) is continuous over \( X_j \) for each \( j \), and hence the second infimum above can be replaced by a minimum. Specifically, let

\[
\phi_j(\lambda_j) \triangleq \min_{x_j \in X_j} c_j(x_j) - \lambda_j x_j, \ j \in \mathbb{N}.
\]

Then, the function \( \phi(\cdot) \) can be compactly rewritten as

\[
\phi(\lambda) = \begin{cases} 
\sum_{j \in \mathbb{N}} \phi_j(\lambda_j) & \text{if } \lambda \in S^\perp, \text{ and} \\
-\infty & \text{otherwise,}
\end{cases}
\]

where \( S^\perp \triangleq \{ \lambda \in \Lambda | \sum_{j \in \mathbb{N}} \lambda_j y_j = 0, \ \forall y \in S \} \) is the orthogonal complement of \( S \). It is easy to see that \( S^\perp \) is a subspace of \( \Lambda \). The dual problem \( (D) \) can then be rewritten using (19) as

\[
(D) \ W = \sup_{\lambda \in S^\perp} \sum_{j \in \mathbb{N}} \phi_j(\lambda_j)
\]

Proposition 4 (Weak duality). \( C(x) \geq \phi(\lambda) \) for any \( x \in F \) and any \( \lambda \in \Lambda \).

Thus, \( V \geq W \).

Proof. The claim trivially holds by (19) if \( \lambda \notin S^\perp \). So suppose that \( \lambda \in S^\perp \). Then, from (19), we get

\[
\phi(\lambda) = \sum_{j \in \mathbb{N}} \phi_j(\lambda_j) \leq \sum_{j \in \mathbb{N}} \left( c_j(x_j) - \lambda_j x_j \right) = \sum_{j \in \mathbb{N}} c_j(x_j) - \sum_{j \in \mathbb{N}} \lambda_j x_j = \sum_{j \in \mathbb{N}} c_j(x_j) = C(x).
\]

Here, the inequality follows from (18) and the penultimate equality holds because \( \lambda \in S^\perp \) and \( x \in S \). This proves the claim. \( \Box \)
Corollary 5. Suppose $x^* \in F$ and $\lambda^* \in \Lambda$ are such that $C(x^*) = \phi(\lambda^*)$. Then $x^*$ is optimal to $(P)$ and $\lambda^*$ is optimal to $(D)$.

Proof. Straightforward, hence omitted. □

Complementary slackness is studied below using the right-derivative $c_j^+$ and the left-derivative $c_j^-$ of the function $c_j$ defined here.

Definition 6. [Right-derivative] (See Section 8A of [22].) The right-derivative of the function $c_j$ at $a_j \leq x_j < b_j$ is given by

$$c_j^+(x_j) \triangleq \lim_{z \uparrow x_j} \frac{c_j(z) - c_j(x_j)}{z - x_j}.$$  

The right-derivative $c_j^+(b_j)$ is defined to equal $\infty$.

Definition 7. [Left-derivative] (See Section 8A of [22].) The left-derivative of the function $c_j$ at $a_j < x_j \leq b_j$ is given by

$$c_j^-(x_j) \triangleq \lim_{z \downarrow x_j} \frac{c_j(z) - c_j(x_j)}{z - x_j}.$$  

The left-derivative $c_j^-(a_j)$ is defined to equal $-\infty$.

The reader is referred to Section 8A in [22], Section 24 in [20], and Section 9.1 in [5] for detailed discussions and proofs of the properties of left- and right-derivatives that are summarize next. Since functions $c_j$ are convex over $[a_j, b_j]$, both the right- and the left-derivatives are finite at every point in the open interval $(a_j, b_j)$. The right-derivative can be $-\infty$ at the left endpoint $a_j$; similarly, the left-derivative can be $+\infty$ at the right-endpoint $b_j$. In this context, a flow $x \in X$ is called regular if $c_j^-(x_j) < \infty$ and $c_j^+(x_j) > -\infty$ for all $j \in \mathbb{N}$ (see Definition 9.2 in Section 9.1 of [5]). The right-derivative is right-continuous over $[a_j, b_j]$ and the left-derivative is left-continuous over $(a_j, b_j]$. Finally, both these derivatives are nondecreasing.

Definition 8 (Complementary slackness). (As in Section 9.7 in [5] on finite-dimensional monotropic programs.) An $x \in X$ and a $\lambda \in \Lambda$ are said to be complementary if, for all $j \in \mathbb{N}$, we have that

$$c_j^-(x_j) \leq \lambda_j \leq c_j^+(x_j).$$  

Remark 9. As in finite-dimensional monotropic programs, an $x^* \in X$ and a $\lambda^* \in \Lambda$ are complementary if and only if $x_j^*$ is the minimizer in (18) for all $j \in \mathbb{N}$.

Proposition 10 (Complementary slackness). Suppose an $x^* \in F$ and a $\lambda^* \in S^\perp$ are complementary. Then $x^*$ is optimal to $(P)$, $\lambda^*$ is optimal to $(D)$, and $V = W$.

Suppose $x^* \in F$ is optimal to $(P)$, $\lambda^* \in S^\perp$ is optimal to $(D)$, and $V = W$. Then $x^*$ and $\lambda^*$ are complementary.

Proof. Suppose $x^* \in F$ and $\lambda^* \in S^\perp$ are complementary. Since $\lambda^* \in S^\perp$, 

$$\phi(\lambda^*) = \sum_{j \in \mathbb{N}} \left( c_j(x_j^*) - \lambda_j^* x_j^* \right) = \sum_{j \in \mathbb{N}} c_j(x_j^*) - \sum_{j \in \mathbb{N}} \lambda_j^* x_j^* = C(x^*).$$  

Here, the first equality follows by Remark 9; the second equality holds because the two series converge; finally, the second series is zero because $\lambda^* \in S^\perp$. This implies that
\( x^* \) and \( \lambda^* \) are optimal to \((P)\) and \((D)\) by Corollary 5 and that their corresponding optimal values are equal.

Conversely, suppose that \( x^* \in F \) and \( \lambda^* \in S^\perp \) are optimal to \((P)\) and \((D)\), respectively, and that \( \phi(\lambda^*) = C(x^*) \). Since \( \lambda^* \in S^\perp \), we have, from Equation (19) that,

\[
\phi(\lambda^*) = \sum_{j \in \mathbb{N}} \phi_j(\lambda^*_j) = \sum_{j \in \mathbb{N}} \min_{x_j \in X_j} \left( c_j(x_j) - \lambda^*_j x_j \right).
\]

Moreover, \( C(x^*) = \sum_{j \in \mathbb{N}} c_j(x_j^*) = \sum_{j \in \mathbb{N}} c_j(x_j^*) - \sum_{j \in \mathbb{N}} \lambda^*_j x_j^* = \sum_{j \in \mathbb{N}} \left( c_j(x_j^*) - \lambda^*_j x_j^* \right) \). Here, the second equality follows because

\[
\sum_{j \in \mathbb{N}} \lambda^*_j x_j^* = 0 \quad \text{and the third equality holds because the two series converge. Combining these three pieces, we obtain,}
\]

\[
\sum_{j \in \mathbb{N}} \left( c_j(x_j^*) - \lambda^*_j x_j^* \right) = \sum_{j \in \mathbb{N}} \min_{x_j \in X_j} \left( c_j(x_j) - \lambda^*_j x_j \right).
\]

This means that \( c_j(x_j^*) - \lambda^*_j x_j^* = \min_{x_j \in X_j} \left( c_j(x_j) - \lambda^*_j x_j \right) \) for all \( j \in \mathbb{N} \). Remark 9 then implies that \( x^* \) and \( \lambda^* \) are complementary. \( \square \)

A planning horizon approach is developed in the next section to establish conditions under which \( V = W \).

### 4. Absence of a duality gap

For each integer \( N \geq 1 \), define the projection of \( S \) into \( \mathbb{R}^N \) as

\[
S_N \triangleq \{(x_1, x_2, \ldots, x_N) | x \in S \} \subset \mathbb{R}^N.
\]

Note that \( S_N \) is a (closed) subspace of \( \mathbb{R}^N \); by arbitrarily extending \( S_N \) with a sequence of real numbers, it will, at times, be convenient to view it as a subspace of \( \mathbb{R}^\infty \). Viewed this way, \( S_N \) can be alternatively defined as

\[
S_N \triangleq \{(x_1, x_2, \ldots, x_N, y_{N+1}, y_{N+2}, \ldots) | x \in S, \ y \in \mathbb{R}^N \} \subset \mathbb{R}^\infty.
\]

In order to avoid unnecessary additional notation, \( S_N \) will denote the above subspace of \( \mathbb{R}^\infty \) and also the corresponding subspace of \( \mathbb{R}^N \). Although this is an abuse of notation, the meaning should be clear from context. This practice is standard in countably infinite optimization (see, for example, [27, 28]).

**Lemma 11.** For any integer \( N \geq 1 \), \( S_N \) is a closed subspace of \( \mathbb{R}^N \).

**Proof.** Let \( \{x(n)\} \in S_N \subset \mathbb{R}^\infty \) be a convergent sequence with limit \( \bar{x} \in \mathbb{R}^\infty \). We show that \( \bar{x} \in S_N \subset \mathbb{R}^N \). Since \( S_N \) is a closed subspace of \( \mathbb{R}^N \), it is clear that \( (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N) \in S_N \subset \mathbb{R}^N \). This means that there is some \( y^* \in S \) such that \( (y_1^*, y_2^*, \ldots, y_N^*) = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N) \) and \( (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N; y_{N+1}^*, y_{N+2}^*, \ldots) \in S_N \subset \mathbb{R}^N \). This shows that \( (\bar{x}_1, \ldots, \bar{x}_N; \bar{x}_{N+1}, \ldots) \in S_N \subset \mathbb{R}^N \) as per the definition of \( S_N \subset \mathbb{R}^N \) in formula (28). \( \square \)

Let \( N_n \geq 1 \), for \( n = 1, 2, \ldots \), be any strictly increasing sequence of positive integers and let \( S_{N_n} \) be the corresponding sequence of projections of \( S \). Consider the corresponding sequence of finite-dimensional monotropic optimization problems

\[
(P_n) \quad V_n \triangleq \inf_{x \in S_{N_n}} C_n(x) \triangleq \sum_{j=1}^{N_n} c_j(x_j)
\]

\[
(30) \quad x \in S_{N_n}, \quad x_j \in X_j, \quad j = 1, 2, \ldots, N_n.
\]
Lemma 12. If \((P)\) has a feasible solution, then \((P_n)\) has a feasible solution.

Proof. If \(x\) is feasible to \((P)\), then its truncation \((x_1, x_2, \ldots, x_{N_n}) \in \mathbb{R}^{N_n}\) is feasible to \((P_n)\). \(\square\)

Let \(X(n) \triangleq \prod_{j=1}^{N_n} X_j\) denote the truncation of \(X\) to \(\mathbb{R}^{N_n}\). Let \(F_n \triangleq S_{N_n} \cap X(n) \subset \mathbb{R}^{N_n}\) denote the feasible region of \((P_n)\).

Lemma 13. If \((P_n)\) has a feasible solution, then it has an optimal solution.

Proof. The feasible region \(F_n = S_{N_n} \cap X(n) \subset \mathbb{R}^{N_n}\) of \((P_n)\) equals the intersection of a closed set \(S_{N_n} \subset \mathbb{R}^{N_n}\) and a compact set \(X(n) \subset \mathbb{R}^{N_n}\); it is therefore compact. The objective in \((P_n)\) is continuous over \(X(n)\) and hence also over \(F_n\). The result then follows by Corollary 2.35 in [1]. \(\square\)

Definition 14. [Kuratowski set limits] Let \(A_1, A_2, \ldots\) be a sequence of sets in a metric space \(\Omega\).

1. (Lower limit; see Section 29.I on page 335 of [13]) A point \(p \in \Omega\) belongs to the lower limit of the sequence \(\{A_n\}\) if, for every open ball \(B(p)\) with center \(p\), there is some index \(m\) large enough such that \(B(p) \cap A_n \neq \emptyset\) for all \(n \geq m\). We then write \(p \in \text{Li} A_n\).

2. (Upper limit; see Section 29.III on page 337 of [13]) A point \(p \in \Omega\) belongs to the upper limit of the sequence \(\{A_n\}\) if, for every open ball \(B(p)\) with center \(p\), there are infinitely many \(n\) with \(B(p) \cap A_n \neq \emptyset\). We then write \(p \in \text{Us} A_n\).

3. (Limit; see Section 29.VI on page 339 of [13]) The sequence \(\{A_n\}\) is said to converge to a set \(A \in \Omega\) if \(\text{Li} A_n = A = \text{Us} A_n\). We then write \(\lim A_n = A\).

Remark 15. (See formula 2a in Section 29.VI on page 339 of [13]) With the notation described in the above definitions, if there is a sequence of points \(p_n \in A_n\) such that \(\bar{p} = \lim_{n \to \infty} p_n\), and if the sequence of sets \(\{A_n\}\) converges, then \(\bar{p} \in \text{Us} A_n\).

Remark 16. (See formula 8 in Section 29.VI on page 339 of [13]) With the notation described in the above definitions, if \(A_1 \supset A_2 \supset \ldots\), then \(\lim A_n = \bigcap_n A_n\), where \(\bar{A}_n\) denotes the closure of \(A_n\).

Lemma 17. We have, \(\mathbb{R}^N \supset S_{N_1} \supset S_{N_2} \supset S_{N_3} \supset \ldots\). Moreover, \(S \subseteq \bigcap_n S_{N_n}\). If \(S\) is closed, then \(S \supset \bigcap_n S_{N_n}\) and hence \(S = \bigcap_n S_{N_n}\).

Proof. For the first claim about set inclusions, fix any integer \(n \geq 1\). Recall as in (28) that \(S_{N_{n+1}} = \{(x_1, x_2, \ldots, x_{n+1}, y_{N_{n+1}+1}, y_{N_{n+1}+2}, \ldots) | x \in S, y \in \mathbb{R}^N\} \subset \mathbb{R}^N\). Thus, if the vector \((x_1, \ldots, x_{N_{n+1}}, y_{N_{n+1}+1}, y_{N_{n+1}+2}, \ldots)\) is in \(S_{N_{n+1}}\), it is also in \(S_{N_n}\). This proves the first claim.

For the second claim, suppose \(x \in S\). Then, by (28), \(x \in S_{N_n} \subset \mathbb{R}^N\) for each \(n\). Consequently, \(x \in \bigcap_n S_{N_n}\).

For the third claim, suppose \(x \in \bigcap_n S_{N_n} \subset \mathbb{R}^N\). Thus, for each \(n\), there is a \(z(n) \in S\) such that \((x_1, x_2, \ldots, x_{N_n}) = (z_1(n), z_2(n), \ldots, z_{N_n}(n))\) by definition of \(S_{N_n}\). Moreover, the sequence \(\{z(n)\}\) is in \(S\) converges to \(x\). Since \(S\) is closed, we have that \(x \in S\). \(\square\)

Proposition 18 (Primal value convergence). Suppose \((P)\) has a feasible solution and \(S\) is closed. Then, problems \((P_n)\) have optimal solutions and \(\lim_{n \to \infty} V_n = V\).
Proof. The proof is similar to the proof of Berge’s maximum principle [4]. Since (P) has a feasible solution, Lemma 12 implies that (P_n) has a feasible solution. Lemma 13 thus means that (P_n) has an optimal solution, which we denote by x^*(n). It will, at times, be convenient to view x^*(n) as belonging to F_n = (S_{N_n} \cap X) \subset X \subset \mathbb{R}^N by seeing S_{N_n} as a subspace of \mathbb{R}^N. Since the sequence \{x^*(n)\} belongs to the compact set X, it has a convergent subsequence. We denote this subsequence by \{x^*(n_m)\}
and use \tilde{x} \in X to denote its limit. Since x^*(n_m) \in S_{N_{n_m}} \subset \mathbb{R}^N for each m, Remarks 15 and 16; Lemmas 11 and 17; and the assumption that S is closed imply that the limit \tilde{x} \in S. In other words, \tilde{x} \in X \cap S = F; that is, \tilde{x} is feasible to (P).

It is shown next by contradiction that \tilde{x} is optimal to (P). So suppose not. Since \tilde{x} is not optimal and since (P) has a feasible solution, there exists a feasible solution \hat{x} such that \sum_{j \in \mathbb{N}} c_j(\hat{x}_j) < \sum_{j \in \mathbb{N}} c_j(\tilde{x}_j). Construct another subsequence \{\hat{x}(n_m)\} \in X such that

\[\tilde{x}(n_m) = \begin{cases} x^*_j(n_m), & j = 1, 2, \ldots, N_{n_m}, \\
\hat{x}_j, & j = N_{n_m} + 1, N_{n_m} + 2, \ldots. \end{cases}\]

This new subsequence also converges to \tilde{x}. Thus, the earlier strict inequality can be rewritten as \sum_{j \in \mathbb{N}} c_j(\hat{x}_j) < \sum_{j \in \mathbb{N}} c_j(\lim_{n \to \infty} \tilde{x}_j(n_m)). Using continuity of functions c_j(\cdot), we further rewrite this inequality as \sum_{j \in \mathbb{N}} c_j(\hat{x}_j) < \sum_{j \in \mathbb{N}} \lim_{n \to \infty} c_j(\tilde{x}_j(n_m)). Now recall from the proof of Lemma 1 that the series \sum_{j \in \mathbb{N}} c_j(x_j) is uniformly convergent over X.

Thus, the series and the limit above can be interchanged (see Theorem 9.7 in [3]) to yield \sum_{j \in \mathbb{N}} c_j(\hat{x}_j) < \lim_{m \to \infty} \sum_{j \in \mathbb{N}} c_j(\tilde{x}_j(n_m)). Now splitting each one of the two series into two parts, we obtain

\[
\sum_{j=1}^{N_{n_m}} c_j(\hat{x}_j) + \sum_{j=N_{n_m}+1}^{\infty} c_j(\hat{x}_j) < \lim_{m \to \infty} \left[ \sum_{j=1}^{N_{n_m}} c_j(\tilde{x}_j(n_m)) + \sum_{j=N_{n_m}+1}^{\infty} c_j(\tilde{x}_j(n_m)) \right].
\]

Then, substituting for \tilde{x}(n_m) from its definition in (32), we get

\[
\sum_{j=1}^{N_{n_m}} c_j(\hat{x}_j) + \sum_{j=N_{n_m}+1}^{\infty} c_j(\hat{x}_j) < \lim_{m \to \infty} \left[ \sum_{j=1}^{N_{n_m}} c_j(x^*_j(n_m)) + \sum_{j=N_{n_m}+1}^{\infty} c_j(\tilde{x}_j) \right].
\]

This implies that there exists an m^* such that

\[
\sum_{j=1}^{N_{n_{m^*}}} c_j(\hat{x}_j) + \sum_{j=N_{n_{m^*}}+1}^{\infty} c_j(\hat{x}_j) < \sum_{j=1}^{N_{n_{m^*}}} c_j(x^*_j(n_{m^*})) + \sum_{j=N_{n_{m^*}}+1}^{\infty} c_j(\tilde{x}_j).
\]

That is, \sum_{j=1}^{N_{n_{m^*}}} c_j(\hat{x}_j) < \sum_{j=1}^{N_{n_{m^*}}} c_j(x^*_j(n_{m^*})). This contradicts the optimality of x^*_j(n_{m^*}) to (P_{n_{m^*}}). This establishes that \tilde{x} is optimal to (P).

To prove value convergence, establish that every convergent subsequence of optimal values V_n converges to the same limit and that this limit is V. This would imply that the sequence V_n converges to V as required. So let V_{n_m} be such a convergent subsequence and let \{x^*_j(n_m)\} \in F_{n_m} = (S_{N_{n_m}} \cap X) \subset X \subset \mathbb{R}^N be the corresponding subsequence of optimal solutions to (P_{n_m}). This subsequence of optimal solutions has a further convergent subsequence, which is denoted by \{\hat{x}^*_j(n_{m'}\}). Denote the limit of this convergent subsequence by \hat{x}. Then, as shown above, \hat{x} is optimal to (P). Fix
any \( \epsilon > 0 \). Then H1 implies that there exists a positive integer \( T \) large enough such that
\[
\sum_{j=N_{mT}+1}^{\infty} c^*_j < \epsilon/2.
\]
Thus, for any \( t \) large enough such that \( N_{mt} > N_{mT} \), we have,
\[
\left| V - V_{n_m} \right| = \left| \sum_{j \in \mathbb{N}} c_j(x_j) - \sum_{j=1}^{N_{n_m}} c_j(x^*_j(n_m)) \right|
\leq \sum_{j=1}^{N_{n_m}} \left| c_j(x_j) - c_j(x^*_j(n_m)) \right| + 2 \sum_{j=N_{mT}+1}^{\infty} c^*_j
\leq \sum_{j=1}^{N_{n_m}} \left| c_j(x_j) - c_j(x^*_j(n_m)) \right| + \epsilon/2.
\]

Now, since \( \lim_{t \to \infty} x^*_j(n_m) = \bar{x}_j \), and since functions \( c_j(\cdot) \) are continuous, there exists a positive integer \( K \) large enough such that
\[
\left| c_j(x_j) - c_j(x^*_j(n_m)) \right| < \epsilon/(2N_{n_m})
\]
for all \( t \geq K \). Combining this observation with the inequality in (35), we see that
\[
\left| V - V_{n_m} \right| < \epsilon \quad \text{for all sufficiently large } t.
\]
Consequently, \( V_{n_m} \) converges to \( V \) as \( t \to \infty \). Since \( V_{n_m} \) is a further subsequence of the convergent subsequence \( V_{n_m} \), it must converge to the same limit as \( V_{n_m} \). In other words, \( \lim_{m \to \infty} V_{n_m} = V \). Thus, every convergent subsequence of \( V_n \) converges to \( V \). This completes the proof.

The dual of the finite-dimensional monotropic program \( (P_n) \) is given by
\[
(37) \quad (D_n) \quad W_n \triangleq \sup_{\lambda \in S_{N_n}^+} \sum_{j=1}^{N_n} \phi_j(\lambda_j)
\]
\[
(38) \quad \lambda \in S_{N_n}^+.
\]
Here, \( S_{N_n}^+ \triangleq \left\{ \lambda \in \mathbb{R}^{N_n} \mid \sum_{j=1}^{N_n} \lambda_j x_j = 0, \forall x \in S_{N_n} \right\} \) is the subspace of \( \mathbb{R}^{N_n} \) that is orthogonal to \( S_{N_n} \).

**Proposition 19 (Zero duality gap, and dual value convergence).** Suppose that \( (P) \) has a feasible solution and that \( S \) is closed. Suppose \( W \geq W_n \) for each \( n \). Then \( V = W \). Moreover, \( \lim_{n \to \infty} W_n = W \).

Proof. Combining weak duality with the hypothesis that \( W \geq W_n \) yields \( V \geq W \geq W_n \). Furthermore, since \( (P_n) \) and \( (D_n) \) are finite-dimensional monotropic programs, we know that \( W_n = V_n \) (see the zero duality gap result in Proposition 9.19 in [5]). Then using the fact from Proposition 18 that \( \lim_{n \to \infty} V_n = V \), we get, \( V \geq W \geq V \).

This proves that \( V = W \).

For the second part, the aforementioned zero duality gap for finite-dimensional monotropic programs implies that \( \lim_{n \to \infty} W_n = \lim_{n \to \infty} V_n \). Then, Propositions 18 and the above zero duality gap result imply that \( \lim_{n \to \infty} W_n = V = W \).

The next lemma shows that \( W \geq W_n \) if costs are nonnegative. This sufficient condition often holds in network flow problems and in CILPs that arise from infinite-horizon planning applications in Operations Research and Economics (see [8, 9, 10, 11, 17]). Nonnegativity of costs has previously been used to show a zero duality gap.

This manuscript is for review purposes only.
result in infinite dimensional linear programs that are more general than CILPs [12].

**Lemma 20.** Suppose that, for each \( j \in \mathbb{N} \), the cost function \( c_j(\cdot) \) is nonnegative over \( X_j \). Then \( W \geq W_n \), for each \( n \).

**Proof.** For each \( \lambda(n) \in S^1_{X_n} \), define \( \xi(\lambda(n)) \) as the corresponding price constructed by appending \( \lambda(n) \) with a sequence of zeros. Specifically,

\[
\xi_j(\lambda(n)) = \begin{cases} 
\lambda_j(n), & j = 1, 2, \ldots, N_n \\
0, & j = N_n + 1, N_n + 2, \ldots 
\end{cases}
\]

Observe that \( \xi(\lambda(n)) \in S^1 \) because for any \( x \in S \), we have,

\[
\sum_{j \in \mathbb{N}} \xi_j(\lambda(n))x_j = \sum_{j=1}^{N_n} \xi_j(\lambda(n))x_j + \sum_{j=N_n+1}^{\infty} \xi_j(\lambda(n))x_j
\]

\[
= \sum_{j=1}^{N_n} \lambda_j(n)x_j + \sum_{j=N_n+1}^{\infty} \xi_j(\lambda(n))x_j = 0.
\]

Here, the first sum is zero because \( \lambda(n) \in S^1_{X_n} \) and the second series is zero because \( \xi_j(\lambda(n)) = 0 \) for \( j = 1 + N_n, 2 + N_n, \ldots \).

Now define \( \Xi_n \triangleq \{\xi(\lambda(n))|\lambda(n) \in S^1_{X_n}\} \subset S^1 \). Consequently,

\[
W_n = \sup_{\lambda(n) \in S^1_{X_n}} \sum_{j=1}^{N_n} \phi_j(\lambda_j(n)) = \sup_{\lambda(n) \in S^1_{X_n}} \left[ \sum_{j=1}^{N_n} \min_{x_j \in X_j} \left( c_j(x_j) - \lambda_j(n)x_j \right) \right]
\]

\[
= \sup_{\xi(\lambda(n)) \in \Xi_n} \left[ \sum_{j=1}^{N_n} \min_{x_j \in X_j} \left( c_j(x_j) - \xi_j(\lambda(n))x_j \right) \right]
\]

\[
\leq \sup_{\xi(\lambda(n)) \in \Xi_n} \left[ \sum_{j \in \mathbb{N}} \min_{x_j \in X_j} \left( c_j(x_j) - \xi_j(\lambda(n))x_j \right) \right]
\]

\[
\leq \sup_{\lambda \in S^1} \left[ \sum_{j \in \mathbb{N}} \min_{x_j \in X_j} \left( c_j(x_j) - \lambda_jx_j \right) \right] = \sup_{\lambda \in S^1} \sum_{j \in \mathbb{N}} \phi_j(\lambda) = W.
\]

Here, the first equality is simply the definition of \( W_n \). The second equality is obtained by substituting for \( \phi_j(\cdot) \) from (18). The third equality is derived by rewriting the problem equivalently in terms of variables \( \xi(\lambda(n)) \). The first inequality holds because \( \xi_j(\lambda(n)) = 0 \) for \( j = N_n + 1, N_n + 2, \ldots \), and functions \( c_j(\cdot) \) are nonnegative. The second inequality holds because the latter problem is a relaxation of the former. The two equalities in (45) follow by the definition of \( W \).

\section{5. Strong duality.}

**Lemma 21.** If \((P)\) has a regular feasible solution, then finite-dimensional dual problems \((D_n)\) have optimal solutions, denoted by \( \lambda^*(n) \).

**Proof.** Since \((P)\) has a regular feasible solution, its projection into \( \mathbb{R}^{N_n} \) is feasible to \((P_n)\) by Lemma 12 and is also regular. By Lemma 13, \((P_n)\) has an optimal solution, say \( x^*(n) \). Then, by strong duality in finite-dimensional monotropic programs (see Proposition 9.18 from [5]), \((D_n)\) has an optimal solution \( \lambda^*(n) \in S^1_{X_n} \) that is complementary to \( x^*(n) \).
THEOREM 22 (Strong duality). Suppose that (P) has a regular feasible solution and that S is closed (thus (P) has an optimal solution by Lemma 1). Let \( \{x^*(n), \lambda^*(n)\} \), for \( n = 1, 2, \ldots, \) be a sequence of pairs of complementary optimal solutions to \((P_n)\) and \((D_n)\) as in Lemmas 13 and 21. Suppose there exists a sequence \( v \in \Lambda \) such that \( |\lambda_j^*(n)| \leq v_j \) for all \( j \in \mathbb{N} \) and all \( n \). Suppose \( S^\perp \) is closed in the product topology on \( \Lambda \). Then (D) has an optimal solution (with \( W = V \)).

Proof. Define the set \( Z \triangleq \prod_{j \in \mathbb{N}} [-v_j, v_j] \), which is in \( \Lambda \) by H2 and H3 because \( v \in \Lambda \). This is compact by Tychoff’s theorem. By appending solutions \( \lambda^*(n) \) with a sequence of zeros, we view them as elements of subspace \( \Lambda \), and in particular, as members of the set \( Z \).

As in the proof of Proposition 18, we continue to use \( x^*(n) \) to also denote its extension in \( F_n = (S_{N_n} \cap X) \times X \times \mathbb{N} \). Since \( X \) and \( Z \) are compact, the sequence \( \{x^*(n), \lambda^*(n)\} \subseteq X \times Z \) has a convergent subsequence \( \{x^*(n_k), \lambda^*(n_k)\} \). We denote its limit by \((\bar{x}, \bar{\lambda}) \in X \times Z \). As shown in Proposition 18, \( \bar{x} \in F \). Since \( Z \subseteq \Lambda \), we know that \( \bar{\lambda} \in \Lambda \). More specifically, we prove that \( \bar{\lambda} \in S^\perp \). To see this, note that the sequence \( \{\lambda^*(n)\} \in \Lambda \) in fact belongs to \( S^\perp \). This holds because, for any \( y \in S \),

\[
\sum_{j \in \mathbb{N}} \lambda_j^*(n) y_j = \sum_{j=N_n}^{N_{n+1}} \lambda_j^*(n) y_j + \sum_{j=N_{n+1}+1}^{\infty} \lambda_j^*(n) y_j = 0. \quad \text{Here, the last equality holds because}
\]

\[
\sum_{j=1}^{N_n} \lambda_j^*(n) y_j = 0 \quad \text{as} \quad (\lambda_1^*(n), \lambda_2^*(n), \ldots, \lambda_{N_n}^*(n)) \in S_{N_n}^\perp, \quad \text{and} \quad \lambda_j^*(n) = 0 \quad \text{for} \quad j \geq N_n + 1.
\]

Since \( S^\perp \) is assumed to be closed, the limit \( \bar{\lambda} \) of the sequence \( \{\lambda^*(n)\} \) in \( S^\perp \) must belong to \( S^\perp \).

It is shown next that \( \bar{x} \in F \) and \( \bar{\lambda} \in S^\perp \) are complementary in \((P)\) and \((D)\). This implies, by the first claim in Proposition 10, that \( \bar{\lambda} \) is optimal to \((D)\).

To show that \( \bar{x} \) and \( \bar{\lambda} \) are complementary, we need to prove that

\[
c_j^-(\bar{x}_j) \leq \bar{\lambda}_j \leq c_j^+(\bar{x}_j),
\]

for every \( j \in \mathbb{N} \). If \( j \) is such that the interval \( X_j \) is a single point, that is, if \( a_j = b_j \), then \( \bar{x}_j = a_j = b_j \). Consequently, \( c_j^-(\bar{x}_j) = -\infty \) and \( c_j^+(\bar{x}_j) = \infty \); the inequality in (46) then holds for any real number \( \bar{\lambda}_j \). We therefore only need to focus on indices \( j \) for which the interval \( X_j \) is not a single point. For these indices, we prove (46) by contradiction. So that there exists either a \( j \in \mathbb{N} \) such that

\[
c_j^-(\bar{x}_j) > \bar{\lambda}_j
\]

or a \( j \in \mathbb{N} \) such that

\[
c_j^+(\bar{x}_j) < \bar{\lambda}_j.
\]

First suppose that it is the former. Then \( \bar{x}_j \neq a_j \) because \( c_j^-(a_j) = -\infty \) and this contradicts (47). We consider two other subcases: (A) there exists a subsequence \( x_j^*(n_{k_t}) \) such that \( x_j^*(n_{k_t}) \leq \bar{x}_j \) for all \( t \); and (B) there is no such subsequence and hence \( x_j^*(n_{k_t}) > \bar{x}_j \) for all \( k \) large enough. In subcase (A), \( x_j^*(n_{k_t}) \uparrow \bar{x}_j \) as \( t \to \infty \).

Then, since the left-derivative is left-continuous over \((a_j, b_j)\), we have \( c_j^-(\bar{x}_j) = c_j^- (\uparrow x_j^*(n_{k_t})) = \lim_{t \to \infty} c_j^- (x_j^*(n_{k_t})) \). Moreover, \( \bar{\lambda}_j = \lim_{t \to \infty} \lambda_j^*(n_{k_t}) \). These two observations, when combined with (47), yield that \( \lim_{t \to \infty} c_j^- (x_j^*(n_{k_t})) > \lim_{t \to \infty} \lambda_j^*(n_{k_t}) \). Thus there must be some \( t \) for which \( c_j^- (x_j^*(n_{k_t})) > \lambda_j^*(n_{k_t}) \). But this contradicts the fact that
the pair \( x_j^*(n_k) \), \( \lambda_j^*(n_k) \) is complementary in \((P_{n_k})\) and \((D_{n_k})\). In subcase (B), we have that \( c_j^- (x_j^*(n_k)) \geq c_j^- (\bar{x}_j) \) for all \( k \) large enough because the left-derivative is nonincreasing. Moreover, \( \lim_{k \to \infty} \lambda_j^*(n_k) = \bar{\lambda}_j \). Combining these two observations with (47) we obtain \( c_j^+ (x_j^*(n_k)) > \lambda_j^*(n_k) \) for some \( k \) large enough. Again, this contradicts the fact that the pair \( x_j^*(n_k), \lambda_j^*(n_k) \) is complementary in \((P_{n_k})\) and \((D_{n_k})\).

Now suppose that it is the latter. Then \( \bar{x}_j \neq b_j \) because \( c_j^+(b_j) = +\infty \) and this contradicts (48). We consider two other subcases: (A) there exists a subsequence \( x_j^*(n_k) \) such that \( x_j^*(n_k) \geq \bar{x}_j \) for all \( t \); and (B) there is no such subsequence and hence \( x_j^*(n_k) < \bar{x}_j \) for all \( k \) large enough. In subcase (A), \( x_j^*(n_k) \downarrow \bar{x}_j \) as \( t \to \infty \). Then, since the right-derivative is right-continuous over \([a_j, b_j]\), we have, \( c_j^+(\bar{x}_j) = c_j^+ (x_j^*(n_k)) = \lim_{t \to \infty} c_j^+ (x_j^*(n_k)) \). Moreover, \( \bar{\lambda}_j = \lim_{k \to \infty} \lambda_j^*(n_k) \). These two observations, when combined with (48), yield that \( \lim_{k \to \infty} c_j^+ (x_j^*(n_k)) < \lim_{k \to \infty} \lambda_j^*(n_k) \). Thus there must be some \( k \) for which \( c_j^+ (x_j^*(n_k)) < \lambda_j^*(n_k) \). But this contradicts the fact that the pair \( x_j^*(n_k), \lambda_j^*(n_k) \) is complementary in \((P_{n_k})\) and \((D_{n_k})\). In subcase (B), we have that \( c_j^+ (x_j^*(n_k)) \leq c_j^+ (\bar{x}_j) \) for all \( k \) large enough because the right-derivative is nondecreasing. Moreover, \( \lim_{k \to \infty} \lambda_j^*(n_k) = \bar{\lambda}_j \). Combining these two observations with (48) we obtain \( c_j^+ (x_j^*(n_k)) < \lambda_j^*(n_k) \) for some \( k \) large enough. Again, this contradicts the fact that the pair \( x_j^*(n_k), \lambda_j^*(n_k) \) is complementary in \((P_{n_k})\) and \((D_{n_k})\). 

6. Countably infinite linear programs. As in [25], consider the following class of linear optimization problems:

\[
(CILP) \inf \sum_{j \in \mathbb{N}} c_j x_j
\]

\[
\sum_{j \in \mathbb{N}} d_{ij} x_j = b_i, \quad i = 1, 2, \ldots
\]

\[
0 \leq x_j \leq u_j, \quad j = 1, 2, \ldots
\]

These problems can be seen as special cases of the CILPs considered in [9, 24] wherein the formulation did not include explicit upper bounds on variables but nevertheless required the variables to be bounded without loss of optimality for strong duality. We work with the bounded-variable version here to be consistent with our framework in Section 2. Here, \( c \in \mathbb{R}^\mathbb{N} \), \( b \in \mathbb{R}^\mathbb{N} \), and \( 0 < u \in \mathbb{R}^\mathbb{N} \) are given sequences of real numbers and \( D = \{ d_{ij} \} \) is a doubly-infinite matrix of real numbers. As in [24, 25], assume that each row and each column of \( D \) includes a finite number of nonzero entries. This assumption is prevalent in duality results on CILPs and it holds in most problems in Operations Research. Under this structure of the constraint coefficient matrix, assume without loss of generality that there is an increasing sequence of positive integers \( \{ M_n \} \) with the following property: variables \( x_j \) indexed by \( j = M_n + 1, M_n + 2, \ldots \) do not appear in the equality constraints (50) indexed by \( i = 1, 2, \ldots, n \). Thus, the ith equality constraints (50) can be equivalently written as \( \sum_{j=1}^{M_i} d_{ij} x_j = b_i \).

We first convert this problem into a countably infinite monotropic program by employing a process identical to that in finite-dimensional LPs (see Section 9.7 in [5]). Introduce the sequence \( z = (z_1, z_2, \ldots) \) of variables to rewrite the ith constraint in (50) as \( \sum_{j=1}^{M_i} d_{ij} x_j - z_i = 0 \), and add constraints \( b_1 \leq z_i \leq b_i, \) for \( i = 1, 2, \ldots \). Use the
(52) \[ S \triangleq \left\{ (x, z) \mid \sum_{j=1}^{M_i} d_{ij} x_j - z_i = 0, \ i = 1, 2, \ldots \right\}. \]

Also let intervals \( X_j \triangleq [0, u_j] \) for \( j \in \mathbb{N} \) and \( Z_i \triangleq [b_i, b_i] \). Then, problem (CILP) can be equivalently rewritten as the countably infinite monotropic program (MONO-CILP)

\[ \inf_{x \in X} \sum_{j \in \mathbb{N}} c_j x_j \]

(54) \((x, z) \in S, \]

(55) \( x_j \in X_j, \ j = 1, 2, \ldots, \]

(56) \( z_i \in Z_i, \ i = 1, 2, \ldots. \]

Hypothesis H1 from Section 2 then reduces to:

**H1.** the series \( \sum_{j \in \mathbb{N}} \max_{x_j \in X_j} |c_j x_j| \) of nonnegative terms is finite.

Several applications where H1 holds, including CILP formulations of infinite-horizon planning problems, countable state MDPs, and robust optimization, are discussed in [17, 10, 11, 24, 25]. The space \( \Lambda \) of dual sequences \( \lambda = (\lambda_1, \lambda_2, \ldots) \) and the space \( \Theta \) of dual sequences \( \theta = (\theta_1, \theta_2, \ldots) \) are then characterized by the hypotheses

**H2.** the series \( \sum_{j \in \mathbb{N}} |\lambda_j||u_j| + \sum_{i \in \mathbb{N}} |\theta_i||b_i| \) of nonnegative terms is finite; and

**H3.** the series \( \sum_{j \in \mathbb{N}} |\lambda_j y_j| + \sum_{i \in \mathbb{N}} |\theta_i w_i| \) of nonnegative terms is finite for every \((y, w) \in S,\]

Let \( X \triangleq \prod_{j \in \mathbb{N}} X_j \) and \( Z \triangleq \prod_{i \in \mathbb{N}} Z_i \). The Lagrangian function in \( (14) \) then reduces to

\[ L((x, z), (y, w); \lambda, \theta) = \sum_{j \in \mathbb{N}} (c_j x_j - \lambda_j x_j) - \sum_{i \in \mathbb{N}} \theta_i z_i + \sum_{j \in \mathbb{N}} \lambda_j y_j + \sum_{i \in \mathbb{N}} \theta_i w_i, \]

for \( x \in X, z \in Z, (y, w) \in S, \lambda \in \Lambda, \) and \( \theta \in \Theta \). Now, similar to \( (12) \), we define,

\[ \phi_j(\lambda_j) \triangleq \min_{x_j \in X_j} (c_j x_j - \lambda_j x_j), \]

\[ \psi_i(\theta_i) \triangleq \min_{z_i \in \mathbb{Z}_i} \theta_i z_i = -\theta_i b_i, \]

because \( Z_i = \{b_i\} \).

Then the dual problem (D) from Section 2 reduces to

\[ \sup_{\lambda \in \Lambda} \inf_{\theta \in \Theta} \left\{ \sum_{j \in \mathbb{N}} \phi_j(\lambda_j) - \sum_{i \in \mathbb{N}} b_i \theta_i \right\}. \]

(60) \( (\lambda, \theta) \in S^\perp, \]

where

\[ S^\perp \triangleq \left\{ (\lambda, \theta) \in \Lambda \times \Theta \mid \sum_{j \in \mathbb{N}} \lambda_j y_j + \sum_{i \in \mathbb{N}} \theta_i w_i = 0, \ \forall (y, w) \in S \right\}. \]

The weak duality result in Proposition 4 and complementary slackness result in Proposition 10 then hold for the pair (MONO-CILP) and (D-MONO-CILP).
In order to apply our zero duality gap and strong duality results from Sections 4 and 5, we first define appropriate finite-dimensional projections of (MONO-CILP). Toward this end, let \( N_n \triangleq M_n + n \), for all positive integers \( n \). We define the projection subspaces

\[
S_{N_n} \triangleq \left\{ (x_1, \ldots, x_{M_n}; z_1, \ldots, z_n) \in \mathbb{R}^{N_n} \bigg| (x, z) \in S \right\}.
\]

**Lemma 23.** Consider the subspace

\[
\sigma_{N_n} \triangleq \left\{ (x_1, \ldots, x_{M_n}; z_1, \ldots, z_n) \in \mathbb{R}^{N_n} \bigg| \sum_{j=1}^{M_n} d_{ij} x_j - z_i = 0, \ i = 1, 2, \ldots, n \right\}.
\]

Then, \( S_{N_n} = \sigma_{N_n} \).

**Proof.** First suppose that \( (x_1, x_2, \ldots, x_{M_n}; z_1, z_2, \ldots, z_n) \in S_{N_n} \). Then, \((x, z) \in S\) and hence \( \sum_{j=1}^{M_n} d_{ij} x_j - z_i = 0 \) for all \( i \); in particular, \( \sum_{j=1}^{M_n} d_{ij} x_j - z_i = 0 \) for \( i = 1, 2, \ldots, n \).

Thus, the vector \((x_1, x_2, \ldots, x_{M_n}; z_1, z_2, \ldots, z_n) \in \sigma_{N_n} \). This shows that \( S_{N_n} \subseteq \sigma_{N_n} \).

Now suppose that \( (x_1, x_2, \ldots, x_{M_n}; z_1, z_2, \ldots, z_n) \in \sigma_{N_n} \). We use this to construct a solution \((x_1, \ldots, x_{M_n}, x_{M_n+1}, x_{M_n+2}, \ldots; z_1, \ldots, z_n, z_{n+1}, z_{n+2}, \ldots) \in S\) as follows.

Set \( x_{M_n+1} = x_{M_n+2} = \ldots = 0 \) and \( z_{n+k} = \sum_{j=1}^{M_n} d_{(n+k)j} x_j \). As a result, the vector \((x_1, x_2, \ldots, x_{M_n}; z_1, z_2, \ldots, z_n) \in S_{N_n} \). Thus, \( \sigma_{N_n} \subseteq S_{N_n} \).

These two observations show that \( S_{N_n} = \sigma_{N_n} \).

In view of the above lemma, we equivalently use \( \sigma_{N_n} \) instead of \( S_{N_n} \) to formulate the finite-dimensional monotropic programming projections of (MONO-CILP) as

\[
(MONO-CILP(n)) \quad \inf \sum_{j=1}^{M_n} c_j x_j \quad \text{subject to} \quad \sum_{j=1}^{M_n} d_{ij} x_j - z_i = 0, \ i = 1, 2, \ldots, n, \quad 0 \leq x_j \leq u_j, \ j = 1, 2, \ldots, M_n, \quad z_i = b_i, \ i = 1, 2, \ldots, n.
\]

As a side note, these projections are familiar, finite-dimensional LPs; this is the benefit of employing \( \sigma_{N_n} \) instead of \( S_{N_n} \) to write these problems. We have,

**Lemma 24.** The subspace \( S \) defined in (52) is closed.

**Proof.** Consider a convergent sequence \( \{(x^n, z^n)\} \in S \) with limit \((\bar{x}, \bar{z})\). We need to show that \((\bar{x}, \bar{z}) \in S \). Consider any fixed \( i \). We have,

\[
\sum_{j=1}^{M_n} d_{ij} \bar{x}_j - \bar{z}_i = \sum_{j=1}^{M_n} d_{ij} \left( \lim_{n \to \infty} x^n_j \right) - \lim_{n \to \infty} z^n_i = \lim_{n \to \infty} \left( \sum_{j=1}^{M_n} d_{ij} x^n_j - z^n_i \right) = 0.
\]

Here, the last equality holds because \((x^n, z^n) \in S \). Since \( i \) was arbitrary, \((\bar{x}, \bar{z}) \in S \) as required.

This manuscript is for review purposes only.
Thus, as long as (MONO-CILP) has a feasible solution, optimal values of the finite-dimensional projections (MONO-CILP(n)) converge to the optimal value of (MONO-CILP) as per Proposition 18. We also note, from the fundamental theorem of linear algebra (see page 198 of [30]), that
\[ \sigma_{\lambda_n} = \{ (\lambda_1, \ldots, \lambda_{M_n} ; \theta_1, \ldots, \theta_n) | \sum_{i=1}^{n} d_{ij} \theta_i + \lambda_j = 0, \quad j = 1, \ldots, M_n \}. \]
Thus, the dual of (MONO-CILP(n)) is given by
\[
(D-MONO-CILP(n)) \sup_{\{ \lambda_n \}} \sum_{j=1}^{M_n} \phi_j (\lambda_j) - \sum_{i=1}^{n} b_i \theta_i
\]
\[
(D-MONO-CILP(n)) \sup_{\{ \lambda_n \}} \sum_{i=1}^{n} d_{ij} \theta_i + \lambda_j = 0, \quad j = 1, \ldots, M_n.
\]
If \( c_j \geq 0 \) for all \( j \in \mathbb{N} \), then optimal values of these finite-dimensional duals provide a lower bound on the optimal value of (D-MONO-CILP) as per Lemma 20. Thus, in this case, if (MONO-CILP) has a feasible solution, then there is no duality gap between (MONO-CILP) and (D-MONO-CILP) by Proposition 19.

The remaining two hypotheses from the strong duality result in Theorem 22 require that the optimal variables in (D-MONO-CILP(n)) be bounded independently of \( n \) and the orthogonal subspace \( S^\perp \) be closed. We therefore first characterize \( S^\perp \) and show that it is closed. We need additional notation to achieve this. So recall (MONO-CILP) as per Proposition 18. We also note, from the fundamental theorem of linear algebra (see page 198 of [30]), that
\[
S^\perp = \{ (\lambda, \theta) \in \Lambda \times \Theta | \sum_{i \in J_i} d_{ij} \theta_i + \lambda_j = 0, \quad j = 1, 2, \ldots \}.
\]

**Lemma 25.** The orthogonal subspace \( S^\perp \) in (62) can be equivalently rewritten as
\[
S^\perp = \{ (\lambda, \theta) \in \Lambda \times \Theta | \sum_{i \in J_i} d_{ij} \theta_i + \lambda_j = 0, \quad j = 1, 2, \ldots \}.
\]

**Proof.** Suppose \((\lambda, \theta) \in S^\perp\) as defined in (72). Let \((y, w) \in S\). Then, we have,
\[
\sum_{j \in \mathbb{N}} \lambda_j y_j + \sum_{i \in \mathbb{N}} \theta_i w_i = \sum_{i \in \mathbb{N}} \sum_{j \in J_i \setminus J_{i-1}} \lambda_j y_j + \sum_{i \in \mathbb{N}} \theta_i w_i
\]
\[
= \sum_{i \in \mathbb{N}} \sum_{j \in J_i \setminus J_{i-1}} \left( - \sum_{k \in I_j} d_{kj} \theta_k \right) y_j + \sum_{i \in \mathbb{N}} \theta_i w_i
\]
\[
= - \sum_{i \in \mathbb{N}} \theta_i \left( \sum_{j \in J_i} d_{ij} y_j \right) + \sum_{i \in \mathbb{N}} \theta_i w_i = - \sum_{i \in \mathbb{N}} \theta_i w_i + \sum_{i \in \mathbb{N}} \theta_i w_i = 0.
\]
Here, the first equality is obtained by reordering terms in the series \( \sum_{j \in \mathbb{N}} \lambda_j y_j \), which is allowed because the series converges absolutely by H3. The second equality holds
because \( \lambda_j = - \sum_{k \in I_j} d_{kj} \theta_k \) as per (72). The third equality is obtained by collecting terms according to their indices \( i \). This shows that \((\lambda, \theta) \in S^\perp\) as defined in (62).

Now, for the reverse, suppose that \((\lambda, \theta) \in S^\perp\) as defined in (62). We will show by contradiction that \((\lambda, \theta) \in S^\perp\) as defined in (72). So suppose not. Then, there must exist some \( j \), say \( j = 1 \) without loss of generality, such that \( \lambda_1 + \sum_{i \in I_1} d_{ij} \theta_i = \epsilon_1 \neq 0 \).

Now consider a \((y, w)\) defined by \( y_1 = 1, y_j = 0 \) for \( j \neq 1 \); and \( w_i = d_{ij} \) for all \( i \). This implies that \((y, w) \in S\) because, for any \( i \), we have, \( \sum_{j \in I_i} d_{ij} y_j = d_{ij} w_i \).

Moreover, we obtain, \( \sum_{j \in \mathbb{N}} \lambda_j y_j + \sum_{i \in \mathbb{N}} \theta_i w_i = \lambda_1 + \sum_{i \in I_1} \theta_i d_{i1} = \lambda_1 + \sum_{i \in I_1} \theta_i d_{i1} = \epsilon_1 \neq 0 \).

This contradicts the fact that \((\lambda, \theta) \in S^\perp\) as defined in (62). \(\square\)

This lemma can be viewed as a countably infinite extension of one piece of the fundamental theorem of linear algebra and hence it may be of independent interest.

This lemma allows us to view the constraint \((\lambda, \theta) \in S^\perp\) in (D-MONO-CILP) in the more familiar manner as \( \sum_{i \in I_j} d_{ij} \theta_i + \lambda_j = 0 \), for \( j = 1, 2, \ldots \), using the transpose of the doubly infinite matrix \( D \).

**Lemma 26.** The orthogonal subspace defined in (72) is closed.

**Proof.** Suppose \( \{(\lambda^n, \theta^n)\} \subseteq S^\perp \) is a convergent sequence with limit \( (\bar{\lambda}, \bar{\theta}) \). We need to show that \( (\bar{\lambda}, \bar{\theta}) \in S^\perp \). Fix any \( j \). We have,

\[
(76) \quad \sum_{i \in I_j} d_{ij} \bar{\theta}_i + \bar{\lambda}_j = \sum_{i \in I_j} d_{ij} \left( \lim_{n \to \infty} \theta^n_i \right) + \lim_{n \to \infty} \lambda^n_j = \lim_{n \to \infty} \left( \sum_{i \in I_j} d_{ij} \theta^n_i + \lambda^n_j \right) = 0.
\]

Here, the first equality holds by definition of \( \bar{\theta} \) and \( \bar{\lambda} \). The second equality holds because the set \( I_j \) is finite. The last equality holds because \( (\lambda^n, \theta^n) \in S^\perp \). Since \( j \) was arbitrary, \( (\bar{\lambda}, \bar{\theta}) \in S^\perp \). \(\square\)

Thus, Theorem 22 implies that as long as (MONO-CILP) has a feasible solution, strong duality holds between (MONO-CILP) and (D-MONO-CILP) when optimal dual variables in (D-MONO-CILP(n)) are bounded independently of \( n \). This essentially recovers Theorem 3.2 in [9]. See [9, 24, 25] for several applications where variables in the finite-dimensional dual problems are bounded in this way without loss of optimality.

### 7. Convex minimum cost flow problems on infinite networks.

Recall the description of a minimum cost flow problem in countably infinite networks from the author’s recent work in [17]. Let \( \mathcal{N} \) denote a countable set of nodes. An arc from node \( i \in \mathcal{N} \) to node \( j \in \mathcal{N} \) is denoted by the ordered tuple \((i, j)\); let \( \mathcal{A} \) denote the countable set of all such arcs. Assume the standard regularity condition that each node has finite in- and out-degree (see [23, 29]). The resulting infinite directed network is denoted by \( \mathcal{G} \triangleq (\mathcal{N}, \mathcal{A}) \).

A real number \( s_i \), called source, is assigned to each node \( i \in \mathcal{N} \). If \( s_i > 0 \), node \( i \) is called a supply node and a net flow of \( s_i \) needs to be pushed out from this node; if \( s_i = 0 \), node \( i \) is called a transshipment node; finally, if \( s_i < 0 \), node \( i \) is called a demand node and a net flow of \(-s_i\) needs to be delivered to this node. The largest amount of flow that can be carried through arc \((i, j) \in \mathcal{A}\) is denoted by \( 0 < u_{ij} < \infty \); these are called flow capacities. Arc flows are denoted by \( x_{ij} \), for \((i, j) \in \mathcal{A} \). For each arc \((i, j) \in \mathcal{A} \), let \( c_{ij} : [0, u_{ij}] \to \mathbb{R} \) be a real-valued, continuous, and convex function.
The cost of carrying a flow of $x_{ij}$ through arc $(i,j) \in A$ equals $c_{ij}(x_{ij})$. We make the natural assumptions that $c_{ij}(0) = 0$ and flow costs are nondecreasing over $[0, u_{ij}]$.

Use $\mathbb{R}^A$ to denote the set of all sequences $x \triangleq \{x_{ij}\}_{(i,j) \in A}$ of real numbers indexed by the arcs in $A$. The goal is to find a flow $x \in \mathbb{R}^A$ that satisfies the supply, demand, and transshipment requirements at all nodes; abides by the arc flow capacities; and achieves this at minimum total cost. This minimum cost network flow problem can then be formulated as

\[
\text{(NET-FLOW)} \quad \inf \ C(x) = \sum_{(i,j) \in A} c_{ij}(x_{ij})
\]

\[
\sum_{\{j|i,j\} \in A} x_{ij} - \sum_{\{j|i,j\} \in A} x_{ji} = s_i, \quad \forall i \in N,
\]

\[
0 \leq x_{ij} \leq u_{ij}, \quad \forall (i,j) \in A,
\]

\[
x \in \mathbb{R}^A.
\]

A minimum cost flow problem on a finite network can be viewed as a finite-dimensional monotropic optimization problem. In particular, the finite-dimensional subspace in this monotropic optimization problem corresponds to the so-called circulation subspace of an equivalent minimum cost flow problem on a finite network wherein all nodes are transshipment nodes. This can be achieved by inserting a single artificial (transshipment) node and several artificial arcs into the original network flow problem. For each supply or demand node in the original network, there is an artificial arc that starts at the artificial node and ends at the supply or demand node. The lower and upper bounds on the flow through this arc equal the supply quantity or demand quantity. The cost on this arc is set arbitrarily. The original supply or demand node is then converted into a transshipment node. See Figure 4.2 and Section 4.1.3 in [5]. Unfortunately, a naive implementation of this transformation does not work in general in the countably infinite case because it would call for an artificial node whose total degree is infinite. An alternative approach is thus proposed here.

For each supply or demand node in the original network, insert an infinite sequence of arcs and transshipment nodes that push flow into this node. The lower and upper bounds on the flows through these arcs equal the supply or demand quantity. All supply and demand nodes in the original network are thus converted into transshipment nodes. The resulting problem can then be seen as a minimum cost circulation problem on a countably infinite network.

Additional notation is needed to make the above conversion rigorous and to ensure that the resulting circulation problem is indeed a special case of countably infinite monotropic optimization. Let $M \subseteq N$ denote the set of supply or demand nodes in the original network. For each node $i \in M$, let $N_i$ denote the corresponding countable set of artificial nodes in the circulation problem. Use $N \triangleq N \bigcup \left( \bigcup_{i \in N} N_i \right)$ to denote the countable set of nodes in the circulation problem. Similarly, for each node $i \in M$, let $A_i$ denote the corresponding countable set of artificial arcs in the circulation problem. Let $\tilde{A} \triangleq \tilde{A} \bigcup \left( \bigcup_{i \in N} A_i \right)$ denote the countable set of arcs in the circulation problem. If $k \in M$, then set the flow bounds for all corresponding artificial arcs $(i,j) \in A_k$ as $s_k \leq x_{ij} \leq s_k$. The flow costs for all artificial arcs are set to zero — that is, $c_{ij}(x_{ij}) = 0$ for all $(i,j) \in \tilde{A} \setminus A$.

Now consider the following minimum cost circulation problem in a countably
infinite network:

\[ (81) \quad \text{(CIRC-FLOW)} \quad \inf_{x_{ij} \in \mathcal{A}} \sum_{(i,j) \in \mathcal{A}} c_{ij}(x_{ij}) \]

\[ (82) \quad \sum_{j(i,j) \in \mathcal{A}} x_{ij} - \sum_{i(j,i) \in \mathcal{A}} x_{ji} = 0, \quad \forall i \in \mathcal{N}, \]

\[ (83) \quad 0 \leq x_{ij} \leq u_{ij}, \quad \forall (i,j) \in \mathcal{A}, \]

\[ (84) \quad s_k \leq x_{ij} \leq s_k, \quad \forall (i,j) \in \mathcal{A}_k, \quad \forall k \in \mathcal{M}, \]

\[ (85) \quad x \in \mathbb{R}^d. \]

Note that if some flow in \( \mathbb{R}^d \) is optimal to (CIRC-FLOW), then its restriction to \( \mathbb{R}^d \) is optimal to (NET-FLOW); these two solutions have identical optimal objective values in the two problems. Conversely, if some flow in \( (85) \) is optimal to (NET-FLOW), then its extension to \( \mathbb{R}^d \), obtained by setting the flows in artificial arcs to the appropriate supply or demand values, is optimal to (CIRC-FLOW); again, these two solutions have identical optimal values. In this sense, the two problems are equivalent. Furthermore, (CIRC-FLOW) can be seen as a countably infinite monotropic optimization problem where variables \( x_{ij} \) are indexed with arcs \( (i,j) \in \mathcal{A} \). This is achieved by defining

\[ (86) \quad X_{ij} \triangleq \begin{cases} [0, u_{ij}], & \text{if } (i,j) \in \mathcal{A}, \\ [s_k, s_k], & \text{if } (i,j) \in \mathcal{A}_k \text{ for some } k \in \mathcal{M}, \end{cases} \]

and the circulation subspace \( S \triangleq \left\{ x \in \mathbb{R}^d \mid \sum_{(j(i,j) \in \mathcal{A})} x_{ij} - \sum_{(i(j,i) \in \mathcal{A})} x_{ji} = 0, \forall i \in \mathcal{N} \right\} \). Note that the interval \( X_{ij} \) reduces to the single point \( s_k \) if arc \( (i,j) \) belongs to the set of artificial arcs \( \mathcal{A}_k \) for some supply or demand node \( k \in \mathcal{M} \). This poses no additional analytical difficulties in our duality results below (see page 410 of Section 9.1 from [5] for a discussion of the left- and the right-derivative of a convex function over a degenerate, single-point interval).

Hypotheses H1, H2, H3 from Section 2 then reduce to:

**H1.** the series \( \sum_{(i,j) \in \mathcal{A}} c_{ij}(u_{ij}) \) of nonnegative terms is finite;

**H2.** the series \( \sum_{(i,j) \in \mathcal{A}} |\lambda_{ij}| u_{ij} + \sum_{k \in \mathcal{M}} \sum_{(i,j) \in \mathcal{A}_k} |\lambda_{ij}| s_k \) of nonnegative terms is finite;

and

**H3.** the series \( \sum_{(i,j) \in \mathcal{A}} |\lambda_{ij} y_{ij}| \) of nonnegative terms is finite for each \( y \in S \).

Recall that hypotheses H2 and H3 characterize the subspace \( \Lambda \subset \mathbb{R}^d \) where the dual sequences \( \lambda \triangleq \{ \lambda_{ij} \}_{(i,j) \in \mathcal{A}} \) reside. We also let \( X \triangleq \prod_{(i,j) \in \mathcal{A}} X_{ij} \). The Lagrangian function in (14) then reduces to

\[ (87) \quad L(x, y; \lambda) \triangleq \sum_{(i,j) \in \mathcal{A}} \left( c_{ij}(x_{ij}) - \lambda_{ij} x_{ij} \right) + \sum_{(i,j) \in \mathcal{A}} \lambda_{ij} y_{ij}, \]

for all \( x \in X, y \in S, \) and \( \lambda \in \Lambda \). Consequently, (18) reduces to

\[ (88) \quad \phi_{ij}(\lambda_{ij}) \triangleq \min_{x_{ij} \in X_{ij}} \left( c_{ij}(x_{ij}) - \lambda_{ij} x_{ij} \right), \quad \forall (i,j) \in \mathcal{A}. \]
Then the dual problem can be written as

\[
(D-FLOW) \quad \sup_{(i,j)\in \bar{A}} \sum_{(i,j)\in \bar{A}} \phi_{ij}(\lambda_{ij})
\]

\[
\lambda \in S^\perp,
\]

where

\[
S^\perp \triangleq \{ \lambda \in \Lambda \mid \sum_{(i,j)\in \bar{A}} \lambda_{ij}y_{ij} = 0, \forall y \in S \}.
\]

Our weak duality result in Proposition 4 and complementary slackness result in Proposition 10 then hold for the pair \((CIRC-FLOW)\) and \((D-FLOW)\).

A suitable definition of finite-dimensional projections of the circulation subspace is now needed to derive zero duality gap and strong duality results. Toward this end, since the circulation problem is defined on a countably infinite network, it can be viewed as a layered problem. That is, as explained in [17, 29], this countably infinite network can be partitioned into mutually exclusive and exhaustive layers of nodes indexed by positive integers \(n\). The first layer includes an arbitrarily chosen, single node; the second layer includes all nodes that are connected to the node in the first layer; the third layer includes all nodes that are connected to the nodes in the second layer, and so on. As in [17], let \(\bar{L}_n\) denote the set of nodes in the \(n\)th layer and let \(\bar{L}_n\) denote the set of nodes in the first \(n\) layers. Also, let \(\bar{A}_n\) denote the set of arcs connected to the nodes in the first \(n\) layers and \(N_n \triangleq \#\bar{A}_n\) be the number of these arcs.

Now, for each positive integer \(n\), define the projection \(S_{N_n}\) of \(S\) into \(\mathbb{R}^{N_n}\) as the set of ordered tuples of flows \(x_{ij}\) in arcs \((i,j)\in \bar{A}_n\), where each tuple is obtained by truncating some \(x \in S\). This leads to the sequence of finite-dimensional monotropic optimization problems

\[
(CIRC-FLOW(n)) \quad \inf_{(i,j)\in \bar{A}_n} \sum_{(i,j)\in \bar{A}_n} c_{ij}(x_{ij})
\]

\[
x \in S_{N_n},
\]

\[
x_{ij} \in X_{ij}, (i,j) \in \bar{A}_n.
\]

**Lemma 27.** The circulation subspace \(S\) is closed.

**Proof.** Follows easily by showing that the limit \(\bar{x} \in \mathbb{R}^\bar{A}\) of a convergent sequence \(x^n \in S\) also lies in \(S\). Details omitted.

This discussion implies that our primal value convergence result in Proposition 18 applies to \((CIRC-FLOW)\). That is, as long as \((CIRC-FLOW)\) has a feasible solution, optimal values of the finite-dimensional problems \((CIRC-FLOW(n))\) converge to the optimal value of \((CIRC-FLOW)\). Moreover, since all costs are nonnegative, Lemma 20 implies that there is no duality gap between \((CIRC-FLOW)\) and \((D-FLOW)\) by Proposition 19.

Recall that the two main remaining hypotheses from the strong duality result in Theorem 22 are that the finite-dimensional dual optimal solutions be bounded and the subspace \(S^\perp\) be closed. Orthogonal subspace \(S^\perp\) for \((CIRC-FLOW)\) is therefore investigated next.
Lemma 28. The orthogonal subspace \( S^\perp \) of the circulation subspace is characterized by

\[
S^\perp = \left\{ \lambda \in \Lambda \left| \lambda_{ij} = \pi_i - \pi_j, \ \forall (i,j) \in \bar{A}, \text{ for some } \pi \in \mathbb{R}^\bar{N} \right. \right\}.
\]

(Here, \( \pi_i \) is interpreted as the “price” of node \( i \).)

Proof. Recall the original definition of the orthogonal subspace from Equation (91). The required result can be derived as a special case/corollary of Lemma 25. An alternative proof sketch below, however, is more revealing of the network structure at hand.

First, if there is a \( \lambda \in S^\perp \) as defined in (95), then this \( \lambda \) also belongs to \( S^\perp \) as defined in (91). To see this, suppose \( \lambda \in \Lambda \) is such that there is some \( \pi \in \mathbb{R}^\bar{N} \) with \( \lambda_{ij} = \pi_i - \pi_j \) for every \( (i,j) \in \bar{A} \). Consider any \( y \in S \). Then,

\[
\sum_{(i,j)\in \bar{A}} \lambda_{ij} y_{ij} = \sum_{i\in \bar{N}} \sum_{j(i,j)\in \bar{A}} \lambda_{ij} y_{ij} = \sum_{i\in \bar{N}} \sum_{j(i,j)\in \bar{A}} (\pi_i - \pi_j) y_{ij} = \sum_{i\in \bar{N}} \pi_i \left( \sum_{j(i,j)\in \bar{A}} y_{ij} - \sum_{\{j|j(j,i)\in \bar{A}\}} y_{ji} \right) = 0.
\]

Here, the first equality is simply a rearrangement of terms in the series, which is allowed because the series converges absolutely by H3. This rearrangement is obtained by summing over nodes first and then summing over arcs — this second sum only includes outgoing arcs at each node to avoid double-counting. The second equality was obtained by substituting \( \pi_i - \pi_j \) for \( \lambda_{ij} \). The third equality was obtained by collecting common terms that multiply \( \pi_i \) for each node \( i \in \bar{N} \). This shows that \( \lambda \in S^\perp \) as defined in (91).

Second, if there is a \( \lambda \in S^\perp \) as defined in (91), then this \( \lambda \) also belongs to \( S^\perp \) as defined in (95). That is, this \( \lambda \) can be expressed as \( \lambda_{ij} = \pi_i - \pi_j \) for every \( (i,j) \in \bar{A} \) using some \( \pi \in \mathbb{R}^\bar{N} \). The existence of one such \( \pi \) is in fact constructively demonstrated here. Assume, without loss of generality, that the first node in \( \bar{N} \) has at least one outgoing arc. Define a set of nodes \( P \) that initially includes only the first node in \( \bar{N} \); that is, \( P = \{1\} \). Set \( \pi_1 \) arbitrarily. We will refer to \( P \) as the set of priced nodes. Now set \( \pi_j = \pi_1 - \lambda_{1j} \) for all nodes \( j \in \bar{N} \setminus \{1\} \) such that \( (1,j) \in \bar{A} \), and add these priced nodes \( j \) to set \( P \). We call this step “fanning out from node 1.” Now partition \( P \) into two subsets: \( F \subseteq P \) is the subset of priced nodes that have been fanned out from, and \( P \setminus F \) is the set of priced nodes that have not. Thus, at this point, \( F = \{1\} \). Repeat this procedure for each node in \( P \setminus F \). Specifically, for node \( i \in F \), set \( \pi_j = \pi_i - \lambda_{ij} \) for all nodes \( j \in \bar{N} \setminus P \) such that \( (i,j) \in \bar{A} \), and add these priced nodes \( j \) to \( P \setminus F \). If, at any point during this process, \( P \setminus F \) becomes empty, pick any node \( i \) from \( \bar{N} \setminus P \) that has an arc \( (i,j) \in \bar{A} \) such that \( j \in P \). Let \( \pi_i = \pi_j + \lambda_{ij} \); add node \( i \) to \( P \setminus F \) and continue. Since the network includes infinitely many nodes, this process never terminates, but it does imply the existence of a \( \pi \) as required. Observe that if the network does not include any (undirected) cycles, then this \( \pi \) is clearly consistent with the given \( \lambda \). However, if the network includes one or more (undirected) cycles, then more care is needed in proving this consistency. Rather than proving this in complete detail (which requires additional, tedious notation), we sketch the idea in this proof using a concrete example of a finite network with nodes \( \{1,2,3\} \) and arcs \( (1,2), (1,3), (2,3) \).
If one applied the above procedure to this network starting with \( \pi_1 = 0 \), then one would obtain, \( \pi_2 = -\lambda_1 \) and \( \pi_3 = -\lambda_1 \). Then, to ensure consistency, one needs to check whether or not \( \pi_2 - \pi_3 = \lambda_{23} \). Subtracting \( \pi_3 \) from \( \pi_2 \) yields \( \pi_2 - \pi_3 = \lambda_{13} - \lambda_{12} \).

Thus, the procedure leads to a consistent solution if and only if \( \lambda_{13} - \lambda_{12} = \lambda_{23} \); that is, if and only if \( \lambda_{13} - \lambda_{12} - \lambda_{23} = 0 \). This is where the fact that \( \lambda \) is in the subspace orthogonal to the circulation subspace comes in handy. In particular, note that the flow \( x_{12} = -1 \), \( x_{13} = 1 \), and \( x_{23} = -1 \) is in the circulation subspace of this network. Consequently, since \( \lambda \) must be orthogonal to this flow, we must have that \( \lambda_{13} - \lambda_{12} - \lambda_{23} = 0 \). This idea works more generally even when there are multiple (undirected) cycles; an appropriate flow in the circulation subspace can be identified by setting flows in all arcs other than the cycle under consideration to zero. \( \Box \)

**Lemma 29.** The orthogonal subspace \( S^\perp \) defined in (95) in Lemma 28 is closed.

**Proof.** Let \( \{ \lambda^n \} \) be a convergent sequence in \( S^\perp \) with limit \( \bar{\lambda} \in \mathbb{R}^N \). Thus, there is a corresponding sequence \( \{ \pi^n \} \subset \mathbb{R}^N \) such that, for each \( n \), \( \lambda_{ij}^n = \pi^n_i - \pi^n_j \) for all \( (i, j) \in \mathcal{A} \). Such a sequence \( \{ \pi^n \} \) may not in general be convergent. However, the procedure outlined for recovering a \( \pi \in \mathbb{R}^N \) from a given \( \lambda \in S^\perp \) in the proof of Lemma 28 ensures that the sequence \( \{ \pi^n \} \) is convergent as long as the order in which nodes are priced is invariant over \( n \). Suppose that \( \bar{\pi} \in \mathbb{R}^N \) is its limit. Then, for each \( (i, j) \in \mathcal{A} \), we have, \( \lambda_{ij} = \lim_{n \to \infty} \lambda_{ij}^n = \lim_{n \to \infty} (\pi^n_i - \pi^n_j) = \bar{\pi}_i - \bar{\pi}_j \). This shows that \( \bar{\lambda} \in S^\perp \). Consequently, \( S^\perp \) is closed as claimed. \( \Box \)

Thus, Theorem 22 reduces to the statement that as long as (CIRC-FLOW) has a regular feasible solution, strong duality holds between (CIRC-FLOW) and (D-FLOW), if optimal solutions to the duals of (CIRC-FLOW(n)) are uniformly bounded independently of \( n \). This essentially recovers the strong duality Theorem 5.2 from [17]; also see that paper for applications in Operations Research and Economics where this boundedness property of dual variables holds.

**8. Separable convex programs with linear constraints.** We briefly outline how our results from Sections 3, 4, and 5 can be applied to problems of the form

\[
\text{(CONV)} \inf \; \sum_{j \in \mathbb{N}} c_j(x_j); \sum_{j \in \mathbb{N}} d_{ij}x_j = b_i, \; i = 1, 2, \ldots; \; 0 \leq x_j \leq u_j, \; j = 1, 2, \ldots.
\]

Here, for each \( j \in \mathbb{N} \), \( c_j(\cdot) \) is a real-valued, convex, and continuous function over the closed and bounded interval \([0, u_j]\). The notation in the equality constraints is as defined for CILPs in Section 6. Specifically, (CONV) is a convex extension of (CILP) from Section 6. It can also be viewed as a generalization of the network flow problem (NET-FLOW) from Section 7 and also of the hypernetwork flow problems studied in [17]. Problem (CONV) can be seen as a countably infinite monotropic program by equivalently redefining the equality constraints and by employing the subspace (52) as we did for (CILP). Moreover, it is also possible to write the dual of (CONV) by using H1, H2, and H3; then the above duality results on monotropic programs apply to this primal-dual pair.

**References.**


This manuscript is for review purposes only.