An Infinite Dimensional Linear Programming Algorithm for Deterministic Semi-Markov Decision Processes on Borel Spaces

Diego Klabjan
Department of Mechanical and Industrial Engineering
University of Illinois at Urbana-Champaign
Urbana, IL
klabjan@uiuc.edu

Daniel Adelman
Graduate School of Business
University of Chicago
Chicago, IL
dan.adelman@gsb.uchicago.edu

July 11, 2006

Abstract

We devise an algorithm for solving the infinite dimensional linear programs that arise from general deterministic semi-Markov decision processes on Borel spaces. The algorithm constructs a sequence of approximate primal/dual solutions that converge to an optimal one. The innovative idea is to approximate the dual solution with continuous piecewise linear ridge functions that naturally represent functions defined on a high dimensional domain as linear combinations of functions defined on only a single dimension. This approximation gives rise to a primal/dual pair of semi-infinite programs, for which we show strong duality. In addition, we prove various properties of the underlying ridge functions.

1 Introduction

Linear programming is a classical approach to the solution of Markov decision processes (MDP), and dates back to the pioneering work of Ghellinck (1960), D'Epenoux (1960), and Manne (1960) for MDP's with finite state and action spaces. Recently, there has been substantial interest in MDP's on general Borel spaces, e.g. see Hernández-Lerma and Lasserre (1996), Hernández-Lerma and Lasserre (1999) and references therein. In this setting, the linear programs are no longer defined on finite spaces, but infinite ones. Unlike finite linear programs, which enjoy the wide availability of solution software, solution methodologies for infinite linear programs are rare. Typically, algorithms are specially designed for particular classes of problems. One exception to this is the paper by Hernández-Lerma and Lasserre (1998a).

Klabjan and Adelman (2006) develop infinite dimensional linear programming theory for semi-Markov decision processes (SMDP) on Borel spaces and long-run time average expected cost criterion, with particular emphasis on the special case in which the state transitions are deterministic. Included in this theory are general conditions for strong duality and the existence of an optimal stationary, deterministic control policy. Adelman and Klabjan (2005b) apply this theory to the generalized joint replenishment problem, which includes many classical problems in inventory control for which the duality and existence questions have remained open for many years.

In this paper, we devise a new algorithm for solving the primal/dual pair of infinite dimensional linear programs that arise from deterministic semi-Markov decision problems with long-run time average cost criterion. We show that under some mild conditions the algorithm is convergent. Our algorithm constructs a sequence of measures that converge weakly to an optimal solution to the primal problem, and a corresponding sequence of dual feasible functions that converge to an optimal solution to the dual problem. A key contribution of our work is in specific forms of functional approximations to the feasible solutions of the dual problem. In Klabjan and Adelman (2006) we showed how to construct an optimal stationary, deterministic control policy from an optimal primal solution.

Bellman and Dreyfus (1959) were the first to consider functional approximations, in their case Legendre polynomials, in the solution of the dynamic programming optimality equation defined on continuous spaces. More recently, Johnson *et al.* (1993) considered the use of multivariate spline functions to approximate the value function. The idea is to discretize the state space, evaluate the value function at the grid points, and then use multivariate splines to interpolate between grid points. Because the number of multivariate basis functions grows exponentially with the number of state space dimensions, Chen *et al.* (1999) consider an experimental design approach from statistics, but with multivariate spline functions of lower polynomial order. These authors consider only the finite-horizon case, work directly with the optimality equations instead of the linear programming formulations as we do here, and do not show convergence. Recently, de Farias and Van Roy (2003, 2004) have considered linear programming for approximate dynamic programming. The key idea, which comes from Schweitzer and Seidmann (1985), is to approximate the value function with a linear combination of basis functions. There has been considerable recent interest in using this approach. However, there remains a fundamental open question: how to generate basis functions? Until now, all research in approximate dynamic programming has assumed a fixed set of basis functions, chosen a priori by the human modeler. Ours is the first paper, to our knowledge, to provide an algorithm that automates the generation of basis functions. Our algorithm dynamically constructs basis functions as the algorithm proceeds, in the limit converging in such a way as to close the gap between the approximate value function and the true optimal value function. The key is limiting the search for new basis functions to a particular well-structured class that has the power to approximate, arbitrarily closely, any bounded measurable function.

Hernández-Lerma and Lasserre (1998b) propose a convergent algorithm for solving the infinite dimensional linear programs resulting from MDPs. Our algorithm differs from theirs in fundamental ways. Whereas we approximate the dual problem, they approximate the primal problem. They consider countable dense subsets of measures for aggregating the variables and countable dense subsets of functions to aggregate constraints of the primal problem. In contrast, we work with an explicit form of functional approximations, which are computationally tractable and still give a convergent algorithm under some mild conditions. Linear programming for semi-infinite problems is considered by Tadic *et al.* (2006), where the problem is reformulated as a stochastic semi-infinite problem, which is then solved by Monte-Carlo simulation.

Our main idea is to approximate the bias function with piecewise linear functions. In our context, a piecewise linear function is always continuous and it has a finite number of breakpoints. The key question is how to define multivariate piecewise linear functions. The definition must allow for an easy encoding of such functions and it should be easy to capture properties such as continuity, convexity, and monotonicity. Such properties are important in applications of approximate dynamic programming, see e.g. Powell and Topaloglu (2003), to ensure computational tractability and economic sensibility. The multivariate spline functions, see e.g. Wang (1994) or Bojanov *et al.* (1983), are very complex and it is extremely difficult to impose the aforementioned properties. For this reason we took a different approach where the mapping from the multi-dimensional space to the 1-dimensional space is carried via the notion of ridge functions, see e.g. Cheney and Light (1999). In the 1-dimensional space we then use piecewise linear functions or 1-splines to approximate continuous functions. It is well known that higher order splines have better convergence properties in other settings and they are smooth. However, in our setting they would yield nonlinear, very hard to solve programs. By using piecewise linear functions often the resulting programs are linear mixed integer programs, which can be solved using readily available optimization software.

Our algorithm begins with a collection of piecewise linear functions $f_j : \mathbb{R} \to \mathbb{R}$, which correspond to 1-splines. Given a Borel state space X of high dimension, corresponding to function j we map each state x into a scalar $r^j x$, where r^j is called a *ridge vector* having dimension equal to that of X. We then approximate the bias function, $u(\cdot)$, with the summed *ridge function*, i.e.

$$u(x) \approx \sum_j f_j(r^j x) \qquad x \in X.$$

By choosing r^j to be different from the unit vectors, we capture interactions between components of the state space. Letting b^j denote the vector of breakpoints for function f_j, we thus start the algorithm with a collection $\{r^j, b^j\}_j$. By substituting the above approximation into the dual infinite program, the slopes between breakpoints for each function f_j essentially become decision variables in what turns out to be a semi-infinite linear program. We show that this program is solvable and has strong duality. Using the primal solution of this semi-infinite program, we then solve an auxiliary problem that finds additional breakpoints and/or ridge vectors, and then we repeat the whole procedure. We show that primal and dual solutions generated by this algorithm converge to an optimal solution under some mild assumptions. In Adelman and Klabjan (2005a), we discuss how to efficiently implement the basic algorithm discussed here on the generalized joint replenishment problem, including how to construct feasible control policies in practice that are (near) optimal.

The paper is organized as follows. In Section 2 we give a brief overview of infinite dimensional linear programming for deterministic SMDPs. In Section 3 we define ridge functions and prove various properties about them. We show in Section 4 how to formulate the problem of finding optimal weights for the ridge function as a semi-infinite linear program, prove a strong duality result, and provide an algorithm for solving it within arbitrary precision. Then in Section 5 we give the main algorithm and prove convergence.

2 Overview of Infinite Dimensional Linear Programming for Deterministic SMDPs

In this section we review the linear programming theory developed in Klabjan and Adelman (2006) and Adelman and Klabjan (2005b).

Given a Borel space Z we denote by $\mathbb{C}(Z)$ the set of all continuous functions on Z and by $\mathbb{B}(Z)$ the set of all Borel measurable bounded functions on Z. The set of all finite signed Borel measures on Z is denoted by $\mathbb{M}(Z)$. All these three sets can be equipped with a norm and the last two become Banach spaces. Let also $\mathcal{B}(Z)$ be the Borel σ-algebra in Z.

2.1 Formulation

Consider a deterministic SMDP defined on a state space X and action space A, both assumed to be Borel spaces. For each $x \in X$, let $A(x) \subseteq A$ be a non-empty Borel subset that specifies the set of admissible actions from state x. We denote the collection of state-action pairs as $K = \{(x, a) : x \in X, a \in A(x)\}$, assumed to be a Borel subset of $X \times A$. Upon taking action a in state x, a cost $c(x, a)$ is incurred and then the system transitions to some state $s(x, a)$ after a time duration of length $\tau(x, a)$, all with probability one. We assume that $c : K \to \mathbb{R}$, $s : K \to X$, and $\tau : K \to [0, \infty)$ are measurable on K. Let $\{x_n, a_n, t_n\}_{n=0,1,\ldots} \in (K \times [0, \infty))^\infty$ denote any infinite sequence of state-action pairs and transition times. Suppose $f : X \to A$ is a measurable decision function that specifies for every $x \in X$ some action $a \in A(x)$. Define the long-run average cost of the system under control f, starting from an initial state $x_0 \in X$, as

$$J(f, x_0) = \limsup_{N\to\infty} \frac{\sum_{n=0}^N c(x_n, f(x_n))}{\sum_{n=0}^N t_n}.$$

The problem

$$J(x_0) = \inf_{f:X\to A} J(f, x_0)$$

finds an optimal decision rule f^* from starting state x_0. One of the main questions in Markov control processes is under what conditions does there exist an f^* such that $J^* = J(f^*, x_0) = J(x_0)$ for every

$x_0 \in X$? Such a decision rule is said to be *long-run time average optimal*, in the class of stationary deterministic decision rules, from every starting state.

More generally, rather than restricting the class of policies to deterministic decision rules $f : X \to A$, we could pose the existence question over all admissible, non-anticipatory policies $\pi \in \Pi$, including randomized history-dependent ones. It follows from Klabjan and Adelman (2006) that stationary deterministic policies still suffice for optimality.

2.2 Infinite Dimensional Linear Programming Theory

We reformulate the problem of finding an optimal policy as an infinite dimensional linear program.

We consider the following primal/dual linear programs on the spaces $(\mathbb{M}(K), \mathbb{B}(K))$, $(\mathbb{R} \times \mathbb{M}(X), \mathbb{R} \times \mathbb{B}(X))$. The primal problem (P) is

$$\min \int_K c(x,a)\mu(\mathrm{d}(x,a)) \tag{1a}$$

$$\int_K \tau(x,a)\mu(\mathrm{d}(x,a)) = 1 \tag{1b}$$

$$\mu((B \times A) \cap K) - \mu(\{(x,a) \in K : s(x,a) \in B\}) = 0 \qquad \text{for every } B \in \mathcal{B}(X) \tag{1c}$$

$$\mu \geq 0, \mu \in \mathbb{M}(K) \tag{1d}$$

and the corresponding dual problem (D) reads

$$\begin{aligned} &\max \rho \\ &\tau(x,a)\rho + u(x) - u(s(x,a)) \leq c(x,a) \qquad \text{for every } (x,a) \in K \\ &\rho \in \mathbb{R}, u \in \mathbb{B}(X)\,. \end{aligned} \tag{2}$$

We denote by $\inf(P)$ and $\sup(D)$ the optimal values of the primal and dual programs, respectively. We call constraints (1c) the *flow balance constraints*.

Next we provide a set of assumptions under which strong duality holds between these two programs.

Assumption B1. *τ is continuous, nonnegative, and bounded on K.*

Assumption B2. *c is lower semi-continuous and nonnegative on K.*

Assumption B3. *$c(x,a) + \tau(x,a) \geq 1$ for every $(x,a) \in K$.*

Here the right-hand side can be changed to any $\epsilon > 0$, but we normalize to 1 for convenience.

Assumption B4. *$\{a \in A(x) : c(x,a) + \tau(x,a) \leq r\}$ is compact for every $x \in X, r \in \mathbb{R}$.*

Assumption B5. *There exists a decision rule $f : X \to A$ and initial state $x_0 \in X$ such that $J(f, x_0) < \infty$.*

Assumption B6. *s is continuous on K.*

Assumption B7. *K is compact.*

The following assumption says that all states communicate with bounded cost and time.

Assumption B8. *There exist constants $C < \infty$, $\Gamma < \infty$ such that for every measurable subset $S \subseteq X$ there is a decision rule $f : X \setminus S \to A$ with the property that for every $x' \in X \setminus S$ there exists a finite integer N and a set of states $x_0, x_1, \ldots, x_N$ with*

- *$x_0 = x'$,*
- *$a_n = f(x_n) \in A(x_n)$ for every $n = 0, \ldots, N-1$,*
- *$x_{n+1} = s(x_n, a_n)$ for every $n = 0, \ldots, N-1$,*

- $x_N \in S$,
- $\sum_{n=0}^{N-1} c(x_n, a_n) \le C$, *and*
- $\sum_{n=0}^{N-1} \tau(x_n, a_n) \le \Gamma$.

This communication assumption holds for a wide variety of deterministic continuous time infinite time horizon inventory problems, Adelman and Klabjan (2005b). For example, the classical joint replenishment problem has this property.

The following important theorems are given in Adelman and Klabjan (2005b).

Theorem 1. (Strong duality) Under Assumptions B1–B8, there exists an optimal primal/dual solution pair $(\mu^*, (\rho^*, u^*)) \in (\mathbb{M}(K), (\mathbb{R}, \mathbb{B}(X)))$ such that $\inf(P) = \sup(D)$ and complementary slackness holds, i.e. for μ^*-almost all $(x, a) \in K$ we have

$$\tau(x,a)\rho^* + u^*(x) = c(x,a) + u^*(s(x,a)).$$

Since (P) and (D) are solvable, instead of denoting by $\inf(P), \sup(D)$ their respective values, we denote them by $\min(P)$ and $\max(D)$. The next theorem states that there is an optimal deterministic stationary policy.

Theorem 2. (General existence result) Under Assumptions B1–B8, there exists a decision rule $f^* : X \to A$ such that

$$J(x) = J(f^*, x) = J^* \quad \text{for all } x \in X.$$

In the present work we give an algorithm that solves (P) and (D). The main idea is to approximate u by certain functions, which we introduce next.

3 Ridge Functions

Let $b_0^j < b_1^j < \cdots < b_{m_j}^j < b_{m_j+1}^j$ for $j \in [n] = \{1, 2, \ldots, n\}$ be real numbers in an interval $[-\Omega, \Omega]$ except that $b_0^j < -\Omega$ and $b_{m_j+1}^j > \Omega$. Each such ordered set of numbers is denoted by b^j. For each $j \in [n]$ and $i \in [m_j]$ let $H_i^j : [-\Omega, \Omega] \to \mathbb{R}$ be defined as

$$H_i^j(z) = \begin{cases} \frac{z - b_{i-1}^j}{b_i^j - b_{i-1}^j} & b_{i-1}^j \le z \le b_i^j \\ \frac{b_{i+1}^j - z}{b_{i+1}^j - b_i^j} & b_i^j \le z \le b_{i+1}^j \\ 0 & \text{otherwise.} \end{cases}$$

Note that $H_i^j(b_i^j) = 1$ and $\operatorname{supp}(H_i^j) = (b_{i-1}^j, b_{i+1}^j)$, see Figure 1. By $\operatorname{supp}(\cdot)$ we denote the support set of a function. These functions are known as *hat* functions or also *B-splines of degree 1* and they form the basis among the spline functions of degree 1, see e.g. Bojanov *et al.* (1983, pp. 33-35).

If $n = 1$, then we have a single ordered set of numbers b^1. It is easy to see that for any weights $w_i^1, i \in [m_1]$ the function $f(z) = \sum_{i=1}^{m_1} w_i^1 H_i^1(z)$ is piecewise linear. We next summarize some basic properties of these functions, which are easy to prove.

Proposition 1. The following properties hold for every $i \in [m_1]$.

1. f is continuous piecewise linear with breakpoints b_i^1 and $f(b_i^1) = w_i^1$.
2. The slope of f in $\left[b_i^1, b_{i+1}^1\right]$ is
$$\frac{w_{i+1}^1 - w_i^1}{b_{i+1}^1 - b_i^1}\,.$$

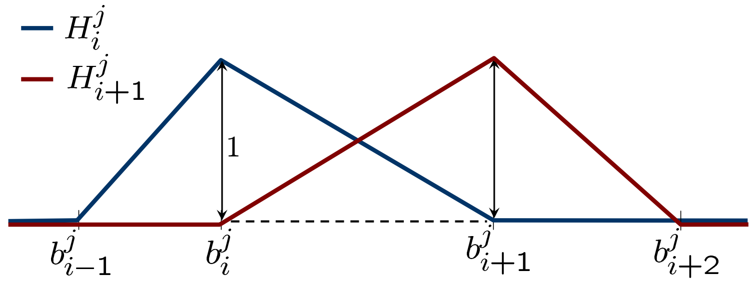

Figure 1: H_i^j and H_{i+1}^j functions

3. If $z \in \left[b_i^1, b_{i+1}^1\right]$, then

$$f(z) = w_i^1 H_i^1(z) + w_{i+1}^1 H_{i+1}^1(z) = w_i^1 \frac{b_{i+1}^1 - z}{b_{i+1}^1 - b_i^1} + w_{i+1}^1 \frac{z - b_i^1}{b_{i+1}^1 - b_i^1}\,.$$

The next theorem is fundamental for this work and it establishes a form of a density property for a specific family of functions within $\mathbb{B}(\bar{X})$.

Theorem 3. Let $f \in \mathbb{B}(\bar{X})$, where $\bar{X} \subseteq \mathbb{R}^q$ for some positive integer q is compact. Let $\mu \in \mathbb{M}(\bar{X})$, and let $\epsilon > 0, \delta > 0$. Then there exist

- a positive integer $n < \infty$ and $\Omega \in \mathbb{R}_+$,
- vectors $r^1, r^2, \ldots, r^n$ in $\mathbb{R}^q$ with $\|r^j\|_\infty = 1$ for every $j \in [n]$,
- ordered sets $b^1, b^2, \cdots, b^n$ with $b_i^j \in [-\Omega, \Omega]$ for every $j \in [n], i \in [m_j]$, where m_j is a positive finite integer,
- and weights $w_i^j \in \mathbb{R}, j \in [n], i \in [m_j]$

such that

$$\|f - \sum_{j=1}^{n} \sum_{i=1}^{m_j} w_i^j H_i^j(< r^j, \cdot >)\|_\infty \leq \epsilon$$

except on a subset E of $\bar{X}$ with $|\mu(E)| \leq \delta$. Here $< x, y >$ denotes the standard inner product in $\mathbb{R}^q$.

Proof. Let $\epsilon > 0, \delta > 0$ be arbitrary. Since μ is a finite signed measure, its total variation $|\mu|$ is a finite positive measure. Therefore $|\mu|$ is regular (see e.g. Rudin (1986, pp. 47-48)) and by Lusin's theorem (see e.g. Rudin (1986, pp. 55)) there exists a subset $E \subseteq \bar{X}$ with $|\mu|$ measure less than or equal to δ and a continuous function $g : \bar{X} \to \mathbb{R}$ such that g and f differ only on E and $\|g\|_\infty \leq \|f\|_\infty < \infty$. Note that $|\mu(E)| \leq |\mu|(E) \leq \delta$.

We call a subset of $\mathbb{C}(Y)$ fundamental if its linear span is dense. Since $\bar{X}$ is compact and $r \in \mathbb{R}^q$, $\|r\|_\infty = 1$ we have for all $x \in \bar{X}$

$$|\ rx\ | \leq \|r\|_2 \cdot \|x\|_2 \leq \sqrt{q}\|x\|_2 \leq \sqrt{q} \max_{x' \in \bar{X}} \|x'\|_2 \equiv \Omega < \infty\,.$$

The set of all continuous piecewise linear functions on $Y = [-\Omega, \Omega]$, denoted by PL, is fundamental in $\mathbb{C}(Y)$ (see e.g. Davis (1975, pp. 122-125)). Then the set $\{h(< r, \cdot >) : h \in PL, \|r\|_\infty = 1\}$ is fundamental in

$\mathbb{C}(\bar{X})$, see e.g. Cheney and Light (1999, pp. 168). Hence by definition and $g \in \mathbb{C}(\bar{X})$, there exist $\lambda_1, \ldots, \lambda_n$ and $r^1, r^2, \ldots, r^n \in \mathbb{R}^q$ with $\|r^j\|_\infty = 1$ for every $j \in [n]$, and $h_1, \ldots, h_n \in PL$ such that

$$\|g - \sum_{j=1}^{n} \lambda_j h_j(< r^j, \cdot >)\|_\infty \leq \epsilon\,.$$

The hat functions form a basis in PL (see e.g. Bojanov *et al.* (1983, pp. 33-35)) and therefore for each j we have $h_j = \sum_{i=1}^{m_j} \bar{w}_{ij} H_i^j$. By defining $w_i^j = \bar{w}_{ij}\lambda_j$ we obtain the claim. □

Let $u : \bar{X} \to \mathbb{R}$ be defined by

$$u(x) = \sum_{j=1}^{n}\sum_{i=1}^{m_j} w_i^j H_i^j(r^j x)\,, \tag{3}$$

where $\bar{X} \subseteq \mathbb{R}^q$. We call vectors r the *ridge vectors* and u is called the *ridge function*. We refer to b's as *breakpoint sets* and to w's as the *weights*. Without loss of generality we assume that all ridge vectors are nonzero.

We remark that given a ridge function u the ridge vectors are not uniquely defined. Consider, for example, a ridge function with two ridge vectors r_1, r_2 that are linearly dependent. It is easy to see that in this case u can also be written as a ridge function with only a single ridge vector r_1 but different breakpoints. Unfortunately this observation cannot be generalized to more than 2 arbitrarily chosen linearly dependent vectors. If $r^2 = \lambda_1 r^1 + \lambda_2 r^2$, where $\lambda_1 \neq 0, \lambda_2 \neq 0$, then in general the corresponding ridge function cannot be written as a ridge function with only 2 piecewise linear functions. Additional such degenerate cases are given in the next section.

By Theorem 3, ridge functions can approximate arbitrarily close dual feasible functions in (D), except on subsets of arbitrarily small measure. We show in Section 5 a slightly stronger version. Namely, there is an optimal solution to (D), which is the limsup of feasible ridge functions, though not necessarily a ridge function.

3.1 Convexity, Monotonicity, and Boundness

In this section we study convexity, monotonicity and boundness of ridge functions. The first two properties are important in approximate dynamic programming, Powell and Topaloglu (2003). We first address convexity. Given a set $C \subseteq \mathbb{R}^q$ we denote by $\text{int}(C)$ its interior.

For every $j \in [n]$ let $\underline{m}_j$ be the smallest index with the property that there exists $x \in \bar{X}$ such that $b^j_{\underline{m}_j} < r^j x < b^j_{\underline{m}_j+1}$. In addition, for every $j \in [n]$ let $\overline{m}_j$ be the largest index with the property that there exists $x \in \bar{X}$ such that $b^j_{\overline{m}_j} < r^j x < b^j_{\overline{m}_j+1}$. Let $R = \{(j,i) | j \in [n]\,, \underline{m}_j + 1 \leq i \leq \overline{m}_j\}$. For each $(j,i) \in R$ let $S_i^j = \{(l,k) \in R |$ there exists $\lambda^{lj}_{ki} \neq 0$ such that $(r^j, b^j_i) = \lambda^{lj}_{ki}(r^l, b^l_k)\}$. Note that $(j,i) \in S_i^j$.

Proposition 2. Let $\bar{X}$ be a convex full-dimensional set. Then u is convex if and only if

$$\sum_{(l,k)\in S_i^j} \lambda^{lj}_{ki} \cdot \frac{w^l_k - w^l_{k-1}}{b^l_k - b^l_{k-1}} \leq \sum_{(l,k)\in S_i^j} \lambda^{lj}_{ki} \cdot \frac{w^l_{k+1} - w^l_k}{b^l_{k+1} - b^l_k} \qquad \text{for each } (j,i) \in R\,. \tag{4}$$

Note that if $n = 1$, then, by using Proposition 1 and the fact $S_i^1 = \{(1,i)\}, \lambda^{11}_{ii} = 1$, these are the standard conditions for convexity of piecewise linear functions. In addition, if all ridge vectors r are linearly independent, then $S_i^j = \{(j,i)\}$ and therefore (4) state convexity conditions for every individual piecewise linear function. The proof is given in Appendix A.

Next we study monotonicity. In our context, a function u is monotone if $u(x) \leq u(y)$ whenever $x \leq y$. Let $X_i^j = \{x \in \text{int}(\bar{X})\,|b^j_i \leq r^j x < b^j_{i+1}\}$ and $\Upsilon = \{(k_1, \ldots, k_n) | X^1_{k_1} \cap X^2_{k_2} \cap \cdots \cap X^n_{k_n} \neq \emptyset\}$. For every $(k_1, \ldots, k_n) \in \Upsilon$ we denote $\zeta(k_1, \ldots, k_n) = \{j \in [n] : r^j x = b^j_{k_j}$ for every $x \in X^1_{k_1} \cap X^2_{k_2} \cap \cdots \cap X^n_{k_n}\}$.

Proposition 3. Let us assume that $\bar{X}$ is full-dimensional and convex. Ridge function u is monotone in $\operatorname{int}(\bar{X})$ if and only if

$$\sum_{j\in[n]\setminus\zeta(k_1,\ldots,k_n)} r_i^j \frac{w_{k_j+1}^j - w_{k_j}^j}{b_{k_j+1}^j - b_{k_j}^j} \le 0 \tag{5}$$

for every $i \in [q]$ and for every $(k_1, k_2, \ldots, k_n) \in \Upsilon$.

The proof is given in Appendix B. We remark that even though u is continuous, monotonicity in the interior of $\bar{X}$ does not necessarily imply monotonicity in $\bar{X}$. This statement holds in $\mathbb{R}$ but not in higher dimensions.

Consider a ridge function u, which is bounded in the infinity norm. Next we address the question if this implies that all the weights w are uniformly bounded. In general the answer is no since cancelations can occur. On the positive side, we have the following result.

Theorem 4. Let us assume that $\bar{X}$ is full-dimensional and convex. Let the ridge function u with ridge vectors r and breakpoint sets b be such that $\|u\|_\infty \le M < \infty$. Then there exist ridge vectors $\widetilde{r}$, breakpoint sets $\widetilde{b}$, and weights $\widetilde{w}$ such that the ridge function $\widetilde{u}$ defined with respect to $\widetilde{r}, \widetilde{b}$, and $\widetilde{w}$ has the following properties:

1. $u(x) = \widetilde{u}(x)$ for every $x \in \bar{X}$, and
2. there exists a constant $B = B(M, r, b, \bar{X}) < \infty$ such that $|\widetilde{w}_i^j| < B$ for every i and j.

The proof is very long and technical and it is presented in Appendix C. As a consequence we obtain that the set of all ridge functions for fixed ridge vectors and breakpoint sets is sequentially compact.

Corollary 1. Let $\bar{X}$ be full-dimensional and convex. Let $\{u_k\}_k$ be a sequence of ridge functions with respect to identical fixed ridge vectors and breakpoint sets and such that $\|u_k\|_\infty \le M < \infty$ for each k. Then there exists a subsequence $\{k_l\}_l$ such that u defined by $u(x) = \lim_{l\to\infty} u_{k_l}(x)$ is a ridge function with the same ridge vectors and breakpoint sets.

Proof. Let us interpret u_k as ridge functions $\tilde{u}_k$ with respect to $\tilde{r}$ and $\tilde{b}$, which are given as in Theorem 4. It follows from Appendix C that $\tilde{r}$ and $\tilde{b}$ are identical for all k. Let $\tilde{w}(k)$ denote the weights of $\tilde{u}_k$. Then all $\tilde{w}(k)_i^j$ are uniformly bounded for every i, j and k by Theorem 4. Since the number of ridge vectors and breakpoints is the same for all $\tilde{u}_k$, it follows that there is a subsequence $\{k_l\}_l$ such that $\lim_{l\to\infty} \tilde{w}(k_l)_i^j = \hat{w}_i^j$. It is easy to see that the ridge function $\tilde{u}$ defined with respect to $\tilde{r}, \tilde{b}, \hat{w}$ has the property $\tilde{u}(x) = \lim_{l\to\infty} \tilde{u}_{k_l}(x)$ for every $x \in \bar{X}$. It is also easy to see from the proof of Theorem 4 that we can transform $\tilde{u}$ back to a ridge function u with respect to r and b for some weights w. $\square$

4 Finding Optimal Weights

In this section we assume that the ridge vectors and the breakpoint sets are fixed. We show here how to find a ridge function u, i.e. weights w, that gives the largest objective value in (D).

We need the following two additional assumptions.

Assumption B9. *The set $\{c(x,a) | (x,a) \in K\} \subset \mathbb{R}$ is compact.*

Assumption B10. *X is a convex, full-dimensional subset in $\mathbb{R}^q$ and $A \subseteq \mathbb{R}^q$.*

The last assumption comes from Corollary 1. Note that we require c to be lower semi-continuous, which in general does not imply Assumption B9. The assumption $A \subseteq \mathbb{R}^q$ is for ease of notation. It suffices to assume that $A \subseteq \mathbb{R}^s$ for some integer $1 \le s < \infty$. In the rest of the paper we assume B1–B10.

We consider functions $u : X \to \mathbb{R}$ given by (3), where w are unknowns. The problem of obtaining the ridge function of this type that gives the largest dual objective value is

$$\sup \hat{\rho} \tag{6a}$$

$$\tau(x,a)\hat{\rho} + \sum_{j=1}^{n}\sum_{i=1}^{m_j} w_i^j \left[H_i^j(r^j s(x,a)) - H_i^j(r^j x)\right] \le c(x,a) \qquad \text{for every } (x,a) \in K \tag{6b}$$

$$\hat{\rho} \in \mathbb{R}, w \text{ unrestricted}\,. \tag{6c}$$

For every $x \in X$ and $j \in [n]$ let $l_j(x) \in [m_j]$ be defined as $b^j_{l_j(x)} \le r^j x < b^j_{l_j(x)+1}$ and for every $(x,a) \in K, j \in [n]$ let $t_j(x,a) = l_j(s(x,a))$. Then it is easy to see that (6) can be rewritten as the following semi-infinite linear program, which we denote by (DW).

$$\begin{aligned}
&\sup \hat{\rho} \\
&\tau(x,a)\hat{\rho} + \sum_{j=1}^{n}\left[H^j_{t_j(x,a)}(r^j s(x,a)) w^j_{t_j(x,a)} + H^j_{t_j(x,a)+1}(r^j s(x,a)) w^j_{t_j(x,a)+1}\right] \\
&\qquad - \sum_{j=1}^{n}\left[H^j_{l_j(x)}(r^j x) w^j_{l_j(x)} + H^j_{l_j(x)+1}(r^j x) w^j_{l_j(x)+1}\right] \le c(x,a) \qquad \text{for every } (x,a) \in K \\
&\hat{\rho} \in \mathbb{R}, w \text{ unrestricted}
\end{aligned} \tag{7}$$

The primal problem (PW) of (DW) is

$$\inf \sum_{(x,a)\in T} c(x,a) z_{x,a} \tag{8a}$$

$$\sum_{(x,a)\in T} \tau(x,a) z_{x,a} = 1 \tag{8b}$$

$$\sum_{\substack{(x,a)\in T \\ t_j(x,a)\in\{i,i-1\}}} H_i^j(r^j s(x,a)) z_{x,a} - \sum_{\substack{(x,a)\in T \\ l_j(x)\in\{i,i-1\}}} H_i^j(r^j x) z_{x,a} = 0 \qquad j \in [n]\,, i \in [m_j] \tag{8c}$$

$$z \ge 0, \operatorname{supp}(z) = T, |T| < \infty\,. \tag{8d}$$

(8b) correspond to the $\hat{\rho}$ variable and (8c) correspond to the w variables.

The following theorem yields solvability and strong duality.

Theorem 5. (PW) and (DW) are solvable and there is no duality gap.

In order to prove this result, we need first to state an auxiliary result. Let

$$J = \{(c(x,a), \tau(x,a), \{H_i^j(r^j s(x,a)) - H_i^j(r^j x)\}_{j\in[n], i\in[m_j]}) \in \mathbb{R}^{2+\sum_{j=1}^{n} m_j} | (x,a) \in K\}\,.$$

This set corresponds to the right-hand side in (DW) and all the constraint coefficient functions. We denote by M^c the conic hull of J.

The next result is a generalization of a result from Glashoff and Gustafson (1983, pp. 71). The proof is similar to the one in this book and it is therefore omitted.

Lemma 1. Suppose that J is compact and (DW) has a Slater point, i.e. there exist $\hat{\rho}, w$ such that (7) are strict inequalities for every $(x,a) \in K$. Then the moment cone M^c is closed.

Proof of Theorem 5. We use the following result from Glashoff and Gustafson (1983, pp. 79). If (PW) is consistent and has a finite value, and the moment cone M^c is closed, then there is no duality gap, and (PW) is solvable.

We first argue that M^c is closed. By Lemma 1 it suffices to show that J is compact and that there is a Slater point. The latter is easy by considering for example $\hat{\rho} = -1, w = 0$. To establish compactness, note that by Assumption B1, τ is continuous, and clearly the hat functions are continuous. From this it follows that the projection of J to all the coordinates but the first one is compact. Recall also that K is compact by Assumption B7. By Assumption B9, the projection to the first coordinate is compact. We conclude that J is compact.

Next we show that (PW) is consistent. Let us pick two states $\bar{x} \in X, \tilde{x} \in X, \bar{x} \neq \tilde{x}$. By using Assumption B8 twice, first with $x' = \bar{x}, S = \{\tilde{x}\}$ and then with $x' = \tilde{x}, S = \{\bar{x}\}$, it follows that there exists a sequence T of state-action pairs $(x_1, a_1), (x_2, a_2), \ldots, (x_{\bar{N}}, a_{\bar{N}})$ and $(x_{\bar{N}+1}, a_{\bar{N}+1}), (x_{\bar{N}+2}, a_{\bar{N}+2}), \ldots, (x_{\hat{N}}, a_{\hat{N}}), \hat{N} \leq 2N$ in K such that $x_1 = x_{\hat{N}} = \bar{x}, x_{\bar{N}} = \tilde{x}$ and $s(x_i, a_i) = x_{i+1}$ for $i = 1, 2, \ldots, \hat{N} - 1$. We can assume that all these states are different, except x_1 and $x_{\hat{N}}$ (otherwise we replace these states with a 'cyclic' subset having this property). Now we defined a feasible solution z to (PW) as follows. For every $i = 1, 2, \ldots, \hat{N} - 1$ let $z(x_i, a_i) = 1/\sum_{j=1}^{\hat{N}-1} \tau(x_j, a_j)$ and z is 0 otherwise. This z clearly has finite support T and clearly by definition it satisfies (8b). Note that for every $j \in [n], i \in [m_j]$ we have

$$\{(x, a) \in T : b_i^j \leq r^j s(x, a) \leq b_{i+1}^j\} = \{(x, a) \in T : b_i^j \leq r^j x \leq b_{i+1}^j\}\,.$$

This property together with z having equal value on T it shows (8c).

(DW) is clearly consistent and therefore by weak duality (PW) has a finite value. This shows that there is no duality gap and that (PW) is solvable.

Next we establish solvability of (DW). We first argue that there exists a constant $\hat{M} < \infty$ such that $\|u\|_\infty \leq \hat{M}$ for every u that is feasible to (DW). Let $\tilde{x} \in X$ be fixed and $x \in X, x \neq \tilde{x}$ be arbitrary. Let $(x_1, a_1), (x_2, a_2), \ldots, (x_{\hat{N}}, a_{\hat{N}})$ be as above with $x = \bar{x}$. If u is feasible to (DW), then $u + \theta$ is feasible as well for any $\theta \in \mathbb{R}$. Therefore without loss of generality we assume that $u(\tilde{x}) = 0$. After applying (6b) or equivalently (2) for $(x_i, a_i), i = 1, 2, \ldots, \bar{N} - 1$ and summing all the inequalities we obtain

$$u(x) \leq \sum_{i=1}^{\bar{N}-1} c(x_i, a_i) - \rho \sum_{i=1}^{\bar{N}-1} \tau(x_i, a_i) + u(\tilde{x}) \leq C + \sup(\mathrm{DW}) \cdot \Gamma\,, \tag{9}$$

where C and Γ are as in Assumption B8 and $\sup(DW)$ denotes the optimal value of (DW). By summing (6b) for $(x_i, a_i), i = \bar{N}, \bar{N} + 1, \ldots, \hat{N} - 1$ we similarly obtain

$$0 = u(\tilde{x}) \leq \sum_{i=\bar{N}}^{\hat{N}-1} c(x_i, a_i) - \rho \sum_{i=\bar{N}}^{\hat{N}-1} \tau(x_i, a_i) + u(x) \leq C + \sup(\mathrm{DW}) \cdot \Gamma + u(x)\,. \tag{10}$$

Thus we can take $\hat{M} = C + \sup(\mathrm{DW}) \cdot \Gamma$.

There exists a sequence $\{(u_k, \rho_k)\}_k$ feasible to (DW) such that $\lim_{k\to\infty} \rho_k = \min(\mathrm{PW})$, where $\min(\mathrm{PW})$ is the optimal value of (PW). By Corollary 1 there is a subsequence $\{k_l\}_l$ such that $u(x) = \lim_{l\to\infty} u_{k_l}(x)$ is again a ridge function for some weights w. Since all u_{k_l} have the same ridge vectors and breakpoint sets (they are feasible to the same linear program (7)), it follows that u also has the same ridge vectors and breakpoint sets. By definition all u_k satisfy (6b) and therefore by taking the limit operator on both sides of (6b) we obtain that $(u, \min(\mathrm{PW}))$ is feasible to (6) with respect to weights w. Clearly then $(u, \min(\mathrm{PW}))$ is an optimal solution to (DW). $\square$

From Theorem 5 it follows that $\min(PW) = \max(DW) \leq \max(D) = \min(P)$. In addition it follows that there exists a constant $B < \infty$ such that without loss of generality we can impose $-B\mathbf{1} \leq w \leq B\mathbf{1}$, where $\mathbf{1}$ is the vector of appropriate dimension consisting of all 1s. The proof of Theorem 4 also gives a constructive way to compute B a priori.

Strong duality also gives the following complementary slackness result.

Theorem 6 (Complementary slackness)**.** If $(\hat{\rho}^*, w^*)$ is optimal to (DW) and z^* is optimal to (PW), then

$$c(x, a) = \tau(x, a)\hat{\rho}^* + \sum_{j=1}^{n} \sum_{i=1}^{m_j} (w^*)_i^j \left[H_i^j(r^j s(x, a)) - H_i^j(r^j x)\right] \tag{11}$$

for every $(x,a) \in \operatorname{supp}(z^*)$.

On the other hand, if $(\hat\rho^*, w^*)$ is feasible to (DW), z^* is feasible to (PW), and they satisfy (11), then $(\hat\rho^*, w^*)$ is optimal to (DW) and z^* is optimal to (PW).

Next we discuss the separation problem. Given $\hat\rho, w$, we want to find the most violated constraint (7) or assert that none exists.

The separation problem reads

$$\Phi(\hat\rho, w) = \min_{(x,a)\in K} \left(c(x,a) - \hat\rho\tau(x,a) - \sum_{j=1}^{n}\sum_{i=1}^{m_j} w_i^j \left[H_i^j(r^j s(x,a)) - H_i^j(r^j x) \right] \right) . \tag{12}$$

Since K is compact, the above minimum is always attained (because c is lower semi-continuous). Furthermore, Φ is concave.

We next briefly argue that the piecewise linear functions in (12) can easily be modeled with additional binary and continuous variables.

Let $\bar\mu, 0 \le \bar\mu \le \mathbf{1}$ and $\tilde\mu, 0 \le \tilde\mu \le \mathbf{1}$ be auxiliary continuous variables such that for some i, k and each j

$$r^j x = \bar\mu_i^j b_i^j + (1-\bar\mu_i^j) b_{i+1}^j \tag{13}$$

$$r^j(s(x,a)) = \tilde\mu_k^j b_k^j + (1-\tilde\mu_k^j) b_{k+1}^j \,. \tag{14}$$

Then

$$\sum_{j=1}^{n} \left[H^j_{l_j(x)}(r^j x) w^j_{l_j(x)} + H^j_{l_j(x)+1}(r^j x) w^j_{l_j(x)+1} \right] = \sum_{j=1}^{n}\sum_{i=1}^{m_j} w_i^j \bar\mu_i^j \text{ , and}$$

$$\sum_{j=1}^{n} \left[H^j_{t_j(x,a)}(r^j s(x,a)) w^j_{t_j(x,a)} + H^j_{t_j(x,a)+1}(r^j s(x,a)) w^j_{t_j(x,a)+1} \right] = \sum_{j=1}^{n}\sum_{i=1}^{m_j} w_i^j \tilde\mu_i^j \,.$$

From (7) we obtain

$$\sum_{j=1}^{n}\sum_{i=1}^{m_j} w_i^j \left[H_i^j(r^j s(x,a)) - H_i^j(r^j x) \right] = \sum_{j=1}^{n}\sum_{i=1}^{m_j} w_i^j (\tilde\mu_i^j - \bar\mu_i^j) \,.$$

The requirements (13) and (14) can easily be modeled with linear constraints (except for $s(x,a)$), see e.g. Nemhauser and Wolsey (1988) or Croxton *et al.* (2003). If c, τ, s, and the requirement $(x,a) \in K$ can be modeled by linear constraints, then the entire separation problem is a mixed integer programming problem.

If we want to impose u to be convex, where X is convex and full-dimensional, we proceed as follows. For each $j \in [n]$ we first compute $\phi_j^* = \max_{x\in X} r^j x$ and $\varphi_j^* = \min_{x\in X} r^j x$. Next we compute $\underline{m}_j$ and $\overline{m}_j$ by considering ϕ_j^*, φ_j^*, respectively. Finally, we form the set R and we add constraints (4). Note that we have only polynomially many constraints.

Imposing monotonicity again under the assumption that X is convex and full-dimensional is more difficult since it is not easy to determine Υ and the corresponding ζ. Since the number of constraints (5) can be exponential, we need to dynamically generate them. Given w, we next show how to either find a violated (5) or assert that the incumbent u is monotone.

We proceed in two steps. We first solve the following optimization problem.

$$\omega = \max u(x_1) - u(x_2) \tag{15a}$$

$$x_1 \le x_2 \tag{15b}$$

$$r^j x_1 \text{ and } r^j x_2 \text{ in the same interval for every } j \in [n] \tag{15c}$$

$$x_1 \in X, x_2 \in X \tag{15d}$$

As above it can easily be argued that the objective function and constraints (15c) can be modeled as linear mixed integer type constraints. A closer look at the proof of Proposition 3 reveals that $\omega \le 0$ if and only if u is monotone. Thus if $\omega \le 0$, the current u is monotone and we quit the dynamic constraint generation procedure. If x_1^*, x_2^* are optimal to (15), then let $k_j, j \in [n]$ be defined as $b^j_{k_j} \le r^j x_1^* \le b^j_{k_j+1}$ and $b^j_{k_j} \le r^j x_2^* \le b^j_{k_j+1}$. From (15c) it follows that such $k_j, j \in [n]$ exist. Note also that $(k_1, k_2, \dots, k_n) \in \Upsilon$.

In the second step we find $\zeta(k_1, \dots, k_n)$. For each $j \in [n]$ we solve

$$\begin{aligned} \gamma^j &= \max r^j x \\ b^{\bar{j}}_{k_{\bar{j}}} &\le r^{\bar{j}} x \le b^{\bar{j}}_{k_{\bar{j}}+1} \qquad \bar{j} \in [n] \\ & x \in X\,. \end{aligned}$$

Now it is clear that $\zeta(k_1, \dots, k_n) = \{j \in [n] : \gamma^j > b^{\bar{j}}_{k_{\bar{j}}}\}$. Finally, one of the constraints (5) is violated for an $i \in [q]$.

4.1 Solving (DW)

In this section we present a row generation algorithm for solving (DW) within a given tolerance. The algorithm is based on the row generation algorithm for general semi-infinite algorithms, see e.g. Goberna and López (1998). Let $\delta > 0$ be given. The restricted master problem (RMP) consists of the problem formed by considering only a finite subset of states $(x, a) \in K$ together with the bounds $-B\mathbf{1} \le w \le B\mathbf{1}$ from Section 3. Note that the resulting problem is a finite linear program. The algorithm finds a maximum objective value solution to (DW) that satisfies all constraints (7) within δ. We require that the separation problem is solved within the tolerance $\delta/2$.

The algorithm is given in Algorithm 1. Given $\hat{\rho}$ and w, for any $(x, a) \in K$ we denote by $\pi_{\hat{\rho},w}(x, a)$ the right-hand side of (7) minus the left-hand side. The initial set S that gives a finite objective value to the RMP can be selected as follows. Let us pick any two states $\bar{x} \in X, \tilde{x} \in X, \bar{x} \ne \tilde{x}$. By using Assumption B8 twice, we obtain a set S of state-action pairs $(x_1, a_1), (x_2, a_2), \dots, (x_{\bar{N}}, a_{\bar{N}})$ and $(x_{\bar{N}+1}, a_{\bar{N}+1}), (x_{\bar{N}+2}, a_{\bar{N}+2}), \dots, (x_{\hat{N}}, a_{\hat{N}})$, $\hat{N} \le 2N$ in K such that $x_1 = x_{\hat{N}} = \bar{x}, x_{\bar{N}} = \tilde{x}$ and $s(x_i, a_i) = x_{i+1}$ for $i = 1, 2, \dots, \hat{N} - 1$. It is easy to see that this S has the desired property.

1: Start with a finite subset S of state-action pairs such that the RMP over S has a finite objective value.
2: **loop**
3: Solve the RMP over S and let $\hat{\rho}, w$ be an optimal solution.
4: Find a finite subset $\bar{S}$ consisting of $(x, a) \in K$ such that $\pi_{\hat{\rho},w}(x, a) \le \Phi(\hat{\rho}, w) + \frac{\delta}{2}$.
5: **if** $\pi_{\hat{\rho},w}(x, a) \ge -\delta$ for every $(x, a) \in \bar{S}$ **then**
6: Stop; $\hat{\rho}, w$ is the solution.
7: **else**
8: $S = S \cup \bar{S}$.
9: **end if**
10: **end loop**

Algorithm 1: The row generation algorithm for solving (DW)

The key step in this algorithm is step 4. If this separation problem can be solved efficiently, then many iterations of the algorithm can be performed. In view of the discussion in the previous section, the optimization problem corresponding to this step is either a linear or a nonlinear mixed integer program. If c and τ are piecewise linear, the resulting problem is a linear mixed integer program (for details see Adelman and Klabjan (2005a)). The computational experiments in Adelman and Klabjan (2005a) demonstrate that in the case of the inventory routing problem, this step can be performed efficiently.

The next theorem establishes that this is a finite algorithm and it produces the desired solution.

Proposition 4. Algorithm 1 stops in a finite number of steps. The resulting solution has the property $\pi_{\hat{\rho},w}(x,a) \geq -\delta$ for all $(x,a) \in K$, i.e. it is optimal with δ *dual feasibility tolerance.* In addition, if Θ is the obtained objective value upon the algorithm completion, then

$$\max(DW) \leq \Theta \leq \max(DW) + \delta(1 + \max(DW))\,.$$

Proof. In iteration k we denote the optimal dual solution to the RMP by $\hat{\rho}_k, w(k)$ and by $\bar{S}_k$ the set produced in step 4. We also denote by S_k the set of all selected state-action pairs at the end of iteration k, which is S at the end of iteration k. The RMP is a finite linear program with an optimal solution. Let z^k be the corresponding primal solution, which is clearly feasible to (PW).

Let us assume that the algorithm does not stop. Since $w \leq B\mathbf{1}$ are always in the RMP, we have that $\hat{\rho}_k, w(k)$ are included in a compact set. Therefore there is a convergent subsequence. For ease of notation we assume that the entire sequence is convergent. Let $\hat{\rho}^* = \lim_{k\to\infty} \hat{\rho}_k$ and $w^* = \lim_{k\to\infty} w(k)$. We claim that $\Phi(\hat{\rho}^*, w^*) \geq -\delta/2$.

Note that Φ is concave and $\infty > \Phi > -\infty$. Therefore Φ is continuous. We first show $\pi_{\hat{\rho}^*,w^*}(x_k, a_k) \geq 0$ for every k, where $(x_k, a_k) \in \bar{S}_k$. Let k be fixed and select $(x_k, a_k) \in \bar{S}_k \subseteq S_k$. For every $p > k$ we have $(x_k, a_k) \in S_p$ since we do not remove constraints from the RMP. Therefore $\pi_{\hat{\rho}_p,w_p}(x_k, a_k) \geq 0$ for every $p \geq k$. By taking the limit as p goes to infinity we obtain $\pi_{\hat{\rho}^*,w^*}(x_k, a_k) \geq 0$.

Now we show $\Phi(\hat{\rho}^*, w^*) \geq -\delta/2$. For any k we have

$$\begin{aligned}
\Phi(\hat{\rho}^*, w^*) &= \Phi(\hat{\rho}_k, w(k)) + (\Phi(\hat{\rho}^*, w^*) - \Phi(\hat{\rho}_k, w(k))) \\
&\geq \pi_{\hat{\rho}_k,w(k)}(x_k, a_k) - \frac{\delta}{2} + (\Phi(\hat{\rho}^*, w^*) - \Phi(\hat{\rho}_k, w(k))) \\
&\geq \left(\pi_{\hat{\rho}_k,w(k)}(x_k, a_k) - \pi_{\hat{\rho}^*,w^*}(x_k, a_k)\right) - \frac{\delta}{2} + (\Phi(\hat{\rho}^*, w^*) - \Phi(\hat{\rho}_k, w(k)))\,.
\end{aligned}$$

Here we have used step 4 of the algorithm. In the limit as k goes to infinity, the last two terms go to 0 since Φ is continuous. Since τ, c, and the hat functions, which correspond to the coefficients in (7), are bounded (Assumptions B9, B1, and B7), it follows that the first two terms go to 0 as well. We conclude that $\Phi(\hat{\rho}^*, w^*) \geq -\delta/2$.

Since $-\delta < -\delta/2$, there exists a k such that $\Phi(\hat{\rho}_k, w(k)) \geq -\delta$. These yields that $\pi_{\hat{\rho}_k,w(k)}(x_k, a_k) \geq \Phi(\hat{\rho}_k, w(k)) \geq -\delta$, which shows that the algorithm stops at the latest in iteration k. To show the second claim we only need to observe that $\pi_{\hat{\rho}_k,w(k)}(x, a) \geq \Phi(\hat{\rho}_k, w(k)) \geq -\delta$ for every $(x, a) \in K$. Therefore $\hat{\rho}_k, w(k)$ is a solution with δ dual feasibility tolerance.

It remains to show the bound on the objective value. Let $o = \max(DW)$. Let us denote by (PPW),(PDW) the same problem as (PW),(DW) except that the cost is defined as $c(x,a)+\delta$ for every $(x,a) \in K$, respectively. Note that $(\hat{\rho}_k, w(k)), z^k$ is the optimal primal/dual pair of (PPW),(PDW). By definition

$$\Theta = \sum_{(x,a)\in \mathrm{supp}(z^k)} (c(x,a) + \delta) z^k_{x,a}\,.$$

Since z^k is feasible to (PW) it follows that

$$o \leq \sum_{(x,a)\in \mathrm{supp}(z^k)} c(x,a) z^k_{x,a} \leq \sum_{(x,a)\in \mathrm{supp}(z^k)} (c(x,a) + \delta) z^k_{x,a} = \Theta\,.$$

If z^* is an optimal solution to (PW), then it is feasible to (PPW) and therefore

$$\Theta = \sum_{(x,a)\in \mathrm{supp}(z^k)} (c(x,a) + \delta) z^k_{x,a} \leq \sum_{(x,a)\in \mathrm{supp}(z^*)} (c(x,a) + \delta) z^*_{x,a} = o + \delta \sum_{(x,a)\in \mathrm{supp}(z^*)} z^*_{x,a}\,. \tag{16}$$

By Assumption B3 we have $1 \leq c(x,a) + \tau(x,a)$ for every $(x,a) \in K$ and therefore

$$\sum_{(x,a)\in \mathrm{supp}(z^*)} z^*_{x,a} \leq \sum_{(x,a)\in \mathrm{supp}(z^*)} c(x,a) z^*_{x,a} + \sum_{(x,a)\in \mathrm{supp}(z^*)} \tau(x,a) z^*_{x,a} = o + 1\,. \tag{17}$$

Combining together (16) and (17) yields the final statement. □

5 The Algorithm

Here we present and analyze the overall algorithm. The algorithm in each iteration first finds the optimal weights for fixed ridge vectors and breakpoints, i.e. we apply Algorithm 1 to solve (DW). In the second phase we generate new breakpoints or ridge vectors.

For any function $z : K \to \mathbb{R}$ with $\mathrm{supp}(z) < \infty$, $r \in \mathbb{R}^q$, and $\bar{b} = (\bar{b}_1, \bar{b}_2, \bar{b}_3)$ with $\bar{b}_1 < \bar{b}_2 < \bar{b}_3$ let

$$g(z, r, \bar{b}) = \Big| \sum_{\substack{(x,a)\in \mathrm{supp}(z) \\ \bar{b}_1 \le rx \le \bar{b}_3}} H_{\bar{b}}(rx) z_{x,a} - \sum_{\substack{(x,a)\in \mathrm{supp}(z) \\ \bar{b}_1 \le rs(x,a) \le \bar{b}_3}} H_{\bar{b}}(rs(x,a)) z_{x,a} \Big| \ .$$

Here we denote by $H_{\bar{b}}$ the hat function on breakpoints $\bar{b}_1, \bar{b}_2, \bar{b}_3$. This function measures the flow inbalance given in (8c).

By Theorem 3 it suffices to consider ridge vectors with the infinity norm of 1. Therefore for every $x \in X$ and every ridge vector r with $\|r\|_\infty \le 1$ we have by Cauchy-Schwartz

$$| \, rx \, | \le \|r\|_2 \cdot \|x\|_2 \le \sqrt{q} \cdot \mathrm{diam}(X) \ ,$$

where $\mathrm{diam}(X) < \infty$ is the diameter of X. Therefore we can select $\Omega = \sqrt{q} \cdot \mathrm{diam}(X)$ (see Section 3 for the role of Ω). We denote by $-\Omega \le \bar{b} \le \Omega$ the requirement $-(\Omega+1) \le \bar{b}_1 < \bar{b}_2 < \bar{b}_3 \le (\Omega+1)$ and $-\Omega \le \bar{b}_2 \le \Omega$. (See the discussion in Section 3 for the reason to allow $\bar{b}_1 < -\Omega$ and $\bar{b}_3 > \Omega$.)

The overall algorithm is given in Algorithm 2. Given a set of ridge vectors and the corresponding breakpoints, the algorithm in step 5 finds the optimal weights w (up to a certain precision). Next we find either new breakpoints or new ridge vectors that violate the flow balance constraints (8c). These are then added to the set of the existing ridge vectors and breakpoints. We assume that the problem of finding the most violated flow balance constraint can be solved within an arbitrary precision.

1: Let $\{\epsilon_i\}_{i=0}^\infty, \{\delta_i\}_{i=0}^\infty$ be such that $\epsilon_i > 0, \delta_i > 0$ for every i and $\lim_{i\to\infty} \epsilon_i = \lim_{i\to\infty} \delta_i = 0$.
2: $s = 0$
3: Let S_0 be a set of initial ridge vectors and the corresponding breakpoints, which can be an empty set.
4: **loop**
5: Run Algorithm 1 with precision δ_s. Let $\hat{\rho}, w$ be the resulting optimal solution with δ_s dual feasibility tolerance.
6: Let z^s be the corresponding primal solution to (PW) with the same objective value. We denote the resulting ridge function as u^s.
7: Find a finite subset $\bar{S}_s$ of $(r, \bar{b}), \|r\|_\infty \le 1, -\Omega \le \bar{b} \le \Omega$ such that

$$g(z^s, r, \bar{b}) \ge \max_{\substack{\|\tilde{r}\|_\infty \le 1 \\ -\Omega \le \tilde{b} \le \Omega}} g(z^s, \tilde{r}, \tilde{b}) - \epsilon_s$$

for every $(r, \bar{b}) \in \bar{S}_s$.
8: For every $(r, \bar{b}) \in \bar{S}_s$ either add $(r, \bar{b})$ to S_s if r is not already in S_s, or add $\bar{b}$ to an already present ridge vector r in S_s.
9: $s = s + 1$
10: **end loop**

Algorithm 2: Optimal solution algorithm

In step 7 we assume that $\max_{\substack{\|\tilde{r}\|_\infty \le 1 \\ -\Omega \le \tilde{b} \le \Omega}} g(z, \tilde{r}, \tilde{b})$ is attained. The following lemma whose proof is given in Appendix D establishes this.

Lemma 2. If $\alpha = \sup_{\substack{\|\tilde{r}\|_\infty \le 1 \\ -\Omega \le \tilde{b} \le \Omega}} g(z, \tilde{r}, \tilde{b})$, then there exists $r, \|r\|_\infty \le 1$ and $b, -\Omega \le b \le \Omega$ such that $\alpha = g(z, r, b)$.

Similarly to Algorithm 1, the separation step 7 is very important. Typically it yields a nonlinear optimization problem, which needs to be solved approximately. In Adelman and Klabjan (2005a) in the context of inventory routing this step is solved heuristically yet it guarantees to separate the current solution. Even though $\max_{\substack{\|\tilde{r}\|_\infty \leq 1 \\ -\Omega \leq \tilde{b} \leq \Omega}} g(z^s, \tilde{r}, \tilde{b})$ is not computed within an arbitrary precision, the computational experiments show convergence after a relatively low number of iterations.

In order to prove the convergence of the algorithm we need to state some definitions and statements from measure theory.

Definition 1. A sequence $\{\mu_m\}_m$ of measures on K converges *weakly* to a measure μ on K if

$$\int_K u \,\mathrm{d}\mu_m \xrightarrow{m\to\infty} \int_K u \,\mathrm{d}\mu \tag{18}$$

for every $u \in \mathbb{C}(K)$.

The *total variation norm* of a finite signed measure μ on K is defined as $\|\mu\|_{\mathrm{TV}} = |\mu|(K)$, where $|\mu|$ is the total variation of μ. If μ is a positive measure, then $\|\mu\|_{\mathrm{TV}} = \mu(K)$. Note that since K is compact, the set of all continuous functions is equivalent to the set of all continuous bounded functions, which are considered in a typical definition of weak convergence.

Every solution z to (PW) induces a measure γ^z on K defined by

$$\gamma^z(\bar{K}) = \sum_{(x,a)\in \bar{K}\cap \mathrm{supp}(z)} z_{x,a}$$

for every Borel subset $\bar{K} \subseteq K$. By definition, for every Borel subset $\bar{K}$ and every Borel measurable function g we have

$$\int_{\bar{K}} g \,\mathrm{d}\gamma^z = \sum_{(x,a)\in \bar{K}\cap \mathrm{supp}(z)} g(x,a) z_{x,a}\,.$$

Every vector $\bar{b}$ with $-\Omega \leq \bar{b} \leq \Omega$ can be viewed as a vector in $[-(\Omega+1), \Omega+1]^3$, which is a compact set. Therefore every sequence of $\bar{b}$'s contains a convergence subsequence with the limit $\tilde{b}$. However $\tilde{b}$ might not have the property $\tilde{b}_1 < \tilde{b}_2 < \tilde{b}_3$. It can happen that for example $\tilde{b}_1 = \tilde{b}_2 = \tilde{b}_3$ and therefore we cannot define the hat function on this triplet. The next theorem states that the algorithm produces an optimal solution to (P) and (D) as long as the generated breakpoints $\bar{b}$ "stay apart" for at least one convergent subsequence.

Theorem 7. Suppose that there exists a subsequence $\{\bar{b}^{s_k}\}_k$ with $(r^{s_k}, \bar{b}^{s_k}) \in \bar{S}_{s_k}$ such that $\lim_{k\to\infty} \bar{b}^{s_k} = \tilde{b}$ and $\tilde{b}_1 < \tilde{b}_2 < \tilde{b}_3$. Then there exists a subsequence s_i such that $\gamma^{z^{s_i}}$ converge weakly to a measure μ, which is optimal to (P). If $u(x) = \limsup_i u^{s_i}$, then u is optimal to (D).

Observe that the corresponding $\{r^{s_k}\}_k$ converge on a subsequence since their infinity norm is less than or equal to 1. We can restate the condition of the theorem as follows. There is at least one accumulation point $(\tilde{r}, \tilde{b})$ in $\cup_{s=1}^\infty \bar{S}_s$ with the property $\tilde{b}_1 < \tilde{b}_2 < \tilde{b}_3$. Note that $\cup_{s=1}^\infty \bar{S}_s$ is the set of all generated breakpoints and ridge vectors. Before proving this theorem, we need to state a few results. The following corollary is a special case of Corollary 3.6 in Klabjan and Adelman (2006), which in turn relies on the Prohorov's theorem (see e.g. Billingsley (1968)).

Proposition 5. Let Γ be a family of nonnegative measures on a Borel compact space K. Assume that there exists a constant $M < \infty$ such that $0 < \|\mu\|_{\mathrm{TV}} < M$. Then for each sequence $\{\mu_n\}_n$ in Γ there is a subsequence $\{\mu_m\}_m$ and a measure μ such that $\{\mu_m\}_m$ converges weakly to μ.

This proposition holds without requiring that K be compact but additional restrictions need to be imposed. The following result is shown in Hernández-Lerma and Lasserre (1999), page 225.

Lemma 3. Let $g \geq 0$ be lower semi-continuous. If $\{\mu_i\}_i$ converges weakly to a measure μ on K, then

$$\int_K g \,\mathrm{d}\mu \leq \liminf_i \int_K g \,\mathrm{d}\mu_i\,.$$

Definition 2. A sequence of functions $\{f_s\}_s$ on a set $\hat{X}$ *continuously converge* to a function f if for every convergent sequence $\{x^s\}_s$ in $\hat{X}$ with $\lim_{s\to\infty} x^s = x$ we have $\lim_{s\to\infty} f_s(x^s) = f(x)$.

The following proposition is proven in Langen (1981).

Proposition 6. Let $\{\mu^s\}_s$ be a sequence of finite positive measures on $\hat{X}$ that converge weakly to a finite positive measure μ. Let $\{f_s\}_s, \{p_s\}_s$ be measurable functions from $\hat{X}$ to $\mathbb{R}$ such that $|f_s| \leq p_s$ for every s. Let $\{f_s\}_s$ continuously converge to f and let $\{p_s\}_s$ continuously converge to p. In addition, let $\lim_{s\to\infty} \int_{\hat{X}} p_s \, \mathrm{d}\mu_s = \int_{\hat{X}} p \, \mathrm{d}\mu < \infty$. Then

$$\lim_{s\to\infty} \int_{\hat{X}} f_s \, \mathrm{d}\mu_s = \int_{\hat{X}} f \, \mathrm{d}\mu \, .$$

Proof of Theorem 7. Let Θ_s be the optimal objective value at step 5. The optimal objective value of the underlying (PW) and (DW) is denoted by o_s. By Proposition 4 we have

$$o_s \leq \Theta_s \leq o_s + \delta_s(1 + o_s) \, . \tag{19}$$

We first show how to find the subsequence s_i. Let $(r^{s_k}, \bar{b}^{s_k}) \in \bar{S}_{s_k}$ and $\tilde{b}$ be as in the statement of the theorem. There is a subsequence in $\{r^{s_k}\}_k$ that converges to $\tilde{r}, \|\tilde{r}\| \leq 1$. For ease of notation we assume that the entire sequence has this property, i.e. all of the subsequences that follow are subsequences of $\{s_k\}_k$. There is a $\Pi > 0$ such that $\delta_s < \Pi$ for every s. From (19) we obtain

$$0 \leq \Theta_s \leq o_s + \Pi(1 + o_s) \leq \max(D) + \Pi(1 + \max(D)) < \infty \tag{20}$$

and therefore the sequence Θ_s is bounded. There is a subsequence $\{\bar{s}_l\}_l$ such that $\lim_{l\to\infty} \Theta_{\bar{s}_l} = \varphi$.

Next we use Proposition 5 to find a weakly convergent subsequence in $\{\gamma^{z^{\bar{s}_l}}\}_l$. First we note that by (20) we have

$$\int_K c \, \mathrm{d}\gamma^{z^s} \leq \int_K (c + \delta_s) \, \mathrm{d}\gamma^{z^s} = \Theta_s \leq \max(D) + \Pi(1 + \max(D))$$

for every s. From $1 \leq \tau + c \leq \tau + c + \delta_s$ we have

$$\|\gamma^{z^s}\|_{\mathrm{TV}} = \int_K \mathrm{d}\gamma^{z^s} \leq \int_K \tau \, \mathrm{d}\gamma^{z^s} + \int_K (c + \delta_s) \, \mathrm{d}\gamma^{z^s} = 1 + \Theta_s \leq 1 + \max(D) + \Pi(1 + \max(D))$$

for every s, where we have again used (20). This shows that we can apply Proposition 5 to find a weakly convergent subsequence $\{\gamma^{z^{s_l}}\}_l$ in $\{\gamma^{z^{\bar{s}_l}}\}_l$. Let μ be the corresponding limit.

Function u defined by $u(x) = \limsup_l u^{s_l}$ is well defined. To see this we apply (9) and (10) with $c + \delta_s$ instead of c. By using (20) we obtain

$$\|u^{s_l}\|_\infty \leq C + \delta_{s_l} \cdot N + \Theta_{s_l} \cdot \Gamma \leq C + \Pi \cdot N + \Gamma(\max(D) + \Pi(1 + \max(D))) \, ,$$

which shows that u is well defined.

We next show that (φ, u) is feasible to (D), μ is feasible to (P), and that they have equal objective values. For simplicity we assume that $\{s_l\}_l$ corresponds to the entire sequence.

We first show that (φ, u) is feasible to (D). Note that u^s is optimal with δ_s dual feasibility tolerance. Therefore u^s satisfies (6b) with $c + \delta_s$, which reads

$$\Theta_s \tau(x, a) + u^s(x) \leq c(x, a) + \delta_s + u^s(s(x, a))$$

for every $(x, a) \in K$. After applying the lim sup operator on both sides and using $\lim_{s\to\infty} \delta_s = 0$ we obtain

$$\varphi \tau(x, a) + u(x) - u(s(x, a)) \leq c(x, a)$$

for every $(x, a) \in K$. This shows the statement.

Next we argue that $\varphi = \int_K c \, \mathrm{d}\mu$, which shows that (φ, u) and μ have the same objective values. By Lemma 3 and Assumption B2 we obtain

$$\int_K c \, \mathrm{d}\mu \leq \liminf_s \int_K c \, \mathrm{d}\gamma^{z^s} \leq \liminf_s \Theta_s = \varphi \,, \tag{21}$$

which shows the claim.

It remains to show that μ is feasible to (P), i.e. it satisfies (1b) and (1c). Since z^s satisfies (8b) and τ is continuous, we have

$$1 = \int_K \tau \, \mathrm{d}\gamma^{z^s} \xrightarrow{s \to \infty} \int_K \tau \, \mathrm{d}\mu = 1 \,,$$

which shows (1b).

The proof for (1c) is more complicated. Let $L_1 : \mathbb{M}(K) \to \mathbb{M}(X)$ be the linear operator defined by

$$(L_1 \phi)(B) = \phi((B \times A) \cap K) - \phi(\{(x, a) \in K : s(x, a) \in B\})$$

for every $B \in \mathcal{B}(X)$. It is easy to see that this is a measure in X. The following is shown in Klabjan and Adelman (2006). If $\{\alpha_s\}_s$ converge weakly to a measure α on K, then for every $v \in \mathbb{C}(X)$ we have

$$\int_X v \, \mathrm{d}L_1 \alpha_s \xrightarrow{s \to \infty} \int_X v \, \mathrm{d}L_1 \alpha \,. \tag{22}$$

We argue that $L_1 \mu$ is a finite signed measure. By definition of L_1 it suffices to show that μ is a finite measure. This follows from $1 \leq \tau + c$, $\int_K \tau \, \mathrm{d}\mu = 1$, and (21).

We need to show that $L_1 \mu = 0$. By the Riesz representation theorem for finite signed measures (see e.g. Rudin (1986)) this is equivalent to showing that $\int_X v \, \mathrm{d}L_1 \mu = 0$ for every $v \in \mathbb{C}(X)$. By Theorem 3 it suffices to show that $\int_X u \, \mathrm{d}L_1 \mu = 0$ for every ridge function u. If we denote $h^r_{\bar{b}} = H_{\bar{b}}(< r, \cdot >)$, then it suffices to show $\int_X h^r_{\bar{b}} \, \mathrm{d}L_1 \mu = 0$ for every ridge vector r and three-tuple $\bar{b} = (\bar{b}_1, \bar{b}_2, \bar{b}_3), \bar{b}_1 < \bar{b}_2 < \bar{b}_3$. Note that by definition of L_1 we have $g(z, r, \bar{b}) = | \int_X h^r_{\bar{b}} \, \mathrm{d}L_1 \gamma^z |$.

From (22) and since $h^r_{\bar{b}}$ are continuous it follows

$$\int_X h^r_{\bar{b}} \, \mathrm{d}L_1 \gamma^{z^s} \xrightarrow{s \to \infty} \int_X h^r_{\bar{b}} \, \mathrm{d}L_1 \mu \tag{23}$$

for every $r, \bar{b}$.

From step 7 of the algorithm we obtain

$$| \int_X h^r_{\bar{b}} \, \mathrm{d}L_1 \gamma^{z^s} | \leq \max_{\substack{\|r\|_\infty \leq 1 \\ -\Omega \leq \bar{b} \leq \Omega}} g(z^s, r, \bar{b}) \leq | \int_X h^{r^s}_{\bar{b}^s} \, \mathrm{d}L_1 \gamma^{z^s} | + \epsilon_s \,. \tag{24}$$

Let us fix s and consider $(r^s, \bar{b}^s) \in S_s$. Then from the algorithm (see (8c)) it follows that $\int_X h^{r^s}_{\bar{b}^s} \, \mathrm{d}L_1 \gamma^{z^p} = 0$ for every $p > s$. By considering the limit as p goes to infinity and from (23) we get

$$\int_X h^{r^s}_{\bar{b}^s} \, \mathrm{d}L_1 \mu = 0 \,. \tag{25}$$

The following is elementary to establish. Since $r^s \to \tilde{r}$ and $\bar{b}^s \to \tilde{b}$, and $\tilde{b}_1 < \tilde{b}_2 < \tilde{b}_3$, then $h^{r^s}_{\bar{b}^s}$ continuously converge to $h^{\tilde{r}}_{\tilde{b}}$. This is the only place where we need the assumption $\tilde{b}_1 < \tilde{b}_2 < \tilde{b}_3$. Note that if this is violated, then $h^{r^s}_{\bar{b}^s}$ converge pointwise to a function however the convergence is not continuous.

By considering $p_s = 1$ for every s and $p = 1$ we can use Proposition 6 with $\{\gamma^{z^s}\}_s$ and μ. We first obtain

$$\int_K h^{r^s}_{\bar{b}^s}(x) \, \mathrm{d}\gamma^{z^s}(x, a) \xrightarrow{s \to \infty} \int_K h^{\tilde{r}}_{\tilde{b}}(x) \, \mathrm{d}\mu(x, a) \tag{26}$$

by taking $f_s(x,a) = h^{r^s}_{\bar{b}^s}(x)$. Next we obtain

$$\int_K h^{r^s}_{\bar{b}^s}(s(x,a))\,\mathrm{d}\gamma^{z^s}(x,a) \xrightarrow{s\to\infty} \int_K h^{\tilde{r}}_{\tilde{b}}(s(x,a))\,\mathrm{d}\mu(x,a) \tag{27}$$

by considering $f_s(x,a) = h^{r^s}_{\bar{b}^s}(s(x,a))$ in Proposition 6. Since s is continuous by Assumption B6, it is easy to see that $h^{r^s}_{\bar{b}^s}(s(x,a))$ is continuously convergent. After subtracting (26) from (27) we obtain

$$|\int_X h^{r^s}_{\bar{b}^s}\,\mathrm{d}L_1\gamma^{z^s}| \xrightarrow{s\to\infty} |\int_X h^{\tilde{r}}_{\tilde{b}}\,\mathrm{d}L_1\mu|\,. \tag{28}$$

By the Lebesgue convergence theorem with the dominant function being the constant 1 it follows

$$\int_X h^{r^s}_{\bar{b}^s}\,\mathrm{d}L_1\mu \xrightarrow{s\to\infty} \int_X h^{\tilde{r}}_{\tilde{b}}\,\mathrm{d}L_1\mu\,.$$

Combining with (25) we obtain $\int_X h^{\tilde{r}}_{\tilde{b}}\,\mathrm{d}L_1\mu = 0$. This equality and (28) yield

$$\lim_{s\to\infty} |\int_X h^{r^s}_{\bar{b}^s}\,\mathrm{d}L_1\gamma^{z^s}| = 0\,.$$

In turn from (24) it follows $\lim_{s\to\infty} |\int_X h^{r}_{\bar{b}}\,\mathrm{d}L_1\gamma^{z^s}| = 0$. Finally, from (23) we get $\int_X h^{r}_{\bar{b}}\,\mathrm{d}L_1\mu = 0$, which shows the entire theorem. □

In order to guarantee convergence and in line with the stated assumption in Theorem 7, it is desirable in step 7 of Algorithm 2 to find new breakpoints that are far apart from the already generated breakpoints. In Adelman and Klabjan (2005a) this is achieved by using the strategy of adding a new breakpoint if this breakpoint is at least a given parameter away from the previously generated breakpoints. If such a breakpoint cannot be found, then a new ridge vector r is found.

Acknowledgement

We are indebted to two anonymous referees whose feedback has significantly improved the exposition and readability of this manuscript. We are particulary thankful to one of the referees that pointed out a flaw in an early version of the document. The second author gratefully acknowledges the financial support of the University of Chicago, Graduate School of Business.

References

Adelman, D. and Klabjan, D. (2005a). Computing optimal policies in generalized joint replenishment. Working paper, draft, University of Chicago, Graduate School of Business.

Adelman, D. and Klabjan, D. (2005b). Duality and existence of optimal policies in generalized joint replenishment. *Mathematics of Operations Reserach*, **30**, 28–50.

Bellman, R. E. and Dreyfus, S. E. (1959). Functional approximations and dynamic programming. *Mathematical Tables and Other Aids to Computation*, **13**, 247–251.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley and Sons.

Bojanov, B., Hakopian, H., and Sahakian, A. (1983). *Spline Functions and Multivariate Interpolations*. Kluwer Academic Publishers.

Chen, V. C., Ruppert, D., and Shoemaker, C. A. (1999). Applying experimental design and regression splines to high-dimensional continuous-state stochastic dynamic programming. *Operations Research*, **47**, 38–53.

Cheney, W. and Light, W. (1999). *A Course in Approximation Theory*. Brooks/Cole Publishing Company.

Croxton, K., Gendron, B., and Magnanti, T. (2003). A comparison of mixed-integer programming models for nonconvex piecewise linear cost minimization problems. *Management Science*, **49**, 1268–1273.

Davis, P. (1975). *Interpolation and Approximation*. Dover Publications.

de Farias, D. and Van Roy, B. (2004). On constraint sampling for the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research*, **29**, 462–478.

de Farias, D. P. and Van Roy, B. (2003). The linear programming approach to approximate dynamic programming. *Operations Research*, **51**, 850–856.

D'Epenoux, F. (1960). Sur un problème de production et de stockage dans l'aléatoire. *Revue Francaise de Recherche Opérationnelle*, **14**, 3–16.

Ghellinck, G. D. (1960). Les problèmes de décisions séquentielles. *Cahiers du Centre d'Etudes de Recherche Opérationnelle*, **2**, 161–179.

Glashoff, K. and Gustafson, S.-A. (1983). *Linear Optimization and Approximation*. Springer-Verlag.

Goberna, M. and López, M. (1998). *Linear Semi-Infinite Optimization*. John Wiley & Sons.

Hernández-Lerma, O. and Lasserre, J. (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer-Verlag.

Hernández-Lerma, O. and Lasserre, J. (1998a). Approximation schemes for infinite linear programs. *SIAM J. Optim.*, **8**, 973–988.

Hernández-Lerma, O. and Lasserre, J. (1998b). Linear programming approximations for Markov control processes in metric spaces. *Acta Applicandae Mathematicae*, **51**, 123–139.

Hernández-Lerma, O. and Lasserre, J. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer-Verlag.

Johnson, S. A., Stedinger, J. R., Shoemaker, C. A., Li, Y., and Tejada-Guibert, J. A. (1993). Numerical solution of continuous-state dynamic programs using linear and spline interpolation. *Operations Research*, **41**, 484–500.

Klabjan, D. and Adelman, D. (2006). Existence of optimal policies for semi-Markov decision processes using duality for infinite linear programming. *SIAM Journal on Control and Optimization*, **44**, 2104–4122.

Langen, H.-J. (1981). Convergence of dynamic programming models. *Mathematics of Operations Research*, **6**, 493–512.

Manne, A. (1960). Linear programming and sequential decisions. *Management Science*, **6**, 259–267.

Nemhauser, G. and Wolsey, L. (1988). *Integer and combinatorial optimization*. John Wiley & Sons.

Powell, W. B. and Topaloglu, H. (2003). Stochastic programming in transportation and logistics. In A. Ruszczynski and A. Shapiro, editors, *Handbook in Operations Research and Management Science: Stochastic Programming*. Elsevier, Amsterdam.

Rudin, W. (1986). *Real and Complex Analysis*. McGraw-Hill.

Schweitzer, P. and Seidmann, A. (1985). Generalized polynomial approximations in Markovian decision processes. *Journal of Mathematical Analysis and Applications*, **110**, 568–582.

Tadic, V., Meyn, S., and Tempo, R. (2006). Randomized algorithms for semi-infinite programming problems. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*. Springer Verlag.

Wang, R.-H. (1994). *Multivariate Spline Functions and Their Applications*. Kluwer Academic Publishers.

A Proof of Proposition 2

Let $\{r^j\}_{j\in[n]} = \cup_{p=1}^{v} R_p$, where if $r^k \in R_p, r^l \in R_p$, then r^k and r^l are linearly dependent and if $r^k \in R_p, r^l \in R_{\bar{p}}, p \neq \bar{p}$, then r^k and r^l are linearly independent. This corresponds to the partition of the ridge vectors into sets of pairwise linearly dependent vectors. For each $p \in [v]$ we arbitrarily select and fix a vector $r^{p^*} \in R_p$, which represents the set R_p. For each $r^j \in R_p$ we denote $r^j = \mu_j r^{p*}$. Using this new notation we get

$$S_i^j = \{(l,k) \in R|\, \frac{b_k^l}{\mu_l} = \frac{b_i^j}{\mu_j}\}\,. \tag{29}$$

Observe also that if $(l,k) \in S_i^j$ for $j \in R_p$, then $\lambda_{ki}^{lj} = \mu_l$.

Now we introduce new variables $y_p = r^{p^*}x$ and we rewrite the ridge function as

$$u(x) = \sum_{p=1}^{v}\sum_{j\in R_p}\sum_{i=1}^{m_j} w_i^j H_i^j(r^j x) = \sum_{p=1}^{v}\sum_{j\in R_p}\sum_{i=1}^{m_j} w_i^j H_i^j(\mu_j y_p) = \sum_{p=1}^{v} u_p(y_p)\,,$$

where $u_p(y_p) = \sum_{j\in R_p}\sum_{i=1}^{m_j} w_i^j H_i^j(\mu_j y_p)$. Note that u_p is now a function of a single variable. We further break u_p into $u_p(y_p) = \sum_{j\in R_p} u_p^j(\mu_j y_p)$, where $u_p^j(y_p) = \sum_{i=1}^{m_j} w_i^j H_i^j(\mu_j y_p)$.

It is obvious that u_p^j are piecewise linear functions with breakpoints b_i^j/μ_j and the corresponding value w_i^j. We conclude that all of these functions have slopes $\mu_j \cdot (w_i^j - w_{i-1}^j)/(b_i^j - b_{i-1}^j)$. Note that u_p^j might not be convex. Since u_p^j are piecewise linear and the sum of piecewise linear functions is piecewise linear, so is u_p. The breakpoints of u_p are $\{b_i^j/\mu_j\}_{j\in[n],i\in[m_j]}$, which is the union of all the breakpoints of functions u_p^j.

We now proceed in two steps. First we show that u is convex if and only if u_p are convex for all $p \in [v]$. Then we show that u_p is convex if and only if (4) holds.

Claim 1. The ridge function u is convex if and only if u_p are convex for all $p \in [v]$.

If u_p's are convex for every p, then clearly u is convex (the sum of convex functions is a convex function and a superposition of a convex and a linear function is convex). Let now u be convex.

Let $p \in [v]$ be fixed and let us select an arbitrary $(j,i) \in R$ with $j \in R_p$. Let also $(l,k) \in S_i^j$ be arbitrary, which implies that $l \in R_p$. We show that

$$u_p(\lambda y_p + (1-\lambda)\bar{y}_p) \leq \lambda u_p(y_p) + (1-\lambda)u_p(\bar{y}_p) \tag{30}$$

for any $\lambda, 0 \leq \lambda \leq 1$ and a particular choice of y_p and $\bar{y}_p$ satisfying

$$b_{k-1}^l \leq \mu_l y_p < b_k^l < \mu_l \bar{y}_p \leq b_{k+1}^l\,. \tag{31}$$

This suffices to prove the statement since u_p is piecewise linear with breakpoints b_i^j/μ_j and (l,k) being arbitrary with $l \in R_p$.

By definition of R, there exists $\bar{x} \in \text{int}(\bar{X})$ with $r^j\bar{x} = b_i^j$. Here we use the assumption that $\bar{X}$ is full-dimensional and the facts $\min\{r^j x : x \in \bar{X}\} = \inf\{r^j x : x \in \text{int}(\bar{X})\}, \max\{r^j x : x \in \bar{X}\} = \sup\{r^j x : x \in \text{int}(\bar{X})\}$. Let $\bar{S} = \{l :$ there exists k such that $(l,k) \in S_i^j\}$ (for simplicity we write $\bar{S}$ instead of $\bar{S}_i^j$). Note that $\bar{S} \subseteq R_p$.

We show that there exists $x \in \text{int}(\bar{X})$ with $r^j x = b_i^j$ and for every q not in $\bar{S}$ there exists $o(q)$ with $b_{o(q)}^q < r^q x < b_{o(q)+1}^q$. Let $q \notin \bar{S}$ be arbitrary. If r^q and r^j are linearly dependent, then since $q \notin \bar{S}$ any

x with $r^j x = b^j_i$ has the property $r^q x \neq b^q_{\bar{i}}$ for every $\bar{i}$. Let now r^q and r^j be linearly independent. Then as in the proof of Claim 5 (see Appendix C) we argue that there exists $x_q \in \text{int}(\bar{X})$ and $\bar{o}_q \in [m_q]$ with $b^q_{\bar{o}_q} < r^q x_q < b^q_{\bar{o}_q+1}$. Then again as in the proof of Claim 5 we can combine the various x_q into a vector x with the desired property. Note that for every $q \in \bar{S}$, $r^q x$ is one of the breakpoints.

Let

$$\begin{aligned}\epsilon &= \min_{q \notin \bar{S}} \{b^q_{o(q)+1} - r^q x, r^q x - b^q_{o(q)}\} > 0\,,\\ \delta_1 &= \min_{(l,k)\in S^j_i} \{b^l_{k+1} - b^l_k, b^l_k - b^l_{k-1}\}\,.\end{aligned}$$

Consider $f : \bar{X} \to \mathbb{R}^n$ defined by

$$f(w) = \begin{cases} r^p w & p \notin \bar{S} \\ r^p x & p \in \bar{S}\,, \end{cases}$$

which is clearly continuous. Therefore there exists δ_2 such that if $\|y - x\|_\infty \leq \delta_2$, then $\|f(y) - f(x)\|_\infty \leq \epsilon$, which is equivalent to $b^q_{o(q)} \leq r^q y \leq b^q_{o(q)+1}$ for every $q \notin \bar{S}$. In addition there exists δ_3 such that if $\|x - y\|_\infty \leq \delta_3$, then $y \in \text{int}(\bar{X})$.

Consider now $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Since by assumption $r^j \neq 0$, there exists $s \in [j]$ with $r^j_s \neq 0$. We construct $\alpha_1 = x - \delta e_s, \alpha_2 = x + \delta e_s$. Since $\delta \leq \delta_3$ we have $\alpha_1 \in \text{int}(\bar{X}), \alpha_2 \in \text{int}(\bar{X})$. Due to $\delta \leq \delta_2$ we have $b^q_{o(q)} \leq r^q \alpha_1 \leq b^q_{o(q)+1}$ and $b^q_{o(q)} \leq r^q \alpha_2 \leq b^q_{o(q)+1}$ for every $q \notin \bar{S}$. Without loss of generality we assume that $r^j_s > 0$ (otherwise we switch α_1 and α_2). Since $\delta \leq \delta_1$ we have $b^l_{k-1} \leq r^l \alpha_1 < b^l_k < r^l \alpha_2 \leq b^l_{k+1}$ for every $(l,k) \in S^j_i$.

We select $y_p = r^{p^*} \alpha_1$ and $\bar{y}_p = r^{p^*} \alpha_2$, which implies that $y_p = r^l \alpha_1/\mu_l, \bar{y}_p = r^l \alpha_2/\mu_l$. We conclude that (31) is satisfied. Let us pick $0 < \lambda < 1$. Then since u is convex we have

$$u(\alpha_3) \leq \lambda u(\alpha_1) + (1 - \lambda) u(\alpha_2)\,, \tag{32}$$

where $\alpha_3 = \lambda \alpha_1 + (1 - \lambda)\alpha_2$. We have $b^q_{o(q)} \leq r^q \alpha_3 \leq b^q_{o(q)+1}$ for every $q \notin \bar{S}$ and for $(l,k) \in S^j_i$ the value $r^l \alpha_3$ lies in either $(r^l \alpha_1, b^l_k]$ or $[b^l_k, r^l \alpha_2)$.

It now follows that (32) is equivalent to (30). (For $q \notin \bar{S}$ the terms cancel out since the function is linear on $[b^q_{o(q)}, b^q_{o(q)+1}]$.) This completes the proof of the claim.

Claim 2. Given a fixed $p \in [v]$ the function u_p is convex if and only if (4) holds.

Consider a set of n piecewise linear functions and their sum $\bar{u}$. Clearly $\bar{u}$ is piecewise linear and its breakpoints are the union of all individual breakpoints. Image that this union is sorted and let us discuss the underlying slopes and convexity of $\bar{u}$.

Consider first a breakpoint of $\bar{u}$, which is not equal to any other breakpoint of the n piecewise linear functions. Let us assume that this is the breakpoint of the j'th function. Function $\bar{u}$ is convex around this breakpoint if and only if the slope to the left is less than or equal to the slope on the right. Only the j'th function changes slope at this breakpoint and therefore convexity in this case is equivalent to requiring that the j'th function is convex around this breakpoint.

Now assume that a breakpoint b equals to several breakpoints, each one being a breakpoint of a different function. Let us denote by C the set of these functions. The functions outside of C do not change slope at b. The slope of $\bar{u}$ to the left and right equals to the sum of the left and right slopes of the functions in C, respectively. Therefore $\bar{u}$ is convex around b if and only if the sum over C of the slopes on the left is less than or equal to the sum over C of the slopes on the right. Note that the right and left slopes of the functions not in C cancel out and therefore it suffices to consider only the slopes of the functions in C.

Let us now interpret this discussion in terms of functions u_p and u^j_p. The breakpoints in question are $\{b^j_i/\mu_j\}_{j\in[n], i\in[m_j]}$. If several breakpoints collide, then this corresponds precisely to the definition of S^j_i, see (29). The slopes of u^j_p are $\mu_j \cdot (w^j_i - w^j_{i-1})/(b^j_i - b^j_{i-1})$. The statement now easily follows.

B Proof of Proposition 3

We first assume that u is monotone in $\mathrm{int}(\bar{X})$ and we show that (5) holds. Let $i \in [q]$ and let $(k_1, k_2, \dots, k_n) \in \Upsilon$ be arbitrary. Let $Y = X^1_{k_1} \cap X^2_{k_2} \cap \cdots \cap X^n_{k_n}$. Then there exists $x \in Y$ such that $b^j_{k_j} < r^j x < b^j_{k_j+1}$ for every $j \in [n] \setminus \zeta(k_1, \dots, k_n)$. To see this consider the following vectors. By definition of Υ and $\zeta(k_1, \dots, k_n)$ for each $j \in [n] \setminus \zeta(k_1, \dots, k_n)$ there exists $x_j \in \mathrm{int}(\bar{X})$ such that $b^j_{k_j} < r^j x_j < b^j_{k_j+1}$. Since $\bar{X}$ is convex, we can take

$$x = \frac{1}{|\,[n] \setminus \zeta(k_1, \dots, k_n)|} \sum_{j \in [n] \setminus \zeta(k_1, \dots, k_n)} x_j \, .$$

Let $\epsilon = \min_{j \in [n] \setminus \zeta(k_1, \dots, k_n)} \{b^j_{k_j+1} - r^j x, r^j x - b^j_{k_j}\} > 0$. Consider now $f : \bar{X} \to \mathbb{R}^n$ defined by $f(y) = (r^1 y, \dots, r^n y)$. Clearly f is continuous in the infinity norm. Therefore there exists $\delta_1 > 0$ such that for every $y \in Y$ with $\|y - x\|_\infty \leq \delta$ it follows that $\|f(y) - f(x)\|_\infty \leq \epsilon$. By definition of ϵ and f for any such y we have $b^j_{k_j} \leq r^j y \leq b^j_{k_j+1}$ for every $j \in [n] \setminus \zeta(k_1, \dots, k_n)$. For every $j \in \zeta(k_1, \dots, k_n)$ we have $r^j y = b^j_{k_j}$. Since $x \in \mathrm{int}(\bar{X})$, there exists $\delta_2 > 0$ such that if $\|y - x\|_\infty \leq \delta_2$, then $y \in \mathrm{int}(\bar{X})$. Let $\delta = \min\{\delta_1, \delta_2\}$.

Consider now $y = x + \delta e_i$. Clearly $\|y - x\|_\infty \leq \delta_1$ and $\|y - x\|_\infty \leq \delta_2$. These implies that $y \in Y$. For every $z_1 \in Y, z_2 \in Y$ we have

$$u(z_1) - u(z_2) = \sum_{j \in [n] \setminus \zeta(k_1, \dots, k_n)} \frac{w^j_{k_j} r^j (z_1 - z_2) - w^j_{k_j+1} r^j (z_1 - z_2)}{b^j_{k_j+1} - b^j_{k_j}} \, . \tag{33}$$

The terms corresponding to $j \in \zeta(k_1, \dots, k_n)$ cancel out by definition of $\zeta(k_1, \dots, k_n)$.

Since $x \leq y$ it follows that $u(x) \leq u(y)$. In turn we have

$$0 \leq u(y) - u(x) = -\delta \sum_{j \in [n] \setminus \zeta(k_1, \dots, k_n)} r^j_i \frac{w^j_{k_j+1} - w^j_{k_j}}{b^j_{k_j+1} - b^j_{k_j}} \, ,$$

which follows from (33) and it shows (5).

Now let us prove the reverse statement. We assume that (5) holds and we need to show monotonicity. Let $x \in \mathrm{int}(\bar{X}), y \in \mathrm{int}(\bar{X})$ with $x \leq y$. We need to show that $u(x) \leq u(y)$.

Let $x \in X^1_{k_1} \cap X^2_{k_2} \cap \cdots \cap X^n_{k_n}$ and $y \in X^1_{\bar{k}_1} \cap X^2_{\bar{k}_2} \cap \cdots \cap X^n_{\bar{k}_n}$ for $(k_1, \dots, k_n) \in \Upsilon, (\bar{k}_1, \dots, \bar{k}_n) \in \Upsilon$. We define $l(x, y) = \sum_{j=1}^n |k_j - \bar{k}_j|$ and we prove the claim by induction on $l(x, y)$, which can have only a finite number of integer values.

Let us first consider $l(x, y) = 0$, i.e. $(k_1, \dots, k_n) = (\bar{k}_1, \dots, \bar{k}_n)$. For each $i \in [q]$ we multiply (5) by $y_i - x_i \geq 0$ and we sum all the resulting inequalities. If we add constant terms corresponding to $j \in \zeta(k_1, \dots, k_n)$, it is easy to see that we obtain $u(x) \leq u(y)$.

Let $w \geq 1$ be an integer. We now assume that for every $\bar{x} \leq \bar{y}$ with $l(\bar{x}, \bar{y}) \leq w - 1$ we have $u(\bar{x}) \leq u(\bar{y})$. Consider $x \leq y$ with $l(x, y) = w$. Since $w \geq 1$, there exists $t \in [n]$ such that $|k_t - \bar{k}_t| \geq 1$.

Consider first $r^t x < r^t y$ and $r^t y > b^t_{\bar{k}_t}$. Let us define

$$\lambda = \left(1 + \frac{r^t y - b^t_{\bar{k}_t}}{r^t y - r^t x}\right) / 2 \, .$$

It is easy to see that $0 < \frac{r^t y - b^t_{\bar{k}_t}}{r^t y - r^t x} < \lambda < 1$. Consider $z = \lambda x + (1 - \lambda) y$, which is in $\mathrm{int}(\bar{X})$ by convexity of $\bar{X}$. Since $x \leq y$, it follows that $x \leq z \leq y$. By definition of λ it follows that $r^t x < r^t z < b^t_{\bar{k}_t}$. Therefore $l(z, y) \leq w - 1$ and $l(x, z) \leq w - 1$. By induction hypothesis we have $u(x) \leq u(z) \leq u(y)$. In turn this implies $u(y) - u(x) = u(y) - u(z) + u(z) - u(x) \geq 0$, which shows the statement.

If $r^t x < r^t y$ and $r^t y = b^t_{\bar{k}_t}$, then we can use any $0 < \lambda < 1$ and proceed as above. The case $r^t x > r^t y$ is identical by reversing the role of x and y. This completes the proof.

C Proof of Theorem 4

The following lemma will be used on several occasions.

Lemma 4. If $A \in \mathbb{R}^{k\times k}$ is an invertible matrix, then the set $\{x \in \mathbb{R}^k : l \leq Ax \leq u\}$ is bounded for every given vectors $l \in \mathbb{R}^k, u \in \mathbb{R}^k$.

Proof. We have

$$\{x \in \mathbb{R}^k : l \leq Ax \leq u\} = \{A^{-1}y : l \leq y \leq u, y \in \mathbb{R}^k\} .$$

The function $f : \mathbb{R}^k \to \mathbb{R}^k$ defined by $f(y) = A^{-1}y$ is continuous and the set $\{y \in \mathbb{R}^k : l \leq y \leq u\}$ is compact. Since in $\mathbb{R}^k$ the image of a compact space of a continuous function is compact, the claim follows. □

In addition, we need the following lemma, which can be easily proven by induction and the recursive definition of the determinant.

Lemma 5. Let $b^j : \mathbb{R}^q \to \mathbb{R}^n$ be differentiable functions for $j \in [n]$. If $h : \mathbb{R}^q \to \mathbb{R}$ is defined as $h(x) = \det\left[b^1, b^2, \ldots, b^n\right]$, then for $i = 1, \ldots, q$ we have

$$\frac{\partial h}{\partial x_i} = \det\left[\frac{\partial b^1}{\partial x_i}, b^2, \ldots, b^n\right] + \det\left[b^1, \frac{\partial b^2}{\partial x_i}, b^3, \ldots, b^n\right] + \cdots + \det\left[b^1, b^2, \ldots, \frac{\partial b^n}{\partial x_i}\right] .$$

Proof of Theorem 4. We have already observed that if two ridge vectors are linearly dependent, then one of them can be eliminated. Therefore without loss of generality we assume that all ridge vectors are pairwise linearly independent. For simplicity we assume that $\overline{m}_j = m_j$ and $\underline{m}_j = 1$ (otherwise we simply remove all the remaining breakpoints without affecting u).

We assume that $\|u\|_\infty \leq M < \infty$. For each $j \in [n]$ and $i \in [m_j]$ we define

$$v_i^j = \frac{w_{i+1}^j - w_i^j}{b_{i+1}^j - b_i^j} .$$

Let $\bar{X}_i^j = \{x \in \operatorname{int}(\bar{X})\,|b_i^j < r^j x < b_{i+1}^j\}$ and $\Psi = \{(k_1, \ldots, k_n)|\bar{X}_{k_1}^1 \cap \bar{X}_{k_2}^2 \cap \cdots \cap \bar{X}_{k_n}^n \neq \emptyset\}$. Note that since $\bar{X}$ is full-dimensional, its interior is non empty. The proof is broken down into several claims.

Claim 3. There exists a ridge function $\hat{u}$ with weights $\hat{w}$ such that $u(x) = \hat{u}(x)$ for every $x \in X$ and $|\hat{w}_{k_j}^j| < M$ for every $j = 1, \ldots, n$ and a fixed $k_j \in [m_j]$.

For any numbers $g_1, \ldots, g_n$ such that $\sum_{j=1}^n g_j = 0$ we have $u(x) = \sum_{j=1}^n \left[\sum_{i=1}^{m_j} w_i^j H_i^j(r^j x) - g_j\right]$. Let us now fix $y \in X$ and let $b_{k_j}^j \leq r^j y < b_{k_j+1}^j$. We now set $g_j = w_{k_j}^j r^j y$ for $j = 1, \ldots, n-1$ and $g_n = -\sum_{j=1}^{n-1} w_{k_j}^j r^j y$. Let us denote $h^j(t) = \sum_{i=1}^{m_j} w_i^j H_i^j(t) - g_j$. Then h^j is a piecewise linear function. For $j = 1, \ldots, n-1$ it is easy to see that h^j can be defined as having a breakpoint at $r^j y$ and in addition $h^j(r^j y) = 0$. Therefore we can assume that the weight corresponding to $r^j y$ is equal to zero. For $j = n$ we can still define a new breakpoint at $r^n y$ but however in this case $h^n(r^n y) = u(y)$. Therefore we can assume that the weight corresponding to this new breakpoint is $u(y)$. Since $|u(y)| \leq M$, this completes the proof of Claim 3.

Claim 4. Let $\bar{j} \in [n]$ be fixed and let $\{r^j : j \in S(\bar{j})\}$ be a maximal set of linearly independent vectors with $\bar{j} \in S(\bar{j})$. Every $r^j, j \notin S(\bar{j})$ is a linear combination of vectors from $S(\bar{j})$ and therefore there exist λs such that $r^j = \sum_{k\in S(\bar{j})} \lambda_k^j r^k$. Then for every $(k_1, \ldots, k_n) \in \Psi$ there exists a constant $B_1 = B_1(M, r, b, \bar{X})$ such that

$$|v_{k_{\bar{j}}}^{\bar{j}} + \sum_{l\in[n]\setminus S(\bar{j})} \lambda_{\bar{j}}^l v_{k_l}^l| \leq B_1 .$$

Note that $|S(\bar{j})| \geq 2$ for every $\bar{j} \in [n]$. For simplicity we assume that $\bar{j} = 1$ and that $S(1) = \{1, 2, \dots, s\}$. It means that $r^1, \dots, r^s$ are linearly independent and for $j = s+1, \dots, n$ we have $r^j = \sum_{k=1}^{s} \lambda_k^j r^k$. For any $x \in \bar{X}$ such that $b_{k_j}^j \leq r^j \leq b_{k_j+1}^j$ for each $j \in [n]$ we have

$$u(x) = \sum_{j=1}^{n} \frac{w_{k_j}^j b_{k_j+1}^j - w_{k_j+1}^j b_{k_j}^j}{b_{k_j+1}^j - b_{k_j}^j} + \sum_{j=1}^{s} \left[\left(v_{k_j}^j + \sum_{l=s+1}^{n} \lambda_j^l v_{k_l}^l \right) r^j x \right] = a + b(x) \, .$$

Here a denotes the first summation and $b(x)$ the second summation.

Let us now fix $x \in \bar{X}_{k_1}^1 \cap \bar{X}_{k_2}^2 \cap \cdots \cap \bar{X}_{k_n}^n$, which by definition is non empty. Since $r^1, \dots, r^s$ are linearly independent, there exist $\alpha_1, \dots, \alpha_s$ such that the matrix A defined by $\{r_{\alpha_g}^j\}_{\substack{j=1,\dots,s \\ g=1,\dots,s}}$ is nonsingular. Since $\bar{X}_{k_1}^1 \cap \bar{X}_{k_2}^2 \cap \cdots \cap \bar{X}_{k_n}^n$ is open, there exists $\epsilon > 0$ such that $y_g = x + \epsilon e_{\alpha_g} \in \bar{X}_{k_1}^1 \cap \bar{X}_{k_2}^2 \cap \cdots \cap \bar{X}_{k_n}^n$ for every $g = 1, \dots, s$. Then we have

$$u(y_g) = a + b(y_g) = a + b(x) + b(y_g - x) = a + b(x) + \epsilon b(e_g) = u(x) + \epsilon b(e_g) \, .$$

Since $u(y_g)$ and $u(x)$ are bounded by M, it follows that $b(e_g)$ is bounded for every g, where the bounds depend on ϵ, i.e. on $\bar{X}$ and M. Now we apply Lemma 4 with $A = \{r_{\alpha_g}^j\}_{\substack{j=1,\dots,s \\ g=1,\dots,s}}$, $x_j = v_{k_j}^j + \sum_{l=s+1}^{n} \lambda_j^l v_{k_l}^l$ for every j, and l, u correspond to the underlying bounds. We obtain that $v_{k_j}^j + \sum_{l=s+1}^{n} \lambda_j^l v_{k_l}^l$ are bounded for every j. In particular it holds for $j = 1$, which shows Claim 4. Note that the resulting constant depends on $(k_1, \dots, k_n)$. However, since Ψ is finite we can take the maximum of all these values.

Claim 5. Let $\bar{j} \in [n]$ and $\bar{i} \in \{2, 3, \dots, m_j - 1\}$ be fixed. Consider $Y = \{x \in \mathrm{int}(\bar{X}) : r^{\bar{j}} x = b_{\bar{i}}^{\bar{j}}\}$. Then there exists $x \in Y$ such that for every $j \in [n]\,, j \neq \bar{j}$ there exist k_j with $b_{k_j}^j < r^j x < b_{k_j+1}^j$.

Observe that for each j we have $\inf\{r^j x : x \in \mathrm{int}(\bar{X})\} = \min\{r^j x : x \in \bar{X}\}, \sup\{r^j x : x \in \mathrm{int}(\bar{X})\} = \max\{r^j x : x \in \bar{X}\}$ and therefore Y is non empty. We prove this claim in two steps.

Suppose that $Y \subseteq \{x \in \mathrm{int}(\bar{X}) : r^j x = b_{k_j}^j\}$ for a $j, j \neq \bar{j}$ and $k_j \in [m_j]$. We show by contradiction that this is not possible. There exist $x_1, \dots, x_{q-1}$ in $\mathbb{R}^q$ such that $r^{\bar{j}}, x_1, \dots, x_{q-1}$ are an orthonormal base in $\mathbb{R}^q$. Let $x \in Y$. There exists $\epsilon > 0$ such that $x + \epsilon x_i \in \mathrm{int}(\bar{X})$ for every $i = 1, \dots, q-1$. We have $r^{\bar{j}}(x + \epsilon x_i) = b_{\bar{i}}^{\bar{j}} + \epsilon r^{\bar{j}} x_i = b_{\bar{i}}^{\bar{j}}$. This shows that $x + \epsilon x_i \in Y$ for every i. By assumption then $r^j(x + \epsilon x_i) = b_{k_j}^j$ for every i. In addition we have $r^j x = b_{k_j}^j$, which yields that $r^j x_i = 0$ for every i. Due to orthonormality we have that there exist $\beta_l, l = 1, 2, \dots, q-1$ and β such that $r^j = \sum_{l=1}^{q-1} \beta_l x_l + \beta r^{\bar{j}}$. After multiplying this equation by x_i we obtain that $\beta_i = 0$ for every i. Hence $r^j = \beta r^{\bar{j}}$. If $\beta = 0$, then $r^j = 0$, which we excluded. Hence $\beta \neq 0$, but now r^j and $r^{\bar{j}}$ are linearly dependent, which we have ruled out as well. We have a contradiction.

By the above statement, for every $j, j \neq \bar{j}$ there exists $x_j \in Y$ such that $b_{k_j}^j < r^j x_j < b_{k_j+1}^j$ for a $k_j \in [m_j]$. For simplicity we assume that $\bar{j} = 1$. We first construct an $x \in Y$ such that the property holds for $j = 2$ and $j = 3$. If $r^2 x_2 = r^2 x_3$ or $r^3 x_2 = r^3 x_3$, then we can either take $x = x_2$ or $x = x_3$. Assume now that $r^2 x_2 \neq r^2 x_3$ and $r^3 x_2 \neq r^3 x_3$. Consider $y = \lambda x_2 + (1 - \lambda) x_3$ for a λ that is to be determined. There is a finite number of λs such that $r^2 y$ and $r^3 y$ are breakpoints. Therefore there exists a $\lambda \in (0, 1)$ such that $r^2 y$ and $r^3 y$ are not breakpoints. For this particular λ we set $x = y$. Note that since $\bar{X}$ is convex the resulting y is in $\bar{X}$.

By using the same approach for all j it is easy to obtain the desired x. This completes the proof of Claim 5.

Claim 6. Let $\bar{j}, \bar{i}$ be as in Claim 5. In addition, let us select k_j for $j \in [n]\,, j \neq \bar{j}$ as in Claim 5. Then there exist $x_1 \in \text{int}(\bar{X}), x_2 \in \text{int}(\bar{X})$ such that

$$\begin{aligned} b^{\bar{j}}_{\bar{i}-1} &< r^{\bar{j}} x_1 < b^{\bar{j}}_{\bar{i}} \\ b^{\bar{j}}_{\bar{i}} &< r^{\bar{j}} x_2 < b^{\bar{j}}_{\bar{i}+1} \\ b^{j}_{k_j} &< r^{j} x_1 < b^{j}_{k_j+1} \qquad j \in [n], j \neq \bar{j} \\ b^{j}_{k_j} &< r^{j} x_2 < b^{j}_{k_j+1} \qquad j \in [n], j \neq \bar{j}\,. \end{aligned}$$

We can restate the statement of the claim as $(k_1, k_2, \ldots, k_{\bar{j}-1}, \bar{i}-1, k_{\bar{j}+1}, \ldots, k_n) \in \Psi$ and $(k_1, k_2, \ldots, k_{\bar{j}-1}, \bar{i}, k_{\bar{j}+1}, \ldots, k_n) \in \Psi$, see Figure 2.

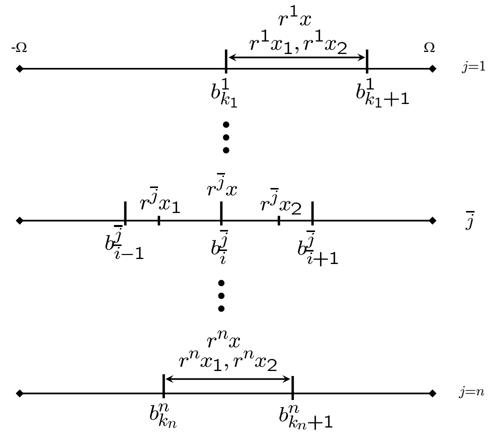

Figure 2: Statement of Claim 6

Let x be as in Claim 5. We define

$$\epsilon_1 = \frac{1}{2} \min\{ \min_{j\in[n], j\neq\bar{j}} \{b^j_{k_j+1} - r^j x\}, \min_{j\in[n], j\neq\bar{j}} \{r^j x - b^j_{k_j}\}, b^{\bar{j}}_{\bar{i}+1} - b^{\bar{j}}_{\bar{i}}, b^{\bar{j}}_{\bar{i}} - b^{\bar{j}}_{\bar{i}-1}\} > 0\,.$$

Consider $F : \mathbb{R}^q \to \mathbb{R}^n$ defined by $F(y) = (r^1 y, r^2 y, \ldots, r^n y)$. F is clearly continuous at x in the infinity norm. Therefore there exists $\delta_1 > 0$ such that if $\|y - x\|_\infty \leq \delta_1$, then $\|F(y) - F(x)\|_\infty \leq \epsilon_1$. The later requirement by definition of ϵ_1 implies $b^{\bar{j}}_{\bar{i}-1} < r^{\bar{j}} y < b^{\bar{j}}_{\bar{i}+1}$ and $b^j_{k_j} < r^j y < b^j_{k_j+1}$ for every $j \in [n], j \neq \bar{j}$.

There exists $\delta_2 > 0$ such that if $\|y - x\|_\infty \leq \delta_2$, then $y \in \text{int}(\bar{X})$. Let us pick any $z \neq 0$ such that $r^{\bar{j}} z > 0$ and we define $\delta = \min\{\delta_1, \delta_2\}/\|z\|_\infty$. Consider now $\bar{y} = x \pm \delta z$. Clearly $\|\bar{y} - x\|_\infty = \delta\|z\|_\infty = \min\{\delta_1, \delta_2\}$. Therefore $\bar{y} \in \text{int}(\bar{X})$ and $b^j_{k_j} < r^j \bar{y} < b^j_{k_j+1}$ for every $j \in [n], j \neq \bar{j}$. For $\bar{j}$ we have $b^{\bar{j}}_{\bar{i}-1} < r^{\bar{j}} \bar{y} < b^{\bar{j}}_{\bar{i}+1}$. In addition

$$\begin{aligned} r^{\bar{j}}(x + \delta x) &= b^{\bar{j}}_{\bar{i}} + \delta r^{\bar{j}} z > b^{\bar{j}}_{\bar{i}}, \\ r^{\bar{j}}(x - \delta x) &= b^{\bar{j}}_{\bar{i}} - \delta r^{\bar{j}} z < b^{\bar{j}}_{\bar{i}}\,. \end{aligned}$$

Therefore we can take $x_1 = x - \delta z$ and $x_2 = x + \delta z$, which completes the proof of Claim 6.

Claim 7. For any $\bar{j} \in [n]$, if there exists a constant $B_2 = B_2(M, r, b, \bar{X})$ such that $|w^{\bar{j}}_l| \leq B_2, |w^{\bar{j}}_k| \leq B_2$ for two fixed $l \in [m_{\bar{j}}], k \in [m_{\bar{j}}]$, then there exists a constant $B = B(M, r, b, \bar{X})$ such that $|w^{\bar{j}}_i| \leq B$ for every $i \in [m_{\bar{j}}]$.

The claim states that if two weights are bounded for each j, then all the weights are bounded. By Claim 3 we know that for each j we have a bound on a single weight and therefore this claim leaves us only with a 'single degree of unboundness' for each j.

Let us fix $\bar{j} \in [n]$ and $\bar{i} \in \{2, 3, \ldots, m_{\bar{j}} - 1\}$ as in Claim 6. Next we use Claim 4 with $(k_1, k_2, \ldots, k_{\bar{j}-1}, \bar{i} - 1, k_{\bar{j}+1}, \ldots, k_n)$ and $(k_1, k_2, \ldots, k_{\bar{j}-1}, \bar{i}, k_{\bar{j}+1}, \ldots, k_n)$. We obtain that

$$|v^{\bar{j}}_{\bar{i}-1} + \sum_{l \in [n] \setminus S(\bar{j})} \lambda^l_{\bar{j}} v^l_{k_l}| \leq B_1$$
$$|v^{\bar{j}}_{\bar{i}} + \sum_{l \in [n] \setminus S(\bar{j})} \lambda^l_{\bar{j}} v^l_{k_l}| \leq B_1 ,$$

for a constant B_1. It follows that $v^{\bar{j}}_{\bar{i}-1} - v^{\bar{j}}_{\bar{i}}$ is bounded. If we denote $a^j_i = 1/(b^j_{i+1} - b^j_i)$, then

$$v^{\bar{j}}_{\bar{i}-1} - v^{\bar{j}}_{\bar{i}} = -a^{\bar{j}}_{\bar{i}-1} w^{\bar{j}}_{\bar{i}-1} + (a^{\bar{j}}_{\bar{i}} + a^{\bar{j}}_{\bar{i}-1}) w^{\bar{j}}_{\bar{i}} - a^{\bar{j}}_{\bar{i}} w^{\bar{j}}_{\bar{i}+1}$$

is bounded.

Let $\tilde{A} \in \mathbb{R}^{m_{\bar{j}} \times (m_{\bar{j}} - 2)}$ be the matrix

$$\begin{bmatrix} -a^{\bar{j}}_1 & a^{\bar{j}}_1 + a^{\bar{j}}_2 & a^{\bar{j}}_2 & & & \\ & -a^{\bar{j}}_2 & a^{\bar{j}}_2 + a^{\bar{j}}_3 & -a^{\bar{j}}_3 & & \\ & \ddots & \ddots & \ddots & & \\ & & -a^{\bar{j}}_{m_{\bar{j}}-2} & a^{\bar{j}}_{m_{\bar{j}}-2} + a^{\bar{j}}_{m_{\bar{j}}-1} & -a^{\bar{j}}_{m_{\bar{j}}-1} & \end{bmatrix}$$

that is triagonal. All the remaining entries are 0. It follows that $\tilde{A}(w^{\bar{j}}_1, \ldots, w^{\bar{j}}_{m_{\bar{j}}})$ are bounded. After removing any two columns from $\tilde{A}$, which corresponds to having two bounded columns, the resulting matrix A is diagonally dominant and therefore invertible. By Lemma 4 we get that all the remaining $w^{\bar{j}}$ are bounded, which completes the proof of Claim 7.

For each $j \in [n]$ let us fix a $\tilde{k}_j \in [m_j]$ such that $w^j_{\tilde{k}_j} \neq 0$. We define $\tilde{w}^j_i = w^j_i / w^j_{\tilde{k}_j}$. By combining Claim 7 and Claim 3 we obtain that we can assume that there exists a constant $B_3 = B_3(M, r, b, \bar{X})$ such that $|\tilde{w}^j_i| \leq B_3$ for every j, i. We can rewrite the ridge function u as

$$u(x) = \sum_{j=1}^{n} w^j_{\tilde{k}_j} \sum_{i=1}^{m_j} \tilde{w}^j_i H^j_i(r^j x) = \sum_{j=1}^{n} w^j_{\tilde{k}_j} f_j(r^j x) ,$$

where f_j is piecewise linear and bounded. Let $Q = \{j \in [n] : m_j = 2\}$. Note that for any $j \in Q$ by our assumption this means that r^j maps $\bar{X}$ into a single subinterval. In other words, for $j \in Q$ we have that f_j is a linear function. The rest of the proof considers two cases; Q is empty and Q is non empty.

Claim 8. If $Q = \emptyset$, then there exist $x^1, \ldots, x^n \in \bar{X}$ such that the matrix with the entries $\{f_j(r^j x^i)\}_{j \in [n], i \in [n]}$ is nonsingular.

In order to prove this claim, we first show that there exist $s_1, s_2, \ldots, s_n \in [q]$ and $\alpha_t = (k^t_1, k^t_2, \ldots, k^t_n) \in \Psi$ for $t = 1, 2, \ldots, n$ such that the matrix D defined by $\{v^j_{k^t_j} r^j_{s_j}\}_{j \in [n], t \in [n]}$ is nonsingular. We construct these vectors, i.e. indices, iteratively with respect to j.

Let us assume that we have $j - 1$ linearly independent vectors, i.e. we have $s_1, s_2, \ldots, s_n \in [q]$ and $\alpha_1, \alpha_2, \ldots, \alpha_{j-1} \in \Psi$. Without loss of generality we assume that the submatrix consisting of the first $j - 1$ columns is nonsingular. In other words, the matrix $B = \{v^{\bar{j}}_{k^t_{\bar{j}}} r^{\bar{j}}_{s_{\bar{j}}}\}_{\bar{j} \in [j-1], t \in [j-1]}$ is nonsingular.

We consider $f_j(r^j x)$. If $v^j_1 = v^j_2 = \cdots = v^j_{m_j}$, then $f_j(r^j x)$ is a linear function, which is a contradiction to $Q = \emptyset$. Let $v^j_{\bar{k}_j} \neq v^j_{\bar{k}_j+1}$. By Claim 6 with $\bar{j} = j$ and $\bar{i} = \bar{k}_j + 1$ we get that there exist $\bar{k}_1, \ldots, \bar{k}_{j-1}, \bar{k}_{j+1}, \ldots, \bar{k}_n$ such that $\alpha = (\bar{k}_1, \ldots, \bar{k}_{j-1}, \bar{k}_j, \bar{k}_{j+1}, \ldots, \bar{k}_n) \in \Psi$ and $\beta = (\bar{k}_1, \ldots, \bar{k}_{j-1}, \bar{k}_j + 1, \bar{k}_{j+1}, \ldots, \bar{k}_n) \in \Psi$. Let us select s_j such that $r^j_{s_j} \neq 0$.

Consider the vector $\gamma_1 = \left(v^j_{k_j} r^j_{s_j}\right)_{k_j \in \alpha}$. If this vector is linearly independent from the currently constructed $j-1$ vector, then we select this vector as the new jth vector and we proceed to the next iteration.

Let $\bar{B}$ be the matrix obtained from B by appending the first $j-1$ coordinates of γ_1 and adding the column corresponding to s_j. Since we assume that γ_1 is linearly dependent from the remaining vectors, it follows that $\det(\bar{B}) = 0$. Consider now $\gamma_2 = \left(v^j_{k_j} r^j_{s_j}\right)_{k_j \in \beta}$. Let $\tilde{B}$ be the matrix obtained from B by appending the first $j-1$ coordinates of γ_2 and adding the column corresponding to s_j. Since α and β differ only in the single coordinate, it is easy to see that

$$\det(\tilde{B}) = \det(\bar{B}) \pm r^j_{s_j}(v^j_{\bar{k}_j} - v^j_{\bar{k}_j+1}) = \pm r^j_{s_j}(v^j_{\bar{k}_j} - v^j_{\bar{k}_j+1}) \neq 0\,,$$

where the sign depends on the parity of the last column in $\tilde{B}$. The last term is nonzero since by choice $r^j_{s_j} \neq 0$ and $v^j_{\bar{k}_j} \neq v^j_{\bar{k}_j+1}$.

This shows that we can construct j linear independent vectors. Repeating this procedure n times we obtain the desired result.

Consider now the function $F : \Delta \subseteq \mathbb{R}^{qn} \to \mathbb{R}$ defined by $F(x^1, \ldots, x^n) = \det\{f_j(r^j x^i)\}_{j\in[n], i\in[n]}$. Here Δ is defined by the requirement $x^t \in \{x \in \operatorname{int}(X) : b^j_{k^t_j} < r^j x < b^j_{k^t_j+1}$ for every $j \in [n]\}$, which are nonempty open sets by definition of α_t. Then by Lemma 5 we have that

$$\frac{\partial F}{\partial x^1_{s_1} \partial x^2_{s_2} \ldots \partial x^n_{s_n}} = \frac{1}{w^1_{\tilde{k}_1} \cdot w^2_{\tilde{k}_2} \cdots w^n_{\tilde{k}_n}} \det D \neq 0\,.$$

Therefore F is not anywhere 0 and in turn there exist $x^1, \ldots, x^n$ in the respective sets such that $F(x^1, \ldots, x^n) \neq 0$, which is equivalent to the statement of Claim 8.

By using Claim 8 and Lemma 4 we conclude that if $Q = \emptyset$, then there exists $\tilde{u}$ such that $u = \tilde{u}$ and all weights of $\tilde{u}$ are bounded. Note that in this case the ridge vectors and breakpoints of u and $\tilde{u}$ differ only due to the procedure described in Claim 3 and the reduction that eliminates pairwise linearly dependent ridge vectors.

It remains to consider the case $Q \neq \emptyset$. Let $\bar{j} \in Q$. By definition the corresponding i can have only values 1 and 2. We first assume that $r^{\bar{j}}$ is a linear combination of the remaining ridge vectors. Let $r^{\bar{j}} = \sum_{k\in S} \mu_k r^k$, where $S \neq \emptyset$ and $\mu_k \neq 0$ for every $k \in S$. Let us pick a fixed $\tilde{j} \in S$ and for $j \in [n]\,, j \neq \bar{j}$ we define

$$\hat{w}^j_i = \begin{cases} w^j_i + \frac{w^{\bar{j}}_2 - w^{\bar{j}}_1}{b^{\bar{j}}_2 - b^{\bar{j}}_1} b^j_i \mu_j & j \in S \setminus \{\tilde{j}\} \\ w^j_i & j \in [n] \setminus S, j \neq \bar{j} \\ w^j_i + \frac{w^{\bar{j}}_2 - w^{\bar{j}}_1}{b^{\bar{j}}_2 - b^{\bar{j}}_1} b^j_i \mu_j + w^{\bar{j}}_1 - b^{\bar{j}}_1 \frac{w^{\bar{j}}_2 - w^{\bar{j}}_1}{b^{\bar{j}}_2 - b^{\bar{j}}_1} & j = \tilde{j}\,. \end{cases}$$

Let us define $\tilde{u}$ as the ridge function with the same ridge vectors and breakpoints as u except that the ridge vector $r^{\bar{j}}$ and the corresponding breakpoints are left out. In addition, the weights of $\tilde{u}$ are defined based on $\hat{w}$. A long but straight forward calculation shows that $u(x) = \tilde{u}(x)$ for every $x \in \bar{X}$.

As a consequence we can assume that $\bar{j} \in Q$ is not a linear combination of the remaining ridge vectors. For ease of notation we assume that $\bar{j} = 1$. Let $r^1, r^2, \ldots, r^s$ be a maximal set of linearly independent vectors. Note that clearly $s \leq q$. Then there exists $y \in \mathbb{R}^q$ such that $r^1 y = 1, r^2 y = 0, \ldots, r^s y = 0$. For $s+1 \leq j \leq n$ we have $r^j = \sum_{k=2}^s \mu_k r^k$. Note that r^1 is not present in this summation due to the assumption that r^1 is not a linear combination of the remaining ridge vectors. In turn we obtain that $r^j y = 0$ for $s+1 \leq j \leq n$. Thus we conclude that $r^j y = 0$ for $j = 2, \ldots, n$.

Let us select $x \in \operatorname{int}(\bar{X})$. There exists $\epsilon > 0$ such that if $\|z - x\|_\infty \leq \epsilon$, then $z \in \bar{X}$. Consider $z = x + \epsilon y / \|y\|_\infty$. By definition of ϵ we have $z \in \bar{X}$. By choice of y we have that $r^j z = r^j x$ for every $j = 2, \ldots, n$. Therefore $r^j z$ and $r^j x$ fall in the same subinterval for $j = 2, \ldots, n$. For $j = 1$ we have a single subinterval and therefore this holds also for $j = 1$.

As a consequence we have $u(z) = u(x) + \frac{\epsilon}{\|y\|_\infty} \frac{w_2^1 - w_1^1}{b_2^1 - b_1^1}$. Since $u(x)$ and $u(z)$ are bounded it follows that $w_2^1 - w_1^1$ is bounded. By Claim 3 we can assume that one of them is bounded, which in turn implies that the other weight is bounded as well.

This shows that the two weights corresponding to each $\bar{j} \in Q$ are bounded. In turn we can reduce the general case to the case $Q = \emptyset$, which is covered by Claim 8. This completes the proof. □

D Proof of Lemma 2

Proof. It is easy to see that $g(z, \tilde{r}, \tilde{b}) \leq 2 \cdot \mid \operatorname{supp}(z) \mid \cdot \max_{(x,a)\in \operatorname{supp}(z)} z_{x,a} < \infty$, which shows that $\alpha < \infty$.

By the definition of supremum, for every integer $n, n > 0$ there exist r^n, $\|r^n\|_\infty \leq 1$ and $b^n, -\Omega \leq b^n \leq \Omega$ such that $g(z, r^n, b^n) \leq \alpha - 1/n$. We can view b^n as elements in $[-(\Omega+1), \Omega+1]^3$. As a result there exists a convergent subsequence of r^n and b^n. Without loss of generality we assume that the entire sequence is convergent. Therefore let $lim_{n\to\infty} r^n = r$ and $lim_{n\to\infty} b^n = b$. Next we consider several cases.

Case 1) Let first $b_1 < b_2 < b_3$.

We have

$$g(z, r, b) = \lim_{n\to\infty} g(z, r^n, b^n) = \alpha$$

since in this case g is continuous in the neighborhood of b and r.

Case 2) Let now $b_1 = b_2 = b_3$.

It is easy to see that for every (x, a) we have

$$\lim_{n\to\infty} H_{b^n}(r^n x) = \begin{cases} 1 & rx = b_2 \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad \lim_{n\to\infty} H_{b^n}(r^n s(x,a)) = \begin{cases} 1 & rs(x,a) = b_2 \\ 0 & \text{otherwise.} \end{cases}$$

Let $A_1 = \{(x, a) \in \operatorname{supp}(z) \,|\, rx = b_2\}$ and $A_2 = \{(x, a) \in \operatorname{supp}(z) \,|\, rs(x, a) = b_2\}$. In turn since z has finite support we have

$$\begin{aligned} \alpha = \lim_{n\to\infty} g(z, r^n, b^n) &= \Big| \sum_{(x,a)\in \operatorname{supp}(z)} z_{x,a} \lim_{n\to\infty} H_{b^n}(r^n x) - \sum_{(x,a)\in \operatorname{supp}(z)} z_{x,a} H_{b^n}(r^n s(x,a)) \Big| \\ &= \Big| \sum_{(x,a)\in A_1} z_{x,a} - \sum_{(x,a)\in A_2} z_{x,a} \Big| \ . \end{aligned} \tag{34}$$

Let $u_1(x, a) = x$ and $u_2(x, a) = s(x, a)$. For $i = 1, 2$ we define the following sets:

$$\begin{aligned} S_n^i &= \{r^n u_i(x,a) : (x,a) \in \operatorname{supp}(z)\} \\ A_n^i &= S_n^i \cap [b_1^n, b_3^n] \\ \bar{A}^i(x,a) &= \{n \in \mathbb{N} : r^n u_i(x,a) \in A_n^i\} \\ C^i &= \{(x,a) \in \operatorname{supp}(z) : \mid \bar{A}^i(x,a) \mid < \infty\} \\ D^i &= \{(x,a) \in \operatorname{supp}(z) : \mid \bar{A}^i(x,a) \mid = \infty\} \end{aligned}$$

Values not in $A_n^1 \cup A_n^2$ do not contribute towards $g(z, r^n, b^n)$ and therefore they can be neglected. Note that $\cap_{i=1}^n A_n^i$ is either $\emptyset$ or b_2.

Let $(x, a) \in D^i$. Then for infinitely many n we have $r^n u_i(x, a) \in A_n^i$. As n goes to infinity, this yields $ru_i(x, a) = b_2$. In other words, $D^1 \subseteq A^1, D^2 \subseteq A^2$. Now it is easy to see that $D^1 = A^1, D^2 = A^2$. Note also that the complement of A_i equals to C^i.

Let

$$M = \max_{\substack{(x,a)\in C^1 \cup C^2 \\ i=1,2}} \bar{A}^i(x,a) < \infty \, .$$

By definition for every $n > M$ we have $r^n u_i(x,a) \notin A_n^i$ for every $(x,a) \in C_1 \cup C_2$. If $ru_i(x,a) = \lim_{k\to\infty} r^k u_i(x,a) \in (b_1^M, b_3^M)$, then we contradict the definition of M. We conclude that for $(x,a) \in C_1 \cup C_2$ we have $ru_i(x,a) \notin (b_1^M, b_3^M)$ for $i = 1,2$.

Consider now $\hat{b}$ defined as $\hat{b}_1 = b_1^M, \hat{b}_2 = b_2, \hat{b}_3 = b_3^M$. Based on this definition if $(x,a) \in D_i$, then $ru_i(x,a) = \hat{b}_2$ and in turn $H_{\hat{b}}(ru_i(x,a)) = 1$. On the other hand if $(x,a) \in C_i$, then $ru_i(x,a) \notin (\hat{b}_1, \hat{b}_3)$ and therefore $H_{\hat{b}}(ru_i(x,a)) = 0$. Now it immediately follows that $g(z,r,\hat{b}) = \alpha$ based on (34), which completes this case.

Case 3) Let now $b_1 = b_2 < b_3$ or $b_1 < b_2 = b_3$.

This case is a combination of the previous two cases and therefore it can be proven in a similar way. □