# An Algorithm for Approximating Convex Pareto Surfaces Based on Dual Techniques

## Rasmus Bokrantz
Optimization and Systems Theory, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden;
and RaySearch Laboratories, SE-111 34 Stockholm, Sweden, bokrantz@kth.se

## Anders Forsgren
Optimization and Systems Theory, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden,
andersf@kth.se

We consider the problem of approximating Pareto surfaces of convex multicriteria optimization problems by a discrete set of points and their convex combinations. Finding the scalarization parameters that optimally limit the approximation error when generating a single Pareto optimal solution is a nonconvex optimization problem. This problem can be solved by enumerative techniques but at a cost that increases exponentially with the number of objectives. We present an algorithm for solving the Pareto surface approximation problem that is practical with 10 or less conflicting objectives, motivated by an application to radiation therapy optimization. Our enumerative scheme is, in a sense, dual to a family of previous algorithms. The proposed technique retains the quality of the best previous algorithm in this class while solving fewer subproblems. A further improvement is provided by a procedure for discarding subproblems based on reusing information from previous solves. The combined effect of the enhancements is empirically demonstrated to reduce the computational expense of solving the Pareto surface approximation problem by orders of magnitude. For problems where the objectives have positive curvature, an improved bound on the approximation error is demonstrated using transformations of the initial objectives with strictly increasing and concave functions.

*Key words*: multicriteria optimization; Pareto optimality; sandwich algorithms; radiation therapy; Hausdorff distance

## 1. Introduction

Multicriteria optimization (MCO) deals with optimization problems involving multiple mutually conflicting objectives (see, e.g., the monographs Miettinen 1999, Ehrgott 2005). Generally, there is no feasible solution that is optimal with respect to all objectives simultaneously. Instead, a well-balanced trade-off between objectives is sought within the Pareto optimal set: the set encompassed by the feasible solutions such that an improvement in one objective can only be achieved through a sacrifice in another. One approach to exploring the Pareto optimal set is to precompute a finite number of solutions without human interaction, and then use a navigation tool to form convex combinations between discrete solutions in real time (see Monz et al. 2008, Eskelinen et al. 2010). This technique is by design restricted to problems with convex constraints.

We are particularly interested in an application of the described technique to treatment planning for intensity-modulated radiation therapy (IMRT). There is a rich literature of methods that recognize IMRT planning as an MCO problem (see, e.g., Cotrutz et al. 2001; Küfer et al. 2003; Lahanas et al. 2003; Craft et al. 2005, 2007). Clinical evaluations have demonstrated that such methods have the potential of improving both manual planning time and treatment quality (Thieke et al. 2007, Hong et al. 2008, Craft et al. 2012). In view of this application, we limit ourselves to generating Pareto optimal points by replacing the vector-valued objective function of the initial problem with a weighted sum of its components. This method is extensively used throughout the field of MCO and is the de facto standard for IMRT optimization (Hunt et al. 2002, Ahnesjö et al. 2006). We further limit ourselves to consider convex objectives, so that the Pareto optimal set forms a connected surface in the boundary of a convex set (Romeijn et al. 2004). For a description of convex criteria that are commonly used in IMRT planning together with nonconvex criteria that can be reformulated as convex (see Romeijn et al. 2004).

Within this context, we consider the problem of generating a discrete set of Pareto optimal solutions so that their convex combinations are a representative approximation of the Pareto surface. The majority

of methods for this problem use polyhedral approximations of the Pareto surface to ensure an even spread of points over this set (see, e.g., the review by Ruzika and Wiecek 2005). Such methods can be broadly classified as inner approximation methods (Chernykh 1995, Schandl et al. 2002, Craft 2010); outer approximation methods (Benson 1998), and sandwich methods (Solanki et al. 1993, Klamroth et al. 2003, Craft et al. 2006, Rennen et al. 2011, Shao and Ehrgott 2008b, Ehrgott et al. 2011). In inner approximation methods, the lower boundary of the convex hull of the current set of Pareto optimal points is used to approximate the Pareto surface from within. Outer approximation methods instead find supporting hyperplanes to the Pareto surface and construct an approximation by taking the intersection of halfspaces associated with these hyperplanes. Sandwich methods make combined use of these two techniques, thereby enclosing the Pareto surface in a hypervolume between two converging approximations. Of related interest is also a technique for obtaining a Pareto set representation by approximating the nondominated set of its dual problem. This approach has been demonstrated for multiobjective linear programs in an outer approximation version (Ehrgott et al. 2012) and a sandwich approximation version (Shao and Ehrgott 2008a).

In this paper, we focus on sandwich algorithms because the maximum distance between the inner and outer approximation can be used to steer generation of new points toward parts of the Pareto surface that currently lack accurate representation, and to provide an upper bound on the approximation error. This property makes sandwich algorithms favorable for large-scale problems where the computational cost of each optimization limits the number of Pareto optimal solutions that are practically computable.

The computational expense of a sandwich algorithm increases exponentially with the number of objectives. This behavior is due to what is called the curse of dimensionality: in direct sampling of a distribution of data, the number of samples required to maintain a given level of accuracy increases exponentially with the number of variables (Bellman 1961). As a consequence applications with more than six objectives have, to the best of our knowledge, previously not been reported. The number of objectives commonly encountered in IMRT planning, on the other hand, range to about 10 (Craft and Bortfeld 2008, Spalke et al. 2009). Current practice for high-dimensional cases is to sample weights uniformly at random. This technique is well known to be inadequate for generating an even distribution of points from all parts of the Pareto surface (Das and Dennis 1997).

Motivated by these shortcomings, we develop methods for making sandwich algorithms tractable to a wider range of problem formulations. We devote particular attention to the mathematical programming and computational geometry aspects of the problem. For many problems within these two fields the primal and dual formulations have equivalent complexity. However, as noted by Bremner et al. (1998, p. 2) with respect to polytope duality: "For a particular class of polytopes and a fixed algorithm, one transformation may be much easier than its dual." Arguing that this is the case for sandwich algorithms, we give an algorithm that in a sense is dual to a family of previous algorithms in the literature. Our main contribution is a scheme that retains the quality of the best previous algorithm while achieving a more benign ratio between computational effort and problem dimension. The presented algorithm also generalizes sandwich algorithms to be compatible with cone-based preference models (see, e.g., Engau 2009, Hunt and Wiecek 2003, Monz 2006). Ordering cones have proven useful in multicriteria IMRT planning for excluding parts of the Pareto surface that are known a priori not to be of interest (Serna et al. 2009).

## 2. Preliminaries

### 2.1. Notation and Terminology
We denote by $e$ the vector of ones with its dimension defined by its context. We treat sets of points and matrices interchangeably when convenient; the rows of the matrix are the elements of the corresponding set. The shorthand $(\cdot)_+$ is used to denote $\max\{\cdot, 0\}$. We denote the optimal value of an optimization problem $P$ by $\text{optval}(P)$. For a function $f$ and a subset $S$ of its domain, we denote by $f(S)$ the image $\{f(s): s \in S\}$. For a set $S$, we denote by $\text{conv}(S)$ its convex hull. For two sets $S_1$ and $S_2$, we denote by $S_1 + S_2$ their Minkowski sum. Minkowski addition between a set $S$ and a singleton set $\{s\}$ is denoted by $S + s$. A hyperplane $\{z: a^T z = b\}$ with nonzero normal $a$ and offset $b$ is denoted by $H(a, b)$. With each hyperplane, we associate a closed positive, a closed negative, an open positive, and an open negative halfspace, defined by substituting, respectively, "$\geq$," "$\leq$," "$>$," and "$<$" for the equality in the hyperplane equation. The $k$-dimensional intersection between a polyhedron and one of its supporting hyperplanes is called a $k$-face. A 0-face is called a vertex, a 1-face an edge, an $(n-2)$-face a ridge, and an $(n-1)$-face a facet. Unless otherwise stated, a normal vector to a polyhedral face is assumed to be oriented inward.

### 2.2. Problem Formulation
We consider multiobjective optimization problems of the form

$$(\text{MOP}) \quad \begin{aligned} &\underset{x}{\text{minimize}} \quad f(x) = (f_1(x) \cdots f_n(x))^T \\ &\text{subject to} \quad x \in X = \{x: c(x) \leq 0\}, \end{aligned} \quad (1)$$

involving $n \geq 2$ objective functions $f_i: \mathbb{R}^m \to \mathbb{R}$ to be minimized over a feasible region $X \subseteq \mathbb{R}^m$ defined by a vector $c: \mathbb{R}^m \to \mathbb{R}^k$ of constraint functions. We denote by $Z$ the image of the feasible region under the objective function mapping, i.e., $Z = f(X)$. We refer to the $m$-dimensional space of which $X$ is a subset as the decision space and to the $n$-dimensional space of which $Z$ is a subset as the objective space. Throughout, the feasible region $X$ is assumed to be nonempty, and the functions $f$ and $c$ are assumed to be convex and bounded on $X$. The feasible region is a convex set because sublevel sets of convex functions are convex. Because $f$ and $X$ are convex, MOP is a convex optimization problem.

### 2.3. Notion of Optimality

The solution set to MOP is the set of nondominated feasible points. Dominance relations between points in objective space are defined with respect to the partial order induced by some ordering cone $C$, which we require to be closed, pointed ($C \cap (-C) \subseteq \{0\}$), and convex.

DEFINITION 2.1 (NONDOMINANCE). Let $x^*$ be feasible to MOP. Then, $x^*$ is nondominated if there exists no $x$ in $X$ such that $f(x^*) \in f(x) + C \setminus \{0\}$.

To easily distinguish between decision space and objective space we refer to a nondominated solution $x^*$ as efficient whereas the corresponding objective vector $f(x^*)$ is called Pareto optimal. We refer to the set of all efficient solutions as the efficient set and to the set of all Pareto optimal objective vectors as the Pareto set.

We restrict ourselves to consider polyhedral ordering cones generated by some matrix $Q$, i.e., $C = \{Q\mu: \mu \geq 0\}$. Instead of specifying $Q$ directly, we prefer to use the set of admissible trade-off rates between objectives, which is the dual cone $C^* = \{z: y^T z \geq 0, \forall y \in C\}$. Let $T$ be the symmetric $n \times n$ matrix with unit diagonal and nonnegative off-diagonal elements $t_{ij}$ such that the reciprocal of $t_{ij}$ is the maximum acceptable increase in $f_i$ for a unit decrease in $f_j$. Then, $C^*$ is the polyhedral cone generated by $T$, and $C$ is the dual cone to $C^*$, i.e., $C = \{z: Tz \geq 0\}$, here using that $C = C^{**}$ by convexity and closedness of $C$ (Stoer and Witzgall 1970, p. 53). By construction, $C^* \subseteq \mathbb{R}^n_+$, so that $\mathbb{R}^n_+ \subseteq C$ (Stoer and Witzgall 1970, p. 56). Taking $T$ and $Q$ to be the identity matrix, so that $C = C^* = \mathbb{R}^n_+$, gives dominance in the conventional Pareto sense.

### 2.4. The Weighting Method

Computing a Pareto optimal point typically involves reformulating the initial vector-valued problem to a parameter-dependent scalar problem. We restrict ourselves to consider scalarization using the weighting method, where an $n$-vector $w$ of weights such that $w \in C^*$ is introduced to yield the single-objective optimization problem

$$(\text{SUM}(w)) \quad \begin{array}{ll} \underset{x}{\text{minimize}} & w^T f(x) \\ \text{subject to} & x \in X. \end{array}$$

This is a convex optimization problem because $C^* \subseteq \mathbb{R}^n_+$ and nonnegative linear combinations preserve convexity. The vector $w$ is throughout assumed to be normalized so that $e^T w = 1$.

Problems MOP and SUM($w$) are related as follows. Sufficient conditions for a point $x^*$ to be an efficient solution to MOP is that $x^*$ is an optimal solution to SUM($w$) for some $w$ in $C^*$ such that $w > 0$ (Miettinen 1999, Theorem 3.1.2), or that $x^*$ is a unique optimal solution to SUM($w$) for some $w$ in $C^*$ (Miettinen 1999, Theorem 3.1.3). A sufficient condition for uniqueness is that all objectives are strictly convex. Any $x^*$ that is optimal to SUM($w$) for some $w$ in $C^*$ defines a hyperplane $H(w, f(x^*))$ that supports the feasible objective space $Z$ at $f(x^*)$. To see this, observe that $f(x^*) \in Z \cap H(w, f(x^*))$ and that the intersection between $Z$ and the open negative halfspace associated with $H(w, f(x^*))$ is empty, or otherwise $x^*$ would not be optimal to SUM($w$). A necessary condition for a point $x^*$ to be an efficient solution to MOP is, by convexity, that there exists $w$ in $C^*$ such that $x^*$ is an optimal solution to SUM($w$) (Miettinen 1999, Theorem 3.1.4). Finding any point on the Pareto surface consequently decomposes to solving SUM($w$) with $w$ being normal to the Pareto surface at the sought-after point.

## 3. The Sandwich Algorithm

### 3.1. The Algorithmic Idea

A generic sandwich algorithm based on the weighting method is stated in Algorithm 3.1, and schematically illustrated for a bi-objective problem in Figure 1. This algorithm generates a set of points such that their convex hull constitutes an approximation of the efficient set with approximation error below some $\varepsilon > 0$. The computational goal is to construct the approximation with as few solves as possible. The algorithm avoids assessing the quality of the approximation of the efficient set in the typically high-dimensional decision space by a mapping to objective space. The resulting image is evaluated with respect to its approximation of the Pareto surface. An upper bound on the approximation error is calculated as the distance between polyhedral inner and outer approximations of the Pareto surface. The weighting vector in the next weighted-sum problem to be solved is the normal to the inner approximation at the point where the upper bound is attained. This choice corresponds to

**(a) Iteration 1**   **(b) Iteration 2**   **(c) Iteration 3**
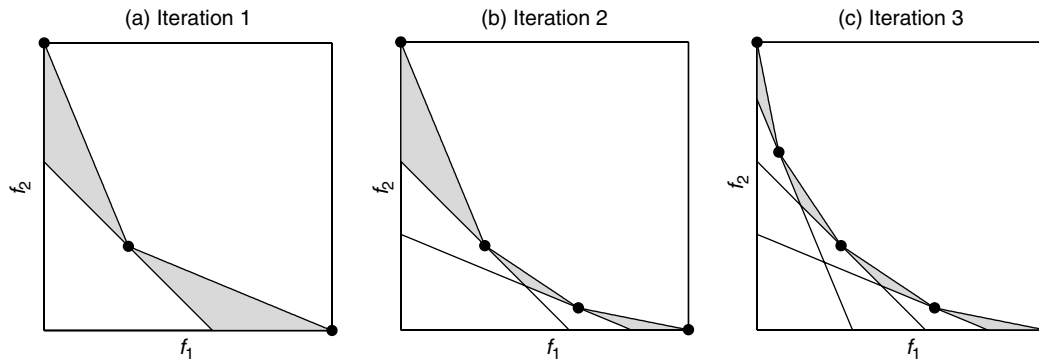
**Figure 1    First Three Iterations of the Sandwich Algorithm Applied to a Bi-Objective Program**

*Notes.* The points indicate Pareto optimal objective vectors. The Pareto surface is enclosed in the shaded region between the inner and outer approximations.

the greedy strategy of maximizing the decrease in the approximation error.

The first two steps in Algorithm 3.1 normalize the range of each objective function to avoid bias toward objectives of large magnitude. We use the pragmatic approach of normalizing each objective function with respect to its minimum and maximum value during the $n$ initial solves of SUM($w$).

**Algorithm 3.1** (The Sandwich Algorithm)
  **for** $i = 1, \ldots, n$ **do**
    solve SUM($t_i$) with $i$ being the $i$th extreme ray
    of $C^*$;
  normalize $f$ to $[0, 1]^n$;
  solve SUM($w$) with $w = (1/n)e$;
  construct inner and outer approximations $Z_{in}$ and
    $Z_{out}$ of the Pareto surface;
  **while** not converged **do**
    compute an upper bound on the approximation
      error;
    **if** the upper bound is below $\varepsilon$ **then**
      converged; **continue**;
    solve SUM($w$) with $w$ normal to $Z_{in}$ where the
      upper bound is attained;
    update $Z_{in}$ and $Z_{out}$;
  **end**

### 3.2.   Relation to Previous Work

Sandwich algorithms originate from methods for approximating univariate convex functions by piecewise linear and continuous upper and lower approximations (see, e.g., Fruhwirth et al. 1989, Rote 1992). Such methods partition the real line into contiguous intervals by iteratively evaluating the function and its derivative. A function evaluation is performed at each iteration to partition the interval with greatest approximation error into two subintervals, thereby improving the overall error measure. Different implementations of this scheme can be categorized according to partitioning rule (see, e.g., the review by Rote 1992). Two notable examples are the maximum error rule and the chord rule. The maximum error rule adds a

new point with abscissa equal to that of the vertex of its lower bounding approximation, see Figure 2(a). With the chord rule, a new point is added with tangent equal to the slope of the upper bounding approximation, see Figure 2(b).

For the problem of approximating a Pareto surface, the analog of a partitioning rule is a method for generating scalar subproblems. A general-dimensional analog of the maximum error rule is provided by Benson's outer approximation algorithm (Benson 1998). This method has been extended to a sandwich technique for multiobjective linear programs by Shao and Ehrgott (2008b), and further generalized to the convex nonlinear case by Ehrgott et al. (2011). Here, a reference point is first selected in the interior of the feasible region in objective space. Pareto optimal objective vectors are then generated at the intersection of the Pareto surface and a line segment with endpoints defined by the reference point and a vertex of the outer approximation. This step requires solving scalar problems with the initial objectives posed as constraints and an auxiliary variable as objective.

In this paper, we focus on a general-dimensional analog of the chord rule because this leads to scalarization using the weighting method. General-dimensional sandwich algorithms based on the weighting

**(a) The maximum error rule**   **(b) The chord rule**
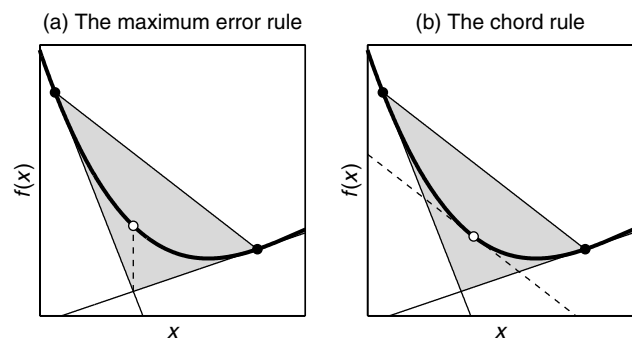
**Figure 2    Two Partitioning Rules for Approximation of a Convex Univariate Function by Piecewise Linear Upper and Lower Bounds**

*Note.* The partitioning point is indicated in white.

method have previously been suggested by Solanki et al. (1993), Craft et al. (2006), and Rennen et al. (2011), where the last two can be viewed as enhanced versions of the preceding one(s). A complicating issue for this class of methods is the fact that a vector normal to the convex hull of a discrete set of Pareto optimal points can have negative components in dimensions beyond two. Hence, if such a normal vector is used directly as a weighting vector in SUM($w$), this vector may lie outside $C^*$, and the resulting optimal solution may thus not be efficient to MOP. The algorithm of Solanki et al. (1993) handles this complication by introducing bounds on the allowed deviation from the Pareto surface. In the algorithm of Craft et al. (2006), a heuristic method is instead used to transform mixed normals into nonnegative ones, to make better use of each optimization. The algorithm of Rennen et al. (2011) avoids mixed normals altogether by augmenting the convex hull representation of the inner approximation through setwise summation with the nonnegative orthant. Any normal to the resulting polyhedron is nonnegative (Rennen et al. 2011, Lemma 2).

For the univariate case, the maximum error rule and the chord rule have been shown by Rote (1992) to be theoretically equivalent under geometric duality of convex functions. This equivalence does not extend to the general-dimensional case because although the chord rule is invariant to affine transformations, the maximum error rule becomes biased by its reference point (Rote 1992). The methods also differ in their practical implementations. Rote notes that the introduction of additional constraints necessitated by the maximum error rule may destroy some inherent structure of the initial problem such as the constraint structure of a maximum flow problem.

In Rennen et al. (2011), the algorithms of Solanki et al. (1993), Craft et al. (2006), and Rennen et al. (2011) are empirically evaluated on a suite of test problems. This study indicates that the algorithm of Rennen et al. (2011) generates well-distributed points on the Pareto surface and provides a corresponding rapid improvement in bound on the approximation error, whereas this is generally not the case for the algorithms of Solanki et al. (1993) and Craft et al. (2006). Based on these findings, we use the algorithm of Rennen et al. (2011) as a single benchmark to the algorithm proposed in this paper, both in the theoretical exposition and the numerical experiments.

### 3.3. Polyhedral Approximations
The key result that makes construction of polyhedral approximations of the Pareto surface possible is that the set $Z_+ = Z + C$ is a convex set whenever MOP is a convex problem. This result is a generalization of the result that a function is convex if and only if its epigraph is a convex set to the case of epigraphs induced by convex cones (Pennanen and Eckstein 1997). Other proofs of this result for the special case $C = \mathbb{R}^n_+$ can be found in Romeijn et al. (2004), and Craft (2010). Convexity of $Z_+$ implies that the convex hull of any discrete set of points in $Z_+$ is an inner approximation of this set, and that the intersection of any set of closed positive halfspaces associated with supporting hyperplanes to $Z_+$ an outer approximation. In particular, polyhedral approximations of $Z_+$ can be constructed as follows.

DEFINITION 3.1 (INNER AND OUTER APPROXIMATIONS). Let $D$ be a discrete set of points that are efficient to MOP and optimal to SUM($w$) with re-BBB spect to some set $W$ of weighting vectors in $C^*$. Then, $Z_{in} = \{P^T\lambda + Q^T\mu: \lambda, \mu \geq 0, e^T\lambda = 1\}$, where $P = f(D)$, is an inner approximation of $Z_+$, and $Z_{out} = \{z: Wz \geq r\}$, where $r$ is the vector of pairwise scalar products between the rows of $P$ and the rows of $W$, an outer approximation of $Z_+$, in the sense that $Z_{in} \subseteq Z_+ \subseteq Z_{out}$.

### 3.4. Quantifying the Approximation Error
Quality measures of discrete representations of the Pareto set have been reviewed by Sayin (2000). The definition we give quantifies the coverage of the Pareto set in terms of a relaxation of the nondominance criterion.

DEFINITION 3.2 (APPROXIMATION ERROR). Let $D$ be a discrete set of feasible points to MOP. Then, the approximation error of $D$ is the minimum $\varepsilon$ such that for any efficient $x^*$, there exists $x$ in conv($D$) such that $f(x^*) \in f(x) + (C - \varepsilon e)$.

Unfortunately, explicit knowledge of the Pareto set is required to compute this quantity. Because the entire Pareto set is unknown in general, we work with the upper bound on the approximation error provided by the minimum $\varepsilon$ such that $Z_{out} \subseteq (Z_{in} - \varepsilon e)$. This upper bound is equivalent to the Hausdorff distance

$$d_{\text{Haus}}(Z_{in}, Z_{out}) = \max_{z \in Z_{out}} \min_{z' \in Z_{in}} d(z, z'),$$

with respect to the one-sided distance function

$$d(z, z') = \max_{i=1, \ldots, n} (z'_i - z_i)_+.$$

Computing $d_{\text{Haus}}(Z_{in}, Z_{out})$ requires solving the linear bilevel program

$$\underset{z}{\text{maximize}} \left\{ \begin{array}{l} \underset{\eta, \lambda, \mu}{\text{minimize}} \quad \eta \\ \text{subject to} \quad \eta e \geq P^T\lambda + Q^T\mu - z, \\ \qquad\qquad\quad e^T\lambda = 1, \\ \qquad\qquad\quad \eta, \lambda, \mu \geq 0. \end{array} \right\} \quad (2)$$

subject to $Wz \geq r.$

Linear bilevel problems have been shown to be NP-hard and inapproximable within any constant factor in polynomial time (Dempe 2002, Theorem 3.12). Problems within this class may be exactly solvable by enumerative techniques in moderate problem dimensions, whereas finding the optimum to large-scale instances is in general not tractable.

# 4. The Vertex Enumerative Algorithm

## 4.1. Solution by Enumerating the Vertices of the Outer Approximation

We propose to solve (2) by enumerating the extreme points of its feasible region and solving a linear programming subproblem for each extreme point found. The extreme points of the outer approximation are its finitely many vertices. Validity of the proposed method relies on a combination of well-known results. For completeness, we summarize these in Proposition 4.1. A proof is given in §6 under more general conditions.

PROPOSITION 4.1. *An optimal solution to the bilevel program* (2) *occurs at a vertex of the polyhedron defined by the constraints of its outer level program.*

As a direct consequence of Proposition 4.1, the optimal value of (2) is given by $\max_{v \in V} \text{optval}(\text{PLP}(v))$, where $V$ denotes the set of vertices of the outer approximation and where

$$\underset{\eta, \lambda, \mu}{\text{minimize}} \quad \eta \tag{3a}$$

$$(\text{PLP}(v)) \qquad \text{subject to} \quad \eta e \geq P^T \lambda + Q^T \mu - v, \tag{3b}$$

$$e^T \lambda = 1, \tag{3c}$$

$$\eta, \lambda, \mu \geq 0. \tag{3d}$$

We postpone for the moment how to enumerate the vertices of $Z_{\text{out}}$ and first discuss some properties of the above linear program.

## 4.2. Identifying the Next Weighting Vector

Having solved all instances of (3), we turn to identifying the weighting vector of the next weighted-sum problem that is to be solved. Similar lines of reasoning have previously been applied to a related problem by Craft (2010). The linear programming dual to PLP($v$) is

$$\underset{\pi, \rho}{\text{maximize}} \quad \rho - v^T \pi \tag{4a}$$

$$(\text{DLP}(v)) \qquad \text{subject to} \quad P\pi \geq \rho e, \tag{4b}$$

$$Q\pi \geq 0, \tag{4c}$$

$$e^T \pi \leq 1, \tag{4d}$$

$$\pi \geq 0. \tag{4e}$$

It is straightforward to verify that PLP($v$) is feasible and its objective value bounded from below. Therefore, DLP($v$) is feasible and its objective value is bounded from above. Moreover, for any primal-dual optimal solution $(\eta, \lambda, \mu, \pi, \rho)$ to PLP($v$) and DLP($v$) that is associated with some $v$ in $V$ such that optval(PLP($v$)) > 0, $\pi$ lies in $C^*$ and is normal to the inner approximation at $y = P^T \lambda + Q^T \mu$. This claim is made precise in Proposition 4.2. The next weighting vector is thus the vector of optimal dual variables $\pi$ to the instance of DLP($v$) with maximum optimal value.

PROPOSITION 4.2. *Let* $(\eta, \lambda, \mu, \pi, \rho)$ *denote a primal-dual optimal solution to* PLP($v$) *and* DLP($v$) *defined by some vertex $v$ in $V$ such that $\eta > 0$. Then, the hyperplane $H(\pi, \rho)$ supports $Z_{\text{in}}$ at $y = P^T \lambda + Q^T \mu$ and has normal vector $\pi$ in $C^* \setminus \{0\}$.*

PROOF. Feasibility and boundedness of PLP($v$) and DLP($v$) by strong duality for linear programming imply that optval(PLP($v$)) = optval(DLP($v$)), or equivalently

$$\eta = \rho - v^T \pi. \tag{5}$$

This result, the assumption that $\eta > 0$, (4b), and (4e) together imply that $\pi \neq 0$. The set $H(\pi, \rho)$ thus forms a hyperplane in objective space with normal $\pi$ in $C^* \setminus \{0\}$ by feasibility with respect to (4c). To show that $H(\pi, \rho)$ supports $Z_{\text{in}}$ at $y$, it remains to show that $Z_{\text{in}}$ is entirely contained in the closed positive halfspace associated with $H(\pi, \rho)$ and that $y$ is contained in $H(\pi, \rho)$. Take any $\bar{y}$ in $Z_{\text{in}}$ parametrized by some $\bar{\mu}$ and $\bar{\lambda}$ that are feasible to (3), i.e., $\bar{y} = P^T \bar{\lambda} + Q^T \bar{\mu}$. Then,

$$\pi^T \bar{y} = \pi^T (P^T \bar{\lambda} + Q^T \bar{\mu}) \geq \pi^T P^T \bar{\lambda} \geq \rho \bar{\lambda}^T e = \rho, \tag{6}$$

by (4c), (3d), (4b), and (3c), which yields the first part of the statement. Forming the scalar product between $\pi$ and $y$ gives

$$\pi^T y = \pi^T (P^T \lambda + Q^T \mu) \leq \pi^T (\eta e + v)$$
$$\leq \eta + \pi^T v = \rho, \tag{7}$$

by (3b), (4d), and (5). Inserting $y$ in (6) gives that $\pi^T y \geq \rho$. Therefore, all inequalities in (7) are tight and $y \in H(\pi, \rho)$. $\square$

## 4.3. Reducing the Number of Subproblems to be Solved

The number of linear programming subproblems of the form (3) that needs to be solved to compute the optimal value of (2) can be reduced by bounding from above the optimal value of some of the subproblems. Consider a sequence of solves of PLP($v$) over $v$ in $V$ and let $\beta^*$ denote the maximum optimal value obtained. Then, any vertex $\bar{v}$ in $V$ that

is optimal to (2) must satisfy $\text{optval}(\text{PLP}(\bar{v})) \geq \beta^*$. Therefore, for any vertex $v$ in $V$ and scalar $\beta$ such that $\text{optval}(\text{PLP}(v)) \leq \beta \leq \beta^*$, the instance $\text{PLP}(v)$ need not be solved. An upper bound on the optimal value of an instance of (3) is provided by the following result.

PROPOSITION 4.3. *For an iteration of the sandwich algorithm, let $Z_{\text{out}} = \{z: Wz \geq r\}$ denote the outer approximation, $V$ its vertex set, and $\text{PLP}(v)$ an instance of (3) defined by some $v$ in $V$. Let also the corresponding notation with superscript "$+$" apply to the subsequent iteration. Then, for any $v$ in $V^+$, it holds that*

$$\text{optval}(\text{PLP}^+(v))$$
$$\leq \begin{cases} \text{optval}(\text{PLP}(v)) & \text{if } v \in V \\ \max_{\bar{v} \in E} \text{optval}(\text{PLP}(\bar{v})) & \text{otherwise,} \end{cases} \quad (8)$$

*where $E$ is the extreme point set of the unique edge of $Z_{\text{out}}$ that contains $v$.*

PROOF. First suppose that $v \in V$ and let $(\eta, \lambda, \mu)$ be an optimal solution to $\text{PLP}(v)$. Then, $(\eta, (\lambda^T \ 0)^T, \mu)$ is feasible to $\text{PLP}^+(v)$ with objective value $\text{optval}(\text{PLP}(v))$. Hence, the if clause in (8) follows. Suppose that $v \notin V$. Because $v \in V^+$, the system $W^+ v \geq r^+$ is satisfied with equality in exactly $n$ linearly independent rows. Similarly, $v \notin V$ and $Z_{\text{out}}^+ \subseteq Z_{\text{out}}$ implies that the system $Wv \geq r$ is satisfied with equality in at most $n-1$ linearly independent rows. Because $W^+$ is $W$ augmented with one additional row, the system $W^+ v \geq r^+$ is satisfied with equality in exactly $n-1$ linearly independent rows. The point $v$ is thus contained in an edge of $Z_{\text{out}}$. By an argument analogous to that in the proof of Proposition 6.1, the maximum optimal value of the inner level linear program in (2) taken over all points in this edge occurs at one of its extreme points, which yields the otherwise clause in (8). □

### 4.4. Enumerating the Vertices of the Outer Approximation

We enumerate the vertices of the outer approximation by representing this set as a polytope and converting its halfspace representation to a vertex representation. To perform the latter of these two steps, we use the fact that vertex enumeration is equivalent to a convex hull problem under polar duality between points and hyperplanes defined by a reciprocation $H(a, b) \mapsto (a_1/b \cdots a_n/b)^T$ about the unit sphere (Preparata and Shamos 1985). We first define a duality relation between polytopes and then outline the vertex enumerative scheme.

DEFINITION 4.1 (POLYTOPE DUALITY). Let $A$ be a polytope that contains the origin in its strict interior. Then, the polytope $A^* = \{z: y^T z \leq 1, \forall y \in A\}$ is the polar dual of $A$.

Polar duality defines a bijection between the facets of a polytope and the vertices of its dual. This correspondence is inclusion reversing in the sense that two facets incident on a common ridge uniquely correspond to two vertices contained in a common edge (Grünbaum 2003, Theorem 3.4.4). Polar duality is moreover a reflexive transformation, so that twice dualizing a polytope gives back the initial polytope (Grünbaum 2003, Theorem 3.4.9).

We use the above theory to enumerate the vertices of the outer approximation with the following steps: (i) augment the outer approximation with $n$ sufficiently large upper bounds so that the resulting set is closed and bounded; (ii) identify a point in the interior of the resulting primal polytope, e.g., the arithmetic mean of the vertices of the inner approximation; (iii) translate the coordinate system so that this point is the origin; (iv) dualize the outward oriented bounding hyperplanes of the primal polytope; (v) solve for the convex hull of the resulting points, thus obtaining a halfspace representation of the dual polytope; (vi) dualize the facet-inducing hyperplanes of the dual polytope; (vii) translate the resulting points back into the initial coordinate system; and (viii) remove any point that satisfies any of the auxiliary upper bounds with equality. The resulting set of points forms the vertices of the outer approximation.

### 4.5. Performing the Polyhedral Computations Online

The problem in step (v) in §4.4 is the online convex hull problem: we are given points one at a time and after receiving each point, we are to compute the convex hull of the points received so far. The variant of this problem in which all input points are known in advance is called the offline convex hull problem.

We solve the online convex hull problem by maintaining a graph representation of the current convex hull with facets as nodes and ridges between adjacent facets as edges. We make the mild assumption that the vertices of the dual polytope are nondegenerate, i.e., no $(n+1)$-tuple of points lie in a common hyperplane. By this assumption, any facet of the dual polytope is an $(n-1)$-simplex incident on exactly $n$ ridges, and dually, exactly $n$ edges of the primal polytope are incident on any common vertex (Grünbaum 2003). Nondegeneracy can be simulated using standard perturbation techniques (see, e.g., Edelsbrunner 1987, p. 185).

The facet graph is updated using a beneath-and-beyond step (see, e.g., Preparata and Shamos 1985, Edelsbrunner 1987). In brief, one such step processes a new point by partitioning the facets of the current convex hull into disjoint sets of visible and obscured facets. A facet is visible if it contains the new point in its associated open negative halfspace. Obscured

facets are defined conversely. One visible facet is first identified. Remaining visible facets are then found by a depth-first search through adjacent visible facets, using the fact that the set of visible facets forms a connected subgraph. A cone of new facets is created from the new point to all ridges on which one visible and one obscured facet are incident. The visible facets are finally deleted.

Efficiently identifying the first visible facet is nontrivial in the general online version of the convex hull problem. However, because the vertex $v$ of the outer approximation that, in a given iteration, was found to be most distant from the inner approximation cannot be a vertex of the outer approximation in the subsequent iteration, a visible facet for our problem is immediately available as the facet dual to $v$. With an online convex hull algorithm, $p$ iterations of the sandwich algorithm requires $p$ beneath-and-beyond steps. The straightforward solution of calling an offline algorithm in every iteration requires $\sum_{k=1}^{p} k = p(p+1)/2$ beneath-and-beyond steps.

The upper-bounding technique of §4.3 can be incorporated with the described convex hull method as follows. With each facet of the dual polytope that is dual to a vertex $v$ of the outer approximation, we attach the current best upper bound on optval(PLP($v$)). This upper bound is updated whenever PLP($v$) is solved. At the creation of a new facet, its upper bound is initialized as the maximum over the upper bounds attached to any of the two facets incident on the ridge that induces the new facet. Validity of this update rule follows from the observation that the ridge is dual to an edge of the outer approximation defined in Proposition 4.3 by the incidence-reversing property of polytope duality.

# 5. Comparison with the Facet Enumerative Algorithm

## 5.1. Solution by Enumerating the Facets of the Inner Approximation

Problem (2) is in the algorithm of Rennen et al. (2011) solved by enumerating the facet-inducing hyperplanes of the inner approximation. Let $F$ denote the set of facet-inducing hyperplanes of $Z_{\text{in}}$ and take any hyperplane $H(\pi, \rho)$ in $F$. Then, if $\lambda$ and $\mu$ in (2) are restricted to values such that $P^T \lambda + Q^T \mu \in H(\pi, \rho)$ and the normal $\pi$ is normalized so that $e^T \pi = 1$, the optimal value function of the inner level linear program in (2) decomposes by algebraic manipulations into $\rho - \pi^T z$. The optimal value of (2) is thus given by $\max_{H(\pi, \rho) \in F} \text{optval}(\text{LP}(\pi, \rho))$, where

$$(\text{LP}(\pi, \rho)) \quad \begin{array}{l} \underset{z}{\text{maximize}} \quad \rho - \pi^T z \\ \text{subject to} \quad Wz \geq r. \end{array} \quad (9)$$

The normal $\pi$ of the hyperplane at which the maximum is attained is taken as weighting vector in the next iteration of the sandwich algorithm. The set $F$ is determined by computing the convex hull of the union of $P$ and the set $\{p + \theta q : p \in P, q \in Q\}$, where $\theta > 0$ is a sufficiently large scalar. The facet enumerative algorithm can be enhanced with an upper-bounding procedure completely analogous to that outlined for the vertex enumerative algorithm.

## 5.2. Correspondence Between Algorithms
The vertex enumerative algorithm and the facet enumerative algorithm are both methods for removing the nonlinearity of (2) by replacing a variable with a fixed value. The two resulting linear programming subproblems (4) and (9) both geometrically correspond to maximizing the projective distance between a hyperplane and a point, subject to the constraints that the point is contained in the outer approximation and that the hyperplane contains the inner approximation in its positive halfspace. The hyperplane constitutes the free variable in (4) whereas the point is the free variable in (9). These two problems are essentially equivalent under point-hyperplane duality. We therefore prefer to view the vertex enumerative algorithm and the facet enumerative algorithm as a primal-dual pair of methods. To further illustrate this duality relation, we re-examine the example shown in Figure 1 in a dual space defined by reciprocation about an interior point of the inner approximation, see Figure 3. From this viewpoint, the vertex enumerative algorithm proceeds by enumerating the facets of the inner polyhedron, just as the facet enumerative algorithm does in primal space. Note though that the vertex enumerative algorithm is not the facet enumerative algorithm applied in dual space, because polar duality is a nonlinear mapping.

It may be observed that the hyperplane $H(\pi, \rho)$ considered in (9) induces an $(n-1)$-face of the inner approximation, whereas the corresponding hyperplane considered in (4) induces a general $k$-face. The dimensionality $k$ here depends on the choice of linear programming algorithm used to solve (4). For a simplex method that converges to vertex solutions, the solution to (4) satisfies $n$ of the components of (4b) and (4c) with equality, and therefore, $k = n - 1$. If instead using an interior point method that converges to the analytic center of the optimal face, the number of binding constraints and hence the dimensionality $k$ may be lower. For a given pair of inner and outer approximations, the vertex enumerative and the facet enumerative approaches thus provide an identical upper bound on the approximation error, but the weighting vector returned by the two approaches need not be equal.
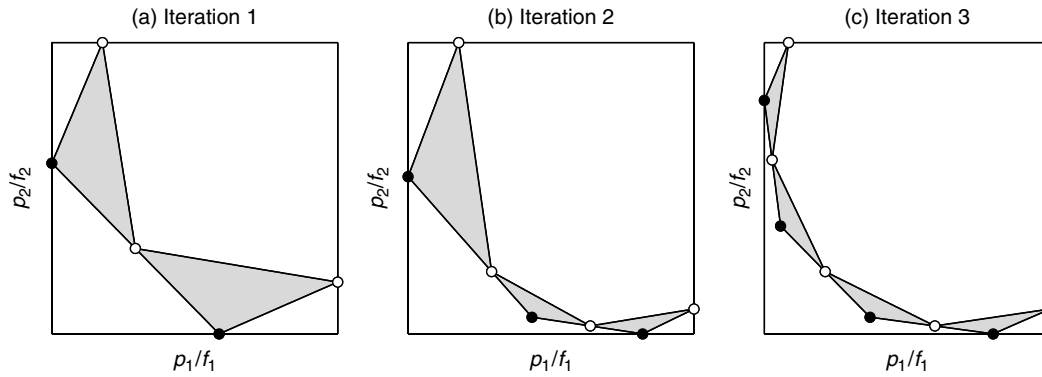
**Figure 3    The Polar Dual of the Pareto Surface Approximation Shown in Figure 1**

*Notes.* Polar duality is defined by a reciprocation about a circle with center point $p$ contained in the interior of the inner approximation. Vertices corresponding to edges of the inner approximation are indicated in black. Vertices corresponding to edges of the outer approximation are indicated in white.

### 5.3.  Computational Complexity

Disregarding an inevitable solve of a SUM($w$) problem, the computational cost of an iteration of the vertex enumerative algorithm consists of the cost of enumerating the vertices of the outer approximation and the cost of solving (2) by a sequence of linear programming subproblems. These two costs are directly proportional to the number of visible facets of the dual polytope (Edelsbrunner 1987) and the number of subproblems of the form (3) that are solved. Both are bounded from above by the number of facets of the dual polytope. A tight upper bound on the number of facets of a convex hull of $k$ points in an $n$-dimensional Euclidean space has been proved by McMullen (1970), namely,

$$\varphi(k, n) = \binom{k - \left\lfloor \frac{n+1}{2} \right\rfloor}{k - n} + \binom{k - \left\lfloor \frac{n+2}{2} \right\rfloor}{k - n}.$$

Then, because the dual polytope in the $k$th iteration of the vertex enumerative scheme is the convex hull of $2n + k + 1$ points, the total cost for $p$ iterations is bounded by

$$\sum_{k=2n+1}^{2n+p+1} O\big(\varphi(k, n)\big) \leq O\big(p\varphi(2n+p+1, n)\big). \tag{10}$$

In the $k$th iteration of the facet enumerative scheme, the polytope representation of the inner approximation is the convex hull of $(k+1)(n+1)$ points. By analogous reasoning, its total cost for $p$ iterations is thus bounded by

$$\sum_{k=n+1}^{n+p+1} O\big(\varphi((k+1)(n+1), n)\big)$$

$$\leq O\big(p\varphi((p+1)(n+1), n)\big). \tag{11}$$

Figure 4 illustrates the worst-case complexity of the vertex enumerative scheme and the facet enumerative
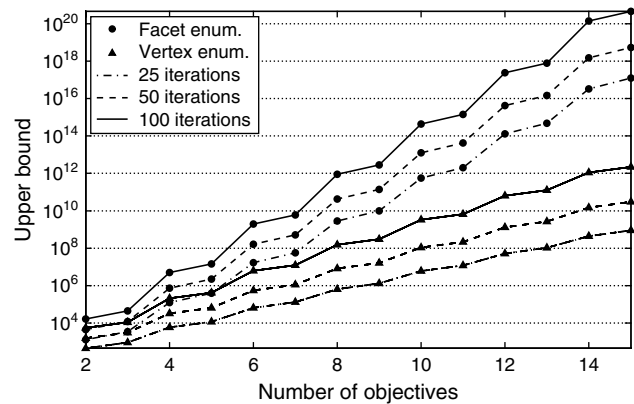


**Figure 4    Upper Bound on Number of Beneath-and-Beyond Steps and Number of Linear Programming Solves as a Function of Number of Objectives and Total Number of Iterations of the Sandwich Algorithm**

scheme according to (10) and (11), respectively, as a function of $n$ and at various fixed number of iterations $p$.

## 6.  Sandwich Approximations Under Monotonic Transformations

Sandwich approximations under transformations with strictly increasing functions have been studied in the univariate case by Siem et al. (2008). The benefit of this technique is twofold: it allows for tighter bounds on the approximation error, and it extends the applicability of sandwich methods to problems with non-convex objectives. Generalization of this technique to the $n$-dimensional case has been provided by Rennen et al. (2011). For completeness, we outline how monotonic transformations can be utilized in combination with the vertex enumerative approach taken in this paper.

Dispense with the convexity assumption on $f$ and consider a strictly increasing function $h: \mathbb{R}^n \to \mathbb{R}^n$ such

that $h \circ f$ is convex. Then, the transformed problem

$$
\begin{aligned}
&\underset{x}{\text{minimize}} && h \circ f(x) = \left(h_1 \circ f_1(x) \cdots h_n \circ f_n(x)\right)^T \\
&\text{subject to} && x \in X = \{x \colon c(x) \le 0\},
\end{aligned}
\tag{12}
$$

is a convex multiobjective program of the same form as problem (1). Let the notation of §§2 and 3 with accent "$\wedge$" apply to problem (12), e.g., $\hat{Z} = h(f(X))$. Let also the superscript "$-1$" denote sets under the inverse transformation $h^{-1} \colon \mathbb{R}^n \to \mathbb{R}^n$, e.g., $\hat{Z}_+^{-1} = h^{-1}(\hat{Z}_+)$. Note that $h^{-1}$ always exists because $h$ is strictly increasing. Then, by strict monotonicity of $h$, the efficient sets of (1) and (12) coincide, so that $Z_+ = \hat{Z}_+^{-1}$, and the inverse images $\hat{Z}_{\text{in}}^{-1}$ and $\hat{Z}_{\text{out}}^{-1}$ are inner and outer approximations of $Z_+$, i.e., $\hat{Z}_{\text{in}}^{-1} \subseteq Z_+ \subseteq \hat{Z}_{\text{out}}^{-1}$. This result is proven in Rennen et al. (2011, Proposition 5). The set $Z_+$ may consequently be approximated by applying the sandwich method in transformed objective space.

Now suppose that $f$ is convex and that $h$ is a strictly increasing and concave function such that $h \circ f$ is convex. Then, $h^{-1}$ is convex, and it holds that

$$
Z_{\text{in}} \subseteq \hat{Z}_{\text{in}}^{-1} \subseteq Z_+ \subseteq \hat{Z}_{\text{out}}^{-1} \subseteq Z_{\text{out}}.
$$

Proof of this claim is provided in Rennen et al. (2011, Propositions 6 and 7). The distance $d_{\text{Haus}}(\hat{Z}_{\text{in}}^{-1}, \hat{Z}_{\text{out}}^{-1})$ is thus a tighter bound on the current approximation error than our previous bound $d_{\text{Haus}}(Z_{\text{in}}, Z_{\text{out}})$, as illustrated for a bi-objective program in Figure 5.

Calculating the improved upper bound amounts to solving the nonlinear bilevel program

$$
\underset{z}{\text{maximize}}
\begin{cases}
\underset{\eta,\,\lambda,\,\mu}{\text{minimize}} & \eta \\
\text{subject to} & \eta e \ge h^{-1}(\hat{P}^T \lambda) \\
& \qquad + Q^T \mu - z, \\
& e^T \lambda = 1, \\
& \eta,\,\lambda,\,\mu \ge 0.
\end{cases}
\tag{13a}
$$

$$
\text{subject to} \quad \hat{W} h(z) \ge \hat{r}.
\tag{13b}
$$

The inner level program is a convex optimization problem because $h^{-1}$ is convex and the sublevel set of any convex function is convex. The feasible region of the outer level program is a convex set because $h$ is concave and the superlevel set of any concave function is convex. Similar to problem (2), problem (13) thus corresponds to maximizing a convex function over a convex set. The vertex enumerative approach is however not directly applicable to this problem because its feasible region is not polyhedral.

To make the vertex enumerative algorithm valid for (13), we replace the nonlinear constraints of the outer level program with a system of linear inequalities defined by a first-order Taylor series expansion of (13b) about some set of points that satisfy at least one row of this system with equality. Let $\hat{w}_j^T$ denote the $j$th row in $\hat{W}$ and $\hat{r}_j$ the $j$th element in $\hat{r}$. Then, the linearized constraints are given by the union of linear inequalities of the form

$$
\hat{w}_j^T \left(h(p) + \nabla h(p)^T (z - p)\right) \ge \hat{r}_j,
\tag{14}
$$

over all active constraints $j$ at all points $p$ considered during the linearization. Denote the linearized outer approximation by $\hat{Z}_{\overline{\text{out}}}^{-1}$. Then, if assuming that the linearization is performed about at least all points in $h(P)$, it holds that $\hat{Z}_{\overline{\text{out}}}^{-1} \subseteq Z_{\text{out}}$.

By convexity of the feasible region of (13), the linearization results in a relaxation of this problem, thus making the optimal value of the relaxed program an upper bound on the optimal value of (13). The relaxed problem is a convex maximization problem over a polyhedral set that can be solved using a direct analogue of the vertex enumerative algorithm for (2). In this case, the subproblems are nonlinear
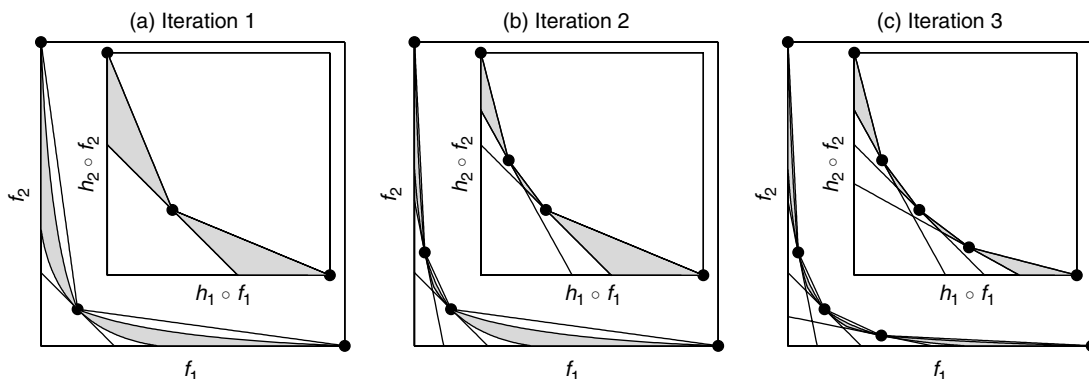


**Figure 5** **First Three Iterations of the Sandwich Algorithm Applied to a Bi-Objective Program Under a Transformation with a Concave and Strictly Increasing Function $h$**

*Notes.* The discrete points indicate Pareto optimal objective vectors. The Pareto surface is enclosed in the shaded region between the inner and outer approximations. The insert depicts the Pareto surface approximation in the transformed objective space.

programs of the form

$$\underset{\eta,\,\lambda,\,\mu}{\text{minimize}} \quad \eta \tag{15a}$$

$$\text{subject to} \quad \eta e \geq h^{-1}(\hat{P}^T \lambda) + Q^T \mu - v, \tag{15b}$$

$$e^T \lambda = 1, \tag{15c}$$

$$\eta, \lambda, \mu \geq 0, \tag{15d}$$

defined over all vertices $v$ of $\hat{Z}_{\text{out}}^{-1}$. To prove validity of the vertex enumerative approach in this setting, we show that the algorithm correctly solves the relaxation of (13), and that the Karush-Kuhn-Tucker (KKT) multipliers associated with (15b) define a valid weighting vector. These two results are summarized by the following two propositions.

**PROPOSITION 6.1.** *An optimal solution of the relaxed bilevel program obtained by substituting the linearization* (14) *for the constraints of the outer level program in* (13) *occurs at a vertex of the polyhedron defined by the linearized constraints.*

**PROOF.** The optimal value function $g \colon \mathbb{R}^n \to \mathbb{R}$ of the inner level program in (13) is given by

$$g(z) = \min_{\substack{\mu,\,\lambda \geq 0 \\ e^T \lambda = 1}} d\bigl(h^{-1}(\hat{P}^T \lambda) + Q^T \mu, z\bigr).$$

The argument in the above minimization is jointly convex in $\mu$, $\lambda$, and $z$ because $h^{-1}$ is convex, because composition with affine functions preserves convexity, and because the pointwise maximum of any number of convex functions is convex. Minimization of the form

$$g(\varepsilon) = \inf_{x \in S} g(x, \varepsilon),$$

is convex in $\varepsilon$ whenever $g$ is jointly convex in $x$ and $\varepsilon$, $g$ is bounded from below on $S$, and $S$ is nonempty and convex (Fiacco and Kyparisis 1986, Proposition 2.1). By a change of sign in the objective function, problem (13) thus amounts to minimizing a concave function over a convex set. Because every global and local minimum value of a concave function is either attained at an extreme point of its feasible domain or the function is unbounded from below on a feasible ray (Rockafellar 1970, Theorem 32.3), the proof reduces to showing that the objective value of (13) is nonincreasing on any ray in the polyhedron $\hat{Z}_{\text{out}}^{-1}$ defined by the linearized constraints. Let $(\bar{W}, \bar{r})$ be a matrix-vector pair such that $\hat{Z}_{\text{out}}^{-1} = \{\bar{W} z \geq \bar{r}\}$. Let also $\bar{z}$ and $\bar{p}$ be vectors such that $\{\bar{z} + \alpha \bar{p} \colon \alpha \geq 0\}$ is a ray in $\hat{Z}_{\text{out}}^{-1}$ (i.e., $\bar{z} \in \hat{Z}_{\text{out}}^{-1}$, $\bar{p} \neq 0$, and $\bar{W} \bar{p} \geq 0$) and let $(\eta, \lambda, \mu)$ be an optimal solution to the inner level linear program in (13) with respect to $z = \bar{z}$. Then, because $\hat{Z}_{\text{out}}^{-1} \subseteq Z_{\text{out}}$, every row vector in $\bar{W}$ lies in $C^*$, so that $\bar{p} \in C$. Therefore, there exists $\bar{\mu} \geq 0$ such that $\bar{p} = Q \bar{\mu}$.

For such $\bar{\mu}$, $(\eta, \lambda, \mu + \alpha \bar{\mu})$ is a feasible point with objective value $\eta$ to the inner level linear program in (13) with respect to $z = \bar{z} + \alpha \bar{p}$. The objective value of (13) at any point on the ray is thus bounded from above by the objective value in the point $\bar{z}$ from which the ray emanates, and the proof is complete. □

Note that we have not used strict monotonicity of $h^{-1}$ in the above proof. The proof for Proposition 4.1 thus follows as a corollary by specializing $h$ to the identity function.

We now turn to showing that the KKT multipliers associated with (15b) define a normal to the inner approximation at the point where the maximum approximation error is attained.

**PROPOSITION 6.2.** *Let* $(\eta^*, \lambda^*, \mu^*)$ *such that* $\eta^* > 0$ *denote an optimal solution to* (15). *Let also* $\pi^*$ *and* $\rho^*$ *denote KKT multipliers at* $(\eta^*, \lambda^*, \mu^*)$ *associated with the constraints* (15b) *and* (15c), *respectively. Then, the hyperplane* $H(\pi^*, \rho^*)$ *supports* $\hat{Z}_{\text{in}}^{-1}$ *at* $y = h^{-1}(\hat{P}^T \lambda^*) + Q^T \mu^*$ *and has normal vector* $\pi^*$ *in* $C^* \setminus \{0\}$.

**PROOF.** Slater's condition (see, e.g., Boyd and Vandenberghe 2004, p. 226) is satisfied by strict feasibility and convexity of (15). Therefore, strong duality holds for problem (15) and its Lagrange dual problem. A primal-dual optimal solution to this pair of problems is under strong duality exactly characterized by the KKT conditions (see, e.g., Boyd and Vandenberghe 2004, pp. 243–244). The KKT conditions for (15) are identical with the corresponding conditions for a linearization of this problem of the form

$$\underset{\eta,\,\lambda,\,\mu}{\text{minimize}} \quad \eta$$

$$\begin{aligned}
\text{subject to} \quad & \eta e \geq h^{-1}(\hat{P}^T \lambda^*) + \nabla h^{-1}(\hat{P}^T \lambda^*)^T \\
& \quad \cdot \hat{P}^T (\lambda - \lambda^*) + Q^T \mu - v, \\
& e^T \lambda = 1, \\
& \eta, \lambda, \mu \geq 0.
\end{aligned} \tag{16}$$

By linearity, strong duality holds for this problem and its dual. The pair $(\pi^*, \rho^*)$ is thus an optimal solution to both the Lagrange dual problem to (15) and the linear programming dual of (16). The problem (16) can be put in the form of problem (3) by the substitution

$$P = \nabla h^{-1}(\hat{P}^T \lambda^*)^T \hat{P}^T, \tag{17}$$

and a translation of the coordinates in objective space according to

$$\bar{z} = z + \nabla h^{-1}(\hat{P}^T \lambda^*)^T \hat{P}^T \lambda^* - h^{-1}(\hat{P}^T \lambda^*), \tag{18}$$

where $z$ denotes an initial coordinate and $\bar{z}$ its translated counterpart.

By Proposition 4.1, the pair $(\pi^*, \rho^*)$ defines a hyperplane $H(\pi^*, \rho^*)$ with normal in $C^* \setminus \{0\}$ that supports

the linearized inner approximation at $\bar{y} = P^T \lambda^* + Q^T \mu^*$. Reversing the translation (18) and inserting (17) gives a point $y = h^{-1}(\hat{P}^T \lambda^*) + Q^T \mu^*$. Because $H(\pi^*, \rho^*)$ supports a linearization of $\hat{Z}_{\text{in}}^{-1}$, the statement follows by convexity of this set. □

# 7. Numerical Results

## 7.1. Test Problems

We evaluate the proposed algorithm with respect to three test problems. All problems are scalable in the number of objectives and can be made to comply with the assumptions stated in §2.2 by introducing some sufficiently large upper bounds on the variables.

PROBLEM 7.1. This is a randomly generated extension of test case 1 in Rennen et al. (2011) of the form

$$\underset{x}{\text{minimize}} \quad (x_1 \cdots x_n)^T$$
$$\text{subject to} \quad \sum_{j \neq i} (x_j - a_j)^2 - x_i \leq 0, \quad i = 1, \ldots, n,$$

where $a$ is an $n$-vector of integers drawn uniformly at random from $\{1, \ldots, n\}$. No bounds on the trade-off rate between objectives were imposed for this problem.

PROBLEM 7.2. This problem is of the form

$$\underset{x}{\text{minimize}} \quad \left(x_1^2 \cdots x_n^2\right)^T$$
$$\text{subject to} \quad \sum_{i=1}^{n} \left(x_i - (1 + \varepsilon)\right)^2 \leq 1,$$

for some small constant $\varepsilon > 0$. Alongside this problem, we also introduce the function $h(x) = (\sqrt{x_1} \cdots \sqrt{x_n})^T$, so that a transformation $h \circ f$ results in a problem with linear objectives. No bounds on the trade-off rate between objectives were imposed for this problem.

PROBLEM 7.3. This is an example of an IMRT optimization problem for a head and neck cancer case. Data for this problem was exported from the RayStation treatment planning system (RaySearch Laboratories, Stockholm, Sweden). The goal of IMRT is to deliver a highly conformal radiation dose to the tumor volume (see, e.g., Romeijn et al. 2008, Ehrgott et al. 2008, Ahnesjö et al. 2006). Target coverage must be traded against the sparing of radiosensitive organs in its vicinity. We consider the problem of optimizing incident energy fluence. This problem was posed of the form of (1) by assigning objectives and constraints to each anatomical structure. All objective and constraint functions were constructed as one-sided quadratic penalties of the deviations in voxel dose from a reference dose level, as made explicit in the

appendix. A bound $t_{ij} = 10^{-2}$ on the trade-off rate between all pairs of objectives $(i, j)$ was introduced to focus on the high-curvature region of the Pareto surface.

## 7.2. Computational Cost

The computational cost of the vertex enumerative algorithm and the facet enumerative algorithm was evaluated numerically with respect to Problems 7.1 and 7.3. We report the results of applying these two algorithms, in conjunction with, and without, the proposed upper-bounding procedure (abbreviated by the prefix "online"). Both algorithms were implemented in C++ using identical linear algebra routines and interfaced with Matlab. Quadratic programs, quadratically constrained linear programs, and quadratically constrained quadratic programs were solved using CPLEX 10.2 (ILOG, Sunnyvale, California) with default settings. Linear programs were solved using the primal simplex method built into SNOPT 7.2 (Stanford Business Software, Inc., Stanford, California), with problems sorted in descending order with respect to available upper bounds. These solves are amenable to parallelization, but for ease of comparison, all computations were run under 64-bit Linux on a single Intel Xeon 3 GHz processor core with hyperthreading disabled and with 32 GB of memory. A timeout of three hours was set for all processes, which kept overall running time reasonable.

The convex hull representation of the inner approximation was empirically observed to be a degenerate polytope, manifesting as multiple faces induced by near-identical hyperplanes. Because multiple solves over such hyperplanes do not contribute considerably to the solution of (2), any hyperplane identified as a duplicate within a tolerance of $10^{-5}$ was disregarded.

For each problem and algorithm we report number of beneath-and-beyond steps, number of linear programming solves, and CPU time, summed over 50 iterations of the sandwich algorithm. In addition, we report the upper bound $d_{\text{Haus}}(Z_{\text{in}}, Z_{\text{out}})$ on the approximation error as a function of iteration number. The numerical results for Problems 7.1 and 7.3 are summarized in Figures 6 and 7, respectively. We stress that our research implementation is not optimized for speed and the reported running times are for comparative purposes only.

The depicted results show that the vertex and the facet enumerative scheme are equivalent in terms of approximation quality. In terms of computational load, the combined effect of the vertex enumerative scheme and the proposed upper-bounding procedure results in an improvement that increases with problem dimension. For the two studied problems, the proposed enhancements translate into a reduction in the number of linear programming solves by one order
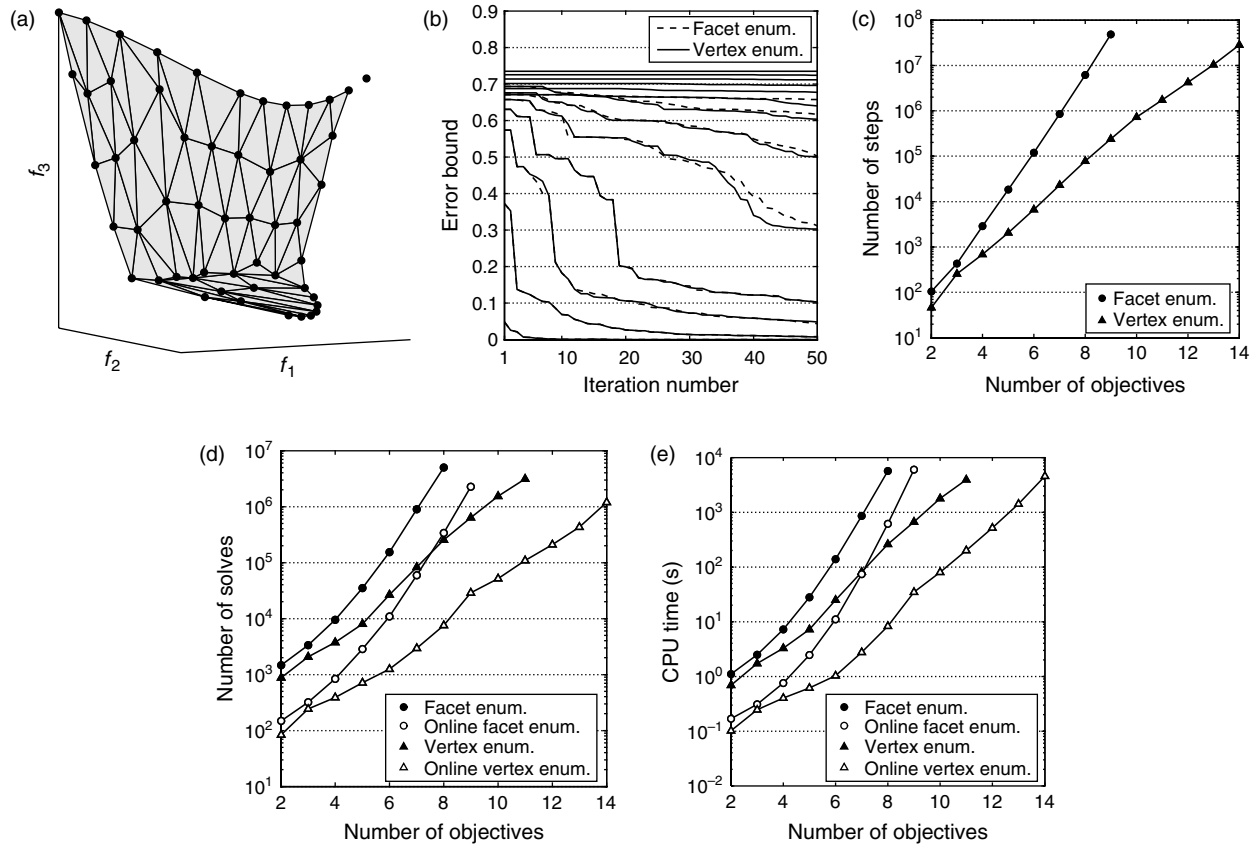
**Figure 6** **Numerical Results for 50 Iterations of the Sandwich Algorithm Applied to Problem 7.1: (a) Pareto Surface Representation at $n = 3$; (b) Upper Bound on the Approximation Error as a Function of Number of Objectives $n$, with the Lowermost Curve Corresponding to $n = 2$ and the Uppermost Curve Corresponding to $n = 14$; (c) Total Number of Beneath-and-Beyond steps vs. $n$; (d) Total Number of Linear Programming Solves vs. Surfaces $n$; (e) Total CPU Time vs. $n$**

of magnitude for dimensions beyond two, and by two orders of magnitude for dimensions beyond five. Correspondingly, the number of dimensions tractable at computational times within the order of minutes increases from about 6 to 11.

### 7.3. Approximation Quality

The quality of the Pareto surface approximation generated by a sandwich method was evaluated with respect to Problems 7.1 and 7.2. For Problem 7.1, we report the current bound on the approximation error given by $d_{\text{Haus}}(Z_{\text{in}}, Z_{\text{out}})$. We also report the exact approximation error according to Definition 3. Because all objectives are linear, this quantity coincides with the distance $d_{\text{Haus}}(Z_{\text{in}}, Z_+)$. The corresponding values are reported for Problem 7.2. The exact approximation error for this problem is given by the distance $d_{\text{Haus}}(\hat{Z}_{\text{in}}^{-1}, Z_+)$, because the composition $h \circ f$ is linear and $h$ is the inverse of $f$ on the domain $\mathbb{R}_{++}^n$. We also report the improved upper bound given by the distance $d_{\text{Haus}}(\hat{Z}_{\text{in}}^{-1}, \hat{Z}_{\text{out}}^{-1})$. The system (13b) was linearized about all points in $P \cup V(\hat{Z}_{\text{out}})^{-1}$, where $V(\hat{Z}_{\text{out}})^{-1}$ denotes the image of the vertex set of $\hat{Z}_{\text{out}}$ under the inverse transformation $h^{-1}$. The numerical

results for Problems 7.1 and 7.2 are summarized in Figures 8 and 9, respectively.

We can make several observations based on the depicted results. The distance $d_{\text{Haus}}(Z_{\text{in}}, Z_{\text{out}})$ is a rather pessimistic bound on the actual approximation error for the two studied problems, increasingly so with increasing problem dimension. The discrepancy between the actual approximation error and its bounds stems from nonlinearity of Problems 7.1 and 7.2. For multiobjective linear problems, the set $Z_+$ is polyhedral and the bound $d_{\text{Haus}}(Z_{\text{in}}, Z_{\text{out}})$ hence tight. Figure 9 demonstrates that positive curvature in the objectives can be partially compensated for by applying a concave transformation. The resulting improved bound is however not tight, both due to nonlinearity in the constraints of Problem 7.2 and the relaxation of the outer approximation into a polyhedral set.

## 8. Summary and Discussion

We have proposed a sandwich algorithm for approximating Pareto surfaces of convex multiobjective optimization problems based on enumerating the vertices
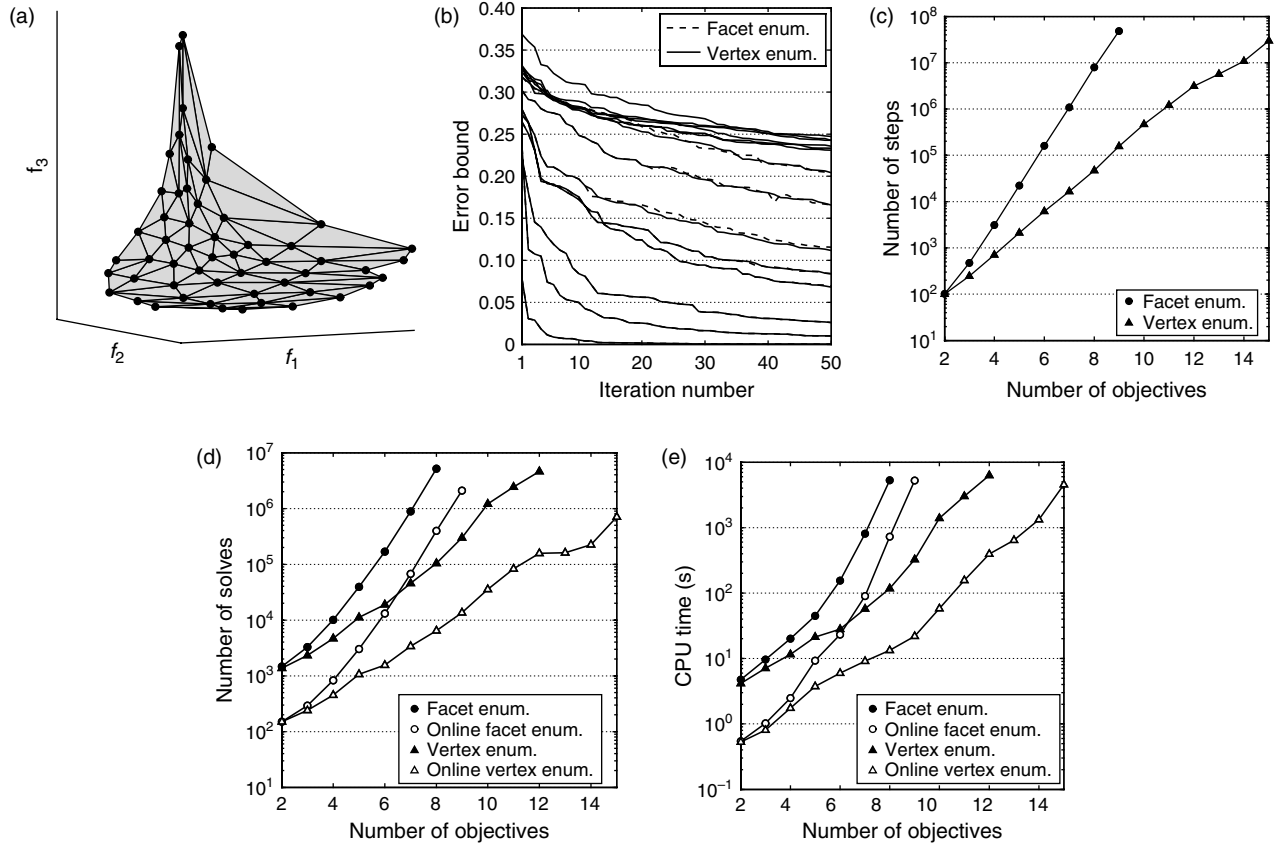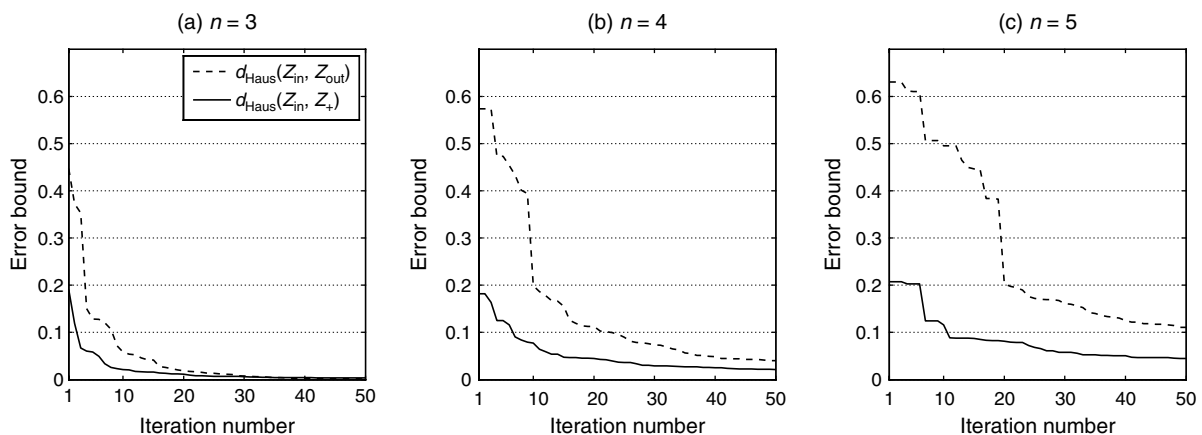
**Figure 7** Numerical Results for 50 Iterations of the Sandwich Algorithm Applied to Problem 7.3: (a) Pareto Surface Representation at $n = 3$; (b) Upper Bound on the Approximation Error as a Function of Number of Objectives $n$, with the Lowermost Curve Corresponding to $n = 2$ and the Uppermost Curve Corresponding to $n = 15$; (c) Total Number of Beneath-and-Beyond Steps vs. $n$; (d) Total Number of Linear Programming Solves vs. $n$; (e) Total CPU Time vs. $n$

of an outer polyhedral approximation of the Pareto surface. Additionally, an upper-bounding procedure was presented to reduce the number of subproblems required to solve the nonconvex optimization problem of calculating an upper bound on the current approximation error. This technique was made possible by implementing the polyhedral computations in an online fashion.

The proposed vertex enumerative method was contrasted to the previously suggested facet enumerative algorithm of Rennen et al. (2011). These two algorithms are both exact methods for maximizing the



**Figure 8** Numerical Results for the Sandwich Algorithm Applied to Problem 7.1: Upper Bounds on the Approximation Error vs. Iteration Number
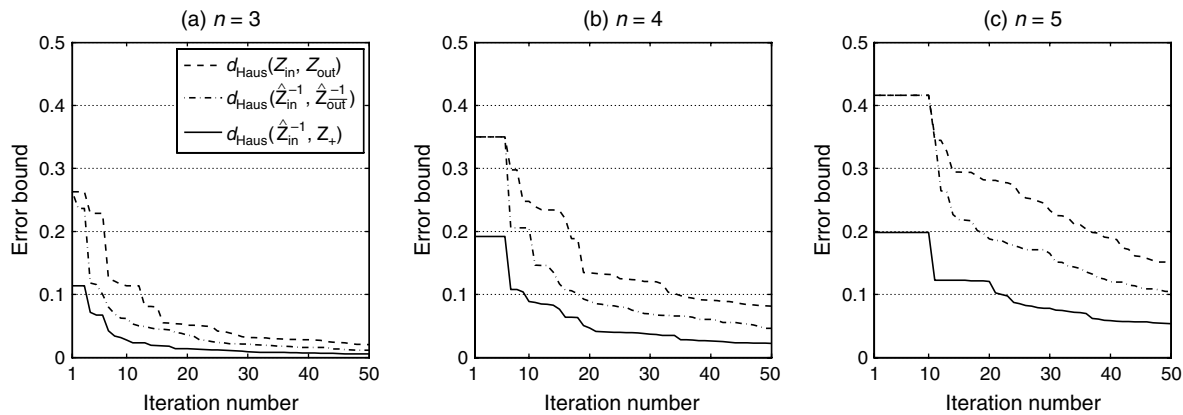
**Figure 9**     **Numerical Results for the Sandwich Algorithm Applied to Problem 7.2: Upper Bounds on the Approximation Error vs. Iteration Number**

improvement in the approximation error when generating a single Pareto optimal solution. As a result, the two methods are equivalent in terms of quality of output, as was verified experimentally. The vertex enumerative scheme was shown to provide an improvement in both worst-case complexity and practical performance of the sandwich algorithm. This improvement can be attributed to the fact that the vertex enumerative approach handles the normal vectors of the inner approximation—which is the more structurally complex polyhedron of the inner and outer approximations—as free variables in the linear programming subproblems. In the facet enumerative approach, these normal vectors are instead explicitly stated in the subproblem, leading to more costly polyhedral computations and a larger number of subproblems.

We also studied a previously described technique for yielding tighter bounds on the approximation error by concave transformations of the initial objectives. The vertex enumerative algorithm was shown to be compatible with this technique under a relaxation of the outer approximation into a polyhedral set. The resulting approximation scheme was demonstrated numerically to improve on the bound on the approximation error for the current Pareto surface representation.

With respect to the wider family of convex Pareto surface approximations techniques, the method developed in this paper falls within a class of methods that generalizes the chord rule for approximating convex univariate functions. Although previous methods within this class use halfspace representations of the inner and outer approximations, in resemblance with inner approximation methods, the proposed method instead uses vertex descriptions of these two polyhedrons, similar to previous outer approximation methods. Halfspace and vertex representations are equivalent under point-hyperplane duality. We therefore view the proposed technique as a natural dual to the previous algorithms in this class.

We conclude by summarizing the implications for the IMRT application. There is yet no widely accepted consensus on acceptable computational time for generating a discrete representation of the Pareto surface for this application. However, judging by a recent clinical evaluation (Craft et al. 2012) where total planning time was on the order of 10 minutes, running times that exceed much beyond the order of minutes appear unrealistic. In view of the numerical results for the enhancements proposed in this paper, it appears practical to solve the Pareto surface approximation problem for up to about 10 objectives. This limit coincides with the range of problem formulations commonly encountered in IMRT planning. We therefore envisage that sandwich algorithms will allow for better resolved models of the viable treatment options in the form of more accurately represented Pareto surfaces throughout the spectrum of problem formulations encountered in IMRT optimization.

### Appendix. Formulation of Problem 7.3
The patient volume was discretized into $5 \times 5 \times 5$ mm$^3$ volume elements (voxels) and the beam planes into $1 \times 1$ cm$^2$ surface elements (bixels). Dose kernels for five coplanar photon beams at equispaced gantry angles were computed using a pencil beam convolution technique based on singular value decomposition, similar to Bortfeld et al. (1993). The problem was posed of the form of (1) by taking the elements of $x$ to be the energy fluence per bixel and introducing a nonnegativity bound $x \geq 0$. All objectives and constraints

were modeled by minimum dose or maximum dose functions. A maximum dose function is defined as

$$f(x) = \sum_{i \in V} \Delta v_i (p_i^T x - d^{\text{ref}})_+^2, \tag{19}$$

where $V$ indexes the voxels included in the anatomical structure to which the function is assigned, $\Delta v_i$ denotes the relative volume of the $i$th voxel, and $p_i$ is a pencil beam kernel such that $p_i^T x$ is the planned dose in the $i$th voxel. Reversing the sign of the two terms in the argument of the plus function in (19) gives a minimum dose function.

The target structure was assigned with a minimum dose objective and a maximum dose objective, both with $d^{\text{ref}} = 70$ Gy, and a minimum dose constraint with $d^{\text{ref}} = 63$ Gy. A maximum dose objective was introduced with $d^{\text{ref}} = 0$ Gy for each healthy structure contained in the projection of the target volume onto the beam planes. The resulting number of objectives was 15. A constraint on global maximum dose at $d^{\text{ref}} = 77$ Gy was introduced by sampling 2% of all voxels in the patient volume uniformly at random, so to keep running times reasonable. The problem was posed as an inequality constrained quadratic program with 5,416 variables and 6,937 linear constraints by introducing auxiliary variables (see Carlsson et al. 2006). Scaling in the number of objectives was performed by aggregating positively correlated objectives. Each objective was first optimized individually. Objectives for healthy structures were then aggregated into composite functions being the direct sum of all constituent functions by iteratively grouping together the two objectives showing maximum degree of monotone association, as determined by Spearman rank correlation (see, e.g., Kendall 1962).

## References

Ahnesjö A, Hårdemark B, Isacsson U, Montelius A (2006) The IMRT information process—Mastering the degrees of freedom in external beam therapy. *Phys. Med. Biol.* 51(13):R381–R402.

Bellman R (1961) *Adaptive Control Processes: A Guided Tour* (Princeton University Press, Princeton, NJ).

Benson H (1998) An outer approximation algorithm for generating all efficient extreme points in the outcome set of a multiple objective linear programming problem. *J. Global Optim.* 13(1):1–24.

Bortfeld T, Schlegel W, Rhein B (1993) Decomposition of pencil beam kernels for fast dose calculations in three-dimensional treatment planning. *Med. Phys.* 20(2):311–318.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).

Bremner D, Fukuda K, Marzetta A (1998) Primal-dual methods for vertex and facet enumeration. *Discrete Comput. Geom.* 20(3):333–357.

Carlsson F, Forsgren A, Rehbinder H, Eriksson K (2006) Using eigenstructure of the Hessian to reduce the dimension of the intensity modulated radiation therapy optimization problem. *Ann. Oper. Res.* 148(1):81–94.

Chernykh O (1995) Approximating the Pareto-hull of a convex set by polyhedral sets. *Comp. Math. Math. Phys.* 35(8):1033–1039.

Cotrutz C, Lahanas M, Kappas C, Baltas D (2001) A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy. *Phys. Med. Biol.* 46(8):2161–2175.

Craft D (2010) Calculating and controlling the error of discrete representations of Pareto surfaces in convex multi-criteria optimization. *Phys. Medica* 26(4):184–191.

Craft D, Bortfeld T (2008) How many plans are needed in an IMRT multi-objective plan database? *Phys. Med. Biol.* 53(11):2785–2796.

Craft D, Halabi T, Bortfeld T (2005) Exploration of tradeoffs in intensity-modulated radiotherapy. *Phys. Med. Biol.* 50(24):5857–5868.

Craft D, Halabi T, Shih H, Bortfeld T (2006) Approximating convex Pareto surfaces in multiobjective radiotherapy planning. *Med. Phys.* 33(9):3399–3407.

Craft D, Halabi T, Shih H, Bortfeld T (2007) An approach for practical multiobjective IMRT treatment planning. *Int. J. Radiat. Oncol.* 69(5):1600–1607.

Craft D, Hong T, Shih H, Bortfeld T (2012) Improved planning time and plan quality through multi-criteria optimization for intensity modulated radiation therapy. *Int. J. Radiat. Oncol.* 82(1):e83–e90.

Das I, Dennis E (1997) A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. *Struct. Multidiscip. O.* 14(1):63–69.

Dempe S (2002) Nonconvex optimization and its applications. *Foundation of Bilevel Programming* (Kluwer Academic Publishers, Dordrecht, The Netherlands).

Edelsbrunner H (1987) *Algorithms in Combinatorial Geometry* (Springer-Verlag, Berlin).

Ehrgott M (2005) *Multicriteria Optimization, Lecture Notes in Economics and Mathematical Systems*, Vol. 491, 2nd ed. (Springer-Verlag, Berlin).

Ehrgott M, Löhne A, Shao L (2012) A dual variant of Benson's "outer approximation algorithm" for multiple objective linear programming. *J. Global Optim.* 52(4):757–778.

Ehrgott M, Shao L, Schöbel A (2011) An approximation algorithm for convex multi-objective programming problems. *J. Global Optim.* 50(3):397–416.

Ehrgott M, Güler Ç, Hamacher H, Shao L (2008) Mathematical optimization in intensity modulated radiation therapy. *4OR-Q. J. Oper. Res.* 6(3):199–262.

Engau A (2009) Tradeoff-based decomposition and decision-making in multiobjective programming. *Eur. J. Oper. Res.* 199(3):883–891.

Eskelinen P, Miettinen K, Klamroth K, Hakanen J (2010) Pareto navigator for interactive nonlinear multiobjective optimization. *OR Spectrum* 32(1):211–227.

Fiacco A, Kyparisis J (1986) Convexity and concavity properties of the optimal value function in parametric nonlinear programming. *J. Optim. Theory Appl.* 48(1):95–126.

Fruhwirth B, Bukkard R, Rote G (1989) Approximation of convex curves with application to the bicriterial minimum cost flow problem. *Eur. J. Oper. Res.* 42(3):326–338.

Grünbaum B (2003) *Convex Polytopes*, 2nd ed. (Springer-Verlag, New York).

Hong T, Craft D, Carlsson F, Bortfeld T (2008) Multicriteria optimization in intensity-modulated radiation therapy treatment planning for locally advanced cancer of the pancreatic head. *Int. J. Radiat. Oncol.* 72(4):1208–1214.

Hunt B, Wiecek M (2003) Cones to aid decision making in multicriteria programming. Tanino T, Tanaka T, Inuiguchi M, eds. *Multi-Objective Programming and Goal Programming* (Springer-Verlag, Berlin), 153–158.

Hunt M, Hsiung C, Spirou S, Chui C, Amols H, Ling C (2002) Evaluation of concave dose distributions created using an inverse planning system. *Int. J. Radiat. Oncol.* 54(3):953–962.

Kendall M (1962) *Rank Correlation Methods* (Griffin, London).

Küfer K-H, Scherrer A, Monz M, Alonso F, Trinkaus H, Bortfeld T, Thieke C (2003) Intensity-modulated radiotherapy—A large scale multi-criteria programming problem. *OR Spectrum* 25(2):223–249.

Klamroth K, Tind J, Wiecek M (2003) Unbiased approximation in multicriteria optimization. *Math. Method Oper. Res.* 56(3):413–437.

Lahanas M, Schreibmann E, Baltas D (2003) Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. *Phys. Med. Biol.* 48(17):2843–2871.

McMullen P (1970) The maximum number of faces of a convex polytope. *Mathematika* 17(2):179–184.

Miettinen K (1999) *Nonlinear Multiobjective Optimization* (Kluwer Academic Publishers, Dordrecht, The Netherlands).

Monz M (2006) Pareto navigation—Interactive multiobjective optimisation and its application in radiotherapy planning. Ph.D. thesis, Kaiserslautern University of Technology, Kaiserslautern, Germany.

Monz M, Kufer K-H, Bortfeld T, Thieke C (2008) Pareto navigation—Algorithmic foundation of interactive multi-criteria IMRT planning. *Phys. Med. Biol.* 53(4):985–998.

Pennanen T, Eckstein J (1997) Generalized Jacobians of vector-valued convex functions. Technical Report RRR 6-97, Rutgers University, New Brunswick, NJ.

Preparata F, Shamos M (1985) *Computational Geometry—An Introduction* (Springer-Verlag, New York).

Rennen G, van Dam E, den Hertog D (2011) Enhancement of sandwich algorithms for approximating higher-dimensional convex Pareto sets. *INFORMS J. Comput.* 23(4):493–517.

Rockafellar R (1970) *Convex Analysis* (Princeton University Press, Princeton, NJ).

Romeijn E, Dempsey J (2008) Intensity modulated radiation therapy treatment plan optimization. *TOP* 16(2):215–243.

Romeijn E, Dempsey J, Li J (2004) A unifying framework for multi-criteria fluence map optimization models. *Phys. Med. Biol.* 49(10):1991–2013.

Rote G (1992) The convergence rate of the sandwich algorithm for approximating convex functions. *Computing* 48(3):337–361.

Ruzika S, Wiecek M (2005) Approximation methods in multiobjective programming. *J. Optimiz. Theory App.* 126(3):473–501.

Sayin S (2000) Measuring the quality of discrete representations of efficient sets in multiple objective mathematical programming. *Math. Program.* 87(3):543–560.

Schandl B, Klamroth K, Wiecek M (2002) Norm-based approximation in multicriteria programming. *Comput. Math. Appl.* 44(7):925–942.

Serna J, Monz M, Küfer K-H, Thieke C (2009) Trade-off bounds for the Pareto surface approximation in multi-criteria IMRT planning. *Phys. Med. Biol.* 54(20):6299–6311.

Shao L, Ehrgott M (2008a) Approximating the nondominated set of an MOLP by approximately solving its dual problem. *Math. Method Oper. Res.* 68(2):469–492.

Shao L, Ehrgott M (2008b) Approximately solving multiobjective linear programmes in objective space and an application in radiotherapy treatment planning. *Math. Method Oper. Res.* 68(3):257–276.

Siem A, den Hertog D, Hoffmann A (2008) The effect of transformations on the approximation of univariate (convex) functions with applications to Pareto curves. *Eur. J. Oper. Res.* 189(2):347–362.

Solanki R, Appino P, Cohon J (1993) Approximating the non-inferior set in multiobjective linear programming problems. *Eur. J. Oper. Res.* 68(3):356–373.

Spalke T, Craft D, Bortfeld T (2009) Analyzing the main trade-offs in multiobjective radiation therapy treatment planning databases. *Phys. Med. Biol.* 54(12):3741–3754.

Stoer J, Witzgall C (1970) *Convexity and Optimization in Finite Dimensions I* (Springer-Verlag, Berlin).

Thieke C, Küfer K-H, Monz M, Scherrer A, Alonso F, Oelfke U, Huber P, Debus J, Bortfeld T (2007) A new concept for interactive radiotherapy planning with multicriteria optimization: First clinical evaluation. *Radiother. Oncol.* 85(2):292–298.