CHALMERS |  UNIVERSITY OF GOTHENBURG

PREPRINT 2009:3

Approximating the Pareto Optimal Set using a Reduced Set of Objective Functions

PETER LINDROTH
MICHAEL PATRIKSSON
ANN-BRITH STRÖMBERG

Department of Mathematical Sciences
Division of Mathematics
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Göteborg Sweden 2009

Preprint 2009:3

Approximating the Pareto Optimal Set using a Reduced Set of Objective Functions

Peter Lindroth, Michael Patriksson,
Ann-Brith Strömberg

Department of Mathematical Sciences
Division of Mathematics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Göteborg, Sweden
Göteborg, January 2009

Preprint 2009:3
ISSN 1652-9715

Matematiska vetenskaper
Göteborg 2009

Approximating the Pareto Optimal Set using a Reduced Set of Objective Functions

Peter Lindroth*† Michael Patriksson† Ann-Brith Strömberg†

Abstract

Real-world applications of multi-objective optimization often involve numerous objective functions. But while such problems are in general computationally very hard, it is often not necessary to find the Pareto optimal set exactly. A significantly smaller computational burden thus motivates the loss of precision if the size of the latter can be estimated. We describe a method for an optimal reduction of the set of objectives yielding a smaller problem whose Pareto optimal set is as similar as possible, w.r.t. Hausdorff distance, to the original Pareto optimal set. Using a new characterization of Pareto optimality, we derive a program whose solution represents an optimal reduction. We also propose an approximate, and computationally less demanding, formulation which utilizes correlations between the objectives and separates into two parts. The method is illustrated with a graphical example. Numerical results from an industrial instance concerning the configuration of heavy-duty trucks are also reported, demonstrating the usefulness of the method developed. The results show that multi-objective problems can be simplified with an induced error which can be measured.

Keywords: Multiple objective programming, many-objective optimization, Pareto optimality, objective space reduction, engineering optimization

1 Introduction and Pareto optimality

Engineering design problems typically involve the handling of a number of more or less conflicting criteria. Such a problem can mathematically be formulated as a multi-objective optimization problem with the standard notation

$$\min_{\mathbf{x}\in X} \{f_1(\mathbf{x}), \ldots, f_k(\mathbf{x})\}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ denotes a vector of decision variables, $X \subseteq \mathbb{R}^n$ is the set of feasible decision vectors (or the decision space), and each $f_i : X \to \mathbb{R}$ is an objective function to be minimized. The vector of objective functions is denoted $\mathbf{f} = \{f_1, \ldots, f_k\}$ and we define $\mathcal{K} = \{1, \ldots, k\}$. The objective space, Z, is defined as being the image of the decision space, i.e., $Z = \mathbf{f}(X) = \{\mathbf{z} = \mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in X\}$. If the objective functions are at least partially in conflict, i.e., there is no feasible decision vector minimizing all objectives simultaneously, then the optimal solution to (1) is not well-defined since there exists no natural complete ordering of vectors. However, there exists a set of decision vectors in which the best solution by rational judgments must be contained regardless of the relative importance of each single objective, namely the *Pareto optimal set*. In the sequel, we will refer to the problem (1) as the *original problem*.

Definition 1.1 (Pareto optimality) *Consider the problem* (1). *A vector* $\mathbf{x}^* \in X$ *is defined as* Pareto optimal *if there exists no vector* $\mathbf{x} \in X$ *such that* $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*)$, $i \in \mathcal{K}$, *and* $f_j(\mathbf{x}) < f_j(\mathbf{x}^*)$ *for at least one* $j \in \mathcal{K}$. *An objective vector* $\mathbf{z}^* = \mathbf{f}(\mathbf{x}^*)$ *is called* Pareto optimal *if the corresponding vector* $\mathbf{x}^*$ *is Pareto optimal. The set of Pareto optimal decision vectors* $\mathbf{x}^* \in X$ *is denoted by* $\mathcal{P} \subseteq X$.

Definition 1.2 (Domination) *A vector* $\mathbf{x} \in X$ *is said to* dominate *a vector* $\mathbf{y} \in X$ *if* $f_i(\mathbf{x}) \leq f_i(\mathbf{y})$, $i \in \mathcal{K}$, *and* $f_j(\mathbf{x}) < f_j(\mathbf{y})$ *for at least one* $j \in \mathcal{K}$.

The set of Pareto optimal objective vectors is bounded from below by the *ideal* vector, $\mathbf{z}^{\text{ideal}}$, and from above by the *nadir* vector, $\mathbf{z}^{\text{nadir}}$, defined as follows:

*Volvo 3P, Chassis and vehicle dynamics, SE-405 08, Gothenburg, Sweden. Corresponding author. E-mail: peter.lindroth@volvo.com, Phone: +46 31 3231063, Fax: +46 31 226203

†Mathematical Sciences, Chalmers University of Technology, and Mathematical Sciences, University of Gothenburg, SE-412 96 Gothenburg, Sweden. E-mail: {peter.lindroth, mipat, anstr}@chalmers.se

DEFINITION 1.3 (IDEAL AND NADIR VECTORS) *The* ideal *vector* $\mathbf{z}^{\text{ideal}} \in \mathbb{R}^k$ *and the* nadir *vector* $\mathbf{z}^{\text{nadir}} \in \mathbb{R}^k$ *consist of the component-wise minimum and maximum, respectively, of each objective over the Pareto optimal set, i.e.,*

$$\mathbf{z}^{\text{ideal}} = \Big(\min_{\mathbf{x}\in\mathcal{P}} f_1(\mathbf{x}), \ldots, \min_{\mathbf{x}\in\mathcal{P}} f_k(\mathbf{x})\Big), \qquad \mathbf{z}^{\text{nadir}} = \Big(\max_{\mathbf{x}\in\mathcal{P}} f_1(\mathbf{x}), \ldots, \max_{\mathbf{x}\in\mathcal{P}} f_k(\mathbf{x})\Big).$$

DEFINITION 1.4 (WEAK PARETO OPTIMALITY) *Consider the problem* (1). *A vector* $\mathbf{x}^* \in X$ *is defined as* weakly Pareto optimal *if there exists no other vector* $\mathbf{x} \in X$ *such that* $f_i(\mathbf{x}) < f_i(\mathbf{x}^*)$, $i \in \mathcal{K}$. *An objective vector* $\mathbf{z}^* = \mathbf{f}(\mathbf{x}^*)$ *is called* weakly Pareto optimal *if the corresponding vector* $\mathbf{x}^*$ *is weakly Pareto optimal. The set of weakly Pareto optimal vectors is denoted by* $\mathcal{P}_{\text{w}}$.

We now consider instances of the problem (1) with a discrete and bounded feasible set, $X = \{\mathbf{x}^j \mid j \in \mathcal{N}\}$, $\mathcal{N} = \{1, \ldots, N\}$. For such problems, we give an equivalent formulation of Pareto optimality which appears to be new. We start by introducing this characterization, which is more explicit than Definition 1.1 and is in some cases better suited for modeling practical optimization problems.

PROPOSITION 1.5 (PARETO SYSTEM) *Let* $N \geq 1$ *be an arbitrary integer,* $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$, *and* $M \gg 1$. *Then a vector* $\mathbf{x}^* \in X$ *is Pareto optimal in* (1) *if and only if the system*

$$f_i(\mathbf{x}^*) < f_i(\mathbf{x}^j) + M(1 - u_{ij}), \qquad j \in \mathcal{N},\ i \in \mathcal{K}, \tag{2a}$$
$$\textstyle\sum_{i\in\mathcal{K}} f_i(\mathbf{x}^*) \leq \sum_{i\in\mathcal{K}} f_i(\mathbf{x}^j) + M(1 - u_{0j}), \qquad j \in \mathcal{N}, \tag{2b}$$
$$\textstyle\sum_{i\in\{0\}\cup\mathcal{K}} u_{ij} \geq 1, \qquad j \in \mathcal{N}, \tag{2c}$$
$$u_{ij} \in \{0, 1\}, \qquad j \in \mathcal{N},\ i \in \{0\} \cup \mathcal{K}. \tag{2d}$$

is consistent.

PROOF. The Definition 1.1 of Pareto optimality can be restated as follows: for $\mathbf{x}^* \in X$ to be Pareto optimal, it is required that there is no $\mathbf{x} \in X$ such that $f_i(\mathbf{x}) \leq f_i(\mathbf{x}^*)$, $i \in \mathcal{K}$, and $\sum_{i\in\mathcal{K}} f_i(\mathbf{x}) < \sum_{i\in\mathcal{K}} f_i(\mathbf{x}^*)$ hold. Equivalently, for each $\mathbf{x} \in X$, no more than k out of these $k+1$ constraints may be fulfilled for $\mathbf{x}^* \in X$ to be Pareto optimal, or, by expressing the complement, for each $j \in \mathcal{N}$, at least one out of the $k+1$ constraints $f_i(\mathbf{x}^j) > f_i(\mathbf{x}^*)$, $i \in \mathcal{K}$, and $\sum_{i\in\mathcal{K}} f_i(\mathbf{x}^j) \geq \sum_{i\in\mathcal{K}} f_i(\mathbf{x}^*)$ must be fulfilled. The latter is equivalent to the system (2) being consistent. ∎

REMARK 1.6 *Pareto optimality can be replaced by weak Pareto optimality in Proposition* 1.5 *if the strict inequalities* (2a) *are replaced by (non-strict) inequalities and the inequalities* (2b) *are removed.*

By solving (1), we mean to find the set $\mathcal{P} \subseteq X$. As the number k of objectives increases the task of finding (a good approximation of) $\mathcal{P}$ becomes increasingly difficult (cf. [8]; cf. also [3] which claims the opposite for some special problems). For the instances of (1) of our interest the number k of objectives is large. Therefore our approach to solve these problems is based on the selection of a smaller set of objective functions. We next present an overview of such approaches.

1.1 Previous work

The notion of *redundant* (or *non-essential*) objectives was introduced by Gal and Leberling [12], referring to objective functions whose removal from the problem formulation would not affect the Pareto optimal set. It is proved that for linear multi-objective programs (i.e., with all f_i linear and X polyhedral) an objective whose cost vector is a positive linear combination of some other cost vectors is redundant. From an application point of view, however, it is noted in [11] that dropping redundant functions often will affect the final solution(s) obtained from standard multi-criteria decision making methods. In [17] it is stated that for non-linear problems, and especially in connection with so called interactive solution methods, it is more suitable to define redundancy based on "conflicts" between objectives; here, "no conflict" between a pair of objectives means that all feasible decision vectors ($\mathbf{x} \in X$) are sorted equally by the two objectives. Agrell [1] proposes a different definition of conflict, requiring the sortings to be equivalent only over the efficient set. More definitions of conflict can be found in [4]. Measures of interdependency are defined in [5]. Equal or opposite sorting over the decision space is required for two objectives to be interdependent. Deb and Saxena [8] propose a method for reducing the set of objectives based on the

Principal Component Analysis (PCA) technique; roughly speaking, the objectives in a reduced problem are the ones that retain as much variation as possible over the original objective space Z.

The above-mentioned works aim at a reformulation of the problem (1) such that the Pareto optimal set is retained. However, after a reduction of redundant objectives their number may still be too large. We introduce a measure of "partial redundancy" among objectives. Based on this measure we construct models with significally less than k objectives, having a lower computational complexity, and for which the Pareto optimal sets are *similar* to that of (1).

In [4] a method for reducing the set of objectives is presented. The aim is to drop objectives such that for all $\mathbf{x}, \mathbf{y} \in X$, $\mathbf{x}$ dominates $\mathbf{y}$ in the reduced problem if and only if $\mathbf{x}$ dominates $\mathbf{y}$ in the original problem. The method is extended to allow some changes in the dominance structure, leading to a change in the Pareto optimal set. The aim of this method is similar to ours. However, their focus lies on the dominance structure in the entire decision space X rather than on the minimization of the differences between the respective Pareto optimal sets.

The concept of *partial weighting* was introduced by Koski and Silvennoinen [14]. It can be seen as a generalization of the traditional scalarization technique [17] for solving multi-objective optimization problems: some of the objectives are replaced by their weighted sum. How the objectives should be grouped is however not clearly proposed. It is shown that the Pareto optimal set for the reduced problem, $\mathcal{P}^\beta$, is a subset of $\mathcal{P}$, but the authors do not analyze or characterize the vectors of $\mathcal{P}$ that are lost. We contribute with both a method for the selection of objectives for the reduced problem and characterizations of the Pareto optimal vectors in the original and reduced problems.

1.2 Motivation

The main contribution of this work is a practical method for simplifying a multi-objective optimization problem without losing too much information. In the literature, usually very strong assumptions are made when reducing the number of objective functions; it is often required that the Pareto optimal set is retained after the reduction. Practical engineering design problems usually include many objective functions and it may be well motivated to sacrifice some precision of the Pareto optimal set if the associated problem becomes significantly smaller and some measure of the error made is produced. We aim at finding a smaller representation—in terms of numbers of objectives—of (1) such that the respective Pareto optimal sets are as similar as possible. Our method can be used to preprocess optimization problems with many objectives and many feasible decision vectors. By studying a subset of the decision vectors, small enough to enable an exhaustive search for the Pareto optimal set, our method reduces the problem such that the difference between the respective Pareto optimal sets are held at a minimum. The reduced number of objectives can then be evaluated with respect to the complete set of decision vectors.

1.3 Outline

In Section 2 we investigate some relations between Pareto optimal sets when modifying the set of objective functions and propose distance measures between these. By showing that the Pareto optimal set shrinks when applying certain reduction rules we motivate the introduction of a *dominance tolerance* parameter. We also introduce a *centrality* parameter, aiming at focusing the approximation on the most important part of the resulting set. The number of objective functions and the value of the centrality parameter are fixed in the model, whence the variables defining the objectives in the reduced problem and the tolerance parameter constitute the decision variables for which we search optimal values.

In Section 3 we utilize Proposition 1.5 to define a binary linear program (an ideal model) whose solution represents an optimally reduced problem. The complexity of this model is far too high for a problem of practical industrial size to be solved in reasonable time. Therefore, in Section 4 we consider an approximation of the ideal model that can be solved efficiently.

In Section 5 and 6 the approximate model is applied to an illustrating example and to a realistic model, inspired by industrial applications, concerning configurations of heavy-duty trucks.

2 Reducing the set of objective functions

In this section we study how the Pareto optimal set varies with the set of objective functions. We propose a quality measure for the Pareto optimal set of the reduced problem and introduce a parameter for which

we search a value with the purpose of retaining the similarity between the *central* regions (cf. Section 2.3) of the respective Pareto optimal sets.

Reducing a multi-objective optimization problem by dropping objectives may induce an error, i.e., difference between the Pareto optimal set of the original problem and that of the reduced problem. In [6], a review of quality measures for an approximate Pareto optimal set (here denoted $\hat{\mathcal{P}}$) is presented. These are mostly used to evaluate metaheuristics. The authors note that there exist no standard measures. Further, most proposals aim at evaluating the performance of proposed heuristics for which the size of $\hat{\mathcal{P}}$ is significantly smaller than the size of $\mathcal{P}$. A good approximation normally means that $\hat{\mathcal{P}}$ is well spread over $\mathcal{P}$ and that the points in $\hat{\mathcal{P}}$ are also near-optimal, meaning that each point in $\hat{\mathcal{P}}$ is close to some point in $\mathcal{P}$. One measure, introduced in [7] and also used in [20], is defined by the metric

$$\mathit{Dist2}(\hat{\mathcal{P}}) = \max_{\mathbf{x}\in\mathcal{P}} \left\{ \min_{\mathbf{y}\in\hat{\mathcal{P}}} c(\mathbf{x},\mathbf{y}) \right\}, \tag{3}$$

where the function $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ measures the *closeness* of a pair of points. This means that *Dist2* is the largest deviation of a point in $\mathcal{P}$ from its nearest point in $\hat{\mathcal{P}}$. We wish to find out how well $\mathcal{P}^\beta$ (the Pareto optimal set of the reduced problem) approximates $\mathcal{P}$, but also the opposite: how well $\mathcal{P}$ approximates $\mathcal{P}^\beta$. Therefore we extend the measure *Dist2* to the well-known Hausdorff distance measure, which measures how distant two non-empty compacts sets are in a certain metric. With $d(\cdot,\cdot)$ denoting the distance metric, the Hausdorff distance, between the closed, discrete, and non-empty sets E and F, is defined as

$$d_H(E,F) = \max\left\{ \max_{\mathbf{u}\in E}\min_{\mathbf{v}\in F} d(\mathbf{u},\mathbf{v}); \max_{\mathbf{v}\in F}\min_{\mathbf{u}\in E} d(\mathbf{v},\mathbf{u}) \right\}. \tag{4}$$

Given the multi-objective optimization problem (1), the question of how many objective functions that are required to characterize $\mathcal{P}$ is raised in [9]. For strictly quasi-convex objectives it is shown that for $n = 2$, three objectives are sufficient. However, without assuming convexity, or even continuity, of the objective functions it is clear that, in principle, one objective function is enough, e.g., the indicator function

$$\xi(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \notin \mathcal{P}, \\ 0, & \text{if } \mathbf{x} \in \mathcal{P}. \end{cases}$$

Since our method is intended to be applied to a subset of the complete decision space—hoping that the reduced set of functions are good representatives also in the complete decision space—the indicator function is not a wise choice.

In machine learning language [18], the wish is to find a hypothesis (a set of new objective functions) that fits the entire data (the complete set of decision vectors) well, i.e., leads to approximately the same Pareto optimal set, and not just fits the data measured (the smaller subset of the set of decision vectors). For this reason, hypotheses that over-fit the data must be avoided, meaning that there exists an alternative hypothesis that explains the measured data worse but explains the entire data better. In our application, the set of all hypotheses are all functions $\mathbf{g} : X \to \mathbb{R}^r$, where $r < k$ is the number of objectives in the reduced problem. In this set of functions we want to find the function $\mathbf{g}$ that approximates the Pareto optimal set the best over the entire set of data. We have chosen to limit the set of hypotheses to those for which the objective functions of the reduced problem are linear combinations of the original ones. This leads to a fairly simple hypothesis that still seems reasonable. Furthermore, we have restricted the weights of the linear combinations to be uniform so that any set of two or more objectives may be replaced by their mean only. Thus, we end up with a procedure similar to that introduced in [14]: a partial weighting of the original objectives.

2.1 The feasible choices of objectives for the reduced problem

Let the power set, i.e., the set of all subsets, of $\mathcal{K}$ be denoted by $2^{\mathcal{K}} = \{K_1, \ldots, K_{2^k}\}$. Our aim is to find a subset of $2^{\mathcal{K}}$ with at most r elements, represented by $\{s_1, \ldots, s_r\} \subset \hat{\mathcal{K}} = \{1, \ldots, 2^k\}$, and defining which collections of original objective functions to aggregate into objectives in the reduced problem. In other words, the aim is to reduce the set $\{f_1, \ldots, f_k\}$ of objective functions to $\{g_{s_1}, \ldots, g_{s_r}\}$, where $r < k, g_{s_j} = \frac{1}{|K_{s_j}|}\sum_{i\in K_{s_j}} f_i$, and where the set K_{s_j} indicates which of the original objectives that are included in the objective g_{s_j} in the reduced problem. We also require that each objective function f_i is included among the terms of some linear combination of functions g_{s_j}, i.e., $\cup_{j=1}^{r} K_{s_j} = \mathcal{K}$. We introduce

the binary variables

$$\beta_p = \begin{cases} 1, & \text{if collection } p \text{ of objectives is chosen,} \\ 0, & \text{otherwise,} \end{cases} \qquad p \in \hat{\mathcal{K}}. \tag{5}$$

Let $A \in \mathbb{B}^{k \times 2^k}$ be an incidence matrix with elements

$$a_{ip} = \begin{cases} 1, & \text{if objective } f_i \text{ is contained in collection } p, \\ 0, & \text{otherwise,} \end{cases} \qquad i \in \mathcal{K},\ p \in \hat{\mathcal{K}}.$$

The feasible choices of collections, for a prescribed maximal number r of objectives in the reduced problem, are then defined by all vectors $\boldsymbol{\beta}$ fulfilling

$$\begin{aligned} A\boldsymbol{\beta} &\geq \mathbf{1}^k, \\ \boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}^{2^k} &\leq r, \\ \boldsymbol{\beta} &\in \{0,1\}^{2^k}, \end{aligned} \tag{6}$$

$\mathbf{1}^n$ denoting the n-vector of ones.

In the following, we will instead of the previous notation $\hat{\mathcal{P}}$ denote the Pareto optimal set of the reduced problem defined by $\boldsymbol{\beta}$ as $\mathcal{P}^\beta$.

2.2 Relations between objectives and Pareto optimal sets

We next investigate how the set of Pareto optimal decision vectors varies with the set of objective functions. Dropping an objective function (illicit according to our rules) will reduce the set of weakly Pareto optimal points for problems (1) with convex objective functions; this is a consequence of, e.g., Corollary 1 in [15]. It is true also for general problems, guaranteed by the following proposition.

PROPOSITION 2.1 *Consider problem* (1). *Let* $\mathcal{K}^\beta = \{j \in \mathcal{K} \,|\, \beta_j = 1\}$ *indicate a subset of the objective functions. Let* $\mathcal{P}_{\mathrm{w}} \subseteq X$ ($\mathcal{P}^\beta_{\mathrm{w}} \subseteq X$) *be the set of weakly Pareto optimal vectors corresponding to the set* $\mathcal{K}$ ($\mathcal{K}^\beta$). *Then,* $\mathcal{P}^\beta_{\mathrm{w}} \subseteq \mathcal{P}_{\mathrm{w}}$.

PROOF. Suppose that $\mathbf{y}^* \in \mathcal{P}^\beta_{\mathrm{w}}$. This implies that $\nexists \mathbf{y} \in X$ such that $f_i(\mathbf{y}) < f_i(\mathbf{y}^*)$, $i \in \mathcal{K}^\beta$. Since $\mathcal{K}^\beta \subseteq \mathcal{K}$, $\nexists \mathbf{y} \in X$ such that $f_i(\mathbf{y}) < f_i(\mathbf{y}^*)$, $i \in \mathcal{K}$, and thus, $\mathbf{y}^* \in \mathcal{P}_{\mathrm{w}}$. ∎

REMARK 2.2 *Note that the analogous statement to Proposition* 2.1 *where weak Pareto optimality is replaced by Pareto optimality is false. Consider* (1) *with* $k = n = 2, \{f_1(\mathbf{x}), f_2(\mathbf{x})\}$ *over* X, *with* $f_1(\mathbf{x}) = x_1$, $f_2(\mathbf{x}) = x_2$, *and* $X = \{(1,2);(2,1);(3,1)\}$. *Then* $\mathcal{P} = \{(1,2);(2,1)\}$. *By dropping* f_1 (*i.e., for* $\beta = (0,1)$) *we obtain* $\mathcal{P}^\beta = \{(2,1);(3,1)\} \not\subseteq \mathcal{P}$.

The following proposition (proved in [14]) guarantees, however, that the set of Pareto optimal solutions is reduced or kept constant when a subset of the objective functions is replaced by their positively weighted sum. We conclude that if constructing a reduced problem by replacing certain subsets of objective functions with their respective weighted means, then $\mathcal{P}^\beta \subseteq \mathcal{P}$ holds.

PROPOSITION 2.3 *Consider problem* (1). *Let* $\mathcal{K}^\beta$ *indicate the reduced set of objective functions where the first two objectives are replaced by a positively weighted sum, i.e.,* $\{w_1 f_1 + w_2 f_2, f_3, \ldots, f_k\}$, $w_1, w_2 > 0$. *Let* $\mathcal{P} \subseteq X$ ($\mathcal{P}^\beta \subseteq X$) *be the set of Pareto optimal vectors corresponding to* $\mathcal{K}$ ($\mathcal{K}^\beta$). *Then,* $\mathcal{P}^\beta \subseteq \mathcal{P}$.

2.3 Centrality

Reducing the set of objective functions through partial weighting typically leads to the loss of extreme Pareto optimal solutions, i.e., solutions with a very good value in one objective but arbitrarily poor values in others. These are probably not attractive anyway when a final solution is picked from the set of Pareto optimal solutions. In order to obtain a set $\mathcal{P}^\beta$ that differs the least from $\mathcal{P}$ for the solutions that are likely to be chosen as final solutions, a centrality parameter $\rho \in [0,1]$ is defined, and the vectors $\mathbf{x} \in E$, where E is any subset of X (e.g. $\mathcal{P}$ or $\mathcal{P}^\beta$), are partitioned into a *ρ-central* and a *non-ρ-central* part.

DEFINITION 2.4 (ρ-CENTRALITY) *Consider the multi-objective optimization problem* (1). *The* ρ-central *part* $E^\rho \subseteq E$ *of a set* $E \subseteq X$ *is defined as*

$$E^\rho = \left\{ \mathbf{x} \in E \;\middle|\; f_i(\mathbf{x}) \leq (1-\rho) z_i^{\text{nadir}} + \rho z_i^{\text{ideal}},\ i \in \mathcal{K} \right\}. \tag{7}$$

PROPOSITION 2.5 *For any subset* $E \subseteq X$ *it holds that* $E^\rho = E \cap X^\rho$.

In words, the ρ-central part of the Pareto optimal set, $\mathcal{P}^\rho = \mathcal{P} \cap X^\rho$, consists of the Pareto optimal vectors that have no objective function value relative to the span of that objective over $\mathcal{P}$, closer than ρ to any component of the *nadir* vector. As special cases we have for a set $E \subseteq X$, $E^0 = \{\mathbf{x} \in E \;|\; \mathbf{f}(\mathbf{x}) \leq \mathbf{z}^{\text{nadir}}\}$ and $E^1 = \{\mathbf{x} \in E \;|\; \mathbf{f}(\mathbf{x}) \leq \mathbf{z}^{\text{ideal}}\}$ ($= \emptyset$ if the objectives are partially in conflict). Figure 1 illustrates the concept of centrality.

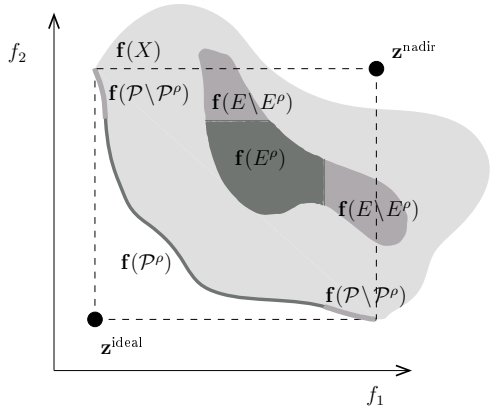


Figure 1: An illustration of ρ-centrality in the objective space for two objective functions. Here, $\rho \approx 0.2$ and E is a general subset of X.

2.4 Dominance tolerance

We want to approximate the Pareto optimal set as well as is possible using the principles described above. Therefore, a tolerance parameter $\tau \geq 0$ is introduced with the aim of enlarging the Pareto optimal set of the reduced problem, so that the resulting set differ the least from the Pareto optimal set corresponding to the original set of objectives. We define the *τ-Pareto optimal set*, $\mathcal{P}_\tau$, as follows.

DEFINITION 2.6 (τ-PARETO OPTIMALITY) *Consider the problem* (1). *For* $\tau \geq 0$, *a vector* $\mathbf{x}^* \in X$ *is defined as* τ-Pareto optimal *if there exists no vector* $\mathbf{x} \in X$ *such that* $f_i(\mathbf{x}) + \tau \leq f_i(\mathbf{x}^*)$, $i \in \mathcal{K}$, *and* $f_j(\mathbf{x}) + \tau < f_j(\mathbf{x}^*)$ *for at least one* $j \in \mathcal{K}$. *An objective vector* $\mathbf{z}^* = \mathbf{f}(\mathbf{x}^*)$ *is called* τ-Pareto optimal *if the corresponding vector* $\mathbf{x}^*$ *is* τ*-Pareto optimal. The set of* τ*-Pareto optimal decision vectors is denoted by* $\mathcal{P}_\tau \subseteq X$.

By construction, $\mathcal{P} \subseteq \mathcal{P}_{\tilde{\tau}} \subseteq \mathcal{P}_\tau$ holds for all $\tau \geq \tilde{\tau} \geq 0$. Moreover, we write $\mathcal{P}_\tau^\beta$ for $\left(\mathcal{P}^\beta\right)_\tau$.

2.5 Summarizing our goal

We now summarize our goal: Given the problem (1), the number $r < k$ of objective functions in the reduced problem and the value $\rho \in [0,1]$ of the centrality parameter, we wish to find the optimal set of collections represented by $\{s_1, \ldots, s_r\} \subset \hat{\mathcal{K}}$, and the value of the tolerance parameter $\tau \geq 0$, such that the Hausdorff distance[1] $d_H(\mathbf{f}(\mathcal{P}^\rho), \mathbf{f}(\mathcal{P}_\tau^{\beta,\rho}))$ is minimized (see Figure 2). The problem is mathematically

[1] We write $\mathcal{P}_\tau^{\beta,\rho}$ for $(\mathcal{P}_\tau^\beta)^\rho$.

stated as

$$\begin{aligned} \underset{\boldsymbol{\beta},\tau}{\text{minimize}} \quad & \delta(\boldsymbol{\beta},\tau) := d_H(\mathbf{f}(\mathcal{P}^{\rho}), \mathbf{f}(\mathcal{P}_{\tau}^{\beta,\rho})), \\ \text{subject to} \quad & A\boldsymbol{\beta} \geq \mathbf{1}^k, \\ & \boldsymbol{\beta}^{\mathrm{T}}\mathbf{1}^{2^k} \leq r, \\ & \boldsymbol{\beta} \in \{0,1\}^{2^k}, \\ & \tau \geq 0. \end{aligned} \tag{8}$$

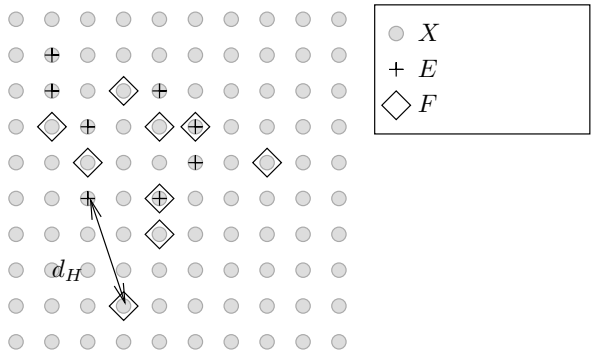


Figure 2: An illustration of the Hausdorff distance in the Euclidean metric between the two discrete sets $E \subseteq X$ and $F \subseteq X$.

3 The ideal model

Starting from the characterization of Pareto optimality in Proposition 1.5, we construct the "ideal" model for solving the problem (8) as a binary linear program.

3.1 An explicit formulation of Pareto optimality

Clearly, $\mathcal{P}_{\tau}^{\beta,\rho}$ depends on the choices of $\boldsymbol{\beta}$ and τ. Thus, a well-posed optimization problem requires an explicit formulation of Pareto optimality. Such a formulation is provided through Proposition 1.5. In the following subsections we develop explicit constraints that separate τ-Pareto optimal and non-τ-Pareto optimal vectors. We start with a formulation of the general problem (1) which is, in Section 3.2, applied to the reduced problem.

Using the system (2), we construct a consistent system of inequalities which is used to partition the set X into a Pareto optimal and a non-Pareto optimal set. Consistency of (2a), (2b) and (2d) is required and binary variables w_ℓ, $\ell \in \mathcal{N}$, are introduced indicating whether also (2c) holds for a specific vector $\mathbf{x}^\ell$. We let the vectors $\mathbf{x}^j$ and $\mathbf{x}^\ell$ be two specific vectors in X and define the decision variables

$$u_{ij\ell} = \begin{cases} 1, & \text{if } f_i(\mathbf{x}^\ell) < f_i(\mathbf{x}^j), \\ 0, & \text{otherwise,} \end{cases} \qquad j,\ell \in \mathcal{N},\ i \in \mathcal{K},$$

$$u_{0j\ell} = \begin{cases} 1, & \text{if } \sum_{i\in\mathcal{K}} f_i(\mathbf{x}^\ell) \leq \sum_{i\in\mathcal{K}} f_i(\mathbf{x}^j), \\ 0, & \text{otherwise,} \end{cases} \qquad j,\ell \in \mathcal{N},$$

$$v_{j\ell} = \begin{cases} 1, & \text{if } \mathbf{x}^j \text{ does not dominate } \mathbf{x}^\ell \text{ (i.e., if } \sum_{i\in\{0\}\cup\mathcal{K}} u_{ij\ell} \geq 1), \\ 0, & \text{if } \mathbf{x}^j \text{ dominates } \mathbf{x}^\ell, \end{cases} \qquad j,\ell \in \mathcal{N},$$

$$w_\ell = \begin{cases} 1, & \text{if } \mathbf{x}^\ell \in \mathcal{P} \text{ (i.e., if } v_{j\ell} = 1\ \forall j), \\ 0, & \text{if } \mathbf{x}^\ell \notin \mathcal{P}, \end{cases} \qquad \ell \in \mathcal{N}.$$

Using these variable declarations we formulate, for each $\ell \in \mathcal{N}$, the following system of inequalities:

$$
\begin{aligned}
-Mu_{ij\ell} \le f_i(\mathbf{x}^\ell) - f_i(\mathbf{x}^j) < M(1-u_{ij\ell}), &\qquad j \in \mathcal{N},\ i \in \mathcal{K}, &\text{(9a)}\\
-Mu_{0j\ell} < \textstyle\sum_{i\in\mathcal{K}} f_i(\mathbf{x}^\ell) - \sum_{i\in\mathcal{K}} f_i(\mathbf{x}^j) \le M(1-u_{0j\ell}), &\qquad j \in \mathcal{N}, &\text{(9b)}\\
v_{j\ell} \le \textstyle\sum_{i\in\{0\}\cup\mathcal{K}} u_{ij\ell} \le (k+1)v_{j\ell}, &\qquad j \in \mathcal{N}, &\text{(9c)}\\
Nw_\ell \le \textstyle\sum_{j\in\mathcal{N}} v_{j\ell} \le w_\ell + N - 1, & &\text{(9d)}\\
u_{ij\ell},\, v_{j\ell},\, w_\ell \in \{0,1\}, &\qquad j \in \mathcal{N},\ i \in \{0\}\cup\mathcal{K}. &\text{(9e)}
\end{aligned}
$$

We have the following result:

PROPOSITION 3.1 *For the problem* (1)*, the system* (9) *of inequalities partitions the vectors* $\mathbf{x}^\ell \in X$, $\ell \in \mathcal{N}$*, into a Pareto optimal and a non-Pareto optimal set, which are distinguished by the values of the variables* w_ℓ, $\ell \in \mathcal{N}$.

REMARK 3.2 *The strict inequalities in* (9a) *can be replaced by (non-strict) inequalities if the constant*

$$\xi = \min\left\{|f_i(\mathbf{x}^j) - f_i(\mathbf{x}^\ell)| :\ i \in \mathcal{K},\ j,\ell \in \mathcal{N},\ |f_i(\mathbf{x}^j) - f_i(\mathbf{x}^\ell)| > 0\right\} > 0,$$

is added on the left-hand side of each inequality. Analogous definitions can be used to eliminate the strict inequalities in (9b)*,* (10a)*,* (11a)*,* (11b)*,* (15a)*, and* (19a).

REMARK 3.3 *If the goal is to find the Pareto optimal set* $\mathcal{P} \subseteq X$*, half of the inequalities in* (9) *can be replaced by an objective function, according to*

$$
\begin{aligned}
\text{maximize} \quad & \textstyle\sum_{\ell\in\mathcal{N}} w_\ell, & & &\\
\text{subject to} \quad & f_i(\mathbf{x}^\ell) - f_i(\mathbf{x}^j) < M(1-u_{ij\ell}), & j,\ell \in \mathcal{N},\ i \in \mathcal{K}, &\text{(10a)}\\
& \textstyle\sum_{i\in\mathcal{K}} f_i(\mathbf{x}^\ell) - \sum_{i\in\mathcal{K}} f_i(\mathbf{x}^j) \le M(1-u_{0j\ell}), & j,\ell \in \mathcal{N}, &\text{(10b)}\\
& v_{j\ell} \le \textstyle\sum_{i\in\{0\}\cup\mathcal{K}} u_{ij\ell}, & j,\ell \in \mathcal{N}, &\text{(10c)}\\
& Nw_\ell \le \textstyle\sum_{j\in\mathcal{N}} v_{j\ell}, & \ell \in \mathcal{N}, &\text{(10d)}\\
& u_{ij\ell},\, v_{j\ell},\, w_\ell \in \{0,1\}, & j,\ell \in \mathcal{N},\ i \in \{0\}\cup\mathcal{K}. &\text{(10e)}
\end{aligned}
$$

3.2 An explicit formulation of Pareto optimality for the reduced problem

There are 2^k possible objectives in the reduced problem; the ones chosen are indicated by the values of β_p, $p \in \hat{\mathcal{K}} = \{1,\dots,2^k\}$. However, we are interested in τ-Pareto optimality for the reduced problem, wherefore $\{f_1,\dots,f_k\}$ in (9) cannot be directly replaced with $\{\beta_1 g_1,\dots,\beta_{2^k} g_{2^k}\}$. The reason is that if some objective function h_i maps all $\mathbf{x} \in X$ to the same value, then for any $\mathbf{x}^* \in X$ and $\tau > 0$ there exists no $\mathbf{x} \in X$ such that $h_i(\mathbf{x}) + \tau \le h_i(\mathbf{x}^*)$; thus all $\mathbf{x} \in X$ will be τ-Pareto-optimal. This will be the case for every objective function g_p corresponding to $\beta_p = 0$ (since if so, $\beta_p g_p \equiv 0$). Therefore the system (9) must be modified so that all inequalities involving terms of an objective g_p such that $\beta_p = 0$ become redundant. We redefine the u, v and w variables according to

$$
\begin{aligned}
u_{pj\ell} &= \begin{cases} 1, & \text{if } g_p(\mathbf{x}^\ell) < g_p(\mathbf{x}^j) + \tau, \text{ and collection } p \text{ is chosen,}\\ 0, & \text{otherwise,}\end{cases} && j,\ell \in \mathcal{N},\ p \in \hat{\mathcal{K}},\\
u_{0j\ell} &= \begin{cases} 1, & \text{if } \textstyle\sum_{p\in\hat{\mathcal{K}}} \beta_p g_p(\mathbf{x}^\ell) \le \sum_{p\in\hat{\mathcal{K}}} (\beta_p g_p(\mathbf{x}^j) + \tau),\\ 0, & \text{otherwise,}\end{cases} && j,\ell \in \mathcal{N},\\
v_{j\ell} &= \begin{cases} 1, & \text{if } \mathbf{x}^j \text{ does not } \tau\text{-dom. } \mathbf{x}^\ell \text{ (i.e., if } \textstyle\sum_{p\in\{0\}\cup\hat{\mathcal{K}}} u_{pj\ell} \ge 1),\\ 0, & \text{if } \mathbf{x}^j \ \tau\text{-dominates } \mathbf{x}^\ell,\end{cases} && j,\ell \in \mathcal{N},\\
w_\ell &= \begin{cases} 1, & \text{if } \mathbf{x}^\ell \in \mathcal{P}_\tau^\beta \text{ (i.e., if } v_{j\ell} = 1\ \forall j),\\ 0, & \text{if } \mathbf{x}^\ell \notin \mathcal{P}_\tau^\beta,\end{cases} && \ell \in \mathcal{N}.
\end{aligned}
$$

Letting β_p, $p \in \hat{\mathcal{K}}$, define the objectives for the reduced system, the following system for $\ell \in \mathcal{N}$, possesses a solution with $w_\ell = 1$ if and only if $\mathbf{x}^\ell \in \mathcal{P}_\tau^\beta$:

$$-Mu_{pj\ell} \le M(1-\beta_p) + \beta_p g_p(\mathbf{x}^\ell) - \left[\beta_p g_p(\mathbf{x}^j) + \tau\right] < 2M(1-u_{pj\ell}), \qquad j \in \mathcal{N},\ p \in \hat{\mathcal{K}}, \tag{11a}$$

$$-Mu_{0j\ell} < \textstyle\sum_{p\in\hat{\mathcal{K}}} \beta_p g_p(\mathbf{x}^\ell) - \sum_{p\in\hat{\mathcal{K}}} \left[\beta_p g_p(\mathbf{x}^j) + \tau\right] \le M(1-u_{0j\ell}), \qquad j \in \mathcal{N}, \tag{11b}$$

$$v_{j\ell} \le \textstyle\sum_{p\in\{0\}\cup\hat{\mathcal{K}}} u_{pj\ell} \le (r+1)v_{j\ell}, \qquad j \in \mathcal{N}, \tag{11c}$$

$$Nw_\ell \le \textstyle\sum_{j\in\mathcal{N}} v_{j\ell} \le w_\ell + N - 1, \tag{11d}$$

$$u_{pj\ell},\, v_{j\ell},\, w_\ell \in \{0,1\}, \qquad j \in \mathcal{N},\ p \in \{0\}\cup\hat{\mathcal{K}}. \tag{11e}$$

The main difference between the systems (9) and (11) is the parameter τ which is introduced to define the τ-Pareto optimal set. Note also the difference between the inequalities (9a) and (11a), where terms are added to make sure that $u_{pj\ell} = 0$ whenever $\beta_p = 0$. Also, the constant $k+1$ from (9c) is in (11c) replaced by $r+1$. From the arguments above we have the following result.

PROPOSITION 3.4 *Let the set $\{g_1, \ldots, g_{2^k}\}$ consist of the potential objective functions for a reduced version of* (1) *and let $\boldsymbol{\beta} \in \{0,1\}^{2^k}$ indicate which (at most r) of the objectives are chosen. Then a feasible solution to* (11) *partitions the vectors $\mathbf{x}^\ell \in X$, $\ell \in \mathcal{N}$ into a τ-Pareto optimal and a non-τ-Pareto optimal set.*

3.3 An explicit formulation of centrality

The objective function $d_H(\mathbf{f}(\mathcal{P}^\rho), \mathbf{f}(\mathcal{P}_\tau^{\beta,\rho}))$ to minimize over the decision variables $\boldsymbol{\beta}$ and τ depends on the centrality parameter ρ. Therefore, centrality according to Definition 2.4 must also be characterized explicitly using constraints. In the following subsections we derive a system of linear inequalities which partitions the set X into a ρ-central and a non-ρ-central part. Provided that $\dim(\mathcal{P}) = k$, using the replacement

$$f_i(\mathbf{x}) := \frac{f_i(\mathbf{x}) - z_i^{\text{ideal}}}{z_i^{\text{nadir}} - z_i^{\text{ideal}}}, \quad i \in \mathcal{K}, \tag{12}$$

it follows that $\mathbf{f}(X) \subseteq \mathbb{R}_+^k$ and $\mathbf{f}(\mathbf{x}) \in [0,1]^k$, $\mathbf{x} \in \mathcal{P}$. Proposition 3.5 then guarantees the existence of a linear inequality system characterizing ρ-centrality.

PROPOSITION 3.5 *In* (1)*, assume that $f_i : X \to \mathbb{R}_+$, $i \in \mathcal{K}$, let $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$, and let $\rho \in [0,1]$ and $M \gg 1$. Let $\mathcal{P} \subseteq X$ be indicated by the binary variables $w_\ell = 1$ if and only if $\mathbf{x}^\ell \in \mathcal{P}$. Then, $\mathbf{x}^\ell \in X$ is ρ-central if and only if for each $i \in \mathcal{K}$, $m \in \mathcal{N}$ there exists a $j \in \mathcal{N}$ such that*

$$f_i(\mathbf{x}^\ell) \le (1-\rho) w_j f_i(\mathbf{x}^j) + \rho f_i(\mathbf{x}^m) + M(1-w_m). \tag{13}$$

PROOF. First, observe that the definition (7) of centrality can be rewritten as follows: For a vector $\mathbf{x}^\ell \in X$ to be ρ-central, a necessary and sufficient condition is that for each $i \in \mathcal{K}$, the inequality

$$f_i(\mathbf{x}^\ell) \;\le\; (1-\rho) \max_{j:\, w_j=1} f_i(\mathbf{x}^j) + \rho \min_{m:\, w_m=1} f_i(\mathbf{x}^m), \tag{14}$$

holds. Let $\hat{\jmath} \in \arg\max_{j\in\mathcal{N}}\{f_i(\mathbf{x}^j) \mid w_j = 1\}$ and $\hat{m} \in \arg\min_{m\in\mathcal{N}}\{f_i(\mathbf{x}^m) \mid w_m = 1\}$. Note that $\{j \mid w_j = 1\}$ is non-empty since the set X is closed and bounded and therefore must contain Pareto optimal vectors. Now, the equivalence between (13) and (14) follows by the next arguments, in which the terms involving $\mathbf{x}^j$ and the terms involving $\mathbf{x}^m$ are studied separately.

- $\Leftarrow$ If (14) holds for $\hat{\jmath}$ then (13) holds for at least one $j \in \mathcal{N}$. If (14) holds for $\hat{m}$ then (13) holds for all $m \in \mathcal{N}$, since for all m such that $w_m = 1$, $f_i(\mathbf{x}^m) \ge f_i(\mathbf{x}^{\hat{m}})$ and also it holds whenever $w_m = 0$.
- $\Rightarrow$ If (13) holds for some $j \in \mathcal{N}$, it must hold for $\hat{\jmath}$ since for all j, $f_i(\mathbf{x}^{\hat{\jmath}}) \ge w_j f_i(\mathbf{x}^j)$. The right-hand side of (13) is smallest for $m = \hat{m}$, since $M \gg 1$ and $w_m = 1$ for at least one index $m \in \mathcal{N}$. If the inequality holds for all $m \in \mathcal{N}$, it holds for $\hat{m}$ and thus (14) holds. ∎

A characterization similar to that in Proposition 3.5 can be formulated without the assumption of non-negativity of f_i (through addition of more large constants); however, the scaling (12) may also be a numerically wise choice.

We introduce the four sets of binary variables

$$b_{ijm\ell} = \begin{cases} 1, & \text{if } f_i(\mathbf{x}^\ell) \le (1-\rho)w_j f_i(\mathbf{x}^j) + \rho f_i(\mathbf{x}^m) + M(1-w_m), \\ 0, & \text{otherwise,} \end{cases} \qquad j,m,\ell \in \mathcal{N},\ i \in \mathcal{K},$$

$$c_{ij\ell} = \begin{cases} 1, & \text{if } \sum_{m\in\mathcal{N}} b_{ijm\ell} \ge N, \\ 0, & \text{otherwise,} \end{cases} \qquad j,\ell \in \mathcal{N},\ i \in \mathcal{K},$$

$$e_{i\ell} = \begin{cases} 1, & \text{if } \sum_{j\in\mathcal{N}} c_{ij\ell} \ge 1, \\ 0, & \text{otherwise,} \end{cases} \qquad \ell \in \mathcal{N},\ i \in \mathcal{K},$$

$$a_\ell = \begin{cases} 1, & \text{if } \mathbf{x}^\ell \in X^\rho \text{ (i.e., if } \sum_{i\in\mathcal{K}} e_{i\ell} \ge k), \\ 0, & \text{if } \mathbf{x}^\ell \notin X^\rho, \end{cases} \qquad \ell \in \mathcal{N},$$

and formulate, for each $\ell \in \mathcal{N}$, the following inequality system, which has a solution with $a_\ell = 1$ if and only if $\mathbf{x}^\ell \in X$ is ρ-central (i.e., if and only if $\mathbf{x}^\ell \in X^\rho$):

$$-2Mb_{ijm\ell} < f_i(\mathbf{x}^\ell) - \left((1-\rho)w_j f_i(\mathbf{x}^j) + \rho f_i(\mathbf{x}^m) + M(1-w_m)\right), \qquad j,m \in \mathcal{N},\ i \in \mathcal{K}, \tag{15a}$$

$$M(1-b_{ijm\ell}) \ge f_i(\mathbf{x}^\ell) - \left((1-\rho)w_j f_i(\mathbf{x}^j) + \rho f_i(\mathbf{x}^m) + M(1-w_m)\right), \qquad j,m \in \mathcal{N},\ i \in \mathcal{K}, \tag{15b}$$

$$Nc_{ij\ell} \le \textstyle\sum_{m\in\mathcal{N}} b_{ijm\ell} \le c_{ij\ell} + N - 1, \qquad j \in \mathcal{N},\ i \in \mathcal{K}, \tag{15c}$$

$$e_{i\ell} \le \textstyle\sum_{j\in\mathcal{N}} c_{ij\ell} \le Ne_{i\ell}, \qquad i \in \mathcal{K}, \tag{15d}$$

$$ka_\ell \le \textstyle\sum_{i\in\mathcal{K}} e_{i\ell} \le a_\ell + k - 1, \tag{15e}$$

$$b_{ijm\ell},\ c_{ij\ell},\ e_{i\ell},\ a_\ell \in \{0,1\}, \qquad j,m \in \mathcal{N},\ i \in \mathcal{K}. \tag{15f}$$

PROPOSITION 3.6 *For the problem* (1) *with objective functions* $f_i : X \to \mathbb{R}_+$, $i \in \mathcal{K}$, *the vectors* $\mathbf{x}^\ell \in X$, $\ell \in \mathcal{N}$, *are partitioned into a* ρ*-central* (X^ρ) *and a non-*ρ*-central* ($X \setminus X^\rho$) *part by the system* (15) *of inequalities.* ∎

Combining the systems (9) and (15) and introducing the binary variables

$$\eta_\ell = \begin{cases} 1, & \text{if } \mathbf{x}^\ell \in \mathcal{P}^\rho, \\ 0, & \text{if } \mathbf{x}^\ell \notin \mathcal{P}^\rho, \end{cases} \qquad \ell \in \mathcal{N},$$

and the constraints

$$w_\ell + a_\ell - 1 \le 2\eta_\ell \le w_\ell + a_\ell, \qquad \ell \in \mathcal{N}, \tag{16a}$$

$$\eta_\ell \in \{0,1\}, \qquad \ell \in \mathcal{N}, \tag{16b}$$

finally yields a system with feasible solutions that fulfil $\eta_\ell = 1$ for $\ell \in \mathcal{N}$ such that $\mathbf{x}^\ell \in \mathcal{P}^\rho$ ($\mathbf{x}^\ell$ is Pareto optimal and ρ-central) and $\eta_\ell = 0$ for $\ell \in \mathcal{N}$ such that $\mathbf{x}^\ell \in X \setminus \mathcal{P}^\rho$.

3.4 The explicit formulation of centrality applied to the reduced problem

We next turn back to the model with a reduced set of objectives. Our aim is then to decide which vectors $\mathbf{x} \in X$ that are ρ-central with respect to the (unknown) set of objectives $\{g_{s_1}, \ldots, g_{s_r}\}$. Let us denote this set by $X^{\beta,\rho} \subseteq X$. The Definition 2.4 of centrality implies that we can set up the centrality inequalities for the (known) set of objectives $\{\beta_1 g_1, \ldots, \beta_{2^k} g_{2^k}\}$, since for all p with $\beta_p = 0$ the corresponding inequality in (7) is fulfilled. However, it is not possible to use analogous variable definitions, since this would lead to non-linear constraints corresponding to (15a). Instead, we rewrite the inequality (13) as:

$$\exists j \in \mathcal{N}: \qquad f_i(\mathbf{x}^\ell) \le (1-\rho)f_i(\mathbf{x}^j) - M(1-w_j) + \rho f_i(\mathbf{x}^m) + M(1-w_m), \quad i \in \mathcal{K},\ m \in \mathcal{N}. \tag{17}$$

Replacing f_i by $\beta_p g_p$, then there exists a $j \in \mathcal{N}$ such that

$$\beta_p g_p(\mathbf{x}^\ell) \le (1-\rho)\beta_p g_p(\mathbf{x}^j) - M(1-w_j) + \rho\beta_p g_p(\mathbf{x}^m) + M(1-w_m), \quad p \in \hat{\mathcal{K}},\ m \in \mathcal{N}. \tag{18}$$

Analogously to Section 3.3, we use the following variable declarations for $p \in \hat{\mathcal{K}}$, and $j, m, \ell \in \mathcal{N}$ (the first definition is modified in order to avoid the non-linearity):

$$b_{pjm\ell} = \begin{cases} 1, & \text{if } \beta_p g_p(\mathbf{x}^\ell) \leq (1-\rho)\beta_p g_p(\mathbf{x}^j) - M(1-w_j) + \rho\beta_p g_p(\mathbf{x}^m) + M(1-w_m), \\ 0, & \text{otherwise,} \end{cases}$$

$$c_{pj\ell} = \begin{cases} 1, & \text{if } \sum_{m\in\mathcal{N}} b_{pjm\ell} \geq N, \\ 0, & \text{otherwise,} \end{cases}$$

$$e_{p\ell} = \begin{cases} 1, & \text{if } \sum_{j\in\mathcal{N}} c_{pj\ell} \geq 1, \\ 0, & \text{otherwise,} \end{cases}$$

$$a_\ell = \begin{cases} 1, & \text{if } \mathbf{x}^\ell \in X^{\beta,\rho} \text{ (i.e., if } \sum_{p\in\hat{\mathcal{K}}} e_{p\ell} \geq 2^k), \\ 0, & \text{if } \mathbf{x}^\ell \notin X^{\beta,\rho}. \end{cases}$$

For the set of objectives $\{\beta_1 g_1, \ldots, \beta_{2^k} g_{2^k}\}$ the ρ-centrality system corresponding to (15) is then expressed as:

$$-2Mb_{pjm\ell} < \beta_p g_p(\mathbf{x}^\ell) - \big((1-\rho)\beta_p g_p(\mathbf{x}^j) - M(1-w_j) + \rho\beta_p g_p(\mathbf{x}^m) + M(1-w_m)\big), \quad j, m, \ell \in \mathcal{N},\ p \in \hat{\mathcal{K}}, \tag{19a}$$

$$M(1-b_{pjm\ell}) \geq \beta_p g_p(\mathbf{x}^\ell) - \big((1-\rho)\beta_p g_p(\mathbf{x}^j) - M(1-w_j) + \rho\beta_p g_p(\mathbf{x}^m) + M(1-w_m)\big), \quad j, m, \ell \in \mathcal{N},\ p \in \hat{\mathcal{K}}, \tag{19b}$$

$$Nc_{pj\ell} \leq \sum_{m\in\mathcal{N}} b_{pjm\ell} \leq c_{pj\ell} + N - 1, \quad j, \ell \in \mathcal{N},\ p \in \hat{\mathcal{K}}, \tag{19c}$$

$$e_{p\ell} \leq \sum_{j\in\mathcal{N}} c_{pj\ell} \leq Ne_{p\ell}, \quad \ell \in \mathcal{N},\ p \in \hat{\mathcal{K}}, \tag{19d}$$

$$2^k a_\ell \leq \sum_{p\in\hat{\mathcal{K}}} e_{p\ell} \leq a_\ell + 2^k - 1, \quad \ell \in \mathcal{N}, \tag{19e}$$

$$b_{pjm\ell},\ c_{pj\ell},\ e_{p\ell},\ a_\ell \in \{0,1\}, \quad j, m, \ell \in \mathcal{N},\ p \in \hat{\mathcal{K}}. \tag{19f}$$

As in the previous section, we combine the systems (11) and (19) and introduce the binary variables

$$\eta_\ell = \begin{cases} 1, & \text{if } \mathbf{x}^\ell \in \mathcal{P}_\tau^{\beta,\rho}, \\ 0, & \text{if } \mathbf{x}^\ell \notin \mathcal{P}_\tau^{\beta,\rho}, \end{cases} \qquad \ell \in \mathcal{N},$$

and the constraints

$$w_\ell + a_\ell - 1 \leq 2\eta_\ell \leq w_\ell + a_\ell, \quad \ell \in \mathcal{N}, \tag{20a}$$

$$\eta_\ell \in \{0,1\}, \quad \ell \in \mathcal{N}, \tag{20b}$$

to form a consistent system such that $\eta_\ell = 1$ if and only if $\mathbf{x}^\ell \in X$ is ρ-central and τ-Pareto optimal in the reduced problem (i.e., if $\mathbf{x}^\ell \in \mathcal{P}_\tau^{\beta,\rho}$), $\ell \in \mathcal{N}$.

3.5 Formulating the ideal model

Using the explicit formulation of Pareto optimality together with ρ-centrality, the ideal formulation of the problem (8) can now be stated. Assume that the number of ρ-central Pareto optimal points in the original problem (1) is Q and define $\mathcal{Q} = \{1, \ldots, Q\}$. Then, without any loss of generality we assume that $\mathcal{P}^\rho = \{\mathbf{x}^1, \ldots, \mathbf{x}^Q\}$. The pairwise distances in the objective space between all pairs of points $\mathbf{x}^q \in \mathcal{P}^\rho$ and $\mathbf{x}^\ell \in X$ are denoted by $d_{ql} = ||\mathbf{f}(\mathbf{x}^q) - \mathbf{f}(\mathbf{x}^\ell)||$, $q \in \mathcal{Q}$, $\ell \in \mathcal{N}$. Introducing an auxiliary variable $\theta \in \mathbb{R}_+$ the objective in (8) can then be formulated as

$$\begin{aligned} &\text{minimize} && \theta, \\ &\text{subject to} && \theta \geq \min_{\substack{\ell\in\mathcal{N}: \\ \eta_\ell = 1}} d_{q\ell}, \quad q \in \mathcal{Q}, \\ & && \theta \geq \min_{q\in\mathcal{Q}} d_{q\ell}, \quad \ell \in \mathcal{N} : \eta_\ell = 1, \end{aligned} \tag{21}$$

where the variables η_ℓ implicitly depend on the decision variables $\boldsymbol{\beta}$ and τ. The problem (21) can be expressed as

$$\begin{aligned} &\text{minimize} && \theta, \\ &\text{subject to} && \theta \geq \min_{\ell \in \mathcal{N}} \big\{(1-\eta_\ell)M + d_{q\ell}\big\}, \quad q \in \mathcal{Q}, \\ &&& \theta \geq \min_{q \in \mathcal{Q}} \big\{d_{q\ell} - (1-\eta_\ell)M\big\}, \quad \ell \in \mathcal{N}, \end{aligned} \tag{22}$$

where $M \geq \max\big\{d_{q\ell} \mid q \in \mathcal{Q},\ \ell \in \mathcal{N}\big\}$. The min-operators are then replaced using the binary variables

$$y_{q\ell} = \begin{cases} 1, & \text{if } \theta \geq (1-\eta_\ell)M + d_{q\ell}, \\ 0, & \text{otherwise,} \end{cases} \qquad q \in \mathcal{Q},\ \ell \in \mathcal{N},$$

$$z_{q\ell} = \begin{cases} 1, & \text{if } \theta \geq d_{q\ell} - (1-\eta_\ell)M, \\ 0, & \text{otherwise,} \end{cases} \qquad q \in \mathcal{Q},\ \ell \in \mathcal{N},$$

and, finally, the ideal model (8) is formulated as the binary linear program

$$\begin{aligned} &\text{minimize} && \theta, \\ &\text{subject to} && (1-\eta_\ell)M + d_{ql} - \theta \leq 2M(1-y_{q\ell}), && q \in \mathcal{Q},\ \ell \in \mathcal{N}, && \text{(23a)} \\ &&& -(1-\eta_\ell)M + d_{ql} - \theta \leq M(1-z_{q\ell}), && q \in \mathcal{Q},\ \ell \in \mathcal{N}, && \text{(23b)} \\ &&& \textstyle\sum_{\ell \in \mathcal{N}} y_{q\ell} \geq 1, && q \in \mathcal{Q}, && \text{(23c)} \\ &&& \textstyle\sum_{q \in \mathcal{Q}} z_{q\ell} \geq 1, && \ell \in \mathcal{N}, && \text{(23d)} \\ &&& y_{q\ell},\ z_{q\ell} \in \{0,1\}, && q \in \mathcal{Q},\ \ell \in \mathcal{N}, && \text{(23e)} \\ &&& \tau \geq 0, && && \text{(23f)} \\ &&& \boldsymbol{\beta} \text{ satisfies (6)}, && && \text{(23g)} \\ &&& (u, v, w, \boldsymbol{\beta}, \tau) \text{ satisfies (11)}, && && \text{(23h)} \\ &&& (b, c, e, a) \text{ satisfies (19)}, && && \text{(23i)} \\ &&& (w, a, \eta) \text{ satisfies (20)}. && && \text{(23j)} \end{aligned}$$

The program (23) has in the order of $N^3 2^k$ binary variables and constraints, where the magnitude is settled by the explicit formulation of centrality. This program cannot be solved in reasonable time for practical values of N and k.

What we have done until now is to reformulate our goal into an exact model based on an explicit characterization of Pareto optimality. Now when we know how this looks like, and from where the complexity arises, we can study efficient approximations of the exact model. Therefore, in the next section we formulate an approximate model that separates the optimization over the decision variables $\boldsymbol{\beta}$ and τ, resulting in a sequence of two fairly easily solved problems.

4 The approximate model

The underlying characteristic of the ideal model (23) that yields the high complexity is that the optimization is made simultaneously over $\boldsymbol{\beta}$ and τ. To accomplish this, τ-Pareto optimality as well as ρ-centrality must enter the model via constraints, as presented in Section 3. This leads to an explosion of binary variables and constraints. In an approximate model to be presented we separate the optimization over the variables $\boldsymbol{\beta}$ and τ. The only constraints from the ideal model that are kept intact are (6), which describe the collection of objectives. The rest of the constraints—including the large sets of constraints (11) and (19)—will no longer be necessary (and neither will the variables introduced in these systems). Instead a certain objective function is used in a first problem (cf. (26) on page 14) which is aimed at evaluating and deciding on a good collection of objectives described by $\boldsymbol{\beta}$ without making use of the tolerance parameter τ. Then simple polynomial algorithms are used to find the (0-)Pareto optimal ρ-central part of X for the reduced problem and after that a second problem (cf. (27) on page 14) is solved to find the optimal value of τ given the already chosen value of $\boldsymbol{\beta}$.

Obviously, a solution found using the approximate model might not be optimal in (23), since decisions are made before all information is known, i.e., $\boldsymbol{\beta}$ is selected before it is known exactly how τ will affect

the Pareto optimal set. Also, there is no measure of optimality provided for the approximate solution. However, the mechanism for selecting $\boldsymbol{\beta}$ is sensible, and the complexity of the problem formulation decreases enormously. This motivates the use of the approximate formulation.

4.1 Correlation between objectives

The *correlation coefficient* of two objective functions f_i and f_j over a set of decision vectors $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$ is defined as

$$\hat{\rho}(f_i, f_j) = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \in [-1, 1], \quad i, j \in \mathcal{K}, \tag{24}$$

where

$$s_{ij} = \frac{1}{N}\sum_{\ell \in \mathcal{N}} \left(f_i(\mathbf{x}^\ell) - \tfrac{1}{N}\sum_{m \in \mathcal{N}} f_i(\mathbf{x}^m) \right) \left(f_j(\mathbf{x}^\ell) - \tfrac{1}{N}\sum_{m \in \mathcal{N}} f_j(\mathbf{x}^m) \right), \quad i, j \in \mathcal{K}.$$

The value of the pairwise correlation coefficients between two objective functions gives a measure of how similar the functions evaluate the set $X = \{\mathbf{x}^1, \ldots, \mathbf{x}^N\}$. If the correlation is perfect, i.e., $\hat{\rho}(f_i, f_j) = 1$, then the Pareto optimal set $\mathcal{P}$ would be unchanged if the pair of objectives was replaced by its mean; one of the objectives in the pair would actually be redundant[2]. Using this reasoning we identify an *almost redundant* objective function as a function having a *high correlation* with some other objective function. With this as a starting point we will in Section 4.2 derive the first problem to solve in the approximate model.

The objective when deciding on the optimal value of $\boldsymbol{\beta}$ is to maximize the least correlation coefficient between each original objective function $f_i \in \{f_1, \ldots, f_k\}$ and its most similar (with respect to correlation) objective function in the reduced set of objectives $\{g_{s_1}, \ldots, g_{s_r}\}$. A rough illustration of this objective is given in Figure 3 by representing $\{f_1, \ldots, f_k\}$ by vectors and with small/large angles between positively/negatively correlated f_i's. The objective is then to find at most r sets K_{s_j}, $j = 1, \ldots, r$, in the power set of $\mathcal{K}$, each corresponding to a collection of objectives f_i, $i \in K_{s_j}$, such that the maximum distance between a vector in the collection and the mean of these vectors is at minimum.

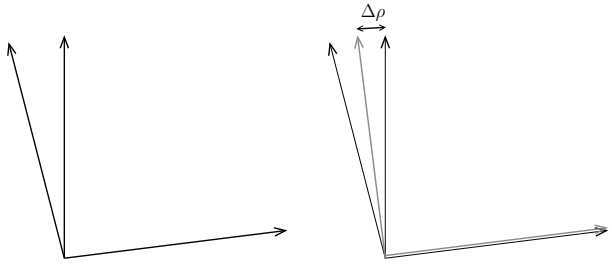


Figure 3: An abstract illustration of original objective functions (black), optimally aggregated (with $r = 2$) objective functions (gray), and of the "correlation error", $\Delta\rho$, for this aggregation.

4.2 Finding the optimal collections

In the ideal model, $2^{\mathcal{K}}$ denotes the power set of $\mathcal{K}$ in which each element represents a set of original objectives that may be dropped from the system and replaced by their weighted mean. To reduce the computational complexity of the model, we introduce a threshold, α, and only allow collections of original objectives such that the correlation between any two objectives in the collection is at least α. For $\alpha \in [-1, 1]$ we define the restricted set of collection candidates as

$$\mathcal{K}^\alpha := \left\{ \ell \in \hat{\mathcal{K}} \,|\, \hat{\rho}(f_i, f_j) \geq \alpha, \; \forall i, j \in K_\ell \right\}, \tag{25}$$

where $K_\ell \in 2^{\mathcal{K}}$, $\ell \in \hat{\mathcal{K}}$. As previously we want to reduce the set of original objectives $\{f_1, \ldots, f_k\}$ to the smaller set $\{g_{s_1}, \ldots, g_{s_r}\}$, where $g_{s_j} = \frac{1}{|K_{s_j}|}\sum_{i \in K_{s_j}} f_i$, $s_j \in \mathcal{K}^\alpha$. Analogously to Section 2 the decision

[2]Note that $\hat{\rho}(f_i, f_j) = 1$ is not a necessary condition for $\mathcal{P}$ to remain intact; e.g., if $f_i(x) = x$ is redundant in a multi-objective optimization problem, then so is $f_j(x) = x^2$, but in general $\hat{\rho}(f_i, f_k) \neq \hat{\rho}(f_j, f_k)$ for $k \notin \{i, j\}$.

vector $\boldsymbol{\beta} \in \{0,1\}^{|\mathcal{K}^\alpha|}$ indicates which functions g_{s_j} to choose and the incidence matrix $A \in \mathbb{B}^{k \times |\mathcal{K}^\alpha|}$ defines which original objectives f_i belong to each collection g_{s_j} of functions.

Let $\psi_p := \min_{i \in K_p} \hat{\rho}(g_p, f_i)$, and introduce an auxiliary variable $z \in \mathbb{R}$ and a scalar $M \geq 2$. The problem of choosing the best set of at most r collections is formulated as the binary linear program

$$\begin{aligned} \underset{z,\beta}{\text{maximize}} \quad & z, \\ \text{such that} \quad & z \leq \psi_p + (1 - \beta_p)M, \quad p \in \mathcal{K}^\alpha, \\ & A\boldsymbol{\beta} \geq \mathbf{1}^k, \\ & \boldsymbol{\beta}^{\mathrm{T}} \mathbf{1}^{|\mathcal{K}^\alpha|} \leq r, \\ & \boldsymbol{\beta} \in \{0,1\}^{|\mathcal{K}^\alpha|}, \end{aligned} \tag{26}$$

where $r < k$. For $\alpha = -1$ this problem, which is a set covering problem [19] with additional complicating constraints, has 2^k binary variables. However, the number of variables can be substantially reduced by employing a larger value of α, whence the complexity of (26) will not be a serious issue.

An observation to be made is given in Proposition 4.1, whose result is quite obvious from Figure 3.

PROPOSITION 4.1 *Let the set of optimal solutions to* (26) *be denoted by* $\mathcal{B}$. *Then for some* $(z^*, \boldsymbol{\beta}^*) \in \mathcal{B}$, *exactly* r *objectives are used, i.e,* $\sum_{p \in \mathcal{K}^\alpha} \beta_p^* = r$.

PROOF. Let $s = 0$ and suppose that $(z^s, \boldsymbol{\beta}^s)$ is optimal in (26) with $\sum_{p \in \mathcal{K}^\alpha} \beta_p^s < r$. We have that $r < k \leq |\mathcal{K}^\alpha|$ since, for all $\alpha \in [0,1]$, it holds that $i \in \mathcal{K}^\alpha$, for each $i \in \mathcal{K}$ corresponding to an original objective f_i. Without loss of generality, let the k first elements in the set $\mathcal{K}^\alpha$ correspond to the sets $K_\ell = \{\ell\}$, $\ell \in \mathcal{K}$. Then, $\psi_\ell = 1$, $\ell = 1, \ldots, k$, and there exists an element $q \in \mathcal{K}$ such that the objective f_q is not selected as an objective for the reduced problem, which implies that $\beta_q^s = 0$. Letting $\beta_p^{s+1} = \beta_p^s$ for $p \in \mathcal{K}^\alpha \setminus \{q\}$ and $\beta_q^{s+1} = 1$, it follows that $(z^{s+1}, \boldsymbol{\beta}^{s+1}) \in \mathcal{B}$, where $z^{s+1} = z^s$, and that $\sum_{p \in \mathcal{K}^\alpha} \beta_p^{s+1} = \sum_{p \in \mathcal{K}^\alpha} \beta_p^s + 1$. Let $s = s + 1$ and repeat until $\sum_{p \in \mathcal{K}^\alpha} \beta_p^s = r$. The result follows. ∎

4.3 Finding the optimal dominance tolerance

Solving (26) yields an optimal solution $(z^*, \boldsymbol{\beta}^*)$, which defines the objectives to use in the reduced problem. Observe that "optimal" here means with respect to the approximate model (26) and that $(z^*, \boldsymbol{\beta}^*)$ may be non-optimal with respect to (23). By pairwise comparisons between all the points in X, the Pareto optimal set with respect to the selected objectives, $\mathcal{P}^{\beta^*}$, is then extracted. Further, by checking the ρ-centrality characterization from Definition 2.4 for all vectors $\mathbf{x} \in \mathcal{P}^{\beta^*}$ the ρ-central part $\mathcal{P}^{\beta^*,\rho} \subseteq \mathcal{P}^{\beta^*}$ is extracted. The idea is then—starting from $\mathcal{P}_0^{\beta^*,\rho}$—to increase the value of the tolerance parameter τ from zero and find the value τ^* for which the Hausdorff distance between $\mathbf{f}(\mathcal{P}^\rho)$ and $\mathbf{f}(\mathcal{P}_\tau^{\beta^*,\rho})$ is minimized. The second problem is thus to

$$\begin{aligned} \underset{\tau}{\text{minimize}} \quad & d_H(\mathbf{f}(\mathcal{P}^\rho), \mathbf{f}(\mathcal{P}_\tau^{\beta^*,\rho})), \\ \text{such that} \quad & \tau \geq 0. \end{aligned} \tag{27}$$

The solution process of the approximate model is illustrated in Figure 4.

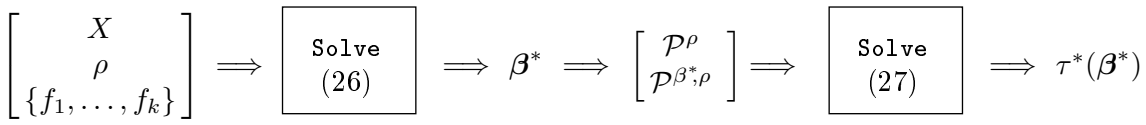


Figure 4: The solution procedure for the approximate model including inputs and outputs of the problems (26) and (27).

Changing the value of τ changes the Pareto set that ρ-centrality relates to in Definition 2.4. This implies a difficulty since the value of τ will not be known until after (27) has been solved. The ρ-centrality partitioning could be done in each iteration of the solution procedure for (27), defined in Algorithm 4.1, but to receive an easily solved optimization problem over τ (presented below) we make the following

approximation: Fix the set to relate to for ρ-centrality by approximating $\mathcal{P}_\tau^{\beta^*}$ by $\mathcal{P}$ in the ρ-centrality partitioning, i.e., use (cf. Definition 2.4)

$$\mathcal{P}_\tau^{\beta^*,\rho} :\approx \left\{ \mathbf{x} \in \mathcal{P}_\tau^{\beta^*} \;\middle|\; g_p(\mathbf{x}) \leq (1-\rho) \max_{\mathbf{y}\in\mathcal{P}} g_p(\mathbf{y}) + \rho \min_{\mathbf{y}\in\mathcal{P}} g_p(\mathbf{y}),\ \forall p \in \mathcal{K}^\alpha \right\}. \tag{28}$$

From Section 2 it is clear that $\mathcal{P}^{\beta^*} \subseteq \mathcal{P}$. It is also clear from Definition 2.4 that $\mathcal{P}^\rho \subseteq \mathcal{P}$. We also conclude that $\mathcal{P}^{\beta^*,\rho} \subseteq \mathcal{P}^\rho$, since $\mathcal{P}^{\beta^*,\rho} = \mathcal{P}^{\beta^*} \cap X^\rho \subseteq \mathcal{P} \cap X^\rho = \mathcal{P}^\rho$, where the first equality relies on the approximation (28), since the set X^ρ depends on the objective functions. These relations are illustrated in Figure 5 for the objective space.

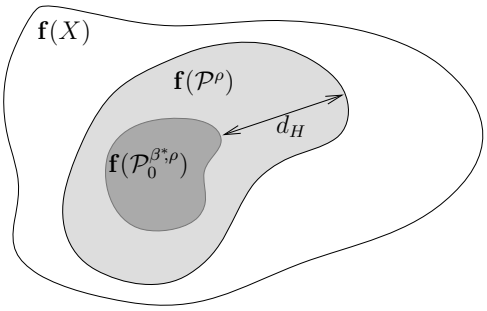


Figure 5: Originating from a subset of $\mathcal{P}^\rho$, the problem (27) enlarges the inner set by increasing the value of τ.

We define the function $\phi_1(\tau)$ as the maximum distance from a point in $\mathcal{P}^\rho$ to its nearest point in $\mathcal{P}_\tau^{\beta^*,\rho}$ and $\phi_2(\tau)$ as the maximum distance from a point in $\mathcal{P}_\tau^{\beta^*,\rho}$ to its nearest point in $\mathcal{P}^\rho$, i.e.,

$$\phi_1(\tau) = \max_{\mathbf{x}\in\mathcal{P}^\rho} \Big\{ \min_{\mathbf{y}\in\mathcal{P}_\tau^{\beta^*,\rho}} d\big(\mathbf{f}(\mathbf{x}), \mathbf{f}(\mathbf{y})\big) \Big\}, \tag{29}$$

$$\phi_2(\tau) = \max_{\mathbf{y}\in\mathcal{P}_\tau^{\beta^*,\rho}} \Big\{ \min_{\mathbf{x}\in\mathcal{P}^\rho} d\big(\mathbf{f}(\mathbf{y}), \mathbf{f}(\mathbf{x})\big) \Big\}. \tag{30}$$

Then, the problem (8) can, for a fixed value $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, be rewritten as that to minimize $\delta(\boldsymbol{\beta}^*, \tau)$ over $\tau \geq 0$, where

$$\delta(\boldsymbol{\beta}^*, \tau) = d_H(\mathbf{f}(\mathcal{P}^\rho), \mathbf{f}(\mathcal{P}_\tau^{\beta^*,\rho})) = \max\left\{\phi_1(\tau), \phi_2(\tau)\right\}. \tag{31}$$

Since increasing the value of τ will increase the size of $\mathcal{P}_\tau^{\beta^*,\rho}$, $\phi_1(\tau)$ is a monotonically decreasing lower semi-continuous function. Analogously, $\phi_2(\tau)$ describes a monotonically increasing upper semi-continuous function. The sets $\mathcal{P}^\rho$ and $\mathcal{P}_\tau^{\beta^*,\rho}$ are discrete whence the functions $\phi_1(\tau)$ and $\phi_2(\tau)$ are piecewise constant. Thus, $\delta(\boldsymbol{\beta}^*, \tau)$ is a piecewise constant quasi-convex [2] function of τ. Figure 6 illustrates the functions $\phi_1(\tau)$, $\phi_2(\tau)$ (thus also the function $\delta(\boldsymbol{\beta}^*, \tau) = \max\{\phi_1(\tau), \phi_2(\tau)\}$), and the monotonically increasing upper semi-continuous function

$$\phi(\tau) := \phi_2(\tau) - \phi_1(\tau). \tag{32}$$

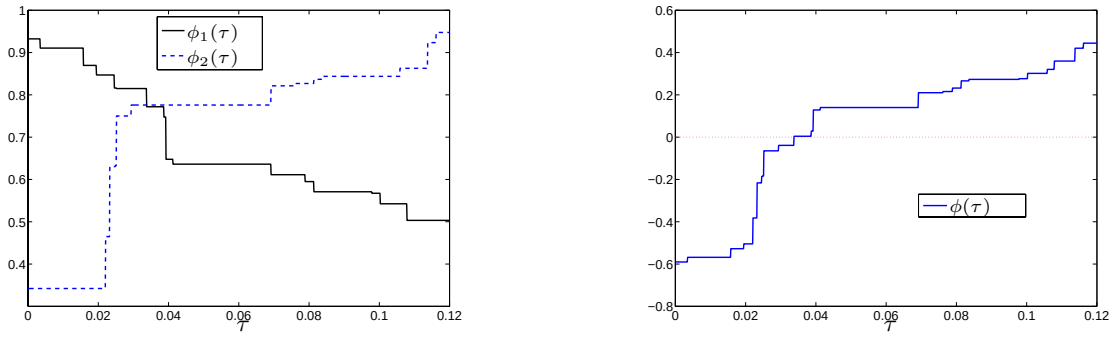
We are interested in finding

$$\tau^* \in T_\delta := \arg\min_{\tau\geq 0}\ \delta(\boldsymbol{\beta}^*, \tau). \tag{33}$$

If $\phi_1(0) \leq \phi_2(0)$ then $\tau^* = 0$. For the case when $\phi_1(0) > \phi_2(0)$ we define, for $\varepsilon > 0$,

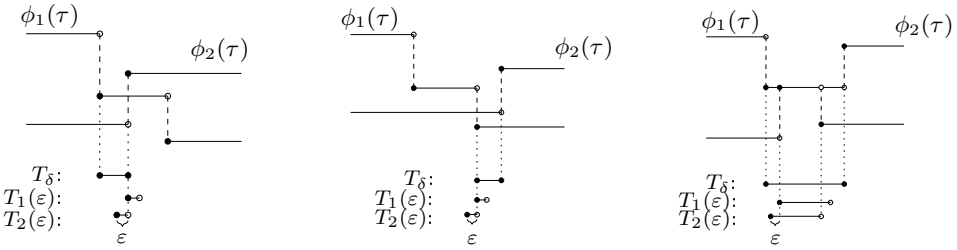
$$T_1(\varepsilon) = \{\tau \geq 0 \mid \phi(\tau+\varepsilon) \geq 0, \phi(\tau) \leq 0\} \quad \text{and} \quad T_2(\varepsilon) = \{\tau \geq 0 \mid \phi(\tau) \geq 0, \phi(\tau-\varepsilon) \leq 0\}. \tag{34}$$

Then, either $T_1(\varepsilon) \subseteq T_\delta$, $T_2(\varepsilon) \subseteq T_\delta$, or both hold, as illustrated in Figure 7.



(a) $\phi_1(\tau)$ and $\phi_2(\tau)$ (b) $\phi(\tau)$

Figure 6: Typical appearances of the functions: (a) $\phi_1(\tau)$ and $\phi_2(\tau)$ defined in (29) and (30), respectively (and $\delta(\boldsymbol{\beta}^*, \tau)$ defined in (31)); (b) $\phi(\tau)$ defined in (32). For visibility reasons, the graphs of all the functions have been closed.



(a) $T_2(\varepsilon) \subseteq T_\delta$ (b) $T_1(\varepsilon) \subseteq T_\delta$ (c) $T_1(\varepsilon) \subseteq T_\delta \supseteq T_2(\varepsilon)$

Figure 7: Illustration of the sets T_δ, $T_1(\varepsilon)$, and $T_2(\varepsilon)$ for three different cases.

For a sufficiently small value of $\varepsilon > 0$ it then suffices to find $\tau_1 \in T_1(\varepsilon)$ and $\tau_2 \in T_2(\varepsilon)$, and then choose $\tau^* \in \{\tau_1, \tau_2\}$ such that $\tau^* \in T_\delta$. This is done using a simple bisection technique on $\phi(\tau)$, for which an algorithm in pseudo-code is given in Algorithm 4.1. In principle, since X is a discrete set of points, there exists a value of $\varepsilon > 0$ that guarantees an optimal solution to (27) using Algorithm 4.1. This value is, however, a priori unknown.

The main output from the algorithm is τ^*, the (approximate) optimal value of the tolerance parameter τ (given the already chosen value of $\boldsymbol{\beta}^*$ from Section 4.2). Other outputs are the resulting Hausdorff distance $\delta\left(\boldsymbol{\beta}^*, \tau^*(\boldsymbol{\beta}^*)\right)$ and $error^{\text{ub}}$, which denotes the maximum error in $\delta(\boldsymbol{\beta}^*, \tau)$ due to a possibly too large value of $\varepsilon > 0$.

Instead of solving over τ for just one value of $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, it is possible to create a pool of candidate $\boldsymbol{\beta}$'s (e.g. by solving the program (26) repeatedly, and in each iteration add a constraint that cuts away the previous solution) and compute $\tau^*(\boldsymbol{\beta})$ for each value of $\boldsymbol{\beta}$ in the pool. By comparing $\delta(\boldsymbol{\beta}, \tau^*(\boldsymbol{\beta}))$ for all $\boldsymbol{\beta}$'s in the pool, the pair $(\boldsymbol{\beta}, \tau)$ that yields the lowest value of $\delta(\boldsymbol{\beta}, \tau)$ can be selected as an approximate solution to the main reduction problem (8). By increasing the pool, the objective function value decreases monotonically and—for a sufficiently large pool—an optimal solution to (8) will eventually be found, provided that $\varepsilon > 0$ is large enough.

5 An illustrating example

The results derived in the previous section are here illustrated for an application of the approximate model to a small example before—in Section 6—it is applied to a larger industrial problem. The example indicates that "similar" objectives may be aggregated with only a minor loss of precision of $\mathcal{P}$, whereas the loss will be large if "non-similar" objectives are forced to be aggregated. However, it also indicates that the utilization of the tolerance parameter τ can substantially repair the damages from unsuitable aggregations. The model has been implemented in `MATLAB` [16] in combination with `AMPL` [10] and the

Algorithm 4.1 Pseudo-code for the τ-optimization algorithm.

```
input phi1, phi2, phi, epsilon
pick tau_1 = 0
pick tau_2 big enough s.t. phi(tau_2) > 0
until tau_2 - tau_1 < epsilon {
    let tau_3 = (tau_1 + tau_2)/2
    if phi(tau_3) < 0 {
        let tau_1 = tau_3 }
    elseif phi(tau_3) > 0 {
        let tau_2 = tau_3 }
    else {
        output tau_star = tau_3
        delta_star = phi1(tau_star)
        error^ub = 0 }
    }
output tau_star in argmin_{tau_1,tau_2} max(phi1(tau),phi2(tau))
output delta_star = max(phi1(tau_star),phi2(tau_star))
output error^ub = phi(tau_2) - phi(tau_1)
```

CPlex solver [13] for solving problem (I).

The example instance of (1) is defined by

$$f_1(\mathbf{x}) := (x_1 - 6)^2 + 4\,(x_2 - 4)^2, \qquad f_4(\mathbf{x}) := \tfrac{25}{4}\,(x_1 - 4)^2 + \tfrac{9}{4}\,(x_2 + 5)^2,$$
$$f_2(\mathbf{x}) := 4\,(x_1 - 5)^2 + \tfrac{9}{4}\,(x_2 - 5)^2, \qquad f_5(\mathbf{x}) := (x_1 + 3)^2 + (x_2 + 3)^2,$$
$$f_3(\mathbf{x}) := 4\,(x_1 - \tfrac{11}{2})^2 + 4\,(x_2 + 3)^2, \qquad X := \{x_1 \times x_2 \mid x_i \in \{-10, -9.75, -9.5, \ldots, 10\},\ i = 1, 2\}.$$

Figure 8 shows level curves of the objective functions and (the convex hull of) the Pareto optimal subset $\mathcal{P} \subseteq X$. The pairwise correlations, defined in (24), of the objective functions $f_1, \ldots, f_5$ are shown in Table 1.

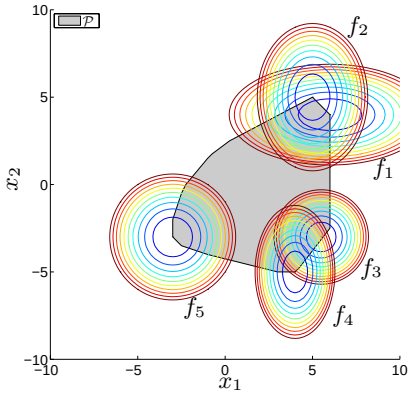


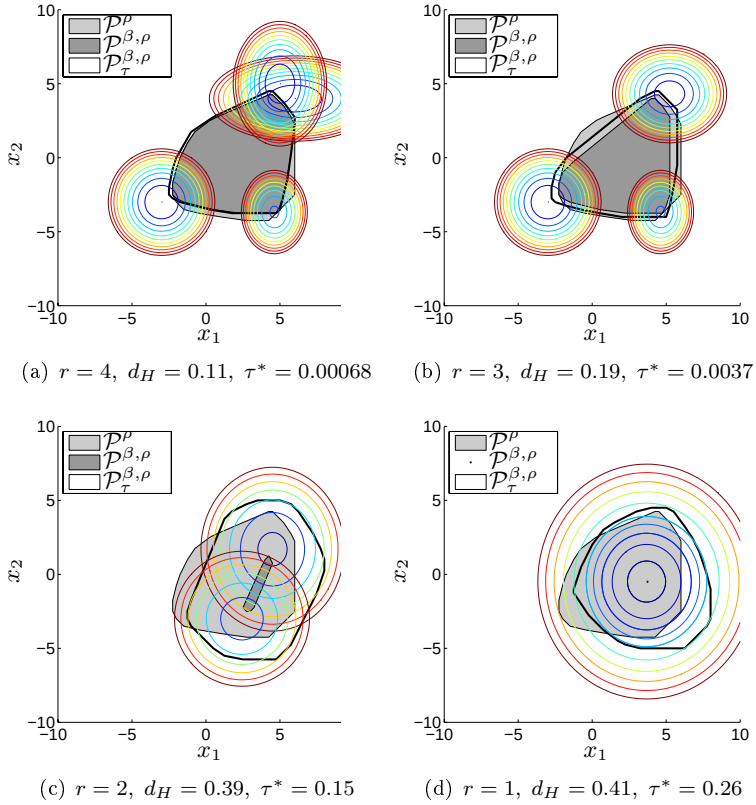
Figure 8: Level curves of the five objective functions and the Pareto optimal subset $\mathcal{P} \subseteq X$.

	f_1	f_2	f_3	f_4	f_5
f_1	1	0.74	0.13	0.11	-0.28
f_2		1	0.63	0.69	-0.35
f_3			1	0.97	0.15
f_4				1	0.09
f_5					1

Table 1: Correlation coefficients for the objective functions in the example.

Solving the approximate model for the example problem and with the number of objectives in the reduced problem restricted to 4, 3, 2, and 1, respectively, leads to the results shown in Figures 9(a)–(d). These figures show the level curves of the respective aggregated objective functions, and the resulting (convex hulls of the) Pareto optimal sets for the respective reduced problems, with and without a tolerance $\tau > 0$, as compared to the original $\mathcal{P}^\rho$. We employed the value $\rho = 0.15$. Note that the Hausdorff distance is measured in the 5-dimensional objective space.

Figures 9(a)–(d) clearly indicate that as long as only objectives with large pairwise correlations are



(a) $r = 4,\ d_H = 0.11,\ \tau^* = 0.00068$ (b) $r = 3,\ d_H = 0.19,\ \tau^* = 0.0037$

(c) $r = 2,\ d_H = 0.39,\ \tau^* = 0.15$ (d) $r = 1,\ d_H = 0.41,\ \tau^* = 0.26$

Figure 9: Illustration of results for different numbers of objective functions in the reduced problem.

aggregated, the loss of precision of the Pareto optimal set will be small. Also, by allowing a tolerance $\tau > 0$, the similarities between the original and reduced Pareto optimal set can be retained.

6 An industrial application

The approximate model developed in Section 4 has been applied to an industrial problem regarding the configuration of heavy-duty trucks. Models and data have been supplied by Volvo 3P, which is the business unit responsible for among other things product development for the trucks brands in the Volvo Group. Typically, trucks are very customer adapted depending on differences in the environments in which the trucks are to be used and for what transport missions. For this reason, the trucks are highly modularized, which results in an enormous number of possible configurations. The background to the problem approached is that there is no wish to produce a truck with a configuration that is worse than some other configuration in all possible quality measures. However, different customers may appreciate the quality measures differently. Thus, the target for the truck company should be to offer the customers a number of configurations that are in some sense well distributed over the Pareto optimal set of trucks. The first step is to identify this Pareto optimal set.

Here, a simple linear multi-body system model of a truck has been programmed in `MATLAB` [16]. The model is illustrated in Figure 10. The complete decision (or, configuration) space X' consists of all possible combinations of a number of cabs, front axle installations, rear axle installations, and superstructures. In this example $|X'| = 1296$.

From the outputs of the model, 12 quality measures (i.e., objective functions in the multi-objective optimization problem) have been defined, concerning, e.g., durability, driver environment, load comfort, and exposition of road wear.

Solving the original multi-objective optimization problem (1), where $\{f_1, \ldots, f_{12}\}$ represent the 12 quality measures and X consists of the 1296 feasible configurations, showed that the proportion of originally Pareto optimal solutions is $\frac{|\mathcal{P}|}{|X'|} \approx 0.19$. The pairwise correlations between the objectives are shown in Table 2, where it is clear that some of the objectives are positively correlated, some are negatively

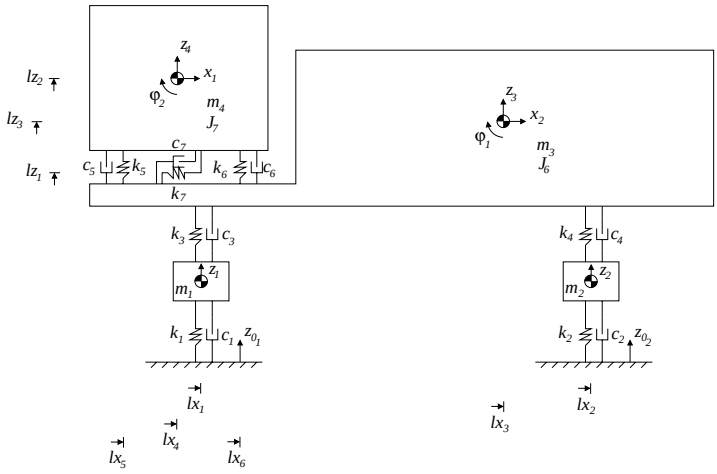


Figure 10: The truck model used. A configuration is defined by the positions lx and lz, the spring constants k, the damping constants c, the masses m and the moments of inertia J. Inputs to the model are road excitations z_0 and outputs are displacements, velocities, accelerations and forces in the degrees of freedom x, z, and φ.

correlated, and some are rather independent of each other.

	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}
f_1	1	-0.24	0.29	-0.06	0.19	0.78	-0.15	0.25	-0.31	0.39	-0.08	-0.21
f_2		1	0.63	-0.40	0.19	-0.17	0.22	0.37	0.13	-0.45	-0.22	0.09
f_3			1	0.06	0.77	0.39	0.11	0.61	-0.18	-0.19	-0.29	-0.10
f_4				1	0.68	-0.03	0.13	0.05	0.15	0.14	0.18	0.11
f_5					1	0.28	0.18	0.49	-0.03	-0.04	-0.09	0.01
f_6						1	-0.24	0.22	-0.45	0.39	-0.07	-0.28
f_7							1	0.44	0.80	0.08	0.50	0.81
f_8								1	0.06	-0.39	-0.32	0.19
f_9									1	0.22	0.73	0.81
f_{10}										1	0.76	0.21
f_{11}											1	0.58
f_{12}												1

Table 2: Correlation coefficients for the objective functions in the industrial application.

For the definition of τ-Pareto optimality to make sense it is important that the objective functions are of approximately the same magnitude. As introduced in Section 3.3, the objective functions are scaled and each $f_i(\mathbf{x})$ replaced by the expression in (12), so that $f_i(\mathbf{x}) \in [0,1]$, $i \in \mathcal{K}$, $\mathbf{x} \in \mathcal{P}$.

In the numerical experiments—when nothing else is specified—the parameter set employed is $r = 6$, $\rho = 0.15$, $\alpha = 0$, a Euclidean distance measure, and $\varepsilon = 10^{-4}$.

To simulate a real application of the method developed, X' is randomly partitioned into two sets, X and X^{ref}, such that $|X| \approx |X^{\text{ref}}|$. The set X is used to find the values of $\boldsymbol{\beta}$ and τ that defines the reduced problem. X^{ref} is a reference set, to which the reduced problem is applied and for which the quality of the resulting ρ-central and τ-Pareto optimal set $\mathcal{P}_\tau^{\beta,\rho}$ is measured. Through this procedure the risk of over-fitting the data (cf. Section 2) is reduced. In Table 3 the results for different values of r are shown. It presents the Hausdorff distances $d_H(\mathcal{P}^\rho, \mathcal{P}_{\tau^*}^{\beta^*,\rho})$ (as defined in (4)) and the mean distance over all points in $\mathcal{P}^\rho$ and $\mathcal{P}_{\tau^*}^{\beta^*,\rho}$ to their respective nearest points in $\mathcal{P}_{\tau^*}^{\beta^*,\rho}$ and $\mathcal{P}^\rho$, respectively, defined as

$$d_{\text{mean}}(\mathcal{P}^\rho, \mathcal{P}_{\tau^*}^{\beta^*,\rho}) = \frac{1}{|\mathcal{P}^\rho| \cdot |\mathcal{P}_{\tau^*}^{\beta^*,\rho}|} \sum_{\mathbf{x}' \in \mathcal{P}^\rho} \sum_{\mathbf{y}' \in \mathcal{P}_{\tau^*}^{\beta^*,\rho}} \left\{ \min_{\mathbf{y} \in \mathcal{P}_{\tau^*}^{\beta^*,\rho}} ||\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{y})||; \min_{\mathbf{x} \in \mathcal{P}^\rho} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y}')|| \right\}. \quad (35)$$

The distances are computed with respect to both X (i.e., the quantities d_H and d_{mean}) and X^{ref} (i.e., the quantities d_H^{ref} and $d_{\text{mean}}^{\text{ref}}$). Also the optimal tolerance value, τ^*, and the optimal objective value, z^*, in problem (26) are presented. All values in the table are averages over 10 runs.

r	d_H	d_{mean}	d_H^{ref}	$d_{\text{mean}}^{\text{ref}}$	τ^*	z^*
11	0.18	0.002	0.29	0.005	0	0.95
10	0.23	0.004	0.38	0.009	0	0.93
9	0.26	0.011	0.44	0.015	0	0.93
8	0.36	0.021	0.47	0.022	0	0.92
7	0.46	0.034	0.76	0.068	0.008	0.91
6	0.46	0.041	0.75	0.103	0.016	0.89
5	0.56	0.068	1.05	0.170	0.033	0.77
4	0.53	0.085	0.98	0.207	0.078	0.69
3	0.58	0.113	0.93	0.216	0.105	0.55
2	0.56	0.106	0.83	0.213	0.190	0.44
1	0.52	0.099	0.75	0.152	0.236	0.18

Table 3: Numerical results for varying numbers r of aggregated objectives.

There is no sharp breakpoint from which the error increases dramatically when r is decreased, thus it is hard to draw any general conclusions on how many objective functions that are required for a reasonably small error. Also, it is not possible to draw any clear conclusions on how the errors will turn out based on the correlation matrix. To enable a better interpretation of the numbers in Table 3, we present in Table 4—in the first two columns—the objective values (non-scaled) for the pair $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ of solutions defining the Hausdorff distance (cf. (4) and Figure 11), according to

$$(\hat{\mathbf{x}}, \hat{\mathbf{y}}) \in \arg\max_{\mathbf{x},\mathbf{y}} \left\{ \max_{\mathbf{x}\in\mathcal{P}^\rho} \min_{\mathbf{y}\in\mathcal{P}_{\tau^*}^{\beta^*,\rho}} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})||; \max_{\mathbf{y}\in\mathcal{P}_{\tau^*}^{\beta^*,\rho}} \min_{\mathbf{x}\in\mathcal{P}^\rho} ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})|| \right\}, \tag{36}$$

for a particular run resulting in the distance $||\mathbf{f}(\hat{\mathbf{x}}) - \mathbf{f}(\hat{\mathbf{y}})|| = 0.585$.

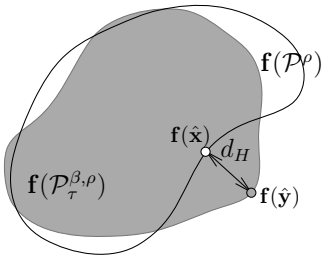


Figure 11: Illustration of the active pair of points $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ defining the Hausdorff distance.

The third column presents the difference between the largest and the smallest value of each function over the configuration space, defined as $\text{Span}_i = \max_{\mathbf{x}\in X} f_i(\mathbf{x}) - \min_{\mathbf{x}\in X} f_i(\mathbf{x})$. The fourth column presents the difference between function values for the pair $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ relative to Span_i, defined as $\text{Diff}(\mathbf{x}, \mathbf{y})_i = \frac{|f_i(\mathbf{x}) - f_i(\mathbf{y})|}{\text{Span}_i}$.

Table 4 shows that the difference between a point $\hat{\mathbf{x}}$ in the ρ-central part of the original Pareto optimal set and a point $\hat{\mathbf{y}}$ in the ρ-central part of the τ-Pareto optimal set of the reduced problem may be quite large. For the pair $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ such that $||\mathbf{f}(\hat{\mathbf{x}}) - \mathbf{f}(\hat{\mathbf{y}})|| = 0.585$, the relative differences between the points are over 20% for three of the objectives. We define

$$(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}) \in \arg \min_{\mathbf{x}\in\mathcal{P}^\rho,\ \mathbf{y}\in\mathcal{P}_{\tau^*}^{\beta^*,\rho}} \left| \, ||\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})|| - d_{\text{mean}}(\mathcal{P}^\rho, \mathcal{P}_{\tau^*}^{\beta^*,\rho}) \, \right| \tag{37}$$

to be the pair whose distance is closest to the mean distance. Table 5 indicates that even though the maximum difference may be large, the differences for a pair $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}})$, are small. For the pair $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}})$ in

i	$f_i(\hat{\mathbf{x}})$	$f_i(\hat{\mathbf{y}})$	Span_i	$\text{Diff}(\hat{\mathbf{x}},\hat{\mathbf{y}})_i$
1	1.905	1.990	0.69	0.12
2	2.251	2.380	0.58	0.22
3	1.280	1.338	0.79	0.07
4	2.721	2.783	1.15	0.05
5	1.447	1.546	1.29	0.08
6	1.632	1.662	0.88	0.03
7	-15550	-15550	1000	0.00
8	27460	27440	3940	0.01
9	189.8	191.3	43.8	0.04
10	738.6	822.9	303	0.28
11	1.061	1.100	0.39	0.10
12	2.486	3.267	3.42	0.23

Table 4: Difference between the pair $(\hat{\mathbf{x}},\hat{\mathbf{y}})$ defining the maximum Hausdorff distance $d_H = 0.585$.

i	$f_i(\widetilde{\mathbf{x}})$	$f_i(\widetilde{\mathbf{y}})$	Span_i	$\text{Diff}(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}})_i$
1	1.893	1.896	0.69	0.01
2	2.275	2.257	0.58	0.03
3	1.338	1.302	0.79	0.04
4	2.741	2.750	1.15	0.01
5	1.416	1.392	1.29	0.02
6	1.556	1.530	0.88	0.03
7	-15550	-15550	1000	0.00
8	27620	27647	3940	0.01
9	191.2	190.9	43.8	0.01
10	732.6	739.6	303	0.02
11	1.057	1.068	0.39	0.03
12	2.427	2.539	3.42	0.03

Table 5: Difference between the pair $(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}})$ whose distance is closest to the mean distance $d_{\text{mean}} = 0.093$.

Table 5 the relative difference is only a few percent for each of the objectives and with a maximum value of 4% and a mean value of 2%.

Table 6 shows how the value of the centrality parameter ρ affects the results. As expected, the distances

ρ	d_H	d_{mean}	d_H^{ref}	$d_{\text{mean}}^{\text{ref}}$	τ^*
0	0.53	0.044	0.69	0.051	0.017
0.05	0.49	0.050	0.89	0.131	0.022
0.10	0.48	0.047	0.81	0.117	0.021
0.15	0.46	0.041	0.75	0.103	0.016
0.20	0.37	0.034	0.65	0.092	0.015
0.25	0.34	0.028	0.77	0.147	0.014
0.30	0.22	0.018	0.78	0.195	0.007

Table 6: Numerical results when varying the centrality parameter ρ, for $r = 6$.

tend to decrease with an increased value of the centrality parameter, since more extreme Pareto optimal solutions are then filtered out, which the reduced problem does not catch. However, the pros of using a larger value of ρ must be balanced against the cons: a large value of ρ means that a smaller part of the Pareto optimal set is assumed to be interesting. Table 6 also shows that the values of the distance measures for the reference set X^{ref} do not decrease as nicely as for the set X. A plausible explanation for this is that for larger values of ρ, $|\mathcal{P}^\rho|$ is smaller, which increases the risk of over-fitting the decision variables when deciding on the reduced problem.

7 Conclusions

The motivation behind this work is the fact that industrial multi-objective optimization problems often are computationally hard to solve. Therefore, one must sometimes be content with a good approximation of $\mathcal{P}$. We have created two mathematical models for reducing the original problem by decreasing the cardinality of the set of objective functions using aggregations and such that the precision loss is minimized. The first model, leading to an exact solution given the stated problem, is based on an explicit characterization of Pareto optimality. However, the complexity of this model is far too high for applications to a practical case. The second model represents an approximation of this problem through a separation of the optimization into two sequential problems in such a way that the complexity is substantially reduced.

The second model has been applied to an industrial application which shows that it is possible to reduce the number of objectives in an optimization problem while getting measures of the sizes of the errors thereby induced. We have not found any clear a priori indicators on how many objectives are

required for reasonably small errors as in terms of, e.g., the correlation matrix for the objective functions. It is hard to measure a priori how much could be earned in reduction of the problem size for a certain level of the error. Since the error seems to increase rather smoothly, we suggest that the model should be applied to different numbers of objectives in the reduced problem before deciding on a suitable reduction. The decision maker (or, the problem owner) should decide on how large error is tolerable for the specific application and how much the problem-specific improvement in computation time by the reduction is worth. The numerical experiments also show that even if the Hausdorff distance between the respective Pareto optimal sets of the original and reduced problem is large, the distance between an average pair of points in these sets might be small.

To summarize, we have developed a method for problem reduction that can be applied to any multi-objective optimization problem. The approximation focuses on the most interesting part of the decision space, i.e., the solutions that are (near-)Pareto optimal. The outcome of the reduction is both a simplified problem formulation to use instead of the original one and a measure of the error thereby induced.

References

[1] P. Agrell, *On redundancy in multi criteria decision making*, European Journal of Operational Research, 98 (3) (1997), pp. 571–586.

[2] M. Bazaraa, H. Sherali, and C. Shetty, *Nonlinear Programming Theory and Algorithms, Third edition*, Wiley, Hoboken, NJ, USA, 2006.

[3] D. Brockhoff, T. Friedrich, N. Hebbinghaus, C. Klein, F. Neumann, and E. Zitzler, *Do additional objectives make a problem harder?*, in Genetic and Evolutionary Computation Conference (GECCO 2007), New York, NY, USA, 2007, ACM Press.

[4] D. Brockhoff and E. Zitzler, *On objective conflicts and objective reduction in multiple criteria optimization*, TIK Report 243, Institut für Technische Informatik und Kommunikationsnetze, ETH Zürich, Switzerland, Feb. 2006.

[5] C. Carlsson and R. Fullér, *Multiple criteria decision making: The case for interdependence*, Computers & Operations Research, 22 (3) (1995), pp. 251–260.

[6] W. M. Carlyle, J. W. Fowler, E. S. Gel, and B. Kim, *Quantitative comparison of approximate solution sets for bi-criteria optimization problems*, Decision Sciences, 34 (1) (2003), pp. 63–82.

[7] P. Czyżak and A. Jaszkiewicz, *Pareto simulated annealing—a metaheuristic technique for multiple-objective combinatorial optimization*, Journal of Multi-Criteria Decision Analysis, 7 (1998), pp. 34–47.

[8] K. Deb and D. Saxena, *On finding Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems*, KanGAL Report 2005011, Department of Mechanical Engineering, IIT Kanpur, India, 2005.

[9] M. Ehrgott and S. Nickel, *On the number of criteria needed to decide Pareto optimality*, Mathematical Methods of Operations Research, 55 (2002), pp. 329–345.

[10] R. Fourer, D. Gay, and B. Kerninghan, *AMPL – A Modeling Language for Mathematical Programming*, Brooks/Cole, Toronto, Canada, 2003.

[11] T. Gal and T. Hanne, *Consequences of dropping nonessential objectives for the application of MCDM-methods*, European Journal of Operational Research, 119 (2) (1999), pp. 373–378.

[12] T. Gal and H. Leberling, *Redundant objective functions in linear vector maximization problems and their determination*, European Journal of Operational Research, 1 (1977), pp. 176–184.

[13] ILOG Inc. *ILOG CPLEX 10.110*, `www.ilog.com/products/cplex`.

[14] J. Koski and R. Silvennoinen, *Norm methods and partial weighting in multicriterion optimization of structures*, International Journal for Numerical Methods in Engineering, 24 (1987), pp. 1101–1121.

[15] T. Lowe, J.-F. Thisse, J. Ward, and R. Wendell, *On efficient solutions to multiple objective mathematical programs*, Management Science, 30 (1984), pp. 1346–1349.

[16] The Mathworks. *MATLAB 7.3.0 (R2006b)*, `www.mathworks.se/products`.

[17] K. Miettinen, *Nonlinear Multiobjective Optimization*, Kluwer Academic Publishers, Boston, MA, USA, 1998.

[18] T. Mitchell, *Machine Learning*, McGraw-Hill, Singapore, 1997.

[19] G. Nemhauser and L. Wolsey, *Integer and Combinatorial Optimization*, Wiley, New York, NY, USA, 1988.

[20] A. Viana and J. Pinho de Sousa, *Using metaheuristics in multiobjective resource constrained project scheduling*, European Journal of Operational Research, 120 (2000), pp. 359–374.