Robust and Stochastically Weighted Multi-Objective Optimization Models and Reformulations *

Jian Hu Sanjay Mehrotra
Department of Industrial Engineering and Management Sciences
Northwestern University
Evanston, IL 60208
jianhu@northwestern.edu, mehrotra@iems.northwestern.edu

(Original draft 10/12/2010)

July 29, 2011

Abstract

We introduce and study a family of models for multi-expert multi-objective/criteria decision making. These models use a concept of weight robustness to generate a risk averse decision. In particular, the multi-expert multi-criteria robust weighted sum approach (McRow) introduced in this paper identifies a (robust) Pareto optimum decision that minimizes the worst case weighted sum of objectives over a given weight region. The corresponding objective value, called the robust-value of a decision, is shown to be increasing and concave in the weight set. Compact reformulations of the models with polyhedral and conic descriptions of the weight regions. The McRow model is developed further for stochastic multi-expert multi-criteria decision making by allowing ambiguity or randomness in the weight region as well as the objective functions. The properties of the proposed approach is illustrated with a few examples. The usefulness of the stochastic (McRow) model is demonstrated using a disaster planning example and an agriculture revenue management example.

Key Words: Pareto Optimality, Multi-Criterion Optimization, Multi-Expert Optimization, Robust Optimization, Weighted Sum Method, McRow

*The research is partially supported by the National Science Foundation grant CMMI-0928936 and Office of Naval Research grant ONR-N00014-09-10518

1 Introduction

Multi-objective optimization models arise in a large number of decision situations such as finance, energy, education, construction, transportation, ecology, communication, etc. Keeney and Raiffa (1976). For some recent applications see Liu et al. (2010), Marcenaro-Gutierreza et al. (2010), Kandil et al. (2010), Stepanov and Smith (2009), Côtéa et al. (2007), Alba et al. (2007), Mehrotra and Kim (2011). Multi-objective models have also emerged recently that involve stochastic objectives (see for example, Abdelaziz et al. (2007), Amjady et al. (2009), Lounis and Vanier (2000), Chen et al. (2010), Zhang et al. (2007), He et al. (2010), Hu et al. (2011)).

In multi-objective optimization a decision maker must consider several competing objectives before reaching a decision. Since it is not possible to simultaneously optimize for all objectives, in the simplest and widely used approach each objective is given a positive weight by its importance and the decision is arrived at by optimizing a weighted sum objective (see Steuer (1986), Miettinen (1999), Ehrgott (2005)). An optimal solution of a weighted sum function is a Pareto optimal (efficient, non-dominated) solution, i.e., this solution is not strictly dominated by any other solution. Alternative Pareto optimal solutions can be generated by using different weights under some mild conditions (see Miettinen (1999)). Methods such as the simple multi-attribute rating technique and point allocation (Edwards (1977), von Winterfeldt and Edwards (1986)), the swing weighting (von Winterfeldt and Edwards (1986)), the trade-off weighting (Keeney and Raiffa (1976)), and the analytic hierarchy process (Saaty (1980)) have been proposed to solicit decision maker's relative importance of decision criteria. Multi-objective models often also involve input from multiple experts (see Tsiporkovaa (2006) and references therein). In multi-expert decision making differing opinions of multiple experts must also be considered before arriving at a decision.

Current approaches to model based multi-expert multi-objective decision making has several shortcomings. There is no clear guidance to the decision maker to select a solution among (possibly a very large number of) generated Pareto optimal solutions. In multi-expert optimization the relative weights given by experts can differ significantly, since experts with differing opinion often assign different relative weights to objectives. Furthermore, there may be ambiguity in the weights provided by an expert, as it is often not easy to decide relative weights for each objective. In addition, studies have shown that relative weights provided by the same expert may depend on the elicitation approaches (see Schoemaker and Waid (1982), Borcherding et al. (1991), Weber and Borcherding (1993), Pöyhönen and Hämäläinen (2001)). In a model driven decision technique, there is a need to consistently handle these issues.

The problem of multi-expert weight aggregation has been considered in the literature. A survey of methods, taxonomy, and difficulties in arriving at decisions in these situations are discussed in Matsatsinis and Samaras (2001). Analytical methods for generating a consensus weight vector include the use of geometric mean of the weights (Dyer and Forman (1992)), recursive hierarchical aggregation method Tsiporkovaa (2006), or use the "relative-consensus expert weights" of expert weights (see for example Davis (1973), Davis (1996)). Logical rule based models have also been proposed for the aggregation purpose (for example, see Canfora and Troiano (2004) and references therein).

Contributions of this Paper. In this paper we propose and develop a minimax framework for multi-expert multi-objective decision making. We call this framework (McRow) (Multi-criteria Robust optimization with weight set). The proposed (McRow) models provide a way to address difference of opinion among experts, and errors in weight computations when multiple methods are used to elicit weights from an expert. Instead of a priori arriving at a consensus weight vector, the (McRow) framework finds a decision by defining the problem over a set of weights. This weight set can be defined using the alternative weight estimates from the same expert, or weights given by different experts. In the latter case the weight set may be viewed as the "disagreement set." We show that the optimum (McRow) objective value is an increasing concave function of the disagreement set. Managerially, this suggests the notion of a "cost of disagreement," or "dissatisfaction value." (McRow) can be viewed as a risk averse approach to decision making with respect to the dissatisfaction value.

More specifically, by using the concept of a robust Pareto optimum solution (see Section 2) we show that the robust value of (McRow) models is an increasing concave function of the weight set. We give compact reformulations of the (McRow) models as single mathematical programs. These reformulations are given for linear as well as conic weight regions. The (McRow) framework has the additional advantage of incorporating uncertainty in weights, and extending to the cases where functions are specified with uncertain parameters. For the stochastic problems we consider the cases where (i) the weights are given by uncertainty sets defined using the first and second moments; (ii) the weights are random and possibly correlated with the randomness of the objective functions. The tractability of (i) is shown. The model for (ii) is shown to be a two-stage stochastic program. Finally, for completeness we give reformulations of both deterministic and stochastic objective problems when l_p-norm and Tchebycheff functions are used to scalarize multiple objectives (see Miettinen (1999)).

Some recent work in multi-objective optimization defines the concept of solution robustness by considering stability of the optimal solution with respect to errors in the objective function parameters (see Soares et al. (2009); Cromvik et al. (2011) and references therein). Soares et al. (2009) define the concept of absolute and relative robustness as the expected value of the difference of a given composite utility function to a utility value calculated at a reference point, where the expectation is taken with respect to random function parameters. Cromvik et al. (2011) take a min-max approach over function parameters in their definition. An alternative definition of robust solution is considered in Deb and Gupta (2005) In contrast, we consider inconsistency and ambiguity in decision maker's ability to provide relative weights to each objective to define the concept of weight-robustness. The (Sr-McRow) model in this paper uses parameter and weight uncertainty together to define the concept of a robust stochastic Pareto optimum solution.

This paper is organized as follows. Section 2 introduces the concept of robust Pareto optimality and robust value. We describe the relationship between Pareto optimality and the robust Pareto optimality and also discuss a geometric interpretation of robust Pareto optimality in this section. Here we show that the robust value function is an increasing concave function of the weight set. Sections 3 gives reformulations of the robust weighted sum model with polyhedral weight regions. Section 4 gives reformulations of (McRow) when the weight region $\mathcal{W}$ is a cone. This section also considers the case where the regions are given by an ellipsoid. Section 5 further develop the (McRow) concept for stochastic multi-objective optimization and introduces a moment robust (Mr-McRow) and stochastically robust (Sr-McRow) Models. This section also introduces the concept of stochastic Pareto optimality under various assumptions on the description of the weight set. Section 6 gives two applications of the (Sr-McRow) model. The first example problem is that of identifying a suitable alternative to reduce the impact of mud volcano disaster in Sidoarjo, Indonesia. The second example illustrates the use of (Sr-McRow) framework by viewing a revenue management problem as a stochastic multi-objective optimization problem. In particular, detailed data from US department of agriculture (USDA) is used for risk averse decision making for aggregate crop planning for farmers in Illinois. The reformulations of (McRow, Mr-McRow, Sr-McRow) models with l_p-norm and Tchebycheff function norm are given in Appendix B. Some known definition and a few proofs are presented in Appendices A and C.

2 Robust Pareto Optimality

Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a nonempty set and $f(x) = (f_1(x), \ldots, f_m(x))^T : \mathcal{X} \to \mathbb{R}^m$ be a vector function. A deterministic multi-objective optimization problem is written as follows:

$$\min_{x \in \mathcal{X}} \; f(x). \tag{MOO}$$

Given weights $w \in \mathbb{R}^m_+$, the weighted sum version of (MOO) (e.g., see Zadeh (1963) and Gass and Saaty (1955)) is given as

$$\min_{x \in \mathcal{X}} \; w^T f(x). \tag{WSO}$$

Let

$$\mathcal{W}_f = \{w \in \mathbb{R}^m_+ \mid \|w\|_1 = 1\}$$

and let $\mathcal{W} \subseteq \mathcal{W}_f$ represent a nonempty set of weights. This construction ensures that all weights in the set $\mathcal{W}$ are non-negative and they add to one. We consider a robust version of (WSO) as follows.

Definition 2.1 *Let $\mathcal{W}$ be a closed set. An optimal solution of*

$$M(\mathcal{W}) = \min_{x \in \mathcal{X}} \max_{w \in \mathcal{W}} \; w^T f(x) \tag{McRow}$$

is called a $\mathcal{W}$-robust (in short, robust) Pareto optimal solution of (MOO)*. The function value $M(\mathcal{W})$ is called $\mathcal{W}$-robust (in short, robust) value of* (MOO)*.*

Observe that for $\mathcal{W}_1 \subseteq \mathcal{W}_2$, $M(\mathcal{W}_2) - M(\mathcal{W}_1) \geq 0$. In a multi-expert decision making the size of the weight set $\mathcal{W}$ indicates the magnitude of disagreement between experts. With this interpretation of $\mathcal{W}$, the robust value is a "cost of disagreement" which according to this observation is increasing with increase in disagreement.

2.1 Minimizing Decision Dissatisfaction

In the introduction we discussed several technical motivations for using (McRow). In this subsection we give a managerial motivation for (McRow) formulation of (MOO). An expert would be most satisfied if a decision is optimum for the weights provided by this expert. We may define an expert's dissatisfaction as the difference between the best objective achievable for this expert to the objective value in the proposed decision. We may now consider a model that finds a solution minimizing the worst weighted sum of dissatisfaction values of experts. More formally, assume that we have N decision makers. Let w^i be the weight vector specified by decision maker i, $i = 1, \ldots, N$. Let $\hat{f}^i(x) = {w^i}^T f(x)$ and $\hat{f}^{i*} := \min_{x\in\mathcal{X}} \hat{f}^i(x)$, be the best objective value of (MOO) from the perspective of decision maker i. Let $\bar{f}^i(x) := \hat{f}^i(x) - \hat{f}^{i*}$ and $\bar{f}(x) := (\bar{f}^1(x), \ldots, \bar{f}^N(x))^T$. Note that $\bar{f}^i(x) \geq 0$ for all $x \in \mathcal{X}$. Now our decision problem is

$$\min_{x\in\mathcal{X}} \max_{\bar{w}\in\bar{\mathcal{W}}} \bar{w}^T \bar{f}(x),$$

where the set $\bar{\mathcal{W}}$ may have a further interpretation of describing the "relative importance" of different decision makers. The weight $(\frac{1}{N}, \ldots, \frac{1}{N})$ gives equal importance to all decision makers. We may further use the l_p-norm or the Tchebycheff approaches when constructing the dissatisfaction functions $\bar{f}(x)$ (see Appendix B).

2.2 Properties of the McRow Model

We first show that an optimum solution of (McRow) is a Pareto optimum solution of (MOO). For completeness the definition of a Pareto optimal solution and related concepts are stated in Appendix A. Throughout e denotes the vector whose elements are all 1; e_i denotes the vector whose ith element is 1 and others are 0.

Theorem 2.2 *Let $\mathcal{W} \subseteq \mathcal{W}_f$ be a closed set and x^* be a solution of* (McRow)*. Then,*

1. *x^* is a weakly Pareto optimal solution of* (MOO)*.*
2. *If all components $w_i > 0$ for every $w \in \mathcal{W}$, then x^* is a Pareto optimal solution of* (MOO)*.*
3. *If x^* is a unique optimal solution of* (McRow) *below, then x^* is a Pareto optimal solution of* (MOO)*.*
4. *If all components $w_i > 0$ for every $w \in \mathcal{W}$, then x^* is properly Pareto optimal for* (MOO)*.*

Proof: 1. If x^* is not weakly Pareto optimal, there exists $\hat{x} \in \mathcal{X}$ such that $f_i(\hat{x}) < f_i(x^*)$ for all $i = 1, \ldots, m$. Then we have a contradiction as

$$\min_{x\in\mathcal{X}} \max_{w\in\mathcal{W}} w^T f(x) \leq \max_{w\in\mathcal{W}} w^T f(\hat{x}) < \max_{w\in\mathcal{W}} w^T f(x^*) = \min_{x\in\mathcal{X}} \max_{w\in\mathcal{W}} w^T f(x). \tag{2.1}$$

The last equality in (2.1) holds because x^* is a solution of (McRow).

2. If x^* is not Pareto optimal, there exists $\hat{x} \in \mathcal{X}$ such that $f_i(\hat{x}) \leq f_i(x^*)$ for all $i = 1, \ldots, m$ and $f_j(\hat{x}) < f_j(x^*)$ for at least one index $j \in \{1, \ldots, m\}$. It is obvious that the arguments for the contradiction in (2.1) still hold for $\mathcal{W} \subseteq \{w \in \mathbb{R}^m \,|\, w_i > 0,\ i = 1, \ldots, m\}$.

3. Since each w_i is not strictly positive, we have

$$\max_{w \in \mathcal{W}} w^T f(\hat{x}) \leq \max_{w \in \mathcal{W}} w^T f(x^*),$$

if x^* is not Pareto optimal. Then $\hat{x}$ is another optimal solution of (McRow). This is contradictory that x^* is a unique optimal solution of (McRow).

4. We have shown that x^* is Pareto optimal in Part 2. Let us prove that x^* is properly Pareto optimal. Denote

$$\kappa = (m-1) \max_{\substack{w \,\in\, \mathcal{W} \\ i,\, j \,\in\, \{1, \dots, m\}}} \frac{w_j}{w_i}.$$

By the assumption, there exists a constant $\epsilon > 0$ such that $\epsilon \leq w_i < 1$ $(i = 1, \dots, m)$. It means that $\kappa < (m-1)/\epsilon$. Now suppose that x^* is not properly Pareto optimal. Then for some $i \in \{1, \dots, m\}$ (which we fix) and for $x \in \mathcal{X}$ such that $f_i(x^*) > f_i(x)$ we have

$$f_i(x^*) - f_i(x) > \kappa(f_j(x) - f_j(x^*)),$$

for all $j \in \{1, \dots, m\}$ such that $f_j(x^*) < f_j(x)$. For a given $w \in \mathcal{W}$, we write

$$f_i(x^*) - f_i(x) > (m-1)\frac{w_j}{w_i}(f_j(x) - f_j(x^*)).$$

Multiplying the both sides by $w_i/(m-1) > 0$, we obtain

$$\frac{w_i}{m-1}(f_i(x^*) - f_i(x)) > w_j(f_j(x) - f_j(x^*))\big(> 0 \geq w_l(f_l(x) - f_l(x^*))\big),$$

where l differs from the fixed index i and the specified indices j. It follows that

$$w_i(f_i(x^*) - f_i(x)) > \sum_{\substack{j \,=\, 1 \\ j \,\neq\, i}}^{m} w_j(f_j(x) - f_j(x^*)),$$

or equivalently

$$w^T f(x^*) > w^T f(x).$$

Since w is arbitrary, we have

$$\min_{x \in \mathcal{X}} \max_{w \in \mathcal{W}} w^T f(x) = \max_{w \in \mathcal{W}} w^T f(x^*) > \max_{w \in \mathcal{W}} w^T f(x) \geq \min_{x \in \mathcal{X}} \max_{w \in \mathcal{W}} w^T f(x).$$

The contradiction shows that x^* is properly Pareto optimal. □

The following proposition shows that we can take the convex hull of $\mathcal{W}$ in (McRow) without changing our problem.

Proposition 2.3 *For given sets $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{W} \subseteq \mathbb{R}^m$, x^* is an optimum solution of (McRow) if* *and only if it is an optimum solution of the following problem:*

$$\min_{x \in \mathcal{X}} \max_{w \in conv(\mathcal{W})} w^T f(x). \tag{2.2}$$

Proof: It is convenient to consider the reformulation of (McRow), $\min\{\eta \,|\, (\eta, x) \in \mathcal{G}\}$, where

$$\mathcal{G} = \left\{(\eta, x) \in \mathbb{R}^{n+1} \,|\, x \in \mathcal{X}, w^T f(x) \leq \eta, \forall w \in \mathcal{W}\right\}.$$

Also denote

$$\mathcal{G}' = \left\{(\eta, x) \in \mathbb{R}^{n+1} \,|\, x \in \mathcal{X}, w^T f(x) \leq \eta, \forall w \in \text{conv}\,(\mathcal{W})\right\}.$$

Obviously $\mathcal{G}' \subseteq \mathcal{G}$. We now claim $\mathcal{G} \subseteq \mathcal{G}'$. This is trivial when $\mathcal{G}$ is an empty set. For $(\eta, x) \in \mathcal{G}$, we want to show $(\eta, x) \in \mathcal{G}'$. For an arbitrarily given $w \in \text{conv}\,(\mathcal{W})$, there exist $w^i \in \mathcal{W}$, $i = 1, \ldots, k$, such that $w = \sum_{i=1}^k \lambda_i w^i$ where $\lambda_i \geq 0$ $(i = 1, \ldots, k)$ and $\sum_{i=1}^k \lambda_i = 1$ for some $k \leq m + 1$ by Carathéodory's theorem. ${w^i}^T f(x) \leq \eta$ $(i = 1, \ldots, k)$ should be satisfied for $(\eta, x) \in \mathcal{G}$. Thus $w^T f(x) = \sum_{i=1}^k \lambda_i {w^i}^T f(x) \leq \sum_{i=1}^k \lambda_i \eta = \eta$. It follows that $(\eta, x) \in \mathcal{G}'$. Thus, $\mathcal{G} \subseteq \mathcal{G}'$. □

Because of Proposition 2.3, we will assume that $\mathcal{W}$ is a closed convex subset of $\mathcal{W}_f$ in the rest of this paper. We now show that the robust value function $M(\mathcal{W})$ is a concave function of $\mathcal{W}$. Let $\lambda\mathcal{W} = \{\lambda w \mid w \in \mathcal{W}\}$ and denote the Minkowski sum of two sets as $\mathcal{W}_1 \oplus \mathcal{W}_2 = \{w_1 + w_2 \mid w_1 \in \mathcal{W}_1, w_2 \in \mathcal{W}_2\}$. Also for given closed convex sets $\mathcal{W}_1 \subseteq \mathcal{W}_2 \subseteq \mathbb{R}^m$, let $\mathcal{W}(\lambda) = ((1-\lambda)\mathcal{W}_1 \oplus \lambda\mathcal{W}_2) \cap \mathcal{W}_f$ for some $\lambda \in [0, 1]$.

Theorem 2.4 *The function $M(\mathcal{W}(\lambda))$ is a nondecreasing and concave function of λ.*

The following technical propositions showing that $\mathcal{W}(\lambda)$ is nested with increasing λ, and the "linearity" of Minkowski sum of two convex sets are needed in the proof of Theorem 2.4.

Proposition 2.5 *For $0 \leq \lambda_1 \leq \lambda_2 \leq 1$, $\mathcal{W}(\lambda_1) \subseteq \mathcal{W}(\lambda_2)$.*

Proof: Let $\mathcal{G}(\lambda) = (1-\lambda)\mathcal{W}_1 \oplus \lambda\mathcal{W}_2$ and thus $\mathcal{W}(\lambda) = \mathcal{G}(\lambda) \cap \mathcal{W}_f$. It is trivial to see that $\mathcal{G}(\lambda_1)$, $\mathcal{G}(\lambda_2)$ are convex and $\mathcal{W}_1 \subseteq \mathcal{G}(\lambda_1)$ and $\mathcal{W}_1 \subseteq \mathcal{G}(\lambda_2)$. By definition, for any $w \in \mathcal{W}(\lambda_1)$, there exist $w_1 \in \mathcal{W}_1$ and $w_2 \in \mathcal{W}_2$ such that

$$w = (1-\lambda_1)w_1 + \lambda_1 w_2 = \left(1 - \frac{\lambda_1}{\lambda_2}\right) w_1 + \frac{\lambda_1}{\lambda_2}((1-\lambda_2)w_1 + \lambda_2 w_2).$$

Since w_1, $(1-\lambda_2)w_1 + \lambda_2 w_2 \in \mathcal{G}(\lambda_2)$, we have that $w \in \mathcal{G}(\lambda_2)$ and hence $\mathcal{G}(\lambda_1) \subseteq \mathcal{G}(\lambda_2)$. It also follows that $\mathcal{W}(\lambda_1) = \mathcal{G}(\lambda_1) \cap \mathcal{W}_f \subseteq \mathcal{G}(\lambda_2) \cap \mathcal{W}_f = \mathcal{W}(\lambda_2)$. □

Proposition 2.6 (Theorem 3.2 in Rockafellar (1972)) *Let $\mathcal{W}$ be a convex set. For $\lambda_1, \lambda_2 \geq 0$, it follows that $(\lambda_1 + \lambda_2)\mathcal{W} = \lambda_1\mathcal{W} \oplus \lambda_2\mathcal{W}$.*

Proof: (Theorem 2.4). Choose λ_1, λ_2 such that $0 \leq \lambda_1 \leq \lambda_2 \leq 1$. It follows by Proposition 2.5 that $\mathcal{W}(\lambda_1) \subseteq \mathcal{W}(\lambda_2)$ so that $M(\mathcal{W}(\lambda_1)) \leq M(\mathcal{W}(\lambda_2))$. We now prove the concavity of $M(\mathcal{W}(\cdot))$. Let

$\mathcal{G}(\lambda) = (1-\lambda)\mathcal{W}_1 \oplus \lambda\mathcal{W}_2$ and thus $\mathcal{W}(\lambda) = \mathcal{G}(\lambda) \cap \mathcal{W}_f$. For $\delta \in [0,1]$, we have

$$\begin{aligned}
&\mathcal{G}((1-\delta)\lambda_1 + \delta\lambda_2) \\
&= (1 - ((1-\delta)\lambda_1 + \delta\lambda_2))\mathcal{W}_1 \oplus ((1-\delta)\lambda_1 + \delta\lambda_2)\mathcal{W}_2 \\
&= ((1-\delta)(1-\lambda_1) + \delta(1-\lambda_2))\mathcal{W}_1 \oplus ((1-\delta)\lambda_1 + \delta\lambda_2)\mathcal{W}_2 \\
&= (1-\delta)((1-\lambda_1)\mathcal{W}_1 \oplus \lambda_1\mathcal{W}_2) \oplus \delta((1-\lambda_2)\mathcal{W}_1 \oplus \lambda_2\mathcal{W}_2) \qquad \text{(by Propositon 2.6)} \\
&= (1-\delta)\mathcal{G}(\lambda_1) \oplus \delta\mathcal{G}(\lambda_2).
\end{aligned}$$

Since $(1-\delta)(\mathcal{G}(\lambda_1) \cap \mathcal{W}_f) \oplus \delta(\mathcal{G}(\lambda_2) \cap \mathcal{W}_f) \subseteq (1-\delta)\mathcal{W}_f \oplus \delta\mathcal{W}_f = \mathcal{W}_f$ and $(1-\delta)(\mathcal{G}(\lambda_1) \cap \mathcal{W}_f) \oplus \delta(\mathcal{G}(\lambda_2) \cap \mathcal{W}_f) \subseteq (1-\delta)\mathcal{G}(\lambda_1) \oplus \delta\mathcal{G}(\lambda_2)$, we also have

$$((1-\delta)\mathcal{G}(\lambda_1) \oplus \delta\mathcal{G}(\lambda_2)) \cap \mathcal{W}_f \supseteq (1-\delta)\mathcal{W}(\lambda_1) \oplus \delta\mathcal{W}(\lambda_2).$$

Thus, it follows that

$$\begin{aligned}
M(\mathcal{W}((1-\delta)\lambda_1 + \delta\lambda_2)) &= \min_{x\in\mathcal{X}} \max_{w \in \mathcal{G}((1-\delta)\lambda_1+\delta\lambda_2) \cap \mathcal{W}_f} w^T f(x) \\
&= \min_{x\in\mathcal{X}} \max_{w \in ((1-\delta)\mathcal{G}(\lambda_1) \oplus \delta\mathcal{G}(\lambda_2)) \cap \mathcal{W}_f} w^T f(x) \\
&\geq \min_{x\in\mathcal{X}} \max_{w \in (1-\delta)\mathcal{W}(\lambda_1) \oplus \delta\mathcal{W}(\lambda_2)} w^T f(x) \\
&= \min_{x\in\mathcal{X}} \max_{\substack{w^1 \in \mathcal{W}(\lambda_1) \\ w^2 \in \mathcal{W}(\lambda_2)}} (1-\delta){w^1}^T f(x) + \delta {w^2}^T f(x) \\
&\geq (1-\delta) \min_{x\in\mathcal{X}} \max_{w^1 \in \mathcal{W}(\lambda_1)} {w^1}^T f(x) + \delta \min_{x\in\mathcal{X}} \max_{w^2 \in \mathcal{W}(\lambda_2)} {w^2}^T f(x) \\
&= (1-\delta)M(\mathcal{W}(\lambda_1)) + \delta M(\mathcal{W}(\lambda_2)).
\end{aligned}$$

□

The definition of $\mathcal{W}(\lambda)$ and due to Proposition 2.5 the weight (disagreement) set is parameterized by λ. Theorem 2.4 describes the nondecreasing concavity of the cost of disagreement (robust value of (MOO)). This suggests that according to the (McRow) model there is a diminishing rate of increase in the cost disagreement (dissatisfaction) with increasing disagreement set.

2.3 A Geometric Interpretation of Robust Pareto Optimality

Denote $f(\mathcal{X}) = \{y \in \mathbb{R}^m \,|\, y = f(x), \forall x \in \mathcal{X}\}$ and rewrite (McRow) as

$$\min_{y \in f(\mathcal{X})} \left\{ \phi(y) = \max_{w \in \mathcal{W}} w^T y \right\}. \tag{2.3}$$

The function $\phi(\cdot)$ is a support function of the set $\mathcal{W}$. Thus, $y^T w = \phi(y)$ is a supporting hyperplane of the set $\mathcal{W}$ for each normal vector $y \in f(\mathcal{X})$. Let $\hat{w}(y)$ be an intersection point of this supporting hyperplane and the set $\mathcal{W}$. The point $\hat{w}(y)$ is also an optimal solution of the inner maximization problem of (2.3). A robust Pareto optimal solution is a minimizer of the inner product of vectors $f(x)$ and $\hat{w}(f(x))$.

Let us assume that the order of the minimization and maximization problems in (2.3) can be switched, i.e., there is no gap between (2.3) and the following problem

$$\max_{w \in \mathcal{W}} \left\{ \psi(w) = \min_{y \in f(\mathcal{X})} w^T y \right\}. \tag{2.4}$$

By Sion's minimax theorem (Sion (1958) and Komiya (1988)), this assumption is satisfied if $\mathcal{X}$ is a compact convex set and $f(\cdot)$ is lower semicontinuous and quasi-convex on $\mathcal{X}$. Let $\hat{y}(w)$ be the minimizer of the inner minimization problem of (2.4). Thus, $\mathcal{Y}(\mathcal{W}) = \{\hat{y}(w) \mid w \in \mathcal{W}\}$ is a set of Pareto optimal objective vectors of (MOO). Moreover, it follows from Theorem 3.1.4 in Miettinen (1999) that all the Pareto optimal objective vectors of (MOO) can be attained when (MOO) is convex and $\mathcal{W}_f$ is chosen as the weight region. $\psi(\cdot)$ can also be regarded as a support function of the set $f(\mathcal{X})$ for the surface of the downside boundary[1]. $w^T y = \psi(w)$ is a supporting hyperplane of set $f(\mathcal{X})$ at point $\hat{y}(w)$ for each normal vector $w \in \mathcal{W}$. Denote $y^\ell(w)$ as the intersection point of this hyperplane and line $\ell = \{y \in \mathbb{R}^m \mid y_1 = \cdots = y_m\}$, i.e., $y^\ell(w)$ is the solution of the following equations:

$$\begin{cases} w^T y = \psi(w), \\ y_1 = \cdots = y_m. \end{cases}$$

We have $y^\ell(w) = (\psi(w), \ldots, \psi(w))$ for $w^T e = 1$. Recall that problem (2.4) maximizes $\psi(w)$, which means that (McRow) pushes the intersection point $y^\ell(w)$ upward along line ℓ. Figure 1 gives a 2-dimensional example to explain the geometry of robust Pareto soutions. For two weight vector, $w_1, w_2 \in \mathcal{W}$, let $h(w_1)$ and $h(w_2)$ denote the corresponding hyperplanes. The intersections of line ℓ and these hyperplanes are labelled by $y^\ell(w_1) = (\psi(w_1), \psi(w_1))$ and $y^\ell(w_2) = (\psi(w_2), \psi(w_2))$. Since $\psi(w_2) > \psi(w_1)$, w_2 is preferable to w_1 in this case. Figure 1 also draws the optimal solution w^* of problem (2.4). An observation is that (McRow) solution pushes the optimal objective vector $\hat{y}(w)$ close to line ℓ while approaching $\hat{y}(w^*)$. This indicates that (McRow) recommends a solution which trades off different objectives to achieve a balance in their values.

3 Reformulation with Polyhedral Weight Region

We now give tractable reformulations of (McRow) as equivalent mathematical programs. If the vector-valued function $f(\cdot)$ is convex, these reformulations are convex optimization problems. We start by assuming that $\mathcal{W}$ is a polytope. The simplest situation is when the set it is taken as a convex hull of a finite number of weights. Alternatively, such construction is possible when the set is constructed by taking perturbations along different rays from a reference weight center. If extreme points of $\mathcal{W}$ are known as in these examples, we can easily solve (McRow) by a reformulation using the following proposition. Because of Proposition 2.3, this proposition is also applicable for the model where $\mathcal{W}$ is described by the convex hull of a finite set of points.

Proposition 3.1 *Let $w^1, \ldots, w^d$ be all the extreme points of a polytope $\mathcal{W}$. Then* (McRow) *is*

[1] The support function of set $f(\mathcal{X})$ is traditionally defined as $\max_{y \in f(\mathcal{X})} w^T y$ for the surface of the upside boundary.

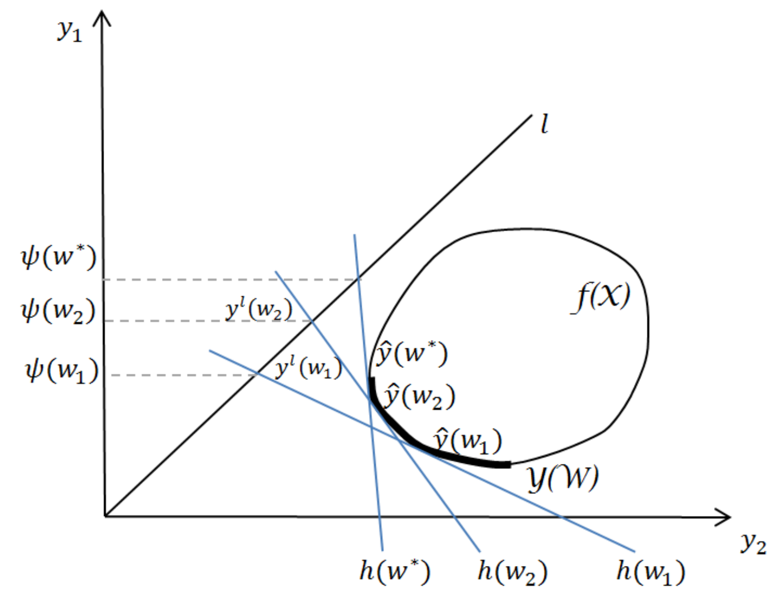

Figure 1: A Geometric Interpretation of Robust Pareto Optimality

equivalent to

$$\begin{aligned} \min_{x,\zeta}\ & \zeta \\ s.t.\ & {w^i}^T f(x) \le \zeta, \qquad i = 1, \ldots, d, \\ & x \in \mathcal{X}. \end{aligned} \tag{3.1}$$

Proof: It directly follows by Proposition 2.3 since $\mathcal{W}$ is the convex hull of its extreme points. □

A perturbation region around a fixed reference point is a typical approach to construct uncertain parameters of optimization problems (see Ben-Tal et al. (2009)). Let $\hat{w}$ be a reference point and Q be a $m \times k$ coefficient matrix used to construct a perturbation region around $\hat{w}$. We denote

$$\mathcal{W} = \{w \in \mathbb{R}^m \,|\, w = \hat{w} + Qv,\ v \in \mathcal{V}\}, \tag{3.2}$$

where $\mathcal{V}$ is a subset of

$$\mathcal{V}_f = \{v \in \mathbb{R}^k \,|\, \hat{w} + Qv \ge 0,\ e^T(\hat{w} + Qv) = 1\}. \tag{3.3}$$

The use of $\mathcal{V}_f$ in defining the perturbation region ensures that $\mathcal{W} \subseteq \mathcal{W}_f$ in (3.2). Note that $\mathcal{W} = \mathcal{V}$ when $\hat{w}$ is 0 and Q is a $m \times m$ identity matrix. Hence (3.2) also allows us to define weights directly, i.e., without using a reference weight $\hat{w}$ and a perturbation matrix. When $\mathcal{W}$ is given by linear constraints, in general it is not easy to enumerate all the extreme points of $\mathcal{W}$. Furthermore, the number of extreme points of $\mathcal{W}$ grows exponentially as the dimension m increases. We now give a reformulation of (McRow) for the case where $\mathcal{W}$ is defined by linear equality and inequality constraints.

Theorem 3.2 *Let* $\mathcal{W}$ *be denoted as in* (3.2) *and*

$$\mathcal{V} = \{v \in \mathcal{V}_f \,|\, Av = a,\ Bv \ge b\} \tag{3.4}$$

be a nonempty set. Then (McRow) *is equivalent to*

$$\begin{aligned}\min_{x,y,z,s,\zeta}\ & \hat{w}^T s + \zeta - a^T y - b^T z \\ s.t.\ & Q^T s + A^T y + B^T z = 0, \\ & f(x) - s - \zeta e \le 0, \\ & x \in \mathcal{X},\ z \ge 0.\end{aligned} \tag{3.5}$$

Proof: Assume $a \in \mathbb{R}^i$ and $b \in \mathbb{R}^j$. Let

$$\widehat{A} = \begin{bmatrix} I_{k\times k} \\ 0_{i\times k} \\ 0_{j\times k} \end{bmatrix}, \qquad \widehat{c} = \begin{bmatrix} 0_{k\times 1} \\ a \\ b \end{bmatrix},$$

and polyhedral cone $\mathbf{K} = \{(v, y, z) \in \mathbb{R}^{k+i+j} \mid Av = y,\ Bv \ge z\}$. Thus, we can rewrite $\mathcal{V}$ in (3.4) as

$$\mathcal{V} = \{v \in \mathcal{V}_f \mid \widehat{A}v + \widehat{c} \in \mathbf{K}\}.$$

It implies that the perturbation set in (3.4) is a special case of the general conical definition (4.1) which will be studied in Section 4. The remaining proof of this theorem is a special case of the proof of Theorem 4.1. □

In special cases the perturbation set $\mathcal{V}$ is given as a simplex $\{v \in \mathcal{V}_f \mid e^T v \le 1,\ v \ge 0\}$, l_1-norm $\{v \in \mathcal{V}_f \mid \|v\|_1 \le 1\}$, or l_∞-norm $\{v \in \mathcal{V}_f \mid \|v\|_\infty \le 1\}$. The following corollary gives the corresponding reformulations.

Corollary 3.3 *Let* $\mathcal{W}$ *be denoted in* (3.2)*.*

1. *If* $\mathcal{V} = \{v \in \mathcal{V}_f \mid e^T v \le 1,\ v \ge 0\}$*, then* (McRow) *is equivalent to*

$$\begin{aligned}\min_{x,s,\zeta,\eta}\ & \hat{w}^T s + \zeta + \eta \\ s.t.\ & Q^T s - \eta e \le 0, \\ & f(x) - s - \zeta e \le 0, \\ & x \in \mathcal{X},\ \eta \ge 0.\end{aligned} \tag{3.6}$$

2. *If* $\mathcal{V} = \{v \in \mathcal{V}_f \mid \|v\|_1 \le 1\}$*, then* (McRow) *is equivalent to*

$$\begin{aligned}\min_{x,s,\zeta,\eta}\ & \hat{w}^T s + \zeta + \eta \\ s.t.\ & -\eta e \le Q^T s \le \eta e \\ & f(x) - s - \zeta e \le 0, \\ & x \in \mathcal{X},\ \eta \ge 0.\end{aligned} \tag{3.7}$$

3. If $\mathcal{V} = \{v \in \mathcal{V}_f \mid \|v\|_\infty \le 1\}$, *then* (McRow) *is equivalent to*

$$\begin{aligned} \min_{x,y,z,s,\zeta} \; & \hat{w}^T s + \zeta + e^T y + e^T z \\ s.t. \; & Q^T s + y - z = 0, \\ & f(x) - s - \zeta e \le 0, \\ & x \in \mathcal{X}, \; y \ge 0, \; z \ge 0. \end{aligned} \tag{3.8}$$

Proof: See Appendix C.1. □

Remark 3.4 *In* (3.7) *the constraint* $-\eta e \le Q^T s \le \eta e$ *can also be written as* $\|Q^T s\|_\infty \le \eta$. *Hence, we reformulate* (3.7) *as*

$$\begin{aligned} \min_{x,s,\zeta} \; & \hat{w}^T s + \zeta + \|Q^T s\|_\infty \\ s.t. \; & f(x) - s - \zeta e \le 0, \\ & x \in \mathcal{X}. \end{aligned} \tag{3.9}$$

Similarly, (3.8) *is also equivalent to*

$$\begin{aligned} \min_{x,s,\zeta} \; & \hat{w}^T s + \zeta + \|Q^T s\|_1 \\ s.t. \; & f(x) - s - \zeta e \le 0, \\ & x \in \mathcal{X}. \end{aligned} \tag{3.10}$$

Theorem 2.4 showed the concavity of the robust value of (MOO) for general nested weight sets. Below we give a specific construction of the polyhedral sets and restate the concavity result of this case.

Corollary 3.5 *For a given* $d \ge 0$, *let*

$$\mathcal{V}(\lambda) = \{v \in \mathcal{V}_f \mid Av = a, Bv \ge b - \lambda d\}$$

be nonempty, and $\mathcal{W}(\lambda) := \{w \in \mathbb{R}^m \mid w = \hat{w} + Qv, \; v \in \mathcal{V}(\lambda)\}$. *Then, the robust value* $M(\mathcal{W}(\lambda))$ *of McRow is a nondecreasing concave function of* λ.

Proof: See Appendix C.2. □

3.1 Example: Using McRow with Ambiguous Weights

We consider a textbook (MOO) problem from Korhonen et al. (1997) and Ehrgott (2005) to illustrate some of the properties of (McRow):

$$\begin{array}{llrrrrrrrl} \min & f_1(x) = & & -11x_2 & -11x_3 & -12x_4 & -9x_5 & -9x_6 & +9x_7 & \\ \min & f_2(x) = & -11x_1 & & -11x_3 & -9x_4 & -12x_5 & -9x_6 & +9x_7 & \\ \min & f_3(x) = & -11x_1 & -11x_2 & & -9x_4 & -9x_5 & -12x_6 & -12x_7 & \\ \text{s.t.} & & x_1 & +x_2 & +x_3 & +x_4 & +x_5 & +x_6 & +x_7 & = 1, \\ & & & & & & & & x & \ge 0. \end{array} \tag{3.11}$$

Assume that the analytic hierarchy process (Saaty (1980)) is used to assign weights for the three objectives in (3.11). We assume that the decision maker is unsure about the relative importance of f_1 and f_2. Consequently, the decision maker allows three possible scale indices of this pair as $f_1/f_2 = 1/2$, 1 or 2. Similarly, the decision maker is unsure about the relative importance of f_1 and f_3 and assigns f_1/f_3 as 2, 3 or 4. This ambiguous assessment of relative importance of these objectives produces nine combinations. The first nine rows in Table 1 list the weight vectors for these combinations. Note that although the analytic hierarchy process asks for an integer scale of $1, \dots, 9$, we allow some non-integer scales to ensure consistency of relative weights in the weight matrix (not shown here).

Table 1: Comparison of Different Approaches to Problem (3.11)

f_1/f_2	f_1/f_3	Relative Weights	Value of f	Optimal Solution
1/2	2	(0.2857, 0.5714, 0.1429)	(-9, -12, -9)	(0, 0, 0, 0, 1, 0, 0)
1/2	3	(0.3, 0.6, 0.1)	(-9, -12, -9)	(0, 0, 0, 0, 1, 0, 0)
1/2	4	(0.3077, 0.6154, 0.0769)	(-9, -12, -9)	(0, 0, 0, 0, 1, 0, 0)
1	2	(0.4, 0.4, 0.2)	(-12, -9, -9)	(0, 0, 0, 1, 0, 0, 0)
1	3	(0.4286, 0.4286, 0.1429)	(-12, -9, -9)	(0, 0, 0, 1, 0, 0, 0)
1	4	(0.4444, 0.4444, 0.1111)	(-12, -9, -9)	(0, 0, 0, 1, 0, 0, 0)
2	2	(0.5, 0.25, 0.25)	(-12, -9, -9)	(0, 0, 0, 1, 0, 0, 0)
2	3	(0.5455, 0.2727, 0.1818)	(-12, -9, -9)	(0, 0, 0, 1, 0, 0, 0)
2	4	(0.5714, 0.2857, 0.1429)	(-12, -9, -9)	(0, 0, 0, 1, 0, 0, 0)
1	2.8845	(0.4261, 0.4261, 0.1477)	(-12, -9, -9)	(0, 0, 0, 1, 0, 0, 0)
		(McRow)	(-10.8, -10.2, -9)	(0, 0, 0, 0.6, 0.4, 0, 0)

There are two classic approaches in the literature: (1) to solve the weighted sum approach of (MOO) with each weight vector and let the decision maker choose one of these different solutions; (2) to aggregate ambiguous pairwise judgements f_1/f_2 and f_1/f_3 to obtain a unique weight vector. Table 1 enumerates values of vector function f and optimal solutions for these two approaches. The geometric mean method is commonly employed in the analytic hierarch process to combine different pairwise judgements (Ramanathan and Ganesh (1994)). In our case, the geometric means of f_1/f_2 and f_1/f_3, 1 and 2.8845, which is also given at the tenth row of Table 1.

We use the convex hull of the nine weight vectors to construct the weight region $\mathcal{W}$ for (McRow). Table 1 also gives values of vector function f and optimal solution of (McRow). In the above example (McRow) produces a unique optimal solution mitigating the conflict between the first three cases and the next six cases. The objective function values of (McRow) are less extreme. This property to attempt to balance values of different objectives was discussed in Section 2.3. Note that the weights in rows 4-6 and 10 of Table 1 also has an alternative optimum solutions $(0, 0, 0, 0, 1, 0, 0)$, i.e., the optimum is not unique. While in this situation $(0, 0, 0, 0.6, 0.4, 0, 0)$ is also an optimum solution (through convex combination of extreme points), the model does not provide any guidance/reason for choosing this solution.

It is worth pointing out another issue about the sensitivity of pairwise judgements used in the analytic hierarchy approach on the aggregate weights calculated in Table 1. Instead of $f_1/f_2 = 1$, consider using $f_1/f_2 = 0.9$ and 1.1. The nonzero elements of the optimal solution of (McRow) remain at $(x_4, x_5) = (0.6, 0.4)$. Although the aggregation approach gives very similar weight vectors $(0.4197, 0.4347, 0.1455)$ and $(0.4319, 0.4184, 0.1497)$, the nonzero element of the corresponding optimal solutions jumps from $x_5 = 1$ to $x_4 = 1$. The (McRow) model is more robust in this aspect.

4 Reformulation with Conic Weight Region

Section 3 discussed (McRow) with a polyhedral weight set. In many practical situations the weights may be available as a simple or general confidence ellipsoids. This is the case, for example, in problems where the weights are obtained by questionnaire surveys where opinion from a large number of participants is desired. Such survey based decision making is discussed in Aretoulis et al. (2009) and Prato and Herath (2007) for problems in ecological and project management. An ellipsoidal region is a natural choice when considering confidence regions used to specify $\mathcal{W}$. To see that let $v^1, \ldots, v^q \in \mathcal{V}_f$ be q weights recommended by a survey and view $v^1, \ldots, v^q$ as independent and identically distributed samples from an unknown distribution of weights. Then, we can use a multivariate statistical method to build a $100(1-\alpha)\%$ confidence region $\{v \in \mathcal{V}_f \mid q(v-\bar{v})^T S^{-1}(v-\bar{v}) \leq \chi^2_{\alpha,m}\}$ for the expected weight, where $\bar{v}$ is the mean of $v^1, \ldots, v^q$, S is the covariance matrix of those vectors and $\chi^2_{\alpha,m}$ is the $100(1-\alpha)\%$ critical value of the chi-square distribution with m degrees of freedom (see Yang and Trewn (2004)). Note that such an approach also addresses the issue of outliers. Such ellipsoidal sets can be described as second order cones, which are a special case of general convex cones. Conic representations give a general approach to describe such sets and their parameterizations.

4.1 McRow with a Single Conic Weight Region

We consider a rather general case where the perturbation set $\mathcal{V}$ is given by a conic representation

$$\mathcal{V} = \{v \in \mathcal{V}_f \mid \exists u \in \mathbb{R}^l \text{ s.t. } Av + Bu + c \in \mathbf{K}\}, \tag{4.1}$$

where $\mathbf{K}$ is a closed convex cone in $\mathbb{R}^h$, $A_{h\times k}$ and $B_{h\times l}$ are given matrices, and $c \in \mathbb{R}^h$ is a fixed vector. The following theorem gives a reformulation of (McRow) under this setting.

Theorem 4.1 *Let $\mathcal{W}$ be denoted in* (3.2) *and the perturbation set $\mathcal{V}$ be given by* (4.1). *When $\mathbf{K}$ is not a polyhedral cone, assume that there exist $\hat{v} \in \mathcal{V}$ and $\hat{u} \in \mathbb{R}^l$ such that $\hat{w}^T + Q\hat{v} > 0$ and $A\hat{v} + B\hat{u} + c \in int(\mathbf{K})$. Then* (McRow) *is equivalent to*

$$\begin{aligned} \min_{x,y,s,\zeta} \quad & \hat{w}^T s + \zeta + c^T y \\ s.t. \quad & Q^T s + A^T y = 0, \\ & B^T y = 0, \\ & f(x) - s - \zeta e \leq 0, \\ & x \in \mathcal{X},\ y \in \mathbf{K}_*. \end{aligned} \tag{4.2}$$

where $\mathbf{K}_* = \{z \in \mathbb{R}^h : z^T y \geq 0,\ \forall y \in \mathbf{K}\}$ is the cone dual to $\mathbf{K}$.

Proof: Let us represent the condition $v \in \mathcal{V}_f$ in a conic formulation as $A'v + c' \in \mathbf{K}'$ where $A' = \begin{bmatrix} Q \\ 0_{1\times k} \end{bmatrix}$, $c' = \begin{bmatrix} \hat{w} \\ 1 \end{bmatrix}$, and $\mathbf{K}' = \{(v,\eta) \in \mathbb{R}^{k+1} \,|\, v^T e = \eta,\ v \geq 0\}$ (thus $\mathbf{K}'_* = \{(z,\zeta) \in \mathbb{R}^{m+1} : z \geq -\zeta e\}$). We rewrite the inner maximization problem of (McRow) as

$$\begin{aligned} \max_{v,u}\ & \hat{w}^T f(x) + v^T Q^T f(x) \\ \text{s.t. } & A'v + c' \in \mathbf{K}', \\ & Av + Bu + c \in \mathbf{K}. \end{aligned}$$

The corresponding dual problem is

$$\begin{aligned} \min_{y,z}\ & \hat{w}^T f(x) + {c'}^T z + c^T y \\ \text{s.z. } & {A'}^T z + A^T y = -Q^T f(x), \\ & B^T y = 0, \\ & z \in \mathbf{K}'_*,\ y \in \mathbf{K}_*. \end{aligned}$$

Letting $z = \begin{bmatrix} s - f(x) \\ \zeta \end{bmatrix}$, we further have

$$\begin{aligned} \min_{s,z,\zeta}\ & \hat{w}^T s + \zeta + c^T y \\ \text{s.t. } & Q^T s + A^T y = 0, \\ & B^T y = 0, \\ & f(x) - s - \zeta e \leq 0, \\ & y \in \mathbf{K}_*. \end{aligned}$$

Since the Slater condition holds, the above primal and dual problems have the same optimal value by the conic duality theorem (see A.2.1 in Ben-Tal et al. (2009)). Note that the Slater condition is unnecessary if $\mathbf{K}$ is a polyhedral cone. Finally, we can obtain (4.2) by replacing the inner maximization problem of (McRow) by its dual problem. □

Let us apply Theorem 4.1 to a special case that the perturbation set $\mathcal{V}$ is denoted in l_p-norm $(1 < p < \infty)$.

Corollary 4.2 *Let $\mathcal{W}$ be denoted in* (3.2) *and the perturbation set be $\mathcal{V} = \{v \in \mathcal{V}_f \,|\, \|v\|_p \leq 1\}$ $(1 < p < \infty)$, where $\|\cdot\|_p$ represents the l_p-norm of a vector. Assume that there exist $\hat{v} \in \mathcal{V}$ such that $\hat{w}^T + Q\hat{v} > 0$ and $\|\hat{v}\|_p < 1$. Then* (McRow) *is equivalent to*

$$\begin{aligned} \min_{x,s,\zeta}\ & \hat{w}^T s + \zeta + \|Q^T s\|_q \\ s.t.\ & f(x) - s - \zeta e \leq 0, \\ & x \in \mathcal{X}, \end{aligned} \tag{4.3}$$

where $q = \frac{p}{p-1}$ (or $\frac{1}{p} + \frac{1}{q} = 1$).

Proof: See Appendix C.3. □

Remark 4.3 *Corollary 4.2 is also applicable to the two cases that* $p = 1$ *and* ∞ *without the Slater condition. Readers may refer to* (3.9) *and* (3.10)*.*

We now consider a practical application of Corollary 4.2 where the weight region is given as an ellipsoid

$$\mathcal{V}_e(\bar{v}, S, \gamma) = \left\{ v \in \mathbb{R}^k \,\middle|\, (v - \bar{v})^T S^{-1} (v - \bar{v}) \leq \gamma^2 \right\}, \tag{4.4}$$

where $\gamma > 0$, $\bar{v} \in \mathcal{V}_f$, and S is a positive definite matrix, i.e., assume that the weight perturbations are given by an ellipsoid. It is easy to see that there exists an interior point in $(\mathcal{V}_f \cap \mathcal{V}_e(\bar{v}, S, \gamma))$. Since S is a positive definite matrix, there exists a $k \times k$ full rank matrix denoted as $S^{1/2}$ such that $S = \left(S^{1/2}\right)^T S^{1/2}$. Let $S^{-1/2} = \left(S^{1/2}\right)^{-1}$ in the later statements.

We now study the case where the perturbation set is denoted by $\mathcal{V} = \mathcal{V}_f \cap \mathcal{V}_e(\bar{v}, S, \gamma)$. Let $\tilde{v} = \gamma^{-1} S^{-1/2} (v - \bar{v})$, $\tilde{w} = \hat{w} + Q\bar{v}$, and $\tilde{Q} = rQS^{1/2}$. Then we rewrite the perturbation set $\mathcal{V}_e(\bar{v}, S, \gamma) \cap \mathcal{V}_f$ as

$$\widetilde{\mathcal{V}} = \{\tilde{v} \in \widetilde{\mathcal{V}}_f \mid \|\tilde{v}\|_2 \leq 1\},$$

where

$$\widetilde{\mathcal{V}}_f = \{\tilde{v} \in \mathbb{R}^k \mid \tilde{w} + \tilde{Q}\tilde{v} \geq 0, e^T(\tilde{w} + \tilde{Q}\tilde{v}) = 1\}.$$

It follows by Corollary 4.2 that (McRow) is equivalent to

$$\begin{aligned} \min_{x,s,\zeta} \ & \tilde{w}^T s + \zeta + \|\tilde{Q}^T s\|_2 \\ \text{s.t.} \ & f(x) - s - \zeta e \leq 0, \\ & x \in \mathcal{X}. \end{aligned}$$

Now substituting $\hat{w}$, Q, $\bar{v}$, S, and γ for $\tilde{w}$ and $\tilde{Q}$ we have

$$\begin{aligned} \min_{x,s,\zeta,\eta} \ & \hat{w}^T s + \zeta + \eta \\ \text{s.t.} \ & \gamma\|S^{1/2} Q^T s\|_2 + \bar{v}^T Q^T s - \eta \leq 0, \\ & f(x) - s - \zeta e \leq 0, \\ & x \in \mathcal{X}. \end{aligned} \tag{4.5}$$

The following corollary states that the robust value of (MOO) in this setting is a nondecreasing concave function of γ for fixed $\bar{v}$ and S.

Corollary 4.4 *Let* $\mathcal{V}_f$ *be denoted in* (3.3)*. For a fixed* $\bar{v} \in \mathcal{V}_f$ *and* $S \succ 0$*, let* $\mathcal{W}(\gamma) = \{w \in \mathbb{R}^m \mid w = \hat{w} + Qv,\ v \in \mathcal{V}_f \cap \mathcal{V}_e(\bar{v}, S, \gamma)\}$*, where* $\mathcal{V}_e(\bar{v}, S, \gamma)$ *is defined in* (4.4)*. The robust value* $M(\mathcal{W}(\gamma))$ *is a nondecreasing concave function of* γ*.*

Proof: See Appendix C.4. □

4.2 Example: Using McRow with Ellipsoidal Weight Region

We will continue with the example in Section 3.1. Assume that the weights given in Table 1 are calculated from a survey, and instead of taking their convex hull an ellipsoidal weight region is generated as follows. Assume that these nine weight vectors are from a survey sampling method, and our problem is to find risk averse solution at $(1-\alpha)\%$ level of confidence. The sample mean and sample covariance matrix of these vectors are

$$\bar{w} = (0.4204, 0.4298, 0.1518), \quad \Sigma = \begin{bmatrix} 0.0114 & -0.0144 & 0.0030 \\ -0.0144 & 0.0203 & -0.0059 \\ 0.0030 & -0.0059 & 0.0029 \end{bmatrix}.$$

After using the fact that the weights sum to one, and using the reference weight vector and coefficient matrix as $\hat{w} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$, $Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}$ the $(1-\alpha)\%$ confidence region can be described by taking a 2-dimensional perturbation subset of $\mathcal{V}_f := \{v \in \mathbb{R}^2_+ \,|\, e^T v \leq 1\}$ using $\bar{v} = (0.4204, 0.4298)$, $S = \begin{bmatrix} 0.0114 & -0.0144 \\ -0.0144 & 0.0203 \end{bmatrix}$, in $\mathcal{V}_e\left(\bar{v}, S, \sqrt{\chi^2_{\alpha,2}\,/\,9}\right)$.

Table 2: (McRow) with Ellipsoidal confidence Weight Region

α	Value of f	Optimal Solution
0.10	(-10.7076, -10.2924, -9)	(0, 0, 0, 0.5692, 0.4308, 0, 0)
0.05	(-10.7093, -10.2907, -9)	(0, 0, 0, 0.5698, 0.4302, 0, 0)
0.01	(-10.7115, -10.2885, -9)	(0, 0, 0, 0.5705, 0.4295, 0, 0)
0.005	(-10.7121, -10.2879, -9)	(0, 0, 0, 0.5707, 0.4293, 0, 0)

Table 2 shows the remarkable stability of results for different values of α. Moreover, these solutions are consistent with the McRow solution produced by the polyhedral weight region used in Section 3.1. We remark that the solution stability is due to the fact that the confidence region calculated in this example (see the magnitude of covariance matrix Σ) is small. This stability of solution in the presence of small data errors is a desired property of a good model.

4.3 Multiple Conic Weight Regions

Let cone $\mathbf{K}$ be the direct product of simpler cones $\mathbf{K}^1, \ldots, \mathbf{K}^r$. We write

$$\mathcal{V} = \{v \in \mathcal{V}_f \,|\, \exists u^1, \ldots, u^r \in \mathbb{R}^l \text{ s.t. } A^i v + B^i u^i + c^i \in \mathbf{K}^i, i = 1, \ldots, r\}. \tag{4.6}$$

The following corollary shows the tractable reformulation in this case.

Corollary 4.5 *Let $\mathcal{W}$ be denoted in* (3.2) *and the perturbation set $\mathcal{V}$ be given by* (4.6). *Suppose that there exist $\hat{v} \in \mathcal{V}$ and $\hat{u}^1, \ldots, \hat{u}^r \in \mathbb{R}^l$ such that $\hat{w}^T + Q\hat{v} > 0$ and $A^i\hat{v} + B^i\hat{u}^i + c \in int(\mathbf{K}^i)$ for*

$i = 1, \ldots, r$. *Then* (McRow) *is equivalent to*

$$
\begin{aligned}
\min_{x,s,\zeta,y^1,\ldots,y^r} \quad & \hat{w}^T s + \zeta + \sum_{i=1}^{r} {c^i}^T y^i \\
s.t. \quad & {B^i}^T y^i = 0, \qquad i = 1, \ldots, r, \\
& Q^T s + \sum_{i=1}^{r} {A^i}^T y^i = 0, \\
& f(x) - s - \zeta e \leq 0, \\
& y^i \in \mathbf{K}_*^i, \qquad i = 1, \ldots, r, \\
& x \in \mathcal{X},
\end{aligned}
\tag{4.7}
$$

where $\mathbf{K}_*^i$ *is the cone dual to* $\mathbf{K}^i$.

Proof: Refer to the proof of Theorem 4.1. □

Let cone $\mathbf{K}$ be the union of $\mathbf{K}^1, \ldots, \mathbf{K}^r$. We write

$$
\mathcal{V} = \{v \in \mathcal{V}_f \,|\, \exists i \in \{1, \ldots, r\}, u \in \mathbb{R}^l \text{ s.t. } A^i v + B^i u + c^i \in \mathbf{K}^i\}. \tag{4.8}
$$

This situation may arise if decision makers have two or more very different views on the importance of objectives and they cluster around different regions described as separate cones. The following theorem gives a reformulation of the resulting (McRow) model. A proof using constructions similar to that in the proof of Theorem 4.1 is given in Appendix C.5.

Theorem 4.6 *Let* $\mathcal{W}$ *be denoted in* (3.2) *and the perturbation set* $\mathcal{V}$ *be given by* (4.8)*. Suppose that there exist* $\hat{v} \in \mathcal{V}$ *and* $\hat{u} \in \mathbb{R}^l$ *such that* $\hat{w}^T + Q\hat{v} > 0$ *and* $A^i\hat{v} + B^i\hat{u}^i + c \in int(\mathbf{K}^i)$ *for some* $i \in \{1, \ldots, r\}$. *Then* (McRow) *is equivalent to*

$$
\begin{aligned}
\min_{x,s,\zeta,\eta,y^1,\ldots,y^r} \quad & \hat{w}^T s + \zeta + \eta \\
s.t. \quad & Q^T s + {A^i}^T y^i = 0, \quad i = 1, \ldots, r, \\
& {B^i}^T y^i = 0, \quad i = 1, \ldots, r, \\
& {c^i}^T y^i - \eta \leq 0, \quad i = 1, \ldots, r, \\
& f(x) - s - \zeta e \leq 0, \\
& y^i \in \mathbf{K}_*^i, \quad i = 1, \ldots, r, \\
& x \in \mathcal{X}.
\end{aligned}
\tag{4.9}
$$

5 Multi-Objective Stochastic Optimization

In many practical situations objective functions of (MOO) and/or relative importance of the objectives are stochastic. Several application papers where such models have arisen were cited in the introduction. We now consider such problems. Let $F(x, \xi) = (F_1(x, \xi), \ldots, F_m(x, \xi))^T : \mathcal{X} \times \Xi \to \mathbb{R}^m$ be a vector-valued function. Here Ξ is the support of the random parameter ξ. A common approach (see Abdelaziz et al. (2007)) to stochastic multi-objective optimization problem is to take expectation of the stochastic objective functions

$$\min_{x \in \mathcal{X}} \; \{f(x) = \mathbb{E}[F(x, \xi)]\}. \tag{5.1}$$

In this case we can still use (WSO) if deterministic weights are unrelated to ξ, to minimize the weighted sum expected-value objectives. We now model the situation where decision makers choose weights dependent of the random scenario. We need to consider a stochastic version of (WSO).

Definition 5.1 *Denote Ω as a state set consisting of various conditions and situations which determine decision makers' choice of weights. Let $\Lambda \in (\Omega, \mathcal{F}, \mathcal{W}_f, \mu)$ be random vector-valued mapping from the state space Ω to the weight region $\mathcal{W}_f$. An optimal solution of*

$$\min_{x \in \mathcal{X}} \mathbb{E}\left[\Lambda^T F(x, \xi)\right] \tag{SWSO}$$

is called a stochastic Pareto optimal solution.

Conceptually the stochastic Pareto optimality builds a stochastic optimization problem minimizing the mean of the stochastic weighted sum of the objectives. The concept of stochastic Pareto optimality extends deterministic weights used in classical multi-objective techniques to a random vector expressing uncertainty of criteria weights. When the random weight vector Λ is independent of the random scenario ξ, (SWSO) is simplified as

$$\min_{x \in \mathcal{X}} \mathbb{E}\left[\Lambda^T\right] f(x). \tag{MWSO}$$

(MWSO) can be regarded as a special case of (WSO). It is possible that Λ and ξ are correlated, since decision makers may give different importance to objective functions depending on their realized values. If Λ and ξ are totally correlated, then we can regard the weight vector as a vector-valued mapping from the support of the random scenario to a weight region, $\Lambda(\cdot) : \Xi \to \mathcal{W}_f$. In this case we may rewrite (SWSO) as

$$\min_{x \in \mathcal{X}} \mathbb{E}\left[\Lambda(\xi)^T F(x, \xi)\right], \tag{FWSO}$$

which is a stochastic optimization problem (see Ruszczyński and Shapiro (2003) and references therein). (FWSO) provides a flexible framework allowing us to consider different weights related to random scenario. We now consider the robust versions of stochastic (MOO). The definition of weight region $\mathcal{W}$ in (McRow) ignores the fact that, in some cases, decision makers are more inclined to specific weight combinations, by implicitly assuming that the evaluations of all decision makers are treated equally. In many practical settings, though, such a condition is not satisfied; a hierarchical group decision is a typical example, since the decisions of the group leaders are often emphasized. In our stochastic construction we allow to put a measure μ to represents the influence of each decision maker.

5.1 Robust Expected-Value Weighted Sum Model

We first study a robust version of (MWSO). We use the ideas of distributional robustness from Delage and Ye (2010) to describe our uncertainty set. First consider a moment robust (McRow) as follows:

$$\min_{x\in\mathcal{X}}\max_{\mu\in\mathcal{D}}\mathbb{E}_\mu\left[\Lambda^T\right]f(x). \tag{Mr-McRow}$$

Here $\mathcal{D}$ is a set of possible probability distributions of the random weight vector Λ. Denote a vector functional on a measure space

$$h(\mu)=\mathbb{E}_\mu[\Lambda]. \tag{5.2}$$

Then we write (Mr-McRow) as

$$\min_{x\in\mathcal{X}}\max_{w\in h(\mathcal{D})} w^T f(x). \tag{5.3}$$

The optimal solution x^* of (Mr-McRow) is a robust Pareto optimal solution of (MOO) with respect to the weight region $h(\mathcal{D})$.

We now study two special cases, which in the literature are well-known optimization problems under moment uncertainty (see Delage and Ye (2010) and references therein). Denote the weight region $\mathcal{W}$ by the perturbation set $\mathcal{V}$ as (3.2). Let $V:\Omega\to\mathcal{V}_f$ be random perturbation vector and thus $\Lambda=\hat{w}+QV$.

Weight Region with the First Moment Uncertainty

Denote $\bar{v}$ as the sample mean of q i.i.d. samples of weights and S as their sample covariance matrix. It is obvious that $\bar{v}\in\mathcal{V}_f$. Now assume that S is a positive definite matrix. We have mentioned that the i.i.d. samples may be obtained by questionnaire surveys in Section 4. Denote

$$\mathcal{D}_1(\mathcal{V}_f,\bar{v},S,\gamma)=\left\{\mu\in\mathcal{M}\;\middle|\;\begin{array}{l}\mu(V\in\mathcal{V}_f)=1\\ (\mathbb{E}_\mu[V]-\bar{v})^T S^{-1}(\mathbb{E}_\mu[V]-\bar{v})\le\gamma^2\end{array}\right\},$$

where $\mathcal{M}$ is the set of all probability measures on the space $(\Omega,\mathcal{F})$. The second condition in the set $\mathcal{D}_1(\mathcal{V}_f,\bar{v},S,\gamma)$ is a confidence region of the mean of V. Let $\gamma^2=\chi^2_{\alpha,m}/q$ where $\chi^2_{\alpha,m}$ is $100(1-\alpha)\%$ critical value of the chi-square distribution with m degrees of freedom, as shown in Section 4.2. Let us consider an optimization problem

$$\min_{x\in\mathcal{X}}\max_{\mu\in\mathcal{D}_1(\mathcal{V}_f,\bar{v},S,\gamma)}(\hat{w}+Q\mathbb{E}_\mu[V])^T f(x). \tag{5.4}$$

The following theorem gives an equivalent reformulation of (5.4).

Theorem 5.2 *Given $\gamma\ge 0$ and $S\succ 0$,* (5.4) *is equivalent to*

$$\min_{x\in\mathcal{X}}\max_{v\in\mathcal{V}_f\cap\mathcal{V}_e(\bar{v},S,\gamma)}(\hat{w}+Qv)^T f(x), \tag{5.5}$$

which can be reformulated as (4.5).

Proof: Denote $v(\mu)=\mathbb{E}_\mu[V]$ and $\widehat{\mathcal{D}}_1=\{\mu\in\mathcal{M}\,|\,v(\mu)\in\mathcal{V}_f,(v(\mu)-\bar{v})^T S^{-1}(v(\mu)-\bar{v})\le\gamma^2\}$. For a given $x\in\mathcal{X}$, denote $\theta(x)$ as the optimal value of the inner maximization problem of (5.4). We now claim that $\theta(x)$ is equal to

$$\hat{\theta}(x)=\max_{\mu\in\widehat{\mathcal{D}}_1}(\hat{w}+Qv(\mu))^T f(x). \tag{5.6}$$

It is easy to see that $\widehat{\mathcal{D}}_1 \supseteq \mathcal{D}_1(\mathcal{V}_f, \bar{v}, S, \gamma)$. Hence, $\theta(x) \leq \hat{\theta}(x)$. Now let μ^* be an optimal solution of (5.6) and $v^* = \mathbb{E}_{\mu^*}[V]$. We then choose Dirac measure δ_{v^*} for which mass one lies at the point v^*. It follows that $\delta_{v^*} \in \mathcal{D}_1(\mathcal{V}_f, \bar{v}, S, \gamma)$ and then

$$\hat{\theta}(x) = (\hat{w} + Qv(\mu^*))^T f(x) = (\hat{w} + Qv(\delta_{v^*}))^T f(x) \leq \theta(x).$$

We can now complete the proof since (5.6) is equivalent to the inner maximization problem of (5.5) for any $x \in \mathcal{X}$. □

Weight with First and Second Moment Uncertainty

(Mr-McRow) under the first moment uncertainty only considers a confidence region of the mean of V. However, it does not construct a relationship of the sample covariance matrix S and the true covariance matrix of V. Delage and Ye (2010) give a constraint parameterized by $\kappa \geq 1$

$$\mathbb{E}_\mu[(V - \bar{v})(V - \bar{v})^T] \preceq \kappa S, \tag{5.7}$$

which requires the covariance matrix $\mathbb{E}[(V - \bar{v})(V - \bar{v})^T]$ to be in a positive semi-definite cone bounded by a matrix inequality. Using the idea proposed by Delage and Ye (2010), we represent the weight uncertainty by the following distribution set

$$\mathcal{D}_2(\mathcal{V}_f, \bar{v}, S, \gamma, \kappa) = \left\{ \mu \in \mathcal{M} \;\middle|\; \begin{array}{l} \mu(V \in \mathcal{V}_f) = 1 \\ (\mathbb{E}_\mu[V] - \bar{v})^T S^{-1} (\mathbb{E}_\mu[V] - \bar{v}) \leq \gamma^2 \\ \mathbb{E}_\mu[(V - \bar{v})(V - \bar{v})^T] \preceq \kappa S \end{array} \right\}.$$

Note that (Mr-McRow) under the first moment uncertainty is only a special case by choosing $\mathcal{D}_2(\mathcal{V}_f, \bar{v}, S, \gamma, \infty)$. Delage and Ye (2010) further discuss the relationship of sample size q and parameters $\bar{u}$, S, γ, and κ in $\mathcal{D}_2(\mathcal{V}_f, \bar{v}, S, \gamma, \kappa)$, which provides a probabilistic guarantee for the true distribution of the weight vector lying in $\mathcal{D}_2(\mathcal{V}_f, \bar{v}, S, \gamma, \kappa)$.

Let us now consider the following optimization problem

$$\min_{x \in \mathcal{X}} \max_{\mu \in \mathcal{D}_2(\mathcal{V}_f, \bar{v}, S, \gamma, \kappa)} (\hat{w} + Q\mathbb{E}_\mu[V])^T f(x). \tag{5.8}$$

We adapt Lemma 1 in Delage and Ye (2010) to our case in the following theorem which represents (5.8) as a tractable reformulation. For completeness we also adapt the proof given by Delage and Ye (2010) to this case in Appendix C.6.

Theorem 5.3 *Given $\gamma \geq 0$, $\kappa \geq 1$, and $S \succ 0$, (5.8) is equivalent to*

$$\begin{aligned} \min_{x, y, \zeta, \eta, Y} \; & \zeta + \eta \\ s.t. \; & \zeta \geq (\hat{w} + Qv)^T f(x) - v^T y - v^T Y v, \quad \forall v \in \mathcal{V}_f, \\ & \eta \geq (\kappa S + \bar{v}\bar{v}^T) \bullet Y + \bar{v}^T y + \gamma \| S^{1/2}(2Y\bar{v} + y) \|_2, \\ & x \in \mathcal{X}, \; Y \succeq 0. \end{aligned} \tag{5.9}$$

Proof: See Appendix C.6. □

5.2 Stochastically Robust Weighted Sum Model

Let $\mathcal{G}$ to be the collection of all closed convex subsets of $\mathcal{W}_f$ and $\mathcal{W}(\cdot) : \Xi \to \mathcal{G}$ be a set-valued mapping. A robust version of general (SWSO) is proposed as a stochastically robust (McRow):

$$\min_{x\in\mathcal{X}} \mathbb{E}\left[\max_{w\in\mathcal{W}(\xi)} w^T F(x,\xi)\right]. \tag{Sr-McRow}$$

We will call a solution of (Sr-McRow) is a robust stochastic Pareto optimal solution. (Sr-McRow) introduces a robust concept to hedge the risk resulting from uncertain weight regions that are scenario dependent. We can also view (Sr-McRow) as an extension of (FWSO) where decision makers provide a weight region for each scenario instead of a weight vector. We first discuss an alternative model that also leads to (Sr-McRow). Rewrite (SWSO) as

$$\min_{x\in\mathcal{X}} \mathbb{E}\left[\mathbb{E}\left[\Lambda^T \,|\, \xi\right] F(x,\xi)\right].$$

Let $\mathcal{D}(\xi)$ be a set of possible conditional distribution of the random weight vector Λ for a given ξ. A corresponding robust version is

$$\min_{x\in\mathcal{X}} \mathbb{E}\left[\max_{\mu\in\mathcal{D}(\xi)} \mathbb{E}_\mu\left[\Lambda^T\right] F(x,\xi)\right].$$

Analogous to (5.3), we represent the above problem as (Sr-McRow)

$$\min_{x\in\mathcal{X}} \mathbb{E}\left[\max_{w\in h(\mathcal{D}(\xi))} w^T F(x,\xi)\right]. \tag{5.10}$$

We now give reformulations of (Sr-McRow) for a specific description of the weight region $\mathcal{W}(\xi)$. Let us represent $\mathcal{W}(\xi)$ for each $\xi \in \Xi$ by

$$\mathcal{W}(\xi) = \{w \in \mathbb{R}^m \,|\, w = \hat{w}(\xi) + Q(\xi)v, v \in \mathcal{V}(\xi)\}. \tag{5.11}$$

Similarly, denote the perturbation region by

$$\mathcal{V}_f(\xi) = \{v \in \mathbb{R}^k \,|\, \hat{w}(\xi) + Q(\xi)v \geq 0,\ e^T(\hat{w}(\xi) + Q(\xi)v) = 1\}, \tag{5.12}$$

$$\mathcal{V}(\xi) = \{v \in \mathcal{V}_f(\xi) \,|\, \exists u \in \mathbb{R}^l \text{ s.t. } A(\xi)v + B(\xi)u + c(\xi) \in \mathbf{K}(\xi)\}. \tag{5.13}$$

Assume that for each $\xi \in \Xi$ there exist $v(\xi)$ and $u(\xi)$ such that $\hat{w}(\xi) + Q(\xi)v(\xi) > 0$, $e^T(\hat{w}(\xi) + Q(\xi)v(\xi)) = 1$, and $A(\xi)v(\xi) + B(\xi)u(\xi) + c(\xi) \in \text{int}(\mathbf{K}(\xi))$. Thus it follows from Theorem 4.1 that (Sr-McRow) can be represented as

$$\min_{x\in\mathcal{X}} \mathbb{E}[Z(x,\xi)], \tag{5.14}$$

where $Z(x,\xi)$ is given as follows:

$$
\begin{aligned}
Z(x,\xi) := \min_{y,s,\zeta}\ & \hat{w}(\xi)^T s + \zeta + c(\xi)^T y \\
\text{s.t. } & Q(\xi)^T s + A(\xi)^T y = 0, \\
& B(\xi)^T y = 0, \\
& F(x,\xi) - s - \zeta e \le 0, \\
& y \in \mathbf{K}_*(\xi),
\end{aligned}
\tag{5.15}
$$

where $\mathbf{K}_*(\xi)$ is the dual cone of $\mathbf{K}(\xi)$. One may compute the expectation in (5.14) as a sample average approximation. Let Ξ represent as a finite set $\{\xi^1,\dots,\xi^r\}$ with respective probabilities $p_1,\dots,p_r$. Then one may reformulate (5.14) as a large-scale optimization problem as follows

$$
\begin{aligned}
\min_{x,y^1,\dots,y^r s^1,\dots,s^r,\zeta^1,\dots,\zeta^r}\ & \sum_{i=1}^{r} p_i \left(\hat{w}(\xi^i)^T s^i + \zeta^i + c(\xi^i)^T y^i\right) \\
\text{s.t. } & Q(\xi^i)^T s^i + A(\xi^i)^T y^i = 0, \quad i = 1,\dots,r, \\
& B(\xi^i)^T y^i = 0, \quad i = 1,\dots,r, \\
& F(x,\xi^i) - s^i - \zeta^i e \le 0, \quad i = 1,\dots,r, \\
& y^i \in \mathbf{K}_*(\xi^i), \quad i = 1,\dots,r, \\
& x \in \mathcal{X}.
\end{aligned}
\tag{5.16}
$$

The above construction also shows that (Sr-McRow) has the structure of a two-stage stochastic conic program. For each scenario, decision makers suggest different weight regions. The second stage problem considers a maximum of the weighted sum of the objectives over these weight regions. These problems can be solved by an extensive formulation of the model, or using Benders decomposition type algorithm (see Mehrotra and Özevin (2010) and references therein).

5.3 Example: Using Sr-McRow with Weights Related to Random Scenario

We modify the deterministic problem (3.11) in Section 3.1 to illustrate the implications of stochastic weights in multi-objective decision making. Consider three scenarios ξ = -1, 0, and 1 with equal probability $p_1 = p_2 = p_3 = 1/3$. We modify objective functions in (3.11) as

$$
\begin{array}{lrrrrrrr}
F_1(x,\xi) = & & -11x_2 & -11x_3 & -(12-\xi)x_4 & -9x_5 & -9x_6 & +9x_7 \\
F_2(x,\xi) = & -11x_1 & & -11x_3 & -(9+4\xi)x_4 & -(12+\xi)x_5 & -(9-4\xi)x_6 & +9x_7 \\
F_3(x,\xi) = & -11x_1 & -11x_2 & & -9x_4 & -(9-4\xi)x_5 & -(12-\xi)x_6 & -12x_7
\end{array}
$$

Note that the objectives $f_j(x)$ (j = 1, 2, and 3) in (3.11) is equal to $\mathbb{E}[F_j(x,\xi)]$.

We first keep the weight region denoted in Section 3.1 as the convex hull of the first nine ambiguous weight vectors in Table 1. Note that even when using a fixed weight region (unrelated to random objective function scenarios) (Sr-McRow) is different from a model that will first take the expectation of random functions and convert the stochastic multi-objective optimization to a deterministic problem (McRow). This is because the order of expectation and maximization cannot be interchanged in general. In our example the (Sr-McRow) optimal solution is (0, 0, 0, 0.2517, 0.2943, 0.4540, 0) and the objective value of f is $(-9.75402, -9.88182, -10.3612)$. Compared to the results given by (McRow) (see Table 1), the decision variable x_6 in this solution now gives more contribution because of its variability. This suggests that the suggested solution is more robust with respect to variability in this variable.

In order to understand the affect of random weights on solutions we randomize pairwise judgements of f_1 and f_2 as $f_1/f_2 = 2^{\xi}$ and keep those of f_1 and f_3 as $f_1/f_3 = 2$, 3, and 4 as before. Thus, in this case, those nine weight vector are divided as three groups by different ratios of f_1/f_2, each of which responds to a scenario. We now use the convex hull of each group as set-valued weight mapping $\mathcal{W}(\xi)$ of the random parameter ξ. Correspondingly, we obtain an optimal solution as (0, 0, 0, 0, 0.6692, 0.3308, 0) and the value of f as (-8.9991, -11.0064, -9.9915). This suggests that the dependence of a decision makers stochastic weight to that of random parameters in the objective functions can not be ignored.

6 Example Applications of the Sr-McRow Framework

In this section we illustrate the potential use of (Sr-McRow) using two application examples. The first example is a multi-criteria group decision problem from a disaster planning application. All group members give different criteria weights and attributes, which can be interpreted as a scenario in (Sr-McRow). The aggregation weights used to produce the final objective function are the probability associated with each scenario. Used in this way, the (Sr-McRow) framework generalizes the concept of additive social welfare function considered by Harsanyi (1955) and Keeney (1976) and the weighted arithmetic mean method in Ramanathan and Ganesh (1994).

In the second example, we consider the use of (Sr-McRow) to a yield management problem. The goal here is to find sales or production volumes of multiple commodities competing for the same limited resource that maximize total revenue or profit. In the multi-criteria stochastic optimization model the competing commodities are regarded as criteria and the prices or demands are used as trade-off weights. The random prices and yield are due to uncertain marketing factors and environmental conditions.

6.1 Multi-Criteria Group Decision Making

Utomo et al. (2009) present a multi-criteria group decision problem for selecting a good plan to reduce the impact of mud volcano disaster in Sidoarjo, Indonesia. The input to the decision problem is given by five stakeholders expressing different views. These stakeholders are: community, government, engineer, sponsor, and Non-Governmental Organizations (NGO). They evaluate six alternative plans from technical, economic, and social/environmental perspectives. Table 3 from Utomo et al. (2009) gives the relative criteria weights and attributes provided by each decision maker. We now build a (Sr-McRow) model for this problem.

We use a consensus measure μ to represents the influence of each decision maker, i.e., the evaluation of decision maker k is weighted by μ_k. In the social welfare function in Harsanyi (1955) and Keeney (1976) the measure μ_k is called a "scaling constant". Let $m(=3)$ be number of criteria, $n(=6)$ be number of available plans, and $\ell(=5)$ be the number of decision makers. Let a_{ij}^k, the attribute of plan j under criterion i given by decision maker k, represent the attribute data in Table 3. The multi-criteria group decision problem is given as

$$\max_{(x_1,\ldots,x_n)\in\mathcal{X}} \sum_{k=1}^{\ell} \mu_k \min_{(w_1,\ldots,w_m)\in\mathcal{W}^k} \sum_{i=1}^{m}\sum_{j=1}^{n} w_i u_i^k(a_{ij}^k) x_j, \tag{6.1}$$

where $\mathcal{X} = \{x \in \{0,1\}^n \mid e^T x = 1\}$ is a decision region and $u_i^k(\cdot)$ is the utility function for criterion i given by decision maker k. To simply this example, we use linear utility functions, i.e., $u_i^k(a_{ij}^k) = a_{ij}^k$. We use the criteria weights given in Table 3 as reference weights and allow a perturbation around this weight vector $\bar{w}^k$ to represent the weight regions $\mathcal{W}^k$ for our numerical study as follows:

$$\mathcal{W}^k = \left\{ w \in \mathcal{W}_f \;\middle|\; \|w - \bar{w}^k\|_2 \le \gamma \right\}, \quad k = 1, \ldots, 5.$$

Table 3: Stakeholder Evaluation Based-on Their Own Preferences

Stakeholder	Criterion	Weight	P1	P2	P3	P4	P5	P6
	Tech	0.2316	0.0631	0.0055	0.0397	0.0460	0.0151	0.0623
Community	Econ	0.0719	0.0039	0.0375	0.0119	0.0041	0.0105	0.0042
	Socio-Enviro	0.6965	0.1022	0.0232	0.1056	0.3433	0.0912	0.0311
	Tech	0.3483	0.0949	0.0083	0.0597	0.0692	0.0227	0.0937
Government	Econ	0.1485	0.0080	0.0774	0.0245	0.0084	0.0216	0.0086
	Socio-Enviro	0.5032	0.0738	0.0167	0.0763	0.2481	0.0659	0.0225
	Tech	0.7093	0.1932	0.0168	0.1215	0.1409	0.0463	0.1907
Engineer	Econ	0.2141	0.0115	0.1116	0.0353	0.0121	0.0312	0.0125
	Socio-Enviro	0.0766	0.0112	0.0025	0.0116	0.0377	0.0100	0.0034
	Tech	0.0833	0.0227	0.0020	0.0143	0.0165	0.0054	0.0224
Sponsor	Econ	0.7235	0.0388	0.3771	0.1193	0.0409	0.1054	0.0421
	Socio-Enviro	0.1932	0.0283	0.0064	0.0293	0.0952	0.2530	0.0086
	Tech	0.1679	0.0457	0.0040	0.0288	0.0334	0.0110	0.0452
NGO	Econ	0.0807	0.0043	0.0421	0.0133	0.0046	0.0118	0.0047
	Socio-Enviro	0.7514	0.1102	0.0250	0.1139	0.3704	0.0984	0.0336

For a given μ and γ model (6.1) can be solved by (4.5) and (5.16). However, given the simplicity of our problem we can rank order all the plans by fixing x to the different possible choices. Table 4 gives results for different values of γ and the value of μ constructed to reflect different level of importance given to the sponsor. In the baseline case all stakeholders have equal importance (same value for each $\mu_k = 0.2, k = 1, \ldots 5$). We then parameterize μ_k by increasing sponsor's importance $\mu_4 = \pi \in [0,1]$, and taking $(1-\pi)/4$, for the other stakeholders $k = 1, 2, 3$, and 5.

First we note that in this example the rank order of different plans is rather stable for different choices of γ, consequently error in criteria weight will not lead to a significantly different decision recommendation. This suggests that the weight region parametric analysis in (McRow) and (Sr-McRow) may be useful to check for decision robustness for possible uncertainty in criteria weights. The results in Table 4 also highlight the role a sponsor may play in swaying the decision. Note that the second plan, gets an increasingly improved ranking with increasing π. In fact, the results show that the decision is eventually narrowed between plan P2 and P4 depending provided the sponsor gets the relative importance of 0.4 or greater, otherwise plan P4 is favored.

Table 4: Ranks of Policies

π	γ	P1	P2	P3	P4	P5	P6
0.2	0	3	4	2	1	5	6
0.2	0.2	3	5	2	1	4	6
0.2	0.4	3	5	2	1	4	6
0.4	0	5	2	3	1	4	6
0.4	0.2	5	2	3	1	4	6
0.4	0.4	4	2	3	1	5	6
0.6	0	5	1	4	2	3	6
0.6	0.2	5	1	3	2	4	6
0.6	0.4	5	1	3	2	4	6

6.2 Stochastic Revenue Management Planning

We now consider an aggregate crop planning problem to illustrate the use of (Sr-McRow) in revenue/yield management. We use data available for US Department of Agriculture for our purpose. In the basic problem farmers need to decide the proportion of planted areas of corn, soybean, and wheat in next year. Future crop yields and prices are two crucial factors in their decision. Crop yields are highly related to temperature, precipitation, and soil condition (see Thompson (1975); Reed and Riggins (1982); Kantanantha et al. (2010)), while crop prices depend on world market supplies and demands (see Hoffman et al. (2007)). We want the planting decisions to be hedged (risk averse) against the price and yield uncertainties.

Yearly crop yields and monthly prices are regarded as functions defined on an appropriately defined probability space. We express crop prices as set-valued functions because of their much larger variability and frequency of available data (monthly data) than that for the crop yields (yearly data). Our crop planning model is written as

$$\max_{x\in\mathcal{X}} \mathbb{E}\left[G(x,\xi) = \min_{w\in\mathcal{W}(\xi)} w^T(Y(\xi)\circ x)\right], \tag{6.2}$$

where " $\circ$ " means the Hadamard product of two vectors, the set $\mathcal{X} = \{x \in \mathbb{R}^3_+ \mid e^T x = 1\}$ represents the available proportions of planning corn, soybean and wheat, $Y(\cdot)$ is a 3-dimensional vector-valued yield function, and $\mathcal{W}(\cdot) \subseteq \mathbb{R}^3_+$ is the set-valued price function. Model (6.2) interprets the crop prices as weights. Note that, since these weights are physical dollars, they should not be normalized, i.e. it is unnecessary to require that $e^T w = 1$ for all $w \in \mathcal{W}(\xi)$ w.p.1.

Table 5: Yields and Prices of Corn, Soybean, and Wheat

	Yield			Mean of Price		Covariance Matrix of Price		
Corn	Soybean	Wheat	Corn	Soybean	Wheat			
						0.217508	0.296205	0.332016
157	51.5	56	3.830833	9.974167	5.120833	0.296205	0.462691	0.453347
						0.332016	0.453347	0.564091
						0.084908	0.013121	0.147413
174	46	56	3.745	10.0525	5.304167	0.013121	0.448019	0.066873
						0.147413	0.066873	0.327091
						0.245117	0.611158	0.195542
179	47	64	4.78	11.31083	8.019167	0.611158	1.881274	0.827926
						0.195542	0.827926	2.176424
						0.028341	0.095615	0.059626
175	43.5	55	3.390833	7.7425	5.759167	0.095615	1.079169	1.141169
						0.059626	1.141169	1.424474
						0.106908	0.058551	0.101322
163	48	67	2.280833	5.648333	4.033333	0.058551	0.075914	0.033589
						0.101322	0.033589	0.117839
						0.009958	0.018292	-0.00435
143	46.5	61	1.964167	5.95	3.359167	0.018292	0.139567	-0.02804
						-0.00435	-0.02804	0.009158
						0.092042	0.48665	0.047408
180	50	59	2.465	7.555	3.568333	0.48665	2.641875	0.2466
						0.047408	0.2466	0.042781
						0.007685	-0.01217	0.005896
164	37	65	2.2675	6.084167	3.445	-0.01217	0.295074	0.013179
						0.005896	0.013179	0.063808
						0.041374	0.089214	0.117603
135	43	49	2.130833	4.933333	3.405833	0.089214	0.249289	0.252106
						0.117603	0.252106	0.387774
						0.004389	0.002583	0.000775
152	45	61	1.893333	4.43	2.83	0.002583	0.05645	-0.0139
						0.000775	-0.0139	0.007283
						0.032724	0.03471	0.013824
151	44	57	1.859167	4.7325	2.568333	0.03471	0.051985	0.005004
						0.013824	0.005004	0.023814
						0.022058	0.024613	0.010204
140	42	60	1.885	4.5675	2.575833	0.024613	0.071769	0.033065
						0.010204	0.033065	0.021091
						0.072274	0.139104	0.073519
141	44	48	2.204167	5.925	2.903333	0.139104	0.283558	0.125158
						0.073519	0.125158	0.107172
						0.013608	0.042835	0.028977
129	43	61	2.599167	7.4025	3.7125	0.042835	0.395502	0.08166
						0.028977	0.08166	0.073002
						0.377547	0.189901	0.203347
136	40.5	38	3.551667	7.274167	4.766667	0.189901	0.139841	0.064114
						0.203347	0.064114	0.262122
						0.067602	0.100958	0.124065
113	39	49	2.5575	5.85	4.0925	0.100958	0.159333	0.189467
						0.124065	0.189467	0.257352
						0.081141	0.171488	-0.00754
156	45.5	56	2.409167	6.105	3.515833	0.171488	0.380892	-0.0294
						-0.00754	-0.0294	0.051058
						0.032139	0.053472	0.019178
130	43	44	2.216667	6.044167	3.205833	0.053472	0.147624	-0.00992
						0.019178	-0.00992	0.053924
						0.040291	0.028537	0.044394
149	43	54	2.285833	5.570833	3.3825	0.028537	0.038458	0.025481
						0.044394	0.025481	0.073035
						0.001991	0.003641	-0.00134
107	37.5	32	2.329167	5.599167	2.738333	0.003641	0.015808	-0.02164
						-0.00134	-0.02164	0.103747

Yearly local yields and monthly national prices of these corn, soyabean and wheat crops for the past twenty years were obtained from Quick Stats 2.0, an agricultural statistics database, provided by the National Agricultural Statistics Service (USDA (2011)). Table 5 gives the crop yield Y^i $(i = 1, \ldots, 20)$ as a number of bushels harvested per acre and gives the means $\bar{w}^i$ and covariance matrices S^i of crop price in dollars per bushel. In this example, the weight region $\mathcal{W}(\cdot)$ is constructed with the $100(1-\alpha)\%$ prediction ellipsoidal region of crop price. For simplicity we assume that the monthly prices are normally distributed. Thus, from Chew (1966), the prediction regions of prices are given as

$$\mathcal{W}^i = \left\{ w \in \mathbb{R}^3_+ \;\middle|\; (w - \bar{w}^i)^T {S^i}^{-1} (w - \bar{w}^i) \leq \frac{m(n-1)(n+1)}{n(n-m)} F(\alpha, m, n-m) \right\}, \quad i = 1, \ldots, 20,$$

where m (=3, three types of crops), n (=12, number of monthly price data for each year) and $F(\alpha, d_1, d_2)$ is the $100(1-\alpha)\%$ critical value of the F distribution with d_1 degrees of freedom numerator and d_2 degrees of freedom denominator are used to construct this prediction region.

Assume that the yearly data are i.i.d samples. By the following proposition, we solve model (6.2), using the Sample Average Approximation method as

$$\max_{x \in \mathcal{X}} \frac{1}{N} \sum_{i=1}^{N} \min_{w \in \mathcal{W}^i} w^T (Y^i \circ x). \tag{6.3}$$

Proposition 6.1 *Suppose $G(x, \cdot)$ to be an integrable function for each $x \in \mathcal{X}$. Let v^* and $\mathcal{S}^*$ be the optimal value and the set of optimal solutions of* (6.2)*. Let v_n and $\mathcal{S}_N$ be those of* (6.3)*. Then, $v_n \to v^*$ and $\mathbb{D}(\mathcal{S}_N, \mathcal{S}^*) = \max_{s \in \mathcal{S}_N} \min_{t \in \mathcal{S}^*} \|s - t\|_2 \to 0$ w.p.1 as $N \to \infty$.*

Proof: Let $g(x)$ be the expected value of $G(x, \xi)$ and $g_N(x)$ be the same average approximation of $g(x)$. Since $G(x, \xi)$ is concave w.p.1 and integrable for each $x \in \mathcal{X}$, it follows by Corollary 3 in Shapiro (2003) that g_N converges to g uniformly on $\mathcal{X}$ w.p.1 as N increases to ∞. Then the remainder of the proof follows from Proposition 6 in Shapiro (2003). □

Table 6: Proportion of Planted Areas of Corn, Soybean, and Wheat

α	Corn	Soybean	Wheat
1	1	0	0
0.2	0.6498	0.3502	0
0.15	0.5362	0.4406	0.0232
0.1	0.4175	0.4487	0.1338
Allocation in 2011	0.5641	0.4016	0.0343

We use (4.5) and (5.16) to solve model (6.3). Note that all ζ's in (4.5) and (5.16) should be set to 0 to solve (6.3), since as mentioned above the normalization of $\mathcal{W}$ is not needed in this example. Table 6 shows the proportion of planted areas in Illinois 2011 and the optimal solutions of model (6.3) for different values of α. Note that the weight regions shrink to the mean of price when α is equal to 1. A more robust risk averse solution results from a larger weight region as α decreases. The results suggest a greater diversification of planted crops in this case. Interestingly, the result for $\alpha = .15$, i.e., 85% prediction region is very close to the actual Illinois statewide plantation for 2011. Observe that if we use the mean of price ($\alpha = 1$) as weight, farmers in Illinois will only plant corn. This is obviously not the case. We point out that the above model is illustrative and ignores several additional decision parameters/variables such as demand, transportation costs, competing state producing similar commodities, and corp export potential. Nevertheless, it emphasizes the possible application of our (Sr-McRow) framework to revenue management.

7 Concluding Remarks

We have introduced a risk averse (minimax) multi-criteria robust optimization with weight set (McRow) framework for model based multi-expert multi-objective decision making. Using this framework we have defined the concepts of robust Pareto optimality, robust stochastic Pareto optimality and robust value of a multi-objective optimization problem. The robust value has an interesting interpretation of cost of disagreement (or dissatisfaction). The increasing concave property of the robust value function suggests that the (McRow) framework has a natural managerial interpretation of the cost of disagreement as a function of the magnitude of disagreement. We give reformulations of the (McRow) models as single mathematical programs for problems with polyhedral, conic, and uncertain weight sets. We also give models and their reformulations for problems with multiple stochastic objectives where the weights are potentially correlated with the objective functions. With the help of a textbook numerical example and we illustrate various properties of the (McRow) family of models. In particular, we find that the solutions recommended by (McRow) are stable with respect to changes in the weight set. We also illustrate the impact of correlation in the randomness of weights with the randomness in the objective function parameters. Finally, the use of proposed concepts and techniques are illustrated using a real-world multi-criteria group decision problem in project selection, and a revenue management application example from the crop planting decision making.

References

F. B. Abdelaziz, B. Aounib, and R. E. Fayedha. Multi-objective stochastic programming for portfolio selection. *European Journal of Operational Research*, 177(3):1811–1823, 2007.

E. Alba, B. Dorronsoro, F. Luna, A. J. Nebro, P. Bouvry, and L. Hogie. A cellular multi-objective genetic algorithm for optimal broadcasting strategy in metropolitan manets. *Computer Communications*, 30(4):685–697, 2007.

N. Amjady, J. Aghaei, and H. Shayanfar. Stochastic multiobjective market clearing of joint energy and reserves auctions ensuring power system security. *IEEE Transactions on Power Systems*, 24 (4):1841–1854, 2009.

G. N. Aretoulis, G. P. Kalfakakou, and F. Z. Striagka. Construction material supplier selection under multiple criteria. *Operational Research*, 2009. inprint.

A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

K. Borcherding, T. Eppel, and D. von Winterfeldt. Comparison of weighting judgements in multi-attribute utility measurement. *Management Science*, 37(12):1603–1619, 1991.

V. Bowman. On the relationship of the tchebycheff norm and the efficient frontier of multiple-criteria objectives. In H. Thiriez and S. Zionts, editors, *Multiple-Criteria Decision Making*, pages 76–85. Springer-Verlag, New York, 1976.

G. Canfora and L. Troiano. A model for opinion agreement and confidence in multi-expert multi-criteria decision making. *Mathware & Soft Computing*, 2:67–82, 2004.

V. Chankong and Y. Y. Haim. *Multiobjective Decision Making: Theory and Methodology*. Elsevier Science Publishing, 1983.

A. Chen, J. Kim, L. Seungjae, and K. Youngchan. Stochastic multi-objective models for network design problem. *Expert Systems with Applications: An International Journal*, 37(2):1608–1619, 2010.

V. Chew. Confidence, prediction, and tolerance regions for the multivariate normal distribution. *Journal of the American Statistical Association*, 61(315):605–617, 1966.

P. Côtéa, L. Parrotta, and R. Sabourinb. Multi-objective optimization of an ecological assembly model. *Ecological Informatics*, 2(1):23–31, 2007.

C. Cromvik, P. Lindroth, M. Patriksson, and A.-B. Strömberg. A new robustness index for multi-objective optimization based on a user perspective. Technical report, Department of Mathematical Sciences, Chalmers University of Ttechnology and University of Gothenburg., Gothenburg, Sweden, 2011.

J. Davis. Group decision and social interactions: A theory of social decision schemes. *Psychological Review*, 80:97–125, 1973.

J. Davis. Group decision making and quantitative judgments: a consensus model. In *Understanding Group Behaviour Consensual Action by Small Groups*, volume 1, chapter 3, pages 35–59. Lawrence Erlbaum, Mahwah, NJ, 1996.

K. Deb and H. Gupta. Searching for robust pareto-optimal solutions in multi-objective optimization. *Evolutionary Multi-Criterion Optimization*, pages 150–164, 2005.

E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 2010. DOI: 10.1287/opre.1090.0741.

R. Dyer and E. Forman. Group decision support with the analytic hierarchy process. *Decision Support Systems*, 8:99–124, 1992.

W. Edwards. How to use multiattribute utility measurement for social decision making. *IEEE Transactions on Systems Man and Cybernetics*, 7:326–340, 1977.

M. Ehrgott. *Multicriteria Optimization*. Springer, 2005.

S. Gass and T. Saaty. The computational algorithm for the parametric objective function. *Naval Research Logistics Quarterly*, 2:39–45, 1955.

A. Geoffrion. Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications*, 22:618–630, 1968.

J. Harsanyi. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63:309–321, 1955.

S. He, J. Chen, W. Xu, Y. Sun, T. Preetha, and X. Shen. A stochastic multiobjective optimization framework for wireless sensor networks. *EURASIP Journal on Wireless Communications and Networking*, 2010:1–10, 2010.

L. A. Hoffman, S. H. Irwin, and J. Toasa. Forecast performance of futures price models for corn, soybeans, and wheat. Number 9889 in 2007 Annual Meeting, Portland, Oregon TN, July 29-August 1 2007. American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association).

J. Hu, T. Homem-de-Mello, and S. Mehrotra. Risk adjusted budget allocation models with application in homeland security. To appear in IIE Transactions, 2011.

A. Kandil, K. El-Rayes, and O. El-Anwar. Optimization research: Enhancing the robustness of large-scale multiobjective optimization in construction. *Journal of Construction Engineering and Management*, 136(1):17–25, 2010.

N. Kantanantha, N. Serban, and P. Griffin. Yield and price forecasting for stochastic crop decision planning. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(3):362–380, 2010.

R. Keeney. A group preference axiomatization with cardinal utility. *Management Science*, 23(2): 140–145, 1976.

R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: preferences and value tradeoffs*. John Wiley & Sons, New York, 1976.

H. Komiya. Elementary proof for sion's minimax theorem. *Kodai Mathematical Journal*, 11(1):5–7, 1988.

P. Korhonen, S. Salo, and R. Steure. A heuristic for estimating nadir criterion values in multiple objective linear programming. *Operations Research*, 45(5):751–757, 1997.

P. Liu, E. N. Pistikopoulos, and Z. Li. A multi-objective optimization approach to polygeneration energy systems design. *AIChE Journal*, 56(5):1218–1234, 2010.

Z. Lounis and D. J. Vanier. A multiobjective and stochastic system for building maintenance management. *computer-aided civil and infrastructure engineering*, 15:320–329, 2000.

O. D. Marcenaro-Gutierreza, M. Luqueb, and F. Ruizb. An application of multiobjective programming to the study of workers satisfaction in the spanish labour market. *European Journal of Operational Research*, 203(2):430–443, 2010.

N. F. Matsatsinis and A. P. Samaras. Mcda and preference disaggregation in group decision support systems. *European Journal of Operational Research*, 130:414–429, 2001.

S. Mehrotra and K. Kim. Outcome based state budget allocation for diabetes prevention programs. *Technical Report*, 2011.

S. Mehrotra and G. Özevin. Convergence of a weighted barrier decomposition algorithm for two-stage stochastic programming with discrete support. *SIAM Journal on Optimization*, 20(5): 2474–2486, 2010.

K. M. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, 1999.

M. Pöyhönen and R. P. Hämäläinen. Theory and methodology on the convergence of multiattribute weighting methods. *European Journal of Operational Research*, 129:569–585, 2001.

T. Prato and G. Herath. Multiple-criteria decision analysis for integrated catchment management. *Ecological Economics*, 63(2-3):627–632, 2007.

R. Ramanathan and L. S. Ganesh. Group preference aggregation methods employed in ahp: An evaluation and an intrinsic process for deriving members' weightages. *European Journal of Operational Research*, 79(2):249–265, 1994.

M. R. Reed and S. K. Riggins. Corn yield response: A micro-analysis. *North Central Journal of Agricultural Economics*, 4(2):91–94, 1982.

R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1972.

A. Ruszczyński and A. Shapiro, editors. *Handbooks in Operations Research and Management Science: Stochastic Programming*. Elsevier Science Publishing, 2003.

T. L. Saaty. *The analytc hierarchy process*. McGraw-Hill, 1980.

P. J. H. Schoemaker and C. C. Waid. An experimental comparison of different approaches to determining weights in additive utility models. *Management Science*, 28(2):182–196, 1982.

A. Shapiro. On duality theory of conic linear problems. In M. A. Goberna and M. A. López, editors, *Semi-Infinite Programming: Recent Advances*, pages 135–165. Kluwer Academic Publishers, 2001.

A. Shapiro. Monte Carlo sampling methods. In A. Ruszczynski and A. Shapiro., editors, *Handbook of Stochastic Optimization*. Elsevier Science Publishers B.V., Amsterdam, Netherlands, 2003.

M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8:171–176, 1958.

G. Soares, R. L. S. Adriano, C. Maia, L. Jaulin, and J. A. Vasconcelos. Robust multi-objective team22 problem: A case study of uncertainties in design optimization. *IEEE Transactions on Magnetics*, 45(3):1028–1031, 2009.

A. Stepanov and J. M. Smith. Multi-objective evacuation routing in transportation networks. *European Journal of Operational Research*, 198(2):435–446, 2009.

R. Steuer. *Multiple Criteria Optimization: Theory, Computation, and Application*. John Wiley & Sons, 1986.

L. M. Thompson. Weather variability, climatic change, and grain production. *Science*, 188:535–541, 1975.

V. Tsiporkovaa, E.and Boeva. Multi-step ranking of alternatives in a multi-criteria and multi-expert decision making environment. *Information Sciences*, 176:2673–2697, 2006.

USDA. *Quick Stats 2.0.* National Agricultural Statistics Service, July 2011. URL `http://www.nass.usda.gov/Data_and_Statistics/Quick_Stats/index.asp`.

C. Utomo, A. Idrus, M. Napiah, and M. F. Khamidi. Agreement options on multi criteria group decision and negotiation. *World Academy of Science, Engineering and Technology*, 50, 2009.

D. von Winterfeldt and W. Edwards. *Decision Analysis and Behavioral Research.* Cambridge University Press, 1986.

M. Weber and K. Borcherding. Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research*, 67:1–12, 1993.

K. Yang and J. Trewn. *Multivariate statistical methods in quality management.* McGraw-Hill, 2004.

P. L. Yu and G. Leitmann. Compromise solutions, domination structures, and salukvadze's solution. *Journal of Optimization Theory and Applications*, 13(3):362–378, 1974.

L. Zadeh. Optimality and non-scalar-valued performance criteria. *IEEE Transactions on Automatic Control*, 8:59–60, 1963.

Q. Zhang, S. Maeda, and T. Kawachi. Stochastic multiobjective optimization model for allocating irrigation water to paddy fields. *Paddy and Water Environment*, 5(2):93–99, 2007.

A Pareto Optimality

Pareto optimality (efficiency, nondominance) defines an order to compare with the objectives. In the following we introduce three widely-used concepts of Pareto optimality (see Miettinen (1999)).

Definition A.1 (Pareto Optimality) *A point $x^* \in \mathcal{X}$ is called Pareto optimal if there exists no point $x \in \mathcal{X}$ such that $f_i(x) \leq f_i(x^*)$ for all $i = 1, \ldots, m$, and $f_j(x) < f_j(x^*)$ for at least one index $j \in \{1, \ldots, m\}$.*

Definition A.2 (Weak Pareto Optimality) *A point $x^* \in \mathcal{X}$ is called weakly Pareto optimal if there exists no point $x \in \mathcal{X}$ such that $f_i(x) < f_i(x^*)$ for all $i = 1, \ldots, m$.*

Definition A.3 (Proper Pareto Optimality (Geoffrion (1968))) *A point $x^* \in \mathcal{X}$ is called properly Pareto optimal if it is Pareto optimal and if there is a real number $M > 0$ such that for all $i \in \{1, \ldots, m\}$ and $x \in \mathcal{X}$ satisfying $f_i(x) < f_i(x^*)$ there exists at least an index $j \in \{1, \ldots, m\}$ such that $f_j(x^*) < f_j(x)$ and*

$$\frac{f_i(x^*) - f_i(x)}{f_j(x) - f_j(x^*)} \leq M.$$

B Other Weighted Sum Approaches Using Robustness

In the literature the weighted l_p-metric ($p \in [1,\infty)$) (Yu and Leitmann (1974) and Chankong and Haim (1983)) and the weighted Tchebycheff approaches (Bowman (1976)) have also been studied as extensions to (WSO) (see Miettinen (1999)). We now present the corresponding robust weighted versions of these extensions.

Denote $g(x) = f(x) - f^*$ where f^* is an ideal point. Here, $f^* \le f(x)$ for all $x \in \mathcal{X}$. For example, f_i^* $(i = 1,\dots,m)$ may be chosen as the optimal value of the single objective problem $\min_{x\in\mathcal{X}} f_i(x)$. For a stochastic (MOO), denote $G(x,\xi) = F(x,\xi) - F^*(\xi)$. The ideal point $F^*(\cdot)$ is a function of the random scenario ξ and $F^*(\xi) \le F(x,\xi)$ for all $x \in \mathcal{X}$ w.p.1.

B.1 Robust Versions of Weighted l_p-metric Models

The weighted l_p-metric approach is described as

$$\min_{x\in\mathcal{X}} \left(w^T g(x)^p\right)^{1/p}. \tag{Wl_pO}$$

Here the power function of a vector is denoted as the power of each of its components, i.e., $y^p = (y_1^p,\dots,y_m^p)$ for $y \in \mathbb{R}^m$.

Robust Weighted l_p-metric Model

Write the robust weighted l_p-metric model as

$$\min_{x\in\mathcal{X}} \max_{w\in\mathcal{W}} \left(w^T g(x)^p\right)^{1/p}. \tag{RWl_pO}$$

Robust Expected-Value Weighted l_p-metric Model

In Section 5.1 we have presented an approach to regard weights as a random vector $\Lambda \in (\Omega, \mathcal{F}, \mathcal{W}_f, \mu)$ independent of the random scenario ξ. Using this idea, we build the following program which uses the mean of the random weight vector to trade off multiple objectives:

$$\min_{x\in\mathcal{X}} \left(\mathbb{E}_\mu[\Lambda^T] g(x)^p\right)^{1/p}. \tag{MWl_pO}$$

A robust version of (MWl_pO) is given as

$$\min_{x\in\mathcal{X}} \max_{\mu\in\mathcal{D}} \left(\mathbb{E}_\mu[\Lambda^T] g(x)^p\right)^{1/p}. \tag{RMWl_pO}$$

Robust Stochastically Weighted l_p-metric Model

We now discuss a case with totally correlated random weight vector and random scenario in objectives. Recall that the weight vector is regarded as a function of the random scenario ξ. The stochastically weighted l_p-metric model is denoted as

$$\min_{x\in\mathcal{X}} \left(\mathbb{E}\left[\Lambda(\xi)^T G(x,\xi)^p\right]\right)^{1/p}. \tag{FWl_pO}$$

A corresponding robust version of (FWl_pO) is written as

$$\min_{x\in\mathcal{X}}\left(\mathbb{E}\left[\max_{w\in\mathcal{W}(\xi)} w^T G(x,\xi)^p\right]\right)^{1/p}. \tag{RSWl_pO}$$

We now present reformulations of the three robust weighted l_p-metric models. The following proposition shows that these models are not different from the corresponding robust weighted sum models and hence the conclusions in Sections 3-5 are still applicable.

Lemma B.1 *For any $t > 0$, $\min_{z\in\mathcal{Q}}(\phi(z))^t$ is equivalent to $(\min_{z\in\mathcal{Q}}\phi(z))^t$.*

Proof: It is obvious that

$$\min_{z\in\mathcal{Q}}(\phi(z))^t \geq \left(\min_{z\in\mathcal{Q}}\phi(z)\right)^t.$$

Let z^* be an optimal solution of $\min_{z\in\mathcal{Q}}(\phi(z))^t$, i.e., $(\phi(z^*))^t \leq (\phi(z))^t$ for all $z \in \mathcal{Q}$. It follows that $\phi(z^*) \leq \phi(z)$ for all $z \in \mathcal{Q}$. Hence, we have

$$\min_{z\in\mathcal{Q}}(\phi(z))^t = (\phi(z^*))^t \leq \left(\min_{z\in\mathcal{Q}}\phi(z)\right)^t.$$

□

Proposition B.2

1. (RWl_pO) *is equivalent to*

$$\left(\min_{x\in\mathcal{X}}\max_{w\in\mathcal{W}} w^T g(x)^p\right)^{1/p}. \tag{B.1}$$

2. (RMWl_pO) *is equivalent to*

$$\left(\min_{x\in\mathcal{X}}\max_{\mu\in\mathcal{D}} \mathbb{E}_\mu[\Lambda^T] g(x)^p\right)^{1/p}. \tag{B.2}$$

3. (RSWl_pO) *is equivalent to*

$$\left(\min_{x\in\mathcal{X}}\mathbb{E}\left[\max_{w\in\mathcal{W}(\xi)} w^T G(x,\xi)^p\right]\right)^{1/p}. \tag{B.3}$$

Proof: Part 1 and Part 3 obviously hold by Lemma B.1. Referring to (5.3), we know that (RMWl_pO) is a special case of (RWl_pO). Then, Part 2 can be proved in the same way. □

B.2 Robust Versions of Weighted Tchebycheff Model

The weighted Tchebycheff formulation, or weighted minimax model, is given as

$$\min_{x\in\mathcal{X}}\max_{i\in\{1,\ldots,m\}} w_i g_i(x). \tag{WTO}$$

Let us now propose robust versions of (WTO) as following.

Robust Weighted Tchebycheff Model
We write the robust weighted Tchebycheff model as

$$\min_{x\in\mathcal{X}} \max_{\substack{i \in \{1,\ldots,m\} \\ w \in \mathcal{W}}} w_i g_i(x). \tag{RWTO}$$

Robust Expected-Value Weighted Tchebycheff Model
Similarly, for the independent random weight vector $\Lambda \in (\Omega, \mathcal{F}, \mathcal{W}_f, \mu)$, we write a program

$$\min_{x\in\mathcal{X}} \mathbb{E}[\Lambda_i] g_i(x), \tag{MWTO}$$

and its robust version

$$\min_{x\in\mathcal{X}} \max_{\substack{i \in \{1,\ldots,m\} \\ \mu \in \mathcal{D}}} \mathbb{E}_\mu[\Lambda_i] g_i(x). \tag{RMWTO}$$

Robust Stochastically Weighted Tchebycheff Model
We now describe the stochastically weighted Tchebycheff model for totally correlated random wight vector and random scenario as

$$\min_{x\in\mathcal{X}} \mathbb{E}\left[\max_{i\in\{1,\ldots,m\}} \Lambda_i(\xi) G_i(x,\xi)\right], \tag{FWTO}$$

and its robust version

$$\min_{x\in\mathcal{X}} \mathbb{E}\left[\max_{\substack{i \in \{1,\ldots,m\} \\ w \in \mathcal{W}(\xi)}} w_i G_i(x,\xi)\right]. \tag{RSWTO}$$

Although we can solve the robust weighted Tchebycheff models using results in Sections 3-5, the following proposition presents easier approaches for these special cases.

Proposition B.3

1. *Let* $w_i^* = \max_{w\in\mathcal{W}} w_i$. (RWTO) *is equivalent to*

$$\begin{aligned} \min_{x,\zeta}\ & \zeta \\ s.t.\ & \zeta \geq w_i^* g_i(x), \quad i = 1,\ldots,m, \\ & x \in \mathcal{X}. \end{aligned} \tag{B.4}$$

2. *Let* $\bar{w}_i^* = \max_{\mu\in\mathcal{D}} \mathbb{E}_\mu[\Lambda_i]$. (RMWTO) *is equivalent to*

$$\begin{aligned} \min_{x,\zeta}\ & \zeta \\ s.t.\ & \zeta \geq \bar{w}_i^* g_i(x), \quad i = 1,\ldots,m, \\ & x \in \mathcal{X}. \end{aligned} \tag{B.5}$$

3. *Let* $w_i^*(\xi) = \max_{w\in\mathcal{W}(\xi)} w_i$ *for* $\xi \in \Xi$. (RSWTO) *is equivalent to*

$$\min_{x\in\mathcal{X}} \mathbb{E}[H(x,\xi)], \tag{B.6}$$

where $H(x,\xi)$ *is the optimal value of the second stage problem*

$$\begin{aligned} &\min_{\zeta} \ \zeta \\ &\text{s.t. } \zeta \geq w_i^*(\xi) G_i(x,\xi), \quad i = 1,\ldots,m. \end{aligned} \tag{B.7}$$

Proof: Let us first prove Part 1. (RWTO) can be reformulated as

$$\begin{aligned} &\min_{x,\zeta} \ \zeta \\ &\text{s.t. } \zeta \geq \max_{w\in\mathcal{W}} w_i g_i(x), \quad i = 1,\ldots,m, \\ &\quad\ x \in \mathcal{X}. \end{aligned}$$

Since $\mathcal{W} \subseteq \mathbb{R}_+^m$ and $g_i(\cdot) \geq 0$ over the set $\mathcal{X}$, we further have

$$\max_{w\in\mathcal{W}} w_i g_i(x) = g_i(x) \max_{w\in\mathcal{W}} w_i = w_i^* g_i(x),$$

for a given $x \in \mathcal{X}$.

Part 2 and Part 3 also follow a similar proof argument. □

Remark B.4 *Proposition B.3 shows that* (RWTO) *is actually a special case of* (WTO). *Let* $\hat{w}_i^* = w_i^* / \sum_{i=1}^m w_i^*$. *Then* (B.4) *can be written in a formulation of* (WTO)

$$\min_{x\in\mathcal{X}} \max_{i\in\{1,\ldots,m\}} \hat{w}_i^* g_i(x).$$

C Proofs

C.1 Proof of Corollary 3.3

1. In this case the inner maximization problem of (McRow) is

$$\begin{aligned} &\max_{v} \ \hat{w}^T f(x) + v^T Q^T f(x) \\ &\text{s.t. } e^T v \leq 1, \\ &\quad\ v \geq 0, \\ &\quad\ v \in \mathcal{V}_f. \end{aligned}$$

Thus, we can directly obtain (3.6) by Theorem 3.2.

2. Introduce two nonnegative variables v_+ and v_- and let $v = v_+ - v_-$. Let $v' = \begin{bmatrix} v_+^T & -v_-^T \end{bmatrix}^T$

and $Q' = \begin{bmatrix} Q & Q \end{bmatrix}$. Now reformulate the inner maximization problem of (McRow) as

$$\begin{aligned}
\max_{v'} \ & \hat{w}^T f(x) + v'^T Q'^T f(x) \\
\text{s.t.} \ & \begin{bmatrix} -e^T & e^T \end{bmatrix} v' \geq -1 \\
& \begin{bmatrix} I_{k\times k} & 0_{k\times k} \end{bmatrix} v' \geq 0 \\
& \begin{bmatrix} 0_{k\times k} & -I_{k\times k} \end{bmatrix} v' \geq 0 \\
& v' \in \mathcal{V}'_f,
\end{aligned}$$

where $\mathcal{V}'_f = \{v \in \mathbb{R}^{2k}, \hat{w} + Q'v \geq 0,\ e^T(\hat{w} + Q'v) = 1\}$. Hence, by Theorem 3.2, (McRow) in this case is equivalent to

$$\begin{aligned}
\min_{x,z,s,\zeta} \ & \hat{w}^T s + \zeta - \begin{bmatrix} -1 & 0_{1\times 2k} \end{bmatrix} z \\
\text{s.t.} \ & Q'^T s + \begin{bmatrix} -e & I_{k\times k} & 0_{k\times k} \\ e & 0_{k\times k} & -I_{k\times k} \end{bmatrix} z = 0, \\
& f(x) - s - \zeta e \leq 0, \\
& x \in \mathcal{X},\ z \geq 0.
\end{aligned}$$

Let η be equal to the first component of vector z. Thus we can obtain (3.7).

3. In this case the inner maximization problem of (McRow) is written as

$$\begin{aligned}
\max_{v} \ & \hat{w}^T f(x) + v^T Q^T f(x) \\
\text{s.t.} \ & -e \leq v \leq e \\
& v \in \mathcal{V}_f.
\end{aligned}$$

Again, by Theorem 3.2, we obtain the reformulation of (McRow) as

$$\begin{aligned}
\min_{x,s,t,\zeta} \ & \hat{w}^T s + \zeta + e^T t \\
\text{s.t.} \ & Q^T s + \begin{bmatrix} I_{k\times k} & -I_{k\times k} \end{bmatrix} t = 0, \\
& f(x) - s - \zeta e \leq 0, \\
& x \in \mathcal{X},\ t \geq 0.
\end{aligned}$$

Letting $t = \begin{bmatrix} y \\ z \end{bmatrix}$, we rewrite the above problem as (3.8).

C.2 Proof of Corollary 3.5

Let $M(\lambda)$ be the objective value of (McRow) with the nest perturbation weight region $\mathcal{V}(\lambda)$. It follows from Theorem 3.2 that

$$\begin{aligned} M(\lambda) = \min_{x,y,z,s,\zeta} \ & \hat{w}^T s + \zeta - a^T y - b^T z + \lambda d^T z \\ \text{s.t. } & Q^T s + A^T y + B^T z = 0, \\ & f(x) - s - \zeta e \leq 0, \\ & x \in \mathcal{X},\ z \geq 0. \end{aligned}$$

is a nondecreasing concave function of λ for $d^T z \geq 0$.

C.3 Proof of Corollary 4.2

Lemma C.1 *Let $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then l_q-norm cone*

$$\mathbf{K}_* = \{(y, \eta) \in \mathbb{R}^{k+1} \mid \|y\|_q \leq \eta\}$$

is the dual of l_p-norm cone

$$\mathbf{K} = \{(v, \tau) \in \mathbb{R}^{k+1} \mid \|v\|_p \leq \tau\}.$$

Proof: By Hölder's inequality, we have $v^T y + \tau\eta \geq v^T y + \|v\|_p \|y\|_q \geq v^T y + |v^T y| \geq 0$. Hence, $\mathbf{K}_*$ is contained in the dual cone of $\mathbf{K}$. Suppose that (y, η) is an arbitrary element of the dual cone of $\mathbf{K}$. Thus, for any $(x, \tau) \in \mathbf{K}$, we have $v^T y + \tau\eta \geq 0$. If $y = 0$, it is trivial to see that η should be nonnegative so that $(y, \eta) \in \mathbf{K}_*$. Now let $y \neq 0$. Choose $v_i = -|y_i|^{q-1}\text{sign}(y_i)$ for $i = 1, \ldots, m$ and $\tau = \|y^{(q-1)}\|_p = \|y\|_q^{q/p}$ where $\text{sign}(\cdot)$ is a signum function. It follows that $\|y\|_q^q \leq \eta \|y\|_q^{q/p}$, which is equivalent to $\|y\|_q \leq \eta$. Hence, $\mathbf{K}_*$ contains the dual cone of $\mathbf{K}$. □

We now prove Corollary 4.2. Let

$$A = \begin{pmatrix} I_{k\times k} \\ 0_{1\times k} \end{pmatrix}, \quad B = 0_{(k+1)\times l}, \quad c = \begin{pmatrix} 0_{k\times 1} \\ 1 \end{pmatrix}, \quad \text{and } \mathbf{K} = \{(v, \tau) \in \mathbb{R}^{k+1} \mid \|v\|_p \leq \tau\},$$

in the conic representation (4.1). By Lemma C.1, we know that the dual cone of $\mathbf{K}$ is l_q-norm cone $\mathbf{K}_* = \{(y, \eta) \in \mathbb{R}^{k+1} \mid \|y\|_q \leq \eta\}$. Thus, the corresponding dual problem is

$$\begin{aligned} \min_{y,s,\zeta} \ & \hat{w}^T s + \zeta + \eta \\ \text{s.t. } & Q^T s + y = 0, \\ & \|y\|_q \leq \eta, \\ & f(x) - s - \zeta e \leq 0, \end{aligned}$$

which we can substitute for the inner maximization problem of (McRow) to obtain (4.3), since the Slater condition holds by assumption.

C.4 Proof of Corollary 4.4

Let $M(\gamma)$ be the objective value of (McRow) with the nest perturbation weight set $\mathcal{V}(\bar{v}, S, \gamma)$. For given $\bar{v} \in \mathcal{V}_f$ and $S \succ 0$, it follows by (4.5) that

$$
\begin{aligned}
M(\gamma) = \min_{x,s,\zeta,\eta} \ & \hat{w}^T s + \gamma \|S^{1/2} Q^T s\|_2 + \bar{v}^T Q^T s + \zeta \\
\text{s.t. } & f(x) - s - \zeta e \leq 0, \\
& x \in \mathcal{X}.
\end{aligned}
$$

It is obvious that $M(\cdot)$ is a nondecreasing concave function.

C.5 Proof of Proposition 4.6

Denote

$$
\mathcal{V}^i = \{v \in \mathbb{R}^m \mid \exists u \in \mathbb{R}^l \text{ s.t. } A^i v + B^i u + c^i \in \mathbf{K}^i\}, \quad i = 1, \ldots, r.
$$

By Proposition 2.3, we can substitute region $\mathcal{V}_f \cap \operatorname{conv}\left(\cup_{i=1,\ldots,r} \mathcal{V}^i\right)$ for $\mathcal{V}$ denoted by (4.8) in the corresponding (McRow). Built on $\mathcal{V}_f \cap \operatorname{conv}\left(\cup_{i=1,\ldots,r} \mathcal{V}^i\right)$, inner maximization problem of (McRow) is represented as

$$
\begin{aligned}
\max_{v,\lambda,v^1,\ldots,v^r} \ & \hat{w}^T f(x) + v^T Q^T f(x) \\
\text{s.t. } & v = [v^1 \ldots v^r]\lambda, \\
& v^i \in \mathcal{V}^i, \quad i = 1, \ldots, r, \\
& v \in \mathcal{V}_f, \\
& e^T \lambda = 1, \\
& \lambda \geq 0.
\end{aligned}
$$

By expressing $\mathcal{V}^i$ by constraints and letting $t^i = \lambda_i v^i$ and $q^i = \lambda_i u^i$, we rewrite the above maximization problem as

$$
\begin{aligned}
\max_{\lambda,t^1,\ldots,t^r,q^1,\ldots,q^r} \ & \hat{w}^T f(x) + \sum_{i=1}^r {t^i}^T Q^T f(x) \\
\text{s.t. } & A^i t^i + B^i q^i + c^i e_i^T \lambda \in \mathbf{K}^i, \quad i = 1, \ldots, r, \\
& \hat{w} + Q \sum_{i=1}^r t^i \geq 0, \\
& e^T \left(\hat{w} + Q \sum_{i=1}^r t^i \right) = 1, \\
& e^T \lambda = 1, \\
& \lambda \geq 0.
\end{aligned}
$$

The corresponding dual problem is

$$\begin{aligned}\min_{z,\zeta,\eta,y^1,\ldots,y^r} \quad & \hat{w}^T f(x) + \hat{w}^T z + (1 - e^T\hat{w})\zeta + \eta \\ \text{s.t.}\ & Q^T z - Q^T e\zeta + {A^i}^T y^i = -Q^T f(x), \quad i = 1,\ldots,r, \\ & {B^i}^T y^i = 0, \quad i = 1,\ldots,r, \\ & {c^i}^T y^i - \eta \le 0, \quad i = 1,\ldots,r, \\ & y^i \in \mathbf{K}^i_*, \quad i = 1,\ldots,r, \\ & z \ge 0.\end{aligned}$$

Letting $z = s + \zeta e - f(x)$, we have

$$\begin{aligned}\min_{s,\zeta,\eta,y^1,\ldots,y^r} \quad & \hat{w}^T s + \zeta + \eta \\ \text{s.t.}\ & Q^T s + {A^i}^T y^i = 0, \quad i = 1,\ldots,r, \\ & {B^i}^T y^i = 0, \quad i = 1,\ldots,r, \\ & {c^i}^T y^i - \eta \le 0, \quad i = 1,\ldots,r, \\ & f(x) - s - \zeta e \le 0, \\ & y^i \in \mathbf{K}^i_*, \quad i = 1,\ldots,r.\end{aligned}$$

In addition, by the assumption, it follows that there exists an interior point in the region $\mathcal{V}_f \cap \operatorname{conv}\left(\cup_{i=1,\ldots,r}\mathcal{V}^i\right)$. Thus, the above primal and dual problems have the same optimal value. We can obtain (4.9) by replacing the inner maximization problem of (McRow) by its dual problem.

C.6 Proof of Theorem 5.3

The following proof of Theorem 5.3 follows the arguments used in (Delage and Ye, 2010). We first describe the inner maximization problem of (5.8) as a semi-infinite conic linear problem

$$\max_{\mu} \int_{\mathcal{V}_f} (\hat{w} + Qv)^T f(x)\mu(dv) \tag{C.1a}$$

$$\text{s.t.} \int_{\mathcal{V}_f} \mu(dv) = 1, \tag{C.1b}$$

$$\int_{\mathcal{V}_f} (v - \bar{v})(v - \bar{v})^T \mu(dv) \preceq \kappa S, \tag{C.1c}$$

$$\int_{\mathcal{V}_f} \begin{bmatrix} S & (v - \bar{v}) \\ (v - \bar{v})^T & \gamma^2 \end{bmatrix} \mu(dv) \succeq 0, \tag{C.1d}$$

$$\mu \in \mathcal{M}. \tag{C.1e}$$

Let $\zeta \in \mathbb{R}$, $Y \in \mathbb{R}^{k\times k}$ be the Lagrangian multipliers of constraints (C.1b) and (C.1c) respectively and $Z \in \mathbb{R}^{k\times k}$, $z \in \mathbb{R}^k$, and $\rho \in \mathbb{R}$ to form together a matrix which is the dual variable associated with constraint (C.1d). Then the dual problem of (C.1) is constructed as

$$\min_{\zeta,\rho,z,Y,Z} \ (\kappa S - \bar{v}\bar{v}^T) \bullet Y + \zeta + S \bullet Z - 2\bar{v}^T z + \gamma^2 \rho \tag{C.2a}$$

$$\text{s.t. } v^T S v + 2v^T(z - Y\bar{v}) + \zeta - (\hat{w} + Qv)^T f(x) \geq 0, \quad \forall v \in \mathcal{V}_f, \tag{C.2b}$$

$$Y \succeq 0, \tag{C.2c}$$

$$\begin{bmatrix} Z & z \\ z^T & \rho \end{bmatrix} \succeq 0. \tag{C.2d}$$

Let $(\zeta^*, \rho^*, z^*, Y^*, Z^*)$ be an optimal solution of (C.2). Constraint (C.2d) implies that $\rho^* \geq 0$. If $\rho^* = 0$, we claim that $z^* = 0$. Indeed, if z^* is not 0, we have that ${z^*}^T z^* > 0$ and

$$\begin{bmatrix} z^* \\ q \end{bmatrix}^T \begin{bmatrix} Z^* & z^* \\ z^{*T} & \rho^* \end{bmatrix} \begin{bmatrix} z^* \\ q \end{bmatrix} = {z^*}^T Z^* z^* - 2{z^*}^T z^* q < 0,$$

for $q > \frac{{z^*}^T Z^* z^*}{2{z^*}^T z^*}$. This contradicts constraint (C.2d). Finally, $Z^* = 0$ is an optimal solution since it minimizes the objective. Therefore, when $\rho^* = 0$, replacing $y = 2(z - Y\bar{v})$, we reduces the objective (C.2a) to

$$(\kappa S - \bar{v}\bar{v}^T) \bullet Y + \zeta = \zeta + (\kappa S + \bar{v}\bar{v}^T) \bullet Y + \bar{v}^T y + \gamma \| S^{1/2}(2Y\bar{v} + y)\|_2. \tag{C.3}$$

Now assume that $\rho^* > 0$. By applying Schur complement, constraint (C.2d) is equivalent to $Z \succeq \frac{1}{\rho} z z^T$. Since $S \succeq 0$, we have $Z^* = \frac{1}{\rho} z z^T$. It remains to solve for $\rho^* > 0$ in the one dimensional convex optimization problem: $\min_{\rho > 0} \frac{1}{\rho} z^T S z + \gamma^2 \rho$. The corresponding optimal solution is $\rho^* = \frac{1}{\gamma}\sqrt{z^T S z}$. Again, by replacing $y = 2(z - Y\bar{v})$, we reduces the objective (C.2a) to

$$\zeta + (\kappa S + \bar{v}\bar{v}^T) \bullet Y + \bar{v}^T y + \gamma \| S^{1/2}(2Y\bar{v} + y)\|_2. \tag{C.4}$$

Therefore, (C.2) can be rewritten as

$$\min_{\zeta,\eta,y,Y} \ \zeta + \eta \tag{C.5a}$$

$$\text{s.t. } \zeta \geq (\hat{w} + Qv)^T f(x) - v^T y - v^T S v, \quad \forall v \in \mathcal{V}_f, \tag{C.5b}$$

$$\eta \geq (\kappa S + \bar{v}\bar{v}^T) \bullet Y + \bar{v}^T y + \gamma \| S^{1/2}(2Y\bar{v} + y)\|_2, \tag{C.5c}$$

$$Y \succeq 0. \tag{C.5d}$$

It is easy to show that $\gamma \geq 0$, $\kappa \geq 1$, and $S \succ 0$ are sufficient to ensure that the Dirac measure $\delta_{\bar{v}}$ lies in the relative interior of $\mathcal{D}_2(\mathcal{V}_f, \bar{v}, S, \gamma, \kappa)$, the feasible region of the inner maximization problem of (5.8). Based on the weaker version of Proposition 3.4 in Shapiro (2001), we can conclude that there is no duality gap between problems (C.1) and (C.5).