SCORE Allocations for Bi-objective Ranking and Selection

GUY FELDMAN and SUSAN R. HUNTER, Purdue University

Consider the context of multi-objective optimization via simulation (MOOvS), that is, multi-objective optimization problems in which the objective functions are known only through dependent Monte Carlo estimators. The solution to this problem is a non-dominated (Pareto) set. We consider the special case of MOOvS on finite sets with two objectives, called bi-objective ranking and selection (R&S). We exploit the special structure of this problem to derive (1) the asymptotically optimal simulation budget allocation, which we characterize as the solution to a bi-level optimization problem, and (2) an easily-implementable SCORE (Sampling Criteria for Optimization using Rate Estimators) sampling framework that approximates the optimal allocation when the number of design points, or systems, is large. Notably, our allocations account for dependence between the objectives and balance the probabilities associated with two types of misclassification error: misclassification by exclusion (MCE), in which a Pareto system is estimated as non-Pareto, and misclassification by inclusion (MCI), in which a non-Pareto system is estimated as Pareto. Like much of the R&S literature, our focus is on the case in which the simulation observations are bivariate normal.

Toward (2) and assuming normality, we show that in a certain asymptotic regime, the optimal allocations to non-Pareto systems in (1) are inversely proportional to a single intuitive measure called the score. Perhaps surprisingly, in this asymptotic regime, the optimal allocation exclusively controls for MCI events. We also provide an iterative algorithm that repeatedly estimates the score to determine how the available simulation budget should be expended across the competing systems. Our numerical experience with the resulting SCORE framework is promising. For problems of up to ten thousand systems, we are able to identify the optimal allocation to negligible error within a few seconds on a personal computer.

Key Words and Phrases: multi-objective, optimization via simulation, simulation optimization, ranking and selection

1. INTRODUCTION

The optimization via simulation (OvS) problem, also called the simulation optimization problem, is a nonlinear optimization problem in which the objective and constraint functions can only be observed with error as output from a Monte Carlo simulation. Such problems tend to arise when computer models are used to design complex systems under uncertainty — an increasingly popular practice [Powers et al. 2012]. Since the OvS formulation is quite general, OvS problems arise frequently in a variety of application areas, including agriculture [Hunter and McClosky 2016], energy [Marmidis et al. 2008; Subramanyan et al. 2011], and transportation [Osorio and Bierlaire 2013]. For additional examples and a library of OvS problems, see simopt.org [Henderson and Pasupathy 2016].

Methods to solve the OvS problem are often categorized by whether the feasible set contains categorical, integer-ordered, continuous, or mixed variables [Pasupathy and Henderson 2006]. Solution methods can further be categorized by the number of performance measures posed as objectives and constraints. In the presence of a single objective and deterministic constraints, mature solution methods are available for all types of feasible sets. For an overview of these methods and entry points into this literature, see Pasupathy and Ghosh [2013] and Fu [2015]. Recently, solution methods for a single objective with stochastic constraints have been proposed in the case of categorical variables [Andradóttir and Kim 2010; Lee et al. 2012; Pasupathy et al. 2014] and integer-ordered variables [Luo and Lim 2013; Nagaraj and Pasupathy 2016; Park and Kim 2015]. For methods with continuous variables, see, e.g., Ruszczyński and

Author’s addresses: G. Feldman, Department of Statistics, Purdue University, Haas Hall, 250 N. University Street, West Lafayette, IN 47907; email: gfeldman@purdue.edu; S. R. Hunter, School of Industrial Engineering, Purdue University, Grissom Hall, 315 N. Grant Street, West Lafayette, IN 47907; email: susan-hunter@purdue.edu.
Shapiro [2003], Homem-de-Mello and Bayraksan [2015], and references therein. However despite its mature development in the analogous deterministic context [Miettinen 1999, for example], few papers in the OvS literature provide solution methods in the presence of multiple simultaneous objectives – a problem we call the multi-objective optimization via simulation (MOOvS) problem.

We formulate the MOOvS problem as

$$\text{Problem } M : \min_{x \in D} (\mathbb{E}[G_1(x, \xi)], \ldots, \mathbb{E}[G_d(x, \xi)])$$

where $D \subseteq \mathbb{R}^q$ is a known feasible set, $\xi$ represents a random quantity, and each objective can be estimated as output from a Monte Carlo simulation. Since there may not exist a single point $x \in D$ that minimizes all objectives simultaneously, the solution to Problem $M$ is called the Pareto set of non-dominated points; that is, the set of all points for which no other point is “better” on all objectives. Since the Pareto set may be large and difficult to identify in full, Butler et al. [2001] propose utility functions as a method to solve Problem $M$ when the feasible set $D$ is finite. However recently, increased computing power has fueled interest in identifying the entire Pareto set [Eichfelder 2008].

We consider the context of solving Problem $M$ when the goal is to identify the entire Pareto set, the feasible set $D$ is finite or contains categorical variables, and there are two objectives. Methods to solve OvS problems in which $D$ is finite are often called ranking and selection (R&S) methods. Such methods require the feasible set to be small enough to permit simulation of each design point or “system” in $D$; see Kim and Nelson [2006] for an overview. R&S methods are broadly divided into two types: methods that provide a probabilistic guarantee on the quality of the returned solution with a finite sample size, and methods that provide a guarantee on sampling efficiency [Pasupathy and Ghosh 2013]. We fall in the latter category: we seek a method to identify the Pareto set that provides a rigorous guarantee on sampling efficiency; as such, we do not provide a finite-sample probabilistic guarantee on the quality of the returned solution.

1.1. Questions Answered

To explore in more detail what we mean by a guarantee on sampling efficiency, consider a simple algorithm to solve Problem $M$: (1) allocate some non-zero proportion of a total sampling budget to each system, (2) sample and construct estimators of the objectives for each system, (3) return the estimated set of non-dominated systems as the estimated Pareto set. Ideally, the estimated Pareto set returned at the end of this procedure will be equal to the true Pareto set; if not, a misclassification occurs. Under mild regularity conditions, as the total sampling budget tends to infinity, the probability of a misclassification decays to zero. Then we ask, what proportion of the total sampling budget should be allocated to each system to maximize the rate of decay of the probability of misclassification, as the sampling budget tends to infinity?

As may be expected given prior work answering a similar question in other contexts — notably Glynn and Juneja [2004] for unconstrained OvS, Szechtman and Yücesan [2008] for feasibility determination, and Hunter and Pasupathy [2013], Pasupathy et al. [2014] for stochastically constrained OvS — we characterize the asymptotically optimal sampling budget as the solution to a bi-level optimization problem in which the “outer” problem is concave maximization, and the “inner” problems are convex minimization. Importantly, our allocation accounts for correlation between the objectives, and balances the probabilities associated with two types of misclassification error: misclassification by exclusion (MCE), in which a Pareto system is estimated as non-Pareto, and misclassification by inclusion (MCI), in which a non-Pareto system is estimated as Pareto. However, solving for the optimal allocation may be computationally burden-
some. Then we then ask, **what is the asymptotically optimal sampling allocation when the number of non-Pareto systems is large?**

As the number of non-Pareto systems tends to infinity in a certain rigorous sense, we find that the optimal allocation for a non-Pareto system is inversely proportional to a single intuitive measure called the **score**. When the random vectors corresponding to the objectives are bivariate normal, which is our focus, the score is the squared standardized “distance” between the non-Pareto system and the Pareto frontier. Consistent with Pasupathy et al. [2014], we determine the relative allocations to the suboptimal systems by their scores. We call these allocations the bi-objective Sampling Criteria for Optimization using Rate Estimators (SCORE) allocations.

We also highlight a key insight from this work that may be surprising: **when the number of non-Pareto systems is large relative to the number of Pareto systems, the optimal allocation exclusively controls for the probability of an MCI event.** To understand why this is true, for now, let MCE be the event that a Pareto is falsely excluded by another Pareto, while MCI is the event that a non-Pareto is falsely included in the Pareto set, whether it is estimated as dominating a Pareto or not. Then loosely speaking, as the number of non-Pareto systems tends to infinity, the Pareto systems each compete with more and more non-Pareto systems. Thus the Pareto systems receive a larger proportion of sample than the non-Pareto systems. Indeed, they receive so much more sample that the probability of a Pareto system falsely excluding another Pareto system is small relative to the probability of a non-Pareto being falsely included in the estimated Pareto set. Thus the Pareto front appears “fixed” relative to the non-Pareto systems, and the optimal allocation exclusively controls for MCI events.

While it could be argued that the primary insight of this work is theoretical, we also include a heuristic sampling allocation that is based on the theory. The SCORE sampling allocation we propose is fast and accurate for 20 to 10,000 systems. We also provide a sequential framework for implementation, and compare the performance of the SCORE sampling allocation with other popular allocation schemes in the literature.

### 1.2. Other Relevant Work

When the goal in solving Problem $M$ is to identify the entire Pareto set, few solution methods have been proposed in the OvS literature. Arguably, the most well-known and popular method is the Multi-objective Optimal Computing Budget Allocation (MOCBA) [Lee et al. 2010], which is a multi-objective version of the popular Optimal Computing Budget Allocation [Chen et al. 2000] for a finite feasible set. Other recent works include (1) M-MOBA [Branke and Zhang 2015], a multi-objective version of the small-sample expected value of information (EVI) procedures in Chick et al. [2010] for a finite feasible set; (2) MO-COMPASS [Li et al. 2015], which is a multi-objective version of COMPASS [Xu et al. 2010] for integer-ordered feasible sets; (3) Huang and Zabinsky [2014], who provide a branch-and-bound algorithm for an integer-ordered or continuous feasible sets; and (4) Kim and Ryu [2011b], Kim and Ryu [2011a], and Ryu et al. [2009] who provide methods for continuous feasible sets.

Given our context of finite feasible sets, the only methods appropriate for comparison with our sequential algorithm are MOCBA [Lee et al. 2010] and M-MOBA [Branke and Zhang 2015]. We compare the performance of our sequential algorithm with these methods in Section 8. We find that our numerical implementation of the SCORE allocation performs well relative to its competitors on our test problems. SCORE appears to be a fast and accurate heuristic allocation scheme for bi-objective R&S, inspired by theoretical allocations that have limiting optimality guarantees on efficiency.
We consider the MOOvS problem with two objectives on a finite set. That is, we solve

\[ \text{Problem } B : \quad \text{Find } \arg \min_{x \in \mathcal{D}} \left( \mathbb{E}[G(x, \xi)], \mathbb{E}[H(x, \xi)] \right), \]

where \( \mathcal{D} \) is a known finite set and \( \xi \) is a random quantity. Since \( \mathcal{D} \) is finite, we index the points \( x \in \mathcal{D} \) by \( i = 1, \ldots, r \) and, for notational simplicity, define \( g_i := \mathbb{E}[G(x_i, \xi)] \) and \( h_i := \mathbb{E}[H(x_i, \xi)] \) for all “systems” \( i \in \{1, \ldots, r\} \). The objective vector \( (g_i, h_i) \in \mathbb{R}^2 \) for each system is unknown, but may be estimated as a vector of sample means.

Remark 1.1. We would be remiss not to mention the popular evolutionary algorithms for solving MOOvS problems, e.g., those cited in Zhou et al. [2011]. However, these methods tend not to provide the efficiency and solution quality guarantees standard in the OvS literature [Ólafsson 2006] and which we seek. We also note that multi-objective algorithms exist in the context of the multi-armed bandit problem [Yahyaa et al. 2014c; 2014a; 2014b].

Remark 1.2. A preliminary version of this work appears in Hunter and Feldman [2015]. Also, Hunter and McClosky [2016] contains an asymptotically optimal allocation for the case of two independent objectives in the context of a plant breeding application. This paper is a significant outgrowth of Hunter and Feldman [2015] and subsumes the allocation provided in Hunter and McClosky [2016] for independent objectives. Feldman et al. [2015] provides analogous MOOvS methods on finite sets for more than two objectives. Since the methods in Feldman et al. [2015] are more computationally burdensome than those we propose, we do not advocate using the methods of Feldman et al. [2015] in the bi-objective case. Thus we do not include these methods in numerical comparisons.

2. PROBLEM SETTING AND FORMULATION

We now provide a formal problem statement, describe notational conventions used in the paper, and outline our assumptions. Due to space constraints, unless otherwise noted in the text, proofs for all results appear in the Online Appendix.

2.1. Problem Statement

We consider the MOOvS problem with two objectives on a finite set. That is, we solve

\[ \text{Problem } B : \quad \text{Find } \arg \min_{x \in \mathcal{D}} \left( \mathbb{E}[G(x, \xi)], \mathbb{E}[H(x, \xi)] \right), \]

where \( \mathcal{D} \) is a known finite set and \( \xi \) is a random quantity. Since \( \mathcal{D} \) is finite, we index the points \( x \in \mathcal{D} \) by \( i = 1, \ldots, r \) and, for notational simplicity, define \( g_i := \mathbb{E}[G(x_i, \xi)] \) and \( h_i := \mathbb{E}[H(x_i, \xi)] \) for all “systems” \( i \in \{1, \ldots, r\} \). The objective vector \( (g_i, h_i) \in \mathbb{R}^2 \) for each system is unknown, but may be estimated as a vector of sample means.

There may not be a single system \( i \) that minimizes both objectives in Problem \( B \). Thus the solution to Problem \( B \) is the set of non-dominated systems

\[ \mathcal{P} := \{ \text{systems } i : \not \exists \text{ system } k \in \{1, \ldots, r\} \text{ such that } k \preceq i \}, \]

where system \( k \) dominates system \( i \), written \( k \preceq i \), if and only if \( g_k \leq g_i \) and \( h_k \leq h_i \), or \( g_k < g_i \) and \( h_k < h_i \).

Now consider a method to solve Problem \( B \) in which we allocate a proportion \( \alpha_i > 0 \) of the total sampling budget to each system \( i = 1, \ldots, r \), where \( \sum_{i=1}^{r} \alpha_i = 1 \). Once the total sampling budget has been expended, we return the estimated Pareto \( \mathcal{P} \), constructed as follows. Let the vector of sample means after \( n \) samples be \( (\hat{G}_i(n), \hat{H}_i(n)) := (\frac{1}{n} \sum_{k=1}^{n} G_{ik}, \frac{1}{n} \sum_{k=1}^{n} H_{ik}) \) for all \( i = 1, \ldots, r \), and define \( (\hat{G}_i, \hat{H}_i) := (\hat{G}(\alpha_i n), \hat{H}(\alpha_i n)) \) as the estimators of \( g_i \) and \( h_i \) after scaling the total sample size \( n \) by the proportional allocation to system \( i \), \( \alpha_i > 0 \). Then

\[ \hat{\mathcal{P}} := \{ \text{systems } i : \not \exists \text{ system } k \in \{1, \ldots, r\} \text{ such that } k \preceq i \}, \]

where \( k \preceq i \) if and only if \( \hat{G}_k \leq \hat{G}_i \) and \( \hat{H}_k \leq \hat{H}_i \), or \( \hat{G}_k < \hat{G}_i \) and \( \hat{H}_k < \hat{H}_i \).

If \( \hat{\mathcal{P}} \neq \mathcal{P} \), then at least one system has been misclassified, that is, a Pareto system has been falsely estimated as non-Pareto, or a non-Pareto system has been falsely
estimated as Pareto. As the sampling budget tends to infinity, the probability of misclassification tends to zero. Then we ask, what sampling budget \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_r) \) maximizes the rate of decay of the probability of misclassification?

Remark 2.1. While we focus on allocating the sample to maximize the rate of decay of the probability of misclassification, one could also allocate to minimize the expected number of misclassifications. In Hunter and McClosky [2016], it was shown that these two objectives result in identical asymptotic allocations. We anticipate that a similar result holds in the context of this paper.

2.2. Notational Conventions

When it is reasonable to do so, uppercase letters denote random variables, lowercase letters denote fixed quantities, script letters denote sets, and vectors are written in bold. We often use \( i \leq r \) as shorthand for \( i = 1, 2, \ldots, r \). For a set \( \mathcal{C} \), the cardinality of \( \mathcal{C} \) is denoted \( |\mathcal{C}| \). For a function \( f \), let \( \nabla f(x) \) be the gradient of \( f \) with respect to \( x \), and \( f'(x) \) the derivative of \( f \) with respect to \( x \). For a sequence of real numbers \( \{a_n\} \), we say that \( a_n = O(1) \) if \( \{a_n\} \) is bounded, that is, if there exists \( c > 0 \) with \( |a_n| < c \) for all \( n \). Further, \( a_n = \Theta(1) \) if \( 0 < \lim \inf a_n \leq \lim \sup a_n < \infty \). We use \( \text{iff} \) for “if and only if.” Solutions to optimization problems are often denoted with an asterisk, e.g., \( x^* \).

2.3. Assumptions

In what follows, we assume that the set of non-Pareto systems is nonempty. To estimate the unknown quantities \( g_i \) and \( h_i \), we assume we obtain replicates of the random vector \((G_i, H_i)\) from each system. We also assume the following.

**Assumption 1.** Random vectors \((G_i, H_i)\) are mutually independent for all \( i \leq r \).

That is, we develop a model to guide sampling that does not specifically account for correlation between systems, such as the correlation that would arise with the use of common random numbers (CRN). However, our model does not preclude the use of CRN during implementation. We also require the following technical assumption which is standard in optimal allocation literature, since it ensures all systems in the Pareto set are distinguishable on each objective with a finite sample size.

**Assumption 2.** We assume \( g_i \neq g_k \) and \( h_i \neq h_k \) for all \( i \in \mathcal{P} \) and \( k \leq r, k \neq i \).

Since we employ a large deviations (LD) analysis in Section 3, we require the following Assumptions 3 and 4, included here for completeness. We refer the reader to Dembo and Zeitouni [1998] for further explanation. Let \( \langle \cdot, \cdot \rangle \) denote the dot product, and let \( \Lambda_G^{(n)}(\theta) = \log \mathbb{E}[e^{\theta G_i(n)}] \), \( \Lambda_H^{(n)}(\theta) = \log \mathbb{E}[e^{\theta H_i(n)}] \), and \( \Lambda_{(G_i, H_i)}^{(n)}(\theta) = \log \mathbb{E}[e^{\theta (G_i(n), H_i(n))}] \) be the cumulant generating functions of \( G_i(n) \), \( H_i(n) \), and \( (G_i(n), H_i(n)) \), respectively, where \( \theta \in \mathbb{R}^2 \). Let the effective domain of \( f(\cdot) \) be \( \mathcal{D}_f = \{x : f(x) < \infty\} \), and its interior \( \mathcal{D}_f^\circ \). We make the following standard assumption.

**Assumption 3.** For each system \( i = 1, 2, \ldots, r \),

1. the limit \( \Lambda_{(G_i, H_i)}(\theta) = \lim_{n \to \infty} \frac{1}{n} \Lambda_{(G_i, H_i)}^{(n)}(n\theta) \) exists as an extended real number for all \( \theta \in \mathbb{R}^2 \), where \( \Lambda_{G_i}(\theta) = \lim_{n \to \infty} \frac{1}{n} \Lambda_G^{(n)}(n\theta) \) and \( \Lambda_{H_i}(\theta) = \lim_{n \to \infty} \frac{1}{n} \Lambda_H^{(n)}(n\theta) \) for all \( \theta \in \mathbb{R} \);
2. the origin belongs to the interior of \( \mathcal{D}_{\Lambda_{(G_i, H_i)}} \);
3. \( \Lambda_{(G_i, H_i)}(\theta) \) is strictly convex and \( C^\infty \) on \( \mathcal{D}_{\Lambda_{(G_i, H_i)}}^\circ \);
4. \( \Lambda_{(G_i, H_i)}(\theta) \) is steep, that is, for any sequence \( \{\theta_n\} \in \mathcal{D}_{\Lambda_{(G_i, H_i)}} \) converging to a boundary point of \( \mathcal{D}_{\Lambda_{(G_i, H_i)}} \), \( \lim_{n \to \infty} |\nabla \Lambda_{(G_i, H_i)}(\theta_n)| = \infty \).
Assumption 3 implies that by the Gärtner-Ellis theorem, the probability measure governing \((G_i(n), H_i(n))\) obeys the large deviations principle (LDP) with good, strictly convex rate function \(I_i(x, y) = \sup_{\theta \in \mathbb{R}^2} \{\theta \cdot (x, y) - \Lambda_{G_i, H_i}(\theta)\}\) [Dembo and Zeitouni 1998, p.44]. Let \((x, y) \in \mathcal{F}_{G_i, H_i}^n = \text{int}\{\Lambda_{G_i, H_i}(\theta) : \theta \in \mathcal{D}_{\Lambda_{G_i, H_i}}\}\), and let \(\mathcal{F}_{G_i, H_i}^n\) denote the closure of the convex hull of the set \(\{(g_i, h_i) : (g_i, h_i) \in \mathbb{R}^2, i \in \{1, 2, \ldots, r\}\}\).

ASSUMPTION 4. The closure of the convex hull of all points \((g_i, h_i) \in \mathbb{R}^2\) is a subset of the intersection of the interiors of the effective domains of the rate functions \(I_i(x, y)\) for all \(i = 1, 2, \ldots, r\), that is, \(\bigcap_{i=1}^r \mathcal{F}_{G_i, H_i}^n\).

3. CHARACTERIZATION OF THE OPTIMAL BUDGET ALLOCATION

Given that our goal is to determine the sample allocation \(\alpha\) that maximizes the rate of decay of the probability of misclassification, we first formulate the misclassification event in a way that facilitates analysis. We then analyze the rate of decay of the probability of misclassification, as a function of \(\alpha\), and provide a characterization of the optimal budget allocation as the solution to a bi-level optimization problem. To avoid mathematical complications, we assume \(n\alpha > 1\) for all \(i \leq r\) in this section.

3.1. Formulation of the Misclassification Event

Recall that a misclassification event occurs if, after expending a total of \(n\) samples, the estimated Pareto set, \(\hat{P}\), is not equal to the true Pareto set, \(P\). If \(\hat{P} \neq P\), then at least one of two events occurs: (1) a Pareto system was estimated as non-Pareto, which we call misclassification by exclusion (MCE), or (2) a non-Pareto system was estimated as Pareto, which we call misclassification by inclusion (MCI). Therefore we can formulate the misclassification event as \(MC := MCE \cup MCI\), where

\[
\begin{align*}
MCE & := \bigcup_{i \in P} \bigcup_{k \leq r, k \neq i} (\hat{G}_k \leq \hat{G}_i) \cap (\hat{H}_k \leq \hat{H}_i); \\
MCI & := \bigcup_{j \in P^c} \bigcap_{k \leq r, k \neq j} (\hat{G}_j \leq \hat{G}_k) \cup (\hat{H}_j \leq \hat{H}_k).
\end{align*}
\]

While the MCE event is easy to analyze, the MCI event contains dependence that is difficult to analyze. To overcome this difficulty, we reformulate the MCI event.

To reformulate the MCI event, first, label the true Pareto systems by the ordering of their means, so that \(g_1 < g_2 < \ldots < g_{p-1} < g_p < g_{p+1}\) and \(h_0 > h_1 > h_2 > \ldots > h_{p-1} > h_p\), where \(p = |P|\) is the number of Pareto systems, \(g_{p+1} := \infty\), and \(h_0 := \infty\). Then the objective values for the true Pareto systems are \((g_i, h_i)\) for \(\ell = 1, \ldots, p\), where we use \(\ell\) exclusively to index the Pareto systems. An example of this labeling appears in Figure 1. Now construct the phantom Pareto systems as the coordinates \((g_{\ell+1}, h_{\ell})\) for \(\ell = 0, \ldots, p\), where we also place phantom Pareto systems at \((g_1, \infty)\) and \((\infty, h_p)\) for a total of \(p + 1\) phantom Pareto systems. We now use the phantom Pareto systems to construct
an MCI event in an MCE-like fashion: an MCI event occurs if and only if a non-Pareto system is estimated as dominating a phantom Pareto system. Notice from Figure 1 that there is some “overlap” between MCE and MCI events — if a non-Pareto system is falsely included in the estimated Pareto set by falsely dominating a Pareto system, then MCE and MCI events occur together.

To rewrite the MCI event rigorously using the phantom Pareto systems, we must estimate the objective values for the phantom Pareto systems. Define \( \hat{G}[\ell] \) and \( \hat{H}[\ell] \) as the \( \ell \)th order statistics of the estimated objectives of the true Pareto set. Thus \( \hat{G}[1] < \cdots < \hat{G}[p-1] < \hat{G}[p] < \hat{G}[p+1] \) and \( \hat{H}[0] > \hat{H}[1] > \hat{H}[2] > \ldots > \hat{H}[p] \), where \( \hat{G}[p+1] := \infty \) and \( \hat{H}[0] := \infty \) for all \( n \). Now the estimated objective values for the phantom Pareto systems are \( (\hat{G}[\ell+1], \hat{H}[\ell]) \) for \( \ell = 0, 1, \ldots, p \). Define misclassification by dominating a phantom system as \( \text{MCL}^- \) and \( \text{MCL}^+ \), which, assuming the limits exist, implies

\[
\lim_{n \to \infty} \frac{1}{n} \log P\{\text{MC}\} = \min \left( \lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCE}\}, \lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCI}^-\} \right). \tag{1}
\]

In what follows, we analyze the rate of decay of \( P\{\text{MCE}\} \) and \( P\{\text{MCI}^-\} \) separately.

First, consider the rate of decay of \( P\{\text{MCE}\} \) in equation (1), since it is the most straightforward. For brevity, for all \( i \in \mathcal{P}, k \leq r, k \neq i \), define the rate function

\[
R_i(\alpha_i, \alpha_k) := \inf_{x_k \leq x_i, y_k \leq y_i} \alpha_i I_i(x_i, y_i) + \alpha_k I_k(x_k, y_k).
\]

The following Lemma 3.2 states the rate of decay of \( P\{\text{MCE}\} \) in terms of the “pairwise” rates \( R_i(\alpha_i, \alpha_k) \) corresponding to a system \( k \) dominating a Pareto system \( i \). We do not provide a proof for Lemma 3.2 since it follows directly from an analysis similar to that in Glynn and Juneja [2004] and Hunter [2011]. This result also appears in Li [2012].

**Lemma 3.2.** The rate of decay of \( P\{\text{MCE}\} \) is

\[
- \lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCE}\} = \min_{i \in \mathcal{P}} \min_{k \leq r, k \neq i} R_i(\alpha_i, \alpha_k).
\]

Lemma 3.2 states that the rate of decay of \( P\{\text{MCE}\} \) is determined by the slowest rate for the probability that a Pareto system is falsely dominated by some other system.

Now let us turn our attention to the term corresponding to \( \text{MCI}^- \) in (1). The analysis for the rate of decay of the probability of an \( \text{MCI}^- \) event is a bit more involved: in addition to the possibility that a non-Pareto system \( j \) is estimated as dominating a phantom Pareto system, the Pareto systems themselves may be estimated out of “order.” In what follows, we do not directly state the rate of decay of the probability of \( \text{MCI}^- \). Instead, we show that the events corresponding to \( \text{MCI}^- \) in which the Pareto
systems are also estimated out of order have rates of decay faster than the rate of decay of MCE, and thus can never be the unique minimum rate in (1).

To explicitly denote the ordering of the Pareto systems, we require the following notation. First, recall that the Pareto systems are labeled in “order” from 1, 2, . . . , p. Then we define the ordered list \( \mathcal{O} := \{(1, 1), (2, 2), \ldots, (p, p)\} \) as the positions of the true Pareto set on each objective, where the first objective is labeled from smallest to largest, and the second objective is labeled from largest to smallest. Now define \( \hat{\mathcal{O}} \) as the ordered list of estimated positions of the true Pareto set. Thus the event that the Pareto set is estimated in the correct order is \( \mathcal{O} = \hat{\mathcal{O}} \). Further, let \( S \) be an ordered set of \( p \) elements of the form \( \{(x_1, y_1), (x_2, y_2), \ldots, (x_p, y_p)\} \) where the \( x \) and \( y \) coordinates are separately drawn without replacement from the set of Pareto indices \( \{1, 2, \ldots, p\} \). The set \( S \) corresponds to a (fixed) realized instance of \( \hat{\mathcal{O}} \).

Define \( \text{MCI}_{ph} \) without order statistics, \( \text{MCI}_{ph}^* := \bigcup_{j \in \mathcal{P}} \bigcup_{\alpha = 0}^p (\hat{G}_j \leq \hat{G}_{j+1}) \cap (\hat{H}_j \leq \hat{H}_{j+1}) \), where \( \hat{G}_{p+1} := \infty, \hat{H}_0 := \infty \) for all \( n \). Then the event \( \text{MCI}_{ph}^* \cap \hat{\mathcal{O}} = \mathcal{O} \) is the event that the Pareto set is estimated “in order,” and a non-Pareto system is falsely included in the estimated Pareto set. The following lemma states that only this event can be a binding minimum in the overall rate of decay of MC in (1).

**Lemma 3.3.** The rate of decay of \( P\{MC\} \) is

\[
- \lim_{n \to \infty} \frac{1}{n} \log P\{MC\} = \min \left( \lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCE}\}, \lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^* \cap \hat{\mathcal{O}} = \mathcal{O}\} \right).
\]

Notice that the probability that the Pareto set is estimated in order, \( P\{\hat{\mathcal{O}} = \mathcal{O}\} \), becomes certain in the limit. Thus loosely speaking, unless the \( \text{MCI}_{ph}^* \) event and the \( \hat{\mathcal{O}} = \mathcal{O} \) events are colluding, the event \( \hat{\mathcal{O}} = \mathcal{O} \) should not affect the rate of decay of the probability of \( \text{MCI}_{ph}^* \). As it turns out, this intuition is correct — the events are not colluding, and the rate of decay of \( \text{MCI}_{ph}^* \) is not affected. For all \( j \in \mathcal{P}^e, \ell = 0, \ldots, p \), define the rate function corresponding to the \( \text{MCI}_{ph}^* \) event as

\[
R_j(\alpha_j, \alpha_\ell, \alpha_{\ell+1}) := \begin{cases} 
\inf_{x_j \leq x_1} & \alpha_j I_j(x_j, y_j) + \alpha_1 I_1(x_1, y_1) & \text{if } \ell = 0 \\
\inf_{x_j \leq x_{\ell+1}, y_j \leq y_1} & \alpha_j I_j(x_j, y_j) + \sum_{i=\ell}^{\ell+1} \alpha_i I_i(x_i, y_i) & \text{if } \ell \in \{1, \ldots, p-1\} \\
\inf_{y_j \leq y_p} & \alpha_j I_j(x_j, y_j) + \alpha_p I_p(x_p, y_p) & \text{if } \ell = p,
\end{cases}
\]

where \( \alpha_0 := 1 \) and \( \alpha_{p+1} := 1 \). The following Theorem 3.4 essentially states that we can retrieve the overall rate of decay of \( P\{MC\} \) by considering only the rate of decay of \( P\{\text{MCE}\} \) and the rate of decay of \( P\{\text{MCI}_{ph}^* \} \).

**Theorem 3.4.** The rate of decay of the probability of misclassification is

\[
- \lim_{n \to \infty} \frac{1}{n} \log P\{MC\} = \min \left( \min_{i \in \mathcal{P}} \min_{k \in \mathcal{P}, k \neq i} R_i(\alpha_i, \alpha_k), \min_{\ell = 0, \ldots, p} \min_{j \in \mathcal{P}^e} R_j(\alpha_j, \alpha_\ell, \alpha_{\ell+1}) \right).
\]

Therefore according to Theorem 3.4, the rate of decay of the probability of a misclassification event is the minimum rate among all pairwise MCE events and all “pairwise” MCI events, where the “pairs” in the MCI event are constructed by considering the event that a non-Pareto \( j \in \mathcal{P}^e \) falsely dominates a phantom Pareto \( \ell \in \{0, 1, \ldots, p\} \). Notice that in Theorem 3.4 we only consider pairwise MCE events between Pareto systems. We can make this simplification because a non-Pareto system dominates a Pareto system if and only if it also dominates a phantom Pareto system.
3.3. Optimal Allocation Strategy

We now consider the rate of decay of $P\{MC\}$ in Theorem 3.4 as a function of the sampling allocation $\alpha$. To maximize the rate of decay of $P\{MC\}$, we solve the following Problem $Q$, defined as

\[
\text{Problem } Q: \quad \text{maximize } z \quad \text{s.t.} \\
R_i(\alpha_i, \alpha_k) \geq z \text{ for all } i, k \in \mathcal{P} \text{ such that } k \neq i, \\
R_j(\alpha_j, \alpha_\ell, \alpha_\ell+1) \geq z \text{ for all } j \in \mathcal{P}_c, \ell = 0, \ldots, p, \\
\sum_{i=1}^r \alpha_i = 1, \quad \alpha_i \geq 0 \text{ for all } i \leq r.
\]

Given a value of $(\alpha_i, \alpha_k)$, the value of $R_i(\alpha_i, \alpha_k)$ is obtained by solving

\[
\text{Problem } R_{ik}^p: \quad \text{minimize } \alpha_i I(x, y_i) + \alpha_k I_k(x_k, y_k) \quad \text{s.t. } x_k \leq x_i, \quad y_k \leq y_i,
\]

and given a value of $(\alpha_j, \alpha_\ell, \alpha_\ell+1)$, the value of $R_j(\alpha_j, \alpha_\ell, \alpha_\ell+1)$ is obtained by solving

\[
\text{Problem } R_{j\ell}: \quad \text{minimize } \alpha_j I_j(x_j, y_j) + \alpha_\ell I_\ell(x_\ell, y_\ell) \quad \text{s.t. } (x_\ell - x_{\ell+1})_p \leq 0, \quad (y_\ell - y_{\ell+1})_p \leq 0.
\]

As a matter of notation, we distinguish Problems $R_{ik}^p$ and $R_{j\ell}$ as strictly convex optimization problems in $(x_i, y_i, x_k, y_k)$ and $(x_j, y_j, x_\ell, y_\ell, x_{\ell+1}, y_{\ell+1})$, respectively, while $R_i(\alpha_i, \alpha_k)$ and $R_j(\alpha_j, \alpha_\ell, \alpha_\ell+1)$ are their respective objective values at optimality.

In the sections that follow, Problem $R_{ik}$ plays a prominent role. Thus we briefly discuss the properties of Problem $Q$, then we provide a more in-depth look at the properties of Problem $R_{j\ell}$.

3.3.1. Properties of Problem $Q$. Problem $Q$ has $p \times (p-1)$ constraints corresponding to controlling the rate of decay of $P\{MC\}$ and $(r-p) \times (r+1)$ constraints corresponding to controlling the rate of decay of $P\{MCI_{p\ell}\}$. Also, $R_i(\alpha_i, \alpha_k)$ and $R_j(\alpha_j, \alpha_\ell, \alpha_\ell+1)$ are concave functions of $(\alpha_i, \alpha_k)$ and $(\alpha_j, \alpha_\ell, \alpha_\ell+1)$, respectively, for all $i, k \in \mathcal{P}_c, j \in \mathcal{P}_c$, and $\ell = 0, 1, \ldots, p$ [Boyd and Vandenberghe 2004, p. 81]. Slater's condition [Boyd and Vandenberghe 2004] holds for Problem $Q$, so the KKT conditions are necessary and sufficient for global optimality.

3.3.2. Properties of Problem $R_{j\ell}$. Along with primal feasibility, the following KKT conditions are necessary and sufficient for global optimality in the strictly convex minimization Problem $R_{j\ell}$. Let $(x_j^*, y_j^*, x_\ell^*, y_\ell^*, x_{\ell+1}^*, y_{\ell+1}^*)$ be the solution to Problem $R_{j\ell}$, where $(x_0^*, y_0^*) := (0, 0), (x_{p+1}^*, y_{p+1}^*) := (0, 0)$. Letting $\lambda_x \geq 0$ and $\lambda_y \geq 0$ be dual variables in Problem $R_{j\ell}$, we have complementary slackness conditions $\lambda_x(x_j^* - x_{\ell+1}^*) = 0$ if $\ell \neq p$, $\lambda_y(y_j^* - y_{\ell+1}^*) = 0$ if $\ell \neq 0$; and stationarity conditions

\[
\begin{align*}
\alpha_j \frac{\partial I_j(x_j^*, y_j^*)}{\partial x_j} + \lambda_x \|e_p\| = 0, & \quad \alpha_\ell \frac{\partial I_\ell(x_\ell^*, y_\ell^*)}{\partial y_\ell} + \lambda_y \|e_0\| = 0, \\
\alpha_j \frac{\partial I_j(x_j^*, y_j^*)}{\partial y_j} = 0 & \text{ if } \ell \neq 0, \quad \alpha_\ell \frac{\partial I_\ell(x_\ell^*, y_\ell^*)}{\partial x_\ell} - \lambda_y \|e_0\| = 0, \\
\alpha_{\ell+1} \frac{\partial I_{\ell+1}(x_{\ell+1}^*, y_{\ell+1}^*)}{\partial y_{\ell+1}} & = 0 \text{ if } \ell \neq p, \quad \alpha_{\ell+1} \frac{\partial I_{\ell+1}(x_{\ell+1}^*, y_{\ell+1}^*)}{\partial x_{\ell+1}} = 0.
\end{align*}
\]

In the solution to Problem $R_{j\ell}$, $x_j^*, y_j^*, x_\ell^*, y_\ell^*, x_{\ell+1}^*$, and $y_{\ell+1}^*$ are each functions of $(\alpha_j, \alpha_\ell, \alpha_{\ell+1})$. We suppress this dependence in our notation unless it is required for clarity. When the dependence must explicitly be denoted, for brevity, we use

\[
\begin{align*}
\mathbf{s}_i^k(\alpha_j, \alpha_\ell, \alpha_{\ell+1}) := (x_j^*(\alpha_j, \alpha_\ell, \alpha_{\ell+1}), y_j^*(\alpha_j, \alpha_\ell, \alpha_{\ell+1})), & \quad \mathbf{s}_k^j(\alpha_j, \alpha_\ell, \alpha_{\ell+1}) := (x_\ell^*(\alpha_j, \alpha_\ell, \alpha_{\ell+1}), y_\ell^*(\alpha_j, \alpha_\ell, \alpha_{\ell+1})).
\end{align*}
\]
Now notice that under Assumptions 2–4, from the KKT conditions for Problem $R_{j\ell}$, the value of the rate function $I_j(z_j^*(\alpha_j, \alpha_{\ell+1})) > 0$ at optimality in Problem $R_{j\ell}$. This result is stated formally in Lemma 3.5; we omit the proof.

**Lemma 3.5.** If $\alpha_j > 0, \alpha_{\ell} > 0$, and $\alpha_{\ell+1} > 0$, then $I_j(z_j^*(\alpha_j, \alpha_{\ell+1})) > 0$ in Problem $R_{j\ell}$ for all non-Pareto systems $j \in \mathcal{P}$ and all $\ell \in \{0, \ldots, p\}$.

We include a lemma regarding the locations of the optimal solutions in Problem $R_{j\ell}$ in Online Appendix D.

### 4. Limiting Approximation to the Optimal Allocation

Since Problem $Q$ is a bi-level optimization problem, it may take some time to solve for the optimal allocation using Problem $Q$ when the number of systems is large. While the computational time could be reduced by solving the inner problems in parallel, we believe it is useful to see if the optimal allocation can be simplified for large problem instances. In this section, we send the number of non-Pareto systems to infinity while keeping the number of Pareto systems finite and equal to $p$, which enables us to write the relative allocations between the non-Pareto systems in closed form. We remind the reader that, like much of the R&S literature, our emphasis is on the case in which the underlying distributions are normal. Thus we make a normality assumption in Section 4.1. This assumption simplifies the proofs and assists our intuition regarding dependence.

#### 4.1. Preliminaries and Assumptions for the Limiting Allocation

Recall that $r = p + |\mathcal{P}|$ is the total number of systems, and only $|\mathcal{P}|$ will tend to infinity while $p$ remains constant. We make three assumptions on the way non-Pareto systems are added to ensure a meaningful limiting allocation.

**Assumption 5.** The mean values $(g_j, h_j)$ satisfy $\inf\{|h_j - h_i| : i \in \mathcal{P}\} > \epsilon$ and $\inf\{|g_j - g_i| : i \in \mathcal{P}\} > \epsilon$ for some $\epsilon > 0$ and all $j \in \mathcal{P}$.

**Assumption 6.** There exists a compact set $\mathcal{C}_1 \subset \mathbb{R}^2$ such that $(g_k, h_k) \in \mathcal{C}_1$ for all $k = 1, \ldots, r$, and such that $\mathcal{C}_1 \subset \mathcal{P}$, (See Assumption 4 for notation.)

**Remark 4.1.** Given that systems are added to the compact set $\mathcal{C}_1$, since all the rate functions are strictly convex with a unique minimum at the location of the mean, there exists an “extended” compact set $\overline{\mathcal{C}_1}$ that also contains the locations of the solutions to all Problems $R_{j\ell}$. Let $\beta$ be the diameter of a circle that covers $\mathcal{C}_1$.

**Assumption 7.** The random vector $(G_k, H_k)$ is a bivariate normal random vector with parameters $(g_k, h_k, \sigma_{g_k}^2, \sigma_{h_k}^2, \rho_k)$ for all $k = 1, \ldots, r$. The corresponding rate function for $(\bar{G}_k(n), \bar{H}_k(n))$ is of the form $I_k(x, y) = \frac{1}{2(1-\rho_k^2)}(\frac{(x-g_k)^2}{\sigma_{g_k}^2} - 2\rho_k(x-g_k)(y-h_k) + \frac{(y-h_k)^2}{\sigma_{h_k}^2})$. We further assume there exist constants $\sigma_{a}^2, \sigma_{b}^2$, and $\rho_0$ such that $0 < \sigma_{a}^2 \leq \sigma_{g_k}^2 \leq \sigma_{b}^2 < \infty$, $0 < \sigma_{a}^2 \leq \sigma_{h_k}^2 \leq \sigma_{b}^2 < \infty$, and $|\rho_k| \leq \rho_0$ where $0.5 < \rho_0 < 1$, for all $k = 1, \ldots, r$.

Assumption 5 ensures the systems do not systematically approach the Pareto front. Assumption 6 ensures that the systems that are added continue to “compete” with the Pareto systems and do not become irrelevant in the limit. Assumptions 5 and 6 are analogous to assumptions in Pasupathy et al. [2014]. We differ from Pasupathy et al. [2014] in Assumption 7. While Pasupathy et al. [2014] assume the rate functions have upper and lower bounding quadratics on a compact set (a mild assumption), we simplify the analysis by assuming, like much of the R&S literature, that the rate functions are normal. We conjecture that the following analysis holds in the more general case.
of bounding quadratics, but we do not show it. Note that we require that the systems be added to $\mathcal{C}$ in such a way that their corresponding rate functions cannot become too shallow (less than $\sigma_j^2$) or too steep (larger than $\sigma_j^2$), and so that they cannot degenerate to a single dimension ($|\rho_k| = 1$). Assumption 7 subsumes Assumptions 3 and 4.

We write $R_j(\alpha_j, \alpha_\ell, \alpha_{\ell+1})$ under Assumption 7 in Proposition 4.2. For brevity, define the indicators $[g_{j\ell}] := [x_\ell > 0, \ell \neq p]$ and $[\sigma_{j\ell}] := [x_\ell > 0, \ell \neq 0]$ at optimality in Problem $R_{j\ell}$. Intuitively, $[g_{j\ell}] > 0$ means that non-Pareto $j$ “plays” with phantom Pareto $\ell$ on the $q$ objective via the Pareto $\ell + 1$, and $[\sigma_{j\ell}] > 0$ means that non-Pareto $j$ “plays” with phantom Pareto $\ell$ on the $h$ objective via Pareto $\ell$. Notice that the expression for $R_j(\alpha_j, \alpha_\ell, \alpha_{\ell+1})$ simplifies to one of three cases: the one-dimensional rate corresponding to system $j$ “beating” Pareto $\ell + 1$ on objective $g$, the one-dimensional rate corresponding to system $j$ “beating” Pareto $\ell$ on objective $h$, or a bivariate rate of $j$ dominating the phantom Pareto $\ell$. Since $\tilde{G}_{\ell+1}$ and $\tilde{H}$ are independent, only $\rho_j$ appears in the rate.

**Proposition 4.2.** Under Assumption 7, if $\alpha_j > 0, \alpha_\ell > 0, \alpha_{\ell+1} > 0$,

1. the rate function $R_j(\alpha_j, \alpha_\ell, \alpha_{\ell+1})$ equals

   \[
   \left[\frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2}ight] (g_j - g_{\ell+1})^2 g_{j\ell} - 2 \rho_j \frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2} \frac{\alpha_j}{\alpha_j^2 + \alpha_\ell^2} \left(\alpha_j^2 + \alpha_\ell^2\right) [g_{j\ell}] (g_j - g_{\ell+1}) (h_j - h_{\ell+1}) |[\sigma_{j\ell}]| + \left[\frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2}ight] (h_j - h_{\ell+1})^2 [\sigma_{j\ell}] \\
   2 \left[\left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2}\right) \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) - \rho_j^2 \frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2} [g_{j\ell}] [\sigma_{j\ell}]\right],
   \]

   \[
   [g_{j\ell}] > 0, [\sigma_{j\ell}] = 0 \text{ iff } \ell \neq p, g_j > g_{\ell+1}, h_j \leq h_{\ell+1} + \frac{(g_j - g_{\ell+1}) \rho_j \sigma_{j\ell}}{\alpha_j + \alpha_{\ell+1}},
   \]

   \[
   [g_{j\ell}] = 0, [\sigma_{j\ell}] > 0 \text{ iff } \ell \neq 0, h_j > h_{\ell+1}, g_j \leq g_{\ell+1} + \frac{(h_{\ell+1} - h_j) \sigma_{j\ell}}{\alpha_j + \alpha_{\ell+1}},
   \]

   \[
   [g_{j\ell}] > 0, [\sigma_{j\ell}] > 0 \text{ iff } \ell \notin \{0, p\}, g_j > g_{\ell+1} + \frac{(h_j - h_{\ell+1}) \sigma_{j\ell}}{\alpha_j + \alpha_{\ell+1}},
   \]

   \[
   h_j > h_{\ell+1} + \frac{(g_j - g_{\ell+1}) \rho_j \sigma_{j\ell}}{\alpha_j + \alpha_{\ell+1}}.
   \]

2. the rate functions corresponding to systems $j, \ell$, and $\ell + 1$ are

   \[
   I_j(x_j^*, y_j^*) = \frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2} \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2}\right)^2 - \rho_j^2 \frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2} \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) [g_{j\ell}] (g_j - g_{\ell+1})^2 [\sigma_{j\ell}],
   \]

   \[
   - 2 \rho_j^2 \frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2} \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) \left(1 - \rho_j^2 \frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) [g_{j\ell}] (g_j - g_{\ell+1}) (h_j - h_{\ell+1}) [g_{j\ell}] [\sigma_{j\ell}],
   \]

   \[
   + \frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2} \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right)^2 - \rho_j^2 \frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2} \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) [g_{j\ell}] (h_j - h_{\ell+1}) [\sigma_{j\ell}] [\sigma_{j\ell}] \left(2 \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) - \rho_j^2 \frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2} [g_{j\ell}] [\sigma_{j\ell}] [\sigma_{j\ell}] \right) \right]^{-1},
   \]

   \[
   I_\ell(x_\ell^*, y_\ell^*) = \frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2} \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_\ell^2}\right)^2 \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) (h_j - h_{\ell+1})^2 [\sigma_{j\ell}] [\sigma_{j\ell}] \left[\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right] [g_{j\ell}] [\sigma_{j\ell}] [\sigma_{j\ell}] \left(2 \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) - \rho_j^2 \frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2} [g_{j\ell}] [\sigma_{j\ell}] [\sigma_{j\ell}] \right) \right]^{-1},
   \]

   \[
   I_{\ell+1}(x_{\ell+1}^*, y_{\ell+1}^*) = \frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2} \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right)^2 \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) (g_j - g_{\ell+1})^2 [\sigma_{j\ell}] [\sigma_{j\ell}] \left[\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right] [g_{j\ell}] [\sigma_{j\ell}] [\sigma_{j\ell}] \left(2 \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) \left(\frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2}\right) - \rho_j^2 \frac{\alpha_j^2}{\alpha_j^2 + \alpha_{\ell+1}^2} [g_{j\ell}] [\sigma_{j\ell}] [\sigma_{j\ell}] \right) \right]^{-1}.
   \]

While the value of $I_j(x_j^*(\alpha_j, \alpha_\ell, \alpha_{\ell+1}))$ is always strictly positive at optimality by Lemma 3.5, it may be that $I_j(x_j^*(\alpha_j, \alpha_\ell, \alpha_{\ell+1})) = 0$ or $I_{\ell+1}(x_{\ell+1}^*(\alpha_j, \alpha_\ell, \alpha_{\ell+1})) = 0$, in which case we say that system $\ell$ or system $\ell + 1$ “falls out” of the rate function in Problem $R_{j\ell}$, respectively. This fact raises the possibility that a particular Pareto system $\ell$ “falls out” of the rate function in both Problem $R_{j\ell-1}$ and Problem $R_{j\ell}$, in which
case the non-Pareto system $j$ does not “play” with the Pareto $\ell$ at all. The following Lemma 4.3 states that such a case is impossible.

**Lemma 4.3.** Under Assumption 7, if $\alpha_j > 0$, $\alpha_{\ell-1} > 0$, $\alpha_\ell > 0$, $\alpha_{\ell+1} > 0$, then $\max \{I_{\ell}(z_{\ell}^*(\alpha_j, \alpha_{\ell-1}, \alpha_\ell)), I_{\ell}(z_{\ell}^*(\alpha_j, \alpha_\ell, \alpha_{\ell+1}))\} > 0$ for all non-Pareto systems $j \in P^c$ and Pareto systems $\ell \in \{1, \ldots, p\}$.

### 4.2. Allocation to Non-Pareto Systems

Since we send the number of non-Pareto systems to infinity, we relax the constraints in Problem $Q$ that pertain only to Pareto systems and MCE events. Thus in this section, we concern ourselves not with Problem $Q$, but with its relaxation:

Problem $\tilde{Q}$: maximize $\tilde{z}$ s.t.

\[
R_j(\tilde{\alpha}_j, \tilde{\alpha}_\ell, \tilde{\alpha}_{\ell+1}) \geq \tilde{z} \quad \text{for all } j \in P^c, \ell = 0, \ldots, p, \\
\sum_{j=1}^p \tilde{\alpha}_j = 1, \quad \tilde{\alpha}_i \geq 0 \quad \text{for all } i \leq r.
\]

The KKT conditions are necessary and sufficient for global optimality in Problem $\tilde{Q}$. We first use Problem $\tilde{Q}$ to derive insights on the optimal allocation as the number of non-Pareto systems tends to infinity. In §4.4, we show that under mild conditions, for a large enough set of non-Pareto systems, the solutions to Problems $Q$ and $\tilde{Q}$ are equal.

Since they play a prominent role in the results that follow, we present the KKT conditions for Problem $\tilde{Q}$ in Theorem 4.4. In Theorem 4.4, Parts (1) and (2) essentially ensure the existence of a binding constraint in Problem $\tilde{Q}$ for each non-Pareto system $j$ and Pareto system $\ell$, respectively. The result in equation (5) determines the relative allocation between a Pareto system $\ell$ and the non-Pareto systems $j$ that “play” with it. The result in (6) determines the relative allocations between the Pareto systems. The result in (7) determines the relative allocations between the non-Pareto systems.

**Theorem 4.4.** Let $\lambda_{j\ell} \geq 0$ for all $j \in P^c$ and $\ell = 0, \ldots, p$ be dual variables associated with Problem $\tilde{Q}$. Then at optimality in Problem $\tilde{Q}$, $\tilde{\alpha}_j > 0$ for all $i \leq r$ and

1. For all non-Pareto systems $j \in P^c$, there exists a phantom Pareto system $\ell^*(j) \in \{0, \ldots, p\}$ such that $\lambda_{j\ell^*(j)} > 0$, which implies $\tilde{z}^* = R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*) > 0$;
2. For all Pareto systems $\ell \in \{1, \ldots, p\}$, there exists a non-Pareto system $j^*(\ell) \in P^c$ such that $\lambda_{j^*(\ell)\ell} I_\ell(z_{\ell}^*(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*)) > 0$, which implies $\tilde{z}^* = \min \{R_{j^*(\ell)}(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*), R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*)\} > 0$;
3. For all Pareto systems $\ell, \ell' \in \{1, \ldots, p\}$,

\[
\sum_{j \in P^c} \frac{\lambda_{j\ell-1} I_\ell(z_{\ell}^*(\tilde{\alpha}_j^*, \tilde{\alpha}_{\ell-1}^*, \tilde{\alpha}_\ell^*))) + \lambda_{j\ell} I_\ell(z_{\ell}^*(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*))}{\sum_{\ell'=0}^\ell \lambda_{j\ell'} I_{\ell'}(z_{\ell'}^*(\tilde{\alpha}_j^*, \tilde{\alpha}_{\ell'}^*, \tilde{\alpha}_{\ell'+1}^*))} = 1;
\]

\[
\sum_{j \in P^c} \lambda_{j\ell-1} I_j(z_{j}(\tilde{\alpha}_j^*, \tilde{\alpha}_{j-1}^*, \tilde{\alpha}_j^*)) + \lambda_{j\ell} I_j(z_{j}(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*)) = 1.
\]

4. For all non-Pareto systems $j, j' \in P^c$,

\[
\sum_{\ell=0}^p \frac{\lambda_{j\ell} I_j(z_{j}(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*))}{\sum_{\ell=0}^\ell \lambda_{j\ell} I_{j'}(z_{j'}(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*))} = 1.
\]

**Proof.** In addition to $\lambda_{j\ell} \geq 0$ for all $j \in P^c, \ell = 0, \ldots, p$, let $\nu$ be a dual variable. Then we have the complementary slackness conditions $\lambda_{j\ell}(R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*) - \tilde{z}^*) = 0$. 


for all $j \in P^c$, $\ell = 0, \ldots, p$; and the (simplified) stationarity conditions

$$
\sum_{j \in P^c} \{\lambda_{j\ell-1} I_j(\bar{q}_{\alpha_j}, \bar{q}_{\alpha_{\ell-1}}, \bar{q}_{\alpha_{\ell+1}}) + \lambda_{j\ell} I_j(\bar{q}_{\alpha_j}, \bar{q}_{\alpha_{\ell}}, \bar{q}_{\alpha_{\ell+1}})\} = \nu \quad \forall \ell = 1, \ldots, p; \quad (8)
$$

$$
\sum_{\ell=0}^p \lambda_{j\ell} I_j(\bar{q}_{\alpha_j}, \bar{q}_{\alpha_{\ell}}, \bar{q}_{\alpha_{\ell+1}}) = \nu \quad \forall j \in P^c; \quad (9)
$$

$$
\sum_{j \in P^c} \sum_{\ell=0}^p \lambda_{j\ell} = 1. \quad (10)
$$

See Online Appendix G for a complete proof. □

Observe that as the number of non-Pareto systems added according to Assumptions 5–7 grows, the overall rate of decay of $\text{MCI}_{P^c}$ in Problem $Q$ will decrease. If this fact is not intuitive, it can be seen by noticing that adding non-Pareto systems that compete with the Pareto set in a non-trivial way implies that we are adding binding constraints to Problem $Q$, which necessarily decreases its optimal value. Thus under our assumptions, as $|P^c| \to \infty$, $\tilde{z}^* \to 0$. (Notice that now, quantities such as $\tilde{z}^*$, $\lambda_{j\ell}$ for all $j \in P^c$, $\ell \in \{0, \ldots, p\}$ and $\bar{q}_k$ for all $k \leq r$ are also functions of $r$ and could be denoted as $\tilde{z}^*(r)$, $\lambda_{j\ell}(r)$ and $\bar{q}_k(r)$, respectively. We often suppress this notation unless it is helpful for clarity.)

The following Proposition 4.6 states the rate at which $\tilde{z}^* \to 0$. Before we state the proposition, we present Lemma 4.5, which provides useful bounds on $R_j(\bar{q}_j, \bar{q}_\ell, \bar{q}_{\ell+1})$ and rate functions that comprise $R_j(\bar{q}_j, \bar{q}_\ell, \bar{q}_{\ell+1})$. A proof for Proposition 4.6 follows from Lemma 4.5 and appears in Online Appendix I.

**Lemma 4.5.** Under Assumptions 5–7, the following bounds hold for each $j \in P^c$ and each $\ell \in \{0, \ldots, p\}$:

$$
\bar{q}_j \left(\frac{\epsilon^2 \sigma_j^2 (1-p_0)}{2 \sigma} \right) \left[\frac{I_{[0]}(\alpha_j)}{1+\alpha_j/\alpha_{\ell+1}}\right] + \frac{I_{[\ell]}(\alpha_j)}{1+\alpha_j/\alpha_{\ell+1}} \leq R_j(\bar{q}_j, \bar{q}_\ell, \bar{q}_{\ell+1}) \leq R_j(\bar{q}_j, \bar{q}_\ell, \bar{q}_{\ell+1}) \quad (11)
$$

$$
\frac{\epsilon^2 \sigma_j^2 (1-p_0)^2}{2 \sigma} \left[\frac{I_{[0]}(\alpha-j)}{1+\alpha_j/\alpha_{\ell+1}}\right] - \frac{2 p_0 I_{[0]}(\alpha_j)}{1+\alpha_j/\alpha_{\ell+1}} \leq I_j(\bar{q}_j, \alpha_j, \alpha_{\ell+1}) \quad (12)
$$

$$
\frac{\epsilon^2 \sigma_j^2 (1-p_0)^2}{2 \sigma} \left[\frac{I_{[0]}(\alpha-j)}{1+\alpha_j/\alpha_{\ell+1}}\right] \leq I_j(\bar{q}_j, \alpha_j, \alpha_{\ell+1}) \quad (13)
$$

**Proposition 4.6.** Under Assumptions 5–7, as $|P^c| \to \infty$, $\tilde{z}^* = O(1/|P^c|)$.

For each Pareto system $\ell \in \{1, \ldots, p\}$ and for each value of $r$, let $P^c(\ell, r)$ denote the set of non-Pareto systems that have a binding constraint with system $\ell$ in Problem $Q(r)$, where in this case, we explicitly denote the dependence of Problem $Q$ on the number of systems $r$. That is, for all $\ell \in \{1, \ldots, p\}$ and all $r$, define

$$
P^c(\ell, r) := \{j \in P^c : \lambda_{j\ell-1} I_j(\bar{q}_{\alpha_j}, \bar{q}_{\alpha_{\ell-1}}, \bar{q}_{\alpha_{\ell+1}}) + \lambda_{j\ell} I_j(\bar{q}_{\alpha_j}, \bar{q}_{\alpha_{\ell}}, \bar{q}_{\alpha_{\ell+1}}) > 0\}.
$$

If $j \in P^c(\ell, r)$, we say that non-Pareto system $j$ “binds with” Pareto system $\ell$ in Problem $Q(r)$. By Theorem 4.4, $|P^c(\ell, r)| \geq 1$ for all $\ell \in \{1, \ldots, p\}$ and all $r$. Further, by Lemma 4.3, since the rate function for Pareto system $\ell$ cannot “drop out” of both $R_j(\bar{q}_j, \bar{q}_{\ell-1}, \bar{q}_\ell)$ and $R_j(\bar{q}_j, \bar{q}_\ell, \bar{q}_{\ell+1})$, it follows that for all $r$, all $j \in P^c$, and all $\ell \in \{1, \ldots, p\}$, we have $I_{[\ell]}(\alpha_{j}) + I_{[\ell]}(\alpha_{j}) > 0$. If we also have $j \in P^c(\ell, r)$, then $\lambda_{j\ell-1} I_{[\ell]}(\alpha_{j}) + \lambda_{j\ell} I_{[\ell]}(\alpha_{j}) > 0$. This result follows because for $j$ to bind with $\ell$, we require it not to “fall out” of the binding constraint in Problem $Q(r)$.
Using the notion of \(\mathcal{P}_c(\ell, r)\), for a particular value of \(r\) in the sequence, Lemma 4.7 provides relationships between the key quantities in Problem \(\mathcal{Q}(r)\) and the KKT conditions in Theorem 4.4.

**Lemma 4.7.** Under Assumptions 5–7, the following statements hold for each \(r < \infty\) and its corresponding Problem \(\mathcal{Q}(r)\):

1. There exist \(\kappa_{1a} > 0, \kappa_{1b} < \infty\) such that \(\kappa_{1a}/z^* \leq 1/\tilde{\alpha}_j^* + 1/\tilde{\alpha}_\ell^* \leq \kappa_{1b}/z^*\) for all \(j \in \mathcal{P}_c(\ell, r), \ell \in \{1, \ldots, p\}\).
2. If \(j \in \mathcal{P}_c(\ell, r) \cap \mathcal{P}_c(\ell', r)\), there exists \(\kappa_2 < \infty\) such that \((1 + \tilde{\alpha}_j^*/\tilde{\alpha}_\ell^*) < \kappa_2(1 + \tilde{\alpha}_{j'}^*/\tilde{\alpha}_{\ell'}^*)\) for all \(\ell, \ell' \in \{1, \ldots, p\}\).
3. There exists \(\kappa_3 > 0\) such that \(\tilde{\alpha}_j^2/\tilde{\alpha}_\ell^2 \geq \kappa_3(\sum_{\ell=0}^p \lambda_{j\ell})/(\sum_{\ell=0}^p \lambda_{j'\ell})\) for all \(j, j' \in \mathcal{P}_c\).
4. There exists \(\kappa_4 > 0\) such that, for all \(\ell, \ell' \in \{1, \ldots, p\}\),
   \[
   \frac{\tilde{\alpha}_\ell^*}{\tilde{\alpha}_{\ell'}^*} \geq \frac{\sum_{j \in \mathcal{P}_c(\ell, r)} \lambda_{j\ell-1}(l_{j\ell-1} + \lambda_{j\ell}l_{j\ell})}{\sum_{j' \in \mathcal{P}_c(\ell', r)} \lambda_{j'\ell-1}(l_{j'\ell-1} + \lambda_{j'\ell}l_{j'\ell})}.
   \]
5. There exist \(\kappa_{5a} > 0, \kappa_{5b} < \infty\) such that, for all \(\ell \in \{1, \ldots, p\}\) and all \(j \in \mathcal{P}_c\),
   \[
   \sum_{j \in \mathcal{P}_c(\ell, r)} \lambda_{j\ell-1}(l_{j\ell-1} + \lambda_{j\ell}l_{j\ell}) \leq \frac{\tilde{\alpha}_j^2}{\tilde{\alpha}_\ell^2} \leq \frac{\sum_{j \in \mathcal{P}_c(\ell, r)} \lambda_{j\ell-1}(l_{j\ell-1} + \lambda_{j\ell}l_{j\ell})}{\sum_{j' \in \mathcal{P}_c(\ell, r)} \lambda_{j'\ell-1}(l_{j'\ell-1} + \lambda_{j'\ell}l_{j'\ell})}.
   \]
6. There exist \(\kappa_{6a} > 0, \kappa_{6b} < \infty\) such that \(\kappa_{6a} \sum_{\ell=0}^p \lambda_{j\ell} \leq \tilde{\alpha}_j^2 / \sum_{\ell=1}^p \tilde{\alpha}_\ell^2 \leq \kappa_{6b} \sum_{\ell=0}^p \lambda_{j\ell}\) for all \(\ell \in \{1, \ldots, p\}\), \(j \in \mathcal{P}_c\).

In Lemma 4.7, we have stated the relationships between key quantities, such as \(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*,\) and \(\lambda_{j\ell}\) for all \(j \in \mathcal{P}_c, \ell \in \{1, \ldots, p\}\), based on whether the non-Pareto \(j\) binds with the Pareto \(\ell\) for a particular \(r\). To make stronger statements about the limiting relative allocations, we now require notions of the limiting allegiances between non-Pareto and Pareto as the number of non-Pareto tends to infinity.

First, we require that there exists a ”primary” binding Pareto for each non-Pareto system. Specifically, when a new non-Pareto \(j\) enters Problem \(\mathcal{Q}(r)\), it will bind with one or more Pareto systems, via a phantom Pareto system. In the sequence of problems indexed by \(r\), we assume that there exists a value \(r_{j0}\) such that for all \(r \geq r_{j0}\), the non-Pareto \(j\) binds with the phantom Pareto \(\ell^*\). This \(\lambda_{j\ell^*}(r) > 0\) for all \(r \geq r_{j0}\), and then \(j \in \mathcal{P}_c(\ell^*, r)\), or \(j \in \mathcal{P}_c(\ell^*+1, r)\) for all \(r \geq r_{j0}\). We further assume that for all other phantom Pareto systems \(\ell\) that bind with \(j\), there exists a constant \(\kappa_0 < \infty\) such that \(\lambda_{j\ell}(r) \leq \kappa_0 \lambda_{j\ell^*}(r)\) for all \(r \geq r_{j0}\). Using the shadow price interpretation of \(\lambda_{j\ell}\)'s, this assumption implies that the greatest gain to the rate \(z^*\) will be achieved by perturbing the constraint associated with the rate for \(j\) and \(\ell^*\), to within a constant. If \(j\) binds with \(\ell^*\) in this way, we say that \(\ell^*\) is the primary phantom Pareto for non-Pareto system \(j\). Of course \(\ell^*\) is a function of \(j\), again for simplicity, we omit this notation.

While this assumption may feel somewhat artificial, we believe that it will arise naturally, for example, when non-Pareto systems are added according to a uniform distribution (provided the requirements of our assumptions are maintained). To understand why, consider what it means for one non-Pareto system \(j\) to bind with more than one phantom Pareto system. In a scenario with multiple Pareto and only one non-Pareto \(j_1\), the non-Pareto \(j_2\) will bind with all of the phantom Pareto due to Theorem 4.4. However as new non-Pareto are added uniformly across \(\ell\), the new non-Pareto bind with the phantom Pareto “closest” to them, and \(j_1\) will cease to bind with phantom Pareto that are “far away” from it — those phantom Pareto will bind with other, closer, non-Paretos. Therefore intuitively, non-Pareto binding with multiple phantom Pareto arises when there are not very many non-Pareto, or when the non-Pareto are not “evenly distributed,” as might arise when all non-Pareto are uniquely dominated.
by the same Pareto. Therefore we anticipate that the number of non-Pareto systems with multiple phantom Paretos decreases as non-Pareto systems are added “evenly.”

In addition to assuming each non-Pareto has a primary phantom Pareto, we also require that the cardinality of the set of non-Pareto systems binding with each Pareto increase to infinity. Specifically, for all $\ell \in \{1, \ldots, p\}$, define $P^c(\ell)$ as

$$P^c(\ell) := \{j \in P^c : \ell^* \in \{\ell - 1, \ell\}, \lim_{r \to \infty} \|\lambda_{j^*+1}(r)\| > 0, \|\lambda_{j^*}(r)\| > 0, \|\lambda_{j^*}(r)\| \geq 0, \|\lambda_{j^*+1}(r)\| > 0\}.$$  

Intuitively, $P^c(\ell)$ is the set of non-Pareto systems that repeatedly bind with system $\ell$ in Problem $\tilde{Q}(r)$ via phantom Pareto $\ell^* \in \{\ell - 1, \ell\}$ as $r \to \infty$. Notice that $\|P^c(\ell, r)\| \to \infty$ does not imply that $\|P^c(\ell, r)\| \to \infty$, since we can imagine pathological cases in which new systems are added to $P^c(\ell, r)$, but later become binding with a different Pareto system in high enough numbers that $\|P^c(\ell, r)\|$ remains bounded. However, $\|P^c(\ell, r)\| \to \infty$ implies $\|P^c(\ell, r)\| \to \infty$. Therefore in what follows, we send $\|P^c(\ell, r)\| \to \infty$ for all $\ell \in \{1, \ldots, p\}$. This condition ensures that the number of non-Pareto systems that bind with each Pareto system in Problem $\tilde{Q}(r)$ goes to infinity as $r$ grows. To ensure “evenness” of non-Pareto systems, we require that the cardinality of each set $P^c(\ell)$ remain within a constant of the total number of non-Pareto systems.

We formalize these assumptions below, where $\{r\}$ denotes the sequence of the total number of competing systems. We numerically evaluate such a regime in §4.5.

**Assumption 8.** We assume the following: (1) For each $j \in P^c$, there exists a phantom Pareto system $\ell^*$ and constants $r_{j0}, \kappa_0 < \infty$ such that $\lambda_{j_\ell} > 0$ for all $\ell \geq r_{j0}$ and $\lambda_{j_\ell} \leq \kappa_0 \lambda_{j_\ell}$ for all $\ell \in \{0, \ldots, p\}$, $r \geq r_{j0}$. (2) There exists $k > 0$ such that $\|P^c(\ell)\| \geq k\|P^c\|$ for all $\ell \in \{1, \ldots, p\}$ and all $r$.

Under the regularity conditions in Assumption 8, we ensure that each system receives nonzero sample in the limit. Thus we extend the results from Lemma 4.7 to the stronger results on the relative allocations in the following Theorem 4.8.

**Theorem 4.8.** Let $\ell^*$ be the primary phantom Pareto for non-Pareto system $j$. Under Assumptions 5–8, the following statements hold:

1. Let $j \in P^c$. There exists $\tau_1 > 0$ and $r_{j1}$ such that $\lambda_{j_\ell}^2 / \lambda_{j_\ell} \geq \tau_1 \|\lambda_{j_\ell+1}\|$ and $\lambda_{j_{\ell+1}}^2 / \lambda_{j_{\ell+1}} \geq \tau_1 \|\lambda_{j_\ell+1}\|$ for all $\ell \geq r_{j1}$.  
2. Let $j \in P^c$. For all $\ell \geq r_{j1}$, there exists $\tau_2 > 0$ such that $I_j(\lambda_{j_\ell}^*/\lambda_{j_\ell}^*) \geq \tau_2$.  
3. Let $j, j' \in P^c$. There exists $\tau_3 < \infty$ and $r_{j,j'}$, such that $\lambda_{j_\ell}^*/\lambda_{j_\ell}^* \leq \tau_3$ for all $\ell \geq r_{j,j'}$.  
4. As $\|P^c(\ell)\| \to \infty$ for all $\ell \in \{1, \ldots, p\}$, $\lambda_{j_\ell}^* \Rightarrow \lambda_{j_\ell} \in \{1/\|P^c\|\}$ for all $j \in P^c$.  
5. There exists $\tau_5 < \infty$ and $\tau_5$ such that $\lambda_{j_\ell}^*/\lambda_{j_\ell}^* \leq \tau_5$ for all $\ell, \ell' \in \{1, \ldots, p\}$ and all $\ell \geq r_{j_\ell}$.

6. As $\|P^c(\ell)\| \to \infty$ for all $\ell \in \{1, \ldots, p\}$, $\lambda_{j_\ell}^*/\lambda_{j_\ell}^* \to 0$ for all $j \in P^c$ and all $\ell \in \{1, \ldots, p\}$.  
7. As $\|P^c(\ell)\| \to \infty$ for all $\ell \in \{1, \ldots, p\}$, $I_\ell(\lambda_{j_\ell}(\alpha_j, \alpha_\ell, \alpha_{\ell+1})) = O(\lambda_{j_\ell}^*/\lambda_{j_\ell}^*)$ and $I_{\ell+1}(\lambda_{j_\ell+1}(\alpha_j, \alpha_\ell, \alpha_{\ell+1})) = O(\lambda_{j_\ell}^*/\lambda_{j_\ell}^*)$ in Problem $R_{j_\ell}$ for all $j \in P^c$, $\ell \in \{0, \ldots, p\}$.  
8. As $\|P^c(\ell)\| \to \infty$ for all $\ell \in \{1, \ldots, p\}$, $(x_{j_\ell}^*, y_{j_\ell}^*) \to (g_\ell, h_\ell)$ if $\ell \neq 0$ and $(x_{j_{\ell+1}}^*, y_{j_{\ell+1}}^*) \to (g_{\ell+1}, h_{\ell+1})$ if $\ell \neq 0$ in Problem $R_{j_\ell}$ for all $j \in P^c$, $\ell \in \{0, \ldots, p\}$.

The primary results in Theorem 4.8 appear in Parts (4), (6), and (8). Part (4) essentially states that, as we add more and more non-Pareto systems that compete with each Pareto system, then the allocation to each non-Pareto system decreases to zero. Because each Pareto system is now competing with a large number of non-Pareto, Part (6) states that in the limit, the Pareto systems will receive much more sample than non-Pareto systems. Finally, Part (8) states that in the limit in Problem $R_{j_\ell}$, the rate functions corresponding to both Pareto and non-Pareto systems tend to zero, while we know by
Part (2) that the rate function corresponding to system $j$ remains positive. Thus in the limit, the rate function corresponding to $j$ is evaluated over the cone in which $j$ would dominate $\ell$, $x_j \leq g_{\ell+1}$, $y_j \leq h_\ell$. Loosely speaking, the Pareto systems receive enough sample in this asymptotic regime that, relative to the non-Paretos, the Pareto frontier appears "fixed."

This last result leads us directly to the main result of the paper, presented in the following Theorem 4.9. We do not provide a proof; notice that it follows by dividing the rate function through by $\hat{\alpha}_j^*$ and applying Theorem 4.8 Parts (7) and (8).

**Theorem 4.9.** **Under Assumptions 5–8, as $|P^c(\ell)| \to \infty$ for all $\ell \in \{1, \ldots, p\}$,**

$$R_j(\hat{\alpha}_j^*, \hat{\alpha}_j^*, \hat{\alpha}_{\ell+1}^*)/\hat{\alpha}_j^* = \inf_{x_j \leq g_{\ell+1}, y_j \leq h_\ell} I_j(x_j, y_j) \quad \text{for all } j \in P^c, \ell = 0, \ldots, p.$$  

To see the implications of Theorem 4.9, define the score $S_j$ as

$$S_j := \min_{\ell \in \{0, \ldots, p\}} \{ \inf_{x_j \leq g_{\ell+1}, y_j \leq h_\ell} I_j(x_j, y_j) \} \quad \text{for all } j \in P^c.$$  

Then it follows that for all $j \in P^c$, $\tilde{z}^* = \min_{\ell \in \{0, \ldots, p\}} R_j(\hat{\alpha}_j^*, \hat{\alpha}_j^*, \hat{\alpha}_{\ell+1}^*)$, and we have

$$\tilde{z}^*/\hat{\alpha}_j^* = \min_{\ell \in \{0, \ldots, p\}} \{ \inf_{x_j \leq g_{\ell+1}, y_j \leq h_\ell} I_j(x_j, y_j) \} \quad \text{for all } j \in P^c.$$  

Therefore the relative allocations between the non-Pareto systems are determined by the score, which is written formally in the following Theorem 4.10.

**Theorem 4.10.** **Under Assumptions 5–8, as $|P^c(\ell)| \to \infty$ for all $\ell \in \{1, \ldots, p\}$,** then

$$\frac{\hat{\alpha}_j^*}{\hat{\alpha}_k^*} = \frac{S_k}{S_j} = \min_{\ell \in \{0, \ldots, p\}} \{ \inf_{x_k \leq g_{\ell+1}, y_k \leq h_\ell} I_k(x_k, y_k) \} \quad \text{for all } j, k \in P^c.$$  

Under Assumption 7, $\inf_{x_j \leq g_{\ell+1}, y_j \leq h_\ell} I_j(x_j, y_j)$ is a quadratic program with box constraints. The following proposition presents its closed-form solution.

**Proposition 4.11.** **Define the univariate rate functions $I_{g_k}(x) := (x - g_k)^2/(2\sigma_{g_k}^2)$ and $I_{h_k}(y) := (y - h_k)^2/(2\sigma_{h_k}^2)$ for all $k \leq r$. Then under Assumption 7, for all non-Pareto systems $j \in P^c$, the score $S_j = \min_{\ell = 0, \ldots, p} S_j(\ell)$, where**

$$S_j(\ell) := \begin{cases} I_{g_j}(g_{\ell+1}) & \text{if } \ell \neq p, g_j > g_{\ell+1}, h_j \leq h_\ell + \rho_j \frac{\sigma_{h_j}}{\sigma_{g_j}} (g_j - g_{\ell+1}); \\ I_{h_j}(h_\ell) & \text{if } \ell \neq 0, h_j > h_\ell, g_j \leq g_{\ell+1} + \rho_j \frac{\sigma_{h_j}}{\sigma_{g_j}} (h_j - h_\ell); \\ I_j(g_{\ell+1}, h_\ell) & \text{if } \ell \notin \{0, p\}, g_j > g_{\ell+1} + \rho_j \frac{\sigma_{h_j}}{\sigma_{g_j}} (h_j - h_\ell), h_j > h_\ell + \rho_j \frac{\sigma_{h_j}}{\sigma_{g_j}} (g_j - g_{\ell+1}). \\ \end{cases}$$

Thus in our asymptotic regime, the relative allocations between the non-Pareto systems can be expressed in closed-form, where the allocation to a particular non-Pareto system is inversely proportional to its scaled “distance” from the Pareto set.

### 4.3. Allocation to Pareto Systems

While we are able to solve for the relative allocations between the non-Pareto systems, we also require a sense of how much to allocate to the Pareto systems. The following Theorem 4.12 states that the allocations to the Pareto systems also tend to zero, albeit at a slower rate than the non-Pareto systems.

**Theorem 4.12.** **Under Assumptions 5–8, as $|P^c(\ell)| \to \infty$ for all $\ell \in \{1, \ldots, p\}$,**

$$\hat{\alpha}_i^* = \Theta(1/\sqrt{|P^c|}) \quad \text{for all Pareto systems } i \in P.$$  

Theorem 4.12 only gives us a sense of the allocation to the Pareto systems in the limit; to implement an allocation, we require heuristics, discussed in Section 5.
4.4. Equivalence of Allocations When the Number of Non-Pareto Systems is Large

Recall that all results presented in §4.2 and §4.3 pertain to Problem $\tilde{Q}$ and not to the original characterization of the optimal allocation as the solution to Problem $Q$. The following Theorem 4.13 states that as the number of non-Pareto systems grows, the optimal allocation provided by Problem $\tilde{Q}$ is equal to that provided by Problem $Q$.

**Theorem 4.13.** Under Assumptions 5–8, for large enough $|P_c|$, $\tilde{\alpha}^* = \alpha^*$.

Intuitively, Theorem 4.13 holds because in the limiting regime, the Pareto systems receive enough sample that the rate functions corresponding to MCE events between Pareto systems cannot be the unique minimum in the overall rate of decay of the probability of MC.

4.5. Numerical Evaluation of the Limiting Regime

We have shown that under some regularity conditions, as the number of non-Pareto systems tends to infinity, the rate of decay of $P\{\text{MCE}\}$ becomes non-binding in Problem $Q$. To numerically evaluate this effect, we created a Pareto front containing five Pareto systems on a circle of radius six, placed at equally spaced angles. This spacing guarantees the minimum separation between Pareto systems on both objectives is greater than 0.5. (Thus $|g_i - g_k| > 0.5, |h_i - h_k| > 0.5$ for all $i,k \in P$.)

Fixing this Pareto front, we generated non-Pareto systems by one of two methods: uniform or normal. In the uniform method, non-Paretos were generated uniformly in a circle centered at $(100,100)$ with radius six. In the normal method, non-Pareto systems were generated according to an independent bivariate normal distribution with both means equal to 100 and both standard deviations equal to three. Thus the majority of systems were generated within six units of the mean. In both methods, non-Pareto systems less than 0.25 units away from the Pareto front were rejected — this condition ensures that Assumption 5 is satisfied, and that the rate is large enough for us to solve for the optimal allocation numerically by solving Problem $Q$. To give a sense of what these problem instances look like, Figures 2 and 5 show example problem instances in which 445 non-Pareto systems were added according to the uniform and normal methods, respectively. All systems were assumed to have bivariate normal rate functions under Assumption 7 with independent objectives and unit variance.

As non-Pareto systems were added to fifty problem instances of each type, uniform and normal, we solved Problem $Q$ for the optimal allocation. We then created two types of plots: Figures 3 and 6, which show the percent of problem instances with binding MCE constraints, and Figures 4 and 7, which show box plots of the percent of the dual variable values associated with MCE constraints at optimality in Problem $Q$. To better understand what we mean by the percent of dual variable values associated with MCE constraints, in Problem $Q$, suppose we have $\lambda^p_{ik}$ as the dual variables corresponding to MCE constraints and $\lambda^c_{j\ell}$ for all $j \in P_c, \ell \in \{0, \ldots, p\}$ as the dual variables corresponding to MCI constraints. Then the percent of dual variable values associated with MCE constraints is $\sum_{i \in P} \sum_k \sum_{i \neq k} \lambda^p_{ik} / (\sum_{i \in P} \sum_k \sum_{i \neq k} \lambda^p_{ik} + \sum_{\ell=0}^p \sum_{j \in P_c} \lambda^c_{j\ell})$ at optimality.

While it could be argued that the problems we have generated are somewhat artificial, given the nicely spaced Pareto front and its “buffer” away from the non-Pareto systems, we believe there is an important message in Figures 2–7. If the systems in the Pareto front can be viewed as coming from a distribution, the distribution will affect the rate at which the limiting regime kicks in. When the systems are generated according to the normal method, they cluster together near the mean at $(100,100)$. Thus many of the non-Paretos are some distance away from the Pareto frontier, and the limit kicks in slower. When the systems are generated according to the uniform
100% with MCE Binding

100% of dual to MCE

G. Feldman and S. R. Hunter

5. THE SCORE ALLOCATION

We now describe a heuristic allocation for implementation that is based on the theory presented in §4, and which we call the SCORE (Sampling Criteria for Optimization using Rate Estimators) allocation. As in the “original” SCORE allocation of Pasupathy et al. [2014], we use Theorem 4.10 to determine the relative allocations among the Pareto systems. We now describe the method by which we solve for the allocations to the Pareto systems.

First, notice that in the limiting version of Problem $\tilde{Q}$, there is exactly one binding constraint corresponding to MCI for each system $j$. Letting the phantom Pareto $\ell^* = \arg\min_{\ell \in \{0, \ldots, p\}} S_j(\ell)$, the constraints corresponding to $j$ dominating all phantom Paretos other than $\ell^*$ are non-binding, which implies the dual variables $\lambda_{j\ell^*} = 0$ for all $\ell = 0, \ldots, p, \ell \neq \ell^*$, and all $j \in \mathcal{P}^c$. Then for all non-Paretos $j, j' \in \mathcal{P}^c$, the KKT condition in (7) implies that as $|\mathcal{P}^c(\ell)| \to \infty$ for all $\ell \in \{1, \ldots, p\}$,

$$\frac{\lambda_{j\ell^*}}{\lambda_{j'\ell^*}} = \frac{I_j(\gamma_j^*)/\sum_{\ell \in \mathcal{P}^c} \alpha_j^*}{I_{j'}(\gamma_{j'}^*)/\sum_{\ell \in \mathcal{P}^c} \alpha_{j'}^*),}$$

so $\lambda_{j\ell^*} = \alpha_j^*/\sum_{k \in \mathcal{P}^c} \alpha_k^* = S_j^{-1}/\sum_{k \in \mathcal{P}^c} S_k^{-1}$ is the portion of the “non-Pareto” simulation budget allocated to system $j$ for all $j \in \mathcal{P}^c$. Then it follows that $\alpha_j^* = \lambda_{j\ell^*}(1 - \sum_{i=1}^{p} \alpha_i^*).

It is tempting to create a heuristic allocation by solving a reduced version of Problem $\tilde{Q}$ that only includes one constraint for each $j$ and its primary phantom Pareto, $\ell^*$. However, there are some drawbacks of this approach. First, the assumptions of the lim-
iting score regime may not be satisfied, and some Pareto systems may receive a falsely low allocation by not including constraints corresponding to MCE. Since constraints corresponding to MCE involve only Pareto systems, including these constraints in the allocation heuristic may yield much better allocations without adding much computational complexity. Second, while such a version of Problem \( Q \) would have reduced complexity, when the number of Pareto systems is large, the number of constraints still grows with the number of non-Pareto systems, albeit linearly. We avoid these issues by creating a new reduced version of Problem \( Q \) that includes at least one constraint corresponding to MCI for each Pareto system and includes all constraints corresponding to MCE, as follows.

First, for each phantom Pareto system \( \ell \in \{0, \ldots, p\} \), find
\[
    j^{*}_{\ell} := \underset{j \in \mathcal{P}}{\arg\min} \left\{ S_{j}(\ell) : S_{j}(\ell) \neq I_{g_{\ell}}, j \in \mathcal{P} \right\} \text{ if } \ell \neq 0,
\]
\[
    j^{*}_{\ell+1} := \underset{j \in \mathcal{P}}{\arg\min} \left\{ S_{j}(\ell) : S_{j}(\ell) \neq I_{h}, j \in \mathcal{P} \right\} \text{ if } \ell \neq p,
\]
and let \( \mathcal{J}^{*}(\ell) := \{ j_{\ell}^{*} \} \cup \{ j_{\ell+1}^{*} \} \), where \( j_{0}^{*}(0) := \emptyset \) and \( j_{p+1}^{*}(p) := \emptyset \). The set \( \mathcal{J}^{*}(\ell) \) contains up to two of the “closest” non-Pareto systems to phantom Pareto \( \ell \), in terms of their scores. Because it is possible for a non-Pareto to “play” with only one Pareto through the phantom, we ensure that we retain at least one non-Pareto that “plays” with each Pareto \( \ell \) and \( \ell + 1 \) in the set \( \mathcal{J}^{*}(\ell) \). Then the SCORE allocation results from solving the following reduced problem for the allocations to the Pareto systems:

\[
    \text{Problem } Q_{\mathcal{J}} : \quad \max \ z \quad \text{s.t.} \quad R_{i}(\alpha_{i}, \alpha_{k}) \geq z \text{ for all } i, k \in \mathcal{P} \text{ such that } k \neq i,
\]
\[
    R_{j}^{*}(\ell)(\lambda_{j}^{*}(\ell) - (1 - \sum_{\ell=1}^{p} \alpha_{\ell}) \alpha_{\ell}, \alpha_{\ell+1}) \geq z \text{ for all } j^{*}(\ell) \in \mathcal{J}^{*}(\ell), \ell \in \{0, \ldots, p\}
\]
\[
    \sum_{\ell=1}^{p} \alpha_{\ell} \leq 1, \quad \alpha_{i} \geq 0 \text{ for all } i \in \mathcal{P}.
\]

In Problem \( Q_{\mathcal{J}} \), for \( j^{*}(\ell) \in \mathcal{J}^{*}(\ell) \), \( \lambda_{j}^{*}(\ell) \) is the dual variable corresponding to \( j^{*}(\ell) \) and the primary phantom Pareto for \( j^{*}(\ell) \), which we call \( \ell^{*} \). Thus in this context, \( \ell^{*} = \arg\min_{\ell \in \{0, \ldots, p\}} S_{j^{*}(\ell)}(\ell) \) for each \( j^{*}(\ell) \in \mathcal{J}^{*}(\ell) \). Notice that the complexity of Problem \( Q_{\mathcal{J}} \) depends only on the number of Paretos, regardless of the number of non-Paretos.

Since Problem \( Q_{\mathcal{J}} \) is a bi-level optimization problem, to speed up the computation, we use the “closed-form” expressions of the rate functions corresponding to MCI in Proposition 4.2 Part (1). The following Proposition 5.1 provides corresponding closed-form expressions for the rate functions corresponding to MCE. Notice that the rate functions for MCE appear similar in form to those of MCI, except that the correlation \( \rho_{i} \) and \( \rho_{k} \) for both Pareto systems appears.

**Proposition 5.1.** Under Assumption 7, the rate function corresponding to the MCE event for systems \( i, k \in \mathcal{P} \) is
\[
    R_{i}(\alpha_{i}, \alpha_{k}) = \begin{cases} 
        \frac{\alpha_{i}^{2}}{\sigma^{2}_{i}/\alpha_{i} + \sigma^{2}_{k}/\alpha_{k}} & \text{if } g_{i} \leq g_{k}, \ h_{i} \geq h_{k} - (g_{k} - g_{i}), \ \frac{\alpha_{i}^{2}}{\sigma^{2}_{i}/\alpha_{i} + \sigma^{2}_{k}/\alpha_{k}} \\
        \frac{\alpha_{i}^{2}}{\sigma^{2}_{i}/\alpha_{i} + \sigma^{2}_{k}/\alpha_{k}} & \text{if } g_{i} \leq h_{k}, \ h_{i} \geq g_{k} - (h_{k} - h_{i}), \ \frac{\alpha_{i}^{2}}{\sigma^{2}_{i}/\alpha_{i} + \sigma^{2}_{k}/\alpha_{k}} \\
        \frac{\sigma^{2}_{i} + \sigma^{2}_{k}}{\sigma^{2}_{i} + \sigma^{2}_{k}} (h_{k} - h_{i})^{2} - 2 \left[ \frac{\sigma^{2}_{i} + \sigma^{2}_{k}}{\sigma^{2}_{i} + \sigma^{2}_{k}} \alpha_{i}^{2} \right] (g_{k} - g_{i}) (h_{k} - h_{i}) + \left[ \frac{\sigma^{2}_{i} + \sigma^{2}_{k}}{\sigma^{2}_{i} + \sigma^{2}_{k}} \alpha_{i}^{2} \right] (g_{k} - g_{i})^{2} & \text{otherwise}.
    \end{cases}
\]

### 6. Time to Solve for the Score Allocation Versus Optimality Gap
In practice, a decision-maker’s choice of simulation budget allocation method is influenced by the amount of time it takes to solve for the allocation, as well as how close
that allocation is to the optimal allocation. We now give a sense of how our proposed allocations perform on these metrics as the number of systems increases. (In this section, we assume all rate functions are known.)

For a population of ten problems generated according to the uniform method from §4.5, the following Table I reports the average wall-clock time it took to solve for each allocation, the average rate $z$ achieved by the resulting allocation, and the average optimality gap. We kept the same Pareto set as in §4.5, but we allowed the non-Paretos to be closer to the Pareto set. Instead of rejecting non-Pareto systems that were less than 0.25 units away, we rejected non-Paretos that were less than 0.05 units away. Thus the problems are more realistic while keeping the full Problem $Q$ solvable for up to a thousand systems. The specified allocation models are “BVN True,” in which we solve the full Problem $Q$ for the asymptotically optimal allocation; “BVN Independent,” in which we solve the full Problem $Q$, except we model the correlation between the objectives for all systems as $\rho_k = 0$ for all $k \leq r$; SCORE; and equal allocation, which is provided for reference.

<table>
<thead>
<tr>
<th>No. of Sys. ($r$)</th>
<th>Metric</th>
<th>BVN True</th>
<th>BVN Indep.</th>
<th>SCORE</th>
<th>Equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Time</td>
<td>0.07 sec</td>
<td>0.06 sec</td>
<td>0.04 sec</td>
<td>0 sec</td>
</tr>
<tr>
<td></td>
<td>Rate $z \times 10^{-4}$</td>
<td>54.96</td>
<td>44.63</td>
<td>52.52</td>
<td>13.43</td>
</tr>
<tr>
<td></td>
<td>Opt. Gap $\times 10^{-4}$</td>
<td>0 $^a$</td>
<td>10.33</td>
<td>2.44</td>
<td>41.53</td>
</tr>
<tr>
<td>100</td>
<td>Time</td>
<td>0.53 sec</td>
<td>0.47 sec</td>
<td>0.06 sec</td>
<td>0 sec</td>
</tr>
<tr>
<td></td>
<td>Rate $z \times 10^{-4}$</td>
<td>12.97</td>
<td>11.83</td>
<td>11.31</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Opt. Gap $\times 10^{-4}$</td>
<td>0</td>
<td>1.14</td>
<td>1.66</td>
<td>12.31</td>
</tr>
<tr>
<td>500</td>
<td>Time</td>
<td>43.16 sec</td>
<td>25.06 sec</td>
<td>0.08 sec</td>
<td>0 sec</td>
</tr>
<tr>
<td></td>
<td>Rate $z \times 10^{-4}$</td>
<td>1.65</td>
<td>1.37</td>
<td>1.46</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Opt. Gap $\times 10^{-4}$</td>
<td>0</td>
<td>0.28</td>
<td>0.19</td>
<td>1.63</td>
</tr>
<tr>
<td>1,000</td>
<td>Time</td>
<td>18.09 min</td>
<td>12.31 min</td>
<td>0.13 sec</td>
<td>0 sec</td>
</tr>
<tr>
<td></td>
<td>Rate $z \times 10^{-4}$</td>
<td>0.95</td>
<td>0.81</td>
<td>0.82</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>Opt. Gap $\times 10^{-4}$</td>
<td>0</td>
<td>0.14</td>
<td>0.13</td>
<td>0.94</td>
</tr>
<tr>
<td>2,000</td>
<td>Time</td>
<td>$&gt; 6$ hr</td>
<td>$&gt; 6$ hr</td>
<td>0.22 sec</td>
<td>0 sec</td>
</tr>
<tr>
<td></td>
<td>Rate $z \times 10^{-4}$</td>
<td>— $^b$</td>
<td>—</td>
<td>0.45</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>Opt. Gap $\times 10^{-4}$</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5,000</td>
<td>Time</td>
<td>$&gt; 6$ hr</td>
<td>$&gt; 6$ hr</td>
<td>0.48 sec</td>
<td>0 sec</td>
</tr>
<tr>
<td></td>
<td>Rate $z \times 10^{-4}$</td>
<td>—</td>
<td>—</td>
<td>0.22</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>Opt. Gap $\times 10^{-4}$</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>10,000</td>
<td>Time</td>
<td>$&gt; 6$ hr</td>
<td>$&gt; 6$ hr</td>
<td>0.92 sec</td>
<td>0 sec</td>
</tr>
<tr>
<td></td>
<td>Rate $z \times 10^{-4}$</td>
<td>—</td>
<td>—</td>
<td>0.11</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>Opt. Gap $\times 10^{-4}$</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: All computing performed in MATLAB R2015b on a 2.5 GHz Intel Core i7 processor with 16GB 1600MHz DDR3 memory.

$^a$The optimality gap of the true allocation is to the precision of the solver.

$^b$The symbol ‘—’ indicates that data is unavailable due to the large computational time.

Interestingly, from Table I, the BVN Independent allocation is not much faster to calculate than the BVN True allocation. Further, it often yields an average optimality gap that is larger than that of SCORE, which is faster to compute. Thus we do not consider the BVN Independent allocation in further numerical experiments. Table I also seems to show that SCORE is an extremely competitive allocation scheme whether the number of systems is small (on the order of 20 systems) or very large (on the order of 10,000 systems). Further, SCORE is fast — on average, it takes less than a second to solve for the SCORE allocation with 10,000 systems.
7. A SEQUENTIAL ALGORITHM FOR IMPLEMENTATION

Since the SCORE allocation framework requires knowledge of the rate functions, which we clearly do not know in advance, we now present a sequential algorithm, Algorithm 1, that can be used for implementation. The broad idea of this algorithm can be stated as follows: (1) obtain an initial amount of sample \( \delta_0 \) to “learn” about the problem structure and estimate the SCORE allocation; (2) use the estimated SCORE allocation as a probability distribution from which to obtain the next \( \delta \) samples; (3) update the estimated optimal allocation and return to step (2). This algorithm proceeds until some total sampling budget specified by the user has been expended. Since implementing such a stopping rule is trivial, we write the sequential algorithm as nonterminating.

ALGORITHM 1: A sequential algorithm to sample from systems using the proposed allocations

Require: Number of initial samples \( \delta_0 \geq 2 \); number of samples between allocation vector updates \( \delta \geq 1 \); and a minimum-sample proportion \( \alpha_c > 0 \) that is small relative to \( 1/r \).

1: Initialize: collect \( \delta_0 \) samples from each system \( k \leq r \) and set \( n = r \times \delta_0 \), \( n_k = \delta_0 \) for \( k \leq r \).
2: Update the estimated parameters \( \hat{H}_k, \hat{G}_k, \hat{\sigma}_k^2, \hat{\delta}_k \), and \( \hat{\rho}_k \) for all \( k \leq r \); update the estimated rate function \( \hat{I}_k(x_k, y_k) \) for all \( k \leq r \) using the estimated parameters.
3: Calculate the estimated SCORE allocations, \( \alpha_{kn} \), by using the estimators calculated in Step 2 to solve an estimated version of Problem \( Q_k \). (Alternatively, we can use the estimators from Step 2 to solve an estimated version of Problem \( Q \).)
4: for \( m = 1, \ldots, \delta \) do
5: Select a system \( K_m \) from which to obtain the next sample, where each \( K_m \) is an iid random variable with probability mass function \( \alpha_i \) and support \( \{1, 2, \ldots, r\} \).
6: Collect one sample from system \( K_m \) and update \( n_{K_m} = n_{K_m} + 1 \).
7: end for
8: Set \( n = n + \delta \) and update \( \hat{\alpha}_n = \{n_1/n, n_2/n, \ldots, n_r/n\} \). Set \( \delta^+ = 0 \).
9: for \( k = 1, \ldots, r \) do
10: If \( n_k/n < \alpha_c \), collect one sample from system \( k \), then set \( n_k = n_k + 1 \) and \( \delta^+ = \delta^+ + 1 \).
11: end for
12: Set \( n = n + \delta^+ \) and go to Step 2.

While the broad idea of Algorithm 1 is simple, there is one parameter we have not discussed: the minimum-sample proportion \( \alpha_c \). This parameter, which should be small relative to \( 1/r \), ensures that each system is sampled infinitely often as the sequential algorithm progresses — thereby preventing the sequential algorithm from failing to sample from some systems in the limit.

8. NUMERICAL PERFORMANCE OF SEQUENTIAL ALLOCATIONS

In this section we evaluate the performance of sequential versions of the proposed allocations on four test problems.

8.1. Test Problems

We constructed four test problems, shown in Figures 8–11, as follows. First, we generated two instances of true system performances by uniformly generating 100 systems in a circle of radius six, centered at \((100, 100)\). The system objective values in the first test problem correspond to the circle centers in Figures 8 and 9, and the system objective values in the second test problem correspond to the circle centers in Figures 10 and 11. Then we set all systems to have unit variances, but the systems in Figures 8 and 10 have correlation \( \rho_k = -0.5 \) for all \( k \leq r \), while the systems in Figures 9 and 11 have correlation \( \rho_k = 0.5 \) for all \( k \leq r \). At optimality in Problem \( Q \), the test problems in Figures 8 and 9 have a low percent of dual variable values associated with MCE.
constraints, while the test problems in Figures 8 and 9 have a high percent of dual variable values associated with MCE constraints. In Figures 8–11, the asymptotically optimal allocations are proportional to the size of the circle. While there is no obvious “visible” difference in the optimal allocations with positive versus negative correlations, the allocations do differ slightly. A listing of the \((g_k, h_k)\) values for all \(k \leq r\) for each of the test problems is provided in Online Appendix P.

8.2. Estimated Expected Number of Misclassifications
For each algorithm BVN True, SCORE, MOCBA, M-MOBA, and equal allocation, we ran one hundred independent sample paths on each of the test problems in Figures 8–11. For each algorithm, we calculated the average number of misclassifications, false exclusions, and false inclusions across the sample paths, as a function of sample size. Note that for a particular sample path, the sequence containing the number of misclassifications as a function of the sample size \(n\) is autocorrelated. In all implementations of Algorithm 1, which include all sample paths of the BVN True and SCORE allocations, we used parameter settings \(\delta_0 = 5, \delta = 50,\) and \(\alpha_{\epsilon} = 10^{-8}\). In our implementation of MOCBA [Lee et al. 2010], we used parameter settings \(N_0 = \delta_0 = 5, \Delta = \delta = 50,\) and \(\delta = \Delta/2 = 25\), where \(\delta\) is the maximum number of samples one system can receive in a given iteration. In our implementation of M-MOBA [Branke and Zhang 2015], we set \(n_0 = \delta_0 = 5\) and \(\tau = \delta = 50\), where \(\tau\) is the amount of sample given to the alternative with the largest probability of changing the Pareto set. Our stopping rule in M-MOBA is the sampling budget rule. The resulting performance of the algorithms is reported in Figures 12–15.

In terms of overall percent of systems misclassified, all algorithms except equal allocation seem to perform about equally well, with the SCORE allocation and BVN True performing particularly well. Since the optimality guarantees on the BVN True allocation are asymptotic, it is not clear that allocating according to BVN True will perform better than other allocation schemes for finite \(n\). However, BVN True seems consistently to perform well, and the SCORE allocation tracks the BVN True allocation closely in all of the graphs presented in Figures 12–15. Importantly, since Test Problems 1 and 2 have a high percent of dual constraints to MCE — implying the assumptions required in the limiting SCORE framework may not hold — we do not notice a significant loss of quality in the SCORE allocation in Figures 12 and 13.

9. CONCLUDING REMARKS
In conclusion, SCORE is a fast, approximately optimal allocation heuristic for bi-objective R&S derived from an allocation framework that is asymptotically optimal.
Fig. 12. Test 1: For 100 sample paths per algorithm, the graphs show the average % of systems misclassified (MC), % of Paretos falsely excluded (FE), and % of non-Paretos falsely included (FI), respectively.

Fig. 13. Test 2: For 100 sample paths per algorithm, the graphs show the average % of systems misclassified (MC), % of Paretos falsely excluded (FE), and % of non-Paretos falsely included (FI), respectively.

Fig. 14. Test 3: For 100 sample paths per algorithm, the graphs show the average % of systems misclassified (MC), % of Paretos falsely excluded (FE), and % of non-Paretos falsely included (FI), respectively.

Fig. 15. Test 4: For 100 sample paths per algorithm, the graphs show the average % of systems misclassified (MC), % of Paretos falsely excluded (FE), and % of non-Paretos falsely included (FI), respectively.
in a certain rigorous sense. We mention here that we are aware of issues with estimating rate functions in a general context [Glynn and Juneja 2011; 2015]. However our numerical experience in the case of normal rate functions has been overwhelmingly positive [Pasupathy et al. 2014; Hunter and McClosky 2016]. Finally, it is not clear that our methods for bi-objective R&S extend cleanly to multi-objective R&S. We rely heavily on the definition of phantom Pareto systems, which seem to be easily constructed only in two dimensions. For further insights to the multi-objective case, see Feldman et al. [2015].

ACKNOWLEDGMENTS

The authors would like to thank Juergen Branke and Wen Zhang for providing M-MOBA code, Loo Hay Lee for providing MOCBA code, and Raghu Pasupathy for interesting discussions relevant to this work.

REFERENCES


Received May 2016
A. PROOF OF THEOREM 3.1

First, note that

\[ MC = MCE \cup (MCI \cap MCE^c) \]

and

\[ MC_{ph} = MCE \cup (MCI_{ph} \cap MCE^c). \]

\((\Rightarrow)\) To show \(MC\) implies \(MC_{ph}\), it is sufficient to show that \(MCI \cap MCE^c\) implies \(MCI_{ph}\). Suppose \(MCI \cap MCE^c\). Let \(j \in \mathcal{P}^c\) and \(j \in \mathcal{P}\) be a non-Pareto system falsely estimated as Pareto. Then for each \(i \in \mathcal{P}\), \(\hat{G}_j \leq \hat{G}_i\) or \(\hat{H}_j \leq \hat{H}_i\). Thus \((\hat{G}_j, \hat{H}_j) \in \cap_{i \in \mathcal{P}}((\hat{G}_i, \hat{H}_i) : (\hat{G}_\ell \leq \hat{G}_i) \cup (\hat{H}_\ell \leq \hat{H}_i)).\) Since \(MCE^c\), no Pareto systems dominate other Pareto systems. Therefore \((\hat{G}_j, \hat{H}_j) \in \cup_{\ell=0,\ldots,n}((\hat{G}_\ell, \hat{H}_\ell) : (\hat{G}_\ell \leq \hat{G}_{\ell+1} \cap (\hat{H}_\ell \leq \hat{H}_{\ell+1})).\) That is, \(j\) lies in the union of the southwest quadrants defined by origins at the estimated phantom Pareto systems. Therefore \(MCI_{ph}\) occurs.

\((\Leftarrow)\) To show \(MC_{ph}\) implies \(MC\), it is sufficient to show that \(MCI_{ph} \cap MCE^c\) implies \(MC\). Suppose \(MCI_{ph} \cap MCE^c\). Then all Pareto systems are estimated as Pareto. From the set of all \(j \in \mathcal{P}^c\) dominating some estimated phantom Pareto system, there exists \(j^* \in \mathcal{P}^c\) such that \(j^* \in \mathcal{P}\). (Otherwise, if there exists no such \(j^*\), then each \(j \in \mathcal{P}^c\) is dominated by some \(i \in \mathcal{P}\), and \(MCI_{ph}\) does not occur.) Therefore \(\mathcal{P} \neq \mathcal{P}\), which implies \(MC\).

B. PROOF OF LEMMA 3.3

By the law of total probability,

\[ P\{MCI_{ph}\} = P\{MCI_{ph} \cap \hat{O} = \emptyset\} + \sum_{S \neq \emptyset} P\{MCI_{ph} \cap \hat{O} = S\}. \]

Consider the first term on the right hand side, \(P\{MCI_{ph} \cap \hat{O} = \emptyset\}\). Since \(\hat{O} = \emptyset\) occurs, we may write \(MCI_{ph}\) without order statistics so that

\[ P\{MCI_{ph}\} = P\{MCI_{ph} \cap \hat{O} = \emptyset\} + \sum_{S \neq \emptyset} P\{MCI_{ph} \cap \hat{O} = S\}. \]

Assuming the limits exist, by the principle of the slowest term [Ganesh et al. 2004, Lemma 2.1],

\[ -\lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph}\} \]

\[ = -\lim_{n \to \infty} \frac{1}{n} \log \left( P\{MCI_{ph} \cap \hat{O} = \emptyset\} + \sum_{S \neq \emptyset} P\{MCI_{ph} \cap \hat{O} = S\} \right) \]

\[ = \min\left( -\lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph} \cap \hat{O} = \emptyset\}, \min_{S \neq \emptyset} -\lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph} \cap \hat{O} = S\} \right). \]

Then from equation (1),

\[ -\lim_{n \to \infty} \frac{1}{n} \log P\{MC\} = \min\left( -\lim_{n \to \infty} \frac{1}{n} \log P\{MCE\}, -\lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph} \cap \hat{O} = \emptyset\}, \min_{S \neq \emptyset} -\lim_{n \to \infty} \frac{1}{n} \log P\{MCI_{ph} \cap \hat{O} = S\} \right). \]

\[ (15) \]
We now show that \( \min_{\delta \neq 0} - \lim_{n \to \infty} \frac{1}{n} \log P[MCI_{ph} \cap \hat{\delta} = S] \) is never the binding minimum in equation (15). First, it is clear that any value of \( S \) that results in an MCE event will have a rate function that is greater than or equal to the corresponding rate function for MCE. Now consider values of \( S \) that do not result in MCE, such that all Pareto systems are estimated as Pareto, but may be estimated in the wrong order, e.g., \( S = \{(2, 2), (1, 1), (3, 3), \ldots, (p, p)\} \), where Pareto systems 1 and 2 have exchanged positions. In this case, it is sufficient to consider only instances of \( S \) that contain pair-wise exchanges, and further, it is sufficient to consider instances of \( S \) in which there is exactly one “pair exchange.” For any \( \ell_1, \ell_2 \) indexing the Pareto set such that \( \ell_1 < \ell_2 \), a pair exchange occurs if \((G_{\ell_2} \leq \hat{G}_{\ell_2}) \cap (H_{\ell_1} \leq \hat{H}_{\ell_1})\). Let \( S_{(\ell_1, \ell_2)} \) denote an ordering with exactly one pair exchange where \( \ell_1 \) and \( \ell_2, \ell_1 < \ell_2 \), are the Pareto systems whose places have been exchanged. Thus

\[
\min_{\delta \neq 0} - \lim_{n \to \infty} \frac{1}{n} \log P[MCI_{ph} \cap \hat{\delta} = S] \\
\geq \min\left\{ - \lim_{n \to \infty} \frac{1}{n} \log P[MCE], - \lim_{n \to \infty} \frac{1}{n} \log P[MCI_{ph} \cap \hat{\delta} = S(\ell_1, \ell_2)] \right\} \\
\geq \min\left\{ - \lim_{n \to \infty} \frac{1}{n} \log P[MCE], - \lim_{n \to \infty} \frac{1}{n} \log P[\hat{\delta} = S(\ell_1, \ell_2)] \right\}.
\]

Now note that

\[
- \lim_{n \to \infty} \frac{1}{n} \log P[\hat{\delta} = S(\ell_1, \ell_2)] \geq \min_{\ell_1, \ell_2, P} \inf_{x_1 \leq x_1, y_1 \leq y_2} \alpha_{\ell_1} I_{\ell_1}(x_{\ell_1}, y_{\ell_1}) + \alpha_{\ell_2} I_{\ell_2}(x_{\ell_2}, y_{\ell_2}).
\]

Since the infimum in equation (16) is a convex minimization problem with a nonempty interior, the KKT conditions [Boyd and Vandenberghe 2004] are necessary and sufficient for global optimality. In addition to primary feasibility conditions, for \( \lambda_x \geq 0 \) and \( \lambda_y \geq 0 \) we have the complementary slackness conditions \( \alpha_x (x_{\ell_1} - x_{\ell_2}^*) = 0 \) and \( \lambda_y (y_{\ell_1} - y_{\ell_2}^*) = 0 \), and the stationarity conditions

\[
\alpha_{\ell_1} \frac{\partial I_{\ell_1}(x_{\ell_1}, y_{\ell_1})}{\partial x_{\ell_1}} - \lambda_x = 0, \quad \alpha_{\ell_2} \frac{\partial I_{\ell_1}(x_{\ell_1}, y_{\ell_1})}{\partial y_{\ell_1}} + \lambda_y = 0, \\
\alpha_{\ell_1} \frac{\partial I_{\ell_1}(x_{\ell_1}, y_{\ell_1})}{\partial x_{\ell_2}} + \lambda_x = 0, \quad \alpha_{\ell_2} \frac{\partial I_{\ell_1}(x_{\ell_1}, y_{\ell_1})}{\partial y_{\ell_2}} - \lambda_y = 0.
\]

Since \( \lambda_x = \lambda_y = 0 \) implies that \( (x_{\ell_1}^*, y_{\ell_1}^*) = (g_{\ell_1}, h_{\ell_1}) \) and \( (x_{\ell_2}^*, y_{\ell_2}^*) = (g_{\ell_2}, h_{\ell_2}) \), which is infeasible, then it must be the case that \( \lambda_x > 0 \) or \( \lambda_y > 0 \). Since both \( \ell_1 \) and \( \ell_2 \) are Pareto systems, without loss of generality, suppose that \( \lambda_y > 0 \) so that \( y_{\ell_1}^* = y_{\ell_2}^* \). Then continuing from line (16),

\[
(16) \geq \min_{\ell_1, \ell_2, P} \inf_{x_{\ell_1} \leq x_{\ell_1}, y_{\ell_1} = y_{\ell_2}} \alpha_{\ell_1} I_{\ell_1}(x_{\ell_1}, y_{\ell_1}) + \alpha_{\ell_2} I_{\ell_2}(x_{\ell_2}, y_{\ell_2}) \\
\geq \min_{\ell_1, \ell_2, P} \inf_{x_{\ell_1} \leq x_{\ell_1}, y_{\ell_2} \leq y_{\ell_1}} \alpha_{\ell_1} I_{\ell_1}(x_{\ell_1}, y_{\ell_1}) + \alpha_{\ell_2} I_{\ell_2}(x_{\ell_2}, y_{\ell_2}) \geq - \lim_{n \to \infty} \frac{1}{n} \log P[MCE],
\]

and the result follows.

C. PROOF OF THEOREM 3.4

Starting from Lemma 3.3, we show that the rate of decay of \( P[MCI_{ph} \cap \hat{\delta} = \emptyset] \) is either equal to \( \min_{j \in \mathcal{P}} \min_{\ell_0 = \ldots, \ell_{j-1} \in \mathcal{P}} R_j(\alpha_j, \ell_0, \ell_{j-1}) \), or greater than or equal to the rate of decay of \( P[MCE] \). The substitution regarding the rate of decay of \( P[MCE] \) follows from Lemma 3.2 and by noting that \( j \in \mathcal{P} \) falsely dominates \( i \in \mathcal{P} \) if and only if it also falsely dominates a phantom Pareto system. In the proof that follows, we show that whenever
\[ -\lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^c \cap \hat{\mathcal{O}} = \emptyset\} \text{ is not equal to} \min_{j \in \mathcal{P}^c} \min_{\ell = 0, \ldots, p} R_j(\alpha_j, \ell, \alpha_{\ell+1}), \text{the rate of decay of} \ P\{\text{MCI}_{ph}^c \cap \hat{\mathcal{O}} = \emptyset\} \text{ is bounded below by the rate of decay of} \ P\{\text{MCE}\} \text{ between two Pareto systems. Thus we only need to consider the minimum MCE rate function across all} i \in \mathcal{P} \text{ and} \ k \in \mathcal{P}, k \neq i. \]

We now prove the results regarding the rate of decay of \( P\{\text{MCI}_{ph}^c \cap \hat{\mathcal{O}} = \emptyset\} \). Note that we may write the rate function for \( P\{\text{MCI}_{ph}^c \cap \hat{\mathcal{O}} = \emptyset\} \) as

\[ -\lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCI}_{ph}^c \cap \hat{\mathcal{O}} = \emptyset\} = \min_{j \in \mathcal{P}^c} \min_{\ell \in \{0, \ldots, p\}} \left( -\lim_{n \to \infty} \frac{1}{n} \log P\{(\check{G}_j \leq \hat{G}_{\ell+1}) \cap (\check{H}_j \leq \hat{H}_\ell) \cap \left[ \cap_{i=1}^{p-1}(\hat{G}_i \leq \check{G}_{\ell+1}) \cap (\hat{H}_{i+1} \leq \check{H}_\ell) \right] \right). \] (17)

For some \( j \in \mathcal{P}^c \) and \( \ell \in \{0, \ldots, p\} \), consider the inner rate function from line (17),

\[ -\lim_{n \to \infty} \frac{1}{n} \log P\{(\check{G}_j \leq \hat{G}_{\ell+1}) \cap (\check{H}_j \leq \hat{H}_\ell) \cap \left[ \cap_{i=1}^{p-1}(\hat{G}_i \leq \check{G}_{\ell+1}) \cap (\hat{H}_{i+1} \leq \check{H}_\ell) \right] \}. \] (18)

Then it can be shown that

\[
(18) = \begin{cases} 
\inf_{x_j \leq x_j^*} \alpha_j I_j(x_j, y_j) + \sum_{i=1}^{p} \alpha_i I_i(x_i, y_i) & \text{if} \ \ell = 0 \\
\inf_{x_j \leq x_j^* \cap y_j \leq y_j^*} \alpha_j I_j(x_j, y_j) + \sum_{i=1}^{p} \alpha_i I_i(x_i, y_i) & \text{if} \ \ell \in \{1, \ldots, p-1\} \\
\inf_{x_j \leq x_j^* \cap y_j \leq y_j^*} \alpha_j I_j(x_j, y_j) + \sum_{i=1}^{p} \alpha_i I_i(x_i, y_i) & \text{if} \ \ell = p.
\end{cases}
\] (19)

Since all problems in (19) are convex minimization problems where Slater’s condition holds [Boyd and Vandenberghe 2004], the KKT conditions are necessary and sufficient for global optimality. Then from (19), for \( \lambda_{x_i} \geq 0; \lambda_{y_i} \geq 0; \lambda_{x_{\ell+1}} \geq 0 \) for all \( i = 1, \ldots, p-1 \), and \( \lambda_{y_{\ell+1}} \geq 0 \) for all \( i = 2, \ldots, p \), the KKT conditions for complementary slackness are

\[
\lambda_{x_j}(x_j^* - x_j^* \cap y_j \leq y_j^*) = 0 \text{ if } \ell \neq p; \\
\lambda_{y_j}(y_j^* - y_j^*) = 0 \text{ if } \ell \neq 0; \\
\lambda_{x_i}(x_i^* - x_i^* \cap y_i \leq y_i^*) = 0 \forall i = 1, \ldots, p-1; \\
\lambda_{y_i}(y_i^* - y_i^*) = 0 \forall i = 1, \ldots, p-1;
\]

and, for all \( i \in \{2, \ldots, p-1\} \), the stationarity conditions are

\[
\alpha_i \frac{\partial I_i^*(x_i^*, y_i^*)}{\partial x_i} + \lambda_{x_i} I_i^*(x_i^*, y_i^*) = 0, \\
\alpha_i \frac{\partial I_i^*(x_i, y_i^*)}{\partial x_i} + \lambda_{x_{\ell+1}} = 0, \\
\alpha_{\ell+1} \frac{\partial I_{\ell+1}^*(x_{\ell+1}^*, y_{\ell+1}^*)}{\partial x_{\ell+1}} - \lambda_{x_{\ell+1}} - \lambda_{x_{\ell+1}} = 0, \\
\alpha_{\ell+1} \frac{\partial I_{\ell+1}^*(x_{\ell+1}, y_{\ell+1}^*)}{\partial x_{\ell+1}} - \lambda_{x_{\ell+1}} - \lambda_{x_{\ell+1}} = 0, \\
\alpha_p \frac{\partial I_p^*(x_p^*, y_p^*)}{\partial y_p} - \lambda_{y_{\ell+1}} = 0, \\
\alpha_p \frac{\partial I_p^*(x_p^*, y_p^*)}{\partial y_p} = 0.
\]

Now suppose that at optimality in (19), \( x_i^* = x_i^* \cap y_i \leq y_i^* \) for some \( i \in \{1, \ldots, p-1\} \). Then removing constraints in (19), for \( \ell \in \{0, \ldots, p\} \), and for some \( i \in \{1, \ldots, p-1\} \),

\[ (19) \geq \inf_{x_i \geq x_i^* \cap y_i \leq y_i^*} \alpha_i I_i(x_i, y_i) + \alpha_{i+1} I_{i+1}(x_{i+1}, y_{i+1}) \]

\[ \geq \inf_{x_i \geq x_i^* \cap y_i \leq y_i^*} \alpha_i I_i(x_i, y_i) + \alpha_{i+1} I_{i+1}(x_{i+1}, y_{i+1}) \geq -\lim_{n \to \infty} \frac{1}{n} \log P\{\text{MCE}\}. \]

By a similar argument, it can also be shown that if \( y_i^* = y_{i-1}^* \) for any \( i \in \{2, \ldots, p\} \), then the rate in (19) is bounded below by the rate of decay of \( P\{\text{MCE}\} \) for some \( i, k \in \mathcal{P} \).

Therefore it is sufficient to consider only \( \lambda_{x_i} = 0 \) for all \( i = 1, \ldots, p-1 \) and \( \lambda_{y_i} = 0 \).
for all $i = 2, \ldots, p$; otherwise, the rate would never be the unique minimum in the overall rate of decay of $P\{MC\}$. Thus we consider only the case $(x_i^*, y_i^*) = (h_i, g_i)$ for all $i = 1, \ldots, p, i \neq \ell, i \neq \ell + 1$, and it is sufficient to search for $x_{\ell+1}$ in $(g_{\ell}, g_{\ell+2})$ and $y_{\ell}$ in $(h_{\ell-1}, h_{\ell})$. However, since we have already shown that $x_{\ell+1}^* \notin (g_{\ell}, g_{\ell+2})$ and $y_{\ell}^* \notin (h_{\ell-1}, h_{\ell})$ results in a rate of decay of $P\{MCI_{ph} \cap \emptyset = \emptyset\}$ that is bounded below by the rate of decay of $P\{MCE\}$, it is sufficient to simplify (19) to

$$\inf_{x_j \leq x_1} \alpha_j I_j(x_j, y_j) + \alpha_1 I_1(x_1, y_1) \quad \text{if } \ell = 0$$

$$\inf_{x_j \leq x_{\ell+1}, y_j \leq y_{\ell}} \alpha_j I_j(x_j, y_j) + \sum_{i=\ell}^{\ell+1} \alpha_i I_i(x_i, y_i) \quad \text{if } \ell \in \{1, \ldots, p-1\}$$

$$\inf_{y_j \leq y_p} \alpha_j I_j(x_j, y_j) + \alpha_p I_p(x_p, y_p) \quad \text{if } \ell = p,$$

which we previously defined as $R_j(\alpha_j, \alpha_\ell, \alpha_{\ell+1})$.

**D. STATEMENT OF LEMMA D.1 WITH PROOF**

**Lemmas D.1.** Suppose $\alpha_j > 0, \alpha_\ell > 0$, and $\alpha_{\ell+1} > 0$. At optimality in Problem $R_{j\ell}$, the following hold.

1. If $\ell \in \{1, \ldots, p-1\}$, then $g_{\ell+1} \leq x_{\ell+1}^*, h_{\ell} \leq y_{\ell}^*,$ and $x_j^* \leq g_j$ or $y_j^* \leq h_j$. Further, if $\lambda_x = 0$, then $y_j^* \leq h_j$, and if $\lambda_y = 0$, then $x_j^* \leq g_j$.
2. If $\ell = 0$, then $g_{\ell+1} < x_{\ell+1}^* < g_j$.
3. If $\ell = p$, then $h_{\ell} < y_{\ell}^* < h_j$.

**Proof.** Proof of (1). By the convexity of the rate function $I_{\ell+1}(\cdot)$, it holds that

$$(\nabla I_{\ell+1}(g_{\ell+1}, h_{\ell+1}) - \nabla I_{\ell+1}(x_{\ell+1}^*, y_{\ell+1}^*))^T((g_{\ell+1}, h_{\ell+1}) - (x_{\ell+1}^*, y_{\ell+1}^*)) \geq 0,$$

which, together with the KKT conditions, implies

$$\frac{\partial I_{\ell+1}(x_{\ell+1}^*, y_{\ell+1}^*)}{\partial x_{\ell+1}} (g_{\ell+1} - x_{\ell+1}^*) \geq 0. \quad \text{Thus } g_{\ell+1} \leq x_{\ell+1}^*.$$

A similar proof shows that

$$-\frac{\partial I_{\ell+1}(x_{\ell+1}^*, y_{\ell+1}^*)}{\partial y_{\ell+1}} (h_{\ell} - y_{\ell}^*) \geq 0. \quad \text{Thus } h_{\ell} \leq y_{\ell}^*.$$

Using similar logic, by the convexity of $I_j(\cdot)$, we get $g_j \leq x_j^*$ and $y_j^* \leq h_j$. Since the KKT conditions imply both partial derivatives are non-negative, then it must be the case that $x_j^* \leq g_j$ or $y_j^* \leq h_j$, and the implications when $\lambda_x = 0$ or $\lambda_y = 0$ follow.

Proof of (2). When $\ell = 0$, we know that $g_{\ell+1} < g_j$. The KKT conditions simplify to

$$\lambda_x (x_j^* - x_{\ell+1}^*) = 0$$

and

$$\alpha_j \frac{\partial I_j(x_j^*, y_j^*)}{\partial x_j} + \lambda_x = 0, \quad \alpha_j \frac{\partial I_j(x_j^*, y_j^*)}{\partial y_j} = 0,$$

$$\alpha_{\ell+1} \frac{\partial I_{\ell+1}(x_{\ell+1}^*, y_{\ell+1}^*)}{\partial x_{\ell+1}} - \lambda_x = 0, \quad \alpha_{\ell+1} \frac{\partial I_{\ell+1}(x_{\ell+1}^*, y_{\ell+1}^*)}{\partial y_{\ell+1}} = 0.$$

Since $\lambda_x = 0$ implies that $(x_j^*, y_j^*) = (g_j, h_j)$ and $(x_{\ell+1}^*, y_{\ell+1}^*) = (g_{\ell+1}, h_{\ell+1})$, which is an infeasible point, then it must be the case that $\lambda_x > 0$. Therefore $x_j^* = x_{\ell+1}^*$. By the strict convexity of the rate functions, it follows that $g_{\ell+1} < x_j^* = x_{\ell+1}^* < g_j$.

Proof of (3). This proof follows in a similar manner to that of part (2). □

**E. PROOF OF PROPOSITION 4.2**

First, recall that

$$\text{Problem } R_{j\ell} : \quad \text{minimize } \alpha_j I_j(x_j, y_j) + \alpha_\ell I_\ell(x_\ell, y_\ell) \mathbb{I}_{[\ell \neq \emptyset]} + \alpha_{\ell+1} I_{\ell+1}(x_{\ell+1}, y_{\ell+1}) \mathbb{I}_{[\ell \neq \emptyset]}$$

$$\text{s.t. } (x_j - x_{\ell+1}) \mathbb{I}_{[\ell \neq \emptyset]} \leq 0, \quad (y_j - y_{\ell}) \mathbb{I}_{[\ell \neq \emptyset]} \leq 0.$$
Using this information and updating the stationarity conditions (2)–(4) yields

\[
\alpha_{\ell} \left[ \frac{(x_{\ell}^* - g_{\ell})}{\sigma_{\ell}} \right] - \lambda_{\ell} I_{\ell}^{(p \neq 0)} = 0, \quad \alpha_{\ell} \left[ \frac{(y_{\ell}^* - h_{\ell})}{\sigma_{\ell}} \right] - \lambda_{\ell} I_{\ell}^{(p \neq 0)} = 0.
\]

We also have the complementary slackness conditions \( \lambda_{\ell}(x_{\ell}^* - x_{\ell+1}^*) = 0 \) if \( \ell \neq p \) and \( \lambda_{\ell}(y_{\ell}^* - y_{\ell+1}^*) = 0 \) if \( \ell \neq 0 \).

Notice that if \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} = \lambda_{\ell} I_{\ell}^{(p = 0)} = 0 \), then we have primal infeasibility. We now consider three cases: (1) \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} > 0, \lambda_{\ell} I_{\ell}^{(p = 0)} = 0 \); (2) \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} = 0, \lambda_{\ell} I_{\ell}^{(p = 0)} > 0 \); and (3) \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} > 0, \lambda_{\ell} I_{\ell}^{(p = 0)} > 0 \).

**Case 1: Suppose \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} > 0 \) and \( \lambda_{\ell} I_{\ell}^{(p = 0)} = 0 \).** Then \( (x_{\ell}^*, y_{\ell}^*) = (g_{\ell}, h_{\ell}) \), and

\[
x_{\ell}^* = x_{\ell+1} = \frac{(\alpha_{\ell}/\sigma_{\ell}^2)g_{\ell} + (\alpha_{\ell+1}/\sigma_{\ell+1}^2)(g_{\ell+1})}{\alpha_{\ell}/\sigma_{\ell}^2 + \alpha_{\ell+1}/\sigma_{\ell+1}^2}, \quad y_{\ell}^* = h_{\ell+1} + \rho_{\ell+1}\frac{\sigma_{h_{\ell}}}{\sigma_{\ell}^2 + \alpha_{\ell+1}/\sigma_{\ell+1}^2},
\]

which implies the results in Parts (1) and (2) in this case. Notice that if \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} > 0 \) and \( \lambda_{\ell} I_{\ell}^{(p = 0)} = 0 \), primal feasibility of \( y_{\ell}^* \) implies \( h_{\ell} \leq h_{\ell+1} + \rho_{\ell+1}\frac{\sigma_{h_{\ell}}}{\sigma_{\ell}^2 + \alpha_{\ell+1}/\sigma_{\ell+1}^2} \).

**Case 2: Suppose \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} = 0 \) and \( \lambda_{\ell} I_{\ell}^{(p = 0)} > 0 \).** Then \( (x_{\ell+1}^*, y_{\ell+1}^*) = (g_{\ell+1}, h_{\ell+1}) \) and

\[
x_{\ell+1}^* = g_{\ell+1} + \rho_{\ell+1}\frac{\sigma_{h_{\ell}}}{\sigma_{\ell}^2 + \alpha_{\ell}/\sigma_{\ell}^2}, \quad y_{\ell+1}^* = h_{\ell+1} + \rho_{\ell+1}\frac{\sigma_{h_{\ell}}}{\sigma_{\ell}^2 + \alpha_{\ell}/\sigma_{\ell}^2},
\]

which implies the results in Parts (1) and (2) in this case. Notice that if \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} = 0 \) and \( \lambda_{\ell} I_{\ell}^{(p = 0)} > 0 \), then primal feasibility of \( x_{\ell}^* \) implies \( g_{\ell} \leq \rho_{\ell+1} + \frac{\sigma_{h_{\ell}}}{\sigma_{\ell}^2 + \alpha_{\ell}/\sigma_{\ell}^2} \).

**Case 3: Suppose \( \lambda_{\ell} I_{\ell}^{(p \neq 0)} > 0 \) and \( \lambda_{\ell} I_{\ell}^{(p = 0)} > 0 \).** Then

\[
x_{\ell}^* = x_{\ell+1} = \frac{(\alpha_{\ell}/\sigma_{\ell}^2)g_{\ell} + (\alpha_{\ell+1}/\sigma_{\ell+1}^2)(g_{\ell+1})}{\alpha_{\ell}/\sigma_{\ell}^2 + \alpha_{\ell+1}/\sigma_{\ell+1}^2} + \frac{\alpha_{\ell}g_{\ell}}{\alpha_{\ell}/\sigma_{\ell}^2 + \alpha_{\ell+1}/\sigma_{\ell+1}^2} \frac{(\alpha_{\ell}/\sigma_{\ell}^2)h_{\ell} + (\alpha_{\ell+1}/\sigma_{\ell+1}^2)h_{\ell+1}}{\alpha_{\ell}/\sigma_{\ell}^2 + \alpha_{\ell+1}/\sigma_{\ell+1}^2} + \frac{\alpha_{\ell+1}g_{\ell+1}}{\alpha_{\ell}/\sigma_{\ell}^2 + \alpha_{\ell+1}/\sigma_{\ell+1}^2}.
\]
\[ y_j^* = y_t^* = \frac{\alpha_j}{(1-\rho_j^t)\sigma_g,\sigma_j} \left[ \frac{\alpha_j \rho_j_t}{\sigma_g,\sigma_j} \right] + \frac{\alpha_j}{(1-\rho_j^t)\sigma_g,\sigma_j} \left( \frac{\alpha_j}{\sigma_j} g^{*}_{j+1} \right) \]

\[ x_t^* = g_t + \rho_t \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} \left[ \frac{\alpha_j}{(1-\rho_j^t)\sigma_g,\sigma_j} \right] + \frac{\alpha_j}{(1-\rho_j^t)\sigma_g,\sigma_j} \left( \frac{\alpha_j}{\sigma_j} g^{*}_{j+1} \right) \]

\[ y_{t+1}^* = h_{t+1} + \rho_{t+1} \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} \left[ \frac{\alpha_j}{(1-\rho_j^t)\sigma_g,\sigma_j} \right] + \frac{\alpha_j}{(1-\rho_j^t)\sigma_g,\sigma_j} \left( \frac{\alpha_j}{\sigma_j} g^{*}_{j+1} \right) \]

which implies the results in Parts (1) and (2) in this case. Algebra reveals that \( x_t^* > g_t+1 \) and \( y_t^* > h_t \) imply \( g_t > g_t+1 + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (h_t - h_t) \) and \( h_t > h_t + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (g_t - g_t+1) \).

We now show that the indicators depend on the locations of the systems.

**Case 1a:** Suppose \( \ell \neq p, g_j > g_{t+1}, h_j \leq h_t + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (g_j - g_{t+1}) \). First, since \( \ell \neq p, g_j > g_{t+1} \), it follows that \( x_t^* = g_t, x_t^* = g_t+1 \) is infeasible. If we nonetheless have \( x_t^* = g_t+1 \), then \( \lambda_{x_t} [\ell \neq p] > 0 \) and \( \lambda_{y_t} [\ell \neq p] > 0 \). However from Case 2 above, this implies \( g_j \leq g_t+1 + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (h_j - h_t) (\frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} \frac{\alpha_j}{\sigma_j}) \) and \( h_j > h_t \), which provides a contradiction. Therefore it must be the case that \( x_t^* = g_t+1 \) and \( \lambda_{x_t} [\ell \neq p] > 0 \). From Case 3 above, if \( \lambda_{y_t} [\ell \neq p] > 0 \), we also have a contradiction. Therefore \( \lambda_{x_t} [\ell \neq p] > 0 \) and \( \lambda_{y_t} [\ell \neq p] = 0 \).

**Case 2a:** Suppose \( \ell \neq 0, h_j > h_t, g_j \leq g_t+1 + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (h_j - h_t) (\frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} \frac{\alpha_j}{\sigma_j}) \). First, since \( \ell \neq 0, h_j > h_t \), then \( y_t^* = h_t, y_t^* = h_t \) is infeasible. If we nonetheless have \( y_t^* = h_t \), then \( \lambda_{y_t} [\ell \neq p] > 0 \) and \( \lambda_{x_t} [\ell \neq p] = 0 \). However the results of Case 1 above provide a contradiction. Therefore it must be the case that \( y_t^* > h_t \), and hence \( \lambda_{y_t} [\ell \neq p] > 0 \). Since the results of Case 3 also provide a contradiction, it follows that \( \lambda_{x_t} [\ell \neq p] = 0 \) and \( \lambda_{y_t} [\ell \neq p] > 0 \).

**Case 3a:** Suppose \( \ell \notin \{0, p\}, g_j > g_{t+1} + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (h_j - h_t) (\frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} \frac{\alpha_j}{\sigma_j}) \), and \( h_j > h_t + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (g_j - g_{t+1}) (\frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} \frac{\alpha_j}{\sigma_j}) \). Combining the results of Cases 1 and 1a, we have shown that \( \lambda_{x_t} [\ell \neq p] > 0 \) and \( \lambda_{y_t} [\ell \neq p] = 0 \) if and only if \( \ell \neq p, g_j > g_{t+1}, h_j \leq h_t + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (g_j - g_{t+1}) (\frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} \frac{\alpha_j}{\sigma_j}) \). Combining the results of Cases 2 and 2a, we have shown that \( \lambda_{x_t} [\ell \neq p] = 0 \) and \( \lambda_{y_t} [\ell \neq p] > 0 \) if and only if \( \ell \neq 0, h_j > h_t, g_j \leq g_{t+1} + \rho_j \frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} (h_j - h_t) (\frac{\sigma g,\sigma_h}{\sigma g,\sigma_h} \frac{\alpha_j}{\sigma_j}) \).
\[
\rho_j \sigma_j / \sigma_i = (\sigma_j^2 / \alpha_i) + \sigma_i^2 / \alpha_i + \sigma_i^2 / \alpha_i.
\]

Therefore in Case 3a, the only remaining possibility is that \(\lambda_\ell \in [\ell \neq p] > 0\) and \(\lambda_\ell^{(p \neq 0)} > 0\).

Finally, as a point of interest, we have that the values of the dual variables are

\[
\lambda_x = -\alpha_j \frac{\partial l_j(x_j^\ast, y_j^\ast)}{\partial x_j} = \alpha_\ell + 1 \frac{\partial l_\ell+1(x_\ell+1, y_\ell+1)}{\partial x_\ell+1}
\]

and

\[
\lambda_y = -\alpha_j \frac{\partial l_j(x_j^\ast, y_j^\ast)}{\partial y_j} = \alpha_\ell \frac{\partial l_\ell(x_\ell+1, y_\ell+1)}{\partial y_\ell+1},
\]

where

\[
\frac{\partial l_\ell(x_\ell^+, y_\ell^+)}{\partial x_\ell} = -\left( \frac{1}{\alpha_j} \right) \left[ \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_j^2}{\alpha_\ell+1} \left( g_j - g_\ell+1 \right) (h_j - h_\ell+1) \right] \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_j^2}{\alpha_\ell+1} \right)
\]

and

\[
\frac{\partial l_\ell(x_\ell^+, y_\ell^+)}{\partial y_\ell} = -\left( \frac{1}{\alpha_j} \right) \left[ \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_j^2}{\alpha_\ell+1} \left( h_j - h_\ell+1 \right) \right] \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_j^2}{\alpha_\ell+1} \right)
\]

F. PROOF OF LEMMA 4.3

We do not provide a proof when \(\ell \in \{1, p\}\). To prove the result when \(\ell \in \{2, \ldots, p-1\}\), for a contradiction, suppose that \(\ell \in \{2, \ldots, p-1\}\), and that the rate function \(I_\ell\) “drops out” of both problems. Let \(\lambda_\ell(\ell - 1), \lambda_\ell(\ell - 1)\) be the dual variables for Problem \(R_{j\ell-1}\) and let \(\lambda_x(\ell), \lambda_y(\ell)\) be the dual variables for Problem \(R_{j\ell}\).

Then in Problem \(R_{j\ell-1}\), from Lemma D.1, we have \(\lambda_x(\ell, \ell - 1, \alpha_\ell) = (g_\ell, h_\ell)\), which implies \(\lambda_x(\ell - 1) = 0, \lambda_y(\ell - 1) > 0\). Thus in Problem \(R_{j\ell-1}\),

\[
x_j^\ast(\alpha_j, \alpha_{\ell-1}, \alpha_\ell) \leq g_\ell, \quad h_\ell < h_{\ell-1} \leq y_j^\ast(\alpha_j, \alpha_{\ell-1}, \alpha_\ell) = y_j^\ast(\alpha_j, \alpha_{\ell-1}, \alpha_\ell) \leq h_j.
\]

Likewise, in Problem \(R_{j\ell}\), we have \(y_j^\ast(\alpha_j, \alpha_\ell, \alpha_{\ell+1}) = (g_\ell, h_\ell)\), which implies \(\lambda_y(\ell) > 0, \lambda_y(\ell) = 0\). Thus in Problem \(R_{j\ell}\),

\[
g_\ell < g_{\ell+1} \leq x_j^\ast(\alpha_j, \alpha_{\ell}, \alpha_{\ell+1}) = x_j^\ast(\alpha_j, \alpha_{\ell}, \alpha_{\ell+1}) \leq g_j, \quad y_j^\ast(\alpha_j, \alpha_{\ell}, \alpha_{\ell+1}) \leq h_\ell.
\]

Putting (21) and (22) together, along with Assumption 2, we have that \(g_{\ell+1} < g_\ell \leq g_{\ell+1} < g_j\) and \(h_{\ell+1} \leq h_j, h_{\ell-1} < h_\ell \leq h_j\), which implies Pareto systems \(\ell - 1, \ell, \ell + 1\) dominate non-Pareto system \(j\) (see Figure 16).

![Fig. 16. The figure shows Pareto systems \(\ell - 1, \ell, \ell + 1\), and the associated phantom Pareto systems when the Pareto system \(\ell \in \{2, \ldots, p-1\}\). If the rate function \(I_\ell(\cdot)\) “drops out” of both problems, then the non-Pareto system \(j\) must be in the shaded region of the figure.](image)

From Proposition 4.2, \(\mathbb{I}_{[\ell \neq p]} = 0\) in Problem \(R_{j\ell-1}\), \(\mathbb{I}_{[\ell \neq p]} = 0\) in Problem \(R_{j\ell}\), and

\[
g_{\ell+1} < g_j \leq \frac{(h_\ell - h_{\ell+1}) p_j \sigma_j \sigma_h / \alpha_j}{\sigma_j^2 / \alpha_j + \sigma_h^2 / \alpha_\ell}, \quad h_{\ell-1} \leq h_j \leq \frac{(g_j - g_{\ell+1}) p_j \sigma_j \sigma_h / \alpha_j}{\sigma_j^2 / \alpha_j + \sigma_h^2 / \alpha_{\ell+1}}.
\]

(23)
If \( \rho_j \leq 0 \), the conditions in (23) cannot hold together, which provides a contradiction in this case. Now suppose \( \rho_j > 0 \). Then

\[
\frac{(h_j - h_{\ell-1})p_j\sigma_{j\ell}/\alpha_j}{\sigma_{j\ell}/\alpha_j + \sigma_{j\ell+1}/\alpha_{j+1}} \geq g_j - g_\ell > g_j - g_{\ell+1}, \quad \frac{(g_{j+1} - g_{\ell+1})p_j\sigma_{j\ell}/\alpha_j}{\sigma_{j\ell}/\alpha_j + \sigma_{j\ell+1}/\alpha_{j+1}} \geq h_j - h_\ell > h_j - h_{\ell-1},
\]

which implies

\[
\rho_j > \max\left\{ \frac{(g_{j+1} - g_{\ell+1})\sigma_{j\ell}/\alpha_j}{(h_j - h_{\ell-1})\sigma_{j\ell}/\alpha_j + (g_{j+1} - g_{\ell+1})\sigma_{j\ell+1}/\alpha_{j+1}}, \frac{(h_j - h_{\ell-1})\sigma_{j\ell+1}/\alpha_{j+1}}{(g_{j+1} - g_{\ell+1})\sigma_{j\ell}/\alpha_j + (h_j - h_{\ell-1})\sigma_{j\ell+1}/\alpha_{j+1}} \right\}. \tag{24}
\]

Since \( \max\{ \frac{(g_{j+1} - g_{\ell+1})\sigma_{j\ell}/\alpha_j}{(h_j - h_{\ell-1})\sigma_{j\ell}/\alpha_j + (g_{j+1} - g_{\ell+1})\sigma_{j\ell+1}/\alpha_{j+1}}, \frac{(h_j - h_{\ell-1})\sigma_{j\ell+1}/\alpha_{j+1}}{(g_{j+1} - g_{\ell+1})\sigma_{j\ell}/\alpha_j + (h_j - h_{\ell-1})\sigma_{j\ell+1}/\alpha_{j+1}} \} \geq 1 \), (24) cannot hold; we have a contradiction.

**G. PROOF OF THEOREM 4.4**

*Proof that \( \tilde{z}^* > 0, \tilde{\alpha}_j^* > 0 \) for all \( i \leq r \).* First, note that \( \tilde{\alpha}_k = 1/r \) for all \( k = 1, \ldots, r \) is a feasible solution for Problem \( \tilde{Q} \) that results in \( \tilde{z} > 0 \) under Lemma 3.5; therefore \( \tilde{z}^* > 0 \). If \( \tilde{\alpha}_j^* = 0 \) for some \( j \in \mathcal{P}^c \) in Problem \( \tilde{Q} \), then \( \tilde{z} = 0 \). Thus it must be the case that \( \tilde{\alpha}_j^* > 0 \) for all \( j \in \mathcal{P}^c \).

Now let \( j \in \mathcal{P}^c \), and consider \( R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*) \). If \( \ell = p \) and \( \tilde{\alpha}_p = 0 \), then \( \tilde{z} = 0 \). If \( \ell \neq p \) and \( \tilde{\alpha}_\ell = 0 \), then \( \tilde{z} = 0 \). Further, if there exists \( \ell \in \{1, \ldots, p-1\} \) such that \( \max\{\tilde{\alpha}_\ell, \tilde{\alpha}_{\ell+1}\} = 0 \), then \( \tilde{z} = 0 \). Therefore \( \tilde{\alpha}_1^* > 0, \tilde{\alpha}_p^* > 0 \), and for all \( \ell \in \{1, \ldots, p-1\} \), we must have \( \max\{\tilde{\alpha}_\ell, \tilde{\alpha}_{\ell+1}\} > 0 \) in each \( R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*) \).

Suppose \( p \geq 3 \) and for a contradiction, suppose there exists \( \ell \in \{2, \ldots, p-1\} \) such that \( \tilde{\alpha}_\ell^* = 0 \). Then \( \tilde{\alpha}_{\ell+1}^* > 0 \). Letting \( j \in \mathcal{P}^c \) and \( \tilde{\alpha}_j^* > 0 \), we have

\[
\tilde{z}^* \leq R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*) = \inf_{x_j \leq s_{\ell+1}, y_j \leq y_\ell} \tilde{\alpha}_j^* I_j(x_j, y_j) + \tilde{\alpha}_\ell^* I_\ell(x_\ell, y_\ell) + \tilde{\alpha}_{\ell+1}^* I_{\ell+1}(x_{\ell+1}, y_{\ell+1})
\]

\[
\leq \inf_{x_j \leq x_\ell, y_j \leq y_\ell} \tilde{\alpha}_j^* I_j(x_j, y_j) + \tilde{\alpha}_\ell^* I_\ell(x_\ell, y_\ell) + \tilde{\alpha}_{\ell+1}^* I_{\ell+1}(x_{\ell+1}, y_{\ell+1})
\]

\[
= \inf_{x_j \leq x_\ell, y_j \leq y_\ell} \tilde{\alpha}_j^* I_j(x_j, y_j) + \tilde{\alpha}_\ell^* I_\ell(x_\ell, y_\ell)
\]

\[
= \inf_{x_j \leq x_\ell, y_j \leq y_\ell} \tilde{\alpha}_j^* I_j(x_j, y_j) = 0,
\]

where the rate function in line (25) is the rate function corresponding to an MCE event in which non-Pareto system \( j \) is falsely estimated as dominating the Pareto system \( \ell \). Therefore since \( \tilde{z}^* > 0 \), we have a contradiction, and \( \tilde{\alpha}_j^* > 0 \) for all \( \ell = 1, \ldots, p \).

**KKT Conditions.** Let \( \nu, \lambda_\ell \geq 0 \) for all \( j \in \mathcal{P}^c, \ell = 0, \ldots, p \) be dual variables. Then the complementary slackness conditions are \( \lambda_j(R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*) - \tilde{z}^*) = 0 \) for all \( j \in \mathcal{P}^c, \ell = 0, \ldots, p \); and the stationarity conditions are

\[
\sum_{j \in \mathcal{P}^c} \left( \lambda_{\ell+1} \frac{\partial R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_{\ell+1}^*)}{\partial \alpha_j} + \lambda_\ell \frac{\partial R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*)}{\partial \alpha_j} \right) = \nu \quad \forall \ell = 1, \ldots, p; \tag{26}
\]

\[
\sum_{\ell=0}^p \lambda_\ell \frac{\partial R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*)}{\partial \alpha_j} = \nu \quad \forall j \in \mathcal{P}^c; \tag{27}
\]

\[
\sum_{j \in \mathcal{P}^c} \sum_{\ell=0}^p \lambda_\ell = 1. \tag{28}
\]
Proof of Part (1). Using the KKT conditions for Problem $R_{j\ell}$, we can simplify the stationarity conditions in (27). Then

$$
\frac{\partial R_{j}(\alpha_j, \omega_j, \omega_{j+1})}{\partial \alpha_j} = I_j(x_j^{*}, y_j^{*}) + \alpha_j \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_j} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_j} \right) + \alpha_{\ell} \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_{\ell}} \right) \mathbb{I}[\ell \neq 0] + \alpha_{\ell+1} \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_{\ell+1}} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_{\ell+1}} \right) \mathbb{I}[\ell \neq p]
$$

$$
= I_j(x_j^{*}, y_j^{*}) + \lambda_y \mathbb{I}[\ell \neq 0] \left( \frac{\partial y_j^{*}}{\partial \alpha_j} - \frac{\partial x_j^{*}}{\partial \alpha_j} \right) + \lambda_x \mathbb{I}[\ell \neq 0] \left( \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} - \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} \right)
$$

$$
= I_j(x_j^{*}, y_j^{*}) + \lambda_y \mathbb{I}[\ell \neq 0] \left( \frac{\partial y_j^{*}}{\partial \alpha_j} - \frac{\partial x_j^{*}}{\partial \alpha_j} \right) + \lambda_x \mathbb{I}[\ell \neq 0] \left( \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} - \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} \right)
$$

$$
= I_j(x_j^{*}(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1}), y_j^{*}(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1})) = I_j(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1}),
$$

(29)

where the penultimate equality holds by noticing that if $\lambda_y \mathbb{I}[\ell \neq 0] > 0$, then $x_j^{*} = x_j^{*}$, implying $\frac{\partial x_j^{*}}{\partial \alpha_j} = \frac{\partial x_j^{*}}{\partial \alpha_j}$. Likewise, if $\lambda_x \mathbb{I}[\ell \neq 0] > 0$, then $y_j^{*} = y_j^{*}$, implying $\frac{\partial y_j^{*}}{\partial \alpha_{\ell}} = \frac{\partial y_j^{*}}{\partial \alpha_{\ell}}$.

From (28), at least one of the dual variables in the sum must be strictly positive, and by Lemma 3.5, for all $j \in \mathcal{P}^c$, $I_j(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1})) > 0$ for all $\ell = 0, \ldots, p$. Therefore $\nu > 0$, which implies that for all $j \in \mathcal{P}^c$,

$$
\sum_{\ell=0}^{p} \lambda_j I_j(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1})) > 0.
$$

Thus for every $j \in \mathcal{P}^c$, there exists a phantom Pareto system $\ell^{*}(j) \in \{0, \ldots, p\}$ such that $\lambda_{j^{*}(j)} > 0$, and Part (1) of the Theorem holds.

Proof of Part (2). Using the KKT conditions for Problem $R_{j\ell-1}$, we have

$$
\frac{\partial R_{j}(\alpha_j, \omega_{j-1}, \omega_{j+1})}{\partial \alpha_j} = I_j(x_j^{*}, y_j^{*}) + \alpha_j \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_j} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_j} \right) + \alpha_{\ell} \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_{\ell}} \right) \mathbb{I}[\ell \neq 0] + \alpha_{\ell+1} \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_{\ell+1}} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_{\ell+1}} \right) \mathbb{I}[\ell \neq p]
$$

$$
= I_j(x_j^{*}, y_j^{*}) + \lambda_x \mathbb{I}[\ell \neq 0] \left( \frac{\partial x_j^{*}}{\partial \alpha_j} - \frac{\partial x_j^{*}}{\partial \alpha_j} \right) + \lambda_y \mathbb{I}[\ell \neq 0] \left( \frac{\partial y_j^{*}}{\partial \alpha_{\ell}} - \frac{\partial y_j^{*}}{\partial \alpha_{\ell}} \right)
$$

$$
= I_j(x_j^{*}(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1}), y_j^{*}(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1})) = I_j(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1}),
$$

where the penultimate equality holds by noticing that if $\lambda_x > 0$, then $x_j^{*} = x_j^{*}$, implying $\frac{\partial x_j^{*}}{\partial \alpha_j} = \frac{\partial x_j^{*}}{\partial \alpha_j}$. Likewise, if $\lambda_y > 0$, then $y_j^{*} = y_j^{*}$, implying $\frac{\partial y_j^{*}}{\partial \alpha_{\ell}} = \frac{\partial y_j^{*}}{\partial \alpha_{\ell}}$.

Using similar logic, from the KKT conditions for Problem $R_{j\ell}$, we have

$$
\frac{\partial R_{j}(\alpha_j, \omega_{j-1}, \omega_{j+1})}{\partial \alpha_j} = I_j(x_j^{*}, y_j^{*}) + \alpha_j \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_j} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_j} \right) + \alpha_{\ell} \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_{\ell}} \right) \mathbb{I}[\ell \neq 0] + \alpha_{\ell+1} \left( \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial x_j^{*}} \frac{\partial x_j^{*}}{\partial \alpha_{\ell+1}} + \frac{\partial I_j(x_j^{*}, y_j^{*})}{\partial y_j^{*}} \frac{\partial y_j^{*}}{\partial \alpha_{\ell+1}} \right) \mathbb{I}[\ell \neq p]
$$

$$
= I_j(x_j^{*}, y_j^{*}) + \lambda_y \mathbb{I}[\ell \neq 0] \left( \frac{\partial y_j^{*}}{\partial \alpha_j} - \frac{\partial x_j^{*}}{\partial \alpha_j} \right) + \lambda_x \mathbb{I}[\ell \neq 0] \left( \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} - \frac{\partial x_j^{*}}{\partial \alpha_{\ell}} \right)
$$

$$
= I_j(x_j^{*}(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1}), y_j^{*}(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1})) = I_j(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1}),
$$

Then since $\nu > 0$, from (26), for each Pareto system $\ell = 1, \ldots, p$

$$
\sum_{j \in \mathcal{P}^c} [\lambda_j I_j(\tilde{\alpha}_j, \tilde{\alpha}_{\ell-1}, \tilde{\alpha}_{\ell+1})) + \lambda_j I_j(\tilde{\alpha}_j, \tilde{\alpha}_{\ell}, \tilde{\alpha}_{\ell+1}))] > 0.
$$
Thus at least one constraint is binding for each Pareto system at optimality, and the result in Part (2) follows.

Proof of Parts (3) and (4). Updating the stationarity conditions in (26)–(28), we have
\[
\sum_{j \in \mathcal{P}} h_j \left( \lambda_{j \ell} - 1 \right) I_j(\tilde{a}_j^*, \tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*) = \nu \quad \forall \ell = 1, \ldots, p; \quad (30)
\]
\[
\sum_{\ell = 0}^p \lambda_{j \ell} I_j(\tilde{a}_j^*, \tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*) = \nu \quad \forall j \in \mathcal{P}; \quad (31)
\]
\[
\sum_{j \in \mathcal{P}} h_j \sum_{\ell = 0}^p \lambda_{j \ell} = 1. \quad (32)
\]
Substituting (31) into (30) yields (5); dividing yields (6) and (7).

H. PROOF OF LEMMA 4.5

We use Proposition 4.2 Part (2) and Assumptions 5–7 to derive bounds on the values of
\[ R_j(\tilde{a}_j^*, \tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*) \]
\[ I_j(\tilde{a}_j^*, \tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*) \]
\[ I_{\ell+1}(\tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*; \alpha_j, \alpha_\ell, \alpha_{\ell+1}) \] for the lower bound, we have
\[
R_j(\tilde{a}_j^*, \tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*) = \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_\ell^2}{\alpha_\ell} \right) (g_j - g_{\ell+1})^2 \left[ \frac{2}{\alpha_j} \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_\ell^2}{\alpha_\ell} \right) \right] (h_j - h_{\ell+1})^2 \left[ \frac{2}{\alpha_j} \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_\ell^2}{\alpha_\ell} \right) \right]
\]
\[
= \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_\ell^2}{\alpha_\ell} \right) \left[ \frac{1 - \rho_j}{\alpha_j} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right] \left[ \frac{1 - \rho_j}{\alpha_j} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right] \left[ \frac{1 - \rho_j}{\alpha_j} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right] \left[ \frac{1 - \rho_j}{\alpha_j} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right] \left[ \frac{1 - \rho_j}{\alpha_j} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right] \left[ \frac{1 - \rho_j}{\alpha_j} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right]
\]
By similar logic, the upper bound follows from the fact that
\[
R_j(\tilde{a}_j^*, \tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*) \leq \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_\ell^2}{\alpha_\ell} \right) \left[ \frac{1}{\alpha_j + \alpha_{\ell+1}} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right] \left[ \frac{1}{\alpha_j + \alpha_{\ell+1}} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right] \left[ \frac{1}{\alpha_j + \alpha_{\ell+1}} + \frac{1}{\alpha_j + \alpha_{\ell+1}} \right]
\]
Now we prove the bound on \[ I_j(\tilde{a}_j^*, \tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*) \]. First, for a lower bound, we have
\[
I_j(\tilde{a}_j^*, \tilde{a}_\ell^*, \tilde{a}_{\ell+1}^*) = \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_\ell^2}{\alpha_\ell} \right) \left[ \frac{2}{\alpha_j} \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_\ell^2}{\alpha_\ell} \right) \right] (g_j - g_{\ell+1})^2 \left[ \frac{2}{\alpha_j} \left( \frac{\sigma_j^2}{\alpha_j} + \frac{\sigma_\ell^2}{\alpha_\ell} \right) \right]
\]
\[\begin{align*}
&\geq \frac{1}{\alpha_j^2} \left( \frac{\sigma_h^2}{2\alpha_j^2} \right)\left\{ \left[ \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 - \frac{\sigma_h^2}{\alpha_j^2} \right] (g_j - g_{j+1})^2 l_{[j\mid j+1]} + \left[ \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 - \frac{\sigma_h^2}{\alpha_j^2} \right] (h_j - h_{j+1})^2 l_{[j\mid j+1]} \right\} \\
&\geq \frac{1}{\alpha_j^2} \left( \frac{\sigma_h^2}{2\alpha_j^2} (1 - \rho_j^2) \right) \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 l_{[j\mid j+1]} - 2\rho_j \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right) l_{[j\mid j+1]} \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 l_{[j\mid j+1]} \\
&\geq \frac{1}{\alpha_j^2} \left( \frac{\sigma_h^2}{2\alpha_j^2} (1 - \rho_j^2) \right) \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 l_{[j\mid j+1]} - 2\rho_j \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right) l_{[j\mid j+1]} \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 l_{[j\mid j+1]},
\end{align*}\]

and the result follows. For an upper bound, by similar logic, we have

\[
I_j(g_j((\widehat{\alpha}_j^*, \widehat{\alpha}_j^*, \widehat{\alpha}_j^*_{j+1})))
\]

\[
\leq \frac{1}{\alpha_j^2} \left( \frac{\sigma_h^2}{2\alpha_j^2} (1 - \rho_j^2) \right) \left\{ \left[ \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 - \frac{\sigma_h^2}{\alpha_j^2} \right] (g_j - g_{j+1})^2 l_{[j\mid j+1]} + \left[ \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 - \frac{\sigma_h^2}{\alpha_j^2} \right] (h_j - h_{j+1})^2 l_{[j\mid j+1]} \right\} \\
\leq \frac{1}{\alpha_j^2} \left( \frac{\sigma_h^2}{2\alpha_j^2} (1 - \rho_j^2) \right) \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 l_{[j\mid j+1]} + 2\rho_j \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right) l_{[j\mid j+1]} \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 l_{[j\mid j+1]},
\]

and the result follows.

Now we prove the bounds on \(I_j(g_j((\widehat{\alpha}_j^*, \widehat{\alpha}_j^*, \widehat{\alpha}_j^*_{j+1})))\). First, for the lower bound we have

\[
I_j(g_j((\widehat{\alpha}_j^*, \widehat{\alpha}_j^*, \widehat{\alpha}_j^*_{j+1}))) = \left( \frac{\sigma_h^2}{\alpha_j^2} (\gamma_j + \rho_j g_j - h_j) \right)^2 l_{[j\mid j+1]} + \left( \frac{\sigma_h^2}{\alpha_j^2} (\gamma_j + \rho_j g_j - h_j) \right)^2 l_{[j\mid j+1]} \]

\[
\leq \frac{1}{\alpha_j^2} \left( \frac{\sigma_h^2}{2\alpha_j^2} \right) \left\{ \left[ \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 - \frac{\sigma_h^2}{\alpha_j^2} \right] (g_j - g_{j+1})^2 l_{[j\mid j+1]} + \left[ \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 - \frac{\sigma_h^2}{\alpha_j^2} \right] (h_j - h_{j+1})^2 l_{[j\mid j+1]} \right\} \\
\leq \frac{1}{\alpha_j^2} \left( \frac{\sigma_h^2}{2\alpha_j^2} \right) \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 l_{[j\mid j+1]} + 2\rho_j \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right) l_{[j\mid j+1]} \left( \frac{1}{\alpha_j^2} + \frac{1}{\alpha_j} \right)^2 l_{[j\mid j+1]},
\]
\[
\geq \left(\frac{\sigma^2}{2\pi^6}\right) \left(\frac{1}{\pi^6}\right) \left\{ \frac{1}{\pi^6} \right\} \left(\frac{1}{\pi^6} \right) \left(\frac{1}{\pi^6} \right) \left(\frac{1}{\pi^6} \right) \left(\frac{1}{\pi^6} \right) \left(\frac{1}{\pi^6} \right)
\]

For the upper bound, we have

\[
I_{\ell}(\mathcal{A}((\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*))) \leq \left(\frac{\sigma^2}{2\pi^6}\right) \left(\frac{1}{\pi^6}\right) \left\{ \frac{1}{\pi^6} \right\} \left(\frac{1}{\pi^6} \right) \left(\frac{1}{\pi^6} \right) \left(\frac{1}{\pi^6} \right) \left(\frac{1}{\pi^6} \right) \left(\frac{1}{\pi^6} \right)
\]

We omit the proof for the bounds on \( I_{\ell+1}(\mathcal{A}((\alpha_j, \alpha_\ell, \alpha_{\ell+1})) \), since it is similar to the proof for the bounds on \( I_{\ell}(\mathcal{A}((\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*))) \).

I. PROOF OF PROPOSITION 4.6

Using the upper bound in (11), at optimality in Problem \( \bar{Q} \) for each non-Pareto \( j \in \mathcal{P}^\ell \) and Pareto \( \ell \in \{1, \ldots, p\} \), letting \( c_1 := \frac{\beta^2\sigma^2(1+\rho_0)}{2\sigma^2(1-\rho_0)^2} \), we have

\[
|\mathcal{P}^\ell| \bar{z}^* \leq \sum_{j \in \mathcal{P}^\ell} R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*) \leq \sum_{j \in \mathcal{P}^\ell} \tilde{\alpha}_j^* c_1 \frac{(1+\tilde{\alpha}_j^*/\tilde{\alpha}_\ell^*)|\bar{x}_j|+(1+\tilde{\alpha}_j^*/\tilde{\alpha}_{\ell+1}^*)|\bar{x}_{\ell+1}|}{(1+\tilde{\alpha}_j^*/\tilde{\alpha}_\ell^*)+(1+\tilde{\alpha}_j^*/\tilde{\alpha}_{\ell+1}^*)}
\]

which implies \( \bar{z}^* \leq 2c_1/|\mathcal{P}^\ell| \). Thus \( \bar{z}^* = O(1/|\mathcal{P}^\ell|) \).

J. PROOF OF LEMMA 4.7

We provide proofs for the parts as follows.

Proof of Part (1). Let \( \ell \in \{1, \ldots, p\} \) be a Pareto system. Using the lower bound in (11), for all \( j \in \mathcal{P}^\ell \), we have

\[
\bar{z}^* \geq \frac{1}{2}(R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*)|\bar{x}_j|+R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*)|\bar{x}_{\ell+1}|)
\]

Now from Problem \( \bar{Q} \) and using the upper bound in (11), for all \( j \in \mathcal{P}^\ell, \ell \in \{1, \ldots, p\} \),

\[
\bar{z}^* \leq R_j(\tilde{\alpha}_j^*, \tilde{\alpha}_\ell^*, \tilde{\alpha}_{\ell+1}^*) \leq \inf_{x_j \leq g_{j+1}, y_j \leq y_{j+1}} \tilde{\alpha}_j^* I_j(x_j, y_j) + \tilde{\alpha}_\ell^* I_{\ell+1}(x_\ell, y_\ell) \leq \frac{c_1}{1+\tilde{\alpha}_j^*/\tilde{\alpha}_\ell^*}
\]

Let \( \kappa_{1a} := \frac{\sigma^2(1+\rho_0)}{4\sigma^2} \), \( \kappa_{1b} := \frac{\beta^2\sigma^2(1+\rho_0)}{2\sigma^2(1-\rho_0)^2} \). Combining (33) and (34) yields the result.

Proof of Part (2). This part follows directly from Part (1).
Proof of Part (3). Let \( j, j' \in \mathcal{P}_c \). Using the KKT condition in (7), the bounds in (12), and the results in Part (1) and Part (2), we have

\[
1 = \frac{\sum_{\ell=0}^p \lambda_{j\ell} I_j(\tilde{x}_j^\star, \tilde{\alpha}_j^\star, \tilde{\alpha}_{j+1}^\star)}{\sum_{\ell=0}^p \lambda_{j\ell} I_j(\tilde{x}_j^\star, \tilde{\alpha}_j^\star, \tilde{\alpha}_{j+1}^\star)} \geq \left( \frac{\beta^2 \sigma_{b}^4(1 + \rho_b^2)}{\varepsilon^2 \sigma_{a}^4(1 - \rho_a^2)^2} \right) \sum_{\ell=0}^p \frac{\lambda_{j\ell}}{\sum_{\ell=0}^p \lambda_{j\ell}} \left[ \frac{\sum_{\ell=0}^p \lambda_{j\ell}}{\sum_{\ell=0}^p \lambda_{j\ell} I_j(\tilde{x}_j^\star, \tilde{\alpha}_j^\star, \tilde{\alpha}_{j+1}^\star)} \right] \geq \left( \frac{\beta^2 \sigma_{b}^4(1 + \rho_b^2)}{\varepsilon^2 \sigma_{a}^4(1 - \rho_a^2)^2} \right) \sum_{\ell=0}^p \frac{\lambda_{j\ell}}{\sum_{\ell=0}^p \lambda_{j\ell}} \left[ \frac{\sum_{\ell=0}^p \lambda_{j\ell}}{\sum_{\ell=0}^p \lambda_{j\ell} I_j(\tilde{x}_j^\star, \tilde{\alpha}_j^\star, \tilde{\alpha}_{j+1}^\star)} \right] \geq \left( \frac{\beta^2 \sigma_{b}^4(1 + \rho_b^2)}{\varepsilon^2 \sigma_{a}^4(1 - \rho_a^2)^2} \right) \sum_{\ell=0}^p \frac{\lambda_{j\ell}}{\sum_{\ell=0}^p \lambda_{j\ell}} \left[ \frac{\sum_{\ell=0}^p \lambda_{j\ell}}{\sum_{\ell=0}^p \lambda_{j\ell} I_j(\tilde{x}_j^\star, \tilde{\alpha}_j^\star, \tilde{\alpha}_{j+1}^\star)} \right],
\]

where the step for the numerator in (35) follows from line (34). Letting the constant 
\[
\kappa_3 := \frac{\beta^2 \sigma_{b}^4(1 + \rho_b^2)}{\varepsilon^2 \sigma_{a}^4(1 - \rho_a^2)^2} \min\{\kappa_2, \rho_b^2, (\kappa_2 - \rho_b)^2\} \frac{2\varepsilon^2}{\kappa_2^2} \sum_{\ell=0}^p \lambda_{j\ell},
\]
implies the result.

Proof of Part (4). From (6) and (13)–(14), for all \( \ell, \ell' \in \{1, \ldots, p\} \), we have

\[
1 = \frac{\sum_{j \in \mathcal{P}_{\ell}(r)} \lambda_{j\ell-1} I_j(\tilde{x}_j^\star, \tilde{\alpha}_{j-1}^\star, \tilde{\alpha}_j^\star)}{\sum_{j' \in \mathcal{P}_{\ell'}(r')} \lambda_{j'\ell'-1} I_{j'}(\tilde{x}_{j'}^\star, \tilde{\alpha}_{j'-1}^\star, \tilde{\alpha}_{j'}^\star)} \geq \left( \frac{\varepsilon^2 \sigma_{a}^4(1 + \rho_a^2)^2}{2 \rho_a^2} \right) \sum_{j \in \mathcal{P}_{\ell}(r)} \lambda_{j\ell-1} I_j(\tilde{x}_j^\star, \tilde{\alpha}_{j-1}^\star, \tilde{\alpha}_j^\star) \geq \left( \frac{\rho_a^2 \varepsilon^2 \sigma_{a}^4(1 + \rho_a^2)^2}{2 \rho_a^2} \right) \sum_{j \in \mathcal{P}_{\ell}(r)} \lambda_{j\ell-1} I_j(\tilde{x}_j^\star, \tilde{\alpha}_{j-1}^\star, \tilde{\alpha}_j^\star) \geq \left( \frac{\rho_a^2 \varepsilon^2 \sigma_{a}^4(1 + \rho_a^2)^2}{2 \rho_a^2} \right) \sum_{j \in \mathcal{P}_{\ell}(r)} \lambda_{j\ell-1} I_j(\tilde{x}_j^\star, \tilde{\alpha}_{j-1}^\star, \tilde{\alpha}_j^\star),
\]

which, letting \( \kappa_4 := \frac{\rho_a^2 \varepsilon^2 \sigma_{a}^4(1 + \rho_a^2)^2}{2 \rho_a^2} \), implies the result.

Proof of Part (5). The KKT condition in (5) is equivalent to

\[
\frac{\lambda_{j\ell-1} I_j(\tilde{x}_j^\star, \tilde{\alpha}_{j-1}^\star, \tilde{\alpha}_j^\star)}{\sum_{\ell' = 0}^p \lambda_{j\ell'} I_j(\tilde{x}_j^\star, \tilde{\alpha}_{j-1}^\star, \tilde{\alpha}_j^\star)} = 1.
\]

Then for \( j \in \mathcal{P}_{\ell}(r) \), a lower bound on the inner term in the summation in (36) is

\[
\frac{\lambda_{j\ell-1} I_j(\tilde{x}_j^\star, \tilde{\alpha}_{j-1}^\star, \tilde{\alpha}_j^\star)}{\sum_{\ell' = 0}^p \lambda_{j\ell'} I_j(\tilde{x}_j^\star, \tilde{\alpha}_{j-1}^\star, \tilde{\alpha}_j^\star)} \geq \left( \frac{\varepsilon^2 \sigma_{a}^4(1 + \rho_a^2)^2}{2 \rho_a^2} \right) \sum_{\ell = 0}^p \frac{\lambda_{j\ell}}{\sum_{\ell = 0}^p \lambda_{j\ell}} \left[ \frac{\sum_{\ell = 0}^p \lambda_{j\ell}}{\sum_{\ell = 0}^p \lambda_{j\ell} I_j(\tilde{x}_j^\star, \tilde{\alpha}_j^\star, \tilde{\alpha}_{j+1}^\star)} \right] \geq \left( \frac{\varepsilon^2 \sigma_{a}^4(1 + \rho_a^2)^2}{2 \rho_a^2} \right) \sum_{\ell = 0}^p \frac{\lambda_{j\ell}}{\sum_{\ell = 0}^p \lambda_{j\ell}} \left[ \frac{\sum_{\ell = 0}^p \lambda_{j\ell}}{\sum_{\ell = 0}^p \lambda_{j\ell} I_j(\tilde{x}_j^\star, \tilde{\alpha}_j^\star, \tilde{\alpha}_{j+1}^\star)} \right].
\]
and, as before, applying Part (3) yields the result.

We provide proofs for the theorem parts as follows.

K. PROOF OF THEOREM 4.8

We prove theorems for the parts as follows.

Proof of Part (1). From the proof of Lemma 4.7 Part (5), for all \( r \) such that \( j \in \mathcal{P}_r(\ell, r) \), \( \ell \in \{1, \ldots, p\} \) we have \( \tilde{\alpha}_j^{2}/\tilde{\alpha}_j^{2} \geq \kappa_{n_{j, r}} |\lambda_{j, \ell, r - 1} [g_{j, \ell} - 1] + \lambda_{j, \ell} [g_{j, \ell} r] | / \sum_{j' = 0}^{p} |\lambda_{j, \ell} r] |. \) Since \( \ell \) is the primary phantom Pareto for \( j \), then \( j \in \mathcal{P}_c(\ell) \) or \( j \in \mathcal{P}_c(\ell + 1) \). If \( j \in \mathcal{P}_c(\ell) \), then for large enough \( r \), we have

\[
\frac{\tilde{\alpha}_j^{2}}{\alpha_j^{2}} \geq \left( \frac{\kappa_{n_{j, r}}}{(p + 1) \kappa_0} \right) \lambda_{j, \ell, r - 1} r] [g_{j, \ell} - 1] + \lambda_{j, \ell} r] [g_{j, \ell} r] ] \geq \left( \frac{\kappa_{n_{j, r}}}{(p + 1) \kappa_0} \right) r [g_{j, \ell} r] ].
\]

If \( j \in \mathcal{P}_c(\ell + 1) \), then for large enough \( r \), we have

\[
\frac{\tilde{\alpha}_j^{2}}{\alpha_j^{2}} \geq \left( \frac{\kappa_{n_{j, r}}}{(p + 1) \kappa_0} \right) \lambda_{j, \ell, r - 1} r] [g_{j, \ell} - 1] + \lambda_{j, \ell} r - 1 r] [g_{j, \ell} r] ] \geq \left( \frac{\kappa_{n_{j, r}}}{(p + 1) \kappa_0} \right) r [g_{j, \ell} r] ].
\]

Letting \( \tau_j := \frac{\kappa_{n_{j, r}}}{(p + 1) \kappa_0} \) implies the result.
Proof of Part (2). From (12) and using Part (1), for all \( r \geq r_{j1} \) we have

\[
I_j(\tilde{s}_j^*(\tilde{a}_{j}^*, \tilde{a}_{j}^*, \tilde{a}_{j}^* + \alpha)) \geq \left( \frac{e^2 \sigma^2_j(1-\rho_j^2)}{2 \sigma_j^2} \right) \left[ \frac{\|g_{j,\tau}^\epsilon\|}{1 + \frac{\lambda_j^T g_{j,\tau}^\epsilon}{\alpha_j^T \alpha_j^T + 1}} \right] \geq \left( \frac{e^2 \sigma^2_j(1-\rho_j^2)}{2 \sigma_j^2} \right) \left[ \frac{\|g_{j,\tau}^\epsilon\|}{1 + \frac{\lambda_j^T g_{j,\tau}^\epsilon}{\alpha_j^T \alpha_j^T + 1}} \right].
\]

Letting \( \tau_2 := \frac{e^2 \sigma^2_j(1-\rho_j^2)}{2 \sigma_j^2(1+\tau_1)^2} \) implies the result.

Proof of Part (3). From Assumption 8 and Lemma 4.7 Part (3), there exists \( \tau_{5a} \) such that, for all \( r \geq \max\{r_{j0}, r_{j1}\} \), \( \tilde{a}_j^2 / \tilde{a}_j^2 \geq \tau_{5a} \lambda_{j^r} / \lambda_{j^r} \). Under Assumptions 5–7, there exists \( b < \infty \) such that \( I_j(\tilde{s}_j^*(\tilde{a}_{j}^*, \tilde{a}_{j}^*, \tilde{a}_{j}^* + \alpha)) \leq b \) for all \( j \in \mathcal{P}_c \), all \( \ell \in \{0, \ldots, p\} \), and all \( r \).

Using Part (2) and equation (31), for all \( j \in \mathcal{P}_c \) and all \( r \geq \max\{r_{j0}, r_{j1}\} \),

\[
\lambda_{j^r} \tau_2 \leq \nu = \sum_{\ell=0}^p \lambda_j I_j(\tilde{s}_j^*(\tilde{a}_{j}^*, \tilde{a}_{j}^*, \tilde{a}_{j}^* + \alpha)) \leq (p + 1) \lambda_0 \nu, \tag{38}
\]

which implies \( \lambda_{j^r} = \Theta(\nu) \). Now for all \( j, j' \in \mathcal{P}_c \) and all \( r \geq \max\{r_{j0}, r_{j1}, r_{j'0}, r_{j'1}\} \),

\[
\frac{\tilde{a}_j^2}{\tilde{a}_j^2} \geq \tau_{3a} \lambda_{j^r} / \lambda_{j'^r} \geq \tau_{3a} \tau_2 \nu / (p + 1) \lambda_0 \nu,
\]

which implies the result.

Proof of Part (4). From Part (3), it follows that \( \tilde{a}_j^* \tau_{3} > \tilde{a}_j^* \), for all \( j \in \mathcal{P}_c \) and \( j' \in \mathcal{P}_c \), which implies \( \sum_{\ell=1}^p \tilde{a}_j^* + |\mathcal{P}_c| \tilde{a}_j^* \tau_{3} > \sum_{\ell=1}^p \tilde{a}_j^* + \sum_{j' \in \mathcal{P}_c} \tilde{a}_j^* = 1 \). Then it follows that \( \tilde{a}_j^* < ((1 - \sum_{\ell=1}^p \tilde{a}_j^*) \tau_{3}) / |\mathcal{P}_c| \), which implies the upper bound for the result. Likewise, for a lower bound, \( \tilde{a}_j^* < \tau_{5a} \tilde{a}_j^* \) for all \( j \in \mathcal{P}_c \) and \( j' \in \mathcal{P}_c \) implies \( \sum_{\ell=1}^p \tilde{a}_j^* + |\mathcal{P}_c| \tilde{a}_j^* \tau_{3} > \sum_{\ell=1}^p \tilde{a}_j^* + \sum_{j' \in \mathcal{P}_c} \tilde{a}_j^* = 1 \). Therefore \( \tilde{a}_j^* > ((1 - \sum_{\ell=1}^p \tilde{a}_j^*) \tau_{3}^{-1}) / |\mathcal{P}_c| \), which implies the lower bound.

Proof of Part (5). First, notice that summing over \( \ell \in \{1, \ldots, p\} \) and \( j \in \mathcal{P}_c \) in (38) implies that \( \nu = \Theta(1/|\mathcal{P}_c|) \), and hence \( \lambda_{j^r} = \Theta(1/|\mathcal{P}_c|) \) for all \( j \in \mathcal{P}_c \), and there exists \( \tau_{5a} > 0 \) such that \( \lambda_{j^r} \geq \tau_{5a} / |\mathcal{P}_c| \).

From Assumption 8 and Lemma 4.7 Part (4), for all \( \ell, \ell' \in \{1, \ldots, p\} \), we have

\[
\frac{\tilde{a}_j^2}{\tilde{a}_j^2} \geq \frac{\kappa_4}{2} \sum_{j \in \mathcal{P}_c} \frac{\|g_{j,\tau}^\epsilon\|}{\|g_{j,\tau}^\epsilon\|} \geq \frac{\kappa_4}{2} \sum_{j \in \mathcal{P}_c} \frac{\|g_{j,\tau}^\epsilon\|}{\|g_{j,\tau}^\epsilon\|} \geq \frac{\kappa_4}{2} \frac{|\mathcal{P}_c|\tau_{5a}}{|\mathcal{P}_c|} \geq \frac{\kappa_4 \tau_{5a} K}{2},
\]

which, letting \( \tau_3 := 2/(\kappa_4 \tau_{5a} K) \), implies the result.

Proof of Part (6). Since \( \lambda_{j^r} = \Theta(1/|\mathcal{P}_c|) \) for all \( j \in \mathcal{P}_c \), there exists \( \tau_{6a} \) such that \( \lambda_{j^r} \leq \tau_{6a} / |\mathcal{P}_c| \). From Assumption 8 and Part (5),

\[
\frac{\tilde{a}_j^2}{\tilde{a}_j^2} \geq \frac{\kappa_5a}{p + 1} \frac{\|g_{j,\tau}^\epsilon\|}{\|g_{j,\tau}^\epsilon\|} \geq \frac{\kappa_5a \tau_{5a} K}{(p + 1) \kappa_0 \lambda_{j^r}} \geq \frac{\kappa_5a \tau_{5a} K}{(p + 1) \kappa_0 \tau_{6a}} |\mathcal{P}_c|,
\]

which implies the result.

Proof of Part (7). This result follows from the bounds in (13) and (14) and Part (6).

Proof of Part (8). This result follows from Part (7).
L. PROOF OF PROPOSITION 4.11
Under our assumptions, the KKT conditions are necessary and sufficient for global optimality when finding $\inf_{x \leq g_1, y \leq h} I_j(x, y)$; for notational simplicity, we drop the subscripts on the variables $x$ and $y$. Let $\lambda_g \geq 0$ and $\lambda_h \geq 0$ be dual variables associated with the first and second constraints, respectively. In addition to primal feasibility, we have the complementary slackness conditions $\lambda_g (x - g + 1) = 0$ if $\ell \neq p$ and $\lambda_h (y - h) = 0$ if $\ell = 0$, and the stationarity conditions

$$
\frac{1}{(1 - p)} \left( \frac{x - g}{\sigma_j} - \frac{\rho_j (y - h)}{\sigma_j \sigma_j} \right) = -\lambda_g \mathbb{I}_{\ell \neq p}, \quad \frac{1}{(1 - p)} \left( \frac{y - h}{\sigma_j} - \frac{\rho_j (x - g)}{\sigma_j \sigma_j} \right) = -\lambda_h \mathbb{I}_{\ell = 0}.
$$

Since $\lambda_g \geq 0$ and $\lambda_h \geq 0$, it follows that

$$
x \leq g_j + \rho_j \sigma_j (y - h_j) \quad \text{and} \quad y \leq h_j + \rho_j \sigma_j (x - g_j).
$$

(39)

We now consider three cases: $\lambda_g \mathbb{I}_{\ell \neq p} > 0$ and $\lambda_h \mathbb{I}_{\ell \neq 0} = 0$, $\lambda_g \mathbb{I}_{\ell \neq p} = 0$ and $\lambda_h \mathbb{I}_{\ell \neq 0} > 0$, and $\lambda_g \mathbb{I}_{\ell \neq p} > 0$ and $\lambda_h \mathbb{I}_{\ell \neq 0} > 0$.

**Case 1:** $\lambda_g \mathbb{I}_{\ell \neq p} > 0$ and $\lambda_h \mathbb{I}_{\ell \neq 0} = 0$. Then $(x^*, y^*) = (g_{\ell + 1}, h_j + \rho_j \sigma_j \sigma_j (g_{\ell + 1} - g_j))$, which implies that the optimal value of the rate function is $I_j(x^*, y^*) = I_{g_j}(g_{\ell + 1})$. By (39) and primal feasibility, respectively, $g_j > g_{\ell + 1}$ and $h_j \leq h_\ell \leq \rho_j \sigma_j \sigma_j (g_j - g_{\ell + 1})$.

**Case 2:** $\lambda_g \mathbb{I}_{\ell \neq p} = 0$ and $\lambda_h \mathbb{I}_{\ell \neq 0} > 0$. Then $(x^*, y^*) = (g_j + \rho_j \sigma_j \sigma_j (h_j - h_\ell), h_\ell)$, which implies that the optimal value of the rate function is $I_j(x^*, y^*) = I_{h_j}(h_\ell)$. By (39) and primal feasibility, respectively, $h_j > h_\ell$ and $g_j \leq g_{\ell + 1} + \rho_j \sigma_j \sigma_j (h_j - h_\ell)$.

**Case 3:** $\lambda_g \mathbb{I}_{\ell \neq p} > 0$ and $\lambda_h \mathbb{I}_{\ell \neq 0} > 0$. Then $(x^*, y^*) = (g_{\ell + 1}, h_\ell)$, which implies that the optimal value of the rate function is $I_j(x^*, y^*) = I_{g_{\ell + 1}, h_\ell}$. By (39), it follows that $g_j > g_{\ell + 1} + \rho_j \sigma_j \sigma_j (h_j - h_\ell)$ and $h_j > h_\ell + \rho_j \sigma_j \sigma_j (g_j - g_{\ell + 1})$.

It can also be shown that the rate functions depend on the locations of the systems, as was shown in the proof of Proposition 4.2. We do not provide the details of the proof in this case.

M. PROOF OF THEOREM 4.12
From the proof of Theorem 4.8 Part (6), there exists $\eta_{1a} > 0$ such that $\eta_{1a} \check{\alpha}_j^2 | \mathbb{P}^c | \leq \check{\alpha}_j^2$.

(5)

From Lemma 4.7 Part (5), it follows that

$$
\frac{\check{\alpha}_j^2}{\hat{\alpha}_j^2} \leq \kappa_{5b} \sum_{j \in \mathbb{P}^c, \ell \neq p} \lambda_{\ell - 1} I_{[j - 1]} / \sum_{j = 0}^p \lambda_{\ell} \leq \kappa_{5b} | \mathbb{P}^c | 2 \kappa_{10} \lambda_{\ell} / \lambda_{\ell},
$$

so that letting $\eta_{1b} := 2 \kappa_{5b} \kappa_{10}$ implies $\eta_{1a} \check{\alpha}_j^2 | \mathbb{P}^c | \leq \check{\alpha}_j^2 \leq \eta_{1b} \hat{\alpha}_j^2 | \mathbb{P}^c |$. Summing across $j \in \mathbb{P}^c$ yields $\eta_{1a} \sum_{j \in \mathbb{P}^c} \check{\alpha}_j^2 \leq \check{\alpha}_j^2 \leq \eta_{1b} \hat{\alpha}_j^2$. Applying that $\check{\alpha}_j^2 = \Theta(1/| \mathbb{P}^c |)$ from Theorem 4.8 Part (4) yields the result.

N. PROOF OF THEOREM 4.13
Recall that the only difference between Problems $Q$ and $\tilde{Q}$ is that Problem $Q$ includes constraints corresponding to the rate functions for MCE events between all pairs of Pareto systems. Recall that for all $i, k \in \mathbb{P}, k \neq i$, we have

$$
R_i(\alpha_i, \alpha_k) = \inf_{x_k \leq x_i, y_k \leq y_i} \alpha_i I_i(x_i, y_i) + \alpha_k I_k(x_k, y_k).
$$

Under our assumptions, notice that there exists a constant $a > 0$ such that $\inf_{x_k \leq x_i, y_k \leq y_i} I_i(x_i, y_i) + I_k(x_k, y_k) > a$ for all $i, k \in \mathbb{P}$. From Theorem 4.8 Part (5),
for large enough values of $|\mathcal{P}|$, $\tilde{\alpha}_i^*/\tilde{\alpha}_k^* \leq \tau_5$ for all $i, k \in \mathcal{P}$. Thus supposing without loss of generality that $\tau_5 > 1$,

$$R_i(\tilde{\alpha}_i^*, \tilde{\alpha}_k^*)/\tilde{\alpha}_i^* = \inf_{x_k \leq x_i, y_k \leq y_i} I_i(x_i, y_i) + (\tilde{\alpha}_k^*/\tilde{\alpha}_i^*) I_k(x_k, y_k) \geq \inf_{x_k \leq x_i, y_k \leq y_i} I_i(x_i, y_i) + \tau_5^{-1} I_k(x_k, y_k) \geq \tau_5^{-1} a > 0.$$ 

However, since it follows from Theorem 4.8 that $\tilde{z}^* = \Theta(1/|\mathcal{P}|)$ while $\tilde{\alpha}_i^* = \Theta(1/\sqrt{|\mathcal{P}|})$, then $\tilde{z}^*/\tilde{\alpha}_i^* \to \infty$. Thus for large enough $|\mathcal{P}|$, the Pareto systems receive a large enough portion of the allocation that the constraints corresponding to MCE in Problem $Q$ are not binding, and the result holds.

**O. PROOF OF PROPOSITION 5.1**

Under Assumption 7, Problem $R_{i,k}^p$ is a quadratic program with linear constraints. The KKT conditions are necessary and sufficient for global optimality. Let $\lambda_x^p \geq 0$ and $\lambda_y^p \geq 0$ be dual variables. In addition to primal feasibility, we have the complementary slackness conditions $\lambda_x^p(x_k - x_i) = 0$ and $\lambda_y^p(y_k - y_i) = 0$, and the stationarity conditions

$$\alpha_i \frac{\partial I_i(x_i, y_i)}{\partial x_i} - \lambda_x^p = 0 \quad \alpha_k \frac{\partial I_k(x_k, y_k)}{\partial x_k} + \lambda_x^p = 0 \quad \alpha_i \frac{\partial I_i(x_i, y_i)}{\partial y_i} - \lambda_y^p = 0 \quad \alpha_k \frac{\partial I_k(x_k, y_k)}{\partial y_k} + \lambda_y^p = 0,$$

which simplify to

$$\frac{\alpha_i}{(1-\rho_i^2)} \left( \frac{x_i - g_i}{\sigma_{g_i}} - \frac{\rho_i(y_i - h_i)}{\sigma_{g_i} \sigma_{h_i}} \right) = \lambda_x^p \quad \frac{\alpha_k}{(1-\rho_k^2)} \left( \frac{x_k - g_k}{\sigma_{g_k}} - \frac{\rho_k(y_k - h_k)}{\sigma_{g_k} \sigma_{h_k}} \right) = -\lambda_x^p \quad \frac{\alpha_i}{(1-\rho_i^2)} \left( \frac{y_i - h_i}{\sigma_{h_i}} - \frac{\rho_i(x_i - g_i)}{\sigma_{g_i} \sigma_{h_i}} \right) = \lambda_y^p \quad \frac{\alpha_k}{(1-\rho_k^2)} \left( \frac{y_k - h_k}{\sigma_{h_k}} - \frac{\rho_k(x_k - g_k)}{\sigma_{g_k} \sigma_{h_k}} \right) = -\lambda_y^p.$$

Since $\lambda_x^p \geq 0$ and $\lambda_y^p \geq 0$, then

$$x_i \geq g_i + \rho_i \frac{\sigma_{g_i}}{\sigma_{h_i}} (y_i - h_i) \quad y_i \geq h_i + \rho_i \frac{\sigma_{h_i}}{\sigma_{g_i}} (x_i - g_i) \quad x_k \leq g_k + \rho_k \frac{\sigma_{g_k}}{\sigma_{h_k}} (y_k - h_k) \quad y_k \leq h_k + \rho_k \frac{\sigma_{h_k}}{\sigma_{g_k}} (x_k - g_k).$$

Since $i, k \in \mathcal{P}$, we cannot have $\lambda_x^p = 0$ and $\lambda_y^p = 0$. We consider three cases, as follows.

**Case 1: $\lambda_x^p > 0$ and $\lambda_y^p > 0$**

Then $x_i^* = x_k^*$; solving, we find

$$x_i^* = x_k^* = \frac{(\alpha_i/\sigma_{g_i}^2) g_i + (\alpha_k/\sigma_{g_k}^2) g_k}{\alpha_i/\sigma_{g_i}^2 + \alpha_k/\sigma_{g_k}^2} \quad y_i^* = y_k^* = \frac{\rho_i (\sigma_{g_i}/\sigma_{h_i}) (\alpha_i/\sigma_{g_i}^2 + \alpha_k/\sigma_{g_k}^2) (g_k - g_i)}{\alpha_i/\sigma_{g_i}^2 + \alpha_k/\sigma_{g_k}^2}.$$ 

which implies $g_i \leq g_k$. Primal feasibility further implies

$$h_i \geq h_k - (g_k - g_i) \left( \rho_i (\sigma_{g_i}/\sigma_{h_i}) (\alpha_i/\sigma_{g_i}^2 + \alpha_k/\sigma_{g_k}^2) (g_k - g_i) \right).$$

Substituting into the objective function yields the result in this case.

**Case 2: $\lambda_x^p > 0$ and $\lambda_y^p = 0$**

Then $y_i^* = y_k^*$; solving, we find

$$x_i^* = g_i + \rho_i \frac{\sigma_{g_i}}{\sigma_{h_i}} \frac{(\alpha_i/\sigma_{g_i}^2) (h_i - h_i)}{\alpha_i/\sigma_{g_i}^2 + \alpha_k/\sigma_{g_k}^2} \quad y_i^* = y_k^* = \frac{(\alpha_i/\sigma_{g_i}^2) h_i + (\alpha_k/\sigma_{g_k}^2) h_k}{\alpha_i/\sigma_{g_i}^2 + \alpha_k/\sigma_{g_k}^2}.$$ 

$$x_k^* = g_k - \rho_k \frac{\sigma_{g_k}}{\sigma_{h_k}} \frac{(\alpha_i/\sigma_{g_i}^2) (h_i - h_i)}{\alpha_i/\sigma_{g_i}^2 + \alpha_k/\sigma_{g_k}^2}.$$
which implies \( h_i \leq h_k \). Primal feasibility further implies

\[
g_i \geq g_k - (h_k - h_i) \left( \frac{\rho_i(\sigma_{g_k}/\sigma_{h_i})(\alpha_i/\sigma_{h_i}^2) + \rho_i(\sigma_{g_k}/\sigma_{h_i})(\alpha_k/\sigma_{h_k}^2)}{\alpha_i/\sigma_{h_i}^2 + \alpha_k/\sigma_{h_k}^2} \right).
\]

Substituting into the objective function yields the result in this case.

**Case 3:** \( \lambda_i^0 > 0 \) and \( \lambda_k^0 > 0 \). Then \( x_i^* = x_k^* \) and \( y_i^* = y_k^* \); solving, we find

\[
x_i^* = x_k^* = \frac{\left( \sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k \right) (\sigma_{h_i}^2/\alpha_i + \sigma_{h_k}^2/\alpha_k) - (\rho_i\sigma_{g_i}\sigma_{h_i}/\alpha_i + \rho_k\sigma_{g_k}\sigma_{h_k}/\alpha_k)^2}{\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k - \rho_i\sigma_{g_i}\sigma_{h_i}/\alpha_i + \rho_k\sigma_{g_k}\sigma_{h_k}/\alpha_k} (h_k - h_i)
\]

\[
y_i^* = y_k^* = \frac{\left( \sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k \right) (\sigma_{h_i}^2/\alpha_i + \sigma_{h_k}^2/\alpha_k) - (\rho_i\sigma_{g_i}\sigma_{h_i}/\alpha_i + \rho_k\sigma_{g_k}\sigma_{h_k}/\alpha_k)^2}{\sigma_i^2/\alpha_i + \sigma_k^2/\alpha_k - \rho_i\sigma_{g_i}\sigma_{h_i}/\alpha_i + \rho_k\sigma_{g_k}\sigma_{h_k}/\alpha_k} (g_i - g_k)
\]

Substituting into the objective function yields the result.

It can also be shown that the rate functions depend on the locations of the systems, as was shown in the proof of Proposition 4.2. We do not provide the details of the proof in this case.

**P. TEST PROBLEMS**

In Section 8.1, we consider test problems having the following objective function values for the systems, \((g_k, h_k)\) for all \(k \leq r\).

**P.1. Test Problems 1 and 2**

\[
\begin{align*}
102.637584243830, & \quad 97.9407933298666 \\
95.7051438379384, & \quad 102.76860985494 \\
102.25775394492, & \quad 97.6354102475420 \\
103.220921219786, & \quad 100.622586819871 \\
101.012085253317, & \quad 104.463510539031 \\
101.914384981138, & \quad 97.3885735879622 \\
95.553440813858, & \quad 102.462093034316 \\
99.8642556184995, & \quad 103.621906627193 \\
99.8691298026735, & \quad 105.195861379648 \\
95.2816376908915, & \quad 103.323661380141 \\
96.9166532810687, & \quad 98.1973392980161 \\
101.693003090605, & \quad 105.625464250530 \\
104.47761875384, & \quad 96.5696207713976 \\
99.6902711004535, & \quad 94.501519331645 \\
98.1913326499132, & \quad 98.5706164804242 \\
96.3599298766039, & \quad 103.296837233218 \\
100.86511636904, & \quad 98.381080983246 \\
103.194181463619, & \quad 96.9960640919551 \\
99.5010611990529, & \quad 95.1233892952049 \\
104.99400945041, & \quad 99.3581864607797 \\
102.122503735952, & \quad 103.200782983455 \\
90.768690184718, & \quad 98.2846337186736 \\
101.11000198240, & \quad 105.625464250530 \\
104.47761875384, & \quad 96.5696207713976 \\
99.6902711004535, & \quad 94.501519331645 \\
98.1913326499132, & \quad 98.5706164804242 \\
96.3599298766039, & \quad 103.296837233218 \\
100.86511636904, & \quad 98.381080983246 \\
103.194181463619, & \quad 96.9960640919551 \\
99.5010611990529, & \quad 95.1233892952049 \\
104.99400945041, & \quad 99.3581864607797 \\
102.122503735952, & \quad 103.200782983455 \\
102.637584243830, & \quad 97.9407933298666 \\
95.7051438379384, & \quad 102.76860985494 \\
102.25775394492, & \quad 97.6354102475420 \\
103.220921219786, & \quad 100.622586819871 \\
101.012085253317, & \quad 104.463510539031 \\
101.914384981138, & \quad 97.3885735879622 \\
95.553440813858, & \quad 102.462093034316 \\
99.8642556184995, & \quad 103.621906627193 \\
99.8691298026735, & \quad 105.195861379648 \\
95.2816376908915, & \quad 103.323661380141 \\
96.9166532810687, & \quad 98.1973392980161 \\
101.693003090605, & \quad 105.625464250530 \\
104.47761875384, & \quad 96.5696207713976 \\
99.6902711004535, & \quad 94.501519331645 \\
98.1913326499132, & \quad 98.5706164804242 \\
96.3599298766039, & \quad 103.296837233218 \\
100.86511636904, & \quad 98.381080983246 \\
103.194181463619, & \quad 96.9960640919551 \\
99.5010611990529, & \quad 95.1233892952049 \\
104.99400945041, & \quad 99.3581864607797 \\
102.122503735952, & \quad 103.200782983455 \
\end{align*}
\]
P.2. Test Problems 3 and 4

(100.939212597913, 94.5154358824747)
(97.8254313414512, 98.172841094010)
(94.0823223630584, 99.1163218930572)
(97.961490293601, 102.21009955028)
(100.53697647137, 99.188120966508)
(100.101667237351, 99.9755314308252)
(101.45064870579, 100.88867424273)
(95.296453683518, 102.4307410149)
(103.415849689433, 98.2910258202288)
(101.823184416312, 97.781432709317)
(98.451287720632, 99.7769241175329)
(98.1493022364179, 94.788614536425)
(94.2787365653163, 97.95158717002)
(98.9207052676011, 105.69109045121)
(103.909159580843, 96.218144329311)
(99.9036453818685, 97.935893591807)
(94.217204351440, 101.23504551504)
(100.118045439091, 100.9988680588)
(104.149765043428, 96.4957502971374)
(99.7297122487128, 98.8975394591573)
(102.398035205915, 102.887231143845)
(97.7117146285451, 97.0571362360859)
(105.103779054345, 100.45312153805)
(96.053064989149, 99.312966104852)
(96.54508294498, 97.856845184996)
(96.6409796759792, 101.60325497851)
(100.758972805841, 99.79402209991)
(101.15936802606, 98.9937614851489)
(102.31665148918, 94.6044487640181)
(104.68647444360, 103.201423746106)
(96.463878789349, 97.905973075366)
(95.111294341657, 102.17906322232)
(100.152322906727, 102.94096526398)
(105.158120772548, 103.02309697426)

(102.716369875272, 95.4334015982139)
(95.7699672092644, 98.745784148117)
(101.269340379399, 96.5736829921744)
(100.607245966295, 98.098048327701)
(98.1615300150245, 97.3504038647590)
(99.9506543550752, 94.799402227434)
(99.8464839799573, 103.996865144026)
(102.315317483047, 94.739388543954)
(101.187330678711, 100.20821145246)
(98.2325079171963, 99.7302250638134)
(98.584805098539, 94.9992899441466)
(100.206780184380, 95.804983636570)
(97.831843227179, 97.437261715673)
(101.350384548042, 94.1658147716179)
(97.325307085260, 98.475900235179)
(104.65396851019, 100.694296558004)
(98.5302788002185, 102.32373221581)
(95.6396319113402, 103.19893397421)
(97.361188427103, 95.451926305504)
(102.19885433754, 104.33159681897)
(94.6547007601706, 102.952043290603)
(95.5199558196539, 97.885045795857)
(100.422524872004, 97.3680759343192)
(104.157604685498, 101.635292348204)
(100.70407256824, 98.948672266674)
(96.913506916631, 101.775062258845)
(97.7667107866592, 102.029030253916)
(98.5932628596196, 102.43514950501)
(102.969618603541, 97.064140095156)
(101.341931586957, 97.418098185284)
(103.712907145583, 98.129096122361)
(101.270501966862, 99.7475110030962)
(96.886005486454, 95.914578553128)
(100.875542440197, 99.458836987221)

\( g \approx 105 \)