

Correlation Robust Stochastic Optimization

Shipra Agrawal* Yichuan Ding† Amin Saberi‡ Yinyu Ye§

July 6, 2009

Abstract

We consider a robust model proposed by Scarf, 1958, for stochastic optimization when only the marginal probabilities of (binary) random variables are given, and the correlation between the random variables is unknown. In the robust model, the objective is to minimize expected cost against worst possible joint distribution with those marginals. We introduce the concept of *correlation gap* to compare this model to the stochastic optimization model that ignores correlations and minimizes expected cost under independent Bernoulli distribution. We identify a class of functions, using concepts of summable cost sharing schemes from game theory, for which the correlation gap is well-bounded and the robust model can be approximated closely by the independent distribution model. As a result, we derive efficient approximation factors for many popular cost functions, like submodular functions, facility location, and Steiner tree. As a byproduct of our analysis, we also improve some existing results in the areas of social welfare maximization and existence of Walrasian equilibria, which may be of independent interest.

1 Introduction

Stochastic optimization models decision making under uncertain or unknown problem data. We consider stochastic optimization problems in which the uncertain variable is the “demand” set. For example, in stochastic network design problems, the random variable is the subset of source-destination pairs to be connected; in stochastic facility location problem, the random variable is the subset of potential clients that will have a demand; and in stochastic set cover problem, it is the subset of elements that need to be covered. In general, such a stochastic program can be expressed as

$$\min_{x \in C} \mathbb{E}[f(x, S)], \quad (1)$$

where x is the decision variable which lies in a constrained set C , and the random subset $S \subseteq V$ cannot be observed before the decisions x is made. $f(x, S)$ is the *cost function* which depends on both the decision x and the outcome scenario S . The objective of stochastic programming is to minimize the expected cost, which depends on the joint distribution of items in V .

In stochastic optimization, it is typically assumed that the distribution of random variable is either known or can be sampled from [1, 3, 14]. In this model, sample average approximation (SAA) has been used give approximation algorithms for many two-stage stochastic discrete optimization problems, including stochastic set cover [14], uncapacitated facility location [14], and

*Email: shipra@stanford.edu. Computer Science and Engineering, Stanford University, Stanford, CA 94305, USA. Research supported in part by Boeing.

†Email: y7ding@stanford.edu. Department of Management Science and Engineering, Stanford University, Stanford, CA 94305, USA.

‡Email: saberi@stanford.edu. Department of Management Science and Engineering, Stanford University, Stanford, CA 94305, USA.

§Email: yinyu-ye@stanford.edu. Management Science and Engineering and, by courtesy, Electrical Engineering, Stanford, CA 94305, USA. Research supported in part by Boeing.

Steiner tree problem [7]. Those models are suitable when one does not have access to a lot of time invariant reliable statistical information. In this paper, we study the problem when information about a part of the distribution (marginals) is known.

In the case when only estimates of marginal probabilities p_i of each element are available, a common heuristic is to assume that the distribution of random set S is a product distribution. In other words, each element i may appear in S independently with a given probability p_i . For example, see [8, 9]. However, there is a conventional wisdom that ignoring correlations can have catastrophic consequences. Examples can be constructed such that the cost of the solution optimized against the independent distribution performs very poorly once certain correlations are introduced.

To address such problems, Scarf (1958, [13]) proposed a correlation-robust or distributionally-robust stochastic model, which minimizes the expected cost over distributions having a fixed marginal probability p_i for each $i \in V$, but with any possible correlations. For a problem instance $(f, V, \{p_i\})$, we wish to find

$$\min_{x \in C} g(x), \quad (2)$$

where $g(x)$ is the expected cost under worst-case distribution when decision x has been made, given by

$$\begin{aligned} \max_{\mathcal{D}} \quad & \mathbb{E}_{\mathcal{D}}[f(x, S)] \\ \text{s.t.} \quad & \sum_{S: i \in S} \mathbb{P}_{\mathcal{D}}(S) = p_i. \quad \forall i \in V. \end{aligned} \quad (3)$$

We believe this is a very useful model because it takes advantage of the stochasticity of the input, and at the same time efficiently utilizes the available information. On the other hand, it defines an exponential size linear program which makes the problem potentially difficult to solve.

A natural question is how much risk it involves to simply ignore the correlations and minimize the expected cost of independent distribution instead of the worst case distribution. Or, in other words, how well the stochastic optimization model with independent distribution approximates the correlation robust model. The focus of this paper is to study this *correlation gap*. For a particular problem instance $(f, V, \{p_i\})$ and a decision x , we define the correlation gap as the ratio between the expected cost $\mathbb{E}[f(x, S)]$ under the worst case distribution and that under the independent distribution on S . Our main result is to characterize a wide class of functions for which the correlation gap can be well bounded. We also provide counter-examples showing large correlation gap for various other classes of functions.

Correlation gap has many interesting implications for stochastic optimization problems. A small upper bound on correlation gap allows relaxation of the stochastic optimization problem under any distribution, including the worst case distribution model (2), to the product distribution case which is often more efficient to solve either by sampling or by other algorithmic techniques [8, 9]. Further, in many real data collection scenarios, practical constraints can make it very difficult (or costly) to learn the complete information about correlations in data. In those cases, the correlation gap can provide a guideline to decide how important it is to spend resources on learning these correlations. In other words, it measures the “value of correlations” in the statistical data.

Below, we summarize our key results:

- *A class of functions with bounded correlation gap:* For functions $f(x, S)$ that are monotone in S and have a cross-monotone, β -budget balance, (weak) η -summable cost-sharing scheme, we show that the correlation gap is upper bounded by $\eta\beta\frac{e}{e-1}$. This will give correlation gap bounds (and matching approximation factors for robust model) of $e/(e-1)$ for submodular functions, $O(\log(n))$ for facility location, and $O(\log^2 n)$ for Steiner forest, where $n = |V|$, the size of ground set.

- *Hardness results:* We show examples with correlation gap of $O(2^n)$ for functions supermodular in S , $O(\sqrt{n}/\sqrt{\log(n)})$ for monotone subadditive functions in S , and $e/(e-1)$ for submodular functions. These examples will also prove corresponding lower bounds on approximation factors that can be achieved by substituting independent distribution for the robust model.
- *Polynomial-time algorithm for supermodular functions:* We analytically characterize the worst case distribution when function $f(x, S)$ is supermodular in the random set, and consequently give a polynomial-time algorithm for the correlation robust model provided f is convex in x .
- *New results for welfare maximization problems:* As a byproduct, our result provides a $\frac{1}{\eta\beta}(1-1/e)$ -approximation algorithm for the well-studied problem of social welfare maximization in combinatorial auctions, when the utility functions are identical and admit $\eta\beta$ -cost-sharing scheme. Notably, this implies $(1-1/e)$ -approximation for submodular utility functions, matching the current best approximation factor (Vondrak, 2008 [15]) for this case.

Our analysis also yields a simple counterexample for the conjecture by Bikhchandani [2] that markets that have buyers with identical submodular utilities admit a Walrasian price equilibria.

The rest of the paper is organized as follows. To begin, Section 2 will provide a mathematical definition of correlation gap, and examples showing large correlation gap for certain classes of cost functions. In Section 3, we present our main technical theorem that upper bounds the correlation gap for a wide class of cost functions, and discuss its implications on various stochastic optimization problems and the welfare maximization problem. The proof of this theorem is presented in Section 4. Finally, in Section 5, we end with a direct solution of correlation robust model for supermodular functions.

2 Correlation Gap

For a problem instance $(f, V, \{p_i\})$ and at a given decision x , we define correlation gap as the ratio κ between the expected cost of the worst case distribution and that of the independent distribution, i.e.,

$$\kappa := \frac{\mathbb{E}_{\mathcal{D}^R}[f(x, S)]}{\mathbb{E}_{\mathcal{D}^I}[f(x, S)]}, \quad (4)$$

where \mathcal{D}^I is the independent Bernoulli distribution (also called product distribution) with marginals $\{p_i\}$, and \mathcal{D}^R is the worst-case distribution (as given by (3)).

Suppose for some particular cost function f , the correlation gap can be bounded above by $\bar{\kappa}$ for all x , then it is not difficult to show that the decision obtained assuming independent distribution will give a $\bar{\kappa}$ -approximate solution to the robust optimization problem. More precisely, let x_I is the optimal solution to the stochastic optimization problem (1) with independent Bernoulli distribution, and x_R is the optimal solution to the correlation robust problem (2). Then,

$$\begin{aligned} g(x_I) &= \mathbb{E}_{\mathcal{D}^R}[f(x_I, S)], \quad \text{and} \\ g(x_R) &= \mathbb{E}_{\mathcal{D}^R}[f(x_R, S)] \geq \mathbb{E}_{\mathcal{D}^I}[f(x_R, S)] \geq \mathbb{E}_{\mathcal{D}^I}[f(x_I, S)] \end{aligned}$$

Using the bound on correlation gap at x_I , this implies

$$g(x_I) \leq \bar{\kappa} g(x_R)$$

Unfortunately, for general cost functions, the correlation gap and hence the corresponding approximation factor can be large in order of n , as demonstrated by the following examples.

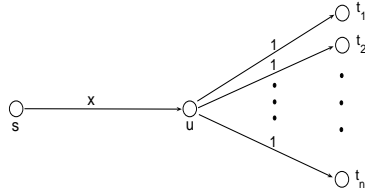


Figure 1: An example with exponential correlation gap

Example 1. (*Minimum cost flow: $O(2^n)$ correlations gap for supermodular functions*)

(Sketch) Consider a two-stage minimum cost flow problem as in Figure 1. There is a single source s , and n sinks t_1, t_2, \dots, t_n . Each sink t_i has a probability $p_i = \frac{1}{2}$ to request a demand, and then a unit flow has to be sent from s to t_i . Each arc (u, t_i) has a fixed capacity 1, but the capacity of arc (s, u) needs to be purchased at a cost $c^I(x)$ in the first stage, and a higher cost $c^{II}(x)$ in the second stage after the set of demand requests is revealed. $c^I(x)$, $c^{II}(x)$ are given as

$$c^I(x) = \begin{cases} x, & x \leq n-1 \\ n+2, & x = n \end{cases} \quad c^{II}(x) = 2^n x.$$

Given the first stage decision x , the cost of edges that need to be bought in the second stage to serve a set S of requests is given by: $f(x, S) = c^I(x) + c^{II}(|S| - x)^+ = c^I(x) + 2^n(|S| - x)^+$. It is easy to check that $f(x, S)$ is supermodular in S for any given x , i.e. $f(x, S \cup i) - f(x, S) \geq f(x, T \cup i) - f(x, T)$ for any $S \supseteq T$. The objective is to minimize the total expected cost $c^I(x) + \mathbb{E}[f(x, S)]$. If the decision maker assumes independent demands from the sinks, then $x_I = n - 1$ minimizes the expected cost, and the expected cost is n ; however, for the worst case distribution the expected cost of this decision will be $g(x_I) = 2^{n-1} + n - 1$ (when $\Pr(V) = \Pr(\emptyset) = 1/2$ and all other scenario have zero probability), which is exponentially higher than that of independent distribution. Hence, the correlation gap is exponentially high. On the other hand, the risk-averse strategy is to use the robust solution $x_R = n$, which leads to a cost $g(x_R) = n + 1$. Thus, approximation ratio $g(x_I)/g(x_R) = O(2^n)$. \square

Example 2. (*Stochastic set cover: $O(\sqrt{n}/\sqrt{\log(n)})$ correlation gap for subadditive functions*)

(Sketch) Consider a set cover problem with elements $V = \{1, \dots, n\}$. Each item $j \in V$ has a marginal probability of $1/K$ to appear in the random set S . The covering sets are defined as follows. Consider a partition of V into $K = \sqrt{n}$ sets A_1, \dots, A_K each containing K elements. The covering sets are all the sets in the cartesian product $A_1 \times \dots \times A_K$. Each set has unit cost. Then, cost of covering a set S is given by subadditive function

$$c(S) = \max_{i=1, \dots, K} |S \cap A_i| \quad \forall S \subseteq V.$$

The worst case distribution with marginal probabilities $p_i = 1/K$ is one where probabilities $\Pr(S) = 1/K$ for $S = A_i$, $i = 1, 2, \dots, K$, and $\Pr(S) = 0$ otherwise. The expected value of $c(S)$ under this distribution is K . For independent distribution, $c(S) = \max_i \zeta_i$, where $\zeta_i = |S \cap A_i|$ are independent $(K, 1/K)$ -binomially distributed random variables. By central limit theory, for large n , ζ_i s are approximately normally distributed with mean 1 and variance $1 - 1/K$. Also, it is known that for K i.i.d. standard normal random variables $\{Y_k\}$, $\lim_{K \rightarrow \infty} (2 \log K)^{-1/2} \max_{1 \leq k \leq K} Y_k = 1$ almost surely [6]. This implies that $\mathbb{E}[\max\{\zeta_i\}]$ is $O(\sqrt{\log n})$. So the correlation gap is of order $O(\sqrt{n}/\sqrt{\log(n)})$.

To obtain approximation lower bound for two-stage stochastic set cover instance, extend the above instance as follows. Let the first stage cost of a covering set to be $w^I = (1 + \epsilon)\sqrt{\log n}/\sqrt{n}$ for some small $\epsilon > 0$, and the second stage cost to be $w^{II} = 1$. For a given first stage cover

x , let $B(x)$ be the set of elements covered by x , then $f(x, S) = w^I x + c(S - B(x))$. Using above analysis for function $c(S)$, the optimal solution for independent distribution will be to buy no (or very few) sets in the first stage giving $\mathbb{E}[f(x, S)] = O(\sqrt{\log(n)})$ for independent distribution, but $O(\sqrt{n})$ cost for worst case distribution. On the other hand, the optimal solution considering worst case distribution is to cover all the elements in the first stage giving slightly more than $O(\sqrt{\log(n)})$ cost in the worst case. Thus, approximation ratio $g(x_I)/g(x_R) = O(\sqrt{n}/\sqrt{\log(n)})$. \square

3 A class of functions with low correlation gap

A key contribution of our paper is to identify a class of cost functions for which the correlation gap is well bounded. To our interest, many popular cost functions including submodular functions, facility location, Steiner forest, etc. belong to this class, which will lead to efficient approximations for these problems.

We derive our characterization using concepts of cost-sharing. A cost-sharing scheme is a function defining how to share the cost of a service among the serviced customers. We consider the class of cost functions f such that for every feasible x , there exists some cost-sharing scheme for allocating the cost $f(x, S)$ among members of set S with (a) β -budget balance (b) weak cross-monotonicity, and (c) weak η -summability. Below we precisely state these properties. Since we assume that x can take any fixed value, we will abbreviate $f(x, S)$ as $f(S)$ for simplicity when clear from the context.

A cost-sharing scheme is *cross-monotonic* if it satisfies the property that everyone is better off when the set of people who receive the service expands [10]. Roughgarden et al [11] introduced an additional property of *summability* for cost-sharing schemes. Here, we will define a slightly weaker version of these properties by requiring them to hold only for given ordering on a subset of V . More precisely, we define a cost-sharing scheme as a function $\chi(i, S, \sigma_S)$ that, for each element $i \in S$ and ordering σ_S on S , specifies the share of i in S . The three properties of budget-balance, weak cross-monotonicity and weak summability are now stated as follows:

1. *β -budget balance*: For all S , and orderings σ_S on S :

$$f(S) \geq \sum_{i=1}^{|S|} \chi(i, S, \sigma_S) \geq \frac{f(S)}{\beta}$$

2. *Weak cross-monotonicity*: $\chi(i, S, \sigma_S) \geq \chi(i, T, \sigma_T)$, if $S \subseteq T$, and $\sigma_S \preceq \sigma_T$. Here, $\sigma_S \preceq \sigma_T$ means that the ordering σ_S is a prefix of ordering σ_T .

3. *Weak η -summability*: For all S , and orderings σ_S :

$$\sum_{\ell=1}^{|S|} \chi(i_\ell, S_\ell, \sigma_{S_\ell}) \leq \eta f(S)$$

where i_ℓ is the ℓ^{th} element and S_ℓ is the set of the first ℓ members of S according to ordering σ_S . And, σ_{S_ℓ} is the restriction of σ_S on S_ℓ . Note that this is a weaker requirement than the conventional definition of summability, where a single cost-sharing function $\chi(i, S)$ must satisfy the given inequality for *all* orderings on the ground set [11].

We may re-emphasize that any cost-sharing scheme satisfying the conventional definition of β -budget-balance, cross-monotonicity and η -summability (as in [10, 11]) will always satisfy the above weaker conditions. However, this relaxation to weak conditions could give significant

savings in approximation factors for some cases. For example, submodular functions satisfy the above weak conditions with $\eta = 1$ and $\beta = 1$ for the incremental cost-sharing scheme:

$$\chi(i, S, \sigma_S) = f(S_i) - f(S_{i-1})$$

where S_i is the set of the first i members of S according to ordering σ_S . On the other hand, for the conventional definition of summability, a lower bound of $\eta \geq O(\log n)$ was shown for submodular functions in [11].

Let us call a cost-sharing scheme satisfying the above three properties an (η, β) -cost-sharing scheme. Our main result is the following theorem, which we will prove in the next section:

Theorem 1. *For any instance $(f, V, \{p_i\})$, if for all feasible x , the cost function $f(x, S)$ is non-decreasing in S and has an (η, β) -cost-sharing method for elements in S , then the correlation gap is bounded as $\eta\beta \left(\frac{e}{e-1}\right)$.*

As described in Section 2, this gives following corollary for approximating the correlation robust model:

Corollary 1.1. *For instances $(f, V, \{p_i\})$ as defined in Theorem 1, an $\eta\beta\frac{e}{e-1}$ approximate solution for correlation robust optimization problem can be constructed by solving the corresponding stochastic optimization problem under independent distribution.*

Further, it is easy to show that for these functions, the variance under independent distribution is bounded by $O\left(\frac{\eta^2\beta^2}{p^2}\right)$. Thus, if the cost function is convex in x , these stochastic optimization problems may be solved efficiently using sample average approximation (SAA) method [1]. For specific problems, the structural simplicity provided by independent distribution may even eliminate the need of using sample average approximation.

Before moving on to the proof of Theorem 1, let us briefly discuss its implications for various stochastic optimization problems, and for a seemingly unrelated problem of welfare maximization in combinatorial auctions:

3.1 Stochastic optimization with submodular functions

A function $h : 2^V \rightarrow \mathbb{R}$ is submodular if $h(S \cup i) - h(S) \leq h(T \cup i) - h(T)$ for all $S \supseteq T$, and $i \in V$. These cost functions are characterized by diminishing marginal costs, which is common for resource allocation problems where a resource can be shared by multiple users and thereby the marginal cost decreases as number of users increases. As discussed earlier, for submodular functions $\eta = 1, \beta = 1$. Therefore, Theorem 1 directly leads to the following corollary:

Corollary 1.2. *If the cost function $f(x, S)$ is monotone and submodular in S for all feasible x , then for any instance $(f, V, \{p_i\})$, the correlation gap is bounded by the constant $\frac{e}{e-1}$.*

The next example shows the $e/(e-1)$ bound is tight for submodular functions.

Example 3. (*Tightness*) Let $V := \{1, 2, \dots, n\}$, define $f(S) = 1$ if $S \neq \emptyset$, and $f(\emptyset) = 0$. Let each item has a probability $p = \frac{1}{n}$. Then the worst case distribution is $Pr(\{i\}) = 1/n$ for each $i \in V$, with expected value 1. The independent distribution has an expected cost $1 - (1 - \frac{1}{n})^n \rightarrow 1 - 1/e$ as $n \rightarrow \infty$. \square

3.2 Stochastic Uncapacitated Facility Location (SUFL)

In two-stage stochastic facility location problem, any facility $j \in F$ can be bought at a low cost w_j^I in the first stage, and higher cost $w_j^{II} > w_j^I$ in the second stage, that is, after the random set $S \subseteq V$ of cities to be served is revealed. The decision maker's problem is to decide $x \in \{0, 1\}^{|F|}$,

the facilities to be build in the first stage so that the total expected cost $\mathbb{E}[f(x, S)]$ of facility location is minimized. (refer to [14] for further details).

Given a first stage decision x , the cost function $f(x, S) = w^I \cdot x + c(x, S)$, where $c(x, S)$ is the cost of deterministic UFL for set S of customers and set F of facilities such that the facilities in x are available at 0 cost. For deterministic UFL cost function there exists a cross-monotonic, 3-budget balanced, $\log n$ summable cost-sharing scheme [12]. Therefore, using Theorem 1, we get following bound on correlation gap:

Corollary 1.3. *The correlation gap for Stochastic uncapacitated facility location is bounded by $O(\log n)$.*

This observation reduces our robust facility location problem to the well-studied stochastic UFL problem under known (independent Bernoulli) distribution [14] at the expense of an $O(\log n)$ approximation factor.

3.3 Stochastic Steiner Tree (SST)

In the two-stage stochastic Steiner tree problem, we are given a graph $G = (V, E)$. An edge $e \in E$ can be bought at cost w_e^I in the first stage. The random set S of terminals to be connected are revealed in the second stage. More edges may be bought at a higher cost w_e^{II} , $e \in E$ in the second stage after observing the actual set of terminals. Here, decision variable x is the edges to be bought in the first stage, and cost function $f(x, S) = w^I \cdot x + c(x, S)$, where $c(x, S)$ is the Steiner tree cost function for set S given that the edges in x are already bought. Since a $\log^2(n)$ -summable, 2-budget balanced cost sharing method is known for this cost function [12, 4], we can conclude:

Corollary 1.4. *The correlation gap for Stochastic Steiner tree is bounded by $O(\log^2 n)$.*

This observation reduces our robust problem to the well-studied (for example see [7]) SST problem under known (independent Bernoulli) distribution at the expense of an $O(\log^2 |V|)$ -approximation factor.

3.4 Welfare Maximization Problem

Finally, Theorem 1 extends some existing results for *social welfare maximization* in combinatorial auctions. Consider the problem of maximizing total utility achieved by partitioning n goods among K players each with utility function $f(S)$ for subset S of goods¹. The optimal welfare OPT is obtained by following integer program:

$$\begin{aligned} \max_{\alpha} \quad & \sum_S \alpha_S f(S) \\ \text{s.t.} \quad & \sum_{S:i \in S} \alpha_S = 1, \quad \forall i \in V \\ & \sum_S \alpha_S = K \\ & \alpha_S \in \{0, 1\}, \quad \forall S \subseteq V \end{aligned} \tag{5}$$

Observe that on relaxing the integrality constraints on α and scaling it by $1/K$, the above problem reduces to that of finding the worst-case distribution α^* (i.e. one that maximizes expected value $\sum_S \alpha_S f(S)$ of function f) such that the marginal probability $\sum_{S:i \in S} \alpha_S$ of each element is $1/K$. Therefore:

$$\text{OPT} \leq \mathbb{E}_{\alpha^*}[Kf(S)]$$

Consequently, the correlation gap bound in Theorem 1 leads to the following corollary for welfare maximization problems:

¹A more general formulation of this problem that is often considered in the literature allows non-identical utility functions for various players

Corollary 1.5. *For welfare maximization problems with n goods and K players with identical utility functions f , the randomized algorithm that assigns goods independently to each of the K players with probability $1/K$ gives $\frac{1}{\eta\beta}(1 - \frac{1}{e})$ approximation to the optimal partition; given that function f is non-decreasing and admits an (η, β) -cost-sharing scheme.*

Since $\eta = 1, \beta = 1$ for submodular functions, the above result matches the $1 - 1/e$ approximation factor provided by Vondrak [15] for this problem in case of identical monotone submodular functions.

The reader may observe that even though approximating the worst case distribution directly provides a matching approximation for the corresponding welfare maximization problem, the converse is not true. In addition to having uniform probabilities $p_i = 1/K$, solutions for welfare maximization approximate the integer program (5), where as the worst case distribution requires solving the corresponding LP relaxation. The latter is a strictly harder problem unless the integrality gap is 0. A notable example is the above-mentioned case of identical submodular functions. This case was studied by Bhikchandani [2] in context of Walrasian equilibria who conjectured a 0 integrality gap for this problem implying the existence of Walrasian equilibria. However, in appendix B we show a simple counter-example with non-zero integrality gap (11/12) for this problem. As a byproduct, this counter-example proves that even for identical submodular valuation functions, Walrasian equilibria may not exist.

4 Proof of Theorem 1

For a problem instance $(f, V, \{p_i\})$, use $\mathcal{L}(f, V, \{p_i\})$ and $\mathcal{I}(f, V, \{p_i\})$ to denote the expected cost of worst-case distribution and independent Bernoulli distribution respectively. In this section, we prove our main technical result that the correlation gap

$$\frac{\mathcal{L}(f, V, \{p_i\})}{\mathcal{I}(f, V, \{p_i\})} \leq \eta\beta \frac{e}{(e-1)}$$

when f is monotone and admits (η, β) cost-sharing in S . As before, we will abbreviate $f(x, S)$ as $f(S)$ for simplicity.

The proof is structured as follows. We first focus on special instances of the problem in which all p_i 's are equal to $1/K$ for some integer K , and the worst case distribution is a “ K -partition-type” distribution. That is, the worst case distribution divides the elements of V into K disjoint sets $\{A_1, \dots, A_K\}$, and each occurs with probability $1/K$. Observe that for such instances, the expected value under worst case distribution is $\mathcal{L}(f, V, \{p_i\}) = \frac{1}{K} \sum_i f(A_i)$. In Lemma 1, we show that for such “nice” instances the correlation gap is bounded by $\eta\beta \frac{e}{e-1}$. Then, in Lemma 2 we use a “split” operation to reduce any given instance of our problem to a nice instance such that the reduction can only increase the correlation gap. This will show that the bound $\eta\beta \frac{e}{e-1}$ for nice instances is an upper bound for any instance of the problem, thus concluding the proof of the theorem.

Lemma 1. *For instances $(f, V, \{p_i\})$ such that (a) $f(S)$ is non-decreasing and admits an (η, β) -cost-sharing method (b) marginal probabilities p_i are all equal to $1/K$ for some integer K , and (c) the worst case distribution is a K -partition-type distribution, the correlation gap can be bounded as:*

$$\frac{\mathcal{L}(f, V, \{1/K\})}{\mathcal{I}(f, V, \{1/K\})} \leq \eta\beta \frac{e}{(e-1)}$$

Proof. Let the optimal K -partition corresponding to the worst case distribution is $\{A_1, A_2, \dots, A_K\}$. Assume w.l.o.g that $f(A_1) \geq f(A_2) \geq \dots \geq f(A_K)$. Fix an order σ on elements of V such that

for all k , the elements in A_k come before A_{k-1} . Let χ is the (η, β) cost-sharing. Then by weak η -summability of χ , for this ordering:

$$\mathcal{I}(f, V, \{1/K\}) = \mathbb{E}_{S \subseteq V} [f(S)] \geq \frac{1}{\eta} \mathbb{E}_{S \subseteq V} \left[\sum_{l=1}^{|S|} \chi(i_l, S_l, \sigma_{S_l}) \right]. \quad (6)$$

where the expected value is taken over independent distribution.

Denote $\phi(V) := \mathbb{E}_{S \subseteq V} \left[\sum_{l=1}^{|S|} \chi(i_l, S_l, \sigma_{S_l}) \right]$. We will show that

$$\phi(V) \geq (1-p)\phi(V \setminus A_1) + \frac{1}{\beta} f(A_1)$$

Recursively using this inequality will prove the result.

For any $S \subseteq V$, denote $S_{-1} = S \cap (V \setminus A_1)$, $S_1 = S \cap A_1$. Then,

$$\phi(V) = \mathbb{E}_S \left[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l, \sigma_{S_l}) \right] + \mathbb{E}_S \left[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S_l, \sigma_{S_l}) \right] \quad (7)$$

Since elements in A_1 come after the elements in $V \setminus A_1$ in ordering σ , we have that for $l > |S_{-1}|$, $S_l \subseteq S_{-1} \cup A_1$. Also, since $\sigma_{S_l} \prec \sigma_{S_{-1} \cup A_1}$, using cross-monotonicity of χ , the second term above can be bounded as:

$$\mathbb{E}_S \left[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S_l, \sigma_{S_l}) \right] \geq \mathbb{E}_S \left[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S_{-1} \cup A_1, \sigma_{S_{-1} \cup A_1}) \right]. \quad (8)$$

Because S_{-1} and S_1 are mutually independent, for any fixed S_{-1} , each $i \in A_1$ will have the same conditional probability $p = 1/K$ of appearing in S_1 . Therefore,

$$\begin{aligned} \mathbb{E}_S \left[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S_{-1} \cup A_1, \sigma_{S_{-1} \cup A_1}) \right] &= \mathbb{E}_{S_{-1}} \left[\mathbb{E} \left[\sum_{l=|S_{-1}|+1}^{|S|} \chi(i_l, S_{-1} \cup A_1, \sigma_{S_{-1} \cup A_1}) \mid S_{-1} \right] \right] \\ &= p \mathbb{E}_{S_{-1}} \left[\sum_{i \in A_1} \chi(i, S_{-1} \cup A_1, \sigma_{S_{-1} \cup A_1}) \right] \end{aligned} \quad (9)$$

Similarly, for $l \leq |S_{-1}|$, $S_l \subseteq S_{-1}$. Again, using independence and cross-monotonicity, analyze the first term in the right hand side of (7),

$$\begin{aligned} \mathbb{E}_S \left[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l, \sigma_{S_l}) \right] &= \mathbb{E}_{S_{-1}} \left[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l, \sigma_{S_l}) \right] \\ &\geq (1-p) \mathbb{E}_{S_{-1}} \left[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l, \sigma_{S_l}) \right] + p \mathbb{E}_{S_{-1}} \left[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l \cup A_1, \sigma_{S_l \cup A_1}) \right] \\ &= (1-p) \phi(V \setminus A_1) + p \mathbb{E}_{S_{-1}} \left[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l \cup A_1, \sigma_{S_l \cup A_1}) \right]. \end{aligned} \quad (10)$$

Based on (7), (9) and (10), and the fact that the cost-sharing method is β -budget balanced, we deduce

$$\begin{aligned} \phi(V) &= (1-p) \phi(V \setminus A_1) + p \mathbb{E}_{S_{-1}} \left[\sum_{l=1}^{|S_{-1}|} \chi(i_l, S_l \cup A_1, \sigma_{S_l \cup A_1}) + \sum_{i \in A_1} \chi(i, S_{-1} \cup A_1, \sigma_{S_{-1} \cup A_1}) \right] \\ &\geq (1-p) \phi(V \setminus A_1) + \frac{1}{\beta} p \mathbb{E}_{S_{-1}} [f(S_{-1} \cup A_1)] \\ &\geq (1-p) \phi(V \setminus A_1) + \frac{1}{\beta} p f(A_1), \end{aligned} \quad (11)$$

The last inequality follows from monotonicity of f . Expanding the above recursive inequality for A_2, \dots, A_K , we get

$$\phi(V) \geq \frac{1}{\beta} p \sum_{k=1}^K (1-p)^{k-1} f(A_k), \quad (12)$$

Since $f(A_k)$ is decreasing in k , and $p = 1/K$ by simple arithmetic one can show

$$\phi(V) \geq \frac{1}{\beta} \cdot \sum_{k=1}^K p f(A_k) \cdot \frac{(\sum_{k=1}^K (1-p)^{k-1})}{K} \geq \frac{1}{\beta} \cdot \left(1 - \frac{1}{e}\right) \cdot \sum_{k=1}^K p f(A_k)$$

By definition of $\phi(V)$, this gives:

$$\mathcal{I}(f, V, \{1/K\}) \geq \frac{1}{\eta\beta} \left(1 - \frac{1}{e}\right) \mathcal{L}(f, V, \{1/K\}).$$

□

Next, we reduce a general problem instance to an instance satisfying the three properties required in Lemma 1. We use the following split operation.

Split: Given a problem instance $(f, V, \{p_i\})$, and integers $\{n_i \geq 1, i \in V\}$, define a new instance $(f', V', \{p'_j\})$ as follows: split each item $i \in V$ into n_i copies i_1, i_2, \dots, i_{n_i} , and assign a marginal probability of $p'_{i_k} = \frac{p_i}{n_i}$ to each copy. Let V' denote the new ground set containing all the duplicates. Define the new cost function $f' : 2^{V'} \rightarrow \mathbb{R}$ as:

$$f'(S') = f(\Pi(S')), \text{ for all } S' \subseteq V', \quad (13)$$

where $\Pi(S') \subseteq V$ is the original subset of elements whose duplicates appear in S' , i.e. $\Pi(S') = \{i \in V \mid i_k \in S' \text{ for some } k \in \{1, 2, \dots, n_i\}\}$.

The split operation has following properties. Their proofs, which are somewhat involved, will be given in Appendix A.

Property 1. *If $f(S)$ is a non-decreasing function in S with an (η, β) -cost sharing scheme, then so is f' .*

Property 2. *If $f(S)$ is non-decreasing in S , then splitting does not change the worst case expected value, that is:*

$$\mathcal{L}(f, V, \{p_i\}) = \mathcal{L}(f', V', \{p'_j\})$$

Property 3. *If $f(S)$ is non-decreasing in S , then splitting can only decrease the expected value over independent distribution:*

$$\mathcal{I}(f, V, \{p_i\}) \geq \mathcal{I}(f', V', \{p'_j\}).$$

Using these properties of split operation, it is straightforward to reduce any instance to a “nice” instance.

Lemma 2. *Any problem instance $(f, V, \{p_i\})$ with function f that is monotone in S and admits (η, β) cost-sharing scheme, can be reduced to an instance $(f', V, \{1/K\})$ using splitting such that the new instance satisfies the conditions required in Lemma 1 and the correlation gap*

$$\frac{\mathcal{L}(f', V, \{1/K\})}{\mathcal{I}(f', V, \{1/K\})} \geq \frac{\mathcal{L}(f, V, \{p_i\})}{\mathcal{I}(f, V, \{p_i\})}.$$

Proof. Suppose that the worst case distribution for instance $(f, V, \{p_i\})$ is not a partition-type distribution. Then, we split every element i that appears in two different sets until the distribution becomes a partition. Since each set formed after splitting contains at least one copy of i , by definition of function f' , this splitting does not change the expected function value. By Property 2 of Split operation, the worst case expected values for the two instances (before and after splitting) must be the same, so this partition forms the worst case distribution for the new instance. Then, we can further split until the marginal probability of each new element is $1/K$ for some large enough integer K ². By the properties of Split operation, the resulting cost function f' obtained by splitting will be monotone with $\eta\beta$ -cost sharing scheme, and the correlation gap can only become larger on splitting. Thus, all conditions are satisfied by the reduced instance. □

²Such an integer K can always be reached assuming p_i s are rational.

5 Supermodular functions

In the end, we directly consider the correlation robust model for cost functions $f(x, S)$ which are supermodular in S . As shown in Section 2, the correlation gap for these cost functions can be exponentially high, so independent distribution does not give a good approximation to the worst case distribution. However, it is easy to characterize the worst case distribution and directly solve the correlation robust model in this case.

Lemma 3. *Given that function $f : 2^V \rightarrow \mathbb{R}$ is supermodular, the worst case distribution over S has the following closed form*

$$\Pr(S) = \begin{cases} p_n & \text{if } S = S_n \\ (p_i - p_{i+1}) & \text{if } S = S_i, 1 \leq i \leq n-1 \\ 1 - p_1 & \text{if } S = \emptyset \\ 0 & \text{o.w.} \end{cases}$$

where $n = |V|$; i is the i^{th} member of V and S_i is the set of first i members of V , both with respect to a specific ordering over V such that $p_1 \geq \dots \geq p_n$.

The lemma is simple to prove, a proof appears in appendix C. Lemma 3 implies following corollary for solving the robust optimization problem.

Corollary 3.1. *For cost functions $f(x, S)$ that are supermodular in S for any feasible x , the robust optimization problem is simply formulated as:*

$$\min_{x \in C} p_n f(x, S^n) + \sum_{i=1}^{n-1} (p_i - p_{i+1}) f(x, S^i) + (1 - p_1) f(x, \emptyset)$$

Thus, if $f(x, S)$ is convex in x and C is a convex set, then it is a convex optimization problem and can be solved efficiently.

Acknowledgements The authors would like to thank Ashish Goel and Mukund Sundarajan for many useful insights on the problem.

References

- [1] S. Ahmed, A. Shapiro, and E. Shapiro. The sample average approximation method for stochastic programs with integer recourse. *SIAM Journal of Optimization*, 12:479–502, 2002.
- [2] S. Bikhchandani and J. W. Mamer. Competitive equilibrium in an exchange economy with indivisibilities. *Journal of Economic Theory*, 74(2):385–413, June 1997.
- [3] M. Charikar, C. Chekuri, and M. Pál. Sampling bounds for stochastic optimization. In *APPROX-RANDOM*, pages 257–269, 2005.
- [4] S. Chawla, T. Roughgarden, and M. Sundararajan. Optimal cost-sharing mechanisms for steiner forest problems. In *In Proceedings of the 2nd Workshop on Internet and Network Economics (WINE)*, 2006.
- [5] J. Edmonds. Submodular functions, matroids, and certain polyhedra. pages 11–26, 2003.
- [6] J. Galambos. *The Asymptotic Theory of Extreme Order Statistics*. Wiley, New York, 1978, 1987.

- [7] A. Gupta, M. Pal, R. Ravi, and A. Sinha. Boosted sampling: Approximation algorithms for stochastic optimization. In *In Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 417–426, 2004.
- [8] J. Kleinberg, Y. Rabani, and E. Tardos. Allocating bandwidth for bursty connections. *SIAM J. Comput*, 30:2000, 1997.
- [9] R. H. Möhring, A. S. Schulz, and M. Uetz. Approximation in stochastic scheduling: the power of LP-based priority policies. *J. ACM*, 46(6):924–942, 1999.
- [10] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.
- [11] T. Roughgarden and M. Sundararajan. New trade-offs in cost-sharing mechanisms. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 79–88, New York, NY, USA, 2006. ACM.
- [12] T. Roughgarden and M. Sundararajan. Optimal efficiency guarantees for network design mechanisms. In *IPCO '07: Proceedings of the 12th international conference on Integer Programming and Combinatorial Optimization*, pages 469–483, Berlin, Heidelberg, 2007. Springer-Verlag.
- [13] H. E. Scarf. A min-max solution of an inventory problem. *Studies in The Mathematical Theory of Inventory and Production*, pages 201–209, 1958.
- [14] C. Swamy and D. B. Shmoys. Sampling-based approximation algorithms for multi-stage stochastic optimization. In *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 357–366, Washington, DC, USA, 2005. IEEE Computer Society.
- [15] J. Vondrak. Optimal approximation for the submodular welfare problem in the value oracle model. In *STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 67–74, New York, NY, USA, 2008. ACM.

A Properties of Split Operation

Property 1 If $f(S)$ is non-decreasing in S with an (η, β) -cost sharing scheme, then so is f' .

Proof. Monotonicity holds since for any $S' \subseteq T' \subseteq V'$, $\Pi(S') \subseteq \Pi(T')$:

$$f'(S') = f(\Pi(S')) \leq f(\Pi(T')) = f'(T')$$

Given the original (η, β) cost-sharing method χ for f , we construct a cost-sharing method χ' for the new instance as follows: cost-share χ' coincides with the original scheme χ for the sets without duplicates, but for a set with duplicates, it assigns the cost-share solely to the copy with smallest index (as per the input ordering). That is, any $S' \subseteq V'$, ordering $\sigma_{S'}$, and item i_k (k -th copy of item i) in S' , allocate cost-shares as follows:

$$\chi'(i_k, S', \sigma_{S'}) = \begin{cases} \chi(i, S, \sigma_S) & \text{if } k = \operatorname{argmin}_{\sigma_{S'}} \{j : i_j \in S'\} \\ 0 & \text{if } k \neq \operatorname{argmin}_{\sigma_{S'}} \{j : i_j \in S'\} \end{cases} \quad (14)$$

where $S = \Pi(S')$, σ_S is the ordering of lowest index copies in $\sigma_{S'}$. It is easy to see that the property of β -budget-balance carries through to the new cost sharing scheme.

For cross-monotonicity, consider any $S' \subseteq T'$, ordering $\sigma_{S'} \prec \sigma_{T'}$, and let $\Pi(S') = S$, $\Pi(T') = T$. For any $i' \in T'$ which is not a lowest indexed copy $\chi(i', T', \sigma_{T'}) = 0$, so that the

condition is automatically satisfied. Let i' is one of the lowest indexed copies in T' , then it must have been a lowest indexed copy in S' . This is because $\sigma'_{S'} \preceq \sigma'_{T'}$, i.e. new elements are added only in the end of the ordering $\sigma'_{S'}$ to get $\sigma'_{T'}$. By similar reasoning, if $\sigma'_{S'} \preceq \sigma'_{T'}$ then the ordering σ_S, σ_T of the lowest indexed copies in S', T' satisfies $\sigma_S \preceq \sigma_T$ as well. Thus,

$$\chi(i', T', \sigma'_{T'}) = \chi(i, T, \sigma_T) \leq \chi(i, S, \sigma_S) = \chi(i', S', \sigma'_{S'})$$

Weak η -summability holds since

$$\sum_{\ell=1}^{|S'|} \chi'(i'_\ell, S'_\ell, \sigma_{S'}) = \sum_{j=1}^{|S|} \chi(i_j, S_j, \sigma_S) \leq \eta f(S) = \eta f'(S')$$

□

Property 2 If the cost function $f(\cdot)$ is non-decreasing in S , then the splitting procedure does not change the worst-case expected value. That is:

$$\mathcal{L}(f, V, \{p_i\}) = \mathcal{L}(f', V', \{p'_j\})$$

Proof. For any fixed x , the worst case expected cost is the optimal value of following linear program, where $\{\alpha_S\}_{S \subseteq V}$ represents a distribution over subsets of set V :

$$\begin{aligned} \mathcal{L}(f, V, \{p_i\}) = \max_{\alpha} \quad & \sum_S \alpha_S f(x, S) \\ \text{s.t.} \quad & \sum_{S: i \in S} \alpha_S = p_i, \quad \forall i \in V \\ & \sum_S \alpha_S = 1 \\ & \alpha_S \geq 0, \quad \forall S \subseteq V. \end{aligned} \tag{15}$$

Suppose item 1 is split into n_1 pieces, and each piece is assigned a probability $\frac{p_1}{n_1}$. Let $\{\alpha_S\}$ denote the optimal solution for the instance $(f, V, \{p_i\})$, then we can construct a solution for the new instance $(f', V', \{p'_j\})$ which has the same objective value by assigning non-zero probabilities to only those sets which have no duplicates.

$$\forall S' \subseteq V', \quad \alpha'_{S'} = \begin{cases} \alpha_{S'} & \text{if } S' \text{ contains no copies of item 1} \\ \frac{p_1}{n_1} \alpha_{S'} & \text{if } S' \text{ contains exactly one copy of item 1} \\ 0 & \text{if } S' \text{ contains more than one copy of item 1} \end{cases}$$

One can verify that $\{\alpha'_{S'}\}$ is a feasible distribution (i.e., feasible to the linear program (15)) for the new instance $(f', V', \{p'_j\})$, and has the same objective value as $\mathcal{L}(f, V, \{p_i\})$. Hence, $\mathcal{L}(f, V, \{p_i\}) \leq \mathcal{L}(f', V', \{p'_j\})$.

For the other direction, consider an optimal solution $\{\alpha'_{S'}\}$ of the new instance. It is easy to see that there exists an optimal solution $\{\alpha'_{S'}\}$ that $\alpha'_{S'} = 0$ for all S' that contain more than one copy of item 1. To see this, assume for contradiction that some set with non-zero probability has two copies of item 1. By definition of f' , removing one copy will not decrease the function value. Then, because of monotonicity of f' , we can move out one copy to another set T that has no copy of item 1. Such T always exists since the probabilities of copies of item 1 must sum up to $p_1 \leq 1$. So, we can assume that in the optimal solution $\alpha'_{S'} = 0$ for any set S' containing more than one copy. Thus, we can set $\alpha_S = \alpha'_{S'}$ where S is the corresponding original set for any $S' \subseteq V'$. That forms a feasible solution for original instance with same objective value as $\mathcal{L}(f', V', \{p'_j\})$. We can apply the argument recursively for all the items to prove the lemma. □

Next, we prove that the expected cost under independent Bernoulli distribution can only decrease by the split operation.

Property 3 If $f(\cdot)$ is non-decreasing, then after splitting

$$\mathcal{I}(f', V', \{p'_j\}) \leq \mathcal{I}(f, V, \{p_i\}).$$

Proof. Let $(f', V', \{p'_j\})$ denote the new instance by splitting item 1 into n_1 pieces. Denote

$$\Lambda := \{S' \subseteq V' \mid S' \text{ contains at least one copy of } 1\},$$

and denote $\pi = \Pr(S' \in \Lambda)$. Consider the expected cost under independent Bernoulli distribution, by independence,

$$\begin{aligned} \mathcal{I}(f', V', \{p'_j\}) &= \mathbb{E}_{S'} [f'(S') I(S' \in \Lambda)] + \mathbb{E}_{S'} [f'(S') I(S' \notin \Lambda)] \\ &= \pi \mathbb{E}_{S \subseteq V \setminus \{1\}} [f(S \cup \{1\})] + (1 - \pi) \mathbb{E}_{S \subseteq V \setminus \{1\}} [f(S)]. \\ &\leq p_1 \mathbb{E}_{S \subseteq V \setminus \{1\}} [f(S \cup \{1\})] + (1 - p_1) \mathbb{E}_{S \subseteq V \setminus \{1\}} [f(S)] \\ &= \mathcal{I}(f, V, \{p_i\}). \end{aligned}$$

The second last inequality holds because $\pi = 1 - (1 - \frac{p_1}{n_1})^{n_1} \leq p_1$, and $f(S) \leq f(S \cup \{1\})$ by monotonicity. \square

B $\frac{11}{12}$ Integrality gap for SWM with identical submodular valuations

Let $V = \{1, 2, 3, 4, 5, 6\}$, $K = 3$, and construct a monotone submodular value function as

$$f(S) = \begin{cases} 0 & \text{if } S = \emptyset \\ 2 & \text{if } |S| = 1 \\ 3 & \text{if } |S \cap \{1, 2, 3\}| = 1 \text{ and } |S \cap \{4, 5, 6\}| = 1 \\ 4 & \text{if } |S \cap \{1, 2, 3\}| \geq 2 \text{ or } |S \cap \{4, 5, 6\}| \geq 2 \end{cases}$$

Then the optimal fractional solution to the LP relaxation of (5) is given by

$$\alpha_{\{1,2\}} = \alpha_{\{2,3\}} = \alpha_{\{1,3\}} = 0.5, \quad \alpha_{\{4,5\}} = \alpha_{\{5,6\}} = \alpha_{\{4,6\}} = 0.5,$$

with an optimal value 12; but the optimal integer solution will have an optimal value 11. So there is an 11/12 integrality gap.

C Proof of Lemma 3

For any fixed x , the worst case expected cost is the optimal value of following linear program, where $\{\alpha_S\}_{S \subseteq V}$ represents a distribution over subsets of set V :

$$\begin{aligned} \mathcal{L}(f, V, \{p_i\}) &= \max_{\alpha} \sum_S \alpha_S f(x, S) \\ \text{s.t.} \quad &\sum_{S: i \in S} \alpha_S = p_i, \quad \forall i \in V \\ &\sum_S \alpha_S = 1 \\ &\alpha_S \geq 0, \quad \forall S \subseteq V. \end{aligned} \tag{16}$$

It is easy to verify that

$$\alpha^* = \begin{cases} p_n & \text{if } S = S_n \\ (p_i - p_{i+1}) & \text{if } S = S_i, 1 \leq i \leq n-1 \\ 1 - p_1 & \text{if } S = \emptyset \\ 0 & \text{o.w.} \end{cases}$$

is a feasible solution to (16). Next we show that it is actually the optimal solution. The dual of linear program (16) is:

$$\begin{aligned} \min_{\gamma, \lambda} \quad & \gamma + p^T \lambda \\ \text{s.t.} \quad & f(S) - \sum_{i \in S} \lambda_i \leq \gamma, \quad \forall S. \end{aligned} \tag{17}$$

Consider the problem in λ for a given value of γ . This problem is to minimize a linear function with positive coefficients (p_i) over the supermodular polyhedron (of supermodular function $f(S) - \gamma$). Minimizing a linear function over a supermodular polyhedron can be solved by a greedy procedure [5], with the optimal value given by $\sum_{i=1}^n p_i (f(S_i) - f(S_{i-1}))$. Then (17) can be rewritten as

$$\begin{aligned} \min_{\gamma} \quad & \gamma + p_n f(S^n) + \sum_{i=1}^{n-1} (p_i - p_{i+1}) f(S^i) - p_1 f(\emptyset) \\ \text{s.t.} \quad & f(\emptyset) \leq \gamma. \end{aligned}$$

The optimal solution for above is $\gamma = f(\emptyset)$, therefore optimal value:

$$p_n f(S^n) + \sum_{i=1}^{n-1} (p_i - p_{i+1}) f(S^i) + (1 - p_1) f(\emptyset) = \sum_S \alpha_S^* f(S)$$

This proves the lemma.