Supermodularity, Curvature, and Convex Relaxations in a Class of Quadratic Binary Optimization Problems

Bismark Singh

School of Mathematical Sciences, University of Southampton, SO17 1BJ, UK

Abstract

We study the combinatorial structure of a quadratic set function $F(S)$ arising from a class of binary optimization models within the family of undesirable facility location problems. Despite strong empirical evidence of nested optimal solutions in previously studied real-world instances, we show that $F(S)$ is, in general, neither submodular nor supermodular. To quantify deviation from modularity, we derive analytical bounds on the total curvature of $F(S)$ and, under mild assumptions, show that it lies between zero and one. Next, we identify structural regimes—based on continuous relaxations of the feasible region and parameter values—where our inherently combinatorial problem reduces to a simple convex quadratic problem admitting a closed-form optimal solution. Further, this relaxed formulation not only exhibits remarkable connections to notions of proportional fairness but also restores the supermodularity of $F(S)$. Our results reconcile empirically observed nestedness with theoretical non-modularity, contributing new structural insight into this class of combinatorial quadratic optimization models.

Keywords: Supermodularity, Total Curvature, Quadratic Optimization, Convex Relaxation

1. Introduction

Consider the following set-valued optimization model that we derive from the mixed-integer quadratic program originally proposed by Schmitt and Singh (2024).

$$z^* = \min_{S} F(S) \tag{1a}$$

$$\text{s.t.} \quad |S| \le B \tag{1b}$$

where

$$F(S) = \min_{x \in \mathcal{X}_S} \sum_{s \in S} C_s \left(1 - \frac{\sum_{i \in I} W_{is} x_{is}}{C_s}\right)^2 + \sum_{s \in J \setminus S} C_s \tag{2a}$$

$$\text{where } \mathcal{X}_S = \Big\{ x :$$

$$\sum_{i \in I} W_{is} x_{is} \le C_s \qquad \forall s \in S \tag{2b}$$

$$\sum_{s \in S} x_{is} = 1 \qquad \forall i \in I \tag{2c}$$

$$x_{is} = 0 \qquad \forall i \in I,\ s \in J \setminus S \tag{2d}$$

$$x_{is} \in \{0, 1\} \qquad \forall i \in I,\ s \in S \Big\}. \tag{2e}$$

Email address: b.singh@southampton.ac.uk (Bismark Singh)

We refer to model (1) as the *master problem*, and to model (2) as the *subproblem*. In model (2), $S \subseteq 2^J$, where 2^J denotes the power set of $J \subset \mathbb{N}$. The parameters $W_{ij}, C_j \in \mathbb{R}^+$, $\forall i \in I, j \in J$. The set S denotes a candidate subset of facilities, from the ground finite set J, that may be selected (i.e., opened) as determined by constraint (2e). The remaining facilities are necessarily unselected (i.e., closed) from equation (2d). Constraint (2c) provides an assignment of each user $i \in I$ to the open facilities, subject to the capacities of the open facilities given by constraint (2b). Here, $\mathcal{X}_S$ denotes the set of feasible x satisfying constraints (2b)–(2e); i.e., the feasible region of the subproblem. Then, the master problem in model (1) seeks to minimize the optimal value of the subproblem, over all possible subsets of the ground set J of bounded cardinality B.

Finding a feasible point $x \in \mathcal{X}_S$ is known to be $\mathcal{NP}$-complete (Martello and Toth, 1990). Although for problems with "sufficient capacity", feasible solutions are easily obtainable, finding optimal solutions $\arg\min_{\mathcal{X}_S} F(S)$ remains $\mathcal{NP}$-hard (Schmidt and Singh, 2025); here, sufficient capacity denotes the restriction $C_j \geq \sum_{i \in I} W_{ij}, \forall j \in J$, and we detail this in Section 2. Such hardness results motivate the set-theoretic formulation given by model (1) over the traditional integer programming formulation of Schmitt and Singh (2024). The computational tractability of this combinatorial optimization problem defined over subsets of facilities depends crucially on whether $F(S)$ belongs to well-studied classes of set functions.

As background, we recall some classical complexity results for set-function optimization. Unconstrained minimization of submodular functions (or equivalently, unconstrained maximization of supermodular functions) is polynomial-time solvable; while, unconstrained maximization of general submodular functions (or equivalently, unconstrained minimization of general supermodular functions) is $\mathcal{NP}$-hard. However, for the cardinality-constrained case, maximizing a monotone non-decreasing submodular function (or equivalently, minimizing a monotone non-increasing supermodular function) admits a greedy algorithm with a $(1 - 1/e)$ approximation guarantee, even though the general problem without monotonicity remains $\mathcal{NP}$-hard. For an overview of supermodular function optimization, see Fujishige (2005); the above-mentioned classical greedy approximation guarantee is due to Nemhauser et al. (1978).

The aim of this work is to investigate theoretical properties of the function $F(S)$, and, thus, characterize the master problem. Since we are minimizing $F(S)$, supermodularity is the property of primary interest to us. Real-world test instances (from Bavaria, Germany and Hampshire, UK) display a striking nestedness of optimal solutions which suggests a fundamental and unexplored combinatorial structure of the problem; we present a numerical illustration in Appendix Appendix A.1. However, we show in Section 3.2 that, in general, $F(S)$ is *not* supermodular. This contrast, between observed nestedness in practice and provable lack of supermodularity in generality, forms the central theoretical tension that motivates our work. By comparison, establishing the lack of submodularity is more straightforward as we demonstrate both through examples and structural arguments in Section 3.1.

The following are the key contributions of this work.

(i) We establish that $F(S)$ is, in general, neither submodular nor supermodular, despite strong empirical evidence of nested optimal solutions in real-world instances.

(ii) We derive analytical guarantees on the total curvature of $F(S)$ thereby enabling a measure of deviation from modularity.

(iii) We identify families of parameter regimes, as well as relaxed feasible regions under which the combinatorial structure of $F(S)$ admits a convex quadratic reformulation. We show that this reformulation has a closed-form solution and recovers supermodularity.

The structure of the rest of this work is as follows. As preparation, Section 2 introduces the conventions and basic notions that form the foundation of our analysis. Section 3 then develops our results on the absence of both submodularity and supermodularity. In Section 4, we present positive results on the the total curvature of $F(S)$ as a finer measure of deviation from these classical structures. Section 5 highlights regimes and parameterizations where more favorable structural properties arise. Finally, we conclude in Section 6.

2. Preliminaries

We begin this section with the conventions and notation that we use throughout this work.

2.1. Conventions and Feasibility

Let $\mathcal{F} \subseteq 2^J$ denote the family of facility sets $S \subseteq J$ for which model (2) is feasible. For any $S \in \mathcal{F}$, let $u^S = \frac{\sum_{i\in I} W_{ij} x_{ij}^S}{C_j}, \forall j \in J$ denote an optimal utilization for model (2), with the convention $u_j^S = 0$ for $j \notin S$, where x^S denotes an optimal solution of the set $\mathcal{X}_S$ defined by constraints (2b)-(2e). Then, the value of the objective function in equation (2a) is $F(S) = \sum_{j\in S} C_j \left(1 - u_j^S\right)^2 + \sum_{s\in J\setminus S} C_s$ for $S \in \mathcal{F}$.

Throughout this work, we use the convention that $F(S) = \infty$ when $\mathcal{X}_S = \emptyset$, $\forall S \subseteq J$ (i.e., for $S \notin \mathcal{F}$). When $S = \emptyset$ we have no assignments to conduct as all the x variables are set to zero from constraint (2d); then, $F(S) = \sum_{j\in J} C_j$. For sake of exposition, unless explicitly stated otherwise, any statement that involves u^S or $F(S)$ tacitly assumes feasible sets $S \in \mathcal{F}$. For instance, later in Definition 4, it is understood that we restrict ourselves to indices $\left\{j \in J : \{j\} \in \mathcal{F} \text{ and } J \setminus \{j\} \in \mathcal{F}\text{: } u_j^{\{j\}} > 0\right\}$. Further, the following upward-closed property of feasibility holds.

Proposition. For any $S \in \mathcal{F}$ we have $S \cup \{j\} \in \mathcal{F}$ for all $j \notin S$.

Proof. Proof: A feasible solution for the set $S \cup \{j\}$ is constructed from the optimal utilization of the extended set $S \cup \{j\}$ by not utilizing the additional facility j. □ □

To address the issue of non-empty feasible sets $\mathcal{X}_S$ for a given $S \subseteq J$, throughout this work we distinguish between two capacity regimes.

(a) When capacities are sufficiently large such that each facility can accommodate by itself the entire weight of users (i.e., $C_j \geq \sum_{i\in I} W_{ij}, \forall j \in J$), non-emptiness of the set $\mathcal{X}_S$ is assured for any S. Then, $\mathcal{F} = 2^J$. Specifically, in this case, all singleton sets (i.e., $S = \{j\}$) are feasible; hence, by Proposition 1, all other sets are also feasible. We call this regime as the *sufficient-capacity* regime. As we mention in Section 1, finding a feasible point within $\mathcal{X}_S$ is trivial in this regime, however determining the optimal $F(S)$ is $\mathcal{NP}$-complete (Schmidt and Singh, 2025).

(b) When capacities are not large enough, some facilities within the selected subsets S may have less capacity than required to accommodate the total weight of user. Thus, feasibility is not guaranteed for arbitrary subsets. However, it follows from Proposition 1 that given $S \in \mathcal{F}$ and $S \subset T \subseteq J$, we have $T \in \mathcal{F}$. A necessary condition to ensure feasibility of the entire set of users (i.e., $S = J$) is $\sum_{i\in I} \min_{j\in J} W_{ij} \leq \sum_{j\in J} C_j$. We assume this throughout this work, else the problem is vacuous. We call this regime as the *insufficient-capacity* regime. In this regime, even finding a feasible point within $\mathcal{X}_S$ is $\mathcal{NP}$-complete (Martello and Toth, 1990).

Due to this fundamentally different structure of $F(S)$, the dichotomy of $\mathcal{X}_S$ is important. Hence, we distinguish our structural results, proofs, and counterexamples between the two regimes. Next, for completeness, we provide some basic definitions of classical properties of set functions.

2.2. Classical Set-Function Properties

Definition 1 (Monotonicity). Consider the finite discrete set J. A function $F : 2^J \to \mathbb{R}^+$ is monotonically non-decreasing if for all $S \subseteq T \subseteq J$, $F(S) \leq F(T)$ while it is monotonically non-increasing if for all $S \subseteq T \subseteq J$, $F(S) \geq F(T)$. □

Definition 2 (Super/sub-additivity). Consider the finite discrete set J. A function $F : 2^J \to \mathbb{R}^+$ is superadditive if for all $S, T \subseteq J$, we have $F(S \cup T) \geq F(S) + F(T)$.

Analogously, F is subadditive if for all $S, T \subseteq J$, we have $F(S \cup T) \leq F(S) + F(T)$. □

Definition 3 (Super/sub-modularity). Consider the finite discrete set J. Then, the function $F: 2^J \to \mathbb{R}^+$ is supermodular if for all $S \subseteq J$ and $j, j^* \in J \setminus S$, where $j \neq j^*$, we have $F(S \cup \{j^*\}) + F(S \cup \{j\}) \leq F(S \cup \{j, j^*\}) + F(S)$.

Analogously, F is submodular if for all $S \subseteq J$ and $j, j^* \in J \setminus S$, where $j \neq j^*$, we have $F(S \cup \{j^*\}) + F(S \cup \{j\}) \geq F(S \cup \{j, j^*\}) + F(S)$. □

Definition 4 (Total curvature). Consider the finite discrete set J. The total curvature of a monotonically non-increasing function $F: 2^J \to \mathbb{R}^+$ is given by $c_F = 1 - \min_{j \in J} \frac{F(J \setminus \{j\}) - F(J)}{F(\emptyset) - F(\{j\})}$. □

Definition 5 (Monotonicity of u under removal). Let $u_j^S = \frac{\sum_{i \in I} W_{ij} x_{ij}^S}{C_j}, \forall j \in J$ denote an optimal utilization of facility j of model (2) with the convention $u_j^S = 0$ for $j \notin S$. Then, we say that the utilizations are monotonic under removal at S if for every $S \subseteq J$ and $j^* \in S$ the following holds: $u_j^S \leq u_j^{S \setminus \{j^*\}}, \forall j \in S \setminus \{j^*\}$. □

Below, we state two two propositions that we use later in this work. Proposition 2 is a special case of a more general result of composition of functions, see, (Topkis, 1998, Lemma 2.6.2), while Proposition 3 follows from elementary calculus.

Proposition. Let $a > 0$ and $C_s > 0, \forall s \in S > 0$. Define $H(T) = \frac{a}{\sum_{s \in T} C_s}$ on the domain $\{T \subseteq J : T \neq \emptyset\}$. Then, H is supermodular.

Proof. Proof: Consider distinct $j, j' \in J$ and $S \subseteq J \setminus \{j, j'\}$ with $S \neq \emptyset$. Let $y = \sum_{s \in S} C_s > 0$. Then,

$$
\begin{aligned}
& H(S \cup \{j, j'\}) - H(S \cup \{j\}) - H(S \cup \{j'\}) + H(S) \\
& \quad = a \left(\frac{1}{y + C_j + C_{j'}} - \frac{1}{y + C_j} - \frac{1}{y + C_{j'}} + \frac{1}{y} \right) && \text{(3a)} \\
& \quad = \frac{a \, C_j C_{j'} \, (C_j + C_{j'} + 2y)}{y \, (y + C_j) \, (y + C_{j'}) \, (y + C_j + C_{j'})} && \text{(3b)} \\
& \quad \geq 0. && \text{(3c)}
\end{aligned}
$$

Thus, from Definition 3, $H(T)$ is supermodular on all non-empty sets T. □ □

Proposition. Define $G(y) = \frac{y^2}{a} - 2y$ for $a > 0$. Then $G(y)$ is decreasing for all $y \leq a$.

Proof. Proof: We have $G'(y) = \frac{2y}{a} - 2 \leq 0$ for all $y \leq a$. □ □

In the proceeding sections, we examine whether the set-function F inherits any of the classical structural properties defined above. We begin with two basic results in this section.

Lemma 1. *Consider model* (2)*. Then, the function $F(S)$ is monotonically non-increasing according to Definition 1.*

Proof. Proof: If $F(S) < \infty$ (i.e., $\mathcal{X}(S) \neq \emptyset$), then an optimal solution for $\mathcal{X}(S)$ remains feasible (though possibly suboptimal) for $\mathcal{X}(S \cup k)$. Hence, $F(S \cup \{k\}) \leq F(S)$. If $F(S) = \infty$ (i.e., $\mathcal{X}(S) = \emptyset$), then the result trivially follows. □ □

Lemma 2. *Consider model* (2)*. Then, the function $F(S)$ is subadditive according to Definition 2.*

Proof. Proof: Consider sets $S, T \subseteq J$. Then, from Lemma 1 we have $F(S \cup T) \leq F(S)$ and $F(S \cup T) \leq F(T)$; thus, $F(S \cup T) \leq \min\{F(S), F(T)\}$. Since, $F(\cdot)$ is always non-negative, it follows that $F(S \cup T) \leq F(S) + F(T)$. □ □

3. Structural Properties of $F(S)$

Since monotonicity and subadditivity are relatively weak properties, we next ask whether the set function $F(S)$ admits stronger structures. As we mention in Section 1, supermodularity is a natural structure to explore here since the function value decreases as more facilities are added (see, Lemma 1), and the "marginal decrease" may become larger when additional facilities are already present. This intuition often underlies the appearance of supermodularity in combinatorial facility location models (see, e.g., Chudak and Williamson (2004)). In contrast, submodularity is less plausible here, since the marginal decrease need not shrink as more facilities are opened. However, as we show, $F(S)$ inherits neither submodularity nor supermodularity. We begin with the easier case of disproving submodularity.

3.1. Lack of Submodularity

Theorem 1. *Consider model (2). Then, the function $F(S)$ is not submodular in S for both the sufficient and insufficient-capacity regimes.*

Proof. Proof: We provide families of instances — for both the sufficient- and insufficient-capacity regimes — that violate submodularity.

1. *Sufficient-capacity*: Fix an integer $k \geq 2$. Let $I = \{i_1, \dots, i_k\}$ and $J = \{j_1, j_2\}$. Set all weights equal: $W_{ij} = 1$ for all $i \in I, j \in J$. Set both capacities equal $C_{j_1} = C_{j_2} = k$. We have $F(S) = \sum_{j \in J} C_j + \min_{x \in \mathcal{X}_S} \sum_{s \in S}(F(S) - 2L_s + \frac{L_s^2}{C_s})$, where $L_s = \sum_{i \in I} W_{is} x_{is}$.
   (i) $F(\emptyset) = C_{j_1} + C_{j_2} = 2k$.
   (ii) $F(\{j_1\}) = C_{j_2} + C_{j_1}\big(1 - \frac{L_1}{C_{j_1}}\big)^2$. Since the capacity of the open facility j_1 is sufficient to accommodate all users, we have $L_1 = k$; thus, the second term in $F(\{j_1\})$ is $k(1 - k/k)^2 = 0$. Hence, $F(\{j_1\}) = k$. By symmetry, $F(\{j_2\}) = k$.
   (iii) $F(\{j_1, j_2\}) = k\Big(1 - \frac{a}{k}\Big)^2 + k\Big(1 - \frac{b}{k}\Big)^2 = \frac{a^2 + b^2}{k}$, where a users be assigned to the open facility j_1 and the remaining $b = k - a$ users are assigned to the other open facility j_2. Minimizing this expression over the integers $a + b = k$, we have $a = \lfloor k/2 \rfloor$, and $b = \lceil k/2 \rceil$, hence $F(\{j_1, j_2\}) = \frac{\lfloor k/2 \rfloor^2 + \lceil k/2 \rceil^2}{k}$.

   From Definition 3, consider $S = \emptyset, j = \{j_1\}$ and $j^* = \{j_2\}$. Then, $F(S \cup \{j^*\}) + F(S \cup \{j\}) = 2k < \frac{\lfloor k/2 \rfloor^2 + \lceil k/2 \rceil^2}{k} + 2k = F(S \cup \{j, j^*\}) + F(S)$. Thus, the submodular inequality in Definition 3 does not hold.

2. *Insufficient-capacity*: Fix parameters $0 < \varepsilon < \alpha \leq \frac{1}{2}$. Let $I = \{i_1, i_2, i_3\}$ and $J = \{j_1, j_2, j_3, j_4\}$ with capacities $C_j = 1$ for all $j \in J$. Set weights W_{ij} as follows $W_{i_1 j_3} = W_{i_2 j_3} = \alpha; W_{i_1 j_1} = W_{i_2 j_2} = W_{i_3 j_3} = 1$ and $W_{ij} = \varepsilon$ for all other (i, j).

   Then, capacities of all j are insufficient since $\sum_{j \in J} W_i > C_j$. We have $F(S) = \min_{x \in \mathcal{X}_s} \sum_{s \notin S} C_s + \sum_{s \in S} C_s \Big(1 - \frac{L_s}{C_s}\Big)^2$, where $L_s = \sum_{i \in I} W_{is} x_{is}$. Consider $S = \{j_3, j_4\}$, $j = \{j_1\}$, and $j^* = \{j_2\}$.
   (i) $F(\{j_3, j_4\}) = (1 - 2\alpha)^2 + (1 - 1)^2 + C_{j_1} + C_{j_2} = (1 - 2\alpha)^2 + 2$, where i_1, i_2 are assigned to j_3 and i_3 to j_4.
   (ii) $F(\{j_1, j_3, j_4\}) = (1 - 1)^2 + (1 - \alpha)^2 + (1 - 1)^2 + C_{j_2} = (1 - \alpha)^2 + 1$, where i_1, i_2, i_3 are assigned to j_1, j_4, j_3, respectively.
   (iii) $F(\{j_2, j_3, j_4\}) = (1 - \alpha)^2 + 1$ since this case is symmetric to (ii).
   (iv) $F(\{j_1, j_2, j_3, j_4\}) = (1 - 1)^2 + (1 - 1)^2 + (1 - 0)^2 + (1 - 1)^2 = 1$, where i_1, i_2, i_3 are assigned to j_1, j_2, j_4, respectively with j_3 left unused.

   From Definition 3, $F(S \cup j^*) + F(S \cup j) = 2((1 - \alpha)^2 + 1) < (1 - 2\alpha)^2 + 3 = F(S \cup \{j, j^*\}) + F(S)$. Thus, the submodular inequality in Definition 3 does not hold.

□ □

Corollary 1. *Consider model (2). Then, the function $F(S)$ is not superadditive according to Definition 2.*

Proof. Proof: Consider the instance given in Theorem 1. For the sufficient-capacity regime, we have $F(\{j_1\})+F(\{j_2\}) = 2k > \frac{\lfloor k/2\rfloor^2+\lceil k/2\rceil^2}{k} = F(\{j_1, j_2\})$ which contradicts superadditivity. For the insufficient-capacity regime, we have $F(\{j_1, j2\}) + F(\{j_3, j_4\}) = \Big((1-2\varepsilon)^2 + (1-\varepsilon)^2 + 2\Big) + \Big((1-2\alpha)^2 + 2\Big) > 1 = F(\{j_1, j_2, j_3, j_4\})$ which again contradicts superadditivity. □ □

While the lack of submodularity is easy to establish by simple families of counterexamples as demonstrated by Theorem 1, the case of supermodularity is deeper.

3.2. Lack of Supermodularity

In contrast to Section 3.1, the proofs we present below are considerably more involved requiring separate handling of both the capacity regimes. We first present a few additional results that lead us to our key result of this section.

Lemma 3. *Consider $|I| = 5, |J| = 4$ with $\alpha, \delta \in \mathbb{R}^+$ defining the following instance of model (2): $W_{i_1,j_1} = W_{i_5,j_2} = \alpha, W_{i_2,j_3} = W_{i_3,j_4} = W_{i_4,j_3} = W_{i_4,j_4} = \frac{\alpha}{2}, W_{i_1,j_4} = W_{i_5,j_4} = \frac{\alpha}{20}$, and all other $W_{ij} = 1$; capacities $C_{j_1} = C_{j_2} = C_{j_3} = 5\alpha$, and $C_{j_4} = (5+\delta)\alpha$. Then, in this sufficient-capacity regime, the inequality*

$$F\big(\{j_1, j_3, j_4\}\big) + F\big(\{j_2, j_3, j_4\}\big) \;>\; F\big(\{j_1, j_2, j_3, j_4\}\big) + F\big(\{j_3, j_4\}\big).$$

holds for all $\alpha \geq 20$ and $\delta \in \big(\frac{1}{30}, \frac{19}{30}\big)$.

Proof. Proof: For $S \subseteq J$, define

$$L_s(x) = \sum_{i\in I} W_{is}x_{is}, \qquad \Phi(S, x) = \sum_{s\in S}(-2L_s(x) + \frac{L_s(x)^2}{C_s}), \qquad \Phi^\star(S) = \min_{x\in\mathcal{X}_S} \Phi(S, x).$$

Then, $F(S) = \sum_{j\in J} C_j + \Phi^\star(S)$. From Proposition 3, for $y \leftarrow L_s(x)$ and $a \leftarrow C_s$, $\Phi(S, x)$ decreases with increasing L_s if $L_s(x) \leq C_s$ for all s. Specifically, for the given instance, we have

$$\sum_{i\in I} W_{ij_1} = \alpha + 4, \quad \sum_{i\in I} W_{ij_2} = \alpha + 4, \quad \sum_{i\in I} W_{ij_3} = \alpha + 3, \quad \sum_{i\in I} W_{ij_4} = 1.1\alpha + 1.$$

Thus, for $\alpha \geq 20$ and $\delta > 0$, we have

$$C_{j_1} = 5\alpha \geq \alpha + 4, \quad C_{j_2} = 5\alpha \geq \alpha + 4, \quad C_{j_3} = 5\alpha \geq \alpha + 3, \quad C_{j_4} = (5+\delta)\alpha \geq 1.1\alpha + 1;$$

i.e., $L_s(x) \leq C_s$ holds for all $s \in J$ demonstrating that the capacities are sufficient and the hypothesis of Proposition 3 holds. Thus, an optimal solution to $\Phi^\star(S)$ (and, hence, to $F(S)$) is obtained by assigning each user i to a facility from the set $\arg\max_{s\in S} W_{is}$. Below, we determine this optimal assignment.

Since for $\alpha \geq 20$, the ordering $\alpha > \alpha/2 > \alpha/20 \geq 1$ holds, each of the following users has a unique maximum-weight facility:

$$i_1 \to j_1 \;\text{ if } j_1 \in S, \qquad i_2 \to j_3 \;\text{ if } j_3 \in S, \qquad i_3 \to j_4 \;\text{ if } j_4 \in S, \qquad i_5 \to j_2 \;\text{ if } j_2 \in S.$$

The only tie (i.e., multiple solutions for the argmax) is for i_4, since $W_{i_4,j_3} = W_{i_4,j_4} = \alpha/2$. We resolve this tie by comparing the quadratic terms in Φ (the linear terms are equal for both assignments). For brevity, let $\beta = \frac{\alpha}{5+\delta}$. Below we compute the loads explicitly by first excluding i_4 (denoted $\widetilde{L}_s$), then assigning i_4 to either j_3 or j_4, and finally comparing the resulting quadratic penalties.

(i) Consider $S = \{j_3, j_4\}$. Without i_4:

$$\widetilde{L}_3 = W_{i_2,j_3} = \frac{\alpha}{2}, \qquad \widetilde{L}_4 = W_{i_3,j_4} + W_{i_1,j_4} + W_{i_5,j_4} = \frac{\alpha}{2} + \frac{\alpha}{20} + \frac{\alpha}{20} = 0.60\alpha.$$

If $i_4 \to j_3$, then $L_3 = \widetilde{L}_3 + \frac{\alpha}{2} = \alpha$ and $L_4 = \widetilde{L}_4 = 0.60\alpha$. If $i_4 \to j_4$, then $L_3 = \widetilde{L}_3 = 0.50\alpha$ and $L_4 = \widetilde{L}_4 + \frac{\alpha}{2} = 1.10\alpha$. The quadratic terms in $\Phi(S)$ for both these scenarios (the linear terms are equal for both) are

$$\frac{L_3^2}{C_{j_3}} + \frac{L_4^2}{C_{j_4}} = \begin{cases} \dfrac{\alpha^2}{5\alpha} + \dfrac{(0.60\alpha)^2}{(5+\delta)\alpha}, & i_4 \to j_3, \\ = \dfrac{(0.50\alpha)^2}{5\alpha} + \dfrac{(1.10\alpha)^2}{(5+\delta)\alpha}, & i_4 \to j_4. \end{cases}$$

Then, $i_4 \to j_3$ is better if and only if

$$\frac{\alpha}{5} + \frac{0.36\,\alpha}{5+\delta} \;\le\; \frac{\alpha}{20} + \frac{1.21\,\alpha}{5+\delta} \iff \delta \;\le\; \frac{2}{3}.$$

Hence,

$$L_3 = \alpha, L_4 = 0.60\alpha \;\text{ if }\; \delta < \tfrac{2}{3}; \qquad L_3 = 0.50\alpha,\; L_4 = 1.10\alpha \;\text{ if }\; \delta > \tfrac{2}{3}.$$

Thus,

$$\Phi^\star(\{j_3, j_4\}) = \begin{cases} (-2\alpha + \frac{\alpha}{5}) + (-1.2\alpha + 0.36\beta) = -3.0\alpha + 0.36\beta, & \delta < \frac{2}{3}, \\ (-\alpha + \frac{\alpha}{20}) + (-2.2\alpha + 1.21\beta) = -3.15\alpha + 1.21\beta, & \delta > \frac{2}{3}. \end{cases} \tag{4}$$

(ii) Consider $S = \{j_1, j_3, j_4\}$. Without i_4:

$$\widetilde{L}_1 = W_{i_1,j_1} = \alpha, \qquad \widetilde{L}_3 = W_{i_2,j_3} = \frac{\alpha}{2}, \qquad \widetilde{L}_4 = W_{i_3,j_4} + W_{i_5,j_4} = \frac{\alpha}{2} + \frac{\alpha}{20} = 0.55\,\alpha.$$

If $i_4 \to j_3$, then $L_1 = \alpha$, $L_3 = \widetilde{L}_3 + \frac{\alpha}{2} = \alpha$, $L_4 = \widetilde{L}_4 = 0.55\alpha$. If $i_4 \to j_4$, then $L_1 = \alpha$, $L_3 = \widetilde{L}_3 = 0.50\alpha$, $L_4 = \widetilde{L}_4 + \frac{\alpha}{2} = 1.05\alpha$. The quadratic terms in $\Phi(S)$ for both these scenarios (all linear terms and the quadratic term for j_1 are equal) is

$$\frac{L_3^2}{C_{j_3}} + \frac{L_4^2}{C_{j_4}} = \begin{cases} \dfrac{\alpha^2}{5\alpha} + \dfrac{(0.55\alpha)^2}{(5+\delta)\alpha}, & i_4 \to j_3, \\ \dfrac{(0.50\alpha)^2}{5\alpha} + \dfrac{(1.05\alpha)^2}{(5+\delta)\alpha}, & i_4 \to j_4. \end{cases}$$

Then, $i_4 \to j_3$ is better if and only if

$$\frac{\alpha}{5} + \frac{0.3025\,\alpha}{5+\delta} \;\le\; \frac{\alpha}{20} + \frac{1.1025\,\alpha}{5+\delta} \iff \delta \;\le\; \frac{1}{3}.$$

Hence,

$$L_1 = \alpha,\; L_3 = \alpha,\; L_4 = 0.55\alpha \;\text{ if }\; \delta < \tfrac{1}{3}; \qquad L_1 = \alpha,\; L_3 = 0.50\alpha,\; L_4 = 1.05\alpha \;\text{ if }\; \delta > \tfrac{1}{3}.$$

Thus,

$$\Phi^\star(\{j_1, j_3, j_4\}) = \begin{cases} (-2\alpha + \frac{\alpha}{5}) + (-2\alpha + \frac{\alpha}{5}) + (-1.1\alpha + 0.3025\beta) \\ = -4.7\alpha + 0.3025\beta, \quad \delta < \frac{1}{3}, \\ (-2\alpha + \frac{\alpha}{5}) + (-\alpha + \frac{\alpha}{20}) + (-2.1\alpha + 1.1025\beta) \\ = -4.85\alpha + 1.1025\beta, \quad \delta > \frac{1}{3}. \end{cases} \tag{5}$$

(iii) Consider $S = \{j_2, j_3, j_4\}$. Without i_4:

$$\widetilde{L}_2 = W_{i_5,j_2} = \alpha, \qquad \widetilde{L}_3 = W_{i_2,j_3} = \frac{\alpha}{2}, \qquad \widetilde{L}_4 = W_{i_3,j_4} + W_{i_1,j_4} = \frac{\alpha}{2} + \frac{\alpha}{20} = 0.55\,\alpha.$$

If $i_4 \to j_3$, then $L_2 = \alpha$, $L_3 = \widetilde{L}_3 + \frac{\alpha}{2} = \alpha$, $L_4 = \widetilde{L}_4 = 0.55\alpha$. If $i_4 \to j_4$, then $L_2 = \alpha$, $L_3 = \widetilde{L}_3 = 0.50\alpha$, $L_4 = \widetilde{L}_4 + \frac{\alpha}{2} = 1.05\alpha$. The quadratic terms in $\Phi(S)$ for both these scenarios (all linear terms and the quadratic term for j_2 are equal) are

$$\frac{L_3^2}{C_{j_3}} + \frac{L_4^2}{C_{j_4}} = \begin{cases} \dfrac{\alpha^2}{5\alpha} + \dfrac{(0.55\alpha)^2}{(5+\delta)\alpha}, & i_4 \to j_3, \\ \dfrac{(0.50\alpha)^2}{5\alpha} + \dfrac{(1.05\alpha)^2}{(5+\delta)\alpha}, & i_4 \to j_4. \end{cases}$$

Then, $i_4 \to j_3$ is better if and only if

$$\frac{\alpha}{5} + \frac{0.3025\,\alpha}{5+\delta} \le \frac{\alpha}{20} + \frac{1.1025\,\alpha}{5+\delta} \iff \delta \le \frac{1}{3}.$$

Hence,

$$L_2 = \alpha,\ L_3 = \alpha,\ L_4 = 0.55\alpha \ \text{ if } \ \delta < \tfrac{1}{3}; \qquad L_2 = \alpha,\ L_3 = 0.50\alpha,\ L_4 = 1.05\alpha \ \text{ if } \ \delta > \tfrac{1}{3}.$$

Thus,

$$\Phi^\star(\{j_2, j_3, j_4\}) = \begin{cases} (-2\alpha + \frac{\alpha}{5}) + (-2\alpha + \frac{\alpha}{5}) + (-1.1\alpha + 0.3025\beta) \\ \ = -4.7\alpha + 0.3025\beta, \quad \delta < \frac{1}{3}, \\ (-2\alpha + \frac{\alpha}{5}) + (-\alpha + \frac{\alpha}{20}) + (-2.1\alpha + 1.1025\beta) \\ \ = -4.85\alpha + 1.1025\beta, \quad \delta > \frac{1}{3}. \end{cases} \tag{6}$$

(iv) Consider $S = \{j_1, j_2, j_3, j_4\}$. Without i_4:

$$\widetilde{L}_1 = W_{i_1,j_1} = \alpha, \qquad \widetilde{L}_2 = W_{i_5,j_2} = \alpha, \qquad \widetilde{L}_3 = W_{i_2,j_3} = \frac{\alpha}{2}, \qquad \widetilde{L}_4 = W_{i_3,j_4} = \frac{\alpha}{2}.$$

If $i_4 \to j_3$, then $L_1 = \alpha$, $L_2 = \alpha$, $L_3 = \widetilde{L}_3 + \frac{\alpha}{2} = \alpha$, $L_4 = \widetilde{L}_4 = \frac{\alpha}{2}$. If $i_4 \to j_4$, then $L_1 = \alpha$, $L_2 = \alpha$, $L_3 = \widetilde{L}_3 = \frac{\alpha}{2}$, $L_4 = \widetilde{L}_4 + \frac{\alpha}{2} = \alpha$. The quadratic terms in $\Phi(S)$ for both these scenarios (all linear terms and the quadratic terms for j_1 and j_2 are equal) are:

$$\frac{L_3^2}{C_{j_3}} + \frac{L_4^2}{C_{j_4}} = \begin{cases} \dfrac{\alpha^2}{5\alpha} + \dfrac{(0.50\alpha)^2}{(5+\delta)\alpha}, & i_4 \to j_3, \\ \dfrac{(0.50\alpha)^2}{5\alpha} + \dfrac{\alpha^2}{(5+\delta)\alpha}, & i_4 \to j_4. \end{cases}$$

Then, $i_4 \to j_3$ is better if and only if

$$\frac{\alpha}{5} + \frac{0.25\,\alpha}{5+\delta} \le \frac{\alpha}{20} + \frac{\alpha}{5+\delta} \iff \delta \le 0,$$

which cannot hold since $\delta > 0$. Hence,

$$L_1 = \alpha,\ L_2 = \alpha,\ L_3 = 0.50\alpha,\ L_4 = \alpha \quad \text{for all } \delta > 0.$$

Thus,

$$\begin{aligned} \Phi^\star(\{j_1, j_2, j_3, j_4\}) &= (-2\alpha + \tfrac{\alpha}{5}) + (-2\alpha + \tfrac{\alpha}{5}) + (-\alpha + \tfrac{\alpha}{20}) + (-2\alpha + \beta) \\ &= -6.55\alpha + \beta. \end{aligned} \tag{7}$$

Finally, let

$$\Gamma(\alpha,\delta) = F(\{j_1,j_3,j_4\}) + F(\{j_2,j_3,j_4\}) - F(\{j_1,j_2,j_3,j_4\}) - F(\{j_3,j_4\}) \tag{8a}$$
$$= \Phi^\star(\{j_1,j_3,j_4\}) + \Phi^\star(\{j_2,j_3,j_4\}) - \Phi^\star(\{j_1,j_2,j_3,j_4\}) - \Phi^\star(\{j_3,j_4\}). \tag{8b}$$

Since $\alpha \geq 20$, we need to determine the range of δ for which $\Gamma(\alpha,\delta)$ is strictly positive.

Plugging the formulas calculated in equations (4)-(7) in equation (8) yields

$$\Gamma(\delta,\alpha) = \begin{cases} \alpha\left(\frac{3}{20} - \frac{151}{200(5+\delta)}\right), & 0 < \delta < \frac{1}{3}, \\ \alpha\left(-\frac{3}{20} + \frac{169}{200(5+\delta)}\right), & \frac{1}{3} \leq \delta < \frac{2}{3}, \\ -\frac{\alpha}{200(5+\delta)}, & \delta \geq \frac{2}{3}. \end{cases} \tag{9}$$

Then, for $\alpha \geq 20$, it follows from plain algebra that $\Gamma(\delta,\alpha) > 0$ for $\delta \in (\frac{1}{30}, \frac{19}{30})$ which completes the proof. □ □

The following corollary to Lemma 3 assists in developing an insufficient-capacity regime family.

Corollary 2. *Consider the instance of model (2) defined in Lemma 3. For each j, define*

$$\eta_j = \frac{\max_S \sum_{i\in I} W_{ij}\, x_{ij}^S}{C_j}, \qquad \textit{and} \qquad \tau_j = \frac{\sum_{i\in I} W_{ij}}{C_j},$$

where $S \in \{\{j_3,j_4\}, \{j_1,j_3,j_4\}, \{j_2,j_3,j_4\}, \{j_1,j_2,j_3,j_4\}\} \subseteq \mathcal{F}$ is the collection of subsets considered in Lemma 3 and x_{ij}^S are the corresponding optimal assignments. Then, under the hypothesis of Lemma 3, $\eta_j < \tau_j$ for all $j \in J$.

Proof. Proof: For brevity, let $S_1 = \{j_3,j_4\}, S_2 = \{j_1,j_3,j_4\}, S_3 = \{j_2,j_3,j_4\}$ and $S_4 = \{j_1,j_2,j_3,j_4\}$. From the proof of Lemma 3, for each j, we have the following.

(i) For j_1:

$$\sum_i W_{ij_1} x_{ij_1}^{S_1} = 0, \quad \sum_i W_{ij_1} x_{ij_1}^{S_2} = \alpha, \quad \sum_i W_{ij_1} x_{ij_1}^{S_3} = 0, \quad \sum_i W_{ij_1} x_{ij_1}^{S_4} = \alpha.$$

Hence,

$$\eta_{j_1} = \frac{\max\{0,\,\alpha,\,0,\,\alpha\}}{5\alpha} = \frac{\alpha}{5\alpha} = \frac{1}{5}.$$

(ii) For j_2:

$$\sum_i W_{ij_2}\, x_{ij_2}^{S_1} = 0, \quad \sum_i W_{ij_2}\, x_{ij_2}^{S_2} = 0, \quad \sum_i W_{ij_2}\, x_{ij_2}^{S_3} = \alpha, \quad \sum_i W_{ij_2}\, x_{ij_2}^{S_4} = \alpha.$$

Hence,

$$\eta_{j_2} = \frac{\max\{0,\,0,\,\alpha,\,\alpha\}}{5\alpha} = \frac{1}{5}.$$

(iii) For j_3:

$$\sum_i W_{ij_3}\, x_{ij_3}^{S_1} = \begin{cases} \alpha, & \delta < \frac{2}{3}, \\ \frac{\alpha}{2}, & \delta > \frac{2}{3}, \end{cases} \quad \sum_i W_{ij_3}\, x_{ij_3}^{S_2} = \begin{cases} \alpha, & \delta < \frac{1}{3}, \\ \frac{\alpha}{2}, & \delta > \frac{1}{3}, \end{cases}$$

$$\sum_i W_{ij_3}\, x_{ij_3}^{S_3} = \begin{cases} \alpha, & \delta < \frac{1}{3}, \\ \frac{\alpha}{2}, & \delta > \frac{1}{3}, \end{cases} \qquad \sum_i W_{ij_3}\, x_{ij_3}^{S_4} = \frac{\alpha}{2}.$$

Hence, for every $\delta < \frac{2}{3}$ (and, thus, for $\delta \in (\frac{1}{30}, \frac{19}{30})$ as in the hypothesis),

$$\eta_{j_3} = \frac{\max\{\alpha,\,\alpha \text{ or } \frac{\alpha}{2},\,\alpha \text{ or } \frac{\alpha}{2},\,\frac{\alpha}{2}\}}{5\alpha} = \frac{\alpha}{5\alpha} = \frac{1}{5}.$$

(iv) For j_4:

$$\sum_i W_{ij_4}\ x^{S_1}_{ij_4} = \begin{cases} 0.60\,\alpha, & \delta < \frac{2}{3}, \\ 1.10\,\alpha, & \delta > \frac{2}{3}, \end{cases} \quad \sum_i W_{ij_4}\ x^{S_2}_{ij_4} = \begin{cases} 0.55\,\alpha, & \delta < \frac{1}{3}, \\ 1.05\,\alpha, & \delta > \frac{1}{3}, \end{cases}$$

$$\sum_i W_{ij_4}\ x^{S_3}_{ij_4} = \begin{cases} 0.55\,\alpha, & \delta < \frac{1}{3}, \\ 1.05\,\alpha, & \delta > \frac{1}{3}, \end{cases} \qquad \sum_i W_{ij_4}\ x^{S_4}_{ij_4} = 1.00\,\alpha.$$

Hence

$$\eta_{j_4} = \frac{1}{(5+\delta)\alpha} \cdot \max\Big\{\ 0.60\alpha \text{ or } 1.10\alpha,\ \ 0.55\alpha \text{ or } 1.05\alpha,\ \ 0.55\alpha \text{ or } 1.05\alpha,\ \ 1.00\alpha\ \Big\}.$$

For $\delta \in \left(\frac{1}{30}, \frac{19}{30}\right)$ as in the hypothesis, we have:

$$\eta_{j_4} = \begin{cases} \dfrac{1.00}{5+\delta}, & \delta \in \left(\frac{1}{30}, \frac{1}{3}\right), \\ \dfrac{1.05}{5+\delta}, & \delta \in \left[\frac{1}{3}, \frac{19}{30}\right). \end{cases}$$

The quantity τ is directly calculated via

$$\sum_i W_{ij_1} = \alpha + 4, \quad \sum_i W_{ij_2} = \alpha + 4, \quad \sum_i W_{ij_3} = \alpha + 3, \quad \sum_i W_{ij_4} = 1.1\alpha + 1,$$

yielding

$$\tau_{j_1} = \tau_{j_2} = \frac{\alpha+4}{5\alpha} = \tfrac{1}{5} + \tfrac{4}{5\alpha}, \quad \tau_{j_3} = \tfrac{\alpha+3}{5\alpha} = \tfrac{1}{5} + \tfrac{3}{5\alpha}, \quad \tau_{j_4} = \tfrac{1.1\alpha+1}{(5+\delta)\alpha} = \tfrac{1.1}{5+\delta} + \tfrac{1}{(5+\delta)\alpha}.$$

Since $\alpha \geq 20$ and $\delta \in (\frac{1}{30}, \frac{19}{30})$, we have $\eta_j < \tau_j$ for all $j \in J$. □ □

Lemma 4. *Consider the same instance of model (2) as in Lemma 3 except with capacities C'_j set to θ_j times the previous capacities C_j (i.e., $C'_j = \theta_j C_j$ with all else remaining unchanged). Choose θ_j with $\eta_j < \theta_j < \tau_j$; such a θ_j always exists from Corollary 2. Then, this instance is in the insufficient-capacity regime, and the inequality*

$$F(\{j_1, j_3, j_4\}) + F(\{j_2, j_3, j_4\}) > F(\{j_1, j_2, j_3, j_4\}) + F(\{j_3, j_4\})$$

continues to hold for all $\alpha \geq 20$ and $\delta \in \left(\frac{1}{30}, \frac{19}{30}\right)$.

Proof. Proof: We reuse the notation from Lemma 3. For $S \subseteq J$ and any assignment $x \in \mathcal{X}_S$, let

$$L_s(x) = \sum_{i\in I} W_{is} x_{is}, \qquad \Phi(S,x) = \sum_{s\in S}\Big(-2L_s(x) + \frac{L_s(x)^2}{C'_s}\Big), \qquad \Phi^\star(S) = \min_{x\in\mathcal{X}_S} \Phi(S,x).$$

Hence, $F(S) = \sum_{j\in J} C'_j + \Phi^\star(S)$. Let $\bar{x}^S$ be the optimal assignments from Lemma 3, and for brevity let $\bar{L}_j(S) = L_j(\bar{x}^S)$. From Corollary 2 there exists θ_j with $\eta_j < \theta_j < \tau_j$. Setting $C'_j = \theta_j C_j$ yields, for every j,

$$\bar{L}_j(S) \leq \max_{S'} \bar{L}_j(S') = \eta_j C_j < \theta_j C_j = C'_j.$$

Thus, the previously optimal assignment, $\bar{x}^S$, is feasible for the considered instance with capacities C'. Further, $\sum_i W_{ij} = \tau_j C_j > \theta_j C_j = C'_j, \forall j \in J$. Thus, the singletons $\{j\}$ are infeasible for model (2) demonstrating that we are in the insufficient-capacity regime. Next, we determine the optimal assignments for this instance.

Since $\bar{L}_s(S) < C'_s$ for all s, S, the capacities remain nonbinding and the setting of Lemma 3 applies for $\alpha \geq 20$. Then, the arg max is unique for

$$i_1 \to j_1 \ \text{ if } j_1 \in S, \qquad i_2 \to j_3 \ \text{ if } j_3 \in S, \qquad i_3 \to j_4 \ \text{ if } j_4 \in S, \qquad i_5 \to j_2 \ \text{ if } j_2 \in S.$$

The only tie (i.e., multiple solutions for the argmax) is again for i_4, since $W_{i_4,j_3} = W_{i_4,j_4} = \alpha/2$. Similar to the proof of Lemma 3, we resolve this tie by comparing the quadratic terms in Φ (the linear terms are equal for both assignments). Below we compute the loads explicitly by first excluding i_4 (denoted $\widetilde{L}_s$), then assigning i_4 to either j_3 or j_4, and finally comparing the resulting quadratic penalties.

(i) Consider $S = \{j_3, j_4\}$. Without i_4:

$$\widetilde{L}_3 = W_{i_2,j_3} = \frac{\alpha}{2}, \qquad \widetilde{L}_4 = W_{i_3,j_4} + W_{i_1,j_4} + W_{i_5,j_4} = \frac{\alpha}{2} + \frac{\alpha}{20} + \frac{\alpha}{20} = 0.60\alpha.$$

If $i_4 \to j_3$, then $L_3 = \widetilde{L}_3 + \frac{\alpha}{2} = \alpha$ and $L_4 = \widetilde{L}_4 = 0.60\alpha$. If $i_4 \to j_4$, then $L_3 = \widetilde{L}_3 = 0.50\alpha$ and $L_4 = \widetilde{L}_4 + \frac{\alpha}{2} = 1.10\alpha$. The quadratic terms in $\Phi(S)$ for both these scenarios (the linear terms are equal for both) are

$$\frac{L_3^2}{C'_{j_3}} + \frac{L_4^2}{C'_{j_4}} = \begin{cases} \dfrac{\alpha^2}{5\alpha\,\theta_3} + \dfrac{(0.60\alpha)^2}{(5+\delta)\alpha\,\theta_4}, & i_4 \to j_3, \\ \dfrac{(0.50\alpha)^2}{5\alpha\,\theta_3} + \dfrac{(1.10\alpha)^2}{(5+\delta)\alpha\,\theta_4}, & i_4 \to j_4. \end{cases}$$

Then, $i_4 \to j_3$ is better if and only if

$$\frac{\alpha}{5\theta_3} + \frac{0.36\,\alpha}{(5+\delta)\theta_4} \;\le\; \frac{\alpha}{20\theta_3} + \frac{1.21\,\alpha}{(5+\delta)\theta_4} \iff \theta_4 \;\le\; \frac{17}{3}\cdot\frac{\theta_3}{5+\delta}.$$

Hence,

$$L_3 = \alpha, L_4 = 0.60\alpha \;\text{ if }\; \theta_4 \le \tfrac{17}{3}\tfrac{\theta_3}{5+\delta}; \qquad L_3 = 0.50\alpha,\; L_4 = 1.10\alpha \;\text{ if }\; \theta_4 > \tfrac{17}{3}\tfrac{\theta_3}{5+\delta}.$$

Thus,

$$\Phi^\star(\{j_3, j_4\}) = \begin{cases} (-2\alpha + \frac{\alpha}{5\theta_3}) + (-1.2\alpha + \frac{0.36\,\alpha}{(5+\delta)\theta_4}), & \theta_4 \le \frac{17}{3}\frac{\theta_3}{5+\delta}, \\ (-\alpha + \frac{\alpha}{20\theta_3}) + (-2.2\alpha + \frac{1.21\,\alpha}{(5+\delta)\theta_4}), & \theta_4 > \frac{17}{3}\frac{\theta_3}{5+\delta}. \end{cases} \tag{10}$$

(ii) Consider $S = \{j_1, j_3, j_4\}$. Without i_4:

$$\widetilde{L}_1 = W_{i_1,j_1} = \alpha, \qquad \widetilde{L}_3 = W_{i_2,j_3} = \frac{\alpha}{2}, \qquad \widetilde{L}_4 = W_{i_3,j_4} + W_{i_5,j_4} = \frac{\alpha}{2} + \frac{\alpha}{20} = 0.55\,\alpha.$$

If $i_4 \to j_3$, then $L_1 = \alpha$, $L_3 = \widetilde{L}_3 + \frac{\alpha}{2} = \alpha$, $L_4 = \widetilde{L}_4 = 0.55\alpha$. If $i_4 \to j_4$, then $L_1 = \alpha$, $L_3 = \widetilde{L}_3 = 0.50\alpha$, $L_4 = \widetilde{L}_4 + \frac{\alpha}{2} = 1.05\alpha$. The quadratic terms in $\Phi(S)$ for both these scenarios (all linear terms and the quadratic term for j_1 are equal) is

$$\frac{L_3^2}{C'_{j_3}} + \frac{L_4^2}{C'_{j_4}} = \begin{cases} \dfrac{\alpha^2}{5\alpha\,\theta_3} + \dfrac{(0.55\alpha)^2}{(5+\delta)\alpha\,\theta_4}, & i_4 \to j_3, \\ \dfrac{(0.50\alpha)^2}{5\alpha\,\theta_3} + \dfrac{(1.05\alpha)^2}{(5+\delta)\alpha\,\theta_4}, & i_4 \to j_4. \end{cases}$$

Then, $i_4 \to j_3$ is better if and only if

$$\frac{\alpha}{5\theta_3} + \frac{0.3025\,\alpha}{(5+\delta)\theta_4} \;\le\; \frac{\alpha}{20\theta_3} + \frac{1.1025\,\alpha}{(5+\delta)\theta_4} \iff \theta_4 \;\le\; \frac{16}{3}\cdot\frac{\theta_3}{5+\delta}.$$

Hence,

$$L_1 = \alpha, \quad L_3 = \alpha, \quad L_4 = 0.55\alpha \quad \text{if } \theta_4 \le \tfrac{16}{3}\tfrac{\theta_3}{5+\delta},$$

and,

$$L_1 = \alpha, \quad L_3 = 0.50\alpha, \quad L_4 = 1.05\alpha \quad \text{if } \theta_4 > \tfrac{16}{3}\tfrac{\theta_3}{5+\delta}.$$

Thus,

$$\Phi^\star(\{j_1, j_3, j_4\}) = \begin{cases} (-2\alpha + \frac{\alpha}{5\theta_1}) + (-2\alpha + \frac{\alpha}{5\theta_3}) + (-1.1\alpha + \frac{0.3025\,\alpha}{(5+\delta)\theta_4}), & \theta_4 \le \frac{16}{3}\frac{\theta_3}{5+\delta}, \\ (-2\alpha + \frac{\alpha}{5\theta_1}) + (-\alpha + \frac{\alpha}{20\theta_3}) + (-2.1\alpha + \frac{1.1025\,\alpha}{(5+\delta)\theta_4}), & \theta_4 > \frac{16}{3}\frac{\theta_3}{5+\delta}. \end{cases} \tag{11}$$

(iii) Consider $S = \{j_2, j_3, j_4\}$. Without i_4:

$$\widetilde{L}_2 = W_{i_5,j_2} = \alpha, \qquad \widetilde{L}_3 = W_{i_2,j_3} = \frac{\alpha}{2}, \qquad \widetilde{L}_4 = W_{i_3,j_4} + W_{i_1,j_4} = \frac{\alpha}{2} + \frac{\alpha}{20} = 0.55\,\alpha.$$

If $i_4 \to j_3$, then $L_2 = \alpha$, $L_3 = \widetilde{L}_3 + \frac{\alpha}{2} = \alpha$, $L_4 = \widetilde{L}_4 = 0.55\alpha$. If $i_4 \to j_4$, then $L_2 = \alpha$, $L_3 = \widetilde{L}_3 = 0.50\alpha$, $L_4 = \widetilde{L}_4 + \frac{\alpha}{2} = 1.05\alpha$. The quadratic terms in $\Phi(S)$ for both these scenarios (all linear terms and the quadratic term for j_2 are equal) are

$$\frac{L_3^2}{C'_{j_3}} + \frac{L_4^2}{C'_{j_4}} = \begin{cases} \dfrac{\alpha^2}{5\alpha\,\theta_3} + \dfrac{(0.55\alpha)^2}{(5+\delta)\alpha\,\theta_4}, & i_4 \to j_3, \\ \dfrac{(0.50\alpha)^2}{5\alpha\,\theta_3} + \dfrac{(1.05\alpha)^2}{(5+\delta)\alpha\,\theta_4}, & i_4 \to j_4. \end{cases}$$

Then, $i_4 \to j_3$ is better if and only if

$$\frac{\alpha}{5\theta_3} + \frac{0.3025\,\alpha}{(5+\delta)\theta_4} \le \frac{\alpha}{20\theta_3} + \frac{1.1025\,\alpha}{(5+\delta)\theta_4} \iff \theta_4 \le \frac{16}{3} \cdot \frac{\theta_3}{5+\delta}.$$

Hence,

$$L_2 = \alpha,\ L_3 = \alpha,\ L_4 = 0.55\alpha \quad \text{if } \theta_4 \le \tfrac{16}{3}\tfrac{\theta_3}{5+\delta}$$

and,

$$L_2 = \alpha,\ L_3 = 0.50\alpha,\ L_4 = 1.05\alpha \quad \text{if } \theta_4 > \tfrac{16}{3}\tfrac{\theta_3}{5+\delta}.$$

Thus,

$$\Phi^\star(\{j_2, j_3, j_4\}) = \begin{cases} (-2\alpha + \frac{\alpha}{5\theta_2}) + (-2\alpha + \frac{\alpha}{5\theta_3}) + (-1.1\alpha + \frac{0.3025\,\alpha}{(5+\delta)\theta_4}), & \theta_4 \le \frac{16}{3}\frac{\theta_3}{5+\delta}, \\ (-2\alpha + \frac{\alpha}{5\theta_2}) + (-\alpha + \frac{\alpha}{20\theta_3}) + (-2.1\alpha + \frac{1.1025\,\alpha}{(5+\delta)\theta_4}), & \theta_4 > \frac{16}{3}\frac{\theta_3}{5+\delta}. \end{cases} \tag{12}$$

(iv) Consider $S = \{j_1, j_2, j_3, j_4\}$. Without i_4:

$$\widetilde{L}_1 = W_{i_1,j_1} = \alpha, \qquad \widetilde{L}_2 = W_{i_5,j_2} = \alpha, \qquad \widetilde{L}_3 = W_{i_2,j_3} = \frac{\alpha}{2}, \qquad \widetilde{L}_4 = W_{i_3,j_4} = \frac{\alpha}{2}.$$

If $i_4 \to j_3$, then $L_1 = \alpha$, $L_2 = \alpha$, $L_3 = \widetilde{L}_3 + \frac{\alpha}{2} = \alpha$, $L_4 = \widetilde{L}_4 = \frac{\alpha}{2}$. If $i_4 \to j_4$, then $L_1 = \alpha$, $L_2 = \alpha$, $L_3 = \widetilde{L}_3 = \frac{\alpha}{2}$, $L_4 = \widetilde{L}_4 + \frac{\alpha}{2} = \alpha$. The quadratic terms in $\Phi(S)$ for both these scenarios (all linear terms and the quadratic terms for j_1 and j_2 are equal) are:

$$\frac{L_3^2}{C'_{j_3}} + \frac{L_4^2}{C'_{j_4}} = \begin{cases} \dfrac{\alpha^2}{5\alpha\,\theta_3} + \dfrac{(0.50\alpha)^2}{(5+\delta)\alpha\,\theta_4}, & i_4 \to j_3, \\ \dfrac{(0.50\alpha)^2}{5\alpha\,\theta_3} + \dfrac{\alpha^2}{(5+\delta)\alpha\,\theta_4}, & i_4 \to j_4. \end{cases}$$

Then, $i_4 \to j_4$ is better if and only if

$$\frac{\alpha}{5\theta_3} + \frac{0.25\,\alpha}{(5+\delta)\theta_4} \ge \frac{\alpha}{20\theta_3} + \frac{\alpha}{(5+\delta)\theta_4} \iff \theta_4 \ge \frac{5\,\theta_3}{5+\delta}.$$

Hence,

$$L_1 = \alpha,\ L_2 = \alpha,\ L_3 = 0.50\alpha,\ L_4 = \alpha \quad \text{if} \quad \theta_4 \ge \tfrac{5\,\theta_3}{5+\delta}.$$

Thus,

$$\Phi^\star(\{j_1, j_2, j_3, j_4\}) = (-2\alpha + \tfrac{\alpha}{5\theta_1}) + (-2\alpha + \tfrac{\alpha}{5\theta_2}) + (-\alpha + \tfrac{0.25\,\alpha}{5\theta_3}) + (-2\alpha + \tfrac{\alpha}{(5+\delta)\theta_4}). \tag{13}$$

Finally, let

$$\Gamma(\alpha,\delta,\theta) = F(\{j_1,j_3,j_4\}) + F(\{j_2,j_3,j_4\}) - F(\{j_1,j_2,j_3,j_4\}) - F(\{j_3,j_4\}) \tag{14a}$$
$$= \Phi^\star(\{j_1,j_3,j_4\}) + \Phi^\star(\{j_2,j_3,j_4\}) - \Phi^\star(\{j_1,j_2,j_3,j_4\}) - \Phi^\star(\{j_3,j_4\}). \tag{14b}$$

We need to find the range where $\Gamma(\alpha,\delta,\theta)$ is strictly positive. When θ_4 is chosen such that case (i) selects $i_4 \to j_3$, cases (ii)–(iii) select $i_4 \to j_3$, and case (iv) selects $i_4 \to j_4$, (i.e., $\frac{5\,\theta_3}{5+\delta} \le \theta_4 \le \frac{16}{3}\frac{\theta_3}{5+\delta}$ and $\theta_4 \le \frac{17}{3}\frac{\theta_3}{5+\delta}$), then plugging the formulas calculated in equations (10)-(13) in equation (14) yields

$$\Gamma(\alpha,\delta,\theta) = \alpha\left(\tfrac{0.15}{\theta_3} - \tfrac{0.755}{(5+\delta)\theta_4}\right). \tag{15}$$

From equation (15), $\Gamma(\alpha,\delta,\theta) > 0$ if $\dfrac{0.15}{\theta_3} > \dfrac{0.755}{(5+\delta)\theta_4}$. Since $\eta_j < \theta_j < \tau_j$ and the interval $\left[\frac{5\,\theta_3}{5+\delta}, \frac{16}{3}\frac{\theta_3}{5+\delta}\right]$ is nonempty, there exist values for θ satisfying these conditions. Thus, for all $\alpha \ge 20$ and $\delta \in \left(\frac{1}{30}, \frac{19}{30}\right)$, the inequality in the lemma's statement holds true. □ □

With this background, we now present the main result of this section.

Theorem 2. *Consider model (2). Then, the function $F(S)$ is not supermodular in S for both the sufficient- and insufficient-capacity regimes.*

Proof. Proof: Lemma 3 and Lemma 4 provide instances of sufficient- and insufficient-capacity regimes, respectively, where the supermodular inequality in Definition 3 does not hold.

Concretely, for Lemma 3, consider $\alpha = 1{,}000$ and $\delta = 0.34$, which satisfy its hypothesis. For these parameters, we have (with numerical values rounded to one decimal place)

$$\begin{aligned} F(\{j_1,j_3,j_4\}) + F(\{j_2,j_3,j_4\}) &= 15{,}696.5 + 15{,}696.5 && (16a)\\ &= 31{,}393.0 && (16b)\\ &> 31{,}384.7 && (16c)\\ &= 13{,}977.3 + 17{,}407.4 && (16d)\\ &= F(\{j_1,j_2,j_3,j_4\}) + F(\{j_3,j_4\}). && (16e)\end{aligned}$$

Thus, the supermodularity inequality is violated with a gap of 8.3.

Similarly, for Lemma 4, consider $u = 1{,}000$ and $\delta = 0.34$, with $\theta_{j_1} = \theta_{j_2} = 0.2004, \theta_{j_3} = 0.2003, \theta_{j_4} = 0.20615$ which satisfies the hypothesis of Lemma 4. For these parameters, we have (with numerical values rounded to one decimal place)

$$\begin{aligned} F(\{j_1,j_3,j_4\}) + F(\{j_2,j_3,j_4\}) &= 1{,}255.5 + 1{,}255.5 && (17a)\\ &= 2{,}511.0 && (17b)\\ &> 2{,}492.4 && (17c)\\ &= 260.4 + 2232.0 && (17d)\\ &= F(\{j_1,j_2,j_3,j_4\}) + F(\{j_3,j_4\}). && (17e)\end{aligned}$$

Thus, the supermodularity inequality is still violated, now with a gap of 18.6. □ □

To summarize, our analysis establishes that $F(S)$ is neither submodular nor supermodular, for both sufficient- and insufficient-capacity regimes, despite the empirical nestedness observed in practice. This absence of classical structure motivates the search for measures of deviation from modularity.

4. Total Curvature

In this section, we study the total curvature of F which provides a measure of how far F lies from the extremal cases of modularity.

Theorem 3. *Assume the monotonicity under removal in Definition 5 holds. Then, the total curvature of $F(S)$ in model (2) satisfies $c_F \in [0,1]$.*

Proof. Proof: Applying Definition 5 with $S = J$ and a fixed $j^* \in J$ yields

$$u_j^J \le u_j^{J\setminus\{j^*\}}, \qquad \forall j \in J \setminus \{j^*\}. \tag{18}$$

Further, $0 \le u_j^J \le 1$ for all $j \in J$, $0 \le u_j^{J\setminus\{j^*\}} \le 1$ for all $j \in J \setminus \{j^*\}$, and, by convention, $u_{j^*}^{J\setminus\{j^*\}} = 0$. From equation (18), for this fixed j^*, we have

$$1 - u_j^J \ge 1 - u_j^{J\setminus\{j^*\}}, \qquad \forall j \in J \setminus \{j^*\}, \tag{19a}$$

$$C_j(1 - u_j^J)^2 \ge C_j(1 - u_j^{J\setminus\{j^*\}})^2, \qquad \forall j \in J \setminus \{j^*\}, \tag{19b}$$

$$\sum_{j\in J\setminus\{j^*\}} C_j(1 - u_j^J)^2 \ge \sum_{j\in J\setminus\{j^*\}} C_j(1 - u_j^{J\setminus\{j^*\}})^2. \tag{19c}$$

Further, by recursively applying equation (18), for any fixed $j \in J$, we have

$$u_j^J \le u_j^{\{j\}}, \tag{20a}$$

$$2u_j^J - (u_j^J)^2 \le 2u_j^{\{j\}} - (u_j^{\{j\}})^2. \tag{20b}$$

Here, equation (20a) follows from equation (18) by successively removing each $j^* \in J \setminus \{j\}$, yielding the chain $u_j^J \le u_j^{J\setminus\{j_1^*\}} \le \cdots \le u_j^{J\setminus(J\setminus\{j\})} = u_j^{\{j\}}$. Then equation (20b) follows from Proposition 3 by applying the decreasing function g with $a = 1$ on both sides.

Next, we calculate the numerator and denominator of the total curvature in Definition 4 for this j^*. For the numerator, we have:

$$F\left(J \setminus \{j^*\}\right) - F(J) = \sum_{j\ne j^*} C_j \Big[\big(1 - u_j^{J\setminus\{j^*\}}\big)^2 - \big(1 - u_j^J\big)^2\Big] + C_{j^*}\big[1 - \big(1 - u_{j^*}^J\big)^2\big] \tag{21a}$$

$$= \sum_{j\ne j^*} C_j \big(u_j^{J\setminus\{j^*\}} - u_j^J\big)\big(u_j^{J\setminus\{j^*\}} + u_j^J - 2\big) + C_{j^*}\big(2u_{j^*}^J - (u_{j^*}^J)^2\big) \tag{21b}$$

$$\le C_{j^*}\big(2u_{j^*}^J - (u_{j^*}^J)^2\big). \tag{21c}$$

Here, equations (21a) and (21b) follow directly by expanding the squares and rearranging the terms. Equation (21c) follows from equation (18) (since $j \ne j^*$) and the fact that u is upper bounded by one; thus, each summand in the first term of equation (21b) is non-positive.

For the denominator, with $F(\emptyset) = \sum_{j\in J} C_j$, we have,

$$F(\emptyset) - F(\{j^*\}) = \sum_{j\in J} C_j - \Big(C_{j^*}(1 - u_{j^*}^{\{j^*\}})^2 + \sum_{j\in J\setminus\{j^*\}} C_j\Big) \tag{22a}$$

$$= C_{j^*}\Big[1 - (1 - u_{j^*}^{\{j^*\}})^2\Big] \tag{22b}$$

$$= C_{j^*}\big(2u_{j^*}^{\{j^*\}} - (u_{j^*}^{\{j^*\}})^2\big). \tag{22c}$$

Thus, for this j^*, we have,

$$\rho_{j^*} = \frac{F\big(J \setminus \{j^*\}\big) - F(J)}{F(\emptyset) - F(\{j^*\})}, \tag{23a}$$
$$\leq \frac{C_{j^*}\big(2u_{j^*}^J - (u_{j^*}^J)^2\big)}{C_{j^*}\big(2u_{j^*}^{\{j^*\}} - (u_{j^*}^{\{j^*\}})^2\big)}, \tag{23b}$$
$$= \frac{2u_{j^*}^J - (u_{j^*}^J)^2}{2u_{j^*}^{\{j^*\}} - (u_{j^*}^{\{j^*\}})^2} \tag{23c}$$
$$\leq 1, \tag{23d}$$

where equation (23b) follows by direct substitution from equations (21) and (22), equation (23c) follows directly, while equation (23d) follows from equation (20) with $j = j^*$. Since both the numerator and denominator of ρ_j are positive from Lemma 1, we have $\rho_{j^*} \geq 0$. As j^* was chosen arbitrarily, we have $0 \leq \rho_j \leq 1$ for all $j \in J$. Thus,

$$c_F = 1 - \min_{j \in J} \rho_j \in [0, 1].$$

□ □

Theorem 3 shows that, if the monotonicity assumption holds, the total curvature c_F of our set function is guaranteed to lie within the canonical interval $[0, 1]$. Hence — even though $F(S)$ lacks submodularity and supermodularity — it still admits a bounded curvature measure that quantifies its deviation from modularity. In the following section, we examine the extent to which this positive result depends on the monotonicity of utilizations under removal, and explore special cases of model (2) where this property is indeed guaranteed.

5. Special Cases and Extended Feasible Regions

As we show in Section 3, our considered set function $F(S)$ is neither submodular nor supermodular, in general, for both the sufficient- and insufficient-capacity regimes. The derivation of the total curvature in Theorem 3 relies on the monotonicity of utilizations under removal (Definition 5) which also does not hold true in general; we present a simple numerical illustration in Example 2 in Appendix Appendix A.2. These negative results suggest that, in full generality, $F(S)$ resists breaking its combinatorial structure. However, certain structural settings restore these strong properties. In this section, we uncover special cases where utilizations behave monotonically under removal, allowing us to derive curvature bounds and even supermodularity. Thus, although the general model lacks submodularity and supermodularity, the special cases reveal an unexpectedly rich structure. To analyze such special cases, we first introduce an extended feasible region of $F(S)$ that serves as a technical tool throughout this section.

5.1. Extended Feasible Region and Monotonicity

Definition 6. Consider model (2) with $W_{is} = W_i > 0, \forall i \in I, s \in S$. Then, we define the fractional-feasible assignment set as follows

$$\mathcal{X}_S^{\text{frac}} = \Big\{ x \in [0,1]^{I \times S} : \sum_{s \in S} x_{is} = 1, \forall i \in I; \sum_{i \in I} W_i x_{is} \leq C_s, \forall s \in S \Big\} \supseteq \mathcal{X}_s.$$

The set $\mathcal{X}_S^{\text{frac}}$ frequently appears in the combinatorial optimization literature on facility location problems, see, e.g., Shmoys et al. (1997); An et al. (2017). Further, it allows the monotonicity guarantee.

Theorem 4. *Consider model (2) with $\mathcal{X}_S$ relaxed to $\mathcal{X}_S^{\text{frac}}$. Then, the optimal utilizations of model (2) are monotonic under removal of $j^* \subseteq J$ under all $S \subseteq \mathcal{F}$; i.e., Definition 5 holds.*

Proof. Proof: Under the hypothesis, the utilization of an open facility $s \subset J$ is given by $u_s^S = \frac{\overline{W} x_{is}^S}{C_s}$. Thus, $\sum_{s\in S} u_s C_s = \sum_{i\in I, s\in S} W_i x_{is}^S = \sum_{i\in I} W_i$, where the last equality holds from equation (2e). Let $L_s^S = \sum_{i\in I} W_i x_{is}^S \in [0, C_s]$ and let $\overline{W} = \sum_{i\in I} W_i = \sum_{s\in S} L_s^S = \sum_{s\in S} u_s C_s$ (this quantity is independent of set S). Then, expanding the objective function (2a) for a given S, we have

$$F(S) = \sum_{s\in S}\Big(C_s - 2L_s^S + \frac{(L_s^S)^2}{C_s}\Big) = \Big(\sum_{s\in S} C_s - 2\overline{W}\Big) + \sum_{s\in S}\frac{(L_s^S)^2}{C_s};$$

thus, minimizing $F(S)$ is equivalent to

$$\min\left\{\sum_{s\in S}\frac{y_s^2}{C_s}, \quad \text{s.t. } 0 \le y_s \le C_s,\ \sum_{s\in S} y_s = \overline{W}\right\}. \tag{24}$$

By the Cauchy–Schwarz inequality, $\overline{W}^2 = \big(\sum_{s\in S} y_s\big)^2 \le \big(\sum_{s\in S} C_s\big)\big(\sum_{s\in S} y_s^2/C_s\big)$. Thus,

$$\sum_{s\in S}\frac{y_s^2}{C_s} \ \ge\ \frac{\overline{W}^2}{\sum_{s\in S} C_s},$$

with equality holding if and only if $\frac{y_s}{\sqrt{C_s}} = \bar{u}\sqrt{C_s}, \forall s \in S$ for some $\bar{u} > 0$. Thus, a minimizer for model (24) has the property:

$$y_s^S = \bar{u}^S C_s, \qquad u_s^S = \bar{u}^S \quad \text{for all } s \in S, \quad \text{where} \quad \bar{u}^S = \frac{\overline{W}}{\sum_{k\in S} C_k} \in [0,1].$$

Applying this result to the sets S and $S \setminus \{j^*\}$ we have,

$$u_j^S = \frac{\overline{W}}{\sum_{s\in S} C_s} \quad \forall s \in J, \qquad u_j^{S\setminus\{j^*\}} = \frac{\overline{W}}{\sum_{s\in S\setminus\{j^*\}} C_s} = \frac{\overline{W}}{\sum_{s\in S} C_s - C_{j^*}} \quad \forall s \in S \setminus \{j^*\}.$$

Since $C_{j^*} > 0$, we have $u_j^{S\setminus\{j^*\}} \ge u_j^S$ for all $j \in S \setminus \{j^*\}$ which is exactly the monotonicity property in Definition 5. □ □

The proof of Theorem 4 uses the fact that, under the hypothesis, the utilizations of all facilities in the set S are equal. This fact is also shown via Theorem 2 of Schmitt and Singh (2024) holds which proves so-called proportional fairness for such a setting; i.e., if model (2) is feasible in this special case, then there exists an optimal solution with equal utilization for all $j \in S$. Specifically, this utilization is the same as the utilization we compute in Theorem 4. In this sense, Theorem 4 provides a different proof of Theorem 2 of Schmitt and Singh (2024) by employing the Cauchy–Schwarz inequality.

5.2. *Projected Load-Based Convex Reformulation*

The equalization of utilization for a fixed set S hints at a deeper structural simplification. In particular, it suggests that the relevant assignment decision variables, x, are understandable entirely in terms of facility loads, L_s, instead of assignments. We next formalize this mapping and show that the new feasible region in this projected space coincides with a simple polytope.

Proposition. Consider model (2) with $\mathcal{X}_S$ relaxed to $\mathcal{X}_S^{\text{frac}}$. Let $L_s(x) = \sum_{i\in I} W_i x_{is}, \forall s \in S$, and consider the S-dimensional polytope

$$\mathcal{L}_S = \Big\{L \in \mathbb{R}_+^S : \sum_{s\in S} L_s = \sum_{i\in I} W_i; L_s \le C_s, \forall s \in S\Big\},$$

where $\overline{W} = \sum_{i\in I} W_i$. Then, $\{L(x) : x \in \mathcal{X}_S^{\text{frac}}\} \ = \ \mathcal{L}_S$.

Proof. Proof: (⇒) Consider a $x \in \mathcal{X}_S^{\text{frac}}$. Then, $L_s(x) \geq 0$ and

$$\sum_{s\in S} L_s(x) = \sum_{s\in S} \overline{W} x_{is} = \overline{W} \sum_{s\in S} x_{is} = \overline{W}.$$

Further, $L_s(x) \leq C_s$ holds directly by the definition of $L_s(x)$. Hence $L(x) \in \mathcal{L}_S$.

(⇐) Consider a $L \in \mathcal{L}_S$. Define, $x_{is} = \frac{L_s}{\overline{W}}, \forall i \in I, s \in S$. Then, $x_{is} \in [0,1]$ since $L_s \in [0, \overline{W}]$. Further, $\sum_{s\in S} x_{is} = \frac{\sum_{s\in S} L_s}{\overline{W}} = \frac{\overline{W}}{\overline{W}} = 1$ holds for all $i \in I$. Finally, since $L_s \leq C_s$, we have $\overline{W} x_{is} \leq C_s \implies \sum_{i\in I} W_i x_{is} \leq C_s$. Hence, $x \in \mathcal{X}_S^{\text{frac}}$. □ □

The equivalence of the feasible regions $\mathcal{X}_S^{\text{frac}}$ and $\mathcal{L}_S$ is powerful since it reduces the combinatorial structure of assignments into a basic convex optimization problem. Leveraging this polyhedral description, we now solve for the optimal loads explicitly. As we show next, the optimal load is proportional to capacities with the proportionality constant depending only on the sum of the weights.

Proposition. Consider model (2) with $\mathcal{X}_S$ relaxed to $\mathcal{X}_S^{\text{frac}}$. Assume the capacity-sufficient regime holds (i.e., $\sum_{s\in S} C_s \geq \sum_{i\in I} W_i = \overline{W}$ for every nonempty $S \subseteq J$) for model (2). Then, $\min_L \left\{ \sum_{s\in S} \frac{L_s^2}{C_s} \ : \ L \in \mathcal{L}_S \right\} = \frac{\overline{W}^2}{\sum_{s\in S} C_s}$, where $\mathcal{L}_S$ is as defined in Proposition 4. Further, the optimal objective function value is attained at $L_s^* = \frac{\overline{W} C_s}{\sum_{r\in S} C_r}$, $\forall s \in S$.

Proof. Proof: The feasible region $\mathcal{L}_S$ is the following linear system:

$$L_s \geq 0, \qquad \forall s \in S, \tag{25a}$$

$$\sum_{s\in S} L_s = \sum_{i\in I} W_i, \tag{25b}$$

$$L_s \leq C_s, \qquad \forall s \in S. \tag{25c}$$

Let $\mu_s \geq 0$, $\lambda \in \mathbb{R}$, and $\nu_s \geq 0$ denote the Lagrange multipliers for constraints (25a), (25b), and (25c), respectively. Then, the considered optimization model is a strictly convex quadratic program with linear constraints; hence, it has a unique optimal solution. Thus, the Karush-Kuhn-Tucker (KKT) optimality conditions — given by primal feasibility (25a)-(25c), dual feasibility, and the following stationarity plus complementary slackness conditions — are necessary and sufficient.

$$\frac{2L_s}{C_s} - \lambda - \mu_s + \nu_s = 0, \qquad \forall s \in S, \tag{26a}$$

$$\mu_s L_s = 0, \qquad \forall s \in S, \tag{26b}$$

$$\nu_s (L_s - C_s) = 0, \qquad \forall s \in S. \tag{26c}$$

Under the hypothesis $\sum_{i\in I} W_i \leq \sum_{s\in S} C_s$, we distinguish two cases.

- Case (i): $\sum_{i\in I} W_i < \sum_{s\in S} C_s$. Assume that $0 < L_s < C_s$ for all s (i.e., L_s is in the interior of $\mathcal{L}_s$). Then by complementary slackness (26b) and (26c), we have $\mu_s = \nu_s = 0$ for all s. Thus,

$$L_s = \tfrac{\lambda}{2} C_s, \qquad \forall s \in S, \tag{27a}$$

$$\implies \lambda = \frac{2\overline{W}}{\sum_{r\in S} C_r}, \tag{27b}$$

$$\implies L_s^* = \frac{(\overline{W}) C_s}{\sum_{r\in S} C_r}, \qquad \forall s \in S. \tag{27c}$$

Here, equation (27a) follows from the stationarity condition (26a), equation (27b) follows by enforcing constraint (25b) which yields a candidate solution, L_s^*, given by equation (27c). Then, this candidate solution L_s^* satisfies the assumption $0 < L_s^* < C_s$ for all s; hence, it is a valid feasible solution satisfying all the KKT conditions. Further, by strict convexity of the objective function, it is the unique optimal solution; thus, validating the assumption. The corresponding optimal objective function value follows by direct substitution.

- Case (ii): $\sum_{i\in I} W_i = \sum_{s\in S} C_s$. Then, $L_s^* = C_s$ is the unique solution. The corresponding optimal objective function value follows by direct substitution.

□ □

5.3. *Supermodularity of the Relaxed Model*

With the closed-form characterization of the optimal loads from Section 5.2, and employing results from Section 2, we now directly substitute these into the definition of $F(S)$. This yields an explicit formula for $F(S)$ which immediately leads us to the supermodularity of $F(S)$ under the capacity-sufficient regime.

Theorem 5. *Consider model (2) with $\mathcal{X}_S$ relaxed to $\mathcal{X}_S^{\text{frac}}$. If the capacity-sufficient regime holds, then the set function $F(S)$ is supermodular on the family of nonempty $S \subseteq J$.*

Proof. Proof: Consider a fixed $S \subseteq \mathcal{F}$. Under the hypothesis, we have

$$F(S) = \sum_{j\in J} C_j - 2\overline{W} + \min_{x\in X_S^{\text{frac}}} \frac{\sum_{i\in I, s\in S}(W_i x_{is})^2}{C_s}, \tag{28}$$

where $\overline{W} = \sum_{i\in I} W_i$. Projecting $\sum_{i\in I} W_i x_{is}$ to loads L_s, under the hypothesis, it follows from Proposition 4 that the feasible region $\mathcal{X}_S^{\text{frac}}$ is equivalent to the polytope $\mathcal{L}_S$. Then, the optimal value of the minimizing term in equation (28) is given by Proposition 5, and we have

$$F(S) = \sum_{j\in J} C_j - 2\overline{W} + \frac{\overline{W}^2}{\sum_{s\in S} C_s}. \tag{29a}$$

In equation (29a), the first two terms on the right-hand-side are independent of S; hence, proving supermodularity is equivalent to proving supermodularity of the third term alone. Consider $a = \overline{W}$; then, it follows from Proposition 2 that $F(S)$ is supermodular. □ □

Thus, supermodularity holds in the relaxed space. However, the ultimate question is whether such a property persists under the binary restrictions of the original model. We answer this question affirmatively in the next section.

5.4. *Binary Realizability*

Since $\mathcal{X}_S \subseteq \mathcal{X}_S^{\text{frac}}$, naturally, any feasible binary assignment, x^S, is also feasible for its continuous relaxation; i.e., optimizing over $\mathcal{X}_S^{\text{frac}}$ provides a lower bound for $F(S)$. The proportional load identified in Theorem 5 achieves the minimum in $\mathcal{X}_S^{\text{frac}}$. Thus, any binary assignment $x^S \in \mathcal{X}_S$ that induces exactly this proportional loads achieves the same objective value as the lower bound, and, hence, is optimal for $F(S)$. The remaining question is whether such a binary assignment exists. Corollary 3 provides a positive answer.

Corollary 3. *Consider model (2) with $W_{is} = W_i > 0, \forall i \in I, s \in S$. Assume there exists an optimal solution $x^S \in \mathcal{X}_S$ such that $\sum_{i\in I} W_i x_{is}^S = \frac{C_s}{\sum_{j\in S} C_j}\overline{W}$, $\forall s \in S$; i.e., the optimal utilization, $u^S = \frac{\overline{W}}{\sum_{j\in S} C_j}$, is the same for all $s \in S$. Then, the set function $F(S)$ is supermodular on the family of non-empty $S \subseteq J$.*

Proof. Proof: For a fixed $S \in \mathcal{F}$ consider the optimal solution $x^S \in \mathcal{X}_S$ given in the hypothesis. Then, the value of $F(S)$ is the same as that calculated in equation (29a) and the rest of the proof of Theorem 5 goes through. Thus, $F(S)$ is supermodular even with the binary restrictions in place. □ □

Corollary 3 reduces the challenge to an implementability question: is such a binary assignment actually achievable? The following lemmas identify two arithmetic conditions — equal weights and rationally proportional weights — under which this proportional allocation is attainable. In other words, in these special cases, the set function $F(S)$ is supermodular even under the binary restriction.

Lemma 5. *In Corollary 3, let $W_i = W > 0, \forall i \in I$, and let the capacity-sufficient regime hold. For a fixed $S \subseteq J$, define*

$$K_s \;=\; |I|\frac{C_s}{\sum_{j\in S} C_j}, \qquad s \in S.$$

If $K_s \in \mathbb{Z}_{\geq 0}, \forall s \in S$ then the optimal load (or, equivalently utilization) defined in Corollary 3 is attainable.

Proof. Proof: Under the hypothesis, for a fixed S, partition the set of I users into disjoint subsets with K_s users assigned to facility s. Such a partition exists because each K_s is an integer and $\sum_{s\in S} K_s = |I|$. Let $x_{is}^S = 1$ if user i is placed in the partition corresponding to facility s, and $x_{is}^S = 0$ otherwise. Then, for any $s \in S$,

$$\sum_{i\in I} W_i x_{is}^S = \sum_{i\in I_s} W_i = \sum_{i\in I_s} W = W|I_s| = WK_s = W|I|\frac{C_s}{\sum_{j\in S} C_j} = \frac{C_s}{\sum_{jinS} C_j}\sum_{i\in I} W_i$$

which equals the optimal load in Corollary 3. Equivalently, the utilization at s is $\frac{\sum_{i\in I} W_i}{\sum_{j\in S} C_j}$ which is the same as in Corollary 3. Thus, the optimal value is attainable from the set $\mathcal{X}_S$. □ □

Lemma 6. *In Corollary 3, let $W_i = M_i\epsilon, \forall i \in I$ with $M_i \in \mathbb{Z}_{>0}$ and $\epsilon > 0$, and let the capacity-sufficient regime hold. For a fixed $S \subseteq J$, define*

$$Q_s \;=\; \frac{C_s}{\sum_{j\in S} C_j}\cdot\frac{1}{\epsilon}\sum_{i\in I} W_i, \qquad s \in S.$$

If $Q_s \in \mathbb{Z}_{\geq 0}, \forall s \in S$, and if the multiset $\{M_i : i \in I\}$ can be partitioned into disjoint subsets whose sums equal Q_s for each $s \in S$, then the optimal load (or, equivalently utilization) defined in Corollary 3 is attainable.

Proof. Proof: Under the hypothesis, for a fixed S, partition the set I of users into disjoint subsets such that the total number of users assigned to facility s is Q_s. Such a partition exists by the hypothesis that each Q_s is an integer and that $\{M_i : i \in I\}$ admits partitions with the required bin-sums. Let $x_{is}^S = 1$ if user i is placed in the partition corresponding to facility s, I_s, and $x_{is}^S = 0$ otherwise. Then, for any $s \in S$, we have

$$\sum_{i\in I} W_i x_{is}^S \;=\; \sum_{i\in I_s} W_i \;=\; \sum_{i\in I_s} M_i\epsilon \;=\; \epsilon\sum_{i\in I_s} M_i \;=\; \epsilon\, Q_s \;=\; \frac{C_s}{\sum_{j\in S} C_j}\sum_{i\in I} W_i$$

which equals the optimal load in Corollary 3. Equivalently, the utilization at s is $\frac{\sum_{i\in I} W_i}{\sum_{j\in S} C_j}$ which is the same as in Corollary 3. The corresponding optimal utilization is attained as well. Thus, the optimal value is attainable from the set $\mathcal{X}_S$. Example 3 in the appendix provides a numerical illustration of the partition in Lemma 6. □ □

6. Conclusions

We investigated the combinatorial structure of a set function $F(S)$ arising from an existing quadratic optimization model with binary assignment decision variables. This model has been studied in the integer programming community through its interaction between the convex quadratic objective and classical facility-location polytopes. Applications of this model have also been studied in waste-management. However, our work took a different perspective by analyzing the foundational structure of the induced set function itself.

Despite strong empirical evidence of nested optimal solutions in practice, we established that $F(S)$ is neither submodular nor supermodular in general. These fundamental results reveal the structural fragility of the problem in its most general form. To address this negative result, we analytically derived the total curvature as a measure of deviation from modularity. We proved that, under suitable assumptions, the curvature is bounded by one; this property partially explains the empirically observed nestedness. Further, more positively, we identified structural regimes — based on continuous relaxations and special parameterizations of the problem data — under which $F(S)$ admits monotonic utilizations, closed-form representations, and even supermodularity. We showed how these regimes not only recover an elegant structure of a rich combinatorial problem but also connected directly with proportional fairness concepts that are pervasive in optimization and engineering contexts.

Our analysis suggests several avenues for future research. Developing further structural regimes or special cases — such as, via Voronoi orderings or dominance relationships — may provide a richer explanation of the nestedness observed in real-world instances. Algorithmic implications of our result further warrant research, e.g., developing approximation guarantees for greedy algorithms utilizing the bounded curvature and the supermodularity of special cases. Finally, given the applied demonstrations of this problem in previous works, our theoretical insights could pave way for new applications in undesirable facility location problems.

Acknowledgment

I am grateful to Milan Studený for fruitful discussions and, in particular, assistance in a numerical illustration of Lemma 3.

References

An, H.C., Singh, M., Svensson, O., 2017. LP-based algorithms for capacitated facility location. SIAM Journal on Computing 46, 272–306. doi:10.1137/151002320.

Chudak, F.A., Williamson, D.P., 2004. Improved approximation algorithms for capacitated facility location problems. Mathematical Programming 102, 207–222. doi:10.1007/s10107-004-0524-9.

Fujishige, S., 2005. Submodular functions and optimization. volume 58. Elsevier.

Martello, S., Toth, P., 1990. Knapsack problems: algorithms and computer implementations. John Wiley & Sons, Inc., New York.

Nemhauser, G.L., Wolsey, L.A., Fisher, M.L., 1978. An analysis of approximations for maximizing submodular set functions—I. Mathematical Programming 14, 265–294. doi:10.1007/bf01588971.

Schmidt, M., Singh, B., 2025. The balanced facility location problem: Complexity and heuristics. INFORMS Journal on Computing doi:10.1287/ijoc.2024.0693.

Schmitt, C., Singh, B., 2024. Quadratic optimization models for balancing preferential access and fairness: Formulations and optimality conditions. INFORMS Journal on Computing 36, 1150–1167. doi:10.1287/ijoc.2022.0308.

Shmoys, D.B., Tardos, É., Aardal, K., 1997. Approximation algorithms for facility location problems, in: Proceedings of the twenty-ninth annual ACM symposium on theory of computing, Springer Berlin Heidelberg. pp. 265–274. doi:10.1007/3-540-44436-x_4.

Topkis, D.M., 1998. Supermodularity and complementarity. Frontiers of economic research, Princeton University Press, Princeton, N.J. doi:10.1515/9781400822539.

Appendix A. Appendix

Appendix A.1. Real-world instances with nested optimal solutions

A key motivation for this work was that previously available real-world instances of model (2) display a strong nestedness property of optimal solutions. Indeed, in our computational experiments conducted on several randomized instances of model (2) this nestedness is observed. We do not present all of these results; they are direct runs from the publicly available dataset of Schmidt and Singh (2025). Since monotonocity of the function $F(S)$ is relatively easy to establish (as we show in Section 3 of the main text), such a nested structure supports a possible supermodular property. In the following numerical illustration, we detail the optimal solutions from the county of Hampshire in the UK using publicly available data on recycling centers; see, `https://github.com/35436506/Hampshire_HWRC_closure/tree/main`.

Example 1. Consider an instance of model (2) available at this GitHub page consisting of $|I| = 234, |J| = 26$. Here, the set J denotes all the recycling centers in Hampshire, UK while I denotes all the ZIP codes in this county. Then, W_{ij} denotes preference-weighted population weights for an assignment from i to j while C_j denotes the capacity of a recycling center. For details on the practical aspects of this instance, see the above-mentioned GitHub page.

This instance is capacity-insufficient in the sense that not all singleton sets (i.e., $S = \{j\}$) are feasible; however, $F(S)$ is finite as an optimal assignment does exist even for $|S| = 1$. We compute $F(S)$ using Gurobi to an optimality gap of 0% (due to the problem's small-scale, this is computationally easy even on a laptop). Table A.1 reports the selected facilities for representative budgets, $|S|$, illustrating the nested structure in optimal solutions of the function $F(S)$ for each budget value.

We further ran an additional experiment on the dataset provided in Schmidt and Singh (2025) from the Nürnberg-Erlangen-Fürth area in Germany. This instance has $|I| = 234, |J| = 26$ and is the smallest instance reported in in Schmidt and Singh (2025). This instance is strongly capacity-insufficient in the sense all singleton sets (i.e., $S = \{j\}$) except one are infeasible; thus, $F(S)$ is still finite. Here, we again observe the nested property although not for all budget values.

Appendix A.2. Numerical illustrations

Example 2. Consider an instance of model (2) with $I = \{i_1, i_2, i_3\}$ and $J = \{j_1, j_2, j_3\}$ with $C_j = 1, \forall j$ and weights given by:

$$W = \begin{pmatrix} W_{i_1j_1} & W_{i_1j_2} & W_{i_1j_3} \\ W_{i_2j_1} & W_{i_2j_2} & W_{i_2j_3} \\ W_{i_3j_1} & W_{i_3j_2} & W_{i_3j_3} \end{pmatrix} = \begin{pmatrix} 0.60 & 0.55 & 0.40 \\ 0.55 & 0.45 & 0.60 \\ 0.50 & 0.60 & 0.55 \end{pmatrix}.$$

For $S = J$, an optimal assignment yields utilizations

$$u^J_{j_1} = \tfrac{0.60}{1} = 0.60, \qquad u^J_{j_2} = \tfrac{0.60}{1} = 0.60, \qquad u^J_{j_3} = \tfrac{0.60}{1} = 0.60.$$

(i) Remove j_1: $S = J \setminus \{j_1\} = \{j_2, j_3\}$. An optimal assignment yields utilizations

$$u^{J\setminus\{j_1\}}_{j_2} = \tfrac{0.60}{1} = 0.60, \qquad u^{J\setminus\{j_1\}}_{j_3} = \tfrac{0.40+0.60}{1} = 1.00.$$

(ii) Remove j_2: $S = J \setminus \{j_2\} = \{j_1, j_3\}$. An optimal assignment yields utilizations

$$u^{J\setminus\{j_2\}}_{j_1} = \tfrac{0.55}{1} = 0.55, \qquad u^{J\setminus\{j_2\}}_{j_3} = \tfrac{0.40+0.55}{1} = 0.95.$$

(iii) Remove j_3: $S = J \setminus \{j_3\} = \{j_1, j_2\}$. An optimal assignment yields utilizations

$$u^{J\setminus\{j_3\}}_{j_1} = \tfrac{0.50}{1} = 0.50, \qquad u^{J\setminus\{j_3\}}_{j_2} = \tfrac{0.55+0.45}{1} = 1.00.$$

Table A.1: Chain of optimal selections of model (2) for the real-world case of Hampshire, UK. For each $|S|$, the set of open facilities is nested. For details, see Example 1 and the GitHub page cited in Appendix Appendix A.1.

$\lvert S\rvert$	Facility
1	{Basingstoke}
2	{Basingstoke, Portsmouth}
3	{Basingstoke, Portsmouth, Southampton}
4	{Basingstoke, Portsmouth, Southampton, Eastleigh}
5	{Basingstoke, Portsmouth, Southampton, Eastleigh, Netley}
6	{Basingstoke, Portsmouth, Southampton, Eastleigh, Netley, Gosport}
7	{Basingstoke, Portsmouth, Southampton, Eastleigh, Netley, Gosport, Andover}
8	{Basingstoke, ..., Andover, Aldershot}
9	{Basingstoke, ..., Aldershot, Segensworth}
10	{Basingstoke, ..., Segensworth, Efford}
11	{Basingstoke, ..., Efford, Farnborough}
12	{Basingstoke, ..., Farnborough, Havant}
13	{Basingstoke, ..., Havant, Winchester}
14	{Basingstoke, ..., Winchester, Waterlooville}
15	{Basingstoke, ..., Waterlooville, Bordon}
16	{Basingstoke, ..., Bordon, Hedge End}
17	{Basingstoke, ..., Hedge End, Somerley}
18	{Basingstoke, ..., Somerley, Hartley Wintney}
19	{Basingstoke, ..., Hartley Wintney, Casbrook}
20	{Basingstoke, ..., Casbrook, Alton}
21	{Basingstoke, ..., Alton, Petersfield}
22	{Basingstoke, ..., Petersfield, Bishops Waltham}
23	{Basingstoke, ..., Bishops Waltham, Hayling Island}
24	{Basingstoke, ..., Hayling Island, Fair Oak}
25	{Basingstoke, ..., Fair Oak, Alresford}
26	{Basingstoke, ..., Alresford, Marchwood}

For the second and third cases above, we have $u_{j_1}^J = 0.60 > u_{j_1}^{J\setminus\{j_2\}} = 0.55$ and $u_{j_1}^J = 0.60 > u_{j_1}^{J\setminus\{j_3\}} = 0.50$, respectively, which violate Definition 5. □

Example 3. Consider an instance of model (2) with $|I| = 5$ and $|J| = 3$ with $C_{j_1} = 60, C_{j_2} = 50, C_{j_3} = 40$ and weights given by

$$W_1 = 30 \quad W_2 = 40, \quad W_3 = 20 \quad W_4 = 50, \quad W_5 = 10.$$

. Let $S = \{j_1, j_2, j_3\}$. Under the hypothesis of Lemma 6, consider $\epsilon = 10$ with $M_1 = 3, M_2 = 4, M_3 = 2, M_4 = 5, M_5 = 1$. Then, the corresponding Q values are $Q_{j_1} = 6, Q_{j_2} = 5, Q_{j_3} = 4$; thus, $Q \in \mathbb{Z}_{\geq 0}$.

We seek to partition the multiset $\{M_i\} = \{3, 4, 2, 5, 1\}$ into three groups with sums $6, 5, 4$, respectively. A valid partition is

$$\{M_4 = 5, M_5 = 1\} \to j_1, \quad \{M_1 = 3, M_3 = 2\} \to j_2, \quad \{M_2 = 4\} \to j_3.$$

Let $x_{is}^S = 1$ if user i is assigned to the bin for facility s and $x_{is}^S = 0$ otherwise. Then the respective resulting loads on the three facilities are:

$$W_4 + W_5 = 50 + 10 = 60 = Q_{j_1}\epsilon;$$
$$W_1 + W_3 = 30 + 20 = 50 = Q_{j_2}\epsilon;$$
$$W_2 = 40 = Q_{j_3}\epsilon = L_{j_3},$$

demonstrating the attainability of the optimal solution of Corollary 3; i.e., the integer and fractional optimal solutions coincide. The corresponding utilization is 1 for all $s \in S$, providing $F(S) = 0$. □