Proceedings of Machine Learning Research 129:1–25, 2020 ACML 2020

Polytime Decomposition of Generalized Submodular Base Polytopes with Efficient Sampling

Aadirupa Saha AADIRUPA@IISC.AC.IN
Indian Institute of Science, Bangalore, India

Editors: Sinno Jialin Pan and Masashi Sugiyama

Abstract

We consider the problem of efficient decomposition of a given point x in an n-dimensional convex polytope into convex combination of its extreme points. Besides the widespread scopes of the problem in theory of convex polytopes in mathematics, the problem also has applications in online combinatorial optimization problems. Towards this we first propose a general class of convex polytopes–*Generalized Submodular Base Polytopes* (GSBPs)–that includes several well known convex polytopes as its special case including permutahedron, k-forest, spanning tree, combinatorial subset choice polytopes. We next propose a general decomposition algorithm for the above class of GSBPs that uses the novel idea of first decomposing the given point into at most n *face centers*, and further decomposing each face center into *extreme points* of their corresponding faces. In addition, we discover a few special class of *partition-respecting* and *symmetric* GSBPs for which the above two steps could be performed in respectively $O(n^2 + nT(f))$ and $O(n^2T(f))$ [1] time. We also give a complete characterization of the underlying submodular function f, for which the associated GSBP satisfies the above properties. One interesting fact is that we show that the support of the resulting decomposition with our proposed algorithm is only *poly*(n) in the number of extreme points which respects *efficient sampling* from the resulting distribution. Finally we corroborate our theoretical results with empirical evaluations.

1. Introduction

The theory of convex polytopes has many applications across mathematics and computer science Gao and Lauder (2001); Ewald (2012); Meurant (2014). Of course the core problem has significance in mathematics in its own right. In machine learning community a very relevant application is online learning in combinatorial decision spaces, e.g. in online routing, job scheduling, subset selection, resource allocation etc., Rahmanian et al. (2016); Dai et al. (2017); Audibert et al. (2013); Hazan et al. (2016a); Hazan and Kale (2009); Chen et al. (2008); Rakhlin et al. (2010); Gopalan et al. (2014); Neu and Valko (2014) where the goal is to find the "optimal (profit maximizing) decision" in an online fashion. However classical algorithms like Exponential-Weight or Weighted-Majority algorithm Freund and Schapire (1997); Arora et al. (2012), fail to give real time performances as they require to maintain a weight distribution over the decision simplex, which is of combinatorially large dimension in the current setup Littlestone and Warmuth (1994); Kale (2007). Here precisely the polytope decomposition becomes useful as now it suffices to maintain a point in a suitably chosen polytope whose extreme points has an one-to-one mapping to each point in the decision

1. $T(f)$ is the unit time to evaluate the submodular function f at any input point.

space, and one can recover back the desired distribution by decomposing that point into convex combination of the extreme points of the polytopes; for instance the Alg. 1 of Suehiro et al. (2012), *PermELearn* algorithm Helmbold and Warmuth (2009), or more generally LDOMD algorithm Rajkumar and Agarwal (2014) etc.

Related Works. In recent years, the above problem was studied for some specific polytopes arising out of different combinatorial structures, e.g. rankings, spanning trees, k-forests, job scheduling, subset selection, shortest paths etc Koolen et al. (2010); Helmbold and Warmuth (2009); Yasutake et al. (2011, 2012); Rajkumar and Agarwal (2014); Fujita et al. (2013); Hazan et al. (2016b); Bubeck (2011); Fujita et al. (2014); Audibert et al. (2013, 2014). Many of these polytopes turn out to be associated with an underlying submodular function, known as submodular base polytopes (SBPs). Suehiro et al. (2012) give a general $O(n^6 + n^5T(f))$ time algorithm for decomposing a point in a SBP into its extreme points by solving a submodular minimization problem ($T(f)$ being the unit time to evaluate the submodular function f). Yasutake et al. (2011) consider the permutahedron polytope, and give a $O(n^2)$ decomposition algorithm for the same. Generalizing above to any cardinality based SBPs (of which permutahedron is a special case), Suehiro et al. (2012) also give an $O(n^2)$ time decomposition algorithm. In another line of work Hoeksma et al. (2014) consider the single machine scheduling polytopes (SMSP) which in general are not SBPs, and give a $O(n^2)$ time decomposition for the same.

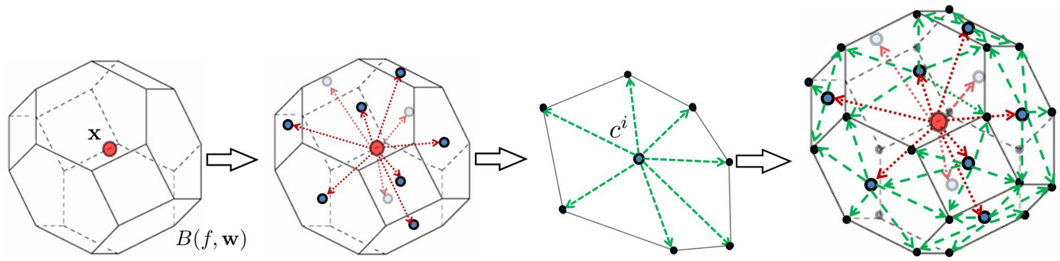
From the above chain of developments, the following questions thus arise naturally: *Are there interesting generalizations of SBPs that respects an $O(poly(n))$ decomposition? Can we characterize the structural properties of such a polytope under which such decompositions are possible? What is the common intuition behind generalizing the above algorithms?*

Contributions. Our specific contributions are as follows:

1. We conceptualize a class of convex polytopes: *generalized submodular base polytopes* (GSBP), that generalizes several well studied class of convex polytopes including permutahedrons Yasutake et al. (2012), spanning-trees, k-subsets, SBPs, and SMSPs Suehiro et al. (2012); Hoeksma et al. (2014); Rajkumar and Agarwal (2014); Koolen et al. (2010) etc.

2. **Polytime decomposition into face centers (Sec. 4):** Towards this, we first define a special class of *partition-respecting* GSBP (Def. 7), and give an $O(n^2 + nT(f))$ time efficient algorithm that decomposes any point $\mathbf{x}$ of the GSBP into convex combination of at most n of its face centers, say $\mathbf{c}^1, \mathbf{c}^2, \ldots, \mathbf{c}^t \in \mathbb{R}^n$, such that $\mathbf{x} = \sum_{i=1}^{t} \lambda^i \mathbf{c}^i$, where $\lambda^i \in [0,1]$, $\forall i \in [t]$, $\sum_{i=1}^{t} \lambda^i = 1$, $t \le n$ (see Algorithm 2 and Thm. 11). We also give a complete characterization of the underlying submodular function of any partition-respecting GSBP (Thm. 10).

3. **Polytime decomposition of face centers into extreme points (Sec. 5):** Our end goal being decomposing $\mathbf{x}$ in terms of extreme points of the GSBP, we further propose to decompose each face center $\mathbf{c}^i$ into convex combinations of the extreme points of its corresponding face. Towards this, we introduce notion of structural symmetries of GSBP (Def. 18), under which any of its face center can shown to be $O(\text{poly}(n)\, T(f))$ time decomposable (Thm. 19). We also analyze two such special symmetric structures: *circular and reflexive symmetry* (Def. 22 and 25) along with a



(a) **Input:** Any point $\mathbf{x}$ in GSBP $B(f,\mathbf{w}) \subset \mathbb{R}^n$, which satisfies two structural properties: 1. *Partition Respecting*, and 2. *Symmetric*. (Def 7 and 18)

(b) **Step 1.** Decompose $\mathbf{x} \in B(f,\mathbf{w})$ into convex combination of at most n face centers: $\mathbf{c}^1, \mathbf{c}^2, \ldots, \mathbf{c}^t \in \mathbb{R}^n$, such that $\mathbf{x} = \sum_{i=1}^{t} \lambda^i \mathbf{c}^i$ (Alg. 2, Sec. 4)

(c) **Step 2.** Further decompose each $\mathbf{c}^i$ into convex combination of extreme points of it face: $\mathbf{e}_1^i, \mathbf{e}_2^i, \ldots, \mathbf{e}_{\tau^i}^i$, such that $\mathbf{c}^i = \sum_{j=1}^{\tau^i} \widetilde{\lambda}^i \mathbf{e}_j^i$

(d) **Final Output.** $\mathbf{x}$ as convex combination of N extreme points of $B(f,\mathbf{w})$, say $E := \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\}$: $\mathbf{x} = \sum_{i=1}^{t} \lambda^i \Big(\sum_{j=1}^{\tau^i} \widetilde{\lambda}^i \mathbf{e}_j^i \Big) = \sum_{j=1}^{N} \gamma^j \mathbf{e}_j$.

Figure 1: Proposed algorithm of this work: GSBP Decomposition in a nutshell

complete characterization of the underlying submodular functions under them (Thm. 24, 27), and give an $O(n^2 T(f))$ decomposition (Thm. 23, 26).Fig. 1 summarizes the key steps of our proposed algorithm.

4. **Efficient Sampling:** Often times the application might demand the support of the decomposition of $\mathbf{x}$ to be 'small', precisely $O(\text{poly}(n))$ in order to efficiently sample an extreme point from the resulting convex combination (which induces a distribution) of $\mathbf{x}$ (Sec. 2.2). We show that our above algorithm is not only polytime executable, but it also decomposes $\mathbf{x}$ into only *poly*(n) number of extreme points aiding efficient sampling of the resulting extreme points (Thm. 23, 26).

In summary, we develop a fundamental understanding of the complexity of a class of polytope decomposition problem in terms of their structural properties.

Organization. We introduce the preliminaries and the formal problem statement in Sec. 2. We formally define *generalized submodular base polytopes* (GSBP) in Sec. 3. Our proposed algorithms are presented in Sec. 4 and 5 – Fig. 1 summarizes the key steps: Step-1 is described in Sec. 4 which decomposes the given point into the GSBP face centers, following which Sec. 5 details how to further decompose each face center into convex combination of extreme points of their respective faces. Sec. 7 concludes with some future directions. [2]

2. Preliminaries and Problem Statement

Notations. For any $n \in \mathbb{Z}_+$, let $[n] = \{1, ..., n\}$. For a set $S \subseteq [n]$, we denote the cardinality of S by $|S|$, the power set of S by 2^S, $S(i)$ denotes the i^{th} element when elements of (ordered)

2. All proofs are deferred to the appendix due to space constraints.

set S are sorted in the increasing order, and the set of all permutations of the elements of set S by $\Sigma_S = \{\sigma : S \to [|S|] \mid \sigma \text{ is a bijective mapping}\}$; for any $\sigma \in \Sigma_S$ and $i \in S$, $\sigma(i) \in [|S|]$ denotes the position of $i \in S$ under σ, and $\forall j \in [|S|]$, $\sigma^{-1}(j) \in S$ denotes the item in S ranked at position j. $\mathbf{1}_n$ denotes the n-dimensional all-ones vector.

Definition 1 (Ordered partition) *For $0 < d \le n$, an* ordered d-partition *of $[n]$ is a tuple $(S_1, \ldots, S_d)$, where $S_k \subseteq [n]$ $\forall k \in [d]$ with $S_k \neq \emptyset$ $\forall k$, $S_k \cap S_{k'} = \emptyset$ $\forall k \neq k'$, and $\cup_{k=1}^d S_k = [n]$.*

Definition 2 (Permutations respecting an ordered partition) *A permutation $\sigma \in \Sigma_{[n]}$* respects *an ordered d-partition $P = (S_1, \ldots, S_d)$ of $[n]$ if for all $1 \le k < k' \le d$, $i \in S_k, i' \in S_{k'} \implies \sigma(i) < \sigma(i')$. We denote the set of all permutations in $\Sigma_{[n]}$ respecting P as $\Sigma_{[n]}^P$. It is easy to note that, $|\Sigma_{[n]}^P| = \prod_{k=1}^d |S_k|!$.*

2.1. Problem Statement

Given a *generalized submodular base polytope* (GSBP) $B(f, \mathbf{w}) \subset \mathbb{R}^n$ (see Def. 3), and a fixed point $\mathbf{x} \in B(f, \mathbf{w})$, the problem is to find a $O(poly(n))$-time computable representation of $\mathbf{x}$ into convex combination of 'small number of' extreme points of $B(f, \mathbf{w})$. More precisely, if $E(B) := \{\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_N\}$ denotes the extreme points of GSBP $B(f, \mathbf{w})$, can we find a convex combination $\gamma^1, \gamma^2, \ldots, \gamma^N \in [0,1]$ with $\sum_{j=1}^N \gamma^j = 1$, $\mathbf{x} = \sum_{j=1}^N \gamma^j \mathbf{e}_j$? However, note that, in general $B(f, \mathbf{w})$ can have at most $n!$ many extreme points, i.e. $N = O(n^n)$, in which case computing the above representation is infeasible! Thus we also want the support of $\boldsymbol{\gamma} = (\gamma^1, \ldots, \gamma^N)$ to be small: precisely, $\|\boldsymbol{\gamma}\|_0 = O(poly(n))$.

Results in a nutshell. We propose a $O(n^2 T(f))$ time algorithm for the problem ($T(f)$: unit time required to evaluate f), for any GSBP with two "nice properties": 1 *Partition respecting* and 2. *Symmetry*. See Sec. 3 for all the definitions. The main building blocks of our algorithm is given in Fig. 1.

2.2. One Potential Application in Online learning with Combinatorial Decision spaces

We briefly describe a well studied problem in online learning where polytope decomposition algorithms are hugely relevant.

Setup. Consider a combinatorially large decision space $\mathcal{D}$, say set of 2^n subsets of a given set of size n, or maybe $n!$ possible number of permutations of n items, or even the set of spanning tress of a given graph with n vertices etc. Each decision point is associated to some unknown reward and the learner's goal to identify the 'optimal profit maximizing' decision in the hindsight by playing an online game: At each iteration, the learner chooses an decision point following which a noisy reward vector of the decision model is revealed to the learner based on which the learner chooses to update which point to play in the next round–now how to play this game efficiently when the number of choices in the decision space $\mathcal{D}$ is combinatorially large in n?

We here describe a general algorithm (high level idea) that is used to deal with this class of problems Suehiro et al. (2012); Rajkumar and Agarwal (2014), as described in Algorithm 1. It precisely maintains a point $\mathbf{x}^t$ in a suitably chosen convex polytope $\widehat{\mathcal{D}}$, and plays at

Algorithm 1 Online Learning from Combinatorial Decision Spaces: A generic approach with **Polytope Decomposition**

1: **input:** A suitable mapping $\phi : \mathcal{D} \mapsto \mathbb{R}^d$, d : positive integer ($\mathcal{D}$ being the combinatorial decision space)
2: **init:** $\widehat{\mathcal{D}} \leftarrow$ Convex-Hull($\mathcal{D}$) [convex polytope of interest], Note: Extreme points of $\widehat{\mathcal{D}}$ has 1-1 mapping to $\mathcal{D}$
3: Choose any point $\mathbf{x}^1 \in \widehat{\mathcal{D}}$
4: **for** $t = 1, 2, \ldots$ **do**
5:     **Decompose** $\mathbf{x}^t$ **into extreme points of** $\widehat{\mathcal{D}}$ **as:** $\mathbf{x}^t = \sum_{j \in \mathcal{D}} \gamma^j \phi(j)$**, where** $\sum_{j=1}^{N} \gamma^j = 1$**,** $\gamma^j \in [0, 1]$
6:     Play a decision $j \in \mathcal{D}$ with probability γ^j
7:     Receive environment feedback and update $\mathbf{x}^t \mapsto \mathbf{x}^{t+1}$
8: **end for**

each round decomposing $\mathbf{x}^t$ as convex combinations of its extreme points (Line 5)–where our *polytope decomposition* routine plays the pivotal role. Further the support of $\boldsymbol{\gamma} = (\gamma^1, \ldots, \gamma^N)$ being small, i.e. $\|\boldsymbol{\gamma}\|_0 = O(poly(n))$, Line 6 (sampling) can be executed efficiently which would have been otherwise impossible (Sec. 2.1).

3. Generalized Submodular Base Polytopes

In this section, we introduce the notion of generalized SBPs, which will be used throughout the rest of the paper.

Definition 3 (Generalized submodular base polytope (GSBP)) *Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function, with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$. We define the* generalized submodular base polytope (GSBP) $B(f, \mathbf{w})$ *as*

$$B(f,\mathbf{w}) = \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, \sum_{i \in S} w_i x_i \le f(S) \ \forall S \subset [n], \sum_{i \in [n]} w_i x_i = f([n]) \right\}$$

Clearly, $B(f, \mathbf{w})$ lies in an $(n-1)$-dimensional subspace of $\mathbb{R}^n$, and is defined by the intersection of $(2^n - 2)$ half-spaces (corresponding to the inequalities for $S \subset [n], S \neq \emptyset$) with this subspace. It is easy to verify that $B(f, \mathbf{w})$ is *convex* from its definition.

Definition 4 (Extreme points of GSBPs.) *The extreme points of a polytope are those which cannot be represented as a convex combination of any other points of the polytope Ziegler (1995). For $B(f, \mathbf{w})$, each permutation $\sigma \in \Sigma_{[n]}$ gives rise to an extreme point $\mathbf{e}^\sigma \in \mathbb{R}^n$ with coordinates:*

$$e_i^\sigma = \frac{1}{w_i}\Big(f\big(\{i' \in [n] \mid \sigma(i') \le \sigma(i)\}\big) - f\big(\{i' \in [n] \mid \sigma(i') < \sigma(i)\}\big) \Big) \ \forall i \in [n]. \qquad (1)$$

Any $n-1$-dimensional polytope has faces of dimension ranging from 0 to $n-1$; the extreme points are faces of dimension 0 (zero degrees of freedom for the points belonging to that face), edges are faces of dimension 1 (one degree of freedom), and the polytope itself is a face of dimension $n-1$ (clearly the coordinates of any arbitrary point belonging the polytope has $n-1$ degress of freedom, as could be verified from Def. 3). Now just as all extreme points of $B(f, \mathbf{w})$ (i.e. 0-dimensional faces) can be viewed as coordinates arising from permutations (where note that each permutation defines an n partition of the set $[n]$), for any $d \in [n]$, all $(n-d)$-dimensional faces can be viewed as arising from all possible ordered d-partitions. More formally any ordered d-partition $P = (S_1, \ldots, S_d)$ of $[n]$ gives rise to an $(n-d)$-dimensional face $\mathcal{F}_P \subset \mathbb{R}^n$ such that

$$\mathcal{F}_P = \left\{ \mathbf{x} \in B(f) \Big| \sum_{i \in \cup_{k'=1}^{k} S_{k'}} x_i = f\big(\cup_{k'=1}^{k} S_{k'}\big) \ \ \forall k \in [d-1] \right\}. \tag{2}$$

Note that, in general, it is possible for two or more ordered d-partitions to give rise to the same face. It is worth noting that, $\mathcal{F}_P$ being a face of a convex polytope $B(f, \mathbf{w})$, it is also convex, as follows from the above definition as well.

Definition 5 (Faces of GSBPs.) *For GSBP $B(f, \mathbf{w})$, each ordered d-partition $P = (S_1, \ldots, S_d)$ of $[n]$ gives rise to an $(n-d)$-dimensional face $\mathcal{F}_P \subset \mathbb{R}^n$:*

$$\mathcal{F}_P = \left\{ \mathbf{x} \in B(f, \mathbf{w}) \,\Bigg|\, \sum_{i \in \cup_{k'=1}^{k} S_{k'}} w_i x_i = f\big(\cup_{k'=1}^{k} S_{k'}\big) \ \ \forall k \in [d-1] \right\}. \tag{3}$$

Definition 6 (Face P-center of a GSBP.) *Given any ordered d-partition $P = (S_1, \ldots, S_d)$, the face P-center of the face $\mathcal{F}_P$ is defined as $\mathbf{c}^P = \dfrac{\sum_{\sigma \in \Sigma_{[n]}^P} \mathbf{e}^\sigma}{\prod_{k=1}^{d} |S_k|!}$.*

It is easy to see $c_i^P = \dfrac{\sum_{\sigma' \in \Sigma(\sigma, S_k)} e_i^{\sigma'}}{|S_k|!}$, for each coordinate $i \in S_k, \ k \in [d]$, and any $\sigma \in \Sigma_{[n]}^P$.

Remark 1 *The above class of GSBP subsumes a number of well studied polytopes, including k-subsets, spanning trees, permutahedron (more generally truncated permutahedons) Suehiro et al. (2012), k-forest, single machine scheduling polytope (SMSP) amongst many. Few specific examples are described in more details below, we also show empirical evaluations on fours GSBPs in experiments (Sec. 6).*

Example 1 (Permutahedron Rajkumar and Agarwal (2014)) *A well studied generalized submodular base polytope is the permutahedron, which is a GSBP $B(f, \mathbf{w})$ associated with the function $f : 2^{[n]} \to \mathbb{R}$ defined as $f(S) = \sum_{i=1}^{|S|}(n+1-i) \ \forall S \subseteq [n]$, with $\mathbf{w} = \mathbf{1}_n$. It is well known that permutahedron has $n!$ extreme points given by $\left\{ (\sigma(1), \ldots, \sigma(n))^\top \,|\, \sigma \in \Sigma_{[n]} \right\}$. It can be easily verified from Eqn. 1 that $e_i^\sigma = n+1-\sigma(i)$, and therefore each permutation σ yields a distinct extreme point. Just to understand a toy example, let $n = 4$ and consider the ordered 2-partition $P = (S_1, S_2)$ of $\{1, 2, 3, 4\}$, where $S_1 = \{3, 4\}$, $S_2 = \{1, 2\}$. Then from Eqn. 3 and 6 the face of $B(f)$ induced by P is given by $\mathcal{F}_P = \left\{ \mathbf{x} \in B(f) \,\big|\, \sum_{i \in \{3,4\}} x_i = f\big(\{3, 4\}\big) = 7 \right\}$, and the corresponding face P-center of $\mathcal{F}^P$ is $\mathbf{c}^P = (1.5, 1.5, 3.5, 3.5)^\top$.*

Example 2 (Single machine scheduling polytope (SMSP) Hoeksma et al. (2014)) *Given a vector $\mathbf{w} \in \mathbb{R}^n_+$, the single machine scheduling polytope (SMSP) can be viewed as an example of a GBSP $B(f, \mathbf{w})$, where the submodular function $f : 2^{[n]} \to \mathbb{R}$ is defined as:*

$$f(S) = (\sum_{i \in S} w_i)\left(\sum_{i \in [n] \setminus S} w_i - \frac{\sum_{i \in S} w_i}{2}\right) \ \ \forall S \subseteq [n].$$

Suppose $\mathbf{w} \in \mathbb{R}^n_+$ denotes the processing times of n jobs, and $\sigma \in \Sigma_{[n]}$ be a scheduling of the n jobs, then the half completion time of job i is defined as $h_i^\sigma = \sum_{i'=1}^{\sigma(i)-1} w_{\sigma^{-1}(i')} + \frac{w_i}{2}$. The SMSP $B(f, \mathbf{w})$ can be viewed as convex hull of all vectors representing the remaining time after half-completion of the set of n jobs, when scheduled on a single machine according to some scheduling $\sigma \in \Sigma_{[n]}$, without idle time, non-preemptively. Thus the set of extreme points of SMSP $B(f, \mathbf{w})$, computed using Eqn. 1, becomes $\{\mathbf{x} \in \mathbb{R}^n | x_i = \sum_{i \in [n]} w_i - h_i^\sigma \ \forall i \in [n], \sigma \in \Sigma_{[n]}\}$. Again consider a simple example $n = 3$: Here the set $B(f, \mathbf{w})$ is $\{\mathbf{x} \in \mathbb{R}^3 \,|\, x_i = \sum_{i \in [3]} w_i - h_i^\sigma \ \ \forall i \in [3], \sigma \in \Sigma_{[3]}\}$. Let $P = (S1, S2)$ be a partition of the set $\{1, 2, 3\}$, where $S1 = \{2, 3\}$ and $S2 = \{1\}$. Then from Eqn. 3, we get:

$$\mathcal{F}_P = \left\{\mathbf{x} \in B(f, \mathbf{w}) \Big| \sum_{i \in \{2,3\}} w_i x_i = (w_2 + w_3)\left(\frac{w_2}{2} + \frac{w_3}{2} + w_1\right)\right\},$$

and the corresponding face P-center of $\mathcal{F}_P$ turns out: $c_1^P = \frac{w_1}{2}$, and $c_2^P = c_3^P = w_1 + \frac{w_2 + w_3}{2}$.

4. Decomposing Points in Partition Respecting GSBPs into P-Centers

We now describe our main algorithm for decomposing any point $\mathbf{x} \in B(f, \mathbf{w})$ of a *partition-respecting* GSBP into a convex combination of its extreme points. Our first step is to decompose the given point into at most n of its face centers: $\mathbf{c}^1, \mathbf{c}^2, \ldots, \mathbf{c}^t$, $t \leq n$ (see **Step**-1 of Fig. 1). Following definitions will prove to be useful, before going into the details of the actual algorithm (see Algorithm 2).

4.1. Partition-Respecting GSBPs

Definition 7 (Partition-respecting GSBP) *Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and let $\mathbf{w} \in \mathbb{R}^n$. We define the GSBP $B(f, \mathbf{w})$ to be* partition-respecting *if for all $0 < d \leq n$, and all ordered d-partitions $P = (S_1, \ldots, S_d)$ of $[n]$, the P-center $\mathbf{c}^P$ satisfies the following:*

$$\forall k \in [d]: \qquad i, i' \in S_k \implies c_i^P = c_{i'}^P \tag{4}$$
$$\forall 0 < k < k' \leq [d]: \quad i \in S_k, i' \in S_{k'} \implies c_i^P \geq c_{i'}^P. \tag{5}$$

Theorem 8 (Face P-Center of Partition-respecting GSBP) *If GSBP $B(f, \mathbf{w})$ is* partition-respecting, *then it can be shown that for all ordered d-partitions $P = (S_1, \ldots, S_d)$, $0 < d \leq n$, for any $i \in S_k$, $k \in [d]$,*

$$c_i^P = \frac{f(\cup_{k'=1}^k S_{k'}) - f(\cup_{k'=1}^{k-1} S_{k'})}{\sum_{i \in S_k} w_i}. \tag{6}$$

Proposition 9 *Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and let $\mathbf{w} \in \mathbb{R}^n$. Then for all $\sigma \in \Sigma_{[n]}$, $e^{\sigma}_{\sigma^{-1}(1)} \ge e^{\sigma}_{\sigma^{-1}(2)} \ge \cdots \ge e^{\sigma}_{\sigma^{-1}(n)}$, if and only if for all $d \in [n]$, and each ordered d-partition $P = (S_1, \ldots, S_d)$ of $[n]$, $c_i^P \ge c_{i'}^P$, where $i \in S_k, i' \in S_{k'}$, $0 < k < k' \le [d]$.*

Example 3 (Permutahedron is Partition-Respecting) *Let $n = 4$ and $f : 2^{[4]} \to \mathbb{R}$ be the submodular function associated with the* Permutahedron *in R^4: $f(S) = \sum_{i=1}^{|S|}(5-i)\ \forall S$. Consider the ordered 2-partition $P = (S_1, S_2)$ of $\{1,2,3,4\}$, where $S_1 = \{3,4\}$, $S_2 = \{1,2\}$. Then the face of $B(f)$ induced by P is given by $\mathcal{F}_P = \Big\{\mathbf{x} \in B(f) \mid \sum_{i \in \{3,4\}} x_i = f\big(\{3,4\}\big) = 7\Big\}$. Also note that here $c_3^P = c_4^P = \frac{f(\{3,4\}) - f(\emptyset)}{\sum_{i \in S_1} w_i} = 3.5$, and $c_1^P = c_2^P = \frac{f([4]) - f(\{3,4\})}{\sum_{i \in S_2} w_i} = 1.5$, hence $c_3^P = c_4^P > c_1^P = c_2^P$. It can be verified that this is the case for all ordered partitions P, and the permutahedron is order-respecting.*

Theorem 10 (Characterization of partition-respecting GSBPs) *Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$. If the GSBP $B(f, \mathbf{w})$ is partition-respecting, then f satisfies the following conditions:*

$$\forall \{i,j\} \subseteq [n] : f(\{i,j\}) = \left(\frac{f(\{i\})}{(w_i - w_j)} + \frac{f(\{j\})}{(w_j - w_i)}\right)(w_i + w_j)$$

$$\forall \{i,j,k\} \subseteq [n] : f(\{i,j,k\}) = \Big(\sum_{i' \in \{i,j,k\}} \frac{f(\{i'\})}{\prod_{j' \in \{i,j,k\} \setminus \{i'\}} (w_{i'} - w_{j'})}\Big)(w_i + w_j + w_k)$$

$$\vdots$$

$$f([n]) = \left(\sum_{i'=1}^{n} \frac{f(\{i'\})}{\prod_{j' \in [n] \setminus \{i'\}} (w_{i'} - w_{j'})}\right)\Big(\sum_{i'=1}^{n} w_{i'}\Big).$$

4.2. Algorithm for Decomposition into P-Centers

We now describe our algorithm for decomposing any point $\mathbf{x}$ in a partition-respecting GSBP as a convex combination of at most n of its face P-centers. The details is given in Algorithm 2. Without loss of generality we assume $x_1 \ge x_2 \ge \cdots \ge x_n$ (one can always sort the coordinates of $\mathbf{x}$ in the decreasing order). The algorithm starts by setting $\mathbf{x}^1 = \mathbf{x}$, and proceeds iteratively: At any iteration $t < n$, if $\mathbf{x}^t$ hits one of the face P-center of $B(f, \mathbf{w})$, the algorithm terminates; otherwise $\mathbf{x}^t$ is partially represented by a suitably chosen face P-center $\mathbf{c}^t$ scaled by a constant $\widetilde{\lambda}^t$, and the remaining part of $\mathbf{x}^t$ is projected back to $B(f, \mathbf{w})$ setting $\mathbf{x}^{t+1} = \frac{\mathbf{x}^t - \widetilde{\lambda}^t \mathbf{c}^t}{1 - \widetilde{\lambda}^t}$. Theorem 11 shows that our proposed algorithm terminates in at most n iterations, as $\mathbf{x}^t$ eventually hits the center of the GSBP $B(f, \mathbf{w})$ itself, note this point is given by the vector $\big(\frac{f([n])}{\sum_{i \in [n]} w_i}\big)\mathbf{1}$ as per Thm. 8.

Theorem 11 (Correctness and running time of Algorithm 2) *Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$ such that the GSBP $B(f, \mathbf{w})$ is partition-respecting. Then Algorithm 2 decomposes any point $\mathbf{x} \in B(f, \mathbf{w})$ as a convex combination of at most n face P-centers of $B(f, \mathbf{w})$ in $O(n^2 + nT(f))$ time, where $T(f)$ denotes the unit time to evaluate the function f.*

Proof (sketch): Proof follows using the following key observations:

Algorithm 2 Decomposition of partition-respecting submodular base polytopes into at most n face P-centers

input: $\mathbf{x} \in B(f, \mathbf{w})$, such that $x_1 \geq x_2 \geq \cdots \geq x_n$, and $B(f, \mathbf{w})$ is partition-respecting

init: $t = 1$, $\mathbf{x}^1 = \mathbf{x}$, $\lambda^0 = 0$, $J^1 = \{i \in \{2, 3, \dots, n\} \mid x^1_{i-1} > x^1_i\}$

While $J^t \neq \emptyset$

– Compute the center $\mathbf{c}^t$ of the face associated to the partition induced by J^t:

$$c^t_i = \frac{f(\cup_{k'=1}^{k} S^t_{k'}) - f(\cup_{k'=1}^{k-1} S^t_{k'})}{\sum_{i \in S^t_k} w_i} \;\; \forall i \in S_k,\, k \in [|J^t| + 1],$$

where $S^t_k = \{J^t(k-1), J^t(k-1)+1, \dots, J^t(k)-1\}$, $J^t(0) = 1$, $J^t(|J^t|+1) = n+1$

– Set: $\widetilde{\lambda}^t = \min_{i \in J^t} \left\{ \frac{x^t_{i-1} - x^t_i}{c^t_{i-1} - c^t_i} \right\}$

– If $\widetilde{\lambda}^t = 1$ break

– Update:

$\mathbf{x}^{t+1} = \frac{\mathbf{x}^t - \widetilde{\lambda}^t \mathbf{c}^t}{1 - \widetilde{\lambda}^t}$, $J^{t+1} = \{i \in J^t \mid x^{t+1}_{i-1} > x^{t+1}_i\}$

$\lambda^t = \left(1 - \sum_{i=1}^{t-1} \lambda^i\right) \widetilde{\lambda}^t$, $t = t + 1$

End

$\mathbf{c}^t = \mathbf{x}^t$

$\lambda^t = 1 - \left(\sum_{j=1}^{t-1} \lambda^j\right)$

output: Face P-centers $\mathbf{c}^1, \dots, \mathbf{c}^t$, and $\lambda^1, \dots, \lambda^t \in [0, 1]$, $\sum_{i=1}^t \lambda^i = 1$, such that $\mathbf{x} = \sum_{i=1}^t \lambda^i \mathbf{c}^i$

Lemma 12 *At any iteration $i \in [t]$ 2, $\sum_{j=1}^n \mathbf{x}^i = f([n])$, and $\sum_{j \in T^i_k} \mathbf{x}^i_j < f(T^i_k)$ for all $k \in [|J^i| + 1]$.*

Lemma 13 *At any iteration $i \in [t]$ of Algorithm 2, $x^i_1 \geq x^i_2 \geq \cdots \geq x^i_n$, $0 \leq \widetilde{\lambda}^i \leq 1$.*

Lemma 14 *For any $i \in [t]$, $0 \leq \lambda^i \leq 1$, and $\sum_{i=1}^t \lambda^i = 1$.*

Lemma 15 *After t iterations, $\mathbf{x} = \sum_{i=1}^t \lambda^i \mathbf{c}^i$*

Lemma 16 *For any iteration $i \in [t-1]$, $|J^i| \geq |J^{i+1}| + 1$, hence $1 \leq t \leq n$.*

The proof of each of the above claims are given in in the supplementary. Lem. 28 and 29 lead to Lem. 30 and 31 which proves the correctness of Alg. 2. Lem. 32 shows the algorithm can run for at most n iterations, and within any iteration $i \in [t]$, all steps can be computed in $O(n)$ time except the computation of c^i which in general can take $T(f)$ time, thus making the total runtime complexity $O(n^2 + nT(f))$. Complete proof is given in the supplementary. ■

Remark 2 *Clearly, given an input point $\mathbf{x} \in B(f, \mathbf{w})$, Algorithm 2 decomposes it as convex combination of $t \leq n$ face center: $\mathbf{x} = \sum_{i=1}^t \lambda^i \mathbf{c}^i$. Note however our end goal is to decompose $\mathbf{x}$ in terms of the extreme points of the GSBP, but according to Defn. 6, any face P-center of a GSBP is itself a convex combination (more precisely average) of all its extreme points, i.e.*

$$\mathbf{c}^P = \frac{\sum_{\sigma \in \Sigma^P_{[n]}} \mathbf{e}^\sigma}{\prod_{k=1}^d |S_k|!} = \sum_{j=1}^{|\Sigma^P_{[n]}|} \widetilde{\lambda} \mathbf{e}_j, \tag{7}$$

where we denote by $\mathbf{e}_j$ *the* j^{th} *extreme point of the face corresponding to partition* P *and* $\widetilde{\lambda} = \dfrac{1}{\prod_{k=1}^{d}|S_k|!}$. *Note* $\sum_{j=1}^{|\Sigma_{[n]}^P|}\widetilde{\lambda} = 1$, *so it is a valid convex combination. Hence this itself yields the desired final decomposition of* $\mathbf{x}$ *as:* $\mathbf{x} = \sum_{i=1}^{t}\lambda^i\mathbf{c}^i = \sum_{i=1}^{t}\lambda^i\Big(\sum_j \widetilde{\lambda}^i\mathbf{e}_j^i\Big) = \sum_{k\in E(B)}\gamma_k\mathbf{e}_k$, *where* $\sum_{k\in E(B)}\gamma_k = 1$ *with* $\gamma^j \in [0,1], \forall j \in E(B)$, *which gives the desired decomposition, at least theoretically. However this might be computationally infeasible as the possible number of extreme points of a face corresponding to a d-partition* $P = (S_1,\ldots,S_d)$ *can be combinatorially many:* $|\Sigma_{[n]}^P| = \prod_{k=1}^{d}|S_k|!$, *so Eqn. 7 could hard to compute for large* n. *This motivates us to find a polytime decomposition of face* P*-centres into* $O(poly(n))$ *extreme points. Next section shows how that is actually possible for a class of GSBPs.*

5. Decomposing Face P-Centers of GSBP into few Extreme Points and Efficient Sampling

We now proceed to explain **Step**-2 of our decomposition routine, as given in Fig. 1. Recall our final objective is to represent $\mathbf{x}$ as a convex combination of small number of extreme points of $B(f,\mathbf{w})$. As argued in Rem. 2, our objective is to express each P-center as a convex combination of small number of extreme points of its associated face, for which we introduce the notion of structural symmetry of a GSBP under which any face center of GSBP can be efficiently decomposed in $O(\text{poly}(n)\,T(f))$-time, and analyze two of its special cases, namely *circular* (Def. 22) and *reflexive* symmetry (Def. 25), which along with the *partition-respecting* property can be shown to yield $O(n^2\,T(f))$ decomposition for any such GSBP.

Definition 17 (Shifts of a permutation restricted to a subset) *Given any permutation* $\sigma \in \Sigma_{[n]}$, *we define the shifts of the permutation* σ *restricted to a* $S \subseteq [n]$, *denoted by* $\Sigma(\sigma,S)$, *as the set of all permutations obtained by permuting the elements in* S, *keeping the rankings of the rest of the elements in* $[n]$ *fixed to that in* σ. *In other words,* $\Sigma(\sigma,S) = \{\sigma' \in \Sigma_{[n]} \mid \sigma'(i) = \sigma(i) \quad \forall i \in [n]\setminus S\}$. *Note that* $|\Sigma(\sigma,S)| = |S|!$.

Definition 18 (Symmetry in a GSBP) *Let* $f : 2^{[n]} \to \mathbb{R}$ *be a submodular function with* $f(\emptyset) = 0$, *and* $\mathbf{w} \in \mathbb{R}^n$. *We say the GSBP* $B(f,\mathbf{w})$ *is symmetric if for every* $d \in [n]$, *and all ordered d-partition* $P = (S_1,\ldots,S_d)$ *of* $[n]$, *for any* $\sigma \in \Sigma_{[n]}^P$, *and* $k \in [d]$,

$$c_i^P = \frac{\sum\limits_{\sigma' \in R(\sigma,S_k)} e_i^{\sigma'}}{|R(\sigma,S_k)|} \qquad \forall i \in S_k,$$

where $R(\sigma,S_k) \subset \Sigma(\sigma,S_k)$ denotes a subset of shifts of permutation σ restricted to set S_k s.t. $|R(\sigma,S_k)| = O(\text{poly}(|S_k|))$. In particular we call $B(f,\mathbf{w})$ to be *c-symmetric* if there exists a constant $c \geq 0$, such that $|R(\sigma,S_k)| = |S_k|^c, \forall k \in [d]$.

Theorem 19 (Decomposition of face P-centers of *c-symmetric GSBPs*) *Any face* P*-center of a* c-symmetric *GSBP* $B(f,\mathbf{w})$ *(for some constant* $c \geq 0$*)* $B(f,\mathbf{w})$ *can be decomposed into the extreme points of its corresponding face in* $O(n^{c+1}\,T(f))$ *time. Moreover, any such extreme point can be randomly sampled in* $O(n^c\,T(f))$ *time.*

Proof (sketch): By Defn. 18, for every $0 < d \leq n$, and all ordered d-partition $P = (S_1, \dots, S_d)$ of $[n]$, for any $\sigma \in \Sigma_{[n]}^P$, and $k \in [d]$, $c_i^P = \frac{\sum_{\sigma' \in R(\sigma, S_k)} e_i^{\sigma'}}{|R(\sigma, S_k)|} \quad \forall i \in S_k$, with $|R(\sigma, S_k)| = |S_k|^c \; \forall k \in [d]$. Hence for all $i \in S_k$, the i^{th} coordinate of c^P can be represented as an average of the i^{th} coordinates of at most $|S_k|^c$ extreme points of the $B(f, \mathbf{w})$, whose underlying $O(|S_k|^c)$ permutations can be computed in $O(|S_k|^c)$ time, and using (1) all n coordinates of such extreme points can be evaluated in $O(n|S_k|^c T(f))$ time. The first claim now follows since for all $k \in [d]$, $|S_k| \leq n$ (as $\sum_{k=1}^d |S_k| = n$).

It is worth noting that if we are interested in randomly sampling only one such extreme point, for all $i \in S_k$, we can compute the feasible set of permutations in $O(|S_k|^c)$ time, and randomly sample one out of them. For this particular permutation, we can compute all the i^{th} coordinates of its corresponding extreme point in $|S_k| \, T(f)$ time. Repeating this for each $k \in [d]$ the desired extreme point can be computed in $O(n \, T(f))$ time. Thus the total sampling time complexity becomes $O(n^c \, T(f))$. Complete proof is given in the supplementary. ■

Theorem 20 (Decomposition of *symmetric*, partition-respecting GSBP) *Any point of a* c-symmetric*, partition respecting GSBP $B(f, \mathbf{w})$ (for some integer constant $c \geq 0$), can be decomposed into a convex combination of its extreme points in $O(n^2 + n \, T(f) + n^{c+2} \, T(f))$ time. Moreover, any such extreme point can be sampled in $O(n^2 + n \, T(f) + n^{c+1} \, T(f))$ time.*

Next we discuss two special symmetry structures of GSBPs: *Circular* and *Reflexive* symmetry.

5.1. Circular Symmetry

Definition 21 (Circular shifts of a permutation restricted to a subset) *Consider a subset S of the set $[n]$. Given any permutation $\sigma \in \Sigma_{[n]}$, we define the circular shifts of the permutation σ restricted to the subset S, denoted by $C(\sigma, S)$, as the set of all permutations obtained by applying circular shifts on the rankings of the elements in S, keeping the rankings of the rest of the elements of $[n]$ fixed to that in σ. In other words, $C(\sigma, S) = \{\sigma_1, \sigma_2, \cdots, \sigma_{|S|}\} \subset \Sigma_{[n]}$, where $\sigma_k(S(i)) = \sigma(S((i + k - 1) \mod |S|)) \;\; \forall i \in [|S|]$, and $\sigma_k(j) = \sigma(j) \;\; \forall j \in [n] \setminus S, \forall k \in [|S|]$. Note that, here $\sigma_1 = \sigma$, $|C(\sigma, S)| = |S|$, and $C(\sigma, S) \subset \Sigma(\sigma, S)$.*

Definition 22 (Circular symmetry) *Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$. We define the GSBP $B(f, \mathbf{w})$ to be* circular symmetric *if for every $0 < d \leq n$, and all ordered d-partition $P = (S_1, \dots, S_d)$ of $[n]$, for any $\sigma \in \Sigma_{[n]}^P$, and $k \in [d]$,*

$$c_i^P = \frac{\sum_{\sigma' \in C(\sigma, S_k)} e_i^{\sigma'}}{|S_k|} \qquad \forall i \in S_k. \tag{8}$$

Remark. Any circular symmetric GSBP is 1-symmetric.

Theorem 23 (Decomposition of circular symmetric partition-respecting GSBP) *Any point of a circular symmetric, partition-respecting GSBP $B(f, \mathbf{w})$ can be decomposed into a convex combination of its extreme points in $O(n^3 \, T(f))$ time. Moreover any such extreme point can be randomly sampled in $O(n^2 \, T(f))$ time.*

Example 4 *Permutahedron (Example 1) is a circular symmetric and partition respecting GSBP, and thus could be efficiently decomposed in $O(n^3T(f))$ time as per Thm. 23, and any such extreme point can be sampled in $O(n^2T(f))$ time. Note $T(f) = O(1)$ for the underlying f for Permutahedrons. In fact any cardinality based GSBP (i.e. whose underlying $f(S)$ is a function of $|S|$) can be shown to have* Circular Symmetry *property, for example the k-Subset Polytope discussed in our experiments (Sec. 6), and respect $O(n^3T(f))$ decomposition.*

Theorem 24 (Characterization of partition-respecting, circular symmetric GSBPs) *If the GSBP $B(f, \mathbf{w})$ is partition-respecting and circular symmetric then f satisfies all the constraints of Thm. 10, and one additional property: $\forall \{i, j, k\} \subseteq [n] : f(\{i\})(w_j - w_k) + f(\{j\})(w_k - w_i)$, where the last property actually induces the circular-symmetry in $B(f, \mathbf{w})$.*

5.2. Reflexive Symmetry

Definition 25 (Reflexive symmetry) *Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$. We define the GSBP $B(f, \mathbf{w})$ to be* reflexive symmetric *if for all ordered d-partition $P = (S_1, \ldots, S_d)$ of $[n]$, $0 < d \leq n$, for every $\sigma \in \Sigma_{[n]}^P$ and $k \in [d]$,*

$$c_i^P = \frac{e_i^\sigma + e_i^{\sigma_R^{S_k}}}{2}, \;\; \forall i \in S_k, \tag{9}$$

where $\sigma_R^{S_k}$ denotes the reverse permutation of σ restricted to set S_k, i.e., $\sigma_R^{S_k}(i) := n_{k-1} + (n_k - \sigma(i) + 1) \quad \forall i \in S_k$, and $\sigma_R^{S_k}(i) := \sigma(i) \quad \forall i \in [n] \setminus S_k$, $n_k = |\cup_{k'=1}^{k} S_{k'}| \;\; \forall k \in [d]$ [3]*. Note, any reflexive symmetric GSBP is 0-symmetric.*

Theorem 26 (Decomposition of reflexive symmetric, partition-respecting GSBP) *Any point of a reflexive symmetric, partition-respecting GSBP $B(f, \mathbf{w})$, can be decomposed into a convex combination of its extreme points in $O(n^2\, T(f))$ time. Moreover any such extreme point can be randomly sampled in $O(n\, T(f))$ time.*

Example 5 (Efficient decomposition for single machine scheduling polytope) *Our algorithm yields $O(n^2T(f))$ decomposition for SMSPs (see Example 2) Hoeksma et al. (2014). Note in this case $T(f) = O(n)$ (this could be $O(1)$ at a larger space complexity if $\sum_{i\in S} w_i$ are precomputed for all S), and for every $d \in [n]$, and all ordered d-partition $P = (S_1, \ldots, S_d)$ of $[n]$, $c_i^P = \frac{e_i^\sigma + e_i^{\sigma_R^{S_k}}}{2} = \big(\sum_{i\in[n]} w_i\big) - \big(\sum_{j \in \cup_{k' \leq k-1} S'_k} w_j + \sum_{j\in S_k} \frac{w_j}{2}\big) \;\; \forall i \in S_k, \sigma \in \Sigma_{[n]}^P, k \in [d]$. Thus our algorithm respects efficient decomposition of SMSPs.*

Theorem 27 (Characterization of partition-respecting, reflexive symmetric GSBPs) *If the GSBP $B(f, \mathbf{w})$ is partition-respecting and reflexive symmetric then f satisfies all the constraints of Thm. 10 and an additional $\forall \{i, j, k\} \subseteq [n] : \frac{f(\{i\})}{w_i}(w_j - w_k) + \frac{f(\{j\})}{w_j}(w_k - w_i)$, where the last property actually induces the reflexive-symmetry in $B(f, \mathbf{w})$.*

3. For example if $n = 6$, $P = (S_1, S_2)$ with $S_1 = \{1, 2, 3\}$, $S_2 = \{4, 5, 6\}$. Then for $\sigma = (1, 2, 3, 4, 5, 6)$, we have $\sigma_R^{S_1} = (3, 2, 1)$ and $\sigma_R^{S_2} = (6, 5, 4)$.

6. Experiments

In this section, we compare the empirical performances of our decomposition algorithm (Alg. 2) over different GSBPs. The details of the experiment setups are given below:

Constructing GSBPs $B(f, \mathbf{w})$. We use the following four different generalized submodular base polytopes for our experiments:

1. Permutahedron: As described in Example 1.

2. SMSP: As described in Example 2.

3. k-Subset: The problem of k-subsets is a generalization of experts problem, where each combinatorial structure corresponds to a set of k experts among n expert, each k-set being represented by a sum of k-corresponding canonical basis vectors. The associated submodular modular function of this polytope can be represented as: $f : 2^{[n]} \mapsto \mathbb{R}$ such that $f(S) = \begin{cases} |S| & \text{if } |S| \le k \\ k & \text{otherwise} \end{cases}$, for any subset $S \subseteq [n]$.

4. Spanning-Tree: We consider undirected spanning trees $G = (V, E)$. In this case the extreme points of the GSBP is represented by an edge encoded format: $\{\mathbf{x} \in \{0,1\}^{|E|} \mid$ the edges reperesented by $\{x(i) = 1\}$ forms a spanning tree of G$\}$. The associated submodular function in this case is given by $f : 2^{|E|} \mapsto \mathbb{R}$ such that $f(S) = |V(S)| - c(S)$, with $V(S)$ being the vertex set of the subgraph induced by the edges in S, and $c(S)$ being the number of the connected components in the induced subgraph. This polyhedron is known as spanning tree polyhedron Suehiro et al. (2012).

We choose $k = \frac{n}{2}$ for the k-subset polytope, and generate a random graph with n vertices for the spanning tree experiments. All the reported results are averaged across 50 runs (i.e. over 50 different initial points $\mathbf{x}$).

6.1. Decomposition of different GSBPs with dimension (n)

We first run experiments to analyze the yielded number of decomposed points of our proposed algorithm over different GSBPs. The first two figures of Fig. 2 (top left and bottom left) respectively show the number of decomposed face-centres (Sec. 4) and extreme-points (Sec. 5) on *Permutahedron* with dimensionality $n = 10, 40, 80, 150, 200$.

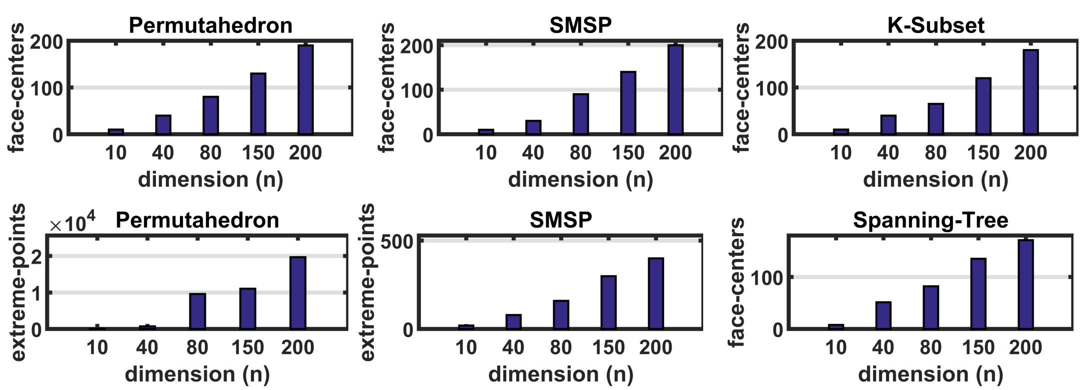


Figure 2: No of decomposition points with increasing dimensionality (n) for four GSBPs

Note that for each n the algorithm yields roughly $O(n)$ decomposed face centres on an average. The next two figures (in the middle) of Fig. 2 shows the same plots for *SMSP* polytopes. In this case, while the pattern for the number of decomposed face centres remains same as Permutahedrons, note the final number of decomposed extreme points for SMSPs are much less than that of Permutahedrons—the reason being the later being *circular symmetric* (Def. 22) each face centre can be decomposed up to $O(n)$ extreme points, whereas the former having *reflexsive symmetry* (Def. 25) here any face-center can be decomposed in to just 2 extreme points and consequently resulting into much smaller number of extreme points. The last two figures (top and bottom of extreme right) again show the number of decomposed face centres with increasing dimension for *k-subset* and *spanning-tree* polytopes, where again it shows the number of decomposed face-centres are $O(n)$ on an average.

6.2. Runtime comparison across different GSBPs

We also empirically evaluate the execution runtime (in seconds) of Alg. 2 for different GSBPs. As expected, and justified theoretically in Thm. 10, Fig. 3 shows that the average decomposition time varies roughly as $O(n^2)$: For example, the runtime for $n = 100$ is about 4 times that for $n = 50$; similarly that for $n = 150$ and $n = 200$ is respectively are about ~ 2 and ~ 4 times higher than the corresponding average runtime at $n = 100$.

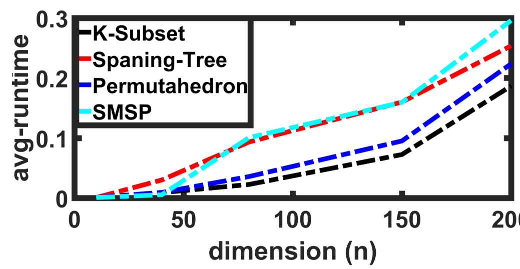


Figure 3: Runtime (in seconds) comparison across four different GSBPs

7. Conclusion and Future Works

We consider the problem of decomposing a given point of a convex polytope into a convex combination of its extreme points for a new class of *generalized submodular base polytopes* (GSBPs), that subsumes both SBPs and SMSPs as its special case. We first give a $O(n^2 + nT(f))$ polytime algorithm for decomposing any point of a GSBP into a small number of its face centers under a *partition-respecting* property. Moreover, we define and characterize the notions of *symmetry* associated with a GSBP, under which any face center GSBP is shown to be $O\Big(poly(n)T(f)\Big)$ efficiently decomposible into convex combination of its extreme points. Our algorithm also is shown to perform in $O(n^2T(f))$ time under special structures of *circular and reflexive symmetry*. Our work unifies several previous polytope decompostion algorithms including Suehiro et al. (2012) and Hoeksma et al. (2014), and develop a fundamental understanding of the problem complexity in terms of structural properties of the polytopes.

Future Works. It would be interesting to analyze the regret guarantees of online combinatorial optimization problems with our decomposition routine as blackbox, e.g. for Rajkumar and Agarwal (2014) etc. Understanding the problem complexity for some special polytopes, e.g. associahedron or zonotopes Doker (2011), or may be in the absence of the *symmetry partition respecting* structures, and modelling more real world applications with submodularity constraints to find practicability of our results are other interesting directions.

References

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2013.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2014.

Sébastien Bubeck. Introduction to online optimization. *Lecture Notes*, pages 1–86, 2011.

Yiling Chen, Lance Fortnow, Nicolas Lambert, David M Pennock, and Jennifer Wortman. Complexity of combinatorial market makers. In *Proceedings of the 9th ACM conference on Electronic commerce*, pages 190–199. ACM, 2008.

Hanjun Dai, Elias B Khalil, Yuyu Zhang, Bistra Dilkina, and Le Song. Learning combinatorial optimization algorithms over graphs. *arXiv preprint arXiv:1704.01665*, 2017.

Jeffrey Samuel Doker. *Geometry of Generalized Permutohedra*. PhD thesis, Citeseer, 2011.

Günter Ewald. *Combinatorial convexity and algebraic geometry*, volume 168. Springer Science & Business Media, 2012.

Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

Takahiro Fujita, Kohei Hatano, and Eiji Takimoto. Combinatorial online prediction via metarounding. In *Algorithmic Learning Theory*, pages 68–82. Springer, 2013.

Takahiro Fujita, Kohei Hatano, Shuji Kijima, and Eiji Takimoto. Online linear optimization over permutations with precedence constraints. 2014.

Shuhong Gao and Alan GB Lauder. Decomposition of polytopes and polynomials. *Discrete & Computational Geometry*, 26(1):89–104, 2001.

Aditya Gopalan, Shie Mannor, and Yishay Mansour. Thompson sampling for complex online problems. In *International Conference on Machine Learning*, pages 100–108, 2014.

Elad Hazan and Satyen Kale. Beyond convexity: Online submodular minimization. In *Advances in Neural Information Processing Systems*, pages 700–708, 2009.

Elad Hazan, Tomer Koren, Roi Livni, and Yishay Mansour. Online learning with low rank experts. In *Conference on Learning Theory*, pages 1096–1114, 2016a.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016b.

David P. Helmbold and Manfred K. Warmuth. Learning permutations with exponential weights. *Journal of Machine Learning Research*, 10:1705–1736, 2009.

Ruben Hoeksma, Bodo Manthey, and Marc Uetz. Decomposition algorithm for the single machine scheduling polytope. In *Combinatorial Optimization*, pages 280–291. Springer, 2014.

Satyen Kale. *Efficient algorithms using the multiplicative weights update method*. Princeton University, 2007.

Wouter M. Koolen, Manfred K. Warmuth, and Jyrki Kivinen. Hedging structured concepts. In *COLT*, 2010.

Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.

Gerard Meurant. *Handbook of convex geometry*. Elsevier, 2014.

Gergely Neu and Michal Valko. Online combinatorial optimization with stochastic decision sets and adversarial losses. In *Advances in Neural Information Processing Systems*, pages 2780–2788, 2014.

Holakou Rahmanian, David P Helmbold, and SVN Vishwanathan. Online learning of combinatorial objects via extended formulation. *arXiv preprint arXiv:1609.05374*, 2016.

Arun Rajkumar and Shivani Agarwal. Online decision-making in general combinatorial spaces. In *NIPS*, pages 3482–3490, 2014.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Random averages, combinatorial parameters, and learnability. In *Advances in Neural Information Processing Systems*, pages 1984–1992, 2010.

Daiki Suehiro, Kohei Hatano, Shuji Kijima, Eiji Takimoto, and Kiyohito Nagano. Online prediction under submodular constraints. In *ALT*, 2012.

Shota Yasutake, Kohei Hatano, Shuji Kijima, Eiji Takimoto, and Masayuki Takeda. Online linear optimization over permutations. In *ISAAC*, pages 534–543, 2011.

Shota Yasutake, Kohei Hatano, Eiji Takimoto, and Masayuki Takeda. Online rank aggregation. In *ACML*, 2012.

Günter M. Ziegler. *Lectures on Polytopes*. Springer, 1995.

Supplementary: Polytime Decomposition of Generalized Submodular Base Polytopes with Efficient Sampling

Appendix A. Supplement to Section 4.1

Theorem 8 [Face P-Center of Partition-respecting GSBP]. Let $f : 2^{[n]} \rightarrow \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and let $\mathbf{w} \in \mathbb{R}^n$ such that the GSBP $B(f, \mathbf{w})$ is *partition-respecting*. Then for all ordered d-partitions $P = (S_1, \ldots, S_d)$, $0 < d \leq n$, for any $i \in S_k$, $k \in [d]$,

$$c_i^P = \frac{f(\cup_{k'=1}^{k} S_{k'}) - f(\cup_{k'=1}^{k-1} S_{k'})}{\sum_{i \in S_k} w_i}.$$

Proof Recall from Eqn. 3,

$$\mathcal{F}_P \ = \ \left\{ \mathbf{x} \in B(f, \mathbf{w}) \ \middle| \ \sum_{i \in \cup_{k'=1}^{k} S_{k'}} w_i x_i = f\big(\cup_{k'=1}^{k} S_{k'} \big) \ \ \forall k \in [d-1] \right\}.$$

Hence for any $\mathbf{x} \in \mathcal{F}_P$, $\sum_{i \in \cup_{k'=1}^{k} S_{k'}} x_i = f(\cup_{k'=1}^{k} S_{k'})$, or equivalently $\sum_{i \in S_k} x_i = f(\cup_{k'=1}^{k} S_{k'}) - f(\cup_{k'=1}^{k-1} S_{k'})$, and in particular $\sum_{i \in S_k} c_i^P = f(\cup_{k'=1}^{k} S_{k'}) - f(\cup_{k'=1}^{k-1} S_{k'})$. Now as the GSBP $B(f, \mathbf{w})$ is partition respecting GSBP, from Eqn. 4 we get for all $i \in S_k$,

$$c_i^P = \frac{f(\cup_{k'=1}^{k} S_{k'}) - f(\cup_{k'=1}^{k-1} S_{k'})}{\sum_{i \in S_k} w_i}.$$

■

Theorem 9. Let $f : 2^{[n]} \rightarrow \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and let $\mathbf{w} \in \mathbb{R}^n$. Then for all $\sigma \in \Sigma_{[n]}$, $e^{\sigma}_{\sigma^{-1}(1)} \geq e^{\sigma}_{\sigma^{-1}(2)} \geq \cdots \geq e^{\sigma}_{\sigma^{-1}(n)}$, if and only if for all $0 < d \leq n$ and each ordered d-partition $P = (S_1, \ldots, S_d)$ of $[n]$, $c_i^P \geq c_{i'}^P$, where $i \in S_k, i' \in S_{k'}$, $0 < k < k' \leq [d]$.

Proof First let us consider that for all $\sigma \in \Sigma_{[n]}$, $e^{\sigma}_{\sigma^{-1}(1)} \geq e^{\sigma}_{\sigma^{-1}(2)} \geq \cdots \geq e^{\sigma}_{\sigma^{-1}(n)}$, i.e. coordinates of all extreme points of the GSBP are sorted in the decreasing order of their ranking corresponding to the extreme point. Now recall from the definition of P-center of any a face $\mathcal{F}_P$ associated to the d-ordered partition $P = (S_1, \ldots, S_d)$, $0 < d \leq n$, the i^{th} coordinate of the face center c^P is given by

$$c_i^P \ = \ \frac{1}{|S_k|!} \sum_{\sigma' \in \Sigma(\sigma, S_k)} e_i^{\sigma'} \ \ \forall i \in S_k, \ k \in [d], \ \text{for any } \sigma \in \Sigma_{[n]}.$$

Now suppose n_k denotes the total number of elements in the set $\cup_{k'=1}^{k} S_{k'}$, $\forall k \in [d]$, i.e. $n_k = | \cup_{k'=1}^{k} S_{k'}|$, and let $n_0 = 0$. Note that for any element $i \in S_k$, for all $\sigma \in \Sigma_{[n]}^P$, $\sigma(i) \in \{(n_{k-1} + 1), (n_{k-1} + 2), \cdots, n_k\}$, and $\forall i \in \cup_{k'=1}^{k} S_{k'}, \sigma(i) \leq n_k$. Then equivalently coordinates of C^P can be defined as

$$c^P_{\sigma^{-1}(i)} \ = \ \frac{1}{|S_k|!} \sum_{\sigma' \in \Sigma(\sigma, S_k)} e^{\sigma'}_{\sigma'^{-1}(i)} \ \ \forall \sigma^{-1}(i) \in S_k,$$

or equivalently,

$$c^P_{\sigma^{-1}(i)} \;=\; \frac{1}{|S_k|!}\sum_{\sigma'\in\Sigma(\sigma,S_k)} e^{\sigma'}_{\sigma'^{-1}(i)} \quad \forall i \in \{(n_{k-1}+1),(n_{k-1}+2),\cdots,n_k\}.$$

Since we have assumed that for all $\sigma\in\Sigma_{[n]}$, $e^\sigma_{\sigma^{-1}(1)} \ge e^\sigma_{\sigma^{-1}(2)} \ge \cdots \ge e^\sigma_{\sigma^{-1}(n)}$, the above result implies that if $i\in\{(n_{k-1}+1),(n_{k-1}+2),\cdots,n_k\}$, and $i'\in\{(n_{k'-1}+1),(n_{k'-1}+2),\cdots,n_{k'}\}$, then $c^P_{\sigma^{-1}(i)} \ge c^P_{\sigma^{-1}(i')}$ whenever $0<k<k'\le[d]$. Note that whenever $i\in\{(n_{k-1}+1),(n_{k-1}+2),\cdots,n_k\}$, $\sigma^{-1}(i)\in S_k$, thus we get $c^P_i \ge c^P_{i'}$, where $i\in S_k, i'\in S_{k'}$, $0<k<k'\le[d]$.

Next let us assume that for all $0<d\le n$ and each ordered d-partition $P=(S_1,\ldots,S_d)$ of $[n]$, $c^P_i\ge c^P_{i'}$, where $i\in S_k, i'\in S_{k'}$, $0<k<k'\le[d]$. Clearly when $d=n$, each d-ordered partition essentially boils down to a single permutation, say $\sigma\in\Sigma_{[n]}$, and gives rise to the extreme point e^σ (which can also be seen as a 0-dimensional face) of the GSBP. Clearly in this case, $|S_k|=1\ \forall k\in[n]$, more specifically $S_k=\{\sigma^{-1}(k)\}$. Hence we get, for all $\sigma\in\Sigma_{[n]}$, $e^\sigma_{\sigma^{-1}(1)} \ge e^\sigma_{\sigma^{-1}(2)} \ge \cdots \ge e^\sigma_{\sigma^{-1}(n)}$. ■

Theorem 10 [Characterization of partition-respecting GSBPs]. Let $f: 2^{[n]}\to\mathbb{R}$ be a submodular function with $f(\emptyset)=0$, and $\mathbf{w}\in\mathbb{R}^n$. If the GSBP $B(f,\mathbf{w})$ is partition-respecting, then f satisfies the following conditions:

$$\begin{aligned}
\forall\{i,j\}\subseteq[n]:\ f(\{i,j\}) &= \left(\frac{f(\{i\})}{(w_i-w_j)}+\frac{f(\{j\})}{(w_j-w_i)}\right)(w_i+w_j)\\
\forall\{i,j,k\}\subseteq[n]:\ f(\{i,j,k\}) &= \left(\sum_{i'\in\{i,j,k\}}\frac{f(\{i'\})}{\prod_{j'\in\{i,j,k\}\setminus\{i'\}}(w_{i'}-w_{j'})}\right)(w_i+w_j+w_k)\\
&\ \vdots\\
f([n]) &= \left(\sum_{i'=1}^{n}\frac{f(\{i'\})}{\prod_{j'\in[n]\setminus\{i'\}}(w_{i'}-w_{j'})}\right)\Big(\sum_{i'=1}^{n}w_{i'}\Big).
\end{aligned}$$

Proof Assume that $B(f,\mathbf{w})$ is partition-respecting, and consider any d-ordered partition $P=(S_1,\cdots,S_d)$, $0<d\le n$. Recall the definition of *partition-respecting* GSBP from section 4.1, we will first analyze the first condition of Definition 7 given by Eqn. 4. Clearly when $d=n$, or $|S_k|=1\ \forall k\in[d]$, Eqn. 4 trivially holds. Now let $|S_1|=2$, in general we can assume $S_1=\{i,j\}\subseteq[n]$. Now since $B(f,\mathbf{w})$ is partition-respecting, from Eqn. 6, we get $c^P_1=\frac{f(S_1)-f(\emptyset)}{|S_1|}\ \forall k\in S$, thus using Eqn. 1 and the partition respecting property,

$$\begin{aligned}
&c^P_i = \frac{f(\{i\})-f(\emptyset)+f(\{i,j\})-f(\{j\})}{2w_i} = \frac{f(\{i,j\})-f(\emptyset)}{(w_i+w_j)}\\
&\Longrightarrow f(\{i\})-f(\{j\}) = 2w_i\frac{f(\{i,j\})}{(w_i+w_j)} - f(\{i,j\})\\
&\Longrightarrow \frac{f(\{i,j\})}{(w_i+w_j)} = \left(\frac{f(\{i\})}{(w_i-w_j)}+\frac{f(\{j\})}{(w_j-w_i)}\right).
\end{aligned}\tag{10}$$

Now instead consider $|S_1| = 3$, such that $S_1 = \{i,j,k\} \subseteq [n]$. Again since $B(f,\mathbf{w})$ is partition-respecting, as before using Eqn. 1 and partition respecting property,

$$\begin{aligned}
c_i^P &= \frac{2f(\{i\}) + (f(\{i,j\}) - f(\{j\})) + (f(\{i,k\}) - f(\{k\})) + 2(f(\{i,j,k\}) - f(\{j,k\}))}{6w_i} \\
&= \frac{f(\{i,j,k\})}{(w_i + w_j + w_k)} \\
\Longrightarrow & 2f(\{i\}) - f(\{j\}) - f(\{k\}) + f(\{i,j\} + f(\{i,k\}) - 2f(\{j,k\}) = \frac{f(\{i,j,k\})6w_i}{(w_i + w_j + w_k)} - 2f(\{i,j,k\})
\end{aligned}$$

Applying Eqn. 10 in the above expression one can get,

$$f(\{i,j,k\}) = \left(\sum_{i' \in \{i,j,k\}} \frac{f(\{i'\})}{\prod_{j' \in \{i,j,k\} \setminus \{i'\}} (w_{i'} - w_{j'})} \right) (w_i + w_j + w_k)$$

Similarly considering different cardinalities of S_1, one can derive the expressions of f on any set $S \subseteq [n]$, and finally considering $S_1 = [n]$, and combining this with Eqn. 10 we get $f([n]) = \left(\sum_{i'=1}^n \frac{f(\{i'\})}{\prod_{j' \in [n] \setminus \{i'\}} (w_{i'} - w_{j'})} \right) \left(\sum_{i'=1}^n w_{i'} \right)$. Note that instead of S_1, we could have worked with any other partition set S_k $k \in [d]$, $0 < d \leq 1$ which essentially would given same the expression of $f(\cup_{k'=1}^k S_{k'})$ that we get considering $S_1 = \cup_{k'=1}^k S_{k'}$. ■

Appendix B. Supplement to Section 4.2

Theorem 11 [Correctness and running time of Algorithm 2]. Let $f : 2^{[n]} \rightarrow \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$ such that the GSBP $B(f,\mathbf{w})$ is partition-respecting. Then Algorithm 2 decomposes any point $\mathbf{x} \in B(f,\mathbf{w})$ as a convex combination of at most n face P-centers of $B(f,\mathbf{w})$ in $O(n^2 + nT(f))$ time, where $T(f)$ denotes the unit time to evaluate the function f.

Proof We proof the result for SBPs, i.e. when $\mathbf{w} = \mathbf{1}$. Similar argument follows for GSBPs as well. First let us assume that for all iteration $i \in [t]$, $T_k^i = \cup_{k'=1}^k S_{k'}^i$ $\forall k \in [|J^i| + 1]$. The above theorem can be proved using the following lemmas:

Lemma 28 *At any iteration $i \in [t]$ of Algorithm 2, $\sum_{j=1}^n \mathbf{x}^i = f([n])$, and $\sum_{j \in T_k^i} \mathbf{x}_j^i < f(T_k^i)$ for all $k \in [|J^i| + 1]$.*

Proof We give an inductive proof of the above claim. Recall from Definition of SBP if $\mathbf{x} \in B(f)$, then $\sum_{i \in S} x_i \leq f(S)$ $\forall S \subset [n]$, and $\sum_{i \in [n]} x_i = f([n])$. Clearly, for $i = 1$, $\mathbf{x}^1 = \mathbf{x} \in B(f)$, hence the claim follows.

Next let us assume that at any iteration $i \in [t-1]$, $\sum_{j=1}^n \mathbf{x}^i = f([n])$. Then $\sum_{j=1}^n \mathbf{x}_j^{i+1} = \frac{\sum_{j=1}^n \mathbf{x}_j^i - \widetilde{\lambda}^i \sum_{j=1}^n \mathbf{c}_j^i}{1 - \widetilde{\lambda}^i} = f([n])$, as $c^i \in B(f)$, and hence $\sum_{j=1}^n \mathbf{c}_j^i = f([n])$.

Now assume that at any iteration $i \in [t-1]$, $\sum_{j\in T_k^i} \mathbf{x}_j^i < f(T_k^i)$, for all $k \in [|J^i|+1]$. Then $\sum_{j\in T_k^i} \mathbf{x}_j^{i+1} = \frac{\sum_{j\in T_k^i} \mathbf{x}_j^i - \widetilde{\lambda}^i \sum_{j\in T_k^i} \mathbf{c}_j^i}{1-\widetilde{\lambda}^i} < f(T_k^i)$, as definition $\sum_{j\in T_k^i} \mathbf{c}_j^i = f(T_k^i)$. Now it is easy to check from line 14 of Algorithm 2, J^{i+1} is a subset of J^i, and hence the partition defined by J^i must be a refinement of that defined by J^{i+1}. Thus we can claim that $\sum_{j\in T_k^{i+1}} \mathbf{x}_j^{i+1} < f(T_k^{i+1})$, for all $k \in [|J^{i+1}|+1]$ for all $i \in [t-1]$. ∎

Lemma 29 *At any iteration $i \in [t]$ of Algorithm 2, $x_1^i \ge x_2^i \ge \cdots \ge x_n^i$, $0 \le \widetilde{\lambda}^i \le 1$.*

Proof We again give an inductive proof of the above claim. Clearly, for $i=1$, $x_1^1 \ge x_2^1 \ge \cdots \ge x_n^1$ holds from the definition of $\mathbf{x}$. Using this we first show that $0 \le \widetilde{\lambda}^1 \le 1$. Note as $x_1^1 \ge x_2^1 \ge \cdots \ge x_n^1$, clearly $\widetilde{\lambda}^1 \ge 0$, as follows from the definition of $\widetilde{\lambda}^1$ in line (11), in fact unless $J^1 = \emptyset$, and $t=1$, $\lambda^1 > 0$. Also it is easy to see that in case $\mathbf{c}^1 = \mathbf{x}^1$, $\widetilde{\lambda}^1 = 1$, then $t=1$, and the algorithm terminates as evident from line (12). Now in case if $\widetilde{\lambda}^1 > 1$, this implies $(x_{j-1}^1 - x_j^1) > (c_{j-1}^1 - c_j^1) > 0 \quad \forall j \in J^1$. Again from Eqn. 6, we know that $\sum_{j\in T_k^1} c_j^1 = f(T_k^1) \ \forall k \in [|J^1|+1]$, and from Lemma 28, $\sum_{j\in T_k^1} x_j^1 \le f(\cup_{k'=1}^k T_k^1)$. Now since $\mathbf{c}^t \ne \mathbf{x}^t$, $x_1^1 \ge x_2^1 \ge \cdots \ge x_n^1$, and both $\mathbf{c}^1$ and $\mathbf{x}^1$ take uniform values within any set S_k^1 of the partition defined by J^1, let S_b^1 ($b \in [|J^1|+1]$) be any such set such that $x_j^1 \ne c_j^1$, $j \in S_b^1$. Then if $x_j^1 > c_j^1$ for $j \in S_b^1$, and $x_j^1 = c_j^1$, $\forall j \in S_{b+1}^1, ..., S_{|J^1|+1}^1$ this implies $x_j^1 > c_j^1 \ \forall j \in T_{b-1}^1$ since $(x_{j-1}^1 - x_j^1) > (c_{j-1}^1 - c_j^1) \ \forall j \in J^1$, and hence $\sum_{j=1}^n x_j^1 \ge f([n])$ which gives a contradiction according to Lemma 28. Next if we assume that $x_j^1 < c_j^1$ for $j \in S_b^1$, and $x_j^1 > c_j^1$, $\forall j \in S_1^1, ..., S_{b-1}^1$, then $\sum_{j=1}^n x_j^1 = f([n])$ only if $\sum_{j\in T_{b-1}^1} x_j^1 \ge \sum_{j\in T_{b-1}^1} c_j^1 = f(T_{b-1}^1)$, which gives a contradiction according to Lemma 28. Hence $\widetilde{\lambda}^1 < 1$.

Next using the fact that if $t > 1$, then $0 \le \widetilde{\lambda}^1 < 1$, we show that $x_1^2 \ge x_2^2 \ge \cdots \ge x_n^2$. First consider all the coordinates $j \in \{2, 3, ...n\}$ where $x_{j-1}^1 = x_j^1$, clearly in this case $c_{j-1}^1 = c_j^1$ too, hence,

$$x_{j-1}^2 - x_j^2 = \frac{x_{j-1}^1 - \widetilde{\lambda}^1 c_{j-1}^1}{1-\widetilde{\lambda}^1} - \frac{x_j^1 - \widetilde{\lambda}^1 c_j^1}{1-\widetilde{\lambda}^1} = \frac{x_{j-1}^1 - x_j^1}{1-\widetilde{\lambda}^1} - \frac{\widetilde{\lambda}^1 (c_{j-1}^1 - c_j^1)}{1-\widetilde{\lambda}^1} = 0,$$

and hence $x_{j-1}^2 = x_j^2$. Now consider $j \in \{2, 3, ..., n\}$ such that $x_{j-1}^1 > x_j^1$, by definition of c^1, in this case $c_{j-1}^1 > c_j^1$, thus

$$x_{j-1}^2 - x_j^2 = \frac{(x_{j-1}^1 - x_j^1) - \widetilde{\lambda}^1 (c_{j-1}^1 - c_j^1)}{1-\widetilde{\lambda}^1} = (c_{j-1}^1 - c_j^1) \frac{\frac{(x_{j-1}^1 - x_j^1)}{(c_{j-1}^1 - c_j^1)} - \widetilde{\lambda}^1}{1-\widetilde{\lambda}^1} > 0,$$

where the last inequality follows from the definition of $\widetilde{\lambda}^1$ (see line (10) of algorithm 2), and the fact that $\widetilde{\lambda}^1 < 1$. Thus we get $x_1^2 \ge x_2^2 \ge \cdots \ge x_n^2$ using which along with Lemma 28 we can similarly prove that $0 \le \lambda^2 \le 1$ and so on $\forall i \in [t]$. ∎

Lemma 30 *For any $i \in [t]$, $0 \le \lambda^i \le 1$, and $\sum_{i=1}^t \lambda^i = 1$.*

Proof Clearly if the algorithm terminates in the first iteration $x^1 = c^1$, $t = 1$, and $\lambda^1 = \widetilde{\lambda}^1 = 1$. Now suppose $t > 1$, then at any iteration $i \in [t-1]$ it is easy to check that, $\lambda^i = \big(1 - \sum_{k=1}^{i-1} \lambda^k\big)\widetilde{\lambda}^i = \prod_{k=1}^{i-1}(1 - \widetilde{\lambda}^k)\widetilde{\lambda}^i$. Next to prove that, $0 \leq \lambda^t \leq 1$, we show $\sum_{i=1}^{t-1} \lambda^i \leq 1$, this is true because $\sum_{i=1}^{t-1} \lambda^i = \sum_{i=1}^{t-1} \big(\prod_{k=1}^{i-1}(1-\widetilde{\lambda}^k)\widetilde{\lambda}^i\big) = \widetilde{\lambda}^1 + (1-\widetilde{\lambda}^1)\big(\widetilde{\lambda}^2 + (1-\widetilde{\lambda}^2)\big(\widetilde{\lambda}^3 + (1-\widetilde{\lambda}^3)\big(...(\widetilde{\lambda}^{t-2} + (1-\widetilde{\lambda}^{t-2})(\widetilde{\lambda}^{t-1})))...\big)\big) \leq 1$, as for $0 \leq y, z \leq 1$, $((1-y)z + y) \leq 1$. Now from Lemma 29, we see that $0 \leq \widetilde{\lambda}^i \leq 1$, $\forall i \in [t-1]$, which implies $0 \leq \lambda^i \leq 1$, $\forall i \in [t-1]$.

From the above arguments it is easy to prove that $\sum_{i=1}^{t} \lambda^i = 1$. Clearly if $t = 1$, $\lambda^1 = 1$, $c^1 = \mathbf{x}^1 = \mathbf{x}$. Now let $t > 1$, clearly by definition of λ^t in line (19), $\sum_{i=1}^{t} \lambda^i = 1$. ∎

Lemma 31 $\mathbf{x} = \sum_{i=1}^{t} \lambda^i \mathbf{c}^i$

Proof Recall from Lemma 30, $\forall i \in [t-1]$, $\lambda^i = \big(1 - \sum_{k=1}^{i-1} \lambda^k\big)\widetilde{\lambda}^i = \prod_{k=1}^{i-1}(1 - \widetilde{\lambda}^k)\widetilde{\lambda}^i$. Now, $\forall i \in [t-1]$, $\mathbf{x}^{i+1} = \frac{\mathbf{x}^i - \widetilde{\lambda}^i \mathbf{c}^i}{1-\widetilde{\lambda}^i}$, which gives $\mathbf{x}^i = \widetilde{\lambda}^i \mathbf{c}^i + \mathbf{x}^{i+1}(1-\widetilde{\lambda}^i)$. Thus $x^1 = \widetilde{\lambda}^1 \mathbf{c}^1 + \mathbf{x}^2(1-\widetilde{\lambda^1}) = \lambda^1 \mathbf{c}^1 + \big(\widetilde{\lambda}^2 \mathbf{c}^2 + \mathbf{x}^3(1-\widetilde{\lambda^2})\big)(1-\widetilde{\lambda^1}) = \lambda^1 \mathbf{c}^1 + \lambda^2 \mathbf{c}^2 + \mathbf{x}^3(1-\widetilde{\lambda^2})(1-\widetilde{\lambda^1}) = \cdots = \sum_{i=1}^{t-1} \lambda^i c^i + \mathbf{x}^t \prod_{i=1}^{t-1}(1-\widetilde{\lambda}^i) = \sum_{i=1}^{t} \lambda^i c^i$, as for $i = t$, $\mathbf{c}^t = \mathbf{x}^t$, and $\lambda^t = (1 - \sum_{i=1}^{t-1} \lambda^i) = \prod_{i=1}^{t-1}(1-\widetilde{\lambda}^i)$. ∎

Lemma 32 *For any iteration $i \in [t-1]$, $|J^i| \geq |J^{i+1}| + 1$, hence $1 \leq t \leq n$.*

Proof Clearly if $J^1 = \emptyset$, or $\widetilde{\lambda}^1 = 1$, $t = 1$. Suppose $t > 1$, and consider any $i \in [t-1]$, clearly in that case $J^i \neq \emptyset$, and $\widetilde{\lambda}^i \neq 1$. Now let $k = \mathrm{argmin}_{j \in J^1}\{\frac{(x^i_{j-1} - x^i_j)}{(c^i_{j-1} - c^i_j)}\}$, then $(x^{i+1}_{k-1} - x^{i+1}_k) = \frac{x^i_{k-1} - \widetilde{\lambda}^i c^i_{k-1}}{1-\widetilde{\lambda}^i} - \frac{x^i_k - \widetilde{\lambda}^i c^i_k}{1-\widetilde{\lambda}^i} = \frac{\frac{x^i_{k-1} - x^i_k}{c^i_{k-1} - ^i_k} - \widetilde{\lambda}^i}{1-\widetilde{\lambda}^i} = 0$, hence $J^{i+1} = J^i \setminus \{k\}$, and $|J^i| = |J^{i+1}| + 1$. Clearly, if k is not unique $|J^i| > |J^{i+1}| + 1$. ∎

Thus Lemma 32 shows that Algorithm 2 runs in at most n iterations, within any iteration $i \in [t]$, all steps can be computed in $O(n)$ time except the computation of c^i which in general can take $T(f)$ time, thus total running time of the algorithm is $O(n^2 + nT(f))$. ∎

Appendix C. Supplement to Section 5

Theorem19 [Decomposition of face P-centers of *c-symmetric GSBPs* into extreme points] Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$ such that the GSBP $B(f, \mathbf{w})$ is *c-symmetric*, where $c \geq 0$ is a constant. Then any face P-center of $B(f, \mathbf{w})$ can be decomposed into a convex combination of the extreme points of its corresponding face in $O(n^{c+1}\, T(f))$ time, moreover any such extreme point can be randomly sampled in $O(n^c\, T(f))$ time.

Proof Recall from the Definition 18, a GSBP $B(f, \mathbf{w})$ is called c-symmetric if for every $0 < d \leq n$, and all ordered d-partition $P = (S_1, \ldots, S_d)$ of $[n]$, for any $\sigma \in \Sigma^P_{[n]}$, and $k \in [d]$,

$c_i^P = \frac{\sum\limits_{\sigma' \in R(\sigma, S_k)} e_i^{\sigma'}}{|R(\sigma, S_k)|} \quad \forall i \in S_k$, such that there exists a constant $c \geq 0$, with $|R(\sigma, S_k)| = |S_k|^c \; \forall k \in [d]$. Hence for all $i \in S_k$, the i^{th} coordinate of c^P can be represented as an average of the i^{th} coordinates of at most $|S_k|^c$ extreme points of the $B(f, \mathbf{w})$, whose underlying $O(|S_k|^c)$ permutations can be computed in $O(|S_k|^c)$ time, and using Eqn. 1 all n coordinates of such extreme points can be evaluated in $O(n|S_k|^c T(f))$ time. Now since for all $k \in [d]$, $|S_k| \leq n$, and $\sum_{k=1}^{d} |S_k| = n$, combinedly decomposing all n coordinates of c^P can take $O(n^{c+1} T(f))$ time, in the worst case.

Note that instead if we are interested in randomly sampling only one such extreme point, for all $i \in S_k$, we can compute the feasible set of permutations in $O(|S_k|^c)$ time, and randomly sample one out of them. For this particular permutation, we can compute all the i^{th} coordinates of its corresponding extreme point in $|S_k| T(f)$ time. Repeating this for each $k \in [d]$, combinedly all n coordinates of the desired extreme point can be computed in $O(n\, T(f))$ time. Thus the total running time complexity for randomly sampling one such extreme point is given by $O(n^c T(f))$. ■

Theorem 20 [Decomposition of *c-symmetric GSBPs*, partition-respecting GSBPs into extreme points] Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$ such that the GSBP $B(f, \mathbf{w})$ is *c-symmetric*, where $c \geq 0$ is a constant. Then any point of $B(f, \mathbf{w})$ can be decomposed into a convex combination of its extreme points in $O(n^2 + n\, T(f) + n^{c+2}\, T(f))$ time, moreover any such extreme point can be randomly sampled in $O(n^2 + n\, T(f) + n^{c+1}\, T(f))$ time.

Proof The proof follows from the results of Theorem 11 and 19. ■

Appendix D. Supplement to Section 5.1

Theorem 23 [Decomposition of partition-respecting, circular symmetric GSBPs into extreme points]. Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$ such that the GSBP $B(f, \mathbf{w})$ is partition-respecting and circular symmetric. Then any point of $B(f, \mathbf{w})$ can be decomposed into a convex combination of its extreme points in $O(n^3\, T(f))$ time, moreover any such extreme point can be randomly sampled in $O(n^2\, T(f))$ time.

Proof The proof follows from the proof of Theorem 20, note that in this case $c = 1$. ■

Theorem 24 [Characterization of partition-respecting, circular symmetric GSBPs] Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$. If the GSBP $B(f, \mathbf{w})$ is partition-respecting and circular symmetric then f satisfies the following conditions:

$$\forall \{i, j, k\} \subseteq [n] : \quad f(\{i\})(w_j - w_k) + f(\{j\})(w_k - w_i) + f(\{k\})(w_i - w_j) = 0$$

$$\forall \{i, j\} \subseteq [n] : \quad f(\{i, j\}) = \big(f(\{i\}) - f(\{j\})\big) \frac{w_i + w_j}{w_i - w_j}$$

$$\forall\{i,j,k\} \subseteq [n]: \quad f(\{i,j,k\}) \;=\; \big((f(\{i\}) - f(\{j\})) + (f(\{i,j\}) - f(\{j,k\}))\big)\frac{w_i + w_j + w_k}{2w_i - (w_j + w_k)}$$

$$\vdots$$

$$f([n]) = \frac{(f(\{1\}) - f(\{2\})) + (f(\{1,2\}) - f(\{2,3\})) + \cdots + (f([n-1]) - f([n]\setminus\{1\}))}{\big((n-1)w_1 - \sum_{j\in[n-1]} w_j\big)} \sum_{j\in[n]} w_j$$

Proof This can be proved similar to the proof of theorem 10. Given that $B(f, \mathbf{w})$ is partition-respecting, consider any d-ordered partition $P = (S_1, \cdots, S_d)$, $0 < d \le n$. Recall the definition of *partition-respecting* GSBP from section 4.1, we will first analyze the first condition of Definition 7 given by Eqn. 4. Clearly when $d = n$, or $|S_k| = 1 \ \forall k \in [d]$, Eqn. 4 trivially holds. Now let $|S_1| = 2$, in general we can assume $S_1 = \{i, j\} \subseteq [n]$. Now since $B(f, \mathbf{w})$ is partition-respecting, from Eqn. 6, we get $c_1^P = \frac{f(S_1) - f(\emptyset)}{\sum_{j\in S_1} w_j} \ \forall k \in S$, thus using Eqn. 1 and from the partition respecting property, for any $\tau \in \Sigma_{[n]}^P$,

$$\begin{aligned}
\frac{\sum_{\sigma\in C(\tau)} e_i^\sigma}{|S_1|} &= \frac{f(S_1) - f(\emptyset)}{\big(\sum_{j\in S_1} w_j\big)} \\
\implies \frac{f(\{i\}) - f(\emptyset) + f(\{i,j\}) - f(\{j\})}{2w_i} &= \frac{f(\{i,j\}) - f(\emptyset)}{(w_i + w_j)} \\
\implies f(\{i\}) - f(\{j\}) &= 2w_i \frac{f(\{i,j\})}{(w_i + w_j)} - f(\{i,j\}) \\
\implies f(\{i,j\}) &= \big(f(\{i\}) - f(\{j\})\big)\frac{w_i + w_j}{w_i - w_j}.
\end{aligned} \tag{11}$$

Now instead consider $|S_1| = 3$, such that $S_1 = \{i,j,k\} \subseteq [n]$. Again since $B(f, \mathbf{w})$ is partition-respecting and circular symmetric, as before using Eqn. 1 and the partition respcting property, for any $\tau \in \Sigma_{[n]}^P$, we get,

$$\begin{aligned}
\frac{\sum_{\sigma\in C(\tau)} e_i^\sigma}{|S_1|} &= \frac{f(S_1) - f(\emptyset)}{\big(\sum_{j\in S_1} w_j\big)} \\
\frac{f(\{i\}) + (f(\{i,j\}) - f(\{j\})) + (f(\{i,j,k\}) - f(\{j,k\}))}{3w_i} &= \frac{f(\{i,j,k\})}{w_i + w_j + w_k} \\
\implies f(\{i,j,k\}) &= \big((f(\{i\}) - f(\{j\})) + (f(\{i,j\}) - f(\{j,k\}))\big)\frac{w_i + w_j + w_k}{2w_i - (w_j + w_k)}
\end{aligned} \tag{12}$$

Similarly applying Eqn. 8 for $j \in S_1$ we get,

$$f(\{i,j,k\}) = \big((f(\{j\}) - f(\{k\})) + (f(\{j,k\}) - f(\{i,k\}))\big)\frac{w_i + w_j + w_k}{2w_j - (w_i + w_k)} \tag{13}$$

Now equating Eqn. 12 and 13, applying Eqn. 11 we get $f(\{i\})(w_j - w_k) + f(\{j\})(w_k - w_i) + f(\{k\})(w_i - w_j) = 0$. Similarly considering different cardinalities of S_1, one can derive the expressions of f on any set $S \subseteq [n]$, and finally considering $S_1 = [n]$, and using Eqn. 11 we get $f([n]) = \left(\frac{(f(\{1\})-f(\{2\}))+(f(\{1,2\})-f(\{2,3\}))+\cdots+(f([n-1])-f([n]\setminus\{1\}))}{\left((n-1)w_1-\sum_{j\in[n-1]} w_j\right)}\sum_{j\in[n]} w_j\right)\left(\sum_{i'=1}^{n} w_{i'}\right)$. Note that instead of S_1, we could have worked with any other partition set S_k $k \in [d]$, $0 < d \leq 1$ which essentially would given same the expression of $f(\cup_{k'=1}^{k} S_{k'})$ that we get considering $S_1 = \cup_{k'=1}^{k} S_{k'}$.

The result now follows by expressing right hand side of every constraint in terms of the singletons. ∎

Appendix E. Supplement to Section 5.2

Proof The proof follows from the proof of Theorem 19, note that in this case $c = 0$. ∎

Theorem 26 [Decomposition of partition-respecting, reflexive symmetric GSBPs into extreme points]. Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$ such that the GSBP $B(f, \mathbf{w})$ is partition-respecting and reflexive symmetric. Then any point of $B(f, \mathbf{w})$ can be decomposed into a convex combination of its extreme points in $O(n^2 T(f))$ time, moreover any such extreme point can be randomly sampled in $O(n T(f))$ time.

Proof The proof follows from the proof of Theorem 20, note that in this case $c = 0$. ∎

Theorem 27 [Characterization of partition-respecting, reflexive symmetric GSBPs]. Let $f : 2^{[n]} \to \mathbb{R}$ be a submodular function with $f(\emptyset) = 0$, and $\mathbf{w} \in \mathbb{R}^n$. If the GSBP $B(f, \mathbf{w})$ is partition-respecting and reflexive symmetric then f satisfies the following conditions:

$$\forall \{i,j,k\} \subseteq [n] : \quad \frac{f(\{i\})}{w_i}(w_j - w_k) + \frac{f(\{j\})}{w_j}(w_k - w_i) + \frac{f(\{k\})}{w_k}(w_i - w_j) = 0$$

$$\begin{aligned}
\forall \{i,j\} \subseteq [n] : \ f(\{i,j\}) &= \big(f(\{i\}) - f(\{j\})\big)\frac{w_i + w_j}{w_i - w_j} \\
\forall \{i,j,k\} \subseteq [n] : \ f(\{i,j,k\}) &= \big(f(\{i\}) - f(\{j,k\})\big)\frac{w_i + w_j + w_k}{w_i - (w_j + w_k)} \\
&\ \vdots
\end{aligned}$$

$$f([n]) = \big(f(\{n\}) - f([n-1])\big)\frac{\sum_{j\in[n]} w_j}{w_1 - \sum_{j\in[n-1]} w_j}$$

Proof This can be proved similar to the proof of theorem 10. Given that $B(f, \mathbf{w})$ is partition-respecting, consider any d-ordered partition $P = (S_1, \cdots, S_d)$, $0 < d \leq n$. Recall the definition of *partition-respecting* GSBP from section 4.1, we will first analyze the first condition of Definition 7 given by Eqn. 4. Clearly when $d = n$, or $|S_k| = 1$ $\forall k \in [d]$, Eqn. 4 trivially holds. Now let $|S_1| = 2$, in general we can assume $S_1 = \{i, j\} \subseteq [n]$. Now since $B(f, \mathbf{w})$ is partition-respecting, from Eqn. 6, we get $c_1^P = \frac{f(S_1) - f(\emptyset)}{\sum_{j\in S_1} w_j}$ $\forall k \in S$, thus using Eqn. 1 and using the partition respecting property, for any $\sigma \in \Sigma_{[n]}^P$,

$$\begin{aligned}
&\frac{e_i^{\sigma} + e_i^{\sigma_R^{S_k}}}{2} = \frac{f(S_1) - f(\emptyset)}{\left(\sum_{j \in S_1} w_j\right)} \\
\Longrightarrow &\frac{f(\{i\}) - f(\emptyset) + f(\{i,j\}) - f(\{j\})}{2w_i} = \frac{f(\{i,j\}) - f(\emptyset)}{(w_i + w_j)} \\
\Longrightarrow &f(\{i\}) - f(\{j\}) = 2w_i \frac{f(\{i,j\})}{(w_i + w_j)} - f(\{i,j\}) \\
\Longrightarrow &f(\{i,j\}) = \big(f(\{i\}) - f(\{j\})\big) \frac{w_i + w_j}{w_i - w_j}. \qquad (14)
\end{aligned}$$

Now instead consider $|S_1| = 3$, such that $S_1 = \{i,j,k\} \subseteq [n]$. Again since $B(f, \mathbf{w})$ is partition-respecting and reflexive symmetric, as before using Eqn. 1 and from partition respecting property, consider any $\sigma \in \Sigma^P_{[n]}$ such that $\sigma(i) = 1$[4], we get,

$$\begin{aligned}
&\frac{e_i^{\sigma} + e_i^{\sigma_R^{S_k}}}{2} = \frac{f(S_1) - f(\emptyset)}{\left(\sum_{j \in S_1} w_j\right)} \\
&\frac{f(\{i\}) + (f(\{i,j,k\}) - f(\{j,k\}))}{2w_i} = \frac{f(\{i,j,k\})}{w_i + w_j + w_k} \\
\Longrightarrow &f(\{i,j,k\}) = \big(f(\{i\}) - f(\{j,k\})\big) \frac{w_i + w_j + w_k}{w_i - (w_j + w_k)} \qquad (15)
\end{aligned}$$

Using similar argument for $j \in S_1$ we get,

$$f(\{i,j,k\}) = \big(f(\{j\}) - f(\{i,k\})\big) \frac{w_i + w_j + w_k}{w_j - (w_i + w_k)} \qquad (16)$$

Now equating Eqn. 15 and 16, applying Eqn. 14 we get $\frac{f(\{i\})}{w_i}(w_j - w_k) + \frac{f(\{j\})}{w_j}(w_k - w_i) + \frac{f(\{k\})}{w_k}(w_i - w_j) = 0$. Similarly considering different cardinalities of S_1, one can derive the expressions of f on any set $S \subseteq [n]$, and finally considering $S_1 = [n]$, and using Eqn. 14 we get $f([n]) = \left(\big(f(\{n\}) - f([n-1])\big)\frac{\sum_{j \in [n]} w_j}{w_1 - \sum_{j \in [n-1]} w_j}\right)\left(\sum_{i'=1}^{n} w_{i'}\right)$. Note that instead of S_1, we could have worked with any other partition set S_k $k \in [d]$, $0 < d \leq 1$ which essentially would given same the expression of $f(\cup_{k'=1}^{k} S_{k'})$ that we get considering $S_1 = \cup_{k'=1}^{k} S_{k'}$.

Similar to the proof of Thm. 24, the result now follows by expressing right hand side of every constraint in terms of the singletons. ■

4. We get the same results considering $\sigma(i) = 2$ or $\sigma(i) = 3$ as well.