

Submodular Dominance and Applications

Frederick Qiu ✉

Department of Computer Science, Princeton University, NJ, USA

Sahil Singla ✉

School of Computer Science, Georgia Tech, Atlanta, GA, USA

Abstract

In submodular optimization we often deal with the expected value of a submodular function f on a distribution \mathcal{D} over sets of elements. In this work we study such submodular expectations for negatively dependent distributions. We introduce a natural notion of negative dependence, which we call *Weak Negative Regression* (WNR), that generalizes both Negative Association and Negative Regression. We observe that WNR distributions satisfy *Submodular Dominance*, whereby the expected value of f under \mathcal{D} is at least the expected value of f under a product distribution with the same element-marginals.

Next, we give several applications of Submodular Dominance to submodular optimization. In particular, we improve the best known submodular prophet inequalities, we develop new rounding techniques for polytopes of set systems that admit negatively dependent distributions, and we prove existence of contention resolution schemes for WNR distributions.

2012 ACM Subject Classification Theory of computation → Approximation algorithms analysis; Mathematics of computing → Probabilistic algorithms; Theory of computation → Algorithmic game theory

Keywords and phrases Submodular Optimization, Negative Dependence, Negative Association, Weak Negative Regression, Submodular Dominance, Submodular Prophet Inequality

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2022.44

Category APPROX

1 Introduction

A function $f : 2^U \rightarrow \mathbb{R}$ on universe $U = \{1, \dots, n\}$ is *submodular* if it satisfies $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$ for all $S, T \subseteq U$. These functions capture the concept of diminishing returns, and are therefore useful in many fields such as machine learning, operations research, mechanism design, and combinatorial optimization; see books [20, 3, 36, 31].

Although f is a discrete function, for many applications it is useful to define a continuous relaxation $f_{\text{cont}} : [0, 1]^n \rightarrow \mathbb{R}$ of f , since that allows us to use techniques from continuous optimization. Here, by a relaxation we mean that f_{cont} equals f at the indicator vectors of the sets, i.e., $f_{\text{cont}}(\mathbf{1}_S) = f(S)$ for all $S \subseteq U$. A standard way to define such continuous relaxations is to first define a probability distribution $\mathcal{D}(\mathbf{x})$ over subsets of U with element-marginals $\mathbf{x} \in [0, 1]^n$, and then define $f_{\text{cont}}(\mathbf{x})$ to be the expectation with respect to this distribution, i.e., $f_{\text{cont}}(\mathbf{x}) := \mathbb{E}_{S \sim \mathcal{D}(\mathbf{x})}[f(S)]$, where S is a random set drawn from $\mathcal{D}(\mathbf{x})$. For example, the popular *multilinear relaxation* $F(\mathbf{x})$ is defined by taking $\mathcal{D}(\mathbf{x})$ to be the product distribution with marginals \mathbf{x} . Other examples include the convex closure relaxation $f^-(\mathbf{x})$ (which is equivalent to the Lovász extension for submodular functions), the concave closure relaxation $f^+(\mathbf{x})$, and the relaxation $f^*(\mathbf{x})$ [38]. Studying the properties of submodular expectations for these distributions has been a fruitful direction, which has led us to several optimal/approximation algorithms for submodular optimization [3, 39, 7, 18, 2, 16].

Given the success of the above continuous relaxations, it is natural to ask what other continuous relaxations, or equivalently, what other submodular expectations and distributions $\mathcal{D}(\mathbf{x})$ could be defined that are useful for new or improved applications. In this work, we study



© Frederick Qiu and Sahil Singla;

licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022).

Editors: Amit Chakrabarti and Chaitanya Swamy; Article No. 44; pp. 44:1–44:21



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

submodular expectations for negatively dependent distributions and use them to improve the best known submodular prophet inequalities, to develop new rounding techniques, and to design contention resolution schemes for negatively dependent distributions.

1.1 Submodular Dominance

Since the multilinear extension F is commonly employed in combinatorial optimization, one avenue to explore other continuous relaxations is by comparing them to F .

► **Definition 1** (Submodular Dominance). *A distribution \mathcal{D} over 2^U with marginals $\mathbf{x} \in [0, 1]^n$ satisfies Submodular Dominance if for every submodular function $f : 2^U \rightarrow \mathbb{R}$,*

$$\mathbb{E}_{S \sim \mathcal{D}}[f(S)] \geq F(\mathbf{x}) .$$

Shao [37] studied a similar concept that he called a comparison theorem, which involved a subclass of submodular functions. Christofides and Vaggelatou [13] later studied what they called the supermodular ordering, which is essentially equivalent to Submodular Dominance. Both viewed the problem through the lens of probability theory, whereas we approach it from the standpoint of combinatorial optimization.

It is not difficult to see how one might apply Submodular Dominance, e.g., it immediately yields an algorithm to round multilinear extension subject to feasibility constraints. However, Submodular Dominance implies a much wider variety of results in stochastic settings, where most of our current understanding relies on the independence of random variables. By relating product distributions to more complex distributions, Submodular Dominance allows us to improve existing results and study more general problems.

1.2 Negative Dependence and Submodular Dominance

Positive correlations can only decrease the expectations of submodular functions due to their diminishing marginal returns, so we turn our attention to negatively dependent distributions. Pemantle initiated a systematic study of such distributions in [32]. In this work, we introduce the following generalization of Negative Association (NA) and Negative Regression (NR), two popular notions of negative dependence (details in Section 2).

► **Definition 2** (WNR). *A distribution \mathcal{D} over 2^U satisfies Weak Negative Regression (WNR) if for any $i \in U$ and any monotone function $f : 2^U \rightarrow \mathbb{R}$,¹*

$$\mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus i) \mid i \in S] \leq \mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus i) \mid i \notin S] . \quad (1)$$

Equivalently, \mathcal{D} is WNR if $S \setminus i$ conditioned on $i \notin S$ stochastically dominates $S \setminus i$ conditioned on $i \in S$ for all $i \in U$. This captures an intuitive notion of negative dependence where conditioning on including an element lowers the probability of other inclusion events. WNR distributions satisfy Submodular Dominance as well as many desirable closure properties.

Submodular Dominance for Negatively Dependent Distributions. Christofides and Vaggelatou [13] proved that NA distributions over continuous random variables satisfy Submodular Dominance for a continuous generalization of submodular functions. We strengthen their result in Section 3 in the setting of Bernoulli random variables from NA to WNR distributions, a strict superset of the union of NA and NR distributions.

¹ A function f is monotone if it satisfies $f(S) \leq f(T)$ for all $S \subseteq T$. Elements should be taken as singleton sets depending on context, e.g., $S \setminus i$ means $S \setminus \{i\}$.

► **Theorem 3.** *WNR distributions satisfy Submodular Dominance.*

It turns out that there exist distributions that satisfy Submodular Dominance but are not WNR. This raises the question: what conditions are necessary for Submodular Dominance? We first recall the classic notion of Negative Cylinder Dependence (see, e.g., [21]).

► **Definition 4 (NCD).** *A distribution \mathcal{D} over 2^U with marginals \mathbf{x} satisfies Negative Cylinder Dependence (NCD) if for any $T \subseteq U$,*²

$$\Pr_{S \sim \mathcal{D}}[T \subseteq S] \leq \Pr_{S \sim \mathbf{x}}[T \subseteq S] \quad \text{and} \quad \Pr_{S \sim \mathcal{D}}[T \subseteq S^c] \leq \Pr_{S \sim \mathbf{x}}[T \subseteq S^c] .$$

NCD can be interpreted as saying that any subset of elements are negatively correlated.

► **Theorem 5.** *All distributions that satisfy Submodular Dominance are NCD.*

This can be useful when Submodular Dominance is an easier property to prove. For example, the distribution arising from randomized swap rounding can be shown to satisfy Submodular Dominance via a straightforward convexity argument, but a direct proof that the distribution is NCD is more involved [10]; this theorem shows that such results follow due to a natural relationship between Submodular Dominance and negative dependence rather than any algorithm specific properties.

Although NCD is necessary for Submodular Dominance, it is insufficient on its own. While this insufficiency result was previously known [11],³ we strengthen it by constructing an example of an NCD distribution which violates Submodular Dominance and is additionally homogeneous, meaning it is distributed only on sets of the same size. Such distributions occur often enough to be of interest, e.g., distributions over the bases of a matroid.

1.3 Applications

Besides being a natural question, Submodular Dominance has several applications.

Submodular Prophet Inequalities. The Prophet Inequality is a classical problem where a gambler sees the realizations of non-negative random variables one-by-one, choosing a random variable in an online fashion and attempting to maximize its value. The celebrated result of Krengel, Sucheston, and Garling [28, 29] demonstrates a $1/2$ prophet inequality, meaning that just knowing the distributions in advance is enough to obtain $1/2$ the expectation obtained by the prophet that knows all the realizations in advance.

Motivated by applications to mechanism design, several works extended the $1/2$ prophet inequality to gamblers selecting multiple random variables subject to a packing constraint to maximize a linear objective function, e.g., [24, 8, 1, 27, 34]. The *Submodular Prophet Inequality* (SPI) was introduced by Rubinstein and Singla [35] as a further generalization to submodular objective functions to capture combinatorial applications.

One significant complication in SPI is that beyond simple Bernoulli settings, we deal with expectations that are no longer taken over product distributions. Chekuri and Livanos [9] obtain an efficient⁴ $c \cdot (1 - e^{-b}) \cdot (e^{-b} - \epsilon)$ SPI for set systems with solvable polytopes⁵ and

² $S \sim \mathbf{x}$ means S is sampled from the product distribution with marginals \mathbf{x} .

³ Observing that certain randomized rounding algorithms give rise to distributions satisfying both Submodular Dominance and NCD, Chekuri, Vondrák, and Zenklusen [11] remarked that there exist NCD distributions which violate Submodular Dominance, so NCD was not sufficient for Submodular Dominance. Our Theorem 5 shows the other direction, that Submodular Dominance implies NCD.

⁴ We use efficient to mean algorithms that run in probabilistic polynomial time.

⁵ The polytope $\mathcal{P}_{\mathcal{I}}$ of a set system \mathcal{I} is formed by taking the convex hull of the indicator vectors of maximal independent sets in \mathcal{I} , and is solvable if linear objective functions can be efficiently maximized over it.

44:4 Submodular Dominance and Applications

an efficient (b, c) -selectable greedy online contention resolution scheme (OCRS) for product distributions (see formal definitions in Section 4). Crucially, their result loses a factor of $e^{-b} - \epsilon$ to handle the non-product distributions of SPI. We use Submodular Dominance to re-analyze the performance of greedy OCRSs in Section 4.3, which allows us to save this factor of $e^{-b} - \epsilon$ and improve the best known SPIs.

► **Theorem 6** (Submodular Prophet Inequalities). *For fixed $\epsilon > 0$, if a set system $\mathcal{I} \subseteq 2^U$ has a solvable polytope and an efficient (b, c) -selectable greedy OCRS for product distributions:*

- *There is an efficient $c \cdot (1 - e^{-b} - \epsilon)$ SPI for monotone non-negative submodular functions.*
- *There is an efficient $c/4 \cdot (1 - e^{-b} - \epsilon)$ SPI for general non-negative submodular functions.*

Combining with known greedy OCRSs, this implies efficient SPIs as given in Table 1.

■ **Table 1** Submodular Prophet Inequalities for different feasibility constraints.

Feasibility Constraint	Prior Best [9]		Our Results	
	Monotone	General	Monotone	General
Uniform Matroid of rank $k \rightarrow \infty$	1/4.30	1/17.20	$1 - 1/e - \epsilon$	1/6.33
Matroid	1/7.39	1/29.54	1/5.02	1/20.07
Matching	1/9.49	1/37.93	1/6.75	1/27.00
Knapsack	1/17.41	1/69.64	1/13.40	1/53.60

It is known that even for offline monotone submodular maximization over uniform matroids, no efficient algorithm can do better than a $(1 - 1/e)$ -approximation [30]. Thus, we obtain the first optimal efficient $1 - 1/e - \epsilon$ monotone SPI over large rank uniform matroids.

Submodular Maximization. Another application is sampling from WNR distributions as a randomized rounding technique where the integral solution obtains at least the value of the fractional solution in expectation. A common method in submodular optimization is to first maximize the multilinear extension, which Vondrák [39] showed can be done for downward-closed set systems with solvable polytopes. For matroids, we know of methods which round the fractional solutions to sets without losing value [6, 10, 11], but set systems with solvable polytopes are far more general than matroids. Thus, the challenge in going beyond matroids is rounding the multilinear extension. By Submodular Dominance, an algorithm that efficiently generates a WNR distribution for a polytope automatically rounds the multilinear extension, which we show has immediate consequences for submodular maximization (details in Section 5.1).

► **Theorem 7** (Submodular Maximization). *Let $f : 2^U \rightarrow \mathbb{R}_{\geq 0}$ be a monotone submodular function. If a downward-closed set system $\mathcal{I} \subseteq 2^U$ has a solvable polytope and efficiently admits WNR distributions, there exists an efficient algorithm that returns $T \in \mathcal{I}$ such that $\mathbb{E}[f(T)] \geq (1 - 1/e - o(1)) \cdot \max_{S \in \mathcal{I}} f(S)$.*

Adaptivity Gaps for Stochastic Probing. A natural generalization of submodular maximization is by adding stochasticity: replace elements by random variables called items. Such problems are often known as Stochastic Probing [22, 2, 23, 5, 17]. In addition to knowing the distributions of the items, we also allow algorithms to learn the realization of an item after selecting it. This opens up the concept of adaptive algorithms, which modify their behavior

conditioned on such realizations. Though adaptivity can result in better algorithms, it also introduces significant complexity; for example, a decision tree can be of exponential size. Therefore, non-adaptive algorithms may be preferable if their performance is comparable to that of the optimal adaptive algorithm, a concept known as the adaptivity gap. By sampling from WNR distributions to round the multilinear extension, we adapt the analysis of the adaptivity gap upper bound by Asadpour and Nazerzadeh [2] from matroids to any set system for which WNR distributions exist (details in Section 5.2).

► **Theorem 8** (Stochastic Probing). *For a downward-closed set system \mathcal{I} that admits WNR distributions, the adaptivity gap for Stochastic Probing is upper-bounded by $\frac{e}{e-1}$.*

Contention Resolution Schemes. Contention resolution schemes (CRS) are another randomized rounding technique, with the concept being formally introduced by [12] for submodular maximization. (Similar but less thoroughly explored notions appear in earlier works such as [4].) Since submodular maximization usually occurs via approximations of the multilinear extension, CRSs have generally been studied with respect to product distributions. Recently, Dughmi [14, 15] initiated the study of CRSs for non-product distributions because of their applications in settings such as the Matroid Secretary Problem. We extend the CRS of [12] for matroids from product distributions to WNR distributions, which gives possible directions to generalize our understanding of CRSs (details in Section 5.3).

► **Theorem 9** (Contention Resolution Schemes). *For a matroid \mathcal{M} , there exists a $(1 - 1/e)$ -selectable CRS for any WNR distribution with marginals $\mathbf{x} \in \mathcal{P}_{\mathcal{M}}$.*

2 WNR and Other Negatively Dependent Distributions

In this section, we first discuss popular notions of negative dependence, and then introduce WNR and study its various properties. Lengthier proofs are deferred to Appendix A.1.

► **Definition 10** (NA). *A distribution \mathcal{D} over 2^U satisfies Negative Association (NA) if for any monotone $f, g : 2^U \rightarrow \mathbb{R}$ depending on disjoint sets of elements, $\text{Cov}_{S \sim \mathcal{D}}[f(S), g(S)] \leq 0$.*

This property is very similar to the Positive Association (PA) condition in the FKG inequality, with the main difference being the reversed inequality and disjoint sets. Although the FKG Inequality gives a straightforward condition to check for PA, no analogous result exists for NA, and in general, it is difficult to prove that a distribution is NA [26, 32].

Another way to define negative dependence is based on the idea that conditioning on “larger” inclusion events should reduce the probability of other inclusion events.

► **Definition 11** (NR). *A distribution \mathcal{D} over 2^U satisfies Negative Regression (NR) if for any sets $R_-, R_+, T \subseteq U$ such that $R_- \subseteq R_+ \subseteq T$ and any monotone function $f : 2^U \rightarrow \mathbb{R}$,*

$$\mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus T) \mid (S \cap T) = R_+] \leq \mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus T) \mid (S \cap T) = R_-] .$$

Equivalently, \mathcal{D} is NR if $S \setminus T$ conditioned on $S \cap T = R_-$ stochastically dominates $S \setminus T$ conditioned on $S \cap T = R_+$ for all $R_- \subsetneq R_+ \subseteq T \subseteq U$. It turns out that NR is also a difficult property to check.

Since NA and NR are both natural forms of negative dependence, it is surprising that the exact relationship between them is unknown. While it is known that NA does not imply NR, it is conjectured that NR implies NA [32]. One might then ask whether we can generalize

44:6 Submodular Dominance and Applications

NA and NR to get the best of both worlds: a weaker notion of negative dependence that is easier to check while still satisfying many desirable properties. This is our motivations for defining WNR distributions.

We first reformulate the WNR condition in terms of covariance.

▷ **Claim 12.** The WNR condition (1) is equivalent to $\text{Cov}_{S \sim \mathcal{D}}[f(S \setminus i), \mathbb{1}_{i \in S}] \leq 0$.

Proof. Let \mathcal{D} have marginals \mathbf{x} . The covariance inequality is equivalent to $\mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus i) \cdot \mathbb{1}_{i \in S}] \leq \mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus i)] \mathbb{E}_{S \sim \mathcal{D}}[\mathbb{1}_{i \in S}]$. Then observe that if we expand LHS by conditioning on i , the expectation conditioned on $i \notin S$ is 0 due to the indicator. We can also simplify RHS using $\mathbb{E}_{S \sim \mathcal{D}}[\mathbb{1}_{i \in S}] = x_i$, so the covariance inequality is equivalent to

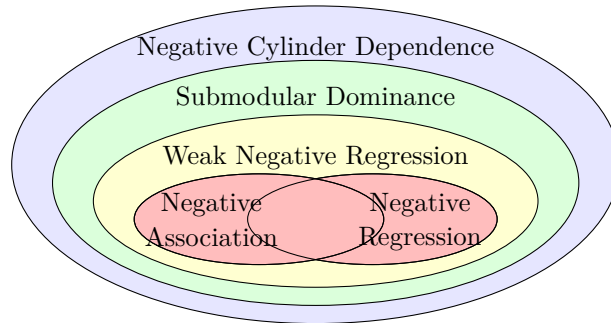
$$x_i \cdot \mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus i) \cdot \mathbb{1}_{i \in S} \mid i \in S] \leq \mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus i)] \cdot x_i .$$

Canceling the x_i and noting that the indicator on LHS is always 1, we have $\mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus i) \mid i \in S] \leq \mathbb{E}_{S \sim \mathcal{D}}[f(S \setminus i)]$. This is equivalent to the WNR condition since the expectation is just a weighted sum of the conditional expectation on $i \in S$ and $i \notin S$. ◀

Next, we prove that WNR distributions generalize NA and NR distributions.

► **Proposition 13.** *NA and NR imply WNR, and WNR implies NCD, but the reverse implications do not hold. In other words, the union of NA and NR distributions is a strict subset of WNR distributions, which is a strict subset of NCD distributions.*

Proof. The WNR condition is a special case of the NR condition when $T := \{i\}$, and by Claim 12, it is also a special case of the NA condition when $g(S) := \mathbb{1}_{i \in S}$, so both NA and NR imply WNR. For strict containment, we give a WNR distribution that is neither NA nor NR in Appendix A.1. Theorems 3 and 5 and the example distributions in Section 3 demonstrate that WNR distributions are a strict subset of NCD distributions. ◀



■ **Figure 1** Hierarchy of negative dependence and its relation to Submodular Dominance.

Finally, we observe that WNR satisfies two closure properties, proved in Appendix A.1.

► **Definition 14** (Projection). *Let \mathcal{D} be a distribution over 2^U . Its projection onto $U' \subseteq U$ is the distribution which samples $S \sim \mathcal{D}$ and returns $S \cap U'$.*

► **Definition 15** (Products). *Let \mathcal{A} and \mathcal{B} be distributions over 2^A and 2^B for disjoint A, B . Their product distribution independently samples $S \sim \mathcal{A}$ and $T \sim \mathcal{B}$, then returns $S \cup T$.*

► **Proposition 16.** *WNR is closed both under projection and under products.*

Joag-Dev and Proschan [26] showed that NA distributions are NCD, closed under projection, closed under products, and closed under taking monotone functions of disjoint subsets of variables.⁶ Since WNR shares three of these properties with NA and requires a weaker condition while also generalizing NR, it appears to be a useful notion of negative dependence.

3 Towards a Characterization of Submodular Dominance

3.1 WNR is a Sufficient Condition

► **Theorem 3.** *WNR distributions satisfy Submodular Dominance.*

Proof. We prove by induction on the number of elements. The base case of 1 element is trivial because the marginals fully specify the distribution.

We will show that any WNR distribution \mathcal{D} over $2^{[k]}$ with marginals \mathbf{x} satisfies Submodular Dominance, assuming by induction that all WNR distributions over $2^{[k-1]}$ satisfy Submodular Dominance. We assume that $x_k \neq 0, 1$ because otherwise, we can interpret \mathcal{D} as a distribution over $2^{[k-1]}$ and trivially be done.

Let $\mathcal{D} \setminus k$ and $\mathbf{x} \setminus k$ denote the projections of \mathcal{D} and \mathbf{x} onto $[k-1]$. Let \mathcal{D}_k denote the distribution which samples $S \sim \mathcal{D} \setminus k$, then returns $S \cup k$ w.p. x_k and returns S otherwise, i.e., \mathcal{D}_k is \mathcal{D} but with element k sampled independently.

Let $f : 2^{[k]} \rightarrow \mathbb{R}$ be a submodular function. To prove Submodular Dominance, we will show the following inequalities hold:

$$\mathbb{E}_{S \sim \mathcal{D}} [f(S)] \stackrel{\text{Claim 18}}{\geq} \mathbb{E}_{S \sim \mathcal{D}_k} [f(S)] \stackrel{\text{Claim 17}}{\geq} \mathbb{E}_{S \sim \mathbf{x}} [f(S)] . \quad (2)$$

► **Claim 17.** The second inequality of (2) holds, i.e., $\mathbb{E}_{S \sim \mathcal{D}_k} [f(S)] \geq \mathbb{E}_{S \sim \mathbf{x}} [f(S)]$.

Proof. Since k is independently sampled in both \mathcal{D}_k and \mathbf{x} , we can write

$$\begin{aligned} \mathbb{E}_{S \sim \mathcal{D}_k} [f(S)] &= \mathbb{E}_{S \sim \mathcal{D} \setminus k} [x_k \cdot f(S \cup k) + (1 - x_k) \cdot f(S)] \quad \text{and} \\ \mathbb{E}_{S \sim \mathbf{x}} [f(S)] &= \mathbb{E}_{S \sim \mathbf{x} \setminus k} [x_k \cdot f(S \cup k) + (1 - x_k) \cdot f(S)] . \end{aligned}$$

Convex combinations of submodular functions are submodular, $\mathcal{D} \setminus k$ is WNR by closure under projection (Proposition 16), and the marginals of $\mathcal{D} \setminus k$ are equal to the marginals of $\mathbf{x} \setminus k$. Therefore, by the induction hypothesis,

$$\mathbb{E}_{S \sim \mathcal{D} \setminus k} [x_k \cdot f(S \cup k) + (1 - x_k) \cdot f(S)] \geq \mathbb{E}_{S \sim \mathbf{x} \setminus k} [x_k \cdot f(S \cup k) + (1 - x_k) \cdot f(S)] ,$$

which implies $\mathbb{E}_{S \sim \mathcal{D}_k} [f(S)] \geq \mathbb{E}_{S \sim \mathbf{x}} [f(S)]$. ◁

► **Claim 18.** The first inequality of (2) holds, i.e., $\mathbb{E}_{S \sim \mathcal{D}} [f(S)] \geq \mathbb{E}_{S \sim \mathcal{D}_k} [f(S)]$.

Proof. Expanding expectations for \mathcal{D}_k (as in the proof of Claim 17) and moving it to LHS,

$$\mathbb{E}_{S \sim \mathcal{D}} [f(S) - x_k \cdot f(S \cup k) - (1 - x_k) \cdot f(S \setminus k)] \geq 0 .$$

⁶ This last property is useful in the setting of continuous random variables studied in [26] because properties closed under convolutions are extremely powerful. However, it is not so relevant in the discrete settings we study.

44:8 Submodular Dominance and Applications

Conditioning on k yields

$$x_k \cdot \mathbb{E}_{S \sim \mathcal{D}} [f(S \cup k) - x_k \cdot f(S \cup k) - (1 - x_k) \cdot f(S \setminus k) \mid k \in S] \\ + (1 - x_k) \cdot \mathbb{E}_{S \sim \mathcal{D}} [f(S \setminus k) - x_k \cdot f(S \cup k) - (1 - x_k) \cdot f(S \setminus k) \mid k \notin S] \geq 0 ,$$

which simplifies to

$$x_k(1 - x_k) \left(\mathbb{E}_{S \sim \mathcal{D}} [f(S \cup k) - f(S \setminus k) \mid k \in S] + \mathbb{E}_{S \sim \mathcal{D}} [f(S \setminus k) - f(S \cup k) \mid k \notin S] \right) \geq 0 .$$

Dividing out $x_k(1 - x_k)$ and moving the second term to RHS gives

$$\mathbb{E}_{S \sim \mathcal{D}} [f(S \cup k) - f(S \setminus k) \mid k \in S] \geq \mathbb{E}_{S \sim \mathcal{D}} [f(S \cup k) - f(S \setminus k) \mid k \notin S] . \quad (3)$$

Let $f_k(S) := f(S \cup k) - f(S \setminus k)$. f_k does not depend on k , and by the submodularity of f , $-f_k$ is a monotone function. Thus, (3) is directly implied by the WNR condition (1). \triangleleft

Claims 17 and 18 complete the proof of Theorem 3. \blacktriangleleft

The following proposition, which we prove in Appendix A.2, shows that WNR is not a necessary condition for Submodular Dominance.

► Proposition 19. *The distribution \mathcal{D} which samples uniformly from $\emptyset, \{1\}, \{2\}, \{1, 2\}, \{1, 3\}, \{2, 3\}$ satisfies Submodular Dominance, but \mathcal{D} violates WNR for $f(S) := \max(\mathbb{1}_{1 \in S}, \mathbb{1}_{2 \in S})$ and $i = 3$.*

3.2 NCD is a Necessary Condition

Since WNR is not equivalent to Submodular Dominance, we search for necessary conditions to better understand the relationship between negative dependence and Submodular Dominance.

► Theorem 5. *All distributions that satisfy Submodular Dominance are NCD.*

Proof. Let \mathcal{D} be a distribution over 2^U with marginals \mathbf{x} which satisfies Submodular Dominance. For any $T \subseteq U$, consider the functions

$$f_T(S) := 1 - \mathbb{1}_{T \subseteq S^c} \quad \text{and} \quad g_T(S) := |S \cap T| - \mathbb{1}_{T \subseteq S} .$$

Equivalently, f_T, g_T are the rank functions of the uniform matroids of rank 1 and $|T| - 1$ over the ground set T . The only fact about matroid rank functions we use here is that all matroid rank functions are submodular (though one can also easily check that f_T, g_T are submodular via the definition of submodularity). Since subtracting a linear function from a submodular function results in a submodular function, $-\mathbb{1}_{T \subseteq S}$ and $-\mathbb{1}_{T \subseteq S^c}$ are submodular functions. Thus, by Submodular Dominance we have

$$-\Pr_{S \sim \mathcal{D}} [T \subseteq S] = \mathbb{E}_{S \sim \mathcal{D}} [-\mathbb{1}_{T \subseteq S}] \geq \mathbb{E}_{S \sim \mathbf{x}} [-\mathbb{1}_{T \subseteq S}] = -\Pr_{S \sim \mathbf{x}} [T \subseteq S] \quad \text{and} \\ -\Pr_{S \sim \mathcal{D}} [T \subseteq S^c] = \mathbb{E}_{S \sim \mathcal{D}} [-\mathbb{1}_{T \subseteq S^c}] \geq \mathbb{E}_{S \sim \mathbf{x}} [-\mathbb{1}_{T \subseteq S^c}] = -\Pr_{S \sim \mathbf{x}} [T \subseteq S^c] .$$

Multiplying both sides by -1 yields the definition of NCD (Definition 4). \blacktriangleleft

The following two propositions, which we prove in Appendix A.2, show that NCD is not a sufficient condition for Submodular Dominance.

► **Proposition 20.** *The distribution \mathcal{D} over $2^{[4]}$ which chooses uniformly at random $i \in [4]$, then returns w.p. $1/2$ either i or $[4] \setminus i$, is NCD. However, \mathcal{D} violates Submodular Dominance for the submodular function $f(S) := \min(2, |S|)$.*

► **Proposition 21.** *The distribution \mathcal{D} over $2^{[8]}$ which chooses uniformly at random $i \in A := \{1, 2, 3, 4\}$ and $j \in B := \{5, 6, 7, 8\}$, then returns w.p. $1/2$ either $i \cup (B \setminus j)$ or $(A \setminus i) \cup j$, is NCD. However, \mathcal{D} violates Submodular Dominance for the submodular function $f(S) := \min(2, |S \cap A|)$.*

Thus, the class of distributions which satisfy Submodular Dominance is a strict subset of NCD distributions and a strict superset of WNR distributions. It is unclear whether the “right” answer will turn out to be a useful notion of negative dependence.

4 Applications to Submodular Prophet Inequalities

In SPI, we have *items* U , which are discrete random variables with disjoint images and arbitrary probability mass functions. We denote realizations of items as *elements*. WLOG, let the image of $i \in U$ be $\{ij : j \in [m]\}$, and let the realization of i be ij w.p. p_{ij} . Let $E := [n] \times [m]$ denote the set of elements. The distributions of each item are independent and known to us in advance.

We are given a set system $\mathcal{I} \subseteq 2^U$ and a submodular objective function $f : 2^E \rightarrow \mathbb{R}_{\geq 0}$. Notice that while the items are independent, the elements do not follow a product distribution. As we are optimizing over the element-space, this is a non-trivial complication.

Each item arrives one-by-one. When an item arrives, we learn its realization, and must choose whether to accept or reject it. The set of accepted items must be in \mathcal{I} , and the goal is to maximize f on the realizations of the accepted items. The arrival order is chosen by an *almighty adversary*, who knows in advance the outcomes of all randomness, such as the item realizations, the decisions of our algorithm, etc.

If there exists an α -competitive algorithm compared to the prophet, we say there is an α SPI. Rubinstein and Singla [35] proved $\Omega(1)$ SPIs over matroids, and Chekuri and Livanos [9] refined their analysis to obtain better constants, as well as results for a broader range of set systems. We further improve upon their approach using Submodular Dominance, obtaining results such as the first tight SPI for large rank uniform matroids.

4.1 Core Approach: SPI for Bernoulli Items

Before tackling the full problem, it is helpful to first consider a simplified version, *Bernoulli SPI*, where each item i is only a Bernoulli random variable, taking value 1 w.p. p_i and taking value 0 otherwise. Here, there is no notion of elements (or rather, elements are effectively synonymous with items), so we consider a submodular objective function $f : 2^U \rightarrow \mathbb{R}_{\geq 0}$. As in the full problem, we have a set system constraint $\mathcal{I} \subseteq 2^U$.

This is quite similar to the problem of submodular maximization from Section 5.1, but with a stochastic component (each item being usable only w.p. p_i) and an online component (items are revealed one-by-one). Therefore, it makes sense to borrow the high level approach of optimizing the multilinear extension F , then rounding the fractional solution. Since \mathcal{I} is a discrete constraint and F is a continuous function, the following relaxation is useful in offline submodular maximization:

► **Definition 22.** *For any downward-closed set system $\mathcal{I} \subseteq 2^U$, its polytope $\mathcal{P}_{\mathcal{I}} \subseteq [0, 1]^n$ is the convex hull of the indicator vectors representing the maximal sets of \mathcal{I} .*

44:10 Submodular Dominance and Applications

For Bernoulli SPI, we consider a modified polytope $\mathcal{P}'_{\mathcal{I}} := \{\mathbf{x} \in [0, 1]^n : \mathbf{x} \in \mathcal{P}_{\mathcal{I}}, x_i \leq p_i \forall i \in U\}$. Since the fractional solution corresponds to a distribution over \mathcal{I} , the additional constraint $x_i \leq p_i$ ensures that no item is included more often than it takes value 1. It turns out that we can efficiently optimize F over $\mathcal{P}'_{\mathcal{I}}$ under mild conditions.

As for rounding, we can use *online contention resolution schemes* (OCRS). OCRSs function in the following setting: we have a set system $\mathcal{I} \subseteq 2^U$ and a distribution \mathcal{D} over 2^U with marginals \mathbf{x} . Let items $i \in S$ for some $S \sim \mathcal{D}$ be called *active*. The items then arrive one-by-one in adversarial order. When item i arrives, we learn whether it is active, and if so, must decide to accept or reject it, subject to the set of accepted items being in \mathcal{I} . An OCRS $\pi_{\mathcal{I}, \mathcal{D}}$ is an algorithm that plays this game. The following notion is a way to measure the performance of an OCRS.

► **Definition 23** (*(b, c)-selectable OCRS*). For $b, c \in [0, 1]$, a set system $\mathcal{I} \subseteq 2^U$, and a distribution \mathcal{D} over 2^U with marginals $\mathbf{x} \in b \cdot \mathcal{P}_{\mathcal{I}}$, an OCRS $\pi_{\mathcal{I}, \mathcal{D}}$ is *(b, c)-selectable* if the probability of $\pi_{\mathcal{I}, \mathcal{D}}$ accepting i is at least $c \cdot x_i$ for all $i \in U$. If $b = 1$, we say $\pi_{\mathcal{I}, \mathcal{D}}$ is *c-selectable*.

Feldman, Svensson, and Zenklusen [19] obtained the following approximation result for rounding via *greedy OCRSs* (we omit the definition as it is not relevant).

► **Proposition 24** ([19]). For a set system $\mathcal{I} \subseteq 2^U$, a monotone submodular function $f : 2^U \rightarrow \mathbb{R}_{\geq 0}$, and $\mathbf{x} \in b \cdot \mathcal{P}_{\mathcal{I}}$, applying a *(b, c)-selectable greedy OCRS* to $S \sim \mathbf{x}$ obtains $T \in \mathcal{I}$ such that $\mathbb{E}_{S \sim \mathbf{x}}[f(T)] \geq c \cdot F(\mathbf{x})$. Further, the *greedy OCRS* can be efficiently modified such that even for non-monotone f , the *modified greedy OCRS* obtains $T \in \mathcal{I}$ such that $\mathbb{E}_{S \sim \mathbf{x}}[f(T)] \geq c/4 \cdot F(\mathbf{x})$.

In Bernoulli SPI, there is no notion of elements and we optimize over items. Thus, the distribution of active items is already a product distribution, and simply applying greedy OCRSs already yields approximation results using Proposition 24.

4.2 Generalizing to Arbitrary Discrete Random Variables

We now return to the full version of SPI. Again, let U be the set of items, $\mathcal{I} \subseteq 2^U$ be a downward-closed set system with a solvable polytope, $E := [n] \times [m]$ be the set of elements, $\mathbf{p} \in [0, 1]^{nm}$ be the element realization probabilities, and $f : 2^E \rightarrow \mathbb{R}_{\geq 0}$ be a submodular function. The first step is to compute a fractional solution. Chekuri and Livanos [9] define the polytope

$$\mathcal{P}''_{\mathcal{I}} := \{\mathbf{x} \in [0, 1]^{nm} : \exists \mathbf{z} \in \mathcal{P}_{\mathcal{I}} \text{ satisfying } \sum_j x_{ij} = z_i \forall i \in U, x_{ij} \leq p_{ij} \forall ij \in E\}.$$

Here, the summation constraint is a natural relaxation of $\mathcal{P}'_{\mathcal{I}}$ from the item-space to the element-space. Chekuri and Livanos prove a series of results⁷ which culminates in the following (note that we do not require monotonicity of f):

► **Proposition 25** ([9]). Let OPT be the expectation obtained by the prophet, $\mathcal{I} \subseteq 2^U$ be a set system with a solvable polytope, and $f : 2^E \rightarrow \mathbb{R}_{\geq 0}$ be a submodular function. Then for any fixed $\epsilon > 0$, we can efficiently compute $\mathbf{x} \in b \cdot \mathcal{P}''_{\mathcal{I}}$ such that $F(\mathbf{x}) \geq (1 - e^{-b} - \epsilon) \cdot \text{OPT}$.

⁷ See Section 3 of [9] for details, in particular, Claim 3.4, Theorem 1.3, and Remark 3.7.

It remains to round the fractional solution. While it is fairly straightforward to convert a (b, c) -selectable greedy OCRS for the item-space to a (b, c) -selectable greedy OCRS for the element-space (indeed, we do this in Algorithm 1), we cannot obtain approximation results directly from Proposition 24 like in the Bernoulli case because the distribution of elements is not a product distribution. Chekuri and Livanos handle this by incurring an additional loss of $e^{-b} - \epsilon$ to “mask” the elements under a product distribution. We save this factor by re-analyzing a simpler algorithm.

4.3 Improved Analysis

Let $\mathbf{x} \in b \cdot \mathcal{P}_{\mathcal{I}}''$ be the solution computed as per Proposition 25. Define $x_i := \sum_j x_{ij}$, and define $\vec{x} := (x_i : i \in U)$. By definition of $\mathcal{P}_{\mathcal{I}}''$ and the fact that $\mathbf{x} \in b \cdot \mathcal{P}_{\mathcal{I}}''$, we have that $\vec{x} \in b \cdot \mathcal{P}_{\mathcal{I}}$, so let $\pi_{\mathcal{I}, \vec{x}}$ be an efficient (b, c) -selectable greedy OCRS.

We first consider monotone f . Our rounding algorithm is almost identical to [9, Algorithm 1], removing some steps that our improved analysis demonstrates to be unnecessary.

■ **Algorithm 1** MONOTONE ROUNDING $(U, E, \mathbf{p}, f, \mathbf{x}, \pi_{\mathcal{I}, \vec{x}})$.

```

TALG = ∅
for  $t \leftarrow 1$  to  $n$  do
  Let  $i \in U$  be the item that arrives on day  $t$ 
  Let  $ij \in E$  be the realization of  $i$ 
  With probability  $x_{ij}/p_{ij}$ , reveal active  $i$  to  $\pi_{\mathcal{I}, \vec{x}}$ , otherwise reveal inactive  $i$  to  $\pi_{\mathcal{I}, \vec{x}}$ 
  if  $\pi_{\mathcal{I}, \vec{x}}$  accepts  $i$  then
    |  $T_{\text{ALG}} \leftarrow T_{\text{ALG}} \cup \{ij\}$ 
  end
end
Return  $T_{\text{ALG}}$ 

```

Denote element ij as *active* when ij is the realization of i , and Algorithm 1 reveals active i to $\pi_{\mathcal{I}, \vec{x}}$. Since the elements do not follow a product distribution, we cannot apply Proposition 24 even though the algorithm acts like a greedy OCRS. However, we provide a new analysis which states that a c -approximation for product distributions implies a c -approximation for the following wider class of distributions.

► **Definition 26.** A product of singletons distribution over 2^E with marginals $\mathbf{x} \in [0, 1]^{nm}$ such that $\sum_j x_{ij} \leq 1$ for all i is a distribution which independently samples 0 or 1 elements from each set $\{ij : j \in [m]\}$ according to the marginals \mathbf{x} .

It is not difficult to see that the active elements follow a product of singletons distribution with marginals \mathbf{x} . The following lemma, which we prove in Appendix A.3, draws a connection between product of singletons distributions and product distributions.

► **Lemma 27.** Let \mathcal{D} be a product of singletons distribution over 2^E with marginals $\mathbf{x} \in [0, 1]^{nm}$. Let $x_i := \sum_j x_{ij}$, and let $\vec{x} := (x_i : i \in U)$. For any $\mathbf{u} \in [m]^n$, let $E_{\mathbf{u}} := \{iu_i : i \in U\}$ and let $\mathcal{D}_{\mathbf{u}}$ be a product distribution over $2^{E_{\mathbf{u}}}$ with marginals \vec{x} . Then for any $g : 2^E \rightarrow \mathbb{R}$,

$$\mathbb{E}_{S \sim \mathcal{D}}[g(S)] = \sum_{\mathbf{u} \in [m]^n} \left(\mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}}[g(S)] \cdot \prod_{i \in U} \frac{x_{iu_i}}{x_i} \right). \quad (4)$$

In simpler terms, $\mathcal{D}_{\mathbf{u}}$ is the distribution which samples $S \sim \mathcal{D}$, then replaces each element $ij \in S$ with the element iu_i . Lemma 27 states that any product of singletons distribution with marginals \mathbf{x} can be written as a convex combination of product distributions with marginals \vec{x} .

44:12 Submodular Dominance and Applications

► **Lemma 28.** For monotone f , Algorithm 1 returns T_{ALG} such that $\mathbb{E}[f(T_{\text{ALG}})] \geq c \cdot F(\mathbf{x})$.

Proof. Let \mathcal{D} be the distribution of active elements. While the adversary sees the item realizations and which items Algorithm 1 will reveal as active to $\pi_{\mathcal{I}, \vec{x}}$, the adversary cannot influence \mathcal{D} because the decisions to reveal active i do not depend on the item ordering.

Therefore, it is valid for us to “partition” the outcomes of randomness contributing to \mathcal{D} . Since \mathcal{D} is a product of singletons distribution with marginals \mathbf{x} , Lemma 27 tells us that there exists a partition such that each part is a product distribution $\mathcal{D}_{\mathbf{u}}$ over $2^{E_{\mathbf{u}}}$ with marginals \vec{x} . We fix some $\mathbf{u} \in [m]^n$ and analyze the performance of Algorithm 1 on the subset of randomness corresponding to the distribution $\mathcal{D}_{\mathbf{u}}$.

Since $\mathcal{D}_{\mathbf{u}}$ is a product distribution, $\vec{x} \in b \cdot \mathcal{P}_{\mathcal{I}}$, and the algorithm copies the acceptances of $\pi_{\mathcal{I}, \vec{x}}$, Algorithm 1 acts exactly like a (b, c) -selectable greedy OCRS over $\mathcal{D}_{\mathbf{u}}$. Most importantly, $\mathcal{D}_{\mathbf{u}}$ being product distribution means we can apply Proposition 24 to get

$$\mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}} [f(T_{\text{ALG}})] \geq c \cdot \mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}} [f(S)] .$$

T_{ALG} is implicitly a randomized function of S , so we can rewrite LHS as

$$\mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}} \left[\mathbb{E}[f(T_{\text{ALG}}) \mid S = S'] \right] \geq c \cdot \mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}} [f(S)] ,$$

where the inner expectation is taken over the possible randomization of the underlying greedy OCRS $\pi_{\mathcal{I}, \vec{x}}$ and the adversarial ordering of the items. As this holds for any \mathbf{u} , weighting the inequality and summing over all $\mathbf{u} \in [m]^n$ yields

$$\sum_{\mathbf{u} \in [m]^n} \left(\mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}} \left[\mathbb{E}[f(T_{\text{ALG}}) \mid S = S'] \right] \cdot \prod_{i \in U} \frac{x_{iu_i}}{x_i} \right) \geq \sum_{\mathbf{u} \in [m]^n} \left(c \cdot \mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}} [f(S)] \cdot \prod_{i \in U} \frac{x_{iu_i}}{x_i} \right) .$$

Factoring out the c on RHS, then applying Lemma 27 to the functions $\mathbb{E}[f(T_{\text{ALG}}) \mid S = S']$ and $f(S)$ simplifies to $\mathbb{E}_{S \sim \mathcal{D}} [f(T_{\text{ALG}})] \geq c \cdot \mathbb{E}_{S \sim \mathcal{D}} [f(S)]$.

A distribution which samples only sets of size 0 or 1 is WNR because conditioning on inclusion of an element excludes all other elements. Further, products of WNR distributions are WNR (Proposition 16). Thus, product of singletons distributions are WNR, and applying Submodular Dominance (Theorem 3) on \mathcal{D} gives the following and completes the proof:

$$\mathbb{E}_{S \sim \mathcal{D}} [f(T_{\text{ALG}})] \geq c \cdot \mathbb{E}_{S \sim \mathcal{D}} [f(S)] \geq c \cdot F(\mathbf{x}) . \quad \blacktriangleleft$$

► **Remark 29.** For general f , we can replace the greedy OCRS $\pi_{\mathcal{I}, \vec{x}}$ by its efficient modification mentioned in Proposition 24, then just repeat the proof of Lemma 28. We lose an additional factor of $1/4$ when we invoke Proposition 24 on the modified greedy OCRS, which gives us a $c/4$ -approximation algorithm when f is not monotone.

► **Theorem 6 (Submodular Prophet Inequalities).** For fixed $\epsilon > 0$, if a set system $\mathcal{I} \subseteq 2^U$ has a solvable polytope and an efficient (b, c) -selectable greedy OCRS for product distributions:

- There is an efficient $c \cdot (1 - e^{-b} - \epsilon)$ SPI for monotone non-negative submodular functions.
 - There is an efficient $c/4 \cdot (1 - e^{-b} - \epsilon)$ SPI for general non-negative submodular functions.
- Combining with known greedy OCRSSs, this implies efficient SPIs as given in Table 1.

Proof. Combining Proposition 25 and Lemma 28 gives us a $c \cdot (1 - e^{-b} - \epsilon)$ SPI for monotone non-negative submodular functions, and, as noted in Remark 29, a similar argument gives us a $c/4 \cdot (1 - e^{-b} - \epsilon)$ SPI for general non-negative submodular functions. From [19], we can efficiently construct greedy OCRSSs satisfying the following properties:

- $(b, 1 - b)$ -selectable over matroids, for $b \in [0, 1]$.
- (b, e^{-2b}) -selectable over matchings, for $b \in [0, 1]$.
- $(b, \frac{1-2b}{2-2b})$ -selectable over knapsacks, for $b \in [0, 1/2]$.
- $(1 - o(1))$ -selectable over uniform matroids of rank $k \rightarrow \infty$ [9].

To obtain the results in Table 1, we simply choose b which maximizes $c \cdot (1 - e^{-b} - \epsilon)$. ◀

5 Applications to Rounding

A common problem setting is optimization constrained to some feasible set system \mathcal{I} .

► **Definition 30.** For a downward-closed set system $\mathcal{I} \subseteq 2^U$, its polytope $\mathcal{P}_{\mathcal{I}} \subseteq [0, 1]^n$ is the convex hull of the indicator vectors representing the maximal sets of \mathcal{I} .

Under mild conditions, we can efficiently optimize over the polytope, then round the fractional solution $\mathbf{x} \in \mathcal{P}_{\mathcal{I}}$ to an integral solution $S \in \mathcal{I}$. It is natural to think of \mathbf{x} as a distribution over \mathcal{I} with those marginals. If these distributions exhibit certain properties, then sampling can be an effective rounding technique. We give results for set systems which satisfy the following property:

► **Definition 31.** A set system $\mathcal{I} \subseteq 2^U$ admits WNR distributions if for any $\mathbf{x} \in \mathcal{P}_{\mathcal{I}}$, there exists a WNR distribution over \mathcal{I} with marginals \mathbf{x} . If we can efficiently sample from these distributions, we say \mathcal{I} efficiently admits WNR distributions.

5.1 Submodular Maximization

For a set system $\mathcal{I} \subseteq 2^U$ and a monotone submodular function $f : 2^U \rightarrow \mathbb{R}_{\geq 0}$, a classical optimization problem is to efficiently find $T \in \mathcal{I}$ such that $f(T)$ is a good approximation of $\max_{S \in \mathcal{I}} f(S)$. We start by optimizing over $\mathcal{P}_{\mathcal{I}}$.

► **Proposition 32** ([39]). For any set system \mathcal{I} with a solvable polytope, we can efficiently compute $\mathbf{x} \in \mathcal{P}_{\mathcal{I}}$ such that $F(\mathbf{x}) \geq (1 - 1/e - o(1)) \cdot \max_{S \in \mathcal{I}} f(S)$.

Now, we want to round \mathbf{x} to an integral solution with at least value $F(\mathbf{x})$. Pipage [6] and randomized swap rounding [10] achieve this for matroid polytopes, but it is unclear how to extend it. Submodular Dominance gives new approaches for submodular maximization.

► **Theorem 7** (Submodular Maximization). Let $f : 2^U \rightarrow \mathbb{R}_{\geq 0}$ be a monotone submodular function. If a downward-closed set system $\mathcal{I} \subseteq 2^U$ has a solvable polytope and efficiently admits WNR distributions, there exists an efficient algorithm that returns $T \in \mathcal{I}$ such that $\mathbb{E}[f(T)] \geq (1 - 1/e - o(1)) \cdot \max_{S \in \mathcal{I}} f(S)$.

Proof. We compute $\mathbf{x} \in \mathcal{P}_{\mathcal{I}}$ as per Proposition 32, then sample from a WNR distribution over \mathcal{I} with marginals \mathbf{x} . By Theorem 3, this returns $T \in \mathcal{I}$ such that $\mathbb{E}[f(T)] \geq F(\mathbf{x}) \geq (1 - 1/e - o(1)) \cdot \max_{S \in \mathcal{I}} f(S)$. ◀

5.2 Adaptivity Gaps for Stochastic Probing

Stochastic Probing is a generalization of submodular maximization with randomized inputs. Elements are replaced by *items*, and we *probe* items to learn their realizations. The goal is to maximize the expectation of a function over the realizations of probed items. We consider a simple version of the problem where we have a monotone submodular function $f : 2^U \rightarrow \mathbb{R}_{\geq 0}$ and each item X_i contains element i independently w.p. p_i and is empty otherwise.

As probing reveals information, we differentiate between *adaptive algorithms*, which behave differently conditioned on the realizations of probed items, and *non-adaptive algorithms*.

► **Definition 33.** *The adaptivity gap is the ratio between the expectations obtained by the optimal adaptive algorithm and optimal non-adaptive algorithm.*

Asadpour and Nazerzadeh [2] give a tight result that the adaptivity gap for stochastic probing subject to a matroid constraint is $\frac{e}{e-1}$. The approach is to first define an auxiliary function f' , where $f'(S)$ is the expectation of f upon probing items $\{X_i : i \in S\}$. It turns out that the multilinear extension F' of f' satisfies the property that $\max_{\mathbf{x} \in \mathcal{P}_{\mathcal{I}}} F'(\mathbf{x})$ is a $(1 - 1/e)$ -approximation of the expectation obtained by the optimal adaptive algorithm.⁸

With this approximation result, the idea is to use \mathbf{x} to design a non-adaptive algorithm. Simply probing each item w.p. x_i may violate the matroid constraint, so Asadpour and Nazerzadeh design non-adaptive algorithms using pipage rounding. We go beyond matroids by designing non-adaptive algorithms using WNR distributions.

► **Theorem 8 (Stochastic Probing).** *For a downward-closed set system \mathcal{I} that admits WNR distributions, the adaptivity gap for Stochastic Probing is upper-bounded by $\frac{e}{e-1}$.*

Proof. Our analysis follows that of [2] until we obtain $\operatorname{argmax}_{\mathbf{x} \in \mathcal{P}_{\mathcal{I}}} F'(\mathbf{x})$. As \mathcal{I} admits WNR distributions, there exists a WNR distribution \mathcal{D} over \mathcal{I} with marginals \mathbf{x} . By Theorem 3, we have $\mathbb{E}_{S \sim \mathcal{D}}[f'(S)] \geq F'(\mathbf{x})$. Therefore, the non-adaptive algorithm which samples $S \sim \mathcal{D}$ and probes $\{X_i : i \in S\}$ obtains at least $F'(\mathbf{x})$ in expectation. No adaptive algorithm can obtain expectation greater than $\frac{e}{e-1} \cdot F'(\mathbf{x})$, so $\frac{e}{e-1}$ upper-bounds the adaptivity gap. ◀

5.3 Contention Resolution Schemes

► **Definition 34 (CRS).** *A contention resolution scheme (CRS) for a set system $\mathcal{I} \subseteq 2^U$ and a distribution \mathcal{D} over 2^U with marginals \mathbf{x} is a (possibly randomized) mapping $\pi_{\mathcal{I}, \mathcal{D}} : 2^U \rightarrow \mathcal{I}$ such that for all $S \subseteq U$, we have $\pi_{\mathcal{I}, \mathcal{D}}(S) \subseteq S$.*

Contention resolution schemes have applications to submodular maximization as a rounding technique. The following is the simplest measure of performance for a CRS.

► **Definition 35 (c -selectable CRS).** *For $c \in [0, 1]$, a set system $\mathcal{I} \subseteq 2^U$, and a distribution \mathcal{D} over 2^U with marginals $\mathbf{x} \in \mathcal{P}_{\mathcal{I}}$, a CRS $\pi_{\mathcal{I}, \mathcal{D}}$ is c -selectable if $\Pr_{S \sim \mathcal{D}}[i \in \pi_{\mathcal{I}, \mathcal{D}}(S)] \geq c \cdot x_i$ for all $i \in U$.*

In submodular maximization, rounding fractional solutions is closely related to the multilinear extension, so study of CRSs is primarily centered around product distributions. However, as Dughmi [14, 15] recently showed, CRSs over non-product distributions have applications in settings such as the Matroid Secretary Problem. We use Submodular Dominance to extend a selectability result to WNR distributions, which provides a direction by which other CRS results may be generalized to correlated distributions.

► **Theorem 9 (Contention Resolution Schemes).** *For a matroid \mathcal{M} , there exists a $(1 - 1/e)$ -selectable CRS for any WNR distribution with marginals $\mathbf{x} \in \mathcal{P}_{\mathcal{M}}$.*

Proof. [12] demonstrated this result for product distributions via strong LP duality. Following the same idea, Dughmi [14] reduced this result to proving Submodular Dominance:

⁸ We omit many of the finer details because our result does not alter this part of the analysis. Section 3 of [2] covers this in depth.

► **Proposition 36** ([14]). *For a matroid \mathcal{M} and a distribution \mathcal{D} over 2^U with marginals $\mathbf{x} \in \mathcal{P}_{\mathcal{M}}$, there exists a $(1 - 1/e)$ -selectable CRS if every submodular function $f : 2^U \rightarrow \mathbb{R}$ satisfies $\mathbb{E}_{S \sim \mathcal{D}}[f(S)] \geq \mathbb{E}_{S \sim \mathbf{x}}[f(S)]$.*

Theorem 9 follows directly from Proposition 36 and Theorem 3. ◀

6 Conclusion and Open Questions

In this paper, we explore Submodular Dominance and its applications. In the process, we introduce a notion of negative dependence that we refer to as Weak Negative Regression (WNR), which is a natural generalization of both Negative Association (NA) and Negative Regression (NR) and may be of use in other applications. We prove that WNR distributions satisfy Submodular Dominance, and that all distributions satisfying Submodular Dominance also satisfy Negative Cylinder Dependence (NCD). Finally, we give a variety of applications for Submodular Dominance, improving the best known submodular prophet inequalities, developing new rounding techniques, and generalizing results for contention resolution schemes to negatively dependent distributions.

Sampling for More General Set Systems. Although our results for negatively distributions satisfying Submodular Dominance already have several applications, their usage could be broadened further by finding new techniques to generate negatively dependent distributions. An interesting future direction is to design algorithms to sample from negatively dependent distributions for more general set systems. For example, can we efficiently sample from a WNR/NA/NR distribution for any marginals in a given matroid polytope? We remark that [33] claimed such a result for NA distributions, but later, a gap in their proof was found. Max-entropy distributions over matroids are also not negatively dependent in general, as it is known that there exist matroids for which the uniform distribution (which is entropy-maximizing without constrained marginals) exhibits positive correlations [25].

Approximate Submodular Dominance. While we showed that NCD distributions do not always satisfy Submodular Dominance, one question is whether these weaker notions of negative dependence obtain constant-factor approximation variants of Submodular Dominance; that is, for a non-negative submodular function f , what distributions satisfy $\mathbb{E}_{S \sim \mathcal{D}}[f(S)] \geq O(1) \cdot F(\mathbf{x})$? What about for monotone f ? Another direction is to generalize Submodular Dominance to a larger class of functions. XOS functions are functions that can be expressed as the maximum of a collection of linear functions, and are a strict superset of submodular functions. While no non-product distribution satisfies “XOS Dominance” (consider $\max(X_1 + X_2, 1)$ and $\max(X_1, X_2)$, which are both XOS; the former decreases in expectation if X_1 and X_2 are negatively correlated, the latter if X_1 and X_2 are positively correlated), we might similarly ask if approximate versions hold for XOS functions.

Concentration Inequalities. Another important direction is understanding concentration inequalities for negatively dependent distributions. Submodular Dominance demonstrates that the expectation of negatively dependent distributions behaves favorably compared to product distributions, but we may also be interested in whether these distributions are concentrated around their mean. We know dimension-dependent concentration inequalities for arbitrary Lipschitz functions over NR distributions [21]. Proving dimension-independent concentration inequalities for submodular functions is an interesting future direction.

References

- 1 Saeed Alaei. Bayesian combinatorial auctions: Expanding single buyer mechanisms to many buyers. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 512–521, 2011.
- 2 Arash Asadpour and Hamid Nazerzadeh. Maximizing stochastic monotone submodular functions. *Manag. Sci.*, 62(8):2374–2391, 2016. doi:10.1287/mnsc.2015.2254.
- 3 Francis R. Bach. Learning with submodular functions: A convex optimization perspective. *Found. Trends Mach. Learn.*, 6(2-3):145–373, 2013.
- 4 Nikhil Bansal, Nitish Korula, Viswanath Nagarajan, and Aravind Srinivasan. Solving packing integer programs via randomized rounding with alterations. *Theory Comput.*, 8(1):533–565, 2012. doi:10.4086/toc.2012.v008a024.
- 5 Domagoj Bradac, Sahil Singla, and Goran Zuzic. (Near) optimal adaptivity gaps for stochastic multi-value probing. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA*, pages 49:1–49:21, 2019.
- 6 Gruia Călinescu, Chandra Chekuri, Martin Pál, and Jan Vondrák. Maximizing a submodular set function subject to a matroid constraint (extended abstract). In *Integer Programming and Combinatorial Optimization, 12th International IPCO Conference, Ithaca, NY, USA, June 25-27, 2007, Proceedings*, pages 182–196, 2007.
- 7 Gruia Calinescu, Chandra Chekuri, Martin Pal, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011.
- 8 Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 311–320, 2010.
- 9 Chandra Chekuri and Vasilis Livanos. On submodular prophet inequalities and correlation gap. In Ioannis Caragiannis and Kristoffer Arnsfelt Hansen, editors, *Algorithmic Game Theory - 14th International Symposium, SAGT 2021, Aarhus, Denmark, September 21-24, 2021, Proceedings*, volume 12885 of *Lecture Notes in Computer Science*, page 410. Springer, 2021. URL: <https://link.springer.com/content/pdf/bbm%3A978-3-030-85947-3%2F1.pdf>.
- 10 Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Dependent randomized rounding for matroid polytopes and applications. *arXiv preprint arXiv:0909.4348*, 2009.
- 11 Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Dependent randomized rounding via exchange properties of combinatorial structures. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 575–584, 2010.
- 12 Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Submodular function maximization via the multilinear relaxation and contention resolution schemes. *SIAM J. Comput.*, 43(6):1831–1879, 2014.
- 13 Tasos Christofides and Eutichia Vaggelatou. A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis*, 88:138–151, January 2004. doi:10.1016/S0047-259X(03)00064-2.
- 14 Shaddin Dughmi. The outer limits of contention resolution on matroids and connections to the secretary problem. In *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, pages 42:1–42:18, 2020.
- 15 Shaddin Dughmi. Matroid secretary is equivalent to contention resolution. In *13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 - February 3, 2022, Berkeley, CA, USA*, pages 58:1–58:23, 2022.

- 16 Alina Ene and Huy L. Nguyen. Constrained submodular maximization: Beyond $1/e$. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 248–257, 2016.
- 17 Hossein Esfandiari, Amin Karbasi, and Vahab S. Mirrokni. Adaptivity in adaptive submodularity. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT 2021, 15-19 August 2021, Boulder, Colorado, USA*, volume 134 of *Proceedings of Machine Learning Research*, pages 1823–1846. PMLR, 2021. URL: <http://proceedings.mlr.press/v134/esfandiari21a.html>.
- 18 Moran Feldman, Joseph Naor, and Roy Schwartz. A unified continuous greedy algorithm for submodular maximization. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 570–579, 2011.
- 19 Moran Feldman, Ola Svensson, and Rico Zenklusen. Online contention resolution schemes. In Robert Krauthgamer, editor, *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1014–1033. SIAM, 2016. doi:10.1137/1.9781611974331.ch72.
- 20 Satoru Fujishige. *Submodular functions and optimization*. Elsevier, 2005.
- 21 Kevin Garbe and Jan Vondrak. Concentration of lipschitz functions of negatively dependent variables. *arXiv preprint arXiv:1804.10084*, 2018.
- 22 Anupam Gupta and Viswanath Nagarajan. A stochastic probing problem with applications. In *Integer Programming and Combinatorial Optimization - 16th International Conference, IPCO 2013, Valparaíso, Chile, March 18-20, 2013. Proceedings*, pages 205–216, 2013.
- 23 Anupam Gupta, Viswanath Nagarajan, and Sahil Singla. Algorithms and adaptivity gaps for stochastic probing. In *Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1731–1747, 2016.
- 24 Mohammad Taghi Hajiaghayi, Robert D. Kleinberg, and Tuomas Sandholm. Automated online mechanism design and prophet inequalities. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada*, pages 58–65, 2007.
- 25 Mark Jerrum. Two remarks concerning balanced matroids. *Comb.*, 26(6):733–742, 2006. doi:10.1007/s00493-006-0039-5.
- 26 Kumar Joag-Dev and Frank Proschan. Negative association of random variables with applications. *The Annals of Statistics*, 11(1):286–295, 1983. doi:10.1214/aos/1176346079.
- 27 Robert Kleinberg and S. Matthew Weinberg. Matroid prophet inequalities. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, page 123–136, 2012.
- 28 Ulrich Krengel and Louis Sucheston. Semiamarts and finite values. *Bull. Am. Math. Soc.*, 1977.
- 29 Ulrich Krengel and Louis Sucheston. On semiamarts, amarts, and processes with finite value. *Advances in Prob.*, 4:197–266, 1978.
- 30 George L Nemhauser and Laurence A Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of operations research*, 3(3):177–188, 1978.
- 31 Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007.
- 32 Robin Pemantle. Towards a theory of negative dependence. *Journal of Mathematical Physics*, 41(3):1371–1390, 2000. doi:10.1063/1.533200.
- 33 Yuval Peres, Mohit Singh, and Nisheeth K. Vishnoi. Random walks in polytopes and negative dependence. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA*, pages 50:1–50:10, 2017.
- 34 Aviad Rubinfeld. Beyond matroids: secretary problem and prophet inequality with general constraints. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 324–332, 2016.

- 35 Aviad Rubinfeld and Sahil Singla. Combinatorial prophet inequalities. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 1671–1687, 2017.
- 36 Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.
- 37 Qi-Man Shao. A comparison theorem on moment inequalities between negatively associated and independent random variables. *Journal of Theoretical Probability*, 13(2):343–356, 2000.
- 38 Jan Vondrák. *Submodularity in combinatorial optimization*. PhD thesis, Univerzita Karlova, Matematicko-fyzikální fakulta, 2007.
- 39 Jan Vondrák. Optimal approximation for the submodular welfare problem in the value oracle model. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 67–74, 2008.

A Missing Proofs

A.1 Weak Negative Regression Proofs

► **Proposition 13.** *NA and NR imply WNR, and WNR implies NCD, but the reverse implications do not hold. In other words, the union of NA and NR distributions is a strict subset of WNR distributions, which is a strict subset of NCD distributions.*

Proof. We proved in Section 2 that the union of NA and NR distributions is a subset of WNR, and that WNR distributions are a subset of NCD distributions. Appendix A.2 provides example distributions demonstrating strict containment of WNR in NCD.

To show strict containment of NA and NR in WNR, we consider the following distribution \mathcal{D} , which was given by Joag-Dev and Proschan [26].⁹ We treat \mathcal{D} as a distribution over Bernoulli random variables $\mathbf{X} = (X_1, X_2, X_3, X_4)$ to simplify notation.

■ **Table 2** A distribution which is WNR, but not NA or NR.

\mathcal{D}	(X_1, X_2)					
		(0, 0)	(0, 1)	(1, 0)	(1, 1)	Marginal
(X_3, X_4)	(0, 0)	0.0577	0.0623	0.0623	0.0577	0.24
	(0, 1)	0.0623	0.0677	0.0677	0.0623	0.26
	(1, 0)	0.0623	0.0677	0.0677	0.0623	0.26
	(1, 1)	0.0577	0.0623	0.0623	0.0577	0.24
	Marginal	0.24	0.26	0.26	0.24	

\mathcal{D} violates NA because $\text{Cov}_{\mathbf{X} \sim \mathcal{D}}[X_1 X_2, X_3 X_4] > 0$, and \mathcal{D} violates NR because the conditional expectation $\mathbb{E}_{\mathbf{X} \sim \mathcal{D}}[X_1 X_2 \mid X_3 = 1, X_4]$ is increasing in X_4 .

By observing that the value in column 2 is larger than in column 4 for any row of Table 2, we see that for any $x_3, x_4 \in \{0, 1\}$, the conditional expectation $\mathbb{E}_{\mathbf{X} \sim \mathcal{D}}[X_2 \mid X_3 = x_3, X_4 = x_4, X_1]$ is decreasing in X_1 . Therefore, we can convert the distribution \mathcal{D} conditioned on $X_1 = 0$ into the distribution \mathcal{D} conditioned on $X_1 = 1$ by only transferring probability mass “downwards,” which cannot increase the expectation of a monotone function $f : \{0, 1\}^4 \rightarrow \mathbb{R}$. Thus, for any such function which does not depend on X_1 ,

⁹ They studied NA distributions over continuous random variables, and gave this distribution as an example that Negative Orthant Dependence (this is equivalent to NCD for Bernoulli random variables) does not imply NA.

$$\mathbb{E}_{\mathbf{X} \sim \mathcal{D}} [f(\mathbf{X}) \mid X_1 = 1] \leq \mathbb{E}_{\mathbf{X} \sim \mathcal{D}} [f(\mathbf{X}) \mid X_1 = 0] ,$$

which is the WNR condition. Since X_1, X_2 and $(X_1, X_2), (X_3, X_4)$ are both exchangeable, we can repeat this analysis for all X_i . Thus, \mathcal{D} is WNR, but neither NA nor NR. ◀

► **Proposition 16.** *WNR is closed both under projection and under products.*

Proof. Closure under projection follows trivially because the WNR condition (1) is satisfied for all monotone functions, and monotone functions restricted to a subset of elements are still monotone.

For closure under products, let \mathcal{A} and \mathcal{B} be WNR distributions over 2^A and 2^B for disjoint A, B , and let \mathcal{D} be their product. WLOG, fix $i \in A$ and a monotone function $f : 2^{A \cup B} \rightarrow \mathbb{R}$. Since \mathcal{A} and \mathcal{B} are independent, $\mathbb{E}_{S \sim \mathcal{D}} [f(S \setminus i) \mid S \cap B = T] = \mathbb{E}_{S \sim \mathcal{A}} [f((S \setminus i) \cup T)]$. Since \mathcal{A} is WNR and $f(S \cup T)$ is still a monotone function,

$$\mathbb{E}_{S \sim \mathcal{D}} [f(S \setminus i) \mid S \cap B = T, i \in S] \leq \mathbb{E}_{S \sim \mathcal{D}} [f(S \setminus i) \mid S \cap B = T, i \notin S] .$$

Taking expectations over $S \cap B$ gives $\mathbb{E}_{S \sim \mathcal{D}} [f(S \setminus i) \mid i \in S] \leq \mathbb{E}_{S \sim \mathcal{D}} [f(S \setminus i) \mid i \notin S]$, completing the proof. ◀

A.2 Submodular Dominance Example Distributions

► **Proposition 19.** *The distribution \mathcal{D} which samples uniformly from $\emptyset, \{1\}, \{2\}, \{1, 2\}, \{1, 3\}, \{2, 3\}$ satisfies Submodular Dominance, but \mathcal{D} violates WNR for $f(S) := \max(\mathbb{1}_{1 \in S}, \mathbb{1}_{2 \in S})$ and $i = 3$.*

Proof. Using the definition of \mathcal{D}_k from the proof of Theorem 3, notice that \mathcal{D}_1 is a product distribution. Therefore, we only need the analysis in Claim 18 to follow, which only requires that the WNR condition (1) holds for $i = 1$. \mathcal{D} conditioned on $1 \in S$ samples $\emptyset, \{2\}$, and $\{3\}$ w.p. $1/3$, and \mathcal{D} conditioned on $1 \notin S$ samples $\emptyset, \{2\}$, and $\{2, 3\}$ w.p. $1/3$, which cannot obtain lower expectation for a monotone function, so the WNR condition holds for $i = 1$ and \mathcal{D} satisfies Submodular Dominance.

\mathcal{D} violates WNR for f and $i = 3$ because \mathcal{D} conditioned on $3 \in S$ always samples either 1 or 2, whereas \mathcal{D} conditioned on $3 \notin S$ can sample \emptyset . Thus, there exist distributions which satisfy Submodular Dominance but violate WNR. ◀

► **Proposition 20.** *The distribution \mathcal{D} over $2^{[4]}$ which chooses uniformly at random $i \in [4]$, then returns w.p. $1/2$ either i or $[4] \setminus i$, is NCD. However, \mathcal{D} violates Submodular Dominance for the submodular function $f(S) := \min(2, |S|)$.*

Proof. Notice that \mathcal{D} is identical under permutations of elements. Further, because i or $[4] \setminus i$ is returned with equal probability, we have the property that for any $T \subseteq [4]$,

$$\Pr_{S \sim \mathcal{D}} [T \subseteq S] = \Pr_{S \sim \mathcal{D}} [T \subseteq S^c] .$$

Therefore, it is sufficient to show that $\Pr_{S \sim \mathcal{D}} [1, 2 \in S] \leq 1/4$, $\Pr_{S \sim \mathcal{D}} [1, 2, 3 \in S] \leq 1/8$, and $\Pr_{S \sim \mathcal{D}} [1, 2, 3, 4 \in S] \leq 1/16$ to prove \mathcal{D} is NCD.

- For $1, 2 \in S$, we need to choose $i = 3, 4$, then return $[4] \setminus i$. This occurs w.p. $1/2 \cdot 1/2 = 1/4$.
- For $1, 2, 3 \in S$, we need to choose $i = 4$, then return $[4] \setminus i$. This occurs w.p. $1/4 \cdot 1/2 = 1/8$.
- There is no way for $1, 2, 3, 4 \in S$.

44:20 Submodular Dominance and Applications

Thus, \mathcal{D} is NCD. f is a matroid rank function, so it is submodular. Letting \mathbf{x} be the marginals of \mathcal{D} , a simple expected value computation shows that $\mathbb{E}_{S \sim \mathcal{D}}[f(S)] = 12/8 < 13/8 = F(\mathbf{x})$, so \mathcal{D} violates Submodular Dominance. \blacktriangleleft

► **Proposition 21.** *The distribution \mathcal{D} over $2^{[8]}$ which chooses uniformly at random $i \in A := \{1, 2, 3, 4\}$ and $j \in B := \{5, 6, 7, 8\}$, then returns w.p. $1/2$ either $i \cup (B \setminus j)$ or $(A \setminus i) \cup j$, is NCD. However, \mathcal{D} violates Submodular Dominance for the submodular function $f(S) := \min(2, |S \cap A|)$.*

Proof. This example extends the previous example to a homogeneous distribution. Similar to the previous example, \mathcal{D} is closed under permutations of A , permutations of B , and swaps of A with B . Since we already showed any $T \subseteq A$ satisfies the NCD condition, it is sufficient to show that $\Pr_{S \sim \mathcal{D}}[1, 5 \in S] \leq 1/4$, $\Pr_{S \sim \mathcal{D}}[1, 2, 5 \in S] \leq 1/8$, $\Pr_{S \sim \mathcal{D}}[1, 2, 5, 6 \in S] \leq 1/16$, and $\Pr_{S \sim \mathcal{D}}[1, 2, 3, 5 \in S] \leq 1/16$ (since only sets of size 4 are drawn, if $|T| > 4$ it automatically satisfies the NCD condition).

- For $1, 5 \in S$, we need to choose $i = 1$ and $j = 6, 7, 8$, then return $i \cup (B \setminus j)$, or choose $i = 2, 3, 4$ and $j = 5$, then return $(A \setminus i) \cup j$. This occurs w.p. $2 \cdot 1/4 \cdot 3/4 \cdot 1/2 = 3/16 \leq 1/4$.
- For $1, 2, 5 \in S$, we need to choose $i = 3, 4$ and $j = 5$, then return $(A \setminus i) \cup j$. This occurs w.p. $1/2 \cdot 1/4 \cdot 1/2 = 1/16 \leq 1/8$.
- There is no way for $1, 2, 5, 6 \in S$.
- For $1, 2, 3, 5 \in S$, we need to choose $i = 4$ and $j = 5$, then return $(A \setminus i) \cup j$. This occurs w.p. $1/4 \cdot 1/4 \cdot 1/2 = 1/32 \leq 1/16$.

Thus, \mathcal{D} is NCD, and we again have $\mathbb{E}_{S \sim \mathcal{D}}[f(S)] = 12/8 < 13/8 = \mathbb{E}_{S \sim \mathbf{x}}[f(S)]$, so \mathcal{D} violates Submodular Dominance. \blacktriangleleft

A.3 Product of Singletons is a Convex Combination of Product Distributions

► **Lemma 27.** *Let \mathcal{D} be a product of singletons distribution over 2^E with marginals $\mathbf{x} \in [0, 1]^{nm}$. Let $x_i := \sum_j x_{ij}$, and let $\vec{x} := (x_i : i \in U)$. For any $\mathbf{u} \in [m]^n$, let $E_{\mathbf{u}} := \{iu_i : i \in U\}$ and let $\mathcal{D}_{\mathbf{u}}$ be a product distribution over $2^{E_{\mathbf{u}}}$ with marginals \vec{x} . Then for any $g : 2^E \rightarrow \mathbb{R}$,*

$$\mathbb{E}_{S \sim \mathcal{D}}[g(S)] = \sum_{\mathbf{u} \in [m]^n} \left(\mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}} [g(S)] \cdot \prod_{i \in U} \frac{x_{iu_i}}{x_i} \right). \quad (4)$$

Proof. For some weights p_S , we can rewrite RHS of (4) as

$$\sum_{S \subseteq E} g(S) \cdot p_S.$$

Our approach is to show that $p_T = \Pr_{S \sim \mathcal{D}}[S = T]$ for any $T \subseteq E$. Then the summation is simply the expectation of g over \mathcal{D} and we are finished.

Fix some set $T \subseteq E$. The distributions for which T is in the image of $\mathcal{D}_{\mathbf{u}}$ are those where for all $ij \in T$, $u_i = j$. Therefore,

$$p_T = \sum_{\substack{\mathbf{u} \in [m]^n \\ u_i = j \forall ij \in T}} \left(\Pr_{S \sim \mathcal{D}_{\mathbf{u}}} [S = T] \cdot \prod_{i \in U} \frac{x_{iu_i}}{x_i} \right).$$

Let $T^* \subseteq U$ be a set where $i \in T^*$ if there exists some j for which $ij \in T$. Then,

$$p_T = \sum_{\substack{\mathbf{u} \in [m]^n \\ u_i = j \forall ij \in T}} \left(\prod_{i \in T^*} x_i \prod_{i \notin T^*} (1 - x_i) \cdot \prod_{i \in T^*} \frac{x_{iu_i}}{x_i} \prod_{i \notin T^*} \frac{x_{iu_i}}{x_i} \right).$$

We combine the products over $i \in T^*$, and move the first product over $i \notin T^*$ outside the summation as the inner term does not depend on \mathbf{u} .

$$p_T = \prod_{i \notin T^*} (1 - x_i) \cdot \sum_{\substack{\mathbf{u} \in [m]^n \\ u_i = j \forall ij \in T}} \left(\prod_{i \in T^*} x_{iu_i} \prod_{i \notin T^*} \frac{x_{iu_i}}{x_i} \right).$$

Because the summation is restricted to \mathbf{u} where $u_i = j$ for all $ij \in T$, the coordinates of \mathbf{u} for $i \in T^*$ can only take one value. Thus, the product over $i \in T^*$ is always the same, and can be factored out of the summation.

$$p_T = \prod_{i \notin T^*} (1 - x_i) \cdot \prod_{ij \in T} x_{ij} \cdot \sum_{\substack{\mathbf{u} \in [m]^n \\ u_i = j \forall ij \in T}} \left(\prod_{i \notin T^*} \frac{x_{iu_i}}{x_i} \right).$$

As we just observed, the summation only enforces a condition on $i \in T^*$, so we sum up over all possible $u_i \in [m]$ for $i \notin T^*$. We can rewrite this as

$$\begin{aligned} p_T &= \prod_{i \notin T^*} (1 - x_i) \cdot \prod_{ij \in T} x_{ij} \cdot \prod_{i \notin T^*} \left(\sum_{j \in [m]} \frac{x_{ij}}{x_i} \right) \\ &= \prod_{i \notin T^*} (1 - x_i) \cdot \prod_{ij \in T} x_{ij} \cdot \prod_{i \notin T^*} \left(\frac{1}{x_i} \cdot \sum_{j \in [m]} x_{ij} \right) \\ &= \prod_{i \notin T^*} (1 - x_i) \cdot \prod_{ij \in T} x_{ij} \cdot \prod_{i \notin T^*} \left(\frac{1}{x_i} \cdot x_i \right) \\ &= \prod_{i \notin T^*} (1 - x_i) \cdot \prod_{ij \in T} x_{ij} \\ &= \Pr_{S \sim \mathcal{D}} [S = T]. \end{aligned}$$

Since these computations follow for any $T \subseteq E$, we have

$$\mathbb{E}_{S \sim \mathcal{D}} [g(S)] = \sum_{S \subseteq E} g(S) \cdot p_S = \sum_{\mathbf{u} \in [m]^n} \left(\mathbb{E}_{S \sim \mathcal{D}_{\mathbf{u}}} [g(S)] \cdot \prod_{i \in U} \frac{x_{iu_i}}{x_i} \right),$$

which completes the proof. ◀