γ-weakly θ-up-concavity: A Unified Framework for Non-Convex Optimization Beyond DR-Submodular and OSS Functions

Mohammad Pedramfar
Mila – Quebec AI Institute / McGill University
mohammad.pedramfar@mila.quebec

Vaneet Aggarwal
Purdue University
vaneet@purdue.edu

Abstract

Many objectives in machine learning and combinatorial optimization are non-concave. We introduce γ-weak θ-up-concavity, a first-order condition for nonnegative monotone functions in which a scaling function θ lets the curvature change with the size of the solution. The resulting class strictly contains monotone DR-submodular functions and One-Sided Smooth (OSS) functions. It also contains functions whose marginal gains first increase and then decrease, and functions that vanish faster than any power of the norm at the origin. We construct a surrogate gradient from gradients along the segment from the origin to the current point and prove an upper-linearization inequality with a point-dependent approximation coefficient. Restricting actions to the closed convex hull of the coordinate-wise maximal feasible points makes this coefficient uniformly positive, with a value determined by the curvature parameters and the geometry of the feasible set. Standard reductions to linear optimization then give offline approximation algorithms and online algorithms with regret guarantees. The coefficient matches the optimal one for DR-submodular maximization and improves the best known coefficients for OSS maximization over matroid constraints whenever the OSS parameter is positive.

1 Introduction

Maximizing nonlinear objectives over convex sets arises in influence maximization, recommender systems, experimental design, feature selection, and diversity maximization [1, 7, 23, 34]. We consider problems of the form $\max_{\mathbf{x}\in\mathcal{K}} f(\mathbf{x})$, where $\mathcal{K}$ is a convex set and f belongs to a class $\mathcal{F}$ of nonnegative functions. Such objectives are often non-concave and the problems are often NP-hard, so we look for a polynomial-time algorithm that finds $\mathbf{y} \in \mathcal{K}$ with $f(\mathbf{y}) \geq \alpha \max_{\mathbf{x}\in\mathcal{K}} f(\mathbf{x})$, where the *approximation coefficient* $\alpha \in (0, 1]$ is as large as possible.

Several relaxations of concavity keep these problems tractable in continuous domains, including continuous DR-submodularity [7, 16], up-concavity [24, 28, 33], and one-sided smoothness (OSS) [14, 37, 38]. They have led to approximation algorithms for offline and online optimization under downward-closed and matroid constraints. Pedramfar and Aggarwal [28] introduced *upper-linearizable* function classes and meta-algorithms that transfer guarantees for linear objectives to these classes by bounding approximation regret with the regret of suitable linear surrogates. Upper-linearizable classes include DR-submodular functions and phase retrieval objectives; finding broad, verifiable classes with good upper-linearization constants is open.

We introduce γ-*weak* θ-*up-concavity*, a first-order condition for monotone objectives that strictly generalizes γ-weakly DR-submodular functions, OSS functions, and other weakly concave families. It bounds increments of f by gradient inner products weighted by a monotone scaling function θ, which describes how the curvature changes with the scale of the solution. In many allocation

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

problems, marginal gains first increase and then decrease because of saturation or interference; we call this *accumulating-then-diminishing returns*. Examples include recommendation and influence systems, where engagement may accelerate as exposure accumulates, and allocation problems in which marginal gains are very small at first. Some objectives are also *flat at the origin*: they vanish faster than any power of $\|\mathbf{x}\|_1$ as $\mathbf{x} \to \mathbf{0}$. OSS models allow some of these behaviors but do not describe scale-dependent curvature in general. Appendix D gives functions with both behaviors that are γ-weakly θ-up-concave but neither DR-submodular nor σ-OSS for any parameters.

We show that every γ-weakly θ-up-concave function is upper-linearizable, with coefficients that depend only on the curvature parameters and the geometry of the feasible set. For $\theta(\mathbf{x}) = \|\mathbf{x}\|_1^\kappa$ and a matroid constraint, maximizing over the independence polytope has coefficient $1 - \exp\big(-\frac{\gamma}{2^\kappa(\kappa+1)}\big)$; for $\kappa = 0$ and $\gamma = 1$ this is the optimal coefficient $1 - 1/e$ for DR-submodular maximization. Every nonnegative monotone σ-OSS function satisfies our condition with $\gamma = 1$ and $\theta(\mathbf{x}) = \|\mathbf{x}\|_1^{2\sigma}$ (Theorem 3), which gives the coefficient $1 - \exp\big(-\frac{1}{2^{2\sigma}(2\sigma+1)}\big)$; for $\sigma > 0$ this improves the best known coefficient for general monotone OSS objectives [14, 37] (Section 6). For general downward-closed sets, the coefficient depends on a geometric ratio of the feasible set (Definition 1) and can be smaller than that of [14].

Our main contributions are as follows.

- **A unified curvature model.** We define the class (Section 3) and relate it to existing ones. The inclusions relating OSS functions to this class (Theorem 3) and the separation examples, including multivariate ones, are new.
- **Upper-linearization guarantees.** Every γ-weakly θ-up-concave function is upper-linearizable. When θ is small near the origin, a coefficient that is uniform over $\mathcal{K}$ can be zero, so we prove an inequality whose coefficient depends on the point $\mathbf{x}$ and run the base algorithm on $\mathcal{K}^*$, the closed convex hull of the maximal points of $\mathcal{K}$ (Definition 1), where this coefficient is bounded away from zero. We also give the coefficient for any closed convex set between $\mathcal{K}^*$ and $\mathcal{K}$ (Proposition 1).
- **Improved guarantees under matroid constraints.** Using the geometry of matroid polytopes, we obtain the coefficients stated above; for σ-OSS functions, they improve prior coefficients when $\sigma > 0$ (Section 6).
- **Offline and online optimization via reductions.** Through the reductions of [29, 28], upper-linearizability gives offline and online algorithms; the new ingredients are the surrogate gradient of Theorem 4 and a query algorithm that estimates it without bias from one stochastic gradient. This gives (i) a projection-free offline first-order algorithm with sample complexity $\mathcal{O}(\epsilon^{-2})$; (ii) projection-free algorithms with adaptive regret $\mathcal{O}(T^{1/2})$ under full-information and $\mathcal{O}(T^{2/3})$ under semi-bandit stochastic first-order feedback; and (iii) dynamic regret $\mathcal{O}(\sqrt{T(1+P_T(\mathbf{u}))})$ under full-information stochastic first-order feedback for comparator sequences $\mathbf{u}$ in $\mathcal{K}^*$, where P_T is the path length.

2 Background and Notation

We model online optimization as a repeated game over T rounds between an agent and an adversary on a convex domain $\mathcal{K} \subseteq \mathbb{R}^d$; both players know T and $\mathcal{K}$. An *oblivious* adversary fixes the functions $(f_t)_{t=1}^T$ and their query oracles before the game starts, while a *fully adaptive* adversary may choose f_t and its query oracle after observing $\mathbf{x}_1, \dots, \mathbf{x}_t$. In round t, the agent selects an action $\mathbf{x}_t \in \mathcal{K}$, the adversary provides a query oracle for the reward function $f_t \in \mathcal{F}$, and the agent queries the oracle at points $\mathbf{y}_{t,i}$, $1 \le i \le k_t$, for some $k_t \ge 0$. A query oracle $\mathcal{Q}_f$ for f is a possibly random map that receives a point $\mathbf{y} \in \mathcal{K}$ and returns information about f near $\mathbf{y}$. The gradient oracle returns $\mathcal{Q}_f(\mathbf{y}) = \nabla f(\mathbf{y})$. More generally, $\mathcal{Q}_f$ is a *stochastic first-order* oracle for f if $\mathbb{E}[\mathcal{Q}_f(\mathbf{y})] = \nabla f(\mathbf{y})$, where the expectation is conditional on the history of the game and on the query point $\mathbf{y}$. The oracle is bounded by B if $\|\mathcal{Q}_f(\mathbf{y})\| \le B$ always holds, where $\|\cdot\|$ is the Euclidean norm. An algorithm is *semi-bandit*

Meta-algorithm 1: Online Maximization By Quadratization - `OMBQ`$(\mathcal{A}, \mathcal{G}, h)$

Input : Base algorithm $\mathcal{A}$, Query algorithm $\mathcal{G}$, Mapping $h : \mathcal{K} \to \mathcal{K}$.

for $t = 1, 2, \dots, T$ **do**

- Let $\mathbf{x}_t$ be the action chosen by $\mathcal{A}$.
- **Play:** $\mathbf{y}_t = h(\mathbf{x}_t)$.
- **Query:** Run query algorithm $\mathcal{G}$ at $\mathbf{x}_t$ to obtain gradient estimate $\mathbf{g}_t$.
- **Update:** Pass reward vector $\mathbf{g}_t$ to $\mathcal{A}$ to update its state.

if it queries the oracle only at its action $\mathbf{x}_t$, and *full-information* otherwise. For a function class $\mathcal{F}$, $\mathrm{Adv}_1(\mathcal{F})$ denotes the set of fully adaptive adversaries that select functions from $\mathcal{F}$ and provide deterministic first-order oracles, and $\mathrm{Adv}_1^{\mathrm{o}}(\mathcal{F}, B)$ the set of oblivious adversaries that provide stochastic first-order oracles bounded by B.

For a maximization problem, $\alpha \in (0, 1]$, $1 \le a \le b \le T$, and a compact set $\mathcal{U} \subseteq \mathcal{K}^T$, the α-regret of an algorithm $\mathcal{A}$ against a set of adversaries Adv is

$$\mathcal{R}^{\mathcal{A}}_{\alpha,\mathrm{Adv}}(\mathcal{U})[a,b] := \sup_{\mathcal{B}\in\mathrm{Adv}} \mathbb{E}\left[\alpha \max_{\mathbf{u}=(\mathbf{u}_1,\cdots,\mathbf{u}_T)\in\mathcal{U}} \sum_{t=a}^{b} f_t(\mathbf{u}_t) - \sum_{t=a}^{b} f_t(\mathbf{x}_t)\right],$$

where the expectation is over the randomness of the algorithm and of the query oracles. We drop α when it equals 1, and assume that the functions are nonnegative when $\alpha < 1$. With $a = 1$, $b = T$, and $\mathcal{U} = \mathrm{diag}(\mathcal{K}^T) := \{(\mathbf{x}, \cdots, \mathbf{x}) \mid \mathbf{x} \in \mathcal{K}\}$, this is the *static regret*. With $a = 1$, $b = T$, and $\mathcal{U}$ a single comparator sequence $\mathbf{u}$, it is the *dynamic regret* [44, 39], which is usually bounded in terms of the path length $P_T(\mathbf{u}) := \sum_{i=1}^{T-1} \|\mathbf{u}_i - \mathbf{u}_{i+1}\|$. The *adaptive regret* is $\mathcal{AR}^{\mathcal{A}}_{\alpha,\mathrm{Adv}} := \max_{1\le a\le b\le T} \mathcal{R}^{\mathcal{A}}_{\alpha,\mathrm{Adv}}(\mathrm{diag}(\mathcal{K}^T))[a,b]$ [17]. We omit a, b, and $\mathcal{U}$ when they are clear from the context or do not matter.

We use the upper-linearizable classes of [28]. Let $\mathcal{K} \subseteq \mathbb{R}^d$ be a convex set. A class $\mathcal{F}$ of nonnegative functions over $\mathcal{K}$ is *upper-linearizable* if there are maps $\mathfrak{g} : \mathcal{F} \times \mathcal{K} \to \mathbb{R}^d$ and $h : \mathcal{K} \to \mathcal{K}$ and constants $0 < \alpha \le 1$ and $\beta > 0$ such that[1]

$$\alpha f(\mathbf{y}) - f(h(\mathbf{x})) \le \beta \langle \mathfrak{g}(f, \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \tag{1}$$

for all $f \in \mathcal{F}$ and $\mathbf{x}, \mathbf{y} \in \mathcal{K}$. Here $\mathfrak{g}(f, \mathbf{x})$ acts as a surrogate gradient, h maps the action of the base algorithm to the action played, and α is the resulting approximation coefficient. Given a stochastic first-order oracle $\mathcal{Q}_f$ for each $f \in \mathcal{F}$, an algorithm $\mathcal{G}$ is a *query algorithm* for $\mathfrak{g}$ if $\mathcal{G}(\mathcal{Q}_f, \mathbf{x})$ is an unbiased estimate of $\mathfrak{g}(f, \mathbf{x})$; it may query $\mathcal{Q}_f$ several times and at several points. Upper-linearizable classes can be optimized with the meta-algorithm `OMBQ` (Meta-algorithm 1), which has the following guarantee.

Theorem 1 (Theorem 1 in Pedramfar and Aggarwal [28])**.** *Let $\mathcal{K} \subseteq [0, 1]^d$ and let $\mathcal{A}(\mathcal{K})$ be a deterministic algorithm for online optimization with semi-bandit feedback over $\mathcal{K}$. Also let $\mathcal{F}$ be a function class over $\mathcal{K}$ that is upper-linearizable with $\mathfrak{g} : \mathcal{F} \times \mathcal{K} \to \mathbb{R}^d$ and $h : \mathcal{K} \to \mathcal{K}$. Let $\mathcal{G}$ be a query algorithm for $\mathfrak{g}$ and let $\mathcal{A}' = \mathtt{OMBQ}(\mathcal{A}(\mathcal{K}), \mathcal{G}, h)$. If $\mathcal{G}$ returns an unbiased estimate of $\mathfrak{g}$ and the output of $\mathcal{G}$ is bounded by B_1, then*

$$\mathcal{R}^{\mathcal{A}'}_{\alpha,\mathrm{Adv}_1^{\mathrm{o}}(\mathcal{F},B_1)} \le \beta \mathcal{R}^{\mathcal{A}(\mathcal{K})}_{1,\mathrm{Adv}_1(\mathcal{L}(\mathcal{K})[B_1])},$$

where α and β are the upper-linearization constants and $\mathcal{L}(\mathcal{K})[B_1]$ denotes the class of linear functions $\ell : \mathcal{K} \to \mathbb{R}$ such that $\|\nabla \ell\| \le B_1$.

Pedramfar and Aggarwal [28] showed that concave functions and several classes of up-concave functions are upper-linearizable. The class introduced in the next section satisfies a version of (1) in which α depends on $\mathbf{x}$.

3 γ-Weakly θ-Up-Concave Functions

This section defines the class and relates it to existing ones. We say that a function $\phi : [0, 1]^d \to \mathbb{R}_{\ge 0}$ is *flat at the origin* if $\phi(\mathbf{x})/\|\mathbf{x}\|_1^n \to 0$ as $\mathbf{x} \to \mathbf{0}$ for every $n > 0$.

A function $f : \mathcal{K} \to \mathbb{R}$ is *monotone* if $f(\mathbf{y}) \ge f(\mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{K}$ with $\mathbf{y} \ge \mathbf{x}$. We write C^1 and C^2 for the continuously differentiable and twice continuously differentiable functions.

Let $\theta : [0, 1]^d \to \mathbb{R}_{\ge 0}$ be a continuous monotone function where $\theta(\mathbf{x}) \ne 0$ for all $\mathbf{x} \in [0, 1]^d \setminus \{\mathbf{0}\}$. Given $0 < \gamma \le 1$, we say a C^1 monotone function $f : [0, 1]^d \to \mathbb{R}_{\ge 0}$ is *γ-weakly θ-up-concave* if

$$\tfrac{\gamma\,\theta(\mathbf{x})}{\theta(\mathbf{y})} \langle \nabla f(\mathbf{y}), \mathbf{y} - \mathbf{x}\rangle \le f(\mathbf{y}) - f(\mathbf{x}) \le \tfrac{\theta(\mathbf{y})}{\gamma\,\theta(\mathbf{x})} \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle, \tag{2}$$

[1] Upper-linearizability generalizes concavity and is used for maximization; lower-linearizability is the analogous notion for minimization.

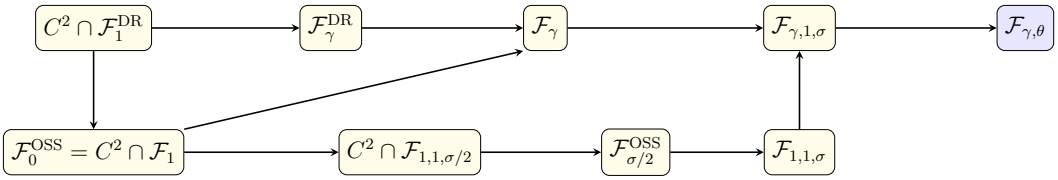

Figure 1: Containment relations among the function classes introduced in Section 3. An arrow $A \to B$ indicates that $A \subseteq B$. The arrow into $\mathcal{F}_{\gamma,\theta}$ means that $\mathcal{F}_{\gamma,1,\sigma}$ is the special case $\theta(\mathbf{x}) = \|\mathbf{x}\|_1^\sigma$.

for all $\mathbf{x}, \mathbf{y} \in [0,1]^d$ with $\mathbf{y} \geq \mathbf{x} \neq \mathbf{0}$. If $\theta(\mathbf{0}) > 0$, the inequalities also hold for $\mathbf{x} = \mathbf{0}$ by continuity, taking $\mathbf{x} = \epsilon\mathbf{y}$ with $\epsilon \downarrow 0$. We write $\mathcal{F}_{\gamma,\theta}$ for this class. The ratio $\theta(\mathbf{y})/\theta(\mathbf{x})$ lets the curvature bounds change multiplicatively with scale. The class is closed under nonnegative combinations.

Lemma 1. *If $f, g \in \mathcal{F}_{\gamma,\theta}$ and $a, b \geq 0$, then $af + bg \in \mathcal{F}_{\gamma,\theta}$.*

Proof. The function $h = af + bg$ is C^1, nonnegative and monotone, and $\nabla h = a\nabla f + b\nabla g$. Multiplying the inequalities (2) for f by a and those for g by b, and adding them, gives (2) for h. □

The case $\theta(\mathbf{x}) = \|\mathbf{x}\|_p^\sigma$ with $p \geq 1$ and $\sigma \geq 0$ is important (for $\sigma = 0$, $\theta \equiv 1$, including at the origin). We call such functions γ-*weakly* (p,σ)-*up-concave*; they satisfy

$$\frac{\gamma \|\mathbf{x}\|_p^\sigma}{\|\mathbf{y}\|_p^\sigma} \langle \nabla f(\mathbf{y}), \mathbf{y} - \mathbf{x} \rangle \leq f(\mathbf{y}) - f(\mathbf{x}) \leq \frac{\|\mathbf{y}\|_p^\sigma}{\gamma \|\mathbf{x}\|_p^\sigma} \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle. \tag{3}$$

for all $\mathbf{x}, \mathbf{y} \in [0,1]^d$ with $\mathbf{y} \geq \mathbf{x} \neq \mathbf{0}$. We use $\mathcal{F}_{\gamma,p,\sigma}$ to denote this class of functions.

Normalized Diminishing Returns. The next theorem, proved in Appendix B, gives a sufficient condition in terms of a normalized gradient.

Theorem 2. *Let $\theta : [0,1]^d \to \mathbb{R}_{\geq 0}$ be monotone, continuous, and satisfy $\theta(\mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$. Suppose $f : [0,1]^d \to \mathbb{R}_{\geq 0}$ is a C^1 monotone function where, for all $\mathbf{x}, \mathbf{y} \in [0,1]^d$ with $\mathbf{y} \geq \mathbf{x} \neq 0$,*

$$\frac{\gamma \nabla f(\mathbf{y})}{\theta(\mathbf{y})} \leq \frac{\nabla f(\mathbf{x})}{\theta(\mathbf{x})} \tag{4}$$

coordinate-wise. Then f is γ-weakly θ-up-concave.

For $\gamma = 1$, condition (4) says that the normalized gradient $\nabla f/\theta$ is nonincreasing along positive directions; for $\gamma < 1$, it may increase by a factor of at most $1/\gamma$. The gradient ∇f itself may increase, as long as θ grows accordingly. Condition (4) is sufficient but not necessary.

Choosing θ and γ. In a model, θ describes how the scale of the solution affects marginal gains, and γ how far the model may deviate from exact normalized diminishing returns. Polynomial growth of marginal gains suggests $\theta(\mathbf{x}) = \|\mathbf{x}\|_p^\sigma$; marginal gains that are extremely small at low allocation levels suggest a θ that is flat at the origin, as in Example 1. For an explicit f, Theorem 2 can be used to check a proposed pair. If f is C^2 and θ is differentiable on $[0,1]^d \setminus \{\mathbf{0}\}$, then $\partial_j(\partial_i f/\theta) \leq 0$ on $[0,1]^d \setminus \{\mathbf{0}\}$ for all i, j implies (4) with $\gamma = 1$, because $\nabla f/\theta$ is then coordinate-wise nonincreasing along every segment from $\mathbf{x}$ to $\mathbf{y}$ with $\mathbf{y} \geq \mathbf{x} \neq \mathbf{0}$. More generally, any lower bound $\gamma \in (0,1]$ on the ratios $(\partial_i f(\mathbf{x})/\theta(\mathbf{x}))/(\partial_i f(\mathbf{y})/\theta(\mathbf{y}))$ over all i and $\mathbf{y} \geq \mathbf{x} \neq \mathbf{0}$ with $\partial_i f(\mathbf{y}) > 0$ satisfies (4). For example, γ-weakly DR-submodular functions satisfy (4) with constant θ, and nonnegative monotone σ-OSS functions are 1-weakly θ-up-concave with $\theta(\mathbf{x}) = \|\mathbf{x}\|_1^{2\sigma}$ (Theorem 3).

Relation to Existing Classes. For constant θ, that is, for $\mathcal{F}_{\gamma,p,0}$, condition (2) is γ-weak up-concavity [41, 28]: for all $\mathbf{x} \leq \mathbf{y}$ in $[0,1]^d$,

$$\gamma \langle \nabla f(\mathbf{y}), \mathbf{y} - \mathbf{x} \rangle \leq f(\mathbf{y}) - f(\mathbf{x}) \leq \tfrac{1}{\gamma} \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle. \tag{5}$$

We write $\mathcal{F}_\gamma$ for the class of nonnegative monotone γ-weakly up-concave functions. It contains the γ-weakly DR-submodular functions [16, 32], defined by

$$\gamma \nabla f(\mathbf{y}) \leq \nabla f(\mathbf{x}), \tag{6}$$

for all $\mathbf{x}, \mathbf{y} \in [0,1]^d$ with $\mathbf{y} \geq \mathbf{x}$. This is (4) with constant θ; ordinary DR-submodularity is the case $\gamma = 1$. We write $\mathcal{F}_\gamma^{\mathrm{DR}}$ for the class of nonnegative monotone γ-weakly DR-submodular functions.

For $\sigma \geq 0$, a C^2 function f is *σ-One-Sided Smooth (σ-OSS)* [14, 37] if

$$\frac{1}{2}\mathbf{u}^\top \nabla^2 f(\mathbf{x})\mathbf{u} \leq \sigma \frac{\|\mathbf{u}\|_1}{\|\mathbf{x}\|_1}\mathbf{u}^\top \nabla f(\mathbf{x}) \tag{7}$$

for all $\mathbf{x} \in [0,1]^d \setminus \{\mathbf{0}\}$ and $\mathbf{u} \geq \mathbf{0}$. We write $\mathcal{F}_\sigma^{\mathrm{OSS}}$ for the class of nonnegative monotone σ-OSS functions, as in [14]. Unlike DR-submodularity, the σ-OSS condition allows marginal gains to increase (Lemma 2). The next theorem, proved in Appendix C, relates it to $(1,\sigma)$-up-concavity.

Theorem 3. *For any $\sigma \geq 0$, we have $C^2 \cap \mathcal{F}_{1,1,\sigma} \subseteq \mathcal{F}_\sigma^{\mathrm{OSS}} \subseteq \mathcal{F}_{1,1,2\sigma}$.*

For $\sigma > 0$, the exponent 2σ cannot be reduced: for every $\beta \in [0, 2\sigma)$, some nonnegative monotone σ-OSS function is not in $\mathcal{F}_{1,1,\beta}$ (Proposition 2 in Appendix C).

Figure 1 shows the resulting inclusions, using also that $\mathcal{F}_0^{\mathrm{OSS}}$ contains $C^2 \cap \mathcal{F}_1^{\mathrm{DR}}$ [14]. Appendix D constructs functions that are 1-weakly θ-up-concave but neither OSS nor γ-weakly (p,σ)-up-concave for any parameters; they are flat at the origin and can have accumulating-then-diminishing returns (Lemma 4). The following example is of this kind.

Example 1 (Allocation with slow start and saturation)**.** *Let x_i be the amount allocated to channel i (e.g., advertising exposure), let $\mathbf{a} \in \mathbb{R}^d_{>0}$ describe the effective exposure produced by each channel, and let $M > 0$. Define*

$$f(\mathbf{x}) = \int_0^{\mathbf{a}^\top \mathbf{x}} \exp\Big(-\frac{1}{t} - Mt\Big)\, dt, \qquad \theta(\mathbf{x}) = \exp\Big(-\frac{1}{\mathbf{a}^\top \mathbf{x}}\Big) \text{ for } \mathbf{x} \neq \mathbf{0}, \qquad \theta(\mathbf{0}) = 0.$$

Then $\nabla f(\mathbf{x})/\theta(\mathbf{x}) = e^{-M\mathbf{a}^\top \mathbf{x}}\mathbf{a}$ is coordinate-wise nonincreasing, so f is 1-weakly θ-up-concave by Theorem 2. As a function of the effective exposure $s = \mathbf{a}^\top \mathbf{x}$, the marginal gain $e^{-1/s - Ms}$ is exponentially small for small s, increases for $s < 1/\sqrt{M}$, and decreases for $s > 1/\sqrt{M}$; both phases occur on $[0,1]^d$ when $M > \|\mathbf{a}\|_1^{-2}$. Since f is flat at the origin, it is neither σ-OSS nor γ-weakly (p,σ)-up-concave for any parameters (Lemma 5 in Appendix D).

4 γ-Weakly θ-Up-Concave Functions are Upper-Linearizable

Our main result, Theorem 4, shows that γ-weakly θ-up-concave functions are upper-linearizable. It uses the following quantities.

$\mathcal{K}$	$\mathcal{K}^*$
$[0,1]^d$	$\{\mathbf{1}\}$
$\{\mathbf{x} \in [0,1]^d : \|\mathbf{x}\|_1 \leq 1\}$	$\{\mathbf{x} \in [0,1]^d : \|\mathbf{x}\|_1 = 1\}$
Matroid independence polytope	Matroid basis polytope
$\{\mathbf{x} \in [0,1]^d : \mathbf{a}^\top \mathbf{x} \leq B\}$	$\{\mathbf{x} \in [0,1]^d : \mathbf{a}^\top \mathbf{x} = B\}$

Table 1: Examples of $\mathcal{K}$ and corresponding $\mathcal{K}^*$ (last row: $\mathbf{a} > \mathbf{0}$, $0 < B \leq \sum_i a_i$).

Definition 1. *Let $\mathcal{K} \subseteq [0,1]^d$ be a nonempty, closed, convex set. We define the* maximal subset $\mathcal{K}^{\mathrm{m}} \subseteq \mathcal{K}$ *to be the set of points $\mathbf{v} \in \mathcal{K}$ such that there is no $\mathbf{u} \in \mathcal{K}$ with $\mathbf{u} \geq \mathbf{v}$ and $\mathbf{u} \neq \mathbf{v}$. We define $\mathcal{K}^* = \overline{\mathrm{conv}}(\mathcal{K}^{\mathrm{m}})$, the closed convex hull of $\mathcal{K}^{\mathrm{m}}$. For a set $A \subseteq \mathbb{R}^d$, $\mathrm{conv}(A)$ is its convex hull and $\overline{\mathrm{conv}}(A)$ the closure of the convex hull. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, $\mathbf{u} \vee \mathbf{v}$ is the coordinate-wise maximum. For a continuous monotone function $\theta : [0,1]^d \to \mathbb{R}_{\geq 0}$, we define*

$$r_\theta(\mathcal{K}) = \min_{\mathbf{v} \in \mathcal{K}^*} \theta(\mathbf{v}), \quad R_\theta(\mathcal{K}) = \max_{\mathbf{v},\mathbf{u} \in \mathcal{K}} \theta(\mathbf{v} \vee \mathbf{u}) = \max_{\mathbf{v},\mathbf{u} \in \mathcal{K}^*} \theta(\mathbf{v} \vee \mathbf{u}).$$

For $\theta(\mathbf{v}) = \|\mathbf{v}\|_p$ with $p \geq 1$, we write r_p and R_p.

For $\mathcal{K} = [0,1]^d$, we have $\mathcal{K}^* = \mathcal{K}^{\mathrm{m}} = \{\mathbf{1}\}$, so $r_\theta = R_\theta$ for every θ; Table 1 lists further examples.

Every point of the compact set $\mathcal{K}$ is dominated by a point of $\mathcal{K}^{\mathrm{m}}$, so a continuous monotone function on $\mathcal{K}$ attains its maximum on $\mathcal{K}^{\mathrm{m}} \subseteq \mathcal{K}^* \subseteq \mathcal{K}$, and the optimization may be restricted to $\mathcal{K}^*$.

Theorem 4. *Let $\mathcal{K} \subseteq [0,1]^d$ be a closed convex set containing the origin and at least one other point, and let $\mathcal{F} = \mathcal{F}_{\gamma,\theta}$ be the class of γ-weakly θ-up-concave functions. Let $R_\theta = R_\theta(\mathcal{K})$ and $l(r,\mathbf{x}) := \frac{-\gamma}{R_\theta}\int_r^1 \theta(s\mathbf{x})ds$, and let $\mathfrak{g} : \mathcal{F} \times \mathcal{K} \to \mathbb{R}^d$ be the functional defined by*

$$\mathfrak{g}(f,\mathbf{x}) = \int_0^1 e^{l(r,\mathbf{x})} \nabla f(r\mathbf{x})dr,$$

for all $f \in \mathcal{F}$ and $\mathbf{x} \in \mathcal{K}$. Then $\langle \mathfrak{g}(f,\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle \geq \alpha_{\mathbf{x}} f(\mathbf{y}) - f(\mathbf{x})$, for all $\mathbf{y} \in \mathcal{K}$ and $\mathbf{x} \in \mathcal{K}$, where $\alpha_{\mathbf{x}} = 1 - \exp\left(-\frac{\gamma}{R_\theta}\int_0^1 \theta(s\mathbf{x})ds\right)$.

Proof. Throughout the proof, fix $f \in \mathcal{F}$ and $\mathbf{x}, \mathbf{y} \in \mathcal{K}$, and write $R = R_\theta$. Note that $R > 0$, since $\mathcal{K}$ contains a point $\mathbf{v} \neq \mathbf{0}$ and $R \geq \theta(\mathbf{v} \vee \mathbf{v}) = \theta(\mathbf{v}) > 0$.

Preliminaries on l. Recall that $l(r,\mathbf{x}) = -\frac{\gamma}{R}\int_r^1 \theta(s\mathbf{x})ds$ for $r \in [0,1]$. Since $s \mapsto \theta(s\mathbf{x})$ is continuous and nonnegative, the fundamental theorem of calculus gives

$$l(1,\mathbf{x}) = 0, \qquad l(r,\mathbf{x}) \leq 0, \qquad \frac{\partial}{\partial r} l(r,\mathbf{x}) = \frac{\gamma}{R}\theta(r\mathbf{x}), \qquad \frac{d}{dr} e^{l(r,\mathbf{x})} = \frac{\gamma}{R}\theta(r\mathbf{x}) e^{l(r,\mathbf{x})}. \tag{8}$$

Integrating the last identity in (8) over $r \in [0,1]$ and using $l(1,\mathbf{x}) = 0$, we obtain

$$\int_0^1 \frac{\gamma}{R}\theta(r\mathbf{x}) e^{l(r,\mathbf{x})} dr = e^{l(1,\mathbf{x})} - e^{l(0,\mathbf{x})} = 1 - \exp\left(-\frac{\gamma}{R}\int_0^1 \theta(s\mathbf{x})ds\right) = \alpha_{\mathbf{x}}. \tag{9}$$

We also introduce the shorthand

$$J := \int_0^1 \frac{\gamma}{R}\theta(r\mathbf{x}) e^{l(r,\mathbf{x})} f(r\mathbf{x}) dr. \tag{10}$$

An upper bound on $\langle \mathfrak{g}(f,\mathbf{x}), \mathbf{x}\rangle$. By the definition of $\mathfrak{g}$ and linearity of the integral,

$$\langle \mathfrak{g}(f,\mathbf{x}), \mathbf{x}\rangle = \int_0^1 e^{l(r,\mathbf{x})} \langle \nabla f(r\mathbf{x}), \mathbf{x}\rangle dr = \int_0^1 e^{l(r,\mathbf{x})} \frac{d}{dr} f(r\mathbf{x})\, dr, \tag{11}$$

where we used the chain rule $\frac{d}{dr} f(r\mathbf{x}) = \langle \nabla f(r\mathbf{x}), \mathbf{x}\rangle$. Both $r \mapsto e^{l(r,\mathbf{x})}$ and $r \mapsto f(r\mathbf{x})$ are C^1 on $[0,1]$, so we may integrate (11) by parts. Using (8) and (10),

$$\langle \mathfrak{g}(f,\mathbf{x}), \mathbf{x}\rangle = e^{l(1,\mathbf{x})} f(\mathbf{x}) - e^{l(0,\mathbf{x})} f(\mathbf{0}) - \int_0^1 f(r\mathbf{x}) \frac{d}{dr} e^{l(r,\mathbf{x})} dr = f(\mathbf{x}) - e^{l(0,\mathbf{x})} f(\mathbf{0}) - J. \tag{12}$$

Since $f(\mathbf{0}) \geq 0$ and $e^{l(0,\mathbf{x})} > 0$, (12) implies

$$\langle \mathfrak{g}(f,\mathbf{x}), \mathbf{x}\rangle \leq f(\mathbf{x}) - J. \tag{13}$$

A pointwise lower bound. We claim that, for every $r \in [0,1]$,

$$\langle \nabla f(r\mathbf{x}), \mathbf{y}\rangle \geq \frac{\gamma}{R}\theta(r\mathbf{x})\big(f(\mathbf{y}) - f(r\mathbf{x})\big). \tag{14}$$

Fix $r \in [0,1]$ and let $\mathbf{w} = \mathbf{y} \vee (r\mathbf{x})$ be the coordinate-wise maximum. Since $\mathcal{K}$ is convex and contains $\mathbf{0}$ and $\mathbf{x}$, it contains $r\mathbf{x}$, and therefore, by the definition of R_θ,

$$\theta(\mathbf{w}) = \theta(\mathbf{y} \vee (r\mathbf{x})) \leq R. \tag{15}$$

For each coordinate i, if $y_i \geq r x_i$ then $w_i - r x_i = y_i - r x_i \in [0, y_i]$, and if $y_i < r x_i$ then $w_i - r x_i = 0$. Hence

$$\mathbf{0} \leq \mathbf{w} - r\mathbf{x} \leq \mathbf{y}. \tag{16}$$

Since f is monotone, $\nabla f(r\mathbf{x}) \geq \mathbf{0}$. Together with (16), this gives

$$\langle \nabla f(r\mathbf{x}), \mathbf{y}\rangle \geq \langle \nabla f(r\mathbf{x}), \mathbf{w} - r\mathbf{x}\rangle. \tag{17}$$

Now we consider two cases.

Case $r\mathbf{x} \neq \mathbf{0}$. The pair $(r\mathbf{x}, \mathbf{w})$ satisfies $\mathbf{w} \geq r\mathbf{x} \neq \mathbf{0}$, and $\theta(\mathbf{w}) \geq \theta(r\mathbf{x}) > 0$. The upper inequality in (2) for this pair reads $f(\mathbf{w}) - f(r\mathbf{x}) \leq \frac{\theta(\mathbf{w})}{\gamma\theta(r\mathbf{x})}\langle \nabla f(r\mathbf{x}), \mathbf{w} - r\mathbf{x}\rangle$, that is,

$$\langle \nabla f(r\mathbf{x}), \mathbf{w} - r\mathbf{x}\rangle \geq \frac{\gamma\theta(r\mathbf{x})}{\theta(\mathbf{w})}\big(f(\mathbf{w}) - f(r\mathbf{x})\big). \tag{18}$$

Since $\mathbf{w} \geq r\mathbf{x}$ and f is monotone, $f(\mathbf{w}) - f(r\mathbf{x}) \geq 0$. Hence, by (15), replacing $\theta(\mathbf{w})$ by the larger number R can only decrease the right-hand side of (18). Moreover, $\mathbf{w} \geq \mathbf{y}$, so $f(\mathbf{w}) \geq f(\mathbf{y})$. Combining these facts with (17) and (18),

$$\langle \nabla f(r\mathbf{x}), \mathbf{y}\rangle \geq \frac{\gamma\theta(r\mathbf{x})}{\theta(\mathbf{w})}\big(f(\mathbf{w}) - f(r\mathbf{x})\big) \geq \frac{\gamma\theta(r\mathbf{x})}{R}\big(f(\mathbf{w}) - f(r\mathbf{x})\big) \geq \frac{\gamma\theta(r\mathbf{x})}{R}\big(f(\mathbf{y}) - f(r\mathbf{x})\big),$$

which is (14).

Case $r\mathbf{x} = \mathbf{0}$. Then $\mathbf{w} = \mathbf{y}$. If $\theta(\mathbf{0}) = 0$, the right-hand side of (14) is 0, and the left-hand side is nonnegative because $\nabla f(\mathbf{0}) \geq \mathbf{0}$ and $\mathbf{y} \geq \mathbf{0}$. If $\theta(\mathbf{0}) > 0$ and $\mathbf{y} \neq \mathbf{0}$, then (2) holds for the pair $(\mathbf{0}, \mathbf{y})$ by continuity (see Section 3), and the argument of the previous case applies verbatim. If $\mathbf{y} = \mathbf{0}$, both sides of (14) are 0.

Integrating the lower bound. Multiplying (14) by $e^{l(r,\mathbf{x})} > 0$ and integrating over $r \in [0, 1]$, we get

$$\langle \mathfrak{g}(f, \mathbf{x}), \mathbf{y} \rangle = \int_0^1 e^{l(r,\mathbf{x})} \langle \nabla f(r\mathbf{x}), \mathbf{y} \rangle dr \geq f(\mathbf{y}) \int_0^1 \frac{\gamma}{R} \theta(r\mathbf{x}) e^{l(r,\mathbf{x})} dr - J = \alpha_{\mathbf{x}} f(\mathbf{y}) - J, \quad (19)$$

where the last equality uses (9) and (10).

Combining the bounds. Subtracting (13) from (19), the term J cancels and we obtain

$$\langle \mathfrak{g}(f, \mathbf{x}), \mathbf{y} - \mathbf{x} \rangle = \langle \mathfrak{g}(f, \mathbf{x}), \mathbf{y} \rangle - \langle \mathfrak{g}(f, \mathbf{x}), \mathbf{x} \rangle \geq \big(\alpha_{\mathbf{x}} f(\mathbf{y}) - J\big) - \big(f(\mathbf{x}) - J\big) = \alpha_{\mathbf{x}} f(\mathbf{y}) - f(\mathbf{x}). \ \square$$

The surrogate gradient $\mathfrak{g}(f, \mathbf{x})$ uses ∇f only on the segment from $\mathbf{0}$ to $\mathbf{x}$, which lies in $\mathcal{K}$ because $\mathbf{0} \in \mathcal{K}$. The proof uses R_θ only through $\theta(\mathbf{y} \vee (r\mathbf{x})) \leq R_\theta$, so R_θ may be replaced throughout (in l, $\mathfrak{g}$, the query algorithm and α) by any known upper bound $\overline{R}_\theta \geq R_\theta(\mathcal{K})$, for example $\overline{R}_\theta = \theta(\mathbf{1})$.

When $\inf_{\mathbf{x} \in \mathcal{K}} \alpha_{\mathbf{x}} > 0$, this infimum gives a uniform approximation coefficient on $\mathcal{K}$.

Corollary 1. *If* $\theta \equiv 1$, *then* $\inf_{\mathbf{x} \in \mathcal{K}} \alpha_{\mathbf{x}} = 1 - e^{-\gamma}$, *the best known coefficient for* γ*-weakly up-concave functions [28, 41].*

Proof. If $\theta \equiv 1$, then $R_\theta = \max_{\mathbf{v}, \mathbf{u} \in \mathcal{K}} \theta(\mathbf{v} \vee \mathbf{u}) = 1$ and $\int_0^1 \theta(s\mathbf{x}) ds = 1$ for every $\mathbf{x} \in \mathcal{K}$, so $\alpha_{\mathbf{x}} = 1 - e^{-\gamma}$ for every $\mathbf{x} \in \mathcal{K}$. $\square$

If $\theta(\mathbf{0}) = 0$, then $\alpha_{\mathbf{0}} = 0$, so the infimum over $\mathcal{K}$ is 0. Restricting the actions to a subset of $\mathcal{K}$ that is bounded away from the origin keeps $\alpha_{\mathbf{x}}$ bounded away from 0.

Corollary 2. *Under the assumptions of Theorem 4, we have* $\mathbf{0} \notin \mathcal{K}^*$ *and*

$$\alpha := 1 - \exp\left(-\frac{\gamma}{R_\theta} \inf_{\mathbf{x} \in \mathcal{K}^*} \int_0^1 \theta(s\mathbf{x}) ds \right) > 0.$$

Moreover, for all $f \in \mathcal{F}$ *and* $\mathbf{x}, \mathbf{y} \in \mathcal{K}^*$,

$$\langle \mathfrak{g}(f, \mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq \alpha f(\mathbf{y}) - f(\mathbf{x}).$$

The proof is in Appendix E.

For $\theta(\mathbf{z}) = \|\mathbf{z}\|_p^\sigma$ with $p \geq 1$ and $\sigma \geq 0$, we have $R_\theta(\mathcal{K}) = (R_p(\mathcal{K}))^\sigma$, $r_\theta(\mathcal{K}) = (r_p(\mathcal{K}))^\sigma$, and $\alpha = 1 - \exp\left(-\frac{\gamma}{\sigma+1} \left(\frac{r_p}{R_p} \right)^\sigma \right)$.

Section 5 shows that this inequality on $\mathcal{K}^*$ suffices to obtain these coefficients for maximization over $\mathcal{K}$.

5 Main Algorithm

We now combine Theorem 4 with the meta-algorithm `OMBQ` of [28] (Section 2). For our class, `OMBQ` uses the surrogate gradient $\mathfrak{g}$ of Theorem 4 and the following query algorithm, which estimates $\mathfrak{g}$ without bias from a single stochastic gradient.

Definition 2. *Let* $\mathcal{K} \subseteq [0, 1]^d$ *be a nonempty, closed, convex set that contains the origin and at least one other point. For* $\mathbf{x} \in \mathcal{K}$, *let* $l(r, \mathbf{x}) := \frac{-\gamma}{R_\theta} \int_r^1 \theta(s\mathbf{x}) ds$ *as in Theorem 4. Given a stochastic first-order query oracle* $\mathcal{Q}_f$ *for* f, *we define* $\mathcal{G}(\mathcal{Q}_f, \mathbf{x}) := e^{l(Z,\mathbf{x})} \mathcal{Q}_f(Z\mathbf{x})$, *where* Z *is sampled uniformly from* $[0, 1]$ *using fresh randomization. For* $\theta(\mathbf{x}) = \|\mathbf{x}\|_p^\sigma$, *the weight is explicit:* $l(Z, \mathbf{x}) = -\frac{\gamma \|\mathbf{x}\|_p^\sigma}{(\sigma+1) R_\theta} (1 - Z^{\sigma+1})$. *For a general* θ, *we assume that this one-dimensional integral can be evaluated.*

The base algorithm plays only points of $\mathcal{K}^*$. On $\mathcal{K}^*$, the point-dependent inequality of Theorem 4 becomes the uniform inequality of Corollary 2, which gives the following regret bound, proved in Appendix F.

Theorem 5. *Let $\mathcal{K} \subseteq [0,1]^d$ be a closed convex set containing the origin and at least one other point, and let $\mathcal{F} = \mathcal{F}_{\gamma,\theta}$ be the class of γ-weakly θ-up-concave functions. Also let $\mathcal{G}$ be as above and let $h = \mathrm{Id}_{\mathcal{K}^*}$. Then $\mathcal{G}$ is a query algorithm for $\mathfrak{g}$, which is bounded by B_1 whenever $\mathcal{Q}_f$ is bounded by $B_1 \geq 0$. Moreover, if $\mathcal{A}(\mathcal{K}^*)$ is a deterministic algorithm for online optimization with semi-bandit feedback over $\mathcal{K}^*$ and $\mathcal{A}' = \texttt{OMBQ}(\mathcal{A}(\mathcal{K}^*), \mathcal{G}, h)$, then*

$$\mathcal{R}^{\mathcal{A}'}_{\alpha,\mathrm{Adv}^{\mathrm{o}}_1(\mathcal{F},B_1)} \leq \mathcal{R}^{\mathcal{A}(\mathcal{K}^*)}_{1,\mathrm{Adv}_1(\mathcal{L}(\mathcal{K}^*)[B_1])},$$

where $\alpha = 1 - \exp\left(-\frac{\gamma}{R_\theta} \inf_{\mathbf{x}\in\mathcal{K}^*} \int_0^1 \theta(s\mathbf{x})ds\right)$. The inequality holds for every interval and every comparator set $\mathcal{U} \subseteq (\mathcal{K}^*)^T$, used on both sides. For static and adaptive regret, the nonlinear benchmark on the left may equivalently be maximized over $\mathcal{K}$, while the linear comparator domain on the right remains $\mathcal{K}^*$.

Remark 1. *The base algorithm needs a set oracle for $\mathcal{K}^*$: a projection oracle for* OGA*, or a separation oracle for the Separation-Oracle OGA (*SO-OGA*) of Section 7. Computing $\mathcal{K}^*$ is a problem in vector optimization: if $\mathcal{K}$ is a rational polytope, the classical Benson algorithm [3, 22] describes $\mathcal{K}^*$ in finite time, possibly exponential in the description size of $\mathcal{K}$ (Appendix H). Alternatively, the algorithm may run on any closed convex set $\mathcal{K}'$ with $\mathcal{K}^* \subseteq \mathcal{K}' \subseteq \mathcal{K}$ for which the required oracle is available; Proposition 1 gives the resulting coefficient, which can be zero, for example for $\mathcal{K}' = \mathcal{K}$ when $\theta(\mathbf{0}) = 0$. For matroid independence polytopes, $\mathcal{K}^*$ is the basis polytope (Section 6).*

The next proposition is proved in Appendix G.

Proposition 1. *Let $\mathcal{K} \subseteq [0,1]^d$ be a closed convex set containing the origin and at least one other point, let $\mathcal{F} = \mathcal{F}_{\gamma,\theta}$, let $R_\theta = R_\theta(\mathcal{K})$, and let $\mathcal{K}'$ be a closed convex set with $\mathcal{K}^* \subseteq \mathcal{K}' \subseteq \mathcal{K}$. Define*

$$\alpha(\mathcal{K}') := 1 - \exp\left(-\frac{\gamma}{R_\theta} \inf_{\mathbf{x}\in\mathcal{K}'} \int_0^1 \theta(s\mathbf{x})ds\right).$$

Then the inequality of Corollary 2 holds with $\mathcal{K}'$ and $\alpha(\mathcal{K}')$ in place of $\mathcal{K}^*$ and α. If $\alpha(\mathcal{K}') > 0$, the same is true for Theorem 5; if $\alpha(\mathcal{K}') = 0$, there is no approximation guarantee. Moreover, $\alpha(\mathcal{K}') \leq \alpha(\mathcal{K}^*)$. If $\theta(\mathbf{x}) = \|\mathbf{x}\|_p^\sigma$, then $\alpha(\mathcal{K}') = 1 - \exp\left(-\frac{\gamma}{\sigma+1}\left(\min_{\mathbf{x}\in\mathcal{K}'} \|\mathbf{x}\|_p / R_p(\mathcal{K})\right)^\sigma\right)$, which equals $1 - e^{-\gamma}$ when $\sigma = 0$. If θ is L-Lipschitz with respect to a norm $\|\cdot\|$ and every point of $\mathcal{K}'$ is within distance ε of $\mathcal{K}^*$ in this norm, then $\max\{0,\ 1 - (1 - \alpha(\mathcal{K}^*))\exp(\frac{\gamma L\varepsilon}{2R_\theta})\} \leq \alpha(\mathcal{K}') \leq \alpha(\mathcal{K}^*)$.

Since every action lies in $\mathcal{K}^*$, the coefficient $\alpha_{\mathbf{x}}$ of Theorem 4 at each played point is at least the α of Corollary 2. Algorithm 2 gives the pseudo-code when the base algorithm is Online Gradient Ascent (OGA); Section 7 uses other base algorithms.

Algorithm 2: OMBQ with Online Gradient Ascent

Input : Oracles $\mathcal{Q}_{f_t}$, projection $\mathcal{P}_{\mathcal{K}^*}$, step sizes (η_t).
Let $\mathbf{x}_1$ be a point in $\mathcal{K}^*$.
for $t = 1, 2, \ldots, T$ **do**
  Play $\mathbf{x}_t$.
  Sample $Z_t \sim \mathrm{Unif}[0,1]$;
  $\mathbf{o}_t \leftarrow \mathcal{Q}_{f_t}(Z_t\mathbf{x}_t)$.
  $\mathbf{g}_t \leftarrow e^{l(Z_t,\mathbf{x}_t)}\mathbf{o}_t$.
  $\mathbf{x}_{t+1} \leftarrow \mathcal{P}_{\mathcal{K}^*}(\mathbf{x}_t + \eta_t\mathbf{g}_t)$

Feedback model. The base algorithm $\mathcal{A}(\mathcal{K}^*)$ receives $\mathbf{g}_t$ as the gradient of a linear reward at its own action $\mathbf{x}_t$, so it is semi-bandit for the surrogate linear problem. The composed algorithm $\mathcal{A}' = \texttt{OMBQ}(\mathcal{A}(\mathcal{K}^*), \mathcal{G}, \mathrm{Id})$ plays $\mathbf{x}_t \in \mathcal{K}^*$ but queries $\mathcal{Q}_{f_t}$ at $Z_t\mathbf{x}_t$ with Z_t uniform on $[0,1]$, which in general differs from $\mathbf{x}_t$. Hence $\mathcal{A}'$ is a first-order full-information algorithm for the original problem, and f_t and $\mathcal{Q}_{f_t}$ must be defined on all of $\mathcal{K}$. Section 7 obtains a semi-bandit algorithm by applying the conversion SFTT of [28] to $\mathcal{A}'$.

Corollary 3. *Under the assumptions of Theorem 5, with α as there, assume that $\mathcal{Q}_f$ is a stochastic first-order oracle bounded by $B_1 > 0$ for any $f \in \mathcal{F}$. With step sizes $\eta_t = \frac{D}{B_1\sqrt{t}}$, where $D \geq \mathrm{diam}(\mathcal{K}^*)$, the static regret of Algorithm 2 satisfies $\mathcal{R}_{\alpha,\mathrm{Adv}^{\mathrm{o}}_1(\mathcal{F},B_1)} \leq \frac{3}{2}B_1D\sqrt{T}$.*

6 γ-weakly $(1,\sigma)$-up-concave Functions with Matroid Constraint Sets

We recall the standard definitions.

Definition 3 (Matroid)**.** *A* matroid *is a pair $\mathcal{M} = (V, \mathcal{I})$, where V is a finite set, called the* ground set *and $\mathcal{I} \subseteq 2^V$ is a non-empty family of subsets, called the* independent sets*, satisfying:*

1. ***Downward-closed property:*** *If $A \in \mathcal{I}$ and $B \subseteq A$, then $B \in \mathcal{I}$.*

2. ***Exchange property:*** *If $A, B \in \mathcal{I}$ and $|A| < |B|$, then there exists an element $e \in B \setminus A$ such that $A \cup \{e\} \in \mathcal{I}$.*

A maximal independent set is called a basis*, and all bases have the same cardinality, called the* rank *of the matroid, denoted by $\rho(\mathcal{M})$.*

Definition 4 (Independence and Basis Polytopes)**.** *Let $\mathcal{M} = (V, \mathcal{I})$ be a matroid with $V = \{1, \cdots, d\}$. The* independence polytope *and* basis polytope *are defined as*

$$P_{\mathcal{I}} := \operatorname{conv}\{\mathbf{1}_S : S \in \mathcal{I}\} \subseteq [0,1]^d, \quad P_{\mathcal{B}} := \operatorname{conv}\{\mathbf{1}_B : B \text{ is a basis of } \mathcal{M}\},$$

where $\mathbf{1}_S \in \{0,1\}^d$ denotes the indicator vector of S.

The next theorem, proved in Appendix I, identifies $(P_{\mathcal{I}})^*$ and the radii r_1 and R_1 of Definition 1.

Theorem 6. *Let $\mathcal{M}$ be a nonempty matroid on ground set $V = \{1, \ldots, d\}$, and let $P_{\mathcal{I}}$ and $P_{\mathcal{B}}$ denote its independence and basis polytopes, respectively. Then $P_{\mathcal{I}}$ is a closed, downward-closed convex set, and*

$$P_{\mathcal{B}} = (P_{\mathcal{I}})^{\mathrm{m}} = (P_{\mathcal{I}})^* = P_{\mathcal{I}} \cap \{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\|_1 = \rho(\mathcal{M})\},$$
$$r_1(P_{\mathcal{I}}) = r_1(P_{\mathcal{B}}) = \rho(\mathcal{M}), \qquad R_1(P_{\mathcal{I}}) = R_1(P_{\mathcal{B}}) \le 2\rho(\mathcal{M}).$$

Combining Theorem 6 with Corollary 2 gives the following.

Corollary 4. *Let $\mathcal{M}$ be a nonempty matroid on ground set $V = \{1, \ldots, d\}$ with $\rho(\mathcal{M}) \ge 1$, and let $\mathcal{K} = P_{\mathcal{I}}$. Then maximizing any function in the class $\mathcal{F}_{\gamma,1,\sigma}$ over $\mathcal{K}$ admits both offline and online algorithms with approximation coefficient $\alpha = 1 - \exp\Big(-\frac{\gamma}{2^\sigma(\sigma+1)}\Big)$.*

Proof. By Theorem 6, we have $r_1(\mathcal{K}) = \rho(\mathcal{M})$ and $R_1(\mathcal{K}) \le 2\rho(\mathcal{M})$. Substituting $p = 1$ and $\frac{r_1}{R_1} \ge \frac{1}{2}$ into the formula for α after Corollary 2, which is increasing in r_1/R_1, yields the stated approximation coefficient. Finally, Theorem 5 establishes α-regret bounds for the class $\mathcal{F}_{\gamma,1,\sigma}$ over $\mathcal{K}$ with this value of α. □

Remark 2. *Achieving this coefficient requires set oracles for $P_{\mathcal{B}}$ (see Remark 1). Since $P_{\mathcal{B}}$ is the intersection of $P_{\mathcal{I}}$ with a hyperplane (Theorem 6), membership and separation oracles for $P_{\mathcal{B}}$ follow directly from those for $P_{\mathcal{I}}$, and other oracles, such as linear optimization oracles, can be obtained with the ellipsoid method. The algorithm may use $\overline{R}_\theta = (2\rho(\mathcal{M}))^\sigma$ in place of R_θ (Section 4), since $R_1(P_{\mathcal{I}}) \le 2\rho(\mathcal{M})$ by Theorem 6; this gives the coefficient of Corollary 4 without computing R_1.*

For $\sigma = 0$, this is the $1 - e^{-\gamma}$ coefficient for $\mathcal{F}_{\gamma,1,0} = \mathcal{F}_\gamma$ [28]. With Theorem 3, it gives the following for σ-OSS functions.

Corollary 5. *Let $\mathcal{M}$ be a nonempty matroid on ground set $V = \{1, \ldots, d\}$ with $\rho(\mathcal{M}) \ge 1$ and let $\mathcal{K} = P_{\mathcal{I}}$. Then maximizing any nonnegative monotone σ-OSS function over $\mathcal{K}$ admits both offline and online algorithms with approximation coefficient $\alpha = 1 - \exp\Big(-\frac{1}{2^{2\sigma}(2\sigma+1)}\Big)$.*

Proof. By Theorem 3, $\mathcal{F}_\sigma^{\mathrm{OSS}} \subseteq \mathcal{F}_{1,1,2\sigma}$, and Corollary 4 with $\gamma = 1$ and exponent 2σ gives the coefficient. □

For $\sigma = 0$, this is the $1 - 1/e$ coefficient for 0-OSS functions; here $\mathcal{F}_0^{\mathrm{OSS}} = C^2 \cap \mathcal{F}_{1,1,0} = C^2 \cap \mathcal{F}_1$, and Corollary 4 covers all of $\mathcal{F}_{1,1,0} = \mathcal{F}_1$ [28]. For general σ, the closest results are those of Ghadiri et al. [14] for monotone σ-OSS functions over downward-closed polytopes, with coefficient $\alpha_2(\sigma) = \sup_{t\in[0,1)} \big(1 - \exp\big(-(1-t)\big(\frac{t}{t+1}\big)^{2\sigma}\big)\big)$. Corollary 5 gives $\alpha_1(\sigma) = 1 - \exp\big(-\frac{1}{2^{2\sigma}(2\sigma+1)}\big)$. Appendix J shows that $\alpha_1(\sigma) \ge \alpha_2(\sigma)$, with equality only at $\sigma = 0$, and that $\alpha_1(\sigma)/\alpha_2(\sigma) \to e/2$ as $\sigma \to \infty$. This comparison uses $r_1/R_1 \ge 1/2$ for matroid polytopes (Theorem 6). For a general downward-closed polytope $\mathcal{K}$, Theorem 3 and Corollary 2 give $1 - \exp\big(-\frac{1}{2\sigma+1}\big(\frac{r_1(\mathcal{K})}{R_1(\mathcal{K})}\big)^{2\sigma}\big)$, which,

for $\sigma > 0$, is smaller than $\alpha_2(\sigma)$ when $r_1(\mathcal{K})/R_1(\mathcal{K})$ is sufficiently small. Ghadiri et al. [14] also give a separate guarantee under an additional third-order derivative-sign condition; the comparison above is for general monotone σ-OSS functions. All coefficients in this section concern the continuous problem over $P_{\mathcal{I}}$.

Zhang et al. [37] study online optimization of monotone OSS functions. Their Frank–Wolfe-type algorithm recovers the ratio of [14] under an additional smoothness condition, and their online-gradient-ascent algorithm, which does not need it, has ratio at most $\alpha_3(\sigma) = \sup_{t\in(0,1)} \big(1 + (\frac{t+1}{t})^{4\sigma}\big)^{-1} = (1 + 2^{4\sigma})^{-1}$. We have $\alpha_1(\sigma) > \alpha_3(\sigma)$ for all $\sigma \geq 0$, and $\alpha_1(\sigma)/\alpha_3(\sigma) \to \infty$ as $\sigma \to \infty$.

7 Results for Other Settings

This section applies the meta-algorithms of [29, 28] to Theorem 5 with existing base algorithms. For a single fixed function, `OTB` (Theorem 8 in [28]) gives an offline algorithm, and `SFTT` converts full-information algorithms into semi-bandit ones. We use the version of `SFTT` in Theorem 7 of [28], whose conditions hold here: $\mathcal{F}_{\gamma,\theta}$ is closed under convex combinations (Lemma 1), $\texttt{OMBQ}(\mathcal{A}, \mathcal{G}, \mathrm{Id})$ makes one query per round that does not depend on that round's observations, and f is B_1-Lipschitz on $\mathcal{K}$ when $\mathcal{Q}_f$ is bounded by B_1.

We use two deterministic semi-bandit base algorithms. Separation-Oracle Online Gradient Ascent (`SO-OGA`) is projection-free and has optimal adaptive regret [13, 28]. As in [28], it runs within the affine hull of $\mathcal{K}^*$, which may have empty interior (as for the basis polytope), and is given affine coordinates, a point $\mathbf{c} \in \mathrm{relint}(\mathcal{K}^*)$, a radius $r > 0$ with $B_r(\mathbf{c}) \cap \mathrm{aff}(\mathcal{K}^*) \subseteq \mathcal{K}^*$, and a deterministic separation oracle; the hidden constants below depend on these. If $\mathcal{K}^*$ is a single point, or if $B_1 = 0$ (so that every f is constant on $\mathcal{K}$), the problem is trivial. Improved Ader (`IA`) has dynamic regret with optimal dependence on the path length [39]. Applying the framework to `SO-OGA` and `IA` gives the following bounds; the proofs follow Corollaries 7 and 8 in [28].

Theorem 7. *Under the assumptions of Theorem 5, we have $\mathcal{AR}^{\mathcal{A}}_{\alpha,\mathrm{Adv}^{\circ}_1(\mathcal{F},B_1)} \leq \mathcal{O}(B_1 T^{1/2})$, where $\mathcal{A} = \texttt{OMBQ}(\texttt{SO-OGA}, \mathcal{G}, \mathrm{Id})$ is a first-order full-information algorithm with a single query per round. Moreover, $\mathcal{AR}^{\mathcal{A}_{\text{semi-bandit}}}_{\alpha,\mathrm{Adv}^{\circ}_1(\mathcal{F},B_1)} \leq \mathcal{O}(B_1 T^{2/3})$, where $\mathcal{A}_{\text{semi-bandit}} = \texttt{SFTT}(\mathcal{A})$. Moreover, the sample complexity of the offline first-order algorithm $\texttt{OTB}(\mathcal{A})$ is $\mathcal{O}(\epsilon^{-2})$; here $\texttt{OTB}(\mathcal{A})$ outputs an iterate X chosen uniformly at random, and the guarantee is $\mathbb{E}[f(X)] \geq \alpha \max_{\mathbf{x}\in\mathcal{K}} f(\mathbf{x}) - \epsilon$.*

Theorem 8. *Under the assumptions of Theorem 5, we have*

$$\mathcal{R}^{\mathcal{A}}_{\alpha,\mathrm{Adv}^{\circ}_1(\mathcal{F},B_1)}(\mathbf{u}) = \mathcal{O}(B_1\sqrt{T(1 + P_T(\mathbf{u}))})$$

for every comparator sequence $\mathbf{u} \in (\mathcal{K}^*)^T$, where $\mathcal{A} = \texttt{OMBQ}(\texttt{IA}, \mathcal{G}, \mathrm{Id})$ is also a first-order full-information algorithm with a single query per round.

8 Conclusion

We introduced γ-weakly θ-up-concave functions, a class that strictly contains monotone DR-submodular and OSS functions and includes functions that are flat at the origin or have accumulating-then-diminishing returns. These functions are upper-linearizable, which gives offline and online guarantees and, under matroid constraints, improved coefficients for general monotone OSS objectives. Whether these coefficients are tight is an open question.

Acknowledgments and Disclosure of Funding

This work was in part supported by U.S. National Science Foundation under grant CCF-2149588, IVADO and CIFAR.

References

[1] Abbassi, Z., Mirrokni, V. S., and Thakur, M. (2013). Diversity maximization under matroid constraints. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 32–40.

[2] Bedi, A. S., Rajawat, K., Aggarwal, V., and Koppel, A. (2021). Escaping saddle points for successive convex approximation. *IEEE Transactions on Signal Processing*, 70:307–321.

[3] Benson, H. P. (1998). An Outer Approximation Algorithm for Generating All Efficient Extreme Points in the Outcome Set of a Multiple Objective Linear Programming Problem. *Journal of Global Optimization*, 13(1):1–24.

[4] Beznosikov, A., Samokhin, V., and Gasnikov, A. (2025). Distributed saddle point problems: lower bounds, near-optimal and robust algorithms. *Optimization Methods and Software*, pages 1–18.

[5] Bian, A., Levy, K., Krause, A., and Buhmann, J. M. (2017a). Continuous DR-submodular maximization: Structure and algorithms. In *Advances in Neural Information Processing Systems*.

[6] Bian, A. A., Mirzasoleiman, B., Buhmann, J., and Krause, A. (2017b). Guaranteed Non-convex Optimization: Submodular Maximization over Continuous Domains. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*.

[7] Bian, Y., Buhmann, J., and Krause, A. (2019). Optimal continuous dr-submodular maximization and applications to provable mean field inference. In *International Conference on Machine Learning*, pages 644–653. PMLR.

[8] Calinescu, G., Chekuri, C., Pál, M., and Vondrák, J. (2011). Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766.

[9] Chen, L., Hassani, H., and Karbasi, A. (2018). Online continuous submodular maximization. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*.

[10] Chen, S., Du, D., Yang, W., Xu, D., and Gao, S. (2023). Continuous non-monotone DR-submodular maximization with down-closed convex constraint. *arXiv preprint arXiv:2307.09616*.

[11] Dixit, R. (2024). *Gradient Based Methods for Nonconvex Optimization: A Dynamical Systems Viewpoint*. PhD thesis, Rutgers The State University of New Jersey, School of Graduate Studies.

[12] Fazel, M. and Sadeghi, O. (2023). Fast first-order methods for monotone strongly dr-submodular maximization. In *SIAM Conference on Applied and Computational Discrete Algorithms (ACDA23)*.

[13] Garber, D. and Kretzu, B. (2022). New projection-free algorithms for online convex optimization with adaptive regret guarantees. In *Proceedings of Thirty Fifth Conference on Learning Theory*, pages 2326–2359. PMLR.

[14] Ghadiri, M., Santiago, R., and Shepherd, B. (2025). Beyond submodular maximization via one-sided smoothness. *Mathematical Programming*, pages 1–37.

[15] Hassani, H., Karbasi, A., Mokhtari, A., and Shen, Z. (2020). Stochastic conditional gradient++: (non)convex minimization and continuous submodular maximization. *SIAM Journal on Optimization*, 30(4):3315–3344.

[16] Hassani, H., Soltanolkotabi, M., and Karbasi, A. (2017). Gradient methods for submodular maximization. In *Advances in Neural Information Processing Systems*.

[17] Hazan, E. and Seshadhri, C. (2009). Efficient learning algorithms for changing environments. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 393–400. Association for Computing Machinery.

[18] Lee, D., Ho-Nguyen, N., and Lee, D. (2023). Non-smooth, hölder-smooth, and robust submodular maximization. *arXiv preprint arXiv:2210.06061*.

[19] Liao, Y., Wan, Y., Yao, C., and Song, M. (2023). Improved Projection-free Online Continuous Submodular Maximization. *arXiv preprint arXiv:2305.18442*.

[20] Lu, Y., Jadav, H., Pedramfar, M., Singh, R., and Aggarwal, V. (2026). Upper linearizability of online non-monotone dr-submodular maximization over down-closed convex sets. In *Proceedings of the International Conference on Machine Learning (ICML)*.

[21] Lu, Y., Pedramfar, M., and Aggarwal, V. (2025). Decentralized projection-free online upper-linearizable optimization with applications to DR-submodular optimization. *Transactions on Machine Learning Research*.

[22] Löhne, A. (2011). *Vector Optimization with Infimum and Supremum*. Springer Science & Business Media.

[23] Mirzasoleiman, B., Jegelka, S., and Krause, A. (2018). Streaming non-monotone submodular maximization: Personalized video summarization on the fly. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

[24] Mitra, S., Feldman, M., and Karbasi, A. (2021). Submodular+ concave. *Advances in Neural Information Processing Systems*.

[25] Mokhtari, A., Hassani, H., and Karbasi, A. (2020). Stochastic conditional gradient methods: From convex minimization to submodular maximization. *The Journal of Machine Learning Research*, 21(1):4232–4280.

[26] Mualem, L. and Feldman, M. (2023). Resolving the approximability of offline and online non-monotone DR-submodular maximization over general convex sets. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*.

[27] Niazadeh, R., Golrezaei, N., Wang, J., Susan, F., and Badanidiyuru, A. (2023). Online learning via offline greedy algorithms: Applications in market design and optimization. *Management Science*, 69(7):3797–3817.

[28] Pedramfar, M. and Aggarwal, V. (2024). From linear to linearizable optimization: A novel framework with applications to stationary and non-stationary DR-submodular optimization. *Advances in Neural Information Processing Systems*, 37:37626–37664. Latest version available at arXiv:2405.00065.

[29] Pedramfar, M. and Aggarwal, V. (2026). A unified framework for analyzing meta-algorithms in online convex optimization. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[30] Pedramfar, M., Nadew, Y. Y., Quinn, C. J., and Aggarwal, V. (2024a). Unified projection-free algorithms for adversarial DR-submodular optimization. In *The Twelfth International Conference on Learning Representations*.

[31] Pedramfar, M., Quinn, C. J., and Aggarwal, V. (2023). A unified approach for maximizing continuous DR-submodular functions. In *Thirty-seventh Conference on Neural Information Processing Systems*.

[32] Pedramfar, M., Quinn, C. J., and Aggarwal, V. (2024b). A unified approach for maximizing continuous γ-weakly dr-submodular functions. Technical report, Optimization Online, `https://optimization-online.org/?p=25915`.

[33] Pedramfar, M., Quinn, C. J., and Aggarwal, V. (2025). Uniform wrappers: Bridging concave to quadratizable functions in online optimization. In *Advances in Neural Information Processing Systems*, volume 38.

[34] Umrawal, A. K., Quinn, C. J., and Aggarwal, V. (2023). A community-aware framework for social influence maximization. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(4):1253–1262.

[35] Wan, Z., Zhang, J., Chen, W., Sun, X., and Zhang, Z. (2023). Bandit multi-linear dr-submodular maximization and its applications on adversarial submodular bandits. In *International Conference on Machine Learning*.

[36] Wilder, B. (2018). Equilibrium computation and robust optimization in zero sum games with submodular structure. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

[37] Zhang, H., Xu, D., Gai, L., and Zhang, Z. (2024a). Online learning under one sided σ-smooth function. *Journal of Combinatorial Optimization*, 47(5):76.

[38] Zhang, H., Xu, D., Zhang, Y., and Zhang, Z. (2023a). Parallelized maximization of nonmonotone one-sided σ-smooth function. *Computers and Electrical Engineering*, 105:108478.

[39] Zhang, L., Lu, S., and Zhou, Z.-H. (2018). Adaptive online learning in dynamic environments. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

[40] Zhang, M., Chen, L., Hassani, H., and Karbasi, A. (2019). Online continuous submodular maximization: From full-information to bandit feedback. In *Advances in Neural Information Processing Systems*, volume 32.

[41] Zhang, Q., Deng, Z., Chen, Z., Hu, H., and Yang, Y. (2022). Stochastic continuous submodular maximization: Boosting via non-oblivious function. In *Proceedings of the 39th International Conference on Machine Learning*.

[42] Zhang, Q., Deng, Z., Chen, Z., Zhou, K., Hu, H., and Yang, Y. (2023b). Online learning for non-monotone DR-submodular maximization: From full information to bandit feedback. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*.

[43] Zhang, Q., Wan, Z., Deng, Z., Chen, Z., Sun, X., Zhang, J., and Yang, Y. (2024b). Boosting gradient ascent for continuous DR-submodular maximization. *arXiv preprint arXiv:2401.08330*.

[44] Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pages 928–936.

A Related Work

DR-submodular and up-concave maximization. Maximization of continuous DR-submodular functions over convex sets has been studied in offline and online settings with two main approaches. Frank–Wolfe-type methods have been adapted to stochastic and online settings [5, 6, 26, 31, 10, 25, 15, 9, 27, 40, 42, 30]. Boosting methods maximize an auxiliary objective whose maximizers approximately maximize the original function [16, 9, 12, 41, 35, 19, 43].

Continuous DR-submodular functions are concave along positive directions [8, 6], which led Wilder [36] to define *up-concave* functions. Offline maximization of up-concave functions is studied in [18]. Our class contains monotone DR-submodular and up-concave functions, and our guarantees come from upper-linearization rather than from problem-specific algorithms.

One-sided smoothness (OSS). Ghadiri et al. [14] introduced one-sided smoothness for diversity-type objectives and other monotone nonconcave functions over downward-closed polytopes. It strictly generalizes twice-differentiable DR-submodularity, and Ghadiri et al. [14] obtain offline guarantees with variants of continuous greedy. Zhang et al. [37] give online algorithms and regret bounds for OSS objectives. Under matroid constraints, our coefficient is strictly larger than the general OSS coefficient of [14] for $\sigma > 0$, with asymptotic ratio $e/2$, and has better dependence on σ than the online-gradient-ascent coefficient of [37]. Our class strictly contains the OSS class, and our matroid coefficients need no third-order conditions; Section 6 gives the detailed comparison. Non-monotone OSS functions are studied in [38]. Outside these structured classes, most results in continuous nonconcave optimization concern first- or second-order stationarity rather than global approximation ratios [2, 11, 4].

Upper-linearizable framework. Pedramfar and Aggarwal [28] introduced upper-linearizable functions and meta-algorithms that convert algorithms for linear optimization into algorithms for nonconcave objectives, including DR-submodular and up-concave functions. The framework transfers offline approximation guarantees and static, dynamic, and adaptive regret bounds, under first-order, zeroth-order, semi-bandit, and bandit feedback. Later work gave uniform wrapper constructions for regret conversions, improved guarantees under limited feedback, and decentralized and projection-free versions [33, 21, 20]. We show that a broader class of monotone objectives is upper-linearizable, so the first-order consequences (full-information and semi-bandit) apply to it. The zeroth-order and bandit conversions of [28] also require the smoothed functions to remain in the class, which we have not shown for general θ.

B Proof of Theorem 2

Proof. Fix $\mathbf{x}, \mathbf{y} \in [0,1]^d$ with $\mathbf{y} \geq \mathbf{x} \neq \mathbf{0}$, and define $\mathbf{z}_t = \mathbf{x} + t(\mathbf{y} - \mathbf{x})$ for $t \in [0,1]$. Since f is C^1, the fundamental theorem of calculus applied to $t \mapsto f(\mathbf{z}_t)$ gives

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle \nabla f(\mathbf{z}_t), \mathbf{y} - \mathbf{x} \rangle dt. \tag{20}$$

Since $\mathbf{x} \leq \mathbf{z}_t \leq \mathbf{y}$ coordinate-wise and θ is monotone,

$$0 < \theta(\mathbf{x}) \leq \theta(\mathbf{z}_t) \leq \theta(\mathbf{y}), \tag{21}$$

where $\theta(\mathbf{x}) > 0$ because $\mathbf{x} \neq \mathbf{0}$. In particular $\mathbf{z}_t \neq \mathbf{0}$, so (4) may be applied to the pairs $(\mathbf{z}_t, \mathbf{y})$ and $(\mathbf{x}, \mathbf{z}_t)$. This gives, coordinate-wise,

$$\frac{\gamma \nabla f(\mathbf{y})}{\theta(\mathbf{y})} \leq \frac{\nabla f(\mathbf{z}_t)}{\theta(\mathbf{z}_t)} \qquad \text{and} \qquad \frac{\gamma \nabla f(\mathbf{z}_t)}{\theta(\mathbf{z}_t)} \leq \frac{\nabla f(\mathbf{x})}{\theta(\mathbf{x})}. \tag{22}$$

Since f is monotone, all its partial derivatives are nonnegative, so $\nabla f(\mathbf{x}) \geq \mathbf{0}$ and $\nabla f(\mathbf{y}) \geq \mathbf{0}$. Multiplying the first inequality in (22) by $\theta(\mathbf{z}_t) > 0$, and then using $\theta(\mathbf{z}_t) \geq \theta(\mathbf{x})$ from (21) and $\nabla f(\mathbf{y}) \geq \mathbf{0}$, we get

$$\nabla f(\mathbf{z}_t) \geq \frac{\gamma \theta(\mathbf{z}_t)}{\theta(\mathbf{y})} \nabla f(\mathbf{y}) \geq \frac{\gamma \theta(\mathbf{x})}{\theta(\mathbf{y})} \nabla f(\mathbf{y}). \tag{23}$$

Multiplying the second inequality in (22) by $\theta(\mathbf{z}_t)/\gamma > 0$, and then using $\theta(\mathbf{z}_t) \le \theta(\mathbf{y})$ from (21) and $\nabla f(\mathbf{x}) \ge \mathbf{0}$, we get

$$\nabla f(\mathbf{z}_t) \le \frac{\theta(\mathbf{z}_t)}{\gamma\theta(\mathbf{x})}\nabla f(\mathbf{x}) \le \frac{\theta(\mathbf{y})}{\gamma\theta(\mathbf{x})}\nabla f(\mathbf{x}). \tag{24}$$

Since $\mathbf{y} - \mathbf{x} \ge \mathbf{0}$, taking the inner product of the coordinate-wise inequalities (23) and (24) with $\mathbf{y} - \mathbf{x}$ preserves them:

$$\frac{\gamma\theta(\mathbf{x})}{\theta(\mathbf{y})}\langle\nabla f(\mathbf{y}), \mathbf{y} - \mathbf{x}\rangle \le \langle\nabla f(\mathbf{z}_t), \mathbf{y} - \mathbf{x}\rangle \le \frac{\theta(\mathbf{y})}{\gamma\theta(\mathbf{x})}\langle\nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle.$$

The outer terms do not depend on t. Integrating over $t \in [0, 1]$ and using (20) gives (2). Since f is also C^1, nonnegative and monotone, f is γ-weakly θ-up-concave. □

C Proof of Theorem 3

Proof. We prove the two inclusions separately. Throughout, we use that $\|\mathbf{x} + \mathbf{u}\|_1 = \|\mathbf{x}\|_1 + \|\mathbf{u}\|_1$ for $\mathbf{x}, \mathbf{u} \ge \mathbf{0}$, since the ℓ_1 norm of a nonnegative vector is the sum of its coordinates.

Proof of $C^2 \cap \mathcal{F}_{1,1,\sigma} \subseteq \mathcal{F}_\sigma^{\mathrm{OSS}}$. Let $f \in C^2 \cap \mathcal{F}_{1,1,\sigma}$. By definition, f is nonnegative and monotone, so it remains to verify (7). Fix $\mathbf{x} \in (0, 1)^d$ and $\mathbf{u} \ge \mathbf{0}$. Choose $\varepsilon > 0$ such that $\mathbf{x} + t\mathbf{u} \in (0, 1)^d$ for all $t \in [0, \varepsilon)$, and define $g(t) = f(\mathbf{x} + t\mathbf{u})$ for $t \in [0, \varepsilon)$. By the chain rule,

$$g'(t) = \langle\nabla f(\mathbf{x} + t\mathbf{u}), \mathbf{u}\rangle, \qquad g''(t) = \mathbf{u}^\top\nabla^2 f(\mathbf{x} + t\mathbf{u})\mathbf{u}. \tag{25}$$

Let $h \in (0, \varepsilon)$. The pair $(\mathbf{x}, \mathbf{x} + h\mathbf{u})$ satisfies $\mathbf{x} + h\mathbf{u} \ge \mathbf{x} \ne \mathbf{0}$, so the upper inequality in (3) with $\gamma = 1$ and $p = 1$ gives

$$g(h) - g(0) \le \left(\frac{\|\mathbf{x} + h\mathbf{u}\|_1}{\|\mathbf{x}\|_1}\right)^\sigma \langle\nabla f(\mathbf{x}), h\mathbf{u}\rangle = (1 + hc)^\sigma\, h\, g'(0), \qquad \text{where } c := \frac{\|\mathbf{u}\|_1}{\|\mathbf{x}\|_1}, \tag{26}$$

and we used $\|\mathbf{x} + h\mathbf{u}\|_1 = \|\mathbf{x}\|_1 + h\|\mathbf{u}\|_1$ and (25) at $t = 0$. Since f is C^2, Taylor's theorem gives

$$g(h) - g(0) = hg'(0) + \tfrac{1}{2}h^2 g''(0) + o(h^2) \qquad \text{as } h \downarrow 0. \tag{27}$$

Substituting (27) into (26), subtracting $hg'(0)$ from both sides and dividing by $h^2 > 0$, we obtain

$$\tfrac{1}{2}g''(0) + \frac{o(h^2)}{h^2} \le \frac{(1 + hc)^\sigma - 1}{h}\, g'(0). \tag{28}$$

As $h \downarrow 0$, the fraction on the right of (28) converges to the derivative of $h \mapsto (1 + hc)^\sigma$ at $h = 0$, which is σc, and the term $o(h^2)/h^2$ converges to 0. Taking the limit in (28) and using (25) at $t = 0$ gives

$$\frac{1}{2}\mathbf{u}^\top\nabla^2 f(\mathbf{x})\mathbf{u} \le \sigma\frac{\|\mathbf{u}\|_1}{\|\mathbf{x}\|_1}\langle\nabla f(\mathbf{x}), \mathbf{u}\rangle, \tag{29}$$

which is (7) for $\mathbf{x} \in (0, 1)^d$. Now let $\mathbf{x} \in [0, 1]^d \setminus \{\mathbf{0}\}$ be arbitrary, and let $\mathbf{x}_k = (1 - \frac{1}{k})\mathbf{x} + \frac{1}{2k}\mathbf{1}$ for $k \ge 2$. Each coordinate of $\mathbf{x}_k$ lies in $[\frac{1}{2k}, 1 - \frac{1}{2k}]$, so $\mathbf{x}_k \in (0, 1)^d$, and $\mathbf{x}_k \to \mathbf{x}$. Inequality (29) holds at every $\mathbf{x}_k$. Both sides of (29) are continuous in $\mathbf{x}$ on $[0, 1]^d \setminus \{\mathbf{0}\}$, because ∇f and $\nabla^2 f$ are continuous and $\|\mathbf{x}\|_1 > 0$ there. Letting $k \to \infty$ shows that (29) holds at $\mathbf{x}$. Hence $f \in \mathcal{F}_\sigma^{\mathrm{OSS}}$.

Proof of $\mathcal{F}_\sigma^{\mathrm{OSS}} \subseteq \mathcal{F}_{1,1,2\sigma}$. Let $f \in \mathcal{F}_\sigma^{\mathrm{OSS}}$. Then f is nonnegative, monotone and C^2, so it remains to verify (3) with $\gamma = 1$, $p = 1$ and exponent 2σ. Fix $\mathbf{x}, \mathbf{y} \in [0, 1]^d$ with $\mathbf{y} \ge \mathbf{x} \ne \mathbf{0}$. For $t \in [0, 1]$, define

$$\phi(t) = \mathbf{x} + t(\mathbf{y} - \mathbf{x}), \qquad g(t) = f(\phi(t)), \qquad N(t) = \|\phi(t)\|_1.$$

Since $\mathbf{x} \ge \mathbf{0}$ and $\mathbf{y} - \mathbf{x} \ge \mathbf{0}$, we have

$$N(t) = \|\mathbf{x}\|_1 + t\|\mathbf{y} - \mathbf{x}\|_1, \qquad N'(t) = \|\mathbf{y} - \mathbf{x}\|_1, \qquad \|\mathbf{x}\|_1 = N(0) \le N(t) \le N(1) = \|\mathbf{y}\|_1, \tag{30}$$

and in particular $N(t) > 0$, so $\phi(t) \ne \mathbf{0}$. By the chain rule,

$$g'(t) = \langle\nabla f(\phi(t)), \mathbf{y} - \mathbf{x}\rangle, \qquad g''(t) = (\mathbf{y} - \mathbf{x})^\top\nabla^2 f(\phi(t))(\mathbf{y} - \mathbf{x}). \tag{31}$$

Since f is monotone, all its partial derivatives are nonnegative, and since $\mathbf{y} - \mathbf{x} \geq \mathbf{0}$, (31) gives

$$g'(t) \geq 0 \qquad \text{for all } t \in [0,1]. \tag{32}$$

Applying (7) at the point $\phi(t) \neq \mathbf{0}$ with $\mathbf{u} = \mathbf{y} - \mathbf{x} \geq \mathbf{0}$, and using (30) and (31), we get

$$g''(t) \leq 2\sigma \frac{N'(t)}{N(t)} g'(t). \tag{33}$$

Define $q(t) = g'(t)/N(t)^{2\sigma}$. By the quotient rule,

$$q'(t) = \frac{g''(t)N(t)^{2\sigma} - 2\sigma N(t)^{2\sigma-1}N'(t)g'(t)}{N(t)^{4\sigma}} = \frac{1}{N(t)^{2\sigma}} \left(g''(t) - 2\sigma \frac{N'(t)}{N(t)} g'(t) \right),$$

and by (33), $q'(t) \leq 0$. Hence q is nonincreasing on $[0,1]$, so

$$q(1) \leq q(t) \leq q(0) \qquad \text{for all } t \in [0,1]. \tag{34}$$

For the upper bound, the right inequality in (34), together with $q(0) \geq 0$ (by (32)) and $N(t) \leq N(1)$ (by (30)), gives

$$g'(t) = N(t)^{2\sigma} q(t) \leq N(t)^{2\sigma} q(0) \leq N(1)^{2\sigma} q(0) = \left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{x}\|_1} \right)^{2\sigma} g'(0). \tag{35}$$

For the lower bound, the left inequality in (34), together with $q(1) \geq 0$ and $N(t) \geq N(0)$, gives

$$g'(t) = N(t)^{2\sigma} q(t) \geq N(t)^{2\sigma} q(1) \geq N(0)^{2\sigma} q(1) = \left(\frac{\|\mathbf{x}\|_1}{\|\mathbf{y}\|_1} \right)^{2\sigma} g'(1). \tag{36}$$

By the fundamental theorem of calculus, $f(\mathbf{y}) - f(\mathbf{x}) = g(1) - g(0) = \int_0^1 g'(t)dt$. Integrating (35) and (36) over $t \in [0,1]$ and using (31) at $t = 0$ and $t = 1$, we obtain

$$\left(\frac{\|\mathbf{x}\|_1}{\|\mathbf{y}\|_1} \right)^{2\sigma} \langle \nabla f(\mathbf{y}), \mathbf{y} - \mathbf{x} \rangle \leq f(\mathbf{y}) - f(\mathbf{x}) \leq \left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{x}\|_1} \right)^{2\sigma} \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle,$$

which is (3) with $\gamma = 1$, $p = 1$ and exponent 2σ. Hence $f \in \mathcal{F}_{1,1,2\sigma}$. □

Remark 3. *In the proof above, the inequality $q(1) \leq q(0)$ from* (34)*, i.e.,*

$$\frac{g'(1)}{\|\mathbf{y}\|_1^{2\sigma}} \leq \frac{g'(0)}{\|\mathbf{x}\|_1^{2\sigma}}$$

directly implies that

$$\langle \nabla f(\mathbf{y}), \mathbf{y} - \mathbf{x} \rangle \leq \left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{x}\|_1} \right)^{2\sigma} \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle.$$

Proposition 2. *For every $\sigma > 0$, $\beta \in [0, 2\sigma)$ and $\gamma \in (0,1]$, there is a nonnegative, monotone, C^2, σ-OSS function on $[0,1]^d$ that is not γ-weakly $(1,\beta)$-up-concave. In particular, the exponent 2σ in Theorem 3 cannot be replaced by a smaller one.*

Proof. Fix $\sigma > 0$, $\beta \in [0, 2\sigma)$ and $\gamma \in (0,1]$. For $\delta \in (0,1]$, define

$$f_\delta(\mathbf{x}) = (\mathbf{1}^\top \mathbf{x} + \delta)^{1+2\sigma} - \delta^{1+2\sigma}, \qquad \mathbf{x} \in [0,1]^d. \tag{37}$$

Write $s = \mathbf{1}^\top \mathbf{x}$; on $[0,1]^d$ we have $s = \|\mathbf{x}\|_1$. Since $s + \delta \geq \delta > 0$, the function f_δ is C^2 on $[0,1]^d$, and

$$\nabla f_\delta(\mathbf{x}) = (1+2\sigma)(s+\delta)^{2\sigma} \mathbf{1}, \qquad \nabla^2 f_\delta(\mathbf{x}) = 2\sigma(1+2\sigma)(s+\delta)^{2\sigma-1} \mathbf{1}\mathbf{1}^\top. \tag{38}$$

Since $s \geq 0$, we have $f_\delta \geq 0$, and since $\nabla f_\delta \geq \mathbf{0}$, the function f_δ is monotone.

We first show that f_δ is σ-OSS. Fix $\mathbf{x} \neq \mathbf{0}$, so that $s > 0$, and $\mathbf{u} \geq \mathbf{0}$, so that $\mathbf{1}^\top \mathbf{u} = \|\mathbf{u}\|_1$. By (38),

$$\frac{1}{2} \mathbf{u}^\top \nabla^2 f_\delta(\mathbf{x}) \mathbf{u} = \sigma(1+2\sigma)(s+\delta)^{2\sigma-1} \|\mathbf{u}\|_1^2, \tag{39}$$

$$\sigma \frac{\|\mathbf{u}\|_1}{\|\mathbf{x}\|_1} \mathbf{u}^\top \nabla f_\delta(\mathbf{x}) = \sigma(1+2\sigma) \frac{(s+\delta)^{2\sigma}}{s} \|\mathbf{u}\|_1^2. \tag{40}$$

The right-hand side of (39) equals $s/(s+\delta)$ times the right-hand side of (40). Since the latter is nonnegative and $0 < s/(s+\delta) \le 1$, (39) is at most (40), which is (7). This also covers $\mathbf{u} = \mathbf{0}$, where both sides are 0. So $f_\delta \in \mathcal{F}_\sigma^{\mathrm{OSS}}$ for every $\delta \in (0, 1]$.

Next, suppose that f_δ is γ-weakly $(1, \beta)$-up-concave for some $\delta \in (0, 1]$. Take $\mathbf{x} = \delta \mathbf{e}_1$ and $\mathbf{y} = \mathbf{1}$, so that $\mathbf{y} \ge \mathbf{x} \ne \mathbf{0}$, $\|\mathbf{x}\|_1 = \delta$ and $\|\mathbf{y}\|_1 = d$. By (37), the left-hand side of the upper inequality in (3) is

$$f_\delta(\mathbf{y}) - f_\delta(\mathbf{x}) = (d+\delta)^{1+2\sigma} - (2\delta)^{1+2\sigma}, \tag{41}$$

and, by (38) with $s = \delta$ and $\langle \mathbf{1}, \mathbf{y} - \mathbf{x} \rangle = d - \delta$, its right-hand side is

$$\frac{1}{\gamma}\left(\frac{d}{\delta}\right)^{\beta} (1+2\sigma)(2\delta)^{2\sigma}(d-\delta) = \frac{(1+2\sigma)2^{2\sigma} d^{\beta}(d-\delta)}{\gamma}\, \delta^{2\sigma-\beta}. \tag{42}$$

The upper inequality therefore requires (41) to be at most (42). As $\delta \downarrow 0$, (41) converges to $d^{1+2\sigma} > 0$, while (42) converges to 0 because $2\sigma - \beta > 0$. Hence the inequality fails for all sufficiently small δ, and for such δ the function f_δ is not γ-weakly $(1, \beta)$-up-concave. □

D Examples Beyond DR-Submodular and OSS Functions

A function has accumulating-then-diminishing returns if its marginal reward first increases and then decreases. Such returns already occur for γ-weakly $(1, \sigma)$-up-concave and σ-OSS functions (Lemma 2). Flatness at the origin (Section 3) provides a separation from these classes: by Lemma 3, a nonzero function in $\mathcal{F}_{\gamma,\theta}$ that is flat at the origin is neither γ'-weakly (p, σ)-up-concave nor σ-OSS for any admissible parameters. Lemmas 4 and 5 construct such functions with a scaling θ that is itself flat at the origin; these functions depend on $\mathbf{x}$ only through $\|\mathbf{x}\|_1$ or $\mathbf{a}^\top \mathbf{x}$. Proposition 3 gives a multivariate family whose members need not depend on $\mathbf{x}$ through a single linear statistic.

Lemma 2. *Fix $\sigma \ge 1$ and $M > 0$, and define*

$$f(\mathbf{x}) = \int_0^{\|\mathbf{x}\|_1} t^\sigma e^{-Mt}\, dt.$$

Then $f \in \mathcal{F}_{1,1,\sigma} \cap \mathcal{F}_\sigma^{\mathrm{OSS}}$. Moreover, f exhibits monotone accumulating-then-diminishing returns: its marginal reward as a function of $s = \|\mathbf{x}\|_1$ is increasing for $s < \sigma/M$ and decreasing for $s > \sigma/M$. Both phases occur on $[0, 1]^d$ when $M > \sigma/d$.

Proof. Let $\theta(\mathbf{x}) = \|\mathbf{x}\|_1^\sigma$, and define $F(s) = \int_0^s t^\sigma e^{-Mt} dt$ for $s \ge 0$, so that $f(\mathbf{x}) = F(\|\mathbf{x}\|_1)$. On $[0, 1]^d$ we have $\|\mathbf{x}\|_1 = \mathbf{1}^\top \mathbf{x}$. By the chain rule and the fundamental theorem of calculus, with $s = \|\mathbf{x}\|_1$,

$$\nabla f(\mathbf{x}) = F'(s)\mathbf{1} = s^\sigma e^{-Ms}\mathbf{1}. \tag{43}$$

Thus f is C^1, $f \ge 0$, and f is monotone because $\nabla f \ge \mathbf{0}$. For $\mathbf{x} \ne \mathbf{0}$, (43) gives

$$\frac{\nabla f(\mathbf{x})}{\theta(\mathbf{x})} = e^{-Ms}\mathbf{1}. \tag{44}$$

If $\mathbf{y} \ge \mathbf{x} \ne \mathbf{0}$, then $\|\mathbf{y}\|_1 \ge \|\mathbf{x}\|_1$, so by (44), $\nabla f(\mathbf{y})/\theta(\mathbf{y}) \le \nabla f(\mathbf{x})/\theta(\mathbf{x})$ coordinate-wise. This is (4) with $\gamma = 1$, and Theorem 2 gives $f \in \mathcal{F}_{1,1,\sigma}$.

Next, for $s > 0$ we have $F''(s) = s^{\sigma-1} e^{-Ms}(\sigma - Ms)$. Since $\sigma \ge 1$, the factor $s^{\sigma-1}$ is continuous on $[0, \infty)$, so F'' extends continuously to $s = 0$ and F is C^2 on $[0, \infty)$. Hence f is C^2 on $[0, 1]^d$, with $\nabla^2 f(\mathbf{x}) = F''(\|\mathbf{x}\|_1)\mathbf{1}\mathbf{1}^\top$. Since $f \in C^2 \cap \mathcal{F}_{1,1,\sigma}$, Theorem 3 gives $f \in \mathcal{F}_\sigma^{\mathrm{OSS}}$.

Finally, the marginal reward along $s = \|\mathbf{x}\|_1$ is $F'(s) = s^\sigma e^{-Ms}$, and for $s > 0$,

$$\frac{d}{ds} \log F'(s) = \frac{\sigma}{s} - M. \tag{45}$$

By (45), F' is increasing for $s < \sigma/M$ and decreasing for $s > \sigma/M$. On $[0, 1]^d$, s ranges over $[0, d]$, so both phases occur when $\sigma/M < d$, that is, when $M > \sigma/d$. □

Recall from Section 3 that $\phi : [0, 1]^d \to \mathbb{R}_{\ge 0}$ is flat at the origin if $\lim_{\mathbf{x} \to \mathbf{0}} \phi(\mathbf{x})/\|\mathbf{x}\|_1^n = 0$ for all $n > 0$.

Lemma 3. *Let $f \in \mathcal{F}_{\gamma,\theta}$ for some $\gamma \in (0,1]$ and θ and suppose that f is flat at the origin and not identically zero. Then θ is flat at the origin. In particular, f is not γ'-weakly (p,σ)-up-concave for any $\gamma' \in (0,1]$, $p \geq 1$ and $\sigma \geq 0$ and f is not σ-OSS for any $\sigma \geq 0$.*

Proof. Define $F(t) = f(t\mathbf{1})$ for $t \in [0,1]$. Since f is monotone and $\mathbf{1} \geq \mathbf{x}$ for every $\mathbf{x} \in [0,1]^d$, we have $f(\mathbf{1}) = \max_{\mathbf{x}\in[0,1]^d} f(\mathbf{x})$, and $f(\mathbf{1}) > 0$ because f is nonnegative and not identically zero. Since f is flat at the origin, $F(t) \to 0$ as $t \downarrow 0$. Hence there is $t_0 \in (0,1)$ such that

$$f(\mathbf{1}) - F(t) \geq \tfrac{1}{2} f(\mathbf{1}) \qquad \text{for all } t \in (0, t_0]. \tag{46}$$

Fix $t \in (0, t_0]$. The pair $\mathbf{x} = t\mathbf{1}$, $\mathbf{y} = \mathbf{1}$ satisfies $\mathbf{y} \geq \mathbf{x} \neq \mathbf{0}$, so the upper inequality in (2) gives

$$f(\mathbf{1}) - F(t) \leq \frac{\theta(\mathbf{1})}{\gamma\theta(t\mathbf{1})} \langle \nabla f(t\mathbf{1}), (1-t)\mathbf{1} \rangle = \frac{\theta(\mathbf{1})(1-t)}{\gamma\theta(t\mathbf{1})} F'(t), \tag{47}$$

where we used $F'(t) = \langle \nabla f(t\mathbf{1}), \mathbf{1} \rangle$. Combining (46) and (47), and using $1 - t \leq 1$, we obtain

$$F'(t) \geq c\,\theta(t\mathbf{1}) \quad \text{for all } t \in (0, t_0], \qquad \text{where } c := \frac{\gamma f(\mathbf{1})}{2\theta(\mathbf{1})} > 0. \tag{48}$$

For $t \in (0, t_0]$, since $F(0) = f(\mathbf{0}) \geq 0$ and $F' \geq c\,\theta(\cdot\,\mathbf{1}) \geq 0$ on $(0, t_0]$ by (48), the monotonicity of θ gives

$$F(t) \geq \int_0^t F'(u)du \geq \int_{t/2}^t F'(u)du \geq c \int_{t/2}^t \theta(u\mathbf{1})du \geq \frac{ct}{2}\,\theta\Big(\frac{t}{2}\mathbf{1}\Big). \tag{49}$$

Hence $\theta(\frac{t}{2}\mathbf{1}) \leq \frac{2F(t)}{ct}$. Now fix $n > 0$. Since f is flat at the origin and $\|t\mathbf{1}\|_1 = dt$, we have $F(t) = o(t^{n+1})$ as $t \downarrow 0$. Therefore $\theta(\frac{t}{2}\mathbf{1}) = o(t^n)$, that is,

$$\theta(\tau\mathbf{1}) = o(\tau^n) \qquad \text{as } \tau \downarrow 0. \tag{50}$$

Finally, let $\mathbf{x} \to \mathbf{0}$ and write $r = \|\mathbf{x}\|_1$. For $r \leq 1$, every coordinate of $\mathbf{x}$ is at most r, so $\mathbf{x} \leq r\mathbf{1} \in [0,1]^d$, and by monotonicity of θ and (50), $\theta(\mathbf{x}) \leq \theta(r\mathbf{1}) = o(r^n) = o(\|\mathbf{x}\|_1^n)$. Since $n > 0$ was arbitrary, θ is flat at the origin.

For the last claim, suppose that $f \in \mathcal{F}_{\gamma',p,\sigma}$ for some $\gamma' \in (0,1]$, $p \geq 1$ and $\sigma \geq 0$. The first part of the proof, applied with $\theta'(\mathbf{x}) = \|\mathbf{x}\|_p^\sigma$, shows that θ' is flat at the origin. However, $\|\mathbf{x}\|_p \geq \|\mathbf{x}\|_\infty \geq \|\mathbf{x}\|_1/d$, so $\theta'(\mathbf{x})/\|\mathbf{x}\|_1^{\sigma+1} \geq d^{-\sigma}/\|\mathbf{x}\|_1 \to \infty$ as $\mathbf{x} \to \mathbf{0}$, which contradicts flatness with $n = \sigma + 1$. Hence $f \notin \mathcal{F}_{\gamma',p,\sigma}$. If f were σ-OSS, then Theorem 3 would give $f \in \mathcal{F}_{1,1,2\sigma}$, which we have just excluded. □

Lemma 4. *Let*

$$\theta(\mathbf{x}) = \exp\!\left(-\frac{1}{\|\mathbf{x}\|_1}\right), \qquad \theta(\mathbf{0}) = 0.$$

Also let $q : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a continuous non-increasing function with $q(0) > 0$, and define

$$f(\mathbf{x}) = \int_0^{\|\mathbf{x}\|_1} \exp\!\left(-\frac{1}{t}\right) q(t)\,dt.$$

Then the following are true.

(i) f is 1-weakly θ-up-concave.

(ii) f is flat at the origin.

(iii) For any $p \geq 1$, $\sigma \geq 0$ and $\gamma \in (0,1]$, f is not γ-weakly (p,σ)-up-concave.

(iv) For any $\sigma \geq 0$, f is not σ-OSS.

(v) f may exhibit accumulating-then-diminishing returns.

Choice of $q(s)$	Transition equation	Transition point $s_\star$
e^{-Ms}	$Ms_\star^2 = 1$	$s_\star = M^{-1/2}$
e^{-Ms^α}, $\alpha > 0$	$M\alpha s_\star^{\alpha+1} = 1$	$s_\star = (M\alpha)^{-1/(\alpha+1)}$
$e^{-M(s+s^2)}$	$Ms_\star^2(1+2s_\star) = 1$	unique positive root
$(1+s)^{-\beta}$, $\beta > 0$	$\dfrac{\beta s_\star^2}{1+s_\star} = 1$	$s_\star = \dfrac{1+\sqrt{1+4\beta}}{2\beta}$
$\log(e+s)^{-\beta}$, $\beta > 0$	$\dfrac{\beta s_\star^2}{(e+s_\star)\log(e+s_\star)} = 1$	unique positive root
e^{-Me^s}	$Ms_\star^2 e^{s_\star} = 1$	unique positive root

Table 2: Examples of decay profiles q (with $M > 0$) for which the marginal reward $e^{-1/s}q(s)$ first increases and then decreases. The transition point $s_\star$ solves $\frac{1}{s_\star^2} = -\frac{q'(s_\star)}{q(s_\star)}$. The profile q defines f; an algorithm observes only values or gradients of f.

Proof. Throughout, write $s = \|\mathbf{x}\|_1 = \mathbf{1}^\top \mathbf{x}$ for $\mathbf{x} \in [0,1]^d$, and $F(s) = \int_0^s e^{-1/t} q(t)dt$, so that $f(\mathbf{x}) = F(s)$. The integrand $t \mapsto e^{-1/t}q(t)$, extended by 0 at $t = 0$, is continuous on $[0,\infty)$, so F is C^1 with $F'(s) = e^{-1/s}q(s)$.

(i): By the chain rule,

$$\nabla f(\mathbf{x}) = F'(s)\mathbf{1} = e^{-1/s}q(s)\mathbf{1}, \tag{51}$$

so f is C^1, $f \geq 0$, and f is monotone because $\nabla f \geq \mathbf{0}$. The function θ is continuous and monotone, with $\theta(\mathbf{x}) > 0$ for $\mathbf{x} \neq \mathbf{0}$. For $\mathbf{x} \neq \mathbf{0}$, (51) gives $\nabla f(\mathbf{x})/\theta(\mathbf{x}) = q(s)\mathbf{1}$. If $\mathbf{y} \geq \mathbf{x} \neq \mathbf{0}$, then $\|\mathbf{y}\|_1 \geq \|\mathbf{x}\|_1$, and since q is nonincreasing, $\nabla f(\mathbf{y})/\theta(\mathbf{y}) \leq \nabla f(\mathbf{x})/\theta(\mathbf{x})$ coordinate-wise. Theorem 2 with $\gamma = 1$ gives (i).

(ii): Fix $n > 0$ and choose $N > n$. Since $e^{-1/t}/t^N \to 0$ as $t \downarrow 0$, there is $\delta > 0$ with $e^{-1/t} \leq t^N$ for all $t \in (0,\delta]$. Since q is nonincreasing, $q(t) \leq q(0)$. Hence, for $s \leq \delta$,

$$f(\mathbf{x}) = \int_0^s e^{-1/t}q(t)dt \leq q(0)\int_0^s t^N dt = \frac{q(0)}{N+1}s^{N+1}, \tag{52}$$

so $f(\mathbf{x})/\|\mathbf{x}\|_1^n \leq \frac{q(0)}{N+1}s^{N+1-n} \to 0$ as $\mathbf{x} \to \mathbf{0}$, because $N + 1 - n > 0$.

(iii) and (iv): Since q is continuous and $q(0) > 0$, there is $\eta > 0$ with $q > 0$ on $[0,\eta)$, so $f(\mathbf{x}) > 0$ whenever $0 < \|\mathbf{x}\|_1 < \eta$, and f is not identically zero. By (i) and (ii), Lemma 3 applies and gives (iii) and (iv).

(v): Let $q(s) = e^{-\psi(s)}$, where $\psi : [0,\infty) \to \mathbb{R}$ is continuous and nondecreasing, differentiable on $(0,\infty)$, and such that $s \mapsto s^2\psi'(s)$ is strictly increasing on $(0,\infty)$ with range containing 1; then q is continuous, nonincreasing and $q(0) = e^{-\psi(0)} > 0$. The calculation below only differentiates at $s > 0$. We have $F'(s) = e^{-1/s-\psi(s)}$, and for $s > 0$,

$$F''(s) = \left(\frac{1}{s^2} - \psi'(s)\right)e^{-1/s-\psi(s)} = \frac{1 - s^2\psi'(s)}{s^2}\,e^{-1/s-\psi(s)}. \tag{53}$$

Let $s_\star$ be the unique solution of $s_\star^2\psi'(s_\star) = 1$. Since $s \mapsto s^2\psi'(s)$ is strictly increasing, (53) shows that $F'' > 0$ on $(0,s_\star)$ and $F'' < 0$ on $(s_\star,\infty)$. Thus the marginal reward F' increases on $(0,s_\star)$ and decreases on $(s_\star,\infty)$. On $[0,1]^d$, where $s \in [0,d]$, both phases occur when $s_\star < d$. Table 2 lists examples of q of this form. □

Lemma 5. *Let $\mathbf{a} \in \mathbb{R}^d_{>0}$, let $M > 0$, and let f and θ be as in Example 1. Then the following are true.*

(i) f is 1-weakly θ-up-concave.

(ii) f is flat at the origin.

(iii) For any $p \geq 1$, $\sigma \geq 0$ and $\gamma \in (0,1]$, f is not γ-weakly (p,σ)-up-concave, and for any $\sigma \geq 0$, f is not σ-OSS.

(iv) As a function of $s = \mathbf{a}^\top \mathbf{x}$, the marginal gain of f is increasing for $s < 1/\sqrt{M}$ and decreasing for $s > 1/\sqrt{M}$; both phases occur on $[0,1]^d$ when $M > \|\mathbf{a}\|_1^{-2}$.

Proof. Throughout, write $s = \mathbf{a}^\top \mathbf{x}$ and $F(s) = \int_0^s \exp(-1/t - Mt)dt$, so that $f(\mathbf{x}) = F(s)$. The integrand, extended by 0 at $t = 0$, is continuous on $[0, \infty)$, so F is C^1 with $F'(s) = \exp(-1/s - Ms)$.

(i): By the chain rule,

$$\nabla f(\mathbf{x}) = F'(s)\mathbf{a} = \exp\Big(-\frac{1}{s} - Ms\Big)\mathbf{a}, \tag{54}$$

so f is C^1 on $[0,1]^d$, $f \geq 0$, and f is monotone since $\nabla f \geq \mathbf{0}$. The function θ is continuous and monotone, and $\theta(\mathbf{x}) > 0$ for $\mathbf{x} \neq \mathbf{0}$, because $\mathbf{a} \in \mathbb{R}^d_{>0}$ gives $s > 0$. For $\mathbf{x} \neq \mathbf{0}$, (54) gives $\nabla f(\mathbf{x})/\theta(\mathbf{x}) = e^{-Ms}\mathbf{a}$. If $\mathbf{y} \geq \mathbf{x} \neq \mathbf{0}$, then $\mathbf{a}^\top \mathbf{y} \geq \mathbf{a}^\top \mathbf{x}$, so $e^{-M\mathbf{a}^\top \mathbf{y}} \leq e^{-M\mathbf{a}^\top \mathbf{x}}$ and $\nabla f(\mathbf{y})/\theta(\mathbf{y}) \leq \nabla f(\mathbf{x})/\theta(\mathbf{x})$ coordinate-wise. Theorem 2 with $\gamma = 1$ gives (i).

(ii): Let $A = \|\mathbf{a}\|_\infty$, so that $0 \leq s \leq A\|\mathbf{x}\|_1$. For $0 < t \leq s$, we have $\exp(-1/t - Mt) \leq e^{-1/t} \leq e^{-1/s}$, hence

$$f(\mathbf{x}) \leq se^{-1/s} \leq A\|\mathbf{x}\|_1 \exp\Big(-\frac{1}{A\|\mathbf{x}\|_1}\Big), \tag{55}$$

where the second inequality uses that $u \mapsto ue^{-1/u}$ is increasing on $(0, \infty)$. For every $n > 0$, the right-hand side of (55) divided by $\|\mathbf{x}\|_1^n$ tends to 0 as $\mathbf{x} \to \mathbf{0}$, because $e^{-1/u}$ tends to 0 faster than any power of u as $u \downarrow 0$.

(iii): Since $F'(s) > 0$ for $s > 0$, the function f is not identically zero. By (i) and (ii), Lemma 3 gives (iii).

(iv): For $s > 0$, $F''(s) = (1/s^2 - M)F'(s)$, and $F'(s) > 0$. Hence F' is increasing for $s < 1/\sqrt{M}$ and decreasing for $s > 1/\sqrt{M}$. On $[0,1]^d$, s ranges over $[0, \|\mathbf{a}\|_1]$, so both phases occur when $1/\sqrt{M} < \|\mathbf{a}\|_1$, that is, when $M > \|\mathbf{a}\|_1^{-2}$. □

Proposition 3. *Let $c > 0$, let $\theta(\mathbf{x}) = \exp(-c/\|\mathbf{x}\|_1)$ for $\mathbf{x} \neq \mathbf{0}$ and $\theta(\mathbf{0}) = 0$, and let $g : [0,1]^d \to \mathbb{R}_{>0}$ be a C^2 function that is coordinate-wise nonincreasing and DR-submodular, i.e., $\partial^2_{ij} g \leq 0$ for all i, j. Define $F = \theta g$, with $F(\mathbf{0}) = 0$. Then F is C^1, and F is monotone if and only if*

$$-\partial_i \log g(\mathbf{x}) \leq \frac{c}{\|\mathbf{x}\|_1^2} \quad \textit{for all } i \textit{ and all } \mathbf{x} \neq \mathbf{0}. \tag{56}$$

If (56) *holds, then F is* 1*-weakly θ-up-concave.*

Proof. For $\mathbf{x} \neq \mathbf{0}$, write $s = \|\mathbf{x}\|_1 = \mathbf{1}^\top \mathbf{x} > 0$. Since $\partial_i s = 1$ and $\theta(\mathbf{x}) = e^{-c/s}$, the chain rule gives $\partial_i \theta(\mathbf{x}) = \frac{c}{s^2}\theta(\mathbf{x})$, and the product rule gives

$$\partial_i F(\mathbf{x}) = \theta(\mathbf{x})\partial_i g(\mathbf{x}) + g(\mathbf{x})\partial_i \theta(\mathbf{x}) = \theta(\mathbf{x}) q_i(\mathbf{x}), \qquad \text{where } q_i(\mathbf{x}) := \partial_i g(\mathbf{x}) + \frac{c}{s^2} g(\mathbf{x}). \tag{57}$$

Regularity. Since g is C^2 on the compact set $[0,1]^d$, both g and ∇g are bounded there. Since $e^{-c/s}/s^2 \to 0$ as $s \downarrow 0$, (57) gives $\nabla F(\mathbf{x}) \to \mathbf{0}$ as $\mathbf{x} \to \mathbf{0}$. Moreover, $0 \leq F(\mathbf{x}) \leq \theta(\mathbf{x}) \max g$ and $\theta(\mathbf{x})/\|\mathbf{x}\|_1 \to 0$, so F is differentiable at $\mathbf{0}$ with $\nabla F(\mathbf{0}) = \mathbf{0}$. Hence F is C^1 on $[0,1]^d$, and $F \geq 0$.

Monotonicity. A C^1 function on $[0,1]^d$ is monotone if and only if all its partial derivatives are nonnegative. Since $\nabla F(\mathbf{0}) = \mathbf{0}$ and $\theta(\mathbf{x}) > 0$ for $\mathbf{x} \neq \mathbf{0}$, (57) shows that F is monotone if and only if $q_i(\mathbf{x}) \geq 0$ for all i and all $\mathbf{x} \neq \mathbf{0}$. Dividing by $g(\mathbf{x}) > 0$, the inequality $q_i(\mathbf{x}) \geq 0$ is equivalent to $-\partial_i g(\mathbf{x})/g(\mathbf{x}) \leq c/s^2$, which is (56).

Up-concavity. Assume (56), so that F is monotone. For $\mathbf{x} \neq \mathbf{0}$, differentiating q_i and using $\partial_j s = 1$ gives

$$\partial_j q_i(\mathbf{x}) = \partial^2_{ij} g(\mathbf{x}) + \frac{c}{s^2}\partial_j g(\mathbf{x}) - \frac{2c}{s^3} g(\mathbf{x}). \tag{58}$$

Each term in (58) is nonpositive: $\partial^2_{ij} g \le 0$ by DR-submodularity, $\partial_j g \le 0$ because g is nonincreasing, and $g > 0$. So $\nabla q_i(\mathbf{x}) \le \mathbf{0}$ for all i and all $\mathbf{x} \ne \mathbf{0}$. Now let $\mathbf{y} \ge \mathbf{x} \ne \mathbf{0}$. Every point $\mathbf{x} + t(\mathbf{y} - \mathbf{x})$ with $t \in [0, 1]$ is at least $\mathbf{x}$ and hence nonzero, so q_i is C^1 along this segment, and

$$q_i(\mathbf{y}) - q_i(\mathbf{x}) = \int_0^1 \langle \nabla q_i(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt \le 0,$$

because $\nabla q_i \le \mathbf{0}$ and $\mathbf{y} - \mathbf{x} \ge \mathbf{0}$. By (57), $q_i = \partial_i F / \theta$, so $\nabla F(\mathbf{y})/\theta(\mathbf{y}) \le \nabla F(\mathbf{x})/\theta(\mathbf{x})$ coordinate-wise. Since θ is continuous and monotone with $\theta(\mathbf{x}) > 0$ for $\mathbf{x} \ne \mathbf{0}$, Theorem 2 with $\gamma = 1$ shows that F is 1-weakly θ-up-concave. □

Remark 4. *Condition* (56) *holds whenever g is large compared to its partial derivatives. Indeed, if $|\partial_i g| \le L$ on $[0, 1]^d$ for some $L > 0$ and all i, and $g \ge d^2 L/c$, then, since $\|\mathbf{x}\|_1 \le d$,*

$$-\partial_i \log g(\mathbf{x}) = \frac{-\partial_i g(\mathbf{x})}{g(\mathbf{x})} \le \frac{L}{d^2 L/c} = \frac{c}{d^2} \le \frac{c}{\|\mathbf{x}\|_1^2}.$$

Thus, if $h : [0, 1]^d \to \mathbb{R}_{\ge 0}$ is C^2, coordinate-wise nonincreasing and DR-submodular with $|\partial_i h| \le L$ for some $L > 0$ and all i, then $g = h + d^2 L/c$ satisfies the assumptions of Proposition 3, and $F = \theta g$ is 1-weakly θ-up-concave. For example, $h(\mathbf{x}) = d - \frac{1}{2}\|\mathbf{x}\|_2^2$ satisfies these conditions with $L = 1$, so

$$F(\mathbf{x}) = \exp\Big(-\frac{c}{\|\mathbf{x}\|_1}\Big)\Big(d + \frac{d^2}{c} - \frac{1}{2}\|\mathbf{x}\|_2^2\Big)$$

is 1-weakly θ-up-concave. Here $\partial_i F(\mathbf{x})/\theta(\mathbf{x}) = \frac{c}{\|\mathbf{x}\|_1^2} g(\mathbf{x}) - x_i$, so for $d \ge 2$ the direction of $\nabla F(\mathbf{x})$ changes with $\mathbf{x}$. In particular, F is not of the form $\phi(\mathbf{a}^\top \mathbf{x})$ for any $\mathbf{a} \in \mathbb{R}^d$ and differentiable $\phi : \mathbb{R} \to \mathbb{R}$, and its level sets are not parallel hyperplanes. Since F is flat at the origin and not identically zero, Lemma 3 shows that it is neither σ-OSS nor γ-weakly (p, σ)-up-concave for any choice of parameters. Functions in classes with different scaling functions can also be combined. If $f_j \in \mathcal{F}_{\gamma_j, \theta_j}$ for $j = 1, \dots, m$, let $\gamma = \min_j \gamma_j$ and $\Theta = \prod_j \theta_j$. For $\mathbf{y} \ge \mathbf{x} \ne \mathbf{0}$, each ratio $\theta_j(\mathbf{y})/\theta_j(\mathbf{x})$ is at least 1, so $\Theta(\mathbf{y})/\Theta(\mathbf{x}) \ge \theta_j(\mathbf{y})/\theta_j(\mathbf{x})$ for every j. Since the gradient inner products in (2) *are nonnegative, the inequalities for (γ_j, θ_j) imply those for (γ, Θ), so each f_j belongs to $\mathcal{F}_{\gamma, \Theta}$. By Lemma 1, every nonnegative combination $\sum_j a_j f_j$ also belongs to $\mathcal{F}_{\gamma, \Theta}$.*

E Proof of Corollary 2

Proof. Under the assumptions of Theorem 4, $\mathcal{K} \subseteq [0, 1]^d$ is closed and convex and contains a point other than $\mathbf{0}$, and θ is continuous and monotone with $\theta(\mathbf{x}) > 0$ for $\mathbf{x} \ne \mathbf{0}$.

A point of $\mathcal{K}$ with maximal support. Let I be the set of coordinates i such that $x_i > 0$ for some $\mathbf{x} \in \mathcal{K}$. Since $\mathcal{K} \subseteq [0, 1]^d$ contains a point other than $\mathbf{0}$, the set I is nonempty. For each $i \in I$, choose $\mathbf{x}^{(i)} \in \mathcal{K}$ with $x_i^{(i)} > 0$, and let

$$\mathbf{q} = \frac{1}{|I|} \sum_{i \in I} \mathbf{x}^{(i)}. \tag{59}$$

Since $\mathcal{K}$ is convex, $\mathbf{q} \in \mathcal{K}$. For $i \in I$, every term in (59) is nonnegative and the term $\mathbf{x}^{(i)}$ has a positive i-th coordinate, so $q_i > 0$. For $i \notin I$, every point of $\mathcal{K}$ has i-th coordinate 0, so $q_i = 0$. Let $\delta = \min_{i \in I} q_i > 0$.

Every maximal point has ℓ_1 norm at least δ. Let $\mathbf{v} \in \mathcal{K}^{\mathrm{m}}$. Suppose, for contradiction, that $v_i < q_i$ for all $i \in I$. For $i \notin I$, we have $v_i = 0 = q_i$ because $\mathbf{v} \in \mathcal{K}$. Hence $\mathbf{q} \ge \mathbf{v}$ and $\mathbf{q} \ne \mathbf{v}$ (since $I \ne \emptyset$), and $\mathbf{q} \in \mathcal{K}$, which contradicts the maximality of $\mathbf{v}$. Therefore $v_i \ge q_i \ge \delta$ for some $i \in I$, and since $\mathbf{v} \ge \mathbf{0}$,

$$\mathbf{1}^\top \mathbf{v} \ge v_i \ge \delta \qquad \text{for all } \mathbf{v} \in \mathcal{K}^{\mathrm{m}}. \tag{60}$$

$\mathbf{0} \notin \mathcal{K}^*$. By (60), $\mathcal{K}^{\mathrm{m}}$ is contained in the halfspace $H = \{\mathbf{x} : \mathbf{1}^\top \mathbf{x} \ge \delta\}$. Since H is closed and convex, it also contains $\mathcal{K}^* = \overline{\mathrm{conv}}(\mathcal{K}^{\mathrm{m}})$. Since $\mathbf{0} \notin H$, we get $\mathbf{0} \notin \mathcal{K}^*$.

The infimum is positive. Define $\Phi(\mathbf{x}) = \int_0^1 \theta(s\mathbf{x}) ds$ for $\mathbf{x} \in [0, 1]^d$. If $\mathbf{x} \ne \mathbf{0}$, then $s\mathbf{x} \ne \mathbf{0}$ for $s \in (0, 1]$, so the continuous function $s \mapsto \theta(s\mathbf{x})$ is positive on $(0, 1]$ and $\Phi(\mathbf{x}) > 0$. Since θ is

continuous on the compact set $[0,1]^d$, it is uniformly continuous there, and hence Φ is continuous. The set $\mathcal{K}^*$ is closed (by definition) and contained in $[0,1]^d$ (since $\mathcal{K}^{\mathrm{m}} \subseteq \mathcal{K}$ and $\mathcal{K}$ is closed and convex), so it is compact. Therefore Φ attains its minimum over $\mathcal{K}^*$ at some point $\mathbf{x}^\star \in \mathcal{K}^*$. Since $\mathbf{x}^\star \neq \mathbf{0}$, we get $\inf_{\mathbf{x}\in\mathcal{K}^*} \Phi(\mathbf{x}) = \Phi(\mathbf{x}^\star) > 0$, and therefore $\alpha = 1 - \exp(-\gamma\Phi(\mathbf{x}^\star)/R_\theta) > 0$.

The inequality. Since $\mathcal{K}$ is closed and convex and $\mathcal{K}^{\mathrm{m}} \subseteq \mathcal{K}$, we have $\mathcal{K}^* \subseteq \mathcal{K}$. Hence Theorem 4 applies to $\mathbf{x}, \mathbf{y} \in \mathcal{K}^*$ and gives

$$\langle \mathfrak{g}(f,\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \geq \alpha_{\mathbf{x}} f(\mathbf{y}) - f(\mathbf{x}).$$

For $\mathbf{x} \in \mathcal{K}^*$, we have $\Phi(\mathbf{x}) \geq \inf_{\mathbf{z}\in\mathcal{K}^*} \Phi(\mathbf{z})$, and $t \mapsto 1 - \exp(-\gamma t/R_\theta)$ is nondecreasing, so $\alpha_{\mathbf{x}} \geq \alpha$. Since $f(\mathbf{y}) \geq 0$, it follows that $\alpha_{\mathbf{x}} f(\mathbf{y}) \geq \alpha f(\mathbf{y})$, which gives the claimed inequality. □

F Proof of Theorem 5

Proof. Recall that $\mathcal{G}(\mathcal{Q}_f, \mathbf{x}) = e^{l(Z,\mathbf{x})}\mathcal{Q}_f(Z\mathbf{x})$, where Z is drawn uniformly from $[0,1]$ using fresh randomization. Throughout, we use only the oracle assumption of Section 2: the oracle output at a query point $\mathbf{y}$ has conditional mean $\nabla f(\mathbf{y})$, given the history of the game and the query point.

$\mathcal{G}$ is an unbiased query algorithm for $\mathfrak{g}$. Fix $\mathbf{x} \in \mathcal{K}$. Since $\mathcal{K}$ is convex and contains $\mathbf{0}$ and $\mathbf{x}$, the query point $Z\mathbf{x}$ lies in $\mathcal{K}$. The query point is determined by Z, so conditional unbiasedness gives $\mathbb{E}[\mathcal{Q}_f(Z\mathbf{x}) \mid Z] = \nabla f(Z\mathbf{x})$. By the tower rule of conditional expectation,

$$\begin{aligned}\mathbb{E}\left[\mathcal{G}(\mathcal{Q}_f,\mathbf{x})\right] &= \mathbb{E}\Big[e^{l(Z,\mathbf{x})}\,\mathbb{E}[\mathcal{Q}_f(Z\mathbf{x}) \mid Z]\Big] = \mathbb{E}\left[e^{l(Z,\mathbf{x})}\nabla f(Z\mathbf{x})\right] \\ &= \int_0^1 e^{l(r,\mathbf{x})}\nabla f(r\mathbf{x})dr = \mathfrak{g}(f,\mathbf{x}),\end{aligned} \tag{61}$$

where the third equality holds because Z has density 1 on $[0,1]$.

Boundedness. By (8), $l(Z,\mathbf{x}) \leq 0$, so $0 < e^{l(Z,\mathbf{x})} \leq 1$. If $\|\mathcal{Q}_f(\mathbf{z})\| \leq B_1$ for all $\mathbf{z} \in \mathcal{K}$, then, since $Z\mathbf{x} \in \mathcal{K}$,

$$\|\mathcal{G}(\mathcal{Q}_f,\mathbf{x})\| = e^{l(Z,\mathbf{x})}\|\mathcal{Q}_f(Z\mathbf{x})\| \leq B_1. \tag{62}$$

Setting up the regret bound. Fix an adversary in $\mathrm{Adv}_1^{\mathrm{o}}(\mathcal{F}, B_1)$, an interval $[a,b]$ and a compact comparator set $\mathcal{U} \subseteq (\mathcal{K}^*)^T$. Since the adversary is oblivious, the functions $f_1,\ldots,f_T$ are fixed before the game. Each f_t is continuous and $\mathcal{U}$ is compact, so we may fix, before the game, a deterministic maximizer

$$\mathbf{u}^\star = (\mathbf{u}_1^\star,\ldots,\mathbf{u}_T^\star) \in \operatorname*{argmax}_{\mathbf{u}\in\mathcal{U}} \sum_{t=a}^{b} f_t(\mathbf{u}_t). \tag{63}$$

Let $\mathbf{x}_t \in \mathcal{K}^*$ be the action played at round t, let $\mathbf{o}_t$ be the oracle output at the query point $Z_t\mathbf{x}_t$, and let $\mathbf{g}_t = e^{l(Z_t,\mathbf{x}_t)}\mathbf{o}_t$ be the vector passed to $\mathcal{A}(\mathcal{K}^*)$. Let $\mathcal{H}_t$ denote the history before the round-t sampling step. Since $\mathcal{A}(\mathcal{K}^*)$ is deterministic, $\mathbf{x}_t$ is $\mathcal{H}_t$-measurable, and Z_t is uniform on $[0,1]$ and independent of $\mathcal{H}_t$, because it is drawn with fresh randomization. Conditional unbiasedness with respect to the complete pre-query history gives

$$\mathbb{E}[\mathbf{o}_t \mid \mathcal{H}_t, Z_t] = \nabla f_t(Z_t\mathbf{x}_t). \tag{64}$$

Therefore, by the tower rule,

$$\begin{aligned}\mathbb{E}[\mathbf{g}_t \mid \mathcal{H}_t] &= \mathbb{E}\Big[e^{l(Z_t,\mathbf{x}_t)}\,\mathbb{E}[\mathbf{o}_t \mid \mathcal{H}_t, Z_t]\,\Big|\,\mathcal{H}_t\Big] = \mathbb{E}\Big[e^{l(Z_t,\mathbf{x}_t)}\nabla f_t(Z_t\mathbf{x}_t)\,\Big|\,\mathcal{H}_t\Big] \\ &= \int_0^1 e^{l(r,\mathbf{x}_t)}\nabla f_t(r\mathbf{x}_t)dr = \mathfrak{g}(f_t,\mathbf{x}_t),\end{aligned} \tag{65}$$

where the third equality uses that $\mathbf{x}_t$ is $\mathcal{H}_t$-measurable and that, given $\mathcal{H}_t$, Z_t is still uniform on $[0,1]$.

From the nonlinear to the linear regret. Since $\mathbf{x}_t, \mathbf{u}_t^\star \in \mathcal{K}^*$, Corollary 2 gives, for every t,

$$\alpha f_t(\mathbf{u}_t^\star) - f_t(\mathbf{x}_t) \leq \langle \mathfrak{g}(f_t,\mathbf{x}_t), \mathbf{u}_t^\star - \mathbf{x}_t\rangle. \tag{66}$$

The vector $\mathbf{u}_t^\star - \mathbf{x}_t$ is $\mathcal{H}_t$-measurable, so by (65) and the tower rule,

$$\mathbb{E}\langle \mathfrak{g}(f_t, \mathbf{x}_t), \mathbf{u}_t^\star - \mathbf{x}_t\rangle = \mathbb{E}\Big[\langle \mathbb{E}[\mathbf{g}_t \mid \mathcal{H}_t], \mathbf{u}_t^\star - \mathbf{x}_t\rangle\Big] = \mathbb{E}\langle \mathbf{g}_t, \mathbf{u}_t^\star - \mathbf{x}_t\rangle. \tag{67}$$

Taking expectations in (66), using (67), summing over $t \in [a, b]$, and using (63), we obtain

$$\mathbb{E}\left[\alpha \max_{\mathbf{u}\in\mathcal{U}} \sum_{t=a}^{b} f_t(\mathbf{u}_t) - \sum_{t=a}^{b} f_t(\mathbf{x}_t)\right] \le \mathbb{E}\sum_{t=a}^{b}\langle \mathbf{g}_t, \mathbf{u}_t^\star - \mathbf{x}_t\rangle \le \mathbb{E}\left[\max_{\mathbf{v}\in\mathcal{U}} \sum_{t=a}^{b}\langle \mathbf{g}_t, \mathbf{v}_t\rangle - \sum_{t=a}^{b}\langle \mathbf{g}_t, \mathbf{x}_t\rangle\right]. \tag{68}$$

Bounding the linear regret. Since $\mathcal{A}(\mathcal{K}^*)$ is deterministic and receives only $\mathbf{g}_1, \dots, \mathbf{g}_{t-1}$ before choosing $\mathbf{x}_t$, for every realization of the sequence $(\mathbf{g}_t)_{t=1}^T$ the actions $\mathbf{x}_t$ are exactly those that $\mathcal{A}(\mathcal{K}^*)$ plays against the fixed sequence of linear rewards $\mathbf{v} \mapsto \langle \mathbf{g}_t, \mathbf{v}\rangle$. By (62), these rewards belong to $\mathcal{L}(\mathcal{K}^*)[B_1]$. Hence the worst-case linear regret bound of $\mathcal{A}(\mathcal{K}^*)$ applies pathwise, and the quantity inside the expectation on the right-hand side of (68) is at most $\mathcal{R}^{\mathcal{A}(\mathcal{K}^*)}_{1,\mathrm{Adv}_1(\mathcal{L}(\mathcal{K}^*)[B_1])}(\mathcal{U})[a, b]$ for every realization. Taking expectations, and then the supremum over oblivious adversaries on the left-hand side, proves the claimed inequality.

Comparators in $\mathcal{K}$. For static and adaptive regret, the comparator on $[a, b]$ is a single point $\mathbf{u}$ used in all rounds $t \in [a, b]$. Let $\mathbf{u} \in \mathcal{K}$. The set $\{\mathbf{w} \in \mathcal{K} : \mathbf{w} \ge \mathbf{u}\}$ is compact and nonempty, so $\|\cdot\|_1$ attains its maximum on it at some point $\mathbf{v}$. This $\mathbf{v}$ lies in $\mathcal{K}^{\mathrm{m}}$: if $\mathbf{z} \in \mathcal{K}$ satisfied $\mathbf{z} \ge \mathbf{v}$ and $\mathbf{z} \ne \mathbf{v}$, then $\mathbf{z} \ge \mathbf{u}$ and $\|\mathbf{z}\|_1 > \|\mathbf{v}\|_1$, a contradiction. Thus $\mathbf{v} \in \mathcal{K}^{\mathrm{m}} \subseteq \mathcal{K}^*$ and $\mathbf{v} \ge \mathbf{u}$, and since every f_t is monotone, $\sum_{t=a}^b f_t(\mathbf{v}) \ge \sum_{t=a}^b f_t(\mathbf{u})$. Hence $\max_{\mathbf{u}\in\mathcal{K}} \sum_{t=a}^b f_t(\mathbf{u}) = \max_{\mathbf{u}\in\mathcal{K}^*} \sum_{t=a}^b f_t(\mathbf{u})$, and the left-hand side may be maximized over $\mathcal{K}$. □

G Proof of Proposition 1

Proof. Throughout, write $R = R_\theta$ and $\Phi(\mathbf{x}) = \int_0^1 \theta(s\mathbf{x})ds$, so that

$$\alpha_{\mathbf{x}} = 1 - \exp\left(-\frac{\gamma}{R}\Phi(\mathbf{x})\right), \qquad \alpha(\mathcal{K}') = 1 - \exp\left(-\frac{\gamma}{R}\inf_{\mathbf{x}\in\mathcal{K}'}\Phi(\mathbf{x})\right). \tag{69}$$

The map $t \mapsto 1 - \exp(-\gamma t/R)$ is nondecreasing on $\mathbb{R}$.

The inequality of Corollary 2 on $\mathcal{K}'$. Fix $\mathbf{x}, \mathbf{y} \in \mathcal{K}'$. Since $\mathcal{K}' \subseteq \mathcal{K}$, Theorem 4 gives

$$\langle \mathfrak{g}(f, \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \ge \alpha_{\mathbf{x}} f(\mathbf{y}) - f(\mathbf{x}). \tag{70}$$

Since $\mathbf{x} \in \mathcal{K}'$, we have $\Phi(\mathbf{x}) \ge \inf_{\mathbf{z}\in\mathcal{K}'}\Phi(\mathbf{z})$, and by (69) and the monotonicity of $t \mapsto 1 - \exp(-\gamma t/R)$, $\alpha_{\mathbf{x}} \ge \alpha(\mathcal{K}')$. Since $f(\mathbf{y}) \ge 0$, it follows that $\alpha_{\mathbf{x}} f(\mathbf{y}) \ge \alpha(\mathcal{K}')f(\mathbf{y})$, and (70) gives

$$\langle \mathfrak{g}(f, \mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \ge \alpha(\mathcal{K}')f(\mathbf{y}) - f(\mathbf{x}) \qquad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{K}'. \tag{71}$$

The regret bound on $\mathcal{K}'$. The proof of Theorem 5 in Appendix F uses three facts about $\mathcal{K}^*$: the inequality of Corollary 2, that $\mathcal{K}^* \subseteq \mathcal{K}$ (so that the query points $Z_t\mathbf{x}_t$ lie in $\mathcal{K}$), and that every point of $\mathcal{K}$ is dominated by a point of $\mathcal{K}^*$. The set $\mathcal{K}'$ has the same three properties: the first is (71), the second holds by assumption, and the third holds because $\mathcal{K}^{\mathrm{m}} \subseteq \mathcal{K}^* \subseteq \mathcal{K}'$. Hence that proof applies verbatim with $\mathcal{K}'$ and $\alpha(\mathcal{K}')$ in place of $\mathcal{K}^*$ and α. If $\alpha(\mathcal{K}') = 0$, the resulting bound compares $\sum_t f_t(\mathbf{x}_t)$ with 0, so it gives no approximation guarantee.

Comparison with $\mathcal{K}^*$. Since $\mathcal{K}^* \subseteq \mathcal{K}'$, an infimum over $\mathcal{K}'$ is at most the infimum over $\mathcal{K}^*$. By (69) and monotonicity, $\alpha(\mathcal{K}') \le \alpha(\mathcal{K}^*)$.

The case $\theta(\mathbf{x}) = \|\mathbf{x}\|_p^\sigma$. Suppose first that $\sigma > 0$. Since $\|s\mathbf{x}\|_p = s\|\mathbf{x}\|_p$ for $s \ge 0$,

$$\Phi(\mathbf{x}) = \|\mathbf{x}\|_p^\sigma \int_0^1 s^\sigma ds = \frac{\|\mathbf{x}\|_p^\sigma}{\sigma + 1}. \tag{72}$$

Since $t \mapsto t^\sigma$ is nondecreasing on $[0, \infty)$, we have $R_\theta(\mathcal{K}) = \max_{\mathbf{v},\mathbf{u}\in\mathcal{K}} \|\mathbf{v} \vee \mathbf{u}\|_p^\sigma = R_p(\mathcal{K})^\sigma$, and $\inf_{\mathbf{x}\in\mathcal{K}'} \|\mathbf{x}\|_p^\sigma = (\inf_{\mathbf{x}\in\mathcal{K}'} \|\mathbf{x}\|_p)^\sigma$. The set $\mathcal{K}'$ is closed and contained in $[0, 1]^d$, hence compact, and $\|\cdot\|_p$ is continuous, so the infimum is a minimum. Substituting into (69) gives

$$\alpha(\mathcal{K}') = 1 - \exp\left(-\frac{\gamma}{\sigma + 1}\left(\frac{\min_{\mathbf{x}\in\mathcal{K}'} \|\mathbf{x}\|_p}{R_p(\mathcal{K})}\right)^\sigma\right).$$

If $\sigma = 0$, then $\theta \equiv 1$, so $\Phi \equiv 1$ and $R = 1$, and (69) gives $\alpha(\mathcal{K}') = 1 - e^{-\gamma}$.

The Lipschitz case. Assume that θ is L-Lipschitz with respect to $\|\cdot\|$, and let

$$I^* := \inf_{\mathbf{v} \in \mathcal{K}^*} \Phi(\mathbf{v}), \qquad \text{so that} \qquad 1 - \alpha(\mathcal{K}^*) = \exp\left(-\frac{\gamma I^*}{R}\right). \tag{73}$$

Fix $\mathbf{x} \in \mathcal{K}'$ and $\eta > 0$. Since $\mathbf{x}$ is within distance ε of $\mathcal{K}^*$, there is $\mathbf{v} \in \mathcal{K}^*$ with $\|\mathbf{x} - \mathbf{v}\| \le \varepsilon + \eta$. For every $s \in [0, 1]$, the Lipschitz property gives

$$\theta(s\mathbf{x}) \ge \theta(s\mathbf{v}) - L\|s\mathbf{x} - s\mathbf{v}\| = \theta(s\mathbf{v}) - Ls\|\mathbf{x} - \mathbf{v}\| \ge \theta(s\mathbf{v}) - Ls(\varepsilon + \eta). \tag{74}$$

Integrating (74) over $s \in [0, 1]$, and using $\int_0^1 s\, ds = \frac{1}{2}$ and $\Phi(\mathbf{v}) \ge I^*$,

$$\Phi(\mathbf{x}) \ge \Phi(\mathbf{v}) - \frac{L(\varepsilon + \eta)}{2} \ge I^* - \frac{L(\varepsilon + \eta)}{2}. \tag{75}$$

The left-hand side of (75) does not depend on η, so letting $\eta \downarrow 0$ gives $\Phi(\mathbf{x}) \ge I^* - L\varepsilon/2$. Also $\Phi(\mathbf{x}) \ge 0$ because $\theta \ge 0$. Taking the infimum over $\mathbf{x} \in \mathcal{K}'$,

$$\inf_{\mathbf{x} \in \mathcal{K}'} \Phi(\mathbf{x}) \ge \max\left\{0,\ I^* - \frac{L\varepsilon}{2}\right\}. \tag{76}$$

By (69) and monotonicity, (76) implies $\alpha(\mathcal{K}') \ge 1 - \exp\big(-\frac{\gamma}{R}\max\{0, I^* - \frac{L\varepsilon}{2}\}\big)$. We now compare this with the claimed bound. If $I^* \ge L\varepsilon/2$, then by (73),

$$1 - \exp\left(-\frac{\gamma}{R}\Big(I^* - \frac{L\varepsilon}{2}\Big)\right) = 1 - \exp\left(-\frac{\gamma I^*}{R}\right)\exp\left(\frac{\gamma L\varepsilon}{2R}\right) = 1 - (1 - \alpha(\mathcal{K}^*))\exp\left(\frac{\gamma L\varepsilon}{2R}\right),$$

and this quantity is nonnegative, so it equals the maximum in the claim. If $I^* < L\varepsilon/2$, the lower bound above is $1 - e^0 = 0$, while $1 - (1 - \alpha(\mathcal{K}^*))\exp(\frac{\gamma L\varepsilon}{2R}) = 1 - \exp(\frac{\gamma}{R}(\frac{L\varepsilon}{2} - I^*)) < 0$, so the maximum in the claim is 0. In both cases, $\alpha(\mathcal{K}') \ge \max\{0, 1 - (1 - \alpha(\mathcal{K}^*))\exp(\frac{\gamma L\varepsilon}{2R})\}$. The upper bound $\alpha(\mathcal{K}') \le \alpha(\mathcal{K}^*)$ was shown above. □

H Separation Oracle for $\mathcal{K}^*$ via Benson-Type Methods

Let $\mathcal{K} \subseteq [0, 1]^d$ be a nonempty rational polyhedron. Recall that

$$\mathcal{K}^{\mathrm{m}} = \{\mathbf{x} \in \mathcal{K} : \nexists\, \mathbf{y} \in \mathcal{K},\ \mathbf{y} \ge \mathbf{x},\ \mathbf{y} \ne \mathbf{x}\}, \qquad \mathcal{K}^* = \overline{\mathrm{conv}}(\mathcal{K}^{\mathrm{m}}).$$

We assume that $\mathcal{K}$ is given by a finite rational inequality description, and we use exact linear-programming and polyhedral-enumeration subroutines.

Vector optimization formulation. Consider

$$(\mathsf{P}) \qquad \min_{\le \mathbb{R}^d_+} (-I_d)\mathbf{x} \quad \text{s.t. } \mathbf{x} \in \mathcal{K}.$$

Then $\mathbf{x} \in \mathcal{K}$ is a minimizer if and only if $\mathbf{x} \in \mathcal{K}^{\mathrm{m}}$.

For a linear map P and set S, write $P[S] = \{P\mathbf{x} \mid \mathbf{x} \in S\}$. We use the following standard result about the classical primal Benson algorithm (see [3, 22]).

Theorem 9 (Primal Benson, bounded case)**.** *Let $\mathcal{K}$ be a nonempty bounded polyhedron. For*

$$\min_{\le C} P\mathbf{x} \quad \textit{s.t. } \mathbf{x} \in \mathcal{K}, \quad C = \mathbb{R}^d_+,$$

the primal Benson algorithm terminates finitely and returns a finite set $\mathcal{E} \subseteq \mathcal{K}$ of minimizers such that

$$\mathrm{conv}(P[\mathcal{E}]) + C = P[\mathcal{K}] + C.$$

Applying with $P = -I_d$, we obtain $\mathcal{E} \subseteq \mathcal{K}^{\mathrm{m}}$ with

$$\mathcal{K}^* = \mathrm{conv}(\mathcal{E}).$$

Indeed, let $\mathbf{x} \in \mathcal{K}^{\mathrm{m}}$. The identity above with $P = -I_d$ gives $-\mathbf{x} = -\mathbf{v} + \mathbf{c}$ for some $\mathbf{v} \in \mathrm{conv}(\mathcal{E}) \subseteq \mathcal{K}$ and $\mathbf{c} \ge \mathbf{0}$. Thus $\mathbf{v} \ge \mathbf{x}$, and maximality of $\mathbf{x}$ forces $\mathbf{v} = \mathbf{x}$. Hence $\mathcal{K}^{\mathrm{m}} \subseteq \mathrm{conv}(\mathcal{E})$, and since $\mathrm{conv}(\mathcal{E})$ is closed and convex, $\mathcal{K}^* \subseteq \mathrm{conv}(\mathcal{E})$. Conversely, $\mathcal{E} \subseteq \mathcal{K}^{\mathrm{m}}$ gives $\mathrm{conv}(\mathcal{E}) \subseteq \mathcal{K}^*$.

Separation oracle (primal version). Membership in $\mathcal{K}^*$ reduces to convex hull feasibility. If it is infeasible, Farkas' lemma gives a separating hyperplane.

The numbers of generated vertices and faces can be exponential in the description size of $\mathcal{K}$, so this construction is finite but potentially expensive, and it does not give a polynomial-time oracle for $\mathcal{K}^*$.

Remark 1 and Proposition 1 describe the alternative of running the algorithm on a larger set $\mathcal{K}'$.

I Proof of Theorem 6

Proof. Write $\rho = \rho(\mathcal{M})$ and, for $S \subseteq V$, let $\mathbf{1}_S$ be its indicator vector. We use repeatedly that, for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d_{\geq 0}$,

$$\|\mathbf{x} + \mathbf{y}\|_1 = \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1 \qquad \text{and} \qquad \|\mathbf{1}_S\|_1 = |S|. \tag{77}$$

(a) $P_{\mathcal{I}}$ is closed, convex and downward-closed. As the convex hull of finitely many points, $P_{\mathcal{I}}$ is compact and convex. For downward-closedness, we first show that one coordinate can be decreased. Let $\mathbf{x} \in P_{\mathcal{I}}$, write $\mathbf{x} = \sum_{A \in \mathcal{I}} \lambda_A \mathbf{1}_A$ with $\lambda_A \geq 0$ and $\sum_A \lambda_A = 1$, fix a coordinate i, and fix $t \in [0, 1]$. Note that

$$x_i = \sum_{A \in \mathcal{I}:\, i \in A} \lambda_A. \tag{78}$$

For every $A \in \mathcal{I}$, the set $A \setminus \{i\}$ is independent by the downward-closed property of matroids, so $(1-t)\mathbf{1}_A + t\mathbf{1}_{A\setminus\{i\}} \in P_{\mathcal{I}}$. Taking the convex combination of these points with weights λ_A, we obtain a point of $P_{\mathcal{I}}$, namely

$$\sum_A \lambda_A \big((1-t)\mathbf{1}_A + t\mathbf{1}_{A\setminus\{i\}}\big) = \mathbf{x} - t\sum_{A \ni i} \lambda_A \mathbf{e}_i = \mathbf{x} - t x_i \mathbf{e}_i,$$

where we used $\mathbf{1}_A - \mathbf{1}_{A\setminus\{i\}} = \mathbf{e}_i$ if $i \in A$ and $= \mathbf{0}$ otherwise, and then (78). As t ranges over $[0, 1]$, the i-th coordinate $(1-t)x_i$ takes every value in $[0, x_i]$, while the other coordinates are unchanged. Now let $\mathbf{x} \in P_{\mathcal{I}}$ and $\mathbf{0} \leq \mathbf{y} \leq \mathbf{x}$. Starting from $\mathbf{x}$, apply this operation to the coordinates $i = 1, \ldots, d$ in turn, lowering the i-th coordinate from x_i to y_i (with $t = 1 - y_i/x_i$ if $x_i > 0$, and no change if $x_i = 0$, in which case $y_i = 0$). Each step stays in $P_{\mathcal{I}}$, and the final point is $\mathbf{y}$. Hence $\mathbf{y} \in P_{\mathcal{I}}$.

(b) The ℓ_1 norm on $P_{\mathcal{I}}$. Every independent set A is contained in a maximal independent set, i.e., a basis, since V is finite; all bases have ρ elements. Hence

$$|A| \leq \rho \text{ for all } A \in \mathcal{I}, \text{ with equality if and only if } A \text{ is a basis.} \tag{79}$$

Let $\mathbf{x} = \sum_A \lambda_A \mathbf{1}_A \in P_{\mathcal{I}}$ be any representation as a convex combination. By (77) and (79),

$$\|\mathbf{x}\|_1 = \sum_A \lambda_A |A| \leq \rho. \tag{80}$$

If $\mathbf{x} \in P_{\mathcal{B}}$, we may choose the representation using bases only, and then (80) holds with equality. Conversely, if $\|\mathbf{x}\|_1 = \rho$, then every A with $\lambda_A > 0$ has $|A| = \rho$, so it is a basis by (79), and $\mathbf{x} \in P_{\mathcal{B}}$. Since bases are independent, $P_{\mathcal{B}} \subseteq P_{\mathcal{I}}$, and therefore

$$P_{\mathcal{B}} = P_{\mathcal{I}} \cap \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_1 = \rho\}. \tag{81}$$

(c) Every point of $P_{\mathcal{I}}$ is dominated by a point of $P_{\mathcal{B}}$. Let $\mathbf{x} = \sum_A \lambda_A \mathbf{1}_A \in P_{\mathcal{I}}$. For each A with $\lambda_A > 0$, choose a basis $B_A \supseteq A$, and let $\mathbf{y} = \sum_A \lambda_A \mathbf{1}_{B_A}$. Then $\mathbf{y} \in P_{\mathcal{B}}$, and $\mathbf{y} \geq \mathbf{x}$ because $\mathbf{1}_{B_A} \geq \mathbf{1}_A$ for each A.

(d) $(P_{\mathcal{I}})^{\mathrm{m}} = P_{\mathcal{B}} = (P_{\mathcal{I}})^*$. Let $\mathbf{x} \in (P_{\mathcal{I}})^{\mathrm{m}}$. By (c), there is $\mathbf{y} \in P_{\mathcal{B}} \subseteq P_{\mathcal{I}}$ with $\mathbf{y} \geq \mathbf{x}$, and maximality of $\mathbf{x}$ forces $\mathbf{y} = \mathbf{x}$, so $\mathbf{x} \in P_{\mathcal{B}}$. Conversely, let $\mathbf{x} \in P_{\mathcal{B}}$ and suppose that $\mathbf{y} \in P_{\mathcal{I}}$ satisfies $\mathbf{y} \geq \mathbf{x}$ and $\mathbf{y} \neq \mathbf{x}$. Then $\mathbf{y} - \mathbf{x} \geq \mathbf{0}$ is nonzero, and by (77), $\|\mathbf{y}\|_1 = \|\mathbf{x}\|_1 + \|\mathbf{y} - \mathbf{x}\|_1 > \rho$, which contradicts (80). Hence $\mathbf{x} \in (P_{\mathcal{I}})^{\mathrm{m}}$, and $(P_{\mathcal{I}})^{\mathrm{m}} = P_{\mathcal{B}}$. Since $P_{\mathcal{B}}$ is closed and convex, $(P_{\mathcal{I}})^* = \overline{\mathrm{conv}}(P_{\mathcal{B}}) = P_{\mathcal{B}}$. The same argument, with $P_{\mathcal{B}}$ in place of $P_{\mathcal{I}}$, shows that no point of $P_{\mathcal{B}}$ is strictly dominated by another point of $P_{\mathcal{B}}$, so $(P_{\mathcal{B}})^{\mathrm{m}} = (P_{\mathcal{B}})^* = P_{\mathcal{B}}$. Together with (81), this proves the first line of the theorem.

(e) $r_1(P_{\mathcal{I}}) = r_1(P_{\mathcal{B}}) = \rho$. By (d) and (81), $r_1(P_{\mathcal{I}}) = \min_{\mathbf{v} \in P_{\mathcal{B}}} \|\mathbf{v}\|_1 = \rho$ and $r_1(P_{\mathcal{B}}) = \min_{\mathbf{v} \in P_{\mathcal{B}}} \|\mathbf{v}\|_1 = \rho$.

(f) $R_1(P_{\mathcal{I}}) = R_1(P_{\mathcal{B}}) \le 2\rho$. Since $P_{\mathcal{B}} \subseteq P_{\mathcal{I}}$, we have $R_1(P_{\mathcal{B}}) \le R_1(P_{\mathcal{I}})$. Conversely, let $\mathbf{u}, \mathbf{v} \in P_{\mathcal{I}}$ and, by (c), choose $\mathbf{u}', \mathbf{v}' \in P_{\mathcal{B}}$ with $\mathbf{u}' \ge \mathbf{u}$ and $\mathbf{v}' \ge \mathbf{v}$. Then $\mathbf{u}' \vee \mathbf{v}' \ge \mathbf{u} \vee \mathbf{v} \ge \mathbf{0}$, so $\|\mathbf{u} \vee \mathbf{v}\|_1 \le \|\mathbf{u}' \vee \mathbf{v}'\|_1 \le R_1(P_{\mathcal{B}})$. Taking the maximum over $\mathbf{u}, \mathbf{v}$ gives $R_1(P_{\mathcal{I}}) \le R_1(P_{\mathcal{B}})$. Finally, for $\mathbf{u}, \mathbf{v} \in P_{\mathcal{I}}$, each coordinate satisfies $\max\{u_i, v_i\} \le u_i + v_i$ because $u_i, v_i \ge 0$, so by (77) and (80),

$$\|\mathbf{u} \vee \mathbf{v}\|_1 \le \|\mathbf{u} + \mathbf{v}\|_1 = \|\mathbf{u}\|_1 + \|\mathbf{v}\|_1 \le 2\rho.$$

Taking the maximum over $\mathbf{u}, \mathbf{v} \in P_{\mathcal{I}}$ gives $R_1(P_{\mathcal{I}}) \le 2\rho$. □

J Comparison with Prior Approximation Ratios for σ-OSS Maximization

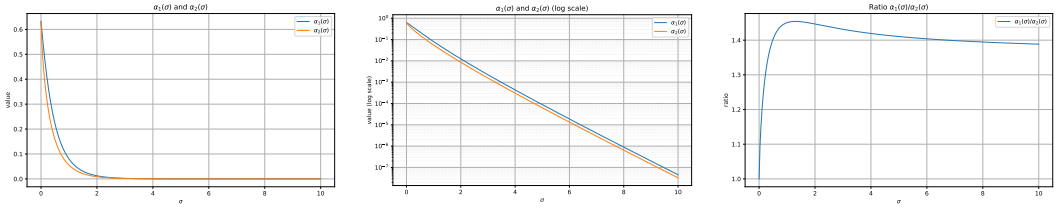

Figure 2: Comparison of $\alpha_1(\sigma)$ and $\alpha_2(\sigma)$, the same values on a logarithmic scale, and their ratio $\alpha_1(\sigma)/\alpha_2(\sigma)$.

Theorem 10. *Let*

$$\alpha_1(\sigma) = 1 - \exp\Big(-\frac{1}{2^{2\sigma}(2\sigma+1)}\Big)$$

and

$$\alpha_2(\sigma) = \sup_{t\in[0,1)} \left(1 - \exp\left(-(1-t)\left(\frac{t}{t+1}\right)^{2\sigma}\right)\right),$$

for $\sigma \ge 0$. Then we have $\alpha_1(\sigma) \ge \alpha_2(\sigma)$ with equality only at $\sigma = 0$. Moreover, we have

$$\lim_{\sigma\to\infty} \frac{\alpha_1(\sigma)}{\alpha_2(\sigma)} = \frac{e}{2}.$$

Proof. For $\sigma \ge 0$, let

$$A(\sigma) = \frac{1}{2^{2\sigma}(1+2\sigma)}, \qquad B(\sigma) = \max_{0\le t<1} (1-t)\Big(\frac{t}{1+t}\Big)^{2\sigma}.$$

Note that $\alpha_1(\sigma) = 1 - \exp(-A(\sigma))$ and $\alpha_2(\sigma) = 1 - \exp(-B(\sigma))$. We will prove that

$$A(\sigma) \ge B(\sigma) \quad \text{for all } \sigma \ge 0,$$

with equality if and only if $\sigma = 0$, and that

$$\lim_{\sigma\to\infty} \frac{A(\sigma)}{B(\sigma)} = \frac{e}{2}.$$

Given these two facts, $A(\sigma) \to 0$ and $0 \le B(\sigma) \le A(\sigma)$ imply that $B(\sigma) \to 0$ and therefore

$$\begin{aligned}
\lim_{\sigma\to\infty} \frac{\alpha_1(\sigma)}{\alpha_2(\sigma)} &= \lim_{\sigma\to\infty} \frac{1-\exp(-A(\sigma))}{1-\exp(-B(\sigma))} \\
&= \lim_{\sigma\to\infty} \frac{1-\exp(-A(\sigma))}{A(\sigma)} \frac{B(\sigma)}{1-\exp(-B(\sigma))} \frac{A(\sigma)}{B(\sigma)} \\
&= \lim_{\sigma\to\infty} \frac{A(\sigma)}{B(\sigma)} = \frac{e}{2},
\end{aligned}$$

which completes the proof.

Proof of $A(\sigma) \ge B(\sigma)$: For $\sigma = 0$, we have

$$A(0) = 1, \qquad B(0) = \sup_{0\le t<1} (1-t) = 1,$$

so equality holds. Now assume $\sigma > 0$. Set

$$f_\sigma(t) := (1-t)\Big(\frac{t}{1+t}\Big)^{2\sigma}, \qquad 0 \le t < 1.$$

Since $f_\sigma(0) = 0$ and $f_\sigma(t) \to 0$ as $t \to 1^-$, the maximum is attained at some interior point $t_\sigma \in (0,1)$. Differentiating,

$$\frac{d}{dt}\log f_\sigma(t) = -\frac{1}{1-t} + 2\sigma\Big(\frac{1}{t} - \frac{1}{1+t}\Big) = -\frac{1}{1-t} + \frac{2\sigma}{t(1+t)}.$$

Thus t_σ satisfies

$$-\frac{1}{1-t_\sigma} + \frac{2\sigma}{t_\sigma(1+t_\sigma)} = 0,$$

or equivalently,

$$2\sigma(1-t_\sigma) = t_\sigma(1+t_\sigma). \tag{82}$$

Using (82),

$$1 + 2\sigma = 1 + \frac{t_\sigma(1+t_\sigma)}{1-t_\sigma} = \frac{1+t_\sigma^2}{1-t_\sigma}.$$

Therefore

$$\frac{B(\sigma)}{A(\sigma)} = 2^{2\sigma}(1+2\sigma)(1-t_\sigma)\Big(\frac{t_\sigma}{1+t_\sigma}\Big)^{2\sigma} = (1+t_\sigma^2)\Big(\frac{2t_\sigma}{1+t_\sigma}\Big)^{2\sigma}.$$

Again by (82),

$$\log\frac{B(\sigma)}{A(\sigma)} = \log(1+t_\sigma^2) + \frac{t_\sigma(1+t_\sigma)}{1-t_\sigma}\log\Big(\frac{2t_\sigma}{1+t_\sigma}\Big) = h(t_\sigma),$$

where

$$h(t) := \log(1+t^2) + \frac{t(1+t)}{1-t}\log\Big(\frac{2t}{1+t}\Big), \qquad 0 < t < 1.$$

We claim that $h(t) < 0$ for all $t \in (0,1)$. Set

$$u = \frac{1-t}{1+t} \in (0,1).$$

Then

$$t = \frac{1-u}{1+u}, \qquad \frac{2t}{1+t} = 1-u, \qquad 1+t^2 = \frac{2(1+u^2)}{(1+u)^2}, \qquad \frac{t(1+t)}{1-t} = \frac{1-u}{u(1+u)}.$$

Thus

$$h(t) = \log\Big(\frac{2(1+u^2)}{(1+u)^2}\Big) + \frac{1-u}{u(1+u)}\log(1-u).$$

Using $\log(1-u) \le -u$ for $0 < u < 1$, we get

$$h(t) \le \log\Big(\frac{2(1+u^2)}{(1+u)^2}\Big) - \frac{1-u}{1+u}.$$

Define

$$\phi(u) := \log\Big(\frac{2(1+u^2)}{(1+u)^2}\Big) - \frac{1-u}{1+u}, \qquad 0 < u \le 1.$$

Differentiating,

$$\begin{aligned}
\phi'(u) &= \frac{d}{du}\Big[\log(2(1+u^2)) - 2\log(1+u)\Big] - \frac{d}{du}\Big(\frac{1-u}{1+u}\Big). \\
&= \frac{2u}{1+u^2} - \frac{2}{1+u} + \frac{2}{(1+u)^2} \\
&= \frac{2u(1+u)^2 - 2(1+u)(1+u^2) + 2(1+u^2)}{(1+u)^2(1+u^2)} \\
&= \frac{2u(1+2u+u^2) - 2(1+u+u^2+u^3) + 2(1+u^2)}{(1+u)^2(1+u^2)} \\
&= \frac{4u^2}{(1+u)^2(1+u^2)} > 0
\end{aligned}$$

Hence ϕ is increasing on $(0,1]$. Since $\phi(1) = \log 1 - 0 = 0$, it follows that $\phi(u) < 0$, for all $0 < u < 1$. Therefore $h(t) < 0$ for every $t \in (0,1)$, and so

$$\log \frac{B(\sigma)}{A(\sigma)} = h(t_\sigma) < 0.$$

Thus, we have shown that

$$A(\sigma) > B(\sigma), \quad \text{for all } \sigma > 0.$$

Together with the case $\sigma = 0$, this proves

$$A(\sigma) \geq B(\sigma), \quad \text{for all } \sigma \geq 0,$$

with equality if and only if $\sigma = 0$.

Computing the limit of $A(\sigma)/B(\sigma)$: It remains to compute the asymptotic ratio. Define

$$\eta_\sigma := \frac{1}{1 - t_\sigma}.$$

Then $t_\sigma = 1 - \frac{1}{\eta_\sigma}$, and so

$$t_\sigma(1 + t_\sigma) = \left(1 - \frac{1}{\eta_\sigma}\right)\left(2 - \frac{1}{\eta_\sigma}\right) = 2 - \frac{3}{\eta_\sigma} + \frac{1}{\eta_\sigma^2}.$$

Substituting into (82) gives

$$\frac{2\sigma}{\eta_\sigma} = 2 - \frac{3}{\eta_\sigma} + \frac{1}{\eta_\sigma^2}.$$

By multiplying both sides by $\eta_\sigma/2$ and rearranging, we see

$$\eta_\sigma = \sigma + \frac{3}{2} - \frac{1}{2\eta_\sigma}.$$

Since $t_\sigma \in (0,1)$, we have $\eta_\sigma > 1$. Therefore

$$|\eta_\sigma - \sigma| = \frac{3}{2} - \frac{1}{2\eta_\sigma} < 2.$$

Now

$$B(\sigma) = (1 - t_\sigma)\Big(\frac{t_\sigma}{1 + t_\sigma}\Big)^{2\sigma} = \frac{1}{\eta_\sigma}\left(\frac{1 - \eta_\sigma^{-1}}{2 - \eta_\sigma^{-1}}\right)^{2\sigma} = \frac{2^{-2\sigma}}{\eta_\sigma}\left(1 - \frac{1}{2\eta_\sigma - 1}\right)^{2\sigma}.$$

Since $|\eta_\sigma - \sigma| < 2$, we see that

$$\begin{aligned}\lim_{\sigma\to\infty} 2\sigma \log\left(1 - \frac{1}{2\eta_\sigma - 1}\right) &= \lim_{\sigma\to\infty} \frac{2\sigma}{2\eta_\sigma - 1}(2\eta_\sigma - 1)\log\left(1 - \frac{1}{2\eta_\sigma - 1}\right) \\ &= \lim_{\sigma\to\infty} \frac{\log\left(1 - \frac{1}{2\eta_\sigma - 1}\right)}{\frac{1}{2\eta_\sigma - 1}} = -1,\end{aligned}$$

where the last equality follows from the standard limit $\lim_{x\to 0} \frac{\log(1-x)}{x} = -1$. Therefore

$$\lim_{\sigma\to\infty} \frac{B(\sigma)}{2^{-2\sigma}/\sigma} = \lim_{\sigma\to\infty} \frac{\sigma}{\eta_\sigma}\left(1 - \frac{1}{2\eta_\sigma - 1}\right)^{2\sigma} = \lim_{\sigma\to\infty}\left(1 - \frac{1}{2\eta_\sigma - 1}\right)^{2\sigma} = e^{-1}.$$

Thus

$$B(\sigma) \sim \frac{2^{-2\sigma}}{e\,\sigma}.$$

On the other hand,

$$A(\sigma) = \frac{1}{2^{2\sigma}(1 + 2\sigma)},$$

so

$$\frac{A(\sigma)}{2^{-2\sigma}/\sigma} = \frac{\sigma}{1 + 2\sigma} \longrightarrow \frac{1}{2}.$$

Therefore

$$\frac{A(\sigma)}{B(\sigma)} = \frac{A(\sigma)}{2^{-2\sigma}/\sigma} \cdot \frac{2^{-2\sigma}/\sigma}{B(\sigma)} \longrightarrow \frac{1}{2} \cdot e = \frac{e}{2}.$$

Hence

$$\lim_{\sigma\to\infty} \frac{A(\sigma)}{B(\sigma)} = \frac{e}{2}.$$

□

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims stated in the abstract and introduction are consistent with the main result established in the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The computational dependence on a set oracle for $\mathcal{K}^*$ is discussed in Remark 1; Proposition 1 explains when a larger action domain yields a zero coefficient; Definition 2 states the integration assumption for a general scaling; and Section 7 specifies the feedback and comparator restrictions of the online guarantees.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: All the setup assumptions are provided in the paper.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [N/A]

Justification: This paper primarily focuses on theoretical results.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [N/A]

   Justification: The paper does not focus on experiments.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A]

   Justification: This paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A]

   Justification: This paper does not include experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A]

   Justification: This paper focuses on the theoretical framework.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: To the best of our knowledge, the research in this paper conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This paper is primarily theoretical and aims to advance the current literature. We do not anticipate direct societal impacts from the work

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The research in this paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: This paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The methods developed and analyzed in this paper do not use LLMs as a substantive component.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.