SUBMODULAR MAXIMIZATION WITH PREDICTIONS: THE EXACT PRICE OF ERROR

Anonymous authors
Paper under double-blind review

ABSTRACT

Maximizing a monotone submodular function under a cardinality constraint K has a classical approximation guarantee, but in many subset-selection tasks the objective cannot be evaluated while the set is chosen, and an algorithm maximizes a learned model of the objective, a surrogate. Prior work bounds the surrogate's error on function values and is pessimistic: no algorithm with polynomially many queries retains a constant fraction of the optimum. The classical greedy guarantee rests on marginal gains, and we measure the error there: the surrogate may distort any marginal gain by at most a factor η. We consider deterministic algorithms that query only the surrogate and determine three prices. The information price is $\frac{1}{\eta}$: with unlimited queries, no deterministic algorithm guarantees more, and exhaustive search on the surrogate attains it. Predictive greedy, the greedy algorithm applied to the surrogate, pays exactly $\rho_K(\eta)$, an explicit function of η and K that tends to $1-e^{-1/\eta}$; the classical approximate-greedy bound is not tight for a surrogate. The price of the budget accounts for the rest: within greedy's query budget, no deterministic algorithm, whatever sets it queries, guarantees more than $\rho_K(\eta)$, so predictive greedy is optimal there. The error along the greedy path gives a per-instance guarantee and, on learned surrogates for feature selection, influence maximization, and summarization, ranks surrogates by their true performance.

1 INTRODUCTION

Maximizing a monotone submodular function under a cardinality constraint K is a classical problem with a provable approximation guarantee (Nemhauser et al., 1978), arising in feature selection (Das & Kempe, 2011), sensor placement (Krause et al., 2008), and influence maximization (Kempe et al., 2003). The guarantee applies when the objective can be evaluated. In many applications, however, it cannot be evaluated while the set is chosen: data selection aims at accuracy on unseen data (Wei et al., 2015), summarization at quality judged by readers (Lin & Bilmes, 2011), and influence maximization at spread on the true network (Balkanski et al., 2017). An algorithm therefore maximizes a surrogate, an approximate version of the objective. One question follows: does maximizing the surrogate still guarantee a constant fraction of the optimum, and how does the guarantee depend on the surrogate's error? Prior work answers pessimistically: a surrogate learned from samples can be useless for optimization (Balkanski et al., 2017), and with a surrogate off by a small factor on every set (Horel & Singer, 2016) or adversarially noisy (Hassidim & Singer, 2017), no algorithm with polynomially many queries retains a constant fraction. All three measure the surrogate's error on values, but the classical guarantee rests on marginal gains. Measured there, the error has a price, and we compute it exactly. No deterministic algorithm with the same query budget beats greedy.

Greedy, the algorithm behind the classical guarantee, starts from the empty set and adds, K times, the element of largest marginal gain (Nemhauser et al., 1978). Greedy uses nothing but marginal gains. Applied to the surrogate, greedy compares predicted marginal gains; we call this *predictive greedy*. Figure 1 shows why the place of measurement matters. Measured on values, A is the better surrogate, within 10% of f against 30% for B. Yet predictive greedy finds the optimum with B and only 0.83 of the optimum with A. The loss comes at the second step, where b truly adds twice what c adds. Surrogate A predicts 1.7 for b and 1.85 for c, each set value still within 10% but b's gain 43% too low, so A adds c. Surrogate B, 30% high everywhere, keeps the true order and adds b. A small error in set values can be a large error in a marginal gain. For predictive greedy, value accuracy is

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

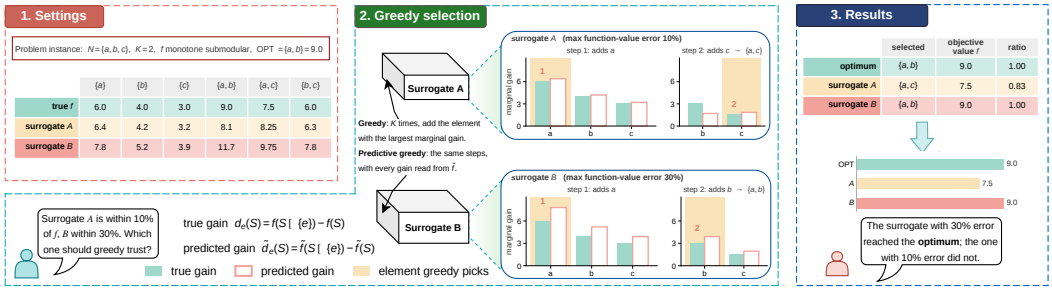

Figure 1: Toy example: $N = \{a, b, c\}$, $K = 2$, f monotone submodular with $f(\emptyset) = 0$, optimum $\{a, b\}$ with $f = 9.0$. **Settings** lists f and the two surrogates on every nonempty set of size at most K; the largest relative value error is 10% for A and 30% for B. **Greedy selection** plots the marginal gains at each step, true $d_e(S) = f(S \cup \{e\}) - f(S)$ in green and predicted $\tilde{d}_e(S) = \tilde{f}(S \cup \{e\}) - \tilde{f}(S)$ in pink, with the added element shaded. **Results** evaluate the selections with the true f: A reaches 7.5 and B reaches the optimum.

neither necessary, as B shows, nor sufficient: however accurate the values, there are instances on which predictive greedy's share of the optimum shrinks like $\frac{1}{K}$ as K grows (Hassidim & Singer, 2017). Suppose instead that every predicted marginal gain is within a fixed factor of the true one. The classical analysis rests on one fact: at every step some element of the optimal set closes at least one K-th of the gap between the current value and the optimum, and greedy picks the element that closes the most. With the factor, the element predictive greedy picks may close less, but at most by that factor, and K such steps still leave a constant fraction of the optimum, smaller for a larger factor (Goundan & Schulz, 2007). Two questions follow, both settled when f is known (Nemhauser et al., 1978; Nemhauser & Wolsey, 1978) and, until now, open with a surrogate:

How much of the optimum does predictive greedy guarantee as a function of the surrogate's marginal-gain error, and does any deterministic algorithm with the same query budget do better?

We analyze deterministic algorithms at three levels of query budget. (1) The budget is unlimited. An adversary with an uninformative surrogate gives an upper bound on the best guarantee: the algorithm commits to its output before the adversary chooses the objective. An exchange argument shows that exhaustive search over predicted values attains this bound. (2) The budget is greedy's own, with at most K elements per query. We determine greedy's guarantee exactly and show that no deterministic algorithm with this budget does better. The exact value follows the linear-programming route of Nemhauser et al. (1978, Section 4): linear inequalities satisfied by every run, a dual solution certifying the lower bound, and an instance attaining it. The classical analysis of Goundan & Schulz (2007) treats every step as an independent worst case. For a surrogate, this analysis is loose because the errors at all steps come from one function. We add inequalities on true gains that link each step to the next: adding two elements in either order gives the same total predicted gain. The dual multipliers change only at integer values of the error; the exact value is therefore piecewise in the error. Optimality within the budget rests on a hiding argument: we build a surrogate blind to any single optimal element. A deterministic algorithm therefore cannot tell this surrogate from an uninformative one unless a query contains two optimal elements. We choose the optimal set so that no query does. (3) The budget is polynomial, with degree independent of K and at most K elements per query. Blinding the surrogate to larger groups of optimal elements extends the hiding argument.

Technical contributions. We answer the two questions with three prices: the information price and the price greedy pays, both exact, and the price of the budget, exact within greedy's own budget. The information price is $1/\eta$, the best approximation ratio of any deterministic algorithm with unlimited queries (Proposition 3). The price greedy pays is $\rho_K(\eta)$, a closed form that tends to $1 - e^{-1/\eta}$ as K grows and exceeds the classical bound for every $\eta > 1$ and $K \geq 2$ (Theorem 1); to our knowledge, the first exact approximation ratio of greedy under a marginal-gain error. Within its budget of nK queries on sets of at most K elements, greedy is optimal among deterministic algorithms (Theorem 2), and with n^c such queries for any constant c, no deterministic algorithm guarantees more than a bound with the same limit (Theorem 3). The gap between $\rho_K(\eta)$ and $1/\eta$ is therefore the price of the budget, not of greedy. We order three errors, $\eta^{\mathrm{sel}} \leq \eta^{\mathrm{tr}} \leq \eta$: the classical bound certifies a run through its selection error, and the exact value requires the global error.

2 MODEL AND PRELIMINARIES

We study monotone submodular maximization under a cardinality constraint in the prediction-only regime: the algorithm cannot evaluate the objective while it selects and queries a surrogate $\tilde{f}$ instead. A ground set N of n elements carries a monotone submodular objective $f : 2^N \to \mathbb{R}_{\geq 0}$ with $f(\emptyset) = 0$ and a budget K, $1 \leq K \leq n$. The marginal gain of $e \notin S$ is $d_e(S) = f(S \cup \{e\}) - f(S)$; monotone means $d_e(S) \geq 0$, and submodular means $d_e(T) \leq d_e(S)$ whenever $S \subseteq T$ and $e \notin T$. The problem is to maximize $f(S)$ subject to $|S| \leq K$, and O^* denotes an optimal set. The surrogate is a set function $\tilde{f} : 2^N \to \mathbb{R}$ with $\tilde{f}(\emptyset) = 0$ and marginal gains $\tilde{d}_e(S)$ defined likewise; the algorithm is judged by the true value $f(S)$ of the set it returns. We place *no structural assumption* on $\tilde{f}$: a trained model or a cross-validation estimate need not be submodular. Each evaluation of $\tilde{f}$ is one query, and a class of algorithms is described by its number of queries, as a function of n and K, and by the largest set it queries. Predictive greedy is the greedy algorithm of Nemhauser et al. (1978) applied to the surrogate: from $S^0 = \emptyset$, for $t = 0, \ldots, K-1$ it adds $e_t \in \arg\max_{e \notin S^t} \tilde{d}_e(S^t)$, with ties broken adversarially. It runs all K steps, even when every remaining predicted gain is zero, and makes $nK - K(K-1)/2$ queries, each on a set of size at most K. For an algorithm A with output $A^{\tilde{f}}$ on the instance $(f, \tilde{f})$, the approximation ratio at error η is $\alpha(A) = \inf f(A^{\tilde{f}})/f(O^*)$ over instances on n elements with marginal-gain error at most η (Definition 1) and $f(O^*) > 0$, for predictive greedy also over tie resolutions; for a class $\mathcal{A}$, $\alpha^*(\mathcal{A}) = \sup_{A \in \mathcal{A}} \alpha(A)$ is the best approximation ratio the class admits. Both ratios depend on n, K and η, which the notation suppresses, and $\rho_K(\eta) = \alpha(\text{predictive greedy})$. Table 2 in Appendix A.1 collects the notation.

The error is measured on marginal gains, multiplicatively and in both directions.

Definition 1 (Marginal-gain error)**.** *A surrogate $\tilde{f}$ has marginal-gain error $\eta = \eta_u \eta_o$, for some $\eta_u, \eta_o > 0$ with $\eta \geq 1$, if $\frac{d_e(S)}{\eta_u} \leq \tilde{d}_e(S) \leq \eta_o \, d_e(S)$, for all $S \subseteq N$, $e \notin S$.*

The two factors are the sides of one band, and $\tilde{d}_e(S) = 0$ exactly when $d_e(S) = 0$, so $\tilde{f}$ is monotone. Error η means error at most η, since error η implies error η' for every $\eta' \geq \eta$; the error is exactly (η_u, η_o) when both bounds are attained at gains with $d_e(S) > 0$. Only the product matters: rescaling $\tilde{f}$ by $c > 0$ moves (η_u, η_o) to $(\eta_u/c, c\eta_o)$ and changes no comparison between predicted values, hence no run of predictive greedy (Lemma 2). Predictive greedy need not know η. The multiplicative form follows learning-augmented scheduling (Lattanzi et al., 2020; Zhao et al., 2022; Azar et al., 2021). Two facts fix this form: the error must be bounded and on marginal gains rather than values.

Proposition 1 (No guarantee without a bounded error)**.** *For every deterministic algorithm with arbitrary query access to $\tilde{f}$ and every $n \geq 2K$ there is an instance satisfying every assumption of this section except a bound on η on which the output T, $|T| \leq K$, satisfies $f(T) \leq \frac{K}{n-K}$ OPT.*

Proposition 2 (Value accuracy is neither sufficient nor necessary for predictive greedy)**.** *Call $\tilde{f}$ value-accurate at level $\epsilon \in (0,1)$ if $(1-\epsilon)f(S) \leq \tilde{f}(S) \leq (1+\epsilon)f(S)$ for all S. (i) For every ϵ and every $n \geq 2$ there is an instance with $\tilde{f}$ value-accurate at level ϵ and $\eta = \infty$. (ii) For every $M > 0$ and every f not identically zero, $\tilde{f} = (1+M)f$ is value-accurate at no level below M, yet $\eta = 1$. (iii) If $\tilde{f}$ has marginal-gain error η, some positive rescaling of $\tilde{f}$ is value-accurate at level $\frac{\eta-1}{\eta+1}$.*

By (i) no guarantee in η follows from value accuracy; by (iii) the marginal-gain condition is stronger. Definition 1 is the only assumption on the surrogate in what follows. Proofs are in Appendix B.

RELATED WORK

Four lines of work are close to ours. The first optimizes from samples $\{(S_i, f(S_i))\}$ alone: in the PMAC model of Balcan & Harvey (2011), coverage functions are learnable to within $1 + \epsilon$ (Badanidiyuru et al., 2012), yet no algorithm obtains a constant factor for them from polynomially many samples (Balkanski et al., 2017); bounded curvature restores a constant (Balkanski et al., 2016; 2017). The second queries a noisy oracle, $\tilde{f}(S) = \xi_S f(S)$ with ξ_S drawn once per set from a light-tailed distribution: averaging over nearby sets gives a ratio arbitrarily close to $1 - 1/e$ for $K = \Omega(\log \log n)$, and other estimators cover smaller K (Hassidim & Singer, 2017), extended to matroid constraints and non-monotone objectives (Bhawalkar et al., 2025), while adversarial multipliers deny every polynomial-query algorithm a constant factor, at any fixed error (Hassidim & Singer,

2017). The third assumes adversarial error small on every set, $(1-\epsilon)f(S) \le F(S) \le (1+\epsilon)f(S)$: no polynomial-query algorithm obtains a constant factor for $\epsilon \ge n^{-1/2+\beta}$, while greedy keeps one for $\epsilon \le \delta/K$ with a constant δ, a tight threshold (Horel & Singer, 2016). The fourth is the learning-augmented framework we adopt: an algorithm receives a predicted input and its guarantee degrades with a stated error (Purohit et al., 2018; Lykouris & Vassilvitskii, 2021); on submodular maximization, Agarwal & Balkanski (2024) predict the insertion and deletion times of a dynamic stream with the objective queryable, and Antoniadis et al. (2025) take a predicted solution as input for linear objectives. Our regime has no fallback: the algorithm never observes the objective while it runs, and without a bound on the error no deterministic algorithm has a guarantee (Proposition 1).

The first three lines measure error on values, which leaves marginal gains unconstrained (Proposition 2(i)); ours is on marginal gains. The closest prior analysis is greedy with an approximate step: Goundan & Schulz (2007) bound greedy when each chosen element is within a factor of the best, with tightness shown only for exact steps, and Hassidim & Singer (2017) record the constant factor under multiplicative error on marginal gains as well known and note that additive error admits none. Under Definition 1 every step of predictive greedy is such a step, so this guarantee applies. We give its exact worst case, show that no deterministic algorithm with the same query budget improves on it, and identify the error a real surrogate lets one measure.

3 THEORY

This section proves the three prices. Section 3.1 removes the query budget: a modular instance and one exchange inequality give the information price $1/\eta$. Section 3.2 fixes the budget to greedy's own: the classical guarantee, the path-independence lemma the classical analysis misses, and the exact value $\rho_K(\eta)$. Section 3.3 shows that no deterministic algorithm with that budget beats $\rho_K(\eta)$ and bounds every polynomial budget. Section 3.4 shows which error is measurable on a real surrogate.

3.1 THE INFORMATION PRICE IS $1/\eta$

This subsection removes the query budget. A deterministic algorithm may evaluate $\tilde{f}$ on every subset and compute without limit; it never sees f. The best guarantee is $1/\eta$ once $n \ge 2K$, and exhaustive search over predicted values attains it.

Proposition 3 (Upper bound for unbounded queries)**.** *Let $1 \le K < n$ and $\eta_u, \eta_o > 0$ with $\eta_u\eta_o = \eta \ge 1$. (i) For every deterministic algorithm with arbitrary query access to $\tilde{f}$ and output of size at most K, some instance with error exactly (η_u, η_o) makes the output T satisfy $f(T) \le \frac{K}{K+(\eta-1)\min\{K,n-K\}} f(O^*)$, which equals $\frac{1}{\eta} f(O^*)$ for $n \ge 2K$. (ii) Every set S maximizing $\tilde{f}$ over K-subsets satisfies, on every instance, $f(S) \ge \frac{K}{K+(\eta-1)|O^*\setminus S|} f(O^*) \ge \frac{K}{K+(\eta-1)\min\{K,n-K\}} f(O^*)$.*

For $n \ge 2K$ both directions take three lines. Give the algorithm $\tilde{f}(S) = c|S|$; its queries and its output T are then fixed before f is chosen. Take a K-set O disjoint from T and let f be modular with weight $c\eta_u$ on O and c/η_o elsewhere: Definition 1 holds with both factors attained, $f(O) = cK\eta_u$, and $f(T) \le cK/\eta_o = f(O)/\eta$. Conversely, adding the elements of any set one at a time and applying the band to each step gives $f(S)/\eta_u \le \tilde{f}(S) \le \eta_o f(S)$, so a maximizer $\hat{S}$ of $\tilde{f}$ satisfies $f(\hat{S}) \ge \tilde{f}(\hat{S})/\eta_o \ge \tilde{f}(O^*)/\eta_o \ge f(O^*)/\eta$. The attainment direction is also the observation of Horel & Singer (2016) that a maximizer of a function within $1 \pm \epsilon$ of f is a $\frac{1-\epsilon}{1+\epsilon}$-approximation, applied through Proposition 2(iii) with $\epsilon = (\eta-1)/(\eta+1)$. The per-instance bound in (ii), which also covers $K < n < 2K$, rests on one exchange estimate at a predicted maximizer; (i) uses the same modular instance for every $K < n$ (Appendix C). With $K = 3$, $\eta = 3/2$, $n = 6$, $c = 1$ and $(\eta_u, \eta_o) = (3/2, 1)$, the instance gives $f(T) = 3$ against $f(O) = 9/2$: the ratio $2/3 = 1/\eta$.

The bound concerns information, not computation: exhaustive search evaluates $\binom{n}{K}$ predicted sets, polynomial in n only with exponent K, and Section 3.3 bounds what n^c queries reach when c does not grow with K. At $\eta = 1$ the price is 1; as $\eta \to \infty$ it tends to 0, and without a bound on η Proposition 1 rules out every constant. At $K = 1$, predictive greedy attains $1/\eta$, so it is optimal among deterministic algorithms with no restriction on queries. For $K \ge 2$ its worst case sits strictly below $1/\eta$ exactly when $\eta < K$; Section 3.2 gives the exact value.

3.2 PREDICTIVE GREEDY PAYS EXACTLY $\rho_K(\eta)$

The approximation ratio of predictive greedy at marginal-gain error η is $\rho_K(\eta) = \min_{0\le j\le K-1} V_j(\eta)$, a piecewise formula that equals $1/\eta$ once $\eta \ge K$. For $\eta > 1$ it exceeds the classical bound $1 - \big(1 - 1/(\eta K)\big)^K$.

Proposition 4 (Guarantee of predictive greedy). *Let T be the output of predictive greedy on an instance with marginal-gain error η, and write $L_K(x) = 1-\big(1-\frac{1}{xK}\big)^K$ for the classical approximate-greedy bound at error x. Then $f(T) \ge L_K(\eta)\, f(O^*) \ge \big(1 - e^{-1/\eta}\big) f(O^*)$.*

Definition 1 gives, at every step t of predictive greedy with state S^t and choice e_t, $d_{e_t}(S^t) \ge (1/\eta)\max_{e\notin S^t} d_e(S^t)$, the approximate-oracle condition of Goundan & Schulz (2007) with factor η; their Theorem 1 yields $L_K(\eta)$ (Appendix D). That analysis treats each step as an independent worst case, as if a different surrogate answered each step. One surrogate cannot do that: its value on $S \cup \{e, e'\}$ is one number whichever element is added first, so the band on the four gains involved ties the error at S to the error after e is added (Lemma 1). The approximation ratio $\rho_K(\eta) = \alpha(\text{predictive greedy})$ therefore exceeds $L_K(\eta)$ for $\eta > 1$. It involves only the $2K$ elements of $T \cup O^*$, so ρ_K is the same for every $n \ge 2K$ (Appendix D).

Lemma 1 (Path independence). *Let $(f, \tilde f)$ have marginal-gain error η, $S \subseteq N$, and $e \ne e'$ with $e, e' \notin S$ and $\tilde d_e(S) \ge \tilde d_{e'}(S)$. Then (i) $d_e(S\cup\{e'\}) \ge \frac{1}{\eta}\, d_{e'}(S\cup\{e\})$ and (ii) $\big(1-\frac{1}{\eta}\big)\, d_{e'}(S\cup\{e\}) \ge d_{e'}(S) - d_e(S)$.*

Proof. Adding e then e' or e' then e reaches the same $\tilde f(S\cup\{e,e'\})$, so $\tilde d_{e'}(S\cup\{e\}) - \tilde d_e(S\cup\{e'\}) = \tilde d_{e'}(S) - \tilde d_e(S) \le 0$. Applying the lower band to the left term and the upper band to the right term, $d_{e'}(S\cup\{e\}) \le \eta_u \tilde d_{e'}(S\cup\{e\}) \le \eta_u \tilde d_e(S\cup\{e'\}) \le \eta\, d_e(S\cup\{e'\})$, which is (i). The same two orders for f give $d_{e'}(S) - d_e(S) = d_{e'}(S\cup\{e\}) - d_e(S\cup\{e'\})$, and substituting (i) gives (ii). □

Write $d = d_e(S)$ for the gain of the chosen element e, and $g = d_{e'}(S)$, $h = d_{e'}(S\cup\{e\})$ for a competitor's gain before and after the step. Then (ii) and submodularity give $d - g/\eta \ge (1 - 1/\eta)(g - h) \ge 0$. The left side is how far the step is from the worst the band allows; the middle is how much the competitor loses. A step cannot be worst ($d = g/\eta$) and also make the competitor lose ($h < g$). The classical analysis assumes both at every step; the exact value accounts for the trade-off.

For $K \ge 2$ and $\eta \ge 1$, let $k_1 = (K-1)\eta + 1$, $q = (K-1)\eta/k_1$ and $V_j(\eta) = 1 - q^j\big(1 - (K - j)/(K\eta)\big)$ for $0 \le j \le K-1$.

Theorem 1 (Tight approximation ratio of predictive greedy). *For every $K \ge 2$, $\eta \ge 1$ and $n \ge 2K$, under adversarial tie breaking, $\rho_K(\eta) = \min_{0\le j\le K-1} V_j(\eta)$; the minimum is attained by V_j on $\eta \in [K-j, K-j+1]$, with $V_0 = \frac{1}{\eta}$ on $[K, \infty)$, so the breakpoints are the integers $2, \dots, K$. In particular $\rho_K(\eta) = \frac{1}{\eta}$ exactly when $\eta \ge K$; $\rho_2 = \min\big\{\frac{1}{\eta}, \frac{3}{2(\eta+1)}\big\}$; and ρ_3 equals $\frac{16\eta+3}{3(2\eta+1)^2}$, $\frac{7}{3(2\eta+1)}$, $\frac{1}{\eta}$ on $[1,2]$, $[2,3]$, $[3,\infty)$.*

Proof sketch. Normalize $f(O^*) = 1$ and enumerate $O^* = \{o_1, \dots, o_K\}$. A run determines the chosen true gains d_t and the true gains $g_{t,i}$ of o_i at state S^t; the output is $\sum_{t<K} d_t$. Every run satisfies four families of linear inequalities: $\sum_i g_{t,i} \ge 1 - \sum_{s<t} d_s$ (monotonicity and submodularity), $d_t \ge \frac{g_{t,i}}{\eta}$ (selection under the band), $g_{t+1,i} \le g_{t,i}$ (submodularity), and $\big(1 - \frac{1}{\eta}\big) g_{t+1,i} \ge g_{t,i} - d_t$ (Lemma 1(ii)), so the linear program minimizing $\sum_t d_t$ subject to them bounds every run from below. For $\eta \in [K-j, K-j+1]$, Appendix D gives nonnegative weights on the four families whose weighted sum reads $\sum_t d_t \ge V_j$; the weights vanish at the interval endpoints, which is why the minimizing index changes at the integers. For each j the appendix also constructs an instance on $2K$ elements attaining V_j: for j steps, the path-independence inequality is tight and the chosen gains $\frac{q^t}{k_1}$ shrink by q each step, and for the remaining $K - j$ steps the selection inequality is tight with chosen gains $\frac{q^j}{K\eta}$; the chosen gains sum to V_j. □

Corollary 1 (Limit as $K \to \infty$)**.** *Fix $\eta \geq 1$. Both $L_K(\eta)$ and $\rho_K(\eta)$ converge to $1 - e^{-1/\eta}$ from above: L_K decreases strictly in K, and $\rho_K(\eta) = \frac{1}{\eta}$ for $K \leq \eta$ and decreases strictly from $K = \lfloor \eta \rfloor$ on. With $c(\eta) = \frac{e^{-1/\eta}(2\eta - 1)}{2\eta^2}$,*

$$\rho_K(\eta) = 1 - e^{-1/\eta} + \frac{c(\eta)}{K} + O_\eta\Big(\frac{1}{K^2}\Big), \qquad \rho_K(\eta) - L_K(\eta) = \frac{e^{-1/\eta}(\eta - 1)}{\eta^2 K} + O_\eta\Big(\frac{1}{K^2}\Big).$$

The second expansion is what path independence buys: the exact ratio exceeds the classical bound by a $1/K$ term that vanishes at $\eta = 1$. For $\eta > 1$ the attaining surrogates are monotone but not submodular; requiring a submodular surrogate defines a different ratio, equal to $\rho_K(\eta)$ for $\eta \geq K$ and not determined for $1 < \eta < K$ (Appendix D). For $\eta < K$ the exact value lies strictly below the line $1/\eta$ of Section 3.1; Section 3.3 attributes the gap to the query budget.

Example. At $K = 3$ and $\eta = 3/2$ the classical guarantee is $L_3 = 386/729 \approx 0.53$, while $V_0 = 2/3$, $V_1 = 7/12$, $V_2 = 9/16$ give $\rho_3 = 9/16$. The attaining run has chosen gains $1/4$, $3/16$, $1/8$ against competitor gains $1/3$, $1/4$, $3/16$: path independence is tight at every step and selection only at the last. As $K \to \infty$, $\rho_K(3/2) \to 1 - e^{-2/3} \approx 0.49$.

3.3 The gap is the price of the budget

For $\eta < K$ the two values differ. This subsection shows that the difference is the price of the query budget: within greedy's own budget no deterministic algorithm beats $\rho_K(\eta)$, and under any polynomial budget the guarantee is bounded by a curve with the same limit $1 - e^{-1/\eta}$.

Theorem 2 (Optimality of predictive greedy within its query budget)**.** *Let $K \geq 2$, $\eta > 1$ and $n \geq K^4$, and let $\mathcal{A}_{\text{lin}}$ be the class of deterministic algorithms that make at most nK queries to $\tilde{f}$, each on a set of size at most K, and output a set of size at most K; any deterministic tie resolution inside an algorithm is part of the algorithm, and predictive greedy belongs to $\mathcal{A}_{\text{lin}}$. For every $A \in \mathcal{A}_{\text{lin}}$ and every split $\eta_u, \eta_o > 0$ with $\eta_u \eta_o = \eta$, there is an instance on n elements with error exactly (η_u, η_o) on which the output T satisfies $f(T) \leq \rho_K(\eta)\, f(O^*)$. Hence $\alpha^*(\mathcal{A}_{\text{lin}}) = \rho_K(\eta)$ for every $n \geq K^4$: predictive greedy is exactly optimal within its own query budget.*

Proof sketch. The upper bound is a hiding argument. Appendix E.2 builds, on a ground set $B \sqcup O$ with $|O| = K$, an instance whose true objective follows the attaining trajectory of Section 3.2 and whose surrogate depends only on $|S|$ on every set S with $|S| \leq K$ and $|S \cap O| \leq 1$; every K-subset of B has value $\rho_K(\eta) f(O)$, and the error is exactly (η_u, η_o). Run A against that size profile, a legal surrogate in its own right: its at most nK queries $S_1, \ldots, S_M$ and its output T_0 are then fixed. Complete T_0 to a K-set T. Join two elements of $N \setminus T$ by an edge when some S_i contains both; each query contributes at most $\binom{K}{2}$ edges, so a standard bound on the independence number, proved in Appendix G, gives an independent set of size K once $n \geq B_K = K^2(K-1)^2 + 3K - 1$, and $K^4 \geq B_K$. Take O inside it. Then no query meets O twice, every query of A receives the size-profile answer on the instance built around this O, and A outputs $T_0 \subseteq T \subseteq B$, of value at most $\rho_K(\eta) f(O^*)$ by monotonicity. The lower bound is Theorem 1, since $n \geq K^4 \geq 2K$. □

At $K = 3$ and $\eta = 3/2$, no deterministic algorithm with $3n$ queries on sets of at most three elements guarantees more than $9/16$ of the optimum once $n \geq 81$; the appendix threshold is $B_3 = 44$.

Remark 1 (The ground-set threshold)**.** *The bound $n \geq K^4$ is where the hiding argument starts to work, not where optimality starts. For $n \leq K + 2$ (and $n \leq 5$ at $K = 2$) exhaustive search over the $\binom{n}{K}$ predicted sets fits the budget nK and guarantees $1/\eta > \rho_K(\eta)$ for $1 < \eta < K$, so the least n_0 from which $\alpha^*(\mathcal{A}_{\text{lin}}) = \rho_K(\eta)$ satisfies $K + 3 \leq n_0 \leq B_K$, and $6 \leq n_0 \leq 9$ at $K = 2$. Whether n_0 grows like K^4 is open (Appendix E.2).*

Theorem 3 (Bounded-query hardness)**.** *Let $c \geq 0$ and $\tau = \lceil c \rceil + 1$, let $K > \tau$, $n \geq 4K^{c+2}$ and $\eta > 1$ with $\eta \geq \frac{K-1}{K-\tau}$. For every deterministic algorithm that makes at most n^c queries to $\tilde{f}$, each on a set of size at most K, and every split $\eta_u \eta_o = \eta$, there is an instance with error exactly (η_u, η_o) on which the output T satisfies*

$$\frac{f(T)}{f(O^*)} \;\leq\; H_{K,\tau}(\eta) \;:=\; 1 - \Big(1 - \frac{1}{\eta(K-\tau)+1}\Big)^K, \qquad H_{K,\tau}(\eta) \xrightarrow{K\to\infty} 1 - e^{-1/\eta}.$$

The same hiding argument, with the surrogate blind to every set that contains at most τ optimal elements, covers the larger budget; the price is $H_{K,\tau}$ in place of ρ_K (Appendix E.3). At $\tau = 1$ the constant is $U_K(\eta) = 1 - q^K$, the value of the tight family of Section 3.4 at its global error, and for every $K \geq 2$ and $\eta > 1$ the four curves are ordered $L_K(\eta) \leq \rho_K(\eta) < U_K(\eta) = H_{K,1}(\eta)$. At finite parameters $H_{K,\tau}$ can exceed $1/\eta$, so the two upper bounds combine as $\min\{H_{K,\tau}(\eta), 1/\eta\}$; the content of the theorem is the limit, which coincides with that of ρ_K. The cap on query size is part of the query model: with cnK queries of arbitrary size the class optimum still lies between $\rho_K(\eta)$ and $1/\eta$, and as $n \to \infty$ it exceeds $\rho_K(\eta)$ by at most $1/(K(e^{K-1} - K - 1))$, an additive term exponentially small in K (Appendix E.4). The budget, not the algorithm, is what separates $\rho_K(\eta)$ from $1/\eta$; Section 3.4 shows which error certifies a run.

3.4 A GUARANTEE FROM THE ERROR ONE CAN MEASURE

The global error η is defined on every set, including the many a run never visits, so it cannot be read off a run. The selection error is defined on the run alone: at each step, how much better the best available element was than the one taken. Once the objective is evaluated afterwards, that number is computable, and it bounds the value of the run by the classical bound, exactly.

Definition 2 (Selection error and trajectory error)**.** *For a run of predictive greedy with states $S^0, \ldots, S^{K-1}$ and picks $e_0, \ldots, e_{K-1}$, write $M_t = \max_{e \notin S^t} d_e(S^t)$ and $g_t = d_{e_t}(S^t)$, and set*

$$a_t = \begin{cases} \dfrac{M_t}{g_t}, & g_t > 0, \\ 1, & M_t = g_t = 0, \\ \infty, & g_t = 0 < M_t, \end{cases} \qquad \eta_{\mathrm{sel}} = \max\{1, a_0, \ldots, a_{K-1}\},$$

with the convention $L_K(\infty) = 0$. The trajectory error η_{tr} of the run is $\max\{1, \inf \eta'_u \eta'_o\}$ over the pairs (η'_u, η'_o) for which the band of Definition 1 holds at the states $S^0, \ldots, S^{K-1}$.

A step with $g_t = 0 < M_t$ is harmful: the run took a useless element while a useful one existed, and no bound in η_{sel} survives it.

Proposition 5 (Per-run certificate)**.** *For every run of predictive greedy on a surrogate with error η, $\eta_{\mathrm{sel}} \leq \eta_{\mathrm{tr}} \leq \eta$, and the output T satisfies $f(T) \geq L_K(\eta_{\mathrm{sel}})\, f(O^*) \geq L_K(\eta_{\mathrm{tr}})\, f(O^*) \geq L_K(\eta)\, f(O^*)$. For every $K \geq 2$ and $\hat{a} > 1$ there is an instance on $2K$ elements whose adversarial-tie run has $\eta_{\mathrm{sel}} = \eta_{\mathrm{tr}} = \hat{a}$ and output value exactly $L_K(\hat{a})\, f(O^*)$.*

The bound is the induction of Proposition 4, which uses the error only through the ratios a_t; the ordering and the tight family are in Appendix F. The certificate therefore cannot be sharpened for either run-measured error, while the exact value of Theorem 1 needs the global error, because Lemma 1 uses the band off the run; on the attaining run of Theorem 1 the two coincide, $\eta_{\mathrm{sel}} = \eta$.

The certificate needs the true gains at the K states of the run, at most nK values: an audit after the run, not a fallback during it. A harmful zero step gives $\eta_{\mathrm{sel}} = \infty$ and certifies nothing. Outside the model the number is only a diagnostic; Section 4 reports where certificates hold.

3.5 THE THRESHOLD, AND WHAT IS OPEN

Only Theorem 2 needs $n \geq K^4$; Theorem 1 holds for every $n \geq 2K$. The threshold comes from one step of the hiding argument. After the algorithm's transcript is fixed against the size profile, the optimum O must be a K-set that no query contains twice, an independent K-set in the graph whose edges join elements queried together.With nK queries of at most K elements that graph has at most $nK\binom{K}{2}$ edges, and a standard bound on the independence number yields the independent set exactly from $n \geq B_K = K^2(K-1)^2 + 3K - 1$ on, with $K^4 \geq B_K$ (Appendix G). The step cannot be skipped for these instances: their surrogate is not a function of $|S|$ once a set holds two optimal elements. The family of Theorem 3 tolerates τ optimal elements per query, but it attains only $H_{K,\tau}$, not ρ_K. Some threshold is necessary: for $n \leq K + 2$ exhaustive search over the $\binom{n}{K}$ predicted sets fits the budget nK and guarantees $1/\eta > \rho_K(\eta)$. So the least n_0 from which $\alpha^*(\mathcal{A}_{\mathrm{lin}}) = \rho_K(\eta)$ satisfies $K + 3 \leq n_0 \leq B_K$, and $6 \leq n_0 \leq 9$ at $K = 2$; whether n_0 grows like K^4 is open. In practice the theorem covers $K \leq n^{1/4}$, about $K \leq 8$ at $n = 5000$; beyond that range the exact price of predictive greedy stands and only its optimality within the budget is unproved.

4 EXPERIMENTAL STUDY

We evaluate predictive greedy on three applications and two controls, with three aims. The first is to compare the three readings of the bound of Proposition 4, at η_{sel}, η_{tr} and η, which the chain $\eta_{\mathrm{sel}} \leq \eta_{\mathrm{tr}} \leq \eta$ of Section 3.4 orders. The second is to measure how far the certificate falls below the realized ratio on real surrogates. The third is to measure how close real networks come to the worst case of Theorem 1, which the constructed instances attain by design. Das & Kempe (2011) evaluated the submodularity ratio in the same way, comparing the bound computed on the selected sets with the realized value. Figure 2 plots ratios against η, p and K; Table 1 lists the per-instance values; Appendix H gives data, protocol and baselines.

Setup. The surrogates come from a learned model, a partial observation and a proxy objective. In feature selection, $f(S)$ is the held-out accuracy of a decision tree trained on the features S, and $\tilde{f}(S)$ the cross-validation accuracy of the same tree on the training split alone. In the one-hop case of influence maximization (Kempe et al., 2003), $f(S)$ counts the nodes within one hop of S in the full network, and $\tilde{f}(S)$ the same count in an observed network that keeps each edge independently with probability p. In summarization, $f(S)$ is the ROUGE-1 F-score of S against the reference summary, and $\tilde{f}$ one of three submodular objectives of Lin & Bilmes (2011), coverage, diversity or facility location, computed from the article alone. The two controls are the constructed instances of Theorem 1 and Remark 4, run through the same pipeline as the real data, and a stepwise adversary on the real networks, which selects at every step the element of smallest true gain among those with true gain at least $1/\eta$ times the largest. The adversary has the per-step budget of the worst case without its structure: a selection rule with $\eta_{\mathrm{sel}} \leq \eta$, not a surrogate.

Measurement. At every visited state the largest true gain is computed exactly, so η_{sel} is measured. Coverage is monotone submodular, so the network bounds are certificates; held-out accuracy and ROUGE-1 are neither, so their bounds are reference values. Ratios are $f(T)/f(G)$, with T the output of predictive greedy and G that of greedy on f; G is feasible, so each is an upper estimate of $f(T)/f(O^*)$. Ratios therefore never fall below certificates; we report the distance above them.

Certificates on real surrogates. Pooled over the four networks at $K = 30$ and $p = 0.5$ (per network in Table 1), the median ratio is 0.957; the certificate at the median η_{sel} is 0.105 and at the median η_{tr} 0.004. The global error of Definition 1 is infinite on each of the 160 email runs over all p (Appendix H.2); the median certificate of these runs is 0.154. A zero-gain step beside a positive candidate makes the certificate zero; this happens on 46 of the 640 runs and zeroes the government medians at $p = 0.5$ and 0.6. From $p = 0.9$ to 0.2 (panel c) the median ratio falls from at least 0.99 to 0.83–0.93 and the median certificate from 0.45–0.60 to below 0.07, monotonically on every network; with K (panel d) the certificate decays while the network ratio stays above 0.95. The distance is a property of the instances, not slack in the bound: by Remark 4, some run attains $L_K(\eta_{\mathrm{sel}})$ for every value of η_{sel}. At $p = 0.5$ predictive greedy keeps 0.94–0.99 of the value of greedy on f; ranking by observed degree keeps 0.66–0.89.

Objectives outside the model. ROUGE-1 F violates submodularity on 2.1 % and monotonicity on 7.1 % of 70,560 enumerated triples; at $K = 7$ the chosen feature has nonpositive true gain on 52 % of the steps on wine and 47 % on breast cancer, and at $K = 5$ the chosen sentence on 12–19 %. Such steps make η_{sel} infinite, so for these families Table 1 prints the maximum over positive-gain steps and counts the infinite runs. The ratios are 0.94–1.00 for feature selection and 0.63–0.71 for summarization. Against four classical selectors, greedy on the surrogate leads on airline (0.999 against 0.990) and ties within 0.005 elsewhere; against lead-K, only coverage leads at every K.

Where the worst case occurs. On each of the 80 constructed instances the realized ratio equals the predicted value; 47 of them are the crosses in (a) and 20 the stars in (b). The two settings differ most at $\eta = 1$: the instance of Theorem 1 realizes $19/27$ at $K = 3$ through adversarial ties alone, while on every network the adversary, restricted to ties, realizes 1.000. In (b), $L_{30}(\eta)$ certifies the adversary's runs, and their ratio stays 0.27–0.65 above it for every η from 1.25 to 5; a one-step lookahead lowers the ratio by at most 0.08, and random admissible choices raise it by 0.01–0.15. The adversary comes closest at $K = 5$ and $\eta = 3$ on the politician network, 0.385 against $L_5(3) = 0.292$, and the gap widens with K. On real structure the budget alone does not reach the worst case.

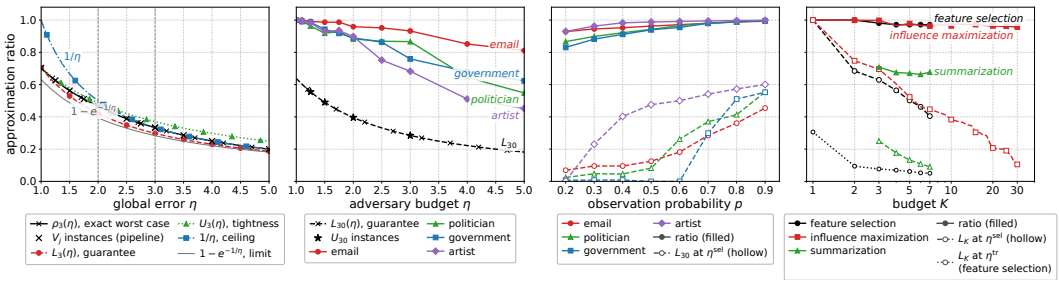

Figure 2: **(a)** The approximation ratio ρ_3 (Theorem 1), the guarantee L_3, the bound U_3 (Remark 19), the ceiling $1/\eta$ and the large-K limit against the global error η; crosses: the constructed instances of Theorem 1. **(b)** The stepwise adversary on four real networks at $K = 30$ against $L_{30}(\eta)$; stars: the instances of Remark 4. **(c)** Observed-network surrogates at $K = 30$ against the observation probability p. **(d)** The three real families against the budget K; for feature selection also L_K at the median η_{tr} (dotted). In (c) and (d), filled markers are median ratios and hollow markers L_K at the median η_{sel}.

Table 1: Predictive greedy on real surrogates, with two references.

	Instance	f	$\tilde{f}$	Ratio med/q25	Error $\eta_{\mathrm{sel}}/\eta_{\mathrm{tr}}$	L_K at $\eta_{\mathrm{sel}}/\eta_{\mathrm{tr}}$	$\infty/d_t{\leq}0$ count/%	Baseline
Reference	Tightness instance, K=3, $\hat{a}$=3/2	✓	×	0.529/–	1.5/1.5	0.529/0.529	0/0	–
	Stepwise adversary, η=3, K=30	✓	–	0.813/0.683	3/–	0.285/–	0/0	–
Networks	email-Eu-core	✓	✓	0.963/0.960	7.5/78.5	0.125/0.013	3/0.5	0.889
	Facebook politician	✓	✓	0.942/0.936	11.7/200	0.082/0.005	0/0.0	0.679
	Facebook government	✓	✓	0.940/0.933	∞/684	0/0.001	10/1.7	0.660
	Facebook artist	✓	✓	0.988/0.987	1.57/1407	0.476/0.001	0/0.0	0.794
Features	wine	×	×	0.944/0.917	1.38/10.1	(0.536/0.095)	25/51.9	0.944
	breast cancer	×	×	0.955/0.940	2/12.2	(0.405/0.079)	30/46.7	0.955
	digits (20 features)	×	×	0.959/0.923	2.9/51.6	(0.297/0.019)	14/8.1	0.963
	airline	×	×	0.999/0.998	1.55/26.5	(0.492/0.037)	0/0.0	0.990
Summaries	coverage	×	✓	0.712/0.635	3.54/21	(0.252/0.047)	177/19.2	0.580
	diversity	×	✓	0.662/0.569	9.9/34.8	(0.097/0.028)	131/12.0	0.580
	facility location	×	✓	0.628/0.535	9.34/35.9	(0.103/0.028)	168/17.5	0.580

✓: f monotone submodular (L_K certifies); $\tilde{f}$ submodular. Ratio $f(T)/f$(greedy on f), upper estimate, run medians. ∞: zero-gain step beside a positive candidate. Parentheses: reference, f outside the model. Baseline: best selector, degree ranking or lead-K, same denominator. Networks: $p = 0.5$, $K = 30$; Features: $K = 7$; Summaries: $K = 5$. Summaries: η_{tr} after scale matching (Appendix H.1).

5 Conclusion

We revisit the submodular optimization problem for unknown objectives and measure a surrogate's error on marginal gains, where greedy reads it and priced that error exactly for deterministic algorithms: an information price that no query budget removes, a price predictive greedy pays, and a price of the budget that no algorithm within greedy's budget avoids. The classical bound evaluated at the selection error of a run is a certificate, tight at every cardinality. The pessimism of prior work is thus a property of where the error was measured, not of surrogates: a bound on values leaves the marginal gains free; a bound on the marginal gains leaves greedy a constant fraction of the optimum, and no deterministic algorithm with the same budget does better.

Limitations and open questions. The results concern deterministic algorithms and worst cases over all surrogates with a given error, so on learned surrogates, whose errors are not arranged against the algorithm, observed ratios sit well above the prices. The exact price depends on the global error, which no run reveals; a run reveals its selection error, which certifies only the classical bound. The certificate presumes a monotone submodular objective. Three questions remain open: the ground-set size from which optimality within the budget holds, the value for randomized algorithms and for queries of arbitrary size, and the exact price under a submodular surrogate.

AI USE STATEMENT

The problem, the model, the error definition, the statements of the main results and major math proofs are the authors' own. Generative AI tools assisted in formulating some mathematical proofs, in providing ingredients for some proofs, in feedback on the experimental methodology, and in co-writing experimental code scripts. We have not used generative AI tools to generate major math proofs and synthetic data sets, to clean or reformat datasets, or for translation. Additionally, we used generative AI tools to polish some parts of the paper and to cross-check references. We have reviewed all AI-assisted work, and all AI-assisted code was run and edited by the authors. We take responsibility for the final content of this work.

REPRODUCIBLE STATEMENT

All experiments in Section 4 and the experimental appendix can be reproduced from the anonymous repository at `https://anonymous.4open.science/r/sub-modular-optimization-experiment-4051`. It contains the implementation of predictive greedy, the baselines, the stepwise adversary and the three error measures.

REFERENCES

Arpit Agarwal and Eric Balkanski. Learning-augmented dynamic submodular maximization. *Advances in Neural Information Processing Systems*, 37:14148–14176, 2024.

Antonios Antoniadis, Marek Eliás, Adam Polak, and Moritz Venzin. Approximation algorithms for combinatorial optimization with predictions. In *International Conference on Learning Representations*, volume 2025, pp. 49039–49052, 2025.

Yossi Azar, Stefano Leonardi, and Noam Touitou. Flow time scheduling with uncertain processing time. In *proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1070–1080, 2021.

Ashwinkumar Badanidiyuru, Shahar Dobzinski, Hu Fu, Robert Kleinberg, Noam Nisan, and Tim Roughgarden. Sketching valuation functions. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pp. 1025–1035. SIAM, 2012.

Maria-Florina Balcan and Nicholas JA Harvey. Learning submodular functions. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pp. 793–802, 2011.

Eric Balkanski, Aviad Rubinstein, and Yaron Singer. The power of optimization from samples. *Advances in Neural Information Processing Systems*, 29, 2016.

Eric Balkanski, Aviad Rubinstein, and Yaron Singer. The limitations of optimization from samples. In *Proceedings of the 49th annual acm sigact symposium on theory of computing*, pp. 1016–1027, 2017.

Kshipra Bhawalkar, Yang Cai, Zhe Feng, Christopher Liaw, and Tao Lin. A unified approach to submodular maximization under noise. In *Advances in Neural Information Processing Systems 38 (NeurIPS 2025)*, pp. 166355–166383, 2025.

Abhimanyu Das and David Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. *arXiv preprint arXiv:1102.3975*, 2011.

Pranava R Goundan and Andreas S Schulz. Revisiting the greedy approach to submodular set function maximization. *Optimization online*, 1, 2007.

Avinatan Hassidim and Yaron Singer. Submodular optimization under noise. In *Conference on Learning Theory*, pp. 1069–1122. PMLR, 2017.

Thibaut Horel and Yaron Singer. Maximization of approximately submodular functions. *Advances in neural information processing systems*, 29, 2016.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146, 2003.

Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9(2), 2008.

Silvio Lattanzi, Thomas Lavastida, Benjamin Moseley, and Sergei Vassilvitskii. Online scheduling via learned weights. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1859–1877. SIAM, 2020.

Hui Lin and Jeff Bilmes. A class of submodular functions for document summarization. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pp. 510–520, 2011.

Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. *Journal of the ACM (JACM)*, 68(4):1–25, 2021.

George L Nemhauser and Laurence A Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of operations research*, 3(3):177–188, 1978.

George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming*, 14(1):265–294, 1978.

Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ml predictions. *Advances in Neural Information Processing Systems*, 31, 2018.

Kai Wei, Rishabh Iyer, and Jeff Bilmes. Submodularity in data subset selection and active learning. In *International conference on machine learning*, pp. 1954–1963. PMLR, 2015.

Tianming Zhao, Wei Li, and Albert Y Zomaya. Real-time scheduling with predictions. In *2022 IEEE Real-Time Systems Symposium (RTSS)*, pp. 331–343. IEEE, 2022.

A Appendix

A.1 Notation

Table 2: Summary of notation.

Symbol	Meaning
Problem	
N, n	Ground set and its size, $n = \lvert N \rvert$
K	Cardinality budget; the number of elements selected
f	True objective: normalized, monotone, submodular; not evaluated by the algorithm
$d_e(S)$	Marginal gain of e at S, $f(S \cup \{e\}) - f(S)$
OPT	$\max_{\lvert S \rvert \le K} f(S)$
O^*	An optimal set, $\lvert O^* \rvert \le K$, $f(O^*) = \mathrm{OPT}$
$A^{\tilde{f}}$	Output of algorithm A on the instance $(f, \tilde{f})$
$\alpha(A)$, $\alpha^*(\mathcal{A})$	Worst-case ratio of A at error η, and its supremum over a class $\mathcal{A}$
Surrogate and error	
$\tilde{f}$	Surrogate for f; evaluable on any set, $\tilde{f}(\emptyset) = 0$
$\tilde{d}_e(S)$	Predicted marginal gain, $\tilde{f}(S \cup \{e\}) - \tilde{f}(S)$
η_u, η_o	Underestimation and overestimation factors (Definition 1)
η	Marginal-gain error, $\eta = \eta_u \eta_o \ge 1$; $\eta = 1$: every marginal gain correct up to a common scale
η^{tr}	Trajectory error: Definition 1 restricted to the states greedy visits
η^{sel}	Selection error: $\max\{1, \max_t M_t/g_t\}$ (∞ if $g_t = 0 < M_t$), measured after a run
ϵ	Value error, $\max_S \lvert \tilde{f}(S) - f(S) \rvert / f(S)$, used for comparison only
Algorithm and queries	
S^t, e_t	Set after t steps of predictive greedy, and the element it adds at step t
M_t, g_t	Largest true gain available at S^t, and the true gain of e_t
$\mathcal{A}_{\mathrm{lin}}$	Deterministic algorithms using at most nK queries, each on a set of size at most K, with output of size at m
B_K	$K^2(K-1)^2 + 3K - 1$; ground-set threshold of Theorem 2
Guarantees	
$L_K(x)$	Classical bound $1 - \left(1 - \frac{1}{xK}\right)^K$ (Proposition 4)
k_1, q	$k_1 = (K-1)\eta + 1$, $q = (K-1)\eta/k_1$
V_j	$1 - q^j\left(1 - \frac{K-j}{K\eta}\right)$, $0 \le j \le K-1$
$\rho_K(\eta)$	Exact worst-case ratio of predictive greedy, $\min_{0 \le j \le K-1} V_j$ (Theorem 1)
$U_K(\eta)$	$1 - q^K$; value of the tight family of Proposition 4 at its global error
$H_{K,\tau}(\eta)$	Polynomial-query bound $1 - \left(1 - \frac{1}{\eta(K-\tau)+1}\right)^K$ (Theorem 3)

B Proofs for Section 2

This appendix proves the scaling lemma behind the remark that only the product η matters, and Propositions 1 and 2. Throughout, O^* is an optimal set.

B.1 Scaling

Lemma 2 (Scaling). *Fix a monotone submodular f and write $\mathcal{F}(\eta_u, \eta_o)$ for the set of surrogates that satisfy Definition 1 with the factors (η_u, η_o). For every $c > 0$ the map $\tilde{f} \mapsto c\tilde{f}$ is a bijection from $\mathcal{F}(\eta_u, \eta_o)$ onto $\mathcal{F}(\eta_u/c,\, c\eta_o)$, and the two classes have the same product η. The map leaves f untouched and preserves every comparison between predicted values, so it leaves each run of a comparison-based algorithm unchanged, predictive greedy in particular. Consequently the worst-case ratio of such an algorithm depends on (η_u, η_o) only through η.*

Proof. Fix $c > 0$ and $\tilde{f} \in \mathcal{F}(\eta_u, \eta_o)$. Multiplying the band of Definition 1 by c gives, for every S and every $e \notin S$,

$$\frac{d_e(S)}{\eta_u/c} \;\le\; c\,\tilde{d}_e(S) \;\le\; (c\eta_o)\,d_e(S),$$

and $c\tilde{f}(\emptyset) = 0$. The two new factors are positive and their product is $(\eta_u/c)(c\eta_o) = \eta \ge 1$, so $c\tilde{f} \in \mathcal{F}(\eta_u/c,\, c\eta_o)$. The same computation with $1/c$ in place of c maps $\mathcal{F}(\eta_u/c,\, c\eta_o)$ into $\mathcal{F}(\eta_u, \eta_o)$ and inverts the map, so the map is a bijection onto.

Predicted values are multiplied by the positive number c, so at every state the order of the predicted gains is unchanged, and with it the set of maximizers. An algorithm whose queries and choices depend on the surrogate only through comparisons of predicted values therefore produces the same run on $\tilde{f}$ and on $c\tilde{f}$ under the same tie-breaking rule, and returns the same set.

Finally, let (η_u, η_o) and (η'_u, η'_o) be positive pairs with the same product η. With $c = \eta_u/\eta'_u > 0$ one has $\eta_u/c = \eta'_u$ and $c\eta_o = \eta/\eta'_u = \eta'_o$, so the map sends $\mathcal{F}(\eta_u, \eta_o)$ onto $\mathcal{F}(\eta'_u, \eta'_o)$. Since f and the output set are both unchanged, the ratio $f(T)/f(O^*)$ is unchanged, and the two worst-case ratios coincide. □

Remark 2 (Prescribed splits)**.** *The hardness constructions of Sections 3.3 and* **??** *prescribe a split (η_u, η_o). By Lemma 2 they extend to every positive split with the same product: rescaling the surrogate moves the split without moving f, the run of a comparison-based algorithm, or the ratio. An algorithm that reads predicted values rather than comparisons of them is not covered by the run-invariance clause of the lemma; for such an algorithm the split is handled by rescaling the construction itself, which is how the theorems are stated.*

B.2 Proof of Proposition 1

Proof. Fix $K \ge 1$ and $n \ge 2K$, and let A be a deterministic algorithm with arbitrary query access to $\tilde{f}$. Put

$$\gamma = \frac{K^2}{n(n-K)} \in (0, 1], \qquad \tilde{f}(S) = |S|,$$

where $\gamma \le 1$ because $n \ge 2K$ gives $n(n-K) \ge 2K \cdot K > K^2$. For a K-set $O \subseteq N$ define

$$f_O(S) = |S \cap O| + \gamma\,|S \setminus O|.$$

The pair is admissible and has finite error. The function f_O is modular with coefficients in $\{1, \gamma\} \subseteq (0, 1]$, so it is nonnegative, monotone and submodular with $f_O(\emptyset) = 0$, and $\tilde{f}(\emptyset) = 0$. Every predicted single-element gain is $\tilde{d}_e(S) = 1$ and every true one is $d_e(S) \in \{1, \gamma\}$, so the largest ratio $d_e(S)/\tilde{d}_e(S)$ is 1 and the largest ratio $\tilde{d}_e(S)/d_e(S)$ is $1/\gamma$. The band of Definition 1 therefore holds with $(\eta_u, \eta_o) = (1, 1/\gamma)$ and with no pair of smaller product, so the error of the pair is

$$\eta = \frac{1}{\gamma} = \frac{n(n-K)}{K^2},$$

finite for every n and unbounded as $n \to \infty$. Also $\max_{|S| \le K} f_O(S) = f_O(O) = K$, because $\gamma \le 1$ makes the K largest modular coefficients exactly those of O; so $O^* = O$ may be taken and $f_O(O^*) = K$.

The output is fixed before the instance. The surrogate $\tilde{f}$ does not depend on O, so every answer A receives is determined before O is chosen. The whole query transcript, and therefore the output T with $|T| \le K$, is the same for every O. Since $n \ge 2K$ and $|T| \le K$, a K-set $O \subseteq N \setminus T$ exists. For that O the sets T and O are disjoint, so

$$\frac{f_O(T)}{f_O(O^*)} = \frac{\gamma|T|}{K} \;\le\; \gamma = \frac{K^2}{n(n-K)} \;\le\; \frac{K}{n-K}.$$

Letting $n \to \infty$ with K fixed drives the right-hand side to 0, so no constant worst-case ratio survives when η is unconstrained. □

Remark 3 (Randomized algorithms)**.** *The same construction bounds randomized algorithms in expectation. Let A be randomized, keep $\gamma = K^2/(n(n-K))$ and $\tilde{f}(S) = |S|$, and let O be a uniformly random K-subset of N, independent of the random string of A. The transcript still does not depend on O, so the output T is independent of O, and the hypergeometric mean gives $\mathbb{E}|T \cap O| = \mathbb{E}|T| \cdot K/n \le K^2/n$. Bounding $f_O(T) \le |T \cap O| + \gamma K$ and dividing by $f_O(O^*) = K$,*

$$\mathbb{E}\Big[\frac{f_O(T)}{f_O(O^*)}\Big] \;\le\; \frac{K}{n} + \gamma = \frac{K(n-K)+K^2}{n(n-K)} = \frac{K}{n-K}.$$

Since $\min_O \le \mathbb{E}_O$, a single K-set O attains the bound in expectation over the random string. The deterministic bound of Proposition 1 is stronger by the additive term K/n, which is the price of averaging.

Remark 4 (The rate in η)**.** *The choice $\gamma = K^2/(n(n-K))$ only matches the constant in the statement. The same construction with any $\gamma \in (0,1]$ gives $f(T) \le \gamma f(O^*)$ for a deterministic algorithm, with $\eta = 1/\gamma$. So without an upper bound on η no algorithm has any constant guarantee, and the guarantee of every algorithm degrades at least as fast as $1/\eta$. Proposition 3 shows that this rate is exact.*

B.3 Proof of Proposition 2

Proof. Part (i). Fix $\epsilon \in (0,1)$, let $N \supseteq \{a,b\}$ and define

$$f(\emptyset) = 0, \quad f(\{a\}) = f(\{b\}) = 1, \quad f(\{a,b\}) = 1+\epsilon, \qquad \tilde{f}(\emptyset) = 0, \quad \tilde{f}(\{a\}) = \tilde{f}(\{b\}) = \tilde{f}(\{a,b\}) = 1,$$

both extended to N by $f(S) = f(S \cap \{a,b\})$ and $\tilde{f}(S) = \tilde{f}(S \cap \{a,b\})$, which adds $n-2$ elements of identically zero gain under both functions and reaches any prescribed $n \ge 2$. The function f is normalized and monotone, because $\epsilon > 0$, and it is submodular, because $d_a(\{b\}) = d_b(\{a\}) = \epsilon \le 1 = d_a(\emptyset) = d_b(\emptyset)$ and the added elements have zero gain everywhere. The surrogate is value-accurate at level ϵ: on the singletons $\tilde{f} = f$, and on $\{a,b\}$ one has $\tilde{f}/f = 1/(1+\epsilon) \ge 1-\epsilon$ because $(1-\epsilon)(1+\epsilon) = 1-\epsilon^2 \le 1$, while $\tilde{f}/f \le 1 \le 1+\epsilon$. Yet

$$\tilde{d}_b(\{a\}) = \tilde{f}(\{a,b\}) - \tilde{f}(\{a\}) = 0 \qquad \text{while} \qquad d_b(\{a\}) = \epsilon > 0,$$

so no finite η_u satisfies $d_b(\{a\})/\eta_u \le \tilde{d}_b(\{a\})$, that is, $\eta = \infty$.

Part (ii). Let f be monotone submodular and not identically zero, and let $\tilde{f} = (1+M)f$ with $M > 0$. Every set S with $f(S) > 0$ witnesses $|\tilde{f}(S) - f(S)|/f(S) = M > \epsilon$ for every $\epsilon < M$, so $\tilde{f}$ is value-accurate at no level below M. Its marginal gains are $\tilde{d}_e(S) = (1+M)\, d_e(S)$ for every S and every $e \notin S$, so Definition 1 holds with $\eta_u = 1/(1+M)$ and $\eta_o = 1+M$, and $\eta = \eta_u \eta_o = 1$. Since $1 + M > 0$, the maximizers of $\tilde{d}_e(S)$ and of $d_e(S)$ coincide at every state, so whichever of them tie breaking picks, the picked element has maximum true gain: predictive greedy on $\tilde{f}$ is greedy on f, and its selection error (Definition 2) is $\eta^{\mathrm{sel}} = 1$.

Part (iii). Fix a nonempty $S = \{e_1, \dots, e_m\}$ and write $S_i = \{e_1, \dots, e_i\}$ with $S_0 = \emptyset$. Since $f(\emptyset) = \tilde{f}(\emptyset) = 0$,

$$f(S) = \sum_{i=1}^{m} d_{e_i}(S_{i-1}), \qquad \tilde{f}(S) = \sum_{i=1}^{m} \tilde{d}_{e_i}(S_{i-1}).$$

Each term satisfies $d_{e_i}(S_{i-1})/\eta_u \le \tilde{d}_{e_i}(S_{i-1}) \le \eta_o\, d_{e_i}(S_{i-1})$ by Definition 1, and all terms are nonnegative because f is monotone. Summing gives

$$\frac{f(S)}{\eta_u} \;\le\; \tilde{f}(S) \;\le\; \eta_o\, f(S) \qquad \text{for all } S \subseteq N, \tag{1}$$

the empty set being covered by $f(\emptyset) = \tilde{f}(\emptyset) = 0$. Now set

$$\epsilon = \frac{\eta - 1}{\eta + 1} \in [0,1), \qquad c = \frac{2\eta_u}{\eta + 1} > 0.$$

Then $c/\eta_u = 2/(\eta+1) = 1-\epsilon$ and $c\eta_o = 2\eta/(\eta+1) = 1+\epsilon$, so multiplying equation 1 by c yields $(1-\epsilon)f(S) \le c\tilde{f}(S) \le (1+\epsilon)f(S)$ for all $S \subseteq N$. The rescaled surrogate $c\tilde{f}$ is therefore value-accurate at level ϵ. For $\eta = 1$ this reads $c\tilde{f} = f$; for $\eta > 1$ the level lies in $(0,1)$. □

Remark 5 (Part (iii) and the $1/\eta$ ceiling)**.** *Part (iii) uses only that f is monotone and that both functions vanish on the empty set; it does not use submodularity. Combined with the observation of Horel & Singer (2016) that an α-approximation for a function within $1 \pm \epsilon$ of f is an $\alpha\frac{1-\epsilon}{1+\epsilon}$-approximation for f, it gives a second proof of the attainment direction of Proposition 3: with $\epsilon = (\eta-1)/(\eta+1)$ one has $\frac{1-\epsilon}{1+\epsilon} = 1/\eta$, and rescaling does not change the maximizer of $\tilde{f}$.*

C Proofs for Section 3.1: the information price

This appendix proves Proposition 3 for all $1 \le K < n$, records the case $K = 1$, and treats randomized algorithms, which are outside the scope of the main text. Throughout, $(f, \tilde{f})$ is an instance in the sense of Section 2, $\eta_u, \eta_o > 0$ with $\eta_u\eta_o = \eta \ge 1$, and O^* is an optimal K-set. The value band equation 1, $f(S)/\eta_u \le \tilde{f}(S) \le \eta_o f(S)$ for every $S \subseteq N$, is used repeatedly; it needs only that f is monotone and that both functions vanish on the empty set.

C.1 The exchange bound

Lemma 3 (Exchange bound at a predicted maximizer)**.** *Let $1 \le K \le n$, let S be a K-set maximizing $\tilde{f}$ among all K-subsets of N, let O be any K-subset of N, and put $a = |O \setminus S|$, $A = f(S)$, $Z = f(S \cup O)$ and $D = \sum_{b\in O\setminus S} d_b(S)$. Writing $r_s = f(S) - f(S \setminus \{s\})$ for $s \in S$ and $R = \sum_{s\in S} r_s$,*

$$0 \;\le\; R \;\le\; A, \qquad K\,D \;\le\; \eta\,a\,R \;\le\; \eta\,a\,A, \qquad Z \;\le\; A + D \;\le\; \Big(1 + \frac{\eta a}{K}\Big)A.$$

Proof. Each $r_s = d_s(S \setminus \{s\})$ is nonnegative by monotonicity. Enumerate $S = \{s_1, \dots, s_K\}$ and put $S_i = \{s_1, \dots, s_i\}$. Telescoping, $A = \sum_{i=1}^{K} d_{s_i}(S_{i-1})$, and since $S_{i-1} \subseteq S \setminus \{s_i\}$, submodularity gives $d_{s_i}(S_{i-1}) \ge r_{s_i}$; summing, $A \ge R \ge 0$.

Fix $b \in O \setminus S$ and $s \in S$. The set $(S \setminus \{s\}) \cup \{b\}$ has exactly K elements, so predicted optimality of S gives $\tilde{d}_b(S \setminus \{s\}) \le \tilde{d}_s(S \setminus \{s\})$. With submodularity of f and the two sides of Definition 1,

$$d_b(S) \;\le\; d_b(S \setminus \{s\}) \;\le\; \eta_u\,\tilde{d}_b(S \setminus \{s\}) \;\le\; \eta_u\,\tilde{d}_s(S \setminus \{s\}) \;\le\; \eta_u\eta_o\,d_s(S \setminus \{s\}) \;=\; \eta\,r_s. \quad (2)$$

Equivalently, $\eta r_s - d_b(S)$ is the sum of four nonnegative quantities, contributed by the upper band, by predicted exchange optimality, by the lower band and by submodularity. The deletions range over all K elements of S, those in $S \cap O$ included. Summing equation 2 over all aK pairs (b, s) gives $KD \le \eta aR$, and $R \le A$ gives $KD \le \eta aA$.

Finally, enumerate $O \setminus S = \{b_1, \dots, b_a\}$ and telescope f along the chain from S to $S \cup O$: each increment $d_{b_j}(S \cup \{b_1, \dots, b_{j-1}\})$ is at most $d_{b_j}(S)$ by submodularity, so $Z - A \le D$. Together with $D \le \eta aA/K$ this is the last display. □

C.2 Proof of Proposition 3

Part (ii), the per-instance guarantee. Let S maximize $\tilde{f}$ over sets of size exactly K; the surrogate is monotone (Section 2), so a smaller predicted maximizer can be completed to size K without loss. Let O be an optimal K-set and put $a = |O \setminus S|$, $A = f(S)$, $B = f(O)$, $Z = f(S \cup O)$. Put $X = \tilde{f}(S \cup O) - \tilde{f}(S)$ and $Y = \tilde{f}(S \cup O) - \tilde{f}(O)$. Predicted optimality gives $\tilde{f}(S) \ge \tilde{f}(O)$, that is $X \le Y$. Telescoping the lower side of the band along a chain from S to $S \cup O$ gives $(Z - A)/\eta_u \le X$; telescoping the upper side along a chain from O to $S \cup O$ gives $Y \le \eta_o(Z - B)$. Together,

$$\frac{Z - A}{\eta_u} \;\le\; X \;\le\; Y \;\le\; \eta_o(Z - B), \qquad \text{hence} \qquad B \;\le\; \Big(1 - \frac{1}{\eta}\Big)Z + \frac{A}{\eta}.$$

The algorithm need not have queried $S \cup O$, and $\tilde{f}$ need not be submodular; only the band is used. The coefficient $1 - 1/\eta$ is nonnegative, so Lemma 3 may be substituted:

$$B \;\le\; \Big(1 - \frac{1}{\eta}\Big)\Big(1 + \frac{\eta a}{K}\Big)A + \frac{A}{\eta} \;=\; \Big(1 + \frac{(\eta - 1)a}{K}\Big)A,$$

which is the per-instance bound. Since $a \le K$ and $O \setminus S \subseteq N \setminus S$ has at most $n-K$ elements, $a \le \min\{K, n-K\}$, which gives the second inequality of (ii). With $\gamma = 1 - 1/\eta$, $E = \gamma Z + A/\eta - B$, $H = A + D - Z$, $J = \eta aR - KD$ and $W = A - R$, the whole argument is one nonnegative combination,

$$\Big(1 + \frac{(\eta-1)a}{K}\Big)A - B \;=\; E + \gamma H + \frac{\gamma}{K}J + \frac{\gamma\eta a}{K}W, \tag{3}$$

an identity in $A, B, Z, D, R, a, K, \eta$ as expanding the right side shows whose four terms are nonnegative by the display above and by Lemma 3.

Part (i), the matching instance. Feed the algorithm $\tilde{f}(S) = c|S|$ for a constant $c > 0$; its transcript and its output are fixed before f is chosen. Complete the output to a K-set S if necessary, and let f be modular with weight c/η_o on each element of S and $c\eta_u$ on each element outside S; the outside weight is the larger one because $\eta \ge 1$. Every set increment is a sum of weights, so the error equals (η_u, η_o) exactly (this uses $K < n$, so that both weights occur). With $a_0 = \min\{K, n-K\}$, an optimal K-set takes a_0 heavy and $K - a_0$ light elements, so

$$\frac{f(S)}{f(O^*)} \;=\; \frac{Kc/\eta_o}{(K-a_0)\,c/\eta_o + a_0\,c\eta_u} \;=\; \frac{K}{K + (\eta-1)a_0},$$

and an output smaller than K is only worse. For $n \ge 2K$ the ratio is $1/\eta$, the value of the main text. □

At $n = K$, excluded from the proposition, the whole ground set is the only K-set, every algorithm that outputs N is optimal, and the formula gives 1. With $K = 3$, $\eta = 3/2$ and $n = 3, 4, 5, 6$, the value $K/(K + (\eta-1)\min\{K, n-K\})$ is 1, 6/7, 3/4, 2/3; the per-instance bound at $|O^* \setminus S| = 0, 1, 2, 3$ takes the same four values.

Remark 6 (The case $K = 1$)**.** *At $K = 1$ predictive greedy returns a singleton $\{e\}$ with $\tilde{f}(\{e\}) \ge \tilde{f}(\{o\})$ for every $o \in N$, so for the optimal singleton $O^* = \{o\}$, by equation 1,*

$$f(\{e\}) \;\ge\; \frac{\tilde{f}(\{e\})}{\eta_o} \;\ge\; \frac{\tilde{f}(\{o\})}{\eta_o} \;\ge\; \frac{f(\{o\})}{\eta_u\eta_o} \;=\; \frac{f(O^*)}{\eta}.$$

The matching instance of part (i) applies verbatim at $K = 1$ and any $n \ge 2$, where the formula gives $1/\eta$, so at $K = 1$ predictive greedy is optimal among deterministic algorithms with no restriction on the query budget.

C.3 Randomized algorithms

Randomized algorithms are outside the scope of the main text. The modular instance bounds them as follows, and the bound is not the randomized value.

Proposition 6 (Randomized upper bound)**.** *Let $K \ge 1$ and $n \ge 2K$. Fix a randomized algorithm with arbitrary query access to $\tilde{f}$ and output of size at most K. Then some instance with error exactly (η_u, η_o) makes the output T satisfy*

$$\mathbb{E}\big[f(T)\big] \;\le\; \Big(\big(1 - \tfrac{K}{n}\big)\frac{1}{\eta} + \frac{K}{n}\Big) f(O^*),$$

the expectation taken over the algorithm's randomness.

Proof. Fix $c > 0$. For a K-set O let f_O be modular with weight $c\eta_u$ on O and c/η_o elsewhere, and let $\tilde{f}(S) = c|S|$; the pair has error exactly (η_u, η_o) and $f_O(O) = cK\eta_u$ is its optimum. Let O be uniform among the K-subsets of N, independent of the algorithm's random string. Fix a value of the string. The answers reveal nothing about O, so the output T is a fixed set independent of O. Write $m = |T| \le K$. Each element of T lies in O with probability K/n, so $\mathbb{E}_O|T \cap O| = mK/n$ and $\mathbb{E}_O|T \setminus O| = m(1 - K/n)$. Hence

$$\mathbb{E}_O\left[\frac{f_O(T)}{f_O(O)}\right] \;=\; \frac{m(1-\frac{K}{n})\frac{c}{\eta_o} + \frac{mK}{n}c\eta_u}{cK\eta_u} \;=\; \frac{m}{K}\Big(\big(1-\tfrac{K}{n}\big)\frac{1}{\eta} + \frac{K}{n}\Big) \;\le\; \big(1-\tfrac{K}{n}\big)\frac{1}{\eta} + \frac{K}{n}.$$

The bound holds for every value of the string. Averaging over the string and exchanging with the finite average over O, some O attains at most the average, and $(f_O, \tilde{f})$ for that O is the required instance. □

Lemma 4 (Uniformly random subsets)**.** *Let f be monotone submodular with $f(\emptyset) = 0$ and let R be a uniformly random K-subset of N. Then $\mathbb{E}[f(R)] \geq \frac{K}{n} f(N)$.*

Proof. Let π be a uniformly random ordering of N, let P_i be the set of its first i elements, and write $\mu_i = \mathbb{E}[f(P_i) - f(P_{i-1})]$. Then $\sum_{i=1}^{n} \mu_i = f(N)$ and $\mathbb{E}[f(R)] = \sum_{i=1}^{K} \mu_i$, since P_K is a uniformly random K-subset. For $i \leq j$, $\mu_i \geq \mu_j$: conditioned on the element e in position i or j, the prefix before it is a uniformly random subset of $N \setminus \{e\}$ of size $i-1$ or $j-1$, a uniformly random $(i-1)$-subset can be sampled inside a uniformly random $(j-1)$-subset, and d_e is nonincreasing in its argument. A nonincreasing sequence with sum $f(N)$ has its first K terms summing to at least $\frac{K}{n} f(N)$. □

The deterministic value and the randomized value do not coincide in general. At $n = 3$, $K = 2$, $\eta = 3$, Proposition 3 gives the deterministic value $1/2$, while outputting a uniformly random 2-subset guarantees expected value at least $\frac{2}{3} f(N) \geq \frac{2}{3} f(O^*)$ on every instance by Lemma 4, without any query. No general formula for the randomized value is known; at $\eta = 1$ both values equal 1.

C.4 Scope

The attainment quantifier in Proposition 3 is that one algorithm, exhaustive search over predicted values, guarantees the stated ratio on every instance. Lemma 3 uses the aK single-exchange comparisons, which a set that is locally optimal for $\tilde{f}$ under single swaps also satisfies; part (ii) also uses $\tilde{f}(S) \geq \tilde{f}(O)$, which such a set need not satisfy, so the guarantee is stated for exhaustive search only. Exhaustive search evaluates $\binom{n}{K}$ predicted sets, polynomial in n only with exponent K; nothing here yields an algorithm with n^c queries for a c independent of K, which is the regime of Theorem 3.

D Proofs for Section 3.2: the exact approximation ratio of predictive greedy

D.1 Conventions and a counting lemma

Four conventions are in force throughout the appendix. First, f is monotone submodular with $f(\emptyset) = 0$ and $\tilde{f}(\emptyset) = 0$, and submodularity is used in its diminishing-returns form: $d_e(T) \leq d_e(S)$ for $S \subseteq T$ and $e \notin T$. Second, an optimal set is written O^*, $|O^*| \leq K$, and instances are normalized so that $f(O^*) = 1$ whenever a ratio is computed. Third, the tie-breaking rule of a greedy run is part of the instance: "adversarial ties" means that when several elements maximize the predicted gain, the one named by the construction is taken. Fourth, with $k_1 = (K-1)\eta + 1$ and $q = (K-1)\eta/k_1$ as in Section 3.2, $U_K(\eta) = 1 - q^K$ denotes the value of the family of Appendix **??** read at its global error; it is used for comparison only.

Several constructions below are counting functions, so the following observation is recorded once.

Lemma 5 (Counting functions)**.** *Let $N = N_1 \cup \cdots \cup N_r$ be a partition and let $h(S) = H(|S \cap N_1|, \ldots, |S \cap N_r|)$ for a function H on the box $\prod_i \{0, 1, \ldots, |N_i|\}$. Write $\Delta_i H(v) = H(v + u_i) - H(v)$ for the forward difference in coordinate i (u_i the ith unit vector), defined when $v_i < |N_i|$. If $\Delta_i H \geq 0$ for every i and every v, and $\Delta_\ell \Delta_i H \leq 0$ for every pair (i, ℓ), including $i = \ell$, and every v at which both differences are defined, then h is monotone and submodular.*

Proof. Let $v(S)$ denote the count vector of S. For $e \in N_i$ with $e \notin S$ one has $d_e(S) = \Delta_i H(v(S))$, which is nonnegative by hypothesis, so h is monotone. For $S \subseteq T$ and $e \in N_i \setminus T$, the vectors satisfy $v(S) \leq v(T)$ coordinatewise and $v(T)_i < |N_i|$. Walk from $v(S)$ to $v(T)$ by single unit steps, raising coordinate i last: at each step the quantity $\Delta_i H$ changes by one second difference $\Delta_\ell \Delta_i H \leq 0$, and every intermediate vector has ith coordinate at most $v(T)_i < |N_i|$, so $\Delta_i H$ is defined all along. Hence $d_e(T) = \Delta_i H(v(T)) \leq \Delta_i H(v(S)) = d_e(S)$. □

D.2 THE GUARANTEE (PROPOSITION 4)

The statement is Theorem 1 of Goundan & Schulz (2007) restated in the model of Section 2; the proof is included for completeness. It is written for the sharper bound $f(T) \geq L_K(\eta_{\mathrm{sel}}) f(O^*)$ in the selection error η_{sel} of Definition 2 (Section 3.4); since $\eta_{\mathrm{sel}} \leq \eta$ and L_K is nonincreasing, the bound of Proposition 4 follows.

Let $S^0 = \emptyset, S^1, \ldots, S^K$ be the states of a run of predictive greedy with picks $e_0, \ldots, e_{K-1}$ and selection error η_{sel}, and put $r_t = f(O^*) - f(S^t)$.

Step 1: one greedy step covers a $1/(\eta_{\mathrm{sel}} K)$ share of the residual. Let $t < K$ be a step with $d_{e_t}(S^t) > 0$. Enumerate $O^* \setminus S^t = \{o_1, \ldots, o_m\}$, $m \leq K$, and telescope:

$$f(O^* \cup S^t) - f(S^t) = \sum_{i=1}^{m} \Big[f\big(S^t \cup \{o_1, \ldots, o_i\}\big) - f\big(S^t \cup \{o_1, \ldots, o_{i-1}\}\big) \Big] \;\leq\; \sum_{i=1}^{m} d_{o_i}(S^t),$$

one application of submodularity per term. Monotonicity gives $f(O^* \cup S^t) \geq f(O^*)$, so $r_t \leq \sum_{i=1}^{m} d_{o_i}(S^t) \leq K \max_{e \notin S^t} d_e(S^t)$. At a step with $g_t = d_{e_t}(S^t) > 0$, Definition 2 gives $M_t = \max_{e \notin S^t} d_e(S^t) = a_t g_t \leq \eta_{\mathrm{sel}}\, d_{e_t}(S^t)$ whenever η_{sel} is finite, hence

$$d_{e_t}(S^t) \;\geq\; \frac{r_t}{\eta_{\mathrm{sel}} K}, \qquad \text{and so} \qquad r_{t+1} = r_t - d_{e_t}(S^t) \;\leq\; \Big(1 - \frac{1}{\eta_{\mathrm{sel}} K}\Big) r_t.$$

Step 2: the zero-gain steps. Monotonicity makes $g_t \geq 0$ always, so the steps not covered by Step 1 are those with $g_t = 0$. If such a step is harmful ($M_t > 0$), then $a_t = \infty$, $\eta_{\mathrm{sel}} = \infty$, $L_K(\infty) = 0$, and there is nothing to prove. If it is benign ($M_t = 0$), the covering inequality $r_t \leq K M_t$ of Step 1, whose derivation does not use positivity of the chosen gain, reads $r_t \leq 0$: the run has already reached the optimal value, and $r_s = 0$ for every $s \geq t$ by monotonicity. So when η_{sel} is finite, the contraction of Step 1 applies at every step, with the residual 0 from the first zero step onwards.

Step 3: unrolling. With $r_0 = f(O^*)$, iterating K times gives $r_K \leq \big(1 - \frac{1}{\eta_{\mathrm{sel}} K}\big)^K f(O^*)$, hence

$$f(S^K) = f(O^*) - r_K \;\geq\; \Big(1 - \big(1 - \tfrac{1}{\eta_{\mathrm{sel}} K}\big)^K\Big) f(O^*) = L_K(\eta_{\mathrm{sel}})\, f(O^*).$$

Since $u = 1/(\eta_{\mathrm{sel}} K) \in (0, 1]$ and $1 - u \leq e^{-u}$, also $L_K(\eta_{\mathrm{sel}}) \geq 1 - e^{-1/\eta_{\mathrm{sel}}}$.

Step 4: the three error measures. At every state S^t of the run,

$$\max_{e \notin S^t} d_e(S^t) \;\leq\; \eta_u \max_{e \notin S^t} \tilde{d}_e(S^t) \;=\; \eta_u\, \tilde{d}_{e_t}(S^t) \;\leq\; \eta_u \eta_o\, d_{e_t}(S^t),$$

the first step by the lower band at the maximizing element, the equality by the greedy choice, the last step by the upper band at (S^t, e_t). Read as the step ratio a_t of Definition 2, this gives $a_t \leq \eta_u \eta_o$ for the global band and $a_t \leq \eta_{\mathrm{tr}}$ for the band restricted to the states of the run, at every step with $g_t > 0$; at a step with $g_t = 0$ the display forces $M_t = 0$ and $a_t = 1$. Taking the maximum over steps gives $\eta_{\mathrm{sel}} \leq \eta_{\mathrm{tr}} \leq \eta$ ($\eta_{\mathrm{tr}} \leq \eta$ because the states of the run are a subset of all sets), the chain stated in Section 3.4. The map $x \mapsto 1 - 1/(xK)$ increases on $x \geq 1$ with values in $[0, 1)$, so L_K decreases, and $L_K(\eta_{\mathrm{sel}}) \geq L_K(\eta_{\mathrm{tr}}) \geq L_K(\eta)$.

Remark 7 (Stepwise product bound)**.** *Retaining the per-step ratios individually, the same unrolling gives*

$$f(S^K) \;\geq\; \Big(1 - \prod_{t=0}^{K-1} \big(1 - \tfrac{1}{K a_t}\big)\Big) f(O^*), \qquad 1/\infty = 0,$$

which degrades a harmful zero step to a contraction factor of 1 *instead of collapsing the whole bound to* 0*.*

D.3 PATH INDEPENDENCE (LEMMA 1)

Let $S \subseteq N$ and $e, e' \notin S$ with $e \neq e'$, and assume $\tilde{d}_e(S) \geq \tilde{d}_{e'}(S)$.

The exchange identity. Expanding $\tilde{f}(S\cup\{e,e'\})-\tilde{f}(S)$ in the two orders, first e then e' and first e' then e,

$$\tilde{d}_e(S)+\tilde{d}_{e'}(S\cup\{e\}) \;=\; \tilde{f}(S\cup\{e,e'\})-\tilde{f}(S) \;=\; \tilde{d}_{e'}(S)+\tilde{d}_e(S\cup\{e'\}).$$

This uses nothing but the definition of a set function. Subtracting and applying the hypothesis,

$$\tilde{d}_{e'}(S\cup\{e\})-\tilde{d}_e(S\cup\{e'\})=\tilde{d}_{e'}(S)-\tilde{d}_e(S) \;\le\; 0. \tag{4}$$

Part (i). Chain the two bands of Definition 1 around equation 4:

$$d_{e'}(S\cup\{e\}) \;\le\; \eta_u\,\tilde{d}_{e'}(S\cup\{e\}) \;\le\; \eta_u\,\tilde{d}_e(S\cup\{e'\}) \;\le\; \eta_u\eta_o\,d_e(S\cup\{e'\}) \;=\; \eta\,d_e(S\cup\{e'\}),$$

the first inequality by the lower band at $(S\cup\{e\},e')$, the second by equation 4, the third by the upper band at $(S\cup\{e'\},e)$. Dividing by η gives (i).

Part (ii). The same two-order expansion for f reads $d_e(S)+d_{e'}(S\cup\{e\})=d_{e'}(S)+d_e(S\cup\{e'\})$, that is $d_{e'}(S)-d_e(S)=d_{e'}(S\cup\{e\})-d_e(S\cup\{e'\})$. Substituting (i) in the form $-d_e(S\cup\{e'\})\le -\frac{1}{\eta}d_{e'}(S\cup\{e\})$,

$$d_{e'}(S)-d_e(S) \;\le\; d_{e'}(S\cup\{e\})-\frac{1}{\eta}\,d_{e'}(S\cup\{e\}) \;=\; \Big(1-\frac{1}{\eta}\Big)d_{e'}(S\cup\{e\}),$$

which is (ii). Only $\eta\ge 1$ is used beyond the bands, and only to keep the factor $1-1/\eta$ nonnegative when the inequality is applied later.

D.4 The exact approximation ratio (Theorem 1)

Roadmap. Lemma 1, in the sharp form of Section 3.2, limits how much a single step can lose now and cost later at the same time. The two-phase trajectory of the attaining instances below is the shape that limit leaves available. The multipliers equation 6 turn that shape into a lower bound valid for every run, the instances equation 10 attain it, and Appendix D.6 reads the same certificate backwards and recovers the trajectory from the equality case. The validity of the four constraint families is in Appendix D.5.

Throughout this subsection $K\ge 2$, $\eta\ge 1$, $f(O^*)=1$, and

$$k_1=(K-1)\eta+1, \qquad q=\frac{(K-1)\eta}{k_1}=1-\frac{1}{k_1}, \qquad V_j(\eta)=1-q^{\,j}\Big(1-\frac{K-j}{K\eta}\Big).$$

Note $k_1\ge K\ge 2$, so $q\in(0,1)$, and $(1-q)k_1=1$.

The reduced linear program. Every run of predictive greedy induces $d_t=d_{e_t}(S^t)$ for $t=0,\dots,K-1$ and, after enumerating $O^*=\{o_1,\dots,o_K\}$,

$$g_{t,i}=\begin{cases} d_{o_i}(S^t), & o_i\notin S^t,\\ 0, & o_i\in S^t, \end{cases} \qquad 0\le t\le K,\ 1\le i\le K.$$

Appendix D.5 derives, one family at a time and with the case $e_t\in O^*$ treated explicitly, that this vector satisfies

$$\begin{aligned} \textstyle\sum_i g_{t,i}+\sum_{s<t} d_s \;&\ge\; 1 \quad (\mathrm{sum}(t)),\\ d_t-g_{t,i}/\eta \;&\ge\; 0 \quad (\mathrm{pred}(t,i)),\\ g_{t,i}-g_{t+1,i} \;&\ge\; 0 \quad (\mathrm{mono}(t,i)),\\ \big(1-\tfrac{1}{\eta}\big)g_{t+1,i}-g_{t,i}+d_t \;&\ge\; 0 \quad (\mathrm{cons}(t,i)), \end{aligned} \tag{5}$$

together with $d_t\ge 0$, $g_{t,i}\ge 0$, and that $\sum_{t<K} d_t=f(S^K)$ by telescoping. Consequently the minimum of $\sum_{t<K} d_t$ over the polyhedron equation 5 is at most $\rho_K(\eta)$: the polyhedron contains every realizable run and possibly more.

Lower bound: greedy is never worse than $\min_j V_j$. Fix $0 \le j \le K-1$ and let $\eta \in [K-j, K-j+1]$ (in $[K,\infty)$ when $j=0$). Write $M = K\eta - (K-j)$ and define nonnegative numbers $\lambda_S(t)$, $\lambda_P(t)$, $\lambda_C(t)$ (one per constraint family, the same for every index i by the symmetry of equation 5 in i) as follows. For $j \ge 1$,

$$\begin{aligned} \lambda_S(0) &= \frac{q^{j-1}M}{Kk_1}, & \lambda_S(t) &= \frac{q^{j-1-t}M}{k_1^2} \quad (1 \le t \le j-1), \\ \lambda_S(j) &= \frac{K-j+1-\eta}{k_1}, & \lambda_C(t) &= \frac{q^{j-1-t}M}{Kk_1} \quad (t \le j-1), \\ \lambda_C(t) &= \frac{K-1-t}{K(\eta-1)} \quad (t \ge j), & \lambda_P(t) &= \frac{\eta-(K-t)}{K(\eta-1)} \quad (t \ge j), \end{aligned} \tag{6}$$

with $\lambda_S(t) = 0$ for $t > j$ and $\lambda_P(t) = 0$ for $t < j$; for $j = 0$,

$$\begin{aligned} \lambda_S(0) &= \frac{1}{\eta}, & \lambda_S(t) &= 0 \quad (t \ge 1), \\ \lambda_C(t) &= \frac{K-1-t}{K(\eta-1)}, & \lambda_P(t) &= \frac{\eta-(K-t)}{K(\eta-1)} \quad (0 \le t \le K-1). \end{aligned} \tag{7}$$

All multipliers of the family mono are 0: the lower bound does not use that the optimal gains decrease along the run.

Nonnegativity. On the segment, $\eta \ge K-j \ge 1$ and $K\eta > K-j$, so $M > 0$; since $q \in (0,1)$ and $k_1 > 0$, the multipliers built from powers of q are positive. For $\lambda_S(j)$, the segment condition $\eta \le K-j+1$ is $K-j+1-\eta \ge 0$, with equality at the right endpoint. For $\lambda_C(t)$ with $t \ge j$, $K-1-t \ge 0$. For $\lambda_P(t)$ with $t \ge j$, $\eta-(K-t) \ge \eta-(K-j) \ge 0$, with equality at $t=j$ at the left endpoint. The denominator $\eta-1$ is positive whenever $j \le K-2$, since then $\eta \ge K-j \ge 2$. In the remaining case $j = K-1$, whose segment is $[1,2]$, the range $t \ge j$ contains only $t = K-1$, where $K-1-t = 0$ and $\lambda_C(K-1)$ is defined to be 0, while $\lambda_P(K-1) = \frac{\eta-1}{K(\eta-1)}$ is defined by cancellation to be $\frac{1}{K}$. So equation 6 has no singularity on any segment.

The weighted-sum identity. For an arbitrary vector $(d_t, g_{t,i})$, not necessarily feasible, consider

$$\begin{aligned} \Sigma \;:=\; & \sum_{t=0}^{j} \lambda_S(t)\Big[\sum_i g_{t,i} + \sum_{s<t} d_s - 1\Big] \;+\; \sum_{t=j}^{K-1} \sum_i \lambda_P(t)\Big[d_t - \frac{g_{t,i}}{\eta}\Big] \\ & + \sum_{t=0}^{K-1} \sum_i \lambda_C(t)\Big[\big(1-\tfrac{1}{\eta}\big)g_{t+1,i} - g_{t,i} + d_t\Big]. \end{aligned} \tag{8}$$

The last summand of the third group is void, because $\lambda_C(K-1) = 0$. Collecting coefficients in Σ gives three families of scalar identities. The coefficient of $g_{t,i}$, for $0 \le t \le K$, is $\lambda_S(t) - \lambda_P(t)/\eta - \lambda_C(t) + (1-1/\eta)\lambda_C(t-1)$, with the conventions $\lambda_C(-1) = 0$, $\lambda_C(K) = 0$ and $\lambda_S(t) = \lambda_P(t) = 0$ outside the ranges listed above; the coefficient of d_t is $\sum_{s>t} \lambda_S(s) + K\big(\lambda_P(t) + \lambda_C(t)\big)$; the constant is $-\sum_{t \le j} \lambda_S(t)$. The claim is that these equal 0, 1 and $-V_j$ respectively, so that $\Sigma = \sum_{t<K} d_t - V_j$ identically.

(a) The coefficient of $g_{t,i}$ vanishes. Take $j \ge 1$ first. For $t = 0$ one has $\lambda_C(-1) = 0$ and $\lambda_P(0) = 0$, so the coefficient is $\lambda_S(0) - \lambda_C(0) = 0$, the two entries of equation 6 being the same number. For $1 \le t \le j-1$, again $\lambda_P(t) = 0$, and dividing the coefficient by $q^{j-1-t}M/(Kk_1) > 0$ turns it into

$$\frac{K}{k_1} - 1 + \Big(1 - \frac{1}{\eta}\Big)q = \frac{K-k_1}{k_1} + \frac{\eta-1}{\eta} \cdot \frac{(K-1)\eta}{k_1} = \frac{(K-1)(1-\eta) + (\eta-1)(K-1)}{k_1} = 0,$$

using $K - k_1 = (K-1)(1-\eta)$. For $t = j$, write $u = K-j$, so that $\eta \in [u, u+1]$ and $M = K\eta - u$, and multiply the coefficient by $K\eta k_1(\eta-1) > 0$:

$$\underbrace{K\eta(\eta-1)(u+1-\eta)}_{\text{from } \lambda_S(j)} \underbrace{-\,k_1(\eta-u)}_{\text{from } \lambda_P(j)/\eta} \underbrace{-\,(u-1)\eta k_1}_{\text{from } \lambda_C(j)} \underbrace{+\,(\eta-1)^2(K\eta-u)}_{\text{from } (1-1/\eta)\lambda_C(j-1)}.$$

The two middle terms combine as $-k_1\big[(\eta-u) + (u-1)\eta\big] = -k_1 u(\eta-1)$, so with $A = \eta - 1$ the whole expression is $A\big[K\eta(u+1-\eta) + A(K\eta - u) - k_1 u\big]$, and the bracket expands to

$$K\eta u + K\eta - K\eta^2 + K\eta^2 - K\eta - \eta u + u - k_1 u = u\big[K\eta - \eta + 1 - k_1\big] = 0.$$

For $j+1 \le t \le K-1$ one has $\lambda_S(t) = 0$, and multiplying by $K\eta(\eta-1)$ the coefficient becomes

$$-(\eta - K + t) - \eta(K-1-t) + (\eta-1)(K-t) = 0.$$

For $t = K$ only the last term survives and $\lambda_C(K-1) = 0$. The multiplication by $\eta - 1$ is legitimate on every segment with $j \le K-2$, where $\eta \ge 2$; on the segment $j = K-1$, which is $[1,2]$, both sides of each identity are rational functions of η continuous at $\eta = 1$ after the cancellation recorded above, so the identity at $\eta = 1$ is the limit of the identity on $(1,2]$. For $j = 0$ the first computation is replaced by the single case $t = 0$: multiplying by $K\eta(\eta-1)$, $K(\eta-1) - (\eta - K) - \eta(K-1) = 0$, and the cases $1 \le t \le K-1$ and $t = K$ are as before.

(b) The coefficient of d_t equals 1. For $t \ge j$ the sum $\sum_{s>t} \lambda_S(s)$ is empty, and

$$K\big(\lambda_P(t) + \lambda_C(t)\big) = K \cdot \frac{(\eta - K + t) + (K - 1 - t)}{K(\eta - 1)} = 1,$$

which at $j = K-1$, $\eta = 1$ reads $K(\frac{1}{K} + 0) = 1$ with the cancelled values. For $t = j-1$ (so $j \ge 1$), $\lambda_P(j-1) = 0$ and the only surviving λ_S is $\lambda_S(j)$, so the coefficient is

$$\frac{K - j + 1 - \eta}{k_1} + K \cdot \frac{M}{Kk_1} = \frac{(K - j + 1 - \eta) + (K\eta - K + j)}{k_1} = \frac{(K-1)\eta + 1}{k_1} = 1.$$

For $t \le j-2$ the value is unchanged from $t+1$ to t: subtracting the two coefficients and using $\lambda_P(t) = \lambda_P(t+1) = 0$,

$$\lambda_S(t+1) + K\big[\lambda_C(t) - \lambda_C(t+1)\big] = \frac{q^{j-2-t}M}{k_1^2} + K \cdot \frac{Mq^{j-2-t}}{Kk_1}(q-1) = 0,$$

because $q - 1 = -1/k_1$. Downward induction from $t = j-1$ gives the value 1 for every $t \le j-1$.

(c) The constant equals $-V_j$. For $j \ge 1$, the middle multipliers form a geometric sum with ratio q, summed with $1 - q = 1/k_1$:

$$\sum_{t=1}^{j-1} \lambda_S(t) = \frac{M}{k_1^2} \cdot \frac{1 - q^{j-1}}{1 - q} = \frac{M\big(1 - q^{j-1}\big)}{k_1},$$

so that, using $M + (K - j + 1 - \eta) = k_1$,

$$\sum_{t=0}^{j} \lambda_S(t) = \frac{1}{k_1}\Big[\frac{Mq^{j-1}}{K} + M - Mq^{j-1} + (K - j + 1 - \eta)\Big] = \frac{1}{k_1}\Big[k_1 - Mq^{j-1}\frac{K-1}{K}\Big].$$

The definition of q gives $\frac{K-1}{K} = \frac{qk_1}{K\eta}$, so $Mq^{j-1}\frac{K-1}{K} = \frac{Mq^j k_1}{K\eta}$ and

$$\sum_{t=0}^{j} \lambda_S(t) = 1 - \frac{Mq^j}{K\eta} = 1 - q^j\,\frac{K\eta - (K-j)}{K\eta} = V_j.$$

For $j = 0$ the sum is the single term $\lambda_S(0) = 1/\eta = V_0$.

Conclusion of the lower bound. By (a), (b), (c), identity equation 8 reads $\sum_{t<K} d_t - V_j = \Sigma$ for every vector $(d_t, g_{t,i})$. At a point feasible for equation 5 every bracket in Σ is nonnegative and every multiplier is nonnegative, so $\Sigma \ge 0$ and $\sum_{t<K} d_t \ge V_j$. Since every run of predictive greedy at error η gives such a feasible point with objective $f(S^K)/f(O^*)$, $\rho_K(\eta) \ge V_j(\eta)$ on the segment of j.

Which V_j is smallest. For $0 \le i \le K-2$,

$$V_i - V_{i+1} = q^i\Big[q - \frac{q(K-i-1)}{K\eta} - 1 + \frac{K-i}{K\eta}\Big] = q^i\Big[-\frac{1}{k_1} + \frac{1}{K\eta}\Big(1 + \frac{K-i-1}{k_1}\Big)\Big] = \frac{q^i\,(K - i - \eta)}{K\eta k_1}, \tag{9}$$

where the middle step uses $q - 1 = -1/k_1$ and $(K-i) - q(K-i-1) = 1 + \frac{K-i-1}{k_1}$, and the last step uses $k_1 - K\eta = 1 - \eta$. On the segment $\eta \in [K-j, K-j+1]$ the sign of $K - i - \eta$ is nonnegative for $i \le j-1$ and nonpositive for $i \ge j$, so the sequence $V_0, \dots, V_{K-1}$ decreases up to

index j and increases after it: $\min_i V_i = V_j$ there, and neighbouring segments agree at the integer endpoints. The breakpoints are therefore the integers $2, \dots, K$, and $\rho_K(\eta) \geq \min_i V_i(\eta)$ for every $\eta \geq 1$. At $\eta \geq K$ the active index is $j = 0$ and $\min_i V_i = V_0 = 1/\eta$, while at $\eta < K$ the active index is at least 1 and equation 9 with $i = 0$ gives $V_1 < V_0$, so $\min_i V_i = 1/\eta$ exactly when $\eta \geq K$.

Closed forms for small K. The displayed forms of Theorem 1 are specializations of $\min_j V_j$; for instance at $K = 3$ and $j = 2$, $V_2 = 1 - \frac{4\eta^2}{(2\eta+1)^2} \cdot \frac{3\eta-1}{3\eta} = \frac{16\eta+3}{3(2\eta+1)^2}$. For $K = 4$ the value is $\frac{135\eta^2+36\eta+4}{4(3\eta+1)^3}$, $\frac{21\eta+2}{2(3\eta+1)^2}$, $\frac{13}{4(3\eta+1)}$, $\frac{1}{\eta}$ on $[1, 2]$, $[2, 3]$, $[3, 4]$, $[4, \infty)$.

Upper bound: the value V_j is attained. Fix $K \geq 2$, $0 \leq j \leq K - 1$, and η with $\eta \geq K - j$, and fix any split $\eta_u \eta_o = \eta$ with $\eta_u, \eta_o > 0$. Partition a ground set of $n = 2K$ elements into

$$C \ (|C| = j), \qquad P \ (|P| = K - j), \qquad O \ (|O| = K),$$

write $x = |S \cap C|$, $z = |S \cap P|$, $y = |S \cap O|$, and put

$$\delta_j = \frac{q^j}{K\eta}, \qquad W(0) = \frac{k_1}{K\eta_u}, \qquad W(y) = \frac{(K-y)\eta_o}{K} \ \ (1 \leq y \leq K), \qquad \chi(y) = \mathbf{1}[y < K].$$

The instance is

$$f(S) = 1 - q^x\Big(1 - \frac{y}{K}\Big) + z\,\delta_j\,\chi(y), \qquad \tilde{f}(S) = W(0) - q^x W(y) + z\,\eta_o \delta_j\,\chi(y). \tag{10}$$

Both are counting functions of (x, z, y), so Lemma 5 applies with $r = 3$. Two normalizations are immediate: $f(\emptyset) = 1 - 1 = 0$ and $\tilde{f}(\emptyset) = W(0) - W(0) = 0$.

Marginal gains. Directly from equation 10, for $y < K$,

$$\Delta_C f = q^x(1-q)\frac{K-y}{K}, \qquad \Delta_P f = \delta_j, \qquad \Delta_O f = \begin{cases} q^x/K, & y \leq K-2, \\ q^x/K - z\delta_j, & y = K-1, \end{cases}$$

and at $y = K$ all three gains are 0. On the predictor side, $\Delta_C \tilde{f} = q^x(1-q)W(y)$ and $\Delta_P \tilde{f} = \eta_o \delta_j$, while

$$\Delta_O \tilde{f} = \begin{cases} q^x\big(W(0) - \frac{(K-1)\eta_o}{K}\big), & y = 0, \\ q^x\eta_o/K, & 1 \leq y \leq K-2, \\ q^x\eta_o/K - z\eta_o\delta_j, & y = K-1. \end{cases}$$

Two identities are used repeatedly: $(1-q)k_1 = 1$ and

$$W(0) - \frac{(K-1)\eta_o}{K} = \frac{(K-1)\eta + 1}{K\eta_u} - \frac{(K-1)\eta_u\eta_o}{K\eta_u} = \frac{1}{K\eta_u}. \tag{11}$$

Monotonicity and submodularity of f. The three first differences are nonnegative: $\Delta_C f \geq 0$ and $\Delta_P f > 0$ are immediate from $q \in (0, 1)$, and at $y = K - 1$

$$\Delta_O f = \frac{q^x}{K} - z\,\frac{q^j}{K\eta} \ \geq \ 0 \iff \eta\, q^{\,x-j} \ \geq \ z,$$

which holds because $x \leq j$ makes $q^{\,x-j} \geq 1$ and $z \leq K - j \leq \eta$ by the standing assumption $\eta \geq K - j$. This is the only place where the segment condition is consumed, and it is where the breakpoints come from. The second differences are $\Delta_C\Delta_C f = -q^x(1-q)^2\frac{K-y}{K} \leq 0$; $\Delta_O\Delta_C f = \Delta_C\Delta_O f = -q^x(1-q)/K \leq 0$ for every $y \leq K-1$; $\Delta_P\Delta_C f = \Delta_C\Delta_P f = \Delta_P\Delta_P f = 0$; $\Delta_O\Delta_P f = \Delta_P\Delta_O f = 0$ for $y \leq K-2$ and $-\delta_j \leq 0$ at $y = K-1$; $\Delta_O\Delta_O f = 0$ for $y \leq K-3$ and $-z\delta_j \leq 0$ at $y = K-2$. Lemma 5 then gives that f is monotone and submodular.

The error is exactly (η_u, η_o). Divide each predicted gain by the corresponding true gain, using equation 11 in the case $y = 0$ of direction O. Every single-element ratio $\tilde{d}/d$ with $d > 0$ takes one of three values:

$$\frac{\Delta_C\tilde{f}}{\Delta_C f}\Big|_{y=0} = W(0) = \frac{k_1}{K\eta_u}, \qquad \frac{\Delta_O\tilde{f}}{\Delta_O f}\Big|_{y=0} = \frac{1}{\eta_u}, \qquad \eta_o \ \text{in all remaining cases},$$

the remaining cases being direction C with $y \geq 1$, direction P, and direction O with $y \geq 1$; wherever a true gain vanishes, the predicted gain vanishes with it. The three values are ordered by $k_1 - K = (K-1)(\eta-1) \geq 0$ and $K\eta - k_1 = \eta - 1 \geq 0$, that is $\frac{1}{\eta_u} \leq \frac{k_1}{K\eta_u} \leq \eta_o$. The outer two are attained, so $\max \tilde{d}/d = \eta_o$ and $\max d/\tilde{d} = \eta_u$: the pair sits exactly on the band of Definition 1.

The optimum. $f(O) = f(0,0,K) = 1$, and for every S with $|S| \leq K$ one has $f(S) \leq 1$: at $y = K$ the value is exactly 1, and at $y \leq K-1$, $f(S) = 1 - q^x \frac{K-y}{K} + z\delta_j \leq 1$ because $z\delta_j \leq (K-j)\frac{q^j}{K\eta} \leq \frac{q^x}{K} \leq q^x \frac{K-y}{K}$, using $z \leq K - j \leq \eta$ and $q^j \leq q^x$. So $f(O^*) = 1$ with $O^* = O$.

The per-step gain table. At each state of the run, the true and predicted gains of the three kinds of candidate are:

step t	state (x,z,y)	candidate	true gain	predicted gain	picked
$t < j$	$(t,0,0)$	$e \in C$	q^t/k_1	$q^t/(K\eta_u)$	yes
		$e \in P$	$q^j/(K\eta)$	$q^j/(K\eta_u)$	no
		$e \in O$	q^t/K	$q^t/(K\eta_u)$	no (tie)
$t \geq j$	$(j, t-j, 0)$	$e \in P$	$q^j/(K\eta)$	$q^j/(K\eta_u)$	yes
		$e \in O$	q^j/K	$q^j/(K\eta_u)$	no (tie)

The entries follow from the gain formulas above: at $y = 0$ a C-step has true gain $q^t(1-q) = q^t/k_1$ and predicted gain $q^t(1-q)W(0) = q^t/(K\eta_u)$; a P-step has predicted gain $\eta_o\delta_j = q^j/(K\eta_u)$; an O-step at $y = 0$ has predicted gain $q^x/(K\eta_u)$ by equation 11.

The greedy induction. Claim: for $0 \leq t \leq K$ the state after t steps is $(x,z,y) = (\min(t,j),\ \max(t-j,0),\ 0)$. It holds at $t = 0$. Assume it at $t < j$. By the table the maximum predicted gain at that state is $q^t/(K\eta_u)$, attained by every remaining element of C and every element of O, while every element of P scores $q^j/(K\eta_u)$, strictly smaller because $q \in (0,1)$ and $j > t$. Adversarial tie breaking picks an element of C, so the state becomes $(t+1,0,0)$. Now assume the claim at t with $j \leq t < K$. The state is $(j, t-j, 0)$, C is exhausted, and the predicted gains of the remaining P elements and of the O elements are both $q^j/(K\eta_u)$: the step is a tie and adversarial tie breaking picks an element of P, available because $t - j < K - j = |P|$. After K steps $S^K = C \cup P$ and

$$f(S^K) = 1 - q^j + (K-j)\,\delta_j = 1 - q^j\Big(1 - \frac{K-j}{K\eta}\Big) = V_j(\eta),$$

which is $V_j(\eta) f(O^*)$ because $f(O^*) = 1$. Choosing j with $\eta \in [K-j, K-j+1]$ gives $\rho_K(\eta) \leq \min_i V_i(\eta)$, which together with the lower bound is Theorem 1.

The role of the ties. Every one of the K steps above is a tie with an element of O, so adversarial tie breaking is load bearing. On the steps $t \geq j$ the tie is forced by the band: the prediction constraint of equation 5 holds with equality at the optimum, so $\eta_o\, d_P = d_O/\eta_u$, and the largest value the band allows a P element and the smallest it allows an O element coincide. Whether the same value is the infimum over runs with strict predicted preferences is not shown here; Theorem 1 is stated under adversarial tie breaking.

Remark 8 (The gap to L_K is strict)**.** *Fix $K \geq 2$ and $\eta > 1$. Dropping the families* mono *and* cons *from equation 5 leaves the program behind Proposition 4, whose optimum is $L_K(\eta)$ and is attained only at $d_t = \frac{r_t}{\eta K}$ and $g_{t,i} = \frac{r_t}{K}$ for every t and i, where $r_t = 1 - \sum_{s<t} d_s$: at each step $d_t \geq \max_i g_{t,i}/\eta \geq \sum_i g_{t,i}/(\eta K) \geq r_t/(\eta K)$, with equality only when all $g_{t,i}$ equal r_t/K, and choosing d_t above this floor at any step lowers every later residual, hence lowers r_K and raises the objective $\sum_t d_t = 1 - r_K$. At that unique point $r_{t+1} = r_t\big(1 - \frac{1}{\eta K}\big) < r_t$, so*

$$\Big(1 - \frac{1}{\eta}\Big) g_{t+1,i} \;<\; \Big(1 - \frac{1}{\eta}\Big) g_{t,i} \;=\; g_{t,i} - d_t,$$

and $\mathrm{cons}(t,i)$ fails. The optimum of equation 5, which by Theorem 1 equals $\rho_K(\eta)$, is therefore strictly larger than $L_K(\eta)$.

Remark 9 (Ground-set size)**.** *Write $\rho_{n,K}(\eta)$ for the approximation ratio over instances on a ground set of size n. Restricting a worst instance to $T \cup O^*$ leaves the run legal and the ratio unchanged,*

and padding with elements of identically zero true and predicted gain preserves the band and the run, so $\rho_{n,K}(\eta) = \rho_{2K,K}(\eta)$ for every $n \geq 2K$; ρ_K denotes this common value throughout.

Proof. Both inequalities come from two operations on instances, each of which preserves the model, the error band, the optimum, the greedy run and the output value.

Padding. Let $(f, \tilde{f})$ be an instance on N_0 with $|N_0| = n_0 \geq K$, and let $N \supseteq N_0$ with $|N| = n \geq n_0$. Extend by $f'(S) = f(S \cap N_0)$ and $\tilde{f}'(S) = \tilde{f}(S \cap N_0)$. For $e \in N_0$ the marginal gains are $d'_e(S) = d_e(S \cap N_0)$ and $\tilde{d}'_e(S) = \tilde{d}_e(S \cap N_0)$, and for $e \in N \setminus N_0$ both vanish identically. Since $S \subseteq T$ implies $S \cap N_0 \subseteq T \cap N_0$, monotonicity and the diminishing-returns form of submodularity pass from f to f', and $f'(\emptyset) = \tilde{f}'(\emptyset) = 0$. The band of Definition 1 holds for $(f', \tilde{f}')$: for $e \in N \setminus N_0$ it reads $0 \leq 0 \leq 0$, and for $e \in N_0$ it is the band of the original instance at the state $S \cap N_0$; the extremal ratios of the original instance are still attained, so the error is exactly the same pair. Every K-set S has $f'(S) = f(S \cap N_0)$, so the optimal value is unchanged and O^* is still optimal. Finally, each new element has predicted gain 0 at every state, which is the smallest value a predicted gain can take, so at every step the element selected by the original run still maximizes the predicted gain, possibly after a tie that adversarial tie breaking resolves in its favour. The run and its value are therefore unchanged, and every ratio realized at ground-set size n_0 is realized at size n: $\rho_{n,K}(\eta) \leq \rho_{n_0,K}(\eta)$ for $n \geq n_0$.

Restriction. Let $(f, \tilde{f})$ be an instance on N with $|N| = n$, let $T = S^K$ be the output of a run of predictive greedy, and let O^* be an optimal K-set. Put $N' = T \cup O^*$, so $K \leq |N'| \leq 2K$, and restrict both functions to the subsets of N'. The restriction of f is monotone submodular with value 0 at $\emptyset$, and the band holds for every $S \subseteq N'$ and $e \in N' \setminus S$ because it holds in the larger instance. The set O^* lies in N', its value is unchanged, and no K-subset of N' exceeds it because none of N does, so O^* is optimal for the restricted instance and the normalization is unchanged. Each pick $e_t \in T \subseteq N'$ maximizes $\tilde{d}_e(S^t)$ over all $e \in N \setminus S^t$, hence in particular over $e \in N' \setminus S^t$, so the same sequence of picks is a legal adversarial-tie run on the restricted instance and returns the same value. Every ratio realized at size n is therefore realized at some size $n' \leq 2K$, and padding raises n' to exactly $2K$ without changing the ratio, so $\rho_{2K,K}(\eta) \leq \rho_{n,K}(\eta)$.

For $n \geq 2K$ the padding step also gives $\rho_{n,K}(\eta) \leq \rho_{2K,K}(\eta)$, and the two inequalities together give equality. Since the instances equation 10 live on $2K$ elements and the bound of the certificate is valid on every ground set, Theorem 1 computes this common value. □

D.5 Validity of the reduced constraints

This subsection supplies the step consumed by Appendix D.4: every constraint of equation 5 is a valid inequality, so the linear program is a relaxation of the set of realizable runs and its optimum is a lower bound on the approximation ratio for every K.

Let f be monotone submodular with $f(\emptyset) = 0$, let $\tilde{f}$ be an arbitrary set function with $\tilde{f}(\emptyset) = 0$ obeying the band of Definition 1, and let $e_0, \ldots, e_{K-1}$ and $S^0 = \emptyset, \ldots, S^K$ be the picks and states of a run of predictive greedy, so that $e_t \in \arg\max_{e \notin S^t} \tilde{d}_e(S^t)$. Fix an optimal set $O^* = \{o_1, \ldots, o_K\}$ with $f(O^*) = 1$ (if $|O^*| < K$, set $g_{t,i} = 0$ for the unused indices; every case (a) below covers them). Put $d_t = d_{e_t}(S^t)$ and

$$g_{t,i} = \begin{cases} d_{o_i}(S^t), & o_i \notin S^t, \\ 0, & o_i \in S^t, \end{cases}$$

the second line being the *zeroing convention*, which handles a step at which greedy selects an element of O^*: from that step on the corresponding variable is 0. Finally $r_t = 1 - \sum_{s<t} d_s$, which telescopes to $r_t = f(O^*) - f(S^t)$.

Throughout, $t \in \{0, \ldots, K-1\}$, $i \in \{1, \ldots, K\}$, and the three cases are (a) $o_i \in S^t$; (b) $o_i \notin S^t$ and $o_i \neq e_t$; (c) $o_i = e_t$, the case $e_t \in O^*$.

Lemma 6 (Signs). *$d_t \geq 0$ and $g_{t,i} \geq 0$ for all t, i.*

Proof. Each is a marginal gain of f or, under the zeroing convention, 0; this uses monotonicity of f. □

Lemma 7 (Coverage, $\mathrm{sum}(t)$)**.** $\sum_{i=1}^{K} g_{t,i} \geq r_t$ *for every* t.

Proof. Let $I_t = \{i : o_i \notin S^t\}$, so that $\sum_i g_{t,i} = \sum_{i \in I_t} d_{o_i}(S^t)$ by the zeroing convention; the indices of already picked optimal elements drop out of the sum. Order $I_t = \{i_1, \ldots, i_m\}$ and telescope

$$f(S^t \cup O^*) - f(S^t) = \sum_{\ell=1}^{m} \Big[f\big(S^t \cup \{o_{i_1}, \ldots, o_{i_\ell}\}\big) - f\big(S^t \cup \{o_{i_1}, \ldots, o_{i_{\ell-1}}\}\big) \Big] \leq \sum_{\ell=1}^{m} d_{o_{i_\ell}}(S^t),$$

one application of submodularity per term. Monotonicity gives $f(S^t \cup O^*) \geq f(O^*)$, so $\sum_i g_{t,i} \geq f(O^*) - f(S^t) = r_t$. □

Lemma 8 (Greedy choice inside the band, $\mathrm{pred}(t,i)$)**.** $d_t \geq g_{t,i}/\eta$ *for every* t, i.

Proof. Case (a): $g_{t,i} = 0$ and $d_t \geq 0$. Case (b): chain

$$d_t = d_{e_t}(S^t) \;\geq\; \frac{\tilde{d}_{e_t}(S^t)}{\eta_o} \;\geq\; \frac{\tilde{d}_{o_i}(S^t)}{\eta_o} \;\geq\; \frac{d_{o_i}(S^t)}{\eta_u \eta_o} = \frac{g_{t,i}}{\eta},$$

the first inequality by the upper band at (S^t, e_t), the second by the greedy choice (o_i is eligible at step t), the third by the lower band at (S^t, o_i). Writing $P = \tilde{d}_{e_t}(S^t)$ and $Q = \tilde{d}_{o_i}(S^t)$, the same chain says that the slack decomposes into three nonnegative terms,

$$d_t - \frac{g_{t,i}}{\eta} = \underbrace{\frac{\eta_o d_t - P}{\eta_o}}_{\text{upper band at } (S^t, e_t)} + \underbrace{\frac{P - Q}{\eta_o}}_{\text{greedy}} + \underbrace{\frac{\eta_u Q - g_{t,i}}{\eta_u \eta_o}}_{\text{lower band at } (S^t, o_i)}. \tag{12}$$

Case (c): $g_{t,i} = d_{e_t}(S^t) = d_t$ and the claim reads $d_t \geq d_t/\eta$, which holds because $d_t \geq 0$ and $\eta \geq 1$. □

Lemma 9 (Single-element submodularity, $\mathrm{mono}(t,i)$)**.** $g_{t+1,i} \leq g_{t,i}$ *for every* t, i.

Proof. Case (a): $o_i \in S^t \subseteq S^{t+1}$, so both sides are 0. Case (b): $o_i \notin S^{t+1}$ as well, so $g_{t+1,i} = d_{o_i}(S^{t+1}) \leq d_{o_i}(S^t) = g_{t,i}$ by diminishing returns. Case (c): $o_i = e_t \in S^{t+1}$, so $g_{t+1,i} = 0$ while $g_{t,i} = d_t \geq 0$. □

Lemma 10 (Path independence, $\mathrm{cons}(t,i)$)**.** $(1 - \frac{1}{\eta}) g_{t+1,i} \geq g_{t,i} - d_t$ *for every* t, i.

Proof. Case (a): the left side is 0 and the right side is $-d_t \leq 0$. Case (b): abbreviate $d = d_t$, $g = g_{t,i}$, $h = g_{t+1,i}$ and, for the predicted gains, $P = \tilde{d}_{e_t}(S^t)$, $Q = \tilde{d}_{o_i}(S^t)$, $R = \tilde{d}_{o_i}(S^{t+1})$. Because the four values of a set function on S^t, $S^{t+1} = S^t \cup \{e_t\}$, $S^t \cup \{o_i\}$ and $S^t \cup \{e_t, o_i\}$ determine both orders of insertion, the gains of e_t at $S^t \cup \{o_i\}$ are $d + h - g$ (true) and $P + R - Q$ (predicted). The claim is then the sum of three nonnegative terms:

$$\Big(1 - \frac{1}{\eta}\Big) h - g + d \;=\; \underbrace{\frac{\eta_o(d + h - g) - (P + R - Q)}{\eta_o}}_{\text{upper band at } (S^t \cup \{o_i\},\, e_t)} + \underbrace{\frac{P - Q}{\eta_o}}_{\text{greedy}} + \underbrace{\frac{\eta_u R - h}{\eta_u \eta_o}}_{\text{lower band at } (S^{t+1},\, o_i)}, \tag{13}$$

an identity in the six symbols, as expanding the right side shows; the first brace is nonnegative by Definition 1 at the off-trajectory state $S^t \cup \{o_i\}$, the second because o_i is eligible at step t and e_t was chosen, the third by the lower band. (Equivalently, this is part (ii) of Lemma 1 with its certificate written out.) Case (c): $g_{t+1,i} = 0$ by the zeroing convention and $g_{t,i} - d_t = 0$, so the constraint holds with equality; Lemma 1, which requires $e \neq e'$, is not invoked. □

Lemma 11 (Relaxation)**.** *For every instance* $(f, \tilde{f})$ *satisfying Definition 1 with* $\eta = \eta_u \eta_o$*, every optimal set* O^* *and every run of predictive greedy, the vector* $(d_t, g_{t,i})$ *is feasible for equation 5 and its objective value equals* $f(S^K)/f(O^*)$*. Hence the optimum of equation 5 is a lower bound on the approximation ratio of predictive greedy at budget* K *and error* η*.*

Proof. Feasibility is Lemmas 6 to 10. The objective is $\sum_{t<K} d_t = f(S^K)$ by telescoping, and $f(O^*) = 1$ by the normalization. □

Remark 10 (What is and is not claimed)**.** *Lemma 11 gives one direction only. The reverse direction, that every feasible point of equation 5 is realized by some instance, is not claimed; the attaining instances of Appendix D.4 supply what the theorem needs instead.*

Remark 11 (Which error measure each family tolerates)**.** *The families* $\mathrm{sum}(t)$ *and* $\mathrm{mono}(t,i)$ *use no error notion at all. The family* $\mathrm{pred}(t,i)$ *uses the band only at the greedy state* S^t*, so it stays valid with* η *replaced by* η_{tr}*, and, read directly as a ratio of true gains at greedy states, with* η *replaced by* η_{sel}*. The family* $\mathrm{cons}(t,i)$ *invokes Lemma 1, whose proof uses the band at the state* $S^t \cup \{o_i\}$*, off the greedy trajectory; so* cons *is not covered by either restricted measure. This is why the* L_K *bound, whose proof uses only* sum *and* pred*, holds at* η_{sel} *(Section 3.4), while the exact value of Theorem 1 is stated for the global* η*.*

Remark 12 (Which hypothesis each family consumes)**.** $\mathrm{sum}(t)$ *uses submodularity and monotonicity of* f *and no property of* $\tilde{f}$*.* $\mathrm{pred}(t,i)$ *uses the greedy choice and both bands, and no property of* f *beyond monotonicity.* $\mathrm{mono}(t,i)$ *uses submodularity, and monotonicity in case (c).* $\mathrm{cons}(t,i)$ *uses the greedy choice and both bands, and monotonicity in cases (a) and (c). No family uses submodularity of* $\tilde{f}$*, consistent with the predictor being an arbitrary set function.*

D.6 Structure of the attaining runs

With the variables of Appendix D.4, the equality case of the certificate determines the run.

Proposition 7 (Structure of attaining runs)**.** *Let* $K \geq 2$*, normalize* $f(O^*) = 1$*, and let either* $\eta \in (K-j, K-j+1)$ *for some* $1 \leq j \leq K-1$*, or* $j = 0$ *and* $\eta > K$*. If an adversarial-tie run of predictive greedy attains* $f(S^K) = \rho_K(\eta) = V_j(\eta)$*, then the variables it induces are*

$$d_t = \begin{cases} q^t/k_1, & t < j, \\ q^j/(K\eta), & t \geq j, \end{cases} \qquad g_{t,i} = \frac{q^{\min(t,j)}}{K} \quad \textit{for } 0 \leq t \leq K \textit{ and every } i.$$

Proof. Throughout, η lies in the interior of a segment, $\gamma = 1 - \frac{1}{\eta}$ and $r_t = 1 - \sum_{s<t} d_s$, so that $r_0 = 1$ and $r_{t+1} = r_t - d_t$. Fix an attaining run; Appendix D.5 makes the vector $(d_t, g_{t,i})$ feasible for equation 5. Nothing below assumes that the run avoids O^* or that the $g_{t,i}$ are equal across i; both come out of the argument.

Step 1: which multipliers are positive. On the interior of the segment, $M > 0$, $q \in (0,1)$, $k_1 > 0$ and $\eta > 1$ (for $j \leq K-2$, $\eta > K - j \geq 2$; for $j = K-1$ the segment is $(1,2)$). Reading the signs off equation 6 and equation 7: $\lambda_S(t) > 0$ exactly for $0 \leq t \leq j$ (the entries with $t \leq j-1$ are positive multiples of M, and $\lambda_S(j) = (K - j + 1 - \eta)/k_1 > 0$ strictly inside the segment); $\lambda_C(t) > 0$ exactly for $0 \leq t \leq K-2$; $\lambda_P(t) > 0$ exactly for $j \leq t \leq K-1$ (its numerator is $\eta - (K-t) \geq \eta - (K-j) > 0$); every multiplier of mono is 0. At an integer η one of these strict inequalities becomes an equality, which is the content of Remark 13.

Step 2: every supported constraint holds with equality. Identity equation 8 reads $\sum_{t<K} d_t - V_j = \Sigma$, and Σ is a sum of terms $\lambda_r \,\mathrm{slack}_r$ with $\lambda_r \geq 0$ and $\mathrm{slack}_r \geq 0$ at a feasible point. The run attains the value, so every term vanishes, so

$$\mathrm{sum}(t)\ (0 \leq t \leq j), \qquad \mathrm{cons}(t,i)\ (0 \leq t \leq K-2), \qquad \mathrm{pred}(t,i)\ (j \leq t \leq K-1)$$

hold with equality for every i. The constraints with zero multiplier, $\mathrm{sum}(t)$ for $t > j$, $\mathrm{cons}(K-1,i)$ and every $\mathrm{mono}(t,i)$, stay valid but are not forced; Step 7 uses two of them as inequalities.

Step 3: the geometric phase, $t < j$. Let $t < j$ (empty when $j = 0$). Both $\mathrm{sum}(t)$ and $\mathrm{sum}(t+1)$ carry indices at most j, so $\sum_i g_{t,i} = r_t$ and $\sum_i g_{t+1,i} = r_{t+1}$. Summing the K equalities $\mathrm{cons}(t,i)$ over i (legitimate because $t \leq j-1 \leq K-2$),

$$\gamma \sum_i g_{t+1,i} - \sum_i g_{t,i} + K d_t = 0, \qquad \text{that is} \qquad \gamma\, r_{t+1} = r_t - K d_t.$$

Substituting $r_{t+1} = r_t - d_t$ gives $d_t(K - \gamma) = r_t/\eta$, and $K - \gamma = k_1/\eta$, so $d_t = r_t/k_1$ and $r_{t+1} = q\, r_t$. With $r_0 = 1$, $r_t = q^t$ and $d_t = q^t/k_1$ for every $t < j$, and $r_j = q^j$.

Step 4: the switch at $t = j$. The index j lies in the support of both λ_S and λ_P, so $\mathrm{pred}(j, i)$ holds with equality for every i, which is $g_{j,i} = \eta\, d_j$, and $\mathrm{sum}(j)$ holds with equality, which is $\sum_i g_{j,i} = r_j = q^j$. Hence $d_j = q^j/(K\eta)$ and $g_{j,i} = q^j/K$ for every i. The K numbers $g_{j,i}$ come out equal; this is a conclusion of the equality case, not a symmetry imposed on the program.

Step 5: the frozen phase, $j \le t \le K-2$. Both $\mathrm{pred}(t, i)$ and $\mathrm{pred}(t+1, i)$ hold with equality, so $g_{t,i} = \eta d_t$ and $g_{t+1,i} = \eta d_{t+1}$, and $\mathrm{cons}(t, i)$ holds with equality, so $\gamma\, g_{t+1,i} = g_{t,i} - d_t = (\eta - 1)d_t$ while also $\gamma\, g_{t+1,i} = (\eta - 1)d_{t+1}$. Since $\eta > 1$ this forces $d_{t+1} = d_t$ and $g_{t+1,i} = g_{t,i}$. With Step 4, $d_t = q^j/(K\eta)$ and $g_{t,i} = q^j/K$ for $j \le t \le K-1$ and every i.

Step 6: the optimal marginals before the switch. For $t < j$ the equality $\mathrm{cons}(t, i)$ reads $g_{t,i} = d_t + \gamma\, g_{t+1,i}$. Going downwards from $g_{j,i} = q^j/K$ and assuming $g_{t+1,i} = q^{t+1}/K$,

$$g_{t,i} = \frac{q^t}{k_1} + \gamma\,\frac{q^{t+1}}{K} = \frac{q^t}{K}\cdot\frac{K + (K-1)(\eta-1)}{k_1} = \frac{q^t}{K},$$

because $K + (K-1)(\eta-1) = k_1$. Downward induction gives $g_{t,i} = q^t/K$ for every $t \le j$ and every i.

Step 7: the last state. The multiplier $\lambda_C(K-1)$ is 0, so Step 2 does not force $\mathrm{cons}(K-1, i)$; the constraint is valid all the same (Lemma 10), and so is $\mathrm{mono}(K-1, i)$. Write $d = d_{K-1}$, $g = g_{K-1,i}$ and $h = g_{K,i}$. In the sharp form of Section 3.2 the two constraints read $d - \frac{g}{\eta} \ge (1 - \frac{1}{\eta})(g - h)$ and $g - h \ge 0$. Step 2 forces $\mathrm{pred}(K-1, i)$ with equality, that is $d = g/\eta$, so the left-hand side of the first display is 0; since $\eta > 1$ the two displays leave $h = g = q^j/K$. Every variable of the run is therefore determined, and the values are those stated. As a check, $\sum_{t<K} d_t = (1 - q^j) + (K - j)q^j/(K\eta) = V_j(\eta)$, and the gain table of Appendix D.4 shows that the instances equation 10 follow exactly this trajectory. □

What the argument delivers and what it leaves open. Every $g_{K,i}$ equals $q^j/K > 0$, while the zeroing convention sets $g_{K,i} = 0$ whenever $o_i \in S^K$; so an attaining run in this range of η selects no optimal element. The conclusion is about the vector $(d_t, g_{t,i})$ only: the pair $(f, \tilde{f})$ is not determined by it, and neither are the true marginals of the candidates the run does not select. In particular the largest true marginal at a state, the quantity that Definition 2 compares the chosen gain against, is not pinned by this argument. At the breakpoints $\eta \in \{2, \dots, K\}$ and at $\eta = 1$ the argument stops at Step 1; an integer $\eta > K$ lies inside the last segment and is covered.

Remark 13 (Breakpoints as an active-constraint switch)**.** *The integers at which the minimizing index changes are the error levels where the multipliers carrying Proposition 7 degenerate: $\lambda_P(j)$ vanishes at the left endpoint of the segment of j (for $j \le K-2$; at $j = K-1$ that endpoint is $\eta = 1$, where the expression cancels to $\frac{1}{K}$) and $\lambda_S(j)$ vanishes at its right endpoint, so at each breakpoint $2, \dots, K$ fewer constraints are forced. The uniqueness does not extend across a breakpoint: at $K = 3$ and $\eta = 2$ the two chosen-gain sequences $(\frac{1}{5}, \frac{2}{15}, \frac{2}{15})$ and $(\frac{1}{5}, \frac{4}{25}, \frac{8}{75})$, from the $j = 1$ and the $j = 2$ instances, both sum to $\frac{7}{15} = \rho_3(2)$.*

Remark 14 (Ties, and what near-equality forces)**.** *In the range of Proposition 7, attaining the exact value forces a predicted-gain tie with an optimal element at every step. The slacks of $\mathrm{pred}(t, i)$ and $\mathrm{cons}(t, i)$ decompose into three nonnegative terms each, equation 12 and equation 13, whose middle term in both cases is $(\tilde{d}_{e_t}(S^t) - \tilde{d}_{o_i}(S^t))/\eta_o$; Steps 2 and 7 put both slacks at zero, so $\tilde{d}_{e_t}(S^t) = \tilde{d}_{o_i}(S^t)$ at every step: the selected element and every optimal element carry the same predicted gain. This is why adversarial tie breaking is a hypothesis of Theorem 1 and not a convenience. The same certificate controls near-equality: a run whose output is $V_j(\eta) + \varepsilon$ has $0 \le \mathrm{slack}_r \le \varepsilon/\lambda_r$ for every constraint r with $\lambda_r > 0$, since the weighted sum of the slacks equals ε. That bound holds one run at a time; no uniform statement over η and K is derived from it here.*

D.7 A SUBMODULAR SURROGATE

Three comparisons place ρ_K among the neighbouring quantities. First, L_K and ρ_K differ by $O(1/K)$ at fixed η, by the expansion of Appendix D.8. Second, the family of Appendix **??**, charged the global error η, realizes only $U_K(\eta) = 1 - q^K$, which is strictly above V_{K-1} for every $\eta > 1$:

$$U_K - V_{K-1} = q^{K-1}\Big(1 - \frac{1}{K\eta}\Big) - q^K = q^{K-1}\Big(\frac{1}{k_1} - \frac{1}{K\eta}\Big) = \frac{q^{K-1}(\eta - 1)}{K\eta k_1},$$

using $1 - q = 1/k_1$ and $K\eta - k_1 = \eta - 1$. Third, the predictor $\tilde{f}$ of the instances equation 10 is monotone, since its gains are at least $1/\eta_u$ times the true gains, but for $\eta > 1$ it is not submodular: by the gain table, the predicted gain of an element of O at the empty set is $1/(K\eta_u)$, while the predicted gain of a second element of O at a singleton of O is η_o/K, which is larger exactly when $\eta > 1$.

Requiring the surrogate to be submodular therefore defines a different ratio. Write $\rho^{\mathrm{sub}}_{n,K}(\eta)$ for the approximation ratio of predictive greedy over the instances on n elements whose predictor is monotone submodular as well. At $\eta = 1$ the band forces $\tilde{f} = f$, so the two models coincide. For $\eta > 1$ the instances equation 10 and the family of Appendix **??** are inadmissible in the submodular-surrogate model, so neither $\rho_K(\eta)$ nor $U_K(\eta)$ transfers to it by the arguments above; the restriction and padding of Remark 9 do transfer, since they preserve submodularity of $\tilde{f}$, so $\rho^{\mathrm{sub}}_{n,K}(\eta)$ is again independent of n for $n \geq 2K$. Since the submodular-surrogate instances form a subclass of all instances, $\rho^{\mathrm{sub}}_{n,K}(\eta) \geq \rho_K(\eta)$. The following family bounds it from above.

Lemma 12 (An explicit family with a submodular surrogate)**.** *Fix $K \geq 2$, $\eta \geq 1$, a split $\eta_u, \eta_o \geq 1$ with $\eta_u\eta_o = \eta$, and an integer $0 \leq m \leq K - 1$. Put $r = 1 - \frac{1}{K}$, $D = 1 + (\eta - 1)r^m$ and $d_t = r^{\min\{t,m\}}/(KD)$ for $0 \leq t \leq K-1$. On the ground set $N = B \sqcup O$ with $B = \{b_0, \ldots, b_{K-1}\}$ and $O = \{o_1, \ldots, o_K\}$ write, for $T \subseteq N$,*

$$b(T) = \sum_{b_t \in T} d_t, \qquad c(T) = \sum_{t<m,\ b_t \in T} d_t, \qquad y(T) = |T \cap O|,$$

and define

$$f(T) = b(T) + \frac{y(T)}{K}\big(1 - c(T)\big), \qquad \tilde{f}(T) = \eta_o\, b(T) + \frac{y(T)}{K}\Big(\frac{\eta_o}{D} - \eta_o\, c(T)\Big). \tag{14}$$

Then f and $\tilde{f}$ are both normalized, monotone and submodular; the smallest admissible error factors of Definition 1 are exactly (η_u, η_o); O is optimal with $f(O) = 1$; and selecting $b_0, \ldots, b_{K-1}$ in this order is a legal run of predictive greedy under adversarial ties, of value

$$W_m(K, \eta) := f(B) = \frac{K - m\, r^m}{K\big(1 + (\eta - 1)r^m\big)}.$$

Consequently $\rho^{\mathrm{sub}}_{n,K}(\eta) \leq \min_{0 \leq m \leq K-1} W_m(K, \eta)$ for every $n \geq 2K$. Moreover, with

$$\beta_m = 1 + \frac{K - m - 1}{K\,(1 - r^{m+1})} \quad (0 \leq m \leq K - 2), \qquad \beta_{-1} = \infty, \qquad \beta_{K-1} = 1,$$

the breakpoints β_m decrease strictly in m, and for $\beta_m \leq \eta \leq \beta_{m-1}$ the minimum of the family is attained at W_m. In particular $\beta_0 = K$ and $W_0(K, \eta) = 1/\eta$, so $\rho^{\mathrm{sub}}_{n,K}(\eta) = \rho_K(\eta) = 1/\eta$ for every $\eta \geq K$ and every $n \geq 2K$.

Proof. Monotonicity and submodularity. Both functions are multilinear polynomials of degree at most two in the indicator variables of the elements:

$$f(T) = \sum_{b_t \in T} d_t + \frac{1}{K} \sum_{o_i \in T} 1 - \frac{1}{K} \sum_{o_i \in T} \sum_{t<m,\ b_t \in T} d_t,$$

and $\tilde{f}$ has the same form with the coefficients $\eta_o d_t$, $\eta_o/(KD)$ and $-\eta_o d_t/K$. For a set function of this form, two distinct elements $e \neq e'$ and a set S containing neither, the second difference $d_e(S \cup \{e'\}) - d_e(S)$ equals the coefficient of the product of the two indicators, so the function is submodular exactly when every such coefficient is nonpositive. Here every second-order coefficient is $-d_t/K \leq 0$ or $-\eta_o d_t/K \leq 0$, so both functions are submodular, and both vanish on $\emptyset$. With $y = y(T)$ and $c = c(T)$ for the current set T, the marginal gains are

$$\begin{aligned}
b_t,\ t < m: \quad & d_{b_t}(T) = d_t\Big(1 - \frac{y}{K}\Big), & & \tilde{d}_{b_t}(T) = \eta_o\, d_t\Big(1 - \frac{y}{K}\Big), \\
b_t,\ t \geq m: \quad & d_{b_t}(T) = d_t, & & \tilde{d}_{b_t}(T) = \eta_o\, d_t, \\
o_i: \quad & d_{o_i}(T) = \frac{1 - c}{K}, & & \tilde{d}_{o_i}(T) = \frac{\eta_o\,(1/D - c)}{K}.
\end{aligned}$$

Since c ranges over $[0, c_{\max}]$ with $c_{\max} = \sum_{t<m} d_t = \frac{1-r^m}{D}$, a geometric sum with ratio r and $1 - r = \frac{1}{K}$, and since $1 - c_{\max} = \frac{\eta r^m}{D} > 0$ and $\frac{1}{D} - c_{\max} = \frac{r^m}{D} > 0$, every marginal gain is nonnegative and both functions are monotone.

The error factors. On a B-element the predicted gain is exactly η_o times the true gain, the two vanishing together when $t < m$ and $y = K$. On an O-element the ratio of predicted to true gain is $\phi(c) = \eta_o \frac{1/D - c}{1-c}$, which is nonincreasing in c because $1/D \le 1$, and so ranges over $[\phi(c_{\max}), \phi(0)] = [\eta_o r^m/(\eta r^m),\ \eta_o/D] = [1/\eta_u,\ \eta_o/D]$, with $\eta_o/D \le \eta_o$. Hence $d_e(T)/\eta_u \le \tilde{d}_e(T) \le \eta_o\, d_e(T)$ on every edge. The upper endpoint is attained on every B-edge of positive gain, and the lower endpoint on an O-edge at a set that contains $b_0, \dots, b_{m-1}$ and at most $K-1$ elements of O, whose true gain $(1 - c_{\max})/K$ is positive (for $m = 0$ one has $c \equiv 0$ and $D = \eta$, and every O-edge has ratio exactly $1/\eta_u$). So no smaller pair of factors is admissible.

The optimum. $f(O) = \frac{K}{K}(1 - 0) = 1$. For $|T| \le K$, since $d_t \le d_0 = \frac{1}{KD} \le \frac{1}{K}$ and $c(T) \ge 0$, $f(T) \le \frac{|T \cap B|}{K} + \frac{|T \cap O|}{K} \le 1$. So O is optimal and the normalization $f(O^*) = 1$ holds.

The run. Let $S_t = \{b_0, \dots, b_{t-1}\}$, so $y(S_t) = 0$. For $t \le m$, $c(S_t) = \sum_{s<t} d_s = \frac{1-r^t}{D}$ and $\tilde{d}_{o_i}(S_t) = \frac{\eta_o}{K}\big(\frac{1}{D} - \frac{1-r^t}{D}\big) = \frac{\eta_o r^t}{KD} = \eta_o d_t$; for $t > m$, $c(S_t) = c_{\max}$ and $\tilde{d}_{o_i}(S_t) = \frac{\eta_o r^m}{KD} = \eta_o d_t$. The unselected B-elements b_s, $s \ge t$, have predicted gain $\eta_o d_s \le \eta_o d_t$, since d_s is nonincreasing in s. So at every step b_t is a predicted maximizer, tied with every o_i and possibly with later B-elements, and adversarial tie breaking may select it. The output is B, of value

$$f(B) = \sum_{t<K} d_t = \frac{1}{KD}\Big[K(1 - r^m) + (K - m)r^m\Big] = \frac{K - m\, r^m}{KD} = W_m(K, \eta).$$

Ground-set size. The instance has $2K$ elements, so $\rho^{\mathrm{sub}}_{2K,K}(\eta) \le W_m(K, \eta)$ for every m, and the padding of Remark 9 preserves submodularity of $\tilde{f}$, so the bound holds for every $n \ge 2K$.

The lower envelope. Write $D_m = 1 + (\eta - 1)r^m$, so that $W_m = \frac{K - m r^m}{K D_m}$ and

$$W_m - W_{m+1} = \frac{(K - m r^m)D_{m+1} - \big(K - (m+1)r^{m+1}\big)D_m}{K\, D_m D_{m+1}}.$$

Expanding the numerator and using $r^{m+1} - r^m = -r^m/K$ and $(m+1)r - m = 1 - \frac{m+1}{K}$,

$$(K - m r^m)D_{m+1} - \big(K - (m+1)r^{m+1}\big)D_m = r^m\Big[\frac{K - m - 1}{K} - (\eta - 1)\big(1 - r^{m+1}\big)\Big].$$

The denominators are positive, so for $0 \le m \le K-2$ one has $W_m \ge W_{m+1}$ exactly when $\eta \le \beta_m$. The numerator $K - m - 1$ of $\beta_m - 1$ decreases in m and its denominator $K(1 - r^{m+1})$ increases, so β_m decreases strictly in m; moreover $\beta_0 = 1 + \frac{K-1}{K(1-r)} = K$ and $\beta_{K-2} = 1 + \frac{1}{K(1-r^{K-1})} > 1$. If $\beta_m \le \eta \le \beta_{m-1}$, then $\eta \le \beta_{m'}$ for every $m' < m$, which gives $W_0 \ge W_1 \ge \cdots \ge W_m$, and $\eta \ge \beta_{m'}$ for every $m' \ge m$, which gives $W_m \le W_{m+1} \le \cdots \le W_{K-1}$; so W_m is the minimum of the family. Finally, for $m = 0$ one has $c \equiv 0$, $D = \eta$ and $W_0 = \frac{K}{K\eta} = \frac{1}{\eta}$, which is the minimum of the family for $\eta \ge \beta_0 = K$; hence $\rho^{\mathrm{sub}}_{n,K}(\eta) \le 1/\eta$ there. In the other direction $\rho^{\mathrm{sub}}_{n,K}(\eta) \ge \rho_{n,K}(\eta) = \rho_K(\eta)$ by Remark 9, and $\rho_K(\eta) = 1/\eta$ for $\eta \ge K$ by Theorem 1, the minimizing index being $j = 0$. The two ratios therefore coincide for every $\eta \ge K$. □

The value of $\rho^{\mathrm{sub}}_{n,K}(\eta)$ for $1 < \eta < K$ is not determined here; at $K = 3$ and $\eta = 3/2$ the family gives $\min_m W_m = W_2 = 19/33$, against $\rho_3(3/2) = 9/16$.

D.8 Asymptotics (Corollary 1)

Fix $\eta \ge 1$ and write $m = \lfloor \eta \rfloor \ge 1$. Two elementary bounds are used: for $0 < u \le \frac{1}{2}$,

$$0 \ \le\ -\ln(1 - u) - u \ =\ \sum_{i \ge 2} \frac{u^i}{i} \ \le\ \frac{u^2}{2(1-u)} \ \le\ u^2, \tag{15}$$

and for $0 < w < 1$,

$$\ln(1-w) + \frac{w}{1-w} = \sum_{i\geq 1} w^i\Big(1-\frac{1}{i}\Big) \;>\; 0. \tag{16}$$

The sequence L_K. Let $x = 1/\eta \in (0,1]$ and let $K > x$ be real. With $w = x/K$,

$$\frac{\partial}{\partial K}\, K\ln\Big(1-\frac{x}{K}\Big) = \ln\Big(1-\frac{x}{K}\Big) + \frac{x}{K-x} = \ln(1-w) + \frac{w}{1-w} \;>\; 0$$

by equation 16, so $(1-\frac{x}{K})^K$ strictly increases in K and $L_K(\eta)$ strictly decreases in K. By equation 15 with $u = \frac{1}{\eta K}$, $\big|K\ln(1-\frac{1}{\eta K}) + \frac{1}{\eta}\big| \leq \frac{1}{\eta^2 K} \to 0$, so $L_K(\eta) \to 1 - e^{-1/\eta}$ from above.

The plateau. If $2 \leq K \leq m$ then $\eta \geq m \geq K$, so Theorem 1 gives $\rho_K(\eta) = 1/\eta$.

The limit of ρ_K. Let $K \geq m$. The index $j = K - m$ satisfies $0 \leq j \leq K-1$ and $\eta \in [K-j, K-j+1] = [m, m+1]$, so it is a minimizing index and $\rho_K(\eta) = V_{K-m}(\eta) = 1 - q^{\,K-m}\big(1-\frac{m}{K\eta}\big)$ with $q = 1 - \frac{1}{k_1}$. Since $k_1 \geq K$, equation 15 with $u = \frac{1}{k_1}$ gives $\big|(K-m)\ln(1-\frac{1}{k_1}) + \frac{K-m}{k_1}\big| \leq \frac{K-m}{k_1^2} \to 0$ and $\frac{K-m}{k_1} \to \frac{1}{\eta}$. Hence $q^{\,K-m} \to e^{-1/\eta}$, while $1 - \frac{m}{K\eta} \to 1$, so $\rho_K(\eta) \to 1 - e^{-1/\eta}$.

The first step off the plateau. Suppose $m \geq 2$ and put $K = m+1$, so that $\eta < K$ and the minimizing index is $j = 1$. By equation 9 with $i = 0$,

$$\rho_{m+1}(\eta) = V_1(\eta) = V_0(\eta) - \frac{K-\eta}{K\eta k_1} = \frac{1}{\eta} - \frac{m+1-\eta}{(m+1)\eta k_1} < \frac{1}{\eta} = \rho_m(\eta),$$

the drop being strictly positive because $\eta < m+1$.

Strict decrease past the plateau. Let $K \geq m+1$ be real and put

$$P(K) = q^{\,K-m}\Big(1-\frac{m}{K\eta}\Big), \qquad x = (K-1)\eta + 1, \qquad q = 1 - \frac{1}{x},$$

so that $\rho_K = 1 - P(K)$ at every integer $K \geq m$. Here $x > 1$ and $K\eta \geq (m+1)m > m$, so $P(K) > 0$. Differentiating $\ln P = (K-m)\ln((x-1)/x) + \ln((K\eta - m)/(K\eta))$ and using $dx/dK = \eta$ together with $x - 1 = (K-1)\eta$,

$$\frac{\partial}{\partial K}\ln\frac{x-1}{x} = \eta\Big(\frac{1}{x-1} - \frac{1}{x}\Big) = \frac{\eta}{x(x-1)} = \frac{1}{(K-1)x},$$
$$\frac{\partial}{\partial K}\ln\frac{K\eta - m}{K\eta} = \frac{\eta}{K\eta - m} - \frac{1}{K} = \frac{m}{K(K\eta - m)},$$

so that

$$\frac{\partial \ln P}{\partial K} = \ln\Big(1-\frac{1}{x}\Big) + \frac{K-m}{(K-1)x} + \frac{m}{K(K\eta - m)}. \tag{17}$$

Bounding the logarithm termwise,

$$-\ln\Big(1-\frac{1}{x}\Big) = \sum_{i\geq 1}\frac{1}{ix^i} \;\leq\; \frac{1}{x} + \frac{1}{2}\sum_{i\geq 2}\frac{1}{x^i} = \frac{1}{x} + \frac{1}{2x(x-1)},$$

so, multiplying equation 17 by $x(K-1) > 0$ and using $\frac{K-1}{2(x-1)} = \frac{1}{2\eta}$,

$$x(K-1)\,\frac{\partial \ln P}{\partial K} \;\geq\; (K-m) - (K-1) + L - \frac{1}{2\eta} \;=\; L - (m-1) - \frac{1}{2\eta}, \qquad L := \frac{m(K-1)x}{K(K\eta - m)}.$$

Two observations finish the estimate. First, $x - (K\eta - m) = (K-1)\eta + 1 - K\eta + m = m + 1 - \eta > 0$, so $\frac{x}{K\eta - m} > 1$ and

$$L = \frac{m(K-1)}{K}\cdot\frac{x}{K\eta - m} \;>\; m - \frac{m}{K}.$$

Second, $K \geq m+1$ and $\eta \geq m$ give

$$\frac{m}{K} + \frac{1}{2\eta} \;\leq\; \frac{m}{m+1} + \frac{1}{2m} \;=\; 1 - \frac{1}{m+1} + \frac{1}{2m} \;\leq\; 1,$$

the last step because $2m \geq m+1$ for $m \geq 1$; equivalently $m - \frac{m}{K} \geq m - 1 + \frac{1}{2\eta}$. Combining the two, $L > m - 1 + \frac{1}{2\eta}$ and $\partial \ln P / \partial K > 0$ on $[m+1, \infty)$. So P strictly increases there and $\rho_K > \rho_{K+1}$ for every integer $K \geq m+1$.

Conclusion. For $2 \leq K \leq m$ the value is the constant $1/\eta$; when $m \geq 2$ the step from $K = m$ to $K = m+1$ is a strict drop; and from $K = m+1$ on the sequence strictly decreases. Hence ρ_K equals $1/\eta$ for $K \leq \lfloor \eta \rfloor$ and decreases strictly from $K = \lfloor \eta \rfloor$ on. A strictly decreasing sequence stays above its limit, so $\rho_K > 1 - e^{-1/\eta}$ for every $K \geq \max\{2, m\}$, and on the plateau $\rho_K = 1/\eta > 1 - e^{-1/\eta}$ because $1 - e^{-y} < y$ for $y > 0$. The limit is therefore approached from above.

The $1/K$ expansion. Fix $\eta \geq 1$, write $m = \lfloor \eta \rfloor$, and expand the active branch V_{K-m} as $K \to \infty$. From $k_1 = K\eta - (\eta - 1)$,

$$\frac{1}{k_1} = \frac{1}{K\eta} \cdot \frac{1}{1 - \frac{\eta - 1}{K\eta}} = \frac{1}{K\eta} + \frac{\eta - 1}{K^2\eta^2} + O_\eta\Big(\frac{1}{K^3}\Big),$$

so, using $\ln(1-u) = -u - \frac{u^2}{2} + O(u^3)$ with $u = \frac{1}{k_1}$ and multiplying by $K - m$,

$$\ln q^{K-m} = -\frac{1}{\eta} + \frac{A}{K} + O_\eta\Big(\frac{1}{K^2}\Big), \qquad A = \frac{m\eta - \eta + \frac{1}{2}}{\eta^2}.$$

Exponentiating and multiplying by the second factor of V_{K-m}, which is $1 - \frac{m}{K\eta}$ exactly,

$$q^{K-m}\Big(1 - \frac{m}{K\eta}\Big) = e^{-1/\eta}\Big(1 + \frac{1}{K}\Big(A - \frac{m}{\eta}\Big)\Big) + O_\eta\Big(\frac{1}{K^2}\Big), \qquad A - \frac{m}{\eta} = -\frac{2\eta - 1}{2\eta^2},$$

the coefficient m cancelling. Hence

$$\rho_K(\eta) = 1 - e^{-1/\eta} + \frac{c(\eta)}{K} + O_\eta\Big(\frac{1}{K^2}\Big), \qquad c(\eta) = \frac{e^{-1/\eta}(2\eta - 1)}{2\eta^2},$$

and the $1/K$ coefficient does not depend on m, so at an integer $\eta \geq 2$, where two adjacent branches are active, both give the same coefficient. The same expansion for L_K, where $K \ln(1 - \frac{1}{\eta K}) = -\frac{1}{\eta} - \frac{1}{2\eta^2 K} + O_\eta(K^{-2})$, gives the coefficient $c_L(\eta) = \frac{e^{-1/\eta}}{2\eta^2}$, and for $U_K = 1 - q^K$, where $K \ln(1 - \frac{1}{k_1}) = -\frac{1}{\eta} - \frac{2\eta - 1}{2\eta^2 K} + O_\eta(K^{-2})$, the coefficient is $c(\eta)$ itself, so at fixed η

$$\rho_K(\eta) - L_K(\eta) = \frac{e^{-1/\eta}(\eta - 1)}{\eta^2 K} + O_\eta\Big(\frac{1}{K^2}\Big), \qquad U_K(\eta) - \rho_K(\eta) = O_\eta\Big(\frac{1}{K^2}\Big).$$

The remainders are taken at a fixed error level; no uniformity in η is asserted.

The $1/K^2$ coefficient of $U_K - \rho_K$. Fix $\eta > 1$, write $m = \lfloor \eta \rfloor \geq 1$ and $\varepsilon = 1/K$. The quantity $1/k_1 = \varepsilon/(\eta - (\eta - 1)\varepsilon)$ is analytic in ε at 0 and vanishes there, so $\ln q$ has a Taylor expansion

$$\ln q = -b_1\varepsilon - b_2\varepsilon^2 - b_3\varepsilon^3 + O_\eta(\varepsilon^4), \qquad b_1 = \frac{1}{\eta}, \qquad b_2 = \frac{\eta - 1}{\eta^2} + \frac{1}{2\eta^2} = \frac{2\eta - 1}{2\eta^2},$$

the first two coefficients being those computed above; the value of b_3 is not needed. Multiplying by $K - m = \varepsilon^{-1} - m$ and by $K = \varepsilon^{-1}$,

$$(K-m)\ln q = -b_1 + (mb_1 - b_2)\varepsilon + (mb_2 - b_3)\varepsilon^2 + O_\eta(\varepsilon^3), \qquad K \ln q = -b_1 - b_2\varepsilon - b_3\varepsilon^2 + O_\eta(\varepsilon^3).$$

Exponentiating with $e^z = 1 + z + \frac{z^2}{2} + O(z^3)$ applied to the $O(\varepsilon)$ parts of the exponents,

$$q^{K-m} = e^{-1/\eta}\Big[1 + (mb_1 - b_2)\varepsilon + \Big(mb_2 - b_3 + \frac{(mb_1 - b_2)^2}{2}\Big)\varepsilon^2\Big] + O_\eta(\varepsilon^3),$$

$$q^K = e^{-1/\eta}\Big[1 - b_2\varepsilon + \Big(-b_3 + \frac{b_2^2}{2}\Big)\varepsilon^2\Big] + O_\eta(\varepsilon^3).$$

Multiplying the first expansion by the exact factor $1 - \frac{m}{K\eta} = 1 - mb_1\varepsilon$ and using $(mb_1 - b_2) - mb_1 = -b_2$ and $\frac{(mb_1 - b_2)^2}{2} - mb_1(mb_1 - b_2) = -\frac{m^2 b_1^2}{2} + \frac{b_2^2}{2}$,

$$q^{K-m}\Big(1 - \frac{m}{K\eta}\Big) = e^{-1/\eta}\Big[1 - b_2\varepsilon + \Big(mb_2 - b_3 - \frac{m^2 b_1^2}{2} + \frac{b_2^2}{2}\Big)\varepsilon^2\Big] + O_\eta(\varepsilon^3).$$

Subtracting the expansion of q^K, the constant terms, the ε terms and the terms b_3 and $\frac{b_2^2}{2}$ cancel, and

$$U_K(\eta) - \rho_K(\eta) = q^{K-m}\Big(1 - \frac{m}{K\eta}\Big) - q^K = \frac{c'(\eta)}{K^2} + O_\eta\Big(\frac{1}{K^3}\Big),$$
$$c'(\eta) = e^{-1/\eta}\,\frac{\lfloor\eta\rfloor\,\big(2\eta - \lfloor\eta\rfloor - 1\big)}{2\eta^2}.$$

Since $1 \le m \le \eta$, we have $2\eta - m - 1 \ge \eta - 1 > 0$, so $c'(\eta) > 0$ for every $\eta > 1$: at a fixed error level the exact value improves on U_K by a strictly positive multiple of $1/K^2$. At an integer $\eta \ge 2$ the two adjacent branches $m = \eta$ and $m = \eta - 1$ both give $m(2\eta - m - 1) = \eta(\eta - 1)$, so c' is continuous in η although each of its factors is piecewise. As before, the remainder is taken at a fixed error level.

E Proofs for Section 3.3: the price of the budget

This appendix proves Theorem 2 (Appendix E.2), Theorem 3 (Appendix E.3) and the statement on queries of arbitrary size (Appendix E.4). All three rest on one mechanism: a surrogate whose answers on the sets a bounded algorithm can query do not depend on where the optimum is, so that the algorithm's transcript is fixed before the optimum is placed. The two lemmas that carry the mechanism are recorded first.

E.1 Two lemmas on transcripts

Lemma 13 (Canonical transcript)**.** *Let A be a deterministic algorithm that queries the surrogate on sets from a family $\mathcal{S}$ and returns a set of size at most K, the number Q of queries being arbitrary. Let κ be a set function on $\mathcal{S}$, the canonical oracle, and let $\tilde{f}$ be a surrogate. Run A against κ and let $S_1, \dots, S_Q$ be the queries and T_0 the output it produces; these are fixed sets, determined by A and κ alone. If $\tilde{f}(S_i) = \kappa(S_i)$ for every $i = 1, \dots, Q$, then the run of A against $\tilde{f}$ issues exactly the queries $S_1, \dots, S_Q$ in the same order, receives exactly the same answers, and returns T_0.*

Proof. Induct on i. Suppose the first $i-1$ queries of the run against $\tilde{f}$ are $S_1, \dots, S_{i-1}$ and the first $i-1$ answers agree with the canonical ones. A deterministic algorithm's ith query is a function of the answers received so far, so it is again S_i, and by hypothesis $\tilde{f}(S_i) = \kappa(S_i)$, so the ith answer agrees as well. The base case $i = 1$ holds because the first query is a constant of A. After Q steps all answers agree, so the output, again a function of the answers, is T_0. □

The order of the two steps matters and is used in every proof below: the transcript is taken against the canonical oracle, which does not involve the hidden optimum O, and only then is O placed so that $\tilde{f}$ agrees with κ on that transcript. An argument that first fixes O and then inspects the queries the algorithm issues against $\tilde{f}$ would not make those queries independent of O.

Lemma 14 (Counting)**.** *Let $1 \le \tau < K \le n$ be integers, fix $S \subseteq N$ with $|S| \le K$, and let O be a uniformly random K-subset of the n-element ground set N. Then*

$$\Pr\big[|S \cap O| > \tau\big] \;\le\; \binom{K}{\tau+1}\Big(\frac{K}{n}\Big)^{\tau+1} \;\le\; \frac{K^{2\tau+2}}{(\tau+1)!\,n^{\tau+1}}. \tag{18}$$

Proof. The event $\{|S \cap O| > \tau\}$ is the union, over the $(\tau+1)$-subsets $R \subseteq S$, of the events $\{R \subseteq O\}$, and

$$\Pr[R \subseteq O] = \prod_{i=0}^{\tau} \frac{K-i}{n-i} \;\le\; \Big(\frac{K}{n}\Big)^{\tau+1},$$

since $(K-i)n \le K(n-i)$ for $n \ge K$. There are at most $\binom{|S|}{\tau+1} \le \binom{K}{\tau+1}$ such subsets R, so a union bound gives the first inequality, and $\binom{K}{\tau+1} \le K^{\tau+1}/(\tau+1)!$ gives the second. The integrality of τ enters at this step, which is the reason for taking $\tau = \lceil c \rceil + 1$ rather than $\tau = c + 1$ at real c in Theorem 3. □

E.2 Exact optimality within the greedy budget (Theorem 2)

The proof rests on four lemmas. The first builds the hard family, a pair of truncated residual sequences on a two-class count grid; the second shows that its predictor is a function of $|S|$ alone on precisely the sets a transcript of $\mathcal{A}_{\mathrm{lin}}$ can meet; the third computes its value; the fourth places the hidden optimum inside the ground set so that no query of a fixed transcript sees more than one of its elements.

Lemma 15 (The double-residual family)**.** *Fix $K \geq 2$, $\eta > 1$, $n \geq 2K$ and a split $\eta_u, \eta_o > 0$ with $\eta_u \eta_o = \eta$. Put $k_1 = (K-1)\eta + 1$, $q = (K-1)\eta/k_1$, $j = \max\{0, K - \lfloor \eta \rfloor\}$, $Q = q^j$, $\delta = Q/(K\eta)$, $C = k_1/K$ and $[z]_+ = \max\{z, 0\}$, and define two residual sequences on the integers $x \geq 0$ by*

$$r_x = \begin{cases} q^x, & 0 \leq x \leq j, \\ \big[Q - (x-j)\delta\big]_+, & x \geq j, \end{cases} \qquad h_x = \begin{cases} \dfrac{K-1}{K}\, q^x, & 0 \leq x \leq j, \\ \Big[\dfrac{K-1}{K}\, Q - (x-j)\delta\Big]_+, & x \geq j, \end{cases} \tag{19}$$

the two branches agreeing at $x = j$. Let the ground set be $N = B \sqcup O$ with $|O| = K$ and $|B| = n - K$; for $S \subseteq N$ write $x = |S \cap B|$, $y = |S \cap O|$ and $c_y = (K-y)/(K-1)$, and set $f(S) = F(x, y)$ and $\tilde{f}(S) = H(x, y)/\eta_u$, where

$$F(x,y) = \begin{cases} 1 - r_x, & y = 0, \\ 1 - c_y\, h_x, & 1 \leq y \leq K, \end{cases} \qquad H(x,y) = \begin{cases} C - r_x - (\eta - 1)h_x, & y = 0, \\ C - \eta\, c_y\, h_x, & 1 \leq y \leq K. \end{cases} \tag{20}$$

Then f is monotone, submodular and normalized, $0 \leq f \leq 1$, $f(O) = 1$ and O maximizes f among the sets of size at most K; and $\tilde{f}$ is normalized and satisfies, on every edge of the Boolean lattice, $\Delta f/\eta_u \leq \Delta \tilde{f} \leq \eta_o \Delta f$, both endpoints being attained on edges of positive gain, so the smallest admissible error factors of Definition 1 are exactly (η_u, η_o). Neither F nor H involves n: every $n \geq 2K$ restricts the same table to its available count grid.

Proof. Write $p_x = r_x - r_{x+1}$ and $s_x = h_x - h_{x+1}$. Three groups of one-dimensional facts carry the whole argument.

(i) $0 \leq s_x \leq p_x$, both sequences are nonincreasing, and their steps are nonincreasing. On the geometric phase, $x + 1 \leq j$, we have $p_x = q^x(1-q) = q^x/k_1$ and $s_x = \frac{K-1}{K} p_x$, so $0 \leq s_x \leq p_x$ and both steps decrease geometrically in x. At the junction $x = j - 1$, using $qk_1 = (K-1)\eta$,

$$p_{j-1} = \frac{q^{j-1}}{k_1} = \frac{Q}{(K-1)\eta}, \qquad s_{j-1} = \frac{K-1}{K}\, p_{j-1} = \frac{Q}{K\eta} = \delta.$$

On the linear phase, and as long as neither residual is truncated, $p_x = s_x = \delta$, and $\delta \leq Q/((K-1)\eta) = p_{j-1}$: the steps of r drop at the junction and those of h stay constant there. Since $\frac{K-1}{K} Q \leq Q$, the sequence h reaches 0 no later than r; at the step where h vanishes its step is $h_x \leq \delta$ and the step of r is δ or the remaining value r_x, and afterwards $s_x = 0$. At the step where r vanishes, $p_x = r_x \leq \delta$ and $s_x = 0$. Hence both step sequences are nonnegative and nonincreasing throughout and $s_x \leq p_x$ everywhere.

(ii) $r_x - h_x \geq h_x/(K-1) \geq 0$, with equality in the first inequality for $x \leq j$, and $r_x - h_x$ is nonincreasing in x. For $x \leq j$ both sides of the first inequality equal q^x/K. For $x > j$ and before any truncation, $r_x - h_x = Q/K$ while $h_x/(K-1) = Q/K - (x-j)\delta/(K-1) \leq Q/K$. Once $h_x = 0$ the inequality reads $r_x \geq 0$. For monotonicity, $r_x - h_x = q^x/K$ decreases on $x \leq j$, equals Q/K on the untruncated part of the linear phase and, from the first index x with $h_x = 0$ on, equals r_x, which satisfies $r_x = Q - (x-j)\delta \leq Q - \frac{K-1}{K} Q = Q/K$ there.

(iii) No truncation occurs before $x = K$. Since $K - j = \min\{K, \lfloor \eta \rfloor\} \leq \eta$ we get

$$(K-j)\,\delta = \frac{(K-j)\,Q}{K\eta} \;\leq\; \frac{Q}{K} \;\leq\; \frac{(K-1)\,Q}{K},$$

whence $r_K = Q - (K-j)\delta \geq Q/K > 0$ and $h_K = \frac{K-1}{K} Q - (K-j)\delta \geq \frac{K-2}{K} Q \geq 0$.

The first differences of F on the count grid are

$$\Delta_x F(x,0) = p_x, \qquad \Delta_x F(x,y) = c_y\, s_x \quad (1 \le y \le K),$$
$$\Delta_y F(x,0) = r_x - h_x, \qquad \Delta_y F(x,y) = \frac{h_x}{K-1} \quad (1 \le y < K),$$

all nonnegative by (i) and (ii). Each is nonincreasing in x: this is (i) for the first two and (ii) together with the monotonicity of h for the last two. Each is nonincreasing in y: passing from $y = 0$ to $y = 1$ replaces p_x by $c_1 s_x = s_x \le p_x$ in the first row and $r_x - h_x$ by $h_x/(K-1)$ in the second, which is (ii), and c_y decreases in y. So the count grid has nonnegative, coordinatewise nonincreasing first differences, and Lemma 5 lifts it to a monotone submodular set function f with $f(\emptyset) = F(0,0) = 0$. Moreover $F(0,K) = 1$ and $0 \le F \le 1$ on the whole grid, so $f(O) = 1$ and O maximizes f under the cardinality constraint.

On the predictor side the four edge classes carry

$$\begin{aligned}
B\text{-edge}, y = 0: &\quad \Delta F = p_x, && \Delta H = p_x + (\eta-1)s_x,\\
B\text{-edge}, 1 \le y \le K: &\quad \Delta F = c_y s_x, && \Delta H = \eta\, c_y s_x,\\
O\text{-edge}, y = 0: &\quad \Delta F = r_x - h_x, && \Delta H = r_x - h_x,\\
O\text{-edge}, 1 \le y < K: &\quad \Delta F = \frac{h_x}{K-1}, && \Delta H = \frac{\eta\, h_x}{K-1}.
\end{aligned} \tag{21}$$

In the first row $0 \le s_x \le p_x$ gives $\Delta F \le \Delta H \le \eta\, \Delta F$; the other rows realize the ratio exactly 1 or exactly η. With $\tilde f = H/\eta_u$ this is $\Delta f/\eta_u \le \Delta \tilde f \le \eta_o\, \Delta f$ on every edge, and zero-gain edges have $\Delta H = 0$. Finally

$$H(0,0) = C - 1 - (\eta-1)\frac{K-1}{K} = \frac{k_1 - K - (\eta-1)(K-1)}{K} = 0,$$

so $\tilde f(\emptyset) = 0$. At $x = 0$ the third row has gain $r_0 - h_0 = 1/K > 0$ and ratio 1, that is $\Delta \tilde f = \Delta f/\eta_u$, and the fourth row has gain $h_0/(K-1) = 1/K > 0$ and ratio η, that is $\Delta \tilde f = \eta_o\, \Delta f$: both endpoints of the band are attained on edges of positive gain, so no smaller pair of factors is admissible. □

Lemma 16 (Small-set indistinguishability and the rigid trajectory)**.** *In the family of Lemma 15 define*

$$\widehat H_s = \begin{cases} C\,(1-q^s), & 0 \le s \le j,\\ C\,(1-Q) + \eta\,(s-j)\,\delta, & j \le s \le K. \end{cases} \tag{22}$$

Then:

1. *$H(x,0) = \widehat H_x$ for $0 \le x \le K$ and $H(x,1) = \widehat H_{x+1}$ for $x+1 \le K$; consequently $\tilde f(S) = \widehat H_{|S|}/\eta_u$ whenever $|S| \le K$ and $|S \cap O| \le 1$, so on every set that a transcript with queries of size at most K and at most one optimal element can meet, the predictor depends on $|S|$ alone.*
2. *On larger sets the predictor is not a function of $|S|$: at $K = 3$, $\eta = 3/2$ the two six-element sets with $(x,y) = (6,0)$ and $(5,1)$ receive $H = 61/48$ and $H = 4/3$.*
3. *At every state $(t,0)$ with $t < K$ the predicted marginals of all remaining elements equal $q^{\min(t,j)}/(K\eta_u)$, so adversarial tie breaking keeps predictive greedy inside B for all K steps; the true selected gains are q^t/k_1 for $t < j$ and $q^j/(K\eta)$ for $j \le t < K$, and every element of O still outside the run keeps true marginal $q^{\min(t,j)}/K$, so the trajectory of Proposition 7 is realized exactly.*

Proof. (1) For $0 \le x \le j$, using $K + (\eta-1)(K-1) = k_1$,

$$H(x,0) = C - q^x\Big(1 + (\eta-1)\frac{K-1}{K}\Big) = C - q^x\frac{k_1}{K} = C\,(1-q^x).$$

For $j \le x \le K$ no truncation has occurred, by (iii) in the proof of Lemma 15, and the same grouping gives

$$H(x,0) = C - \big(Q - (x-j)\delta\big) - (\eta-1)\frac{K-1}{K}\big(Q - (x-j)\delta\big) = C\,(1-Q) + \eta\,(x-j)\,\delta.$$

For the second identity, $c_1 = 1$, so $H(x,1) = C - \eta h_x$. With $Cq = \eta(K-1)/K$ one gets, for $x+1 \le j$, $H(x,1) = C - Cq\,q^x = C(1-q^{x+1}) = \widehat{H}_{x+1}$, and for $x \ge j$,

$$H(x,1) = C - \eta\frac{K-1}{K}Q + \eta(x-j)\delta = C - CqQ + \eta(x-j)\delta = C(1-Q) + \eta(x+1-j)\delta,$$

because $CQ(1-q) = Q/K = \eta\delta$. A set S with $|S| \le K$ and $|S \cap O| \le 1$ has $(x,y) = (|S|,0)$ or $(|S|-1,1)$, and in both cases $H(x,y) = \widehat{H}_{|S|}$.

(2) At $K = 3$ and $\eta = 3/2$ one has $k_1 = 4$, $q = 3/4$, $j = 2$, $Q = 9/16$, $\delta = 1/8$ and $C = 4/3$, so $r_6 = Q - 4\delta = 1/16$ and $h_5 = h_6 = [\frac{3}{8} - \frac{3}{8}]_+ = 0$; hence $H(6,0) = \frac{4}{3} - \frac{1}{16} = \frac{61}{48}$ while $H(5,1) = \frac{4}{3} - \eta h_5 = \frac{4}{3}$.

(3) By (1), the predicted value at a state $(t,0)$ with $t < K$ is $\widehat{H}_t/\eta_u$, and adding any element, of B or of O, produces a set of size $t+1 \le K$ meeting O in at most one element, of predicted value $\widehat{H}_{t+1}/\eta_u$. The common marginal is therefore $(\widehat{H}_{t+1} - \widehat{H}_t)/\eta_u$, which equals $Cq^t(1-q) = q^t/K$ divided by η_u for $t < j$ and $\eta\delta = Q/K = q^j/K$ divided by η_u for $t \ge j$, that is $q^{\min(t,j)}/(K\eta_u)$ in both cases. All predicted marginals tie, so adversarial tie breaking may select an element of B at every step. The true gain of that step is p_t, which is q^t/k_1 for $t < j$ and $\delta = q^j/(K\eta)$ for $j \le t < K$, and the true marginal of an element of O at the state $(t,0)$ is $r_t - h_t = q^{\min(t,j)}/K$. □

Lemma 17 (The value of the family)**.** *In the family of Lemma 15, for every $n \ge 2K$ and every K-set $T \subseteq B$,*

$$\frac{f(T)}{f(O)} = F(K,0) = 1 - Q + (K-j)\,\delta = V_j(\eta) = \rho_K(\eta). \tag{23}$$

Proof. By (iii) in the proof of Lemma 15 no truncation occurs before $x = K$, so $r_K = Q - (K-j)\delta$ and $F(K,0) = 1 - r_K = 1 - Q + (K-j)\delta$. Since $Q = q^j$ and $\delta = Q/(K\eta)$ this is $1 - q^j\big(1 - \frac{K-j}{K\eta}\big) = V_j(\eta)$. For the last equality, the adjacent-branch difference equation 9 shows that V_i is minimized over $0 \le i \le K-1$ at $i = \max\{0, K - \lfloor\eta\rfloor\} = j$, with ties exactly at integer η, and Theorem 1 gives $\rho_K(\eta) = \min_i V_i(\eta)$. Finally $f(O) = 1$ by Lemma 15. The value is exact at each $n \ge 2K$ and no limit is taken. □

Lemma 18 (An independent K-set in the query graph)**.** *Put $B_K = K^2(K-1)^2 + 3K - 1$.*

1. *Every graph on m vertices with e edges has an independent set of size at least*

$$\sum_v \frac{1}{\deg(v)+1} \;\ge\; \frac{m^2}{2e+m}. \tag{24}$$

2. *Let $K \ge 2$ and $n \ge B_K$, let N be a ground set of n elements, let $S_1, \dots, S_M$ with $M \le nK$ be subsets of N each of size at most K, and let $T \subseteq N$ with $|T| = K$. Then there is a K-set $O \subseteq N \setminus T$ with $|O \cap S_i| \le 1$ for every i.*

Every $n \ge K^4$ qualifies, because $K^4 - B_K = K(2K-3)(K+1) + 1 > 0$.

Proof. (1) In a uniformly random ordering of the vertices, the vertices that precede all their neighbours form an independent set, and v belongs to it with probability $1/(\deg(v)+1)$, so the expected size of that independent set is $\sum_v 1/(\deg(v)+1)$ and some ordering attains at least this value; this is the Caro–Wei bound (**?**). By Cauchy–Schwarz, $\sum_v \frac{1}{\deg(v)+1} \cdot \sum_v(\deg(v)+1) \ge m^2$, and $\sum_v(\deg(v)+1) = 2e + m$, which gives equation 38.

(2) On the vertex set $N \setminus T$, of size $m = n - K$, join two elements by an edge whenever some S_i contains both. Each S_i contributes at most $\binom{K}{2}$ edges, so $e \le nK\binom{K}{2}$ and $2e \le nK^2(K-1)$. By (1) there is an independent set of size at least $m^2/(2e+m)$, and this exceeds $K-1$ as soon as $m^2 > (K-1)(2e+m)$, for which it suffices that

$$(n-K)^2 > (K-1)\big(nK^2(K-1) + n - K\big), \qquad \text{that is} \qquad (n-K)(n-2K+1) > nK^2(K-1)^2,$$

where we used $(n-K)^2-(K-1)(n-K)=(n-K)(n-2K+1)$. The difference of the two sides is

$$(n-K)(n-2K+1)-nK^2(K-1)^2=n\,(n-B_K)+K(2K-1), \tag{25}$$

which is positive for every $n \geq B_K$; at $n=B_K-1$ it equals $(K-1)^2(2-K^2)<0$, so B_K is exactly where this estimate starts to work. Hence for $n \geq B_K$ the graph has an independent set of size at least K, and any K-subset O of it lies in $N \setminus T$ and meets every S_i in at most one element. Finally $K^4-B_K=K^2\big(K^2-(K-1)^2\big)-3K+1=2K^3-K^2-3K+1=K(2K-3)(K+1)+1>0$ for $K \geq 2$, so $n \geq K^4$ implies $n \geq B_K$. □

Proof of Theorem 2. Fix $K \geq 2$, $\eta>1$, $n \geq K^4$, a split $\eta_u\eta_o=\eta$ and an algorithm $A \in \mathcal{A}_{\text{lin}}$. Let N be a ground set of n elements.

The canonical run. Run A against the canonical oracle $\kappa(S)=\widehat{H}_{|S|}/\eta_u$ of equation 22, which is defined for every S with $|S| \leq K$; every query of A has size at most K, so the answer is always available. Since A is deterministic and its tie resolution is part of it, this run fixes a query sequence $S_1,\dots,S_M$ with $M \leq nK$ and an output T_0 with $|T_0| \leq K$, all of them before, and independently of, any choice of hidden optimum. Pad T_0 to a K-set T.

The choice of O. By Lemma 24 applied to $S_1,\dots,S_M$ and T, and since $n \geq K^4 \geq B_K$, there is a K-set $O \subseteq N \setminus T$ with $|O \cap S_i| \leq 1$ for every i.

The instance. Build the family of Lemma 15 on $N=B \sqcup O$ with this O. By that lemma, f is monotone, submodular and normalized, $\tilde{f}(\emptyset)=0$, and the smallest admissible error factors of Definition 1 are exactly (η_u,η_o), so the error product is exactly η and the prescribed split is realized.

The actual run reproduces the canonical one. Every canonical query satisfies $|S_i| \leq K$ and $|S_i \cap O| \leq 1$, so by Lemma 16(1) its true predicted value $\tilde{f}(S_i)$ equals the canonical answer $\widehat{H}_{|S_i|}/\eta_u$. Lemma 13 then applies: A run on $(f,\tilde{f})$ issues exactly $S_1,\dots,S_M$ and outputs T_0.

The value. By construction $O \cap T=\emptyset$ and $T_0 \subseteq T$, so $T \subseteq B$ and $T_0 \subseteq B$. Monotonicity of f and Lemma 17 give

$$\frac{f(T_0)}{f(O^*)} \;\leq\; \frac{f(T)}{f(O)}=F(K,0)=\rho_K(\eta).$$

The matching direction. Predictive greedy, retaining previously queried values, spends $\sum_{t=0}^{K-1}(n-t)=nK-K(K-1)/2 \leq nK$ queries, each on a set of size at most K, and outputs a set of size at most K, so it belongs to $\mathcal{A}_{\text{lin}}$; by Theorem 1 and Remark 9 its ratio is at least $\rho_K(\eta)$ on every instance on $n \geq 2K$ elements with error at most η, and $n \geq K^4 \geq 2K$. The hard instance above has error exactly η and so belongs to that class. Together the two directions give $\alpha^*(\mathcal{A}_{\text{lin}})=\rho_K(\eta)$. □

Remark 15 (The budget property actually used)**.** *The only property of the budget used in the proof is the edge bound $e \leq M\binom{K}{2}$ of Lemma 24: any budget M with $(n-K)(n-2K+1)>MK(K-1)^2$ is covered by the same choice of O, and $M=nK$ gives equation 39. By contrast, covering the budget nK on the scale of Theorem 3 needs $n^c \geq nK$, hence $c>1$ and $\tau=\lceil c \rceil+1 \geq 3$, and delivers only the weaker constant $H_{K,3}(\eta)>H_{K,1}(\eta)$; running the argument at the exact budget is what keeps the $\tau=1$ constant.*

Remark 16 (Where the threshold stands)**.** *Exhaustive search over all K-subsets stays within the budget nK exactly when $\binom{n}{K} \leq nK$, that is for $n \leq 5$ at $K=2$ and for $n \leq K+2$ at $K \geq 3$, since $\binom{n}{K}/n=\binom{n-1}{K-1}/K$ increases with n. On every instance with error at most η the predicted argmax is worth at least $f(O^*)/\eta$ (Proposition 3), which exceeds $\rho_K(\eta)$ for $1<\eta<K$ because $V_0(\eta)-V_1(\eta)=\frac{K-\eta}{K\eta k_1}>0$ by equation 9. So the least n_0 from which $\alpha^*(\mathcal{A}_{\text{lin}})=\rho_K(\eta)$ satisfies $6 \leq n_0 \leq 9$ at $K=2$ and $K+3 \leq n_0 \leq B_K \leq K^4$ at $K \geq 3$; here $B_2=9$ and $B_3=44$. The quartic exponent of B_K comes from nK queries, at most $\binom{K}{2}$ pairs each, and the need for an independent K-set. Whether n_0 really grows like K^4, or more slowly, is open.*

Remark 17 (Two endpoints)**.** *The restriction on n cannot be dropped from statements of this kind: at $n=4$, $K=2$, $\eta=3/2$, querying all six pairs and returning the predicted argmax stays within the nK-query budget and guarantees $2/3=1/\eta$, above $\rho_2(3/2)=3/5$, so for small n the class contains exhaustive search and no greedy-optimality reading survives; the bounds on the*

least admissible n are those of Remark 26. At $\eta = 1$ the same conclusion holds by the route of Appendix E.3: the calibration reads $\bar{\theta} = 1$, where the family of Theorem 3 degenerates to $\tilde{f} = f$ with error exactly 1, and $U_K(1) = L_K(1) = \rho_K(1)$; that family's predictor is likewise a function of $|S|$ on balanced small sets, which is all that the choice of O in Lemma 24 and the proof above use, so for $n \geq K^4$ predictive greedy is exactly optimal within $\mathcal{A}_{\text{lin}}$ at $\eta = 1$ as well, continuously matching Theorem 2 as $\eta \downarrow 1$. The case $K = 1$ sits outside the statement, which requires $K \geq 2$: there predictive greedy is exactly optimal among all deterministic algorithms with no budget restriction at all (Remark 6).

E.3 Bounded-query hardness (Theorem 3)

Fix an integer τ with $1 \leq \tau < K$ and a design parameter $\theta \geq 1$, and put

$$a_\theta = 1 - \frac{1}{\theta K}, \qquad A_\theta = a_\theta^\tau \, \frac{K}{K-\tau}, \qquad B_\theta = a_\theta^{1-\tau},$$
$$\delta(\theta) = A_\theta B_\theta - 1 = \frac{\tau - 1/\theta}{K-\tau}, \qquad \Psi(\theta) = \theta\big(1+\delta(\theta)\big) = \frac{\theta K - 1}{K-\tau}.$$

The middle identity is immediate from $A_\theta B_\theta = a_\theta K/(K-\tau) = (K-1/\theta)/(K-\tau)$. The pair built from θ in Lemma 19 below has smallest admissible error factors $\sqrt{\theta}A_\theta$ and $\sqrt{\theta}B_\theta$, whose product is $\Psi(\theta)$, and Ψ inverts in closed form: the design parameter that realizes a prescribed error η is

$$\bar{\theta} = \Psi^{-1}(\eta) = \frac{\eta(K-\tau)+1}{K},$$

which is at least 1 exactly when $\eta \geq (K-1)/(K-\tau)$. This is the hypothesis of Theorem 3.

The hard family. Fix integers $1 \leq \tau < K < n$ and a real $\theta \geq 1$, write $a = a_\theta$, $A = A_\theta$, $B = B_\theta$, and for a hidden K-set $O \subseteq N$ and $S \subseteq N$ put $x = |S \setminus O|$ and $y = |S \cap O|$. Define

$$F_O(S) = \theta^{-1/2}\Big[1 - a^x\Big(1 - \frac{y}{K}\Big)\Big], \qquad G_O(S) = \begin{cases} 1 - a^{x+y}, & y \leq \tau, \\ 1 - a^{x+\tau}\,\dfrac{K-y}{K-\tau}, & y \geq \tau, \end{cases} \tag{26}$$

together with the balanced-region profile $\widehat{G}(s) = 1 - a^s$. The two branches of G_O agree at $y = \tau$, where both give $1 - a^{x+\tau}$. Call a set S *balanced* for O when $|S \cap O| \leq \tau$. The instance is $f = F_O$ and $\tilde{f} = G_O$; the prefactor $\theta^{-1/2}$ places the symmetric split and cancels from every ratio below.

Lemma 19 (The hard family)**.** *With the notation of equation 26:*

1. *F_O is normalized, monotone and submodular on N; it satisfies $0 \leq \sqrt{\theta}F_O \leq 1$, its unique maximizer among sets of size at most K is O, with $F_O(O) = \theta^{-1/2}$, and every T with $|T| \leq K$ and $T \cap O = \emptyset$ satisfies $F_O(T)/F_O(O) = 1 - a^{|T|} \leq 1 - a^K = L_K(\theta)$.*
2. *$G_O(\emptyset) = 0$, and the smallest factors for which the pair (F_O, G_O) satisfies Definition 1 are $\eta_o^{\text{act}} = \sqrt{\theta}A$ and $\eta_u^{\text{act}} = \sqrt{\theta}B$, both at least 1, with product $\eta^{\text{act}} = \theta AB = \frac{\theta K - 1}{K-\tau} = \Psi(\theta)$.*
3. *If S is balanced for O, then $G_O(S) = \widehat{G}(|S|)$.*
4. *For every split $\eta_u\eta_o = \Psi(\theta)$ with $\eta_u, \eta_o > 0$, the rescaled predictor βG_O with $\beta = \eta_o/(\sqrt{\theta}A)$ has error exactly (η_u, η_o) against F_O, and $\beta\widehat{G}$ is again a function of $|S|$ alone.*

Proof. Both F_O and G_O depend on S only through the pair (x, y), so Lemma 5 applies to them with the two-class partition $N = (N \setminus O) \cup O$.

Item 1. Write $\bar{F} = \sqrt{\theta}F_O$, so $\bar{F}(x,y) = 1 - a^x(1 - y/K)$. Its first differences are $\Delta_x\bar{F} = a^x(1-a)(1-y/K)$ and $\Delta_y\bar{F} = a^x/K$, nonnegative on every feasible edge because $0 < a < 1$ and $y \leq K$, and its second differences are $\Delta_x^2\bar{F} = -a^x(1-a)^2(1-y/K)$, $\Delta_y\Delta_x\bar{F} = \Delta_x\Delta_y\bar{F} = -a^x(1-a)/K$, $\Delta_y^2\bar{F} = 0$, all nonpositive. Lemma 5 then makes F_O a normalized monotone submodular function; normalization is $\bar{F}(0,0) = 0$. Also $0 \leq \bar{F} \leq 1$, and $\bar{F}(x,y) = 1$ holds

exactly when $a^x(1-y/K)=0$, that is, when $y=K$. A set of size at most K with $|S\cap O|=K$ equals O, so O is the unique maximizer among such sets and $\max_{|S|\le K} F_O(S)=F_O(O)=\theta^{-1/2}$. If $|T|\le K$ and $T\cap O=\emptyset$ then $x=|T|$ and $y=0$, so $F_O(T)/F_O(O)=1-a^{|T|}\le 1-a^K$, and $1-a^K=1-(1-\frac{1}{\theta K})^K=L_K(\theta)$.

Item 2. Clearly $G_O(\emptyset)=1-a^0=0$. On an edge of the count grid with positive true gain put $\rho=\Delta G_O/(\sqrt{\theta}\,\Delta F_O)=\Delta G_O/\Delta\bar{F}$, the predicted-to-true gain ratio of that edge. Subtracting the two displays in equation 26 gives the following exhaustive table, in which every power a^x cancels, so that the table is the same for all $n>K$:

Direction	Starting state	ρ
x	$0\le y\le\tau$	$a^yK/(K-y)$
x	$\tau\le y<K$	A
y	$0\le y\le\tau-1$	a^y/θ
y	$\tau\le y<K$	A

The four entries are computed directly. In the first row both endpoints lie in the first branch, so $\Delta G_O=a^{x+y}(1-a)$ while $\Delta\bar{F}=a^x(1-a)(1-y/K)$, giving $\rho=a^yK/(K-y)$. In the second row both endpoints lie in the second branch, so $\Delta G_O=a^{x+\tau}(1-a)\frac{K-y}{K-\tau}$ and $\rho=a^\tau K/(K-\tau)=A$. In the third row $\Delta G_O=a^{x+y}(1-a)$ and $\Delta\bar{F}=a^x/K$, so $\rho=a^y(1-a)K=a^y/\theta$ by $1-a=1/(\theta K)$. In the fourth row $\Delta G_O=a^{x+\tau}\frac{1}{K-\tau}$ and $\Delta\bar{F}=a^x/K$, so $\rho=a^\tau K/(K-\tau)=A$. An x-edge starting at $y=K$ has zero true and zero predicted gain, so it carries no band constraint and is consistent with Definition 1; no y-edge starts at $y=K$. The edge from $y=\tau-1$ to $y=\tau$ is in the third row and the edge from $y=\tau$ to $y=\tau+1$ is in the fourth, so the junction of the two branches needs no separate treatment.

Write $h_y=a^yK/(K-y)$ for the first row. Then

$$\frac{h_{y+1}}{h_y}-1=\frac{a(K-y)-(K-y-1)}{K-y-1}=\frac{(\theta-1)K+y}{\theta K(K-y-1)}\;\ge\;0,$$

so $1=h_0\le h_y\le h_\tau=A$ for $0\le y\le\tau$; in particular $A\ge 1$. The third row a^y/θ decreases from $1/\theta$ at $y=0$ to $a^{\tau-1}/\theta=1/(\theta B)$ at $y=\tau-1$, and $B=a^{1-\tau}\ge 1$. Since $A\ge 1$ and $1/(\theta B)\le 1$, the largest and the smallest values of ρ over the whole grid are A and $1/(\theta B)$, and both are attained on actual edges because $n>K$ and $1\le\tau<K$. In the notation of Definition 1, where the band reads $1/\eta_u\le\tilde{d}_e(S)/d_e(S)\le\eta_o$ and $\tilde{d}_e(S)/d_e(S)=\sqrt{\theta}\,\rho$, the smallest admissible factors of the pair are therefore $\eta_o^{\mathrm{act}}=\sqrt{\theta}A$ and $\eta_u^{\mathrm{act}}=\sqrt{\theta}B$, both at least 1, and their product is $\theta AB=\Psi(\theta)$.

Item 3. If $|S\cap O|=y\le\tau$ then S falls in the first branch, where $x+y=|S|$ and $G_O(S)=1-a^{|S|}=\widehat{G}(|S|)$.

Item 4. Replacing G_O by βG_O multiplies every ρ by β, so the largest predicted-to-true ratio becomes $\beta\sqrt{\theta}A=\eta_o$ and the smallest becomes $\beta\sqrt{\theta}/(\theta B)=\eta_o/(\theta AB)=\eta_o/\Psi(\theta)=1/\eta_u$; the product is unchanged, which is the instance of Lemma 2 used here. The canonical profile $\widehat{G}$ is multiplied by the same constant and remains a function of $|S|$ alone, so the rescaling carries no information about the identity of O. □

Proof of Theorem 3. The deterministic statement. Let A be deterministic, make at most $Q=n^c$ queries of size at most K, and return a set of size at most K. Fix $\theta\ge 1$ and a rescaling constant $\beta>0$ for the moment, both to be chosen in the calibration below, and let $S_1,\dots,S_Q$ and T_0 be the queries and output of A against the canonical oracle $S\mapsto\beta\widehat{G}(|S|)$ (Lemma 13). Let O be a uniformly random K-subset of N. Summing equation 18 over the Q canonical queries and adding $\Pr[T_0\cap O\ne\emptyset]\le\mathbb{E}|T_0\cap O|=|T_0|K/n\le K^2/n$,

$$\Pr\big[\exists\, i:\ |S_i\cap O|>\tau\ \text{ or }\ T_0\cap O\ne\emptyset\big]\;\le\;\frac{K^{2\tau+2}}{(\tau+1)!\,n^{\tau+1-c}}+\frac{K^2}{n}.$$

Under $n \geq 4K^{c+2}$ and $\tau = \lceil c \rceil + 1$ the exponent $\tau + 1 - c \geq 2 > 0$, so $n^{\tau+1-c} \geq 4^{\tau+1-c}K^{(c+2)(\tau+1-c)}$ and the first term is at most

$$\frac{K^{\,2\tau+2-(c+2)(\tau+1-c)}}{(\tau+1)!\,4^{\tau+1-c}} = \frac{K^{\,c(c+1-\tau)}}{(\tau+1)!\,4^{\tau+1-c}} \;\leq\; \frac{1}{16\,(\tau+1)!} \;\leq\; \frac{1}{32},$$

because $\tau \geq c+1$ makes the exponent $c(c+1-\tau)$ of K nonpositive, $\tau+1-c \geq 2$ gives $4^{\tau+1-c} \geq 16$, and $\tau \geq 1$ gives $(\tau+1)! \geq 2$. The second term is $K^2/n \leq 1/(4K^c) \leq 1/4$. The sum is at most $9/32 < 1$, so at least one K-set O makes every canonical query balanced and misses T_0; fix such an O. By Lemma 19(3) every S_i receives the canonical answer on βG_O, so by Lemma 13 the run of A against βG_O outputs T_0, and $T_0 \cap O = \emptyset$ with $|T_0| \leq K$ gives, by Lemma 19(1),

$$\frac{f(T_0)}{f(O^*)} \;\leq\; 1 - a_\theta^K = L_K(\theta).$$

Calibration. Choose $\theta = \bar{\theta} = \Psi^{-1}(\eta) = (\eta(K-\tau)+1)/K$, which is at least 1 exactly under the hypothesis $\eta \geq (K-1)/(K-\tau)$. By Lemma 19(2) the pair's smallest admissible factors then have product $\Psi(\bar{\theta}) = \eta$ exactly, and by item (4) any prescribed split $\eta_u\eta_o = \eta$ is realized by the rescaled predictor βG_O with $\beta = \eta_o/(\sqrt{\bar{\theta}}A)$. That rescaling leaves the true objective F_O untouched, hence changes no ratio of true values, and it leaves the canonical profile $\beta\widehat{G}$ a function of $|S|$ alone, which is all that the counting argument above uses. Since $\bar{\theta}K = \eta(K-\tau)+1$, the value bound becomes

$$L_K(\bar{\theta}) = 1 - \Big(1 - \frac{1}{\bar{\theta}K}\Big)^K = 1 - \Big(1 - \frac{1}{\eta(K-\tau)+1}\Big)^K = H_{K,\tau}(\eta),$$

the constant of the theorem.

The limits. Since $\eta\tau \geq 1$ gives $\bar{\theta} = \eta - \frac{\eta\tau-1}{K} \leq \eta$ and L_K is decreasing, $H_{K,\tau}(\eta) \geq L_K(\eta)$, so Theorem 3 and Proposition 4 do not cross at any finite K. With c and η fixed, $K/(\eta(K-\tau)+1) \to 1/\eta$, so $K\ln\big(1 - \frac{1}{\eta(K-\tau)+1}\big) \to -1/\eta$ and $H_{K,\tau}(\eta) \to 1 - e^{-1/\eta}$. □

Remark 18 (Randomized algorithms)**.** *For a randomized algorithm the same bound holds in expectation up to an additive term. Fix the random seed r; then A_r is deterministic, with canonical queries $S_1^{(r)}, \ldots, S_Q^{(r)}$ and canonical output $T_0^{(r)}$, and let E_r be the event, depending on O, that all of its canonical queries are balanced. On E_r the run against βG_O returns $T_0^{(r)}$ by Lemma 13. Every output T has at most K elements, so its counts satisfy $x \leq K$, and $a_\theta \in (0,1)$ gives $a_\theta^x \geq a_\theta^K$, hence $1 - a_\theta^x(1-y/K) \leq 1 - a_\theta^K + y/K$, while off E_r the output ratio is at most 1. Hence, pointwise in O and in r,*

$$\frac{f_O(T)}{f_O(O^*)} \;\leq\; 1 - a_\theta^K + \frac{|T_0^{(r)} \cap O|}{K} + \mathbf{1}[E_r^c].$$

Averaging over a uniform O bounds the middle term by $|T_0^{(r)}|K/(Kn) \leq K/n$ and, by equation 18 summed over the $Q = n^c$ canonical queries of the seed r, the last one by $K^{2\tau+2}/((\tau+1)!\,n^{\tau+1-c})$. Averaging over r, exchanging the two finite expectations and using $\min_O \leq \mathbb{E}_O$ produces a single K-set O, hence a single instance, on which

$$\mathbb{E}\Big[\frac{f_O(T)}{f_O(O^*)}\Big] \;\leq\; L_K(\theta) + \varepsilon_n, \qquad \varepsilon_n = \frac{K}{n} + \frac{K^{2\tau+2}}{(\tau+1)!\,n^{\tau+1-c}}.$$

Only the averaging direction of the minimax principle is used. The additive term does not vanish on its own: at the smallest permitted $n = 4K^{c+2}$ and integer c its second summand equals $1/(16(c+2)!)$, a constant, so a limit statement for randomized algorithms selects $n(K)$ with $\varepsilon_n \to 0$, for example $n \geq 4K^{c+3}$.

Remark 19 (The case $\tau = 1$)**.** *At $\tau = 1$, which under $\tau = \lceil c \rceil + 1$ means $c = 0$, the hardness constant is U_K: with $k_1 = (K-1)\eta + 1$,*

$$H_{K,1}(\eta) = 1 - \Big(1 - \frac{1}{\eta(K-1)+1}\Big)^K = 1 - \Big(1 - \frac{1}{k_1}\Big)^K = 1 - q^K = U_K(\eta),$$

because $1 - 1/k_1 = (k_1-1)/k_1 = (K-1)\eta/k_1 = q$. For every $K \geq 2$ and $\eta > 1$ the four curves are then ordered

$$L_K(\eta) \;\leq\; \rho_K(\eta) \;<\; U_K(\eta) = H_{K,1}(\eta),$$

the first inequality by Proposition 4, the second by $\rho_K \le V_{K-1}$ and the difference $U_K - V_{K-1} = \frac{q^{K-1}(\eta-1)}{K\eta k_1} > 0$ computed in Appendix D.7.

Remark 20 (The qualifier on the query class)**.** *Combining Proposition 4 with Theorem 3 pins the bounded-query optimum at error η to $1 - e^{-1/\eta}$ asymptotically in K, for query classes that contain predictive greedy; $nK \le n^c$, for instance integer $c \ge 2$ with $K \le n$, is a sufficient condition. The qualifier cannot be dropped. At $c = 0$ a deterministic algorithm allowed a single query of size at most K can be forced to output a set of value zero: run it against the predictor that is identically zero, so that it issues a fixed query Q_0 with $|Q_0| \le K$, receives the answer 0 and returns a fixed set T with $|T| \le K$; for $n \ge 2K + 2$ pick two elements $e_1, e_2 \notin Q_0 \cup T$, let f be modular with weights 1 on e_1 and e_2 and 0 elsewhere, and $\tilde{f}$ modular with weights 1 on e_1, 2 on e_2 and 0 elsewhere. The predicted-to-true ratios are 1 on e_1, 2 on e_2 and $0 = 0$ elsewhere, so the error is exactly 2; the new predictor still answers 0 on Q_0, so the run is unchanged and returns T, and $e_1, e_2 \notin T$ gives $f(T) = 0$ while $f(O^*) \ge 1$. At error 2 the ratio of that class is therefore 0, not $1 - e^{-1/2}$; this does not collide with Theorem 3, which caps the ratio from above and does not assert that the cap is attained. At finite K the gap between $L_K(\eta)$ and $H_{K,\tau}(\eta)$ is of order $(c+1)/K$: writing $s = \eta\tau - 1 > 0$, so that $\bar{\theta} = \eta - s/K$, the expansion $(1 - \frac{1}{\theta K})^K = e^{-1/\theta}\big(1 - \frac{1}{2\theta^2 K} + O(K^{-2})\big)$ at $\theta = \eta$ and at $\theta = \bar{\theta}$ gives $K\big(H_{K,\tau}(\eta) - L_K(\eta)\big) \to e^{-1/\eta}(\eta\tau - 1)/\eta^2$, which is of order $c + 1$ at fixed η and already nonzero at $\tau = 1$.*

Remark 21 (The size cap is essential for this construction)**.** *The family equation 26 cannot be freed of the cap on query sizes: one round of the n queries $N \setminus \{e\}$ separates membership in O, because the two $(n-1)$-set answer classes of G_O differ. If $e \in O$ then $S = N \setminus \{e\}$ has $x = n - K$ and $y = K - 1 \ge \tau$, so the second branch gives $G_O(S) = 1 - a^{n-K+\tau}/(K - \tau)$, while if $e \notin O$ then $x = n - K - 1$ and $y = K$, and the same branch gives $G_O(S) = 1$. Hardness against queries of arbitrary size needs a different construction; Appendix E.4 records one direction and its limits.*

E.4 Queries of arbitrary size at a linear budget

Theorem 3 restricts every query to at most K elements, and Theorem 2 inherits that restriction; by Lemma 16(2) the restriction is essential for the family used there, whose predictor is not a function of $|S|$ on larger sets. Removing it costs an additive term exponentially small in K.

Proposition 8 (Arbitrary query sizes at linear budget)**.** *Let $K \ge 3$, $\eta > 1$ and $c \ge 1$, and let $\alpha_n(K, \eta)$ denote the best approximation ratio of the deterministic algorithms that make at most cnK queries to $\tilde{f}$ of arbitrary size and output a set of size at most K, over the instances on n elements with monotone submodular normalized f and error at most η. Then*

$$\rho_K(\eta) \;\le\; \alpha_n(K, \eta) \;\le\; \frac{1}{\eta} \qquad \text{for every } n \ge 2K,$$

and

$$\limsup_{n\to\infty} \alpha_n(K, \eta) \;\le\; \min\Big\{\frac{1}{\eta},\, W_K(\eta)\Big\}, \qquad 0 < W_K(\eta) - \rho_K(\eta) < \frac{1}{K\big(e^{K-1} - K - 1\big)},$$

where $W_K(\eta)$ is the explicit constant of Lemma 22. Consequently

$$\rho_K(\eta) \;\le\; \liminf_{n\to\infty} \alpha_n(K, \eta) \;\le\; \limsup_{n\to\infty} \alpha_n(K, \eta) \;\le\; \min\Big\{\frac{1}{\eta},\, \rho_K(\eta) + \frac{1}{K(e^{K-1} - K - 1)}\Big\}:$$

even with arbitrary-size queries, predictive greedy is optimal among algorithms of its own oracle complexity up to an additive term exponentially small in K. For $\eta \ge K$ all four quantities equal $1/\eta$.

Lemma 20 (The plateau family for arbitrary query sizes)**.** *Fix $K \ge 3$, $\eta > 1$ and a split $\eta_u, \eta_o > 0$ with $\eta_u \eta_o = \eta$, keep $k_1 = \eta(K-1) + 1$ and $q = 1 - 1/k_1$, and put $\nu = \eta/(\eta - 1)$ and*

$$\Psi_0(t) = (K\eta - t - 1)\,\nu^t - K(\eta - 1), \qquad m = \min\{z \in \mathbb{Z}_{\ge 1} :\; \Psi_0(z) \le 0\}. \tag{27}$$

Then m is well defined and $\eta(K-1) < m < K\eta$. Set $B_m = \eta(\nu^m - 1) - m > 0$ and $d = (\nu^m/K - 1)/B_m$; then

$$\begin{gathered} d - \frac{1}{m+1} = \frac{-\Psi_0(m)}{K(m+1)B_m} \;\ge\; 0, \qquad \frac{1}{m} - d = \frac{\nu\,\Psi_0(m-1)}{KmB_m} \;\ge\; 0, \\ d - \frac{1}{K\eta} = \frac{m - \eta(K-1)}{K\eta B_m} \;>\; 0. \end{gathered} \tag{28}$$

With $j = \max\{0, \min\{K-1, K+1-\lceil\eta\rceil\}\}$, $Q = q^j$, $D = Qd$ and $t^* = j + m$, define on the integers $x \geq 0$

$$r_x = \begin{cases} q^x, & x \leq j, \\ Q - (x-j)D, & j < x \leq t^*, \\ 0, & x > t^*, \end{cases} \qquad g_x = \begin{cases} \dfrac{q^x}{K}, & x \leq j, \\ \eta D - \Big(\eta D - \dfrac{Q}{K}\Big)\nu^{x-j}, & j < x \leq t^*, \\ 0, & x > t^*, \end{cases} \tag{29}$$

and $a_x = r_x - g_x$. On a ground set $N = B \sqcup O$ with $|O| = K$ and $|B| = n - K$, write $x = |S \cap B|$, $y = |S \cap O|$, $c_y = (K-y)/(K-1)$ and $C = k_1/K$, and set $f(S) = F(x,y)$, $\tilde{f}(S) = H(x,y)/\eta_u$ with

$$F(x,y) = \begin{cases} 1 - r_x, & y = 0, \\ 1 - c_y\, a_x, & y \geq 1, \end{cases} \qquad H(x,y) = \begin{cases} C - r_x - (\eta-1)a_x, & y = 0, \\ C - \eta\, c_y\, a_x, & y \geq 1. \end{cases} \tag{30}$$

Then f is monotone, submodular and normalized, $0 \leq f \leq 1$, $f(O) = 1$ and O maximizes f among the sets of size at most K; $\tilde{f}(\emptyset) = 0$ and every edge satisfies $\Delta f/\eta_u \leq \Delta\tilde{f} \leq \eta_o\, \Delta f$ with both endpoints attained on edges of positive gain, so the smallest admissible error factors of Definition 1 are exactly (η_u, η_o). Moreover

$$a_x - a_{x+1} = \frac{g_{x+1}}{\eta} \quad \textit{and, equivalently,} \quad H(x+1,0) = H(x,1) \quad \textit{for every } x \geq 0. \tag{31}$$

Proof. The truncation index. The function $\nu^{-t}\Psi_0(t) = (K\eta - t - 1) - K(\eta-1)\nu^{-t}$ is strictly concave in t, equals $K - 1 > 0$ at $t = 0$ and tends to $-\infty$, so m is well defined. Since $\nu = 1 + 1/u$ with $u = \eta - 1 > 0$, the elementary inequality $(1 + 1/u)^{u+1} > e$ gives $\nu^\eta > e$, hence $\nu^{\eta(K-1)} = (\nu^\eta)^{K-1} > e^{K-1} > K$, and $\Psi_0(\eta(K-1)) = (\eta-1)\big(\nu^{\eta(K-1)} - K\big) > 0$. Concavity of $\nu^{-t}\Psi_0(t)$ and positivity at $t = 0$ and at $t = \eta(K-1)$ make Ψ_0 positive on the whole interval in between, so $m > \eta(K-1)$; and $\Psi_0(t) < 0$ for every $t \geq K\eta - 1$, since there the first summand is nonpositive and the second is negative, so $m < K\eta$. Also $\nu^m > \nu^{\eta(K-1)} > e^{K-1}$, whence $B_m = \eta\nu^m - \eta - m > \eta(e^{K-1} - 1 - K) > 0$ for $K \geq 3$.

The three identities equation 28. Multiplying out,

$$(m+1)\Big(\frac{\nu^m}{K} - 1\Big) - B_m = \nu^m\, \frac{m + 1 - K\eta}{K} + (\eta - 1) = -\frac{\Psi_0(m)}{K},$$

which is the first identity after dividing by $(m+1)B_m$, and $\Psi_0(m) \leq 0$ by the definition of m. Similarly

$$B_m - m\Big(\frac{\nu^m}{K} - 1\Big) = \nu^m\, \frac{K\eta - m}{K} - \eta = \frac{\nu\, \Psi_0(m-1)}{K},$$

using $\nu(\eta - 1) = \eta$, which gives the second identity; here $\Psi_0(m-1) > 0$ by the minimality of m when $m \geq 2$, and $\Psi_0(0) = K - 1 > 0$ when $m = 1$. Finally $K\eta\big(\nu^m/K - 1\big) - B_m = m - \eta(K-1) > 0$ gives the third.

The closing step. Solving $r_{t^*} = g_{t^*}$ for the plateau slope, that is $Q - mD = \eta D - (\eta D - Q/K)\nu^m$, yields $D = Q(\nu^m/K - 1)/B_m = Qd$, so the definition of d is exactly the one that makes the two sequences meet at t^* and vanish together afterwards. Moreover $r_{t^*} = Q(1 - md) \geq 0$ is $d \leq 1/m$ and $r_{t^*} \leq D$ is $d \geq 1/(m+1)$, both supplied by equation 28.

One-dimensional facts. On the plateau, $g_{x+1} = \nu(g_x - D)$ by inspection of equation 29, using $\nu(\eta - 1) = \eta$. The third identity of equation 28 says $\eta D > Q/K$, so on the plateau the subtracted term in g_x is positive and increasing in x: g is positive and decreasing there, ending at $g_{t^*} = r_{t^*} \geq 0$, and $g_x < \eta D$ throughout. Next, equation 31: on the geometric phase $a_x - a_{x+1} = q^x(1-q)\frac{K-1}{K} = q^x\frac{K-1}{Kk_1}$ and $g_{x+1}/\eta = q^{x+1}/(K\eta)$ agree because $qk_1 = (K-1)\eta$; on the plateau

$$a_x - a_{x+1} = D - (g_x - g_{x+1}) = D - \big(g_x - \nu(g_x - D)\big) = (\nu - 1)(g_x - D) = \frac{\nu - 1}{\nu}\, g_{x+1} = \frac{g_{x+1}}{\eta},$$

using $(\nu - 1)/\nu = 1/\eta$; and for $x \geq t^*$ both sides vanish, since $a_{t^*} = r_{t^*} - g_{t^*} = 0$ and $g_{x+1} = 0$ there. Consequently a is nonincreasing, and since $a_{t^*} = 0$ and a vanishes beyond t^*, $a \geq 0$

everywhere. The steps of r are nonincreasing: they decrease geometrically up to the junction, where $r_{j-1} - r_j = q^{j-1}/k_1 = Q/((K-1)\eta) \geq D$ because $d \leq 1/m < 1/(\eta(K-1))$, then equal D on the plateau, then equal $r_{t^*} \leq D$ and then 0. The steps of r dominate those of a: on the plateau $a_x - a_{x+1} = g_{x+1}/\eta < D$ since $g_{x+1} < \eta D$, on the geometric phase $g_{x+1}/\eta = q^x \frac{K-1}{Kk_1} \leq q^x/k_1$, and for $x \geq t^*$ the step of a vanishes. Finally $Kg_x \geq r_x$ throughout: equality holds on the geometric phase, and on the plateau $Kg_{j+\ell} - r_{j+\ell}$ is concave in ℓ (a linear function minus a positive multiple of ν^ℓ), equals 0 at $\ell = 0$ and equals $(K-1)r_{t^*} \geq 0$ at $\ell = m$, so it is nonnegative in between; beyond t^* both sides vanish.

Legality. The first differences of F on the count grid are

$$\Delta_x F(x,0) = r_x - r_{x+1}, \qquad \Delta_x F(x,y) = c_y \frac{g_{x+1}}{\eta} \quad (1 \leq y \leq K),$$
$$\Delta_y F(x,0) = r_x - a_x = g_x, \qquad \Delta_y F(x,y) = \frac{a_x}{K-1} \quad (1 \leq y < K),$$

all nonnegative, each nonincreasing in x by the facts just proved, and each nonincreasing in y: passing from $y = 0$ to $y = 1$ replaces $r_x - r_{x+1}$ by g_{x+1}/η, which is smaller, and g_x by $a_x/(K-1)$, which is smaller because $Kg_x \geq r_x$; and c_y decreases in y. Lemma 5 lifts the grid to a monotone submodular set function f with $f(\emptyset) = F(0,0) = 0$; moreover $F(0,K) = 1$ and $0 \leq F \leq 1$ on the whole grid, so $f(O) = 1$ and O maximizes f under the cardinality constraint. The four edge classes carry

$$\begin{aligned}
B\text{-edge, } y = 0: &\quad \Delta F = r_x - r_{x+1}, &\quad \Delta H = (r_x - r_{x+1}) + (\eta - 1)\frac{g_{x+1}}{\eta},\\
B\text{-edge, } y \geq 1: &\quad \Delta F = c_y \frac{g_{x+1}}{\eta}, &\quad \Delta H = \eta\, c_y \frac{g_{x+1}}{\eta},\\
O\text{-edge, } y = 0: &\quad \Delta F = g_x, &\quad \Delta H = g_x,\\
O\text{-edge, } 1 \leq y < K: &\quad \Delta F = \frac{a_x}{K-1}, &\quad \Delta H = \frac{\eta\, a_x}{K-1},
\end{aligned}$$

so $\Delta F \leq \Delta H \leq \eta\, \Delta F$ holds in the first row because $0 \leq g_{x+1}/\eta \leq r_x - r_{x+1}$, and with ratio exactly 1 or exactly η in the others. Also $H(0,0) = C - 1 - (\eta - 1)\frac{K-1}{K} = 0$, so $\tilde{f}(\emptyset) = 0$. At $x = 0$ the third row has gain $g_0 = 1/K > 0$ and realizes $\Delta\tilde{f} = \Delta f/\eta_u$, and the fourth has gain $a_0/(K-1) = 1/K > 0$ and realizes $\Delta\tilde{f} = \eta_o\, \Delta f$, so both endpoints of the band are attained and the smallest admissible factors are exactly (η_u, η_o).

The equivalence in equation 31. Since $c_1 = 1$, the identity $H(x+1,0) = H(x,1)$ reads $r_{x+1} + (\eta - 1)a_{x+1} = \eta a_x$, that is $r_{x+1} = \eta(a_x - a_{x+1}) + a_{x+1}$; as $a_{x+1} = r_{x+1} - g_{x+1}$, this is exactly $a_x - a_{x+1} = g_{x+1}/\eta$. □

Remark 22 (The truncation rule)**.** *The rule $m = \lceil K\eta \rceil - 1$ is not admissible in general: at $K = 3$ and $\eta = 667/500$ it gives $m = 4$ and a negative r_{t^*}, a negative marginal on the O-edges, whereas the rule equation 27 gives $m = 3$ at the same parameters. The bounds $1/(m+1) \leq d \leq 1/m$ of equation 28, which are what makes the closing step of Lemma 20 legal, are consequences of the sign pattern $\Psi_0(m-1) > 0 \geq \Psi_0(m)$ that defines m.*

Lemma 21 (Leakage)**.** *Let $n \geq 2K$ and let O be a uniformly random K-subset of N; build the family of Lemma 20 on N with hidden optimum O, and let $\kappa(S) = H(|S|,0)/\eta_u$ be the canonical oracle, which does not involve O. Then:*

1. *$\tilde{f}(S) = \kappa(S)$ whenever $|S \cap O| \leq 1$, and also whenever $|S| > t^* + K$; so $\tilde{f}(S) \neq \kappa(S)$ forces $|S \cap O| \geq 2$ and $|S| \leq t^* + K$.*
2. *For each fixed S, $\Pr[\tilde{f}(S) \neq \kappa(S)] \leq K^2(t^* + K)^2/(2n^2)$.*
3. *For any list of at most cnK fixed sets, the probability that $\tilde{f}$ and κ disagree on at least one of them is at most $cK^3(t^* + K)^2/(2n)$, which tends to 0 as $n \to \infty$.*

Proof. (1) By equation 31, $H(x,1) = H(x+1,0)$ for every $x \geq 0$, so on the region $|S \cap O| \leq 1$ the value $H(x,y)$ depends only on $x + y = |S|$ and equals $H(|S|,0)$, at every size. If $|S| > t^* + K$

then $x \geq |S| - K > t^*$, so $r_x = g_x = a_x = 0$ and $H(x, y) = C = H(|S|, 0)$ for every y, both functions having saturated.

(2) Let $s = |S| \leq t^* + K$. Then

$$\Pr\big[|S \cap O| \geq 2\big] \;\leq\; \binom{s}{2} \frac{K(K-1)}{n(n-1)} \;\leq\; \frac{s^2}{2} \cdot \frac{K^2}{n^2} \;\leq\; \frac{K^2(t^*+K)^2}{2n^2},$$

where the middle inequality uses $K(K-1)/(n(n-1)) \leq K^2/n^2$, which is $K \leq n$. Sets with $|S| > t^* + K$ never disagree by (1).

(3) Union bound over at most cnK sets: $cnK \cdot K^2(t^*+K)^2/(2n^2) = cK^3(t^*+K)^2/(2n)$. □

Lemma 22 (Value and the exponential gap)**.** *In the family of Lemma 20, $t^* \geq K$ and every K-set disjoint from O has value*

$$W_K(\eta) = 1 - Q + (K-j)\,Q\,d,$$

and for $K \geq 3$

$$W_K(\eta) - \rho_K(\eta) = \frac{(K-j)\,Q\,\big(m - \eta(K-1)\big)}{K\eta\,\big(\eta(\nu^m - 1) - m\big)} \;\in\; \Big(0,\; \frac{1}{K(e^{K-1} - K - 1)}\Big). \tag{32}$$

At $K = 3$, $\eta = 3/2$ one has $m = 4$, $d = 13/58$ and $W_3(3/2) = 523/928$, against $\rho_3(3/2) = 9/16 = 522/928$.

Proof. Since $m > \eta(K-1) > K - 1$ and m is an integer, $m \geq K$ and $t^* = j + m \geq K$, so $r_K = Q - (K-j)D$ lies on the geometric phase or the plateau and $F(K, 0) = 1 - r_K = 1 - Q + (K-j)Qd = W_K(\eta)$; a K-set disjoint from O has count $(K, 0)$, and $f(O) = 1$.

For the gap, $j = \max\{0, \min\{K-1, K+1-\lceil\eta\rceil\}\}$ is a minimizing branch of V by equation 9: for non-integer η it equals $\max\{0, K - \lfloor\eta\rfloor\}$, and at integer η the two branches $K - \eta$ and $K - \eta + 1$ tie. Hence $\rho_K(\eta) = V_j(\eta) = 1 - Q + (K-j)Q/(K\eta)$ by Theorem 1, and

$$W_K(\eta) - \rho_K(\eta) = (K-j)\,Q\Big(d - \frac{1}{K\eta}\Big) = \frac{(K-j)\,Q\,\big(m - \eta(K-1)\big)}{K\eta B_m}$$

by the third identity of equation 28. It is positive because $1 \leq K - j$, $Q > 0$ and $m > \eta(K-1)$. For the upper bound, $K - j \leq \eta$, $Q \leq 1$ and $m - \eta(K-1) < K\eta - \eta(K-1) = \eta$ bound the numerator by η^2, while $\nu^m > e^{K-1}$ and $m < K\eta$ give $B_m > \eta(e^{K-1} - 1 - K)$, so the quotient is smaller than $\eta^2/\big(K\eta \cdot \eta(e^{K-1} - K - 1)\big) = 1/(K(e^{K-1} - K - 1))$, which is positive for $K \geq 3$.

At $K = 3$, $\eta = 3/2$: $\nu = 3$, and $\Psi_0(1) = 6$, $\Psi_0(2) = 12$, $\Psi_0(3) = 12$, $\Psi_0(4) = -42$, so $m = 4$; then $B_m = \frac{3}{2}(81 - 1) - 4 = 116$ and $d = (27 - 1)/116 = 13/58$. Also $k_1 = 4$, $q = 3/4$, $\lceil\eta\rceil = 2$, $j = 2$, $Q = 9/16$, so $W_3(3/2) = 1 - \frac{9}{16} + \frac{9}{16} \cdot \frac{13}{58} = \frac{523}{928}$, and $\rho_3(3/2) = \frac{16\eta+3}{3(2\eta+1)^2} = \frac{9}{16} = \frac{522}{928}$. □

Proof of Proposition 8. Lower bound. Predictive greedy makes $nK - K(K-1)/2 \leq cnK$ queries, each of size at most K, and outputs a set of size K, so it belongs to the class; by Theorem 1 and Remark 9 its ratio is at least $\rho_K(\eta)$ on every instance on $n \geq 2K$ elements with error at most η. Hence $\alpha_n \geq \rho_K(\eta)$ for $n \geq 2K$.

The bound $1/\eta$. For $n \geq 2K$, Proposition 3 provides, for every deterministic algorithm with arbitrary query access to $\tilde{f}$ and output of size at most K, an instance of the class on which the ratio is at most $1/\eta$, so $\alpha_n \leq 1/\eta$.

The bound $W_K(\eta)$. Fix a deterministic A in the class and a ground set N of n elements. Run A against the canonical oracle κ of Lemma 21, which is defined on all sets and does not involve O; since A is deterministic this fixes a query sequence $S_1, \ldots, S_M$ with $M \leq cnK$ and an output T_0 with $|T_0| \leq K$, independently of O. Now draw O uniformly at random among the K-subsets of N and instantiate the family of Lemma 20 with that O; by that lemma the resulting pair $(f, \tilde{f})$ is an instance of the class, with error exactly η and any prescribed split. Let E be the event that $\tilde{f}$ and κ agree on all of $S_1, \ldots, S_M$ and that $T_0 \cap O = \emptyset$. By Lemma 21(3) and the union bound $\Pr[T_0 \cap O \neq \emptyset] \leq K^2/n$,

$$\Pr[\bar{E}] \;\leq\; \varepsilon_n := \frac{cK^3(t^*+K)^2}{2n} + \frac{K^2}{n}.$$

On E, Lemma 13 applies: A issues exactly $S_1, \dots, S_M$ on $(f, \tilde{f})$ and outputs T_0, which is disjoint from O and can be padded inside B to a K-set, so by monotonicity and Lemma 22 $f(T_0)/f(O^*) \le F(K,0) = W_K(\eta)$. Off E the ratio is at most 1. Averaging over O,

$$\mathbb{E}_O\Big[\frac{f(A^{\tilde{f}})}{f(O^*)}\Big] \;\le\; W_K(\eta) + \varepsilon_n,$$

so some choice of O yields an instance of the class on which the ratio is at most $W_K(\eta)+\varepsilon_n$. Hence $\alpha_n \le W_K(\eta) + \varepsilon_n$ for every n, and $\varepsilon_n \to 0$ as $n \to \infty$ with K, η and c fixed, so $\limsup_n \alpha_n \le W_K(\eta)$. The same averaging applies verbatim to a randomized algorithm measured in expectation over its own randomness, since the bound holds for each fixed realization of its coins.

Combining the three bounds with the gap estimate equation 32 gives the displayed chain. For $\eta \ge K$, Theorem 1 gives $\rho_K(\eta) = 1/\eta$, so the lower and upper ends coincide. □

Remark 23 (The asymptotics in n cannot be dropped)**.** *For $n < K + t^*$ the bound $\alpha_n \le W_K(\eta)$ fails on the family of Lemma 20: no subset of N then has size larger than $t^* + K$, so nothing is saturated and the region $|S \cap O| \ge 2$, $|S| \le t^* + K$, on which the predicted values do depend on O, covers the large sets. A reverse greedy, which starts from N and removes elements one at a time, reads O off those sets and returns it, achieving ratio exactly 1. This is why Proposition 8 is stated asymptotically in n, and why the quantity bounded there is a limit superior rather than a value at a fixed n.*

F Proofs for Section 3.4: the measurable error

This appendix proves the ordering $\eta_{\mathrm{sel}} \le \eta_{\mathrm{tr}} \le \eta$, records where the bound of Proposition 5 is proved, and constructs, for every K, the family on which the certificate is exact. It closes with the reason predictive greedy runs all K steps.

F.1 The three rulers

Lemma 23 (Ordering)**.** *Let the surrogate have finite error (η_u, η_o) in the sense of Definition 1 and put $\eta = \eta_u\eta_o$. Then every run of predictive greedy satisfies $\eta_{\mathrm{sel}} \le \eta_{\mathrm{tr}} \le \eta$. In particular the case $a_t = \infty$ cannot occur for such a surrogate.*

Proof. The band of Definition 1 holds at every set, hence at every state of the run, so the pair (η_u, η_o) is admissible for the run and the infimum in Definition 2 is at most η. Since $\eta \ge 1$, the truncation gives $\eta_{\mathrm{tr}} \le \eta$.

Let (η'_u, η'_o) be admissible for the run, fix a step t, and let $e^* \notin S^t$ maximize $d_e(S^t)$, so $M_t = d_{e^*}(S^t)$. Then

$$M_t \;\le\; \eta'_u\, \tilde{d}_{e^*}(S^t) \;\le\; \eta'_u\, \tilde{d}_{e_t}(S^t) \;\le\; \eta'_u\eta'_o\, d_{e_t}(S^t) = \eta'_u\eta'_o\, g_t,$$

the first inequality by the lower side of the band at (S^t, e^*), the second by the greedy choice of e_t, the third by the upper side of the band at (S^t, e_t). If $g_t > 0$ this reads $a_t = M_t/g_t \le \eta'_u\eta'_o$. If $g_t = 0$ it forces $M_t \le 0$, hence $M_t = 0$ because f is monotone, and $a_t = 1$. The case $g_t = 0 < M_t$ is therefore excluded, and in both remaining cases $a_t \le \max\{1, \eta'_u\eta'_o\}$. Taking the maximum over the K steps gives $\eta_{\mathrm{sel}} \le \max\{1, \eta'_u\eta'_o\}$, and taking the infimum over admissible pairs gives $\eta_{\mathrm{sel}} \le \eta_{\mathrm{tr}}$. □

F.2 The bound of Proposition 5

The inequality $f(T) \ge L_K(\eta_{\mathrm{sel}}) f(O^*)$ is proved in Appendix D.2, Steps 1 to 3, where the error enters only through the ratios a_t and a harmful zero step is handled by the convention $L_K(\infty) = 0$. Since L_K is nonincreasing on $[1, \infty]$, Lemma 23 gives $L_K(\eta_{\mathrm{sel}}) \ge L_K(\eta_{\mathrm{tr}}) \ge L_K(\eta)$, which is the chain displayed in the proposition. The per-step form of the same induction, which keeps the ratios a_t separate, is Remark 7.

F.3 THE CERTIFICATE IS EXACT FOR EVERY K

Theorem 4 (Tightness for every K)**.** *For every $K \geq 2$ and every $\hat{a} > 1$ there is an instance $(f, \tilde{f})$ on $2K$ elements with $\eta_{\mathrm{sel}} = \eta_{\mathrm{tr}} = \hat{a}$ on the adversarial-tie run of predictive greedy, whose output value is exactly $L_K(\hat{a})\, f(O^*)$. Hence Proposition 5 is an equality on this family for every K, not only in the limit.*

Proof. Fix $K \geq 2$ and $\hat{a} > 1$, and put

$$a = 1 - \frac{1}{\hat{a}K} \in (0,1), \qquad N = B \cup O, \qquad |B| = |O| = K, \qquad n = 2K.$$

For $S \subseteq N$ write $x = |S \cap B|$ and $y = |S \cap O|$, and define the pair $(f, \tilde{f})$ by

$$f(S) = F(x,y) = 1 - a^x\Big(1 - \frac{y}{K}\Big),$$
$$\tilde{f}(S) = G(x,y) = \begin{cases} 1 - a^x, & y = 0, \\ 1 - a^x + a^x\Big[(1-a) + \dfrac{(y-1)a}{K-1}\Big], & 1 \leq y \leq K. \end{cases} \tag{33}$$

Both are counting functions of (x, y), so Lemma 5 applies with the two blocks B and O. Note $F(0,0) = 0$ and $G(0,0) = 0$.

Step 1: a closed form for the second branch of G. For $1 \leq y \leq K$,

$$1 - a^x + a^x(1-a) + \frac{(y-1)a^{x+1}}{K-1} = 1 - a^{x+1} + \frac{(y-1)a^{x+1}}{K-1} = 1 - a^{x+1}\,\frac{K-y}{K-1}.$$

In particular $G(x, K) = 1$ for every x, and the two branches agree at $y = 1$: the second branch gives $1 - a^{x+1}$, which is $G(x,0) + a^x(1-a)$.

Step 2: the four gain identities. Writing Δ_x and Δ_y for the forward differences in x and in y, the four cases are computed directly from equation 33 and Step 1:

$$\Delta_x F(x,y) = a^x(1-a)\,\frac{K-y}{K}, \qquad \Delta_y F(x,y) = \frac{a^x}{K}, \tag{34}$$
$$\Delta_x G(x,0) = a^x(1-a), \qquad \Delta_y G(x,0) = a^x(1-a), \tag{35}$$
$$\Delta_x G(x,y) = a^{x+1}(1-a)\,\frac{K-y}{K-1}, \qquad \Delta_y G(x,y) = \frac{a^{x+1}}{K-1}, \tag{36}$$

the last row for $1 \leq y \leq K$ in the x direction and for $1 \leq y \leq K-1$ in the y direction. The two entries of equation 35 are the same number because both read $(1 - a^{x+1}) - (1 - a^x)$: the first crosses no branch, and the second crosses from $G(x,0)$ to $G(x,1) = 1 - a^{x+1}$ by Step 1. Dividing equation 35 and equation 36 by the corresponding entries of equation 34 gives the four ratios, with $R := aK/(K-1)$:

$$\frac{\Delta_x G(x,0)}{\Delta_x F(x,0)} = 1, \qquad \frac{\Delta_y G(x,0)}{\Delta_y F(x,0)} = \frac{a^x(1-a)}{a^x/K} = K(1-a) = \frac{1}{\hat{a}},$$

and, in the range $1 \leq y \leq K-1$ where both are defined,

$$\frac{\Delta_x G(x,y)}{\Delta_x F(x,y)} = \frac{a^{x+1}(1-a)(K-y)/(K-1)}{a^x(1-a)(K-y)/K} = R, \qquad \frac{\Delta_y G(x,y)}{\Delta_y F(x,y)} = \frac{a^{x+1}/(K-1)}{a^x/K} = R,$$

where $K(1-a) = 1/\hat{a}$ is the definition of a. At $y = K$ the only remaining direction is x, and both gains vanish: $\Delta_x F(x,K) = 0$ by equation 34 and $\Delta_x G(x,K) = 0$ by Step 1. So a zero true gain always comes with a zero predicted gain, as Definition 1 demands.

Step 3: monotonicity and submodularity of f. The two first differences in equation 34 are nonnegative for $0 \leq y \leq K$ and $x \geq 0$, since $a \in (0,1)$. The four second differences are

$$\Delta_x\Delta_x F = -a^x(1-a)^2\,\frac{K-y}{K} \leq 0, \qquad \Delta_y\Delta_x F = \Delta_x\Delta_y F = -\frac{a^x(1-a)}{K} \leq 0, \qquad \Delta_y\Delta_y F = 0.$$

With $F(0,0)=0$, Lemma 5 gives that f is normalized, monotone and submodular on the whole of 2^N. The surrogate $\tilde{f}$ is monotone, by equation 35 and equation 36, but it is not submodular:

$$\Delta_y G(x,1) - \Delta_y G(x,0) = a^x\Big(\frac{a}{K-1} - (1-a)\Big) > 0 \qquad \text{because } a > 1 - \frac{1}{K},$$

which $\hat{a} > 1$ guarantees. The model of Section 2 permits this, since it places no structural assumption on $\tilde{f}$.

Step 4: the global error of the pair. By Step 2 the set of single-element ratios $\tilde{d}_e(S)/d_e(S)$, over all S and all $e \notin S$ with $d_e(S) > 0$, is exactly $\{1, R, 1/\hat{a}\}$. Two sign computations order it:

$$R - 1 = \frac{aK-(K-1)}{K-1} = \frac{K - \frac{1}{\hat{a}} - K + 1}{K-1} = \frac{1-1/\hat{a}}{K-1} > 0, \qquad 1 - \frac{1}{\hat{a}} = \frac{\hat{a}-1}{\hat{a}} > 0,$$

both strict because $\hat{a} > 1$. So the largest ratio is R and the smallest is $1/\hat{a}$: the band of Definition 1 holds with $(\eta_u, \eta_o) = (\hat{a}, R)$, both factors attained, and with no pair of smaller product. The error of the pair is therefore

$$\eta = \eta_u \eta_o = \hat{a}\,\frac{aK}{K-1} = \frac{K(\hat{a} - \frac{1}{K})}{K-1} = \frac{K\hat{a}-1}{K-1}, \tag{37}$$

using $\hat{a}\,a = \hat{a} - 1/K$.

Step 5: the optimum. Since $1 - F(x,y) = a^x(K-y)/K \geq 0$, one has $F \leq 1$ everywhere, and $F(0,K) = 1$. As $|O| = K$, the value $f(O^*) = 1$ is attained at O, and $O^* = O$ may be taken.

Step 6: the greedy induction under adversarial ties. The claim is that for $0 \leq t \leq K$ the state after t steps satisfies $S^t \subseteq B$ and $|S^t| = t$, that is $(x,y) = (t,0)$. For $t = 0$ this is $S^0 = \emptyset$. Assume it at $t < K$. At the state $(t,0)$ every candidate is either a remaining element of B, whose predicted gain is $\Delta_x G(t,0) = a^t(1-a)$, or an element of O, whose predicted gain is $\Delta_y G(t,0) = a^t(1-a)$. These are equal by equation 35, so every candidate is a maximizer and the step is a tie among all $2K - t$ of them. Adversarial tie breaking selects an element of B, which is possible because $t < K = |B|$, so S^{t+1} has $(x,y) = (t+1,0)$. After K steps $S^K = B$ and

$$f(S^K) = F(K,0) = 1 - a^K = 1 - \Big(1 - \frac{1}{\hat{a}K}\Big)^K = L_K(\hat{a}),$$

which is $L_K(\hat{a}) f(O^*)$ by Step 5.

Step 7: the two run-dependent errors equal $\hat{a}$. At the state $S^t = (t,0)$ the true gains are $a^t(1-a) = a^t/(\hat{a}K)$ for an element of B and a^t/K for an element of O, by equation 34. Since $\hat{a} > 1$, the largest true gain at that state is a^t/K, while the chosen element has gain $a^t/(\hat{a}K) > 0$. Every step therefore has positive chosen gain, the stepwise a_t of Definition 2 reduce to the ratio form, and

$$a_t = \frac{\max_{e \notin S^t} d_e(S^t)}{d_{e_t}(S^t)} = \frac{a^t/K}{a^t/(\hat{a}K)} = \hat{a}$$

at every one of the K steps, so $\eta_{\text{sel}} = \hat{a}$. For η_{tr} only the states of the run matter, and at those states $y = 0$, so by Step 2 the only two ratios that occur are 1 in the direction x and $1/\hat{a}$ in the direction y. A pair (η'_u, η'_o) is admissible for the run exactly when $\eta'_o \geq 1$ and $\eta'_u \geq \hat{a}$, so the smallest admissible product is $\hat{a}$, and $\eta_{\text{tr}} = \max\{1, \hat{a}\} = \hat{a}$ as well. Comparing with equation 37,

$$\eta - \hat{a} = \frac{K\hat{a} - 1 - (K-1)\hat{a}}{K-1} = \frac{\hat{a}-1}{K-1} > 0,$$

so on this family the selection and trajectory rulers agree and are strictly smaller than the global one.

Step 8: the same family read at fixed global error. Solving equation 37 for $\hat{a}$ gives $\hat{a} = (\eta(K-1)+1)/K = k_1/K$, hence $a = 1 - \frac{1}{\hat{a}K} = 1 - \frac{1}{k_1} = q$ and

$$f(S^K) = 1 - a^K = 1 - \Big(1 - \frac{1}{\eta(K-1)+1}\Big)^K = U_K(\eta),$$

the curve of Appendix D.1. So the instances that make Proposition 5 an equality at $\eta_{\text{sel}} = \hat{a}$ realize only $U_K(\eta)$ when they are charged the global error η, which is strictly above $\rho_K(\eta)$ for $\eta > 1$ (Appendix D.7).

Steps 1 to 8 give the theorem: for every $K \geq 2$ and every $\hat{a} > 1$ the pair equation 33 on $2K$ elements has monotone submodular f, finite error, an adversarial-tie greedy run with $\eta_{\text{sel}} = \eta_{\text{tr}} = \hat{a}$, and output value exactly $L_K(\hat{a}) f(O^*)$. □

F.4 WHY THE RUN DOES NOT STOP EARLY

Remark 24 (The stopping variant)**.** *A variant of predictive greedy that halts as soon as the maximum predicted gain is zero obtains none of the bounds above; it is limited to the product bound of Remark 7 over the executed steps. Two elements already price the difference. Let $N = \{a, b\}$, let $f(S) = |S|$, which is modular, monotone and normalized, and let $\tilde{f}(S) = \min\{|S|, 1\}$. At $S^0 = \emptyset$ both predicted gains equal 1, and the tie is resolved in favour of a. At $S^1 = \{a\}$ the only remaining predicted gain is 0, so the stopping variant halts and returns $\{a\}$ with $f(\{a\}) = 1$ against $f(\{a, b\}) = 2$, the ratio $1/2$, while $L_2(1) = 3/4$. The executed step has $M_0 = g_0 = 1$ and therefore $a_0 = 1$ in Definition 2, so the selection error measured before stopping does not bound the ratio of a stopped run. Predictive greedy executes both steps, picks b at the second one, and returns the optimum.*

G PROOFS FOR SECTION 3.5: THE GROUND-SET THRESHOLD

Theorem 2 needs $n \geq K^4$ at one step only: after the transcript of the algorithm is fixed, the hidden optimum must be a K-set that no query contains twice. This appendix proves the estimate behind that step, records the property of the budget it uses, and locates the threshold between the two bounds that are known.

G.1 AN INDEPENDENT K-SET IN THE QUERY GRAPH

Lemma 24 (Independent K-set)**.** *Put $B_K = K^2(K-1)^2 + 3K - 1$.*

1. *Every graph on m vertices with e edges has an independent set of size at least*
$$\sum_v \frac{1}{\deg(v)+1} \;\geq\; \frac{m^2}{2e+m}. \tag{38}$$
2. *Let $K \geq 2$ and $n \geq B_K$, let N be a ground set of n elements, let $S_1, \ldots, S_M$ with $M \leq nK$ be subsets of N each of size at most K, and let $T \subseteq N$ with $|T| = K$. Then there is a K-set $O \subseteq N \setminus T$ with $|O \cap S_i| \leq 1$ for every i.*

Every $n \geq K^4$ qualifies, because $K^4 - B_K = K(2K-3)(K+1) + 1 > 0$.

Proof. (1) In a uniformly random ordering of the vertices, the vertices that precede all their neighbours form an independent set, and v belongs to it with probability $1/(\deg(v)+1)$, so the expected size of that independent set is $\sum_v 1/(\deg(v)+1)$ and some ordering attains at least this value; this is the Caro–Wei bound, and the argument is the one of the probabilistic lens on Turán's theorem in **?**. By Cauchy–Schwarz, $\sum_v \frac{1}{\deg(v)+1} \cdot \sum_v(\deg(v)+1) \geq m^2$, and $\sum_v(\deg(v)+1) = 2e + m$, which gives equation 38.

(2) On the vertex set $N \setminus T$, of size $m = n - K$, join two elements by an edge whenever some S_i contains both. Each S_i contributes at most $\binom{K}{2}$ edges, so $e \leq nK\binom{K}{2}$ and $2e \leq nK^2(K-1)$. By (1) there is an independent set of size at least $m^2/(2e+m)$, and this exceeds $K-1$ as soon as $m^2 > (K-1)(2e+m)$, for which it suffices that

$$(n-K)^2 > (K-1)\big(nK^2(K-1)+n-K\big), \qquad \text{that is} \qquad (n-K)(n-2K+1) > nK^2(K-1)^2,$$

where we used $(n-K)^2 - (K-1)(n-K) = (n-K)(n-2K+1)$. The difference of the two sides is

$$(n-K)(n-2K+1) - nK^2(K-1)^2 = n\,(n - B_K) + K(2K-1), \tag{39}$$

which is positive for every $n \geq B_K$; at $n = B_K - 1$ it equals $(K-1)^2(2-K^2) < 0$, so B_K is exactly where this estimate starts to work. Hence for $n \geq B_K$ the graph has an independent set of size at least K, and any K-subset O of it lies in $N \setminus T$ and meets every S_i in at most one element. Finally $K^4 - B_K = K^2\big(K^2 - (K-1)^2\big) - 3K + 1 = 2K^3 - K^2 - 3K + 1 = K(2K-3)(K+1) + 1 > 0$ for $K \geq 2$, so $n \geq K^4$ implies $n \geq B_K$. □

G.2 What the argument uses

Remark 25 (The budget property actually used)**.** *The only property of the budget used in the proof of Theorem 2 is the edge bound $e \le M\binom{K}{2}$ of Lemma 24: any budget M with $(n-K)(n-2K+1) > MK(K-1)^2$ is covered by the same choice of O, and $M = nK$ gives equation 39. By contrast, covering the budget nK on the scale of Theorem 3 needs $n^c \ge nK$, hence $c > 1$ and $\tau = \lceil c \rceil + 1 \ge 3$, and delivers only the weaker constant $H_{K,3}(\eta) > H_{K,1}(\eta)$; running the argument at the exact budget is what keeps the $\tau = 1$ constant. The step cannot be skipped for the instances of Lemma 15: by Lemma 16(2) their surrogate is not a function of $|S|$ once a set holds two optimal elements, so a query that meets O twice can distinguish the instance from the size profile.*

G.3 Where the threshold stands

Remark 26 (The two ends)**.** *Exhaustive search over all K-subsets stays within the budget nK exactly when $\binom{n}{K} \le nK$, that is for $n \le 5$ at $K = 2$ and for $n \le K + 2$ at $K \ge 3$, since $\binom{n}{K}/n = \binom{n-1}{K-1}/K$ increases with n. On every instance with error at most η the predicted argmax is worth at least $f(O^*)/\eta$ (Proposition 3), which exceeds $\rho_K(\eta)$ for $1 < \eta < K$ because $V_0(\eta) - V_1(\eta) = \frac{K-\eta}{K\eta k_1} > 0$ by equation 9. So the least n_0 from which $\alpha^*(\mathcal{A}_{\mathrm{lin}}) = \rho_K(\eta)$ satisfies $6 \le n_0 \le 9$ at $K = 2$ and $K + 3 \le n_0 \le B_K \le K^4$ at $K \ge 3$; here $B_2 = 9$ and $B_3 = 44$. The quartic exponent of B_K comes from nK queries, at most $\binom{K}{2}$ pairs each, and the need for an independent K-set. Whether n_0 really grows like K^4, or more slowly, is open.*

Remark 27 (Small n, $\eta = 1$ and $K = 1$)**.** *The restriction on n cannot be dropped from statements of this kind: at $n = 4$, $K = 2$, $\eta = 3/2$, querying all six pairs and returning the predicted argmax stays within the nK-query budget and guarantees $2/3 = 1/\eta$, above $\rho_2(3/2) = 3/5$, so for small n the class contains exhaustive search and no greedy-optimality reading survives. At $\eta = 1$ the conclusion of Theorem 2 holds by the route of Appendix E.3: the calibration reads $\bar{\theta} = 1$, where the family of Theorem 3 degenerates to $\tilde{f} = f$ with error exactly 1, and $U_K(1) = L_K(1) = \rho_K(1)$; that family's predictor is likewise a function of $|S|$ on balanced small sets, which is all that Lemma 24 and the proof of Theorem 2 use, so for $n \ge K^4$ predictive greedy is exactly optimal within $\mathcal{A}_{\mathrm{lin}}$ at $\eta = 1$ as well, continuously matching Theorem 2 as $\eta \downarrow 1$. The case $K = 1$ sits outside the theorem, which requires $K \ge 2$: there predictive greedy is exactly optimal among all deterministic algorithms with no budget restriction at all (Remark 6).*

H Experimental details

H.1 Protocol

Every run is one trajectory of predictive greedy on $\tilde{f}$ with $K_{\max}$ steps; smaller budgets are read as prefixes, since the steps of greedy do not depend on K. Ties are broken by element index. On the constructed instances the indices are assigned so that index order realizes the adversarial tie breaking of Section 3. At every visited state S_t the true and predicted gains of all remaining candidates are recorded; M_t is the largest true gain and g_t the true gain of the chosen element. The selection error follows Definition 2: $a_t = M_t/g_t$ when $g_t > 0$, $a_t = 1$ when $M_t = g_t = 0$, $a_t = \infty$ when $g_t = 0 < M_t$, and $\eta_{\mathrm{sel}} = \max\{1, a_0, \dots, a_{K-1}\}$ with $L_K(\infty) = 0$. Held-out accuracy and ROUGE-1 are not monotone, so a chosen element may have negative true gain; such a step is treated as $g_t = 0$. For these two families the paper also reports the finite-step diagnostic, the same maximum over the steps with $g_t > 0$, together with the number of runs whose stepwise value is infinite.

The trajectory error is measured on the visited states only. Over all candidate pairs $(d, \tilde{d})$ with $d \ge \varepsilon$ and $\tilde{d} \ge \varepsilon$, $\eta_{\mathrm{tr}} = \max\{1, \max d/\tilde{d}\} \cdot \max\{1, \max \tilde{d}/d\}$, where ε is one quantization unit of the objective: $1/n_{\mathrm{test}}$ for accuracy, one node for coverage, and 0.005 for ROUGE-1 F. Pairs below ε are dropped, so η_{tr} is a lower estimate and $L_K(\eta_{\mathrm{tr}})$ an upper estimate. For summarization the three surrogates have different scales, so before trimming $\tilde{d}$ is multiplied by $c = \max_e d_e(\emptyset)/\max_e \tilde{d}_e(\emptyset)$, one constant per article and surrogate; the constant affects η_{tr} only, never the run.

The denominator of every ratio is greedy on f. For accuracy and coverage it uses the exact maximizer at every step; for ROUGE-1 F it uses lazy evaluation, checked against the exact scan in Section H.5. Predictive greedy uses lazy evaluation where $\tilde{f}$ is submodular, that is for coverage and the three summarization objectives, and the exact maximizer otherwise. A ratio can exceed one on an intermediate prefix, because the denominator is the greedy prefix and not the optimum of that length: this happens on 346 of the 19,200 network rows (1.8 %, at most 1.0087) and on 2 of the 640 rows at $K = 30$.

H.2 Applications

Feature selection. The four datasets are listed in Table 3. The airline passenger satisfaction data are the public test split of the Kaggle dataset; the cleaning follows the earlier version of this work: the two identifier columns are dropped, the 83 missing arrival delays are replaced by the column mean, the five categorical columns are label-encoded, and three fixed thresholds (flight distance above 3736.5, departure delay above 800 minutes, arrival delay above 650 minutes) remove 601 rows, leaving 25,375 rows and 22 features. The other three datasets are the scikit-learn copies of breast cancer, wine and digits, the last restricted to its first 20 pixel features. Each split is `train_test_split` with test size 0.2 and random state equal to the split index $0, \ldots, 29$. $f(S)$ is the held-out accuracy of `DecisionTreeClassifier` with random state 42 trained on S; $\tilde{f}(S)$ is the mean five-fold cross-validation accuracy of the same estimator on the training split; $f(\emptyset)$ and $\tilde{f}(\emptyset)$ are the accuracies of the majority-class predictor. The surrogate is constructed from the training arrays alone, and a behavioural probe runs on every split: permuting the held-out labels changes f and leaves every value of $\tilde{f}$ bit-identical. Every run was repeated with `GradientBoostingClassifier` in place of the tree; the median ratio at $K = 7$ is at least as high on all four datasets (airline 1.000 against 0.999, breast cancer 0.974 against 0.955, digits 0.970 against 0.959, wine 0.958 against 0.944).

Table 3: Feature selection datasets. $\varepsilon = 1/n_{\text{test}}$ is one accuracy quantization unit.

Dataset	rows	features	n_{train}	n_{test}	ε
airline	25,375	22	20,300	5,075	2.0×10^{-4}
breast cancer	569	30	455	114	8.8×10^{-3}
digits	1,797	20	1,437	360	2.8×10^{-3}
wine	178	13	142	36	2.8×10^{-2}

Influence maximization. The four networks are the email-Eu-core network (**??**) and the artist, politician and government page networks of GEMSEC Facebook (**?**), all from SNAP. Their sizes are 1,005 nodes and 25,571 directed edges; 5,908 nodes and 41,729 edges; 7,057 nodes and 89,455 edges; and 50,515 nodes and 819,306 edges. $f(S)$ is the number of nodes in S or pointed to by S; the undirected networks are stored with both directions. An observed network keeps every input edge independently with probability $p \in \{0.2, 0.3, \ldots, 0.9\}$, an undirected edge as a whole; 20 observed networks are drawn per p with seeds $0, \ldots, 19$, and $\tilde{f}$ is the same count on the observed network. The observation model drops edges independently and does not model missing blocks of the network. Greedy on f reaches 831, 2,794, 3,827 and 15,777 nodes at $K = 30$ on the four networks.

Summarization. The articles are the first 100 files of the business, sport and technology categories of the BBC News Summary corpus (**?**); the corpus files carry reference summaries for business only, so the sport and technology references are taken from the CSV release of the same corpus, matched by full text; on business the two sources agree word for word on all 99 matched articles. One technology article matches two CSV rows with different summaries and is skipped, leaving 299. The title is the first sentence of the ground set, and the body is split at sentence-final punctuation; ground sets have 6 to 97 sentences, median 15, and K runs from 3 to $\min\{7, n-2\}$, so 299 articles enter at $K = 3$, 293 at $K = 5$ and 279 at $K = 7$. $f(S)$ is ROUGE-1 F of the concatenated sentences against the reference summary, computed without stemming or stopword removal; on 1,624 candidate summaries the implementation agrees with `rouge-score` to the last digit. The three surrogates use only the article: saturated coverage with saturation 0.25, the diversity reward with $\max\{2, \lfloor 0.2n \rfloor\}$

clusters from farthest-first traversal, and facility location; similarities are cosines of term-frequency vectors (Lin & Bilmes, 2011). The surrogates are monotone submodular by construction, and no violation appears in the enumeration of Section H.5.

H.3 THE OPTIMUM PROXY

On the breast cancer data the optimum can be enumerated up to $K = 4$ on every split. Table 4 gives the median over splits of $f(G)/f(O^*)$ and $f(T)/f(O^*)$. At $K = 4$ the proxy inflates the reported ratio by $1/0.982$, that is by 1.8 % in the median.

Table 4: Breast cancer, enumerated optimum, medians over 10 splits.

K	1	2	3	4
$f(O^*)$	0.886	0.956	0.974	0.987
$f(G)/f(O^*)$	1.000	1.000	0.982	0.982
$f(T)/f(O^*)$	0.966	0.973	0.959	0.946

H.4 BASELINES

Feature selection. Four classical selectors run on the training split with the same tree as downstream classifier: `SelectKBest` with the F-statistic, recursive feature elimination with the tree, mutual information, and extra-trees importances (**?**). Every method is scored by f on the same held-out split and normalized by greedy on f, over the same 30 splits. At $K = 7$ the median ratios of greedy on $\tilde{f}$ and of the best selector are 0.999 and 0.990 on airline, 0.944 and 0.944 on wine, 0.955 and 0.955 on breast cancer, and 0.959 and 0.963 on digits. Greedy on $\tilde{f}$ is at least as good as the best selector at all seven budgets on airline, at six on wine, at four on digits and at two on breast cancer; on the three small datasets the differences at $K = 7$ are at most 0.005.

Summarization. Lead-K takes the first K elements of the ground set, title included; the body-only variant starts at the first body sentence. At $K = 5$ their median ratios are 0.580 and 0.657, against 0.712, 0.662 and 0.628 for coverage, diversity and facility location. Coverage exceeds both variants at every K from 3 to 7. Diversity exceeds the body-only variant at $K \leq 5$ and trails it by less than 0.01 at $K = 6$ and 7; facility location exceeds it at $K \leq 4$ and trails it by up to 0.07 at $K = 7$. Table 1 prints the title-included variant.

Influence maximization. Two baselines use only observable information: the K nodes of largest out-degree in the observed network, one run per observed network, and K uniformly random nodes with ten fixed seeds. At $K = 30$ and $p = 0.5$ the degree ranking reaches 0.889, 0.679, 0.660 and 0.794 of greedy on f on email, politician, government and artist, and the random baseline 0.541, 0.156, 0.182 and 0.062; predictive greedy reaches 0.963, 0.942, 0.940 and 0.988.

H.5 OBJECTIVES OUTSIDE THE MODEL

For 30 articles (every tenth) the first eight sentences were enumerated: all sets $A \subseteq B = A \cup \{b\}$ with $|A| \leq 3$ and all elements e, 70,560 triples. ROUGE-1 F violates $d_e(A) \geq d_e(B)$ on 2.14 % of the triples, by at most 0.020, and violates $d_e(A) \geq 0$ on 7.12 %; the three surrogates violate neither. Held-out accuracy is not monotone either: at $K = 7$ with the tree, the share of chosen features with nonpositive true gain is 51.9 % on wine, 46.7 % on breast cancer, 8.1 % on digits and 0.0 % on airline, and the number of runs with an infinite stepwise η_{sel} is 25, 30, 14 and 0 of 30. For summarization at $K = 5$ the shares are 19.2 %, 12.0 % and 17.5 % for coverage, diversity and facility location, with 177, 131 and 168 of 293 runs infinite. Lazy evaluation of greedy on ROUGE-1 F, which is not submodular, is checked against the exact scan on every article: the two agree at $K \leq 5$ and differ on about 1 % of the articles at $K = 6$ and $K = 7$, by at most 0.026 in ROUGE-1 F, in the direction that lowers the denominator.

H.6 CONSTRUCTED INSTANCES

The instances of Theorem 1 are built for $K = 3$ at $\eta \in \{1.00, 1.05, \ldots, 3.00\} \cup \{3.5, 4, 5, 8\}$ and for $K = 4$ at $\eta \in \{1, 1.25, 1.5, 1.75, 2, 2.5, 3, 3.5, 4, 5\}$; at the integer breakpoints both adjacent members of the family are built, giving 60 runs. The instances of Remark 4 are built for $K \in \{3, 5, 7, 10, 30\}$ and $\hat{a} \in \{1.25, 1.5, 2, 3\}$, giving 20 runs. All instances are constructed in exact rational arithmetic with the error band split as $(\eta_u, \eta_o) = (1, \eta)$, and each is checked on its lattice before the run: monotone, submodular, inside the band, $d = 0$ exactly when $\tilde{d} = 0$, $f(O) = 1$ optimal among sets of size at most K, and greedy under adversarial ties following the intended trajectory. On every one of the 80 runs the realized ratio equals the closed form, η_{sel} equals η on the first family, and $\eta_{\mathrm{sel}} = \eta_{\mathrm{tr}} = \hat{a}$ on the second. The global error of the second family is $(\hat{a}K - 1)/(K - 1)$, larger than $\hat{a}$, so these instances are not drawn on the global axis of panel (a).

H.7 THE STEPWISE ADVERSARY

At state S_t on a real network the admissible set is $\{e : d_e(S_t) \geq d_{\max}/\eta\}$. Three selection rules were run: the worst rule picks the admissible element of smallest true gain, ties by index; the lookahead rule picks the admissible element that minimizes $f(S_t \cup \{e\}) + \max_{e'} d_{e'}(S_t \cup \{e\})$, on email and politician only; the random rule picks uniformly among admissible elements, with ten seeds. Each rule runs one trajectory of 30 steps per network and budget; $K = 5$ and $K = 10$ are prefixes. The budgets are $\eta \in \{1.25, 1.5, 2, 3\}$ for all three rules and $\eta \in \{1, 1.1, 1.25, 1.5, 1.75, 2, 2.5, 3, 4, 5\}$ for the worst and random rules, which panel (b) uses. On every one of the 1,320 rows the realized η_{sel} is at most η, and on every worst-rule row the ratio is at least $L_K(\eta)$. At $\eta = 1$ only ties are admissible and every ratio is 1.000. The worst rule takes the smallest admissible integer gain on 42.3 % of its steps, the lookahead rule on 52.1 %, the random rule on 15.0 %; the admissible set is a single element on 2.6 to 5.4 % of the steps. Lookahead is lower than the worst rule in 18 of 24 network, budget and K cells, by 0.00 to 0.08; random is higher by 0.01 to 0.15 and stays at or above 0.81 at $K = 30$. The worst rule is nearest the guarantee at small K and large η: at $K = 5$ and $\eta = 3$ its ratio is 0.385 on politician, 0.413 on artist, 0.499 on government and 0.714 on email against $L_5(3) = 0.292$; at $K = 30$ and $\eta = 5$ it is 0.452, 0.548, 0.623 and 0.811 on artist, politician, government and email against $L_{30}(5) = 0.182$.

H.8 RUNS BY OBSERVATION PROBABILITY

Table 5 gives the data behind panel (c): for every network and p, the median ratio and the certificate L_{30} at the median stepwise η_{sel} over the 20 observed networks. The certificate is zero where the median η_{sel} is infinite. The number of runs with an infinite η_{sel}, from $p = 0.2$ to 0.9, is 4, 2, 6, 3, 2, 1, 0, 0 on email; 2, 1, 0, 0, 0, 0, 0, 0 on politician; 1, 0, 2, 10, 12, 0, 0, 0 on government; and 0 on artist throughout.

Table 5: Influence maximization at $K = 30$: median ratio and certificate L_{30} at the median η_{sel}, 20 observed networks per cell.

		p							
		0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
email	ratio	0.928	0.945	0.952	0.963	0.972	0.980	0.989	0.993
	cert.	0.069	0.095	0.095	0.125	0.182	0.285	0.361	0.455
politician	ratio	0.867	0.899	0.922	0.942	0.966	0.981	0.988	0.997
	cert.	0.024	0.047	0.046	0.082	0.262	0.370	0.413	0.554
government	ratio	0.831	0.882	0.911	0.940	0.955	0.980	0.990	0.996
	cert.	0.008	0.008	0.008	0	0	0.299	0.510	0.553
artist	ratio	0.930	0.963	0.983	0.988	0.992	0.995	0.997	0.999
	cert.	0.003	0.230	0.403	0.476	0.500	0.541	0.573	0.601

H.9 THE PER-STEP CERTIFICATE

For every network run the per-step certificate of Remark **??**, $1 - \prod_{t \le K}(1 - 1/(a_t K))$ with the factor 1 for $a_t = \infty$, was computed from the recorded M_t and g_t of the K steps. On all 19,200 rows the ratio is at least the per-step certificate, and the per-step certificate is at least $L_K(\eta_{\mathrm{sel}})$, up to the rounding of the printed values. Table 6 gives the medians at $K = 30$; pooled over the four networks at $p = 0.5$ the median is 0.536, against 0.105 for L_{30} at the median η_{sel}, and it falls from 0.636 at $K = 5$ to 0.536 at $K = 30$ while L_K at the median η_{sel} falls from 0.524 to 0.105. The 46 runs with a zero-gain step beside a positive candidate, whose $L_{30}(\eta_{\mathrm{sel}})$ is zero, have per-step certificates from 0.39 to 0.56. On the constructed instances of Remark 4 every a_t equals $\hat{a}$, so the per-step certificate, $L_K(\hat{a})$ and the realized ratio coincide.

Table 6: Influence maximization at $K = 30$: median per-step certificate, 20 observed networks per cell.

p	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
email	0.392	0.432	0.471	0.495	0.510	0.542	0.573	0.598
politician	0.429	0.482	0.506	0.540	0.570	0.596	0.610	0.627
government	0.403	0.452	0.498	0.533	0.557	0.591	0.612	0.625
artist	0.502	0.544	0.574	0.594	0.608	0.617	0.626	0.634

H.10 REPRODUCIBILITY

All runs are deterministic given the seeds stated above. Every result file is regenerated by one script per experiment, and every number in Section **??**, Table 1 and this appendix is read from those files. Wall-clock times on one workstation: the 640 network runs take 691 s on three processes; the 120 tree runs 451 s and the 120 gradient-boosting runs about 4.4 hours; the summarization runs 6 s; the constructed instances 1 s; the adversary 91 s. No run was skipped, subsampled or truncated. The network rows at $p \in \{0.3, 0.5, 0.8\}$, the summarization rows and the tree rows reproduce the rows of the first-round pipeline exactly (7,200, 4,377 and 840 rows). Code and data will be released with the paper.