

A Curvature Criterion for Cannibalization Feedback

One inequality decides whether scarcity feedback destroys diminishing returns, and for a propensity derived from mechanism it reduces to a condition on logistics timing

Dayan Ibrar

THEORY, WITH ONE MEASURED SECTION · PREPRINT, NOT PEER REVIEWED

Theory, with one measured section. Every closed-form statement here is proved for an explicitly stated model and independently re-evaluated in fifty-digit arithmetic; Section 7's ratios are measured against ex-

haustive optima. No claim here is an observation of any fleet. Every figure carries its kind: conceptual, illustrative computation, or measured.

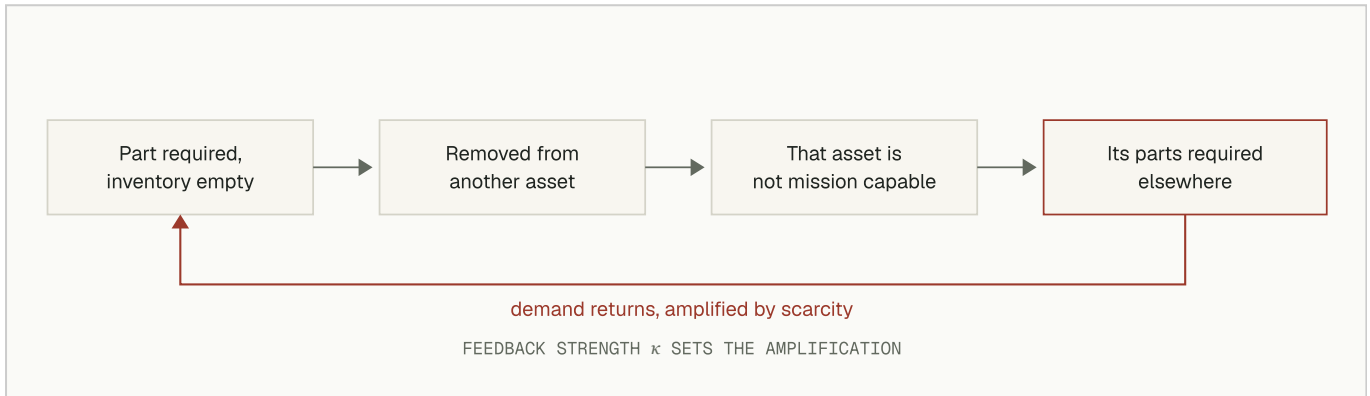


FIGURE 1 · CONCEPTUAL

The mechanism behind the collapse. A shortfall does not simply go unmet; it creates further demand, which is why the response to capacity is not proportional. This is the structure Section 3.4 parameterises, not an observation of any unit.

Abstract

Sustainment organizations that maintain fleets of repairable assets face a mechanism ordinary scheduling theory omits. When a part is needed and inventory is empty, maintainers remove it from another asset, which disables that asset, which creates further demand. The practice is documented across the services and inconsistently tracked [15]. A shortfall therefore creates demand, and positive feedback of that kind does not give a proportional response.

This note asks a single question about such a model: does the feedback destroy the diminishing-returns structure that approximation algorithms for scheduling rely on? The answer is a one-line inequality, and the main content is that it is the *same* inequality under every balance law we consider, with only a threshold moving.

Write the balance between uncovered failure mass and the disabled fraction, let $b(x)$ be the share of disabled assets disabled by someone else's shortage, and let the schedule enter through a coverage function. Then expected readiness is a scalar response composed with that coverage function, and the response is concave — so readiness is submodular and greedy keeps its guarantee — if and only if

$$2b'(x) + x b''(x)$$

where x , disabled fraction, X/n .

clears a threshold set by the balance law: **zero** under mean-field balance, and $2/(1-x)^3$ once a removal is required to find a serviceable donor. For a propensity derived from the mechanism rather than assumed, $b(x) = \beta_0 e^{-\lambda x} + a x$, the mean-field condition collapses to the closed form $a \geq \beta_0 \lambda$, which is necessary and sufficient and is decided entirely at the origin. Since the transit coefficient a is proportional to the ratio of part-removal time to repair time, **the structural question is a logistics-timing question and not a behavioural one** — and it is decided in the direction that makes greedy safe exactly when cannibalization is a net loss of readiness.

Two negative results bound this. The transfer of submodularity through the composition is classical [12], and the bound relating the submodularity ratio to a ratio of extreme response slopes is an instance of a published implication [5], [6] — we state it because we need the parameter as a property of the response, and we show that where this note applies it, a curvature result from 2017 dominates it under a weaker constraint [10]. And submodularity fails for a reason unrelated to the feedback: mission requirements are conjunctive, and when readiness is the probability that at least m of n assets are simultaneously capable, early

inductions show increasing returns exactly when $m - 1 > (n - 1)q_{lo}$.

Finally, measured against exhaustive optima, greedy clears the constant appropriate to its feasible region across all three ways the structure fails here. That is the practically useful finding and it is the one we did not predict.

Code and data. Everything here is reproducible from <https://github.com/DayanEbrar0X/readiness-critical-point>, whose *REPRODUCE.md* maps each number to the single test that asserts it. The criteria are implemented twice over, in Rust and in fifty-digit Python, because one implementation agreeing with itself is evidence about nothing.

The evidence is in the main paper. The simulator that measures these criteria against a fleet, and the three panels of published operational statistics that estimate the sign they turn on, are in *Readiness at the Critical Point*, at <https://www.dayanibrar.com/papers/readiness-at-the-critical-point>. This note is self-contained as theory and makes no empirical claim that paper does not carry the evidence for.

1. The question, and why its answer is narrow

1.1 What the feedback does

An organization keeps a fleet of assets mission capable. It has a fixed number of depot slots per period, a spare inventory of finite depth, and assets that degrade under use. The standard treatment is a scheduling problem: predict which assets will fail, induct them before they do, subject to capacity.

That treatment omits the dominant dynamic. When a part is needed and the shelf is empty, it is frequently removed from another asset. The second asset is then not mission capable. If its part is later needed elsewhere, the process repeats. The practice is documented across the services, and the Government Accountability Office has reported it as both widespread and inconsistently tracked, with one service not recording it at all [15].

The structural point is that the consequence of a shortfall is additional demand. That is positive feedback, and it does not produce a proportional response. It produces one regime in which capacity changes barely move readiness, and another in which they move it enormously.

1.2 The only question this note answers

Scheduling algorithms for problems of this shape rely on **diminishing returns**: the marginal value of inducting one more asset falls as more are already inducted. Formally, that the objective is a monotone submodular set function, which buys $1 - 1/e$ for greedy under a cardinality constraint [1], $1/2$ for a single matroid and $1/(p + 1)$ for an intersection of p matroids [2], and $1 - 1/e$ for a single matroid via continuous greedy [3]. Those constants differ, and which applies depends on the feasible region, so we are careful below about which one is in play. The cardinal-

ity constant is optimal under $P \neq NP$ by Feige's inapproximability result for maximum coverage [19].

So: **does the cannibalization cascade destroy diminishing returns?**

The expectation this work began with was that it does, dramatically — that a cascade is the canonical way for marginal returns to grow rather than shrink. The algebra says otherwise, and says so in a form sharp enough to be useful: whether the structure survives is decided by the sign of one quantity, which can be computed from a measured propensity curve without simulating anything.

1.3 What this note claims, and what it does not

Four things, each marked by how it is established.

1. **Expected readiness is a scalar response composed with a coverage function** (Section 3, *derived*), so every structural claim about the scheduling problem is a claim about the curvature of one function of one variable.
2. **A family of exact curvature criteria** decides submodularity for any propensity under either balance law (Propositions 1 and 2, Section 5, *proved*). The left-hand side is always $2b'(x) + x b''(x)$; only the threshold moves.
3. **For a propensity derived from the mechanism the criterion has a closed form**, $a \geq \beta o \lambda$, necessary and sufficient and decided at the origin (Proposition 4, Section 5, *proved*). This is the result we think is worth having: it converts a question about curvature into a question about how long a part removal takes relative to a repair.
4. **Greedy is robust past where its guarantees stop** (Section 7, *measured against exhaustive optima*). Across all three structural failures identified here it clears the constant appropriate to its feasible region, which we did not expect and did not look for.

What this note does **not** claim. It does not claim the model describes a fleet; the companion paper measures several of its hypotheses as unmet, and Section 8 lists which. It does not claim priority for the composition result or for the ratio bound, both of which are in the literature and are cited to it. And it does not claim that the criteria predict a readiness curve: the companion paper measures a second source of curvature they do not account for, and Section 8 states the consequence.

2. Related work

Submodular maximization. Nemhauser, Wolsey and Fisher established the $1 - 1/e$ guarantee for greedy under a cardinality constraint [1]; the companion paper gives $1/2$ for a single matroid and $1/(p + 1)$ for an intersection of p matroids [2]. Călinescu, Chekuri, Pál and Vondrák achieved $1 - 1/e$ for a single matroid with continuous greedy and pipage rounding [3]. Feige's inapproximability for maximum coverage makes the cardinality constant optimal [19]. The distinctions matter here be-

cause Section 7 measures against three different regions and the constants differ.

Composition. That a nondecreasing concave function composed with a monotone submodular function is monotone submodular is classical, stated as Theorem 1 of Lin and Bilmes [12]. Lemma 1 below is that result, restated on a bounded interval, and we claim no credit for it. For a *modular* inner function the equivalence holds with its converse — concavity is necessary as well as sufficient — as Theorem 4.5.1 of Bilmes's course notes [13], which is the direct citation for the first caution in Section 4.

Approximate submodularity, and the priority of Theorem 1.

Das and Kempe introduced the submodularity ratio γ and proved a $1 - e^{-\gamma}$ guarantee under a cardinality constraint — Theorem 6 of [4], stated for the ratio taken over the greedy set, which the global γ lower-bounds. Theorem 1 below bounds γ by a ratio of extreme response slopes, and **it is not new**. El Halabi, Bach and Cevher call a set function ρ -submodular when every marginal on a larger set is at least ρ times the same marginal on a subset, and prove that ρ -submodularity implies $\gamma \geq \rho$ — Definition 4 and Proposition 8 of [5], which uses the same letter. Bogunovic, Zhao and Cevher state the implication as $\gamma \geq 1 - \tilde{\alpha}$ for an inverse generalized curvature (Proposition 1 of [6]). All Theorem 1 adds is the mean-value-theorem observation that a composition is ρ -submodular with $\rho = \inf h' / \sup h'$, which we have not found written down and which nobody would call a theorem. We state it because we need the parameter as a property of the response.

Bounds of that shape are standard: Elenberg, Khanna, Dimakis and Negahban prove $\gamma \geq m/M$ for a restricted-strongly-concave, restricted-smooth likelihood (Theorem 1 of [7]), and Bian, Buhmann, Krause and Tschitschek sharpen the greedy constant to a tight $(1/\alpha)(1 - e^{-\alpha\gamma})$ in terms of γ and a generalized curvature (Theorem 1 of [8]). Neither treats a composition. The nearest is Proposition 2b of [8], which bounds γ for a determinantal objective that is an exponential of a submodular function — a convex transformation of exactly our kind — but by eigenvalue interlacing rather than by any property of the exponential.

Curvature, which is the refinement this question actually needs. Where the response is concave, Lemma 1 already gives submodularity and the relevant quantity is curvature rather than γ . Conforti and Cornuéjols's total curvature improves the greedy constant to $(1 - e^{-c})/c$ under a cardinality constraint and $1/(1 + c)$ for a single matroid [9]; Sviridenko, Vondrák and Ward obtained the optimal $1 - c/e$ and defined a generalized curvature for arbitrary monotone functions [10]. Iyer, Jegelka and Bilmes compute curvature for precisely the concave-over-modular class and are explicit that curvature and the submodularity ratio are distinct quantities [11]. Section 6 records a consequence against this note: when the coverage function is modular, the slope ratio of Theorem 1 is *exactly* the complement of the generalized curvature of [10], so Theorem 1 restates curvature theory there rather than adding to it.

Cascading failure. Positive feedback whose offspring mean rises with system loading is the shape of the power-systems cascade literature, and the closest prior art to the model of Section 3 is Dobson, Carreras, Lynch and Newman's branching-process treatment of blackouts [18]. Our feedback mechanism is the same shape and we claim no novelty for it. The contribution here is connecting it to an optimization guarantee.

Inventory and cannibalization. Sherbrooke's METRIC model finds multi-echelon repairable inventory and, like most of that literature, treats demand as exogenous [17]. The feature driving the analysis here is that demand is endogenous. On cannibalization specifically, and this matters for how the results should be read: Fisher and Brennan simulated competing cannibalization policies under spares, repair and resource constraints and found the question to be a policy comparison rather than a fixed hazard, with the outcome turning on how long a removal takes relative to a repair [16]. **That is the same dial Proposition 4 below identifies**, reached from mechanism rather than from simulation, which we regard as the strongest external support any result here has. It also means the model below treats as a pure loss mechanism something that literature reports as having a benefit side, and that is a restriction rather than a finding.

3. The model, and the reduction to one variable

3.1 Objects

A fleet of n assets. Each asset is serviceable or not. A schedule S is a set of maintenance inductions drawn from a ground set of candidate actions, and the feasible schedules form the intersection of two partition matroids — one bounding depot throughput per period, one bounding each asset to a single induction. Nothing below depends on the matroid structure except Section 7's constants, so we keep it in the background.

3.2 Coverage

An induction protects an asset against a set of anticipated failure causes, and different inductions may protect against overlapping causes. Write $Q(S)$ for the failure mass covered by schedule S , normalized so $Q(\emptyset) = 0$.

Q is monotone submodular, not modular, and the distinction is load-bearing twice over. It is submodular because two inductions protecting against the same part lot do not protect twice; and the overlap is the entire source of combinatorial difficulty, since without it the problem is a transportation problem solvable in polynomial time. Everything from Lemma 1 onward assumes Q monotone submodular.

Write $P(S) = Q(G) - Q(S)$ for the uncovered mass, where G is the full ground set, and $\pi = P(S)/n$ for the uncovered mass per asset. Then π is affine decreasing in $Q(S)$, by definition.

3.3 The feedback

Let x be the disabled fraction of the fleet. Let $b(x)$ be the share of disabled assets disabled because a part was taken from them — the *induced* share — as against disabled by their own failure.

The uncovered mass π drives primary disablement; the induced share adds to it.

Mean-field balance. If every cannibalization event disables a fresh asset, the disabled fraction satisfies

$$x(1 - b(x)) = \pi \quad (1)$$

where x , disabled fraction, X/n ; π , uncovered failure mass per asset, $P(S)/n$.

so the disabled fraction consistent with uncovered mass π is the solution of (1).

Donor-corrected balance. A removal needs a donor that still holds the part, and a recipient that is not already down. Both conditions scale with the capable fraction, which multiplies both terms:

$$(\pi + b(x)x)(1 - x) = x \quad (2)$$

where x , disabled fraction, X/n ; π , uncovered failure mass per asset, $P(S)/n$.

A propensity derived from the mechanism. Neither balance law says what b is. Two competing effects do. As the fleet degrades, the assets that are *already down* become the obvious donors, so a removal stops costing a flyer — a self-limiting

channel that decays with x . Against that, a part in transit keeps its recipient down for as long as the move takes — a transit channel that grows with x . Writing the first as exponential decay and the second as linear growth:

$$b(x) = \beta_0 e^{-\lambda x} + a x \quad (3)$$

where β_0 , baseline propensity for one failure to induce another in a healthy fleet; x , disabled fraction, X/n .

with β_0 the propensity at $x = 0$, λ the rate at which donor scarcity bites, and a the transit coefficient, which is proportional to the ratio of part-removal time to repair time. Two limits recover known cases: $\lambda \rightarrow 0$ removes donor scarcity and leaves the linear law $b(x) = \beta_0(1 + \kappa x)$ with $\kappa = a/\beta_0$, and $a \rightarrow 0$ leaves pure self-limitation.

3.4 Readiness is a scalar response composed with coverage

Let $x^*(\pi)$ be the relevant root of the balance law in use, and let readiness be the expected capable fraction, $R = 1 - x^*$. Composing with the coverage map,

$$R(S) = h(Q(S)), \quad h(q) = 1 - x^*((Q(G) - q)/n)$$

where x^* , stable disabled fraction, the smaller root of equation (2); q , covered failure mass, the argument of the response h ; $Q(S)$, covered failure mass; monotone, normalized and submodular; $R(S)$, expected mission-capable fraction under schedule S ; h , scalar response: readiness as a function of covered mass; n , fleet size; S , a schedule; in the submodularity-ratio definition, the disjoint test set.

so every structural claim about the scheduling problem is a claim about the curvature of h , a function of one variable. That is the reduction the rest of this note depends on, and it is

worth being explicit about what it assumes: that the schedule affects the equilibrium only through π . Section 8 reports that the companion paper's simulator violates this, and why.

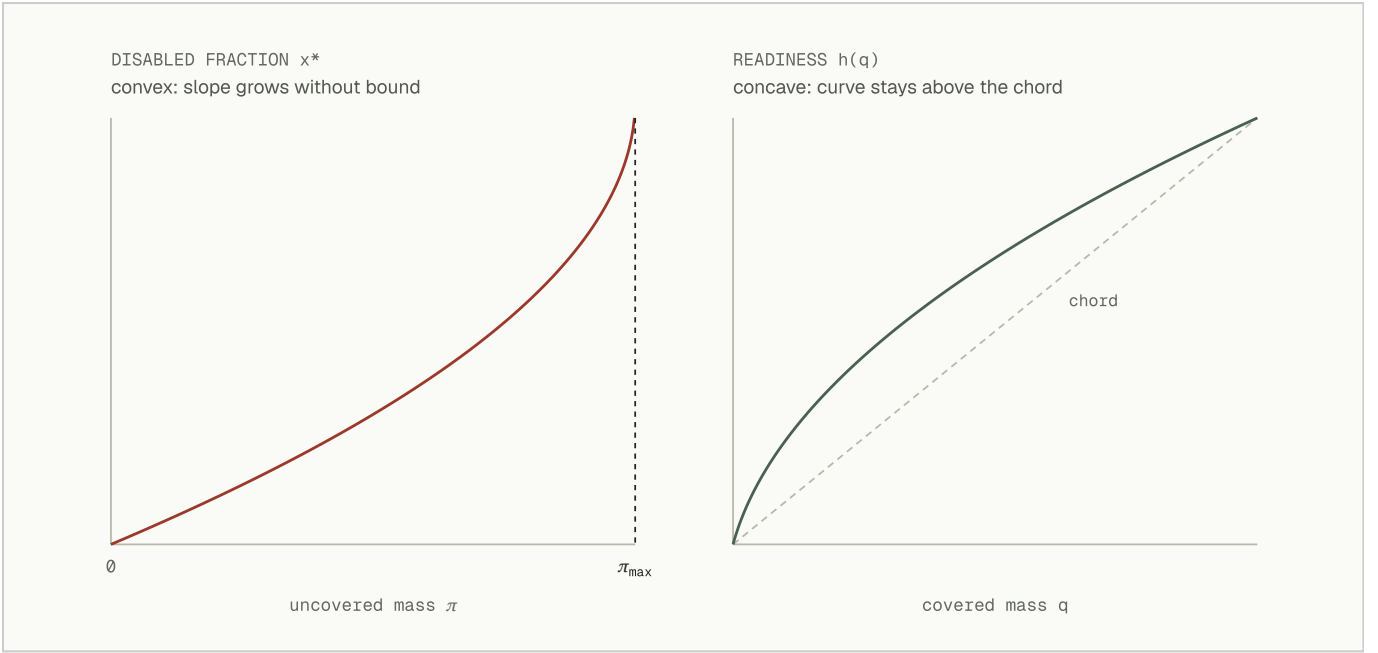


FIGURE 3 · ILLUSTRATIVE COMPUTATION

Why the cascade does not break submodularity under the mean-field law. Computed from equation (2) with baseline propensity 0.45 and feedback strength 2.2. Disabled fraction is convex in uncovered mass and its slope diverges at the fold. Readiness is that quantity negated under a decreasing reparameterisation, which flips the curvature, so Lemma 1 gives submodularity. Two caveats the figure cannot show. At these same parameters Proposition 4 says the donor-corrected law is convex and Section 10 says it has no fold at all. And the reparameterisation is affine only under the coverage map of Section 5.1, which Section 9 measures as false.

4. Concavity of the response transfers submodularity

Lemma 1. Let $f(S) = h(v(S))$ with v monotone, normalized and submodular, and h nondecreasing and concave on an interval containing the range of v . Then f is monotone submodular.

Proof. For $L \subseteq M$ and $a \notin M$, write $\delta_L = v(L \cup a) - v(L)$ and $\delta_M = v(M \cup a) - v(M)$. Submodularity of v gives $\delta_L \geq \delta_M$, and monotonicity gives $v(L) \leq v(M)$. The marginals of f are $h(v(L) + \delta_L) - h(v(L))$ and $h(v(M) + \delta_M) - h(v(M))$. Shorten the first increment to δ_M and then move it right: concavity of h gives that an increment of fixed length is nonincreasing in its left endpoint, and nondecreasingness gives that shortening an increment does not increase it. Both steps evaluate h inside $[v(L), v(M \cup a)]$, which is inside the stated interval. Hence the first marginal is at least the second. ■

Two cautions about the converse, because both are easy to get wrong.

First, **monotonicity of v is necessary and is not implied by nonnegativity**. Take the ground set $\{a, b\}$ with modular v given by $v(\emptyset) = 0$, $v(a) = 1$, $v(b) = -0.9$, $v(ab) = 0.1$, which is not monotone, and $h = \sqrt{\cdot}$. Then $f(a) + f(b) = 1 + 0 = 1$ while $f(ab) + f(\emptyset) = \sqrt{0.1} = 0.316$, so submodularity holds here; but reversing the signs to $v(\emptyset) = 10$, $v(a) = 11$, $v(b) = 7$, $v(ab) = 8$ — all nonnegative — gives $f(a) + f(b) = 5.962$ against $f(ab) + f(\emptyset) = 5.991$, and submodularity fails. The hypothesis is monotone, not nonnegative. For a modular inner function, concavity of h is also *necessary*, which is Theorem 4.5.1 of [13].

Second, **submodularity of v is equally necessary**, and it is a separate hypothesis rather than a consequence. Take $v(\emptyset) = 0$, $v(a) = v(b) = 1$, $v(ab) = 10$, which is monotone, normalized and supermodular, with $h = \sqrt{\cdot}$ increasing and concave. Then $f(a) + f(b) = 2 < 3.162 = f(ab) + f(\emptyset)$. Both hypotheses carry weight; neither can be dropped.

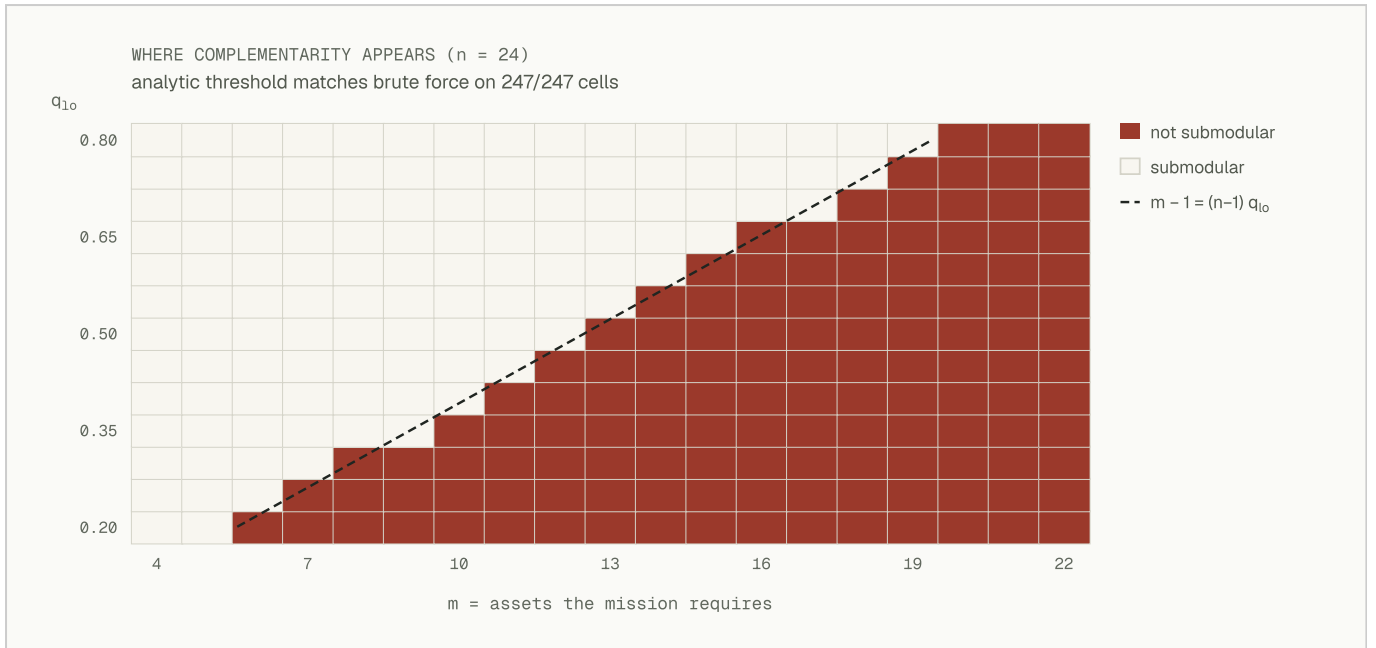


FIGURE 6 • ILLUSTRATIVE COMPUTATION

Proposition 2 as a map. Filled cells are parameter settings where some marginal increases, so the conjunctive objective is not submodular; in empty cells every marginal decreases. The check is exhaustive over all covered counts, not just the first induction. The dashed line marks equality in the analytic threshold $m - 1 > (n - 1) q_{10}$, which concerns the first induction alone, and it agrees with the exhaustive check on all 247 cells shown. The weaker condition that m lies between the expected capable counts is not sufficient, and the cells between the two conditions are exactly where that fails.

5. The criteria

Everything in this section is a statement about the curvature of h , and the point is that one quantity appears in all of them.

Proposition 1 (mean-field balance). Under balance (1), with $g(x) = x(1 - b(x))$ increasing on the branch of interest, the response h is concave if and only if

$$2b'(x) + x b''(x) \geq 0$$

where x , disabled fraction, X/n .

Proof. Differentiating g twice gives $g'' = -2b'(x) - x b''(x)$. On an increasing branch $x^* = g^{-1}$, so $d^2x^*/d\pi^2 = -g''/(g')^3$, and $h = 1 - x^*$ composed with the affine map $\pi = (Q(G) - q)/n$ gives $h'' = (1/n^2) g''/(g')^3$. Since $g' > 0$ on the branch, $h'' \leq 0$ if and only if $g'' \geq 0$, which is the stated inequality. ■

For the linear law $b(x) = \beta o(1 + \kappa x)$ this reduces to $2\beta o \kappa \geq 0$, so under mean-field balance the whole question is the **sign of κ** : does the propensity to cannibalize rise with the disabled fraction?

Proposition 2 (donor-corrected balance). Under balance (2), h is concave if and only if

$$2b'(x) + x b''(x) \geq 2/(1 - x)^3$$

where x , disabled fraction, X/n .

Proof. Solving (2) for π gives $\pi(x) = x/(1 - x) - b(x)x$. Then $\pi' = 1/(1 - x)^2 - b - x b'$ and $\pi'' = 2/(1 - x)^3 - 2b' - x b''$. The response is the inverse composed with an affine map, so concavity of h is $\pi'' \geq 0$, which rearranges to the stated inequality. ■

The same left-hand side, a different threshold, and the threshold is at least two — since $2/(1 - x)^3 \geq 2$ on $[0, 1)$. For the linear law the condition becomes $\beta o \kappa (1 - x)^3 \geq 1$. At $\beta o = 0.45$ and $\kappa = 2.2$ the left side at $x = 0.2$ is 1.98 against a threshold of 3.906 , so requiring a serviceable donor **reverses** the verdict at those parameters. That is a near miss rather than a rout, and we say so: a slightly stronger feedback puts it back on the submodular side.

Proposition 3 (a decreasing propensity breaks the transfer, and what replaces it). If $2b'(x) + x b''(x) < 0$ on $[0, x_{hi}]$ — which the linear law with $\kappa < 0$ satisfies — then h is convex there, Lemma 1 does not apply, and f need not be submodular. If in addition the coverage function is **modular**, f is supermodular. With overlapping coverage it is in general neither.

Proof. Convexity is immediate from Proposition 1, whose criterion fails at every point. A convex increasing function of a modular function is supermodular. For the third claim, let two inductions cover the same mass, so $v(a) = v(b) = v(ab)$ with $v(\emptyset) = 0$; then the supermodular inequality reduces to $h(v(a)) \leq h(0)$, which fails because h increases. ■

Proposition 4 (the closed form, and the result of this note). Under mean-field balance with the mechanism-derived propensity (3), the response is concave across the whole operating range if and only if

$$a \geq \beta_0 \lambda$$

where β_0 , baseline propensity for one failure to induce another in a healthy fleet.

Proof. From (3), $b' = a - \beta_0 \lambda e^{-\lambda x}$ and $b'' = \beta_0 \lambda^2 e^{-\lambda x}$, so

$$2b' + x b'' = 2a + \beta_0 \lambda e^{-\lambda x} (\lambda x - 2)$$

where β_0 , baseline propensity for one failure to induce another in a healthy fleet; x , disabled fraction, X/n .

and Proposition 1's condition is $2a \geq G(x)$ with $G(x) = \beta_0 \lambda e^{-\lambda x} (2 - \lambda x)$. Differentiating, $G' = \beta_0 \lambda^2 e^{-\lambda x} (\lambda x - 3)$, which is negative on $[0, 3/\lambda)$, so G falls there; beyond $2/\lambda$ we have $G \leq 0$ and the condition holds for free; and on $(3/\lambda, \infty)$ the function G turns back **up** but remains nonpositive, so it never again approaches its value at the origin. Hence $\sup G = G(0) = 2\beta_0 \lambda$ and the condition over the whole range is $2a \geq 2\beta_0 \lambda$. ■

The step most easily waved through is the last one — that G rising again past $3/\lambda$ is harmless because it is negative there — so

$$\pi \leq \pi_{\max} = (1 - \beta_0)^2 / (4 \beta_0 \kappa), \quad \text{at which} \quad x_c = (1 - \beta_0) / (2 \beta_0 \kappa)$$

where $\pi_{\max}$, critical uncovered mass, the fold beyond which no healthy equilibrium exists; β_0 , baseline propensity for one failure to induce another in a healthy fleet; x_c , disabled fraction at the fold, where the two equilibria collide; κ , feedback strength: how much scarcity raises the chance of the next removal; π , uncovered failure mass per asset, $P(S)/n$.

Below $\pi_{\max}$ the smaller root is the relevant one, $dx^*/d\pi = \Delta^{-1/2}$ with $\Delta = (1 - \beta_0)^2 - 4\beta_0 \kappa \pi$, and the response is concave in π . At $\pi_{\max}$ the discriminant vanishes and the two roots collide: a saddle-node bifurcation.

Proof. Direct from the quadratic formula, with $\Delta = 4\beta_0 \kappa (\pi_{\max} - \pi)$. Under the flow $\dot{x} = \pi - g(x)$ the derivative at the smaller root is $-\sqrt{\Delta} < 0$, so it is the attracting one up to the fold. ■

Two restrictions on Proposition 5 that are easy to miss. The fold is at a disabled *fraction*, so it is physically meaningful only when $x_c \leq 1$, which needs $\kappa \geq (1 - \beta_0)/(2 \beta_0)$; below that, readiness is already negative before the tipping point is reached and the bifurcation belongs to the equations rather than to any fleet. And $\pi_{\max}$ can exceed one, in which case it is unreachable because π is a fraction. Both are stated because the companion paper's measured parameters fall in exactly those regions.

Theorem 1 (the submodularity ratio is bounded by the slope ratio). Let $f(S) = h(v(S))$ with v monotone, normalized and submodular, and h differentiable and strictly increasing on an interval containing the range of v . Let $\rho = \inf h' / \sup h'$ over that interval. Then the submodularity ratio of [4] satisfies $\gamma \geq \rho$.

Proof. Fix L and disjoint S . For each $a \in S$ the mean value theorem gives $f(L \cup a) - f(L) = [v(L \cup a) - v(L)] h'(\xi_a) \geq [v(L \cup$

it is checked numerically rather than asserted: scanning G over $[0, 40/\lambda]$ in fifty-digit arithmetic returns its maximum at the origin, equal to $2\beta_0 \lambda$ exactly.

Why this is the result worth having. The condition is **global but decided at a single point**, so there is no operating range to check. And a is proportional to the ratio of part-removal time to repair time, so the inequality says:

> Greedy keeps its guarantee when a part removal is slow relative to a repair.

That is a statement about logistics timing, not about maintainer behaviour, and it is the same dial Fisher and Brennan identified in 1986 by simulation [16]. It also points the uncomfortable way: slow removals are the regime in which cannibalization *costs* readiness, so the structure survives exactly where the practice is harmful and fails where it helps.

Proposition 5 (the fold). Under mean-field balance with the linear law and $\kappa > 0$, equation (1) is quadratic in x and has a real root only for

$a) - v(L)] \cdot \inf h'$. Submodularity of v gives $\sum_{a \in S} [v(L \cup a) - v(L)] \geq v(L \cup S) - v(L)$, and one more application of the mean value theorem bounds $f(L \cup S) - f(L) \leq [v(L \cup S) - v(L)] \cdot \sup h'$. Dividing gives a ratio of at least ρ for every admissible pair. ■

Remark (priority). The inequality is not ours. The hypotheses make f what [5] calls ρ -submodular, and ρ -submodularity implies $\gamma \geq \rho$ by telescoping — Definition 4 and Proposition 8 of [5], restated as $\gamma \geq 1 - \check{\alpha}$ in Proposition 1 of [6]. The proof above is the one-line verification that a composition satisfies that definition with $\rho = \inf h' / \sup h'$. Note that passing to γ **discards information**: ρ -submodularity is strictly stronger, the converse of Proposition 8 of [5] being false, which is one reason the bound is loose in Section 6.

Corollary 1. Under a **cardinality** constraint, greedy returns at least $(1 - e^{-\rho}) \cdot OPT$, combining Theorem 1 with [4].

The cardinality restriction is a limitation of γ , not modesty about the statement. Feldman and Ward show that for every $\gamma < 1$ there is a monotone γ -weakly submodular function and a *simple partition matroid* on which greedy has no constant-factor guarantee at all (Theorem 1.1 of [14]). So no γ -only bound transfers to a matroid, let alone to the matroid intersection of Section 3.1. Sharper constants need more than γ : the tight guar-

antee in terms of γ and a generalized curvature is $(1/\alpha)(1 - e^{-\alpha\gamma})$ (Theorem 1 of [8]).

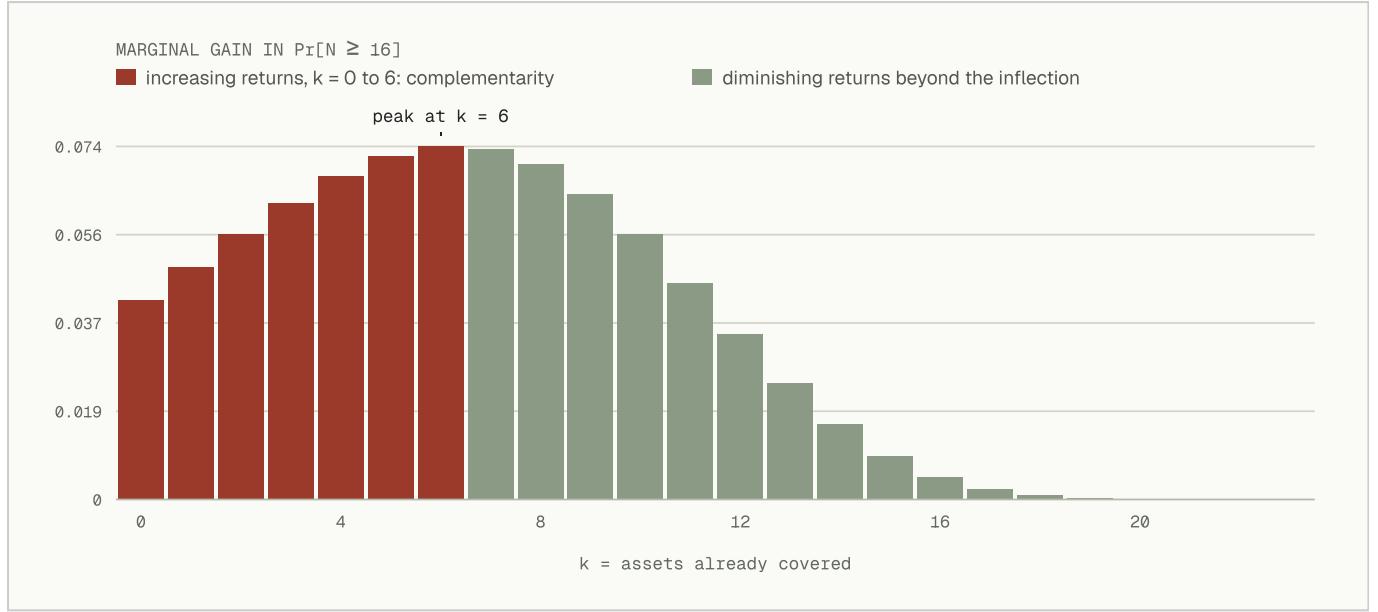


FIGURE 5 • ILLUSTRATIVE COMPUTATION

Where submodularity actually fails. Marginal gain in the probability that at least 16 of 24 assets are simultaneously capable, as each successive asset is upgraded from capability 0.55 to 0.95. The marginal rises through $k = 6$ before falling, so early inductions show increasing returns. That is complementarity, and it comes from the conjunctive form of the requirement rather than from the cascade. Computed from the stated parameters; not observed on a fleet.

6. Where submodularity actually fails

The feedback does not break the structure under the mean-field law with a rising propensity. Something else does, for an unrelated reason, and it breaks it harder.

6.1 Conjunctive mission requirements

Mission requirements are not fractional. An operator needs some number of assets simultaneously capable. Let $N(S)$ be the number of capable assets and

$$R_m(S) = \Pr[N(S) \geq m]$$

where R_m , conjunctive readiness: the probability that at least m assets are capable; $N(S)$, number of capable assets under schedule S ; m , the number of assets the mission requires simultaneously; S , a schedule; in the submodularity-ratio definition, the disjoint test set.

Assume homogeneous failure probabilities so coverage reduces to a count, that each induction protects one previously unprotected asset, and per-asset capability q_{hi} when covered and q_{lo} when not, independent across assets, with $0 < q_{lo} < q_{hi}$. The strict positivity of q_{lo} is not cosmetic: at $q_{lo} = 0$ the identity below has both sides zero for every $m \geq 3$, and the threshold fires on a flat marginal. Then $N(S)$ is Poisson-binomial, R_m depends only on the covered count k , and the relevant h lives on $\{0, \dots, n\}$ with unit increments, where concavity and the discrete increment condition coincide.

Upgrading one asset from q_{lo} to q_{hi} gives the identity

$$h(k+1) - h(k) = (q_{hi} - q_{lo}) \cdot \Pr[M_k = m-1]$$

where q_{lo} , per-asset capability probability when the asset is not covered; q_{hi} , per-asset capability probability when the asset is covered; M_k , number of capable assets among the other $n-1$, a Poisson-binomial count; h , scalar response: readiness as a function of covered mass; m , the number of assets the mission requires simultaneously; k , a budget, or the count of assets already covered.

where M_k counts the other $n-1$ assets. So the marginal increases at k exactly when $\Pr[M_{k+1} = m-1] > \Pr[M_k = m-1]$.

Proposition 6. The first induction shows increasing returns if and only if

$$m-1 > (n-1)q_{lo}$$

where q_{lo} , per-asset capability probability when the asset is not covered; n , fleet size; m , the number of assets the mission requires simultaneously.

Proof. Expanding at $k=0$ with $B \sim \text{Bin}(n-2, q_{lo})$, the difference reduces to $(q_{hi} - q_{lo})(\Pr[B = m-2] - \Pr[B = m-1])$, and the ratio $\Pr[B=m-1]/\Pr[B=m-2]$ equals $[(n-m)/(m-1)] \cdot q_{lo}/(1-q_{lo})$, which is one exactly at $q_{lo} = (m-1)/(n-1)$. ■

Exhaustive exact-rational search over $n \leq 14$, all $1 \leq m \leq n$ and q_{lo}, q_{hi} on twentieths confirms the threshold with no slack for $q_{lo} > 0$, and finds 1 560 mismatches at $q_{lo} = 0$ with $m \geq 3$ — which is why the hypothesis is stated.

This is complementarity, and it comes from the conjunction rather than from the cascade. The second modelling assumption above removes coverage overlap, so the failure found here is separate from overlap and not a consequence of it. On a worked instance with $n = 24, m = 16, q_{lo} = 0.55, q_{hi} = 0.95$, the threshold gives $(24-1)(0.55) + 1 = 13.65 < 16$, so returns increase initially; computed, the marginal rises through $k = 6$, peaks there, and first falls at $k = 7$.

6.2 What the ratio bound is worth here, and what dominates it

On that same instance the discrete slope ratio over the full range is $\rho = 4.779 \times 10^{-5}$, because the tail probability flattens almost

completely at large k , against a directly computed $\gamma = 0.665$. So **Theorem 1 is four orders of magnitude loose here**, and direct computation of γ is the right tool. Restricting to the convex window $k \leq 6$ gives $\rho = 0.566$ against $\gamma = 0.721$ for the matching budget.

It is also **redundant** here, and this is the sharper point. The inner function is a count, hence modular, and for a modular inner function the discrete ρ coincides *exactly* with the complement of the generalized curvature of [10] — we verified $\rho = 1 - c$ to machine precision on this instance. Theorem 8.1 of [10] then gives greedy a ρ -approximation under **any single matroid**, a constant no smaller than $1 - e^{-\rho}$ under a weaker constraint than Corollary 1 assumes. So on the one instance where this note applies Theorem 1, a 2017 curvature result dominates it, and we state Corollary 1 only because $1 - e^{-\gamma}$ is the form in which this literature is usually quoted.

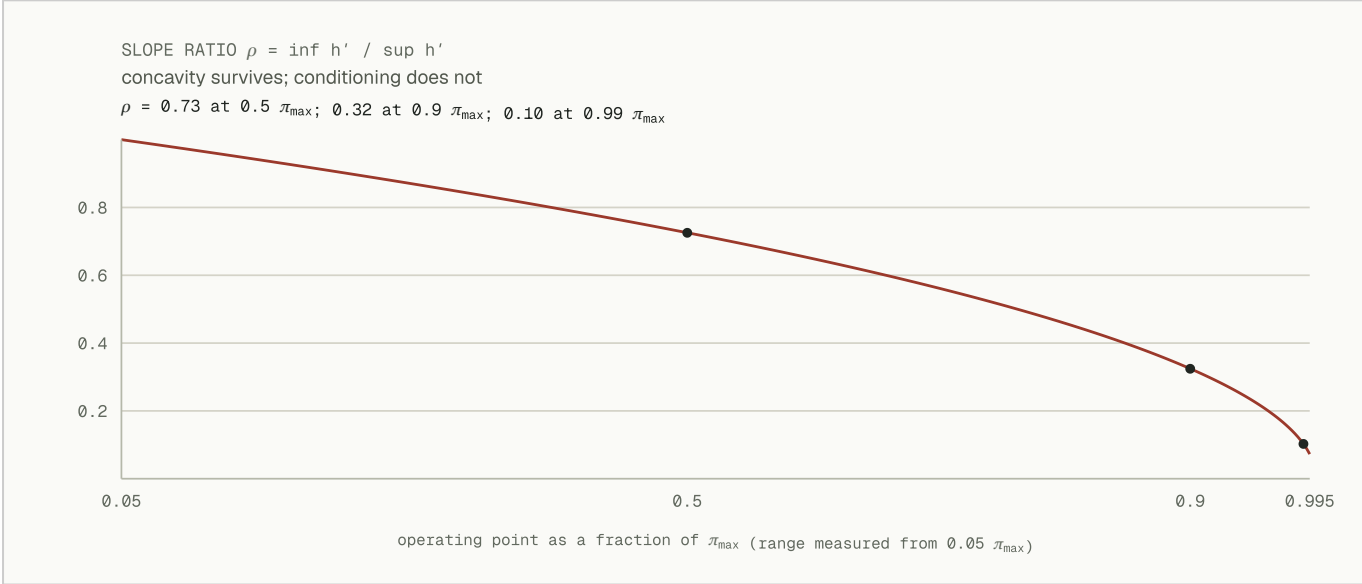


FIGURE 4 • ILLUSTRATIVE COMPUTATION

What the bifurcation actually costs. The ratio of extreme response slopes falls toward zero as the operating point approaches the fold, so the marginal value of an induction varies by a growing factor across the range. Under the mean-field law of Section 3.4 the objective stays submodular throughout and becomes stiff. Stiffness is an estimation problem, which Figure 8 quantifies. Section 9 reports that the fold this figure approaches is not reachable at a measured propensity.

7. Greedy is robust past its guarantees

This is the one measured section, and it is the most useful thing here for someone deciding where to spend effort. Sections 5 and

6 identify three distinct ways the structure greedy relies on can fail. Each was tested against **exhaustive optima** on instances small enough to solve exactly.

STRUCTURAL FAILURE	FEASIBLE REGION	WHAT GREEDY IS OWED	WORST MEASURED RATIO
Conjunctive objective (6.1)	cardinality, one matroid	nothing; not submodular	0.9317 , costing 0.0220 of mission probability
Coverage coupled by routing	cardinality	nothing; coverage not even monotone	0.7851
Coverage coupled by routing	two matroids, the real region	nothing	0.4964

STRUCTURAL FAILURE	FEASIBLE REGION	WHAT GREEDY IS OWED	WORST MEASURED RATIO
Convex response (Proposition 3)	cardinality	nothing at all	no breach observed

The feasible region decides whether this is a strong claim or a weak one. Under a cardinality bound, not one instance in any of these families fell below $1 - 1/e$, a constant greedy is entitled to only when the objective is monotone submodular, which none of them is. On the intersection of two partition matroids — the region the problem actually has — tight routing coupling drops the mean ratio to *0.734*, the worst case to *0.4964*, and **11 of 120 instances fall below the cardinality constant**. Greedy still clears $1/3$, which is the constant two-matroid greedy would be owed if the objective were submodular [2].

So the corrected statement has two halves and both belong in it. Greedy *does* breach the cardinality constant on the real region under tight routing, on about nine per cent of instances. And it stays above the constant appropriate to that region. The robustness survives in weakened form: comfortably above the right constant, with a thin margin rather than no margin.

Part of that gap is budget rather than structure. The cardinality rows select six inductions; the two-matroid region has rank eight, and both were run against leg capacity calibrated to a budget of six, so the harder region also selects a third more inductions against the same capacity. Re-running the cardinality case at a matched budget of eight gives a worst ratio of *0.7069*, which accounts for about a quarter of the published gap — not the half an earlier analysis claimed. What survives the matched comparison is the part that matters: breaches of the cardinality constant occur only on the two-matroid region.

Why two matroids are harder is worth a sentence. Under a cardinality bound greedy can substitute a crowded induction for an uncrowded one freely; when both matroids restrict where the budget may be spent, that substitution is often infeasible and a congested route cannot be avoided.

The honest summary. The structural concerns in this note are real, and the practical stakes attached to them are small on the easy feasible region and moderate on the real one. That is a conclusion we did not expect and did not look for, and it is the strongest support the practical thesis has: the difficulty lands on estimation rather than on the scheduler, not because a guarantee protects the scheduler but because greedy clears the constant appropriate to its region even where the structure supporting that constant has gone.

8. Limitations, and what the companion paper measures as unmet

Stated at length because several of this note's hypotheses are measured as false elsewhere, and a reader should not have to find that out from the other paper.

The reduction assumes the schedule acts only through π . Section 3.4's composition requires that covering failure mass changes the equilibrium through the uncovered-mass term and nothing else. In the companion paper's simulator it does not: protecting an asset removes its failures, removes it as a source of removal demand, *and* relieves the shared spares and depot pool, and the third changes the propensity law itself. So the balance law is misspecified under coverage, and a measured readiness curve carries curvature from a source the criteria do not account for. **This is the largest gap, and it is not the feedback law.** Deriving $b(x; S)$ when protection changes the donor pool is the open problem we would most like solved.

The propensity is a modelling choice, and the headline depends on it. Proposition 4's closed form is exact for the law (3) and says nothing about any other. The companion paper exhibits a strictly subcritical saturating propensity, with no fold anywhere, under which readiness is not submodular — so submodularity here is a fact about a propensity curve, not about cannibalization.

The measured sign is the opposite of the one assumed. Under the donor policies implemented in the companion paper, the fitted propensity *decreases* in the disabled fraction, which is Proposition 3's regime rather than Proposition 1's. On three panels of published operational statistics the sign is positive, as the model needs — in all nine specifications the three admit — and it survives removing unit-specific linear trends on two of the three. No panel is strong in every way: the one with the cleanest provenance is the one whose significance is coincident trend, the second rests on a document the authors have not read, and the third, which is the largest and carries the right regressor, was digitised from charts that print no values. Neither can calibrate β_0 or λ , because published cannibalization rates have a different denominator from $b(x)$, which is a share of the disabled fraction.

The fold is not reachable at measured parameters. Across the companion paper's grid the largest $\beta_0 \kappa$ observed is *0.164* against roughly *0.4* for a reachable fold, and the shortfall is a factor of 2.4 to 7.9 under mean-field balance and 21 to 60 under the donor-corrected one. Proposition 5 is a correct statement about a region of parameter space that those fleets do not occupy.

Proposition 4's own prediction misplaces the measured transition. Fitting the donor-limited law per regime, λ lands on the edge of the scan in four of seven regimes, meaning the data does not identify it. Of the three where it is identified, the criterion is right **once**.

Theorem 1 is not ours, and is dominated where applied. Sections 2 and 5 state the priority; Section 6.2 states the domination.

The matroid results are quoted, not proved here. Section 7 measures ratios against exhaustive optima; it does not prove a guarantee for the two-matroid region, and by [14] no γ -only argument could.

No fleet experiment. Nothing here was run on an actual fleet. Section 7's instances are synthetic and small enough to solve exactly, which is what makes the ratios exact rather than estimated.

9. Open problems

1. **Derive the coverage map.** What is π as a function of covered mass — or rather, what is $b(x; S)$ — when protecting an asset removes its failures, removes it as a source of removal demand, and relieves the shared repair and spares pool? Until this is modelled, no criterion here can be tested against a fleet, and the decomposition of Section 3.4 is a convenience rather than a result.
2. **A balance law that both derives from the mechanism and predicts the curve.** Law (3) has the right two channels and Proposition 4's threshold still misplaces the measured transition. Whether a corrected coverage map suffices, or the propensity needs a third channel, is open.

3. **Does the first-induction threshold characterize submodularity globally?** Proposition 6 settles the first induction. Whether the first marginal failing to increase implies that no marginal increases is verified exhaustively for small fleets in exact arithmetic but not proved, and the obvious monotonicity route fails.
4. **Which propensity curves satisfy the criterion empirically?** Proposition 1 gives the mathematical condition, that $x^2 b'(x)$ be nondecreasing. Which curves real fleets exhibit is a question about cannibalization behaviour, and it decides whether any of this describes anything.
5. **Where is the donor-corrected response monotone, concave and physical at once?** Under balance (2) the fold condition and the concavity condition are governed by the same inequality family, so the two can no longer be chosen independently. Characterize the region in (β_0, κ) on which the response is simultaneously single-valued, concave in covered mass, and confined to $[0, 1]$.
6. **Why is greedy so robust?** Section 7 measures it clearing the constant appropriate to each region across three structural failures. A bound that applies to these objectives without assuming submodularity would be worth more to a practitioner than anything else here.

References

- [1] G. L. Nemhauser, L. A. Wolsey, M. L. Fisher. An analysis of approximations for maximizing submodular set functions I. *Mathematical Programming*, 14:265–294, 1978. doi.org/10.1007/BF01588971
- [2] G. L. Nemhauser, L. A. Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*, 3(3):177–188, 1978. doi.org/10.1287/moor.3.3.177
- [3] G. Călinescu, C. Chekuri, M. Pál, J. Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011. doi.org/10.1137/080733991
- [4] A. Das, D. Kempe. Approximate submodularity and its applications: sub-set selection, sparse approximation and dictionary selection. *Journal of Machine Learning Research*, 19(3):1–34, 2018. jmlr.org/papers/v19/16-534.html
- [5] M. El Halabi, F. Bach, V. Cevher. Combinatorial penalties: which structures are preserved by convex relaxations? *AISTATS*, PMLR 84:1551–1560, 2018. arxiv.org/abs/1710.06273
- [6] I. Bogunovic, J. Zhao, V. Cevher. Robust maximization of non-submodular objectives. *AISTATS*, PMLR 84:890–899, 2018. arxiv.org/abs/1802.07073
- [7] E. R. Elenberg, R. Khanna, A. G. Dimakis, S. Negahban. Restricted strong convexity implies weak submodularity. *Annals of Statistics*, 46(6B):3539–3568, 2018. doi.org/10.1214/17-AOS1679
- [8] A. A. Bian, J. M. Buhmann, A. Krause, S. Tschichatschek. Guarantees for greedy maximization of non-submodular functions with applications. *ICML*, PMLR 70:498–507, 2017. arxiv.org/abs/1703.02100
- [9] M. Conforti, G. Cornuéjols. Submodular set functions, matroids and the greedy algorithm: tight worst-case bounds and some generalizations of the Rado-Edmonds theorem. *Discrete Applied Mathematics*, 7(3):251–274, 1984. doi.org/10.1016/0166-218X(84)90003-9
- [10] M. Sviridenko, J. Vondrák, J. Ward. Optimal approximation for submodular and supermodular optimization with bounded curvature. *Mathematics of Operations Research*, 42(4):1197–1218, 2017. doi.org/10.1287/moor.2016.0842
- [11] R. Iyer, S. Jegelka, J. Bilmes. Curvature and optimal algorithms for learning and minimizing submodular functions. *Advances in Neural Information Processing Systems* 26, 2013. arxiv.org/abs/1311.2110
- [12] H. Lin, J. Bilmes. A class of submodular functions for document summarization. *Proceedings of ACL-HLT*, 510–520, 2011. aclanthology.org/P11-1052/
- [13] J. A. Bilmes. *Submodular functions, optimization, and applications to machine learning*. Course notes, EE596b, University of Washington, Spring 2016, Lecture 4, Theorems 4.5.1 and 4.5.4. people.ece.uw.edu/bilmes/classes/ee596b_spring_2016/lecture4_print.pdf
- [14] M. Feldman, J. Ward. On maximizing a weakly submodular function over a matroid constraint via the greedy algorithm. 2026. arxiv.org/abs/2609.05008
- [15] U.S. General Accounting Office (now the Government Accountability Office). *Military aircraft: services need strategies to reduce cannibalizations*. GAO-02-86, 2001. www.gao.gov/products/gao-02-86

- [16] W. W. Fisher, J. J. Brennan. The performance of cannibalization policies in a maintenance system with spares, repair, and resource constraints. *Naval Research Logistics Quarterly*, 33(1):1–15, 1986. doi.org/10.1002/nav.3800330102
- [17] C. C. Sherbrooke. METRIC: a multi-echelon technique for recoverable item control. *Operations Research*, 16(1):122–141, 1968. doi.org/10.1287/opre.16.1.122
- [18] I. Dobson, B. A. Carreras, V. E. Lynch, D. E. Newman. Complex systems analysis of series of blackouts: cascading failure, critical points, and self-organization. *Chaos*, 17(2):026103, 2007. doi.org/10.1063/1.2737822
- [19] U. Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998. doi.org/10.1145/285055.285059