THE GREEDY HEURISTIC FOR THE ℓ_0-CONSTRAINED FEATURE-BASED NEWSVENDOR HAS NO APPROXIMATION GUARANTEE

KOREA SUPERINTELLIGENCE LABS

ABSTRACT. Yuan and Wang (arXiv:2609.01544) select features for a linear newsvendor decision rule by minimising the empirical newsvendor (pinball) loss plus a ridge term under a hard sparsity constraint $\|\omega\|_0 \le d$, and propose a greedy heuristic for the resulting support-selection problem $\min_{|S|\le d} G(S)$. They leave the approximation quality of the greedy iterate as an open question, noting that G is not submodular. We answer the question in the negative. On an explicit instance with three features, four samples, budget $d = 2$ and unit costs, every greedy run, under every tie-breaking rule, ends at a support of value at least $\frac{1}{2}$ while the optimal support has value at most $1/\gamma$, for every ridge parameter $\gamma \ge 8$; the loss ratio is therefore at least $\gamma/2$ and no constant-factor guarantee exists. The same instance breaks the guarantee in the gain form used by the weak-submodularity literature: the greedy gain is strictly below $1 - 1/e$ times the optimal gain for every $\gamma \ge 8$, and G has increasing marginal returns. At $\gamma = 8$ all seven values of G are determined exactly, each by a matching primal–dual pair, and the ratio is exactly 6. We also explain why the question was genuinely open: the standard lower bound on the submodularity ratio requires a differentiable loss with a finite restricted-smoothness constant, and the pinball loss has none. Every value certificate rests on a weak-duality inequality proved from first principles; strong duality is never assumed. All theorems are verified in Lean 4 with Mathlib.

1. INTRODUCTION

1.1. **The problem and the greedy heuristic.** The feature-based newsvendor chooses an order quantity as a function of an observed feature vector. Yuan and Wang [1] study the sparse linear version of this problem: given samples $(x_i, y_i)_{i=1}^n$ with $x_i \in \mathbb{R}^D$ and demand $y_i \in \mathbb{R}$, a back-ordering cost $b > 0$, a holding cost $h > 0$, a ridge parameter $\gamma > 0$ and a sparsity budget d, they minimise

$$F(\omega, c) = \frac{1}{n}\sum_{i=1}^{n}\Big[b \max\{0,\ y_i - \omega^\top x_i - c\} + h \max\{0,\ \omega^\top x_i + c - y_i\}\Big] + \frac{1}{2\gamma}\|\omega\|_2^2 \tag{1}$$

over $\omega \in \mathbb{R}^D$, $c \in \mathbb{R}$ subject to $\|\omega\|_0 \le d$. The bracket is the newsvendor loss of the residual; with $\tau = b/(b+h)$ it is $(b+h)$ times the check (pinball) loss ρ_τ of quantile regression [6], and $b = h$ gives the absolute loss. The intercept c is free and is not counted by the sparsity constraint.

Yuan and Wang rewrite the problem over supports. For $S \subseteq \{1, \dots, D\}$ let

$$G(S) = \inf\{F(\omega, c) :\ \operatorname{supp}\omega \subseteq S\}, \tag{2}$$

so that the problem is $\min_{|S|\le d} G(S)$. Their theorem titled *Minimax Reformulation* expresses $G(S)$ through a concave maximisation over a polytope,

$$G(S) = \max\Big\{-z^\top \tfrac{\gamma}{2n^2} X_S X_S^\top z - \tfrac{1}{n} z^\top y :\ z \in [-b, h]^n,\ \textstyle\sum_i z_i = 0\Big\}, \tag{3}$$

where X_S is the $n \times |S|$ submatrix of the feature matrix X. Their greedy heuristic (the algorithm titled *Greedy Algorithm*) starts from $S_0 = \emptyset$ and, for $k = 0, \dots, d-1$, picks

$$j_k \in \arg\min_{j \notin S_k} G(S_k \cup \{j\}), \qquad S_{k+1} = S_k \cup \{j_k\}, \tag{4}$$

and returns the optimal (ω, c) supported on S_d. Immediately after the algorithm they write (we replace two internal cross-references by brackets):

> By [the minimax reformulation], it can be shown that the set function $G(S)$ is non-increasing in S, and the objective of iterates in [the greedy algorithm] is monotonically decreasing. It is an open question on the approximation guarantees of the final iterate

from the greedy algorithm, as the set function G does not satisfy the submodular assumption commonly used in literature.

Throughout we refer to the statements of [1] by their titles rather than by their numbers: the preprint is at its first version, and equation and theorem numbers of a preprint are not stable across versions. For the reader of version 1 (numbers obtained by compiling its arXiv source on 2026-09-03): the sparse problem $\min\{F(\omega, c) : \|\omega\|_0 \le d\}$ is its display (3), the support form $\min_{|S|\le d} G(S)$ is its display (5), the reformulation (3) is its Theorem 3 with display (6), and the heuristic is its Algorithm 2.

1.2. **Why no guarantee was available in print.** The natural route to a guarantee for (4) is weak submodularity. For the monotone gain $f(S) = G(\emptyset) - G(S)$, Das and Kempe [2] define the submodularity ratio $\gamma_{U,k}(f)$ (Definition 2.3 of the arXiv version) as the smallest value of $\sum_{x\in S}\big(f(L\cup\{x\}) - f(L)\big)/\big(f(L\cup S) - f(L)\big)$ over $L \subseteq U$ and S disjoint from L with $|S| \le k$, and prove that forward selection attains at least $(1 - e^{-\gamma_{S^{\mathrm{FR}},k}})$ times the optimal gain for the squared-multiple-correlation objective (their Theorem 3.2). This is a conditional statement: it produces a number only once the submodularity ratio of the objective at hand is bounded below. The standard lower bound is due to Elenberg, Khanna, Dimakis and Negahban [3]: if l is restricted strongly concave with parameter m and restricted smooth with parameter M on the relevant sparse domains, then $\gamma_{U,k} \ge m_{|U|+k}/\widetilde{M}_{|U|+1}$ (their Theorem 1). Their definition of restricted strong concavity and restricted smoothness (their Definition 3) is stated through the gradient $\nabla l(x)$,

$$-\tfrac{m_\Omega}{2}\|y-x\|_2^2 \;\ge\; l(y) - l(x) - \langle \nabla l(x), y-x\rangle \;\ge\; -\tfrac{M_\Omega}{2}\|y-x\|_2^2,$$

and their standing assumption is verbatim "We assume a differentiable function $l : \mathbb{R}^p \to \mathbb{R}$"; the section is introduced with "if the concavity parameter is bounded away from 0 and the smoothness parameter is finite, then the submodularity ratio is also bounded away from 0". (Statement numbers of [2] and [3] are those of version 2 of each arXiv e-print, obtained by compiling the sources on 2026-09-03; the quotations were checked against the same sources.)

The newsvendor objective (1) is piecewise linear in (ω, c) plus a ridge term. At every kink $y_i = \omega^\top x_i + c$ its subgradient jumps, and the lower inequality above fails for every finite M, whichever subgradient is substituted for $\nabla l(x)$: the deficit $l(y) - l(x) - \langle g, y-x\rangle$ is of first order in $\|y-x\|$ along a direction crossing the kink, not of second order. So the restricted smoothness constant is $M = \infty$, the bound reads $\gamma_{U,k} \ge 0$, and the guarantee degenerates to $1 - e^0 = 0$. The ridge term supplies strong concavity on the ω coordinates, i.e. the numerator; it does nothing for the denominator. Among greedy methods Yuan and Wang cite the sparse ridge regression work of Xie and Deng [4]; the greedy guarantee proved there runs through the closed-form dual of the squared loss and is squared-loss only, as its authors say. The preprint [1] itself does not cite the weak-submodularity literature: its bibliography names neither [2] nor [3] and contains no work with "submodular" in its title. Two days after its submission it had no record in OpenAlex, so no citation of it could be searched. In a local full-text corpus of arXiv e-prints (373 text shards) we found 125 papers containing the phrase "submodularity ratio" and 1,796 containing "quantile regression", and none containing both; likewise none containing both "newsvendor" and "submodularity ratio". None of the 291 titles citing [2] or [3] in OpenAlex (272 and 19 respectively by the `cites` filter, read 2026-09-03; the second is an undercount, since only the journal record of [3] exists there) contains a word indicating a non-smooth loss. The nearest work we found, Barber and Sidky [7], proves a restricted-strong-convexity property, in a form of their own, for sparse high-dimensional quantile regression, whose loss is not differentiable; but it mentions submodularity only in its bibliography, where it cites [3], and makes no greedy or approximation-ratio claim. We read these negatives as "not found", not as proof of absence. The question is therefore open in print, in its qualitative form: not only the value of the worst-case ratio, but whether any ratio exists.

1.3. **Results.** We answer the question in the negative on one explicit instance, at the smallest size where the greedy heuristic can fail at all ($D = 3$, $d = 2$: for $d = 1$ the greedy rule is the exhaustive search over singletons and G is non-increasing by its definition as an infimum, and for $d \ge D$ it returns the full support). Take $n = 4$, $b = h = 1$, ridge parameter γ left free, and

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & -1 & -1 \end{pmatrix}, \qquad y = \begin{pmatrix} 2 \\ 0 \\ 0 \\ -2 \end{pmatrix}, \tag{5}$$

with the three columns indexed $0, 1, 2$ and the four samples $1, \dots, 4$. Write $S_{\text{gr}} = \{j_0, j_1\}$ for the support returned by any run of (4), and S_{opt} for an optimal support of size at most 2.

(A) For every $\gamma \geq 8$ and every greedy run, $G(S_{\text{gr}}) \geq \frac{1}{2}$ while $G(\{1,2\}) \leq 1/\gamma$, so $\frac{\gamma}{2}\, G(\{1,2\}) \leq G(S_{\text{gr}})$ (Theorem 4.2). The loss ratio $G(S_{\text{gr}})/G(S_{\text{opt}})$ is at least $\gamma/2$: no constant-factor approximation guarantee holds for the greedy heuristic.

(B) For every $\gamma \geq 8$ and every greedy run, the greedy gain is strictly below the classical submodular bound: $G(\emptyset) - G(S_{\text{gr}}) < (1 - e^{-1})\big(G(\emptyset) - G(\{1,2\})\big)$ (Theorem 4.6). A monotone submodular gain satisfies the reverse inequality [5], so no $(1 - 1/e)$ guarantee holds for G either.

(C) At $\gamma = 8$ the seven values of G are exactly $G(\emptyset) = 1$, $G(\{0\}) = \frac{3}{4}$, $G(\{1\}) = G(\{2\}) = 1$, $G(\{0,1\}) = G(\{0,2\}) = \frac{3}{4}$, $G(\{1,2\}) = \frac{1}{8}$, each certified by a matching primal–dual pair, and every greedy run satisfies $G(S_{\text{gr}}) = 6\, G(\{1,2\})$ (Theorem 4.7).

Two supporting statements make the mechanism explicit: the first greedy iterate is forced onto column 0 for every $\gamma \geq 8$ under any tie-breaking rule (Proposition 4.1), and G has increasing marginal returns for every $\gamma > 1$ (Proposition 4.4), which proves the assertion of [1] that G is not submodular; that assertion is made there without proof or witness. Proposition 4.8 records the controls that show no hypothesis is vacuous.

The construction is simple. Columns 1 and 2 form a parity pair, $y = x^{(1)} + x^{(2)}$, so the support $\{1, 2\}$ fits the demand exactly and pays only the ridge; but each of them alone is a flat direction of the absolute loss, so on its own it is worth nothing. Column 0 is a decoy: it lowers the objective on its own, so the greedy rule takes it first, and then it is trapped, because the dual certificates for $\{0,1\}$ and $\{0,2\}$ can be chosen orthogonal to the columns involved, which makes their value $\frac{1}{2}$ independent of γ. Every value certificate is a pair of explicit rational vectors checked against a weak-duality inequality proved from first principles (Section 2); strong duality, attainment and numerical solvers are never relied upon.

1.4. **What remains open.** The supremum of the loss ratio over all instances is $+\infty$ by (A), so the quantitative questions that remain are finite ones: the exact maximum of the ratio over a named finite grid of instances, the infimum of the gain ratio $(G(\emptyset) - G(S_{\text{gr}}))/(G(\emptyset) - G(S_{\text{opt}}))$ (which on the present instance tends to $\frac{1}{2}$ as $\gamma \to \infty$ and may or may not be 0 over all instances), and a positive guarantee derived from the dual (3), which is a smooth concave quadratic over a polytope and so may admit the restricted-smoothness side that the primal denies. Section 6 states these with their computational cost.

2. Preliminaries

2.1. **The objective and the dual.** Fix $n, D \geq 1$, a feature matrix $X \in \mathbb{R}^{n \times D}$ with rows x_i and columns $x^{(j)}$, a demand vector $y \in \mathbb{R}^n$, costs $b, h > 0$ and $\gamma > 0$. The newsvendor loss of a residual r is $\rho(r) = b \max\{0, r\} + h \max\{0, -r\}$, and F is (1), i.e. $F(\omega, c) = \frac{1}{n}\sum_i \rho(y_i - \omega^\top x_i - c) + \frac{1}{2\gamma}\|\omega\|_2^2$. For $S \subseteq \{0, \dots, D-1\}$, $G(S)$ is the infimum (2); it is finite because $F \geq 0$ and $\omega = 0$ is admissible. The dual objective and feasible set of (3) are

$$\Phi_S(z) = -\frac{\gamma}{2n^2}\sum_{j \in S}\Big(\sum_{i=1}^n z_i x_{ij}\Big)^2 - \frac{1}{n}\sum_{i=1}^n z_i y_i, \qquad \mathcal{Z} = \Big\{ z \in \mathbb{R}^n : \ -b \leq z_i \leq h \text{ for all } i,\ \textstyle\sum_i z_i = 0 \Big\},$$

so that $\sum_{j \in S}(\sum_i z_i x_{ij})^2 = \|X_S^\top z\|_2^2$ and Φ_S is the maximand in (3). Note that $0 \in \mathcal{Z}$ and $\Phi_S(0) = 0$.

2.2. **Weak duality and the certificate principle.** We use only the easy half of the minimax reformulation, proved directly.

Lemma 2.1 (pointwise Fenchel inequality). *If $-b \leq z \leq h$ then $-zr \leq \rho(r)$ for every $r \in \mathbb{R}$.*

Proof. If $r \geq 0$ then $\rho(r) = br \geq -zr$ because $z \geq -b$; if $r < 0$ then $\rho(r) = -hr \geq -zr$ because $z \leq h$. □

Lemma 2.2 (weak duality)**.** *Let $z \in \mathcal{Z}$ and let ω be supported on S. Then $\Phi_S(z) \leq F(\omega, c)$ for every $c \in \mathbb{R}$. Consequently $\Phi_S(z) \leq G(S)$ for every $z \in \mathcal{Z}$, and $G(S) \leq F(\omega, c)$ for every ω supported on S.*

Proof. Write $r_i = y_i - \omega^\top x_i - c$ and $A_j = \sum_i z_i x_{ij}$. By Lemma 2.1, $-\sum_i z_i r_i \leq \sum_i \rho(r_i)$. Expanding, $\sum_i z_i r_i = \sum_i z_i y_i - \sum_j \omega_j A_j - c \sum_i z_i$, and the last term vanishes because $\sum_i z_i = 0$; since ω is supported on S the middle sum runs over $j \in S$ only. Finally, for each $j \in S$ the completed square $\frac{1}{2\gamma}\big(\omega_j + \frac{\gamma}{n} A_j\big)^2 \geq 0$ gives $-\frac{\gamma}{2n^2} A_j^2 \leq \frac{\omega_j A_j}{n} + \frac{\omega_j^2}{2\gamma}$. Dividing the first inequality by n and adding these over $j \in S$ yields $\Phi_S(z) \leq F(\omega, c)$. The two consequences follow by taking the infimum over (ω, c) and the definition of G. □

Corollary 2.3 (certificate principle)**.** *If ω is supported on S, $z \in \mathcal{Z}$, and $F(\omega, c) = \Phi_S(z) = v$, then $G(S) = v$.*

Every exact value in this paper is proved by exhibiting such a pair; every one-sided bound by exhibiting one of the two witnesses. In particular, the identity (3) of [1], which rests on strong duality, is never used.

2.3. **Greedy runs.** The rule (4) is nondeterministic when the minimum is attained more than once. For $D = 3$ and $d = 2$ a greedy run is a pair (j_0, j_1) with $j_1 \neq j_0$ such that

$$G(\{j_0\}) \leq G(\{k\}) \text{ for all } k, \qquad G(\{j_0, j_1\}) \leq G(\{j_0, k\}) \text{ for all } k \neq j_0. \tag{6}$$

All statements below are quantified over every pair satisfying (6), i.e. over every tie-breaking rule.

3. The instance

Let X, y be as in (5), $n = 4$, $b = h = 1$, so that $\rho(r) = |r|$ and $F(\omega, c) = \frac{1}{4}\sum_i |y_i - \omega^\top x_i - c| + \frac{1}{2\gamma}\|\omega\|^2$. We write G_γ for G at ridge parameter γ when the dependence matters. The columns are $x^{(0)} = (1, 0, 0, 0)$, $x^{(1)} = (1, 1, -1, -1)$, $x^{(2)} = (1, -1, 1, -1)$, and $y = x^{(1)} + x^{(2)}$. Five dual vectors and two primal policies suffice:

$$z_a = (-1, 1, -1, 1), \qquad z_b = (-1, -1, 1, 1), \qquad z_c = (0, 0, -1, 1),$$
$$z_d = (0, -1, 0, 1), \qquad z_e = (-1, \tfrac{3}{4}, \tfrac{3}{4}, -\tfrac{1}{2}),$$
$$\omega_{\mathrm{dec}} = (2, 0, 0),\ c = 0, \qquad \omega_{\mathrm{par}} = (0, 1, 1),\ c = 0.$$

All five dual vectors lie in $\mathcal{Z}$ (entries in $[-1, 1]$, sum zero). The orthogonality relations are $z_a \perp x^{(1)}$, $z_b \perp x^{(2)}$, $z_c \perp x^{(0)}, x^{(1)}$, $z_d \perp x^{(0)}, x^{(2)}$, and the linear terms are $-\frac{1}{4} z^\top y = 1, 1, \frac{1}{2}, \frac{1}{2}, \frac{1}{4}$ for $z_a, \ldots, z_e$ respectively.

Lemma 3.1 (the flat directions)**.** *For every $\gamma > 0$, $G_\gamma(\emptyset) = G_\gamma(\{1\}) = G_\gamma(\{2\}) = 1$.*

Proof. The policy $\omega = 0$, $c = 0$ has $F = \frac{1}{4}(2 + 0 + 0 + 2) = 1$ and is supported on every S. For the dual side, $\Phi_\emptyset(z_a) = 1$ (no quadratic term); $\Phi_{\{1\}}(z_a) = 1$ because $z_a \perp x^{(1)}$; and $\Phi_{\{2\}}(z_b) = 1$ because $z_b \perp x^{(2)}$. Corollary 2.3 applies in each case, for every $\gamma > 0$, since the quadratic terms vanish identically. □

Lemma 3.2 (the decoy and the parity pair)**.** *For every $\gamma > 0$, $G_\gamma(\{0\}) \leq \frac{1}{2} + \frac{2}{\gamma}$ and $G_\gamma(\{1, 2\}) \leq \frac{1}{\gamma}$.*

Proof. The residuals of ω_{dec} are $y - 2x^{(0)} = (0, 0, 0, -2)$, so $F(\omega_{\mathrm{dec}}, 0) = \frac{1}{2} + \frac{4}{2\gamma}$. The residuals of ω_{par} vanish, so $F(\omega_{\mathrm{par}}, 0) = \frac{2}{2\gamma}$. Apply Lemma 2.2. □

Lemma 3.3 (the trap)**.** *For every $\gamma > 0$, $G_\gamma(\{0, 1\}) \geq \frac{1}{2}$ and $G_\gamma(\{0, 2\}) \geq \frac{1}{2}$.*

Proof. z_c is orthogonal to $x^{(0)}$ and $x^{(1)}$, so $\Phi_{\{0,1\}}(z_c) = -\frac{1}{4} z_c^\top y = \frac{1}{2}$ with no dependence on γ; likewise $\Phi_{\{0,2\}}(z_d) = \frac{1}{2}$. Apply Lemma 2.2. □

Lemma 3.3 is the whole point of the construction: a dual certificate chosen in the null space of $X_S^\top$ has a value that does not decay as γ grows, whereas the optimal support $\{1,2\}$ pays only the ridge, $O(1/\gamma)$.

4. Results

Proposition 4.1 (greedy is forced onto the decoy)**.** *Let $\gamma \geq 8$. If j_0 satisfies the first condition of* (6)*, i.e. $G_\gamma(\{j_0\}) \leq G_\gamma(\{k\})$ for all k, then $j_0 = 0$.*

Proof. By Lemma 3.2, $G_\gamma(\{0\}) \leq \frac{1}{2} + \frac{2}{\gamma} \leq \frac{3}{4}$ for $\gamma \geq 8$, while $G_\gamma(\{1\}) = G_\gamma(\{2\}) = 1$ by Lemma 3.1. So neither 1 nor 2 can be a minimiser of $k \mapsto G_\gamma(\{k\})$. ☐

Theorem 4.2 (the loss ratio is unbounded)**.** *Let $\gamma \geq 8$ and let (j_0, j_1) be any greedy run, i.e. any pair satisfying* (6)*. Then*

$$\tfrac{1}{2} \leq G_\gamma(\{j_0, j_1\}), \qquad G_\gamma(\{1,2\}) \leq \tfrac{1}{\gamma}, \qquad \tfrac{\gamma}{2}\, G_\gamma(\{1,2\}) \leq G_\gamma(\{j_0, j_1\}).$$

Proof. By Proposition 4.1, $j_0 = 0$, so $j_1 \in \{1,2\}$ and $G_\gamma(\{j_0, j_1\}) \geq \frac{1}{2}$ by Lemma 3.3. The second inequality is Lemma 3.2. Multiplying it by $\gamma/2 > 0$ gives $\frac{\gamma}{2} G_\gamma(\{1,2\}) \leq \frac{1}{2} \leq G_\gamma(\{j_0, j_1\})$. ☐

Remark 4.3 (the ratio form). Since $\{1,2\}$ is admissible, $G_\gamma(S_{\mathrm{opt}}) \leq G_\gamma(\{1,2\})$. Moreover $G_\gamma(S) > 0$ for every S: for $t > 0$ small, $tz_c \in \mathcal{Z}$ and $\Phi_S(tz_c) = \frac{t}{2} - t^2 \frac{\gamma}{32} \sum_{j \in S} (z_c^\top x^{(j)})^2 > 0$. Hence Theorem 4.2 gives $G_\gamma(S_{\mathrm{gr}})/G_\gamma(S_{\mathrm{opt}}) \geq G_\gamma(S_{\mathrm{gr}})/G_\gamma(\{1,2\}) \geq \gamma/2$ for every $\gamma \geq 8$ and every greedy run, and the ratio is unbounded over the class of instances of [1]. Note that the second condition of (6) is used only to force $j_1 \neq j_0$; the trap catches both choices of j_1. The positivity argument of this remark is not part of the formal development; the multiplicative form in the theorem is.

Proposition 4.4 (G has increasing marginal returns)**.** *For every $\gamma > 1$,*

$$G_\gamma(\emptyset) - G_\gamma(\{2\}) \;<\; G_\gamma(\{1\}) - G_\gamma(\{1,2\}).$$

Proof. The left side is $1 - 1 = 0$ by Lemma 3.1; the right side is at least $1 - \frac{1}{\gamma} > 0$ by Lemmas 3.1 and 3.2. ☐

In terms of the gain $f(S) = G(\emptyset) - G(S)$, Proposition 4.4 says $f(\{2\}) - f(\emptyset) < f(\{1,2\}) - f(\{1\})$ with $\emptyset \subseteq \{1\}$: the marginal value of column 2 increases when column 1 is present, so f is not submodular. This is the assertion of [1] quoted in Section 1.1, now with a witness.

Remark 4.5 (the submodularity ratio of the instance is zero). With $L = \emptyset$ and $S = \{1,2\}$ in the definition of $\gamma_{U,k}$ recalled in Section 1.2, the numerator is $f(\{1\}) + f(\{2\}) = 0$ by Lemma 3.1 while the denominator $f(\{1,2\}) \geq 1 - 1/\gamma$ is positive for $\gamma > 1$. So $\gamma_{\emptyset,2}(f) = 0$ on this instance, and the bound $1 - e^{-\gamma_{U,k}}$ of [2] is not merely unavailable but equal to 0 here. This observation is drawn in prose from the formal facts of Lemmas 3.1 and 3.2; it is not itself a formal statement.

Theorem 4.6 (the greedy gain is below the submodular bound)**.** *Let $\gamma \geq 8$ and let (j_0, j_1) be any greedy run. Then*

$$G_\gamma(\emptyset) - G_\gamma(\{j_0, j_1\}) \;<\; (1 - e^{-1})\big(G_\gamma(\emptyset) - G_\gamma(\{1,2\})\big).$$

Proof. By Lemma 3.1 and Theorem 4.2 the left side is at most $1 - \frac{1}{2} = \frac{1}{2}$ and the bracket on the right is at least $1 - \frac{1}{\gamma} \geq \frac{7}{8}$. Since $e > \frac{7}{3}$ we have $e^{-1} < \frac{3}{7}$ and $1 - e^{-1} > \frac{4}{7}$, so the right side exceeds $\frac{4}{7} \cdot \frac{7}{8} = \frac{1}{2}$. (The formal proof takes $e > \frac{7}{3}$ from the Mathlib bound $2.7182818283 < e$.) ☐

For a monotone submodular set function f with $f(\emptyset) = 0$ and a cardinality constraint, Nemhauser, Wolsey and Fisher [5] proved that the greedy set satisfies $f(S_{\mathrm{gr}}) \geq (1 - 1/e) \max_{|S| \leq d} f(S)$; choosing the element of largest marginal gain is the same as (4). Theorem 4.6 shows that the gain of the present G violates this inequality, strictly, for every $\gamma \geq 8$ and every tie-breaking rule, since $\max_{|S| \leq 2} f(S) \geq f(\{1,2\})$. This is the form in which [2] and [3] state their guarantees.

Theorem 4.7 (the seven values at $\gamma = 8$, and the exact ratio 6)**.** *At $\gamma = 8$,*

$$G_8(\emptyset) = 1, \quad G_8(\{0\}) = \tfrac{3}{4}, \quad G_8(\{1\}) = G_8(\{2\}) = 1, \quad G_8(\{0,1\}) = G_8(\{0,2\}) = \tfrac{3}{4}, \quad G_8(\{1,2\}) = \tfrac{1}{8},$$

and every greedy run (j_0, j_1) at $\gamma = 8$ satisfies $G_8(\{j_0, j_1\}) = 6\, G_8(\{1,2\})$.

Proof. Three values are Lemma 3.1. For the other four, Corollary 2.3 with the following pairs; the quadratic dual term is $\frac{8}{32}\|X_S^\top z\|^2 = \frac{1}{4}\|X_S^\top z\|^2$.

- $S = \{0\}$: $F(\omega_{\mathrm{dec}}, 0) = \frac{1}{2} + \frac{4}{16} = \frac{3}{4}$; $z_b^\top x^{(0)} = -1$, so $\Phi_{\{0\}}(z_b) = -\frac{1}{4} + 1 = \frac{3}{4}$.
- $S = \{0,1\}$: the same primal point; $z_a^\top x^{(0)} = -1$, $z_a \perp x^{(1)}$, so $\Phi_{\{0,1\}}(z_a) = \frac{3}{4}$.
- $S = \{0,2\}$: the same primal point; $z_b^\top x^{(0)} = -1$, $z_b \perp x^{(2)}$, so $\Phi_{\{0,2\}}(z_b) = \frac{3}{4}$.
- $S = \{1,2\}$: $F(\omega_{\mathrm{par}}, 0) = \frac{2}{16} = \frac{1}{8}$; $z_e^\top x^{(1)} = z_e^\top x^{(2)} = -\frac{1}{2}$ and $-\frac{1}{4} z_e^\top y = \frac{1}{4}$, so $\Phi_{\{1,2\}}(z_e) = -\frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} = \frac{1}{8}$.

For the ratio, Proposition 4.1 gives $j_0 = 0$, so $\{j_0, j_1\}$ is $\{0,1\}$ or $\{0,2\}$, of value $\frac{3}{4} = 6 \cdot \frac{1}{8}$. □

The optimal support of size at most 2 at $\gamma = 8$ is $\{1,2\}$, the unique minimiser among the seven values, and the greedy ratio on this instance at $\gamma = 8$ is exactly $(3/4)/(1/8) = 6$.

Proposition 4.8 (controls)**.** *At $\gamma = 8$:*

(i) *a greedy run exists, namely $(j_0, j_1) = (0,1)$: both conditions of* (6) *hold;*
(ii) *the ratio is not at least* 7*: $7\, G_8(\{1,2\}) \le G_8(\{0,1\})$ is false;*
(iii) *the ratio is not at most* 5*: $G_8(\{0,1\}) \le 5\, G_8(\{1,2\})$ is false;*
(iv) *the trap is strict: $G_8(\{1,2\}) < G_8(\{0,1\})$ and $G_8(\{0\}) < G_8(\{1\})$.*

Proof. All four are read off the seven values of Theorem 4.7: (i) $\frac{3}{4} \le \frac{3}{4}, 1, 1$ and $\frac{3}{4} \le \frac{3}{4}, \frac{3}{4}$; (ii) $\frac{7}{8} > \frac{3}{4}$; (iii) $\frac{3}{4} > \frac{5}{8}$; (iv) $\frac{1}{8} < \frac{3}{4}$ and $\frac{3}{4} < 1$. □

The controls guard against a misplaced quantifier: (i) shows that the hypotheses of Theorems 4.2, 4.6 and 4.7 are satisfiable, and (ii)–(iv) show that the value 6 is neither an artefact of a too-weak nor of a too-strong bound.

Remark 4.9 (larger d and D). Appending zero columns to X leaves every value of G on the original supports unchanged and makes the new columns worthless, so the unboundedness of Theorem 4.2 persists for every $D \ge 3$ at $d = 2$; a sharp statement for general d is not attempted here. This remark is not part of the formal development.

5. Computations outside the formal development

The results of Section 4 are complete as stated; this section reports how the instance was found and what the exact values look like at other ridge parameters. Nothing here is used in the proofs above, and nothing here is machine-checked.

5.1. **An exact-rational dual solver.** The maximisation in (3) is a concave quadratic over the box $[-b, h]^n$ cut by one hyperplane. For $n = 4$ we enumerated all $3^4 = 81$ active-set patterns (each coordinate at its lower bound, free, or at its upper bound), solved each KKT system exactly over $\mathbb{Q}$, and kept the pattern whose solution satisfies the sign conditions. Because the objective is concave, a KKT point is a global maximiser, so the value found is $\max \Phi_S$ exactly. The primal optimum was then recovered as $\omega_S = -\frac{\gamma}{n} X_S^\top z$ with c minimised over the residual breakpoints, and the solver asserts that the primal value equals the dual value on every support — a check that never failed. (This is where strong duality is used: in the exploration only, never in a proof.)

5.2. **The values at five ridge parameters.** Table 1 lists the exact values returned by the solver. The loss ratio is $G(S_{\mathrm{gr}})/G(\{1,2\})$ with $S_{\mathrm{gr}} = \{0,1\}$ (the run $(0,2)$ gives the same value), and the gain ratio is $(G(\emptyset) - G(S_{\mathrm{gr}}))/(G(\emptyset) - G(\{1,2\}))$.

At $\gamma \in \{8, 16, 32, 64\}$ the table is consistent with the closed forms

$$G(\{0\}) = G(\{0,1\}) = G(\{0,2\}) = \tfrac{1}{2} + \tfrac{2}{\gamma}, \qquad G(\{1,2\}) = \tfrac{1}{\gamma},$$

γ	$G(\emptyset)$	$G(\{0\})$	$G(\{1\})$	$G(\{2\})$	$G(\{0,1\})$	$G(\{0,2\})$	$G(\{1,2\})$	loss ratio	gain ratio
4	1	7/8	1	1	7/8	7/8	1/4	7/2	1/6
8	1	3/4	1	1	3/4	3/4	1/8	6	2/7
16	1	5/8	1	1	5/8	5/8	1/16	10	2/5
32	1	9/16	1	1	9/16	9/16	1/32	18	14/31
64	1	17/32	1	1	17/32	17/32	1/64	34	10/21

TABLE 1. Exact values of G on the instance (5), computed by the exact-rational solver. Only the line $\gamma = 8$ is machine-checked (Theorem 4.7).

hence with a loss ratio $\frac{\gamma}{2} + 2$ and a gain ratio $\frac{\gamma-4}{2(\gamma-1)} \to \frac{1}{2}$. Of these, the inequalities $G(\{0\}) \le \frac{1}{2} + \frac{2}{\gamma}$, $G(\{1,2\}) \le \frac{1}{\gamma}$ and $G(\{0,1\}), G(\{0,2\}) \ge \frac{1}{2}$ hold for every $\gamma > 0$ and are formal (Lemmas 3.2 and 3.3); the equalities are formal at $\gamma = 8$ only (Theorem 4.7), and at the other three parameters rest on the solver. The closed forms are not claimed for any other γ. At $\gamma = 4$ the decoy is still preferred ($\frac{7}{8} < 1$) but the primal bound $\frac{1}{2} + \frac{2}{\gamma} = 1$ is no longer tight.

5.3. **A partial search over decoy columns.** To see whether the clean parity construction is extremal among small integer instances, we fixed columns $1, 2$ and y as in (5) and replaced column 0 by each of the $5^4 = 3{,}125$ vectors in $\{-2, \dots, 2\}^4$, at $\gamma \in \{4, 8, 16\}$ (about 0.3 s per instance for all seven supports). The largest tie-free greedy ratio found was $325/32 = 10.15625$ at $\gamma = 16$, for the decoy column $(2, -2, -1, 0)$ (reported by the founding record, not reproduced here), above the value 10 of the parity construction at the same γ. So the maximum of the ratio over a grid of integer instances is not attained by the construction of Section 3, and any statement about a grid maximum must name the grid.

6. Open questions, with their cost

(1) *A named finite grid.* What is $\max G(S_{\mathrm{gr}})/G(S_{\mathrm{opt}})$ over all instances with $D = 3$, $n = 4$, $b = h = 1$, $X \in \{-2, \dots, 2\}^{4\times 3}$, $y \in \{-2, \dots, 2\}^4$ and $\gamma \in \{1, 2, 4, 8, 16\}$? Without symmetry reduction this is $5^{16} \times 5$ instances at roughly 0.3 s each, of the order of 6×10^7 CPU-hours, and is not attempted; the symmetry group (column and row permutations, sign flips, scaling) must be quotiented first. The partial search above shows the answer exceeds 10 at $\gamma = 16$.

(2) *The infimum of the gain ratio.* Is $\inf\,(G(\emptyset) - G(S_{\mathrm{gr}}))/(G(\emptyset) - G(S_{\mathrm{opt}}))$ over all instances equal to 0? The present instance gives $\frac{1}{2}$ in the limit. Driving it to 0 needs a decoy whose own gain is ε while $G(\{0,1\})$ and $G(\{0,2\})$ stay within $O(\varepsilon)$ of $G(\emptyset)$; the γ-free certificates used here cannot do this, so a γ-dependent dual family is required. We price this at a few CPU-minutes of search plus a short formal proof; it is not attempted here.

(3) *A positive guarantee from the dual.* The maximand of (3) is a smooth concave quadratic in z over a polytope, so the smoothness side that the primal lacks may exist on the dual side. A ratio bound derived there would be a theorem rather than a counterexample, and would have to be reconciled with the unbounded loss ratio of Theorem 4.2; it can therefore only concern a different ratio, a restricted class of instances, or a bound depending on γ.

7. Verification

Propositions 4.1, 4.4 and 4.8 and Theorems 4.2, 4.6 and 4.7, together with Lemmas 2.1–3.3 and Corollary 2.3 as used in them, are verified in Lean 4 [8] with Mathlib [9]. The development has two files. The first defines, for arbitrary n and D over $\mathbb{R}$, the loss ρ, the objective F, the support predicate, the dual objective Φ_S and feasible set $\mathcal{Z}$, and $G(S)$ literally as the infimum (2) (no attainment is assumed), and proves Lemma 2.2 and Corollary 2.3. The second defines the instance (5), the five dual vectors and two policies, the two greedy conditions (6) as predicates on (γ, j_0, j_1), and proves the six statements above, each quantified over every γ in the stated range and every (j_0, j_1) satisfying the conditions.

Every numeric step is a `norm_num` evaluation on rational literals (the vector components are exposed by definitional evaluation lemmas); Theorem 4.6 uses the Mathlib bound `Real.exp_one_gt_d9`. No `native_decide`, no `sorry` and no additional axiom is used: for each of the twelve declarations listed in Appendix A (the ten pinned statements plus weak duality and the certificate principle), `#print axioms` reports exactly `propext`, `Classical.choice` and `Quot.sound`. The computations of Section 5 are not formalised. Repository reference to be added at submission.

Appendix A. The statements as machine-checked

The declarations below are quoted verbatim from the formal development, docstrings included. `Gg g S` is $G_\gamma(S)$ at $\gamma = \mathtt{g}$ on the instance (with $b = h = 1$), `Step1 g j₀` and `Step2 g j₀ j₁` are the two conditions of (6), and `Fin 3` indexes the columns $0, 1, 2$. The labels "(5)" and "(6)" in two docstrings of the first file are the names of the source's internal LaTeX labels for its sparse problem and its support form, not its printed display numbers, which in version 1 are (3) and (5) (Section 1.1).

Vocabulary (first file):

```
/-- The newsvendor loss of a residual `r = y − ωᵀx − c`: back-ordering cost `b` on a
shortage, holding cost `h` on an overage.  This is the bracket of the source's `F`. -/
noncomputable def pin (b h r : ℝ) : ℝ := b * max 0 r + h * max 0 (-r)

/-- The source's empirical objective `F(ω,c)` (its Problem (5)). -/
noncomputable def obj (X : Fin n → Fin D → ℝ) (y : Fin n → ℝ) (b h g : ℝ)
    (w : Fin D → ℝ) (c : ℝ) : ℝ :=
  (∑ i, pin b h (y i - (∑ j, w j * X i j) - c)) / n + (∑ j, (w j) ^ 2) / (2 * g)

/-- `w` is supported on `S`. -/
def Supported (S : Finset (Fin D)) (w : Fin D → ℝ) : Prop := ∀ j, j ∉ S → w j = 0

/-- The source's dual objective (its minimax reformulation of (6)):
`−zᵀ(γ/(2n²))X_S X_Sᵀ z − (1/n) zᵀ y`, written with `X_S X_Sᵀ` expanded as
`Σ_{j ∈ S} (Σ_i z_i X_{ij})²`. -/
noncomputable def dual (X : Fin n → Fin D → ℝ) (y : Fin n → ℝ) (g : ℝ)
    (S : Finset (Fin D)) (z : Fin n → ℝ) : ℝ :=
  -(g / (2 * (n : ℝ) ^ 2)) * (∑ j ∈ S, (∑ i, z i * X i j) ^ 2) - (∑ i, z i * y i) / n

/-- The dual feasible set `{z ∈ [−b,h]ⁿ : Σ z = 0}`. -/
def Feas (b h : ℝ) (z : Fin n → ℝ) : Prop := (∀ i, -b ≤ z i ∧ z i ≤ h) ∧ ∑ i, z i = 0

/-- `G(S) = min { F(ω,c) : supp ω ⊆ S }`, taken as an infimum. -/
noncomputable def Gval (X : Fin n → Fin D → ℝ) (y : Fin n → ℝ) (b h g : ℝ)
    (S : Finset (Fin D)) : ℝ :=
  sInf {t : ℝ | ∃ w c, Supported S w ∧ obj X y b h g w c = t}
```

Lemma 2.2 and Corollary 2.3 (`dual_le_obj`, `Gval_eq_of_witnesses`):

```
/-- **Weak duality** for the source's minimax reformulation: every dual feasible `z` gives a
lower bound on `F(ω,c)` for every `ω` supported on `S`.  No attainment, no strong duality. -/
theorem dual_le_obj {X : Fin n → Fin D → ℝ} {y : Fin n → ℝ} {b h g : ℝ}
    {w : Fin D → ℝ} {c : ℝ} {z : Fin n → ℝ} {S : Finset (Fin D)}
    (hn : 0 < n) (hg : 0 < g) (hz : Feas b h z) (hw : Supported S w) :
    dual X y g S z ≤ obj X y b h g w c

/-- A primal point and a dual point of the same value pin `G(S)` **exactly**. -/
theorem Gval_eq_of_witnesses {X : Fin n → Fin D → ℝ} {y : Fin n → ℝ} {b h g : ℝ}
    {w : Fin D → ℝ} {c : ℝ} {z : Fin n → ℝ} {S : Finset (Fin D)} {v : ℝ}
```

```
    (hn : 0 < n) (hg : 0 < g)
    (hbdd : BddBelow {t : ℝ | ∃ w c, Supported S w ∧ obj X y b h g w c = t})
    (hw : Supported S w) (hp : obj X y b h g w c = v)
    (hz : Feas b h z) (hd : dual X y g S z = v) : Gval X y b h g S = v
```

The instance and the greedy conditions (second file):

```
/-- The feature matrix of the instance: column `0` is the decoy, columns `1`, `2` the parity
pair. -/
def Xi : Fin 4 → Fin 3 → ℝ := ![![1, 1, 1], ![0, 1, -1], ![0, -1, 1], ![0, -1, -1]]

/-- The demand vector; `y = x⁽¹⁾ + x⁽²⁾`. -/
def yi : Fin 4 → ℝ := ![2, 0, 0, -2]

/-- `G(S)` for this instance at ridge parameter `g`, with `b = h = 1`. -/
noncomputable def Gg (g : ℝ) (S : Finset (Fin 3)) : ℝ := Gval Xi yi 1 1 g S

/-- The source's first greedy iterate: `j₀ ∈ argmin_j G({j})`. -/
def Step1 (g : ℝ) (j₀ : Fin 3) : Prop := ∀ k : Fin 3, Gg g {j₀} ≤ Gg g {k}

/-- The source's second greedy iterate: `j₁ ∈ argmin_{k ≠ j₀} G({j₀, k})`. -/
def Step2 (g : ℝ) (j₀ j₁ : Fin 3) : Prop :=
  j₁ ≠ j₀ ∧ ∀ k : Fin 3, k ≠ j₀ → Gg g {j₀, j₁} ≤ Gg g {j₀, k}
```

Proposition 4.1 (greedy_first_is_decoy):

```
/-- **Greedy is forced into the decoy.**  For every `γ ≥ 8` the first greedy iterate is the
decoy column `0`, whatever the tie-breaking rule. -/
theorem greedy_first_is_decoy {g : ℝ} (hg : 8 ≤ g) {j₀ : Fin 3} (h : Step1 g j₀) : j₀ = 0
```

Theorem 4.2 (greedy_ratio_unbounded):

```
/-- **The greedy iterate is trapped at `1/2`.**  For every `γ ≥ 8`, every greedy run of the
source's algorithm ends at a support whose objective is at least `1/2`, while the optimum over
supports of size `≤ 2` is at most `1/γ`.  The loss ratio `G(S_greedy)/G(S_opt)` is therefore at
least `γ/2`: the greedy heuristic admits **no** constant-factor approximation guarantee. -/
theorem greedy_ratio_unbounded {g : ℝ} (hg : 8 ≤ g) {j₀ j₁ : Fin 3}
    (h1 : Step1 g j₀) (h2 : Step2 g j₀ j₁) :
    (1 : ℝ) / 2 ≤ Gg g {j₀, j₁} ∧ Gg g {1, 2} ≤ 1 / g ∧
      (g / 2) * Gg g {1, 2} ≤ Gg g {j₀, j₁}
```

Proposition 4.4 (not_submodular):

```
/-- **`G` has increasing marginal returns.**  With `S = ∅ ⊆ T = {1}` and `j = 2`,
`G(∅) − G({2}) = 0` but `G({1}) − G({1,2}) ≥ 1 − 1/γ > 0`: the gain function
`f(S) = G(∅) − G(S)` that the weak-submodularity literature works with is **not** submodular
for this `G`, which is exactly the property the source says is missing. -/
theorem not_submodular {g : ℝ} (hg : 1 < g) :
    Gg g ∅ - Gg g {2} < Gg g {1} - Gg g {1, 2}
```

Theorem 4.6 (greedy_below_one_minus_inv_e):

```
/-- **Greedy on `G` breaks the classical submodular guarantee.**  For a monotone submodular
gain the greedy set satisfies `f(S_greedy) ≥ (1 − 1/e) · f(S_opt)`; here, for every `γ ≥ 8`,
the greedy gain is *strictly below* that bound.  So no `(1 − 1/e)` guarantee — and a fortiori
no submodularity of the gain function — can hold for the source's `G`. -/
theorem greedy_below_one_minus_inv_e {g : ℝ} (hg : 8 ≤ g) {j₀ j₁ : Fin 3}
    (h1 : Step1 g j₀) (h2 : Step2 g j₀ j₁) :
    Gg g ∅ - Gg g {j₀, j₁} < (1 - Real.exp (-1)) * (Gg g ∅ - Gg g {1, 2})
```

Theorem 4.7 (values_at_eight, greedy_ratio_eight):

```
/-- **All seven values of `G` at `γ = 8`**, each pinned by a matching primal–dual pair.  The
optimum over supports of size `≤ 2` is `G({1,2}) = 1/8`, every greedy run ends at `3/4`, and
the exact worst-case ratio on this instance is `(3/4)/(1/8) = 6`. -/
theorem values_at_eight :
    Gg 8 ∅ = 1 ∧ Gg 8 {0} = 3 / 4 ∧ Gg 8 {1} = 1 ∧ Gg 8 {2} = 1 ∧
      Gg 8 {0, 1} = 3 / 4 ∧ Gg 8 {0, 2} = 3 / 4 ∧ Gg 8 {1, 2} = 1 / 8

/-- **The exact greedy ratio at `γ = 8` is `6`**: every greedy run ends at `3/4 = 6 · (1/8)`,
and `{1,2}` is the optimal support of size `≤ 2` (the seven values of `values_at_eight` are
`1, 3/4, 1, 1, 3/4, 3/4, 1/8`). -/
theorem greedy_ratio_eight {j₀ j₁ : Fin 3} (h1 : Step1 8 j₀) (h2 : Step2 8 j₀ j₁) :
    Gg 8 {j₀, j₁} = 6 * Gg 8 {1, 2}
```

Proposition 4.8 (`greedy_run_exists`, `ratio_not_seven`, `ratio_not_five`, `trap_is_strict`):

```
/-- Non-vacuity: a greedy run exists at `γ = 8`, namely `j₀ = 0`, `j₁ = 1`. -/
theorem greedy_run_exists : Step1 8 0 ∧ Step2 8 0 1

/-- Negative control (too large): the greedy ratio at `γ = 8` is **not** `≥ 7`. -/
theorem ratio_not_seven : ¬ (7 * Gg 8 {1, 2} ≤ Gg 8 {0, 1})

/-- Negative control (too small): the greedy ratio at `γ = 8` is **not** `≤ 5`. -/
theorem ratio_not_five : ¬ (Gg 8 {0, 1} ≤ 5 * Gg 8 {1, 2})

/-- Control: the trap is real — greedy's set is strictly worse than the optimum, and the decoy
strictly wins the first round, so no hypothesis of `greedy_ratio_unbounded` is vacuous. -/
theorem trap_is_strict : Gg 8 {1, 2} < Gg 8 {0, 1} ∧ Gg 8 {0} < Gg 8 {1}
```

References

[1] Z. Yuan, J. Wang, Variable selection for feature-based newsvendor, arXiv:2609.01544v1 (2026).

[2] A. Das, D. Kempe, Submodular meets spectral: greedy algorithms for subset selection, sparse approximation and dictionary selection, Proceedings of the 28th International Conference on Machine Learning (ICML 2011), 1057–1064; arXiv:1102.3975v2.

[3] E. R. Elenberg, R. Khanna, A. G. Dimakis, S. Negahban, Restricted strong convexity implies weak submodularity, The Annals of Statistics 46 (2018), no. 6B, 3539–3568, doi:10.1214/17-AOS1679; arXiv:1612.00804v2.

[4] W. Xie, X. Deng, Scalable algorithms for the sparse ridge regression, SIAM Journal on Optimization 30 (2020), no. 4, 3359–3386, doi:10.1137/19M1245414; arXiv:1806.03756.

[5] G. L. Nemhauser, L. A. Wolsey, M. L. Fisher, An analysis of approximations for maximizing submodular set functions – I, Mathematical Programming 14 (1978), no. 1, 265–294, doi:10.1007/BF01588971.

[6] R. Koenker, G. Bassett, Regression quantiles, Econometrica 46 (1978), no. 1, 33–50, doi:10.2307/1913643.

[7] R. F. Barber, E. Y. Sidky, Convergence for nonconvex ADMM, with applications to CT imaging, Journal of Machine Learning Research 25 (2024), 38:1–38:46; arXiv:2006.07278.

[8] L. de Moura, S. Ullrich, The Lean 4 theorem prover and programming language, Automated Deduction – CADE 28, Lecture Notes in Computer Science 12699 (2021), 625–635, doi:10.1007/978-3-030-79876-5_37.

[9] The mathlib Community, The Lean mathematical library, Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), 367–381, doi:10.1145/3372885.3373824.