SIAM J. COMPUT.
Vol. 38, No. 6, pp. 2382–2410


# DYNAMIC PROGRAMMING OPTIMIZATION OVER RANDOM DATA: THE SCALING EXPONENT FOR NEAR-OPTIMAL SOLUTIONS*

DAVID J. ALDOUS†, CHARLES BORDENAVE‡, AND MARC LELARGE§

**Abstract.** A very simple example of an algorithmic problem solvable by dynamic programming is to maximize, over $A \subseteq \{1, 2, \ldots, n\}$, the objective function $|A| - \sum_i \xi_i \mathbb{1}(i \in A, i+1 \in A)$ for given $\xi_i > 0$. This problem, with random $(\xi_i)$, provides a test example for studying the relationship between optimal and near-optimal solutions of combinatorial optimization problems. We show that, amongst solutions differing from the optimal solution in a small proportion $\delta$ of places, we can find near-optimal solutions whose objective function value differs from the optimum by a factor of order $\delta^2$ but not of smaller order. We conjecture this relationship holds widely in the context of dynamic programming over random data, and Monte Carlo simulations for the Kauffman–Levin NK model are consistent with the conjecture. This work is a technical contribution to a broad program initiated in [D. J. Aldous and A. G. Percus, *Proc. Natl. Acad. Sci. USA*, 100 (2003), pp. 11211–11215] of relating such scaling exponents to the algorithmic difficulty of optimization problems.

**Key words.** dynamic programming, local weak convergence, Markov chain, near-optimal solutions, optimization, probabilistic analysis of algorithms, scaling exponent

**AMS subject classifications.** 68Q25, 90C39, 60J05

**DOI.** 10.1137/070709037

## 1. Introduction and motivation.

**1.1. Near-optimal solutions in combinatorial optimization.** Consider a combinatorial optimization problem which is "size $n$" in the sense that a feasible solution $\mathbf{x} = (x_i, 1 \leq i \leq n)$ consists of $n$ elements (e.g., edges of a graph; binary digits) subject to some constraints, and the objective function $f(\mathbf{x})$ is akin to a sum over $i$ of costs or rewards associated with each $x_i$. In such a setting one can define the relative distance between the structure of a feasible solution $\mathbf{x}$ and the optimal solution $\mathbf{x}^*$ by

$$\delta_n(\mathbf{x}) = n^{-1}|\{i : x_i \neq x_i^*\}|,$$

and the relative difference in objective function is $n^{-1}|f(\mathbf{x}) - f(\mathbf{x}^*)|$. So the quantity

$$\varepsilon_n(\delta) := \min\{n^{-1}|f(\mathbf{x}) - f(\mathbf{x}^*)| \ : \delta_n(\mathbf{x}) \geq \delta\} \tag{1}$$

measures how close we can get to the optimal value using feasible solutions which have nonnegligibly different structure from the optimal solution. A program initiated in [3] is to study this quantity for combinatorial optimization problems over *random* data. In this setting $\varepsilon_n(\delta)$ becomes a random variable, but in many cases one expects that

*Received by the editors November 26, 2007; accepted for publication (in revised form) December 17, 2008; published electronically March 20, 2009.
http://www.siam.org/journals/sicomp/38-6/70903.html
†Department of Statistics, University of California, Berkeley, CA 94270 (aldous@stat.berkeley.edu). This author's research was supported by NSF grant DMS0704159.
‡Université de Toulouse & CNRS, Institut de Mathématiques de Toulouse, F-31062 Toulouse, France (charles.bordenave@math.univ-toulouse.fr). This author's research was supported by NSF grant CCF-050023.
§INRIA-ENS, 75005 Paris, France (marc.lelarge@ens.fr).

as $n \to \infty$ there is a *deterministic* limit function $\varepsilon(\delta)$. Motivation for this program is a conjecture that (within some suitable class of problems)

$$\varepsilon(\delta) \asymp \delta^\alpha \text{ as } \delta \to 0$$

for some *scaling exponent* $\alpha$, whose value is robust under model details, and that for "algorithmically easy" problems we have $\alpha = 2$ (which of course mimics the behavior we expect by calculus for smooth functions $f : \mathbb{R}^d \to \mathbb{R}$) whereas for "algorithmically hard" problems we have $\alpha > 2$. Here is the previous evidence in support of this conjecture.

(i) *Traveling salesman problem* and *minimum matching problem* [3]. In the random link (mean-field) model, a cavity method analysis (nonrigorous but generally regarded as accurate) enables one to compute $\varepsilon(\delta)$ numerically and to observe scaling exponent $\alpha = 3$. In the random Euclidean model, Monte Carlo simulations suggest the same $\alpha = 3$.

(ii) *Minimum spanning tree*. Here we expect $\alpha = 2$. This is proved in [2] for the $d \geq 2$ dimensional random Euclidean model and also for a "disordered lattice" model.

The purpose of this paper is to consider some problems which are algorithmically easy to solve via dynamic programming, and where we therefore expect $\alpha = 2$. We first give a trivial but instructive case (section 1.2) and then describe a prototypical "interesting" case, the Kauffman–Levin NK model (section 1.3). Here both a heuristic argument and simulations suggest $\alpha = 2$, but we do not have a proof. Our main focus is on giving a complete analysis of a simple nontrivial model (section 1.4), where we are required to pick a subset $A \subseteq [n] := \{1, 2, \ldots, n\}$ of items with a reward of 1 per item picked and independent and identically distributed (i.i.d.) costs $\xi_i$ incurred if both items $i$ and $i+1$ are picked. Theorem 2 establishes $\alpha = 2$ for this specific model. In these dynamic programming examples and the minimum spanning tree example, the key structural property is that the near-optimal solutions attaining the minimum in (1) differ from the optimal solution via only "local changes," each local change affecting only a number of items which remains $O(1)$ as $\delta \to 0$. It is natural to speculate that this structural property corresponds quite generally to the $\alpha = 2$ case.

*Related work.* We do not know any other lines of research in theoretical computer science which are close to the topic of this paper. A recent survey of average-case complexity of NP problems is given in [7]. Interest in the average-case gap between optimal and second-optimal solutions arises in several contexts; see, e.g., [5]. Closer in spirit is the statistical physics of disordered systems, where for low temperatures the Gibbs distribution on configurations concentrates on near-minimal-cost configurations. In the context of random energy models (the precise analogue of optimization over random data), two random picks from the Gibbs distribution over the same random choice of energy are called *replicas*, and study of such replicas and their overlaps is a central theme of the *replica method* [15, 17]. So that topic studies the structural difference between two *typical* near-optimal configurations, whereas we study the *maximal* (over all near-optimal configurations) structural difference from the optimal configuration. Our mathematical arguments are much less sophisticated than those in statistical physics, but there are some intriguing parallels, described briefly in section 5.2.

**1.2. A trivial example.** Let $(X_i, i \geq 1)$ be i.i.d. real-valued random variables with continuous density $h(x)$ and $\mathrm{E}X < \infty$. For each $n$ consider the problem of finding

$$M_n = \max_{A \subseteq [n]} \sum_{i \in A} (X_i - 1).$$

The maximum is obviously obtained by choosing $A = \{i : X_i > 1\}$ and then as $n \to \infty$

$$n^{-1} M_n \to \mathrm{E}(X_1 - 1)^+ \text{ a.s.}$$

Fix $0 < \delta < 1$. It is also obvious that the subset $A'$ that minimizes

$$M'_n = \max_{A' \subseteq [n]} \sum_{i \in A'} (X_i - 1)$$

$$\text{subject to } |A' \triangle A| \geq \delta n$$

is the subset $A' = A \triangle D$, where $D$ is the set of indices of the $\lceil \delta n \rceil$ smallest values of $|X_i - 1|$. So as $n \to \infty$

$$n^{-1}(M_n - M'_n) \to_{L_1} \varepsilon(\delta) := \int_{1-a(\delta)}^{1+a(\delta)} |x - 1| h(x)\ dx,$$

where $a(\delta)$ is defined by

$$\delta = \int_{1-a(\delta)}^{1+a(\delta)} h(x)\ dx.$$

So by continuity of $h(x)$, and assuming $0 < h(1) < \infty$, as $\delta \downarrow 0$ we have

$$a(\delta) \sim \tfrac{\delta}{2h(1)}; \quad \varepsilon(\delta) \sim a^2(\delta) h(1) \sim \tfrac{\delta^2}{4h(1)}, \tag{2}$$

which is the desired "scaling exponent = 2" result.

*Discussion.* (i) This example illustrates a feature that arises in other examples, that proving $\alpha = 2$ reduces to showing that the density of a certain measure at a certain point is finite and nonzero. In nontrivial examples the measure in question arises in the *analysis* of the problem rather than the statement of the problem: see Lemma 19 below and Proposition 8 of [2].

(ii) In this example we could see the form of the best near-optimal solution by inspection, but a systematic method is to use Lagrange multipliers. In this example, introduce a parameter $\theta > 0$ and consider for each $n$

$$A_\theta := \arg\max_A \left( \sum_{i \in A} (X_i - 1) + \theta |A \triangle A^*| \right),$$

where $A^* = \{i : X_i > 1\}$ is the optimal solution. By inspection the solution is

$$A_\theta = \{i : \ 1 - \theta \leq X_i \leq 1 \text{ or } 1 + \theta \leq X_i\}.$$

Although now $|A_\theta \triangle A^*|$ is random, we can use the law of large numbers to obtain existence of the limits

$$\delta(\theta) := \lim_{n \to \infty} n^{-1} |A^* \triangle A_\theta| = \int_{1-\theta}^{1+\theta} h(x)\ dx,$$

$$\varepsilon(\theta) := \lim_{n \to \infty} n^{-1} \left( \sum_{i \in A^*} (X_i - 1) - \sum_{i \in A_\theta} (X_i - 1) \right) = \int_{1-\theta}^{1+\theta} |x - 1|\ dx.$$

By the interpretation of Lagrange multipliers, this is an implicit function representation of $\varepsilon$ as a function of $\delta$ and rederives the limit (2) above.

011000110100101110100010011010101011011110001010110 10 $\mathbf{x}^N$
01100011011010111010001001101010001110100000101011010 $\mathbf{y}$

FIG. 1. *Excursions of lengths $l = 3$ and 11. Here $K = 2$.*

**1.3. The NK model.** The Kauffman–Levin NK model of random fitness landscape has attracted extensive literature in statistical physics [10, 19] and has been studied by probabilists [9, 11]. For our version of the model we fix $K \geq 2$. We seek to minimize, over binary sequences $\mathbf{x} = (x_1, \dots, x_N)$, the objective function $H_N(\mathbf{x}) = \sum_{i=1}^{N-K} W_i(x_i, x_{i+1}, \dots, x_{i+K})$, where the values $(W_i(b_0, b_1, \dots, b_K) : i \geq 1, \mathbf{b} \in \{0,1\}^{K+1})$ are independent exponential(1) random variables. This is algorithmically easy via dynamic programming. Write $\mathbf{x}^N$ for the minimizing sequence. By subadditivity there is an a.s. limit $N^{-1}H_N(\mathbf{x}^N) \to c_K$. For a general sequence $\mathbf{y} = \mathbf{y}^N$ write

$$\delta_N(\mathbf{y}) = N^{-1}|\{1 \leq i \leq N-K : (y_i, \dots, y_{i+K}) \neq (x_i^N, \dots, x_{i+K}^N)\}|,$$

$$\varepsilon_N(\mathbf{y}) = N^{-1}(H_N(\mathbf{y}) - H_N(\mathbf{x}^N))$$

and then set

$$\varepsilon_N(\delta) = \min\{\varepsilon_N(\mathbf{y}) : \ \delta_N(\mathbf{y}) \geq \delta\}. \tag{3}$$

We expect existence of a deterministic limit

$$\varepsilon(\delta) = \text{ a.s.-}\lim_{N\to\infty} \varepsilon_N(\delta).$$

*A heuristic analysis.* The purpose of this section is to give a heuristic argument for $\varepsilon(\delta) \asymp \delta^2$. Given $i$ and $l \geq K+1$, consider the set of sequences $\mathbf{y}$ such that

$$(y_j, \dots, y_{j+K}) = (x_j^N, \dots, x_{j+K}^N) \quad \forall j \notin [i+1, i+l],$$

$$(y_j, \dots, y_{j+K}) \neq (x_j^N, \dots, x_{j+K}^N) \quad \forall j \in [i+1, i+l].$$

Over this set, let $D_{i,l}$ be the minimum of $H_N(\mathbf{y}) - H_N(\mathbf{x}^N)$ and let $\mathbf{y}^{(i,l)}$ be the minimizing sequence. The distribution of $D_{i,l}$ essentially depends only on $l$, not on $i$ or $N$; write $f_l(0+)$ for its density at $0+$. Let us assume

$$\sum_{l \geq K+1} l^2 f_l(0+) = A < \infty. \tag{4}$$

It is intuitively clear how to choose a sequence $\mathbf{y}$ which minimizes $\varepsilon_N(\mathbf{y})$ for a given $\delta$. Just fix a small $\eta > 0$ and create a sequence of "excursion" away from $\mathbf{x}^N$ as follows. For each pair $(i,l)$ such that $D_{i,l} < \eta l$, choose $\mathbf{y}$ to equal $\mathbf{y}^{(i,l)}$ on the sites $[i+K+1, i+l]$; set $\mathbf{y} = \mathbf{x}^N$ elsewhere. See Figure 1.

With this scheme, $\delta$ will be the mean length of possible excursions starting from a given site, that is,

$$\delta \sim \sum_{l \geq K+1} l \cdot \eta l f_l(0+).$$

Table 1

*Monte Carlo simulations with $K = 3, N = 10,000$; 1000 repeats. These are exact optimizations done by introducing a Lagrange multiplier $\theta$ which penalizes matching $(K+1)$-tuples. We find $c_3 = 0.3065$.*

| $\theta$ | $\delta$ | $\varepsilon$ | $\varepsilon/\delta^2$ | $\mathrm{E}L_\delta$ |
|---|---|---|---|---|
| 0.002 | 0.0397 | $4.85 \cdot 10^{-5}$ | 0.0308 | 10.9 |
| 0.004 | 0.0774 | $2.00 \cdot 10^{-4}$ | 0.0334 | 11.0 |
| 0.008 | 0.147 | $7.69 \cdot 10^{-4}$ | 0.0354 | 11.3 |
| 0.016 | 0.266 | $2.75 \cdot 10^{-3}$ | 0.0388 | 11.8 |

And $\varepsilon$ is the mean increment of $H_N$ associated with possible excursions starting from a given site, that is,

$$\varepsilon \sim \sum_{l \geq K+1} (\eta l/2) \cdot \eta l f_l(0+).$$

In other words $\delta \sim A\eta,\ \varepsilon \sim A\eta^2/2$, giving $\varepsilon \sim (2A)^{-1}\delta^2$, which is the desired "scaling exponent $= 2$" result.

Why should assumption (4) be true? Well, for large $l$ we expect central limit behavior: $D_l \approx \mathrm{Normal}(\mu l, \sigma^2 l)$ for some $\mu > 0$ and $0 < \sigma^2 < \infty$. This in turn suggests that $f_l(0+)$ should decrease at least geometrically fast in $l$.

Note that the optimizing $\mathbf{y}^N$ in (3) will have (in the $N \to \infty$ limit) some distribution $L_\delta$ of excursion lengths. The heuristic argument predicts that as $\delta \downarrow 0$ we have $L_\delta \xrightarrow{d} L$, where the limit distribution has $\mathrm{P}(L = l) \propto l f_l(0+)$ and $\mathrm{E}L < \infty$.

Simulations (Table 1) with $K = 3$ are consistent with both the predicted scaling exponent 2 and the prediction of existence of a $\delta \downarrow 0$ limit distribution $L$ for excursion lengths. Making a rigorous proof seems difficult, and so we turn to a simpler example.

**1.4. Main model and results.** Let $(\xi_i, i \geq 1)$ be i.i.d. copies of a strictly positive random variable $\xi$, and write $G(x) = \mathrm{P}(\xi \leq x)$. Define the *benefit* function

$$f_n(A) = \left( |A| - \sum_{i=1}^{n-1} \xi_i \mathbb{1}(i \in A, i+1 \in A) \right), \quad A \subseteq \{1, 2, \ldots, n\}, \tag{5}$$

where $\mathbb{1}(B) = \mathbb{1}_B$ denotes the indicator random variable associated with an event $B$. Intuitively, we choose a set $A$ of items, getting reward 1 from each item chosen but paying cost $\xi_i$ if we choose both $i$ and $i+1$; we seek to maximize benefit = reward − cost. So we shall study

$$M_n := \max_{A \subseteq \{1,2,\ldots,n\}} f_n(A). \tag{6}$$

To simplify exposition we will assume

$$G \text{ has bounded continuous density } g \text{ with } g(\tfrac{1}{2}) > 0, \tag{7}$$

which implies

$$0 < G(\tfrac{1}{2}) < 1, \tag{8}$$

though we suspect that Theorems 1 and 2 remain true under some much weaker nondegeneracy assumptions. See section 5.1 for further remarks.

We will first prove the following.

THEOREM 1. *There exists* $\frac{1}{2} \leq c \leq 1$ *such that, a.s. and in* $L^1$*,*

$$\lim_{n\to\infty} n^{-1} M_n = c.$$

*The constant* $c$ *is given by the forthcoming formula* (31)*. If* $\xi$ *is an exponential random variable with parameter* $\lambda > 0$*, then*

$$c = (1 - e^{-\lambda})^{-1} - \lambda^{-1}.$$

We record the explicit value of $c$ only in the exponential case, but one could use formula (31) to obtain explicit values for other standard distributions.

We now formalize the setup in the introduction. The optimization problem (6) has a solution, a random subset $A_n^{\text{opt}} \subseteq \{1, 2, \ldots, n\}$, and Corollary 4 will show the solution is unique with probability $\to 1$ as $n \to \infty$. Define the random variable:

$$\varepsilon_n(\delta) := \min\left\{n^{-1}(f_n(A_n^{\text{opt}}) - f_n(B)) : |B \triangle A_n^{\text{opt}}| \geq \delta n\right\}, \tag{9}$$

where the minimum is over all subsets $B \subset \{1, \ldots, n\}$ such that the symmetric difference with $A_n^{\text{opt}}$ is at least $\delta n$. Our main result is the following.

THEOREM 2. $\bar{\varepsilon}(\delta) := \lim_n \mathrm{E}\varepsilon_n(\delta)$ *exists for all* $0 < \delta < 1$*, and*

$$\limsup_{\delta \downarrow 0} \delta^{-2} \bar{\varepsilon}(\delta) < \infty, \tag{10}$$

$$\liminf_{\delta \downarrow 0} \delta^{-2} \bar{\varepsilon}(\delta) \ > 0. \tag{11}$$

We now outline the key ideas in the proof and the organization of this paper.

Dynamic programming over i.i.d. data is essentially just study of a related Markov chain (section 2.2), and in our model there are simple *inclusion criteria* for whether item $i$ is in the optimal solution. The inclusion criterion involves two Markov chains (one looking left, one looking right) and the cost $\xi_i$ (Table 2 and Lemma 5). By considering the related infinite-time *stationary* Markov chain and using the same inclusion criteria, we can define a random subset $A^{\text{opt}} \subset \mathbb{Z}$ interpretable as the solution of an infinite optimization problem (section 2.3). The $n \to \infty$ limit benefit in Theorem 1 is just the mean benefit per item using $A^{\text{opt}}$ in the infinite problem (section 2.4).

Study of $\varepsilon_n(\delta)$ is an "optimization under constraint" problem, most naturally handled via introduction of a Lagrange multiplier $\theta$. So the $B_n^{\text{opt}}$ attaining the maximum in (9) can be studied as above by introducing a more complicated Markov chain parametrized by $\theta$ (section 4.1), finding the inclusion criteria (Table 3), formulating the parallel optimization under constraint problem, and observing that $\bar{\varepsilon}(\delta)$ is representable via functions $\delta(\theta), \varepsilon(\theta)$ defined in terms of the stationary distribution of the more complicated Markov chain (Proposition 12). Without trying to write details, it seems intuitively clear that the methodology above could be implemented in more general dynamic programming models such as the NK model of section 1.3. However, to complete the argument we need to analyze the $\theta \to 0$ behavior of the functions $\delta(\theta), \varepsilon(\theta)$. Even in our simple model, we do not have any useful explicit expression for the needed stationary distribution, so we proceed via inequalities rather than using the exact formulas. For the upper bound (10) we just identify a "local configuration" which can be replaced by a different local configuration at small extra cost (section 3). For the lower bound, we decompose the process into blocks by breaking at certain special configurations, and then we get bounds on the chance that $B_n^{\text{opt}}$ differs from

| | $g$ | $-$ | $(g+1)$ | $-$ | | | | $-$ | $(d-1)$ | $-$ | $d$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $A_1$ | $\circ$ | $-$ | $\bullet$ | $-$ | $\circ$ | $\ldots$ | $\bullet$ | $-$ | $\circ$ | $-$ | $\bullet$ |
| $A_2$ | $\bullet$ | $-$ | $\circ$ | $-$ | $\bullet$ | $\ldots$ | $\circ$ | $-$ | $\bullet$ | $-$ | $\circ$ |

FIG. 2. *Included items marked* $\bullet$*, excluded items marked* $\circ$.

$A_n^{\text{opt}}$ on a block and bounds on the mean decrease in benefit if it does differ (section 4.5). But these arguments rely on the particular combinatorial structure of our special model. It is not clear how readily they can be extended to general models.

## 2. Analysis of optimal solutions.

**2.1. Nonuniqueness.** In the case $n = 2$, if $\xi_1 > 1$, then both $\{1\}$ and $\{2\}$ attain the maximum value 1 of the optimization problem (6): the optimizing set is not unique. Corollary 4 shows that, provided some $\xi_i + \xi_{i+1} < 1$ is less than 1, the optimizing set $A_n^{\text{opt}}$ is unique, and by assumption (8) this proviso holds with probability $\to 1$ as $n \to \infty$. After this section we generally ignore the possibility of nonuniqueness.

We start with some terminology that will also be used later. For an integer interval $[g, d]$ with $d - g + 1$ even, the two *complementary alternating subsets* $A_1, A_2$ are as shown in Figure 2.

LEMMA 3. *Let* $n \geq 2$*. For almost all realizations of* $\xi_1, \ldots, \xi_{n-1}$*, the following are equivalent:*

(a) *The subset maximizing* (6) *is not unique.*

(b) $n$ *is even and the only optimal solutions are the two complementary alternating subsets of* $[1, n]$*.*

(c) $n$ *is even and* $M_n = n/2$*.*

*Proof.* Either of (b), (c) implies (a), so it is enough to show (a) implies (b) and (c). Suppose distinct subsets $B_1$ and $B_2$ attain the maximum. Then a.s. the values of $\xi_i$ used in the optimal sum are identical, that is,

$$\{i : \ (i, i+1) \subset B_1\} = \{i : \ (i, i+1) \subset B_2\} := \mathcal{S}, \text{ say.} \tag{12}$$

First suppose $\mathcal{S}$ is empty. Then each of $B_1$ and $B_2$ has only isolated elements. But amongst such sets, the maximum of (6) is attained (for $n$ odd) uniquely by the alternating subset giving $M_n = (n+1)/2$, or (for $n$ even) only by the complementary alternating subsets. So $\mathcal{S}$ empty implies (b) and (c). For general $\mathcal{S}$, take some $i \in B_1 \triangle B_2$, and then take the maximal interval $i \in [g, d] \subset [1, n]$ which is disjoint from $\mathcal{S}$. Repeating the argument above, the restrictions of $B_1$ and $B_2$ to $[g, d]$ must be complementary alternating subsets. If $[g, d] \neq [1, n]$, then either $d + 1$ or $g - 1$ is in $\mathcal{S}$—say $d + 1$—and so $d + 1 \in B_1 \cap B_2$. But exactly one of $B_1, B_2$ contains $d$, contradicting (12). So $[g, d] = [1, n]$, and so $\mathcal{S}$ is empty. ∎

COROLLARY 4. *If* $\xi_i + \xi_{i+1} < 1$ *for some* $1 \leq i \leq n - 2$*, then a.s.* $A_n^{opt}$ *is unique.*

*Proof.* Fix $i$ with $\xi_i + \xi_{i+1} < 1$ and let $B$ be the alternating subset of $[1, n]$ containing $i$ and $i+2$. Replacing $B$ by $B \cup \{i+1\}$ increases the benefit by $1 - \xi_i - \xi_{i+1} > 0$, so $B$ cannot be optimal, and the result follows from Lemma 3. ∎

**2.2. Dynamic programming.** Finding the maximum value and the maximizing subset of (6) is algorithmically easy by dynamic programming, as follows. Define

$$V_{n,i}^L = \max_{i \in A \subseteq \{1, \ldots, i-1, i\}} \left( |A| - \sum_{j=1}^{i-1} \xi_j \mathbb{1}(j \in A, j+1 \in A) \right), \tag{13}$$

$$W_{n,i}^L = \max_{i \notin A \subseteq \{1,\dots,i-1,i\}} \left( |A| - \sum_{j=1}^{i-1} \xi_j \mathbb{1}(j \in A, j+1 \in A) \right), \tag{14}$$

which differ in that the former requires $i \in A$ and the latter requires $i \notin A$. The superscripts $L$ here and $R$ later indicate *left* and *right*. Note that in fact $V_{n,i}^L, W_{n,i}^L$ above and $X_{n,i}^L$ below do not depend on $n$, but the notation is useful to distinguish from the limit process $X_i^L$ later.

From (13), (14) we see $V_{n,1}^L = 1, W_{n,1}^L = 0$, and by induction over $1 \le i$

$$V_{n,i+1}^L = 1 + \max(V_{n,i}^L - \xi_i, W_{n,i}^L),$$
$$W_{n,i+1}^L = \max(V_{n,i}^L, W_{n,i}^L),$$

the two terms in the max indicating the choice of using or not using element $i$. Then $M_n = \max(V_{n,n}^L, W_{n,n}^L)$, and by examining which max term is used at each stage leading to $M_n$ we can recover the optimizing subset $A_n^{\mathrm{opt}}$.

We now describe an alternative, more useful way to obtain $A_n^{\mathrm{opt}}$. First, consider the evolution rule for the process

$$X_{n,i}^L := V_{n,i}^L - W_{n,i}^L \tag{15}$$

as $i$ increases; the rule is

$$\begin{aligned} X_{n,i+1}^L &= 1 + \max(0, X_{n,i}^L - \xi_i) - \max(0, X_{n,i}^L) \\ &= 1 + \max(-X_{n,i}^L, -\xi_i)\mathbb{1}(X_{n,i}^L \ge 0). \end{aligned} \tag{16}$$

One can check by induction that $0 \le X_{n,i}^L \le 1$ and thus rewrite the recursion as

$$X_{n,i+1}^L = \max(1 - X_{n,i}^L, 1 - \xi_i).$$

For $n$ fixed we define the right processes analogously:

$$V_{n,i}^R = \max_{i \in A \subseteq \{i,i+1,\dots,n\}} \left( |A| - \sum_{j=i}^{n-1} \xi_j \mathbb{1}(j \in A, j+1 \in A) \right), \tag{17}$$

$$W_{n,i}^R = \max_{i \notin A \subseteq \{i,i+1,\dots,n\}} \left( |A| - \sum_{j=i}^{n-1} \xi_j \mathbb{1}(j \in A, j+1 \in A) \right), \tag{18}$$

with $V_{n,n}^R = 1, W_{n,n}^R = 0$. Observe that the evolution rule for the process

$$X_{n,i}^R := V_{n,i}^R - W_{n,i}^R \tag{19}$$

as $i$ decreases does not depend on $n$. In fact, we have

$$X_{n,i-1}^R = \max(1 - X_{n,i}^R, 1 - \xi_{i-1}). \tag{20}$$

The point is that we can determine the optimizing random set $A_n^{\mathrm{opt}}$ in terms of the quantities above. Fix $i$, consider the quantities $(X_{n,i}^L, V_{n,i}^L, W_{n,i}^L)$, $\xi_i$, and $(X_{n,i+1}^R, V_{n,i+1}^R, W_{n,i+1}^R)$, and drop subscripts. We have four choices of whether to include

TABLE 2
*Inclusion criteria for $i, i+1$ in $A_n^{opt}$.*

| $-i-(i+1)-$ | Absolute benefit | Relative benefit | When used |
|---|---|---|---|
| $-\bullet--\bullet-$ | $V^L+V^R-\xi$ | $X^L+X^R-\xi$ | $\xi<\min(X^L,X^R)$ |
| $-\bullet--\circ-$ | $V^L+W^R$ | $X^L$ | if $X^R<\min(X^L,\xi)$ |
| $-\circ--\bullet-$ | $W^L+V^R$ | $X^R$ | if $X^L<\min(X^R,\xi)$ |
| $-\circ--\circ-$ | $W^L+W^R$ | 0 | never |

(marked as $\bullet$ in Table 2) or exclude (marked as $\circ$ in Table 2) items $i$ and $i+1$ in the optimal set $A_n^{\rm opt}$. For each choice, the table shows the absolute benefit of that choice and then the relative benefit (relative to the choice to exclude both items). For each $i$ the optimal $A_n^{\rm opt}$ will contain, in positions $(i, i+1)$, the combination with the largest relative benefit, and the final column indicates the criteria for use of each combination. (The case of nonuniqueness of $A_n^{\rm opt}$, Lemma 3, is the case where $X_i^L$ and $X_i^R$ alternate between 0 and 1 throughout the interval $[1,n]$, and where we have equalities $X_i^L = X_{i+1}^R < \xi_i$. Outside this case, one of the three strict inequalities holds. We ignore the nonuniqueness possibility in the summary below.)

We summarize the argument above as follows.

LEMMA 5. *For each $n$ define $X_{n,i}^L, 1\le i\le n$, and $X_{n,i}^R, 1\le i\le n$, by*

$$X_{n,1}^L = 1;\quad X_{n,i+1}^L = \max(1-X_{n,i}^L, 1-\xi_i),\quad 1\le i\le n-1, \tag{21}$$

$$X_{n,n}^R = 1;\quad X_{n,i-1}^R = \max(1-X_{n,i}^R, 1-\xi_{i-1}),\quad 2\le i\le n. \tag{22}$$

*Then $A_n^{opt}$ is the random subset of $\{1,2,\dots,n\}$ specified by the following: for each $1\le i\le n-1$,*

$$\text{if } \xi_i<\min(X_{n,i}^L, X_{n,i+1}^R),\quad \text{then}\quad i\in A_n^{opt},\quad i+1\in A_n^{opt},$$

$$\text{if } X_{n,i+1}^R<\min(X_{n,i}^L, \xi_i),\quad \text{then}\quad i\in A_n^{opt},\quad i+1\notin A_n^{opt},$$

$$\text{if } X_{n,i}^L<\min(X_{n,i+1}^R, \xi_i),\quad \text{then}\quad i\notin A_n^{opt},\quad i+1\in A_n^{opt}.$$

Let us emphasize two points:

- whether or not $i\in A_n^{\rm opt}$ depends only on the three random variables $X_{n,i}^L$, $\xi_i, X_{n,i+1}^R$;
- the only place where the value of $n$ enters is as the boundary condition $X_{n,n}^R = 1$.

In the next section, we show how to define a corresponding stationary process

$$((X_i^L, \xi_i, X_{i+1}^R),\ -\infty<i<\infty).$$

By applying the specification in Lemma 5 to this process, we will define a set $A^{\rm opt}\subseteq\mathbb{Z}$ which will be shown (Lemma 8) to be the limit of $A_n^{\rm opt}$. As a consequence, we will be able to derive the limit of $M_n/n$.

**2.3. A stationary Markov chain and the infinite limit problem.** The recursion (21) specifies a Markov chain on the continuous state space $[0,1]$ with transitions

$$x\to\max(1-x, 1-\xi), \tag{23}$$

where $\xi$ has distribution function $G$. Write $F(x) = \mathrm{P}(X^L \le x)$ for a stationary distribution function for this chain. Then

$$\begin{aligned} F(x) &= \mathrm{P}(\max(1 - X^L, 1 - \xi) \le x) \\ &= \mathrm{P}(\min(X^L, \xi) > 1 - x) \\ &= \overline{G}(1 - x)\overline{F}(1 - x), \end{aligned}$$

where for any distribution function $F$ we write $\overline{F}(x) = 1 - F(x)$. Iterating this identity once gives

$$F(x) = \overline{G}(1 - x)\left(1 - \overline{G}(x)\overline{F}(x)\right),$$

and solving this equation gives

$$F(x) = \frac{\overline{G}(1 - x)\overline{G}(x)}{1 - \overline{G}(x)\overline{G}(1 - x)}. \tag{24}$$

The assumption (7) that $G$ has a density implies that $F$ has a density, so in what follows we do not need to distinguish carefully between weak and strict inequalities for random variables with these distributions.

Now consider the infinite line graph, with vertices $-\infty < i < \infty$ and with i.i.d. edge-costs $\xi_i$ on edge $(i, i + 1)$ such that $\mathrm{P}(\xi_0 + \xi_1 < 1) > 0$, which is ensured by the condition $\overline{G}(1/2) < 1$.

LEMMA 6. *The recursion*

$$X^L_{i+1} = \max(1 - X^L_i, 1 - \xi_i), \quad -\infty < i < \infty, \tag{25}$$

*defines uniquely a joint distribution for* $((\xi_i, X^L_i), -\infty < i < \infty)$ *in which* $(X^L_i)$ *is the stationary Markov chain with transition kernel* (23) *and stationary distribution* (24). *And*

$$X^L_i = \phi(\dots, \xi_{i-2}, \xi_{i-1}) \tag{26}$$

*for a certain function* $\phi$ *not depending on* $i$.

*Proof.* Having proved existence and uniqueness of the stationary distribution at (24), it remains only to prove the measurability property (26). Iterating (25) once shows

$$1 - \xi_i \le X^L_{i+1} \le \max(1 - \xi_i, \xi_{i-1}). \tag{27}$$

So outside the event $\{1 - \xi_i < \xi_{i-1}\}$ the value of $X^L_{i+1}$ depends only on $(\xi_{i-1}, \xi_i)$ and not on the value of $X^L_i$. So inductively on $Q \ge 1$ there exists a measurable function $\phi_Q$ such that

$$X^L_1 = \phi_Q(\xi_{-2Q-1}, \xi_{-2Q}, \dots, \xi_0) \text{ outside } \cap_{q=-Q}^{0} \{1 - \xi_{2q} < \xi_{2q-1}\}.$$

Now (26) follows because $\mathrm{P}\left(\cap_{q=-Q}^{0}\{1 - \xi_{2q} < \xi_{2q-1}\}\right) = (\mathrm{P}(\xi_0 + \xi_1 > 1))^{Q+1} \to 0.$ ∎

If we define an "$i$ decreasing" process by

$$X^R_i = \phi(\dots, \xi_{i+2}, \xi_{i+1}, \xi_i), \tag{28}$$

then $(X_i^R)$ satisfies the analogous recursion

$$X_i^R = \max(1 - X_{i+1}^R, 1 - \xi_i), \quad -\infty < i < \infty, \tag{29}$$

and is distributed as the same stationary Markov chain. Hence we have a rigorous definition of a unique (in distribution) stationary process $((X_i^L, \xi_i, X_{i+1}^R),\ -\infty < i < \infty)$ satisfying (25), (29) which we will call the *triple process*. Note that from (26), (28)

$$\text{for each } i \text{ the three random variables } X_i^L, \xi_i, X_{i+1}^R \text{ are independent.} \tag{30}$$

LEMMA 7. *Let $(X_i^L, \xi_i, X_{i+1}^R)$, $-\infty < i < \infty$, be the stationary triple process. Then there is a random subset $A^{opt}$ of $\mathbb{Z}$ specified by the following: for each $-\infty < i < \infty$,*

$$\begin{aligned}
&\textit{if } \xi_i < \min(X_i^L, X_{i+1}^R), \quad \textit{then} \quad i \in A^{opt}, \quad i+1 \in A^{opt},\\
&\textit{if } X_{i+1}^R < \min(X_i^L, \xi_i), \quad \textit{then} \quad i \in A^{opt}, \quad i+1 \notin A^{opt},\\
&\textit{if } X_i^L < \min(X_{i+1}^R, \xi_i), \quad \textit{then} \quad i \notin A^{opt}, \quad i+1 \in A^{opt}.
\end{aligned}$$

*Proof.* We need only check that the definition of $A^{\text{opt}}$ is consistent, in that the criterion for item $i$ to be excluded should be the same whether we look at the pair $(i, i+1)$ or the pair $(i-1, i)$. (Of course this is intuitively clear from the consistency in the finite setting of Lemma 5, but let us give an algebraic verification anyway.) We need to check

$$\{X_i^L < \min(X_{i+1}^R, \xi_i)\} \stackrel{?}{=} \{X_i^R < \min(X_{i-1}^L, \xi_{i-1})\}.$$

Using the recursions (29), (25) for $X_i^R$ and $X_i^L$, we need to check

$$\{\max(1-X_{i-1}^L, 1-\xi_{i-1}) < \min(X_{i+1}^R, \xi_i)\} \stackrel{?}{=} \{\max(1-X_{i+1}^R, 1-\xi_i) < \min(X_{i-1}^L, \xi_{i-1})\}.$$

But these are equal by applying the transformation $u \to 1-u$ to the right-hand side. ∎

Because the rule defining $A^{\text{opt}}$ is translation-invariant, the augmented triple process

$$((X_i^L, \xi_i, X_{i+1}^R, \mathbb{1}(i \in A^{\text{opt}})),\ -\infty < i < \infty)$$

is also stationary. The next lemma shows this process is the limit of the corresponding finite-$n$ process. The mode of convergence can be viewed as a very elementary case of *local weak convergence* [4] of random graphical structures. In other words, it asserts that relative to a random time-origin the finite processes approximate the limit process.

LEMMA 8. *Let $U_n$ be uniform on $\{1, \dots, n\}$. As $n \to \infty$*

$$\begin{aligned}
&((X_{n,U_n+i}^L, \xi_{U_n+i}, X_{n,U_n+i+1}^R, \mathbb{1}(U_n + i \in A_n^{opt})),\ -\infty < i < \infty)\\
&\quad \stackrel{d}{\to}\ ((X_i^L, \xi_i, X_{i+1}^R, \mathbb{1}(i \in A^{opt})),\ -\infty < i < \infty),
\end{aligned}$$

*where the left-hand side is defined arbitrarily for $U_n + i \notin \{1, \dots, n\}$ and where convergence in distribution is with respect to the usual product topology on infinite sequence space.*

*Proof.* Because the $X$'s are bounded and the $\xi$'s are i.i.d., the sequence of processes is tight in the product topology. Write

$$((\hat{X}_i^L, \hat{\xi}_i, \hat{X}_{i+1}^R, \mathbb{1}(i \in \hat{A}^{\text{opt}})),\ -\infty < i < \infty)$$

for a subsequential weak limit. Clearly $(\hat{\xi}_i) \overset{d}{=} (\xi_i)$. Because for each $n$ the process $(X_{n,i}^L, \xi_i)$ satisfies recursion (21), the limit $(\hat{X}_i^L, \hat{\xi}_i)$ satisfies this recursion, and so by the "uniqueness of joint distribution" assertion of Lemma 6, $(\hat{X}_i^L, \hat{\xi}_i) \overset{d}{=} (X_i^L, \xi_i)$. Applying the same argument to $X^R$ we deduce

$$((X_{U_n+i}^L, \xi_{U_n+i}, X_{n,U_n+i+1}^R),\ -\infty < i < \infty) \overset{d}{\to} ((X_i^L, \xi_i, X_{i+1}^R),\ -\infty < i < \infty).$$

For fixed $i_0$ the event $i_0 \in A^{\text{opt}}$ is a function of the limit process, the function implied by Lemma 7, and by a standard fact [6, Theorem 5.2] it is enough to check that this function is a.s. continuous with respect to the limit process. But this just requires that the probability of an *equality* between some two of $X_{i_0}^L, \xi_{i_0}, X_{i_0+1}^R$ should be zero, which follows from their independence (30) and existence of densities (7), (24). ∎

**2.4. Proof of Theorem 1.** Because

$$M_n = \sum_{i=1}^n \mathbb{1}(i \in A_n^{\text{opt}}) - \sum_{i=1}^{n-1} \xi_i \mathbb{1}(i \in A_n^{\text{opt}}, i+1 \in A_n^{\text{opt}}),$$

we can write

$$n^{-1}\mathrm{E}M_n = \mathrm{P}(U_n \in A_n^{\text{opt}}) - \mathrm{E}\xi_{U_n}\mathbb{1}(U_n \in A_n^{\text{opt}}, U_n+1 \in A_n^{\text{opt}})\mathbb{1}(U_n \neq n)$$

and then by Lemma 8

$$n^{-1}\mathrm{E}M_n \to c := \mathrm{P}(0 \in A^{\text{opt}}) - \mathrm{E}\xi_0 \mathbb{1}(0 \in A^{\text{opt}}, 1 \in A^{\text{opt}}).$$

Note that clearly $c \leq 1$; the other inequality $c \geq 1/2$ holds because the subset $\{1, 3, 5, \ldots\}$ is a feasible choice.

We now exploit the *method of bounded differences* [12] in a very routine way. We observe that $M_n = m_n(\xi_1, \ldots, \xi_n)$ for a certain function $m_n$ with the property

changing any one argument of $m_n(z_1, \ldots, z_n)$ changes the value of $m_n(\cdot)$ by at most 1

This property holds because $A_n^{\text{opt}}$ will never contain a pair $(i, i+1)$ for which $\xi_i > 1$. And this property implies the well-known Azuma–Hoeffding inequality of the form (see, e.g., [16])

$$\mathrm{P}(|M_n - \mathrm{median}(M_n)| \geq t) \leq 4\exp(-\tfrac{t^2}{4n}).$$

It is now routine to use this large deviation inequality to establish the a.s. and $L^1$ convergence of $n^{-1}M_n$ to $c$.

To evaluate $c$, abbreviate $(X_0^L, \xi_0, X_1^R)$ to $(X^L, \xi, X^R)$ and use the Lemma 7 definition of $A^{\text{opt}}$ to write

$$\begin{aligned}\mathrm{P}(0 \in A^{\text{opt}}) &= 1 - \mathrm{P}(X^L < \min(X^R, \xi)) \\ &= 1 - \tfrac{1}{2}(1 - \mathrm{P}(\xi < \min(X^L, X^R))) \text{ by symmetry} \\ &= \tfrac{1}{2} + \tfrac{1}{2}\mathrm{P}(\xi < \min(X^L, X^R))\end{aligned}$$

and then

$$c = \tfrac{1}{2} + \tfrac{1}{2}\mathrm{P}(\xi < \min(X^L, X^R)) - \mathrm{E}\xi \mathbb{1}(\xi < \min(X^L, X^R)). \tag{31}$$

Recall that $X^L$, $\xi$, and $X^R$ are independent and that $X^L$ and $X^R$ have common distribution $F$ given in terms of $G$ by (24). So (31) constitutes a formula for $c$ in terms of the underlying distribution function $G$ of $\xi$.

We now evaluate $c$ in the special case where $\xi$ has the exponential($\lambda$) distribution:

$$\overline{G}(x) = e^{-\lambda x}, \quad 0 < x < \infty,$$

so that, from formula (24), we have

$$F(x) = \frac{e^{-\lambda(1-x)}(1-e^{-\lambda x})}{1-e^{-\lambda}} = \frac{e^{\lambda x}-1}{e^{\lambda}-1}.$$

We deduce

$$\begin{aligned}
\mathrm{P}(\xi < \min(X^L, X^R)) &= \int_0^1 \lambda e^{-\lambda u} P^2(X^L > u)\, du \\
&= \frac{\lambda}{(e^\lambda - 1)^2} \int_0^1 e^{-\lambda u} \left(e^\lambda - e^{\lambda u}\right)^2 \, du \\
&= \frac{\lambda}{(e^\lambda - 1)^2} \left( e^{2\lambda} \int_0^1 e^{-\lambda u} du - 2e^\lambda + \int_0^1 e^{\lambda u} du \right) \\
&= \frac{e^{2\lambda} - 2\lambda e^\lambda - 1}{(e^\lambda - 1)^2},
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\xi \mathbb{1}(\xi < \min(X^L, X^R)) &= \int_0^1 u \lambda e^{-\lambda u} P^2(X^L > u)\, du \\
&= \frac{\lambda}{(e^\lambda - 1)^2} \int_0^1 u e^{-\lambda u} \left(e^\lambda - e^{\lambda u}\right)^2 \, du \\
&= \frac{\lambda}{(e^\lambda - 1)^2} \left( e^{2\lambda} \int_0^1 u e^{-\lambda u} du - e^\lambda + \int_0^1 u e^{\lambda u} du \right) \\
&= \frac{1}{\lambda (e^\lambda - 1)^2} \left( e^{2\lambda} - (\lambda^2 + 2) e^\lambda + 1 \right).
\end{aligned}$$

Combining,

$$\begin{aligned}
c &= \frac{1}{2} + \frac{\frac{\lambda}{2}(e^{2\lambda} - 2\lambda e^\lambda - 1) - (e^{2\lambda} - (\lambda^2 + 2)e^\lambda + 1)}{\lambda (e^\lambda - 1)^2} \\
&= \frac{1}{1 - e^{-\lambda}} - \frac{1}{\lambda}.
\end{aligned}$$

**3. The upper bound in Theorem 2.** Local weak convergence (Lemma 8 above and Lemma 11 below) provides one sense in which the $n \to \infty$ limit of the solution $A_n^{\rm opt}$ of the size-$n$ optimization problem is $A^{\rm opt}$. A logically different sense is provided by coupling, as follows. Part of the stationary triple process is the doubly infinite i.i.d. sequence $(\ldots, \xi_{-1}, \xi_0, \xi_1, \xi_2, \ldots)$. For each $n$ use these same random variables $\xi_1, \ldots, \xi_n$ to construct $A_n^{\rm opt}$. Because of boundary effects it is not always true that $A^{\rm opt} \cap [1, n] = A_n^{\rm opt}$. But we expect the sets to coincide "away from the boundary," and Lemma 9(b) below provides one expression of this equality. We call this technique *localization.*

**3.1. Optimality properties of $A^{\rm opt}$.** Lemma 7 gave a concise definition of $A^{\rm opt}$ but did not explicitly identify its optimality properties. Lemma 9 below will relate $A^{\rm opt}$ to certain finite optima and thereby allow us to deduce some explicit properties.

The benefit function $f_n(A)$ and its maximum value $M_n$ defined at (5), (6) refer to subsets of $[1, n]$, and it is convenient to make the corresponding definitions for an arbitrary interval $[\ell, m]$:

$$f_{[\ell,m]}(A) := |A| - \sum_{i=\ell}^{m-1} \xi_i \mathbb{1}(i \in A, i+1 \in A), \quad A \subseteq \{\ell, \ell+1, \ldots, m\}, \tag{32}$$

$$M_{[\ell,m]} := \max_{A \subseteq \{\ell, \ell+1, \ldots, m\}} f_{[\ell,m]}(A), \tag{33}$$

and denote by $A_{[\ell,m]}^{\rm opt}$ the corresponding optimizing set.

LEMMA 9. (a) *If $\xi_{i-1} + \xi_i \leq 1$, then $i \in A^{opt}$.*

(b) *If $\ell < m$ and $\xi_{\ell-1} + \xi_\ell \leq 1$ and $\xi_{m-1} + \xi_m \leq 1$, then $A_{[\ell,m]}^{opt}$ is unique and*

$$A^{opt} \cap [\ell, m] = A_{[\ell,m]}^{opt}. \tag{34}$$

*If, furthermore, $[\ell, m] \subseteq [1, n]$, then $A_n^{opt} \cap [\ell, m] = A_{[\ell,m]}^{opt}$ (interpreting $\xi_0 = 0$ if $\ell = 1$).*

(c) *If both $i, i+1 \in A^{opt}$, then $\xi_i \leq 1$.*

(d) *If $\xi_i + \xi_{i+1} > 1$, then $i, i+1$ and $i+2$ together cannot belong to $A^{opt}$.*

(e) *Let $k \geq 2$. If $[g, g+2k-1]$ is an interval such that $\xi_g > \xi_{g+1} > \cdots > \xi_{g+2k-1} > \xi_{g+2k}$ and*

$$\xi_j + \xi_{j+1} > 1, \quad g \leq j \leq g+2k-2,$$

*then $A^{opt} \cap [g, g+2k-1]$ must be one of the two complementary alternating sequences in $[g, g+2k-1]$.*

*Proof.* (a) If $\xi_{i-1} + \xi_i \leq 1$, then $X_i^L \geq 1 - \xi_{i-1} \geq \xi_i$, and hence from the Lemma 7 definition we see that for any possible value of $X_{i+1}^R$, we have $i \in A^{\rm opt}$.

(b) First note that both $\ell$ and $m$ are in $A_{[\ell,m]}^{\rm opt}$; otherwise adding each element would increase $f_{[\ell,m]}(A_{[\ell,m]}^{\rm opt})$ by at least $1 - \xi_\ell$ and $1 - \xi_{m-1}$, respectively. Next note that by rewriting Lemma 5 (which concerns the special case $[\ell, m] = [1, n]$) for general $[\ell, m]$, we have a construction of $A_{[\ell,m]}^{\rm opt}$ in terms of processes $X_{[\ell,m],i}^L$ and $X_{[\ell,m],i}^R$ for $\ell \leq i \leq m$ defined by the recursions analogous to (21), (22). By (a) both $\ell$ and $m$ are in $A^{\rm opt}$. We have now shown $X_{[\ell,m],\ell}^L = 1 - \xi_{\ell-1} = X_\ell^L$ and $X_{[\ell,m],m-1}^R = 1 - \xi_{m-1} = X_{m-1}^R$; because the restricted and unrestricted processes have the same boundary conditions and satisfy the same recursions over $[l, m]$ they must agree throughout the interval. Finally, because both endpoints $\ell$ and $m$ are in $A_{[\ell,m]}^{\rm opt}$

it cannot fit the "complementary alternating sequences" criteria for nonuniqueness (Lemma 3). The same argument works for $A_n^{\mathrm{opt}}$.

One could prove (c), (d), (e) algebraically from the definition of $A^{\mathrm{opt}}$, but it is more intuitive to exploit the finite optimality criterion as follows. From assumption (8) there are infinitely many $\ell$ with $\xi_{\ell-1}+\xi_\ell \le 1$, and so for any given $i$ there is a (random) long interval $[\ell, m]$ containing $i$ for which by (b) $A^{\mathrm{opt}} \cap [\ell, m] = A^{\mathrm{opt}}_{[\ell,m]}$. In other words, the restriction of $A^{\mathrm{opt}}$ to $[\ell, m]$ is the solution of the finite optimization problem (32), and we can derive its properties by considering the effect of local changes. Now (c) and (d) follow from the following observations:

(for (c)): if $i, i+1$ are in $A^{\mathrm{opt}}$, then removing $i+1$ will give a relative benefit of at least $\xi_i - 1$.

(for (d)): if $i, i+1, i+2$ are all in $A^{\mathrm{opt}}$, then removing $i+1$ will give a relative benefit of $\xi_i + \xi_{i+1} - 1$.

For (e), consider $j \in [g, g+2k]$. By (d) we cannot have $\{j, j+1, j+2\} \subset A^{\mathrm{opt}}$. If $j$ and $j+1$ but not $j+2$ are in $A^{\mathrm{opt}}$, then deleting $j+1$ while adding $j+2$ would increase the benefit by at least $\xi_j - \xi_{j+2} > 0$, which is impossible. It follows that we cannot have $\{j, j+1\} \subset A^{\mathrm{opt}}$. Thus $A^{\mathrm{opt}} \cap [g, g+2k-1]$ contains only isolated elements. It is now easy to check that one can change $A^{\mathrm{opt}} \cap [g, g+2k-1]$ into one of the alternating sequences on $[g, g+2k-1]$ in such a way that the cardinality does not decrease, and the end items $g, g+2k-1$ change (if at all) only from included to excluded. Thus the change can only increase the benefit; appealing to the uniqueness property (b) in a larger interval establishes (e). ∎

**3.2. Proof of upper bound.** In this section we prove the bound

$$\limsup_{\delta \downarrow 0} \delta^{-2} \limsup_n \mathrm{E}\varepsilon_n(\delta) < \infty \tag{35}$$

via a simple construction of near-optimal sets. We first describe a particular configuration. Let $g, d \in \mathbb{Z}$ such that $d - g = 2k$ for some $k \ge 2$, and consider the sets $A$ and $B$ below:

| | $g$ | – | $(g+1)$ | – | | | | | | | $(g+2k-1)$ | – | $g+2k$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | • | – | ○ | – | • | – | ○ | … | • | – | ○ | – | • |
| $B$ | • | – | • | – | ○ | – | • | … | ○ | – | • | – | • |

where $|A \triangle B| = 2k-1$ and the difference between the benefits of $A$ and $B$ is

$$\begin{aligned} f_{[g,g+2k]}(A) - f_{[g,g+2k]}(B) &= (k+1) - ((k+2) - \xi_g - \xi_{g+2k-1}) \\ &= \xi_g + \xi_{g+2k-1} - 1. \end{aligned} \tag{36}$$

Now fix $k \ge 2$ and $\alpha > 0$ such that $\alpha k < 1/2$. Consider the event $\Omega_g$ defined by

$$\xi_g > \xi_{g+1} > \xi_{g+2} > \cdots > \xi_{g+2k-2} > \xi_{g+2k-1} > \tfrac{1}{2} > \xi_{g+2k};$$
$$\xi_{g-1} + \xi_g < 1; \quad \xi_{g+2k-1} + \xi_{g+2k} < 1.$$

By assumption (7) this event has nonzero probability. If this event occurs, Lemma 9(a) shows that $A^{\mathrm{opt}}$ contains $g$ and $g+2k$, and then Lemma 9(e) implies that $A^{\mathrm{opt}} \cap [g, g+2k]$ is the set $A$ above. By applying Lemma 9(b) to $[\ell, m] = [g, g+2k]$ we have the analogue in the finite $n$ setting: if $\Omega_g$ occurs for $[g, g+2k] \subset [1, n]$, then

$A_n^{\text{opt}} \cap [g, g+2k]$ is the set $A$ above. So if we change $A_n^{\text{opt}}$ by replacing pattern $A$ by pattern $B$ on such an interval, then from (36) the decrease in benefit equals $\xi_g + \xi_{g+2k-1} - 1 > 0$. Now define

$$\Omega_g^{(\alpha)} = \Omega_g \cap \{1 < \xi_g + \xi_{g+2k-1} < 1 + 2k\alpha\},$$

$$q(\alpha) = \mathrm{P}(\Omega_g^{(\alpha)}),$$

$$r(\alpha) = \mathrm{E}(\xi_g + \xi_{g+2k-1} - 1)\mathbb{1}(\Omega_g^{(\alpha)})$$

so that $r(\alpha)$ is the unconditional mean increase in benefit from the possible change, now performed only if event $\Omega_g^{(\alpha)}$ happens. Using assumption (7) we see that $(\xi_g + \xi_{g+2k-1})$ restricted to $\Omega_g^{(\alpha)}$ has a continuous density which is nonzero at 1, which easily implies that for fixed $k$

$$q(\alpha) \sim \bar{q}\alpha, \quad r(\alpha) \sim \bar{r}\alpha^2 \text{ as } \alpha \downarrow 0 \tag{37}$$

for constants $\bar{q}, \bar{r} \in (0, \infty)$.

Given $n$ and the optimal set $A_n^{\text{opt}}$, construct a near-optimal set $B_n^{(\alpha)}$ as follows. Let $g_1 = 1$ and let

$$[g_1, g_1 + 2k], \quad [g_2, g_2 + 2k], \quad [g_3, g_3 + 2k], \dots, [g_{j_n}, g_{j_n} + 2k]$$

be the adjacent disjoint intervals in $[1, n]$ containing $2k + 1$ integers. For each such $g = g_j$, if event $\Omega_g^{(\alpha)}$ occurs, then on $[g, g + 2k]$ replace pattern $A$ by pattern $B$.

Letting $n \to \infty$ and using the weak law of large numbers, we get

$$\tfrac{1}{n}|B_n^{(\alpha)} \triangle A_n^{\text{opt}}| \to 2kq(\alpha)/(2k+1) \text{ in probability,}$$

$$\tfrac{1}{n}(f_n(A_n^{\text{opt}}) - f_n(B_n^{(\alpha)})) \to r(\alpha)/(2k+1) \text{ in probability.}$$

If $\frac{1}{n}|B_n^{(\alpha)} \triangle A_n^{\text{opt}}| \le kq(\alpha)/(2k+1)$, then redefine $B_n^{(\alpha)}$ to be the empty set. Then (taking $k = 3$ for concreteness)

$$\tfrac{1}{n}|B_n^{(\alpha)} \triangle A_n^{\text{opt}}| \ge 3q(\alpha)/7,$$

$$\lim_n \tfrac{1}{n}(\mathrm{E}f_n(A_n^{\text{opt}}) - \mathrm{E}f_n(B_n^{(\alpha)})) = r(\alpha)/7.$$

The upper bound (35) now follows from the $\alpha \to 0$ asymptotics (37).

## 4. Proof of Theorem 2: The lower bound.

### 4.1. Analysis of near-optimal solutions: The quintuple process.

Throughout section 4 we fix a constant $\tau > 0$ such that

$$G\left(\tfrac{1}{2} - \tau\right) > 0. \tag{38}$$

Such a constant exists by assumption (7). To study near-optimal solutions, fix a Lagrange multiplier $\theta$ such that

$$0 < \theta < \tau. \tag{39}$$

We will derive the existence of, and derive an exact expression for, the function $\bar{\varepsilon}(\delta) = \lim_n \mathrm{E}\varepsilon_n(\delta)$ when $\delta$ is sufficiently small. The expression is an implicit function representation $\bar{\varepsilon}(\delta(\theta)) = \varepsilon(\theta)$ via two functions $\varepsilon(\theta), \delta(\theta)$ defined (49), (50) in terms of the stationary distribution of a certain *quintuple process*.

We study the modified optimization problem in which we get an extra reward $\theta$ for choosing an item which is not in $A_n^{\mathrm{opt}}$ or for not choosing an item which is in $A_n^{\mathrm{opt}}$:

$$\max_{A\subseteq[n]} \left( |A| - \sum_{i=1}^{n} \xi_i \mathbb{1}(i \in A, i+1 \in A) + \theta |A \triangle A_n^{\mathrm{opt}}| \right). \tag{40}$$

To study this we modify (13), (14) to

$$\widetilde{V}_{n,i}^L = \max_{i\in A\subseteq\{1,2,\dots,i\}} \left( |A| - \sum_{j=1}^{i-1} \xi_j \mathbb{1}(j, j+1 \in A) + \theta \left| (A \triangle A_n^{\mathrm{opt}}) \cap \{1,2,\dots,i\} \right| \right), \tag{41}$$

$$\widetilde{W}_{n,i}^L = \max_{i\notin A\subseteq\{1,2,\dots,i\}} \left( |A| - \sum_{j=1}^{i-1} \xi_j \mathbb{1}(j, j+1 \in A) + \theta \left| (A \triangle A_n^{\mathrm{opt}}) \cap \{1,2,\dots,i\} \right| \right). \tag{42}$$

We also define $\widetilde{M}_n = \max(\widetilde{V}_{n,n}^L, \widetilde{W}_{n,n}^L)$ and write $B_n^{\mathrm{opt}}$ for the corresponding optimizing set. Note that these quantities depend on $\theta$. Analogous to the definition (15) of $X_{n,i}^L$ we define

$$Z_{n,i}^L := \widetilde{V}_{n,i}^L - \widetilde{W}_{n,i}^L.$$

Then as the analogue of (16) we can obtain the recursion

$$Z_{n,i+1}^L = 1 - \min(Z_{n,i}^L, \xi_i)\mathbb{1}(Z_{n,i}^L > 0) + \theta J_{n,i+1},$$

where

$$Z_{n,1}^L = 1 + \theta J_{n,1},$$
$$J_{n,i} = \mathbb{1}(i \notin A_n^{\mathrm{opt}}) - \mathbb{1}(i \in A_n^{\mathrm{opt}}).$$

Recall from section 2.3 the stationary triple process $((X_i^L, \xi_i, X_{i+1}^R), -\infty < i < \infty)$ and define

$$J_i = \mathbb{1}(i \notin A^{\mathrm{opt}}) - \mathbb{1}(i \in A^{\mathrm{opt}}).$$

Just as the stationary triple process is interpretable (Lemma 8) as an $n \to \infty$ limit of the process $(X_{n,i}^L, \xi_i, X_{n,i+1}^R)$, we want to define a process which will be the limit of $(Z_{n,i}^L, X_{n,i}^L, \xi_i, X_{n,i+1}^R)$. So define a *quadruple process* $(Z_i^L, X_i^L, \xi_i, X_{i+1}^R)$ to be a process such that

(i) $(X_i^L, \xi_i, X_{i+1}^R)$ evolves as the triple process,

(ii) $Z_i^L$ satisfies the recursion

$$Z_{i+1}^L = 1 - \min(Z_i^L, \xi_i)\mathbb{1}(Z_i^L > 0) + \theta J_{i+1}. \tag{43}$$

Recall $0 < \theta < \tau$.

LEMMA 10. *The quadruple process* $((Z_i^L, X_i^L, \xi_i, X_{i+1}^R), -\infty < i < \infty)$ *has a unique stationary distribution, for which*

$$Z_i^L = \psi(\dots, \xi_{i-2}, \xi_{i-1}, \xi_i, X_{i+1}^R) \tag{44}$$

*for a certain function* $\psi$ *not depending on* $i$. *On the event* $\{\xi_{i-1} + \xi_i \le 1 - \tau\}$, *we have*

$$X_{i+1}^L = 1 - \xi_i, Z_{i+1}^L = 1 - \xi_i + \theta J_{i+1}. \tag{45}$$

*Proof.* Recursion (43) implies $Z_{i+1}^L \ge 1 - \xi_i + \theta J_{i+1}$. Thus iterating once (43) and using this last inequality, we obtain

$$1 - \xi_i + \theta J_{i+1} \le Z_{i+1}^L \le 1 - \min(1 - \xi_{i-1} + \theta J_i, \xi_i) \mathbb{1}(1 - \xi_{i-1} + \theta J_i > 0) + \theta J_{i+1}.$$

Thus, on the event $\{\xi_{i-1} + \xi_i \le 1 - \theta\}$ we have $Z_{i+1}^L = 1 - \xi_i + \theta J_{i+1}$ and also, by (27), we have $X_{i+1}^L = 1 - \xi_i$, establishing (45). Assumption (7) implies that the event $\{\xi_{i-1} + \xi_i \le 1 - \tau\}$ occurs for infinitely many $i < 0$, so in particular $K := \max\{i < 0 : \xi_{i-1} + \xi_i \le 1 - \tau\}$ is finite. By the recursion (43) we can write $Z_0^L$ in the form

$$Z_0^L = \psi^1(\xi_{K+1}, \xi_{K+2}, \dots, \xi_{-1}; Z_{K+1}^L; J_{K+2}, J_{K+3}, \dots, J_0)$$

for some function $\psi^1$. Then by (45) with $Z_i^L = Z_{K+1}^L$ we can rewrite as

$$Z_0^L = \psi^2(\xi_K, \xi_{K+1}, \xi_{K+2}, \dots, \xi_{-1}; J_{K+1}, J_{K+2}, J_{K+3}, \dots, J_0).$$

By the definition of $A^{\text{opt}}$, each $J_i$ is a function of $X_i^L, \xi_i, X_{i+1}^R$, and then from the recursions for $X_i^L$ and $X_i^R$

$$Z_0^L = \psi^3(\xi_K, \xi_{K+1}, \xi_{K+2}, \dots, \xi_0; X_{K+1}^L, X_1^R).$$

By (45) with $X_i^L = X_{K+1}^L$ this is of the form

$$Z_0^L = \psi(\dots, \xi_{-2}, \xi_{-1}, \xi_0, X_1^R).$$

Now (44) defines a stationary version of the quadruple process. ∎

Just as $X_{n,i}^R$ was the "looking right" analogue of the "looking left" process $X_{n,i}^L$, we can define a "looking right" process $Z_{n,i}^R$ analogous to $Z_{n,i}^L$ as follows. Define

$$\widetilde{V}_{n,i}^R = \max_{i \in A \subseteq \{i, i+1, \dots, n\}} \left( |A| - \sum_{j=i}^{n-1} \xi_j \mathbb{1}(j, j+1 \in A) + \theta \left| (A \triangle A^{\text{opt}}) \cap \{i, i+1, \dots, n\} \right| \right), \tag{46}$$

$$\widetilde{W}_{n,i}^R = \max_{i \notin A \subseteq \{i, i+1, \dots, n\}} \left( |A| - \sum_{j=i}^{n-1} \xi_j \mathbb{1}(j, j+1 \in A) + \theta \left| (A \triangle A^{\text{opt}}) \cap \{i, i+1, \dots, n\} \right| \right). \tag{47}$$

Then the difference $Z_{n,i}^R = \widetilde{V}_{n,i}^R - \widetilde{W}_{n,i}^R$ satisfies the recursion

$$Z_{n,i}^R = 1 - \min(Z_{n,i+1}^R, \xi_i) \mathbb{1}(Z_{n,i+1}^R > 0) + \theta J_{n,i}; \quad Z_{n,n}^R = 1 + \theta J_{n,n}.$$

TABLE 3
*Inclusion criteria for $i, i+1$ in $B_n^{opt}$.*

| $-i-(i+1)-$ | Absolute benefit | Relative benefit | When used |
|---|---|---|---|
| $-\bullet--\bullet-$ | $\widetilde{V}^L + \widetilde{V}^R - \xi$ | $Z^L + Z^R - \xi$ | $\xi < \min(Z^L - \theta J_i,$ |
| | $+\theta(\mathbb{1}_{i \notin A^{\text{opt}}} + \mathbb{1}_{i+1 \notin A^{\text{opt}}})$ | $-\theta(J_i + J_{i+1})$ | $Z^R - \theta J_{i+1})$ |
| $-\bullet--\circ-$ | $\widetilde{V}^L + \widetilde{W}^R$ | $Z^L - \theta J_i$ | $(Z^R - \theta J_{i+1})^+$ |
| | $+\theta(\mathbb{1}_{i \notin A^{\text{opt}}} + \mathbb{1}_{i+1 \in A^{\text{opt}}})$ | | $< \min(Z^L - \theta J_i, \xi)$ |
| $-\circ--\bullet-$ | $\widetilde{W}^L + \widetilde{V}^R$ | $Z^R - \theta J_{i+1}$ | $(Z^L - \theta J_i)^+$ |
| | $+\theta(\mathbb{1}_{i \in A^{\text{opt}}} + \mathbb{1}_{i+1 \notin A^{\text{opt}}})$ | | $< \min(Z^R - \theta J_{i+1}, \xi)$ |
| $-\circ--\circ-$ | $\widetilde{W}^L + \widetilde{W}^R$ | $0$ | otherwise |
| | $+\theta(\mathbb{1}_{i \in A^{\text{opt}}} + \mathbb{1}_{i+1 \in A^{\text{opt}}})$ | | |

Recall that $B_n^{\text{opt}}$ attains $\max_{A \subseteq \{1,\dots,n\}}(|A| - \sum_{i=1}^{n-1} \xi_i \mathbb{1}(i \in A, i+1 \in A) + \theta |A \triangle A_n^{\text{opt}}|)$. As in section 2.2, we can write down the benefits of each of the four possible choices for including or excluding items $i$ and $i+1$, and thereby obtain criteria for which combination is used in $B_n^{\text{opt}}$. See Table 3, in which $(Z_{n,i}^L, \xi_i, Z_{n,i+1}^R)$ is abbreviated to $(Z^L, \xi, Z^R)$ and the $n$ subscript is dropped. It should now be clear that the stationary quadruple process can be extended to a *stationary quintuple process*

$$(Z_i^L, X_i^L, \xi_i, X_{i+1}^R, Z_{i+1}^R), \quad -\infty < i < \infty,$$

in which $Z^R$ satisfies the recursion

$$Z_i^R = 1 - \min(Z_{i+1}^R, \xi_i)\mathbb{1}(Z_{i+1}^R > 0) + \theta J_i, \quad -\infty < i < \infty,$$

satisfied by $Z_{n,i}^R$. By "reflection symmetry" between $Z^R$ and $Z^L$, the functional relationship (44) holds for $Z^R$ in reflected form with the same function $\psi$:

$$Z_i^R = \psi(\dots, \xi_{i+1}, \xi_i, \xi_{i-1}, X_{i-1}^L). \tag{48}$$

We can now use the stationary quintuple process to define a random subset $B^{\text{opt}} \subset \mathbb{Z}$ by specifying that, for each pair $(i, i+1)$, we use the one of the four choices which has the largest relative benefit in Table 3. Analogously to Lemma 7 one can check that this definition is consistent. The local weak convergence property (Lemma 8) extends to the present setting as follows.

LEMMA 11. *Let $U_n$ be uniform on $\{1,\dots,n\}$. As $n \to \infty$*

$$((Z_{n,U_n+i}^L, X_{n,U_n+i}^L, \xi_{U_n+i}, X_{n,U_n+i+1}^R, Z_{n,U_n+i+1}^R,$$
$$\mathbb{1}(U_n + i \in A_n^{opt}), \mathbb{1}(U_n + i \in B_n^{opt})))_{-\infty<i<\infty}$$
$$\xrightarrow{d} ((Z_i^L, X_i^L, \xi_i, X_{i+1}^R, Z_{i+1}^R, \mathbb{1}(i \in A^{opt}), \mathbb{1}(i \in B^{opt})))_{-\infty<i<\infty}.$$

*Proof.* The proof repeats the proof of Lemma 8, using (44), (48) to incorporate the $(Z^L, Z^R)$ terms. In order to incorporate the $B^{\text{opt}}$ component, we need to check that the function $\mathbb{1}(0 \in B^{\text{opt}})$ is a.s. continuous with respect to the stationary distribution of $(Z_0^L, X_0^L, \xi_0, X_1^R, Z_1^R)$. From Table 3, we get that $\{0 \in B^{\text{opt}}\} = \{Z_0^L - \theta J_0 > \min(\xi_0, \max(Z_1^R - \theta J_1, 0)\}$. Hence, it requires that the probability of an equality between some of two $Z_0^L - \theta J_0, \xi_0, Z_1^R - \theta J_1$ is zero. We check only that $\mathrm{P}(Z_0^L - \theta J_0 = \xi_0) = 0$. The recursion satisfied by $Z_0^L$ reads $Z_0^L - \theta J_0 =$

$1 - \min(Z^L_{-1}, \xi_{-1})\mathbb{1}(Z^L_{-1} > 0)$. Thus, arguing as in the proof of Lemma 10, $Z^L_0 - \theta J_0$ is a function of $(Z^L_{K+1}, \xi_{K+1}, \ldots, \xi_{-1}, J_{K+1}, J_{K+2}, \ldots, J_{-1})$ with $K = \max\{i < 0 : \xi_{i-1} + \xi_i \le 1 - \tau\}$. Since $Z^L_{K+1} = 1 - \xi_K + \theta J_{K+1}$ and $J_i \in \{-1, 1\}$, we deduce by recursion that there exists a pair of integers $(i_0, n)$ with $K \le i_0 \le -1$ and $-K \le n \le K$ such that $Z^L_0 \in \{1 - \xi_{i_0} + n\theta, \xi_{i_0} + n\theta\}$. The independence of $\xi_i$ and $\xi_0$ for $i < 0$ and assumption (7) imply that $\mathrm{P}(Z^L_0 - \theta J_0 = \xi_0) = 0$. $\square$

Now define

$$\delta(\theta) = \mathrm{P}(\{0 \in A^{\mathrm{opt}}\} \triangle \{0 \in B^{\mathrm{opt}}\}), \tag{49}$$

$$\begin{aligned}\varepsilon(\theta) = {} & \mathrm{P}(0 \in A^{\mathrm{opt}}) - \mathrm{E}\xi_0 \mathbb{1}(0 \in A^{\mathrm{opt}}, 1 \in A^{\mathrm{opt}}) - \mathrm{P}(0 \in B^{\mathrm{opt}}) \\ & + \mathrm{E}\xi_0 \mathbb{1}(0 \in B^{\mathrm{opt}}, 1 \in B^{\mathrm{opt}}).\end{aligned} \tag{50}$$

So $\delta(\theta)$ is the proportion of items at which $A^{\mathrm{opt}}$ and $B^{\mathrm{opt}}$ differ, and $\varepsilon(\theta)$ is the difference in mean benefit per item between $A^{\mathrm{opt}}$ and $B^{\mathrm{opt}}$. By Lemma 11,

$$\begin{aligned}\frac{1}{n}\mathrm{E}|A^{\mathrm{opt}}_n \triangle B^{\mathrm{opt}}_n| & = \mathrm{E}|\mathbb{1}(U_n \in A^{\mathrm{opt}}_n) - \mathbb{1}(U_n \in B^{\mathrm{opt}}_n)| \\ & \to \mathrm{P}(\{0 \in A^{\mathrm{opt}}\} \triangle \{0 \in B^{\mathrm{opt}}\}) = \delta(\theta),\end{aligned} \tag{51}$$

and similarly the mean benefits satisfy

$$n^{-1}(\mathrm{E}f_n(A^{\mathrm{opt}}_n)) - \mathrm{E}f_n(B^{\mathrm{opt}}_n)) \to \varepsilon(\theta). \tag{52}$$

PROPOSITION 12. *Let $\widetilde{M}_n = f_n(B^{opt}_n)$ be the benefit associated with $B^{opt}_n$; then a.s. and in $L^1$*

$$\lim_{n\to\infty} n^{-1}|B^{opt}_n \triangle A^{opt}_n| = \delta(\theta), \tag{53}$$

$$\lim_{n\to\infty} n^{-1}(M_n - \widetilde{M}_n) = \varepsilon(\theta). \tag{54}$$

*Moreover for any choice $B'_n$ satisfying (53) in $L^1$, the associated benefit $M'_n = f_n(B'_n)$ satisfies*

$$\liminf_n n^{-1}\mathrm{E}(M_n - M'_n) \ge \varepsilon(\theta).$$

*Proof.* The convergence assertions (53), (54) follow from (51), (52) and the same concentration argument used in the proof of Theorem 1; we will not repeat the details. By construction, for any $B'_n$ the associated reward $M'_n$ satisfies

$$M'_n + \theta|B'_n \triangle A^{\mathrm{opt}}_n| \le \widetilde{M}_n + \theta|B^{\mathrm{opt}}_n \triangle A^{\mathrm{opt}}_n|.$$

Then because both $(B^{\mathrm{opt}}_n)$ and $(B'_n)$ satisfy (53), we see that

$$\mathrm{E}M'_n \le \mathrm{E}\widetilde{M}_n + o(n). \qquad \square$$

*Discussion.* For $0 < \theta < \tau$ and for $\delta = \delta(\theta)$, Proposition 12 implies that the limit $\bar{\varepsilon}(\delta) = \lim_n \mathrm{E}\varepsilon_n(\delta)$ exists and that

$$\bar{\varepsilon}(\delta(\theta)) = \varepsilon(\theta).$$

So to prove Theorem 2 it should be enough to prove

$$\delta(\theta) \sim \alpha\theta, \quad \varepsilon(\theta) \sim \beta\theta^2 \text{ as } \theta \to 0 \tag{55}$$

for positive constants $\alpha, \beta$. Now the definitions (49), (50) enable us to rewrite (using Table 3) $\delta(\theta)$ and $\varepsilon(\theta)$ in terms of the stationary distribution $(Z_0^L, X_0^L, \xi_0, X_1^R, Z_1^R)$ of the quintuple process, as

$$\delta(\theta) = \mathrm{P}\left(\{X_0^L > \min(X_1^R, \xi_0)\} \bigtriangleup \{Z_0^L - \theta J_0 > \min((Z_1^R - \theta J_1)^+, \xi_0)\}\right), \tag{56}$$

$$\begin{aligned}\varepsilon(\theta) &= \mathrm{P}(X_0^L > \min(X_1^R, \xi_0)) - \mathrm{P}(Z_0^L - \theta J_0 > \min((Z_1^R - \theta J_1)^+, \xi_0)) \\ &\quad - \mathrm{E}\xi_0\left(\mathbb{1}(\xi_0 < \min(X_0^L, X_1^R)) - \mathbb{1}(\xi_0 < \min(Z_0^L - \theta J_0, Z_1^R - \theta J_1))\right).\end{aligned}$$

So if we had an explicit formula for the stationary distribution $(Z_0^L, X_0^L, \xi_0, X_1^R, Z_1^R)$, then we could derive an explicit formula for $\delta(\theta)$ and $\varepsilon(\theta)$ and seek to prove (55) by calculus. But we do not have such an explicit formula—note the independence property (30) of the triple process does not hold for the quintuple process—and we have not completely succeeded in that program. We could prove the $\delta(\theta) \sim \alpha\theta$ part of (55), though we use only the weaker upper bound, proved by a simpler argument in section 4.2. To handle $\varepsilon(\theta)$ we show how to rewrite $\delta(\theta)$ and $\varepsilon(\theta)$ in a different way (Proposition 18) that allows us to derive inequalities, which will establish the stated form of Theorem 2.

**4.2. Existence of the limit function $\bar{\varepsilon}(\delta)$.** There is a minor technical point we deal with first. We expect intuitively that the function $\delta(\theta)$ should be continuous monotone, but neither property is obvious. If there were small values of $\delta$ which were not of the form $\delta = \delta(\theta)$ for some $\theta$, then we cannot use Proposition 12 to establish existence of a limit $\bar{\varepsilon}(\delta)$. Instead we outline an argument (reusing previous methods) to prove more abstractly (Lemma 13) that the limit $\bar{\varepsilon}(\delta)$ always exists. We could have started the proof of Theorem 2 this way, but we wanted to emphasize the Lagrange multiplier approach as more useful for calculation.

LEMMA 13. $\bar{\varepsilon}(\delta) := \lim_n \mathrm{E}\varepsilon_n(\delta)$ *exists for each* $0 < \delta < 1$.

Note that $\varepsilon_n(\delta)$ is a priori nondecreasing in $\delta$, and hence $\bar{\varepsilon}(\cdot)$ is nondecreasing.

*Outline proof.* Fix $0 < \delta < 1$. Let $B_n^{(\delta)}$ attain the minimum in the definition (9) of $\varepsilon_n(\delta)$. Set $\bar{\varepsilon}_*(\delta) = \liminf_n \mathrm{E}\varepsilon_n(\delta)$. There exists a subsequence (of the subsequence of $n$ attaining the liminf) in which the local weak convergence (Lemma 8) of $A_n^{\mathrm{opt}}$ to $A^{\mathrm{opt}}$ extends to joint convergence of $B_n^{(\delta)}$ to some limit random set $B^{(\delta)}$. The analogues of (49), (50) with $B^{(\delta)}$ in place of $B^{\mathrm{opt}}$ equal $\delta$ and $\bar{\varepsilon}_*(\delta)$. For arbitrary $n$, start with the restriction ($B_n^*$, say) of $B^{(\delta)}$ to $[1, n]$ and then show that by modifying $B_n^*$ near the endpoints we can construct $B_n^{**}$ satisfying $|B_n^{**} \bigtriangleup A_n^{\mathrm{opt}}| \geq \delta n$ and $\mathrm{E}[n^{-1}(f_n(A_n^{\mathrm{opt}}) - f_n(B_n^{**}))] \to \bar{\varepsilon}_*(\delta)$. ∎

The following lemma (to be proved in section 4.4) allows us to complete the proof of Theorem 2.

LEMMA 14. *There exist positive constants* $C_1, C_2$ *such that, for all* $0 < \theta < \tau$,

$$\delta(\theta) \leq C_1\theta, \tag{57}$$

$$\varepsilon(\theta) \geq C_2\theta^2. \tag{58}$$

We now finish the proof of Theorem 2. Recall that Proposition 12 showed $\bar{\varepsilon}(\delta(\theta)) = \varepsilon(\theta)$, and that (Lemma 13) $\bar{\varepsilon}(\cdot)$ is a nondecreasing function. Using (57)

$$\bar{\varepsilon}(C_1\theta) \geq \bar{\varepsilon}(\delta(\theta)) = \varepsilon(\theta) \geq C_2\theta^2,$$

and setting $\delta = C_1\theta$ gives $\bar{\varepsilon}(\delta) \geq C_2\delta^2/C_1^2$. This establishes the lower bound (11) and completes the proof of Theorem 2.

**4.3. A cycle formula representation.**

LEMMA 15. *If $\xi_{i-1} + \xi_i < 1 - \tau$, then $i \in A^{opt}$ and $i \in B^{opt}$.*

*Proof.* Suppose $\xi_{i-1} + \xi_i < 1 - \tau$. Lemma 9(a) showed $i \in A^{\text{opt}}$. Recall that $B_n^{\text{opt}}$ maximizes (40). If $i \notin A$, then the increase in the benefit at (40) obtained by including $i$ is at least $1 - \xi_i - \xi_{i-1} - \theta$, so by our standing assumption (39) the increase is positive, and so $i \in B_n^{\text{opt}}$. Letting $n \to \infty$ and using Lemma 11 gives the same conclusion for $B^{\text{opt}}$. ∎

We next need a lemma (analogous to Lemma 9(b)) giving conditions under which we can "localize" $A^{\text{opt}}$ and $B^{\text{opt}}$ by forcing them to coincide with the optimal sets $A_n^{\text{opt}}$ and $B_n^{\text{opt}}$ for the optimization problem on $[1, n]$ for suitable $n$, which we now write as $t - 1$.

LEMMA 16. *Let $t \geq 2$. Suppose $\xi_{i-1} + \xi_i < 1 - \tau$ for each of $i = 0, 1, t - 1, t$. Then the following hold:*

(a) *$A^{opt}$ and $B^{opt}$ contain $\{0, 1, t - 1, t\}$.*

(b) *The restrictions of $A^{opt}$ and $B^{opt}$ to $[1, t - 1]$ coincide with $A_{t-1}^{opt}$ and $B_{t-1}^{opt}$.*

(c) *For any $B \subseteq \{1, 2, \ldots, t - 1\}$, either $B = A_{t-1}^{opt}$ or $f_{t-1}(B) < f_{t-1}(A_{t-1}^{opt})$.*

(d) *In particular, either $A_{t-1}^{opt} = B_{t-1}^{opt}$ or $f_{t-1}(A_{t-1}^{opt}) > f_{t-1}(B_{t-1}^{opt})$.*

*Proof.* (a) follows from Lemma 15. Observe that $A_{t-1}^{\text{opt}}$ and $B_{t-1}^{\text{opt}}$ contain 1 and $t-1$, because $\xi_1 < 1 - \tau$ and $\xi_{t-2} < 1 - \tau$. If we consider the solutions $A_{[\ell,m]}^{\text{opt}}$, $B_{[\ell,m]}^{\text{opt}}$ for some interval $[\ell, m]$ strictly containing $[0, t]$, then they contain 1 and $t - 1$ by the argument for Lemma 15. Thus by optimality the restrictions of $A_{[\ell,m]}^{\text{opt}}$ and $B_{[\ell,m]}^{\text{opt}}$ to $[1, t - 1]$ must coincide with $A_{t-1}^{\text{opt}}$ and $B_{t-1}^{\text{opt}}$. So (b) follows from weak convergence, Lemma 11. And (c) follows from the uniqueness result, Lemma 3. ∎

We start by quoting a standard form (cf. [8, Exercise 6.3.4]) of Kac's identity for stationary processes.

LEMMA 17. *Let $(\Xi_i, -\infty < i < \infty)$ be a stationary ergodic sequence on some state space, let $\mathrm{P}(\Xi_1 \in \bar{D}) > 0$, and let $h(\Xi_1)$ be real-valued and integrable. For any $t_0 \geq 1$, define $T = t_0 \min\{i \geq 2 : \Xi_{it_0} \in \bar{D})$. Then*

$$\mathrm{E}h(\Xi_1) = \mathrm{E}\left[\mathbb{1}(\Xi_1 \in \bar{D}) \sum_{i=1}^{T-1} h(\Xi_i)\right].$$

We apply this to $\Xi_i = (Z_i^L, X_i^L, \xi_i, \xi_{i-1}, \xi_{i-2}, X_{i+1}^R, Z_{i+1}^R)$, $t_0 = 3$, and

$$D := \{\xi_{-1} + \xi_0 < 1 - \tau,\ \xi_0 + \xi_1 < 1 - \tau\} = \{\Xi_1 \in \bar{D}\} \tag{59}$$

for suitable $\bar{D}$, making the $T$ in Lemma 17 be

$$T = 3 \min\{t \geq 2 : \xi_{3t-2} + \xi_{3t-1} < 1 - \tau,\ \xi_{3t-1} + \xi_{3t} < 1 - \tau\}. \tag{60}$$

Now definition (49) says $\delta(\theta) = \mathrm{E}h(\Xi_0)$ for

$$h(\Xi_0) = \mathbb{1}(\{0 \in A^{\text{opt}}\} \triangle \{0 \in B^{\text{opt}}\}).$$

So $\sum_{i=1}^{T-1} h(\Xi_i)$ equals the cardinality of $A^{\text{opt}} \triangle B^{\text{opt}}$ restricted to $[1, T-1]$. On the event $D$, Lemma 16 identifies this restriction as $A^{\text{opt}}_{T-1} \triangle B^{\text{opt}}_{T-1}$, so Kac's identity gives (61) below. Similarly, definition (50) says $\varepsilon(\theta) = \mathrm{E}h(\Xi_0)$ for

$$h(\Xi_0) = \mathbb{1}(0 \in A^{\text{opt}}) - \xi_0 \mathbb{1}(0 \in A^{\text{opt}}, 1 \in A^{\text{opt}}) - \mathbb{1}(0 \in B^{\text{opt}}) + \xi_0 \mathbb{1}(0 \in B^{\text{opt}}, 1 \in B^{\text{opt}}),$$

and on the event $D$ the sum $\sum_{i=1}^{T-1} h(\Xi_i)$ equals the difference $f_{T-1}(A^{\text{opt}}_{T-1}) - f_{T-1}(B^{\text{opt}}_{T-1})$ between the benefits. This establishes (62), and the final assertion (63) follows from Lemma 16(d). To summarize:

PROPOSITION 18. *Let $D$ be the event* (59) *and let $T$ be the random time* (60). *Then*

$$\delta(\theta) = \mathrm{E}[\mathbb{1}_D \times |A^{opt}_{T-1} \triangle B^{opt}_{T-1}|], \tag{61}$$

$$\varepsilon(\theta) = \mathrm{E}[\mathbb{1}_D \times (f_{T-1}(A^{opt}_{T-1}) - f_{T-1}(B^{opt}_{T-1}))], \tag{62}$$

$$\textit{on } D, \textit{ either } A^{opt}_{T-1} = B^{opt}_{T-1} \textit{ or } f_{T-1}(A^{opt}_{T-1}) - f_{T-1}(B^{opt}_{T-1}) > 0. \tag{63}$$

**4.4. An integration lemma.** Let us rewrite the difference in (62) as

$$W(\theta) := f_{T-1}(A^{\text{opt}}_{T-1}) - f_{T-1}(B^{\text{opt}}_{T-1})$$

to emphasize its dependence on $\theta$; and note $D$ does not depend on $\theta$. The key ingredient in the proof of the lower bound is the following lemma, to be proved in section 4.5.

LEMMA 19. *There exists $C_3 > 0$ such that for all $0 < \theta < \tau$, for all $k \geq 0$, and $x > 0$,*

$$\mathrm{P}(T \geq k,\ 0 < \mathbb{1}_D W(\theta) < x) \leq C_3 x(k+1)\mathrm{P}(T \geq k).$$

Taking $k = 0$ in this lemma we get

$$\mathrm{P}(0 < \mathbb{1}_D W(\theta) < x) \leq C_3 x. \tag{64}$$

Recall a simple integration lemma (for a more general result see [2, Lemma 6(a)]).

LEMMA 20. *Let $V \geq 0$ be a real-valued random variable such that*

$$\mathrm{P}(0 < V < x) \leq Cx, \quad 0 < x < \infty.$$

*Then*

$$\mathrm{E}V \geq \frac{[\mathrm{P}(V > 0)]^2}{2C}.$$

By (64) and Lemma 20, we get

$$\begin{aligned} \varepsilon(\theta) &= \mathrm{E}(\mathbb{1}_D W(\theta)) \text{ by (62)} \\ &\geq \frac{[\mathrm{P}(W(\theta)\mathbb{1}_D > 0)]^2}{2C_3}. \end{aligned} \tag{65}$$

To finish the proof of (58), we need the following lemma.

LEMMA 21. *There exists a positive constant $C_4$ such that, for all $0 < \theta < \tau$,*

$$\mathrm{P}(W(\theta)\mathbb{1}_D > 0) \geq C_4\theta. \tag{66}$$

*Proof.* By assumption (7) we may assume that the constant $\tau$ at (38) is such that

$$\inf_{1/2-2\tau<x<1/2\tau} g(x) > 0, \tag{67}$$

where $g$ is the density function for $\xi_i$. Consider the following event:

$$\Omega(\theta) = \{\xi_{-1} \in (0,1/2),\ \xi_0 \in (0,1/2-\tau),\ \xi_1 \in (1/2-\tau,1/2),$$
$$\xi_2 \in (1-\xi_1-\theta, 1-\xi_1),\ \xi_3 \in (0,1/2-2\tau)\}.$$

Using (67) there exists $C_4 > 0$ such that

$$\mathrm{P}(\Omega(\theta)) \geq C_4\theta.$$

Assume this event $\Omega(\theta)$ happens. Then $\xi_{-1}+\xi_0 \leq 1-\tau$, $\xi_0+\xi_1 \leq 1-\tau$, $1-\theta < \xi_1+\xi_2 < 1$, and $\xi_2+\xi_3 \leq 1-\tau$. So $D$ happens and, using Lemma 9(a), we have $\{1,2,3\} \in A^{\mathrm{opt}}$, and by Lemma 16(b) the same holds true for $A^{\mathrm{opt}}_{T-1}$. Still assuming $\Omega(\theta)$ occurs, we see that for $B = A^{\mathrm{opt}}_{T-1}\backslash\{2\}$, we have $f_{T-1}(A^{\mathrm{opt}}_{T-1}) - f_{T-1}(B) = 1-\xi_1-\xi_2 \in (0,\theta)$ and therefore $f_{T-1}(B)+\theta|A^{\mathrm{opt}}_{T-1} \triangle B| > f_{T-1}(A^{\mathrm{opt}}_{T-1})$, implying $0 < W(\theta)$ by (63). In particular

$$\mathrm{P}(W(\theta)\mathbb{1}_D > 0) \geq \mathrm{P}(\Omega(\theta)) \geq C_4\theta,$$

and we have proved assertion (66). ∎

From (65) and (66), we directly get the second assertion (58) of Lemma 14. We now show how to obtain the first assertion of Lemma 14. Recall that, by definition, we have

$$f_{T-1}(B^{\mathrm{opt}}_{T-1}) + \theta|A^{\mathrm{opt}}_{T-1} \triangle B^{\mathrm{opt}}_{T-1}| \geq f_{T-1}(A^{\mathrm{opt}}_{T-1});$$

hence we get $\theta T > \theta|A^{\mathrm{opt}}_{T-1} \triangle B^{\mathrm{opt}}_{T-1}| \geq W(\theta)$. In particular, by Proposition 18, we have $D \cap \{W(\theta) > 0\} \subset D \cap \{\theta T > W(\theta) > 0\}$. Also, by Lemma 19, we have

$$\begin{aligned}
\delta(\theta) &\leq \mathrm{E}[T\mathbb{1}_D\mathbb{1}(W(\theta) > 0)] \text{ by (61)}\\
&\leq \sum_j j\mathrm{P}(T \geq j, \theta j > W(\theta) > 0)\\
&\leq C_3\theta \sum_j j^2(j+1)\mathrm{P}(T \geq j)\\
&\leq C_3\theta\mathrm{E}[(T+1)^4],
\end{aligned}$$

and $T/3$ has a geometric distribution so that assertion (57) of Lemma 14 follows.

**4.5. Proof of Lemma 19.** Write $W = W(\theta)$. Consider the random collection

$$\mathcal{B}(T-1) := \{B \subseteq \{1,2,\ldots,T-1\} : \ B \neq A^{\text{opt}}_{T-1}, 1 \in B, T-1 \in B\}.$$

By Proposition 18

$$\text{on } D, \text{ either } W = 0 \text{ or } W \geq \min_{B \in \mathcal{B}(T-1)} (f_{T-1}(A^{\text{opt}}_{T-1}) - f_{T-1}(B)) > 0. \tag{68}$$

Our first goal is to derive a lower bound (Lemma 24) for the right-hand side of (68) in terms of the $\xi_i$'s. Until the end of the proof of Lemma 24, we are working on the event $D$.

Let $C = \arg\min_{B \in \mathcal{B}(T-1)} (f_{T-1}(A^{\text{opt}}_{T-1}) - f_{T-1}(B))$ be the optimal perturbation of $A^{\text{opt}}$ on $[1, T-1]$. For any subinterval $\mathcal{I} = [\ell, m] \subseteq [1, T-1]$ write $\mathcal{I}_e = [\max(\ell - 1, 1), \min(m+1, T-1)]$. Decompose $A^{\text{opt}} \triangle C$ as $\cup_i \mathcal{I}_i$, where the $\mathcal{I}_i$'s are disjoint maximal intervals of $A^{\text{opt}} \triangle C$. Then

$$\begin{aligned} f_{T-1}(A^{\text{opt}}) - f_{T-1}(C) &= \sum_i \left( f_{(\mathcal{I}_i)_e}(A^{\text{opt}} \cap (\mathcal{I}_i)_e) - f_{(\mathcal{I}_i)_e}(C \cap (\mathcal{I}_i)_e) \right) \\ &= \sum_i \left( f_{T-1}(A^{\text{opt}}_{T-1}) - f_{T-1}(C_i) \right), \end{aligned}$$

where $C_i = (A^{\text{opt}}_{T-1} \cap {\mathcal{I}_i}^c) \cup (C \cap (\mathcal{I}_i)_e)$. This implies that $A^{\text{opt}} \triangle C$ is a *single* subinterval $\mathcal{I}$ of $[1, T-1]$.

We now look at the possible perturbations of $A^{\text{opt}}$ on the interval $[0, T]$. Recall that we are working on the event $D$, and that $A^{\text{opt}}$ contains $0, 1, T-1, T$. Let $L_0, L_1, \ldots, L_K$ be the maximal subintervals $[a, b] \subseteq A^{\text{opt}} \cap [0, T]$ for which $b > a$, that is, with at least two elements. So we can partition $[0, T]$ as $L_0 \cup S_0 \cup L_1 \cup S_1 \cup \ldots \cup L_K$, where the $S_k$'s are the complementary intervals. We call the $L_k$'s *lakes* and the $S_k$'s *switches*.

LEMMA 22. *Let $L = [a, b]$ be a lake. For any set $B \in \mathcal{B}(T-1)$ such that $B \cap L^c = A^{\text{opt}} \cap L^c$ and hence $B \cap L \neq A^{\text{opt}} \cap L$, we have*

$$f_{T-1}(A^{\text{opt}}) - f_{T-1}(B) \geq \min\left\{1 - \xi_a, 1 - \xi_{b-1}, \min_{a \leq i \leq b-1} 1 - \xi_{i-1} - \xi_i \right\} > 0. \tag{69}$$

*Proof.* First suppose $B$ is obtained by removing from $A^{\text{opt}}$ a single item. If this item is $a$, we have $f_{T-1}(A^{\text{opt}}) - f_{T-1}(B) = 1 - \xi_a$; if it is $b$, we have $f_{T-1}(A^{\text{opt}}) - f_{T-1}(B) = 1 - \xi_{b-1}$, and if it is $i \in (a, b)$, then we have $f_{T-1}(A^{\text{opt}}) - f_{T-1}(B) = 1 - \xi_{i-1} - \xi_i$. So by optimality of $A^{\text{opt}}_{T-1}$ the first inequality in (69) holds for these $B$, and Lemma 16 implies the last inequality in (69). Now recall that Lemma 9(c) shows $\min_{a \leq i \leq b-1} 1 - \xi_i \geq 0$. So construct a general $B$ by removing items from $A^{\text{opt}}$ one by one, and for items after the first the benefit can only decrease. So the first inequality holds generally. ∎

LEMMA 23. *Let $S = [a, b]$ be a switch and $S_e = [a-1, b+1]$. For any set $B$ such that $B \cap S_e^c = A^{\text{opt}} \cap S_e^c$ and $B \cap S \neq A^{\text{opt}} \cap S$, we have*

$$\begin{aligned} f_{S_e}(A^{\text{opt}}) - f_{S_e}(B) \geq \min\Bigg\{ &\min_{a-1 \leq i < j \leq b} \xi_i + \xi_j - 1, \min_{a-1 \leq i \leq b} \xi_i, \min_{a \leq i \leq b} \xi_i - \xi_{a-2}, \\ &\min_{a \leq i \leq b} \xi_i - \xi_{b+1}, 1 - \xi_{a-2} - \xi_{b+1} \Bigg\}. \end{aligned}$$

| | $a$ | | | | $u$ | | | | | | $v$ | | | | $b$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A^{\text{opt}}$ | ○ | . . . | ● | – | ○ | – | ● | . . . | ● | – | ○ | – | ● | . . . | ○ |
| $B$ | ○ | . . . | ● | – | ● | – | ○ | . . . | ○ | – | ● | – | ● | . . . | ○ |

FIG. 3. *Case* [○ . . . ○], *where* $a \le u \le v \le b$. *Benefit change* $= \xi_{u-1} + \xi_v - 1$.

| | $a$ | | | | $u$ | | | | | | | | $v$ | | | | $b$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A^{\text{opt}}$ | ○ | . . . | ● | – | ○ | – | ● | . . . | ● | – | ○ | – | ● | – | ○ | . . . | ○ |
| $C$ | ○ | . . . | ● | – | ● | – | ○ | . . . | ○ | – | ● | – | ○ | – | ○ | . . . | ○ |

FIG. 4. *Case* [○ . . . ●], *where* $a \le u \le v \le b-1$. *Benefit change* $= \xi_{u-1}$.

| | $a$ | | | | $u$ | | | | | | | | $v$ | | | | $b$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A^{\text{opt}}$ | ○ | . . . | ○ | – | ● | – | ○ | – | ● | . . . | ● | – | ○ | – | ● | . . . | ○ |
| $D$ | ○ | . . . | ○ | – | ○ | – | ● | – | ○ | . . . | ○ | – | ● | – | ● | . . . | ○ |

FIG. 5. *Case* [● . . . ○], *where* $a+1 \le u \le v \le b$. *Benefit change* $= \xi_v$.

TABLE 4

| $[u, v]$ | $[a-1, \circ]$ | $[a-1, \bullet]$ | $[\circ, b+1]$ | $[\bullet, b+1]$ | $[a-1, b+1]$ |
|---|---|---|---|---|---|
| benefit change | $\xi_v - \xi_{a-2}$ | $1 - \xi_{a-2}$ | $\xi_u - \xi_{b+1}$ | $1 - \xi_{b+1}$ | $1 - \xi_{a-2} - \xi_{b+1}$ |

*Proof.* By construction a switch starts and ends with items not in $A^{\text{opt}}$, and the two items before and after the switch are in $A^{\text{opt}}$. Moreover, Table 2 shows that two adjacent items cannot both be not in $A^{\text{opt}}$, so the items in a switch $[a, b]$ must strictly alternate between in and not in $A^{\text{opt}}$, as illustrated in Figure 3.

We first consider a set $B$ obtained from $A^{\text{opt}}$ by flipping all items in some subinterval $[u, v]$ of $[a, b]$. There are four cases, corresponding to whether the endpoints $u, v$ are in or not in $A^{\text{opt}}$. We exhibit three cases in Figures 3, 4, and 5, labeled as, e.g., [● . . . ○], together with the value of the benefit change $f_{S_e}(A^{\text{opt}}) - f_{S_e}(B)$. In the fourth case [● . . . ●], the benefit change equals 1.

We also need to consider cases where the flipped subinterval $[u, v]$ has $u = a-1$ or $v = b+1$ or both. There are five cases, indicated in Table 4.

Now consider any subset $B$ satisfying the hypothesis of Lemma 23. Decompose $A^{\text{opt}} \triangle B$ into disjoint maximal intervals $\mathcal{I}_i$. It is easy to check that the benefit change between $A^{\text{opt}}$ and $B$ is just the sum of the separate benefit changes between $A^{\text{opt}}$ and $A^{\text{opt}}$ with interval $\mathcal{I}_i$ flipped. Thus the minimum over $B$ is attained by one of the cases we have considered, establishing the lemma. ∎

LEMMA 24. *Set* $w = \min_{1 \le i < j \le T-1} \{|\xi_i + \xi_j - 1|; \xi_i; |1 - \xi_i|; |\xi_i - \xi_j|\}$. *On the event* $D$, *either* $W = 0$ *or* $W \ge w$.

*Proof.* We need only consider the case $W > 0$. Recall that $C = \arg\min_{B \in \mathcal{B}(T-1)} (f_{T-1}(A^{\text{opt}}_{T-1}) - f_{T-1}(B))$ is such that $A^{\text{opt}} \triangle C$ is a *single* subinterval $\mathcal{I}$ of $[1, T-1]$. It is enough to show that $C$ satisfies the assumptions of Lemmas 22 (for some lake) or the assumptions of Lemma 23 (for some switch), for then the lower bound $w$ follows from the lower bounds in those lemmas.

We argue by contradiction: if false, then $\mathcal{I}$ intersects some lake and some adjacent switch, say $L_k$ and $S_k$ (the case of $L_k$ and $S_{k-1}$ is similar). So there exists $a < b < c$ such that $b = \sup L_k$ and $\mathcal{I} = [a, c]$. Now check the following:

If $c \in A^{\text{opt}}$, then $f(B) - f(C) = 1$, for $B := C \cup \{b, b+2, b+4, \ldots, c\} \backslash \{b+1, b+3, \ldots, c-1\}$.

If $c \notin A^{\text{opt}}$, then $f(B) - f(C) = \xi_c$, for $B := C \cup \{b, b+2, b+4, \ldots, c-1\} \backslash \{b+1, b+3, \ldots, c\}$.

Either case contradicts the optimality of $C$. $\square$

We may now complete the proof of Lemma 19. The key point is that the bound $w$ in Lemma 24 does not depend on $\theta$. From Lemma 24,

$$\begin{aligned}
&\mathrm{P}(T \geq 3k,\ 0 < W(\theta)\mathbb{1}_D < x) \\
&\leq \mathrm{P}(T \geq 3k,\ D; 0 < w < x) \leq \mathrm{P}(T \geq 3k,\ w < x) \\
&\leq \mathrm{P}\left(T \geq 3k,\ \min_{1 \leq i < j \leq T-1} |\xi_i + \xi_j - 1| < x\right) + \mathrm{P}\left(T \geq 3k,\ \min_{1 \leq i \leq T-1} \xi_i < x\right) \\
&\quad + \mathrm{P}\left(T \geq 3k,\ \min_{1 \leq i \leq T-1} |\xi_i - 1| < x\right) + \mathrm{P}\left(T \geq 3k,\ \min_{1 \leq i < j \leq T-1} |\xi_i - \xi_j| < x\right).
\end{aligned}$$

The four terms on the right-hand side are treated similarly: we will just study the final term and will prove that there exists $C > 0$ independent of $k$ such that

$$\mathrm{P}\left(\min_{1 \leq i < j \leq T-1} |\xi_i - \xi_j| < x \,\middle|\, T \geq 3k\right) \leq C(k+1)x. \tag{70}$$

The effect of conditioning on the event $\{T \geq 3k\}$ is that each nonoverlapping triple $(\xi_{3m}, \xi_{3m+1}, \xi_{3m+2})$ is conditioned to satisfy either $\{\xi_{3m} + \xi_{3m+1} \geq 1 - \tau\} \cup \{\xi_{3m+1} + \xi_{3m+2} \geq 1 - \tau\}$ or $\{\xi_{3m} + \xi_{3m+1} < 1 - \tau,\ \xi_{3m+1} + \xi_{3m+2} < 1 - \tau\}$ (for $m = T$). It follows that, for any $i < j$,

$$\mathrm{P}((\xi_i, \xi_j) \in \cdot | T > j) \leq a^{-2} \mathrm{P}((\xi_i, \xi_j) \in \cdot), \tag{71}$$

where

$$\begin{aligned}
a = \min\big(&\mathrm{P}(\{\xi_0 + \xi_1 \geq 1 - \tau\} \cup \{\xi_1 + \xi_2 \geq 1 - \tau\}), \\
&\mathrm{P}(\xi_0 + \xi_1 < 1 - \tau, \xi_1 + \xi_2 < 1 - \tau)\big).
\end{aligned}$$

From assumption (7) the density of $\xi_j - \xi_i$ is bounded by some constant $b$, and so

$$\begin{aligned}
&\mathrm{P}\left(\min_{1 \leq i < j \leq T-1} |\xi_i - \xi_j| < x \,\middle|\, T \geq 3k\right) \\
&\leq \sum_{i<j} \mathrm{P}(|\xi_i - \xi_j| < x, T \geq j | T \geq 3k) \\
&= \sum_{i<j} \mathrm{P}(|\xi_i - \xi_j| < x | T \geq \max(j+1, 3k)) \mathrm{P}(T \geq j+1 | T \geq 3k) \\
&\leq b a^{-2} x \sum_{j \geq 3k-1} (j-1) \mathrm{P}(T \geq j+1 | T \geq 3k) \\
&\leq b a^{-2} x \sum_{j \geq k} 3(j+1) \mathrm{P}(T \geq 3j | T \geq 3k)
\end{aligned}$$

$$= ba^{-2}x \sum_{j \geq k} 3(j+1)\mathrm{P}(T \geq 3(j-k))$$

$$\leq ba^{-2}x(k\mathrm{E}[T] + \mathrm{E}[T(T+1)]),$$

where we used the fact that $T/3$ has a geometric distribution. This concludes the proof of Lemma 19.

## 5. Final remarks.

**5.1. Technical assumptions on $G$.** We stated a single assumption (7) on $G$. What we actually used was three consequences of this assumption:

- $\mathrm{P}(\xi < 1/2) > 0$, which implies $\mathrm{P}(\xi_i + \xi_{i+1} < 1) > 0$, was used in Lemma 15 and thereby throughout section 4 (because it implies $i \in A^{\mathrm{opt}}$) to implement "localization" arguments.
- $\mathrm{P}(\xi \leq 1/2) < 1$ was used in section 3.2 to show $P(\Omega_g) > 0$. Note that if $\mathrm{P}(\xi \leq 1/2) = 1$, then the optimization problem is degenerate in that the optimal $A_n^{\mathrm{opt}} = \{1, 2, \ldots, n\}$.
- $\xi_1 + \xi_2$ has density bounded below in some interval $(1, 1+\eta)$, which was used in section 3.2 to obtain (37).

The latter two are used only in a convenient way to exhibit one near-optimal set. The "localization" arguments essentially just require one to find some event of positive probability involving $(\xi_{-k}, \ldots, \xi_k)$ which forces items 0 and 1 to be in (or not in) $A^{\mathrm{opt}}$. Lemma 15 is just a simple way to exhibit such an event. So we expect Theorem 2 to remain true under much weaker assumptions on $G$.

**5.2. Parallels with the cavity method.** The arguments in this paper in the context of i.i.d.-DP (dynamic programming) may be compared with the more sophisticated arguments from the statistical physics *cavity method* [14], as reformulated in more probabilistic language in [1, 4], whose prototype example we take to be the analysis of the traveling salesman problem (TSP) in the "mean-field" model of geometry where there are $n$ points and each of the $\binom{n}{2}$ interpoint links has random length. Of course *algorithmically* DP and TSP are quite different, but there are striking parallels between the analysis of optimal solutions of i.i.d.-DP and mean-field-TSP, as follows:

- There are $n \to \infty$ limits for the random data; in DP this is just the obvious infinite i.i.d. sequence, while for mean-field-TSP it is a certain random infinite tree.
- The "inclusion criterion" for i.i.d.-DP involves $X_i^L, X_{i+1}^R$ and the edge-cost $\xi_i$. Finite-$n$ TSP has of course no simple inclusion criteria, but in the $n \to \infty$ limit of mean-field-TSP there is an analogous criterion for inclusion of an edge $(i,j)$ in terms of quantities $Z_i^L, Z_j^R$ and the edge-length $\xi_{ij}$. Each $Z$ is interpreted (cf. (19) for DP) as the difference between costs of two optimal solutions (subject to different local constraints) on one side of the tree.
- The distribution we use for $X$ in i.i.d.-DP, the stationary distribution of a Markov chain, is the solution of an equation with abstract structure $X \stackrel{d}{=} h(\xi, X^1)$. The distribution we use for $Z$ in mean-field-TSP, by a recursion on the limit tree, is the solution of an equation with abstract structure $Z \stackrel{d}{=} h(\xi; Z^1, Z^2, Z^3, \ldots)$, where the $Z^j$'s are i.i.d. copies of the unknown distribution $Z$.

These parallels provide a glimpse of how the analogue of Theorem 1, a formula for the asymptotic expected cost in mean-field-TSP, may be derived (the original nonrigorous argument was in [13]; a rigorous proof was given only recently via more combinatorial methods [18]). The analogue of Theorem 2 for mean-field-TSP, using Lagrange multipliers as in this paper, and leading to a nonrigorous argument that the scaling exponent equals 3, was given in [3].

**Acknowledgment.** We thank Vlada Limic for discussions regarding the NK model and Guilhem Semerjian for explaining to us an alternative proof of the $\delta(\theta) \sim \alpha\theta$ part of (55).

## REFERENCES

[1] D. J. Aldous, *The $\zeta(2)$ limit in the random assignment problem*, Random Structures Algorithms, 18 (2001), pp. 381–418.
[2] D. J. Aldous, C. Bordenave, and M. Lelarge, *Near-minimal spanning trees: A scaling exponent in probability models*, Ann. Inst. H. Poincaré Probab. Stat., 44 (2008), pp. 962–976.
[3] D. J. Aldous and A. G. Percus, *Scaling and universality in continuous length combinatorial optimization*, Proc. Natl. Acad. Sci. USA, 100 (2003), pp. 11211–11215.
[4] D. J. Aldous and J. M. Steele, *The objective method: Probabilistic combinatorial optimization and local weak convergence*, in Probability on Discrete Structures, Encyclopaedia Math. Sci. 110, Springer-Verlag, Berlin, 2004, pp. 1–72.
[5] R. Beier and B. Vöcking, *Typical properties of winners and losers in discrete optimization*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing, ACM, New York, 2004, pp. 343–352.
[6] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.
[7] A. Bogdanov and L. Trevisan, *Average-case complexity*, Found. Trends Theor. Comput. Sci., 2 (2006), pp. 1–106.
[8] R. Durrett, *Probability: Theory and Examples*, 3rd ed., Brooks/Cole, Pacific Grove, CA, 2005.
[9] R. Durrett and V. Limic, *Rigorous results for the $NK$ model*, Ann. Probab., 31 (2003), pp. 1713–1753.
[10] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York, 1993.
[11] V. Limic and R. Pemantle, *More rigorous results on the Kauffman-Levin model of evolution*, Ann. Probab., 32 (2004), pp. 2149–2178.
[12] C. McDiarmid, *On the method of bounded differences*, in Surveys in Combinatorics, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser. 141, Cambridge University Press, Cambridge, UK, 1989, pp. 148–188.
[13] M. Mézard and G. Parisi, *A replica analysis of the travelling salesman problem*, J. Physique, 47 (1986), pp. 1285–1296.
[14] M. Mézard and G. Parisi, *The cavity method at zero temperature*, J. Statist. Phys., 111 (2003), pp. 1–34.
[15] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*, World Sci. Lecture Notes Phys. 9, World Scientific, Teaneck, NJ, 1987.
[16] M. Talagrand, *A new look at independence*, Ann. Probab., 24 (1996), pp. 1–34.
[17] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians. Cavity and Mean Field Models*, Ergeb. Math. Grenzgeb. (3), Springer-Verlag, Berlin, 2003.
[18] J. Wästlund, *The Travelling Salesman Problem in the Stochastic Mean Field Model*, http://www.math.chalmers.se/~wastlund/.
[19] E. D. Weinberger, *Local properties of Kauffman's N-k model: A tunably rugged energy landscape*, Phys. Rev. A, 44 (1991), pp. 6399–6413.