

An Analysis of Graph Traversal with a Hindsight Oracle

J. R. Landers 

August 2026

Abstract

Clairvoyance does not imply deliberation. A hindsight oracle may know every future reward yet traverse its decision graph greedily. Its output can miss the globally best path; that best path may also be a poor learning target after selection from an exponentially rich family. We separate these effects. For multiplicative rewards with transition losses, a calibration-invariant certificate compares surrogate-greedy traversal with any path. A uniform complexity bound then identifies exceptional hindsight paths explainable by policy-class selection. Their synthesis distinguishes optimization error from the more elusive premium of clairvoyance. A stock-selection oracle grounds the discussion, but the results concern arbitrary layered decision systems.

1 QUESTION AND CONTRIBUTION

Knowing the future and optimizing over it are different powers, although their footprints can look remarkably alike. A hindsight oracle can possess the first while using a local rule that lacks the second. Conversely, an unrestricted optimizer can select a path whose success depends on so many realized contingencies that it is not a useful object to imitate. This note follows the observed gap to its two sources: the rule that traversed the graph and the comparator waiting at the end.

Contributions and scope. The argument proceeds in two movements. First, we give a pathwise certificate for multiplicative traversal with transition losses. It decomposes the gap to any comparator into movement, surrogate distortion, and approximate maximization, and Theorem 4 proves the certificate sharp for every prescribed sequence of these quantities. Second, we place a description cost on policies selected in hindsight and combine the resulting comparator restriction with the traversal certificate. The concentration step itself is the familiar prior-weighted union bound; the contribution is its use with the sharp path certificate to attribute an oracle gap without treating an exceptional realized path as an automatically valid learning target. A fully enumerated two-asset example lets the three comparators part ways in plain view.

Relation to neighboring theories. Metrical task systems and smoothed online optimization compare causal decisions with an offline optimum under service and movement costs [1, 2, 5]. Learning-augmented MTS additionally asks for consistency with fallible predictions and robustness when those predictions fail [10]. Our object is different: it is an offline diagnostic of a fixed, possibly clairvoyant but myopic rule. Advice complexity measures the future information an online method must receive to attain a competitive guarantee [11]. Here $K(\pi)$ is not a communication budget. It measures how surprising it was to select policy π from a prespecified class after the same realization was observed. These distinctions locate the note between competitive analysis and statistical model selection.

2 THE GRAPH AND THE ORACLE

Fix a finite horizon $T \in \mathbb{N}$, an initial vertex x_0 in a nonempty finite set $\mathcal{X}_0$, and nonempty finite layers $\mathcal{X}_1, \dots, \mathcal{X}_T$. Assume every pair in $\mathcal{X}_{t-1} \times \mathcal{X}_t$ is an available directed edge. At layer $t \in \{1, \dots, T\}$, choosing $x_t \in \mathcal{X}_t$ earns reward $r_t(x_t)$ for a map $r_t : \mathcal{X}_t \rightarrow (0, \infty)$; traversing the edge retains $a_t(x_{t-1}, x_t)$ for a map $a_t : \mathcal{X}_{t-1} \times \mathcal{X}_t \rightarrow (0, 1]$. A path $x = (x_1, \dots, x_T)$ therefore belongs to $\mathfrak{X}_T = \prod_{t=1}^T \mathcal{X}_t$ and has value

$$J(x) = \prod_{t=1}^T a_t(x_{t-1}, x_t) r_t(x_t). \quad (1)$$

The edge factor may represent movement, reconfiguration, latency, or physical loss. All logarithms in this note are natural. Taking logarithms turns (1) into an additive shortest-path objective:

$$-\log J(x) = \sum_{t=1}^T \{-\log r_t(x_t) - \log a_t(x_{t-1}, x_t)\}. \quad (2)$$

Taking logarithms reveals the additive structure concealed by compounding. Equation (2) therefore shows why shortest-path language is natural even though the objective is written multiplicatively.

The oracle ranks layer t using a positive surrogate $s_t : \mathcal{X}_t \rightarrow (0, \infty)$. Its resulting path is $g = (g_1, \dots, g_T) \in \mathfrak{X}_T$, and each choice need only be approximately greedy:

$$s_t(g_t) \geq \eta_t \max_{x \in \mathcal{X}_t} s_t(x), \quad 0 < \eta_t \leq 1. \quad (3)$$

Here η_t is the achieved approximation factor; exact maximization has $\eta_t = 1$. Smaller values absorb discretization, restricted candidate generation, and approximate maximization.

The score in (3) ranks destination vertices, not complete outgoing edges. This asymmetry is intentional. The motivating oracle ranks a target by its realized return over one interval, then incurs turnover and execution loss when that target differs from its current position. Thus a_t affects the realized path value but remains invisible to the ranking rule. A transition-aware heuristic can instead use any positive local score conditional on g_{t-1} ; after that predecessor is fixed, the same analysis applies with that local score as s_t . The present formulation isolates the common case in which ranking and execution are separate modules.

Stock motif. The market makes the distinction vivid. At a trading date, a vertex can denote a target portfolio and an edge a rebalance. Gross price change supplies r_t ; execution and turnover enter through a_t . A hindsight score might rank known close-to-close returns while positions are held open-to-open. The oracle sees the future but follows the best-looking destination, not the best complete route. The graph remembers the cost of moving even when the ranking rule does not. Cover’s universal portfolio competes with the best constant rebalanced portfolio in hindsight, while switching portfolios enlarge that comparator class [6, 7]. Our unrestricted path comparator is richer still, making comparator complexity part of the question. The same graph also describes machine reconfiguration, cache placement, routing, or sequential experimental design.

3 A PATHWISE CERTIFICATE FOR GREEDY TRAVERSAL

The useful quantity is not absolute score error, but the cross-sectional distortion between true reward and surrogate. What matters is whether the score preserves the relative geometry of a layer. Put

$$h_t(x) = \frac{r_t(x)}{s_t(x)}, \quad \Delta_t = \frac{\max_{x \in \mathcal{X}_t} h_t(x)}{\min_{x \in \mathcal{X}_t} h_t(x)} \geq 1, \quad (4)$$

and define the movement cost of a path by

$$D(x) = - \sum_{t=1}^T \log a_t(x_{t-1}, x_t).$$

Lemma 1 (One-step distortion). *For every comparator choice $y_t \in \mathcal{X}_t$,*

$$\frac{r_t(g_t)}{r_t(y_t)} \geq \frac{\eta_t}{\Delta_t}.$$

Proof. Since $r_t = s_t h_t$, approximate greediness gives

$$\frac{r_t(g_t)}{r_t(y_t)} = \frac{s_t(g_t) h_t(g_t)}{s_t(y_t) h_t(y_t)} \geq \eta_t \frac{\min_{x \in \mathcal{X}_t} h_t(x)}{\max_{x \in \mathcal{X}_t} h_t(x)}.$$

□

Theorem 2 (Surrogate-greedy path certificate). *For every comparator path $y = (y_1, \dots, y_T) \in \mathfrak{X}_T$,*

$$\log \frac{J(y)}{J(g)} \leq D(g) - D(y) + \sum_{t=1}^T (\log \Delta_t - \log \eta_t). \quad (5)$$

Consequently, if $x^ \in \arg \max_{x \in \mathfrak{X}_T} J(x)$ is a globally optimal path,*

$$\log \frac{J(x^*)}{J(g)} \leq D(g) + \sum_{t=1}^T (\log \Delta_t - \log \eta_t). \quad (6)$$

Proof. Multiply Lemma 1 over t , insert the transition factors from (1), and take logarithms. Equation (6) follows from $D(x^*) \geq 0$. □

The certificate is pathwise, computable without finding x^* , and calibration-invariant: replacing s_t by $c_t s_t$ for any $c_t > 0$ leaves Δ_t unchanged. It turns an opaque failure of hindsight into quantities that can be inspected. Approximate value functions also yield performance bounds for suboptimal multi-period investment policies [8]; here the bound instead separates observed movement from surrogate distortion.

In the stock motif, $D(g)$ measures turnover and slippage, η_t captures whole-share rounding or candidate pruning, and Δ_t measures the mismatch between ranking and holding intervals. Reporting them separately identifies why the global path was missed.

Corollary 3 (Growth-rate optimality). *Consider a sequence of such problems with horizon $T \rightarrow \infty$, and let $g = g^{(T)}$ and $x^* = x^{*,(T)}$ denote the corresponding greedy and optimal paths. Here $f(T) = o(T)$ means $f(T)/T \rightarrow 0$. If*

$$D(g) = o(T), \quad \sum_{t \leq T} \log \Delta_t = o(T), \quad \sum_{t \leq T} -\log \eta_t = o(T),$$

then

$$\frac{1}{T} \log \frac{J(x^*)}{J(g)} \rightarrow 0.$$

Thus greedy and optimal have the same asymptotic geometric growth rate, although their terminal values need not converge.

Theorem 4 (Sharpness of the certificate). *Fix $T \in \mathbb{N}$ and prescribed numbers $\eta_t \in (0, 1]$, $\Delta_t \geq 1$, and $d_t \geq 0$ for $1 \leq t \leq T$. There is an instance with two vertices per layer, an approximate-greedy path g satisfying (3), and a globally optimal path y such that*

$$D(g) = \sum_{t=1}^T d_t, \quad D(y) = 0, \quad \log \frac{J(y)}{J(g)} = \sum_{t=1}^T (d_t + \log \Delta_t - \log \eta_t).$$

Hence no uniformly smaller right-hand side can depend only on $D(g) - D(y)$, $(\eta_t)_{t=1}^T$, and $(\Delta_t)_{t=1}^T$; because $D(y) = 0$, the same conclusion holds for (6).

Proof. Let $\mathcal{X}_t = \{g_t, y_t\}$ and set

$$s_t(g_t) = \eta_t, \quad s_t(y_t) = 1, \quad h_t(g_t) = 1, \quad h_t(y_t) = \Delta_t, \quad r_t = s_t h_t.$$

Then g_t satisfies (3), with ties resolved toward g_t when $\eta_t = 1$, and the distortion in (4) is exactly Δ_t . Set $g_0 = y_0 = x_0$, assign $a_t(g_{t-1}, g_t) = e^{-d_t}$ and $a_t(y_{t-1}, y_t) = 1$, and set every other edge efficiency to 1. Since $r_t(y_t) = \Delta_t \geq \eta_t = r_t(g_t)$ and no edge efficiency exceeds 1, y is globally optimal. Finally,

$$J(g) = \prod_{t=1}^T e^{-d_t} \eta_t, \quad J(y) = \prod_{t=1}^T \Delta_t,$$

which gives the stated equality. $\square$

The construction leaves no hidden slack. Theorem 4 shows that stronger conclusions require additional structure, such as metric geometry, temporal regularity, curvature, or restricted paths.

4 A FULLY ENUMERATED STOCK EXPERIMENT

Consider two assets, A and B , over six holding periods, a world small enough that computation can hide nothing. Their gross reward vectors, in temporal order, are

$$r(A) = (1.18, .97, 1.16, .98, 1.15, .99), \quad r(B) = (.99, 1.17, .98, 1.16, .97, 1.18).$$

The surrogate score vectors are

$$s(A) = (1.20, 1.10, .90, .90, 1.10, 1.10), \quad s(B) = (.90, .95, 1.15, 1.10, .95, .90).$$

Let the first position be free to establish. Thereafter a switch retains 0.96 of wealth and staying retains all of it, so $a_1(x_0, x_1) = 1$ and, for $2 \leq t \leq 6$, $a_t(x_{t-1}, x_t)$ equals .96 when $x_{t-1} \neq x_t$ and 1 otherwise. Exact score maximization gives the greedy path $g = AABBA$.

All $2^6 = 64$ paths can be enumerated, so no solver stands between the assumptions and their consequences. The unrestricted optimum is $x^* = ABABAB$, which changes asset five times. To represent a preference for describable persistence, put a Markov prior on paths: the first asset is uniform, and the probability of switching at each later period is $\rho = .1$. If $N(x)$ denotes the number of switches in path x , then

$$P_\rho(x) = \frac{1}{2} \rho^{N(x)} (1 - \rho)^{5-N(x)}, \quad K_\rho(x) = -\log P_\rho(x).$$

With regularization strength $\beta = .05$, exhaustive maximization of $\log J(x) - \beta K_\rho(x)$, equivalently $J(x) P_\rho(x)^\beta$, selects $x_\beta = ABBBBB$. This comparator leaves some hindsight profit behind, but compresses five switches to one.

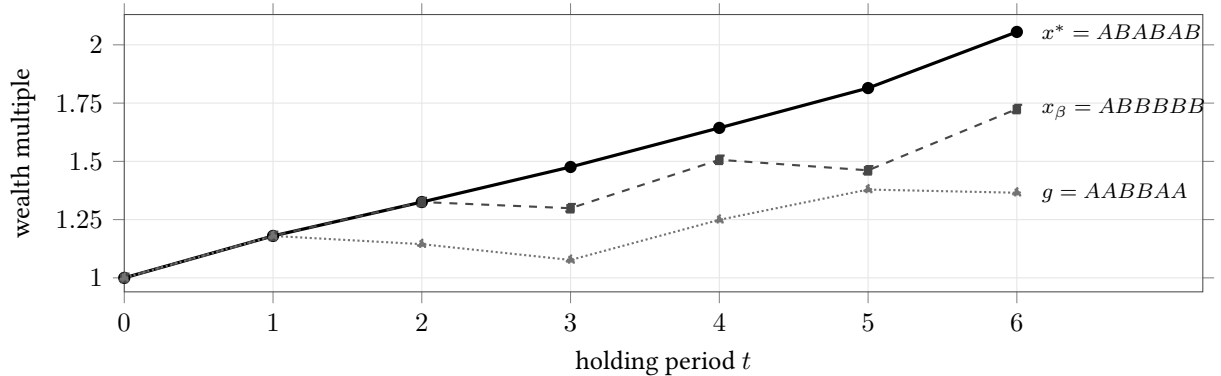


Figure 1: Cumulative wealth for the raw optimum, the description-penalized optimum, and score-greedy traversal. The paths agree briefly, then reveal three different meanings of “best.”

Figure 1 separates the two phenomena. Terminal wealth is 2.056 for x^* , 1.725 for x_β , and 1.365 for g . For the comparison with x^* , the realized log gap is .409. The certificate terms are $D(g) - D(x^*) = -.122$ and $\sum_t \log \Delta_t = 1.292$, with $\eta_t = 1$ throughout, giving the valid upper bound 1.169. Movement favors the greedy path because it switches less; the loss is attributable to surrogate distortion. Complexity regularization then asks a different question and retains much of the raw opportunity with a simpler route. The experiment is illustrative rather than an estimate of market performance, but every displayed quantity follows from exhaustive enumeration.

5 WHY THE BEST PATH MAY BE THE WRONG TARGET

Global optimality settles a deterministic question: which path scored highest on this realized graph? Learning asks the harder question: which rule could have selected a valuable path without storing the answer? An unrestricted hindsight optimum may thread a sequence of singular events, forming a route that is perfect precisely because it need never be walked again.

Let Y be the random environment generating the layer rewards and transition efficiencies, and let π be a causal policy, meaning a map from information available through layer $t - 1$ to an action in $\mathcal{X}_t$. For a realization of Y , the policy induces a path whose value is denoted by $J(\pi; Y)$. Write

$$L(\pi; Y) = \log J(\pi; Y).$$

The probability $\Pr$ and expectation $\mathbb{E}$ below are taken with respect to the law of Y . Fix, before observing Y , a probability mass function Q on a countable policy class $\mathcal{P}$, with $Q(\pi) > 0$ for every $\pi \in \mathcal{P}$ and $\sum_{\pi \in \mathcal{P}} Q(\pi) = 1$. Define the description cost

$$K(\pi) = -\log Q(\pi).$$

This is an Occam or minimum-description-length device [3, 4]: common, simple policies receive more mass; a policy that reproduces one exceptional path by a lookup table receives very little. The path remains real, but its explanation grows expensive. Tracking-the-best-expert bounds similarly charge a comparator for changing experts across segments [9]; $K(\pi)$ permits a broader charge for the full policy description.

Theorem 5 (Uniform hindsight-selection bound). *Suppose that, for each fixed $\pi \in \mathcal{P}$, there is a constant $V_\pi > 0$ such that the centered log value satisfies, for every $u \geq 0$,*

$$\Pr(L(\pi; Y) - \mathbb{E}L(\pi; Y) \geq u) \leq \exp\left(-\frac{u^2}{2V_\pi}\right).$$

Then, for every confidence level $\delta \in (0, 1)$, with probability at least $1 - \delta$ over Y , simultaneously for all $\pi \in \mathcal{P}$,

$$L(\pi; Y) - \mathbb{E}L(\pi; Y) \leq \sqrt{2V_\pi \left(K(\pi) + \log \frac{1}{\delta}\right)}. \quad (7)$$

Proof. Apply the assumed tail bound to policy π with failure probability $\delta Q(\pi)$, then sum those probabilities over $\mathcal{P}$. $\square$

If each layer has exactly $A \geq 2$ actions and the A^T recorded action sequences receive uniform prior mass A^{-T} , then such a sequence has $K = T \log A$. If also $V_\pi = \sigma^2 T$ for some $\sigma > 0$, its selection allowance in (7) is exactly

$$\sigma T \sqrt{2 \log A + \frac{2}{T} \log \frac{1}{\delta}},$$

which is asymptotic to $\sigma T \sqrt{2 \log A}$ for fixed δ . The allowance is linear in the horizon. An immense compounded payoff can therefore coexist with little evidence for a compact rule. The exceptional path

can be entirely real without being reproducible. This is the mathematical version of the stock oracle's spectacular but unteachable route: it is genuinely optimal after the year is known; what remains doubtful is its status as a learning target.

For a complexity budget $B \geq 0$, two replacements are natural whenever the displayed maxima are attained. The *budgeted oracle*

$$\pi_B^* \in \arg \max_{\pi: K(\pi) \leq B} L(\pi; Y)$$

asks for the best path generated by a policy of limited description. The *certified oracle* is any policy

$$\pi_{\text{cert}}^* \in \arg \max_{\pi \in \mathcal{P}} \left\{ L(\pi; Y) - \sqrt{2V_\pi \left(K(\pi) + \log \frac{1}{\delta} \right)} \right\}.$$

Neither declares simple policies morally superior. They merely require hindsight to account for every answer it was allowed to inspect.

6 THE ORACLE-GAP DECOMPOSITION

The two threads now meet. Fix a realization of Y and suppress it from the notation. Let x^* be an unrestricted maximizer of J over $\mathfrak{X}_T$, and let x_B^* be the path induced on this realization by the budgeted policy π_B^* . Define the *clairvoyance premium*

$$H_B = \log J(x^*) - \log J(x_B^*) \geq 0.$$

The exact identity is $\log[J(x^*)/J(g)] = H_B + \log[J(x_B^*)/J(g)]$.

Proposition 6 (Optimization–learnability decomposition). *Under the hypotheses of Theorem 2,*

$$\log \frac{J(x^*)}{J(g)} \leq H_B + D(g) - D(x_B^*) + \sum_{t=1}^T (\log \Delta_t - \log \eta_t). \quad (8)$$

Proof. Insert $J(x_B^*)$ into the ratio $J(x^*)/J(g)$ and apply Theorem 2 to the second factor. $\square$

Equation (8) bounds the raw oracle gap through four distinct contributions:

$$\underbrace{H_B}_{\text{clairvoyance}} + \underbrace{D(g) - D(x_B^*)}_{\text{movement}} + \underbrace{\sum_t \log \Delta_t}_{\text{surrogate distortion}} + \underbrace{\sum_t -\log \eta_t}_{\text{approximation}}.$$

Only the last three concern traversal quality. The first asks whether the comparator was ever an honest target.

Factorized plausibility. The framework is closed under a factorized plausibility penalty. Let P be a strictly positive probability mass function on $\mathfrak{X}_T$, with initial mass function q_1 on $\mathcal{X}_1$ and conditional mass functions $q_t(\cdot \mid x_{t-1})$ on $\mathcal{X}_t$, such that

$$P(x) = q_1(x_1) \prod_{t=2}^T q_t(x_t \mid x_{t-1}).$$

For regularization strength $\beta \geq 0$, $J_\beta(x) = J(x)P(x)^\beta$ again has form (1), with

$$\tilde{a}_1(x_0, x_1) = a_1(x_0, x_1)q_1(x_1)^\beta, \quad \tilde{a}_t(x_{t-1}, x_t) = a_t(x_{t-1}, x_t)q_t(x_t \mid x_{t-1})^\beta \quad (2 \leq t \leq T).$$

Plausibility is therefore not an ornament attached after optimization. When its description is sequential, it becomes part of the graph. The Markov penalty in Section 4 is precisely this construction with $P = P_\rho$ and regularization strength $\beta = .05$.

7 PERSPECTIVE

A hindsight oracle can be too weak when greedy misses coordinated paths, or too strong when unrestricted optimization elevates a path whose only concise description is “the one that happened to win.” The same oracle can therefore fail by seeing too little structure or by being permitted too much hindsight. Theorems 2 and 5 separate the geometric and algorithmic question from the statistical and descriptive one.

For stocks, the raw optimum maps realized opportunity; the complexity-bounded optimum maps opportunity available to a describable rule. Greedy’s distance from the latter is the meaningful algorithmic question. More generally, the theory asks not only whether greedy survives hindsight, but whether hindsight itself survives scrutiny.

REFERENCES

- [1] A. Borodin, N. Linial, and M. E. Saks, “An optimal on-line algorithm for metrical task systems,” *J. ACM*, 39(4):745–763, 1992.
- [2] G. Goel and A. Wierman, “An online algorithm for smoothed online convex optimization,” *SIGMETRICS Perform. Eval. Rev.*, 47(2):6–8, 2019.
- [3] D. A. McAllester, “Some PAC-Bayesian theorems,” *Machine Learning*, 37:355–363, 1999.
- [4] J. Rissanen, “Modeling by shortest data description,” *Automatica*, 14(5):465–471, 1978.
- [5] L. Zhang, W. Jiang, S. Lu, and T. Yang, “Revisiting smoothed online learning,” *NeurIPS*, 34:13599–13612, 2021.
- [6] T. M. Cover, “Universal portfolios,” *Mathematical Finance*, 1(1):1–29, 1991.
- [7] Y. Singer, “Switching portfolios,” *Int. J. Neural Syst.*, 8(4):445–455, 1997.
- [8] S. Boyd, M. T. Mueller, B. O’Donoghue, and Y. Wang, “Performance bounds and suboptimal policies for multi-period investment,” *Found. Trends Optim.*, 1(1):1–72, 2014.
- [9] M. Herbster and M. K. Warmuth, “Tracking the best expert,” *Machine Learning*, 32(2):151–178, 1998.
- [10] N. Christianson, J. Shen, and A. Wierman, “Optimal robustness-consistency tradeoffs for learning-augmented metrical task systems,” in *Proc. AISTATS*, PMLR 206:9377–9399, 2023.
- [11] H.-J. Böckenhauer, D. Komm, R. Kráľovič, R. Kráľovič, and T. Mömke, “On the advice complexity of online problems,” in *Proc. ISAAC*, LNCS 5878:331–340, 2009.