JOTA manuscript No.
(will be inserted by the editor)

Go Forth!

Simple Detection of Incomplete Meta-Learning by Algorithms Performing Limited Exploration on a Rugged Landscape

Peter Cotton and Lazy Boofhead

Received: date / Accepted: date

Abstract It is shown that an algorithm given two chances to improve its position on the path of an exponentiated Ornstein-Uhlenbeck (OU) process should not choose its final position between the first two locations. It is sometimes easy, therefore, to diagnose failure of an algorithm to learn the optimal policy.

The proof introduces the notion of rapidity on an OU bridge, and complements similar results in managerial science that are used as metaphors for complex but ill-defined business and strategy problems.

Keywords Titles · Sections · Formulas · More

Mathematics Subject Classification (2000) 49J53 · 49K99 · more

Lazy Boofhead

Playing Fortnite

'cause he's got tenure

1 Introduction

When exploration of a space is performed, we generally anticipate some back and forth motion in the sequence of points that are sampled. In a minimal example, an univariate optimizer that has sampled two points might often decide to pick a third inside the interval created by the first two.

Yet here an exception will be exhibited, and thus a synthetic test defined where failure to grasp the essence of the task can be very easily recognized. This can assist the benchmarking of few-shot exploration algorithms.

Secondary motivations are also considered, in the remainder of this introduction.

1.1 Management science

In *Searching for Good Policies* [3] by Steven Callander, the relevance to management science of this style of problem is discussed. One can borrow Callander's idea of using stochastic process paths to represent ill-posed business and strategy decisions. However here we consider three-shot optimization, whereas in the task set by Callander, only two function evaluations are allowed.

The construction of difficult sequential exploration problems is not unusual in the management science literature, and a recent review is provided by Baumann, Schmidt and Stieglitz [2]. They are, according to the authors, a "useful metaphor for the multidimensional search spaces that are spanned by complex problems".

Change can be difficult and costly. A professional golfer might embark on an effort to reconstruct his swing, but doing it too often might end her career. A startup might pivot from one product emphasis to another, though funding may run out before the forth product is attempted. A family might move from one region to another seeking a higher quality of life, but doing it too often may prove counter-productive.

A pharmaceutical company faces a very heavy price of exploration. The Ethereum network ponders a radical change from proof of work to proof of stake. The author of the Python programming language, Guido van Rossum, decided that his language needed to move to a new place (Python 3 is not backward compatible with Python 2) - though we don't expect breaking changes of this type too often.

The maxim to "move outside the bounds of existing experience", which we formalize, might be worthy of consideration in some of these situations.

1.2 Sequential time-series algorithm optimization

A more concrete motivation for the present work is development of adapting, online algorithms for live time-series prediction.

Of interest is the case where an algorithm adjusts key hyper-parameters, or parameters, a small number of times to meet the demands of a fast-paced setting where computational resources are at a premium. This task is the subject of research using a new platform. Live "microprediction" contests present a microcosm for the challenges [7]. In an attempt to face those, some algorithms employ a space-filling curve approach to exploration.

Recall that a space-filling curve is a map from some subset of $\mathcal{R}$, usually the unit interval, into some subset of $\mathcal{R}^d$, usually a hypercube. For example a Morton curve is most easily described by its inverse: the intertwining the digits in the binary expansions of d variables.

Peano curves are sometimes preferred in optimization [15]. Properties of pre-images of space-filling curves were considered by Darzentas, Nemirovsky and Yudin [8]. A recent approach by Lera and Sergeyev ([9], [14]) adapts a search using multiple DIRECT-like Lipschitz estimates.

In this work the test functions that are not Lipschitz (though they are Hölder continuous for exponents less than 1/2). This laboratory setting may inform selection of meta-learning techniques, especially if they are to be applied to rugged fitness functions. A de-noising model containing five parameters can illustrate the approach, and its operation is depicted in Figure 1. (For completeness the code is made available in the Python microfilter package [6] but these details are not important to the point being made.)

Figure 2 exhibits some pre-images of three-dimensional likelihood function sub-spaces for the de-noising model under a Morton curve - to give the reader a sense of the nature of those real-world fitness functions.

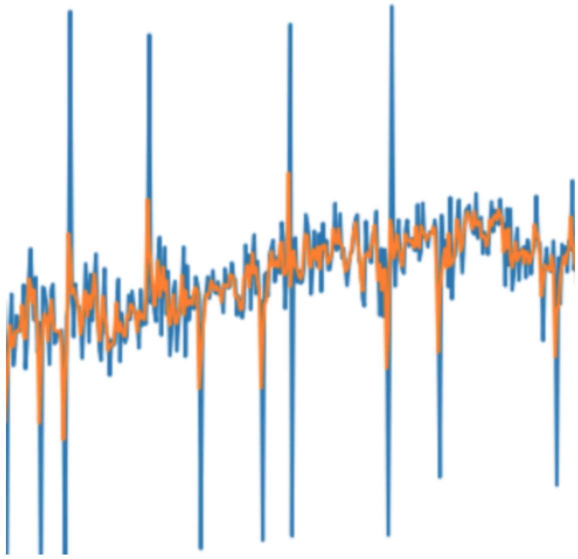

Fig. 1 An ad-hoc filter (smoothed series in orange, original in blue) designed to minimize chasing of outliers. Five parameters control the measurement noise model. Two parameters control the update to the anchor, shown in orange. No known online fitting procedure is known in this example. Instead, the filter makes a relatively expensive moves to a new "location" periodically in parameter space. Of particular relevance to the discussion are the pre-images of likelihood functions under a space-filling curve, shown in Figure 2.

1.3 Related work

Arguably, few-shot exploration is under-served by examples admitting analytical treatment, and clean criteria for judging ongoing advances.

However in [10] the problem of optimizing a Brownian motion on the interval $[0, 1]$ is considered. In Calvin [4] a setup similar Callander [3] is assumed, insofar as the objective functions are generated by Brownian motion.

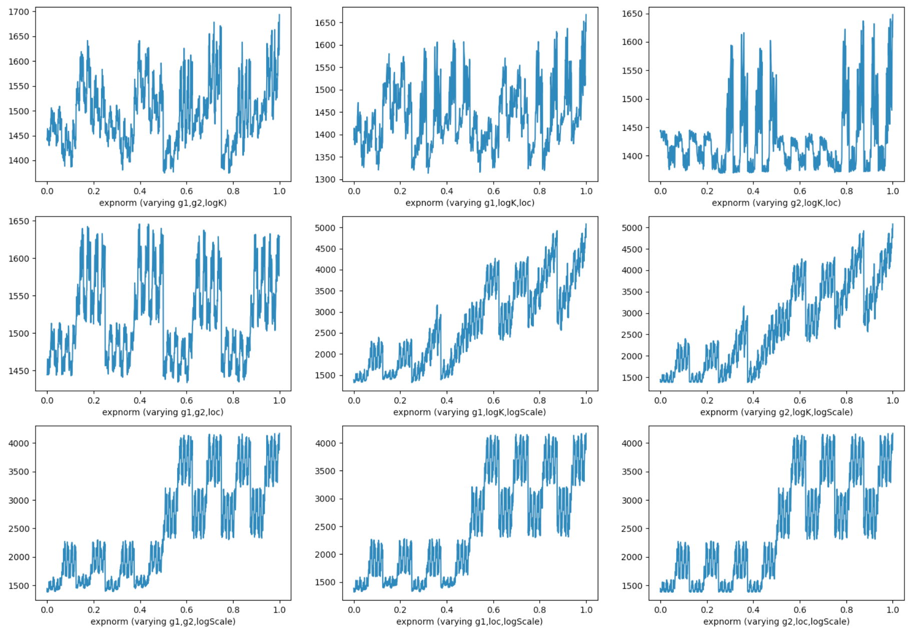

Fig. 2 Trivariate log-likelihood functions for a non-linear filter composed with the inverse of a Morton space-filling z-curve. A speculative exploration strategy models the local behaviour of these paths as exponentiated Ornstein-Uhlenbeck processes. The diagnostic provided in this paper can be used to independently assess the meta-learning capability of few-shot exploration strategies.

In both cases Wiener measure defines the test suite so there is some similarity to our landscape choices.[1] But in both cases, the intent is the analysis of performance for many-trial optimization algorithms - as compared with analytic results for very limited exploration.

[1] The Ornstein-Uhlenbeck process can be formulated as dynamically scaled, time-dilated Brownian motion, making an even closer connection to other test suites and their definitions of mean performance.

In [10] approximations to Thompson sampling in Bayesian optimization are anticipated. The authors make a remark that is probably relevant to applications of the current work. Many of the fitness functions to which their approach may be applied are not real-world examples but rather, are created as intermediate constructs during other optimizations.

There are ways in which the optimal policy presented here is reminiscent of the study of optimizers, insofar as sample choice is dictated in part by identification of a characteristic scale. However, usually that notion rests on estimation of a Lipschitz coefficient (central to many optimizers such as the DLIB package [13], or DIRECT [11] descendants surveyed in [12]) rather than the characteristic time-scale of a mean-reverting process.

The exploration problem considered is a special case of a continuous arm bandit problem [1]. But the emphasis in that literature has not, to date, centered on situations where we wish to make the most of a small number of samples. Nor, to my knowledge, have analytical solutions for sampling policy been supplied.

2 A test of meta-learning ability

We propose tasking general purpose meta-learning few-trial exploration algorithms with a sequence of difficult, rugged landscapes - each one generated synthetically.

The algorithm's choices of samples may, over repeated experiments, reveal the law of the process generating the objective functions (landscapes). However in practice there is no requirement, or assumption, that any candidate algorithm will perform that inference explicitly - and nor will they be provided with knowledge of how the landscapes were created.

Instead, it is anticipated that most algorithms will directly learn strategy choices instead, in a formal or heuristic manner, based on their success or otherwise.

2.1 Generative model for objective functions

Our test objection functions, or "landscapes" are exponential Ornstein-Uhlenbeck paths. To create one, first draw standard normal $b \sim N(0,1)$. Then sample one path of the Ornstein-Uhlenbeck (OU) process for $t > 0$:

$$dX_t = -\kappa X_t dt + \sqrt{2\kappa} dW_t \tag{1}$$

starting at $X_0 = b$. Use a second independent but otherwise identical process, also starting at $X_0 = b$, to sample $t < 0$ traveling left. Define the test function $f(t) = \exp(X_t)$, where X_t refers to the union.[2]

The excursions of the process $Y_t := \exp(X_t)$ help make this problem more life-like in many settings, where there can be some notion of locality but also super-linear reward. For instance a product that hits a real sweet spot might greatly outsell one that is proximate, but uninspiring to consumers.

[2] This construction avoids discussion of a process beginning at $-\infty$.

2.2 Preview of the optimal policy

One might make a few common sense observations about the challenge presented, adopting the position of one who has learned the law of the generating process.

The "mountain range" $f(t)$ is a fixed path of Y_t and everywhere positive, exhibiting the occasional high peak where excursions of the OU process get magnified by the exponential function. A strategy might find a high value on the first or second try, in which event it seems intuitive that a smaller step (or none at all) might be taken so as to trade-off the exploratory value against the high likelihood of "falling down the mountain".

Another perspective on the problem is provided by the mean variance trade-off. The mean value of the exponential of a gaussian variable includes half the variance, thus providing an incentive for exploration to increase the variance of X_{t_1} and X_{t_2}. On the other hand, the mean reversion drags us toward the line $Y_t = 1$ and, unless $b < 0$, that means down. A high value of $X(0) = b$ suggests we not risk any serious departure.

At the other extreme, if b is less than zero, we're better off forgetting it altogether and moving as far away as possible. That choice to flee, informally $t_1 = \infty$, will be formalized in Section 5.1.

If we don't flee, it will be shown that once $t_1 > 0$ has been chosen we need never choose $t_2 \in (0, t_1)$. The only time we can do so, without penalty, is on a set of measure zero where $X_{t_1} = b$. Even then we are no better off than the best outside choice.

Thus it will be established that it is sufficient to consider strategies where we choose either $t_2 \leq 0$ or $t_2 \geq t_1$, depending on whether the discovered value at t_1 is larger than at t_0 or smaller.

3 Gaussian process updates

We begin our consideration of the optimal policy, to which a meta-learning algorithm might eventually hope to converge to.

To this end we assume the generative model is known. Then as a new game begins, the observations $f(t_0), f(t_1)$ betray the posterior distribution of $X_t = \log f(t)$ using the customary calculation for gaussian processes. The prior is the zero mean gaussian process with OU kernel $k(t_1, t_2) = e^{-\kappa|t_2 - t_1|})$.

Defining $\lambda(t) := e^{-\kappa t}$ the gaussian posterior for X_t after observation of $f(t_0) = e^b$ and $f(t_1) = e^d$ say, has posterior mean

$$\begin{aligned}
\mu_{b,d}(t_2) &= (k(t_2, t_0)\ k(t_2, t_1))) \begin{pmatrix} k(t_0, t_0) & k(t_0, t_1) \\ k(t_1, t_0) & k(t_1, t_1) \end{pmatrix}^{-1} \begin{pmatrix} \log f(t_0) \\ \log f(t_1) \end{pmatrix} \\
&= \frac{1}{1 - \lambda_{01}^2} (\lambda_{02}\ ,\ \lambda_{12}) \begin{pmatrix} 1 & -\lambda_{01} \\ -\lambda_{01} & 1 \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix} \\
&= \frac{\lambda_{02} - \lambda_{12}\lambda_{01}}{1 - \lambda_{01}^2} b + \frac{\lambda_{12} - \lambda_{02}\lambda_{01}}{1 - \lambda_{01}^2} d
\end{aligned}$$

where λ_{ij} is shorthand for $\lambda(|t_i - t_j|) = e^{-\kappa|t_i - t_j|}$ and, in avoidance of doubt, $\mu_{b,d}(t_2)$ means $E\left[(X(t_2)|f(t_0) = e^b, f(t_1) = e^d\right]$. Notice that when $t_2 > t_1$ $\lambda_{01}\lambda_{12} = \lambda_{02}$ and the first term vanishes leaving only a dependency on the closest discovered value - as we expect given the independent increments of the OU process.

Also, b can be reused since by translational symmetry of the law of the ExpOU process, an optimal policy can choose a first sample point $t_0 = 0$ without loss of generality, discovering $f(0) = e^b$. We'll assume the first location is $(0, e^b)$ throughout.

The conditional variance calculation for X_{t_2} is very similar, with similar simplification. It does not depend on the discovered values $f(t_0)$ or $f(t_1)$:

$$\nu(t_2) = 1 - \frac{1}{1-\lambda_{01}^2} \left(\lambda(|t_2 - t_0|) \ , \ \lambda_{12} \right) \begin{pmatrix} 1 & -\lambda_{01} \\ -\lambda_{01} & 1 \end{pmatrix} \begin{pmatrix} \overbrace{\lambda(|t_2 - t_0|)}^{=\lambda_{02}} \\ \lambda(|t_2 - t_1|) \end{pmatrix}$$

emphasizing the only terms depending on choice t_2. These formulas apply in the limit $t \to \infty$ if we use the convention $\lambda(\infty) = 0$, though not as written to the case $\lambda_{01} = 1$.

3.0.1 Symmetric bridge

When $t_2 \in (t_0, t_1)$ we'll say we are "on the bridge" and we have $\lambda_{02}\lambda_{12} = \lambda_{01}$. We'll be interested in the special case $b = d$ when the mean simplifies to:

$$\mu_{b,b}(t_2) = b\frac{\lambda_{02} + \lambda_{12}}{1 + \lambda_{01}} \tag{2}$$

The variance reduces to

$$\nu(t_2) = \frac{(1-\lambda_{02}^2)(1-\lambda_{12}^2)}{1-\lambda_{01}^2} \tag{3}$$

3.1 Mean performance

Hereafter b is treated as a parameter for the task of choosing t_1 and t_2. The algorithm should subsequently maximize $u = E[cf(t_1) + f(t_2)]$. We'll carry c for now (although the problem with $c = 0$ happens to be our primary interest).

In choosing the third point t_2 given distinct t_0 and t_1, we are mindful that mean of the process $Y(t_2)$ picks up half the variance $\frac{1}{2}\nu_{b,d}$ and thus that final task amounts to choosing t_2 maximizing $\mu_{b,d}(t_2) + \frac{1}{2}\nu(t_2)$. However, choice of t_1 will impact the expectation of b and d in the first term.

4 Optimal two-shot policy

In analogy to Calendar [3], though with a different fitness function, we consider the case where only two function evaluations are made. As this will be reused, we use 0 and t to denote the first and second locations, where $t > 0$ without any loss of generality.

Conditioned on $X_0 = b$ the mean and variance of X_t are given by $be^{-\kappa t}$ and $\left(1 - e^{-2\kappa t}\right)$ respectively. Thus maximizing (the logarithm of) $E[Y_t]$ yields a choice

$$\begin{aligned}
t^* &= \arg\max_{t\geq 0} \log E[e^{X(t)}] \\
&= \arg\max_{t\geq 0} \left(b\lambda(t) + \frac{1}{2}(1 - \lambda(t)^2) \right) \\
&= \arg\max_{t\geq 0} -\frac{1}{2}\left(\lambda(t) - b\right)^2 + \frac{1}{2}\left(b^2 + 1\right)
\end{aligned}$$

but recalling that $\lambda(t) := e^{-\kappa t}$ take values only in $[0, 1]$, this can only be maximized by choosing $\lambda(t) = b$, if $b \in (0, 1)$. There are three regions demanding

Region	$\lambda(t^*)$	t^*	$\log E[e^{X_{t^*}}]$	Decision
$b < 0$	0	∞	$\frac{1}{2}$	Reset
$0 < b < 1$	b	$-\frac{1}{\kappa}\log(b)$	$\frac{1}{2}\left(b^2+1\right)$	Explore
$b > 1$	1	0	b	Stay

Table 1 Optimal second and final sample point t when exploring an ExpOU process whose value at the first sample point $t_0 = 0$ is discovered to be e^b.

qualitatively different behaviour. Bringing these together, the utility of the two shot problem is given by $\zeta(b)$ where

$$\zeta(x) = \begin{cases} e^{\frac{1}{2}} & x < 0 \qquad (\textit{flee to } \infty) \\ e^{\frac{1}{2}(x^2+1)} & 0 \le x \le 1 \\ e^x & x > 1 \qquad (\textit{stay put}) \end{cases} \tag{4}$$

because if $b < 0$ we as best to run to infinity. We let $\pi^{(1)}(b)$ denote the function taking b to the optimal choice t^* in the two-shot problem.

$$\pi^{(1)}(b) = \begin{cases} \infty & b < 0 \\ -\frac{1}{\kappa}\log(b) & 0 \le b \le 1 \\ 0 & b > 1 \end{cases} \tag{5}$$

The strategy and outcome is summarized in Table 1.

5 Optimal three-shot policy

We redeploy the two-shot result when evaluating a subset of all three-shot strategies, and prove that this subset is sufficient.

5.1 Game formalization

First make a superficial game modification to avoid awkwardness in discussion of resets. Note that $t_0 = 0$ without loss of generality and $t_1 > 0$. Let $\mathcal{R}^+$ denote the positive real line extended with $+\infty$ only. Let $\bar{\mathcal{R}}$ denote the entire real line extended with $-\infty, +\infty$.

A player may choose $t_1 = \infty$. If they do, then a second half extended line shall appear, with a new path created independently of the first, and a sample collected at the origin. Then one of the two worlds will vanish. The world that ceases to be part of game is the one with the lowest discovered value (it doesn't matter how a tie is broken).

A player may choose $t_2 = \infty$ also, in which case we assign a final value e^η where $\eta \sim N(0,1)$ is a fresh, independent draw. It should be clear that these modifications are made in the interest of brevity and do not change the strategy.[3]

5.2 Policies

Let $\mathcal{C}^+ = \mathcal{R}^+ \times \mathcal{R}$ denote the half Cartesian plane with only abscissa extended. An endgame policy $\pi : \mathcal{R} \times \mathcal{C}^+ \to \bar{\mathcal{R}}$ takes the first discovered value b (equal to $Y_{t_0=0}$) and the second position and value $(t_1, d = Y_{t_1})$ into a third choice $t_2 \in \bar{\mathcal{R}}$.

[3] Alternatively, we could describe limits of strategies choosing finite t values. Or one could map $(0, \infty]$ to a an interval. However, both possibilities seem more cumbersome.

Let $\bar{\mathcal{R}}+$ denote the real line extended with $+\infty$ only. A policy Π is a function taking $\mathcal{R}$ into a tuple comprising a first choice $t_1 \in \mathcal{R}+$ and an endgame policy.

$$\Pi : b \mapsto (t_1, \pi)$$

Thus Π results in the sequence $0, t_1, t_2$ where $t_2 = \pi(b, (t_1, d))$ when b and d are revealed.

5.3 Outside policies

A policy $\Pi = (t_1, \pi)$ is said to be an outside strategy if $\pi(b, (t_1, d)) \in \bar{\mathcal{R}} \backslash (0, t_1)$ for all b and d. All strategies with $t_1 = \infty$ are also classified as "outside".

5.4 Canonical outside policies

Given the considerations in Section 4 it is only necessary to consider outside strategies whose endgame policies take on the following form:

$$\pi^{out}(b, (t_1, d)) = \begin{cases} -\pi^{(1)}(b) & d \leq b \\ t_1 + \pi^{(1)}(b) & d > b \end{cases} \tag{6}$$

where $\pi^{(1)}$ is give in Equation 5.

Without presupposing that an outside strategy will yield as high a utility as a strategy admitting the possibility of $t_0 < t_2 < t_1$, we will first consider the optimal outside strategies, beginning with the simplest case $b < 0$. The following will be useful.

Lemma 5.1 *Suppose $f(x)$ is constant for $x \leq 0$ and strictly monotonically increasing for $x > 0$. Suppose $\eta \sim N(\mu, \nu)$. Then $E[f(\eta)]$ is monotonic in both μ and ν.*

Proof Both observations follow immediately from changes of variables (translation and scaling, respectively).

5.5 Poor first location

Claim The best outside strategy when $b < 0$ resets by taking $t_1 \to \infty$, thus accepting the performance of the two-shot game, namely $E[\xi(\eta)]$ for $\eta \sim N(0, 1)$.

Proof This follows from Lemma 5.1 as

$$X_{t_1} \sim N\left(be^{-\kappa t_1}, 1 - e^{-2\kappa t_1}\right) \tag{7}$$

so $E[\xi(X_{t_1})]$ strictly increases as $t_1 \to \infty$.

5.6 Middling first guess - lower bound

Next consider outside policies restricted to $b > 0$. Though Equation 7 still applies, Lemma 5.1 does not help us since the mean and variance move in opposite directions as t_1 is varied.

It is obvious, however, that if $b \in (0, 1)$ a move will be made, since even if we were to forfeit one opportunity move (i.e. taking $t_1 = 0$) the solution from Section 4 dictates that we choose $t_2 = -\frac{1}{\kappa} \log(b)$.

Therefore, taking $t_1 = -\frac{1}{\kappa}\log(b)$ must be no worse than standing pat, since we could always choose to stop afterwards. But then following this (likely sub-optimal) policy to it's logical conclusion, we must apply 5.

$$\Pi^{lb} = b \mapsto \begin{cases} \left(\infty, \pi^{(1)}\right) \ , \ b \leq 0 \\ \left(-\frac{1}{\kappa}\log(b), \pi^{out}\right) \ , \ b > 0 \end{cases}$$

with a slight abuse of notation. Here $\pi^{(1)}$ refers to the map $(b, (\infty, d)) \mapsto \pi^{(1)}(d)$ which applies the two-shot problem solution to the second revealed value. Substituting into 7 we can assess the mean performance of this policy easily. Half the time the value is the value of the two-shot problem. But if $b > 0$ we instead use

$$X_{t_1} \sim N\left(b^2, 1 - b^2\right) \tag{8}$$

and $E[\xi(X_{t_1})]$ could be computed by quadrature.

5.7 Optimal outside policy $b \in (0,1)$

However we can clearly do better. For instance for $b \in (0,1)$ we can evaluate choices of t_1 by integrating the value of the two-shot game against the distribution of $X(t_1)$. If X_{t_1} is worse than b we can revert, so the integrand is floored at $e^{\frac{1}{2}(b^2+1)}$, the value of the two-shot game in the case $b \in (0,1)$.

$$u(\pi_{out}; t_1) = cE[e^{X_{t_1}}] + E\left[\begin{cases} e^{\frac{1}{2}(b^2+1)}, X_{t_1} \leq b \\ e^{\frac{1}{2}(X_{t_1}^2+1)}, X_{t_1} \in (b,1) \\ e^{X_{t_1}}, X_{t_1} \geq b \end{cases}\right] \tag{9}$$

Alternatively we can write

$$u(\pi_{out}) = cE[e^{X_{t_1}}] + P(X_{t_1} < b)\xi(b) + P(X_{t_1} > b)E[\xi(X_{t_1})|X_{t_1} > b]$$

It is certainly possible to wield analytical solutions for u, using integrals of the type

$$I(\mu, \sigma; x_1, x_2; a_0, a_1, a_2) := \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} e^{a_0 + a_1 x + a_2 x^2} dx$$

as noted in the appendix.

5.8 Optimal outside policy $b > 1$

Similar remarks apply to the case $b > 1$, except that the fallback is to stay put. Thus integrands are floored by e^b.

5.9 Inside strategies

If considering only outside strategies we have thus far avoided consideration of a third evaluation occurring in between our first two. Any strategy admitting $t_2 \in (0, t_1)$ for some value of t_1 will be called an inside strategy.

In what follows we will establish that for every inside strategy which chooses $t_2 \in (0, t_1)$ there is an alternative outside choice that has equal or better utility.

5.10 When inside strategies help the most

To remind ourselves that these considerations might apply to humans as well as machines, we refer to the interval $(0, t_1)$ as the *comfort zone*. It is bookended by prior experience. We first establish that backtracking to inside the comfort zone is most likely to help in the measure zero event when the first and second evaluations $X(0) = b$ and $X(t_1) = d$ take identical values.

Claim If an inside strategy ever has highest utility in the case $d \neq b$, there will also be an inside strategy with the highest utility when $d = b$

Proof Suppose an inside strategy is useful when $d < b$. It must be superior to choosing an outside strategy with $t_2 < 0$, where the relevant information is the value b. Similarly if an inside strategy is useful when $b > d$ it must be superior to a choice $t_2 > t_1$, where the relevant information is the value $d > b$. Either way, we are comparing a possible inside choice against an outside choice nearest the higher side of the bridge.

Now by inspection of Equations 2 and 3 the conditional variance for an inside choice $t_2 \in (0, t_1)$ is independent of b and d whereas the conditional mean is increasing in both.

From Lemma 5.1 it follows that raising the far end of the bridge to the same height would make the inside choice more attractive, not less, and it is then clear why an inside strategy that improves our outcome in the case $d \neq b$ will also improve our outcome in the case $b = d$.

5.11 The symmetric case

Although it is unlikely, the case $X(0) = X(t_1)$ is, therefore, seemingly critical to the determination of overall optimal strategy. Section 5.12 establishes a calculus showing that when $b = d$ every inside choice t_2 can be matched by and outside alternative t_2'. But let's first check this is true for $t_2 = \frac{1}{2}t_1$, the middle of the bridge.

Claim We need never choose $t_2 = \frac{1}{2}t_1$ when $d = b$

Proof When $t_2 = \frac{1}{2}t_1$ we have $\lambda_{02}^2 = \lambda_{12}^2 = e^{-\kappa(t_1-t_0)} = \lambda_{01}$. So from Equations 2 and 3

$$X_{t_2} \sim N\left(b\frac{2\sqrt{\lambda_{01}}}{1+\lambda_{01}}, \frac{1-\lambda_{01}}{1+\lambda_{01}}\right) \tag{10}$$

Notice that since $\left(\sqrt{\lambda_{01}}-1\right)^2 > 0$ the $\frac{2\sqrt{\lambda_{01}}}{1+\lambda_{01}} < 1$ as we expect. The bridge "sags" in the middle due to the pull towards zero. Now consider the outside choice $t_2' > t_1$ defined by

$$t_2' := t_1 - \frac{1}{\kappa}\overbrace{\log\left(\frac{2\sqrt{\lambda_{01}}}{1+\lambda_{01}}\right)}^{<0} > t_1$$

In analogy with 7

$$X_{t_2'} \sim N\left(b\lambda_{12'}\ ,\ 1-\lambda_{12'}^2\right)$$

for $\lambda_{12'} = e^{-\kappa(t_2'-t_1)}$. But evidently t_2' was chosen so that $\lambda_{12'} = \frac{2\sqrt{\lambda_{01}}}{1+\lambda_{01}}$ and after some simplification $1-\lambda_{02'}^2 = \frac{1-\lambda_{01}}{1+\lambda_{01}}$ matching Equation 10. It is apparent that the outside choice $X_{t_2'}$ and the middle of the bridge choice X_{t_2} share the same mean and variance. The outside choice $t_{2'}$ can be made instead of $t_2 = \frac{1}{2}t_1$.

As a remark, this is surely promising as we can present a plausibility argument for why $t_2 = \frac{1}{2}t_1$ is the best we can do inside, as follows.

As we travel to the center of the bridge we gain mean performance because the variance increases. We lose because the mean of X_{t_2} decreases, but the gradient of the mean decrease must be zero at the middle of the (symmetric) bridge. On the margin we have nothing to lose (mean wise) and variance to gain by sliding all the way to the middle, rather than stopping at some small distance from the center.

Of course, we'll tighten this intuition.

5.12 Bridge trigonometry

Next we consider an arbitrary choice t_2 on the bridge $(0, t_1)$ and establish a similar result. It helps to reduce subscript clutter so set

$$\lambda_2 = \lambda_{02} = e^{-\kappa(t_2 - t_0)} = e^{-\kappa t_2}$$

and

$$\lambda = \lambda_{12} = e^{-\kappa(t_1 - t_2)}$$

Recall Equation 2:

$$\mu_{b,b}(t_2) = b\frac{\lambda_{02} + \lambda_{12}}{1 + \lambda_{01}} = b\frac{\lambda_2 + \lambda}{1 + \lambda_2 \lambda}$$

Now notice that the λ's interact in a manner resembling relativistic velocity addition. This motivates the definitions:

$$\theta_2 = \tanh^{-1}\left(e^{-\kappa t_2}\right) = \tanh^{-1}(\lambda_1)$$
$$\theta = \tanh^{-1}\left(e^{-\kappa(t_1 - t_2)}\right) = \tanh^{-1}(\lambda)$$

called rapidities. Likewise an outside point t_2' can be defined to have rapidity

$$\theta_2' = \tanh^{-1}\left(e^{-\kappa(t_2'-t_1)}\right) = \tanh^{-1}(\lambda_{12'})$$

so that this gives us a natural way to represent the distribution of $X_{t_2'}$ and reason about its relation to inside points using trigonometric identities. Notice

$$X_{t_2'} \sim N\left(b\tanh(\theta'), \mathrm{sech}(\theta)\right)$$

whereas

$$\mu_{b,b}(t_2) = b\frac{\tanh(\theta_2)+\tanh(\theta)}{1+\tanh(\theta_2)\tanh(\theta)} = b\tanh(\theta_2+\theta)$$

which must also be the mean of t_2' if its rapidity is chosen to be $\theta_2+\theta$. Similarly the conditional variance on the bridge is

$$\begin{aligned}
\nu_{b,d}(t_2) &= b\left(\frac{1}{1-\lambda_2^2}+\frac{\lambda^2}{1-\lambda^2}\right)^{-1} \\
&= b\left(\frac{1}{1-\tanh^2(\theta_2)}+\frac{\tanh^2(\theta)}{1-\tanh^2(\theta)}\right)^{-1} \\
&= b\frac{1}{\cosh^2(\theta_2)+\sinh^2(\theta)} \\
&= b\frac{1}{\frac{1}{2}\left(\cosh(2\theta_2)+\cosh(2\theta)\right)} \\
&\leq b\frac{1}{\cosh(\theta_2+\theta)} \\
&= b\mathrm{sech}(\theta_2+\theta)
\end{aligned}$$

which is the variance for t_2'. In this sequence the inequality follows from convexity of cosh and an application of Jensen's Inequality. Equality holds for $\theta_2 = \theta$ at the middle of the bridge - the case we already considered.

We can now provide a terse rationale for eschewing inside strategies.

Claim There is never a need to choose any point on the bridge if $b = d$.

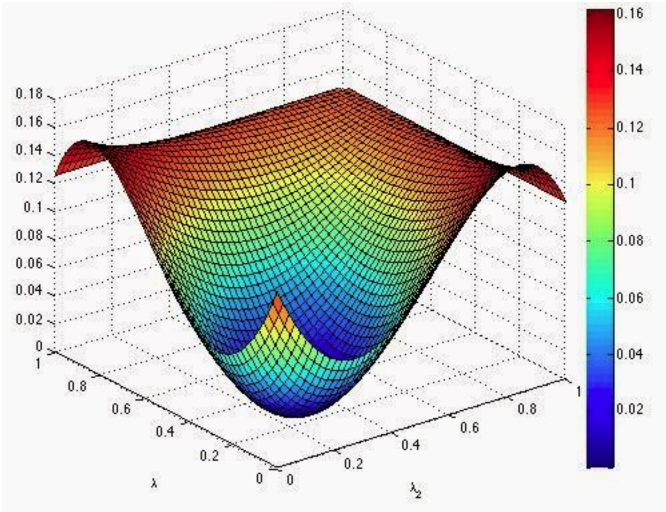

Fig. 3 The increase in utility achieved by moving outside the comfort zone, for the case $X(0) = b = 0.5$. The coordinates $\lambda_2 = e^{-\kappa t_2}$ and $\lambda = e^{-\kappa(t_1 - t_2)}$ for $t_2 \in (0, t_1)$ are employed in keeping with the proof of non-negativity.

Proof Each interior point defines two rapidities θ_2 and θ. It has been shown that an outside point with rapidity $\theta_2 + \theta$ will have the same mean and equal or higher variance. The result follows from the lemma.

A visual representation of this proof is given in Figure 3 where the utility differential between outside and inside strategies is shown in the coordinate system λ, λ_2. This plot assumes a middling first guess $b = 0.5$. The same comparison for the case of a lucky first guess $b = 1.5$ is shown in Figure 4.

5.13 Conclusion

We summarize the above with an argument by contradiction.

Claim Inside strategies are superfluous

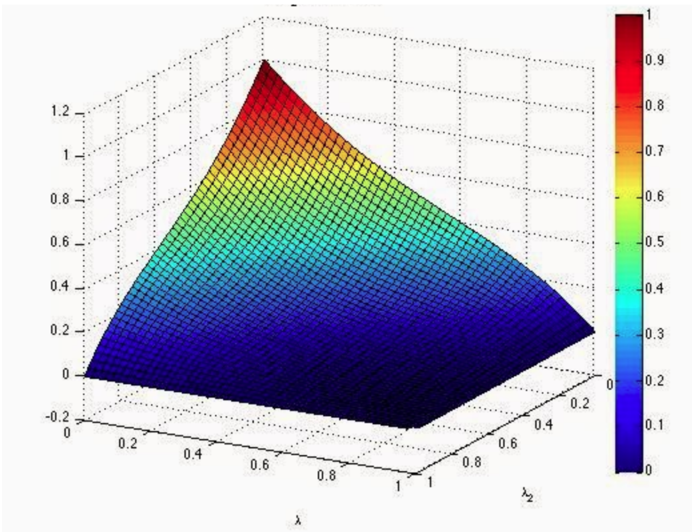

Fig. 4 Similar to Figure 3, we show the utility differential for $X(0) = b = 1.5$. Again, it pays to leave one's comfort zone.

Proof Suppose it is strictly beneficial to sometimes choose $t_2 \in (0, t_1)$, say after discovering $X(t_1) = d > b$. Then it would also be strictly beneficial to choose t_2 had we started with $X(0) = d$. In the light of Lemma 5.1 this contradicts the fact, proven above, that for a symmetric bridge we can exhibit an outside choice t_2' with the same mean and no lessor variance.

Similarly, suppose it is strictly beneficial to sometimes choose t_2 after discovering $X(t_1) = d$ with $d < b$. Then it would also have been strictly beneficial (indeed, even better) to do the same had we discovered $X(t_1) = b$. Again this is a contradiction.

5.14 Summary of the optimal three-shot policy

We have reasoned that inside strategies are superfluous, and therefore the outside strategy dictated indirectly by Equation 5 (or Table 1) and Equation 6 can be followed. To review, the strategy begins with $t_0 = 0$. Then there are two possibilities.

1. If $X(t_0) = b < 0$ we reset by choosing $t_1 \to \infty$, and follow the two evaluation policy thereafter, given by Table 1.
2. Otherwise, we choose $t_1 > 0$. Computing the value t_1 of this move involves a univariate optimization of lengthy closed form expressions whose existence is discussed further in the appendix. Thereafter we lean on the two evaluation policy, but with the highest valuation as a reference. This means:
   (a) Assuming the second evaluation $X(t_1) = d \in (b, 1)$, we then move further to the right to the point $t_2 = t_1 - \frac{1}{\kappa}\log(d)$.
   (b) If $X(t_1) = d \leq b < 1$ we move to $t_2 = \frac{1}{\kappa}\log(b) < 0$
   (c) We stay put if $b > 1$ or $d > 1$, choosing the best one.

6 Conclusions

It has been established that an optimal search policy on a rugged landscape defined by the exponentiated Ornstein-Uhlenbeck process should not backtrack (except on a set of measure zero).

Based on preliminary polling, it seems that people find this to be a counter-intuitive result. When given four possibilities (bridge middle, bridge inner, stay, move outside) only 23 out of 152 of those surveyed chose correctly. The rest were seemingly shy about venturing forth.

Details of the poll are described in [5]. A possible line of future work would improve this survey technique (say in a browser-based game) to determine if humans can learn the maxim "move outside the comfort zone" given repeated exposure to the problem.

The same test applies to allegedly meta-learning few-shot algorithms. This test is designed in anticipation of a rich variety of open-source packages for repeated few-shot exploration. Choosing a third point inside the first two is a simple "tell" that might betray algorithms with otherwise impressive credentials.

Acknowledgements This work was originally inspired by conversations with Steven Callander, who introduced me to the use of rugged landscapes in managerial science and the example in [3], in particular.

Appendix A. Utility computations

Closed form expressions for the mean performance are too lengthy to render, and beside the main point. However, the following identities are sufficient to establish existence.

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} e^{a_1 x + a_2 x^2} dx = \frac{1}{\sqrt{2\pi}}\int_{x_1/\sigma}^{x_2/\sigma} e^{-\frac{1}{2}(u-\mu/\sigma)^2} e^{a_1 \sigma x + a_2 \sigma^2 u^2} du$$

$$\frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{1}{2}(x-\mu)^2} e^{a_1 x + a_2 x^2} dx = \frac{1}{\sqrt{2\pi}}\int_{x_1-\mu}^{x_2-\mu} e^{-\frac{1}{2}u^2} e^{(a_1\mu + a_2\mu^2)+(a_1+2a_2\mu^2)u + a_2 u^2} du$$

$$\frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{1}{2}x^2} e^{a_1 x + a_2 x^2} dx = \frac{1}{\sqrt{p}}\frac{1}{\sqrt{2\pi}}\int_{x_1\sqrt{p}}^{x_2\sqrt{p}} e^{-\frac{1}{2}u^2} e^{\frac{a_1}{\sqrt{p}}u} du$$

$$\frac{1}{\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{1}{2}x^2} e^{a_1 x} dx = e^{\frac{1}{2}a_1^2}\frac{1}{\sqrt{2\pi}}\int_{x_1-a_1}^{x_2-a_1} e^{-\frac{1}{2}u^2} du$$

In the second to last equality $p = 1 - 2a_2$ and validity requires $p > 0$.

The case $b > 1$ is similar.

Appendix B. Morton curve on $\mathcal{R}$

An explanation of Figure 2.

let I^d denote a cube $(-1/2, 1/2)^d$ and I the interval $(-1/2, 1/2)$. Let $\phi : I \to I^d$ be a space filling curve - a function with dense image in I^d. Assume F is defined on R^d. Assume a bijection $\gamma : I = (-1/2, 1/2) \to \mathcal{R}$ implying a bijection $\mathit{\Gamma} : I^d \to R^d$ which is simply the tensor product of the univariate transforms γ applied coordinatewise. Define f with domain $\mathcal{R}$ by

$$f = F \circ \mathit{\Gamma} \circ \phi \circ \gamma^{-1}$$

(In optimization the tactic has, in the past, been motivated by an assumption that the univariate function f will satisfy a Lipschitz condition, or more generally a Hoelder continuity condition.

$$|f(u) - f(v)| < K|u - v|^\alpha$$

for some constants K and α).

References

1. Rajeev Agrawal. The Continuum-Armed Bandit Problem. *SIAM Journal on Control and Optimization*, 33(6):1926–1951, 1995.
2. Oliver Baumann, Jens Schmidt, and Nils Stieglitz. Effective Search in Rugged Performance Landscapes: A Review and Outlook. *Journal of Management*, 45(1), 2019.
3. Steven Callander. Searching for Good Policies. *American Political Science Review*, 105(4):643–662, 2011.
4. James M. Calvin. Average performance of a class of adaptive algorithms for global optimization. *Annals of Applied Probability*, 7(3), 1997.
5. Peter Cotton. A test of human intuition for mean-reverting processes.
6. Peter Cotton. Microfilter Python package, 2019.
7. Peter Cotton. Microprediction.Org, July 2020.
8. John Darzentas, A. S. Nemirovsky, and D. B. Yudin. Problem Complexity and Method Efficiency in Optimization. *The Journal of the Operational Research Society*, 35(5), 1984.
9. Hime A. E Oliveira and Antonio Petraglia. Global optimization using space-filling curves and measure-preserving transformations. In *Advances in Intelligent and Soft Computing*, volume 96 AISC, 2011.
10. Jean Bastien Grill, Michal Valko, and Rémi Munos. Optimistic optimization of a Brownian. In *Advances in Neural Information Processing Systems*, volume 2018-December, 2018.
11. D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1), 1993.
12. Donald R. Jones and Joaquim R.R.A. Martins. The DIRECT algorithm: 25 years Later, 2021.
13. Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10, 2009.
14. Daniela Lera and Yaroslav D. Sergeyev. GOSH: derivative-free global optimization using multi-dimensional space-filling curves. *Journal of Global Optimization*, 71(1), 2018.

15. G. Peano. Sur une courbe, qui remplit toute une aire plane. *Mathematische Annalen*, 36(1), 1890.