Using Landscape Theory to Measure Learning Difficulty for Adaptive Agents

Christopher H. Brooks[1] and Edmund H. Durfee[2]

[1] Computer Science Department
University of San Francisco
2130 Fulton St.
San Francisco, CA 94118
cbrooks@usfca.edu
[2] EECS Department
University of Michigan
1101 Beal Ave. Ann Arbor, MI 48109-2110
durfee@umich.edu

Abstract. In many real-world settings, particularly economic settings, an adaptive agent is interested in maximizing its cumulative reward. This may require a choice between different problems to learn, where the agent must trade optimal reward against learning difficulty. A landscape is one way of representing a learning problem, where highly rugged landscapes represent difficult problems. However, ruggedness is not directly measurable. Instead, a proxy is needed.
We compare the usefulness of three different metrics for estimating ruggedness on learning problems in an information economy domain. We empirically evaluate the ability of each metric to predict ruggedness and use these metrics to explain past results showing that problems that yield equal reward when completely learned yield different profits to an adaptive learning agent.

1 Introduction

In many problems, such as learning in an economic context, an adaptive agent that is attempting to learn how to act in a complex environment is interested in maximizing its cumulative payoff; that is, optimizing its performance over time. In such a case, the agent must make a tradeoff between the long-term value of information gained through learning and the short-term cost incurred in gathering information about the world. This tradeoff is typically referred to in the machine learning literature as the *exploration-exploitation tradeoff* [9]. If an agent can estimate the amount of learning needed to produce an improvement in performance, it can then decide whether to learn or, more generally, what it should learn. However, making this estimate requires that an adaptive agent know something about the relative difficulty of the problems it can choose to learn.

In this paper, we demonstrate how metrics from landscape theory can be applied to a particular agent learning problem, namely that of an agent learning the prices of information goods. A landscape is a way of representing the relative quality of solutions that lie near each other within some topology. We begin by describing our past work on the problem and appeal to a pictorial description to explain these results. Following this, we provide some background on landscapes and metrics for assessing their ruggedness, or difficulty. We then empirically evaluate two metrics, distribution of optima and autocorrelation, and show how these metrics can explain our previous results. We conclude by summarizing and discussing opportunities for future work.

2 Summarizing Price Schedule Learning Performance

In our previous work [1], we studied the problem of an adaptive agent selling information goods to an unknown consumer population. This agent acted as a monopolist and was interested in maximizing its cumulative profit. We assumed that the learning algorithm (amoeba [8], a direct search method) was a fixed feature of the agent. The adaptive agent's decision problem involved selecting a particular price schedule to learn, where this schedule served as an approximate model of consumer preferences. These schedules are summarized in Table 1.

Pricing Schedule	**Parameters**	**Description**
Pure Bundling	b	Consumers pay a fixed price b for access to all N articles.
Linear Pricing	p	Consumers pay a fixed price p for each article purchased.
Two-part Tariff	f, p	Consumers pay a subscription fee f, along with a fixed price p for each article
Mixed Bundling	b, p	Consumers have a choice between a per-article price p and a bundle price b
Block Pricing	p_1, p_2, m	Consumers pay a price p_1 for the first m articles ($m < N$), and a price p_2 for remaining articles.
Nonlinear Pricing	$p_1, ..., p_N$	Consumers pay a different price p_i for each article i.

Table 1. This table presents the parameters of six pricing schedules, ordering in terms of increasing complexity. More complex schedules allow a producer to capture a greater fraction of potential consumer surplus by fitting demand more precisely, but require longer to learn, since they have more parameters.

We found that simple schedules were learned more easily, but yielded lower profit per period once learned. More complex schedules took longer to learn, but yielded higher profits per period after learning. We ran experiments comparing the performance of six different pricing schedules (a sample is shown in Figure 1) and found that moderately complex two-parameter schedules tended to perform best in the short to medium-run, where learning is most important. In addition, the relative performance of the different schedules changes as the total number of articles (N) that the producer could offer was varied.

An agent that had these curves, which we call learning profiles, could then apply decision theory to determine which schedule to select at every iteration. The details of this are discussed in our previous work [1]; essentially, an agent must compare the expected cumulative profits gained from each schedule and select the one that yields the highest profits.

There are two problems with this approach. The experiments above do not explain *why* one schedule outperforms another, or why the relative performance of the schedules changes as the number of articles is varied. For example, two-part tariff and mixed bundling yield the same profits under perfect information, yet learning producers accrued higher profits per article with two-part tariff than with mixed bundling when $N = 100$. This leads us to ask both why these schedules have different learning profiles and why mixed bundling's performance depends upon N, whereas two-part tariff's performance seems not to. One way of explaining this is through an appeal to pictorial representations, such as Figure 3, where we see that two-part tariff has a single hill, whereas mixed bundling has a large plateau. As N increases, the size of this plateau grows, and so a large 'flat area' in the landscape is introduced, thereby thwarting an adaptive agent that employs a hill-climbing method.

Another complication is the number of consumers C in the population. For small values of C, performance on most schedules is lower (per consumer) than for large values of C. The conjectural argument is that large values of C tend to "smooth out" the landscape by producing a more uniform distribution of consumer preferences. Figure 2 shows an example of this for the one-dimensional linear pricing problem as N and C are varied. This type of pictorial argument is helpful as a visualization aid, but it is not particularly rigorous, and cannot be easily applied to functions with more than two inputs. A more precise measure of why these problems are different is needed.

Using the learning profile to determine a problem's complexity is also a problem in the many cases where an adaptive agent does not have this complete learning profile. Instead, it might have some sample problems, and need to use these to compare learning problems directly.

In order to estimate the difficulty of a learning problem when the learning profile is either uninformative or not available, we draw on results from landscape theory. Much of the recent work on landscape theory has taken place within the context of the theoretical study of genetic algorithms (GAs). In this case, the problem is to construct landscapes of varying difficulty to serve as benchmarks for a particular GA. Our problem is different; we assume that the landscape (the learning problem, or the mapping from inputs of the price schedule to profits) is determined by an external process (such as the composition of the consumer population) and our job is to characterize how hard it is. Rather than generating a landscape with particular features and claiming that these features make it

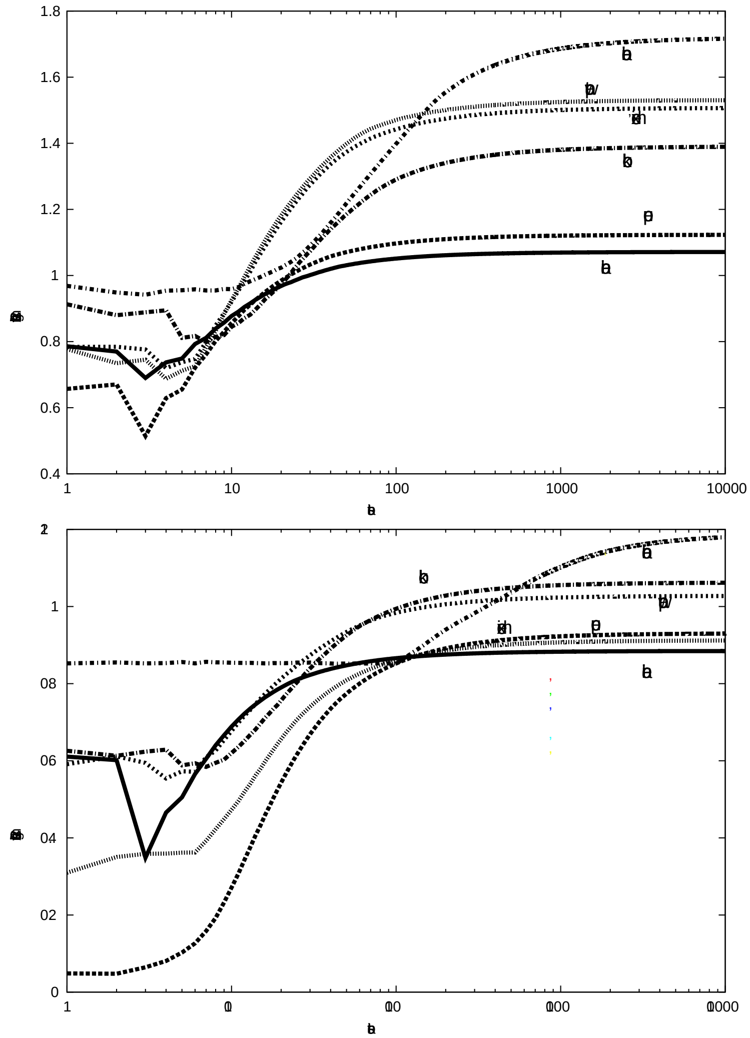

Fig. 1. Learning curves for six price schedules when a monopolist offers N=10 (left) and N=100 (right) articles. The schedules are: Linear pricing, where each article has the same price p; pure bundling, where a consumer pays b for access to all articles, two-part tariff, where a consumer pays a subscription fee f, plus a per-article price p for each article purchased, mixed bundling, where a consumer can choose between the per-article price p and the bundle price b, block pricing, where the consumer pays p_1 for each of the first i articles purchased and p_2 for each remaining article, and nonlinear pricing, where the consumer pays a different price p_i for each article purchased. The x axis is number of iterations (log scale) and the y axis is average cumulative profit per article, per customer.

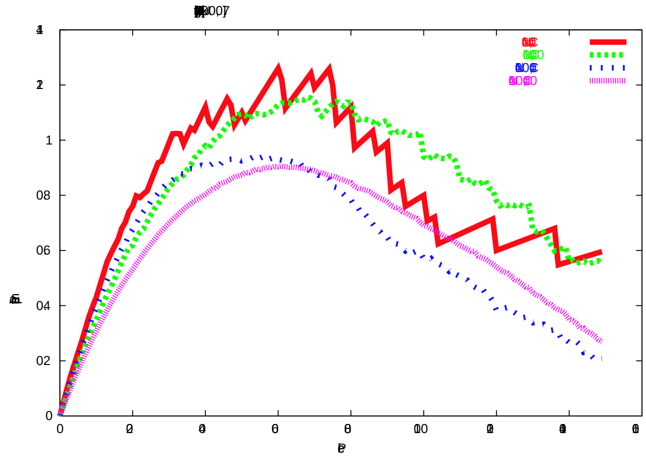

Fig. 2. Linear pricing landscapes for a small (10) and large (100) N and C.

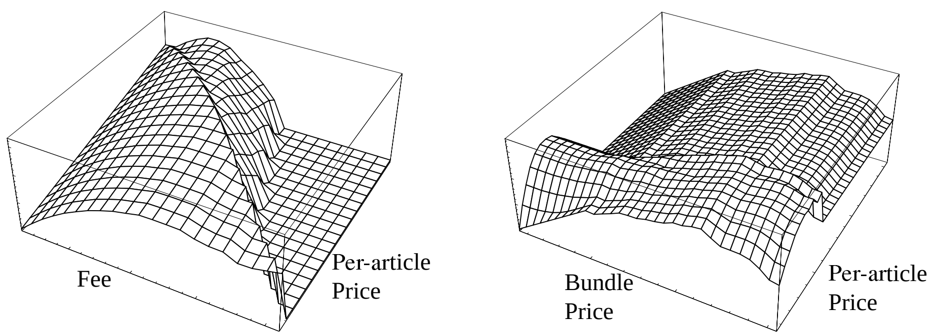

Fig. 3. Two-part tariff (left) and mixed bundling (right) landscapes. Even though they have the same number of parameters and the same optimal profit, their landscapes are very different.

difficult, we want to characterize the features of existing landscapes and identify sets of features that make adaptation difficult.

In the following section, we provide a context for this work, identifying some key results concerning landscape theory.

3 A Review of Landscape Theory

The concept of a *landscape* is a familiar one in evolutionary biology, optimization, and artificial intelligence. Figure 3 appeals pictorially to the concept. A landscape is visualized as a surface with one or more peaks, where each point on the landscape corresponds to a solution. Optimizing a function is cast as locating the highest peak. This idea is simple, yet extremely powerful. It allows a wide range of seemingly dissimilar problems to be cast in a common framework. In particular, the selection of a set of price schedule parameters that maximizes profit is equivalent to finding the global peak of a profit landscape.

The primary distinction that is made is between those landscapes that are *smooth* and those that are *rugged*. A smooth landscape is one that is easy to ascend; the optima can be located without much effort, and there are typically few local optima. A rugged landscape is one that contains discontinuities, many local optima, and other features that make it difficult for a local search algorithm (that is, one that is not able to see the entire landscape at once) to find the global optimum.

The notion of a rugged landscape has received a great deal of attention in the complex systems community. Kauffman [7] was one of the first researchers to describe a landscape's mathematical properties with respect to a search algorithm. (The concept was originally proposed by Sewall Wright [11] as a model for explaining natural selection.) Hordjik [3] and Jones [4], among others, tighten up Kauffman's concepts and apply more mathematical rigor.

A landscape consists of two components: an objective function F, and a neighborhood relation R that indicates the elements of the domain of F that are adjacent to each other.

F is the function that an agent is interested in optimizing. The input of F is indicated by the vector $\mathbf{x}$, where $\mathbf{x}$ can contain numeric (either real-valued or integer) elements or symbolic elements. Since this is an optimization problem, F maps into the reals; the goal of the problem is to find an $\mathbf{x}$ that maximizes F. In our price-setting problems, $\mathbf{x}$ is the parameters of the price schedule, and $F(\mathbf{x})$ is the resultant profit.

R is a neighborhood relation that, for any $\mathbf{x}$, returns those elements that are adjacent to $\mathbf{x}$. This provides a topology over F, and allows us to describe it as a surface. The choice of neighborhood relation may be exogenously determined by the input variables, or it may be endogenously determined by the user, depending upon the domain. If one is optimizing a price schedule and the inputs of $\mathbf{x}$ are the parameters of that schedule, each category, it is natural to define $R(\mathbf{x})$ to be the schedule one gets when one parameter in the schedule is increased or decreased by a set amount. In other problems, such as the traveling salesman problem, the problem may be encoded in a number of different ways, leading to different R relations and, subsequently, different landscapes that may be easier or harder to search. We treat the neighborhood relation as exogenously given, since we are searching over pricing parameters defined on either the reals or the integers, which have natural neighborhood relations.

Jones [4] presents a slightly different formulation of the R relation which depends upon the algorithm being used to traverse the landscape. Essentially, Jones' R is the successor relationship generated by a search algorithm; for a given state, R gives all the states that can be reached in one step for a particular algorithm. This formulation works well for Jones's purposes, which involve developing a theory for genetic algorithms, but it makes it difficult to compare two different landscapes and ask whether one is intrinsically easier or harder.

The distinction is that Jones couples the neighborhood relationship explicitly to the particulars of the search algorithm being used, whereas we assume that there are landscapes which have a natural neighborhood relationship. For example, prices occur on the real line, making a neighborhood relationship based on adjacency on the real line a natural choice. Since some price schedules induce profit landscapes that appear to be intrinsically easier than others to optimize, we would like our definition to capture this.

There have also been a variety of metrics proposed for comparing problem difficulty for genetic algorithms, in addition to the metrics described below. These metrics include fitness distance correlation [5] and epistatis variance and correlation [6]. Fitness distance correlation is similar to the autocorrelation metric we describe below. It examines how closely correlated neighboring points in a landscape are. Epistatis is a biological term that refers to the amount of 'interplay' between two input variables. If there is no epistasis, then each input variable can be optimized independently, whereas a large amount of epistasis (as is found in most NP-hard problems) means that the optimal choice for one of the inputs to F depends upon the choices for other values of F. Epistasis is most useful and easily measured when evolutionary algorithms are being employed for optimization.

4 Applying Landscape Theory to Price Schedule Learning

By estimating a landscape's ruggedness, an adaptive agent can then construct an estimate of how long it will take it to find an optimum and the learning cost associated with finding an optimum. However, ruggedness cannot be induced directly. Instead, it must be inferred from other landscape characteristics. When one is using a generative model such as the NK model to build landscapes, it is fine to build your model so that a parameter such as K can be used to tune those features of the landscape that make it difficult to optimize of and then claim that K is the amount of 'ruggedness.' However, when landscapes are provided exogenously, this is not an option. Instead, an agent must look at the measurable characteristics of a landscape and use this to estimate ruggedness. In this section, we consider three possible observable landscape characteristics and study their efficacy as estimators of ruggedness, using amoeba as a measure of actual ruggedness. These measures are also applied to the two-part tariff and mixed bundling landscapes and used to quantitatively explain the result we argued for pictorially in Figure 3.

4.1 Number of Optima

One metric that is sometimes discussed [6] for determining the difficulty of finding the global optimum of a particular landscape is the number of optima it contains. Intuitively, if a landscape contains a single optimum, it should be easy to locate. The exception to this is a landscape that contains a single narrow peak and a large plateau. (See Figure 4 for an example of this.) Similarly, if a landscape contains a large number of optima, it will be harder, particularly when using a hill-climbing algorithm, to find the global optimum.

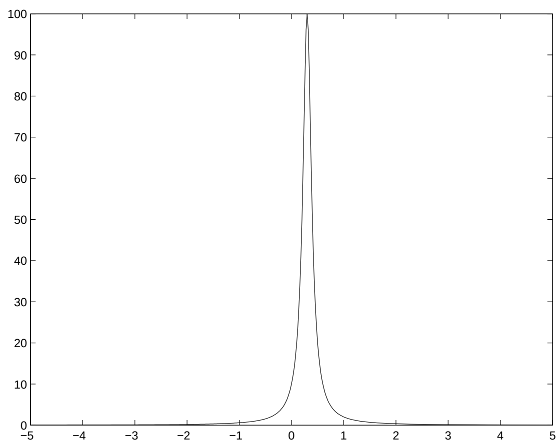

Fig. 4. A landscape with a single peak, but large plateaus, which present a challenge for hill-climbing algorithms.

If we accept that the number of optima is an indicator of ruggedness, and therefore of learning difficulty, we must then ask how we can determine the number of optima a landscape contains. In general, it is not possible to exhaustively search a landscape, at least one with more than a couple of dimensions. In addition, the profit landscapes we examine in this domain have continuous dimensions. In addition, the landscapes typically contain ridges and discontinuities. This makes the use of a standard hillclimbing algorithm to find optima a rather arbitrary exercise. Since the input dimensions are continuous and landscape contains ridges, the number of optima found will depend upon the granularity of the hillclimbing algorithm (how large a step it takes), rather than any intrinsic feature of the landscape. Given this problem, it is more useful to look at the *distribution* of optima, rather than the number of optima.

4.2 Distribution of Optima

In addition to practical problems in calculating the number of optima on profit landscapes with continuous inputs, there are deeper problems in using the number of optima as an estimator of ruggedness. A great deal of information is lost if optima are simply counted. A landscape with a large number of optima that are all clustered together at the top of a hill would seem to be qualitatively different (and less rugged) than one in which the optima are evenly spaced throughout the landscape. Again, thinking in terms of basins of attraction may make this easier to understand. In general, we would conjecture that the more the distribution of basin sizes tends toward uniform, the more rugged the landscape is. In Figure 3, we can see that the optima for two-part tariff are clustered on a hill, whereas mixed bundling contains a large plateau. This is a possible explanation for two-part tariff's being more easily learned, and yielding higher cumulative profit. In this section, we validate that argument by estimating the distribution of optima for two-part tariff and mixed bundling landscapes.

We will find the distribution of optima in a landscape by finding the distribution of basin sizes. To do this, we use the following technique, inspired by a similar approach used by Garnier and Kallel [2]. First, note that for any point in a landscape, we can find the optimum of the basin it resides in by using a steepest-ascent hillclimbing algorithm. We choose a random set p of starting points and use steepest-ascent to locate the corresponding optima. The distribution will be stored in a sequence β. For each optimum i, we calculate β_i, which is the number of points in p that lead to that optimum. Each element of β will correspond to an optimum i, and so the value of element β_i is the number of points from p that are in the basin of optimum i. If we sum all of the elements of β, we have a value for the 'size' of the landscape. (If we normalize this size, then the elements of β are percentages.) By fitting a sorted β to a distribution, we can estimate the clusteredness of the landscape's optima.

For simplicity, we fit β to an exponential distribution $e^{\lambda x}$, where the magnitude of λ indicates the clusteredness of the optima. The exponential distribution is convenient because it has only one free parameter, λ, which governs the clusteredness of the optima. When $\lambda = 0$, optima are uniformly distributed; as $abs(\lambda)$ increases, the distribution becomes more clustered. By using an exponential distribution, we can then fit $log(\beta)$ to a line, where the slope of the line is λ. Of course, we have no *a priori* reason (other than observation) to assume that an exponential distribution is the correct distribution. Future work will consider more complicated distributions, particularly ones such as the Beta distribution (not to be confused with our list β of basin sizes) that have "heavy tails."

To determine the distribution of optima for two-part tariff and mixed bundling landscapes, we performed the following experiment. For each schedule, we generated 10 landscapes using $N = 10$ and 10 landscapes using $N = 100$. Consumers were generated identically to those in the experiments summarized in Figure 1. For each landscape we chose $p = 1000$ points and ran a steepest-ascent hillclimbing algorithm to determine β. The value of p was determined by using a χ-square test, as in [2]; a p was selected, a distribution generated, and then the expected and actual distributions were compared. p was then increased and this comparison repeated. When increases in p no longer produced significant gains in confidence, p was taken to be "large enough."

Figure 5 compares the log distributions of β (averaged over 10 landscapes) for two-part tariff and mixed bundling for $N = 10$ and $N = 100$. Previously, we argued that two-part tariff performed well because the optima were clustered on a hill, whereas mixed bundling contained a large flat region that served as a set

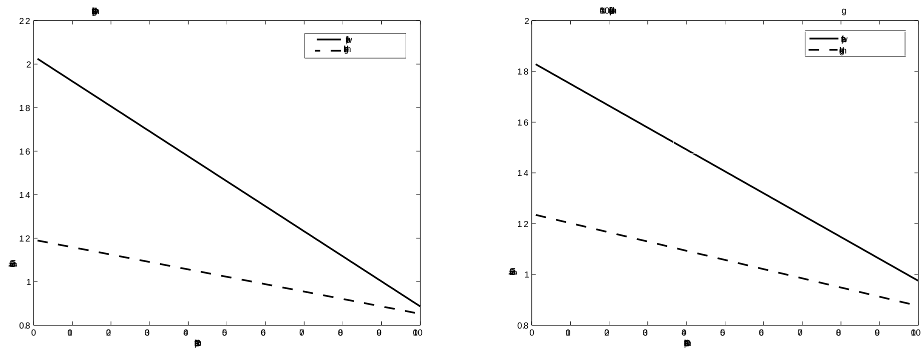

Fig. 5. Distribution of basin sizes for two-part tariff and mixed bundling. The x axis ranks basins sorted by size and the y axis (log scale) is the size of a particular basin. Each line is the fitted distribution of basin sizes. The left figure is for $N = 10$ and the right is for $N = 100$.

of optima. Figure 5 supports this argument; λ is an order of magnitude larger for two-part tariff than for mixed bundling.

We can also see that the optima for two-part tariff become less clustered as we move from $N = 10$ to $N = 100$. A closer examination of Figure 3 shows that the two-part tariff optima are located along a ridge; as N increases, this ridge grows, since larger values for the fee will yield positive profit. This spreads out the optima and reduces the magnitude of λ.

The distribution of optima for mixed bundling does not change as significantly as we move from $N = 10$ to $N = 100$. Recall that mixed bundling offers consumers a choice between per-article and bundle pricing. This creates a large plateau in the landscape where the per-article price is too high, and so all consumers buy the bundle. A small change in per-article price gives an adaptive agent no change in profit. As we increase N, this plateau takes up a larger portion of the landscape, but the optima on this plateau (really just the flat portion of the plateau) retain their *relative* basin sizes. This points out a weakness in using this normalized approach: we are measuring the fraction of a landscape occupied by each basin, rather than an absolute measure of its size, which increases with N. Measuring the absolute size of each landscape is a difficult thing; it clearly affects the learning process, but it is hard to do without making arbitrary assumptions.

4.3 Autocorrelation coefficient

Estimating the distribution of optima is a useful technique for explaining why two-part tariff outperforms mixed bundling, but there are other questions about landscape ruggedness that have been raised in this article, such as the role of N (the number of articles) and C (the number of consumers) in affecting ruggedness. In this section, we explain these differences in ruggedness using *autocorrelation.*

Hordjik [3] describes the use of autocorrelation as a method of measuring the ruggedness of a landscape. To construct this measurement, one conducts a random walk over a landscape, retaining all the $(x, F(x))$ pairs. This series of pairs is then treated not as a sequence of steps but instead as a series of state transitions generated by a Markov process. We can then apply a technique from time series analysis known as autocorrelation to find an estimate of ruggedness. What we wish to know is how well the last n points allow us to predict the value of the $n+1$th point. More importantly, we wish to know the largest t for which the $n+t$th point can be predicted from the nth point. The larger t is, the less rugged the landscape, since a learner will have a great deal of information that it can use to predict the next point. A small t indicates a rapidly changing landscape in which past observations have little predictive value, which is what is meant by ruggedness.

To be more precise, we begin by recalling that the covariance between two series of points X and Y $(Cov(XY) = E[XY] - \mu_X \mu_Y)$ is an indicator of how closely related these series are. If covariance is normalized by dividing by the product of the standard deviations of X and Y, then we have the *correlation* between X and Y, denoted $\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$.

Autocorrelation is a closely related concept, except that instead of comparing two series, we are going to compare a series to *itself*, shifted by a time lag, as a way of measuring the change in the process over time. The autocorrelation of points i steps apart in a series y is defined as: $\rho_i = \frac{E[y_t y_{t+i}] - E[y_t]E[t_{t+i}]}{Var(y_t)}$.

Autocorrelation allows us to determine the correlation length of a landscape [10], which we will use as an indicator of ruggedness. Correlation length is the largest i for which there is a significantly nonzero autocorrelation.

We compare the autocorrelation of two-part tariff and mixed bundling landscapes as N and C are varied. In addition, we consider two different sorts of paths: one collected through a steepest-ascent algorithm, which indicates ruggedness during optimization, and one collected through a random walk over the landscape, which serves as an overall characterization of ruggedness. This will help us to understand whether particular values of N and C play a part in the learnability of two-part tariff and mixed bundling.

The experiment works as follows: we generate a random profit landscape (using the distribution of consumers that generated the learning curves in Figure 1 and varying N and C between 10 and 100). We then choose 1000 random points on the landscape and run a steepest-ascent hill-climbing algorithm from each point until an optimum is reached. We then compute the autocorrelation over that path for all window sizes from 1 to 40 and average the results to get a mean autocorrelation (for each window size) during optimization for this landscape. This is then averaged across 10 landscapes, giving an average autocorrelation during optimization for each schedule. Next, for each landscape, we conduct a random walk of length 1000 and measure the autocorrelation over this walk with window sizes from 1 to 40. These random walk autocorrelations are then averaged across 10 landscapes to yield a random walk autocorrelation for each schedule.

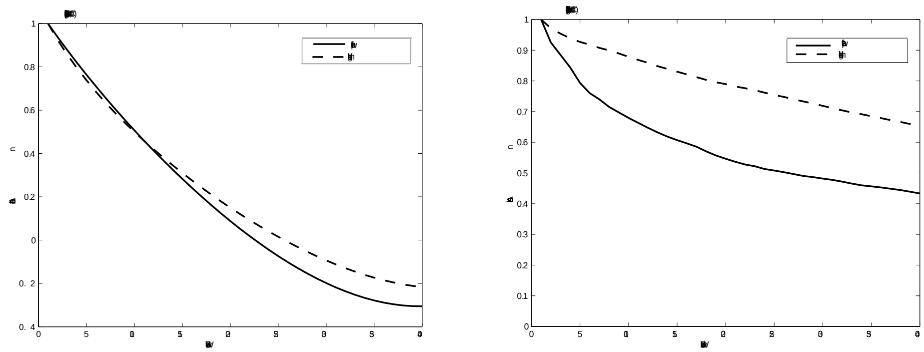

Fig. 6. Autocorrelation as a function of window size for two-part tariff and mixed bundling (N=10, C=10). The left figure uses a path generated by steepest ascent, and the right uses a path generated by a random walk.

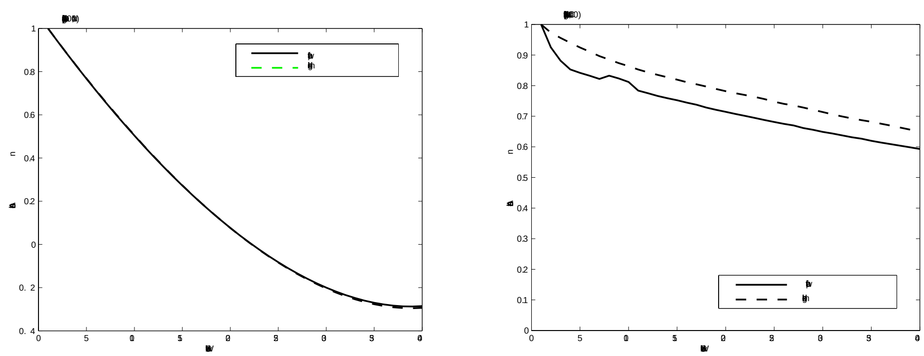

Fig. 7. Autocorrelation as a function of window size for two-part tariff and mixed bundling (N=10, C=100). The left figure uses a path generated by steepest ascent, and the right uses a path generated by a random walk.

Figures 6, 7, 8, and 9 compare autocorrelation over both random walks and optimization paths for $N = \{10, 100\}$ and $C = \{10, 100\}$. From these figures, we can draw several conclusions. First, the significant window size is much smaller when optimizing on either landscape than when performing a random walk; this should not be surprising, since the whole point of optimizing is to change one's state, hopefully in a useful direction. It is interesting that both landscapes produced very similar autocorrelations when optimizing, indicating that the difference in learning difficulty is probably not due to a difference in the ability to effectively reach optima. Instead, our previous conclusion that distribution of optima was more uniform for mixed bundling (meaning also that it is more difficult to move between optima) gains credence.

Second, we note that, for random walks, mixed bundling shows little change as N and C are varied, while two-part tariff improves when either N or C are

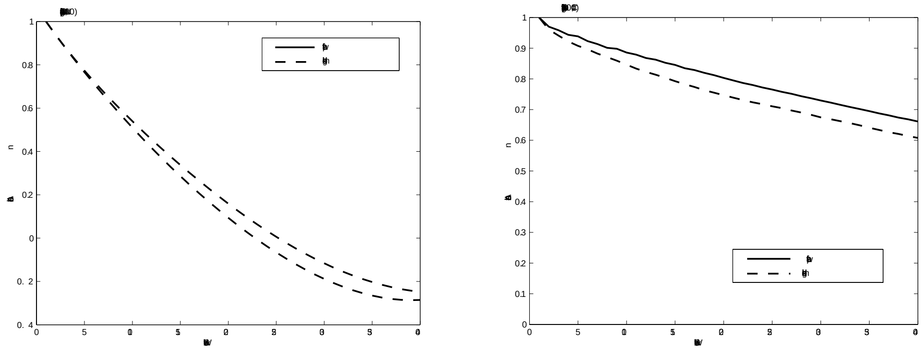

Fig. 8. Autocorrelation as a function of window size for two-part tariff and mixed bundling (N=100, C=10). The left figure uses a path generated by steepest ascent, and the right uses a path generated by a random walk.

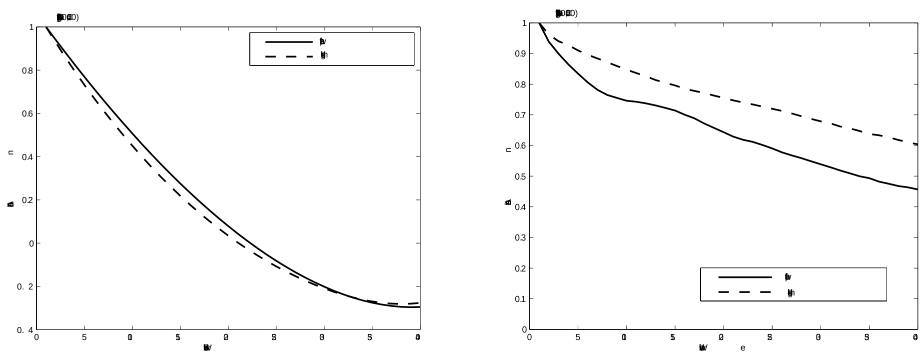

Fig. 9. Autocorrelation as a function of window size for two-part tariff and mixed bundling (N=100, C=100). The left figure uses a path generated by steepest ascent, and the right uses a path generated by a random walk.

increased. We expected increasing either variable to improve autocorrelation, since increasing C appears to reduce the size of discontinuities in the landscape. It is unclear why mixed bundling does not show the same improvement; however, we can conclude that it is less resistant to a change in these parameters.

We also can see that autocorrelation is significantly higher for mixed bundling than it is for two-part tariff, although only for random walks. It is very similar when optimizing. We conjecture that the reason for this is due to the large plateaus seen in the mixed bundling landscape; a random walk along a plateau will have an autocorrelation of 1.

Finally, we note that the effective window size actually decreases for two-part tariff when both N and C are 100 over the case where one variable is 10 and the other 100. This may be due to a 'stretching' of the landscape as N is increased; the effective range of the fee parameter of two-part tariff grows with N.

In summary, these experiments help us to understand quantitatively what we were previously able to explain only through a reliance on pictures and an appeal to metaphors. There clearly seems to be a correlation between optima distribution and learning difficulty with regard to two-part tariff and mixed bundling. There is also some evidence that increasing C and N reduces landscape ruggedness, although not necessarily in a way that affects learning performance for an algorithm such as amoeba.

5 Conclusions and future work.

In this article, we have described the problem of learning in an environment where cumulative reward is the measure of performance and stressed the need for an adaptive agent to consider what it chooses to learn as a way of optimizing its total reward. We have argued that landscapes are a useful representation for an agent's learning problem and applied the analysis of landscapes to a particular learning problem, that of learning price schedules in an information economy. We showed that two metrics, distribution of optima and autocorrelation, can be calculated and used as estimates of ruggedness, and further used these metrics to explain results in our previous work. By using these metrics to estimate the difficulty of different landscapes, an adaptive agent can thereby make a more informed decision as to which learning problem it will choose to solve.

There are many possible directions for future research. One particular avenue included the extension of this analysis to learning in nonstationary environments. That is, where the landscape an agent is adapting to changes over time. In this case, we would like to characterize how this change affects the difficulty of the agent's learning problem. In particular, we are interested in problems where one agent's learning affects the learning problem of another agent, and providing tools by which agents can minimize their impact on each other's learning. Measuring this impact is a necessary step toward solving that problem.

6 Acknowledgments

This work was supported in part by NSF grants IIS-9872057 and IIS-0112669.

References

1. Christopher H. Brooks, Robert S. Gazalle, Rajarshi Das, Jeffrey O. Kephart, Jeffrey K. MacKie-Mason, and Edmund H. Durfee. Model selection in an information economy: Choosing what to learn. *Computational Intelligence*, 2002. to appear.
2. Josselin Garnier and Leila Kallel. How to detect all attraction basins of a function. In *Theoretical Aspects of Evolutionary Computation*, pages 343–365. Springer-Verlag, 2000.
3. Wim Hordjik. A measure of landscapes. *Evolutionary Computation*, 4(4):336–366, 1996.

4. Terry Jones. *Evolutionary Algorithms, Fitness Landscapes, and Search.* PhD thesis, University of New Mexico, May 1995.
5. Terry Jones and Stephanie Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In Larry Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 184–192, San Francisco, 1995. Morgan Kauffman.
6. L. Kallel, B. Naudts, and C.R. Reeves. Properties of fitness functions and search landscapes. In L. Kallel, B. Naudts, and A. Rogers, editors, *Theoretical Aspects of Evolutionary Computation*, pages 175–206. Springer-Verlag, 2000.
7. Stuart Kauffman. *Origins of Order: Self-organization and Selection in Evolution.* Oxford University Press, New York, 1993.
8. William H. Press et al. *Numerical Recipes.* Cambridge University Press, 1992.
9. Sebastian Thrun. The role of exploration in learning control. In *Handbook of Intelligent Control: Neural, Fuzzy and Adaptive Systems.* Van Nostrand Reinhold, Florence, Kentucky, 1992.
10. E. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 63:325–336, 1990.
11. Sewall Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings of the 6th Congress on Genetics*, page 356, 1932.