Efficient Algorithms for Universal Portfolios

Adam Kalai*
CMU
Department of Computer Science
akalai@cs.cmu.edu

Santosh Vempala†
MIT
Department of Mathematics and
Laboratory for Computer Science
vempala@math.mit.edu

Abstract

A constant rebalanced portfolio is an investment strategy which keeps the same distribution of wealth among a set of stocks from day to day. There has been much work on Cover's Universal algorithm, which is competitive with the best constant rebalanced portfolio determined in hindsight [3, 9, 2, 8, 14, 4, 5, 6]. While this algorithm has good performance guarantees, all known implementations are exponential in the number of stocks, restricting the number of stocks used in experiments [9, 4, 2, 5, 6]. We present an efficient implementation of the Universal algorithm that is based on non-uniform random walks that are rapidly mixing [1, 12, 7]. This same implementation also works for non-financial applications of the Universal algorithm, such as data compression [6] and language modeling [10].

1. Introduction

A constant rebalanced portfolio (CRP) is an investment strategy which keeps the same distribution of wealth among a set of stocks from day to day. That is, the proportion of total wealth in a given stock is the same at the beginning of each day. Recently there has been work on on-line investment strategies which are competitive with the best CRP determined in hindsight [3, 9, 2, 8, 14, 4, 5, 6]. Specifically, the daily performance of these algorithms on a market approaches that of the best CRP for that market, chosen in hindsight, as the lengths of these markets increase without bound.

As an example of a useful CRP, consider the following market with just two stocks [9, 5]. The price of one stock remains constant, and the price of the other stock alternately halves and doubles. Investing in a single stock will not increase the wealth by more than a factor of two. However, a $(\frac{1}{2}, \frac{1}{2})$ CRP will increase its wealth exponentially. At the end of each day it trades stock so that it has an equal worth in each stock. On alternate days the total value will change by a factor of $\frac{1}{2}(1) + \frac{1}{2}(\frac{1}{2}) = \frac{3}{4}$ and $\frac{1}{2}(1) + \frac{1}{2}(2) = \frac{3}{2}$, thus increasing total worth by a factor of $9/8$ every two days.

The main contribution of this paper is an efficient implementation of Cover's UNIVERSAL algorithm for portfolios [3], which Cover and Ordentlich [4] show that, in a market with n stocks, over t days,

$$\frac{\text{performance of UNIVERSAL}}{\text{performance of best CRP}} \geq \frac{1}{(t+1)^{n-1}}.$$

By performance, we mean the return per dollar on an investment. The above ratio is a decreasing function of t. However, the average per-day ratio, $(1/(t+1)^{n-1})^{1/t}$, increases to 1 as t increases without bound. For example, if the best CRP makes one and a half times as much as we do over a day of 22 years, it is only making a factor of $1.5^{1/22} \approx 1.02$ as much as we do per year. In this paper, we do not consider the Dirichelet$(1/2, \ldots, 1/2)$ UNIVERSAL [4] which has the better guaranteed ratio of $2\sqrt{1/(t+1)^{n-1}}$.

All previous implementations of Cover's algorithm are exponential in the number of stocks with run times of $O(t^{n-1})$. Blum and Kalai have suggested a randomized approximation based on sampling portfolios from the uniform distribution [2]. However, in the worst case, to have a high probability of performing almost as well as UNIVERSAL, they require $O(t^{n-1})$ samples. We show that by sampling portfolios from a non-uniform distribution, only polynomially many samples are required to have a high probability of performing nearly as well as UNIVERSAL. This non-uniform sampling can be achieved by random walks on the simplex of portfolios.

2. Notation and Definitions

A price relative for a given stock is the nonnegative ratio of closing price to opening price during a given day. If the

*Supported by an IBM Distinguished Graduate Fellowship.
†Supported by NSF CAREER award CCR-9875024.

market has n stocks and trading takes place during T days, then the market's performance can be expressed by T price relative vectors, $(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T), \vec{x}_i \in \Re^m_+$, where x^j_i is the nonnegative price relative of the jth stock for the ith day.

A portfolio is simply a distribution of wealth among the stocks. The set of portfolios is the $(n-1)$-dimensional simplex,

$$\Delta = \{\vec{b} \in \Re^n \,|\, \sum_{j=1}^{N} b^j = 1 \wedge b^j \geq 0\}.$$

The CRP investment strategy for a particular portfolio $\vec{b}$, CRP$_{\vec{b}}$, redistributes its wealth at the end of each day so that the proportion of money in the jth stock is b^j. An investment using a portfolio $\vec{b}$ during a day with price relatives $\vec{x}$ increases one's wealth by a factor of $\vec{b} \cdot \vec{x} = \sum b^j x^j$. Therefore, over t days, the wealth achieved by CRP$_{\vec{b}}$ is,

$$P_t(\vec{b}) = \prod_{i=1}^{t} \vec{b} \cdot \vec{x}_i.$$

Finally, we let μ be the uniform distribution on Δ.

3. Universal portfolios

Before we define the universal portfolio, suppose you just want a strategy that is competitive with respect to the best single stock. In other words, you want to maximize the worst-case ratio of your wealth to that of the best stock. In this case, a good strategy is simply to divide your money among the n stocks and let it sit. You will always have at least $\frac{1}{n}$ times as much money as the best stock. Note that this deterministic strategy achieves the expected wealth of the randomized strategy that just places all its money in a random stock.

Now consider the problem of competing with the best CRP. Cover's universal portfolio algorithm is similar to the above. It splits its money evenly among all CRPs and lets it sit in these CRP strategies. (It does not transfer money between the strategies.) Likewise, it always achieves the expected wealth of the randomized strategy which invests all its money in a random CRP. In particular, the bookkeeping works as follows:

Definition 1 (UNIVERSAL) *The universal portfolio algorithm at time t has portfolio $\vec{u}_t$, which for stock j is, on the first day $u^j_0 = 1/n$, and on the end of the tth day,*

$$u^j_t = \frac{\int_\Delta v^j P_t(\vec{v}) d\mu(\vec{v})}{\int_\Delta P_t(\vec{v}) d\mu(\vec{v})}, \; i = 1, 2, \ldots$$

(Recall that μ is the uniform distribution over the $(n-1)$-dimensional simplex of portfolios, Δ.)

This is the form in which Cover defines the algorithm. He also notes [4] that UNIVERSAL achieves the average performance of all CRPs, i.e.,

$$\text{Perf. of UNIVERSAL} = \prod_{t=1}^{T} \vec{u}_{t-1} \cdot \vec{x}_t = \int_\Delta P_T(\vec{v}) d\mu(\vec{v})$$

4. An efficient algorithm

Unfortunately, the straightforward method of evaluating the integral in the definition of UNIVERSAL takes time exponential in the number of stocks. Since UNIVERSAL is really just an average of CRP's, it is natural to approximate the portfolio by sampling [2]. However, with uniform sampling, one needs $O(t^{n-1})$ samples in order to have a high probability of performing as well as UNIVERSAL, which is still exponential in the number of stocks. Here we show that, with non-uniform sampling, we can approximate the portfolio efficiently. With high probability ($1-\delta$), we can achieve performance of at least $(1-\epsilon)$ times the performance of UNIVERSAL. The algorithm is polynomial in $1/\epsilon$, $\log(1/\delta)$, n (the number of stocks), and T (the number of days).

The key to our algorithm is sampling according to a biased distribution. Instead of sampling according to μ, the uniform distribution on Δ, we sample according to ν_t, which weights portfolios in proportion to their performance, i.e.,

$$d\nu_t(\vec{b}) = \frac{P_t(\vec{b})}{\int_\Delta P_t(\vec{v}) d\mu(\vec{v})}$$

In the next section, we show how to efficiently sample from this biased distribution.

UNIVERSAL can be thought of as computing each component of the portfolio by taking the expectation of draws from ν_t, i.e.,

$$u^j_t = \int_\Delta v^j \, d\nu_t(\vec{v}) = \mathrm{E}_{\vec{v} \in \nu_t}\left[v^j\right] \tag{1}$$

Thus our sampling implementation of UNIVERSAL averages draws from ν_t:

Definition 2 *(**Universal biased sampler**) The Universal biased sampler, with m samples, on the end of day t chooses a portfolio $\vec{a}_t$ as the average of m portfolios drawn independently from ν_t.*

Now, we apply Chernoff bounds to show that with high probability, for each j, a^j_t closely approximates u^j_t. In order to ensure that this biased sampling will get us a^j_t/u^j_t close to 1, we need to ensure that u^j_t isn't too small:

Lemma 1 *For all $j \leq n$ and $t \leq T$, $u^j_t \geq 1/(n+t)$.*

Proof. WLOG let $j = 1$ and $x_1^1 = x_2^1 = \ldots x_t^1 = 0$, because this makes u_t^1 smallest. Now, u_t^1 is a random variable between 0 and 1 (see (1)), and the expectation of a random variable $0 \le X \le 1$ is $E[X] = \int_0^1 \text{Prob}(X \ge z)dz$. Thus,

$$u_t^1 = \mathrm{E}_{\vec{v}\in\nu_t}\left[v^1\right] = \int_0^1 \nu_t\left(\{\vec{v}|v^1 \ge z\}\right)dz.$$

Furthermore, $\{\vec{v}|v^1 \ge z\} = (z, 0, \ldots, 0) + (1-z)\Delta$, is a shrunken simplex of volume $(1-z)^{n-1}$ times the volume of Δ, since Δ has dimension $n-1$. The average performance of portfolios in this set is $(1-z)^t$ times the average over Δ, because for each of t days, a portfolio in this set $(z, 0, \ldots, 0) + (1-z)\vec{b}$ performs $(1-z)$ as well as the corresponding portfolio $\vec{b} \in \Delta$. So the probability of this set under ν_t is $(1-z)^{n-1}(1-z)^t$ and,

$$u_t^1 = \int_0^1 (1-z)^{n-1}(1-z)^t dz = 1/(n+t).$$

□

Combining this lemma with Chernoff bounds, we get:

Theorem 1 *With $m = 2T^2(n+T)\log(nT/\delta)/\epsilon^2$ samples, the Universal biased sampler performs at least $(1-\epsilon)$ as well as Universal, with probability at least $1-\delta$.*

Proof. Say each u_t^j is approximated by a_t^j. Furthermore, suppose each $a_t^j \ge u_t^j(1-\epsilon/T)$. Then, on any individual day, the performance of the $\vec{a}_t$ is at least $(1-\epsilon/T)$ times as good as the performance of $\vec{u}_t$. Thus, over T days, our approximation's performance must be at least $(1-\epsilon/T)^T \ge 1-\epsilon$ times the performance of UNIVERSAL.

The multiplicative Chernoff bound for approximating a random variable $0 \le X \le 1$, with mean $\bar{X}$, by the sum S of m independent draws is,

$$\Pr\left[S < (1-\alpha)\bar{X}m\right] \le e^{-m\bar{X}\alpha^2/2}.$$

In our case, we are approximating each u_t^j by m samples, our lemma shows that the expectation of $u_t^j = X$ is $\bar{X} \ge 1/(j+t) \ge 1/(n+T)$, and we want to be within $\alpha = \epsilon/T$. Since this must hold for nT different u_t^j's, it suffices for,

$$e^{-m\epsilon^2/(2T^2(n+T))} \le \frac{\delta}{nT},$$

which holds for the number of samples m chosen in the theorem. □

The biased sampler will actually sample from a distribution that is close to ν_t, call it p_t, with the property that

$$\int_\Delta |\nu_t(b) - p_t(b)|db \le \epsilon_0$$

for any desired $\epsilon_0 > 0$ in time proportional to $\log\frac{1}{\epsilon_0}$. It is not hard to verify that the estimates from p_t and ν_t differ by at most a factor of $(1+\epsilon_0)$. By applying Chernoff bounds as described above the Universal biased sampler performs at least $(1-\epsilon)(1-\epsilon_0)$ as well as Universal (note that ϵ_0 is exponentially small).

5. The biased sampler

In this section we describe a random walk for sampling from the simplex with probability density proportional to

$$f(\vec{b}) = P_t(\vec{b}) = \prod_{i=1}^{t} \vec{b}\cdot\vec{x}_i.$$

Before we do this, note that sampling from the uniform distribution over the simplex is easy: pick $n-1$ reals $x_1, \ldots, x_{n-1}$ uniformly at random between 0 and 1 and sort them into $y_1 \le \ldots \le y_{n-1}$; then the vector $(y_1, y_2 - y_1, \ldots, y_{n-1} - y_{n-2}, 1 - y_{n-1})$ is uniformly distributed on the simplex.

There is another (less efficient) way. Start at some point x in the simplex. Pick a random point y within a small distance δ of x. If y is also in the simplex, then move to y; if it is not, then try again. The *stationary* distribution of a random walk is the distribution on the points attained as the number of steps tends to infinity. Since this random walk is *symmetric*, i.e. the probability of going from x to y is equal to the probability of going from y to x, the distribution of the point reached after t steps tends to the uniform distribution. In fact, in a polynomial number of steps, one will reach a point whose distribution is nearly uniform on the simplex.

In our case, we have the additional difficulty that the desired distribution is not the uniform distribution. Although the distribution induced by f can be quite different from the uniform density, it has the following nice property.

Lemma 2 *The function $f(\vec{b})$ is log-concave for nonnegative vectors.*

Proof. The function $-\log f$ is convex. The derivative of $\log f$ at $\vec{b}$ is is the vector $f'(\vec{b}) = (\frac{b_1}{f(\vec{b})}, \frac{b_2}{f(\vec{b})}, \ldots, \frac{b_n}{f(\vec{b})})$. The matrix F'' of second derivatives has i,jth entry $-\frac{b_i b_j}{f(\vec{b})^2}$. Thus $F'' = -f'^T f'$ is a negative semidefinite matrix, implying that $\log f$ is a concave function in the positive orthant. □

The symmetric random walk described above can be modified to have any desired target distribution. This is called the Metropolis filter [13], and can be viewed as a combination of the walk with rejection sampling: If the walk is at x and chooses the point y as its next step, then move to y with probability $\min(1, \frac{f(y)}{f(x)})$ and do nothing with the remaining probability (i.e. try again). Lovasz and Simonovits [12] have shown that this random walk is rapidly mixing, i.e. it attains a distribution close to the stationary one in polynomial time.

For our purpose, however, the following discretized random walk has the best provable bounds on the mixing time. First rotate Δ so that it is on the plane $x = 0$ and scale it by a factor of $1/\sqrt{2}$ so that it has unit diameter. We will only walk on the set of points in Δ whose coordinates are multiples of a fixed parameter $\delta > 0$ (to be chosen below), i.e. points on an axis parallel grid whose "unit" length is δ. Any point on this grid has $2n$ neighbors, 2 along each axis.

1. Start at a (uniformly) random grid point in the simplex.
2. Suppose $X(\tau)$ is the location of the walk at time τ.
3. Let y be a random neighbor of $X(\tau)$.
4. If y is in Δ, then move to it, i.e. set $X(\tau+1) = y$ with probability $p = \min(1, \frac{f(y)}{f(x)})$, and stay put with probability $1-p$ (i.e. $X(\tau+1) = X(\tau)$).

Let the set of grid points be denoted by D. We will actually only sample from the set of grid points in Δ that are not too close to the boundary, namely, each coordinate x_i is at least $\frac{\epsilon}{n+t}$ for a small enough ϵ. For convenience we will assume that each coordinate is at least $\frac{1}{2(n+t)}$. Let this set of grid points be denoted by D. Each grid point x can be associated with a unique axis-parallel cube of length δ centered at x. Call this cube $C(x)$. The step length δ is chosen so that for any grid point x, $f(x)$ is close to $f(y)$ for any $y \in C(x)$.

Lemma 3 *If we choose $\delta < \frac{\log 1+\alpha}{2(n+t)t}$ then for any grid point z in D, and any point $y \in C(z)$, we have*

$$(1+\alpha)^{-1} f(z) \le f(y) \le (1+\alpha) f(z).$$

Proof. Since $y \in C(z)$, $\max_j |y_j - z_j| \le \delta$. For any price relative x^*, the ratio $\frac{y \cdot x^*}{z \cdot x^*}$ is at most $\max_j \frac{y_j}{z_j}$. This can be written as

$$\max_j \frac{z_j + (y_j - z_j)}{z_j} = \max_j (1 + \frac{\delta}{z_j})$$

Since each coordinate is at least $\frac{1}{2(n+t)}$ we have that the ratio is at most $(1 + 2\delta(n+t))$. Thus the ratio $\frac{f(y)}{f(z)}$ is at most $(1+2\delta(n+t))^t$ and the lemma follows. □

The stationary distribution π of the random walk will be proportional to $f(z)$ for each grid point z. Thus when viewed as a distribution on the simplex, for any point y in the simplex,

$$\pi(y)(1+\alpha)^{-1} \le d\nu_t(y) \le \pi(y)(1+\alpha)$$

The main issue is how fast the random walk approaches π. We return to the discrete distribution on the grid points. Let the distribution attained by the random walk after τ steps be p_τ, i.e. $p_\tau(x)$ is the probability that the walk is at the grid point x after τ steps. The progress of the random walk can be measured as the distance between its current distribution p_τ and the stationary distribution as follows:

$$||p_\tau - \pi|| = \sum_{x \in D} |p_\tau(x) - \pi(x)|$$

In [7], Frieze and Kannan derive a bound on the convergence of this random walk which can be used to derive the following bound for our situation.

Theorem 2 *After τ steps of the random walk,*

$$||p_\tau - \pi||^2 \le e^{-\frac{\gamma\tau}{nt^2(n+t)^2}} (n+t)^2$$

where $\gamma > 0$ is an absolute constant.

Corollary 1 *For any $\epsilon_0 > 0$, after $O(nt^2(n+t)^2 \log \frac{n+t}{\epsilon_0})$ steps,*

$$||p_\tau - \pi||^2 \le \epsilon_0.$$

Proof (of theorem). Frieze and Kannan prove that

$$2||p_\tau - \pi||^2 \le e^{-\frac{\gamma\theta\tau\delta^2}{nd^2}} \log \frac{1}{\pi_*} + \frac{M\pi_\theta n d^2}{\gamma\delta^2}$$

where $\gamma > 0$ is a constant, d is the diameter of the convex body in which we are running the random walk, π_* is $\min_{x \in D} \pi(x)$, θ is a parameter between 0 and 1, and

$$\pi_\theta = \sum_{x \in D : \frac{vol(C(x) \cap \Delta)}{vol(C(x))} \le \theta} \pi(x).$$

In words, π_θ is the probability of the grid points whose cubes intersect the simplex in less than θ fraction of their volume. The parameter M is defined[1] as $\max_x p_0(x) \log \frac{p_0(x)}{\pi(x)}$, where p_0 is the initial distribution on the states.

For us the diameter d is 1. We will set $\theta = \frac{1}{4}$ and choose δ small enough so that π_θ is a constant. This can be done for example with any $\delta \le \frac{1}{2(n+t)}$. To see this, consider the simplex blown up by a factor of $\frac{1}{\delta}$ i.e. the set $\frac{1}{\delta}\Delta = \{y | y \ge 0, \sum_i y_i = \frac{1}{\delta}\}$. Now the set of points with integer coordinates correspond to the original grid points. Let B be the set of cubes on the border of this set, i.e. the volume of each cube in B that is in $\frac{1}{\delta}\Delta$ is less than $\frac{1}{4}$. Then by

[1] The M we use here differs slightly from the definition in [7], where $M = \max_x \frac{p_0(x)}{\pi(x)} \log \frac{p_0(x)}{\pi(x)}$. However, the theorem holds with either choice of M.

blowing up further by 1 unit, we get a set that contains all these cubes. But the ratio of the volumes is

$$\frac{(\frac{1}{\delta}+1)^n}{(\frac{1}{\delta})^n} = (1+\delta)^n.$$

Also, the performance of these border grid points can only be $(1+\delta)^t$ better than the corresponding (non-blown up) points in the corresponding points. Thus $\pi_\theta \le (1+\delta)^{n+t} < 2$ for $\delta \le \frac{1}{2(n+t)}$.

Thus the bound above on the distance to stationary becomes

$$2||p_\tau - \pi||^2 \le e^{-\frac{\gamma\tau\delta^2}{n}} \log\frac{1}{\pi_*} + \frac{2Mn}{\gamma\delta^2}$$

Next we observe that by our choice of starting point (uniform over the simplex) M is exponentially small. Thus we can ignore the second term in the right hand side. Finally we note that π_* is at least $\delta^n(\frac{\epsilon}{n+t})^t$, which simplifies the inequality to

$$||p_\tau - \pi||^2 \le e^{-\frac{\gamma\tau\delta^2}{n}}(n+t)^2$$

Our choice of δ (= $O(\frac{1}{(n+t)t})$ implies the theorem (with a different γ)). □

5.1. Collecting samples

Although generating the first sample point takes $O^*(nt^2(n+t)^2)$ steps of the random walk, future samples can be collected more efficiently using a trick from [11]. In fact, the position of the random walk can be observed every $O(n^2(n+t)^2)$ steps to obtain sample points with the property that they are pairwise *nearly* independent in the following sense. For any two subsets A, B of the simplex, two sample points u and v satisfy

$$|Pr(u \in A, v \in B) - Pr(u \in A)Pr(v \in B)| \le \epsilon$$

This can be used to reduce the number of samples. We collect completely independent groups of samples, each sample consisting of pairwise independent samples, and then compute the average of the groups.

As an implementation detail, one can do the random walk as follows. Choose an initial point at random on the simplex, as described earlier, and then choosing an individual component at random. Increase (or decrease) that component by δ, and then decrease (or increase) the remaining components by $\delta/(n-1)$. Use the same rejection technique to decide whether to actually take that step in the random walk.

6. Conclusion

We have presented an efficient randomized approximation of the UNIVERSAL algorithm. Not only does the approximation have an expected performance equal to that of UNIVERSAL, but with high probability $(1-\delta)$ it is within $(1-\epsilon)$ times the performance of universal, and runs in time polynomial in $\log\frac{1}{\delta}$, $1/\epsilon$, the number of days, and the number of stocks. With money, it is especially important to achieve this expectation. For example, a 50% chance at 10 million dollars may not be as valuable to most people as a guaranteed 5 million dollars.

While our implementation can be used for applications of UNIVERSAL, such as data compression [6] and language modeling [10], we do not implement it in the case of transaction costs [2] or for the Dirichelet$(1/2,\ldots,1/2)$ UNIVERSAL [4].

References

[1] D. Applegate and R. Kannan Sampling and integration of near log-concave functions. In *Proceedings of the Twenty Third Annual ACM Symposium on Theory of Computing*, 1991.

[2] A. Blum and A. Kalai Universal Portfolios With and Without Transaction Costs. *Machine Learning*, 35:3, pp 193-205, 1999.

[3] Universal portfolios. *Math. Finance*, 1(1):1-29, January 1991.

[4] T.M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2), March 1996.

[5] E. Ordentlich. and T.M. Cover. Online portfolio selection. In *Proceedings of the 9th Annual Conference on Computational Learning Theory*, Desenzano del Garda, Italy, 1996.

[6] T.M. Cover. Universal data compression and portfolio selection. In *Proceedings of the* 37*th IEEE Symposium on Foundations of Computer Science*, pages 534-538, Oct 1996.

[7] A. Frieze and R. Kannan. Log-Sobolev inequalities and sampling from log-concave distributions. Annals of Applied Probability 9, 14-26.

[8] D.P. Foster and R.V. Vohra. Regret in the On-line Decision Problem. *Games and Economic Behavior*, Vol. 29, No. 1/2, pp. 7-35, Nov 1999.

[9] D. Helmbold, R. Schapire, Y. Singer, and M. Warmuth. On-line portfolio selection using multiplicative updates. *Machine Learning: Proceedings of the Thirteenth International Conference*, 1996.

[10] A. Kalai, S. Chen, A. Blum, and R. Rosenfeld. On-line Algorithms for Combining Language Models. In *Proceedings of the International Conference on Accoustics, Speech, and Signal Processing* (ICASSP '99), 1999.

[11] R. Kannan, L. Lovasz and M. Simonovits. Random walks and an $O^*(n^5)$ volume algorithm for convex bodies. Random Structures and Algorithms 11, 1-50.

[12] L. Lovasz and M. Simonovits. On the randomized complexity of volume and diameter. In *Proceedings of the IEEE Symp. on Foundation of Computer Science,*, 1992.

[13] N. Metropolis, Rosenberg, Rosenbluth, Teller and Teller. Equation of state calculation by fast computing machines. In *Journal of Chemical Physics*, 21 (1953), pp 1087-1092.

[14] V. Vovk and C.J.H.C. Watkins. Universal Portfolio Selection. In *Proceedings of the 11th Annual Conference on Computational Learning Theory*, pages 12-23, 1998.