Angewandte Mathematik und
Informatik
Universität zu Köln

Report No. 92.122

On Derandomized Approximation Algorithms

by

Anand Srivastav[1], Peter Stangier

1992

Anand Srivastav

Institut für Informatik
Lehrstuhl Algorithmen und Komplexität
Humboldt Universität zu Berlin
Unter den Linden 6
10099 Berlin
Germany
srivasta@informatik.hu-berlin.de

Peter Stangier

Institut für Informatik

Universität zu Köln
Pohligstr.1
50969 Köln
Germany
stangier@informatik.uni-koeln.de

[1] supported by: Deutsche Forschungs Gemeinschaft

1991 Mathematics Subject Classification: 05B40, 60G42, 68Q25, 68R05
Keywords: Combinatorial discrepancies, randomized algorithm, derandomization, discrete martingales

On Derandomized Approximation Algorithms

Anand Srivastav Peter Stangier

December 1992

Abstract

With the design of powerful randomized algorithms the transformation of a randomized algorithm or probabilistic existence result for combinatorial problems into an efficient deterministic algorithm (called derandomization) became an important issue in algorithmic discrete mathematics. In the last years several interesting examples of derandomization have been published, like discrepancy in hypergraph colouring, packing integer programs and an algorithmic version of the Lovász-Local-Lemma. In this paper the derandomization method of conditional probabilities of Raghavan/Spencer is extended using discrete martingales. As a main result pessimistic estimators are constructed for combinatorial approximation problems involving non-linear objective functions with bounded martingale differences. The theory gives polynomial-time algorithms for the linear and quadratic lattice approximation problem and a quadratic variant of the matrix balancing problem extending results of Spencer, Beck/Fiala and Raghavan. Finally a probabilistic existence result of Erdős on the average graph bisection is transformed into a deterministic algorithm.

1 Introduction

Derandomization, that is the transformation of randomized algorithms or probabilistic existence results into efficient deterministic algorithms, is considered as an important issue in algorithmic discrete mathematics.

The use of advanced concepts from probability theory like discrete martingales, rapidly mixing Markoff chains and Poisson processes in the analysis of combinatorial problems opened a wide range of applications and led in the past five years to several results which are considered as breakthroughs, for example the sharp concentration of the chromatic number of random graphs [33], randomized approximate counting of combinatorial structures[22] [34] or the computation of the volume of convex bodies [15].

On the other hand a theory of derandomization, although successfully developed in some examples, is in its beginning phase.

First principles of derandomization were implicitly introduced by Erdős and Selfridge and explicitely developed by Spencer in his cosine-hyperbolic algorithm

[36]. The basic idea there was the construction of weight functions based on the computation of conditional probabilities. Therefore the efficiency of the method depends on the efficient computation of certain conditional probabilities. Unfortunately in many examples this cannot be done.

Raghavan first overcame this problem by the construction of efficiently computable upper bounds for conditional probabilities, the so called "pessimistic estimators", which mimic the role of conditional probabilities under consideration and established applications to a class of 0-1 integer programs of packing type [31]. Beck [9] gave a derandomized algorithm for a special case of the Lovász-Local-Lemma and interesting parallel counterparts of previously sequential derandomized algorithms have been exhibited by Berger/Rompel [12], Motwani, Naor and Naor [28] and Alon [3] .

One key fact in the conditional probability method is that pessimistic estimators can be constructed, because *linear* objective functions are involved and therefore Černov and Hoeffding type inequalities on the deviation of linear sums of independent random variables from their expectation are available. Obviously this limits the power of the method and the range of applications.

The purpose of this paper is to extend the derandomization method of conditional probabilities to a larger class of combinatorial objective functions, namely those with bounded martingale differences, naturally including linear and quadratic functions and to construct deterministic polynomial-time approximation algorithms for some interesting combinatorial problems.

The key tool in our approach to control non-linear functions are discrete martingales. The application of martingales presented here is related to the methods of Shamir/Spencer [33] and Bollobás [13] who investigated the chromatic number of random graphs. But in this paper we show how to use martingales in the anlysis of a randomized algorithm and for derandomization. The application of the general theory imply the following briefly stated results.

The **Linear Lattice Approximation** problem has been investigated by Spencer [35], Beck and Fiala [10] and Raghavan [31] and is formulated as follows: Given a $n \times r$ matrix $C = (c_{ij})$ with rational entries $0 \leq c_{ij} \leq 1$ for all i, j and a rational vector $p \in [0,1]^r$ the lattice approximation problem is to construct a lattice point $q \in \{0,1\}^r$ such that $||C(p-q)||_\infty$ is small, hence the discrepancies

$$\Delta_i = |\sum_{j=1}^{r} c_{ij}(p_j - q_j)|$$

are small. When the matrix C represents the constraints and the objective of a $0-1$ integer linear program and p is a solution vector of the corresponding linear programming realaxtion, then lattice approximation is exactly the problem of rounding the vector p such that no constraint is violated too much and the objective function value of the rounded $0-1$ vector is close to the real objective function value.

The same question arises in $0-1$ quadratic optimization. Let $c, p \in [0,1]^n$ be rational vectors, D a rational $r \times r$ matrix and C a rational $(n-1) \times r$ matrix with $0 \leq c_{ij} \leq 1$. By the **Quadratic Lattice Approximation** problem we address the problem of finding a lattice point $q \in \{0,1\}^r$ in polynomial-time such that the following conditions hold:

(a) $|c^T(p-q) + p^T Dp - q^T Dq|$ is small

(b) $||C(p-q)||_\infty$ is small.

Without the quadratic term (i.e. $D \equiv 0$) and taking the matrix $[C, c]$ instead of C this is exactly the (linear) lattice approximation problem.

In the linear case Spencer [35] showed the existence of a lattice point with $\Delta_i \leq 6\sqrt{n}$ for all i, while Beck and Fiala [10] gave an algorithm constructing a lattice point with $\Delta_i \leq 2\sqrt{2n \ln 2n}$ for all i. Raghavan [31] gave a derandomized algorithm finding a vector $q \in \{0,1\}^n$ such that $\Delta_i \leq s_i D(s_i, \frac{1}{2n})$, where $s_i = \sum_{j=1}^r c_{ij} p_j$ and $D(s_i, \frac{1}{2n})$ is a function asymptotically better than the Beck-Fiala bound. Raghavan showed in the unweighted case a polynomial-time implementation of his algorithm in the RAM-model, whereas the same problem in the weighted case $0 \leq c_{ij} \leq 1$ remained open. As far as we know no positive results have been discovered for the more difficult quadratic version of the problem.

Let $D = (d_{ij})$. With $d := 2\max_{1 \leq i \leq r} \sum_{j=1}^r |d_{ij}|$ we give for all matrices with not too large trace, i.e. $trace(D) \leq \alpha d\sqrt{n}$, $\alpha \geq 0$, an $O(r^2 n \log n + r^3)$-time algorithm finding a lattice point $q \in \{0,1\}^r$ such that

(a) $|c^T(p-q) + p^T Dp - q^T Dq| \leq 2\sqrt{n \ln 2n} + (3+\alpha)d\sqrt{n}$

(b) $||C(p-q)||_\infty \leq 2\sqrt{n \ln 2n}$.

Especially in the linear weighted case this gives an $O(r^2 n \log n)$-time algorithm finding a lattice point within an improved Beck-Fiala bound.

Furthermore we study an interesting discrepancy problem, which we call the **Dependently Balancing Matrix** problem. This problem is formally similar to the matrix balancing problem posed by Moser and solved by Beck and Spencer [11], but is mathematically different. Given a $n \times n$-matrix $A = (a_{ij})$ with $-1 \leq a_{ij} \leq 1$, $1 \leq i,j \leq n$, the matrix balancing problem is to find row shifts $x = (x_1, \ldots, x_n) \in \{-1,+1\}^n$ and column shifts $y = (y_1, \ldots, y_n) \in \{-1,+1\}^n$ such that the quantity $D(x,y) := |\sum_{i,j=1}^n a_{ij} x_i y_j|$ is small. Here x and y do not depend on each other. Beck and Spencer proved Mosers conjecture and gave a polynomial-time algorithm finding (x,y) with $D(x,y) \leq 2$.

The *dependently* balancing matrix problem is to find $x_i \in \{-1,+1\}$ for all i such that the quadratic form $D(x,x) = |\sum_{i,j=1}^n a_{ij} x_i x_j|$ is minimal. If $a_{ii} = 0$ for all i, $D(x,x)$ may be smaller than $O(n)$, but finding such small discrepancies seems to be a hard problem.

With $d = \max\limits_{1\leq i\leq n} 2\sum\limits_{j=1}^{n}(|a_{ij}| + |a_{ji}|)$ we have an $O(n^3)$-time algorithm finding a

$x \in \{-1,+1\}^n$ such that $D(x,x) \leq 2d\sqrt{n}$. For small d, i.e $d = O(n^{\frac{1}{2}-\alpha})$ where $0 < \alpha \leq \frac{1}{2}$ or $d = O(\log n)$, this is asymptotically better than the greedy bound $2n$.

Finally we consider a problem about average cuts in graphs: Let $G = (V, E)$ be a graph $|V| = 2n$. The objective in the **Graph Bisection** problem is to find a partition of V into equal sized sets $A, B \subseteq V$ ($|A| = |B| = n$) with minimal cut $c(A, B)$. By probabilistic arguments Erdős [17] showed the existence of a bisection $A, B \subseteq V$, $|A| = |B| = n$ with cut value less than $(1 + o(1))\frac{|E|}{2}$.

For dense graphs ($|E| = \Omega(n^{\frac{3}{2}+\alpha})$, $0 < \alpha \leq \frac{1}{2}$) derandomization transfers this existence result into a deterministic $O(n^3)$-time algorithm.

In the next section we define what we mean by derandomization more formally. In the third section pessimistic estimators for combinatorial functions with bounded martingale differences are constructed and the main theorem is proved. In the last section we give the applications mentioned above.

2 The Derandomization Problem

The model of computation troughout this paper is the RAM-model (see [26]). In the RAM-model an algorithm runs in polynomial-time, if the number of elementary arithmetic operations (briefly called running time) is bounded in a polynomial in the number of numbers of the input and the sum of the encoding lengths of numbers appearing during the execution of the algorithm (briefly called space) is polynomially bounded in the input size. Such a polynomial-time algorithm is also a polynomial-time algorithm in the usual Turing machine model (see [20]). Such an algorithm, where the encoding length does not affect the number of arithmetic operations, is often called a strongly polynomial algorithm. Let *log* denote the binary and *ln* the natural logarithm.

Let I denote an instance of a problem with size n, $(\Omega, \mathbb{P})$ a finite probability space and $C(I)$ a rational number. For $\epsilon > 0$ and $0 < \delta < 1$ let $A_{\epsilon,\delta}$ be a randomized algorithm, which for every instance I outputs a (might be empty) set $S \subseteq \Omega$, and a rational number $A(I, S)$. Denote by $\bar{E}(\epsilon)$ an event of the following type: For $C(I) \neq 0$:

$$|A(I,S) - C(I)| > \epsilon C(I) \tag{1}$$

$$A(I,S) - C(I) > \pm\epsilon C(I) \tag{2}$$

$$A(I,S) - C(I) < \pm\epsilon C(I) \tag{3}$$

For $C(I) = 0$:

$$|A(I,S)| > \epsilon \tag{4}$$

$$A(I,S) > \pm\epsilon \tag{5}$$
$$A(I,S) < \pm\epsilon. \tag{6}$$

Let $\bar{E}(\epsilon)$ denote the complementary event, i.e. $|A(I,S) - C(I)| \leq \epsilon C(I)$ etc. In the following we will specify the type of $E(\epsilon)$, if necessary.

Definition 2.1 *$A_{\epsilon,\delta}$ is called a randomized polynomial-time (resp. fully polynomial-time) $\epsilon - \delta -$ approximation algorithm for the event $E(\epsilon)$, if the following conditions are satisfied:*

(i) $\mathbb{P}\left(\bar{E}(\epsilon)\right) \leq \delta < 1$.

(ii) The running time of $A_{\epsilon,\delta}$ is bounded by a polynomial in n and log $\frac{1}{\delta}$ (resp. $n, \frac{1}{\epsilon}$ and log $\frac{1}{\delta}$).

If for all $\epsilon > 0$ and $0 < \delta < \frac{1}{4}$ the algorithm $A_{\epsilon,\delta}$ is a randomized polynomial-time (resp. fully polynomial-time) $\epsilon - \delta -$approximation algorithm, then we call the family $\left(A_{\epsilon,\delta}\right)_{\epsilon,\delta}$ a polynomial-time randomized approximation scheme (PRAS) (resp. a fully polynomial-time randomized approximation scheme (FPRAS)) for $C(I)$.

ϵ is the relative error, δ the confidence parameter and $1 - \delta$ the confidence probability. Definition 2.1 extends Karp's definition of randomized approximation algorithms for counting problems [24]. Note that for counting problems and approximation of type (a) our definition differs from Karp's definition, because we expect in the output of the algorithm besides $A(I,S)$ also a set $S \subseteq \Omega$. But S might be empty, so for counting problems the two definitions are equivalent. The advantage of Definition 2.1 is that sometimes one is not only interested in the approximation value $A(I,S)$, but also in a sample set S on which $A(I,S)$ can be computed, if such a set exists. More important, in combinatorial problems like discrepancies in hypergraph colouring the probabilistic method proves the existence of a non-empty subset S of the sets of all colouring with zero discrepancy ([2], Chapter 12). But since $C(I) \equiv 0$ the only algorithmic interesting problem is to find S or elements of S deterministically. The deterministic counterpart of Definition 2.1 is

Definition 2.2 *Let I be an instance of a problem with size n, Ω a finite set attached to the problem, $C(I)$ a rational number and $\epsilon > 0$. A deterministic algorithm B_ϵ is called a polynomial-time (resp. fully polynomial-time) ϵ-approximation algorithm for the event $E(\epsilon)$ if it outputs a set $S \subseteq \Omega$ and a rational number $A(I,S)$ such that*

(i) $E(\epsilon)$ holds

(ii) The running time of B_ϵ is bounded by a polynomial in n (resp. n and $\frac{1}{\epsilon}$).

If for every $\epsilon > 0$ a polynomial-time (resp. fully polynomial-time) algorithm B_ϵ exists, the family $(B_\epsilon)_\epsilon$ is called a polynomial-time (resp. fully polynomial-time) approximation scheme (PTAS) (resp. (FPTAS)). For optimization problems Definition 1.2 coincides with the defintion of Papadimitriou and Steiglitz ([29], chap.17). Having designed a randomized $\epsilon - \delta-$approximation algorithm, the objective of derandomization is to find $(S, A(I, S))$ in an efficient deterministic way.

Definition 2.3 (Derandomized Approximation) *Let I be a problem instance and $A_{\epsilon,\delta}$ a (fully) polynomial-time randomized $\epsilon - \delta-$approximation algorithm which outputs $(S, A(I, S))$. The derandomization problem is to construct a (fully) polynomial-time $\epsilon-$approximation algorithm B_ϵ finding $(S, A(I, S))$.*

The parameters characterising the input size of our derandomization problem are n, a positive integer L which is the maximal encoding length of rational numbers needed to implement a procedure for the computation of the rational numbers $C(I)$ and $A(I, S)$, the encoding length of the error probability $\log \frac{1}{\delta}$ (and in case of fully polynomial-time approximations also $\frac{1}{\epsilon}$). Usually L is the encoding length of rational matrices or vectors.

The following examples show the complexity status of Derandomized Approximation ranging from polynomial-time solvable to intractable.

Example 1 (Discrepancy of matrices; [2], Chapter 15, Theorem 1.2)
Let $(a_{ij})_{1\leq i,j\leq n}$ be a rational $n \times n$ matrix of reals, where $-1 \leq a_{ij} \leq 1$ for all i, j. The problem is to find signs $x_1, ..., x_n \in \{-1, +1\}$ such that for every i, $|\sum_{j=1}^{n} a_{ij}x_j| \leq \sqrt{2n \ln 2n}$. With $\Omega = \{-1, +1\}^n$, $\mathbb{P}(\omega) = 2^{-n}$, $\epsilon = \sqrt{2n \ln 2n}$ and $\delta = \mathbb{P}(\exists\, i \text{ s.t. } |\sum_{j=1}^{n} a_{ij}x_j| > \epsilon)$ Theorem 1.1, Chapter 12 of [2] shows that $\delta < 1$, hence with $C(I) = 0$ the above procedure is a fully polynomial-time randomized $\epsilon - \delta-$ approximation algorithm and Theorem 1.2, Chapter 15 of [2] solves the derandomization problem for this example.

Example 2 (Two terminal global routing, see [23], [30])
This example represents a class of randomized approximation algorithms, where the probability distribution is not uniform over the (finite) solution set, but drawn from the solution of a linear program.

The problem of finding a global routing of VLSI-circuits in gate-array designs for two terminal nets is stated as follows: Given a rectilinear $n \times n$ grid (where grid-nodes represent an ensemble of circuit elements and the grid-edges are channels for wiring), a collection of two terminal nets $N = \{N_1, ..., N_r\}$ and for each net N_i two possible paths P_1^i and P_2^i, the task is to choose for each net exactly one path minimizing the channel width that is the maximal number of paths using an grid edge over all grid edges. The problem has been studied by Karp et al. [23] and is known to be NP-hard.

It is easily formulated as an integer linear program. Let W_R be the minimal fractional channel width that is the solution of the corresponding LP-relaxation. Raghavan [30] showed that randomized rounding is for any $0 < \delta < 1$ and $\epsilon = \sqrt{3 W_R \ln \frac{2n(n-1)}{\delta}}$ a fully polynomial-time randomized $\epsilon - \delta -$algorithm finding an integral channel width W such that $0 \leq W - W_R \leq \epsilon$ with probability at least $1 - \delta$, provided that $W_R \geq 3 \ln(2n(n-1)\delta^{-1})$. Again this algorithm can be derandomized [31].

Example 3 (Approximation of the permanent [22])
Given a graph $G = (V, E)$ the exact calculation of its permanent is known to be $\#P$-complete, whereas Jerrum and Sinclair established a FPRAS for this problem. Whether their algorithm can be derandomized or not is unknown.

Example 4 (Max-Cut)
Given a graph $G = (V, E)$, $|V| = n$, the Max-Cut problem is to find a cut with maximal number of edges. Let m_{opt} be the values of a maximal cut, Ω be the set of all cuts and $\delta = 1 - \frac{1}{|\Omega|}$, $C(I) = m_{opt}$ and $\epsilon = \frac{1}{1+m_{opt}}$. When every cut is equiprobable, picking a cut randomly builds trivially a fully polynomial-time randomized $\epsilon - \delta -$approximation algorithm. But for this special choice of ϵ derandomization is equivalent to the determination of a maximal cut, which is NP-hard [19].

Remark: Karps definition of a randomized approximation algorithm requires $\delta \leq \frac{1}{4}$. This would exclude trivial randomized algorithms as considered in the Max-Cut example. But for the purpose of derandomization such a restriction is not justified, because there are randomized algorithms having only exponentially small confidence probability, for example Becks algorithmic version of Lovász-Local-Lemma [9], which can be derandomized.

The last example shows that sometimes Derandomized Approximation is intractable.

Example 5
The problem of finding the exact value of the volume of a convex body, given by a membership oracle, is $\#P-$complete. Dyer, Frieze and Kannan [15] established a FPRAS for this problem. But the problem of Derandomized Approximation is intractable according to results of Elekes [16] and Bárány and Füredi [7].

Known problems where randomized approximation algorithms have been derandomized are of the type of Example 1 and 2. In such problems linear combinatorial functions are involved and therefore Černov and Hoeffding type estimates on the tail of the Binomial distribution are sufficient for the construction of pessimistic estimators.

But even in the case of linear objective functions with rational coefficients it remained an unsolved problem how to construct efficient computable pessimistic estimators in the RAM model of computation ([31], p. 138).

In the next section we will extend the method of conditional probabilities to cover derandomization problems with "computable" functions with bounded martingale differences.

3 Pessimistic Estimators and Martingales

Many approximation problems can be formulated as the problem of approximating the expectation of a certain combinatorial function over a discrete finite set. We consider such problems and work for simplicity on discrete sequence spaces. Let $m, n \in \mathbb{N}$ and $\Omega = \{0, 1, ..., m-1\}^n$ the set of all vectors of length n with entries from $\{0, 1, ..., m-1\}$. Take the powerset $\mathcal{P}(\Omega)$ over Ω as the $\sigma-$algebra and let $\mathbb{P}$ be a probability measure on Ω. Denote by $[n]$ and $[m]_0$ the sets $\{1, \ldots, n\}$ and $\{0, 1, \ldots, m-1\}$. Let $E \subseteq \Omega$ be an event such that for a $0 < \delta < 1$, $\mathbb{P}(\bar{E}) \leq \delta$, where $\bar{E}$ is the complement of E. Our task is to find an $\omega \in E$. We partition Ω into two classes calling a point $\omega \in \Omega$ good, if $\omega \in E$ and calling it bad, if $\omega \notin E$. Let us briefly review Spencers method of conditional probabilities for the construction of a good $\omega \in \Omega$. Following Raghavan [31] Spencers idea can be explained in a suggestive way as a walk on a rooted $m-$ary tree $T(m, n)$. The inner nodes on the $i-$th level $(1 \leq i \leq n)$ of $T(m, n)$ represent the setting of ω_i to $0, ..., m-1$ while the leafs correspond exactly to the points of Ω. During the walk the entries of the output vector ω are choosen from $0, ..., m-1$: Let $\mathbb{P}(\bar{E}|\ \omega_i)$ denote the conditional probability, that a bad event will occur when the i-th entry is choosen as ω_i. A pre-deterministic algorithm is the following recursively defined procedure:

Definition 3.1 Algorithm WALK($\bar{E}$)

(a) Initial Step $(i = 1)$
Compute $\omega_1 \in [m]_0$ such that $\mathbb{P}(\bar{E}|\omega_1) = \min_{0 \leq j \leq m-1} \mathbb{P}(\bar{E}|j)$

(b) Induction Step $(i = 2, .., n)$
Compute ω_i with $\mathbb{P}(\bar{E}|\omega_1, \ldots, \omega_{i-1}, \omega_i) = \min_{0 \leq j \leq m-1} \mathbb{P}(\bar{E}|\omega_1, \ldots, \omega_{i-1}, j)$

(c) Output the vector $\omega = (\omega_1, .., \omega_n)$

The straight forward proved and for all applications of the method of conditional probabilites striking observation is that $\omega \in E$, hence ω is a good point of Ω (see [36] or [31]).

The algorithmic interesting and essential question is whether the conditional probabilities $\mathbb{P}(\bar{E}|\ \omega_1, \ldots, \omega_{i-1})$ can be computed in polynomial-time or not. Unfortunately this seems not to be possible, even in simple cases.

Raghavan suggested a method to remove this difficulty. His idea is to construct upper bounds on the conditional probabilities, which play the same role as the conditional probabilities, but are efficiently computable, the so called "pessimistic estimators" [31]. The following definition formalizes the concept of "pessimistic estimators" and can be considered as an extension of Raghavans definition from the case $m = 2$ to general $m \in \mathbb{N}$.

Let $i \in [n]$, $\omega_1, \ldots, \omega_{i-1} \in [m]_0$,

$$C_{ij} = \{\omega' \in \Omega; \omega'_k = \omega_k \text{ } for \text{ } k = 1, \ldots, i-1 \text{ } and \text{ } \omega'_i = j\}$$

and

$$C_i = \{\omega' \in \Omega; \omega'_k = \omega_k \text{ } for \text{ } k = 1, \ldots, i-1\}.$$

Define

$$\mu_{ij}(\omega_1, \ldots, \omega_{i-1}) = \frac{\mathbb{P}(C_{ij})}{\mathbb{P}(C_i)}.$$

Definition 3.2 *Let $U = \{U_{ij}(\omega_1, \ldots, \omega_{i-1});\ i \in [n],\ j \in [m]_0, |\ \omega_1, \ldots, \omega_{i-1} \in [m]_0\}$ be a family of real-valued functions with the convention that U_{1j} denote the j-th function on the first level. The family U is called a pessimistic estimator (resp. weak pessimistic estimator) for the event $\bar{E}$, if for each $i \in [n]$ and $\omega_1, \ldots, \omega_{i-1} \in [m]_0$, the following conditions (i) - (iv) (resp. (i) - (iii)) are satisfied:*

(i) $\mathbb{P}(\bar{E}|\ \omega_1, \ldots, \omega_{i-1}, \omega_i = j) \leq U_{ij}(\omega_1, \ldots, \omega_{i-1})$

(ii) U is $\mathbb{P}$-convex, that means

$$\sum_{j=0}^{m-1} \mu_{ij}(\omega_1, \ldots, \omega_{i-1}) U_{ij}(\omega_1, \ldots, \omega_{i-1}) \leq U_{i,\omega_{i-1}}(\omega_1, \ldots, \omega_{i-2}).$$

(iii) $\min_{0 \leq j \leq m-1} U_{1j} \leq \delta < 1$

(iv) U is computable in polynomial-time, that means each function

$$U_{ij}(\omega_1, \ldots, \omega_{i-1})$$

can be computed in the RAM model of computation in time bounded by a polynomial in n, m and $\log \frac{1}{\delta}$.

Taking the pessimistic estimator instead of the conditional probabilities in the WALK algorithm we obtain indeed a polynomial-time algorithm.

Definition 3.3 *Let $E \subset \Omega$ be an event with $\mathbb{P}(\bar{E}) \leq \delta < 1$ and let U be a pessimistic estimator for $\bar{E}$. Then let $D - WALK(\bar{E})$ the algorithm defined as in Definition 3.1, but where the conditional probabilities have been replaced by the corresponding functions of the pessimistic estimator U.*

Combination of pessimistic estimators of different events are nothing else than sums of the corresponding families and $\mathbb{P}$-convexity implies

Proposition 3.4 *Let $E_i \subset \Omega$, $i = 1, \ldots, l$, be events with $\mathbb{P}(\bar{E}_i) \leq \delta_i < 1$ and $\delta_1 + \ldots + \delta_l < 1$. Let $U^{(i)}$ be a pessimistic estimators for $\bar{E}_i$ and let $U = U^{(1)} + \ldots + U^{(l)}$. Then U is a pessimistic estimator for the event $\bar{E}_1 \vee \ldots \vee \bar{E}_l$.*

Our special probabilistic framework is a n-product of Bernoulli trials and the entries of each $\omega \in \Omega$ are outcomes of n independently casted dices with m-faces, where the i-th face of the j-th dice occurs with probability p_{ij}, $0 \leq p_{ij} \leq 1$, $i \in [n], j \in [m]_0$. Define a probability measure on Ω by

$$\mathbb{P}(\{\omega\}) = \prod_{i=1}^{n} p_{i\omega_i}, \quad (\omega \in \Omega).$$

Then $(\Omega, \mathbb{P})$ is a probability space with the powerset $\mathcal{P}(\Omega)$ of Ω as the sigma algebra.

Let $f : \Omega \to \mathbb{Q}$ be a function and set $C(I) = \mathbb{E}(f)$. Our algorithm, which we call DICE, casts the above defined n-dices independently, hence generate a random point $\omega \in \Omega$ and and outputs $S = \{\omega\}$ along with the value $A(I, S) := f(\omega)$.

We will study under which circumstances $A(I, S) = f(\omega)$ is a good approximation of $C(I) = \mathbb{E}(f)$.

Definition 3.5 *To shorten notation define for $\epsilon > 0$ the events E_a, E_b, E_c*

(i) the above event $E_a(\epsilon)$ as $f(\omega) \geq \mathbb{E}(f) + \epsilon$.

(ii) the below event $E_b(\epsilon)$ as $f(\omega) \leq \mathbb{E}(f) - \epsilon$.

(iii) the concentration event $E_c(\epsilon)$ as $|f(\omega) - \mathbb{E}(f)| \leq \epsilon$.

Remark: If the conditional probabilities $\mathbb{E}(f|\omega_1, \ldots, \omega_i)$ can be computed efficiently, then it is easy to find an $\omega \in \Omega$ with $f(\omega) \leq \mathbb{E}(f)$ or $f(\omega) \geq \mathbb{E}(f)$ by a method similar to the D-WALK algorithm in Definition 3.3: Simply replace $P(\bar{E}|\omega_1, \ldots, \omega_{i-1})$ by $\mathbb{E}(f|\omega_1, \ldots, \omega_{i-1})$ and this gives $\mathbb{E}(f) \geq f(\omega) \geq f(\omega) - \epsilon$ for any $\epsilon > 0$. So the derandomization problem for the events $E_a(\epsilon)$ and $E_b(\epsilon)$ is immediately solved and the consideration of complicated estimates on $P(f < \mathbb{E}(f) - \epsilon|\omega_1 \ldots, \omega_{i-1})$ would be obsolet. But this simple procedure exhibits only a vector ω with $f(\omega) \geq \mathbb{E}(f) - \epsilon$ and *another* vector $\tilde{\omega}$ with $f(\tilde{\omega}) \leq \mathbb{E}(f) + \epsilon$, but in general $\omega \neq \tilde{\omega}$! Hence the algorithm cannot be applied to a combination of above and below events, especially does not find concentrated events. Therefore we are urged to construct pessimistic estimator for the above and below events, because then we will be able to analyse any combination of such events.

For functions f with bounded martingale differences it will turn out that DICE is a randomized approximation algorithm and this fact is nothing but an algorithmic interpretation of the inequality of Azuma [6].

Martingales and the inequality of Azuma swepped into the analysis of combinatorial functions 1987 with the work of Shamir and Spencer and since then has been widely used to discover new results in the theory of random graphs, which yet could not be established without martingales (see Shamir, Spencer [33], Bollobás [13], McDiarmid [27], Rhee, Talagrand [32], and Frieze, Karp, Reed [18]). The most prominent results are probably the theorem of Shamir and Spencer [33] on the sharp concentration of the chromatic number around its expectation and the determination of the expected chromatic number as $(1+o(1))\frac{n}{\log n}$ by Bollobás [13]. We will use martingales and Azumas inequality in a completely different way. Our purpose is to analyse the worst case behaviour of randomized algorithms and to derandomize them via martingales.

To exhibit the combinatorial meaning of Azumas inequality in our context, we follow Shamir and Spencer and generate a filtration of σ-algebras along a sequence of Bernoulli trials. The most important observation, on which all combinatorial applications of Azumas inequality are based, is that the objective function f may change its value passing from one σ-algebra in the filtration to the next one only by a small "local" amount. This can be interpreted as a kind of Lipschitz continuity of f with a small constant ([2], Theorem 4.1). The martingale is constructed as follows:

Say $\omega, \omega' \in \Omega$ are k-equivalent, i.e. $\omega \cong_k \omega'$ if $\omega_j = \omega'_j$ for all $1 \leq j \leq k$. $k-$equivalency defines an equivalence relation on Ω and induces for each k a partition P_k of Ω where $P_n = \{\{\omega\}; \omega \in \Omega\}$ and $P_0 = \{\Omega\}$. Denote by $\mathcal{F}_k$ the $\sigma-$algebra generated by P_k. Then $\{\emptyset, \Omega\} = \mathcal{F}_o \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n = \mathcal{P}(\Omega)$ and $(\mathcal{F}_k)_{k=0}^n$ is a finite filtering of $\sigma-$algebras.

For $f : \Omega \to \mathbb{R}$ denote by $\mathbb{E}(f|\ \mathcal{F}_k)$ the conditional expectation of f with respect to $\mathcal{F}_k$. The sequence $(f_k)_{k=0}^n$ is a Doob's martingale process with $f_0 = \mathbb{E}(f)$ and $f_n = f$. Denote by ϕ_k the martingale differences $\phi_k = f_k - f_{k-1}$ and suppose that there are $d_k \geq 0$ with $||\phi_k||_\infty \leq d_k$ for all k. Let $\Delta := \sum_{k=1}^n d_k^2$, $0 < \delta < 1$ and $\epsilon_i := \sqrt{2\Delta \ln \frac{i}{\delta}}$ for $i = 1, 2$.

Proposition 3.6 (Azuma Inequality [6]) *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, $(\Sigma_k)_{k=1}^n$ a filtering of $\sigma-$algebras with $\Sigma_k \subseteq \Sigma$ for all k. Let $(f_k)_{k=1}^n$ be a martingale with bounded differences $||\phi_k||_\infty \leq d_k$, for all k. Then we have*

(i) $\mathbb{P}(f_n - f_0 \leq -\epsilon_1(\delta)) \leq \exp(-\frac{\epsilon_1^2}{2\Delta}) = \delta$

(ii) $\mathbb{P}(f_n - f_0 \geq \epsilon_1(\delta)) \leq \exp(-\frac{\epsilon_1^2}{2\Delta}) = \delta$

(iii) $\mathbb{P}(|f_n - f_0| \geq \epsilon_2(\delta)) \leq 2\exp(-\frac{\epsilon_2^2}{2\Delta}) = \delta$

Taking $\Sigma_k := \mathcal{F}_k$ and $(\mathcal{F}_k)$ as the filtering defined above Proposition 3.6 proves

Proposition 3.7 *DICE is a randomized fully polynomial-time $\epsilon_1 - \delta -$approximation algorithm for the events $E_a(\epsilon_1)$, $E_b(\epsilon_1)$ and a randomized fully polynomial-time $\epsilon_2 - \delta -$approximation algorithm for the concentration event $E_c(\epsilon_2)$.*

Remark: If $m = 2$ and f is the weighted sum of independent Bernoulli trials, i.e. $f(\omega) = \sum_{i=1}^{n} a_i \omega_i$, $0 \leq a_i \leq 1$, then Černov and Hoeffding type inequalities [27] gives better $\epsilon_i(\delta)$ values than the Azuma inequality. In the forthcoming we will use Azumas inequality having in mind that especially for linear functions all approximation results we will establish can also be proved with slightly better constants. In order to derandomize the algorithm DICE or in other words to find an $\omega \in \Omega$ deterministically such that the events $E_a(\epsilon)$, $E_b(\epsilon)$ or $E_c(\epsilon)$ holds, we define the functions from which the pessimistic estimators are derived.

Definition 3.8 *Let $\Omega = \{0, 1, \dots, m-1\}^n$ and $(\mathcal{F}_k)$ the $\sigma-$algebras generated by the equivalence relation $=_k$ as defined before. Let $f : \Omega \to \mathbb{Q}$ be a function with bounded martingale differences, $\|\mathbb{E}(f|\ \mathcal{F}_k) - E(f|\ \mathcal{F}_{k-1})\|_\infty \leq d_k$ for each $k, d_k \geq 0$. For $i \in [n], j \in [m]_0, \omega_1, \dots, \omega_{i-1} \in [m]_0$ and parameters $\epsilon, t > 0$ define the families of functions $U^{(a)}$, $U^{(b)}$ and $U^{(c)}$ by*

(i) $U_{ij}^{(a)}(\omega_1, \dots, \omega_{i-1}) := e^{-t(\epsilon + \mathbb{E}(f))} e^{\frac{1}{2}t^2(d_{i+1}^2 + \dots + d_n^2)} e^{t\mathbb{E}(f|\omega_1, \dots, \omega_{i-1}, j)}$.

(ii) $U_{ij}^{(b)}(\omega_1, \dots, \omega_{i-1}) := e^{-t(\epsilon - \mathbb{E}(f))} e^{\frac{1}{2}t^2(d_{i+1}^2 + \dots + d_n^2)} e^{-t\mathbb{E}(f|\omega_1, \dots, \omega_{i-1}, j)}$.

(iii) $U_{ij}^{(c)}(\omega_1, \dots, \omega_{i-1}) = (U_{ij}^{(a)} + U_{ij}^{(b)})(\omega_1, \dots, \omega_{i-1})$.

In order to derandomize the algorithm DICE we must show that the families $U^{(a)}$ (resp. $U^{(b)}$, resp. $U^{(c)}$) are pessimistic estimators for the events $\bar{E}_a(\epsilon)$, resp. $\bar{E}_b(\epsilon)$, resp. $\bar{E}_c(\epsilon)$, when the events $\bar{E}_a(\epsilon)$, $\bar{E}_b(\epsilon)$, $\bar{E}_c(\epsilon)$ are $f - \mathbb{E}(f) > \epsilon$, $f - \mathbb{E}(f) < -\epsilon$ and $|f - \mathbb{E}(f)| > \epsilon$.

We first prove for an appropriate choice of the parameters ϵ and t the weak pessimistic estimator property. This will be used to construct pessimistic estimators also in the RAM-model of computation.

Theorem 3.9 *Let $0 < \delta < 1$, $\epsilon_i = \sqrt{2\Delta \ln \frac{i}{\delta}}$ and $t_i = \epsilon_i \Delta^{-1}$, $(i = 1, 2)$. The families $U^{(a)}$, $U^{(b)}$ and $U^{(c)}$ are weak pessimistic estimators for the events $\bar{E}_a(\epsilon_1)$, $\bar{E}_b(\epsilon_1)$ and $\bar{E}_c(\epsilon_2)$.*

In the proof of Theorem 3.9 we need two lemmata.

Lemma 3.10 *For all $t > 0$ and $k \in \{1, \dots, n\}$ we have*

$$\mathbb{E}(e^{t(f_k - f_{k-1})}|\ \mathcal{F}_{k-1}) \leq \exp(\frac{td_k^2}{2}).$$

Lemma 3.10 is an immediate consequence of [33], Lemma 4.

Lemma 3.11 *For $k \in \{1,\dots,n\}$ let $C \in \mathcal{P}_k$ be a partition set. Then we have for all $t > 0$*

$$\mathbb{E}(e^{tf_k}|C) = e^{t\mathbb{E}(f|C)}.$$

Proof. Let $\mathcal{P}_k = \{C_1,\dots,C_l\}$ and let $\mathbf{1}_{C_i}$ the characteristic function of the set C_i. Then $f_k = \sum_{i=1}^{l} \mathbf{1}_{C_i}\mathbb{E}(f|C_i)$.

Hence for $\omega \in C$

$$f_k(\omega) = \mathbb{E}(f|\mathcal{F}_k)(\omega) = \mathbb{E}(f|C),$$

and we have

$$\begin{aligned}\mathbb{E}(e^{tf_k}|C) &= \frac{1}{\mathbb{P}(C)}\sum_{\omega\in C} e^{tf_k(\omega)}\mathbb{P}(\omega)\\ &= \frac{1}{\mathbb{P}(C)}\sum_{\omega\in C} e^{t\mathbb{E}(f|C)}\mathbb{P}(\omega)\\ &= e^{t\mathbb{E}(f|C)}.\end{aligned}$$

□

Proof of Theorem 3.9:
We first consider the event $\bar{E}_a(\epsilon_1)$ which represents "$f - \mathbb{E}(f) > \epsilon_1$". The proofs for the other events are similar.

(i) Let $\epsilon > 0$ be arbitrary. For the upper bound condition we must show for all i and j :

$$\mathbb{P}(\bar{E}_a(\epsilon)|\omega_1,\dots,\omega_{i-1},j) \le U_{ij}^{(a)}(\omega_1,\dots,\omega_{i-1}).$$

Let

$$C_i = \{\omega' \in \Omega;\, \omega'_k = \omega_k \quad \text{for} \quad k = 1,\dots,i-1\}.$$

and

$$C_{ij} = \{\omega' \in \Omega;\, \omega'_k = \omega_k \quad \text{for} \quad k = 1,\dots,i-1 \quad \text{and} \quad \omega'_i = j\}.$$

Then using Lemma 3.10, induction on k and Lemma 3.11 we have for any $t > 0$

$$\begin{aligned}\mathbb{P}(\bar{E}_a(\epsilon)|C_{ij}) &= \mathbb{P}(f - \mathbb{E}(f) > \epsilon|C_{ij})\\ &\le \exp(-t\epsilon)E\left(e^{t(f-\mathbb{E}(f))}|C_{ij}\right)\\ &= \exp(-t\epsilon)E\left[\mathbb{E}(\mathbf{1}_{C_{ij}}e^{t(f-\mathbb{E}(f))}|\mathcal{F}_{n-1})\right]\cdot\mathbb{P}(C_{ij})^{-1}\\ &= \exp(-t\epsilon)E\left[\mathbf{1}_{C_{ij}}e^{t(f_{n-1}-\mathbb{E}(f))}\mathbb{E}(e^{t(f_n-f_{n-1})}|\mathcal{F}_{n-1})\right]\cdot\mathbb{P}(C_{ij})^{-1}\end{aligned}$$

$$\begin{aligned}
&\leq \exp(-t\epsilon)\exp(\frac{1}{2}t^2 d_n^2)\cdot E\left(\mathbf{1}_{C_{ij}}\, e^{t(f_{n-1}-\mathbb{E}(f))}\right)\cdot \mathbb{P}(C_{ij})^{-1}\\
&= \exp(-t\epsilon)\exp(\frac{1}{2}t^2 d_n^2)\cdot E\left(e^{t(f_{n-1}-\mathbb{E}(f)}|C_{ij}\right)\\
&\leq \exp(-t\epsilon)\exp(\frac{1}{2}t^2\sum_{k=i+1}^{n} d_k^2)\cdot E\left(\mathbf{1}_{C_{ij}} e^{t(f_i-\mathbb{E}(f))}\right)\mathbb{P}(C_{ij})^{-1}\\
&= \exp(-t\epsilon)\exp(\frac{t^2}{2}\sum_{k=i+1}^{n} d_k^2)\exp(t\mathbb{E}(f|C_{ij}))\exp(-t\mathbb{E}(f))\\
&= U_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1})\,.
\end{aligned}$$

(ii) $\mathbb{P}$-convexity of $U^{(a)}$ is proved as follows. Again the special choice $\epsilon = \epsilon_1$ is not needed, so let be $\epsilon > 0$ arbitrary and let $\mu_{ij} := \mu_{ij}(\omega_1,\ldots,\omega_{i-1})$ as in Definition 3.2 (ii). Then

$\sum_{j=0}^{m-1} U_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1})\mu_{ij}$

$= \exp(-t\epsilon)\exp(\frac{1}{2}t^2\sum_{k=i+1}^{n} d_k^2)\sum_{j=0}^{m-1}\mu_{ij}\exp(t[\mathbb{E}(f|C_{ij})-\mathbb{E}(f)])$

$= (1)$.

With Lemma 3.11 we have

$$\exp(t\mathbb{E}(f|C_{ij})) = E\left(e^{tf_i}|C_{ij}\right)\ .$$

With this and Lemma 3.10 we have the estimates

$$\begin{aligned}
(1) &= \exp(-t\epsilon)\exp(\frac{1}{2}t^2\sum_{k=i+1}^{n} d_k^2)E\left(e^{tf_i}|C_i\right)\cdot\exp(-t\mathbb{E}(f))\\
&\leq \exp(-t\epsilon)\exp(\frac{1}{2}t^2\sum_{k=i}^{n} d_k^2)E\left(e^{tf_{i-1}}|C_i\right)\exp(-t\mathbb{E}(f))\\
&= \exp(-t\epsilon)\exp(\frac{1}{2}t^2\sum_{k=i}^{n} d_k^2)e^{t[\mathbb{E}(f|C_i)-\mathbb{E}(f)]}\\
&= U_{i-1,\omega_{i-1}}^{(a)}(\omega_1,\ldots,\omega_{i-2})
\end{aligned}$$

(iii) Now we choose special values for ϵ and t. Let $\epsilon := \epsilon_1$ and $t := t_1 = \epsilon_1\Delta^{-1}$. We show the initial condition

$$\min U_{1j}^{(a)} \leq \delta.$$

Let $C_{1j} = \{\omega \in \Omega; \omega_1 = j\}$ and $p_{1j} = \mathbb{P}(C_{1j})$.
Then

$$\begin{aligned}
U_1^{(a)}(\omega_1) &= \min_{0 \leq j \leq m-1} U_{1j}^{(a)} \\
&\leq \sum_{j=0}^{m-1} p_{1j} U_{1j}^{(a)} \\
&= \exp(-t\epsilon)\exp(-t\mathbb{E}(f))\exp(\frac{t^2}{2}\sum_{k=2}^{n} d_k^2)\sum_{j=0}^{m-1} p_{1j}\exp(-t\mathbb{E}(f|C_{1j})) \\
&= \exp(-t\epsilon)\exp(\frac{t^2}{2}\sum_{k=2}^{n} d_k^2)\sum_{j=0}^{m-1} p_{1j}\exp(t[\mathbb{E}(f|C_{1j}) - E(f)]) \\
&= \exp(-t\epsilon)\exp(\frac{t^2}{2}\sum_{k=2}^{n} d_k^2)\sum_{j=0}^{m-1} E\left(e^{t[\mathbb{E}(f|\mathcal{F}_1)-\mathbb{E}(f)]}|C_{1j}\right) \\
&\quad \text{(Lemma 3.11)} \\
&= \exp(-t\epsilon)\exp(\frac{t^2}{2}\sum_{k=2}^{n} d_k^2)E\left(e^{t[\mathbb{E}(f|\mathcal{F}_1)-\mathbb{E}(f)]}|\mathcal{F}_0\right) \\
&\leq \exp(-t(\epsilon - \frac{t}{2}\sum_{k=1}^{n} d_k^2)) \quad \text{(Lemma 3.10)} \\
&= \delta < 1 .
\end{aligned}$$

□

In the next step we construct pessimistic estimators. The idea is the following: Suppose a family $U^{(a)}$ is a weak pessimistic estimator. Suppose furthermore that the $U_{ij}^{(a)}$ can be approximated by polynomial-time computable functions up to a specified precision $\gamma \geq 0$. Let $q(n,m) \geq 1$ be a polynomial with $\delta + \frac{\delta}{q(n,m)} < 1$. Then with $\gamma = \frac{\delta}{(4n-1)q(n,m)}$ we approximate $U_{ij}^{(a)}$ by functions $V_{ij}^{(a)}$ up to the absolute error γ and a family of the form $(V^{(a)} + (4n-1)\gamma)$ is the desired pessimistic estimator in the RAM model.

In case of functions with bounded martingale differences it will be sufficient to approximate the $U_{ij}^{(a)}$, which are compositions of square roots, logarithms and exponential functions, by their Taylor polynomials. But we state the result for the general case.

Definition 3.12 (Approximate Pessimistic Estimators) *Let $\epsilon > 0$, $\delta \in (0,1)$ and $q(n,m)$ a polynomial with $q(n,m) \geq 1$. Let $U^{(a)}, U^{(b)}, U^{(c)}$ be weak pessimistic estimators for the events $E_a(\epsilon), E_b(\epsilon)$ and $E_c(\epsilon)$. Let $V^{(a)}, V^{(b)}$ and $V^{(c)}$ be families of functions computable in the RAM model in time bounded by a polynomial in n, m and $\ln\frac{1}{\delta}$ having the properties* *(a)* $\quad |U_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) - V_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1})| \leq \frac{\delta}{(4n-1)q(n,m)}$

(b) $\quad |U_{ij}^{(b)}(\omega_1,\ldots,\omega_{i-1}) - V_{ij}^{(b)}(\omega_1,\ldots,\omega_{i-1})| \leq \frac{\delta}{(4n-1)q(n,m)}$

(c) $\quad |U_{ij}^{(c)}(\omega_1,\ldots,\omega_{i-1}) - V_{ij}^{(c)}(\omega_1,\ldots,\omega_{i-1})| \leq \frac{\delta}{2(4n-1)q(n)}$,

and define the families $W^{(a)}, W^{(b)}, W^{(c)}$ as follows:

(d) $\quad W_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) = V_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) + \frac{2(2n-i)}{4n-1}\frac{\delta}{q(n,m)}$

(e) $\quad W_{ij}^{(b)}(\omega_1,\ldots,\omega_{i-1}) = V_{ij}^{(b)}(\omega_1,\ldots,\omega_{i-1}) + \frac{2(2n-i)}{4n-1}\frac{\delta}{q(n,m)}$

(f) $\quad W_{ij}^{(c)}(\omega_1,\ldots,\omega_{i-1}) = V_{ij}^{(c)}(\omega_1,\ldots,\omega_{i-1}) + \frac{2n-i}{4n-1}\frac{\delta}{q(n)}$.

Proposition 3.13 *The families $W^{(a)}, W^{(b)}, W^{(c)}$ are pessimistic estimators for the events $\bar{E}_a(\epsilon)$, $\bar{E}_b(\epsilon)$ and $\bar{E}_c(\epsilon)$.*

By Proposition 3.13 D-WALK is a deterministic polynomial-time algorithm finding an $\omega \in \Omega$ such that $f(\omega) \leq \mathbb{E}(f) + \epsilon$ etc.

Proof of Proposition 3.13: Let us consider $W^{(a)}$, the arguments for the other cases are the same:

(i) The upper bound condition is an immediate consequence of Definition 3.2 (i):
Define $\gamma := \frac{\delta}{(4n-1)q(n,m)}$. Then

$$\begin{aligned}
\mathbb{P}(\bar{E}_a(\epsilon))|\omega_1,\ldots,\omega_{i-1},j) &\leq U_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) \\
&\leq V_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) + \gamma \\
&\leq V_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) + \gamma + \gamma(4n-2i-1) \\
&= W_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}).
\end{aligned}$$

(ii) $\mathbb{P}$-convexity follows with Definition 3.2(ii):
Let $\mu_{ij} := \mu_{ij}(\omega_1,\ldots,\omega_{i-1})$ as in Definition 3.2 (ii).

$$\begin{aligned}
\sum_{j=0}^{m-1} \mu_{ij} W_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) &= \sum_{j=0}^{m-1} \mu_{ij} V_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) + 2(2n-i)\gamma \\
&\leq \sum_{j=0}^{m-1} \mu_{ij} U_{ij}^{(a)}(\omega_1,\ldots,\omega_{i-1}) + \gamma + 2(2n-i)\gamma \\
&\leq U_{i-1,\omega_{i-1}}^{(a)}(\omega_1,...,\omega_{i-2}) + (4n-2i+1)\gamma \\
&\leq V_{i-1,\omega_{i-1}}^{(a)}(\omega_1,...,\omega_{i-2}) + (4n-2i+2)\gamma \\
&= W_{i-1,\omega_{i-1}}^{(a)}(\omega_1,...,\omega_{i-2})
\end{aligned}$$

(iii) The initialization follows from the assumption on $U_{1j}^{(a)}$:

$$\begin{aligned}\min_{0\le j\le m-1} W_{1j}^{(a)} &= \min_{0\le j\le m-1} V_{1j}^{(a)} + 2(2n-1)\gamma \\ &\le \min_{0\le j\le m-1} U_{1j}^{(a)} + \gamma + 2(2n-1)\gamma \\ &\le \delta + \frac{\delta}{q(n,m)} < 1.\end{aligned}$$

□

In the next step we approximate the weak pessimistic estimators of Definition 3.8 by Taylor polynomials and construct the W_{ij}. We need the following technical lemma, which assures that in our case approximation by Taylor polynomials is easy. We recall that we defined running time as the number of arithmetic operations.

Lemma 3.14 *(i) Let y be a rational number with encoding length L and $\gamma_1 \in (0,1)$ a positive real number. Let N be a positive integer with $N \ge 7\lceil |y| \rceil + \lfloor \log_3 \frac{1}{\gamma_1} \rfloor$. Then the N-th degree Taylor polynomial $T_N(y) = \sum_{k=0}^{N} \frac{y^k}{k!}$ of $\exp(y)$ has encoding length $O(LN)$, can be computed in $O(N)-$time such that $|\exp(y) - T_N(y)| \le \gamma_1$.*

(ii) Let $x \ge 1$ be a rational number with encoding length L and $\gamma_2 \in (0,1)$ a positive real number. Let N be a positive integer with $N \ge 2\lceil \log \frac{4}{\gamma_2} \rceil$. Then a rational number y with encoding length $O(LN)$ can be computed in $O(N)-$time such that $|\ln\ x - y| \le \gamma_2$.

(iii) Let x be a rational number with encoding length L, $\gamma_3 \in (0,1)$ a positive real number. If $x \ge 1$ let N be a positive integer with $N \ge \lceil \log \frac{x}{\gamma_3} \rceil$ and if $0 < x < 1$ let $N \ge \lceil \log \frac{1}{\gamma_3} \rceil$. Then a rational number y with encoding length $O(L+N)$ can be computed in $O(N)-$time such that $|\sqrt{x} - y| \le \gamma_3$.

(iv) Let $x \ge \frac{4}{3}$ be a rational number with encoding length L and $\gamma_4 \in (0,1)$ a real number. Let N be a positive integer with $N \ge 2\lceil \log \frac{8x}{\gamma_4} \rceil$. Then a rational number y with encoding length $O(LN)$ can be computed in time $O(N)$ such that $|\sqrt{\ln\ x} - y| \le \gamma_4$.

Proof.
For a proof of (i) – (iii) see ([38], Lemma 2.4).

(iv) By (ii) we can find a rational $y_0 \ge \ln\ x$ such that $y_0 - \ln\ x \le \frac{\gamma_4}{2}$ in time $O(N)$. Since $0 < y_0 \le \ln x + \frac{\gamma_4}{2} \le 2x$ we can find by (iii) a rational y with $|y - \sqrt{y_0}| \le \frac{\gamma_4}{2}$ in $O(N)-$time. For y we have the estimates

$$|\sqrt{\ln\ x} - y| \le \sqrt{y_0} - \sqrt{\ln\ x} + |y - \sqrt{y_0}|$$

$$
\begin{aligned}
&\le \frac{\gamma_4}{2} + \frac{y_0 - \ln\ x}{\sqrt{y_0} + \sqrt{\ln\ x}} \\
&\le \frac{\gamma_4}{2}(1 + \frac{1}{2\sqrt{\ln\ x}}) \\
&\le \gamma_4 \qquad (\text{Using } \ x \ge \frac{4}{3}).
\end{aligned}
$$

Furthermore by (ii) and (iii) the encoding length of y is $O(LN)$.

□

Remark In Lemma 3.14 the computation of the integer N involves the calculation of logarithmic terms $\log\frac{1}{\gamma_i}$ with real γ_i. In the RAM model this cannot be carried out. But in our applications of Lemma 3.14 the terms $\log\frac{1}{\gamma_i}$ are bounded by polynomials p_γ in n, m, L and $\log\frac{1}{\delta}$. So we evaluate p_γ and then choose a positive integer N, $N = O(p_\gamma)$, such that N is sufficiently large as required by Lemma 3.14.

Before we proceed to the main theorem of this section, recall that the input size of our derandomization problem is given by n, m, $\log\frac{1}{\delta}$ and L (the maximal encoding length of rational numbers, which are needed to implement an algorithm for the pointwise evaluation of f).

Theorem 3.15 *Let $f : \Omega \longrightarrow \mathbb{Q}$ be a function with bounded martingale differences, $||\mathbb{E}(f|\mathcal{F}_k) - E(f|\mathcal{F}_{k-1})||_\infty \le d_k$ for each $k \in \{1, \ldots, n\}$, where $d_k \ge 0$ are rational numbers. Let $c > 0$ be a constant and let p_1, p_2, p_4 be rational-valued polynomials in n and m with $c \le p_1, p_2, p_4$. Let p_3 be a rational-valued polynomial in L, n and m. Suppose that the following conditions are satisfied.*

(i) $||f||_\infty \le p_1(n, m)$.

(ii) *For each $\omega \in \Omega$ and $k \in \{1, \ldots, n\}$ the conditional expectation $\mathbb{E}(f|\mathcal{F}_k)(\omega)$ can be computed in time bounded by $p_2(n, m)$.*

(iii) *For all $k \in [n]$ and $\omega \in \Omega$, the maximal encoding length of a number appearing in the computation of $\mathbb{E}(f|\mathcal{F}_k)(\omega)$ is $p_3(L, n, m)$.*

(iv) $\sum_{k=1}^{n} d_k^2 \ge \frac{1}{p_4(n, m)}$

Then we have for every $0 < \delta < 1$ and $\epsilon_i = \sqrt{2\sum_{k=1}^{n} d_k^2 \ln\frac{2i}{\delta}}$, $(i = 1, 2)$

(a) *Pessimistic estimators for the events $\bar{E}_a(\epsilon_1)$, $\bar{E}_b(\epsilon_1)$ and $\bar{E}_c(\epsilon_2)$ can be computed in $O(\ \log\frac{1}{\delta}\ [p_1p_4 + \log n]\ + p_2\ )$-time and with $O(\ (\log\frac{1}{\delta})^3 p_3\ [p_1p_4 + n]^2\ )$-space.*

(b) The algorithm D-WALK finds an $\omega \in \Omega$ such that $f(\omega) \leq \mathbb{E}(f) + \epsilon_1$ (resp. $f(\omega) \geq \mathbb{E}(f) - \epsilon_1$, resp. $|f(\omega) - \mathbb{E}(f)| \leq \epsilon_2$) in $O(\ nm(\log \frac{1}{\delta}\ [p_1 p_4 + \log n]\ + p_2)\ )$-time.

Remark

(a) As the running time in Theorem 3.15 does not contain the encoding length we have indeed a strongly polynomial algorithm.

(b) Note that the difference between the approximation parameter ϵ_i in Theorem 3.15 and the probabilistic statement (Proposition 3.7) is the term $\ln \frac{2i}{\delta}$ instead of $\ln \frac{i}{\delta}$. So the approximation is less tight, but this fact will enable us to approximate the weak pessimistic estimators by Taylor polynomials and to apply Proposition 3.13.

Proof. We consider the event $E_a(\epsilon_1)$. The argumentation for the other two events is similar. By Definition 3.8

$$U_{ij}^{(a)}(\omega_1, ..., \omega_{i-1}) = \exp(-t_1[\epsilon_1 - \frac{t_1}{2}(d_{i+1}^2 + ... + d_n^2) + \mathbb{E}(f) - \mathbb{E}(f|\ \omega_1, ..., \omega_{i-1}, j)]).$$

With $\Delta := \sum_{k=1}^{n} d_k^2, \quad \alpha := \Delta^{-1} \sum_{k=i+1}^{n} d_k^2 - 2,$

$$\beta := \sqrt{\frac{2}{\Delta}}(\mathbb{E}(f|\ \omega_1, ..., \omega_{i-1}, j) - \mathbb{E}(f))$$

and

$$x := \alpha \ln \frac{2}{\delta} + \beta \sqrt{\ln \frac{2}{\delta}}$$

the function $U_{ij}^{(a)}$ is rewritten in the simple form

$$U_{ij}^{(a)}(\omega_1, ..., \omega_{i-1}) = \exp(x).$$

We will approximate $\exp(x)$ by suitable Taylor polynomials. ¿From the assumptions (i)-(iv) it follows with

$$p(|x|) := 13\lceil p_1 p_4 \rceil \lceil \log \frac{1}{\delta} \rceil$$

that

$$|x| \leq 1 + |x| \leq p(|x|).$$

Let $\gamma := \frac{\delta}{2(4n-1)}$ and $E := \mathbb{E}(f|\ \omega_1, ..., \omega_{i-1}, j) - \mathbb{E}(f)$. Then

$$\begin{aligned} p(|x|) &= O(p_1 p_4 \log \frac{1}{\delta}), \quad |E| = O(p_1). \\ \log \frac{1}{\gamma} &= O(\log n + \log \frac{1}{\delta}) \text{ and } |\alpha| = O(1). \end{aligned}$$

In the sequel we will use Lemma 3.14. In case the given approximation error γ is bigger than 1, we avoid negative $\ln \frac{1}{\gamma}$ terms in Lemma 3.14 replacing γ by 0.5.

Step 1: Approximation of $\beta\sqrt{\ln \frac{2}{\delta}}$

Computation of E: Let $\omega' := (\omega_1, ..., \omega_{i-1}, j, \omega'_{i+1}, ..., \omega'_n)$, and ω'_k arbitrary for $k \geq i+1$. The conditional expectation $\mathbb{E}(f|\ \mathcal{F}_i)$ is constant on each partition class of P_i, hence $\mathbb{E}(f|\ \omega_1, ..., \omega_{i-1}, j) = \mathbb{E}(f|\ \mathcal{F}_i)(\omega')$ and by condition (iii) E can be computed in time

$$O(p_2). \tag{7}$$

Approximation of $\sqrt{\frac{2}{\Delta}}$:
We wish to find a rational y_1 with

$$|\sqrt{\frac{2}{\Delta}} - y_1| \leq \frac{1}{16}\gamma\delta|E|^{-1}e^{-p(|x|)}. \tag{8}$$

Let $N = 2p(|x|) + \lceil \log \frac{32|E|}{\gamma\delta\Delta} \rceil$.
Since $|E| = O(p_1)$ and $p(|x|) = O(p_1 p_4 \log \frac{1}{\delta})$ we have

$$\lceil \log(\frac{32|E|}{\gamma\delta\Delta}e^{p(|x|)}) \rceil \leq N = O(p_1 p_4 \log \frac{1}{\delta} + \log n).$$

Hence with Lemma 3.14 (iii) we can compute a y_1 such that (8) holds in

$$O(p_1 p_4 \log \frac{1}{\delta} + \log n) \tag{9}$$

time. Define $\beta_1 := y_1 E$. With (2)

$$\begin{aligned} |\beta_1| &= |y_1||E| \\ &\leq |E|\left(\sqrt{\frac{2}{\Delta}} + \frac{\gamma\delta}{16|E|}e^{-p(|x|)}\right) \\ &\leq |E|\sqrt{\frac{2}{\Delta}} + 1 \\ &\leq 2|E|\max(1, \frac{2}{\Delta}) \\ &= O(p_1 p_4), \end{aligned}$$

and

$$\ln|\beta_1| = O(\log p_1 + \log p_4) \tag{10}$$

Approximation of $\sqrt{\ln\frac{2}{\delta}}$:
Let $\gamma_3 := \min(1, \frac{1}{8}\gamma|\beta_1|^{-1}e^{-p(|x|)})$. We want to compute a rational y_2 with

$$|\sqrt{\ln\frac{2}{\delta}} - y_2| \leq \gamma_3$$

With (10) we have

$$2\lceil\log(\frac{2}{\delta}\frac{8}{\gamma_3})\rceil = O\left([p_1p_4 + \log n]\log\frac{1}{\delta}\right).$$

Using this and Lemma 3.14 (iv) we can compute y_2 in

$$O\left([p_1p_4 + \log n]\log\frac{1}{\delta}\right) \tag{11}$$

time.

Approximization of $\beta\sqrt{\ln\frac{2}{\delta}}$:

The time for the computation of $\beta_1 y_2$ is by (7), (9) and (11)

$$O\left(\log\frac{1}{\delta}[p_1p_4 + \log n] + p_2\right) \tag{12}$$

Finally we get the desired estimate:

$$\begin{aligned} |\beta\sqrt{\ln\frac{2}{\delta}} - \beta_1 y_2| &\leq |\beta - \beta_1|\sqrt{\ln\frac{2}{\delta}} + |\sqrt{\ln\frac{2}{\delta}} - y_2|\cdot|\beta_1| \\ &\leq \left(\frac{\gamma}{8}\frac{\delta}{2}\sqrt{\ln\frac{2}{\delta}} + \frac{\gamma}{8}\right)e^{-p(|x|)} \leq \frac{\gamma}{4}e^{-p(|x|)}. \end{aligned}$$

Encoding length of $\beta_1 y_2$:

Since $\mathbb{E}(f|\omega_1,\ldots,\omega_{i-1},j) = \mathbb{E}(f|\mathcal{F}_i)(\omega')$ with $\omega' = (\omega_1,\ldots,\omega_{i-1},j,\omega'_{i+1},\ldots,\omega'_n)$, the encoding length of E is by condition (iii) the theorem $O(p_3)$. By the proof of Lemma 3.14 (iii) y_1 is computed through halving the interval $[0,\frac{2}{\Delta}]$ $O(p_1p_4\log\frac{1}{\delta})$ times. By condition (ii) of the theorem we may assume that the encoding length of the d_k's is $O(p_3)$, hence of $\frac{2}{\Delta}$ it is $O(p_3+n)$ and finally y_1 has encoding length $O(p_3 + n + p_1p_4\log\frac{1}{\delta})$. Therefore $\beta_1 = y_1E$ has encoding length

$$O(p_3 + n + p_1p_4\log\frac{1}{\delta}). \tag{13}$$

According to Lemma 3.14 (iv) the encoding length of y_2 is

$$O\left(\log\frac{1}{\delta}\log\frac{1}{\gamma_3}\right) = O\left((\log\frac{1}{\delta})^2[p_1p_4 + \log n]\right). \tag{14}$$

By (13) and (14) the encoding length of $\beta_1 y_2$ then is

$$O\left((\log\frac{1}{\delta})^2\,[p_1p_4+\log n]+p_3+n\right). \tag{15}$$

Step 2: Approximation of $\alpha\ln\frac{2}{\delta}$
We wish to approximate $\ln\frac{2}{\delta}$ by a rational y_3 with

$$|\ln\frac{2}{\delta}-y_3|\leq\frac{\gamma}{4|\alpha|}e^{-p(|x|)}.$$

Let $N=4p(|x|)+2\lceil\log(\frac{4|\alpha|}{\gamma})\rceil$. Then

$$2\lceil\log\frac{4|\alpha|e^{p(|x|)}}{\gamma}\rceil\leq N=O\left(\log\frac{1}{\delta}\,[p_1p_4+\log n]\right).$$

According to Lemma 3.14 (ii) we can find y_3 in

$$O\left(\log\frac{1}{\delta}\,[p_1p_4+\log n]\right) \tag{16}$$

time.

Encoding length of αy_3 :

By Lemma 3.14 (ii) the encoding length of y_3 is

$$O\left(N\log\frac{1}{\delta}\right)=O\left((\log\frac{1}{\delta})^2\,[p_1p_4+\log n]\right). \tag{17}$$

The encoding length of α is $O(p_3+n)$, hence with (17) the encoding length of αy_3 is

$$O\left((\log\frac{1}{\delta})^2\,[p_1p_4+\log n]+p_3+n\right). \tag{18}$$

Step 3: Approximation of $\exp(x)$

Now with $y:=\alpha y_3+\beta_1 y_2$ we have

$$|x-y|\leq\frac{\gamma}{2}e^{-p(|x|)}. \tag{19}$$

By the mean value theorem there is a z in $[x,y]$ (or in $[y,x]$) with $|e^x-e^y|=e^z|x-y|$. Using (19) and $|x|+1\leq p(|x|)$ we have

$$\begin{aligned}|e^x-e^y| &= |x-y|e^z\\ &\leq |x-y|\exp(|x|+\frac{\gamma}{2})\\ &\leq \frac{\gamma}{2}e^{-p(|x|)}e^{p(|x|)}\\ &= \frac{\gamma}{2}.\end{aligned}$$

We need an estimate on $|y|$. Observe that $|\alpha| \leq 3$, $|y_3| \leq \log(\frac{4}{\delta})$ and as showed in Step 1 $\beta_1 \leq 2|E|\max(1, \frac{2}{\Delta})$. Furthermore $|y_2| \leq \log\frac{4}{\delta}$ and all this together imply

$$|y| \leq \log\frac{4}{\delta}(3 + 2|E|\max(1, \frac{2}{\Delta})).$$

Define $N = 7\lceil\log\frac{4}{\delta}\rceil(3 + 2|E|\max(1, \frac{2}{\Delta})) + \lfloor\log_3\frac{2}{\gamma}\rfloor$. Then

$$7\lceil|y|\rceil + \lfloor\log_3\frac{2}{\gamma}\rfloor \leq N = O(\log(\frac{1}{\delta})[p_1p_4 + \log n]). \tag{20}$$

We apply Lemma 3.14 (i) in order to compute the Taylor polynomial $T_N(y)$ such that

$$|e^y - T_N(y)| \leq \frac{\gamma}{2}:$$

The total time for the computation of y is by (12) and (16) $O(\log\frac{1}{\delta}[p_1p_4 + \log n] + p_2)$. And this together with (20) gives by Lemma 3.14 (i) for the computation of $T_N(y)$ the total time of

$$O(\log\frac{1}{\delta}[p_1p_4 + \log n] + p_2). \tag{21}$$

Encoding length of y and $T_N(y)$:

With (15) and (18) the encoding length of y is bounded by

$$O((\log\frac{1}{\delta})^2 p_3[n + p_1p_4]). \tag{22}$$

With Lemma 3.14 (i), (20) and (22) the encoding length of $T_N(y)$ then is

$$O(N(\log\frac{1}{\delta})^2 p_3[n + p_1p_4]) = O((\log\frac{1}{\delta})^3 p_3[n + p_1p_4]^2). \tag{23}$$

(21) together with (23) proves the assertion (a) of the theorem.

By the assumption of the theorem $\epsilon_1 = \sqrt{2\Delta\ln\frac{2}{\delta}}$. Now take in Definition 3.12 $q(n, m) = 1$ and define the family $W^{(a)}$ by

$$W_{ij}^{(a)}(\omega_1, ..., \omega_{i-1}) = T_N(y) + \frac{(2n - i)\delta}{(4n - 1)},$$

where $T_N(y)$ is the Taylor polynomial of $U_{ij}^{(a)}(\omega_1, ..., \omega_{i-1})$ as constructed above.

Theorem 3.9 and Proposition 3.13 imply that the family $W^{(a)}$ is a pessimistic estimators for the event $\bar{E}_a(\epsilon_1)$.

We have to compute on each level i, $m - 1$ functions $W_{ij}^{(a)}$, hence by (20) the running time of the D-WALK procedure is

$$O(\ nm(\log\frac{1}{\delta}\ [p_1p_4 + \log n]\ + p_2)\ ),$$

and assertion (b) of the theorem is proved.

□

The term for the running time collapses immediately considering special cases, especially when f is a quadratic or a linear function. The following corollary is the basic result for the applications in the next section.

Let $m = 2,\ \Omega = \{0,1\}^n, Q = (q_{ij})$ a $n \times n$-matrix and $c = (c_1, \dots, c_n)$ a vector with $|c_i|, |q_{ij}| \leq 1$ and let L be the encoding length of Q and c. Let f be the quadratic function

$$f(x) = c^T x + x^T Q x,$$

$x \in [0,1]^n$. We assume that the numbers $d_k, k = 1, \dots, n$ are positive rationals with a constant lower bound and satisfying the property

$$|f(x) - f(x')| \leq d_k,$$

if $x_i = x'_i$ for all $i \neq k$.

With $\Delta := \sum_{k=1}^n d_k^2,\ 0 < \delta < 1,\ \epsilon_i := \sqrt{2\Delta \ln \frac{2i}{\delta}}$ we have

Corollary 3.16

(i) Pessimistic estimators for the events $\bar{E}_a(\epsilon_1),\ \bar{E}_b(\epsilon_1),\ \bar{E}_c(\epsilon_2)$ can be computed in $O(n^2 \log \frac{1}{\delta})$-time and requiring $O(Ln^4\ (\log \frac{1}{\delta})^3)$ space.

(ii) The procedures $D - WALK(\bar{E}_a(\epsilon_1))$, resp. $D - WALK(\bar{E}_b(\epsilon_1))$, resp. $D - WALK(\bar{E}_c(\epsilon_2))$ find in $O(n^3 \log \frac{1}{\delta})$ time a $x \in \Omega$ such that $f(x) \leq \mathbb{E}(f) + \epsilon_1$,
resp. $f(x) \geq \mathbb{E}(f) - \epsilon_1$, resp. $|f(x) - \mathbb{E}(f)| \leq \epsilon_2$.

(iii) In the linear case $(Q = 0)$ the time for the computation of the pessimistic estimator is $O(n \log \frac{1}{\delta})$, the running time for the D-WALK procedure is $O(n^2 \log \frac{1}{\delta})$ and the space needed is $O(Ln^2\ (\log \frac{1}{\delta})^3)$.

Proof. (i) By a slight modification of the proof of theorem 4.1 of [2], f possesses martingale differences $||\mathbb{E}(f|F_k) - \mathbb{E}(f|F_{k-1})||_\infty \leq d_k$ for all $k = 1, \dots, n$. Now apply Theorem 3.15 with $p_1 = O(n^2)$, $p_2 = O(n^2)$, $p_3 = L$ and $p_4 =$ *const.*.
(ii) In the linear case take $p_1 = p_2 = O(n)$.

□

4 Applications

Lattice Approximation

An instance of the Quadratic Lattice Approximization problem (QLA) is a symmetric $r \times r$ matrix D, a $(n-1) \times r$ matrix C, rational vectors $c, p \in [0,1]^r$ and an objective function $x \to c^T x + x^T Dx \quad (x \in [0,1]^r)$. The problem is to find a lattice point $q \in \{0,1\}^r$ in polynomial-time such that

(a) $|c^T(p-q) + p^T Dp - q^T Dq|$ is small

(b) $||C(p-q)||_\infty$ is small.

We assume that the entries of D, C are rational numbers and $0 \le c_{ij} \le 1$. Let L be the encoding length of (D, C, c, p). As discussed in the introduction, the quadratic lattice approximization problem has an interesting interpretation in $0-1$ quadratic optimization. When $D \equiv 0$ the problem simplifies to the well known lattice approximization problem (LA). Let $d := 2\max_{1\le i\le r} \sum_{j=1}^r |d_{ij}|$. Derandomization gives the following result

Theorem 4.1 *Let $\alpha \ge 0$ and trace$(D) \le \alpha d\sqrt{n}$. Then the procedure D-WALK finds in $O(\ r^2 n \log n + r^3\ )$-time and requiring $O(Lr^2(r^2 + (\log n)^3) + n)$ space a vector $q \in \{0,1\}^r$ such that*

(i) $|c^T(p-q) + p^T Dp - q^T Dq| \le 2\sqrt{n \ln 2n} + (3+\alpha)d\sqrt{n}$

(ii) $||C(p-q)||_\infty \le 2\sqrt{n \ln 2n}$

Proof. Define $\epsilon_1 = 2\sqrt{n \ln n}$ and $\epsilon_2 = 3d\sqrt{n}$.
Let f be the function $f(x) = x^T Dx,\ x \in [0,1]^r$. Denote by $\bar{E}_0$ the event

$$"|c^T(p-q)| > \epsilon_1",$$

by $\bar{E}_i$, $i = 1, \ldots, n-1$ the events

$$"|\sum_{j=1}^r c_{ij}(p_j - q_j)| > \epsilon_1",$$

and by $\bar{E}_n$ the event

$$"|q^T Dq - \mathbb{E}(f)| > \epsilon_2".$$

In order to apply Corollary 3.16 note that

$$|f(x) - f(x')| \le d,$$

if $x_i = x'_i$ for all $i \ne k$. By Corollary 3.16 (i) we can compute a pessimistic estimator $W^{(n)}$ for the event $\bar{E}_n$ in $O(r^2)$-time and with $O(Lr^4)$ space. By Corollary 3.16 (iii) pessimistic estimators $W^{(i)}$ for the events $\bar{E}_i$, $i = 0, \ldots, n-1$ can be computed in $O(r \log n)$-time and with $O(Lr^2(\log n)^3)$ space. Then by Proposition 3.4 $W := \sum_{i=0}^n W^{(i)}$ is a pessimistic estimator for the event $\bar{E}_0 \vee \ldots \vee \bar{E}_{n-1} \vee \bar{E}_n$ and can be computed in $O(\ rn\log + r^2\ )$-time and with at most $O(Lr^2(r^2 + (\log)^3) + n)$ space.
Hence $D - WALK(\bar{E}_0 \vee \ldots \vee \bar{E}_{n-1} \vee \bar{E}_n)$ finds in $O(\ r^2 n \log n + r^3\ )$-time a vector $q \in \{0,1\}^r$ such that

(a) $|q^T Dq - \mathbb{E}(f)| \leq 3d\sqrt{n}$

(b) $||[C,c](p-q)||_\infty \leq 2\sqrt{n \ln 2n}$.

Let (ξ_i) be the Bernoulli trials under considerations defined trough $\mathbb{P}(\xi_i = 1) = p_i$ and $\mathbb{P}(\xi_i = 0) = 1 - p_i$. Then the expectation $\mathbb{E}(f)$ is:

$$\mathbb{E}(f) = \mathbb{E}(\sum_{i,j=1}^{r} d_{ij}\xi_i\xi_j) = \sum_{i \neq j} d_{ij}p_ip_j + \sum_{i=1}^{r} d_{ii}p_i.$$

But this together with

$$|p^T Dp - \mathbb{E}(f)| \leq \sum_{j=1}^{r} d_{jj}(p_j - p_j^2) \leq \alpha d\sqrt{n},$$

implies the theorem.

□

Remark Theorem 4.1 shows the similarity between the linear and quadratic discrepancy bounds. In the quadratic case we have a $O(d\sqrt{n})$ bound, while in the linear case the algorithmic reachable bound is $O(\sqrt{n \ln n})$ and the existence bound of Spencer is $O(\sqrt{n})$. It is known that Spencers bound is sharp for Hadamard matrices. The interesting question arising here is whether the gap factor d reflects the quadratic behaviour and so is best possible or not. For small d, i.e.
$d = O(n^{\frac{1}{2}-\epsilon})$, $0 < \epsilon \leq \frac{1}{2}$, and if the trace of D is not too large trace our bound is good compared with the greedy bound $O(n)$, which is also the worst case discrepancy (attained for $D = (d_{ij}), d_{ij} = 1$ for all i,j and $p = (\frac{1}{2}, \ldots, \frac{1}{2})$). It would be interesting to exhibit more classes of matrices where lattice approximations beating the $O(n)$ greedy bound are possible.

In the weighted linear case Theorem 4.1 gives an $O(\ r^2 n \log n\ )$-time algorithm achieving discrepancies within $2\sqrt{n \ln 2n}$:

Corollary 4.2 *The procedure D-WALK finds in $O(\ r^2 n \log n\ )$-time and requiring $O(Lr^2(\log n)^3 + n)$ space a vector $q \in \{0,1\}^r$ such that*

$$||[C,c](p-q)||_\infty \leq 2\sqrt{n \ln 2n}.$$

□

Remark Raghavan improved the Beck-Fiala bound using Angluin-Valiant type inequalities. He showed a derandomized algorithm which achieves
$||C(p-q)||_\infty \leq \max_{1 \leq i \leq n} s_i D(s_i, \frac{1}{2n})$, where $s_i = \sum_{j=1}^{r} c_{ij}p_j$. Unfortunately in the weighted case the algorithm has no polynomial-time implementation, because

the numbers $D(s_i, \frac{1}{2n})$ and $e^{c_{ij}}$ cannot be computed efficiently in the RAM model. But using the upper bounds of Raghavan on $D(s_i, \frac{1}{2n})$ (see [31], 1.13 and 1.14) and Taylor approximations the following result can be proved in a similar way as Corollary 4.2:

Theorem 4.3 *Let $\Delta_i = ([C,c]p)_i$. Derandomization gives an $O(\ r^2 n \log n\ )$-time and $O(Lr^2(\log n)^3 + n)$ space algorithm, which finds a vector $q \in \{0,1\}^r$ such that*

(i) $\Delta_i \leq 3\sqrt{s_i \ln 2n}$, *if* $s_i > \ln 4n$ *for all* i.

(ii) $\Delta_i \leq 6 \ln 2n$, *if* $s_i \leq \ln 4n$ *for all* i.

□

Balancing matrices

Let $A = (a_{ij})$ be a $n \times n$ matrix with $a_{ij} = \pm 1$ for all i. Beck and Spencer [11] gave a polynomial-time algorithm finding row shifts $x_i = \pm 1$ and column shifts $y_i = \pm 1, 1 \leq i \leq n$, such that $|\sum_{i,j=1}^{n} a_{ij} x_i y_i| \leq 2$. In this problem the crucial point is that the row shifts and the column shifts do not depend on each other. A formally similar, but mathematical different problem is stated, when the row and column shifts must be the same.

Definition 4.4 (Dependently Matrix Balancing Problem) *Let A be as above.*

(a) What is the minimal number $K(n)$ such that there exist $x_i = \pm 1,\ 1 \leq i \leq n$ with $|\sum_{i,j=1}^{n} a_{ij} x_i x_j| \leq K(n)$?

(b) If the answer in (a) for a certain $K(n)$ is affirmative, can one find the $x_1, \ldots, x_n$ in polynomial time?

With a greedy algorithm the problem in 4.4 can be solved for $K(n) = 2n$. But if $a_{ii} = 0$ for all i, by chance $K(n)$ may be much smaller than $O(n)$. The only yet known result is the following, using martingales and derandomization.

Theorem 4.5 *Let $A = (a_{ij})$ be a $n \times n$ matrix with $a_{ij} = \pm 1$ and $a_{ii} = 0$ for all i,j. Let $d_k := 2\sum_{j=1}^{n}(|a_{kj}| + |a_{jk}|)$ and $\Delta := \sum_{k=1}^{n} d_k^2$.*
Then a vector $x \in \{-1,+1\}^n$ can be found in $O(n^3)$-time such that

$$|\sum_{i,j=1}^{n} a_{ij} x_i x_j| \leq 2\sqrt{\Delta}.$$

Remark: With $d := \max_k d_k$, then the bound is $4d\sqrt{n}$ and for "small" d, i.e. $d = O(n^{\frac{1}{2}-\alpha})$ where $0 < \alpha \leq \frac{1}{2}$, this is asymptotically much better than the greedy bound $2n$.

Proof of Theorem 4.2.:

Define $f(x) := \sum_{i,j=1}^{n} a_{ij} x_i x_j,\ x \in \{-1, +1\}^n$.
If two vectors $x, x' \in \{-1, 1\}^n$ differs only in the k-th position, i.e. $x_k \neq x'_k$, then $|f(x) - f(x')| \leq d_k$.
Let (ξ_i) be random variables with $P(\xi_i = -1) = P(\xi_i = 1) = \frac{1}{2}$. Then the expectation $\mathbb{E}(f)$ is zero:

$$\mathbb{E}(f) = \mathbb{E}(\sum_{i,j=1}^{n} a_{ij}\xi_i\xi_j) = \sum_{i,j=1}^{n} a_{ij}\mathbb{E}(\xi_i)\mathbb{E}(\xi_j) = 0,$$

and each conditional expectation $\mathbb{E}(f|\omega_1, \ldots, \omega_i)$ can be computed in $O(n^2)$-time.
Taking $\delta = \frac{1}{2}$ and observing that $L = 1$, Corollary 3.16 (ii) concludes the proof.

□

Average Graph Bisection

Given a graph $G = (V, E),\ V = \{1, \ldots, 2n\}$, the Graph Bisection problem is to find a partition of V in two disjoint sets $A, B \subseteq V$ such that $|A| = |B| = n$, called a bisection (A, B), such that the cut $c(A, B)$, that is the number of edges between A and B, is minimal. The problem is known to be NP-hard [19] . Erdős [17] showed with simple probabilistic arguments the existence of a bisection (A, B) with cut value c such that $c \leq \frac{|E|}{2}(1 + o(1))$.

Theorem 4.6 *Let $G = (V, E)$ be a graph with $|V| = 2n$.*

(i) Derandomization finds a bisection $\omega \in \Omega$ with $c(\omega) \leq \frac{|E|}{2} + 8d\sqrt{n}$ in $O(n^3)$ time.

(ii) For dense graphs ($|E| = \Omega(n^{\frac{3}{2}+\alpha}),\ 0 < \alpha \leq \frac{1}{2}$) we have $c(\omega) \leq \frac{|E|}{2}(1 + o(1))$.

Proof. : (ii) follows directly from (i) so let us prove (i). Let $\Omega = \{0, 1\}^n$ and $A = (a_{ij})$ the adjacency matrix of G. Each $\omega \in \Omega$ defines via
$A := \{i; \omega_i = 0\},\ B = \{i; \omega_i = 1\}$ a partition of V. For $\omega \in \Omega$ define the cut function c and the counting function b by

$$c(\omega) := \sum_{i,j=1}^{2n} a_{ij} x_i (1 - x_j)$$

and

$$b(\omega) := \sum_{i=1}^{2n} \omega_i.$$

Let $d := \max_{1 \leq i \leq 2n} \sum_{j=1}^{2n} a_{ij}$, the maximal vertex degree.
Our randomized algorithm is simply to flip a fair coin $2n$-times independently in order to determine for each ω_i its value from $\{0, 1\}$ and make the sets $A = \{i; \omega_i = 1\}$, $B = \{i; \omega_i = 0\}$ equal sized, if they are not, in a linear time greedy way by shifting vertices from the bigger set to the smaller one.

Let $\omega, \omega' \in \omega$ with $\omega_i = \omega_i'$ for all $i \neq k$, but $\omega_k \neq \omega_k'$.
If d_k is the degree of the vertex k, then we have

$$|c(\omega) - c(\omega')| \leq d_k \leq d,$$

and

$$|b(\omega) - b(\omega')| \leq 1.$$

Furthermore $\mathbb{E}(c) = \frac{|E|}{2}$ and $\mathbb{E}(b) = n$.
Let $\epsilon_1 := d\sqrt{n \ln 12}$, $\epsilon_2 := 2\sqrt{n \ln 16}$ and let $\bar{E}^{(1)}$, $\bar{E}^{(2)}$ be the events

$$\bar{E}^{(1)} : \text{“}c > \frac{|E|}{2} + \epsilon_1\text{”}$$

and

$$\bar{E}^{(2)}(\epsilon_2) : \text{“}|b - n| > \epsilon_2\text{”}.$$

By Corollary 3.16 (i) (resp. (iii)) pessimistic estimators for the events $\bar{E}^{(1)}$ (resp.$\bar{E}^{(2)}$) can be computed in $O(n^2)$ (resp. $O(n)$) time. The sum of these two pessimistic estimators is according to Proposition 3.4 a pessimistic estimator for the event $\bar{E}^{(1)} \vee \bar{E}^{(2)}$ and the procedure D-WALK($\bar{E}^{(1)} \vee \bar{E}^{(2)}$) finds in $O(n^3)$ time an $\omega \in \Omega$ such that

$$c(\omega) \leq \frac{|E|}{2} + \epsilon_2 \leq \frac{|E|}{2} + 4d\sqrt{n}$$

and

$$|b(\omega) - n| \leq \epsilon_1 \leq 4\sqrt{n}.$$

After applying the linear time greedy procedure we make the sets $A := \{i;\ \omega_i = 0\}$ and $B := \{i;\ \omega_i = 1\}$ equal sized. Then at most $4d\sqrt{n}$ edges can augment the cut, hence we find in $O(n^3)$ time a *bisection* ω, with

$$c(\omega) \leq \frac{|E|}{2} + 8d\sqrt{n}.$$

The space needed is by Corollary 3.16 (i) $O(n^4)$.

□

5 Concluding Remarks

(a) Further applications of this type include the maximal *weighted* k-matching problem in hypergraphs. In [38] previous results of Lovász [25], Aharoni, Erdős and Linial [1] and Raghavan [31] are extended from the *unweighted* to the weighted case.

(b) In the average bisection problem we used the uniform distribution $P(\omega) = 2^{-2n}$ and the consequence was an approximation of $\frac{|E|}{2}$, which might be far away from the minimum bisection. Furthermore only partitioning into two sets was considered. But with the help of convex quadratic optimization a similar approach based on martingales gives good approximations of the optimal (bisection) partition size [37] .

It would be interesting to find other examples, where martingale based derandomization works.

(c) Can one derive in the quadratic lattice approximation problem an $O(\sqrt{n \ln n} + d\sqrt{n})$ (or even better) discrepancy bound for arbitrary matrices D ?

(d) Our algorithms are sequential. The interesting question here is, whether one can parallelize them as Berger/Rompel [12] and Motwani/Naor and Naor [28] showed for some linear problems. This might be possible using estimates on the variation of functions with small martingale differences.

Acknowledgement We would like to thank Joel Spencer for helpfull comments and discussions.

References

[1] R. Aharoni, P. Erdős, N. Linial; *Dual Integer Linear Programs and the Relationship between their Optima.* Proceedings of the 17th ACM Symposium on the Theory of Computing, ACM, New York (1985), 476-483.

[2] N. Alon, J. Spencer, P. Erdős; *The prababilistic method.* John Wiley & Sons, Inc. 1992.

[3] N. Alon; *A parallel algorithmic version of the Local Lemma.* Random Structures and Algorithms, Vol.2, No.4 (1991), 367-378.

[4] D. Angluin, L.G. Valiant: *Fast probabilistic algorithms for Hamiltonion circuits and matchings.* J. Computer and System Sciences, Vol. 18, (1979), 155–193.

[5] S. Arora, C. Lund, R. Motwani, M. Sudan,M. Szegedy: *On the intractability of approximation problems* (early draft), UC berkeley and Stanford University.

[6] K. Azuma, *Weighted sums of certain dependent variables.* Tohoku Math. Journ. 3, (1967), 357-367.

[7] I. Bárány Z. Füredi; *Computing the volume is difficult.* Proc. 18th Annual ACM Symposium on Theory of Computing (1986), 442-447.

[8] H. Bauer; *Probability theory and elements of measure theory.* Academic Press, (1981).

[9] J. Beck; *An algorithmic approach to the Lovász Local LemmaI.* Random Structures and Algorithms, Vol.2, No.4.,(1991), 343-365.

[10] J. Beck, Y. Fiala; *Integer-making theorems.* Discrete Appl. Math. 3 (1991), 1-8.

[11] J. Beck, J. Spencer; *Balancing Matrices with line shifts.* Combinatorica 3, Vol. 3-4, (1983), 299-304.

[12] B. Berger, J. Rompel; *Simulating ($log^c n$)-wise Independence in NC.* Proceeding of FOCS 1989, IEEE Coputer Society Press, Los Alamitos, CA, 2-8.

[13] B. Bollobás; *The chromatic number of random graphs.* Combinatorica 8, (1988), 49-56.

[14] H. Černov; *A measure of asymptotic efficiency for test of a hypothesis based on the sum of observation.* Ann. Math. Stat. 23, (1952), 493-509.

[15] M. E. Dyer, A. M. Frieze, R. Kannan; *A random polynomial time algorithm for approximating the volume of convex bodies.* Proc. 21st Annual Symposium of Theory of Computing (1989), 375-381.

[16] G. Elekes; *A geometric inequality and the complexity of computing volume.* Discrete Computational Geometry 1 (1986), 289-292.

[17] P. Erdős; *On bipartite subgraphs of graphs.* Math. Lapok 18, (1967), 283-288.

[18] A.M. Frieze, R. M. Karp, B. Reed; *When is the assignment bound tight for the asymetric traveling-salesman problem.* Preprint (1992), Research Institute of Discrete Mathematics, University of Bonn.

[19] M. R. Garey, D. S. Johnson, L. Stockmeyer; *Some simplified NP-complete graph problems.* Theoretical Computer Science 1, (1967), 237-267.

[20] M. Grötschel, L. Lovász, A. Schrijver; *Geometric algorithms and combinatorial optimixation.* Springer-Verlag (1988).

[21] W. Hoeffding; *On the distribution of the number of success in independent trials.* Annals of Math. Stat. 27, (1956), 713-721.

[22] M. R. Jerrum, A. J. Sinclair; *Approximating the permanent.* SIAM. J. Computing 18 (1989), 1149-1178.

[23] R. M. Karp, F. T. Leighton, R. L. Rivest, C. D. Thompson, U. V. Vazirani, V. V. Vazirani; *Global wire routing in two-dimensional arrays.* Proc. 24th. Annual Symposium on Foundations of Computer Science (1983), 453-459.

[24] R. M. Karp; *An introduction to randomized algorithms.* Discrete Appl Math. 34 (1991), 165-201.

[25] L. Lovász; *On the ratio of optimal and fractional covers.* Discrete Mathematics, 13 (1975), 383-390.

[26] K. Mehlhorn; *Data structures and algorithms 1: Sorting and Searching.* Sringer-Verlag (1984)

[27] C. McDiarmid; *On the Method of Bounded Differences.* Surveys in Combinatorics, 1989. J. Siemons, Ed.: London Math. Soc. Lectures Notes, Series 141, Cambridge University Press, Cambridge, England 1989. K. Mehlhorn; *Data structures and algorithms 1: Sorting and Searching.* Sringer-Verlag (1984).

[28] R. Motwani, J. Naor, M. Naor; *The probabilistic method yields deterministic parallel algorithms.* Proceedings 30the IEEE Conference on Foundation of Computer Science (FOCS'89), (1989), 8 –13.

[29] C. H. Papadimitriou, K. Steiglitz; *Combinatorial Optimization: Algorithms and Complexity.* Prentice-Hall, Englewood Cliffs NJ, (1982).

[30] P. Raghavan, C. D. Thompson; *Randomized Rounding: A technique for provably good algorithms and algorithmic proofs.* Combinatorica 7 (4), (1987), 365-374.

[31] P. Raghavan; *Probabilistic construction of deterministic algorithms: Approximating packing integer programs.* Jour. of Computer and System Sciences 37, (1988), 130-143.

[32] W. T. Rhee, M. Talagrand; *Martingale inequalities and NP-complete problems.* Mathematics of Operations Research, Vol 12, No. 1, (1987), 177-181.

[33] E. Shamir, J. Spencer; *Sharp concentration of the chromatic number of random graphs $G_{n,p}$.* Combinatorica 7, (1987), 121-129.

[34] A. J. Sinclair, M. R. Jerrum; *Approximate counting uniform generation and rapidly mixing Markov chains.* Information and Computation, Vol. 289, (1989), 93-133.

[35] J. Spencer; *Six standard deviation suffice.* Trans. Amer. Math. Society, Vol. 289, (1985), 679-706.

[36] J. Spencer; *Ten lectures on the probabilistic method.* SIAM, Philadelphia (1987).

[37] A. Srivastav, P. Stangier; *The relationship between fractional and integral graph partitioning.* Working Paper; Research Institute of Discrete Mathematics, University of Bonn, (1992).

[38] A. Srivastav, P. Stangier; *Weighted fractional and integral k-matching in hypergraphs.* Research Institute of Discrete Mathematics, Working Paper, University of Bonn (1992).