Probabilistic analysis of a differential equation for linear programming

Asa Ben-Hur[a,b], **Joshua Feinberg**[c,d], **Shmuel Fishman**[d,e]
& Hava T. Siegelmann[f]

[a)]Biochemistry Department,
Stanford University CA 94305
[b)]Faculty of Industrial Engineering and Management,
Technion, Haifa 32000, Israel.
[c)]Physics Department,
University of Haifa at Oranim, Tivon 36006, Israel
[d)]Physics Department,
Technion, Israel Institute of Technology, Haifa 32000, Israel
[e)]Institute for Theoretical Physics
University of California
Santa Barbara, CA 93106, USA
[f)]Laboratory of Bio-computation,
Department of Computer Science,
University of Massachusetts at Amherst,
Amherst, MA 01003

e-mail addresses: asa@barnhilltechnologies.com, joshua@physics.technion.ac.il, fishman@physics.technion.ac.il, hava@mit.edu
proposed running head: probabilistic analysis of a differential equation for LP
All correspondence should be sent to J. Feinberg. Tel. 972-4-8292041, Fax: 972-4-8221514

Abstract

In this paper we address the complexity of solving linear programming problems with a set of differential equations that converge to a fixed point that represents the optimal solution. Assuming a probabilistic model, where the inputs are i.i.d. Gaussian variables, we compute the distribution of the convergence rate to the attracting fixed point. Using the framework of Random Matrix Theory, we derive a simple expression for this distribution in the asymptotic limit of large problem size. In this limit, we find the surprising result that the distribution of the convergence rate is a scaling function of a single variable. This scaling variable combines the convergence rate with the problem size (i.e., the number of variables and the number of constraints). We also estimate numerically the distribution of the computation time to an approximate solution, which is the time required to reach a vicinity of the attracting fixed point. We find that it is also a scaling function. Using the problem size dependence of the distribution functions, we derive high probability bounds on the convergence rates and on the computation times to the approximate solution.

Keywords: Theory of Analog Computation, Dynamical Systems, Linear Programming, Scaling, Random Matrix Theory.

1 Introduction

In recent years scientists have developed new approaches to computation, some of them based on continuous time analog systems. Analog VLSI devices, that are often described by differential equations, have applications in the fields of signal processing and optimization. Many of these devices are implementations of neural networks [1, 2, 3], or the so-called neuromorphic systems [4] which are hardware devices whose structure is directly motivated by the workings of the brain. In addition there is an increasing number of algorithms based on differential equations that solve problems such as sorting [5], linear programming [6] and algebraic problems such as singular value decomposition and finding of eigenvectors (see [7] and references therein). On a more theoretical level, differential equations are known to simulate Turing machines [8]. The standard theory of computation and computational complexity [9] deals with computation in discrete time and in a discrete configuration space, and is inadequate for the description of such systems. This work may prove useful in the analysis and comparison of analog computational devices (see e.g. [3, 10]).

In a recent paper we have proposed a framework of analog computation based on ODE's that converge exponentially to fixed points [11]. In such systems it is natural to consider the *attracting fixed point as the output*. The input can be modeled in various ways. One possible choice is the initial condition. This is appropriate when the aim of the computation is to decide to which attractor out of many possible ones the system flows (see [12]). The main problem within this approach is related to initial conditions in the vicinity of basin boundaries. The flow in the vicinity of the boundary is slow, resulting in very long computation times. Here, as in [11] the parameters on which the vector field depends are the input, and the initial condition is part of the algorithm. This modeling is natural for optimization problems, where one wishes to find extrema of some function $E(x)$, e.g. by a gradient flow $\dot{x} = \mathrm{grad}E(x)$. An instance of the optimization problem is specified by the parameters of $E(x)$, i.e. by the parameters of the vector field.

The basic entity in our model of analog computation is a set of ODEs

$$\frac{dx}{dt} = F(x), \tag{1}$$

where x is an n-dimensional vector, and F is an n-dimensional smooth vector field, which converges exponentially to a fixed point. Eq. (1) solves a computational problem as follows: Given an instance of the problem, the parameters of the vector field F are set, and it is started from some pre-determined initial condition. The result of the computation is then deduced from the fixed point that the system approaches.

Even though the computation happens in a real configuration space, this model can be considered as either a model with real inputs, as for example the BSS model [13], or as a model with integer or rational inputs, depending what types of values the initial conditions are given. In [11] it was argued that the time complexity in a large class of ODEs is the physical time that is the time parameter of the system. The initial condition there was assumed to be integer or rational. In the present paper, on the other hand, we consider real inputs. More specifically, we will analyze the complexity of a flow for linear programming (LP) introduced in [6]. In the real number model the complexity of solving LP with interior point methods is unbounded [14], and a similar phenomenon occurs for the flow we analyze here. To obtain finite computation times one can either measure the computation time in

terms of a *condition number* as in [15], or impose a distribution over the set of LP instances. Many of the probabilistic models used to study the performance of the simplex algorithm and interior point methods assume a Gaussian distribution of the data [16, 17, 18], and we adopt this assumption for our model. Recall that the worst case bound for the simplex algorithm is exponential whereas some of the probabilistic bounds are quadratic [18].

Two types of probabilistic analysis were carried out in the LP literature: average case and "high probability" behavior [19, 20, 21]. A high probability analysis provides a bound on the computation time that holds with probability 1 as the problem size goes to infinity [21]. In a worst case analysis interior point methods generally require $\mathcal{O}(\sqrt{n}|\log\epsilon|)$ iterations to compute the cost function with ϵ-precision, where n is the number of variables [20]. The high probability analysis essentially sets a limit on the required precision and yields $\mathcal{O}(\sqrt{n}\log n)$ behavior [21]. However, the number of iterations has to be multiplied by the complexity of each iteration which is $\mathcal{O}(n^3)$, resulting in an overall complexity $\mathcal{O}(n^{3.5}\log n)$ in the high probability model [20]. The same factor per iteration appears in the average case analysis as well [19].

In contrast, in our model of analog computation, the computation time is the physical time required by a hardware implementation of the vector field F(x) to converge to the attracting fixed point. We need neither to follow the flow step-wise nor to calculate the vector field F(x) since it is assumed to be realized in hardware and does not require repetitive digital approximations. As a result, the complexity of analog processes does not include the $\mathcal{O}(n^3)$ term as above, and in particular it is lower than the digital complexity of interior point methods. In this set-up we conjecture, based on numerical calculations, that the flow analyzed in this paper has complexity $\mathcal{O}(n\log n)$ on average and with high probability. This is higher than the number of iterations of state of the art interior point methods, but lower than the overall complexity $\mathcal{O}(n^{3.5}\log n)$ of the high probability estimate mentioned above, which includes the complexity of an individual operation.

In this paper we consider a flow for linear programming proposed by Faybusovich [6], for which $F(x)$ is given by (4). Substituting (4) into the general equation (1) we obtain (5), which realizes the Faybusovich algorithm for LP. We consider real inputs that are drawn from a Gaussian probability distribution. For any feasible instance of the LP problem, the flow converges to the solution. We consider the question: Given the probability distribution of LP instances, what is the probability distribution of the convergence rates to the solution? The convergence rate measures the asymptotic computation time: the time to reach an ϵ vicinity of the attractor, where ϵ is arbitrarily small. The main result of this paper, as stated in Theorem (4.1), is that with high probability and on the average, the asymptotic computation time is $\mathcal{O}\left(\sqrt{n}|\log\epsilon|\right)$, where n is the problem size and ϵ is the required precision (see also Corollary (5.1)).

In practice, the solution to arbitrary precision is not always required, and one may need to know only whether the flow (1) or (5) has reached the vicinity of the optimal vertex, or which vertex out of a given set of vertices will be approached by the system. Thus, the non-asymptotic behavior of the flow needs to be considered [11]. In this case, only a heuristic estimate of the computation time is presented, and in Section 6 we conjecture that the associated complexity is $\mathcal{O}\left(n\log n\right)$, as mentioned above.

The rest of the paper is organized as follows: In section 2 the Faybusovich flow is presented along with an expression for its convergence rate. The probabilistic ensemble of the LP instances is presented in section 3. The distribution of the convergence rate of this

flow is calculated analytically in the framework of random matrix theory (RMT) in section 4. In secton 5 we introduce the concept of "high-probability behavior" and use the results of section 4 to quantify the high-probability behavior of our probabilistic model. In section 6 we provide measures of complexity when precision asymptotic in ϵ is not required. Some of the results in sections 6.2-8 are heuristic, supported by numerical evidence. The structure of the distribution functions of parameters that control the convergence is described in section 7 and its numerical verification is presented in section 8. Finally, the results of this work and their possible implications are discussed in section 9. Some technical details are relegated to the appendices. Appendix A contains more details of the Faybusovich flow. Appendix B exposes the details of the analytical calculation of the results presented in section 4, and appendix C contains the necessary details of random matrix theory relevant for that calculation.

2 A flow for linear programming

We begin with the definition of the linear programming problem (LP) and a vector field for solving it introduced by Faybusovich in [6]. The *standard form* of LP is to find

$$\max\{c^T x \ : \ x \in \mathbb{R}^n, Ax = b, x \geq 0\} \tag{2}$$

where $c \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in \mathbb{R}^{m\times n}$ and $m \leq n$. The set generated by the constraints in (2) is a polyheder. If a bounded optimal solution exists, it is obtained at one of its vertices. Let $\mathcal{B} \subset \{1,\dots,n\}$, $|\mathcal{B}| = m$, and $\mathcal{N} = \{1,\dots,n\} \setminus \mathcal{B}$, and denote by $x_\mathcal{B}$ the coordinates with indices from $\mathcal{B}$, and by $A_\mathcal{B}$, the $m \times m$ matrix whose columns are the columns of A with indices from $\mathcal{B}$. A vertex of the LP problem is defined by a set of indices $\mathcal{B}$, which is called a *basic set*, if

$$x_\mathcal{B} = A_\mathcal{B}^{-1} b \geq 0 \, . \tag{3}$$

The components of a vertex are $x_\mathcal{B}$ that satisfy (3), and $x_\mathcal{N} = 0$. The set $\mathcal{N}$ is then called a *non-basic* set. Given a vector field that converges to an optimal solution represented by basic and non-basic sets $\mathcal{B}$ and $\mathcal{N}$, its solution $x(t)$ can be decomposed as $(x_\mathcal{N}(t), x_\mathcal{B}(t))$ where $x_\mathcal{N}(t)$ converges to 0, and $x_\mathcal{B}(t)$ converges to $A_\mathcal{B}^{-1} b$.

In the following we consider the non-basic set $\mathcal{N} = \{1,\dots,n-m\}$, and for notational convenience denote the $m \times m$ matrix $A_\mathcal{B}$ by B and denote $A_\mathcal{N}$ by N, i.e. $A = (N, B)$.

The Faybusovich vector field is a projection of the gradient of the linear cost function onto the constraint set, relative to a Riemannian metric which enforces the positivity constraints $x \geq 0$ [6]. Let $h(x) = c^T x$. We denote this projection by $\operatorname{grad} h$. The explicit form of the gradient is:

$$\operatorname{grad} h(x) = [X - XA^T (AXA^T)^{-1} AX]\, c \, , \tag{4}$$

where X is the diagonal matrix $\mathrm{Diag}(x_1 \dots x_n)$. It is clear from (4) that

$$A \operatorname{grad} h(x) = 0 \, .$$

Thus, the dynamics

$$\frac{dx}{dt} = \operatorname{grad} h(x) \tag{5}$$

preserves the constraint $Ax = b$ in (2). Thus, the faces of the polyheder are invariant sets of the dynamics induced by $\operatorname{grad} h$. Furthermore, it is shown in [6] that the fixed points of $\operatorname{grad} h$ coincide with the vertices of the polyheder, and that the dynamics converges exponentially to the maximal vertex of the LP problem. Since the formal solution of the Faybusovich vector field is the basis of our analysis we give its derivation in Appendix A.

Solving (5) requires an appropriate initial condition - an interior point in this case. This can be addressed either by using the "big-M" method [22], which has essentially the same convergence rate, or by solving an auxiliary linear programming problem [21]. We stress that here, the initial interior point is not an input for the computation, but rather a part of the algorithm. In the analog implementation the initial point should be found by the same device used to solve the LP problem.

The linear programming problem (2) has $n - m$ independent variables. The formal solution shown below, describes the time evolution of the $n-m$ variables $x_{\mathcal{N}}(t)$, in terms of the variables $x_{\mathcal{B}}(t)$. When $\mathcal{N}$ is the non-basic set of an optimal vertex of the LP problem, $x_{\mathcal{N}}(t)$ converges to 0, and $x_{\mathcal{B}}(t)$ converges to $A_{\mathcal{B}}^{-1}b$. Denote by $e^1, \ldots, e^n$ the standard basis of $\mathbb{R}^n$, and define the $n - m$ vectors

$$\mu^i = e^i + \sum_{j=1}^{m} \alpha_{ji} e^j \ , \tag{6}$$

where

$$\alpha_{ji} = -(B^{-1}N)_{ji} \tag{7}$$

is an $m \times (n - m)$ matrix. The vectors μ^i are perpendicular to the rows of A and are parallel to the faces of the polyheder defined by the constraints. In this notation the analytical solution is (see Appendix A):

$$x_i(t) = x_i(0) \exp\left(-\Delta_i t - \sum_{j=1}^{m} \alpha_{ji} \log \frac{x_{j+n-m}(t)}{x_{j+n-m}(0)} \right) \ , \quad i \in \mathcal{N} = \{1, \ldots, n-m\} \tag{8}$$

where $x_i(0)$ and $x_{j+n-m}(0)$ are components of the initial condition, $x_{j+n-m}(t)$ are the $x_{\mathcal{B}}$ components of the solution, and

$$\Delta_i = - < \mu^i, c > = -c_i - \sum_{j=1}^{m} c_j \alpha_{ji} \tag{9}$$

(where $< ., . >$ is the Euclidean inner product).

An important property which relates the signs of the Δ_i and the optimality of the partition of A (into (B, N)) relative to which they were computed is now stated:

Lemma 2.1 [6] For a polyhedron with $\{n-m+1, \ldots, n\}$, a basic set of a maximum vertex,

$$\Delta_i \geq 0 \quad i = 1, \ldots, n - m \ .$$

The converse statement does not necessarily hold. The Δ_i are independent of b. Thus we may have that all Δ_i are positive, and yet the constraint set is empty.

Remark 2.1 Note that the analytical solution is only a formal one, and does not provide an answer to the LP instance, since the Δ_i depend on the partition of A, and only relative to a partition corresponding to a maximum vertex are all the Δ_i positive.

The quantities Δ_i are the convergence rates of the Faybusovich flow, and thus measure the time required to reach the ϵ-vicinity of the optimal vertex, where ϵ is arbitrarily small:

$$T_\epsilon \sim \frac{|\log \epsilon|}{\Delta_{min}}\,, \tag{10}$$

where

$$\Delta_{\min} = \min_i \Delta_i\ . \tag{11}$$

Therefore, if the optimal vertex is required with arbitrary precision ϵ , then the computation time (or complexity) is $\mathcal{O}\left(\Delta_{min}^{-1}|\log\epsilon|\right)$.

In summary, if the Δ_i are small then large computation times will be required. The Δ_i can be arbitrarily small when the inputs are real numbers, resulting in an unbounded computation time. However, we will show that in the probabilistic model, which we define in the next section, "bad" instances are rare, and the flow performs well "with high probability" (see Theorem (4.1) and Corollary (5.1)).

3 The probabilistic model

We now define the ensemble of LP problems for which we analyze the complexity of the Faybusovich flow. Denote by $N(0,\sigma^2)$ the standard Gaussian distribution with 0 mean and variance σ^2. Consider an ensemble in which the components of (A,b,c) are i.i.d. (independent identically distributed) random variables with the distribution $N(0,\sigma^2)$. The model will consist of the following set of problems:

$$\begin{aligned} LPM \ = \ \{(A,b,c) \quad | \quad & (A,b,c) \text{ are i.i.d. variables with the distribution } N(0,\sigma^2) \\ & \text{and the LP problem has a bounded optimal solution}\}\,. \end{aligned} \tag{12}$$

Therefore, we use matrices with a distribution $N(0,\sigma^2)$:

$$f(A) = \frac{1}{\mathcal{Z}_A}\exp\left(-\frac{1}{2\sigma^2}\mathrm{tr}A^TA\right) \tag{13}$$

with normalization

$$\mathcal{Z}_A = \int d^{mn}A\ \exp\left(-\frac{1}{2\sigma^2}\mathrm{tr}A^TA\right) = \left(2\pi\sigma^2\right)^{mn/2}\ . \tag{14}$$

The ensemble (13) factorizes into mn i.i.d. Gaussian random variables for each of the components of A.

The distributions of the vectors c and b are defined by:

$$f(c) = \frac{1}{\mathcal{Z}_c}\exp\left(-\frac{1}{2\sigma^2}c^Tc\right) \tag{15}$$

with normalization

$$\mathcal{Z}_c = \int d^nc\ \exp\left(-\frac{1}{2\sigma^2}c^Tc\right) = \left(2\pi\sigma^2\right)^{n/2}\ , \tag{16}$$

and

$$f(b) = \frac{1}{\mathcal{Z}_b} \exp\left(-\frac{1}{2\sigma^2} b^T b\right) \tag{17}$$

with normalization

$$\mathcal{Z}_b = \int d^m b \ \exp\left(-\frac{1}{2\sigma^2} b^T b\right) = \left(2\pi\sigma^2\right)^{m/2} . \tag{18}$$

With the introduction of a probabilistic model of LP instances $\Delta_{\min}$ becomes a random variable. We wish to compute the probability distribution of $\Delta_{\min}$ for instances with a bounded solution, when $\Delta_{\min} > 0$. We reduce this problem to the simpler task of computing $\mathcal{P}(\Delta_{min} > \Delta | \Delta_{min} > 0)$, in which the condition $\Delta_{min} > 0$ is much easier to impose than the condition that an instance produces an LP problem with a bounded solution. This reduction is justified by the following lemma:

Lemma 3.1

$$\mathcal{P}(\Delta_{\min} > \Delta | \text{LP instance has a bounded maximum vertex}) = \mathcal{P}(\Delta_{\min} > \Delta | \Delta_{\min} > 0). \tag{19}$$

Proof. Let (A, b, c) be an LP instance chosen according to the probability distributions (13), (15) and (17). There is a *unique* orthant (out of the 2^n orthants) where the constraint set $Ax = b$ defines a nonempty polyheder. This orthant is not necessarily the positive orthant, as in the standard formulation of LP.

Let us consider now *any* vertex of this polyheder with basic and non-basic sets $\mathcal{B}$ and $\mathcal{N}$. Its m non-vanishing coordinates $x_{\mathcal{B}}$ are given by solving $A_{\mathcal{B}} x_{\mathcal{B}} = b$. The matrix $A_{\mathcal{B}}$ is full rank with probability 1; also, the components of $x_{\mathcal{B}}$ are non-zero and finite with probability 1. Therefore, in the probabilistic analysis we can assume that $x_{\mathcal{B}}$ is well defined and non-zero. With this vertex we associate the $n - m$ quantities $\Delta_i = -(c_{\mathcal{N}})_i + (c_{\mathcal{B}}^T A_{\mathcal{B}}^{-1} A_{\mathcal{N}})_i$, from (9).

We now show that there is a set of 2^m *equiprobable* instances, which contains the instance (A, b, c), that shares the same vector b and the same values of $\{\Delta_i\}$, when computed according to the given partition. This set contains a unique instance with $x_{\mathcal{B}}$ in the positive orthant. Thus, if $\Delta_{min} > 0$, the latter instance will be the unique member of the set which has a bounded optimal solution.

To this end, consider the set $\mathcal{R}(x_{\mathcal{B}})$ of the 2^m reflections $\mathcal{Q}_l x_{\mathcal{B}}$ of $x_{\mathcal{B}}$, where $\mathcal{Q}_l$ is an $m \times m$ diagonal matrix with diagonal entries ± 1 and $l = 1, 2, ..., 2^m$.

Given the instance (A, b, c) and a particular partition into basic and non-basic sets, we split A columnwise into $(A_{\mathcal{B}}, A_{\mathcal{N}})$ and c into $(c_{\mathcal{B}}, c_{\mathcal{N}})$. Let $\mathcal{S}$ be the set of 2^m instances $((A_{\mathcal{B}} \mathcal{Q}_l, A_{\mathcal{N}}), b, (\mathcal{Q}_l c_{\mathcal{B}}, c_{\mathcal{N}}))$ where $l = 1, \ldots, 2^m$. The vertices $\mathcal{Q}_l x_{\mathcal{B}}$ of these instances, which correspond to the prescribed partition, comprise the set $\mathcal{R}(x_{\mathcal{B}})$, since $(A_{\mathcal{B}} \mathcal{Q}_l)(\mathcal{Q}_l x_{\mathcal{B}}) = b$. Furthermore, all elements in $\mathcal{R}(x_{\mathcal{B}})$ (each of which corresponds to a different instance) have the same set of Δ's, since $\Delta_i = -(c_{\mathcal{N}})_i + [(\mathcal{Q}_l c_{\mathcal{B}})^T (A_{\mathcal{B}} \mathcal{Q}_l)^{-1} A_{\mathcal{N}}]_i$. Because of the symmetry of the ensemble under the reflections $\mathcal{Q}_l$, the probability of all instances in $\mathcal{S}$ is the same.

All the vertices belonging to $\mathcal{R}(x_{\mathcal{B}})$ have the same Δ_i's with the same probability, and exactly one is in the positive orthant. Thus, if $\Delta_{\min} > 0$, the latter vertex is the unique element from $\mathcal{S}$ which is the optimal vertex of an LP problem with a bounded solution.

Consequently, the probability of having any prescribed set of Δ_i's, and in particular, the probability distribution for the Δ_i's given $\Delta_{min} > 0$, is not affected by the event that the LP instance has a bounded optimal solution (i.e., that the vertex is in the positive orthant). In other words, these are independent events. Integration over all instances and taking this way into account all possible sets $\mathcal{S}$ while imposing the requirement $\{\Delta_{\min} > \Delta | \Delta_{\min} > 0\}$ results in (19). ∎

The event $\Delta_{\min} > 0$ corresponds to a specific partition of A into basic and non-basic sets $\mathcal{B}, \mathcal{N}$, respectively. It turns out that it is much easier to analytically calculate the probability distribution of Δ_{min} for a given partition of the matrix A. It will be shown in what follows that in the probabilistic model we defined, $\mathcal{P}(\Delta_{\min} > \Delta | \Delta_{\min} > 0)$ is proportional to the probability that $\Delta_{\min} > \Delta$ for a fixed partition. Let W_j be the event that a partition j of the matrix A is an optimal partition, i.e. all Δ_i are positive (j is an index with range $1, \ldots, \binom{n}{m}$). Let the index 1 stand for the partition where B is taken from the last m columns of A. We now show:

Lemma 3.2 Let $\Delta > 0$ then

$$\mathcal{P}(\Delta_{min} > \Delta | \Delta_{min} > 0) = \mathcal{P}(\Delta_{min} > \Delta | W_1) \ .$$

Proof. Given that $\Delta_{\min} > 0$, there is a unique optimal partition since a non-unique optimal partition occurs only if c is orthogonal to some face of the polyhedron, in which case $\Delta_i = 0$ for some i. Thus we can write:

$$\begin{aligned} \mathcal{P}(\Delta_{min} > \Delta | \Delta_{min} > 0) &= \sum_j \mathcal{P}(\Delta_{min} > \Delta | \Delta_{min} > 0, W_j) \mathcal{P}(W_j) \qquad (20) \\ &= \sum_j \mathcal{P}(\Delta_{min} > \Delta | W_j) \mathcal{P}(W_j) \ , \qquad (21) \end{aligned}$$

where the second equality holds since the event W_j is contained in the event that $\Delta_{min} > 0$. The probability distribution of (A, c) is invariant under permutations of columns of A and c, and under permutations of rows of A. Therefore the probabilities $\mathcal{P}(W_j)$ are all equal, and so are $\mathcal{P}(\Delta_{min} > \Delta | \Delta_{min} > 0, W_j)$, and the result follows.

∎

We define

$$\Delta_{min1} = \min\{\Delta_i \mid \Delta_i \text{ are computed relative to the partition 1}\} \qquad (22)$$

Note that the definition of $\Delta_{\min}$ in equation (11) is relative to the optimal partition. To show that all computations can be carried out for a fixed partition of A we need the next lemma:

Lemma 3.3 Let $\Delta > 0$ then

$$\mathcal{P}(\Delta_{\min} > \Delta | \Delta_{\min} > 0) = \frac{\mathcal{P}(\Delta_{min1} > \Delta)}{\mathcal{P}(\Delta_{min1} > 0)} \ .$$

Proof. The result follows from

$$\mathcal{P}(\Delta_{\min} > \Delta | W_1) = \mathcal{P}(\Delta_{min1} > \Delta | \Delta_{min1} > 0), \tag{23}$$

combined with the result of the previous lemma and the definition of conditional probability. ∎

In view of the symmetry of the joint probability distribution (j.p.d.) of $\Delta_1, \ldots, \Delta_{n-m}$, given by (28) and (32), the normalization constant $\mathcal{P}(\Delta_{min1} > 0)$ satisfies:

$$\mathcal{P}(\Delta_{min1} > 0) = 1/2^{n-m} \, . \tag{24}$$

Remark 3.1 Note that we are assuming throughout this work, that the optimal vertex is unique, i.e., given a partition $(\mathcal{N}, \mathcal{B})$ of A that corresponds to an optimal vertex, the basic components are all non-zero. The reason is that if one of the components of the optimal vertex vanishes, all of its permutations with the $n-m$ components of the non-basic set result in the same value of $c^T x$. Vanishing of one of the components of the optimal vertex requires that b is a linear combination of columns of A, that is an event of zero measure in our probabilistic ensemble. Therefore this case will not be considered in the present work.

4 Computing the distributions of Δ_{min1} and of Δ_{min}

In the following we compute first the distribution of Δ_{min1} and use it to obtain the distribution of Δ_{min} via Lemma (3.3). We denote the first $n-m$ components of c by y, and its last m components by z. In this notation equation (9) for Δ_i takes the form:

$$\Delta_p = -y_p + (z^T B^{-1} N)_p \quad p = 1, \ldots, n-m \, . \tag{25}$$

Our notation will be such that indices

$$i, j, k, \ldots \qquad \text{range over} \qquad 1, 2, \ldots, m$$

and

$$p, q, \ldots \qquad \text{range over} \qquad 1, 2, \ldots, n-m.$$

In this notation, the ensembles (13) and (15) may be written as

$$\begin{aligned} f(A) &= f(N, B) = \frac{1}{\mathcal{Z}_A} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{ij} B_{ij}^2 + \sum_{ip} N_{ip}^2 \right) \right] \\ f(c) &= f(y, z) = \frac{1}{\mathcal{Z}_c} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{i} z_i^2 + \sum_{p} y_p^2 \right) \right] . \end{aligned} \tag{26}$$

We first compute the joint probability distribution (j.p.d.) of $\Delta_1, \ldots, \Delta_{n-m}$ relative to the partition 1. This is denoted by $f_1(\Delta_1, \ldots, \Delta_{n-m})$. Using (25), we write

$$\begin{aligned} f_1(\Delta_1, \ldots, \Delta_{n-m}) &= \int d^{m^2} B \, d^{m(n-m)} N \, d^m z \, d^{n-m} y \\ & \qquad f(N, B) f(y, z) \prod_{q=1}^{n-m} \delta \left(\Delta_q + y_q - \sum_{i,j=1}^{m} z_j (B^{-1})_{ji} N_{iq} \right) , \end{aligned} \tag{27}$$

where $\delta(x)$ is the Dirac delta function. We note that this j.p.d. is not only completely symmetric under permuting the Δ_p's, but is also *independent* of the partition relative to which it is computed.

We would like now to perform the integrals in (27) and obtain a more explicit expression for $f_1(\Delta_1, \ldots, \Delta_{n-m})$. It turns out that direct integration over the y_q's, using the δ function, is not the most efficient way to proceed. Instead, we represent each of the δ functions as a Fourier integral. Thus,

$$\begin{aligned} f_1(\Delta_1, \ldots, \Delta_{n-m}) &= \int d^{m^2}B\, d^{m(n-m)}N\, d^m z\, d^{n-m}y\, \frac{d^{n-m}\lambda}{(2\pi)^{n-m}}\, f(N,B) f(y,z) \\ &\quad \cdot \exp\left[i \sum_q \lambda_q \left(\Delta_q + y_q - \sum_{i,j=1}^{m} z_j (B^{-1})_{ji} N_{iq} \right) \right] . \end{aligned}$$

Integration over N_{ip}, λ_q and y_p is straight forward and yields

$$\begin{aligned} f_1(\Delta_1, \ldots, \Delta_{n-m}) = & \left(\tfrac{1}{2\pi\sigma^2}\right)^{\frac{m^2+n}{2}} \int \frac{d^{m^2}B\, d^m z}{[z^T (B^T B)^{-1} z + 1]^{\frac{n-m}{2}}} \\ & \cdot \exp\left[-\tfrac{1}{2\sigma^2} \left(\textstyle\sum_{ij} B_{ij}^2 + \sum_i z_i^2 + \frac{\sum_p \Delta_p^2}{z^T (B^T B)^{-1} z + 1} \right) \right] . \end{aligned} \tag{28}$$

Here the complete symmetry of $f_1(\Delta_1, \ldots, \Delta_{n-m})$ under permutations of the Δ_p's is explicit, since it is a function of $\sum_p \Delta_p^2$.

The integrand in (28) contains the combination

$$u(B,z) = \frac{1}{z^T (B^T B)^{-1} z + 1} . \tag{29}$$

Obviously, $0 \le u(B,z) \le 1$. It will turn out to be very useful to consider the distribution function $P(u)$ of the random variable $u = u(B,z)$, namely,

$$P(u) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m^2+m}{2}} \int d^{m^2}B\, d^m z\, e^{-\frac{1}{2\sigma^2}\left(\mathrm{tr}\, B^T B + z^T z\right)} \cdot \delta\left(u - \frac{1}{z^T (B^T B)^{-1} z + 1} \right) . \tag{30}$$

Note from (29) that $u(\lambda B, \lambda z) = u(B,z)$. Thus, in fact, $P(u)$ is independent of the (common) variance σ of the Gaussian variables B and z, and we might as well rewrite (30) as

$$P(u) = \left(\frac{\lambda}{\pi}\right)^{\frac{m^2+m}{2}} \int d^{m^2}B\, d^m z\, e^{-\lambda\left(\mathrm{tr}\, B^T B + z^T z\right)} \cdot \delta\left(u - \frac{1}{z^T (B^T B)^{-1} z + 1} \right) , \tag{31}$$

with $\lambda > 0$ an arbitrary parameter.

Thus, if we could calculate $P(u)$ explicitly, we would be able to express the j.p.d. $f_1(\Delta_1, \ldots, \Delta_{n-m})$ in (28) in terms of the one dimensional integral

$$f_1(\Delta_1, \ldots, \Delta_{n-m}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n-m}{2}} \int_0^\infty du\, P(u)\, u^{\frac{n-m}{2}}\, \exp\left[-\frac{1}{2\sigma^2} \left(u \sum_{p=1}^{n-m} \Delta_p^2 \right) \right] , \tag{32}$$

as can be seen by comparing (28) and (30).

In this paper we are interested mainly in the minimal Δ. Thus, we need $f_{min1}(\Delta)$, the probability density of $\Delta_{\min 1}$. Due to the symmetry of $f_1(\Delta_1, \ldots, \Delta_{n-m})$, which is explicit in (32), we can express $f_{min1}(\Delta)$ simply as

$$f_{min1}(\Delta) = (n-m) \int_{\Delta}^{\infty} d\Delta_2 \ \ldots d\Delta_{n-m} f_1(\Delta, \Delta_2, \ldots, \Delta_{n-m}) \,. \tag{33}$$

It will be more convenient to consider the complementary cumulative distribution (c.c.d.)

$$\mathcal{Q}(\Delta) = \mathcal{P}(\Delta_{min1} > \Delta) = \int_{\Delta}^{\infty} f_{min1}(u) du \,, \tag{34}$$

in terms of which

$$f_{min1}(\Delta) = -\frac{\partial}{\partial \Delta} \mathcal{Q}(\Delta) \ . \tag{35}$$

The c.c.d. $\mathcal{Q}(\Delta)$ may be expressed as a symmetric integral

$$\mathcal{Q}(\Delta) = \int_{\Delta}^{\infty} d\Delta_1 \ \ldots d\Delta_{n-m} f_1(\Delta_1, \Delta_2, \ldots, \Delta_{n-m}) \tag{36}$$

over the Δ's, and thus, it is computationally a more convenient object to consider than $f_{min1}(\Delta)$.

From (36) and (32) we obtain that

$$\mathcal{Q}(\Delta) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n-m}{2}} \int_{0}^{\infty} du \, P(u) \, \left(\sqrt{u} \int_{\Delta}^{\infty} dv \, e^{-\frac{1}{2\sigma^2} uv^2} \right)^{n-m} \,, \tag{37}$$

and from (37) one readily finds that

$$\mathcal{Q}(0) = \frac{1}{2^{n-m}} \,, \tag{38}$$

(as well as $\mathcal{Q}(-\infty) = 1$, by definition of $\mathcal{Q}$).

Then, use of the integral representation

$$1 - \mathrm{erf}(x) = \mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} dv \ e^{-v^2} \,, \qquad (x > 0) \tag{39}$$

and (38) leads (for $\Delta > 0$) to

$$\mathcal{Q}(\Delta) = \mathcal{Q}(0) \int_{0}^{\infty} du \, P(u) \, \left(\mathrm{erfc} \, [\Delta \sqrt{\frac{u}{2\sigma^2}} \,] \right)^{n-m} \,. \tag{40}$$

This expression is an *exact* integral representation of $\mathcal{Q}(\Delta)$ (in terms of the yet undetermined probability distribution $P(u)$).

In order to proceed, we have to determine $P(u)$. Determining $P(u)$ for any pair of integers (n, m) in (31) in a closed form is a difficult task. However, since we are interested mainly in the asymptotic behavior of computation times, we will contend ourselves in analyzing the behavior of $P(u)$ as $n, m \to \infty$, with

$$r \equiv m/n < 1 \tag{41}$$

held fixed.

We were able to determine the large n, m behavior of $P(u)$ (and thus of $f_1(\Delta_1, \Delta_2, \ldots, \Delta_{n-m})$ and $\mathcal{Q}(\Delta)$) using standard methods [23, 24] of random matrix theory [25].

This calculation is presented in detail in Appendix B. We show there (see Eq. (125)) that the leading asymptotic behavior of $P(u)$ is

$$P(u) = \sqrt{\frac{m}{2\pi u}}\, e^{-\frac{mu}{2}}\,, \tag{42}$$

namely, $\sqrt{u}$ is simply a Gaussian variable, with variance proportional to $1/\sqrt{m}$. Note that (42) is independent of the width σ, which is consistent with the remark preceding (31).

Substituting (42) in (32), we obtain, with the help of the integral representation

$$\Gamma(z) = \int_0^\infty t^{z-1}\, e^{-t}\, dt \tag{43}$$

of the Γ function, the large n, m behavior of the j.p.d. $f_1(\Delta_1, \ldots, \Delta_{n-m})$ as

$$f_1(\Delta_1, \ldots, \Delta_{n-m}) = \sqrt{m}\sigma\, \Gamma\left(\frac{n-m+1}{2}\right) \left(\frac{1}{\pi}\, \frac{1}{m\sigma^2 + \sum_p \Delta_p^2}\right)^{\frac{n-m+1}{2}}\,. \tag{44}$$

Thus, the Δ's follow asymptoticly a multi-dimensional Cauchy distribution. It can be checked that (44) is properly normalized to 1.

Similarly, by substituting (42) in (40), and changing the variable to $y = \sqrt{mu/2}$, we obtain the large n, m behavior of $\mathcal{Q}(\Delta)$ as

$$\mathcal{Q}(\Delta) = \frac{2\mathcal{Q}(0)}{\sqrt{\pi}} \int_0^\infty dy\, e^{-y^2} \left(\mathrm{erfc}\, \left[\Delta\, \frac{y}{\sqrt{m}\sigma}\, \right]\right)^{n-m}\,. \tag{45}$$

As a consistency check of our large n, m asymptotic expressions, we have verified, with the help of (43), that substituting (44) into (36) leads to (40), with $P(u)$ there given by the asymptotic expression (42).

We are interested in the scaling behavior of $\mathcal{Q}(\Delta)$ in (45) in the limit $n, m \to \infty$. In this large n, m limit, the factor $\left(\mathrm{erfc}\, \left[\Delta\, \frac{y}{\sqrt{m}\sigma}\, \right]\right)^{n-m}$ in (45) decays rapidly to zero. Thus, the integral in (45) will be appreciably different from zero only in a small region around $\Delta = 0$, where the erfc function is very close to 1. More precisely, using $\mathrm{erfc}\, x = 1 - \frac{2x}{\sqrt{\pi}} + \mathcal{O}(x^2)$, we may expand the erfc term in (45) as

$$\left(\mathrm{erfc}\, \left[\Delta\, \frac{y}{\sqrt{m}\sigma}\, \right]\right)^{n-m} = \left(1 - \frac{2y\Delta}{\sqrt{\pi m \sigma^2}} + \cdots\right)^{n-m} \tag{46}$$

(due to the Gaussian damping factor in (45), this expansion is uniform in y). Thus, we see that $\mathcal{Q}(\Delta)/\mathcal{Q}(0)$ will be appreciably different from zero only for values of Δ/σ of the order up to $1/\sqrt{m}$, for which (46) exponentiates into a quantity of $\mathcal{O}(1)$, and thus

$$\mathcal{Q}(\Delta) \simeq \frac{2\mathcal{Q}(0)}{\sqrt{\pi}} \int_0^\infty dy\, e^{-y^2} \exp\left(-\frac{2}{\sqrt{\pi}}\left(\frac{n}{m}-1\right) y\delta\right), \tag{47}$$

where

$$\delta = \frac{\sqrt{m}\Delta}{\sigma} \tag{48}$$

is $\mathcal{O}(m^0)$. Note that m/n is kept finite and fixed. The integral in (47) can be done, and thus we arrive at the explicit scaling behavior of the c.c.d.

$$\mathcal{Q}(\Delta) = \mathcal{Q}(0)\, e^{x_\Delta^2} \operatorname{erfc}(x_\Delta), \tag{49}$$

where

$$x_\Delta = \eta_\Delta(n,m)\Delta, \tag{50}$$

with

$$\eta_\Delta(n,m) = \frac{1}{\sqrt{\pi}}\left(\frac{n}{m}-1\right)\frac{\sqrt{m}}{\sigma}. \tag{51}$$

The c.c.d. $\mathcal{Q}(\Delta)$ depends, in principle, on all the three variables n, m and Δ. The result (49) demonstrates, that in the limit $(n,m) \to \infty$ (with $r = m/n$ held finite and fixed), $\mathcal{Q}(\Delta)$ is a function only of *one scaling variable*: the x_Δ defined in (50).

We have compared (49) and (50) against results of numerical simulations, for various values of n/m. The results are shown in Figures 2 and 3 in Section 8.

Establishing the explicit scaling expression of the probability distribution of the convergence rate constitutes the main result in our paper, which we summarize by the following Theorem:

Theorem 4.1 *Assume that LP problems of the form (2), with the instances distributed according to (13)-(18), are solved by the Faybusovich algorithm (5). Then, in the asymptotic limit $n \to \infty$, $m \to \infty$ with $0 < r = m/n < 1$ kept fixed, the convergence rate Δ_{min} defined by (11) is distributed according to*

$$\mathcal{P}(\Delta_{min} > \Delta | \textit{bounded optimal solution}) = e^{x_\Delta^2} \operatorname{erfc}(x_\Delta), \tag{52}$$

where x_Δ is given by (50).

Proof. $\mathcal{Q}(\Delta) = \mathcal{P}(\Delta_{min1} > \Delta)$ by (34). Therefore, use of (24) and (38), namely,

$$\mathcal{P}(\Delta_{min1} > 0) = \mathcal{Q}(0) = 1/2^{n-m}$$

, and of (49) implies

$$\mathcal{P}(\Delta_{min1} > \Delta) = \frac{1}{2^{n-m}}\, e^{x_\Delta^2} \operatorname{erfc}(x_\Delta), \tag{53}$$

but according to Lemma (3.3),

$$\mathcal{P}(\Delta_{\min} > \Delta | \Delta_{\min} > 0) = \frac{\mathcal{P}(\Delta_{min1} > \Delta)}{\mathcal{P}(\Delta_{min1} > 0)}.$$

Finally, substituting (53) and (24) in the last equation, and use of Lemma (3.1), leads to the statement of the theorem. ∎

From (49) and (50), we can obtain the probability density $f_{min1}(\Delta)$ of $\Delta_{\min 1}$, using (35). In particular, we find

$$f_{min1}(0) = \frac{2\sqrt{m}}{\pi\sigma}\left(\frac{n}{m} - 1\right)\mathcal{Q}(0)\,, \tag{54}$$

which coincides with the expression one obtains for $f_{min1}(0)$ by directly substituting the large (n, m) expression (45) into (35), without first going to the scaling regime $\Delta \sim 1/\sqrt{m}$, where (49) holds.

5 High-probability behavior

In this paper we show that the Faybusovich vector field performs well with high probability, a term that is explained in what follows. Such an analysis was carried out for interior point methods e.g. in [21, 26]. When the inputs of an algorithm have a probability distribution, $\Delta_{\min}$ becomes a random variable. High probability behavior is defined as follows:

Definition 5.1 Let T_n be a random variable associated with problems of size n. We say that $T(n)$ is a *high probability bound* on T_n if for $n \to \infty$ $T_n \leq T(n)$ with probability one.

To show that $1/\Delta_{\min} < \eta(m)$ with high probability is the same as showing that $\Delta_{\min} > 1/\eta(m)$ with high probability. Let $f_{\min}^{(m)}(\Delta|\Delta_{\min} > 0)$ denote the probability density of $\Delta_{\min}$ given $\Delta_{\min} > 0$. The m superscript is a mnemonic for its dependence on the problem size. We make the following observation:

Lemma 5.1 Let $\mathcal{P}(\Delta_{\min} > x|\Delta_{\min} > 0)$ be analytic in x around $x = 0$. Then, $\Delta_{\min} > \left[f_{\min}^{(m)}(0|\Delta_{\min} > 0)g(m)\right]^{-1}$ with high-probability, where $g(m)$ is any function such that $\lim_{m\to\infty} g(m) = \infty$.

Proof. For very small x we have:

$$\mathcal{P}(\Delta_{\min} > x|\Delta_{\min} > 0) \approx 1 - f_{\min}^{(m)}(0|\Delta_{\min} > 0)x\ . \tag{55}$$

We look for $x = x(m)$ such that $\mathcal{P}(\Delta_{\min} > x(m)|\Delta_{\min} > 0) = 1$ with high probability. For this it is sufficient that

$$\lim_{m\to\infty} f_{\min}^{(m)}(0|\Delta_{\min} > 0)x(m) = 0 \tag{56}$$

This holds if

$$x(m) = \left[f_{\min}^{(m)}(0|\Delta_{\min} > 0)g(m)\right]^{-1}\ , \tag{57}$$

where $g(m)$ is any function such that $\lim_{m\to\infty} g(m) = \infty$. ∎

The growth of $g(m)$ can be arbitrarily slow, so from this point on we will ignore this factor. As a corollary to Theorem (4.1) and (54) we now obtain:

Corollary 5.1 Let (A, b, c) be linear programming instances distributed according to (12) then

$$\frac{1}{\Delta_{\min}} = \mathcal{O}(m^{1/2}) \qquad \text{and} \qquad T_\epsilon = \mathcal{O}(m^{1/2}) \tag{58}$$

with high probability.

Proof. According to the results of Section 4, (and more explicitly, from the derivation of (86) in Section 7), $f_{\min}^{(m)}(0|\Delta_{\min} > 0) \sim m^{1/2}$, and the result follows from lemma (5.1) and the definition of T_ϵ (equation (10)). ∎

Remark 5.1 Note that bounds obtained in this method are tight, since they are based on the actual distribution of the data.

Remark 5.2 Note that $f_{\min}^{(m)}(0|\Delta_{\min} > 0) \neq 0$. Therefore, the $\frac{1}{\Delta}$ moment of the probability density function $f_{\min}^{(m)}(\Delta|\Delta_{\min} > 0)$ does not exist.

6 Measures of complexity in the non-asymptotic regime

In some situations one wants to identify the optimal vertex with limited precision.

The term

$$\beta_i(t) = -\sum_{j=1}^{m} \alpha_{ji} \log \frac{x_{j+n-m}(t)}{x_{j+n-m}(0)} \tag{59}$$

in (8), when it is positive, is a kind of "barrier": $\Delta_i t$ in equation (8) must be larger than the barrier before x_i can decrease to zero.

In this section we discuss heuristically the behavior of the barrier $\beta_i(t)$ as the dynamical system flows to the optimal vertex. To this end, we first discuss in rhe following sub-section some relevant probabilistic properties of the vertices of polyheders in our ensemble.

6.1 The typical magnitude of the coordinates of vertices

The flow (5) conserves the constraint $Ax = b$ in (2). Let us split these equations according to the basic and non-basic sets which corresponding to an *arbitrary vertex* as

$$A_{\mathcal{B}} x_{\mathcal{B}} + A_{\mathcal{N}} x_{\mathcal{N}} = b\,. \tag{60}$$

Precisely at the vertex in question $x_{\mathcal{N}} = 0$, of course. However, we may be interested in the vicinity of that vertex, and thus leave $x_{\mathcal{N}}$ arbitrary at this point.

We may consider (60) as a system of equations in the unknowns $x_{\mathcal{B}}$ with parameters $x_{\mathcal{N}}$, with coefficients $A_{\mathcal{B}}, A_{\mathcal{N}}$ and b drawn from the *equivariant* gaussian ensembles (13), (14), (17) and (18). Thus, the components of $x_{\mathcal{B}}$ (e.g., the $x_{j+n-m}(t)$'s in (59) if we are considering the *optimal* vertex) are random variables. The joint probability density for the m random variables $x_{\mathcal{B}}$ is given by Theorem 4.2 of [27] (applied to the particular gaussian ensembles (13), (14), (17) and (18)) as

$$P(x_{\mathcal{B}}; x_{\mathcal{N}}) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\pi^{\frac{m+1}{2}}} \frac{\lambda}{\left(\lambda^2 + x_{\mathcal{B}}^T x_{\mathcal{B}}\right)^{\frac{m+1}{2}}}\,, \tag{61}$$

where

$$\lambda = \sqrt{1 + x_{\mathcal{N}}^T x_{\mathcal{N}}}\,. \tag{62}$$

(Strictly speaking, we should constrain $x_{\mathcal{B}}$ to lie in the positive orthant, and thus multiply (61) by a factor 2^m to keep it normalized. However, since these details do not affect our discussion below, we avoid introducing them below.)

It follows from (61) that the components of $x_{\mathcal{B}}$ are identically distributed, with probability density of any one of the components $x_{\mathcal{B}j} = \zeta$ given by

$$p(\zeta; x_{\mathcal{N}}) = \frac{1}{\pi}\frac{\lambda}{\lambda^2 + \zeta^2} \tag{63}$$

in accordance with a general theorem due to Girko [28].

The main object of the discussion in this sub-section is to estimate the typical magnitude of the m components of $x_{\mathcal{B}}$. One could argue that typically all m components $|x_{\mathcal{B}j}| < \lambda$, since the Cauchy distribution (63) has width λ. However, from (63) we have that Prob($|\zeta| > \lambda$) = 1/2, namely, $|x_{\mathcal{B}j}| < \lambda$ and $|x_{\mathcal{B}j}| > \lambda$ occur with equal probability. Thus, one has to be more careful, and the answer lies in the probability density function for $R = \sqrt{x_{\mathcal{B}}^T x_{\mathcal{B}}}$.

From (61), we find that the probability density function for $R = \sqrt{x_{\mathcal{B}}^T x_{\mathcal{B}}}$ takes the form

$$\Pi\left(|x_{\mathcal{B}}| = R\right) = \frac{2}{\sqrt{\pi}}\,\frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)}\,\frac{1}{\lambda}\,\frac{\left(\frac{R}{\lambda}\right)^{m-1}}{\left[1 + \left(\frac{R}{\lambda}\right)^2\right]^{\frac{m+1}{2}}}\,. \tag{64}$$

For a *finite* fixed value of m, this expression vanishes as $(R/\lambda)^{m-1}$ for $R << \lambda$, attains its maximum at

$$\left(\frac{R}{\lambda}\right)^2 = \frac{m-1}{2}\,, \tag{65}$$

and then and decays like λ/R^2 for $R >> \lambda$. Thus, like the even Cauchy distribution (63), it does not have a second moment.

In order to make (64) more transparent, we introduce the angle θ defined by

$$\tan\theta\,(R) = \frac{R}{\lambda}\,, \tag{66}$$

where $0 \leq \theta \leq \pi/2$. In terms of θ we have

$$\Pi\left(|x_{\mathcal{B}}| = R\right) = \frac{2}{\sqrt{\pi}}\,\frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)}\,\frac{1}{\lambda}\,\cos^2\theta\,\sin^{m-1}\theta\,. \tag{67}$$

(In order to obtain the probability density for θ we have to multiply the latter expression by a factor $dR/d\theta = \lambda/\cos^2\theta$.)

Let us now concentrate on the asymptotic behavior of (67) (or (64)) in the limit $m \to \infty$. Using Stirling's formula

$$\Gamma(x) \sim \sqrt{\frac{2\pi}{x}}\,x^x\,e^{-x} \tag{68}$$

for the large x asymptotic behavior of the Gamma functions, we obtain for $m \to \infty$

$$\Pi\left(|x_{\mathcal{B}}| = R\right) \sim \sqrt{\frac{2m}{\pi\lambda^2}}\, \cos^2\theta\, \sin^{m-1}\theta\,. \tag{69}$$

Clearly, (69) is exponentially small in m, unless $\sin\theta \simeq 1$, which implies

$$\theta = \pi/2 - \delta \tag{70}$$

with $\delta \sim 1/\sqrt{m}$. Thus, writing

$$\delta = \sqrt{\frac{2u}{m}} \tag{71}$$

(with $u << m$), we obtain, for $m \to \infty$,

$$\Pi\left(|x_{\mathcal{B}}| = R\right) \sim \sqrt{\frac{8}{\pi m\lambda^2}}\, u\, e^{-u}\,. \tag{72}$$

In this regime

$$\frac{R}{\lambda} = \tan\theta \simeq \sqrt{\frac{m}{2u}} >> 1\,. \tag{73}$$

The function on the r.h.s. of (72) has its maximum at $u = 1$, i.e., at $R/\lambda = \sqrt{m/2}$ (in accordance with (65)) and has width of $\mathcal{O}(1)$ around that maximum. However, this is not enough to deduce the typical behavior of R/λ, since as we have already commented following (65), $\Pi\left(|x_{\mathcal{B}}| = R\right)$ has long tails and decays like λ/R^2 past its maximum. Thus, we have to calculate the probability that $R > R_0 = \lambda\tan\theta_0$, given R_0. The calculation is straight forward: using (69) and (66) we obtain

$$\mathrm{Prob}(R > R_0) = \int_{R_0}^{\infty} \Pi\left(R\right)\, dR = \sqrt{\frac{2m}{\pi}} \int_{\theta_0}^{\frac{\pi}{2}} \sin^{m-1}\theta\,. \tag{74}$$

Due to the fact that in the limit $m \to \infty$, $\sin^{m-1}\theta$ may be approximated by a gaussian centered around $\theta = \pi/2$ with variance $1/m$, it is clear that

$$\mathrm{Prob}(R > R_0) = \mathrm{Prob}(\theta > \theta_0) \simeq 1\,,$$

unless $\delta_0 = \pi/2 - \theta_0 \sim \sqrt{2u_0/m}$, with $u_0 << m$. Thus, using (70) and (71) we obtain

$$\mathrm{Prob}(R > R_0) = \sqrt{\frac{2m}{\pi}} \int_0^{\delta_0} \cos^{m-1}\delta \sim \frac{1}{\sqrt{\pi}} \int_0^{u_0} e^{-u}\, \frac{du}{\sqrt{u}} = \mathrm{erf}(\sqrt{u_0})\,. \tag{75}$$

Finally, using the definitions of u_0, θ_0 and R_0, we rewrite (75) as

$$\mathrm{Prob}(R > R_0) = \mathrm{erf}\left[\sqrt{\frac{m}{2}}\, \arctan\left(\frac{\lambda}{R_0}\right)\right]\,. \tag{76}$$

From the asymptotic behavior $\mathrm{erf}(x) \sim 1 - e^{-x^2}/x\sqrt{\pi}$ at large x, we see that $\mathrm{Prob}(R > R_0)$ saturates at 1 exponentially fast as R_0 decreases. Consequently, $1 - \mathrm{Prob}(R > R_0) \sim$

$\mathcal{O}(m^0)$ is not negligible only if R_0/λ is large enough, namely, $\sqrt{\frac{m}{2}}\arctan\left(\frac{\lambda}{R_0}\right) \leq 1$, i.e., $R_0/\lambda \geq \sqrt{m/2}$. If R_0/λ is very large, namely, $R_0/\lambda >> \sqrt{m/2}$, which corresponds to a small argument of the error function in (76), where we clearly have $\text{Prob}(R > R_0) \simeq \sqrt{2m/\pi}(\lambda/R_0) << 1$. From these properties of (76) it thus follows that typically

$$\frac{R}{\lambda} \sim \mathcal{O}(\sqrt{m})\,. \tag{77}$$

Up to this point, we have left the parameters $x_{\mathcal{N}}$ unspecified. At this point we select the prescribed vertex of the polyheder. At the vertex itself, $x_{\mathcal{N}} = 0$. Therefore, from (62), we see that $\lambda = 1$. Thus, according to (77), at the vertex, typically

$$R_{\text{vertex}} \sim \mathcal{O}(\sqrt{m})\,. \tag{78}$$

This result obviously holds for any vertex of the polyheder: any partition (60) of the system of equations $Ax = b$ into basic and non-basic sets leads to the same distribution function (61), and at each vertex we have $x_{\mathcal{N}} = 0$.

Thus, clearly, this means that the whole polyheder is typically bounded inside an n-dimensional sphere of radius $R \sim \mathcal{O}(\sqrt{m})$ centered at the origin.

Thus, from (78) and from the rotational symmetry of (61), we conclude that any component of $x_{\mathcal{B}}$ at the optimal vertex, or at any other vertex (with its appropriate basic set $\mathcal{B}$), is typically of $\mathcal{O}(R_{\text{vertex}}/\sqrt{m}) = \mathcal{O}(1)$ (and of course, positive). Points on the polyheder other than the vertices are weighted linear combinations of the vertices with positive weights which are smaller than unity, and as such also have their individual components typically of $\mathcal{O}(1)$.

6.2 Non-asymptotic complexity measures from β_i

Applying the results of the previous subsection to the optimal vertex, we expect the components of $x_{\mathcal{B}}(t)$ (i.e., the $x_{j+n-m}(t)$'s in (59)) to be typically of the same order of magnitude as their asymptotic values $\lim_{t\to\infty} x_{\mathcal{B}}(t)$ at the optimal vertex, and as a result, we expect the barrier $\beta_i(t)$ to be of the same order of magnitude as its asymptotic value $\lim_{t\to\infty}\beta_i(t)$.

Note that, for this reason, in order to determine how the $x_i(t)$ in (8) tend to zero, to leading order, we can safely replace all the $x_{j+n-m}(t)$ by their asymptotic values in $x^*_{\mathcal{B}}$. Thus, in the following we approximate the barrier (59) by its asymptotic value

$$\beta_i = -\lim_{t\to\infty}\sum_{j=1}^{m}\alpha_{ji}\log x_{j+n-m}(t) = -\sum_{j=1}^{m}\alpha_{ji}\log x^*_{j+n-m}\,, \tag{79}$$

where we have also ignored the contribution of the initial condition.

We now consider the convergence time of the solution $x(t)$ of (5) to the optimal vertex. In order for $x(t)$ to be close to the maximum vertex we must have $x_i(t) < \epsilon$ for $i = 1,\ldots,n-m$ for some small positive ϵ. The time parameter t must then satisfy:

$$\exp(-\Delta_i t + \beta_i) < \epsilon\ , \quad \text{for } i = 1,\ldots,n-m. \tag{80}$$

Solving for t, we find an estimate for the time required to flow to the vicinity of the optimal vertex as

$$t > \frac{\beta_i}{\Delta_i} + \frac{|\log\epsilon|}{\Delta_i}\ , \text{for all } i = 1,\ldots,m. \tag{81}$$

We define

$$T = \max_i \left(\frac{\beta_i}{\Delta_i} + \frac{|\log \epsilon|}{\Delta_i} \right) \ , \tag{82}$$

which we consider as the computation time. We denote

$$\beta_{\max} = \max_i \beta_i \ . \tag{83}$$

In the limit of asymptotically small ϵ, the first term in (82) is irrelevant, and the distribution of computation times is determined by the distribution of the Δ_i's stated by Theorem (4.1).

If the asymptotic precision is not required, the first term in (82) may be dominant. To bound this term in the expression for the computation time we can use the quotient $\beta_{\max}/\Delta_{\min}$, where Δ_{min} is defined in (11).

In the probabilistic ensemble used in this work β_{max} and β_{max}/Δ_{min} are random variables, as is Δ_{min}. Unfortunately, we could not find the probability distributions of β_{max} and β_{max}/Δ_{min} analytically as we did for Δ_{min}. In the following section, a conjecture concerning these distributions, based on numerical evidence, will be formulated.

7 Scaling functions

In Section 4 it was shown that in the limit of large (n, m) the probability $\mathcal{P}(\Delta_{min} > \Delta | \Delta_{min} > 0)$ is given by (52). Consequently, $\mathcal{P}(\Delta_{min} < \Delta | \Delta_{min} > 0) \equiv \mathcal{F}^{(n,m)}(\Delta)$ is of the scaling form

$$\mathcal{F}^{(n,m)}(\Delta) = 1 - e^{x_\Delta^2} \operatorname{erfc}(x_\Delta) \ \equiv \mathcal{F}(x_\Delta). \tag{84}$$

Such a scaling form is very useful and informative, as we will demonstrate in what follows. The scaling function $\mathcal{F}$ contains *all* asymptotic information on Δ. In particular, one can extract the problem size dependence of $f_{\min}^{(m)}(0|\Delta_{min} > 0)$ which is required for obtaining a high probability bound using Lemma 5.1. (This has already been shown in Corollary (5.1).) We use the scaling form, equation (84), leading to,

$$f_{min}^{(m)}(0|\Delta_{min} > 0) = \frac{d\mathcal{F}^{(n,m)}(\Delta)}{d\Delta}|_{\Delta=0} = \eta_\Delta(n,m) \frac{\mathcal{F}(x_\Delta)}{dx_\Delta}|_{x_\Delta=0}. \tag{85}$$

This is just $f_{min1}(0)/\mathcal{Q}(0)$. With the help of lemma 5.1, leading to (58) and our finding that $\eta(n, m) \sim \sqrt{m}$, we conclude that with high probability

$$\frac{1}{\Delta_{\min}} = \mathcal{O}(\sqrt{m}). \tag{86}$$

The next observation is that the distribution $\mathcal{F}(x_\Delta)$ is very wide. For large x_Δ it behaves as $1 - \frac{1}{\sqrt{\pi} x_\Delta}$, as is clear from the asymptotic behavior of the erfc function. Therefore it does not have a mean. Since at $x_\Delta = 0$ the slope $d\mathcal{F}/dx_\Delta|_{x_\Delta=0}$ does not vanish, also $1/x_\Delta$ does not have a mean (see Remark (5.2)).

We would like to derive scaling functions like (84) also for the barrier β_{max}, that is the maximum of the β_i defined by (79) and for the computation time T defined by (82). The analytic derivation of such scaling functions is difficult and therefore left for further studies.

Their existence is verified numerically in the next section. In particular for fixed $r = m/n$, we found that

$$\mathcal{P}\left(\frac{1}{\beta_{max}} < \frac{1}{\beta}\right) \equiv \mathcal{F}^{(n,m)}_{1/\beta_{max}}\left(\frac{1}{\beta}\right) = \mathcal{F}_{1/\beta}(x_\beta) \tag{87}$$

and

$$\mathcal{P}\left(\frac{1}{T} < \frac{1}{t}\right) \equiv \mathcal{F}^{(n,m)}_{1/T}\left(\frac{1}{t}\right) = \mathcal{F}_{1/T}(x_T), \tag{88}$$

where β_{max} and T are the maximal barrier and computation time. The scaling variables are

$$x_\beta = \eta_\beta(n,m)\frac{1}{\beta} \tag{89}$$

and

$$x_T = \eta_T(n,m)\frac{1}{t}. \tag{90}$$

The asymptotic behavior of the scaling variables was determined numerically to be

$$\eta_\beta(n,m) \sim m \tag{91}$$

and

$$\eta_T(n,m) \sim m\log m. \tag{92}$$

This was found for constant r. The precise r dependence could not be determined numerically. The resulting high probability behavior for the barrier and computation time is therefore:

$$\beta_{\max} = \mathcal{O}(m), \quad T = \mathcal{O}(m\log m) \tag{93}$$

Note that scaling functions, such as these, immediately provide the average behavior as well (if it exists)

Here, in the calculation of the distribution of computation times it was assumed that these are dominated by the barriers rather then by $|\log\epsilon|$ in (82). The results (87), (88) and (93) are conjectures supported by the numerical calculations of the next section.

8 Numerical simulations

In this section the results of numerical simulations for the distributions of LP problems are presented. For this purpose we generated full LP instances (A, b, c) with the distribution (12). For each instance the LP problem was solved using the linear programming solver of the IMSL C library. Only instances with a bounded optimal solution were kept, and $\Delta_{\min}$ was computed relative to the optimal partition and optimality was verified by checking that $\Delta_{\min} > 0$. Using the sampled instances we obtain an estimate of $\mathcal{F}^{(n,m)}(\Delta) = \mathcal{P}(\Delta_{min} < \Delta|\Delta_{min} > 0)$, and of the corresponding cumulative distribution functions of the barrier $\beta_{\max}$ and the computation time.

As a consistency verification of the calculations we compared $\mathcal{P}(\Delta_{min} < \Delta|\Delta_{min} > 0)$, to $\mathcal{P}(\Delta_{min1} < \Delta|\Delta_{min1} > 0)$ that was directly estimated from the distribution of matrices. For this purpose we generated a sample of A and c according to the probability distributions (13,15) with $\sigma = 1$ and computed for each instance the value of Δ_{min1} (the minimum over

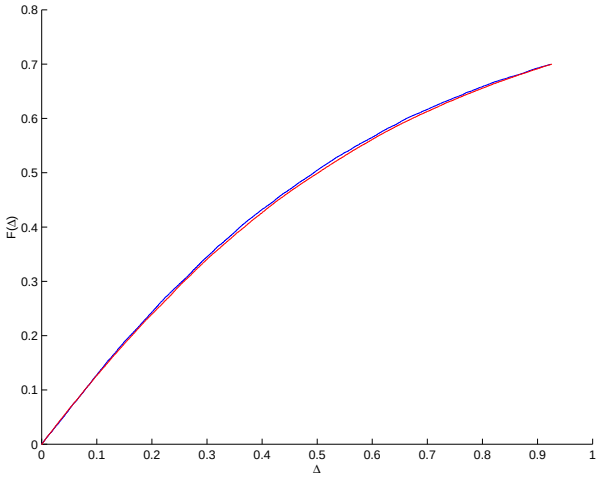

Figure 1: Comparison of $\mathcal{P}(\Delta_{min1} < \Delta | \Delta_{min1} > 0)$ and $\mathcal{P}(\Delta_{\min} < \Delta | \Delta_{\min} > 0)$ for $m = 2, n = 4$.

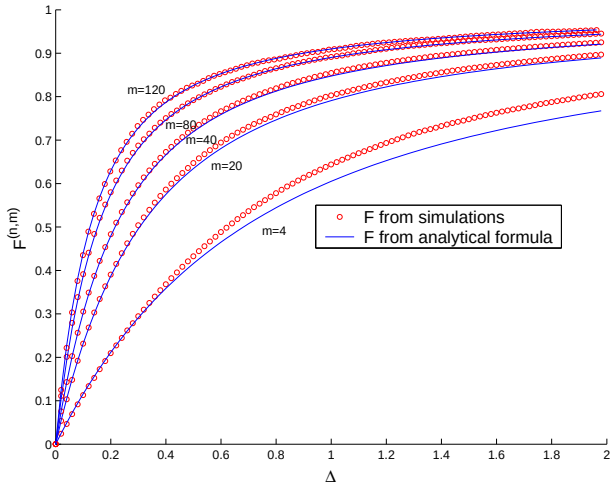

Figure 2: $\mathcal{F}^{(n,m)}(\Delta)$ for $m = 4, 20, 40, 80, 120, n = 2m$. The number of instances was $10^5, 10^5, 40000, 15000, 5800$ respectively. There is very good agreement with the analytical results, improving as m increases.

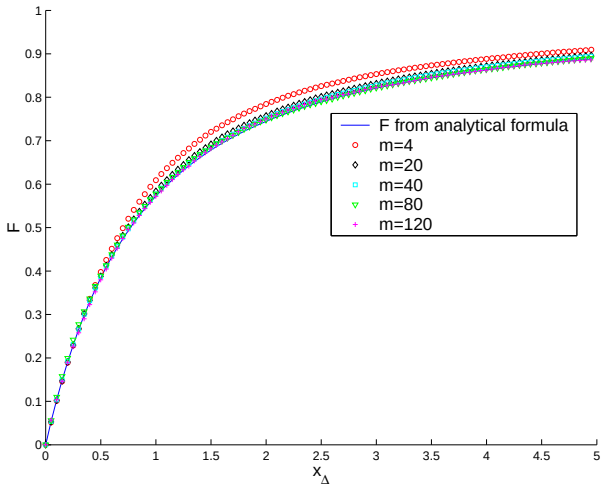

Figure 3: $\mathcal{F}(x_\Delta)$ is plotted against the variable x_Δ, for the same data as Figure 2.

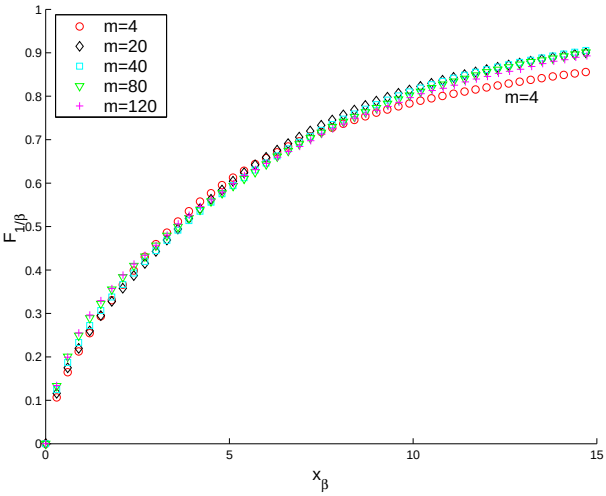

Figure 4: $\mathcal{F}_{1/\beta}(x_\beta)$ as a function of the variable $x_\beta = m/\beta_{\max}$ for the same instances as Figure 2.

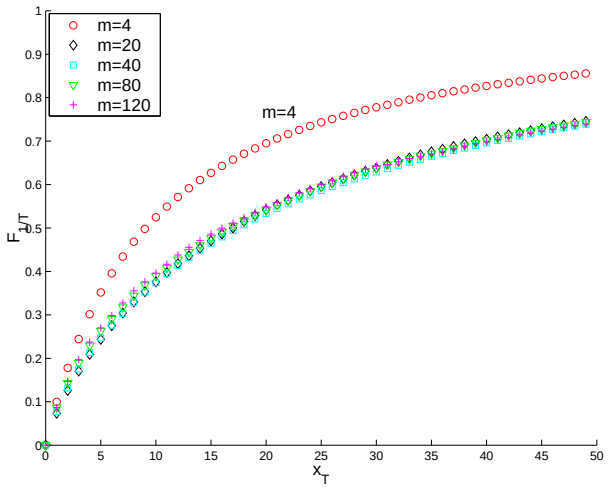

Figure 5: $\mathcal{F}_{1/T}(x_T)$ as a function of the variable $x_T = m \log m/T$ for the same instances as Figure 2.

Δ_i) for a fixed partition of A into (N, B). We kept only the positive values (note that the definition of Δ_{min1} does not require b). The two distributions are compared in Figure 1, with excellent agreement.

Note that estimation of $\mathcal{P}(\Delta_{min1} < \Delta | \Delta_{min1} > 0)$ by sampling from a fixed partition is infeasible for large m and n, since for any partition of A the probability that Δ_{min1} is positive is $2^{-(n-m)}$ (equation (24)). Therefore the equivalence between the probability distributions of $\Delta_{\min}$ and Δ_{min1} cannot be exploited for producing numerical estimates of the probability distribution of $\Delta_{\min}$. Thus we proceed by generating full LP instances, and solving the LP problem as described above.

The problem size dependence was explored while keeping the ratio n/m fixed or while keeping m fixed and varying n. In Figure 2 we plot the numerical estimates of $\mathcal{F}^{(n,m)}(\Delta)$ for varying problem sizes with $n/m = 2$ and compare it with the analytical result, Equation (84). The agreement with the analytical result improves as m is increased, since it is an asymptotic result. The simulations show that the asymptotic result holds well even for $m = 20$. As in the analytical result, in the large m limit we observe that $\mathcal{F}^{(n,m)}(\Delta)$ is not a general function of n, m and Δ, but a scaling function of the form $\mathcal{F}^{(n,m)}(\Delta) = \mathcal{F}(x_\Delta)$ as predicted theoretically in Section 7 (see (84) there). The scaling variable $x_\Delta(m)$ is given by (50). Indeed, Figure 3 demonstrates that $\mathcal{F}^{(n,m)}$ has this form as predicted by Equation (84) with the scaling variable x_Δ.

For the cumulative distribution functions of the barrier $\beta_{\max}$ and of the computation time T we do not have analytical formulas. These distribution functions are denoted by $\mathcal{F}^{(n,m)}_{1/\beta_{\max}}$ and $\mathcal{F}^{(n,m)}_{1/T}$ respectively. Their behavior near zero enables to obtain high probability bounds on $\beta_{\max}$ and T, since for this purpose we need to bound the tails of their distributions, or alternatively, estimate the density of $1/\beta\max$ and $1/T$ at 0. In the numerical estimate of the barrier we collected only positive values, since only these contribute to prolonging the computation time. From Figure 4 we find that $\mathcal{F}^{(n,m)}_{1/\beta_{\max}}$ is indeed a scaling function of the form (87) with the scaling variable x_β of (89). The behavior of the compu-

tation time is extracted from Figure 5. The cumulative function $\mathcal{F}_{1/T}^{(n,m)}$ is found to be a scaling function of the form (88) with the scaling variable x_T of (90). The scaling variables x_β and x_T were found numerically by the requirement that in the asymptotic limit the cumulative distribution $\mathcal{F}$ approaches a scaling form. Such a fitting is possible only if a scaling form exists. We were unable to determine the dependence of the scaling variables x_β and x_T on n/m.

9 Summary and discussion

In this paper we computed the problem size dependence of the distributions of parameters that govern the convergence of a differential equation (Eq.(5)) that solves the linear programming problem [6]. To the best of our knowledge, this is the first time such distributions are computed. In particular, knowledge of the distribution functions enables to obtain the high probability behavior (for example (86)) and (93)), and the moments (if these exist).

The main result of the present work is that the distribution functions of the convergence rate, Δ_{min}, the barrier β_{max} and the computation time T are scaling functions; i.e., in the asymptotic limit of large (n, m), each depends on the problem size only through a scaling variable. These functions are presented in section 7.

The scaling functions obtained here provide *all* the relevant information about the distribution in the large (n, m) limit. Such functions, even if known only numerically, can be useful for the understanding of the behavior for large values of (n, m) that are beyond the limits of numerical simulations. In particular, the distribution function of Δ_{min} was calculated analytically and stated as Theorem (4.1). The relevance of the asymptotic theorem for finite and relatively small problem sizes (n, m) was demonstrated numerically. It turns out to be a very simple function (see (84)). The scaling form of the distributions of β_{max} and of T was conjectured on the basis of numerical simulations.

The Faybusovich flow [6] that is studied in the present work, is defined by a system of differential equations, and it can be considered as an example of the analysis of convergence to fixed points for differential equations. One should note, however, that the present system has a formal solution (8), and therefore it is not typical.

If we require knowledge of the attractive fixed points with arbitrarily high precision (i.e., ϵ of (80) and (82) can be made arbitrarily small), the convergence time to an ϵ-vicinity of the fixed point is dominated by the convergence rate $\Delta_{\min}$. The barrier, that describes the state space "landscape" on the way to fixed points, is irrelevant in this case. Thus, in this limit, the complexity is determined by (86). This point of view is taken in [12].

However, for the solution of some problems (like the one studied in the present work), such high precision is usually not required, and also the non-asymptotic behavior (in ϵ) of the vector field, as represented by the barrier, has an important contribution to the complexity of computing the fixed point.

For computational models defined on the real numbers, worst case behavior can be ill defined and lead to infinite computation times, in particular for interior point methods for linear programming [13]. Therefore, we compute the distribution of computation times for a probabilistic model of linear programming instances rather then an upper bound. Such probabilistic models can be useful in giving a general picture also for traditional discrete problem solving, where the continuum theory can be viewed as an approximation.

A question of fundamental importance is how general is the existence of scaling distributions. Their existence would be analogous to the central limit theorem [29] and to scaling in critical phenomena [30] and in Anderson localization [31, 32]. Typically such functions are universal. In the case of the central limit theorem, for example, under some very general conditions one obtains a Gaussian distribution, irrespectively of the original probability distributions. Moreover it depends on the random variable and the number of the original variables via a specific combination. The Gaussian distribution is a well known example of the so-called stable probability distributions. In the physical problems mentioned above scaling and universality reflect the fact that the systems becomes scale invariant.

A specific challenging problem still left unsolved in the present work is the rigorous calculation of the distributions of $1/\beta_{max}$ and of $1/T$, that is proving the conjectures concerning these distributions. This will be attempted in the near future.

A The Faybusovich vector field

In the following we consider the inner product $< \xi, \eta >_{X^{-1}} = \xi^T X^{-1} \eta$, This inner product is defined on the positive orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i > 0,\ i = 1, \ldots, n\}$, where it defines a Riemannian metric. In the following we denote by $a^i,\ i = 1, \ldots, m$ the rows of A. The Faybusovich vector field is the gradient of h relative to this metric projected to the constraint set [6]. It can be expressed as:

$$\operatorname{grad} h = Xc - \sum_{i=1}^{m} \zeta_i(x) X a^i, \tag{94}$$

where $\zeta_1(x), \ldots, \zeta_m(x)$ make the gradient perpendicular to the constraint vectors, i.e. $A \operatorname{grad} h = 0$, so that $Ax = b$ is maintained by the dynamics. The resulting flow is

$$\frac{dx}{dt} = F(x) = \operatorname{grad} h \tag{95}$$

Consider the functions

$$\Psi_i(x) = \log(x_i) + \sum_{j=1}^{m} \alpha_{ji} \log(x_{j+n-m}) \quad i = 1 \ldots n - m. \tag{96}$$

The Ψ_i are defined such that their equations of motion are easily integrated. This gives $n - m$ equations which correspond to the $n - m$ independent variables of the LP problem. To compute the time derivative of Ψ_i we first find:

$$\nabla \Psi_i = \frac{1}{x_i} e^i + \sum_{j=1}^{m} \frac{\alpha_{ij}}{x_{j+n-m}} e^{j+n-m}\ , \tag{97}$$

and note that the vectors μ^i defined in equation (6) have the following property:

$$< \mu^i, a^j >= 0, \quad i = 1, \ldots, n - m,\ j = 1, \ldots, m$$

Therefore:

$$\begin{aligned}\dot{\Psi}_i(x) &= <\nabla\Psi_i(x), \dot{x}> = <\nabla\Psi_i(x), \mathrm{grad}\, h> \\ &= <\mu^i, c - \sum_{j=1}^{m} \zeta_j(x) a^j> \\ &= <\mu^i, c> \equiv -\Delta_i\end{aligned} \tag{98}$$

This equation is integrated to yield:

$$x_i(t) = x_i(0) \exp\left(-\Delta_i t - \sum_{j=1}^{m} \alpha_{ij} \log \frac{x_{j+n-m}(t)}{x_{j+n-m}(0)} \right) . \tag{99}$$

B The probability distribution $P(u)$

In this Appendix we study the probability distribution function

$$P(u) = \left(\frac{\lambda}{\pi}\right)^{\frac{m^2+m}{2}} \int d^{m^2}B\, d^m z\, e^{-\lambda\left(\mathrm{tr}\, B^T B + z^T z\right)} \cdot \delta\left(u - \frac{1}{z^T (B^T B)^{-1} z + 1} \right) ,$$

defined in (30) and (31) and calculate it in detail explicitly, in the large n, m limit.

We will reconstruct $P(u)$ from its moments. The N-th moment

$$\begin{aligned} k_N &= \int_0^\infty du P(u)\, u^N \\ &= \left(\frac{\lambda}{\pi}\right)^{\frac{m^2+m}{2}} \int d^{m^2}B\, d^m z\, e^{-\lambda\left(\mathrm{tr}\, B^T B + z^T z\right)} \left(\frac{1}{z^T (B^T B)^{-1} z + 1} \right)^N \end{aligned} \tag{100}$$

of $P(u)$ may be conveniently represented as

$$\begin{aligned} k_N &= \left(\frac{\lambda}{\pi}\right)^{\frac{m^2+m}{2}} \frac{1}{\Gamma(N)} \int_0^\infty t^{N-1} e^{-t}\, dt \int d^{m^2}B\, e^{-\lambda \mathrm{tr}\, B^T B} \int d^m z\, e^{-z^T\left(\lambda + \frac{t}{B^T B}\right) z} \\ &= \left(\frac{\lambda}{\pi}\right)^{\frac{m^2}{2}} \frac{1}{\Gamma(N)} \int_0^\infty t^{N-1} e^{-t}\, dt \int d^{m^2}B\, \frac{e^{-\lambda \mathrm{tr}\, B^T B}}{\sqrt{\det\left(\mathbf{1} + \frac{t/\lambda}{B^T B}\right)}} , \end{aligned} \tag{101}$$

where in the last step we have performed Gaussian integration over z.

Recall that $P(u)$ is independent of the arbitrary parameter λ (see the remark preceding (31)). Thus, its N-th moment k_N must also be independent of λ, which is manifest in (101). Therefore, with no loss of generality, and for later convenience, we will henceforth set $\lambda = m$ (since we have in mind taking the large m limit). Thus,

$$\begin{aligned} k_N &= \left(\frac{m}{\pi}\right)^{\frac{m^2}{2}} \frac{1}{\Gamma(N)} \int_0^\infty t^{N-1} e^{-t}\, dt \int d^{m^2}B\, \frac{e^{-m \mathrm{tr}\, B^T B}}{\sqrt{\det\left(\mathbf{1} + \frac{t/m}{B^T B}\right)}} \\ &= \frac{1}{\Gamma(N)} \int_0^\infty t^{N-1} e^{-t}\, \psi\left(\frac{t}{m}\right) dt , \end{aligned} \tag{102}$$

where we have introduced the function

$$\psi(y) = \left(\frac{m}{\pi}\right)^{\frac{m^2}{2}} \int d^{m^2}B \, \frac{e^{-m\mathrm{tr}\, B^T B}}{\sqrt{\det\left(\mathbf{1} + \frac{y}{B^T B}\right)}} \,. \tag{103}$$

Note that

$$\psi(0) = 1 \,. \tag{104}$$

The function $\psi(y)$ is well-defined for $y \geq 0$, where it clearly decreases monotonically

$$\psi'(y) < 0 \,. \tag{105}$$

We would like now to integrate over the rotational degrees of freedom in dB. Any real $m \times m$ matrix B may be decomposed as [24, 25]

$$B = \mathcal{O}_1^T \Omega \mathcal{O}_2 \tag{106}$$

where $\mathcal{O}_{1,2} \in \mathcal{O}(m)$, the group of $m \times m$ orthogonal matrices, and $\Omega = \mathrm{Diag}(\omega_1, \ldots, \omega_m)$, where $\omega_1, \ldots, \omega_m$ are the singular values of B. Under this decomposition we may write the measure dB as [24, 25]

$$dB = d\mu(\mathcal{O}_1) d\mu(\mathcal{O}_2) \prod_{i<j} |\omega_i^2 - \omega_j^2| d^m\omega \,, \tag{107}$$

where $d\mu(\mathcal{O}_{1,2})$ are Haar measures over the appropriate group manifolds. The measure dB is manifestly invariant under actions of the orthogonal group $\mathcal{O}(m)$

$$dB = d(B\mathcal{O}) = d(\mathcal{O}'B) \,, \qquad \mathcal{O}, \mathcal{O}' \in \mathcal{O}(m) \,, \tag{108}$$

as should have been expected to begin with.

Remark B.1 Note that the decomposition (106) is not unique, since $\mathcal{O}_1\mathcal{D}$ and $\mathcal{D}\mathcal{O}_2$, with $\mathcal{D}$ being any of the 2^m diagonal matrices $\mathrm{Diag}\,(\pm 1, \cdots, \pm 1)$, is an equally good pair of orthogonal matrices to be used in (106). Thus, as $\mathcal{O}_1$ and $\mathcal{O}_2$ sweep independently over the group $\mathcal{O}(m)$, the measure (107) over counts B matrices. This problem can be easily rectified by appropriately normalizing the volume $\mathcal{V}_m = \int d\mu(\mathcal{O}_1) d\mu(\mathcal{O}_2)$. One can show that the correct normalization of the volume is

$$\mathcal{V}_m = \frac{\pi^{\frac{m(m+1)}{2}}}{2^m \prod_{j=1}^m \Gamma\left(1 + \frac{j}{2}\right) \Gamma\left(\frac{j}{2}\right)} \,. \tag{109}$$

One simple way to establish (109), is to calculate

$$\int dB \, \exp -\frac{1}{2}\mathrm{tr} B^T B = (2\pi)^{\frac{m^2}{2}} = \mathcal{V}_m \int_{-\infty}^{\infty} d^m\omega \prod_{i<j} |\omega_i^2 - \omega_j^2| \, \exp -\frac{1}{2}\sum_i \omega_i^2 \,.$$

The last integral is a known Selberg type integral [25].

The integrand in (103) depends on B only through the combination $B^TB = \mathcal{O}_2^T\Omega^2\mathcal{O}_2$. Thus, the integrations over $\mathcal{O}_1$ and $\mathcal{O}_2$ in (103) factor out trivially. Thus, we end up with

$$\psi(y) = \mathcal{V}_m \left(\frac{m}{\pi}\right)^{\frac{m^2}{2}} \int_{-\infty}^{\infty} \frac{\prod_{i<j} |\omega_i^2 - \omega_j^2| d^m\omega}{\sqrt{\det\left(\mathbf{1} + \frac{y}{\Omega^2}\right)}} \, e^{-m\mathrm{tr}\,\Omega^2} \,. \tag{110}$$

It is a straight forward exercise to check that (109) is consistent with $\psi(0) = 1$.

Note that in deriving (110) we have made no approximations. Up to this point, all our considerations in this appendix were exact. We are interested in the large n, m asymptotic behavior[1] of $P(u)$ and of its moments. Thus, we will now evaluate the large m behavior of $\psi(y)$ (which is why we have chosen $\lambda = m$ in (102)). This asymptotic behavior is determined by the saddle point dominating the integral over the m singular values ω_i in (110) as $m \to \infty$.

To obtain this asymptotic behavior we rewrite the integrand in (110) as

$$\frac{e^{-S}}{\sqrt{\det\left(\mathbf{1} + \frac{y}{\Omega^2}\right)}} \,,$$

where

$$S = m \sum_{i=1}^{m} \omega_i^2 - \frac{1}{2} \sum_{i<j} \log\left(\omega_i^2 - \omega_j^2\right)^2 . \tag{111}$$

In physical terms, S is the energy (or the action) of the so-called "Dyson gas" of eigenvalues, familiar from the theory of random matrices.

We look for a saddle point of the integral in (110) in which all the ω_i are of $\mathcal{O}(1)$. In such a case, S in (111) is of $\mathcal{O}(m^2)$, and thus e^{-S} overwhelms the factor

$$\frac{1}{\sqrt{\det\left(\mathbf{1} + \frac{y}{\Omega^2}\right)}} = e^{-\frac{m}{2} I(y)} \,,$$

where

$$I(y) = \frac{1}{m} \sum_{i=1}^{m} \log\left(1 + \frac{y}{\omega_i^2}\right) \tag{112}$$

is a quantity of $\mathcal{O}(m^0)$. For later use, note that

$$I(0) = 0 \,. \tag{113}$$

Thus, to leading order in $1/m$, $\psi(y)$ is dominated by the well defined and stable saddle point of S, which is indeed the case.

Simple arguments pertaining to the physics of the Dyson gas make it clear that the saddle point is stable: The "confining potential" term $\sum_i \omega_i^2$ in (111) tends to condense all the ω_i at zero, while the "Coulomb repulsion" term $-\sum_{i<j} \log(\omega_i^2 - \omega_j^2)^2$ acts to keep the $|\omega_i|$ apart. Equilibrium must be reached as a compromise, and it must be stable,

[1] Recall that m and n tend to infinity with the ratio (41), $r = m/n$, kept finite.

since the quadratic confining potential would eventually dominate the logarithmic repulsive interaction for ω_i large enough. The saddle point equations

$$\frac{\partial S}{\partial \omega_i} = 2\omega_i \left[m - \sum_{j \neq i} \frac{1}{\omega_i^2 - \omega_j^2} \right] = 0 \,, \tag{114}$$

are simply the equilibrium conditions between repulsive and attractive interactions, and thus determine the distribution of the $|\omega_i|$.

We will solve (114) (using standard techniques of random matrix theory), and thus will determine the equilibrium configuration of the molecules of the Dyson gas in the next appendix, where we show that the m singular values ω_i condense (non uniformly) into the finite segment (see Eq. (141))

$$0 \leq \omega_i^2 \leq 2$$

(and thus with mean spacing of the order of $1/m$).

To summarize, in the large m limit, $\psi(y)$ is determined by the saddle point of the energy S (111) of the Dyson gas. Thus for large m, according to (110), (111) and (112),

$$\psi(y) \simeq \frac{\mathcal{V}}{2^m} \left(\frac{m}{\pi} \right)^{\frac{m^2}{2}} \exp - \left(S_* + \frac{m}{2} I_*(y) \right) \,,$$

where S_* is the extremal value of (111), and $I_*(y)$ is (112) evaluated at that equilibrium configuration of the Dyson gas, namely,

$$I_*(y) = \frac{1}{m} \sum_{i=1}^{m} \log \left(1 + \frac{y}{\omega_{i*}^2} \right) \,. \tag{115}$$

The actual value of S_* (a number of $\mathcal{O}(m^2)$) is of no special interest to us here, since from (104) and (113) we immediately deduce that in the large m limit

$$\psi(y) \simeq e^{-\frac{m}{2} I_*(y)} \,. \tag{116}$$

Substituting (116) back into (102) we thus obtain the large (n, m) behavior of k_N as

$$k_N \simeq \frac{1}{\Gamma(N)} \int_0^\infty t^{N-1} \, e^{-t - \frac{m}{2} I_* \left(\frac{t}{m} \right)} \, dt \,. \tag{117}$$

The function $I_*(y)$ is evaluated in the next Appendix, and is given in Eq. (145),

$$I_*(y) = -y + \sqrt{y^2 + 2y} + \log \left(y + 1 + \sqrt{y^2 + 2y} \right) \,,$$

which we repeated here for convenience.

The dominant contribution to the integral in (117) comes from values of $t << m$, since the function

$$\phi(t) = t + \frac{m}{2} I_* \left(\frac{t}{m} \right) \,, \tag{118}$$

which appears in the exponent in (117) is monotonously increasing, as can be seen from (144). Thus, in this range of the variable t, using (146), we have

$$\phi(t) = t + \frac{m}{2} I_* \left(\frac{t}{m}\right) = 2\sqrt{2mt} + \frac{t}{2} + \mathcal{O}\left(\frac{1}{\sqrt{m}}\right) . \tag{119}$$

Note that the term $t/2$ in (119) is beyond the accuracy of our approximation for I_*. The reason is that in (142) we used the continuum approximation to the density of singular values, which introduced errors of the orders of $1/m$. Fortunately, this term is not required. The leading order term in the exponential (119) of (117) is just $\sqrt{2mt}$. Consequently, in the leading order (117) reduces to

$$\begin{aligned} k_N &\equiv \int_0^\infty du P(u)\, u^N \simeq \frac{1}{\Gamma(N)} \int_0^\infty t^{N-1} e^{-\sqrt{2mt}}\, dt \\ &= \frac{2\Gamma(2N)}{(2m)^N \Gamma(N)} = \frac{(2N-1)!!}{m^N} . \end{aligned} \tag{120}$$

The moments (120) satisfy Carleman's criterion [33, 34]

$$\sum_{N=1}^{\infty} k_N^{-1/2N} = \infty , \tag{121}$$

which is sufficient to guarantee that these moments define a unique distribution $P(u)$.

Had we kept in (120) the $\mathcal{O}(m^0)$ piece of (119), i.e., the term $t/2$, it would have produced a correction factor to (120) of the form $1 + \mathcal{O}(N^2/m)$. To see this, consider the integral

$$\begin{aligned} \frac{1}{\Gamma(N)} \int_0^\infty t^{N-1} e^{-\sqrt{2mt} - t/2}\, dt &= \frac{2}{(2m)^N \Gamma(N)} \int_0^\infty y^{2N-1} e^{-y - y^2/4m}\, dy \\ &\simeq \frac{2}{(2m)^N \Gamma(N)} \int_0^\infty y^{2N-1} e^{-y} \left(1 - y^2/4m + \cdots\right) dy . \end{aligned}$$

Thus, we can safely trust (120) for moments of order $N << \sqrt{m}$.

The expression in (120) is readily recognized as the $2N$-th moment of a Gaussian distribution defined on the positive half-line. Indeed, the moments of the Gaussian distribution

$$g(x;\mu) = \frac{2\mu}{\sqrt{\pi}}\, e^{-\mu^2 x^2} , \quad x \geq 0 \tag{122}$$

are

$$\langle x^k \rangle = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi}\, \mu^k} . \tag{123}$$

In particular, the even moments of (122) are

$$\langle x^{2N} \rangle = \frac{\Gamma\left(N + \frac{1}{2}\right)}{\sqrt{\pi}\, \mu^{2N}} = \frac{(2N-1)!!}{(2\mu^2)^N} , \tag{124}$$

which coincide with (120) for $2\mu^2 = m$. These are the moments of $u = x^2$ for the distribution $P(u)$ satisfying $P(u)\,du = g(x;\sqrt{m/2})dx$, as can be seen comparing (120) and (124).

Thus, we conclude that the leading asymptotic behavior of $P(u)$ as m tends to infinity is

$$P(u) = \sqrt{\frac{m}{2\pi u}}\, e^{-\frac{mu}{2}}\,, \tag{125}$$

the result quoted in (42).

As an additional check of this simple determination of $P(u)$ from (120), we now sketch how to derive it more formally from the function

$$G(z) = \int_0^\infty \frac{P(u)du}{z-u}\,, \tag{126}$$

known sometimes as the Stieltjes transform of $P(u)$ [33]. $G(z)$ is analytic in the complex z-plane, cut along the support of $P(u)$ on the real axis. We can then determine $P(u)$ from (126), once we have an explicit expression for $G(z)$, using the identity

$$P(u) = \frac{1}{\pi}\,\mathrm{Im}\; G(u - i\epsilon)\,. \tag{127}$$

For z large and off the real axis, and if all the moments of $P(u)$ exist, we can formally expand $G(z)$ in inverse powers of z. Thus,

$$G(z) = \sum_{N=0}^\infty \int_0^\infty \frac{P(u)u^N du}{z^{N+1}} = \sum_{N=0}^\infty \frac{k_N}{z^{N+1}}\,. \tag{128}$$

For the k_N's given by (120), the series (128) diverges. However, it is Borel summable [33]. Borel resummation of (120), making use of

$$\frac{1}{\sqrt{1-x}} = 1 + \sum_{N=1}^\infty \frac{(2N-1)!!}{N!}\left(\frac{x}{2}\right)^N\,,$$

yields

$$G(z) = \frac{1}{z}\int_0^\infty \frac{e^{-t}dt}{\sqrt{1-\frac{2t}{mz}}}\,. \tag{129}$$

Thus,

$$\frac{1}{\pi}\,\mathrm{Im}\; G(u-i\epsilon) = \frac{1}{\pi u}\int_{\frac{mu}{2}}^\infty \frac{e^{-t}dt}{\sqrt{\frac{2t}{mu}-1}} = \sqrt{\frac{m}{2\pi u}}\, e^{-\frac{mu}{2}}\,, \tag{130}$$

which coincides with (125).

C The saddle point distribution of the ω_i

We present in this Appendix the solution of the equilibrium condition (114) of the Dyson gas of singular values

$$\frac{\partial S}{\partial \omega_i^2} = m - \sum_{j\neq i} \frac{1}{\omega_i^2 - \omega_j^2} = 0 \tag{131}$$

(which we repeated here for convenience), and then use it to calculate $I_*(y)$, defined in (115). We follow standard methods [23, 24] of random matrix theory [25]. Let

$$s_i = \omega_i^2 \,, \tag{132}$$

and also define

$$F(w) = \frac{1}{m}\sum_{i=1}^{m} \langle \frac{1}{w - s_i} \rangle = \frac{1}{m} \langle \mathrm{tr} \frac{1}{w - B^T B} \rangle \,, \tag{133}$$

where w is a complex variable. Here the angular brackets denote averaging with respect to the B sector of (13). By definition, $F(w)$ behaves asymptotically as

$$F(w) \underset{w\to\infty}{\longrightarrow} \frac{1}{w} \,. \tag{134}$$

It is clear from (133) that for $s > 0,\ \epsilon \to 0+$ we have

$$F(s - i\epsilon) = \frac{1}{m} \mathrm{P.P.} \sum_{i=1}^{m} \langle \frac{1}{s - s_i} \rangle + \frac{i\pi}{m} \sum_{i=1}^{m} \langle \delta(s - s_i) \rangle \tag{135}$$

where P.P. stands for the principal part. Therefore (from (133)), the average eigenvalue density of $B^T B$ is given by

$$\rho(s) \equiv \frac{1}{m} \sum_{i=1}^{m} \langle \delta(s - s_i) \rangle = \frac{1}{\pi} \,\mathrm{Im}\, F(s - i\epsilon) \,. \tag{136}$$

In the large m limit, the real part of (135) is fixed by (131), namely, setting $s = s_i$,

$$\mathrm{Re}\, F(s - i\epsilon) \equiv \frac{1}{m} \langle \sum_j \frac{1}{s - s_j} \rangle = 1 \,. \tag{137}$$

From the discussion of physical equilibrium of the Dyson gas (see the paragraph preceding (114)), we expect the $\{s_i\}$ to be contained in a single finite segment $0 \leq s \leq a$, with a yet to be determined. This means that $F(w)$ should have a cut (along the real axis, where the eigenvalues of $B^T B$ are found) connecting $w = 0$ and a. Furthermore, $\rho(s)$ must be integrable as $s \to 0+$, since a macroscopic number (i.e., a finite fraction of m) of eigenvalues cannot condense at $s = 0$, due to repulsion. These considerations, together with (137) lead [23, 24] to the reasonable ansatz

$$F(w) = 1 + \left(\frac{p}{w} + q \right) \sqrt{w(w - a)} \,, \tag{138}$$

with parameters p and q. The asymptotic behavior (134) then immediately fixes

$$q = 0 \,, p = -1 \,, \text{and } a = 2 \,. \tag{139}$$

Thus,

$$F(w) = 1 - \sqrt{\frac{w-2}{w}}\,. \tag{140}$$

The eigenvalue distribution of $B^T B$ is therefore

$$\rho(s) = \frac{1}{\pi}\,\mathrm{Im}\;F(s - i\epsilon) = \frac{1}{\pi}\sqrt{\frac{2-s}{s}} \tag{141}$$

for $0 < s < 2$, and zero elsewhere. As a simple check, note that

$$\int_0^2 \rho(s) ds = 1\,,$$

as guaranteed by the unit numerator in (134).

Thus, as mentioned in the previous appendix, ω_i^2, the eigenvalues of $B^T B$, are confined in a finite segment $0 < s < 2$. In the limit $m \to \infty$, they form a continuous condensate in this segment, with non uniform distribution (141).

In an obvious manner, we can calculate S_*, the extremal value of S in (111), by replacing the discrete sums over the s_i by continuous integrals with weights $\rho(s)$ given by (141). We do not calculate S_* explicitly, but merely mention the obvious result that it is a number of $\mathcal{O}(m^2)$. Similarly, from (115) and (141) we obtain

$$I_*(y) = \int_0^2 \rho(s)\,\log\left(1 + \frac{y}{s}\right)\,ds = \frac{1}{\pi}\int_0^2 \sqrt{\frac{2-s}{s}}\,\log\left(1+\frac{y}{s}\right)\,ds\,. \tag{142}$$

Since the continuum approximation for $\rho(s)$ introduces an error of the order $1/m$, an error of similar order is introduced in I_*. It is easier to evaluate $\frac{dI_*(y)}{dy}$, and then integrate back, to obtain $I_*(y)$. We find from (142)

$$\frac{dI_*(y)}{dy} = -F(-y) = -1 + \frac{y+2}{\sqrt{y^2+2y}} = -1 + \sqrt{1 + \frac{2}{y}}\,. \tag{143}$$

It is clear from the last equality in (143) that

$$\frac{dI_*(y)}{dy} > 0 \tag{144}$$

for $y > 0$. Integrating (143), and using (113), $I_*(0) = 0$, to determine the integration constant, we finally obtain

$$I_*(y) = -y + \sqrt{y^2+2y} + \log\left(y + 1 + \sqrt{y^2+2y}\right)\,. \tag{145}$$

From (145) we obtain the limiting behaviors

$$I_*(y) = 2\sqrt{2y} - y + \mathcal{O}(y^{3/2})\,, \qquad 0 \le y << 1\,, \tag{146}$$

and

$$I_*(y) = \log\left(\frac{2y}{e}\right) + \mathcal{O}\left(\frac{1}{y}\right)\,, \qquad y >> 1\,. \tag{147}$$

Due to (143), $I_*(y)$ increases monotonically from $I_*(0) = 0$ to its asymptotic form (147). Note that for $y = t/m$ (as required in (117)), the second term in (146) is $\mathcal{O}(1/m)$ and therefore it is beyond the accuracy of the approximation of this section.

Acknowledgments

It is our great pleasure to thank Arkadi Nemirovski, Eduardo Sontag and Ofer Zeitouni for stimulating and informative discussions. This research was supported in part by the US-Israel Binational Science Foundation (BSF), by the Israeli Science Foundation grant number 307/98 (090-903), by the US National Science Foundation under Grant No. PHY99-07949, by the Minerva Center of Nonlinear Physics of Complex Systems and by the fund for Promotion of Research at the Technion.

References

[1] J. Hertz, A. Krogh, and R. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood City, 1991.

[2] A. Cichocki and R. Unbehauen. *Neural networks for optimization and signal processing*. John Wiley, 1993.

[3] X.B. Liang and J. Wang. A recurrent neural network for nonlinear optimization with a continuously differentiable objective function and bound constraints. *IEEE transaction on neural networks*, 2000.

[4] C. Mead. *Analog VLSI and Neural Systems*. Addison-Wesley, 1989.

[5] R. W. Brockett. Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems. *Linear Algebra and Its Applications*, 146:79–91, 1991.

[6] L. Faybusovich. Dynamical systems which solve optimization problems with linear constraints. *IMA Journal of Mathematical Control and Information*, 8:135–149, 1991.

[7] U. Helmke and J.B. Moore. *Optimization and Dynamical Systems*. Springer Verlag, London, 1994.

[8] M.S. Branicky. Analog computation with continuous ODEs. In *Proceedings of the IEEE Workshop on Physics and Computation*, pages 265–274, Dallas, TX, 1994.

[9] C. Papadimitriou. *Computational Complexity*. Addison-Wesley, Reading, Mass., 1995.

[10] L.O. Chua and G.N. Lin. Nonlinear programming without computation. *IEEE transaction on circuits and systems*, 31(2), 1984.

[11] A. Ben-Hur, H.T. Siegelmann, and S. Fishman. A theory of complexity for continuous time dynamics. Accepted, Journal of Complexity.

[12] H.T. Siegelmann and S. Fishman. Computation by dynamical systems. *Physica D*, 120:214–235, 1998.

[13] L. Blum, F. Cucker, M. Shub, and S. Smale. *Complexity and real Computation*. Springer-Verlag, 1999.

[14] J.F. Traub and H. Wozniakowski. Complexity of linear programming. *Operations Research Letters*, 1:59–62, 1982.

[15] J. Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optimization*, 5(3):506–524, 1995.

[16] S. Smale. On the average number of steps in the simplex method of linear programming. *Math. Programming*, 27:241–262, 1983.

[17] M.J. Todd. Probabilistic models for linear programming. *Mathematics of Operations Research*, 16:671–693, 1991.

[18] R. Shamir. The efficiency of the simplex method: A survey. *Management Science*, 33(3):301–334, 1987.

[19] K.M. Anstreicher, J. Ji, F.A. Potra, and Y. Ye. Probabilistic analysis of an infeasible interior-point algorithm for linear programming. *Mathematics of Operations Research*, 24:176–192, 1999.

[20] Y. Ye. *Interior Point Algorithms: Theory and Analysis*. John Wiley and Sons Inc., 1997.

[21] Y. Ye. Toward probabilistic analysis of interior-point algorithms for linear programming. *Mathematics of Operations Research*, 19:38–52, 1994.

[22] R. Saigal. *Linear Programming*. Kluwer Academic, 1995.

[23] E. Brézin, C. Itzykson, G. Parisi and J. -B. Zuber, Planar diagrams. *Comm. Math. Phys.* **59** (1978) 35.

[24] Some papers that treat random *real* rectangular matrices, such as the matrices relevant for this work, are:
A. Anderson, R. C. Myers and V. Periwal, Complex random surfaces. *Phys. Lett.***B 254**(1991) 89,
Branched polymers from a double scaling limit of matrix models. *Nucl. Phys.* **B 360**, (1991) 463 (Section 3).
J. Feinberg and A. Zee, Renormalizing rectangles and other topics in random matrix theory. *J. Stat. Mech.* **87** (1997) 473-504.
For earlier work see: G.M. Cicuta, L. Molinari, E. Montaldi and F. Riva, Large rectangular random matrices. *J. Math.Phys.* **28** (1987) 1716.

[25] M.L. Mehta. *Random Matrices*. Academic Press, Boston, 2nd edition edition, 1991.

[26] S. Mizuno, M.J. Todd, , and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18:964–981, 1993.

[27] J. Feinberg. On the universality of the probability distribution of the product $B^{-1}X$ of random matrices. arXiv:math.PR/0204312, 2002.

[28] V.L. Girko. On the distribution of solutions of systems of linear equations with random coefficients. *Theory of probability and mathematical statistics*, 2:41–44, 1974.

[29] B.V. Gnedenko and A.N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*. Addison Wesley, Reading, MA, 1954.

[30] K.G. Wilson and J. Kogut, *The renormalization group and the epsilon expansion*. *Phys. Rep.* **12**, (1974) 75.
J. Cardy, *Scaling and Renormalization in Statistical Physics*. Cambridge University Press, Cambridge, 1996.

[31] E. Abrahams, P. W. Anderson, D. C. Licciardelo and T. V. Ramakrishnan, Scaling theory of localization: absence of quantum diffusion in two dimensions. *Phys. Rev. Lett.* **42** (1979) 673;
E. Abrahams, P. W. Anderson, D. S. Fisher and D. J. Thouless, New method for a scaling theory of localization. *Phys. Rev.* **B 22** (1980) 3519.
P. W. Anderson, New method for scaling theory of localization. II. Multichannel theory of a "wire" and possible extension to higher dimensionality. *Phys. Rev.* **B 23** (1981) 4828.

[32] B.L. Altshuler, V.E. Kravtsov and I.V. Lerner, in *Mesoscopic Phenomena in Solids*, ed. B.L. Altshuler, P.A. Lee and R.A. Webb. North Holland, Amsterdam, 1991.

[33] C.M. Bender and S.A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*. Springer Verlag, New York, 2nd edition, 1999. (Chapter 8).

[34] R. Durrett. *Probability: Theory and Examples*. Wadswarth Publishing Co., Belmont, 2nd edition, 1996. (Chapter 2).