Solving Monotone Polynomial Equations

Javier Esparza, Stefan Kiefer, and Michael Luttenberger

Abstract We survey some recent results on iterative methods for approximating the least solution of a system of monotone fixed-point polynomial equations.

1 Introduction

Consider the following problem formulated by Francis Galton in the (politically incorrect) 19th century [26], and quoted by Thomas Harris in his classical text on branching stochastic processes [20]:

> Let $p_0, p_1, p_2 \ldots$ be the respective probabilities that a man has 0, 1, 2, ... sons, let each son have the same probability for sons of his own, and so on. What is the probability that the male line is extinct after r generations, and more generally what is the probability for any given number of descendants in the male line in any given generation?

We are interested here in the probability that the male line *eventually* becomes extinct. A little thought shows that this probability is a solution of the fixed-point equation

$$X = \sum_{n \geq 0} p_n X^n \tag{1}$$

and after some more thought one concludes that it is in fact the smallest solution.

Consider now the following stochastic context-free grammar (i.e., a grammar whose productions are annotated with probabilities) with axiom X:

$$\begin{array}{lll} X \xrightarrow{0.4} XY, & X \xrightarrow{0.6} a & \\ Y \xrightarrow{0.3} XY, & Y \xrightarrow{0.4} YZ, & Y \xrightarrow{0.3} b \\ Z \xrightarrow{0.3} XZ, & Z \xrightarrow{0.7} b & \end{array}$$

Javier Esparza, Stefan Kiefer, Michael Luttenberger
Institut für Informatik, Technische Universität München, 85748 Garching, Germany
e-mail: {esparza,kiefer,luttenbe}@in.tum.de

What is the probability that the grammar eventually generates a word, i.e., a string of non-terminals? Again, it is not to difficult to show that it is equal to the X-component of the least solution of the following system of equations.

$$\begin{aligned} X &= 0.4XY + 0.6 \\ Y &= 0.3XY + 0.4YZ + 0.3 \\ Z &= 0.3XZ + 0.7 \end{aligned} \tag{2}$$

Notice that the vector $(1,1,1)$ is a solution of the system. We will later investigate whether it is the least solution or not.

Equations (1) and (2) are two examples of *monotone systems of polynomial equations* (MSPEs for short). MSPEs are systems of the form

$$\begin{aligned} X_1 &= f_1(X_1,\ldots,X_n) \\ &\vdots \\ X_n &= f_n(X_1,\ldots,X_n) \end{aligned}$$

where $f_1,\ldots,f_n$ are polynomials with *positive* real coefficients. In vector form we denote an MSPE by $X = f(X)$. We call the vector $f(X)$ of polynomials a *monotone system of polynomials*, or MSP. Obviously, a solution of $X = f(X)$ is a fixed-point of $f(X)$, and vice versa. Further, any solution of $X = f(X)$ can be visualized as a point of intersection of the submanifolds defined by the n implicit functions $f_i(X) - X_i = 0$. In particular, when the polynomials of $f(X)$ are quadratic the solutions of $X = f(X)$ correspond to the intersection of n quadrics. Figure 1 shows the graph of such a quadratic MSPE with $n = 2$.

We call MSPEs and MSPs "monotone" because $x \leq x'$ implies $f(x) \leq f(x')$ for every $x, x' \in \mathbb{R}^n_{\geq 0}$. This is a bit imprecise, because not every monotone polynomial has positive coefficients. Perhaps "positive systems" would be a better name, but since we have used the term "monotone" in several papers we stick to it.

MSPEs appear naturally in the analysis of many stochastic models, such as stochastic context-free grammars (with numerous applications to natural language processing [23, 19], and computational biology [24, 5, 4, 22]), probabilistic programs with procedures [9, 2, 13, 11, 10, 12, 14], web-surfing models with back buttons [16, 17], and branching processes [20], a topic in stochastic theory that can be traced back to Galton's problem.

In the last years Etessami and Yannakakis [13] and ourselves [21, 8] have studied the problem of solving MSPEs. This paper gives a succinct—and informal—overview of our results.

2 Some Definitions and Facts

Let $\mathbb{R}_{[0,\infty]}$ denote the set of non-negative reals extended with ∞. We extend the definitions of sum and product as usual: $\infty + k = \infty$ for every $k \in \mathbb{R}_{[0,\infty]}$, $\infty \cdot 0 = 0$,

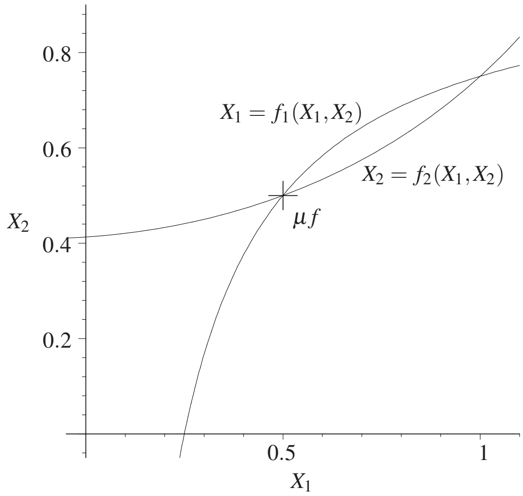

Fig. 1 Graphs of the equations $X_1 = f_1(X_1, X_2)$ and $X_2 = f_2(X_1, X_2)$ with $f_1(X_1, X_2) = X_1 X_2 + \frac{1}{4}$ and $f_2(X_1, X_2) = \frac{1}{6}X_1^2 + \frac{1}{9}X_1 X_2 + \frac{2}{9}X_2^2 + \frac{3}{8}$. There are two real solutions in $\mathbb{R}^2_{[0,\infty]}$, the least one is labelled with μf.

and $\infty \cdot k = \infty$ for every $k \in \mathbb{R}_{[0,\infty]} \setminus \{0\}$. The resulting algebraic structure is the *real semiring*. MSPEs are systems of fixed-point equations over the real semiring.

Given two vectors $u, v \in \mathbb{R}^n_{[0,\infty]}$, we say that $u \le v$ holds if $u_i \le v_i$ holds for every $1 \le i \le n$, where u_i, v_i are the i-th components of u and v, respectively. This is the pointwise order on vectors of reals. The first positive result on MSPEs is a direct consequence of Kleene's theorem:

Theorem 1 (Kleene's fixed-point theorem). *Every MSP $f(X)$ has a least fixed-point μf in $\mathbb{R}^n_{[0,\infty]}$ with respect to the pointwise order. Moreover, the sequence $(\kappa_f^{(k)})_{k \in \mathbb{N}}$ given by*

$$\begin{aligned} \kappa_f^{(0)} &:= 0 \\ \kappa_f^{(k+1)} &:= f(\kappa_f^{(k)}) = f^{k+1}(0) \end{aligned}$$

is non-decreasing with respect to $\le$ (i.e., $\kappa_f^{(k)} \le \kappa_f^{(k+1)}$) and converges to μf.

We call $(\kappa_f^{(k)})_{k \in \mathbb{N}}$ the *Kleene sequence*, and its elements the *Kleene approximants of μf*.

Example 1. For the system (2) we obtain:

$$\begin{array}{lll} \kappa_f^{(0)} = (0,0,0), & \kappa_f^{(1)} = (0.6, 0.3, 0.7), & \\ \kappa_f^{(2)} = (0.672, 0.438, 0.826), & \kappa_f^{(3)} = (0.718, 0.533, 0.867), & \\ \kappa_f^{(4)} = (0.753, 0.600, 0.887), & \kappa_f^{(5)} = (0.781, 0.648, 0.900), & \ldots \end{array}$$

The least solution of a system of *linear* equations (monotone or not) satisfies some good properties that no longer hold for MSPEs. It is easy to show (using for instance Cramer's rule) that if the coefficients are rationals, then the least solution is also rational. However, using Galois theory one can prove that the least solution of a polynomial system may not be expressible by radicals. For instance:

Fact 1. *The least fixed-point of*

$$X = \frac{1}{6}X^6 + \frac{1}{2}X^5 + \frac{1}{3} \,. \tag{3}$$

is not expressible by radicals.

This fact also holds for quadratic systems, i.e., systems in which all polynomials have at most degree 2. Given an MSP f over a set $\mathscr{X}$ of variables, it is easy to construct a quadratic MSP g over a larger set $\mathscr{X} \cup \mathscr{Y}$ such that the projection of μg onto $\mathscr{X}$ is equal to μf. The construction is very similar to the one that brings a context-free grammar in Chomsky normal form. For instance, it "expands" Equation (3) into the system

$$\begin{aligned} X &= \frac{1}{6}XX_5 + \frac{1}{2}XX_4 + \frac{1}{3} \\ X_n &= XX_{n-1} \qquad \text{(for } n = 5,4,3) \\ X_2 &= X^2 \end{aligned}$$

Since this expansion involves only a linear blowup, we can take quadratic MSPEs as a normal form of MSPEs.

The least solution of linear MSPEs is not only rational, but a succinct rational. Consider a system of dimension n (i.e., with n equations) whose coefficients are given as ratios of m-bit integers. It is easy to show using Cramer's rule that the least solution can be written as the quotient of two natural numbers with at most $O(n^2m + n\log n)$ bits. As a consequence, we get

Fact 2. *Let $X = f(X)$ be a linear MSPE of dimension n whose coefficients are given as ratios of m-bit integers. For every component μf_i of the least fixed-point of f: if $0 < \mu f_i < \infty$ then*

$$\frac{1}{2^{O(n^2m+n\log n)}} \le \mu f_i \le 2^{O(n^2m+n\log n)}$$

(where the constant of the Big-Oh notation is independent of f).

Since the least fixed-point of a MSP can be irrational, there is no bound on the number of digits needed to write it down. However, using results of [8] we can still give a lower and an upper bound:

Fact 3. *Let f be a quadratic MSP of of dimension n whose coefficients are given as ratios of m-bit integers. For every component μf_i of the least fixed-point of f: if $0 < \mu f_i < \infty$ then*

$$\frac{1}{2^{m\cdot 2^{O(n)}}} \leq \mu f_i \leq 2^{m\cdot 2^{O(n)}}$$

(where the constant of the Big-Oh notation is independent of f).

So, loosely speaking, while the least fixed-point of a linear system is at most exponential in the dimension of the system, the least solution of a quadratic system is at most double exponential.

It is easy to find examples of quadratic MSPs in which the least fixed-point is rational and double exponential. The n-th component of the least solution of system

$$\begin{aligned} X_1 &= k \\ X_2 &= X_1^2 \\ &\vdots \\ X_n &= X_{n-1}^2 \end{aligned}$$

is equal to $k^{2^{(n-1)}}$.

3 Computational complexity

The fundamental decision problem for MSPs is whether $(\mu f)_i \sim a$ holds for a given MSP f and a component i, where a is some positive rational number and $\sim \in \{\leq, =, \geq\}$. Let us call this problem *MSP-DECISION*. Little is known about its computational complexity. The problem lies in PSPACE:

> Consider e.g. a two-dimensional MSPE $X_1 = f_1(X_1, X_2), X_2 = f_2(X_1, X_2)$. To decide whether $(\mu f)_1 \leq a$ holds one can equivalently decide if the following formula is true:
>
> $$\exists x_1 \in \mathbb{R}, x_2 \in \mathbb{R} : x_1 = f_1(x_1, x_2) \wedge x_2 = f_2(x_1, x_2) \wedge x_1, x_2 \geq 0 \wedge x_1 \leq a$$
>
> Such formulas can be decided in PSPACE, because the first-order theory of the reals is decidable, and its existential fragment is even in PSPACE [3].

For a lower bound, we introduce the problem *SQUARE-ROOT-SUM*:

> Given $k+1$ natural numbers $n_1, \ldots, n_k$ and b, determine whether $\sum_{i=1}^{k} \sqrt{n_i} \leq b$ holds.

The SQUARE-ROOT-PROBLEM is a natural subproblem of many questions in computational geometry. For instance, the length of the boundary of a polygon whose vertices lie in $\mathbb{Z}^2$ is a sum of square roots of integers. It has been a major open problem since the 70s whether SQUARE-ROOT-SUM belongs to NP. The problem can easily be reduced to MSP-DECISION:

> Suppose we are given $n_1 = 2$, $n_2 = 3$, and $b = 3$, and we want to decide if $\sqrt{2} + \sqrt{3} \leq 3$. One would like to come up with an MSP $f(X)$ such that

$(\mu f)_1 = \sqrt{2}, (\mu f)_2 = \sqrt{3}, (\mu f)_3 = \sqrt{2}+\sqrt{3}$, so that deciding $\sqrt{2}+\sqrt{3} \leq 3$ is equivalent to deciding $(\mu f)_3 \leq 3$. One has to be careful though, because for instance the equation $X_1 = X_1^2 + X_1 - 2$ is not an MSPE. It was shown in [13] how to overcome this problem: Instead of encoding e.g. $\sqrt{2}$ directly, it suffices to encode $a + b \cdot \sqrt{2}$ for some rationals a, b.

The least solution of the equation $X = X^2 + (1-\lambda^2 \cdot n)/4$ equals $(1-\lambda\sqrt{n})/2$. So, by choosing for λ a small enough rational number we get a 1-dimensional MSP whose least solution is $a + b \cdot \sqrt{n}$ for some rationals a, b. In our example we can set $\lambda = \frac{1}{\max(2,3)} = \frac{1}{3}$ which leads to the following MSPE.

$$\begin{aligned} X_1 &= X_1^2 + \tfrac{1-\frac{2}{9}}{4} \\ X_2 &= X_2^2 + \tfrac{1-\frac{3}{9}}{4} \\ X_3 &= X_1 + X_2 \end{aligned}$$

Its least solution is

$$\mu f = \left(\frac{1}{2} - \frac{1}{6}\sqrt{2}, \frac{1}{2} - \frac{1}{6}\sqrt{3}, 1 - \frac{1}{6}(\sqrt{2}+\sqrt{3}) \right) .$$

So, the question whether $\sqrt{2}+\sqrt{3} \leq 3$ holds can be translated into the question whether $(\mu f)_3 \geq 1 - \frac{1}{6} \cdot 3 = \frac{1}{2}$ holds.

It follows from this reduction that proving membership of MSP-DECISION in NP would be a major breakthrough.

An interesting issue is the complexity of MSP-DECISION in the Blum-Shub-Smale computational model, in which all operations on rationals take unit time independently of their size. SQUARE-ROOT-SUM can be decided in polynomial time in this model [25], but it is open whether the result extends to MSP-DECISION.

4 Approximating the Least Fixed-Point: Newton's Method

For most practical purposes, the main computational problem concerning MSPs is the approximation of the least fixed-point up to a given accuracy. Kleene's method can be applied, and it is very robust: it always converges when started at 0, for any MSP. On the other hand, the convergence speed of the Kleene sequence can be very poor. Before presenting an example, we define a notion of convergence order that differs from the one commonly used in numerical mathematics, but is particularly natural for computer science.

Let $(a_k)_{k\geq 0}$ be a non-decreasing sequence of vectors over the real semiring such that $\lim_{k\to\infty} a_k = a < \infty$. The *convergence order* of the sequence is the function $\beta : \mathbb{N} \to \mathbb{N}$ defined as follows: $\beta(k)$ is the greatest natural number i such that

$$\frac{\|a - a_k\|}{\|a\|} \leq 2^{-i}$$

where $\|\cdot\|$ is some norm. We say that a sequence has linear, exponential, logarithmic, etc. convergence order if the function $\beta(k)$ grows linearly, exponentially, or logarithmically in k, respectively. Notice that the asymptotic behaviour of $\beta(k)$ is independent of the norm, because all norms are equivalent up to a constant. In the univariate case, $\beta(k)$ is the number of bits of a_k that coincide with the corresponding bits of a (the formalization of this intuition requires some care, like identifying 1 and $0.999\ldots$). For instance, for the sequence $(1-2^{-k})_{k\geq 0}$ we have $\beta(k) = k$, i.e., the first k bits of the k-th element of the sequence coincide with the first k bits of the limit.

Consider now this very simple but at the same time very illustrative quadratic MSPE in one variable:

$$X = 1/2 + 1/2X^2 \tag{4}$$

In Galton's problem, the least solution of this equation gives the extinction probability of an individual's descent line when every individual has 0 or 2 children with probability $1/2$. The least solution is 1. We have:

Fact 4. *The i-th Kleene approximant of $X = 1/2 + 1/2X^2$ satisfies $\kappa^{(i)} \leq 1 - \frac{1}{i+1}$ for every $i \geq 0$. So the Kleene sequence only has logarithmic convergence order.*

Example 2. Here are some of the Kleene iterates.

$$\begin{array}{lll}
\kappa^{(0)} = 0, & \kappa^{(1)} = 0.5, & \kappa^{(2)} = 0.625 \\
\kappa^{(3)} = 0.695, & \kappa^{(4)} = 0.742, & \kappa^{(5)} = 0.775 \\
\quad\ldots & & \\
\end{array}$$

$$\kappa^{(20)} = 0.920,\ \ldots,\ \kappa^{(200)} = 0.990,\ \ldots,\ \kappa^{(2000)} = 0.9990,\ \ldots$$

Faster approximation techniques have been known for a long time. In particular, Newton's method, suggested by Isaac Newton more than 300 years ago, is a standard efficient technique for approximating a zero of a differentiable function. Since the least solution of a fixed-point equation $X = f(X)$ is a zero of $g(X) = f(X) - X$, the method can be applied to search for fixed-points of $f(X)$.

We briefly recall the method for the case of one variable, see Fig. 2 for an illustration. Starting at some value $\nu^{(0)}$ "close enough" to the zero of $g(X)$, we proceed iteratively: given $\nu^{(i)}$, we compute a value $\nu^{(i+1)}$ closer to the zero than $\nu^{(i)}$. For that, we compute the tangent to $g(X)$ passing through the point $(\nu^{(i)}, g(\nu^{(i)}))$, and take $\nu^{(i+1)}$ as the zero of the tangent (i.e., the X-coordinate of the point at which the tangent cuts the X-axis). A little arithmetic leads to:

$$\nu^{(i+1)} = \nu^{(i)} + \frac{f(\nu^{(i)}) - \nu^{(i)}}{1 - f'(\nu^{(i)})}$$

Newton's method can be easily generalized to the multivariate case:

$$\nu^{(i+1)} = \nu^{(i)} + (\mathrm{Id} - f'(\nu^{(i)}))^{-1}(f(\nu^{(i)}) - \nu^{(i)})$$

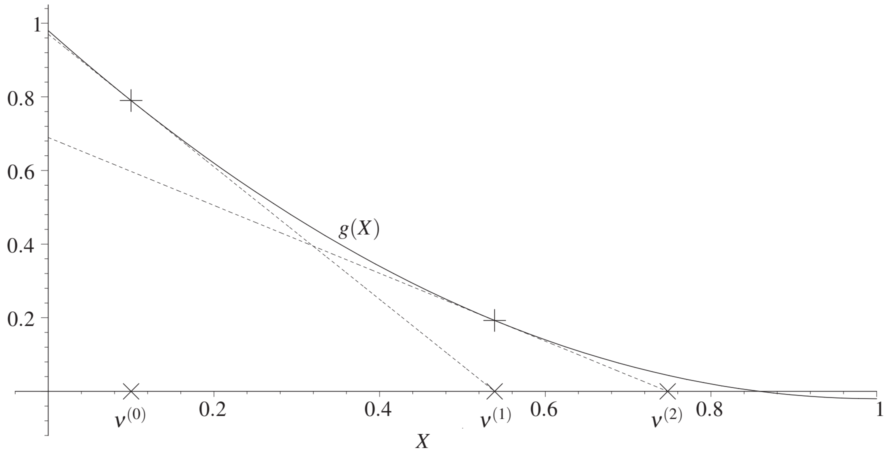

Fig. 2 Newton's method to find a zero of a one-dimensional function $g(X)$

where $f'(X)$ is the Jacobian of f, i.e., the matrix of partial derivatives of f, and Id is the identity matrix.

Notice that Newton's method is not restricted to the real semiring, it can be applied to any differentiable functions over the real field. However, when applied with this generality it is far less robust than Kleene's method: it may converge very slowly, converge only when started at a point very close to the zero—which must be guessed— or even not converge at all.

However, if we apply Newton's method to $f(X) = 1/2 + 1/2X^2$, starting at $\nu^{(0)} = 0$, we obtain:

Fact 5. *The i-th Newton approximant of* $X = 1/2 + 1/2X^2$ *satisfies* $\nu^{(i)} = 1 - \frac{1}{2^i}$ *for every* $i \geq 0$. *The i-th approximant has i correct bits, i.e., the Newton sequence has linear convergence.*

So in this particular example the Newton sequence converges "exponentially faster" than the Kleene sequence. The number of arithmetic operations needed to compute i correct bits of the solution grows polynomially instead of exponentially in i. (Recall, however, that the operations have to be applied to rationals whose length may grow exponentially in the number of iterations.) One can ask whether the good behaviour on this example is just a coincidence, or whether perhaps Newton's method is robust on the real semiring. A number of recent results have shown that (with certain ifs and buts) the latter is the case, and we briefly survey them in the next section.

5 Convergence Order and Thresholds for Newton's Method

The first positive result on the convergence of Newton's method was obtained by Etessami and Yannakakis in [13]. They showed that the method always converges to the least fixed-point starting from $\nu^{(0)} = 0$, and that it converges at least as fast at the Kleene sequence.[1]

Inspired by this positive result, we started to study the convergence order. Given an MSPE $X = f(X)$ whose least solution μf is finite, it is well-known that the convergence order depends critically on the Jacobian matrix at the least fixed-point, i.e., on $f'(\mu f)$. Every textbook proves that the method performs brilliantly when the matrix $(\mathrm{Id} - f'(\mu f))$ is non-singular: it exhibits *exponential* convergence order. So we focused our attention on the singular case, of which $f(X) = 1/2 + 1/2X^2$ is an example. By Fact 5 we can expect at most linear convergence. But perhaps the method converges more slowly on other examples?

It is convenient to start with the special case of *strongly connected* MSPEs. Loosely speaking, an MSPE is strongly connected if every variable depends on any other variable, where dependence is defined as follows. Given two variables X and Y, X *depends on* Y if either Y appears on the right-hand-side of the equation for X, or if there is a variable Z such that X depends on Z and Z depends on Y.

5.1 Strongly Connected MSPEs

We proved the following theorem in [21].

Theorem 2. *Let $f(X)$ be a strongly connected MSP such that μf is finite. There is a number t_f such that for every $i \geq 0$:*

$$\beta(t_f + i) \geq i\ .$$

In particular, the Newton sequence has linear convergence order.

We call t_f the *threshold* of $f(X)$. Loosely speaking, the theorem states that after crossing the threshold (i.e., from the t_f-th approximant onwards) the Newton sequence gains at least one bit of accuracy per iteration. The threshold itself is an upper bound on the number of iterations needed to obtain the first bit of the least fixed-point.

The proof of [21] was based on the following topological property of $\mathbb{R}^n$: if the infimum of the distances between points of two compact sets is 0, then the two sets have at least one common point. As a consequence, it was a purely existential proof, and provided no information on the size of the threshold. In [7] we obtained the following relation between the threshold and the minimal component of μf.

[1] More precisely, Etessami and Yannakakis proved the result for a structured version of the method, and we showed in [8] that this additional structure is not required for convergence (although it is convenient for efficiency).

Theorem 3. *Let $f(X)$ be a quadratic strongly connected MSP of dimension n whose coefficients are given as ratios of m-bit integers. Let μ_{min} be the minimal component of μf. The threshold t_f of Theorem 2 satisfies*

$$t_f \leq 3n^2 m + 2n^2 \left|\log \mu_{min}\right| .$$

In particular, if $f_i(0) > 0$ holds for every $1 \leq i \leq n$, then $t_f \leq 5n^2 m$.

Example 3. Consider again the following MSPE, which was given as Equation (2) on page 2.

$$\begin{aligned} X &= 0.4XY + 0.6 \\ Y &= 0.3XY + 0.4YZ + 0.3 \\ Z &= 0.3XZ + 0.7 \end{aligned}$$

Using a result from [8], slightly stronger than Theorem 3 but technically more difficult to state, one can prove that the threshold of this system satisfies $t_f \leq 6$ for the maximum-norm (i.e., the norm of a vector is the absolute value of its maximal component). So $\beta(14) \geq 8$. After computing 14 Newton iterates we get $\nu^{(14)} = (0.983, 0.974, 0.993)$. As we have computed at least $\beta(14) \geq 8$ bits, we know that μf is at most $\nu^{(14)} + (2^{-8}, 2^{-8}, 2^{-8})$ which is strictly less than 1 in every component. Therefore, the stochastic context-free grammar from the introduction produces a terminal string with probability less than 1.

Combining Theorem 3 with Fact 3 we obtain:

Corollary 1. *Let $X = f(X)$ be a quadratic strongly connected MSPE of dimension n whose coefficients are given as ratios of m-bit integers. The threshold t_f of Theorem 2 satisfies $t_f \in m2^{O(n)}$.*

This corollary gives an exponential bound on the number of iterations needed to compute the first bit of the least fixed-point. It is open whether this bound is tight.

5.2 General MSPEs

The following example shows that an exponential number of iterations is sometimes needed for the first bit, if the MSPE is *not strongly connected*. We give a family of MSPEs in which the number of iterations needed to compute the first bit grows exponentially in the dimension of the system.

Example 4. Consider the following family of MSPEs.

$$
\begin{aligned}
X_1 &= 1/2 + 1/2 \cdot X_1^2 \\
X_2 &= 1/4 \cdot X_1^2 + 1/2 \cdot X_1 X_2 + 1/4 \cdot X_2^2 \\
&\vdots \\
X_n &= 1/4 \cdot X_{n-1}^2 + 1/2 \cdot X_{n-1} X_n + 1/4 \cdot X_n^2
\end{aligned}
\tag{5}
$$

The variable X_i depends on X_j if and only if $j \leq i$. So the dependence graph contains n strongly connected components, one for each variable. The least fixed-point of the system is the vector $(1,1,\ldots,1)$. We show in [21] that $\nu_n^{(2^{n-1})} \leq 1/2$ holds, and so that at least 2^{n-1} iterations of Newton's method are needed to obtain the first bit of X_n. The proof goes as follows. We consider a decomposed version of Newton's method, in which for a given k we perform k iterations of the normal method on the first equation, yielding a lower bound $a_1^{(k)}$ of the first component of the least fixed-point. Then we perform k iterations on the second equation *after setting* $X_1 := a_1^{(k)}$; by monotonicity, this yields a lower bound $a_2^{(k)}$ of the second component. Repeating this procedure we finally obtain a lower bound $a_n^{(k)}$ of the n-th component. It is easy to see that $\nu_i^{(k)} \leq a_i^{(k)}$ holds, i.e, the decomposed method converges at least as fast as the method that performs k iterations on the whole system. Now, let $\delta_i^{(k)} = 1 - a_i^{(k)}$ be the error of the decomposed method. A simple analysis reveals that $\delta_{i+1}^{(k)} \geq \sqrt{\delta_i^{(k)}}$ holds for every $1 \leq i < n$. By Fact 5 we have $\delta_1^{(2^{n-1})} = (1/2)^{2^{n-1}}$, and so we get $\delta_n^{(2^{n-1})} \geq 1/2$, i.e., $\nu_n^{(2^{n-1})} \leq 1/2$.

So, intuitively, the problem of non-strongly connected systems is that the error gets "amplified" when we move up the graph of strongly connected components.

For MSPs that are not strongly connected, Newton's method still has linear convergence order, but a worse rate [8]:

Theorem 4. *Let $f(X)$ be a* clean *(see below) MSP such that μf is finite. There is a number t_f such that for every $i \geq 0$:*

$$\beta(t_f + i \cdot (n+1) \cdot 2^n) \geq i\ .$$

In particular, the Newton sequence has linear convergence order.

In order to make sure that Newton's method stays well-defined (i.e. that the matrix inverses exist) Theorem 4 assumes that the MSP is clean, i.e., $(\mu f)_i > 0$ for all i. An MSP can easily be made clean in linear time by identifying and removing the components with $(\mu f)_i = 0$: $(\mu f)_i = 0$ holds iff $(\kappa^{(n)})_i = 0$.

The rate in Theorem 4 is worse than in the strongly connected case: Newton's method needs (in the worst case) about 2^n iterations per bit, instead of only 1 as in the strongly connected case. This worst case is attained by the MSPE in Equation (5) above, so the exponential rate in Theorem 4 cannot be avoided. Unfortunately, we do not have an upper bound on the threshold t_f in this general case.

5.3 min-max-MSPEs

Theorem 4 forms the basis for the convergence analysis of a recent extension [6] of Newton's method to min-max-MSPEs, i.e., MSPEs where minimum and maximum are allowed as additional operators. Here is an example of a min-max-MSPE:

$$\begin{aligned} X &= \max\{0.7Y + 0.3\ , \quad 0.6XY + 0.4\} \\ Y &= \min\{X\ , \quad 0.8Y^2 + 0.2\} \end{aligned}$$

Such systems arise, for instance, in *extinction games*. Those games add two adversarial players to Galton's setting from the beginning: There are n species, each of which is controlled by one of two players, the *terminator* and the *rescuer*. Each player can apply actions to the individuals controlled by her; an action transforms an individual (probabilistically) into zero or more individuals. The terminator tries to extinguish all individuals, whereas the rescuer tries to save them. Natural questions are: What are optimal strategies[2] for the terminator and the rescuer, and what is the probability of extinction of all individuals, assuming that there is a single initial individual and the players follow optimal strategies?

The MSPE above can be thought of as an equation system for the extinction probabilities of two species X and Y. Species X is controlled by the terminator, whereas Y is controlled by the rescuer. The terminator can apply one of two possible actions to an X-individual: the first one kills the X-individual with probability 0.3, but with probability 0.7 transforms it to a Y-individual; the second action kills the X-individual with probability 0.4, but, with probability 0.6, keeps the X-individual and creates a Y-individual. What can the rescuer do with a Y-individual? She can choose between transforming it to an X-individual and a second action which kills the Y-individual with probability 0.3 and adds another Y-individual with probability 0.7.

It turns out that the X-component (resp. Y-component) of the least solution of the MSPE above equals the extinction probability assuming a single initial X-individual (resp. Y-individual) if both the terminator and the rescuer follow optimal strategies. Such systems also arise in the analysis of recursive simple stochastic games [14, 15].

In order to approximate the least solution of a min-max-MSPE, one could use Kleene iteration. But, as we have seen before (Fact 4), Kleene iteration may converge very slowly even without minimum and maximum. Therefore, in [6] we propose two methods for approximating the least solution of a min-max-MSPE. Both are iterative procedures based on Newton's method.

- The first method linearizes each polynomial appearing in the system (possibly inside a minimum or a maximum expression) by computing the "tangent" at the current iterate. One obtains a min-max-MSP whose polynomials have degree at most 1. Its least fixed-point can be computed exactly by a method from [18] that uses strategy iteration and linear programming. The result is the next iterate.
- The second method linearizes each max-polynomial appearing in the system (possibly inside a minimum expression) by computing the "tangent" at the cur-

[2] A strategy tells a player which action to apply to the individuals controlled by her.

rent iterate. (A special "tie breaking" policy must be adhered to if the current iterate is at the "edge" between two polynomials inside a maximum expression.) One obtains a min-MSP whose polynomials have degree at most 1. Its least fixed-point can be computed exactly by solving a single linear program. The result is the next iterate.

Both methods have at least linear convergence order [6]:

Theorem 5. *Let $f(X)$ be a min-max-MSP such that μf is finite. There is a number t_f such that for every $i \geq 0$:*

$$\beta(t_f + i \cdot m \cdot (n+1) \cdot 2^n) \geq i \ ,$$

where m is the number of possible strategies of the players. In particular, the two extensions of Newton's method have linear convergence order.

The first method converges somewhat faster whereas a single step of the second method is cheaper. The second method also computes ε-optimal strategies for the terminator, i.e., strategies that achieve as extinction probabilities at least the current iterate.

We have used the second method to approximate the extinction probabilities assuming perfect strategies: A population that starts with a single X-individual (resp. Y-individual) becomes extinct with probability 0.475 (resp. 0.250). We have obtained those numbers after performing 3 iterations and then rounding, but in this case those numbers are already the exact solution. The optimal strategy for the terminator is to apply the first action to the X-individuals. The rescuer should choose her second action for her Y-individuals.

6 Conclusions

We have shown that Newton's method is not only efficient but also remarkably robust when applied to monotone systems of fixed-point equations (MSPEs). Unlike for arbitrary systems, the method always converges when started at 0. For strongly connected systems the method always reaches a point, the threshold, after which it is guaranteed to gain at least one bit of accuracy per iteration (in favourable cases it *doubles* the number per iteration). In fewer words, after crossing the threshold the method has linear convergence order with rate 1. If the system is not strongly connected the method still has linear convergence, but the rate deteriorates.

The threshold of the strongly connected case is inversely proportional to the logarithm of the minimal component of the least fixed-point. Therefore, if some kind of analysis can establish that the least fixed-point is not very small, then the method quickly enters the one-bit-per-iteration zone. We still do not have any threshold for the general, non-strongly-connected case.

Newton's method still works for MSPEs that are not strongly connected. We have shown that the convergence order is still linear, albeit the rate may deteriorate exponentially with the dimension.

Newton's method can be extended to min-max-MSPEs, preserving its linear convergence order.

MSPEs appear in a large number of stochastic systems. In [1] we have designed a formal system for establishing the reputation of the individuals of a social network. The reputation of the individuals (defined as the stationary distribution of a Markov chain) is the least solution of a MSPE. These case studies lead to very large MSPEs, and computing their least solutions is an exciting challenge for future research.

References

1. Bouajjani, A., Esparza, J., Schwoon, S., Suwimonteerabuth, D.: SDSIrep: A reputation system based on SDSI. In: Proceedings of TACAS (Tools and Algorithms for the Construction and Analysis of Systems), *LNCS*, vol. 4963, pp. 501–516. Springer (2008)
2. Brázdil, T., Kučera, A., Stražovský, O.: On the decidability of temporal properties of probabilistic pushdown automata. In: Proceedings of STACS'2005, *LNCS*, vol. 3404, pp. 145–157. Springer (2005)
3. Canny, J.: Some algebraic and geometric computations in PSPACE. In: Proceedings of STOC, pp. 460–467 (1988)
4. Dowell, R., Eddy, S.: Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. BMC Bioinformatics **5**(71) (2004)
5. Durbin, R., Eddy, S., Krogh, A., Michison, G.: Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press (1998)
6. Esparza, J., Gawlitza, T., Kiefer, S., Seidl, H.: Approximative methods for monotone systems of min-max-polynomial equations. In: Proceedings of ICALP 2008 (to appear)
7. Esparza, J., Kiefer, S., Luttenberger, M.: On fixed point equations over commutative semirings. In: Proceedings of STACS, LNCS 4397, pp. 296–307. Springer (2007)
8. Esparza, J., Kiefer, S., Luttenberger, M.: Convergence thresholds of Newton's method for monotone polynomial equations. In: Proceedings of STACS, pp. 289–300 (2008)
9. Esparza, J., Kučera, A., Mayr, R.: Model-checking probabilistic pushdown automata. In: Proceedings of LICS 2004, pp. 12–21 (2004)
10. Esparza, J., Kučera, A., Mayr, R.: Quantitative analysis of probabilistic pushdown automata: Expectations and variances. In: Proceedings of LICS 2005, pp. 117–126. IEEE Computer Society Press (2005)
11. Etessami, K., Yannakakis, M.: Algorithmic verification of recursive probabilistic systems. In: Proceedings of TACAS 2005, LNCS 3440, pp. 253–270. Springer (2005)
12. Etessami, K., Yannakakis, M.: Checking LTL properties of recursive Markov chains. In: Proceedings of 2nd Int. Conf. on Quantitative Evaluation of Systems (QEST'05) (2005)
13. Etessami, K., Yannakakis, M.: Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations. In: Proceedings of STACS, pp. 340–352. Springer (2005)
14. Etessami, K., Yannakakis, M.: Recursive Markov decision processes and recursive stochastic games. In: Proceedings of ICALP 2005, *LNCS*, vol. 3580, pp. 891–903. Springer (2005)
15. Etessami, K., Yannakakis, M.: Efficient qualitative analysis of classes of recursive Markov decision processes and simple stochastic games. In: STACS, pp. 634–645 (2006)
16. Fagin, R., Karlin, A., Kleinberg, J., Raghavan, P., Rajagopalan, S., Rubinfeld, R., Sudan, M., Tomkins, A.: Random walks with "back buttons" (extended abstract). In: STOC, pp. 484–493 (2000)

17. Fagin, R., Karlin, A., Kleinberg, J., Raghavan, P., Rajagopalan, S., Rubinfeld, R., Sudan, M., Tomkins, A.: Random walks with "back buttons". Annals of Applied Probability **11**(3), 810–862 (2001)
18. Gawlitza, T., Seidl, H.: Precise relational invariants through strategy iteration. In: CSL, pp. 23–40 (2007)
19. Geman, S., Johnson, M.: Probabilistic grammars and their applications (2002)
20. Harris, T.: The Theory of Branching Processes. Springer (1963)
21. Kiefer, S., Luttenberger, M., Esparza, J.: On the convergence of Newton's method for monotone systems of polynomial equations. In: Proceedings of STOC, pp. 217–226. ACM (2007)
22. Knudsen, B., Hein, J.: Pfold: RNA secondary structure prediction using stochastic context-free grammars. Nucleic Acids Research **31**(13), 3423–3428 (2003)
23. Manning, C., Schütze, H.: Foundations of Statistical Natural Language Processing. MIT Press (1999)
24. Sakabikara, Y., Brown, M., Hughey, R., Mian, I., Sjolander, K., Underwood, R., Haussler, D.: Stochastic context-free grammars for tRNA. Nucleic Acids Research **22**, 5112–5120 (1994)
25. Tiwari, P.: A problem that is easier to solve on the unit-cost algebraic RAM. J. Complexity **8**(4), 393–397 (1992)
26. Watson, H., Galton, F.: On the probability of the extinction of families. J. Anthropol. Inst. Great Britain and Ireland **4**, 138–144 (1874)