Convergence Thresholds of Newton's Method for Monotone Polynomial Equations [*]

{esparza,kiefer,luttenbe}@in.tum.de

Javier Esparza, Stefan Kiefer, and Michael Luttenberger
Institut für Informatik
Technische Universität München, Germany [†]

Abstract

Monotone systems of polynomial equations (MSPEs) are systems of fixed-point equations $X_1 = f_1(X_1, \ldots, X_n)$, $\ldots, X_n = f_n(X_1, \ldots, X_n)$ where each f_i is a polynomial with positive real coefficients. The question of computing the least non-negative solution of a given MSPE $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ arises naturally in the analysis of stochastic models like stochastic context-free grammars, probabilistic pushdown automata, and back-button processes. Etessami and Yannakakis have recently adapted Newton's iterative method to MSPEs. In a previous paper we have proved for strongly connected MSPEs the existence of a threshold $k_{\boldsymbol{f}}$ such that after $k_{\boldsymbol{f}}$ iterations of Newton's method each new iteration computes at least 1 new bit of the solution. However, the proof was purely existential. In this paper we give an upper bound for $k_{\boldsymbol{f}}$ as a function of the maximal and minimal components of the least fixed-point $\mu\boldsymbol{f}$ of $\boldsymbol{f}(\boldsymbol{X})$. Using this result we show that $k_{\boldsymbol{f}}$ is at most single exponential resp. linear for strongly connected MSPEs derived from probabilistic pushdown automata resp. from back-button processes. Further, we prove the existence of a threshold for arbitrary MSPEs after which each new iteration computes at least $1/w2^h$ new bits of the solution, where w, h are the width and height of the DAG of strongly connected components.

1 Introduction

A *monotone system of polynomial equations* (MSPE for short) has the form

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

where $f_1, \ldots, f_n$ are polynomials with *positive* real coefficients. In vector form we denote an MSPE by $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$. We call MSPEs monotone because $\boldsymbol{X} \leq \boldsymbol{X}'$ implies $\boldsymbol{f}(\boldsymbol{X}) \leq \boldsymbol{f}(\boldsymbol{X}')$ for $\boldsymbol{X}, \boldsymbol{X}' \in \mathbb{R}^n_{\geq 0}$. MSPEs appear naturally in the analysis of many stochastic models, like context-free grammars (with numerous applications to natural language processing [19, 15], and computational biology [21, 4, 3, 17]), probabilistic programs with procedures [6, 2, 10, 8, 7, 9, 11], and web-surfing models with back buttons [13, 14].

By Kleene's theorem, a feasible MSPE $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ (i.e., an MSPE with at least one solution) has a least solution $\mu\boldsymbol{f}$; this solution can be irrational and non-expressible by radicals. Given an MSPE and a vector $\boldsymbol{v}$ encoded in

[*] This work was supported by the project *Algorithms for Software Model Checking* of the *Deutsche Forschungsgemeinschaft (DFG)*.

[†] Part of this work was done at the Universität Stuttgart.

binary, the problem whether $\mu \boldsymbol{f} \leq \boldsymbol{v}$ holds is in PSPACE and at least as hard as the SQUARE-ROOT-SUM problem, a well-known problem of computational geometry ([10, 12] for more details).

For the applications mentioned above the most important question is the efficient numerical approximation of the least solution. Finding the least solution of a feasible system $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ amounts to finding the least solution of $\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{0}$ for $\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{X}) - \boldsymbol{X}$. For this we can apply (the multivariate version of) *Newton's method* [20]: starting at some $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$ (we use uppercase to denote variables and lowercase to denote values), compute the sequence

$$\boldsymbol{x}^{(k+1)} := \boldsymbol{x}^{(k)} - (\boldsymbol{F}'(\boldsymbol{x}^{(k)}))^{-1}\boldsymbol{F}(\boldsymbol{x}^{(k)})$$

where $\boldsymbol{F}'(\boldsymbol{X})$ is the Jacobian matrix of partial derivatives.

While in general the method may not even be defined ($\boldsymbol{F}'(\boldsymbol{x}^{(k)})$ may be singular for some k), Etessami and Yannakakis proved in [10, 12] that this is not the case for a more structured method, called *Decomposed Newton's method (DNM),* that decomposes the MSPE into *strongly connected components* (SCCs)[1]. We explain this method in some more detail. In order to define the SCCs of an MSPE, associate to $\boldsymbol{f}$ a graph having the variables $X_1, \ldots, X_n$ as nodes, and the pairs (X_i, X_j) such that X_j appears in f_i as edges. A subset of equations of $\boldsymbol{f}$ is an SCC if its associated subgraph is an SCC of the whole graph. DNM starts by computing k iterations of Newton's method for each bottom SCC of the system. The values obtained for the variables of these SCCs are then "frozen", and their corresponding equations removed. The same procedure is then applied to the new bottom SCCs, again with k iterations, until all SCCs have been processed. Etessami and Yannakakis prove the following properties of DNM:

(a) The Jacobian matrices of all the SCCs remain invertible all the way throughout.

(b) The vector $\boldsymbol{x}^{(k)}$ delivered by the method converges to $\mu \boldsymbol{f}$ when $k \to \infty$ *even if* $\boldsymbol{x}^{(0)} = \boldsymbol{0} = (0, \ldots, 0)^\top$.

Property (b) is in sharp contrast with the non-monotone case, where Newton's method may not converge or may exhibit only *local* convergence, i.e., the method may converge only in a small neighborhood of the zero.

The results of [10, 12] provide no information on the number of iterations needed to compute i *valid bits* of $\mu \boldsymbol{f}$, i.e., to compute a vector $\boldsymbol{\nu}$ such that $\left|\mu \boldsymbol{f}_j - \nu_j\right| / \left|\mu \boldsymbol{f}_j\right| \leq 2^{-i}$ for every $1 \leq j \leq n$. In a former paper [16] we have obtained a first positive result on this problem. We have proved that for every strongly connected MSPE $\boldsymbol{f}$ there exists a threshold $k_{\boldsymbol{f}}$ such that for every $i \geq 0$ the $(k_{\boldsymbol{f}} + i)$-th iteration of Newton's method has at least i valid bits of $\mu \boldsymbol{f}$. Loosely speaking, after reaching the threshold DNM is guaranteed to compute at least 1 new bit of the solution per iteration; we say that DNM converges *linearly with rate 1*.

The problem with this result is that its proof provides no information on $k_{\boldsymbol{f}}$ other than its existence. In this paper we prove that the threshold $k_{\boldsymbol{f}}$ can be chosen as

$$k_{\boldsymbol{f}} = 3n^2 m + 2n^2 \left|\log \mu_{\text{min}}\right|$$

where n is the number of equations of the MSPE, m is such that all coefficients of the MSPE can be given as ratios of m-bit integers, and μ_{min} is the minimal component of the least solution $\mu \boldsymbol{f}$.

It can be objected that $k_{\boldsymbol{f}}$ depends on $\mu \boldsymbol{f}$, which is precisely what Newton's method should compute. However, for MSPEs coming from probabilistic models as the ones listed above we can do far better. The following observations and results help to deal with μ_{min}:

- We obtain a syntactic bound on μ_{min} for probabilistic programs with procedures (having stochastic context-free grammars and back-button stochastic processes as special instances) and prove that in this case $k_{\boldsymbol{f}} \leq n 2^{n+2} m$.

[1] More precisely, the proof also requires the MSPE to be *clean*, see Section 2 for details.

- We show that if every procedure has a non-zero probability of terminating, then $k_{\boldsymbol{f}} \leq 3nm$. This condition always holds in the special case of back-button processes [13, 14]. Hence, our result shows that i valid bits can be computed in time $\mathcal{O}((nm+i) \cdot n^3)$ in the unit cost model of Blum, Shub and Smale [1], where each single arithmetic operation over the reals can be carried out exactly and in constant time. It was proved in [13, 14] by a reduction to a semidefinite programming problem that i valid bits can be computed in poly(i, n, m)-time in the classical (Turing-machine based) computation model. We will not improve this result, because we do not have a proof that round-off errors (which are inevitable on Turing-machine based models) do not crucially affect the convergence of Newton's method. But our result sheds light on the convergence of a practical method to compute $\mu\boldsymbol{f}$.

- Finally, since $\boldsymbol{x}^{(k)} \leq \boldsymbol{x}^{(k+1)} \leq \mu\boldsymbol{f}$ holds for every $k \geq 0$, as Newton's method proceeds it provides better and better lower bounds for μ_{min} and thus for $k_{\boldsymbol{f}}$. To demonstrate this, in the paper we exhibit a concrete MSPE and after a few iterations use our theorem to prove that no component of the solution will reach the value 1 (which no further number of iterations can prove by itself).

The paper contains two further results. In [16] we left open the problem whether DNM converges linearly for non-strongly-connected MSPEs. We prove that this is the case. But the convergence rate is poorer: if h and w are the height and width of the graph of SCCs of $\boldsymbol{f}$, then there is a threshold $\widetilde{k}_{\boldsymbol{f}}$ such that $\widetilde{k}_{\boldsymbol{f}} + i \cdot w \cdot 2^{h+1}$ iterations of DNM compute at least i valid bits of $\mu\boldsymbol{f}$. We also give an example where DNM needs at least $i \cdot 2^h$ iterations for i valid bits.

The final result of the paper brings us back to Etessami and Yannakakis' original motivation for DNM. They introduced the decomposition into SCCs as a tool for proving well-definedness: they showed that the Jacobian exists for all SCCs, which implies that DNM is always defined. Here we prove that the Jacobian of the whole MSPE is guaranteed to exist, whether the MSPE is strongly connected or not. As a consequence, one can safely replace DNM by the standard Newton's method. Still, since DNM can be far more efficient (its iterations concern only SCCs, which can be much smaller than the whole MSPE), and since SCCs play an important part in our threshold analysis, we have formulated our results in terms of DNM.

The paper is structured as follows. In Section 2 we state preliminaries and give some background on Newton's method applied to MSPEs. Sections 3, 5, and 6 contain the three results of the paper. Section 4 contains applications of our main result. We conclude in Section 7. Missing proofs can be found in an appendix.

2 Preliminaries

In this section we introduce our notation used in the following and formalize the concepts mentioned in the introduction.

2.1 Notation

As usual, $\mathbb{R}$ and $\mathbb{N}$ denote the set of real, respectively natural numbers. We assume $0 \in \mathbb{N}$. $\mathbb{R}^n$ denotes the set of n-dimensional real valued *column* vectors and $\mathbb{R}^n_{\geq 0}$ the subset of vectors with non-negative components. We use bold letters for vectors, e.g. $\boldsymbol{x} \in \mathbb{R}^n$, where we assume that $\boldsymbol{x}$ has the components $x_1, \ldots, x_n$. Similarly, the i^{th} component of a function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ is denoted by f_i.

$\mathbb{R}^{m \times n}$ denotes the set of matrices having m rows and n columns. The transpose of a vector or matrix is indicated by the superscript $^\top$. The identity matrix of $\mathbb{R}^{n \times n}$ is denoted by Id.

The *formal Neumann series* of $A \in \mathbb{R}^{m \times m}$ is defined by $A^* = \sum_{k \in \mathbb{N}} A^k$. It is well-known that A^* exists if and only if the spectral radius of A is less than 1, i.e. $\max\{|\lambda| \mid \mathbb{C} \ni \lambda \text{ is an eigenvalue of } A\} < 1$. In the case that A^* exists, we have $A^* = (\mathrm{Id} - A)^{-1}$. The converse does not hold.

The partial order $\leq$ on $\mathbb{R}^n$ is defined as usual by setting $\boldsymbol{x} \leq \boldsymbol{y}$ if $x_i \leq y_i$ for all $1 \leq i \leq n$. Similarly, $\boldsymbol{x} < \boldsymbol{y}$ if $\boldsymbol{x} \leq \boldsymbol{y}$ and $\boldsymbol{x} \neq \boldsymbol{y}$. Finally, we write $\boldsymbol{x} \prec \boldsymbol{y}$ if $x_i < y_i$ for all $1 \leq i \leq n$, i.e., if every component of $\boldsymbol{x}$ is smaller than the corresponding component of $\boldsymbol{y}$.

We use $X_1, \ldots, X_n$ as variable identifiers and arrange them into the vector $\boldsymbol{X}$. In the following n always denotes the number of variables, i.e. the dimension of $\boldsymbol{X}$. While $\boldsymbol{x}, \boldsymbol{y}, \ldots$ denote arbitrary elements in $\mathbb{R}^n$, resp. $\mathbb{R}^n_{\geq 0}$, we write $\boldsymbol{X}$ if we want to emphasize that a function is given w.r.t. these variables. Hence, $\boldsymbol{f}(\boldsymbol{X})$ represents the function itself, whereas $\boldsymbol{f}(\boldsymbol{x})$ denotes its value for some $\boldsymbol{x} \in \mathbb{R}^n$.

If Y is a set of variables and $\boldsymbol{x}$ a vector, then by $\boldsymbol{x}_Y$ we mean the vector obtained by restricting $\boldsymbol{x}$ to the components in Y.

The *Jacobian* of a function $\boldsymbol{f}(\boldsymbol{X})$ with $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ is the matrix $\boldsymbol{f}'(\boldsymbol{X})$ defined by

$$\boldsymbol{f}'(\boldsymbol{X}) = \begin{pmatrix} \frac{\partial f_1}{\partial X_1} & \cdots & \frac{\partial f_1}{\partial X_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial X_1} & \cdots & \frac{\partial f_m}{\partial X_n} \end{pmatrix} .$$

2.2 Monotone Systems of Polynomials

Definition 1. *A function $\boldsymbol{f}(\boldsymbol{X})$ with $\boldsymbol{f} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ is a* monotone system of polynomials (MSP)*, if every component $f_i(\boldsymbol{X})$ is a polynomial in the variables $X_1, \ldots, X_n$ with coefficients in $\mathbb{R}_{\geq 0}$. We call an MSP $\boldsymbol{f}(\boldsymbol{X})$* feasible *if $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{y})$ for some $\boldsymbol{y} \in \mathbb{R}^n_{\geq 0}$.*

Fact 2. *Every MSP $\boldsymbol{f}$ is monotone on $\mathbb{R}^n_{\geq 0}$, i.e. for $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{y}$ we have $\boldsymbol{f}(\boldsymbol{x}) \leq \boldsymbol{f}(\boldsymbol{y})$.*

Since every MSP is continuous, Kleene's fixed-point theorem (see e.g. [18]) applies.

Theorem 3 (Kleene's fixed-point theorem). *Every feasible MSP $\boldsymbol{f}(\boldsymbol{X})$ has a least fixed point $\mu\boldsymbol{f}$ in $\mathbb{R}^n_{\geq 0}$ i.e., $\mu\boldsymbol{f} = \boldsymbol{f}(\mu\boldsymbol{f})$ and, in addition, $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{y})$ implies $\mu\boldsymbol{f} \leq \boldsymbol{y}$. Moreover, the sequence $(\boldsymbol{\kappa}_{\boldsymbol{f}}^{(k)})_{k \in \mathbb{N}}$ with $\boldsymbol{\kappa}_{\boldsymbol{f}}^{(k)} = \boldsymbol{f}^k(\boldsymbol{0})$ is monotonically increasing with respect to $\leq$ (i.e. $\boldsymbol{\kappa}_{\boldsymbol{f}}^{(k)} \leq \boldsymbol{\kappa}_{\boldsymbol{f}}^{(k+1)}$) and converges to $\mu\boldsymbol{f}$.*

In the following we call $(\boldsymbol{\kappa}_{\boldsymbol{f}}^{(k)})_{k \in \mathbb{N}}$ the *Kleene sequence* of $\boldsymbol{f}(\boldsymbol{X})$, and drop the subscript whenever $\boldsymbol{f}$ is clear from the context. Similarly, we sometimes write $\boldsymbol{\mu}$ instead of $\mu\boldsymbol{f}$.

A variable X_i of an MSP $\boldsymbol{f}(\boldsymbol{X})$ is *productive* if $\kappa_i^{(k)} > 0$ for some $k \in \mathbb{N}$. An MSP is *clean* if all its variables are productive. It is easy to see that we have $\kappa_i^{(k)} = 0$ for all $k \in \mathbb{N}$ if $\kappa_i^{(n)} = 0$. Just as in the case of context-free grammars we can determine all productive variables in time linear in the size of $\boldsymbol{f}$.

Notation 4. *In the following, we always assume that an MSP $\boldsymbol{f}$ is clean and feasible. I.e., whenever we write "MSP", we mean "clean and feasible MSP", unless explicitly stated otherwise.*

For the formal definition of the *Decomposed Newton Method (DNM)* (see also Section 1) we need the notion of *dependence* between variables.

Definition 5. *Let $\boldsymbol{f}(\boldsymbol{X})$ be an MSP. X_i* depends directly *on X_k, denoted by $X_i \trianglelefteq X_k$, if $\frac{\partial f_i}{\partial X_k}(\boldsymbol{X})$ is not the zero-polynomial. X_i* depends *on X_k if $X_i \trianglelefteq^* X_k$, where $\trianglelefteq^*$ is the reflexive transitive closure of $\trianglelefteq$. An MSP is* strongly connected *(short: an* scMSP*) if all its variables depend on each other.*

Any MSP can be decomposed into strongly connected components (SCCs), where an SCC S is a maximal set of variables such that each variable in S depends on each other variable in S. The following result for strongly connected MSPs was proved in [10, 12]:

Theorem 6. *Let $\boldsymbol{f}(\boldsymbol{X})$ be an scMSP and define the Newton operator $\mathcal{N}_{\boldsymbol{f}}$ as follows*

$$\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{X}) = \boldsymbol{X} + (\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{X}))^{-1}(\boldsymbol{f}(\boldsymbol{X}) - \boldsymbol{X}) \ .$$

We have:

(1) $\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{x})$ is defined for all $\mathbf{0} \le \boldsymbol{x} \prec \mu\boldsymbol{f}$ (i.e., $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x}))^{-1}$ exists). Moreover, $\boldsymbol{f}'(\boldsymbol{x})^* = \sum_{k \in \mathbb{N}} \boldsymbol{f}'(\boldsymbol{x})^k$ exists for all $\mathbf{0} \le \boldsymbol{x} \prec \mu\boldsymbol{f}$, and so $\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{X}) = \boldsymbol{X} + \boldsymbol{f}'(\boldsymbol{X})^*(\boldsymbol{f}(\boldsymbol{X}) - \boldsymbol{X})$.

(2) The Newton sequence $(\boldsymbol{\nu}_{\boldsymbol{f}}^{(k)})_{k \in \mathbb{N}}$ with $\boldsymbol{\nu}^{(k)} = \mathcal{N}_{\boldsymbol{f}}^k(\mathbf{0})$ is monotonically increasing, bounded from above by $\mu\boldsymbol{f}$ (i.e. $\boldsymbol{\nu}^{(k)} \le \boldsymbol{f}(\boldsymbol{\nu}^{(k)}) \le \boldsymbol{\nu}^{(k+1)} \prec \mu\boldsymbol{f}$), and converges to $\mu\boldsymbol{f}$.

DNM works by substituting the variables of lower SCCs by corresponding Newton approximations that were obtained earlier.

3 A Threshold for scMSPs

In this section we obtain a threshold after which DNM is guaranteed to converge linearly with rate 1.

We showed in [16] that for worst-case results on the convergence of Newton's method it is enough to consider *quadratic* MSPs, i.e., MSPs whose monomials have degree at most 2. The reason is that any MSP (resp. scMSP) $\boldsymbol{f}$ can be transformed into a quadratic MSP (resp. scMSP) $\widetilde{\boldsymbol{f}}$ by introducing auxiliary variables. This transformation is very similar to the transformation of a context-free grammar into Chomsky normal form. The transformation does not accelerate DNM, i.e., DNM on $\boldsymbol{f}$ is at least as fast (in a formal sense) as DNM on $\widetilde{\boldsymbol{f}}$, and so for a worst-case analysis, it suffices to consider quadratic systems. We refer the reader to [16] for details.

We start by defining the notion of "valid bits".

Definition 7. *Let $\boldsymbol{f}(\boldsymbol{X})$ be an MSP. A vector $\boldsymbol{\nu}$ has i* valid bits *of the least fixed point $\mu\boldsymbol{f}$ if $\left|\mu\boldsymbol{f}_j - \nu_j\right| / \left|\mu\boldsymbol{f}_j\right| \le 2^{-i}$ for every $1 \le j \le n$.*

In the rest of the section we prove the following:

Theorem 8. *Let $\boldsymbol{f}(\boldsymbol{X})$ be a quadratic scMSP. Let c_{min} be the smallest nonzero coefficient of $\boldsymbol{f}$ and let μ_{min} and μ_{max} be the minimal and maximal component of $\mu\boldsymbol{f}$, respectively. Let*

$$k_{\boldsymbol{f}} = n \cdot \log \frac{\mu_{max}}{c_{min} \cdot \mu_{min} \cdot \min\{\mu_{min}, 1\}} .$$

Then $\boldsymbol{\nu}^{(\lceil k_{\boldsymbol{f}} \rceil + i)}$ has i valid bits of $\mu\boldsymbol{f}$ for every $i \ge 0$.

Loosely speaking, the theorem states that after $k_{\boldsymbol{f}}$ iterations of Newton's method, every subsequent iteration guarantees at least one more valid bit. It may be objected that $k_{\boldsymbol{f}}$ depends on the least fixed point $\mu\boldsymbol{f}$, which is precisely what Newton's method should compute. However, in the next section we show that there are important classes of MSPs (in fact, those which motivated our investigation), for which bounds on μ_{min} can be easily obtained.

The following corollary is weaker, but less technical in that it avoids a dependence on μ_{max} and c_{min}.

Corollary 9. *Let $\boldsymbol{f}(\boldsymbol{X})$ be a quadratic scMSP of dimension n whose coefficients are given as ratios of m-bit integers. Let μ_{min} be the minimal component of $\mu\boldsymbol{f}$. Let*

$$k_{\boldsymbol{f}} = 3n^2 m + 2n^2 \left|\log \mu_{min}\right| \ .$$

Then $\boldsymbol{\nu}^{(\lceil k_{\boldsymbol{f}} \rceil + i)}$ has at least i valid bits of $\mu\boldsymbol{f}$ for every $i \ge 0$.

Corollary 9 follows from Theorem 8 by a suitable bound on μ_{max} in terms of c_{min} and μ_{min}, and by the inequation $c_{\text{min}} \geq 1/2^m$, see the appendix.

In the rest of the section we sketch the proof of Theorem 8. The proof makes crucial use of the vectors $\boldsymbol{d} \succ \boldsymbol{0}$ such that $\boldsymbol{d} \geq \boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d}$. We call a vector satisfying these two conditions a *cone vector of* $\boldsymbol{f}$ or, when $\boldsymbol{f}$ is clear from the context, just a cone vector. To a cone vector $\boldsymbol{d} = (d_1, \ldots, d_n)$ we associate two parameters, namely the maximum and the minimum of the ratios $\mu\boldsymbol{f}_1/d_1, \mu\boldsymbol{f}_2/d_2, \ldots, \mu\boldsymbol{f}_n/d_n$, which we denote by λ_{max} and λ_{min}, respectively.

In a previous paper we have shown that if $\text{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})$ is singular, then $\boldsymbol{f}$ has a cone vector $\boldsymbol{d}$ ([16], Lemmata 4 and 8). As a first step towards the proof of Theorem 8 we show the following stronger proposition.

Proposition 10. *Any scMSP has a cone vector.*

The second step consists of showing (Proposition 12) that given a cone vector $\boldsymbol{d}$, the threshold $k_{\boldsymbol{f},\boldsymbol{d}} = \log(\lambda_{\text{max}}/\lambda_{\text{min}})$ satisfies the same property as $k_{\boldsymbol{f}}$ in Theorem 8, i.e., $\boldsymbol{\nu}^{(\lceil k_{\boldsymbol{f},\boldsymbol{d}} \rceil + i)}$ has i valid bits of $\mu\boldsymbol{f}$ for every $i \geq 0$.

For that we need the following fundamental property of cone vectors: a cone vector leads to an upper bound on the error of Newton's method.

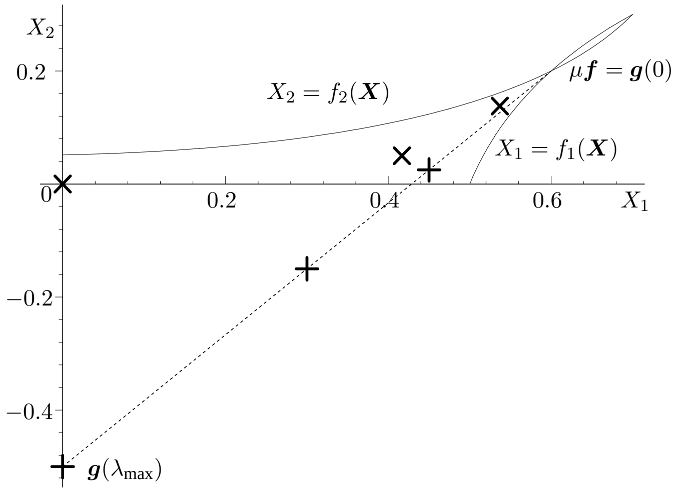
Lemma 11. *Let $\boldsymbol{d}$ be a cone vector of an MSP $\boldsymbol{f}$ and let $\lambda_{max} = \max\{\frac{\mu\boldsymbol{f}_i}{d_i}\}$. Then*

$$\mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)} \leq 2^{-k}\lambda_{max}\boldsymbol{d}.$$

Proof Idea (see Appendix A for a full proof). If we track the ray $\boldsymbol{g}(t) = \mu\boldsymbol{f} - t\boldsymbol{d}$ starting in $\mu\boldsymbol{f}$ and headed in the direction $-\boldsymbol{d}$ (the dashed line in the picture below), then $\boldsymbol{g}(\lambda_{\text{max}})$ is the intersection of $\boldsymbol{g}$ with an axis which is located farthest from $\mu\boldsymbol{f}$. One observes that the center $\boldsymbol{g}(\frac{1}{2}\lambda_{\text{max}})$ of $\boldsymbol{g}(\lambda_{\text{max}})$ and $\mu\boldsymbol{f}$ is always less than or equal to the first Newton iterate $\boldsymbol{\nu}^{(1)}$. This is the first step of the proof.

As soon as this fact is proven, we proceed by repeatedly reallocating the origin into the next Newton iterate and applying the same argument. By induction, one obtains $\boldsymbol{g}(2^{-k}\lambda_{\text{max}}) \leq \boldsymbol{\nu}^{(k)}$ for all $k \in \mathbb{N}$.

The following picture shows the Newton iterates $\boldsymbol{\nu}^{(k)}$ for $0 \leq k \leq 2$ (shape: $\times$) and the corresponding points $\boldsymbol{g}(2^{-k}\lambda_{\text{max}})$ (shape: $+$) located on the ray $\boldsymbol{g}$. Notice that $\boldsymbol{\nu}^{(k)} \geq \boldsymbol{g}(2^{-k}\lambda_{\text{max}})$. □



Now we easily obtain:

Proposition 12. *Let $\boldsymbol{f}(\boldsymbol{X})$ be an scMSP and let $\boldsymbol{d}$ be a cone vector of $\boldsymbol{f}$. Let $k_{\boldsymbol{f},\boldsymbol{d}} = \log\frac{\lambda_{max}}{\lambda_{min}}$, where $\lambda_{max} = \max_j \frac{\mu\boldsymbol{f}_j}{d_j}$ and $\lambda_{min} = \min_j \frac{\mu\boldsymbol{f}_j}{d_j}$. Then $\boldsymbol{\nu}^{(\lceil k_{\boldsymbol{f},\boldsymbol{d}} \rceil + i)}$ has at least i valid bits of $\mu\boldsymbol{f}$ for every $i \geq 0$.*

We now proceed to the third and final step. We have the problem that $k_{\boldsymbol{f},\boldsymbol{d}}$ depends on the cone vector $\boldsymbol{d}$, about which we only know that it exists (Proposition 10). We now sketch how to obtain the threshold $k_{\boldsymbol{f}}$ from Theorem 8, which is independent of any cone vectors, see Appendix A for a full proof.

Consider Proposition 12 and let $\lambda_{\text{max}} = \frac{\mu \boldsymbol{f}_i}{d_i}$ and $\lambda_{\text{min}} = \frac{\mu \boldsymbol{f}_j}{d_j}$. Then the $k_{\boldsymbol{f},\boldsymbol{d}}$ given there equals $\log\left(\frac{d_j}{d_i} \cdot \frac{\mu \boldsymbol{f}_i}{\mu \boldsymbol{f}_j}\right)$. Notice that this gets large when d_i gets small compared to d_j. By Proposition 10 the component d_i cannot be 0 if $\boldsymbol{f}$ is strongly connected. However, MSPs that are *not* strongly connected can have vectors $\boldsymbol{d} \geq \boldsymbol{0}$ with $\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} \leq \boldsymbol{d}$ s.t. some components are 0. One can make $\frac{d_j}{d_i} > 0$ arbitrarily large also for strongly connected MSPs. But when doing this, one can show that the strong connectedness "decreases" in some sense, i.e., there are variables X, Y such that X depends on Y only via a monomial that has a very small coefficient. So, $\frac{d_j}{d_i}$ can be bounded in terms of c_{min}.

4 Stochastic Models

As mentioned in the introduction, several problems concerning stochastic models can be reduced to problems about the least solution $\mu\boldsymbol{f}$ of an MSPE $\boldsymbol{f}$. In these cases, $\mu\boldsymbol{f}$ is a vector of probabilities, and so $\mu_{\text{max}} \leq 1$. Moreover, we can obtain information on μ_{min}, which leads to bounds on the threshold $k_{\boldsymbol{f}}$.

4.1 Probabilistic Pushdown Automata

Our study of MSPs was initially motivated by the verification of probabilistic pushdown automata. A *probabilistic pushdown automaton (pPDA)* is a tuple $\mathcal{P} = (Q, \Gamma, \delta, \mathit{Prob})$ where Q is a finite set of *control states*, Γ is a finite *stack alphabet*, $\delta \subseteq Q \times \Gamma \times Q \times \Gamma^*$ is a finite *transition relation* (we write $pX \hookrightarrow q\alpha$ instead of $(p, X, q, \alpha) \in \delta$), and Prob is a function which to each transition $pX \hookrightarrow q\alpha$ assigns its probability $\mathit{Prob}(pX \hookrightarrow q\alpha) \in (0, 1]$ so that for all $p \in Q$ and $X \in \Gamma$ we have $\sum_{pX \hookrightarrow q\alpha} \mathit{Prob}(pX \hookrightarrow q\alpha) = 1$. We write $pX \overset{x}{\hookrightarrow} q\alpha$ instead of $\mathit{Prob}(pX \hookrightarrow q\alpha) = x$. A *configuration* of $\mathcal{P}$ is a pair qw, where q is a control state and $w \in \Gamma^*$ is a *stack content*. A probabilistic pushdown automaton $\mathcal{P}$ naturally induces a possibly infinite Markov chain with the configurations as states and transitions given by: $pX\beta \overset{x}{\hookrightarrow} q\alpha\beta$ for every $\beta \in \Gamma^*$ iff $pX \overset{x}{\hookrightarrow} q\alpha$. We assume w.l.o.g. that if $pX \overset{x}{\hookrightarrow} q\alpha$ is a transition then $|\alpha| \leq 2$.

pPDAs and the equivalent model of recursive Markov chains have been very thoroughly studied [6, 2, 10, 8, 7, 9, 11]. This work has shown that the key to the analysis of pPDAs are the *termination probabilities* $[pXq]$, where p and q are states, and X is a stack letter, defined as follows (see e.g. [6] for a more formal definition): $[pXq]$ is the probability that, starting at the configuration pX, the pPDA eventually reaches the configuration $q\varepsilon$ (empty stack). It is not difficult to show that the vector of these probabilities is the least fixed point of the MSPE containing the equation

$$\langle pXq \rangle = \sum_{pX \overset{x}{\hookrightarrow} rYZ} x \cdot \sum_{t \in Q} \langle rYt \rangle \cdot \langle tZq \rangle \quad + \sum_{pX \overset{x}{\hookrightarrow} rY} x \cdot \langle rYq \rangle \quad + \sum_{pX \overset{x}{\hookrightarrow} q\varepsilon} x$$

for each triple (p, X, q). Call this quadratic MSPE the *termination MSPE* of the pPDA (we assume that termination MSPEs are clean, and it is easy to see that they are always feasible). We immediately have that if $\boldsymbol{f}$ is a termination MSP, then $\mu_{\text{max}} \leq 1$. We also obtain a lower bound on μ_{min}:

Lemma 13. *Let $\boldsymbol{f}$ be a termination MSP with n variables. Then $\mu_{min} \geq c_{min}^{(2^{n+1}-1)}$.*

Together with Theorem 8 we get an exponential bound for $k_{\boldsymbol{f}}$.

Proposition 14. *Let $\boldsymbol{f}$ be a strongly connected termination MSP with n variables and whose coefficients are expressed as ratios of m-bit numbers. Then $k_{\boldsymbol{f}} \leq n2^{n+2}m$.*

We conjecture that there is a lower bound on $k_{\boldsymbol{f}}$ which is exponential in n for the following reason. We know a family $(\boldsymbol{f}^{(n)})_{n=1,3,5,\ldots}$ of strongly connected MSPs with n variables and irrational coefficients such that $c_{\text{min}}^{(n)} = \frac{1}{4}$ for all n and $\mu_{\text{min}}^{(n)}$ is double-exponentially small in n. Experiments suggest that $\Theta(2^n)$ iterations are needed for the first bit of $\mu\boldsymbol{f}^{(n)}$, but we do not have a proof.

4.2 Strict pPDAs and Back-Button Processes

A pPDA is *strict* if for all $pX \in Q \times \Gamma$ and all $q \in Q$ the transition relation contains a pop-rule $pX \overset{x}{\hookrightarrow} q\epsilon$ for some $x > 0$. Essentially, strict pPDAs model programs in which every procedure has at least one terminating execution that does not call any other procedure. The termination MSP of a strict pPDA is of the form $\boldsymbol{b}(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{l}\boldsymbol{X} + \boldsymbol{c}$ for $\boldsymbol{c} \succ \boldsymbol{0}$. So we have $\mu\boldsymbol{f} \geq \boldsymbol{c}$, which implies $\mu_{\text{min}} \geq c_{\text{min}}$. Together with Theorem 8 we get:

Proposition 15. *Let $\boldsymbol{f}$ be a strongly connected termination MSP with n variables and whose coefficients are expressed as ratios of m-bit numbers. If $\boldsymbol{f}$ is derived from a strict pPDA, then $k_{\boldsymbol{f}} \leq 3nm$.*

Since in most applications m is small, we obtain an excellent convergence threshold.

In [13, 14] a class of stochastic processes is introduced to model the behavior of web-surfers which from the current webpage A can decide either to follow a link to another page, say B, with probability l_{AB}, or to press the "back button" with nonzero probability b_A. These back-button processes correspond to a very special class of pPDAs having one single control state (which in the following we omit), and rules of the form $A \overset{b_A}{\hookrightarrow} \varepsilon$ (press the back button from A) or $A \overset{l_{AB}}{\hookrightarrow} BA$ (follow the link from A to B, remembering A as destination of pressing the back button at B). The termination probabilities, called *revocation* probabilities in [13, 14], are given by the MSPE containing the equation

$$\langle A\rangle \quad = \quad b_A + \sum_{A \overset{l_{AB}}{\hookrightarrow} BA} l_{AB}\langle B\rangle\langle A\rangle \quad = \quad b_A + \langle A\rangle \sum_{A \overset{l_{AB}}{\hookrightarrow} BA} l_{AB}\langle B\rangle$$

for every webpage A. Notice that the revocation probability of a page A is the probability that, when currently visiting an instance of webpage A with $H_0H_1 \ldots H_{n-1}H_n$ the browser history of previously visited pages (H_0 being the startpage of the random user), during the further random exploration of webpages starting from A the random user eventually returns to webpage H_n with $H_0H_1 \ldots H_{n-1}$ being the remaining browser history.

4.3 An Example

As an example of application of Theorem 8 consider the following scMSPE $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$.

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 0.4X_2X_1 + 0.6 \\ 0.3X_1X_2 + 0.4X_3X_2 + 0.3 \\ 0.3X_1X_3 + 0.7 \end{pmatrix}$$

The least solution of the system gives the revocation probabilities of a back-button process with three webpages. For instance, if the surfer is at page 2 it can choose between following links to pages 1 and 3 with probabilities 0.3 and 0.4, respectively, or pressing the back button with probability 0.3.

We wish to know if any of the revocation probabilities is equal to 1. Performing 14 Newton steps (e.g. with Maple) yields an approximation $\boldsymbol{\nu}^{(14)}$ to the termination probabilities with

$$\begin{pmatrix} 0.98 \\ 0.97 \\ 0.992 \end{pmatrix} \leq \boldsymbol{\nu}^{(14)} \leq \begin{pmatrix} 0.99 \\ 0.98 \\ 0.993 \end{pmatrix} .$$

We have $c_{\text{min}} = 0.3$. In addition, since Newton's method converges to $\mu\boldsymbol{f}$ from below, we know $\mu_{\text{min}} \geq 0.97$. Moreover, $\mu_{\text{max}} \leq 1$, as $\mathbf{1} = \boldsymbol{f}(\mathbf{1})$ and so $\mu\boldsymbol{f} \leq \mathbf{1}$. Hence $k_{\boldsymbol{f}} \leq 3 \cdot \log \frac{1}{0.97 \cdot 0.3 \cdot 0.97} \leq 6$. Theorem 8 then implies that $\boldsymbol{\nu}^{(14)}$ has (at least) 8 valid bits of $\mu\boldsymbol{f}$. As $\mu\boldsymbol{f} \leq \mathbf{1}$, the absolute errors are bounded by the relative errors, and since $2^{-8} \leq 0.004$ we know:

$$\mu\boldsymbol{f} \prec \boldsymbol{\nu}^{(14)} + \begin{pmatrix} 2^{-8} \\ 2^{-8} \\ 2^{-8} \end{pmatrix} \prec \begin{pmatrix} 0.994 \\ 0.984 \\ 0.997 \end{pmatrix} \prec \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

So Theorem 8 gives us a proof that all three revocation probabilities are strictly smaller than 1.

Notice also that Newton's method converges much faster than "Kleene's iteration" $\left(\boldsymbol{f}^i(\mathbf{0})\right)_{i \in \mathbb{N}}$. We have $\boldsymbol{\kappa}^{(14)} \prec \left(0.89, 0.83, 0.96\right)^\top$, so $\boldsymbol{\kappa}^{(14)}$ has no more than 4 valid bits in any component, whereas $\boldsymbol{\nu}^{(14)}$ has more than 30 valid bits in each component.

5 Linear Convergence of the Decomposed Newton's Method

When using Newton's method for approximating the least fixed point $\mu\boldsymbol{f}$ of an scMSP $\boldsymbol{f}$, Theorem 8 states that, after $k_{\boldsymbol{f}}$ preparatory iterations of Newton's method, we have at least i bits if additional i iterations are performed. We call this linear convergence with rate 1. Now we show that DNM, which handles non-strongly-connected MSPs, converges linearly as well. We will also give an explicit convergence rate.

Let $\boldsymbol{f}(\boldsymbol{X})$ be any quadratic MSP (again we assume *quadratic* MSPs throughout this section), and let $h(\boldsymbol{f})$ denote the height of the DAG of strongly connected components (SCCs). The convergence rate of DNM crucially depends on this height: In the worst case one needs asymptotically $\Theta(2^{h(\boldsymbol{f})})$ iterations in each component per bit, assuming one performs the same number of iterations in each component.

To get a sharper result, we suggest to perform a different number of iterations in each SCC, depending on its *depth*. The depth of an SCC S is the length of the longest path in the DAG of SCCs from S to a top SCC.

In addition, we use the following notation. For a depth t, we denote by $comp(t)$ the set of SCCs of depth t. Furthermore we define $C(t) := \bigcup comp(t)$ and $C_{>}(t) := \bigcup_{t' > t} C(t')$ and, analogously, $C_{<}(t)$. We will sometimes write $\boldsymbol{v}_t$ for $\boldsymbol{v}_{C(t)}$ and $\boldsymbol{v}_{>t}$ for $\boldsymbol{v}_{C_{>}(t)}$ and $\boldsymbol{v}_{<t}$ for $\boldsymbol{v}_{C_{<}(t)}$, where $\boldsymbol{v}$ is any vector.

Figure 1 shows the Decomposed Newton's Method (DNM) for computing an approximation $\boldsymbol{\nu}$ for $\mu\boldsymbol{f}$, where $\boldsymbol{f}(\boldsymbol{X})$ is any quadratic MSP. The authors of [10] recommend to run Newton's Method in each SCC S until "approximate solutions for S are considered 'good enough' ". Here we suggest to run Newton's Method in each SCC S for a number of steps that depends (exponentially) on the depth of S and (linearly) on a parameter j that controls the precision (see Figure 1).

```
function DNM ($\boldsymbol{f}, j$)
/* The parameter $j$ controls the precision. */
for $t$ from $h(\boldsymbol{f})$ downto $0$
    forall $S \in comp(t)$   /* all SCCs $S$ of depth $t$ */
        $\boldsymbol{\nu}_S := \mathcal{N}_{\boldsymbol{f}_S}^{j \cdot 2^t}(\mathbf{0})$   /* $j \cdot 2^t$ iterations */
        /* apply $\boldsymbol{\nu}_S$ in the depending SCCs */
        $\boldsymbol{f}_{<t}(\boldsymbol{X}) := \boldsymbol{f}_{<t}(\boldsymbol{X})[\boldsymbol{X}_S / \boldsymbol{\nu}_S]$
return $\boldsymbol{\nu}$
```

Figure 1. Decomposed Newton's Method (DNM) for computing an approximation $\boldsymbol{\nu}$ of $\mu\boldsymbol{f}$ (cf. [10])

Recall that $h(\boldsymbol{f})$ was defined as the height of the DAG of SCCs. Similarly we define the width $w(\boldsymbol{f})$ to be $\max_t |comp(t)|$. Notice that $\boldsymbol{f}$ has at most $(h(\boldsymbol{f}) + 1) \cdot w(\boldsymbol{f})$ SCCs. We have the following bound on the number of iterations run by DNM.

Proposition 16. *The function* $\mathrm{DNM}(\boldsymbol{f}, j)$ *of Fig. 1 runs at most* $j \cdot w(\boldsymbol{f}) \cdot 2^{h(\boldsymbol{f})+1}$ *iterations of Newton's method.*

We will now analyze the convergence behavior of DNM asymptotically (for large j). Let $\boldsymbol{\Delta}_S^{(j)}$ denote the error in S when running DNM with parameter j, i.e., $\boldsymbol{\Delta}_S^{(j)} := \boldsymbol{\mu}_S - \boldsymbol{\nu}_S^{(j)}$. Observe that the error $\boldsymbol{\Delta}_t^{(j)}$ can be understood as the sum of two errors:

$$\boldsymbol{\Delta}_t^{(j)} = \boldsymbol{\mu}_t - \boldsymbol{\nu}_t^{(j)} = (\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}_t}^{(j)}) + (\widetilde{\boldsymbol{\mu}_t}^{(j)} - \boldsymbol{\nu}_t^{(j)}) \,,$$

where $\widetilde{\boldsymbol{\mu}_t}^{(j)} := \mu\big(\boldsymbol{f}_t(\boldsymbol{X})[\boldsymbol{X}_{>t}/\boldsymbol{\nu}_{>t}^{(j)}]\big)$, i.e., $\widetilde{\boldsymbol{\mu}_t}^{(j)}$ is the least fixed point of $\boldsymbol{f}_t$ after the approximations from the lower SCCs have been applied. So, $\boldsymbol{\Delta}_t^{(j)}$ consists of the *propagation error* $(\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}_t}^{(j)})$ and the newly inflicted *approximation error* $(\widetilde{\boldsymbol{\mu}_t}^{(j)} - \boldsymbol{\nu}_t^{(j)})$.

The following lemma, technically non-trivial to prove, gives a bound on the propagation error.

Lemma 17 (Propagation error)**.** *There is a constant* $c > 0$ *such that*

$$\|\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}_t}\| \le c \cdot \sqrt{\|\boldsymbol{\mu}_{>t} - \boldsymbol{\nu}_{>t}\|}$$

holds for all $\boldsymbol{\nu}_{>t}$ *with* $\mathbf{0} \le \boldsymbol{\nu}_{>t} \le \boldsymbol{\mu}_{>t}$*, where* $\widetilde{\boldsymbol{\mu}_t} = \mu\big(\boldsymbol{f}_t(\boldsymbol{X})[\boldsymbol{X}_{>t}/\boldsymbol{\nu}_{>t}]\big)$*.*

Intuitively, Lemma 17 states that if $\boldsymbol{\nu}_{>t}$ has k valid bits of $\boldsymbol{\mu}_{>t}$, then $\widetilde{\boldsymbol{\mu}_t}$ has roughly $k/2$ valid bits of $\boldsymbol{\mu}_t$. In other words, (at most) one half of the valid bits are lost on each level of the DAG due to the propagation error.

The following theorem assures that after combining the propagation error and the approximation error, DNM still converges linearly.

Theorem 18. *Let* $\boldsymbol{f}$ *be a quadratic MSP. Let* $\boldsymbol{\nu}^{(j)}$ *denote the result of calling* $\mathrm{DNM}(\boldsymbol{f}, j)$ *(see Figure 1). Then there is a* $k_{\boldsymbol{f}} \in \mathbb{N}$ *such that* $\boldsymbol{\nu}^{(k_{\boldsymbol{f}}+i)}$ *has at least* i *valid bits of* $\mu\boldsymbol{f}$ *for every* $i \ge 0$*.*

We conclude that increasing i by one gives us asymptotically at least one additional bit in each component and, by Proposition 16, costs $w(\boldsymbol{f}) \cdot 2^{h(\boldsymbol{f})+1}$ additional Newton iterations.

The bound above is essentially optimal in the sense that an exponential (in $h(\boldsymbol{f})$) number of iterations is in general needed to obtain an additional bit. To see this consider the following example that we also used in a previous paper [16] for a slightly different purpose.

$$\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X}) = \begin{pmatrix} \frac{1}{4}X_0^2 + \frac{1}{2}X_0X_1 + \frac{1}{4}X_1^2 \\ \vdots \\ \frac{1}{4}X_{h-1}^2 + \frac{1}{2}X_{h-1}X_h + \frac{1}{4}X_h^2 \\ \frac{1}{2} + \frac{1}{2}X_h^2 \end{pmatrix} \tag{1}$$

The only solution of $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ is $(1, \ldots, 1)^\top$. Notice that $\boldsymbol{f}$ has $h+1$ SCCs: $\{X_0\}, \ldots, \{X_h\}$, where $\{X_j\}$ has depth j. The DNM starts at the bottom SCC $\{X_h\}$ and works its way up to $\{X_0\}$. It is easy to show (see [16]) that the propagation error on level t is at least $1 \cdot \sqrt{\Delta_{t+1}^{(j)}}$. The approximation error on level h equals $\Delta_h^{(j)} = 2^{-j \cdot 2^h}$ if DNM is called with parameter j. So, by induction we get $\Delta_t^{(j)} \ge 2^{-j \cdot 2^t}$ for $0 \le t \le h$. We conclude that at least 2^h Newton steps per bit are needed. Notice that this holds even when we approximate only the bottom SCC $\frac{1}{2} + \frac{1}{2}X_h^2$ with Newton's method and solve the other SCCs exactly. Therefore, any method that approximates (1) suffers from an "exponential" amplification of the propagation error. In other words, this example shows that computing $\mu\boldsymbol{f}$ is in general an *ill-conditioned* problem.

6 Newton's Method for General MSPs

Etessami and Yannakakis [10] introduced DNM because they could show that the matrix inverses used by Newton's method exist if Newton's method is run on each SCC separately (see Theorem 6).

In this section we show, maybe surprisingly, that the matrix inverses used by Newton's method exist even if the MSP is *not* decomposed. More precisely we show the following theorem.

Theorem 19. *Let $\boldsymbol{f}(\boldsymbol{X})$ be any MSP, not necessarily strongly connected. Let the Newton operator $\mathcal{N}_{\boldsymbol{f}}$ be defined as before:*

$$\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{X}) = \boldsymbol{X} + (\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{X}))^{-1}(\boldsymbol{f}(\boldsymbol{X}) - \boldsymbol{X})$$

Then the Newton sequence $(\boldsymbol{\nu}_{\boldsymbol{f}}^{(k)})_{k\in\mathbb{N}}$ with $\boldsymbol{\nu}^{(k)} = \mathcal{N}_{\boldsymbol{f}}^{k}(\boldsymbol{0})$ is well-defined (i.e., the matrix inverses exist), monotonically increasing, bounded from above by $\mu\boldsymbol{f}$ (i.e. $\boldsymbol{\nu}^{(k)} \leq \boldsymbol{\nu}^{(k+1)} \prec \mu\boldsymbol{f}$), and converges to $\mu\boldsymbol{f}$.

Theorem 19 relies on a generalized Newton's method for solving fixed point equations over commutative ω-continuous semirings, introduced in [5]. The semiring over the nonnegative reals, $\mathcal{S}_{\mathbb{R}_{\geq 0}} = \langle \mathbb{R}_{\geq 0} \cup \{\infty\}, +, \cdot, 0, 1\rangle$, is such a semiring. In $\mathcal{S}_{\mathbb{R}_{\geq 0}}$ the operations $+$ and $\cdot$ are defined as in the reals with straightforward extensions to ∞, in particular $0 \cdot \infty = 0$ and $a \cdot \infty = \infty$ if $a > 0$. If an infinite sum does not converge to a real number, it is defined to be ∞. So, for a matrix $A \in \mathbb{R}_{\geq 0}^{m\times m}$, the formal Neumann series A^* is always defined in $\mathcal{S}_{\mathbb{R}_{\geq 0}}$, some entries of A^* may be ∞.

A theorem of [5] applied to the semiring $\mathcal{S}_{\mathbb{R}_{\geq 0}}$ yields the following proposition.

Proposition 20 (follows from [5], Theorem 3)**.** *Let $\boldsymbol{f}(\boldsymbol{X})$ be an MSP. Let the Newton operator $\widehat{\mathcal{N}}_{\boldsymbol{f}}$ be defined as follows:*

$$\widehat{\mathcal{N}}_{\boldsymbol{f}}(\boldsymbol{X}) = \boldsymbol{X} + \boldsymbol{f}'(\boldsymbol{X})^*(\boldsymbol{f}(\boldsymbol{X}) - \boldsymbol{X})\,,$$

where $\boldsymbol{f}'(\boldsymbol{X})^*$ is computed in $\mathcal{S}_{\mathbb{R}_{\geq 0}}$ (i.e., may have ∞ as an entry). Then the Newton sequence $(\widehat{\boldsymbol{\nu}}_{\boldsymbol{f}}^{(k)})_{k\in\mathbb{N}}$ with $\widehat{\boldsymbol{\nu}}^{(k)} = \widehat{\mathcal{N}}_{\boldsymbol{f}}^{k}(\boldsymbol{0})$ is monotonically increasing, bounded from above by $\mu\boldsymbol{f}$ and converges to $\mu\boldsymbol{f}$.

Additive inverses are, strictly speaking, not defined. Therefore, in Proposition 20, instead of $\boldsymbol{f}(\boldsymbol{X}) - \boldsymbol{X}$ we should rather write "a vector $\boldsymbol{\delta}(\boldsymbol{X})$ such that $\boldsymbol{X} + \boldsymbol{\delta}(\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{X})$". But the simpler formulation causes no trouble here, because $\boldsymbol{f}(\widehat{\boldsymbol{\nu}}^{(k)}) - \widehat{\boldsymbol{\nu}}^{(k)}$ is positive in each component [5].

In order to show that the Newton scheme from Theorem 19 coincides with the Newton scheme from Proposition 20 we need to show that $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^* = (\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(k)}))^{-1}$. It is sufficient to show that $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*$ does not have ∞ entries, because then clearly $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(k)})) = \mathrm{Id}$. Notice that this is not a trivial consequence of Proposition 20: it could be that $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k)})^*$ has ∞ entries, but the $\widehat{\boldsymbol{\nu}}^{(k)}$ and $\mu\boldsymbol{f}$ do not because the ∞ entries of $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k)})^*$ are cancelled out by matching 0 entries of $\boldsymbol{f}(\widehat{\boldsymbol{\nu}}^{(k)}) - \widehat{\boldsymbol{\nu}}^{(k)}$. What remains to show for Theorem 19 is that this is not the case and $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k)})^*$ has no ∞ entries. The rest of the proof of Theorem 19 can be found in Appendix D.1.

6.1 Convergence Speed

As we now know that Newton's method converges to $\mu\boldsymbol{f}$ for any MSP $\boldsymbol{f}$, we address again the question of convergence *speed*. By exploiting Theorem 18 and Theorem 19 one can show:

Theorem 21. *Let $\boldsymbol{f}$ be any quadratic MSP. Then the Newton sequence $(\boldsymbol{\nu}^{(k)})_{k\in\mathbb{N}}$ is well-defined and converges linearly to $\mu\boldsymbol{f}$. More precisely, there is a $k_{\boldsymbol{f}} \in \mathbb{N}$ such that $\boldsymbol{\nu}^{(k_{\boldsymbol{f}}+i\cdot(h(\boldsymbol{f})+1)\cdot 2^{h(\boldsymbol{f})})}$ has at least i valid bits of $\mu\boldsymbol{f}$ for every $i \geq 0$.*

A proof is given in Appendix D.2. Again, the exponential factor in $2^{h(\boldsymbol{f})}$ cannot be avoided in general. This follows from the example and the discussion at the end of Section 5.

7 Conclusions

We have proved a threshold $k_{\boldsymbol{f}}$ for strongly connected MSPEs. After $k_{\boldsymbol{f}} + i$ steps of DNM we have i bits of accuracy. The threshold $k_{\boldsymbol{f}}$ depends on the representation size of $\boldsymbol{f}$ and on the least solution $\mu\boldsymbol{f}$. Although this latter dependence might seem to be a problem, lower and upper bounds on $\mu\boldsymbol{f}$ can be easily derived for stochastic models (probabilistic programs with procedures, stochastic context-free grammars and back-button processes). In particular, this allows us to show that $k_{\boldsymbol{f}}$ depends linearly on the representation size for back-button processes. We have also shown by means of an example that the threshold $k_{\boldsymbol{f}}$ improves when the number of iterations of DNM increases.

In [16] we left the problem whether DNM converges linearly for non-strongly-connected MSPEs open. We have proven that this is the case, although the convergence rate is poorer: if h and w are the height and width of the graph of SCCs of $\boldsymbol{f}$, then there is a threshold $\widetilde{k}_{\boldsymbol{f}}$ such that $\widetilde{k}_{\boldsymbol{f}} + i \cdot w \cdot 2^{h+1}$ iterations of DNM compute at least i valid bits of $\mu\boldsymbol{f}$. We have also given an example in which DNM needs at least $i \cdot 2^h$ iterations for i valid bits.

Finally, we have shown that the Jacobian of the whole MSPE is guaranteed to exist, whether the MSPE is strongly connected or not.

Acknowledgment. The authors wish to thank Kousha Etessami for very valuable comments.

References

[1] L. Blum, M. Shub, and S. Smale. On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines. *Bulletin of the Amer. Math. Society*, 21(1):1–46, 1989.

[2] T. Brázdil, A. Kučera, and O. Stražovský. On the decidability of temporal properties of probabilistic pushdown automata. In *Proceedings of STACS'2005*, volume 3404 of *LNCS*, pages 145–157. Springer, 2005.

[3] R. Dowell and S. Eddy. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinformatics*, 5(71), 2004.

[4] R. Durbin, S. Eddy, A. Krogh, and G. Michison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[5] J. Esparza, S. Kiefer, and M. Luttenberger. On fixed point equations over commutative semirings. In *Proceedings of STACS*, LNCS 4397, pages 296–307, 2007.

[6] J. Esparza, A. Kučera, and R. Mayr. Model-checking probabilistic pushdown automata. In *Proceedings of LICS 2004*, pages 12–21, 2004.

[7] J. Esparza, A. Kučera, and R. Mayr. Quantitative analysis of probabilistic pushdown automata: Expectations and variances. In *Proceedings of LICS 2005*, pages 117–126. IEEE Computer Society Press, 2005.

[8] K. Etessami and M. Yannakakis. Algorithmic verification of recursive probabilistic systems. In *Proceedings of TACAS 2005*, LNCS 3440, pages 253–270. Springer, 2005.

[9] K. Etessami and M. Yannakakis. Checking LTL properties of recursive Markov chains. In *Proceedings of 2nd Int. Conf. on Quantitative Evaluation of Systems (QEST'05)*, 2005.

[10] K. Etessami and M. Yannakakis. Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations. In *STACS*, pages 340–352, 2005.

[11] K. Etessami and M. Yannakakis. Recursive Markov decision processes and recursive stochastic games. In *Proceedings of ICALP 2005*, volume 3580 of *LNCS*, pages 891–903. Springer, 2005.

[12] K. Etessami and M. Yannakakis. Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations, 2006. Draft journal submission, `http://homepages.inf.ed.ac.uk/kousha/bib_index.html`.

[13] R. Fagin, A. Karlin, J. Kleinberg, P. Raghavan, S. Rajagopalan, R. Rubinfeld, M. Sudan, and A. Tomkins. Random walks with "back buttons" (extended abstract). In *STOC*, pages 484–493, 2000.

[14] R. Fagin, A. Karlin, J. Kleinberg, P. Raghavan, S. Rajagopalan, R. Rubinfeld, M. Sudan, and A. Tomkins. Random walks with "back buttons". *Annals of Applied Probability*, 11(3):810–862, 2001.

[15] S. Geman and M. Johnson. Probabilistic grammars and their applications, 2002.

[16] S. Kiefer, M. Luttenberger, and J. Esparza. On the convergence of Newton's method for monotone systems of polynomial equations. In *Proceedings of STOC*, pages 217–226. ACM, 2007.

[17] B. Knudsen and J. Hein. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Research*, 31(13):3423–3428, 2003.

[18] W. Kuich. *Handbook of Formal Languages*, volume 1, chapter 9: Semirings and Formal Power Series: Their Relevance to Formal Languages and Automata, pages 609 – 677. Springer, 1997.

[19] C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[20] J. Ortega and W. Rheinboldt. *Iterative solution of nonlinear equations in several variables*. Academic Press, 1970.

[21] Y. Sakabikara, M. Brown, R. Hughey, I. Mian, K. Sjolander, R. Underwood, and D. Haussler. Stochastic context-free grammars for tRNA. *Nucleic Acids Research*, 22:5112–5120, 1994.

A Proofs of Section 3

A.1 Proof of Proposition 10

Here is a restatement of Proposition 10.

Proposition 10. *Any scMSP has a cone vector.*

Let $\boldsymbol{f}$ be an scMSP. As mentioned before, it was proved in [16] that $\boldsymbol{f}$ has a cone vector if $\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})$ is singular.

So, it remains to show that $\boldsymbol{f}$ has a cone vector $\boldsymbol{d}$ in the non-singular case, too. In a first step we show that we can relax the requirement $\boldsymbol{d} \succ \mathbf{0}$ to $\boldsymbol{d} > \mathbf{0}$.

Lemma 22. *Let $\boldsymbol{f}$ be an scMSP and let $\boldsymbol{d} > \mathbf{0}$ with $\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} \leq \boldsymbol{d}$. Then $\boldsymbol{d}$ is a cone vector, i.e., $\boldsymbol{d} \succ \mathbf{0}$.*

Proof. Since $\boldsymbol{f}$ is an MSP, every component of $\boldsymbol{f}'(\mu\boldsymbol{f})$ is nonnegative. So,

$$\mathbf{0} \leq \boldsymbol{f}'(\mu\boldsymbol{f})^n\boldsymbol{d} \leq \boldsymbol{f}'(\mu\boldsymbol{f})^{n-1}\boldsymbol{d} \leq \ldots \leq \boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} \leq \boldsymbol{d}.$$

Let w.l.o.g. $d_1 > 0$. As $\boldsymbol{f}$ is strongly connected, there is for all j with $1 \leq j \leq n$ an $r_j \leq n$ s.t. $(\boldsymbol{f}'(\mu\boldsymbol{f})^{r_j})_{j1} > 0$. Hence, $(\boldsymbol{f}'(\mu\boldsymbol{f})^{r_j}\boldsymbol{d})_j > 0$ for all j. With above inequation chain, it follows that $d_j \geq (\boldsymbol{f}'(\mu\boldsymbol{f})^{r_j}\boldsymbol{d})_j > 0$. So, $\boldsymbol{d} \succ \mathbf{0}$. □

Now we can show that there is a cone vector also in the non-singular case.

Lemma 23. *Let $\boldsymbol{f}$ be an scMSP. If $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))^{-1}$ exists, then $\boldsymbol{f}$ has a cone vector.*

Proof. By Lemma 22 it suffices to find a vector $\boldsymbol{d} > \mathbf{0}$ such that $\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} \leq \boldsymbol{d}$. Take any $\boldsymbol{e} \succ \mathbf{0}$ and set $\boldsymbol{d} := (\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))^{-1}\boldsymbol{e}$. Clearly $\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} \leq \boldsymbol{d}$, and so it remains to show $\boldsymbol{d} > \mathbf{0}$. Since $\boldsymbol{e} \succ \mathbf{0}$, it suffices to prove that every entry of $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))^{-1}$ is nonnegative, which we denote by $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))^{-1} \geq \mathbf{0}$. For this recall that $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x}))^{-1} = \boldsymbol{f}'(\boldsymbol{x})^*$ on $G = \{\boldsymbol{x} \mid \mathbf{0} \leq \boldsymbol{x} \prec \mu\boldsymbol{f}\}$, and, hence, $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x}))^{-1} \geq \mathbf{0}$ for $\boldsymbol{x} \in G$.

We now make use of the fact that for every $i, j \in \{1, \ldots, n\}$ we may write $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{X}))^{-1}_{ij}$ as a rational function $r_{ij}(\boldsymbol{X}) = \frac{n_{ij}(\boldsymbol{X})}{d(\boldsymbol{X})}$, where $d(\boldsymbol{X})$ is the determinant of $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{X}))$, and $n_{ij}(\boldsymbol{X})$ is obtained by (1) canceling the j^{th} row and i^{th} column of $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{X}))$, (2) taking the determinant of the resulting submatrix, and (3) multiplying by $(-1)^{i+j}$. As $d(\mu\boldsymbol{f}) \neq 0$, the functions $r_{ij}(\boldsymbol{X})$ are continuous at least on an open ball O centered at $\mu\boldsymbol{f}$. Now, as $G \cap O \neq \emptyset$, we have $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x}))^{-1} \geq \mathbf{0}$ for every $\boldsymbol{x} \in G$, and the r_{ij} are continuous, we get $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))^{-1} \geq \mathbf{0}$. □

This finishes the proof of Proposition 10.

A.2 Proof of Lemma 11

Here is a restatement of Lemma 11.

Lemma 11. *Let $\boldsymbol{d}$ be a cone vector of an MSP $\boldsymbol{f}$ and let $\lambda_{max} = \max\{\frac{\mu\boldsymbol{f}_i}{d_i}\}$. Then*

$$\mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)} \leq 2^{-k}\lambda_{max}\boldsymbol{d}.$$

We handle the base case $k = 1$ of Lemma 11 in the following separate lemma.

Lemma 24. *Let $\boldsymbol{d}$ be a cone vector of a (not necessarily clean) MSP $\boldsymbol{f}$. Let $\lambda_{max} = \max\{\frac{\mu\boldsymbol{f}_i}{d_i}\}$. Then*

$$\mu\boldsymbol{f} - \boldsymbol{\nu}^{(1)} \leq \frac{1}{2}\lambda_{max}\boldsymbol{d}.$$

Proof. We write $\boldsymbol{f}(\boldsymbol{X})$ as a sum

$$\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{c} + \sum_{k=1}^{D} L_k(\boldsymbol{X}, \ldots, \boldsymbol{X})\boldsymbol{X}$$

where D is the degree of $\boldsymbol{f}$, and every L_k is a $(k-1)$-linear map from $(\mathbb{R}^n)^{k-1}$ to $\mathbb{R}^{n\times n}$. Notice that $\boldsymbol{f}'(\boldsymbol{X}) = \sum_{k=1}^{D} k \cdot L_k(\boldsymbol{X}, \ldots, \boldsymbol{X})$. We simply write L for L_1, and $\boldsymbol{h}(\boldsymbol{X})$ for $\boldsymbol{f}(\boldsymbol{X}) - L\boldsymbol{X} - \boldsymbol{c}$.

$$\begin{aligned}
& \tfrac{\lambda_{\max}}{2}\boldsymbol{d} & & \\
= \; & \tfrac{\lambda_{\max}}{2}(L^*\boldsymbol{d} - L^*L\boldsymbol{d}) & & (L^* = \mathrm{Id} + L^*L) \\
\geq \; & \tfrac{\lambda_{\max}}{2}(L^*\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} - L^*L\boldsymbol{d}) & & (\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} \leq \boldsymbol{d}) \\
= \; & \tfrac{\lambda_{\max}}{2}L^*\boldsymbol{h}'(\mu\boldsymbol{f})\boldsymbol{d} & & (\boldsymbol{f}'(\boldsymbol{x}) = \boldsymbol{h}'(\boldsymbol{x}) + L) \\
= \; & L^*\tfrac{1}{2}\boldsymbol{h}'(\mu\boldsymbol{f})\lambda_{\max}\boldsymbol{d} & & \\
\geq \; & L^*\tfrac{1}{2}\boldsymbol{h}'(\mu\boldsymbol{f})\mu\boldsymbol{f} & & \text{(by def. of } \lambda_{\max}\text{: } \lambda_{\max}\boldsymbol{d} \geq \mu\boldsymbol{f}) \\
= \; & L^*\tfrac{1}{2}\textstyle\sum_{k=2}^{D} k \cdot L_k(\mu\boldsymbol{f}, \ldots, \mu\boldsymbol{f})\mu\boldsymbol{f} & & \\
\geq \; & L^*\textstyle\sum_{k=2}^{D} L_k(\mu\boldsymbol{f}, \ldots, \mu\boldsymbol{f})\mu\boldsymbol{f} & & \\
= \; & L^*\boldsymbol{h}(\mu\boldsymbol{f}) & & \\
= \; & L^*(\boldsymbol{f}(\mu\boldsymbol{f}) - L\mu\boldsymbol{f} - \boldsymbol{c}) & & (\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}) + L\boldsymbol{x} + \boldsymbol{c}) \\
= \; & L^*\mu\boldsymbol{f} - L^*L\mu\boldsymbol{f} - L^*\boldsymbol{c} & & (\boldsymbol{f}(\mu\boldsymbol{f}) = \mu\boldsymbol{f}) \\
= \; & \mu\boldsymbol{f} - L^*\boldsymbol{c} & & (L^* = \mathrm{Id} + L^*L) \\
= \; & \mu\boldsymbol{f} - \boldsymbol{\nu}^{(1)} & & (\boldsymbol{\nu}^{(1)} = L^*\boldsymbol{c})
\end{aligned}$$

□

By means of a suitable induction we can extend this last Lemma 24 to an arbitrary number of iterations, yielding the proof of Lemma 11:

Proof. For every $k \geq 0$, define $\boldsymbol{g}_k(\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{X} + \boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}$. We first show that $\boldsymbol{g}_k$ is an MSP (not necessarily clean) for every $k \geq 0$. The only coefficients of $\boldsymbol{g}_k$ that could be negative are those of degree 0. But we have $\boldsymbol{g}_k(\boldsymbol{0}) = \boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)} \geq \boldsymbol{0}$, and so these coefficients are also nonnegative.

Moreover, it follows immediately from the definition that $\mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)} \geq \boldsymbol{0}$ is the least fixed point of $\boldsymbol{g}_k$. Finally, $\boldsymbol{g}_k$ satisfies $\boldsymbol{g}_k'(\mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)})\boldsymbol{d} \leq \boldsymbol{d}$, and so $\boldsymbol{d}$ is also a cone vector of $\boldsymbol{g}_k$.

Let $\lambda_k = \max\{\frac{\mu\boldsymbol{f}_i - \boldsymbol{\nu}_i^{(k)}}{d_i}\}$; in particular, $\lambda_0 = \lambda_{\max}$. We proceed to prove the lemma by induction on k. For $k = 0$ we have by definition $\boldsymbol{\nu}^{(0)} = \boldsymbol{0}$ and $\mu\boldsymbol{f} \leq \lambda_{\max}\boldsymbol{d}$, and we are done. Now let $k \geq 0$ and assume $\mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)} \leq 2^{-k}\lambda_0\boldsymbol{d}$. We have

$$\begin{aligned}
\lambda_k &= \max\{\tfrac{\mu\boldsymbol{f}_i - \boldsymbol{\nu}_i^{(k)}}{d_i}\} \\
&\leq \max\{\tfrac{2^{-k}\lambda_0 d_i}{d_i}\} \quad . \\
&= 2^{-k}\lambda_0.
\end{aligned}$$

Since $\boldsymbol{d}$ is a cone vector of $\boldsymbol{g}_k$, we can apply Lemma 24 to $\boldsymbol{g}_k$ and get:

$$(\mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}_{\boldsymbol{g}_k}^{(1)} \leq \frac{1}{2}\lambda_k\boldsymbol{d}.$$

where $\boldsymbol{\nu}^{(1)}_{\boldsymbol{g}_k}$ denotes the first iteration of Newton's method applied to $\boldsymbol{g}_k$. But we have

$$\begin{aligned}\boldsymbol{\nu}^{(1)}_{\boldsymbol{g}_k} &= \boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*(\boldsymbol{g}_k(\mathbf{0}) - \mathbf{0}) \\ &= \boldsymbol{\nu}^{(k+1)} - \boldsymbol{\nu}^{(k)}\end{aligned}$$

and so

$$\mu\boldsymbol{f} - \boldsymbol{\nu}^{(k+1)} \le \frac{1}{2}\lambda_k\boldsymbol{d} \le 2^{-(k+1)}\lambda_0\boldsymbol{d}. \qquad \square$$

A.3 Proof of Proposition 12

Here is a restatement of Proposition 12.

Proposition 12. *Let $\boldsymbol{f}(\boldsymbol{X})$ be an scMSP and let $\boldsymbol{d}$ be a cone vector of $\boldsymbol{f}$. Let $k_{\boldsymbol{f},\boldsymbol{d}} = \log\frac{\lambda_{max}}{\lambda_{min}}$, where $\lambda_{max} = \max_j \frac{\mu\boldsymbol{f}_j}{d_j}$ and $\lambda_{min} = \min_j \frac{\mu\boldsymbol{f}_j}{d_j}$. Then $\boldsymbol{\nu}^{(\lceil k_{\boldsymbol{f},\boldsymbol{d}}\rceil + i)}$ has at least i valid bits of $\mu\boldsymbol{f}$ for every $i \ge 0$.*

Proof. For all $1 \le j \le n$ the following holds.

$$\begin{aligned}\left(\mu\boldsymbol{f} - \boldsymbol{\nu}^{(\lceil k_{\boldsymbol{f},\boldsymbol{d}}\rceil+i)}\right)_j &\le 2^{-(\lceil k_{\boldsymbol{f},\boldsymbol{d}}\rceil+i)}\lambda_{\text{max}}d_j \\ &\le 2^{-k_{\boldsymbol{f},\boldsymbol{d}}-i}\lambda_{\text{max}}d_j \\ &= \lambda_{\text{min}}d_j \cdot 2^{-i} \\ &\le \mu\boldsymbol{f}_j \cdot 2^{-i} \quad \square\end{aligned}$$

A.4 Proof of Theorem 8

Here is a restatement of Theorem 8.

Theorem 8. *Let $\boldsymbol{f}(\boldsymbol{X})$ be a quadratic scMSP. Let c_{min} be the smallest nonzero coefficient of $\boldsymbol{f}$ and let μ_{min} and μ_{max} be the minimal and maximal component of $\mu\boldsymbol{f}$, respectively. Let*

$$k_{\boldsymbol{f}} = n \cdot \log\frac{\mu_{max}}{c_{min}\cdot\mu_{min}\cdot\min\{\mu_{min},1\}}.$$

Then $\boldsymbol{\nu}^{(\lceil k_{\boldsymbol{f}}\rceil+i)}$ has i valid bits of $\mu\boldsymbol{f}$ for every $i \ge 0$.

Proof. In what follows we shorten $\mu\boldsymbol{f}$ to $\boldsymbol{\mu}$. Let $\boldsymbol{d}$ be a cone vector of $\boldsymbol{f}$ (which exists by Proposition 10). Let $\lambda_j = \frac{\mu_j}{d_j}$ for all $1 \le j \le n$ and assume w.l.o.g. $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n$. By Lemma 11 we have $\boldsymbol{\nu}^{(k)} \ge \boldsymbol{\mu} - 2^{-k}\lambda_1\boldsymbol{d}$. Let $k_j = \log\frac{\lambda_1}{\lambda_j}$. Then we have

$$\nu_j^{(\lceil k_j\rceil+i)} \ge \mu_j - \left(\frac{1}{2}\right)^{k_j+i}\lambda_1 d_j = \mu_j - 2^{-i}\mu_j$$

and $\frac{\mu_j - \nu_j^{(\lceil k_j\rceil+i)}}{\mu_j} \le 2^{-i}$. So it remains to show $k_n \le k_{\boldsymbol{f}}$.

We claim the existence of indices s, t with $1 \le s, t \le n$ such that $f'_{st}(\boldsymbol{\mu}) \ne 0$ and $\log\frac{\lambda_s}{\lambda_t} \ge \frac{1}{n}k_n$. To prove that such s, t exist, we use the fact that $\boldsymbol{f}$ is strongly connected, i.e., that there is a sequence $1 = r_1, r_2, \ldots, r_q = n$ with $q \le n$ and $f'_{r_jr_{j+1}}(\boldsymbol{x}) \ne 0$. Since $\boldsymbol{\mu} \succ \mathbf{0}$, we also have $f'_{r_jr_{j+1}}(\boldsymbol{\mu}) \ne 0$. It follows

$$\frac{\lambda_1}{\lambda_n} = \frac{\lambda_{r_1}}{\lambda_{r_2}}\cdots\frac{\lambda_{r_{q-1}}}{\lambda_{r_q}} \text{ , and so}$$

$$\log\frac{\lambda_1}{\lambda_n} = \log\frac{\lambda_{r_1}}{\lambda_{r_2}} + \cdots + \log\frac{\lambda_{r_{q-1}}}{\lambda_{r_q}}.$$

So there must exist a j s.t.

$$\log \frac{\lambda_{r_j}}{\lambda_{r_{j+1}}} \geq \frac{1}{n-1} \log \frac{\lambda_1}{\lambda_n} \geq \frac{1}{n} k_n$$

and one can choose $s = r_j$ and $t = r_{j+1}$.

Now, $f'_{st}(\boldsymbol{\mu})\boldsymbol{d} \leq \boldsymbol{d}$ implies $f'_{st}(\boldsymbol{\mu}) d_t \leq d_s$, and so

$$f'_{st}(\boldsymbol{\mu}) \leq \frac{d_s}{d_t} = \frac{\lambda_t}{\lambda_s} \cdot \frac{\mu_s}{\mu_t} \leq 2^{-k_n/n} \cdot \frac{\mu_{\text{max}}}{\mu_{\text{min}}} \ . \tag{2}$$

On the other hand, since $\boldsymbol{f}$ is quadratic, $\boldsymbol{f}'$ is a linear mapping such that

$$f'_{st}(\boldsymbol{\mu}) = 2(b_1 \cdot \mu_1 + \cdots + b_n \cdot \mu_n) + l$$

where $b_1, \ldots, b_n$ and l are coefficients of quadratic, respectively linear, monomials of $\boldsymbol{f}$. As $f'_{st}(\boldsymbol{\mu}) \neq 0$, at least one of these coefficients must be nonzero and so greater than or equal to c_{min}. It follows

$$f'_{st}(\boldsymbol{\mu}) \geq c_{\text{min}} \cdot \min\{\mu_{\text{min}}, 1\} \ ,$$

which together with equation (2) yields

$$2^{k_n/n} \leq \frac{\mu_{\text{max}}}{c_{\text{min}} \cdot \mu_{\text{min}} \cdot \min\{\mu_{\text{min}}, 1\}} \quad \text{, and so}$$
$$k_n \leq n \cdot \log \frac{\mu_{\text{max}}}{c_{\text{min}} \cdot \mu_{\text{min}} \cdot \min\{\mu_{\text{min}}, 1\}} \ . \qquad \square$$

A.5 Proof of Corollary 9

Here is a restatement of Corollary 9.

Corollary 9. *Let $\boldsymbol{f}(\boldsymbol{X})$ be a quadratic scMSP of dimension n whose coefficients are given as ratios of m-bit integers. Let μ_{min} be the minimal component of $\mu\boldsymbol{f}$. Let*

$$k_{\boldsymbol{f}} = 3n^2 m + 2n^2 \left|\log \mu_{min}\right| \ .$$

Then $\boldsymbol{\nu}^{(\lceil k_{\boldsymbol{f}} \rceil + i)}$ has at least i valid bits of $\mu\boldsymbol{f}$ for every $i \geq 0$.

First we check the case where $\boldsymbol{f}$ is linear, i.e., all monomials in $\boldsymbol{f}$ have degree at most 1. In this case, Newton's method reaches $\mu\boldsymbol{f}$ after one iteration, so the theorem holds. Consequently, we can assume in the following that $\boldsymbol{f}$ is *strictly quadratic*, meaning that $\boldsymbol{f}$ is quadratic and there is a polynomial in $\boldsymbol{f}$ of degree 2.

In the following lemma we give a bound on μ_{max} in terms of μ_{min}, c_{min} and n. Notice that $\log \mu_{\text{max}}$ is polynomial in terms of those parameters.

Lemma 25. *Let the preconditions of Theorem 8 hold, and let $\boldsymbol{f}$ be strictly quadratic, i.e., nonlinear. Then*

$$\mu_{max} \leq \frac{1}{c_{min}^{3n-2} \cdot \min(\mu_{min}^{2n-2}, 1)} \ .$$

Furthermore, $c_{min} \leq 1$.

Proof. Let w.l.o.g. $\mu_{\text{max}} = (\mu\boldsymbol{f})_1$. The proof is based on the idea that X_1 indirectly depends quadratically on itself. More precisely, by the strong connectedness, X_1 depends (indirectly) on some variable, say X_{i_r}, such that $\boldsymbol{f}_{i_r}$ contains a degree-2-monomial. The variables in that monomial, in turn, depend on X_1. This gives an inequation of the form $(\mu\boldsymbol{f})_1 \geq C \cdot {(\mu\boldsymbol{f})_1}^2$, implying $(\mu\boldsymbol{f})_1 \leq 1/C$.

We give the details in the following. Using the strong connectedness there exists a sequence of variables $X_{i_1}, \dots, X_{i_r}$ and a sequence of monomials $m_{i_1}, \dots, m_{i_r}$ ($1 \leq r \leq n$) with the following properties:

- $X_{i_1} = X_1$,
- m_{i_u} is a monomial appearing in $\boldsymbol{f}_{i_u}$ $\quad (1 \leq u \leq r)$,
- $m_{i_u} = c_{i_u} \cdot X_{i_{u+1}}$ $\quad (1 \leq u \leq r)$,
- $m_{i_r} = c_{i_r} \cdot X_{j_1} \cdot X_{k_1}$ for some variables X_{j_1}, X_{k_1}.

Notice that

$$\begin{aligned} \mu_{\text{max}} = (\mu\boldsymbol{f})_1 &\geq c_{i_1} \cdot \ldots \cdot c_{i_r} \cdot (\mu\boldsymbol{f})_{j_1} \cdot (\mu\boldsymbol{f})_{k_1} \\ &\geq \min(c_{\text{min}}^n, 1) \cdot (\mu\boldsymbol{f})_{j_1} \cdot (\mu\boldsymbol{f})_{k_1} . \end{aligned} \tag{3}$$

Again by strong connectedness, there exists a sequence of variables $X_{j_1}, \dots, X_{j_s}$ and a sequence of monomials $m_{j_1}, \dots, m_{j_{s-1}}$ ($1 \leq s \leq n$) with the following properties:

- $X_{j_s} = X_1$,
- m_{j_u} is a monomial appearing in $\boldsymbol{f}_{j_u}$ $\quad (1 \leq u \leq s-1)$,
- $m_{j_u} = c_{j_u} \cdot X_{j_{u+1}}$ or $m_{j_u} = c_{j_u} \cdot X_{j_{u+1}} \cdot X_{j'_{u+1}}$ for some variable $X_{j'_{u+1}}$ $\quad (1 \leq u \leq s-1)$.

Notice that

$$\begin{aligned} (\mu\boldsymbol{f})_{j_1} &\geq c_{j_1} \cdot \ldots \cdot c_{j_{s-1}} \cdot \min(\mu_{\text{min}}^{s-1}, 1) \cdot (\mu\boldsymbol{f})_1 \\ &\geq \min(c_{\text{min}}^{n-1}, 1) \cdot \min(\mu_{\text{min}}^{n-1}, 1) \cdot (\mu\boldsymbol{f})_1 . \end{aligned} \tag{4}$$

Similarly, there exists a sequence of variables $X_{k_1}, \dots, X_{k_t}$ ($1 \leq t \leq n$) with $X_{k_t} = X_1$ showing

$$(\mu\boldsymbol{f})_{k_1} \geq \min(c_{\text{min}}^{n-1}, 1) \cdot \min(\mu_{\text{min}}^{n-1}, 1) \cdot (\mu\boldsymbol{f})_1 . \tag{5}$$

Combining (3) with (4) and (5) yields

$$\mu_{\text{max}} \geq \min(c_{\text{min}}^{3n-2}, 1) \cdot \min(\mu_{\text{min}}^{2n-2}, 1) \cdot \mu_{\text{max}}^2 ,$$

which implies

$$\mu_{\text{max}} \leq \frac{1}{\min(c_{\text{min}}^{3n-2}, 1) \cdot \min(\mu_{\text{min}}^{2n-2}, 1)} . \tag{6}$$

Now the second statement of the lemma implies the first one. In order to prove the second statement, assume for contradiction $c_{\text{min}} > 1$. This implies $\mu_{\text{min}} > 1$ due to the following reason. Consider the Kleene sequence $\mathbf{0}, \boldsymbol{f}(\mathbf{0}), \boldsymbol{f}^2(\mathbf{0}), \dots$ For all $1 \leq i \leq n$ let b_i be the smallest natural number such that $\big(\boldsymbol{f}^{b_i}(\mathbf{0})\big)_i > 0$. The numbers b_i exist because $\boldsymbol{f}$ is clean and the Kleene sequence converges to $\mu\boldsymbol{f}$. We show by induction on b_i that $\big(\boldsymbol{f}^{b_i}(\mathbf{0})\big)_i > 1$ which, by the monotonicity of the Kleene sequence, implies $\mu_{\text{min}} > 1$. For the inductive step notice that the value $\big(\boldsymbol{f}^{b_i}(\mathbf{0})\big)_i = \boldsymbol{f}_i(\boldsymbol{f}^{b_i-1}(\mathbf{0}))$ is computed as a sum of products of numbers which are either coefficients of $\boldsymbol{f}$ (and hence by assumption greater than 1) or of the form $\big(\boldsymbol{f}^{b_i-1}(\mathbf{0})\big)_j$ for some j. By induction and by the monotonicity of the Kleene sequence, a number of the latter form is either 0 or greater than 1. So, $\big(\boldsymbol{f}^{b_i}(\mathbf{0})\big)_i$ itself must be 0 or greater than 1. By definition of b_i it cannot be 0.

So we have $c_{\text{min}} > 1$ and $\mu_{\text{min}} > 1$. Plugging this into (6) yields $\mu_{\text{max}} \leq 1$. This implies $\mu_{\text{max}} < \mu_{\text{min}}$, contradicting the definition of μ_{max} and μ_{min}. □

Now we can complete the proof of Corollary 9. By Theorem 8 it suffices to show

$$n \cdot \log \frac{\mu_{\text{max}}}{c_{\text{min}} \cdot \mu_{\text{min}} \cdot \min\{\mu_{\text{min}}, 1\}} \leq 3n^2 m + 2n^2 \left|\log \mu_{\text{min}}\right| \ .$$

We have

$$\begin{aligned} & n \cdot \log \frac{\mu_{\text{max}}}{c_{\text{min}} \cdot \mu_{\text{min}} \cdot \min\{\mu_{\text{min}}, 1\}} \\ \leq\ & n \cdot \log \frac{1}{c_{\text{min}}^{3n-1} \cdot \mu_{\text{min}} \cdot \min(\mu_{\text{min}}^{2n-2}, 1)} && \text{(by Lemma 25)} \\ \leq\ & 3n^2 \cdot (-\log c_{\text{min}}) - n \log(\mu_{\text{min}} \cdot \min(\mu_{\text{min}}^{2n-2}, 1)) && \text{(by Lemma 25: } c_{\text{min}} \leq 1\text{)} \\ \leq\ & 3n^2 m - n \log(\mu_{\text{min}} \cdot \min(\mu_{\text{min}}^{2n-2}, 1)) && (c_{\text{min}} \geq 2^{-m})\ . \end{aligned}$$

If $\mu_{\text{min}} \geq 1$ we have $-n \log(\mu_{\text{min}} \cdot \min(\mu_{\text{min}}^{2n-2}, 1)) = -n \log \mu_{\text{min}} \leq 0$, so we are done in this case. If $\mu_{\text{min}} \leq 1$ we have $-n \log(\mu_{\text{min}} \cdot \min(\mu_{\text{min}}^{2n-2}, 1)) = -n \log \mu_{\text{min}}^{2n-1} = n(2n-1) \left|\log \mu_{\text{min}}\right| \leq 2n^2 \left|\log \mu_{\text{min}}\right|$. This completes the proof of Corollary 9.

B Proofs of Section 4

B.1 Proof of Lemma 13

Here is a restatement of Lemma 13.

Lemma 13. *Let $\boldsymbol{f}$ be a termination MSP with n variables. Then $\mu_{min} \geq c_{min}^{(2^{n+1}-1)}$.*

Proof. We prove a stronger result. For every $k \in \{1, \ldots, n\}$, $\boldsymbol{f}$ has k variables $X_1, \ldots, X_k$ such that $\mu \boldsymbol{f}_1, \ldots, \mu \boldsymbol{f}_k \geq c_{\text{min}}^{2^{k+1}-1}$.

We proceed by induction on k. For $k = 1$, observe that, since the MSP is clean, $pX \overset{x}{\hookrightarrow} q\varepsilon$ is a transition of the pPDA for some $\langle pXq \rangle$, and so $[pXq] \geq x \geq c_{\text{min}}$. We call $\langle pXq \rangle$ a sink.

For $k > 1$, let $X_1, \ldots, X_{k-1}$ be variables such that $\mu \boldsymbol{f}_1, \ldots, \mu \boldsymbol{f}_{k-1} \geq c_{\text{min}}^{2^k-1}$. We show that there is a variable X_k such that $\mu \boldsymbol{f}_k \geq c_{\text{min}}^{2^{k+1}-1}$. Let $\widehat{\boldsymbol{f}}$ be the MSP obtained by replacing every occurrence of X_i by $\mu \boldsymbol{f}_i$ for every $i \in \{1, \ldots, k-1\}$; it is easy to see that $\widehat{\boldsymbol{f}}$ is also a termination MSP with $\mu \widehat{\boldsymbol{f}}_k = \mu \boldsymbol{f}_k$ and $\widehat{c}_{\text{min}} \geq c_{\text{min}} (c_{\text{min}}^{2^k-1})^2 = c_{\text{min}}^{2^{k+1}-1}$. So we can choose any sink of $\widehat{\boldsymbol{f}}$ for X_k. □

B.2 Proof of Proposition 14

Here is a restatement of Proposition 14.

Proposition 14. *Let $\boldsymbol{f}$ be a strongly connected termination MSP with n variables and whose coefficients are expressed as ratios of m-bit numbers. Then $k_{\boldsymbol{f}} \leq n 2^{n+2} m$.*

Proof.

$$\begin{aligned} k_{\boldsymbol{f}} &= n \log \tfrac{\mu_{\text{max}}}{c_{\text{min}} \cdot \mu_{\text{min}} \cdot \min\{\mu_{\text{min}}, 1\}} && \text{(Theorem 8)} \\ &\leq -n \log(\mu_{\text{min}}^2 \cdot c_{\text{min}}) && \text{(termination MSP)} \\ &\leq -n \log(c_{\text{min}}^{2 \cdot (2^{n+1}-1)} \cdot c_{\text{min}}) && \text{(Lemma 13)} \\ &\leq -n 2^{n+2} \log c_{\text{min}} \\ &\leq n 2^{n+2} m && (c_{\text{min}} \geq 1/2^m) \quad \square \end{aligned}$$

C Proofs of Section 5

C.1 Proof of Proposition 16

Here is a restatement of Proposition 16.

Proposition 16. *The function* DNM *of Figure 1 runs at most* $j \cdot w(\boldsymbol{f}) \cdot 2^{h(\boldsymbol{f})+1}$ *iterations of Newton's method.*

Proof. The number of iterations of the DNM is $\sum_{t=0}^{h(\boldsymbol{f})} |comp(t)| \cdot j \cdot 2^t$. This can be bounded as follows.

$$\begin{array}{ll} \sum_{t=0}^{h(\boldsymbol{f})} |comp(t)| \cdot j \cdot 2^t & \leq w(\boldsymbol{f}) \cdot j \cdot \sum_{t=0}^{h(\boldsymbol{f})} 2^t \\ & \leq w(\boldsymbol{f}) \cdot j \cdot 2^{h(\boldsymbol{f})+1} \quad \square \end{array}$$

C.2 Proof of Lemma 17

Before proving Lemma 17, we show the following proposition which covers the case of a quadratic, clean, and feasible scMSP

$$\boldsymbol{f}(\boldsymbol{X}) = b(\boldsymbol{X}, \boldsymbol{X}) + l(\boldsymbol{X}) + c,$$

where $b(\boldsymbol{X}, \boldsymbol{Y})$ is a bilinear map, $l(\boldsymbol{X})$ is linear, and c is constant.

Proposition 26. *Let* $\boldsymbol{f}(\boldsymbol{X})$ *satisfy the conditions stated above with* $\mu\boldsymbol{f}$ *its least fixed-point. Then there is a constant* C *such that for all* $0 \leq \boldsymbol{\delta} \leq \mu\boldsymbol{f}$ *it holds*

$$C \cdot \|\boldsymbol{\delta}\|_2^2 \leq \left\|(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})) \cdot \boldsymbol{\delta} + b(\boldsymbol{\delta}, \boldsymbol{\delta})\right\|_2 .$$

Proof. We now discuss the three cases that either (case I) $\boldsymbol{f}$ is linear in $\boldsymbol{X}$, or (case II) $\boldsymbol{f}$ is non-linear and $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))^{-1}$ exists, or (case III) $\boldsymbol{f}$ is non-linear and $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))$ is singular.

Case I: We first consider the case, where the SCC represented by $\boldsymbol{f}$ is linear in $\boldsymbol{X}$. Then $\boldsymbol{f}'(\boldsymbol{X}) \equiv \boldsymbol{f}'(\mathbf{0})$ is constant, $b(\boldsymbol{X}, \boldsymbol{X}) \equiv \mathbf{0}$ and $(\mathrm{Id} - \boldsymbol{f}')^{-1}$ exists, as we consider an SCC. So, we get

$$\left\|(\mathrm{Id} - \boldsymbol{f}'(\mathbf{0}))\boldsymbol{\delta}\right\|_2 \geq \lambda_{\min}(\mathrm{Id} - \boldsymbol{f}'(\mathbf{0})) \cdot \|\boldsymbol{\delta}\|_2 ,$$

where we define $\lambda_{\min}(A)$ to be the smallest absolute value of an eigenvalue of a square matrix A.

Case II: Next, we consider the case where $\boldsymbol{f}$ contains at least one quadratic term in $\boldsymbol{X}$ and $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))$ is invertible. As shown in the proof of Lemma 23, we then have $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))^{-1} \geq \mathbf{0}$. So, we may write

$$\begin{array}{rcl} \left\|(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))\boldsymbol{\delta} + b(\boldsymbol{\delta}, \boldsymbol{\delta})\right\|_2 & \geq & \lambda_{\min}(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})) \cdot \left\|\boldsymbol{\delta} + (\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})^{-1} b(\boldsymbol{\delta}, \boldsymbol{\delta})\right\|_2 \\ & \geq & \lambda_{\min}(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})) \cdot \|\boldsymbol{\delta}\|_2 , \end{array}$$

where we used in the last step that $\boldsymbol{\delta} \geq \mathbf{0}$ and $(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f}))b(\boldsymbol{\delta}, \boldsymbol{\delta})^{-1} \geq \mathbf{0}$.

Case III: Finally, we consider the case where $\boldsymbol{f}$ depends quadratically on $\boldsymbol{X}$ and $\mathrm{Id}-\boldsymbol{f}'(\mu\boldsymbol{f})$ is singular. As $\boldsymbol{f}(\boldsymbol{X})$ is clean and feasible, i.e. $\mu\boldsymbol{f}\succ\boldsymbol{0}$, and quadratically depends on $\boldsymbol{X}$, we know that the first Newton step is well-defined, i.e. $\boldsymbol{f}'(\boldsymbol{0})^*=(\mathrm{Id}-\boldsymbol{f}'(\boldsymbol{0}))^{-1}$ exists. We therefore may write

$$\begin{aligned}\left\|(\mathrm{Id}-\boldsymbol{f}'(\mu\boldsymbol{f}))\boldsymbol{\delta}+b(\boldsymbol{\delta},\boldsymbol{\delta})\right\|_2 &\geq \lambda_{\min}(\mathrm{Id}-\boldsymbol{f}'(\boldsymbol{0}))\left\|\boldsymbol{f}'(\boldsymbol{0})^*((\mathrm{Id}-f'(\mu\boldsymbol{f}))\boldsymbol{\delta}+b(\boldsymbol{\delta},\boldsymbol{\delta}))\right\|_2\\ &= \lambda_{\min}(\mathrm{Id}-\boldsymbol{f}'(\boldsymbol{0}))\left\|(\mathrm{Id}-2\tilde{b}(\mu\boldsymbol{f}))\boldsymbol{\delta}+\tilde{b}(\boldsymbol{\delta},\boldsymbol{\delta}))\right\|_2,\end{aligned}$$

where $\tilde{b}(\boldsymbol{X},\boldsymbol{X}):=\boldsymbol{f}'(\boldsymbol{0})^*b(\boldsymbol{X},\boldsymbol{X})$. Thus, it is sufficient to show that there exists a $\tilde{C}>0$ with

$$\left\|(\mathrm{Id}-2\tilde{b}(\mu))\boldsymbol{\delta}+\tilde{b}(\boldsymbol{\delta},\boldsymbol{\delta}))\right\|_2 \geq \tilde{C}\left\|\boldsymbol{\delta}\right\|_2, \tag{7}$$

as we then may set $C:=\lambda_{\min}(\mathrm{Id}-\boldsymbol{f}'(\boldsymbol{0}))\cdot\tilde{C}$.

We note that $\boldsymbol{f}(\boldsymbol{X})$ and $\tilde{\boldsymbol{f}}(\boldsymbol{X}):=\tilde{b}(\boldsymbol{X},\boldsymbol{X})+\boldsymbol{f}'(\boldsymbol{0})^*\boldsymbol{f}(\boldsymbol{0})$ are equivalent in the sense that both functions have the same set of fixed points, their Newton sequences and nullspaces of $\mathrm{Id}-\boldsymbol{f}'(\boldsymbol{x}^*)$ and $\mathrm{Id}-\tilde{\boldsymbol{f}}'(\boldsymbol{x}^*)$ are identical. These properties are easily checked.

The reason for multiplying by $\boldsymbol{f}'(\boldsymbol{0})^*$ is that this guarantees that no component of $\tilde{b}(\boldsymbol{X},\boldsymbol{X})$ is the zero-polynomial. We are going to need this property shortly.

First let us give an intuition why this property of $\tilde{b}$ holds – we leave the technical details to the reader: as we assume that S is an SCC, every variable of $\boldsymbol{X}$ depends on every other variable of $\boldsymbol{X}$ w.r.t. $\boldsymbol{f}$. Hence, as we consider the case where $\boldsymbol{f}$ contains at least one quadratic term, every variable either directly depends directly on a quadratic term, or there exists a sequence of variables $X_{i_1},X_{i_2},\ldots,X_{i_k}$ such that X_{i_l} depends linearly on $X_{i_{l+1}}$ and X_{i_k} itself depends directly on a quadratic term. All these "linear dependencies" are summarized in $\boldsymbol{f}'(\boldsymbol{0})^*$. Multiplying by $\boldsymbol{f}'(\boldsymbol{0})$ propagates these to the remaining quadratic terms.

Let us introduce the norm

$$\left\|\boldsymbol{y}\right\|_{\mu\boldsymbol{f}}:=\max\{\left|\frac{y_i}{\mu f_i}\right|\}.$$

Remember, we consider a clean scMSP, thus we have $\mu\boldsymbol{f}\succ\boldsymbol{0}$, and $\left\|\cdot\right\|_{\mu\boldsymbol{f}}$ is well-defined. It is straightforward to check that this is indeed a norm. We then define the set of directions

$$D=\{\boldsymbol{d}\in\mathbb{R}^n\mid\boldsymbol{d}\geq\boldsymbol{0},\left\|\boldsymbol{d}\right\|_{\mu\boldsymbol{f}}=1\}.$$

Then we are guaranteed that for every *direction* $\boldsymbol{d}\in D$ the ray $\mu\boldsymbol{f}-r\cdot\boldsymbol{d}$ stays non-negative for $r\in[0,1]$, i.e.

$$\boldsymbol{0}\leq\mu\boldsymbol{f}-r\cdot\boldsymbol{d}\leq\mu\boldsymbol{f}\quad(r\in[0,1]),$$

and, as $\boldsymbol{0}<\boldsymbol{\delta}\leq\mu\boldsymbol{f}$, we have $r_{\boldsymbol{\delta}}^{-1}\cdot\boldsymbol{\delta}\in D$ for $r_{\boldsymbol{\delta}}:=\left\|\boldsymbol{\delta}\right\|_{\mu\boldsymbol{f}}\in[0,1]$.

We now will show that that there exists a $\tilde{C}>0$ – independent of $\boldsymbol{d}$ – such that

$$\left\|r(\mathrm{Id}-\tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\boldsymbol{d}+r^2\tilde{b}(\boldsymbol{d},\boldsymbol{d})\right\|\geq\tilde{C}\cdot r^2$$

for all $r\in[0,1]$ and $\boldsymbol{d}\in D$, which implies Eq. 7.

We set

$$U(r,\boldsymbol{d}):=\frac{\left\|r(\mathrm{Id}-\tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\boldsymbol{d}+r^2\tilde{b}(\boldsymbol{d},\boldsymbol{d})\right\|_2^2}{r^4}=\left\|\tilde{b}(\boldsymbol{d},\boldsymbol{d})\right\|_2^2+\frac{2}{r}\langle\tilde{b}(\boldsymbol{d},\boldsymbol{d}),(\mathrm{Id}-\tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\boldsymbol{d}\rangle+\frac{1}{r^2}\left\|(\mathrm{Id}-\tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\boldsymbol{d}\right\|_2^2,$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean scalar-product. We further define

$$\alpha(\boldsymbol{d}):=\left\|\tilde{b}(\boldsymbol{d},\boldsymbol{d})\right\|_2^2,\quad\beta(\boldsymbol{d}):=\langle\tilde{b}(\boldsymbol{d},\boldsymbol{d}),(\mathrm{Id}-\tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\boldsymbol{d}\rangle,\ \text{and}\ \gamma(\boldsymbol{d}):=\left\|(\mathrm{Id}-\tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\boldsymbol{d}\right\|_2^2.$$

Note that $U(r,\boldsymbol{d}) > 0$ on $(0,1] \times D$, as otherwise $\mu\boldsymbol{f} - r\boldsymbol{d}$ would be a fixed-point of $\boldsymbol{f}(\boldsymbol{X})$, resp. $\tilde{\boldsymbol{f}}(\boldsymbol{X})$, less than $\mu\boldsymbol{f}$.

As D is compact and $U(r,d)$ is continuous on $(0,1] \times D$, the function

$$g(r) := \inf_{\boldsymbol{d}\in D} U(r,\boldsymbol{d})$$

is non-negative and continuous on $(0,1]$, too. Set

$$G(R) := \inf_{R\leq r\leq 1} g(r)$$

for $0 \leq R \leq 1$. We then have $G(R) > 0$ for $0 < R \leq 1$ and G decreases monotonically with $R \to 0$.

Now, if we can show that $G(0) = \inf_{0\leq r\leq 1} g(r) > 0$, our proof will be complete, as we will then have

$$\left\|(\mathrm{Id} - \tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\boldsymbol{\delta} + \tilde{b}(\boldsymbol{\delta},\boldsymbol{\delta})\right\|_2 \geq G(0)\cdot r_{\boldsymbol{\delta}}^2 = G(0)\cdot \|\boldsymbol{\delta}\|_{\mu\boldsymbol{f}}^2 \geq \tilde{C}\,\|\boldsymbol{\delta}\|_2^2$$

by equivalence of norms on $\mathbb{R}^n$ for some appropriate constant $\tilde{C}$.

We proceed by assuming the opposite, i.e. $G(0) = \inf_{r\in[0,1]} g(r) = 0$, and show that this leads to the contradiction that $\mu\boldsymbol{f}$ is not the least fixed-point. With the assumption $\inf_{r\in[0,1]} g(r) = 0$, there has to exist a monotonically decreasing sequence r_i converging to 0 with $g(r_i) \to 0$ for $i \to \infty$.

As $U(r,\boldsymbol{d})$ is continuous, and D compact, we find for every r_i a $\boldsymbol{d}_i \in D$ with $g(r_i) = U(r_i,\boldsymbol{d}_i)$. As $\boldsymbol{d}_i$ is a sequence in the compact set $D \subseteq \mathbb{R}^n$, there exists a convergent subsequence, w.l.o.g. we therefore may assume that the sequence $\boldsymbol{d}_i$ already converges to some $\boldsymbol{d}^* \in D$.

Now, we want to show first that we can refine the sequence $(r_i,\boldsymbol{d}_i)_{i\in\mathbb{N}}$ in such a way that there is a $C_\gamma > 0$ such that $\gamma(\boldsymbol{d}_i) \leq C_\gamma r_i^2$ for all i: By the Cauchy-Schwarz inequation we have $|\beta(\cdot)| \leq \sqrt{\alpha(\cdot)\gamma(\cdot)}$, thus

$$0 \xleftarrow{i\to\infty} g(r_i) \geq \left(\sqrt{\alpha(\boldsymbol{d}_i)} - \frac{\sqrt{\gamma(\boldsymbol{d}_i)}}{r_i}\right)^2 \geq 0.$$

Hence, there has to exist constants $c_\gamma \geq 0$ and $i_\gamma \in \mathbb{N}$ such that for all $i \geq i_\gamma$:

$$\left|\sqrt{\alpha(\boldsymbol{d}_i)} - \frac{\sqrt{\gamma(\boldsymbol{d}_i)}}{r_i}\right| \leq c_\gamma,$$

which in turn implies

$$\frac{\sqrt{\gamma(\boldsymbol{d}_i)}}{r_i} \leq c_\gamma + \alpha(\boldsymbol{d}_i) \leq c_\gamma + \max_{\boldsymbol{d}\in D}\sqrt{\alpha(\boldsymbol{d})} =: C_\gamma \quad \text{ i.e. } \quad \gamma(\boldsymbol{d}_i) \leq C_\gamma^2 \cdot r_i^2.$$

Thus, we have to have $\gamma(\boldsymbol{d}^*) = 0$, i.e. $\boldsymbol{d}^*$ is located in the nullspace of $\mathrm{Id} - f'(\mu\boldsymbol{f})$, implying $\beta(\boldsymbol{d}^*) = 0$ and $\alpha(\boldsymbol{d}^*) \geq c > 0$. As $\boldsymbol{d}^* \in D$, we have $\boldsymbol{d}^* > 0$ – see Lemma 22. Thus by the strong connectivity of $f(\boldsymbol{X})$ we have $\boldsymbol{d}^* \succ \boldsymbol{0}$. Hence, there has to exist an i_0 such that for all $i \geq i_0$, we have $\boldsymbol{d}_i \succ \boldsymbol{0}$, as $\boldsymbol{d}_i$ converges to $\boldsymbol{d}^*$ for all $i \geq i_0$, as $\boldsymbol{d}_i \to \boldsymbol{d}$ and $\tilde{b}$ continuous [2].

We have already stated that $\tilde{b}(\boldsymbol{X},\boldsymbol{X})$ cannot be the zero-polynomial in any component, hence, $\tilde{b}(\boldsymbol{d}_i,\boldsymbol{d}_i) \succ \boldsymbol{0}$ for all $i \geq i_0$, too. So, there also exists a vector $\boldsymbol{c}_{\tilde{b}}$ such that $\tilde{b}(\boldsymbol{d}_i,\boldsymbol{d}_i) \succ \boldsymbol{c}_{\tilde{b}} \succ \boldsymbol{0}$.

Now, as we have

$$g(r_i) = \left\|(\mathrm{Id} - \tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\frac{\boldsymbol{d}_i}{r_i} + \tilde{b}(\boldsymbol{d}_i,\boldsymbol{d}_i)\right\|_2^2 \xrightarrow{i\to\infty} 0,$$

[2] If necessary, we may adjust i_0 suitably.

there has to exist an I_0 such that $\tilde{b}(\boldsymbol{d}_i, \boldsymbol{d}_i) \succ \boldsymbol{c}_{\tilde{b}}$, and $g(r_i) < (\frac{1}{2}\min_{1\leq k\leq n}(\boldsymbol{c}_{\tilde{b}})_k)^2$ for all $i \geq I_0$. This implies $(\mathrm{Id} - \tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\frac{\boldsymbol{d}_i}{r_i} \prec \boldsymbol{0}$.

Consider the vector $\widetilde{\boldsymbol{d}} := \boldsymbol{d}_{I_0}$. As $r_i > 0$, we also have $(\mathrm{Id} - \tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\widetilde{\boldsymbol{d}} \prec \boldsymbol{0}$.

Define

$$\rho := \min\{1, \min_{1\leq k\leq n} \frac{-((\mathrm{Id} - \tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\widetilde{\boldsymbol{d}})_k}{(\tilde{b}(\widetilde{\boldsymbol{d}}, \widetilde{\boldsymbol{d}}))_k}\}.$$

As $(\mathrm{Id} - \tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\widetilde{\boldsymbol{d}} \prec \boldsymbol{0}$ and $\tilde{b}(\widetilde{\boldsymbol{d}}, \widetilde{\boldsymbol{d}}) \succ \boldsymbol{0}$, we have $\rho > 0$. Now for all $0 < r < \rho \leq 1$

$$\begin{aligned} & \tilde{\boldsymbol{f}}(\mu\boldsymbol{f} - r\widetilde{\boldsymbol{d}}) - (\mu\boldsymbol{f} - r\widetilde{\boldsymbol{d}}) \\ = \; & r(\mathrm{Id} - \tilde{\boldsymbol{f}}'(\mu\boldsymbol{f}))\widetilde{\boldsymbol{d}} + r^2\tilde{b}(\widetilde{\boldsymbol{d}}, \widetilde{\boldsymbol{d}}) \\ \prec \; & \boldsymbol{0}. \end{aligned}$$

This means $\tilde{\boldsymbol{f}}(\mu\boldsymbol{f} - r\widetilde{\boldsymbol{d}}) \leq \mu\boldsymbol{f} - r\widetilde{\boldsymbol{d}}$ for $0 < r < \rho$. But $\mu\boldsymbol{f}$ is the least solution of $\tilde{\boldsymbol{f}}(\boldsymbol{X}) \leq \boldsymbol{X}$ over $\mathbb{R}^n_{\geq 0}$ by virtue of the Knaster-Tarski theorem. So we get the desired contradiction. □

Now we can prove Lemma 17, restated here.

Lemma 17. *There is a constant $c > 0$ such that*

$$\|\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}_t}\| \leq c \cdot \sqrt{\|\boldsymbol{\mu}_{>t} - \boldsymbol{\nu}_{>t}\|}$$

holds for all $\boldsymbol{\nu}_{>t}$ with $\boldsymbol{0} \leq \boldsymbol{\nu}_{>t} \leq \boldsymbol{\mu}_{>t}$, where $\widetilde{\boldsymbol{\mu}_t} = \mu\big(\boldsymbol{f}_t(\boldsymbol{X})[\boldsymbol{X}_{>t}/\boldsymbol{\nu}_{>t}]\big)$.

Proof. Let S be an SCC at level t, i.e. $S \in \mathrm{comp}(t)$. S itself does not need to depend on all variables $\boldsymbol{X}_{>t}$. Thus, let $\mathrm{dep}(S)$ be the set of variables on which S really depends on – excluding the variables corresponding to S, i.e. $\boldsymbol{X}_S$. We may then write the MSP $\boldsymbol{f}_S$ – corresponding to S – as $\boldsymbol{f}_S(\boldsymbol{X}_S, \boldsymbol{X}_{\mathrm{dep}(S)})$.

Now, let $\boldsymbol{\mu}_{\mathrm{dep}(S)}$ be the correct (non-negative) least fixed point of $\boldsymbol{f}_{\mathrm{dep}(S)}(\boldsymbol{X}_{\mathrm{dep}(S)})$, and $\boldsymbol{\nu}_{\mathrm{dep}(S)}$ the part of the approximation $\boldsymbol{\nu}_{>t}$ relevant to S. Let

$$\boldsymbol{\varepsilon}_{\mathrm{dep}(S)} = \boldsymbol{\mu}_{\mathrm{dep}(S)} - \boldsymbol{\nu}_{\mathrm{dep}(S)}$$

be the absolute error in the underlying SCCs relevant to S. The propagation error is then

$$\boldsymbol{\delta}_S := \mu\boldsymbol{f}_S(\boldsymbol{X}_s, \boldsymbol{\mu}_{\mathrm{dep}(S)}) - \mu\boldsymbol{f}_S(\boldsymbol{X}_S, \boldsymbol{\nu}_{\mathrm{dep}(S)}).$$

What we are going to show is that there is a $C_S > 0$ such that

$$\|\boldsymbol{\delta}_S\| \leq C_S\sqrt{\left\|\boldsymbol{\varepsilon}_{\mathrm{dep}(S)}\right\|}.$$

Note that this is sufficient to prove the lemma as

$$\|\boldsymbol{\mu}_t - \tilde{\boldsymbol{\mu}}_t\|_\infty = \max_{S\in\mathrm{comp}(t)} \|\boldsymbol{\delta}_S\|_\infty \leq \max_{S\in\mathrm{comp}(t)} (C_S \cdot \sqrt{\left\|\boldsymbol{\varepsilon}_{\mathrm{dep}(S)}\right\|_\infty}) \leq (\max_{S\in\mathrm{comp}(t)} C_S) \cdot \sqrt{\|\boldsymbol{\mu}_{>t} - \boldsymbol{\nu}_{>t}\|_\infty}$$

Because of the equivalence of norms on any vector space of finite dimension over $\mathbb{R}$, we are guaranteed the existence of an appropriate constant such that this holds in any other norm, too.

In the following, we therefore consider a fixed SCC S, and simplify the notation by setting:

- $\boldsymbol{X} := \boldsymbol{X}_S$ – the variables corresponding to the SCC S,
- $\boldsymbol{Y} := \boldsymbol{X}_{\mathrm{dep}(S)}$ – the variables corresponding to the SCCs $\mathrm{dep}(S)$ on which S depends,

- $F(\boldsymbol{X},\boldsymbol{Y}) := \boldsymbol{f}_S(\boldsymbol{X},\boldsymbol{Y})$ – the restriction of the given system $\boldsymbol{f}$ to the SCC S,
- $\boldsymbol{y}^* := \boldsymbol{\mu}_{\text{dep}(S)}$ – the restriction of $\boldsymbol{\mu} = \boldsymbol{\mu f}$ to the variables $\boldsymbol{Y}$,
- $\boldsymbol{x}^* := \mu \boldsymbol{f}_S(\boldsymbol{X}_s, \boldsymbol{\mu}_{\text{dep}(S)})$ – the least fixed point of $F(\boldsymbol{X}, \boldsymbol{\mu}_{\text{dep}(S)})$.
- $\boldsymbol{\varepsilon} := \boldsymbol{\varepsilon}_{\text{dep}(S)}$ – the restriction of the approximation error $\boldsymbol{\varepsilon} = \boldsymbol{\mu} - \boldsymbol{\nu}$, and
- $\boldsymbol{\delta} := \boldsymbol{\delta}_S$ – the error $\boldsymbol{x}^* - \mu\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\nu}_{\text{dep}(S)})$ introduced by replacing $\boldsymbol{y}^*$ by $\boldsymbol{\nu}_{\text{dep}(S)}$.

So we consider the following parameterized MSP

$$F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n : \begin{pmatrix}\boldsymbol{x}\\\boldsymbol{y}\end{pmatrix} \mapsto F(\boldsymbol{x},\boldsymbol{y}) \text{ with } F(\boldsymbol{x},\boldsymbol{y}) = B(\begin{pmatrix}\boldsymbol{x}\\\boldsymbol{y}\end{pmatrix}, \begin{pmatrix}\boldsymbol{x}\\\boldsymbol{y}\end{pmatrix}) + L\begin{pmatrix}\boldsymbol{x}\\\boldsymbol{y}\end{pmatrix} + c,$$

where B is a bilinear map, L is a matrix, and c is a vector.

As we assume that the whole MSP is “clean” and feasible, we have $\boldsymbol{x}^* \succ \mathbf{0}$ and $\boldsymbol{y}^* \succ \mathbf{0}$.

We require some suitable approximation $\boldsymbol{\nu}_{\text{dep}(S)} = \boldsymbol{y}^* - \boldsymbol{\varepsilon}$ of $\boldsymbol{y}^*$, i.e. $\mathbf{0} \prec \boldsymbol{y}^* - \boldsymbol{\varepsilon} \leq \boldsymbol{y}^*$. As $F(\boldsymbol{X}, \boldsymbol{y}^* - \boldsymbol{\varepsilon}) \leq F(\boldsymbol{X}, \boldsymbol{y}^*)$, the Kleene sequence of $F(\boldsymbol{X}, \boldsymbol{y}^* - \boldsymbol{\varepsilon})$ is bounded from above by $\boldsymbol{x}^*$. So, for the least fixed-point $\mu\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\nu}_{\text{dep}(S)}) = \boldsymbol{x}^* - \boldsymbol{\delta}$ of $F(\boldsymbol{X}, \boldsymbol{y}^* - \boldsymbol{\varepsilon})$ we have $\mathbf{0} \leq \boldsymbol{\delta} \leq \boldsymbol{x}^*$. As we assume $\boldsymbol{y}^* - \boldsymbol{\varepsilon} \succ \mathbf{0}$, $F(\boldsymbol{X}, \boldsymbol{y}^* - \boldsymbol{\varepsilon})$ stays clean, hence $\boldsymbol{x}^* - \boldsymbol{\delta} \succ \mathbf{0}$, too.

We are now interested in bounding $\|\boldsymbol{\delta}\|$ by $\|\boldsymbol{\varepsilon}\|$. For this to do, let us rewrite the equation

$$\boldsymbol{x}^* - \boldsymbol{\delta} = F(\boldsymbol{x}^* - \boldsymbol{\delta}, \boldsymbol{y}^* - \boldsymbol{\varepsilon}) \quad \text{as} \quad \boldsymbol{x}^* - \boldsymbol{\delta} = F(\boldsymbol{x}^*, \boldsymbol{y}^*) - F'(\boldsymbol{x}^*, \boldsymbol{y}^*)\begin{pmatrix}\boldsymbol{\delta}\\\boldsymbol{\varepsilon}\end{pmatrix} + B(\begin{pmatrix}\boldsymbol{\delta}\\\boldsymbol{\varepsilon}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\delta}\\\boldsymbol{\varepsilon}\end{pmatrix}).$$

With $\boldsymbol{x}^* = F(\boldsymbol{x}^*, \boldsymbol{y}^*)$, and by moving all terms containing $\boldsymbol{\delta}$ on the right hand side, and the remaining terms on the left hand side, we get

$$F'(\boldsymbol{x}^*, \boldsymbol{y}^*)\begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix} - B(\begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix}, \begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix}) = (\text{Id} - F'(\boldsymbol{x}^*, \boldsymbol{y}^*))\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix} + B(\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}) + 2B(\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix}) \tag{8}$$

We remark that

$$\begin{aligned}(\text{Id} - F'(\boldsymbol{x}^*, \boldsymbol{y}^*))\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix} + B(\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}) &= F(\boldsymbol{x}^* - \boldsymbol{\delta}, \boldsymbol{y}^*) - F(\boldsymbol{x}^*, \boldsymbol{y}^*) + \boldsymbol{\delta}\\ &= F(\boldsymbol{x}^* - \boldsymbol{\delta}, \boldsymbol{y}^*) - (\boldsymbol{x}^* - \boldsymbol{\delta}).\end{aligned}$$

As $0 \leq \boldsymbol{x}^* - \boldsymbol{\delta} \leq \boldsymbol{x}^*$, we have

$$F(\boldsymbol{x}^* - \boldsymbol{\delta}, \boldsymbol{y}^*) - (\boldsymbol{x}^* - \boldsymbol{\delta}) > 0,$$

for $\boldsymbol{\delta} > \mathbf{0}$ – otherwise $\boldsymbol{x}^* - \boldsymbol{\delta}$ would be a fixed point less than $\boldsymbol{x}^*$ of $F(\boldsymbol{X}, \boldsymbol{y}^*)$.

Combining Eq. 8 with $B(\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix}) \geq \mathbf{0}$ and $-B(\begin{pmatrix}\mathbf{0}\\\boldsymbol{\delta}\end{pmatrix}, \begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix}) \leq \mathbf{0}$, we get

$$\begin{aligned}\mathbf{0} &< F(\boldsymbol{x}^* - \boldsymbol{\delta}, \boldsymbol{y}^*) - (\boldsymbol{x}^* - \boldsymbol{\delta})\\ &= (\text{Id} - F'(\boldsymbol{x}^*, \boldsymbol{y}^*))\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix} + B(\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix})\\ &\leq (\text{Id} - F'(\boldsymbol{x}^*, \boldsymbol{y}^*))\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix} + B(\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}) + 2B(\begin{pmatrix}\boldsymbol{\delta}\\\mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix})\\ &= F'(\boldsymbol{x}^*, \boldsymbol{y}^*)\begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix} - B(\begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix}, \begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix})\\ &\leq F'(\boldsymbol{x}^*, \boldsymbol{y}^*)\begin{pmatrix}\mathbf{0}\\\boldsymbol{\varepsilon}\end{pmatrix}.\end{aligned}$$

Thus we have

$$0 < \left\|(\mathrm{Id} - F'(\boldsymbol{x}^*, \boldsymbol{y}^*)) \begin{pmatrix}\boldsymbol{\delta}\\ \mathbf{0}\end{pmatrix} + B(\begin{pmatrix}\boldsymbol{\delta}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\delta}\\ \mathbf{0}\end{pmatrix})\right\| \leq \left\|F'(\boldsymbol{x}^*, \boldsymbol{y}^*) \begin{pmatrix}\mathbf{0}\\ \boldsymbol{\varepsilon}\end{pmatrix}\right\| \leq \left\|F'(\boldsymbol{x}^*, \boldsymbol{y}^*)\right\| \|\boldsymbol{\varepsilon}\| .$$

Now, if we succeed in showing that there is always a constant $C > 0$ such that

$$C \|\boldsymbol{\delta}\|^2 \leq \left\|(\mathrm{Id} - F'(\boldsymbol{x}^*, \boldsymbol{y}^*)) \begin{pmatrix}\boldsymbol{\delta}\\ \mathbf{0}\end{pmatrix} + B(\begin{pmatrix}\boldsymbol{\delta}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\delta}\\ \mathbf{0}\end{pmatrix})\right\|, \tag{9}$$

we will obtain a proof of Lemma 17.

Note that this last inequation does not depend on $\boldsymbol{\varepsilon}$ anymore. Hence, it is sufficient to consider $f(\boldsymbol{X}) := F(\boldsymbol{X}, \boldsymbol{y}^*)$ in the following, forgetting about the underlying SCCs. We write

$$b(\boldsymbol{X}, \boldsymbol{X}) = B(\begin{pmatrix}\boldsymbol{X}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{X}\\ \mathbf{0}\end{pmatrix})$$

for the quadratic part of f.

Then the preceding inequation Eq. 9 may be written as

$$C \|\boldsymbol{\delta}\|^2 \leq \left\|(\mathrm{Id} - f'(\boldsymbol{x}^*))\boldsymbol{\delta} + b(\boldsymbol{\delta}, \boldsymbol{\delta})\right\|, \tag{10}$$

where f' is now the Jacobian of f, i.e. taken only w.r.t. $\boldsymbol{X}$. Because of the equivalence of norms on $\mathbb{R}^n$, we may turn to the Euclidean norm $\|\cdot\|_2$ and apply Proposition 26 to conclude the proof. □

C.3 Proof of Theorem 18

Here is a restatement of Theorem 18.

Theorem 18. *Let $\boldsymbol{f}$ be a quadratic MSP. Let $\boldsymbol{\nu}^{(j)}$ denote the result of calling* $\mathrm{DNM}(\boldsymbol{f}, j)$ *(see Figure 1). Then there is a $k_{\boldsymbol{f}} \in \mathbb{N}$ such that $\boldsymbol{\nu}^{(k_{\boldsymbol{f}}+i)}$ has at least i valid bits of $\mu\boldsymbol{f}$ for every $i \geq 0$.*

We first prove the following lemma which gives a bound on the error on level t.

Lemma 27. *There is a constant $c > 0$ such that*

$$\left\|\boldsymbol{\Delta}_t^{(j)}\right\| \leq 2^{c-j\cdot 2^t} .$$

Proof. It follows from Theorem 8 that $(\widetilde{\boldsymbol{\mu}_t}^{(j)} - \boldsymbol{\nu}_t^{(j)})$, the approximation error at level t, decreases exponentially in the number of iterations, i.e., there is a constant $c_1 > 0$ such that

$$\left\|\widetilde{\boldsymbol{\mu}_t}^{(j)} - \boldsymbol{\nu}_t^{(j)}\right\| \leq 2^{c_1-j\cdot 2^t} . \tag{11}$$

Now we can prove the theorem by induction on t. In the base case ($t = h(\boldsymbol{f})$) there is no propagation error, so the claim of the lemma follows from (11). Let $t < h(\boldsymbol{f})$. Then

$$\begin{aligned}
\left\|\boldsymbol{\Delta}_t^{(j)}\right\| &= \left\|\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}_t}^{(j)} + \widetilde{\boldsymbol{\mu}_t}^{(j)} - \boldsymbol{\nu}_t^{(j)}\right\| \\
&\leq \left\|\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}_t}^{(j)}\right\| + \left\|\widetilde{\boldsymbol{\mu}_t}^{(j)} - \boldsymbol{\nu}_t^{(j)}\right\| \\
&\leq \left\|\boldsymbol{\mu}_t - \widetilde{\boldsymbol{\mu}_t}^{(j)}\right\| + 2^{c_1-j\cdot 2^t} && \text{(by (11))} \\
&\leq c_2 \cdot \sqrt{\left\|\boldsymbol{\Delta}_{>t}^{(j)}\right\|} + 2^{c_1-j\cdot 2^t} && \text{(by Lemma 17)} \\
&\leq c_2 \cdot \sqrt{2^{c_3-j\cdot 2^{t+1}}} + 2^{c_1-j\cdot 2^t} && \text{(by induction hypothesis)} \\
&\leq 2^{c_4-j\cdot 2^t}
\end{aligned}$$

for some constants $c_2, c_3, c_4 > 0$. □

From Lemma 27 we deduce that for each component r of depth t there is a constant c_r such that

$$(\mu\boldsymbol{f}_r - \nu_r^{(j)})/\mu\boldsymbol{f}_r \leq 2^{c_r - j\cdot 2^t} \leq 2^{c_r - j} .$$

Let $k_{\boldsymbol{f}} \geq c_r$ for all $1 \leq r \leq n$. Then

$$(\mu\boldsymbol{f}_r - \nu_r^{(j+k_{\boldsymbol{f}})})/\mu\boldsymbol{f}_r \leq 2^{c_r - (j+k_{\boldsymbol{f}})} \leq 2^{-j} . \quad \square$$

D Proofs of Section 6

D.1 Rest of the Proof of Theorem 19

In the following, if M is a matrix we often write M^i_{jk} resp. M^*_{jk} when we mean $(M^i)_{jk}$ resp. $(M^*)_{jk}$.

The following lemma assures that in order to show that $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*$ has no ∞ entries, it suffices to consider the diagonal elements of the matrix.

Lemma 28. *Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}_{\geq 0}$. Let A^* have an ∞ entry. Then A^* also has an ∞ entry on the diagonal, i.e. $A^*_{ii} = \infty$ for some $1 \leq i \leq n$.*

Proof. By induction on n. The base case $n = 1$ is clear. For $n > 1$ assume w.l.o.g. that $A^*_{1n} = \infty$. We have

$$A^*_{1n} = A^*_{11} \sum_{j=2}^{n} a_{1j} (A_{[2..n,2..n]})^*_{jn} , \tag{12}$$

where by $A_{[2..n,2..n]}$ we mean the square matrix obtained from A by erasing the first row and the first column. To see why (12) holds, think of A^*_{1n} as the sum of weights of paths from 1 to n in the complete graph over the vertices $\{1, \ldots, n\}$. The weight of a path P is the product of the weight of P's edges, and $a_{i_1 i_2}$ is the weight of the edge from i_1 to i_2. Each path P from 1 to n can be divided into two sub-paths P_1, P_2 as follows. The second sub-path P_2 is the suffix of P leading from 1 to n and not returning to 1. The first sub-path P_1, possibly empty, is chosen such that $P = P_1 P_2$. Now, the sum of weights of all possible P_1 equals A^*_{11}, and the sum of weights of all possible P_2 equals $\sum_{j=2}^{n} a_{1j}(A_{[2..n,2..n]})^*_{jn}$. So (12) holds.

As $A^*_{1n} = \infty$, it follows that either A^*_{11} or some $(A_{[2..n,2..n]})^*_{jn}$ equals ∞. In the first case, we are done. In the second case, by induction, there is an i such that $(A_{[2..n,2..n]})^*_{ii} = \infty$. But then also $A^*_{ii} = \infty$, because every entry of $(A_{[2..n,2..n]})^*$ is less or equal the corresponding entry of A^*. □

So it remains to show that $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*_{ss} \neq \infty$ for all $1 \leq s \leq n$. This is done in the proof of Proposition 32 below. There, two cases are considered, depending on whether Newton's method terminates in the s-component or not. The following lemma will be used for the nonterminating case.

Lemma 29. *Let $\mathbf{0} \leq \boldsymbol{\nu} \leq \boldsymbol{f}(\boldsymbol{\nu}) \leq \mu\boldsymbol{f}$. Let S denote an SCC with $\boldsymbol{\nu}_S \prec \mu\boldsymbol{f}_S$. Then the submatrix $\boldsymbol{f}'(\boldsymbol{\nu})^*_{SS}$ does not have ∞ as an entry.*

Proof. Let L denote the set of variables which are not in S but on which a variable in S depends. Let $\boldsymbol{g}(\boldsymbol{X}_S) := \boldsymbol{f}_S(\boldsymbol{X})[\boldsymbol{X}_L/\mu\boldsymbol{f}_L]$. Then $\boldsymbol{g}(\boldsymbol{X}_S)$ is an scMSP with $\mu\boldsymbol{g} = \mu\boldsymbol{f}_S$. As $\boldsymbol{\nu}_S \prec \mu\boldsymbol{g}$, Theorem 6 (1) is applicable, so $\boldsymbol{g}'(\boldsymbol{\nu}_S)^*$ does not have ∞ as an entry. With $\boldsymbol{g}'(\boldsymbol{\nu}_S)^* = \boldsymbol{f}'(\boldsymbol{\nu})^*_{SS}$, the lemma follows. □

The next lemma is a version of Taylor's theorem, which will be used in Lemma 31 below.

Lemma 30 (from [10])**.** *Let $\mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{f}(\boldsymbol{x})$ and let $d, k_1, k_2 \in \mathbb{N}$ with $k_2 \geq k_1$. Then*

$$\boldsymbol{f}^{d+k_2}(\boldsymbol{x}) - \boldsymbol{f}^{d+k_1}(\boldsymbol{x}) \geq \boldsymbol{f}'(\boldsymbol{f}^{k_1}(\boldsymbol{x}))^d(\boldsymbol{f}^{k_2}(\boldsymbol{x}) - \boldsymbol{f}^{k_1}(\boldsymbol{x})) ,$$

where by $\boldsymbol{f}^r(X)$ we mean $\boldsymbol{f}(\boldsymbol{f}^{r-1}(X))$ with $\boldsymbol{f}^0(X) = X$.

Proof. The lemma follows from a generalized form of Taylor's theorem stating:

For an MSP $\boldsymbol{f}$ and $\boldsymbol{v}, \boldsymbol{u} \geq \boldsymbol{0}$:

$$\boldsymbol{f}^d(\boldsymbol{v} + \boldsymbol{u}) \geq \boldsymbol{f}^d(\boldsymbol{v}) + \boldsymbol{f}'(\boldsymbol{v})^d \boldsymbol{u} \,.$$

For the sake of completeness we give a proof of this generalized form of Taylor's theorem, closely following the proof of [10].

For $d = 1$ (induction base) the statement is essentially Taylor's theorem (see e.g. [5]). Let $d \geq 1$. Then, by Taylor's theorem, we have:

$$\begin{array}{ll} \boldsymbol{f}^{d+1}(\boldsymbol{v} + \boldsymbol{u}) & \\ \quad = \boldsymbol{f}(\boldsymbol{f}^d(\boldsymbol{v} + \boldsymbol{u})) & \\ \quad \geq \boldsymbol{f}(\boldsymbol{f}^d(\boldsymbol{v}) + \boldsymbol{f}'(\boldsymbol{v})^d \boldsymbol{u}) & \text{(induction hypothesis)} \\ \quad \geq \boldsymbol{f}^{d+1}(\boldsymbol{v}) + \boldsymbol{f}'(\boldsymbol{f}^d(\boldsymbol{v}))\boldsymbol{f}'(\boldsymbol{v})^d \boldsymbol{u} & \text{(Taylor)} \\ \quad \geq \boldsymbol{f}^{d+1}(\boldsymbol{v}) + \boldsymbol{f}'(\boldsymbol{v})^{d+1} \boldsymbol{u} & \end{array}$$

Lemma 30 itself follows with $\boldsymbol{v} = \boldsymbol{f}^{k_1}(\boldsymbol{x})$ and $\boldsymbol{u} = \boldsymbol{f}^{k_2}(\boldsymbol{x}) - \boldsymbol{f}^{k_1}(\boldsymbol{x})$. □

The following lemma will be used for the case in which Newton's method terminates in some component X_s. It states that if Newton's method terminates in X_s it must have terminated before in some other component on which X_s depends.

Lemma 31. *Let $1 \leq s, l \leq n$. Let $\boldsymbol{f}'(\boldsymbol{X})^*_{ss}$ non-trivially depend on X_l. Let $\boldsymbol{0} \prec \boldsymbol{\nu} \leq \boldsymbol{f}(\boldsymbol{\nu}) \leq \mu\boldsymbol{f}$ and $\boldsymbol{\nu}_s < \mu\boldsymbol{f}_s$ and $\boldsymbol{\nu}_l < \mu\boldsymbol{f}_l$. Then $\widehat{\mathcal{N}}(\boldsymbol{\nu})_s < \mu\boldsymbol{f}_s$.*

Proof. This proof follows closely a proof of [12]. Let $d \geq 0$ s.t. $\boldsymbol{f}'(\boldsymbol{X})^d_{ss}$ depends non-trivially on X_l. Let $m' \geq 0$ s.t. $\boldsymbol{f}^{m'}(\boldsymbol{\nu})_l > \nu_l$. Such an m' exists because with Kleene's theorem the sequence $(\boldsymbol{f}^k(\boldsymbol{\nu}))_{k \in \mathbb{N}}$ converges to $\mu\boldsymbol{f}$. Choose $m \geq m'$ s.t. $\boldsymbol{f}^{m+1}(\boldsymbol{\nu})_s > \boldsymbol{f}^m(\boldsymbol{\nu})_s$. Such an m exists because the sequence $(\boldsymbol{f}^k(\boldsymbol{\nu})_s)_{k \in \mathbb{N}}$ never reaches $\mu\boldsymbol{f}_s$. This is because X_s depends on itself (since $\boldsymbol{f}'(\boldsymbol{X})^*_{ss}$ is not constant 0), and so every increase of the s-component results in an increase of the s-component in some later iteration of the Kleene sequence.

Now we have

$$\begin{array}{ll} \boldsymbol{f}^{d+m+1}(\boldsymbol{\nu}) - \boldsymbol{f}^{d+m}(\boldsymbol{\nu}) & \\ \quad \geq \boldsymbol{f}'(\boldsymbol{f}^m(\boldsymbol{\nu}))^d(\boldsymbol{f}^{m+1}(\boldsymbol{\nu}) - \boldsymbol{f}^m(\boldsymbol{\nu})) & \text{(Lemma 30)} \\ \quad \geq^* \boldsymbol{f}'(\boldsymbol{\nu})^d(\boldsymbol{f}^{m+1}(\boldsymbol{\nu}) - \boldsymbol{f}^m(\boldsymbol{\nu})) & \\ \quad \geq \boldsymbol{f}'(\boldsymbol{\nu})^d \boldsymbol{f}'(\boldsymbol{\nu})^m(\boldsymbol{f}(\boldsymbol{\nu}) - \boldsymbol{\nu}) & \text{(Lemma 30)} \\ \quad = \boldsymbol{f}'(\boldsymbol{\nu})^{d+m}(\boldsymbol{f}(\boldsymbol{\nu}) - \boldsymbol{\nu}) \,. & \end{array}$$

The inequality marked with $*$ is strict in the s-component – this is due to the choice of d and m above. So, with $b = d + m$ we have:

$$(\boldsymbol{f}^{b+1}(\boldsymbol{\nu}) - \boldsymbol{f}^b(\boldsymbol{\nu}))_s > (\boldsymbol{f}'(\boldsymbol{\nu})^b(\boldsymbol{f}(\boldsymbol{\nu}) - \boldsymbol{\nu}))_s \tag{13}$$

For other choices of b inequation (13) also holds, but with $\geq$ instead of $>$. Therefore:

$$\begin{array}{lll} \mu\boldsymbol{f}_s & = \left(\boldsymbol{\nu} + \sum_{i=0}^{\infty}(\boldsymbol{f}^{i+1}(\boldsymbol{\nu}) - \boldsymbol{f}^i(\boldsymbol{\nu}))\right)_s & \text{(Kleene)} \\ & > \left(\boldsymbol{\nu} + \boldsymbol{f}'(\boldsymbol{\nu})^*(\boldsymbol{f}(\boldsymbol{\nu}) - \boldsymbol{\nu})\right)_s & \text{(inequation (13))} \\ & = \left(\widehat{\mathcal{N}}(\boldsymbol{\nu})\right)_s & \qquad \square \end{array}$$

Now we are ready for the central proposition of this proof of Theorem 19.

Proposition 32. *For all $k \geq 0$ the matrix $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k)})^*$ does not have ∞ as entries.*

Proof. Using Lemma 28 it is enough to show that $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k)})^*_{ss} \neq \infty$ for all s.

Case 1: The sequence $(\widehat{\nu}_s^{(k)})_{k\in\mathbb{N}}$ does not terminate, i.e., $\widehat{\nu}_s^{(k)} < \mu\boldsymbol{f}_s$ for all $k \geq 0$. Then obviously this holds for all variables in the SCC of X_s. So Lemma 29 applies, hence $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k)})^*_{ss} \neq \infty$.

Case 2: The sequence $(\widehat{\nu}_s^{(k)})_{k\in\mathbb{N}}$ terminates, i.e., there is a $k_s \geq 1$ such that $\widehat{\nu}_s^{(k_s)} = \widehat{\nu}_s^{(k_s+1)} = \ldots = \mu\boldsymbol{f}_s$. Let k_s be the smallest such number, i.e., $\widehat{\nu}_s^{(k_s-1)} < \widehat{\nu}_s^{(k_s)} = \widehat{\nu}_s^{(k_s+1)} = \mu\boldsymbol{f}_s$. So there is a variable X_u on which X_s depends such that

$$0 < \boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k_s-1)})^*_{su}(\boldsymbol{f}(\widehat{\boldsymbol{\nu}}^{(k_s-1)}) - \widehat{\boldsymbol{\nu}}^{(k_s-1)})_u < \infty\,,$$

where the latter inequality is implied by Proposition 20. This implies $0 < \boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k_s-1)})^*_{su} < \infty$, therefore also $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k_s-1)})^*_{ss} < \infty$. But with Lemma 31, any variable X_l on which $\boldsymbol{f}'(\boldsymbol{X})^*_{ss}$ depends has already terminated one step earlier, i.e. $\widehat{\nu}_l^{(k_s-1)} = \widehat{\nu}_l^{(k_s)}$. Therefore $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k_s)})^*_{ss} = \boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k_s-1)})^*_{ss} < \infty$. As the l-component does not change any further we have $\boldsymbol{f}'(\widehat{\boldsymbol{\nu}}^{(k)})^*_{ss} < \infty$ for all $k \geq k_s$. Since $\boldsymbol{f}'(\boldsymbol{X})$ is monotone and $(\widehat{\boldsymbol{\nu}}^{(k)})_{k\in\mathbb{N}}$ is monotonically increasing, this holds also for $0 \leq k \leq k_s$. □

Combining Proposition 32 with Proposition 20 and the comments below Proposition 20 yields Theorem 19. □

D.2 Proof of Theorem 21

Here is a restatement of Theorem 21.

Theorem 21. *Let $\boldsymbol{f}$ be any quadratic MSP. Then the Newton sequence $(\boldsymbol{\nu}^{(k)})_{k\in\mathbb{N}}$ is well-defined and converges linearly to $\mu\boldsymbol{f}$. More precisely, there is a $k_{\boldsymbol{f}} \in \mathbb{N}$ such that $\boldsymbol{\nu}^{(k_{\boldsymbol{f}}+i\cdot(h(\boldsymbol{f})+1)\cdot 2^{h(\boldsymbol{f})})}$ has at least i valid bits of $\mu\boldsymbol{f}$ for every $i \geq 0$.*

We argue that Theorem 18 assuring linear convergence for DNM essentially carries over to the "undecomposed" method.

The following lemma states that a Newton step is not faster on an SCC, if the values of the lower SCCs are fixed.

Lemma 33. *Let $\boldsymbol{f}$ be an MSP. Let $\mathbf{0} \leq \boldsymbol{\nu} \leq \boldsymbol{f}(\boldsymbol{\nu}) \leq \mu\boldsymbol{f}$. Let S denote an SCC of $\boldsymbol{f}$. Let L denote the set of variables that are not in S, but on which a variable in S depends. Then $(\widehat{\mathcal{N}}_{\boldsymbol{f}}(\boldsymbol{\nu}))_S \geq \widehat{\mathcal{N}}_{\boldsymbol{f}_S[\boldsymbol{X}_L/\boldsymbol{\nu}_L]}(\boldsymbol{\nu}_S)$, where $\widehat{\mathcal{N}}_{\boldsymbol{f}}$ is defined as in Proposition 20.*

Proof.

$$\begin{aligned}
&(\widehat{\mathcal{N}}_{\boldsymbol{f}}(\boldsymbol{\nu}))_S \\
&\quad = \big(\boldsymbol{f}'(\boldsymbol{\nu})^*(\boldsymbol{f}(\boldsymbol{\nu}) - \boldsymbol{\nu})\big)_S \\
&\quad = \boldsymbol{f}'(\boldsymbol{\nu})^*_{SS}(\boldsymbol{f}(\boldsymbol{\nu}) - \boldsymbol{\nu})_S \\
&\quad\quad + \boldsymbol{f}'(\boldsymbol{\nu})^*_{SL}(\boldsymbol{f}(\boldsymbol{\nu}) - \boldsymbol{\nu})_L \\
&\quad \geq \boldsymbol{f}'(\boldsymbol{\nu})^*_{SS}(\boldsymbol{f}(\boldsymbol{\nu}) - \boldsymbol{\nu})_S \\
&\quad = \big((\boldsymbol{f}_S[\boldsymbol{X}_L/\boldsymbol{\nu}_L])'(\boldsymbol{\nu}_S)\big)^*(\boldsymbol{f}_S[\boldsymbol{X}_L/\boldsymbol{\nu}_L](\boldsymbol{\nu}_S) - \boldsymbol{\nu}_S) \\
&\quad = \widehat{\mathcal{N}}_{\boldsymbol{f}_S[\boldsymbol{X}_L/\boldsymbol{\nu}_L]}(\boldsymbol{\nu}_S) \quad \square
\end{aligned}$$

The following lemma states the monotonicity of Newton's method and was proved in [16].

Lemma 34 (Monotonicity of Newton's Method)**.** *Let $\boldsymbol{f}(\boldsymbol{X})$ be an MSP. Then*

$$\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{x}) \leq \mathcal{N}_{\boldsymbol{f}}(\boldsymbol{y}) \text{ for all } \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{y} \leq \boldsymbol{f}(\boldsymbol{y}) \leq \mu\boldsymbol{f}.$$

Lemma 33 and Lemma 34 can be combined to the following lemma stating that $i \cdot (h(\boldsymbol{f}) + 1)$ iterations of the normal Newton's method "dominate" i iterations of a decomposed Newton's method in each SCC.

Lemma 35. *Let $\widetilde{\boldsymbol{\nu}}^{(i)}$ denote the result of a decomposed Newton's method which performs i iterations of Newton's method in each SCC. Let $\boldsymbol{\nu}^{(i)}$ denote the result of i iterations of the normal "undecomposed" Newton's method. Then $\boldsymbol{\nu}^{(i\cdot(h(\boldsymbol{f})+1))} \geq \widetilde{\boldsymbol{\nu}}^{(i)}$.*

Proof. Let $h = h(\boldsymbol{f})$. Let $C(t)$ resp. $C(> t)$ denote the set of variables in an SCC of depth t resp. $> t$. We show by induction on the depth t:

$$\boldsymbol{\nu}_{C(t)}^{(i\cdot(h(\boldsymbol{f})+1-t))} \geq \widetilde{\boldsymbol{\nu}}_{C(t)}^{(i)}$$

Induction base: $t = h(\boldsymbol{f})$. Clear, because for bottom SCCs the two methods are identical.
Let now $t < h(\boldsymbol{f})$. Then

$$\begin{aligned}
&\boldsymbol{\nu}_{C(t)}^{(i\cdot(h+1-t))} \\
&= \mathcal{N}_{\boldsymbol{f}}^{i}(\boldsymbol{\nu}^{(i\cdot(h-t))})_{C(t)} \\
&\geq \mathcal{N}^{i}_{\boldsymbol{f}_{C(t)}[\boldsymbol{X}/\boldsymbol{\nu}_{C(>t)}^{(i\cdot(h-t))}]}(\boldsymbol{\nu}_{C(t)}^{(i\cdot(h-t))}) && \text{(Lemma 33)} \\
&\geq \mathcal{N}^{i}_{\boldsymbol{f}_{C(t)}[\boldsymbol{X}/\widetilde{\boldsymbol{\nu}}_{C(>t)}^{(i)}]}(\boldsymbol{\nu}_{C(t)}^{(i\cdot(h-t))}) && \text{(induction hypothesis)} \\
&\geq \mathcal{N}^{i}_{\boldsymbol{f}_{C(t)}[\boldsymbol{X}/\widetilde{\boldsymbol{\nu}}_{C(>t)}^{(i)}]}(\boldsymbol{0}_{C(t)}) && \text{(Lemma 34)} \\
&= \widetilde{\boldsymbol{\nu}}_{t}^{(i)}
\end{aligned}$$

Now, the lemma itself follows by using Lemma 34 once more. □

As a side note, observe that above proof of Lemma 35 implicitly benefits from the fact that SCCs of the same depth are independent. So, SCCs with the same depth are handled in parallel by the "undecomposed" Newton's method. Therefore, $w(\boldsymbol{f})$, the width of $\boldsymbol{f}$, is irrelevant here (cf. Proposition 16).

Now we can finish the proof of Theorem 21. Let k_2 be the $k_{\boldsymbol{f}}$ of Theorem 18, and let $k_1 = k_2 \cdot (h(\boldsymbol{f})+1) \cdot 2^{h(\boldsymbol{f})}$. Then we have:

$$\begin{aligned}
&\boldsymbol{\nu}^{(k_1+i\cdot(h(\boldsymbol{f})+1)\cdot 2^{h(f)})} \\
&\quad= \boldsymbol{\nu}^{((k_2+i)\cdot(h(\boldsymbol{f})+1)\cdot 2^{h(\boldsymbol{f})})} \\
&\quad\geq \widetilde{\boldsymbol{\nu}}^{((k_2+i)\cdot 2^{h(\boldsymbol{f})})} && \text{(Lemma 35)}
\end{aligned}$$

The approximation $\widetilde{\boldsymbol{\nu}}^{((k_2+i)\cdot 2^{h(\boldsymbol{f})})}$ has at least as many bits as the approximation obtained from running DNM$(\boldsymbol{f}, k_2 + i)$. This is because DNM$(\boldsymbol{f}, k_2 + i)$ runs at most $(k_2 + i) \cdot 2^{h(\boldsymbol{f})}$ iteration in every SCC and Newton's method converges monotonically. So, by Theorem 18, $\boldsymbol{\nu}^{(k_1+i\cdot(h(\boldsymbol{f})+1)\cdot 2^{h(f)})}$ has at least i valid bits of $\mu\boldsymbol{f}$. Therefore, Theorem 21 holds with $k_{\boldsymbol{f}} = k_1$. □