On the Convergence of Newton's Method for Monotone Systems of Polynomial Equations [*]

Stefan Kiefer, Michael Luttenberger, and Javier Esparza
Institute for Formal Methods in Computer Science
Universität Stuttgart, Germany
$\{$kiefersn, luttenml, esparza$\}$@informatik.uni-stuttgart.de

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 $\mathbf{X} = \mathbf{f}(\mathbf{X})$ arises naturally in the analysis of stochastic context-free grammars, recursive Markov chains, and probabilistic pushdown automata. While the Kleene sequence $\mathbf{f}(\mathbf{0}), \mathbf{f}(\mathbf{f}(\mathbf{0})), \ldots$ always converges to the least solution $\mu\mathbf{f}$, if it exists, the number of iterations needed to compute the first i bits of $\mu\mathbf{f}$ may grow exponentially in i. Etessami and Yannakakis have recently adapted Newton's iterative method to MSPEs and proved that the Newton sequence converges at least as fast as the Kleene sequence and exponentially faster in many cases. They conjecture that, given an MSPE of size m, the number of Newton iterations needed to obtain i accurate bits of $\mu\mathbf{f}$ grows polynomially in i and m. In this paper we show that the number of iterations grows linearly in i for *strongly connected* MSPEs and may grow exponentially in m for general MSPEs.

Categories and Subject Descriptors

G.1.5 [**Mathematics of Computing**]: Numerical Analysis

General Terms

Algorithms, Theory, Verification

Keywords

Newton's Method, Fixed-Point Equations, Formal Verification of Software, Probabilistic Pushdown Systems

1. INTRODUCTION

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

$$\begin{array}{rcl} X_1 & = & f_1(X_1, \ldots, X_n) \\ X_2 & = & f_2(X_1, \ldots, X_n) \\ & \vdots & \\ X_n & = & f_n(X_1, \ldots, X_n) \end{array}$$

where $f_1, \ldots, f_n$ are polynomials with *positive* real coefficients. In vector form we denote an MSPE by $\mathbf{X} = \mathbf{f}(\mathbf{X})$. We call MSPEs monotone because $\mathbf{X} \leq \mathbf{X}'$ implies $\mathbf{f}(\mathbf{X}) \leq \mathbf{f}(\mathbf{X}')$ for $\mathbf{X}, \mathbf{X}' \in \mathbb{R}^n_{\geq 0}$. MSPEs appear naturally in different areas of computer science. One of them is the analysis of stochastic context-free grammars, a model widely used in natural language processing [20, 15] and more recently in computational biology [25, 4, 3, 18]. The productions of these grammars are assigned a probability, subject to the restriction that for each variable X the sum of the probabilities of all productions with X on the left-hand side is either 0 or 1. It is easy to see that the probabilities of many events – like for instance the probability that the grammar generates a word containing some terminal or some factor – can be expressed as the least solution of an MSPE. Another application area is the probabilistic verification of recursive Markov chains and probabilistic pushdown automata, two equivalent formalisms combining probability and recursion. These formalisms can be used to analyze probabilistic programs with procedures, and have been extensively studied [6, 1, 10, 8, 7, 9, 11]. Many verification problems for these models reduce to computing the termination probabilities (roughly speaking, the probability that the system terminates when started in a certain configuration), which in turn can be expressed as the least solution of an MSPE. A third example of application is the stochastic process used in [13, 14] to model the behaviour of web users in the presence of a 'back' button; again, basic analysis problems about these processes reduce to computing the least solution of an MSPE.

An MSPE can have zero, one, or many real solutions. We call it *feasible* if it has at least one solution. Solutions can be irrational and non-expressible by radicals. By monotonicity, the set of solutions forms a lattice with respect to the pointwise ordering on $\mathbb{R}^n_{\geq 0}$ (this follows from Knaster-Tarski's theorem), and so a feasible system $\mathbf{X} = \mathbf{f}(\mathbf{X})$ has a least solution $\mu\mathbf{f}$. In this paper we focus on the problem of numerically computing the least solution of a feasible system.

[*] This work was partially supported by the DFG project *Algorithms for Software Model Checking*.

STOC'07, June 11–13, 2007, San Diego, California, USA.
Copyright 2007 ACM 978-1-59593-631-8/07/0006 ...$5.00.

Newton's method.

MSPEs derived from recursive Markov chains or probabilistic pushdown automata are always feasible [10]. Etessami and Yannakakis have studied these systems in [10] and extended some of the result to all feasible MSPEs in [12]. They show that (the multivariate version of) Newton's method for finding zeros of differentiable functions can be used to approximate the least solution. We now proceed to explain this result in some more detail.

Finding the least solution of a feasible system $\mathbf{X} = \mathbf{f}(\mathbf{X})$ amounts to finding the least solution of $\mathbf{F}(\mathbf{X}) = \mathbf{0}$ for $\mathbf{F}(\mathbf{X}) = \mathbf{f}(\mathbf{X}) - \mathbf{X}$. For this we can apply the multivariate version of Newton's method [22]: starting at some $\mathbf{x}^{(0)} \in \mathbb{R}^n$ (we use uppercase to denote variables and lowercase to denote values), compute the sequence

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

where $\mathbf{F}'(\mathbf{X})$ is the Jacobian matrix of partial derivatives. Observe the method may not even be defined because $\mathbf{F}'(\mathbf{x}^{(k)})$ may be singular for some k. It is shown in [10] that this problem can be avoided if (a) the MSPE is previously *cleaned* and (b) the clean MSPE is decomposed into *strongly connected components* (SCCs). We consider these two points in turn. A feasible MSPE is *clean* if all the components of its least solution are positive; an MSPE can be cleaned in linear time by identifying the variables X_i for which the least solution has value 0 and then removing all occurrences of X_i in $\mathbf{f}$ together with the equation $X_i = f_i(X_1, \ldots, X_n)$. In order to define the SCCs of an MSPE, associate to each subset of equations of $\mathbf{f}$ a graph having the variables $X_1, \ldots, X_n$ as node, and the pairs (X_i, X_j) such that X_j appears in f_i as edges. We say that the subset is strongly connected if its associated graph is strongly connected. It is easy to see that every MSPE can be partitioned into SCCs, which can be topologically ordered. Etessami and Yannakakis' *decomposed Newton's method* for a clean system starts by computing k iterations of Newton's method for each bottom SCC of the system. Then the values obtained for the variables of these SCCs are "frozen" and their corresponding equations are removed from the MSPE. This 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 the decomposed method:

- The Jacobian matrices of all the SCCs remain invertible all the way throughout.
- The vector $\mathbf{x}^{(k)}$ delivered by the method converges to the least solution when $k \to \infty$ *even if* $\mathbf{x}^{(0)} = \mathbf{0} = (0, \ldots, 0)^\top$.

It is important to emphasize that this second property 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. In particular, this means that in order to compute the solution the method does not have to guess an appropriate initial value.

While the decomposed Newton's method converges for all feasible MSPEs, almost nothing is known about its convergence rate, i.e., about the number of accurate bits of $\mathbf{x}^{(k)}$ as a function of the number of iterations (see the section on related work for a discussion), or, equivalently, of its inverse, the number of iterations needed to compute a given number of bits of the solution. This is the topic of this paper.

Results.

Let $\mathbf{X} = \mathbf{f}(\mathbf{X})$ be a clean and feasible MSPE and let $\mu\mathbf{f}$ be its least solution. Let $iter_{\mathbf{f}}(i)$ be the least number k such that $\left\|\mu\mathbf{f} - \mathbf{x}^{(k)}\right\| / \|\mu\mathbf{f}\| \leq 2^{-i}$, where $\|\cdot\|$ is some norm. Loosely speaking, $iter_{\mathbf{f}}(i)$ is the number of iterations needed to obtain the first i bits of the solution. Further, let $Iter(m, i)$ be the maximum of $iter_{\mathbf{f}}(i)$ over all clean feasible MSPEs $\mathbf{f}$ of size at most m. (As usual, the size of an MSPE is the number of bits needed to describe the system with coefficients in binary.) Our first and main result concerns $Iter(m, i)$ for strongly connected MSPEs (or, equivalently, the number of iterations required to obtain i bits of accuracy when processing a SCC in the decomposed method). We prove that for strongly connected MSPEs there is a function $g(m)$ such that $Iter(m, i) \leq c \cdot i + g(m)$ for every $m, i \geq 0$. Actually, we even prove that one can choose $c = 1$, i.e., after a certain number of iterations the method computes 1 new bit of the solution per iteration.

Our second result investigates the *threshold* of the decomposed Newton's method, defined as the function $Thr(m) = Iter(m, 1)$. In words, $Thr(m)$ is the maximal number of iterations needed in order to compute the first bit of $\mu\mathbf{f}$ for a system $\mathbf{f}$ of size m. We show that $Thr(m) \in \Omega(2^{d \cdot m})$ for some $d > 0$, i.e., in the worst case the threshold is at least exponential in the size of the system.

Related work.

Etessami and Yannakakis study MSPEs in [10, 12]. They show that the decomposed Newton's method converges at least as fast as the Kleene iteration scheme $\mathbf{x}^{(0)} = \mathbf{0}$ and $\mathbf{x}^{(k+1)} = \mathbf{f}(\mathbf{x}^{(k)})$. (The convergence of this scheme is guaranteed by Kleene's theorem, see for instance [19].)[1] However, as shown in [10], it is easy to construct examples (for instance the monotone equation $X = 1/2X^2 + 1/2$ with 1 as least solution) for which the convergence of Kleene's scheme is unacceptably slow, namely the number of iterations needed to compute i bits of the solution is exponential in i. In fact, this slowness is the motivation of [10] for studying Newton's method. In [12] Etessami and Yannakakis conjecture (Conjecture 26) that for the decomposed Newton's method the function $Iter(m, i)$ is polynomial in both m and i. Our second result refutes the conjecture, but our first and main result proves part of it for strongly connected MSPEs.

There is a large literature on the convergence and convergence rate of Newton's method for arbitrary systems of differentiable functions. A comprehensive reference is Ortega and Rheinboldt's book [22] (see also Chapter 8 of Ortega's course [21] or Chapter 5 of [17] for a brief summary). Several theorems (for instance Theorem 8.1.10 of [21]) prove that the number of accurate bits grows linearly, superlinearly, or even exponentially in the number of iterations, but only under the hypothesis that $\mathbf{F}'(\mathbf{x})$ is non-singular everywhere, in a neighborhood of $\mu\mathbf{f}$, or at least at the point $\mu\mathbf{f}$ itself. However, the matrix $\mathbf{F}'(\mu\mathbf{f})$ can be singular for an MSPE; an example is again given by the equation above.

The general case in which $\mathbf{F}'(\mu\mathbf{f})$ is singular for the solu-

[1] In fact, this result is proved in [10] only for the MSPEs derived from recursive Markov chains. The extension to arbitrary MSPEs is considered in [12].

tion $\mu\mathbf{f}$ the method converges to has been thoroughly studied. In a seminal paper [24], Reddien shows that under certain conditions, the main ones being that the kernel of $\mathbf{F}'(\mu\mathbf{f})$ has dimension 1 and that the initial point is close enough to the solution, Newton's method gains 1 bit per iteration. Decker and Kelly obtain results for kernels of arbitrary dimension, but they require a certain linear map $B(\mathbf{X})$ to be non-singular for all $\mathbf{x} \neq \mathbf{0}$ [2]. Griewank observes in [16] that the non-singularity of $B(\mathbf{X})$ is in fact a strong condition which, in particular, can only be satisfied by kernels of even dimension. He presents a weaker sufficient condition for linear convergence requiring $B(\mathbf{X})$ to be non-singular only at the initial point $\mathbf{x}^{(0)}$, i.e., it only requires to make "the right guess" for $\mathbf{x}^{(0)}$. Unfortunately, none of these results can be directly applied to arbitrary MSPEs. The possible dimensions of the kernel of $\mathbf{F}'(\mu\mathbf{f})$ for an MSPE $\mathbf{f}(\mathbf{X})$ are to the best of our knowledge unknown, and deciding this question seems as hard as those related to the convergence rate[2]. Griewank's result does not apply to the decomposed Newton's method either because the mapping $B(\mathbf{x}^{(0)})$ is always singular for $\mathbf{x}^{(0)} = \mathbf{0}$.

Kantorovich's famous theorem (see e.g. Theorem 8.2.6 of [22] and [23] for an improvement) guarantees global convergence and only requires $\mathbf{F}'$ to be non-singular at $\mathbf{x}^{(0)}$. However, it also requires to find a Lipschitz constant for $\mathbf{F}'$ on a suitable region and some other bounds on $\mathbf{F}'$. These latter conditions are far too restrictive for the applications mentioned above. For instance, the stochastic context-free grammars whose associated MSPEs satisfy Kantorovich's conditions cannot exhibit two productions $X \to aYZ$ and $W \to \varepsilon$ such that $\mathit{Prob}(X \to aYZ) \cdot \mathit{Prob}(W \to \varepsilon) \geq 1/4$. This class of grammars is too contrived to be of use.

Organization of this paper.

The rest of this paper is structured as follows. In Section 2 we give preliminaries and some technical background on MSPEs. In Section 3 we state and prove our main result concerning the linear convergence of the decomposed Newton's method applied to strongly connected MSPEs. In Section 4 we show that $\mathit{Thr}(m)$ is not bounded by a polynomial, which refutes the conjecture of [12]. We conclude in Section 5.

2. PRELIMINARIES

In this section we introduce notations and formalize the concepts mentioned in the introduction.

2.1 Notation

As usual, $\mathbb{R}$ and $\mathbb{N}$ denote the set of real and 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. $\mathbf{x} \in \mathbb{R}^n$, where we assume that $\mathbf{x}$ has the components $x_1, \ldots, x_n$. Similarly, the i^{th} component of a function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ 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 canonical unit vectors of $\mathbb{R}^n$ are denoted by $\mathbf{e}_1, \ldots, \mathbf{e}_n$, e.g. $\mathbf{e}_1 = (1, 0, \ldots, 0)^\top$, and the identity matrix of $\mathbb{R}^{n \times n}$ is Id.

We use $\|\cdot\|$ to denote any norm on $\mathbb{R}^n$. In particular, $\|\cdot\|_\infty$ is the maximum norm, i.e. $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$. We recall the fact that all norms on $\mathbb{R}^n$ are equivalent as for two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ there always exist constants $c, C > 0$ such that $c\|\cdot\|_a \leq \|\cdot\|_b \leq C\|\cdot\|_a$. Finally, let $\|\cdot\|_a$ be a norm on $\mathbb{R}^m$ and $\|\cdot\|_b$ be a norm on $\mathbb{R}^n$, we then set $\|A\|_{b,a} := \max_{\mathbf{x} \in \mathbb{R}^m, \|\mathbf{x}\|_a = 1} \|A\mathbf{x}\|_b$ for $A \in \mathbb{R}^{n \times m}$. As $\{\mathbf{x} \in \mathbb{R}^m \mid \|\mathbf{x}\|_a = 1\}$ is compact and $\|A\mathbf{x}\|_b$ is continuous, $\|A\|_{b,a}$ is always defined. We then have $\|A\mathbf{x}\|_b \leq \|A\|_{b,a} \|\mathbf{x}\|_a$.

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 $A^* = (\mathrm{Id} - A)^{-1}$ (but the existence of $(\mathrm{Id} - A)^{-1}$ does not imply the existence of A^*).

The partial order $\leq$ on $\mathbb{R}^n$ is defined as usual by setting $\mathbf{x} \leq \mathbf{y}$ if $x_i \leq y_i$ for all $i \in \{1, \ldots, n\}$. Similarly, $\mathbf{x} < \mathbf{y}$ if $\mathbf{x} \leq \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$. Finally, we write $\mathbf{x} \prec \mathbf{y}$ if $x_1 < y_i$ for all $i \in \{1, \ldots, n\}$, i.e., if every component of $\mathbf{x}$ is smaller than the corresponding component of $\mathbf{y}$.

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

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

$$\begin{pmatrix} \frac{\partial f_1}{\partial X_1} & \cdots & \frac{\partial f_1}{\partial X_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial X_1} & \cdots & \frac{\partial f_m}{\partial X_n} \end{pmatrix},$$

for which we simply write $\mathbf{f}'(\mathbf{X})$.

2.2 Monotone Systems of Polynomials

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

Fact 1. Every MSP $\mathbf{f}$ is monotone on $\mathbb{R}^n_{\geq 0}$, i.e. for $\mathbf{0} \leq \mathbf{x} \leq \mathbf{y}$ we have $\mathbf{f}(\mathbf{x}) \leq \mathbf{f}(\mathbf{y})$.

Since every MSP is continuous, Kleene's fixed-point theorem (see e.g. [19]) applies.

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

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

As mentioned in the introduction, the convergence of the Kleene sequence can be extremely slow. For this reason in

[2] More precisely, MSPEs with kernels of arbitrary dimension exist, but the cases we know of can be trivially reduced to MSPEs with kernels of dimension 1.

[10] Etessami and Yannakakis present a decomposed Newton's method, which we now define. We first introduce the notions of clean and strongly connected MSPs.

Definition 2. A variable X_i of an MSP $\mathbf{f}(\mathbf{X})$ 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 not hard 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 the MSP.

Definition 3. Let $\mathbf{f}(\mathbf{X})$ be an MSP. X_i *depends directly* on X_k, denoted by $X_i \trianglelefteq X_k$, if $\frac{\partial f_i}{\partial X_k}(\mathbf{X})$ 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.

The following result is proved in [10, 12]:

Theorem 2. *Let $\mathbf{f}(\mathbf{X})$ be a clean feasible scMSP and define the Newton operator $\mathcal{N}_\mathbf{f}$ as follows*

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

We have:

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

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

This result leads to the following *decomposed Newton's method* introduced in [10]. The method starts by decomposing the MSP into strongly connected components (SCCs). Then, the solution for each of the bottom SCCs is approximated by means of a number k of iterations of Newton's method (this can be done by Theorem 2). For every variable X_i of the bottom SCCs, we substitute every occurrence of X_i by $\nu_i^{(k)}$, iterate the procedure with the bottom SCCs of the resulting MSP, and proceed like this until all SCCs have been processed.

To show that this procedure indeed works, we have to prove that the MSP obtained after the substitution also has a least fixed point. For this, let $\mathbf{f}_{\text{app}}(\mathbf{X})$ and $\mathbf{f}_{\mu\mathbf{f}}(\mathbf{X})$ be the result of substituting every occurrence of X_i by $\nu_i^{(k)}$ and $(\mu\mathbf{f})_i$, respectively. Since $\boldsymbol{\nu}^{(k)} \le \mu\mathbf{f}$, we have $\mathbf{f}_{\text{app}}(\mathbf{X}) \le \mathbf{f}_{\mu\mathbf{f}}(\mathbf{X})$, hence the Kleene sequence of $\mathbf{f}_{\text{app}}$ is bounded from above by that of $\mathbf{f}_{\mu\mathbf{f}}$, and we can apply the following proposition:

Proposition 1. *Let $\mathbf{f}(\mathbf{X})$ and $\mathbf{g}(\mathbf{X})$ be two MSPs. Assume that $\mu\mathbf{f}$ exists and $\mathbf{g}(\mathbf{x}) \le \mathbf{f}(\mathbf{x})$ for all $\mathbf{0} \le \mathbf{x} \le \mu\mathbf{f}$. Then $\boldsymbol{\kappa}_\mathbf{g}^{(k)} \le \boldsymbol{\kappa}_\mathbf{f}^{(k)}$, $\mu\mathbf{g}$ exists, and $\mu\mathbf{g} \le \mu\mathbf{f}$.*

Proof. A straightforward induction shows $\boldsymbol{\kappa}_\mathbf{g}^{(k)} \le \boldsymbol{\kappa}_\mathbf{g}^{(k+1)} \le \boldsymbol{\kappa}_\mathbf{f}^{(k+1)} \le \mu\mathbf{f}$ for all $k \in \mathbb{N}$. Hence, $\boldsymbol{\kappa}_\mathbf{g}^{(k)}$ converges. The limit must be $\mu\mathbf{g}$. □

We close this section with the important result that the Newton operator defined by a clean feasible scMSP is monotone, too. This is crucial for several of our proofs.

Lemma 1 (Monotonicity of Newton's Method). *Let $\mathbf{f}(\mathbf{X})$ be a clean feasible scMSP. Then*

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

Proof. Obviously, for $\mathbf{x} \le \mathbf{y}$ we have $\mathbf{f}'(\mathbf{x}) \le \mathbf{f}'(\mathbf{y})$ as every entry of $\mathbf{f}'(\mathbf{x})$ is a monotone polynomial. Hence, $\mathbf{f}'(\mathbf{x})^* \le \mathbf{f}'(\mathbf{y})^*$, too. With this at hand we get:

$$\begin{array}{rl} & \mathcal{N}_\mathbf{f}(\mathbf{y}) \\ = & \mathbf{y} + \mathbf{f}'(\mathbf{y})^*(\mathbf{f}(\mathbf{y}) - \mathbf{y}) \quad \text{(by Theorem 2)} \\ \ge & \mathbf{y} + \mathbf{f}'(\mathbf{x})^*(\mathbf{f}(\mathbf{y}) - \mathbf{y}) \\ & \quad \text{(since } \mathbf{x} \le \mathbf{y} \Rightarrow \mathbf{f}'(\mathbf{x})^* \le \mathbf{f}'(\mathbf{y})^*\text{)} \\ \ge & \mathbf{y} + \mathbf{f}'(\mathbf{x})^*(\mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) - \mathbf{y}) \\ & \quad \text{(since } \mathbf{f}(\mathbf{y}) \ge \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x})(\mathbf{y} - \mathbf{x})\text{)} \\ = & \mathbf{y} + \mathbf{f}'(\mathbf{x})^*((\mathbf{f}(\mathbf{x}) - \mathbf{x}) - (\mathrm{Id} - \mathbf{f}'(\mathbf{x}))(\mathbf{y} - \mathbf{x})) \\ = & \mathbf{y} + \mathbf{f}'(\mathbf{x})^*(\mathbf{f}(\mathbf{x}) - \mathbf{x}) - (\mathbf{y} - \mathbf{x}) \\ & \quad \text{(since } \mathbf{f}'(\mathbf{x})^* = (\mathrm{Id} - \mathbf{f}'(\mathbf{x}))^{-1}\text{)} \\ = & \mathcal{N}_\mathbf{f}(\mathbf{x}). \end{array}$$

At the second inequation above we used a version of Taylor's theorem (see appendix, Lemma 9). □

3. LINEAR CONVERGENCE FOR STRONGLY CONNECTED MSPS

In this section we prove our main result: the number of iterations needed by Newton's method to compute the first i bits of the least solution of a clean feasible scMSP grows linearly in i. More precisely, we show the following theorem:

Theorem 3. *Let $\mathbf{f}(\mathbf{X})$ be a clean feasible scMSP. There exists a $k_\mathbf{f} \in \mathbb{N}$ such that*

$$\frac{\left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(l+k_\mathbf{f})}\right\|}{\|\mu\mathbf{f}\|} \le 2^{-l} \text{ for all } l \in \mathbb{N}.$$

The MSP $f(X) = (X-1)^2 + X$ shows that this bound is tight, as its Newton sequence is $\boldsymbol{\nu}^{(k)} = 1 - 2^{-k}$.

We can easily reformulate the theorem to get the result promised in the introduction:

Corollary 1. *Let $scMSP_m$ be the set of all clean and feasible scMSPs of size at most m. Then there is a function $g : \mathbb{N} \to \mathbb{N}$ such that $\mathit{Iter}(m,i) := \max\{j_\mathbf{f}(i) \mid \mathbf{f} \in scMSP_m\} \le g(m) + i$.*

Proof. For $i \in \mathbb{N}$ let $j_\mathbf{f}(i)$ be the least number such that $\frac{\left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(j_\mathbf{f}(i))}\right\|}{\|\mu\mathbf{f}\|} \le 2^{-i}$. By Theorem 3 we have $j_\mathbf{f}(i) \le i + k_\mathbf{f}$. As scMSP_m is finite, $g(m) := \max\{k_\mathbf{f} \mid \mathbf{f} \in \mathrm{scMSP}_m\}$ exists for all $m \in \mathbb{N}$. Hence, $\mathrm{Iter}(m,i) \le \max\{i + k_\mathbf{f} \mid \mathbf{f} \in \mathrm{scMSP}_m\} = i + g(m)$. □

Theorem 3 implies that from some moment on the number of bits of the current approximation which we know to be correct grows at a rate of one bit per iteration. We say that Newton's method converges *linearly* if the expression of Theorem 3 holds. Using the terms of [17], Theorem 3 states *r-linear* convergence, not *q-linear* convergence which would require $||\mu\mathbf{f} - \boldsymbol{\nu}^{(l+1)}|| \le c \cdot ||\mu\mathbf{f} - \boldsymbol{\nu}^{(l)}||$ for some $c < 1$.

The rest of the section is devoted to the proof of Theorem 3. In Subsection 3.1 we show that it suffices to prove the result for quadratic scMSPs. In Subsection 3.2 we first recall that Newton's method converges quadratically[3], if the

[3] q-quadratical convergence, see [17]

matrix $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))^{-1}$ exists, and then proceed to consider the case in which $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))$ is singular. We show that in this case Newton's method has linear convergence if the kernel of $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))$ contains some vector $\mathbf{d} \succ \mathbf{0}$. Finally, in Subsection 3.3 we prove that such a $\mathbf{d}$ always exists by means of a perturbation method.

3.1 Reduction to Quadratic Polynomials

We reduce MSPs to quadratic MSPs, i.e., to MSPs in which every polynomial $f_i(\mathbf{X})$ has degree at most 2, while not improving the convergence rate of Newton's method.

The idea to reduce the degree of our MSP $\mathbf{f}$ is to introduce auxiliary variables that express quadratic subterms. This can be done repeatedly until all polynomials in the system have reached degree at most 2. The construction is very similar to the one that transforms a context-free grammar into another grammar in Chomsky normal form. The following theorem shows that the transformation does not accelerate the convergence of Newton's method.

Theorem 4. *Let $\mathbf{f}(\mathbf{X})$ be a clean feasible scMSP such that $f_k(\mathbf{X}) = g(\mathbf{X}) + h(\mathbf{X})X_iX_j$ for some $1 \le i, j, k \le n$, where $g(\mathbf{X})$ and $h(\mathbf{X})$ are non-constant polynomials with non-negative coefficients. Let $\widetilde{\mathbf{f}}(\mathbf{X}, Y)$ be the MSP given by*

$$\begin{array}{rcl} \widetilde{f_l}(\mathbf{X}, Y) &=& f_l(\mathbf{X}) \textit{ for every } l \in \{1, \ldots, k-1\} \\ \widetilde{f_k}(\mathbf{X}, Y) &=& g(\mathbf{X}) + h(\mathbf{X})Y \\ \widetilde{f_l}(\mathbf{X}, Y) &=& f_l(\mathbf{X}) \textit{ for every } l \in \{k+1, \ldots, n\} \\ \widetilde{f}_{n+1}(\mathbf{X}, Y) &=& X_iX_j. \end{array}$$

Then the function $b : \mathbb{R}^n \to \mathbb{R}^{n+1}$ given by $b(\mathbf{x}) = (x_1, \ldots, x_n, x_ix_j)^\top$ is a bijection between the set of fixed points of $\mathbf{f}(\mathbf{X})$ and $\widetilde{\mathbf{f}}(\mathbf{X}, Y)$. Moreover, $\widetilde{\boldsymbol{\nu}}^{(k)} \le (\nu_1^{(k)}, \ldots, \nu_n^{(k)}, \nu_i^{(k)}\nu_j^{(k)})^\top$ for all $k \in \mathbb{N}$, where $\widetilde{\boldsymbol{\nu}}^{(k)}$ and $\boldsymbol{\nu}^{(k)}$ are the Newton sequences of $\widetilde{\mathbf{f}}$ and $\mathbf{f}$, respectively.

A proof of this theorem can be found in the appendix.

Since we want to characterize the worst-case behavior of Newton's method, it will therefore suffice to consider quadratic systems in the following sections.

3.2 Properties of Newton's Method for Quadratic MSPs

In the following we assume that $\mathbf{f}(\mathbf{X})$ is a quadratic, clean, and feasible scMSP, and use the following notations.

Notation 1. Let $\mathbf{f}(\mathbf{X}) := B(\mathbf{X}, \mathbf{X}) + L\mathbf{X} + \mathbf{c}$, where $\mathbf{c} = \mathbf{f}(\mathbf{0})$, $L = \mathbf{f}'(\mathbf{0})$, and $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a symmetric bilinear map with $B(\mathbf{X}, \mathbf{X}) = \mathbf{f}(\mathbf{X}) - L\mathbf{X} - c$. We set $B(\mathbf{X})\mathbf{Y} := B(\mathbf{X}, \mathbf{Y})$. Hence, $B(\mathbf{X})\mathbf{Y} = B(\mathbf{Y})\mathbf{X}$, and, in particular, $\mathbf{f}'(\mathbf{X}) = 2B(\mathbf{X}) + L$.

If the matrix $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))$ is non-singular, it is well known that Newton's method converges quadratically (see e.g. [22]), and so Theorem 3 holds. So we only need to consider the case in which $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))$ is singular. However, since most of the proof of the quadratic convergence result is reused in the singular case, we actually consider both cases.

Lemma 2. *For $\mathbf{0} \le \mathbf{x} \prec \mu\mathbf{f}$ we have*

$$\mu\mathbf{f} - \mathcal{N}_{\mathbf{f}}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})^* B(\mu\mathbf{f} - \mathbf{x}, \mu\mathbf{f} - \mathbf{x}).$$

Proof. Set $\mathbf{d} := \mu\mathbf{f} - \mathbf{x}$.

$$\begin{array}{cl} & \mu\mathbf{f} - \mathcal{N}_{\mathbf{f}}(\mathbf{x}) \\ = & \mu\mathbf{f} - \mathbf{x} - \mathbf{f}'(\mathbf{x})^*(\mathbf{f}(\mathbf{x}) - \mathbf{x}) \\ = & \mathbf{d} - \mathbf{f}'(\mathbf{x})^*(\mathbf{f}(\mu\mathbf{f} - \mathbf{d}) - \mu\mathbf{f} + \mathbf{d}) \\ & \quad (\text{since } \mathbf{d} = \mu\mathbf{f} - \mathbf{x}) \\ = & \mathbf{d} - \mathbf{f}'(\mathbf{x})^*(\mathbf{f}(\mu\mathbf{f}) - \mathbf{f}'(\mu\mathbf{f})\mathbf{d} + B(\mathbf{d})\mathbf{d} - \mu\mathbf{f} + \mathbf{d}) \\ & \quad (\text{since } \mathbf{f}(\mathbf{X}) = B(\mathbf{X})\mathbf{X} + L\mathbf{X} + \mathbf{c} \text{ and } B(\mathbf{X}) \text{ linear}) \\ = & \mathbf{d} - \mathbf{f}'(\mathbf{x})^*((\mathrm{Id} - \mathbf{f}'(\mathbf{x}))\mathbf{d} - B(\mathbf{d})\mathbf{d}) \\ & \quad (\text{since } \mathbf{f}'(\mu\mathbf{f}) = \mathbf{f}'(\mathbf{x}) + 2B(\mathbf{d})) \\ = & \mathbf{f}'(\mathbf{x})^* B(\mathbf{d})\mathbf{d} \\ & \quad (\text{since } \mathbf{f}'(\mathbf{x})^*(\mathrm{Id} - \mathbf{f}'(\mathbf{x})) = \mathrm{Id}). \quad \square \end{array}$$

Since B is a symmetric bilinear map, we immediately obtain the following proposition (see also e.g. [22]).

Proposition 2 (Quadratic convergence). *If $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))^{-1}$ exists, then there exists a constant $c \in \mathbb{R}_{>0}$ such that $\left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(k+1)}\right\| \le c \left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(k)}\right\|^2$, and so the Newton sequence finally converges quadratically to $\mu\mathbf{f}$. In particular, if $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))^{-1}$ exists then Theorem 3 holds.*

A proof of Proposition 2 is given in the appendix.

Assume now that $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))$ is singular. In this case, $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))$ has a non-trivial *kernel*.

Definition 4. Let K denote the kernel of $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))$, i.e. $K = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{f}'(\mu\mathbf{f})\mathbf{v} = \mathbf{v}\}$.

A first easy-to-prove but important observation is that Newton's method converges linearly *if it starts at a point in* $\mu\mathbf{f} + K$. This is proved in the next lemma.

Lemma 3. *Assume $\mathbf{x} \in K + \mu\mathbf{f}$ with $0 \le \mathbf{x} \prec \mu\mathbf{f}$. Then $\mu\mathbf{f} - \mathcal{N}(\mathbf{x}) = \frac{1}{2}(\mu\mathbf{f} - \mathbf{x})$.*

Proof. Set $\mathbf{d} = \mu\mathbf{f} - \mathbf{x} \succ \mathbf{0}$. By Lemma 2 we have $\mu\mathbf{f} - \mathcal{N}(\mathbf{x}) = (\mathrm{Id} - \mathbf{f}'(\mathbf{x}))^{-1}B(\mathbf{d})\mathbf{d}$. Since $\mathbf{0} \le \mathbf{x} \prec \mu\mathbf{f}$ we know that $(\mathrm{Id} - \mathbf{f}'(\mathbf{x}))^{-1}$ exists. Consider the following equation.

$$\begin{array}{cl} & (\mathrm{Id} - \mathbf{f}'(\mathbf{x}))\mathbf{d} \\ = & (\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}) + 2B(\mathbf{d}))\mathbf{d} \\ & (\text{because } \mathbf{f}'(\mathbf{x}) = 2B(\mathbf{x}) + L = \mathbf{f}'(\mu\mathbf{f}) - 2B(\mu\mathbf{f} - \mathbf{x})) \\ = & 2B(\mathbf{d})\mathbf{d} \\ & (\text{because } \mathbf{d} \in K \Rightarrow (\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))\mathbf{d} = \mathbf{0}). \end{array}$$

Therefore $\mu\mathbf{f} - \mathcal{N}(\mathbf{x}) = (\mathrm{Id} - \mathbf{f}'(\mathbf{x}))^{-1}B(\mathbf{d})\mathbf{d} = \frac{1}{2}\mathbf{d}$. $\square$

At this point we make crucial use of the monotonicity of Newton's method on scMSPs (Lemma 1) to prove the following sufficient condition for linear convergence.

Proposition 3. *If there exists a $\mathbf{d} \in K$ with $\mathbf{d} \succ \mathbf{0}$, then Theorem 3 holds.*

Proof. Let $\mathbf{x} = \mu\mathbf{f} - \mathbf{d}$, so $\mathbf{x} \in \mu\mathbf{f} + K$. As K is a vector space, we can assume w.l.o.g. that $\mathbf{0} \le \mathbf{x} \prec \mu\mathbf{f}$. Since $\boldsymbol{\nu}^{(k)}$ converges to $\mu\mathbf{f}$, there exists some $k_{\mathbf{f}} \in \mathbb{N}$ such that $\mathbf{x} \le \boldsymbol{\nu}^{(k_{\mathbf{f}})}$. We have

$$\begin{array}{rcll} \mu\mathbf{f} - \boldsymbol{\nu}^{(l+k_{\mathbf{f}})} &=& \mu\mathbf{f} - \mathcal{N}^l(\boldsymbol{\nu}^{(k_{\mathbf{f}})}) & \\ &\le& \mu\mathbf{f} - \mathcal{N}^l(\mathbf{x}) & (\text{Lemma 1}) \\ &=& 2^{-l}(\mu\mathbf{f} - \mathbf{x}) & (\text{Lemma 3}) \\ &\le& 2^{-l}\mu\mathbf{f} & (\mathbf{0} \le \mathbf{x} \prec \mu\mathbf{f}). \quad \square \end{array}$$

3.3 A Perturbation Method to Characterize the Kernel

In this subsection we show that the kernel K always contains some vector $\mathbf{d} \succ \mathbf{0}$. We start by showing that it suffices to find a vector $\mathbf{d} > \mathbf{0}$.

LEMMA 4. *Let $\mathbf{0} < \mathbf{d} \in K$ for a clean feasible scMSP. Then $\mathbf{d} \succ \mathbf{0}$.*

PROOF. Let $\mathbf{0} < \mathbf{d} \in K$. If $\mathbf{d} \succ \mathbf{0}$, we are done. Hence, w.l.o.g., we assume $d_1 = \ldots = d_s = 0$ and $d_{s+1}, \ldots, d_n > 0$. As we have $\mathbf{f}'(\mu\mathbf{f})\mathbf{d} = \mathbf{d}$, we get for $i \in \{1, \ldots, s\}$ that $f_i'(\mu\mathbf{f})\mathbf{d} = 0$. Hence $\frac{\partial f_i}{\partial X_k}(\mu\mathbf{f}) = 0$ for $k \in \{s+1, \ldots, n\}$. But as $\mathbf{0} \prec \mu\mathbf{f}$ this can only be if $\frac{\partial f_i}{\partial X_k}$ is the null polynomial. Therefore none of the variables $X_1, \ldots, X_s$ depends on any of the variables $X_{s+1}, \ldots, X_n$, contradicting our assumption that $\mathbf{f}$ is an scMSP. □

To get an intuition of why the kernel indeed contains a $\mathbf{d} > \mathbf{0}$, consider Fig. 1 (a). It shows the graph of the strongly connected 2-dimensional system $\mathbf{X} = \mathbf{f}(\mathbf{X})$ given by

$$\begin{aligned} X_1 &= \tfrac{1}{4}X_2^2 + \tfrac{1}{4}X_1X_2 + \tfrac{3}{16}X_1^2 + \tfrac{5}{16} \\ X_2 &= \tfrac{1}{8}X_2^2 + \tfrac{1}{4}X_1X_2 + \tfrac{5}{8}. \end{aligned}$$

In this example, $\mu\mathbf{f} = (1,1)^\top$ and the kernel K is the vector space spanned by $(2,1)^\top \succ \mathbf{0}$. The figure also illustrates that, since K is the kernel of $\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f})$, the straight line $\mu\mathbf{f} + K$ is a tangent in $\mu\mathbf{f}$ to the quadrics corresponding to $X_1 = f_1(\mathbf{X})$ and $X_2 = f_2(\mathbf{X})$. So K contains a vector $\mathbf{d} > \mathbf{0}$ iff the tangent has a positive slope. (In higher dimensions, iff the tangent space is slanted towards non-negative coordinates.)

Consider what happens to the point $\mu\mathbf{f}$ when the coefficients of f_2 are slightly decreased by some ε (Fig.1 (b) shows the cases $\varepsilon = 1/2, 1/4, 1/16$). By monotonicity (Proposition 1) we know that no component of $\mu\mathbf{f}$ can increase. The figure indeed suggests that when the quadric $X_2 = f_2(\mathbf{X})$ is scaled down, then the least fixed point $\mu\mathbf{f}_\varepsilon$ of the perturbed system $\mathbf{f}_\varepsilon$ "slides down" along the curve $X_1 = f_1(\mathbf{X})$. For $\varepsilon \to 0$, the vector $\mathbf{u}_\varepsilon := \frac{\mu\mathbf{f} - \mu\mathbf{f}_\varepsilon}{\|\mu\mathbf{f} - \mu\mathbf{f}_\varepsilon\|}$ intuitively converges to the slope of the tangent, and so to a vector $\mathbf{d}$ with $\mathbf{0} < \mathbf{d} \in K$, as desired.

However, proving that $\mathbf{u}_\varepsilon$ converges to a vector in the tangent space turns out to be surprisingly hard, and so we use an indirect but technically easier method. Before explaining it, it is convenient to introduce some notations.

Notation 2.

- Let $\mathbf{f}_\varepsilon(\mathbf{X})$ denote the system $\mathbf{f}_\varepsilon(\mathbf{X}) = \mathbf{f}(\mathbf{X}) - \varepsilon \dfrac{f_n(\mathbf{X})}{(\mu\mathbf{f})_n}\mathbf{e}_n$, where we assume $0 \leq \varepsilon < (\mu\mathbf{f})_n$. Notice that $\mathbf{f}_\varepsilon$ is still a clean feasible scMSP with $\|\mathbf{f} - \mathbf{f}_\varepsilon\| = \left\|\varepsilon \frac{f_n}{(\mu\mathbf{f})_n}\right\| \leq \varepsilon$ on $[\mathbf{0}, \mu\mathbf{f}]^n$.
- Let $\mathbf{g}(\mathbf{X})$ denote the system containing the first $n-1$ equations of $\mathbf{f}(\mathbf{X}) - \mathbf{X}$.
  Notice that the kernel of $\mathbf{g}'(\mu\mathbf{f})$ contains the vectors that are tangent to the first $n-1$ quadrics.
- Let $\mathbf{u}_\varepsilon = \frac{\mu\mathbf{f} - \mu\mathbf{f}_\varepsilon}{\|\mu\mathbf{f} - \mu\mathbf{f}_\varepsilon\|}$. Notice that $\mathbf{u}_\varepsilon > \mathbf{0}$.

The proof proceeds in three steps.

Step (1).

We show that $\|\mathbf{g}'(\mu\mathbf{f})\mathbf{u}_\varepsilon\| \xrightarrow{\varepsilon \to 0} 0$. Of course, this does not prove that $\mathbf{u}_\varepsilon$ converges to some vector in the kernel of $\mathbf{g}'(\mu\mathbf{f})$, i.e., to a tangent to the first $n-1$ quadrics; for this we would have to show additionally that $\lim_{\varepsilon\to 0} \mathbf{u}_\varepsilon$ exists in general.

LEMMA 5. $\|\mathbf{g}'(\mu\mathbf{f})\mathbf{u}_\varepsilon\| \xrightarrow{\varepsilon \to 0} 0$.

A proof is given in the appendix.

Step (2).

We prove the existence of some vector $\mathbf{v}_n \geq \mathbf{0}$ in the kernel of $\mathbf{g}'(\mu\mathbf{f})$. For this, we need the following lemma.

LEMMA 6. *Let U, V be compact subsets of $\mathbb{R}^n$, and let dist$(U,V) = \inf_{\mathbf{u}\in U, \mathbf{v}\in V} \|\mathbf{u} - \mathbf{v}\|$.*
If dist$(U,V) = 0$ then $U \cap V \neq \emptyset$.

Lemma 6 can be easily shown using standard arguments regarding compact sets of $\mathbb{R}^n$, see the appendix for a proof.

LEMMA 7. *Let $\tilde{K}$ denote the kernel of $\mathbf{g}'(\mu\mathbf{f})$, i.e. $\tilde{K} = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{g}'(\mu\mathbf{f})\mathbf{v} = \mathbf{0}\}$. Hence, K, the kernel of $\mathbf{f}'(\mu\mathbf{f}) - \mu\mathbf{f}$, is a subspace of $\tilde{K}$. Further, let S denote the unit sphere $\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}$ and let $S_{\geq \mathbf{0}} = S \cap \mathbb{R}^n_{\geq 0}$.*
Then $\tilde{K} \cap S_{\geq 0} \neq \emptyset$.

Proof Sketch (see appendix for a full proof). Let P be the orthogonal projector that projects $\mathbb{R}^n$ onto $\tilde{K}$. From Lemma 5 it follows for the complementary projector $\mathrm{Id} - P$ that $(\mathrm{Id} - P)\mathbf{u}_\varepsilon \xrightarrow{\varepsilon \to 0} \mathbf{0}$. Let $B_{\tilde{K}} = \{\mathbf{v} \in \tilde{K} \mid \|\mathbf{v}\| \leq 1\}$. Then

$$\mathrm{dist}(S_{\geq 0}, B_{\tilde{K}}) \leq \inf_{0<\varepsilon<(\mu\mathbf{f})_n} \|\mathbf{u}_\varepsilon - P\mathbf{u}_\varepsilon\| = 0.$$

As both $B_{\tilde{K}}$ and $S_{\geq 0}$ are compact, we may apply Lemma 6 to conclude $\tilde{K} \cap S_{\geq 0} \neq \emptyset$. □

Step (3).

Step (1) and (2) prove that for $i = n$ there exists a non-negative vector $\mathbf{v}_i$ tangent to all quadrics of $\mathbf{f}(\mathbf{X}) - \mathbf{X}$ but the i-th. By reordering the quadrics of $\mathbf{f}$, we immediately obtain that the result holds not only for $i = n$ but for every $i \in [1, n]$. We show that this implies the existence of a non-negative vector tangent to *all* quadrics.

LEMMA 8. *If* $\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f})$ *is singular, then the kernel of* $\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f})$ *contains a vector* $\mathbf{d} > \mathbf{0}$.

PROOF. Let $\mathbf{v}_i$ be the vector obtained from Step (3) above. Then the image of $\mathbf{v}_i$ under $\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f})$ is a multiple of $\mathbf{e}_i$. Hence, if none of the $\mathbf{v}_i$ is mapped onto $\mathbf{0}$ by $\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f})$, then their image is a set of n linearly independent vectors. But this would contradict the assumption that $\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f})$ is singular. □

By combining Proposition 3, Lemma 4 and Lemma 8 we obtain a proof of Theorem 3.

4. A LOWER BOUND FOR THE THRESHOLD

In the previous section we showed a strong asymptotic convergence result for the case in which the MSP $\mathbf{f}$ consists

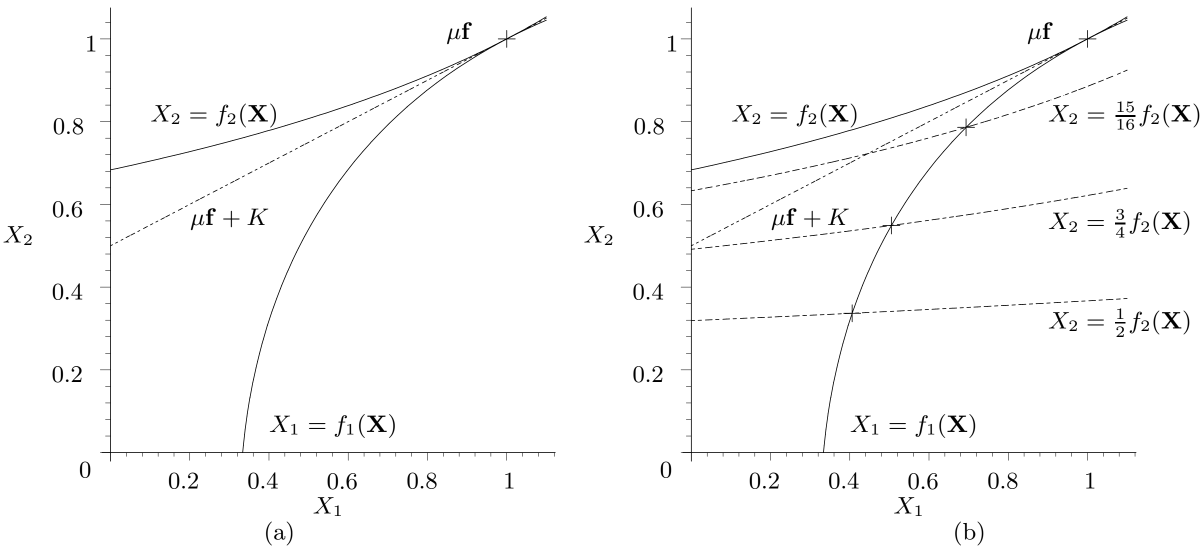

Figure 1: The graph of a 2-dimensional equation system $\mathbf{X} = \mathbf{f}(\mathbf{X})$.

of one single SCC: we eventually gain one bit per iteration. In this section we consider – in some sense – the opposite case: we give concrete systems which consist of n SCCs and for which the decomposed Newton's method converges very badly *initially*. More precisely, we give a family $\{\mathbf{f}_{(n)} \mid n \geq 1\}$ of MSPs with n variables, such that $\Omega(2^n)$ iterations of the decomposed Newton's method are needed for the first valid bit. Consider the following system.

$$\mathbf{X} = \mathbf{f}(\mathbf{X}) = \begin{pmatrix} \frac{1}{2} + \frac{1}{2}X_1^2 \\ \frac{1}{4}X_1^2 + \frac{1}{2}X_1X_2 + \frac{1}{4}X_2^2 \\ \vdots \\ \frac{1}{4}X_{n-1}^2 + \frac{1}{2}X_{n-1}X_n + \frac{1}{4}X_n^2 \end{pmatrix} \tag{1}$$

The only solution of (1) is $\mu\mathbf{f} = (1, \ldots, 1)^\top$. The MSP $\mathbf{f}$ has n SCCs. The decomposed Newton's method starts with the bottom SCC $X_1 = \frac{1}{2} + \frac{1}{2}X_1^2$. The associated Newton sequence is $0, \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \ldots$, i.e., after 2^{n-1} iterations we have an approximation $x_1 = 1 - 2^{-2^{n-1}}$, so the *error* is $d = 2^{-2^{n-1}}$. Now, the decomposed Newton's method continues with the system $X_2 = \frac{1}{4}(1-d)^2 + \frac{1}{2}(1-d)X_2 + \frac{1}{4}X_2^2$. The least solution of this system is $x_2 = 1 + d - 2\sqrt{d}$. Since Newton's method converges to x_2 from below, we know that the error in the Newton estimate will be *at least* $2\sqrt{d} - d \geq \sqrt{d} = 2^{-2^{n-2}}$. Similarly, the error is amplified in all components until the error in x_n is at least $2^{-2^0} = \frac{1}{2}$. In conclusion, there is at most one valid bit in x_n after 2^{n-1} iterations, or, in terms of the introduction, $\mathit{iter}_{\mathbf{f}}(1) \geq 2^{n-1}$. Since the size of the coefficients of the MSP $\mathbf{f}$ does not grow with n, we have $\mathit{Thr}(m) \in 2^{\Omega(m)}$. This refutes a conjecture of [12] (Conjecture 26).

5. SUMMARY AND FUTURE WORK

We have studied several aspects of the convergence of the decomposed Newton's method. On the one hand, in our main theorem we have proved linear asymptotical convergence for strongly connected MSPs at the rate of 1 bit per iteration. On the other hand, we have refuted a conjecture of Etessami and Yannakakis [12] by exhibiting (non-strongly-connected) MSPs for which the decomposed Newton's method fails to produce the first valid bit after 2^{n-1} iterations.

The proof of our linear convergence result relies heavily on the assumption that the MSP is strongly connected. Etessami and Yannakakis [10] showed that having a strongly connected MSP is a sufficient condition for Newton's method to be well-defined. It can be shown that Newton's method for the example in Section 4 is well-defined, but for $n > 1$ one gains less than 1 bit per iteration asymptotically. We believe that SCCs play a crucial role for the convergence speed of Newton's method for MSPs.

Our paper raises several questions for future work:

- Determine the convergence speed of the decomposed Newton's method for arbitrary MSPs.
  We conjecture that the asymptotic convergence is still linear. However, it is easy to deduce from the examples in Section 4 that if the convergence is linear then the rate must be smaller than 1 bit per iteration. We do not know how the rate degrades with an increasing number of SCCs, with n, and/or with the overall size of the system.

- Give an upper bound for the threshold in general MSPs. Some (non-systematic) experiments suggest that the $2^{\Omega(m)}$ lower bound could in fact become a $2^{\Theta(m)}$ tight bound.

- Give an upper bound for the threshold in strongly connected MSPs.
  One can see from the example of Section 4 that the threshold cannot be independent of the scMSP's size, but, again, some experiments suggest that the threshold does not grow as fast with the size as in the general case.

The ultimate goal of our work is to give upper bounds for $\mathit{iter}_{\mathbf{f}}(i)$ in terms of easily observable properties of the MSP $\mathbf{f}$. While this goal seems out of reach for general systems of equations, we think that the very special shape of MSPs gives us good success chances.

6. REFERENCES

[1] 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.

[2] D.W. Decker and C.T. Kelley. Newton's method at singular points I. *SIAM Journal on Numerical Analysis*, 17(1):66–70, 1980.

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

[4] R. Durbin, S.R. Eddy, A. Krogh, and G.J. 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*. 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.R. 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.R. 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] A. Griewank and M.R. Osborne. Newton's method for singular problems when the dimension of the null space is > 1. *SIAM Journal on Numerical Analysis*, 18(1):145–149, 1981.

[17] C.T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. SIAM, 1995.

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

[19] 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.

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

[21] J.M. Ortega. *Numerical Analysis: A Second Course*. Academic Press, New York, 1972.

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

[23] F.A. Potra and V. Pták. Sharp error bounds for Newton's process. *Numerische Mathematik*, 34(1):63–72, 1980.

[24] G.W. Reddien. On Newton's method for singular problems. *SIAM Journal on Numerical Analysis*, 15:993–996, 1978.

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

APPENDIX

A. PROOFS

A.1 Proof of Proposition 2

Now we prove Proposition 2 assuring quadratic convergence in the nonsingular case. It is restated here.

PROPOSITION 2.
If $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))^{-1}$ *exists, then there exists a constant* $c \in \mathbb{R}_{>0}$ *such that* $\left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(k+1)}\right\| \leq c \left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(k)}\right\|^2$, *and so the Newton sequence finally converges quadratically to* $\mu\mathbf{f}$*. In particular, if* $(\mathrm{Id} - \mathbf{f}'(\mu\mathbf{f}))^{-1}$ *exists then Theorem 3 holds.*

PROOF. Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. Then $g(\mathbf{X}) := \left\|(\mathrm{Id} - \mathbf{f}'(\mathbf{X}))^{-1}\right\|$ is a continuous map from $[\mathbf{0}, \mu\mathbf{f}] \subseteq \mathbb{R}^n$ to $\mathbb{R}$. As $[\mathbf{0}, \mu\mathbf{f}]^n$ is compact and $\mathbb{R}$ is Hausdorff, the image $g([\mathbf{0}, \mu\mathbf{f}])$ is compact, too. Hence, $C := \max\{g(\mathbf{x}) | \mathbf{x} \in [\mathbf{0}, \mu\mathbf{f}]\}$ exists. Similarly, we find a constant c with $\|B(\mathbf{x})\mathbf{x}\| \leq c\,\|\mathbf{x}\|^2$ for all $\mathbf{x} \in [\mathbf{0}, \mu\mathbf{f}]$. Hence, we have

$$\|\mu\mathbf{f} - \mathcal{N}(\mathbf{x})\| \leq Cc\,\|\mu\mathbf{f} - \mathbf{x}\|^2$$

for all $\mathbf{x} \in [\mathbf{0}, \mu\mathbf{f}]$.

To prove that Theorem 3 holds in this case, notice that, since $\boldsymbol{\nu}^{(k)}$ converges to $\mu\mathbf{f}$, there is a $k_{\mathbf{f}} \in \mathbb{N}$ such that $\left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(k_{\mathbf{f}})}\right\| \leq \frac{1}{2Cc}$. Then $Cc \left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(k_{\mathbf{f}})}\right\|^2 \leq \frac{1}{2} \left\|\mu\mathbf{f} - \boldsymbol{\nu}^{(k_{\mathbf{f}})}\right\|$ and Theorem 3 follows by induction on l. □

A.2 Proof of Theorem 4

In this appendix we prove Theorem 4, which allows us to focus on quadratic systems for a worst case analysis. First we prove two technical lemmata.

The following lemma is a version of Taylor's theorem for MSPs.

Lemma 9 (Taylor). *Let $\mathbf{x}, \mathbf{d} \in \mathbb{R}^n_{\geq 0}$ and $\mathbf{f}(\mathbf{X})$ be an MSP. Then*

$$\mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x})\mathbf{d} \leq \mathbf{f}(\mathbf{x}+\mathbf{d}) \leq \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x}+\mathbf{d})\mathbf{d}.$$

Proof. It suffices to show this for a multi-variate polynomial $f(\mathbf{X})$ with non-negative coefficients. Consider $g(t) = f(\mathbf{x}+t\mathbf{d})$. We then have

$$\begin{aligned} f(\mathbf{x}+\mathbf{d}) = g(1) \quad &= g(0) + \int_0^1 g'(s)ds \\ &= f(\mathbf{x}) + \int_0^1 f'(\mathbf{x}+s\mathbf{d})\mathbf{d}ds. \end{aligned}$$

The result follows as $f'(\mathbf{x}) \leq f'(\mathbf{x}+s\mathbf{d}) \leq f'(\mathbf{x}+\mathbf{d})$ for $s \in [0,1]$. $\square$

The following lemma can already be found in [5]. It is used in several of our proofs.

Lemma 10. *Let $\mathbf{f}(\mathbf{X})$ be a clean feasible scMSP. We have $\boldsymbol{\kappa}^{(k)} \leq \boldsymbol{\nu}^{(k)} \leq \mathbf{f}(\boldsymbol{\nu}^{(k)})$ for all $k \in \mathbb{N}$.*

Proof. For $k = 0$ this holds by definition. As $\boldsymbol{\nu}^{(k)} \leq \boldsymbol{\nu}^{(k+1)}$ and $\boldsymbol{\kappa}^{(k)} \leq \boldsymbol{\nu}^{(k)}$ by induction, we get $\mathbf{f}(\boldsymbol{\nu}^{(k+1)}) \geq \mathbf{f}(\boldsymbol{\nu}^{(k)}) \geq \mathbf{f}(\boldsymbol{\kappa}^{(k)}) = \boldsymbol{\kappa}^{(k+1)}$. By induction we have $\mathbf{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)} \geq \mathbf{0}$, and so $\boldsymbol{\delta}_{\mathbf{f}}(\boldsymbol{\nu}^{(k)}) \overset{\text{def}}{=} \mathbf{f}'(\boldsymbol{\nu}^{(k)})^*(\mathbf{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \geq \mathbf{0}$. Hence, we may apply Lemma 9 to get:

$$\begin{aligned} \mathbf{f}(\boldsymbol{\nu}^{(k+1)}) \quad &= \quad \mathbf{f}(\boldsymbol{\nu}^{(k)} + \boldsymbol{\delta}_{\mathbf{f}}(\boldsymbol{\nu}^{(k)})) \\ &\geq \quad \mathbf{f}(\boldsymbol{\nu}^{(k)}) + \mathbf{f}'(\boldsymbol{\nu}^{(k)})\boldsymbol{\delta}_{\mathbf{f}}(\boldsymbol{\nu}^{(k)}) \\ &= \quad \boldsymbol{\nu}^{(k)} + (\mathbf{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \\ &\qquad + \mathbf{f}'(\boldsymbol{\nu}^{(k)})\mathbf{f}'(\boldsymbol{\nu}^{(k)})^*(\mathbf{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \\ &= \quad \boldsymbol{\nu}^{(k+1)}. \quad \square \end{aligned}$$

Now we can prove Theorem 4.

Theorem 4. *Let $\mathbf{f}(\mathbf{X})$ be a clean feasible scMSP such that $f_k(\mathbf{X}) = g(\mathbf{X}) + h(\mathbf{X})X_iX_j$ for some $1 \leq i,j,k \leq n$, where $g(\mathbf{X})$ and $h(\mathbf{X})$ are non-constant polynomials with non-negative coefficients. Let $\widetilde{\mathbf{f}}(\mathbf{X}, Y)$ be the MSP given by*

$$\begin{aligned} \widetilde{f}_l(\mathbf{X}, Y) \quad &= \quad f_l(\mathbf{X}) \text{ for every } l \in \{1, \ldots, k-1\} \\ \widetilde{f}_k(\mathbf{X}, Y) \quad &= \quad g(\mathbf{X}) + h(\mathbf{X})Y \\ \widetilde{f}_l(\mathbf{X}, Y) \quad &= \quad f_l(\mathbf{X}) \text{ for every } l \in \{k+1, \ldots, n\} \\ \widetilde{f}_{n+1}(\mathbf{X}, Y) \quad &= \quad X_iX_j. \end{aligned}$$

Then the function $b : \mathbb{R}^n \to \mathbb{R}^{n+1}$ given by $b(\mathbf{x}) = (x_1, \ldots, x_n, x_ix_j)^\top$ is a bijection between the set of fixed points of $\mathbf{f}(\mathbf{X})$ and $\widetilde{\mathbf{f}}(\mathbf{X}, Y)$. Moreover, $\widetilde{\boldsymbol{\nu}}^{(k)} \leq (\nu_1^{(k)}, \ldots, \nu_n^{(k)}, \nu_i^{(k)}\nu_j^{(k)})^\top$ for all $k \in \mathbb{N}$, where $\widetilde{\boldsymbol{\nu}}^{(k)}$ and $\boldsymbol{\nu}^{(k)}$ are the Newton sequences of $\widetilde{\mathbf{f}}$ and $\mathbf{f}$, respectively.

Proof. W.l.o.g. we may assume that $k = 1$. We first show the claim regarding b: if $\mathbf{x}^*$ is a fixed point of $\mathbf{f}$, then $b(\mathbf{x}^*) = (\mathbf{x}^*, x_i^* \cdot x_j^*)$ is a fixed point of $\widetilde{\mathbf{f}}$. Conversely, if $(\mathbf{x}^*, y^*)$ is a fixed point of $\widetilde{\mathbf{f}}$, then we have $y^* = x_i^* \cdot x_j^*$ implying that $\mathbf{x}^*$ is a fixed point of $\mathbf{f}$ and $b(\mathbf{x}^*) = (\mathbf{x}^*, y^*)$. Therefore, the least fixed point $\mu\mathbf{f}$ of $\mathbf{f}$ determines $\mu\widetilde{\mathbf{f}}$, and vice versa.

Now we show that the Newton sequence of $\mathbf{f}$ converges at least as quickly as the Newton sequence of $\widetilde{\mathbf{f}}$ does.

Again, let $\boldsymbol{\delta}_{\mathbf{f}}(\boldsymbol{\nu}^{(k)}) \overset{\text{def}}{=} \mathcal{N}_{\mathbf{f}}(\mathbf{X}) - \mathbf{X} = \mathbf{f}'(\boldsymbol{\nu}^{(k)})^*(\mathbf{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)})$ be the Newton update w.r.t. $\mathbf{f}$. Similarly, let $\widetilde{\boldsymbol{\delta}} \overset{\text{def}}{=} \boldsymbol{\delta}_{\widetilde{\mathbf{f}}}$ be the Newton update for $\widetilde{\mathbf{f}}$. For $\mathbf{x} \in \mathbb{R}^{n+1}$ an $(n+1)$-dimensional vector, we let $\mathbf{x}_{[1,n]}$ denote its restriction to the n first components, i.e. $\mathbf{x}_{[1,n]} = (x_1, \ldots, x_n)^\top$. Then $\widetilde{\boldsymbol{\delta}}$ is the unique solution of this equation system:

$$\begin{pmatrix} \mathrm{Id} - \mathbf{f}'(\mathbf{X}) - \frac{\partial(Y - X_iX_j)h}{\partial\mathbf{X}}\mathbf{e}_1 & -h(\mathbf{X})\mathbf{e}_1 \\ -\frac{\partial X_iX_j}{\partial\mathbf{X}} & 1 \end{pmatrix} \cdot \begin{pmatrix} \widetilde{\boldsymbol{\delta}}_{[1,n]} \\ \widetilde{\delta}_{n+1} \end{pmatrix} = \begin{pmatrix} \mathbf{f}(\mathbf{X}) - \mathbf{X} \\ X_iX_j - Y \end{pmatrix} + \begin{pmatrix} (Y - X_iX_j) \cdot h(\mathbf{X}) \\ 0 \end{pmatrix}.$$

We may solve the last row for $\widetilde{\delta}_{n+1}$ resulting in

$$\widetilde{\delta}_{n+1} = \frac{\partial X_iX_j}{\partial\mathbf{X}}\widetilde{\boldsymbol{\delta}}_{[1,n]} + X_iX_j - Y.$$

Substituting this into the first n equations, one gets

$$\begin{aligned} &\left(\mathrm{Id} - \mathbf{f}'(\mathbf{X}) - \tfrac{\partial(Y - X_iX_j)h}{\partial\mathbf{X}}\mathbf{e}_1\right)\widetilde{\boldsymbol{\delta}}_{[1,n]} \\ &\quad - h(\mathbf{X}) \cdot \left(\tfrac{\partial X_iX_j}{\partial\mathbf{X}}\widetilde{\boldsymbol{\delta}}_{[1,n]} + X_iX_j - Y\right)\mathbf{e}_1 \\ = \quad &\left(\mathrm{Id} - \mathbf{f}'(\mathbf{X}) - (\tfrac{\partial(Y - X_iX_j)h}{\partial\mathbf{X}} - \tfrac{\partial(Y - X_iX_j)}{\partial\mathbf{X}}h)\mathbf{e}_1\right)\widetilde{\boldsymbol{\delta}}_{[1,n]} \\ &\quad - h(\mathbf{X})\,(X_iX_j - Y)\,\mathbf{e}_1 \\ = \quad &\left(\mathrm{Id} - \mathbf{f}'(\mathbf{X}) + (X_iX_j - Y)\tfrac{\partial h}{\partial\mathbf{X}}\mathbf{e}_1\right)\widetilde{\boldsymbol{\delta}}_{[1,n]} \\ &\quad - h(\mathbf{X})(X_iX_j - Y)\mathbf{e}_1 \\ = \quad &\mathbf{f}(\mathbf{X}) - \mathbf{X} + (Y - X_iX_j)h(\mathbf{X})\mathbf{e}_1, \end{aligned}$$

or, after adding $h(\mathbf{X})(X_iX_j - Y)\mathbf{e}_1$ on both sides and then multiplying by $(\mathrm{Id} - \mathbf{f}'(\mathbf{X}))^{-1}$ from the left:

$$\left(\mathrm{Id} + (X_iX_j - Y)(\mathrm{Id} - \mathbf{f}'(\mathbf{X}))^{-1}\frac{\partial h}{\partial\mathbf{X}}\mathbf{e}_1\right)\widetilde{\boldsymbol{\delta}}_{[1,n]} = \boldsymbol{\delta}_{\mathbf{f}}.$$

Note that the update $\widetilde{\boldsymbol{\delta}}_{[1,n]}$ becomes $\boldsymbol{\delta}_{\mathbf{f}}$ if $X_iX_j = Y$. Now, we proceed by induction on k to show $\widetilde{\boldsymbol{\nu}}_{[1,n]}^{(k)} \leq \boldsymbol{\nu}^{(k)}$, where $\widetilde{\boldsymbol{\nu}}^{(k)}$ is the Newton sequence for $\widetilde{\mathbf{f}}$. By definition of the Newton sequence this is true for $k = 0$. For the step we have:

$$\begin{aligned} \widetilde{\boldsymbol{\nu}}_{[1,n]}^{(k+1)} \quad &= \quad \mathcal{N}_{\widetilde{\mathbf{f}}}(\widetilde{\boldsymbol{\nu}}^{(k)})_{[1,n]} \\ &\overset{*}{\leq} \quad \mathcal{N}_{\widetilde{\mathbf{f}}}(\tilde{\mathbf{f}}(\widetilde{\boldsymbol{\nu}}^{(k)}))_{[1,n]} \\ &= \quad \mathcal{N}_{\widetilde{\mathbf{f}}}((\widetilde{\boldsymbol{\nu}}_{[1,n]}^{(k)}, \widetilde{\nu}_i^{(k)} \cdot \widetilde{\nu}_j^{(k)}))_{[1,n]} \\ &= \quad \widetilde{\boldsymbol{\nu}}_{[1,n]}^{(k)} + \widetilde{\boldsymbol{\delta}}((\widetilde{\boldsymbol{\nu}}_{[1,n]}^{(k)}, \widetilde{\nu}_i^{(k)} \cdot \widetilde{\nu}_j^{(k)}))_{[1,n]} \\ &= \quad \widetilde{\boldsymbol{\nu}}_{[1,n]}^{(k)} + \boldsymbol{\delta}_{\mathbf{f}}(\widetilde{\boldsymbol{\nu}}_{[1,n]}^{(k)}) \\ &= \quad \mathcal{N}_{\mathbf{f}}(\widetilde{\boldsymbol{\nu}}_{[1,n]}^{(k)}) \leq \mathcal{N}_{\mathbf{f}}(\boldsymbol{\nu}^{(k)}) \\ &= \quad \boldsymbol{\nu}^{(k+1)} \end{aligned} \quad .$$

At inequation $(*)$ we used the monotonicity of $\mathcal{N}_{\widetilde{\mathbf{f}}}$ combined with Lemma 10, which states $\widetilde{\boldsymbol{\nu}}^{(k)} \leq \widetilde{\mathbf{f}}(\widetilde{\boldsymbol{\nu}}^{(k)})$, hence in particular $\widetilde{\nu}_{n+1}^{(k)} \leq \widetilde{\nu}_i^{(k)}\widetilde{\nu}_j^{(k)}$. $\square$

A.3 Proof of Lemma 5

We first prove the following lemma, which assures some technical properties of $\mathbf{f}_\varepsilon$ as introduced in Notation 2.

Lemma 11. *The least non-negative fixed point $\mu\mathbf{f}_\varepsilon$ of $\mathbf{f}_\varepsilon$ exists. Further, for $0 \leq \varepsilon \leq \varepsilon' < (\mu\mathbf{f})_n$ we have $\mu\mathbf{f} \geq \mu\mathbf{f}_\varepsilon \geq \mu\mathbf{f}_{\varepsilon'}$. Moreover, $\mu\mathbf{f}_\varepsilon$ is continuous in $\varepsilon = 0$, i.e. $\|\mu\mathbf{f} - \mu\mathbf{f}_\varepsilon\| \overset{\varepsilon \to 0}{\to} 0$.*

Proof. The first two claims are consequences of Proposition 1. For the continuity, consider any sequence $(\varepsilon_i)_{i\in\mathbb{N}}$

with $\varepsilon_i \searrow 0$. As $\mu\mathbf{f}_{\varepsilon_i} \leq \mu\mathbf{f}_{\varepsilon_{i+1}} \leq \mu\mathbf{f}$, the sequence $(\mu\mathbf{f}_{\varepsilon_i})_{i\in\mathbb{N}}$ has to converge to some $\tilde{\mu\mathbf{f}}$ less than or equal to $\mu\mathbf{f}$. But $\tilde{\mu\mathbf{f}}$ is a fixed point of $\mathbf{f}$, as:

$$\begin{aligned}\|\mu\mathbf{f}_\varepsilon - \mathbf{f}(\mu\mathbf{f}_\varepsilon)\| \quad &= \|\mu\mathbf{f}_\varepsilon - \mathbf{f}_\varepsilon(\mu\mathbf{f}_\varepsilon) - (\mathbf{f}(\mu\mathbf{f}_\varepsilon) - \mathbf{f}_\varepsilon(\mu\mathbf{f}_\varepsilon))\| \\ &= \|\mathbf{f}(\mu\mathbf{f}_\varepsilon) - \mathbf{f}_\varepsilon(\mu\mathbf{f}_\varepsilon)\| \leq \varepsilon.\end{aligned}$$

This shows that $\lim_{\varepsilon\searrow 0} \|\mu\mathbf{f}_\varepsilon - \mathbf{f}(\mu\mathbf{f}_\varepsilon)\| = 0$.
As $\|\mathbf{X} - \mathbf{f}(\mathbf{X})\|$ is continuous, we therefore have

$$0 = \lim_{\varepsilon\searrow 0} \|\mu\mathbf{f}_\varepsilon - \mathbf{f}(\mu\mathbf{f}_\varepsilon)\| = \left\|\tilde{\mu\mathbf{f}} - \mathbf{f}(\tilde{\mu\mathbf{f}})\right\|.$$

Hence, $\tilde{\mu\mathbf{f}} = \mu\mathbf{f}$, i.e. $\mu\mathbf{f}_\varepsilon$ is continuous in $\varepsilon = 0$. □

Now we can prove Lemma 5 which is restated here.

LEMMA 5. $\|\mathbf{g}'(\mu\mathbf{f})\mathbf{u}_\varepsilon\| \overset{\varepsilon\to 0}{\to} 0$.

PROOF. Let $\mathbf{g}(\mathbf{X}) = \tilde{B}(\mathbf{X})\mathbf{X} + \tilde{L}\mathbf{X} + \tilde{\mathbf{c}}$. We have $\mathbf{g}(\mu\mathbf{f}_\varepsilon) = \mathbf{0}$ for all ε by definition of $\mu\mathbf{f}_\varepsilon$. Hence,

$$\begin{aligned}\mathbf{0} \quad &= \mathbf{g}(\mu\mathbf{f}_\varepsilon) = \mathbf{g}(\mu\mathbf{f} - (\mu\mathbf{f} - \mu\mathbf{f}_\varepsilon)) \\ &= \underbrace{\mathbf{g}(\mu\mathbf{f})}_{=\mathbf{0}} - \mathbf{g}'(\mu\mathbf{f})(\mu\mathbf{f} - \mu\mathbf{f}_\varepsilon) + \tilde{B}(\mu\mathbf{f} - \mu\mathbf{f}_\varepsilon, \mu\mathbf{f} - \mu\mathbf{f}_\varepsilon).\end{aligned}$$

We therefore get

$$\begin{aligned}\|\mathbf{g}'(\mu\mathbf{f})\mathbf{u}_\varepsilon\| \quad &= \left\|\mathbf{g}'(\mu\mathbf{f})\tfrac{\mu\mathbf{f}-\mu\mathbf{f}_\varepsilon}{\|\mu\mathbf{f}-\mu\mathbf{f}_\varepsilon\|}\right\| = \tfrac{\left\|\tilde{B}(\mu\mathbf{f}-\mu\mathbf{f}_\varepsilon, \mu\mathbf{f}-\mu\mathbf{f}_\varepsilon)\right\|}{\|\mu\mathbf{f}-\mu\mathbf{f}_\varepsilon\|} \\ &\leq \max\nolimits_{\mathbf{x}\in[\mathbf{0},\mu\mathbf{f}]^n} \left\|\tilde{B}\right\| \tfrac{\|\mu\mathbf{f}-\mu\mathbf{f}_\varepsilon\|^2}{\|\mu\mathbf{f}-\mu\mathbf{f}_\varepsilon\|}\end{aligned}$$

and this approaches 0 for $\varepsilon \to 0$. □

A.4 Proof of Lemma 6

We now prove Lemma 6 which is restated here.

LEMMA 6. *Let U, V be compact subsets of $\mathbb{R}^n$, and let $dist(U,V) = \inf_{\mathbf{u}\in U,\mathbf{v}\in V} \|\mathbf{u} - \mathbf{v}\|$.*
If $dist(U,V) = 0$ then $U \cap V \neq \emptyset$.

PROOF. Assume U and V are disjoint. Let $\|\cdot\|$ be some norm on $\mathbb{R}^n$ and let $\beta_r(\mathbf{x}) = \{\tilde{\mathbf{x}} \in \mathbb{R}^n \mid \|\mathbf{x} - \tilde{\mathbf{x}}\| < r\}$ denote the open ball located at $\mathbf{x}$ with radius $r > 0$. Then for each pair $(\mathbf{x}, \mathbf{y}) \in U \times V$ we have $\|\mathbf{x} - \mathbf{y}\| > 0$. Set $r(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|$. Every point $\tilde{\mathbf{x}} \in U \cap \beta_{r(\mathbf{x},\mathbf{y})}(\mathbf{x})$ has its distance to $\mathbf{y}$ bounded from below by $r(\mathbf{x}, \mathbf{y})$. Fix some $\mathbf{y} \in V$. Then $\{\beta_{r(\mathbf{x},\mathbf{y})}(\mathbf{x}) \mid \mathbf{x} \in U\}$ is an open covering of U. As U is compact, only finitely many $\beta_{r(\mathbf{x}^{(i)},\mathbf{y})}(\mathbf{x}^{(i)})$ are needed to cover U with $\mathbf{x}^{(i)} \in U$ and $i \in \{1, \dots, K\}$. Hence, every $\mathbf{x} \in U$ has at least the distance

$$m(\mathbf{y}) := \min_{i\in\{1,\dots,K\}} \{r(\mathbf{x}^{(i)}, \mathbf{y})\} > 0$$

to $\mathbf{y}$. Set $r(\mathbf{y}) = \frac{1}{2}m(\mathbf{y})$. Then every $\tilde{\mathbf{y}} \in V \cap \beta_{r(\mathbf{y})}(\mathbf{y})$ has at least distance $r(\mathbf{y})$ to every $\mathbf{x} \in U$. We may use the same construction to realize that there exists some $d > 0$ such that the distance of every pair $(\mathbf{x}, \mathbf{y}) \in U \times V$ is bounded from below by d, contradicting the assumption. □

A.5 Full Proof of Lemma 7

Now we give a full proof of Lemma 7 which is restated here.

LEMMA 7. *Let $\tilde{K}$ denote the kernel of $\mathbf{g}'(\mu\mathbf{f})$, i.e. $\tilde{K} = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{g}'(\mu\mathbf{f})\mathbf{v} = \mathbf{0}\}$. Hence, K, the kernel of $\mathbf{f}'(\mu\mathbf{f}) - \mu\mathbf{f}$, is a subspace of $\tilde{K}$. Further, let S denote the unit sphere $\{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}$ and let $S_{\geq\mathbf{0}} = S \cap \mathbb{R}^n_{\geq 0}$.*
Then $\tilde{K} \cap S_{\geq 0} \neq \emptyset$.

PROOF. Set $M = \mathbf{g}'(\mu\mathbf{f})^\top$ and let $\mathbf{m}_i$ be the i^{th} column vector of M. Assume that M has rank r and that $\mathbf{m}_1, \dots, \mathbf{m}_r$ are linearly independent, otherwise we apply some permutation to $\mathbf{g}$. Construct the vectors $\mathbf{n}_1, \dots, \mathbf{n}_r$ by applying the Gram-Schmidt procedure to $\mathbf{m}_1, \dots, \mathbf{m}_{n-1}$, and arrange them into some matrix $N = (\mathbf{n}_1, \dots, \mathbf{n}_r) \in \mathbb{R}^{n\times r}$. By the Gram-Schmidt procedure $\mathbf{m}_i$ is a linear combination of $\mathbf{n}_1, \dots, \mathbf{n}_{\min\{i,r\}}$ for $1 \leq i \leq n-1$. Hence, we find some upper triangular matrix $C \in \mathbb{R}^{r\times n-1}$ such that $M = NC$. In particular, both N and C have rank r, too.

Now, set $P = \mathrm{Id} - NN^\top$ and $\mathbf{v}_\varepsilon = P\mathbf{u}_\varepsilon$. We want to show that $\|\mathbf{u}_\varepsilon - \mathbf{v}_\varepsilon\| \overset{\varepsilon\to 0}{\to} 0$. First note that, by Lemma 5, $\|\mathbf{g}'(\mu\mathbf{f})\mathbf{u}_\varepsilon\| = \left\|C^\top N^\top \mathbf{u}_\varepsilon\right\| \overset{\varepsilon\to 0}{\to} 0$. This implies $N^\top \mathbf{u}_\varepsilon \overset{\varepsilon\to 0}{\to} \mathbf{0}$, as $C^\top$ has full rank r and thus represents an injective map from $\mathbb{R}^r$ to $\mathbb{R}^{n-1}$. We therefore get $\|\mathbf{u}_\varepsilon - \mathbf{v}_\varepsilon\| = \left\|NN^\top \mathbf{u}_\varepsilon\right\| \leq \|N\| \left\|N^\top \mathbf{u}_\varepsilon\right\| \overset{\varepsilon\to 0}{\to} 0$. Set $B_{\tilde{K}} = \{\mathbf{v} \in \tilde{K} \mid \|\mathbf{v}\| \leq 1\}$. We have $\mathbf{v}_\varepsilon \in B_{\tilde{K}}$ for $0 < \varepsilon < (\mu\mathbf{f})_n$, because P is an orthogonal projector. Hence,

$$\begin{aligned}\mathrm{dist}(S_{\geq 0}, B_{\tilde{K}}) \quad &= \inf\nolimits_{\mathbf{u}\in S_{\geq 0},\mathbf{v}\in B_{\tilde{K}}} \|\mathbf{u} - \mathbf{v}\| \\ &\leq \inf\nolimits_{0<\varepsilon<(\mu\mathbf{f})_n} \|\mathbf{u}_\varepsilon - \mathbf{v}_\varepsilon\| = 0.\end{aligned}$$

As both $B_{\tilde{K}}$ and $S_{\geq 0}$ are compact, we may apply Lemma 6 and conclude $\tilde{K} \cap S_{\geq 0} \neq \emptyset$. □