COMPUTING THE LEAST FIXED POINT OF POSITIVE POLYNOMIAL SYSTEMS*

JAVIER ESPARZA†, STEFAN KIEFER†, AND MICHAEL LUTTENBERGER†

Abstract. We consider equation systems of the form $X_1 = f_1(X_1, \ldots, X_n)$, $\ldots$, $X_n = f_n(X_1, \ldots, X_n)$, where $f_1, \ldots, f_n$ are polynomials with *positive* real coefficients. In vector form we denote such an equation system by $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ and call $\boldsymbol{f}$ a system of positive polynomials (SPP). Equation systems of this kind appear naturally in the analysis of stochastic models like stochastic context-free grammars (with numerous applications to natural language processing and computational biology), probabilistic programs with procedures, web-surfing models with back buttons, and branching processes. The least nonnegative solution $\mu\boldsymbol{f}$ of an SPP equation $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ is of central interest for these models. Etessami and Yannakakis [*J. ACM*, 56 (2009), pp. 1–66] have suggested a particular version of Newton's method to approximate $\mu\boldsymbol{f}$. We extend a result of Etessami and Yannakakis and show that Newton's method starting at $\boldsymbol{0}$ always converges to $\mu\boldsymbol{f}$. We obtain lower bounds on the convergence speed of the method. For so-called *strongly connected* SPPs we prove the existence of a threshold $k_{\boldsymbol{f}} \in \mathbb{N}$ 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 $\mu\boldsymbol{f}$. The proof yields an explicit bound for $k_{\boldsymbol{f}}$ depending only on syntactic parameters of $\boldsymbol{f}$. We further show that for arbitrary SPP equations, Newton's method still converges linearly: there exists a threshold $k_{\boldsymbol{f}}$ and an $\alpha_{\boldsymbol{f}} > 0$ such that for every $i \geq 0$ the $(k_{\boldsymbol{f}} + \alpha_{\boldsymbol{f}} \cdot i)$th iteration of Newton's method has at least i valid bits of $\mu\boldsymbol{f}$. The proof yields an explicit bound for $\alpha_{\boldsymbol{f}}$; the bound is exponential in the number of equations in $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$, but we also show that it is essentially optimal. The proof does not yield any bound for $k_{\boldsymbol{f}}$, but only proves its existence. Constructing a bound for $k_{\boldsymbol{f}}$ is still an open problem. Finally, we also provide a geometric interpretation of Newton's method for SPPs.

Key words. Newton's method, fixed-point equations, polynomial equations, numerical analysis, convergence speed

AMS subject classifications. 41A25, 65H10, 68Q60

DOI. 10.1137/090749591

1. Introduction. We consider equation systems of the form

$$X_1 = f_1(X_1, \ldots, X_n)$$
$$\vdots$$
$$X_n = f_n(X_1, \ldots, X_n),$$

where $f_1, \ldots, f_n$ are polynomials with *positive* real coefficients. In vector form we denote such an equation system by $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$. The vector $\boldsymbol{f}$ of polynomials is called a *system of positive polynomials*, or *SPP* for short. Figure 1.1 shows the graph of a 2-dimensional SPP equation system $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$.

Equation systems of this kind appear naturally in the analysis of stochastic context-free grammars (with numerous applications to natural language processing [19, 28] and computational biology [6, 7, 25, 33]), probabilistic programs with procedures [4, 11, 12, 13, 14, 15, 16], and web-surfing models with back buttons [17, 18].
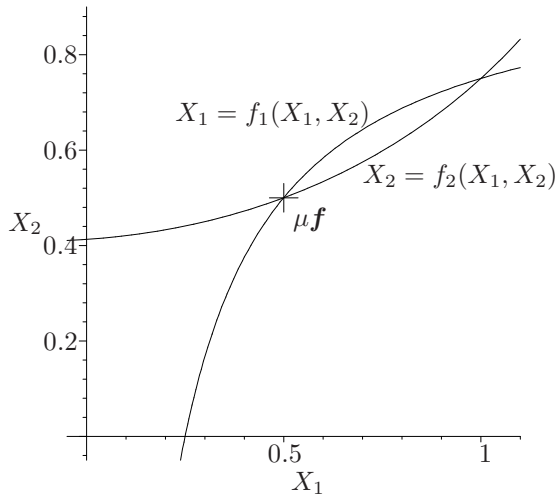

FIG. 1.1. *Graphs of the equations* $X_1 = f_1(X_1, X_2)$ *and* $X_2 = f_2(X_1, X_2)$ *with* $f_1(X_1, X_2) = X_1 X_2 + \frac{1}{4}$ *and* $f_2(X_1, X_2) = \frac{1}{6}X_1^2 + \frac{1}{9}X_1 X_2 + \frac{2}{9}X_2^2 + \frac{3}{8}$. *There are two real solutions in* $\mathbb{R}^2$, *the lesser of which is labeled with* $\mu \boldsymbol{f}$.

More generally, they play an important role in the theory of *branching processes* [2, 21], stochastic processes describing the evolution of a population whose individuals can die and reproduce. The probability of extinction of the population is the least solution of such a system, a result whose history goes back to [35].

Since SPPs have positive coefficients, $\boldsymbol{x} \leq \boldsymbol{x}'$ implies $\boldsymbol{f}(\boldsymbol{x}) \leq \boldsymbol{f}(\boldsymbol{x}')$ for $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}_{\geq 0}^n$; i.e., the functions $f_1, \ldots, f_n$ are monotonic. This allows us to apply Kleene's theorem (see, for instance, [26]) and conclude that a feasible system $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$, i.e., one having at least one nonnegative solution, has a smallest solution $\mu \boldsymbol{f}$. It follows easily from standard Galois theory that $\mu \boldsymbol{f}$ can be irrational and nonexpressible by radicals. The problem of deciding, given an SPP and a rational vector $\boldsymbol{v}$ encoded in binary, whether $\mu \boldsymbol{f} \leq \boldsymbol{v}$ holds, is known to be in PSPACE, and to be at least as hard as two relevant problems: SQUARE-ROOT-SUM and PosSLP. SQUARE-ROOT-SUM is a well-known problem of computational geometry, whose membership in NP is a longstanding open question. PosSLP is the problem of deciding, giving a division-free straight-line program, whether it produces a positive integer (see [16] for more details). PosSLP has been recently shown to play a central role in understanding the Blum–Shub–Smale model of computation, where each single arithmetic operation over the reals can be carried out exactly and in constant time [1].

For the practical applications mentioned above the complexity of determining if $\mu \boldsymbol{f}$ exceeds a given bound is less relevant than the complexity of, given $i \in \mathbb{N}$, computing i *valid bits* of $\mu \boldsymbol{f}$, i.e., computing a vector $\boldsymbol{v}$ such that $\left| \mu \boldsymbol{f}_j - v_j \right| / \left| \mu \boldsymbol{f}_j \right| \leq 2^{-i}$ for every $1 \leq j \leq n$. Given an SPP $\boldsymbol{f}$ and $i \in \mathbb{N}$, deciding whether the first i bits of a component of $\mu \boldsymbol{f}$, say $\mu \boldsymbol{f}_1$, are 0 remains as hard as SQUARE-ROOT-SUM and PosSLP. The reason is that in [16] both problems are reduced to the following one: given $\epsilon > 0$ and an SPP $\boldsymbol{f}$ for which it is known that either $\mu \boldsymbol{f}_1 = 1$ or $\mu \boldsymbol{f}_1 \leq \epsilon$, decide which of the two is the case. So it suffices to take $\epsilon = 2^{-i}$.

In this paper we study the problem of computing i valid bits in the Blum–Shub–Smale model. Since the least fixed point of a feasible SPP $\boldsymbol{f}$ is a solution of $\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{0}$ for $\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{X}) - \boldsymbol{X}$, we can try to apply (the multivariate version of) *Newton's*

method [30]: 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. A first difficulty is that the method might not even be well-defined, because $\boldsymbol{F}'(\boldsymbol{x}^{(k)})$ could be singular for some k. However, Etessami and Yannakakis have recently studied SPPs derived from probabilistic pushdown automata (actually, from an equivalent model called recursive Markov chains) [16], and have shown that a particular version of Newton's method always converges, namely, a version which decomposes the SPP into *strongly connected components* (SCCs)[1] and applies Newton's method to them in a bottom-up fashion. Our first result generalizes Etessami and Yannakakis's: the ordinary Newton method converges for arbitrary SPPs, provided that they are clean (which can be easily achieved).

While these results show that Newton's method can be an adequate algorithm for solving SPP equations, they provide no information on the number of iterations needed to compute i valid bits. To the best of our knowledge (and perhaps surprisingly), the rest of the literature does not contain relevant information either: it has not considered SPPs explicitly, and the existing results have very limited interest for SPPs, since they do not apply even for very simple and relevant SPP cases (see *related work* below). In this paper we obtain upper bounds on the number of iterations that Newton's method needs to produce i valid bits, first for strongly connected and then for arbitrary SPP equations.

For strongly connected SPP equations $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ we prove the existence of 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 $\mu\boldsymbol{f}$. So, loosely speaking, after $k_{\boldsymbol{f}}$ iterations Newton's method is guaranteed to compute at least 1 new bit of the solution per iteration; we say that Newton's method converges at least *linearly* with *rate* 1. Moreover, we show that the threshold $k_{\boldsymbol{f}}$ can be chosen as

$$k_{\boldsymbol{f}} = \lceil 4mn + 3n\max\{0, -\log\mu_{min}\} \rceil,$$

where n is the number of polynomials of the strongly connected SPP, m is such that all coefficients of the SPP can be given as ratios of m-bit integers, and μ_{min} is the minimal component of the least fixed point $\mu\boldsymbol{f}$.

Notice that $k_{\boldsymbol{f}}$ depends on $\mu\boldsymbol{f}$, which is what Newton's method should compute. For this reason we also obtain bounds on $k_{\boldsymbol{f}}$ depending only on m and n. We show that for arbitrary strongly connected SPP equations, $k_{\boldsymbol{f}} = 4mn2^n$ is also a valid threshold. For SPP equations coming from stochastic models, such as the ones listed above, we do far better. First, we show that if every procedure has a nonzero probability of terminating (a condition that always holds for back-button processes [17, 18]), then a valid threshold is $k_{\boldsymbol{f}} = 2m(n + 1)$. Since one iteration requires $\mathcal{O}(n^3)$ arithmetic operations in a system of n equations, we immediately obtain an upper bound on the time complexity of Newton's method in the Blum–Shub–Smale model: for back-button processes, i valid bits can be computed in time $\mathcal{O}(mn^4 + in^3)$. Second, we observe that, 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 $k_{\boldsymbol{f}}$. We exhibit an

---

[1]Loosely speaking, a subset of variables and their associated equations form an SCC if the value of any variable in the subset influences the value of all variables in the subset; see section 2 for details.

SPP for which, using this fact and our theorem, we can prove that no component of the solution reaches the value 1. This cannot be proved by just computing more iterations, no matter how many.

For general SPP equations, not necessarily strongly connected, we show that Newton's method still converges linearly. Formally, we show the existence of a threshold $k_{\boldsymbol{f}}$ and a real number $0 < \alpha_{\boldsymbol{f}}$ such that for every $i \geq 0$ the $(k_{\boldsymbol{f}} + \alpha_{\boldsymbol{f}} \cdot i)$th iteration of Newton's method has at least i valid bits of $\mu\boldsymbol{f}$. So, loosely speaking, after the first $k_{\boldsymbol{f}}$ iterations Newton's method computes new bits of $\mu\boldsymbol{f}$ at a rate of at least $1/\alpha_{\boldsymbol{f}}$ bits per iteration. Unlike the strongly connected case, the proof does not provide any bound on the threshold $k_{\boldsymbol{f}}$: with respect to the threshold the proof is nonconstructive, and finding a bound on $k_{\boldsymbol{f}}$ is still an open problem. However, the proof does provide a bound for $\alpha_{\boldsymbol{f}}$: it shows $\alpha_{\boldsymbol{f}} \leq n \cdot 2^n$ for an SPP with n polynomials. We also exhibit a family of SPPs for which more than $i \cdot 2^{n-1}$ iterations are needed to compute i bits. So $\alpha_{\boldsymbol{f}} \leq n \cdot 2^n$ for every system $\boldsymbol{f}$, and there exists a family of systems for which $2^{n-1} \leq \alpha_{\boldsymbol{f}}$.

Finally, the last result of the paper concerns the geometric interpretation of Newton's method for SPP equations. We show that, loosely speaking, the Newton approximants stay within the hypervolume limited by the hypersurfaces corresponding to each individual equation. This means that a simple geometric intuition of how Newton's method works, extracted from the case of 2-dimensional SPPs, is also correct for arbitrary dimensions. As a byproduct we also obtain a new variant of Newton's method.

Related work. There is a large body of literature on the convergence speed of Newton's method for arbitrary systems of differentiable functions. A comprehensive reference is Ortega and Rheinboldt's book [30] (see also Chapter 8 of Ortega's course [29] or Chapter 5 of [23] for a brief summary). Several theorems (for instance, Theorem 8.1.10 of [29]) prove that the number of valid bits grows linearly, superlinearly, or even exponentially in the number of iterations, but only under the hypothesis that $\boldsymbol{F}'(\boldsymbol{x})$ is nonsingular everywhere, in a neighborhood of $\mu\boldsymbol{f}$, or at least at the point $\mu\boldsymbol{f}$ itself. However, the matrix $\boldsymbol{F}'(\mu\boldsymbol{f})$ can be singular for an SPP, even for the 1-dimensional SPP $f(X) = 1/2X^2 + 1/2$.

The general case in which $\boldsymbol{F}'(\mu\boldsymbol{f})$ may be singular for the solution $\mu\boldsymbol{f}$ that the method converges to has been thoroughly studied. In a seminal paper [32], Reddien shows that under certain conditions, the main ones being that the kernel of $\boldsymbol{F}'(\mu\boldsymbol{f})$ has dimension 1 and that the initial point is close enough to the solution, Newton's method gains 1 bit per iteration. Decker and Kelley obtain results for kernels of arbitrary dimension, but they require a certain linear map $B(\boldsymbol{X})$ to be nonsingular for all $\boldsymbol{x} \neq \boldsymbol{0}$ [5]. Griewank and Osborne observe in [20] that the nonsingularity of $B(\boldsymbol{X})$ is in fact a strong condition which, in particular, can be satisfied only by kernels of even dimension. They present a weaker sufficient condition for linear convergence requiring $B(\boldsymbol{X})$ to be nonsingular only at the initial point $\boldsymbol{x}^{(0)}$; i.e., it requires only to make "the right guess" for $\boldsymbol{x}^{(0)}$. Unfortunately, none of these results can be directly applied to arbitrary SPPs. The possible dimensions of the kernel of $\boldsymbol{F}'(\mu\boldsymbol{f})$ for an SPP $\boldsymbol{f}(\boldsymbol{X})$ are to the best of our knowledge unknown, and deciding this question seems as hard as those related to the convergence rate.[2] Griewank and Osborne's result does not apply to the decomposed Newton method either because the mapping $B(\boldsymbol{x}^{(0)})$ is always singular for $\boldsymbol{x}^{(0)} = \boldsymbol{0}$.

---

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

Kantorovich's famous theorem (see, e.g., [30, Theorem 8.2.6], as well as [31] for an improvement) guarantees global convergence and only requires $\boldsymbol{F}'$ to be nonsingular at $\boldsymbol{x}^{(0)}$. However, it also requires finding a Lipschitz constant for $\boldsymbol{F}'$ on a suitable region and some other bounds on $\boldsymbol{F}'$. These latter conditions are far too restrictive for the applications mentioned above. For instance, the stochastic context-free grammars whose associated SPPs satisfy Kantorovich's conditions cannot exhibit two productions $X \to aYZ$ and $W \to \varepsilon$ such that $\operatorname{Prob}(X \to aYZ) \cdot \operatorname{Prob}(W \to \varepsilon) \geq 1/4$. This class of grammars is too contrived to be of use.

Summarizing, while the convergence of Newton's method for systems of differentiable functions has been intensely studied, the case of SPPs does not seem to have been considered yet. The results obtained for other classes have very limited applicability to SPPs: either they do not apply at all, or they apply only to contrived SPP subclasses. Moreover, these results only provide information about the growth rate of the number of accurate bits, but not about the number itself. For the class of strongly connected SPPs, our thresholds lead to *explicit* lower bounds for the number of accurate bits depending only on syntactical parameters: the number of equations and the size of the coefficients. For arbitrary SPPs we prove the existence of a threshold, while finding explicit lower bounds remains an open problem.

Structure of the paper. Section 2 defines SPPs and briefly describes their applications to stochastic systems. Section 3 presents a short summary of our main theorems. Section 4 proves some fundamental properties of Newton's method for SPP equations. Sections 5 and 6 contain our results on the convergence speed for strongly connected and general SPP equations, respectively. Section 7 shows that the bounds are essentially tight. Section 8 presents our results about the geometrical interpretation of Newton's method, and section 9 contains conclusions. A slightly extended version of this paper is available as a technical report [10].

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 sets of real and natural numbers, respectively. 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 nonnegative components. We use bold letters for vectors, e.g., $\boldsymbol{x} \in \mathbb{R}^n$, where we assume that $\boldsymbol{x}$ has the components $x_1, \dots, x_n$. Similarly, the ith component of a function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ is denoted by f_i. We define $\boldsymbol{0} := (0, \dots, 0)^\top$ and $\boldsymbol{1} := (1, \dots, 1)^\top$, where the superscript $^\top$ indicates the transpose of a vector or a matrix. Let $\|\cdot\|$ denote some norm on $\mathbb{R}^n$. Sometimes we use explicitly the maximum norm $\|\cdot\|_\infty$ with $\|\boldsymbol{x}\|_\infty := \max_{1 \leq i \leq n} |x_i|$.

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 $X_1, \dots, X_n$ 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 $\boldsymbol{x}, \boldsymbol{y}, \dots$ denote arbitrary elements in $\mathbb{R}^n$, we write $\boldsymbol{X}$ if we want to emphasize that a function is given with respect to (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 $S \subseteq \{1, \dots, n\}$ is a set of components and $\boldsymbol{x}$ a vector, then by $\boldsymbol{x}_S$ we mean the vector obtained by restricting $\boldsymbol{x}$ to the components in S.

Let $S \subseteq \{1, \dots, n\}$ and $\overline{S} = \{1, \dots, n\} \setminus S$. Given a function $\boldsymbol{f}(\boldsymbol{X})$ and a vector

$\boldsymbol{x}_S$, then $\boldsymbol{f}[S/\boldsymbol{x}_S]$ is obtained by replacing, for each $s \in S$, each occurrence of $\boldsymbol{X}_s$ by $\boldsymbol{x}_s$ and removing the s-component. In other words, if $\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{f}(\boldsymbol{X}_S, \boldsymbol{X}_{\overline{S}})$, then $\boldsymbol{f}[S/\boldsymbol{x}_S](\boldsymbol{y}_{\overline{S}}) = \boldsymbol{f}_{\overline{S}}(\boldsymbol{x}_S, \boldsymbol{y}_{\overline{S}})$. For instance, if $\boldsymbol{f}(X_1, X_2) = (X_1 X_2 + \frac{1}{2}, X_2^2 + \frac{1}{5})^{\top}$, then $\boldsymbol{f}[\{2\}/\frac{1}{2}] : \mathbb{R} \to \mathbb{R}, X_1 \mapsto \frac{1}{2}X_1 + \frac{1}{2}$.

$\mathbb{R}^{m \times n}$ 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 $A \in \mathbb{R}^{n \times n}$ 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., $\max\{|\lambda| \mid \mathbb{C} \ni \lambda \text{ is an eigenvalue of } A\} < 1$. If A^* exists, then $A^* = (\text{Id} - A)^{-1}$.

The partial derivative of a function $f(\boldsymbol{X}) : \mathbb{R}^n \to \mathbb{R}$ w.r.t. the variable X_i is denoted by $\partial_{X_i} f$. The gradient ∇f of $f(\boldsymbol{X})$ is then defined to be the (row) vector

$$\nabla f := (\partial_{X_1} f, \dots, \partial_{X_n} f).$$

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

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

i.e., the ith row of $\boldsymbol{f}'$ is the gradient of f_i.

2.2. Systems of positive polynomials.

DEFINITION 2.1. *A function $\boldsymbol{f}(\boldsymbol{X})$ with $\boldsymbol{f} : \mathbb{R}^n_{\geq 0} \to \mathbb{R}^n_{\geq 0}$ is a system of positive polynomials (SPP) if every component $f_i(\boldsymbol{X})$ is a polynomial in the variables $X_1, \dots, X_n$ with coefficients in $\mathbb{R}_{\geq 0}$. We call an SPP $\boldsymbol{f}(\boldsymbol{X})$ feasible if $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{y})$ for some $\boldsymbol{y} \in \mathbb{R}^n_{\geq 0}$. An SPP is called* linear *(resp.,* quadratic*) if all polynomials have degree at most 1 (resp., 2).*

FACT 2.2. *Every SPP $\boldsymbol{f}$ is monotone on $\mathbb{R}^n_{\geq 0}$; i.e., for $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{y}$ we have $\boldsymbol{f}(\boldsymbol{x}) \leq \boldsymbol{f}(\boldsymbol{y})$.*

We will need the following lemma, a version of Taylor's theorem.

LEMMA 2.3 (Taylor). *Let $\boldsymbol{f}$ be an SPP and let $\boldsymbol{x}, \boldsymbol{u} \geq \boldsymbol{0}$. Then*

$$\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{u} \leq \boldsymbol{f}(\boldsymbol{x} + \boldsymbol{u}) \leq \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{f}'(\boldsymbol{x} + \boldsymbol{u})\boldsymbol{u}.$$

Proof. It suffices to show this for a multivariate polynomial $f(\boldsymbol{X})$ with nonnegative coefficients. Consider $g(t) = f(\boldsymbol{x} + t\boldsymbol{u})$. We then have

$$f(\boldsymbol{x} + \boldsymbol{u}) = g(1) = g(0) + \int_0^1 g'(s) \, ds = f(\boldsymbol{x}) + \int_0^1 f'(\boldsymbol{x} + s\boldsymbol{u})\boldsymbol{u} \, ds.$$

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

Since every SPP is continuous, Kleene's fixed-point theorem (see, e.g., [26]) applies.

THEOREM 2.4 (Kleene's fixed-point theorem). *Every feasible SPP $\boldsymbol{f}$ 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 $\boldsymbol{\kappa}_{\boldsymbol{f}}^{(k)} = \boldsymbol{f}^k(\boldsymbol{0})$ (where $\boldsymbol{f}^k$ denotes the k-fold iteration of $\boldsymbol{f}$) is monotonically increasing w.r.t. $\leq$ (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 $\mu\boldsymbol{f}$.

An SPP $\boldsymbol{f}(\boldsymbol{X})$ is *clean* if for all variables X_i there is a $k \in \mathbb{N}$ such that $\kappa_i^{(k)} > 0$. It is easy to see that we have $\kappa_i^{(k)} = 0$ for all $k \in \mathbb{N}$ if $\kappa_i^{(n)} = 0$. So we can "clean" an SPP $\boldsymbol{f}(\boldsymbol{X})$ in time linear in the size of $\boldsymbol{f}$ by determining the components i with $\kappa_i^{(n)} = 0$ and removing them.

We will also need the notion of *dependence* between variables.

DEFINITION 2.5. *A polynomial $f(\boldsymbol{X})$ contains a variable X_i if $\partial_{X_i} f(\boldsymbol{X})$ is not the zero-polynomial.*

DEFINITION 2.6. *Let $\boldsymbol{f}(\boldsymbol{X})$ be an SPP. A component i depends directly on a component k if $f_i(\boldsymbol{X})$ contains X_k. A component i depends on k if either i depends directly on k or there is a component j such that i depends on j and j depends on k. The components $\{1, \ldots, n\}$ can be partitioned into strongly connected components (SCCs), where an SCC S is a maximal set of components such that each component in S depends on each other component in S. An SCC is called* trivial *if it consists of a single component that does not depend on itself. An SPP is strongly connected (scSPP, for short) if $\{1, \ldots, n\}$ is a nontrivial SCC.*

2.3. Convergence speed. We will analyze the convergence speed of Newton's method. To this end we need the notion of *valid bits*.

DEFINITION 2.7. *Let $\boldsymbol{f}$ be a feasible SPP. A vector $\boldsymbol{x}$ has i* valid bits *of the least fixed point $\mu\boldsymbol{f}$ if*

$$\frac{\left|\mu\boldsymbol{f}_j - x_j\right|}{\left|\mu\boldsymbol{f}_j\right|} \le 2^{-i}$$

for every $1 \le j \le n$. Let $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ be a sequence with $\boldsymbol{0} \le \boldsymbol{x}^{(k)} \le \mu\boldsymbol{f}$. Then the convergence order *$\beta : \mathbb{N} \to \mathbb{N}$ of the sequence $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ is defined as follows: $\beta(k)$ is the greatest natural number i such that $\boldsymbol{x}^{(k)}$ has i valid bits (or ∞ if such a greatest number does not exist). We will always mean the convergence order of the Newton sequence $(\boldsymbol{\nu}^{(k)})_{k\in\mathbb{N}}$ unless explicitly stated otherwise.*

We say that a sequence has linear, exponential, or logarithmic convergence order if the function $\beta(k)$ grows linearly, exponentially, or logarithmically in k, respectively.

Remark 2.8. Our definition of convergence order differs from the one commonly used in numerical analysis (see, e.g., [30]), where "quadratic convergence" or "Q-quadratic convergence" means that the error e' of the new approximant (its distance to the least fixed point according to some norm) is bounded by $c \cdot e^2$, where e is the error of the old approximant and $c > 0$ is some constant. We consider our notion more natural from a computational point of view, since it directly relates the number of iterations to the accuracy of the approximation. Notice that "quadratic convergence" implies exponential convergence order in the sense of Definition 2.7. In the following we avoid the notion of "quadratic convergence."

2.4. Stochastic models. As mentioned in the introduction, several problems concerning stochastic models can be reduced to problems about the least fixed point $\mu\boldsymbol{f}$ of an SPP $\boldsymbol{f}$. In these cases, $\mu\boldsymbol{f}$ is a vector of probabilities, and so $\mu\boldsymbol{f} \le \boldsymbol{1}$.

2.4.1. Probabilistic pushdown automata. Our study of SPPs was initially motivated by the verification of probabilistic pushdown automata. A *probabilistic pushdown automaton (pPDA)* is a tuple $\mathcal{P} = (Q, \Gamma, \delta, \mathrm{Prob})$, 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 $\mathrm{Prob}(pX \hookrightarrow q\alpha) \in (0, 1]$ so that for all $p \in Q$ and $X \in \Gamma$ we have $\sum_{pX \hookrightarrow q\alpha} \mathrm{Prob}(pX \hookrightarrow q\alpha) = 1$. We write $pX \overset{x}{\hookrightarrow} q\alpha$ instead of $\mathrm{Prob}(pX \hookrightarrow q\alpha) = x$. A *configuration* of $\mathcal{P}$ is a pair qw, where q is a control state and $w \in \Gamma^*$ is a *stack content*. A pPDA $\mathcal{P}$ 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 $\beta \in \Gamma^*$ if and only if $pX \overset{x}{\hookrightarrow} q\alpha$. We assume without loss of generality (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 [4, 11, 12, 13, 14, 15, 16]. These works have 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., [11] 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 solution of the SPP equation system containing the equation

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

for each triple (p, X, q). Call this quadratic SPP the *termination SPP* of the pPDA (we assume that termination SPPs are clean, and it is easy to see that they are always feasible).

2.4.2. 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 SPP of a strict pPDA satisfies $\boldsymbol{f}(\boldsymbol{0}) \succ \boldsymbol{0}$.

In [17, 18] a class of stochastic processes is introduced to model the behavior of web surfers who 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 strict pPDAs having one single control state (which in the following we omit), and rules of the form $A \overset{b_A}{\longrightarrow} \varepsilon$ (press the back button from A) or $A \overset{\ell_{AB}}{\longrightarrow} BA$ (follow the link from A to B, remembering A as the destination of pressing the back button at B). The termination probabilities are given by an SPP equation system containing the equation

$$\langle A \rangle \; = \; b_A + \sum_{A \overset{\ell_{AB}}{\longrightarrow} BA} \ell_{AB} \langle B \rangle \langle A \rangle \; = \; b_A + \langle A \rangle \sum_{A \overset{\ell_{AB}}{\longrightarrow} BA} \ell_{AB} \langle B \rangle$$

for every webpage A. In [17, 18] those termination probabilities are called *revocation* probabilities. The revocation probability of a page A is the probability that, when currently visiting webpage A and having $H_0 H_1 \ldots H_{n-1} H_n$ as the browser history of previously visited pages, then during subsequent surfing from A the random user eventually returns to webpage H_n with $H_0 H_1 \ldots H_{n-1}$ as the remaining browser history.

Example 2.9. Consider the following equation system:

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

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.

3. Newton's method and an overview of our results. In order to approximate the least fixed point $\mu\boldsymbol{f}$ of an SPP $\boldsymbol{f}$ we employ Newton's method.

DEFINITION 3.1. *Let $\boldsymbol{f}$ be a clean and feasible SPP. The Newton operator $\mathcal{N}_{\boldsymbol{f}}$ is defined as follows:*

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

The sequence $(\boldsymbol{\nu}_{\boldsymbol{f}}^{(k)})_{k \in \mathbb{N}}$ with $\boldsymbol{\nu}_{\boldsymbol{f}}^{(k)} = \mathcal{N}_{\boldsymbol{f}}^k(\boldsymbol{0})$ (where $\mathcal{N}_{\boldsymbol{f}}^k$ denotes the k-fold iteration of $\mathcal{N}_{\boldsymbol{f}}$) is called the Newton sequence. *We drop the subscript of $\mathcal{N}_{\boldsymbol{f}}$ and $\boldsymbol{\nu}_{\boldsymbol{f}}^{(k)}$ when $\boldsymbol{f}$ is understood.*

The main results of this paper concern the application of Newton's method to SPPs. We summarize them in this section.

Theorem 4.1 states that the Newton sequence $(\boldsymbol{\nu}^{(k)})_{k \in \mathbb{N}}$ is well-defined (i.e., the inverse matrices $\left(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(k)}) \right)^{-1}$ exist for every $k \in \mathbb{N}$), monotonically increasing and bounded from above by $\mu\boldsymbol{f}$ (i.e., $\boldsymbol{\nu}^{(k)} \le \boldsymbol{f}(\boldsymbol{\nu}^{(k)}) \le \boldsymbol{\nu}^{(k+1)} \le \mu\boldsymbol{f}$), and converges to $\mu\boldsymbol{f}$. This theorem generalizes the result of Etessami and Yannakakis in [16] to arbitrary clean and feasible SPPs and to the ordinary Newton method.

For more quantitative results on the convergence speed, it is convenient to focus on quadratic SPPs. **Theorem 4.12** shows that any clean and feasible SPP can be syntactically transformed into a quadratic SPP without changing the least fixed point and without accelerating Newton's method. This means one can perform Newton's method on the original (possibly nonquadratic) SPP, and convergence will be at least as fast as for the corresponding quadratic SPP.

For quadratic n-dimensional SPPs, one iteration of Newton's method involves $\mathcal{O}(n^3)$ arithmetical operations and $\mathcal{O}(n^3)$ operations in the Blum–Shub–Smale model. Hence, a bound on the number of iterations needed to compute a given number of valid bits immediately leads to a bound on the number of operations. In section 5 we obtain such bounds for *strongly connected* quadratic SPPs. We give different thresholds for the number of iterations and show that when any of these thresholds is reached, Newton's method gains at least one valid bit for each iteration. More precisely, **Theorem 5.10** states the following. Let $\boldsymbol{f}$ be a quadratic, clean, and feasible scSPP, let μ_{min} and μ_{max} be the minimal and maximal component of $\mu\boldsymbol{f}$, respectively, and let the coefficients of $\boldsymbol{f}$ be given as ratios of m-bit integers. Then $\beta(k_{\boldsymbol{f}} + i) \ge i$ holds for all $i \in \mathbb{N}$ and for any of the following choices of $k_{\boldsymbol{f}}$:
1. $4mn + \lceil 3n \max\{0, -\log \mu_{min}\} \rceil$;
2. $4mn2^n$;
3. $7mn$ if $\boldsymbol{f}$ satisfies $\boldsymbol{f}(\boldsymbol{0}) \succ \boldsymbol{0}$;
4. $2m(n+1)$ if $\boldsymbol{f}$ satisfies both $\boldsymbol{f}(\boldsymbol{0}) \succ \boldsymbol{0}$ and $\mu_{max} \le 1$.

We further show that Newton iterations can also be used to obtain a sequence of *upper* approximations of $\mu\boldsymbol{f}$. Those upper approximations converge to $\mu\boldsymbol{f}$, asymptotically as fast as the Newton sequence. More precisely, **Theorem 5.13** states the following: Let $\boldsymbol{f}$ be a quadratic, clean, and feasible scSPP, let c_{min} be the smallest nonzero coefficient of $\boldsymbol{f}$, and let μ_{min} be the minimal component of $\mu\boldsymbol{f}$. Further, for all Newton approximants $\boldsymbol{\nu}^{(k)}$ with $\boldsymbol{\nu}^{(k)} \succ \boldsymbol{0}$, let $\nu_{min}^{(k)}$ be the smallest coefficient of $\boldsymbol{\nu}^{(k)}$. Then

$$\boldsymbol{\nu}^{(k)} \le \mu\boldsymbol{f} \le \boldsymbol{\nu}^{(k)} + \left[\frac{\left\| \boldsymbol{\nu}^{(k)} - \boldsymbol{\nu}^{(k-1)} \right\|_{\infty}}{\left(c_{min} \cdot \min\{\nu_{min}^{(k)}, 1\} \right)^n} \right],$$

where $[s]$ denotes the vector $\boldsymbol{x}$ with $x_j = s$ for all $1 \le j \le n$.

In section 6 we turn to general (not necessarily strongly connected) clean and feasible SPPs. We show in **Theorem 6.5** that Newton's method still converges linearly. Formally, the theorem proves that for every quadratic, clean, and feasible SPP $\boldsymbol{f}$, there is a threshold $k_{\boldsymbol{f}} \in \mathbb{N}$ and $\alpha_{\boldsymbol{f}} > 0$ such that $\beta(k_{\boldsymbol{f}} + \alpha_{\boldsymbol{f}} \cdot i) \ge i$ for all $i \in \mathbb{N}$. With respect to the threshold, our proof is purely existential and does not provide any bound for $k_{\boldsymbol{f}}$. For $\alpha_{\boldsymbol{f}}$ we show an upper bound of $n \cdot 2^n$, i.e., asymptotically at most $n \cdot 2^n$ extra iterations are needed in order to get one new valid bit. Section 7 exhibits a family of SPPs in which one new bit requires at least 2^{n-1} iterations, implying that the bound on $\alpha_{\boldsymbol{f}}$ is essentially tight.

Finally, section 8 gives a geometrical interpretation of Newton's method on quadratic SPP equations. Let R be the region bounded by the coordinate axes and by the quadrics corresponding to the individual equations. **Theorem 8.10** shows that all Kleene and Newton approximations lie within R, i.e., $\boldsymbol{\nu}^{(i)}, \boldsymbol{\kappa}^{(i)} \in R$ for every $i \in \mathbb{N}$.

4. Fundamental properties of Newton's method.

4.1. Effectiveness. Etessami and Yannakakis [16] suggested using Newton's method for SPPs. More precisely, they showed that the sequence obtained by applying Newton's method to the equation system $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ converges to $\mu\boldsymbol{f}$ as long as $\boldsymbol{f}$ is strongly connected. We extend their result to arbitrary SPPs, thereby reusing and extending several proofs of [16].

In Definition 3.1 we defined the Newton operator $\mathcal{N}_{\boldsymbol{f}}$ and the associated Newton sequence $(\boldsymbol{\nu}^{(k)})_{k \in \mathbb{N}}$. In this section we prove the following fundamental theorem on the Newton sequence.

THEOREM 4.1. *Let $\boldsymbol{f}$ be a clean and feasible SPP. Let the Newton operator $\mathcal{N}_{\boldsymbol{f}}$ be defined as in Definition* 3.1:

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

1. *Then the Newton sequence $(\boldsymbol{\nu}^{(k)})_{k \in \mathbb{N}}$ with $\boldsymbol{\nu}^{(k)} = \mathcal{N}_{\boldsymbol{f}}^k(\boldsymbol{0})$ is well-defined (i.e., the matrix inverses exist), monotonically increasing, and bounded from above by $\mu\boldsymbol{f}$ (i.e., $\boldsymbol{\nu}^{(k)} \le \boldsymbol{f}(\boldsymbol{\nu}^{(k)}) \le \boldsymbol{\nu}^{(k+1)} \le \mu\boldsymbol{f}$) and converges to $\mu\boldsymbol{f}$.*
2. *We have $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(k)}))^{-1} = \boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*$ for all $k \in \mathbb{N}$. We also have $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x}))^{-1} = \boldsymbol{f}'(\boldsymbol{x})^*$ for all $\boldsymbol{x} \prec \mu\boldsymbol{f}$.*

The proof of Theorem 4.1 consists of three steps. In the first proof step, we study a sequence generated by a somewhat weaker version of the Newton operator and obtain the following.

PROPOSITION 4.2. *Let $\boldsymbol{f}$ be a feasible SPP. Let the operator $\widehat{\mathcal{N}}_{\boldsymbol{f}}$ be defined as follows:*

$$\widehat{\mathcal{N}}_{\boldsymbol{f}}(\boldsymbol{X}) := \boldsymbol{X} + \sum_{d=0}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{X})^d (\boldsymbol{f}(\boldsymbol{X}) - \boldsymbol{X}) \right) .$$

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

In a second proof step, we show another intermediary proposition, namely, that the star of the Jacobian matrix $\boldsymbol{f}'$ converges for all Newton approximants.

PROPOSITION 4.3. *Let $\boldsymbol{f}$ be clean and feasible. Then the matrix series $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^* := \mathrm{Id} + \boldsymbol{f}'(\boldsymbol{\nu}^{(k)}) + \boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^2 + \cdots$ converges in $\mathbb{R}_{\ge 0}$ for all Newton approximants $\boldsymbol{\nu}^{(k)}$; i.e., there are no ∞ entries.*

In the third and final step, we show that Propositions 4.2 and 4.3 imply Theorem 4.1.

4.1.1. First step. For the first proof step (i.e., the proof of Proposition 4.2), we will need the following generalization of Taylor's theorem.

LEMMA 4.4. *Let $\boldsymbol{f}$ be an SPP, $d \in \mathbb{N}$, $\boldsymbol{0} \le \boldsymbol{u}$, and $\boldsymbol{0} \le \boldsymbol{x} \le \boldsymbol{f}(\boldsymbol{x})$. Then*

$$\boldsymbol{f}^d(\boldsymbol{x} + \boldsymbol{u}) \ge \boldsymbol{f}^d(\boldsymbol{x}) + \boldsymbol{f}'(\boldsymbol{x})^d \boldsymbol{u} .$$

In particular, by setting $\boldsymbol{u} := \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}$ we get

$$\boldsymbol{f}^{d+1}(\boldsymbol{x}) - \boldsymbol{f}^d(\boldsymbol{x}) \ge \boldsymbol{f}'(\boldsymbol{x})^d (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) .$$

Proof. The proof is by induction on d. For $d = 0$ the statement is trivial. Let $d \ge 0$. Then, by Taylor's theorem (Lemma 2.3), we have

$$
\begin{aligned}
\boldsymbol{f}^{d+1}(\boldsymbol{x} + \boldsymbol{u}) &= \boldsymbol{f}(\boldsymbol{f}^d(\boldsymbol{x} + \boldsymbol{u})) \\
&\ge \boldsymbol{f}(\boldsymbol{f}^d(\boldsymbol{x}) + \boldsymbol{f}'(\boldsymbol{x})^d \boldsymbol{u}) && \text{(induction hypothesis)} \\
&\ge \boldsymbol{f}^{d+1}(\boldsymbol{x}) + \boldsymbol{f}'(\boldsymbol{f}^d(\boldsymbol{x}))\boldsymbol{f}'(\boldsymbol{x})^d \boldsymbol{u} && \text{(Lemma 2.3)} \\
&\ge \boldsymbol{f}^{d+1}(\boldsymbol{x}) + \boldsymbol{f}'(\boldsymbol{x})^{d+1}\boldsymbol{u} && (\boldsymbol{f}^d(\boldsymbol{x}) \ge \boldsymbol{x}). \quad \square
\end{aligned}
$$

Lemma 4.4 can be used to prove the following.

LEMMA 4.5. *Let $\boldsymbol{f}$ be a feasible SPP. Let $\boldsymbol{0} \le \boldsymbol{x} \le \mu\boldsymbol{f}$ and $\boldsymbol{x} \le \boldsymbol{f}(\boldsymbol{x})$. Then*

$$\boldsymbol{x} + \sum_{d=0}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{x})^d (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) \right) \le \mu\boldsymbol{f} .$$

Proof. Observe that

(4.1) $$\lim_{d \to \infty} \boldsymbol{f}^d(\boldsymbol{x}) = \mu\boldsymbol{f}$$

because $\boldsymbol{0} \le \boldsymbol{x} \le \mu\boldsymbol{f}$ implies $\boldsymbol{f}^d(\boldsymbol{0}) \le \boldsymbol{f}^d(\boldsymbol{x}) \le \mu\boldsymbol{f}$, and as $(\boldsymbol{f}^d(\boldsymbol{0}))_{d \in \mathbb{N}}$ converges to $\mu\boldsymbol{f}$ by Theorem 2.4, so does $(\boldsymbol{f}^d(\boldsymbol{x}))_{d \in \mathbb{N}}$. We have

$$\boldsymbol{x} + \sum_{d=0}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{x})^d (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) \right) \le \boldsymbol{x} + \sum_{d=0}^{\infty} \left(\boldsymbol{f}^{d+1}(\boldsymbol{x}) - \boldsymbol{f}^d(\boldsymbol{x}) \right) \quad \text{(Lemma 4.4)}$$

$$= \lim_{d \to \infty} \boldsymbol{f}^d(\boldsymbol{x})$$
$$= \mu \boldsymbol{f} \qquad \text{(by (4.1)).} \qquad \square$$

Now we can prove Proposition 4.2.

Proof of Proposition 4.2. First we prove the following inequality by induction on k:

$$\boldsymbol{\nu}^{(k)} \leq \boldsymbol{f}(\boldsymbol{\nu}^{(k)}).$$

The induction base ($k = 0$) is easy. For this step, let $k \geq 0$. Then

$$\boldsymbol{\nu}^{(k+1)} = \boldsymbol{\nu}^{(k)} + \sum_{d=0}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^d (\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \right)$$

$$= \boldsymbol{f}(\boldsymbol{\nu}^{(k)}) + \sum_{d=1}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^d (\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \right)$$

$$\leq \boldsymbol{f}(\boldsymbol{\nu}^{(k)}) + \boldsymbol{f}'(\boldsymbol{\nu}^{(k)}) \sum_{d=0}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^d (\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \right)$$

$$\leq \boldsymbol{f} \left(\boldsymbol{\nu}^{(k)} + \sum_{d=0}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^d (\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \right) \right) \qquad \text{(Lemma 2.3)}$$

$$= \boldsymbol{f}(\boldsymbol{\nu}^{(k+1)}).$$

Now, the inequality $\boldsymbol{\nu}^{(k)} \leq \mu \boldsymbol{f}$ follows from Lemma 4.5 by means of a straightforward induction proof. Hence, it follows that $\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) \leq \boldsymbol{f}(\mu \boldsymbol{f}) = \mu \boldsymbol{f}$. Further we have

(4.2)
$$\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) = \boldsymbol{\nu}^{(k)} + (\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)})$$
$$\leq \boldsymbol{\nu}^{(k)} + \sum_{d=0}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^d (\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \right) = \boldsymbol{\nu}^{(k+1)}.$$

So it remains to show that $(\boldsymbol{\nu}^{(k)})_{k \in \mathbb{N}}$ converges to $\mu \boldsymbol{f}$. As we have already shown that $\boldsymbol{\nu}^{(k)} \leq \mu \boldsymbol{f}$, it suffices to show that $\boldsymbol{\kappa}^{(k)} \leq \boldsymbol{\nu}^{(k)}$ because $(\boldsymbol{\kappa}^{(k)})_{k \in \mathbb{N}}$ converges to $\mu \boldsymbol{f}$ by Theorem 2.4. We proceed by induction on k. The induction base ($k = 0$) is easy. For this induction step, let $k \geq 0$. Then

$$\boldsymbol{\kappa}^{(k+1)} = \boldsymbol{f}(\boldsymbol{\kappa}^{(k)})$$
$$\leq \boldsymbol{f}(\boldsymbol{\nu}^{(k)}) \qquad \text{(induction hypothesis)}$$
$$\leq \boldsymbol{\nu}^{(k+1)} \qquad \text{(by (4.2)).}$$

This completes the proof of Proposition 4.2 and, hence, the first step towards the proof of Theorem 4.1. $\square$

4.1.2. Second step. For the second proof step (i.e., the proof of Proposition 4.3) it is convenient to move to the *extended reals* $\mathbb{R}_{[0,\infty]}$; i.e., we extend $\mathbb{R}_{\geq 0}$ by an element ∞ such that addition satisfies $a + \infty = \infty + a = \infty$ for all $a \in \mathbb{R}_{\geq 0}$ and multiplication satisfies $0 \cdot \infty = \infty \cdot 0 = 0$ and $a \cdot \infty = \infty \cdot a = \infty$ for all $a \in \mathbb{R}_{>0}$. In $\mathbb{R}_{[0,\infty]}$, one can rewrite $\widehat{\mathcal{N}}(\boldsymbol{\nu}^{(k)}) = \boldsymbol{\nu}^{(k)} + \sum_{d=0}^{\infty} \left(\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^d (\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \right)$ as $\boldsymbol{\nu}^{(k)} +$

$\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^* (\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)})$. Notice that Proposition 4.3 does not follow trivially from Proposition 4.2, because ∞ entries of $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*$ could be cancelled out by matching 0 entries of $\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}$.

For the proof of Proposition 4.3 we need several lemmas. The following lemma assures that a starred matrix has an ∞ entry if and only if it has an ∞ entry on the diagonal.

LEMMA 4.6. *Let $A = (a_{ij}) \in \mathbb{R}_{\geq 0}^{n \times n}$. 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. The proof is 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

$$(4.3) \qquad [A^*]_{1n} = [A^*]_{11} \sum_{j=2}^{n} a_{1j} \left[A^*_{[2..n,2..n]} \right]_{jn} ,$$

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 (4.3) 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 subpaths P_1, P_2 as follows. The second subpath P_2 is the suffix of P leading from 1 to n and not returning to 1. The first subpath 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 $\sum_{j=2}^{n} a_{1j} \left[(A_{[2..n,2..n]})^* \right]_{jn}$. So (4.3) holds.

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

The following lemma treats the case that $\boldsymbol{f}$ is strongly connected (cf. [16]).

LEMMA 4.7. *Let $\boldsymbol{f}$ be clean, feasible, and nontrivially strongly connected. Let $\boldsymbol{0} \leq \boldsymbol{x} \prec \mu\boldsymbol{f}$. Then $\boldsymbol{f}'(\boldsymbol{x})^*$ does not have ∞ as an entry.*

Proof. By Theorem 2.4 the Kleene sequence $(\boldsymbol{\kappa}^{(i)})_{i \in \mathbb{N}}$ converges to $\mu\boldsymbol{f}$. Furthermore, $\boldsymbol{\kappa}^{(i)} \prec \mu\boldsymbol{f}$ holds for all i, because as every component depends nontrivially on itself, any increase in any component results in an increase of the same component in a later Kleene approximant. So, we can choose a Kleene approximant $\boldsymbol{y} = \boldsymbol{\kappa}^{(i)}$ such that $\boldsymbol{x} \leq \boldsymbol{y} \prec \mu\boldsymbol{f}$. Notice that $\boldsymbol{y} \leq \boldsymbol{f}(\boldsymbol{y})$. By monotonicity of $\boldsymbol{f}'$ it suffices to show that $\boldsymbol{f}'(\boldsymbol{y})^*$ does not have ∞ as an entry. By Lemma 4.4 (taking $\boldsymbol{x} := \boldsymbol{y}$ and $\boldsymbol{u} := \mu\boldsymbol{f} - \boldsymbol{y}$) we have

$$\boldsymbol{f}'(\boldsymbol{y})^d (\mu\boldsymbol{f} - \boldsymbol{y}) \leq \mu\boldsymbol{f} - \boldsymbol{f}^d(\boldsymbol{y}) .$$

As $d \to \infty$, the right-hand side converges to $\boldsymbol{0}$, because, by Kleene's theorem, $\boldsymbol{f}^d(\boldsymbol{y})$ converges to $\mu\boldsymbol{f}$. So the left-hand side also converges to $\boldsymbol{0}$. Since $\mu\boldsymbol{f} - \boldsymbol{y} \succ \boldsymbol{0}$, every entry of $\boldsymbol{f}'(\boldsymbol{y})^d$ must converge to $\boldsymbol{0}$. Then, by standard facts about matrices (see, e.g., [27]), the spectral radius of $\boldsymbol{f}'(\boldsymbol{y})$ is less than 1, i.e., $|\lambda| < 1$ for all eigenvalues λ of $\boldsymbol{f}'(\boldsymbol{y})$. This, in turn, implies that the series $\boldsymbol{f}'(\boldsymbol{y})^* = \mathrm{Id} + \boldsymbol{f}'(\boldsymbol{y}) + \boldsymbol{f}'(\boldsymbol{y})^2 + \cdots$ converges in $\mathbb{R}_{\geq 0}$; see [27, p. 531]. In other words, $\boldsymbol{f}'(\boldsymbol{y})^*$ and hence $\boldsymbol{f}'(\boldsymbol{x})^*$ do not have ∞ as an entry. $\quad\square$

The following lemma states that Newton's method can terminate in a component s only after certain other components ℓ have reached $\mu\boldsymbol{f}_\ell$.

LEMMA 4.8. *Let $1 \leq s, \ell \leq n$. Let the term $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{ss}$ contain the variable X_ℓ. Let $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{f}(\boldsymbol{x}) \leq \mu \boldsymbol{f}$ and $x_s < \mu \boldsymbol{f}_s$ and $x_\ell < \mu \boldsymbol{f}_\ell$. Then $\widehat{\mathcal{N}}(\boldsymbol{x})_s < \mu \boldsymbol{f}_s$.*

Proof. This proof follows closely a proof of [16]. Let $d \geq 0$ such that $\left[\boldsymbol{f}'(\boldsymbol{X})^d \right]_{ss}$ contains X_ℓ. Let $m' \geq 0$ such that $\boldsymbol{f}^{m'}(\boldsymbol{x}) \succ \boldsymbol{0}$ and $\boldsymbol{f}^{m'}(\boldsymbol{x})_\ell > x_\ell$. Such an m' exists because with Kleene's theorem the sequence $(\boldsymbol{f}^k(\boldsymbol{x}))_{k \in \mathbb{N}}$ converges to $\mu \boldsymbol{f}$. Notice that our choice of m' guarantees $\left[\boldsymbol{f}'(\boldsymbol{f}^{m'}(\boldsymbol{x}))^d \right]_{ss} > \left[\boldsymbol{f}'(\boldsymbol{x})^d \right]_{ss}$.

Now choose $m \geq m'$ such that $\boldsymbol{f}^{m+1}(\boldsymbol{x})_s > \boldsymbol{f}^m(\boldsymbol{x})_s$. Such an m exists because the sequence $(\boldsymbol{f}^k(\boldsymbol{x})_s)_{k \in \mathbb{N}}$ never reaches $\mu \boldsymbol{f}_s$. This is because s depends on itself (since $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{ss}$ 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

$$
\begin{aligned}
\boldsymbol{f}^{d+m+1}&(\boldsymbol{x}) - \boldsymbol{f}^{d+m}(\boldsymbol{x}) \\
&\geq \boldsymbol{f}'(\boldsymbol{f}^m(\boldsymbol{x}))^d (\boldsymbol{f}^{m+1}(\boldsymbol{x}) - \boldsymbol{f}^m(\boldsymbol{x})) \qquad \text{(Lemma 4.4)} \\
&\geq^* \boldsymbol{f}'(\boldsymbol{x})^d (\boldsymbol{f}^{m+1}(\boldsymbol{x}) - \boldsymbol{f}^m(\boldsymbol{x})) \\
&\geq \boldsymbol{f}'(\boldsymbol{x})^d \boldsymbol{f}'(\boldsymbol{x})^m (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) \qquad \text{(Lemma 4.4)} \\
&= \boldsymbol{f}'(\boldsymbol{x})^{d+m} (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) \,.
\end{aligned}
$$

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

$$
(4.4) \qquad (\boldsymbol{f}^{b+1}(\boldsymbol{x}) - \boldsymbol{f}^b(\boldsymbol{x}))_s > (\boldsymbol{f}'(\boldsymbol{x})^b (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}))_s.
$$

Again by Lemma 4.4, inequality (4.4) holds for all $b \in \mathbb{N}$, but with $\geq$ instead of $>$. Therefore

$$
\begin{aligned}
\mu \boldsymbol{f}_s &= \left(\boldsymbol{x} + \sum_{i=0}^{\infty} (\boldsymbol{f}^{i+1}(\boldsymbol{x}) - \boldsymbol{f}^i(\boldsymbol{x})) \right)_s \qquad \text{(Kleene)} \\
&> \left(\boldsymbol{x} + \boldsymbol{f}'(\boldsymbol{x})^* (\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) \right)_s \qquad \text{(inequality (4.4))} \\
&= \left(\widehat{\mathcal{N}}(\boldsymbol{x}) \right)_s. \qquad \square
\end{aligned}
$$

Now we are ready to prove Proposition 4.3.

Proof of Proposition 4.3. Using Lemma 4.6 it is enough to show that $\left[\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^* \right]_{ss} \neq \infty$ for all s. If the s-component constitutes a trivial SCC, then $\left[\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^* \right]_{ss} = 0 \neq \infty$. So we can assume in the following that the s-component belongs to a nontrivial SCC, say S. Let $\boldsymbol{X}_L$ be the set of variables contained by the term $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{ss}$. For any $t \in S$ we have $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{ss} \geq \left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{st} \left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{tt} \left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{ts}$. Neither $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{st}$ nor $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{ts}$ is constant zero, because S is nontrivial. Therefore, $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{ss}$ contains all variables that $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{tt}$ contains, and vice versa, for all $t \in S$. So, $\boldsymbol{X}_L$ is, for all $t \in S$, exactly the set of variables contained by $\left[\boldsymbol{f}'(\boldsymbol{X})^* \right]_{tt}$.

We distinguish two cases.

Case 1. There is a component $\ell \in L$ such that the sequence $(\nu_\ell^{(k)})_{k \in \mathbb{N}}$ does not terminate, i.e., $\nu_\ell^{(k)} < \mu \boldsymbol{f}_\ell$ holds for all k. Then, by Lemma 4.8, the sequence $(\nu_s^{(k)})_{k \in \mathbb{N}}$ cannot reach $\mu \boldsymbol{f}_s$ either. In fact, we have $\boldsymbol{\nu}_S^{(k)} \prec \mu \boldsymbol{f}_S$. Let M denote the set of those components that the S-components depend on, but which do not depend on S. In other words, M contains the components that are "lower" in the directed acyclic graph (DAG) of SCCs than S. Define $\boldsymbol{g}(\boldsymbol{X}_S) := \boldsymbol{f}_S(\boldsymbol{X})[M/\mu \boldsymbol{f}_M]$. Then $\boldsymbol{g}(\boldsymbol{X}_S)$ is an scSPP with $\mu \boldsymbol{g} = \mu \boldsymbol{f}_S$. As $\boldsymbol{\nu}_S^{(k)} \prec \mu \boldsymbol{g}$, Lemma 4.7 is applicable, so $\boldsymbol{g}'(\boldsymbol{\nu}_S^{(k)})^*$ does

not have ∞ as an entry. With $\left[\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*\right]_{SS} \leq \boldsymbol{g}'(\boldsymbol{\nu}_S^{(k)})^*$, we get $\left[\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*\right]_{ss} < \infty$, as desired.

Case 2. For all components $\ell \in L$ the sequence $(\nu_\ell^{(k)})_{k\in\mathbb{N}}$ terminates. Let $i \in \mathbb{N}$ be the least number such that $\nu_\ell^{(i)} = \mu\boldsymbol{f}_\ell$ holds for all $\ell \in L$. By Lemma 4.8 we have $\nu_s^{(i)} < \mu\boldsymbol{f}_s$. But as, according to Proposition 4.2, $(\nu_s^{(k)})_{k\in\mathbb{N}}$ converges to $\mu\boldsymbol{f}_s$, there must exist a $j \geq i$ such that $0 < \left(\boldsymbol{f}'(\boldsymbol{\nu}^{(j)})^*(\boldsymbol{f}(\boldsymbol{\nu}^{(j)}) - \boldsymbol{\nu}^{(j)})\right)_s < \infty$. So there is a component u with $0 < \left[\boldsymbol{f}'(\boldsymbol{\nu}^{(j)})^*\right]_{su}(\boldsymbol{f}(\boldsymbol{\nu}^{(j)}) - \boldsymbol{\nu}^{(j)})_u < \infty$. This implies $0 < \left[\boldsymbol{f}'(\boldsymbol{\nu}^{(j)})^*\right]_{su} < \infty$, and therefore also $\left[\boldsymbol{f}'(\boldsymbol{\nu}^{(j)})^*\right]_{ss} < \infty$. By monotonicity of $\boldsymbol{f}'$, we have $\left[\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*\right]_{ss} \leq \left[\boldsymbol{f}'(\boldsymbol{\nu}^{(j)})^*\right]_{ss} < \infty$ for all $k \leq j$. On the other hand, since $\left[\boldsymbol{f}'(\boldsymbol{X})^*\right]_{ss}$ contains only L-variables and $\boldsymbol{\nu}_L^{(k)} = \mu\boldsymbol{f}_L$ holds for all $k \geq j$, we also have $\left[\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*\right]_{ss} = \left[\boldsymbol{f}'(\boldsymbol{\nu}^{(j)})^*\right]_{ss} < \infty$ for all $k \geq j$. $\qquad\square$

This completes the second intermediary step towards the proof of Theorem 4.1.

4.1.3. Third and final step. Now we can use Propositions 4.2 and 4.3 to complete the proof of Theorem 4.1.

Proof of Theorem 4.1. By Proposition 4.3 the matrix $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*$ has no ∞ entries. Then we clearly have $\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(k)})) = \mathrm{Id}$, so $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(k)}))^{-1} = \boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*$, which is the first claim of part 2 of the theorem. Hence, we also have

$$
\begin{aligned}
\widehat{\mathcal{N}}(\boldsymbol{\nu}^{(k)}) &= \boldsymbol{\nu}^{(k)} + \sum_{d=0}^{\infty}\left(\boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^d(\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)})\right) \\
&= \boldsymbol{\nu}^{(k)} + \boldsymbol{f}'(\boldsymbol{\nu}^{(k)})^*(\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \\
&= \boldsymbol{\nu}^{(k)} + (\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{\nu}^{(k)}))^{-1}(\boldsymbol{f}(\boldsymbol{\nu}^{(k)}) - \boldsymbol{\nu}^{(k)}) \\
&= \mathcal{N}(\boldsymbol{\nu}^{(k)})\,,
\end{aligned}
$$

so we can replace $\widehat{\mathcal{N}}$ by $\mathcal{N}$. Therefore, part 1 of the theorem is implied by Proposition 4.2. It remains to show $(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x}))^{-1} = \boldsymbol{f}'(\boldsymbol{x})^*$ for all $\boldsymbol{x} \prec \mu\boldsymbol{f}$. It suffices to show that $\boldsymbol{f}'(\boldsymbol{x})^*$ has no ∞ entries. By part 1 the sequence $(\boldsymbol{\nu}^{(k)})_{k\in\mathbb{N}}$ converges to $\mu\boldsymbol{f}$. So there is a k' such that $\boldsymbol{x} \leq \boldsymbol{\nu}^{(k')}$. By Proposition 4.3, $\boldsymbol{f}'(\boldsymbol{\nu}^{(k')})^*$ has no ∞ entries, so, by monotonicity, $\boldsymbol{f}'(\boldsymbol{x})^*$ has no ∞ entries either. $\qquad\square$

4.2. Monotonicity.

LEMMA 4.9 (monotonicity of the Newton operator). *Let $\boldsymbol{f}$ be a clean and feasible SPP. Let $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{y} \leq \boldsymbol{f}(\boldsymbol{y}) \leq \mu\boldsymbol{f}$ and let $\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{y})$ exist. Then*

$$
\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{x}) \leq \mathcal{N}_{\boldsymbol{f}}(\boldsymbol{y})\,.
$$

Proof. For $\boldsymbol{x} \leq \boldsymbol{y}$ we have $\boldsymbol{f}'(\boldsymbol{x}) \leq \boldsymbol{f}'(\boldsymbol{y})$, as every entry of $\boldsymbol{f}'(\boldsymbol{X})$ is a monotone polynomial. Hence, $\boldsymbol{f}'(\boldsymbol{x})^* \leq \boldsymbol{f}'(\boldsymbol{y})^*$. With this at hand we get

$$
\begin{aligned}
\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{y}) &= \boldsymbol{y} + \boldsymbol{f}'(\boldsymbol{y})^*(\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{y}) && \text{(Theorem 4.1)} \\
&\geq \boldsymbol{y} + \boldsymbol{f}'(\boldsymbol{x})^*(\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{y}) && (\boldsymbol{f}'(\boldsymbol{y})^* \geq \boldsymbol{f}'(\boldsymbol{x})^*) \\
&\geq \boldsymbol{y} + \boldsymbol{f}'(\boldsymbol{x})^*(\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x}) - \boldsymbol{y}) && \text{(Lemma 2.3)} \\
&= \boldsymbol{y} + \boldsymbol{f}'(\boldsymbol{x})^*((\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) - (\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x}))(\boldsymbol{y} - \boldsymbol{x})) && \\
&= \boldsymbol{y} + \boldsymbol{f}'(\boldsymbol{x})^*(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) - (\boldsymbol{y} - \boldsymbol{x}) && (\boldsymbol{f}'(\boldsymbol{x})^* = \\
& && (\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x}))^{-1}) \\
&= \mathcal{N}_{\boldsymbol{f}}(\boldsymbol{x}) && \text{(Theorem 4.1).} \qquad\square
\end{aligned}
$$

4.3. Exponential convergence order in the nonsingular case. If the matrix $\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})$ is nonsingular, Newton's method has exponential convergence order in the sense of Definition 2.7.

Define for the following lemmas $\boldsymbol{\Delta}^{(k)} := \mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)}$; i.e., $\boldsymbol{\Delta}^{(k)}$ is the error after k Newton iterations. The following lemma bounds $\|\boldsymbol{\Delta}^{(k+1)}\|$ in terms of $\|\boldsymbol{\Delta}^{(k)}\|^2$ if $\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})$ is nonsingular.

LEMMA 4.10. *Let $\boldsymbol{f}$ be a clean and feasible SPP such that $\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})$ is non-singular. Then there is a constant $c > 0$ such that*

$$\left\|\boldsymbol{\Delta}^{(k+1)}\right\|_\infty \le c \cdot \left\|\boldsymbol{\Delta}^{(k)}\right\|_\infty^2 \quad \text{for all } k \in \mathbb{N}.$$

Proof. See [10] or Theorem 4.4 of [34]. □

Lemma 4.10 implies that Newton's method has an exponential convergence order in the nonsingular case. We state this more precisely in the following theorem.

THEOREM 4.11. *Let $\boldsymbol{f}$ be a clean and feasible SPP such that $\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})$ is nonsingular. Then there is a constant $k_{\boldsymbol{f}} \in \mathbb{N}$ such that*

$$\beta(k_{\boldsymbol{f}} + i) \ge 2^i \text{ for all } i \in \mathbb{N}.$$

Proof. We first show that there is a constant $\widetilde{k}_{\boldsymbol{f}} \in \mathbb{N}$ such that

$$(4.5) \qquad \left\|\boldsymbol{\Delta}^{(\widetilde{k}_{\boldsymbol{f}}+i)}\right\|_\infty \le 2^{-2^i} \text{ for all } i \in \mathbb{N}.$$

We can assume w.l.o.g. that $c \ge 1$ for the c from Lemma 4.10. As the $\boldsymbol{\Delta}^{(k)}$ converge to $\boldsymbol{0}$, we can choose $\widetilde{k}_{\boldsymbol{f}} \in \mathbb{N}$ large enough such that $d := -\log\|\boldsymbol{\Delta}^{(\widetilde{k}_{\boldsymbol{f}})}\| - \log c \ge 1$. As $c, d \ge 1$, it suffices to show the following inequality:

$$\left\|\boldsymbol{\Delta}^{(\widetilde{k}_{\boldsymbol{f}}+i)}\right\| \le \frac{2^{-d \cdot 2^i}}{c}.$$

We proceed by induction on i. For $i = 0$, the inequality above follows from the definition of d. Let $i \ge 0$. Then

$$\begin{aligned}
\left\|\boldsymbol{\Delta}^{(\widetilde{k}_{\boldsymbol{f}}+i+1)}\right\| &\le c \cdot \left\|\boldsymbol{\Delta}^{(\widetilde{k}_{\boldsymbol{f}}+i)}\right\|^2 && \text{(Lemma 4.10)} \\
&\le c \cdot \frac{2^{-d \cdot 2^i \cdot 2}}{c^2} && \text{(induction hypothesis)} \\
&= \frac{2^{-d \cdot 2^{i+1}}}{c}.
\end{aligned}$$

Hence, (4.5) is proved.

Choose $m \in \mathbb{N}$ large enough such that $2^{m+i} + \log(\mu\boldsymbol{f}_j) \ge 2^i$ holds for all components j. Thus

$$\begin{aligned}
\Delta_j^{(\widetilde{k}_{\boldsymbol{f}}+m+i)}/\mu\boldsymbol{f}_j &\le 2^{-2^{m+i}}/\mu\boldsymbol{f}_j && \text{(by (4.5))} \\
&= 2^{-(2^{m+i}+\log(\mu\boldsymbol{f}_j))} \\
&\le 2^{-2^i} && \text{(choice of } m).
\end{aligned}$$

So, with $k_{\boldsymbol{f}} := \widetilde{k}_{\boldsymbol{f}} + m$, the approximant $\boldsymbol{\nu}^{(k_{\boldsymbol{f}}+i)}$ has at least 2^i valid bits of $\mu\boldsymbol{f}$. □

This type of analysis has serious shortcomings. In particular, Theorem 4.11 excludes the case where $\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})$ is singular. We will include this case in our convergence analysis in sections 5 and 6. Furthermore, and maybe more severely, Theorem 4.11 does not give any bound on $k_{\boldsymbol{f}}$. We solve this problem for strongly connected SPPs in section 5.

4.4. Reduction to the quadratic case. In this section we reduce SPPs to quadratic SPPs, i.e., to SPPs in which every polynomial $f_i(\boldsymbol{X})$ has degree at most 2, and show that the convergence on the quadratic SPP is no faster than on the original SPP. In the following sections we will obtain convergence speed guarantees of Newton's method on quadratic SPPs. Hence, one can perform Newton's method on the original SPP, and, using the results of this section, convergence is at least as fast as on the corresponding quadratic SPP.

The idea to reduce the degree of our SPP $\boldsymbol{f}$ 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 4.12. *Let $\boldsymbol{f}(\boldsymbol{X})$ be a clean and feasible SPP such that $f_s(\boldsymbol{X}) = g(\boldsymbol{X}) + h(\boldsymbol{X})X_iX_j$ for some $1 \leq i,j,s \leq n$, where $g(\boldsymbol{X})$ and $h(\boldsymbol{X})$ are polynomials with nonnegative coefficients. Let $\widetilde{\boldsymbol{f}}(\boldsymbol{X},Y)$ be the SPP given by*

$$
\begin{aligned}
\widetilde{f}_\ell(\boldsymbol{X},Y) &= f_\ell(\boldsymbol{X}) && \text{for every } \ell \in \{1,\ldots,s-1\}, \\
\widetilde{f}_s(\boldsymbol{X},Y) &= g(\boldsymbol{X}) + h(\boldsymbol{X})Y, \\
\widetilde{f}_\ell(\boldsymbol{X},Y) &= f_\ell(\boldsymbol{X}) && \text{for every } \ell \in \{s+1,\ldots,n\}, \\
\widetilde{f}_{n+1}(\boldsymbol{X},Y) &= X_iX_j.
\end{aligned}
$$

Then the function $b: \mathbb{R}^n \to \mathbb{R}^{n+1}$ given by $b(\boldsymbol{X}) = (X_1,\ldots,X_n,X_iX_j)^\top$ is a bijection between the set of fixed points of $\boldsymbol{f}(\boldsymbol{X})$ and $\widetilde{\boldsymbol{f}}(\boldsymbol{X},Y)$. 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 approximants of $\widetilde{\boldsymbol{f}}$ and $\boldsymbol{f}$, respectively.

Proof. We first show the claim regarding b: if $\boldsymbol{x}$ is a fixed point of $\boldsymbol{f}$, then $b(\boldsymbol{x}) = (\boldsymbol{x}, x_ix_j)$ is a fixed point of $\widetilde{\boldsymbol{f}}$. Conversely, if $(\boldsymbol{x}, y)$ is a fixed point of $\widetilde{\boldsymbol{f}}$, then we have $y = x_ix_j$, implying that $\boldsymbol{x}$ is a fixed point of $\boldsymbol{f}$. Therefore, the least fixed point $\mu\boldsymbol{f}$ of $\boldsymbol{f}$ determines $\mu\widetilde{\boldsymbol{f}}$, and vice versa.

Now we show that the Newton sequence of $\boldsymbol{f}$ converges at least as fast as the Newton sequence of $\widetilde{\boldsymbol{f}}$. In the following we write $\boldsymbol{Y}$ for the $(n+1)$-dimensional vector of variables $(X_1,\ldots,X_n,Y)^\top$ and, as usual, $\boldsymbol{X}$ for $(X_1,\ldots,X_n)^\top$. For an $(n+1)$-dimensional vector $\boldsymbol{x}$, we let $\boldsymbol{x}_{[1,n]}$ denote its restriction to the n first components, i.e., $\boldsymbol{x}_{[1,n]} := (x_1,\ldots,x_n)^\top$. Note that $\boldsymbol{Y}_{[1,n]} = \boldsymbol{X}$. Let $\boldsymbol{e}_s$ denote the unit vector $(0,\ldots,0,1,0\ldots0)^\top$, where the "1" is on the sth place. We have

$$
\widetilde{\boldsymbol{f}}(\boldsymbol{Y}) = \begin{pmatrix} \boldsymbol{f}(\boldsymbol{X}) + \boldsymbol{e}_s h(\boldsymbol{X})(Y - X_iX_j) \\ X_iX_j \end{pmatrix}
$$

and

$$
\widetilde{\boldsymbol{f}}'(\boldsymbol{Y}) = \begin{pmatrix} \boldsymbol{f}'(\boldsymbol{X}) + \boldsymbol{e}_s\partial_{\boldsymbol{X}}h(\boldsymbol{X})(Y - X_iX_j) & \boldsymbol{e}_s h(\boldsymbol{X}) \\ \partial_{\boldsymbol{X}}X_iX_j & 0 \end{pmatrix}.
$$

We need the following lemma.

LEMMA 4.13.   *Let* $\boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$, $\boldsymbol{\delta} = \left(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{z})\right)^{-1}(\boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{z})$, *and* $\widetilde{\boldsymbol{\delta}} = \left(\mathrm{Id} - \widetilde{\boldsymbol{f}}'(\boldsymbol{z}, z_i z_j)\right)^{-1}(\widetilde{\boldsymbol{f}}(\boldsymbol{z}, z_i z_j) - (\boldsymbol{z}, z_i z_j)^\top)$. *Then* $\boldsymbol{\delta} = \widetilde{\boldsymbol{\delta}}_{[1,n]}$.

Proof.

$$\widetilde{\boldsymbol{f}}'(\boldsymbol{z}, z_i z_j) = \begin{pmatrix} \boldsymbol{f}'(\boldsymbol{z}) + \boldsymbol{e}_s h(\boldsymbol{z}) \partial_{\boldsymbol{X}} (Y - X_i X_j)|_{\boldsymbol{Y}=(\boldsymbol{z}, z_i z_j)} & \boldsymbol{e}_s h(\boldsymbol{z}) \\ \partial_{\boldsymbol{X}} X_i X_j|_{\boldsymbol{Y}=(\boldsymbol{z}, z_i z_j)} & 0 \end{pmatrix}$$
$$= \begin{pmatrix} \boldsymbol{f}'(\boldsymbol{z}) - \boldsymbol{e}_s h(\boldsymbol{z}) \partial_{\boldsymbol{X}} (X_i X_j)|_{\boldsymbol{X}=\boldsymbol{z}} & \boldsymbol{e}_s h(\boldsymbol{z}) \\ \partial_{\boldsymbol{X}} X_i X_j|_{\boldsymbol{X}=\boldsymbol{z}} & 0 \end{pmatrix}.$$

We have $(\mathrm{Id} - \widetilde{\boldsymbol{f}}'(\boldsymbol{z}, z_i z_j))\widetilde{\boldsymbol{\delta}} = (\widetilde{\boldsymbol{f}}(\boldsymbol{z}, z_i z_j) - (\boldsymbol{z}, z_i z_j)^\top)$, or equivalently

$$\begin{pmatrix} \mathrm{Id} - \boldsymbol{f}'(\boldsymbol{z}) + \boldsymbol{e}_s h(\boldsymbol{z}) \partial_{\boldsymbol{X}} (X_i X_j)|_{\boldsymbol{X}=\boldsymbol{z}} & -\boldsymbol{e}_s h(\boldsymbol{z}) \\ -\partial_{\boldsymbol{X}} X_i X_j|_{\boldsymbol{X}=\boldsymbol{z}} & 1 \end{pmatrix} \cdot \begin{pmatrix} \widetilde{\boldsymbol{\delta}}_{[1,n]} \\ \widetilde{\delta}_{n+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{z} \\ 0 \end{pmatrix}.$$

Multiplying the last row by $\boldsymbol{e}_s h(\boldsymbol{z})$ and adding to the first n rows yields

$$\left(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{z})\right) \widetilde{\boldsymbol{\delta}}_{[1,n]} = \boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{z}.$$

So we have $\widetilde{\boldsymbol{\delta}}_{[1,n]} = \left(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{z})\right)^{-1} \left(\boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{z}\right) = \boldsymbol{\delta}$, which proves the lemma.   □

Now we proceed by induction on k to show $\widetilde{\boldsymbol{\nu}}^{(k)}_{[1,n]} \leq \boldsymbol{\nu}^{(k)}$, where $\widetilde{\boldsymbol{\nu}}^{(k)}$ is the Newton sequence for $\widetilde{\boldsymbol{f}}$. By definition of the Newton sequence, this is true for $k = 0$. For the step, let $k \geq 0$ and define $\boldsymbol{u} := (\widetilde{\boldsymbol{\nu}}^{(k)}_{[1,n]}, \widetilde{\nu}^{(k)}_i \cdot \widetilde{\nu}^{(k)}_j)^\top$. Then we have

$$
\begin{aligned}
\widetilde{\boldsymbol{\nu}}^{(k+1)}_{[1,n]} &= \mathcal{N}_{\widetilde{\boldsymbol{f}}}(\widetilde{\boldsymbol{\nu}}^{(k)})_{[1,n]} \\
&\overset{(*)}{\leq} \mathcal{N}_{\widetilde{\boldsymbol{f}}}(\boldsymbol{u})_{[1,n]} && \text{(see below)} \\
&= \widetilde{\boldsymbol{\nu}}^{(k)}_{[1,n]} + \left((\mathrm{Id} - \widetilde{\boldsymbol{f}}'(\boldsymbol{u}))^{-1}(\widetilde{\boldsymbol{f}}(\boldsymbol{u}) - \boldsymbol{u})\right)_{[1,n]} \\
&= \widetilde{\boldsymbol{\nu}}^{(k)}_{[1,n]} + (\mathrm{Id} - \boldsymbol{f}'(\widetilde{\boldsymbol{\nu}}^{(k)}_{[1,n]}))^{-1}(\boldsymbol{f}(\widetilde{\boldsymbol{\nu}}^{(k)}_{[1,n]}) - \widetilde{\boldsymbol{\nu}}^{(k)}_{[1,n]}) && \text{(Lemma 4.13)} \\
&= \mathcal{N}_{\boldsymbol{f}}(\widetilde{\boldsymbol{\nu}}^{(k)}_{[1,n]}) \\
&\leq \mathcal{N}_{\boldsymbol{f}}(\boldsymbol{\nu}^{(k)}) && \text{(induction)} \\
&= \boldsymbol{\nu}^{(k+1)}.
\end{aligned}
$$

At the inequality marked with $(*)$ we used the monotonicity of $\mathcal{N}_{\widetilde{\boldsymbol{f}}}$ (Lemma 4.9) combined with Theorem 4.1, which states $\widetilde{\boldsymbol{\nu}}^{(k)} \leq \widetilde{\boldsymbol{f}}(\widetilde{\boldsymbol{\nu}}^{(k)})$; hence in particular $\widetilde{\nu}^{(k)}_{n+1} \leq \widetilde{\nu}^{(k)}_i \widetilde{\nu}^{(k)}_j$. This concludes the proof of Theorem 4.12.   □

5. Strongly connected SPPs. In this section we study the convergence speed of Newton's method on strongly connected SPPs (scSPPs); see Definition 2.6.

5.1. Cone vectors. Our convergence speed analysis makes crucial use of the existence of *cone vectors*.

DEFINITION 5.1.   *Let* $\boldsymbol{f}$ *be an SPP. A vector* $\boldsymbol{d} \in \mathbb{R}^n_{\geq 0}$ *is a* cone vector *if* $\boldsymbol{d} \succ \boldsymbol{0}$ *and* $\boldsymbol{f}'(\mu \boldsymbol{f})\boldsymbol{d} \leq \boldsymbol{d}$.

PROPOSITION 5.2.   *Any clean and feasible scSPP has a cone vector.*

Proof. By Theorem 4.1 $\boldsymbol{f}'(\boldsymbol{x})^*$ exists for all $\boldsymbol{x} \prec \mu\boldsymbol{f}$. So, by fundamental matrix facts [3], the spectral radius of $\boldsymbol{f}'(\boldsymbol{x})$ is less than 1 for all $\boldsymbol{x} \prec \mu\boldsymbol{f}$. As the eigenvalues of a matrix depend continuously on the matrix, the spectral radius of $\boldsymbol{f}'(\mu\boldsymbol{f})$, say ρ, is at most 1. Since $\boldsymbol{f}$ is strongly connected, $\boldsymbol{f}'(\mu\boldsymbol{f})$ is irreducible, and so Perron–Frobenius theory guarantees the existence of an eigenvector $\boldsymbol{d} \succ \boldsymbol{0}$ of $\boldsymbol{f}'(\mu\boldsymbol{f})$ with eigenvalue ρ. So we have $\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} = \rho\boldsymbol{d} \le \boldsymbol{d}$, i.e., the eigenvector $\boldsymbol{d}$ is a cone vector. $\square$

In [10] we prove Proposition 5.2 independently of Perron–Frobenius theory.

5.2. Convergence speed in terms of cone vectors. Now we show that cone vectors play a fundamental role for the convergence speed of Newton's method. The following lemma gives a lower bound of the Newton approximant $\boldsymbol{\nu}^{(1)}$ in terms of a cone vector.

LEMMA 5.3. *Let $\boldsymbol{f}$ be a feasible (not necessarily clean) SPP such that $\boldsymbol{f}'(\boldsymbol{0})^*$ exists. Let $\boldsymbol{d}$ be a cone vector of $\boldsymbol{f}$. Let $\boldsymbol{0} \ge \mu\boldsymbol{f} - \lambda\boldsymbol{d}$ for some $\lambda \ge 0$. Then*

$$\mathcal{N}(\boldsymbol{0}) \ge \mu\boldsymbol{f} - \frac{1}{2}\lambda\boldsymbol{d} \ .$$

Proof. We write $\boldsymbol{f}(\boldsymbol{X})$ as a sum $\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{c} + \sum_{k=1}^{D} \boldsymbol{T}^{(k)}(\boldsymbol{X}, \ldots, \boldsymbol{X})$, where D is the degree of $\boldsymbol{f}$ and, for all $k \in \{1, \ldots, D\}$ and all $i \in \{1, \ldots, n\}$, the component $T_i^{(k)}$ of $\boldsymbol{T}^{(k)}$ is the symmetric k-linear form associated to the degree-k terms of f_i. Let $L^{(k)} : (\mathbb{R}^n)^{k-1} \to \mathbb{R}^{n \times n}$ such that $\boldsymbol{T}^{(k)}(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(k)}) = L^{(k)}(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(k-1)}) \cdot \boldsymbol{X}^{(k)}$. Now we can write

$$\boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{c} + \sum_{k=1}^{D} L^{(k)}(\boldsymbol{X}, \ldots, \boldsymbol{X})\boldsymbol{X} \quad \text{and} \quad \boldsymbol{f}'(\boldsymbol{X}) = \sum_{k=1}^{D} k \cdot L^{(k)}(\boldsymbol{X}, \ldots, \boldsymbol{X}) \ .$$

We write L for $L^{(1)}$, and $\boldsymbol{h}(\boldsymbol{X})$ for $\boldsymbol{f}(\boldsymbol{X}) - L\boldsymbol{X} - \boldsymbol{c}$. We have

$$
\begin{aligned}
\frac{\lambda}{2}\boldsymbol{d} &= \frac{\lambda}{2}(L^*\boldsymbol{d} - L^*L\boldsymbol{d}) && (L^* = \mathrm{Id} + L^*L) \\
&\ge \frac{\lambda}{2}(L^*\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} - L^*L\boldsymbol{d}) && (\boldsymbol{f}'(\mu\boldsymbol{f})\boldsymbol{d} \le \boldsymbol{d}) \\
&= \frac{\lambda}{2}L^*\boldsymbol{h}'(\mu\boldsymbol{f})\boldsymbol{d} && (\boldsymbol{f}'(\boldsymbol{x}) = \boldsymbol{h}'(\boldsymbol{x}) + L) \\
&= L^*\frac{1}{2}\boldsymbol{h}'(\mu\boldsymbol{f})\lambda\boldsymbol{d} && \\
&\ge L^*\frac{1}{2}\boldsymbol{h}'(\mu\boldsymbol{f})\mu\boldsymbol{f} && (\lambda\boldsymbol{d} \ge \mu\boldsymbol{f}) \\
&= L^*\frac{1}{2}\sum_{k=2}^{D} k \cdot L^{(k)}(\mu\boldsymbol{f}, \ldots, \mu\boldsymbol{f})\mu\boldsymbol{f} && \\
&\ge L^*\sum_{k=2}^{D} L^{(k)}(\mu\boldsymbol{f}, \ldots, \mu\boldsymbol{f})\mu\boldsymbol{f} && \\
&= L^*\boldsymbol{h}(\mu\boldsymbol{f}) && \\
&= L^*(\boldsymbol{f}(\mu\boldsymbol{f}) - L\mu\boldsymbol{f} - \boldsymbol{c}) && (\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x}) + L\boldsymbol{x} + \boldsymbol{c}) \\
&= L^*\mu\boldsymbol{f} - L^*L\mu\boldsymbol{f} - L^*\boldsymbol{c} && (\boldsymbol{f}(\mu\boldsymbol{f}) = \mu\boldsymbol{f}) \\
&= \mu\boldsymbol{f} - L^*\boldsymbol{c} && (L^* = \mathrm{Id} + L^*L) \\
&= \mu\boldsymbol{f} - \mathcal{N}(\boldsymbol{0}) && (\mathcal{N}(\boldsymbol{0}) = \boldsymbol{f}'(\boldsymbol{0})^*\boldsymbol{f}(\boldsymbol{0}) = L^*\boldsymbol{c}). \quad \square
\end{aligned}
$$

We extend Lemma 5.3 to arbitrary vectors $\boldsymbol{x}$ as follows.

LEMMA 5.4. *Let $\boldsymbol{f}$ be a feasible (not necessarily clean) SPP. Let $\boldsymbol{0} \le \boldsymbol{x} \le \mu\boldsymbol{f}$ and $\boldsymbol{x} \le \boldsymbol{f}(\boldsymbol{x})$ such that $\boldsymbol{f}'(\boldsymbol{x})^*$ exists. Let $\boldsymbol{d}$ be a cone vector of $\boldsymbol{f}$. Let $\boldsymbol{x} \ge \mu\boldsymbol{f} - \lambda\boldsymbol{d}$ for some $\lambda \ge 0$. Then*

$$\mathcal{N}(\boldsymbol{x}) \ge \mu\boldsymbol{f} - \frac{1}{2}\lambda\boldsymbol{d} .$$

Proof. Define $\boldsymbol{g}(\boldsymbol{X}) := \boldsymbol{f}(\boldsymbol{X} + \boldsymbol{x}) - \boldsymbol{x}$. We first show that $\boldsymbol{g}$ is an SPP (not necessarily clean). The only coefficients of $\boldsymbol{g}$ that could be negative are those of degree 0. But we have $\boldsymbol{g}(\boldsymbol{0}) = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x} \ge \boldsymbol{0}$, and so these coefficients are also nonnegative.

It follows immediately from the definition that $\mu\boldsymbol{f} - \boldsymbol{x} \ge \boldsymbol{0}$ is the least fixed point of $\boldsymbol{g}$. Moreover, $\boldsymbol{g}$ satisfies $\boldsymbol{g}'(\mu\boldsymbol{f} - \boldsymbol{x})\boldsymbol{d} \le \boldsymbol{d}$, and so $\boldsymbol{d}$ is also a cone vector of $\boldsymbol{g}$. Finally, we have $\boldsymbol{0} \ge \mu\boldsymbol{f} - \boldsymbol{x} - \lambda\boldsymbol{d} = \mu\boldsymbol{g} - \lambda\boldsymbol{d}$. So, Lemma 5.3 can be applied as follows.

$$\begin{aligned}
\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{x}) &= \boldsymbol{x} + \boldsymbol{f}'(\boldsymbol{x})^*(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}) \\
&= \boldsymbol{x} + \boldsymbol{g}'(\boldsymbol{0})^*(\boldsymbol{g}(\boldsymbol{0}) - \boldsymbol{0}) \\
&= \boldsymbol{x} + \mathcal{N}_{\boldsymbol{g}}(\boldsymbol{0}) \\
&\ge \boldsymbol{x} + \mu\boldsymbol{g} - \frac{1}{2}\lambda\boldsymbol{d} \qquad\qquad \text{(Lemma 5.3)}\\
&= \mu\boldsymbol{f} - \frac{1}{2}\lambda\boldsymbol{d}. \qquad \square
\end{aligned}$$

By induction we can extend this lemma to the whole Newton sequence.

LEMMA 5.5. *Let $\boldsymbol{d}$ be a cone vector of a clean and feasible SPP $\boldsymbol{f}$ and let $\lambda_{max} = \max_j\{\frac{\mu\boldsymbol{f}_j}{d_j}\}$. Then*

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

Before proving the lemma we illustrate it in the graph shown in Figure 5.1. The dashed line in Figure 5.1 is the ray $\boldsymbol{r}(t) = \mu\boldsymbol{f} - t\boldsymbol{d}$ along a cone vector $\boldsymbol{d}$. Notice that $\boldsymbol{r}(0)$ equals $\mu\boldsymbol{f}$ and that $\boldsymbol{r}(\lambda_{max})$ is the greatest point on the ray that is below $\boldsymbol{0}$. The figure also shows the Newton iterates $\boldsymbol{\nu}^{(k)}$ for $0 \le k \le 2$ (shape: $\times$) and the corresponding points $\boldsymbol{r}(2^{-k}\lambda_{max})$ (shape: $+$) located on the ray $\boldsymbol{r}$. Observe that $\boldsymbol{\nu}^{(k)} \ge \boldsymbol{r}(2^{-k}\lambda_{max})$, as claimed by Lemma 5.5.

Proof of Lemma 5.5. The proof is by induction on k. For the induction base ($k = 0$) we have for all components i

$$(\mu\boldsymbol{f} - \lambda_{max}\boldsymbol{d})_i = \left(\mu\boldsymbol{f} - \max_j\left\{\frac{\mu\boldsymbol{f}_j}{d_j}\right\}\boldsymbol{d}\right)_i \le \mu\boldsymbol{f}_i - \frac{\mu\boldsymbol{f}_i}{d_i}d_i = 0 ,$$

so $\boldsymbol{\nu}^{(0)} = \boldsymbol{0} \ge \mu\boldsymbol{f} - \lambda_{max}\boldsymbol{d}$.

For the induction step, let $k \ge 0$. By the induction hypothesis we have $\boldsymbol{\nu}^{(k)} \ge \mu\boldsymbol{f} - 2^{-k}\lambda_{max}\boldsymbol{d}$. So we can apply Lemma 5.4 to get

$$\boldsymbol{\nu}^{(k+1)} = \mathcal{N}(\boldsymbol{\nu}^{(k)}) \ge \mu\boldsymbol{f} - \frac{1}{2}2^{-k}\lambda_{max}\boldsymbol{d} = \mu\boldsymbol{f} - 2^{-(k+1)}\lambda_{max}\boldsymbol{d} . \qquad \square$$

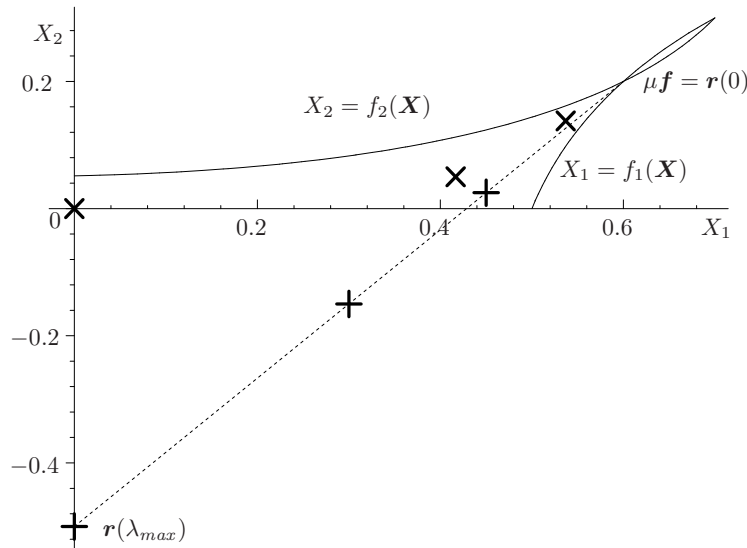The following proposition guarantees a convergence order of the Newton sequence in terms of a cone vector.

FIG. 5.1. *Illustration of Lemma 5.5: The points (+) on the ray r along a cone vector are lower bounds on the Newton approximants (×).*

PROPOSITION 5.6. *Let d be a cone vector of a clean and feasible SPP f and let $\lambda_{max} = \max_j \{\frac{\mu f_j}{d_j}\}$ and $\lambda_{min} = \min_j \{\frac{\mu f_j}{d_j}\}$. Let $k_{f,d} = \lceil \log \frac{\lambda_{max}}{\lambda_{min}} \rceil$. Then $\beta(k_{f,d} + i) \geq i$ for all $i \in \mathbb{N}$.*

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

$$\left(\mu f - \nu^{(k_{f,d}+i)}\right)_j \leq 2^{-(k_{f,d}+i)} \lambda_{max} d_j \qquad \text{(Lemma 5.5)}$$

$$\leq \frac{\lambda_{min}}{\lambda_{max}} 2^{-i} \lambda_{max} d_j \qquad \text{(def. of } k_{f,d})$$

$$= \lambda_{min} d_j \cdot 2^{-i}$$

$$\leq \mu f_j \cdot 2^{-i} \qquad \text{(def. of } \lambda_{min}).$$

Hence, $\nu^{(k_{f,d}+i)}$ has i valid bits of μf. ☐

5.3. Convergence speed independent from cone vectors. The convergence order provided by Proposition 5.6 depends on a cone vector d. While Proposition 5.2 guarantees the existence of a cone vector for scSPPs, it does not give any information on the magnitude of its components. So we do not have any bound yet on the "threshold" $k_{f,d}$ from Proposition 5.6. The following theorem solves this problem.

THEOREM 5.7. *Let f be a quadratic, clean, and feasible scSPP. Let c_{min} be the smallest nonzero coefficient of f and let μ_{min} and μ_{max} be the minimal and maximal component of μf, respectively. Let*

$$k_f = \left\lceil \log \frac{\mu_{max}}{\mu_{min} \cdot (c_{min} \cdot \min\{\mu_{min}, 1\})^n} \right\rceil .$$

Then

$$\beta(k_f + i) \geq i \text{ for all } i \in \mathbb{N}.$$

Before we prove Theorem 5.7 we give an example.

Example 5.8. As an example of an application of Theorem 5.7, consider the scSPP equation of the back-button process of Example 2.9:

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

We wish to know if there is a component $s \in \{1, 2, 3\}$ with $\mu\boldsymbol{f}_s = 1$. Notice that $\boldsymbol{f}(\mathbf{1}) = \mathbf{1}$, so $\mu\boldsymbol{f} \leq \mathbf{1}$. Performing 14 Newton steps (e.g., with Maple) yields an approximation $\boldsymbol{\nu}^{(14)}$ to $\mu\boldsymbol{f}$ 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 $c_{min} = 0.3$. In addition, since Newton's method converges to $\mu\boldsymbol{f}$ from below, we know $\mu_{min} \geq 0.97$. Moreover, $\mu_{max} \leq 1$, as $\mathbf{1} = \boldsymbol{f}(\mathbf{1})$ and so $\mu\boldsymbol{f} \leq \mathbf{1}$. Hence $k_{\boldsymbol{f}} \leq \left\lceil \log \frac{1}{0.97 \cdot (0.3 \cdot 0.97)^3} \right\rceil = 6$. Theorem 5.7 then implies that $\boldsymbol{\nu}^{(14)}$ has 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

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

So Theorem 5.7 yields a proof that $\mu\boldsymbol{f}_s < 1$ for all three components s.

Notice also that the Newton sequence converges much faster than the Kleene sequence $(\boldsymbol{\kappa}^{(k)})_{k \in \mathbb{N}}$. We have $\boldsymbol{\kappa}^{(14)} \prec (0.89, 0.83, 0.96)^{\top}$, so $\boldsymbol{\kappa}^{(14)}$ has no more than 4 valid bits in any component, whereas $\boldsymbol{\nu}^{(14)}$ has, in fact, more than 30 valid bits in each component.

For the proof of Theorem 5.7 we need the following lemma.

LEMMA 5.9. *Let $\boldsymbol{d}$ be a cone vector of a quadratic, clean, and feasible scSPP $\boldsymbol{f}$. Let c_{min} be the smallest nonzero coefficient of $\boldsymbol{f}$ and μ_{min} the minimal component of $\mu\boldsymbol{f}$. Let d_{min} and d_{max} be the smallest and the largest component of $\boldsymbol{d}$, respectively. Then*

$$\frac{d_{min}}{d_{max}} \geq (c_{min} \cdot \min\{\mu_{min}, 1\})^n .$$

Proof. In what follows we shorten $\mu\boldsymbol{f}$ to $\boldsymbol{\mu}$. W.l.o.g. let $d_1 = d_{max}$ and $d_n = d_{min}$. We claim the existence of indices s, t with $1 \leq s, t \leq n$ such that $\boldsymbol{f}'_{st}(\boldsymbol{\mu}) \neq 0$ and

(5.1)
$$\frac{d_{min}}{d_{max}} \geq \left(\frac{d_s}{d_t} \right)^n .$$

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 $q \leq n$ such that $\boldsymbol{f}'_{r_{j+1}r_j}(\boldsymbol{X})$ is not constant zero. As $\boldsymbol{\mu} \succ \mathbf{0}$, we have $\boldsymbol{f}'_{r_{j+1}r_j}(\boldsymbol{\mu}) \neq 0$. Furthermore

$$\frac{d_1}{d_n} = \frac{d_{r_1}}{d_{r_2}} \cdots \frac{d_{r_{q-1}}}{d_{r_q}} , \text{ and so}$$

$$\log \frac{d_1}{d_n} = \log \frac{d_{r_1}}{d_{r_2}} + \cdots + \log \frac{d_{r_{q-1}}}{d_{r_q}} .$$

So there must exist a j such that

$$\log \frac{d_1}{d_n} \le (q-1) \log \frac{d_{r_j}}{d_{r_{j+1}}} \le n \log \frac{d_{r_j}}{d_{r_{j+1}}}, \quad \text{and so}$$

$$\frac{d_n}{d_1} \ge \left(\frac{d_{r_{j+1}}}{d_{r_j}} \right)^n .$$

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

As $\boldsymbol{d}$ is a cone vector, we have $\boldsymbol{f}'(\boldsymbol{\mu})\boldsymbol{d} \le \boldsymbol{d}$ and thus $\boldsymbol{f}'_{st}(\boldsymbol{\mu})d_t \le d_s$. Hence

(5.2) $$\boldsymbol{f}'_{st}(\boldsymbol{\mu}) \le \frac{d_s}{d_t} .$$

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

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

where $b_1, \ldots, b_n$ and ℓ are coefficients of quadratic, respectively, linear, monomials of $\boldsymbol{f}$. As $\boldsymbol{f}'_{st}(\boldsymbol{\mu}) \ne 0$, at least one of these coefficients must be nonzero and so greater than or equal to c_{min}. It follows that $\boldsymbol{f}'_{st}(\boldsymbol{\mu}) \ge c_{min} \cdot \min\{\mu_{min}, 1\}$. So we have

$$(c_{min} \cdot \min\{\mu_{min}, 1\})^n \le \left(\boldsymbol{f}'_{st}(\boldsymbol{\mu}) \right)^n$$
$$\le \left(\frac{d_s}{d_t} \right)^n \qquad \text{(by (5.2))}$$
$$\le \frac{d_{min}}{d_{max}} \qquad \text{(by (5.1)) .} \qquad \square$$

Now we can prove Theorem 5.7.

Proof of Theorem 5.7. By Proposition 5.2, $\boldsymbol{f}$ has a cone vector $\boldsymbol{d}$. Let $d_{max} = \max_j\{d_j\}$, $d_{min} = \min_j\{d_j\}$, $\lambda_{max} = \max_j \left\{ \frac{\boldsymbol{\mu f}_j}{d_j} \right\}$, and $\lambda_{min} = \min_j \left\{ \frac{\boldsymbol{\mu f}_j}{d_j} \right\}$. We have

$$\frac{\lambda_{max}}{\lambda_{min}} \le \frac{\mu_{max} \cdot d_{max}}{\mu_{min} \cdot d_{min}} \qquad \left(\text{as } \lambda_{max} \le \frac{d_{max}}{\mu_{min}} \text{ and } \lambda_{min} \ge \frac{d_{min}}{\mu_{max}} \right)$$
$$\le \frac{\mu_{max}}{\mu_{min} \cdot (c_{min} \cdot \min\{\mu_{min}, 1\})^n} \quad \text{(Lemma 5.9) .}$$

So the statement follows with Proposition 5.6. $\square$

The following consequence of Theorem 5.7 removes some of the parameters on which the $k_{\boldsymbol{f}}$ from Theorem 5.7 depends.

THEOREM 5.10. *Let $\boldsymbol{f}$ be a quadratic, clean, and feasible scSPP, let μ_{min} and μ_{max} be the minimal and maximal component of $\mu\boldsymbol{f}$, respectively, and let the coefficients of $\boldsymbol{f}$ be given as ratios of m-bit integers. Then*

$$\beta(k_{\boldsymbol{f}} + i) \ge i \text{ for all } i \in \mathbb{N}$$

holds for any of the following choices of $k_{\boldsymbol{f}}$:
  1. $\lceil 4mn + 3n \max\{0, -\log \mu_{min}\} \rceil$.
  2. $4mn2^n$.

3. $7mn$ whenever $\boldsymbol{f(0)} \succ \boldsymbol{0}$.
4. $2mn + m$ whenever both $\boldsymbol{f(0)} \succ \boldsymbol{0}$ and $\mu_{max} \leq 1$.

Items 3 and 4 of Theorem 5.10 apply in particular to termination SPPs of strict pPDAs (section 2.4); i.e., they satisfy $\boldsymbol{f(0)} \succ \boldsymbol{0}$ and $\mu_{max} \leq 1$.

To prove Theorem 5.10 we need some relations between the parameters of $\boldsymbol{f}$. We collect them in the following lemma.

LEMMA 5.11. *Let $\boldsymbol{f}$ be a quadratic, clean, and feasible scSPP. With the terminology of Theorems 5.7 and 5.10 the following relations hold:*

1. $c_{min} \geq 2^{-m}$.
2. *If $\boldsymbol{f(0)} \succ \boldsymbol{0}$, then $\mu_{min} \geq c_{min}$.*
3. *If $c_{min} > 1$, then $\mu_{min} > 1$.*
4. *If $c_{min} \leq 1$, then $\mu_{min} \geq c_{min}^{2^n - 1}$.*
5. *If $\boldsymbol{f}$ is strictly quadratic, i.e., nonlinear, then the inequalities $c_{min} \leq 1$ and $\mu_{max} \cdot c_{min}^{3n-2} \cdot \min\{\mu_{min}^{2n-2}, 1\} \leq 1$ hold.*

Proof. We show the relations in turn.

1. The smallest nonzero coefficient representable as a ratio of m-bit numbers is $\frac{1}{2^m}$.
2. As $\boldsymbol{f(0)} \succ \boldsymbol{0}$, in all components i there is a nonzero coefficient c_i such that $f_i(\boldsymbol{0}) = c_i$. We have $\mu\boldsymbol{f} \geq \boldsymbol{f(0)}$, so $\mu\boldsymbol{f}_i \geq f_i(\boldsymbol{0}) = c_i \geq c_{min} > 0$ holds for all i. Hence $\mu_{min} > 0$.
3. Let $c_{min} > 1$. Recall the Kleene sequence $(\boldsymbol{\kappa}^{(k)})_{k \in \mathbb{N}}$ with $\boldsymbol{\kappa}^{(k)} = \boldsymbol{f}^k(\boldsymbol{0})$. We first show by induction on k that for all $k \in \mathbb{N}$ and all components i either $\kappa_i^{(k)} = 0$ holds or $\kappa_i^{(k)} > 1$. For the induction base we have $\boldsymbol{\kappa}^{(0)} = \boldsymbol{0}$. Let $k \geq 0$. Then $\kappa_i^{(k+1)} = f_i(\boldsymbol{\kappa}^{(k)})$ is a sum of products of numbers which are either coefficients of $\boldsymbol{f}$ (and hence by assumption greater than 1) or which are equal to $\kappa_j^{(k)}$ for some j. By induction, $\kappa_j^{(k)}$ is either 0 or greater than 1. So, $\kappa_i^{(k+1)}$ must be 0 or greater than 1.
   By Theorem 2.4, the Kleene sequence converges to $\mu\boldsymbol{f}$. As $\boldsymbol{f}$ is clean, we have $\mu\boldsymbol{f} \succ \boldsymbol{0}$, and so there is a $k \in \mathbb{N}$ such that $\boldsymbol{\kappa}^{(k)} \succ \boldsymbol{1}$. The statement follows with $\mu\boldsymbol{f} \geq \boldsymbol{\kappa}^{(k)}$.
4. Let $c_{min} \leq 1$. We prove the following stronger statement by induction on k: For every k with $0 \leq k \leq n$ there is a set $S_k \subseteq \{1, \dots, n\}$, $|S_k| = k$, such that $\mu\boldsymbol{f}_s \geq c_{min}^{2^k - 1}$ holds for all $s \in S_k$. The induction base $(k = 0)$ is trivial. Let $k \geq 0$. Consider the SPP $\widehat{\boldsymbol{f}}(\boldsymbol{X}_{\{1,\dots,n\} \setminus S_k})$ that is obtained from $\boldsymbol{f}(\boldsymbol{X})$ by removing the S_k-components from $\boldsymbol{f}$ and replacing every S_k-variable in the polynomials by the corresponding component of $\mu\boldsymbol{f}$. Clearly, $\mu\widehat{\boldsymbol{f}} = (\mu\boldsymbol{f})_{\{1,\dots,n\} \setminus S_k}$. By induction, the smallest nonzero coefficient $\widehat{c}_{min}$ of $\widehat{\boldsymbol{f}}$ satisfies $\widehat{c}_{min} \geq c_{min}(c_{min}^{2^k - 1})^2 = c_{min}^{2^{k+1} - 1}$. Pick a component i with $\widehat{f}_i(\boldsymbol{0}) > 0$. Then $\mu\widehat{\boldsymbol{f}}_i \geq \widehat{f}_i(\boldsymbol{0}) \geq \widehat{c}_{min} \geq c_{min}^{2^{k+1} - 1}$. So set $S_{k+1} := S_k \cup \{i\}$.
5. W.l.o.g. let $\mu_{max} = \mu\boldsymbol{f}_1$. The proof is based on the idea that X_1 indirectly depends quadratically on itself. More precisely, as $\boldsymbol{f}$ is strongly connected and strictly quadratic, component 1 depends (indirectly) on some component, say i_r, such that f_{i_r} contains a degree-2-monomial. The variables in that monomial, in turn, depend on X_1. This gives an inequality of the form $\mu\boldsymbol{f}_1 \geq C \cdot \mu\boldsymbol{f}_1^2$, implying $\mu\boldsymbol{f}_1 \cdot C \leq 1$.
   We give the details in the following. As $\boldsymbol{f}$ is strongly connected and strictly quadratic, there exists a sequence of variables $X_{i_1}, \dots, X_{i_r}$ and a sequence of

monomials $m_{i_1}, \ldots, m_{i_r}$ $(1 \le r \le n)$ with the following properties:

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

Notice that

(5.3)
$$\mu_{max} = \mu\boldsymbol{f}_1 \ge c_{i_1} \cdot \ldots \cdot c_{i_r} \cdot \mu\boldsymbol{f}_{j_1} \cdot \mu\boldsymbol{f}_{k_1}$$
$$\ge \min(c_{min}^n, 1) \cdot \mu\boldsymbol{f}_{j_1} \cdot \mu\boldsymbol{f}_{k_1} .$$

Again using that $\boldsymbol{f}$ is strongly connected, there exists a sequence of variables $X_{j_1}, \ldots, X_{j_s}$ and a sequence of monomials $m_{j_1}, \ldots, m_{j_{s-1}}$ $(1 \le s \le n)$ with the following properties:

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

Notice that

(5.4)
$$\mu\boldsymbol{f}_{j_1} \ge c_{j_1} \cdot \ldots \cdot c_{j_{s-1}} \cdot \min(\mu_{min}^{s-1}, 1) \cdot \mu\boldsymbol{f}_1$$
$$\ge \min(c_{min}^{n-1}, 1) \cdot \min(\mu_{min}^{n-1}, 1) \cdot \mu\boldsymbol{f}_1 .$$

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

(5.5)
$$\mu\boldsymbol{f}_{k_1} \ge \min(c_{min}^{n-1}, 1) \cdot \min(\mu_{min}^{n-1}, 1) \cdot \mu\boldsymbol{f}_1 .$$

Combining (5.3) with (5.4) and (5.5) yields

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

or

(5.6)
$$\mu_{max} \cdot \min(c_{min}^{3n-2}, 1) \cdot \min(\mu_{min}^{2n-2}, 1) \le 1 .$$

Now it suffices to show $c_{min} \le 1$. Assume for a contradiction that $c_{min} > 1$. Then, by statement 3, $\mu_{min} > 1$. Plugging this into (5.6) yields $\mu_{max} \le 1$. This implies $\mu_{max} < \mu_{min}$, contradicting the definitions of μ_{max} and μ_{min}. $\quad\square$

Now we are ready to prove Theorem 5.10.

Proof of Theorem 5.10.

1. First we check the case where $\boldsymbol{f}$ is linear; i.e., all polynomials f_i have degree at most 1. In this case, Newton's method reaches $\mu\boldsymbol{f}$ after one iteration, so the statement 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.
   By Theorem 5.7 it suffices to show

$$\log \frac{\mu_{max}}{\mu_{min} \cdot c_{min}^n \cdot \min\{\mu_{min}^n, 1\}} \le 4mn + 3n \max\{0, -\log \mu_{min}\} .$$

We have

$$\log \frac{\mu_{max}}{\mu_{min} \cdot c_{min}^n \cdot \min\{\mu_{min}^n, 1\}}$$

$$\leq \log \frac{1}{c_{min}^{4n-2} \cdot \min\{\mu_{min}^{3n-1}, 1\}} \qquad \text{(Lemma 5.11.5)}$$

$$\leq 4n \cdot \log \frac{1}{c_{min}} - \log(\min\{\mu_{min}^{3n-1}, 1\}) \qquad \text{(Lemma 5.11.5: } c_{min} \leq 1)$$

$$\leq 4mn - \log(\min\{\mu_{min}^{3n-1}, 1\}) \qquad \text{(Lemma 5.11.1)}.$$

If $\mu_{min} \geq 1$, we have $-\log(\min\{\mu_{min}^{3n-1}, 1\}) \leq 0$, so we are done in this case. If $\mu_{min} \leq 1$, we have $-\log(\min\{\mu_{min}^{3n-1}, 1\}) = -(3n-1)\log\mu_{min} \leq 3n \cdot (-\log\mu_{min})$.

2. By statement 1 of this theorem, it suffices to show that $4mn + 3n\max\{0, -\log\mu_{min}\} \leq 4mn2^n$. This inequality obviously holds if $\mu_{min} \geq 1$. So let $\mu_{min} \leq 1$. Then, by Lemma 5.11.3, $c_{min} \leq 1$. Hence, by parts 4 and 1 of Lemma 5.11, $\mu_{min} \geq c_{min}^{2^n-1} \geq 2^{-m(2^n-1)}$. So we have an upper bound on $-\log\mu_{min}$ with $-\log\mu_{min} \leq m(2^n-1)$ and get

$$4mn + 3n\max\{0, -\log\mu_{min}\} \leq 4mn + 3nm(2^n-1)$$
$$\leq 4mn + 4nm(2^n-1) = 4mn2^n.$$

3. Let $\boldsymbol{f}(\boldsymbol{0}) \succ \boldsymbol{0}$. By statement 1 of this theorem it suffices to show that $4mn + 3n\max\{0, -\log\mu_{min}\} \leq 7mn$ holds. By parts 2 and 1 of Lemma 5.11, we have $\mu_{min} \geq c_{min} \geq 2^{-m}$, so $-\log\mu_{min} \leq m$. Hence, $4mn + 3n\max\{0, -\log\mu_{min}\} \leq 4mn + 3nm = 7mn$.

4. Let $\boldsymbol{f}(\boldsymbol{0}) \succ \boldsymbol{0}$ and $\mu_{max} \leq 1$. By Theorem 5.7 it suffices to show that

$$\log \frac{\mu_{max}}{\mu_{min} \cdot c_{min}^n \cdot \min\{\mu_{min}^n, 1\}} \leq 2mn + m.$$

We have

$$\log \frac{\mu_{max}}{\mu_{min} \cdot c_{min}^n \cdot \min\{\mu_{min}^n, 1\}}$$

$$\leq -n\log c_{min} - (n+1)\log\mu_{min} \qquad \text{(as } \mu_{min} \leq \mu_{max} \leq 1)$$

$$\leq -(2n+1)\log c_{min} \qquad \text{(Lemma 5.11.2)}$$

$$\leq 2mn + m \qquad \text{(Lemma 5.11.1)}. \qquad \square$$

5.4. Upper bounds on the least fixed point via Newton approximants. By Theorem 4.1 each Newton approximant $\boldsymbol{\nu}^{(k)}$ is a lower bound on $\mu\boldsymbol{f}$. Theorems 5.7 and 5.10 give us upper bounds on the error $\boldsymbol{\Delta}^{(k)} := \mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)}$. Those bounds can be directly transformed into upper bounds on $\mu\boldsymbol{f}$, as $\mu\boldsymbol{f} = \boldsymbol{\nu}^{(k)} + \boldsymbol{\Delta}^{(k)}$; cf. Example 5.8.

Theorems 5.7 and 5.10 allow us to compute bounds on $\boldsymbol{\Delta}^{(k)}$ even before the Newton iteration has been started. However, this may be more than we actually need. In practice, we may wish to use an iterative method that yields guaranteed lower *and* upper bounds on $\mu\boldsymbol{f}$ that improve during the iteration. The following theorem and its corollary can be used to this end.

THEOREM 5.12. *Let $\boldsymbol{f}$ be a quadratic, clean, and feasible scSPP. Let $\boldsymbol{0} \leq \boldsymbol{x} \leq \mu\boldsymbol{f}$ and $\boldsymbol{x} \leq \boldsymbol{f}(\boldsymbol{x})$ such that $\boldsymbol{f}'(\boldsymbol{x})^*$ exists. Let c_{min} be the smallest nonzero coefficient*

of $\boldsymbol{f}$, and let μ_{min} be the minimal component of $\mu\boldsymbol{f}$. Then

$$\frac{\|\mathcal{N}(\boldsymbol{x}) - \boldsymbol{x}\|_\infty}{\|\mu\boldsymbol{f} - \mathcal{N}(\boldsymbol{x})\|_\infty} \geq (c_{min} \cdot \min\{\mu_{min}, 1\})^n \ .$$

We prove Theorem 5.12 at the end of the section. The theorem can be applied to the Newton approximants.

THEOREM 5.13. *Let $\boldsymbol{f}$ be a quadratic, clean, and feasible scSPP. Let c_{min} be the smallest nonzero coefficient of $\boldsymbol{f}$, and let μ_{min} be the minimal component of $\mu\boldsymbol{f}$. For all Newton approximants $\boldsymbol{\nu}^{(k)}$ with $\boldsymbol{\nu}^{(k)} \succ \boldsymbol{0}$, let $\nu_{min}^{(k)}$ be the smallest coefficient of $\boldsymbol{\nu}^{(k)}$. Then*

$$\boldsymbol{\nu}^{(k)} \leq \mu\boldsymbol{f} \leq \boldsymbol{\nu}^{(k)} + \left[\frac{\left\|\boldsymbol{\nu}^{(k)} - \boldsymbol{\nu}^{(k-1)}\right\|_\infty}{\left(c_{min} \cdot \min\{\nu_{min}^{(k)}, 1\}\right)^n} \right],$$

where $[s]$ denotes the vector $\boldsymbol{x}$ with $x_j = s$ for all $1 \leq j \leq n$.

Proof. Theorem 5.12 applies, due to Theorem 4.1, to the Newton approximants with $\boldsymbol{x} = \boldsymbol{\nu}^{(k-1)}$. So we get

$$\left\|\mu\boldsymbol{f} - \boldsymbol{\nu}^{(k)}\right\|_\infty \leq \frac{\left\|\boldsymbol{\nu}^{(k)} - \boldsymbol{\nu}^{(k-1)}\right\|_\infty}{(c_{min} \cdot \min\{\mu_{min}, 1\})^n}$$

$$\leq \frac{\left\|\boldsymbol{\nu}^{(k)} - \boldsymbol{\nu}^{(k-1)}\right\|_\infty}{\left(c_{min} \cdot \min\{\nu_{min}^{(k)}, 1\}\right)^n} \qquad (\text{as } \boldsymbol{\nu}^{(k)} \leq \mu\boldsymbol{f}) \ .$$

Hence the statement follows from $\boldsymbol{\nu}^{(k)} \leq \mu\boldsymbol{f}$. □

Example 5.14. Consider again the equation $\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{X})$ from Examples 2.9 and 5.8:

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

Again we wish to verify that there is no component $s \in \{1, 2, 3\}$ with $\mu\boldsymbol{f}_s = 1$. Performing 10 Newton steps yields an approximation $\boldsymbol{\nu}^{(10)}$ to $\mu\boldsymbol{f}$ with

$$\begin{pmatrix} 0.9828 \\ 0.9738 \\ 0.9926 \end{pmatrix} \prec \boldsymbol{\nu}^{(10)} \prec \begin{pmatrix} 0.9829 \\ 0.9739 \\ 0.9927 \end{pmatrix} \ .$$

Further, it holds that $\left\|\boldsymbol{\nu}^{(10)} - \boldsymbol{\nu}^{(9)}\right\|_\infty \leq 2 \cdot 10^{-6}$. So we have

$$\frac{\left\|\boldsymbol{\nu}^{(10)} - \boldsymbol{\nu}^{(9)}\right\|_\infty}{\left(c_{min} \cdot \min\{\nu_{min}^{(10)}, 1\}\right)^3} \leq \frac{2 \cdot 10^{-6}}{(0.3 \cdot 0.97)^3} \leq 0.00009,$$

and hence by Theorem 5.13

$$\boldsymbol{\nu}^{(10)} \leq \mu\boldsymbol{f} \leq \boldsymbol{\nu}^{(10)} + [0.00009] \leq \begin{pmatrix} 0.983 \\ 0.974 \\ 0.993 \end{pmatrix}.$$

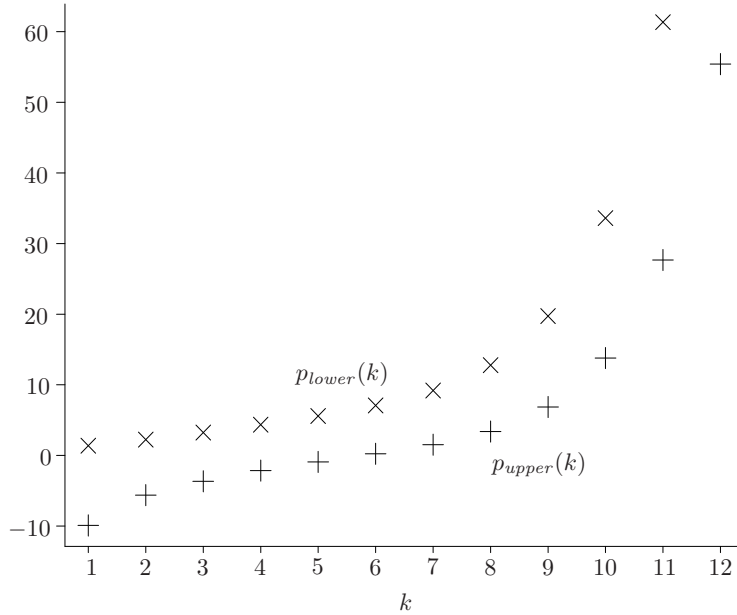In particular we know that $\mu\boldsymbol{f}_s < 1$ for all three components s.

FIG. 5.2. *Number of valid bits of the lower ($\times$) and upper ($+$) bounds on $\mu\boldsymbol{f}_1$; see Example* 5.15.

Example 5.15. Consider again the SPP $\boldsymbol{f}$ from Example 5.14. Setting

$$\boldsymbol{u}^{(k)} := \boldsymbol{\nu}^{(k)} + \left\lceil \frac{\left\|\boldsymbol{\nu}^{(k)} - \boldsymbol{\nu}^{(k-1)}\right\|_\infty}{\left(0.3 \cdot \nu_{min}^{(k)}\right)^3} \right\rceil ,$$

Theorem 5.13 guarantees

$$\boldsymbol{\nu}^{(k)} \leq \mu\boldsymbol{f} \leq \boldsymbol{u}^{(k)} .$$

Let us measure the tightness of the bounds $\boldsymbol{\nu}^{(k)}$ and $\boldsymbol{u}^{(k)}$ on $\mu\boldsymbol{f}$ in the first component. Let

$$p_{lower}(k) := -\log_2(\mu\boldsymbol{f}_1 - \nu_1^{(k)}),$$
$$p_{upper}(k) := -\log_2(u_1^{(k)} - \mu\boldsymbol{f}_1) .$$

Roughly speaking, $\nu_1^{(k)}$ and $u_1^{(k)}$ have $p_{lower}(k)$ and $p_{upper}(k)$ valid bits of $\mu\boldsymbol{f}_1$, respectively. Figure 5.2 shows $p_{lower}(k)$ and $p_{upper}(k)$ for $k \in \{1, \ldots, 11\}$.

It can be seen that the slope of $p_{lower}(k)$ is approximately 1 for $k = 2, \ldots, 6$. This corresponds to the linear convergence of Newton's method according to Theorem 5.7. Since $\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})$ is nonsingular,[3] Newton's method actually has, asymptotically, an exponential convergence order; cf. Theorem 4.11. This behavior can be observed in Figure 5.2 for $k \geq 7$. For p_{upper}, we roughly have (using $\boldsymbol{\nu}^{(k)} \approx \mu\boldsymbol{f}$)

$$p_{upper}(k) \approx p_{lower}(k-1) + \log\left(0.3 \cdot \nu_{min}^{(k)}\right)^3 \approx p_{lower}(k-1) - 5 .$$

[3]In fact, the matrix is "almost" singular, with $\det(\mathrm{Id} - \boldsymbol{f}'(\mu\boldsymbol{f})) \approx 0.006$.

The proof of Theorem 5.12 uses techniques similar to those of the proof of Theorem 5.7, in particular Lemma 5.9.

Proof of Theorem 5.12. By Proposition 5.2, $\boldsymbol{f}$ has a cone vector $\boldsymbol{d}$. Let d_{min} and d_{max} be the smallest and the largest component of $\boldsymbol{d}$, respectively. Let $\lambda_{max} := \max_j\{\frac{\mu\boldsymbol{f}_j - x_j}{d_j}\}$, and w.l.o.g. let $\lambda_{max} = \frac{\mu\boldsymbol{f}_1 - x_1}{d_1}$. We have $\boldsymbol{x} \geq \mu\boldsymbol{f} - \lambda_{max}\boldsymbol{d}$, so we can apply Lemma 5.4 to obtain $\mathcal{N}(\boldsymbol{x}) \geq \mu\boldsymbol{f} - \frac{1}{2}\lambda_{max}\boldsymbol{d}$. Thus

$$\|\mathcal{N}(\boldsymbol{x}) - \boldsymbol{x}\|_\infty \geq (\mathcal{N}(\boldsymbol{x}) - \boldsymbol{x})_1 \geq \mu\boldsymbol{f}_1 - \frac{1}{2}\lambda_{max}d_1 - x_1 = \frac{1}{2}\lambda_{max}d_1 \geq \frac{1}{2}\lambda_{max}d_{min} \ .$$

On the other hand, with Lemma 4.5 we have $\boldsymbol{0} \leq \mu\boldsymbol{f} - \mathcal{N}(\boldsymbol{x}) \leq \frac{1}{2}\lambda_{max}\boldsymbol{d}$ and so $\|\mu\boldsymbol{f} - \mathcal{N}(\boldsymbol{x})\|_\infty \leq \frac{1}{2}\lambda_{max}d_{max}$. Combining those inequalities, we obtain

$$\frac{\|\mathcal{N}(\boldsymbol{x}) - \boldsymbol{x}\|_\infty}{\|\mu\boldsymbol{f} - \mathcal{N}(\boldsymbol{x})\|_\infty} \geq \frac{d_{min}}{d_{max}} \ .$$

Now the statement follows from Lemma 5.9. ☐

6. General SPPs. In section 5 we considered *strongly connected* SPPs; see Definition 2.6. However, it is not always guaranteed that the SPP $\boldsymbol{f}$ is strongly connected. In this section we analyze the convergence speed of two variants of Newton's method that both compute approximations of $\mu\boldsymbol{f}$, where $\boldsymbol{f}$ is a clean and feasible SPP that is not necessarily strongly connected ("general SPPs").

The first one was suggested by Etessami and Yannakakis [16] and is called *decomposed Newton's method (DNM)*. It works by running Newton's method separately on each SCC; see section 6.1. The second one is the regular Newton's method from section 4. We will analyze its convergence speed in section 6.2.

The reason we first analyze DNM is that our convergence speed results about Newton's method for general SPPs (Theorem 6.5) build on our results about DNM (Theorem 6.2). From an efficiency point of view it actually may be advantageous to run Newton's method separately on each SCC. For those reasons DNM deserves a separate treatment.

6.1. Convergence speed of DNM. DNM, originally suggested in [16], works as follows. It starts by using Newton's method for each bottom SCC, say S, of the SPP $\boldsymbol{f}$. Then the corresponding variables $\boldsymbol{X}_S$ are substituted for the obtained approximation for $\mu\boldsymbol{f}_S$, and the corresponding equations $\boldsymbol{X}_S = \boldsymbol{f}_S(\boldsymbol{X})$ are removed. The same procedure is then applied to the new bottom SCCs, until all SCCs have been processed.

Etessami and Yannakakis did not provide a particular criterion for the number of Newton iterations to be applied in each SCC. Consequently, they did not analyze the convergence speed of DNM. We will treat those issues in this section, thereby taking advantage of our previous analysis of scSPPs.

We fix a quadratic, clean, and feasible SPP $\boldsymbol{f}$ for this section. We assume that we have already computed the DAG of SCCs. This can be done in linear time in the size of $\boldsymbol{f}$. To each SCC S we can associate its *depth* t: it is the longest path in the DAG of SCCs from S to a top SCC. Notice that $0 \leq t \leq n-1$. We write $\mathcal{SCC}(t)$ for the set of SCCs of depth t. We define the height $h(\boldsymbol{f})$ as the largest depth of an SCC and the width $w(\boldsymbol{f}) := \max_t |\mathcal{SCC}(t)|$ as the largest number of SCCs of the same depth. Notice that $\boldsymbol{f}$ has at most $(h(\boldsymbol{f}) + 1) \cdot w(\boldsymbol{f})$ SCCs. Further we define the component sets $[t] := \bigcup_{S \in \mathcal{SCC}(t)} S$ and $[>t] := \bigcup_{t' > t}[t']$ and similarly $[< t]$.

```
function DNM (f, i)          /* The parameter i controls the precision. */
    for t from h(f) downto 0
        forall S ∈ SCC(t)                    /* for all SCCs S of depth t */
            ρ_S^(i) := N_{f_S}^{i·2^t}(0)          /* perform i · 2^t Newton iterations */
            f_{[<t]} := f_{[<t]}[S/ρ_S^(i)]        /* apply ρ_S^(i) in the upper SCCs */
    return ρ^(i)
```

FIG. 6.1. *Decomposed Newton's method (DNM) for computing an approximation $\boldsymbol{\rho}^{(i)}$ of $\mu\boldsymbol{f}$.*

Figure 6.1 shows our version of DNM. We suggest running 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 i that controls the precision.

PROPOSITION 6.1. *The function* $\mathrm{DNM}(\boldsymbol{f}, i)$ *of Figure* 6.1 *runs at most* $i \cdot w(\boldsymbol{f}) \cdot 2^{h(\boldsymbol{f})+1} \leq i \cdot n \cdot 2^n$ *iterations of Newton's method.*

Proof. The number of iterations is $\sum_{t=0}^{h(\boldsymbol{f})} |\mathcal{SCC}(t)| \cdot i \cdot 2^t$. This can be estimated as follows:

$$\sum_{t=0}^{h(\boldsymbol{f})} |\mathcal{SCC}(t)| \cdot i \cdot 2^t \leq w(\boldsymbol{f}) \cdot i \cdot \sum_{t=0}^{h(\boldsymbol{f})} 2^t$$
$$\leq w(\boldsymbol{f}) \cdot i \cdot 2^{h(\boldsymbol{f})+1}$$
$$\leq i \cdot n \cdot 2^n \qquad \text{(as } w(\boldsymbol{f}) \leq n \text{ and } h(\boldsymbol{f}) < n\text{).} \qquad \square$$

The following theorem states that DNM has linear convergence order.

THEOREM 6.2. *Let* $\boldsymbol{f}$ *be a quadratic, clean, and feasible SPP. Let* $\boldsymbol{\rho}^{(i)}$ *denote the result of calling* $\mathrm{DNM}(\boldsymbol{f}, i)$ *(see Figure* 6.1*). Let* $\beta_{\boldsymbol{\rho}}$ *denote the convergence order of* $(\boldsymbol{\rho}^{(i)})_{i \in \mathbb{N}}$. *Then there is a* $k_{\boldsymbol{f}} \in \mathbb{N}$ *such that* $\beta_{\boldsymbol{\rho}}(k_{\boldsymbol{f}} + i) \geq i$ *for all* $i \in \mathbb{N}$.

Theorem 6.2 can be interpreted as follows: Increasing i by one yields asymptotically at least one additional bit in each component and, by Proposition 6.1, costs at most $n \cdot 2^n$ additional Newton iterations. Notice that for simplicity we do not take into account here that the cost of performing a Newton step on a single SCC is not uniform, but rather depends on the size of the SCC (e.g., cubically if Gaussian elimination is used for solving the linear systems).

For the proof of Theorem 6.2, let $\boldsymbol{\Delta}^{(i)}$ denote the error when running DNM with parameter i, i.e., $\boldsymbol{\Delta}^{(i)} := \mu\boldsymbol{f} - \boldsymbol{\rho}^{(i)}$. Observe that the error $\boldsymbol{\Delta}^{(i)}$ can be understood as the sum of two errors:

$$\boldsymbol{\Delta}^{(i)} := \mu\boldsymbol{f} - \boldsymbol{\rho}^{(i)} = (\boldsymbol{\mu} - \widetilde{\boldsymbol{\mu}}^{(i)}) + (\widetilde{\boldsymbol{\mu}}^{(i)} - \boldsymbol{\rho}^{(i)}),$$

where $\widetilde{\boldsymbol{\mu}}_{[t]}^{(i)} := \mu\big(\boldsymbol{f}_{[t]}[[>t]/\boldsymbol{\rho}_{[>t]}^{(i)}]\big)$; i.e., $\widetilde{\boldsymbol{\mu}}_{[t]}^{(i)}$ is the least fixed point of $\boldsymbol{f}_{[t]}$ after the approximations from the lower SCCs have been applied. So, $\boldsymbol{\Delta}_{[t]}^{(i)}$ consists of the *propagation error* $(\mu\boldsymbol{f}_{[t]} - \widetilde{\boldsymbol{\mu}}_{[t]}^{(i)})$ (resulting from the error at lower SCCs) and the *approximation error* $(\widetilde{\boldsymbol{\mu}}_{[t]}^{(i)} - \boldsymbol{\rho}_{[t]}^{(i)})$ (resulting from the newly added error of Newton's method on level t).

The following lemma gives a bound on the propagation error.

LEMMA 6.3 (propagation error). *There is a constant* $C_{\boldsymbol{f}} > 0$ *such that*

$$\left\| \mu\boldsymbol{f}_{[t]} - \widetilde{\boldsymbol{\mu}}_{[t]} \right\| \leq C_{\boldsymbol{f}} \cdot \sqrt{\left\| \mu\boldsymbol{f}_{[>t]} - \boldsymbol{\rho}_{[>t]} \right\|}$$

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

Roughly speaking, Lemma 6.3 states that if $\boldsymbol{\rho}_{[>t]}^{(i)}$ has k valid bits of $\mu \boldsymbol{f}_{[>t]}$, then $\widetilde{\boldsymbol{\mu}}_{[t]}^{(i)}$ has at least about $k/2$ valid bits of $\mu \boldsymbol{f}_{[t]}$. 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 proof of Lemma 6.3 is technically involved and, unfortunately, not constructive in that we know nothing about $C_{\boldsymbol{f}}$ except for its existence. Therefore, the statements in this section are independent of a particular norm. The proof of Lemma 6.3 can be found in Appendix A.

The following lemma gives a bound on the error $\|\boldsymbol{\Delta}_{[t]}^{(i)}\|$ on level t, taking both the propagation error and the approximation error into account.

LEMMA 6.4. *There is a* $C_{\boldsymbol{f}} > 0$ *such that* $\|\boldsymbol{\Delta}_{[t]}^{(i)}\| \le 2^{C_{\boldsymbol{f}} - i \cdot 2^t}$ *for all* $i \in \mathbb{N}$.

Proof. Let $\widetilde{\boldsymbol{f}}_{[t]}^{(i)} := \boldsymbol{f}_{[t]}[[> t]/\boldsymbol{\rho}_{[>t]}^{(i)}]$. Observe that the coefficients of $\widetilde{\boldsymbol{f}}_{[t]}^{(i)}$ and thus its least fixed point $\widetilde{\boldsymbol{\mu}}_{[t]}^{(i)}$ are monotonically increasing with i, because $\boldsymbol{\rho}_{[>t]}^{(i)}$ is monotonically increasing as well. Consider an arbitrary depth t and choose real numbers $c_{min} > 0$ and $\mu_{min} > 0$ and an integer i_0 such that, for all $i \ge i_0$, c_{min} and μ_{min} are lower bounds on the smallest nonzero coefficient of $\widetilde{\boldsymbol{f}}_{[t]}^{(i)}$ and the smallest coefficient of $\widetilde{\boldsymbol{\mu}}_{[t]}^{(i)}$, respectively. Let μ_{max} be the largest component of $\mu \boldsymbol{f}_{[t]}$. Let $\widetilde{k} := \lceil n \cdot \log \frac{\mu_{max}}{c_{min} \cdot \mu_{min} \cdot \min\{\mu_{min}, 1\}} \rceil$. Then it follows from Theorem 5.7 that performing $\widetilde{k} + j$ Newton iterations ($j \ge 0$) on depth t yields j valid bits of $\widetilde{\boldsymbol{\mu}}_{[t]}^{(i)}$ for any $i \ge i_0$. In particular, $\widetilde{k} + i \cdot 2^t$ Newton iterations give $i \cdot 2^t$ valid bits of $\widetilde{\boldsymbol{\mu}}_{[t]}^{(i)}$ for any $i \ge i_0$. So there exists a constant $c_1 > 0$ such that, for all $i \ge i_0$,

$$(6.1) \qquad \left\| \widetilde{\boldsymbol{\mu}}_{[t]}^{(i)} - \boldsymbol{\rho}_{[t]}^{(i)} \right\| \le 2^{c_1 - i \cdot 2^t} ,$$

because DNM (see Figure 6.1) performs $i \cdot 2^t$ iterations to compute $\boldsymbol{\rho}_S^{(i)}$, where S is an SCC of depth t. Choose c_1 large enough such that (6.1) holds for all $i \ge 0$ and all depths t.

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 (6.1). Let $t < h(\boldsymbol{f})$. Then

$$\left\| \boldsymbol{\Delta}_{[t]}^{(i)} \right\| = \left\| \mu \boldsymbol{f}_{[t]} - \widetilde{\boldsymbol{\mu}}_{[t]}^{(i)} + \widetilde{\boldsymbol{\mu}}_{[t]}^{(i)} - \boldsymbol{\rho}_{[t]}^{(i)} \right\|$$

$$\le \left\| \mu \boldsymbol{f}_{[t]} - \widetilde{\boldsymbol{\mu}}_{[t]}^{(i)} \right\| + \left\| \widetilde{\boldsymbol{\mu}}_{[t]}^{(i)} - \boldsymbol{\rho}_{[t]}^{(i)} \right\|$$

$$\le \left\| \mu \boldsymbol{f}_{[t]} - \widetilde{\boldsymbol{\mu}}_{[t]}^{(i)} \right\| + 2^{c_1 - i \cdot 2^t} \qquad \text{(by (6.1))}$$

$$\le c_2 \cdot \sqrt{\left\| \boldsymbol{\Delta}_{[>t]}^{(i)} \right\|} + 2^{c_1 - i \cdot 2^t} \qquad \text{(Lemma 6.3)}$$

$$\le c_2 \cdot \sqrt{2^{c_3 - i \cdot 2^{t+1}}} + 2^{c_1 - i \cdot 2^t} \qquad \text{(induction hypothesis)}$$

$$\le 2^{c_4 - i \cdot 2^t}$$

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

Now Theorem 6.2 follows easily.

Proof of Theorem 6.2. From Lemma 6.4 we deduce that for each component $j \in [t]$ there is a c_j such that

$$(\mu \boldsymbol{f}_j - \rho_j^{(i)}) / \mu \boldsymbol{f}_j \le 2^{c_j - i \cdot 2^t} \le 2^{c_j - i} .$$

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

$$(\mu\boldsymbol{f}_j - \rho_j^{(i+k_{\boldsymbol{f}})})/\mu\boldsymbol{f}_j \leq 2^{c_j-(i+k_{\boldsymbol{f}})} \leq 2^{-i} . \qquad \square$$

Notice that, unfortunately, we cannot give a bound on $k_{\boldsymbol{f}}$, mainly because Lemma 6.3 does not provide a bound on $C_{\boldsymbol{f}}$.

6.2. Convergence speed of Newton's method. We use Theorem 6.2 to prove the following theorem for the regular (i.e., not decomposed) Newton sequence $(\boldsymbol{\nu}^{(i)})_{i\in\mathbb{N}}$.

THEOREM 6.5. *Let $\boldsymbol{f}$ be a quadratic, clean, and feasible SPP. There is a threshold $k_{\boldsymbol{f}} \in \mathbb{N}$ such that $\beta(k_{\boldsymbol{f}} + i \cdot n \cdot 2^n) \geq \beta(k_{\boldsymbol{f}} + i \cdot (h(\boldsymbol{f}) + 1) \cdot 2^{h(\boldsymbol{f})}) \geq i$ for all $i \in \mathbb{N}$.*

In the rest of the section we prove this theorem by a sequence of lemmas. The following lemma states that a Newton step is not faster on an SCC if the values of the lower SCCs are fixed.

LEMMA 6.6. *Let $\boldsymbol{f}$ be a clean and feasible SPP. Let $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{f}(\boldsymbol{x}) \leq \mu\boldsymbol{f}$ such that $\boldsymbol{f}'(\boldsymbol{x})^*$ exists. Let S be an SCC of $\boldsymbol{f}$ and let L denote the set of components that are not in S, but on which a variable in S depends. Then $(\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{x}))_S \geq \mathcal{N}_{\boldsymbol{f}_S[L/\boldsymbol{x}_L]}(\boldsymbol{x}_S)$.*

Proof.

$$\begin{aligned}
(\mathcal{N}_{\boldsymbol{f}}(\boldsymbol{x}))_S &= \big(\boldsymbol{f}'(\boldsymbol{x})^*(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x})\big)_S \\
&= \boldsymbol{f}'(\boldsymbol{x})_{SS}^*(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x})_S + \boldsymbol{f}'(\boldsymbol{x})_{SL}^*(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x})_L \\
&\geq \boldsymbol{f}'(\boldsymbol{x})_{SS}^*(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x})_S \\
&= \big((\boldsymbol{f}_S[L/\boldsymbol{x}_L])'(\boldsymbol{x}_S)\big)^*(\boldsymbol{f}_S[L/\boldsymbol{x}_L](\boldsymbol{x}_S) - \boldsymbol{x}_S) \\
&= \mathcal{N}_{\boldsymbol{f}_S[L/\boldsymbol{x}_L]}(\boldsymbol{x}_S). \qquad \square
\end{aligned}$$

Recall Lemma 4.9, which states that the Newton operator $\mathcal{N}$ is monotone. This fact and Lemma 6.6 can be combined in order to obtain the following lemma stating that $i\cdot(h(\boldsymbol{f})+1)$ iterations of the regular Newton's method "dominate" a decomposed Newton's method that performs i Newton steps in each SCC.

LEMMA 6.7. *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 regular Newton's method. Then $\boldsymbol{\nu}^{(i\cdot(h(\boldsymbol{f})+1))} \geq \widetilde{\boldsymbol{\nu}}^{(i)}$.*

Proof. Let $h = h(\boldsymbol{f})$. Let $[t]$ and $[>t]$ again denote the set of components of depth t and $> t$, respectively. We show the following by induction on the depth t:

$$\boldsymbol{\nu}_{[t]}^{(i\cdot(h+1-t))} \geq \widetilde{\boldsymbol{\nu}}_{[t]}^{(i)}.$$

The induction base $(t = h)$ is clear, because for bottom SCCs the two methods are identical. Now let $t < h$. Then

$$\begin{aligned}
\boldsymbol{\nu}_{[t]}^{(i\cdot(h+1-t))} &= \mathcal{N}_{\boldsymbol{f}}^i(\boldsymbol{\nu}^{(i\cdot(h-t))})_{[t]} \\
&\geq \mathcal{N}_{\boldsymbol{f}_{[t]}[[>t]/\boldsymbol{\nu}_{[>t]}^{(i\cdot(h-t))}]}^i(\boldsymbol{\nu}_{[t]}^{(i\cdot(h-t))}) && \text{(Lemma 6.6)} \\
&\geq \mathcal{N}_{\boldsymbol{f}_{[t]}[[>t]/\widetilde{\boldsymbol{\nu}}_{[>t]}^{(i)}]}^i(\boldsymbol{\nu}_{[t]}^{(i\cdot(h-t))}) && \text{(induction hypothesis)} \\
&\geq \mathcal{N}_{\boldsymbol{f}_{[t]}[[>t]/\widetilde{\boldsymbol{\nu}}_{[>t]}^{(i)}]}^i(\boldsymbol{0}_{[t]}) && \text{(Lemma 4.9)} \\
&= \widetilde{\boldsymbol{\nu}}_{[t]}^{(i)} && \text{(definition of } \widetilde{\boldsymbol{\nu}}^{(i)}\text{)}.
\end{aligned}$$

Now, the lemma itself follows by using Lemma 4.9 once more. $\qquad \square$

As a side note, observe that the above proof of Lemma 6.7 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 regular Newton's method. Therefore, $w(\boldsymbol{f})$, the width of $\boldsymbol{f}$, is irrelevant here (cf. Proposition 6.1).

Now we can prove Theorem 6.5.

Proof of Theorem 6.5. Let k_2 be the $k_{\boldsymbol{f}}$ of Theorem 6.2, and let $k_1 = k_2 \cdot (h(\boldsymbol{f})+1) \cdot 2^{h(\boldsymbol{f})}$. Then we have

$$\boldsymbol{\nu}^{(k_1 + i \cdot (h(\boldsymbol{f})+1) \cdot 2^{h(f)})} = \boldsymbol{\nu}^{((k_2 + i) \cdot (h(\boldsymbol{f})+1) \cdot 2^{h(\boldsymbol{f})})}$$

$$\geq \widetilde{\boldsymbol{\nu}}^{((k_2 + i) \cdot 2^{h(f)})} \qquad\qquad \text{(Lemma 6.7)}$$

$$\geq \boldsymbol{\rho}^{(k_2 + i)} ,$$

where the last step follows from the fact that $\mathrm{DNM}(\boldsymbol{f}, k_2 + i)$ runs at most $(k_2 + i) \cdot 2^{h(\boldsymbol{f})}$ iterations in every SCC. By Theorem 6.2, $\boldsymbol{\rho}^{(k_2 + i)}$ and hence $\boldsymbol{\nu}^{(k_1 + i \cdot (h(\boldsymbol{f})+1) \cdot 2^{h(f)})}$ have i valid bits of $\mu \boldsymbol{f}$. Therefore, Theorem 6.5 holds with $k_{\boldsymbol{f}} = k_1$. □

7. Upper bounds on the convergence. In this section we show that the lower bounds on the convergence order of Newton's method that we obtained in the previous section are essentially tight, meaning that an exponential (in n) number of iterations may be needed per bit.

More precisely, we expose a family $(\boldsymbol{f}^{(n)})_{n \geq 1}$ of SPPs with n variables, such that more than $k \cdot 2^{n-1}$ iterations are needed for k valid bits. Consider the following system:

$$(7.1) \qquad \boldsymbol{X} = \boldsymbol{f}^{(n)}(\boldsymbol{X}) = \begin{pmatrix} \frac{1}{2} + \frac{1}{2}X_1^2 \\ \frac{1}{4}X_1^2 + \frac{1}{2}X_1 X_2 + \frac{1}{4}X_2^2 \\ \vdots \\ \frac{1}{4}X_{n-1}^2 + \frac{1}{2}X_{n-1}X_n + \frac{1}{4}X_n^2 \end{pmatrix}.$$

The only solution of (7.1) is $\mu \boldsymbol{f}^{(n)} = (1, \ldots, 1)^\top$. Notice that each component of $\boldsymbol{f}^{(n)}$ is an SCC. We prove the following theorem.

THEOREM 7.1. *The convergence order of Newton's method applied to the SPP $\boldsymbol{f}^{(n)}$ from* (7.1) *(with $n \geq 2$) satisfies*

$$\beta(k \cdot 2^{n-1}) < k \text{ for all } k \in \{1, 2, \ldots\}.$$

In particular, $\beta(2^{n-1}) = 0$.

Proof. We write $\boldsymbol{f} := \boldsymbol{f}^{(n)}$ for simplicity. Let

$$\boldsymbol{\Delta}^{(i)} := \mu \boldsymbol{f} - \boldsymbol{\nu}^{(i)} = (1, \ldots, 1)^\top - \boldsymbol{\nu}^{(i)} .$$

Notice that $(\nu_1^{(i)})_{i \in \mathbb{N}} = (0, \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \ldots)$, which is the same sequence as obtained by applying Newton's method to the 1-dimensional system $X_1 = \frac{1}{2} + \frac{1}{2}X_1^2$. So we have $\Delta_1^{(i)} = 2^{-i}$; i.e., after i iterations we have exactly i valid bits in the first component.

We know from Theorem 4.1 that for all j with $1 \leq j \leq n-1$ we have $\nu_{j+1}^{(i)} \leq f_{j+1}(\boldsymbol{\nu}^{(i)}) = \frac{1}{4}(\nu_j^{(i)})^2 + \frac{1}{2}\nu_j^{(i)}\nu_{j+1}^{(i)} + \frac{1}{4}(\nu_{j+1}^{(i)})^2$ and $\nu_{j+1}^{(i)} \leq 1$. It follows that $\nu_{j+1}^{(i)}$ is at most the least solution of $X_{j+1} = \frac{1}{4}(\nu_j^{(i)})^2 + \frac{1}{2}\nu_j^{(i)}X_{j+1} + \frac{1}{4}(X_{j+1})^2$, and so $\Delta_{j+1}^{(i)} \geq 2\sqrt{\Delta_j^{(i)}} - \Delta_j^{(i)} > \sqrt{\Delta_j^{(i)}}$.

By induction it follows that $\Delta_{j+1}^{(i)} > (\Delta_1^{(i)})^{2^{-j}}$. In particular,

$$\Delta_n^{(k \cdot 2^{n-1})} > \left(\Delta_1^{(k \cdot 2^{n-1})}\right)^{2^{-(n-1)}} = 2^{-k \cdot 2^{n-1} \cdot 2^{-(n-1)}} = 2^{-k}.$$

Hence, after $k \cdot 2^{n-1}$ iterations we have fewer than k valid bits. □

Notice that the proof exploits that an error in the first component gets "amplified" along the DAG of SCCs. One can also show along those lines that computing $\mu \boldsymbol{f}$ is an *ill-conditioned* problem: Consider the SPP $\boldsymbol{g}^{(n,\varepsilon)}$ obtained from $\boldsymbol{f}^{(n)}$ by replacing the first component by $1 - \varepsilon$ where $0 \le \varepsilon < 1$. If $\varepsilon = 0$, then $(\mu \boldsymbol{g}^{(n,\varepsilon)})_n = 1$, whereas if $\varepsilon = \frac{1}{2^{2^{n-1}}}$, then $(\mu \boldsymbol{g}^{(n,\varepsilon)})_n < \frac{1}{2}$. In other words, to get 1 bit of precision of $\mu \boldsymbol{g}$, one needs exponentially in n many bits in $\boldsymbol{g}$. Note that this observation is independent from any particular method to compute or approximate the least fixed point.

8. Geometrical aspects of SPPs. As shown in section 4.4 we can assume that $\boldsymbol{f}$ consists of quadratic polynomials. For quadratic polynomials the locus of zeros is also called a *quadric surface*, or more commonly *quadric*. Quadrics are one of the most fundamental class of hypersurfaces. It is therefore natural to study the quadrics induced by a quadratic SPP $\boldsymbol{f}$ and how the Newton sequence is connected to these surfaces.

Let us write $\boldsymbol{q}$ for $\boldsymbol{f} - \boldsymbol{X}$. Every component q_i of $\boldsymbol{q}$ is also a quadratic polynomial each defining a quadric denoted by

$$Q_i := \{\boldsymbol{x} \in \mathbb{R}^n \mid q_i(\boldsymbol{x}) = f_i(\boldsymbol{x}) - x_i = 0\}.$$

Finding $\mu \boldsymbol{f}$ thus corresponds to finding the least nonnegative point of intersection of these n quadrics Q_i.

Example 8.1. Consider the SPP $\boldsymbol{f}$ given by

$$\boldsymbol{f}(X, Y) = \begin{pmatrix} \frac{1}{2}X^2 + \frac{1}{4}Y^2 + \frac{1}{4} \\ \frac{1}{4}X + \frac{1}{4}XY + \frac{1}{4}Y^2 + \frac{1}{4} \end{pmatrix}$$

leading to

$$q_1(X, Y) = \frac{1}{2}X^2 + \frac{1}{4}Y^2 + \frac{1}{4} - X \quad \text{and} \quad q_2(X, Y) = \frac{1}{4}X + \frac{1}{4}XY + \frac{1}{4}Y^2 + \frac{1}{4} - Y.$$

Using standard techniques from linear algebra, one can show that q_1 defines an ellipse, while q_2 describes a parabola (see Figure 8.1).

Figure 8.1 shows the two quadrics induced by the SPP $\boldsymbol{f}$ discussed in the example above. In Figure 8.1 (a) one can recognize one of the two quadrics as an ellipse, while the other one is a parabola. In this example the Newton approximants (depicted as crosses) stay within the region enclosed by the coordinate axes and the two quadrics as shown in Figure 8.1 (b).

In this section we want to show that the image depicted in the figure in principle is the same for all clean and feasible scSPPs. That is, we show that the Newton (and Kleene) approximants always stay in the region enclosed by the coordinate axes and the quadrics. We characterize this region and study some of the properties of the quadrics restricted to this region. This eventually leads to a generalization of Newton's method (Theorem 8.13). We close the section by showing that this new method converges at least as fast as Newton's method. All missing proofs can be found in Appendix B.
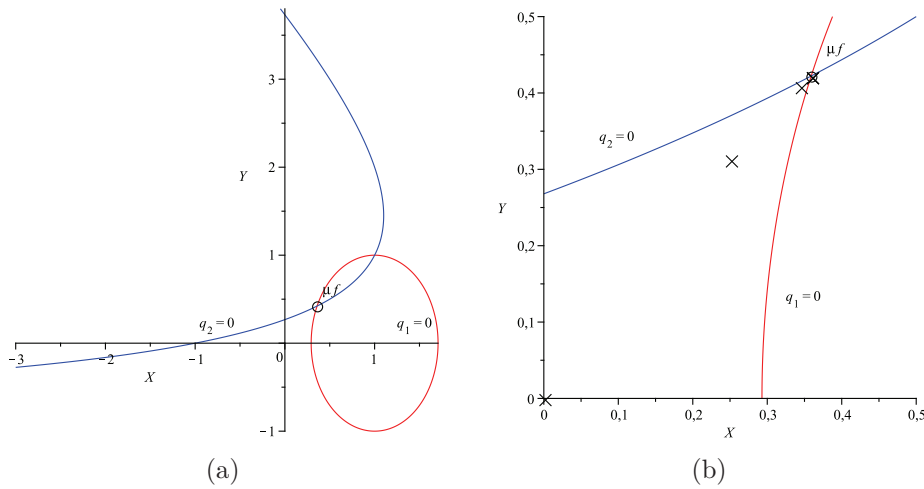
FIG. 8.1. (a) *The quadrics induced by the SPP from Example* 8.1 *with* "$q_1 = 0$" *an ellipse and* "$q_2 = 0$" *a parabola.* (b) *Close-up view of the region important for determining* $\mu\boldsymbol{f}$. *The crosses show the Newton approximants of* $\mu\boldsymbol{f}$.
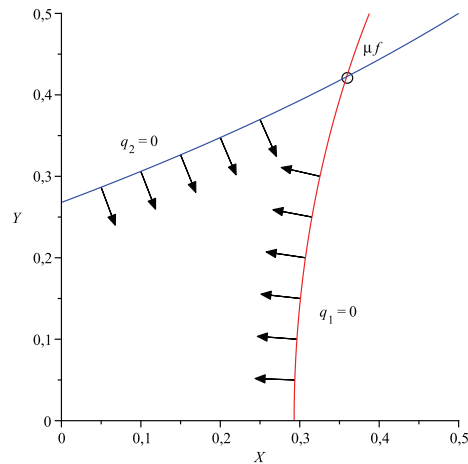

FIG. 8.2. *The normals (scaled down) of the quadrics from Example* 8.1.

Let us start with the properties of the quadrics Q_i. We restrict our attention to the region $[\boldsymbol{0}, \mu\boldsymbol{f})$. For this we set

$$M_i := Q_i \cap [\boldsymbol{0}, \mu\boldsymbol{f}) = \{\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f}) \mid q_i(\boldsymbol{x}) = 0\}.$$

We start by showing that for every $\boldsymbol{x} \in M_i$ the gradient $q_i'(\boldsymbol{x})$ in $\boldsymbol{x}$ at M_i does not vanish. As $q_i'(\boldsymbol{x})$ is perpendicular to the tangent plane in $\boldsymbol{x}$ at M_i, this means that the normal of the tangent plane is determined by $q_i'(\boldsymbol{x})$ (up to orientation). See Figure 8.2 for an example. This will later allow us to apply the implicit function theorem.

LEMMA 8.2. *For every quadric* q_i *induced by a clean and feasible scSPP* $\boldsymbol{f}$ *we have*

$$q_i'(\boldsymbol{x}) = (\partial_{X_1} q_i(\boldsymbol{x}), \partial_{X_2} q_i(\boldsymbol{x}), \dots, \partial_{X_n} q_i(\boldsymbol{x})) \neq \boldsymbol{0} \quad and \quad \partial_{X_i} q_i(\boldsymbol{x}) < 0 \text{ for all } \boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f}).$$

In the following, for $i \in \{1, \ldots, n\}$ we write $\boldsymbol{x}_{-i}$ for the vector $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ and define $(\boldsymbol{x}_{-i}, x_i)$ to also denote the original vector $\boldsymbol{x}$.

We next show that there exists a complete parametrization of "the lower part" of M_i. By "lower part" we refer to the set

$$S_i := \{\boldsymbol{x} \in M_i \mid \forall \boldsymbol{y} \in M_i : (\boldsymbol{x}_{-i} = \boldsymbol{y}_{-i}) \Rightarrow x_i \leq y_i\},$$

i.e., the points $\boldsymbol{x} \in M_i$ such that there is no point $\boldsymbol{y}$ with the same non-i-components but smaller i-component. Taking a look at Figure 8.1, the surfaces S_1 and S_2 are those parts of M_1, respectively, M_2, which delimit that part of $\mathbb{R}^2_{\geq 0}$ shown in Figure 8.1 (b).

If $\boldsymbol{x} \in S_i$, then x_i is the least nonnegative root of the (at most) quadratic polynomial $q_i(X_i, \boldsymbol{x}_{-i})$. As we will see, these roots can also be represented by the following functions.

DEFINITION 8.3. *For a clean and feasible scSPP $\boldsymbol{f}$ we define for all $k \in \mathbb{N}$ the polynomial $h_i^{(k)}$ by*

$$h_i^{(0)}(\boldsymbol{X}_{-i}) := f_i[i/0](\boldsymbol{X}_{-i}), \quad h_i^{(k+1)}(\boldsymbol{X}_{-i}) := f_i[i/h_i^{(k)}(\boldsymbol{X}_{-i})](\boldsymbol{X}_{-i}).$$

The function $h_i(\boldsymbol{X})$ is then defined pointwise by

$$h_i(\boldsymbol{x}_{-i}) := \lim_{k \to \infty} h_i^{(k)}(\boldsymbol{x}_{-i})$$

for all $\boldsymbol{x}_{-i} \in [\boldsymbol{0}, \mu \boldsymbol{f}_{-i}]$.

We show in Appendix B (see Proposition B.1) that the function h_i is well-defined and exists. We therefore can parameterize the surface S_i w.r.t. the remaining variables $\boldsymbol{X}_{-i}$; i.e., h_i is the "height" of the surface S_i above the "ground" $X_i = 0$.

By the preceding proposition the map

$$\boldsymbol{p}_i : [\boldsymbol{0}, \mu \boldsymbol{f}_{-i}] \to [\boldsymbol{0}, \mu \boldsymbol{f}] : \boldsymbol{x}_{-i} \mapsto (x_1, \ldots, x_{i-1}, h_i(\boldsymbol{x}_{-i}), x_{i+1}, \ldots, x_n)$$

gives us a pointwise parametrization of S_i. We want to show that $\boldsymbol{p}_i$ is continuously differentiable. For this it suffices to show that h_i is continuously differentiable, which follows easily from the implicit function theorem (see, e.g., [30]).

LEMMA 8.4. *h_i is continuously differentiable with*

$$\partial_{X_j} h_i(\boldsymbol{x}_{-i}) = \frac{\partial_{X_j} f_i(\boldsymbol{x})}{-\partial_{X_i} q_i(\boldsymbol{x})} = \frac{\partial_{X_j} q_i(\boldsymbol{x})}{-\partial_{X_i} q_i(\boldsymbol{x})} \text{ for } \boldsymbol{x} \in S_i \text{ and } j \neq i.$$

In particular, $\partial_{X_j} h_i$ is monotonically increasing with $\boldsymbol{x}$.

COROLLARY 8.5. *The map*

$$\boldsymbol{p}_i : [\boldsymbol{0}, \mu \boldsymbol{f}_{-i}] \to [\boldsymbol{0}, \mu \boldsymbol{f}] : \boldsymbol{x}_{-i} \mapsto (x_1, \ldots, x_{i-1}, h_i(\boldsymbol{x}_{-i}), x_{i+1}, \ldots, x_n)$$

is continuously differentiable and a local parametrization of the manifold S_i.

Example 8.6. For the SPP $\boldsymbol{f}$ defined in Example 8.1 we can simply solve $q_1(X, Y)$ for X, leading to

$$h_1(Y) = 1 - \sqrt{\frac{1}{2}(1 - Y^2)}.$$

The important point is that by the previous result we know that this function has to be defined on $[0, \mu \boldsymbol{f}_2]$, and differentiable on $[0, \mu \boldsymbol{f}_2)$. Similarly, we get

$$h_2(X) = 2 - \frac{1}{2}X - \frac{1}{2}\sqrt{X^2 - 12X + 12}.$$

Figure 8.1 (b) conveys the impression that the surfaces S_i are convex w.r.t. the parameterizations $\boldsymbol{p}_i$. As we have seen, the functions h_i are monotonically increasing. Thus, in the case of two dimensions the functions h_i even have to be strictly monotonically increasing (as $\boldsymbol{f}$ is strongly connected), so that the surfaces S_i are indeed convex. (Recall that a surface S is convex in a point $\boldsymbol{x} \in S$ if S is located completely on one side of the tangent plane at S in $\boldsymbol{x}$.) But in the case of more than two variables this no longer needs to hold.

Example 8.7. The equation

$$Z = \frac{1}{8}X^2 + \frac{3}{4}XY + \frac{1}{8}Y^2 + \frac{1}{4}$$

is an admissible part of any SPP. It defines the hyperbolic paraboloid depicted in Figure 8.3, which is clearly not convex.
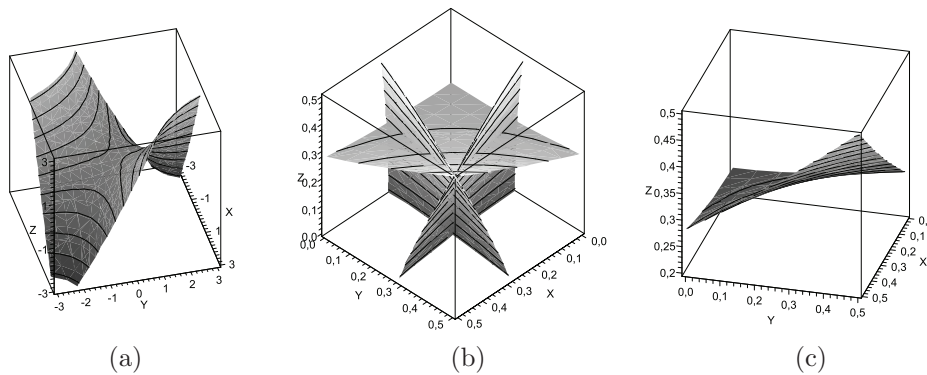

(a) (b) (c)

FIG. 8.3. (a) *The hyperbolic paraboloid defined by* $Z = \frac{1}{8}X^2 + \frac{3}{4}XY + \frac{1}{8}Y^2 + \frac{1}{4}$ *for* $X, Y, Z \in [-10, 10]$. (b) *A visualization of an SPP consisting of three copies of the quadric of* (a) *with* $\mu\boldsymbol{f} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ *the upper apex.* (c) *One of the three quadrics of* (b) *over* $[\boldsymbol{0}, \mu\boldsymbol{f}]$. *Clearly, even limited to this range the surface is not convex.*

Still, as shown in Lemma 2.3 it holds for all $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{y}$ that

$$\boldsymbol{x} + \boldsymbol{f}'(\boldsymbol{x}) \cdot \boldsymbol{y} \leq \boldsymbol{f}(\boldsymbol{x} + \boldsymbol{y}).$$

It now follows (see the following lemma) that the surfaces S_i have the property that for every $\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f})$ the "relevant" part of S_i for determining $\mu\boldsymbol{f}$, i.e., $S_i \cap [\boldsymbol{x}, \mu\boldsymbol{f}]$, is located on the same side of the tangent plane at S_i in $\boldsymbol{x}$ (see Figure 8.4).

LEMMA 8.8. *For all* $\boldsymbol{x} \in S_i$ *we have*

$$\text{for all } \boldsymbol{y} \in S_i \cap [\boldsymbol{x}, \mu\boldsymbol{f}] : q_i'(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) \leq 0.$$

In particular

$$\text{for all } \boldsymbol{y} \in S_i \cap [\boldsymbol{x}, \mu\boldsymbol{f}] : y_i \geq x_i + \sum_{j \neq i} \partial_{X_j} h_i(\boldsymbol{x}_{-i}) \cdot (y_j - x_j).$$

Consider now the set

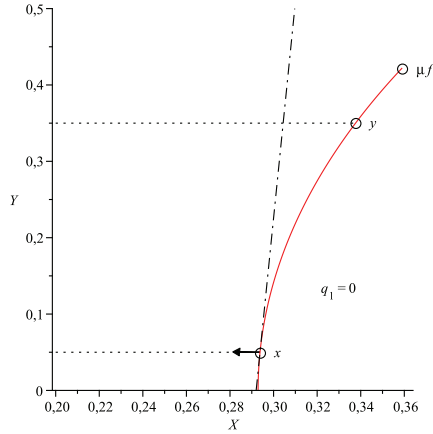$$R := \bigcap_{i=1}^{n} \{\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f}) \mid x_i \leq h_i(\boldsymbol{x}_{-i})\},$$

FIG. 8.4. *The graphic shows the quadric defined by $q_1 = 0$ with the tangent and normal in $\boldsymbol{x}$ at S_1. Every point $\boldsymbol{y}$ of S_1 above $\boldsymbol{x}$ is located on the same side of the the tangent. More precisely, we have $\nabla q_1|_{\boldsymbol{x}} \cdot (\boldsymbol{y} - \boldsymbol{x}) \le 0$.*

i.e., the region of $[\boldsymbol{0}, \mu\boldsymbol{f})$ delimited by the coordinate axes and the surfaces S_i. Note that the gradient $q_i'(\boldsymbol{x})$ for $\boldsymbol{x} \in S_i$ points from S_i into R (see Figure 8.2).

PROPOSITION 8.9. *It holds that*

$$\boldsymbol{x} \in R \Leftrightarrow \boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f}) \wedge \boldsymbol{q}(\boldsymbol{x}) \ge \boldsymbol{0}.$$

From this last result it now easily follows that R is indeed the region of $[\boldsymbol{0}, \mu\boldsymbol{f})$ in which all Newton and Kleene steps are located.

THEOREM 8.10. *Let $\boldsymbol{f}$ be a clean and feasible scSPP. All Newton and Kleene steps starting from $\boldsymbol{0}$ lie within R, i.e.,*

$$\boldsymbol{\nu}^{(i)}, \boldsymbol{\kappa}^{(i)} \in R \quad (\text{for all } i \in \mathbb{N}).$$

Proof. For an scSPP we have $\boldsymbol{\kappa}^{(i)}, \boldsymbol{\nu}^{(i)} \in [\boldsymbol{0}, \mu\boldsymbol{f})$ for all i. Further, $\boldsymbol{\kappa}^{(i)} \le \boldsymbol{\kappa}^{(i+1)} = \boldsymbol{f}(\boldsymbol{\kappa}^{(i)})$ and $\boldsymbol{\nu}^{(i)} \le \boldsymbol{f}(\boldsymbol{\nu}^{(i)})$ holds for all i, too. $\quad\square$

In the rest of this section we will use the results regarding R and the surfaces S_i for interpreting Newton's method geometrically and for obtaining a generalization of Newton's method.

The preceding results suggest another way of determining $\mu\boldsymbol{f}$ (see Figure 8.5): Let $\boldsymbol{x}$ be some point inside of R. We may move from $\boldsymbol{x}$ onto one of the surface S_i by going upward along the line $\boldsymbol{x} + t \cdot \boldsymbol{e}_i$, which gives us the point $\boldsymbol{p}_i(\boldsymbol{x}_{-i}) = (\boldsymbol{x}_{-i}, h_i(\boldsymbol{x}_{-i}))$. As $\boldsymbol{x} \in R$, we have $\boldsymbol{x}, \boldsymbol{p}_i(\boldsymbol{x}_{-i}) \le \mu\boldsymbol{f}$. Consider now the tangent plane

$$T_i|_{\boldsymbol{x}} = \left\{ \boldsymbol{y} \in \mathbb{R}^n \mid q_i'(\boldsymbol{p}_i(\boldsymbol{x}_{-i})) \cdot (\boldsymbol{y} - \boldsymbol{p}_i(\boldsymbol{x}_{-i})) = 0 \right\}$$

at S_i in $\boldsymbol{p}_i(\boldsymbol{x}_{-i})$. Recall that by Lemma 8.8 we have

$$\text{for all } \boldsymbol{y} \in S_i \cap [\boldsymbol{p}_i(\boldsymbol{x}_{-i}), \mu\boldsymbol{f}) : q_i'(\boldsymbol{p}_i(\boldsymbol{x}_{-i})) \cdot (\boldsymbol{y} - \boldsymbol{p}_i(\boldsymbol{x}_{-i})) \le 0;$$

i.e., the part of S_i relevant for determining $\mu\boldsymbol{f}$ is located completely below (w.r.t. $q_i'(()\boldsymbol{p}_i(\boldsymbol{x}_{-i})))$ this tangent plane. By continuity this also has to hold for $\boldsymbol{y} = \mu\boldsymbol{f}$. Hence, when taking the intersection of all the tangent planes T_1 to T_n this gives us
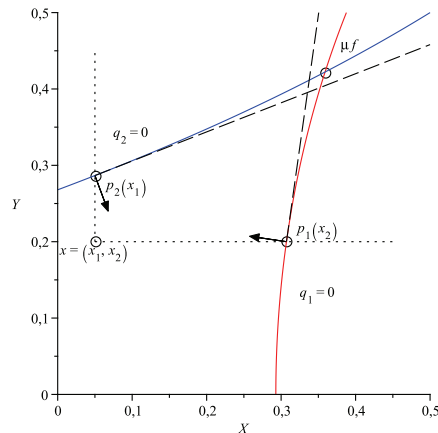
FIG. 8.5. *Given a point $\boldsymbol{x}$ inside of R the intersection of the tangents at the quadrics in the points $p_1(x_2)$, respectively, $p_2(x_1)$, is also located inside of R, yielding a better approximation of $\mu\boldsymbol{f}$.*

again a point $\mathcal{T}(\boldsymbol{x})$ inside of R. That this point $\mathcal{T}(\boldsymbol{x})$ exists and is uniquely determined is shown in the following lemma.

LEMMA 8.11. *Let $\boldsymbol{f}$ be a clean and feasible scSPP. Let $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)} \in [\boldsymbol{0}, \mu\boldsymbol{f})$. Then the matrix*

$$\begin{pmatrix} q_1'(\boldsymbol{x}^{(1)}) \\ \vdots \\ q_n'(\boldsymbol{x}^{(n)}) \end{pmatrix}$$

is regular; i.e., the vectors $\{q_i'(\boldsymbol{x}^{(i)}) \mid i = 1, \ldots, n\}$ are linearly independent.

By this lemma the normals at the quadrics in the points $\boldsymbol{p}_i(\boldsymbol{x}_{-i})$ for $\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f})$ are linearly independent. Thus, there exists a unique point of intersection of tangent planes at the quadrics in these points. Of course, in general the values $h_i(\boldsymbol{x}_{-i})$ can be irrational. The following definition takes this in account by requiring only that underapproximations η_i of $h_i(\boldsymbol{x}_{-i})$ are known.

DEFINITION 8.12. *Let $\boldsymbol{x} \in R$. For $i = 1, \ldots, n$ fix some $\eta_i \in [x_i, h_i(\boldsymbol{x}_{-i})]$, and set $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$. We then let $\mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x})$ denote the solution of*

$$q_i'((\boldsymbol{x}_{-i}, \eta_i))(\boldsymbol{X} - (\boldsymbol{x}_{-i}, \eta_i)) = -q_i((\boldsymbol{x}_{-i}, \eta_i)) \quad (i = 1, \ldots, n).$$

We drop the subscript and simply write $\mathcal{T}$ in the case of $\eta_i = h_i(\boldsymbol{x}_{-i})$ for $i = 1, \ldots, n$.

Note that the operator $\mathcal{T}_{\boldsymbol{x}}$ is the Newton operator $\mathcal{N}$.

THEOREM 8.13. *Let $\boldsymbol{f}$ be a clean and feasible scSPP. Let $\boldsymbol{x} \in R$. For $i = 1, \ldots, n$ fix some $\eta_i \in [x_i, h_i(\boldsymbol{x}_{-i})]$, and set $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$. We then have*

$$\boldsymbol{x} \leq \mathcal{N}(\boldsymbol{x}) \leq \mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x}) \leq \mathcal{T}(\boldsymbol{x}) \leq \mu\boldsymbol{f}.$$

Further, the operator $\mathcal{T}$ is monotone on R; i.e., for any $\boldsymbol{y} \in R$ with $\boldsymbol{x} \leq \boldsymbol{y}$ it holds that $\mathcal{T}(\boldsymbol{x}) \leq \mathcal{T}(\boldsymbol{y})$.

By Theorem 8.13, replacing the Newton operator $\mathcal{N}$ by $\mathcal{T}$ gives a variant of Newton's method which converges at least as fast.

We do not know whether this variant is substantially faster. See Figure 8.6 for a geometrical interpretation of both methods.
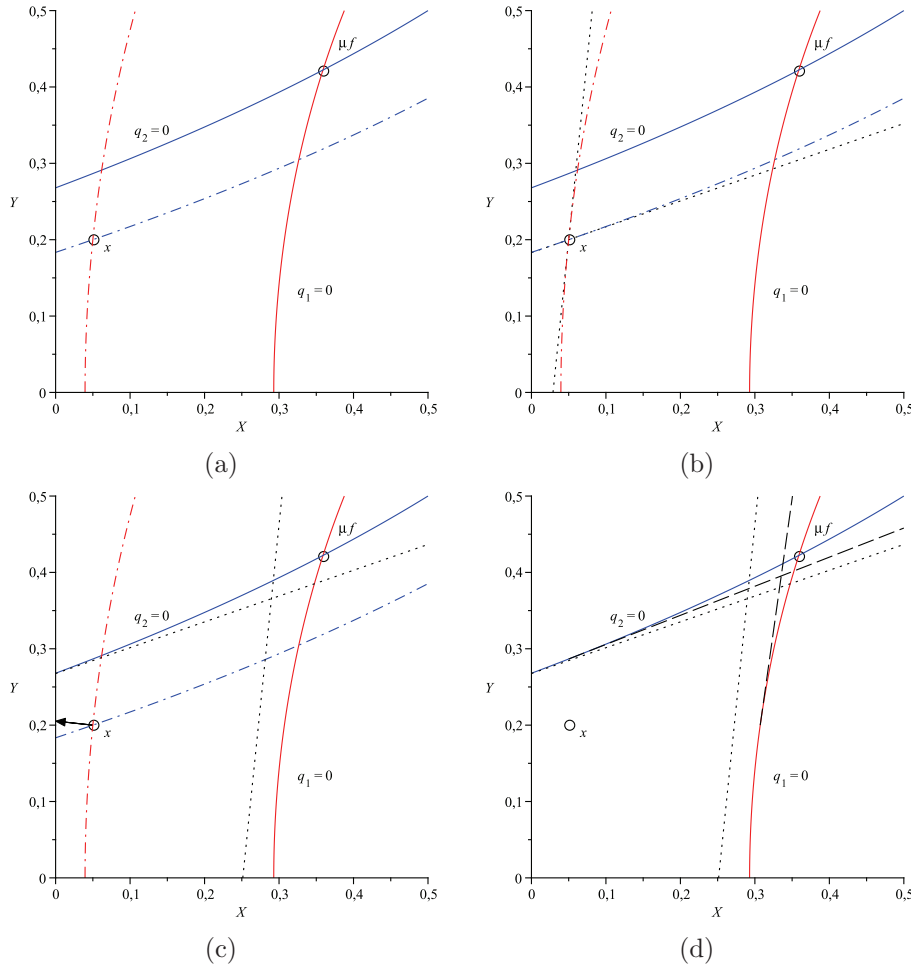
FIG. 8.6. *Geometrical interpretation of Newton's method: (a) Given a point $\boldsymbol{x} \in R$, Newton's method first considers the "enlarged" quadrics defined by $q_i(\boldsymbol{X}) = q_i(\boldsymbol{x})$ (dashed and dotted) which contain the current approximation $\boldsymbol{x}$. (b) Then the tangents in $\boldsymbol{x}$ at these enlarged quadrics are computed (dotted), i.e., $q_i'(\boldsymbol{x}) \cdot (\boldsymbol{X} - \boldsymbol{x}) = 0$. (c) Finally, these tangents are corrected by moving them towards the actual quadrics, i.e., $q_i'(\boldsymbol{x}) \cdot (\boldsymbol{X} - \boldsymbol{x}) = -q_i(\boldsymbol{x})$. The intersection of these corrected tangents gives the next Newton approximation. (d) A comparison between $\mathcal{N}(\boldsymbol{x})$ and $\mathcal{T}(\boldsymbol{x})$: $\mathcal{N}(\boldsymbol{x})$, respectively, $\mathcal{T}(\boldsymbol{x})$, is given by the intersection of the dotted, respectively, dashed lines. Clearly, we have $\mathcal{N}(\boldsymbol{x}) \leq \mathcal{T}(\boldsymbol{x})$.*

9. Conclusions. We have studied the convergence order and convergence rate of Newton's method for fixed-point equations of systems of positive polynomials (SPP equations). These equations appear naturally in the analysis of several stochastic computational models that have been intensely studied in recent years, and they also play a central role in the theory of stochastic branching processes.

The restriction to positive coefficients leads to strong results. For arbitrary polynomial equations Newton's method may not converge or may converge only locally, i.e., when started at a point sufficiently close to the solution. We have extended a result by Etessami and Yannakakis [16] and shown that for SPP equations the method always converges starting at **0**. Moreover, we have proved that the method has at least

linear convergence order and have determined the asymptotic convergence rate. To the best of our knowledge, this is the first time that a lower bound on the convergence order is proved for a significant class of equations with a trivial membership test.[4] Finally, in the case of scSPPs we have also obtained upper bounds on the threshold, i.e., the number of iterations necessary to reach the "steady state" in which valid bits are computed at the asymptotic rate. These results lead to practical tests for checking whether the least fixed point of a strongly connected SPP exceeds a given bound.

It is worth mentioning that in a recent paper we study the behavior of Newton's method when arithmetic operations only have a fixed accuracy [8]. We develop an algorithm for a relevant class of SPPs that computes iterations of Newton's method increasing the accuracy on demand. A simple test applied after each iteration decides if the round-off errors have become too large, in which case the accuracy is increased.

There are still at least two important open questions. The first one is, can one provide a bound on the threshold valid for arbitrary SPPs, and not only for strongly connected ones? Since SPPs cannot be solved exactly in general, we cannot first compute the exact solution for the bottom SCCs, insert it in the SCCs above them, and iterate. We can only compute an approximation, and we are not currently able to bound the propagation of the error. For the second question, say that Newton's method is *polynomial* for a class of SPP equations if there is a polynomial $p(x, y, z)$ such that for every $k \geq 0$ and for every system in the class with n equations and coefficients of size m, the $p(n, m, k)$th Newton approximant $\boldsymbol{\nu}^{(p(n,m,k))}$ has k valid bits. We have proved in Theorem 5.10 that Newton's method is polynomial for strongly connected SPPs $\boldsymbol{f}$ satisfying $\boldsymbol{f}(\mathbf{0}) \succ \mathbf{0}$; for this class one can take $p(n, m, k) = 7mn + k$. We have also exhibited in section 7 a class for which computing the first bit of the least solution takes 2^n iterations. The members of this class, however, are not strongly connected, and this is the fact we have exploited to construct them. So the following question remains open: Is Newton's method polynomial for strongly connected SPPs?

Appendix A. Proof of Lemma 6.3. The proof of Lemma 6.3 is by a sequence of lemmas. The proof of Lemma A.2 and, consequently, the proof of Lemma 6.3 are nonconstructive in the sense that we cannot give a particular $C_{\boldsymbol{f}}$. Therefore, we often use the equivalence of norms, disregard the constants that link them, and state the results in terms of an arbitrary norm.

First we prove the following lemma on cone vectors.

LEMMA A.1. *Let $\boldsymbol{f}$ be a clean and feasible scSPP 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 SPP, every component of $\boldsymbol{f}'(\mu \boldsymbol{f})$ is nonnegative. So,

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

W.l.o.g. let $d_1 > 0$. As $\boldsymbol{f}$ is strongly connected, there is for all j with $1 \leq j \leq n$ an $r_j \leq n$ such that $(\boldsymbol{f}'(\mu \boldsymbol{f})^{r_j})_{j1} > 0$. Hence, $(\boldsymbol{f}'(\mu \boldsymbol{f})^{r_j}\boldsymbol{d})_j > 0$ for all j. With the above inequality chain, it follows that $d_j \geq (\boldsymbol{f}'(\mu \boldsymbol{f})^{r_j}\boldsymbol{d})_j > 0$. So, $\boldsymbol{d} \succ \mathbf{0}$. ☐

The following two lemmas, Lemmas A.2 and A.3, provide a lower bound on $\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{x}\|$ for an "almost-fixed-point" $\boldsymbol{x}$.

LEMMA A.2. *Let $\boldsymbol{f}$ be a quadratic, clean, and feasible SPP without linear terms, i.e., $\boldsymbol{f}(\boldsymbol{X}) = B(\boldsymbol{X}, \boldsymbol{X}) + \boldsymbol{c}$, where B is a bilinear map, and $\boldsymbol{c}$ is a constant vector. Let*

---

[4]Notice the contrast with the classical result stating that if $(\mathrm{Id} - \boldsymbol{f}'(\mu \boldsymbol{f}))$ is nonsingular, then Newton's method has exponential convergence order; here the membership test is highly nontrivial, and, as far as we know, as hard as computing $\mu \boldsymbol{f}$ itself.

$f(X)$ be nonconstant in every component. Let $R \, \dot\cup \, S = \{1, \ldots, n\}$ with $S \neq \emptyset$. Let every component depend on every S-component and not on any R-component. Then there is a constant $C_f > 0$ such that

$$\|f(\mu f - \delta) - (\mu f - \delta)\| \geq C_f \cdot \|\delta\|^2$$

for all δ with $0 \leq \delta \leq \mu f$.

  Proof. With the given component dependencies we can write $f(X)$ as follows:

$$f(X) = \begin{pmatrix} f_R(X) \\ f_S(X) \end{pmatrix} = \begin{pmatrix} B_R(X_S, X_S) + c_R \\ B_S(X_S, X_S) + c_S \end{pmatrix}.$$

A straightforward calculation shows

$$e(\delta) := f(\mu f - \delta) - (\mu f - \delta) = (\mathrm{Id} - f'(\mu f))\delta + B(\delta, \delta) .$$

Furthermore, $\partial_{X_R} f$ is constant zero in all entries, so

$$e_R(\delta) = \delta_R - \partial_{X_S} f_R(\mu f) \cdot \delta_S + B_R(\delta_S, \delta_S),$$
$$e_S(\delta) = \delta_S - \partial_{X_S} f_S(\mu f) \cdot \delta_S + B_S(\delta_S, \delta_S) .$$

Notice that for every real number $r > 0$ we have

$$\min_{0 \leq \delta \leq \mu f, \|\delta\| \geq r} \frac{\|e(\delta)\|}{\|\delta\|^2} > 0 ,$$

because otherwise $\mu f - \delta < \mu f$ would be a fixed point of f. We have to show

$$\inf_{0 \leq \delta \leq \mu f, \|\delta\| > 0} \frac{\|e(\delta)\|}{\|\delta\|^2} > 0.$$

Assume, for a contradiction, that this infimum equals zero. Then there exists a sequence $(\delta^{(i)})_{i \in \mathbb{N}}$ with $0 \leq \delta^{(i)} \leq \mu f, \|\delta^{(i)}\| > 0$ such that $\lim_{i \to \infty} \|\delta^{(i)}\| = 0$ and $\lim_{i \to \infty} \frac{\|e(\delta^{(i)})\|}{\|\delta^{(i)}\|^2} = 0$. Define $r^{(i)} := \|\delta^{(i)}\|$ and $d^{(i)} := \frac{\delta^{(i)}}{\|\delta^{(i)}\|}$. Notice that $d^{(i)} \in \{d \in \mathbb{R}^n_{\geq 0} \mid \|d\| = 1\} =: D$, where D is compact. So some subsequence of $(d^{(i)})_{i \in \mathbb{N}}$, say w.l.o.g. the sequence $(d^{(i)})_{i \in \mathbb{N}}$ itself, converges to some vector $d^* \in D$. By our assumption we have

$$(A.1) \qquad \left\| e(\delta^{(i)}) \right\| / \left\| \delta^{(i)} \right\|^2 = \left\| \frac{1}{r^{(i)}} (\mathrm{Id} - f'(\mu f)) d^{(i)} + B(d^{(i)}, d^{(i)}) \right\| \longrightarrow 0 .$$

As $B(d^{(i)}, d^{(i)})$ is bounded, $\frac{1}{r^{(i)}}(\mathrm{Id} - f'(\mu f))d^{(i)}$ must be bounded, too. Since $r^{(i)}$ converges to 0, $\|(\mathrm{Id} - f'(\mu f))d^{(i)}\|$ must converge to 0, so

$$(\mathrm{Id} - f'(\mu f))d^* = 0 .$$

In particular, $\left((\mathrm{Id} - f'(\mu f))d^*\right)_R = d_R^* - \partial_{X_S} f_R(\mu f) \cdot d_S^* = 0$. So we have $d_S^* > 0$, because $d_S^* = 0$ would imply $d_R^* = 0$, which would contradict $d^* > 0$.

  In the remainder of the proof we focus on f_S. Define the scSPP $g(X_S) := f_S(X)$. Notice that $\mu g = \mu f_S$. We can apply Lemma A.1 to g and d_S^* and obtain $d_S^* \succ 0$. As $f_S(X)$ is nonconstant we get $B_S(d_S^*, d_S^*) \succ 0$. By (A.1), $\frac{1}{r^{(i)}}(\mathrm{Id} - g'(\mu g))d_S^{(i)}$

converges to $-B_S(\boldsymbol{d}_S^*, \boldsymbol{d}_S^*) \prec \boldsymbol{0}$. So there is a $j \in \mathbb{N}$ such that $(\mathrm{Id} - \boldsymbol{g}'(\mu\boldsymbol{g}))\boldsymbol{d}_S^{(j)} \prec \boldsymbol{0}$. Let $\widetilde{\boldsymbol{\delta}} := r\boldsymbol{d}^{(j)}$ for some small enough $r > 0$ such that $\boldsymbol{0} < \widetilde{\boldsymbol{\delta}}_S \le \mu\boldsymbol{g}$ and

$$\begin{aligned}
\boldsymbol{e}_S(\widetilde{\boldsymbol{\delta}}) &= (\mathrm{Id} - \boldsymbol{g}'(\mu\boldsymbol{g}))\widetilde{\boldsymbol{\delta}}_S + B_S(\widetilde{\boldsymbol{\delta}}_S, \widetilde{\boldsymbol{\delta}}_S) \\
&= r(\mathrm{Id} - \boldsymbol{g}'(\mu\boldsymbol{g}))\boldsymbol{d}_S^{(j)} + r^2 B_S(\boldsymbol{d}_S^{(j)}, \boldsymbol{d}_S^{(j)}) \prec \boldsymbol{0} .
\end{aligned}$$

So we have $\boldsymbol{g}(\mu\boldsymbol{g} - \widetilde{\boldsymbol{\delta}}_S) \prec \mu\boldsymbol{g} - \widetilde{\boldsymbol{\delta}}_S$. However, $\mu\boldsymbol{g}$ is the least point $\boldsymbol{x}$ with $\boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{x}$. Thus we get the desired contradiction. $\square$

LEMMA A.3. *Let $\boldsymbol{f}$ be a quadratic, clean, and feasible scSPP. Then there is a constant $C_{\boldsymbol{f}} > 0$ such that*

$$\|\boldsymbol{f}(\mu\boldsymbol{f} - \boldsymbol{\delta}) - (\mu\boldsymbol{f} - \boldsymbol{\delta})\| \ge C_{\boldsymbol{f}} \cdot \|\boldsymbol{\delta}\|^2$$

for all $\boldsymbol{\delta}$ with $\boldsymbol{0} \le \boldsymbol{\delta} \le \mu\boldsymbol{f}$.

Proof. Write $\boldsymbol{f}(\boldsymbol{X}) = B(\boldsymbol{X}, \boldsymbol{X}) + L\boldsymbol{X} + \boldsymbol{c}$ for a bilinear map B, a matrix L, and a constant vector $\boldsymbol{c}$. By Theorem 4.1.2. the matrix $L^* = (\mathrm{Id} - L)^{-1} = (\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{0}))^{-1}$ exists. Define the SPP $\widetilde{\boldsymbol{f}}(\boldsymbol{X}) := L^*B(\boldsymbol{X}, \boldsymbol{X}) + L^*\boldsymbol{c}$. A straightforward calculation shows that the sets of fixed points of $\boldsymbol{f}$ and $\widetilde{\boldsymbol{f}}$ coincide and that

$$\boldsymbol{f}(\mu\boldsymbol{f} - \boldsymbol{\delta}) - (\mu\boldsymbol{f} - \boldsymbol{\delta}) = (\mathrm{Id} - L)\left(\widetilde{\boldsymbol{f}}(\mu\boldsymbol{f} - \boldsymbol{\delta}) - (\mu\boldsymbol{f} - \boldsymbol{\delta})\right) .$$

Further, if $\sigma_n(\mathrm{Id} - L)$ denotes the smallest singular value of $\mathrm{Id} - L$, we have by basic facts about singular values (see [22, Chapter 3]) that

$$\left\|(\mathrm{Id} - L)\left(\widetilde{\boldsymbol{f}}(\mu\boldsymbol{f} - \boldsymbol{\delta}) - (\mu\boldsymbol{f} - \boldsymbol{\delta})\right)\right\|_2 \ge \sigma_n(\mathrm{Id} - L)\left\|\widetilde{\boldsymbol{f}}(\mu\boldsymbol{f} - \boldsymbol{\delta}) - (\mu\boldsymbol{f} - \boldsymbol{\delta})\right\|_2 .$$

Note that $\sigma_n(\mathrm{Id} - L) > 0$ because $\mathrm{Id} - L$ is invertible. So it suffices to show that

$$\left\|\widetilde{\boldsymbol{f}}(\mu\boldsymbol{f} - \boldsymbol{\delta}) - (\mu\boldsymbol{f} - \boldsymbol{\delta})\right\| \ge C_{\boldsymbol{f}} \cdot \|\boldsymbol{\delta}\|^2 .$$

If $\boldsymbol{f}(\boldsymbol{X})$ is linear (i.e., $B(\boldsymbol{X}, \boldsymbol{X}) \equiv \boldsymbol{0}$), then $\widetilde{\boldsymbol{f}}(\boldsymbol{X})$ is constant and we have $\|\widetilde{\boldsymbol{f}}(\mu\boldsymbol{f} - \boldsymbol{\delta}) - (\mu\boldsymbol{f} - \boldsymbol{\delta})\| = \|\boldsymbol{\delta}\|$, so we are done in that case. Hence we can assume that some component of $B(\boldsymbol{X}, \boldsymbol{X})$ is not the zero polynomial. It remains to argue that $\widetilde{\boldsymbol{f}}$ satisfies the preconditions of Lemma A.2. By definition, $\widetilde{\boldsymbol{f}}$ does not have linear terms. Define

$$S := \{i \mid 1 \le i \le n, \ X_i \text{ is contained in a component of } B(\boldsymbol{X}, \boldsymbol{X})\} .$$

Notice that S is nonempty. Let $i_0, i_1, \ldots, i_m, i_{m+1}$ $(m \ge 0)$ be any sequence such that, in $\boldsymbol{f}$, for all j with $0 \le j < m$ the component i_j depends directly on i_{j+1} via a linear term and i_m depends directly on i_{m+1} via a quadratic term. Then i_0 depends directly on i_{m+1} via a quadratic term in $L^m B(\boldsymbol{X}, \boldsymbol{X})$ and hence also in $\widetilde{\boldsymbol{f}}$. So all components are nonconstant and depend (directly or indirectly) on every S-component. Furthermore, no component depends on a component that is not in S, because $L^*B(\boldsymbol{X}, \boldsymbol{X})$ contains only S-components. Thus, Lemma A.2 can be applied, and the statement follows. $\square$

The following lemma gives a bound on the propagation error for the case that $\boldsymbol{f}$ has a single top SCC.

LEMMA A.4. *Let $\boldsymbol{f}$ be a quadratic, clean, and feasible SPP. Let $S \subseteq \{1, \ldots, n\}$ be the single top SCC of $\boldsymbol{f}$. Let $L := \{1, \ldots, n\} \setminus S$. Then there is a constant $C_{\boldsymbol{f}} \ge 0$ such that*

$$\|\mu\boldsymbol{f}_S - \widetilde{\boldsymbol{\mu}}_S\| \le C_{\boldsymbol{f}} \cdot \sqrt{\|\mu\boldsymbol{f}_L - \boldsymbol{x}_L\|}$$

for all $\boldsymbol{x}_L$ with $\boldsymbol{0} \le \boldsymbol{x}_L \le \mu\boldsymbol{f}_L$, where $\widetilde{\boldsymbol{\mu}}_S := \mu\left(\boldsymbol{f}_S[\boldsymbol{X}_L/\boldsymbol{x}_L]\right)$.

Proof. We write $\boldsymbol{f}_S(\boldsymbol{X}) = \boldsymbol{f}_S(\boldsymbol{X}_S, \boldsymbol{X}_L)$ in the following.

If S is a trivial SCC, then $\mu\boldsymbol{f}_S = \boldsymbol{f}_S(\boldsymbol{0}, \mu\boldsymbol{f}_L)$ and $\widetilde{\boldsymbol{\mu}}_S = \boldsymbol{f}_S(\boldsymbol{0}, \boldsymbol{x}_L)$. In this case we have with Taylor's theorem (cf. Lemma 2.3)

$$
\begin{aligned}
\|\mu\boldsymbol{f}_S - \widetilde{\boldsymbol{\mu}}_S\| &= \|\boldsymbol{f}_S(\boldsymbol{0}, \mu\boldsymbol{f}_L) - \boldsymbol{f}_S(\boldsymbol{0}, \boldsymbol{x}_L)\| \\
&\leq \|\partial_{\boldsymbol{X}}\boldsymbol{f}_S(\boldsymbol{0}, \mu\boldsymbol{f}_L) \cdot (\mu\boldsymbol{f}_L - \boldsymbol{x}_L)\| \\
&\leq \|\partial_{\boldsymbol{X}}\boldsymbol{f}_S(\boldsymbol{0}, \mu\boldsymbol{f}_L)\| \cdot \|\mu\boldsymbol{f}_L - \boldsymbol{x}_L\| \\
&= \|\partial_{\boldsymbol{X}}\boldsymbol{f}_S(\boldsymbol{0}, \mu\boldsymbol{f}_L)\| \cdot \sqrt{\|\mu\boldsymbol{f}_L - \boldsymbol{x}_L\|} \cdot \sqrt{\|\mu\boldsymbol{f}_L - \boldsymbol{x}_L\|} \\
&\leq \|\partial_{\boldsymbol{X}}\boldsymbol{f}_S(\boldsymbol{0}, \mu\boldsymbol{f}_L)\| \cdot \sqrt{\|\mu\boldsymbol{f}_L\|} \cdot \sqrt{\|\mu\boldsymbol{f}_L - \boldsymbol{x}_L\|},
\end{aligned}
$$

and the statement follows by setting $C_{\boldsymbol{f}} := \|\partial_{\boldsymbol{X}}\boldsymbol{f}_S(\boldsymbol{0}, \mu\boldsymbol{f}_L)\| \cdot \sqrt{\|\mu\boldsymbol{f}_L\|}$.

Hence, in the following we can assume that S is a nontrivial SCC. Set $\boldsymbol{g}(\boldsymbol{X}_S) := \boldsymbol{f}_S(\boldsymbol{X}_S, \mu\boldsymbol{f}_L)$. Notice that $\boldsymbol{g}$ is an scSPP with $\mu\boldsymbol{g} = \mu\boldsymbol{f}_S$. By applying Lemma A.3 to $\boldsymbol{g}$ and setting $c := 1/\sqrt{C_{\boldsymbol{g}}}$ (the $C_{\boldsymbol{g}}$ from Lemma A.3), we get

$$
\begin{aligned}
\|\mu\boldsymbol{f}_S - \widetilde{\boldsymbol{\mu}}_S\| &\leq c \cdot \sqrt{\|\boldsymbol{g}(\mu\boldsymbol{g} - (\mu\boldsymbol{f}_S - \widetilde{\boldsymbol{\mu}}_S)) - (\mu\boldsymbol{g} - (\mu\boldsymbol{f}_S - \widetilde{\boldsymbol{\mu}}_S))\|} \\
&= c \cdot \sqrt{\|\boldsymbol{f}_S(\widetilde{\boldsymbol{\mu}}_S, \mu\boldsymbol{f}_L) - \widetilde{\boldsymbol{\mu}}_S\|} \\
&= c \cdot \sqrt{\|\boldsymbol{f}_S(\widetilde{\boldsymbol{\mu}}_S, \mu\boldsymbol{f}_L) - \boldsymbol{f}_S(\widetilde{\boldsymbol{\mu}}_S, \boldsymbol{x}_L)\|},
\end{aligned}
$$

and with Taylor's theorem (cf. Lemma 2.3) we obtain

$$
\begin{aligned}
&\leq c \cdot \sqrt{\|\partial_{\boldsymbol{X}_L}\boldsymbol{f}_S(\widetilde{\boldsymbol{\mu}}_S, \mu\boldsymbol{f}_L)(\mu\boldsymbol{f}_L - \boldsymbol{x}_L)\|} \\
&\leq c \cdot \sqrt{\|\partial_{\boldsymbol{X}_L}\boldsymbol{f}_S(\mu\boldsymbol{f}_S, \mu\boldsymbol{f}_L)(\mu\boldsymbol{f}_L - \boldsymbol{x}_L)\|} \\
&\leq c \cdot \sqrt{\|\partial_{\boldsymbol{X}_L}\boldsymbol{f}_S(\mu\boldsymbol{f}_S, \mu\boldsymbol{f}_L)\|} \cdot \sqrt{\|\mu\boldsymbol{f}_L - \boldsymbol{x}_L\|}.
\end{aligned}
$$

So the statement follows by setting $C_{\boldsymbol{f}} := c \cdot \sqrt{\|\partial_{\boldsymbol{X}_L}\boldsymbol{f}_S(\mu\boldsymbol{f}_S, \mu\boldsymbol{f}_L)\|}$.   □

Now we can extend Lemma A.4 to Lemma 6.3, restated here.

LEMMA 6.3.   *There is a constant $C_{\boldsymbol{f}} > 0$ such that*

$$
\left\|\mu\boldsymbol{f}_{[t]} - \widetilde{\boldsymbol{\mu}}_{[t]}\right\| \leq C_{\boldsymbol{f}} \cdot \sqrt{\left\|\mu\boldsymbol{f}_{[>t]} - \boldsymbol{\rho}_{[>t]}\right\|}
$$

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

Proof. Observe that $\mu\boldsymbol{f}_{[t]}, \widetilde{\boldsymbol{\mu}}_{[t]}, \mu\boldsymbol{f}_{[>t]}$, and $\boldsymbol{\rho}_{[>t]}$ do not depend on the components of depth $< t$. So we can assume w.l.o.g. that $t = 0$. Let $\mathcal{SCC}(0) = \{S_1, \ldots, S_k\}$.

For any S_i from $\mathcal{SCC}(0)$, let $\boldsymbol{f}^{(i)}$ be obtained from $\boldsymbol{f}$ by removing all top SCCs except for S_i. Lemma A.3 applied to $\boldsymbol{f}^{(i)}$ guarantees a $C^{(i)}$ such that

$$
\left\|\mu\boldsymbol{f}_{S_i} - \widetilde{\boldsymbol{\mu}}_{S_i}\right\| \leq C^{(i)} \cdot \sqrt{\left\|\mu\boldsymbol{f}_{[>0]} - \boldsymbol{\rho}_{[>0]}\right\|}
$$

holds for all $\boldsymbol{\rho}_{[>0]}$ with $\boldsymbol{0} \leq \boldsymbol{\rho}_{[>0]} \leq \mu\boldsymbol{f}_{[>0]}$. Using the equivalence of norms, w.l.o.g. let the norm $\|\cdot\|$ be the maximum-norm $\|\cdot\|_\infty$. Let $C_{\boldsymbol{f}} := \max_{1 \leq i \leq k} C^{(i)}$. Then we have

$$
\left\|\mu\boldsymbol{f}_{[0]} - \widetilde{\boldsymbol{\mu}}_{[0]}\right\| = \max_{1 \leq i \leq k}\left\|\mu\boldsymbol{f}_{S_i} - \widetilde{\boldsymbol{\mu}}_{S_i}\right\| \leq C_{\boldsymbol{f}} \cdot \sqrt{\left\|\mu\boldsymbol{f}_{[>0]} - \boldsymbol{\rho}_{[>0]}\right\|}
$$

for all $\boldsymbol{\rho}_{[>0]}$ with $\boldsymbol{0} \leq \boldsymbol{\rho}_{[>0]} \leq \mu\boldsymbol{f}_{[>0]}$.   □

Appendix B. Proofs of section 8.

B.1. Proof of Lemma 8.2.

LEMMA 8.2. *For every quadric q_i induced by a clean and feasible scSPP $\boldsymbol{f}$ we have*

$$q_i'(\boldsymbol{x}) = (\partial_{X_1} q_i(\boldsymbol{x}), \partial_{X_2} q_i(\boldsymbol{x}), \ldots, \partial_{X_n} q_i(\boldsymbol{x})) \neq \boldsymbol{0} \quad and \quad \partial_{X_i} q_i(\boldsymbol{x}) < 0 \text{ for all } \boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f}).$$

Proof. As shown by Etessami and Yannakakis in [16] under the above preconditions it holds for all $\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f})$ that $\big(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x})\big)$ is invertible with

$$\big(\mathrm{Id} - \boldsymbol{f}'(\boldsymbol{x})\big)^{-1} = \boldsymbol{f}'(\boldsymbol{x})^*.$$

Thus, we have

$$\boldsymbol{q}'(\boldsymbol{x})^{-1} = \big(\boldsymbol{f}'(\boldsymbol{x}) - \mathrm{Id}\big)^{-1} = -\big(\boldsymbol{f}'(\boldsymbol{x})^*\big),$$

implying that $q_i'(\boldsymbol{x}) \neq \boldsymbol{0}$ for all $\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f})$, as $\boldsymbol{q}'(\boldsymbol{x})$ has to have full rank n in order for $\boldsymbol{q}'(\boldsymbol{x})^{-1}$ to exist. Furthermore, it follows that all entries of $\boldsymbol{q}'(\boldsymbol{x})^{-1}$ are nonpositive, as $\boldsymbol{f}'(\boldsymbol{x})^*$ is nonnegative. Now, as $q_i(\boldsymbol{X}) = f_i(\boldsymbol{X}) - X_i$ and $f_i(\boldsymbol{X})$ is a polynomial with nonnegative coefficients, it holds that

$$q_i'(\boldsymbol{x}) \cdot \boldsymbol{e}_j = \partial_{X_j} q_i(\boldsymbol{x}) = \partial_{X_j} f_i(\boldsymbol{x}) \geq 0$$

for all $j \neq i$ and $\boldsymbol{x} \geq \boldsymbol{0}$. With every entry of $\boldsymbol{q}'(\boldsymbol{x})^{-1}$ nonpositive, and

$$q_i'(\boldsymbol{x}) \cdot \boldsymbol{q}'(\boldsymbol{x})^{-1} = \boldsymbol{e}_i^\top,$$

we conclude that $\partial_{X_i} q_i(\boldsymbol{x}) < 0$. $\square$

B.2. Proof of Lemma 8.4.
We first summarize some properties of the functions h_i.

PROPOSITION B.1. *Let $\boldsymbol{f}$ be a clean and feasible scSPP. Let $\boldsymbol{x}, \boldsymbol{y} \in [\boldsymbol{0}, \mu\boldsymbol{f}]$ with $\boldsymbol{x} \leq \boldsymbol{y}$.*

(a) $0 \leq h_i^{(k)}(\boldsymbol{x}_{-i}) \leq \mu\boldsymbol{f}_i$.
(b) $h_i^{(k)}(\boldsymbol{x}_{-i}) \leq h_i^{(k+1)}(\boldsymbol{x}_{-i})$ *for all $k \in \mathbb{N}$.*
(c) $h_i^{(k)}(\boldsymbol{x}_{-i}) \leq h_i^{(k)}(\boldsymbol{y}_{-i})$ *for all $k \in \mathbb{N}$.*
(d) $h_i(\boldsymbol{x}_{-i}) \leq \mu\boldsymbol{f}_i$, *and h_i is a map from $[\boldsymbol{0}, \mu\boldsymbol{f}_{-i}]$ to $[0, \mu\boldsymbol{f}_i]$. If f_i depends on at least one other variable except X_i, we also have $h_i([\boldsymbol{0}, \mu\boldsymbol{f}_{-i})) \subseteq [\boldsymbol{0}, \mu\boldsymbol{f}_i)$.*
(e) $h_i(\boldsymbol{x}_{-i}) \leq h_i(\boldsymbol{y}_{-i})$.
(f) $f_i(\boldsymbol{x}_{-i}, h_i(\boldsymbol{x}_{-i})) = h_i(\boldsymbol{x}_{-i})$.
(g) *For $x_i = f_i(\boldsymbol{x})$ we have $h_i(\boldsymbol{x}_{-i}) \leq x_i$.*
(h) $h_i(\mu\boldsymbol{f}_{-i}) = \mu\boldsymbol{f}_i$.

Proof. Let $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{y} \leq \mu\boldsymbol{f}$. Using the monotonicity of f_i over $\mathbb{R}_{\geq 0}^n$, we proceed by induction on k.

(a) For $k = 0$ we have

$$0 \leq h_i^{(0)}(\boldsymbol{x}_{-i}) = f_i(0, \boldsymbol{x}_{-i}) \leq f_i(\mu\boldsymbol{f}) = \mu\boldsymbol{f}_i.$$

We then get

$$0 \leq h_i^{(k+1)}(\boldsymbol{x}_{-i}) = f_i(h_i^{(k)}(\boldsymbol{x}_{-i}), \boldsymbol{x}_{-i}) \leq f_i(\mu\boldsymbol{f}) = \mu\boldsymbol{f}_i.$$

(b) For $k = 0$ we have

$$h_i^{(0)}(\boldsymbol{x}_{-i}) = f_i(0, \boldsymbol{x}_{-i}) \leq f_i(h_i^{(0)}(\boldsymbol{x}_{-i}), \boldsymbol{x}_{-i}) = h_i^{(1)}(\boldsymbol{x}_{-i}).$$

Thus

$$h_i^{(k+1)}(\boldsymbol{x}_{-i}) = f_i(h_i^{(k)}(\boldsymbol{x}_{-i}), \boldsymbol{x}_{-i}) \leq f_i(h_i^{(k+1)}(\boldsymbol{x}_{-i}), \boldsymbol{x}_{-i}) = h_i^{(k+2)}(\boldsymbol{x}_{-i})$$

follows.
(c) As $\boldsymbol{x} \leq \boldsymbol{y}$, we have for $k = 0$

$$h_i^{(0)}(\boldsymbol{x}_{-i}) = f_i(0, \boldsymbol{x}_{-i}) \leq f_i(0, \boldsymbol{y}_{-i}) = h_i^{(0)}(\boldsymbol{y}_{-i}).$$

Hence, we get

$$h_i^{(k+1)}(\boldsymbol{x}_{-i}) = f_i(h_i^{(k)}(\boldsymbol{x}_{-i}), \boldsymbol{x}_{-i}) \leq f_i(h_i^{(k)}(\boldsymbol{y}_{-i}), \boldsymbol{y}_{-i}) = h_i^{(k+1)}(\boldsymbol{y}_{-i}).$$

(d) As the sequence $(h_i^{(k)}(\boldsymbol{x}_{-i}))_{k \in \mathbb{N}}$ is monotonically increasing and bounded from above by $\mu \boldsymbol{f}_i$, the sequence converges. Thus, for every $\boldsymbol{x}$ the value

$$h_i(\boldsymbol{x}_{-i}) = \lim_{k \to \infty} h_i^{(k)}(\boldsymbol{x}_{-i})$$

is well-defined; i.e., h_i is a map from $[\boldsymbol{0}, \mu \boldsymbol{f}_{-i}]$ to $[0, \mu f_i]$.
If f_i depends on at least one other variable except X_i, then h_i is a nonconstant power series in this variable with nonnegative coefficients. For $\boldsymbol{x}_{-i} \in [\boldsymbol{0}, \mu \boldsymbol{f}_{-i})$ we thus always have

$$h_i(\boldsymbol{x}_{-i}) < h_i(\mu \boldsymbol{f}_{-i}) = \mu \boldsymbol{f}_i$$

as $\boldsymbol{x}_{-i} \prec \mu \boldsymbol{f}_{-i}$.
(e) This follows immediately from (b).
(f) As f_i is continuous, we have

$$f_i(h_i(\boldsymbol{x}_{-i}), \boldsymbol{x}_{-i}) = f_i \left(\lim_{k \to \infty} h_i^{(k)}(\boldsymbol{x}_{-i}), \boldsymbol{x}_{-i} \right) = \lim_{k \to \infty} h_i^{(k+1)}(\boldsymbol{x}_{-i}) = h_i(\boldsymbol{x}_{-i}),$$

where the last equality holds because of (b).
(g) Using induction similar to (a), replacing $\mu \boldsymbol{f}$ by $\boldsymbol{x}$, one gets $h_i^{(k)}(\boldsymbol{x}_{-i}) \leq x_i$ for all $k \in \mathbb{N}$ as $f_i(\boldsymbol{x}_{-i}) = x_i$. Thus, $h_i(\boldsymbol{x}_{-i}) \leq x_i$ follows similarly to (d).
(h) By definition, we have $\mu \boldsymbol{f} = \lim_{k \to \infty} \boldsymbol{f}^k(\boldsymbol{0})$. For $k = 0$, we have

$$(\boldsymbol{f}^0(\boldsymbol{0}))_i = 0 \leq f_i(0, \mu \boldsymbol{f}_{-i}) = h_i^{(0)}(\mu \boldsymbol{f}_{-i}).$$

We thus get by induction

$$(\boldsymbol{f}^{(k+1)}(\boldsymbol{0}))_i = f_i(\boldsymbol{f}^k(\boldsymbol{0})) \leq f_i(h_i^{(k)}(\mu \boldsymbol{f}_{-i}), \mu \boldsymbol{f}_{-i}) = h_i^{(k+1)}(\mu \boldsymbol{f}_{-i}).$$

Thus, we may conclude that $\mu \boldsymbol{f}_i \leq h_i(\mu \boldsymbol{f}_{-i})$. As $\mu \boldsymbol{f}_i = f_i(\mu \boldsymbol{f})$, we get by virtue of (g) that $h_i(\mu \boldsymbol{f}_{-i}) \leq \mu \boldsymbol{f}_i$, too. $\square$

With Proposition B.1 in hand, we now can show Lemma 8.4.

LEMMA 8.4. *h_i is continuously differentiable with*

$$\partial_{X_j} h_i(\boldsymbol{x}_{-i}) = \frac{\partial_{X_j} f_i(\boldsymbol{x})}{-\partial_{X_i} q_i(\boldsymbol{x})} = \frac{\partial_{X_j} q_i(\boldsymbol{x})}{-\partial_{X_i} q_i(\boldsymbol{x})} \ \text{for } \boldsymbol{x} \in S_i \text{ and } j \neq i.$$

In particular, $\partial_{X_j} h_i$ is monotonically increasing with $\boldsymbol{x}$.

Proof. By Lemma 8.2 the implicit function theorem is applicable for every $\boldsymbol{x} \in S_i$. We therefore find for every $\boldsymbol{x} \in S_i$ a local parametrization $h_{\boldsymbol{x}} : U \mapsto V$ with $h_{\boldsymbol{x}}(\boldsymbol{x}_{-i}) = x_i$. Thus $h_{\boldsymbol{x}}(\boldsymbol{x}_{-i})$ is the least nonnegative solution of $q_i(X_i, \boldsymbol{x}_{-i}) = 0$. By continuity of q_i it is now easily shown that for all $\boldsymbol{y}_{-i} \in U$ it has to hold that $h_{\boldsymbol{x}}(\boldsymbol{y}_{-i})$ is also the least nonnegative solution of $q_i(X_i, \boldsymbol{y}_{-i}) = 0$ (see below). By uniqueness we therefore have $h_{\boldsymbol{x}} = h_i$ and that h_i is continuously differentiable for all $\boldsymbol{x}_{-i} \in [\boldsymbol{0}, \mu \boldsymbol{f}_{-i})$.

For every $\boldsymbol{x}_{-i} \in [\boldsymbol{0}, \mu \boldsymbol{f}_{-i})$ we can solve the (at most) quadratic equation $q_i(X_i, \boldsymbol{x}_{-i}) = 0$. We already know that $h_i(\boldsymbol{x}_{-i})$ is the least nonnegative solution of this equation. So, if there exists another solution, it has to be real, too.

Assume first that this equation has two distinct solutions for some fixed $\boldsymbol{x}_{-i} \in [\boldsymbol{0}, \mu \boldsymbol{f}_{-i})$. Solving $q_i(X_i, \boldsymbol{x}_{-i}) = 0$ thus leads to an expression of the form

$$\frac{-b(\boldsymbol{x}_{-i}) \pm \sqrt{b(\boldsymbol{x}_{-i})^2 - 4a \cdot c(\boldsymbol{x}_{-i})}}{2a}$$

for the solutions where b, c are (at most) quadratic polynomials in $\boldsymbol{X}_{-i}$, c having nonnegative coefficients, and a is a positive constant (leading coefficient of X_i^2 in $q_i(\boldsymbol{X})$). As b and c are continuous, the discriminant $b(\cdot)^2 - 4a \cdot c(\cdot)$ stays positive for some open ball around $\boldsymbol{x}_{-i}$ included inside of U (it is positive in $\boldsymbol{x}_{-i}$ as we assume that we have two distinct solutions). By making U smaller, we may assume that U is this open ball. One of the two solutions must then be the least nonnegative solution. As $h_{\boldsymbol{x}}$ is the least nonnegative solution for $\boldsymbol{x}_{-i}$, and $h_{\boldsymbol{x}}$ is continuous, this also has to hold for some open ball centered at $\boldsymbol{x}_{-i}$. W.l.o.g. U is this ball. So, $h_{\boldsymbol{x}}$ and h_i coincide on U.

We turn to the case that $q_i(X_i, \boldsymbol{x}_{-i}) = 0$ has only a single solution, i.e., $h_i(\boldsymbol{x}_{-i})$. Note that $q_i(\boldsymbol{X})$ is linear in X_i if and only if $q_i(X_i, \boldsymbol{x}_{-i})$ is linear in X_i. Obviously, if q_i linear in X_i, then h_i and $h_{\boldsymbol{x}}$ coincide on U. Thus, consider the case that $q_i(\boldsymbol{X})$ is quadratic in X_i, but $q_i(X_i, \boldsymbol{x}_{-i})$ has only a single solution. This means that $\boldsymbol{x}_{-i}$ is a root of the discriminant, i.e., $b(\boldsymbol{x}_{-i}) - 4ac(\boldsymbol{x}_{-i}) = 0$. As $h_i(\boldsymbol{y}_{-i})$ is a solution of $q_i(X_i, \boldsymbol{y}_{-i}) = 0$ for all $\boldsymbol{y}_{-i} \in U$, the discriminant is nonnegative on U. If it is equal to zero on U, then we again have that h_i is equal to $h_{\boldsymbol{x}}$ on U. Therefore assume that it is positive in some point of U. As the discriminant is continuous, the solutions change continuously with $\boldsymbol{x}_{-i}$. But this implies that for some $\boldsymbol{y}_{-i} \in U$ there are at least two $y_i, y_i^* \in V$ such that $(\boldsymbol{y}_{-i}, y_i)$ and $(\boldsymbol{y}_{-i}, y_i^*)$ are both located on the quadric $q_i(\boldsymbol{X}) = 0$. But this contradicts the uniqueness of $h_{\boldsymbol{x}}$ guaranteed by the implicit function theorem.

Assume now that $\boldsymbol{x} \in S_i$. We then have

$$q_i(\boldsymbol{x}) = q_i(\boldsymbol{x}_{-i}, h_i(\boldsymbol{x}_{-i})) = 0,$$

or equivalently

$$f_i(\boldsymbol{x}_{-i}, h_i(\boldsymbol{x}_{-i})) = h_i(\boldsymbol{x}_{-i}).$$

Calculating the gradient of both in $\boldsymbol{x}$ yields

$$f_i'(\boldsymbol{x}) \cdot \boldsymbol{p}_i'(\boldsymbol{x}_{-i}) = h_i'(\boldsymbol{x}_{-i}).$$

For the Jacobian of $\boldsymbol{p}_i$ we obtain

$$\boldsymbol{p}_i'(\boldsymbol{x}_{-i}) = \begin{pmatrix} \boldsymbol{e}_1^\top \\ \vdots \\ \boldsymbol{e}_{i-1}^\top \\ h_i'(\boldsymbol{x}_{-i}) \\ \boldsymbol{e}_{i+1}^\top \\ \vdots \\ \boldsymbol{e}_n^\top \end{pmatrix}.$$

This leads to

$$\partial_{X_j} f_i(\boldsymbol{x}) + \partial_{X_i} f_i(\boldsymbol{x}) \cdot \partial_{X_j} h_i(\boldsymbol{x}_{-i}) = \partial_{X_j} h_i(\boldsymbol{x}_{-i}),$$

which solved for $\partial_{X_j} h_i$ yields

$$\partial_{X_j} h_i(\boldsymbol{x}_{-i}) = \frac{\partial_{X_j} f_i(\boldsymbol{x})}{-\partial_{X_i} q_i(\boldsymbol{x})}.$$

As $\partial_{X_i} q_i(\boldsymbol{x}) < 0$ and both $\partial_{X_j} f_i$ and $\partial_{X_i} q_i$ monotonically increase with $\boldsymbol{x}$, it follows that $\partial_{X_j} h_i$ also monotonically increases with $\boldsymbol{x}$. Finally, for $j \neq i$ we have that $\partial_{X_j} q_i = \partial_{X_j} f_i$ as $q_i = f_i - X_i$. ☐

B.3. Proof of Lemma 8.8.
LEMMA 8.8. *For all $\boldsymbol{x} \in S_i$ we have*

$$\text{for all } \boldsymbol{y} \in S_i \cap [\boldsymbol{x}, \mu\boldsymbol{f}] : q_i'(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) \leq 0.$$

In particular

$$\text{for all } \boldsymbol{y} \in S_i \cap [\boldsymbol{x}, \mu\boldsymbol{f}] : y_i \geq x_i + \sum_{j \neq i} \partial_{X_j} h_i(\boldsymbol{x}_{-i}) \cdot (y_j - x_j).$$

Proof. Let $\boldsymbol{x} \in S_i$, i.e., $f_i(\boldsymbol{x}) = x_i$. We want to show that

$$q_i'(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) \leq 0$$

for all $\boldsymbol{y} \in S_i \cap [\boldsymbol{x}, \mu\boldsymbol{f}]$. As f_i is quadratic in $\boldsymbol{X}$, we may write

$$\begin{aligned}
0 &= q_i(\boldsymbol{y}) \\
&= -y_i + f_i(\boldsymbol{y}) \\
&= -y_i + \underbrace{f_i(\boldsymbol{x})}_{=x_i} + f_i'(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) + \underbrace{(\boldsymbol{y} - \boldsymbol{x})^\top \cdot A \cdot (\boldsymbol{y} - \boldsymbol{x})}_{\geq 0} \\
&\geq -y_i + x_i + f_i'(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) \\
&= f_i'(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}) - \boldsymbol{e}_i^\top \cdot (\boldsymbol{y} - \boldsymbol{x}) \\
&= q_i'(\boldsymbol{x}) \cdot (\boldsymbol{y} - \boldsymbol{x}),
\end{aligned}$$

where A is a symmetric square matrix with nonnegative components such that the quadric terms of f_i are given by $\boldsymbol{X}^\top A \boldsymbol{X}$.

The second claim is easily obtained by solving this inequality for y_i and recalling that by Lemma 8.4 we have $\partial_{X_j} h_i(\boldsymbol{x}_{-i}) = \frac{\partial_{X_j} q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i}))}{-\partial_{X_i} q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i}))}$ and $\partial_{X_i} q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i})) < 0$. ☐

B.4. Proof of Proposition 8.9.

PROPOSITION 8.9. *It holds that*

$$\boldsymbol{x} \in R \Leftrightarrow \boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f}) \wedge \boldsymbol{q}(\boldsymbol{x}) \geq \boldsymbol{0}.$$

Proof. Let $\boldsymbol{x} \in R$ and $i \in \{1, \ldots, n\}$. Consider the function

$$g(t) := q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i}) + t\boldsymbol{e}_i).$$

As q_i is a quadratic polynomial in $\boldsymbol{X}$, there exists a symmetric square matrix A with nonnegative entries, a vector $\boldsymbol{b}$, and a constant c such that

$$q_i(\boldsymbol{X}) = \boldsymbol{X}^\top A \boldsymbol{X} + \boldsymbol{b}^\top \boldsymbol{X} + c.$$

It then follows that

$$q_i(\boldsymbol{X} + \boldsymbol{Y}) = q_i(\boldsymbol{X}) + q_i'(\boldsymbol{X})\boldsymbol{Y} + \boldsymbol{Y}^\top A \boldsymbol{Y}.$$

With $q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i})) = 0$ this implies

$$g(t) = q_i'(\boldsymbol{p}_i(\boldsymbol{x}_{-i}))t\boldsymbol{e}_i + t^2 \underbrace{\boldsymbol{e}_i^\top A \boldsymbol{e}_i}_{:=a \geq 0} = t \cdot (\partial_{X_i} q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i})) + a \cdot t).$$

As $\boldsymbol{p}_i(\boldsymbol{x}_{-i}) \prec \mu\boldsymbol{f}$ ($\boldsymbol{f}$ is strongly connected and $\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f})$), we know that $\partial_{X_i} q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i})) < 0$. Thus, $g(t)$ has at most two zeros, one at 0, the other for some $t^* \geq 0$.

For the direction ($\Rightarrow$) we have to show only that $x_i \leq h_i(\boldsymbol{x}_{-i})$ implies that $q_i(\boldsymbol{x}) \geq 0$. This now easily follows as $x_i \leq h_i(\boldsymbol{x}_{-i})$ implies that there is a $t' \leq 0$ with $\boldsymbol{p}_i(\boldsymbol{x}_{-i}) + t'\boldsymbol{e}_i = \boldsymbol{x}$. But for this $t' \leq 0$ we have $q_i(\boldsymbol{x}) = g(t') \geq 0$.

Consider therefore the other direction ($\Leftarrow$), that is, $\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f})$ with $\boldsymbol{q}(\boldsymbol{x}) \geq \boldsymbol{0}$. Assume that $\boldsymbol{x} \notin R$; i.e., for at least one i we have $x_i > h_i(\boldsymbol{x}_{-i})$. As $q_i(\boldsymbol{x}) \geq 0$ there has to be a $t'' > 0$ with $\boldsymbol{p}_i(\boldsymbol{x}_{-i}) + t''\boldsymbol{e}_i = \boldsymbol{x}$ and $g(t'') \geq 0$. This implies that $a > 0$ has to hold, as otherwise $g(t)$ would be linear in t and negative for $t > 0$. But then the second root t^* of $g(t)$ has to be positive. Set $\boldsymbol{x}^* = \boldsymbol{p}_i(\boldsymbol{x}_{-i}) + t^*\boldsymbol{e}_i$ with $q_i(\boldsymbol{x}^*) = 0$, too.

A calculation similar to the one above leads to

$$g(t + t^*) = q_i(\boldsymbol{x}^* + t\boldsymbol{e}_i) = t \cdot (\partial_{X_i} q_i(\boldsymbol{x}^*) + a \cdot t).$$

It follows that $\partial_{X_i} q_i(\boldsymbol{x}^*)$ has to be greater than zero for $-t^*$ to be a root (as $a > 0$). But we have shown that $\partial_{X_i} q_i(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f})$. $\square$

B.5. Proof of Lemma 8.11.

LEMMA 8.11. *Let $\boldsymbol{f}$ be a clean and feasible scSPP. Let $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)} \in [\boldsymbol{0}, \mu\boldsymbol{f})$. Then the matrix*

$$\begin{pmatrix} q_1'(\boldsymbol{x}^{(1)}) \\ \vdots \\ q_n'(\boldsymbol{x}^{(n)}) \end{pmatrix}$$

is regular; i.e., the vectors $\{q_i'(\boldsymbol{x}^{(i)}) | i = 1, \ldots, n\}$ are linearly independent.

Proof. Define $\boldsymbol{x} \in [\boldsymbol{0}, \mu\boldsymbol{f})$ by setting

$$x_i := \max\{x_i^{(j)} \mid j = 1, \ldots, n\}.$$

We then have $\boldsymbol{x}^{(i)} \leq \boldsymbol{x}$ for all i, and $\boldsymbol{x} \prec \mu\boldsymbol{f}$. As mentioned above, we therefore have that $\boldsymbol{q}'(\boldsymbol{x})$ is regular with

$$\boldsymbol{q}'(\boldsymbol{x})^{-1} = -\sum_{k \in \mathbb{N}} \boldsymbol{f}'(\boldsymbol{x})^k.$$

As $\boldsymbol{x}^{(i)} \leq \boldsymbol{x}$ it follows that

$$\begin{pmatrix} f_1'(\boldsymbol{x}^{(1)}) \\ \vdots \\ f_n'(\boldsymbol{x}^{(n)}) \end{pmatrix} \leq \boldsymbol{f}'(\boldsymbol{x}).$$

Hence, we also have

$$\sum_{k=0}^{l} \begin{pmatrix} f_1'(\boldsymbol{x}^{(1)}) \\ \vdots \\ f_n'(\boldsymbol{x}^{(n)}) \end{pmatrix}^l \leq \sum_{k=0}^{l} \boldsymbol{f}'(\boldsymbol{x}),$$

implying that

$$\begin{pmatrix} f_1'(\boldsymbol{x}^{(1)}) \\ \vdots \\ f_n'(\boldsymbol{x}^{(n)}) \end{pmatrix}^* \quad \text{and, thus,} \quad \begin{pmatrix} q_1'(\boldsymbol{x}^{(1)}) \\ \vdots \\ q_n'(\boldsymbol{x}^{(n)}) \end{pmatrix}^{-1} \quad \text{exist.}$$

So, the vectors $\{q_1'(\boldsymbol{x}^{(1)}), \ldots, q_n'(\boldsymbol{x}^{(n)})\}$ have to be linearly independent. $\quad\square$

B.6. Proof of Theorem 8.13.

THEOREM 8.13. *Let $\boldsymbol{f}$ be a clean and feasible scSPP. Let $\boldsymbol{x} \in R$. For $i = 1, \ldots, n$ fix some $\eta_i \in [x_i, h_i(\boldsymbol{x}_{-i})]$, and set $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)$. We then have*

$$\boldsymbol{x} \leq \mathcal{N}(\boldsymbol{x}) \leq \mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x}) \leq \mathcal{T}(\boldsymbol{x}) \leq \mu\boldsymbol{f}.$$

Further, the operator $\mathcal{T}$ is monotone on R; i.e., for any $\boldsymbol{y} \in R$ with $\boldsymbol{x} \leq \boldsymbol{y}$ it holds that $\mathcal{T}(\boldsymbol{x}) \leq \mathcal{T}(\boldsymbol{y})$.

Proof. Set

$$\boldsymbol{\pi}_i := (\boldsymbol{x}_{-i}, \eta_i) \quad \text{and} \quad \boldsymbol{h} := (h_1(\boldsymbol{x}_{-1}), \ldots, h_n(\boldsymbol{x}_{-n})).$$

We first show that $\boldsymbol{x} \leq \mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x})$:

$$\begin{aligned}
\mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x}) &= \left(q_i'(\boldsymbol{\pi}_i)\right)_{i=1,\ldots,n}^{-1} \cdot \left(q_i'(\boldsymbol{\pi}_i) \cdot \boldsymbol{\pi}_i - q_i(\boldsymbol{\pi}_i)\right)_{i=1,\ldots,n} \\
&= \left(f_i'(\boldsymbol{\pi}_i)\right)_{i=1,\ldots,n}^* \cdot \left(-q_i'(\boldsymbol{\pi}_i) \cdot \boldsymbol{\pi}_i + q_i(\boldsymbol{\pi}_i)\right)_{i=1,\ldots,n} \\
&= \left(f_i'(\boldsymbol{\pi}_i)\right)_{i=1,\ldots,n}^* \cdot \left(-q_i'(\boldsymbol{\pi}_i) \cdot (\boldsymbol{x} + (\eta_i - x_i) \cdot \boldsymbol{e}_i) + q_i(\boldsymbol{\pi}_i)\right)_{i=1,\ldots,n} \\
&= \underbrace{\left(f_i'(\boldsymbol{\pi}_i)\right)_{i=1,\ldots,n}^*}_{\geq 0 \text{ in every comp.}} \cdot \left(-q_i'(\boldsymbol{\pi}_i) \cdot \boldsymbol{x} - \underbrace{\partial_{X_i} q_i(\boldsymbol{\pi}_i)}_{<0} \cdot \underbrace{(\eta_i - x_i)}_{\geq 0} + \underbrace{q_i(\boldsymbol{\pi}_i)}_{\geq 0}\right)_{i=1,\ldots,n} \\
&\geq \boldsymbol{x}.
\end{aligned}$$

$\mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x})$ is by definition the (unique) solution of the equation system defined by

$$q_i'(\boldsymbol{\pi}_i)(\boldsymbol{X} - \boldsymbol{\pi}_i) = -q_i(\boldsymbol{\pi}_i) \quad (i = 1, \ldots, n).$$

As $\mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x}) \geq \boldsymbol{x}$, we can also consider this system with the origin of the coordinate system moved into $\boldsymbol{x}$, i.e.,

$$q_i'(\boldsymbol{\pi}_i)(\boldsymbol{X} + \boldsymbol{x} - \boldsymbol{\pi}_i) = -q_i(\boldsymbol{\pi}_i) \quad (i = 1, \ldots, n).$$

We show that this system is equivalent to an SPP. For this, we solve these equations for X_i:

$$q_i'(\boldsymbol{\pi}_i)(\boldsymbol{X} + \boldsymbol{x} - \boldsymbol{\pi}_i) = -q_i(\boldsymbol{\pi}_i)$$
$$\Leftrightarrow \quad q_i'(\boldsymbol{\pi}_i)\boldsymbol{X} = -q_i(\boldsymbol{\pi}_i) + q_i'(\boldsymbol{\pi}_i) \underbrace{(\boldsymbol{\pi}_i - \boldsymbol{x})}_{=(\eta_i - x_i)\cdot \boldsymbol{e}_i}$$
$$\Leftrightarrow \quad X_i = \sum_{j \neq i} \frac{\partial_{X_j} q_i(\boldsymbol{\pi}_i)}{-\partial_{X_i} q_i(\boldsymbol{\pi}_i)} \cdot X_j + \frac{q_i(\boldsymbol{\pi}_i)}{-\partial_{X_i} q_i(\boldsymbol{\pi}_i)} + (\eta_i - x_i).$$

Again, we have $\partial_{X_i} q_i(\boldsymbol{\pi}_i) < 0 \leq \partial_{X_j} q_i(\boldsymbol{\pi}_i)$ as $\boldsymbol{\pi}_i \in R$, and $q_i'(\boldsymbol{\pi}_i)$ monotonically increases with η_i. Hence, the above linear equation for X_i is indeed a polynomial with nonnegative coefficients. Denote by $\boldsymbol{f}_{\boldsymbol{\eta}}$ the SPP defined by these linear equations. We then have $\mu \boldsymbol{f}_{\boldsymbol{\eta}} = \mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x}) - \boldsymbol{x}$, as the above equation system has $\mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x}) - \boldsymbol{x} \geq \boldsymbol{0}$ as its unique solution. Further, we know that the Kleene sequence $\left(\boldsymbol{f}_{\boldsymbol{\eta}}^k(\boldsymbol{0})\right)_{k \in \mathbb{N}}$ converges to $\mu \boldsymbol{f}_{\boldsymbol{\eta}}$. We show that all coefficients of $\boldsymbol{f}_{\boldsymbol{\eta}}$ increase with $\boldsymbol{\eta} \to \boldsymbol{h}$. This is straightforward for

$$\frac{\partial_{X_j} q_i(\boldsymbol{\pi}_i)}{-\partial_{X_i} q_i(\boldsymbol{\pi}_i)},$$

as $\partial_{X_i} q_i(\boldsymbol{\pi}_i) < 0 \leq \partial_{X_j} q_i(\boldsymbol{\pi}_i)$, and all of these terms increase with $\eta_i \to h_i(\boldsymbol{x}_{-i})$. Consider therefore

$$0 \geq \frac{q_i(\boldsymbol{\pi}_i)}{-\partial_{X_i} q_i(\boldsymbol{\pi}_i)} + (\eta_i - x_i) = \frac{q_i(\boldsymbol{\pi}_i) - \partial_{X_i} q_i(\boldsymbol{\pi}_i)(\eta_i - x_i)}{-\partial_{X_i} q_i(\boldsymbol{\pi}_i)}.$$

We show that this term increases with η_i. Set $\delta_i := \eta_i - x_i$. We can find a nonnegative, symmetric square matrix A, a vector $\boldsymbol{b}$, and constant c such that

$$q_i(\boldsymbol{X}) = \boldsymbol{X}^\top A \boldsymbol{X} + \boldsymbol{b}^\top \boldsymbol{X} + c \quad \text{and} \quad q_i'(\boldsymbol{X}) = 2\boldsymbol{X}^\top A + \boldsymbol{b}^\top.$$

As $\boldsymbol{\pi}_i = \boldsymbol{x} + \delta_i \boldsymbol{e}_i$, we have

$$q_i(\boldsymbol{\pi}_i) = q_i(\boldsymbol{x} + \delta_i \boldsymbol{e}_i) = q_i(\boldsymbol{x}) + \partial_{X_i} q_i(\boldsymbol{x})\delta_i + \delta_i^2 A_{ii}$$

and

$$\partial_{X_i} q_i(\boldsymbol{\pi}_i) \cdot \delta_i = q_i'(\boldsymbol{x} + \delta_i \boldsymbol{e}_i)\delta_i \boldsymbol{e}_i = \partial_{X_i} q_i(\boldsymbol{x})\delta_i + 2\delta_i^2 A_{ii}.$$

This leads to

$$\frac{q_i(\boldsymbol{\pi}_i) - \partial_{X_i} q_i(\boldsymbol{\pi}_i)\delta_i}{-\partial_{X_i} q_i(\boldsymbol{\pi}_i)} = \frac{q_i(\boldsymbol{x}) - \delta_i^2 A_{ii}}{-\partial_{X_i} q_i(\boldsymbol{x}) - 2\delta_i A_{ii}}.$$

Taking the derivative w.r.t. δ_i yields

$$\frac{-2A_{ii}\delta_i}{-\partial_{X_i} q_i(\boldsymbol{x}) + 2A_{ii}\delta_i} - \frac{q_i(\boldsymbol{x}) - A_{ii}\delta_i^2}{(-\partial_{X_i} q_i(\boldsymbol{x}) - 2A_{ii}\delta_i)^2}(-2A_{ii})$$
$$= \frac{2A_{ii}\partial_{X_i} q_i(\boldsymbol{x})\delta_i + 4A_{ii}^2\delta_i^2 + 2A_{ii}q_i(\boldsymbol{x}) - 2A_{ii}^2\delta_i^2}{(-\partial_{X_i} q_i(\boldsymbol{x}) - 2A_{ii}\delta_i)^2}$$
$$= 2A_{ii} \frac{A_{ii}\delta_i^2 + \partial_{X_i} q_i(\boldsymbol{x})\delta_i + q_i(\boldsymbol{x})}{(-\partial_{X_i} q_i(\boldsymbol{x}) - 2A_{ii}\delta_i)^2}$$
$$= 2A_{ii} \frac{q_i(\boldsymbol{\pi}_i)}{(-\partial_{X_i} q_i(\boldsymbol{\pi}_i))^2}.$$

As $q_i(\boldsymbol{\pi}_i) \geq 0$ and $A_{ii} \geq 0$, it follows that

$$\frac{q_i(\boldsymbol{\pi}_i)}{-\partial_{X_i} q_i(\boldsymbol{\pi}_i)} + (\eta_i - x_i)$$

increases with $\eta_i \to h_i(\boldsymbol{x}_{-i})$. Thus, all coefficients of $\boldsymbol{f}_{\boldsymbol{\eta}}$ increase with $\eta_i \to h_i(\boldsymbol{x}_{-i})$, and so for any $\boldsymbol{\eta}' \in [\boldsymbol{\eta}, \boldsymbol{h}]$ it follows that

$$\boldsymbol{f}_{\boldsymbol{\eta}}(\boldsymbol{y}) \leq \boldsymbol{f}_{\boldsymbol{\eta}'}(\boldsymbol{y}) \text{ for all } \boldsymbol{y} \geq \boldsymbol{0}$$

and

$$\mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x}) - \boldsymbol{x} = \mu \boldsymbol{f}_{\boldsymbol{\eta}} \leq \mu \boldsymbol{f}_{\boldsymbol{\eta}'} = \mathcal{T}_{\boldsymbol{\eta}'}(\boldsymbol{x}) - \boldsymbol{x}.$$

As $\mathcal{N}(\boldsymbol{X}) = \mathcal{T}_{\boldsymbol{x}}(\boldsymbol{X})$ and $\mathcal{T}(\boldsymbol{X}) = \mathcal{T}_{\boldsymbol{h}}(\boldsymbol{X})$, we may therefore conclude that

$$\mathcal{N}(\boldsymbol{x}) \leq \mathcal{T}_{\boldsymbol{\eta}}(\boldsymbol{x}) \leq \mathcal{T}_{\boldsymbol{\eta}'}(\boldsymbol{x}) \leq \mathcal{T}(\boldsymbol{x}).$$

It remains to show that $\mathcal{T}(\boldsymbol{x}) \leq \mu \boldsymbol{f}$. This is equivalent to showing that $\mu \boldsymbol{f}_{\boldsymbol{h}} \leq \mu \boldsymbol{f} - \boldsymbol{x}$. For $\boldsymbol{f}_{\boldsymbol{h}}(\boldsymbol{X})$ we have by definition and Lemma 8.4

$$\left(\boldsymbol{f}_{\boldsymbol{h}}(\boldsymbol{X})\right)_i = \sum_{j \neq i} \frac{\partial_{X_j} q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i}))}{-\partial_{X_i} q_i(\boldsymbol{p}_i(\boldsymbol{x}_{-i}))} X_j + (h_i(\boldsymbol{x}_{-i}) - x_i) = \sum_{j \neq i} \partial_{X_j} h_i(\boldsymbol{x}_{-i}) X_j + (h_i(\boldsymbol{x}_{-i}) - x_i).$$

By virtue of Lemma 8.8 it follows that $\mu \boldsymbol{f}$ is above all the tangents, i.e.,

$$\boldsymbol{f}_{\boldsymbol{h}}(\mu \boldsymbol{f} - \boldsymbol{x}) \leq \mu \boldsymbol{f} - \boldsymbol{x}.$$

By monotonicity of $\boldsymbol{f}_{\boldsymbol{h}}$ we also have

$$\boldsymbol{f}_{\boldsymbol{h}}(\boldsymbol{0}) \leq \boldsymbol{f}_{\boldsymbol{h}}(\mu \boldsymbol{f} - \boldsymbol{x}).$$

A straightforward induction therefore shows that

$$\boldsymbol{f}_{\boldsymbol{h}}^k(\boldsymbol{0}) \leq \mu \boldsymbol{f} - \boldsymbol{x} \quad \text{(for all } k \in \mathbb{N}\text{)},$$

and, thus,

$$\mathcal{T}(\boldsymbol{x}) - \boldsymbol{x} = \mu \boldsymbol{f}_{\boldsymbol{h}} \leq \mu \boldsymbol{f} - \boldsymbol{x}.$$

We turn to the monotonicity of $\mathcal{T}$. Let $\boldsymbol{y} \in R$ with $\boldsymbol{x} \leq \boldsymbol{y}$. Assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are located on the surface S_i, i.e.,

$$h_i(\boldsymbol{x}_{-i}) = x_i \quad \text{and} \quad h_i(\boldsymbol{y}_{-i}) = y_i.$$

The tangent $T_i|_{\boldsymbol{x}}$ at S_i in $\boldsymbol{x}$ is spanned by the partial derivatives of $\boldsymbol{p}_i$ in $\boldsymbol{x}$. The part $T_i|_{\boldsymbol{x}} \cap [\boldsymbol{x}, \mu \boldsymbol{f}]$ relevant for $\mathcal{T}(\boldsymbol{x})$ can therefore be parameterized by

$$\boldsymbol{x} + \sum_{j \neq i} \partial_{X_j} \boldsymbol{p}_i(\boldsymbol{x}) \cdot (u_j - x_j) \quad \text{with} \quad \boldsymbol{u}_{-i} \in [\boldsymbol{x}_{-i}, \mu \boldsymbol{f}_{-i}].$$

Similarly for $T_i|_{\boldsymbol{y}}$.

In particular, for $\boldsymbol{u}_{-i} \in [\boldsymbol{y}_{-i}, \mu\boldsymbol{f}_{-i}]$ both points on the tangents defined by $\boldsymbol{u}_{-i}$ differ only in the ith coordinate (the remaining coordinates are simply $\boldsymbol{u}_{-i}$):

$$t_{\boldsymbol{y}} = y_i + \sum_{j\neq i} \partial_{X_j} h_i(\boldsymbol{y}) \cdot (u_j - y_j), \text{ respectively, } t_{\boldsymbol{x}} = x_i + \sum_{j\neq i} \partial_{X_j} h_i(\boldsymbol{x}) \cdot (u_j - x_j).$$

By Lemma 8.8 we have

$$y_i \geq x_i + \sum_{j\neq i} \partial_{X_j} h_i(\boldsymbol{x}) \cdot (y_j - x_j).$$

From Lemma 8.4 it follows that $\partial_{X_j} h_i(\boldsymbol{y}) \geq \partial_{X_j} h_i(\boldsymbol{x})$. Thus $t_{\boldsymbol{y}} \geq t_{\boldsymbol{x}}$ immediately follows.

Now for $\boldsymbol{x}, \boldsymbol{y} \in R$ with $\boldsymbol{x} \leq \boldsymbol{y}$ we can apply this result to the tangents at S_i in $\boldsymbol{p}_i(\boldsymbol{x}_{-i})$, respectively, $\boldsymbol{p}_i(\boldsymbol{y}_{-i})$, and $\mathcal{T}(\boldsymbol{x}) \leq \mathcal{T}(\boldsymbol{y})$ follows.   ☐

REFERENCES

[1] E. ALLENDER, P. BÜRGISSER, J. KJELDGAARD-PEDERSEN, AND P. B. MILTERSEN, *On the complexity of numerical analysis*, SIAM J. Comput., 38 (2009), pp. 1987–2006.

[2] K. ATHREYA AND P. NEY, *Branching Processes*, Springer-Verlag, New York, 1972.

[3] A. BERMAN AND R. PLEMMONS, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York, 1979.

[4] T. BRÁZDIL, A. KUČERA, AND O. STRAŽOVSKÝ, *On the decidability of temporal properties of probabilistic pushdown automata*, in Proceedings of the 22nd Symposium on Theoretical Aspects of Computer Science (STACS'2005), Lecture Notes in Comput. Sci. 3404, Springer-Verlag, Berlin, 2005, pp. 145–157.

[5] D. W. DECKER AND C. T. KELLEY, *Newton's method at singular points.* I, SIAM J. Numer. Anal., 17 (1980), pp. 66–70.

[6] R. DOWELL AND S. EDDY, *Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction*, BMC Bioinformatics, 5 (2004), article 71.

[7] R. DURBIN, S. EDDY, A. KROGH, AND G. MICHISON, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1998.

[8] J. ESPARZA, A. GAISER, AND S. KIEFER, *Computing least fixed points of probabilistic systems of polynomials*, in Proceedings of the 27th International Symposium on Theoretical Aspects of Computer Science (STACS'2010), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2010, pp. 359–370.

[9] J. ESPARZA, S. KIEFER, AND M. LUTTENBERGER, *Convergence thresholds of Newton's method for monotone polynomial equations*, in Proceedings of the 27th International Symposium on Theoretical Aspects of Computer Science (STACS'2008), 2008, pp. 289–300.

[10] J. ESPARZA, S. KIEFER, AND M. LUTTENBERGER, *Computing the Least Fixed Point of Positive Polynomial Systems*, Technical report, Technische Universität München, 2010; available online at http://arxiv.org/abs/1001.0340.

[11] J. ESPARZA, A. KUČERA, AND R. MAYR, *Model-checking probabilistic pushdown automata*, in Proceedings of the 19th IEEE Symposium on Logic in Computer Science (LICS 2004), IEEE Computer Society Press, Los Alamitos, CA, 2004, pp. 12–21.

[12] J. ESPARZA, A. KUČERA, AND R. MAYR, *Quantitative analysis of probabilistic pushdown automata: Expectations and variances*, in Proceedings of the 20th IEEE Symposium on Logic in Computer Science (LICS 2005), IEEE Computer Society Press, Los Alamitos, CA, 2005, pp. 117–126.

[13] K. ETESSAMI AND M. YANNAKAKIS, *Algorithmic verification of recursive probabilistic systems*, in Proceedings of the 11th International Conference on Tools and Algorithms for Construction and Analysis of Systems (TACAS 2005), Lecture Notes in Comput. Sci. 3440, Springer-Verlag, Berlin, 2005, pp. 253–270.

[14] K. ETESSAMI AND M. YANNAKAKIS, *Checking LTL properties of recursive Markov chains*, in Proceedings of the 2nd International Conference on Quantitative Evaluation of Systems (QEST'05), IEEE Computer Society Press, Washington, DC, 2005, pp. 155–165.

[15] K. ETESSAMI AND M. YANNAKAKIS, *Recursive Markov decision processes and recursive stochastic games*, in Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP 2005), Lecture Notes in Comput. Sci. 3580, Springer-Verlag, Berlin, 2005, pp. 891–903.

[16] K. ETESSAMI AND M. YANNAKAKIS, *Recursive Markov chains, stochastic grammars, and monotone systems of nonlinear equations*, J. ACM, 56 (2009), pp. 1–66.

[17] 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 Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC 2000), ACM, New York, 2000, pp. 484–493.

[18] R. FAGIN, A. KARLIN, J. KLEINBERG, P. RAGHAVAN, S. RAJAGOPALAN, R. RUBINFELD, M. SUDAN, AND A. TOMKINS, *Random walks with "back buttons"*, Ann. Appl. Probab., 11 (2001), pp. 810–862.

[19] S. GEMAN AND M. JOHNSON, *Probabilistic Grammars and Their Applications*, in International Encyclopedia of the Social & Behavioral Sciences, N. J. Smelser and P. B. Baltes, eds., Pergamon, Oxford, 2002, pp. 12075–12082.

[20] A. GRIEWANK AND M. R. OSBORNE, *Newton's method for singular problems when the dimension of the null space is > 1*, SIAM J. Numer. Anal., 18 (1981), pp. 145–149.

[21] T. HARRIS, *The Theory of Branching Processes*, Springer-Verlag, Berlin, 1963.

[22] R. HORN AND C. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.

[23] C. T. KELLEY, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.

[24] S. KIEFER, M. LUTTENBERGER, AND J. ESPARZA, *On the convergence of Newton's method for monotone systems of polynomial equations*, in Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC 2007), ACM, New York, 2007, pp. 217–226.

[25] B. KNUDSEN AND J. HEIN, *Pfold: RNA secondary structure prediction using stochastic context-free grammars*, Nucleic Acids Res., 31 (2003), pp. 3423–3428.

[26] W. KUICH, *Semirings and formal power series: Their relevance to formal languages and automata*, in Handbook of Formal Languages, Vol. 1, Springer-Verlag, Berlin, 1997, Chap. 9, pp. 609–677.

[27] P. LANCASTER AND M. TISMENETSKY, *The Theory of Matrices*, 2nd ed., Academic Press, New York, 1985.

[28] C. MANNING AND H. SCHÜTZE, *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA, 1999.

[29] J. ORTEGA, *Numerical Analysis: A Second Course*, Academic Press, New York, 1972.

[30] J. ORTEGA AND W. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[31] F. POTRA AND V. PTAK, *Sharp error bounds for Newton's process*, Numer. Math., 34 (1980), pp. 63–72.

[32] G. W. REDDIEN, *On Newton's method for singular problems*, SIAM J. Numer. Anal., 15 (1978), pp. 993–996.

[33] Y. SAKABIKARA, M. BROWN, R. HUGHEY, I. MIAN, K. SJOLANDER, R. UNDERWOOD, AND D. HAUSSLER, *Stochastic context-free grammars for tRNA*, Nucleic Acids Res., 22 (1994), pp. 5112–5120.

[34] E. SÜLI AND D. MAYERS, *An Introduction to Numerical Analysis*, Cambridge University Press, Cambridge, UK, 2003.

[35] H. WATSON AND F. GALTON, *On the probability of the extinction of families*, J. Anthropol. Inst. Great Britain and Ireland, 4 (1874), pp. 138–144.