Symbolic-Numeric Computation
D. Wang and L. Zhi, Eds.
Trends in Mathematics, 299–313

On the Extended Iterative Proportional Scaling Algorithm

Ming-Deh Huang and Qing Luo

Abstract. The *iterative proportional scaling algorithm* is generalized to find real positive solutions to polynomial systems of the form: $\sum_{j=1}^{m} a_{sj} p_j = c_s$, $s = 1, \ldots, n$, where $p_j = \pi_j \prod_{s=1}^{n} x_s^{a_{sj}}$ with $a_{sj} \in \mathbb{R}$ and $\pi_j, c_s \in \mathbb{R}_{>0}$. These systems arise in the study of reversible self-assembly systems and reversible chemical reaction networks. Geometric properties of the systems are explored to extend the iterative proportional scaling algorithm. They are also applied to improve the convergent rate of the iterative proportional scaling algorithm when dealing with ill-conditioned systems. Reduction to convex optimization is discussed. Computational results are also presented.

1. Introduction

A real function of the form

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{K} \pi_i x_1^{\alpha_{1i}} x_2^{\alpha_{2i}} \cdots x_n^{\alpha_{ni}},$$

where $\pi_i > 0$ and $\alpha_{ji} \in \mathbb{R}$, is called a *posynomial*. We are interested in finding real positive solutions to posynomial systems of the form:

$$\sum_{j=1}^{m} a_{sj} p_j = c_s, \ s = 1, \ldots, n, \quad \text{where } p_j = \pi_j \prod_{s=1}^{n} x_s^{a_{sj}} \tag{1.1}$$

with $a_{sj} \in \mathbb{R}$, and $\pi_j, c_s \in \mathbb{R}_{>0}$. These systems arise in the study of algorithmic self-assembly. As will be explained in the next section, a solution to such a system corresponds to an equilibrium of a reversible self-assembly system [4] or a reversible chemical reaction network [6]. It will be shown that the real positive solution to such a system is unique when the underlying linear system:

$$\sum_{j=1}^{m} a_{sj} y_j = c_s, \ s = 1, \ldots, n,$$

has a real positive point.

A special case of the system (1.1) is where $a_{nj} = 1$ for all j; then letting $z = x_n$, the problem becomes one of finding a probability function of the form:

$$p_j = \pi'_j z \prod_{s=1}^{n-1} x_s^{a_{sj}}, \; j = 1, \ldots, m, \tag{1.2}$$

that satisfies $\sum_j p_j = 1$ and $\sum_{j=1}^m a_{sj} p_j = c'_s$ for $s = 1, \ldots, n-1$, where $\pi'_j = \frac{\pi_j}{c_n}$ for $j = 1, \ldots, m$, and $c'_s = \frac{c_s}{c_n}$ for $s = 1, \ldots, n-1$. In this case the problem is exactly solving *maximum likelihood equations* in statistics (a good example is on page 114 of [7]). It is known that the positive solution is unique (see [3], also Chapter 4 of [2]). The solution can be found by a numerical algorithm called *iterative proportional scaling* [3]. With this method a simple transformation is applied to the system so that two additional conditions are satisfied: $\sum_{s=1}^n a_{sj} = 1$ and $\sum_{i=s}^n c_s = 1$. The following theorem proven in [3] can then be applied to solve for the solution.

Theorem 1.1. *Consider the system*

$$\sum_{j=1}^m a_{sj} p_j = c_s, \; s = 1, \ldots, n,$$

where

$$p_j = \pi_j \prod_{s=1}^n x_s^{a_{sj}},$$

$a_{sj} \geq 0$, $\sum_{s=1}^n a_{sj} = 1$,$c_s > 0$,*and* $\sum_{i=s}^n c_s = 1$. *Suppose that* $Ax = c$ *has a real positive solution, then the sequence* $\langle p^{(k)} : k = 0, 1, 2, \ldots \rangle$ *with* $p^{(k)} = (p_1^{(k)}, \ldots, p_m^{(k)})$ *and defined by* $p_i^{(0)} = \pi_i$, $p_i^{(n+1)} = p_i^{(n)} \prod_{r=1}^s \left(\frac{c_r}{c_r^{(n)}} \right)^{a_{ri}}$, *where* $c_r^{(n)} = \sum_{i=1}^m a_{ri} p_i^{(n)}$, *converges to the unique positive solution of the system.*

In this paper we generalize the iterative proportional scaling algorithm to solve the posynomial system (1.1). Our approach is to associate the system (1.1) with a parameterized family of systems S_v of the form 1.2 where $v \in \mathbb{R}_{>0}$. We define a function g on a suitable positive real interval such that for v in the interval, $g(v)$ is the value of z in the unique real positive solution to S_v. We show that the function $\frac{g(v)}{v}$ is decreasing in this interval and has a unique fixed point u. Moreover the solution to our system (1.1) can be easily obtained from that of S_u. The fact that g has a unique fixed point and $\frac{g(x)}{x}$ is decreasing allows us to devise a bisection strategy to find u by solving a sequence S_{v_i}, each using the iterative proportional scaling method. We show that if in the solution to our system, $\sum_i p_i = u^{-1}$ and the required precision is ϵ, then the number of times where we apply the iterative proportional scaling method can be bounded by $O(|\log u| + \log(1/\epsilon))$.

The rest of the paper is organized as follows. In Sect. 2 we discuss the geometric perspective of the system (1.1) and its application to self-assembly systems and reversible chemical reaction networks. In Sect. 3 we explore a special geometric property of the system (1.1). This property is applied in this section to improve the convergent rate of iterative proportional scaling for certain ill-conditioned cases.

It is also utilized in the next section in generalizing the proportional scaling algorithm. The details of the extended proportional scaling algorithm are presented in Sect. 4. We implement our algorithm in Mathematica 5.1. We discuss convex optimization as an alternative method for solving (1.1). Computational results are presented in Sect. 6, including comparison of the extended iterative proportional scaling method with convex optimization.

2. Geometric Perspective and Motivation from Self Assembly and Chemical Reaction Network

Given a set $\mathcal{B} = \{v_1, \ldots, v_d\}$ in $\mathbb{R}^m$ and a positive vector $\mu = (\mu_1, \ldots, \mu_d)$, let $V_{I_{\mathcal{B}},\mu}$ denote the set of of $x \in \mathbb{R}^m$ such that $x^{v_i^+} = \mu_i x^{v_i^-}$, for $i = 1, \ldots, d$, where v_i^+ and v_i^- are the nonnegative vectors with disjoint support such that $v_i = v_i^+ - v_i^-$. Suppose $v_1, \ldots, v_d \in \mathbb{Z}^m$. Then $V_{I_{\mathcal{B}},\mu}$ is the zero set of the ideal $I_{\mathcal{B},\mu} = \langle p^{v_i^+} - \mu_i p^{v_i^-} : v_i = v_i^+ - v_i^- \in \mathcal{B}\rangle$ in the polynomial ring $\mathbb{R}[p] = \mathbb{R}[p_1, \ldots, p_m]$. It is called the *deformed toric variety* of $\mathcal{B}$ under μ. When $\mu = (1, \ldots, 1)$, $V_{I_{\mathcal{B}},\mu}$ is simply called the *toric variety* $V_{I_{\mathcal{B}}}$.

For $a = (a_i) \in \mathbb{R}_{>0}$ and $b = (b_i) \in \mathbb{R}^m$, we define $a^b = \prod_i a_i^{b_i}$.

Let $A = (a_{sj})$ be an n by m real matrix and suppose $\mathcal{B} = \{v_1, \ldots, v_d\}$ spans the kernel of A. Let $\pi = (\pi_1, \ldots, \pi_m) \in \mathbb{R}^m_{>0}$, and $\mu_i = \pi^{v_i}$ for $i = 1, \ldots, d$. Then it can be shown that for $c = (c_1, \ldots, c_n)$ with $c_i > 0$ for all i, $p \in V_{I_{\mathcal{B}},\mu} \cap \{y \mid y > 0, Ay = c\}$ if and only if p yields a solution to (1.1). That is:

$$\sum_{j=1}^{m} a_{sj} p_j = c_s, \; s = 1, \ldots, n,$$

where

$$p_j = \pi_j \prod_{s=1}^{n} u_s^{a_{sj}}$$

with some $u \in \mathbb{R}^n_{>0}$.

Note that $V_{I_{\mathcal{B}},\mu} \cap \{y \mid y > 0, Ay = c\}$ is determined by the kernel of A, μ and c. Therefore we may assume without loss of generality that the matrix A is of rank n.

Let A_i be the i-th column of A for $i = 1, \ldots, m$. For $x \in \mathbb{R}^n_{>0}$, let $x^A = (x^{A_1}, \ldots, x^{A_m})$ and $\pi x^A = (\pi_1 x^{A_1}, \ldots, \pi_m x^{A_m})$. Then (1.1) can be rewritten as: $Ap = c$ where $p = \pi u^A$ with $u \in \mathbb{R}^n_{>0}$. Let $U = ker A$. Then $U^\perp = im(A^t)$. Suppose $c = Ap_0$ for some $p_0 \in \mathbb{R}^m_{>0}$. For $u \in \mathbb{R}^n_{>0}$, let $\mu = A^t \log u$, then $\mu \in im(A^t) = U^\perp$, and $p = \pi u^A = \pi e^\mu$ is a solution to (1.1) if and only if $Ap = c = Ap_0$ if and only if $p - p_0 \in U = ker A$. Hence our problem becomes one of finding $\mu \in U^\perp$ with $\pi e^\mu - p_0 \in U$, given positive vectors π and p_0. It follows from Proposition B.1 of [6] that such a μ is unique. Therefore, there is a unique positive solution to (1.1).

In a self-assembly system or a reversible chemical reaction network, we have a collection of species whose concentrations are represented by variables $x_1, \ldots,$

x_m respectively. Each complex is represented by a monomial x^a where $x = (x_i)$ and $a = (a_i)$ with $a_i \in \mathbb{N}$. Each reaction (or *event* in the terminology of [4]) is associated with a binomial $\sigma x^a - \tau x^b$ where σ is the forward rate and τ the backward rate. The rate of change of concentration of the j-th species with time is modeled by $\dot{x_j} = F_j(x)$ where $F_j(x) = -\sum_i(\sigma_i x^{a_i} - \tau_i x^{b_i})v_i(j)$ with $v_i = a_i - b_i$ and $v_i(j)$ is the j-th coordinate of the vector v_i. Let $F = (F_i)$, then the system $\dot{x} = F(x)$ models the dynamics of the mass-action system [4, 6].

Let $\mathcal{B} = \{v_1, \ldots, v_d\}$ in $\mathbb{R}^m$ and $\mu = (\mu_1, \ldots, \mu_d)$, with $\mu_i = \sigma_i/\tau_i$. Then for real positive vectors $p \in \mathbb{R}^m$, $F(p) = 0$ iff $p \in V_{I_{\mathcal{B},\mu}}$ [6]. Moreover if the initial condition is x_0 then the flow determined by the system of differential equations $\dot{x} = F(x)$ satisfies the condition that $x(t) - x_0$ is in the linear span of $\mathcal{B}$. Thus if we choose a matrix A such that $\ker(A) =< v_1, \ldots, v_d >$, then the positive intersection of the deformed toric variety $V_{I_{\mathcal{B},\mu}}$ and the set $\{x \mid x \geq 0, Ax = Ax_0\}$ is precisely the set of positive real equilibria of the flow defined by $\dot{x} = F(x)$ with initial condition x_0.

Example 1. Consider the following reversible chemical reactions:

1. $2H_2 + O_2 = 2H_2O$ with forward rate k_1 and backwards rate k_2,
2. $Cl_2 + H_2 = 2HCl$ with forward rate k_3 and backwards rate k_4.

If we represent the concentrations of O_2, Cl_2, H_2, H_2O and HCl by x_1, x_2, x_3, x_4 and x_5, then the binomial associated with the first reaction is $k_1x_3^2x_1 - k_2x_4^2$; the binomial associated with the second reaction is $k_3x_2x_3 - k_4x_5^2$. Let $x = (x_1, x_2, x_3, x_4, x_5)$ then the dynamical system associated with these two reactions is governed by the following differential equations:

$$\dot{x} = \begin{pmatrix} \dot{x_1} \\ \dot{x_2} \\ \dot{x_3} \\ \dot{x_4} \\ \dot{x_5} \end{pmatrix} = F(x) = \begin{pmatrix} k_2x_4^2x_2 - k_1x_3^2x_1x_2 \\ k_4x_5^2 - k_3x_2x_3 \\ -2(k_1x_3^2x_1x_2 - k_2x_4^2x_2) - (k_3x_2x_3 - k_4x_5^2) \\ 2(k_1x_3^2x_1x_2 - k_2x_4^2x_2) \\ 2(k_3x_2x_3 - k_4x_5^2)\,. \end{pmatrix}$$

Let $B = \{v_1 = (1, 0, 2, -2, 0), v_2 = (0, 1, 1, 0, -2)\}$ and $\mu = (k_2/k_1, k_4/k_3)$. Then $I_{\mathcal{B},\mu} =< x_1x_3 - \frac{k_2}{k_1}x_4^2, x_2x_3 - \frac{k_4}{k_3}x_5^2 >$ and all positive points in $V(I_{\mathcal{B},\mu})$ are equilibria.

Assume that the initial condition of the system $\dot{x} = F(x)$ is $\tilde{x}_0 = (\tilde{x}_1, \ldots, \tilde{x}_5)$ and we choose

$$A = \begin{pmatrix} 1 & 0 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 1 & 1/2 \end{pmatrix}$$

so that the kernel of A is spanned by B. Then the real positive equilibrium is $V(I_{\mathcal{B},\mu}) \bigcap \{x \mid x > 0, Ax = A\tilde{x}_0\}$.

To get the corresponding posynomial system, we need to find one positive vector $\pi = (\pi_1, \ldots, \pi_5)$ such that $\mu_i = \pi^{v_i}$; that is, $\pi_1\pi_3^2\pi_4^{-2} = \frac{k_2}{k_1}, \pi_2\pi_3\pi_5^{-2} = \frac{k_4}{k_3}$. Let $\pi_1 = \pi_2 = \pi_3 = 1, \pi_4 = \sqrt{\frac{k_1}{k_2}}, \pi_5 = \sqrt{\frac{k_3}{k_4}}$; then $\mu_i = \pi^{v_i}$, and the corresponding posynomial system is

$$\begin{pmatrix} 1 & 0 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 1 & 1/2 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = A\tilde{x}_0,$$

where $p_1 = u_1$, $p_2 = u_2$, $p_3 = u_3$, $p_4 = \sqrt{\frac{k_1}{k_2}} u_1^{\frac{1}{2}} u_3$, $p_5 = \sqrt{\frac{k_3}{k_4}} u_2^{\frac{1}{2}} u_3^{\frac{1}{2}}$.

3. Geometric Property

In this section, we show that the first $n-1$ equations in the algebraic system (1.1) can be considered as a curve parameterized by x_n. In other words, let $x_1 = g_1(x_n), \ldots, x_{n-1} = g_{n-1}(x_n)$ be the function derived from the first $n-1$ equation in (1.1). Then the function $f_n(g_1(x_n), \ldots, g_{n-1}(x_n), x_n)$ is increasing in $(0, +\infty)$. We denote the real positive domain as $\mathbb{P}$ in the following proposition.

Proposition 3.1. *Let $c = (c_1, \ldots, c_n) \in \mathbb{P}^n$ and $A = (a_{ij})$ be an n by m real matrix of rank n, and suppose that $Ax = c$ has a real positive solution. Let $\pi = (\pi_1, \ldots, \pi_m) \in \mathbb{P}^m$, and let*

$$f_i(x_1, \ldots, x_n) = \sum_{j=1}^{m} \pi_j a_{ij} x_1^{a_{1j}} \cdots x_n^{a_{nj}} - c_i, \; i = 1, \ldots, n.$$

Then the set of real positive points determined by

$$\begin{cases} f_1(x_1, \ldots, x_n) = 0, \\ f_2(x_1, \ldots, x_n) = 0, \\ \quad \cdots\cdots \\ f_{n-1}(x_1, \ldots, x_n) = 0 \end{cases}$$

forms a curve $\{(X(x_n), x_n) \mid x_n \in \mathbb{R}_{>0}\}$ in $\mathbb{P}^n$ parametrized by x_n, where $X(x_n) = (x_1(x_n), \ldots, x_{n-1}(x_n))$ is such that $f_i(X(x_n), x_n) = 0$ for $i = 1, \ldots, n-1$. Moreover $f_n(X(a), a) \geq f_n(X(b), b)$ for any $a > b > 0$.

The proof of the proposition is in Appendix A.

One application of the proposition is in dealing with an ill-conditioned system where the exponent in one of the variables is unusually large or unusually small. In this case, the convergent rate of the iterative proportional scaling method tends to be slow. We can improve the convergent rate by finding the value of one variable using bisection while finding the values of the others using iterative proportional scaling method as follows.

Consider the system

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & \cdots & \cdots & a_{nm} \\ 1 & 1\cdots & \cdots & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ 1 \end{pmatrix},$$

where

$$p_j = \pi_j \mu \prod_{s=1}^{n} x_s^{a_{sj}}.$$

Such a system can be translated into the system in Theorem 1.1 as discussed before, and thus can be solved by iterative proportional scaling method.

Without loss of generality, assume that a_{1n} is the exponent which is unusually large. From Proposition 3.1, we know that in the positive real domain the first $n-1$ equations in (1.1) determine a curve $\mathcal{L}$ parametrized by x_n. Thus for each real positive x_n there is a unique point $(g(x_n), x_n)$ on $\mathcal{L}$. Moreover f_n is an increasing function along $\mathcal{L}$, that is, $f_n(g(x_n), x_n)$ is increasing. If we can evaluate the function $g(x_n)$, then we can evaluate $f_n(g(x_n), x_n)$. Hence we can approximate the x_n-coordinate of the solution using the bisection method. To evaluate the function g at point $x_n = h$, we observe that after setting $x_n = h$, $g_1(h), \ldots, g_{n-1}(h)$ is the solution of the first $n-1$ equations form an algebraic system in $n-1$ variables of the same form as (1.1), with $\pi_i (1 \le i \le m)$ replaced by $\pi_i h^{a_{in}}$. Thus the reduced system has a unique real positive solution and can be solved by the iterative proportional scaling method. Moreover, the solution is none other than $g(h)$. If $f_n(g(h), h) > c_n$ and $(x_1^*, \ldots, x_n^*)$ is the solution of (1.1), then by Proposition 3.1, we know that $x_n^* < h$, otherwise $x_n^* \ge h$. We can repeat the above procedure to approximate x_n^* as close as we want. A precise description of the improved algorithm is given below:

1. Initially set $startpoint = 0, endpoint = c_n$.
2. Set $x_n = c_n/2$.
3. Let $\pi'_j = \pi_j * x_n^{a_{nj}}$. Solve the following system

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n-1,1} & \cdots & \cdots & a_{n-1,m} \\ 1 & 1\cdots & \cdots & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ 1 \end{pmatrix},$$

where

$$p_j = \pi'_j \mu \prod_{s=1}^{n-1} x_s^{a_{sj}},$$

by the iterative proportional scaling algorithm.

4. Compute $x_1, \ldots, x_{n-1}$ by solving the linear equation

$$\log p_j = \log \pi'_j + \log \mu + \sum_{s=1}^{n-1} a_{sj} \log x_s.$$

5. Compute $p_i = \pi_j \mu \prod_{s=1}^{n} x_s^{a_{sj}}$.
6. If $\sum a_{ni} p_i > c_n$, then let *endpoint* $= x_n$ and repeat from step 2. If $\sum a_{ni} p_i < c_n$, then let *startpoint* $= x_n$ and repeat from step 2. If $\sum a_{ni} p_i = c_n$, then return $(x_1, \ldots, x_n)$ as the solution.

4. The Algorithm

Consider a polynomial system of the form:

$$\sum_{j=1}^{m} a_{sj} p_j = c_s, \; s = 1, \ldots, n, \quad \text{where } p_j = \pi_j \prod_{s=1}^{n} x_s^{a_{sj}} \tag{4.1}$$

with $a_{sj} \in \mathbb{R}$, and $\pi_j, c_s \in \mathbb{R}_{>0}$. Let $\mu^{-1} = \sum_i p_i$, and $q_i = \mu p_i$. Then the system (4.1) is equivalent to

$$\sum_{j=1}^{m} a_{sj} q_j = \mu c_s, \; s = 1, \ldots, n; \quad \sum_i q_i = 1,$$

where

$$q_j = \pi_j \mu \prod_{s=1}^{n} x_s^{a_{sj}}.$$

For $v \in \mathbb{R}_{>0}$, let S_v be the system

$$\sum_{j=1}^{m} a_{sj} q_j = v c_s, \; s = 1, \ldots, n; \quad \sum_i q_i = 1, \tag{4.2}$$

where

$$q_j = \pi_j z \prod_{s=1}^{n} x_s^{a_{sj}}.$$

Note that S_v can be regarded as a posynomial system of the form (4.1) with $q_1, \ldots, q_m$ and z playing the role of $p_1, \ldots, p_m$. Therefore as discussed in Sect. 2, S_v has a unique positive solution if and only if (4.2) has a real positive solution. It is easy to verify that (4.2) has a positive solution if and only if $v \in [\alpha, \beta]$ where $\alpha = \min(\sum p_i)$ and $\beta = \max(\sum p_i)$ under the linear constraints $\sum_{j=1}^{m} a_{sj} p_j = c_s$, $s = 1, \ldots, n$; $p_i \geq 0$, $i = 1, \ldots, m$. Suppose $v \in [\alpha, \beta]$ and $x_1 = \rho_1, \ldots, x_n = \rho_n$, $z = \rho$ is the unique real positive solution to S_v; we define $\rho = g(v)$. We observe that if $v = g(v)$, then the system S_v is equivalent to (4.1). Moreover, the uniqueness of solution of (4.1) implies that the function g has a unique fixed point.

Proposition 4.1. *$g(x)/x$ is decreasing on $[\alpha, \beta]$.*

Proof. Suppose $v_1, v_2 \in [\alpha, \beta]$ and $v_1 < v_2$. Assume that $(x'_1, \ldots, x'_n, g(v_1))$ and $(x''_1, \ldots, x''_n, g(v_2))$ are the solution of S_{v_1} and S_{v_2} respectively. Then $(x'_1, \ldots, x'_n, z_1 = \frac{g(v_1)}{v_1})$ is the solution of

$$\sum_{j=1}^{m} a_{sj} q_j = c_s, \ s = 1, \ldots, n; \quad \sum_i q_i = 1/v_1,$$

where

$$q_j = \pi_j z \prod_{s=1}^{n} x_s^{a_{sj}},$$

and $(x''_1, \ldots, x''_n, z_2 = \frac{g(v_2)}{v_2})$ is the solution of

$$\sum_{j=1}^{m} a_{sj} q_j = c_s, \ s = 1, \ldots, n; \quad \sum_i q_i = 1/v_2,$$

where

$$q_j = \pi_j z \prod_{s=1}^{n} x_s^{a_{sj}}.$$

Now consider $x_1, \ldots, x_n$ as the function of z determined by $\sum_{j=1}^{m} a_{sj} q_j = c_s$, $s = 1, \ldots, n$; where $q_j = \pi_j z \prod_{s=1}^{n} x_s^{a_{sj}}$. From Proposition 3.1 in Sect. 3 we know that $\sum_i q_i$ is an increasing function in z. Let $\sigma(z) = \sum_i q_i(x_1(z), \ldots, x_n(z), z)$. Then $\sigma(z_i) = 1/v_i$ for $i = 1, 2$. Since $1/v_1 > 1/v_2$, we have $z_1 > z_2$. Hence $g(v_1)/v_1 \geq g(v_2)/v_2$ when $v_1 < v_2$. □

We have proved that $g(x)/x$ is decreasing on $[\alpha, \beta]$. We can use bisection search to find the fixed point of $g(x)$ on $[\alpha, \beta]$ (that is, the solution of $g(x)/x = 1$) since $g(x)/x$ is decreasing on $[\alpha, \beta]$.

Below we outline an algorithm for computing the fixed point u.

1. Compute the bound $[\alpha, \beta]$ for $1/\mu = \sum_i p_i$ under the linear constraints

$$\sum_{j=1}^{m} a_{sj} p_j = c_s, \ s = 1, \ldots, n; \quad p_j \geq 0, \ j = 1, \ldots, m,$$

   by linear programming.
2. We start our bisection search for $1/\mu$ at $v_0 = \frac{\alpha+\beta}{2}$ (if $\beta = +\infty$, we can choose v_0 as arbitrary positive number bigger than α). Apply iterative proportional scaling algorithm to solve for:

$$\sum_{j=1}^{m} a_{sj} q_j = v_0 c_s, \ s = 1, \ldots, n; \quad \sum_i q_i = 1,$$

   where

$$q_j = \pi_j z \prod_{s=1}^{n} x_s^{a_{sj}}.$$

3. Compute z through $\log z$ by solving the linear equation

$$\log q_j/\pi_j = \log z + \sum_{s=1}^{n} a_{sj} \log x_s,\ \ j = 1, \ldots, m.$$

4. If $z/v_0 = 1$, then obviously $q_j/v_0, j = 1, \ldots, m$, is the solution of (4.1). If $z/v_0 < 1$, then let $\beta = v_0$ and repeat step 2. If $z/v_0 > 1$, then let $\alpha = v_0$ and repeat step 2 (in this case, if $\beta = +\infty$, we set $v_0 = 2v_0$ when we repeat step 2). By doing this we can proceed to come close to $u = \sum_{i=1}^{n} p_i$ within a precision of ϵ using bisection search and compute $p_i = u * q_i$ is the solution of (4.1).

The convergence of iterative proportional scaling method has been proven in [3] and the convergence of bisection search has been proven in Proposition 4.1. Finally the real positive solution in (4.1) can be found by simple linear algebra once $p_i, i = 1, \ldots, m$, are known.

The running time of our algorithm is closely related with the convergent rate of iterative scaling method. In our algorithm, we apply $O(|\log u| + \log(1/\epsilon))$ times iterative proportional scaling method.

5. Reduction to Convex Optimization

As discussed earlier, our system can be interpreted as finding the intersection

$$V_{I_{\mathcal{B}},\mu} \cap \{y \mid y > 0, Ay = c\},$$

where the kernel of A is the linear span of B. A well-studied case is where we consider the intersection

$$V_{I_{\mathcal{B}}} \cap \{y \mid y > 0, Ay = c\}$$

with the additional assumption that $(1, \ldots, 1)$ is in the row space of A. In this case the unique solution is where the entropy function $-\sum_i p_i \log p_i$ is maximized in the convex set $\{y \mid y > 0, Ay = c\}$ (see p. 115 of [7]). In the more general situation, we do not assume that $(1, \ldots, 1)$ is in the row space of A. We can show that the unique solution to our system is where the function $-\sum_i (x_i \log \frac{x_i}{\pi_i} - x_i + \pi_i)$ is maximized over the convex set $\{y \mid y > 0, Ay = c\}$. The function is the relative entropy function adjusted by the difference in weights between x and π. It is precisely the negative of the *Kullback-Leibler divergence* function $D(x, \pi)$, which is known to be convex (see p. 90 of [1]). Hence the problem can be reduced to a convex optimization problem [1] of minimizing the convex function $D(x, \pi)$ over the convex set $\{x \mid x \geq 0, Ax = b\}$. The objective function $D(x, \pi)$ is not self-concordant (see p. 498 of [1]) and it is difficult to analyze the convergence rate when applying the general convex optimization method. Mathematica 5.1 has implemented the convex optimization as one of the built-in functions, but the error is much bigger than that of the extended iterative method when the entries in matrix A are relatively big. The comparison of computational results of these two algorithms is given in Sect. 6.3.

6. Computation

In this section we discuss some of our computational results.

6.1. Computation of Extended Iterative Proportional Scaling Algorithm

We implement our algorithm with Mathematica 5.1. The experiments show that when the the iterative proportional scaling algorithm runs well in the inner loop, the extended iterative proportional scaling algorithm works well in terms of accuracy and running time. The following is an instance of Example 1 in Sect. 2.

In this example,

$$A = \begin{pmatrix} 2 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 2 & 2 & 1 \end{pmatrix}.$$

Let $\pi = (1, 1, 1, 1, 1)$ and

$$c = \begin{pmatrix} 3.56081 \\ 10.0889 \\ 25.0121 \end{pmatrix}.$$

Then the solution of the above system is

$$(0.445231, 1.67631, 2.66594).$$

With 16625 iterations, the computation result from the extend iterative proportional scaling is

$$(0.445231, 1.67631, 2.66594),$$

which is exactly the actual solution of the system.

6.2. Improvement in the Iterative Proportional Scaling Method

We implement both the improved iterative proportional scaling method for dealing with ill-conditioned systems and the standard iterative proportional scaling method.

Example 2.

$$A = \begin{pmatrix} 40 & 1 & 3 & 3 & 2 \\ 1 & 4 & 3 & 2 & 5 \\ 0 & 1 & 4 & 3 & 1 \end{pmatrix},$$

$\pi = (1, 2, 3, 4, 5)$ and

$$c = \begin{pmatrix} 1.41519 \\ 4.38626 \\ 1.01694 \end{pmatrix}.$$

The value of p_i of the above system is

$$(0.30736 * 10^{-24}, 0.592453, 0.00165342, 0.00599196, 0.399902).$$

The standard iterative proportional scaling method takes 155868 iterations and get the value of p_i as

$$(0.30462 * 10^{-24}, 0.592453, 0.00165358, 0.00599177, 0.399902).$$

The improved iterative proportional scaling method takes 24444 iterations and get the value of p_i as

$$(0.3052 * 10^{-24}, 0.592453, 0.00165354, 0.005918, 0.399902).$$

In these computation results, we see that the output from both methods is close to the correct answer, but the iteration time of the improved iterative proportional method is much shorter.

6.3. Comparison

We compare the computation results of convex optimization method built in Mathematica 5.1 and that of the extended iterative proportional method. Since the optimization is already built in Mathematica 5.1, it will make sense to only compare the computation performance here. When the entries in the matrix A are small, both algorithms work well. When the entries in the matrix A get bigger, the performance of Convex Optimization tends to be unstable while the extended iterative proportional scaling method continues to perform well.

Example 3. Given

$$A = \begin{pmatrix} 6 & 1 & 0 & 6 & 0 \\ 2 & 6 & 3 & 4 & 2 \\ 5 & 3 & 1 & 2 & 1 \end{pmatrix} \text{ and } c = \begin{pmatrix} 1.4173575 \\ 6.876265 \\ 3.2714125 \end{pmatrix},$$

solve the posynomial system $Ap = c$, where $p_j = \prod_{s=1}^{3} x_s^{a_{sj}}$.

The convex optimization method in Mathematica 5.1 reports error for this question while the extended iterative proportional scaling method returns exactly the solution $x_1 = 0.75, x_2 = 1, x_3 = 0.8$, taking 11861 iterations.

Example 4. Given

$$A = \begin{pmatrix} 5 & 1 & 2 & 5 & 0 \\ 7 & 4 & 6 & 5 & 1 \\ 6 & 7 & 0 & 6 & 8 \end{pmatrix} \text{ and } c = \begin{pmatrix} 1.9043664000 \\ 4.9184137600 \\ 3.189678080 \end{pmatrix},$$

solve the posynomial system $Ap = c$, where $p_j = \prod_{s=1}^{3} x_s^{a_{sj}}$.

The convex optimization method in Mathematica 5.1 returns p_i as $p_1 = -5.46438 * 10^{-6}$, $p_2 = 0.356956$, $p_3 = 0.485795$, $p_4 = 4.35763 * 10^{-4}$, $p_5 = 0.115164$, which is obviously a wrong solution because p_1 is negative in the solution returned, while the extended iterative proportional scaling method returns exactly the solution $x_1 = 0.75, x_2 = 1, x_3 = 0.8$ with 60652 iterations.

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[2] W. Fulton. *Introduction to Toric Varieties.* Princeton Unversity Press, 1993.

[3] J. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics*, 43:1470–1480, 1972.

[4] L. Adleman. Toward a general theory of self-assembly. In *Foundation of Nanoscience Self-Assembled Architectures and Devices*, Snowbird, Utah, 2004.

[5] J. L. Troutman. *Variational Calculus with Elementary Convexity.* Springer-Verlag New York, 1983.

[6] M. Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch. Rational Mech. Anal.*, 132:311–370, 1995.

[7] B. Sturmfels. *Solving Systems of Polynomial Equations.* Number 97. American Mathematical Society, 2002.

Appendix A.

Theorem A.1 **([5]).** *Let A be an open set in R^{n+k} and let $f : A \to R^n$ be a C^r function. Write f in the form $f(x, y)$ where x and y are elements of R^k and R^n. Suppose that (a, b) is a point in A such that $f(a, b) = 0$ and the determinant of the $n \times n$ matrix whose elements are the derivatives of the n component functions of f with respect to the n variables, written as y, evaluated at (a, b), is not equal to zero; then there exists a neighborhood B of a in R^n and a unique C^r function $g : B \to R^k$ such that $g(a) = b$ and $f(x, g(x)) = 0$ for all $x \in B$.*

Proposition A.2. *For any $n \times m$ matrix E with rank n, if A is $m \times m$ positive definite matrix, then $A^{-1} - E^T(EAE^T)^{-1}E$ is nonnegative definite.*

Proof. Let $P = \begin{pmatrix} I_m & 0 \\ -(EAE^T)^{-1}E & I_n \end{pmatrix}$; then

$$P^T \begin{pmatrix} A^{-1} & E^T \\ E & EAE^T \end{pmatrix} P = \begin{pmatrix} A^{-1} - E^T(EAE^T)^{-1}E & 0 \\ 0 & EAE^T \end{pmatrix},$$

so $A^{-1} - E^T(EAE^T)^{-1}E$ is nonnegative definite if and only if

$$\begin{pmatrix} A^{-1} & E^T \\ E & EAE^T \end{pmatrix}$$

is nonnegative definite.

Notice that

$$\begin{pmatrix} A^{-1} & 0 \\ E & 0 \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & A^{-1} \end{pmatrix} \begin{pmatrix} A^{-1} & E^T \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} A^{-1} & E^T \\ E & EAE^T \end{pmatrix}.$$

For any vector v, since both A and A^{-1} are positive definite,

$$v \begin{pmatrix} A^{-1} & E^T \\ E & EAE^T \end{pmatrix} v^T = v \begin{pmatrix} A^{-1} & 0 \\ E & 0 \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & A^{-1} \end{pmatrix} \begin{pmatrix} A^{-1} & E^T \\ 0 & 0 \end{pmatrix} v^T \geq 0;$$

thus $A^{-1} - E^T(EAE^T)^{-1}E$ is nonnegative definite. □

Proposition A.3. *Let $c = (c_1, \ldots, c_n) \in \mathbb{P}^n$ and $A = (a_{ij})$ be an n by m real matrix of rank n, and suppose that $Ax = c$ has a real positive solution. Let $\pi = (\pi_1, \ldots, \pi_m) \in \mathbb{P}^m$, and let $f = (f_1, \ldots, f_{n-1})$ be a function mapping $\mathbb{R}^n \to \mathbb{R}^{n-1}$ defined by*

$$f_i = \sum_{j=1}^{m} \pi_j a_{ij} x_1^{a_{1j}} \cdots x_n^{a_{nj}} - c_i, \ i = 1, \ldots, n-1.$$

Then there exists a unique continuously differentiable function $g = (g_1, \ldots, g_{n-1}) : \mathbb{P} \to \mathbb{P}^{n-1}$ such that $f(g(t), t) = 0$ for all $t > 0$.

Proof. Let $C_{(n-1)\times(n-1)} = (c_{ij}) = (\frac{\partial f_i}{\partial x_j})$ for $0 < i, j < n$, and let

$$b_j = \pi_j x_n^{a_{nj}} \prod_{s=1}^{n-1} x_s^{a_{js}}, \ 0 < j < m+1.$$

Then $c_{ij} = x_j^{-1} \sum_{k=1}^{m} a_{ik} a_{jk} b_k \ (i \neq j)$ for $0 < i, j < n$.

Let

$$B = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1,n-1} & a_{2,n-1} & \cdots & a_{m,n-1} \end{pmatrix};$$

then $C = B \operatorname{diag}(b_1, \ldots, b_m) B^T$.

The rank of B is $n-1$ since the rank of A is n and it is easy to prove that C is a positive definite matrix when $(x_1, \ldots, x_n) \in \mathbb{P}^n$, and thus $\det(C) \neq 0$.

Suppose $v = (v_1, \ldots, v_n) \in P^n$ and $f(v) = 0$; then by Theorem A.1 in Appendix and the proof above, there exists a unique different function $g = g(x_n)$ defined in the neighborhood of v such that

$$f(g_1(x_n), \ldots, g_{n-1}(x_n), x_n) = 0.$$

Since for any given $t > 0$, there always exists a unique $u = (u_1, \ldots, u_{n-1})$ such that $f(u, t) = 0$, and from the uniqueness we must have $g(t) = (u_1, \ldots, u_{n-1})$. By Theorem A.1 in Appendix $g(t)$ is a continuously differentiable function in $(0, +\infty)$. □

Proposition A.4. *Let $F : R^{n+m} \to R^m$ be a C^r function. Write F in the form $F(x, y) = (F_1(x, y), \ldots, F_m(x, y))$ for $x \in R^n, y \in R^m$. Suppose $F(a, b) = 0$ and $\det(\frac{\partial(F_1, \ldots, F_m)}{\partial(y_1, \ldots, y_m)})(a, b) \neq 0$. Let $y(x) = (y_1(x), \ldots, y_m(x))$ be the implicit function defined in a neighborhood B of a so that $F(x, y(x)) = 0$ for all $x \in B$. Then for all $x \in B$,*

$$Dy(x) = -[D_y F(x, y)]^{-1} D_x F(x, y),$$

where

$$Dy(x) = \begin{pmatrix} (y_1)'_{x_1} & (y_1)'_{x_2} & \cdots & (y_1)'_{x_n} \\ (y_2)'_{x_1} & (y_2)'_{x_2} & \cdots & (y_2)'_{x_n} \\ \cdots & \cdots & \cdots & \cdots \\ (y_m)'_{x_1} & (y_m)'_{x_2} & \cdots & (y_m)'_{x_n} \end{pmatrix},$$

$$D_xF(x,y)=\begin{pmatrix}(F_1)'_{x_1} & (F_1)'_{x_2} & \cdots & (F_1)'_{x_n}\\ (F_2)'_{x_1} & (F_2)'_{x_2} & \cdots & (F_2)'_{x_n}\\ \cdots & \cdots & \cdots & \cdots\\ (F_m)'_{x_1} & (F_m)'_{x_2} & \cdots & (F_m)'_{x_n}\end{pmatrix},$$

$$D_yF(x,y)=\begin{pmatrix}(F_1)'_{y_1} & (F_1)'_{y_2} & \cdots & (F_1)'_{y_m}\\ (F_2)'_{y_1} & (F_2)'_{y_2} & \cdots & (F_2)'_{y_m}\\ \cdots & \cdots & \cdots & \cdots\\ (F_m)'_{y_1} & (F_m)'_{y_2} & \cdots & (F_m)'_{y_m}\end{pmatrix}.$$

Proof of Proposition 3.1. Let

$$b_j=\pi_j x_n^{a_{nj}}\prod_{s=1}^{n-1}x_s^{a_{js}},\ 0<j<m+1,$$

$$C=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1m}\\ a_{21} & a_{22} & \cdots & a_{2m}\\ \cdots & \cdots & \cdots & \cdots\\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,m}\end{pmatrix},$$

$$U=\begin{pmatrix}b_1 & & \\ & \cdots & \\ & & b_m\end{pmatrix},\ X=\begin{pmatrix}x_1 & & \\ & \cdots & \\ & & x_{n-1}\end{pmatrix},\ v=\begin{pmatrix}a_{n1}\\ a_{n2}\\ \vdots\\ a_{nm}\end{pmatrix}.$$

Then

$$D_xF(x,y)=\begin{pmatrix}(f_1)'_{x_n}\\ (f_2)'_{x_n}\\ \vdots\\ (f_{n-1})'_{x_n}\end{pmatrix}={x_n}^{-1}\begin{pmatrix}\sum_{j=1}^m a_{nj}a_{1j}b_j\\ \sum_{j=1}^m a_{nj}a_{2j}b_j\\ \vdots\\ \sum_{j=1}^m a_{nj}a_{n-1,n}b_j\end{pmatrix}={x_n}^{-1}CUv,$$

$$D_y(F(x,y))=\begin{pmatrix}(f_1)'_{x_1} & (f_1)'_{x_2} & \cdots & (f_1)'_{x_{n-1}}\\ (f_2)'_{x_1} & (f_2)'_{x_2} & \cdots & (f_2)'_{x_{n-1}}\\ \cdots & \cdots & \cdots & \cdots\\ (f_{n-1})'_{x_1} & (f_{n-1})'_{x_2} & \cdots & (f_{n-1})'_{x_{n-1}}\end{pmatrix}=CUC^TX^{-1}.$$

Note that

$$\frac{d\,f_n(X(x_n),X_n)}{dx_n}=\frac{d(\sum_{j=1}^m \pi_j a_{nj}x_1^{a_{1j}}\cdots x_n^{a_{nj}}-c_n)}{dx_n}$$

$$=\sum_{j=1}^m a_{nj}^2\frac{b_j}{x_n}+\sum_{j=1}^m\sum_{i=1}^{n-1}b_ja_{nj}a_{ji}\frac{(x_i)'_{x_n}}{x_i}=x_n^{-1}v^TUv+v^TU^TC^TX^{-1}\begin{pmatrix}(x_1)'_{x_n}\\ \vdots\\ (x_{n-1})'_{x_n}\end{pmatrix}.$$

By Proposition A.4

$$\begin{pmatrix} (x_1)'_{x_n} \\ \vdots \\ (x_{n-1})'_{x_n} \end{pmatrix} = -D_y(F(x,y))^{-1}D_xF(x,y) = (-X(CUC^T)^{-1})(x_n^{-1}CUv).$$

Thus

$$\begin{aligned}\frac{d\,f_n(X(x_n),X_n)}{dx_n} &= x_n^{-1}v^TUv - x_n^{-1}v^TU^TC^T(CUC^T)^{-1}CUv \\ &= x_n^{-1}(v^TUv - v^TU^TC^T(CUC^T)^{-1}CUv) \\ &= x_n^{-1}((Uv)^T(U^{-1} - C^T(CUC^T)^{-1}C)Uv).\end{aligned}$$

By proposition A.2 in Appendix A, $U^{-1} - C^T(CUC^T)^{-1}C$ is nonnegative definite, $(Uv)^T(U^{-1} - C^T(CUC^T)^{-1}C)Uv \geq 0$, when $x_n > 0$, $\frac{d\,f_n(X(x_n),X_n)}{dx_n} \geq 0$. Hence $f_n(X(a),a) \geq f_n(X(b),b)$ for any $a > b > 0$. □

Ming-Deh Huang and Qing Luo
Computer Science Department
University of Southern California
941 W. 37th Place, Los Angeles, CA 90089-0781 USA
e-mail: huang@pollux.usc.edu
qingl@pollux.usc.edu