Numerical and Geometric Properties of a Method for Finding Points on Real Solution Components

Wenyuan Wu*
Chongqing Institute of Green and Intelligent Technology,CAS
wuwenyuan@cigit.ac.cn

Greg Reid†
Department of Applied Mathematics
University of Western Ontario
London, Ontario Canada
reid@uwo.ca

Yong Feng‡
Chongqing Institute of Green and Intelligent Technology,CAS
yongfeng@cigit.ac.cn

ABSTRACT

We consider a critical point method developed in our earlier work for finding certain solution (witness) points on real solution components of real polynomial systems of equations. The method finds points that are critical points of the distance from a plane to the component with the requirement that certain regularity conditions are satisfied. In this paper we analyze the numerical stability of the method. We aim to find at least one well conditioned witness point on each connected component by using perturbation, path tracking and projection techniques. An optimal-direction strategy and an adaptive step size control strategy for path following on high dimensional components are given.

Categories and Subject Descriptors

G.1.0 [**NUMERICAL ANALYSIS**]: General—*Numerical algorithms*

General Terms

Algorithms, Design

Keywords

numerical algebraic geometry, real witness points, singular value, root isolation, perturbation, path tracking

1. INTRODUCTION

This article is a contribution to the development of numerical algorithms for computational algebraic geometry pioneered by Sommese, Wampler, Verschelde and others [17, 2], and recently developed to the real case by Lu [13], Besana et al. [5], Rouillier et al [15] and Hauenstein [10]. In contrast, remarkable developments concerning the computation of real radical of zero dimensional polynomial systems due to Laserre et al [12] are based on moment matrix and numerical semi-definite programming. Recently such moment matrix completion techniques are explored by Zhi et al in [14] for finding at least one real root of a given semi-algebraic system. Furthermore, based on critical point techniques and moment matrix completion, they studied the computation of verified real solutions on components of positive dimensional systems in [21].

Related symbolic approaches for computational real algebraic geometry include Sturm's ancient method for counting real roots of a polynomial and Tarski's real quantifier elimination [18, 16] and cylindrical algebraic decomposition (CAD) introduced by Collins [7] and improved by Hong [11]. Recent improvement of CAD by using triangular decompositions are given in [6] for solving semi-algebraic systems. But the double exponential cost of the CAD algorithm [8] is the main barrier to its application. For more references see [1].

Previously in [20], we introduced a perturbation method aiming to obtain a real witness point on each real connected solution component with high accuracy of systems

$$f = (f_1, f_2, ..., f_k) = 0 \tag{1}$$

of k polynomials from $\mathbb{R}[x_1, ..., x_n]$ satisfying the following regularity assumptions:

A$_1$: $V_{\mathbb{R}}(f_1, f_2, ..., f_i)$ has dimension $n - i$ for $1 \leq i \leq k$.

A$_2$: the ideal $I_i = \langle f_1, f_2, ..., f_i \rangle$ is radical for $1 \leq i \leq k$.

The solution set or variety of $f = 0$ is:

$$V_{\mathbb{R}}(p_1, ..., p_k) = \{x \in \mathbb{R}^n : p_j(x) = 0,\ 1 \leq j \leq k\} \tag{2}$$

As explained in our earlier work, singularities inevitably occur in such methods, and it is important to develop approaches to deal with them. The key in our approach to finding well-conditioned witness points is to track a path of a perturbed system to escape the numerically difficult region near the singularity and project back to the original real variety.

As a sequel of our previous work [20] we consider numerical aspects of the methods in this paper with the assumption that a real witness point on a positive dimensional component is obtained already. The main contributions of this paper are two strategies to ensure the success of this method

*The work is partly supported by the projects cstc2013jjys0002, West Light Foundation of CAS and the grant of Chongqing Key Laboratory of Automated Reasoning and Cognition.

†Research supported in part by an NSERC.

‡Supported by the 973 project 2011CB302402.

SNC'14, July 28-31, 2014, Shanghai, China.

http://dx.doi.org/10.1145/2631948.2631969

which are not provided in [20]. One is to use adapted step size to avoid jumping to another component during tracking the path by the prediction-projection technique. Another is to determine the direction by which we can escape the singularity as soon as possible.

Shub and Smale's condition number analysis of nonlinear systems can be used to estimate the size of the convergent region of a root and the minimal distance to other roots [9, 4]. Based on it Beltran and Leykin gave a rigorous algorithm for homotopy path-following and its complexity analysis to guarantee the convergence and avoid path-jumping [3].

In this paper, we will introduce an alternative way to estimate the root distance. We assume the convergence of Newton iteration during the projection stage and under this assumption we give a more practical method for avoiding the numerical difficulties. It differs from the analysis of Beltran and Leykin in that a rigorous bound is not given on the step size to ensure convergence while Beltran and Leykin give such a bound for certainty at a higher cost. But our method is guaranteed to detect "path jumping". If jumping or divergence happens, we can decrease the step size and eventually it will work for sufficiently small step size.

2. ERROR CONTROL

Our current work starts from the following construction introduced in [20] to find real witness points on $V_{\mathbb{R}}(f)$

$$f = 0, \qquad \sum_{i=1}^{k} \lambda_i \nabla f_i = \mathbf{n} \tag{3}$$

Here $\mathbf{n}$ is a random vector in $\mathbb{R}^n$ and (3) has $n+k$ equations and $n+k$ unknowns $(x, \lambda) = (x_1, ..., x_n, \lambda_1, ..., \lambda_k)$.

For numerical stability, we homogenize the system f by adding a variable x_0 satisfying a new equation $\sum_{i=0}^{n} x_i^2 = 1$ to obtain a *homogenized system* $\bar{f}$ except for the new inhomogeneous equation.

Now we estimate the maximum value of a given polynomial on the hyper-sphere $\sum x_i^2 = 1$. Let $\alpha = (a_0, ..., a_n)$ be the exponent of a monomial, then we define $\alpha^\alpha = a_0^{a_0} \cdots a_n^{a_n}$ and $0^0 = 1$.

PROPOSITION 2.1. *Given a homogenous polynomial* $g = \sum_{i=1}^{T} c_i \boldsymbol{x}^{\alpha_i}$ *of degree* d, *where* $\boldsymbol{x} = (x_0, ..., x_n)$, *an upper bound of* $|g|$ *on* $\sum_{i=0}^{n} x_i^2 = 1$ *in* $\mathbb{R}^{n+1}$ *is* $\sum_{i=1}^{T} \mid c_i \mid \sqrt{\frac{\alpha_i^{\alpha_i}}{d^d}}$.

PROOF. To attain the maximum value of a monomial $\boldsymbol{x}^\alpha = x_0^{a_0} \cdots x_n^{a_n}$ on $\sum x_i^2 = 1$, (without loss of generality we assume each $a_i > 0$), we use the method of Lagrange multipliers $\nabla \boldsymbol{x}^\alpha = \lambda \nabla \sum x_i^2$, i.e.

$$(a_0 \boldsymbol{x}^\alpha / x_0, ..., a_n \boldsymbol{x}^\alpha / x_n) = 2\lambda (x_0, ..., x_n)$$

This implies that $\frac{x_0^2}{a_0} = \frac{x_1^2}{a_1} = \cdots = \frac{x_n^2}{a_n}$. Since $\sum x_i^2 = 1$, we have $x_i^2 = \frac{a_i}{\sum a_i} = a_i/d$. So the maximum value of the monomial $\boldsymbol{x}^\alpha$ on $\sum x_i^2 = 1$ is

$$\sqrt{\frac{\prod_0^n a_i^{a_i}}{d^d}} = \sqrt{\frac{\alpha^\alpha}{d^d}}$$

Thus, $|g| = |\sum c_i \boldsymbol{x}^{\alpha_i}| \le \sum |c_i| \, |\boldsymbol{x}^{\alpha_i}| \le \sum \mid c_i \mid \sqrt{\frac{\alpha_i^{\alpha_i}}{d^d}}$. □

Let $J_{ij} = \partial \bar{f}_i / \partial x_j$. Applying Proposition 2.1 to each element of the matrix $(\|\nabla J_{ij}\|_2)_{(k+1)(n+1)}$ yields a constant matrix. Denote the maximum value of the $(k+1) \times (n+1)$ elements of this matrix by K.

EXAMPLE 2.1. *Consider the implicit curve* $f = \{y^2 + z^2 - (2\,x - x^2)^3 = 0, z - y^2 = 0\}$ *given in [20]. The corresponding homogenized system* $\bar{f}$ *is* $\{u^4 y^2 + u^4 z^2 - 8\,u^3 x^3 + 12\,u^2 x^4 - 6\,u x^5 + x^6, uz - y^2, x^2 + y^2 + z^2 + u^2 - 1\}$. *So the Jacobian*

$$\begin{bmatrix} J_{11} & 2\,u^4 y & J_{13} & 2\,u^4 z \\ 0 & -2\,y & z & u \\ 2\,x & 2\,y & 2\,u & 2\,z \end{bmatrix}$$

where $J_{11} = -24\,u^3 x^2 + 48\,u^2 x^3 - 30\,u x^4 + 6\,x^5$ *and* $J_{13} = 4\,u^3 y^2 + 4\,u^3 z^2 - 24\,u^2 x^3 + 24\,u x^4 - 6\,x^5$.

Using Proposition 2.1 to find the upper bounds of the 2*-norm of the gradient of each* $\mathcal{J}_{ij}$ *yields*

$$\begin{bmatrix} 144.2345348 & 10.44030651 & 92.45156154 & 0.5724334022 \\ 0 & 2.0 & 1.0 & 1.0 \\ 2.0 & 2.0 & 2.0 & 2.0 \end{bmatrix}$$

Thus $K = 144.2345348$. □

2.1 Root Isolation

For a square system $g(x) = \{f_1, ..., f_n, \sum_{i=0}^{n} x_i^2 - 1\}$ with $n+1$ variables, each f_i is homogenous. Suppose its maximum value of $\|\nabla J_{ij}\|_2$ on $\sum_{i=0}^{n} x_i^2 = 1$ is bounded by K.

First we assume there are at least two real roots $\boldsymbol{z}_1, \boldsymbol{z}_2$ and we will estimate the distance between $\boldsymbol{z}_1, \boldsymbol{z}_2$.

Let $\boldsymbol{z}_2 = \boldsymbol{z}_1 + \boldsymbol{h}$ and σ'_{n+1} be the smallest singular values of the Jacobian matrix of g at $\boldsymbol{z}_1$.

By the multivariate fundamental theorem of calculus,

$$g(\boldsymbol{z}_1 + \boldsymbol{h}) - g(\boldsymbol{z}_1) = \left(\int_0^1 \mathcal{J}(\boldsymbol{z}_1 + t\boldsymbol{h}) dt \right) \cdot \boldsymbol{h} = 0$$

where $\mathcal{J}$ is the Jacobian matrix of $g(z)$. So the matrix $D = \int_0^1 \mathcal{J}(\boldsymbol{z}_1 + t\boldsymbol{h}) dt$ must be singular and $\|\mathcal{J}(\boldsymbol{z}_1) - D\|_2 \ge \sigma'_{n+1}$ by Eckart-Young Theorem and consequently $\|\mathcal{J}(\boldsymbol{z}_1) - D\|_{\max} \ge \sigma'_{n+1}/(n+1)$.

Applying the multivariate mean value theorem to $\mathcal{J}_{ij}(\boldsymbol{z}(t))$ yields $\int_0^1 \mathcal{J}_{ij}(\boldsymbol{z}_1 + t\boldsymbol{h}) dt = \mathcal{J}_{ij}(\boldsymbol{z}_1) + \nabla \mathcal{J}_{ij}(c) \cdot \boldsymbol{h}$ where $c = \boldsymbol{z}_1 + \zeta \boldsymbol{h}$ for some $\zeta \in (0, 1)$. Thus, $\sigma'_{n+1}/(n+1) \le \|\mathcal{J}(\boldsymbol{z}_1) - D\|_{\max} \le \max_{ij} \|\nabla \mathcal{J}_{ij}\|_2 \|\boldsymbol{h}\|_2 \le K \|\boldsymbol{h}\|_2$ and consequently

$$\|\boldsymbol{h}\|_2 \ge \frac{\sigma'_{n+1}}{(n+1)K} \tag{4}$$

The estimated distance between two roots is $\frac{\sigma'_{n+1}}{(n+1)K}$. In other words, there is only one real root isolated in the ball centered at $\boldsymbol{z}_1$ with radius $\frac{\sigma'_{n+1}}{(n+1)K}$. This ball is called a *root isolation* of $\boldsymbol{z}_1$. Abusing terminology sometimes we will use "root isolation" to mean the radius of the ball.

Therefore, we have the following lemma.

LEMMA 2.2. *[Root Isolation Lemma] Let* σ'_{n+1} *be the smallest singular value of the Jacobian of* g *at* $\boldsymbol{z}_1$. *Then* $\frac{\sigma'_{n+1}}{(n+1)K(g)}$ *is a root isolation of* $\boldsymbol{z}_1$. *Moreover, if there are only* $k+1$ *nonlinear equations in* g, *then a root isolation is* $\frac{\sigma'_{n+1}}{K(g)\sqrt{(k+1)(n+1)}}$.

Obviously the root isolation is often underestimated. For a better lower bound on root isolations, we consider a rescaling of the system g. The upper bound matrix for $\nabla \mathcal{J}_{ij}$ is denoted by $\mathfrak{G} = \begin{pmatrix} \mathfrak{g}_1 \\ \vdots \\ \mathfrak{g}_{n+1} \end{pmatrix}$ where $\mathfrak{g}_i$ is a row vector.

Let $\theta_i = \|\mathfrak{g}_i\|_\infty$. Then $K(g) = \max\{\theta_1, ..., \theta_{n+1}\}$. Define

$$\tilde{g} = \begin{pmatrix} 1/\theta_1 & & \\ & \ddots & \\ & & 1/\theta_{n+1} \end{pmatrix} \cdot g$$

Then the new system is called a *normalized system* for g and $K(\tilde{g}) = 1$.

Interestingly we will show that $\tilde{g}$ has a larger estimate of root isolation.

PROPOSITION 2.3. *Let σ'_{n+1} and $\tilde{\sigma}_{n+1}$ be the smallest singular value at $\boldsymbol{z}_1$ of the Jacobian of g and its normalized system $\tilde{g}$ respectively. Then*

$$\frac{\sigma'_{n+1}}{K(g)} \leq \frac{\tilde{\sigma}_{n+1}}{K(\tilde{g})}$$

PROOF. Consider $K\tilde{g} = \begin{pmatrix} K/\theta_1 & & \\ & \ddots & \\ & & K/\theta_{n+1} \end{pmatrix} \cdot g$, where $K = K(g)$. The elements of the diagonal matrix Λ are greater than or equal to one.

Let $\mathcal{J}$ be the Jacobian of g at $\boldsymbol{z}_1$. Then $\Lambda \cdot \mathcal{J}$ is the Jacobian of $K\tilde{g}$ at $\boldsymbol{z}_1$. By the definition, there exists a unit vector x_0 such that

$$\sigma'_{n+1} = \min_{\|x\|=1} \|\mathcal{J}x\|_2 = \|\mathcal{J}x_0\|_2$$

Similarly, there exists a unit vector y_0 such that

$$K\tilde{\sigma}_{n+1} = \min_{\|y\|=1} \|\Lambda\mathcal{J}y\|_2 = \|\Lambda\mathcal{J}y_0\|_2$$

Thus, $K\tilde{\sigma}_{n+1} = \|\Lambda\mathcal{J}y_0\|_2 \geq \|\mathcal{J}y_0\|_2 \geq \|\mathcal{J}x_0\|_2 = \sigma_{n+1}$. □

EXAMPLE 2.2. *Reconsider the previous example $g = \{u^4y^2 + u^4z^2 - 8\,u^3x^3 + 12\,u^2x^4 - 6\,ux^5 + x^6, uz - y^2, x^2 + y^2 + z^2 + u^2 - 1, z - 0.01\}$ with $K = 144.2$. It is easy to know $F = f(x, y, 0.01, u)$ has 4 real roots when $z = 0.01$. we want to estimate the distance of the root $p = (u = 0.988, y = -0.0994, x = 0.114)$ to other roots.*

The smallest singular value of the Jacobian at p is 0.141. *Thus, we obtain a root isolation is* 0.000345.

On the other hand, since$\{\theta_1 = 144.23, \theta_2 = 2, \theta_3 = 2, \theta_4 = 1\}$, we construct $\tilde{g} = \{g_i/\theta_i\}$ and the resulting root isolation is 0.000546 *for the normalized system.*

The condition number type of root isolation was given by Blum, Cucker, Shub and Smale and it was improved by Jean-Pierre Dedieu in [9]. To illustrate the difference, we consider the following example.

EXAMPLE 2.3. *Let f is a linear homogenous system with $n + 1$ variables and n equations. Suppose all the complex coefficients have norm one and the 2-norm of the solution x is also one and the smallest singular value of the Jacobian is σ_n. Then the norm of the homogenous system f defined in [4] is $\sqrt{n(n+1)}$ and consequently $sep_{proj}(f, x) \geq \sigma_n/\sqrt{n(n+1)}$ in the sense of "angle distance" $d_p(x, y)$ in Equation (2.4) of [9].*

To use our result, let $g = \{f, (\sum x_i^2 - 1)/2\}$. So $K(g) = 1$ and by Lemma 2.2 a root isolation is $\sigma_n/\sqrt{n+1}$ since there is only one nonlinear equation. Here the distance is in the sense of the standard 2-norm. Note that for any two points x and y on the sphere $\sum x_i^2 - 1 = 0$ we have $2\sin\frac{\theta}{2} = \|x - y\| \geq d_p(x, y) = \sin\theta$. They are very close if the angle θ is small.

#	n	d	dist	est_{cond}	est_{sing}
1	3	3	0.000559	0.525e-4	1.11e-4
2	3	3	0.616e-4	1.52e-9	6.85e-6
3	3	4	2	3.02e-5	1.64e-5
4	3	4	0.121e-4	8.4e-7	2.71e-7
5	4	4	0.0141	1.56e-7	7.47e-4
6	4	4	0.000685	4.85e-6	1.33e-6
7	3	4	0.229	5.17e-4	0.005
8	4	2	0.516	0.0164	0.039
9	5	2	0.167	0.0113	0.0098
10	7	2	0.177	0.0042	0.0052
11	9	2	0.139	0.0022	0.0016
12	10	2	0.321	0.0018	0.0029

Table 1: Root Isolations: n and d are number of variables and maximum degree respectively. Here "dist" is the minimum distance to other roots. The quantities est_{cond} and est_{sing} are the root isolations obtained by the condition number method and our singular value method respectively where est_{cond} has been converted to the 2-norm.

REMARK 2.4. *Our root isolation method only needs local information at this point. Proposition 2.1 giving a more refined "measure" of a polynomial than traditional polynomial "height" or "norm". The upper bound K only depends on system itself, so we only need to compute K once. For the system $g = \{\bar{f}, (\sum x_i^2 - 1)/2\}$, its K is greater than or equal to* 1. *Thus, even if the θ_i corresponding to g_i is less than* 1, *the result is still valid even if we do not rescale g_i. This fact will be used in the later sections.*

We compare our root isolation method with the condition number formula given by Dedieu Theorem 2.2 in [9], on 12 random sparse square polynomial systems. For each system, we choose one root at random, then find the closest root to it to produce the true distance. Applying the condition number formula in Theorem 2 of [9] to $\bar{f}$ and our singular value formula in Lemma 2.2 to $\{\bar{f}, (\sum x_i^2 - 1)/2\}$ separately we show their root isolations in Table 1.

The table indicates that both methods give correct root isolations and our method gives better estimates in 7 examples out of 12. In the rest 5 examples, although $est_{cond} > est_{sing}$ they have the same order of magnitude. For some systems, e.g. $\#2, \#5, \#7$, our method performs very well.

2.2 Direction Determination

We aim to find a "good" path from an ill-conditioned point on a positive dimensional component leading to a well-conditioned region.

Previously, we have shown that root isolations have a close relation with the smallest singular value. Naturally, the key

question to be answered in this section is to find a good direction in a higher (> 1) dimensional tangent spaces in order to increase the smallest singular values along the path on the component. Such a path should lead to a well-conditioned point on the perturbed component which allows us to find a point on the original system with higher accuracy by projection.

Let $f = \{f_1, ..., f_k\}$ be a polynomial system with n variables. Let $\mathcal{J}$ be the Jacobian of a system f at $\boldsymbol{z}_0$ with $\mathcal{J} = U\Sigma V^t$. Let $A = \mathcal{J}\mathcal{J}^t = U\Sigma^2 U^t$. Thus, $A \cdot u = \sigma_k^2 u = \sigma u$ and $u^t u = 1$ where σ_k and u are the smallest singular value and the corresponding singular vector in U.

Imposing a small perturbation on the equations above yields

$$(A+\Delta A)(u+\Delta u) = (\sigma+\Delta\sigma)(u+\Delta u) \text{ and } (u+\Delta u)^t(u+\Delta u) = 1 \tag{5}$$

Ignoring the higher order terms, we have

$$\begin{aligned} Au + \Delta A u + A\Delta u &= \sigma u + \Delta\sigma u + \sigma\Delta u \\ u^t\ \Delta u &= 0 \end{aligned}$$

and

$$\Delta A u = \big((\mathcal{J}+\Delta\mathcal{J})(\mathcal{J}^t+\Delta\mathcal{J}^t) - A\big)\, u = (\Delta\mathcal{J}\mathcal{J}^t + \mathcal{J}\Delta\mathcal{J}^t)u \tag{6}$$

where $\Delta\mathcal{J} = \sum_i \frac{\partial \mathcal{J}}{\partial x_i}\Delta x_i$. Thus, it gives an $(m+1+m) \times (1+m+n)$ homogenous linear system:

$$\begin{pmatrix} u & (\sigma I - A) & -A_1 & \cdots & -A_n \\ & u^t & & & \\ & & & \mathcal{J} & \end{pmatrix} \cdot \begin{pmatrix} \Delta\sigma \\ \Delta u \\ \Delta x_1 \\ \vdots \\ \Delta x_n \end{pmatrix} = 0 \tag{7}$$

where $A_i = (\frac{\partial \mathcal{J}}{\partial x_i}\mathcal{J}^t + \mathcal{J}\frac{\partial \mathcal{J}}{\partial x_i}^t)u$.

LEMMA 2.5. *If the smallest singular value σ_k of $\mathcal{J}$ is strictly less than other singular values. Then the inverse of* $\begin{pmatrix} u & (\sigma I - A) \\ 0 & u^t \end{pmatrix}$ *is* $\begin{pmatrix} u^t & 0 \\ B & u \end{pmatrix}$, *where*

$$B = U \begin{pmatrix} 1/(\sigma-\sigma_1^2) & & & \\ & \ddots & & \\ & & 1/(\sigma-\sigma_{m-1}^2) & \\ & & & 0 \end{pmatrix} U^t$$

where $\sigma = \sigma_k^2$.

PROOF. It is straightforward to verify that $Bu = 0$ and $u^t(\sigma I - A) = 0$. Now we will show that $B(\sigma I - A) + uu^t = I$.

Consider $(B(\sigma I - A) + uu^t)UU^t$ which is equal to $\sigma BUU^t - BAUU^t + uu^tUU^t$.

By substituting the SVDs of A and B, we have

$$U\Sigma_1 U^t - U\Sigma_2 U^t + U\Sigma_3 U^t = I$$

where $\Sigma_1 = \text{diag}\left(\frac{\sigma}{\sigma-\sigma_1^2}, \cdots, \frac{\sigma}{\sigma-\sigma_{m-1}^2}, 0\right)$,

$\Sigma_2 = \text{diag}\left(\frac{\sigma_1^2}{\sigma-\sigma_1^2}, \cdots, \frac{\sigma_{m-1}^2}{\sigma-\sigma_{m-1}^2}, 0\right)$ and $\Sigma_3 = \text{diag}(0, \cdots, 0, 1)$.

□

Equation (7) multiplied by $\begin{pmatrix} u^t & & \\ B & u & \\ & & I \end{pmatrix}$ gives

$$\Delta\sigma = u^t A_1 \Delta x_1 + \cdots + u^t A_n \Delta x_n. \tag{8}$$

Suppose the step size is δ. A good direction aiming to maximize $\Delta\sigma$ leads to the following optimization problem:

$$\begin{aligned} \max \quad & c \cdot \Delta x \\ \text{s.t.} \quad & \mathcal{J} \cdot \Delta x = 0 \\ & \|\Delta x\| = \delta \end{aligned} \tag{9}$$

where $c = (u^t A_1, ..., u^t A_n)$ and $\Delta x = (\Delta x_1, ..., \Delta x_n)^t$.

Let $H = I - \mathcal{J}^t(\mathcal{J}^t)^\dagger$. It is easy to check that $H^t = H$, $H^2 = H$, $H\mathcal{J}^t = 0$ and $\mathcal{J}^\perp H = \mathcal{J}^\perp$, where $\mathcal{J}^\perp$ is the orthogonal complement of $\mathcal{J}$. Thus, Δx belongs to the space spanned by H. And $H(c^t - Hc^t) = 0$ implies that Hc^t is the projection of c in $\langle H \rangle$ and it will maximize the inner product $c \cdot \Delta x$. In addition, the eigenvalues H are $\{0, 1\}$ which means that H is a semi-positive definite matrix and $Hc^t = 0$ if and only if $c \in \langle \mathcal{J} \rangle$. Therefore, we have the following result.

LEMMA 2.6. *Let u be the singular vector corresponding to the smallest singular value, and $A_i = (\frac{\partial \mathcal{J}}{\partial x_i}\mathcal{J}^t + \mathcal{J}\frac{\partial \mathcal{J}}{\partial x_i}^t)u$ and $c = (u^t A_1, ..., u^t A_n)$. If $c \notin \langle \mathcal{J} \rangle$, then the solution to the optimization problem (9) is*

$$\Delta x = \frac{Hc^t}{\sqrt{cHc^t}}\,\delta \quad \textit{and} \quad \Delta\sigma = \sqrt{cHc^t}\,\delta \tag{10}$$

where $H = I - \mathcal{J}^t(\mathcal{J}^t)^\dagger$.

We require that the Jacobian at $\boldsymbol{z}_0$ satisfies "strictly less than" condition in Lemma 2.5 and "$c \notin \langle \mathcal{J} \rangle$" condition in Lemma 2.6. Actually, these conditions hold with high probability because of the randomness in Equation (3).

EXAMPLE 2.4. *Again we use the implicit curve $f = \{y^2 + z^2 - (2\,x - x^2)^3 - 0.0012\}$ to illustrate our method. We focus on the real variety of f near the origin. Suppose we have a starting point $p = (x = -0.02, y = 0.03, z = 0.015)$. The singular value at p is* 0.068. *Fix the step size to be* 0.02. *By Lemma 2.6, the direction for the next step is the unit vector* $(.988, .138, 0.0689)$ *and the new point is* $(-0.00072, 0.031, 0.016)$. *From this point, we can track to additional point on this component by the same idea.*

We use Table 2 to show the changes of the singular values in the first 10 *steps. It indicates that our method can find a new point* $(0.14, 0.12, 0.06)$ *which has better numerical behavior.*

2.3 Step Size Control

In the previous example, we fixed the step size during the path tracking. However, it may lead to a "path jumping" if there is another component nearby. So we discuss an adaptive step size strategy in the section.

For a given system $f(x_1, ..., x_n) = \{f_1, ..., f_k\}$ satisfying the regularity conditions, we homogenize f and suppose its normalized system is

$$g(x_0, x_1, ..., x_n) = \left\{g_1, ..., g_k, \left(\sum x_i^2 - 1\right)/2\right\}$$

with a point $\boldsymbol{z}_0$ on the hyper-sphere $S(0; 1)$. Suppose the dimension $m = n - k$ is greater than one.

#	σ	p
1	0.069	(-0.00072, 0.031, 0.016)
2	0.071	(0.019,0.032,0.016)
3	0.087	(0.038, 0.036, 0.018)
4	0.12	(0.055,0.044,0.022)
5	0.16	(0.071, 0.055, 0.027)
6	0.21	(0.086,0.067,0.033)
7	0.26	(0.099, 0.08, 0.04)
8	0.32	(0.11,0.093,0.047)
9	0.38	(0.13, 0.11, 0.053)
10	0.44	(0.14,0.12,0.06)

Table 2: Our path-tracking technique with fixed step size 0.02 leads to a well-conditioned new point. The column σ gives singular values at each point p.

By the result of previous section, the direction is Hc^t which is in the tangent space. Let S be the orthogonal complement of the vector Hc^t. It is easy to show that the space spanned by $\mathcal{J}$ of dimension $k+1$ is contained in the space S of dimension n. Let $S = \mathcal{J} \oplus T$ where T consists of $m-1$ orthogonal vectors (represented by an $(m-1)\times(n+1)$ matrix) with unit 1-norm. Thus, at $\boldsymbol{z}_0$, tracking in the tangent direction Hc^t is equivalent to tracking the 1-dimensional system $G = \{g, Tx - T\boldsymbol{z}_0\}$. Due to $K = K(G) = 1$ we do not rescale the new equations $Tx - T\boldsymbol{z}_0$ (see Remark 2.4).

Interestingly, the smallest singular value of $\mathcal{J}(G)$ is the same as that of $\mathcal{J}(g)$ at $\boldsymbol{z}_0$.

PROPOSITION 2.7. *Suppose A is an $m \times n$ matrix with smallest singular value $0 < \sigma_m < 1$. Suppose a matrix $B_{t\times n}$ satisfies $AB^t = 0$, $BB^t = I$ and $m + t \le n$. Then the smallest singular value of the matrix $\begin{pmatrix} A \\ B \end{pmatrix}$ is σ_m.*

PROOF. Let the singular value decomposition of A be $U_{m\times m}\begin{pmatrix} \Sigma_{m\times m} & 0 \end{pmatrix}\begin{pmatrix} V_1^t \\ V_2^t \end{pmatrix}$. Thus, V_2^t is the orthogonal complement of A, and consequently B spans a subspace of V_2^t. Let $\langle V_2^t\rangle = \langle B\rangle \oplus \langle C\rangle$, where C is an orthogonal basis. Then we have the following SVD

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} U & \\ & I \end{pmatrix}\begin{pmatrix} \Sigma & 0 \\ I & 0 \end{pmatrix}\begin{pmatrix} V_1^t \\ B \\ C \end{pmatrix}$$

So the new singular values are 1 and the smallest one is still σ_m. □

Let the Jacobian of G be $\mathcal{J}$. Applying the mean value theorem to each element of $\mathcal{J}$ on $S(0;1)$ yields

$$J_{ij}(\boldsymbol{y}) = J_{ij}(\boldsymbol{x}) + \nabla J_{ij}(\zeta_{ij})\cdot(\boldsymbol{y}-\boldsymbol{x}),\ \forall \boldsymbol{x},\boldsymbol{y} \in S(0;1) \quad (11)$$

where $\zeta_{ij} = \boldsymbol{x} + \kappa_{ij}(\boldsymbol{y}-\boldsymbol{x})$ and $\|\nabla J_{ij}(\zeta_{ij})\|_2 \le K = 1$ for $i \le k+1$, $\|\nabla J_{ij}(\zeta_{ij})\|_2 = 0$ for $i > k+1$.

Denote $\left(\nabla J_{ij}(\zeta_{ij})\cdot(\boldsymbol{y}-\boldsymbol{x})\right)_{n\times(n+1)} = \triangle\mathcal{J}$, then we have

$$\begin{aligned} \|\triangle\mathcal{J}\|_2 &= \left\|\left(|\nabla J_{ij}(\zeta_{ij})\cdot(\boldsymbol{y}-\boldsymbol{x})|\right)_{n\times(n+1)}\right\|_2 \\ &\le \left\|\left(\|\nabla J_{ij}(\zeta_{ij})\|_2\|\boldsymbol{y}-\boldsymbol{x}\|_2\right)_{n\times(n+1)}\right\|_2 \\ &\le \sqrt{(k+1)(n+1)}\|\boldsymbol{y}-\boldsymbol{x}\|_2 \end{aligned} \quad (12)$$

Since G is a one dimensional system, naturally the direction is the tangent of the curve at $\boldsymbol{z}_0$. In the *predictor* step, we move distance δ in this direction and in the *corrector* step, the Newton method converges to $\boldsymbol{z}_1$. Next we will estimate the distance between $\boldsymbol{z}_0$ and $\boldsymbol{z}_1$.

Let $\mathcal{J}(\boldsymbol{z}_0) = U\Sigma V^t$. Introduce the new coordinates $u = V^t(x-\boldsymbol{z}_0)$ with the origin at $\boldsymbol{z}_0$ (u_n is the tangent direction). The differential of G at $\boldsymbol{z}_0$ gives

$$0 = dG(x) = \mathcal{J}(\boldsymbol{z}_0)dx = U\Sigma V^t V du = U\Sigma du \quad (13)$$

In the region close to $\boldsymbol{z}_0$, we have

$$0 = (\mathcal{J}(\boldsymbol{z}_0) + \triangle\mathcal{J})dx = U\Sigma du + \triangle\mathcal{J}V du$$

and consequently

$$(I,0)du + \Sigma^{-1}U^{-1}\triangle\mathcal{J}V du = 0 \quad (14)$$

where $\Sigma^{-1} = \begin{pmatrix} 1/\sigma_1 & & \\ & \ddots & \\ & & 1/\sigma_n \end{pmatrix}$.

Denote $\Sigma^{-1}U^{-1}\triangle\mathcal{J}V = (E_{n\times n}, b_{n\times 1})$ and the 2-norm of $\|E\|_2$ and $\|b\|$ are less than or equal to $\|u\|\sqrt{(k+1)(n+1)}/\sigma_n$ by Equation (12).

Consider the coordinate u_n as the independent variable to describe the curve locally. And split $du = (du_0, ..., du_{n-1}, du_n) = (d\vec{u}, du_n)$, where $\vec{u}$ is a vector function of u_n. Then Equation (14) becomes

$$(I+E)d\vec{u} = -b\ du_n \Rightarrow \|d\vec{u}\| \le \frac{\|u\|\sqrt{(k+1)(n+1)}}{\sigma_n - \|u\|\sqrt{(k+1)(n+1)}}\ |du_n| \quad (15)$$

due to the fact $\|(I+E)^{-1}\|_2 \le 1/(1-\|E\|_2)$ when $\|E\|_2 < 1$.

Denote u_n and $\|\vec{u}\|$ by x and $y(x)$ respectively. And let $\mu = \sqrt{(k+1)(n+1)}$. Equation (15) becomes

$$y' \le \frac{\mu\sqrt{x^2+y^2}}{\sigma_n - \mu\sqrt{x^2+y^2}} \text{ and } y(0) = 0 \quad (16)$$

Since when x is sufficiently small, $\sigma_n - \mu\sqrt{x^2+y^2} > 0$ and $x > 0, y > 0$ leads to $\sqrt{x^2+y^2} < x + y$. We obtain

$$0 \le y' < \frac{\mu(x+y)}{\sigma_n - \mu(x+y)} \text{ and } y(0) = 0 \quad (17)$$

The solution to the IVP $z' = \frac{\mu(x+z)}{\sigma_n - \mu(x+z)}$ and $z(0) = 0$ is easily determined as:

$$z = \frac{\sigma_n}{\mu} - x - \frac{\sqrt{\sigma_n^2 - 2\sigma_n\mu\ x}}{\mu} \text{ when } x \le \frac{\sigma_n}{2\mu}. \quad (18)$$

Thus, $y(x)$ is bounded above by $z(x)$ when $x \le \frac{\sigma_n}{2\mu}$.

We set the step size $\delta = \frac{\sigma_n}{2\rho\mu}$, $\rho \ge 1$ and substitute $x = \delta$ in Equation (18). Then

$$\|\boldsymbol{z}_1 - \boldsymbol{z}_0\| < \sqrt{z^2+x^2} = \omega\ \cdot\delta \quad (19)$$

where $\omega = \sqrt{2\,(2\,\rho - 1)\left(2\,\rho - 2\,\sqrt{\rho\ (\rho-1)} - 1\right)}$.

Now we discuss how to choose the step size and how to determine if a jump happens.

By using the same notation, we track the 1-dimensional system $G = \{g, Tx - T\boldsymbol{z}_0\}$ at $\boldsymbol{z}_0$. In order to use the Root Isolation Lemma, we can reduce the system to dimension zero by cutting the curve of G with a hyper-plane which is perpendicular to the tangent Hc^t at distance δ to $\boldsymbol{z}_0$. The equation for this hyper-plane is $L = (cH(x-\boldsymbol{z}_0) - cHc^t\delta)/\sqrt{cHc^t}$.

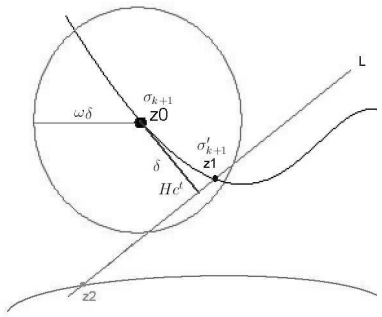

Figure 1: The point z_0 moves δ in the direction Hc^t and $\text{dist}(\boldsymbol{z}_1, \boldsymbol{z}_0) < \omega\delta$

THEOREM 2.8. *Let $g = \{g_1, ..., g_k, (\sum x_i^2 - 1)/2\}$ be a normalized system. Let $\delta = \frac{\sigma_{k+1}}{2\rho\mu}$ where σ_{k+1} is the smallest singular value of the Jacobian of g at $\boldsymbol{z}_0$ and ρ satisfies $2\rho > 3\omega$. We move $\boldsymbol{z}_0$ in the direction Hc^t at distance δ and apply the Newton Iteration to the zero dimensional system $\{G, L\}$. Assume the Newton Iteration converges to $\boldsymbol{z}_1$. Then $\boldsymbol{z}_1$ is on the same component with $\boldsymbol{z}_0$ if and only if* $\text{dist}(\boldsymbol{z}_1, \boldsymbol{z}_0) < \omega\delta$.

PROOF. "⇒" is simply due to Equation (19).

Now we show "⇐".

Let s_{n+1} and s'_{n+1} be the smallest singular values of the Jacobian of the square system $\{G, L\}$ at $\boldsymbol{z}_0$ and $\boldsymbol{z}_1$ respectively. By Proposition 2.7, $\sigma_{k+1} = s_{n+1}$.

We estimate the s'_{n+1} by Weyl's theorem [19]:

$$|s'_{n+1} - \sigma_{k+1}| = |s'_{n+1} - s_{n+1}| \leq \|\triangle\mathcal{J}\|_2 < \mu\omega\delta = \frac{\omega}{2\rho}\sigma_{k+1}$$

Thus, $s'_{n+1} > (1 - \frac{\omega}{2\rho})\sigma_{k+1}$. By the Root Isolation Lemma, we have

$$\text{dist}(\boldsymbol{z}_1, \boldsymbol{z}_2) > s'_{n+1}/\mu > (1 - \frac{\omega}{2\rho})\sigma_{k+1}/\mu \ .$$

Since $2\rho > 3\omega \Leftrightarrow (1 - \frac{\omega}{2\rho})\sigma_{k+1}/\mu > 2\omega\delta$, by our assumption, it leads to $\text{dist}(\boldsymbol{z}_1, \boldsymbol{z}_2) > 2\omega\delta$.

Combined with the condition $\text{dist}(\boldsymbol{z}_1, \boldsymbol{z}_0) < \omega\delta$, we have

$$\text{dist}(\boldsymbol{z}_2, \boldsymbol{z}_0) > \omega\delta$$

Thus, the distance of $\boldsymbol{z}_0$ to other roots of $\{G, L\}$, which are the intersections of L and other components, must be greater than $\omega\delta$ as shown in Figure 1. So there is only one root within distance $\omega\delta$ and it must be the one on the same component as $\boldsymbol{z}_0$. □

REMARK 2.9. *The condition $2\rho > 3\omega$ is a theoretical result, and in practice we may choose ρ as small as possible e.g. $\rho = 1.6$ and then $\omega = 1.0285$. Because we do not know the convergence region of the Newton iteration, we may obtain $\boldsymbol{z}_2$ with* $\text{dist}(\boldsymbol{z}_1, \boldsymbol{z}_0) > \omega\delta$. *In that case, we will decrease the step size. Eventually,* $\text{dist}(\boldsymbol{z}_1, \boldsymbol{z}_0)$ *will be less than $\omega\delta$ when δ is small enough.*

3. ALGORITHM

Following [20], we want to find well-conditioned real witness points of the system

$$F = (F_1, F_2, ..., F_k) = 0 \tag{20}$$

of k polynomials from $\mathbb{R}[x_1, ..., x_n]$ which satisfy the assumptions A_1 and A_2.

To regularize the potential singularities of the given system F, we first perturb the input to yield a nearby system f. By using the method in [20], we first find real witness points of f.

If the Jacobian of some point is near rank deficiency, then we apply the new algorithm FOLLOWCOMP combined with optimal direction and step size control techniques to move the point along $V_{\mathbb{R}}(f)$ until the condition of the Jacobian is tolerable. Finally, we project this new point back $V_{\mathbb{R}}(F)$.

Since we need to do the computation in projective space for better numerical stability, we introduce two transformations:

$$\texttt{H} : \texttt{x} \in \text{ affine space } \rightarrow \texttt{p} \in \text{ projective space}$$

where $p_0 = 1/\sqrt{1 + \|x\|^2}$, and $p_i = x_i/\sqrt{1 + \|x\|^2}$ for $i = 1, ..., n$.

$$\texttt{DH} : \texttt{p} \in \text{ projective space } \rightarrow \texttt{x} \in \text{ affine space}$$

where $x_i = p_i/p_0$ for $i = 1, ..., n$.

Next we describe our new algorithm FOLLOWCOMP.

Algorithm FOLLOWCOMP

Input: f: a perturbed system given by $\{f_1, ..., f_k\} \subset \mathbb{R}[x_1, \ldots, x_n]$
x = an approximate solution of $f = 0$
κ = threshold used in path tracking to control smallest singular value

1. Construct the system $g = \{\bar{f}, \sum x_i^2 - 1\}$ and $p = \texttt{H(x)}$
2. Compute θ_i of g_i and construct $\tilde{g}$ by Proposition 2.3
3. Find σ_{k+1} of $\mathcal{J}_p(\tilde{g})$ by SVD, if $\sigma_{k+1} > \kappa$ then return DH(p)
4. Let $\delta = \frac{\sigma_{k+1}}{3.2\mu}$ where $\mu = \sqrt{(k+1)(n+1)}$
5. Construct the square system $\{G, L\}$
6. Update $p = p + Hc^t\delta$ and let $q = Newton(\{G, L\}, p)$
7. If $\|q - p\| > 1.0285\delta$ then let $\delta = \delta/2$ and go to step 6
   Else let $p = q$ and go to step 3

Output: A well-conditioned point on $V_{\mathbb{R}}(f)$

Usually the poor conditioning region appears close to a singular component or point (e.g. at the intersection of two irreducible components).

Since the dimension of the singular set is lower than the dimension of regular set, the likelihood of leaving this poor conditioning region is quite large. However, the optimal direction technique can not guarantee the termination of the algorithm FOLLOWCOMP for any threshold κ. Since it is still a local method, it may unfortunately oscillate near a local maximum of σ_{k+1} with $\sigma_{k+1} < \kappa$. In practice, we can detect such oscillations and terminate the procedure. Then choose a new random vector in Equation (3) to produce other points.

Although we can not verify the regularity assumption in advance, it can be detected if the perturbed system has no real solutions or the Jacobian is always near rank-deficiency during path tracking.

4. CONCLUSIONS

In this paper we give a preliminary analysis of the numerical stability of our critical point methods. Our experiments applied these methods to the problem of finding at

least one well conditioned witness point on each connected components. Our methods use perturbation, path tracking and projection techniques together with an optimal-direction strategy and an adaptive step size control strategy. These strategies were successful on our test problems for real solutions. As by-products, they can be used in numerical algebraic geometry for computing well-conditioned complex witness points as well.

Future research includes loosening the assumptions A_1 and A_2, to address more degenerate cases. Such research includes methods for deflating higher multiplicity components, and also addressing problems with equations which are sums of squares. We will apply the new strategies to other methods e.g. the closest method to ours, that of Hauenstein [10].

In the paper we assume the convergence of Newton iteration. Other approaches, such as certified numerical homotopy tracking by Beltran and Leykin [3], will be needed to give a rigorous guarantee of convergence to such well-conditioned points.

5. REFERENCES

[1] S. Basu, R. Pollack and M-F Roy, Algorithms in real algebraic geometry. 2nd edition. Algorithms and Computation in Math., 10. Springer-Verlag, 2006.

[2] D.J. Bates, J.D. Hauenstein, A.J. Sommese, and C.W. Wampler, Numerically Solving Polynomial Systems with the Software Package Bertini. In preparation. To be published by SIAM, 2013.

[3] Carlos Beltran, Anton Leykin. Robust Certified Numerical Homotopy Tracking. Foundations of Computational Mathematics April 2013, Volume 13, Issue 2, pp 253-295.

[4] L. Blum, F. Cucker, M. Shub, and S. Samle, Complexity and Real Computation. Springer-Verlag 1998.

[5] G.M Besana, S. DiRocco, J.D. Hauenstein, A.J. Sommese, and C.W. Wampler. Cell decomposition of almost smooth real algebraic surfaces. Num. Algorithms, DOI:10.1007/s11075-012-9646-y, (published online Sept 28, 2012).

[6] C. Chen, J.H. Davenport, J.P. May, M.M. Maza, B. Xia and R. Xiao. *Triangular decomposition of semi-algebraic systems* Journal of Symbolic Computation Vol49-0, pp 3-26, 2013.

[7] G. Collins, Quantifier elimination for real closed fields by cylindrical algebraic decomposition. Springer Lec. Notes Comp. Sci. v33. 515-532, 1975.

[8] J.H. Davenport and J. Heintz, Real quantifier elimination is doubly exponential, J. Symbolic Comp., 5:29-35, 1988.

[9] Jean-Pierre Dedieu, Estimations for the Separation Number of a Polynomial System. J. Symbolic Comp., 24:683-693, 1997.

[10] J. Hauenstein, Numerically computing real points on algebraic sets. Acta Appl. Math. DOI:10.1007/s10440-012-9782-3, (published online September 27, 2012).

[11] H. Hong. Improvement in CAD-Based Quantifer Elimination. Ph.D. thesis, the Ohio State University, Columbus, Ohio, 1990.

[12] J.B. Lasserre, M. Laurent and P. Rostalski. A prolongation-projection algorithm for computing the finite real variety of an ideal. Theoret. Comput. Sci., 410(27-29), 2685-2700, 2009.

[13] Y. Lu, Finding all real solutions of polynomial systems, Ph.D Thesis, University of Notre Dame, 2006. Results of this thesis appear in: (with D.J. Bates, A.J. Sommese, and C.W. Wampler), Finding all real points of a complex curve, Contemporary Mathematics 448, 183–205, 2007.

[14] Y. Ma and L. Zhi. Computing Real Solutions of Polynomial Systems via Low-rank Moment Matrix Completion Proc. 2012 Internat. Symp. Symbolic Algebraic Comput. 249-256, 2012.

[15] F. Rouillier, M.-F. Roy, and M. Safey El Din. Finding at least one point in each connected component of a real algebraic set defined by a single equation. J. Complexity, 16 (4), 716-750, 2000.

[16] A. Seidenberg, A new decision method for elementary algebra, Ann. of Math. 60, 365-374, 1954.

[17] A.J. Sommese and C.W. Wampler. *The Numerical solution of systems of polynomials arising in engineering and science*. World Scientific Press, 2005.

[18] A. Tarski, A decision method for elementary algebra and geometry, Fund. Math. 17, 210-39, 1931.

[19] G. W. Stewart. Perturbation Theory for the Singular Value Decomposition. In SVD and signal processing, II: Algorithms, analysys and applications, pp 99–109, Elsevier, 1990.

[20] Wenyuan Wu and Greg Reid. Finding points on real solution components and applications to differential polynomial systems. ISSAC 2013: 339-346.

[21] Zhengfeng Yang, Lihong Zhi, and Yijun Zhu. Verfied error bounds for real solutions of positive-dimensional polynomial systems In ISSAC'2013 Proc. 2013 Internat. Symp. Symbolic Algebraic Comput. pp. 371-378.