Found Comput Math
DOI 10.1007/s10208-007-9017-6

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Complexity of Bezout's Theorem VI: Geodesics in the Condition (Number) Metric

Michael Shub

Received: 12 March 2007 / Final version received: 28 September 2007 / Accepted: 6 November 2007

Abstract We introduce a new complexity measure of a path of (problems, solutions) pairs in terms of the length of the path in the condition metric which we define in the article. The measure gives an upper bound for the number of Newton steps sufficient to approximate the path discretely starting from one end and thus produce an approximate zero for the endpoint. This motivates the study of short paths or geodesics in the condition metric.

Keywords Approximate zero · Homotopy method · Condition metric

Mathematics Subject Classification (2000) Primary 65H10 · 65H20

1 Introduction

In a series of papers we have studied the complexity of solving systems of homogeneous polynomial equations by applying Newton's method to a homotopy of (system, solution) pairs [7–11]. The latest word in this direction is [1, 2]. A key ingredient is an estimate of the number of Newton steps by the maximum condition number along the path multiplied by the length of the path of solutions. The main result of this paper is to show that the maximum condition number times the length may be replaced by

This work was partly supported by an NSERC Discovery Grant.

M. Shub (✉)
Department of Mathematics, University of Toronto, 40 St. George Street, Toronto, Ontario M5S 2E4, Canada
e-mail: michael.shub@utoronto.ca

the integral of the condition number times the length of the tangent vector to the path, Theorem 3. The result suggests using the condition number to define a Riemannian metric on the solution variety and to study the geodesics of this metric. Finding a geodesic is in itself not easy. So we do not immediately obtain a practical algorithm. Rather the study may help to understand in some systematic fashion the geometry of homotopy algorithms, especially those that attempt to avoid ill-conditioned problems. We note that recentering algorithms in linear programming theory may be seen as adaptively avoiding ill-conditioning. Nesterov and Todd [6] compare the central path of linear programming theory to geodesics in an appropriate metric. In [3] we begin studying distances in the condition metric.

2 Definitions and Theorems

We begin by recalling the context. For every positive integer $l \in \mathbb{N}$, let $H_l \subseteq \mathbb{C}[X_0, \dots, X_n]$ be the vector space of all homogeneous polynomials of degree l. For $(d) := (d_1, \dots, d_n) \in \mathbb{N}^n$, let $\mathcal{H}_{(d)} := \prod_{i=1}^n H_{d_i}$ be the set of all systems $f := (f_1, \dots, f_n)$ of homogeneous polynomials of respective degrees $\deg(f_i) = d_i$, $1 \le i \le n$. So $f : \mathbb{C}^{n+1} \to \mathbb{C}^n$. We denote by $D := \max\{d_i : 1 \le i \le n\}$ the maximum of the degrees.

The solution variety $\hat{V} \subset \mathcal{H}_{(d)} \times (\mathbb{C}^{n+1} \setminus \{0\})$ is the set of points $\{(f, x) | f(x) = 0\}$. Since the equations are homogeneous, for all $\lambda_1, \lambda_2 \in \mathbb{C} \setminus \{0\}$, $\lambda_1 f(\lambda_2 x) = 0$ if and only if $f(x) = 0$. So $\hat{V}$ defines a variety $V \subset \mathbb{P}(\mathcal{H}_{(d)}) \times \mathbb{P}(\mathbb{C}^{n+1})$ where $\mathbb{P}(\mathcal{H}_{(d)})$ and $\mathbb{P}(\mathbb{C}^{n+1})$ are the projective spaces corresponding to $\mathcal{H}_{(d)}$ and $\mathbb{C}^{n+1}$ respectively. $\hat{V}$ and V are smooth. We speak interchangeably of a path (f_t, ζ_t) in $\hat{V}$ and its projection (f_t, ζ_t) in V. Most quantities we define are defined on $\hat{V}$ but are constant on equivalence classes so are defined on V.

For $(g, x) \in \mathbb{P}(\mathcal{H}_{(d)}) \times \mathbb{P}(\mathbb{C}^{n+1})$ let

$$\mu_{\text{norm}}(g, x) = \|g\| \left\| Dg(x)|_{N_x}^{-1} \Delta\left(d_i^{\frac{1}{2}} \|x\|^{d_i - 1}\right) \right\|,$$

where $\|g\|$ is the unitarily invariant norm defined by the unitarily invariant Hermitian structure on $\mathcal{H}_{(d)}$ considered in [7] and sometimes called the Bombieri–Weyl or Kostlan Hermitian structure, $\|x\|$ is the standard norm in $\mathbb{C}^{n+1}$, N_x is the Hermitian complement of x, and $\Delta(a_i)$ for $a_i \in \mathbb{C}, i = 1 \dots n$ is the $n \times n$ diagonal matrix with ith diagonal entry a_i. If $Dg(x)|_{N_x}^{-1}$ does not exist we take $\mu_{\text{norm}} = \infty$. μ_{norm}, also called μ_{proj} in some of our papers, is a normalized version of the condition number which we usually denote by μ. We have also used various notions of distance in projective space. In this paper we use only the Riemannian distance inherited from the Hermitian structure on the vector space, i.e. the angle. We denote this distance by $d(.,.)$. The space and Hermitian structure we are considering will be clear from the context. In previous papers we have paid careful attention to the constants. In this paper we are more cavalier. We begin with an analysis of how the normalized condition number varies.

Proposition 1 *Given $\epsilon > 0$ there is a constant $C > 0$ such that if*

$$g \in \mathbb{P}(\mathcal{H}_{(d)}) \quad \textit{and} \quad \zeta, \eta \in \mathbb{P}(\mathbb{C}^{n+1})$$

with

$$d(\zeta,\eta) < \frac{C}{D^{3/2}\mu_{\text{norm}}(g,\zeta)}.$$

Then $\mu_{\text{norm}}(g,\eta) \le (1+\epsilon)\mu_{\text{norm}}(g,\zeta)$.

Proof Use Proposition 2.3 of [10]. In the notation of that proposition note that $u \le D^{3/2}\mu_{\text{norm}}(g,\eta)\cdot d(\eta,\zeta) < C$. $r_0 \sim d(\eta,\zeta)$, the $\eta(f,x)$ in the definitions of K is ≤ 1 and $\left(\frac{\|\eta\|}{\|\zeta\|}\right)^{D-1} \sim (1+d(\zeta,\eta))^{D-1} < e^c$. □

Proposition 2 *Given* $\epsilon > 0$ *there is a constant* $C > 0$ *such that if*

$$f, g \in \mathbb{P}(\mathcal{H}_{(d)}) \quad \textit{and} \quad \zeta \in \mathbb{P}(\mathbb{C}^{n+1})$$

and

$$d(f,g) < \frac{C}{D^{1/2}\mu_{\text{norm}}(f,\zeta)},$$

then

$$\mu_{\text{norm}}(g,\zeta) \le (1+\epsilon)\mu_{\text{norm}}(f,\zeta).$$

Proof By Proposition 5(b) of [7, Sect. I-3].

$$\begin{aligned}\mu_{\text{norm}}(g,\zeta) &\le \frac{\mu_{\text{norm}}(f,\zeta)(1+d(f,g))}{1-D^{1/2}d(f,g)\mu_{\text{norm}}(f,\zeta)} \\ &\le \frac{\mu_{\text{norm}}(f,\zeta)\left(1+\frac{C}{D^{1/2}\mu_{\text{norm}}(f,\zeta)}\right)}{1-C}.\end{aligned}$$

Recall that $\mu_{\text{norm}}(f,\zeta) \ge 1$. □

Theorem 1 *Given* $\epsilon > 0$ *there is a constant* $C > 0$ *such that if*

$$f, g \in \mathbb{P}(\mathcal{H}_{(d)}) \quad \textit{and} \quad \zeta, \eta \in \mathbb{P}(\mathbb{C}^{n+1})$$

and

$$d(f,g) < \frac{C}{D^{1/2}\mu_{\text{norm}}(f,\zeta)},$$

$$d(\zeta,\eta) < \frac{C}{D^{3/2}\mu_{\text{norm}}(f,\zeta)},$$

then

$$\frac{1}{1+\epsilon}\mu_{\text{norm}}(g,\eta) \le \mu_{\text{norm}}(f,\zeta) \le (1+\epsilon)\mu_{\text{norm}}(g,\eta).$$

Proof Apply Propositions 1 and 2 to prove the left hand inequality. Then given the left hand inequality apply Propositions 1 and 2 again to prove the right hand inequality applying the theorems with g, η in place of f, ζ. Adjust C and ϵ as necessary. □

The next proposition is useful for our Main Theorem 3.

Proposition 3 *Given $\epsilon > 0$ there is a $C > 0$ with the following property*:

Let (f_t, ζ_t) be a C^1 path in V for $t_0 \le t \le t_1$. Define $S_0 = t_0$ and S_i to be the first value of $t \le t_1$ such that

$$\int_{S_{i-1}}^{S_i} \big(\|\dot{f}_t\| + \|\dot{\zeta}_t\|\big) d_t = \frac{C}{D^{3/2}\mu_{\text{norm}}(f_{S_{i-1}}, \zeta_{S_{i-1}})} \quad \textit{or} \quad t_1.$$

Then $S_k = t_1$ for

$$k \le \max\left(1, \frac{(1+\epsilon)}{C} D^{3/2} \int_{t_0}^{t_1} \mu_{\text{norm}}(f_t, \zeta_t)\big(\|\dot{f}_t\| + \|\dot{\zeta}_t\|\big) dt\right)$$

and $\mu_{\text{norm}}(f_{t_1}, \zeta_{t_1}) \le (1+\epsilon)^k \mu_{\text{norm}}(f_{t_0}, \zeta_{t_0})$.

Proof

$$\begin{aligned}\int_{S_{i-1}}^{S_i} \mu_{\text{norm}}(f_t, \zeta_t)\big(\|\dot{f}_t\| + \|\dot{\zeta}_t\|\big) dt &\ge \frac{1}{(1+\epsilon)} \int_{S_{i-1}}^{S_i} \mu_{\text{norm}}(f_{S_{i-1}}, \zeta_{S_{i-1}})\big(\|\dot{f}_t\| + \|\dot{\zeta}_t\|\big) dt \\ &\ge \frac{\frac{1}{1+\epsilon} C}{D^{3/2}} \quad \text{if } S_i < t_1.\end{aligned}$$

Consequently

$$\int_{S_0}^{S_k} \mu_{\text{norm}}(f_t, \zeta_t)\big(\|\dot{f}_t\| + \|\dot{\zeta}_t\|\big) dt \ge \frac{(k-1)C}{(1+\epsilon)D^{3/2}}$$

and

$$(k-1)\frac{C}{(1+\epsilon)D^{3/2}} \le \int_{t_0}^{t_1} \mu_{\text{norm}}(f_{t_1} \zeta_t)\big(\|\dot{f}_t\| + \|\dot{\zeta}_t\|\big) dt,$$

so

$$k - 1 \le \frac{(1+\epsilon)D^{3/2}}{C} \int_{t_0}^{t_1} \mu_{\text{norm}}(f_{t_1} \zeta_t)\big(\|\dot{f}_t\| + \|\dot{\zeta}_t\|\big) dt.$$ □

Since we are working in the metric d we require an approximate zero theorem in this metric. First we prove a lemma. Recall the following quadratic polynomial

$$\psi(u) = 1 - 4u + 2u^2$$

and the definition of the projective Newton iteration

$$N_f(x) = x - (Df(x)|_{x^\perp})^{-1} f(x).$$

Here $x^{\perp}$ is the Hermitian complement to x. $N_f : \mathbb{P}(\mathbb{C}^{n+1}) \to \mathbb{P}(\mathbb{C}^{n+1})$ except that it fails to be defined where $(Df(x)|_{x^{\perp}})^{-1}$ does not exist. If $f(\zeta) = 0$ and $d(N_f^k(x), \zeta) \le \frac{1}{2^{2^k-1}} d(x, \zeta)$ for all positive integers k, then x is called an *approximate zero* of f with associated zero ζ.

Lemma 1 *Let* $u < \frac{3-\sqrt{7}}{4}$. *Let* $f \in P(\mathcal{H}_{(d)})$ *and* $\zeta \in P(\mathbb{C}^{n+1})$ *with* $f(\zeta) = 0$. *If* $d(x, \zeta) \le \frac{u}{D^{3/2}\mu_{\text{norm}}(f,\zeta)}$, *then*

$$d(N_f(x), \zeta) \le \frac{4u}{\psi(2u)} d(x, \zeta).$$

Proof In the range of angles under consideration Tan $d(x, \zeta) \le 2d(x, \zeta)$. So apply Lemma 1 of P263 of [5] to conclude that

$$\begin{aligned} d(N_f(x), \zeta) &\le \text{Tan } d(N_f(x), \zeta) \le \frac{2u}{\psi(2u)} \text{Tan } d(x, \zeta) \\ &\le \frac{4u}{\psi(2u)} d(x, \zeta). \end{aligned}$$

□

Now let u_0 solve the equation

$$\frac{4u_0}{\psi(2u_0)} = \frac{1}{2} \quad \text{or} \quad u_0 = \frac{16 - \sqrt{224}}{16} \sim 0.06458.$$

Theorem 2 (Approximate zero theorem) *Let* $f \in P(\mathcal{H}_{(d)})$, $\zeta \in P\ (\mathbb{C}^{n+1})$ *with* $f(\zeta) = 0$. *If* $d(x, \zeta) < \frac{u_0}{D^{3/2}\mu_{\text{norm}}(f,\zeta)}$ *then* $d(N_f^k(x), \zeta) \le \frac{1}{2^{2^k-1}} d(x, \zeta)$.

Proof By induction.

Suppose

$$\begin{aligned} d(N_f^k(x), \zeta) &\le \left(\frac{4u_0}{\psi(2u_0)}\right)^{2^k-1} d(x, \zeta) \\ &\le \frac{\left(\frac{4u_0}{\psi(2u_0)}\right)^{2^k-1} u_0}{D^{3/2}\mu_{\text{norm}}(f, \zeta)}. \end{aligned}$$

Let $u_1 = \left(\frac{4u_0}{\psi(2u_0)}\right)^{2^k-1} u_0$. Note $u_1 \le u_0$ and so $\psi(2u_1) \ge \psi(2u_0)$. By the lemma

$$\begin{aligned} d(N_f^{k+1}(x), \zeta) &\le \frac{4u_1}{\psi(2u_1)} d(N_f^k(x), \zeta) \\ &\le \frac{4u_1}{\psi(2u_0)} \cdot d(N_f^k(x), \zeta) \end{aligned}$$

$$\leq 4\frac{\left(\frac{4u_0}{\psi(2u_0)}\right)^{2^k-1}u_0}{\psi(2u_0)}\cdot\left(\frac{4u_0}{\psi(2u_0)}\right)^{2^k-1}d(x,\zeta)$$

$$=\left(\frac{4u_0}{\psi(2u_0)}\right)^{2^{k+1}-1}d(x,\zeta).\qquad\square$$

Let (f_t,ζ_t) be a (piecewise) C^1 path in V, $\frac{d}{dt}(f_t,\zeta_t)$ its tangent vector and $\|\frac{d}{dt}(f_t,\zeta_t)\|$ the length of its tangent vector.

Theorem 3 (Main theorem) *There is a constant $C_1>0$, such that: if (f_t,ζ_t) $t_0\leq t\leq t_1$ is a C^1 path in V, then*

$$C_1D^{3/2}\int_{t_0}^{t_1}\mu_{\mathrm{norm}}(f_t,\zeta_t)\left\|\frac{d}{dt}(f_t,\zeta_t)\right\|dt$$

steps of projective Newton method are sufficient to continue an approximate zero x_0 of f_{t_0} with associated zero ζ_0 to an approximate zero x_1 of f_{t_1} with associated zero ζ_1.

Proof Choose $C<\mu_0$ and ϵ small enough such that Theorems 1 and 2 apply, $u=2C(1+\epsilon)<\frac{3-\sqrt{7}}{4}$ and $\frac{4u}{\psi(2u)}<\frac{1}{2(1+\epsilon)}$. Hence, if $d(f,g)<\frac{C}{D^{1/2}\mu_{\mathrm{norm}}(f,\zeta)}$, $d(\zeta,\eta)\leq\frac{C}{D^{3/2}\mu_{\mathrm{norm}}(f,\zeta)}$, $f(\zeta)=0$, $g(\eta)=0$ and $d(x,\zeta)\leq\frac{C}{D^{3/2}\mu_{\mathrm{norm}}(f,\zeta)}$, then $d(N_g(x),\eta)\leq\frac{C}{D^{3/2}\mu_{\mathrm{norm}}(g,\eta)}$. So $N_g(x)$ is an approximate zero of g with associated zero η.

Now apply Proposition 1 to produce $S_0,\dots,S_k$ and x_0 such that

$$d(x_0,\zeta_{t_0})<\frac{C}{D^{3/2}\mu_{\mathrm{norm}}(f_{t_0},\zeta_{t_0})}.$$

Then $x_i=N_{f_{S_i}}(x_i-1)$ is an approximate zero of f_{S_i} with associated zero ζ_{S_i} and

$$d(x_i,\zeta_{S_i})<\frac{C}{D^{3/2}\mu_{\mathrm{norm}}(f_{S_i},\zeta_{S_i})}.\qquad\square$$

Corollary 1 *There is a constant $C_2>0$, such that: if (f_t,ζ_t) $t_0\leq t\leq t_1$ is a C^1 path in V, then*

$$C_2D^{3/2}\int_{t_0}^{t_1}\mu_{\mathrm{norm}}^2(f_t,\zeta_t)\|\dot{f}_t\|dt$$

steps of projective Newton method are sufficient to continue an approximate zero x_0 of f_{t_0} with associated zero ζ_0 to an approximate zero x_1 of f_{t_1} with associated zero ζ_1.

Proof $\|\dot{\zeta}_t\|\leq\mu_{\mathrm{norm}}(f_t,\zeta_t)\|\dot{f}_t\|$. $\square$

By comparison the results of the papers [7] and [10] give for the Theorem 3

$$C_3D^{3/2}\sup_t(\mu_{\mathrm{norm}}(f_t,\zeta_t))\int_{t_0}^{t_1}\left\|\frac{d}{dt}(\zeta_t)\right\|dt$$

steps of projective Newton's method and for the corollary

$$C_3 D^{3/2} \sup_t \big(\mu^2_{\text{norm}}(f_t, \zeta_t)\big) \int_{t_0}^{t_1} \|\dot{f}_t\| dt$$

steps of projective Newton's method for some constant C_3.

Theorem 3 suggests that if we wish to continue a solution $\zeta \epsilon P(\mathbb{C}^{n+1})$ of $f \epsilon P(\mathcal{H}_{(d)})$ to a solution $\eta \epsilon P(\mathbb{C}^{n+1})$ of $g \epsilon P(\mathcal{H}_{(d)})$ an efficient way might be to follow a geodesic joining (f, ζ) to (g, η) in the metric given by the locally Lipschitz-Riemannian structure $\|(\dot{f}, \dot{\zeta})\|^2_k = \mu_{\text{norm}}(f, \zeta)^2(\|\dot{f}\|^2 + \|\dot{\zeta}\|^2)$. We call this structure the *condition* (*number*) *Riemannian structure* and the induced metric (see below) the *condition* (*number*) *metric* and quickly drop the "number" from the names.

Let $\sum' \subset V = \{(f, \zeta) \epsilon V \mid \mu_{\text{norm}}(f, \zeta) = \infty\}$ and $W = V - \sum'$.[1] Note that $\mu^2_{\text{norm}}(f, \zeta)$ is not differentiable everywhere on W.[2] In any case, $\| \; \|_k$ defines a metric d_k on W by $d_k(x, y) = \inf \text{Length}(\gamma)$ over piecewise differentiable paths γ in W joining x to y.

Theorem 4 *W is a locally compact and complete in the metric d_k.*

Lemma 2 *There is a constant $C_4 > 0$ such that (f_t, ζ_t) in a C^1 path in W, $t_0 \le t \le t_1$ of length L in the $\| \; \|_k$ metric, then $\mu_{\text{norm}}(f_{t_1}, \zeta_{t_1}) \le C_4^{D^{3/2}L} \mu_{\text{norm}}(f_{t_0}, \zeta_{t_0})$.*

Proof of Lemma 2 From Proposition 1 it follows that

$$\mu_{\text{norm}}(f_{t_1}, \zeta_{t_1}) \le (1+\epsilon)^{\left(\frac{1+\epsilon}{C}\right) D^{3/2}\sqrt{2}L} \mu_{\text{norm}}(f_{t_0}, \zeta_{t_0}).$$

For an appropriate ϵ, C. Let $C_4 = (1+\epsilon)^{\sqrt{2}\left(\frac{1+\epsilon}{C}\right)}$

Proof of Theorem 2 Fix (f_0, ζ_0) for example $f_{0_i} = \frac{d_i^{1/2}}{n^{1/2}} X_i X_0^{d_i - 1}$, $i = 1, \dots, n$ and $\zeta_0 = (1, 0, \dots, 0)$. Then $\mu_{\text{norm}}(f_0, \zeta_0) = n^{1/2}$. Hence, $\mu_{\text{norm}}(f, \zeta) \le C_4^{D^{3/2} d_k((f,\zeta),(f_0,\zeta_0))} n^{1/2}$.

So any Cauchy sequence in W stays a bounded distance away from $\sum'$, hence in a compact region of W where it converges in the usual metric but also in the metric induced by d_k. Local compactness is obvious.

The condition metric d_k makes W a path metric space in the sense of Gromov [4]. Theorem then allows the application of the Hopf–Rinow theorem from [4] to conclude that any two points of W may be joined by a minimizing geodesic. I thank one of the referees for pointing this reference out to me.

[1] It follows from the condition number theorem see [5] and Theorem 1 that $\mu_{\text{norm}}(f, \zeta)$ is comparable to the reciprocal of the distance of (f, ζ) to $\sum'$ in the metric d. $\ln(\mu_{\text{norm}}(f, \zeta))$ might be an interesting function to study. It has the same asymptotics with respect to $\sum'$ as a Green's function.

[2] Instead of μ_{norm} one might consider a smooth approximation to make the Riemannian geometry easier.

References

1. C. Beltran and L. M. Pardo, Smale's 17th problem: A probabilistic positive solution, *Found. Comput. Math.* **7** (2007), 87–134.
2. C. Beltran and L. M. Pardo, Smale's 17th problem: Average polynomial time to compute affine and projective solutions. Preprint.
3. C. Beltran and M. Shub, Complexity of Bezout's theorem VII: Distance estimates in the condition metric. *Found. Comput. Math.* (2008). doi:10.1007/s10208-007-9018-5.
4. M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhäuser, Basel, 1998.
5. L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and Real Computation*, Springer, Berlin, 1998.
6. Y. E. Nesterov and M. J. Todd, On the Riemannian geometry defined by self-concordant barriers and interior-point methods, *Found. Comput. Math.* **2** (2002), 333–361.
7. M. Shub and S. Smale, Complexity of Bezout's theorem I: Geometrical aspects, *J. Am. Math. Soc.* **6** (1993), 459–501.
8. M. Shub and S. Smale, Complexity of Bezout's theorem II: Volumes and probabilities, in *Computational Algebraic Geometry* (F. Eyssette and A. Galligo, eds.), *Progress in Mathematics*, Vol. 109, pp. 267–285, Birkhäuser, Basel, 1993.
9. M. Shub and S. Smale, Complexity of Bezout's theorem III: Condition number and packing, *J. Complex.* **9** (1993), 4–14.
10. M. Shub and S. Smale, Complexity of Bezout's theorem IV: Probability of success; extensions, *SINUM* **33** (1996), 128–148.
11. M. Shub and S. Smale, Complexity of Bezout's theorem V: Polynomial time, *Theor. Comput. Sci.* **133** (1994), 141–164.