CONDITION NUMBER BOUNDS FOR PROBLEMS WITH INTEGER COEFFICIENTS

GREGORIO MALAJOVICH

Abstract. An apriori bound for the condition number associated to each of the following problems is given: general linear equation solving, minimum squares, non-symmetric eigenvalue problems, solving univariate polynomials, solving systems of multivariate polynomials. It is assumed that the input has integer coefficients and is not on the degenerate locus of the respective problem (i.e. the condition number is finite). Then condition numbers are bounded in terms of the dimension and of the bit-size of the input.

In the same setting, bounds are given for the speed of convergence of the following iterative algorithms: QR without shift for the symmetric eigenvalue problem, and Graeffe iteration for univariate polynomials.

1. Introduction

In most of the numerical analysis literature, complexity and stability of numerical algorithms are usually estimated in terms of the problem instance dimension and of a 'condition number'.

For instance, the complexity of solving an $n \times n$ linear system $Ax = b$ is usually estimated in terms of the dimension n (actually the input size is $n(n+1)$) and of the condition number $\kappa(A) = \| A \|_2 \, \| A^{-1} \|_2$.

There is a set of problems instances with $\kappa(A) = \infty$, and in most cases it makes no sense to attempt solving those problem instances. There are also problem instances (in our case, matrices) close to the locus of degenerate problem instances. Those will have a large condition number, and will be said to be *ill-conditioned*.

It is usually accepted that ill-conditioned problem instances are hard to solve. Thus, for complexity purposes a problem instance with a large condition number should be considered 'large'. Therefore, when considering problems defined for real inputs, a reasonable measure for the input size would be (in our example): $n^2 \log_2 \kappa(A)$. (Compare to [13] Formula 2.1 and paragraph below. See also the discussion in [1], Chapter 3, Section 1).

Another tradition, derived from classical complexity theory and pervasive in several branches of literature (such as linear programming), is to consider the subset of

Date: February 12, 1999.

DEPARTAMENTO DE MATEMÁTICA APLICADA, UNIVERSIDADE FEDERAL DO RIO DE JANEIRO. CAIXA POSTAL 68530, CEP 21945, RIO DE JANEIRO, RJ, BRASIL. E-MAIL: GREGORIO@LABMA.UFRJ.BR. ON LEAVE AT MSRI, 1000 CENTENNIAL DRIVE, BERKELEY CA 94720-5070. E-MAIL: GREGORIO@MSRI.ORG

problems instances with integer coefficients. Hence the input size is the number of coefficients times the bit-size of the largest coefficient (in absolute value).

In this paper, the following classical problems of numerical analysis are considered:

1. Solving a general $n \times n$ system of linear equations.
2. Minimal squares problem for a full-rank matrix.
3. Non-symmetric eigenvalue problem.
4. Solution of one univariate polynomial.
5. Solution of a non-degenerate system of n polynomial equations in n variables.

All those problems share the feature mentioned above: there is a degenerate locus, and problem instances with real coefficients can be as close to the degenerate locus as wished. This implies that they can be arbitrarily ill-conditioned.

However, in Theorems 1 to 5 below, we provide bounds for the condition number of problems instances with integer coefficients and not in the degenerate locus. Those bounds depend on the dimension (size) of the problem instance and on the bit-size of its coefficients.

In the analysis of iterative algorithms, one further considers a certain quantity that can be used to bound the speed of convergence and hence the number of iterations to obtain a given approximation. For instance, for power methods (or QR iteration without shift) in the symmetric eigenvalue problem, one can bound the number of steps in terms of the desired accuracy and of the ratio between different eigenvalues. The farther this number is from 1, the faster is the convergence.

Once again, if input has real coefficients, this quantity can be arbitrarily close to 1. However, explicit bounds for that quantity will be given for inputs with integer coefficients for

6. QR iteration without shift for the Symmetric Eigenvalue Problem.
7. Graeffe iteration for solving univariate polynomials.

The reader should be warned that the results herein are worst case estimates, and are overly pessimistic for application purposes. The main motivation for those results is to convert numerical analysis estimates into 'polynomial time' estimates, not the opposite.

2. Statement of main results

Notation. $\| . \|_2$ stands for the *2-norm*: if $x \in \mathbb{R}^n$ or $\mathbb{C}^n$, then

$$\| x \|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \ .$$

If A is a matrix, then

$$\| A \|_2 = \sup_{\| x \|_2 = 1} \| Ax \|_2 \ .$$

2.1. **Linear equation solving.** The first problem considered is linear equation solving: given an $n \times n$ matrix A and a vector $b \in \mathbb{R}^n$, find $x \in \mathbb{R}^n$ such that $Ax = b$.

Its condition number (with respect to the 2-norm) is defined as

$$\kappa(A) = \| A \|_2 \, \| A^{-1} \|_2 \ .$$

Comprehensive treatment of the perturbation theory for this problem can be found in the literature, such as [3] Section 2.2, [4] Chapter 7, [14] Lecture 12, etc...

Theorem 1. *Let A be an $n \times n$ matrix with integer coefficients. If A is invertible, then*

$$\kappa(A) \le n^{\frac{n}{2}+1} \left(\max_{i,j} |A_{ij}| \right)^n \ .$$

No originality is claimed for Theorem 1. This result is included for completeness and because its proof is elementary, yet illustrates the principle behind the other results.

2.2. **Minimal squares.** The second problem in the list is minimal squares fitting. Let A be an $m \times n$ matrix, $m \ge n$, with full rank, and let $b \in \mathbb{R}^m$. One has to find x to minimize $\| Ax - b \|_2{}^2$.

Let $r = Ax - b$ be the residual, we are minimizing $\| r \|_2{}^2$. Let

$$\sin \theta = \frac{\| r \|_2}{\| b \|_2} \ .$$

According to [3] p. 117 (Compare to [14] Lecture 18 and [4] Section 19.1), the condition number of the linear least squares problem is

$$\kappa_{\mathrm{LS}}(A, b) = \frac{2\kappa(A)}{\cos \theta} + \tan \theta \ \kappa(A)^2 \ .$$

Since we do not assume A to be square, we need to give a new definition for $\kappa(A)$. Let $\sigma_{\mathrm{MAX}}(A)$ and $\sigma_{\mathrm{MIN}}(A)$ be respectively the largest and the smallest singular values of A. Then set

$$\kappa(A) = \frac{\sigma_{\mathrm{MAX}}(A)}{\sigma_{\mathrm{MIN}}(A)} \ .$$

When $m = n$, this definition is equal to the previous one.

The singular locus is now the set of pairs (A, b) such that A does not have full rank (i.e. $\sigma_{\mathrm{MIN}}(A) = 0$) or such that $\| r \|_2 = \| b \|_2$ (i.e. b is orthogonal to the image of A).

The result is:

Theorem 2. *Let A be an $m \times n$ matrix with integer coefficients, and assume that A has full rank. Let $b \in \mathbb{Z}^m$. Set $H = \max_{i,j} (|A_{ij}|, |b_i|)$. Then if b is not orthogonal to the image of A, we have:*

$$\kappa_{LS}(A, b) \le 3 n^{\frac{n}{2}+1} m^{n+\frac{1}{2}} H^{2n+1} \ .$$

2.3. **Non-symmetric eigenvalue problem.** Let A be an $n \times n$ matrix and let λ be a single eigenvalue of A. The condition number of λ depends on the angle between the left and right eigenvectors:

Let x, y be respectively right and left norm-1 eigenvectors of A associated to λ: $Ax = \lambda x$, $y^* A = \lambda y^*$, and $\|x\|_2 = \|y\|_2 = 1$. Then

$$\kappa_{\text{NSE}}(A, \lambda) = \sec(\widehat{x, y}) = \frac{1}{y^* x} \ .$$

See [3] Theorem 4.4 p. 149 for references.

Theorem 3. *Let A be an $n \times n$ matrix with integer coefficients, and let λ be a single eigenvalue of A. Then*

$$\kappa_{NSE}(A, \lambda) \leq n^{3n} 2^{2n} \left(2\sqrt{n} H(A)\right)^{2n^3 - 2n} \ .$$

2.4. **Solving univariate polynomials.** The condition number (in affine space) for solving a univariate polynomial $f(x) = \sum_{i=0}^{d} f_i x^i$ can be defined ([2] page 228) as:

$$\mu(f) = \max_{\zeta \in \mathbb{C} : f(\zeta) = 0} \mu(f, \zeta) \ ,$$

where

$$\mu(f, \zeta) = \frac{\left(\sum_{i=0}^{d} |\zeta|^{2i}\right)^{\frac{1}{2}}}{|f'(\zeta)|} \ .$$

The degenerate locus is the set of polynomials with a multiple root or with a root at infinity.

Theorem 4. *Let $f : x \mapsto \sum_{i=0}^{d} f_i x^i$ be a univariate polynomial with integer coefficients, without multiple roots. Then*

$$\mu(f) \leq 2^{2d^2 - 2} d^{2d} \left(\max |f_i|\right)^{2d^2} \ .$$

2.5. **Solving systems of polynomials.** A similar condition number exists for systems of polynomials. However, for the purpose of condition number theory, it is usually convenient to homogenize the equations and to study the perturbation theory of the 'roots' in complex projective space. This can also be seen as a change of metric, that simplifies the formula of the condition number and of several theorems (See [2] Chapters 10, 12, 13).

Let $f = (f_1, \cdots, f_n)$ be a system of polynomials in variables $x_1, \cdots, x_n$. We homogenize the system by multiplying each coefficient $f_{iJ} x^J = f_{iJ} x_1^{J_1} x_2^{J_2} \cdots x_n^{J_n}$ of f_i by $x_0^{J_0}$, where we choose $J_0 = \deg f_i - (J_1 + \cdots J_n)$. We obtain a system of homogeneous polynomials in $n+1$ variables, that we call $F = (F_1, \cdots, F_n)$. The natural space for the roots of F is projective space $\mathbb{P}^n$, defined as the space of all 'rays'

$$(x_0 : \cdots : x_n) = \{(\lambda x_0, \cdots, \lambda x_n) : \lambda \in \mathbb{C}\} \ .$$

where $x_0, \ldots, x_n$ are not all equal to 0.

Every finite root $(x_1, \cdots, x_n)$ of f corresponds to the projective root of F given by $(1 : x_1 : \cdots : x_n)$, and projective roots of F correspond either to a finite root of f or to a root 'at infinity'.

Suppose that the coefficients of f (hence of F) are made to depend upon a parameter t. The condition number bounds the absolute speed of the roots of F (in projective space) with respect to the absolute speed of the coefficients of F. Recall that the roots ζ of F are in projective space, so their speed vector $\dot{\zeta}$ belongs to the tangent space $T_\zeta \mathbb{P}^n$.

The condition number of F at a root turns out to be:

$$\mu(F, \zeta) = \| F \|_2 \left\| \left(\mathrm{D}F(\zeta)_{|T_\zeta} \right)^{-1} \begin{bmatrix} \| \zeta \|_2{}^{d_1 - 1} & & \\ & \ddots & \\ & & \| \zeta \|_2{}^{d_n - 1} \end{bmatrix} \right\|_2$$

where $\zeta \in \mathbb{C}^{n+1}$ is such that $(\zeta_0 : \cdots : \zeta_n)$ is a root of F (See Proposition 7c in Page 230 of [2]). We did not define the norm of a polynomial yet. Above, $\| \,.\, \|_2$ stands for the unitary invariant norm (See [15] Chapter III-7 or [2] Section 12.1), that is the most reasonable generalization of the 2-norm to spaces of polynomials:

Notation. Let G be a homogeneous degree d polynomial in $n+1$ variables. Then

$$\| G \|_2 = \sqrt{\sum_J \frac{|G_J|^2}{\begin{pmatrix} d \\ J \end{pmatrix}}}$$

where $\begin{pmatrix} d \\ J \end{pmatrix}$ is $\frac{d!}{J_0! \cdots J_n!}$. Let F be a system of homogeneous polynomials. Then

$$\| F \|_2 = \sqrt{\sum \| F_i \|_2{}^2} \,.$$

With these definitions, the number $\mu(F, \zeta)$ is invariant under scalings of F, ζ, and under the action of the unitary group $U(n+1)$, where an element $Q \in SU(n+1)$ acts by $Q : (F, \zeta) \mapsto (F \circ Q, Q\zeta)$.

In order to define the condition number of a system of n equations in n variables, we set:

$$\mu(f) = \max_\zeta \mu(F, \zeta)$$

where ζ ranges over the roots of F. (Another possibility is to restrict ζ to the non-degenerate roots of F. This would make no difference in this paper). The following theorem is true if one restricts ζ to any subset of the roots of F.

Theorem 5. *Let f be a system of n polynomial equations in n variables, with integer coefficients. We write $H(f)$ for the maximum of the absolute value of the coefficients of f, $S(f)$ for the number of non-zero coefficients of f and D for* $\max d_i$. *Assume that $\mu(f)$ is finite. Then*

$$\mu(f) \leq ((n+1) S H(f))^{D^{cn}}$$

where c is an universal constant.

2.6. **Symmetric eigenvalue problem.** Let A be an $n \times n$ real positive symmetric matrix, and let $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0$ be its eigenvalues.

Unlike the non-symmetric eigenvalue problem, the symmetric eigenvalue problem has absolute condition number always equal to 1 (See [3] Theorem 5.1. See also citePARLETT Fact 1.11 p.16).

However, when using an iterative algorithm, the ratio of eigenvalues

$$\rho(A) = \min_{j>i} \frac{\lambda_j}{\lambda_i}$$

may play an important role for estimating convergence. For instance, according to [14] Theorem 28.4, the QR algorithm without shift converges linearly with speed $\frac{1}{\rho(A)}$. Convergence may get slower when $\rho(A) \to 1$. Therefore one can bound the speed of convergence by bounding

$$\nu(A) = \rho(A) - 1 = \min_{j>i} \frac{\lambda_j - \lambda_i}{\lambda_i}$$

above from zero. If $\nu(A) > \delta_0$, then $\rho(A) > 1 + \delta_0$. After $k > \lceil \frac{1}{\delta_0} \rceil$ iterations, one gets

$$\rho(A)^k > 1 + k\delta_0 \geq 2 \ .$$

Thus it suffices to perform $O(\frac{1}{\delta_0} \log_2 \frac{1}{\delta_1})$ iterations to obtain a result with accuracy δ_1.

Also, the quantity $\nu(A)^{-1}$ can also be interpreted as a condition number for the eigenvectors (See [3] Theorem 5.7 p. 208). We will show here that

Theorem 6. *Let A be an $n \times n$ matrix with integer coefficients. Then*

$$\nu(A)^{-1} \geq 8^{-n} (4n)^{-n^2} \left(\max_{i,j} |A_{ij}| \right)^{-2n^2} \quad .$$

2.7. **Graeffe iteration.** Let $f : x \mapsto \sum_{i=0}^{d} f_i x^i$, $f_d = 1$ be a monic univariate polynomial with zeros $\zeta_1, \cdots, \zeta_d$. Those zeros can be ordered such that

$$|\zeta_1| \geq |\zeta_2| \geq \cdots \geq |\zeta_d| \ .$$

The Graeffe operator maps the polynomial $f(x) = \prod_{i=1}^{d} (x - \zeta_i)$ into the polynomial $Gf(x) = (-1)^d f(\sqrt{x}) f(-\sqrt{x}) = \prod_{i=1}^{d} (x - \zeta_i^2)$.

In [10, 11], it is explained how to recover the actual roots of f after a certain number of Graeffe iterations, with a good approximation. The number of required iterations depends on the ratio:

$$\rho(f) = \max_{|\zeta_j| > |\zeta_i|} \frac{|\zeta_j|}{|\zeta_i|} \ .$$

Unlike in Section 2.6, we do not require here that the roots have different absolute value. We consider also the auxiliary quantity

$$\nu(f) = \rho(f) - 1 = \max_{|\zeta_j| > |\zeta_i|} \frac{|\zeta_j| - |\zeta_i|}{|\zeta_i|} \ .$$

By the above definitions, the 'condition number' $\nu(f)^{-1}$ is always finite. In order to recover the roots within relative precision δ, the number of Graeffe iterations to perform is

$$O(\log \nu(f)^{-1} + \log d + \log\log \delta^{-1}) \ .$$

For clarity of exposition, we will show that bound under a special hypothesis: all the roots should be different positive real numbers. For the general case, see [8] and [9]. Also, all estimates here are 'up to the first order', and quadratic error terms will be discarded.

After k steps of Graeffe iteration one obtains the polynomial

$$g(x) = G^k f(x) = \sum_{i=0}^{d} g_i x^i = \prod_{i=1}^{d} (x - \zeta_i^{2^k})$$

with $\rho(g) = \rho(f)^{2^k}$.

Expanding each g_i as the $(d-i)$-th elementary symmetric function of the $\zeta_i^{2^k}$, one obtains

$$\begin{aligned} g_0 &= \sigma_d(\zeta_1^{2^k}, \cdots, \zeta_d^{2^k}) \\ g_1 &= \sigma_{d-1}(\zeta_1^{2^k}, \cdots, \zeta_d^{2^k}) \\ &\vdots \\ g_{d-1} &= \sigma_1(\zeta_1^{2^k}, \cdots, \zeta_d^{2^k}) \\ g_d &= 1 \end{aligned}$$

We can use the special hypothesis to bound

$$(\zeta_1 \zeta_2 \cdots \zeta_{d-i})^{2^k} = g_i(1 + \delta_i)$$

with $|\delta_i| < \frac{2^d}{\rho(g)}$ + h.o.t. Hence

$$\zeta_i^{2^k} = \frac{g_{d-i+1}}{g_{d-i}}(1 + \delta_i')$$

with $|\delta_i'| < \frac{2^{d+1}}{\rho(g)}$ + h.o.t.

Since we assumed the ζ_i are all positive, we can recover them by taking 2^k-th roots

$$\zeta_i = \left(\frac{g_{d-i+1}}{g_{d-i}} \right)^{2^{-k}} (1 + \delta_i'') \ .$$

with $|\delta_i''| \leq \frac{2^{d+1-k}}{\rho(g)}$ + h.o.t.

Now we can use the estimate on $\rho(g) = \rho(f)^{2^k}$ to deduce that $O(\log \nu(f)^{-1} + \log \delta^{-1})$ steps are sufficient to obtain a relative precision δ in the roots. Indeed after $k_1 = \log_2 \nu(f)^{-1}$ steps,

$$\rho(G^{k_1} f) = \rho(f)^{2^{\log_2 \nu(f)^{-1}}} \geq 2 \ .$$

After extra $k_2 = \log_2(d + 1 + \log_2 \delta^{-1})$ steps, one gets

$$\rho(G^{k_1+k_2}(f)) > 2^{2^{\log_2(d+1+\log_2 \delta^{-1})}} = 2^d \delta^{-1} \ .$$

So we can set $k = k_1 + k_2 + 1$, the last 1 to get rid of the high order terms, to deduce that $|\delta_i''| < \delta$.

Theorem 7. *Let $f : x \mapsto \sum_{i=0}^d f_i x^i$ be a polynomial with integer coefficients. Then*

$$\nu(f)^{-1} > (8 \max |f_i|)^{-2d} \ .$$

This says that Graeffe iteration is 'polynomial time', in the sense that we can obtain relative accuracy δ of the roots after

$$O(d \log \max |f_i| + \log \log \delta^{-1})$$

steps.

3. Background material

The proof of Theorems 3 to 7 will make use of the *absolute multiplicative height function* H to bound inequalities involving algebraic numbers.

The construction of the height function H is quite standard in number theory and we refer the reader to [6] Chapter II or to [12] pages 205–214. For applications to complexity theory, see [2] Chapter 7 and [7].

The height function is naturally defined in the projectivization $\mathbb{P}^n(\mathbb{Q}^a)$ of the algebraic numbers $\mathbb{Q}^a$. It returns a real number ≥ 1. We can also extend it to complex projective space $\mathbb{P}^n$ by setting $H(P) = \infty$ when $P \notin \mathbb{P}^n(\mathbb{Q}^a)$. We will adopt this convention in order to simplify the notation of domains and ranges.

Also, if $x = (x_1, \cdots, x_n) \in \mathbb{C}^n$, we can define its height as $H(x) = H(x_1 : \cdots : x_n : 1)$.

We can also define the height of matrices, polynomials and systems of polynomials as the height of the vector of all the coefficients.

The following properties of heights will be used in the sequel. First of all, we can explicitly write the height of a vector with integer coefficients as:

Proposition 1. *If $u \in \mathbb{Z}^n$, then $H(u) = \max_{1 \leq i \leq n} |u_i|$, where $|\,.\,|$ is the standard absolute value.*

Proposition 1 follows from the construction of the height function. One immediate consequence is that if $v \in \mathbb{Q}^n$, then $H(v) = \max |m v_i|, |m|$ where m is the greatest common denominator of the v_i's.

We can use the following fact to bound the height of the roots of an integral polynomial:

Proposition 2. *If $f : x \mapsto f(x) = \sum_{i=0}^d f_i x^i$ is a non-zero polynomial with integer coefficients, and if x is a root of f, then $H(x) \leq 2 \max |f_i|$.*

Proposition 2 is Theorem 5 in [7]. Compare with Theorem 5.9 in [12], where the coefficients of f are algebraic numbers.

We can use a bound on the height to bound absolute values above and below:

Proposition 3. *Let K be an algebraic extension of $\mathbb{Q}$, and let $x \in K$, $x \neq 0$. Then*

$$H(x)^{-\deg[K:\mathbb{Q}]} \leq |x| \leq H(x)^{\deg[K:\mathbb{Q}]} .$$

The height of a vector and of its coordinates can be related by:

Proposition 4.

$$H(x_1) \leq H(x_1, \cdots, x_n) \leq H(x_1)H(x_2)\cdots H(x_n) .$$

Propositions 3 and 4 follow immediately from the construction of the height function. The height function is invariant under permutation of coordinates, and also:

Proposition 5. *Let K be an algebraic extension of $\mathbb{Q}$, and let $g \in Gal_{[K:\mathbb{Q}]}$. Then for any $x \in K$, $H(g(x)) = H(x)$*

Proposition 5 is Lemma 5.10 in [12].

Proposition 6. *Let*

$$\begin{array}{rccl} F = (F_0, \cdots, F_m): & \mathbb{C}^{n_1} \times \mathbb{C}^{n_2} \times \cdots \times \mathbb{C}^{n_k} & \to & \mathbb{C}^{m+1} \\ & P^1, \cdots, P^k & \mapsto & F(P^1, \cdots, P^k) \end{array}$$

be a system of multi-homogeneous polynomials with algebraic coefficients, where each F_i has degree d_j in variables P^j. Let the P^j be algebraic. Then

$$H(F(P)) \leq (\max S(f_i))H(F)H(P^1)^{d_1}\cdots H(P^k)^{d_k} .$$

In the case $k = 1$, this is similar to Theorem 5.6 in [12] (where $\max S(f_i)$ is not given explicitly). For the general case see Theorem 4 in [7].

Proposition 7. *Let*

$$\begin{array}{rccl} G = (G_1, \cdots, G_m): & \mathbb{C}^{n_1} \times \mathbb{C}^{n_2} \times \cdots \times \mathbb{C}^{n_k} & \to & \mathbb{C}^{m} \\ & Q^1, \cdots, Q^k & \mapsto & G(Q^1, \cdots, Q^k) \end{array}$$

be a system of polynomials with algebraic coefficients, where each G_i has degree at most d_j in variables Q^j. Let the Q^j be algebraic. Then

$$H(G(Q)) \leq (\max S(G_i))H(G)H(Q^1)^{d_1}\cdots H(Q^k)^{d_k} .$$

This is Corollary 1 in [7]. Some consequences of this are that $H(\sum_{i=1}^n x_i) \leq n \prod H(x_i)$ and that $H(\prod_{i=1}^n x_i) \leq \prod H(x_i)$.

The following fact follows also from the construction of heights:

Proposition 8. *If x is an algebraic number,*

$$H(x^2) = H(x)^2 .$$

Also, it makes sense to bound the height of the roots of a system of polynomials with respect to the height, size and degree of the system. Corollary 6 in [5] is:

Proposition 9. *[Krick and Pardo] Let $f_1, \cdots, f_r$ $r \leq n$ be polynomials in $\mathbb{Z}[x_1, \cdots, x_r]$ of degree and height bounded by $d \geq n$ and η, respectively, and let V denote the algebraic affine variety defined by: $V = \{x : f_1(x) = \cdots f_r(x) = 0\}$.*

Then V has at most d^n isolated points, and their height verifies:

$$\log_2 H(P) \in d^{O(n)}(\log_2 r + \log_2 \eta) \ .$$

4. Proof of Theorems

Notation. If A is a real (resp. complex) matrix, then A^* is the real (resp. complex) transpose of A, $(A^*)_{ij} = \bar{A}_{ji}$. The same convention will be used for vectors.

The vectors of the canonical basis will be denoted by $e_1 = [1, 0, 0, \cdots]^*$, $e_2 = [0, 1, 0, \cdots]^*$, etc...

4.1. Proof of Theorem 1.

$$\begin{aligned}
\| A \|_2 &= \sup_{\| u \|_2 = 1} \| Au \|_2 \text{ by definition} \\
&\leq \sum_j |u_j| \, \| [A_{1j}, \cdots, A_{nj}]^* \|_2 \text{ by triangular inequality} \\
&\leq \sqrt{n} \max_j \| [A_{1j}, \cdots, A_{nj}]^* \|_2 \text{ since } \| u \|_2 = 1 \\
&\leq n \max_{ij} |A_{ij}|
\end{aligned}$$

Let $A(i, u)$ be the matrix obtained by replacing the i-th column of A by the vector u. Then if $v = A^{-1}u$, Cramer's rule is:

$$v_i = \frac{\det A(i, u)}{\det A} \ .$$

Since A has integer coefficients and $\det A \neq 0$, one can always bound $|v_i| \leq |\det A(i, u)|$. By Hadamard inequality, this implies:

$$\begin{aligned}
|v_i| &\leq \| u \|_2 \max_j \| [A_{1j}, \cdots, A_{nj}]^* \|_2{}^{n-1} \\
&\leq \| u \|_2 (\sqrt{n})^{n-1} \left(\max_{i,j} |A_{ij}| \right)^{n-1}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left\| A^{-1} \right\|_2 &= \sup_{\| u \|_2 = 1} \left\| A^{-1} u \right\|_2 \text{ by definition} \\
&\leq n^{\frac{n}{2}} \left(\max_{i,j} |A_{ij}| \right)^{n-1}
\end{aligned}$$

Combining the bounds for $\| A \|_2$ and $\| A^{-1} \|_2$, one obtains:

$$\kappa(A) \leq n^{\frac{n}{2}+1} \left(\max_{i,j} |A_{ij}| \right)^n \ .$$

4.2. **Proof of Theorem 2.** In order to estimate $\kappa(A)$, we write

$$\begin{aligned} \kappa(A) &= \sqrt{\kappa(A^*A)} \\ &\leq n^{\frac{n}{4}+\frac{1}{2}} H^n m^{n/2} \end{aligned}$$

In order to bound $\cos\theta$, we use the assumption that b is not orthogonal to the image of A. Hence $\| A^*b \|_2 \geq 1$ and the 'normal equation' $A^*Ax = A^*b$ implies:

$$\| A^*Ax \|_2 \geq 1 \ .$$

Therefore,

$$\cos\theta = \frac{\| A^*Ax \|_2}{\| b \|_2} \geq \frac{1}{\| b \|_2} \geq \frac{1}{H\sqrt{m}}$$

and $\frac{1}{\cos\theta}$ and $\tan\theta$ are bounded above by $H\sqrt{m}$. Putting all together,

$$\begin{aligned} \kappa_{\mathrm{LS}}(A,b) &\leq 2n^{\frac{n}{4}+\frac{1}{2}} m^{\frac{n}{2}+\frac{1}{2}} H^{n+1} + n^{\frac{n}{2}+1} m^{n+\frac{1}{2}} H^{2n+1} \\ &\leq 3n^{\frac{n}{2}+1} m^{n+\frac{1}{2}} H^{2n+1} \end{aligned}$$

4.3. **Proof of Theorem 3.**

Lemma 1. *Let B be an $n \times n$ matrix with integer coefficients. Let $p(t) = \det(B - tI) = \sum p_i t^i$. Then*

$$\max |p_i| \leq \left(2\sqrt{n} \max_{i,j} |B_{ij}| \right)^n \ .$$

Proof of Lemma 1.

$$p_i = \sum_C \pm \det C$$

where C ranges over the $(n-i) \times (n-i)$ sub-matrices of B of the form $C_{kl} = Bs_k s_l$ for some $1 \leq s_1 < \cdots < s_{n-1} \leq n$. Hence

$$\begin{aligned} |p_i| &\leq \binom{n}{i} \max_C |\det C| \\ &\leq \binom{n}{i} \left(\sqrt{n-i} \max_{ij} |C_{ij}| \right)^{n-i} \\ &\leq \left(2\sqrt{n} \max_{i,j} |B_{ij}| \right)^n \end{aligned}$$

□

Lemma 2. *Let A be an $n \times n$ matrix with integer coefficients and let λ be an eigenvalue of A. Then*

$$H(\lambda) \leq 2 \left(2\sqrt{n} \max_{i,j} |A_{ij}| \right)^n \ .$$

Proof of Lemma 2. Apply Proposition 2 to the polynomial $p(t)$ from Lemma 1. □

Lemma 3. *Let B be an $n \times n$ matrix with integer coefficients. Let $q(t) = \det(B - tI + te_n^* e_n) = \sum q_i t^i$. Then*

$$\max |q_i| \leq 2 \left(2\sqrt{n} \max_{i,j} |B_{ij}| \right)^n .$$

Proof of Lemma 3. Let $p(t) = \det(B - tI)$ and let $r(t) = \det(\tilde{B} - tI)$ where $\tilde{B}$ is the $(n-1) \times (n-1)$ matrix obtained by deleting the n-th row and the n-th column of B. Then, by multi-linearity of the determinant,

$$p(t) = q(t) \pm t\ r(t) ,$$

hence

$$q(t) = p(t) \pm t\ r(t) .$$

Therefore,

$$\begin{aligned} \max |q_i| &\leq \max |p_i| + \max |r_i| \\ &\leq \left(2\sqrt{n} \max_{i,j} |B_{ij}| \right)^n + \left(2\sqrt{n-1} \max_{i,j} |B_{ij}| \right)^{n-1} \quad \text{(Lemma 1)} \\ &\leq 2 \left(2\sqrt{n} \max_{i,j} |B_{ij}| \right)^n \end{aligned}$$

□

Lemma 4. *Let A be an $n \times n$ matrix with integer coefficients. Let λ be an isolated eigenvalue of A and let x be an eigenvector associated to λ, $Ax = \lambda x$. Then*

$$H(x_1 : \cdots : x_n) \leq n 2^n \left(2\sqrt{n} \max_{i,j} |A_{ij}| \right)^{n^2-1} .$$

Proof of Lemma 4. Assume without loss of generality that the first $n-1$ lines of $A - \lambda I$ are independent. Let $M_1, \ldots, M_i, \ldots, M_n$ be the sub-matrices obtained from $A - \lambda I$ by deleting the last line and the i-th column. Then we can scale x in such a way that

$$x_i = \pm \det M_i .$$

We have $M_n = B_n - \lambda I$. By reordering rows and columns, we obtain for each $i < n$ that M_i is of the form:

$$M_i = B_i - \lambda I + \lambda e_{n-1}^* e_{n-1}$$

where B_i is the sub-matrix of A obtained by deleting the last line and the i-th column.

Set $q^{(i)}(\lambda) = \det M_i = \sum q_j^{(i)} \lambda^j$. Now by Lemma 3,

$$\max |q_j^{(i)}| \leq 2 \left(2\sqrt{n-1} \max_{k,l} |A_{kl}| \right)^{n-1} .$$

We consider now the morphism:

$$\begin{array}{rccc} q : & \mathbb{P} & \to & \mathbb{P}^n \\ & (\lambda : 1) & \mapsto & (q^{(1)}(\lambda) : \cdots : q^{(n)}(\lambda)) \end{array}$$

Then $x = q(\lambda)$ and

$$\begin{aligned} H(x) &= H(q(\lambda)) \\ &\leq nH(q)H(\lambda)^{n-1} \\ &\leq n2\left(2\sqrt{n-1}\max_{i,j}|A_{ij}|\right)^{n-1} 2^{n-1}\left(2\sqrt{n}\max_{i,j}|A_{ij}|\right)^{n(n-1)} \\ &\leq n2^n\left(2\sqrt{n}\max_{i,j}|A_{ij}|\right)^{n^2-1} \end{aligned}$$

the first inequality because of Proposition 6, and the second because of Lemma 2. □

End of the Proof of Theorem 3. Proposition 7 implies

$$H(y^* x) \leq nH(x)H(y) \ .$$

We claim that $\deg[\mathbb{Q}[y^*x] : \mathbb{Q}] \leq n$. Indeed, x and y can be obtained by solving systems of linear equations with coefficients in $\mathbb{Q}[\lambda]$, thus $x_i, y_i \in \mathbb{Q}[\lambda]$. Also, $\overline{y_i} \in \mathbb{Q}[\bar{\lambda}] = \mathbb{Q}[\lambda]$ so $\deg[\mathbb{Q}[y^*x] : \mathbb{Q}] \leq n$ as claimed.

By hypothesis $y^*x \neq 0$. Hence, by Proposition 3,

$$\begin{aligned} |y^*x| &\geq (nH(x)H(y))^{-n} \\ &\geq n^{-3n}2^{-2n}\left(2\sqrt{n}\max_{i,j}|A_{ij}|\right)^{-2n^3+2n} \end{aligned}$$

□

4.4. **Proof of Theorem 4.**

Proof. According to Proposition 2,

$$H(\zeta) \leq 2\max|f_i| \ .$$

Also,

$$H(f_i') \leq dH(f) = d\max|f_i|$$

and according to Proposition 7

$$\begin{aligned} H(f'(\zeta)) &\leq dH(f')H(\zeta)^{d-1} \\ &\leq d^2 2^{d-1}H(f)^d \end{aligned}$$

and hence $|f'(\zeta)| \geq d^{-2d}H(f)^{-d^2}2^{-d(d-1)}$. On the other hand,

$$\begin{aligned} \left(\sum_{i=0}^{d}|\zeta|^{2i}\right)^{\frac{1}{2}} &\leq \sqrt{d+1}\max(1,|\zeta|^d) \\ &\leq \sqrt{d+1}H(\zeta)^{d^2} \\ &\leq \sqrt{d+1}2^{d^2}H(f)^{d^2} \end{aligned}$$

Hence

$$\mu(f,\zeta) \le 2^{2d^2-2} H(f)^{2d^2} d^{2d} \ .$$

□

4.5. **Proof of Theorem 5.**

Lemma 5. *Let A be an $n \times n$ invertible matrix with algebraic coefficients. Then*

$$H(A^{-1}) \le n H(A)^n \ .$$

Proof of Lemma 5. Let $A(i,j)$ be the sub-matrix of A obtained by deleting the i-th row and the j-th column. By Cramer's rule, $(A^{-1})_{ji} = \frac{\det A(i,j)}{\det A}$. Therefore we should define the degree n morphism:

$$\begin{array}{rccc} \varphi : & \mathbb{P}^{n^2} & \rightarrow & \mathbb{P}^{n^2} \\ & \begin{array}{l} (A_{11} : A_{12} : \cdots \\ \quad \cdots : A_{nn} : 1) \end{array} & \mapsto & \begin{array}{l} (\det A(1,1) : \det A(1,2) : \cdots \\ \quad \cdots : \det A(n,n) : \det A) \end{array} \end{array}$$

Then by Proposition 6,

$$\begin{array}{rcl} H(\varphi(A)) & \le & n! H(A)^n H(\varphi) \\ & \le & n! H(A)^n \end{array}$$

□

Let us fix the notations

$$M = \begin{bmatrix} \|\zeta\|_2^{1-d_1} & & & \\ & \ddots & & \\ & & \|\zeta\|_2^{1-d_n} & \\ & & & \|\zeta\|_2^{-1} \end{bmatrix} \begin{bmatrix} \mathrm{D}F(\zeta) \\ \zeta^* \end{bmatrix}$$

and

$$C = \mathrm{D}F(\zeta)^{-1}_{|T_\zeta} \begin{bmatrix} \|\zeta\|_2^{d_1-1} & & \\ & \ddots & \\ & & \|\zeta\|_2^{d_n-1} \end{bmatrix} .$$

Let $\zeta \in \mathbb{C}^{n+1}$ be a fixed representative for a root of F. Any $u \in T_\zeta \mathbb{P}^n$ can be written as a vector in $\mathbb{C}^{n+1}$, orthogonal to ζ. Computing $u = Cv$ is the same as solving $Mu = \begin{bmatrix} v \\ 0 \end{bmatrix}$. The operator C is the same as $(M^{-1})_{|x_{n+1}=0}$. Therefore,

$$\|C\|_2 \le \left\| M^{-1} \right\|_2 \ .$$

Lemma 6. *In the conditions of Theorem 5,*

$$H(M) \le S\sqrt{n+1}^{D-1} H(\zeta)^{2D-2} D H(f) \ .$$

Proof of Lemma 6. We apply Proposition 7 to the system:

$$\zeta, N \mapsto \left(\cdots, N^{d_i - 1} \frac{\partial F_i}{\partial x_j}(\zeta), \cdots, \bar{\zeta}_j, \cdots \right)$$

with $N = \| \zeta \|_2^{-1}$ to obtain

$$H(M) \leq S H(\zeta)^{D-1} H(N)^{D-1} H(\mathrm{D}F) \ .$$

We can bound $H(\mathrm{D}F) \leq D H(F)$ and $H(N) = H(\| \zeta \|_2) = \sqrt{H(\sum |\zeta_i|^2)}$. We can apply Proposition 6 to the map

$$\begin{array}{rccc} \varphi : & \mathbb{C}^{n+1} & \to & \mathbb{C} \\ & \zeta, \bar{\zeta} & \mapsto & \sum \zeta_i, \bar{\zeta}_i \end{array}$$

to get $H(N^2) \leq (n+1) H(\zeta) H(\bar{\zeta})$. Proposition 5 implies $H(\zeta) = H(\bar{\zeta})$, hence:

$$H(N^2) \leq (n+1) H(\zeta)^2$$

and by Proposition 8,

$$H(N) \leq \sqrt{n+1} H(\zeta) \ .$$

Thus, we can estimate that

$$H(M) \leq S \sqrt{n+1}^{D-1} H(\zeta)^{2D-2} D H(f) \ .$$

□

End of the Proof of Theorem 5. By definition of the norm, $\| f \|_2 \leq S H(f)$. By Lemma 5 and Lemma 6, we have:

$$\begin{array}{rcl} H(C) & \leq & (n+1) H(M)^n \\ & \leq & (n+1) S^n \sqrt{n+1}^{nD-n} H(\zeta)^{2nD-2n} D^n H(f)^n \end{array}$$

Knowing that $\deg [\mathbb{Q}[\zeta] : \mathbb{Q}] \leq D^n$, we can use Proposition 3 to deduce that

$$\begin{array}{rcl} \| C \|_2 & \leq & (n+1) H(C) \\ & \leq & (n+1) \left((n+1) S^n \sqrt{n+1}^{nD-n} H(\zeta)^{2nD-2n} D^n H(f)^n \right)^{D^n} \end{array}$$

According to Proposition 9,

$$H(\zeta) \leq n H(f)^{D^{c' n}}$$

where c' is a universal constant. Thus,

$$\begin{array}{rcl} \mu(f, \zeta) & \leq & S H(f) (n+1) \Big((n+1) S^n \sqrt{n+1}^{nD-n} n^{2nD-2n} \\ & & \qquad H(f)^{D^{c' n}(2nD-2n)} D^n H(f)^n \Big)^{D^n} \\ & \leq & ((n+1) S H(f))^{D^{cn}} \end{array}$$

where c is a universal constant.

□

4.6. **Proof of Theorem 6.** Lemma 2 implies:

$$H(\lambda) \leq 2 \left(2\sqrt{n} \max_{i,j} |A_{ij} \right)^{n} .$$

Hence (Proposition 7),

$$H\left(\frac{\lambda_j}{\lambda_i} - 1\right) \leq 8(4n)^n \left(\max_{i,j} |A_{ij}| \right)^{2n} .$$

Thus, by Proposition 3,

$$\left| \frac{\lambda_j}{\lambda_i} - 1 \right| \leq 8^n (4n)^{n^2} \left(\max_{i,j} |A_{ij}| \right)^{2n^2} .$$

4.7. **Proof of Theorem 7.** According to Proposition 2,

$$\begin{aligned} H(\zeta_i) &\leq 2H(f) \\ H(\zeta_j) &\leq 2H(f) \end{aligned}$$

Moreover, $H(|\zeta_i|) \leq H(\zeta_i)$ because $|\zeta_i|^2 = \zeta_i \bar{\zeta_i}$ and $H(\zeta_i) = H(\bar{\zeta_i})$ (Propositions 5, 7 and 8). Thus,

$$\begin{aligned} H\left(\frac{|\zeta_j|}{|\zeta_i|} - 1 \right) &\leq 2H(|\zeta_i|)H(|\zeta_j|) \\ &\leq 8H(f)^2 \end{aligned}$$

It follows that

$$\nu(f)^{-1} \geq (8H(f))^{-2\deg[\mathbb{Q}[|\zeta_i|,|\zeta_j|]:\mathbb{Q}]} \geq (8H(f))^{-2d} .$$

5. Further comments

As mentioned before, a reasonable definition for the 'real complexity' input size is the number of coefficients of a given problem instance, times the logarithm of its condition number.

Theorems 1 to 4 show that the 'real complexity' input size is no worse than a polynomial of the 'classical complexity' input size, for problem instances with integer coefficients. Theorem 5 also, if one considers D^n as part of the input size. It may be possible to replace D^n by the Bézout number $\prod d_i$, that is the number of solutions of a generic system of polynomials.

Since the 'real complexity' of the problems considered can be bound by common numerical analysis techniques, those Theorems provide a scheme to convert 'real complexity' bounds into 'classical complexity' bounds.

The same idea is behind Theorems 6 and 7. In the case of the iterative algorithms considered, the number of iterations for obtaining a certain approximation can also be bounded in terms of a 'condition number'. In the case of problem instances with integer coefficients, the 'condition number' is also polynomially bounded in terms of the input size.

Those Theorems have many features in common, and this is not a coincidence. A more general approach is to interpret the condition number as the inverse of the distance to the degenerate locus. This can be bounded in terms of the height of the problem instance, and in terms of the degenerate locus (degree, dimension, height). However, bound obtained this way will be no sharper and possibly worse than the direct bounds obtained by using the exact expression for the condition number.

This paper was written while the author was visiting MSRI at Berkeley. He wishes to thank MSRI for its generous support. Thanks to Bernard Deconinck, Jennifer Roveno, Paul Gross, Raquel, and very special thanks to Paulo Ney de Souza and family.

References

[1] Dario Bini and Victor Pan. *Polynomial and Matrix Computations, Vol 1.* Progress in Theoretical Computer Science. Birkhäuser, Boston, 1994.

[2] Lenore Blum, Felipe Cucker, Mike Shub, and Steve Smale. *Complexity and Real Computation.* Springer, 1998.

[3] James W. Demmel. *Applied numerical linear algebra.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[4] Nicholas J. Higham. *Accuracy and stability of numerical algorithms.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[5] Teresa Krick and Luis M Pardo. A computational method for diophantine approximation. In *Proceedings of MEGA'94*, Progress in Mathematics. Birkhauser, 1995.

[6] Serge Lang. *Survey of Diophantine Geometry.* Springer, Berlin, 1997. Corrected Second Printing.

[7] Gregorio Malajovich. On a transfer theorem for the $\mathcal{P} \neq \mathcal{NP}$ conjecture, 1998. MSRI Preprint 1998-069.

[8] Gregorio Malajovich and Jorge P. Zubelli. On the geometry of Graeffe iteration, 1997. Informes de Matemática Série B-118, IMPA.

[9] Gregorio Malajovich and Jorge P. Zubelli. Tangent Graeffe iteration, 1998. Informes de Matemática Série B-119, IMPA.

[10] Alexandre Ostrowski. Recherches sur la méthode de Graeffe et les zéros des polynomes et des séries de Laurent. *Acta Math.*, 72:99–155, 1940.

[11] Alexandre Ostrowski. Recherches sur la méthode de Graeffe et les zéros des polynomes et des séries de Laurent. Chapitres III et IV. *Acta Math.*, 72:157–257, 1940.

[12] Joseph A. Silverman. *The Arithmetic of Elliptic Curves.* Graduate Texts in Mathematics 106. Springer, New York, 1986.

[13] Steve Smale. Complexity theory and numerical analysis. *Acta Numerica*, 6:523–551, 1997.

[14] Lloyd N. Trefethen and David Bau III. *Numerical linear algebra.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[15] Hermann Weyl. *The Theory of Groups and Quantum Mechanics.* Dover Publications, Inc, c.1931, 1951.