The condition number of a randomly perturbed matrix

STOC '07

Terence Tao (UCLA)
Van Vu (Rutgers)

Well-conditioned matrices

Suppose one wants to solve the matrix equation $Mx = b$, where M is an $n \times n$ matrix and the vector b is given.

In theory, this problem is solvable quickly (e.g. by Gaussian elimination) whenever M is non-singular.

In practice, computers can only represent a finite subset of the real numbers, and so one must take into account roundoff error. The effect of this error is controlled by the condition number

$$\kappa(M) := \|M\| \|M^{-1}\|$$

where $\|\|$ is the spectral norm. (We adopt the convention $\kappa(M) := \infty$ when M is singular.)

Let $\varepsilon_{\text{machine}}$ which is half of the distance from 1 to the nearest represented number in one's machine (a typical value is 10^{-30}). Then we have the following fundamental result in numerical linear algebra:

Theorem. If $\tilde{x}$ is the numerical solution to $Mx = b$, then

$$\frac{\|\tilde{x} - x\|}{\|x\|} = O\big(\kappa(M)\varepsilon_{\text{machine}}\big).$$

Thus upper bounds on the condition number implies numerical stability in linear algebra. (It also affects the running time of numerical linear algebra algorithms.)

Definition. A matrix M is polynomially well-conditioned if $\kappa(M) = O(n^{O(1)})$.

Suppose M is polynomial size (thus each entry of M is $O(n^{O(1)})$). Then we clearly have $\|M\| = O(n^{O(1)})$. So, being polynomially well-conditioned is usually equivalent to the bound

$$\|M^{-1}\| = O(n^{O(1)}),$$

or equivalently, a lower bound

$$\sigma_n \gg n^{-O(1)}$$

on the least singular value of M.

In theory, ill-conditioned matrices exist:

> **Theorem.** (Alon-Vu, 1996) There exists an invertible matrix M with coefficients ± 1 with $\|M^{-1}\| \gg n^{(\frac{1}{2}+o(1))n}$. In particular, $\kappa(M) \gg n^{(\frac{1}{2}+o(1))n}$.

But in practice, they only seem to arise very rarely.

In fact, linear algebraic algorithms (e.g. the simplex method) frequently run faster (and gives higher accuracy) than the worst case analysis predicts.

Why should this be the case?

The positive effect of noise

Spielman and Teng (2002) proposed the following general explanation:

> **(P)** Let M be an arbitrary $n \times n$ matrix of polynomial size and N_n a non-trivial random $n \times n$ matrix. Then with high probability $M + N_n$ is polynomially well conditioned.

Thus, the inherent measurement or roundoff error in the matrix M itself should cause one to avoid the highly ill-conditioned matrices.

The crucial point here is that M itself may have a large condition number, or even be singular (e.g. $M = 0$).

Continuous and discrete noise

Demmel (1988) established (P) when $M = 0$ and N_n is a Gaussian random matrix. Spielman and Terng (2002) established (P) for arbitrary M of polynomial size and Gaussian random M_n.

In applications to numerical linear algebra, it is more realistic to consider discrete models for the random matrix N_n. In particular we have the Bernoulli random matrix model in which each entry of N_n is ± 1 with independent uniform probability.

With Van Vu, we were able to establish (P) for arbitrary M of polynomial size and for Bernoulli random M_n. More precisely:

Theorem. (T.-Vu, 2007) Let M be polynomial size with integer coefficients, let N_n be a random Bernoulli matrix, and let $A > 0$. Then we have

$$\mathbf{P}(\|(M + N_n)^{-1}\| \geq n^B) \ll n^{-A}$$

if B is sufficiently large depending on A (and on the polynomial size of M).

In particular, by making B a bit bigger, we have $\kappa(M + N_n) = O(n^B)$ with probability $1 - O(n^{-A})$.

For Gaussian noise, the above theorem was proven by Spielman and Terng with $B = A - 1/2$.

The theorem generalises to some other discrete models, where each coordinate a_{jk} of N_n is an independent integer-valued random variable of polynomial size. One needs a large fraction of these random variables to be non-degenerate, e.g. the a_{jk} are symmetric and $\mathbb{P}(a_{jk} = +1) \geq \varepsilon$ for all but $n^{0.01}$ of the coordinates a_{jk} (thus N_n is allowed to have some "frozen" entries). There are more general versions of these results but they get a bit technical to state. One can also allow M to have complex entries instead of integer (this is a work in progress; some results in this direction were obtained recently by Pan and Zhou).

Some ingredients of the proof

Let $M_n := M + N_n$ be the noisy matrix. The goal is to show that $\|M_n^{-1}\| \ll n^B$ with probability $1 - O(n^{-A)})$, for some sufficiently large B. Thus we would like to upper bound the

$$\mathbf{P}(\|M_n v\| \ll n^{-B} \text{for some bounded vector } v)$$

by $O(n^{-A})$.

There are infinitely many unit vectors v, but one can use rounding and only have to deal with those v whose coefficients are a multiple of n^{-B-2} (say).

Some vectors v will be singular (most of the coordinates are rather small). These can be easily dealt with by standard concentration-of-measure, union bound, and ε-net arguments. (This idea was borrowed from Litvak-Pajor-Rudelson-Vershynin (2005).)

Some vectors v will be poor, in the sense that the rows of M_n have only a low probability (e.g. at most n^{-A-4}) of being close to orthogonal to v. These can be dealt with by a conditioning argument of Komlós (1960s), fixing $n-1$ of the rows and looking at the remaining row (which is chosen carefully).

The most difficult case to handle is when v is rich (so the rows of M_n are often close to orthogonal to v) and non-singular.

Inverse Littlewood-Offord theory

To handle this case, we need to understand what vectors v are rich. In the model case when $M = 0$ and N_n is Bernoulli, this question is equivalent to asking for which numbers $v_1, \ldots, v_n$ and a is the concentration probability

$$\mathbb{P}(\pm v_1 \pm v_2 \ldots \pm v_n = a)$$

large, where the $\pm$ are n iid Bernoulli signs. This is the inverse Littlewood-Offord problem. (The forward Littlewood-Offord problem specifies $v_1, \ldots, v_n$ and a and asks to bound the concentration probability.

If the numbers $v_1, \ldots, v_n$ obey many arithmetic relations (e.g. if they are all equal), then the concentration probability tends to be large. But if the $v_1, \ldots, v_n$ are arithmetically "independent" then the concentration probability tends to be low.

There are inverse Littlewood-Offord theorems which quantify this relationship; roughly speaking, they assert that the concentration probability is large if and only if the $v_1, \ldots, v_n$ are mostly concentrated in an arithmetic progression, or a generalised arithmetic progression. These results are inspired by techniques from additive combinatorics, in particular using Fourier analysis and geometry of numbers.

Discretisation of progressions

A key technical lemma is that a generalised arithmetic progression can be "rounded off" to another arithmetic progression, whose elements are well separated from each other. For instance, consider the two-dimensional generalised arithmetic progression

$$P = \{4a + (3 + 10^{-10})b : -10^{-3} \leq a, b \leq 10^{3}\}.$$

This progression contains some very small spacings - as small as 10^{-10}. But one can round this progression off to a one-dimensional arithmetic progression

$$Q = \{n : -7 \times 10^{-3} \leq n \leq 7 \times 10^{-3}\}$$

in the sense that every element of the former is within $O(10^{-7})$ of an element of the latter.

The significance of this rounding operation is that it can convert approximate relations in P to exact relations in Q. For instance, if $x, y, z \in P$ are such that $x + y = z + O(10^{-1})$, and $x', y', z' \in Q$ are their rounded counterparts, then $x' + y'$ is exactly equal to z'.

In practice, this allows us to round off a statement such as "Mv is small" to the statement "Mv' is zero". Ultimately, this reduces the task of controlling condition numbers to the simpler task of controlling the probability that M is invertible. There is some substantial technology (dating back to Kahn, Komlos, and Szemerédi (1995)) to deal with this.