The Robbins phenomenon: p-adic stability of some nonlinear recurrences

Kiran S. Kedlaya
in joint work with Joe Buhler

Department of Mathematics, University of California, San Diego
kedlaya@ucsd.edu
http://math.ucsd.edu/~kedlaya/slides/

Microsoft Research
Redmond, July 24, 2012

Preprint in preparation.

Supported by NSF (grant DMS-1101343), UCSD (Warschawski chair).

Contents

1. p-adic numbers and floating-point arithmetic

2. Condensation of determinants and the Robbins phenomenon

3. The Robbins phenomenon, and some more examples

4. Some notes on the proofs

Contents

1. p-adic numbers and floating-point arithmetic

2. Condensation of determinants and the Robbins phenomenon

3. The Robbins phenomenon, and some more examples

4. Some notes on the proofs

The p-adic numbers

Throughout this talk, $\mathbb{Z}_p$ will be the *ring of p-adic integers*. We may construct $\mathbb{Z}_p$ in one of three equivalent ways.

- Take strings composed of $0, \dots, p-1$ which run infinitely far to the left, performing arithmetic using the usual rules of base p arithmetic. For instance, for $p = 2$, the string $\cdots 11111$ represents an additive inverse of 1.
- Take sequences $(x_1, x_2, \dots)$ in which $x_n \in \mathbb{Z}/p^n\mathbb{Z}$ and $x_{n+1} \equiv x_n \pmod{p^n}$. (That is, take the *inverse limit* of the rings $\mathbb{Z}/p^n\mathbb{Z}$.)
- Take the completion of $\mathbb{Z}$ for the *p-adic absolute value* $|n|_p = p^{-v_p(n)}$, where v_p denotes the *p-adic valuation* (the exponent of p in the prime factorization of n).

The ring $\mathbb{Q}_p = \mathbb{Z}_p[p^{-1}]$ is a field, called the *field of p-adic numbers*. It is the completion of $\mathbb{Q}$ for the p-adic absolute value.

The p-adic numbers

Throughout this talk, $\mathbb{Z}_p$ will be the *ring of p-adic integers*. We may construct $\mathbb{Z}_p$ in one of three equivalent ways.

- Take strings composed of $0, \dots, p-1$ which run infinitely far to the left, performing arithmetic using the usual rules of base p arithmetic. For instance, for $p = 2$, the string $\cdots 11111$ represents an additive inverse of 1.
- Take sequences $(x_1, x_2, \dots)$ in which $x_n \in \mathbb{Z}/p^n\mathbb{Z}$ and $x_{n+1} \equiv x_n \pmod{p^n}$. (That is, take the *inverse limit* of the rings $\mathbb{Z}/p^n\mathbb{Z}$.)
- Take the completion of $\mathbb{Z}$ for the *p-adic absolute value* $|n|_p = p^{-v_p(n)}$, where v_p denotes the *p-adic valuation* (the exponent of p in the prime factorization of n).

The ring $\mathbb{Q}_p = \mathbb{Z}_p[p^{-1}]$ is a field, called the *field of p-adic numbers*. It is the completion of $\mathbb{Q}$ for the p-adic absolute value.

The p-adic numbers

Throughout this talk, $\mathbb{Z}_p$ will be the *ring of p-adic integers*. We may construct $\mathbb{Z}_p$ in one of three equivalent ways.

- Take strings composed of $0, \dots, p-1$ which run infinitely far to the left, performing arithmetic using the usual rules of base p arithmetic. For instance, for $p = 2$, the string $\cdots 11111$ represents an additive inverse of 1.
- Take sequences $(x_1, x_2, \dots)$ in which $x_n \in \mathbb{Z}/p^n\mathbb{Z}$ and $x_{n+1} \equiv x_n \pmod{p^n}$. (That is, take the *inverse limit* of the rings $\mathbb{Z}/p^n\mathbb{Z}$.)
- Take the completion of $\mathbb{Z}$ for the *p-adic absolute value* $|n|_p = p^{-v_p(n)}$, where v_p denotes the *p-adic valuation* (the exponent of p in the prime factorization of n).

The ring $\mathbb{Q}_p = \mathbb{Z}_p[p^{-1}]$ is a field, called the *field of p-adic numbers*. It is the completion of $\mathbb{Q}$ for the p-adic absolute value.

The p-adic numbers

Throughout this talk, $\mathbb{Z}_p$ will be the *ring of p-adic integers*. We may construct $\mathbb{Z}_p$ in one of three equivalent ways.

- Take strings composed of $0, \dots, p-1$ which run infinitely far to the left, performing arithmetic using the usual rules of base p arithmetic. For instance, for $p = 2$, the string $\cdots 11111$ represents an additive inverse of 1.
- Take sequences $(x_1, x_2, \dots)$ in which $x_n \in \mathbb{Z}/p^n\mathbb{Z}$ and $x_{n+1} \equiv x_n \pmod{p^n}$. (That is, take the *inverse limit* of the rings $\mathbb{Z}/p^n\mathbb{Z}$.)
- Take the completion of $\mathbb{Z}$ for the *p-adic absolute value* $|n|_p = p^{-v_p(n)}$, where v_p denotes the *p-adic valuation* (the exponent of p in the prime factorization of n).

The ring $\mathbb{Q}_p = \mathbb{Z}_p[p^{-1}]$ is a field, called the *field of p-adic numbers*. It is the completion of $\mathbb{Q}$ for the p-adic absolute value.

The p-adic numbers

Throughout this talk, $\mathbb{Z}_p$ will be the *ring of p-adic integers*. We may construct $\mathbb{Z}_p$ in one of three equivalent ways.

- Take strings composed of $0, \dots, p-1$ which run infinitely far to the left, performing arithmetic using the usual rules of base p arithmetic. For instance, for $p = 2$, the string $\cdots 11111$ represents an additive inverse of 1.
- Take sequences $(x_1, x_2, \dots)$ in which $x_n \in \mathbb{Z}/p^n\mathbb{Z}$ and $x_{n+1} \equiv x_n \pmod{p^n}$. (That is, take the *inverse limit* of the rings $\mathbb{Z}/p^n\mathbb{Z}$.)
- Take the completion of $\mathbb{Z}$ for the *p-adic absolute value* $|n|_p = p^{-v_p(n)}$, where v_p denotes the *p-adic valuation* (the exponent of p in the prime factorization of n).

The ring $\mathbb{Q}_p = \mathbb{Z}_p[p^{-1}]$ is a field, called the *field of p-adic numbers*. It is the completion of $\mathbb{Q}$ for the p-adic absolute value.

p-adic numbers in number theory

The p-adic numbers were introduced by Hensel in the early 1900s as a way to translate ideas from analysis into number theory. For example, for $p \neq 2$, if $n \in \mathbb{Z}$ is congruent to a perfect square modulo p, it is a square in $\mathbb{Z}_p$, and its square roots can be constructed using an analogue of the Newton-Raphson-Simpson iteration (i.e., finding a root of $f(x) = 0$ using the iteration $x \mapsto x - f(x)/f'(x)$).

More recently, p-adic numbers have also been used profitably in computational number theory (and cryptographic applications). For example, algorithms based on p-adic numbers for computing zeta functions of elliptic and hyperelliptic curves have been considered by Satoh, Mestre, Lauder-Wan, Kedlaya, Denef-Vercauteren, and others, and are implemented in such systems as *Pari*, *Magma*, and *Sage*.

p-adic numbers in number theory

The *p*-adic numbers were introduced by Hensel in the early 1900s as a way to translate ideas from analysis into number theory. For example, for $p \neq 2$, if $n \in \mathbb{Z}$ is congruent to a perfect square modulo *p*, it is a square in $\mathbb{Z}_p$, and its square roots can be constructed using an analogue of the Newton-Raphson-Simpson iteration (i.e., finding a root of $f(x) = 0$ using the iteration $x \mapsto x - f(x)/f'(x)$).

More recently, *p*-adic numbers have also been used profitably in computational number theory (and cryptographic applications). For example, algorithms based on *p*-adic numbers for computing zeta functions of elliptic and hyperelliptic curves have been considered by Satoh, Mestre, Lauder-Wan, Kedlaya, Denef-Vercauteren, and others, and are implemented in such systems as *Pari*, *Magma*, and *Sage*.

p-adic floating-point arithmetic

There is an obvious difficulty in computing with p-adic numbers. Just like real numbers, p-adic numbers are represented by infinite strings and so cannot be stored exactly on a computer.

There are several possible schemes for systematically approximating p-adic numbers with exact rational numbers. The one we consider in this talk is the p-adic analogue of *floating-point arithmetic* (or of *scientific notation*).

Fix a positive integer r (the *maximum relative precision*). We approximate an arbitrary p-adic number by a rational number of the form $p^e m$ where e is an integer (the *exponent*) and m is an integer in the range $\{0, \dots, p^r - 1\}$ not divisible by p (the *mantissa*).

p-adic floating-point arithmetic

There is an obvious difficulty in computing with p-adic numbers. Just like real numbers, p-adic numbers are represented by infinite strings and so cannot be stored exactly on a computer.

There are several possible schemes for systematically approximating p-adic numbers with exact rational numbers. The one we consider in this talk is the p-adic analogue of *floating-point arithmetic* (or of *scientific notation*).

Fix a positive integer r (the *maximum relative precision*). We approximate an arbitrary p-adic number by a rational number of the form $p^e m$ where e is an integer (the *exponent*) and m is an integer in the range $\{0, \dots, p^r - 1\}$ not divisible by p (the *mantissa*).

p-adic floating-point arithmetic

There is an obvious difficulty in computing with p-adic numbers. Just like real numbers, p-adic numbers are represented by infinite strings and so cannot be stored exactly on a computer.

There are several possible schemes for systematically approximating p-adic numbers with exact rational numbers. The one we consider in this talk is the p-adic analogue of *floating-point arithmetic* (or of *scientific notation*).

Fix a positive integer r (the *maximum relative precision*). We approximate an arbitrary p-adic number by a rational number of the form $p^e m$ where e is an integer (the *exponent*) and m is an integer in the range $\{0, \dots, p^r - 1\}$ not divisible by p (the *mantissa*).

Accuracy of approximations

By the *accuracy* of a p-adic floating-point approximation $p^e m$ to a p-adic number x, we will mean the integer

$$\max\{0, v_p(m - p^{-e}x)\}.$$

This counts the number of correct p-adic digits of the mantissa starting from the right. For instance, here are the accuracies of some approximations of -1 when $p = 2$:

$2^0 \cdot 1010111_2$	accuracy 3
$2^0 \cdot 1010101_2$	accuracy 1
$2^0 \cdot 1011100_2$	invalid (last digit should be nonzero)
$2^1 \cdot 1011101_2$	accuracy 0 (wrong exponent)

Addition and multiplication in floating-point arithmetic

Given p-adic floating-point approximations $p^{e_1}m_1, p^{e_2}m_2$ of $x, y \in \mathbb{Q}_p$, we may take $p^{e_1+e_2}m_1m_2$ as a floating-point approximation of xy. The accuracy of this approximation is no less than the minimum accuracy among the original approximations. (One might say that multiplication in floating-point arithmetic is *exact*.)

One can similarly obtain a floating-point approximation to $x + y$ by dividing out the maximum power of p from $p^{e_1}m_1 + p^{e_2}m_2$ and then rounding the mantissa if needed. In case $e_1 < e_2$, the final approximation is $p^{e_1}[m_1 + p^{e_2-e_1}m_2]$, where the brackets denote rounding, and we see that the accuracy is no less than the minimum accuracy among the original approximations. A similar statement holds if $e_1 > e_2$.

Addition and multiplication in floating-point arithmetic

Given p-adic floating-point approximations $p^{e_1} m_1, p^{e_2} m_2$ of $x, y \in \mathbb{Q}_p$, we may take $p^{e_1+e_2} m_1 m_2$ as a floating-point approximation of xy. The accuracy of this approximation is no less than the minimum accuracy among the original approximations. (One might say that multiplication in floating-point arithmetic is *exact*.)

One can similarly obtain a floating-point approximation to $x + y$ by dividing out the maximum power of p from $p^{e_1} m_1 + p^{e_2} m_2$ and then rounding the mantissa if needed. In case $e_1 < e_2$, the final approximation is $p^{e_1}[m_1 + p^{e_2 - e_1} m_2]$, where the brackets denote rounding, and we see that the accuracy is no less than the minimum accuracy among the original approximations. A similar statement holds if $e_1 > e_2$.

Loss of accuracy in *p*-adic arithmetic

If $e_1 = e_2$, we may experience a precision loss when computing a floating-point approximation of $x + y$. This is because $p^{e_1}(m_1 + m_2)$ is only a valid floating-point approximation if $m_1 + m_2$ is not divisible by p. If $v_p(m_1 + m_2) = f > 0$, we must shift a power of p^f from $m_1 + m_2$ into p^{e_1} before rounding; this has the effect of adding f garbage digits at the left of the mantissa.

If one performs a sequence of arithmetic operations using *p*-adic floating-point arithmetic, one may experience progressive loss of accuracy over the course of the computation. The study of such loss of accuracy amounts to a *p*-adic version of the field of *numerical stability*.

In the rest of this talk, we consider some examples of unexpected numerical stability in *p*-adic floating-point arithmetic. These appear to have a deep algebraic origin which is not yet fully understood.

Loss of accuracy in *p*-adic arithmetic

If $e_1 = e_2$, we may experience a precision loss when computing a floating-point approximation of $x + y$. This is because $p^{e_1}(m_1 + m_2)$ is only a valid floating-point approximation if $m_1 + m_2$ is not divisible by p. If $v_p(m_1 + m_2) = f > 0$, we must shift a power of p^f from $m_1 + m_2$ into p^{e_1} before rounding; this has the effect of adding f garbage digits at the left of the mantissa.

If one performs a sequence of arithmetic operations using p-adic floating-point arithmetic, one may experience progressive loss of accuracy over the course of the computation. The study of such loss of accuracy amounts to a p-adic version of the field of *numerical stability*.

In the rest of this talk, we consider some examples of unexpected numerical stability in p-adic floating-point arithmetic. These appear to have a deep algebraic origin which is not yet fully understood.

Loss of accuracy in p-adic arithmetic

If $e_1 = e_2$, we may experience a precision loss when computing a floating-point approximation of $x + y$. This is because $p^{e_1}(m_1 + m_2)$ is only a valid floating-point approximation if $m_1 + m_2$ is not divisible by p. If $v_p(m_1 + m_2) = f > 0$, we must shift a power of p^f from $m_1 + m_2$ into p^{e_1} before rounding; this has the effect of adding f garbage digits at the left of the mantissa.

If one performs a sequence of arithmetic operations using p-adic floating-point arithmetic, one may experience progressive loss of accuracy over the course of the computation. The study of such loss of accuracy amounts to a p-adic version of the field of *numerical stability*.

In the rest of this talk, we consider some examples of unexpected numerical stability in p-adic floating-point arithmetic. These appear to have a deep algebraic origin which is not yet fully understood.

Contents

1. *p*-adic numbers and floating-point arithmetic

2. Condensation of determinants and the Robbins phenomenon

3. The Robbins phenomenon, and some more examples

4. Some notes on the proofs

An identity of Jacobi

Let M be an $n \times n$ matrix. Let A, B, C, D be the determinants of the top left, top right, bottom left, bottom right $(n-1) \times (n-1)$-submatrices of M. Let E be the determinant of the central $(n-2) \times (n-2)$-submatrix of M. Let F be the determinant of M. Then

$$AD - BC = EF.$$

An identity of Jacobi: an example

For example, for

$$M = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 3 & -1 \\ 1 & 1 & 3 \end{pmatrix}$$

we have

$$A = \det \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} = 3, \quad B = \det \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix} = -5,$$
$$C = \det \begin{pmatrix} 0 & 3 \\ 1 & 1 \end{pmatrix} = -3, \quad D = \det \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} = 10,$$
$$E = 3, \quad F = \det M = 9 - 2 + 0 - 3 - (-1) - 0 = 5,$$
$$AD - BC = 30 - 15 = 15 = 3 \cdot 5 = EF.$$

The Dodgson condensation recurrence, with an example

Charles Dodgson (Lewis Carroll) proposed to use Jacobi's identity as a method to compute determinants as follows. Given a square matrix M, we successively compute the connected minors of size k from those of size $k-1$ and $k-2$. (The minors of size 0 are all equal to 1; the minors of size 1 are the entries of M.) This produces a sequence of matrices of decreasing size (hence the name *condensation*), ending with $(\det(M))$. E.g.,

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 & -2 & 2 \\ 1 & -3 & 2 & 1 \\ -2 & 5 & -3 & -2 \\ 1 & 1 & 2 & -1 \end{pmatrix},$$

$$\begin{pmatrix} -7 & -4 & -6 \\ -1 & -1 & -1 \\ -7 & 13 & 7 \end{pmatrix}, \begin{pmatrix} -1 & -1 \\ -4 & -2 \end{pmatrix}, (2).$$

Pros and cons of condensation

Some advantages of condensation:

- It is an $O(n^3)$ algorithm, just like Gaussian elimination.
- All intermediate terms belong to the same ring as the entries of the original matrix.
- For instance, if M has integer entries, one does not encounter any denominators. This helps reduce the size of the numbers involved in the computation (and provides an error check when working by hand).
- Condensation can be carried out in parallel using very little communication.

There is one serious disadvantage, though: condensation does not always work! If one encounters an instance of $AD - BC = EF$ with $E = 0$, one cannot solve for F.

Pros and cons of condensation

Some advantages of condensation:

- It is an $O(n^3)$ algorithm, just like Gaussian elimination.
- All intermediate terms belong to the same ring as the entries of the original matrix.
  - For instance, if M has integer entries, one does not encounter any denominators. This helps reduce the size of the numbers involved in the computation (and provides an error check when working by hand).
  - Condensation can be carried out in parallel using very little communication.

There is one serious disadvantage, though: condensation does not always work! If one encounters an instance of $AD - BC = EF$ with $E = 0$, one cannot solve for F.

Pros and cons of condensation

Some advantages of condensation:

- It is an $O(n^3)$ algorithm, just like Gaussian elimination.
- All intermediate terms belong to the same ring as the entries of the original matrix.
- For instance, if M has integer entries, one does not encounter any denominators. This helps reduce the size of the numbers involved in the computation (and provides an error check when working by hand).
- Condensation can be carried out in parallel using very little communication.

There is one serious disadvantage, though: condensation does not always work! If one encounters an instance of $AD - BC = EF$ with $E = 0$, one cannot solve for F.

Pros and cons of condensation

Some advantages of condensation:

- It is an $O(n^3)$ algorithm, just like Gaussian elimination.
- All intermediate terms belong to the same ring as the entries of the original matrix.
- For instance, if M has integer entries, one does not encounter any denominators. This helps reduce the size of the numbers involved in the computation (and provides an error check when working by hand).
- Condensation can be carried out in parallel using very little communication.

There is one serious disadvantage, though: condensation does not always work! If one encounters an instance of $AD - BC = EF$ with $E = 0$, one cannot solve for F.

Pros and cons of condensation

Some advantages of condensation:

- It is an $O(n^3)$ algorithm, just like Gaussian elimination.
- All intermediate terms belong to the same ring as the entries of the original matrix.
- For instance, if M has integer entries, one does not encounter any denominators. This helps reduce the size of the numbers involved in the computation (and provides an error check when working by hand).
- Condensation can be carried out in parallel using very little communication.

There is one serious disadvantage, though: condensation does not always work! If one encounters an instance of $AD - BC = EF$ with $E = 0$, one cannot solve for F.

Pros and cons of condensation

Some advantages of condensation:

- It is an $O(n^3)$ algorithm, just like Gaussian elimination.
- All intermediate terms belong to the same ring as the entries of the original matrix.
- For instance, if M has integer entries, one does not encounter any denominators. This helps reduce the size of the numbers involved in the computation (and provides an error check when working by hand).
- Condensation can be carried out in parallel using very little communication.

There is one serious disadvantage, though: condensation does not always work! If one encounters an instance of $AD - BC = EF$ with $E = 0$, one cannot solve for F.

Condensation and p-adic numbers

David Robbins noticed that over $\mathbb{F}_p$, one can work around the occurrence of zero denominators by lifting the problem to $\mathbb{Z}$, so that minors which start out equal to 0 have a chance to lift to nonzero values. However, doing this computation exactly requires dealing with unpleasantly large integers.

Since he only wanted an answer over $\mathbb{F}_p$, Robbins proposed to replace exact arithmetic in $\mathbb{Z}$ with floating-point arithmetic in $\mathbb{Q}_p$ using a fairly small relative precision (e.g., one which fits in a machine word). To get an answer from this, one must guarantee that the resulting approximation of the determinant has accuracy at least 1.

Robbins was thus led to test the numerical stability of condensation directly, leading to a surprising observation: accuracy losses in condensation do not appear to accumulate as one might expect!

Condensation and *p*-adic numbers

David Robbins noticed that over $\mathbb{F}_p$, one can work around the occurrence of zero denominators by lifting the problem to $\mathbb{Z}$, so that minors which start out equal to 0 have a chance to lift to nonzero values. However, doing this computation exactly requires dealing with unpleasantly large integers.

Since he only wanted an answer over $\mathbb{F}_p$, Robbins proposed to replace exact arithmetic in $\mathbb{Z}$ with floating-point arithmetic in $\mathbb{Q}_p$ using a fairly small relative precision (e.g., one which fits in a machine word). To get an answer from this, one must guarantee that the resulting approximation of the determinant has accuracy at least 1.

Robbins was thus led to test the numerical stability of condensation directly, leading to a surprising observation: accuracy losses in condensation do not appear to accumulate as one might expect!

Condensation and *p*-adic numbers

David Robbins noticed that over $\mathbb{F}_p$, one can work around the occurrence of zero denominators by lifting the problem to $\mathbb{Z}$, so that minors which start out equal to 0 have a chance to lift to nonzero values. However, doing this computation exactly requires dealing with unpleasantly large integers.

Since he only wanted an answer over $\mathbb{F}_p$, Robbins proposed to replace exact arithmetic in $\mathbb{Z}$ with floating-point arithmetic in $\mathbb{Q}_p$ using a fairly small relative precision (e.g., one which fits in a machine word). To get an answer from this, one must guarantee that the resulting approximation of the determinant has accuracy at least 1.

Robbins was thus led to test the numerical stability of condensation directly, leading to a surprising observation: accuracy losses in condensation do not appear to accumulate as one might expect!

Unexpected numerical stability: an observation of Robbins

Let M be a square matrix with entries in $\mathbb{Z}_p$. Represent each entry with a p-adic floating-point approximation of accuracy at least r, then compute the condensation recurrence using floating-point arithmetic. Let d be the maximum p-adic valuation of any denominator occurring in the recurrence. Let a denote the absolute accuracy of the computed determinant, i.e., the p-adic valuation of its difference from $\det(M)$.

Conjecture (Robbins, 2005)

We have $a \geq r - d$. (Experiments show that this inequality is sharp.)

What is surprising is that d is typically much less than the sum of the accumulated losses of accuracy over individual arithmetic steps!

Theorem (Buhler-K, 2012)

We have $a \geq r - 3d$. (This is proved as a special case of a more general result, more on which later.)

Unexpected numerical stability: an observation of Robbins

Let M be a square matrix with entries in $\mathbb{Z}_p$. Represent each entry with a p-adic floating-point approximation of accuracy at least r, then compute the condensation recurrence using floating-point arithmetic. Let d be the maximum p-adic valuation of any denominator occurring in the recurrence. Let a denote the absolute accuracy of the computed determinant, i.e., the p-adic valuation of its difference from $\det(M)$.

Conjecture (Robbins, 2005)

We have $a \geq r - d$. (Experiments show that this inequality is sharp.)

What is surprising is that d is typically much less than the sum of the accumulated losses of accuracy over individual arithmetic steps!

Theorem (Buhler-K, 2012)

We have $a \geq r - 3d$. (This is proved as a special case of a more general result, more on which later.)

Unexpected numerical stability: an observation of Robbins

Let M be a square matrix with entries in $\mathbb{Z}_p$. Represent each entry with a p-adic floating-point approximation of accuracy at least r, then compute the condensation recurrence using floating-point arithmetic. Let d be the maximum p-adic valuation of any denominator occurring in the recurrence. Let a denote the absolute accuracy of the computed determinant, i.e., the p-adic valuation of its difference from $\det(M)$.

Conjecture (Robbins, 2005)

We have $a \geq r - d$. (Experiments show that this inequality is sharp.)

What is surprising is that d is typically much less than the sum of the accumulated losses of accuracy over individual arithmetic steps!

Theorem (Buhler-K, 2012)

We have $a \geq r - 3d$. (This is proved as a special case of a more general result, more on which later.)

Contents

1. p-adic numbers and floating-point arithmetic
2. Condensation of determinants and the Robbins phenomenon
3. The Robbins phenomenon, and some more examples
4. Some notes on the proofs

Another example: the Somos-4 recurrence

It was observed by Michael Somos that for any x_0, x_1, x_2, x_3 which are units in an integral domain R, if we define the sequence

$$x_{n+4} = \frac{x_{n+1}x_{n+3} + x_{n+2}^2}{x_n} \qquad (n = 0, 1, \dots),$$

then $x_n \in R$ for all n. (This can be proved using elliptic curves.)

Now take $R = \mathbb{Z}_p$. Represent each initial term of the recurrence with a p-adic floating-point approximation of accuracy at least r, then compute the recurrence out to x_n using floating-point arithmetic. Let d be the maximum p-adic valuation of any denominator occurring in the recurrence. Let a denote the absolute accuracy of the computed value of x_n.

Theorem (Buhler-K, 2012)

We have $a \geq r - d$. (Experiments show that this inequality is sharp.)

Another example: the Somos-4 recurrence

It was observed by Michael Somos that for any x_0, x_1, x_2, x_3 which are units in an integral domain R, if we define the sequence

$$x_{n+4} = \frac{x_{n+1}x_{n+3} + x_{n+2}^2}{x_n} \qquad (n = 0, 1, \dots),$$

then $x_n \in R$ for all n. (This can be proved using elliptic curves.)

Now take $R = \mathbb{Z}_p$. Represent each initial term of the recurrence with a p-adic floating-point approximation of accuracy at least r, then compute the recurrence out to x_n using floating-point arithmetic. Let d be the maximum p-adic valuation of any denominator occurring in the recurrence. Let a denote the absolute accuracy of the computed value of x_n.

Theorem (Buhler-K, 2012)

We have $a \geq r - d$. (Experiments show that this inequality is sharp.)

Weak and strong versions of the Robbins phenomenon

One can similarly define a, r, d for any recurrence defined by rational functions over $\mathbb{Z}_p$. If we always have $a \geq r - d$, we say that the recurrence exhibits the *strong Robbins phenomenon*. If we only have $a \geq r - cd$ for some fixed constant c (depending on the recurrence but not the initial terms), we say that the recurrence the *weak Robbins phenomenon with correction factor c*.

For example, the Somos-6 recurrence

$$x_{n+6} = \frac{x_{n+1}x_{n+5} + x_{n+2}x_{n+4} + x_{n+3}^2}{x_n} \qquad (n = 0, 1, \dots)$$

again has unexpected integrality: if $x_0, \dots, x_5$ are units in an integral domain R, then $x_n \in R$ for all n. One observes experimentally that the weak Robbins phenomenon holds with correction factor 2; our results only imply the weak Robbins phenomenon with correction factor 5.

Weak and strong versions of the Robbins phenomenon

One can similarly define a, r, d for any recurrence defined by rational functions over $\mathbb{Z}_p$. If we always have $a \geq r - d$, we say that the recurrence exhibits the *strong Robbins phenomenon*. If we only have $a \geq r - cd$ for some fixed constant c (depending on the recurrence but not the initial terms), we say that the recurrence the *weak Robbins phenomenon with correction factor c*.

For example, the Somos-6 recurrence

$$x_{n+6} = \frac{x_{n+1}x_{n+5} + x_{n+2}x_{n+4} + x_{n+3}^2}{x_n} \qquad (n = 0, 1, \dots)$$

again has unexpected integrality: if $x_0, \dots, x_5$ are units in an integral domain R, then $x_n \in R$ for all n. One observes experimentally that the weak Robbins phenomenon holds with correction factor 2; our results only imply the weak Robbins phenomenon with correction factor 5.

The Laurent phenomenon

There are a large number of recurrences computed by rational functions with the property that their terms can be expressed as Laurent polynomials in the initial data. These recurrences are said to exhibit the *Laurent phenomenon*.

Theorem (Buhler-K, 2012)

Any recurrence which can be shown to exhibit the Laurent phenomenon using the **caterpillar lemma** *of Fomin-Zelevinsky also exhibits the weak Robbins phenomenon for some correction factor. (This factor is explicit but typically not optimal.)*

By contrast, recurrences not exhibiting the Laurent phenomenon typically do not exhibit the weak Robbins phenomenon either; the accuracies of p-adic floating-point approximations exhibit the progressive degradation one would normally expect.

The Laurent phenomenon

There are a large number of recurrences computed by rational functions with the property that their terms can be expressed as Laurent polynomials in the initial data. These recurrences are said to exhibit the *Laurent phenomenon*.

Theorem (Buhler-K, 2012)

Any recurrence which can be shown to exhibit the Laurent phenomenon using the **caterpillar lemma** *of Fomin-Zelevinsky also exhibits the weak Robbins phenomenon for some correction factor. (This factor is explicit but typically not optimal.)*

By contrast, recurrences not exhibiting the Laurent phenomenon typically do not exhibit the weak Robbins phenomenon either; the accuracies of p-adic floating-point approximations exhibit the progressive degradation one would normally expect.

The Laurent phenomenon

There are a large number of recurrences computed by rational functions with the property that their terms can be expressed as Laurent polynomials in the initial data. These recurrences are said to exhibit the *Laurent phenomenon*.

Theorem (Buhler-K, 2012)

Any recurrence which can be shown to exhibit the Laurent phenomenon using the **caterpillar lemma** *of Fomin-Zelevinsky also exhibits the weak Robbins phenomenon for some correction factor. (This factor is explicit but typically not optimal.)*

By contrast, recurrences not exhibiting the Laurent phenomenon typically do not exhibit the weak Robbins phenomenon either; the accuracies of p-adic floating-point approximations exhibit the progressive degradation one would normally expect.

Binomial recurrences

Among recurrences satisfying the Laurent phenomenon, many have the property that the recurrence is computed as the sum of two monomials in prior terms divided by a single prior term. Such recurrences are said to be *binomial*.

Conjecture (Buhler-K, 2012)

Any binomial recurrence which can be shown to exhibit the Laurent phenomenon using a **cluster algebra** *of Fomin-Zelevinsky also exhibits the strong Robbins phenomenon.*

For example, condensation and Somos-4 are governed by cluster algebras. Somos-6 is not (it is not binomial), but the no-middle-term variant

$$x_{n+6} = \frac{x_{n+1}x_{n+5} + x_{n+2}x_{n+4}}{x_n} \qquad (n = 0, 1, \dots)$$

is governed by a cluster algebra, and experimentally exhibits the strong Robbins phenomenon.

Contents

1. *p*-adic numbers and floating-point arithmetic
2. Condensation of determinants and the Robbins phenomenon
3. The Robbins phenomenon, and some more examples
4. Some notes on the proofs

The Laurent phenomenon for Somos-4

The Fomin-Zelevinsky *caterpillar lemma* implies the Laurent phenomenon for Somos-4. Explicitly, one shows that all of

$$x_n, x_{n+1}, x_{n+2}, x_{n+3},$$
$$\frac{x_{n+1}x_{n+3} + x_{n+2}^2}{x_n}, \frac{x_n x_{n+3}^2 + x_{n+2}^3}{x_{n+1}}, \frac{x_n^2 x_{n+3} + x_{n+1}^3}{x_{n+2}}, \frac{x_n x_{n+2} + x_{n+1}^2}{x_{n+3}}$$

are Laurent polynomials in x_0, x_1, x_2, x_3, by induction on n. For example,

$$\begin{aligned}\frac{x_{n+2}x_{n+4} + x_{n+3}^2}{x_{n+1}} &= \frac{x_{n+2}(x_{n+1}x_{n+3} + x_{n+2}^2) + x_n x_{n+3}^2}{x_n x_{n+1}} \\ &= \frac{1}{x_n}\left(x_{n+2}x_{n+3} + \frac{x_{n+2}^2 + x_n x_{n+3}^2}{x_{n+1}}\right)\end{aligned}$$

but any two of $x_n, x_{n+1}, x_{n+2}, x_{n+3}$ generate the unit ideal.

An algebraic model for the Robbins phenomenon

To prove the strong Robbins phenomenon for Somos-4, we introduce an algebraic model of p-adic floating-point arithmetic: compute in parallel a sequence $\{y_n\}$ with $y_n = x_n$ for $n = 0, 1, 2, 3$ but with

$$y_{n+4} = \frac{y_{n+1}y_{n+3}(1 + p^r\epsilon_{n,1}) + y_{n+2}^2(1 + p^r\epsilon_{n,2})}{y_n}$$

for some unknown $\epsilon_{n,1}, \epsilon_{n,2} \in \mathbb{Z}_p$. We then claim that

$$v_p(y_n - x_n) \geq r - \max\{v_p(y_0), \dots, v_p(y_{n-4})\}.$$

The Robbins phenomenon for Somos-4

By modifying the proof of the Robbins phenomenon, we see that if we modify the error terms as follows:

$$y_{n+4} = \frac{y_{n+1}y_{n+3}(1 + y_{n+2}y_n\epsilon_{n,1}) + y_{n+2}^2(1 + y_n y_{n+1} y_{n+3}\epsilon_{n,2})}{y_n}$$

then we have

$$y_n \in \mathbb{Z}[x_0^{\pm}, x_1^{\pm}, x_2^{\pm}, x_3^{\pm}, \epsilon_{i,j}\,(i = 0, \dots, n-4;\, j = 0, 1)].$$

This immediately implies that

$$v_p(y_n - x_n) \geq r - 3\max\{v_p(y_0), \dots, v_p(y_{n-4})\},$$

but we can eliminate the factor of 3 using the fact that no more than one of $y_n, y_{n+1}, y_{n+2}, y_{n+3}$ can have positive valuation.

The weak Robbins phenomenon for condensation

For condensation, if we write the original recurrence as

$$F = \frac{AD - BC}{E},$$

then the modified recurrence can be taken to be

$$\tilde{F} = \frac{\tilde{A}\tilde{D}(1 + \tilde{B}\tilde{C}\tilde{E}\epsilon_*) - \tilde{B}\tilde{C}(1 + \tilde{A}\tilde{D}\tilde{E}\epsilon_*)}{\tilde{E}}$$

and again each term in the recurrence is a polynomial in the matrix entries and the ϵ_*.

This implies the weak Robbins phenomenon with correction factor 3, but in this case it may happen that more than one of $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}, \tilde{E}$ has positive valuation. Since this example is related to cluster algebras, our hope is that the cluster algebra theory can be used to get a better algebraic containment result in order to deduce the strong Robbins phenomenon.

The weak Robbins phenomenon for condensation

For condensation, if we write the original recurrence as

$$F = \frac{AD - BC}{E},$$

then the modified recurrence can be taken to be

$$\tilde{F} = \frac{\tilde{A}\tilde{D}(1 + \tilde{B}\tilde{C}\tilde{E}\epsilon_*) - \tilde{B}\tilde{C}(1 + \tilde{A}\tilde{D}\tilde{E}\epsilon_*)}{\tilde{E}}$$

and again each term in the recurrence is a polynomial in the matrix entries and the ϵ_*.

This implies the weak Robbins phenomenon with correction factor 3, but in this case it may happen that more than one of $\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}, \tilde{E}$ has positive valuation. Since this example is related to cluster algebras, our hope is that the cluster algebra theory can be used to get a better algebraic containment result in order to deduce the strong Robbins phenomenon.