Condition Numbers of Numeric and Algebraic Problems

Stephen Vavasis

1Department of Combinatorics & Optimization
University of Waterloo

2011-Nov-16 / Fields Inst. Workshop on Hybrid Symbolic-Numeric Computation
Outline

1. Condition numbers in general
2. Condition numbers of linear equations
3. Linear least squares
4. Eigenvalues
5. Linear Programming
6. Geometric condition numbers
7. Polynomial evaluation and roots
Condition number definition

Given a real-number problem, that is, a function $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}^n$, the condition number of an instance means its sensitivity to small perturbation. In particular:

$$\text{cond num} = \lim_{\epsilon \to 0} \sup_{\|y\| \leq \epsilon} \frac{\|\Phi(x + y) - \Phi(x)\|}{\|\Phi(x)\| \|y\| / \|x\|}$$

(absolute measurement).

Or perhaps

$$\text{cond num} = \lim_{\epsilon \to 0} \sup_{\|y\| \leq \epsilon} \frac{\|\Phi(x + y) - \Phi(x)\|}{\|y\| / \|x\|}$$
Details to specify:

- Precise definition of input and output
- Relative or absolute? (applies to both the input and output)
- Which part of the data is perturbed?
- What norm is used to measure sensitivity?
Uses of a condition number

- Condition numbers determine the best possible accuracy of the solution in the presence of approximations made by the computation.
- Condition numbers sometimes bound the convergence speed of iterative methods.
- Condition numbers sometimes measure the distance of an instance to singularity.
- Condition numbers sometimes shed light on preconditioning.
Condition numbers in general

Condition numbers and floating-point algorithms

- Condition numbers are properties of an instance (i.e., the data), not any particular algorithm.
- Condition numbers set achievable limits algorithms.
- Condition number analysis may indicate that certain algorithmic choices are unwise.
- Condition numbers often reveal some useful geometric properties of the instance.
Outline

1. Condition numbers in general
2. Condition numbers of linear equations
3. Linear least squares
4. Eigenvalues
5. Linear Programming
6. Geometric condition numbers
7. Polynomial evaluation and roots
Most famous classic example (Von Neumann & Goldstine; Turing) is the condition number of solving linear equations.

Function $\Phi$ is $\Phi(A, b) = A^{-1}b$, $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$.

Sensitivities understood in the relative sense. All data may be perturbed. Matrix norm is induced by vector norm.
Theorem: Suppose \( x = A^{-1}b; \)
\[ x + \Delta x = (A + \Delta A)^{-1}(b + \Delta b) \]

with
\[
\max\left( \frac{\|\Delta A\|}{\|A\|}, \frac{\|\Delta b\|}{\|b\|} \right) \leq \delta.
\]

Then
\[
\frac{\|\Delta x\|}{\|x\|} \leq 2\kappa(A)\delta + O(\delta^2)
\]

where \( \kappa(A) = \|A\| \cdot \|A^{-1}\| \) is the condition number of \( A \).
Condition number of linear equations (cont’d)

- Note \( \kappa(A) \geq 1 \) and \( \kappa(tA) = \kappa(A) \) for all \( t \neq 0 \).
- Specializing to the Euclidean vector norm and its induced matrix norm, \( \kappa(A) = \sigma_1 / \sigma_n \), the ratio of the extremal singular values of \( A \).
- Geometrically: matrix \( A \) maps the \( n \)-ball to an ellipsoid. The condition number is the ratio of the maximum to minimum axis length of this ellipsoid.
Condition number of linear equations (cont’d)

\[
\mathbf{x} \mapsto \begin{pmatrix} 4 & 2 \\ 3 & 2 \end{pmatrix} \mathbf{x}
\]

\[
\kappa \begin{pmatrix} 4 & 2 \\ 3 & 2 \end{pmatrix} = 16.4
\]
The bound on perturbation to \( x \) can be (mostly) achieved.

The bound can be achieved if only \( A \) or only \( b \) is perturbed.

Condition number does not depend on \( b \).
Condition number and distance to singularity

**Theorem:** If $A \in \mathbb{R}^{n \times n}$ is nonsingular, then $1/\kappa(A)$ is the relative distance from $A$ to singular matrices, i.e.,

$$
\frac{1}{\kappa(A)} = \inf \left\{ \frac{\|\Delta A\|}{\|A\|} : A + \Delta A \text{ is singular} \right\}.
$$

**Optimal** $\Delta A$ pushes the smallest singular value to 0.
Theorem: Suppose \( A \in \mathbb{R}^{n \times n} \) is symmetric and positive definite. Then the \( i \)th iterate of the conjugate gradient method for solving \( Ax = b \) satisfies

\[
\|x_i - A^{-1}b\|_A \leq 2\|x_0 - A^{-1}b\|_A \cdot \left( \frac{\sqrt{\kappa(A)-1}}{\sqrt{\kappa(A)+1}} \right)^i.
\]

Note: \( \|x\|_A \) means \( (x^T A x)^{1/2} \).

For a symmetric positive definite \( A \),
\[
\kappa(A) = \frac{\lambda_{\text{max}}(A)}{\lambda_{\text{min}}(A)}.
\]

Steepest descent minimization applied to \( \phi(A) = x^T A x / 2 - b^T x \) is also bounded in terms of condition number.
For nonsymmetric $A$, can apply CG to minimize $\|Ax - b\|$; equivalent to solving symmetric system $A^TAx = A^Tb$.

Note that $\kappa(A^TA) = \kappa(A)^2$.

Other well-known iterative methods for nonsymmetric $Ax = b$, e.g., GMRES, Bi-CGSTAB, are not governed by $\kappa(A)$.
Computing the condition number

- The condition number of the condition number is the condition number (Demmel).
- Means: The sensititivity of the condition number itself with respect to perturbations of $A$ is again $\kappa(A)$.
- In practice, this means that very large condition numbers (greater than $10^{17}$ in Matlab) cannot usually be computed accurately, except for matrices with special structure.
- Even for well-conditioned matrices, computing the condition number is more expensive than solving $Ax = b$. 
Instead of applying conjugate gradient to $Ax = b$, apply it to $CAC^T y = Cb$, where $C$ is a square nonsingular system; $C^T C$ is called the preconditioner.

Want $C$ such that $\kappa(CAC^T) \ll \kappa(A)$.

Too expensive to compute either quantity.

Tradeoff between time to compute/apply $C$ versus $\kappa(CAC^T)$. 
Example of preconditioning

- A symmetric $n \times n$ matrix $A$ is a *weighted Laplacian* if the diagonal entries are nonnegative, the off-diagonal entries are negative, and the row sums are nonnegative.

- Above conditions imply positive semidefiniteness.

- Spielman, Teng and others in a series of papers in the past 10 years found a graph-theoretic preconditioner for weighted Laplacians.

- Consequence is that these systems can be solved in nearly linear time (linear in the number of nonzero entries in $A$).
Extension to finite element stiffness matrices

- Boman, Hendrickson and V. extend Spielman & Teng to finite element discretizations of PDE’s of the form $\nabla \cdot (\sigma \nabla u) = -f$.

- Finite element stiffness matrix $K$ can be factored as $A^T D^{1/2} HD^{1/2} A$, where $A$ is a node-arc incidence matrix, $D$ is diagonal, positive semidefinite ($\iff A^T DA$ is a weighted Laplacian).

- If all cells of the mesh are well-shaped, then $\kappa(H)$ is small, and any preconditioner for $A^T DA$ also works for $K$. 
Outline

1. Condition numbers in general
2. Condition numbers of linear equations
3. Linear least squares
4. Eigenvalues
5. Linear Programming
6. Geometric condition numbers
7. Polynomial evaluation and roots
Problem is: minimize $\|Ax - b\|_2$ given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Equivalent to solving linear equations $A^T Ax = A^T b$ (first order optimality condition).

Assume that $\text{rank}(A) = n$, so solution is unique.

Common algorithms: method of normal equations; QR factorization
**Theorem:** If \( \frac{\|\Delta A\|}{\|A\|} \leq \delta \) and \( \frac{\|\Delta b\|}{\|b\|} \leq \delta \) then

\[
\frac{\|\Delta x\|}{\|x\|} \leq 2\kappa(A)\delta + \frac{\kappa(A)^2\delta\|b - Ax\|}{\|A\| \cdot \|x\|}
\]

due to Wedin. Achievable.

Here, \( \kappa(A) = \sigma_1(A)/\sigma_n(A) \).

If linear systems bound applied to \( A^T Ax = A^T b \), obtain a weaker bound.
Moral 1: Reducing Problem A to Problem B establishes a bound on the condition number, but the bound may be weak.

Moral 2: Solving linear least squares via reduction to linear equations may give a poor answer.
Weighted least squares

- **Weighted** least squares means minimizing \( \| D(Ax - b) \| \) where \( D \) is a positive definite diagonal weight matrix.

- Reduces to ordinary linear least squares under the obvious substitution \( \bar{A} = DA \) and \( \bar{b} = Db \).

- Perturbation bound for ordinary least squares means solution can be arbitrarily inaccurate as \( \kappa(D) \to \infty \).
**Theorem:** There is a bound on $\|x - \hat{x}\|$ that depends on $\lambda_A$ and $\bar{\lambda}_A$ and is independent of $\kappa(D)$.

These quantities were introduced by Stewart, Todd and Dikin independently.

Specialized weighted least-squares algorithms (V.; Hough & V.) can achieve this bound.

These algorithms require that dependence among rows of $A$ is detected infallibly. (Independent rows have a condition number bounded by $\lambda_A$.)
More about $\chi_A$

- $\chi_A = \sup\{\| (A^TDA)^{-1}A^TD \| : D \in \mathcal{D} \}$ and $\bar{\chi}_A = \sup\{\| A(A^TDA)^{-1}A^TD \| : D \in \mathcal{D} \}$.
- $\mathcal{D}$ denotes positive definite diagonal matrices.
- Dikin; Todd; Stewart: this sup is finite.
- Khachiyan; Tunçel: NP-hard to compute or accurately approximate $\chi_A$ or $\bar{\chi}_A$. 

$\mathcal{D}$ denotes positive definite diagonal matrices.
Outline

1. Condition numbers in general
2. Condition numbers of linear equations
3. Linear least squares
4. Eigenvalues
5. Linear Programming
6. Geometric condition numbers
7. Polynomial evaluation and roots
Condition number of eigenvalues

- If $Ax = \lambda x$ and $\lambda$ is not a repeated eigenvalue, then $\lambda$ depends smoothly on $A$. Differentiation shows that if $\lambda + \Delta \lambda$ is the corresponding eigenvalue of $A + \Delta A$, $\|\Delta A\| \leq \delta$, then ... Exercise: FIGURE IT OUT!

- Hint: Normalize eigenvector; expand $(A + \Delta A)(x + \Delta x) = (\lambda + \Delta \lambda)(x + \Delta x)$ to first order; multiply on the left by $y^H$ (normalized left eigenvector of $\lambda$).
Condition number of eigenvalues

- If $Ax = \lambda x$ and $\lambda$ is not a repeated eigenvalue, then $\lambda$ depends smoothly on $A$. Differentiation shows that if $\lambda + \Delta \lambda$ is the corresponding eigenvalue of $A + \Delta A$, $\|\Delta A\| \leq \delta$, then $|\Delta \lambda| \leq \delta/|y^Hx| + O(\delta^2)$.

- Here, $y$ is the normalized left eigenvector of $A$ and $x$ is the normalized right eigenvector.

- Absolute perturbations make sense in this context since could have $\lambda = 0$ when $A \neq 0$.

- If $A$ is a normal matrix, i.e., has an orthonormal basis of eigenvectors, then $\lambda$ is perfectly conditioned since $|y^Hx| = 1$. 
Ill conditioned eigenvalues and pseudoeigenvalues

- If an eigenvalue is ill conditioned ($|y^H x| \approx 0$), then it is sensitive to perturbations of $A$ and hard to compute.

- Eigenvalue criteria applied to Jacobian are often used to test stability of solutions to differential equations and stability of numerical methods for differential equations.

- These criteria are unreliable if they are applied to problems with a highly non-normal Jacobian (Trefethen).
Pseudoeigenvalues

Trefethen and others define an $\epsilon$-pseudoeigenvalue of $A$ to be a complex number $\lambda$ such that $\|(A - \lambda I)^{-1}\| \geq 1/\epsilon$.

Equivalent to: $\lambda$ is a $\epsilon$-pseudoeigenvalue if there exists $\Delta A$ s.t. $\|\Delta A\| \leq \epsilon$ and $\lambda$ is an eigenvalue of $A + \Delta A$.

Large condition number indicates that pseudoeigenvalues can be far from true eigenvalues.

Computational experiments and theoretical analysis show that pseudoeigenvalues can be used reliably in various stability criteria.
Outline

1. Condition numbers in general
2. Condition numbers of linear equations
3. Linear least squares
4. Eigenvalues
5. Linear Programming
6. Geometric condition numbers
7. Polynomial evaluation and roots
Linear programming refers to minimizing $c^T x$
subject to $Ax = b$ and $x \geq 0$, where $A \in \mathbb{R}^{m \times n}$ is
given, as are $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$.

Linear programming admits an elegant duality theory. Two competing algorithms for LP are in
common usage: simplex and interior point. Simplex is not known to be polynomial time.

Three possible outcomes for LP: optimal solution found; problem is unbounded; problem is infeasible.

Optimizer not necessarily unique.
Inputs and outputs for linear programming

- Input is $A, b, c$. Some subtlety here: What if constraints are structured, e.g., $l_i \leq a_i^T x \leq u_i$.
- (At least) two possible outputs from LP: optimal value $c^T x^*$ or optimizer $x^*$ itself.
- Optimizer is more sensitive to perturbation than optimal value.
Example

\[
\begin{align*}
\text{min} & \quad x + (1 + \delta)y \\
\text{s.t.} & \quad 0 \leq x, y \leq 1 \\
& \quad x + y \geq 0.3 \\
\delta &= -0.01
\end{align*}
\]
Example

\[
\begin{align*}
\text{min} & \quad x + (1 + \delta)y \\
\text{s.t.} & \quad 0 \leq x, y \leq 1 \\
& \quad x + y \geq 0.3
\end{align*}
\]
Example

\[
\begin{align*}
\text{min} & \quad x + (1 + \delta)y \\
\text{s.t.} & \quad 0 \leq x, y \leq 1 \\
& \quad x + y \geq 0.3
\end{align*}
\]
Renegar’s condition number

- Start with the *LP feasibility problem*: given $A \in \mathbb{R}^{m \times n}$ such that $\text{rank}(A) = m$ and $b \in \mathbb{R}^m$, find an $x$ such that $Ax = b, x \geq 0$.

- Let the instance be written as $d = (A, b)$. Assume it is feasible. The *condition number* of the instance according to Renegar is

$$C(d) = \frac{\|d\|}{\inf\{\|d - d'\| : d' \text{ is infeasible}\}}.$$ 

- Renegar and Vera relate this condition number to the complexity of finding a solution to LP feasibility.
Followups to Renegar’s work

- Many other problems similar to LP feasibility but with subtly different properties.
- Critique of Renegar’s theorem: $A$ and $b$ scale differently.
- More general critique: condition number should depend on $S = \{x : Ax = b\}$ but not on particular choice of $A$ and $b$. 
Epelman and Freund define a condition number for LP feasibility as $C_2(d) = 1/\text{sym}(H(d), 0)$ where $\text{sym}(D, x) = \sup\{t : x + v \in D \implies x - tv \in D\}$ and $H(d) = \{b\theta - Ax : \theta \geq 0; x \geq 0; |\theta| + \|x\| \leq 1\}$.

Existence of feasible point implies $0 \in \text{int}(H(d))$.

$\text{sym}(D, 0)$ is invariant under invertible linear mappings of $D$.

E & F show that $C_2(d) \leq C(d)$. 
Cheung, Cucker & Peña unify several LP results.

Form the following self-dual feasibility problem:

\[-c^T x + b^T y \geq 0,\]
\[c x_0 - A^T y \geq 0,\]
\[-b x_0 + A x = 0,\]
\[x_0, x \geq 0.\]

Let $\rho$ be the distance of this instance to another instance with a different strictly complementary basis (uniquely determined; Goldman-Tucker 1956).

This distance is equal or related to several previously proposed condition numbers.
V. and Ye proposed an interior point method whose running time depends only on $\chi_A$, $m$, $n$ and not on $b$, $c$.

No other interior point method has this property. Tardos had earlier shown how to get this kind of complexity with the ellipsoid method.

New technique is \textit{layered} least squares, a generalization of weighted least-squares.

Some weights infinitely larger than others.
Is $\bar{\chi}_A$ a condition number?

- Perhaps $\bar{\chi}_A$ is a condition number for linear programming, but no precise statement is known.
- Cheung et al. found relationships between $\bar{\chi}_A$ and other LP condition numbers.
- Note that $\bar{\chi}_A$ depends only on $N(A)$ and not on $A$ itself.
Semidefinite programming instances can have infinite condition number even when there is a unique optimizer.

Example: maximize $x$ subject to $y \geq 1$, $x^2 + y^2 \leq 1$.

Unique feasible point $(0, 1)$ has objective value $0$.

Perturbing constraint to $x^2 + y^2 \leq 1 + \epsilon$ moves optimizer to $(\sqrt{\epsilon}, 1)$.

Hence ratio of perturbation of root to perturbation of data is unbounded as $\epsilon \to 0$. 
In this SDP instance, the *Slater* condition fails.
In this SDP instance, the *Slater* condition fails.
1. Condition numbers in general
2. Condition numbers of linear equations
3. Linear least squares
4. Eigenvalues
5. Linear Programming
6. Geometric condition numbers
7. Polynomial evaluation and roots
A **boundary value problem** means solving a differential equation involving one or more spatial independent variables over a domain $\Omega \subset \mathbb{R}^d$ (usually $d = 1, 2, 3$) given boundary conditions and field values.

Classic example mentioned earlier:
\[ \nabla \cdot (\sigma(x) \nabla u(x)) = -f(x) \] for $x \in \Omega$ with some specification of $u$ (Dirichlet) or $\partial u/\partial n$ (Neumann) at every point on $\partial \Omega$

Special case: $\sigma(x) \equiv 1$ and $f \equiv 0$ called *Laplace’s equation*. 
Geometry and BVP solution

- Difficulty of solving BVP solution depends partly on geometry of $\Omega$. (Many technicalities regarding mesh generation, etc.)
- Clean problem: assume $\Omega$ simply connected in $\mathbb{R}^2$, the BVP is Laplace’s equation, boundary data is Dirichlet.
- Can reduce solution to inversion of integral operators whose condition number depends on the harmonic conjugation operator.
- Harmonic conjugation means: given the real part of a complex analytic function defined on a simply connected open set $\Omega \in \mathbb{C}$, find the imaginary part (unique up to constant additive term).
Example simply connected plane domains

Poorly conditioned domains

Well conditioned domains
In this case, the problem reduces to one in complex analysis: come up with a geometric bound for the harmonic conjugation operator.

In the case that $\Omega$ is convex, harmonic conjugation bounded in terms of aspect ratio of $\Omega$

Conjectures by V. made in 1990s for general case, mostly solved by C. Bishop in the past few years.

Geometric condition number can be bounded in terms of a tree-like decomposition of $\Omega$ into approximately round bodies.

Extension to 3D wide open.
Evaluation of polynomials

- Suppose $p(x) = a_0 + a_1x + \cdots + a_dx^d$ (standard monomial form).
- If $x$ perturbed by an absolute amount $\delta_1$ and $a$ by a relative amount $\delta_2$, then $p(x)$ perturbed by at most $|p'(x)|\delta_1 + \|a\| \cdot \|\text{pwr}(x)\|\delta_2$.
- $\text{pwr}(x) = [1; x; \ldots; x^d]\$
- Example: unstable to use Taylor series to approximate $e^x$ when $x \ll 0$ because $\|a\| \cdot \|\text{pwr}(x)\| \gg |p(x)|$.
- Same polynomial OK when $x \approx 0$.
- Same principles carry over to multivariate polynomials.
A Bernstein-Bézier univariate polynomial is specified by coefficients $a_0, \ldots, a_d$ and given by

$$p(x) = \sum_{i=0}^{d} \binom{d}{i} a_i x^i (1 - x)^{d-i}.$$ 

Argument $x$ usually in $[0, 1]$.

Can show that perturbation to $p(x)$ bounded by $|p'(x)|\delta_1 + \|a\|\delta_2$ for this range of $x$. 

Condition number of a root of a univariate polynomial

- Given $p(x)$ as above, find a root.
- Toh and Trefethen (Gautschi): condition number of root $x^*$ (relative perturbation to $p$, absolute perturbation to $x^*$) is ... Exercise: FIGURE IT OUT!
- Hint: Cauchy-Schwarz inequality needed
Condition number of a root of a univariate polynomial

- Given $p(x)$ as above, find a root.
- Toh and Trefethen (Gautschi): condition number of root $x^*$ (relative perturbation to $p$, absolute perturbation to $x^*$) is $\|a\| \cdot \|\text{pwr}(x^*)\|/|p'(x^*)|$.  
- Blum et al. work with absolute perturbation, so for them the condition number is $\|\text{pwr}(x^*)\|/|p'(x^*)|$.  

For condition of an isolated root, Blum et al. give the following definition of condition number. Suppose $x^*$ is a root of polynomial system $p(x) = 0$ and $\nabla p(x^*)$ (Jacobian matrix) is invertible at $x^*$.

Implicit function theorem means there is a differentiable root function $r(q)$ defined in a neighborhood $O$ of $p$ such that $q(r(q)) = 0$ for all $q \in O$.

Condition number is $\|\nabla r(p)\|$.
Condition number theorem of Blum et al.

- Say that $x^*$ is a *singular* root of $\hat{p}$ if $\nabla p(x^*)$ is singular.

- **Theorem:** The reciprocal condition number of root $x^*$ of polynomial $p$ is within a constant factor (depending on degree) of the minimum distance from $p$ to a polynomial $\hat{p}$ such that $x^*$ is a singular root of $\hat{p}$. 
Roots in $[0, 1]$

- Suppose only roots in $[0, 1]$ are sought. Not sufficient to consider condition numbers of roots in this interval since an ill-conditioned root at $0.5 + \epsilon i$ means slight perturbation of polynomial may move the root to $[0, 1]$.

- E.g., $p(x) = (x - 0.5)^2 + 10^{-9}$.

- Two solutions to this problem: (a) consider all roots, or (b) use a definition of condition number that looks beyond the behavior at the roots.
Finding all roots in $[0, 1]^n$

- Problem arises in geometric computing.
- Suppose $p : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a polynomial, and we wish to compute all roots of $p$ in $[0, 1]^n$.
- Ill-conditioned case is when a small perturbation to $p$ creates a singular root in $[0, 1]^n$.
- The problem considered is not counting the number of roots in $[0, 1]^n$. In that problem, there is an additional ill-conditioned case when one of the roots is very close to $\partial [0, 1]^n$. 

(Srijuntongsiri & V.) Define
\[ \kappa(f) = \|f\| \cdot \sup_{x \in [0,1]^n} \left( \min \left( \frac{1}{\|f(x)\|}, \|\nabla f(x)^{-1}\| \right) \right) \]

In other words, \( f \) is well conditioned if for each point in \([0,1]^n\), either \( \|f(x)\|/\|f\| \) is not too small or \( \nabla f(x) \) is not too ill-conditioned.

Condition number is \( \infty \) if there is an \( x^* \) such that \( f(x^*) = 0 \) and \( \nabla f(x^*) \) is singular.

Any standard norm may be used for \( f \).
Theorem: Suppose \( f, \hat{f} \) are two polynomials such that \( \frac{\|f - \hat{f}\|}{\|f\|} \leq \delta \) where \( \delta \kappa(f) \ll 1 \). Suppose \( x^* \in [0, 1]^n \) and \( f(x^*) = 0 \). Then

\[
\inf_{y \in \hat{f}^{-1}(0)} \|x^* - y\| \leq c_d \delta \kappa(f)
\]
Condition number bounds distance to singularity

- As mentioned earlier, a singular polynomial $\hat{f}$ has a root $\mathbf{x}^* \in [0, 1]^n$ such that $\nabla f(\mathbf{x}^*)$ is singular.

- **Theorem:** For a polynomial $f$, the relative distance to singular polynomials is equal to $1/\kappa(f)$ to within a factor depending on degree.
General theorems about condition numbers

- Suppose the condition number is the relative distance to singularity, where “singularity” means belonging to a semi-algebraic cone of co-dimension at least 1.

- Demmel showed that the mean of the logarithmic condition number is small.

- Bürgisser, Cucker and Lotz showed that the smoothed logarithmic condition number is small.
Smoothed condition number

- Suppose $\psi(f)$ is some kind of complexity measure or condition number. It is a nonnegative real-valued function of an input instance $f \in \mathcal{F}$.
- Worst-case analysis: $\sup_{f \in \mathcal{F}} \psi(f)$.
- Average-case analysis: $E_{f \in \mathcal{F}}[\psi(f)]$
- Smoothed analysis (Spielman & Teng): $\sup_{f \in \mathcal{F}} E_{\hat{f} \in \mathcal{B}(f, \delta)}[\psi(\hat{f})]$
Affine invariance

- Condition number of Srijuntongsiri & V. is not affinely invariant, but many algorithms are.
- Affinely invariant means that $\kappa(f) = \kappa(Af)$ for any nonsingular matrix $A$.
- S. & V. solve this problem in a brute-force manner: define $\hat{\kappa}(f) = \inf \{\kappa(Af) : A \in GL_n(\mathbb{R})\}$; trivially $\hat{\kappa}(f)$ is affinely invariant.
- An affinely invariant algorithm whose complexity is bounded in terms of $\kappa(f)$ is automatically also bounded by $\hat{\kappa}(f)$.
Algorithms to find roots in $[0, 1]^n$

- Problem: Given polynomial $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, find all roots in $[0, 1]^n$.
- Focus is on the case of small degree, small $n$ so that exhaustive search is tractable.
- Important problem in computational geometry.
- Goal: obtain an algorithm whose complexity is bounded by $\kappa(f)$.
An algorithm with no such bound

- Consider classic Toth algorithm. Subdivide $[0, 1]^n$ into $2^n$ subcubes. For each subcube, use a test that either confirms that $f$ has a unique root in the subcube, confirms that there is no root, or else is inconclusive.

- If test is inconclusive, recursively subdivide.

- Test based on interval arithmetic.

- Problem: if a root has a coordinate that is exactly of the form $k/2^l$ (or is very close to such a point) then the Toth algorithm can get stuck (inconclusive at all recursive levels).
KTS Algorithm

- KTS stands for *Kantorovich Test Subdivision* algorithm.
- Also based on recursive subdivision into cubes.
- Uses Kantorovich test and exclusion test.
- Kantorovich test guarantees a unique root that can be found with Newton.
Kantorovich theorem

- Suppose $f : D \to \mathbb{R}^n$ is differentiable, $D$ an open convex subset of $\mathbb{R}^n$.
- Suppose $x_0 \in D$, $\|\nabla f(x_0)^{-1}f(x_0)\| \leq \eta$.
- Suppose $\|\nabla f(x_0)^{-1}(\nabla f(x) - \nabla f(y))\| \leq \omega \|x - y\|$ for all $x, y \in D$.
- If $h = \eta\omega < 1/2$ then $f$ has a root in $B(x_0, (1 - \sqrt{1 - 2h})/\omega)$. This root is unique in $B(x_0, (1 + \sqrt{1 - 2h})/\omega) \cap D$ and is the limit of Newton’s method starting at $x_0$. 
Kantorovich test

- For a polynomial function, it is straightforward to estimate $\eta$ and $\omega$.
- Can determine if the current subcube lies in a disk where there is convergence to a unique root.
- As the subcube gets smaller, the disk stays the same size.
The Bernstein-Bézier form of a polynomial function $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ of individual degree $d$ is
\[
\sum_{i_1=0}^d \cdots \sum_{i_n=0}^d a_{i_1 \cdots i_n} \prod_{j=1}^n x_j^{i_j} (1 - x_j)^{d-i_j} \cdot \binom{d}{i_j}
\]

**Theorem:** $f([0, 1]^n) \subset \text{conv}(a_{0\cdots0}, \ldots, a_{d\cdots d})$.

Follows because monomials are nonnegative and sum to 1.
Exclusion test

- Given a subcube $S = [a_1, b_1] \times \cdots \times [a_n, b_n]$ of $[0, 1]^n$, can define the obvious bijection
  $\pi : [0, 1]^n \to S$ and new polynomial $\tilde{f} : [0, 1]^n \to \mathbb{R}^n$
  by $\tilde{f}(\pi(x)) = f(x)$.

- Rewrite $\tilde{f}$ in B-B form. If hull of control points of $\tilde{f}$
  does not contain $0$ then $f$ has no root in $S$. 
Main theorem

- Suppose $S$ is a subcube of width at most $c_d / \kappa(f)^2$.
- Then either the Kantorovich test is satisfied by $S$, and the region containing the root also contains $S$, or the exclusion test is satisfied by $S$ (or possibly both).
- Means that $S$ does not have to be further subdivided.
- Yields a complexity bound that depends on $\kappa(f)$. 
Extension to the case of one degree of freedom

- Suppose now that \( f : [0, 1]^{n+1} \to \mathbb{R}^n \) is a polynomial system. Roots in this case are, generically, curves in \([0, 1]^{n+1}\).
- This problem is also important in computational geometry: surface-surface intersection problem.
- Define

\[
\kappa(f) = \|f\| \cdot \max_{x \in [0,1]^{n+1}} \min \left( \frac{1}{\|f(x)\|}, \|\nabla f(x)^+\| \right)
\]

- Here, \( B^+ \) denotes \( B^T(BB^T)^{-1} \), the Moore-Penrose inverse of \( B \) in the case that rows of \( B \) are linearly independent.
Srijuntongsiri & V. extend the KTS algorithm to this case with similar complexity bounds.

New problem: tracing curves between subcubes to get connected components.

Kantorovich theorem implies that once subdivision is fine enough, there will be only one possible way to connect curves together.
The KTS algorithm has been implemented in floating point arithmetic.

In floating-point arithmetic, its correctness is not guaranteed.

Can implement KTS in interval arithmetic to guarantee correctness, which raises the cost.

KTS is affinely invariant in exact arithmetic, but not in floating point or interval arithmetic.
Yet another condition number for root-finding

- Possible to reduce the multivariate polynomial rootfinding to eigenvalue/eigenvector computation (see e.g. Jónsson & V.) using resultants.
- Another avenue for defining root-finding condition number: use conditioning of resulting eigenvalue/eigenvector problem
- Jónsson & V.: besides condition number at the root, resulting root-finding condition number deteriorates if the polynomials are close to having a common factor.
Future directions (I)

- Condition-based analysis and floating point arithmetic: Is affine invariance good or bad? For linear system solving, condition number obviously not affinely invariant!
- Multiple ways to define condition number of polynomial rootfinding.
  - Condition number of roots
  - Condition number involving function value and first derivative
  - Condition number involving resultant matrices

Find connections?
Future directions (II)

- Robust optimization and condition numbers
- Lots of work to do on geometric condition numbers.
- Condition number of semidefinite programming. Although Renegar and others have defined this, there are still well-conditioned instances that cause trouble for solvers.
- Preconditioning is not widely used outside of solving $Ax = b$. 