Condition Numbers of Numeric and Algebraic Problems

Stephen Vavasis[1]

[1]Department of Combinatorics & Optimization
University of Waterloo

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

Outline

Outline

Condition number definition

- Given a real-number problem, that is, a function $\Phi : \mathbf{R}^m \longrightarrow \mathbf{R}^n$, the *condition number* of an instance means its sensitivity to small perturbation. In particular: $\text{cond num} = \lim_{\epsilon \to 0} \sup_{\|\mathbf{y}\| \le \epsilon} \frac{\|\Phi(\mathbf{x}+\mathbf{y}) - \Phi(\mathbf{x})\|}{\|\mathbf{y}\|}$ (absolute measurement).
- Or perhaps
$$\text{cond num} = \lim_{\epsilon \to 0} \sup_{\|\mathbf{y}\| \le \epsilon} \frac{\|\Phi(\mathbf{x}+\mathbf{y}) - \Phi(\mathbf{x})\| / \|\Phi(\mathbf{x})\|}{\|\mathbf{y}\| / \|\mathbf{x}\|}$$

Details

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 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

Condition number of linear equations

- Most famous classic example (Von Neumann & Goldstine; Turing) is the condition number of solving linear equations.
- Function Φ is $\Phi(A, \mathbf{b}) = A^{-1}\mathbf{b}$, $A \in \mathbf{R}^{n\times n}$, $\mathbf{b} \in \mathbf{R}^n$.
- Sensitivities understood in the relative sense. All data may be perturbed. Matrix norm is induced by vector norm.

Condition number of linear equations (cont'd)

- **Theorem:** Suppose $\mathbf{x} = A^{-1}\mathbf{b}$; $\mathbf{x} + \Delta\mathbf{x} = (A + \Delta A)^{-1}(\mathbf{b} + \Delta\mathbf{b})$ with

$$\max\left(\frac{\|\Delta A\|}{\|A\|}, \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|}\right) \leq \delta.$$

Then

$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq 2\kappa(A)\delta + O(\delta^2)$$

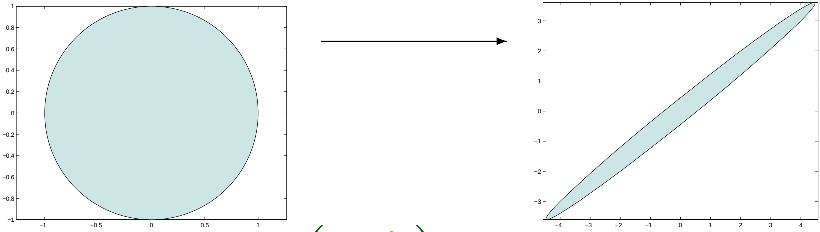
where $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ is the *condition number* of A.

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)

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



$$\kappa \begin{pmatrix} 4 & 2 \\ 3 & 2 \end{pmatrix} = 16.4$$

Condition number of linear equations (cont'd)

- The bound on perturbation to $\mathbf{x}$ can be (mostly) achieved.
- The bound can be achieved if only A or only $\mathbf{b}$ is perturbed.
- Condition number does not depend on $\mathbf{b}$.

Condition number and distance to singularity

- **Theorem:** If $A \in \mathbf{R}^{n \times n}$ 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 ΔA pushes the smallest singular value to 0.

Condition number and iterative methods

- **Theorem:** Suppose $A \in \mathbf{R}^{n\times n}$ is symmetric and positive definite. Then the ith iterate of the conjugate gradient method for solving $A\mathbf{x} = \mathbf{b}$ satisfies
$$\|\mathbf{x}_i - A^{-1}\mathbf{b}\|_A \leq 2\|\mathbf{x}_0 - A^{-1}\mathbf{b}\|_A \cdot \left(\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1}\right)^i.$$
- Note: $\|\mathbf{x}\|_A$ means $(\mathbf{x}^T A\mathbf{x})^{1/2}$.
- For a symmetric positive definite A, $\kappa(A) = \lambda_{\max}(A)/\lambda_{\min}(A)$.
- Steepest descent minimization applied to $\phi(A) = \mathbf{x}^T A\mathbf{x}/2 - \mathbf{b}^T\mathbf{x}$ is also bounded in terms of condition number.

Extension to nonsymmetric systems

- For nonsymmetric A, can apply CG to minimize $\|A\mathbf{x} - \mathbf{b}\|$; equivalent to solving symmetric system $A^T A\mathbf{x} = A^T\mathbf{b}$.
- Note that $\kappa(A^T A) = \kappa(A)^2$.
- Other well-known iterative methods for nonsymmetric $A\mathbf{x} = \mathbf{b}$, 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 $A\mathbf{x} = \mathbf{b}$.

Preconditioning linear equations

- Instead of applying conjugate gradient to $A\mathbf{x} = \mathbf{b}$, apply it to $CAC^T\mathbf{y} = C\mathbf{b}$, where C is a square nonsingular system; C^TC 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 $A^T D^{1/2} H D^{1/2} A$, where A is a node-arc incidence matrix, D is diagonal, positive semidefinite ($\Longrightarrow$ $A^T DA$ 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

Linear least squares

- Problem is: minimize $\|A\mathbf{x} - \mathbf{b}\|_2$ given $A \in \mathbf{R}^{m\times n}$ and $\mathbf{b} \in \mathbf{R}^m$.
- Equivalent to solving linear equations $A^T A\mathbf{x} = A^T\mathbf{b}$ (first order optimality condition).
- Assume that $\mathrm{rank}(A) = n$, so solution is unique.
- Common algorithms: method of normal equations; QR factorization

Condition number of linear least squares

- **Theorem:** If $\|\Delta A\|/\|A\| \leq \delta$ and $\|\Delta \mathbf{b}\|/\|\mathbf{b}\| \leq \delta$ then

$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq 2\kappa(A)\delta + \frac{\kappa(A)^2\delta\|\mathbf{b} - A\mathbf{x}\|}{\|A\| \cdot \|\mathbf{x}\|}$$

  due to Wedin. Achievable.
- Here, $\kappa(A) = \sigma_1(A)/\sigma_n(A)$.
- If linear systems bound applied to $A^T A\mathbf{x} = A^T\mathbf{b}$, obtain a weaker bound.

Solving linear least squares

- 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 $\|D(A\mathbf{x} - \mathbf{b})\|$ where D is a positive definite diagonal weight matrix.
- Reduces to ordinary linear least squares under the obvious substitution $\bar{A} = DA$ and $\bar{\mathbf{b}} = D\mathbf{b}$.
- Perturbation bound for ordinary least squares means solution can be arbitrarily inaccurate as $\kappa(D) \to \infty$.

Weighted least squares (cont'd)

- **Theorem:** There is a bound on $\|\mathbf{x} - \hat{\mathbf{x}}\|$ that depends on χ_A and $\bar{\chi}_A$ 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 χ_A.)

More about χ_A

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

Outline

Condition number of eigenvalues

- If $A\mathbf{x} = \lambda\mathbf{x}$ and λ is not a repeated eigenvalue, then λ 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 $(A + \Delta A)(\mathbf{x} + \Delta\mathbf{x}) = (\lambda + \Delta\lambda)(\mathbf{x} + \Delta\mathbf{x})$ to first order; multiply on the left by $\mathbf{y}^H$ (normalized left eigenvector of λ).

Condition number of eigenvalues

- If $A\mathbf{x} = \lambda\mathbf{x}$ and λ is not a repeated eigenvalue, then λ depends smoothly on A. Differentiation shows that if $\lambda + \Delta\lambda$ is the corresponding eigenvalue of $A + \Delta A$, $\|\Delta A\| \leq \delta$, then $|\Delta\lambda| \leq \delta/|\mathbf{y}^H\mathbf{x}| + O(\delta^2)$.
- Here, $\mathbf{y}$ is the normalized left eigenvector of A and $\mathbf{x}$ 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 λ is perfectly conditioned since $|\mathbf{y}^H\mathbf{x}| = 1$.

Ill conditioned eigenvalues and pseudoeigenvalues

- If an eigenvalue is ill conditioned ($|\mathbf{y}^H\mathbf{x}| \approx 0$), 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 ϵ-*pseudoeigenvalue* of A to be a complex number λ such that $\|(A - \lambda I)^{-1}\| \geq 1/\epsilon$.
- Equivalent to: λ is a ϵ-pseudoeigenvalue if there exists ΔA s.t. $\|\Delta A\| \leq \epsilon$ and λ 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

Linear programming

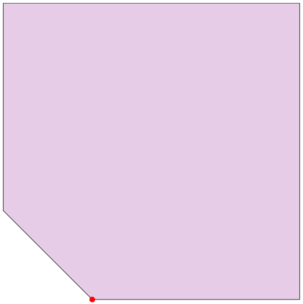
- Linear programming refers to minimizing $\mathbf{c}^T\mathbf{x}$ subject to $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$, where $A \in \mathbf{R}^{m \times n}$ is given, as are $\mathbf{b} \in \mathbf{R}^m$ and $\mathbf{c} \in \mathbf{R}^n$.
- Linear programming admits an elegant duality 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 $A, \mathbf{b}, \mathbf{c}$. Some subtlety here: What if constraints are structured, e.g., $l_i \leq \mathbf{a}_i^T \mathbf{x} \leq u_i$.
- (At least) two possible outputs from LP: optimal value $\mathbf{c}^T \mathbf{x}^*$ or optimizer $\mathbf{x}^*$ itself.
- Optimizer is more sensitive to perturbation than optimal value.

Example

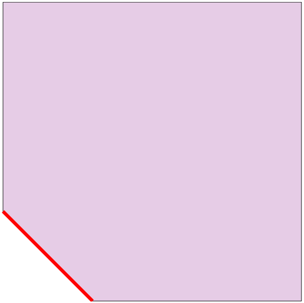
$$\begin{aligned} \min \quad & x + (1+\delta)y \\ \text{s.t.} \quad & 0 \leq x, y \leq 1 \\ & x + y \geq 0.3 \end{aligned}$$



$$\delta = -0.01$$

Example

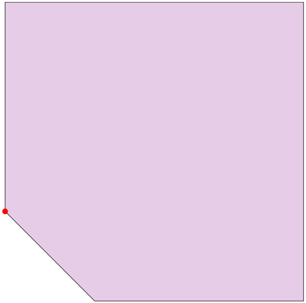
$$\begin{aligned} \min \quad & x + (1+\delta)y \\ \text{s.t.} \quad & 0 \le x, y \le 1 \\ & x + y \ge 0.3 \end{aligned}$$



$\delta = 0$

Example

$$\begin{aligned} \min \quad & x + (1+\delta)y \\ \text{s.t.} \quad & 0 \le x, y \le 1 \\ & x + y \ge 0.3 \end{aligned}$$



$\delta = 0.01$

Renegar's condition number

- Start with the *LP feasibility problem*: given $A \in \mathbf{R}^{m \times n}$ such that $\mathrm{rank}(A) = m$ and $\mathbf{b} \in \mathbf{R}^m$, find an $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$.
- Let the instance be written as $\mathbf{d} = (A, \mathbf{b})$. Assume it is feasible. The *condition number* of the instance according to Renegar is

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

- 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 $\mathbf{b}$ scale differently.
- More general critique: condition number should depend on $S = \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$ but not on particular choice of A and $\mathbf{b}$.

Condition number independent of representation

- Epelman and Freund define a condition number for LP feasibility as $C_2(\mathbf{d}) = 1/\mathrm{sym}(H(\mathbf{d}), \mathbf{0})$ where $\mathrm{sym}(D, \mathbf{x}) = \sup\{t : \mathbf{x} + \mathbf{v} \in D \Longrightarrow \mathbf{x} - t\mathbf{v} \in D\}$ and $H(\mathbf{d}) = \{\mathbf{b}\theta - A\mathbf{x} : \theta \geq 0; \mathbf{x} \geq \mathbf{0}; |\theta| + \|\mathbf{x}\| \leq 1\}$
- Existence of feasible point implies $\mathbf{0} \in \mathrm{int}(H(\mathbf{d}))$.
- $\mathrm{sym}(D, \mathbf{0})$ is invariant under invertible linear mappings of D.
- E & F show that $C_2(\mathbf{d}) \leq C(\mathbf{d})$.

Other forms of linear feasibility

- Cheung, Cucker & Peña unify several LP results.
- Form the following self-dual feasibility problem:

$$\begin{array}{rcrcrcl} & - & \mathbf{c}^T\mathbf{x} & + & \mathbf{b}^T\mathbf{y} & \geq & 0, \\ \mathbf{c}x_0 & & & - & A^T\mathbf{y} & \geq & \mathbf{0}, \\ -\ \mathbf{b}x_0 & + & A\mathbf{x} & & & = & \mathbf{0}, \\ x_0, & & \mathbf{x} & & & \geq & \mathbf{0}. \end{array}$$

- Let ρ 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.

LP complexity bounded by χ_A

- V. and Ye proposed an interior point method whose running time depends only on $\bar{\chi}_A, m, n$ and not on $\mathbf{b}, \mathbf{c}$.
- 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 *layered* least squares, a generalization of weighted least-squares.
- Some weights infinitely larger than others.

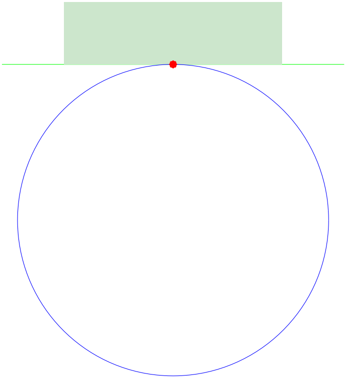
Is $\bar{\chi}_A$ a condition number?

- 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

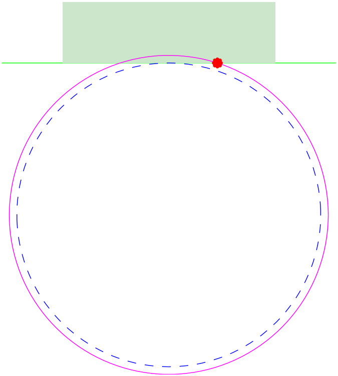
- Semidefinite programming instances can have infinite condition number even when there is a unique optimizer.
- Example: maximize x subject to $y \geq 1$, $x^2 + y^2 \leq 1$.
- Unique feasible point $(0, 1)$ has objective value 0.
- Perturbing constraint to $x^2 + y^2 \leq 1 + \epsilon$ moves optimizer to $(\sqrt{\epsilon}, 1)$.
- Hence ratio of perturbation of root to perturbation of data is unbounded as $\epsilon \to 0$.

Semidefinite programming (cont'd)



In this SDP instance, the *Slater* condition fails.

Semidefinite programming (cont'd)



In this SDP instance, the *Slater* condition fails.

Outline

Boundary value problem

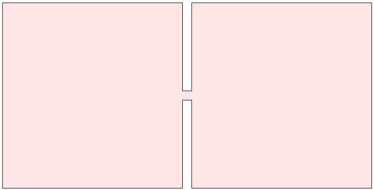
- A *boundary value problem* means solving a differential equation involving one or more spatial independent variables over a domain $\Omega \subset \mathbf{R}^d$ (usually $d = 1, 2, 3$) given boundary conditions and field values.
- Classic example mentioned earlier: $\nabla \cdot (\sigma(x) \nabla u(x)) = -f(x)$ 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

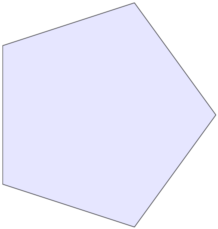
- Difficulty of solving BVP solution depends partly on geometry of Ω. (Many technicalities regarding mesh generation, etc.)
- Clean problem: assume Ω simply connected in $\mathbf{R}^2$, the BVP is Laplace's equation, boundary data is Dirichlet.
- Can reduce solution to inversion of integral operators 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 $\Omega \in \mathbf{C}$, find the imaginary part (unique up to constant additive term).

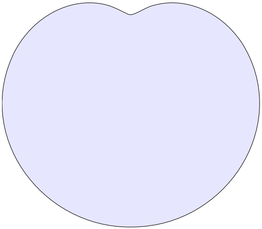
Example simply connected plane domains

Poorly conditioned domains



Well conditioned domains





Complex analysis

- 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 Ω is convex, harmonic conjugation bounded in terms of aspect ratio of Ω
- 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 Ω into approximately round bodies.
- Extension to 3D wide open.

Outline

Evaluation of polynomials

- Suppose $p(x) = a_0 + a_1 x + \cdots + a_d x^d$ (standard monomial form).
- If x perturbed by an absolute amount δ_1 and $\mathbf{a}$ by a relative amount δ_2, then $p(x)$ perturbed by at most $|p'(x)|\delta_1 + \|\mathbf{a}\| \cdot \|\mathrm{pwr}(x)\|\delta_2$.
- $\mathrm{pwr}(x) = [1; x; \ldots; x^d]$
- Example: unstable to use Taylor series to approximate e^x when $x \ll 0$ because $\|\mathbf{a}\| \cdot \|\mathrm{pwr}(x)\| \gg |p(x)|$.
- Same polynomial OK when $x \approx 0$.
- Same principles carry over to multivariate polynomials.

Evaluation of Bernstein-Bézier 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 $|p'(x)|\delta_1 + \|\mathbf{a}\|\delta_2$ 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 $\|\mathbf{a}\| \cdot \|\mathrm{pwr}(x^*)\| / |p'(x^*)|$.
- Blum et al. work with absolute perturbation, so for them the condition number is $\|\mathrm{pwr}(x^*)\| / |p'(x^*)|$.

Multivariate polynomial systems

- For condition of an isolated root, Blum et al. give the following definition of condition number.
- Suppose $\mathbf{x}^*$ is a root of polynomial system $\mathbf{p}(\mathbf{x}) = \mathbf{0}$ and $\nabla\mathbf{p}(\mathbf{x}^*)$ (Jacobian matrix) is invertible at $\mathbf{x}^*$.
- Implicit function theorem means there is a differentiable root function $\mathbf{r}(\mathbf{q})$ defined in a neighborhood O of $\mathbf{p}$ such that $\mathbf{q}(\mathbf{r}(\mathbf{q})) = \mathbf{0}$ for all $\mathbf{q} \in O$.
- Condition number is $\|\nabla\mathbf{r}(\mathbf{p})\|$.

Condition number theorem of Blum et al.

- Say that $\mathbf{x}^*$ is a *singular* root of $\hat{\mathbf{p}}$ if $\nabla \mathbf{p}(\mathbf{x}^*)$ is singular.
- **Theorem:** The reciprocal condition number of root $\mathbf{x}^*$ of polynomial $\mathbf{p}$ is within a constant factor (depending on degree) of the minimum distance from $\mathbf{p}$ to a polynomial $\hat{\mathbf{p}}$ such that $\mathbf{x}^*$ is a singular root of $\hat{\mathbf{p}}$.

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 $\mathbf{p} : \mathbf{R}^n \to \mathbf{R}^n$ is a polynomial, and we wish to compute all roots of $\mathbf{p}$ in $[0, 1]^n$.
- Ill-conditioned case is when an small perturbation to $\mathbf{p}$ 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 $\partial[0, 1]^n$.

Condition number definition

- (Srijuntongsiri & V.) Define $\kappa(\mathbf{f}) = \|\mathbf{f}\| \cdot \sup_{\mathbf{x}\in[0,1]^n} \left(\min\left(\frac{1}{\|\mathbf{f}(\mathbf{x})\|}, \|\nabla\mathbf{f}(\mathbf{x})^{-1}\|\right)\right)$
- In other words, $\mathbf{f}$ is well conditioned if for each point in $[0,1]^n$, either $\|\mathbf{f}(\mathbf{x})\|/\|\mathbf{f}\|$ is not too small or $\nabla\mathbf{f}(\mathbf{x})$ is not too ill-conditioned.
- Condition number is ∞ if there is an $\mathbf{x}^*$ such that $\mathbf{f}(\mathbf{x}^*) = \mathbf{0}$ and $\nabla\mathbf{f}(\mathbf{x}^*)$ is singular.
- Any standard norm may be used for $\mathbf{f}$.

Condition number bounds perturbations to roots

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

$$\inf_{\mathbf{y}\in\hat{\mathbf{f}}^{-1}(\mathbf{0})} \|\mathbf{x}^* - \mathbf{y}\| \leq c_d \delta\kappa(\mathbf{f})$$

Condition number bounds distance to singularity

- As mentioned earlier, a singular polynomial $\hat{\mathbf{f}}$ has a root $\mathbf{x}^* \in [0,1]^n$ such that $\nabla\mathbf{f}(\mathbf{x}^*)$ is singular.
- **Theorem:** For a polynomial $\mathbf{f}$, the relative distance to singular polynomials is equal to $1/\kappa(\mathbf{f})$ 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): $\sup_{f \in \mathcal{F}} E_{\hat{f} \in B(f,\delta)}[\Psi(\hat{f})]$

Affine invariance

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

Algorithms to find roots in $[0, 1]^n$

- Problem: Given polynomial $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}^n$, 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 $\kappa(\mathbf{f})$.

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 $\mathbf{f}$ 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 $\mathbf{f} : D \to \mathbf{R}^n$ is diffble, D an open convex subset of $\mathbf{R}^n$.
- Suppose $\mathbf{x}_0 \in D$, $\|\nabla \mathbf{f}(\mathbf{x}_0)^{-1} \mathbf{f}(\mathbf{x}_0)\| \leq \eta$.
- Suppose $\|\nabla \mathbf{f}(\mathbf{x}_0)^{-1}(\nabla \mathbf{f}(\mathbf{x}) - \nabla \mathbf{f}(\mathbf{y}))\| \leq \omega \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in D$.
- If $h = \eta\omega < 1/2$ then $\mathbf{f}$ has a root in $B(\mathbf{x}_0, (1 - \sqrt{1 - 2h})/\omega)$. This root is unique in $B(\mathbf{x}_0, (1 + \sqrt{1 - 2h})/\omega) \cap D$ and is the limit of Newton's method starting at $\mathbf{x}_0$.

Kantorovich test

- For a polynomial function, it is straightforward to estimate η and ω.
- 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.

Bernstein-Bézier polynomial

- The *Bernstein-Bézier* form of a polynomial function $\mathbf{f} : \mathbf{R}^n \rightarrow \mathbf{R}^n$ of individual degree d is $\sum_{i_1=0}^{d} \cdots \sum_{i_n=0}^{d} \mathbf{a}_{i_1 \cdots i_n} \prod_{j=1}^{n} x_j^{i_j} (1 - x_j)^{d-i_j} \cdot \binom{d}{i_j}$
- **Theorem:** $f([0,1]^n) \subset \mathrm{conv}(\mathbf{a}_{0 \dots 0}, \dots, \mathbf{a}_{d \dots d})$.
- 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 $\tilde{\mathbf{f}} : [0,1]^n \to \mathbf{R}^n$ by $\tilde{\mathbf{f}}(\pi(\mathbf{x})) = \mathbf{f}(\mathbf{x})$.
- Rewrite $\tilde{\mathbf{f}}$ in B-B form. If hull of control points of $\tilde{\mathbf{f}}$ does not contain $\mathbf{0}$ then $\mathbf{f}$ has no root in S.

Main theorem

- Suppose S is a subcube of width at most $c_d/\kappa(\mathbf{f})^2$.
- 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 $\kappa(\mathbf{f})$.

Extension to the case of one degree of freedom

- Suppose now that $\mathbf{f} : [0,1]^{n+1} \to \mathbf{R}^n$ 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
$\kappa(\mathbf{f}) = \|\mathbf{f}\| \cdot \max_{\mathbf{x}\in[0,1]^{n+1}} \min\left(\frac{1}{\|\mathbf{f}(\mathbf{x})\|}, \|\nabla f(\mathbf{x})^{+}\|\right)$
- Here, B^+ denotes $B^T(BB^T)^{-1}$, the Moore-Penrose inverse of B in the case that rows of B are linearly independent.

KTS algorithm for one degree of freedom

- 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.

Interval versus floating point arithmetic

- 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 $A\mathbf{x} = \mathbf{b}$.