The Smoothed Analysis of Algorithms

Daniel A. Spielman
Dept. of Computer Science
Program in Applied Mathematics
Yale University
Outline

Why?

Definitions (by formula, by picture, by example)

Examples:
- Perceptron
- Condition numbers (Gaussian Elimination)
- Simplex Method
- K-means
- Decision trees

What you can do, and how.
Why

Want theorems that explain why algorithms and heuristics work in practice

May be contrived examples on which they fail.

*fail to converge*
*take too long*
*return wrong answer*

Worst-case analysis is not satisfactory.
Attempted fix: Average-case analysis

Measure expected performance on random inputs.

*random graphs*
*random point sets*
*random signals*
*random matrices*
Random is not typical
Critique of Average-Case analysis

Actual inputs might not look random.

Random inputs have very special properties with very high probability.
Smoothed Analysis

Assume randomness/noise in low-order bits

Randomly perturb problem

Problem comes through noisy channel

measurement error
random sampling
arbitrary circumstances
(managers)
Hybrid of worst and average case

Complexity of algorithm: inputs -> time

Worst case: \( C(n) = \max_{x \in \mathbb{R}^n} T(x) \)

Ave case: \( C(n) = \mathbb{E}_{r \in \mathbb{R}^n} [ T(r) ] \)

Smoothed: \( C(n, \sigma) = \max_x \left[ \mathbb{E}_{r \in \mathbb{R}^n} [ T(x + r\sigma \|x\|) ] \right] \)

Gaussian perturbation of std dev \( \sigma \|x\| \)
Smoothed Complexity

\[ C(n, \sigma) = \max_x \left[ \mathbb{E}_{r \in \mathbb{R}^n} \left[ T(x + r\sigma \|x\|) \right] \right] \]

Interpolates between worst and average case

Considers neighborhood of every input

If low, all high complexity is unstable
Complexity Landscape

worst case

average case
Smoothed Complexity Landscape

run time

smoothed complexity

input space
Perceptron Algorithm

Given points $x_i \in \mathbb{R}^d$ and labels $b_i \in \{1, -1\}$

Find $w$ so that $\text{sign}(w^T x_i) = b_i$ for all $i$

Algorithm: iteratively find violated condition, and use it to update $w$

Eventually finds a solution, if one exists
Perceptron Algorithm
(smoothed analysis by Blum-Dunagan ‘02)

Given points $x_i \in \mathbb{R}^d$ and labels $b_i \in \{1, -1\}$
Find $w$ so that $\text{sign}(w^Tx_i) = b_i$ for all $i$

Perturbation:

$$\tilde{x}_i = x_i + r_i \sigma \|x_i\|$$

$r_i$ a Gaussian random vector

Theorem: if solution exists

$$\text{Prob}[\# \text{ steps } > O\left(d^3 m^2 / \sigma^2 \delta^2\right)] < \delta$$
Smoothed Margin

Margin = angle separating pos from neg examples

Block-Novikoff:
Perceptron converges in $O(1/\theta^2)$ iterations.

Blum-Dunagan:

$\text{Prob}[\text{ margin } < \epsilon ] < O(md^{1.5} \epsilon / \sigma)$
Smoothed Margin (simplified)

Assume data is separable by $w_*$ and re-label when perturb

Perturbation:

$$
\tilde{x}_i = x_i + r_i \sigma \|x_i\|
$$

$$
\tilde{b}_i = \text{sign}(w_*^T y_i)
$$

Analysis: is unlikely that any $\tilde{x}_i$ gets close to separating plane
Smoothed Margin (simplified)

Assume data is separable by $w_*$ and re-label when perturb

Perturbation: $\tilde{x}_i = x_i + r_i\sigma \|x_i\|
\tilde{b}_i = \text{sign}(w^T_* y_i)$

Analysis: is unlikely that any $\tilde{x}_i$ gets close to separating plane

$\text{Prob}[\text{dist} < \epsilon] < \epsilon/\sigma$
Smoothed Margin (simplified)

Assume data is separable by $w_*$ and re-label when perturb

Perturbation: $y_i = x_i + r_i \sigma \|x_i\|$
$c_i = \text{sign}(w_*^T y_i)$

Analysis: is unlikely that any $\tilde{x}_i$ gets close to separating plane

$\text{Prob}[\text{dist} < \epsilon] < \epsilon / \sigma$
Explain where going from here

Body text
Condition Numbers

Measure maximum of

$$\frac{\text{norm}(\text{change in output})}{\text{norm}(\text{change in input})}$$

Or, $1/(\text{distance to ill-posed problem})$

$1/\text{margin}$ is a condition number

Perturbed problems usually have small condition number, and are not too close to ill-posed problems
The Matrix Condition Number

\[ \kappa(A) = \| A \| \| A^{-1} \| \]

The ratio of largest to smallest singular values.

Condition number for problem \( Ax = b \)

Focus on \( \| \tilde{A}^{-1} \| = 1 / \sigma_{\text{min}}(\tilde{A}) \)

As \( \| \tilde{A} \| \approx \| A \| \)
Estimating Smoothed Condition Number

Approximately aspect ratio of simplex formed by vectors in columns, and origin.

\[
\frac{\max ||a_i||}{\min \delta_i}
\]
Probability of Large Condition Number

Unlikely, as large $\|a_i\|$ very unlikely, and small $\delta_i$ not too likely, because Gaussian point unlikely near plane.
Smoothed Analysis of $||\tilde{A}^{-1}||$

Edelman:
for standard Gaussian random matrix $G$

$$\Pr [||G^{-1}|| > t] \leq \frac{\sqrt{d}}{t}$$

Sankar-S-Teng:
for $\tilde{A} = A + \sigma G$

$$\Pr [||\tilde{A}^{-1}|| > t] < \frac{??\sqrt{d}}{\sigma t}$$
Geometry of $\|A^{-1}\|$ 

$$A = \begin{pmatrix}
    a_1 \\
    a_2 \\
    \vdots \\
    a_n
\end{pmatrix} \quad A^{-1} = X = \begin{pmatrix}
    x_1 & x_2 & \cdots & x_n
\end{pmatrix}$$

$$Ax_1 = \begin{pmatrix}
    1 \\
    0 \\
    \vdots \\
    0
\end{pmatrix} \quad \Rightarrow \quad \|x_1\| = \frac{1}{\text{dist}(a_1, \text{Span}(a_2, \ldots, a_n))}$$
Geometry of $\|A^{-1}\|$  \hspace{1cm} (A^{-1} = X)

$\|x_1\| = \frac{1}{\text{dist} (a_1, \text{Span} (a_2, \ldots, a_n))}$

$\Pr_{a_1} [\text{dist} (a_1, \text{Span} (a_2, \ldots, a_n)) < \epsilon] \leq \sqrt{\frac{2}{\pi} \frac{\epsilon}{\sigma}}$

$\Pr \left[ \max_i \|x_i\| > t \right] \leq \sqrt{\frac{2}{\pi} \frac{d}{t \sigma}} \hspace{1cm} \text{(union bound)}$

$\Pr [\|A^{-1}\| > t] \leq \sqrt{\frac{2}{\pi} \frac{d^{3/2}}{t \sigma}}$  \hspace{1cm} \text{should be } d^{1/2}
Improving bound on $\|A^{-1}\|$ 

Lemma:

For $\|b\| = 1$ \[ \Pr \left[ \|A^{-1}b\| \geq t \right] \leq \sqrt{\frac{2}{\pi \cdot t \sigma}} \]

Apply to random $b$

\[ \Pr \left[ \|A^{-1}b\| \geq \frac{1}{\sqrt{d}} \|A^{-1}\| \right] \geq \text{const} \]

So

\[ \Pr \left[ \|A^{-1}\| \geq t \right] \leq \frac{\Pr \left[ \|A^{-1}b\| \geq \frac{t}{\sqrt{d}} \right]}{\text{const}} \leq \frac{1.323 \sqrt{d}}{t} \]

correction
Gaussian Elimination w/ Partial Pivoting

\[
\begin{align*}
\text{>> } A &= \text{ randn}(2) \\
A &= \begin{bmatrix}
-0.4326 & 0.1253 \\
-1.6656 & 0.2877
\end{bmatrix} \\
\text{>> } x &= A \backslash b \\
x &= \begin{bmatrix}
-5.6821 \\
-28.7583
\end{bmatrix} \\
\text{>> } b &= \text{ randn}(2,1) \\
b &= \begin{bmatrix}
-1.1465 \\
1.1909
\end{bmatrix} \\
\text{>> } \text{norm}(A*x - b) \\
\text{ans } &= 8.0059e-016
\end{align*}
\]
Gaussian Elimination w/ Partial Pivoting

$$\begin{align*}
&\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
-1 & 1 & 0 & 0 & 0 & 0 & 1 \\
-1 & -1 & 1 & 0 & 0 & 0 & 1 \\
-1 & -1 & -1 & 1 & 0 & 0 & 1 \\
-1 & -1 & -1 & -1 & 1 & 1 & 1 \\
-1 & -1 & -1 & -1 & -1 & 1 & 1 \\
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 1 & 0 & 0 & 0 & 4 \\
0 & 0 & 0 & 1 & 0 & 0 & 8 \\
0 & 0 & 0 & 0 & 1 & 1 & 16 \\
0 & 0 & 0 & 0 & 0 & 0 & 32 \\
\end{bmatrix}
\end{align*}$$

Failed!
Gaussian Elimination w/ Partial Pivoting

```
>> A = 2*eye(70) - tril(ones(70));
>> A(:,70) = 1;
>> b = randn(70,1);
>> x = A \ b;
>> norm(A*x - b)
ans =
 3.5340e+004 Failed!

>> Ap = A + randn(70) / 10^9; Perturb A
>> x = Ap \ b;
>> norm(Ap*x - b)
ans =
 5.8950e-015
```
Gaussian Elimination w/ Partial Pivoting

```matlab
>> b = randn(70,1);
>> x = A \ b;
>> norm(A*x - b)
ans =
    3.5340e+004  Failed!
```

```matlab
>> Ap = A + randn(70) / 10^9;  Perturb A
>> x = Ap \ b;
>> norm(Ap*x - b)
ans =
    5.8950e-015

>> norm(A*x - b)
ans =
    3.6802e-008  Solved original too!
```
Gaussian Elimination with Partial Pivoting

Fast heuristic for maintaining precision, by trying to keep entries small

Pivot not just on zeros, but to move up entry of largest magnitude

\[
\begin{pmatrix}
-1 & 6 & 3 \\
1 & -4 & -1 \\
2 & -3 & -4
\end{pmatrix}
\rightarrow
\begin{pmatrix}
2 & -3 & -4 \\
1 & -4 & -1 \\
-1 & 6 & 3
\end{pmatrix}
\]

\[
\begin{pmatrix}
2 & -3 & -4 \\
0 & -5/2 & 1 \\
0 & 9/2 & 1
\end{pmatrix}
\rightarrow
\begin{pmatrix}
2 & -3 & -4 \\
0 & 9/2 & 1 \\
0 & -5/2 & 1
\end{pmatrix}
\]
“Gaussian elimination with partial pivoting is utterly stable in practice. In fifty years of computing, no matrix problems that excite an explosive instability are know to have arisen under natural circumstances …

Matrices with large growth factors are vanishingly rare in applications.”

Nick Trefethen
Gaussian Elimination with Partial Pivoting

“Gaussian elimination with partial pivoting is utterly stable in practice. In fifty years of computing, no matrix problems that excite an explosive instability are known to have arisen under natural circumstances …

Matrices with large growth factors are vanishingly rare in applications.”

Nick Trefethen

Theorem:

$$\Pr [\text{Growth} > x(n/\sigma)^c] < x^{-\log x}$$

[Sankar-S-Teng]
Simplex Method for Linear Programming

\[
\begin{align*}
\text{max} & \quad c^T x \\
\text{s.t.} & \quad Ax \leq b
\end{align*}
\]

Worst-Case: exponential
Average-Case: polynomial

Widely used in practice
**Smoothed Analysis of Simplex Method**

\[
\begin{align*}
\max & \quad c^T x \\
\text{s.t.} & \quad Ax \leq b
\end{align*}
\]

\[
\begin{align*}
\max & \quad c^T x \\
\text{s.t.} & \quad (A + \sigma \|A\|G)x \leq b
\end{align*}
\]

*G is Gaussian*

**Theorem:** For all \( A, b, c \), simplex method takes expected time polynomial in \( m, n, 1/\sigma \)
Shadow Vertices
Another shadow
Shadow vertex pivot rule

start

z

objective
Theorem: For every plane, the expected size of the shadow of the perturbed tope is $\text{poly}(m,d,1/\sigma)$
Theorem: For every $z$, two-Phase Algorithm runs in expected time $\text{poly}(m, d, 1/\sigma)$
A Local condition for optimality

Vertex on $a_1, \ldots, a_d$ maximizes $z$ iff $z \in \text{cone}(a_1, \ldots, a_d)$
Primal

\[ a_1^T x \leq 1 \]
\[ a_2^T x \leq 1 \]
\[ \vdots \]
\[ a_m^T x \leq 1 \]

Polar

\[ \text{ConvexHull}(a_1, a_2, \ldots, a_m) \]
Polar
Polar Linear Program

\[ \max \alpha \]
\[ \alpha z \in \text{ConvexHull}(a_1, a_2, \ldots, a_m) \]
Shadow vertex pivot rule
Count facets by discretizing to \( N \) directions, \( N \rightarrow \uparrow \)
Count pairs in different facets

\[
\Pr \left[ \text{Different Facets} \right] < \frac{c}{N}
\]

So, expect \( c \) Facets
Expect cone of large angle
Isolate on one Simplex
Integral Formulation

\[ p_{\pi}^{(1)}(\epsilon) = \sum_{\pi \in \Pi(d,m)} \int_{\omega, r} \prod_{i=d+1}^{m} \left( \int_{b_{\pi(i)}} \left[ \langle \omega | b_{\pi(i)} \rangle \leq r \right] \mu_{\pi(i)}(b_{\pi(i)}) \, db_{\pi(i)} \right) \cdot \int_{c_1, \ldots, c_d} \left[ z \in \text{Cone}(\omega, r, c_1, \ldots, c_d) \right] \left[ \text{ang}(z, \partial \text{Simplex}(c_1, \ldots, c_d)) < \epsilon \right] \int_{c_1, \ldots, c_d} \right] \]

\[ \prod_{i=1}^{d} \nu_{\omega, r}^{(c_i)} \text{Vol}(\text{Simplex}(c_1, \ldots, c_d)) \, d\omega \, dr \, dc_1 \cdots dc_d \]
Example:

For \(a\) and \(b\) Gaussian distributed points, given that \(\overline{ab}\) intersects x-axis

\[
\text{Prob}[\theta < \varepsilon] = O(\varepsilon^2)
\]
\[ P_\varepsilon = \Pr[\theta < \varepsilon \mid \overline{ab} \cap \text{axis} \neq \emptyset] \]

\[ = c \int_{\overline{ab} \cap \text{axis} \neq \emptyset}^{\theta < \varepsilon} \mu_0(a)\mu_1(b)dadb \]

**Claim:** For \( \varepsilon < \varepsilon_0 \), \( P_\varepsilon \preceq \varepsilon^2 \)
Change of variables

\[ da \; db = |(u+v)\sin(\theta)| \; du \; dv \; dz \; d\theta \]

\[ \mu_0(a) \rightarrow v_0(u, z, \theta) \]

\[ \mu_1(b) \rightarrow v_1(v, z, \theta) \]
Analysis: For $\varepsilon < \varepsilon_0$, $P_\varepsilon \lesssim \varepsilon^2$

$$P_\varepsilon = c \int_{u,v,z} (u + v) \int_{\theta < \varepsilon} |\sin(\theta)| \nu_0(u, z, \theta) \nu_1(v, z, \theta) du dv dz d\theta$$

Slight change in $\theta$ has little effect on $\nu_i$ for all but very rare $u, v, z$
Distance:

Gaussian distributed corners
Idea: fix by perturbation
Title

Body text

Body text

Body text
Title

Body text

Body text

Body text
Title

Body text

Body text

Body text

Body text
Title

Body text

Body text

Body text

Body text
Title

Body text
Title

Body text
Title

Body text
Title

Body text
Title

Body text
Title

Body text
Title

Body text