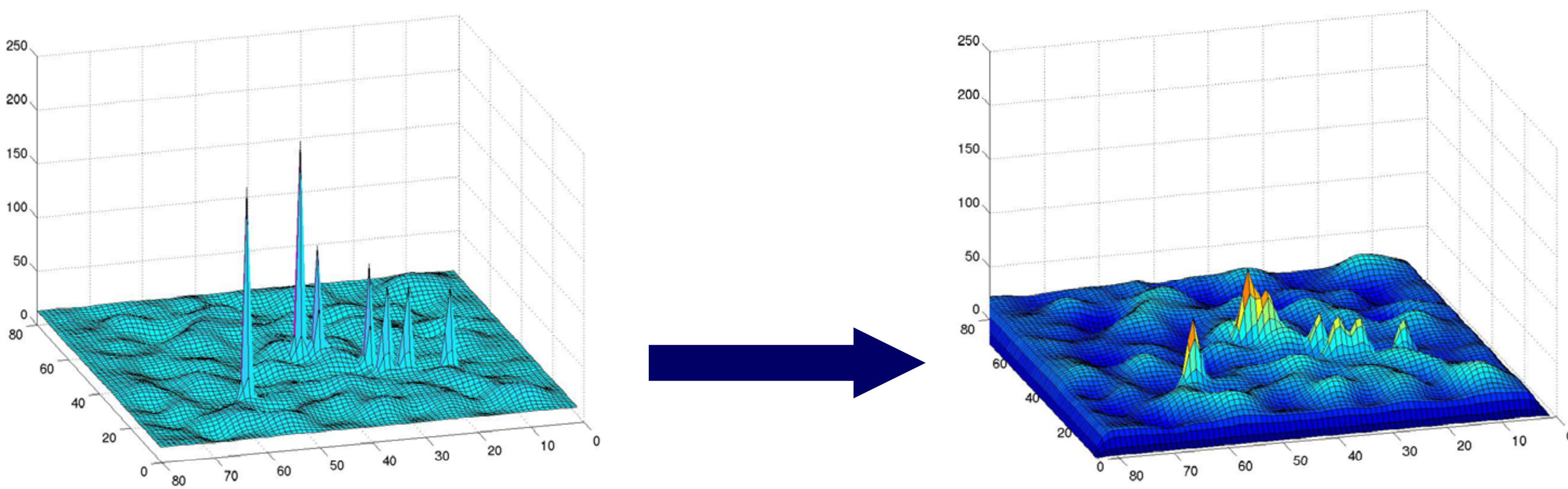
The Smoothed Analysis of Algorithms

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.

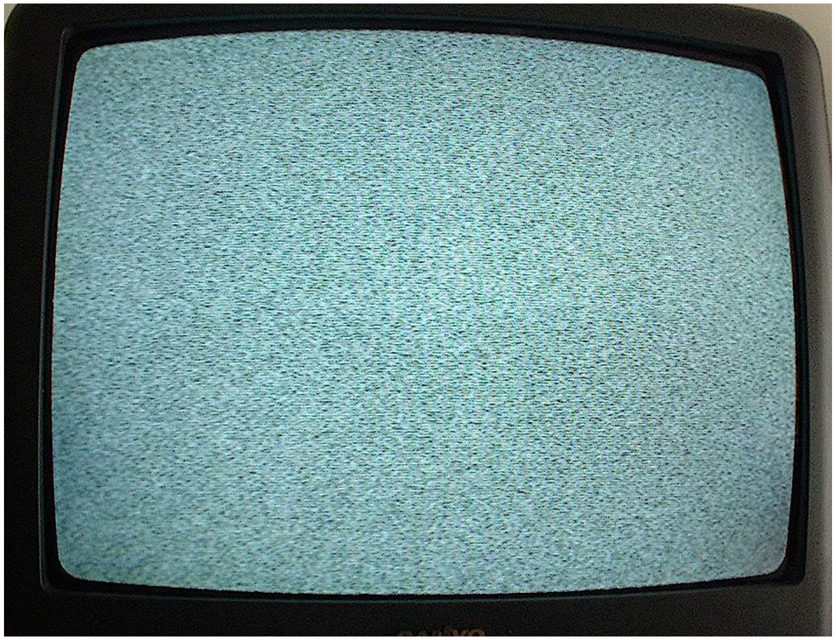
random graphs

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: $C(n) = \mathop{\mathbf{E}}_{r \in \mathbb{R}^n} [\ T(r)\ ]$

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

Gaussian perturbation of std dev $\sigma \|x\|$

Smoothed Complexity

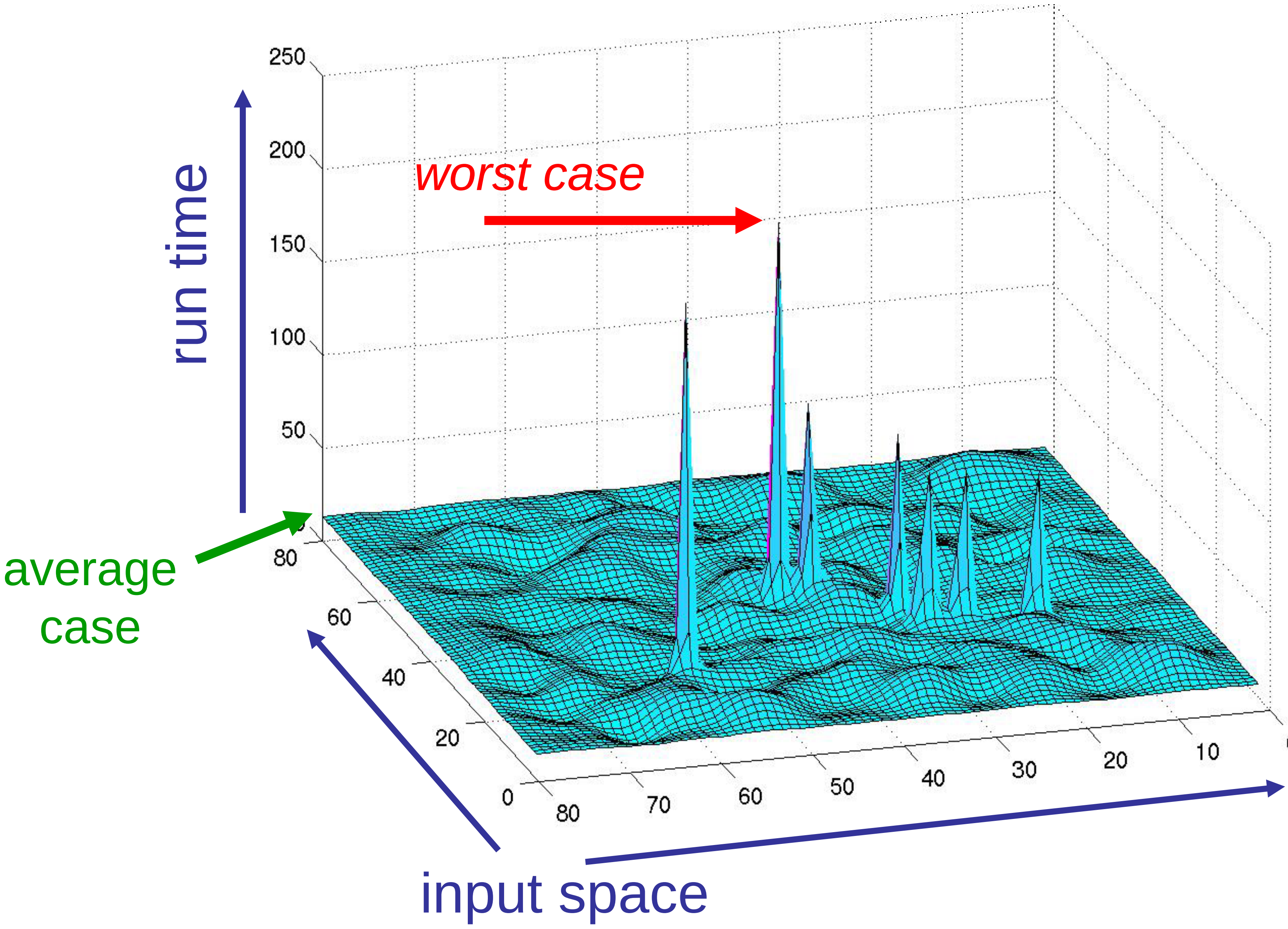
$$C(n,\sigma) = \max_{x} \left[\mathop{\mathbf{E}}_{r \in \mathbb{R}^n} \left[\ T(x + r\sigma \left\| x \right\|)\ \right] \right]$$

Interpolates between worst and average case

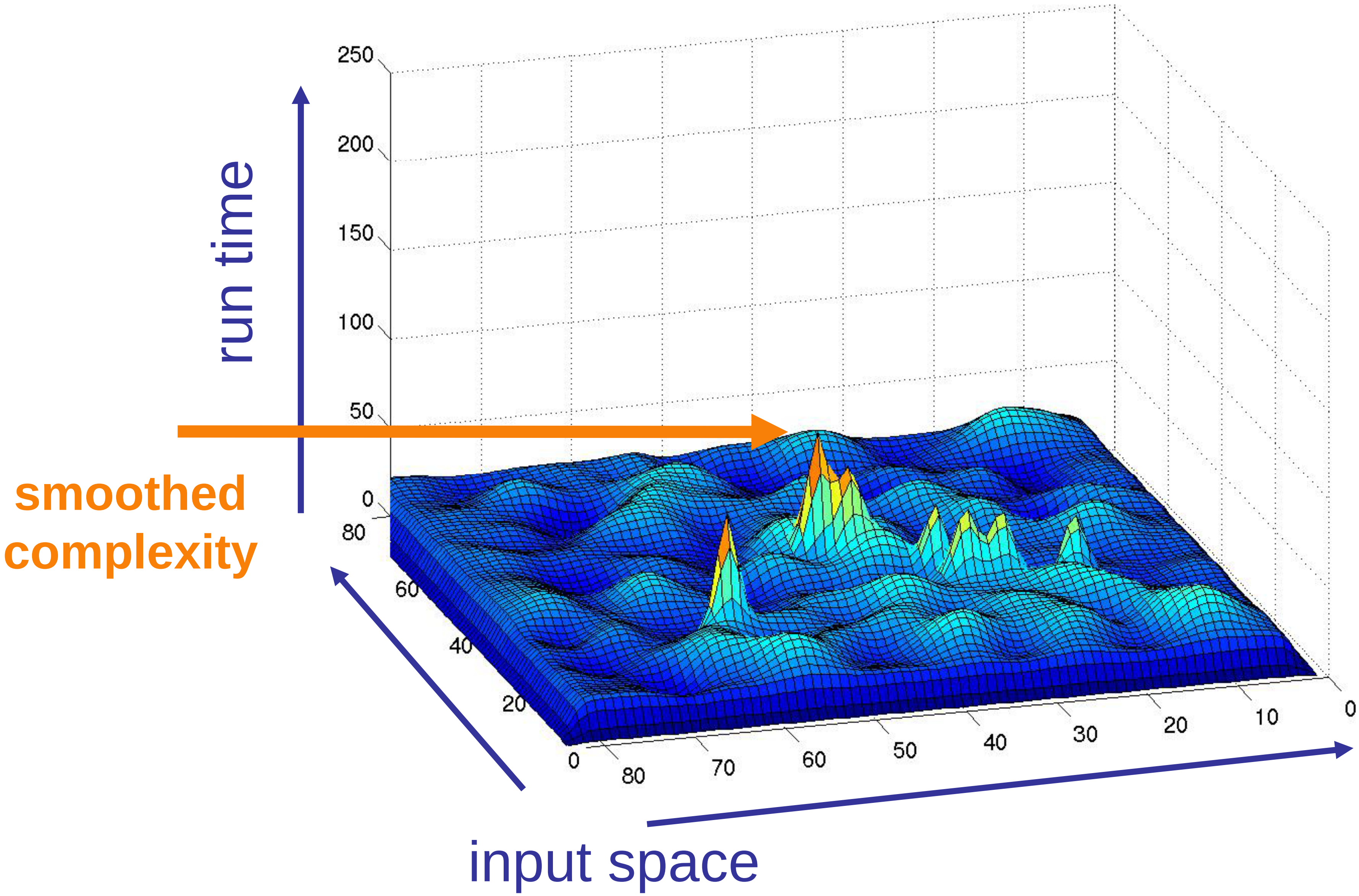
Considers neighborhood of every input

If low, all high complexity is unstable

Complexity Landscape



Smoothed Complexity Landscape



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 $\mathrm{sign}(w^T x_i) = b_i$ for all i

Perturbation:

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

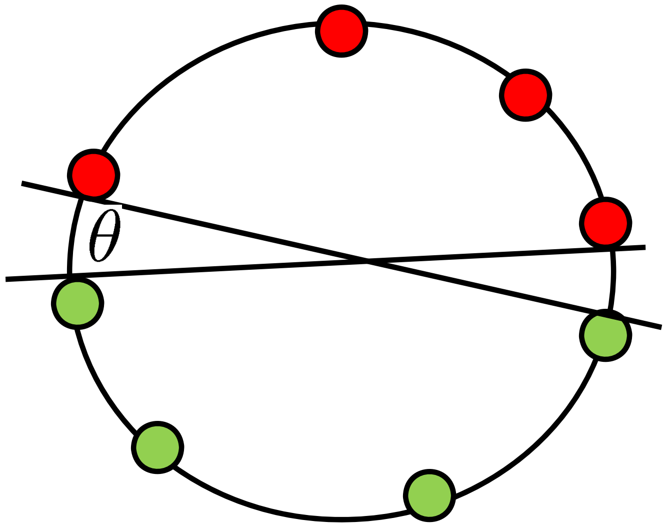
r_i a Gaussian random vector

Theorem: if solution exists

Prob[# steps > $O(d^3 m^2 / \sigma^2 \delta^2)$] < δ

Smoothed Margin

Margin = angle separating pos from neg examples



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

Blum-Dunagan:

Prob[margin < ϵ] < $O(md^{1.5}\epsilon/\sigma)$

Smoothed Margin (simplified)

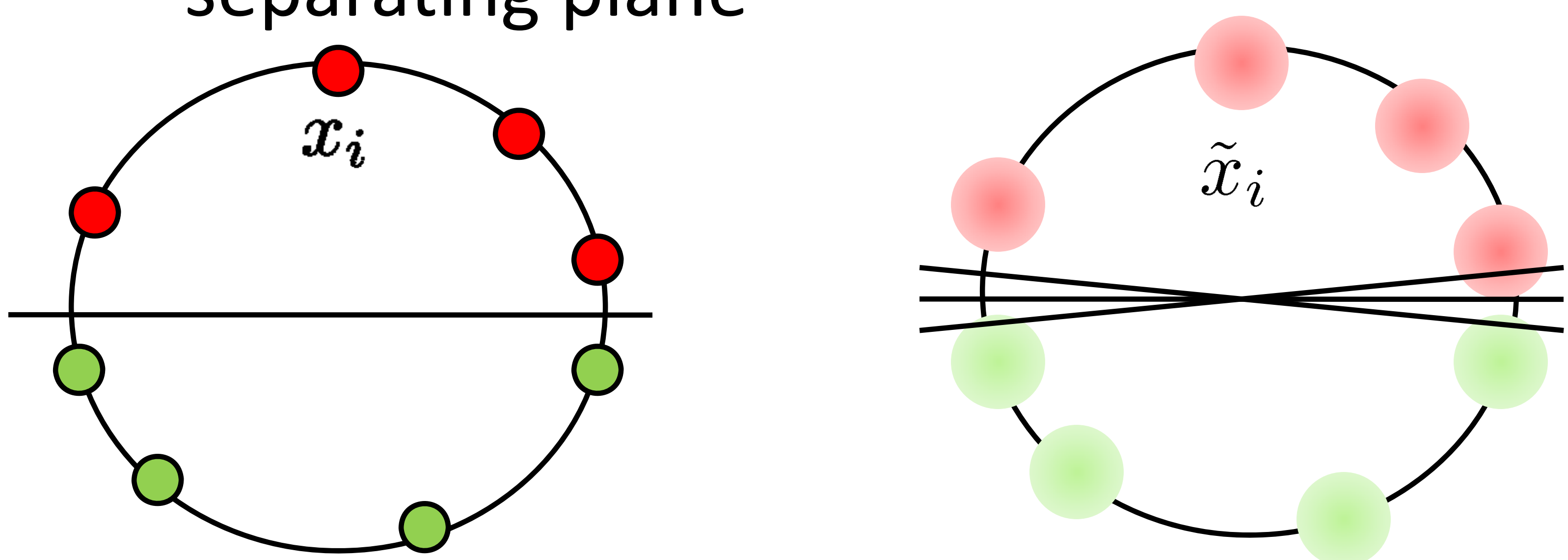
Assume data is separable by w_*
and re-label when perturb

Perturbation:

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

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

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



Smoothed Margin (simplified)

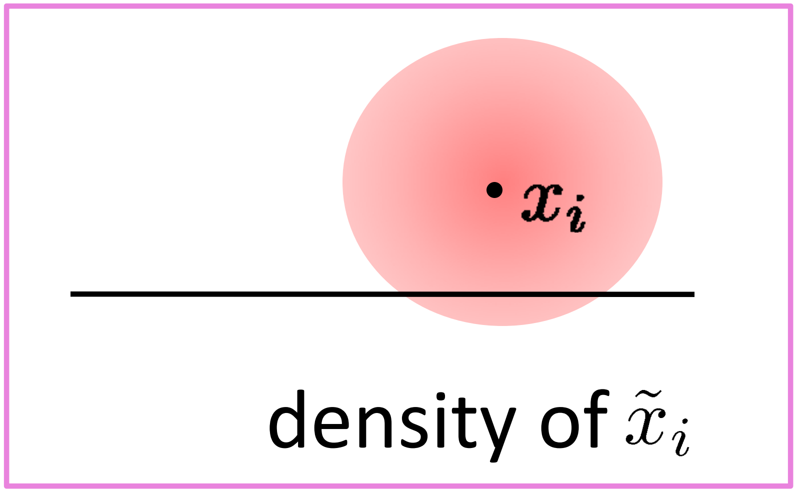
Assume data is separable by w_*
and re-label when perturb

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

Prob[dist < ϵ] < ϵ/σ



Smoothed Margin (simplified)

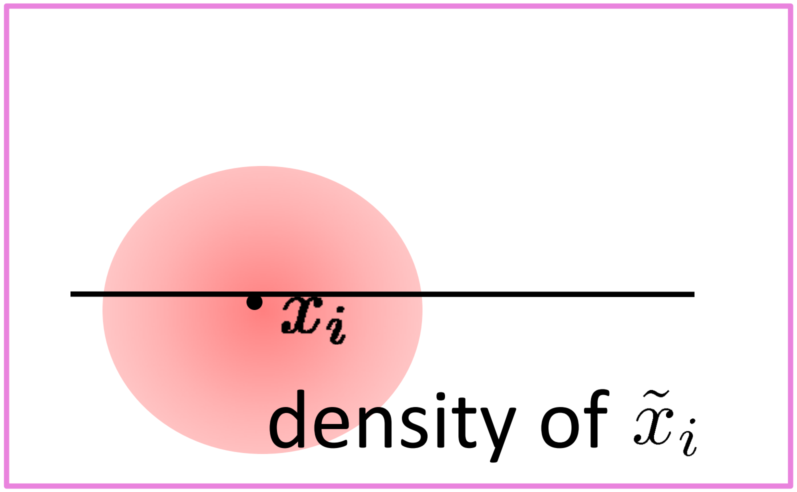
Assume data is separable by w_*
and re-label when perturb

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

Prob[dist < ϵ] < ϵ / σ



Explain where going from here

Body text

Condition Numbers

Measure maximum of

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

Or, 1/(distance to ill-posed problem)

1/margin is a condition number

Perturbed problems usually
have small condition number, and
are not to 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 $Ax = b$

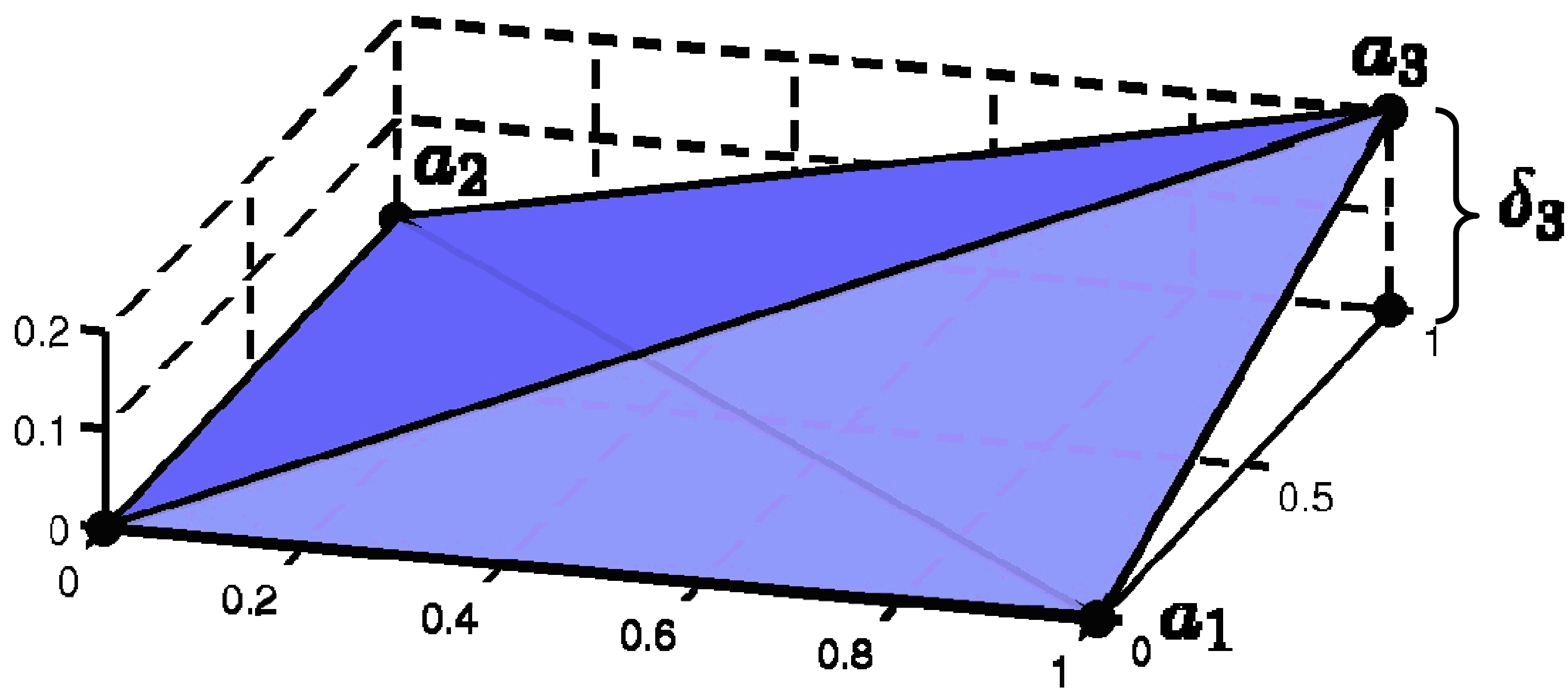
Focus on $||\tilde{A}^{-1}|| = 1/\sigma_{min}(\tilde{A})$

As $||\tilde{A}|| \approx \|A\|$

Estimating Smoothed Condition Number

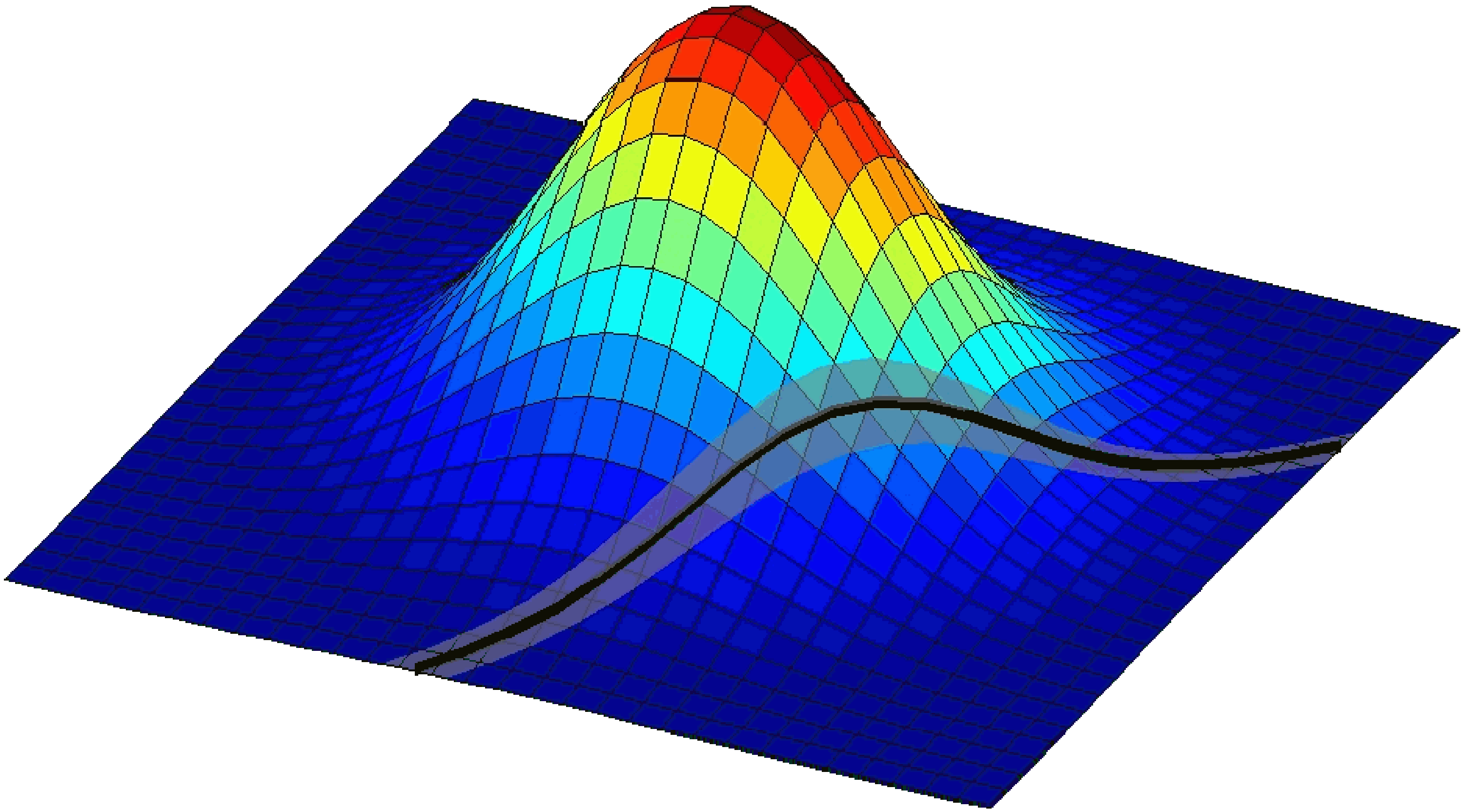
Approximately aspect ratio of simplex formed by vectors in columns, and origin.

$$\frac{\max \|a_i\|}{\min \delta_i}$$



Probability of Large Condition Number

Unlikely, as large $\|a_i\|$ very unlikely,
and small δ_i not too likely,
because Gaussian point unlikely near plane



Smoothed Analysis of $||\tilde{A}^{-1}||$

Edelman:

for standard Gaussian random matrix G

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

Sankar-S-Teng:

for $\tilde{A} = A + \sigma G$

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

Geometry of $\|A^{-1}\|$

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

$$Ax_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow \|x_1\| = \frac{1}{\mathbf{dist}\left(a_1, \mathbf{Span}\left(a_2, \ldots, a_n\right)\right)}$$

Geometry of $\|A^{-1}\|$

$(A^{-1} = X)$

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

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

$$\Pr\left[\max_i \|x_i\| > t\right] \leq \sqrt{\frac{2}{\pi}}\frac{d}{t\sigma}$$

(union bound)

$$\Pr\left[\|A^{-1}\| > t\right] \leq \sqrt{\frac{2}{\pi}}\frac{d^{3/2}}{t\sigma}$$

← **should be $d^{1/2}$**

Improving bound on $\|A^{-1}\|$

Lemma:

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

Apply to random b

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

So

conjecture

$$\Pr\left[\|A^{-1}\| \geq t\right] \leq \frac{\Pr\left[\|A^{-1}b\| \geq \frac{t}{\sqrt{d}}\right]}{\text{const}} \leq \frac{1.823\sqrt{d}}{t}$$

Gaussian Elimination w/ Partial Pivoting

```
>> A = randn(2)

A =

   -0.4326    0.1253
   -1.6656    0.2877

>> b = randn(2,1)

b =

   -1.1465
    1.1909
```

```
>> x = A \ b

x =

   -5.6821
  -28.7583

>> norm(A*x - b)

ans =

  8.0059e-016
```

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!

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 & 1 \\ -1 & -1 & 1 & 0 & 0 & 1 \\ -1 & -1 & -1 & 1 & 0 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & 1 \end{pmatrix} \xrightarrow{\text{GEPP}} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & 16 \\ 0 & 0 & 0 & 0 & 0 & 32 \end{pmatrix}$$

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

```
>> 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

>> 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

$$\begin{pmatrix} -1 & 6 & 3 \\ 1 & -4 & -1 \\ 2 & -3 & -4 \end{pmatrix} \begin{pmatrix} 2 & -3 & -4 \\ 1 & -4 & -1 \\ -1 & 6 & 3 \end{pmatrix}$$

$$\begin{pmatrix} 2 & -3 & -4 \\ 0 & -5/2 & 1 \\ 0 & 9/2 & 1 \end{pmatrix} \begin{pmatrix} 2 & -3 & -4 \\ 0 & 9/2 & 1 \\ 0 & -5/2 & 1 \end{pmatrix}$$

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 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 know to have arisen under natural circumstances …

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

Nick Trefethen

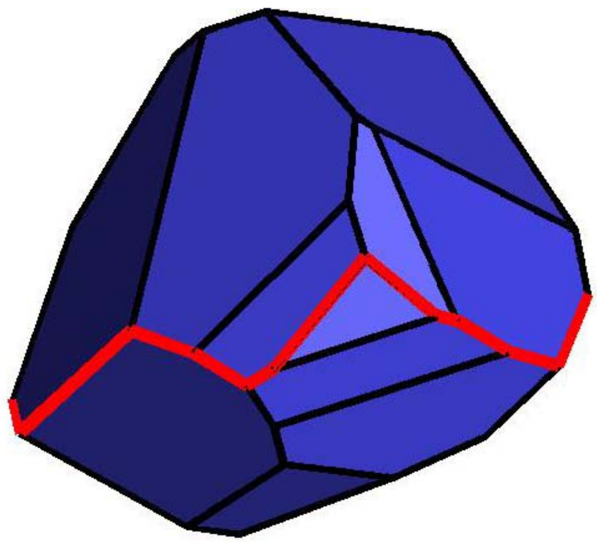
Theorem:

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

[Sankar-S-Teng]

Simplex Method for Linear Programming

$$\max \ c^T x$$
$$\text{s.t.} \quad Ax \leq b$$



Worst-Case: exponential

Average-Case: polynomial

Widely used in practice

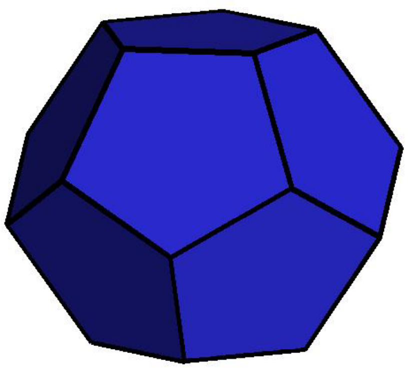
Smoothed Analysis of Simplex Method

max $c^T x$
s.t. $Ax \leq b$

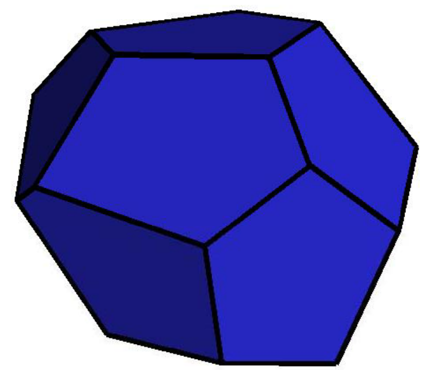
→

max $c^T x$
s.t. $(A + \sigma \|A\| G) x \leq b$

G is Gaussian

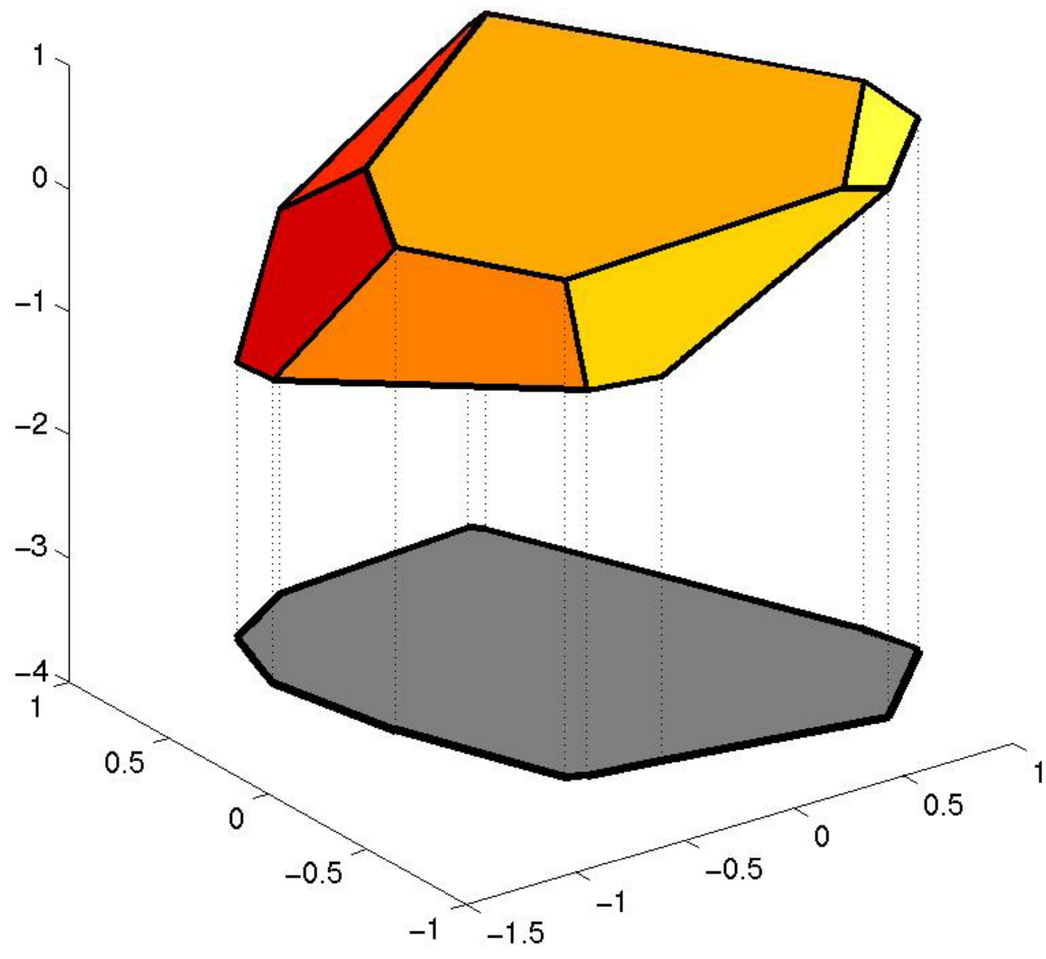


→

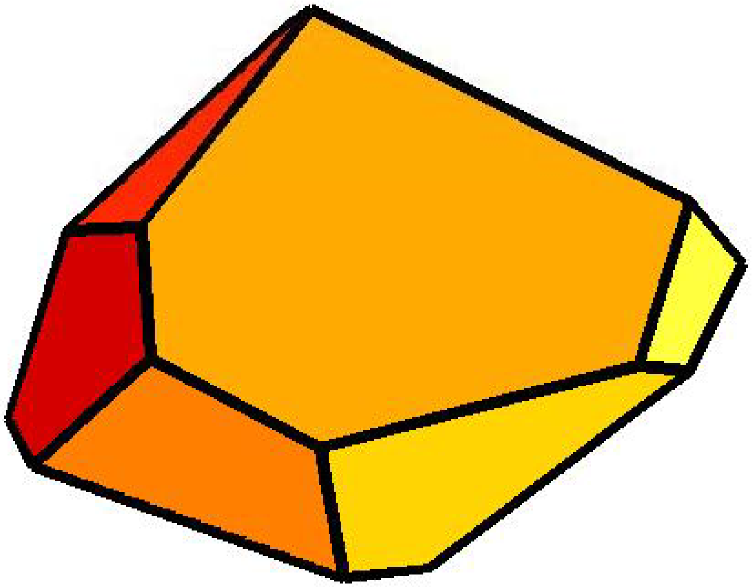


Theorem: For all A, b, c, simplex method takes expected time polynomial in $m, n, 1/\sigma$

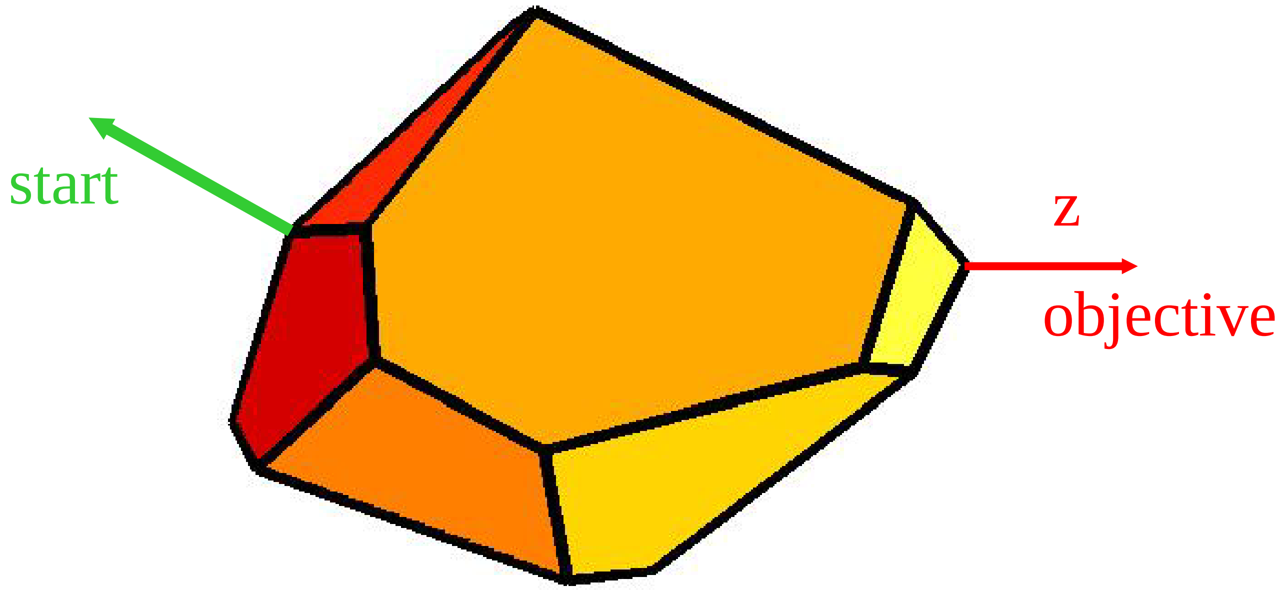
Shadow Vertices



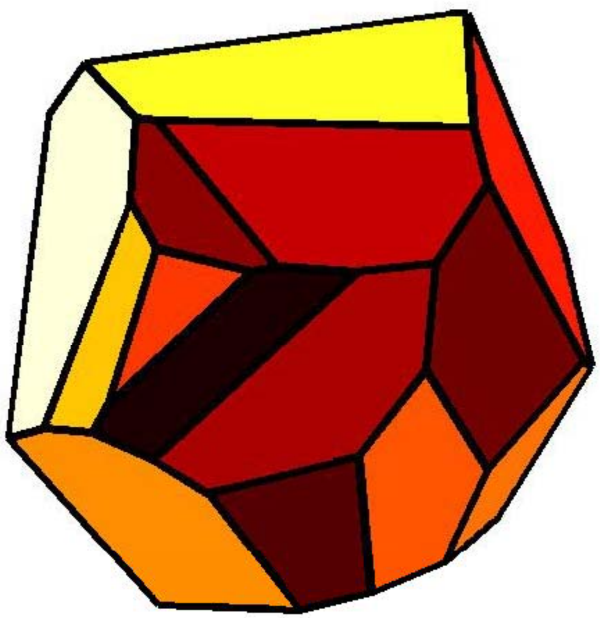
Another shadow

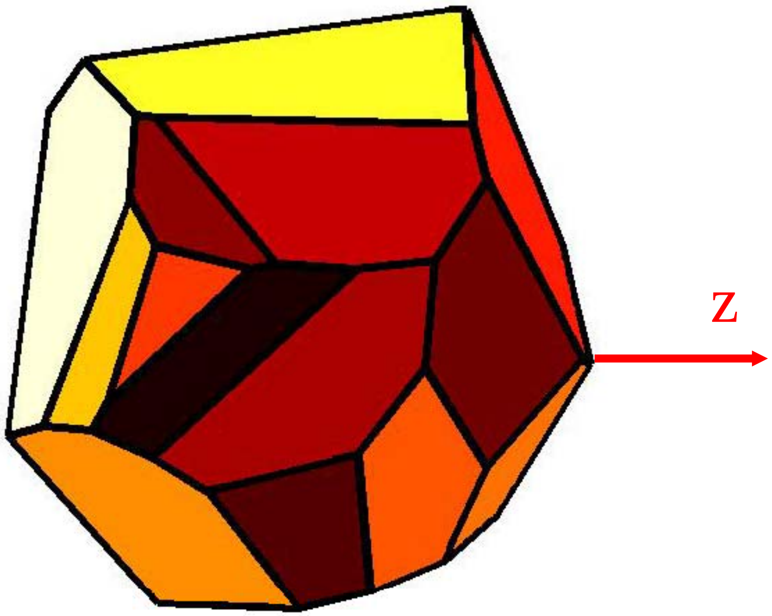


Shadow vertex pivot rule



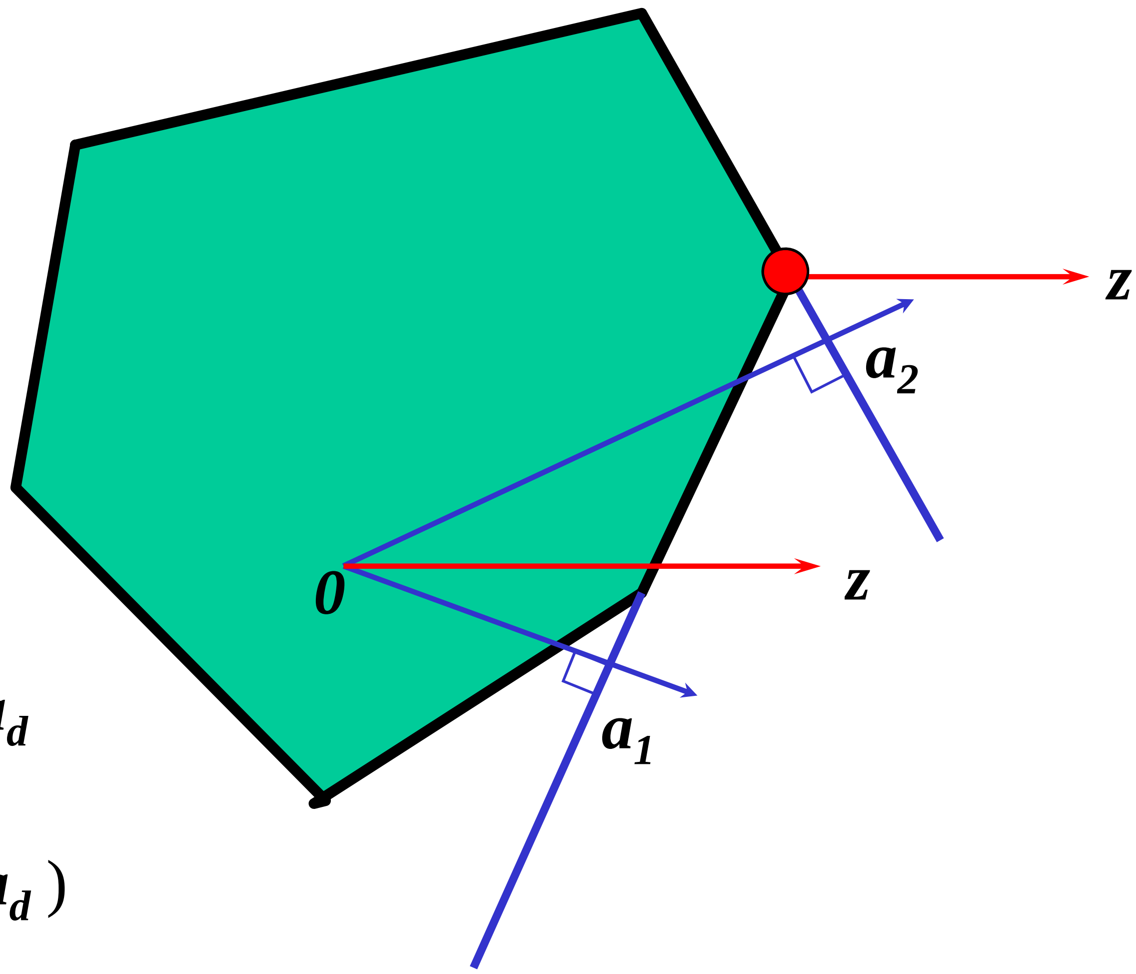
Theorem: For every plane, the expected size of the shadow of the perturbed tope is $poly(m,d,1/\sigma)$





Theorem: For every *z*, two-Phase Algorithm runs in expected time $poly(m,d,1/\sigma)$

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

$\boldsymbol{a}_1^T \boldsymbol{x} \leq 1$

$\boldsymbol{a}_2^T \boldsymbol{x} \leq 1$

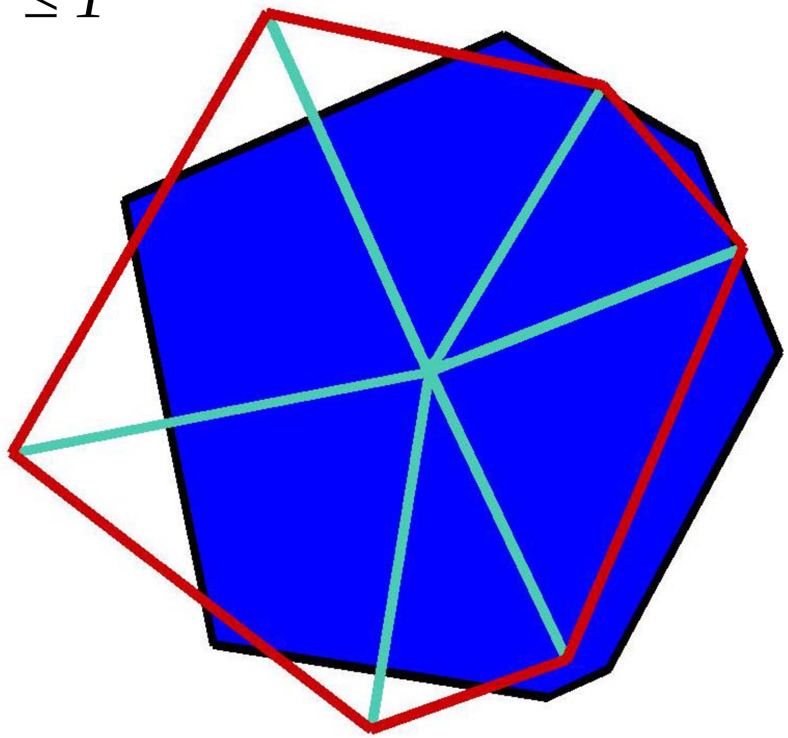
…

$\boldsymbol{a}_m^T \boldsymbol{x} \leq 1$

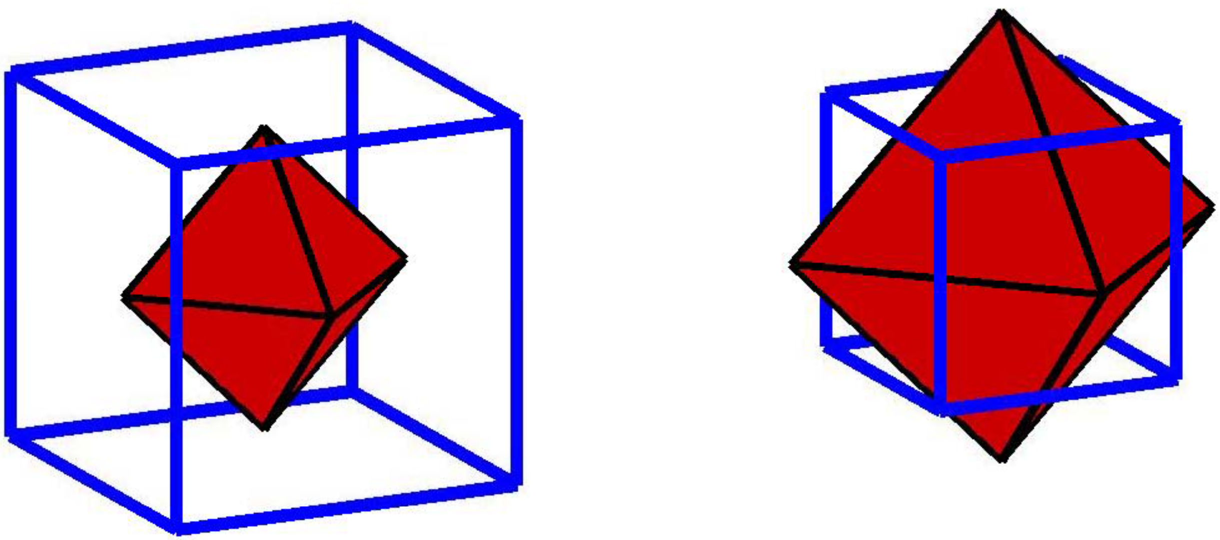
→

Polar

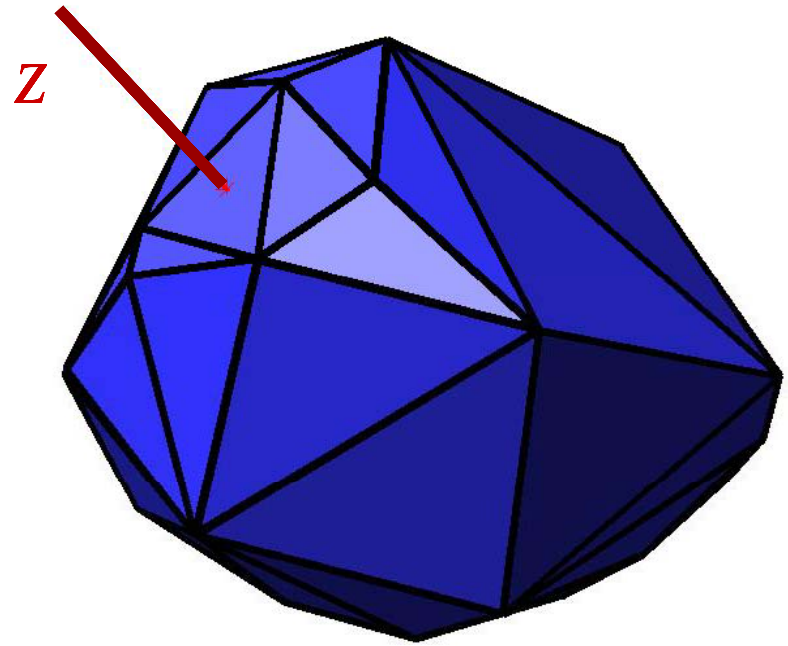
ConvexHull($\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_m$)



Polar

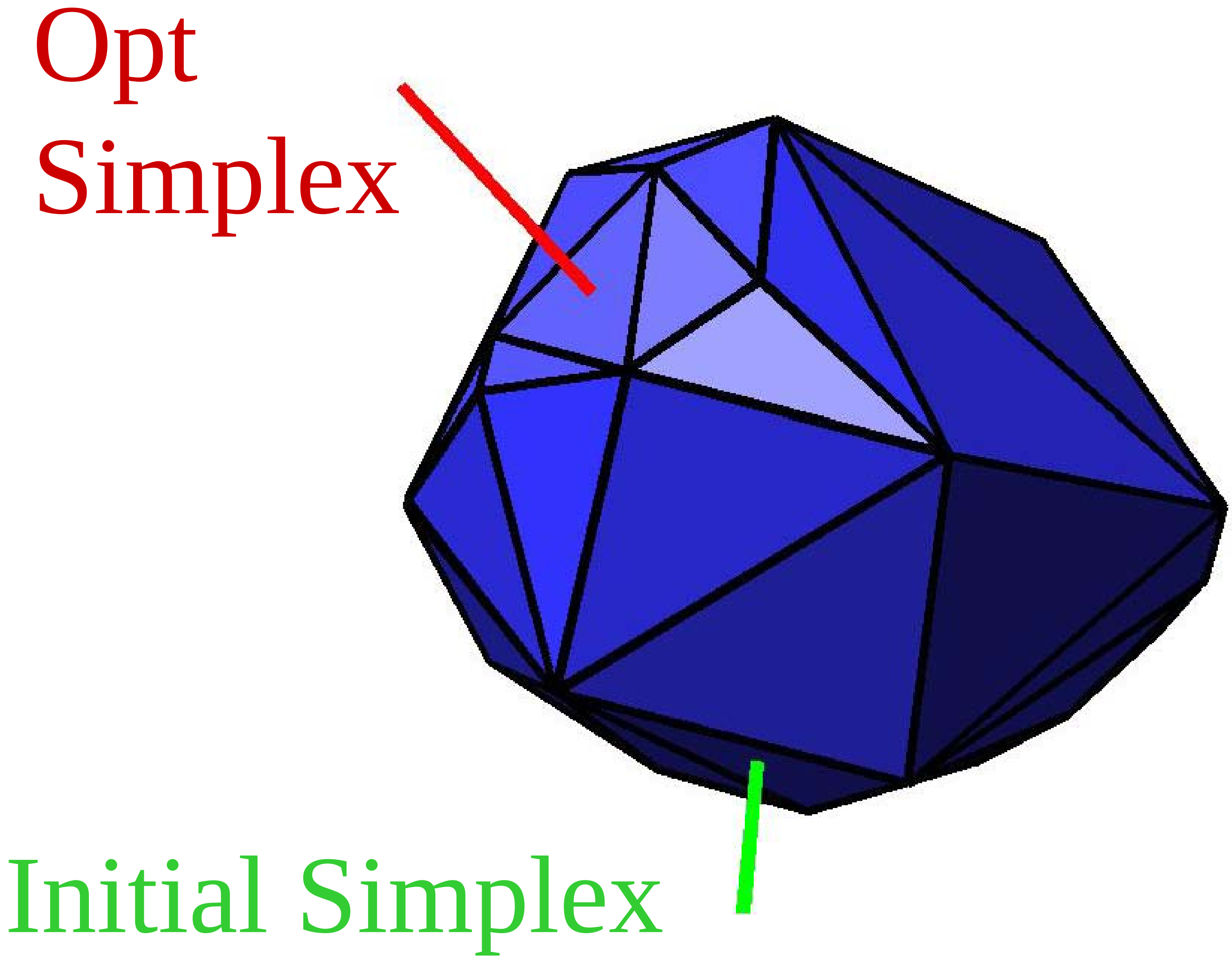


Polar Linear Program

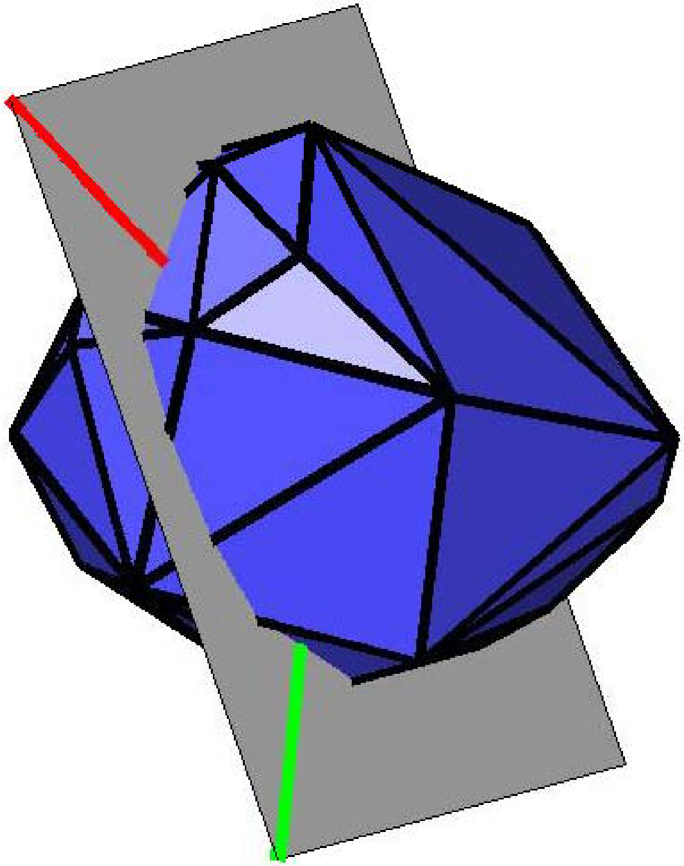


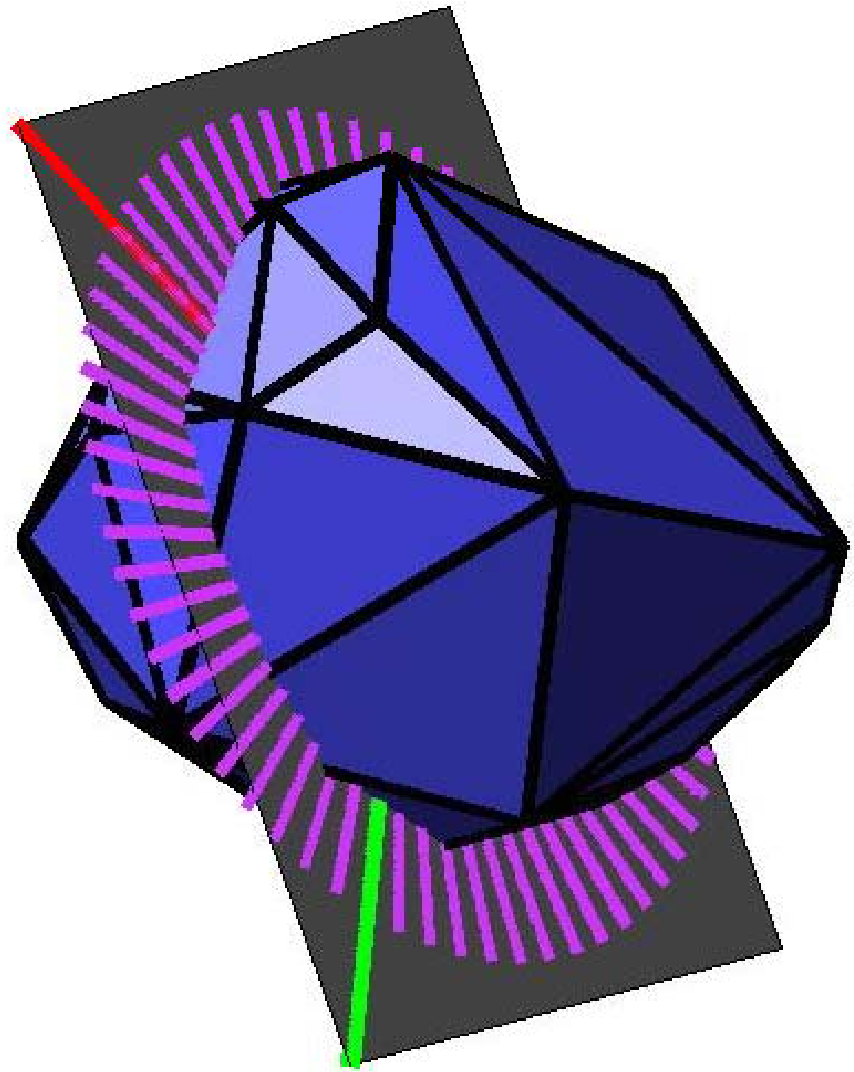
max α

$\alpha \mathbf{z} \in \text{ConvexHull}(\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m)$

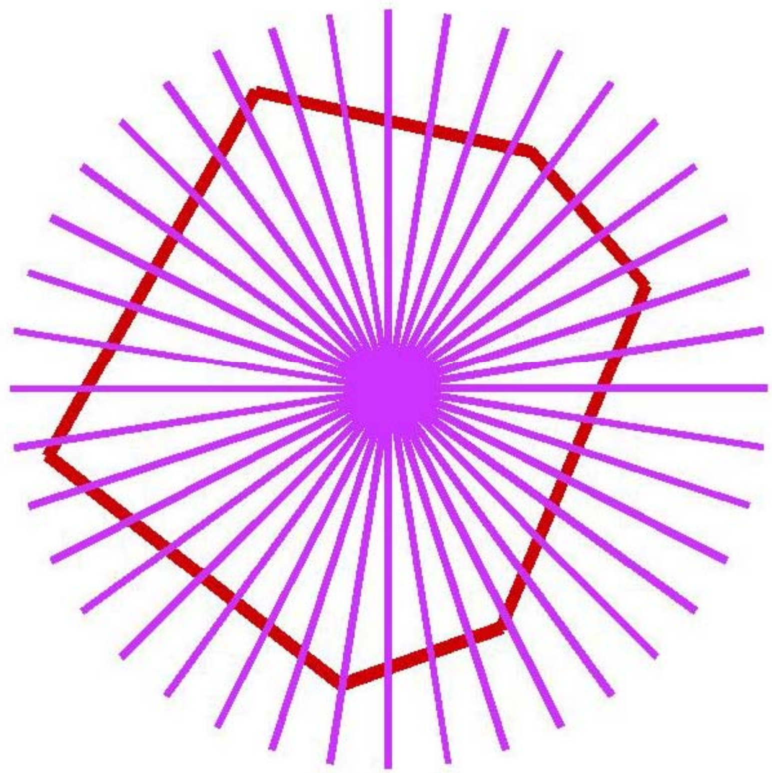

Opt
Simplex
Initial Simplex

Shadow vertex pivot rule

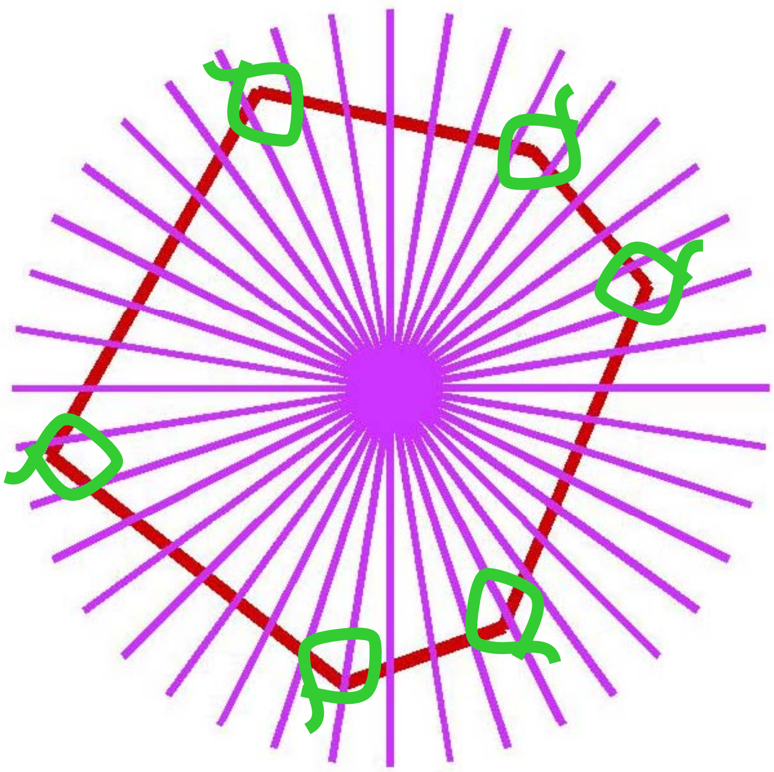




Count facets by discretizing to *N* directions, $N \to \infty$



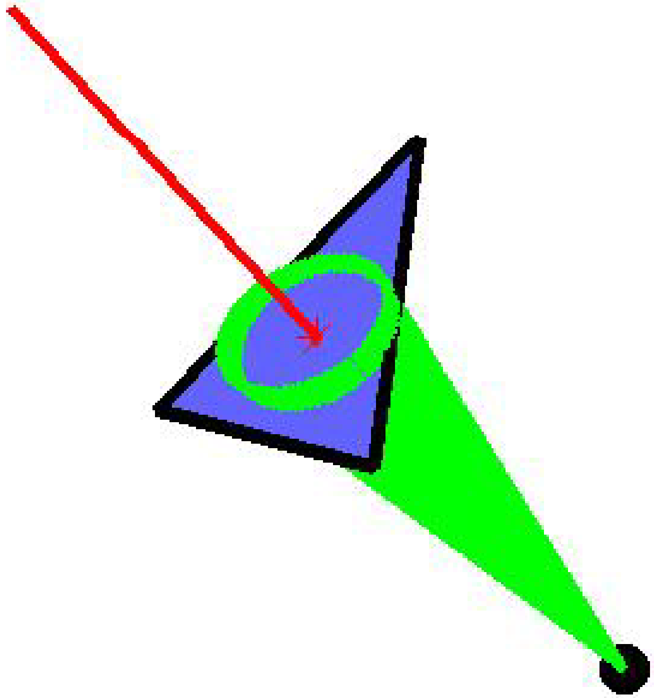
Count pairs in different facets

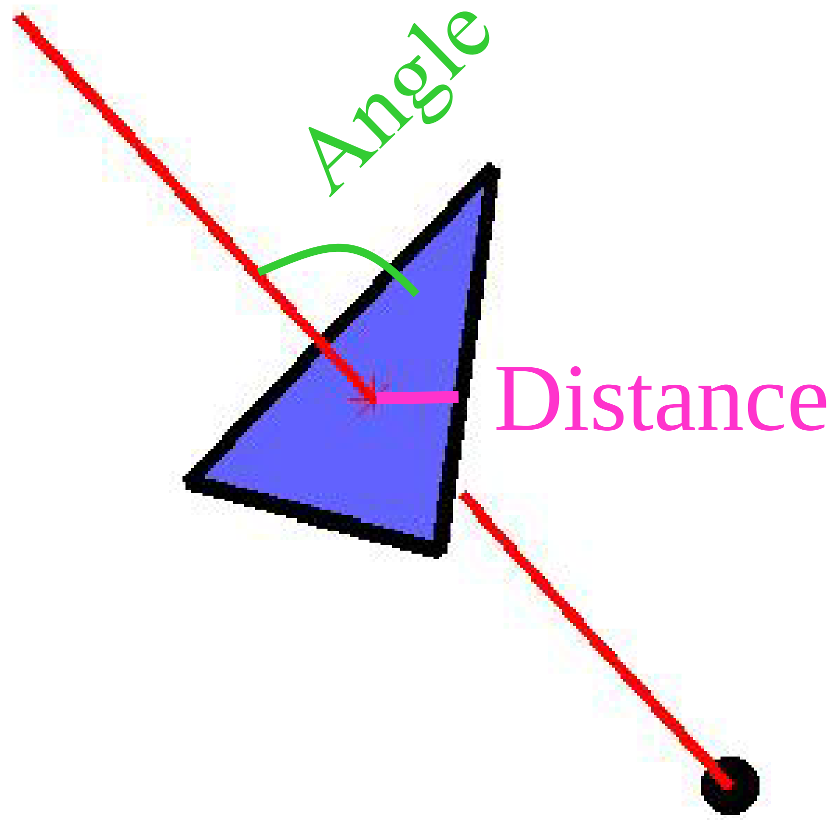


$$\Pr\left[\begin{matrix}\textbf{Different}\\ \textbf{Facets}\end{matrix}\right] < \boldsymbol{c/N}$$

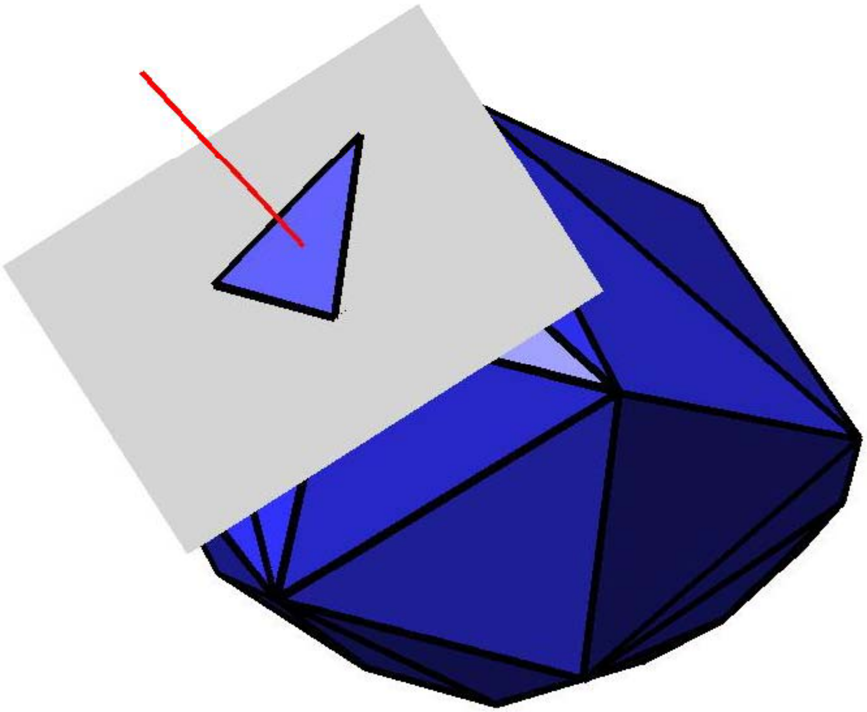
So, expect ***c*** Facets

Expect cone of large angle




Angle
Distance

Isolate on one Simplex



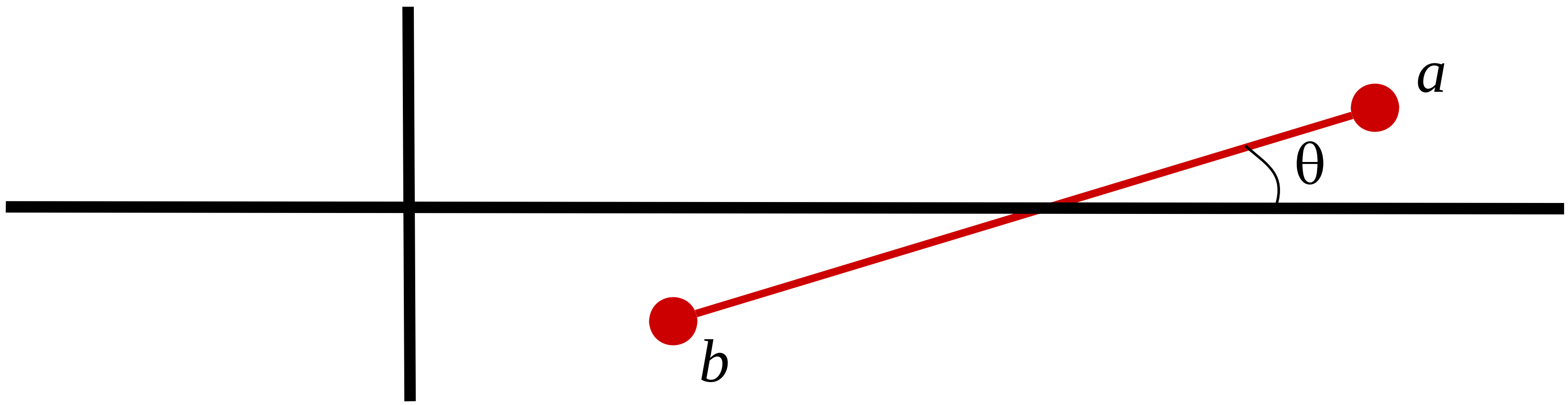
Integral Formulation

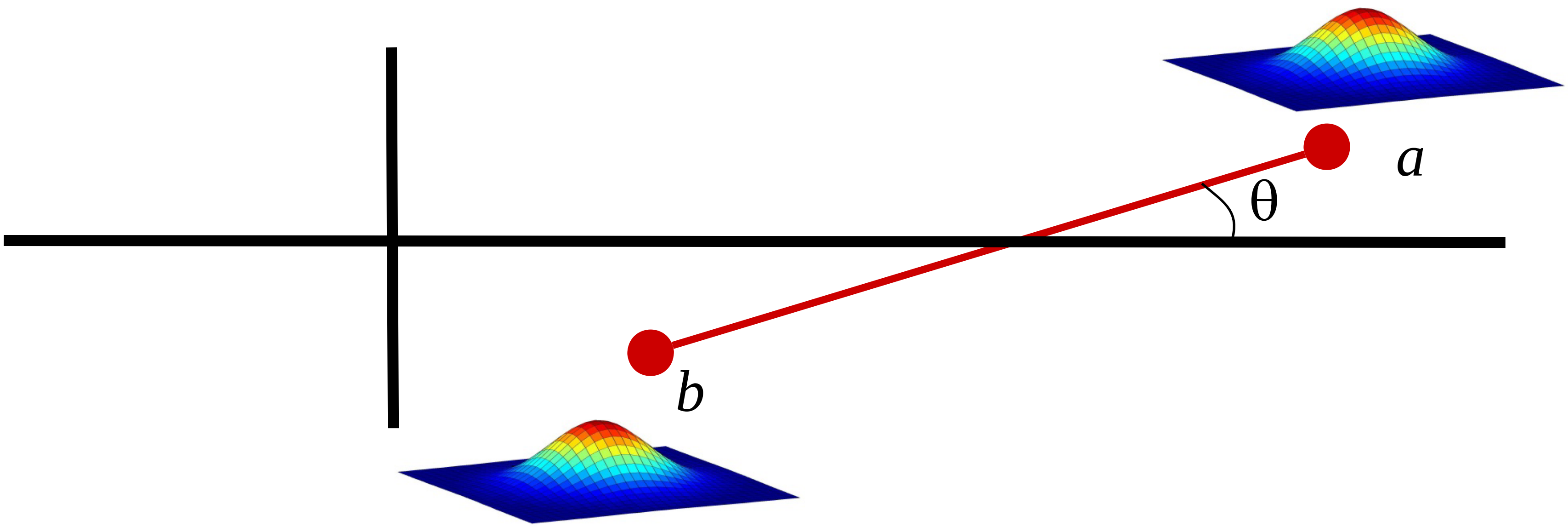
$$
\begin{aligned}
p_{\boldsymbol{z}}^{(1)}(\epsilon) \quad = \quad & \sum_{\pi \in \Pi(d,m)} \int_{\boldsymbol{\omega},r} \prod_{i=d+1}^{m} \left(\int_{\boldsymbol{b}_{\pi(i)}} \left[\langle \boldsymbol{\omega} | \boldsymbol{b}_{\pi(i)} \rangle \leq r \right] \mu_{\pi(i)}(\boldsymbol{b}_{\pi(i)})\, d\boldsymbol{b}_{\pi(i)} \right) \cdot \\
& \int_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_d} \left[\boldsymbol{z} \in \mathbf{Cone}(\boldsymbol{\omega}, r, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_d) \right] \left[\mathbf{ang}\left(\boldsymbol{z}, \partial \mathbf{Simplex}\left(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d \right) \right) < \epsilon \right] \\
& \prod_{i=1}^{d} \nu_i^{\boldsymbol{\omega},r}(\boldsymbol{c}_i) \mathbf{Vol}\left(\mathbf{Simplex}\left(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_d \right) \right)\, d\boldsymbol{\omega}\, dr\, d\boldsymbol{c}_1 \cdots d\boldsymbol{c}_d
\end{aligned}
$$

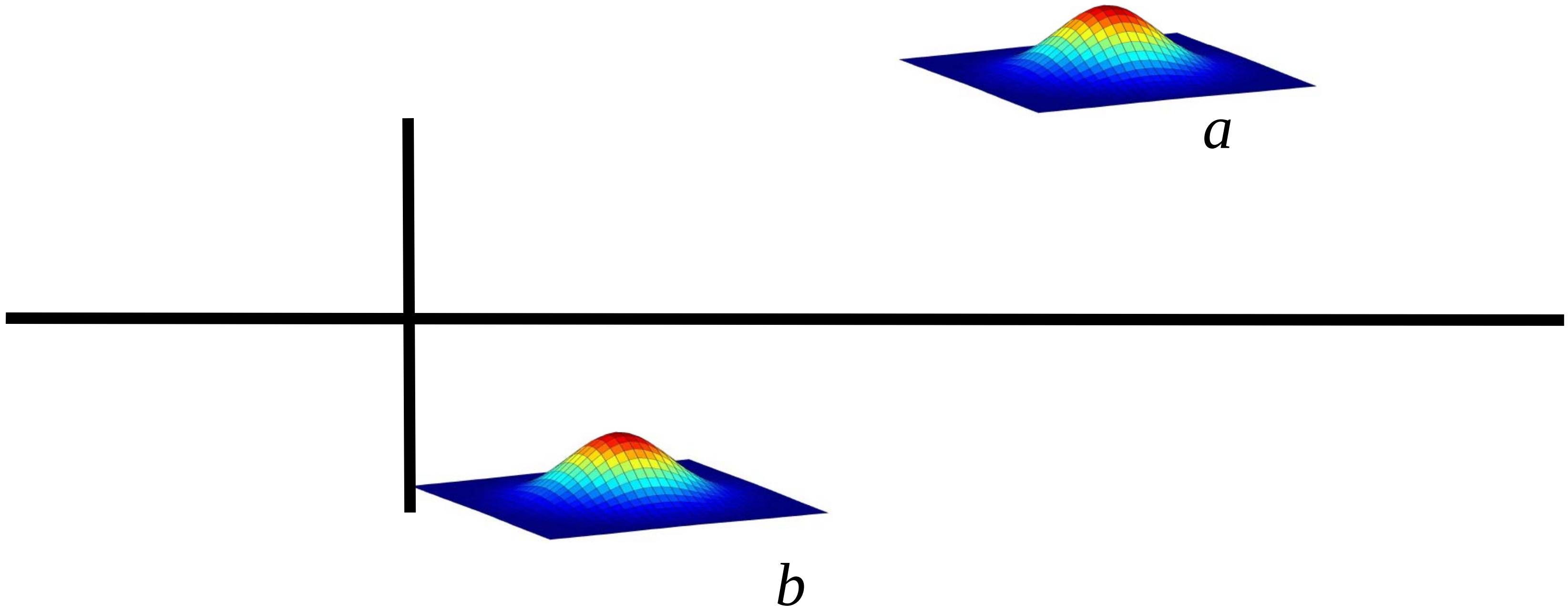
Example:

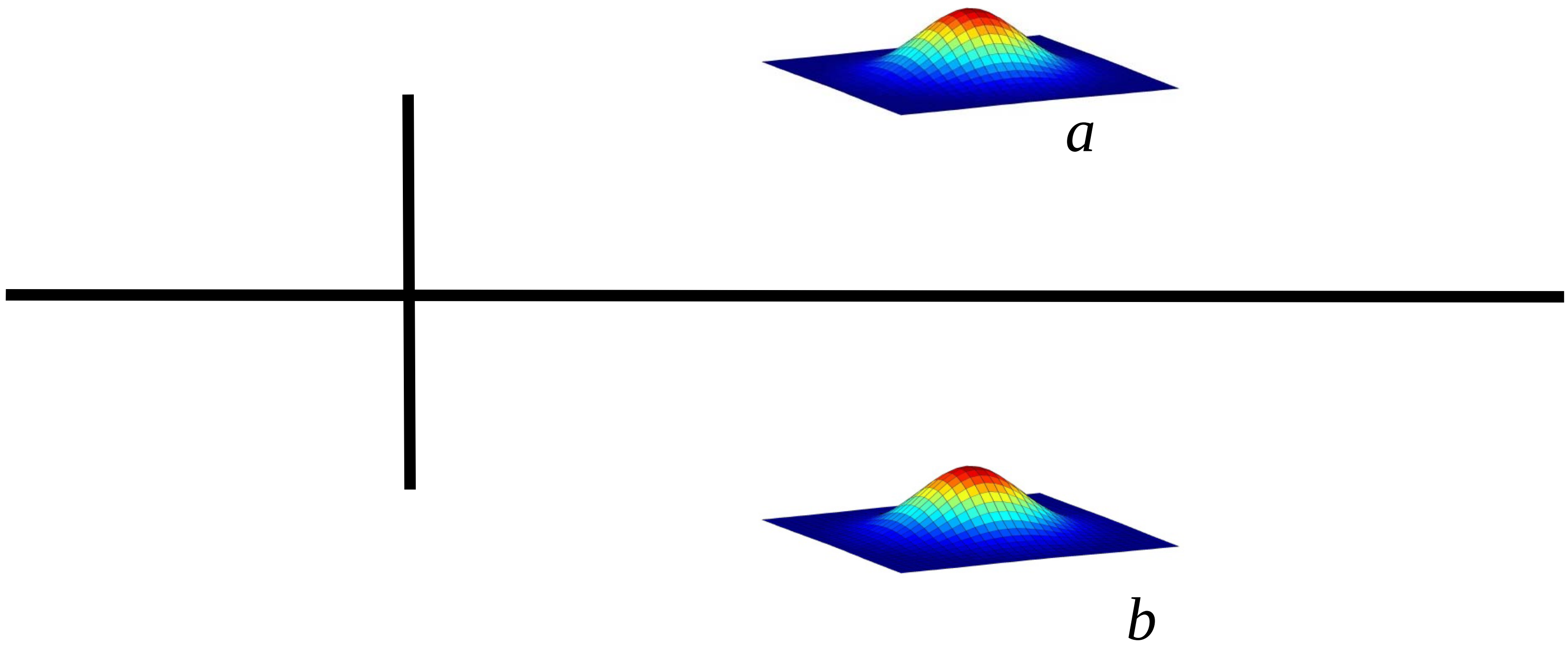
For *a* and *b* Gaussian distributed points, given that $\overline{ab}$ intersects x-axis

Prob[θ < ε] = O(ε²)

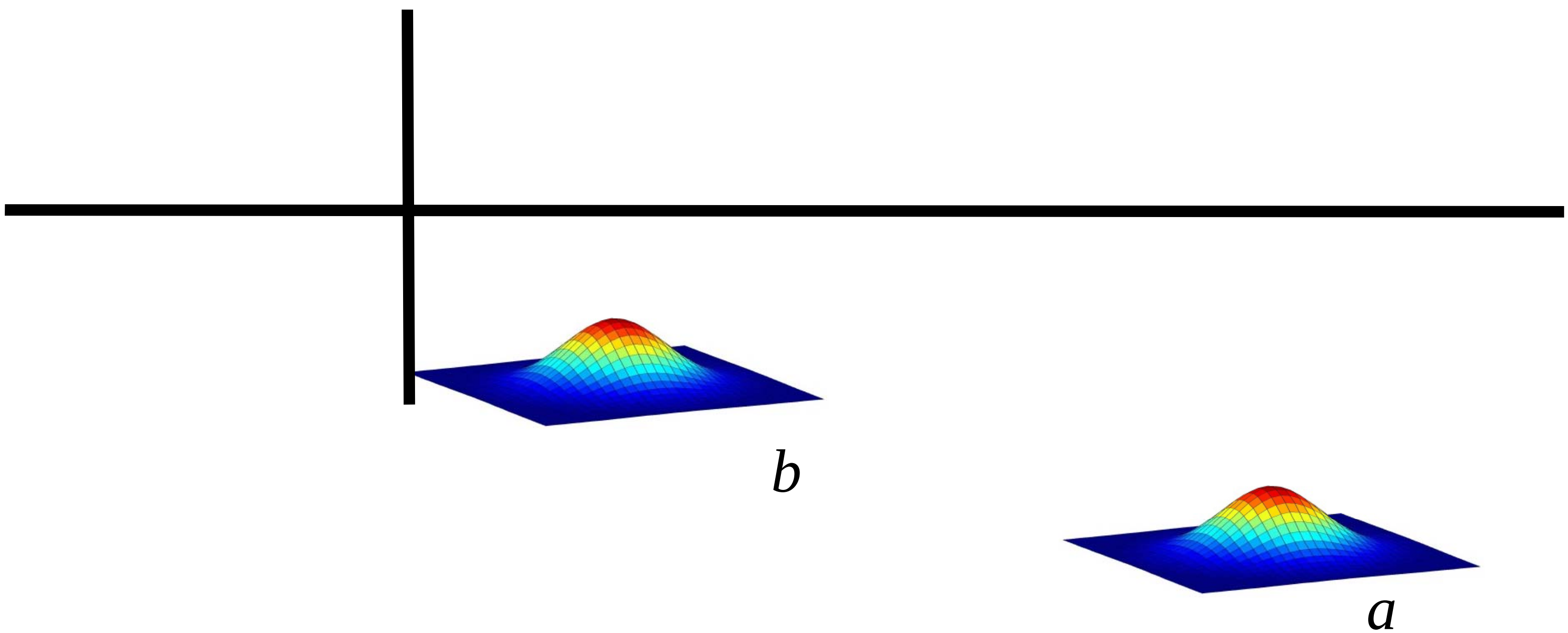


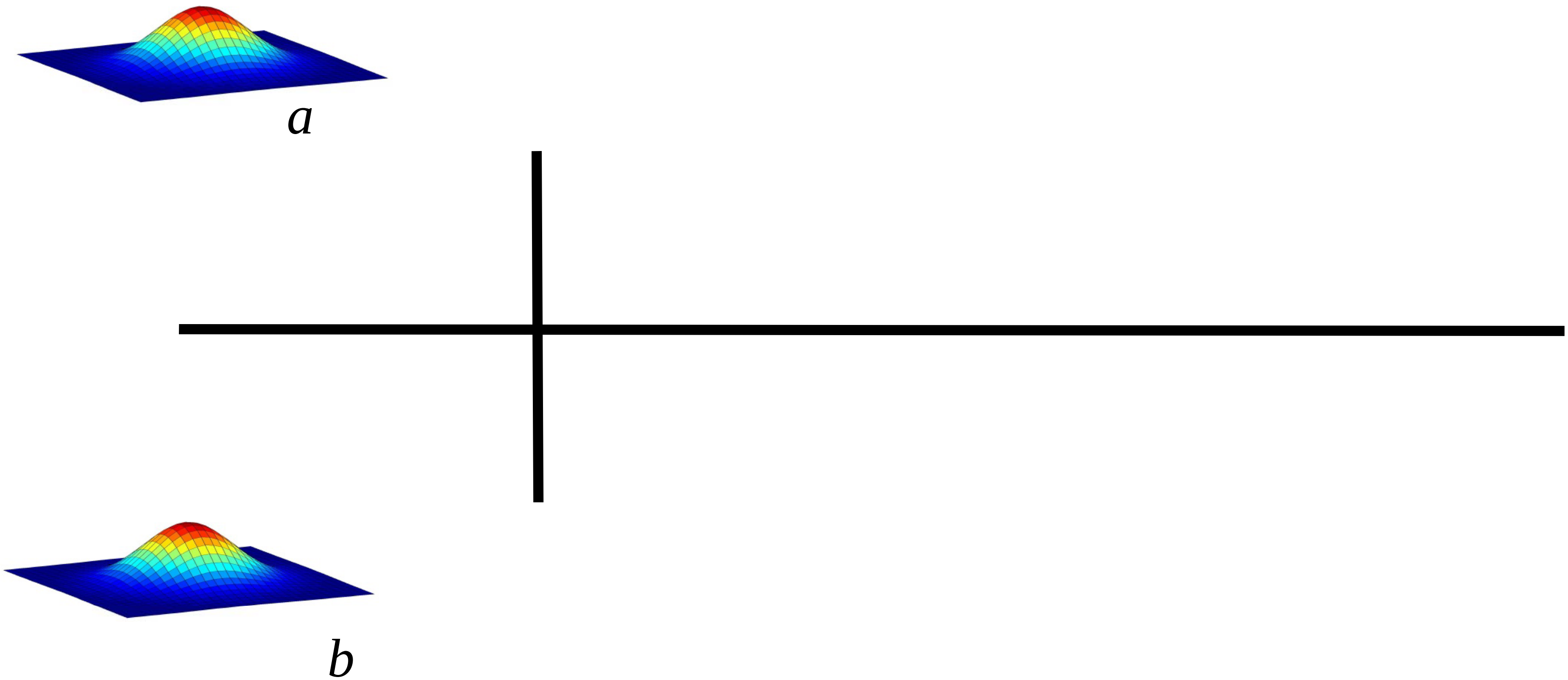

a
θ
b

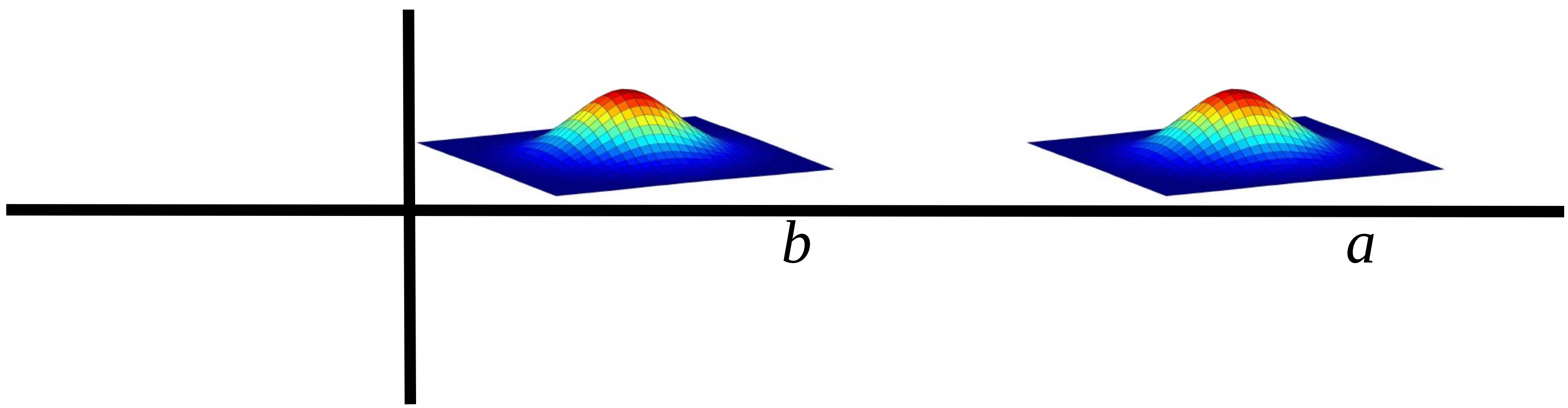

a
b

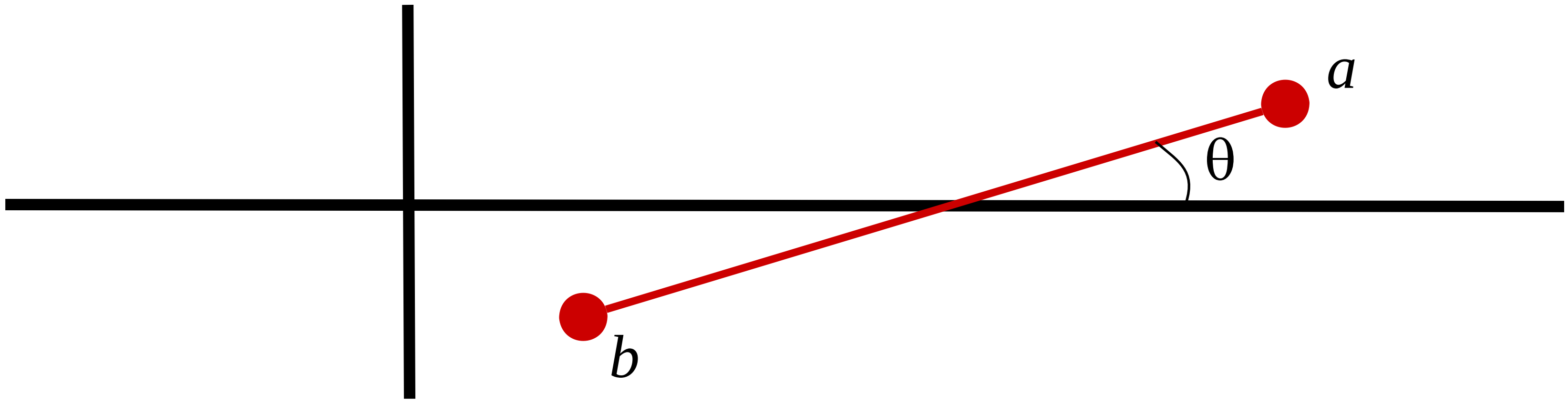


a
b


b
a


a
b


b
a

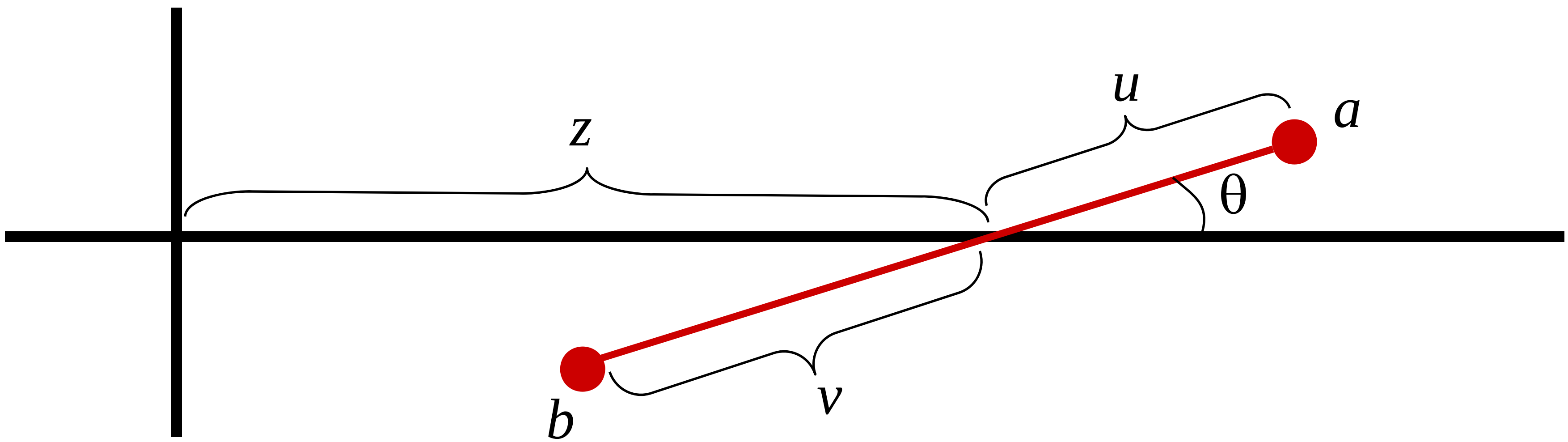


$$P_\varepsilon = \Pr[\theta < \varepsilon \mid \overline{ab} \cap axis \neq \emptyset]$$

$$= c \int_{a,b} [\theta < \varepsilon\,][\overline{ab} \cap axis \neq \emptyset]\, \mu_0(a)\mu_1(b)\,da\,db$$

Claim: **For $\varepsilon < \varepsilon_0$, $P_\varepsilon \lesssim \varepsilon^2$**

Change of variables



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

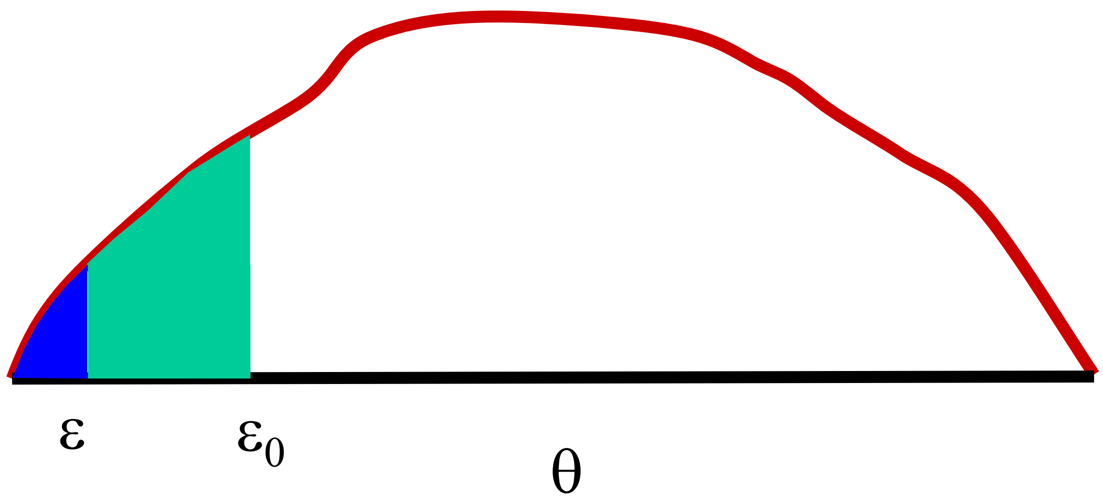
$$\mu_0(a) \longrightarrow \nu_0(u, z, \theta)$$

$$\mu_1(b) \longrightarrow \nu_1(v, z, \theta)$$

Analysis: **For $\varepsilon < \varepsilon_{0,}$ $P_\varepsilon \lesssim \varepsilon^2$**

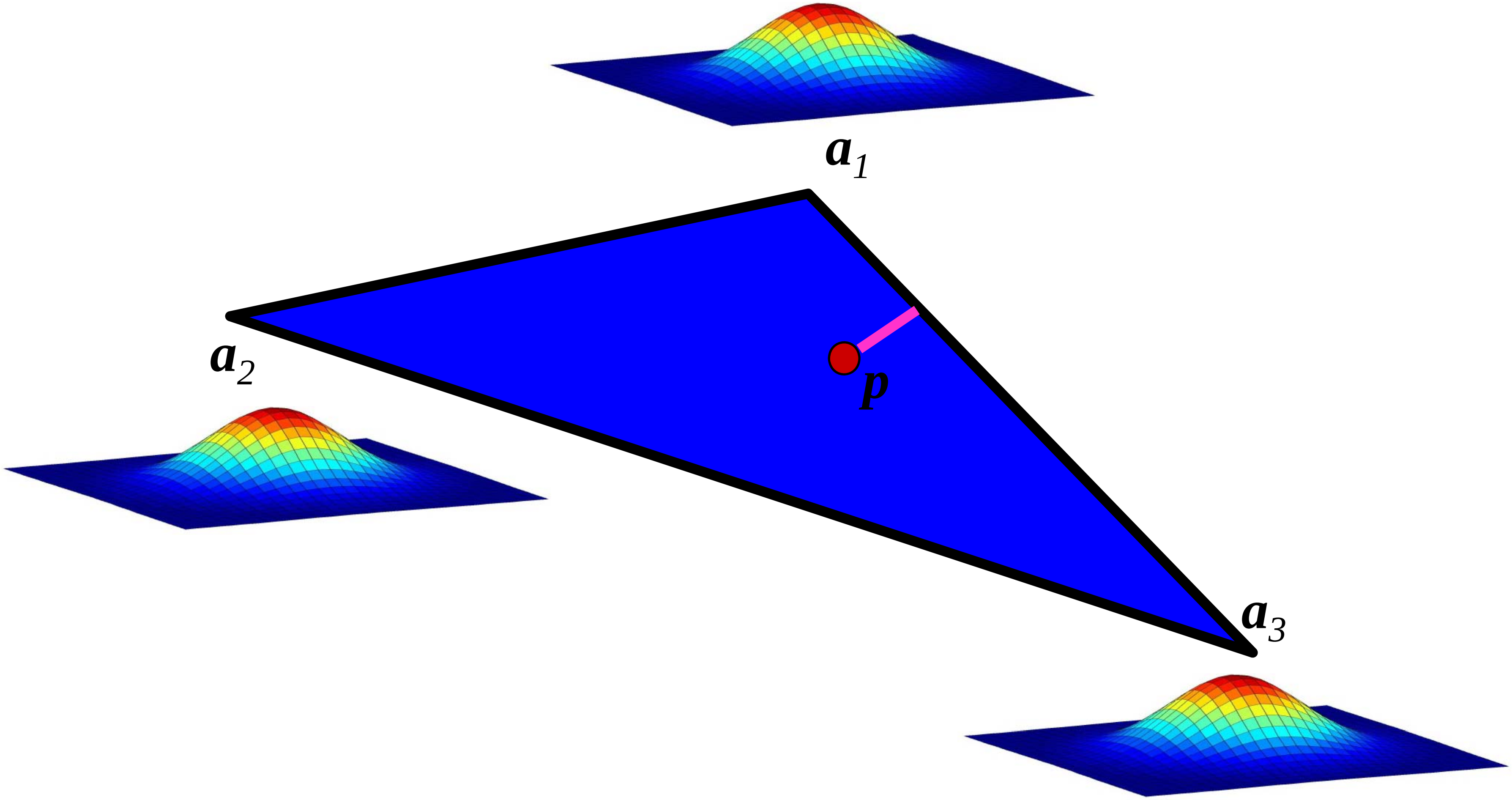
$$P_\varepsilon = c \int_{u,v,z} (u+v) \int_{\theta<\varepsilon} |\sin(\theta)|\, v_0(u,z,\theta)\, v_1(v,z,\theta)\, du\, dv\, dz\, d\theta$$

Slight change in θ has little effect on v_i for all but very rare u,v,z

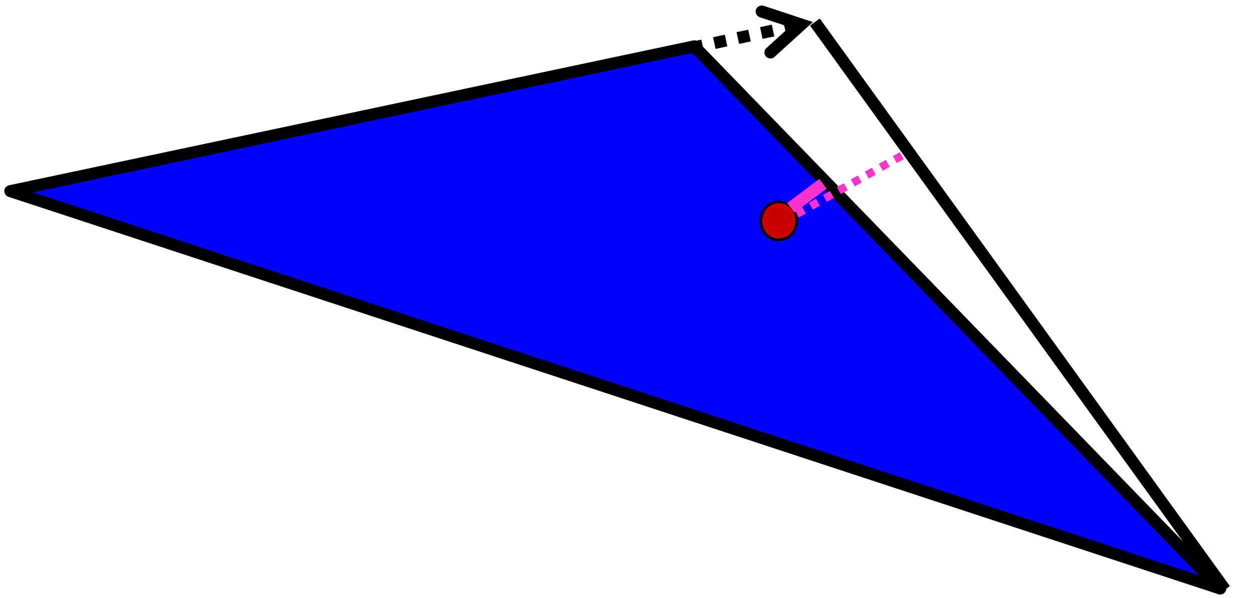


Distance:

Gaussian distributed corners



Idea: fix by perturbation



Title

Body text

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

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