Open Problems about the Simplex Method

Sophie Huiberts
CNRS, LIMOS

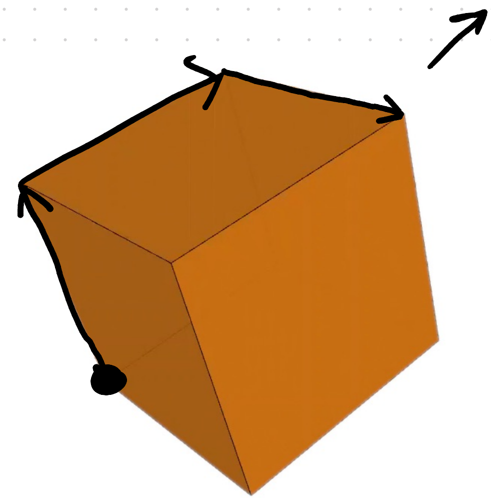
sophie.huiberts.me
@sophiehuiberts@mathstodon.xyz

Linear programming

maximize $c^T x$
subject to $Ax \leq b$

we get $A \in \mathbb{R}^{n \times d}$
$b \in \mathbb{R}^n$
$c \in \mathbb{R}^d$

we compute $x \in \mathbb{R}^d$

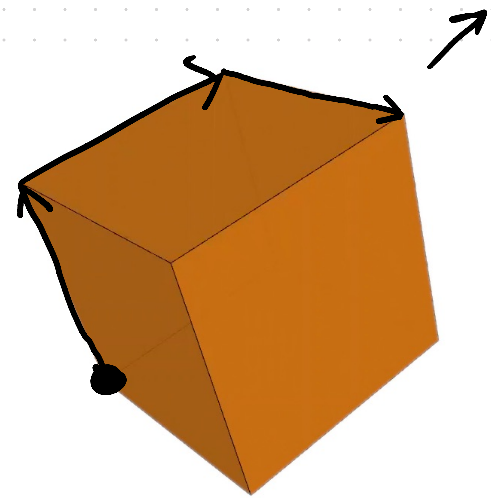


Linear programming

maximize $c^T x$
subject to $Ax \leq b$

we get $A \in \mathbb{R}^{n \times d}$
$b \in \mathbb{R}^n$
$c \in \mathbb{R}^d$

we compute $x \in \mathbb{R}^d$



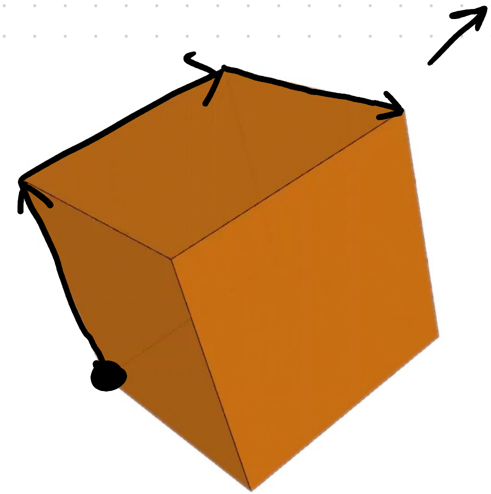
how many pivot steps?

Linear programming

maximize $c^T x$
subject to $Ax \le b$

we get $A \in \mathbb{R}^{n \times d}$
$b \in \mathbb{R}^n$
$c \in \mathbb{R}^d$

we compute $x \in \mathbb{R}^d$



how many pivot steps?
(assuming non-degeneracy)

In practice

The simplex method takes $2(n+d)$ pivot steps to solve an LP.

Worst-case complexity

Theorem The simplex method has exponential worst-case complexity*

*terms and conditions apply

Simplex method is
good in practice
bad in theory

Simplex method is
good in practice
bad in theory

this is a question for science

Simplex method is
good in practice
bad in theory

this is a question for science

many sophisticated papers
are written about this

Average-case analysis

Simplex method is
good in practice
bad in theory

this is a question for science

many sophisticated papers
are written about this

Average-case analysis

Slack ratios

Simplex method is
good in practice
bad in theory

this is a question for science

many sophisticated papers
are written about this

Average-case analysis

Slack ratios

Simplex method is

good in practice

bad in theory

Smoothed analysis

this is a question for science

many sophisticated papers are written about this

Average-case analysis

Slack ratios

Simplex method is
good in practice
bad in theory

Maximum subdeterminants

Smoothed analysis

this is a question for science

many sophisticated papers
are written about this

Average-case analysis

Slack ratios

Simplex method is
good in practice
bad in theory

Maximum subdeterminants

Smoothed analysis

this is a question for science

Circuit ratios

many sophisticated papers
are written about this

Simplex method is
good in practice
bad in theory

this is a question for science

many sophisticated papers
are written about this

Average-case analysis

Slack ratios

Maximum subdeterminants

Smoothed analysis

Circuit ratios

Lattice polytopes

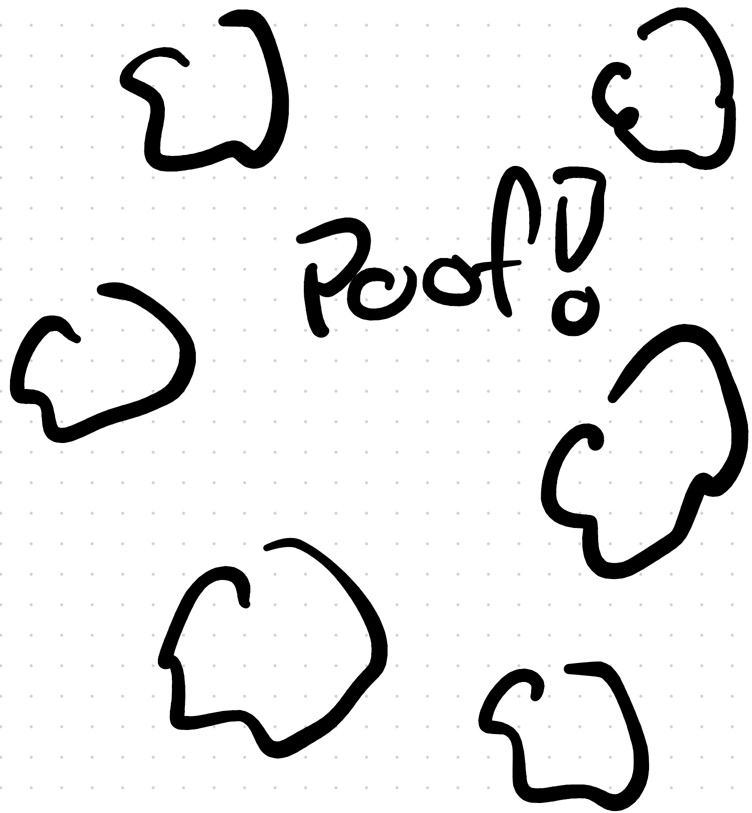
Is our theory any good?

At a conference

At a conference

At a conference


Theory has no value


Poof!


Theory has no value?

Does theory have any bearing on reality?

Instancewise assumptions

slack ratios

Distributional assumptions

smoothed analysis

Slack ratios

Suppose V is the vertex set of feasible region

$$\begin{aligned} &\text{maximize } c^T x \\ &\text{subject to } Ax \le b \end{aligned}$$

Let
$$\kappa = \frac{\min\{(b-Ax)_i : x \in V,\ i=1,\dots,n,\ (Ax)_i < b_i\}}{\max\{\|b-Ax\|_\infty : x \in V\}}$$

Slack ratios

Suppose V is the vertex set of feasible region

$$\begin{aligned} &\text{maximize } c^T x \\ &\text{subject to } Ax \le b \end{aligned}$$

Let $$\kappa = \frac{\min\{(b-Ax)_i : x \in V,\ i=1,\dots,n,\ (Ax)_i < b_i\}}{\max\{\|b-Ax\|_\infty : x \in V\}}$$

After T pivot steps the optimality gap is $\propto e^{-\frac{\kappa T}{d}}$

Slack ratios

Suppose V is the vertex set of feasible region

$$\begin{aligned}&\text{maximize } c^T x\\ &\text{subject to } Ax \le b\end{aligned}$$

Let $$\kappa = \frac{\min\{(b-Ax)_i \,:\, x \in V,\ i=1,\dots,n,\ (Ax)_i < b_i\}}{\max\{\|b-Ax\|_\infty \,:\, x \in V\}}$$

After T pivot steps the optimality gap is $\propto e^{-\frac{\kappa T}{d}}$

Optimal solution after $\frac{nd}{\kappa}\log(n/\kappa)$ steps

This $e^{-\frac{kT}{d}}$ was Dantzig's intuition for developping the simplex method.

never published?

This $e^{-\frac{kT}{d}}$ was Dantzig's intuition
for developping the simplex method.

never published?

Independently rediscovered by Kitahara & Mizuno.

This $e^{-\frac{\kappa T}{d}}$ was Dantzig's intuition
for developping the simplex method.

never published?

Independently rediscovered by Kitahara & Mizuno.

Kuno, Sano & Tsuruda proved computing κ is NP-Hard

A posteriori κ: run simplex, observe slacks

Name	$m \times n$	pivots	est δ	est γ	est bound
afiro	27×59	16	2.615	440.4	3.260×10^{6}
kb2	43×84	61	2.478×10^{-2}	72.15	1.781×10^{8}
sc50b	50×98	49	7.226×10^{-1}	324.9	3.185×10^{7}
blend	74×157	92	1.497×10^{-2}	38.27	5.208×10^{8}
sc105	105×208	103	2.050×10^{-1}	708.9	1.395×10^{9}
scagr7	129×269	140	3.304	6552	1.236×10^{9}
sc205	205×408	236	5.187×10^{-2}	2380	8.891×10^{10}
beaconfd	173×435	102	3.716×10^{-2}	5607	2.798×10^{11}
lotfi	153×461	239	5.474×10^{-3}	4179	1.443×10^{12}
bore3d	233×548	200	2.091×10^{-4}	301.2	5.209×10^{12}
sctap1	300×780	328	1.154×10^{-8}	2.000	1.444×10^{15}
agg2	516×818	218	3.047×10^{-2}	1.676×10^{6}	8.060×10^{14}
scagr25	471×971	742	3.007×10^{-2}	1.118×10^{4}	4.964×10^{12}
standmps	467×1542	279	1.998×10^{-2}	101.1	7.718×10^{10}

Kuno
Sano
Tsuruda

A posteriori κ: run simplex, observe slacks

Name	m×n	pivots	est δ	est γ	est bound
afiro	27×59	16	2.615	440.4	3.260×10^{6}
kb2	43×84	61	2.478×10^{-2}	72.15	1.781×10^{8}
sc50b	50×98	49	7.226×10^{-1}	324.9	3.185×10^{7}
blend	74×157	92	1.497×10^{-2}	38.27	5.208×10^{8}
sc105	105×208	103	2.050×10^{-1}	708.9	1.395×10^{9}
scagr7	129×269	140	3.304	6552	1.236×10^{9}
sc205	205×408	236	5.187×10^{-2}	2380	8.891×10^{10}
beaconfd	173×435	102	3.716×10^{-2}	5607	2.798×10^{11}
lotfi	153×461	239	5.474×10^{-3}	4179	1.443×10^{12}
bore3d	233×548	200	2.091×10^{-4}	301.2	5.209×10^{12}
sctap1	300×780	328	1.154×10^{-8}	2.000	1.444×10^{15}
agg2	516×818	218	3.047×10^{-2}	1.676×10^{6}	8.060×10^{14}
scagr25	471×971	742	3.007×10^{-2}	1.118×10^{4}	4.964×10^{12}
standmps	467×1542	279	1.998×10^{-2}	101.1	7.718×10^{10}

CAN conclude bound is nontrivial

Kuno
Sano
Tsuruda

A posteriori κ: run simplex, observe slacks

Name	m×n	pivots	est δ	est γ	est bound
afiro	27×59	16	2.615	440.4	3.260×10^{6}
kb2	43×84	61	2.478×10^{-2}	72.15	1.781×10^{8}
sc50b	50×98	49	7.226×10^{-1}	324.9	3.185×10^{7}
blend	74×157	92	1.497×10^{-2}	38.27	5.208×10^{8}
sc105	105×208	103	2.050×10^{-1}	708.9	1.395×10^{9}
scagr7	129×269	140	3.304	6552	1.236×10^{9}
sc205	205×408	236	5.187×10^{-2}	2380	8.891×10^{10}
beaconfd	173×435	102	3.716×10^{-2}	5607	2.798×10^{11}
lotfi	153×461	239	5.474×10^{-3}	4179	1.443×10^{12}
bore3d	233×548	200	2.091×10^{-4}	301.2	5.209×10^{12}
sctap1	300×780	328	1.154×10^{-8}	2.000	1.444×10^{15}
agg2	516×818	218	3.047×10^{-2}	1.676×10^{6}	8.060×10^{14}
scagr25	471×971	742	3.007×10^{-2}	1.118×10^{4}	4.964×10^{12}
standmps	467×1542	279	1.998×10^{-2}	101.1	7.718×10^{10}

CAN conclude bound is nontrivial

CAN NOT conclude much else

Kuno
Sano
Tsuruda

Open Questions:

Is κ also NP-Hard to approximate?

can a MIP solver do it anyway?

Open Questions:

Is κ also NP-Hard to approximate?

can a MIP solver do it anyway?

Does a good κ correlate with faster solving?

does presolve improve κ?

Open Questions:

Is κ also NP-Hard to approximate?

can a MIP solver do it anyway?

Does a good κ correlate with faster solving?

does presolve improve κ?

Are these the right questions to ask?

Smooted analysis

Let $\bar{A} \in \mathbb{R}^{n \times d}$ have rows of norm ≤ 1.

$\bar{b} \in [-1, 1]^n$, $c \in \mathbb{R}^d$

Let $\hat{A}$, $\hat{b}$ have iid $N(0, \sigma^2)$ entries

Smooted analysis

Let $\bar{A} \in \mathbb{R}^{n \times d}$ have rows of norm ≤ 1.

$\bar{b} \in [-1, 1]^n$, $c \in \mathbb{R}^d$

Let $\hat{A}, \hat{b}$ have iid $N(0, \sigma^2)$ entries.

Then

$$\max_{\bar{A}, \bar{b}, c} \mathbb{E}_{\hat{A}, \hat{b}} \left[\begin{array}{c} \text{time to solve} \\ \text{maximize } c^T x \\ \text{s.t } (\bar{A} + \hat{A})x \leq \bar{b} + \hat{b} \end{array} \right] \leq \text{poly}(n, d, \sigma^{-1})$$

Why smoothed analysis?

Why smoothed analysis?

independent measurement/numerical errors do not conspire against your algorithm.

Why smoothed analysis?

independent measurement/numerical errors do not conspire against your algorithm.

interpolates between worst case and average case analysis.

Why smoothed analysis?

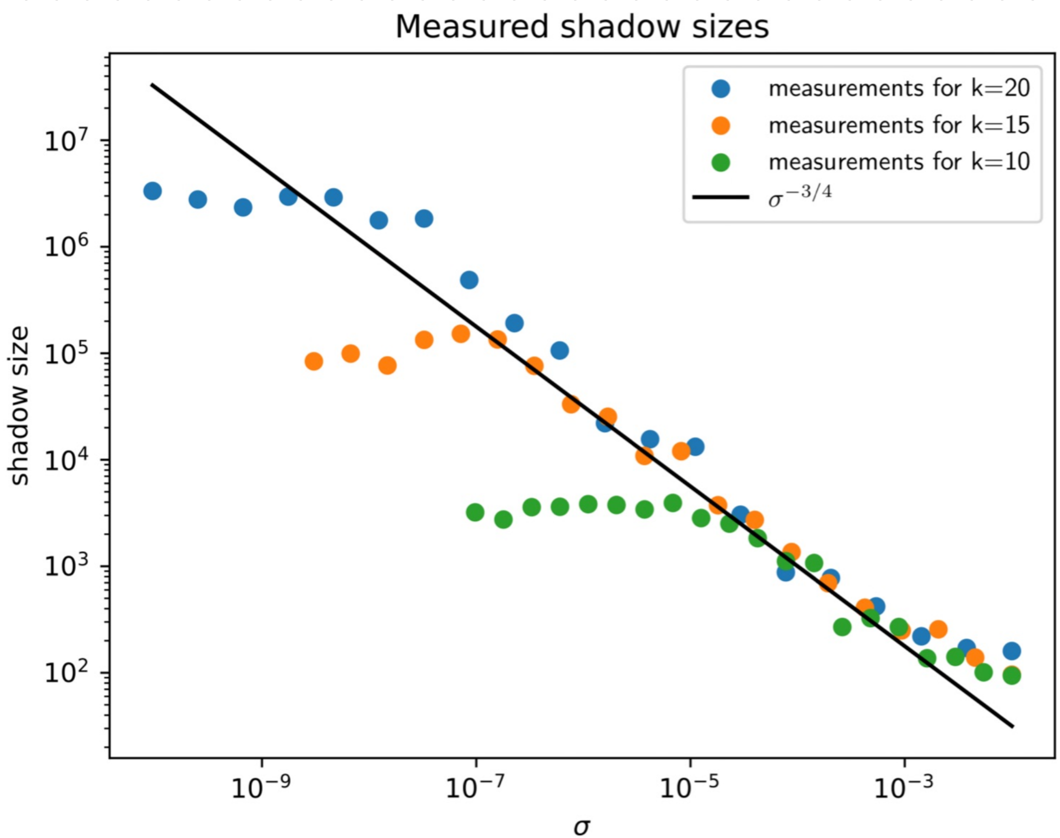
independent measurement/numerical errors
do not conspire against your algorithm.

interpolates between worst case and
average case analysis.

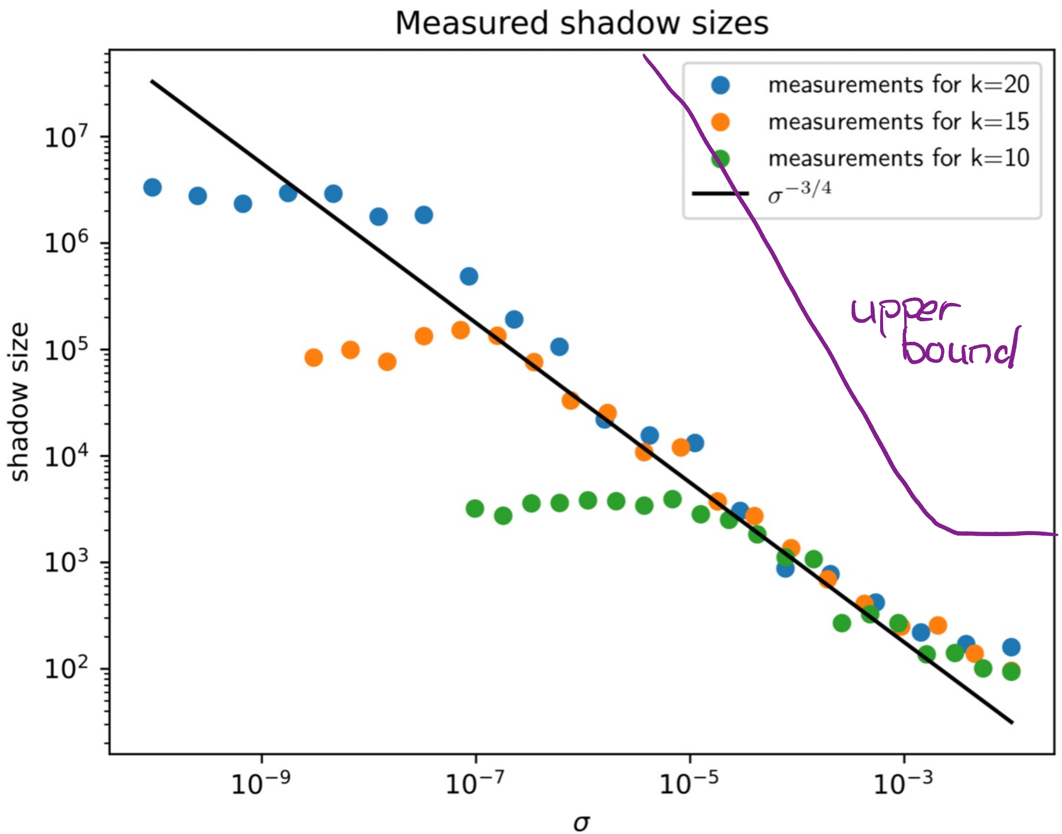
shows algorithm is fast on average
in every large enough neighborhood

Can any of the resulting insights be tested experimentally?

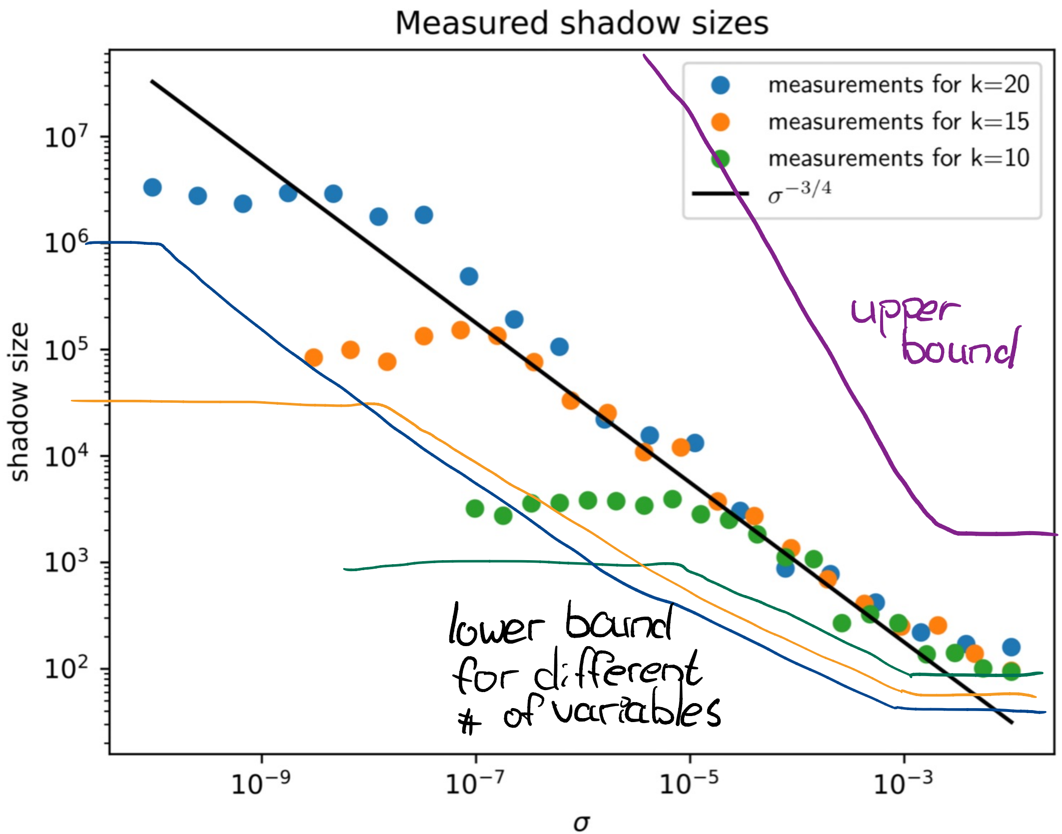
Synthetic data



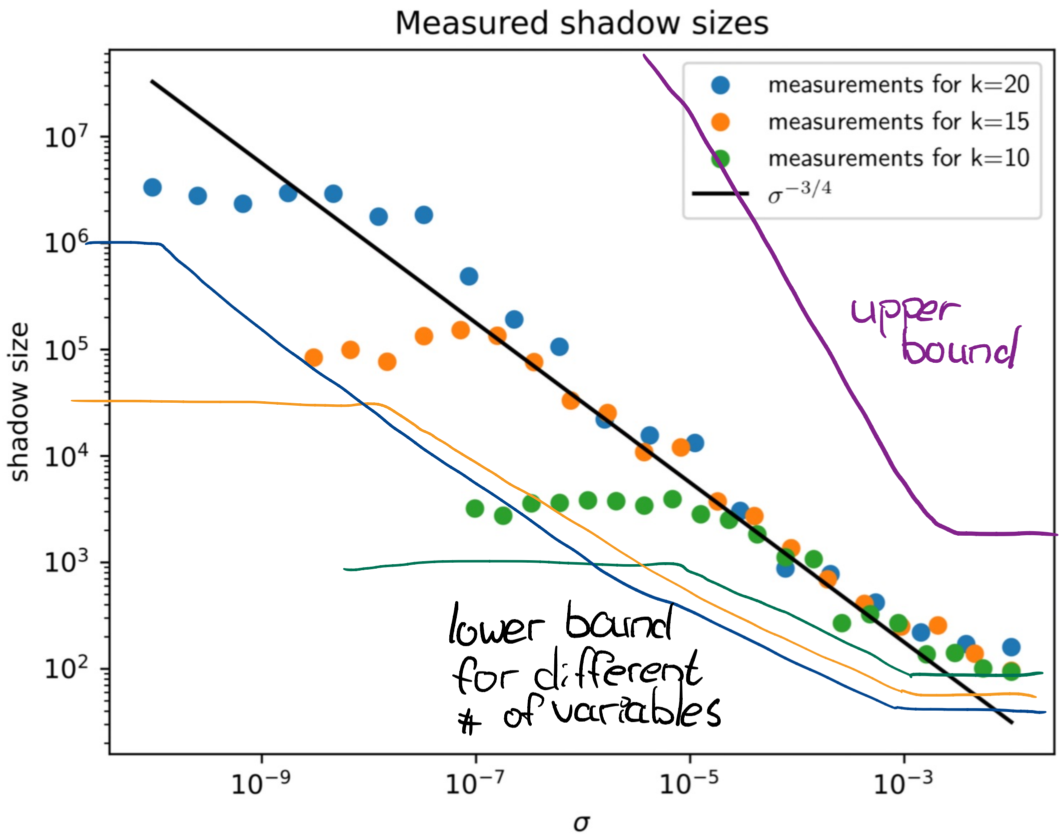
Synthetic data



Synthetic data

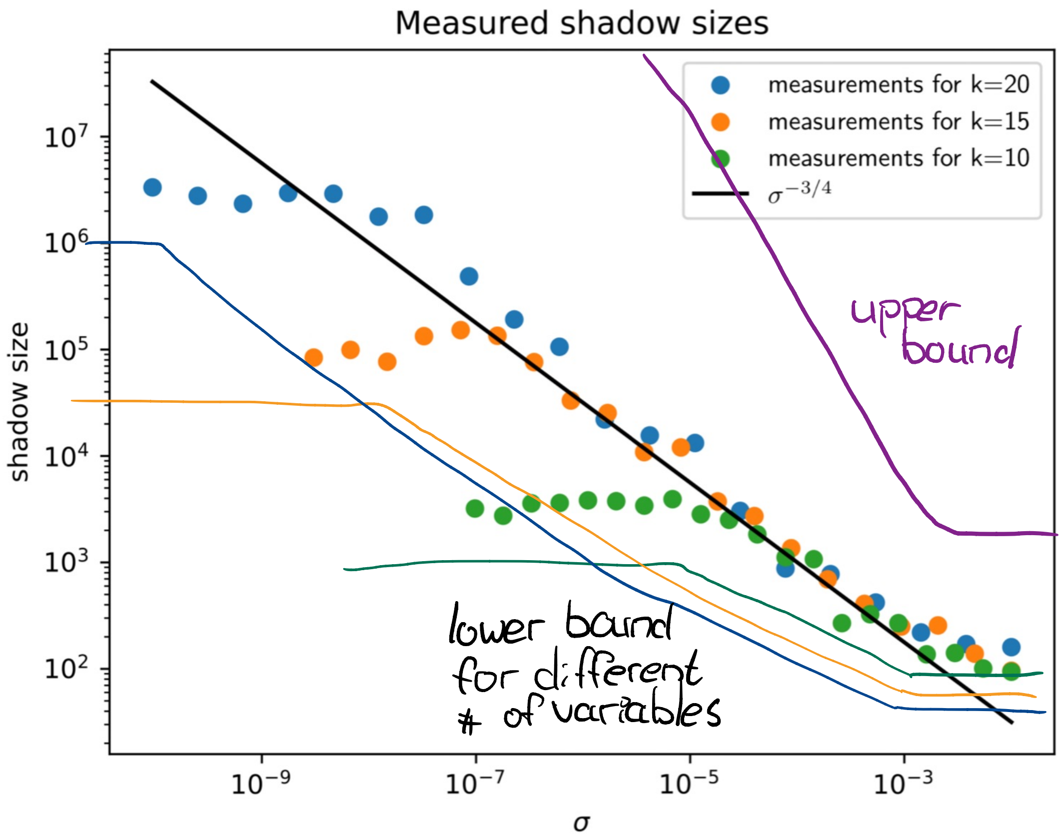


Synthetic data



CAN tell if theorem is tight

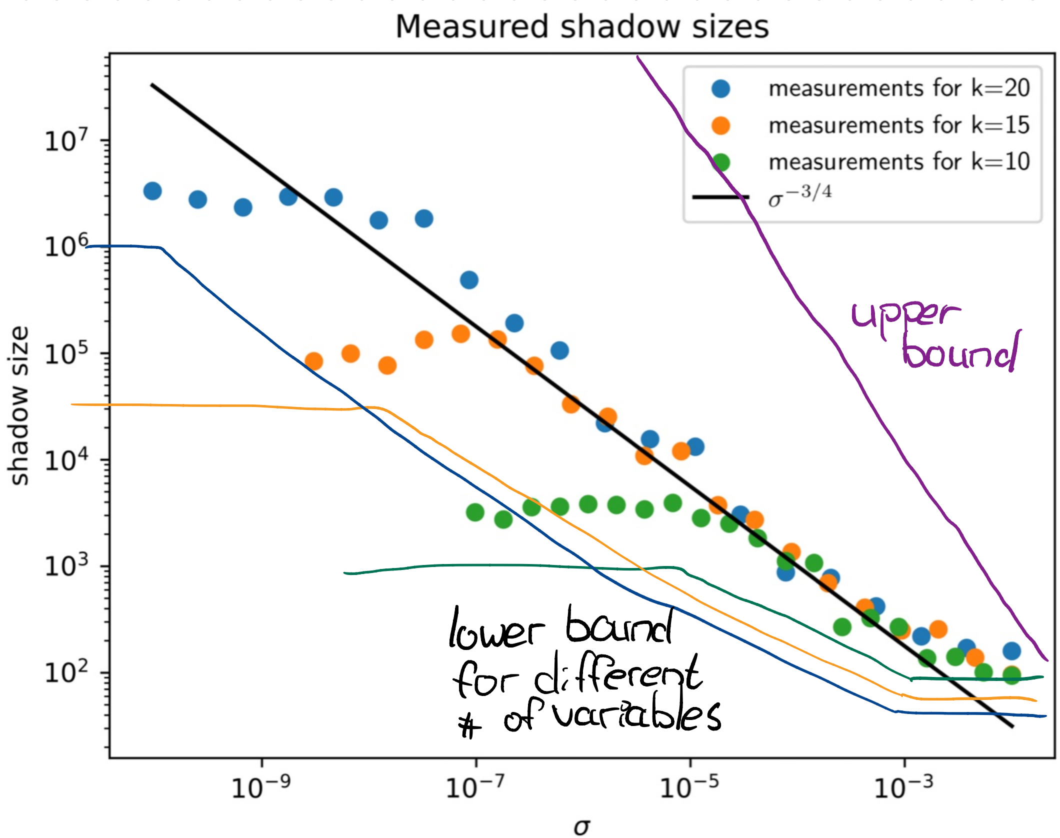
Synthetic data



CAN tell if theorem is tight

CAN NOT tell if theorem is useful

Synthetic data



CAN tell if theorem is tight

CAN NOT tell if theorem is useful

The first linear program (1948)

Given 77 ingredients,
find the cheapest diet
that meets all 9 nutrient needs

The first linear program (1948)

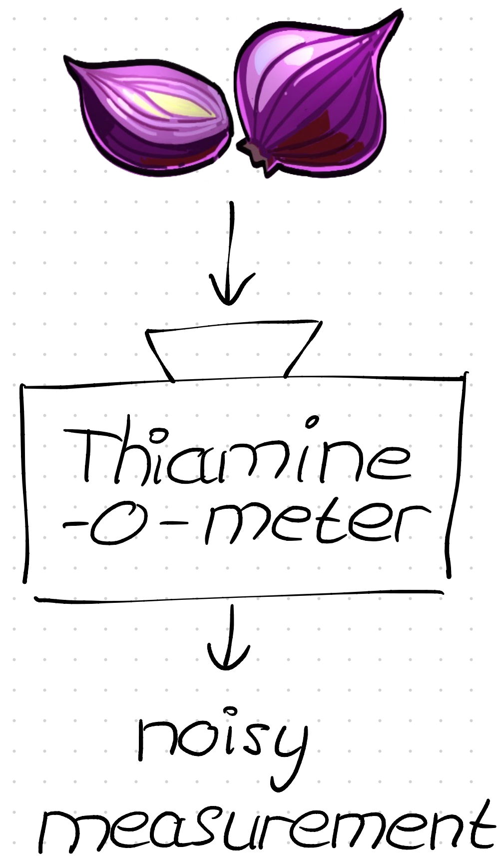
Given 77 ingredients,
find the cheapest diet
that meets all 9 nutrient needs

Side Question:

is this normal to want
and possible to achieve?

The first linear program (1948)

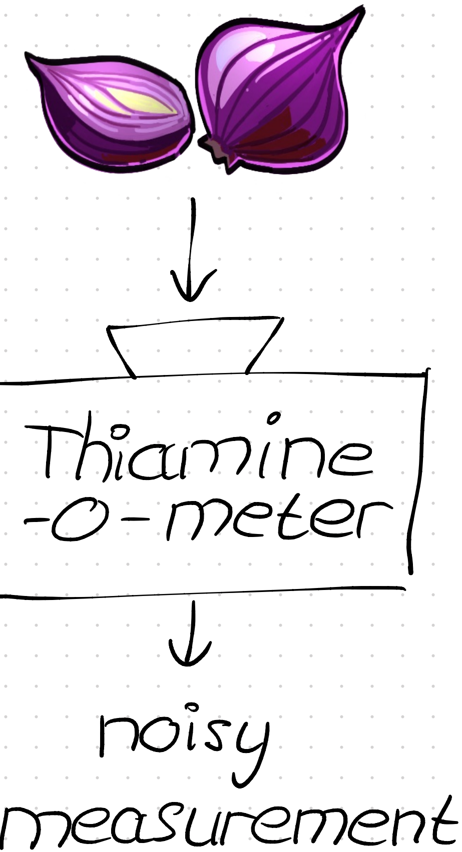
Given 77 ingredients,
find the cheapest diet
that meets all 9 nutrient needs



The first linear program (1948)

Given 77 ingredients,
find the cheapest diet
that meets all 9 nutrient needs

Type of Operations	No. of repetitions
Multiplication	15,315
Division	1,234
Addition of two numbers	14,561
Addition of 77 numbers	190
Addition of 9 numbers	85



The first linear program (1948)

Given 77 ingredients,
find the cheapest diet
that meets all 9 nutrient needs

Type of Operations	No. of repetitions
Multiplication	15,315
Division	1,234
Addition of two numbers	14,561
Addition of 77 numbers	190
Addition of 9 numbers	85

Thiamine
-o-meter

noisy
measurement

Open question: how many
pivot steps is that?

Intermission

History lesson

Mathematical Tables Project 1938 - 1948

450 computers employed

Mathematical Tables Project 1938 - 1948

450 computers employed

To perform these computations with desk machines required 5 computers for 21 days, with 4 hours per day supervision by a mathematician.

Mathematical Tables Project 1938 - 1948

450 computers employed

To perform these computations with desk machines required 5 computers for 21 days, with 4 hours per day supervision by a mathematician.

Mathematical Tables Project 1938 - 1948

450 computers employed

To perform these computations with desk machines required 5 computers for 21 days, with 4 hours per day supervision by a mathematician.

Stopped on Von Neumann's recommendation

Mathematical Tables Project 1938 - 1948

450 computers employed

> To perform these computations with desk machines required 5 computers for 21 days, with 4 hours per day supervision by a mathematician.

Stopped on Von Neumann's recommendation

- Handbook of Mathematical Functions 1964

Mathematical Tables Project 1938 - 1948

450 computers employed

To perform these computations with desk machines required 5 computers for 21 days, with 4 hours per day supervision by a mathematician.

Stopped on Von Neumann's recommendation

- Handbook of Mathematical Functions 1964
- LP's solved by hand until 1952

Linear Programming and Extensions

George B. Dantzig

Dantzig's famous book

STIGLER'S NUTRITION MODEL: AN EXAMPLE OF FORMULATION AND SOLUTION

One of the first applications of the simplex algorithm was to the determination of an adequate diet that was of least cost.[1] In the fall of 1947, J. Laderman of the Mathematical Tables Project of the National Bureau of Standards undertook, as a test of the newly proposed simplex method, the first large-scale computation in this field. It was a system with nine equations in seventy-seven unknowns. Using hand-operated desk calculators, approximately 120 man-days were required to obtain a solution.

The particular problem solved was one which had been studied earlier by G. J. Stigler [1945-1], who had proposed a solution based on the substitution of certain foods by others which gave more nutrition per dollar. He then examined a "handful" of the possible 510 ways to combine the selected foods. He did not claim the solution to be the cheapest but gave good

Linear Programming and Extensions

George B. Dantzig

Dantzig's
famous
book

STIGLER'S NUTRITION MODEL:
AN EXAMPLE OF FORMULATION AND SOLUTION

J. Laderman irst applications of the simplex algorithm was to the deter- adequate diet that was of least cost.[1] In the fall of 1947, 'he Mathematical Tables Project of the National Bureau of Stand... ly proposed simplex method, the first lar... **undertook,** was a system with nine equations in seve... rated desk calculators, approxi-

first large-scale computation in this field.

tion of certain foods by others which gave more nutrition per dollar. He then examined a "handful" of the possible 510 ways to combine the selected foods. He did not claim the solution to be the cheapest but gave good

Mathematical Tables Project 1938 – 1948

450 computers employed

To perform these computations with desk machines required 5 computers for 21 days, with 4 hours per day supervision by a mathematician.

Stopped on Von Neumann's recommendation

- Handbook of Mathematical Functions 1964
- LP's solved by hand until 1952

38. Mathematical Tables Project computers with adding machines

Historical takeaways

38. Mathematical Tables Project computers with adding machines

Human computers played an important role in early comb. opt. history

Historical takeaways

38. Mathematical Tables Project computers with adding machines

Human computers played an important role in early comb. opt. history

Their contributions were made invisible by contemporary white men

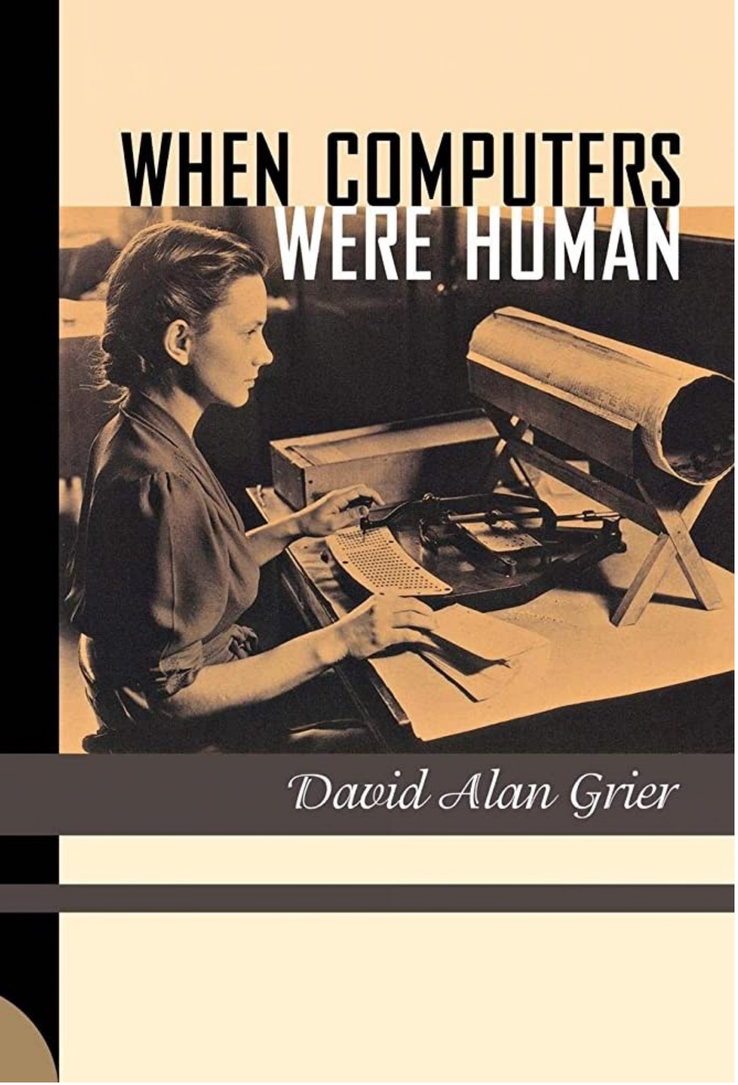
Historical takeaways

38. Mathematical Tables Project computers with adding machines

Human computers played an important role in early comb. opt. history

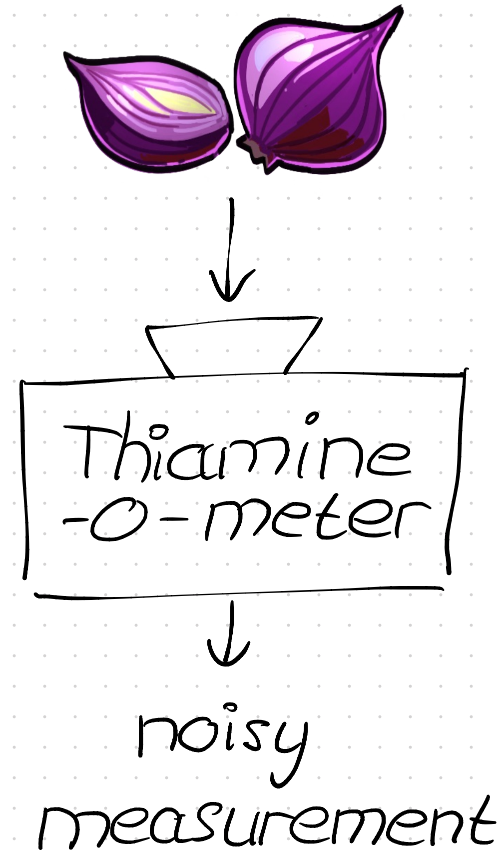
Their contributions were made invisible by contemporary white men

Their demographics are exactly those underrepresented in our field today

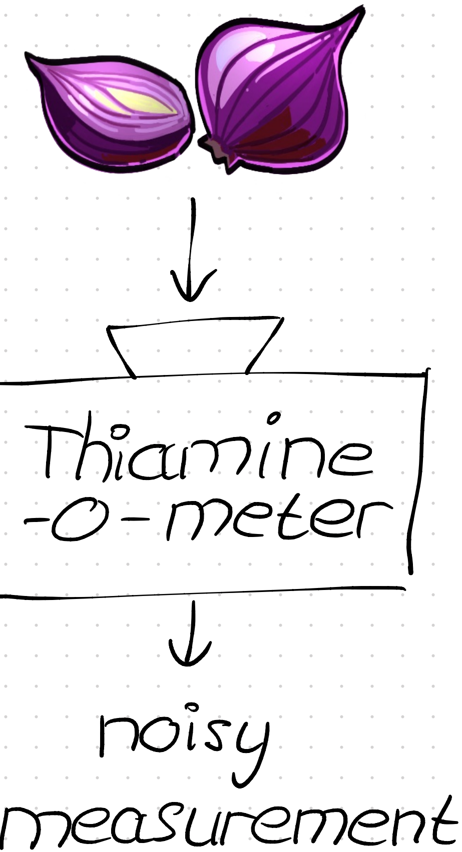


Consider including this history in your lectures

today's source

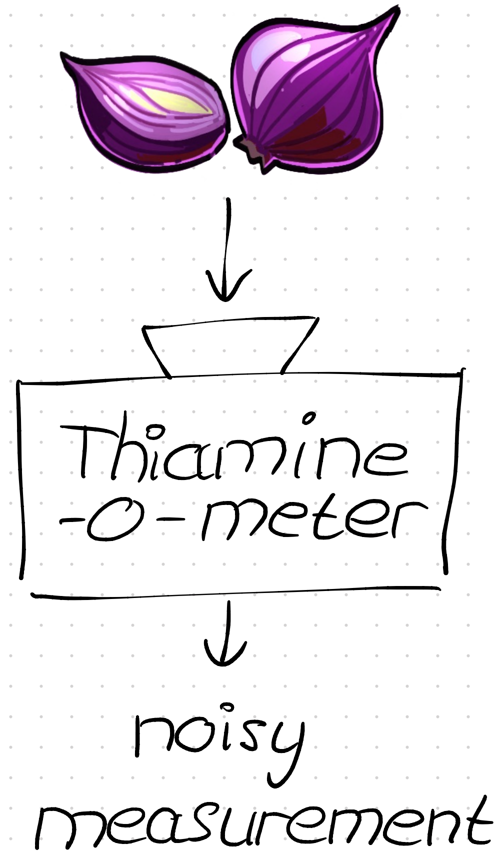

Thiamine
-o-meter
noisy
measurement

Could smoothed analysis have *predicted* that the first computation be fast?



Could smoothed analysis have predicted that the first computation be fast?

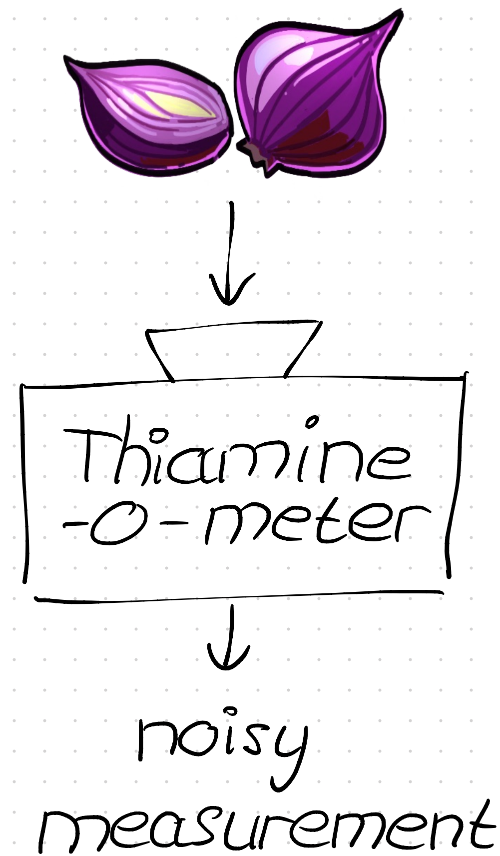
Probably not.



Could smoothed analysis have predicted that the first computation be fast?

Probably not.

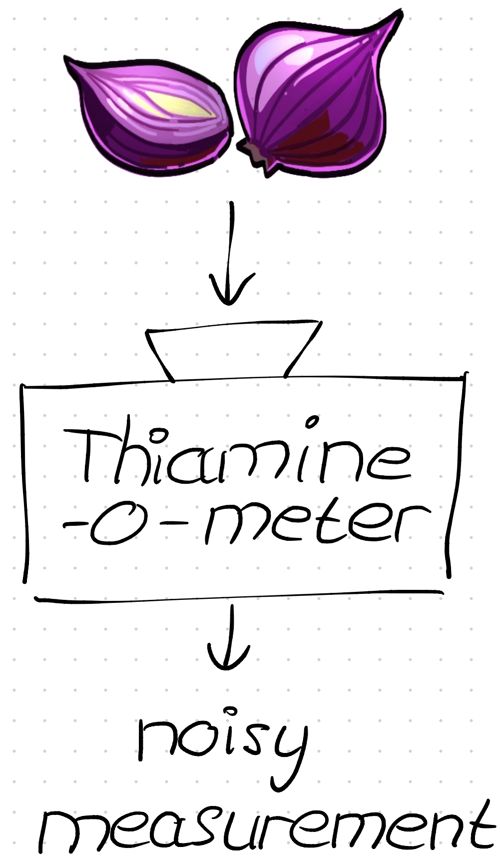
1. different pivot rule



Could smoothed analysis have predicted that the first computation be fast?

Probably not.

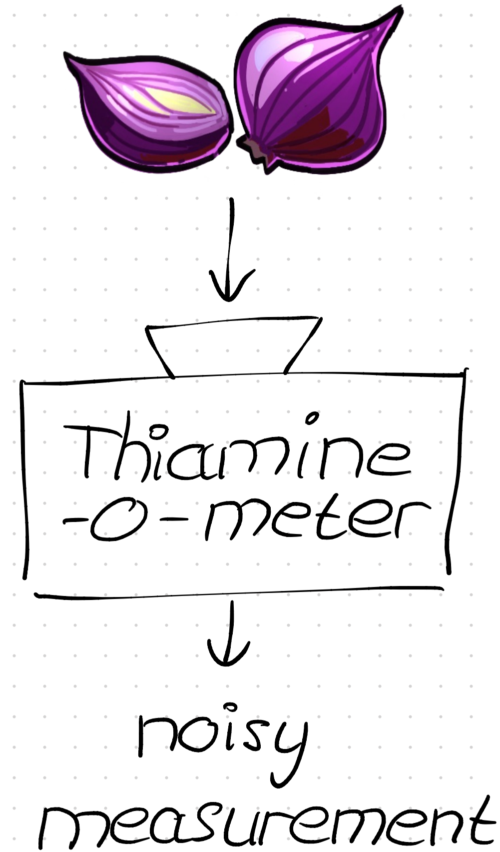
1. different pivot rule
2. different phase 1.



Could smoothed analysis have predicted that the first computation be fast?

Probably not.

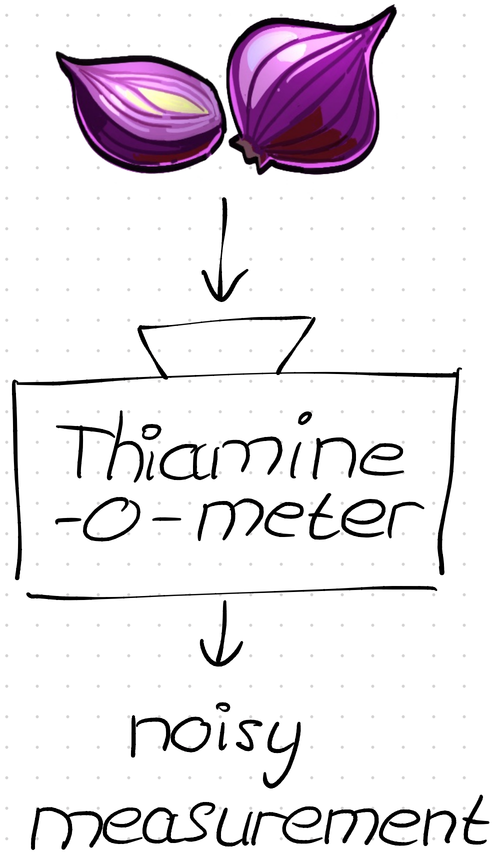
1. different pivot rule
2. different phase 1
3. non-negativity constraints



Could smoothed analysis have predicted that the first computation be fast?

Probably not.

1. different pivot rule
2. different phase 1
3. non-negativity constraints
4. multiplicative error ≥ 15%, but need additive error



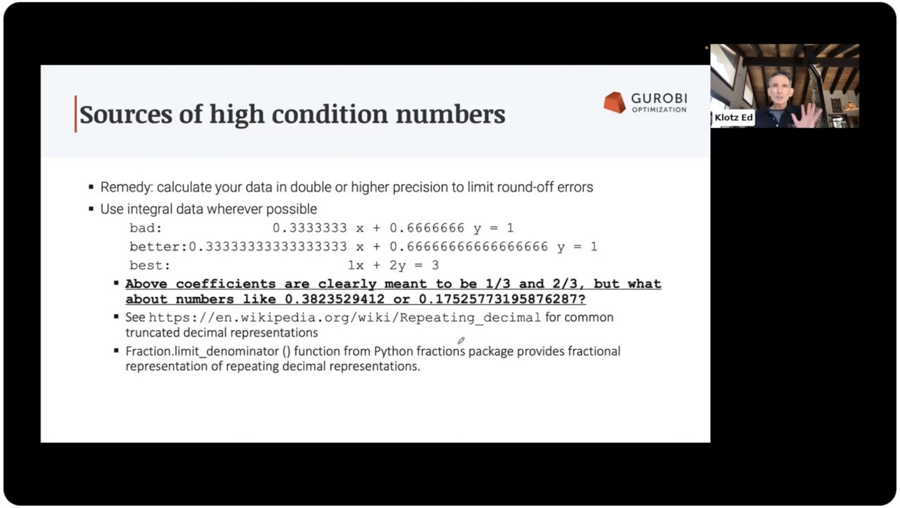
A more fundamental issue?

Smoothed analysis is based on the notion "more noise is better"

A more fundamental issue?

Smoothed analysis is based on the notion
"more noise is better"

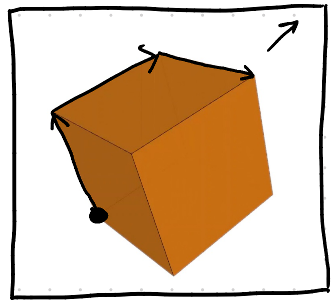
Practitioners say
"less noise is better"



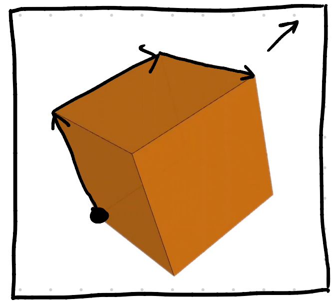
Pinpointing Numerical Issues: Tool for Diagnosis & Explanation of Ill Conditioning in LPs & MILPs

Are we studying the simplex method?

Are we studying the simplex method?

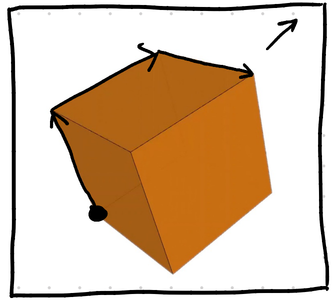
This geometry  doesn't exist.

Are we studying the simplex method?

This geometry  doesn't exist.

What does exist:
- linear algebra
- bound shifting
- bound perturbations
- Harris ratio test

Are we studying the simplex method?

This geometry  doesn't exist.

Paula Harris

What does exist: $\left\{ \begin{array}{l} \text{linear algebra} \\ \text{bound shifting} \\ \text{bound perturbations} \\ \text{Harris ratio test} \end{array} \right.$

Summary:

Our theory are untested
untestable?

Summary:

Our theory are untested
untestable?

We have little/no data

Summary:

Our theory are untested
untestable?

We have little/no data

We disagree with practitioners

Summary:

Our theory are untested

untestable?

We have little/no data

We disagree with practitioners

Are we looking at the wrong algorithm?