Embracing Nonconvexity in Power System Optimization:
Are Local Minima Really So Bad?

Richard Y. Zhang
Univ. of Illinois at Urbana-Champaign
(ryz@illinois.edu)

PSERC Webinar
February 16, 2021

Optimization View of Power System Planning & Operations

In principle: Mathematical optimization

$$\min_{x \in \mathbb{R}^n} \{f(x) : g(x) \leq 0\}$$

In practice: No fast & reliable solution.

Large-scale system

Forces Trade-off

Safety-critical decisions

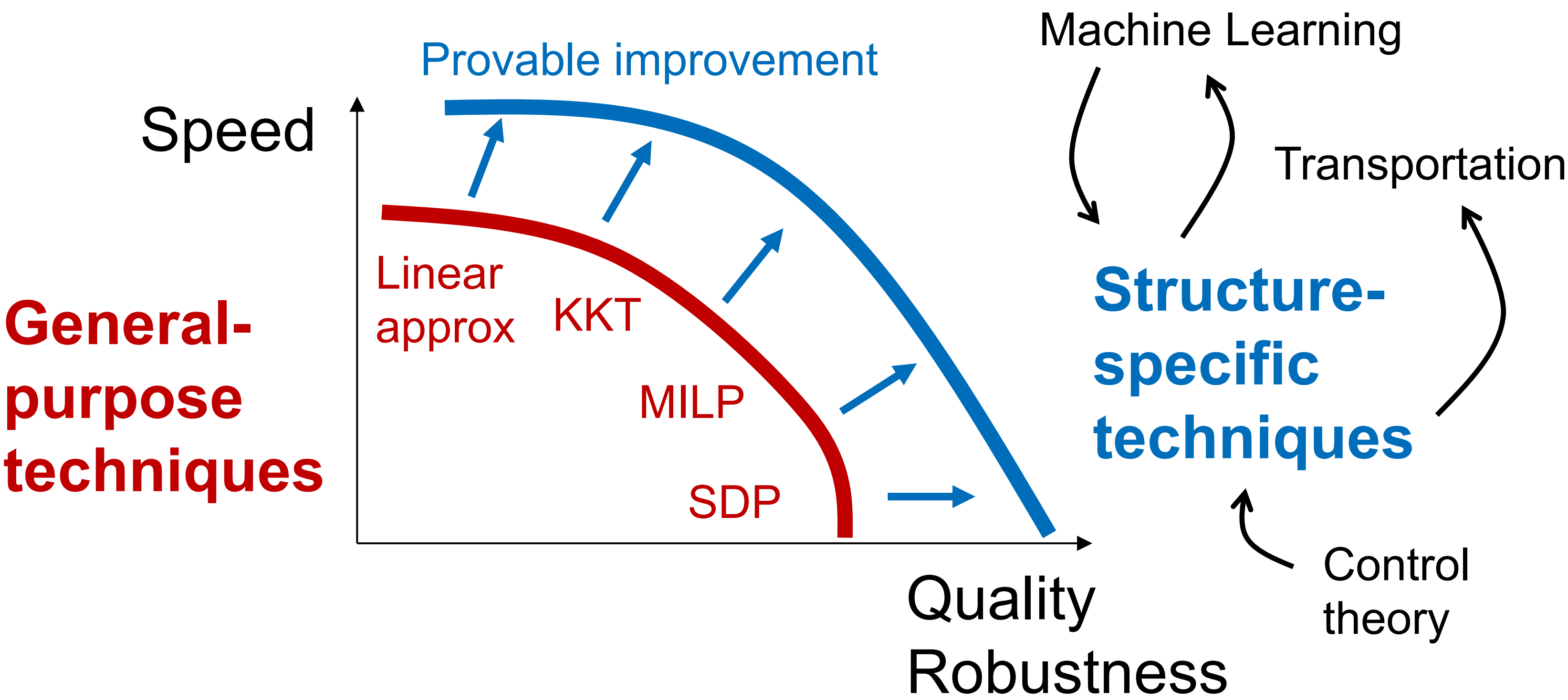
Solution Quality

Robustness

Short span of time

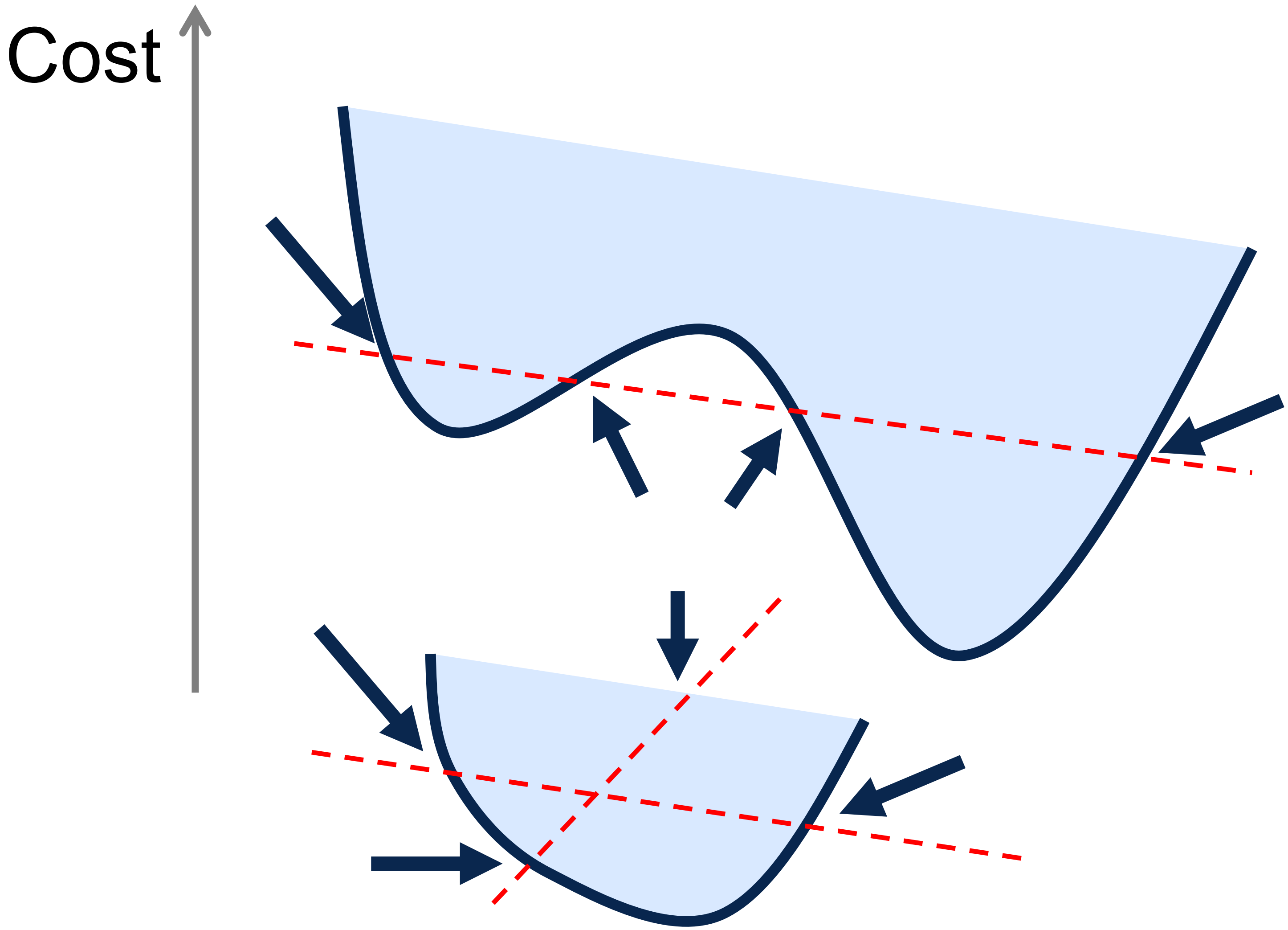
Solution Speed

Our Approach: Structure-Specific Techniques



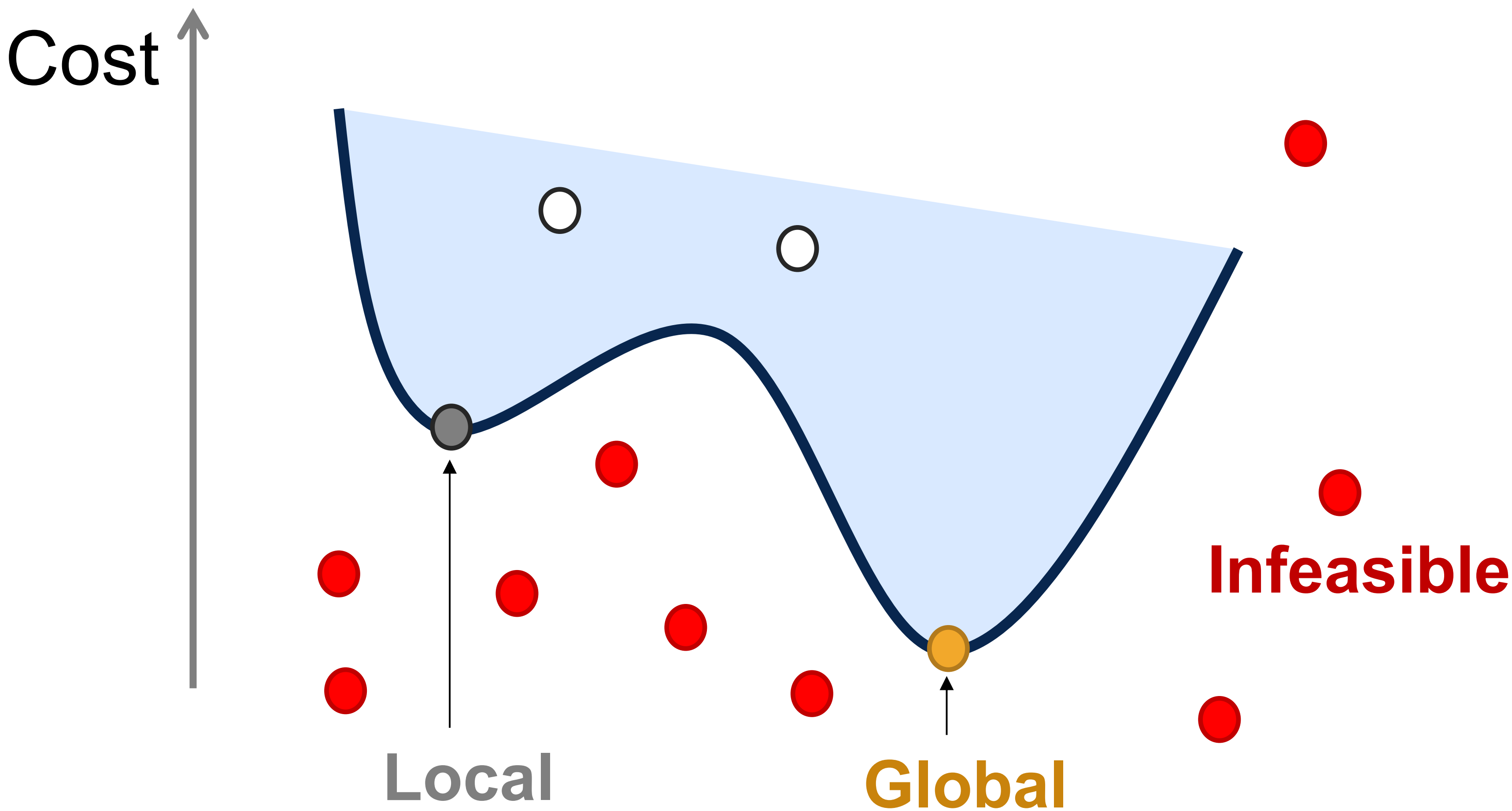
- Absolute improvements by giving up on generality.
- Rigorous notions of structure → Provable guarantees.
- Focus on *structure* for fresh ideas & broader impact.

Fundamental Barrier: Nonconvexity



Convex: Cut boundary at most twice

Getting stuck at local min or infeasibility

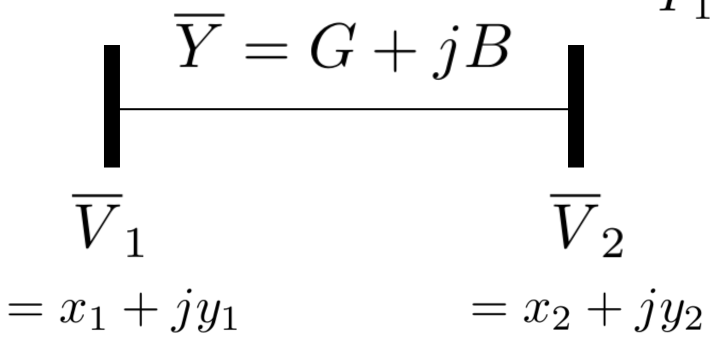


Large-scale algorithms are inherently greedy

All convex problems are alike;
each nonconvex problem
is nonconvex in its own way.

One-size-fits-all is great for convex;
very suspect for nonconvex.

Power Systems: **Quadratic Nonconvexity between Power & Voltage**



$$P_{12} = \mathrm{Re}\{\overline{V}_2 \overline{I}^*_{12}\}, \quad \overline{I}_{12} = \overline{Y}(\overline{V}_1 - \overline{V}_2)$$

$$= \frac{1}{2} \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix}^T \begin{bmatrix} 0 & G & 0 & B \\ G & -2G & -B & 0 \\ 0 & -B & 0 & G \\ B & 0 & G & -2G \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix}$$

$$= v^T A v$$

"Power is proportional to voltage squared"

min gen-cost **s.t** power demand limits

$$\min_{v \in \mathbb{R}^{2n}} \; v^T \underline{A_0} v \quad \text{s.t.} \quad b_k \le v^T \underline{A_k} v \le c_k \quad \text{for all } k$$

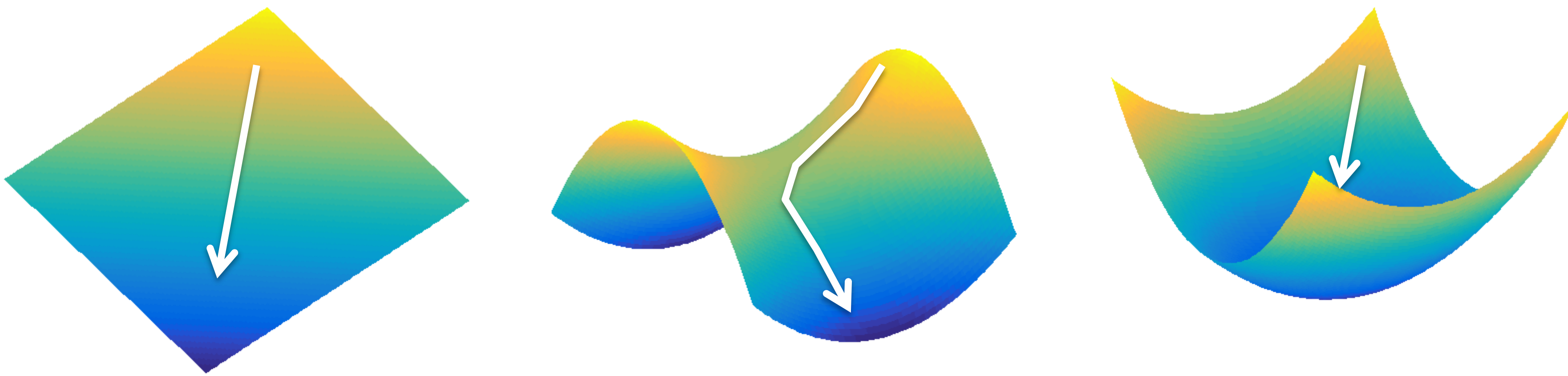
Early 2010s.
Quadratic form
Convex relaxation

Mid-late 2010s.
Specific A_k matrices
Theoretical guarantees

Today: **A Surprising Convex-like Behavior**

$$\min_{v \in \mathbb{R}^{2n}} \quad \sum_k (v^T A_k v - b_k)^2$$

If A_k and b_k satisfy certain properties



Large gradient

Negative curvature

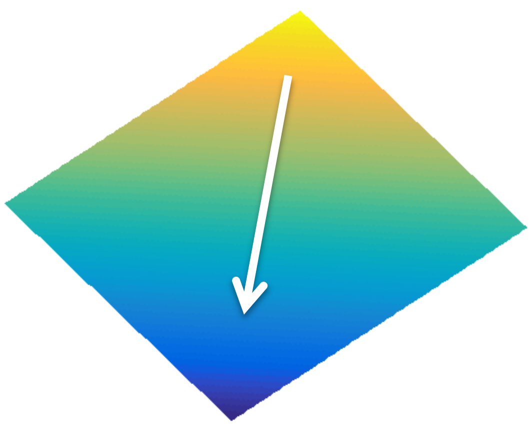
Globally optimal

No spurious local minima guarantees:

- Local optimization yields global optimality
- Nonconvexity is essentially benign

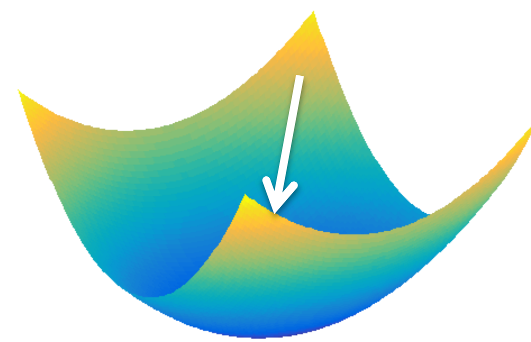
Outline

- Convex relaxation for Optimal Power Flow
- Embracing local minima in State Estimation
- Concluding remarks



Large gradient

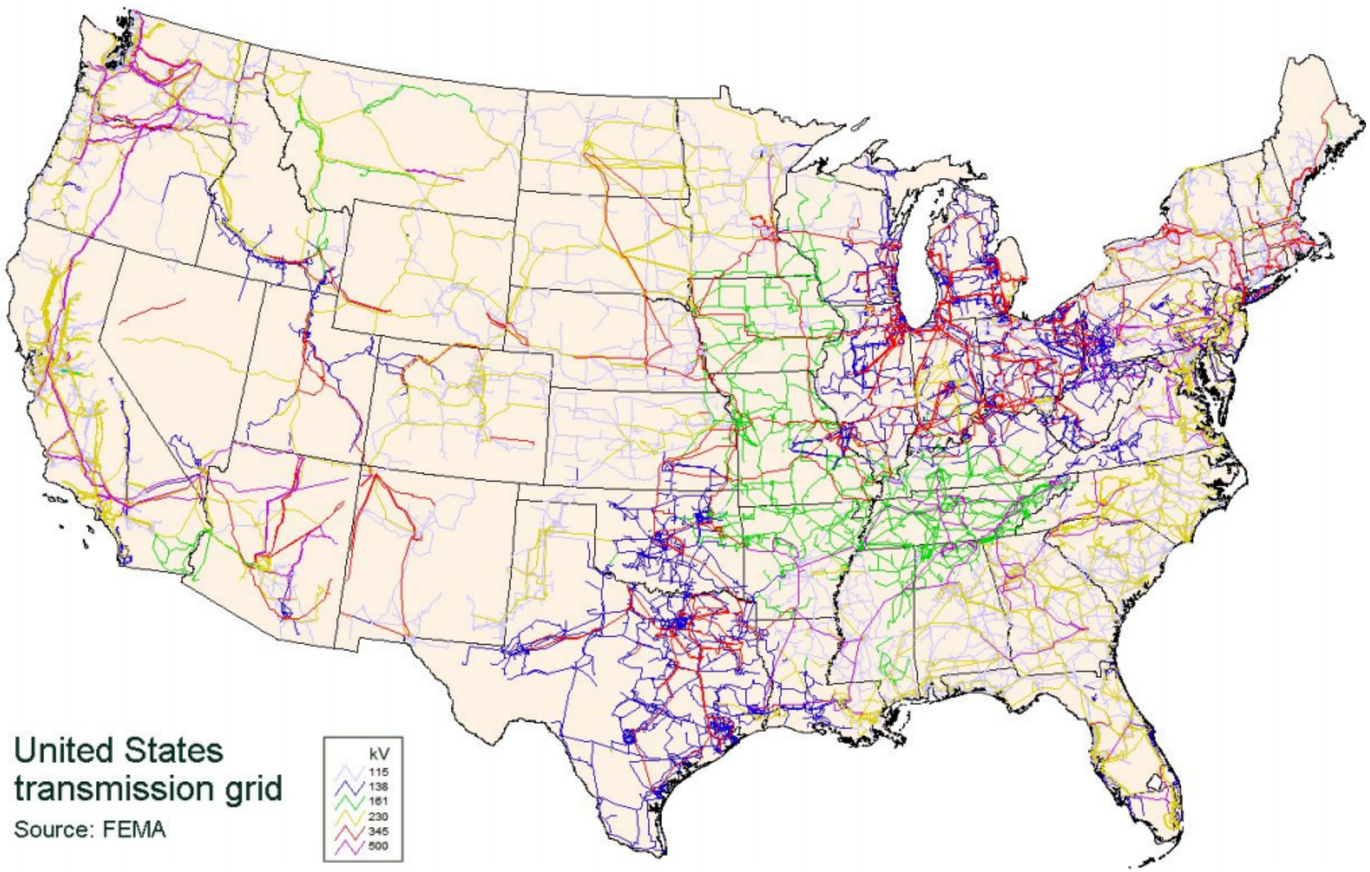
Negative curvature



Globally optimal

Optimal Power Flow on the Electric Grid

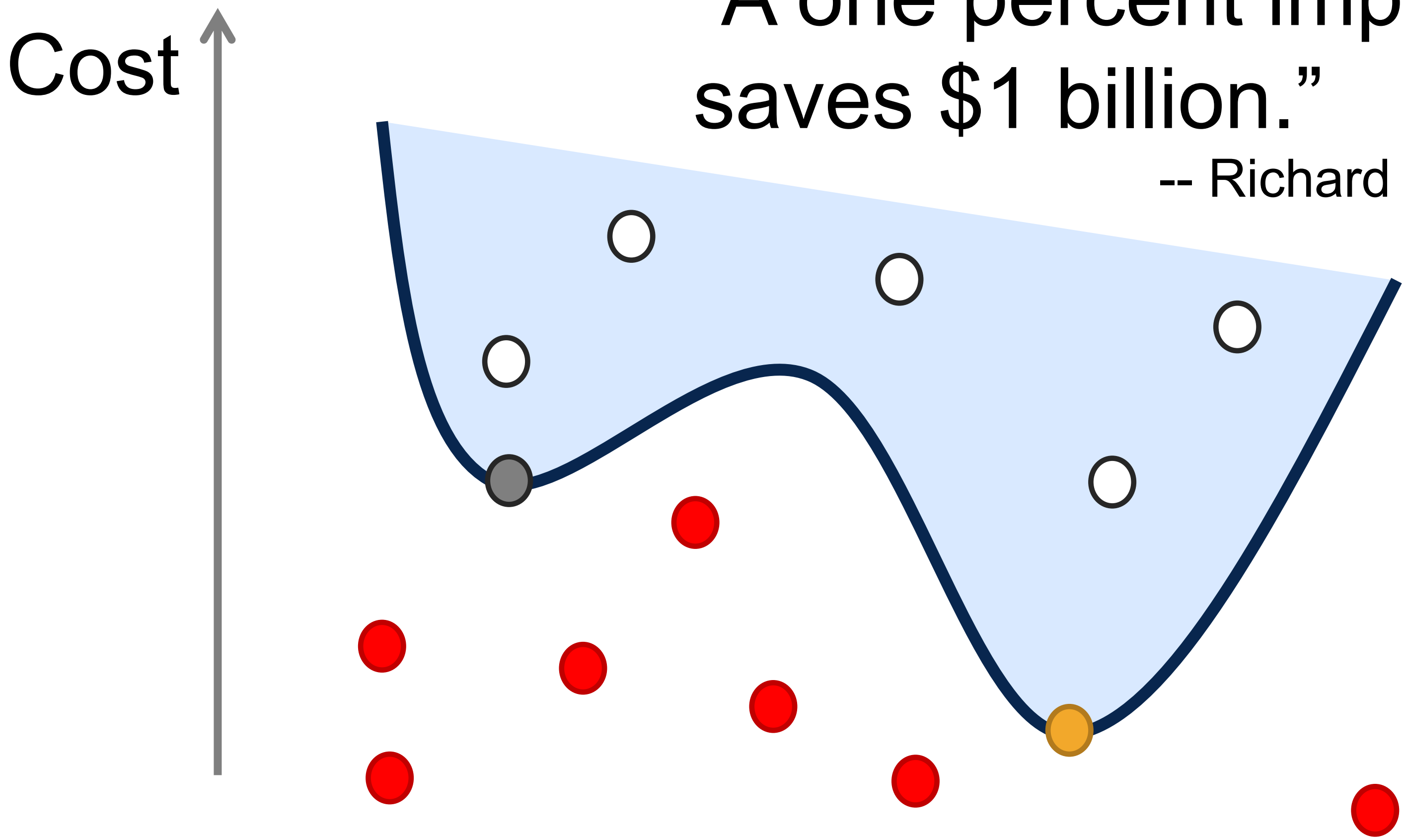
minimize cost of electricity **over** generator dispatch
subject to physics of electricity
reliability & security constraints



Since 1960s: Linearization, Local Optim.

“A one percent improvement saves $1 billion.”

-- Richard O’Neill (FERC)



Elbert, Mittelmann, “Analysis of GO Competition Challenge 1 Final Event Problem Difficulty” (2020)

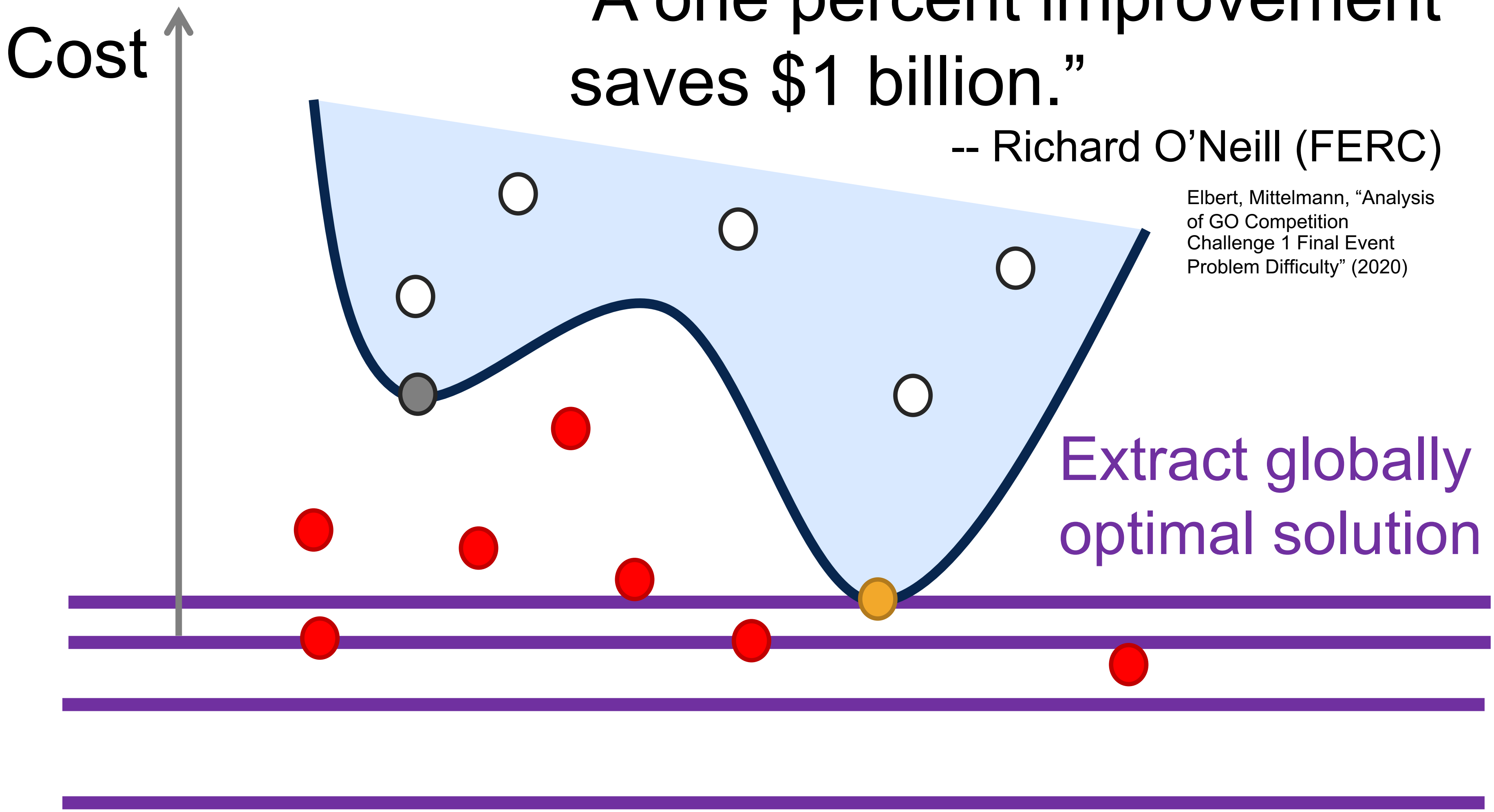
- **Speed:** Very fast and often works very well.
- **Quality & robustness:** No guarantees.
- **Nonconvexity:** Can give very bad or physically impossible solutions.

Early 2010s: Convex relaxation

"A one percent improvement saves $1 billion."

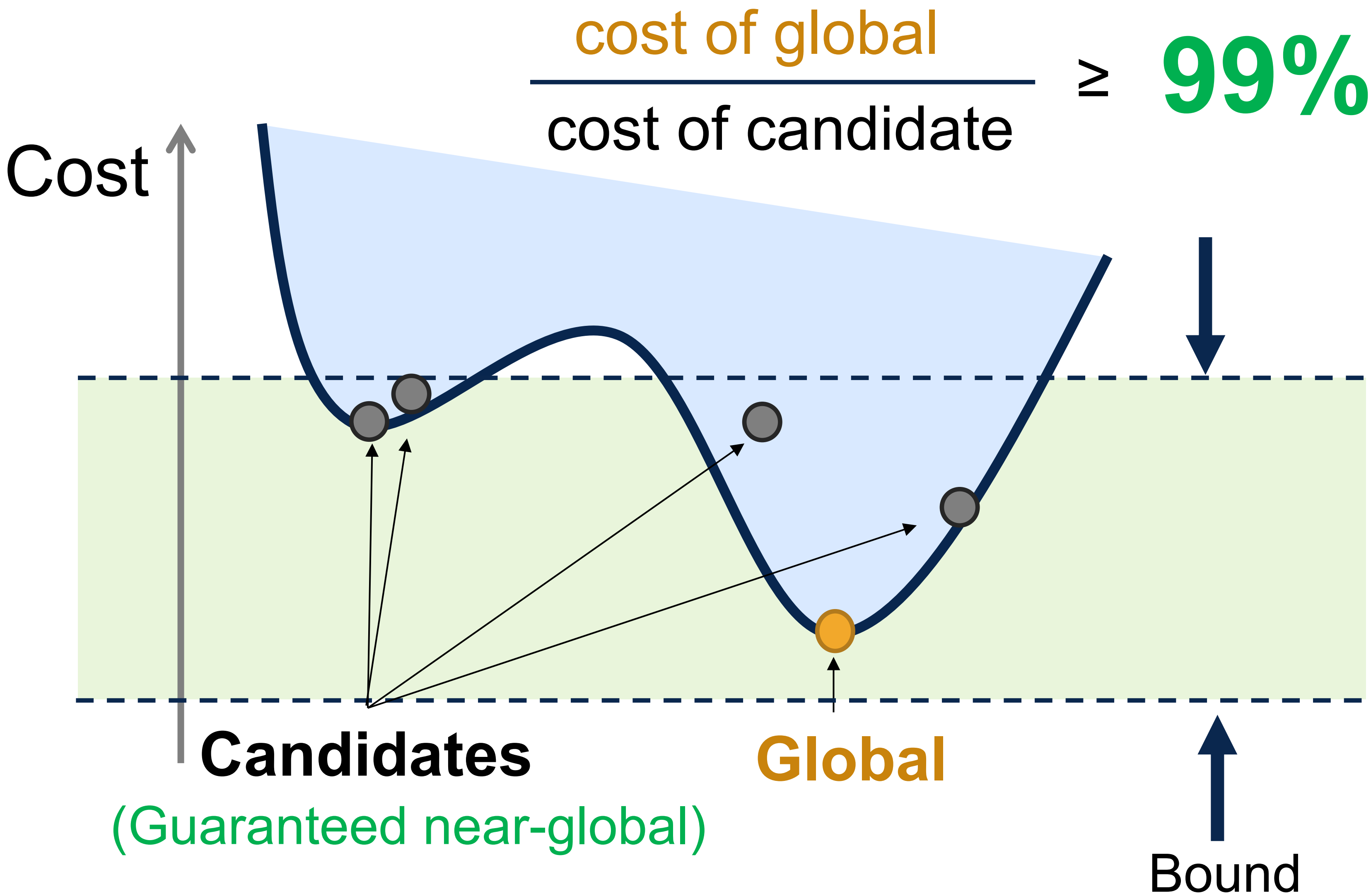
-- Richard O'Neill (FERC)

Elbert, Mittelmann, "Analysis of GO Competition Challenge 1 Final Event Problem Difficulty" (2020)



Convex problem gives provable lower-bound

Combining the two approaches



Brief timeline on convex relaxation

- Issue of nonconvexity (Momoh 1997; Hiskens & Davy 2001)
- SDP relaxation (Jabr 2006; Bai et al. 2008)
- Global guarantees (Lavaei & Low 2012; Sojoudi & Lavaei 2014; Lavaei et al. 2014; Bose et al. 2015; Madani et al. 2017)
- Chordal conversion (Jabr 2012; Molzahn et al. 2013; Bose et al. 2014; Madani et al. 2015)
- Near-global guarantees (Madani et al. 2014)
- Convergent hierarchies (Josz et al. 2014; Josz & Molzahn 2018)

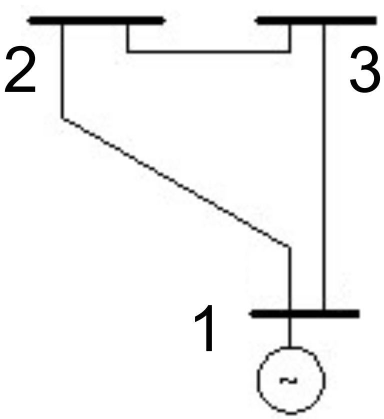
Powerful guarantees on robustness and quality, but very bad worst-case speed

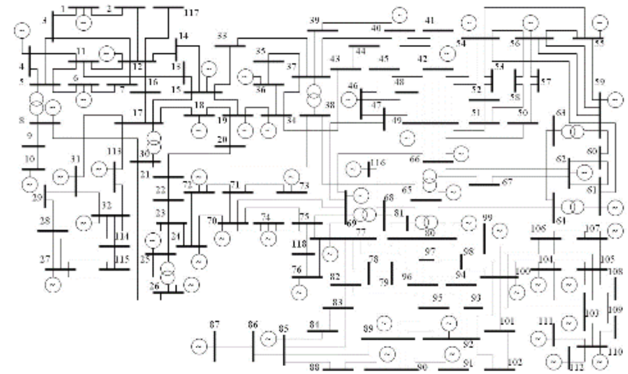
Worst-case Complexity

Time: $O(n^{3.5})$ to $O(n^{6.5})$

Memory: $O(n^2)$ to $O(n^4)$

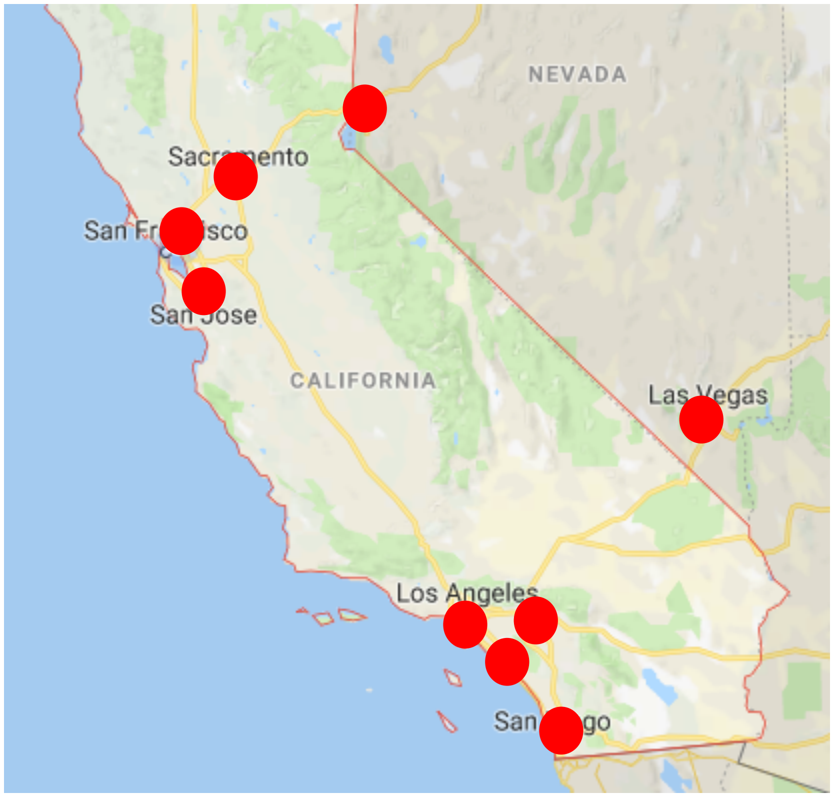
Quadratic Memory & Cubic time

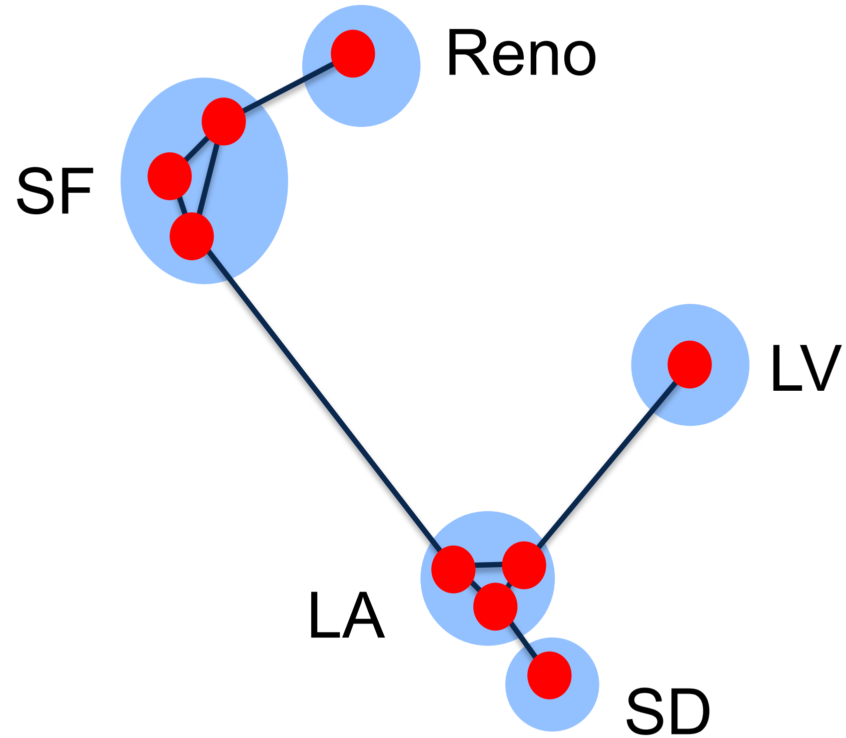




			academic test case		real-world
	$n = 3$	40x →	$n = 120$	25x →	$n = 3{,}000$
RAM n^2	1 MB	1,600x →	1.6 GB	625x →	1 TB
Time n^3	0.1 sec	64,000x →	1.7 hr	15,625x →	3.2 years
	0.001 sec	64,000x →	1 min	15,625x →	11.5 days

Underlying tree-like graph structure:
Decomposition approach

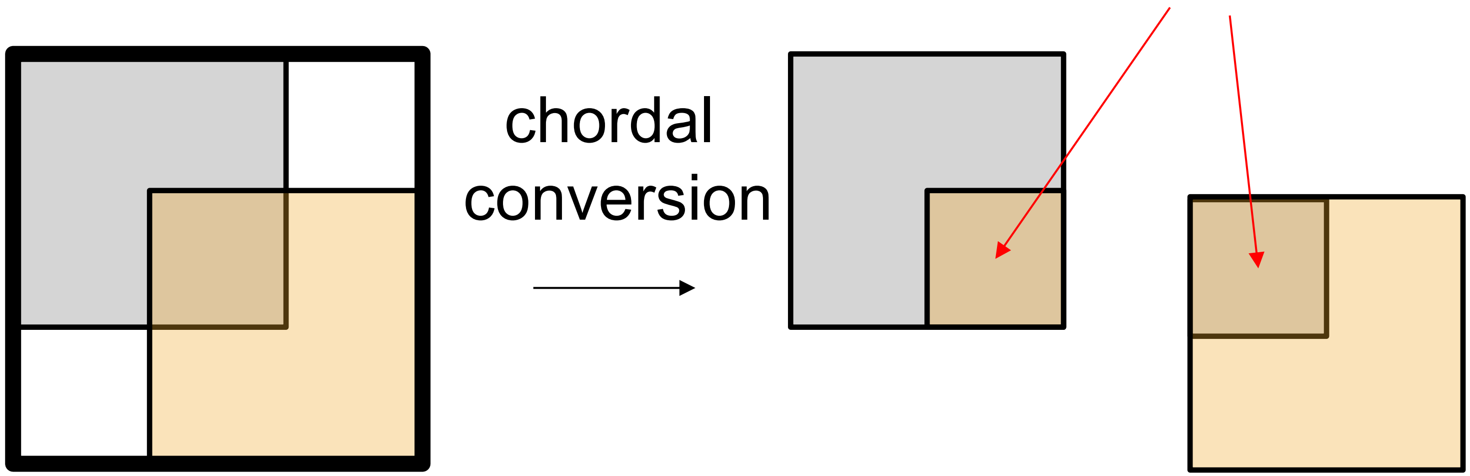




Expansions rare, slow, and expensive. Upgrades faster and cheaper.

Rigorous notions: treewidth; tree decompositions; partial separability.

$(1+\textcolor{blue}{tw})^2$ $\textcolor{red}{n}$ new constraints → Need to enforce overlap constraint



Big matrix variable

$\textcolor{red}{n}^2$ elements

Small matrix variables

$(1+\textcolor{blue}{tw})^2$ $\textcolor{red}{n}$ elements

Partial separability (Griewank & Toit 1982; Sun et al 2014)

Overlap constraints *always* partially separable.

If original constraints partially separable, then $(1+\textcolor{blue}{tw})^6$ $\textcolor{red}{n}$ time per IPM iteration via dynamic prog.

Zhang & Lavaei, *Mathematical Programming* (2020)

Dynamic programming inside an IPM

minimize cost of electricity **over** generator dispatch
subject to physics of electricity
reliability & security constraints

Dualized Clique Tree Conversion with Auxillary Variables

Worst-case		**Empirical**	
Time:	$O(n^{1.5})$	Time:	$O(n)$
Memory:	$O(n)$	Memory:	$O(n)$

Parameterized by how “tree-like” the network is.

Zhang & Lavaei, *Mathematical Programming* (2020)

In Practice: European Power System Model

Parameters

n = 13,659

Security Constraints

m = 40,975

Source: Cédric Josz, Stéphane Fliscounakis, Jean Maeght, Patrick Panciatici (2016)

> 99% globally optimal

< 3 minutes on a laptop.

Zhang & Lavaei, *Mathematical Programming* (2020)

Local Optimization in ARPA-e GO

Parameters

n = 30,000

Security Constraints

m = 300,000,000

arpa·e SM

GRID OPTIMIZATION (GO)
COMPETITION

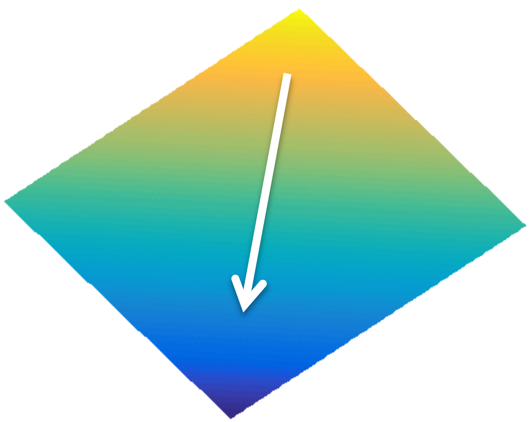
> 99.7% (??) globally optimal < 10 minutes

- **All top scorers used local optimization.**
- "Only provide locally optimality, but seems to be very near globally optimal in practice"
- Similar trends in ML, X-ray imaging.

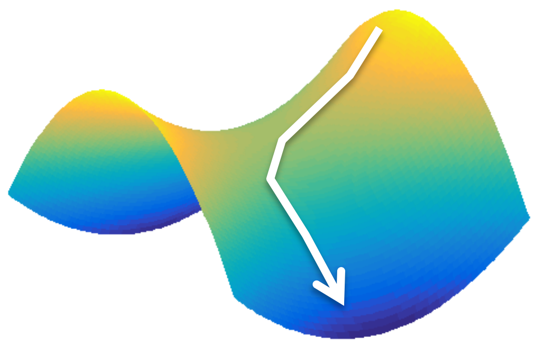
Elbert, Mittelmann, "Analysis of GO Competition Challenge 1 Final Event Problem Difficulty" (2020)
Coffrin, "ARPA-e Grid Competition: SCOPF Overview" (2019)

Outline

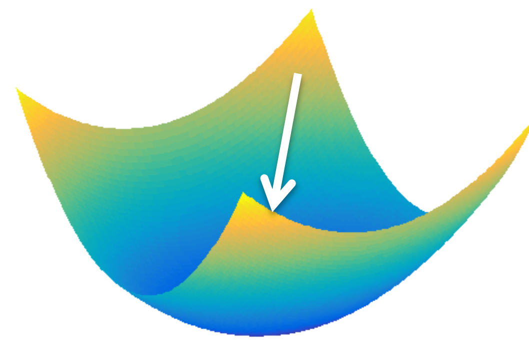
- ~~Convex relaxation for Optimal Power Flow~~
- Embracing local minima in State Estimation
- Concluding remarks



Large gradient

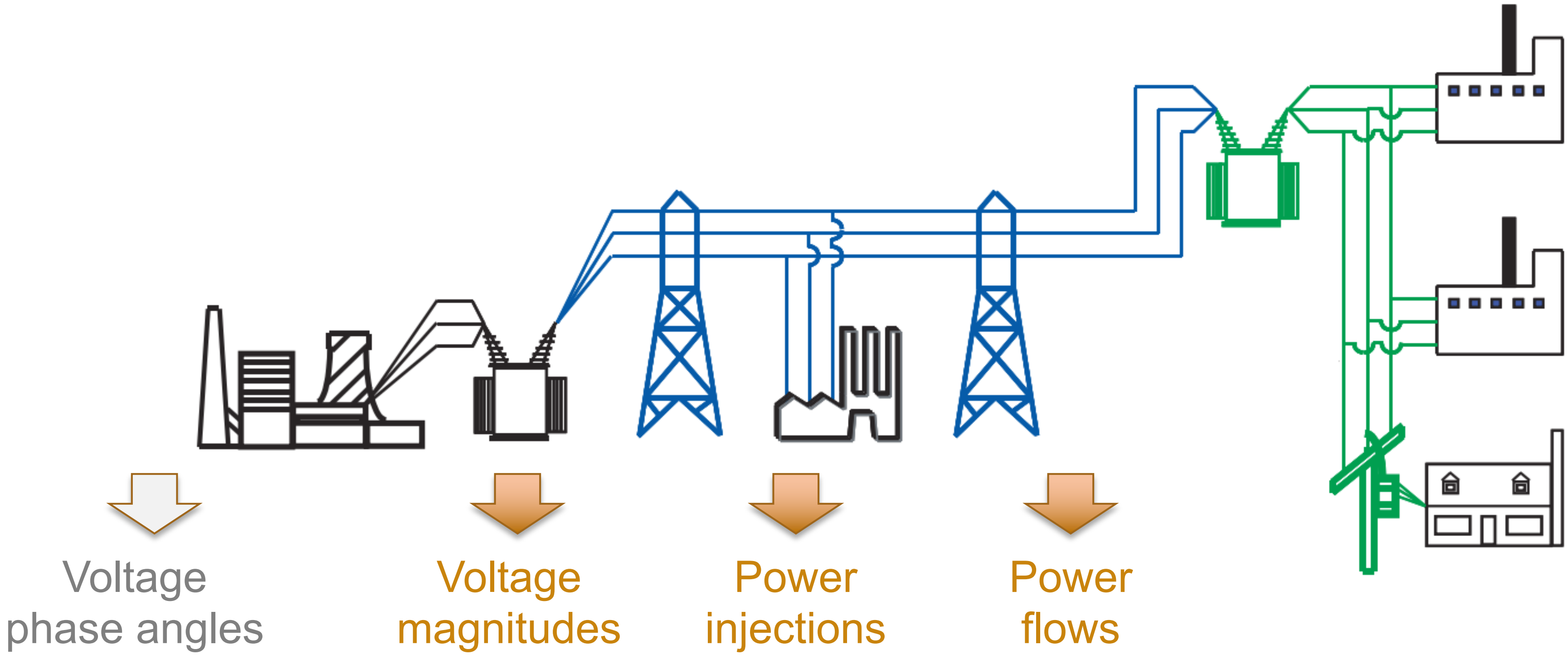


Negative curvature



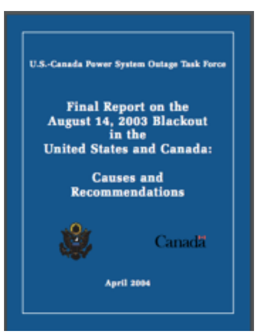
Globally optimal

Power System State Estimation



Estimate all voltage phasors
using incomplete voltage magnitudes
and power measurements.

The August 14th, 2003 Northeast Blackout


U.S.-Canada Power System Outage Task Force

Final Report on the
August 14, 2003 Blackout
in the
United States and Canada:

Causes and
Recommendations

Canada

April 2004

"MISO's state estimator was faulty for most of the period between 12:15 & 15:34 EDT."

--US-Canada Power System Outage Task Force (2004)

Want <u>guarantees</u> on state estimation

Why study state estimation?

Optimal power flow

min gen-cost **s.t** power demand limits

$$\min_{v \in \mathbb{R}^{2n}} \quad v^T A_0 v \quad \text{s.t.} \quad b_k \leq v^T A_k v \leq c_k \quad \text{for all } k$$

State estimation

$$\min_{v \in \mathbb{R}^{2n}} \quad 0 \quad \text{s.t.} \quad b_k = v^T A_k v \quad \text{for all } k$$

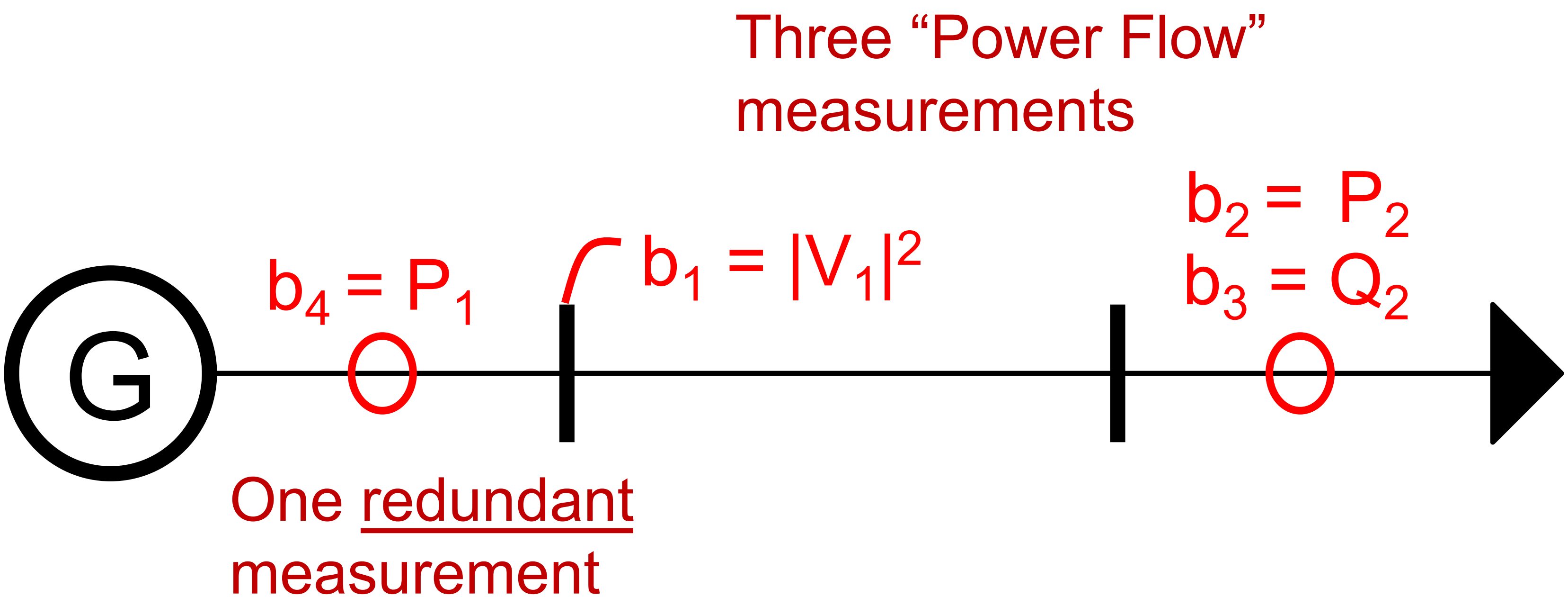
$$\min_{v \in \mathbb{R}^{2n}} \quad \sum_k (v^T A_k v - b_k)^2$$

1. No constraints

2. Same math problem in ML / Imaging

3. State estimation successfully solved using local optimization since 1970s

Risk of plausible but fictitious estimations: A Two-Bus Example

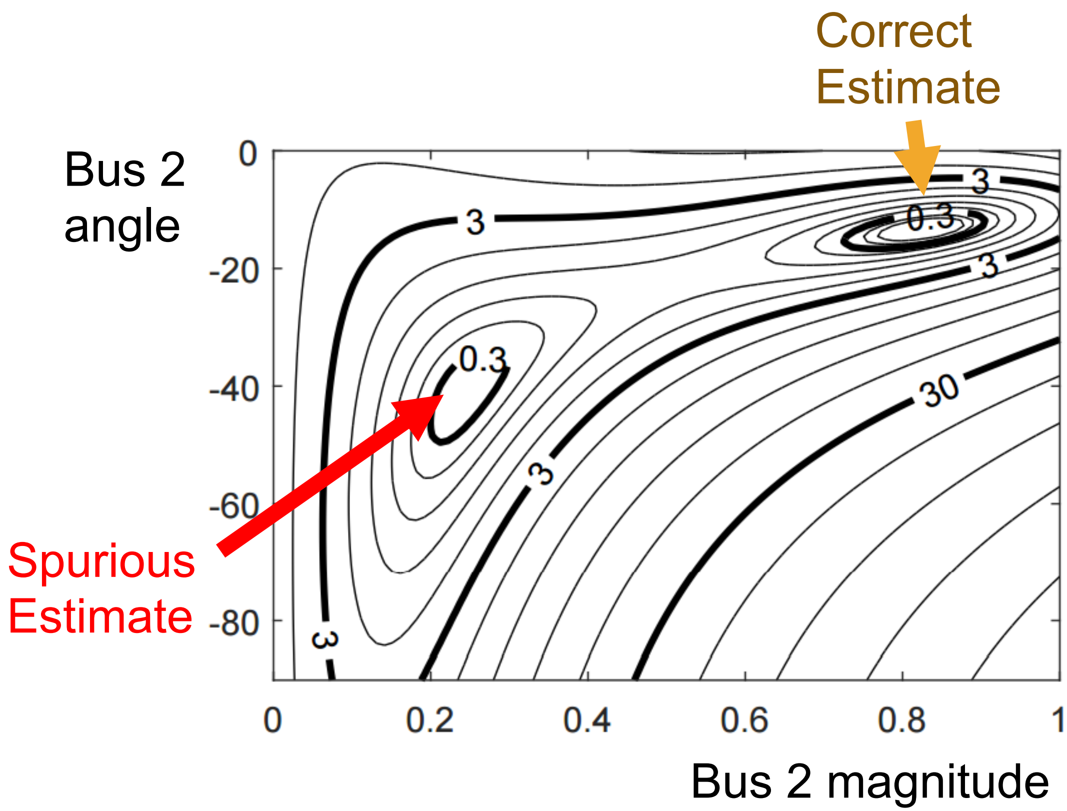


$$\overline{V}_1 = 1, \quad \overline{V}_2 = 0.928\angle - 13.2^\circ, \quad \overline{Y} = \frac{1}{0.01 + j0.1}$$

Zhang, Lavaei, Baldick, *IEEE Trans. Control of Network Systems* (2019)

Existence of bad local minima

$$\min_{v \in \mathbb{R}^{2n}} \quad \sum_k (v^T A_k v - b_k)^2$$



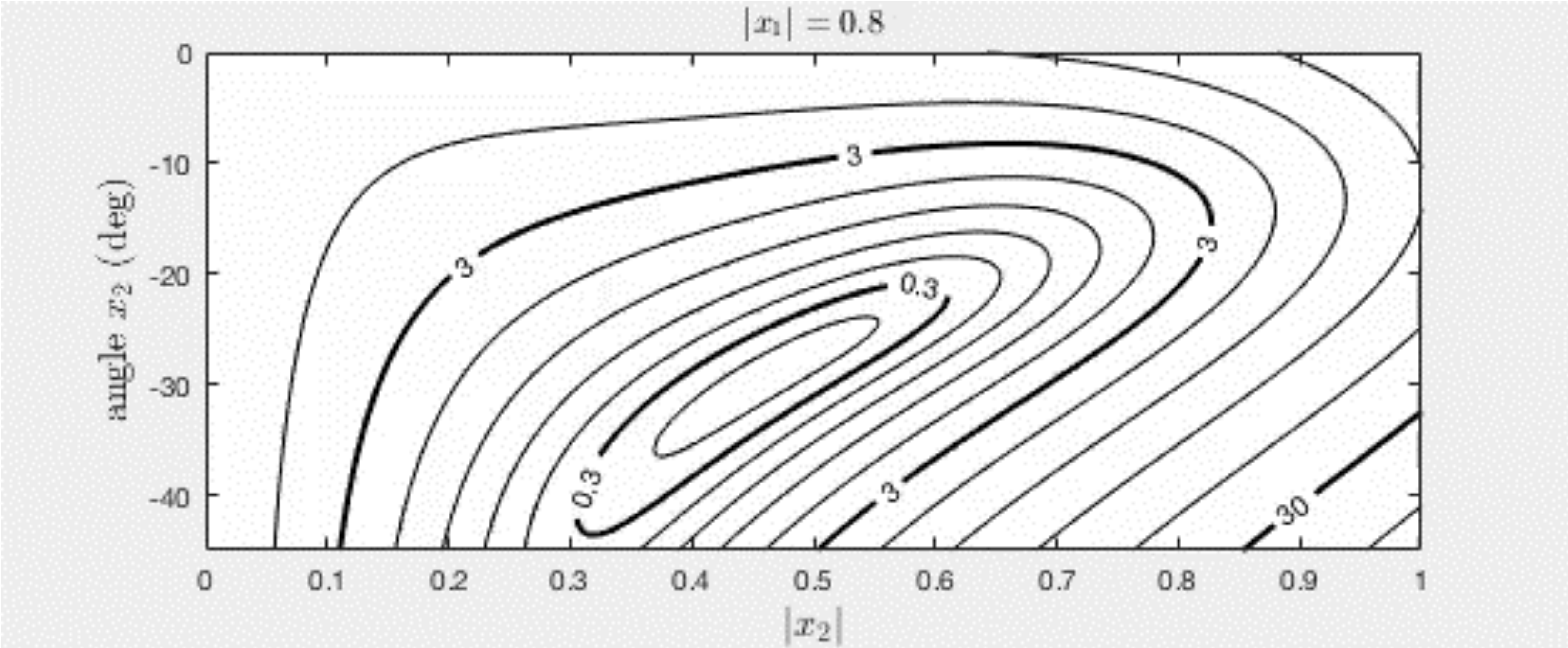
Zhang, Lavaei, Baldick, *IEEE Trans. Control of Network Systems* (2019)

Two plausible but very different estimates

Indeed, we find four critical points, only one of which is the correct estimate

$$\begin{bmatrix} |x_1| \\ |x_2| \\ \measuredangle x_2 \end{bmatrix} \in \left\{ \underbrace{\begin{bmatrix} 1 \\ 0.829 \\ -13.2^\circ \end{bmatrix}}_{\text{Global minim.}}, \underbrace{\begin{bmatrix} 0.870 \\ 0.345 \\ -35.7^\circ \end{bmatrix}}_{\text{Local minim.}}, \underbrace{\begin{bmatrix} 0.846 \\ 0.401 \\ -32.0^\circ \end{bmatrix}}_{\text{Saddle point}}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{\text{Local maxim.}} \right\}$$

Both estimates are plausible!



Zhang, Lavaei, Baldick, *IEEE Trans. Control of Network Systems* (2019)

Confusion with bad data or error

Indeed, we find four critical points, only one of which is the correct estimate

$$\begin{bmatrix} |x_1| \\ |x_2| \\ \measuredangle x_2 \end{bmatrix} \in \left\{ \boxed{\begin{bmatrix} 1 \\ 0.829 \\ -13.2^\circ \end{bmatrix}}, \boxed{\begin{bmatrix} 0.870 \\ 0.345 \\ -35.7^\circ \end{bmatrix}}, \begin{bmatrix} 0.846 \\ 0.401 \\ -32.0^\circ \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$$

Global minim. | Local minim. | Saddle point | Local maxim.

How well do the two estimates match our measurements?

Correct Estimate

$$\begin{bmatrix} 1 \\ 0.829 \\ -13.2^\circ \end{bmatrix} \rightarrow \begin{aligned} F_1(z) - b_1 &= 0 \\ F_2(z) - b_2 &= 0 \\ F_3(z) - b_3 &= 0 \\ F_4(z) - b_4 &= 0 \end{aligned}$$

Perfect match.

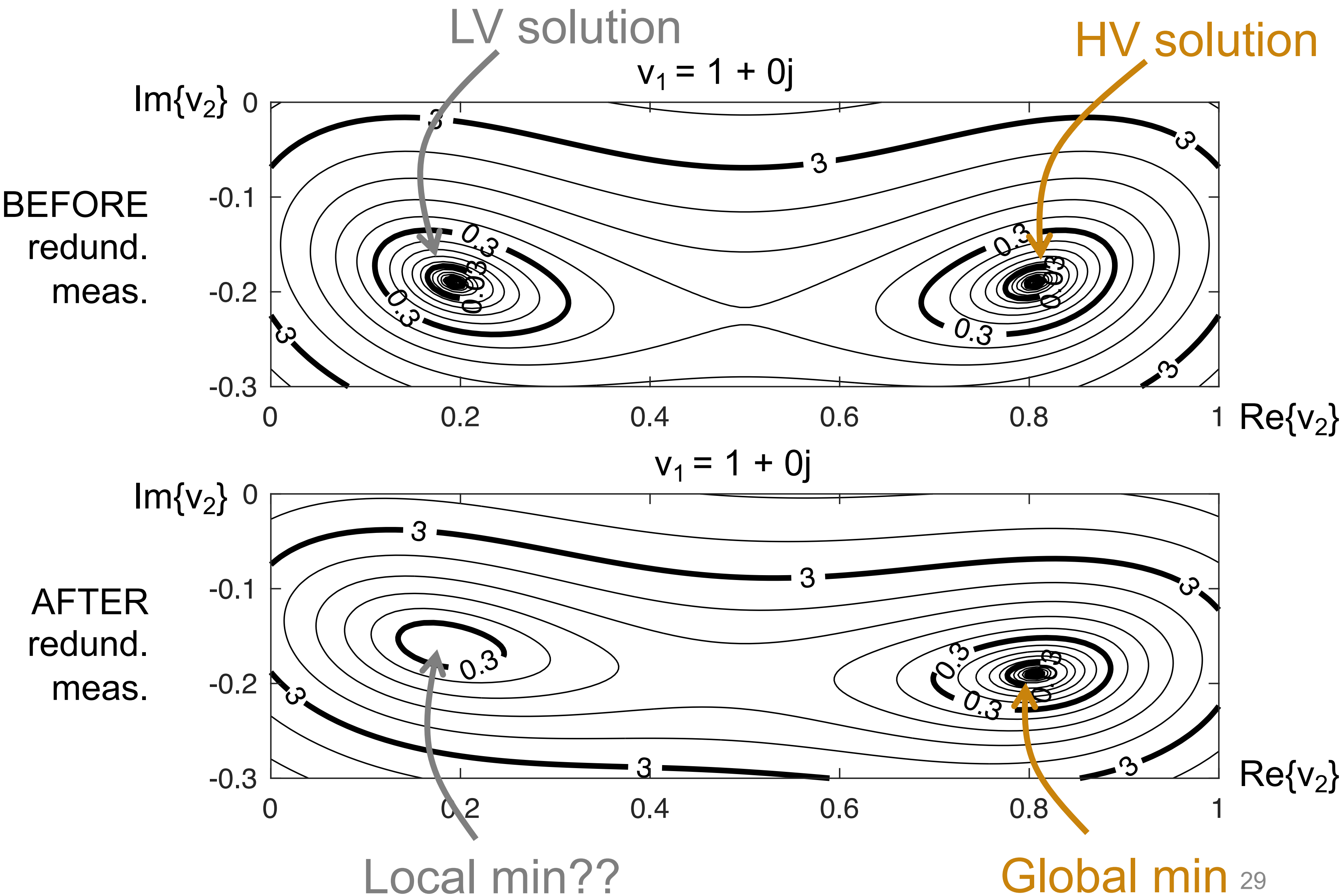
Spurious Estimate

$$\begin{bmatrix} 0.870 \\ 0.345 \\ -35.7^\circ \end{bmatrix} \rightarrow \begin{aligned} F_1(x) - b_1 &= -0.24 \text{ p.u.} \\ F_2(x) - b_2 &= -0.14 \text{ p.u.} \\ F_3(x) - b_3 &= +0.06 \text{ p.u.} \\ F_4(x) - b_4 &= -0.17 \text{ p.u.} \end{aligned}$$

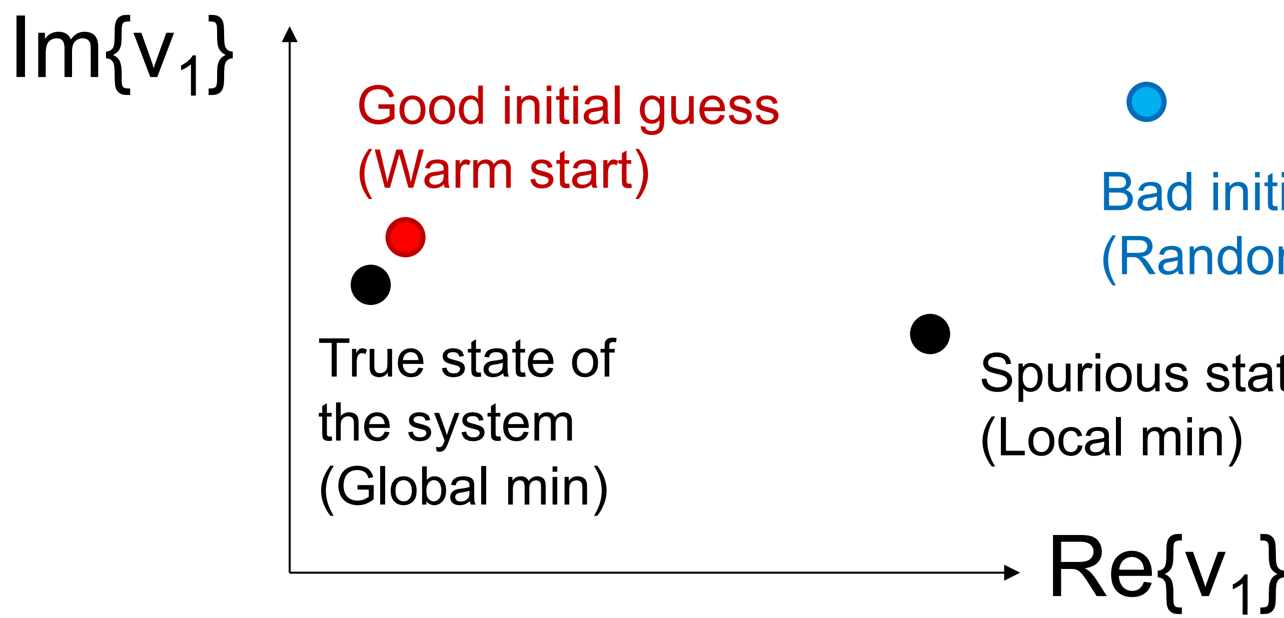
Measurement error? Bad data?

Zhang, Lavaei, Baldick, *IEEE Trans. Control of Network Systems* (2019)

Intuition: **HV / LV Solutions to Power Flow**

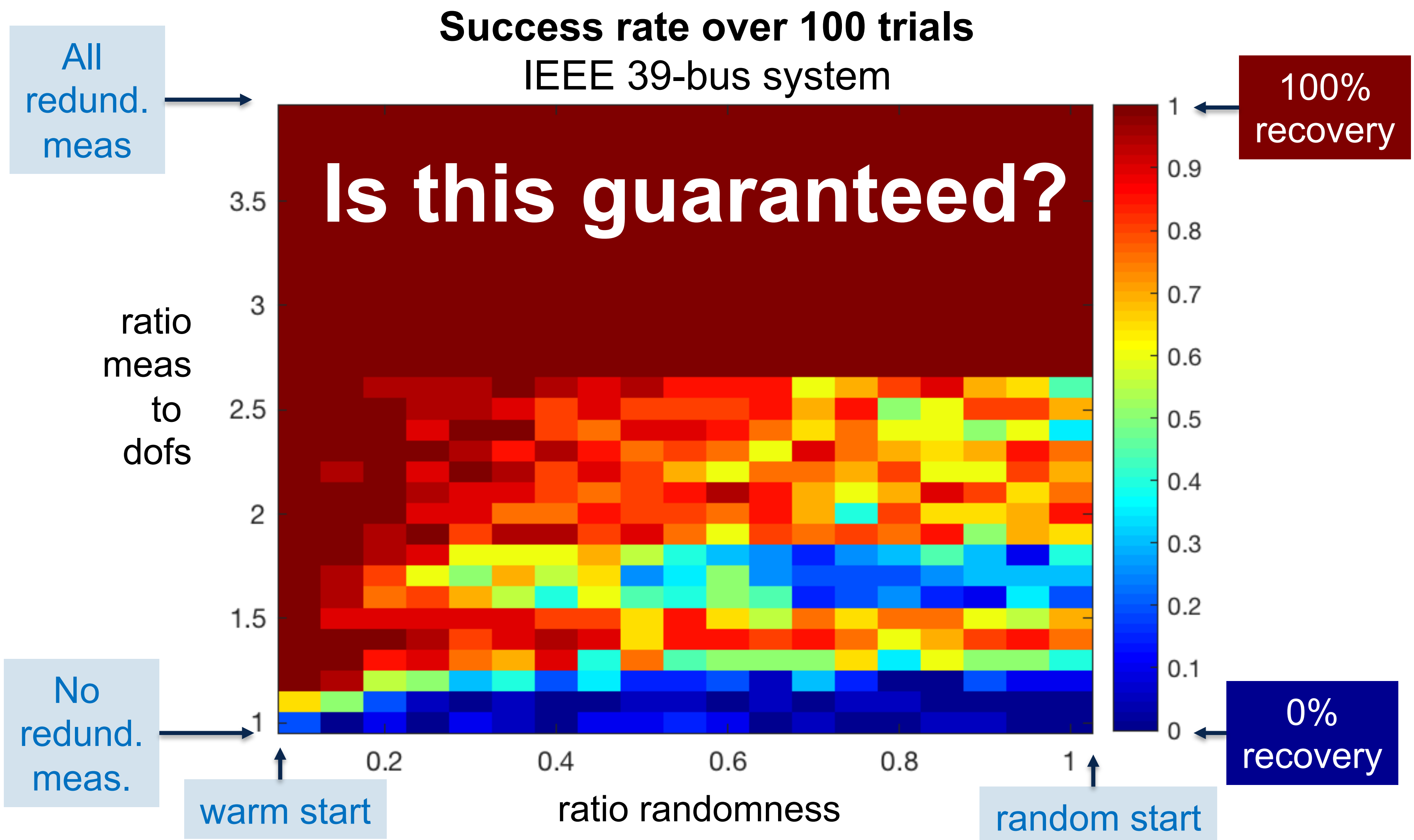


Convergence to bad local minima
Varying quality of the initial guess



Expect a bad initial guess to get stuck at a spurious local min

Experiment results: Bad local minima eliminated by redundant measurements



Zhang, Lavaei, Baldick, *IEEE Trans. Control of Network Systems* (2019)

Prior results – No Spurious Local Minima

Bhojanapalli, Neyshabur, Srebro, *NeurIPS* (2016)
Ge, Lee, Ma, *NeurIPS* (2016)

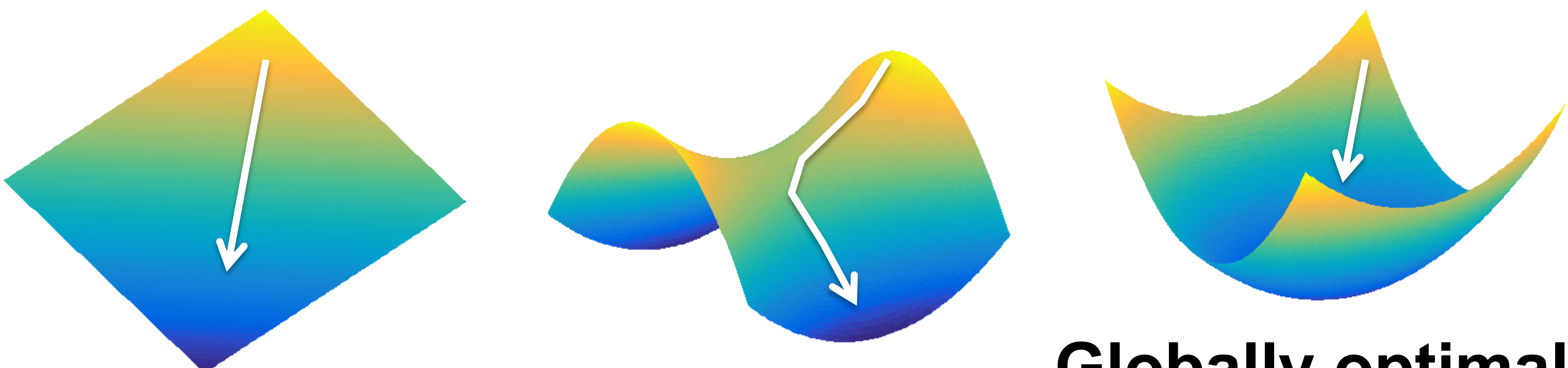
State estimation just the rank r = 1 case

$$\text{minimize} \sum_{i=1}^{m} \left(\text{trace}(U^T A_i U) - b_i\right)^2 \text{ over } U \in \mathbb{R}^{n \times r}$$

"Restricted Isometry Property"

If $A_1, \ldots, A_m$ satisfy δ-RIP (Recht, Fazel, Parrilo 2010)

and $\delta < 1/5$, then $\delta \approx 0.99$ in SE !!

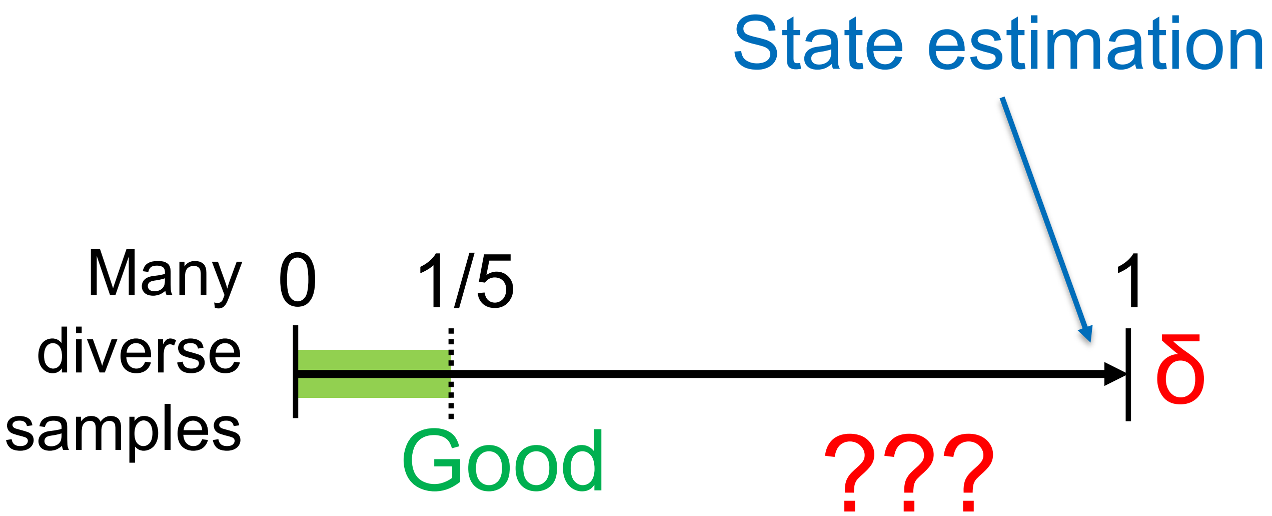


Globally optimal

Can the 1/5 be improved?

(Bhojanapalli et al. 2016) (Ge et al. 2017) (Li & Tang 2017) (Zhu et al. 2017) etc.

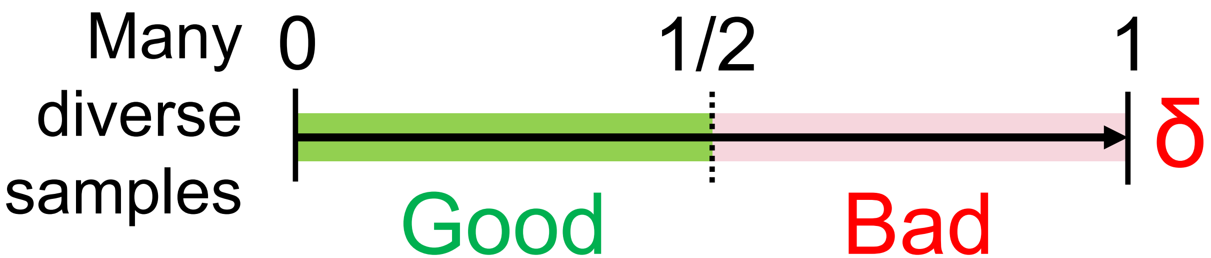
Prior results – No Spurious Local Minima



Main result 1 – Necessary & Sufficient

Let rank $r = 1$.

If $\delta < 1/2$, then no spurious local min.



If $\delta \geq 1/2$, many counterexamples.

First sharp characterization.

Zhang, Sojoudi, Lavaei, *Journal of Machine Learning Research* (2019)

Practical implications for $\delta > 1/2$?

spurious local min	$\not\Rightarrow$	> 0% failure
100% success	$\not\Rightarrow$	no spurious local min

Fundamental limitations of structure-agnostic global guarantees

Zhang, Josz, Sojoudi, Lavaei, *NeurIPS Spotlight* (2018)

Main result 2 – Local Guarantee for State Estimation

(Krumpholz, Clements, Davis 1980)
(Wu & Monticelli 1985)

Let a power system be observable. Then, a neighborhood around the solution has **no spurious local minima**.

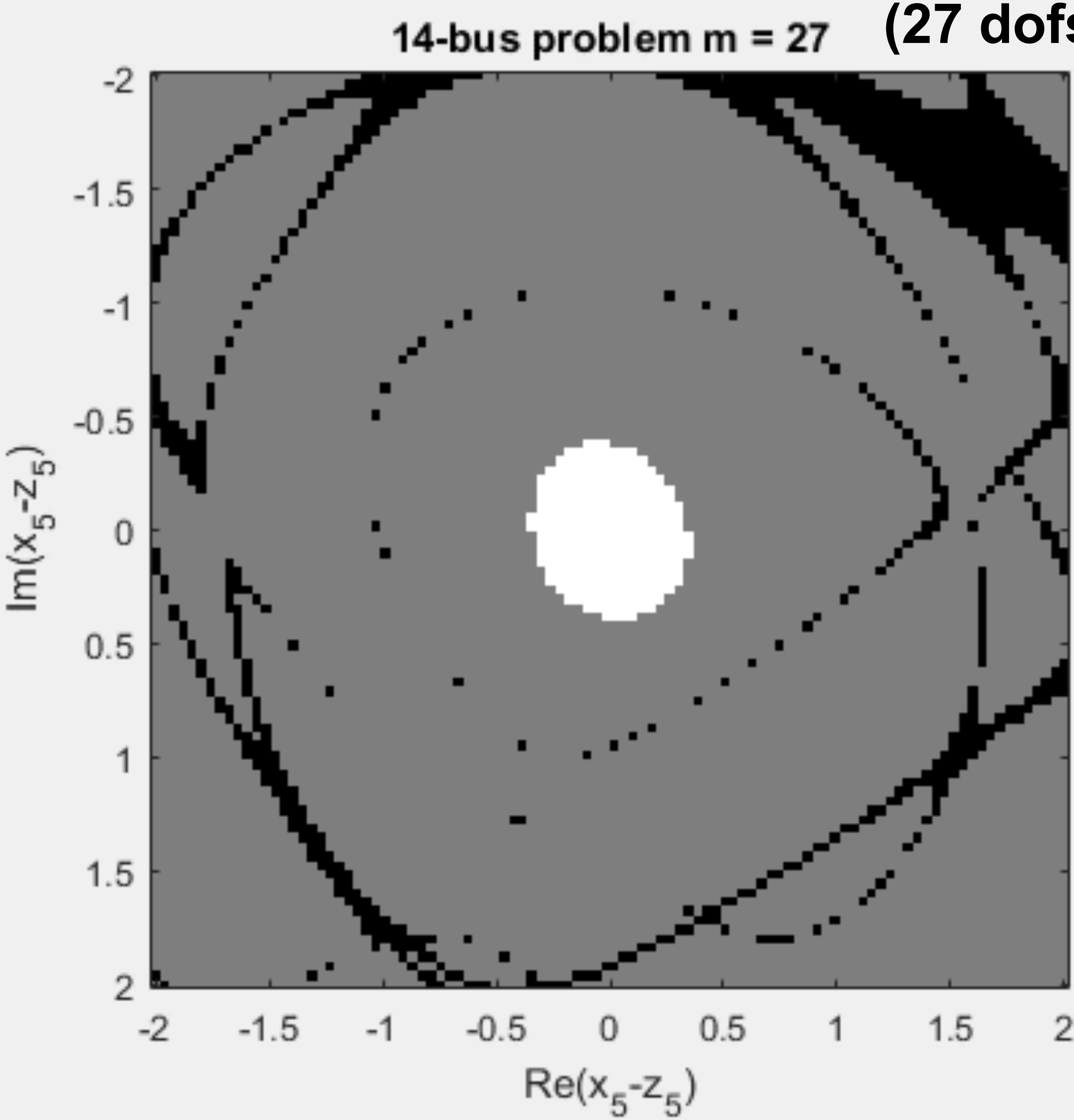
Our neighborhood is very large

Large enough to encompass "typical" initial points

Large enough to be practically useful

Zhang, Lavaei, Baldick, *IEEE Trans. Control of Network Systems* (2018)

(27 dofs)



(white) theorem; (black) possible spurious critical points

Main result 3 – How Many Samples is an Initial Point worth?

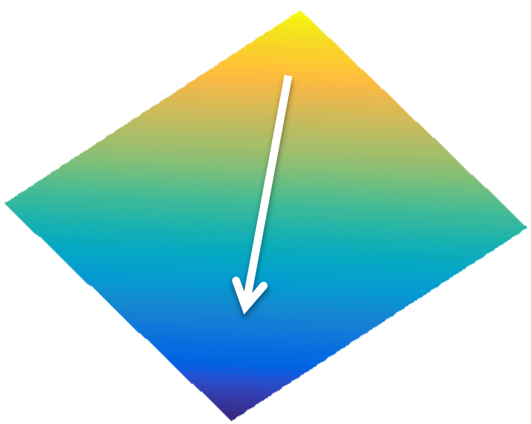
A linear improvement in the quality of the initial guess amounts to a constant factor improvement in the number of required samples.

The better initial guess you have, the fewer samples you need.

Zhang & Zhang, *NeurIPS Spotlight* (2020)

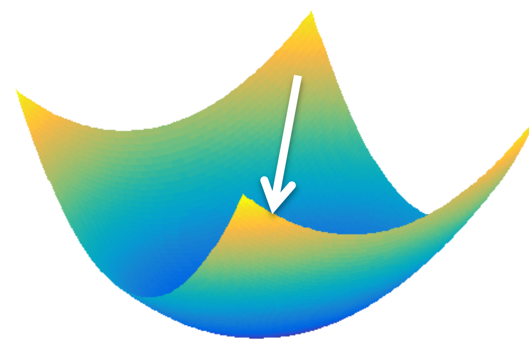
Outline

- ~~Convex relaxation for Optimal Power Flow~~
- ~~Embracing local minima in State Estimation~~
- Concluding remarks



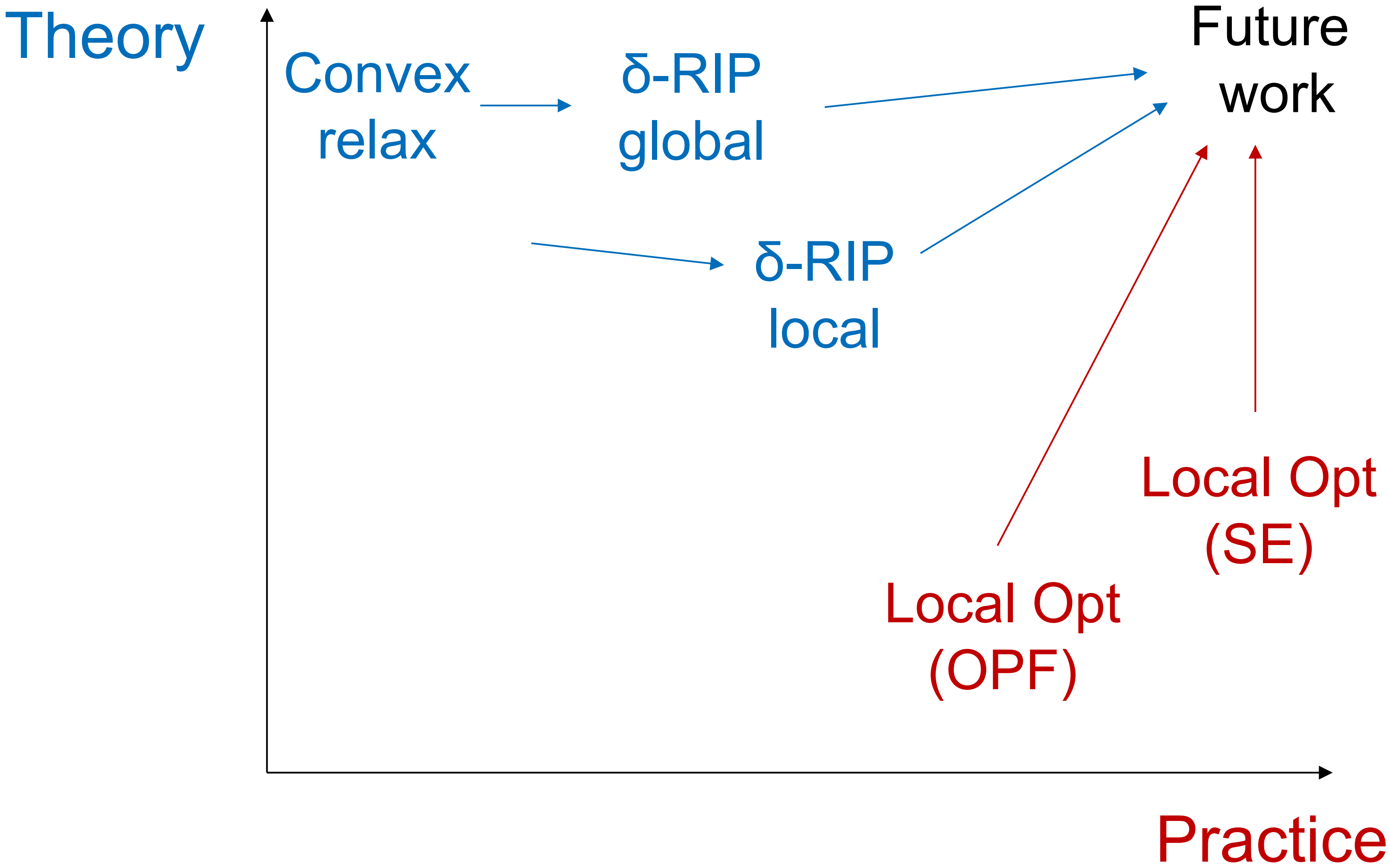
Large gradient

Negative curvature



Globally optimal

Bridging theory and practice



Bringing insights back to OPF

Optimal power flow

min gen-cost **s.t** power demand limits

$$\min_{v \in \mathbb{R}^{2n}} \quad v^T A_0 v \quad \text{s.t.} \quad b_k \leq v^T A_k v \leq c_k \quad \text{for all } k$$

State estimation

$$\min_{v \in \mathbb{R}^{2n}} \quad 0 \quad \text{s.t.} \quad b_k = v^T A_k v \quad \text{for all } k$$

$$\min_{v \in \mathbb{R}^{2n}} \quad \sum_k (v^T A_k v - b_k)^2$$

- Using insights for SE to explain success of OPF
- Developing SE-like techniques for OPF

Conclusions - Thank you!

1. Nonconvexity makes power system optimization hard; issues of local min and infeasibility.
2. Convex relaxation aims to overcome nonconvexity; gains quality and robustness but at cost of speed.
3. Surprising convex-like behavior in nonconvexity between power and voltage. Guarantees for state estimation and variants.
4. Future work: Bring insights back to OPF.

Email: ryz@illinois.edu