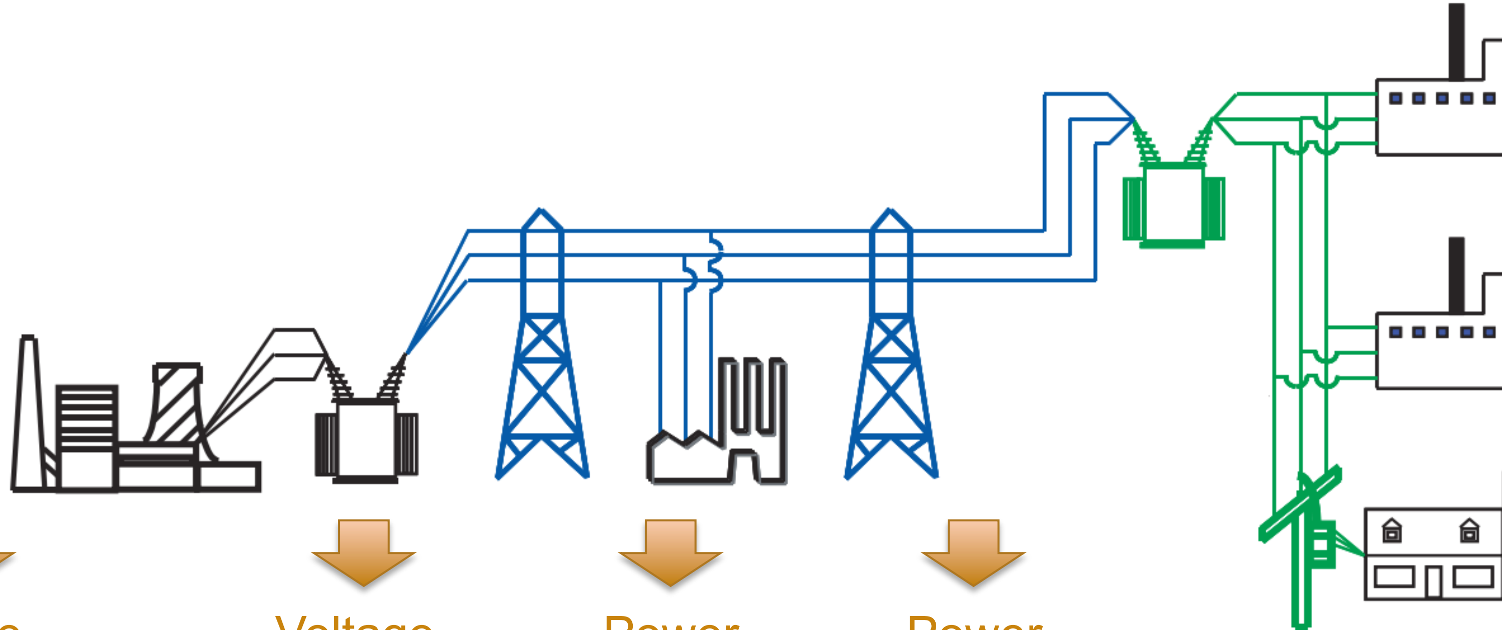
The Dangers of Local Search Algorithms for Power System State Estimation

Richard Y. Zhang
and
Javad Lavaei

Industrial Engineering and Operations Research
University of California, Berkeley

Power System



Voltage phase angles

Voltage magnitudes

Power injections

Power flows

Operator

Minimize energy costs

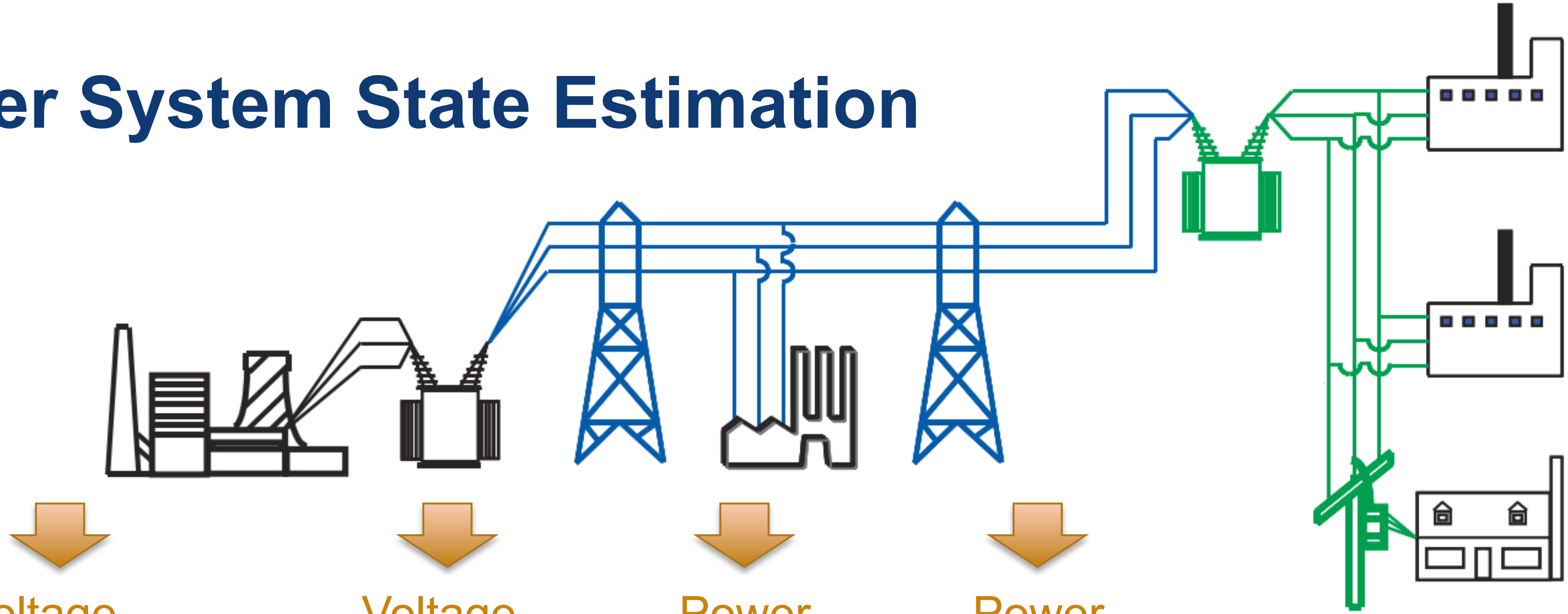
…subject to security constraints.

Monitor & assess system condition

…if needed, take action.

Image Source: U.S.-Canada Power System Outage Task Force, ISO New England

Power System State Estimation



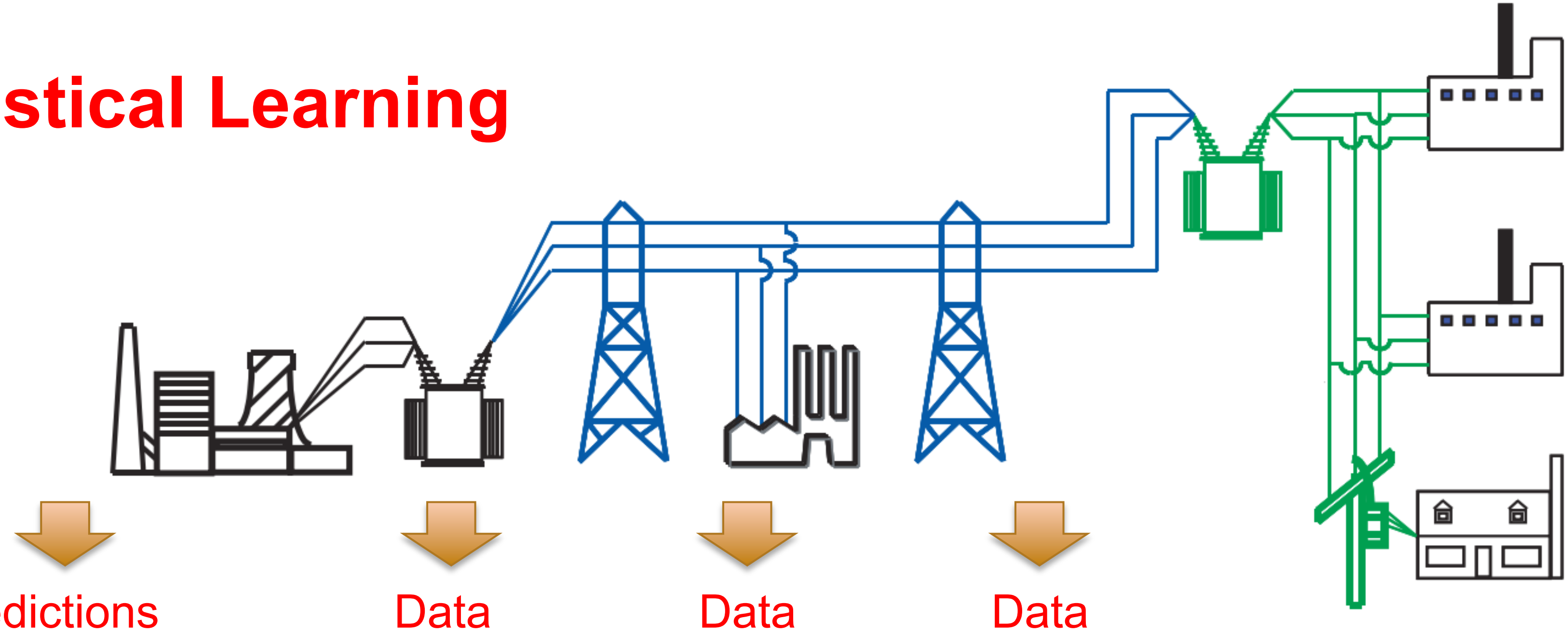
Challenges:

- Some variables cannot be measured.
- Limited number of measurements.
- Corrupted with noise + bad data.

Estimate all state variables using incomplete, inaccurate measurements.

(SE is the system operator’s eyes and ears)

Statistical Learning



Challenges:

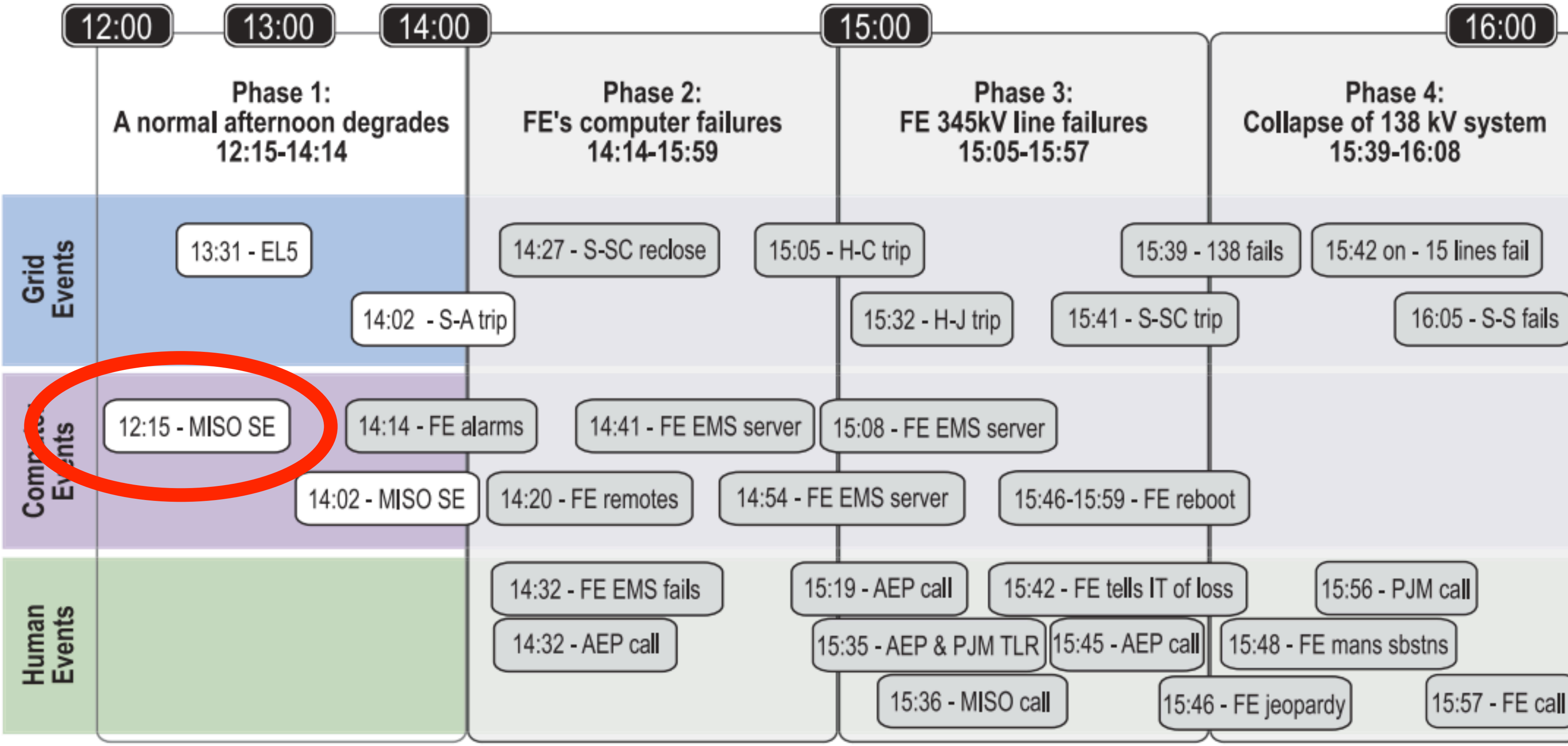
- Predictions cannot be measured.
- Limited amount of data.
- Data corrupted with noise + bad data.

Make predictions using incomplete, inaccurate data.

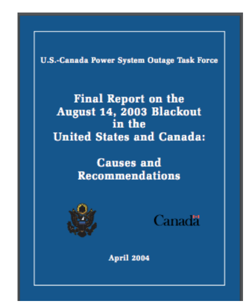
(State estimation is just a special case)

The August 14th, 2003 Northeast Blackout

Figure 5.2. Timeline Phase 1

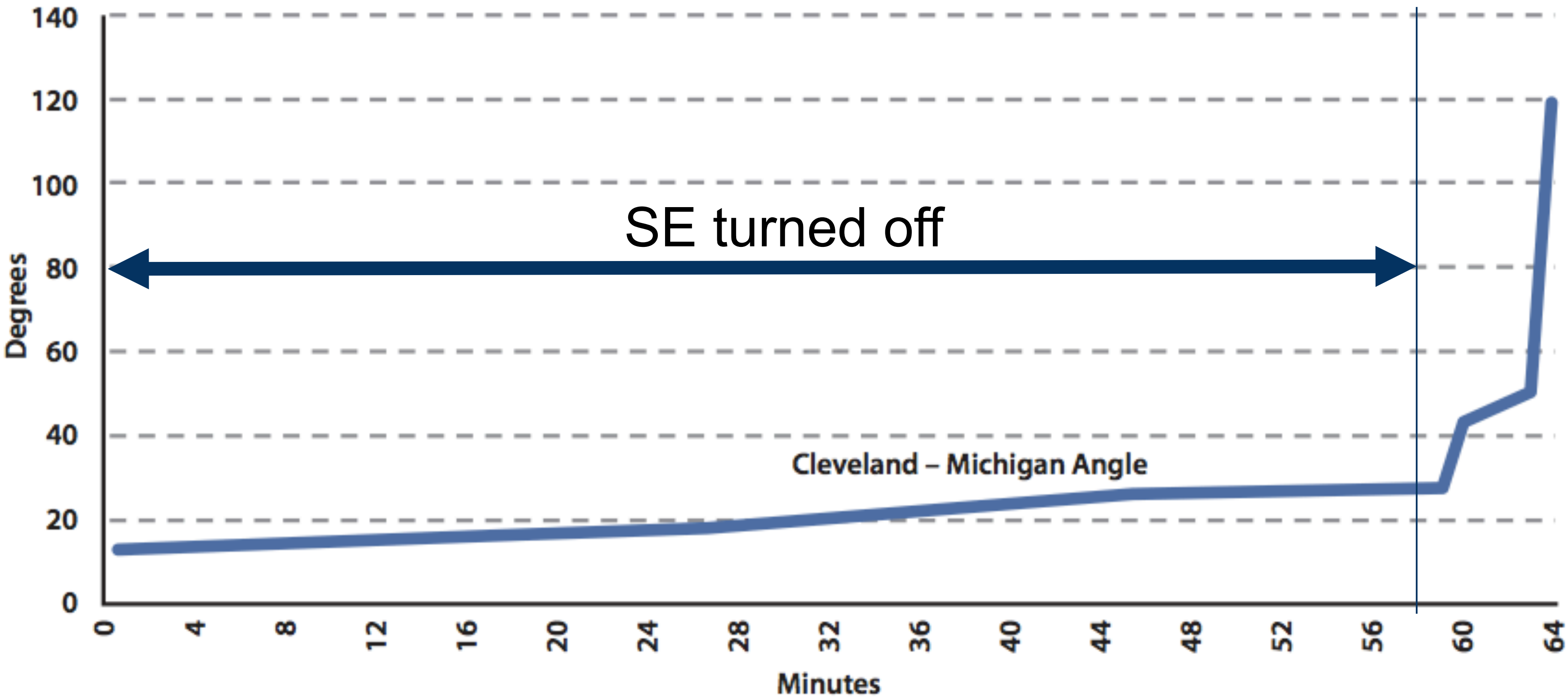


- **MISO's state estimator was inactive for most of the period between 12:15 and 15:34 EDT.**
- Could not identify the system as being on the verge of collapse.



[US-Canada Power System Outage Task Force 2004, p. 48]

Figure 2.3 Cleveland–Michigan Phase Angle Difference Leading Up to the August 2003 Blackout



Source: North American Electric Reliability Corporation Real-Time Application of PMUs to Improve Reliability Task Force, *Real-Time Application of Synchrophasors for Improving Reliability* (Princeton, NJ, 2010), http://www.nerc.com/filez/rapirtf.html.

Image Source: MIT Energy Initialtive, “Future of the Electric Grid Study”, 2011

Integration of Variable Generation

Situational awareness

- Aggregating data on current system status from various sources including EMS/SCADA, load and variable generation forecast systems, and operational planning and/or market results identifying available resources to provide succinct, meaningful displays that support situational awareness.

Real-time reliability/risk assessment

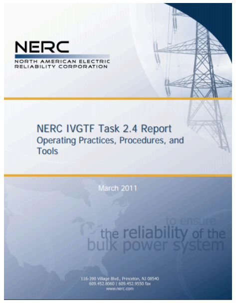
- Evaluation of various dimensions of risk associated with the present and future operating conditions considering elements such as total ramping capability from available resources (supply and demand) and the uncertainty in unit availability, load, and variable generation.

Operator decision support

- Evaluation and recommendation of mitigating actions that can be implemented to solve predicted or realized reliability/security concerns.

- Three key areas of operation support.
- **All rely crucially on good state estimation.**

[NERC IVGTF 2004 Task 2.4, 2011]



Scope

- Focus on State Estimation + classical SCADA measurements.
  - Voltage magnitude, power flows and injection.
  - Classical quadratic formulation due to Schweppe.
  - **Core issue: Quadratic nonconvexity -> Strongly NP-hard.**
- PMUs / Synchrophasors?
  - Quadratic nonconvexity remain (unless all buses have perfect PMUs).
  - Validating consistency (against noisy / bad data) -> Strongly NP-hard.
  - Avoid to keep discussion simple.
- Other measurements (e.g. dq-axis current)?
  - Can be reformulated as quadratic by adding new variables.
  - Quadratic nonconvexity remain -> Strongly NP-hard.
  - Again, avoid to keep discussion simple.

In this talk…

- Review: WLS Estimation
- Local convergence and spurious estimates
- Approaches to avoid local convergence

In this talk…

- Review: WLS Estimation
- Local convergence and spurious estimates
- Approaches to avoid local convergence

Formulation

State Variables (Unknown)

$z \in \mathbb{C}^n$

Voltage phasors

Quadratic Model (Known)

$$F_i(z) = z^* A_i z$$
$$A_i = A_i^*$$

Measurements (Known)

$b_i \in \{b_1, \ldots, b_m\}$

Voltage magnitude, Power measurements

Why quadratic model?

Magnitude-squared is quadratic:

$$|z_1|^2 = z^* \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \ddots \end{bmatrix} z = z^* E_1 z$$

Power is voltage times current, and current is linear wrt voltage, $c = Yz$,

$$\text{Re}\{c_1^* z_1\} = \frac{1}{2} z^* (Y^* E_1 + E_1 Y) z$$

Write each measurement

$$b_i = F_i(z) + \epsilon_i$$

Known: b_i, F_i — Unknown: z, ϵ_i

Model mismatch & measurement error (ϵ_i)

Find estimator $\hat{z} \approx z$ that best explains the measurements.

Weighted Least Squares

[Schweppe 1970]

Minimize the residual sum-of-squares

$$\hat{z} \triangleq \underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} w_i [F_i(x) - b_i]^2$$

Proposition. $\hat{z} \approx z$ is the <u>maximum likelihood estimator</u> if each

$$\epsilon_i = b_i - F_i(z) \qquad (\text{measurement error})$$

is independently & normally distributed with zero mean and variance $1/w_i$.

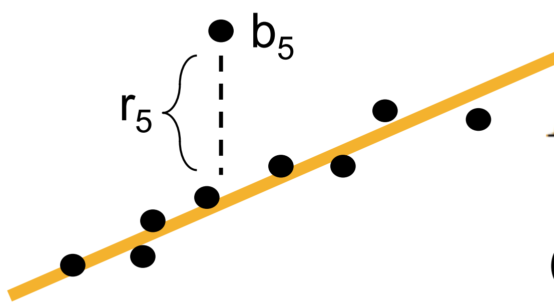
Remark. Must rescale weights w_1 , …, w_m to reflect "trustworthiness" of the data.

- Some data may be bad (variances may be large).
- Bad data are not marked.

Bad Data Detection

If the i-th residual is large, then mark it as bad.

$$r_i \triangleq b_i - F_i(\hat{z})$$



$F_i(\hat{z})$

(Other formulations are also possible)

Solving the Optimization

$$\hat{z} = \underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} w_i [F_i(x) - b_i]^2$$

Nonlinear Least Squares. Schweppe recommended **Gauss-Newton**.

Given initial guess x^0, do k = 0,1,2,…

$$x^{k+1} = \underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} w_i [\underline{F_i(x^k) + \nabla F_i(x^k)(x - x^k)} - b_i]^2$$

Linearize $F_i(x)$ about $x = x^k$

Adopting a step-size rule guarantees convergence.

<u>Gauss-Newton is a local search method.</u>

Other local search methods:

- (Regular) Newton's method,
- Gradient descent,
- Stochastic gradient descent.

Only achieve local optimality.

Global search methods

- Branch & Bound
- Simulated annealing
- Genetic Algorithms

Exponential worst-case time.

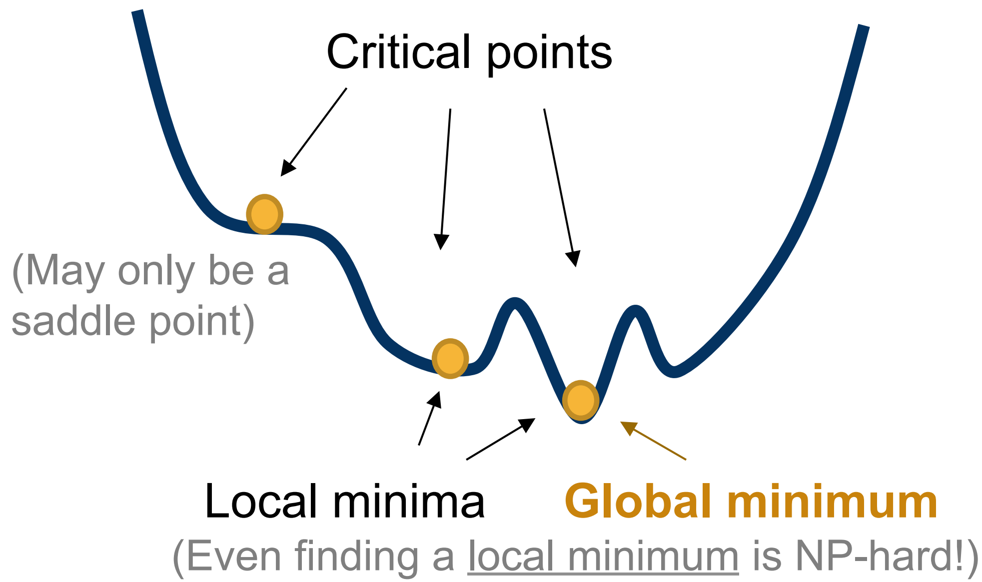
In this talk…

- Review: WLS Estimation
- Local convergence and spurious estimates
- Approaches to avoid local convergence

Local Convergence

$$\hat{z} \triangleq \underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} w_i [F_i(x) - b_i]^2$$

- When $F_i(.)$ is nonlinear, the objective is generally nonconvex.
- Local search can only converge to critical points.
- Finding the global minimum is NP-hard.
- Only the global minimum gives max likelihood estimation.



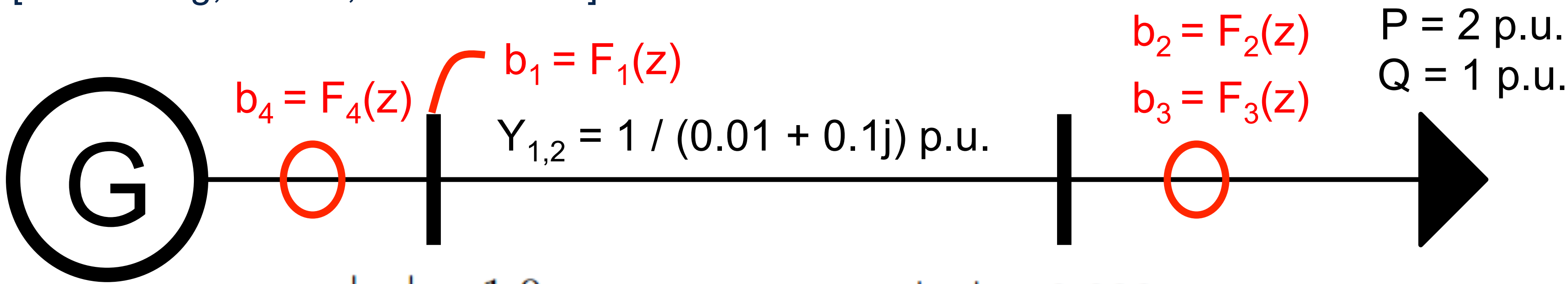
"In the case of statistical and machine learning problems, solving a parameter estimation problem to very high accuracy often yields little to no improvement in actual prediction performance, the real metric of interest in applications."

– Boyd et al.

But in power systems state estimation, inaccuracy can be very dangerous.

Example: Two-Bus System

[R.Y. Zhang, Lavaei, Baldick 2017]



System state: $|z_1| = 1.0$ p.u., $\measuredangle z_1 = 0$ deg; $|z_2| = 0.829$ p.u., $\measuredangle z_2 = -13.2$ deg

Four noise-free measurements:

Bus 1 volt. magn. $b_1 = F_1(z) = z_1^* z_1$

Bus 2 P injection $b_2 = F_2(z) = \mathrm{Re}\,[(Y_{1,2}^*(z_1 - z_2)^* z_2]$

Bus 2 Q injection $b_3 = F_3(z) = \mathrm{Im}\,[Y_{1.2}^*(z_1 - z_2)^* z_2]$

Bus 1 P injection $b_4 = F_4(z) = \mathrm{Re}\,[Y_{1,2}^*(z_2 - z_1)^* z_1]$

Find: Unknown system state z_1, z_2

Given: Model functions $F_1(.), \ldots, F_4(.)$

Noise-free measurements $b_1, \ldots, b_4$

Find: Unknown system state z_1, z_2

Given: Model functions $F_1(.), \ldots, F_4(.)$

Noise-free measurements $b_1, \ldots, b_4$

$$b_1 = F_1(z)$$
$$b_2 = F_2(z)$$
$$b_3 = F_3(z)$$
$$b_4 = F_4(z)$$

Consider nonlinear least-squares

$$\hat{z} = \underset{x_1, x_2 \in \mathbb{C}}{\text{minimize}} \sum_{i=1}^{4} [F_i(x_1, x_2) - b_i]^2$$

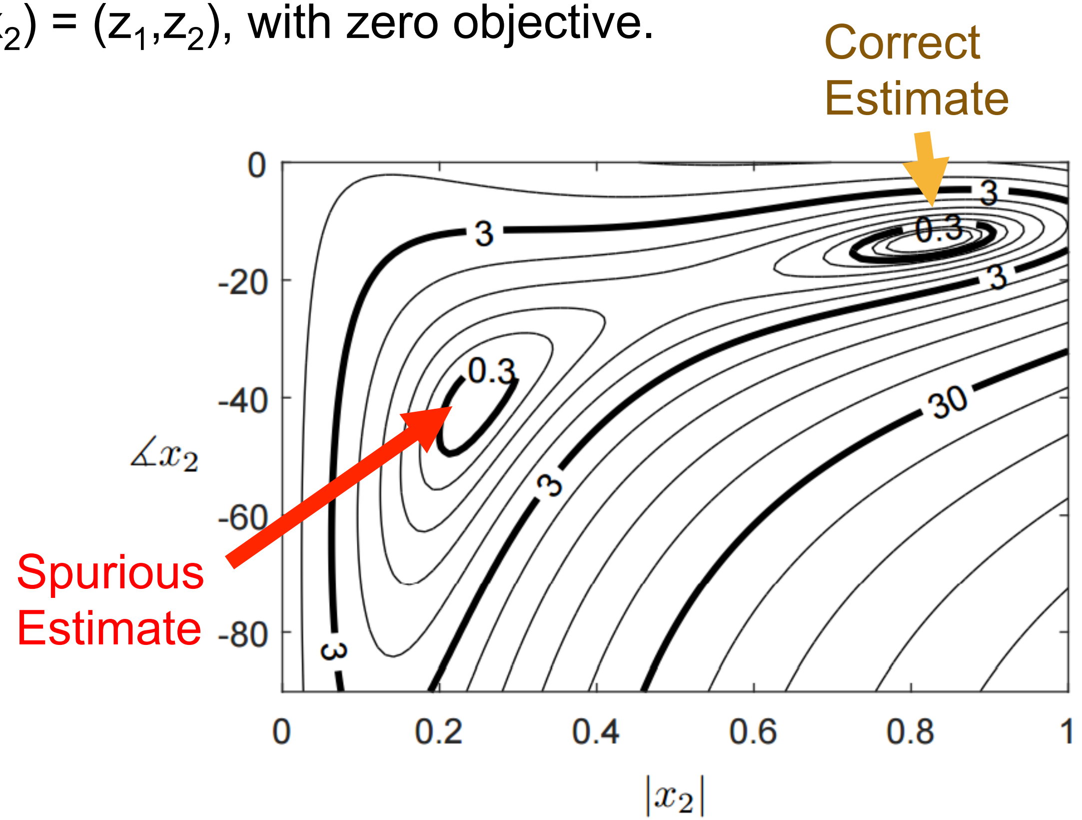
Let's plot the objective function

The global minimizer is $(x_1, x_2) = (z_1, z_2)$, with zero objective.

Problem has 3 dofs:
$|x_1|$, $|x_2|$, and angle x_2.

Let's fix $x_1 = z_1$, and plot over $|x_2|$ and angle x_2.

$$|z_1| = 1.0 \text{ p.u.}$$
$$\measuredangle z_1 = 0 \text{ deg}$$
$$|z_2| = 0.829 \text{ p.u.}$$
$$\measuredangle z_2 = -13.2 \text{ deg}$$



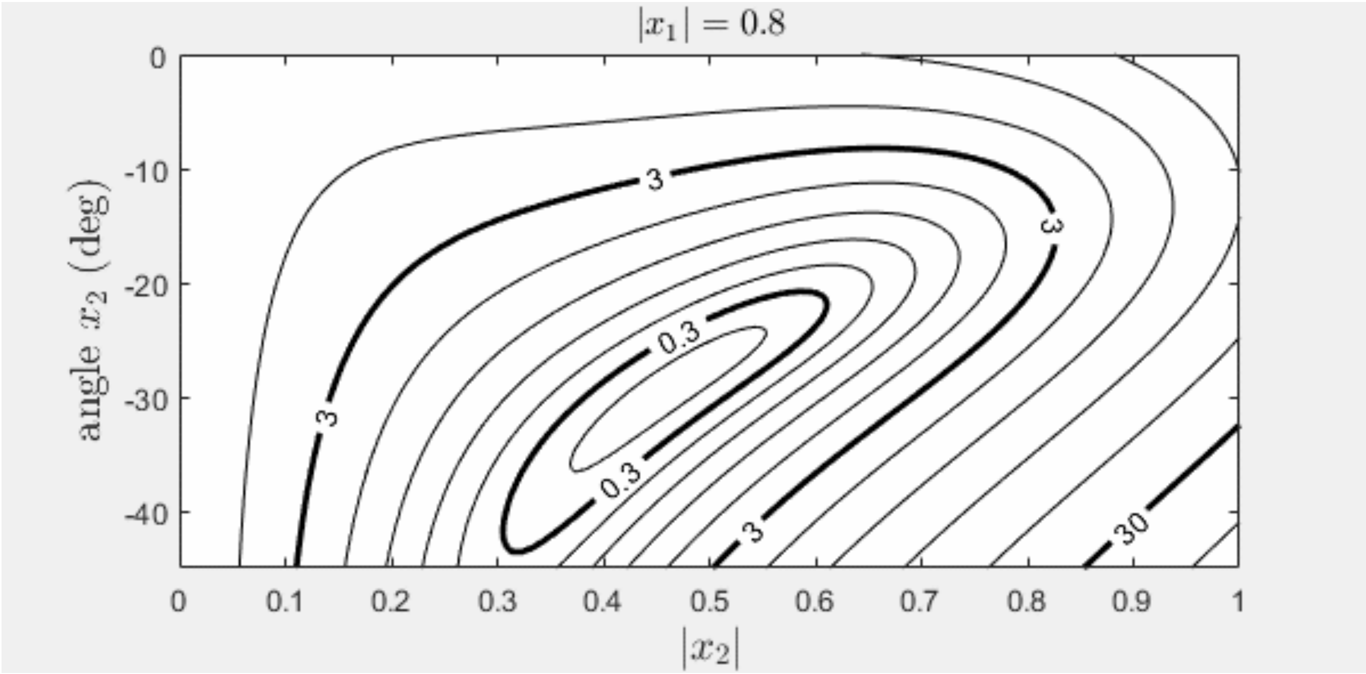
$$\hat{z} \triangleq \underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} [F_i(x) - b_i]^2$$

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

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

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

Both estimates are plausible!



$$\hat{z} \triangleq \underset{x\in\mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} [F_i(x) - b_i]^2$$

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

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

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?

Local convergence gives spurious estimates

[R.Y. Zhang, Lavaei, Baldick 2017]

Can only expect to find <u>critical points</u> of weighted least squares problem

$$\hat{z} \triangleq \underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} [F_i(x) - b_i]^2$$

- Affects even the simplest problems with perfect data.
- Gives plausible but incorrect estimates.
- Misleads re: measurement error / bad data.

Unique closed-form solution:

$$z_1 = \sqrt{F_1(z)},$$

$$\operatorname{Im} z_2 = \frac{F_2(z)\operatorname{Im} Y_{1,2} + F_3(z)\operatorname{Re} Y_{1,2}}{|Y_{12}|^2 z_1},$$

$$\operatorname{Re} z_2 = \frac{F_4(z)/z_1 + z_1 \operatorname{Re} Y_{1,2} + \operatorname{Im} z_2 \operatorname{Im} Y_{1,2}}{\operatorname{Re} Y_{1,2}}.$$

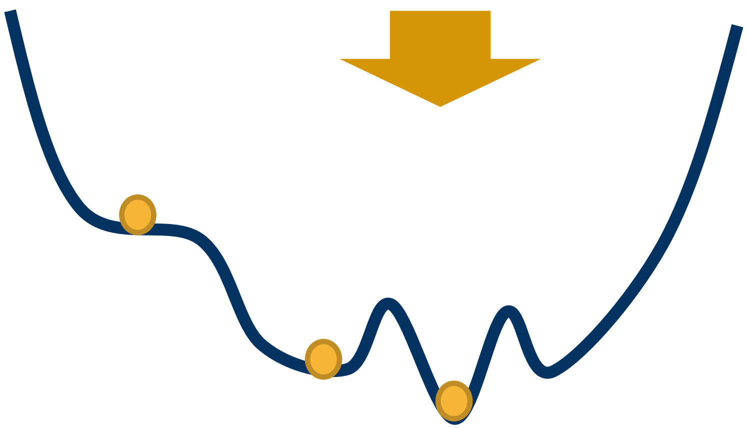
Critical points <u>do not</u> imply:

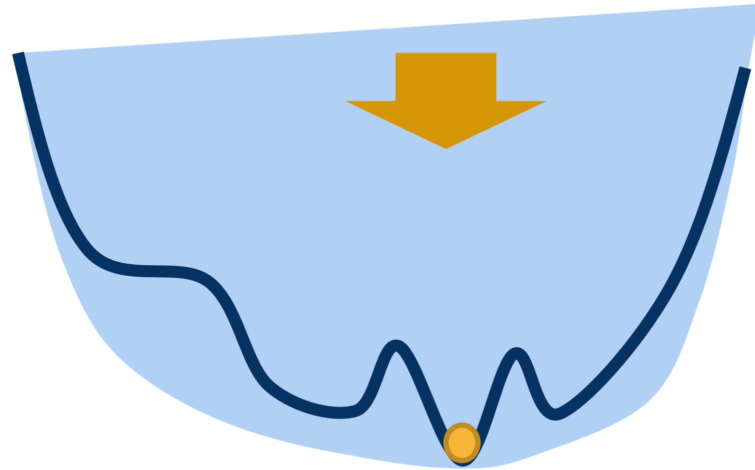
- Nonunique solutions
- Unobservable states
- Inherent “hardness” of the problem

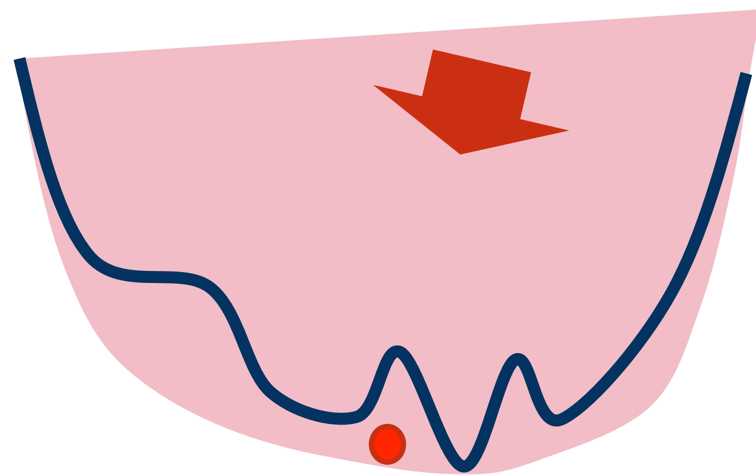
In this talk…

- Review: WLS Estimation
- Local convergence and spurious estimates
- Approaches to avoid local convergence

Convex Relaxations







Consider finding the furthermost point of a nonconvex set.

- Enclose the nonconvex set within a convex set. Then any local minimum is the global minimum (by definition).
- (Success) If that point also lies within the original nonconvex set, then it is a global minimum for the original problem.
- (Failure) If that point does not lie within the original set, then it may be useless.

The Penalized SDP Method

[Madani, Lavaei & Baldick 2016]
[Yu Zhang, Madani, Lavaei 2017]

$$\underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} w_i (x^* A_i x - b_i)^2 \qquad \text{(WLS)}$$

$$\underset{X \succeq 0}{\text{minimize}} \ \operatorname{Tr} CX + \sum_{i=1}^{m} w_i [\operatorname{Tr} A_i X - b_i]^2 \qquad \text{(Relax)}$$

1. Pick special choice of matrix C and solve (Relax)
2. If rank(X) = 1, compute xx* = X and output x.

Theorem (Candes & Recht, Candes & Tao). If A_i are "random", b_i are "noise-free", and m is "sufficiently large", then <u>penalized SDP outputs the global optimum of (WLS)</u> with overwhelming probability.

- Power systems are not "random"; data are seldom "noise-free".
- Nevertheless, often <u>close to global optimum</u>.
- Requires solving an SDP to high accuracy. Complexity may be reduced by exploiting structure. [Andersen, Dahl, Vandenberghe 2014] [Madani, Kalbat, Lavaei 2015] [Zheng et al. 2016] [R.Y. Zhang & Lavaei 2017]

The Penalized SDP Method

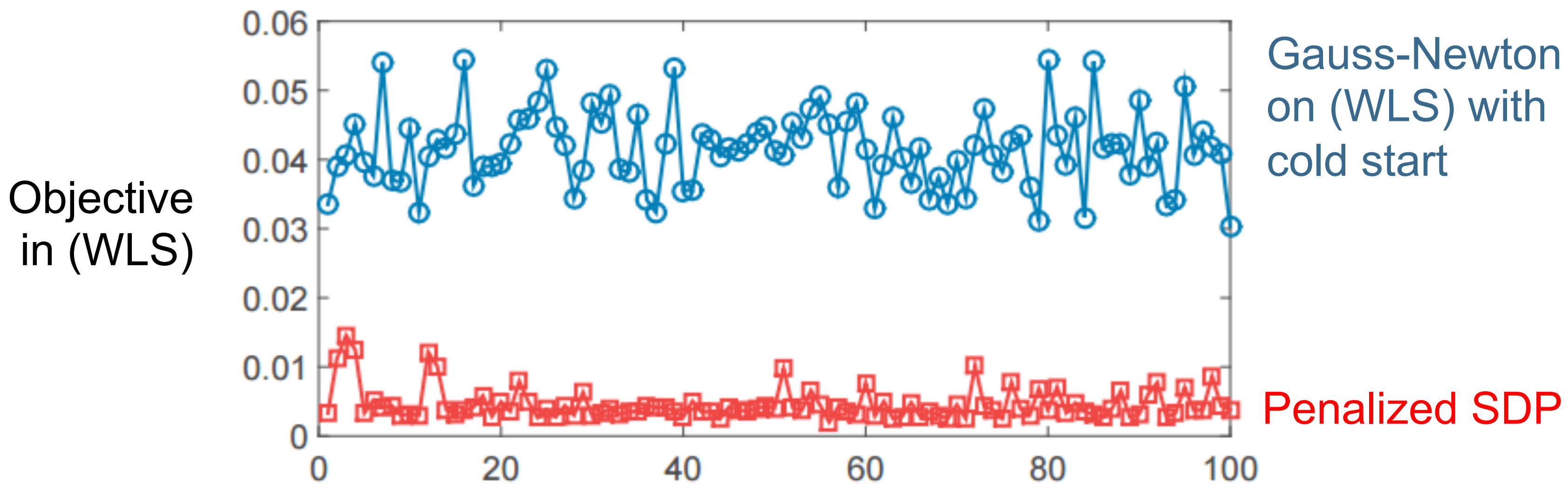
[Madani, Lavaei & Baldick 2016]
[Yu Zhang, Madani, Lavaei 2017]

$$\underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} w_i (x^* A_i x - b_i)^2 \qquad \text{(WLS)}$$

$$\underset{X \succeq 0}{\text{minimize}} \; \operatorname{Tr} CX + \sum_{i=1}^{m} w_i [\operatorname{Tr} A_i X - b_i]^2 \qquad \text{(Relax)}$$

1. Pick special choice of matrix C and solve (Relax)
2. If rank(X) = 1, compute xx* = X and output x.

100 Random instances of IEEE 118-bus system [ZML2017]



Adding Redundant Measurements

[R.Y. Zhang, Lavaei, Baldick 2017]

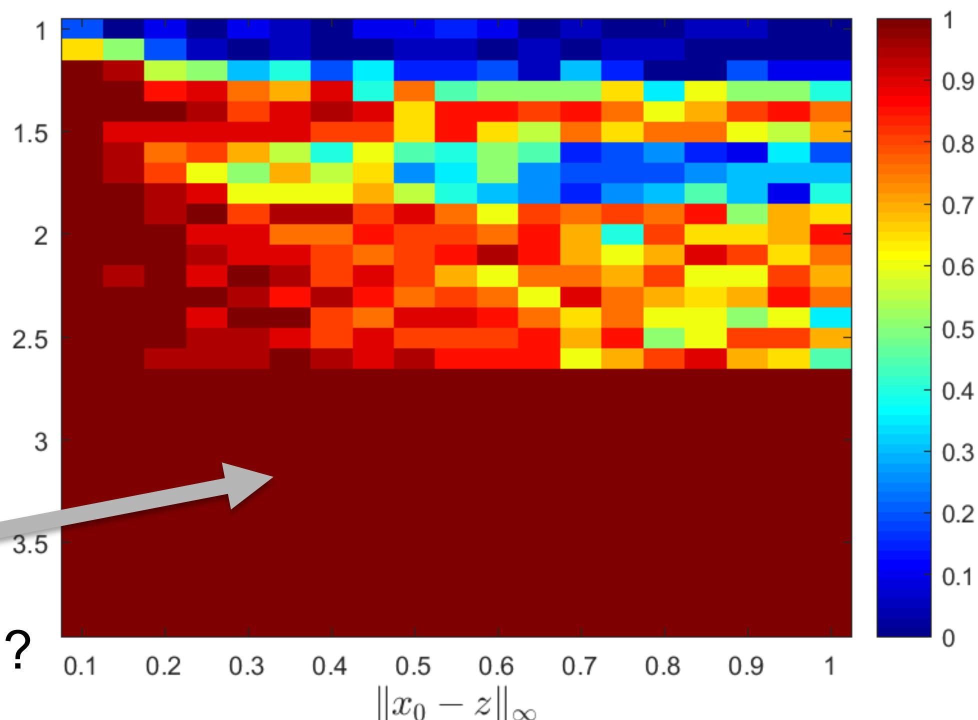
Consider solving WLS with m perfect measurements

$$\underset{x\in\mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m}(x^* A_i x - b_i)^2 \text{ where each } b_i = z^* A_i z$$

using Gauss-Newton with random initial point x^0.

What is the effect of increasing m?

num. meas / num. dof



WLS becomes convex…?

warm-start

cold-start

Success rate over 100 trials for IEEE 39-bus problem

- Begin with power flow constraints
- Randomly add new, perfect measurements without replacement

Adding Redundant Measurements

Consider solving WLS with m perfect measurements

$$\underset{x\in\mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m}(x^* A_i x - b_i)^2 \text{ where each } b_i = z^* A_i z$$

using Gauss-Newton with random initial point x^0.

Theorem (Ge, Lee, Ma 2016). If A_i are "random element-wise", m is "sufficiently large", then after adding a small regularization term, every local minimum is a global minimum to the original problem with overwhelming probability.

- Again, power systems are not "random", data are not "noise-free".
- Should be strongly affected by model / measurement error.
- But much lower time / memory complexity than PSDP.

In Summary…

- State estimation is formulated as nonconvex optimization.
- Classic statistical framework of parameter estimation.
- But local convergence is a significant issue for power systems.
  - Affects all networks, wven with perfect data.
  - Gives plausible but incorrect estimations.
  - Misleads re: measurement error and bad dad.
- Global optimization: Penalized SDP and Redundant Measurements
  - Strong guarantees for random problems.
  - Good empirical performance for the practical problem: near-global-optimal.
  - Future work is to fully understand "why".

Thank you for your attention

Semidefinite Relaxations

Begin with Schweppe's weighted least squares problem:

$$\hat{z} \triangleq \underset{x \in \mathbb{C}^n}{\text{minimize}} \sum_{i=1}^{m} w_i [F_i(x) - b_i]^2 \quad \text{where} \quad F_i(z) = z^* A_i z$$

Define quadratic variable $X = xx^*$ to make quadratic models linear

$$F_i(x) = x^* A_i x = \operatorname{Tr} A_i x x^* = \operatorname{Tr} A_i X.$$

Then,

$$\hat{z}\hat{z}^* = \underset{X = xx^*}{\text{minimize}} \sum_{i=1}^{m} w_i (\operatorname{Tr} A_i X - b_i)^2$$

"X is a rank-1 semidefinite matrix"

Nonconvex

Convex

Classic convex relaxation

$$\underset{X \succeq 0}{\text{minimize}} \operatorname{Tr} C X + \sum_{i=1}^{m} w_i [\operatorname{Tr} A_i X - b_i]^2$$

"X is a semidefinite matrix"

Encourage low-rank solutions