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 Operator

- Monitor & assess system condition
- If needed, take action.

Minimize energy costs
...subject to security constraints.

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)
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.

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


Integration of Variable Generation

• Three key areas of operation support.
• All rely crucially on good state estimation.
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,
\[ c = Y z, \]
\[ \text{Re}\{c_1^* z_1\} = \frac{1}{2} z^* (Y^* E_1 + E_1 Y) z \]

Write each measurement
\[ b_i = F_i(z) + \epsilon_i \]
known
unknown

Model mismatch & measurement error

Find estimator \[ \hat{z} \approx z \] that best explains the measurements.
Weighted Least Squares

Minimize the residual sum-of-squares

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

**Remark.** Must rescale weights \( w_1, \ldots, w_m \) to reflect “trustworthiness” of the data.

**Proposition.** \( \hat{z} \approx z \) is the maximum likelihood estimator if each

\[ \epsilon_i = b_i - F_i(z) \]  

(measurement error)

is independently & normally distributed with zero mean and variance \( 1/w_i \).

**Remark.** Must rescale weights \( w_1, \ldots, w_m \) 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.

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

(Other formulations are also possible)
Solving the Optimization

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

Nonlinear Least Squares. Schweppe recommended Gauss-Newton.
Given initial guess $x^0$, do $k = 0, 1, 2, \ldots$

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

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

Adopting a step-size rule guarantees convergence.

Gauss-Newton is a local search method.

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

\[ \hat{\mathbf{x}} \triangleq \min_{\mathbf{x} \in \mathbb{C}^n} \sum_{i=1}^{m} w_i [F_i(x) - b_i]^2 \]

- 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.
- $\angle z_1 = 0$ deg
- $|z_2| = 0.829$ p.u.
- $\angle z_2 = -13.2$ deg

Four noise-free measurements:
- Bus 1 volt. magn. $b_1 = F_1(z) = z_1^* z_1$
- Bus 2 P injection $b_2 = F_2(z) = \text{Re} \left( (Y_{1,2}^* (z_1 - z_2)^* z_2 \right)$
- Bus 2 Q injection $b_3 = F_3(z) = \text{Im} \left( Y_{1,2}^* (z_1 - z_2)^* z_2 \right)$
- Bus 1 P injection $b_4 = F_4(z) = \text{Re} \left( Y_{1,2}^* (z_2 - z_1)^* z_1 \right)$

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 \( F_1(\cdot), \ldots, F_4(\cdot) \)

Noise-free measurements \( b_1, \ldots, b_4 \)

Consider nonlinear least-squares

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

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

Problem has 3 dofs:

\(|x_1|, |x_2|, \text{ and angle } x_2\).

Let’s fix \( x_1 = z_1 \), and plot over \(|x_2|\) and angle \( x_2 \).

- \(|z_1| = 1.0 \text{ p.u.}\)
- \(\angle z_1 = 0 \text{ deg}\)
- \(|z_2| = 0.829 \text{ p.u.}\)
- \(\angle z_2 = -13.2 \text{ deg}\)
Indeed, we find four critical points, only one of which is the correct estimate.

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

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

Both estimates are plausible!
Indeed, we find four critical points, only one of which is the correct estimate.

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

How well do the two estimates match our measurements?

**Correct Estimate**

\[
\begin{bmatrix}
1 \\
0.829 \\
-13.2^\circ
\end{bmatrix}
\]

- \(F_1(z) - b_1 = 0\)
- \(F_2(z) - b_2 = 0\)
- \(F_3(z) - b_3 = 0\)
- \(F_4(z) - b_4 = 0\)

**Spurious Estimate**

\[
\begin{bmatrix}
0.870 \\
0.345 \\
-35.7^\circ
\end{bmatrix}
\]

- \(F_1(x) - b_1 = -0.24 \text{ p.u.}\)
- \(F_2(x) - b_2 = -0.14 \text{ p.u.}\)
- \(F_3(x) - b_3 = +0.06 \text{ p.u.}\)
- \(F_4(x) - b_4 = -0.17 \text{ p.u.}\)
Local convergence gives spurious estimates
[R.Y. Zhang, Lavaei, Baldick 2017]

Can only expect to find critical points of weighted least squares problem

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

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

Unique closed-form solution:

\[
\begin{align*}
z_1 &= \sqrt{F_1(z)}, \\
\text{Im} z_2 &= \frac{F_2(z)\text{Im} \, Y_{1,2} + F_3(z)\text{Re} \, Y_{1,2}}{|Y_{12}|^2 z_1}, \\
\text{Re} z_2 &= \frac{F_4(z)/z_1 + z_1 \text{Re} \, Y_{1,2} + \text{Im} \, z_2 \text{Im} \, Y_{1,2}}{\text{Re} \, Y_{1,2}}.
\end{align*}
\]

Critical points do not 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
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

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{m} w_i (x^* A_i x - b_i)^2 \\
\text{minimize} & \quad \Tr CX + \sum_{i=1}^{m} w_i [\Tr A_i X - b_i]^2
\end{align*}
\]  
(WLS)  
(Relax)

1. Pick special choice of matrix C and solve (Relax)
2. If \( \text{rank}(X) = 1 \), 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 penalized SDP outputs the global optimum of (WLS) with overwhelming probability.

- Power systems are not “random”; data are seldom “noise-free”.
- Nevertheless, often close to global optimum.
- 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

\[
\min_{x \in \mathbb{C}^n} \sum_{i=1}^{m} w_i (x^* A_i x - b_i)^2 \quad \text{(WLS)}
\]

\[
\min_{X \succeq 0} \text{Tr} CX + \sum_{i=1}^{m} w_i [\text{Tr} A_i X - b_i]^2 \quad \text{(Relax)}
\]

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

Gauss-Newton on (WLS) with cold start

Penalized SDP

[ZML2017]
Adding Redundant Measurements
[R.Y. Zhang, Lavaei, Baldick 2017]

Consider solving WLS with \( m \) perfect measurements

\[
\text{minimize } \sum_{i=1}^{m} (x^* A_i x - b_i)^2 \quad \text{where each } b_i = z^* A_i z
\]

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

What is the effect of increasing \( m \)?

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

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

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, even with perfect data.
  - Gives plausible but incorrect estimations.
  - Misleads re: measurement error and bad data.
- 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:

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

Define quadratic variable \( X = xx^* \) to make quadratic models linear

\[
F_i(x) = x^{*} A_i x = \text{Tr} A_i xx^* = \text{Tr} A_i X.
\]

Then,

\[
\hat{z} \hat{z}^* = \minimize_{X = xx^*} \sum_{i=1}^{m} w_i (\text{Tr} A_i X - b_i)^2
\]

“\( X \) is a rank-1 semidefinite matrix”

Nonconvex

Convex

Classic convex relaxation

\[
\minimize_{X \succeq 0} \text{Tr} CX + \sum_{i=1}^{m} w_i [\text{Tr} A_i X - b_i]^2
\]

“\( X \) is a semidefinite matrix”

Encourage low-rank solutions