Novel Power Grid Reduction Method based on L1 Regularization

Ye Wang, Meng Li, Xinyang Yi, Zhao Song, Michael Orshansky, Constantine Caramanis

University of Texas at Austin

ihywang@utexas.edu

June 3, 2015
Power Grid Reduction: Motivation

- Large-scale power grids
- Simulation time determined by size of power grid: number of ports and elements (R,L,C)
- Speed up simulation/analysis by reducing grid size
  - Focus only on steady-state analysis (DC), hope to extend to other types of analysis (Transient, AC)
Power Grid Reduction: Formulation

- Port relation (Ohm’s law) by $L_G$

\[ L_G v = i \]

- Admittance matrix (Graph Laplacian)

\[ L_G(i, j) = \begin{cases} 
\sum_{k, k \neq i} \omega_{ik}, & \text{if } i = j, \\
-\omega_{ij}, & \text{if } i \neq j \text{ and } \{i, j\} \in E, \\
0, & \text{otherwise.} 
\end{cases} \]

- Goal: want a sparse approximation $L_{G'}$
  - With far fewer non-zeros
  - Preserve similar port relation
State-of-the-Art Methods

- Krylov subspace based methods [FL04,LS06]
  - Project the original system onto a low-rank Krylov subspace for efficiency
- Time-Constant Equilibration Reduction (TICER) [She99]
  - Eliminate low-degree nodes by connecting their neighbors
- Algebraic multigrid methods [SAN03]
  - Reduce the number of nodes and edges simultaneously
- Sampling based spectral sparsification approach [ZFZ14]
  - In time $O(m \log n/\epsilon^2)$, find an $\epsilon$-power approximation $G'$ of $O(n \log n/\epsilon^2)$ edges satisfies:

$$
(1 - \epsilon) v^T L_G v \leq v^T L_{G'} v \leq (1 + \epsilon) v^T L_G v, \quad \forall v \in \mathbb{R}^n.
$$

- They all try to build sparsifiers preserving $L_G v = i$ for all $i \in \mathbb{R}^n$... Is that necessary?
• In practice, currents delivered from ports do not vary unboundedly
  ▶ Peak values of currents can be estimated from system-level description or transistor-level simulation
• The actual space is a small subset of the entire space

• How to utilize the range information not explored before?
  ▶ For more sparsity and accuracy of reduced power grids
Our Main Contribution

- Propose an efficient method that **allows using range information** for better sparsification
- Leverage recent advances of $\ell^1$ regularization to drive sparsity
- We call it graph Sparsification by $\ell^1$ regularization on Laplacian (**SparseLL**)
First Attempt for Sparsification

- Objective function: averaged error in the given range

$$\min_{L_{G'}} \int_{\Omega} \left[ \| (L_G - L_{G'}) v \|_2^2 \right] dv$$

  - Allow to incorporate the range information

- Constraints: sparsity specified by $\ell^0$-norm (number of non-zeros)

  $$\| L_{G'} \|_0 \leq m_0$$

- Non-linear and non-convex in both objective and constraints, hard to solve...
Reformulation as Stochastic Optimization

- Integral discretization by deterministic mesh requires exponential number of samples

\[
\int_{\Omega_V} \left[ \| (L_G - L_{G'}) v \|_2^2 \right] dv \approx \frac{1}{N} \sum_{i=1}^{N} \| (L_G - L_{G'}) v_i \|_2^2
\]

- Randomized discretization leverages fast convergence from stochastic gradient descent (SGD)

- Sample \( v_i \sim \Omega_V \), calculate gradient, update solution
- Converge to optimal solution with rate \( O(\sqrt{1/t}) \) for \( t \) iterations
$\ell^1$ Regularization for Sparsity

- $\ell^0$ constraints are combinatorial and non-convex: result in an NP-hard problem
- $\ell^1$ norm is the tightest while being convex relaxation of $\ell^0$ norm

- Sparsity encouraged by spiky $\ell^1$ norm

![Graphs showing $\ell^0$ and $\ell^1$ norms with non-sparse and sparse solutions]
Complete SparseLL Formulation

- **Objective:** regularized empirical risk function

\[
\min_{L_{G'}} \frac{1}{N} \sum_{i=1}^{N} \left\| (L_G - L_{G'}) v_i \right\|_2^2 + \lambda \left\| L_G' \right\|_1
\]

- Parameter \( \lambda \) controls the degree of sparsity

- **Constraints:**
  
  \[ L_{G'}(i, j) \leq 0, \text{ with } i \neq j, \]
  
  \[ L_{G'} = L_{G'}^T, \]
  
  \[ \sum_{j=1}^{n} L_{G'}(i, j) = 0, \forall i \in \{1, 2, \ldots, n\}. \]
Experimental Results

- Optimizing a sample 17-node power grid

![Graph showing the comparison between Spectral and SparseLL methods for average current error with respect to the number of edges in sparsified graphs.]

- A smaller error while significantly reducing the number of edges

<table>
<thead>
<tr>
<th></th>
<th>Nodes</th>
<th>Edges</th>
<th>Spectral [ZFZ14]</th>
<th>SparseLL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nodes</td>
<td>Edges</td>
<td>Error</td>
<td>Edges</td>
</tr>
<tr>
<td>rand1</td>
<td>100</td>
<td>4000</td>
<td>5.40%</td>
<td>1031</td>
</tr>
<tr>
<td>rand2</td>
<td>500</td>
<td>100000</td>
<td>4.44%</td>
<td>8120</td>
</tr>
<tr>
<td>rand3</td>
<td>1000</td>
<td>400000</td>
<td>4.80%</td>
<td>15114</td>
</tr>
<tr>
<td>ibmpg1</td>
<td>5388</td>
<td>27000</td>
<td>3.80%</td>
<td>6703</td>
</tr>
</tbody>
</table>
• Identified and specified a realistic range for currents
• Formulated power grid reduction as a convex optimization problem
  ▶ With objective as average current error in that range
  ▶ Use $\ell^1$ norm to encourage sparsity
• Solved the problem using an efficient SGD algorithm