Assignment of Electrical Properties to Power Grid Topologies

Eran Schweitzer  
Arizona State University  
eran.schweitzer@asu.edu

Anna Scaglione  
Arizona State University  
anne.scaglione@asu.edu

Kory Hedman  
Arizona State University  
kory.hedman@asu.edu

Abstract—Power systems are often described in terms of graphs, with various properties like load and impedance associated with the nodes and edges. These properties are coupled to the graph’s topology, reflecting the great deal of engineering design in the power system. With the goal of automating the creation of usable synthetic cases, the problem of assigning these properties is considered. It is formulated as a Mixed Integer Program (MIP), which aims to minimize the absolute value of angle difference across a branch. Since the problem quickly balloons in size, a decomposition into smaller zones is explored, that enables scaling the problem to larger system sizes. Experiments demonstrate the efficacy and viability of the approach.

Index Terms—Assignment Problem, DC Powerflow, Synthetic Test Cases, MIP

1. Notation

Sets
- $\mathcal{N}$ Set of buses.
- $\mathcal{L}$ Set of branches.
- $\mathcal{F}_n$ Set of branches with bus $n$ defined as the from bus.
- $\mathcal{T}_n$ Set of branches with bus $n$ defined as the to bus.
- $\mathcal{H}$ Set of zones.
- $\mathcal{B}$ Set of boundary nodes.
- $\mathcal{E}$ Set of boundary edges.
- $\mathcal{P}_1$ Set of nodes with nodal degree 1.
- $\mathcal{D}$ Subset of $\mathcal{N}$ with generation.
- $\mathcal{P}$ Subset of $\mathcal{N}$ with load.

Parameters
- $\delta_{\text{max}}$ Maximum allowable angle difference between adjacent nodes.
- $f_{\text{max}}$ Maximum allowable branch flow.
- $P_{g/d}$ Maximum and minimum allowable generation and load, respectively, at a single bus.
- $b$ Branch susceptances.
- $p$ Nodal power injections: $p_n = P_{g,n} - P_{d,n}$.
- $\lambda$ Large multiplier used to create disjunctive constraints.
- $\alpha[t], \rho[t]$ Dual variable update step size for iteration $t$ in Lagrange relaxation and ADMM respectively.

Subscripts are also used to index variables by their respective sets, so that $f_{\ell}$ is the flow on line $\ell \in \mathcal{L}$. Edges are either referred to by an index, as in $\ell \in \mathcal{L}$, or by the node pair defining the edge, $(f, t) \in \mathcal{L}$. When both index and node pair are needed, we notate it as $\ell : (f, t) \in \mathcal{L}$, meaning that $f$ and $t$ are the end nodes of edge $\ell$.

2. Introduction

The general problem considered in this paper is how to assign properties to a topology, i.e., collection of nodes and edges of a graph. More specifically, the graph considered represents a bulk transmission power system, such that the DC powerflow assumptions [1], [2] are valid. The goal can therefore be translated to assigning the key properties in the DC powerflow—power injections $p$, and susceptances $b$—to a topology, in such a way that the resulting case will exhibit reasonable behavior. The most natural elements to look at when considering how reasonable a DC powerflow solution is, are the angle differences between adjacent nodes and the real power flowing in the branches. Empirically, (see Section 5) the distribution of angle difference appears to fit a Laplacian distribution well, which also serves to motivate the one norm minimization objective in the main formulation.

URL: http://hdl.handle.net/10125/50235  
ISBN: 978-0-9981331-1-9  
(CC BY-NC-ND 4.0)
2.1. Motivation and Related Work

This work is motivated by growing sentiment in the power systems research community that there is a lack of publicly available test cases [3]–[5]. As a result, it is particularly difficult to reproduce many results in the literature. Additionally, easier access to larger number of cases enables more robust testing on ensembles of test systems à la Monte Carlo methods.

The topology of the power grid has been studied quite extensively in the literature. It was first compared to small-world graphs in [6], then to preferential attachment models in [7], and many others since, a review of which can be found in [8]. Other work has built on top of these studies to synthesize topologies. In [3] small world graphs are used as the backbone for construction, while [9] uses a growth model more similar to preferential attachment, and [10] exploits clustering related observations. While these works focus on topology, most of them treat the properties that attach to the topology either briefly, as in [3], or not at all.

A good fit of some structural features of the admittance matrix is insufficient for power systems analysis, because to solve a simple power flow, the additional basic information placing load and generation on the graph is needed. One can easily draw the values of generator and load set-points from realistic distributions, but a completely random assignment will, in general, lead to power flow solutions that differ significantly from realistic ones.

More recent work has emphasized that the assignment problem is far from trivial. In [11] a correlation between the system impedance matrix and location of generators is observed, and [12], [13] explores some of the coupling phenomena between different bus types in the power system. The authors in [13] used their finding to solve the placement problem for generation and load so that the statistical features found in real cases regarding the relationship between degree and generation and load are matched.

Here we look at the problem from another perspective. The main problem with placing the generation, loads and impedance at random is that it results in unrealistic operating points for the grid sample, making the synthetic case useless for any kind of realistic analysis. This motivation is supported by the numerical analysis carried out in Section 5.2. Rather than trying to fit some structural feature that may or may not be directly impacting the power flow solutions, our idea is that one can directly reproduce likely operating points statistics as the objective of the placement problem, starting from a topology and assigning load, generation and impedances through an optimization problem. In short, the novelty of this approach is that, rather than trying to fit statistics of structural properties we are directly targeting the statistics of the operational values.

It should be noted that there is another current approach to building large synthetic systems, as exemplified in [4], where geographic and other publicly available data is utilized to construct systems on a geographical footprint. The approach in [4] is therefore capable of providing far richer cases, in terms of included parameters, at the cost of currently requiring more fine-tuning by hand.

2.2. Problem Scope

This paper attempts to find a somewhat agnostic compromise between all the previously mentioned works. The problem is formulated as follows:

Given a topology, set of power injections, and line susceptances of appropriate size, find a suitable assignment for the elements of the two sets resulting in satisfactory DC powerflow results.

The criterion that we choose to define the suitability of the assignment is that the solution of the DC power flow problem produces a realistic distribution of phase angle differences. Since the topology as well as the distributions of power injections, p, and branch susceptances, b, are assumed correct, the specific method in which they are generated is apart from the assignment method. The main novelty of the work is to provide an automatic way of generating usable powerflow cases given these inputs, with the capability of generating an arbitrary number of cases just as easily as it produces a single one, neglecting the effects of computation time.

The remainder of the paper is structured as follows. An extensive formulation of the problem is presented in Section 3. Since this problem quickly becomes impossibly large, Section 4 introduces methods to decompose the problem into independent, smaller, and computationally manageable subproblems. Section 5 reports results from experiments and Section 6 provides further discussion of these results and highlights the main remaining challenges. Finally, Section 7 discusses future research directions and concludes the paper.

3. Problem Formulation

The following inputs are assumed to be known:

1) The system topology, where each edge has a defined from and to bus.
2) Set of nodal injections p, but without mapping to the topology.
3) Set of branch susceptances b, but without mapping to the topology.
4) Maximum allowable angle difference $\delta_{\text{max}}$, as well as absolute maximum branch flow, $f_{\text{max}}$.

Given these inputs, the assignment problem is solved by determining the entries of permutation matrices $\Pi$ and $Z$, which map the elements of $p$ and $b$, respectively, to the nodes and edges of the topology.

At its core, this is a feasibility problem, since the goal is to find a viable solution rather than the solution. However, the solution space is, at least in theory, very large and not equally weighted, meaning that some solutions are “better” than others. Better should be understood in this context as more closely matching results seen in real cases. In keeping with observations, as well as typical power system
operations, the objective function is chosen to minimize the angle difference magnitude between adjacent buses in the system. This choice can can be seen as simultaneously maximizing stability margins and minimizing system losses. Let us denote $s$ the absolute value of angle differences between adjacent nodes, i.e. if $(f,t)$ are end nodes of an edge $\ell$ then $s_\ell = |\theta_f - \theta_t|$. Our objective can be cast into the following mathematical formulation.

\[
\text{Minimize} \quad \sum_{\ell \in \mathcal{L}} s_\ell \quad \quad \quad \quad (1a)
\]

Subject to \[
-f_{\max} \leq f_\ell \leq f_{\max} \quad \forall \ell \in \mathcal{L} \quad \quad (1b)
\]
\[
-\delta_{\max} \leq \theta_f - \theta_t \leq \delta_{\max} \quad \forall (f,t) \in \mathcal{L} \quad \quad (1c)
\]
\[
f_\ell + b_\ell (\theta_f - \theta_t) + (1 - Z_\ell)M \geq 0 \quad \forall \ell : (f,t), \ell \in \mathcal{L} \quad \quad (1d)
\]
\[
f_\ell + b_\ell (\theta_f - \theta_t) - (1 - Z_\ell)M \leq 0 \quad \forall \ell : (f,t), \ell \in \mathcal{L} \quad \quad (1e)
\]
\[
\sum_{n \in \mathcal{N}} p_n \Pi_{n\bar{n}} + \sum_{\ell \in \mathcal{P}} f_\ell - \sum_{\ell \in \mathcal{S}} f_\ell = 0 \quad \forall n \in \mathcal{N} \quad \quad (1f)
\]
\[
\sum_{n \in \mathcal{N}} \Pi_{n\bar{n}} = 1, \quad \sum_{n \in \mathcal{N}} \Pi_{\bar{n}n} = 1 \quad \forall n \in \mathcal{N} \quad \quad (1g)
\]
\[
\sum_{\ell \in \mathcal{L}} Z_{\ell\ell} = 1, \quad \sum_{\ell \in \mathcal{L}} Z_{\ell\bar{\ell}} = 1 \quad \forall \ell \in \mathcal{L} \quad \quad (1h)
\]
\[
\Pi_{n\bar{n}} = 0 \quad \forall n, \bar{n} : n \in \mathcal{P}_1 \cap p_\bar{n} = 0 \quad \quad (1i)
\]
\[
s_\ell + (\theta_f - \theta_t) \geq 0 \quad \forall \ell : (f,t) \in \mathcal{L} \quad \quad (1j)
\]
\[
s_\ell - (\theta_f - \theta_t) \geq 0 \quad \forall \ell : (f,t) \in \mathcal{L} \quad \quad (1k)
\]
\[
s_\ell \geq 0 \quad \forall \ell \in \mathcal{L} \quad \quad (1l)
\]
\[
\Pi_{n\bar{n}} \in \{0,1\} \quad \forall n, \bar{n} \quad \quad (1m)
\]
\[
Z_{\ell\ell} \in \{0,1\} \quad \forall \ell, \bar{\ell} \quad \quad (1n)
\]

Constraints (1b) and (1c) limit the allowable flows and angle differences, while constraints (1d)–(1f) are the DC power flow equations. The disjunctive formulation in (1d) and (1e) is needed to select entries from susceptances, $b$, based on $Z$, without forming bilinear terms by multiplying $Zb(\theta_f - \theta_t)$ directly. Constraints (1g) and (1h), in conjunction with the binary constraints, force $\Pi$ and $Z$ to be permutation matrices, and constraints (1i)–(1l) are simply a way to enforce $s = |\theta_f - \theta_t|$. Additionally, constraint (1i) is added to prevent leaf nodes (those with degree one) from receiving a zero injection assignment. The reasoning for this decision is that a leaf node with zero injection can be trivially removed from the system and therefore, such an assignment is undesirable.

The disjunctive constraints are bound to slow down the optimization, so it is critical to consider an appropriate $\mathcal{M}$ value. By considering (1b)–(1e), It is clear that any $\mathcal{M}$ satisfying,

\[
\mathcal{M} \geq f_{\max} + \delta_{\max} \max_\ell |b_\ell| \quad \quad (2)
\]

will achieve the desired result, with equality representing the smallest possible value.

---

1. An illustrative example for WECC can be found here: https://www.wecc.biz/Administrative/WECC_BAMap.pdf
multiple iterations the induced subgraphs will become disconnected. For this reason, the second half of the algorithm in Figure 3 checks for disconnected components, and if any are smaller than a predefined size, $N_{\text{min}}$, it finds a neighboring subgraph and combines the two. This procedure could be of interest beyond the scope of this work, as it provides a systematic way of creating zones. Synthetic systems, regardless of how they were created, could use this approach to identify natural borders for different balancing authorities.

After zones are created, an appropriately sized subset of $p$ and $b$ are randomly assigned to each. The only restriction placed on the random assignment is that the zonal imbalance, i.e., the sum of the subset of $p$ in absolute value, be smaller than the maximum import/export capacity for the zone. This is in turn dictated by the number of boundary edges and the maximum allowable flow. For example, for zone $i$ the following would have to hold:

$$\left| \sum_{n\in\mathcal{H}_i} p_n \right| \leq f_{\text{max}} |\mathcal{E}_i|.$$  \hspace{1cm} (3)

This initial assignment greatly simplifies the problem but has the implication that the decomposition cannot be completely equivalent to the extensive formulation. However, for sufficiently large zones, the difference should be minimal. Furthermore, achieving optimality is not really necessary to obtain good statistics (see also Section 5).

4.2. Inter-Zone Flows

Nodes on the boundary of zones are now capable of receiving and sending power on edges that are not represented in their subproblem. To handle this discrepancy, balance constraints (1f) for the boundary nodes are relaxed by adding slack variables corresponding to inter-tie flows, identified as $\beta$. Figure 4 illustrates the basic relationships between zones expressed in the equations that follow. Constraint (1f) for boundary nodes becomes:

$$\sum_{\mathcal{H}} p_n \Pi_n + \sum_{\ell \in \mathcal{B}_i} f_{\ell} - \sum_{\ell \in \mathcal{B}_i} f_{\ell}^{+}$$

$$\sum_{\ell \in \mathcal{B}_i \cap \mathcal{E}_n} \beta_{\ell}^{+} - \sum_{\ell \in \mathcal{B}_i \cap \mathcal{E}_n} \beta_{\ell}^{-} = 0, \forall i \in \mathcal{H}, \forall n \in \mathcal{B}_i \hspace{1cm} (4)$$

where $\mathcal{B}_i$ is the set of boundary nodes for zone $i$, and $\mathcal{E}_i$ the set of boundary edges for zone $i$. To minimize the absolute value of inter-tie flows, the following constraints are added:

$$\beta_{\ell}^{+} \geq \beta_{\ell}^{-} \hspace{1cm} \forall i \in \mathcal{H}, \forall \ell \in \mathcal{E}_i \hspace{1cm} (5a)$$

$$\beta_{\ell}^{+} \geq -\beta_{\ell}^{+} \hspace{1cm} \forall i \in \mathcal{H}, \forall \ell \in \mathcal{E}_i \hspace{1cm} (5b)$$

and the objective becomes,

$$f(s, \beta^{+}, \beta^{-}) = \sum_{\ell \in \mathcal{L}} s_{\ell} + \sum_{\ell \in \mathcal{E}} (\beta_{\ell}^{+} + \beta_{\ell}^{-}).$$  \hspace{1cm} (6)
By defining the sets of edges between any two zones $i$ and $j$ as,

$$
\mathcal{E}_{ij} = \begin{cases} 
\mathcal{E}_i \cap \mathcal{E}_j & i < j \\
0 & \text{otherwise}
\end{cases}
$$

the constraints coupling the zones can be expressed as,

$$
\beta_k^i = \beta_k^j \quad \forall \ell \in \mathcal{E}_{ij}, \forall i, j \in \mathcal{H},
$$

which simply mean that the zones agree on the inter-tie flow. The full problem can therefore be reformulated to:

$$
\begin{align*}
\text{Minimize} & \quad \sum_{i \in \mathcal{H}} f_i(s_i, \beta_i) \\
\text{Subject to} & \quad \beta_k^i = \beta_k^j \quad \forall \ell \in \mathcal{E}_{ij}, \forall i, j \in \mathcal{H} \\
& \quad \text{constraints in (1), (4), and (5)}.
\end{align*}
$$

Since the constraints in (8) are the only ones with “complicating” variables, that is, they are the only ones coupling the zones, they are the obvious choice for relaxation, which is considered in the next two sections.

### 4.3. Solution by Lagrange Relaxation

Recalling that the subvector of $\beta_i$ corresponding to the boundary between zones $i$ and $j$ is denoted as,

$$
\beta^i_{\mathcal{E}_{ij}} = \{ \beta^i_\ell : \ell \in \mathcal{E}_{ij} \},
$$

the Lagrangian for the problem in (9) with (8) relaxed is,

$$
L(s_1, \ldots, s_{|\mathcal{H}|}, \beta_1, \ldots, \beta_{|\mathcal{H}|}) = \sum_{i \in \mathcal{H}} f_i(s_i, \beta_i) + \sum_{i,j \in \mathcal{H}} \nu^i_{\mathcal{E}_{ij}} (\beta^i_{\mathcal{E}_{ij}} - \beta^j_{\mathcal{E}_{ij}}).
$$

2. $\beta^\pm$ are replaced by $\beta$ for notational clarity, since they are only a way of representing the absolute value of $\beta$.

Each subproblem or zone, $k$, can be now solved separately:

$$
\begin{align*}
\text{Minimize} & \quad f(s_k, \beta_k) + \sum_{j \in \mathcal{H}} \nu^T_{\mathcal{E}_{kj}} \beta^k_{\mathcal{E}_{kj}} - \sum_{i \in \mathcal{H}} \nu^T_{\mathcal{E}_{ik}} \beta^k_{\mathcal{E}_{ik}} \\
\text{Subject to} & \quad \text{constraints in (1), (4), and (5) for zone } k \text{ only.}
\end{align*}
$$

Dual variable vector $\nu$ is updated after each iteration $t$ based on the rule:

$$
\nu_{\mathcal{E}_{ij}}[t + 1] := \nu_{\mathcal{E}_{ij}}[t] + \alpha(t) (\beta^i_{\mathcal{E}_{ij}}[t] - \beta^j_{\mathcal{E}_{ij}}[t]) \quad \forall \mathcal{E}_{ij}
$$

where $\alpha[t]$ is a chosen step size. The dual in this way can be seen to represent the “price” of enforcing equality constraint (8), or conversely the benefit of relaxing it with some slack. Due to the non convex nature of each subproblem, the usual guarantees on dual decomposition and gradient decent do not hold. In practice, the problem is found to be highly sensitive to step size $\alpha[t]$. Without very careful tuning, the solution oscillates, so that every element in $\beta$ that is in disagreement with its neighbor alternates sign on each iteration. To address this issue, an alternative decomposition routine is desired.

### 4.4. Solution by ADMM

Robustness is added to the solution via an augmented Lagrangian. The Alternating Direction Method of Multipliers (ADMM) as described in [16] is adopted, where $\beta$ is the communicated variable. This approach is also very similar to the Progressive Hedging method as described in [17]. Instead of updating the dual variable based on the difference between the copies of each element in $\beta$ as done in (13), the update is with respect to the average value,

$$
\omega_{\mathcal{E}_{ij}}^k[t+1] := \omega_{\mathcal{E}_{ij}}^k[t] + \rho[t] (\beta_{\mathcal{E}_{ij}}^i[t] - \beta_{\mathcal{E}_{ij}}^j[t]) \quad \forall \mathcal{E}_{ij}, k \in i, j,
$$

where,

$$
\beta_{\mathcal{E}_{ij}}^k[t] = \frac{\beta_{\mathcal{E}_{ij}}^i[t] + \beta_{\mathcal{E}_{ij}}^j[t]}{2}.
$$

Note that the dual variable $\omega$ is now local to each zone, as opposed to $\nu$, which is global.

In addition to the slightly altered dual update, the augmented Lagrangian adds a quadratic term to the objective, and the new the problem in each zone, $k$ becomes:

$$
\begin{align*}
\text{Minimize} & \quad f(s_k, \beta_k^+, \beta_k^-) + \sum_{j \in \mathcal{H}} (\omega_{\mathcal{E}_{kj}}^k)^T \beta_{\mathcal{E}_{kj}}^k \\
& \quad + \sum_{i \in \mathcal{H}} (\omega_{\mathcal{E}_{ik}}^k)^T \beta_{\mathcal{E}_{ik}}^k + \frac{\rho}{2} \| \beta^k - \bar{\beta}_{\mathcal{E}_{ik}} \|^2 \\
\text{Subject to} & \quad \text{constraints in (1), (4), and (5) for zone } k \text{ only.}
\end{align*}
$$

The main advantage of the ADMM decomposition over the simpler and linear Lagrange relaxation, is that a penalty on a specific distance from a value is possible, which helps prevent oscillations. To achieve the same results without resorting to a quadratic term, constraints similar to (5) are necessary, which alter both of the feasible region, as well as the objective, and introduce more variables.
4.5. Forcing Consensus

Due to the non-convex nature of the problem, the zones are not likely to fully converge to a consensus on \( \beta \). This discrepancy can be significant enough that the final DC powerflow will violate some of the initial constraints, specifically, the flow or angle difference limits.

To resolve this issue, a final iteration is performed where \( P \) and \( Z \) are fixed, meaning that the power injections and susceptances are now strictly parameters of the problem. The goal of the forced consensus is to alter the the power injection vector as little as possible, while still satisfying all the constraints. This is achieved by a quadratic objective in (17). Since the generation and load are now handled individually, instead of parameter \( p \) we have the nodal generation \( P_d \) and nodal consumption \( P_d \). We also use sets \( \mathcal{G} \) and \( \mathcal{D} \) to describe the non-zero entries of \( P_d \) and \( P_d \) respectively.

\[
\text{Minimize} \quad \sum_{n \in \mathcal{G}} (\alpha_n - 1)^2 + \sum_{n \in \mathcal{D}} (\alpha_n - 1)^2 \quad (17a)
\]

Subject to
\[
f_l + b_l (\theta_l - \theta_t) = 0 \quad \forall \ell \in \mathcal{L} \quad (17b)
\]
\[
(\alpha_n P_{d,n} - \alpha_{d,n} P_{d,n}) + \sum_{\ell \in \mathcal{I}_n} f_{\ell} = 0 \quad \forall n \in \mathcal{N} \quad (17c)
\]
\[
- \sum_{\ell \in \mathcal{I}_n} f_{\ell} = 0 \quad \forall n \in \mathcal{N}
\]
\[
- f_{\ell} \leq f_{\ell} \leq f_{\ell}^\max \quad \forall \ell \in \mathcal{L} \quad (17d)
\]
\[
- \delta_{\ell} \leq \theta_f - \theta_t \leq \delta_{\ell} \quad \forall (f, t) \in \mathcal{L} \quad (17e)
\]
\[
P_d \min \leq \alpha_n P_{d,n} \leq P_{d,\max} \quad \forall n \in \mathcal{G} \quad (17f)
\]
\[
P_d \min \leq \alpha_{d,n} P_{d,n} \leq P_{d,\max} \quad \forall n \in \mathcal{D} \quad (17g)
\]
\[
\alpha_{d,n} P_{d,n} \leq \alpha_n P_{d,n} \quad \forall n \in \mathcal{G} \cap \mathcal{D} : P_{d,n} < P_{g,n} \quad (17h)
\]
\[
\alpha_{d,n} P_{d,n} \geq \alpha_n P_{g,n} \quad \forall n \in \mathcal{G} \cap \mathcal{D} : P_{d,n} > P_{g,n} \quad (17i)
\]
\[
\alpha_n, \alpha_{d,n} \geq 0 \quad (17j)
\]

Constraints (17b) and (17d) are the DC powerflow formulation where the injection is explicitly shown as the difference between load and generation. To avoid unreasonable limits, numbers are placed on the both load and generation by constraints (17f) and (17g). Finally, constraints (17i) and (17j) ensure that the relative relation between load and generation on a bus that both does not change.

5. Results

Our procedure can be paired with a topology generation algorithm but can also be used simply to complete and/or resuffle a real case. For testing, the Polish2383wp case available in Matpower [18] is used as a reference case. In one test, the data from the case is simply reshuffled, while in a second one distributions are extracted and used to generate input sets that are placed on an RT-nested-SmallWorld [3] synthetic topology. Implementation is done using the Gurobi solver with Python API. The Polish2383wp case is chosen because it is publicly available, represents a real system, and is sufficiently large that decomposition is absolutely necessary. With an average degree of 2.43, the full problem has over 14 million binary variables.

5.1. Input Data

As stated in Section 3, an edge list, as well as a set of power injections, and line susceptances are needed as inputs to the algorithm. We do not discuss appropriate ways to generate topologies but rather assume that the provided topology is satisfactory.

For the sake of comparison, the distributions of generation, load, and reactance should be similar to the reference case. This is achieved by fitting the empirical cumulative distribution functions from the reference case with a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) [19]. Since this interpolation method is monotone it provides a valid cumulative distribution. At the same time, the result is smoother than an empirical histogram.

In addition to the distributions the following quantities are necessary:
\[
N_{\text{int}} \quad \text{The number of intermediate buses with neither load nor generation.}
\]
\[
N_g \quad \text{Number of buses with generation.}
\]
\[
N_{g \text{gen only}} \quad \text{Number of buses with only generation.}
\]
\[
N_{P_d < P_g} \quad \text{Number of buses with generation and load where generation is greater.}
\]

These can be calculated from fractions extracted from the reference case and the total number of buses in the cases that is being constructed. The procedure for sampling generation \( P_d \) and load \( P_d \) (and therefore power injection) data is as follows:

1. \( N_{\text{int}} \) entries are set to zero for both \( P_d \) and \( P_d \).
2. \( N_{\text{load only}} = |\mathcal{N} - (N_{\text{int}} + N_g) \) load only buses are sampled for the load distribution, corresponding entries in \( P_d \) are zero.
3. \( N_{g \text{gen only}} \) buses are sampled from the generation distribution, corresponding entries in \( P_d \) are zero.
4. For each of the \( N_{P_d < P_g} \) entries, generation is sampled first. Then load samples are drawn until one is found that is smaller than the sampled generation.
5. For each of the \( N_{P_d > P_g} = N_g - (N_{g \text{gen only}} + N_{P_d < P_g}) \) entries, load is sampled first. Then generation samples are drawn until the inequality is satisfied.

The load and generation sampled from their respective distributions will likely not equal. For the simple DC-powerflow neglecting losses, this should be remedied. This

is done using the optimization in (18), which is very similar to (17).

\[
\text{Minimize} \quad \sum_{n \in G} \frac{P_{g,n}}{\max(P_g)} (\alpha_{g,n} - 1)^2 + \sum_{n \in G} \frac{P_{d,n}}{\max(P_d)} (\alpha_{d,n} - 1)^2 \quad (18a)
\]

Subject to

\[
\sum_{n \in G} \alpha_{g,n} P_{g,n} - \sum_{n \in G} \alpha_{d,n} P_{d,n} = 0 \quad (18b)
\]

\[
P_{g,n}^{\text{min}} \leq \alpha_{g,n} P_{g,n} \leq P_{g,n}^{\text{max}} \quad \forall n \in G \quad (18c)
\]

\[
P_{d,n}^{\text{min}} \leq \alpha_{d,n} P_{d,n} \leq P_{d,n}^{\text{max}} \quad \forall n \in G \quad (18d)
\]

\[
\alpha_{d,n} P_{d,n} \leq \alpha_{g,n} P_{g,n} \quad \forall n \in G \cap \mathcal{D} : P_{d,n} < P_{g,n} \quad (18e)
\]

\[
\alpha_{g,n} P_{g,n} \geq \alpha_{d,n} P_{d,n} \quad \forall n \in G \cap \mathcal{D} : P_{d,n} > P_{g,n} \quad (18f)
\]

\[
\alpha_{g,n} \geq 0 \quad (18g)
\]

The optimization finds multipliers for each entry of \(P_g\) and \(P_d\) such that the total load and generation will be equal. The objective attempts to keep these multipliers as close to one as possible, thus preserving the shape of the sampled distribution. Additionally, each term in the objective is weighted with respect to the largest generation or load, which rewards the program for preserving the tails, i.e., prevent it from altering the few large generators or loads too strongly.

Figure 5 shows the distributions of the real case, the original drawn samples, and final input samples to the algorithm following optimization (18). It is evident that the samples are fairly similar in distribution, and are furthermore not greatly affected by the modification.

5.2. Motivation for Optimization: Random Assignment

To motivate the need for applying (1) (or its subsequent decomposition), we first present what happens if the assignment is performed uniformly at random. Power injections and impedance are sampled as described in Section 5.1 and assigned by picking uniformly and independently a bus (for load and generation) or a line (for admittances) from the 3000 node RT-nested-smallWorld topology [3]. The process is repeated 400 times, and for each the maximum angle difference and line flow is recorded. Figure 6 shows the distributions of these maximum values. Note that some of the maximum angle differences in Figure 6a are greater than 90°, which is non-sensical from a physical perspective. The frequent occurrence of unacceptably large extreme values reinforces the idea that some form of optimization is necessary to meet the requirements of a realistic operating point.

5.3. Experiment Setup

Two experiments are considered in comparison to the reference case:
5.4. DC Powerflow Results

The results are evaluated with respect to the values associated with a DC powerflow, namely, the angle differences between adjacent buses and the real power flows. Distributions for these quantities are shown in Figure 8, which provide a strong visual indication that there is good agreement. The majority of the weight in both distributions is centered around zero, however, they exhibit heavy tails, meaning that there are non-negligible extreme values. Table 1 reports these extreme values for the three cases in Figure 8.

Table 1 also shows the result for the RT-nested-SmallWorld case before the forced consensus step in (17) is performed. As the maximum allowable flow was set to 1000 MW in all the problems, there is clearly a violation. The highlights the need for the final forced consensus step. Additionally, it confirms that simple random assignment is not likely to produce satisfactory results. Since the final results meet all the constraints, while intermediate ones do not, we can conclude that the optimization is both necessary and fulfills its function of assigning the parameters in a way that leads to a viable and reasonable DC powerflow result.

6. Discussion

The results presented in the previous section indicate that the proposed algorithm successfully achieves its goal. A closer examination reveals a few issues that merit some further discussion. Table 2 reports the termination criteria values at the end of the optimization as well as run times. Both optimization terminated after the iteration limit of five iterations was reached.

Scalability could be rightfully questioned if for ~3000 nodes the run-time is over five hours. Additionally, a motivation for automating case generation, is to produce multiple
examples, run multiple simulations, and therefore achieve more statistically meaningful results. That is, run Monte-Carlo simulations varying not only parameters but actual systems. To achieve this, a large number of cases need to be generated. If each case requires multiple hours, the time requirement for a hundred could become too onerous. Time is important, therefore, not from an operation critical perspective, but rather in term of user acceptance.

Gurobi’s root relaxation, performed to find a lower bound prior to the branch and bound process is the main culprit for the long computation times. The combination of a quadratic objective and disjunctive constraints (1d) and (1e) lead to a poor and slow relaxation. In combination with the time limit imposed on the optimization, this has the result that larger zones often do not complete the root-relaxation, no new solution is found, and consequently no improvement towards convergence is made.

The observed maximum error is likely a result of these computational issues. When larger errors exist between zones that are incapable of solving their root-relaxation, they cannot improve. An obvious solution is to remove completely or extend the time limit. However, this comes at the cost of longer run-times, which are already longer than desired. Potential improvements in this area are left for future work.

7. Conclusion and Future Work

To fully unlock the potential of synthetic power system cases, creating new cases should be fast enough that preparing several cases is not a daunting task. This paper presented an optimization approach to assigning known power injections and line susceptances to a topology and achieve realistic DC powerflow results. Results on a moderately large sample validate the approach. The solution is only possible by decomposing the problem into smaller zones and iteratively solving using the ADMM method. While the solutions are satisfactory, future work will explore ways to speed up the process to more reasonable times for potential users.

One possibility that will be considered in future work is to relax the binary constraints entirely and replace them with a combination of orthonormal constraints on II and Z, as well as heuristics to project the results back onto the space of permutation matrices. The resulting constraints cease to be linear but perhaps a more complex but continuous feasible region will be more amenable to faster solutions. In parallel, additional elements will be added to the procedure so that more complete cases can be created that also directly address the AC powerflow.

8. Acknowledgements

This work was funded in part by the Advanced Research Projects Agency- Energy (ARPA-E), U.S. Department of Energy, under Award Number DE-AR0000714. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References


