Placing Rotational Inertia in Power Grids

Bala Kameshwar Poolla Saverio Bolognani Florian Dörfler

***Abstract*— A major transition in the operation of electric power grids is the replacement of bulk generation based on synchronous machines by distributed generation based on low-inertia power electronics sources. The accompanying "loss of rotational inertia" and fluctuations by renewable sources jeopardize the system stability, as testified by the ever-growing number of frequency incidents. As a remedy, numerous studies demonstrate how virtual inertia can be emulated through various devices, but few of them address the question of "where" to place this inertia. It is however strongly believed that the placement of virtual inertia hugely impacts system efficiency, as demonstrated by recent case studies. In this article, we carry out a comprehensive analysis in an attempt to address the optimal inertia placement problem. We consider a linear network-reduced power system model along with an $\mathcal{H}_2$ performance metric accounting for the network coherency. The optimal inertia placement problem turns out to be non-convex, yet we provide a set of closed-form global optimality results for particular problem instances as well as a computational approach resulting in locally optimal solutions. We illustrate our results with a three-region power grid case study and compare our locally optimal solution with different placement heuristics.**

I. Introduction

As we retire more and more synchronous machines and replace them by renewable sources interfaced with power electronic devices, the stability of the power grid is jeopardized, which has been recognized as one of the prime concerns by transmission system operators [1]. Both in transmission grids as well as in islanded microgrids, low inertia levels together with uncertain renewable generation lead to large frequency swings.

Not only low levels of inertia are troublesome, but particularly spatially heterogeneous and time-varying inertia profiles can lead to destabilizing effects, as shown in an interesting two-area case study [2]. It is not surprising that rotational inertia has been recognized as a key ancillary service for power system stability, and a plethora of mechanisms have been proposed for the emulation of virtual (or synthetic) inertia [3]–[5] through a variety of devices (ranging from wind turbine control [6] to batteries [7] and flywheels), as well as inertia monitoring schemes [8] and even inertia markets [9]. In this article, we pursue the questions raised in [2] regarding the detrimental effects of spatially heterogeneous inertia profiles, and how they can be alleviated by virtual inertia emulation throughout the grid. In particular, we are interested in the allocation problem "where to place the inertia?" to optimize the system coherency.

The problem of inertia allocation has been hinted at before [2], but we are aware only of the study [10] explicitly addressing the problem. In [10], the grid is modeled by the linearized swing equations, and eigenvalue damping ratios as well as transient overshoots (estimated from the system modes) are chosen as optimization criterion for placing virtual inertia and damping. The resulting problem is highly non-convex, but a sequence of approximations led to some insightful results. In comparison to [10], we focus on network coherency as an alternative performance metric, that is, the amplification of stochastic or impulsive disturbances to a quadratic performance index measured by the $\mathcal{H}_2$ norm [11]. As performance index, we choose a classic coherency criterion penalizing angular differences and absolute frequencies, which has recently been popularized for consensus and synchronization studies [12]–[17] as well as in power system analysis and control [18]–[20]. We feel that this $\mathcal{H}_2$ performance metric is not only more tractable than spectral metrics, but it is also very meaningful for the problem at hand: it measures the effect stochastic fluctuations (caused by loads or renewable generation) as well as discrete events (such as faults or deterministic frequency errors caused by markets) and quantifies their amplification by a coherency index related to frequency volatility. Finally, in comparison to [10], the damping or droop coefficients are not decision variables in our problem setup, since these are determined by the system physics (in case of damping), the outcome of primary reserve markets (in case of primary control), or scheduled according to cost coefficients or ratings [21].

The contributions of this paper are as follows. We provide a comprehensive modeling and analysis framework for the inertia placement problem in power grids to optimize an $\mathcal{H}_2$ coherency index subject to capacity and budget constraints. The optimal inertia placement problem turns out to be non-convex, yet we are able to provide upper and lower bounds on the performance index for general networks as well as a set of closed-form global optimality results for particular problem instances characterized by uniform disturbance over damping ratios and for two-area networks. Moreover, we develop a computational approach based on a gradient formula that allows to find locally optimal solutions for large networks and arbitrary parameter settings. Finally, we illustrate our results with a three-region case study. Our analytic and computational results paint a surprisingly complex picture: they generally show that inertia is placed dominantly at the sites of disturbances yet no other nodes are normally neglected. However, naive solutions such as uniform allocation

This material is supported by ETH start-up funds and SNF Assistant Professor Energy Grant #160573. B.K. Poolla, S. Bolognani, and F. Dörfler are with the Automatic Control Laboratory at the Swiss Federal Institute of Technology (ETH) Zürich, Switzerland. Emails: {bpoolla,bsaverio,dorfler}@ethz.ch.

or placement at maximum capacity deliver very suboptimal results.

The remainder of this section introduces some notation. Section II motivates our system model and the coherency performance index. Section III presents inertial allocation results for general networks and provide explicit results for the two-area networks. Section IV presents a case study on a three-region network. Finally, Section V concludes the paper.

Preliminaries and Notation

Let $\mathbb{1}_n$ and $\mathbb{0}_n$ be the n-dimensional vectors of all ones and zeros. Given an index set $\mathcal{I}$ with cardinality $|\mathcal{I}|$ and a real valued array $\{x_1 \dots x_{|\mathcal{I}|}\}$, we denote by x the vector in $\mathbb{R}^{|\mathcal{I}|}$ obtained by stacking the scalars x_i, and by $\mathrm{diag}\{x_i\}$ the associated diagonal matrix. The vector $\mathbb{e}_i$ is the i-th vector of the canonical basis for $\mathbb{R}^n$.

II. Problem Formulation

A. System Model

Consider a power network modeled by a graph with nodes (buses) $\mathcal{V} = \{1, \dots, n\}$ and edges (transmission lines) $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. We consider a small-signal version of a network-reduced power system model [22], [23], where passive loads are eliminated via Kron reduction [24], and the network is reduced to the sources $i \in \{1,\dots, n\}$ with dynamics

$$m_i \ddot{\theta}_i + d_i \dot{\theta}_i = p_{\text{in},i} - p_{e,i}\,, \quad i \in \{1, \dots, n\}\,, \tag{1}$$

where $p_{\text{in},i}$ and $p_{e,i}$ refer to the power input and electrical power output, respectively. If bus i is a synchronous machine, then (1) describes the electromechanical swing dynamics for the generator rotor angle θ_i [22], [23], and the coefficients $m_i > 0$ and $d_i > 0$ account for the generator's rotational inertia and frequency damping and speed droop control coefficient. If bus i connects to a renewable or battery source interfaced with a power electronics inverter operated in grid-forming mode [25], [26], then θ_i is the voltage phase angle, $d_i > 0$ is the droop control coefficient and $m_i > 0$ accounts for power measurement time constants [27] or arises from virtual inertia emulation [3]–[5]. Finally, the dynamics (1) may also arise from frequency-dependent or actively controlled frequency-responsive loads [23].

Under the assumptions of constant voltage magnitudes, purely inductive lines, and a small signal approximation, the electrical power output at the terminals is given by [23]:

$$p_{e,i} = \sum_{k=1}^{n} b_{ik}(\theta_i - \theta_k), \quad i \in \{1, \dots, n\}\,, \tag{2}$$

where $b_{ik} \geq 0$ is the susceptance between nodes $\{i, k\} \in \mathcal{E}$.

The state space representation of the system (1)-(2) is then

$$\begin{bmatrix} \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -M^{-1}L & -M^{-1}D \end{bmatrix} \begin{bmatrix} \theta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ M^{-1} \end{bmatrix} p_{\text{in}}\,, \tag{3}$$

where $M = \mathrm{diag}\{m_i\}$ and $D = \mathrm{diag}\{d_i\}$ are the diagonal matrices of inertial and damping/droop coefficients, and $L = L^T \in \mathbb{R}^{n\times n}$ is the network Laplacian (or susceptance) matrix with off-diagonal elements $l_{ik} = -b_{ik}$ and diagonals $l_{ii} = \sum_{k=1, k\neq i}^{n} b_{ik}$. The states $(\theta, \omega) \in \mathbb{R}^{2n}$ are the stacked vectors of angles and frequencies and $p_{\text{in}} \in \mathbb{R}^n$ is the net power input – all of which are deviation variables.

B. Coherency performance metric

We consider the linear power system model (3) driven by the inputs $p_{\text{in},i}$ accounting either for faults or non-zero initial values (modeled as impulses) or for fluctuations in renewables and loads. We are interested in the energy expended in returning to the steady-state configuration, expressed as a quadratic cost of the angles and frequency displacements:

$$\frac{1}{2}\int_0^\infty \sum\nolimits_{i,j=1}^{n} \alpha_{ij}(\theta_i(t) - \theta_j(t))^2 + \sum\nolimits_{i=1}^{n} \beta_i \omega_i^2(t)\, dt\,. \tag{4}$$

Here, $\beta_i > 0$ are positive scalars for $i \in \mathcal{V}$, and we assume that the nonnegative scalars $\alpha_{ij} = \alpha_{ji} \geq 0$ induce a connected graph – not necessarily identical with the grid.

Following the interpretation proposed in [18], the above metric (4) can represent a generalized energy in synchronous machines. Indeed, for $\alpha_{ij} = g_{ij}$ (where g_{ij} are the power line conductances) and $\beta_i = m_i$, the metric (4) accounts for the heat losses in the grid lines and the mechanical energy losses in the generators. This case is often referred to as the short-range or local synchronization error in the coherency literature [12], [13] whereas the case $\alpha_{ik} = 1/n$ is the long-range or global synchronization error. Aside from consensus and synchronization studies [12]–[17] the coherency metric (4) is also used in power system analysis and control [18]–[20]. Adopting the state representation introduced in (3), the performance metric (4) can be rewritten as the time-integral $\frac{1}{2}\int_0^\infty y(t)^T y(t) dt$ of the performance output

$$y = \underbrace{\begin{bmatrix} N^{\frac{1}{2}} & 0 \\ 0 & S^{\frac{1}{2}} \end{bmatrix}}_{=C} \begin{bmatrix} \theta \\ \omega \end{bmatrix}\,, \tag{5}$$

where N is the Laplacian matrix of the graph induced by the α_{ij} and S is a diagonal matrix with elements $S_{ii} = \beta_i$. In this notation, the short-range and long-range error corresponds to $N = L$ and $N = \Pi = (I_n - \mathbb{1}_n \mathbb{1}_n^T/n)$, respectively.

In order to model the localization of the disturbances in the grid, we parametrize the input p_{in} as

$$p_{\text{in}} = T^{\frac{1}{2}}\eta, \quad T = \mathrm{diag}\{t_i\}.$$

We therefore obtain the state space model

$$\begin{bmatrix} \dot{\theta} \\ \dot{\omega} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & I \\ -M^{-1}L & -M^{-1}D \end{bmatrix}}_{=A} \begin{bmatrix} \theta \\ \omega \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ M^{-1}T^{1/2} \end{bmatrix}}_{=B} \eta. \tag{6}$$

In the following, we refer to the input/output map (5), (6) as $G = (A, B, C)$. If the inputs η_i are Dirac impulses, then (4) measures the squared $\mathcal{H}_2$ norm [11] of the system G, that we denote as $\|G\|_2$.

There is a number of interpretations of the $\mathcal{H}_2$ norm $\|G\|_2$ of a power system [18]. The relevant ones in our context are:

1) The squared $\mathcal{H}_2$ norm of G measures the energy amplification, i.e., the sum of $\mathcal{L}_2$ norms of the outputs $y_i(t)$, for unit impulses at all inputs $\eta_i(t)=\delta(t)$:
$$\|G\|_2^2 = \sum\nolimits_{i=1}^{n} \int_0^\infty y_i^T(t)\, y_i(t) dt.$$
These impulses can model faults or initial conditions.

2) The squared $\mathcal{H}_2$ norm of G quantifies the steady-state total variance of the output for a system subjected to unit variance stochastic white noise inputs $\eta_i(t)$:
$$\|G\|_2^2 = \lim_{t\to\infty} E\{y^T(t)\, y(t)\}.$$
In our case, the white noise inputs can model stochastic fluctuations of renewable generation or loads.

For a standard (A, B, C) state-space system, we have that
$$\|G\|_2^2 = \text{Trace}(B^T P B)\,, \tag{7}$$
where $P=\int_0^\infty e^{A^T t} C^T C e^{At}\, dt$ is the observability Gramian obtained as solution to the Lyapunov equation
$$PA + A^T P + C^T C = 0\,. \tag{8}$$
Recall that a necessary and sufficient condition for the existence of a positive definite observability Gramian P is that the pair (A, C) is detectable [11]. In our case, we have
$$C \begin{bmatrix} \mathbb{1}_n \\ \mathbb{0}_n \end{bmatrix} = A \begin{bmatrix} \mathbb{1}_n \\ \mathbb{0}_n \end{bmatrix} = \begin{bmatrix} \mathbb{0}_n \\ \mathbb{0}_n \end{bmatrix},$$
that is, the mode $\mathbb{1}_n$ (corresponding to the absolute angle) is not observable. This follows from the shared zero eigenvalue of the Laplacians N and L. On the other hand, the $\mathcal{H}_2$ norm is finite since the mode $[\mathbb{1}_n^T\ \mathbb{0}_n^T]$ is marginally stable and not externally excited; see [14, Lemma 1]. A multiplication of the Lyapunov equation (8) by $\begin{bmatrix}\mathbb{1}_n^T\ \mathbb{0}_n^T\end{bmatrix}$ also yields that $\begin{bmatrix}\mathbb{1}_n^T\ \mathbb{0}_n^T\end{bmatrix} P = \begin{bmatrix}\mathbb{0}_n^T\ \mathbb{0}_n^T\end{bmatrix}$, that is, P is positive semidefinite and the $\mathcal{H}_2$ norm does not account for the absolute angle.

C. Inertia allocation formulation

We assume that each node $i \in \{1, \dots, n\}$ has a nonzero[1] inertial coefficient ($\underline{m_i} > 0$) and we are interested in optimally allocating additional virtual inertia in order to minimize the $\mathcal{H}_2$ norm (4), subject to upper bounds $\overline{m_i}$ at each bus, and total budget constraint m_{bdg}, accounting for the available installation space and the total cost of the storage devices, respectively.

This problem statement is summarized as

$$\begin{aligned}
\underset{P\,,\,m_i}{\text{minimize}} \quad & \|G\|_2^2 = \text{Trace}\left(B^T P B\right) && \text{(9a)}\\
\text{subject to} \quad & \sum\nolimits_{i=1}^{n} m_i \le m_{\text{bdg}} && \text{(9b)}\\
& \underline{m_i} \le m_i \le \overline{m_i}\,, \quad i \in \{1, \dots, n\} && \text{(9c)}\\
& PA + A^T P + C^T C = 0\,, && \text{(9d)}
\end{aligned}$$

where the matrices (A, B, C) arise from the input-output system (5)-(6). Observe the bilinear nature of the constraint equation (9d) featuring products of A and P, and recall from (6) that the optimization variables m_i appear as m_i^{-1} in A. Hence, the problem (9) is highly non-convex and typically also large-scale.

[1]Observe that the case $m_i = 0$ leads to an ill-posed model (1) whose number of algebraic and dynamic state depends on the system parameters.

III. Optimal inertia allocation

In the following, we investigate the optimization problem (9). In particular, we will provide general lower and upper bounds, closed-form results under certain parametric assumptions, general results for a two-area power system, and a gradient formula to determine locally optimal solutions.

A. Analytic closed-form results in the general case

To begin we consider the coordinate transformations $\hat\theta = M^{\frac12}\theta$ and $\hat\omega = M^{\frac12}\omega$ for the sake of symmetrizing the system matrix A in (6). The dynamics of the input-output system defined in (5)-(6) in the transformed variables are then
$$\begin{bmatrix} \dot{\hat\theta} \\ \dot{\hat\omega} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & I \\ -\hat L & -\hat D \end{bmatrix}}_{=\hat A} \begin{bmatrix} \hat\theta \\ \hat\omega \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ M^{-\frac12} T^{\frac12} \end{bmatrix}}_{=\hat B} \eta\,,$$
where $\hat L = M^{-\frac12} L M^{-\frac12}$ and $\hat D = M^{-\frac12} D M^{-\frac12}$ are congruency transforms of L and D. In the transformed coordinates, the performance matrix $C^T C$ is rendered to
$$\hat C^T \hat C = \begin{bmatrix} M^{-\frac12} N M^{-\frac12} & 0 \\ 0 & M^{-\frac12} S M^{-\frac12} \end{bmatrix} = \begin{bmatrix} \hat N & 0 \\ 0 & \hat S \end{bmatrix}.$$
Finally, the Lyapunov equation (8) is rendered to
$$\hat P \begin{bmatrix} 0 & I \\ -\hat L & -\hat D \end{bmatrix} + \begin{bmatrix} 0 & -\hat L \\ I & -\hat D \end{bmatrix} \hat P + \begin{bmatrix} \hat N & 0 \\ 0 & \hat S \end{bmatrix} = 0 \tag{10}$$
where $\hat P$ is the observability Gramian parametrized as
$$\hat P = \hat P^T = \begin{bmatrix} X_1 & X_0 \\ X_0^T & X_2 \end{bmatrix}.$$
In the transformed system, the squared $\mathcal{H}_2$ norm (7) is
$$\begin{aligned}
\|G\|_2^2 &= \text{Trace}(\hat B^T \hat P \hat B) = \text{Trace}(T M^{-1/2} X_2 M^{-1/2}) \\
&= \text{Trace}(T M^{-1} X_2) = \sum\nolimits_{i=1}^{n} \frac{t_i X_{2,ii}}{m_i}\,,
\end{aligned} \tag{11}$$
where we used the ring commutativity of the trace. After multiplying the (1,1) equation of the matrix Lyapunov equation (10) by the Moore-Penrose pseudo-inverse $L^\dagger$ of the Laplacian L, we obtain the (1,1) equation and the (2,2) equations as
$$\text{Trace}(X_0 \hat L \hat L^\dagger) + \text{Trace}(X_0^T \hat L^\dagger \hat L) = \text{Trace}(\hat N \hat L^\dagger)\,, \tag{12a}$$
$$\text{Trace}(X_0 + X_0^T) = -\text{Trace}(\hat S) + 2 \cdot \text{Trace}(D M^{-1} X_2)\,, \tag{12b}$$
where we used the fact that D and M are diagonal and thus commutative matrices in (12b). Recall that $L L^\dagger = \Pi = (I_n - \mathbb{1}_n \mathbb{1}_n^T/n)$ and $\begin{bmatrix}\mathbb{1}_n^T\ \mathbb{0}_n^T\end{bmatrix} P = \begin{bmatrix}\mathbb{0}_n^T\ \mathbb{0}_n^T\end{bmatrix}$. Accordingly, $\hat L \hat L^\dagger = M^{-1/2} \Pi M^{1/2}$ and $\begin{bmatrix}\mathbb{1}_n^T M^{1/2}\ \mathbb{0}_n^T\end{bmatrix} \hat P = \begin{bmatrix}\mathbb{0}_n^T\ \mathbb{0}_n^T\end{bmatrix}$. Equivalently ${X_0}^T M^{1/2} \mathbb{1}_n = \mathbb{0}_n$ and $X_0 \hat L \hat L^\dagger = X_0 M^{-1/2} (I_n - \mathbb{1}_n \mathbb{1}_n^T/n) M^{1/2} = X_0$. Thus, (12a) reduces to
$$\text{Trace}(X_0 + X_0^T) = \text{Trace}(\hat N \hat L^\dagger) = \text{Trace}(N L^\dagger)\,,$$

which in combination with (12b) delivers

$$\text{Trace}(DM^{-1}X_2) = \frac{1}{2}\left(\text{Trace}(M^{-1}S + NL^\dagger)\right). \quad (13)$$

Observe the astonishing similarities between the squared $\mathcal{H}_2$ norm (11) and the constraint equation (13). These relations allow us to state general upper and lower bounds on the problem (9). Furthermore, we can obtain closed-form results for the squared $\mathcal{H}_2$ norm under the following two assumptions:

*Assumption 1 (***Uniform disturbance-damping ratio):** *The ratio t_i/d_i is constant for all $i \in \{1, \dots, n\}$.* □

*Assumption 2 (***Inertia-proportional penalty):** *The frequency penalty matrix S in* (5) *is proportional to the inertia allocation at each node: $S = c \cdot M$ for $c \in \mathbb{R} \geq 0$.*□

Regarding Assumption 1, the droop coefficients d_i are typically scheduled proportionally to the rating of a power source to guarantee fair power sharing [21]. On the other hand, it is not unreasonable to expect the disturbances to scale proportionally to size of a power source. Hence, Assumption 1 can very well be justified. In particular, it holds for uniform damping and disturbances – though it may be restrictive in other scenarios. Regarding Assumption 2, the penalty on frequency with coefficients proportional to additional inertia is akin to penalizing kinetic energy – a reasonable and standard penalty in power systems. Assumption 2 also applies in case of no frequency penalty: $c = 0$. Note that Assumptions 1 and 2 are considered only here. The latter results are independent of these assumptions.

For now we can state the following result:

Theorem 3.1: **(Properties & bounds on performance index)** *Consider the power system model* (5)-(6)*, the squared $\mathcal{H}_2$ norm* (7) *depicting the energy expended to regain synchrony, and the optimal inertia allocation problem* (9)*.*

The following statements hold:

i) *There exists a constant lower (respectively, upper) bound on the $\mathcal{H}_2$ norm $\|G\|_2^2$ such that*

$$\underline{\Psi} W \leq \|G\|_2^2 \leq \overline{\Psi} W\,,$$

where $W = 0.5\left(\text{Trace}(\hat{S} + NL^\dagger)\right)$, $\underline{\Psi} = \dfrac{\min_i t_i}{\max_i d_i}$, $\overline{\Psi} = \dfrac{\max_i t_i}{\min_i d_i}$, i.e., the objective (9a) takes bounded values.

ii) *Under Assumption 1, problem (9) is equivalent to*

$$\underset{m_i}{\text{minimize}} \quad \|G\|_2^2 = \text{Trace}\left(M^{-1}S + NL^\dagger\right). \quad (14)$$

subject to the budget and capacity constraints (9b)*,* (9c)*.*

iii) *Under Assumptions 1 and 2, the $\mathcal{H}_2$ norm $\|G\|_2^2$ is independent of the inertia allocation, that is, any allocation of inertia results in the same energy:*

$$\|G\|_2^2 = \frac{1}{2}\left(n \cdot c + \text{Trace}(NL^\dagger)\right).$$

Theorem 3.1 states some possibly surprising results. Statement *i)* gives explicit upper and lower bounds on the $\mathcal{H}_2$ norm, which though informative may be conservative. Result *ii)* reduces the problem (9) to the simple convex problem (14) which is independent of the Laplacian L, that is, the optimal inertia allocation is *independent* of the network topology in this case. Finally, under the assumptions of statement *iii)*, the $\mathcal{H}_2$ norm is completely independent of the inertial coefficients.

Proof of Theorem 3.1:

i) Let $\underline{t} = \min_i\{t_i\}$, $\overline{t} = \max_i\{t_i\}$, $\underline{d} = \min_i\{d_i\}$, $\overline{d} = \max_i\{d_i\}$. From (11) we obtain the following relations

$$\underline{t}\sum_{i=1}^{n}\frac{X_{2,ii}}{m_i} \leq \|G\|_2^2 \leq \overline{t}\sum_{i=1}^{n}\frac{X_{2,ii}}{m_i}\,,$$

which can be further bounded as

$$\frac{\underline{t}}{\overline{d}}\sum_{i=1}^{n}\frac{d_i X_{2,ii}}{m_i} \leq \|G\|_2^2 \leq \frac{\overline{t}}{\underline{d}}\sum_{i=1}^{n}\frac{d_i X_{2,ii}}{m_i}\,. \quad (15)$$

On comparison with (13), (15) can be rewritten as:

$$\underline{\Psi} W \leq \|G\|_2^2 \leq \overline{\Psi} W\,,$$

where $\underline{\Psi}$, $\overline{\Psi}$, and W are as stated in the theorem.

ii) From Assumption 1, let $\lambda = t_i/d_i > 0$ be constant for all $i \in \{1, \dots, n\}$. Then we can rewrite (11) as

$$\text{Trace}(\hat{B}^T\hat{P}\hat{B}) = \sum_{i=1}^{n}\frac{t_i X_{2,ii}}{m_i} = \lambda \cdot \sum_{i=1}^{n}\frac{d_i X_{2,ii}}{m_i}\,. \quad (16)$$

As λ is a constant over the network and the right-hand side of (16) equals (13) up to the factor λ, the optimization problem (9) is equivalent to the following:

$$\underset{m_i}{\text{minimize}} \quad \frac{\lambda}{2}\cdot\left(\text{Trace}(M^{-1}S) + NL^\dagger\right) \quad (17a)$$

$$\text{subject to} \quad \sum_{i=1}^{n} m_i \leq m_{\text{bdg}} \quad (17b)$$

$$\underline{m_i} \leq m_i \leq \overline{m_i}\,, \quad i \in \{1, \dots, n\}\,. \quad (17c)$$

Observe that we removed the last redundant equation (9d) and the variable P since the objective value of (17) is independent of P. Finally, note that (14) and (17) are equivalent and differ by a constant positive factor $\lambda/2$.

iii) Under Assumption 2, $S = c \cdot M$ for some $c \in \mathbb{R} \geq 0$, the objective (14) is independent of M.

This concludes the proof. ■

B. Explicit results for a two-area network

Previously in Subsection III-A, we provided some results under the possibly restrictive Assumptions 1 and 2 for a general power network model. In this subsection, we focus on a two-area power grid as in [2] for illustrational purposes.

In the case of a two-area system governed by the dynamics in (1) and for the performance metric as defined in (7), it is possible to analytically solve the Lyapunov equation (8) and eliminate P. As a result, we obtain a closed form, but cumbersome expression for the objective (9a) as

$$f(m) = \text{Trace}[B(m)^T P(m) B(m)]$$

subject to the budget and capacity constraints (9b)-(9c), where $f(m)$ is a rational function of polynomials of orders 4 (respectively, 6) in terms of inertial coefficients m_i for the numerator (respectively, denominator). Since the objective becomes constant for an inertia-dependent frequency penalty $S = c \cdot M$, see Theorem 3.1, we focus the the discussion on constant frequency penalties. As the explicit expression is more convoluted than insightful, we will not show it here, but only report the following statements which can be verified by a simple but tedious analysis of the rational function $f(m)$:

1) The problem (9) admits a unique minimizer as a function of the t_i/d_i ratios and frequency penalty S.
2) For sufficiently large $\overline{m_i}$'s, the budget constraint (9b) becomes active, i.e., the optimizers satisfy $m_1^\star + m_2^\star = m_{\text{bdg}}$. In this case, m_2 can be eliminated as $m_2 = m_{\text{bdg}} - m_1$, and (9) can be reduced to a scalar problem.
3) In case of no capacity constraints, for identical t_i/d_i ratios and frequency penalties S, the optimal inertial coefficients are identical $m_1^\star = m_2^\star$. In the general case when $t_i/d_i > t_j/d_j$, then $m_i^* > m_j^*$ (see the example in Figure 1, where we eliminated $m_2 = m_{\text{bdg}} - m_1$).
4) For sufficiently uniform t_i/d_i ratios, the problem (9) is strongly convex. We observe that the cost function $f(m)$ is fairly flat over the feasible set (see Figure 1).
5) For strongly dissimilar t_i/d_i ratios, we observe a less flat cost function. If the disturbance affects only one node, e.g., $t_1 = 1$ and $t_2 = 0$, strong convexity is lost.

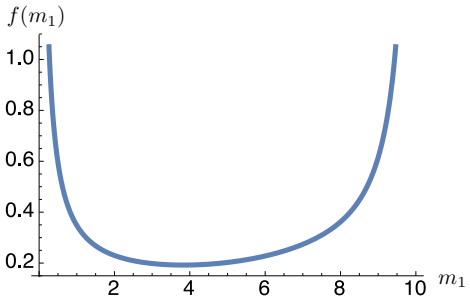

Fig. 1. Trace profile for weakly dissimilar $\frac{t_i}{d_i}$ ratios $\left(\frac{t_2}{d_2} > \frac{t_1}{d_1}\right)$.

From the above facts, we conclude that the input scaling factors t_i play an overriding role in the behaviour of the performance. Hence, some claims in Theorem 3.1 with uniform t_i/d_i ratios may showcase results for an idealistic scenario. To obtain a more complete picture, we linearly vary the disturbance input matrices from $[t_1, t_2] = [0, 1]$ to $[t_1, t_2] = [1, 0]$, that is, from a disturbance localized at node 2 to a disturbance localized at node 1. The resulting optimizers are displayed in Figure 2 showing that inertia is allocated dominantly at the site of the disturbance, which is in line with previous case studies [2], [10]. Notice also that depending on the value of the budget m_{bdg}, the capacity constraints $\overline{m_i}$, and the t_i/d_i ratios, the budget constraint may be active or not. Thus, perhaps surprisingly, sometimes not all inertia resources are allocated. Overall, the two-area

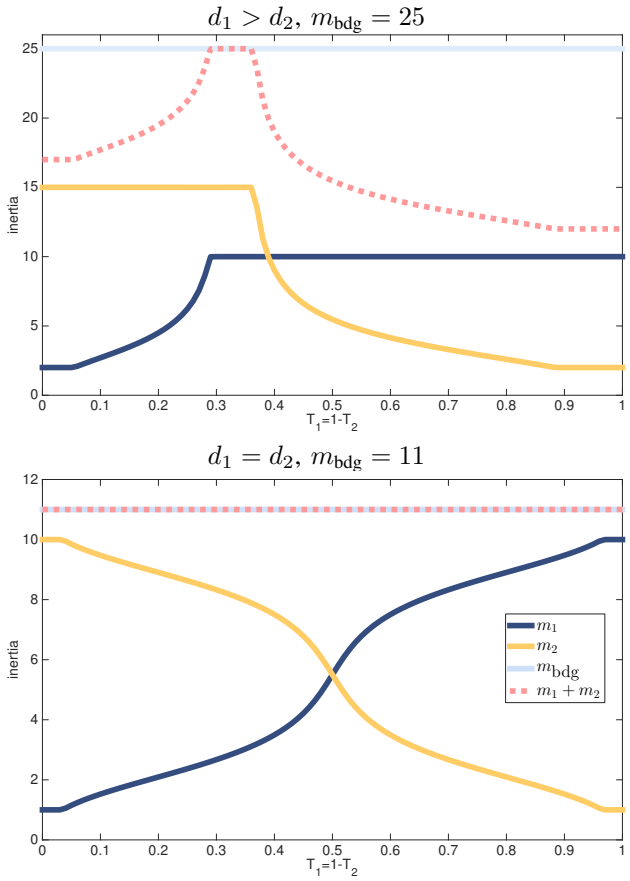

Fig. 2. Optimal inertia allocation for a two-area system with identical frequency penalties $S = I_2$, with non-identical and identical damping coefficients d_i, with disturbances inputs varying from $[t_1, t_2] = [0, 1]$ to $[t_1, t_2] = [1, 0]$, and for two choices of budget m_{bdg}.

case paints a surprisingly complex picture.

C. Perturbation analysis and gradient computation

In Subsections III-A and III-B, we considered a subset of networks that admitted a closed-form solution for the inertia allocation problem (9). Here, we present a computational approach and explicitly calculate the gradient $\nabla f(m)$ of the objective

$$f(m) = \text{Trace}\left(B(m)^T P(m) B(m)\right). \tag{18}$$

Observe that in writing (18) we implicitly assumed that $P = P(m)$ is obtained from the Lyapunov equation (9d). Information on the gradient $\nabla f(m)$ is essential for many computational approaches, for example, the problem (9) can be approached via the *partial* Lagrangian

$$L(m, \overline{\rho}, \underline{\rho}, \varsigma) = \text{Trace}\left(B(m)^T P(m) B(m)\right) + \overline{\rho}^T (m - \overline{m}) + \underline{\rho}^T (\underline{m} - m) + \varsigma \left(\sum_{i=1}^{n} m_i - m_{\text{bdg}}\right),$$

where $\overline{\rho}, \underline{\rho}, \varsigma > 0$ are multipliers, and we did not dualize the Lyapunov constraint (9d). If the gradient of $f(m)$ is explicitly available, then a locally optimal solution can be

computed, for example, via the iterative primal-dual algorithm [28]

$$\begin{aligned} m(k+1) &= m(k) \\ &\quad - \alpha(k)\left(\nabla f(m(k)) + \overline{\rho}(k) - \underline{\rho}(k) + \varsigma(k)\mathbb{1}_n\right), \\ \overline{\rho}(k+1) &= \left[\overline{\rho}(k) + \alpha(k)\left(m(k) - \overline{m}\right)\right]_+, \\ \underline{\rho}(k+1) &= \left[\underline{\rho}(k) + \alpha(k)\left(\underline{m} - m(k)\right)\right]_+, \\ \varsigma(k+1) &= \left[\varsigma(k) + \alpha(k)\left(\sum_{i=1}^{n} m_i(k) - m_{\text{bdg}}\right)\right]_+, \end{aligned}$$

where $k \in \mathbb{Z}_{\geq 0}$ is the iteration index, $\alpha(k) > 0$ is an appropriate step size, and $[\cdot]_+ = \max\{0, \cdot\}$. In general, most computational approaches can be sped up tremendously if the large-scale set of nonlinear (in the decision variables) Lyapunov equations (9d) can be eliminated and included into the gradient information. In the following, we provide an algorithm that achieves so, and uses the routine **Lyap**(A, Q) which returns the matrix P that solves $PA + A^T P + Q = 0$.

Algorithm 1: Gradient computation

Input current value m of the decision variables
Output numerical evaluation g of the gradient $\nabla f(m)$

$A^{(0)} \leftarrow \begin{bmatrix} 0 & I \\ -M^{-1}L & -M^{-1}D \end{bmatrix}$;
$B^{(0)} \leftarrow \begin{bmatrix} 0 \\ M^{-1}T^{1/2} \end{bmatrix}$;
$P^{(0)} \leftarrow$ **Lyap** $\left(A^{(0)}, C^T C\right)$;
for $i = 1, \ldots, n$ **do**
  $\Phi \leftarrow \mathbb{e}_i \mathbb{e}_i^T$;
  $A^{(1)} \leftarrow \begin{bmatrix} 0 & 0 \\ \Phi M^{-2}L & \Phi M^{-2}D \end{bmatrix}$;
  $B^{(1)} \leftarrow \begin{bmatrix} 0 \\ -\Phi M^{-2}T^{1/2} \end{bmatrix}$;
  $P^{(1)} \leftarrow$ **Lyap** $\left(A^{(0)}, P^{(0)}A^{(1)} + {A^{(1)}}^T P^{(0)}\right)$;
  $g_i \leftarrow$ **Trace** $\left(2{B^{(1)}}^T P^{(0)} B^{(0)} + {B^{(0)}}^T P^{(1)} B^{(0)}\right)$;

Theorem 3.2: Consider the objective function (18)*, where* $P(m)$ *is a function of* m *via the Lyapunov equation* (9d)*. The objective function is differentiable for* $m \in \mathbb{R}^n_{>0}$*, and its gradient at* m *is given by Algorithm 1.*

The proof of Theorem 3.2 is partially inspired by [15] and relies on a perturbation analysis of the Lyapunov equation (9d) combined with Taylor and power series expansions.

Proof of Theorem 3.2: In order to compute the gradient of (18) at $m \in \mathbb{R}^n_{>0}$, we make use of the relation

$$\nabla_\mu f(m) = \nabla f(m)^T \mu, \tag{19}$$

where $\nabla_\mu f(m)$ is the directional derivative of f in the direction $\mu \in \mathbb{R}^n$, defined as

$$\nabla_\mu f(m) = \lim_{\delta \to 0} \frac{f(m + \delta\mu) - f(m)}{\delta}, \tag{20}$$

whenever this limit exists.

From (18) we have that

$$f(m + \delta\mu) = \text{Trace}\left(B(m + \delta\mu)^T P B(m + \delta\mu)\right), \tag{21}$$

where P is a solution of the Lyapunov equation

$$PA(m + \delta\mu) + A(m + \delta\mu)^T P + C^T C = 0 \tag{22}$$

and where by $A(m + \delta\mu)$ we denote the system matrix defined in (6), evaluated in $m + \delta\mu$.

The matrices $A(m + \delta\mu)$ and $B(m + \delta\mu)$ can be intended as function of the scalar δ, and can thus be expanded in a Taylor series around $\delta = 0$ as

$$\begin{aligned} A(m + \delta\mu) &= A^{(0)}_{(m,\mu)} + A^{(1)}_{(m,\mu)}\delta + \mathcal{O}(\delta^2), \\ B(m + \delta\mu) &= B^{(0)}_{(m_0,\mu)} + B^{(1)}_{(m_0,\mu)}\delta + \mathcal{O}(\delta^2). \end{aligned} \tag{23}$$

In order to compute the coefficients of the Taylor expansion in (23), we recall the scalar series expansion of $1/(m_i + \delta\mu_i)$ around $\delta = 0$:

$$\frac{1}{(m_i + \delta\mu_i)} = \frac{1}{m_i} - \frac{\delta\mu_i}{m_i^2} + \mathcal{O}(\delta^2).$$

Using the shorthand $\Phi = \text{diag}(\mu_i)$, we therefore have

$$\begin{aligned} A^{(0)}_{(m,\mu)} &= \begin{bmatrix} 0 & I \\ -M^{-1}L & -M^{-1}D \end{bmatrix} \\ A^{(1)}_{(m,\mu)} &= \begin{bmatrix} 0 & 0 \\ \Phi M^{-2}L & \Phi M^{-2}D \end{bmatrix} \\ B^{(0)}_{(m,\mu)} &= \begin{bmatrix} 0 \\ M^{-1}T^{1/2} \end{bmatrix} \\ B^{(1)}_{(m,\mu)} &= \begin{bmatrix} 0 \\ -\Phi M^{-2}T^{1/2} \end{bmatrix}. \end{aligned}$$

Accordingly, the solution to the Lyapunov equation (22) can be exapanded in a power series as

$$P = P(m + \delta\mu) = P^{(0)}_{(m_0,\mu)} + P^{(1)}_{(m_0,\mu)}\delta + \mathcal{O}(\delta^2), \tag{24}$$

and therefore the Lyapunov equation (22) becomes

$$\begin{aligned} &(P^{(0)} + \delta P^{(1)} + \mathcal{O}(\delta^2))(A^{(0)} + \delta A^{(1)} + \mathcal{O}(\delta^2)) + \\ &(A^{(0)} + \delta A^{(1)} + \mathcal{O}(\delta^2))^T (P^{(0)} + \delta P^{(1)} + \mathcal{O}(\delta^2)) + C^T C = 0, \end{aligned}$$

where we dropped the subscript (m, μ) for readability. By collecting terms associated with powers of δ, we obtain two Lyapunov equations determining $P^{(0)}$ and $P^{(1)}$:

$$P^{(0)}A^{(0)} + {A^{(0)}}^T P^{(0)} + C^T C = 0 \tag{25a}$$

$$P^{(1)}A^{(0)} + {A^{(0)}}^T P^{(1)} + (P^{(0)}A^{(1)} + {A^{(1)}}^T P^{(0)}) = 0 \tag{25b}$$

By the same reasoning as used for equation (8), the first Lyapunov equation (25a) is feasible with a positive semidefinite $P^{(0)}$ satisfying $\begin{bmatrix} \mathbb{1}_n^T & \mathbb{0}_n^T \end{bmatrix} P^{(0)} = \begin{bmatrix} \mathbb{0}_n^T & \mathbb{0}_n^T \end{bmatrix}$. The second Lyapunov equation (25b) is feasible by analogous arguments.

Finally, by using (21) together with (23) and (24), we obtain

$$f(m + \delta\mu) = f^{(0)}_{(m,\mu)} + f^{(1)}_{(m,\mu)}\delta + \mathcal{O}(\delta^2)$$

where $f^{(0)}_{(m,\mu)} = f(m)$ and

$$\begin{aligned} f^{(1)}_{(m,\mu)} = \text{Trace}\Big(&2\,{B^{(1)}_{(m,\mu)}}^T P^{(0)}_{(m,\mu)} B^{(0)}_{(m,\mu)} \\ &+ {B^{(0)}_{(m,\mu)}}^T P^{(1)}_{(m,\mu)} B^{(0)}_{(m,\mu)}\Big) \end{aligned} \tag{26}$$

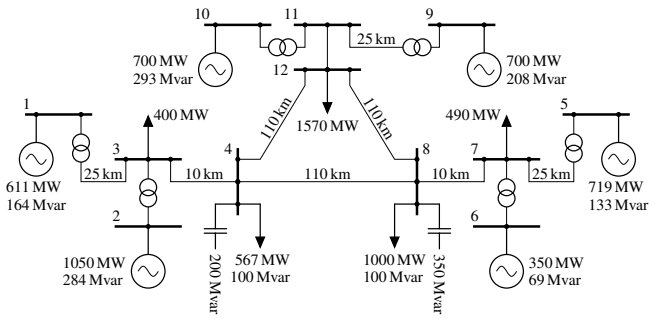

Fig. 3. Three-region test case adopted from [10], [23]

From (20), it clearly follows that $\nabla_\mu f(m) = f^{(1)}_{(m,\mu)}$ as defined in (26), thereby implicitly establishing differentiability of $f(m)$.

This concludes the proof, as the algorithm computes each component of the gradient $\nabla f(m)$ by using the relation (19) with the special choice of $\mu = \mathbb{e}_i$, $i \in \{1, \dots, n\}$, where $\mathbb{e}_i$ is the i-th vector of the canonical base of $\mathbb{R}^n$. ■

IV. Case study: 12-Bus-Three-Region System

In this section, we investigate a 12-bus case study illustrated in Figure 3. The system parameters are based on a modified two-area system from [23, Example 12.6] with an additional third area, as introduced in [10]. After Kron reduction, we obtain a systems of 9 buses, corresponding to the nodes where inertia can be allocated.

We investigate this example computationally using Algorithm 1 to drive standard gradient-based optimization routines, while highlighting parallels to our analytic results. We analyze different parametric scenarios and compare the inertia allocation and the performance of the proposed numerical optimization (which is a locally optimal solution) with two plausible heuristics that one may deduce from the conclusions in [2], [10]: namely the uniform allocation of the available budget, in the absence of capacity constraints, i.e. $m_i = m_{\text{uni}} = m_{\text{bdg}}/n$; or the allocation of the maximum inertia allowed by the bus capacity constraints, in the absence of a budget constraint, i.e. $m = \overline{m}$ (which we set as $\overline{m_i} = 4\underline{m_i}$).

a) Uniform disturbance: We first assume that the disturbance affects all nodes identically, $T = \text{diag}\{\mathbb{1}_9\}/9$. In Figure 4 we consider the case where there are only capacity constraints at each bus, and we compare the different allocations vis-à-vis: the initial inertia $\underline{m}$, a locally optimal solution $m^\star$, and the maximum inertia allocation $\overline{m}$. Figure 5 compares the results in the case where there is only a budget constraint on the total allocation. We compare the initial inertia $\underline{m}$, the locally optimal allocation $m^\star$, and the uniform placement $m_i = m_{\text{uni}}$.

b) Localized disturbance: We then consider the scenario where a localized disturbance affects a particular node, e.g., node 4 with $T = \text{diag}\{0, 0, 0, 1, 0, 0, 0, 0, 0\}$. As in Figures 4 and 5, a comparison of the different inertial allocations and the performance values is presented in Figures 6 and 7 for the cases of capacity and budget constraints.

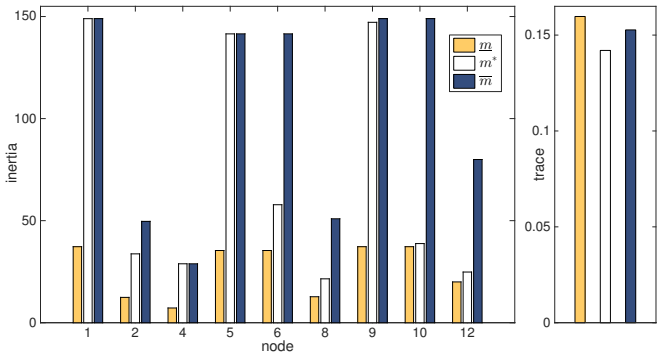

Fig. 4. Uniform disturbance with capacity constraints $m_i \leq \overline{m_i}$.

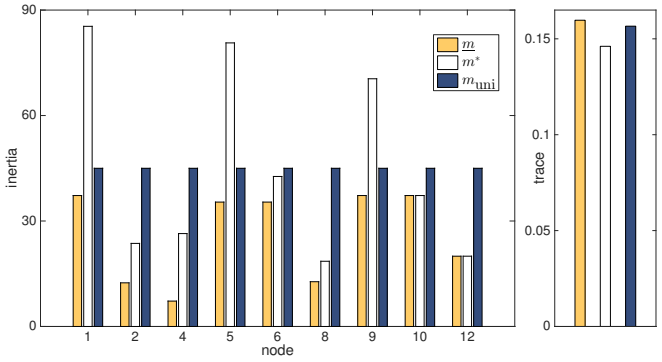

Fig. 5. Uniform disturbance with budget constraint $\sum_{i=1}^n m_i \leq m_{\text{bdg}}$.

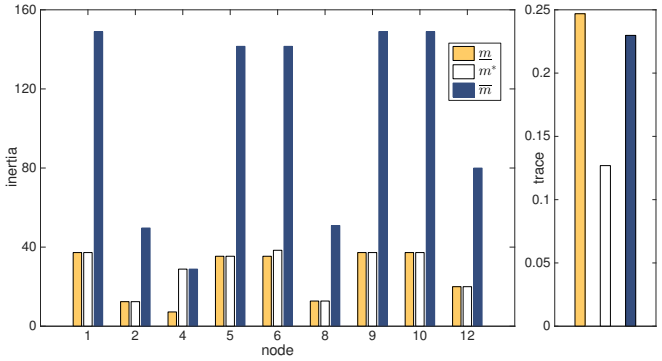

Fig. 6. Disturbance at node 4 with capacity constraints $m_i \leq \overline{m_i}$.

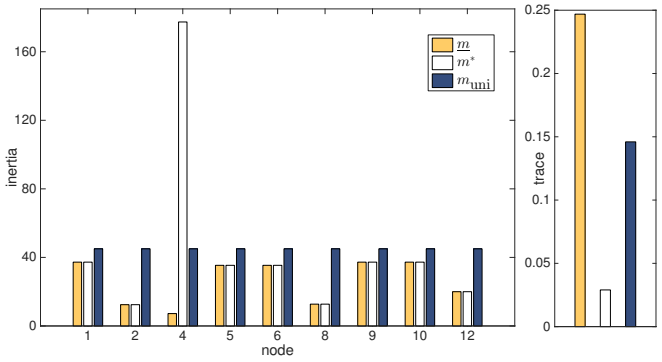

Fig. 7. Disturbance at node 4 with budget constraint $\sum_{i=1}^n m_i \leq m_{\text{bdg}}$.

c) Discussion of results: We infer the following from the above test cases: *i)* The locally optimal solution achieves the best performance among the different heuristics. *ii)* In case of uniform disturbances, the different heuristics as well as the initial allocation perform nearly optimal, which reveals similar features as the flat cost function observed in the two-area case (Section III-B) and the results in Theorem 3.1, indicating that any allocation is optimal under uniformity assumptions. *iii)* In stark contrast is the case of a localized disturbance, where adding inertia dominantly to disturbed node is an optimal choice in comparison to heuristic placements. The latter is also in line with the results presented for the two-area case.

V. Conclusions

We considered the problem of placing virtual inertia in power grids based on an $\mathcal{H}_2$ norm performance metric reflecting the network coherency. This formulation gives rise to a large-scale and non-convex optimization program. For certain problem instances and in the low-dimensional two-area case, we could derive closed-form solutions yielding some, perhaps surprising, insights. Next we developed a computational approach based on an explicit gradient formulation. Finally, we validated our results on a thee-area network to illustrate the effectiveness of our locally optimal inertia allocation and compared it with intuitive heuristics.

Our computational and analytic results are well aligned and suggest insightful strategies for the optimal allocation of virtual inertia. We envision that these results find application in stabilizing low-inertia grids through strategically placed virtual inertia. As part of our ongoing and future work, we also consider the extension to more detailed system models and the allocation of a finite number of virtual inertia units.

Acknowledgements

The authors wish to thank Mihailo Jovanovic, Andreas Ulbig, and Theodor Borsche for their insightful comments on the problem setup and the analysis methods.

References

[1] W. Winter, K. Elkington, G. Bareux, and J. Kostevc, "Pushing the limits: Europe's new grid: Innovative tools to combat transmission bottlenecks and reduced inertia," *Power and Energy Magazine, IEEE*, vol. 13, no. 1, pp. 60–74, Jan 2015.

[2] A. Ulbig, T. S. Borsche, and G. Andersson, "Impact of low rotational inertia on power system stability and operation," in *IFAC World Congress*, vol. 19, no. 1, 2014, pp. 7290–7297.

[3] N. Soni, S. Doolla, and M. C. Chandorkar, "Improvement of transient response in microgrids using virtual inertia," *Power Delivery, IEEE Transactions on*, vol. 28, no. 3, pp. 1830–1838, 2013.

[4] H. Bevrani, T. Ise, and Y. Miura, "Virtual synchronous generators: A survey and new perspectives," *International Journal of Electrical Power & Energy Systems*, vol. 54, pp. 244–254, 2014.

[5] S. D'Arco and J. Suul, "Virtual synchronous machines – classification of implementations and analysis of equivalence to droop controllers for microgrids," in *PowerTech (POWERTECH), 2013 IEEE Grenoble*, June 2013, pp. 1–7.

[6] J. Morren, S. W. H. De Haan, W. L. Kling, and J. A. Ferreira, "Wind turbines emulating inertia and supporting primary frequency control," *IEEE Transactions on Power Systems, 21 (1)*, 2006.

[7] M. Koller, T. Borsche, A. Ulbig, and G. Andersson, "Review of grid applications with the zurich 1 {MW} battery energy storage system," *Electric Power Systems Research*, vol. 120, pp. 128 – 135, 2015.

[8] P. Ashton, C. Saunders, G. Taylor, A. Carter, and M. Bradley, "Inertia estimation of the gb power system using synchrophasor measurements," *Power Systems, IEEE Transactions on*, vol. 30, no. 2, pp. 701–709, March 2015.

[9] E. Ela, V. Gevorgian, A. Tuohy, B. Kirby, M. Milligan, and M. O'Malley, "Market designs for the primary frequency response ancillary service – part i: Motivation and design," *Power Systems, IEEE Transactions on*, vol. 29, no. 1, pp. 421–431, 2014.

[10] T. S. Borsche, T. Liu, and D. J. Hill, "Effects of rotational inertia on power system damping and frequency transients," in *Conference on Decision and Control*, 2015, to appear.

[11] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1996.

[12] E. Lovisari and S. Zampieri, "Performance metrics in the average consensus problem: a tutorial," *Annual Reviews in Control*, vol. 36, no. 1, pp. 26–41, 2012.

[13] B. Bamieh, M. R. Jovanovic, P. Mitra, and S. Patterson, "Coherence in large-scale networks: Dimension-dependent limitations of local feedback," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2235–2249, 2012.

[14] M. Fardad, F. Lin, and M. R. Jovanovic, "Design of optimal sparse interconnection graphs for synchronization of oscillator networks," *Automatic Control, IEEE Transactions on*, vol. 59, no. 9, pp. 2457–2462, 2014.

[15] M. Fardad, X. Zhang, F. Lin, and M. R. Jovanović, "On the properties of optimal weak links in consensus networks," in *Proceedings of the 53rd IEEE Conference on Decision and Control*, Los Angeles, CA, 2014, pp. 2124–2129.

[16] T. Summers, I. Shames, J. Lygeros, and F. Dörfler, "Topology design for optimal network coherence," in *European Control Conference*, 2015, Available at http://arxiv.org/abs/1411.4884. [Online]. Available: http://arxiv.org/abs/1411.4884

[17] M. Siami and N. Motee, "Systemic measures for performance and robustness of large-scale interconnected dynamical networks," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 5119–5124.

[18] E. Sjödin, B. Bamieh, and D. F. Gayme, "The price of synchrony: Evaluating the resistive losses in synchronizing power networks," *IEEE Transactions on Control of Network Systems*, 2015, to appear.

[19] F. Dörfler, M. R. Jovanovic, M. Chertkov, and F. Bullo, "Sparsity-promoting optimal wide-area control of power networks," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2281–2291, September 2014.

[20] X. Wu, F. Dörfler, and M. R. Jovanovic, "Input-output analysis and decentralized optimal control of inter-area oscillations in power systems," *IEEE Transactions on Power Systems*, February 2015, to appear.

[21] F. Dörfler, J. W. Simpson-Porco, and F. Bullo, "Breaking the Hierarchy: Distributed Control & Economic Optimality in Microgrids," *IEEE Transactions on Control of Network Systems*, 2015, to appear.

[22] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Prentice Hall, 1998.

[23] P. Kundur, *Power System Stability and Control*. McGraw-Hill, 1994.

[24] F. Dörfler and F. Bullo, "Kron reduction of graphs with applications to electrical networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 1, pp. 150–163, 2013.

[25] Q.-C. Zhong and T. Hornik, *Control of Power Inverters in Renewable Energy and Smart Grid Integration*. Wiley-IEEE Press, 2013.

[26] J. Schiffer, D. Zonetti, R. Ortega, A. Stankovic, T. Sezi, and J. Raisch, "Modeling of microgrids-from fundamental physics to phasors and voltage sources," *arXiv preprint arXiv:1505.00136*, 2015.

[27] J. Schiffer, D. Goldin, J. Raisch, and T. Sezi, "Synchronization of droop-controlled autonomous microgrids with distributed rotational and electronic generation," in *IEEE Conf. on Decision and Control*, Florence, Italy, Dec. 2013, pp. 2334–2339.

[28] D. P. Bertsekas, *Convex Optimization Algorithms*. Athena Scientific, 2015.