Automatica 55 (2015) 27–36

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Online fault diagnosis for nonlinear power systems[☆]

Wei Pan[a], Ye Yuan[b,1], Henrik Sandberg[c], Jorge Gonçalves[b,d], Guy-Bart Stan[a]

[a] Centre for Synthetic Biology and Innovation and the Department of Bioengineering, Imperial College London, UK

[b] Control Group, Department of Engineering, University of Cambridge, UK

[c] Department of Automatic Control, School of Electrical Engineering, KTH Royal Institute of Technology, Sweden

[d] Luxembourg Centre for Systems Biomedicine, Luxembourg

ARTICLE INFO

Article history:
Received 2 June 2014
Received in revised form
28 September 2014
Accepted 6 February 2015
Available online 17 March 2015

Keywords:
Fault detection and isolation
Power systems
Machine learning

ABSTRACT

This paper considers the problem of automatic fault diagnosis for transmission lines in large scale power networks. Since faults in transmission lines threatens stability of the entire power network, fast and reliable fault diagnosis is an important problem in transmission line protection. This work is the first paper exploiting sparse signal recovery for the fault-diagnosis problem in power networks with nonlinear swing-type dynamics. It presents a novel and scalable technique to detect, isolate and identify transmission faults using a relatively small number of observations by exploiting the sparse nature of the faults. Buses in power networks are typically described by second-order nonlinear swing equations. Based on this description, the problem of fault diagnosis for transmission lines is formulated as a compressive sensing or sparse signal recovery problem, which is then solved using a sparse Bayesian formulation. An iterative reweighted ℓ_1-minimisation algorithm based on the sparse Bayesian learning update is then derived to solve the fault diagnosis problem efficiently. With the proposed framework, a real-time fault monitoring scheme can be built using only measurements of phase angles at the buses.

1. Introduction

Power networks are large-scale spatially distributed systems. Being critical infrastructures, they possess strict safety and reliability constraints. The design of monitoring schemes to diagnose anomalies caused by unpredicted or sudden faults on power networks is thus of great importance (Shahidehpour, Tinney, & Fu, 2005). To be consistent with the international definition of the fault diagnosis problem, the recommendations of the IFAC Technical Committee *SAFEPROCESS* is accordingly employed in what follows. Namely, this work proposes a method to: (1) decide whether there is an occurrence of a fault and the time of this occurrence (*i.e. detection*), (2) establish the location of the detected fault (*i.e. isolation*), and (3) determine the size and time-varying behaviour of the detected fault (*i.e. identification*).

Since power networks are typically large-scale and have nonlinear dynamics, fault diagnosis over transmission lines can be a very challenging problem. This paper draws inspiration from the fields of signal processing and machine learning to combine compressive sensing and variational Bayesian inference techniques so as to offer an efficient method for fault diagnosis.

Most of the literature available on fault diagnosis focuses on systems approximated by linear dynamics (Ding, 2008), with applications in networked system (Dong, Wang, & Gao, 2012), modern complex processes (Yin, Ding, Haghani, Hao, & Zhang, 2012), etc. Beyond linear systems descriptions, the dynamics of buses in power networks can be described by the so-called swing equations where the active power flows are nonlinear functions of the phase angles. Works that have considered fault detection and isolation in power networks include (Mohajerin Esfahani, Vrakopoulou, Andersson, & Lygeros, 2012; Shames, Teixeira, Sandberg, & Johansson, 2011; Zhang, Zhang, Polycarpou, & Parisini, 2014). Shames et al. (2011) focuses on distributed fault detection and isolation using linearised swing dynamics and the faults are considered to be additive. The method developed in Zhang et al. (2014) is used to detect sensor faults assuming that such faults appear as biased faults added to the measurement equation. In Mohajerin

[☆] The material in this paper was partially presented at the 52nd IEEE Conference on Decision and Control, December 10–13, 2013, Florence, Italy. This paper was recommended for publication in revised form by Associate Editor Huijun Gao, under the direction of Editor Ian R. Petersen.

E-mail addresses: w.pan11@imperial.ac.uk (W. Pan), yy311@cam.ac.uk (Y. Yuan), hsan@kth.se (H. Sandberg), jmg77@cam.ac.uk (J. Gonçalves), g.stan@imperial.ac.uk (G.-B. Stan).

[1] Correspondence to: Department of Bioengineering, Imperial College London, SW7 2AZ, UK. Tel.: +44 (0) 207 59 46375.

http://dx.doi.org/10.1016/j.automatica.2015.02.032

Esfahani et al. (2012), a fault detection and isolation residual generator is presented for nonlinear systems with additive faults. The nonlinearities in Mohajerin Esfahani et al. (2012) are not imposed *a priori* on the model structure but treated as disturbances with some known patterns.

To summarise, the works (Ding, 2008; Dong et al., 2012; Shames et al., 2011; Yin et al., 2012) use linear systems to characterise the dynamics of power networks and the faults are assumed to be additive. Though the system dynamics are nonlinear in Mohajerin Esfahani et al. (2012) and Zhang et al. (2014), the faults are still assumed to be additive. The methods developed on the basis of these conservative assumptions yield several problems. Firstly, the linear approximation to nonlinear swing equations can only be used when the phase angles are close to each other. However, when the system is strained and faults appear, phase angles can often be far apart. Therefore, a linear approximation is inappropriate in strained power network situations. Secondly, it is well-known that a large portion of power system faults occurring in transmission lines do not involve additive faults, e.g. a short-circuit fault occurring on the transmission lines between generators would correspond to some changes in the parameters of the nonlinear terms appearing in the swing equation (Kundur, Balu, & Lauby, 1994). Furthermore, the inevitable and frequent introduction of new components in a power network contributes to the vulnerability of transmission lines, which, if not appropriately controlled, can lead to cascading failures (Hines, Balasubramaniam, & Sanchez, 2009; Jiang, Yang, Lin, Liu, & Ma, 2000). Such cascading failures cannot be captured by additive faults. Finally, the methods mentioned above only address fault detection and isolation rather than identification, which is crucial to take appropriate actions when faults occur on transmission lines.

Contributions. The power networks considered in this paper are described by the *nonlinear swing equations* with additive process noise. The faults are assumed to occur on the transmission lines of the power network. The problem of fault diagnosis, i.e. detection, isolation and identification, of such nonlinear power networks is formulated as a compressive sensing or sparse signal recovery problem. To solve this problem we consider a sparse Bayesian formulation of the fault identification problem, which is then casted as a nonconvex optimisation problem. Finally, the problem is relaxed into a convex problem and solved efficiently using an iterative reweighted ℓ_1-minimisation algorithm. The resulting efficiency of the proposed method enables real-time detection of faults in large-scale networks.

Outline. The outline of the paper is as follows. Section 2 introduces the nonlinear model of power networks considered in this paper. Section 3 formulates the fault diagnosis problem as a compressive sensing or sparse signal recovery problem. Section 4 shows how the resulting nonconvex optimisation problem can be relaxed into a convex optimisation problem and solved efficiently using an iterative reweighted ℓ_1-minimisation algorithm. Section 5 applies the method to a power network with 20 buses and 80 transmission lines and, finally, Section 6 concludes and discusses several future problems.

Notation. The notation in this paper is standard. Bold symbols are used to denote vectors and matrices. For a matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$, $\mathbf{A}_{i,j} \in \mathbb{R}$ denotes the element in the ith row and jth column, $\mathbf{A}_{i,:} \in \mathbb{R}^{1\times N}$ denotes its ith row, $\mathbf{A}_{:,j} \in \mathbb{R}^{M\times 1}$ denotes its jth column. For a column vector $\boldsymbol{\alpha} \in \mathbb{R}^{N\times 1}$, α_i denotes its ith element. In particular, $\mathbf{I}_l$ denotes the identity matrix of size $l \times l$. We simply use $\mathbf{I}$ when the dimension is obvious from context. $\|\mathbf{w}\|_1$ and $\|\mathbf{w}\|_2$ denote the ℓ_1 and ℓ_2 norms of the vector $\mathbf{w}$, respectively. $\|\mathbf{w}\|_0$ denotes the ℓ_0 "norm" of the vector $\mathbf{w}$, which counts the number of nonzero elements in the vector $\mathbf{w}$. diag $[\gamma_1, \ldots, \gamma_N]$ denotes a diagonal matrix with principal diagonal elements being $\gamma_1, \ldots, \gamma_N$. $\mathbb{E}(\boldsymbol{\alpha})$ stands for the expectation of stochastic variable $\boldsymbol{\alpha}$.

2. Model formulation

Power systems are examples of complex systems in which generators and loads are dynamically interconnected. Hence, they can be seen as networked systems, where each bus is a node in the network. We assume that all the buses in the network are connected to synchronous machines (motors or generators). The nonlinear model for the active power flow in a transmission line connected between bus i and bus j is given as follows. For $i = 1, \ldots, n$, the behaviour of bus/node i can be represented by the swing equation (Kundur et al., 1994; Shames et al., 2011; Zhang et al., 2014)

$$m_i\ddot{\delta}_i(t) + d_i\dot{\delta}_i(t) - P_{mi}(t) = -\sum_{j\in N_i} P_{ij}(t), \tag{1}$$

where δ_i is the phase angle of bus i, m_i and d_i are the inertia and damping coefficients of the motors and generators, respectively, P_{mi} is the mechanical input power, P_{ij} is the active power flow from bus i to j, and N_i is the neighbourhood set of bus i where bus j and i share a transmission line or communication link.

Considering that there are no power losses nor ground admittances, and letting $V_i = |V_i|e^{\tilde{j}\delta_i}$ be the complex voltage of bus i where $\tilde{j}$ represents the imaginary unit, the active power flow between bus i and bus j, P_{ij}, is given by:

$$P_{ij}(t) = w_{ij}^{(1)} \cos(\delta_i(t) - \delta_j(t)) + w_{ij}^{(2)} \sin(\delta_i(t) - \delta_j(t)), \tag{2}$$

where $w_{ij}^{(1)} = |V_i|\,|V_j|G_{ij}$ and G_{ij} is the branch conductance between bus i and bus j; and $w_{ij}^{(2)} = |V_i|\,|V_j|B_{ij}$ and B_{ij} is the branch susceptance between bus i and bus j.

If we let $\xi_i(t) = \delta_i(t)$ and $\zeta_i(t) = \dot{\delta}_i(t)$, each bus can be assumed to have double integrator dynamics. The dynamics of bus i can thus be written:

$$\dot{\xi}_i(t) = \zeta_i(t), \tag{3}$$

$$\dot{\zeta}_i(t) = u_i(t) + v_i(t), \tag{4}$$

where ξ_i, ζ_i are scalar states, $v_i(t)$ is a known scalar external input, and u_i is the power flow

$$v_i(t) = \frac{P_{mi}(t)}{m_i} \tag{5}$$

$$\begin{aligned} u_i(t) \;=\; & -\frac{d_i}{m_i}\zeta_i(t) - \frac{1}{m_i}\sum_{j\in N_i}[w_{ij}^{(1)} \cos(\xi_i(t) - \xi_j(t)) \\ & + w_{ij}^{(2)} \sin(\xi_i(t) - \xi_j(t))]. \end{aligned} \tag{6}$$

The variables ξ_i and ζ_i can be interpreted as phase and frequency in the context of power networks.

In Shames et al. (2011), the cos(·) terms are neglected (no branch conductance between buses) and it is assumed that phase angles are close to each other. The dynamics in (1) are then linearised to yield

$$m_i\ddot{\delta}_i(t) + d_i\dot{\delta}_i(t) - P_{mi}(t) = -\sum_{j\in N_i} w_{ij}^{(2)}(\delta_i(t) - \delta_j(t)). \tag{7}$$

Each bus i is assumed to have double integrator dynamics as described in (3) and (4). $u_i(t)$ in (6) becomes a linear equation

$$u_i(t) = -\frac{d_i}{m_i}\xi_i(t) - \frac{1}{m_i}\sum_{j\in N_i} w_{ij}^{(2)}(\xi_i(t) - \xi_j(t)). \tag{8}$$

For the linearised system (8), a bus k is faulty if for some functions $f_{\xi k}(t)$ and $f_{\zeta k}(t)$ not identical to zero either $\dot{\xi}_i(t) = \zeta_i(t) + f_{\xi k}(t)$, or $\dot{\zeta}_i(t) = u_i(t) + v_i(t) + f_{\zeta k}(t)$. The functions $f_{\xi k}(t)$ and $f_{\zeta k}(t)$ are referred to as fault signals. Model-based or observer-based fault diagnosis methods are available for power networks (see Shames et al., 2011 and reference therein). However, specific aspects need careful consideration when dealing with fault diagnosis in power networks. Firstly, the simplified linear model can only be used when the phase angles are close to each other. However, when the system is strained and faults appear, phase angles can often be far apart.

In transmission systems the $\sin(\cdot)$ term in (2) is the dominating one. To perform a linearisation, one often assumes "small angle differences" between nodes and hence "small" power flows. This typically works well under normal operation. However, if the power system is under a lot of strain, i.e. if power flows are closer to the theoretical maximum, the angle difference becomes close to 90 degrees and the nonlinearity of the $\sin(\cdot)$ term becomes quite noticeable. In particular, if, in a transient state, the angle difference exceeds 90°, generators typically loose synchrony and trip. This is not captured by linear models. In such circumstances, the linear model cannot be used to approximate the nonlinear model in (1) anymore. Secondly, power networks are highly distributed and interconnected, and more than one transmission line can be faulty at a given time. Thirdly, to be more realistic, some process noise ε_i should be incorporated into the second-order system (1) for each bus i:

$$m_i\ddot{\delta}_i(t) + d_i\dot{\delta}_i(t) - P_{mi}(t) = -\sum_{j=1}^{n} P_{ij}(t) + \varepsilon_i(t). \tag{9}$$

Based on the *swing* equation above, the state space model (3) and (4) can then be rewritten under the form:

$$\dot{\xi}_i(t) = \zeta_i(t), \tag{10}$$

$$\dot{\zeta}_i(t) = u_i(t) + v_i(t) + \varepsilon_i(t), \tag{11}$$

where the noise $\varepsilon_i(t)$ is assumed to be i.i.d. Gaussian with $\mathbb{E}(\varepsilon_i(p)) = 0$, $\mathbb{E}(\varepsilon_i(p)\varepsilon_i(q)) = \epsilon_i^2\delta(p-q)$.

Remark 1. Here we only consider a dynamical system model with process noise ε_i since, in power networks, the measurement noise is small and would typically not have a catastrophic effect on the performance of detection algorithms (Tate & Overbye, 2008). However, we are also currently investigating the case where measurement noise is not neglected. This generalisation is beyond the scope of this paper and will potentially be the subject of a later paper.

3. Problem formulation

Given the model and explanation above, we primarily focus on the following setting in this paper.

Definition 1. If a power network can be described by (10) and (11), the transmission line between bus i and bus j is faulty when $w_{ij}^{(1)}$ changes to a new scalar $w_{ij}^{[\mathbf{f}](1)}$ and/or $w_{ij}^{(2)}$ changes to a new scalar $w_{ij}^{[\mathbf{f}](2)}$, where $w_{ij}^{(1)}$ and $w_{ij}^{(2)}$ are the weights for the cos and sin terms defined in (6).

Based on the considerations above and Definition 1, the problem that we are interested in solving is the following:

Problem 1. Having access to the measurements and the distribution of the noise, how can we detect the occurrence and magnitude of a fault, namely, how can we estimate the magnitude of the errors $w_{ij}^{(1)} - w_{ij}^{[\mathbf{f}](1)}$ and $w_{ij}^{(2)} - w_{ij}^{[\mathbf{f}](2)}$, $\forall i, j$, using the smallest possible number of samples.

In what follows we make the following assumption.

Assumption 1. The power networks described by (10) and (11) are fully measurable, i.e. the phase angles of all the buses can be measured.

3.1. Model transformation

Applying the forward Euler discretisation scheme to (10) and (11) and assuming the discretisation step $t_{k+1} - t_k = \Delta t$ is constant for all k, we obtain the following discrete-time system approximation to the continuous-time system (10) and (11):

$$\frac{\xi_i(t_{k+1}) - \xi_i(t_k)}{\Delta t} = \zeta_i(t_k), \tag{12}$$

$$\frac{\zeta_i(t_{k+1}) - \zeta_i(t_k)}{\Delta t} = u_i(t) + v_i(t) + \eta_i(t_k), \tag{13}$$

where the noise $\eta_i(t_k)$ is assumed to be i.i.d. Gaussian distributed: $\eta_i(t_k) \sim \mathcal{N}(0, \sigma_i^2)$, with $\mathbb{E}(\eta_i(t_p)) = 0$, $\mathbb{E}(\eta_i(t_p)\eta_i(t_q)) = \sigma_i^2\delta(t_p - t_q)$.

Defining the new variable

$$e_i(t_{k+1}) \triangleq -\frac{(\zeta_i(t_{k+1}) - \zeta_i(t_k))}{\Delta t} - \frac{d_i\zeta_i(t_k)}{m_i} + \frac{P_{mi}(t_k)}{m_i}, \tag{14}$$

we have

$$e_i(t_{k+1}) = \frac{1}{m_i}\sum_{j\in N_i}[w_{ij}^{(1)}\cos(\xi_i(t_k) - \xi_j(t_k)) + w_{ij}^{(2)}\sin(\xi_i(t_k) - \xi_j(t_k))] + \eta_i(t_k), \tag{15}$$

where e_i, the power flow measurement, is treated as the output of the system. Since the state variables $\zeta(t_{k+1})$ and $\zeta(t_k)$, the parameters Δt, d_i and m_i, and the input P_{mi} are known, the quantity $e_i(t_{k+1})$ can be computed in real time. It should be noted that "real time" is to be understood as "within the sampling time Δt of the sensors in power generators".

By defining $\mathbf{x}(t_k) = [\xi_1(t_k), \ldots, \xi_N(t_k)]$ we can write (14) into a vector form:

$$e_i(t_{k+1}) = f_i(\mathbf{x}(t_k))\mathbf{w}_i^{\text{true}} + \eta_i(t_k), \tag{16}$$

with

$$f_i(\mathbf{x}(t_k)) = [f_i^{(1)}(\mathbf{x}(t_k)), f_i^{(2)}(\mathbf{x}(t_k))] \in \mathbb{R}^{2n},$$
$$f_i^{(1)}(\mathbf{x}(t_k)) = [\cos(\xi_i(t_k) - \xi_1(t_k)), \ldots, \cos(\xi_i(t_k) - \xi_N(t_k))] \in \mathbb{R}^n,$$
$$f_i^{(2)}(\mathbf{x}(t_k)) = [\sin(\xi_i(t_k) - \xi_1(t_k)), \ldots, \sin(\xi_i(t_k) - \xi_N(t_k))] \in \mathbb{R}^n,$$
$$\mathbf{w}_i^{\text{true}} = [\mathbf{w}_i^{(1)}, \mathbf{w}_i^{(2)}]^{\mathrm{T}} \in \mathbb{R}^{2n},$$
$$\mathbf{w}_i^{(1)} = [w_{i1}^{(1)}, \ldots, w_{iN}^{(1)}] \in \mathbb{R}^n,$$
$$\mathbf{w}_i^{(2)} = [w_{i1}^{(2)}, \ldots, w_{iN}^{(2)}] \in \mathbb{R}^n,$$

where $f_i(\mathbf{x}(t_k))$ represents the transmission functions and $\mathbf{w}_i$ represents the corresponding transmission weights associated to the topology of the network.

Remark 2. In real power systems, a sampling frequency for phasor measurement unit (PMU) as high as 2500 samples per second can be achieved (Phadke & Thorp, 2008). In this case, the sampling time Δt is $4 * 10^{-5}$ second and the Euler discretisation $\frac{\xi_i(t_{k+1}) - \xi_i(t_k)}{\Delta t}$ will typically provide a good approximation of $\dot{\xi}_i(t)$.

3.2. Fault diagnosis problem formulation

As stated in Definition 1, if there are no faults occurring in the transmission lines between bus i and other buses, the dynamics of the power networks will evolve according to (16). The expected output for the next sampling time is defined to be

$$e_i^{[\mathbf{e}]}(t_{k+1}) = f_i(\mathbf{x}(t_k))\mathbf{w}_i^{\text{true}}. \tag{17}$$

From (16) and (17), it is easy to show that $e_i(t_{k+1}) - e_i^{[\mathbf{e}]}(t_{k+1})$ is a stochastic variable with zero mean and variance σ^2. If there are faults occurring in the transmission lines between bus i and other buses, the corresponding transmission weights will change from $\mathbf{w}_i^{\text{true}}$ to $\mathbf{w}_i^{\text{fault}}$. Similar to the definition of $\mathbf{w}_i^{\text{true}}$, $\mathbf{w}_i^{\text{fault}} = [\mathbf{w}_i^{[\mathbf{f}](1)}, \mathbf{w}_i^{[\mathbf{f}](2)}]^{\mathrm{T}}$ where $\mathbf{w}_i^{[\mathbf{f}](1)} = [w_{i1}^{[\mathbf{f}](1)}, \ldots, w_{iN}^{[\mathbf{f}](1)}]$ and $\mathbf{w}_i^{[\mathbf{f}](2)} = [w_{i1}^{[\mathbf{f}](2)}, \ldots, w_{iN}^{[\mathbf{f}](2)}]$. We thus have:

$$e_i^{[\mathbf{f}]}(t_{k+1}) = f_i(\mathbf{x}(t_k))\mathbf{w}_i^{\text{fault}} + \eta_i(t_k), \tag{18}$$

where $e_i^{[\mathbf{f}]}$ is the output when there are faults.

From (17) and (18), it is easy to find that $e_i^{[\mathbf{f}]}(t_{k+1}) - e_i^{[\mathbf{e}]}(t_{k+1})$ is a stochastic variable with mean $f_i(\mathbf{x}(t_k))(\mathbf{w}_i^{\text{fault}} - \mathbf{w}_i^{\text{true}})$ and variance σ^2. Denoting

$$y_i = e_i^{[\mathbf{f}]} - e_i^{[\mathbf{e}]}, \qquad \mathbf{w}_i = \mathbf{w}_i^{\text{fault}} - \mathbf{w}_i^{\text{true}},$$

we have:

$$y_i(t_{k+1}) = f_i(\mathbf{x}(t_k))\mathbf{w}_i + \eta_i(t_k). \tag{19}$$

Remark 3. We formulate the faults identification problem as a linear regression problem. The dependent variable $e_i^{[\mathbf{f}]}(t_{k+1}) - e_i^{[\mathbf{e}]}(t_{k+1})$ is the difference between the expected output and the faulty output; the unknown variable we want to estimate is the difference between the faulty transmission weights and the true transmission weights.

There are three problems of interest based on the formulation in (19): (a) detection of a fault; (b) isolation of a fault, i.e. determination of the type, location and time of occurrence of a fault; and (c) identification of the size and time-varying behaviour of a fault. In the noiseless case, when there are no faults, $\forall i$, y_i and $\mathbf{w}_i$ are both equal to zero. On the other hand, when there are faults, certain y_i are nonzero. So the faults can be *detected* by identifying the entries y_i that are nonzero. However, in the noisy case, even when there are no faults, y_i is nonzero most of the time since it is a stochastic variable with zero mean. This can be interpreted in a probabilistic way by Chebyshev's Inequality: $\mathscr{P}(|e_i(t_{k+1}) - e_i^{[\mathbf{e}]}(t_{k+1})| \geq l\sigma) \leq \frac{1}{l^2}$ where $l \in \mathbb{R}^+$. According to this inequality, when there are no faults, the deviation between true and expected outputs, i.e. $|e_i(t_{k+1}) - e_i^{[\mathbf{e}]}(t_{k+1})|$ cannot be much greater than zero with high probability. On the other hand, when there is a fault, the deviation between faulty and expected outputs, i.e. $|e_i^{[\mathbf{f}]}(t_{k+1}) - e_i^{[\mathbf{e}]}(t_{k+1})|$ should be much greater than zero with high probability.

From an isolation point of view and Chebyshev's inequality, when $|e_i^{[\mathbf{f}]}(t_{k+1}) - e_i^{[\mathbf{e}]}(t_{k+1})|$ is much greater than σ, the fault can be *isolated* with high probability (e.g. if the threshold is set to $l\sigma = 10\sigma$, then the probability is 99%).

If at time t_0 faults have been detected and isolated, the remaining task is to perform fault *identification*, i.e. to identify the location of the faults or equivalently to find the nonzero entries in w_i. Assuming that $M+1$ successive data points, including the initial data point at t_0, are sampled and defining $N = 2n$ and

$$\begin{aligned}
\mathbf{y}_i &\triangleq [y_i(t_1), \ldots, y_i(t_M)]^{\mathrm{T}} \in \mathbb{R}^M, \\
\mathbf{A}_i &\triangleq \begin{bmatrix} f_i^{(1)}(\mathbf{x}(t_0)) & f_i^{(2)}(\mathbf{x}(t_0)) \\ \vdots & \vdots \\ f_i^{(1)}(\mathbf{x}(t_{M-1})) & f_i^{(2)}(\mathbf{x}(t_{M-1})) \end{bmatrix} \\
&= \begin{bmatrix} f_i(\mathbf{x}(t_0)) \\ \vdots \\ f_i(\mathbf{x}(t_{M-1})) \end{bmatrix} \in \mathbb{R}^{M\times N}, \\
\boldsymbol{\eta}_i &\triangleq [\eta_i(t_0), \ldots, \eta_i(t_{M-1})]^{\mathrm{T}} \in \mathbb{R}^M,
\end{aligned} \tag{20}$$

we can write N independent equations of the form:

$$\mathbf{y}_i = \mathbf{A}_i\mathbf{w}_i + \boldsymbol{\eta}_i, \quad (i = 1, \ldots, n). \tag{21}$$

Based on the formulation in (21), our goal is to find $\mathbf{w}_i$ given the output data stored in $\mathbf{y}_i$.

To solve for $\mathbf{w}_i$ in (21) amounts to solving a linear regression problem. This can be done using standard least square approaches. It should be noted that the linear regression problem for bus i in (21) is independent from the linear regression problems for the other buses. In what follows, we will focus on finding the solution to one of these linear regression problem and omit the subscripts i in (21) for simplicity of notation. We thus write

$$\mathbf{y} = \mathbf{A}\mathbf{w} + \boldsymbol{\eta}, \tag{22}$$

where $\mathbf{y}$ is the difference between the faulty measurements and the expected measurements, or namely, the **error measurements**; and $\mathbf{w}$ is the difference between the faulty parameters and the true parameters, or namely, the **faults**. We address this linear regression problem under the following assumption.

Assumption 2. A maximum of S transmission lines are faulty, i.e. $\mathbf{w}$ has at most S non-zero entries. In other words, $\mathbf{w}$ is S-sparse or mathematically, $\|\mathbf{w}\|_0 \leq S$. The constant S is assumed unknown to the system administrator.

Remark 4. Assumption 2 is realistic for small values of S since in the context of a power system, it is typically not the case that all the transmission lines are faulty simultaneously. Furthermore, since buses in power networks are typically sparsely connected the number of faults is typically much smaller than the size of the network n, i.e. $S \ll n$. Therefore $S \ll N = 2n$.

On the other-hand, the size of $\mathbf{y}$ equals to the number of samples needed to identify the location of the faults after the they occur. From a practical viewpoint, the number of samples should be as small as possible. However, standard least square approaches to (22) cannot meet this goal as they require at least $2N$ samples. Moreover, the solution to the standard least square problem is generically dense (hence, violating Assumption 2) and cannot be used to identify which transmission lines are likely to be faulty by identification of the nonzero entries of the estimated $\mathbf{w}^{\text{fault}} - \mathbf{w}^{\text{true}}$.

3.3. Discussion on fault identification

Under the assumption that the system under consideration is identifiable (Němcová, 2010), we cannot get a sparser solution than the true one, as this would contradict the identifiability assumption, i.e. more than one model can equivalently explain the data. In order to search for the sparsest solution $\mathbf{w}$, we impose a penalty on the ℓ_0 norm of $\mathbf{w}$, $\|\mathbf{w}\|_0$, i.e. on the number of nonzero elements in $\mathbf{w}$. With the addition of this ℓ_0 norm penalty, the linear regression problem (22) can be formulated into the following regularised regression problem, which is also known as an ℓ_0-minimisation problem (Candès & Tao, 2005; Donoho, 2006):

$$\hat{\mathbf{w}} = \operatorname*{argmin}_{\mathbf{w}}\{\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 + \rho\,\|\mathbf{w}\|_0\}. \tag{23}$$

In (23), $\mathbf{y}$ is the vector observations, $\mathbf{A}$ is a known regressor matrix, $\mathbf{w}$ is the vector of unknown coefficients and ρ is a tradeoff parameter. Subsequently, one may wonder what the gap between the solution to this ℓ_0-minimisation problem and the true solution is.

To characterise this gap, we shall firstly introduce the following definition.

Definition 2 (*Definition 1 of Donoho & Elad, 2003*)**.** The spark of a given matrix A, i.e., Spark($\mathbf{A}$), is the smallest number of columns of A that are linearly dependent.

Proposition 1 (*Corollary 1 of* Donoho & Elad, 2003**).** *In the noiseless case where $\boldsymbol{\eta} = \mathbf{0}$ for any vector $\mathbf{y} \in \mathbb{R}^M$, there exists one unique signal $\mathbf{w}$, such that $\mathbf{y} = \mathbf{A}\mathbf{w}$ with $\|\mathbf{w}\|_0 = S$ if and only if Spark$(\mathbf{A}) > 2S$.*

Remark 5. It is easy to see that Spark$(\mathbf{A}) \in [2, M+1]$. Therefore, in order to get the unique S-sparse solution $\mathbf{w}$ to $\mathbf{y} = \mathbf{A}\mathbf{w}$, Proposition 1 imposes that $M \geq 2S$.

Corollary 1. *If the number of samples M is greater or equal to 2 times the number of nonzero elements S in the "true" value of $\mathbf{w}$, then the ℓ_0-minimisation solution $\mathbf{w}$ to the equation $\mathbf{y} = \mathbf{A}\mathbf{w}$ will be consistent with the "true" value.*

Proof. Since the sparsest solution can be obtained through ℓ_0-minimisation in (23), this corollary is straightforward from Proposition 1 and Remark 5.

Remark 6. This corollary bridges the gap between the "true" solution and that obtained by ℓ_0-minimisation provided the assumptions of Corollary 1 hold. If these assumptions do not hold, then prior knowledge, additional experiments and/or data points might be required.

3.4. Drawbacks of ℓ_1 relaxation and further motivation for our approach

Unfortunately, obtaining a solution through ℓ_0-minimisation is both numerically unstable and NP-hard. Instead, ℓ_1 relaxation is commonly used since the ℓ_1-norm is the tightest convex relaxation to the ℓ_0-norm (Candes, Wakin, & Boyd, 2008). The ℓ_1 relaxation of the optimisation problem in (23) is

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}}\{\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 + \rho\,\|\mathbf{w}\|_1\}. \tag{24}$$

A sufficient condition for exact reconstruction based on ℓ_1-minimisation is the so called *restricted isometry property* (RIP) (Candès & Tao, 2005). It was shown in Candès and Tao (2005), Candès, Romberg, and Tao (2006) and Dai and Milenkovic (2009) that both convex ℓ_1-minimisations and greedy algorithms lead to exact reconstruction of S-sparse signals if the matrix $\mathbf{A}$ satisfies the RIP condition. One major drawback of the RIP condition is that it can be very difficult to check (combinatorial search). Another related and easier-to-check property is the coherence property. The *coherence* of a matrix $\mathbf{A}$ is defined as $\mu(\mathbf{A}) = \max_{j<k} \frac{|\langle \mathbf{A}_{:,j}, \mathbf{A}_{:,k}\rangle|}{\|\mathbf{A}_{:,j}\|_2 \|\mathbf{A}_{:,k}\|_2}$. It was shown that RIP guarantees *incoherence* of $\mathbf{A}$, i.e. $\mu(\mathbf{A}) \approx 0$, Candès and Tao (2005). This means one is guaranteed that ℓ_1-minimisation solutions are equivalent to the true solution only when $\mathbf{A}$ is near orthogonal, i.e. when the columns of $\mathbf{A}$ are strongly uncorrelated. However, in power networks, correlation between the columns of $\mathbf{A}$ is typically high (close to 1). A different approach thus needs to be considered. We propose hereafter a method intended to solve compressive sensing problems in situations where ℓ_1 relaxations usually do not work (see Pan, Yuan, Gonçalves, & Stan, 2015 for details). Our approach uses a Bayesian formulation to solve (22) (see Tipping, 2001 for details).

4. Bayesian viewpoint on fault diagnosis problem

Bayesian modelling treats all unknowns as stochastic variables with certain probability distributions (Bishop, 2006). For $\mathbf{y} = \mathbf{A}\mathbf{w} + \boldsymbol{\eta}$. The likelihood of the error measurements $\mathbf{y}$ given the faults $\mathbf{w}$ is

$$\mathcal{P}(\mathbf{y}|\mathbf{w}) = \mathcal{N}(\mathbf{y}|\mathbf{A}\mathbf{w}, \sigma^2\mathbf{I}) \propto \exp\left[-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{w}\|^2\right]. \tag{25}$$

Given the likelihood function in (25) and specifying a prior on the faults which is $\mathcal{P}(\mathbf{w}) = \prod_{j=1}^N \mathcal{P}(w_j)$, where w_j is the jth element of the faults vector $\mathbf{w}$, i.e. $w_j \in \mathbf{w}$. We compute the *posterior distribution* over $\mathbf{w}$ via Bayes' rule:

$$\mathcal{P}(\mathbf{w}|\mathbf{y}) = \frac{\mathcal{P}(\mathbf{y}|\mathbf{w})\mathcal{P}(\mathbf{w})}{\int \mathcal{P}(\mathbf{y}|\mathbf{w})\mathcal{P}(\mathbf{w})d\mathbf{w}}.$$

We further define a prior distribution $\mathcal{P}(\mathbf{w})$ as

$$\mathcal{P}(\mathbf{w}) \propto \exp\left[-\frac{1}{2}g(\mathbf{w})\right] = \exp\left[-\frac{1}{2}\sum_{j=1}^N g(w_j)\right], \tag{26}$$

where $g(w_j)$ is an arbitrary function of w_j. We then formulate a *maximum a posteriori* (MAP) estimate on the faults:

$$\begin{aligned}\mathbf{w}_{\mathbf{MAP}} &= \underset{\mathbf{w}}{\operatorname{argmax}}\,\mathcal{P}(\mathbf{w}|\mathbf{y}) \\ &= \underset{\mathbf{w}}{\operatorname{argmin}}\{\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 + \sigma^2 g(\mathbf{w})\},\end{aligned} \tag{27}$$

where $g(\mathbf{w})$ is defined as a penalty function. From a Bayesian viewpoint, MAP estimation is equivalent to a penalised least square (PLS) problem.

In the following sections, we derive a sparse Bayesian formulation of the fault diagnosis problem which is casted into a nonconvex optimisation problem. We relax the nonconvex optimisation problem and develop an iterative reweighted ℓ_1-minimisation algorithm to solve the resulting problem.

4.1. Super Gaussian prior distribution

In practice, the penalty function over the faults $g(\mathbf{w})$ is usually chosen as a concave, non-decreasing function of the faults $|\mathbf{w}|$ that can enforce sparsity constraints over the faults. Since the posterior of the faults given the error measurements $\mathcal{P}(\mathbf{w}|\mathbf{y})$ is highly coupled and non-Gaussian, computing the posterior mean $\mathbb{E}(\mathbf{w}|\mathbf{y})$ for the faults is generally intractable. To alleviate this problem, ideally one would like to approximate $\mathcal{P}(\mathbf{w}|\mathbf{y})$ as a Gaussian distribution from which analytical results can be obtained and efficient algorithms exist (Bishop, 2006). To this end, we may consider *super-Gaussian* priors, which yield a lower bound for the priors $\mathcal{P}(w_j)$. More specifically, if we define $\boldsymbol{\gamma} \triangleq [\gamma_1, \ldots, \gamma_N]^{\mathrm{T}} \in \mathbb{R}_+^N$, we can represent the prior in the following relaxed (variational) form:

$$\mathcal{P}(\mathbf{w}) = \prod_{j=1}^N \mathcal{P}(w_j),\ \ \mathcal{P}(w_j) = \max_{\gamma_j > 0} \mathcal{N}(w_j|0, \gamma_j)\varphi(\gamma_j), \tag{28}$$

where $\varphi(\gamma_j)$ is a nonnegative function which is treated as a hyperprior with γ_j being its associated hyperparameters. Throughout, we call $\varphi(\gamma_j)$ the "*potential function*". This Gaussian relaxation is possible if and only if $\log \mathcal{P}(\sqrt{w_j})$ is concave on $(0, \infty)$. The following theorem provides a justification for the above:

Theorem 1 (*Palmer, Wipf, Kreutz-Delgado, & Rao, 2006*)**.** *A probability density $\mathcal{P}(w_j) \equiv \exp(-g(w_j^2))$ can be represented in the convex variational form: $\mathcal{P}(w_j) = \max_{\gamma_j>0} \mathcal{N}(w_j|0, \gamma_j)\varphi(\gamma_j)$ if and only if $-\log \mathcal{P}(\sqrt{w_j}) = g(w_j)$ is concave on $(0, \infty)$. In this case the potential function takes the following expression: $\varphi(\gamma_j) = \sqrt{2\pi/\gamma_j}\exp\left(g^*\left(\gamma_j/2\right)\right)$ where $g^*(\cdot)$ is the concave conjugate of $g(\cdot)$. A symmetric probability density $\mathcal{P}(w_j)$ is said to be super-Gaussian if $\mathcal{P}(\sqrt{w_j})$ is log-convex on $(0, \infty)$.*

Remark 7. For the *Laplace* prior $\mathcal{P}(w_j) \propto \exp(-\lambda \sum_j |w_j|)$, one can have a *Laplace* potential function $\varphi(\gamma_j) = \exp\left(-1/2|\gamma_j|\right)\sqrt{2\pi|\gamma_j|}$. For the *Student's t* prior $\mathcal{P}(w_j) \propto (b + w_j^2/2)^{-(a+\frac{1}{2})}$, one can have a *Student's t* potential function $\varphi(\gamma) = 1$, when $a, b \to 0$.

For a fixed $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_N]$, we define a relaxed prior which is a joint probability distribution over $\mathbf{w}$ and $\boldsymbol{\gamma}$

$$\begin{aligned}\mathcal{P}(\mathbf{w}; \boldsymbol{\gamma}) &= \prod_{j=1}^{N} \mathcal{N}(w_j|0, \gamma_j)\varphi(\gamma_j) \\ &= \mathcal{P}(\mathbf{w}|\boldsymbol{\gamma})\mathcal{P}(\boldsymbol{\gamma}) \leq \mathcal{P}(\mathbf{w}), \end{aligned} \tag{29}$$

where $\mathcal{P}(\mathbf{w}|\boldsymbol{\gamma}) \triangleq \prod_{j=1}^{N} \mathcal{N}(w_j|0, \gamma_j)$, $\mathcal{P}(\boldsymbol{\gamma}) \triangleq \prod_{j=1}^{N} \varphi(\gamma_j)$.

Now the key question is how to choose the most appropriate $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}} = [\hat{\gamma}_1, \dots, \hat{\gamma}_N]$ to maximise $\prod_{j=1}^{N} \mathcal{N}(w_j|0, \gamma_j)\varphi(\gamma_j)$ such that $\mathcal{P}(\mathbf{w}|\mathbf{y}, \hat{\boldsymbol{\gamma}})$ can be a "good" relaxation to $\mathcal{P}(\mathbf{w}|\mathbf{y})$. Using the product rule for probabilities, we can write the full posterior

$$\begin{aligned}\mathcal{P}(\mathbf{w}, \boldsymbol{\gamma}|\mathbf{y}) &\propto \mathcal{P}(\mathbf{w}|\mathbf{y}, \boldsymbol{\gamma})\mathcal{P}(\boldsymbol{\gamma}|\mathbf{y}) \\ &= \mathcal{N}(\mathbf{m}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{w}}) \times \frac{\mathcal{P}(\mathbf{y}|\boldsymbol{\gamma})\mathcal{P}(\boldsymbol{\gamma})}{\mathcal{P}(\mathbf{y})}. \end{aligned} \tag{30}$$

Since $\mathcal{P}(\mathbf{y})$ is independent of $\boldsymbol{\gamma}$, the quantity

$$\mathcal{P}(\mathbf{y}|\boldsymbol{\gamma})\mathcal{P}(\boldsymbol{\gamma}) = \int \mathcal{P}(\mathbf{y}|\mathbf{w})\mathcal{P}(\mathbf{w}|\boldsymbol{\gamma})\mathcal{P}(\boldsymbol{\gamma})d\mathbf{w}$$

is the prime target for variational methods (Wainwright & Jordan, 2008). This quantity is known as evidence or marginal likelihood. A good way of selecting $\hat{\boldsymbol{\gamma}}$ is to choose it as the minimiser of the sum of the misaligned probability mass, e.g.

$$\begin{aligned}\hat{\boldsymbol{\gamma}} &= \underset{\boldsymbol{\gamma}\geq\mathbf{0}}{\operatorname{argmin}} \int \mathcal{P}(\mathbf{y}|\mathbf{w}) \left|\mathcal{P}(\mathbf{w}) - \mathcal{P}(\mathbf{w}; \boldsymbol{\gamma})\right| d\mathbf{w} \\ &= \underset{\boldsymbol{\gamma}\geq\mathbf{0}}{\operatorname{argmax}} \int \mathcal{P}(\mathbf{y}|\mathbf{w}) \prod_{j=1}^{N} \mathcal{N}(w_j|0, \gamma_j)\varphi(\gamma_j)d\mathbf{w}. \end{aligned} \tag{31}$$

The second equality is a consequence of $\mathcal{P}(\mathbf{w}; \boldsymbol{\gamma}) \leq \mathcal{P}(\mathbf{w})$ (see (29)). The procedure in (31) is referred to as evidence maximisation or type-II maximum likelihood (Tipping, 2001). It means that the marginal likelihood can be maximised by selecting the most probable hyperparameters able to explain the observed data.

Remark 8. By using a Laplace prior (see Remark 7) and the MAP formulation in (27), one can easily obtain the ℓ_1 minimiser in (24), which is a PLS estimate. Therefore, it might be tempting to assume that the Bayesian framework is simply a probabilistic reinterpretation of classical methods since we have just seen that the MAP and PLS estimates are equivalent in the formulation of (27). However, this is not the case. It is sometimes overlooked that the distinguishing element of Bayesian methods is really marginalisation, where instead of seeking to "estimate" all "nuisance" variables in our models, we attempt to integrate them out (Tipping, 2004). In the Bayesian framework, marginal likelihoods have a natural built-in penalty for more complex models. At a certain point, the marginal likelihood will begin to decrease with increasing complexity, and hence, does not intrinsically suffer from the overfitting problems that occur when considering only likelihoods. An intuitive explanation about why the marginal likelihood will begin to decrease with increasing complexity is that, as the complexity of the model increases, the prior will be spread out more thinly across both the "good" models and the "bad" models. Because the marginal likelihood is the likelihood integrated with respect to the prior, spreading the prior across too many models will place too little prior mass on the "good" models, and as a result, cause the marginal likelihood to decrease.

4.2. *Convex relaxation and optimisation for (33)*

We shall now propose an algorithm to compute $\hat{\boldsymbol{\gamma}}$ in (31). From this computed $\hat{\boldsymbol{\gamma}}$ we can obtain an estimation of the posterior mean $\hat{\mathbf{w}}$.

Theorem 2 (Pan et al., 2015). *The optimal hyperparameters $\hat{\boldsymbol{\gamma}}$ in (31) can be achieved by*

$$\hat{\gamma} = \arg\min_{\gamma} \mathcal{L}(\boldsymbol{\gamma}), \tag{32}$$

where

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma}) = \log\left|\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top\right| \\ + \mathbf{y}^\top(\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top)^{-1}\mathbf{y} + \sum_{j=1}^{N} p(\gamma_j), \end{aligned} \tag{33}$$

where $p(\gamma_j) = -2\log\varphi(\gamma_j)$ and $\Gamma = diag\{\boldsymbol{\gamma}\}$. The cost function $\mathcal{L}(\boldsymbol{\gamma})$ is a nonconvex function with respect to $\boldsymbol{\gamma}$.

Before presenting the main results of this section, we introduce an important duality lemma (see Section 4.2 in Jordan, Ghahramani, Jaakkola, & Saul, 1999) which is deeply rooted in convex analysis (Rockafellar, 1996). This duality lemma will be useful for the development of the convex optimisation algorithm in this and the next sections.

Lemma 1. *It is a general fact of convex analysis that a concave function $f(\mathbf{x}) : \mathbb{R}^N \to \mathbb{R}$ can be represented via a* conjugate *or* dual *function as follows $f(\mathbf{x}) = \min_{\mathbf{x}^*}[\langle \mathbf{x}^*, \mathbf{x}\rangle - f^*(\mathbf{x}^*)]$, where the conjugate function f^* can be obtained from the following dual expression: $f^*(\mathbf{x}^*) = \min_{\mathbf{x}}[\langle \mathbf{x}^*, \mathbf{x}\rangle - f(\mathbf{x})]$.*

We can express a nonconvex function $h(\boldsymbol{\gamma})$ as $h(\boldsymbol{\gamma}) = \min_{\boldsymbol{\gamma}^*\geq 0}\langle\boldsymbol{\gamma}^*, \boldsymbol{\gamma}\rangle - h^*(\boldsymbol{\gamma}^*)$, where $h^*(\boldsymbol{\gamma}^*)$ is defined as the concave conjugate of $h(\boldsymbol{\gamma})$ and is given by $h^*(\boldsymbol{\gamma}^*) = \min_{\boldsymbol{\gamma}\geq 0}\langle\boldsymbol{\gamma}^*, \boldsymbol{\gamma}\rangle - h(\boldsymbol{\gamma})$.

Let $h(\boldsymbol{\gamma}) = \log\left|\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top\right| + \sum_{j=1}^{N} p(\gamma_j)$, and assume that $p(\gamma_j)$ is concave with respect to γ_j.[2] Using Lemma 1, we can create a strict upper bounding auxiliary function $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*, \mathbf{w})$ of $\mathcal{L}(\boldsymbol{\gamma})$ in (31),

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*, \mathbf{w}) &\triangleq \langle\boldsymbol{\gamma}^*, \boldsymbol{\gamma}\rangle - h^*(\boldsymbol{\gamma}^*) + \mathbf{y}^\top\left(\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top\right)^{-1}\mathbf{y} \\ &= \frac{1}{\sigma^2}\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 + \sum_{j=1}^{N}\left(\frac{w_j^2}{\gamma_j} + \gamma_j^*\gamma_j\right) - h^*(\boldsymbol{\gamma}^*). \end{aligned} \tag{34}$$

For a fixed $\boldsymbol{\gamma}^*$, we notice that $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*, \mathbf{w})$ is jointly convex in $\mathbf{w}$ and $\boldsymbol{\gamma}$ and can be globally minimised by solving over $\boldsymbol{\gamma}$ and then $\mathbf{w}$. Since $w_j^2/\gamma_j + \gamma_j^*\gamma_j \geq 2w_j\sqrt{\boldsymbol{\gamma}_j^*}$, for any $\mathbf{w}$, $\gamma_j = |w_j|/\sqrt{\gamma_j^*}$ minimises $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*, \mathbf{w})$.

The next step is to find a $\hat{\mathbf{w}}$ that minimises $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*, \mathbf{w})$. When $\gamma_j = |w_j|/\sqrt{\gamma_j^*}$ is substituted into $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*, \mathbf{w})$, $\hat{\mathbf{w}}$ can be obtained by solving the following weighted convex ℓ_1-minimisation problem

$$\begin{aligned}\hat{\mathbf{w}} &= \underset{\mathbf{w}}{\operatorname{argmin}}\left\{\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 + 2\sigma^2\sum_{j=1}^{N} r_j|w_j|\right\} \\ &= \underset{\mathbf{w}}{\operatorname{argmin}}\left\{\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 + 2\sigma^2\sum_{j=1}^{N}\sqrt{\boldsymbol{\gamma}_j^*}|w_j|\right\}, \end{aligned} \tag{35}$$

where $\sqrt{\boldsymbol{\gamma}_j^*}$ are the weights.

[2] This is not a strong assumption since all distributions in Remark 7 satisfy it.

We can then set

$$\gamma_j = \frac{|\hat{w}_j|}{\sqrt{\gamma_j^*}}, \quad \forall j, \tag{36}$$

and, as a consequence, $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*, \mathbf{w})$ will be minimised for any fixed $\boldsymbol{\gamma}^*$.

Now, consider again $\mathcal{L}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*, \mathbf{w})$ in (34). For any fixed $\boldsymbol{\gamma}$ and $\mathbf{w}$, the tightest bound can be obtained by minimising over $\boldsymbol{\gamma}^*$. From the definition of $\boldsymbol{\gamma}^*$, the tightest value of $\boldsymbol{\gamma}^* = \hat{\boldsymbol{\gamma}}^*$ equals the slope at the current $\boldsymbol{\gamma}$ of the function $h(\boldsymbol{\gamma}) \triangleq \log|\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top| + \sum_j p(\gamma_j)$. Using basic principles in convex analysis, we then obtain the following analytic form for the optimiser $\boldsymbol{\gamma}^*$:

$$\begin{aligned}\hat{\boldsymbol{\gamma}}^* &= \nabla_{\boldsymbol{\gamma}}\left(\log|\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top| + \sum_j p(\gamma_j)\right) \\ &= \mathrm{diag}\left[\mathbf{A}^\top\left(\sigma^2\mathbf{I} + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^\top\right)^{-1}\mathbf{A}\right] + p'(\boldsymbol{\gamma}),\end{aligned} \tag{37}$$

where $p'(\boldsymbol{\gamma}) = \left[p'(\gamma_1), \ldots, p'(\gamma_N)\right]^\mathrm{T}$.

The algorithm is then based on successive iterations of (35)–(37) until convergence to $\hat{\boldsymbol{\gamma}}$. We then compute the posterior mean and covariance for the faults as follows:

$$\begin{aligned}\hat{\mathbf{w}} &= \mathbb{E}(\mathbf{w}|\mathbf{y}; \hat{\boldsymbol{\gamma}}) = \hat{\boldsymbol{\Gamma}}\mathbf{A}^\mathrm{T}(\sigma^2\mathbf{I} + \mathbf{A}\hat{\boldsymbol{\Gamma}}\mathbf{A}^\top)^{-1}\mathbf{y}, \\ \boldsymbol{\Sigma}_{\hat{\mathbf{w}}} &= \hat{\boldsymbol{\Gamma}} - \hat{\boldsymbol{\Gamma}}\mathbf{A}^\mathrm{T}(\sigma^2\mathbf{I} + \mathbf{A}\hat{\boldsymbol{\Gamma}}\mathbf{A}^\top)^{-1}\mathbf{A},\end{aligned} \tag{38}$$

where $\hat{\boldsymbol{\Gamma}} = \mathrm{diag}[\hat{\boldsymbol{\gamma}}]$. The above described procedure is summarised in Algorithm 1.

Algorithm 1 Reweighted ℓ_1-minimisation on hyperparameter $\boldsymbol{\gamma}$

Data: Successive observations of $\mathbf{y}$ from t_0 to t_M.
Result: Posterior mean for $\mathbf{w}$.
Step 1 Set iteration count k to zero and initialise each $r_j^{(0)} = \sqrt{\gamma_j^*}$, with randomly chosen initial values for γ_j^*, $\forall j$, e.g. with $\gamma_j^* = 1$, $\forall j$.
Step 2 At the kth iteration, solve the reweighted ℓ_1-minimisation problem

$$\hat{\mathbf{w}}^{(k)} = \underset{\mathbf{w}}{\mathrm{argmin}}\{\|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 + 2\sigma^2\sum_j r_j^{(k)}|w_j|\}, \ \forall j.$$

Step 3 Compute $\gamma_j^{(k)} = \frac{|\hat{w}_j^{(k)}|}{\sqrt{\gamma_j^{*(k)}}}$, $\forall j$.
Step 4 Update $\hat{\boldsymbol{\gamma}^*}^{(k+1)}$ using (37)

$$\hat{\boldsymbol{\gamma}^*}^{(k+1)} = \mathrm{diag}\left[\mathbf{A}^\top\left(\sigma^2\mathbf{I} + \mathbf{A}\Gamma^{(k)}\mathbf{A}^\top\right)^{-1}\mathbf{A}\right] + p'(\boldsymbol{\gamma}^{(k)}).$$

Step 5 Update weights $r_j^{(k+1)}$ for the ℓ_1-minimisation at the next iteration $r_j^{(k+1)} = \sqrt{\hat{\boldsymbol{\gamma}}_j^{*(k+1)}}$.
Step 6 $k \rightarrow k+1$ and iterate Steps 2 to 5 until convergence to some $\hat{\boldsymbol{\gamma}}$.
Step 7 Compute $\hat{\mathbf{w}} = \hat{\Gamma}\mathbf{A}^\mathrm{T}(\sigma^2\mathbf{I} + \mathbf{A}\hat{\Gamma}\mathbf{A}^\top)^{-1}\mathbf{y}$.

It is natural to question the convergence properties of this iterative reweighted ℓ_1-minimisation procedure. Let $\mathcal{A}(\cdot)$ denote a mapping that assigns to every point in $\mathbb{R}_+^N$ the subset of $\mathbb{R}_+^N$ which satisfies Steps 3 and 4 in Algorithm 1. Then the convergence property can be established as follows:

Lemma 2 (*Pan et al., 2015*). *Given the initial point* $\mathbf{a}^{(0)} \in \mathbb{R}_+^N$ *consider the sequence* $\{\mathbf{a}_k\}_{k=0}^\infty$ *obtained by the iterations defined in Steps 3 and 4 of Algorithm 1, i.e. the sequence* $\{\mathbf{a}_k\}_{k=0}^\infty$ *which satisfies* $\mathbf{a}_{k+1} \in \mathcal{A}(\mathbf{a}_k)$. *This sequence is guaranteed to converge to a local minimum (or saddle point) of* $\mathcal{L}_{\boldsymbol{\gamma}}$ *in* (33).

Based on Algorithm 1, we can summarise the fault diagnosis algorithm in Algorithm 2.

Algorithm 2 Diagnosis for faults

```
Set a threshold $\sigma^*$ as indicated in Section 3.2, e.g. $\sigma^* = 10 \times \sigma$;
for $k = 0, \ldots, T$ do
    % $T$ is an integer indicating the number of diagnosis rounds;
    Collect $\xi_i(t_k)$ and $\zeta_i(t_k)$ in (12) and (13)
    for $i = 1, \ldots, N$ do
        Calculate the output data $e_i(t_{k+1})$ in (14);
        Calculate the expected output $e_i^{[\mathbf{e}]}(t_{k+1})$ in (14);
        if $|e_i(t_{k+1}) - e_i^{[\mathbf{e}]}(t_{k+1})| > \sigma^*$ then
            Fault is detected for bus $i$; % {fault detection procedure}
            Compute $y_i(t_{k+1})$ in (19);
            if $|y_i(t_{k+1})| > \sigma^*$ then
                Isolate bus $i$; % {fault isolation procedure}
            end if
        end if
        Set $M \leftarrow k$;
        Apply Algorithm 1 to identify the faults $\hat{\mathbf{w}}_i$; % {fault identification procedure}
    end for
    if $\forall i$, $\|\hat{\mathbf{w}}_i\|_0$ converge to some constant then
        Break;
    end if
end for
An estimate for the faults $\hat{\mathbf{w}}$ in (21), $i = 1, \ldots, n$.
```

Remark 9. If a convex optimisation algorithm is used, no exact zeros will appear in $\hat{\mathbf{w}}$ during the iterations and, strictly speaking, we will typically get a solution with 0-*Sparsity*. However, some of the estimated weights will be very small compared to other weights, e.g. $\pm 10^{-3}$ compared to 1, i.e. the "energy" of the estimated weights will be several orders of magnitude lower than the average "energy", e.g. $\|w_j\|_2^2 \ll \|\mathbf{w}\|_2^2$. Thus a threshold needs to be defined *a priori* to prune the "small" weights at each iteration. An important feature of Algorithm 1 is that it has a low algorithmic complexity since its repeated execution scales as $\mathcal{O}(MN\|\mathbf{w}^{(k)}\|_0)$ (see Candès et al., 2008 and Wipf & Nagarajan, 2010). Since at each iteration certain weights are estimated to be zero, certain dictionary functions spanning the corresponding columns of $\mathbf{A}$ can be pruned out for the next iteration.

5. Numerical study

The effectiveness of our theoretic developments is here illustrated for a randomly generated power network with 20 buses. If all the buses are fully connected, the possible number of transmission lines is 380. We assume that the number of transmission lines is 79 (i.e. we assume that the sparsity of the network is around 20%). Its dynamics can be described by the nonlinear swing equations described in (10) and (11). $w_{ij}^{(1)}$ and $w_{ij}^{(2)}$ are positive real numbers as shown in Fig. 3(a). Let the noise variance $\sigma^2 = 1$. All the parameter values are selected to be similar to those in Kundur et al. (1994) and Pavella, Ernst, and Ruiz-Vega (2000).

Since the sampling frequency is around 50 Hz for the PMU (Kundur et al., 1994; Pavella et al., 2000), we assume the sampling interval to be 20 ms. We thus assume that the discretisation step in Section 3 is performed using a sampling interval $\Delta t = 20$ ms.

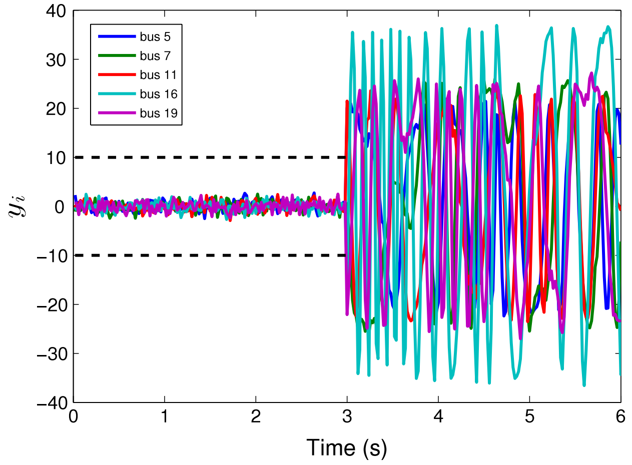


Fig. 1. Time-series of y_i for all buses. The black dashed lines indicate the threshold σ^* in Algorithm 2. The coloured solid lines are the phase angle measurements for bus i, $i = 5, 7, 11, 16, 19$. At time instant $t = 3.02$ s, $|y_5|$, $|y_7|$, $|y_{11}|$, $|y_{16}|$ and $|y_{19}|$ are much greater than σ^* ($\sigma^* = 10$ here). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

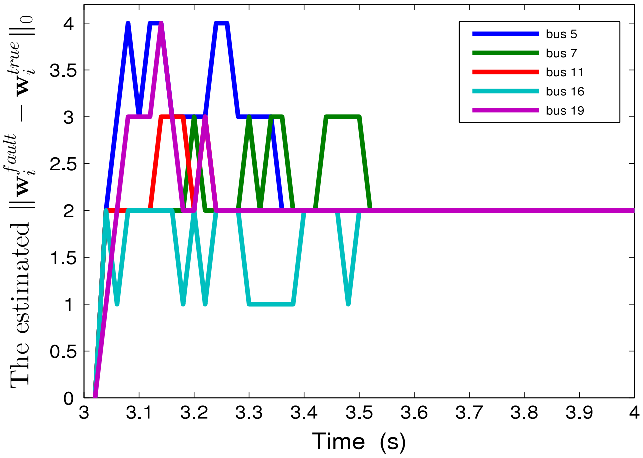
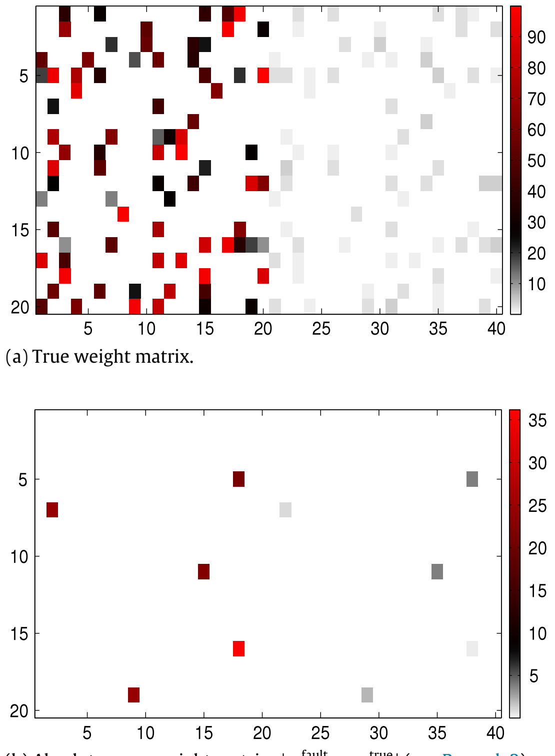


Fig. 2. Time-series of the sparsity of the estimated fault, i.e. $\|\mathbf{w}_i^{\text{fault}} - \mathbf{w}_i^{\text{true}}\|_0$ for bus $i = 5, 7, 11, 16, 19$.

Consider the power networks model in (10) and (11). At time instant $t = 3$ s, there are faults occurring in five transmission lines simultaneously. Specifically, a randomly chosen set of faults can be described as follows: $\forall (i, j) \in \{(5, 18), (7, 2), (11, 15), (16, 18), (19, 9)\}$, $w_{ij}^{(1)}$ and $w_{ij}^{(2)}$ in (6) respectively (which correspond to cos and sin terms) are set to zeros. 5 buses are involved in these transmission lines, i.e. buses 5, 7, 11, 16 and 19. Following the procedure in Algorithm 2, we want to detect and isolate these 5 buses. After detection and isolation, the identification procedure will be performed. We consider $\sigma^* = 10\sigma = 10$ to initialise Algorithm 2.

First, we detect and isolate the buses with $|y_i(t_{k+1})| > \sigma^*$. In Fig. 1, it can be seen that at time instant $t = 3.02$ s (only one sampling time after the faults occur), $|y_5|$, $|y_7|$, $|y_{11}|$, $|y_{16}|$ and $|y_{19}|$ are much greater than σ^* (we set $\sigma^* = 10$ here). Therefore, we can draw the conclusion that buses 5, 7, 11, 16 and 19 are faulty and should be isolated. Next, we identify the faults that occur in the transmission lines connecting the previously isolated buses, i.e. buses 5, 7, 11, 16 and 19. In Fig. 2, the time trajectory of the sparsity of the estimated fault $\|\hat{\mathbf{w}}_i\|_0$, i.e. $\|\mathbf{w}_i^{\text{fault}} - \mathbf{w}_i^{\text{true}}\|_0$ (see Remark 3), for $i = 5, 7, 11, 16, 19$ are depicted starting at the time point $t = 3.02$ s when the faults are detected. We set the pruning threshold (mentioned in Remark 9) to 10^{-3} during the identification procedure of the faults. We define a positive integer n^* to indicate the number of identification rounds which are required to terminate the identification procedure, e.g. $n^* = 10$. As shown in Fig. 2, at time instant $t = 3.52$ s, the sparsity of the estimated fault, i.e. $\|\mathbf{w}_i^{\text{fault}} - \mathbf{w}_i^{\text{true}}\|_0$ for bus $i = 5, 7, 11, 16, 19$ all become equal to 2 and remain unchanged afterwards. At time instant $t = 3.72$ s, only $n^* = 10$ sampling rounds after $t = 3.52$ s, we terminate the identification procedure as the sparsity for all the estimated faults is considered to be stable.



(a) True weight matrix.

(b) Absolute error weight matrix: $|\mathbf{w}_i^{\text{fault}} - \mathbf{w}_i^{\text{true}}|$ (see Remark 3).

Fig. 3. Identification of transmission lines faults: (a) describes the true weight matrix with around 20% nonzero entries. The left half of the matrix corresponds to the weights for cos(·) terms while the right half is for sin(·) terms. (b) Represents the absolute error weight matrix, which is defined as $|\mathbf{w}_i^{\text{fault}} - \mathbf{w}_i^{\text{true}}|$. The nonzero terms in the heat map correspond directly to the faulty transmission lines: (5, 18), (7, 2), (11, 15), (16, 18), (19, 9).

In Fig. 3(a) and (b), we illustrate the true weight matrix and the estimated absolute error matrix $|\mathbf{w}_i^{\text{fault}} - \mathbf{w}_i^{\text{true}}|$. As we can see, all the 5 faults that are occurring in the transmission lines have been identified with high accuracy.

6. Conclusion and discussion

This paper considered the problem of automatic fault diagnosis in large-scale power networks where the buses are described by second-order nonlinear swing equations with process noise. In particular, this work focused on a class of transmission lines faults. We combined tools from compressive sensing and variational Bayesian inference to develop a method to detect, isolate and identify the faults. An illustrative example showed the application of the proposed method to fault diagnosis in nonlinear power networks.

Beyond the results in this paper, some issues still remain for further investigation. This paper assumed that the system is fully measurable. Current work aims to extend the proposed framework to fault diagnosis with partially measured power systems.

Acknowledgements

Mr. Wei Pan acknowledges the support of Microsoft Research through the Ph.D. Scholarship Program. Dr. Ye Yuan and Dr. Jorge Gonçalves acknowledge the support of EPSRC through projects EP/I03210X/1 and EP/G066477/1. Dr. Henrik Sandberg acknowledges the support from the Swedish Research Council through grants 2009-4565 and 2013-5523 and the Swedish Foundation for Strategic Research through the project ICT-Psi. Dr. Guy-Bart Stan gratefully acknowledges the support of the EPSRC Centre for Synthetic Biology and Innovation at Imperial College London through the Science and Innovation award EP/G036004/1.

References

Bishop, C. (2006). *Pattern recognition and machine learning, Vol. 4*. New York: springer.

Candès, E., Romberg, J., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, *59*(8), 1207–1223.

Candès, E., & Tao, T. (2005). Decoding by linear programming. *IEEE Transactions on Information Theory*, *51*(12), 4203–4215.

Candès, E., Wakin, M., & Boyd, S. (2008). Enhancing sparsity by reweighted ℓ_1 minimisation. *Journal of Fourier Analysis and Applications*, *14*(5), 877–905.

Dai, W., & Milenkovic, O. (2009). Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, *55*(5), 2230–2249.

Ding, S. X. (2008). *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer.

Dong, H., Wang, Z., & Gao, H. (2012). Fault detection for Markovian jump systems with sensor saturations and randomly varying nonlinearities. *Circuits and Systems I: Regular Papers, IEEE Transactions on,*, *59*(10), 2354–2362.

Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, *52*(4), 1289–1306.

Donoho, D., & Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proceedings of the National Academy of Sciences*, *100*(5), 2197–2202.

Hines, P., Balasubramaniam, K., & Sanchez, E. C. (2009). Cascading failures in power grids. *IEEE Potentials*, *28*(5), 24–30.

Jiang, J., Yang, J., Lin, Y., Liu, C., & Ma, J. (2000). An adaptive PMU based fault detection/location technique for transmission lines. I. Theory and algorithms. *IEEE Transactions on Power Delivery*, *15*(2), 486–493.

Jordan, M., Ghahramani, Z., Jaakkola, T., & Saul, L. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*(2), 183–233.

Kundur, P., Balu, N. J., & Lauby, M. G. (1994). *Power system stability and control, Vol. 4*. New York: McGraw-hill, no. 2.

Mohajerin Esfahani, P., Vrakopoulou, M., Andersson, G., & Lygeros, J. (2012). A tractable nonlinear fault detection and isolation technique with application to the cyber-physical security of power systems. In *2012 IEEE 51st annual conference on decision and control, CDC, December* (pp. 3433–3438).

Němcová, J. (2010). Structural identifiability of polynomial and rational systems. *Mathematical Biosciences*, *223*(2), 83–96.

Palmer, J., Wipf, D., Kreutz-Delgado, K., & Rao, B. (2006). Variational EM algorithms for non-Gaussian latent variable models. *Advances in Neural Information Processing Systems*, *18*, 1059.

Pan, W., Yuan, Y., Gonçalves, J., & Stan, G. B. (2015). A sparse Bayesian approach to the identification of nonlinear state-space systems. *IEEE Transaction on Automatic Control*, arXiv:1408.3549.

Pavella, M., Ernst, D., & Ruiz-Vega, D. (2000). *Transient stability of power systems: a unified approach to assessment and control*. Springer.

Phadke, A., & Thorp, J. (2008). *Synchronized phasor measurements and their applications*. Springer.

Rockafellar, R. (1996). *Convex analysis, Vol. 28*. Princeton university press.

Shahidehpour, M., Tinney, F., & Fu, Y. (2005). Impact of security on power systems operation. *Proceedings of the IEEE*, *93*(11), 2013–2025.

Shames, I., Teixeira, A. M., Sandberg, H., & Johansson, K. (2011). Distributed fault detection for interconnected second-order systems. *Automatica*, *47*(12), 2757–2764.

Tate, J. E., & Overbye, T. J. (2008). Line outage detection using phasor angle measurements. *IEEE Transactions on Power Systems*, *23*(4), 1644–1652.

Tipping, M. (2001). Sparse Bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, *1*, 211–244.

Tipping, M. (2004). Bayesian inference: an introduction to principles and practice in machine learning. In *Advanced lectures on machine learning* (pp. 41–62). Berlin, Heidelberg: Springer.

Wainwright, M., & Jordan, M. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, *1*(1–2), 1–305.

Wipf, D., & Nagarajan, S. (2010). Iterative reweighted l1 and l2 methods for finding sparse solutions. *IEEE Journal of Selected Topics in Signal Processing*, *4*(2), 317–329.

Yin, S., Ding, S., Haghani, A., Hao, H., & Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control*, *22*(9), 1567–1581.

Zhang, Q., Zhang, X., Polycarpou, M. M., & Parisini, T. (2014). Distributed sensor fault detection and isolation for multimachine power systems. *International Journal of Robust and Nonlinear Control*, *24*(8–9), 1403–1430.

Wei Pan is a Ph.D. candidate at Department of Bioengineering, Imperial College London. He received his Bachelor's degree in Automation from Harbin Institute of Technology and Master's degree in Biomedical Engineering from University of Science and Technology of China. He is interested in nonlinear time series modelling. He is interested in nonlinear time series modelling with applications in biology and quantitative finance. He is the recipient of Dorothy Hodgkin Postgraduate Awards and Microsoft Research Ph.D. Scholarship.

Ye Yuan received his B.Eng. degree (Valedictorian) from the Department of Automation, Shanghai Jiao Tong University in 9.2008, M. Phil. and Ph.D. from the Department of Engineering, Cambridge University in 10.2009 and 2.2012 respectively. He is now a Junior Research Fellow in Darwin College, University of Cambridge and has been holding visiting researcher positions in Caltech, MIT and Imperial College London. His research interests include the unification system identification and machine learning with applications to natural and engineering systems. He is the recipient of Dorothy Hodgkin Postgraduate Awards, Microsoft Research Ph.D. Scholarship, Cambridge Overseas Scholarship, Chinese Government Award for Outstanding Students Abroad and Henry Lester Scholarship, Best Paper Finalist in IEEE ICIA.

Henrik Sandberg received the M.Sc. degree in Engineering Physics and the Ph.D. degree in Automatic Control from Lund University, Lund, Sweden, in 1999 and 2004, respectively. He is an Associate Professor with the Department of Automatic Control, KTH Royal Institute of Technology, Stockholm, Sweden. From 2005 to 2007, he was a Post-Doctoral Scholar with the California Institute of Technology, Pasadena, USA. In 2013, he was a visiting scholar at the Laboratory for Information and Decision Systems (LIDS) at MIT, Cambridge, USA. He has also held visiting appointments with the Australian National University and the University of Melbourne, Australia. His current research interests include secure networked control, power systems, model reduction, and fundamental limitations in control. He was a recipient of the Best Student Paper Award from the IEEE Conference on Decision and Control in 2004 and an Ingvar Carlsson Award from the Swedish Foundation for Strategic Research in 2007. He is currently an Associate Editor of the IFAC Journal Automatica.

Jorge Gonçalves received his Licenciatura (5-year S.B.) degree from the University of Porto, Portugal, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA, all in Electrical Engineering and Computer Science, in 1993, 1995, and 2000, respectively. He then held two postdoctoral positions, first at the Massachusetts Institute of Technology for seven months, and from 2001 to 2004 at the California Institute of Technology with the Control and Dynamical Systems Division. At the Information Engineering Division of the Department of Engineering, University of Cambridge he was a Lecturer from 2004 until 2012, a Reader from 2012 until 2014, and since 2014 he is a Principal Research Associate. From 2005 until 2014 he was a Fellow of Pembroke College, University of Cambridge. From June to December 2010 and January to September 2011 he was a visiting researcher at the University of Luxembourg and California Institute of Technology, respectively. Since 2013 he is a Professor at the Luxembourg Centre for Systems Biomedicine, University of Luxembourg.

Guy-Bart Stan is a Reader in Engineering Design for Synthetic Biology and the head of the "Control Engineering Synthetic Biology" group at the Department of Bioengineering of Imperial College London. He is the recipient of the very prestigious UK Engineering and Physical Sciences Research Council (EPSRC) Fellowship for Growth in Synthetic Biology, directly supporting his research from February 2015 until January 2020. He received his Ph.D. in Applied Sciences (nonlinear dynamical systems and control) from the University of Liège, Belgium in March 2005 and subsequently worked for Philips Applied

Technologies, Leuven, Belgium. From January 2006 until December 2009, he worked as Research Associate in the Control Group of the University of Cambridge, first supported by a Marie Curie Intra-European Fellowship and then by the UK EPSRC.

His current research focus is on the study of core engineering design principles of complex dynamical systems including biological systems and complex networks, and on the development of mathematical modelling, analysis and systems and control engineering methods for such systems. He is author of over 60 peer-reviewed papers and 1 book, and editor of a 2 volumes book on the use of rigorous systems and control engineering methods for solving important problems in systems biology, synthetic biology and complex physical systems.