Müller, Christine H.; Schorning, Kirsten



# A-optimal designs for state estimation in networks

Statistical Papers

**Provided in Cooperation with:**
Springer Nature

*Suggested Citation:* Müller, Christine H.; Schorning, Kirsten (2023) : A-optimal designs for state estimation in networks, Statistical Papers, ISSN 1613-9798, Springer, Berlin, Heidelberg, Vol. 64, Iss. 4, pp. 1159-1186, https://doi.org/10.1007/s00362-023-01435-y

This Version is available at:
https://hdl.handle.net/10419/308575

Statistical Papers (2023) 64:1159–1186
https://doi.org/10.1007/s00362-023-01435-y

REGULAR ARTICLE

# *A*-optimal designs for state estimation in networks

**Christine H. Müller**[1] · **Kirsten Schorning**[1]

Received: 16 December 2022 / Revised: 7 March 2023 / Published online: 24 March 2023


**Abstract**

We consider two models for estimating the expected states of nodes in networks where the observations at nodes are given by random states and measurement errors. In the first model, we assume independent successive observations at the nodes and the design question is how often the nodes should be observed to obtain a precise estimation of the expected states. In the second model, all nodes are observed simultaneously and the design question is to determine the nodes which need larger precision of the measurements than other nodes. Both models lead to the same design problem. We derive explicitly $A$-optimal designs for the most simple network with star configuration. Moreover, we consider the network with wheel configuration and derive some conditions which simplify the numerical calculation of the corresponding $A$-optimal designs.

**Keywords** $A$-optimal designs · Random state models · Network analysis

## 1 Introduction

The design problem addressed in this paper is motivated by a cooperation with electrical engineers who study electrical power distribution grids of medium and low-voltage levels. In a specific distribution grid the question rises where measurements of the electrical power should be taken and how precise these measurements should be in order to get a precise estimation of the state of the grid. Due to high costs, it is not possible to use sensors for measuring the electrical power at each position of the grid and at some positions so-called pseudo measurements have to be used instead, see e.g. Muscas et al. (2014), Schlösser et al. (2014) or Schurtz (2020). These pseudo measurements are typically obtained from historical data or weather data to estimate, for example, the

✉ Kirsten Schorning
schorning@statistik.tu-dortmund.de

Christine H. Müller
cmueller@statistik.tu-dortmund.de

[1] Department of Statistics, TU Dortmund University, 44227 Dortmund, Germany

needed heating energy of a household or the produced energy of a photovoltaic system. It is obvious that these pseudo measurements are less precise than the measurements of sensors. Moreover, they can vary in their precision. For example, a temperature measurement close to a node of the grid is a more precise estimate of the electrical power at this node than a temperature measurement further away. However, the more precise a measurement is the more expensive it is. Hence having cost constrains, the design problem is to determine the necessary measurement precision at the nodes, which is connected to the problem of sensor allocation.

The problem of sensor allocation in distributed systems has often been addressed in research during the past 50 years, see for example the surveys in Kubrusly and Malebranche (1985), Uciński (2022), Duan et al. (2022) or the book of Uciński (2004). In most of the considered cases, methods are developed to minimize a function of the covariance matrix of an appropriate estimator of the system, see e.g. Uciński (2000) and Singhal and Michailidis (2008). In particular, Patan and Patan (2005) use partial differential equations and its simplification to a non-linear models in combination with a steepest descent method to find optimal weights of given support points (i.e. the sensor positions), whereas Uciński (2022) addresses the best selection of sensors in order to obtain a proper estimation of subsets of unknown parameters of a spatiotemporal system modelled by a partial differential equation. The experimental design problem for state estimation in electrical power grids is especially treated for example by Li et al. (2011), Xygkis et al. (2018), and Cao et al. (2022). However, they all solve the problem of allocating $F$ sensor or generator positions out of $E > F$ possible positions by greedy algorithms since the number of positions is high. Azhdari and Ardakan (2022) modify this problem by allocating $E$ components into $F$ groups where each group belongs to a node of the network.

All of these approaches deal with large networks so that approximate solutions can only be found numerically. Moreover, the aim in electrical grids is to estimate unknown expected states at certain positions of the grid. In the present paper, we simplify this state estimation problem so that exact optimal solutions can be found. For this purpose, we consider two specific models: In a first scenario, called Model A, we assume independent univariate observations given by random states and additive measurement errors at given nodes of the grid where the variances of the random states are equal and the same holds for the measurement errors. In this situation, the design problem is given by the question of allocating the observations at the different nodes. However, this approach does not treat the possibility of different precision of measurements at nodes. Moreover, the assumption of independent univariate observations is unrealistic in electrical grids. As soon as several sensors (including pseudo measurements) exist, one would use the simultaneous observations at the sensors placed at the different nodes. Hence, we consider a second scenario, called Model B, where independent simultaneous observations are available at the given nodes of the network for the different observation time points. Consequently, a single observation is a vector given by the random states of the nodes with additive measurement errors consisting of variances depending on the nodes. Here, the design question is at which nodes the variance of the measurement could be high and where not. This concerns the question at which nodes less precise pseudo measurements are sufficient, and where more precise measurements of sensors are necessary. In the following, we show that the design

problem of Model B coincides with the design problem in Model A if nonrandom states are assumed in Model A.

The paper is organized as follows. Section 2 presents the two simple models and how they are related to each other. Section 3 shows how a general result concerning $A$-optimal designs with minimum support can be used to derive $A$-optimal designs in the two models analytically. This result is applied in Sect. 4 to the most simple network, a so-called star network, which nevertheless is often considered in studies for electrical power distribution grids, see e.g. Su and Wang (2020) and Azhdari and Ardakan (2022). In particular, in Sects. 4.2 and 4.3 we study the situation where the whole expected state vector is not identifiable. In Sect. 5, we consider an extension of the star network, given by a wheel. In particular, we derive sufficient conditions for the identifiability of the state vector, which can also be used to reduce the numerical complexity for that type of network. Finally, some extensions of the presented approach are discussed in Sect. 6.

## 2 Simple models for state estimation in networks

We consider a network with $I+1$ nodes $0, \ldots, I$, where node 0 denotes a central node or outgoing node of the electrical power distribution grid. The expected observations $\mathbb{Y}_0, \mathbb{Y}_1, \ldots, \mathbb{Y}_I$ at these nodes depend on the unkown expected states $s_0, s_1, \ldots, s_I$ of the different nodes in the network. The aim is the estimation of these states, contained in the state vector $s = (s_0, s_1, \ldots, s_I)^\top \in \mathbb{R}^{I+1}$ or an appropriate linear aspect $L\,s$ with $L \in \mathbb{R}^{q\times(I+1)}$ using the observation vector $\mathbb{Y} = (\mathbb{Y}_0, \mathbb{Y}_1, \ldots, \mathbb{Y}_I)^\top$. In the situation under consideration, the expected observation $\mathbb{Y}_i$ at a particular node $i$ is both influenced by the corresponding expected state $s_i$ and by the expected states $s_j$ of other nodes $j \neq i$, that are connected to node $i$ ($i = 0, \ldots, I$). More precisely, let $x_{ij}$ be the influence of the state $s_j$ at node $j$ on the expected observation $\mathbb{Y}_i$ taken at a particular node $i$ ($i = 0, \ldots, I$) and denote the matrix storing these influences by $\mathbb{X} = (x_{ij})_{i,j=0,\ldots,I} \in \mathbb{R}^{(I+1)}$. Then the expected observation vector $\mathbb{Y} = (\mathbb{Y}_0, \mathbb{Y}_1, \ldots, \mathbb{Y}_I)^\top$ is given by

$$\mathbb{Y} = \mathbb{X}\,s\,.$$

The matrix $\mathbb{X} \in \mathbb{R}^{(I+1)\times(I+1)}$ is called influence matrix of the network, as it describes the influence of the states on the observations at the different nodes. Note that $\mathbb{X}$ is strongly connected to the adjacency matrix of a network with weighted edges: if the diagonal elements of $\mathbb{X}$ are removed, the resulting matrix describes the structure of the network, where two nodes $i \neq j$ are connected with an edge weighted by $x_{ij}$ if $x_{ij} \neq 0$.
Denoting the $(i+1)$-th unit vector in $\mathbb{R}^{I+1}$ by $u_i$, the observation $y_i$ at node $c_i$ can be rewritten by $u_i^\top \mathbb{X} s$.

Later in the paper, we restrict ourselves to the case, where the influence of the state $s_i$ on the observation at node $i$ is given by $a > 0$, whereas the influence of the states $s_j$ ($j \neq i$) of the adjacent nodes on the expected observation at node $i$ is equal to $b > 0$.

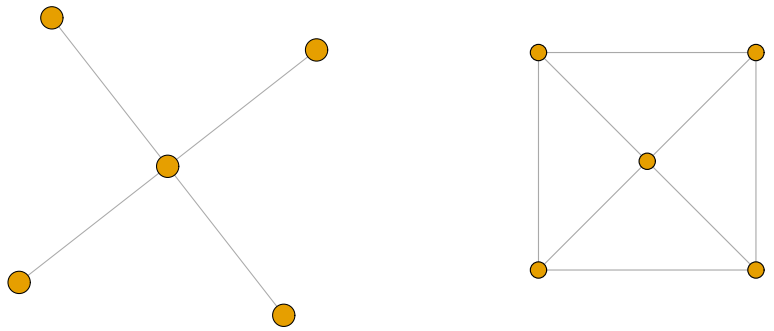

**Fig. 1** Left figure displays a star network consisting of $I + 1 = 5$ nodes, right figure displays a wheel network consisting of $I + 1 = 5$ nodes

Then, the influence matrix $\mathbb{X}$ is of the form

$$\mathbb{X} = a\mathbb{I}_{(I+1)\times(I+1)} + b\mathbb{A} \tag{1}$$

where $\mathbb{I}_{(I+1)}$ denotes the $(I + 1)$-dimensional identity matrix and $\mathbb{A} \in \{0, 1\}^{(I+1)\times(I+1)}$ is the adjacency matrix of the considered (unweighted) network. Moreover, the expected observation $\mathbb{Y}_i$ at node $i$ can be written by

$$\mathbb{Y}_i = u_i^\top \mathbb{X} s = a + b \sum_{j \text{ is connected with node } i} s_j \,. \tag{2}$$

***Example 1*** 1. *Star-Network*. Let node 0 be the center of the network, which is connected to the other nodes $1, \ldots, I$ (see left panel of Fig. 1 for $I = 4$). Let $a$ be the influence of the state $s_i$ on the expected observation at the corresponding node $i$ ($i = 0, \ldots, I$), whereas $b$ denotes the influence of the states $s_j$ ($j \neq i$) of the adjacent nodes on the expected observation taken at node $i$. Using (1), the influence matrix $\mathbb{X}$ is given by

$$\mathbb{X} = \begin{pmatrix} a & b\mathbf{1}_I^\top \\ b\mathbf{1}_I & a\mathbb{I}_{I\times I} \end{pmatrix}, \tag{3}$$

where $\mathbf{1}_I = (1, \ldots, 1)^\top \in \mathbb{R}^I$ and $\mathbb{I}_I$ denotes the $I$-dimensional identity matrix. Consequently, the expected observation $\mathbb{Y}_0$ obtained at the central node 0 is given by

$$\mathbb{Y}_0 = u_0^\top \mathbb{X} s = as_0 + \sum_{i=1}^{I} bs_i,$$

whereas the expected observations at the non-central nodes are of the form

$$\mathbb{Y}_i = u_i^\top \mathbb{X} s = bs_0 + as_i, \quad i = 1, \ldots, I \,.$$

2. *Wheel-Network*. Let node 0 be again the center of the network, which is connected to all other nodes of the network. Moreover, the remaining nodes are connected to two

others nodes (see right panel of Fig. 1 for case $I = 4$). Similar to the situation of the star network, let $a$ be the influence of the state $s_i$ on the expected observation at the corresponding node $i$ ($i = 0, \ldots, I$), whereas $b$ denotes the influence of the states $s_j$ ($j \neq i$) of the adjacent nodes on the expected observation taken at node $i$. Using (1), the influence matrix $\mathbb{X}$ is given by

$$\mathbb{X} = \begin{pmatrix} A & B^\top \\ B & \tilde{\mathbb{X}} \end{pmatrix}, \tag{4}$$

where the matrices $A \in \mathbb{R}^{2\times 2}$ and $B \in \mathbb{R}^{(I-1)\times 2}$ are of the form

$$A = \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \qquad B = b\left(\mathbf{1}_{(I-1)},\, u_1^{I-1} + u_{I-1}^{I-1}\right), \tag{5}$$

where $u_j^{I-1}$ denotes the $j$-th unit vector in $\mathbb{R}^{I-1}$. The matrix $\tilde{\mathbb{X}} \in \mathbb{R}^{(I-1)\times(I-1)}$ is a triadiagonal matrix with main diagonal elements equal to $a$, whereas the lower and upper diagonal elements are equal to $b$, that is

$$\tilde{\mathbb{X}}_{i,j} = \begin{cases} a, & i = j \\ b, & i = j-1 \text{ or } i = j+1 \\ 0, & \text{else} \end{cases}. \tag{6}$$

Based on the structure of the network and on the notation introduced beforehand, the expected observation at the central node is again given by

$$\mathbb{Y}_0 = as_0 + \sum_{i=1}^{I} bs_i,$$

whereas the expected observations at the non-central nodes are of the form

$$\begin{aligned} \mathbb{Y}_i &= u_i^\top \mathbb{X} s = bs_0 + bs_{i-1} + bs_{i+1} + as_i, \quad i = 2, \ldots, I-1, \\ \mathbb{Y}_1 &= u_1^\top \mathbb{X} s = bs_0 + bs_2 + bs_I + as_1, \\ \mathbb{Y}_I &= u_I^\top \mathbb{X} s = bs_0 + bs_{I-1} + bs_1 + as_I. \end{aligned}$$

In practice, observations of the form $\mathbb{Y}$ given in (1) are not available: On the one hand the expected observations $\mathbb{Y}$ might be corrupted by random measurement errors, on the other hand the states at the different nodes might not be fixed to $s$, but also random. Consequently, the vector $s$ only describes the expected state of the network. Nevertheless, the aim of the present paper is to estimate the unknown expected state vector $s$ or a linear aspect $L\,s$ of it using random observations $Y_1, \ldots, Y_N$ at the different nodes of the network. For this purpose, we introduce two different models, called Model A and Model B.

### 2.1 Model A

In the first scenario, we assume that at each time point $n \in \{1, \ldots, N\}$, one observation $Y_n$ at one particular node $i(n) \in \{0, \ldots, I\}$ is available, where $Y_n$ is a linear combination of the random states vector $S_n = (S_{0,n}, \ldots, S_{I,n})^\top$ of the network at that time point and an additive measurement error $E_n$. Furthermore, the distance between two consecutive time points is assumed to be sufficiently large, so that we assume that $Y_1, \ldots, Y_N$ are successive independent univariate observations at different nodes of the network.

Under the assumption that the random state vector $S_n$ at time point $n$ is of the form

$$S_n = s + Z_n,$$

where $s$ is the expected state vector of the network and $Z_1, \ldots, Z_N$ are independent random vectors with mean 0 and covariance matrix $\rho_z \sigma^2 \mathbb{I}_{(I+1)\times(I+1)}$, $\rho_z \geq 0, \sigma^2 \geq 0$, the $n$-th observation $Y_n$ at the node $i(n)$ is given by

$$Y_n = u_{i(n)}^\top \mathbb{X} S_n + E_n = u_{i(n)}^\top \mathbb{X} s + u_{i(n)}^\top \mathbb{X} Z_n + E_n, \tag{7}$$

where $\mathbb{X}$ is the influence matrix, $u_i$ is the $(i+1)$-th unit vector, and the independent measurement errors $E_1, \ldots, E_N$ have mean 0 and variance $\rho_E \sigma^2$. The random elements $E_1, \ldots, E_N$ and $S_1, \ldots, S_N$ are also assumed to be independent. Choosing either $\rho_E = 0$ or $\rho_Z = 0$, we obtain either a model without measurement errors or a model with non-random states at the different nodes, respectively.

The variance of an observation $Y_n$ in model (7) is given by

$$\operatorname{var}(Y_n) = \operatorname{var}(u_{i(n)}^\top \mathbb{X} Z_n + E_n) = \sigma^2 \sigma_{i(n)}^2$$

where the variance $\sigma_{i(n)}^2$ at node $i(n)$ is of the form

$$\sigma_{i(n)}^2 := u_{i(n)}^\top \mathbb{X} \mathbb{X}^\top u_{i(n)} \rho_Z + \rho_E, \quad n = 1, \ldots, N\,. \tag{8}$$

Let $\mathbb{D} := \operatorname{diag}(\sigma_0, \ldots, \sigma_I)$, where $\operatorname{diag}(\sigma_0, \ldots, \sigma_I)$ denotes the diagonal matrix with diagonal elements $\sigma_0, \ldots, \sigma_I$. Using $\frac{1}{\sigma_{i(n)}} u_{i(n)}^\top \mathbb{X} s = u_{i(n)}^\top \mathbb{D}^{-1} \mathbb{X} s$, we define transformed random variables $\widetilde{Y}_n$ by

$$\widetilde{Y}_n := \frac{1}{\sigma_{i(n)}} Y_n = u_{i(n)}^\top \mathbb{D}^{-1} \mathbb{X} s + \widetilde{E}_n\,, \quad n = 1, \ldots, N\,, \tag{9}$$

where $\widetilde{E}_1, \ldots, \widetilde{E}_N$ are independent with mean 0 and variance $\sigma^2$. Note that the model given in (9) is a linear model with homescedastic errors, where the experimental condition at time point $n$ is given by node $i(n)$, $n = 1, \ldots, N$. Hence, setting $\widetilde{Y} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_N)^\top$, $\widetilde{E} = (\widetilde{E}_1, \ldots, \widetilde{E}_N)^\top$, $d = (i(1), \ldots, i(N))$, $\mathbb{U}_d = (u_{i(1)}, \ldots, u_{i(N)})^\top$ and $X_d = \mathbb{U}_d\, \mathbb{D}^{-1} \mathbb{X}$, we obtain

$$\widetilde{Y} = \mathbb{U}_d\, \mathbb{D}^{-1} \mathbb{X}\, s + \widetilde{E} = X_d\, s + \widetilde{E}\,, \tag{10}$$

where the best linear unbiased estimator for an aspect $L\,s$ in (10) is given by

$$L\,\widehat{s}(\widetilde{Y}) = L\,(X_d^\top X_d)^- X_d^\top \widetilde{Y} = L\,(\mathbb{X}^\top \mathbb{D}^{-1} \mathbb{U}_d^\top \mathbb{U}_d\, \mathbb{D}^{-1} \mathbb{X})^- \mathbb{X}^\top \mathbb{D}^{-1} \mathbb{U}_d^\top \widetilde{Y}\,.$$

The corresponding covariance matrix of that estimator is

$$\mathrm{Cov}(L\,\widehat{s}(\widetilde{Y})) = L\,(X_d^\top X_d)^- L^\top = \frac{1}{N} L\,(\mathbb{X}^\top \mathbb{D}^{-1} \mathbb{D}_d\, \mathbb{D}^{-1} \mathbb{X})^- L^\top\,,$$

where $\mathbb{D}_d = \frac{1}{N}\,\mathbb{U}_d^\top \mathbb{U}_d = \mathrm{diag}(\delta_0, \delta_1, \ldots, \delta_I)$ with $\delta_i = \frac{1}{N}\sharp\{n;\ i(n) = i\}$ for $i = 0, 1, \ldots, I$. Note that $\delta_i$ is equal to the relative amount observations taken at node $i$, $i = 0, \ldots, I$. In order to use the established methods of optimal design theory for approximate designs, we further relax the condition on the values of $\delta_0, \delta_1, \ldots, \delta_I$ and assume that

$$\delta = (\delta_0, \ldots, \delta_I) \in \Delta := \{\delta = (\delta_0, \delta_1, \ldots, \delta_I)^\top \in [0, 1]^{I+1};\ \sum_{i=0}^{I} \delta_i = 1\}\,, \quad (11)$$

where the set $\Delta$ denotes the set of all approximate designs $\delta$ with support at the nodes $0, \ldots, I$. If an approximate design $\delta$ is given and $N$ observations can be taken, a rounding procedure is applied to obtain integers $n_0, \ldots, n_I$ from the not necessarily integer valued quantities $\delta_i N$ (see Pukelsheim and Rieder (1992)). Then, the design problem reduces to the determination of an approximate design $\delta = (\delta_0, \ldots, \delta_I) \in \Delta$ such that the covariance matrix $\mathrm{Cov}(L\,\widehat{s}(\widetilde{Y}))$ becomes small in some sense. Since the interest lies in estimating $L\,s$, we are interested in determining the widely used $A$-optimal designs. More precisely, following Pukelsheim (2006), p. 137, a design $\delta^* \in \Delta$ is called $A$-optimal, if it minimizes the trace of the covariance matrix, i.e.

$$\delta^* \in \arg\min\{\mathrm{tr}\ L\,(\mathbb{X}^\top \mathbb{D}^{-1} \mathbb{D}_\delta\, \mathbb{D}^{-1} \mathbb{X})^- L^\top;\ \delta \in \Delta\}\,, \quad (12)$$

with $\mathbb{D}_\delta = \mathrm{diag}(\delta_0, \ldots, \delta_I)$. In the case of non-random states at the different nodes (i.e. $\rho_z = 0$), we set $\rho_E = 1$ without loss of generality and the design problem stated in (12) reduces to

$$\delta^* \in \arg\min\{\mathrm{tr}\ L\,(\mathbb{X}^\top \mathbb{D}_\delta\, \mathbb{X})^- L^\top;\ \delta \in \Delta\}. \quad (13)$$

### 2.2 Model B

For electrical power distribution grids, it is more realistic to assume that for each time point $n \in \{1, \ldots, N\}$, the observation $Y_n$ consists of simultaneous observations at all nodes $i = 0, 1, \ldots, I$ of the network. Hence, $Y_n$ is a $(I + 1)$-dimensional random vector. If the distance between two consecutive time points is sufficiently large, we can still assume that $Y_1, \ldots, Y_N$ are independent random vectors. With the notation of the previous section, the $n$-th observation is a $(I + 1)$-dimensional random vector

$Y_n$ of the form

$$Y_n = \mathbb{X}\, S_n + E_n,$$

with

$$S_n = s + Z_n\,,$$

where the measurement errors $E_1, \ldots, E_N$ and random effects $Z_1, \ldots, Z_N$ are independent random $(I+1)$-dimensional random vectors with mean vector $0_{I+1}$. Additionally, we assume that the components of the measurement error $E_n$ are independent, whereas the entries of the random effect $Z_n$ might be correlated indicating a dependence between the states of different nodes. More precisely, the covariance matrices of $E_n$ and $Z_n$ are assumed to be of the form

$$\mathrm{Cov}(Z_n) = \sigma^2\, \mathbb{D}_Z \text{ with } \mathbb{D}_Z \text{ positive-definite}, \tag{14}$$

$$\mathrm{Cov}(E_n) = \sigma^2\, \mathbb{D}_E \text{ with } \mathbb{D}_E := \mathrm{diag}(\sigma_{0E}^2, \sigma_{1E}^2, \ldots, \sigma_{IE}^2), \tag{15}$$

where the entries of $\mathbb{D}_E$ indicate the different accuracies with which the states are measured at the different nodes.
Then, the covariance matrix of of the observation $Y_n$ is given by

$$\mathrm{Cov}(Y_n) = \mathrm{Cov}(\mathbb{X}\, s + \mathbb{X}\, Z_n + E_n) = \sigma^2\, \mathbb{S} \text{ with } \mathbb{S} := \mathbb{X}\, \mathbb{D}_Z\, \mathbb{X}^\top + \mathbb{D}_E.$$

Since $Y_1, \ldots, Y_N$ are independent, the covariance matrix of the vector of all available observations $Y = (Y_1^\top, \ldots, Y_N^\top)^\top$ satisfies

$$\mathrm{Cov}(Y) = \sigma^2\, \mathbb{S}_* \text{ with } \mathbb{S}_* := \mathbb{I}_{N\times N} \otimes \mathbb{S},$$

where $\otimes$ denotes the Kronecker product and $\mathbb{I}_{N\times N}$ is the $N \times N$ identity matrix. Transforming the vector of observations by

$$\begin{aligned} \widetilde{Y} &:= \mathbb{S}_*^{-1/2}\, Y \\ &= \mathbb{I}_{N\times N} \otimes \mathbb{S}^{-1/2} \begin{pmatrix} \mathbb{X}\, s + \mathbb{X}\, Z_1 + E_1 \\ \vdots \\ \mathbb{X}\, s + \mathbb{X}\, Z_N + E_N \end{pmatrix} = (1_N \otimes \mathbb{S}^{-1/2}\, \mathbb{X})\, s + \widetilde{E} \end{aligned}$$

with

$$\widetilde{E} := \begin{pmatrix} \mathbb{S}^{-1/2}(\mathbb{X}\, Z_1 + E_1) \\ \vdots \\ \mathbb{S}^{-1/2}(\mathbb{X}\, Z_N + E_N) \end{pmatrix},\ \mathrm{Cov}(\widetilde{E}) = \sigma^2 \mathbb{I}_{N\times N} \otimes \mathbb{I}_{(I+1)\times(I+1)},$$

we obtain a linear model with homescedastic errors. Hence, the best linear unbiased estimator for $L\,s$ is given by

$$L\,\widehat{s}(\widetilde{Y}) = L\,\left((1_N \otimes \mathbb{S}^{-1/2}\,\mathbb{X})^\top (1_N \otimes \mathbb{S}^{-1/2}\,\mathbb{X})\right)^-\,(1_N \otimes \mathbb{S}^{-1/2}\,\mathbb{X})^\top \widetilde{Y}.$$

The corresponding covariance matrix is of the form

$$\begin{aligned}\mathrm{Cov}(L\,\widehat{s}(\widetilde{Y})) &= L\,\left((1_N \otimes \mathbb{S}^{-1/2}\,\mathbb{X})^\top (1_N \otimes \mathbb{S}^{-1/2}\,\mathbb{X})\right)^- L^\top \\ &= L\,\left(1_N^\top 1_N \otimes \mathbb{X}^\top \mathbb{S}^{-1}\,\mathbb{X}\right)^- L^\top = \frac{1}{N} L\,\left(\mathbb{X}^\top (\mathbb{X}\,\mathbb{D}_Z\,\mathbb{X}^\top + \mathbb{D}_E)^{-1}\,\mathbb{X}\right)^- L^\top.\end{aligned}$$

If the influence matrix $\mathbb{X}$ of the network is non-singular, the covariance matrix further reduces to

$$\begin{aligned}\mathrm{Cov}(L\,\widehat{s}(\widetilde{Y})) &= \frac{1}{N} L\,\mathbb{X}^{-1}(\mathbb{X}\,\mathbb{D}_Z\,\mathbb{X}^\top + \mathbb{D}_E)\,(\mathbb{X}^\top)^{-1}\,L^\top \\ &= \frac{1}{N}\left(L\,\mathbb{D}_Z\,L^\top + L\,\mathbb{X}^{-1}\,\mathbb{D}_E\,(\mathbb{X}^\top)^{-1}\,L^\top\right) \\ &= \frac{1}{N}\left(L\,\mathbb{D}_Z\,L^\top + L\,(\mathbb{X}^\top\,\mathbb{D}_E^{-1}\,\mathbb{X})^{-1}\,L^\top\right).\end{aligned}$$

The covariance $\mathrm{Cov}(L\,\widehat{s}(\widetilde{Y}))$ directly depends on the matrix $\mathbb{D}_E$ in (15) whose diagonal entries indicate the inaccuracy of the applied measurement procedures at the different nodes. More precisely, if the applied measurement procedure is precise at node $i$, the variance $\sigma_{iE}^2$ will be small ($i = 0, \ldots, I$). In the following, we assume that the quantitive relation of all available measurement procedures to one reference measurement procedure is known, i. e. the constants $c_i = \frac{\sigma_*^2}{\sigma_{iE}^2}$, $i = 0, 1, \ldots, I$, are known, where $\sigma_*^2$ is the variance of the reference measurement procedure. In the context of electrical power distribution grids, that would mean that the practioner does not know the exact precision of a particular measurement procedure, but has knowledge about its relative precision compared to the best available procedure based on sensors.

We now address the design problem of allocating the different available measurement procedures at the different nodes such that the resulting covariance matrix $\mathrm{Cov}(L\,\widehat{s}(\widetilde{Y}))$ becomes small in some sense and such that the estimation of the linear aspect $L\,s$ is precise.

For that purpose, we define the precision of the measurement procedure applied at node $i$ by $\delta_i := \frac{1}{\sigma_{iE}^2}$, $i = 0, \ldots, I$ and set $\mathbb{D}_\delta = \mathrm{diag}(\delta_0, \ldots, \delta_I)$. We further assume that the sum of these precisions is bounded by probably unknown constant $K < \infty$, i.e. $\sum_{i=0}^I \delta_i \le K < \infty$. Note that this can be achieved under the condition that the constants $c_0, \ldots, c_I$ are known. Due to the fact that $\mathbb{D}_\delta = \mathbb{D}_E^{-1}$ and that the formulation of the covariance matrix $\mathrm{Cov}(E_n)$ is in terms of an overall variance $\sigma^2$ (c. f. (15)), we can assume that $\sum_{i=0}^I \delta_i \le 1$ without loss of generality. As the optimal design will be allocated at the boundary of that condition we can restrict ourselves to the side condition $\sum_{i=0}^I \delta_i = 1$ so that the set of admissible designs is given by

$\tilde{\Delta} := \{\delta = (\delta_0, \delta_1, \ldots, \delta_I)^\top \in (0,1)^{I+1};\ \sum_{i=0}^I \delta_i = 1\}$ which is a subset of $\Delta$ for Model A introduced in (11). The reformulation in terms of $\delta$ leads to the design problem

$$\delta^* \in \arg\min\{\text{tr } L\,(\mathbb{X}^\top \mathbb{D}_\delta\, \mathbb{X})^{-1} L^\top;\ \delta \in \tilde{\Delta}\}, \tag{16}$$

which is similar to the design problem (13) in Model A with nonrandom states ($\rho_Z = 0$). Note that in contrast to the situation of (13) the regularity of the influence matrix $\mathbb{X}$ and the restriction to $\tilde{\Delta}$ are necessary to define the design problem stated in (16). Note that a solution of the design problem stated in (16) might not exist due to the fact that the set $\tilde{\Delta}$ is not compact anymore (the boundaries are excluded).

## 3 A general result for *A*-optimal designs in Models A and B

The Models A and B lead to design problems of the form

$$\delta^* \in \arg\min\{\text{tr } L\,(X^\top \mathbb{D}_\delta\, X)^{-1} L^\top;\ \delta \in \tilde{\Delta}\}, \tag{17}$$

where $X$ is an $(I+1) \times (I+1)$-matrix, $\mathbb{D}_d = \text{diag}(\delta_0, \delta_1, \ldots, \delta_I)$, and $\tilde{\Delta} := \{d = (\delta_0, \delta_1, \ldots, \delta_I)^\top \in (0,1)^{I+1};\ \sum_{i=0}^I \delta_i = 1\}$. Since $X$ is a square matrix, the problem at hand is a design problem with minimum support. It is easy to see that the D-optimal design for $L\,s = s$ in this case is given by $\delta_0^* = \delta_1^* = \ldots = \delta_I^* = \frac{1}{I+1}$. However, the A-optimal designs are of a different form. We now assume that $X$ is a non-singular matrix so that its inverse $X^{-1}$ exists. Then the following proposition holds.

**Proposition 1** *If $X$ is non-singular, then the design $\delta^* = (\delta_0^*, \delta_1^*, \ldots, \delta_I^*)$ is a solution of the design problem (17) if and only if*

$$\delta_i^* = \frac{\sqrt{v_i}}{\sum_{j=0}^I \sqrt{v_j}} \quad \textit{with} \quad v_i = u_i^\top (X^{-1})^\top L^\top L\, X^{-1} u_i \quad \textit{for} \quad i = 0, 1, \ldots, I\,, \tag{18}$$

*where $u_i$ is the $(i+1)$-th unit vector in $\mathbb{R}^{I+1}$.*

***Proof*** According to the General Equivalence Theorem for $A$-optimality, see (Pukelsheim 2006), Theorem 7.19, with for $p = -1$, $K^\top = L$, a design $\delta^*$ is $A$-optimal if and only if the inequality

$$\|L\,(X^\top \mathbb{D}_{\delta^*} X)^- x_i\|^2 \le \text{tr } L\,(X^\top \mathbb{D}_{\delta^*} X)^-\, L^\top \tag{19}$$

is satisfied for all nodes $i = 0, 1, \ldots, I$, where the vector $x_i$ is given by

$$x_i = X^\top u_i\,, \quad i = 0, 1, \ldots, I\,,$$

in the situation under consideration.

Using that $X$ is non-singular, we obtain

$$\begin{aligned}
&\operatorname{tr}\, L\,(X^\top \mathbb{D}_{\delta^*}\, X)^{-1} L^\top \\
&= \operatorname{tr}\, \mathbb{D}_{\delta^*}^{-1}\,(X^{-1})^\top\, L^\top L\, X^{-1} = \sum_{i=0}^{I} u_i^\top \mathbb{D}_{\delta^*}^{-1}\,(X^{-1})^\top\, L^\top L\, X^{-1} u_i \\
&= \sum_{i=0}^{I} \frac{\sum_{j=0}^{I}\sqrt{v_j}}{\sqrt{v_i}} u_i^\top\,(X^{-1})^\top\, L^\top L\, X^{-1} u_i = \left(\sum_{i=0}^{I}\sqrt{v_i}\right)^2
\end{aligned}$$

for the right-handside of (19), whereas the left-handside reduces to

$$\begin{aligned}
&\|L\,(X^\top \mathbb{D}_{d^*}\, X)^{-1}\, x_i\,\|^2 \\
&= u_i^\top\, X\, X^{-1}\, \mathbb{D}_{d^*}^{-1}\,(X^{-1})^\top\, L^\top L\, X^{-1}\, \mathbb{D}_{d^*}^{-1}\,(X^{-1})^\top\, X^\top\, u_i \\
&= \frac{\sum_{j=0}^{I}\sqrt{v_j}}{\sqrt{v_i}}\, u_i^\top\,(X^{-1})^\top\, L^\top L\, X^{-1}\; u_i\, \frac{\sum_{j=0}^{I}\sqrt{v_j}}{\sqrt{v_i}} = \left(\sum_{i=0}^{I}\sqrt{v_i}\right)^2
\end{aligned}$$

for all $i = 0, 1, \ldots, I$. Consequently, equality holds in (19) for all $i = 0, \ldots, I$ and the equivalence theorem for A-optimality is satisfied. That provides the assertion. □

Note that $v_i$ defined in (18) are the diagonal elements of the matrix $X^{-1}L^\top L X^{-1}$.

By setting

$$X = \begin{cases} \mathbb{D}^{-1}\,\mathbb{X} & \text{in Model A,} \\ \mathbb{X} & \text{in Model B and Model A with nonrandom states,} \end{cases}$$

Proposition 1 provides a solution of the different design problems stated in the situation of Model A and Model B, respectively.

**Theorem 1** *Let the influence matrix $\mathbb{X}$ be non-singular. Then the A-optimal designs for estimating $L$ s in the Models A and B, i.e. of $\delta^* = (\delta_0^*, \delta_1^*, \ldots, \delta_I^*)$ solving (12), (13), and (16), respectively, are given by*

$$\begin{aligned}
\delta_i^* &= \frac{\sqrt{v_i}}{\sum_{j=0}^{I}\sqrt{v_j}} \quad \textit{with} \\
v_i &= \begin{cases} u_i^\top (\mathbb{X}^{-1})^\top L^\top L\, \mathbb{X}^{-1} u_i\; (u_i^\top \mathbb{X}\mathbb{X}^\top u_i\; \rho_Z + \rho_E) & \textit{in Model A,} \\ u_i^\top (\mathbb{X}^{-1})^\top L^\top L\, \mathbb{X}^{-1} u_i & \textit{in Model B,} \end{cases} \\
&\quad \textit{for} \quad i = 0, 1, \ldots, I.
\end{aligned}$$

***Proof*** The form of $v_i$ in the general Model A follows by

$$v_i = u_i^\top\, \mathbb{D}\,(\mathbb{X}^{-1})^\top L^\top L\, \mathbb{X}^{-1}\, \mathbb{D}\, u_i = \sigma_i\, u_i^\top (\mathbb{X}^{-1})^\top L^\top L\, \mathbb{X}^{-1} u_i\, \sigma_i,$$

where $\mathbb{D} = \text{diag}(\sigma_0, \sigma_1, \ldots, \sigma_I)$ and

$$\sigma_i^2 := u_i^\top \mathbb{X}\mathbb{X}^\top u_i \ \rho_Z + \rho_E.$$

The assertion for Model A with non-random states follows by $\rho_Z = 0$. Note again that the $A$-optimal design for Model B can be obtained by the $A$-optimal design for Model A by setting $\rho_Z = 0$, $\rho_E = 1$, which means that the states are nonrandom. Then the assertion for Model B follows as well. □

Hence as soon as the inverse $\mathbb{X}^{-1}$ of the influence matrix is determined, the $A$-optimal design is available. For large complex networks, this inverse can only be determined numerically. However, for some simple networks as stated in Example 1, $\mathbb{X}^{-1}$ can be calculated analytically. This is the case, for example, for the star network introduced in Example 1, as shown in the next section.

## 4 *A*-optimal designs in a star network

Networks with a star configuartion, shortly star networks, are simple, but realistic networks for electrical power distribution grids, as e.g. Su and Wang (2020) and Azhdari and Ardakan (2022) pointed out. They consist of a central or outgoing node 0 which is connected to all other nodes $i = 1, \ldots, I$ of the network, wheras the other nodes are terminal nodes that are only connected to the central node 0. Therefore, we now concentrate on the situation introduced in the first part of Example 1 with the influence matrix $\mathbb{X}$ of the star network given by

$$\mathbb{X} = \begin{pmatrix} a & b\, 1_I^\top \\ b\, 1_I & a\, \mathbb{I}_{I\times I} \end{pmatrix}. \tag{20}$$

Note again that $a$ describes the influence of the state on the observation at the respective node, whereas $b$ denotes the amount of influence of the states at the adjacent nodes on that observation. We are now interested in the analytic determination of the corresponding $A$-optimal designs if the influence matrix is given by (20). For that purpose, Theorem 1 is only applicable, if the influence matrix $\mathbb{X}$ is non-singular. Therefore, we state an equivalent condition for the non-singularity of $\mathbb{X}$ given in (20) in the following lemma.

**Lemma 1** *For the influence matrix in* (20)*, it holds:*

*a)* $u_i^\top \mathbb{X}\mathbb{X}^\top u_i = \begin{cases} a^2 + I\, b^2 & \text{for } i = 0, \\ a^2 + b^2 & \text{for } i = 1, \ldots, I. \end{cases}$

*b)* $\mathbb{X}$ *is non-singular if and only if* $b^2 \neq \frac{1}{I} a^2$.

***Proof*** a). Let $\widetilde{u}_i$ be the $i$-th unit vector in $\mathbb{R}^I$, $0_I$ the $I$-dimensional vector consisting only of zeros, and $1_I$ the $I$ dimensional vector consisting only of ones. Then it holds

$$u_i^\top \mathbb{X} = \begin{cases} (1, 0_I^\top)\mathbb{X} = (a, b 1_I^\top) & \text{for } i = 0, \\ (0, \widetilde{u}_i^\top)\mathbb{X} = (b, a\, \widetilde{u}_i^\top) & \text{for } i \in \{1, \ldots, I\}. \end{cases}$$

The statement in a) directly follows.
b) Note that the determinant of $\mathbb{X}$ is given by

$$\begin{aligned}\det(\mathbb{X}) &= \det(a\mathbb{I}_{I\times I})\det(a - b\,1_I^\top a^{-1}\mathbb{I}_{I\times I} b\,1_I)\\ &= I\,a^{I-1}\left(\frac{1}{I}a^2 - b^2\right),\end{aligned}$$

which is non-zero if and only if $b^2 \neq \frac{a^2}{I}$. □

Against expectation the influence matrix $\mathbb{X}$ stays non-singular if the influence of the non-central nodes on the central node is equal to the influence of the central node, i. e. $Ib = a$. Instead, the matrix $\mathbb{X}$ becomes singular, if $Ib^2 = a^2$, whereas this combination has no obvious effect on the structure of the star network. Nevertheless, the non-singularity of the influence matrix $\mathbb{X}$ has an direct impact on the availability of an appropriate estimator of the complete expected state vector $s$: $s$ is only identifiable and thus estimable if and only if the influence matrix $\mathbb{X}$ is non-singular. In the next section, we derive the $A$-optimal design for the complete state vector $s$ under the assumption of identifiability.

### 4.1 *A*-Optimal designs for the complete state vector *s* under identifiability

**Lemma 2** *If* $b^2 \neq \frac{1}{I}a^2$ *then*

$$\mathbb{X}^{-1} = \begin{pmatrix} a & b\,1_I^\top \\ b\,1_I & a\,\mathbb{I}_{I\times I}\end{pmatrix}^{-1} = \frac{1}{a^2 - I\,b^2}\begin{pmatrix} a & -b\,1_I^\top \\ -b\,1_I & \frac{a^2 - I\,b^2}{a}\,\mathbb{I}_{I\times I} + \frac{b^2}{a}1_{I\times I}\end{pmatrix},$$

*where* $1_{I\times I} \in \mathbb{R}^{I\times I}$ *is the* $I \times I$*-matrix consisting of ones.*

***Proof*** It is well known that for symmetric matrices $A$ and $C$, where $C$ and $E = A - B^\top C^{-1}B$ are non-singular, it holds

$$\begin{pmatrix} A & B^\top \\ B & C\end{pmatrix}^{-1} = \begin{pmatrix} E^{-1} & -E^{-1}\,B^\top\,C^{-1} \\ -C^{-1}\,B\,E^{-1} & C^{-1} + C^{-1}\,B\,E^{-1}\,B^\top\,C^{-1}\end{pmatrix} \tag{21}$$

(see, e.g., Rencher (1998), p. 407).
Setting $A = a$, $B = b\,1_I$, $C = a\,\mathbb{I}_{I\times I}$, we obtain

$$E = a - b\,1_I^\top\,\frac{1}{a}\,b\,1_I = \frac{1}{a}\,(a^2 - I\,b^2)$$

and thus

$$\begin{pmatrix} a & b\,1_I^\top \\ b\,1_I & a\,\mathbb{I}_{I\times I}\end{pmatrix}^{-1} = \begin{pmatrix} \frac{a}{a^2 - I\,b^2} & -\frac{a}{a^2 - I\,b^2}\,\frac{b}{a}\,1_I^\top \\ -\frac{a}{a^2 - I\,b^2}\,\frac{b}{a}\,1_I & \frac{1}{a}\,\mathbb{I}_{I\times I} + \frac{1}{a^2}\,\frac{a\,b^2}{a^2 - I\,b^2}\,1_I\,1_I^\top\end{pmatrix}$$

$$= \frac{1}{a^2 - I\,b^2} \begin{pmatrix} a & -b\,1_I^\top \\ -b\,1_I & \frac{a^2 - I\,b^2}{a}\,\mathbb{I}_{I\times I} + \frac{b^2}{a} 1_{I\times I} \end{pmatrix}.$$

□

Hence Theorem 1 and Lemma 1, a) provide the following theorem:

**Theorem 2** *If $b^2 \neq \frac{1}{I}a^2$, then the A-optimal design $\delta^* = (\delta_0^*, \delta_1^*, \ldots, \delta_I^*)$ for estimating the expected state vector $s$ in the star network is given by $\delta_0^* = \frac{\sqrt{w}}{I\sqrt{v}+\sqrt{w}}$ and $\delta_i^* = \frac{\sqrt{v}}{I\sqrt{v}+\sqrt{w}}$ for $i = 1, \ldots, I$, where*

$$w = \begin{cases} (a^2 + Ib^2)\,((a^2 + Ib^2)\,\rho_Z + \rho_E) & \text{in Model A,} \\ a^2 + Ib^2 & \text{in Model B,} \end{cases}$$

$$v = \begin{cases} \left(b^2 + \frac{(a^2-(I-1)b^2)^2}{a^2} + (I-1)\frac{b^4}{a^2}\right)\,((a^2 + b^2)\,\rho_Z + \rho_E) & \text{in Model A,} \\ \left(b^2 + \frac{(a^2-(I-1)b^2)^2}{a^2} + (I-1)\frac{b^4}{a^2}\right) & \text{in Model B.} \end{cases}$$

***Proof*** Setting $L = \mathbb{I}_{(I+1)\times(I+1)}$ we obtain by Theorem 1 that

$$\delta_i^* = \frac{\sqrt{v_i}}{\sum_{j=0}^I \sqrt{v_j}} \quad \text{with}$$

$$v_i = \begin{cases} u_i^\top (\mathbb{X}^{-1})^\top\,\mathbb{X}^{-1} u_i\,(u_i^\top \mathbb{X}\mathbb{X}^\top u_i\;\rho_Z + \rho_E) & \text{in Model A,} \\ u_i^\top (\mathbb{X}^{-1})^\top\,\mathbb{X}^{-1} u_i & \text{in Model B,} \end{cases}$$

$$\text{for} \quad i = 0, 1, \ldots, I.$$

Lemma 1 a) provides the additional terms $u_i^\top \mathbb{X}\mathbb{X}^\top u_i\;\rho_Z + \rho_E$ in Model A. The common terms in both models are $u_i^\top (\mathbb{X}^{-1})^\top\,\mathbb{X}^{-1} u_i$, where the inverse $\mathbb{X}^{-1}$ is given by Lemma 2. At first note that the factor $\frac{1}{a^2 - I\,b^2}$ of $\mathbb{X}^{-1}$ cancels out in $\delta_i^*$ so that we only have to consider

$$\mathbb{V} := (a^2 - I\,b^2)\,\mathbb{X}^{-1} = \begin{pmatrix} a & -b\,1_I^\top \\ -b\,1_I & \frac{a^2 - I\,b^2}{a}\,\mathbb{I}_{I\times I} + \frac{b^2}{a} 1_{I\times I} \end{pmatrix}.$$

Here, we get

$$\mathbb{V}^\top u_0 = \mathbb{V}^\top \begin{pmatrix} 1 \\ 0_I \end{pmatrix} = \begin{pmatrix} a \\ -b\,1_I \end{pmatrix} \text{ so that } u_0^\top \mathbb{V}\,\mathbb{V}^\top u_0 = a^2 + I\,b^2$$

and

$$\mathbb{V}^\top u_i = \mathbb{V}^\top \begin{pmatrix} 0 \\ \widetilde{u}_i \end{pmatrix} = \begin{pmatrix} -b \\ \frac{a^2 - I\,b^2}{a}\,\widetilde{u}_i + \frac{b^2}{a}\,1_I \end{pmatrix}$$

so that

$$u_i^\top \mathbb{V}\mathbb{V}^\top u_i = \left( b^2 + \left( \frac{a^2 - I\,b^2}{a} \right)^2 + 2 \left( \frac{a^2 - I\,b^2}{a} \right) \frac{b^2}{a} + I \left( \frac{b^2}{a} \right)^2 \right)$$
$$= \left( b^2 + \left( \frac{a^2 - (I-1)\,b^2}{a} \right)^2 + (I-1) \left( \frac{b^2}{a} \right)^2 \right)$$

for $i = 1, \ldots, I$. Hence these are the common terms for $w$ and $v$ in Model A and Model B, respectively. □

***Remark 1*** Note that $\delta_1^* = \ldots, = \delta_I^*$, i.e. the terminal nodes are treated equally, wheras the central node 0 obtains a different value $\delta_0^*$ in general.

A special case of the star network is $b = 0$, where the states of the adjacent nodes do not influence the observation at a particular node. Then we get $v = w$ in both models so that $\delta_0^* = \delta_1^* = \ldots = \delta_I^* = \frac{1}{I+1}$. Hence, the $A$-optimal design is equal to the $A$-optimal design obtained in the classical model, where $I + 1$ independent levels of one factor are considered.

If the star network only consists of two nodes, i. e. $I = 1$, it also follows $v = w$ in both models so that $\delta_0^* = \delta_1^* = \frac{1}{2}$ for any $b$ with $b^2 \neq a^2$. Hence the design does not depend on the adjacent effect $b$. This is not the case for $I > 1$ which will be considered in detail in the following example.

Note that the $A$-optimal design according to the formula stated in Theorem 2 can also calculated in the case $b^2 = \frac{1}{I}a^2$, i.e. in case of a singular matrix $\mathbb{X}$, since the factor $\frac{1}{a^2 - Ib^2}$ in $\mathbb{X}^{-1}$ cancels out and does not appear in the formula as well as in the proof. Nevertheless, the whole state vector $s$ is not identifiable for $b^2 = \frac{1}{I}a^2$ which is shown in the next section.

***Example 2*** We investigate the behaviour of the $A$-optimal design in dependence on different values of $a$ and $b$ in the situation of Model A, where either nonrandom states are given, i.e. $\rho_Z = 0$, or no measurement errors occur, i.e. $\rho_E = 0$. The $A$-optimal designs only depend on the relationship between $a$ and $b$ and we can set $a = 1$ without loss of generality. Hence, Fig. 2 shows the optimal values for $\delta_0^*$ depending on the quantity $b$ for a star network with $I = 4, 9, 25$ nodes and $a = 1$ for Model A with nonrandom states given by $\rho_Z = 0$ (left-hand side) and for Model A with no measurements errors given by $\rho_E = 0$ (right-hand side). Note that the designs for Model B coincide with those of Model A with nonrandom states, if $\mathbb{X}$ is non-singular. In particular, Fig. 2 shows that the special case $b^2 = \frac{1}{I}$, where the state vector $s$ is not identifiable leads to a smooth continuation of the case $b^2 \neq \frac{1}{I}$. Furthermore, if the influence $b$ of the central node 0 goes to infinity, then the optimal weight $\delta_0$ at the central node 0 goes to zero. This means that only a small proportion of observations should be done at the central node if it has a big influence on its adjacent nodes and vice versa. Surprisingly, the optimal weight $\delta_0^*$ increases for $b^2 < 1/I$ and decreases for $b^2 > 1/I$, i. e., when the value of $b$ is reached where the state vector $s$ is not identifiable. Moreover, Model A with random states and no measurements errors provides larger weights $\delta_0^*$ at the central node than the Model A with nonrandom

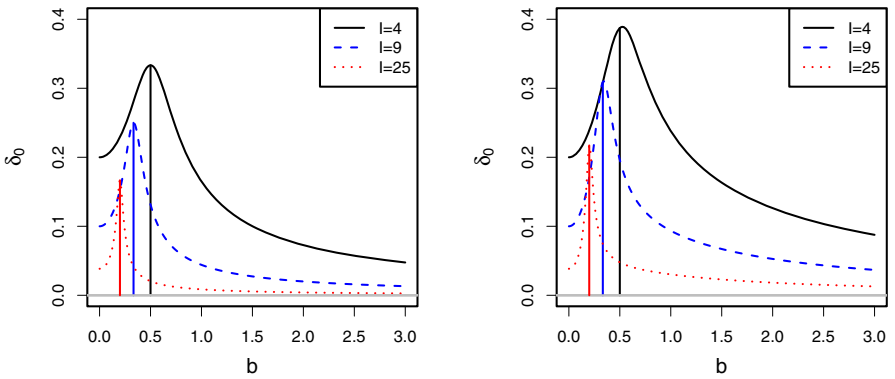


**Fig. 2** Optimal values for $\delta_0$ depending on the quantity $b$, the influence of the central node 0, for $a = 1$. Left: in Model B and Model A with nonrandom states. Right: in Model A without measurement errors. The vertical lines indicate the case $b^2 = \frac{1}{I}$ where the state vector $s$ is not identifiable

states and measurement errors. Probably, this is caused by the increased uncertainty given by the random states.

### 4.2 Nonidentifiability in a star network

As mentioned in Remark 1, the state vector $s$ is not identifiable, as soon as the influence matrix $\mathbb{X}$ becomes singular. In the case of the star network, this is equivalent to the case where $b^2 = \frac{1}{I}a^2$, where the influence matrix

$$\mathbb{X} := \begin{pmatrix} a & b\,1_I^\top \\ b\,1_I & a\,\mathbb{I}_{I\times I} \end{pmatrix} = \begin{pmatrix} a & \frac{1}{\sqrt{I}}\,a\,1_I^\top \\ \frac{1}{\sqrt{I}}\,a\,1_I & a\,\mathbb{I}_{I\times I} \end{pmatrix}$$

is not of full rank, for example, it holds

$$\begin{pmatrix} a & \frac{1}{\sqrt{I}}\,a\,1_I^\top \\ \frac{1}{\sqrt{I}}\,a\,1_I & a\,\mathbb{I}_{I\times I} \end{pmatrix} \begin{pmatrix} 1 \\ -\frac{1}{\sqrt{I}}\,1_I \end{pmatrix} = \begin{pmatrix} a - \frac{a}{I}\,I \\ \frac{1}{\sqrt{I}}\,a\,1_I - a\,\frac{1}{\sqrt{I}}\,1_I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Moreover, the often used aspect

$$\widetilde{L}\,s = \begin{pmatrix} s_1 - s_0 \\ \vdots \\ s_I - s_0 \end{pmatrix} \in \mathbb{R}^I$$

with $\widetilde{L} := (-1_I, \mathbb{I}_{I\times I})$, where the central node 0 is considered as control level, is not identifiable, since

$$\widetilde{L}\begin{pmatrix} 1 \\ -\frac{1}{\sqrt{I}}\,1_I \end{pmatrix} = -1_I - \frac{1}{\sqrt{I}}\,1_I \neq 0_I.$$

However, the aspect

$$L\,s = \begin{pmatrix} s_1 + \frac{1}{\sqrt{I}}\,s_0 \\ \vdots \\ s_I + \frac{1}{\sqrt{I}}\,s_0 \end{pmatrix} \in \mathbb{R}^I \tag{22}$$

with $L := \left(\frac{1}{\sqrt{I}}\,1_I, \mathbb{I}_{I\times I}\right)$ is identifiable since

$$\begin{aligned}
&\frac{1}{a}\left(\sqrt{I}\,1_I,\ -1_{I\times I} + \mathbb{I}_{I\times I}\right)\mathbb{X} = \left(\sqrt{I}\,1_I,\ -1_{I\times I} + \mathbb{I}_{I\times I}\right)\begin{pmatrix} 1 & \frac{1}{\sqrt{I}}\,1_I^\top \\ \frac{1}{\sqrt{I}}\,1_I & \mathbb{I}_{I\times I} \end{pmatrix} \\
&= \left(\sqrt{I}\,1_I - 1_{I\times I}\,\frac{1}{\sqrt{I}}\,1_I + \mathbb{I}_{I\times I}\,\frac{1}{\sqrt{I}}\,1_I\ ,\ \ \sqrt{I}\,1_I\,\frac{1}{\sqrt{I}}\,1_I^\top + (-1_{I\times I} + \mathbb{I}_{I\times I})\right) \\
&= \left(\sqrt{I}\,1_I - \frac{I}{\sqrt{I}}\,1_I + \frac{1}{\sqrt{I}}\,1_I\ ,\ \ 1_{I\times I} - 1_{I\times I} + \mathbb{I}_{I\times I}\right) = L.
\end{aligned}$$

### 4.3 *A*-optimal designs for the always identifiable aspect *L s*

The aim is to determine the $A$-optimal designs for estimating the aspect $L\,s$ given by (22) in case of identifiability and nonidentifiability of $s$ so that, according to (12), the design problem in the general Model A is given by

$$\delta^* \in \arg\min\{\mathrm{tr}\ L\,I(\delta)^-L^\top;\ \delta \in \Delta\} \tag{23}$$

with

$$I(\delta) := \mathbb{X}^\top\mathbb{D}^{-1}\mathbb{D}_\delta\,\mathbb{D}^{-1}\mathbb{X}.$$

At first, we consider the case of nonidentifiability, i.e. $b^2 = \frac{1}{I}a^2$, where only Model A makes sense. For this case, we are now going to prove that the $A$-optimal design $\delta^* = (\delta_0^*, \delta_1^*, \ldots, \delta_I^*)$, i.e. a solution of (23), is given by $\delta_0^* = 0$ and $\delta_1^* = \ldots = \delta_I^* = \frac{1}{I}$ using the equivalence theorem of $A$-optimality. Hence, it is sufficient to consider the information matrix for $\delta^*$ and the corresponding generalized inverse (since the information matrix is not invertible for $\delta^*$). At first, note that Lemma 1 provides for

$b^2 = \frac{1}{I} a^2$

$$u_i^\top \mathbb{X}\mathbb{X}^\top u_i = \begin{cases} a^2 + I\, b^2 = 2a^2 & \text{for } i = 0, \\ a^2 + b^2 = a^2 + \frac{1}{I} a^2 = a^2\, \frac{I+1}{I} & \text{for } i = 1, \ldots, I. \end{cases} \tag{24}$$

**Lemma 3** *Set* $\alpha := \frac{1}{I} \frac{1}{a^2\, \frac{I+1}{I} \rho_Z + \rho_E}$, *then the information matrix* $I(\delta^*)$ *in Model A is given by*

$$I(\delta^*) = a^2\, \alpha \begin{pmatrix} 1 & \frac{1}{\sqrt{I}} 1_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix}.$$

*Moreover, a generalized inverse of* $I(\delta^*)$ *is given by*

$$I(\delta^*)^- = \frac{1}{a^2\, \alpha} \begin{pmatrix} 0 & 0_I^\top \\ 0_I & \mathbb{I}_{\mathbb{I} \times I.} \end{pmatrix}. \tag{25}$$

***Proof*** Since $\delta_0^* = 0$, $\delta_1^* = \ldots = \delta_I^* = \frac{1}{I}$, and $\mathbb{D} = \operatorname{diag}(\sigma_0, \sigma_1, \ldots, \sigma_I)$ with $\sigma_i^2 = a^2\, \frac{I+1}{I} \rho_Z + \rho_E$ for $i = 1, \ldots, I$ according to (24), we obtain

$$\begin{aligned} I(\delta^*) &= \mathbb{X}^\top \begin{pmatrix} 0 & 0_I^\top \\ 0_I & \alpha\, \mathbb{I}_{I \times I} \end{pmatrix} \mathbb{X} \\ &= a^2 \begin{pmatrix} 1 & \frac{1}{\sqrt{I}} 1_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix} \begin{pmatrix} 0 & 0_I^\top \\ 0_I & \alpha\, \mathbb{I}_{I \times I} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{\sqrt{I}} 1_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix} \\ &= a^2 \begin{pmatrix} 1 & \frac{1}{\sqrt{I}} 1_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \frac{\alpha}{\sqrt{I}} 1_I & \alpha\, \mathbb{I}_{I \times I} \end{pmatrix} \\ &= a^2 \begin{pmatrix} \frac{1}{\sqrt{I}} 1_I^\top \frac{\alpha}{\sqrt{I}} 1_I & \frac{1}{\sqrt{I}} 1_I^\top \alpha\, \mathbb{I}_{I \times I} \\ \mathbb{I}_{I \times I} \frac{\alpha}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I}\, \alpha\, \mathbb{I}_{I \times I} \end{pmatrix} = a^2\, \alpha \begin{pmatrix} 1 & \frac{1}{\sqrt{I}} 1_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I.} \end{pmatrix}. \end{aligned}$$

We are now going to show that the matrix $I(\delta^*)^-$ proposed in (25) is a generalized inverse of $I(\delta^*)$ by checking whether $I(\delta^*)\, I(\delta^*)^-\, I(\delta^*) = I(\delta^*)$. Since $a$ and $\alpha$ are multiplicative constants, we do not need to consider them. Without them, we get

$$\begin{aligned} &\begin{pmatrix} 1 & \frac{1}{\sqrt{I}} 1_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix} \begin{pmatrix} 0 & 0_I^\top \\ 0_I & \mathbb{I}_{\mathbb{I} \times I} \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{\sqrt{I}} 1_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \frac{1}{\sqrt{I}} 1_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix} \begin{pmatrix} 0 & 0_I^\top \\ \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{I}} 1_I^\top \frac{1}{\sqrt{I}} 1_I & \frac{1}{\sqrt{I}} 1_I^\top \\ (\mathbb{I}_{I \times I}) \frac{1}{\sqrt{I}} 1_I & \mathbb{I}_{I \times I} \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} 1 & \frac{1}{\sqrt{I}}\, 1_I^\top \\ \frac{1}{\sqrt{I}}\, 1_I & \mathbb{I}_{I\times I} \end{pmatrix}.$$

□

Note that the generalized inverse calculated in Lemma 3 is the Moore-Penrose generalized inverse. We are now able to prove the $A$-optimality of the design $\delta^*$ in case of non-identifiability.

**Theorem 3** *Let the model be given by Model A with influence matrix defined by* (20). *If* $b^2 = \frac{1}{I} a^2$ *and* $L = \left(\frac{1}{\sqrt{I}}\, 1_I,\ \mathbb{I}_{I\times I}\right)$, *the design* $\delta^* = (\delta_0^*, \delta_1^*, \ldots, \delta_I^*)$ *is A-optimal for estimating L s (i. e. a solution of* (23)*) if and only if* $\delta_0^* = 0$ *and* $\delta_1^* = \delta_2^* = \ldots = \delta_I^* = \frac{1}{I}$.

***Proof*** According to the General Equivalence Theorem for $A$-optimality, see Pukelsheim (2006), Theorem 7.19, with for $p = -1$, $K^\top = L$, a design $\delta^*$ is $A$-optimal if and only if the inequality

$$\|L\, I(\delta^*)^- x_i\|^2 \le \operatorname{tr}\, L\, I(\delta^*)^-\, L^\top \tag{26}$$

is satisfied for all nodes $i = 0, 1, \ldots, I$, where the vector $x_i$ is given by

$$x_i = u_i^\top \mathbb{D}^{-1}\mathbb{X}, \quad i = 0, 1, \ldots, I,$$

in the situation under consideration.
We set $\alpha := \frac{1}{I}\frac{1}{a^2\,\frac{I+1}{I}\rho_Z + \rho_E}$, then Lemma 3 provides a reformulation of the right-hand side of inequality (26), namely

$$\operatorname{tr}\, L\, I(\delta^*)^- L^\top = \operatorname{tr}\left[\left(\frac{1}{\sqrt{I}}\, 1_I,\ \mathbb{I}_{I\times I}\right) \frac{1}{a^2\,\alpha} \begin{pmatrix} 0 & 0_I^\top \\ 0_I & \mathbb{I}_{I\times I} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{I}}\, 1_I \\ \mathbb{I}_{I\times I} \end{pmatrix}\right]$$
$$= \operatorname{tr}\left[\left(\frac{1}{\sqrt{I}}\, 1_I,\ \mathbb{I}_{I\times I}\right) \frac{1}{a^2\,\alpha} \begin{pmatrix} 0_I^\top \\ \mathbb{I}_{I\times I} \end{pmatrix}\right] = \operatorname{tr}\left[\frac{1}{a^2\,\alpha}\, \mathbb{I}_{I\times I}\right] = \frac{I}{a^2\,\alpha}.$$

The design matrix containing $x_1, \ldots, x_n$ is of the form

$$\begin{pmatrix} x_0^\top \\ x_1^\top \\ \vdots \\ x_I^\top \end{pmatrix} = X = \mathbb{D}^{-1}\mathbb{X} = \begin{pmatrix} \frac{1}{\sigma_0}\,(a,\ b\, 1_I^\top) \\ \frac{1}{\sigma_1}\,(b,\ a\, \widetilde{u}_1^\top) \\ \vdots \\ \frac{1}{\sigma_I}\,(b,\ a\, \widetilde{u}_I^\top) \end{pmatrix},$$

where $\sigma_1^2 = \ldots = \sigma_I^2 = a^2\,\frac{I+1}{I}\,\rho_z + \rho_E = \frac{1}{I\,\alpha}$ according to (24) and $\widetilde{u}_i$ is the $i$-th unit vector in $\mathbb{R}^I$.

First, we consider the nodes $\{1, \ldots, I\}$ with vectors $x_i = \frac{1}{\sigma_i}(b,\ a\,\widetilde{u}_i^\top)^\top$ with $i = 1, \ldots, I$ and check the inequality given by (26). In this situation, we get

$$\begin{aligned}
&\|L\, I(\delta^*)^- x_i\|^2 \\
&= \left\| \left( \frac{1}{\sqrt{I}}\, 1_I,\ \mathbb{I}_{I\times I} \right) \frac{1}{a^2\,\alpha} \begin{pmatrix} 0 & 0_I^\top \\ 0_I & \mathbb{I}_{I\times I} \end{pmatrix} \frac{1}{\sigma_i} \begin{pmatrix} b \\ a\,\widetilde{u}_i \end{pmatrix} \right\|^2 \\
&= \left\| \left( \frac{1}{\sqrt{I}}\, 1_I,\ \mathbb{I}_{I\times I} \right) \frac{\sqrt{I\,\alpha}}{a^2\,\alpha} \begin{pmatrix} 0 \\ a\,\widetilde{u}_i \end{pmatrix} \right\|^2 \\
&= \left\| \frac{\sqrt{I}}{a^2\,\sqrt{\alpha}}\, a\,\widetilde{u}_i \right\|^2 = \frac{I}{a^4\,\alpha}\, a^2 = \operatorname{tr}\, L\, I(d^*)^-\, L^\top
\end{aligned}$$

for $i = 1, \ldots, I$.
For the central node $i = 0$, we have $x_0 = \frac{1}{\sigma_0}(a, b1_I^\top)^\top$. Using similar arguments, we obtain for the left hand side of inequality (26)

$$\|L\, I(\delta^*)^- x_i\|^2 = \frac{1}{a^2\,\alpha^2} \frac{1}{\sigma_0^2} \le \frac{I}{a^2\,\alpha}$$

where the last inequality follows by the fact that

$$\sigma_0^2 = 2a^2\rho_Z + \rho_E > a^2 \frac{I+1}{I} \rho_Z + \rho_E = \sigma_i^2, \quad i = 1, \ldots, I.$$

Hence, the equivalence theorem for A-optimality provides the assertion. □

In the remaining part of this section, we derive the $A$-optimal designs for $L\,s$ when $s$ is identifiable, i.e., when $b^2 \neq \frac{1}{I} a^2$. This result is both applicable in Model A and Model B, as the influence matrix $\mathbb{X}$ is non-singular.

**Theorem 4** *If $b^2 \neq \frac{1}{I} a^2$, then the A-optimal design $\delta^* = (\delta_0^*, \delta_1^*, \ldots, \delta_I^*)$ for estimating $L\ s$ with $L = \left( \frac{1}{\sqrt{I}}\, 1_I\ ,\ \mathbb{I}_{I\times I} \right)$ is given by $\delta_0^* = \frac{\sqrt{w}}{I\sqrt{v}+\sqrt{w}}$ and $\delta_i^* = \frac{\sqrt{v}}{I\sqrt{v}+\sqrt{w}}$ for $i = 1, \ldots, I$ where*

$$\begin{aligned}
&w = \widetilde{w}\,((a^2 + Ib^2)\ \rho_Z + \rho_E) \ \text{ with } \widetilde{w} = I \left( \frac{a}{\sqrt{I}} - b \right)^2, \\
&v = \widetilde{v}\,((a^2 + b^2)\ \rho_Z + \rho_E) \ \text{ with} \\
&\qquad \widetilde{v} = I \left( \frac{b^2}{a} - \frac{b}{\sqrt{I}} \right)^2 + 2 \left( \frac{b^2}{a} - \frac{b}{\sqrt{I}} \right) \left( \frac{a^2 - Ib^2}{a} \right) + \left( \frac{a^2 - Ib^2}{a} \right)^2.
\end{aligned}$$

***Proof*** According to Theorem 1, we have

$$\delta_i^* = \frac{\sqrt{v_i}}{\sum_{j=0}^I \sqrt{v_j}} \quad \text{with}$$

$$v_i = u_i^\top (\mathbb{X}^{-1})^\top L^\top L \mathbb{X}^{-1} u_i \, (u_i^\top \mathbb{X}\mathbb{X}^\top u_i \;\rho_Z + \rho_E)$$
$$\text{for} \quad i = 0, 1, \ldots, I.$$

Again, Lemma 1 provides the terms $u_i^\top \mathbb{X}\mathbb{X}^\top u_i \quad \rho_Z + \rho_E$. Hence only $u_i^\top (\mathbb{X}^{-1})^\top L^\top L \mathbb{X}^{-1} u_i$ hast to be calculated. Again, the factor $\frac{1}{a^2 - I\,b^2}$ of $\mathbb{X}^{-1}$ cancels out in $\delta_i^*$ so that we only have to consider

$$\mathbb{V} := (a^2 - I\,b^2)\mathbb{X}^{-1} = \begin{pmatrix} a & -b\,1_I^\top \\ -b\,1_I & \frac{a^2 - I\,b^2}{a}\,\mathbb{I}_{I\times I} + \frac{b^2}{a} 1_{I\times I} \end{pmatrix}.$$

Here, we get

$$L\,\mathbb{V}^\top u_0 = L\,\mathbb{V}^\top \begin{pmatrix} 1 \\ 0_I \end{pmatrix}$$
$$= \left(\frac{1}{\sqrt{I}}\,1_I,\, \mathbb{I}_{I\times I}\right) \begin{pmatrix} a \\ -b\,1_I \end{pmatrix} = \left(\frac{a}{\sqrt{I}}\,1_I \,-\, b\,1_I\right) = \left(\frac{a}{\sqrt{I}} \,-\, b\right) 1_I$$

so that

$$\widetilde{w} = u_0^\top \mathbb{V} L^\top L\,\mathbb{V}^\top u_0 = I\,\left(\frac{a}{\sqrt{I}} \,-\, b\right)^2$$

and

$$L\,\mathbb{V}^\top u_i = L\,\mathbb{V}^\top \begin{pmatrix} 0 \\ \widetilde{u}_i \end{pmatrix} = \left(\frac{1}{\sqrt{I}}\,1_I,\, \mathbb{I}_{I\times I}\right) \begin{pmatrix} -b \\ \frac{a^2 - I\,b^2}{a}\,\widetilde{u}_i + \frac{b^2}{a}\,1_I \end{pmatrix}$$
$$= \left(\frac{-b}{\sqrt{I}}\,1_I \,+\, \frac{a^2 - I\,b^2}{a}\,\widetilde{u}_i + \frac{b^2}{a}\,1_I\right) = \left(\frac{b^2}{a} - \frac{b}{\sqrt{I}}\right) 1_I \,+\, \frac{a^2 - I\,b^2}{a}\,\widetilde{u}_i$$

so that

$$\widetilde{v} = u_i^\top \mathbb{V} L^\top L\,\mathbb{V}^\top u_i$$
$$= I\left(\frac{b^2}{a} - \frac{b}{\sqrt{I}}\right)^2 + 2\left(\frac{b^2}{a} - \frac{b}{\sqrt{I}}\right)\left(\frac{a^2 - Ib^2}{a}\right) + \left(\frac{a^2 - Ib^2}{a}\right)^2$$

for $i = 1, \ldots, I$. □

***Remark 2*** Contrary to the estimation of the complete vector $s$ of states in Sect. 4.1, the $A$-optimal designs for estimating $L\,s$ in the case $b^2 \neq \frac{1}{I}a^2$ cannot be extended to the case $b^2 = \frac{1}{I}a^2$ since the values $v$ and $w$ are then equal to zero.

However, as in Sect. 4.1, we get in the case of no influence of adjacent nodes, i.e. $b = 0$, the equalities $\widetilde{w} = \widetilde{v}$ and $w = v$ so that the $A$-optimal design is again given by $\delta_0^* = \delta_1^* = \ldots = \delta_I^* = \frac{1}{I+1}$.

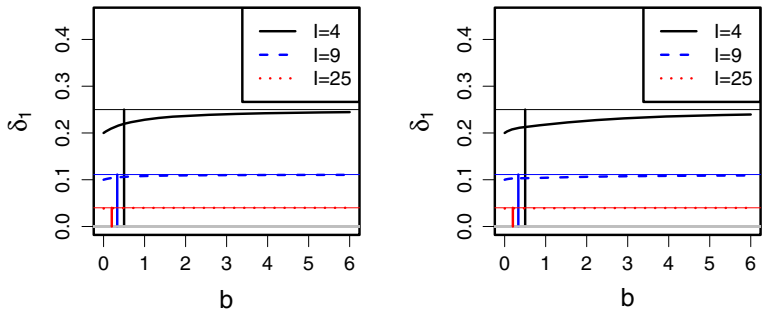


**Fig. 3** Optimal values for $\delta_1 (= \delta_2 = \ldots = \delta_I)$ for estimating $L\,s$ depending on the quantity $b$, the influence of the central node 0, for $a = 1$. Left: in Model B and in Model A with nonrandom states, i.e. $\rho_Z = 0$. Right: in Model A with random states and no measurements errors, i.e. $\rho_E = 0$. The vertical lines indicate the case $b^2 = \frac{1}{I}$ where the state vector $s$ is not identifiable. The thin horizontal lines indicate the weights for $b$ going to infinity

If $I = 1$ and $b^2 \neq \frac{1}{I}a^2 = a^2$, then

$$\begin{aligned}\widetilde{v} &= \left[\left(\frac{b^2-ba}{a}\right)^2 + 2\left(\frac{b^2-ba}{a}\right)\left(\frac{a^2-b^2}{a}\right) + \left(\frac{a^2-b^2}{a}\right)^2\right] \\ &= \frac{1}{a^2}\left[b^2(b-a)^2 + 2b(b-a)(a-b)(a+b) + (a-b)^2(a+b)^2\right] \\ &= \frac{(a-b)^2}{a^2}\left[b^2 - 2b(a+b) + (a+b)^2\right] \\ &= \frac{(a-b)^2}{a^2}\left[b^2 - 2ba - 2b^2 + a^2 + 2ab + b^2\right] = (a-b)^2 = \widetilde{w}\end{aligned}$$

so that $\delta_1^* = \frac{1}{2}$ for any $a$ and $b$ with $b^2 \neq a^2$. This design coincides with the $A$-optimal design for estimating the complete state vector $s$ (see Remark 1). In case of more than one terminal node, i.e. $I > 1$, the $A$-optimal design for estimating the linear aspect $L\,s$ depends on the relationship between $a$ and $b$ as shown in the following example.

***Example 3*** We consider a similar situation as in Example 2. Therefore, we can set $a = 1$ without loss of generality, since the $A$-optimal designs are only influenced by the ratio of the values $a$ and $b$. Figure 3 shows the optimal values for $\delta_1$ for estimating $L\,s$ depending on the quantity $b$ for $I = 4, 9, 25$ terminal nodes and $a = 1$. It shows in particular that the special case $b^2 = \frac{1}{I}$, where the state vector $s$ is not identifiable and the $A$-optimal design is given by $\delta_0^* = 0$ and $\delta_1^* = \ldots \delta_I^* = \frac{1}{I}$, is not a smooth continuation of the case $b^2 \neq \frac{1}{I}$. Furthermore, if the influence $b$ of the central node 0 goes to infinity, the $A$-optimal weights $\delta_1 = \ldots = \delta_I$ at the terminal nodes $i = 1, \ldots, I$ go to $\frac{1}{I}$, i.e. converge to the $A$-optimal weight in the case of no identifiability of $s$. This again means that only a small proportion of observations should be done at the central node if it has a big influence on its adjacent nodes and vice versa. As for estimating the complete state vector $s$, Model A with random states

and no measurements errors provides larger weights $\delta_0^*$ at the central node and thus smaller weights at the terminal nodes than the Model A with nonrandom states and measurement errors. Nevertheless, the differences are relatively small.

## 5 *A*-optimal designs in a wheel network

In this section, we consider the other network introduced in Example 1, namely the wheel network with $I+1$ nodes. In that case, the network consists of the central node 0, which is connected to all remaining nodes, whereas the remaining nodes are both connected to the central node and two other nodes. More precisely, we consider the influence matrix $\mathbb{X}$ of the form

$$\mathbb{X} = \begin{pmatrix} A & B^\top \\ B & \tilde{\mathbb{X}} \end{pmatrix}, \tag{27}$$

where the matrices $A \in \mathbb{R}^{2\times 2}$ and $B \in \mathbb{R}^{(I-1)\times 2}$ are of the form

$$A = \begin{pmatrix} a & b \\ b & a \end{pmatrix}, \qquad B = b\left(\mathbf{1}_{I-1},\, u_1^{I-1} + u_{I-1}^{I-1}\right), \tag{28}$$

and $u_j^{I-1}$ denotes the $j$-th unit vector in $\mathbb{R}^{I-1}$. The matrix $\tilde{\mathbb{X}} \in \mathbb{R}^{(I-1)\times(I-1)}$ contained in $\mathbb{X}$ is a Toeplitz-triadiagonal matrix with main diagonal elements equal to $a$, whereas the lower and upper diagonal elements are equal to $b$, that is,

$$\tilde{\mathbb{X}}_{i,j} = \begin{cases} a, & i = j \\ b, & i = j-1 \text{ or } i = j+1 \\ 0, & \text{else} \end{cases}. \tag{29}$$

We are now interested in determining the $A$-optimal design for estimating the complete state vector $s$. In that case, the influence matrix $\mathbb{X}$ given by (27) must be non-singular, otherwise the state vector will not be identifiable. The following Lemma 5 contains conditions on the influencing values $a > 0$ and $b > 0$ of the network, which ensure the non-singularity of $\mathbb{X}$.

**Theorem 5** *Let $\mathbb{X} \in \mathbb{R}^{(I+1)\times(I+1)}$ be of the form (27) with corresponding matrices $A$ and $B$ of the form given by (28). Then the following statements hold:*

(a) *The eigenvalues of the Toeplitz-tridiagonal matrix $\tilde{\mathbb{X}}$ are given by*

$$\lambda_i = a + 2b\cos(\tfrac{i}{I}\pi),\ i = 1, \ldots, I-1\,. \tag{30}$$

*The eigenvector corresponding to $\lambda_i$ is of the form*

$$v_i = \sqrt{\frac{2}{I}}\left(\sin(\tfrac{i}{I}\pi), \ldots, \sin(\tfrac{i(I-1)}{I}\pi)\right)^\top, i = 1, \ldots, I-1\,. \tag{31}$$

(b) *The matrix* $\tilde{\mathbb{X}}$ *is singular if and only if* $k \in \{1, \ldots, I-1\}$ *exists such that* $\cos(\frac{k}{I}\pi) = -\frac{a}{2b}$. *Moreover, the rank of* $\tilde{\mathbb{X}}$ *is at least* $I-2$.

(c) *Let* $\tilde{\mathbb{X}}$ *be non-singular. Then* $\mathbb{X}$ *is singular if and only if the values* $a, b$ *solve the equation*

$$(a - b^2c_1)(a - b^2c_3) - (b - b^2c_2)^2 = 0\,, \tag{32}$$

*where the values of* $c_1, c_2, c_3$ *are given by*

$$\begin{aligned}
c_1 &= \frac{2}{I}\sum_{i=1}^{I}\frac{1}{\lambda_i}\frac{\sin^2(\frac{i}{2}\pi)\sin^2(\frac{(I-1)i}{2I}\pi)}{\sin^2(\frac{i}{2I}\pi)}\,,\\
c_2 &= \frac{2}{I}\sum_{i=1}^{I}\frac{1}{\lambda_i}\frac{\sin(\frac{i}{2}\pi)\sin(\frac{(I-1)i}{2I}\pi)}{\sin(\frac{i}{2I}\pi)}(\sin(\tfrac{(I-1)i}{2I}\pi) + \sin(\tfrac{i}{2I}\pi))\,,\\
c_3 &= \frac{2}{I}\sum_{i=1}^{I}\frac{1}{\lambda_i}(\sin(\tfrac{i}{I}\pi) + \sin(\tfrac{i(I-1)}{I}\pi))^2\,,
\end{aligned}$$

*and* $1/\lambda_1, \ldots, 1/\lambda_I$ *are the eigenvalues of the matrix* $\tilde{\mathbb{X}}$ *given by* (30).

***Proof*** (a). The statement is a well-known result used in a lot of applications, for instance, in order to solve specific types of differential equations. The result can be found in Noschese et al. (2013) among others.
(b). The matrix $\tilde{\mathbb{X}}$ becomes singular if and only if one of its eigenvalues is equal to zero. The first statement of (b) follows by setting equation (30) equal to zero. The second statement of (b) follows by the fact, that $\tilde{\mathbb{X}}$ has $I-1$ distinct eigenvalues. Consequently, at most one eigenvalue can be equal to zero and the resulting dimension of the eigenvectors corresponding to the non-zero eigenvalues is equal to $I-2$.
(c). Under the assumption that $\tilde{\mathbb{X}}$ is non-singular, the determinant of the influence matrix $\mathbb{X}$ can be reformulated in terms of the schur complement of $\tilde{\mathbb{X}}$, that is,

$$\det(\mathbb{X}) = \det(\tilde{\mathbb{X}})\det(A - B^\top\tilde{\mathbb{X}}^{-1}B)\,,$$

(see Harville (1997), Theorem 13.3.8). Since $\tilde{\mathbb{X}}$ is non-singular, it holds $\det(\tilde{\mathbb{X}}) \neq 0$. Consequently, $\mathbb{X}$ is singular if and only if $\det(A - B^\top\tilde{\mathbb{X}}^{-1}B) = 0$.
We first concentrate on determining $B^\top\tilde{\mathbb{X}}^{-1}B$. Using part (a), the inverse $\tilde{\mathbb{X}}^{-1}$ can be represented in terms of the eigenvalues and eigenvectors of $\tilde{\mathbb{X}}$, that is,

$$\tilde{\mathbb{X}}^{-1} = \sum_{i=1}^{I-1}\frac{1}{\lambda_i}v_i v_i^\top\,, \tag{33}$$

where $\lambda_i$ and $v_i$ are given by (30) and (31), respectively.

Due to the structure of the matrix $B$ given in (28), we obtain

$$B^\top \tilde{\mathbb{X}} B = b^2 \begin{pmatrix} c_1 & c_2 \\ c_2 & c_3 \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

where the value $c_1$ contains the sum of all elements of the matrix $\tilde{\mathbb{X}}^{-1}$, $c_2$ is the sum of the first and last row of $\tilde{\mathbb{X}}^{-1}$, and $c_3$ is the sum of the upper right and upper left and bottom right and bottom left elements of $\tilde{\mathbb{X}}^{-1}$. Using the formula we obtain the expressions for $c_1, c_2$ and $c_3$ given in part (c). □

***Remark 3*** If the matrices $\tilde{\mathbb{X}}$ and $\mathbb{X}$ are non-singular, Theorem 1 can be used to calculate the $A$-optimal design for estimating $s$. In particular, the inverse $\mathbb{X}^{-1}$, which is needed for the determination of the $A$-optimal weights $\delta_0, \delta_1, \ldots, \delta_I$, can be calculated by using formula (21) and the inverse of $\tilde{\mathbb{X}}$ given by (33).

Note that the non-singularity of the Toeplitz-tridiagonal matrix $\tilde{\mathbb{X}}$ is not a necessary condition for the non-singularity of the influence matrix $\mathbb{X}$. If the values $a$ and $b$ result in a singular $\tilde{\mathbb{X}}$, the influence matrix $\mathbb{X}$ can still be non-singular and Theorem 1 be applicable. In case of $\tilde{\mathbb{X}}$ being singular, the influence matrix $\mathbb{X}$ should be partioned in the following way:

$$\mathbb{X} = \begin{pmatrix} \hat{A} & \hat{B}^\top \\ \hat{B} & \hat{\mathbb{X}} \end{pmatrix},$$

where the matrices $\hat{A} \in \mathbb{R}^{3\times 3}$ and $\hat{B} \in \mathbb{R}^{(I-2)\times 3}$ are of the form

$$\hat{A} = \begin{pmatrix} a & b & b \\ b & a & b \\ b & b & a \end{pmatrix}, \qquad \hat{B} = b \left( \mathbf{1}_{I-2},\, u_1^{I-2},\, u_{I-2}^{I-2} \right),$$

and the matrix $\hat{\mathbb{X}}$ is a Toeplitz-tridiagonal matrix of dimension $I-2$. Note that Part a) and b) of Lemma 5 also hold for the matrix $\hat{\mathbb{X}}$. In particular, it follows that $\hat{\mathbb{X}}$ is non-singular, as the fixed values $a, b > 0$ cannot both result in a trivial eigenvalue of $\tilde{\mathbb{X}}$ and $\hat{\mathbb{X}}$. If $\hat{\mathbb{X}}$ is non-singular, $\mathbb{X}$ is non-singular if and only $\det(\hat{A} - \hat{B}^\top \hat{\mathbb{X}}^{-1} \hat{B}) \neq 0$. Note that the inverse of $\hat{\mathbb{X}}$ can be determined by using part a) of Lemma 5.

We conclude this section by considering an example of a wheel network, which is similar to Example 2.

***Example 4*** Let the network be given by the wheel-network with $I = 4, 9, 25$ (non-central) nodes. We investigate the behaviour of the corresponding $A$-optimal designs in dependence of the different values of $a$ and $b$ in the situation of Model A, where either nonrandom states are given, i.e. $\rho_Z = 0$, or no measurement errors occur, i.e. $\rho_E = 0$. In the situations under consideration, the $A$-optimal designs only depend on the relation between the values $a$ and $b$ and we can set $a = 1$ without loss of generality. Thus, for each number of non-central nodes $I$, we vary the value $b$ in the interval $[0, 3]$. In order to exclude the values of $b$, where the influence matrix $\mathbb{X}$ is

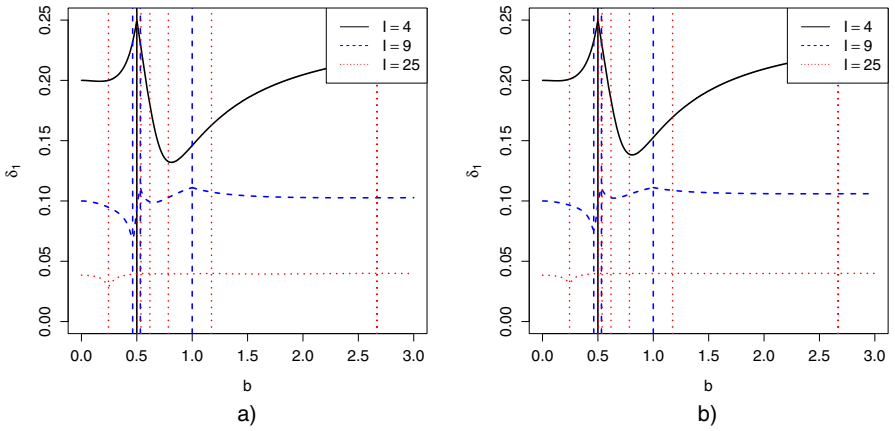


**Fig. 4** $A$-optimal values for $\delta_1$ $(= \delta_2 = \ldots = \delta_I)$ for estimating the state vector $s$ in the wheel network depending on the quantity $b$, the external influence of the other nodes, wheras the influence on itself is given by $a = 1$. Left: in Model B and in Model A with nonrandom states, i.e. $\rho_Z = 0$. Right: in Model A with random states and no measurements errors, i.e. $\rho_E = 0$. The vertical lines indicate the crucial cases where the influence matrix $\mathbb{X}$ is singular (i.e. the state vector $s$ is not identifiable)

singular, we use Lemma 5 and Remark 3. For each $I$ and the corresponding remaining values of $b$ the $A$-optimal designs are calculated numerically using again Remark 3. First, we observe that the numerically calculated $A$-optimal weights $\delta_1^*, \ldots, \delta_I^*$ of the non-central nodes coincide in the wheel network, that is $\delta_1^* = \delta_2^* = \ldots = \delta_I^*$, for $I = 4, 9, 25$. The weight of the central node is then given by $\delta_0^* = 1 - I * \delta_1^*$. Note that the $A$-optimal weights for the non-central nodes also coincide in the star network (see Theorem 4) and it seems that the non-central nodes obtain equal weights as soon as they have a similar adjacency structure.

Figure 4 shows the $A$-optimal values for $\delta_1^*$ depending on the quantity $b$ for the wheel network with $I = 4, 9, 25$ nodes and $a = 1$ for Model A with nonrandom states given by $\rho_Z = 0$ (a)) and for Model A with no measurements errors given by $\rho_E = 0$ (b)). The vertical lines indicate the values of $b$ for which the influence matrix is singular and thus result in a non-identifiable state vector $s$ (independent from the selected design). For $I = 4$, we obtain a singular influence matrix $\mathbb{X}$ for $b = 0.5 \in [0, 3]$. For $I = 9$, the matrix $\mathbb{X}$ is not invertible if $b \in \{0.464, 0.532, 1\} \subset [0, 3]$. If the wheel has $I = 25$ non-central nodes, the influence matrix $\mathbb{X}$ is singular if $b$ is contained in the set $\{0.244, 0.504, 0.538, 0.618, 0.784, 1.174, 2.668\} \subset [0, 3]$.

Similar to the star network, we observe that the weight $\delta_1^*$ is not monotonically increasing with $b$. Moroever, for $b$ going to infinity, the $A$-optimal weight $\delta_1^*$ goes again to $\frac{1}{I}$ in each of the considered cases. Consequently, the weight $\delta_0^*$ tends to zero and a small amount of observations at the central node is sufficient in the wheel network, if the influence of the other nodes is great and vice versa. Comparing the $A$-optimal weight for Model A with random states to the corresponding weight for Model A with nonrandom states, the curves are similar for all cases of nodes $I = 4, 9, 25$ under consideration. In particular, we observe that the weight $\delta_1^*$ are slightly smaller, if random states are considered instead of nonrandom states. Note again that the designs

for Model B coincide with those of Model A with nonrandom states such that the left hand side of Fig. 4 also describes the $A$-optimal weight for $\delta_1$ in Model B.

## 6 Discussion

We have derived a general characterization of $A$-optimal designs for networks where all adjacent nodes have the same influence on the state of a node. For the most simple network, a network with star configuration, we derived the A-optimal designs explicitly. Moreover, we showed that not always all expected states are identifiable and we derived $A$-optimal designs for an aspect of the states which is always identifiable.

Moreover, we considered a more complex network with wheel configuration and derived analytical conditions on the influences of the states which ensure the identifiability of all states.

The star and the wheel configuration lead to similar results: Depending on the influence of the states at adjacent nodes, the $A$-optimal design puts more or less weight at the central node than at the non-central nodes while the non-central nodes get always equal weights. The higher the influence of the adjacent nodes is the smaller should be the weight at the central node. This means in particular that less precise measurements can be used at the central node and the more precise measurements should be used at the non-central nodes when the influence of adjacent nodes is high.

These results can also be used to simplify the numerical calculation of the $A$-optimal design, which is necessary due to the numerical instability of the original design problem. For instance the numerical calculation for a mixture of a star and a wheel network can be simplified if the results about the $A$-optimal designs of the individual networks are used. The remaining problem can then be solved by several optimization algorithms as the multiplicative algorithm developed by Yu (2010) or a Particle Swarm optimization algorithm (see Kennedy and Eberhart (1995) among many others).

For the considered networks, we obtained equal $A$-optimal weights at nodes that have a similar adjacency structure, in particular, they have the same number of adjacent nodes. The proof of that observation will be addressed in further research. Moreover, the presented approach is based on the assumption that the influence of adjacent nodes is the same and known. It is an open problem how to derive optimal designs if this is not the case.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflicts of interest** The authors declare that they have no conflict of interest.

## References

Azhdari A, Ardakan MA (2022) Reliability optimization of multi-state networks in a star configuration with bi-level performance sharing mechanism and transmission losses. Reliabil Eng Syst Saf 226:108556

Cao M, Guo J, Xiao H, Wu L (2022) Reliability analysis and optimal generator allocation and protection strategy of a non-repairable power grid system. Reliabil Eng Syst Saf 222:108443

Duan P, He L, Huang L, Chen G, Shi L (2022) Sensor scheduling design for complex networks under a distributed state estimation framework. Automatica 146:110628

Harville DA (1997) Matrix algebra from a statistician's perspective. Springer, New York

Kennedy J, Eberhart R (1995) Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*. IEEE, vol. 4, pp 1942–1948

Kubrusly C, Malebranche H (1985) Sensors and controllers location in distributed systems: a survey. Automatica 21(2):117–128

Li Q, Negi R, Ilić MD (2011) Phasor measurement units placement for power system state estimation: a greedy approach. In *2011 IEEE power and energy society general meeting*, pp 1–8

Muscas C, Pau M, Pegoraro PA, Sulis S (2014) Effects of measurements and pseudomeasurements correlation in distribution system state estimation. IEEE Trans Instrum Meas 63(12):2813–2823

Noschese S, Pasquini L, Reichel L (2013) Tridiagonal toeplitz matrices: properties and novel applications. Numer Linear Algebra Appl 20(2):302–326

Patan M, Patan K (2005) Optimal observation strategies for model-based fault detection in distributed systems. Int J Control 78(18):1497–1510

Pukelsheim F (2006) Optimal design of experiments. SIAM, Philadelphia

Pukelsheim F, Rieder S (1992) Efficient rounding of approximate designs. Biometrika 79(4):763–770

Rencher AC (1998) Multivariate statistical inference and applications. Wiley series in probability and statistics. Wiley, New York

Schlösser T, Angioni A, Ponci F, Monti A (2014) Impact of pseudo-measurements from new load profiles on state estimation in distribution grids. In *2014 IEEE international instrumentation and measurement technology conference (I2MTC) proceedings*, pp 625–630

Schurtz A (2020) *Verfahren zur Zustandsschätzung und ihr Beitrag zum Engpassmanagement in Mittelspannungsnetzen*. Dissertation, TU Dortmund, Dortmunder Beiträge zu Energiesystemen, Energieeffizienz und Energiewirtschaft, Shaker-Verlag, ISBN 978-3-8440- 7638-7

Singhal H, Michailidis G (2008) Optimal sampling in state space models with applications to network monitoring. SIGMETRICS Perform Eval Rev 36(1):145–156

Su P, Wang G (2020) Reliability of a star configuration power grid system with performance sharing. In *2020 Asia-Pacific international symposium on advanced reliability and maintenance modeling (APARM)*, pap 1–6

Uciński D (2000) Optimal sensor location for parameter estimation of distributed processes. Int J Control 73(13):1235–1248

Uciński D (2004) Optimal measurement methods for distributed parameter system identification. CRC Press, Boca Raton

Uciński D (2022) E-optimum sensor selection for estimation of subsets of parameters. Measurement 187:110286

Xygkis TC, Korres GN, Manousakis NM (2018) Fisher information-based meter placement in distribution grids via the D-optimal experimental design. IEEE Trans Smart Grid 9(2):1452–1461

Yu Y (2010) Monotonic convergence of a general algorithm for computing optimal designs. Ann Stat 38(3):1593–1606