On the relation between non-normality and diameter in linear dynamical networks

Giacomo Baggio and Sandro Zampieri

***Abstract*— Understanding how the "degree" of non-normality of a networked system is connected with the topological structure of the underlying graph is of crucial importance in many areas of the engineering and natural sciences, most notably in the controllability analysis of large-scale networks. This paper explores this relation in terms of the graph *diameter*. More precisely, we derive diameter-dependent upper and lower bounds on network non-normality. Further, we outline a gradient-based optimization procedure to increase the non-normality of a network.**

***Index Terms*— Matrix non-normality, linear dynamical networks, transient amplification, graph diameter.**

I. Introduction

A basic result of linear system theory states that spectrum of the state matrix determines the *asymptotic behavior* of an LTI system. On the other hand, the *transient behavior* of the system depends on the "degree" of non-normality of the latter matrix [1]. The departure from normality of a matrix A can be measured in several different ways [1, Ch. 48], ranging from the condition number of the eigenvector matrix of A to the (Frobenius) norm of the difference $AA^\top - A^\top A$. In this paper, we focus on two particular measures of non-normality for linear dynamical systems that involve the energy of the impulse response of the system. Our main objective is to examine the relation between these non-normality measures and the topological features of the underlying graph.

By means of analytical and numerical results, we show that for positive systems a feature that seems intimately connected with network non-normality is the *diameter* of the underlying graph. In particular, we prove that a positive network driven by a Metzler matrix cannot feature an high degree of non-normality unless it has a large diameter. Conversely, we show that, under certain conditions, a large diameter automatically guarantees an high degree of non-normality. Finally, we develop an algorithm for the maximization of the degree of non-normality of a given linear dynamical network. This procedure leads to networks that exhibit some preferred "anisotropic" directions which are, in turn, connected with the graph diameter.

The main motivation behind the present work comes from the fact that the analysis and classification of networks featuring an high degree of non-normality, as well as the synthesis of such networks via iterative and unsupervised procedures, have witnessed an increase of interest in recent years. In fact, these topics have been investigated in a variety of scientific contexts under different names, for instance:

- in neuroscience, non-normal networks arise in the study of short-term memory capacity of linearized neuronal networks [2]–[4], in the modelling of motor cortex activity [5] and of spontaneously generated activity in the visual cortex [6], [7]. In this context, non-normality is commonly referred as *patterned amplification* and is often measured in terms of length of the so-called *feed-forward chains* of the network;
- in econometrics, non-normality is known as *network volatility* [8], [9] and is tightly connected with issues related to sensitivity and robustness of financial markets [10];
- in control theory, and specifically in the controllability analysis of large-scale networks, non-normal networks have been labelled *anisotropic networks* in [11]. These networks present favourable properties in terms of energy required to steer an initial state to a target one, whereas networks that are close to be normal require a large amount of energy to perform the same task [12]–[15].

To the best of our knowledge, the only works that try to connect a measure of non-normality with a topological feature of the underlying graph are [15], [16], both in the framework of network controllability. In the first paper a relation between the worst-case control energy of positive networks and some notions of network centrality is brought to light. In the second paper, the authors investigate the connection between the latter controllability metric and the graph diameter for a particular class of networks. However, the analysis there is quite restrictive in that the main result, which is a diameter-dependent upper bound, apply to a very special class of networks that essentially consists of acyclic networks. Moreover, a diameter-dependent lower bound is not discussed. Here instead we provide a more complete picture, considering the whole class of positive networks and deriving diameter-dependent upper and lower bounds.

Paper structure: The paper is organized as follows. Section II introduces two measures of network non-normality that will be analyzed in the rest of the paper. In Section III, we derive upper and lower bounds on the first of the latter measures in terms of the diameter of the underlying graph. In Section IV, we outline a gradient-based numerical procedure for maximizing the second non-normality measure. In the same section we illustrate and discuss the numerical results obtained by applying this procedure. Ultimately, in Section

G. Baggio and S. Zampieri are with the Dipartimento di Ingegneria dell'Informazione, Università di Padova, via Gradenigo, 6/B – I-35131 Padova, Italy. E-mails: {baggio, zampi}@dei.unipd.it. This document is an extended version of the conference paper with full proofs.

V, we draw some concluding remarks and list a number of open questions.

Notation and background results: In what follows, $\bar{a}$, $\mathrm{Re}[a]$, and $|a|$ denote the conjugate, real part, and modulus of a complex number $a \in \mathbb{C}$. $A^\top$ and A^* stand for the transpose and conjugate transpose of $A \in \mathbb{C}^{n\times n}$, respectively. Further, $\|A\|^2 := \lambda_{\max}(AA^*)$, $\|A\|_{\mathrm{F}}^2 := \mathrm{tr}(AA^*)$, and $|A|$, denote the operator norm, the Frobenius norm and the absolute value matrix of $A \in \mathbb{C}^{n\times m}$. We define the spectral abscissa of $A \in \mathbb{C}^{n\times n}$, $\alpha(A)$, as the maximum real part of the eigenvalues of A, namely $\alpha(A) := \max \mathrm{Re}[\lambda(A)]$. The symbol $\mathrm{diag}(a_1, a_n, \dots, a_n)$ stands for the diagonal matrix with entries a_1, a_2, $\dots$, a_n on the diagonal. Finally, A_{ij} denotes the (i,j)-th entry of A, $A_{i:}$ the i-th row of A, and $A_{:j}$ the j-th column of A.

We denote by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the directed graph with vertex (or node) set $\mathcal{V} = \{1, 2, \dots, n\}$, edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. The (weighted) adjacency matrix $A \in \mathbb{R}^{n\times n}$ corresponding to the graph $\mathcal{G}$ satisfies $A_{ij} > 0$ iff $(j,i) \in \mathcal{E}$. We denote by $d(k,t)$ the length of a shortest path from the node k to the node t. We say that a path from k to t is *minimal* if its length is equal to $d(k,t)$. Given $\mathcal{K}, \mathcal{T} \subseteq \mathcal{V}$, we denote by $d(\mathcal{K}, \mathcal{T})$ the maximum length of a shortest path from the nodes of $\mathcal{K}$ to the nodes of $\mathcal{T}$, namely

$$d(\mathcal{K}, \mathcal{T}) := \max \{ \, d(k,t) \mid k \in \mathcal{K}, \, t \in \mathcal{T} \, \} .$$

Notice that, in case $\mathcal{K} \equiv \mathcal{T} \equiv \mathcal{V}$, $d(\mathcal{K}, \mathcal{T})$ coincides with the *diameter* of the graph $\mathcal{G}$. For this reason, $d(\mathcal{K}, \mathcal{T})$ will be termed *relative diameter* of the graph $\mathcal{G}$.

A matrix $A \in \mathbb{R}^{n\times n}$ is said to be positive (non-negative, resp.) if all the entries of A are positive (non-negative). A is said to be Metzler if all the off-diagonal entries of A are non-negative. A is said to be (Hurwitz) stable if all the eigenvalues of A have negative real part. From the Perron–Frobenius theory, it follows that a Metzler matrix A has always one real dominant eigenvalue and the right/left eigenvectors corresponding to this eigenvalue are non-negative [17, Chap. 8]. We term the latter eigenvectors right/left Perron vectors of A. A non-negative or Metzler A is said to be irreducible if for every i, j there exists an integer $k > 0$ s.t. $[A^k]_{ij} > 0$ that is, if A represents the adjacency matrix of a graph, there exists a path from node i to node j. The right/left Perron vectors of a non-negative or Metzler irreducible matrix are always strictly positive. Furthermore, it can be shown that the diagonal entries of a stable Metzler matrix are always negative [18, Chap. 6.4].

II. Measures of network non-normality

Consider a network driven by the continuous-time LTI dynamics

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \quad t \geq 0, \; x(0) =: x_0 \in \mathbb{R}^n, \tag{1}$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$, are the vector of node states, control inputs, and outputs at time t, respectively. $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{n\times p}$ are the adjacency matrix of the network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the input matrix, and the output matrix, respectively. Matrices B and C are used to select a subset of input nodes $\mathcal{K} = \{k_i\}_{i=1}^m \subseteq \mathcal{V}$ and a subset of output nodes $\mathcal{T} = \{t_i\}_{i=1}^p \subseteq \mathcal{V}$, so that they are taken of the form

$$B = \begin{bmatrix} \mathbf{e}_{k_1} & \mathbf{e}_{k_2} & \cdots & \mathbf{e}_{k_m} \end{bmatrix}, \quad C = \begin{bmatrix} \mathbf{e}_{t_1}^\top \\ \mathbf{e}_{t_2}^\top \\ \vdots \\ \mathbf{e}_{t_p}^\top \end{bmatrix},$$

where $\{\mathbf{e}_k\}_{k=1}^n$ denote the canonical basis in $\mathbb{R}^n$.

In this paper, we focus on the following two measures of non-normality:

$$\mathrm{nn}_2(A,B,C) := \sup_{t\geq 0} \left\| Ce^{At}B \right\|, \tag{2}$$

$$\mathrm{nn}_{\mathrm{F}}(A,B,C) := \int_0^\infty \left\| Ce^{At}B \right\|_{\mathrm{F}}^2 \mathrm{d}t. \tag{3}$$

Notice that $\mathrm{nn}_2(A,B,C)$ coincides with the worst-case 2-norm of the impulse response of the LTI system (1), whereas $\mathrm{nn}_{\mathrm{F}}(A,B,C)$ with the $\mathcal{H}_2$ norm of the system. The latter norm measures the "energy" of the impulse response of the system (1) [19, Ch. 4]. Note also that the above two measures are well-defined only in case the system has a bounded impulse response. In particular, this is always verified if A is stable.

Remark 1: Notice that expression (2) can be thought of as a measure of non-normality of a matrix A in case $C = B = I$, namely

$$\mathrm{nn}_2(A,I,I) = \sup_{t\geq 0} \left\| e^{At} \right\|$$

This measure of non-normality has been analyzed in many works, see e.g. [1, Chap. 15] wherein several bounds on this index based on the notion of pseudospectrum of a matrix can be found.

Remark 2: In case $C = I$, we have

$$\mathrm{nn}_{\mathrm{F}}(A,B,I) = \int_0^\infty \mathrm{tr}\left(e^{At}BB^\top e^{A^\top t} \right) \mathrm{d}t = \mathrm{tr}(\mathcal{W}_c)$$

where $\mathcal{W}_c$ the controllability Gramian of the system, while, if $B = I$,

$$\mathrm{nn}_{\mathrm{F}}(A,I,C) = \int_0^\infty \mathrm{tr}\left(e^{A^\top t}C^\top Ce^{At} \right) \mathrm{d}t = \mathrm{tr}(\mathcal{W}_o).$$

where $\mathcal{W}_o$ the observability Gramian of the system. We point out, in particular, that the controllability Gramian is related to the amount of energy required to steer the zero state to a target one. In this context, $\mathrm{tr}(\mathcal{W}_c)$ has been analyzed as a measure of network controllability in [20].

Remark 3: As briefly mentioned in the introduction, there are many ways to quantify the non-normality of a matrix and, therefore, the non-normality of a linear dynamical network. Our main motivation behind the choice of the two measures $\mathrm{nn}_2(A,B,C)$ and $\mathrm{nn}_{\mathrm{F}}(A,B,C)$ hinges on the fact that for the first measure it is possible to provide meaningful analytical bounds (Sec. III), while for the second measure a closed-form expression of its gradient can be derived, rendering this measure more suitable for optimization (Sec. IV). We stress however that, although these two measures may look very similar at a first sight, establishing a precise relation between these two still represents a non-trivial open problem.

III. Bounds on network non-normality

In this section, we derive upper and lower bounds on network non-normality for (irreducible) Metzler matrices. We focus, in particular, on the stable case. The bounds relate the non-normality measure $\mathrm{nn}_2(A,B,C)$ to the relative diameter $d(\mathcal{K},\mathcal{T})$ of the underlying graph.

A. Upper bounds

Theorem 1: Consider the linear system in (1) and let $A \in \mathbb{R}^{n\times n}$ be a stable irreducible Metzler matrix. It holds

$$\mathrm{nn}_2(A,B,C) \le \left(\frac{\beta}{a_{\min}}\right)^{d(\mathcal{K},\mathcal{T})}, \tag{4}$$

where $a_{\min}$ is the smallest off-diagonal non-zero entry of A and

$$\beta := \alpha(A) + d_{\max}$$

with $d_{\max}$ is the largest diagonal entry in modulus of A.

Proof: Firstly, we observe that, since A is irreducible, the Perron vectors of A are strictly positive. Let $w, v \in \mathbb{R}^n$ denote the strictly positive left, right (resp.) Perron vectors of A and define

$$D := \mathrm{diag}\left[\frac{v_1}{w_1}, \frac{v_2}{w_2}, \dots, \frac{v_n}{w_n}\right].$$

From sub-multiplicativity of matrix norm $\|\cdot\|$, it follows that, for all $t \ge 0$,

$$\begin{aligned}\left\|Ce^{At}B\right\| &= \left\|CD^{1/2}e^{D^{-1/2}AD^{1/2}t}D^{-1/2}B\right\| \\ &\le \|CD^{1/2}\|\|D^{-1/2}B\| \left\|e^{D^{-1/2}AD^{1/2}t}\right\| \\ &\le \kappa_D^{1/2}\left\|e^{D^{-1/2}AD^{1/2}t}\right\|,\end{aligned} \tag{5}$$

where

$$\kappa_D := \frac{\max_{i\in\mathcal{T}} D_{ii}}{\min_{i\in\mathcal{K}} D_{ii}}.$$

From the proof of [21, Theorem 18, point iii)], we have that[1]

$$DA + A^\top D - 2\alpha(A)D \le 0,$$

or, equivalently,

$$D^{-1/2}AD^{1/2} + D^{1/2}A^\top D^{-1/2} - 2\alpha(A)I \le 0,$$

Thus, from [22, Prop. 5.5.33], it holds

$$\left\|e^{D^{-1/2}AD^{1/2}t}\right\| \le e^{\alpha(A)t}.$$

In view of (5), this in turn implies

$$\mathrm{nn}_2(A) = \sup_{t\ge 0}\left\|Ce^{At}B\right\| \le \kappa_D^{1/2}, \tag{6}$$

since $\alpha(A) < 0$ by assumption. Further, we notice that

$$\kappa_D \le \kappa_v \kappa_w,$$

[1]Here "$\le$" denotes the order relation between symmetric matrices, that is $A \ge B$ if $A - B$ is symmetric and positive semidefinite.

where

$$\kappa_v := \frac{\max_{i\in\mathcal{T}} v_i}{\min_{i\in\mathcal{K}} v_i}, \quad \kappa_w := \frac{\max_{i\in\mathcal{K}} w_i}{\min_{i\in\mathcal{T}} w_i}.$$

Next, we derive an upper bound on κ_v. This bound is inspired by a result of Ostrowski [23] (see also [24, Theorem 2.7]). Consider a path $\mathcal{P} = \{p_1, p_2, \dots, p_r\}$ ($p_i \ne p_{i+1}$ for $i = 1,\dots,r-1$) connecting node p_1 to node p_r where $p_1 = \arg\min_{i\in\mathcal{K}} v_i$ and $p_r = \arg\max_{i\in\mathcal{T}} v_i$, it holds

$$(\alpha(A) - A_{p_ip_i})v_{p_i} = \sum_{j\ne p_i} A_{p_ij}v_j \ge a_{\min}v_{p_{i+1}} > 0,$$

where $a_{\min}$ denotes the smallest non-zero off-diagonal entry of A. From the previous expression, it follows that

$$\frac{v_{p_r}}{v_{p_1}} = \kappa_v \le \prod_{i=1}^{r-1}\frac{\alpha(A) - A_{p_ip_i}}{a_{\min}}.$$

Since $d(\mathcal{K},\mathcal{T}) \ge r-1$, we have

$$\kappa_v \le \left(\frac{\alpha(A) + d_{\max}}{a_{\min}}\right)^{d(\mathcal{K},\mathcal{T})}. \tag{7}$$

By applying the same argument to $A^\top$, we arrive at an identical inequality for κ_w. Eventually, by plugging (7) into (6), we obtain the desired upper bound. ∎

The following corollary provides a more readable version of the previous theorem.

Corollary 1: Consider the linear system in (1) and let $A \in \mathbb{R}^{n\times n}$ be a stable irreducible Metzler matrix. It holds

$$\mathrm{nn}_2(A,B,C) \le \left(\frac{\beta'}{a_{\min}}\right)^{\mathrm{diam}(\mathcal{G})}, \tag{8}$$

where $a_{\min}$ is smallest off-diagonal non-zero entry of A, and

$$\beta' := \min\left\{d_{\max}, d_{\max} - d_{\min} + \max_i R_i\right\},$$

with $d_{\max}$, $d_{\min}$ are the largest, smallest diagonal entry in modulus of A and $R_i = \sum_{j\ne i} A_{ij}$.

Proof: We have that

1) $\beta = \alpha(A) + d_{\max} \le d_{\max}$,
2) $\alpha(A) \le -d_{\min} + \max_i R_i$ by virtue of Gershgorin circle theorem, so that $\beta \le d_{\max} - d_{\min} + \max_i R_i$.

A combination of the latter two bounds yields the desired result. ∎

For the case of networks described by complete graphs, we have the following simpler result, which is *not* a direct consequence of Theorem 1 and Corollary 1.

Proposition 1: Consider the linear system in (1) and let $A \in \mathbb{R}^{n\times n}$ be a stable irreducible Metzler matrix such that $A_{ij} > 0$ for all $i \ne j$. It holds

$$\mathrm{nn}_2(A,B,C) \le \frac{\max\{d_{\max} - d_{\min} + a_{\min}, a_{\max}\}}{a_{\min}}, \tag{9}$$

where $d_{\max}$, $d_{\min}$ denote the largest, smallest (resp.) diagonal entry in modulus of A, and $a_{\max}$, $a_{\min}$ the largest, smallest (resp.) off-diagonal entry of A.

Proof: Let us define the (full) positive matrix

$$P := A + (d_{\max} + \gamma)I, \quad \gamma > 0.$$

Let v be the right Perron vector of P. Since P is positive the following result holds [23, Eq. (10)]

$$\frac{\max_i v_i}{\min_i v_i} \leq \frac{p_{\max}}{p_{\min}},$$

where $p_{\max}$ and $p_{\min}$ are the largest and smallest entry of P, respectively. Further, it holds

$$\begin{aligned} p_{\max} &= \max\left\{\max_i P_{ii}, \max_{i\neq j} P_{ij}\right\} \\ &= \max\left\{d_{\max} + \gamma + \max_i A_{ii}, \max_{i\neq j} A_{ij}\right\} \\ &= \max\left\{d_{\max} + \gamma - d_{\min}, a_{\max}\right\}, \end{aligned}$$

and

$$\begin{aligned} p_{\min} &= \min\left\{\min_i P_{ii}, \min_{i\neq j} P_{ij}\right\} \\ &= \min\left\{d_{\max} + \gamma + \min_i A_{ii}, \min_{i\neq j} A_{ij}\right\} \\ &= \min\left\{\gamma, a_{\min}\right\}. \end{aligned}$$

In view of the above expressions, we have

$$\frac{p_{\max}}{p_{\min}} = \frac{\max\left\{d_{\max} + \gamma - d_{\min}, a_{\max}\right\}}{\min\left\{\gamma, a_{\min}\right\}}.$$

and the minimum is achieved for $\gamma = a_{\min}$. Finally, a reasoning similar to that used in the proof of Theorem 1 completes the proof. ■

Remark 4: The results in Theorem 1, Corollary 1, Proposition 1 assert that the relative diameter $d(\mathcal{K}, \mathcal{T})$ plays a crucial role in increasing the non-normality of a network driven by a stable Metzler state matrix. More specifically, if the network matrix A has bounded entries and the underlying graph has bounded node degrees, then, by virtue of Corollary 1, as n tends to infinity the non-normality of the network can increase only if $d(\mathcal{K}, \mathcal{T})$ increases with n.

The bounds in Theorem 1 and Proposition 1 apply also to systems described by general stable A's (i.e., not necessarily Metzler) as long as their Metzler part,

$$\mathcal{M}(A) := \mathrm{diag}(A) + |A - \mathrm{diag}(A)|,$$

satisfies the assumptions used in these results. This fact is an immediate consequence of the following result, which follows from [24, Theorem 13 and Corollary 14].

Proposition 2: Let $A \in \mathbb{R}^{n\times n}$ be a general matrix and let $\mathcal{M}(A)$ denote its Metzler part. It holds

$$\left\|Ce^{At}B\right\| \leq \left\|Ce^{\mathcal{M}(A)t}B\right\|, \quad \forall t \geq 0.$$

B. Lower bound

Given a path $\mathcal{P}$ connecting, we denote by $a_{\min}(\mathcal{P})$ the minimum entry of A along the path $\mathcal{P}$. Denote by $\mathcal{P}(\mathcal{K}, \mathcal{T})$ the set of all minimal paths from $\mathcal{K}$ and ending in $\mathcal{T}$ whose length is equal to $d(\mathcal{K}, \mathcal{T})$ and define

$$a_{\min}(\mathcal{K}, \mathcal{T}) := \max_{\mathcal{P}\in\mathcal{P}(\mathcal{K},\mathcal{T})} a_{\min}(\mathcal{P}).$$

Theorem 2: Consider the linear system in (1) and let $A \in \mathbb{R}^{n\times n}$ be a stable Metzler matrix. It holds

$$\mathrm{nn}_2(A, B, C) \geq \frac{1}{e(d+1)}\left(\frac{a_{\min}(\mathcal{K}, \mathcal{T})}{d_{\max}}\right)^d, \qquad (10)$$

where $d_{\max}$ is the largest diagonal entry in modulus of A, and $d := d(\mathcal{K}, \mathcal{T})$.

Proof: From [25, Theorem 3.49] the following inequality holds

$$\mathrm{nn}_2(A, B, C) \geq \sup_{s\in\mathbb{C}\,:\,\mathrm{Re}[s]>0} s\left\|C(sI - A)^{-1}B\right\|. \qquad (11)$$

Now pick any $\gamma \geq d_{\max} > 0$ so that $A_\gamma := A + \gamma I$ is non-negative. Then

$$(sI - A)^{-1} = ((s+\gamma)I - A_\gamma)^{-1} = \sum_{u\geq 0}\frac{A_\gamma^u}{(s+\gamma)^{u+1}},$$

Notice that this series converges for all $s \in \mathbb{C}$ such that $\mathrm{Re}[s] > 0$, since $\alpha(A) < 0$. Then

$$\begin{aligned} \left\|C(sI - A)^{-1}B\right\|^2 &= \left\|C\sum_{u\geq 0}\frac{A_\gamma^u}{(s+\gamma)^{u+1}}B\right\|^2 \\ &= \max_{\|x\|=1} x^\top C\sum_{u\geq 0}\frac{A_\gamma^k}{(s+\gamma)^{u+1}}BB^\top\sum_{w\geq 0}\frac{(A_\gamma^\top)^w}{(\bar{s}+\gamma)^{w+1}}C^\top x. \end{aligned}$$

Notice that $BB^\top = \sum_{k\in\mathcal{K}} \mathbf{e}_k\mathbf{e}_k^\top$. Pick a path $\mathcal{P} \in \mathcal{P}(\mathcal{K}, \mathcal{T})$ such that $a_{\min}(\mathcal{P}) = a_{\min}(\mathcal{K}, \mathcal{T})$ and let $\bar{k} \in \mathcal{K}$ be its starting node and $\bar{t} \in \mathcal{T}$ its ending node. Moreover, let j be such that $\mathbf{e}_j^\top C = \mathbf{e}_{\bar{t}}$. Then, choosing $x = \mathbf{e}_j$, we argue that

$$\begin{aligned} &\left\|C(sI - A)^{-1}B\right\|^2 \geq \\ &\geq \sum_{u,w\geq 0}\sum_{k\in\mathcal{K}} \mathbf{e}_j^\top C\frac{A_\gamma^u}{(s+\gamma)^{u+1}}\mathbf{e}_k\mathbf{e}_k^\top\frac{(A_\gamma^\top)^w}{(\bar{s}+\gamma)^{w+1}}C^\top\mathbf{e}_j \\ &= \sum_{u,w\geq 0}\sum_{k\in\mathcal{K}} \mathbf{e}_{\bar{t}}\frac{A_\gamma^u}{(s+\gamma)^{u+1}}\mathbf{e}_k\mathbf{e}_k^\top\frac{(A_\gamma^\top)^w}{(\bar{s}+\gamma)^{w+1}}\mathbf{e}_{\bar{t}} \end{aligned}$$

Choosing $u = d$, $w = d$, and $k = \bar{k}$, in view of the previous relation and inequality (11), it follows that

$$\left\|C(sI_n - A)^{-1}B\right\|^2 \geq \left(\frac{[A_\gamma^d]_{\bar{t}\bar{k}}}{|s+\gamma|^{2(d+1)}}\right)^2,$$

where $[A_\gamma^d]_{\bar{t}\bar{k}}$ is the entry of the matrix A_γ^d at position $(\bar{t}, \bar{k})$. It is easy to see that this number is greater that or equal to $a_{\min}(\mathcal{K}, \mathcal{T})^d$. Finally, taking $\gamma = d_{\max}$ and $s = d_{\max}/d$ we obtain that

$$\mathrm{nn}_2(A, B, C) \geq \frac{d^d}{(d+1)^{d+1}}\left(\frac{a_{\min}(\mathcal{K}, \mathcal{T})}{d_{\max}}\right)^d.$$

Observe now that

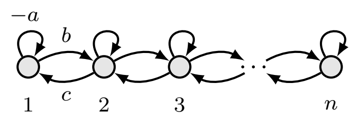
$$\frac{d^d}{(d+1)^{d+1}} \geq \frac{1}{e(d+1)},$$

from which we obtain the thesis. ■

Remark 5: The consequence of the previous result is that, if $a_{\min}(\mathcal{K}, \mathcal{T}) > d_{\max}$ than the index $\mathrm{nn}_2(A, B, C)$ explodes exponentially fast in $d(\mathcal{K}, \mathcal{T})$. This in turn implies that if a network has a large diameter, then, for a particular choice of weights, it is also possible to obtain an high degree of non-normality.

C. A simple example

Consider the following n-dimensional Toeplitz line network ($n \geq 2$)



described by the adjacency matrix

$$A = \begin{bmatrix} -a & c & & \\ b & -a & \ddots & \\ & \ddots & \ddots & c \\ & & b & -a \end{bmatrix} \in \mathbb{R}^{n\times n}, \quad a,\, b,\, c > 0, \quad (12)$$

with $-a + 2\sqrt{bc} < 0$ (stability constraint). Moreover, we set $\mathcal{K} = \{1\}$, $\mathcal{T} = \{n\}$ and suppose, without loss of generality, that $b \geq c$. We recall that the eigenvalues of A have the form

$$\lambda_k(A) = -a + 2\sqrt{bc}\,\cos\left(\frac{k\pi}{n+1}\right),\ k = 1, 2, \ldots, n.$$

In view of Theorem 1 and since $\alpha(A) \leq -a + 2\sqrt{bc}$, we have

$$\mathrm{nn}_2(A, B, C) \leq 2^n \left(\frac{b}{c}\right)^{n/2} =: \overline{\mathrm{nn}}_2,$$

an the right-hand side is always exponentially increasing in n. In view of Theorem 2, we have

$$\mathrm{nn}_2(A, B, C) \geq \frac{1}{e(n+1)} \left(\frac{b}{a}\right)^{n} =: \underline{\mathrm{nn}}_2.$$

Figure 1 shows the log-scale behavior of nn_2, $\overline{\mathrm{nn}}_2$, $\underline{\mathrm{nn}}_2$ as n varies, for a particular choice of the parameters a, b, c that illustrates the effectiveness of the bounds.

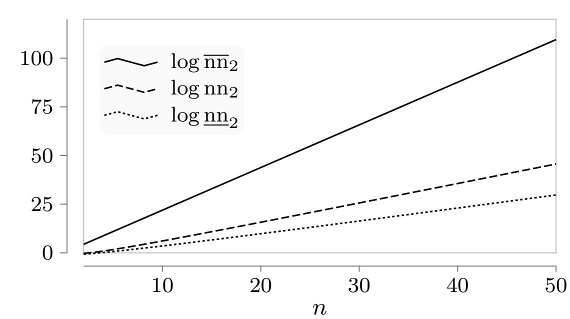


Fig. 1. Log-plot of nn_2, $\overline{\mathrm{nn}}_2$, and $\underline{\mathrm{nn}}_2$ as functions of n, for the Toeplitz line network in Eq. (12) with $a = 1$, $b = 2$, $c = 0.1$.

IV. Synthesis of non-normal networks

In this section, we outline an optimization procedure for increasing the non-normality of a given linear networked system as in (1). In contrast to what done in the previous section, here we focus on the non-normality index $\mathrm{nn}_{\mathrm{F}}(A, B, C)$ which is more amenable to optimization.

A. Problem formulation

We consider the LTI system in (1). The objective is to maximize the non-normality index $\mathrm{nn}_{\mathrm{F}}(A, B, C)$ as defined in (3) w.r.t. matrix A and subject to the following constraints

1) a certain stability margin on A,
2) sparsity constraint on A, depending on the topology induced by $\mathcal{G}$, and
3) upper and lower bounds on the off-diagonal entries of A.

Henceforth, we let $f(A) := \mathrm{nn}_{\mathrm{F}}(A, B, C)$. The aforementioned optimization problem can be formally stated as follows

$$\max_{A\in\mathbb{R}^{n\times n}} f(A) \quad (13)$$
$$\text{s.t. } \alpha(A) < \gamma, \quad (14)$$
$$A_{ij} = 0, \quad \text{if } (i,j) \notin \mathcal{E}, \quad (15)$$
$$\underline{A} \leq |A_{ij}| \leq \overline{A}, \quad \text{if } (i,j) \in \mathcal{E},\ i \neq j, \quad (16)$$

where $\gamma < 0$ is a fixed stability margin and $\underline{A}$, $\overline{A} > 0$, $\underline{A} \leq \overline{A}$. Constraint (14) is the most difficult requirement to take into account, since it depends in a nonlinear way on A. To overcome this issue, we *fix* the spectrum of A, that is we restrict the problem to the set of stable matrices of the form $T^{-1}AT$ with $T \in \mathbb{R}^{n\times n}$ nonsingular and $A \in \mathbb{R}^{n\times n}$ stable s.t. $\alpha(A) \leq \gamma$, describing the network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. In addition, due to the sparsity constraint (15), we further restrict the attention to *diagonal* similarity transformations $D := \mathrm{diag}(d_1, d_2, \ldots, d_n)$, $d_i > 0$, for which constraint (15) is always met. Lastly, note that, up to a rescaling of the elements of D, we may assume $d_1 = 1$, without loss of generality. Thus, defining $A(D) := D^{-1}AD$, we arrive at the following simplified problem

$$\max_{d_2,\ldots,d_n>0} f\left(A(D)\right) \quad (17)$$
$$\text{s.t.} \quad 1/\alpha \leq d_i/d_j \leq \alpha, \quad \text{if } (i,j) \in \mathcal{E}, \quad (18)$$

where $\alpha := \min_{(i,j)\in\mathcal{E}, i\neq j} \alpha_{ij} > 1$ with $\alpha_{ij} := \max\left\{\overline{A}/|A_{ij}|, |A_{ij}|/\underline{A}\right\} > 1$. In order to turn constraint (18) into a linear constraint we define $\delta_i := \log d_i$, so that we have the equivalent problem

$$\max_{\delta_2,\ldots,\delta_n\in\mathbb{R}} f(A(\Delta)) \quad (19)$$
$$\text{s.t.} \quad -\beta \leq \delta_i - \delta_j \leq \beta, \quad \text{if } (i,j) \in \mathcal{E}, \quad (20)$$

where $\Delta := e^{\mathrm{diag}(\delta_1,\ldots,\delta_n)}$ and $\beta := \log\alpha$. As a final remark, by stacking the δ_i's in a vector $\boldsymbol{\delta} := [\delta_1, \delta_2, \ldots, \delta_n]^\top$, we notice that constraint (18) can be written as

$$S^\top \boldsymbol{\delta} \leq \beta \mathbf{1}_N, \quad (21)$$

where $S \in \mathbb{R}^{n\times N}$, $N := |\mathcal{E}|$, coincides with the incidence matrix of the graph $\mathcal{G}$ and $\mathbf{1}_N$ the all-one N-dimensional vector.

B. The algorithm

The procedure we propose for the solution of the constrained maximization problem (19)-(20) is based on a Projected Gradient Ascent (PGA) algorithm, see e.g. [26, Sec. 22.3]. To this aim, we first derive an expression for the partial derivative $\partial f(A(\Delta))/\partial \delta_i$.

Proposition 3: Consider an LTI system as in (1) described by the triple $(A(\Delta), B, C)$. For all $i = 1, 2, \dots, n$, it holds

$$\frac{\partial f(A(\Delta))}{\partial \delta_i} = 2\,\mathrm{tr}\left(\mathcal{W}_c(\Delta)\mathcal{W}_o(\Delta)\Gamma_i\right), \tag{22}$$

where $\mathcal{W}_o(\Delta)$ and $\mathcal{W}_c(\Delta)$ the are observability and controllability Gramians of the system $(A(\Delta), B, C)$, respectively, and

$$\Gamma_i := \Delta^{-1}(A\mathbf{e}_i\mathbf{e}_i^\top - \mathbf{e}_i\mathbf{e}_i^\top A)\Delta. \tag{23}$$

Proof: An application of the chain rule for the derivative of composition of functions [27, Ch. 2.8.1] yields

$$\frac{\partial f(A(\Delta))}{\partial \delta_i} = \mathrm{tr}\left[\left(\frac{\partial f(A(\Delta))}{\partial A(\Delta)}\right)^\top \frac{\partial A(\Delta)}{\partial \delta_i}\right],$$

where

$$\left[\frac{\partial f(A(\Delta))}{\partial A(\Delta)}\right]_{ij} := \frac{\partial f(A(\Delta))}{\partial A(\Delta)_{ij}}.$$

It has been shown in [28, Lemma 3.1] that the latter quantity is given by

$$\frac{\partial f(A(\Delta))}{\partial A(\Delta)} = 2\mathcal{W}_o(\Delta)\mathcal{W}_c(\Delta).$$

So, it remains to show that

$$\frac{\partial A(\Delta)}{\partial \delta_i} = \Gamma_i, \quad i = 2, \dots, n.$$

First, notice that the term δ_i appears in the i-th row and i-th column of $A(\Delta)$ only. So we will restrict the attention to these rows and columns of $A(\Delta)$. The i-th row and i-th column of $A(\Delta)$ have the form

$$A(\Delta)_{i:} = \left[A_{i1}e^{\delta_1-\delta_i} \ \cdots \ A_{ii} \ \cdots \ A_{in}e^{\delta_n-\delta_i}\right],$$
$$A(\Delta)_{:i} = \left[A_{1i}e^{\delta_i-\delta_1} \ \cdots \ A_{ii} \ \cdots \ A_{ni}e^{\delta_i-\delta_n}\right]^\top,$$

respectively. In view of the latter equations, it holds

$$\frac{\partial A(\Delta)_{i:}}{\partial \delta_i} = \left[-A_{i1}e^{\delta_1-\delta_i} \ \cdots \ 0 \ \cdots \ -A_{in}e^{\delta_n-\delta_i}\right],$$
$$\frac{\partial A(\Delta)_{:i}}{\partial \delta_i} = \left[A_{1i}e^{\delta_i-\delta_1} \ \cdots \ 0 \ \cdots \ A_{ni}e^{\delta_i-\delta_n}\right]^\top.$$

It is now a matter of direct computation to conclude that $\partial A(\Delta)/\partial \delta_i = \Gamma_i$, as defined in (23), for all i. ∎

Remark 6: It is worth remarking that the controllability and observability Gramians which appear in the expression of the derivatives (22) can be computed in an efficient and robust way by solving two Lyapunov equations. Indeed, the observability Gramian $\mathcal{W}_o(\Delta)$ is the unique solution of

$$A(\Delta)^\top\mathcal{W}_o(\Delta) + \mathcal{W}_o(\Delta)A(\Delta) = -C^\top C,$$

while the controllability Gramian $\mathcal{W}_c(\Delta)$ is the unique solution of

$$A(\Delta)\mathcal{W}_c(\Delta) + \mathcal{W}_c(\Delta)A(\Delta)^\top = -BB^\top.$$

The latter two solutions always exist since A is stable.

The proposed procedure for the solution of problem (19)-(18) is illustrated in Algorithm 1, where, for $i = 1, 2, \dots, N$, we denoted by

$$\Pi_i(\boldsymbol{\delta}) := \boldsymbol{\delta} + \frac{1}{\|S_{:i}\|^2}(\beta - S_{i:}^\top\boldsymbol{\delta})S_{:i},$$

the Euclidean projections onto the feasible sets defined by the constraint in (21).

Algorithm 1 Maximization of nn_{F} via PGA

Pick A s.t. $\alpha(A) \le \gamma$ *(initial adjacency matrix)*
Set $\varepsilon > 0$ *(stopping condition)*
Set $\eta > 0$ *(gradient ascent step-size)*
Set $\boldsymbol{\delta} \leftarrow \boldsymbol{\delta}_0$, $\boldsymbol{\delta}_{\text{prev}} \leftarrow \boldsymbol{\delta}_{0,\text{prev}}$ *(initialization)*
while $\|A(e^{\mathrm{diag}(\boldsymbol{\delta})}) - A(e^{\mathrm{diag}(\boldsymbol{\delta}_{\text{prev}})})\|_{\mathrm{F}} > \varepsilon$ **do**
    $\boldsymbol{\delta}_{\text{prev}} \leftarrow \boldsymbol{\delta}$
    $\delta_i \leftarrow \delta_i + \eta\,\frac{\partial f(A(e^{\mathrm{diag}(\boldsymbol{\delta})}))}{\partial \delta_i}$, $\;i = 2, \dots, n$
    for $i = 1, 2, \dots, N$ **do**
        if $S_{i:}^\top\boldsymbol{\delta} > \beta$ **then**
            $\boldsymbol{\delta} \leftarrow \Pi_i(\boldsymbol{\delta})$
        end if
    end for
end while

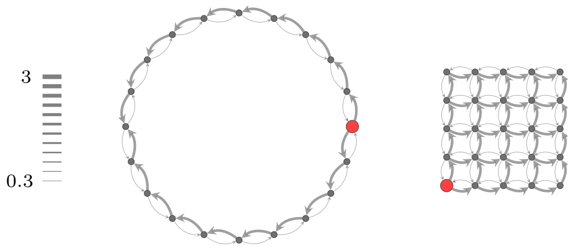
C. Simulations

We tested Algorithm 1 in some different structured and random scenarios. In Fig. 2 the outputs of the algorithm for three structured networks, namely line, cycle, and grid networks, are illustrated. Fig. 3 shows the results for two random network topologies: The Barabási–Albert network with attachment coefficient equal to one (left subplot) and a fixed-degree distribution random network composed of 20% nodes with degree 3 and 80% nodes with degree 2 (right subplot).[2] Observe, in particular, that the first choice yields acyclic random graphs, while the second class of random networks can possibly feature cycles. In all the simulations we picked a non-negative almost symmetric random initialization and we set $\mathcal{K} = \{1\}$ and $\mathcal{T} = \mathcal{V}$.

From the simulations, it can be noticed that the resulting optimized networks exhibit some preferred "anisotropic" directions of maximum length $d(\mathcal{K}, \mathcal{T})$, both in the structured and random case. The obtained numerical results seem therefore in agreement with the analytical results derived in Sec. III for the non-normality measure $\mathrm{nn}_2(A, B, C)$.

[2]We refer to [29] for the precise definitions and the details on the construction of these random networks.

(a) line network ($n = 20$)



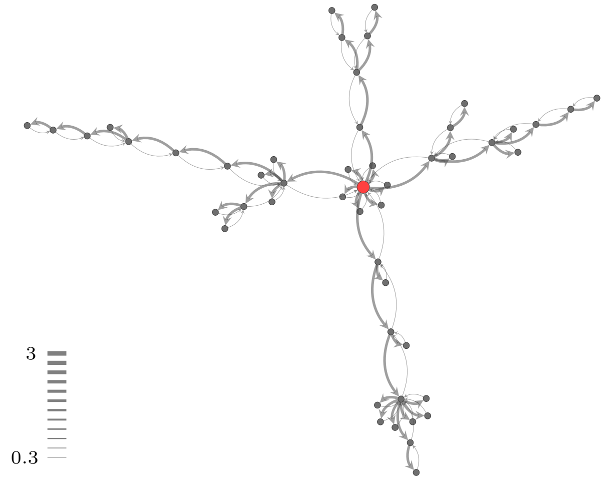
(b) cycle network ($n = 20$) (c) grid network ($n = 25$)

Fig. 2. Simulation results of Algorithm 1 for different structured network topologies, namely line, cycle, and grid networks. Here the thickness of an edge is proportional to the edge weight as indicated in the legend bar on the left. The initialization of A is chosen to be stable, non-negative and almost symmetric (namely, matrix with non-zero off-diagonal entries equal to one plus i.i.d. uniformly distributed noise in $[-0.1, 0.1]$). We used the following setup: $\beta = 1$, $\boldsymbol{\delta}_0 = 0$, $\eta = 0.2$, $\varepsilon = 10^{-6}$, $\mathcal{K} = \{1\}$ (bigger red node in the figures), and $\mathcal{T} = \mathcal{V}$. Self-loops are omitted for clarity.

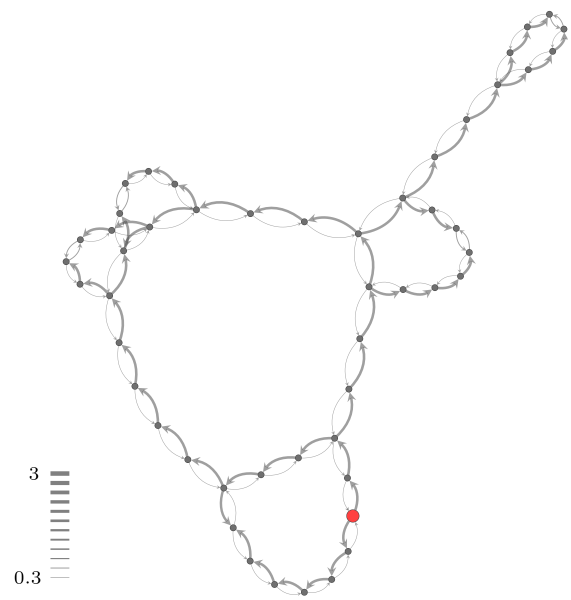
V. Concluding remarks

In this paper, we analyzed the relation between network non-normality and topological network structure. For the case of positive systems, we showed that the network diameter (or, more precisely, a generalization of the latter, called relative diameter) is a topological feature that is strongly connected with the non-normality degree of the dynamical network. This follows both from the upper and lower bounds derived in Sec. III for the measure $\mathrm{nn}_2(A, B, C)$ and from the numerical results illustrated in Sec. IV for the measure $\mathrm{nn}_{\mathrm{F}}(A, B, C)$.

There are many open questions that need to be addressed. An immediate one concerns the derivation of analytical bounds similar to those of Sec. III for the measure $\mathrm{nn}_{\mathrm{F}}(A, B, C)$, whereas a more challenging one concerns the extension of the ideas in this paper to general networks featuring both positive and negative weights.



(a) Barabási-Albert preferential attachment network ($n = 50$)



(b) Random network with degrees 2 and 3 ($n = 50$)

Fig. 3. Simulation results of Algorithm 1 for two different random network topologies, namely an acyclic Barabási–Albert preferential attachment network (attachment coefficient set to one) [29, Ch. 4] and a fixed-degree distribution random network composed of 20% nodes with degree 3 and 80% nodes with degree 2 [29, Ch. 3]. The meaning of arrows/nodes and the initialization of the algorithm are the same of the ones described in the caption of Fig. 2. Network visualization is performed via a force-directed layout.

References

[1] L. N. Trefethen and M. Embree, *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*. Princeton University Press, 2005.

[2] S. Ganguli, D. Huh, and H. Sompolinsky, "Memory traces in dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 105, no. 48, pp. 18 970–18 975, 2008.

[3] S. Ganguli and P. Latham, "Feedforward to the past: The relation between neuronal connectivity, amplification, and short-term memory," *Neuron*, vol. 61, no. 4, pp. 499–501, 2009.

[4] M. S. Goldman, "Memory without feedback in a neural network," *Neuron*, vol. 61, no. 4, pp. 621–634, 2009.

[5] G. Hennequin, T. P. Vogels, and W. Gerstner, "Optimal control of transient dynamics in balanced networks supports generation of complex movements," *Neuron*, vol. 82, no. 6, pp. 1394–1406, 2014.

[6] B. K. Murphy and K. D. Miller, "Balanced amplification: a new mechanism of selective amplification of neural activity patterns," *Neuron*, vol. 61, no. 4, pp. 635–648, 2009.

[7] G. Hennequin, T. P. Vogels, and W. Gerstner, "Non-normal amplification in random balanced neuronal networks," *Physical Review E*, vol. 86, no. 1, p. 011909, 2012.

[8] D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, "The network origins of aggregate fluctuations," *Econometrica*, vol. 80, no. 5, pp. 1977–2016, 2012.

[9] Q. Huang, Y. Yuan, J. Goncalves, and M. A. Dahleh, "$\mathcal{H}_2$ norm based network volatility measures," in *2014 American Control Conference (ACC)*. IEEE, 2014, pp. 3310–3315.

[10] T. Sarkar, M. Roozbehani, and M. A. Dahleh, "Robustness scaling in large networks," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 197–202.

[11] F. Pasqualetti and S. Zampieri, "On the controllability of isotropic and anisotropic networks," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 607–612.

[12] F. Pasqualetti, S. Zampieri, and F. Bullo, "Controllability metrics, limitations and algorithms for complex networks," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 40–52, 2014.

[13] G. Yan, G. Tsekenis, B. Barzel, J.-J. Slotine, Y.-Y. Liu, and A.-L. Barabási, "Spectrum of controlling and observing complex networks," *Nature Physics*, vol. 11, no. 9, pp. 779–786, 2015.

[14] A. Olshevsky, "Eigenvalue clustering, control energy, and logarithmic capacity," *Systems & Control Letters*, vol. 96, pp. 45–50, 2016.

[15] N. Bof, G. Baggio, and S. Zampieri, "On the role of network centrality in the controllability of complex networks," *IEEE Transactions on Control of Network Systems*, vol. 4, pp. 643–653, 2017.

[16] G. Bianchin, F. Pasqualetti, and S. Zampieri, "The role of diameter in the controllability of complex networks," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 980–985.

[17] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[18] H. Minc, *Nonnegative matrices*. New York; Chichester: John Wiley & Sons, 1988.

[19] K. Zhou, J. C. Doyle, and K. Glover, *Robust and optimal control*. Prentice Hall New Jersey, 1996, vol. 40.

[20] T. H. Summers, F. L. Cortesi, and J. Lygeros, "On submodularity and controllability in complex dynamical networks," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 1, pp. 91–101, 2016.

[21] D. Hinrichsen and E. Plischke, "Robust stability and transient behaviour of positive linear systems," *Vietnam Journal of Mathematics*, vol. 35, no. 4, pp. 429–462, 2007.

[22] D. Hinrichsen and A. J. Pritchard, *Mathematical systems theory I: modelling, state space analysis, stability and robustness*. Springer Berlin, 2005, vol. 48.

[23] A. M. Ostrowski, "On the eigenvector belonging to the maximal root of a non-negative matrix," *Proceedings of the Edinburgh Mathematical Society (Series 2)*, vol. 12, no. 02, pp. 107–112, 1960.

[24] S. M. Cioaba and D. A. Gregory, "Principal eigenvectors of irregular graphs," *Electronic Journal of Linear Algebra*, vol. 16, no. 1, p. 31, 2007.

[25] E. Plischke, "Transient effects of linear dynamical systems," Ph.D. dissertation, Universität Bremen, 2005.

[26] E. K. Chong and S. H. Zak, *An introduction to optimization*. John Wiley & Sons, 2013, vol. 76.

[27] K. B. Petersen and M. S. Pedersen, *The matrix cookbook*. Technical University of Denmark, 2008.

[28] J. Vanbiervliet, B. Vandereycken, W. Michiels, S. Vandewalle, and M. Diehl, "The smoothed spectral abscissa for robust stability optimization," *SIAM Journal on Optimization*, vol. 20, no. 1, pp. 156–171, 2009.

[29] R. Durrett, *Random graph dynamics*. Cambridge university press Cambridge, 2007.