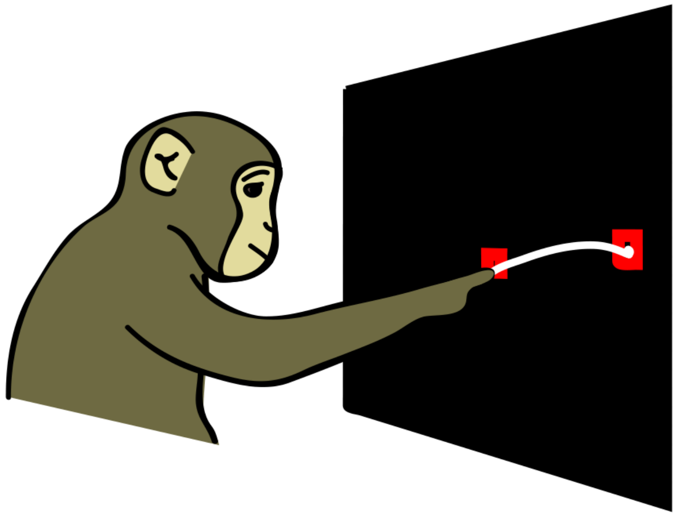
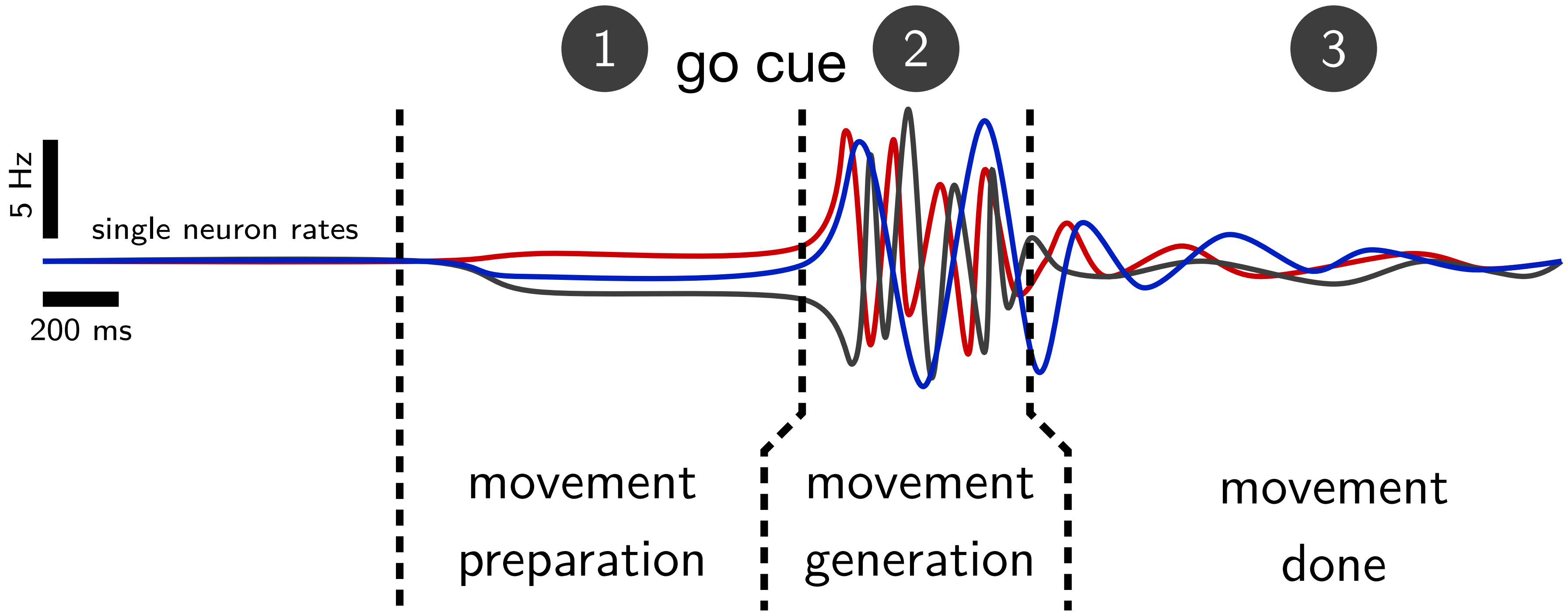
Non-normal dynamics and the efficient information propagation in linear networks

Giacomo Baggio
Virginia Rutten
Guillaume Hennequin
Sandro Zampieri

Motor cortex dynamics: from experiments to dynamical models

[Churchland et al., Nature (2012)]





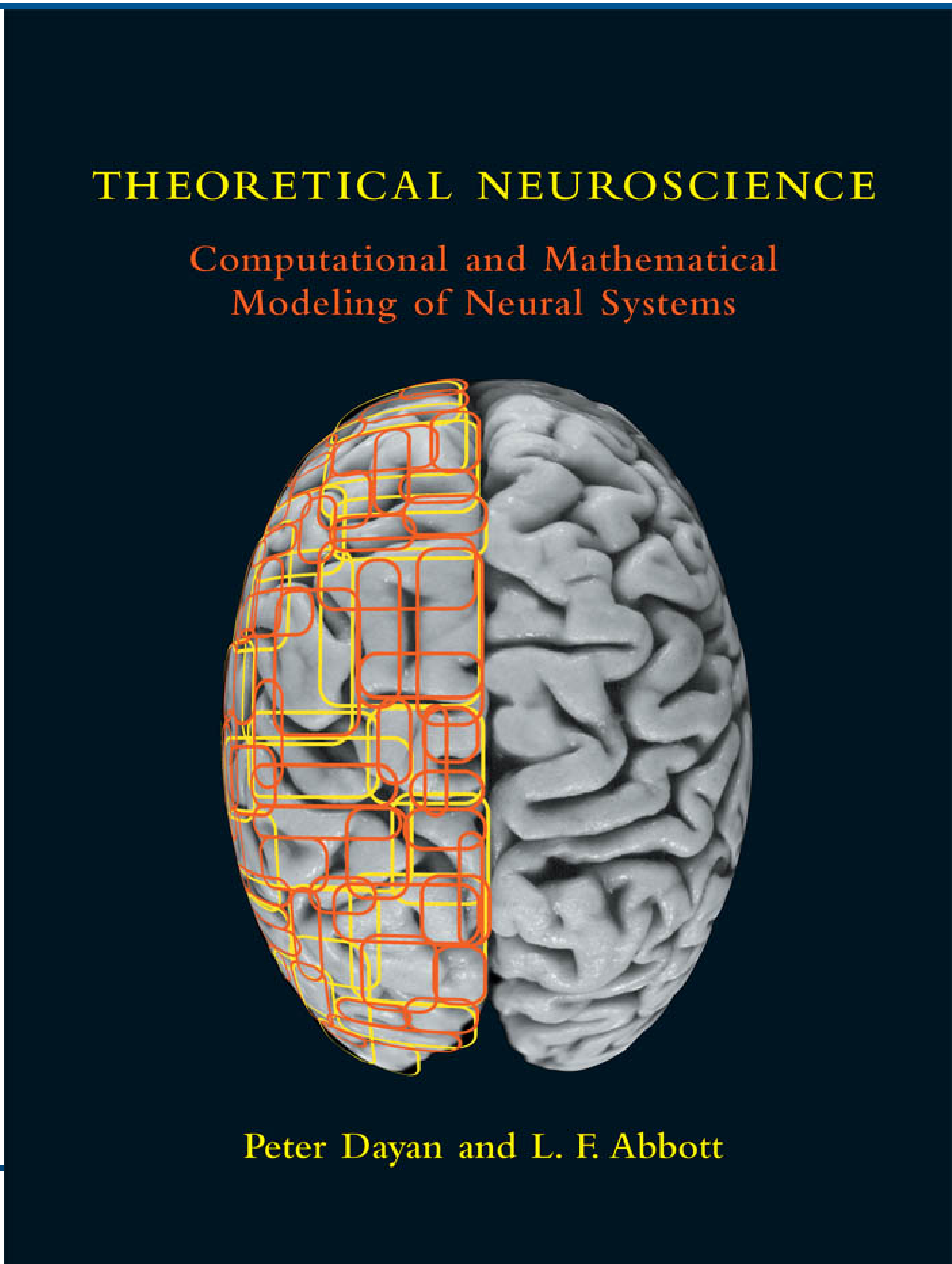
The signal generated is characterized by a “rich” transient and by asymptotic stability target

Introduction to neuronal networks

Peter Dayan and L.F. Abbott
"Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems" The MIT Press, 2001

This is based on a **linearized model** of the interaction dynamics between neurons

The signals are the **spiking rates**



A model for information transmission

The question is how populations of neurons produce **large amplitude transient signals** for information transmission and storage.

In various papers it is emphasized the role the system **non-normality** for the generation of the "best" signals.

1. S. Ganguli, D. Huh, and H. Sompolinsky "Memory traces in dynamical systems." Proceedings of the National Academy of Sciences, vol. 105.48, pp. 18970-18975, 2008.S.
2. Ganguli, and P. Latham, "Feedforward to the Past: The Relation between Neuronal Connectivity, Amplification, and Short-Term Memory", Neuron, Vol. 41, pp. 499-501, 2009.
3. M. S. Goldman, "Memory without Feedback in a Neural Network", Neuron, Vol. 61, pp. 621–634, 2009.
4. G. Hennequin, T. P. Vogels, and W. Gerstner, "Non-normal amplification in random balanced neuronal networks", Physical Review E, Vol. 86, pp. 011909, 2012.
5. G. Hennequin, et al. "Optimal control of transient dynamics in balanced networks supports generation of complex movements." Neuron, 82.6 (2014): 1394-1406.

The key point is to have large transients while keeping the dynamics stable

A model for information transmission

Problems:

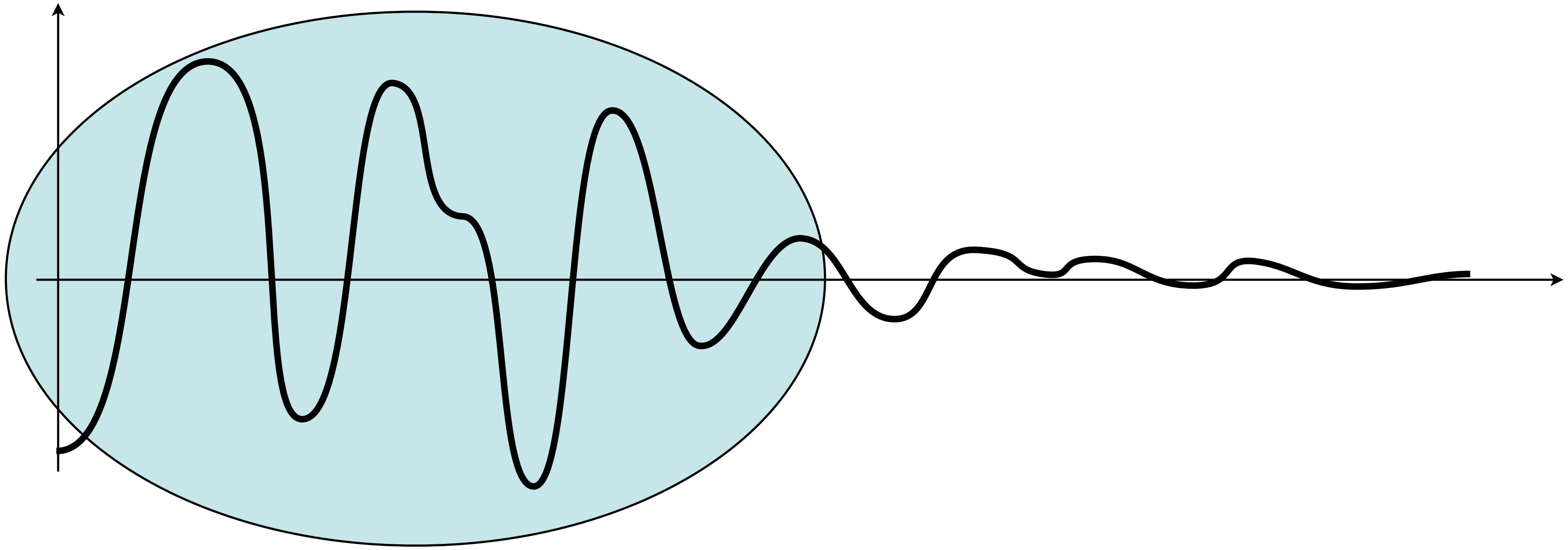
1. Propose a **consistent model** for understanding why non-normality plays a role in making information transmission more efficient.

2. **Quantify** the information transmission efficiency.

3. Verify whether the **model complexity** influences information transmission efficiency.

Generation of "rich" signals

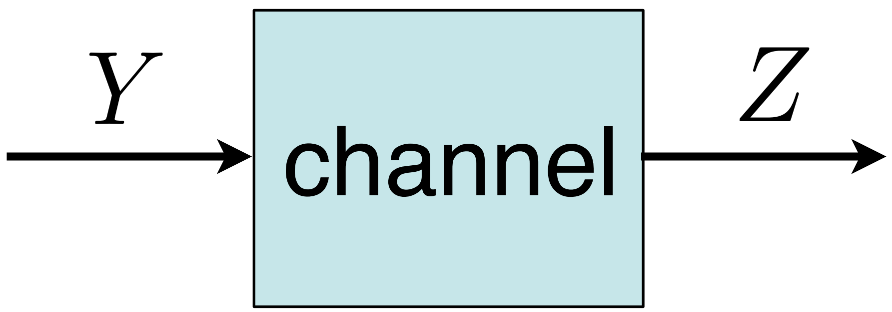
It is important to produce transients that have high energy initially and then converge to zero fast



Channel capacity

Channel model: conditional probability

$$p(Z|Y)$$

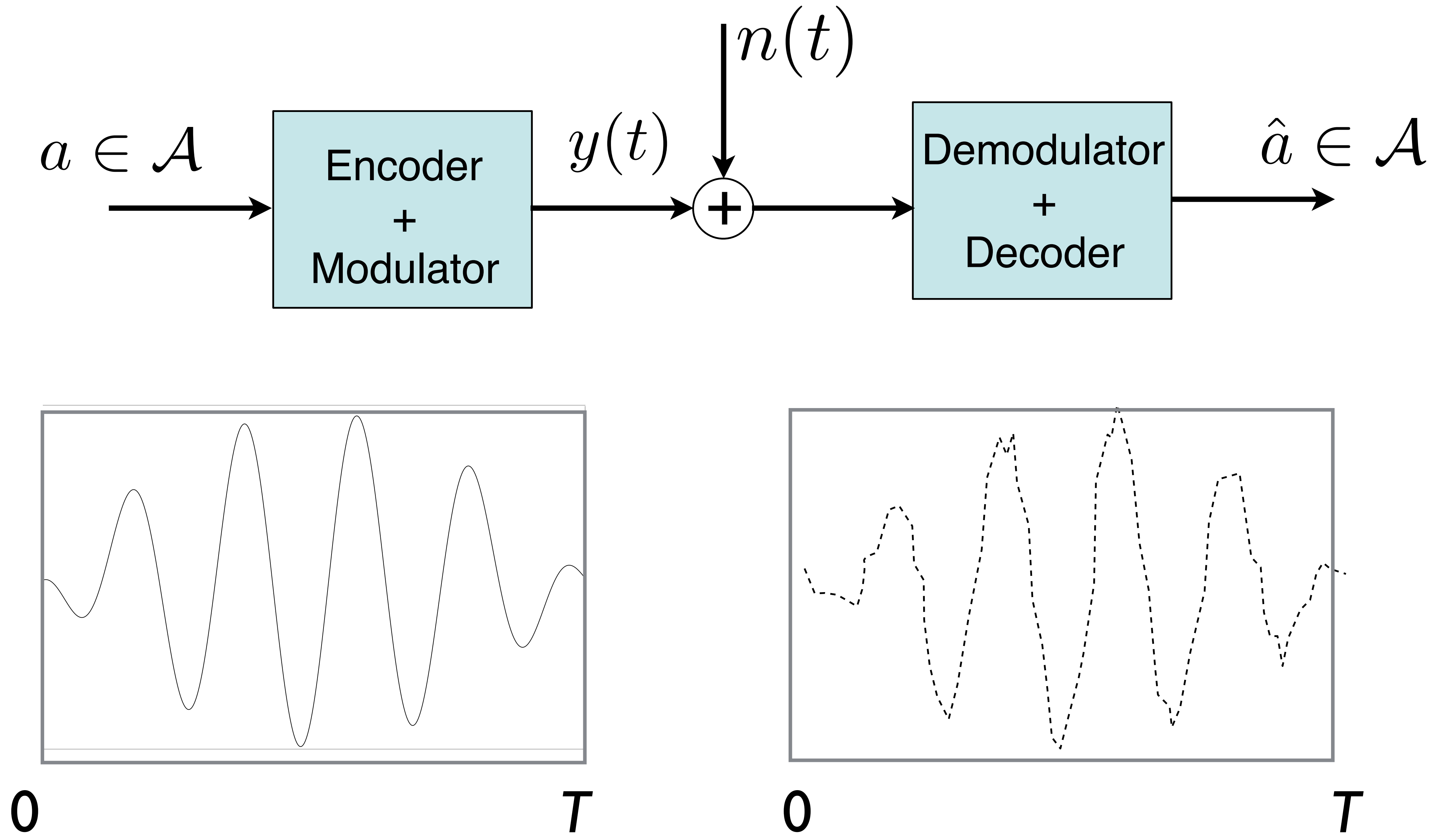


Shannon channel capacity

$$C = \max_{p(Y)} I(Y;Z) = \max_{p(Y)} h(Z) - h(Z|Y)$$

where $I(\cdot)$ is the mutual information and $h(\cdot)$ is the differential entropy

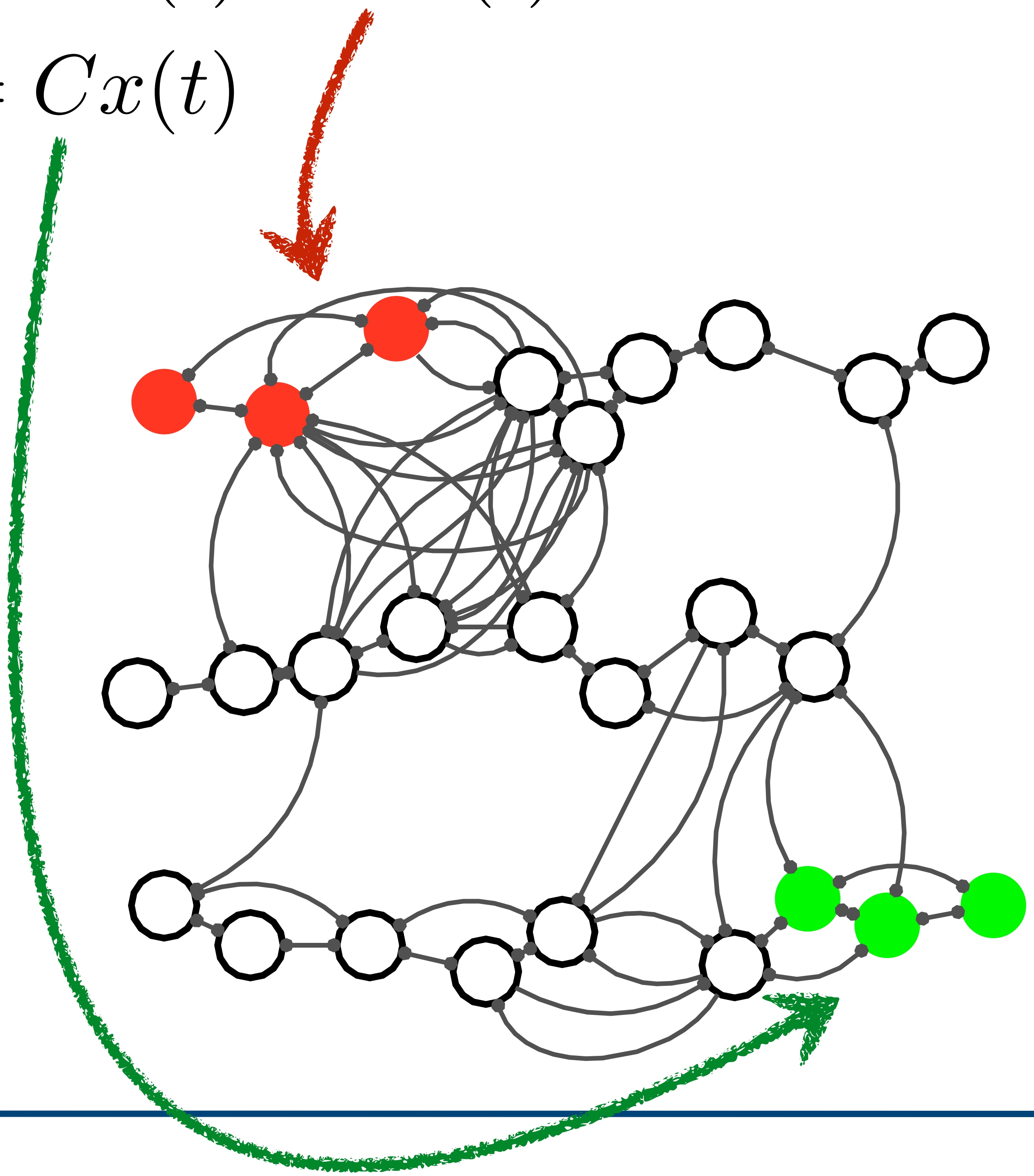
Continuous time channels



The capacity depends on the signal to noise ratio SNR

Modulation by a system transient

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t)$$



$a \in \mathcal{A}$

symbol

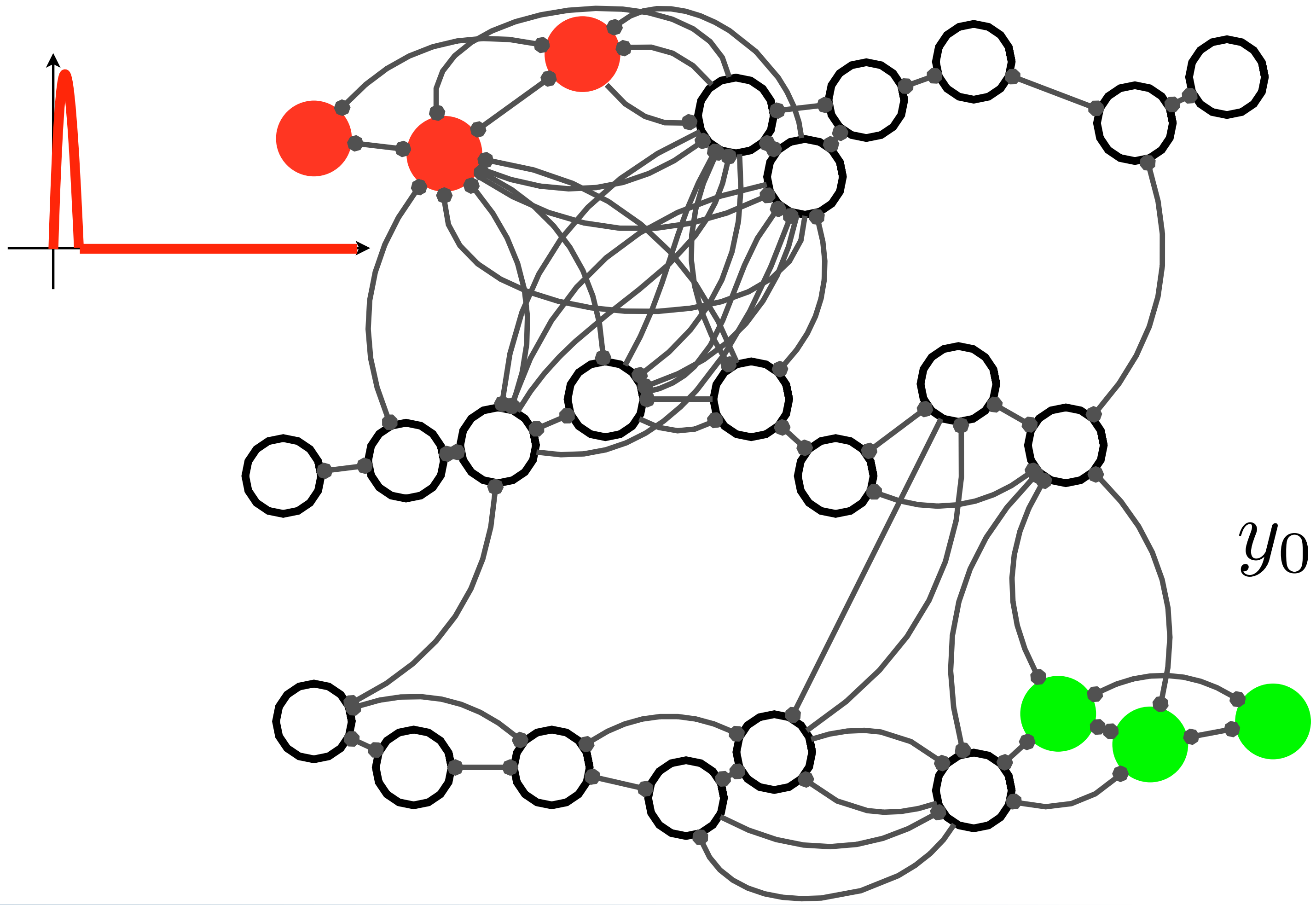
$$u(t) = u_0 \delta(t)$$

$$\|u_0\| \leq P$$

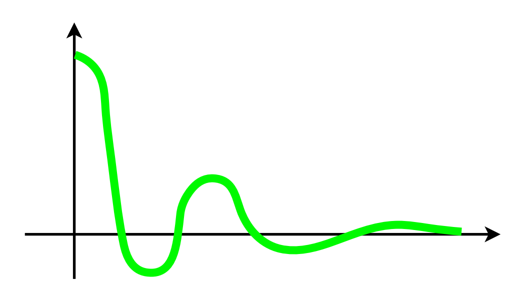
limited power

$$\dot{x}(t) = Ax(t) + Bu(t)$$

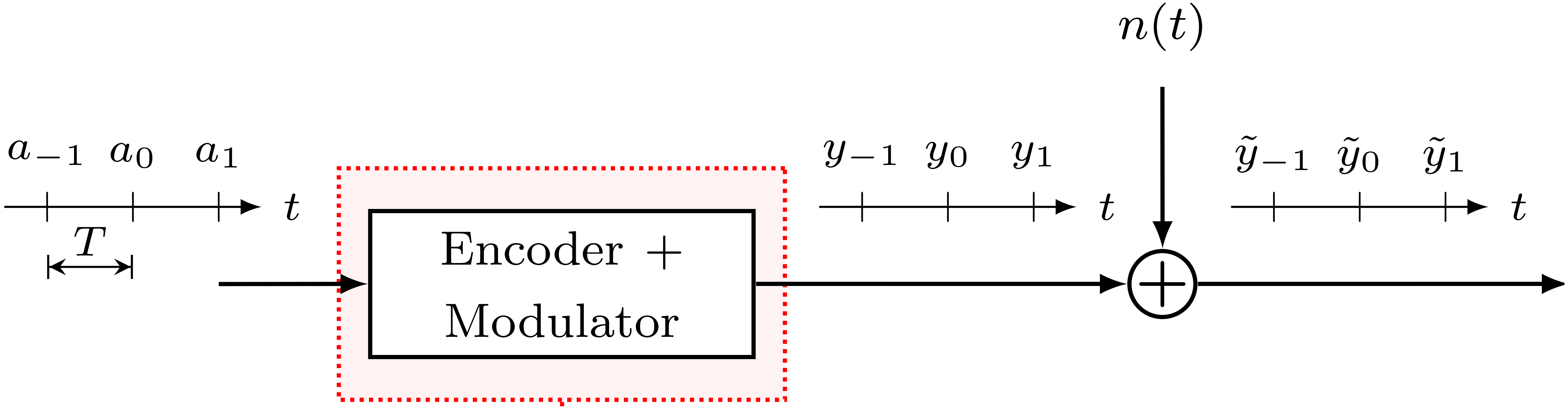
$$y(t) = Cx(t)$$



$$y_0(t) = Ce^{At}Bu_0$$

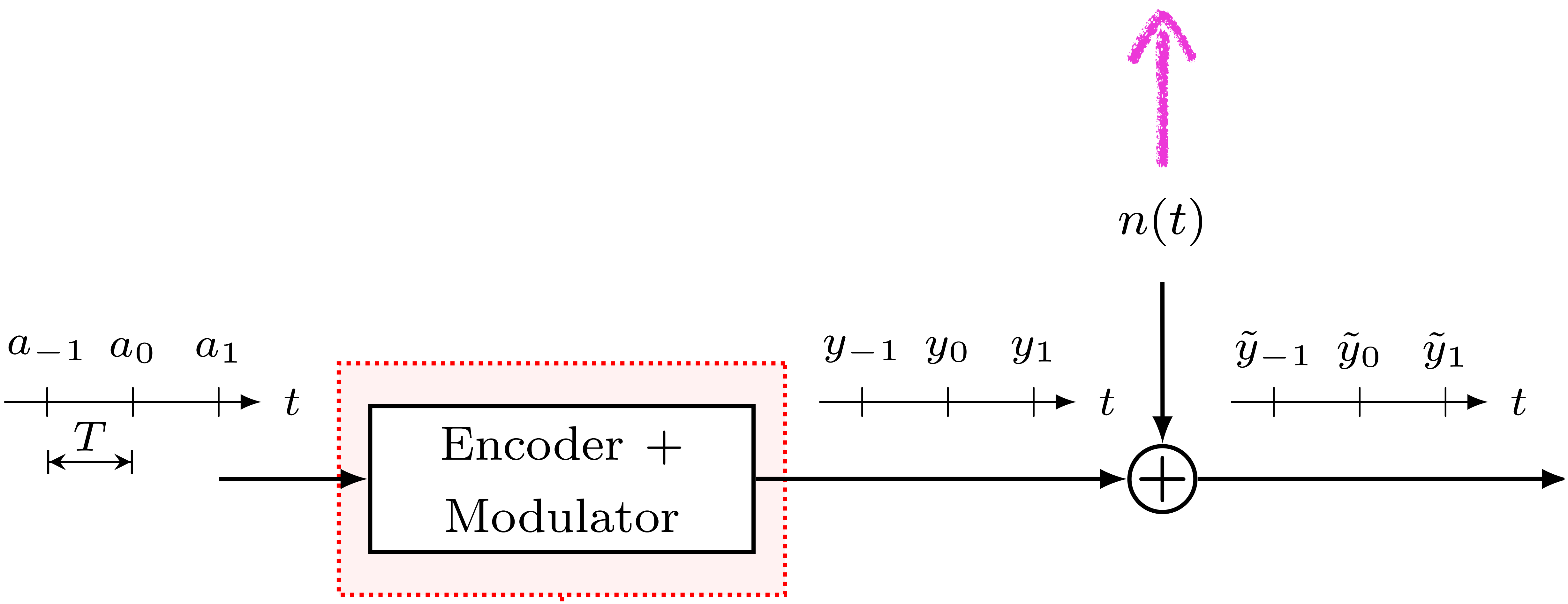


Modulation by a system transient



Modulation by a system transient

σ^2 is the noise power



Modulation by a system transient

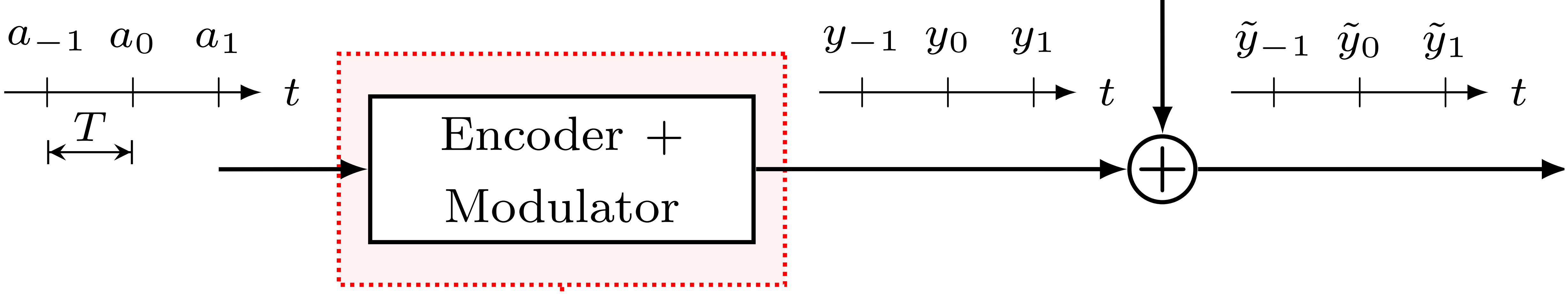
$SNR = P/\sigma^2$
signal to noise ratio

σ^2 is the noise power

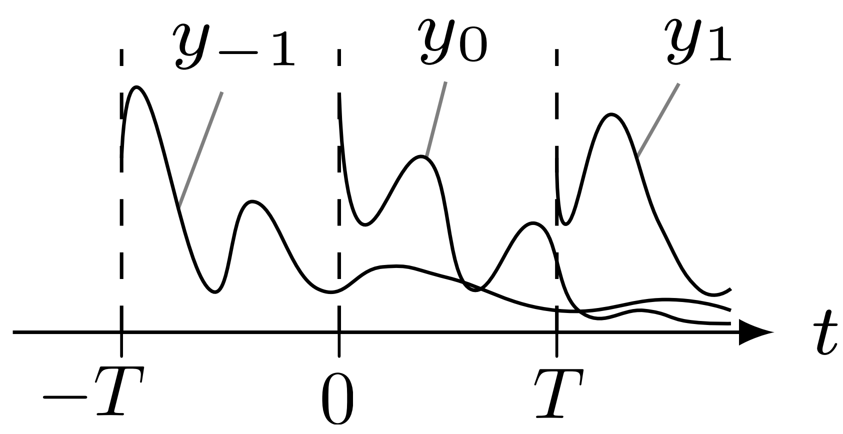
$$u(t) = u_0 \delta(t)$$

$$\|u_0\| \leq P$$

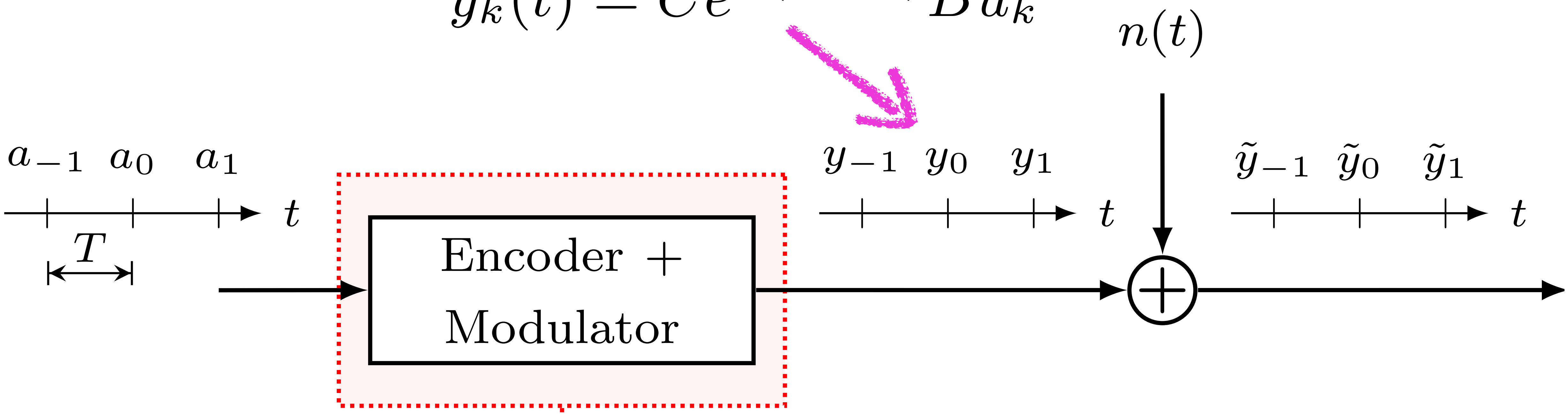
$n(t)$



Modulation by a system transient

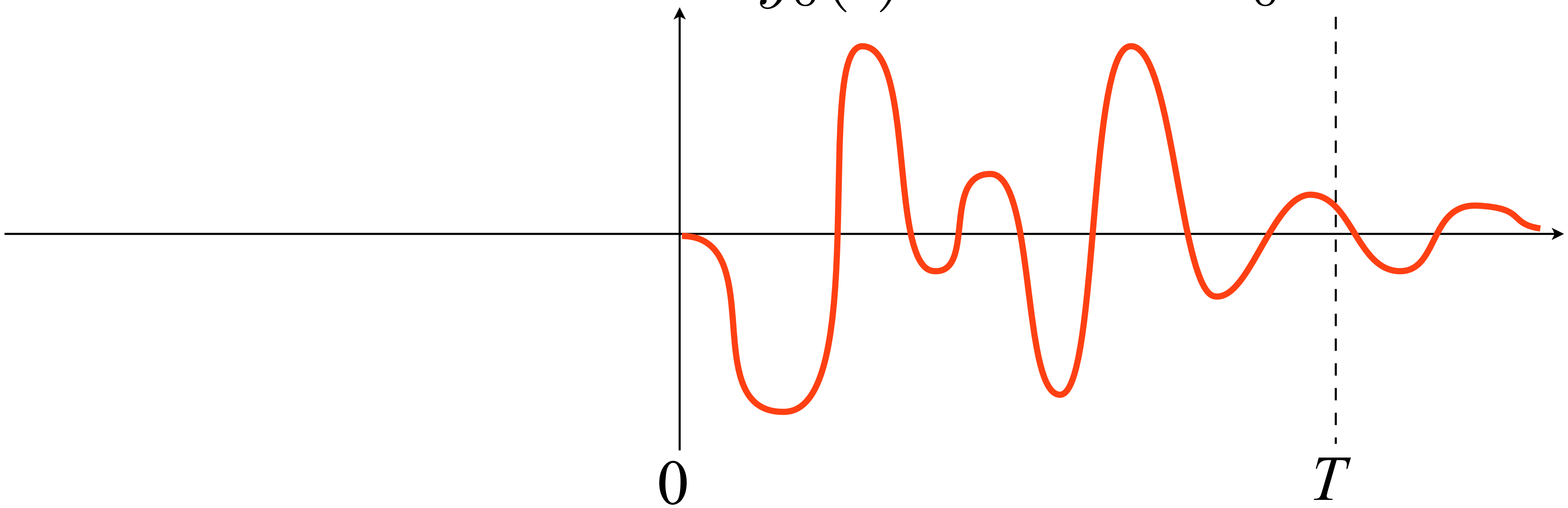


$$y_k(t) = Ce^{A(t-kT)}Bu_k$$

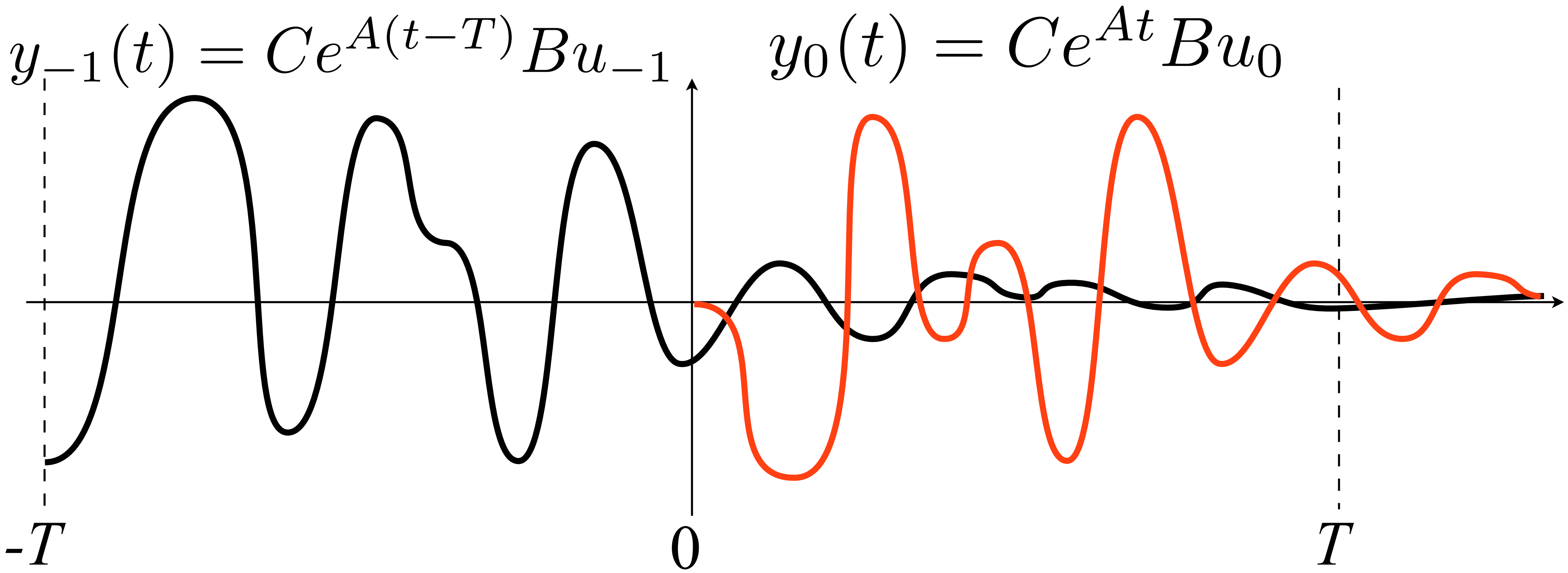


Inter-symbol interference

$$y_0(t) = Ce^{At}Bu_0$$



Inter-symbol interference



Inter-symbol interference: The signal transmitted at time -T, -2T, -3T, ... interfere with the signal transmitted at time 0 and acts as an additional noise.

Expression of the capacity

$$R_T = \frac{1}{2T} \max_{\Sigma \geq 0, \mathrm{tr}\Sigma \leq P} \log_2 \frac{\det\left(\sigma^2 I + \mathbf{O}\mathbf{W}\right)}{\det\left(\sigma^2 I + \mathbf{O}(\mathbf{W} - B\Sigma B^\top)\right)},$$

where

$$\mathbf{O} := \int_0^T e^{A^\top t} C^\top C e^{At}\, dt$$

denotes the $[0, T]$-observability Gramian of the pair (A, C) and

$$\mathbf{W} := \sum_{k=0}^{\infty} e^{AkT} B\Sigma B^\top e^{A^\top kT}$$

denotes the discrete-time controllability Gramian of the pair $(e^{AT}, B\Sigma^{1/2})$.

Expression of the capacity

$$R_T = \frac{1}{2T} \max_{\Sigma \geq 0, \mathrm{tr}\Sigma \leq P} \log_2 \frac{\det\left(\sigma^2 I + \mathbf{O}\mathbf{W}\right)}{\det\left(\sigma^2 I + \mathbf{O}(\mathbf{W} - B\Sigma B^\top)\right)},$$

where

$$\mathbf{O} := \int_0^T e^{A^\top t} C^\top C e^{At}\, dt$$

denotes the $[0, T]$-observability Gramian of the pair (A, C) and

$$\mathbf{W} := \sum_{k=0}^{\infty} e^{AkT} B\Sigma B^\top e^{A^\top kT}$$

denotes the discrete-time controllability Gramian of the pair $(e^{AT}, B\Sigma^{1/2})$.

$$R_{\max} := \max_{T \geq 0} R_T$$

Expression of the capacity

$$R_T = \frac{1}{2T} \max_{\Sigma \geq 0, \mathrm{tr}\Sigma \leq P} \log_2 \frac{\det\left(\sigma^2 I + \mathbf{O}\mathbf{W}\right)}{\det\left(\sigma^2 I + \mathbf{O}(\mathbf{W} - B\Sigma B^\top)\right)},$$

where

$$\mathbf{O} := \int_0^T e^{A^\top t} C^\top C e^{At}\, dt$$

denotes the $[0, T]$-observability Gramian of the pair (A, C) and

$$\mathbf{W} := \sum_{k=0}^{\infty} e^{AkT} B\Sigma B^\top e^{A^\top kT}$$

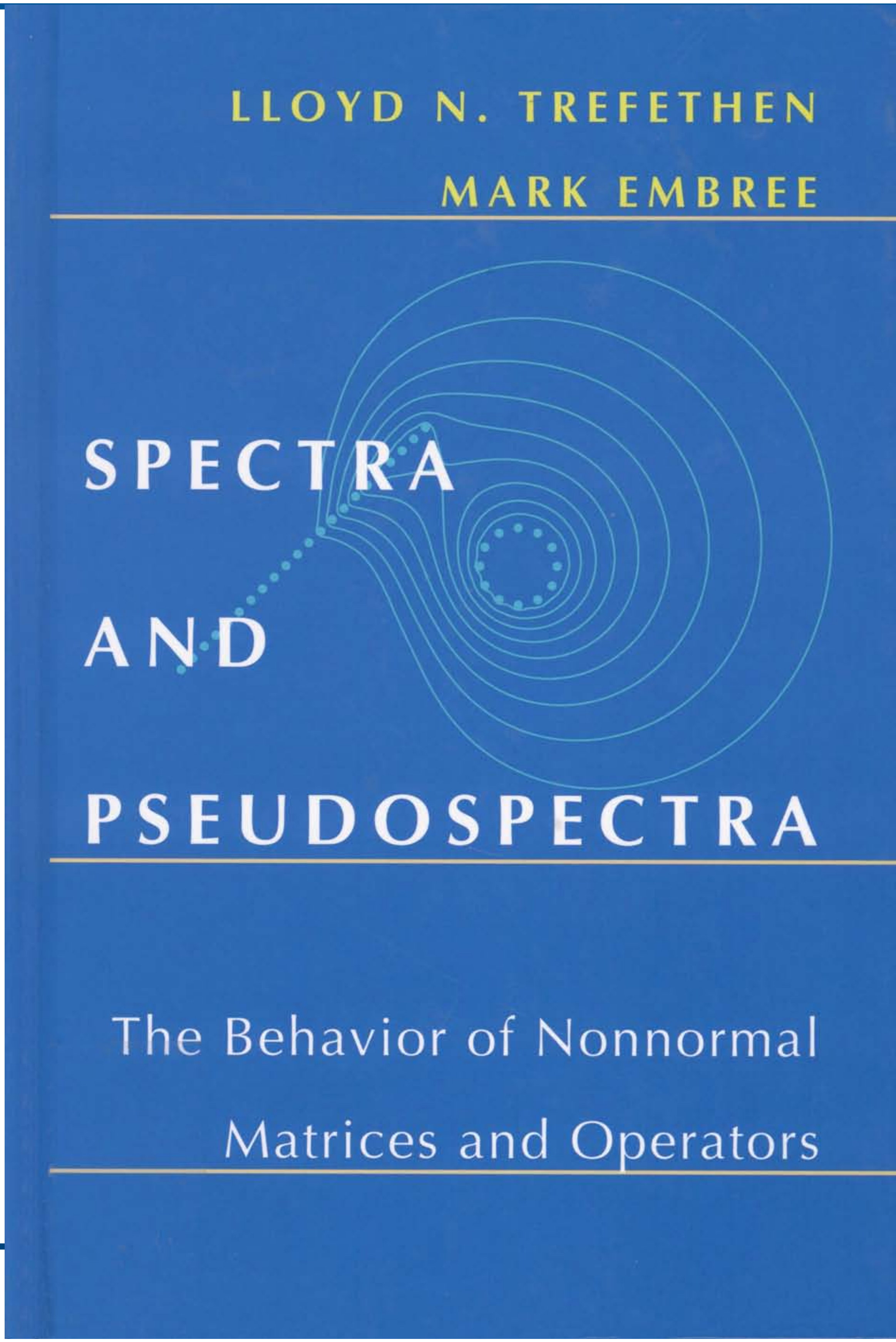
denotes the discrete-time controllability Gramian of the pair $(e^{AT}, B\Sigma^{1/2})$.

$$R_{\max} := \max_{T \geq 0} R_T$$

Depends on $SNR = P/\sigma^2$

Matrix non-normality

A matrix is normal iff $AA^T = A^T A$



The normal case

For normal networks R_T is decreasing in T and so the smaller is T the better the performance we obtain.

Corollary
If $\mathcal{V}_{in} = \mathcal{V}_{out} = \mathcal{V}$ and $A \in \mathbb{R}^{n\times n}$ is a normal and stable matrix, then

$$R_{\max} := \max_{T\geq 0} R_T = R_0 = \frac{1}{\ln 2}\frac{-\mathrm{tr}(A)\,SNR}{SNR - 2\mathrm{tr}(A)}.$$

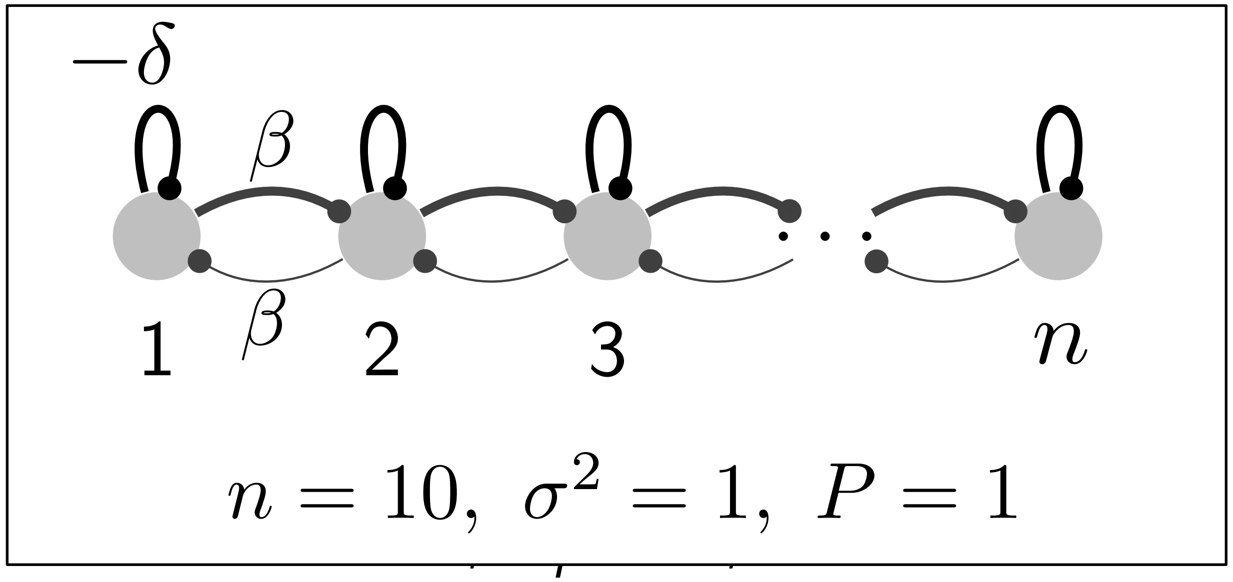
The normal case

For normal networks R_T is decreasing in T and so the smaller is T the better the performance we obtain.

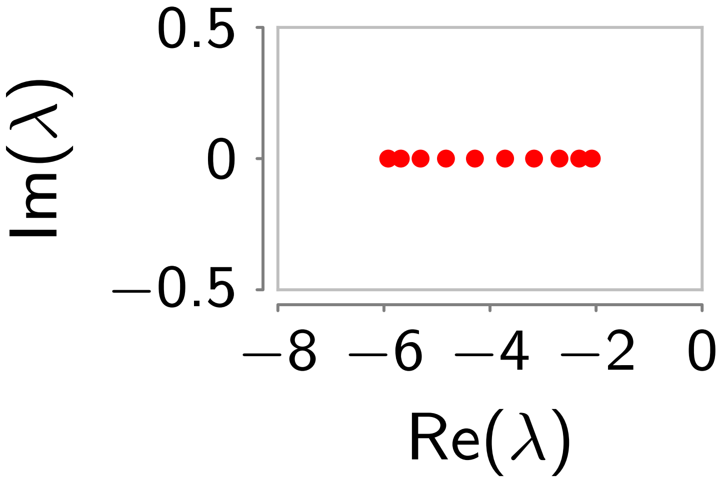
Corollary
If $\mathcal{V}_{in} = \mathcal{V}_{out} = \mathcal{V}$ and $A \in \mathbb{R}^{n \times n}$ is a normal and stable matrix, then

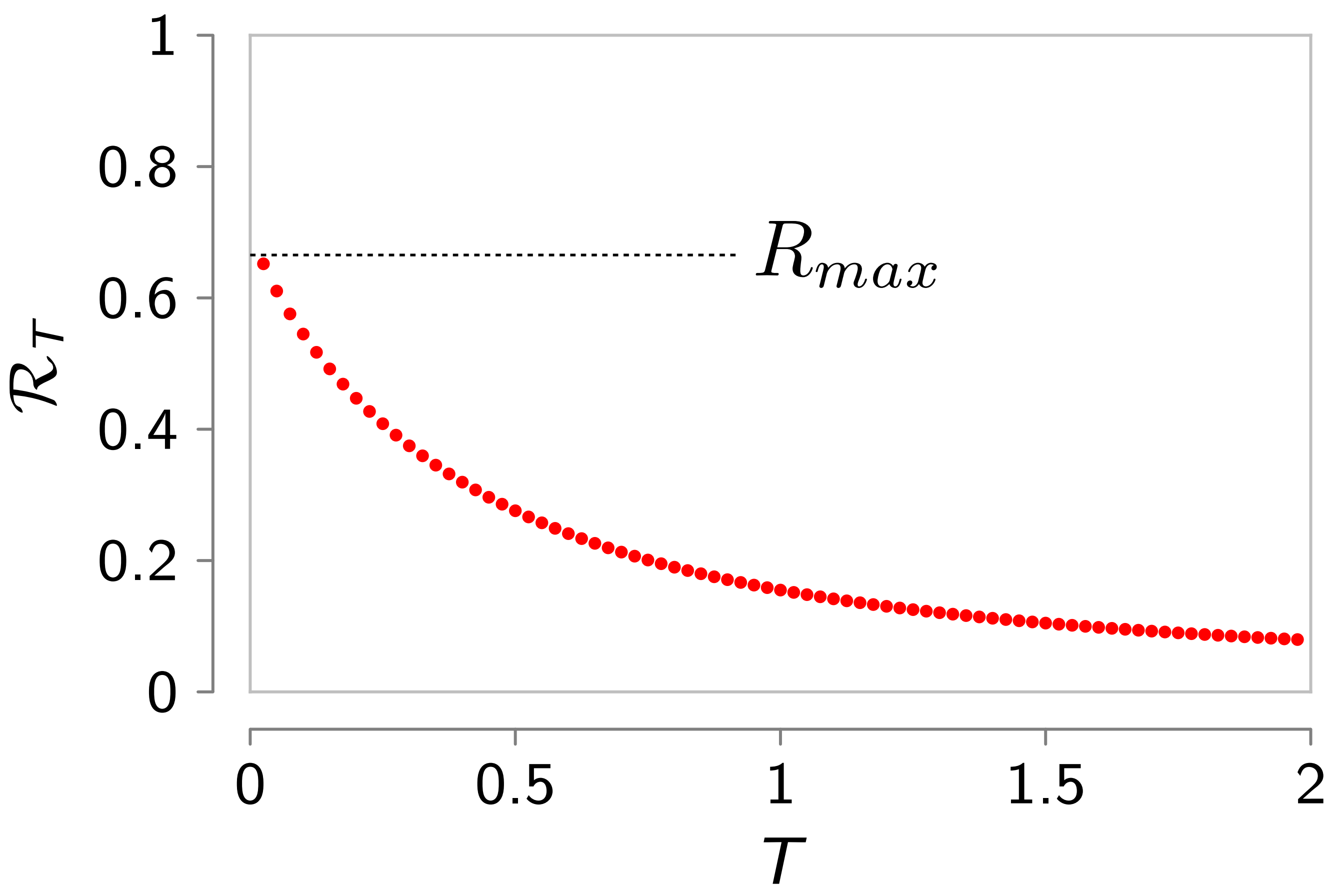
$$R_{\max} := \max_{T \geq 0} R_T = R_0 = \frac{1}{\ln 2} \frac{-\mathrm{tr}(A)\, SNR}{SNR - 2\mathrm{tr}(A)}.$$

Example

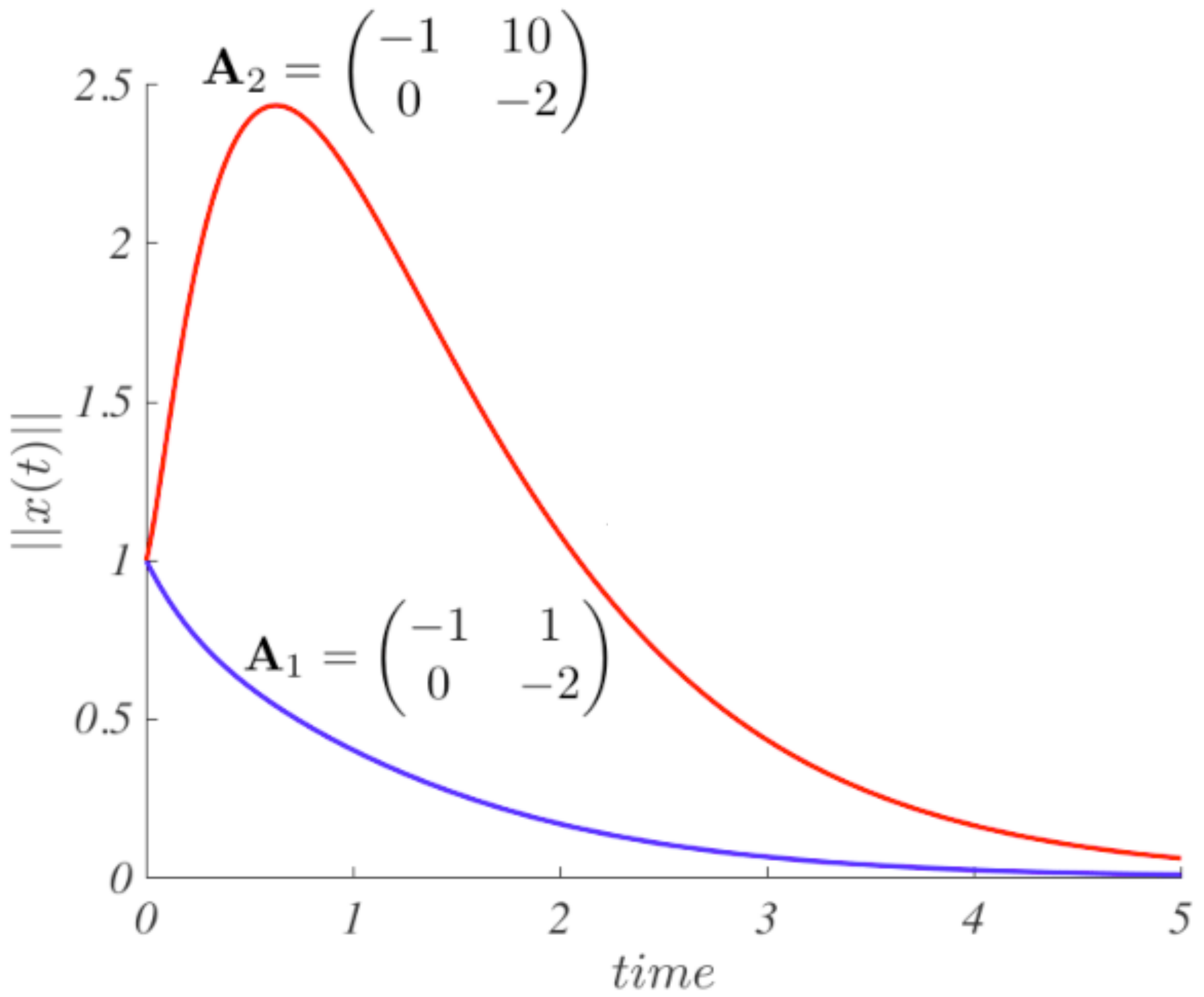


$\delta = 5, \ \beta = 1$





Matrix non-normality



Matrix non-normality

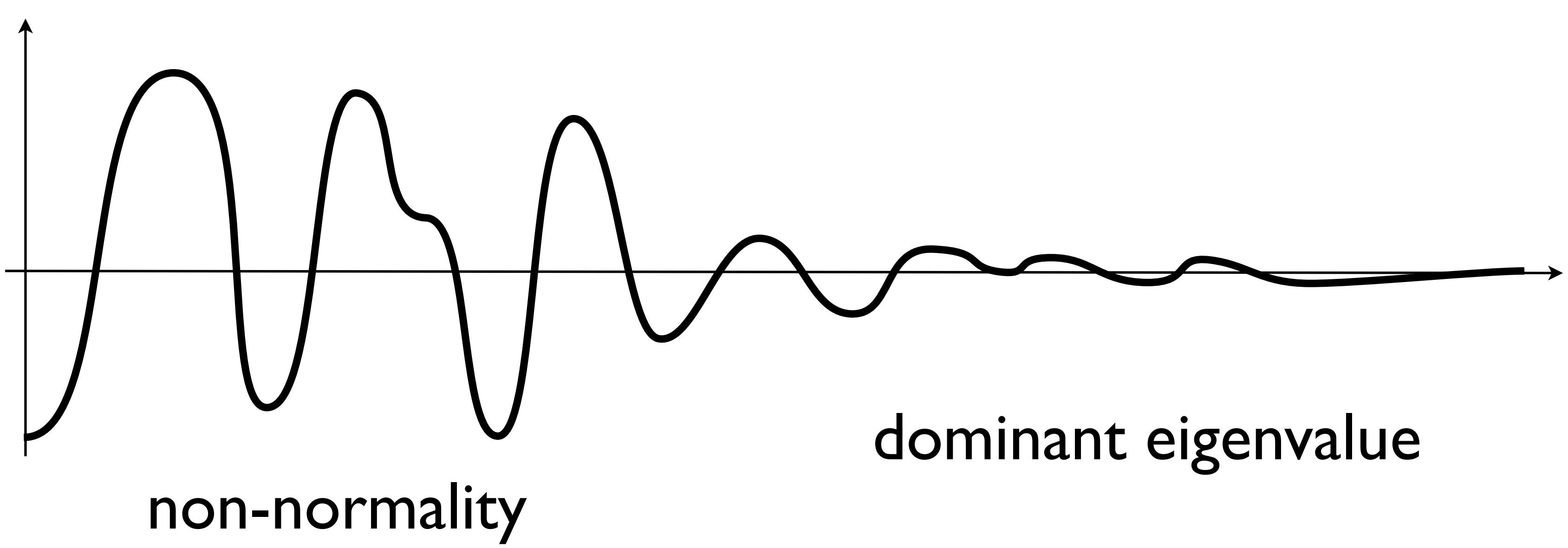
Non-normality has two important features:

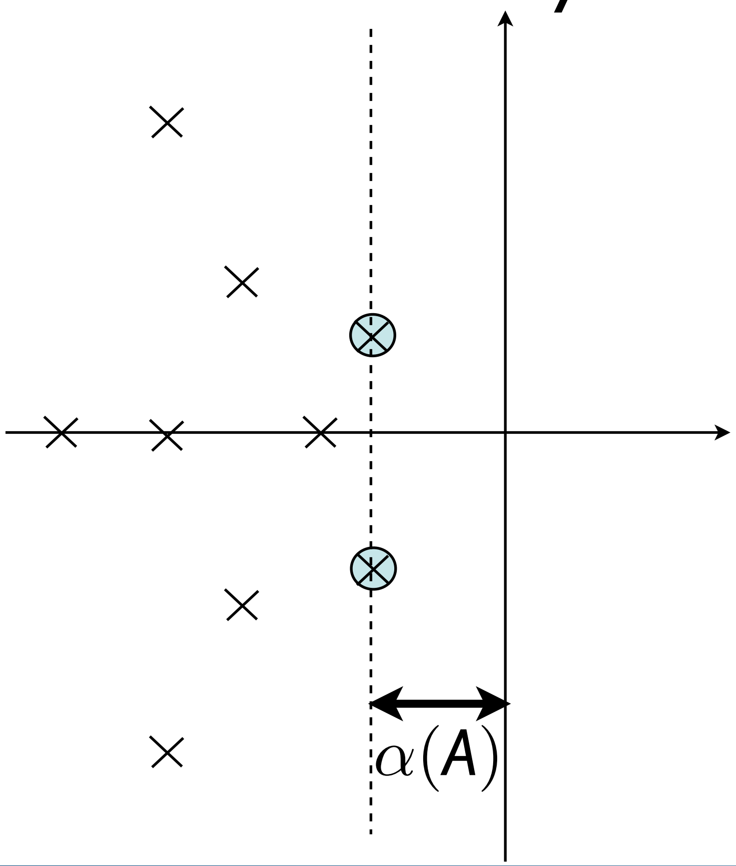
1. The eigenvalues of a highly non-normal matrix A are very sensitive to matrix entries variations.

2. The exponential

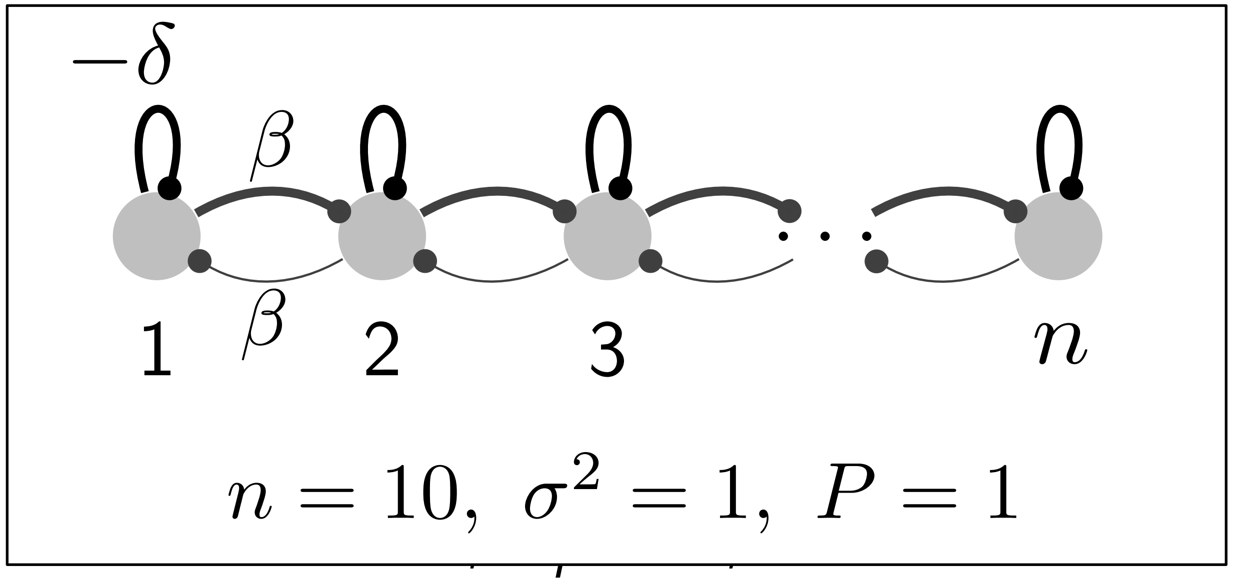
$$e^{At}$$

of a stable highly non-normal matrix A is large for small t and then decays to zero according to the spectral abscissa $\alpha(A)$.

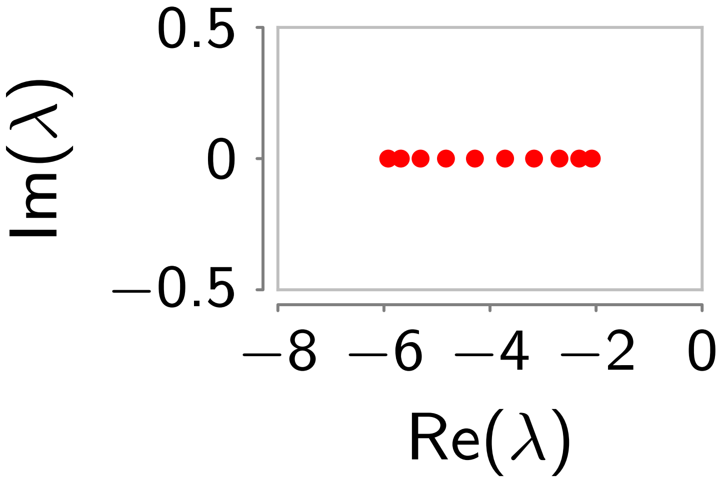


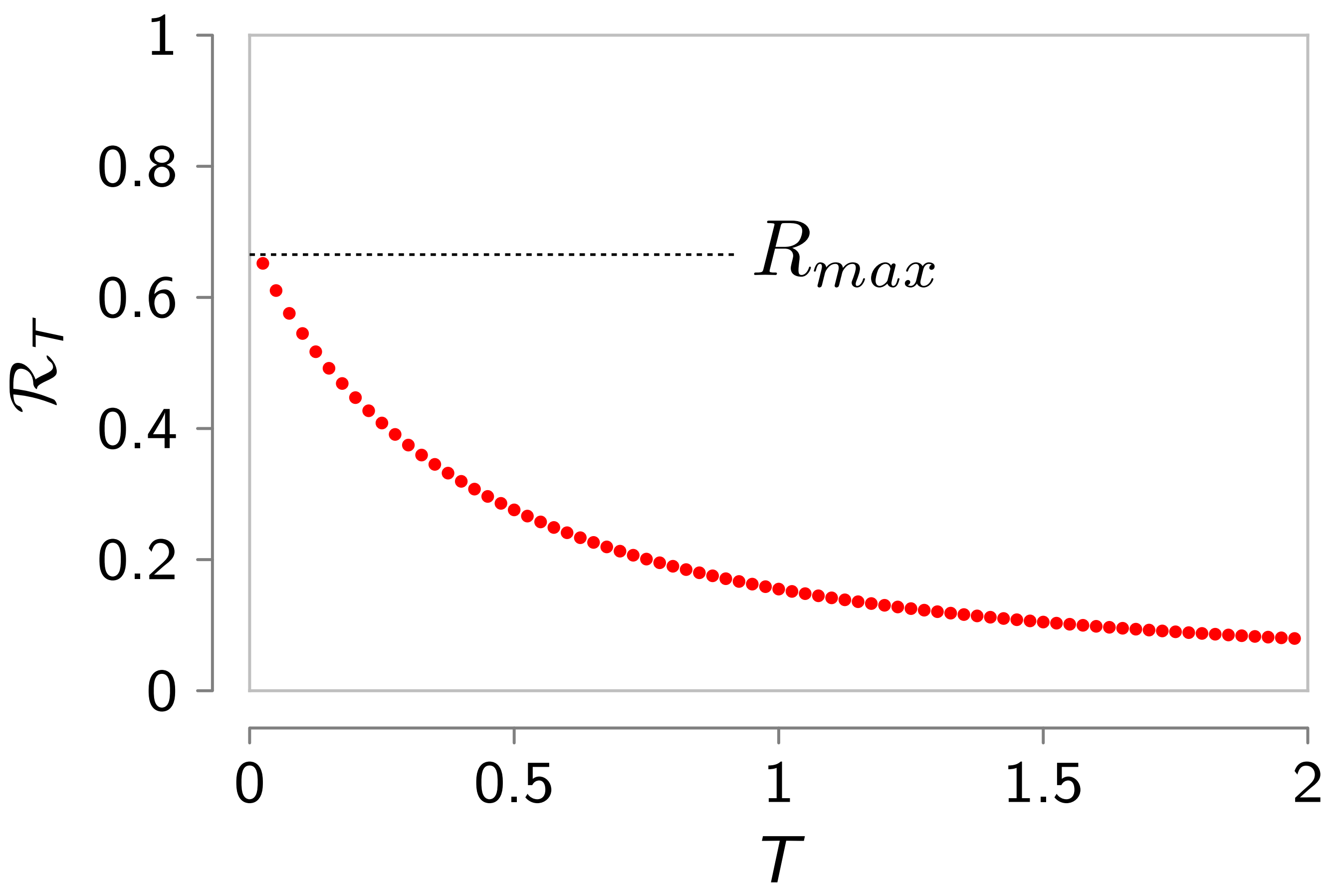


Example

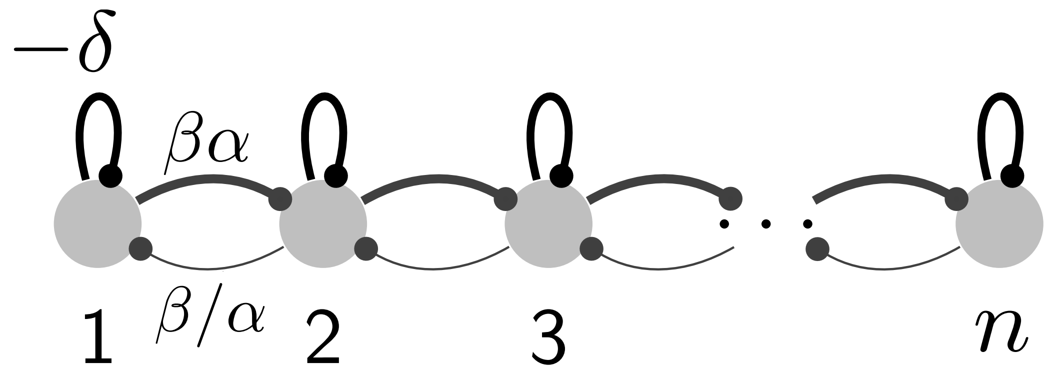


$\delta = 5,\ \beta = 1$



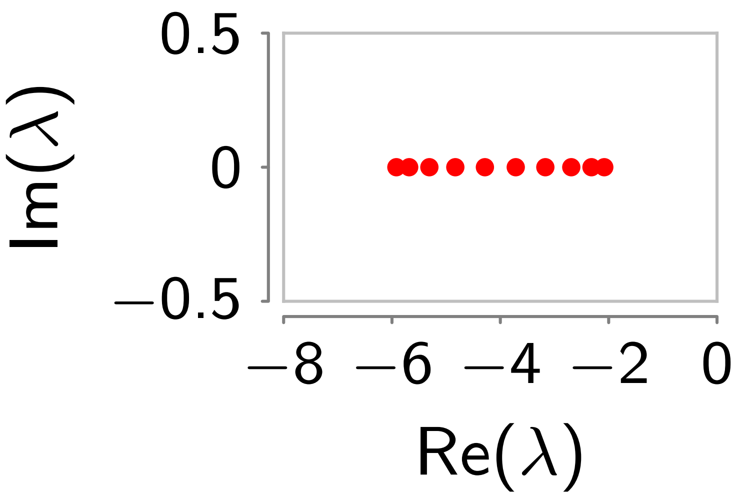


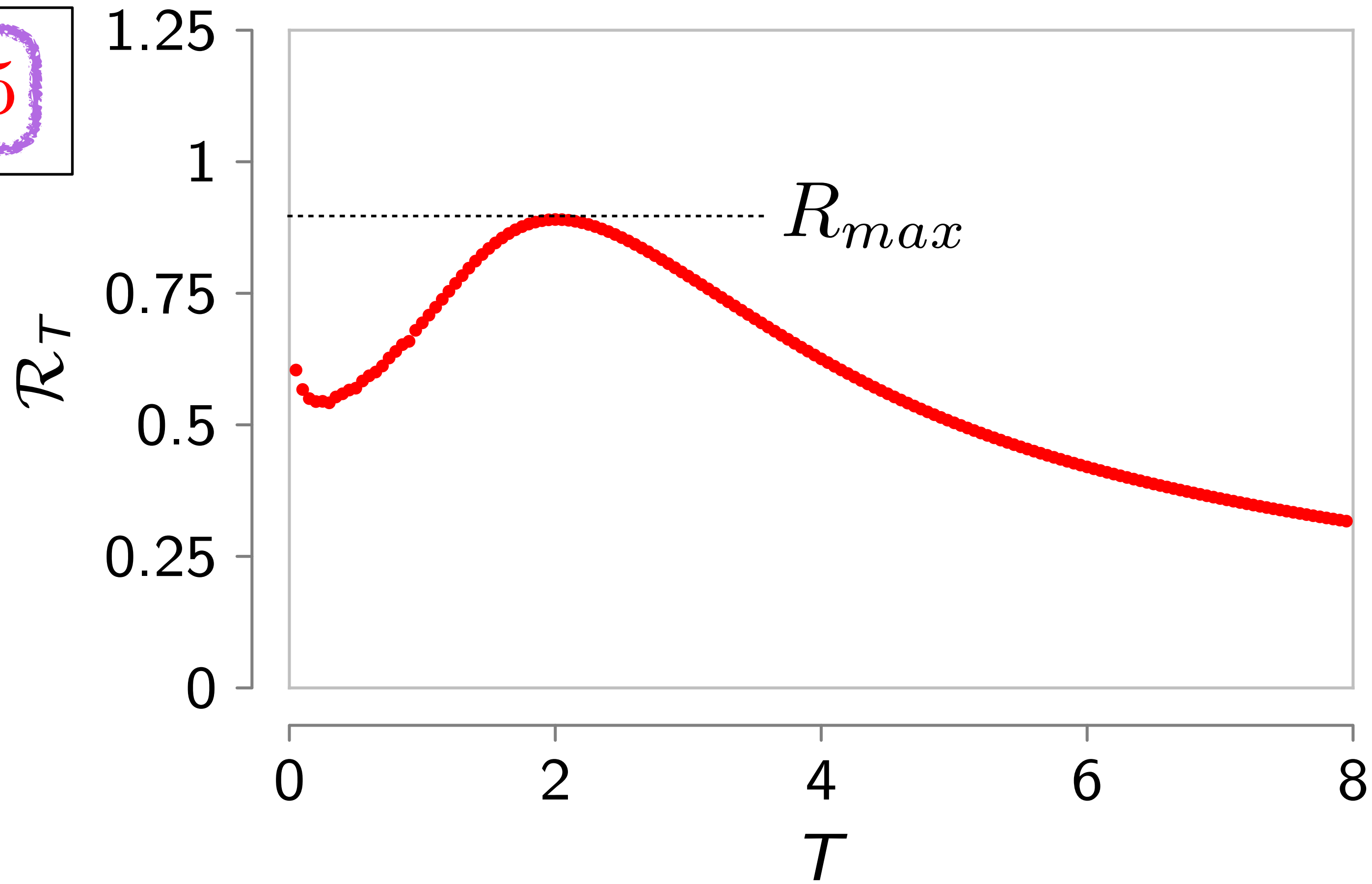
Example: Non normal matrix



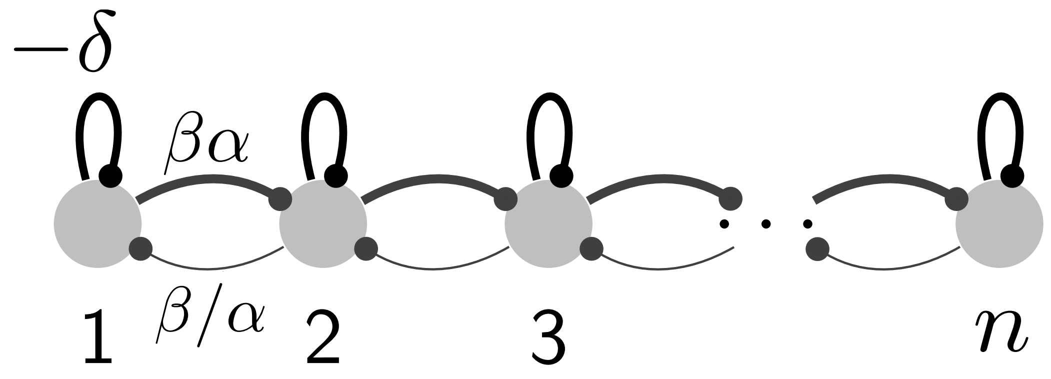
$$n = 10,\ \sigma^2 = 1,\ P = 1$$

$$\delta = 4,\ \beta = 1,\ \alpha = 5$$



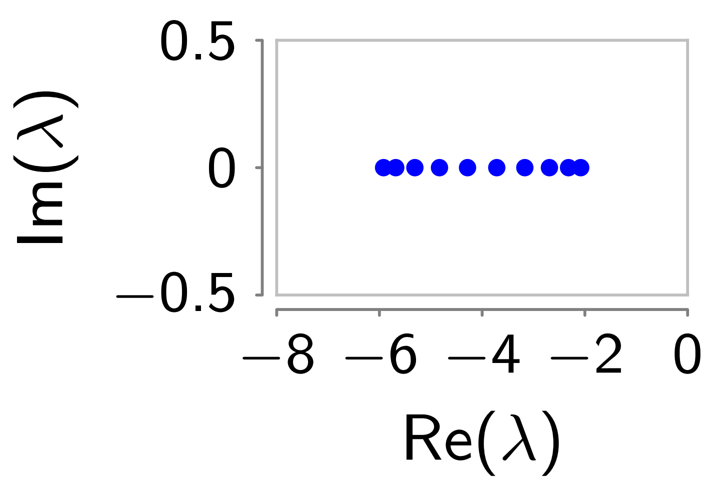


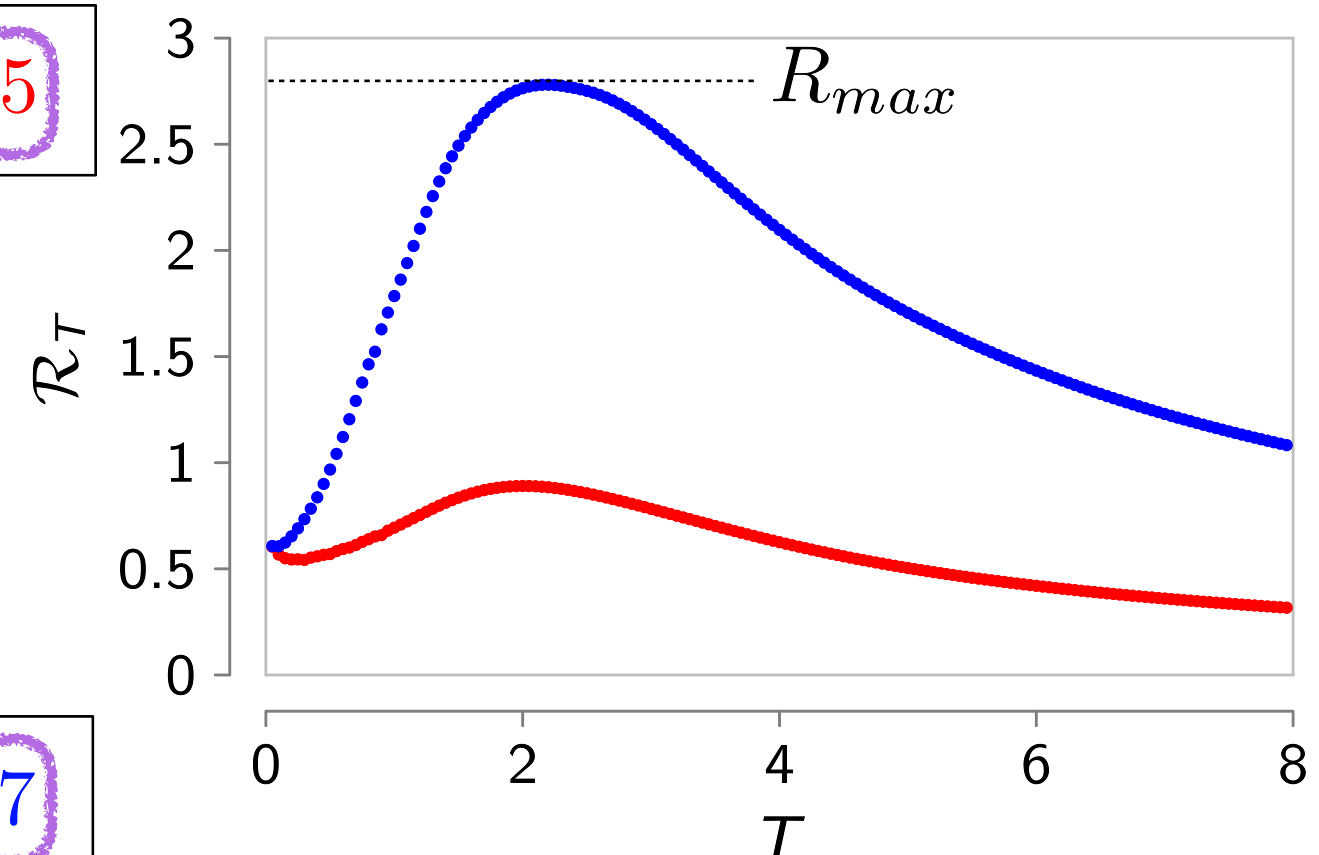
Example: Non normal matrix



$$n = 10, \ \sigma^2 = 1, \ P = 1$$

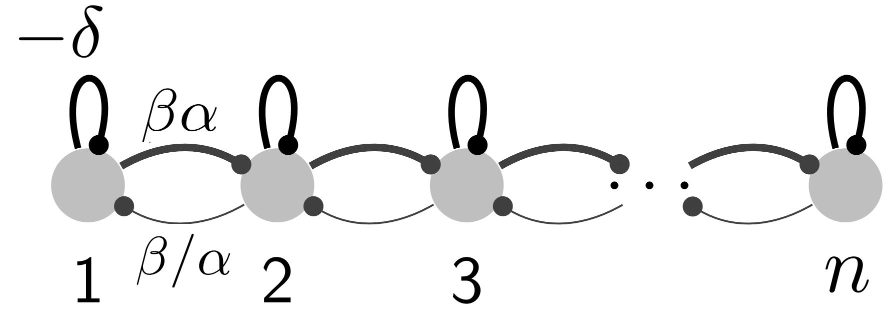
$\delta = 4, \ \beta = 1, \ \alpha = 5$



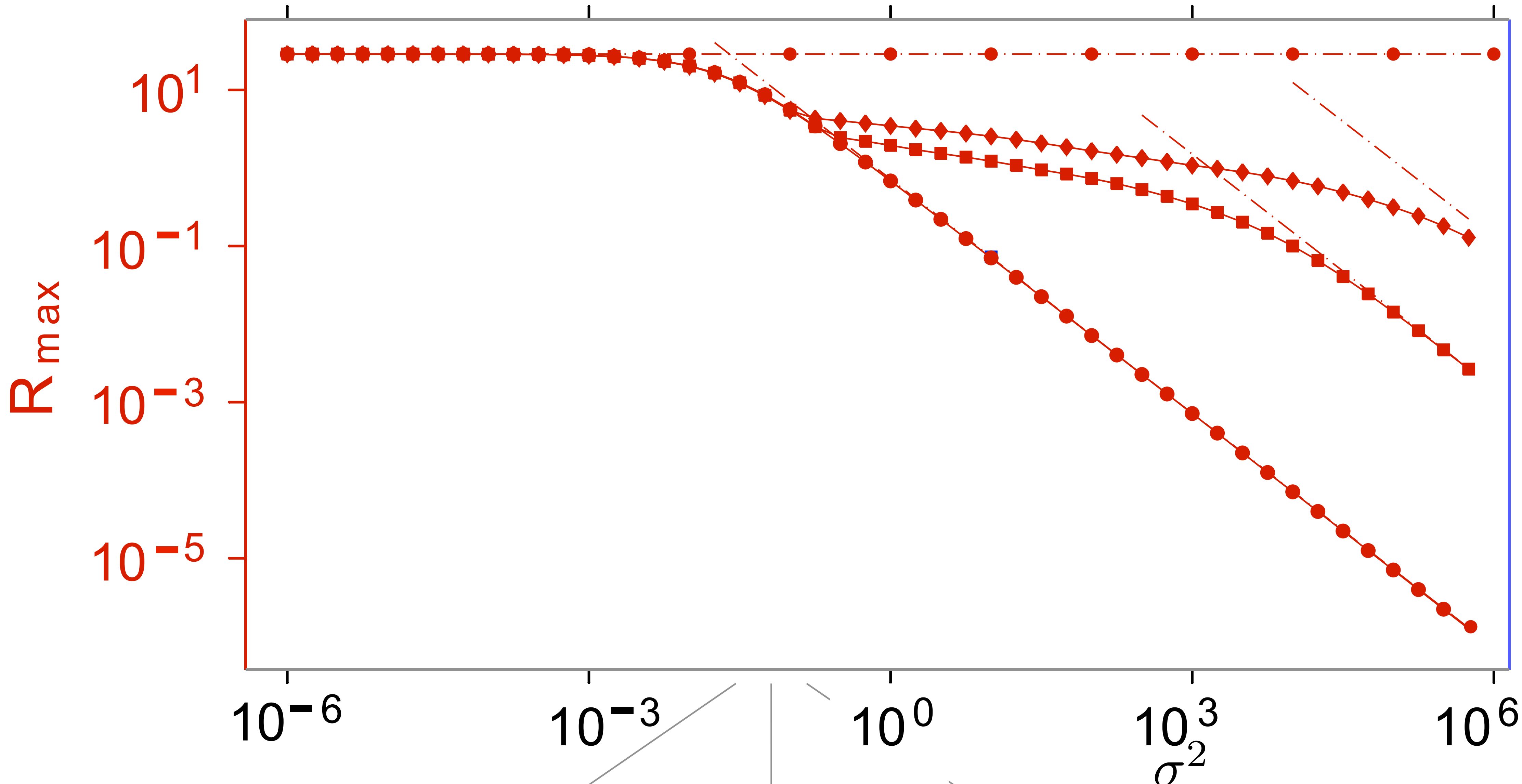


$\delta = 4, \ \beta = 1, \ \alpha = 7$

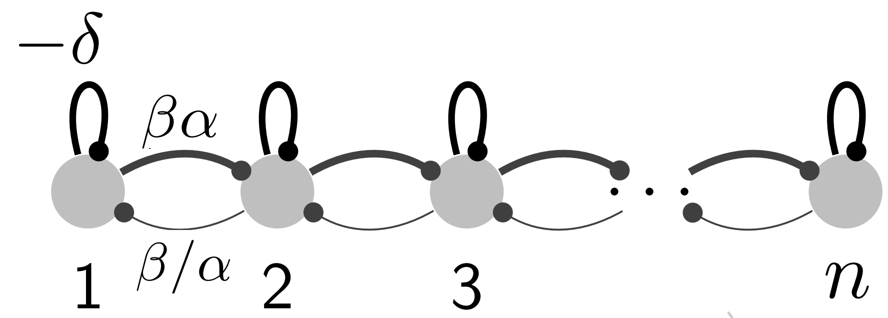
Line network: varying anysotropy strength



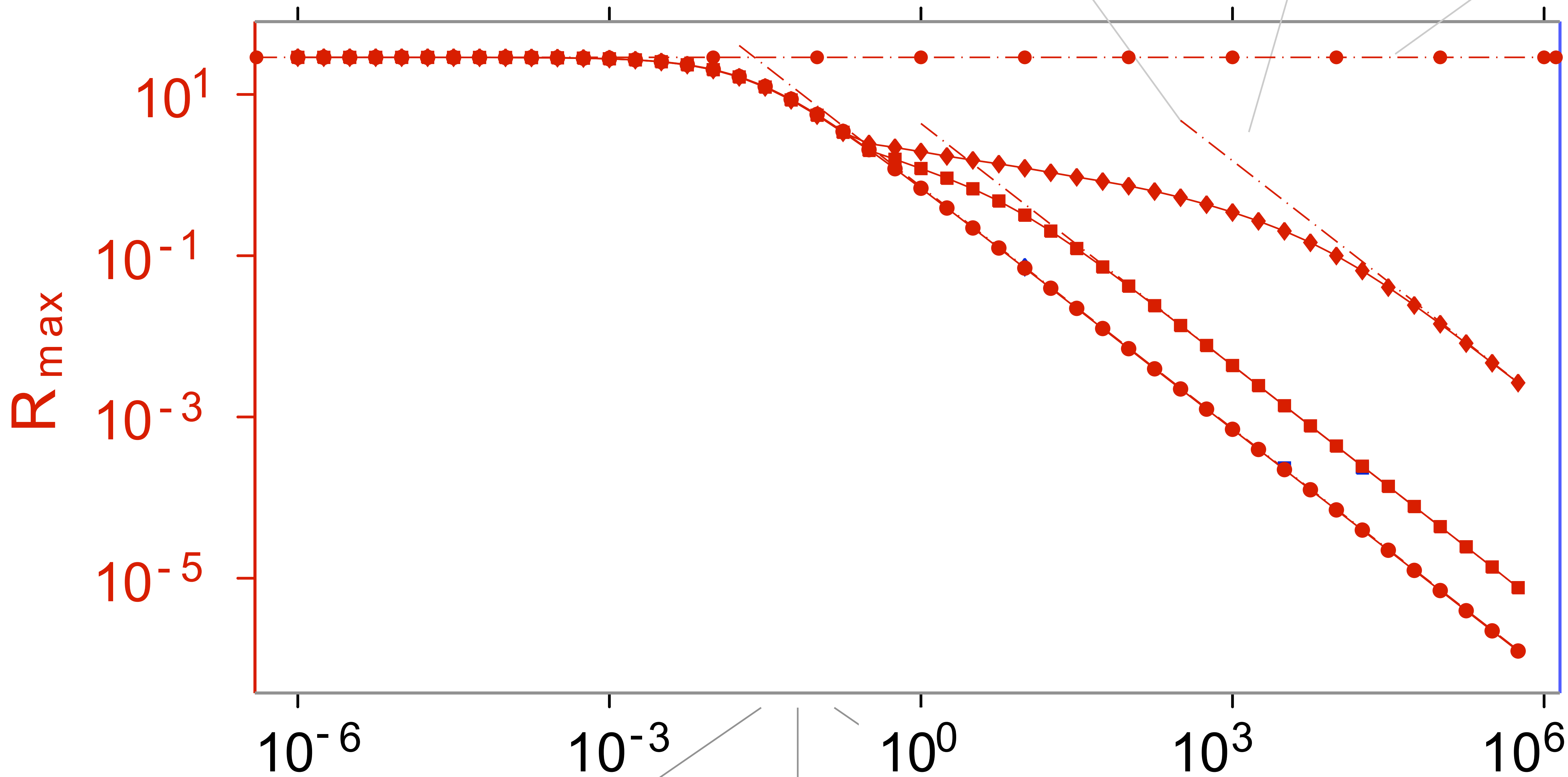
varying α



Line network: varying chain length



varying n



Limit behaviors in the noise

If the noise variance $\sigma^2 \to 0$, then

$$\boxed{R_{\max} \simeq -\frac{1}{\ln 2}\operatorname{tr}(A)}$$

In general

$$\boxed{R_{\max} \leq -\frac{1}{\ln 2}\operatorname{tr}(A)}$$

Limit behaviors in the noise

If the noise variance $\sigma^2 \to \infty$, then

$$R_{\max} \simeq \frac{1}{2\ln 2}\frac{\ell}{\sigma^2}$$

where

$$\ell := \max_{T\geq 0}\frac{\|B^T O_T B\|}{T} \qquad O_T = \int_0^T e^{A^T t}C^T C e^{At}dt$$

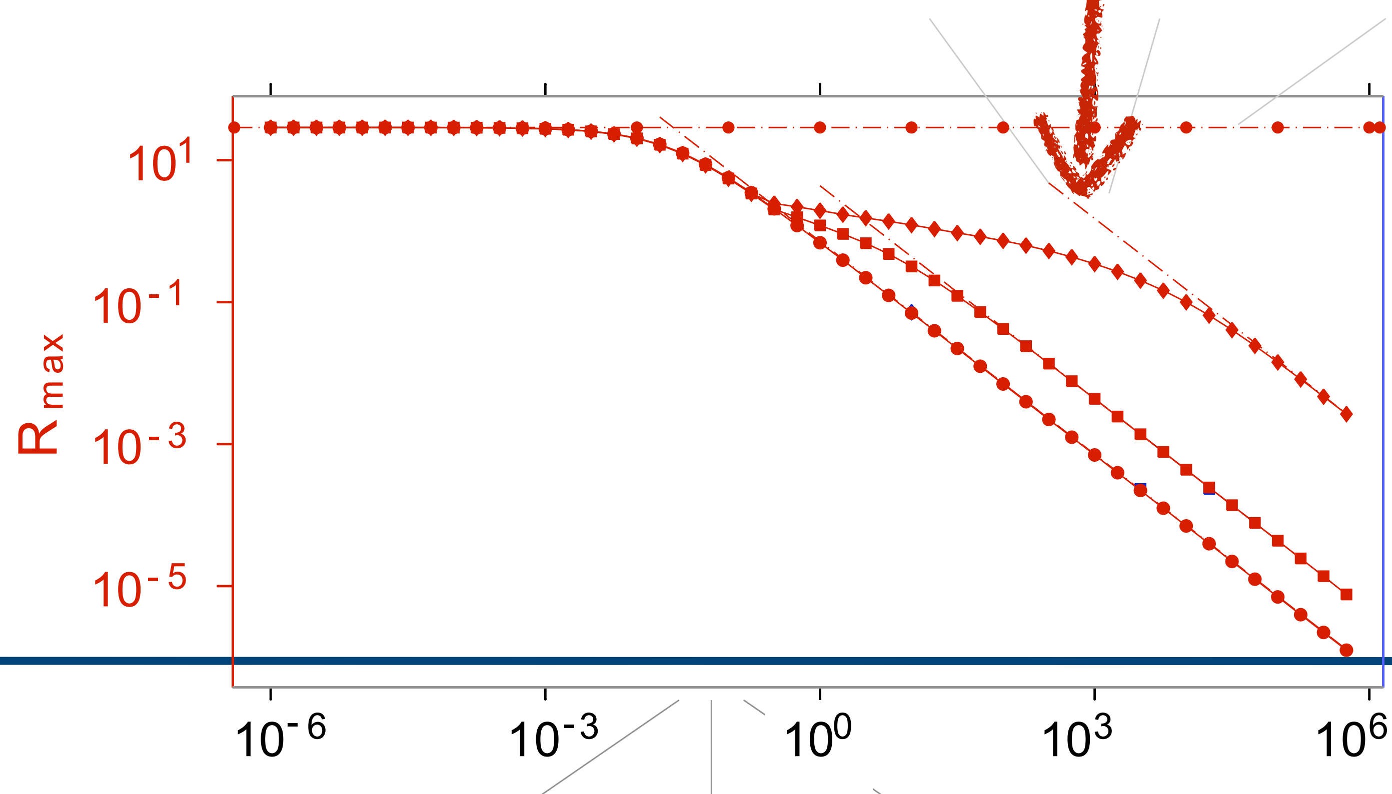
Limit behaviors in the noise

If the noise variance $\sigma^2 \to \infty$, then

$$R_{\max} \simeq \frac{1}{2\ln 2}\frac{\ell}{\sigma^2}$$

where

$$\ell := \max_{T \geq 0} \frac{\|B^T O_T B\|}{T} \qquad O_T = \int_0^T e^{A^T t} C^T C e^{At} dt$$



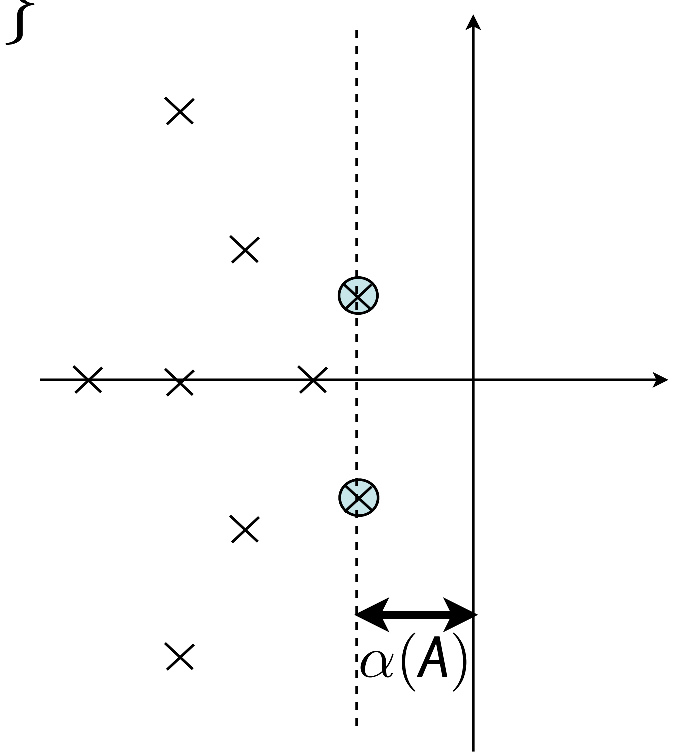
Matrix non-normality

Let

$$\alpha(A) = \max\{Re[\lambda_i] :\ \lambda_i \text{ eigenvalues of } A\}$$

← spectral abscissa

$$\omega(A) = \max\{\lambda_i :\ \lambda_i \text{ eigenvalues of } \frac{A+A^T}{2}\}$$



For normal matrices

$$\omega(A) = \alpha(A)$$

For non-normal matrices we may have

$$\omega(A) \geq \alpha(A)$$

Hence it may happen that $\omega(A) > 0$ also for stable matrices.

Phase transition

Theorem

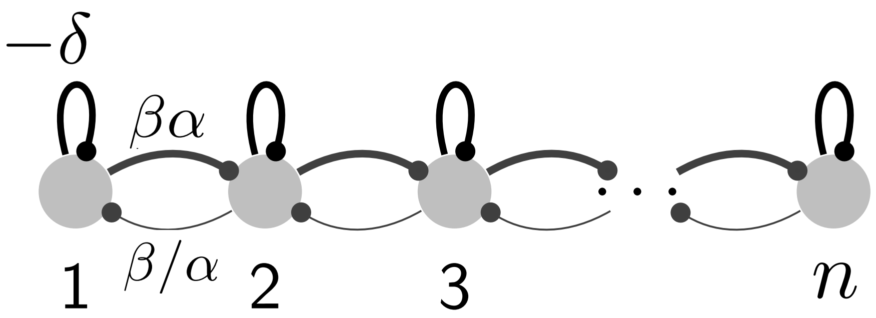
Let $A \in \mathbb{R}^{n \times n}$ be a stable matrix so that $\alpha(A) < 0$.

If $\omega(A) \leq 0$, then

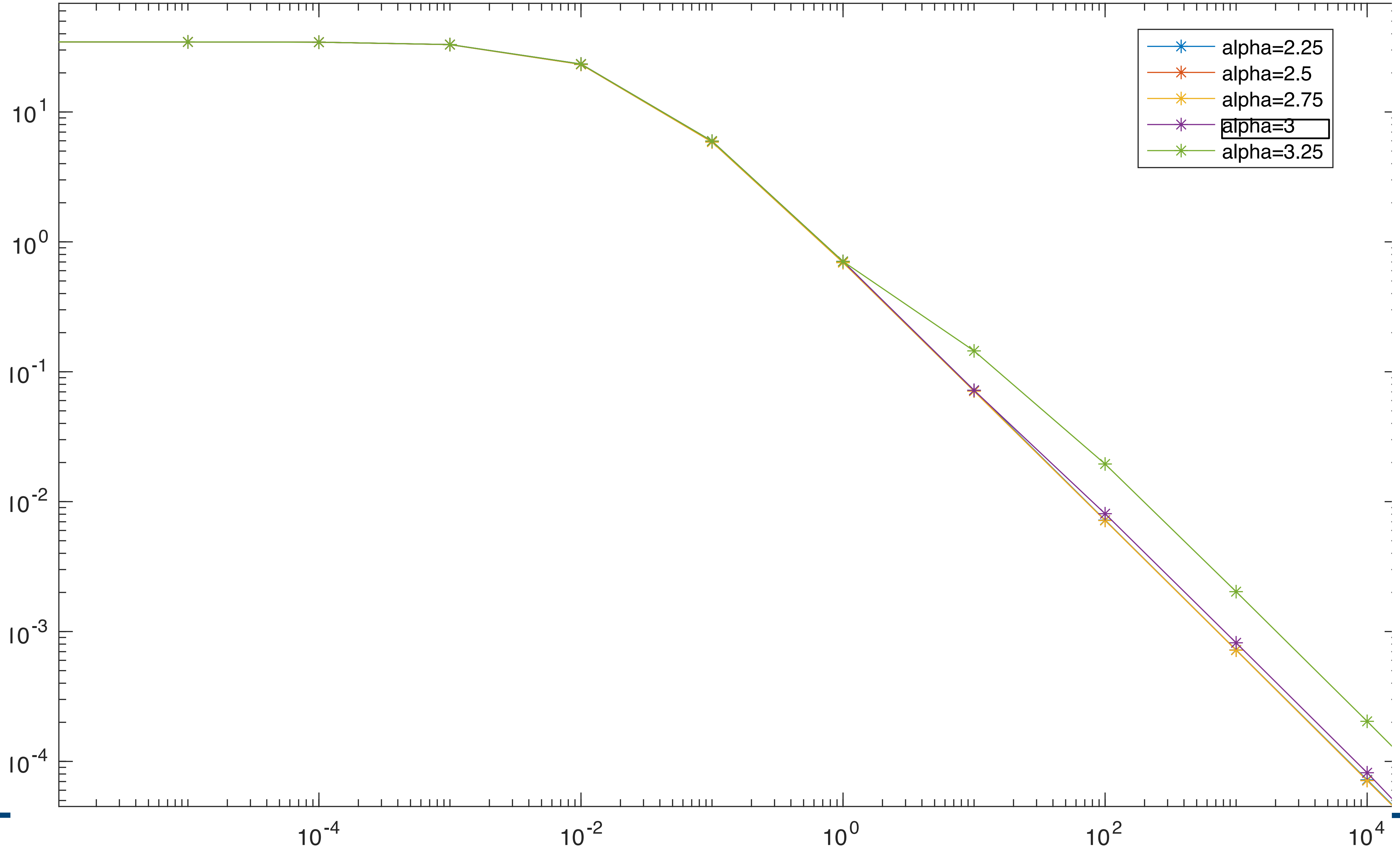
$$\ell(A, B, C) \leq 1$$

If $\omega(A) > 0$ and $B = C = I$, then

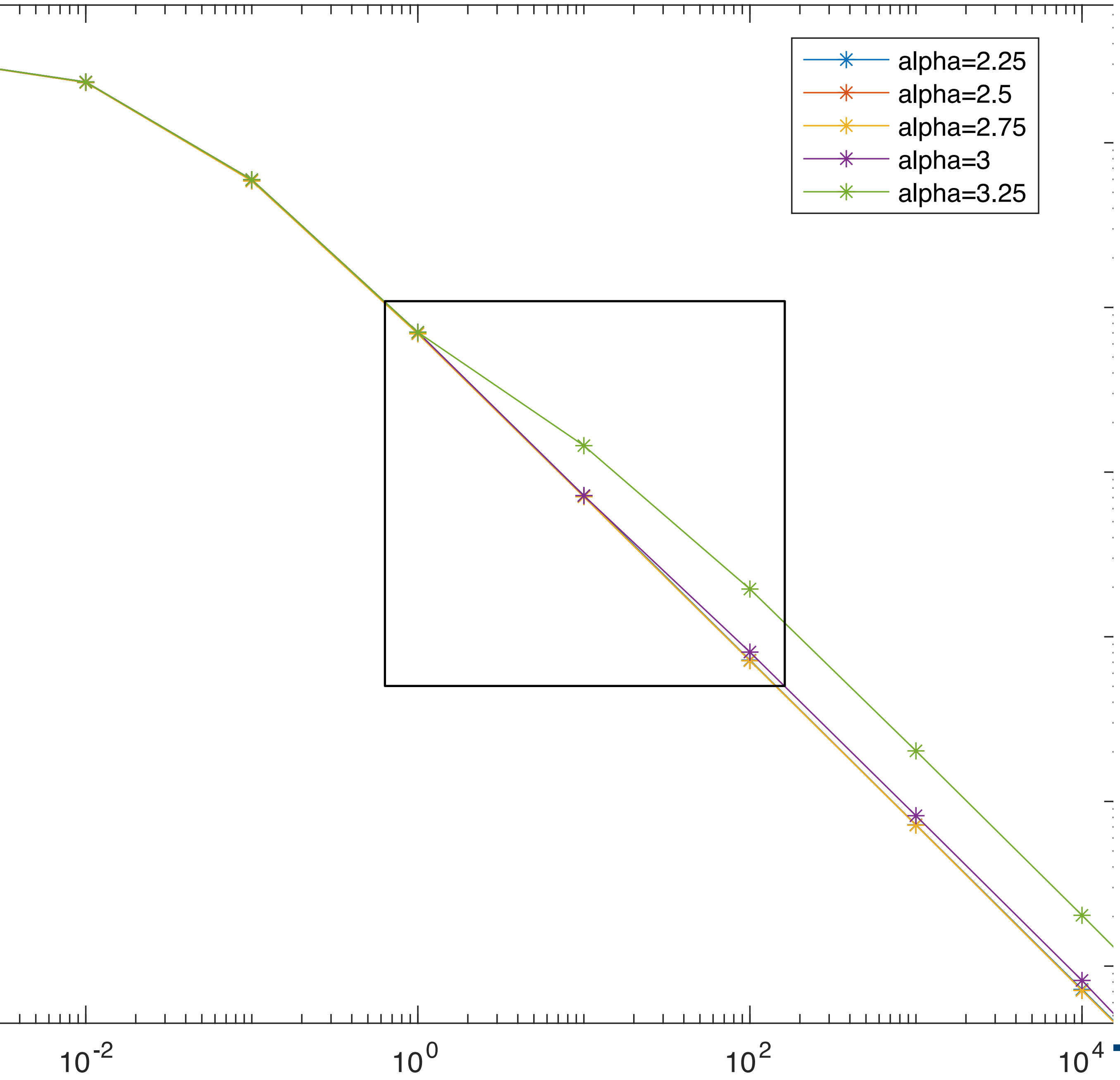
$$\ell(A, B, C) > 1$$



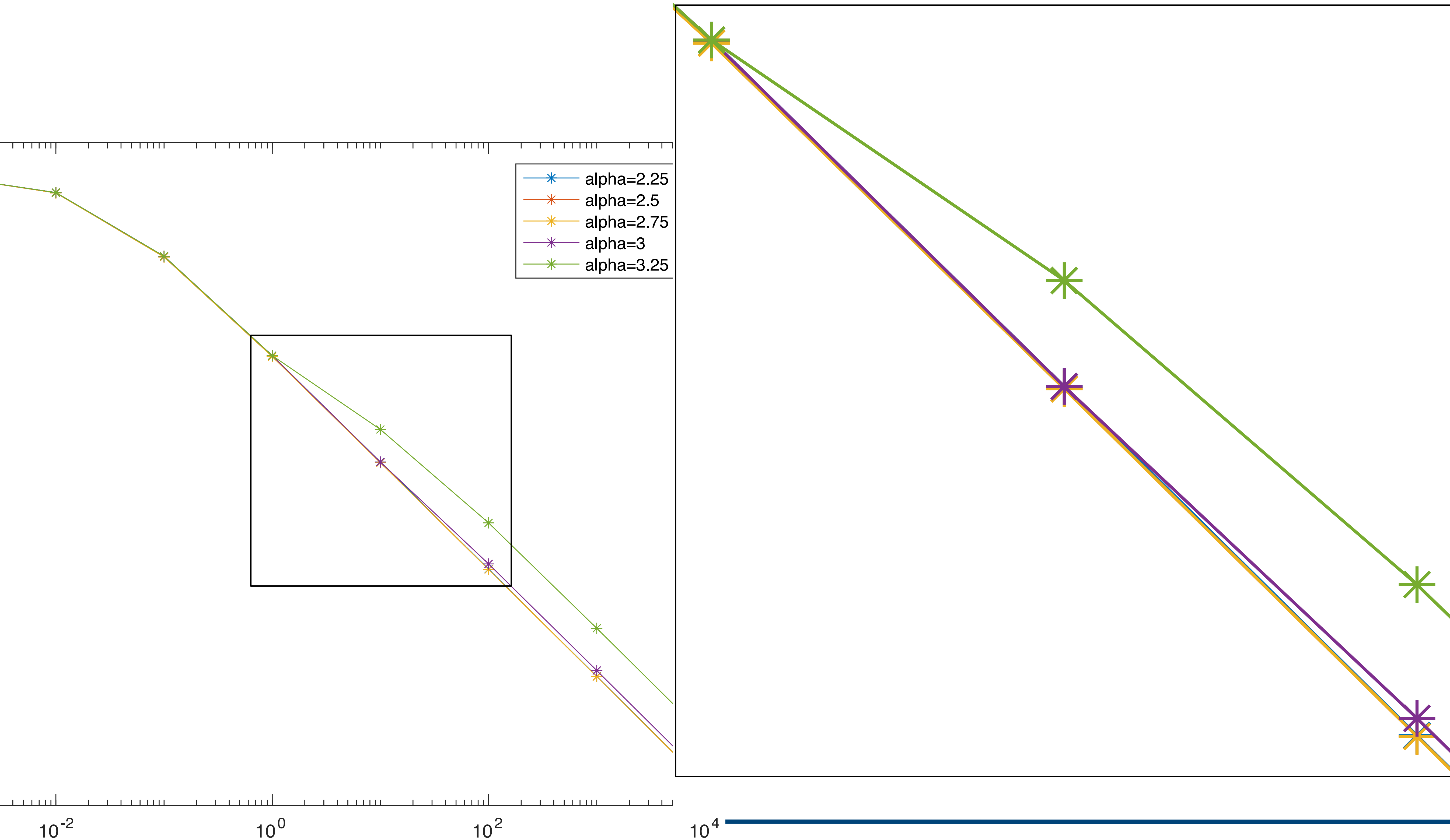
varying α

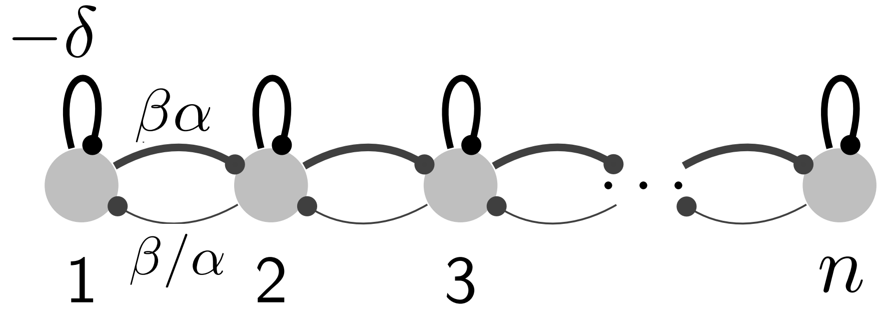


Line network

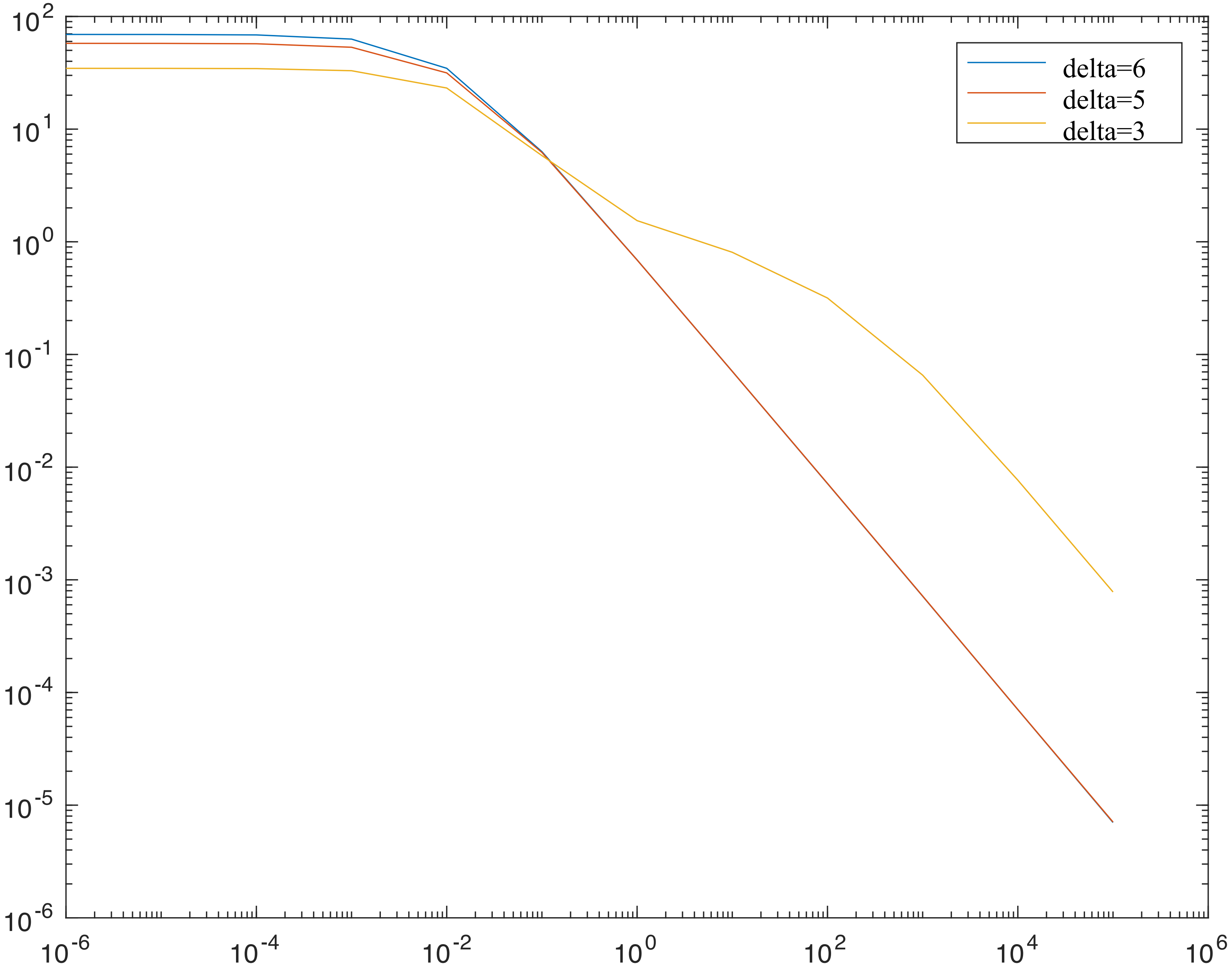


Line network

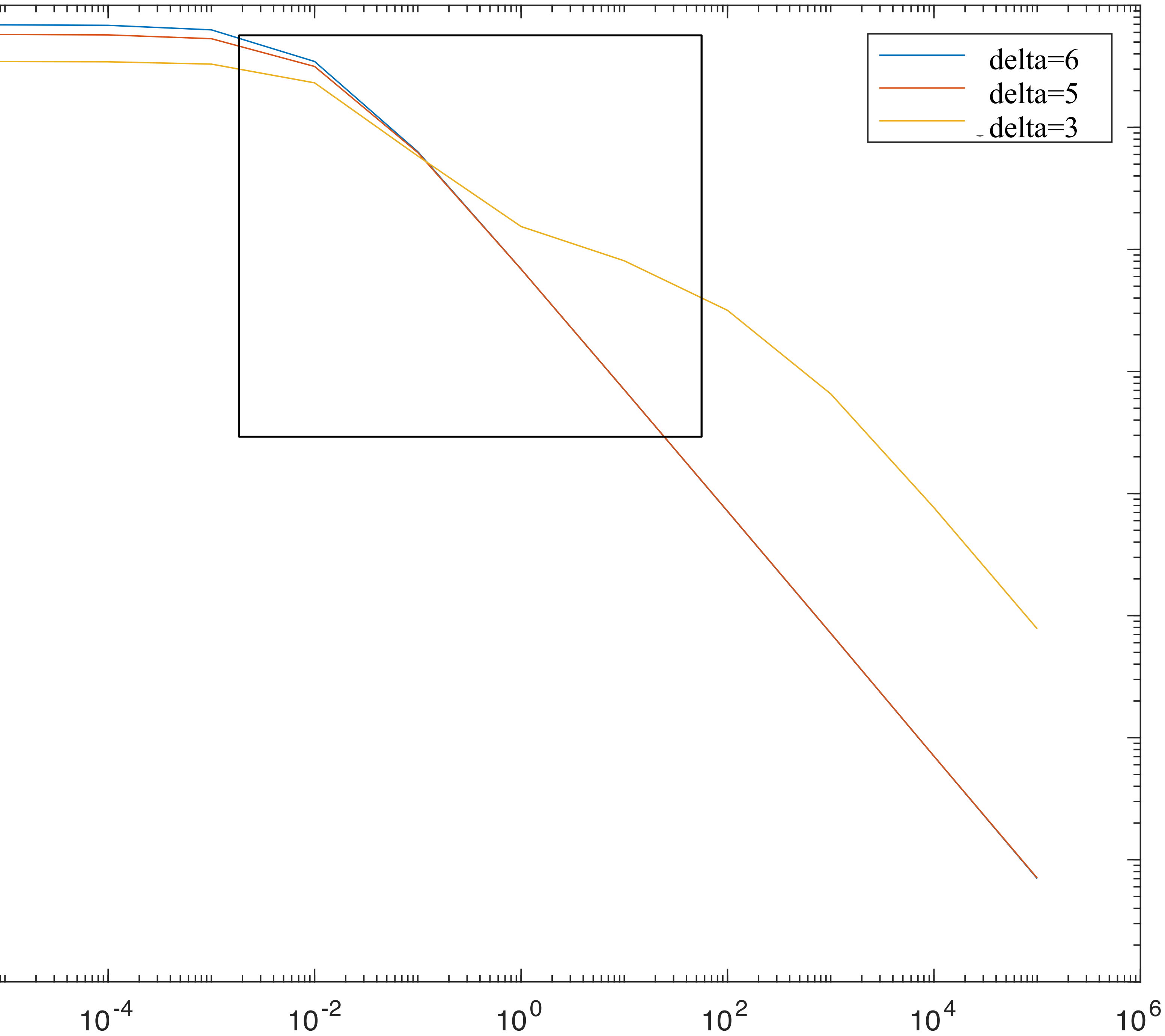




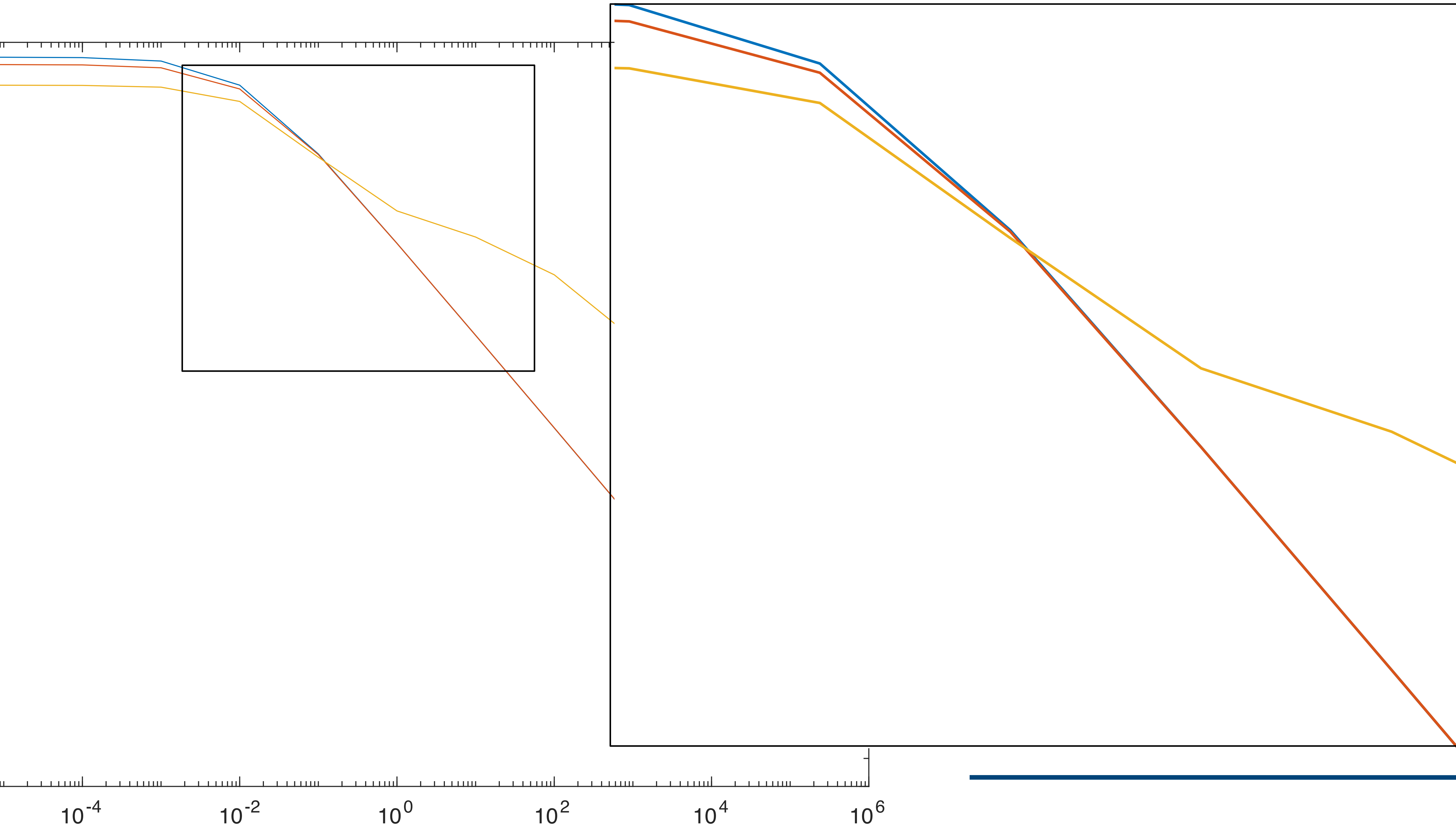
varying δ



Line network



Line network



Example

$$A = \begin{bmatrix} -\delta & 0 & & & \\ \alpha & -\delta & 0 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & -\delta & 0 \\ & & & \alpha & -\delta \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

Example

$$A = \begin{bmatrix} -\delta & 0 & & & \\ \alpha & -\delta & 0 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & -\delta & 0 \\ & & & \alpha & -\delta \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

For small noise $\quad R_{\max} \simeq \dfrac{\delta n}{\ln 2}$

Example

$$A = \begin{bmatrix} -\delta & 0 & & & \\ \alpha & -\delta & 0 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & -\delta & 0 \\ & & & \alpha & -\delta \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

For small noise $R_{\max} \simeq \dfrac{\delta n}{\ln 2}$ $\longrightarrow$ grows in δ

Example

$$A = \begin{bmatrix} -\delta & 0 & & & \\ \alpha & -\delta & 0 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & -\delta & 0 \\ & & & \alpha & -\delta \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

For small noise $R_{\max} \simeq \dfrac{\delta n}{\ln 2}$ → grows in δ

For large noise

$\omega(A) \simeq -\delta + \alpha$ ⇒ $\omega(A) > 0 \text{ iff } \alpha > \delta$

Example

$$A = \begin{bmatrix} -\delta & 0 & & & \\ \alpha & -\delta & 0 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & -\delta & 0 \\ & & & \alpha & -\delta \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

For small noise $\quad R_{\max} \simeq \dfrac{\delta n}{\ln 2} \longrightarrow$ grows in δ

For large noise

$$\omega(A) \simeq -\delta + \alpha \Longrightarrow \omega(A) > 0 \text{ iff } \alpha > \delta$$

$$\ell(A, B, C) \geq \frac{1}{4(2n-1)\sqrt{\pi(n-1)}} \left(\frac{\alpha}{\delta}\right)^{2n-2}$$

Example

$$A = \begin{bmatrix} -\delta & 0 & & & \\ \alpha & -\delta & 0 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & -\delta & 0 \\ & & & \alpha & -\delta \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

For small noise $\quad R_{\max} \simeq \frac{\delta n}{\ln 2}$ → grows in δ

For large noise

$\omega(A) \simeq -\delta + \alpha$ ⇒ $\omega(A) > 0$ iff $\alpha > \delta$

$$\ell(A, B, C) \geq \frac{1}{4(2n-1)\sqrt{\pi(n-1)}} \left(\frac{\alpha}{\delta}\right)^{2n-2}$$

Example

$$A = \begin{bmatrix} -\delta & 0 & & & \\ \alpha & -\delta & 0 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & -\delta & 0 \\ & & & \alpha & -\delta \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

For small noise $R_{\max} \simeq \dfrac{\delta n}{\ln 2}$ → grows in δ

For large noise

$\omega(A) \simeq -\delta + \alpha$ ⇒ $\omega(A) > 0$ iff $\alpha > \delta$

$$\ell(A, B, C) \geq \frac{1}{4(2n-1)\sqrt{\pi(n-1)}} \left(\frac{\alpha}{\delta}\right)^{2n-2}$$ → decreases in δ

Generating non-normal networks

The simplest way to generate highly non-normal matrices is by increasing matrix variability. Let

$$\mathcal{H}(A) := \frac{\max_{(i,j)\in\mathcal{E}}\{|A_{ij}|\}}{\min_{(i,j)\in\mathcal{E}}\{|A_{ij}|\}}$$

PROBLEM: How to increase ℓ while keeping the variability $\mathcal{H}(A)$ bounded.

Generating non-normal networks

Let $A \in \mathbb{R}^{n\times n}$ and $D = \text{diag}\{d_1, d_2, \ldots, d_n\}$ and consider DAD^{-1}

$$[DAD^{-1}]_{ij} = \frac{d_i}{d_j} A_{ij}$$

In the sinple input single output case, if $B = e_1$ and $C = e_n^T$, then

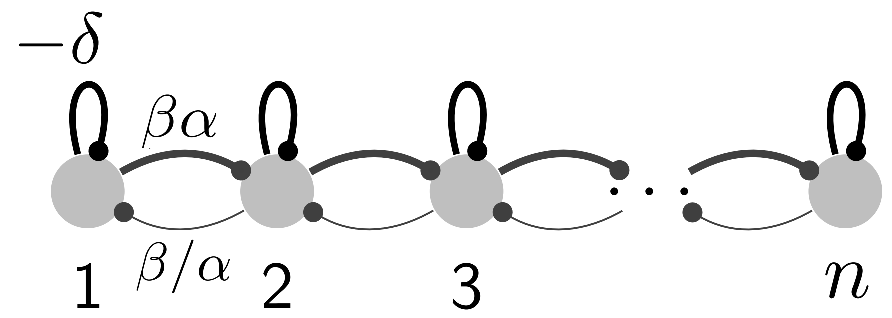
$$\ell(DAD^{-1}) \geq \frac{d_n^2}{d_1^2} \ell(A)$$

Large global variability

Small local variability

$$\mathcal{H}(DAD^{-1}) \leq \max_{(i,j)\in\mathcal{E}} \left\{ \frac{d_j^2}{d_i^2} \right\} \mathcal{H}(A)$$

Generating non-normal networks

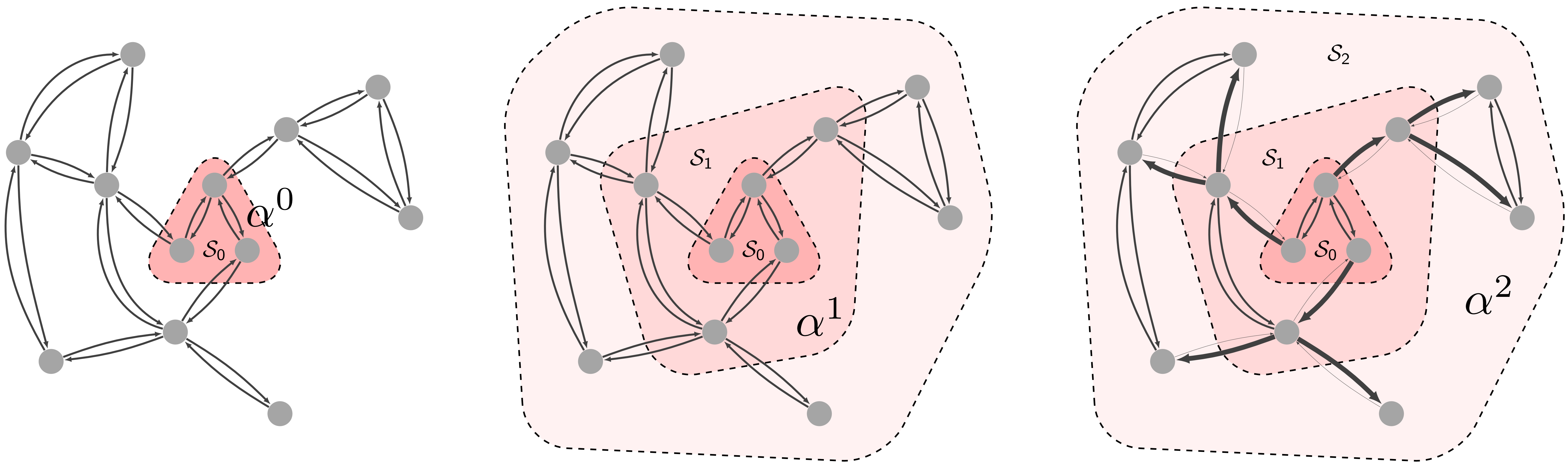


In the line example with $B = e_1$ and $C = e_n^T$, taking $d_i = \alpha^i$ we obtain

$$\ell(DAD^{-1}) \geq \alpha^{2(n-1)} \ell(A)$$ Global variability

$$\mathcal{H}(DAD^{-1}) \leq \alpha \mathcal{H}(A)$$ Local variability

Layering



$$\text{Set } d_i = \alpha^h \text{ for } i \in \mathcal{S}_h \text{ and } \bar{A} := DAD^{-1}$$

Layering

For every edge $(i, j) \in \mathcal{E}$

1. if $i, j \in \mathcal{S}_h$, then $\bar{A}_{ij} = A_{ij}$;
2. if $i \in \mathcal{S}_{h-1}$ and $j \in \mathcal{S}_h$, then $\bar{A}_{ij} = \frac{1}{\alpha} A_{ij}$;
3. if $i \in \mathcal{S}_h$ and $j \in \mathcal{S}_{h-1}$, then $\bar{A}_{ij} = \alpha A_{ij}$.

Layering

For every edge $(i, j) \in \mathcal{E}$

1. if $i, j \in \mathcal{S}_h$, then $\bar{A}_{ij} = A_{ij}$;
2. if $i \in \mathcal{S}_{h-1}$ and $j \in \mathcal{S}_h$, then $\bar{A}_{ij} = \frac{1}{\alpha} A_{ij}$;
3. if $i \in \mathcal{S}_h$ and $j \in \mathcal{S}_{h-1}$, then $\bar{A}_{ij} = \alpha A_{ij}$.

Then

$$\mathcal{H}(\bar{A}) \leq \alpha^2 \mathcal{H}(A)$$

while

$$\ell(\bar{A}) \geq \alpha^{2d} \ell(A)$$

where d is the number of layers in the layering.

Conclusions

1. We proposed a model to quantify the information transmission performance in linear dynamic networks.

2. By introducing the inter-symbol interference, we could highlight the role of non-normality of the dynamics.

3. General non-normal dynamics is hard to study.

4. Low and high noise regimes are more treatable.

5. In the low regime it is convenient to have large trace(A) while non-normality does not play any role.

6. In the high noise regime non-normality plays an important role.

7. In general it is not clear how to build highly non-normal networks. We proposed a method based on the network layering.

Thank you

Matrix non-normality

$$\boldsymbol{A} \in \mathbb{R}^{N \times N}$$

normal matrices

$$\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{A}^\top \boldsymbol{A}$$

$$\boldsymbol{A} = \boldsymbol{U}^* \boldsymbol{D} \boldsymbol{U},$$

$\boldsymbol{U}$ unitary, $\boldsymbol{D}$ diagonal

(spectral decomposition)

$$\boldsymbol{D} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix}$$

$$\sigma(\boldsymbol{A}) = \{\lambda_i\}_{i=1}^N$$

non-normal matrices

$$\boldsymbol{A}\boldsymbol{A}^\top \neq \boldsymbol{A}^\top \boldsymbol{A}$$

$$\boldsymbol{A} = \boldsymbol{U}^* \boldsymbol{T} \boldsymbol{U},$$

$\boldsymbol{U}$ unitary, $\boldsymbol{T}$ triangular

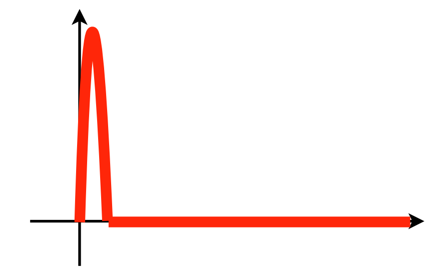
(Schur decomposition)

$$\boldsymbol{T} = \begin{bmatrix} \lambda_1 & & & \\ \star & \lambda_2 & & \\ \star & \star & \ddots & \\ \star & \star & \star & \lambda_N \end{bmatrix}$$

$$\boldsymbol{T} = \boldsymbol{D} + \boldsymbol{N}$$

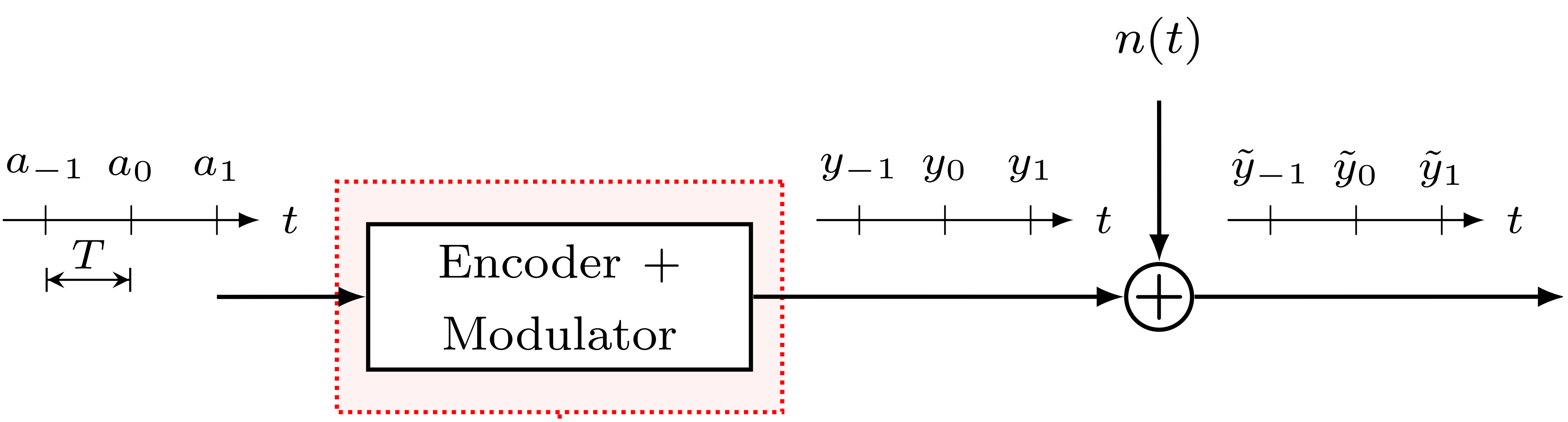
At time $t = 0$ with a symbol $a \in \mathcal{A}$ we associate an impulsive input

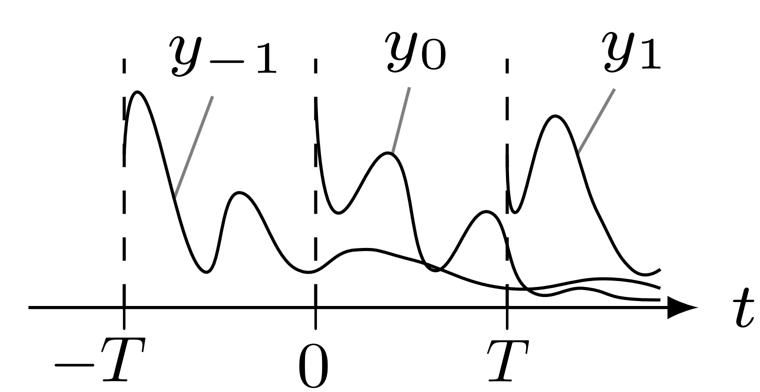
$$u(t) = u_0 \delta(t) \qquad \text{with } \|u_0\| \leq P$$



$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t)$$

$$\Rightarrow \quad y_0(t) = Ce^{At}Bu_0$$



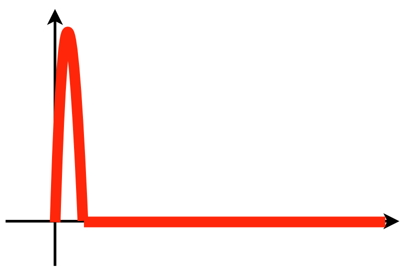


$$y_k(t) = Ce^{A(t-kT)}Bu_k$$

Modulation by a system transient

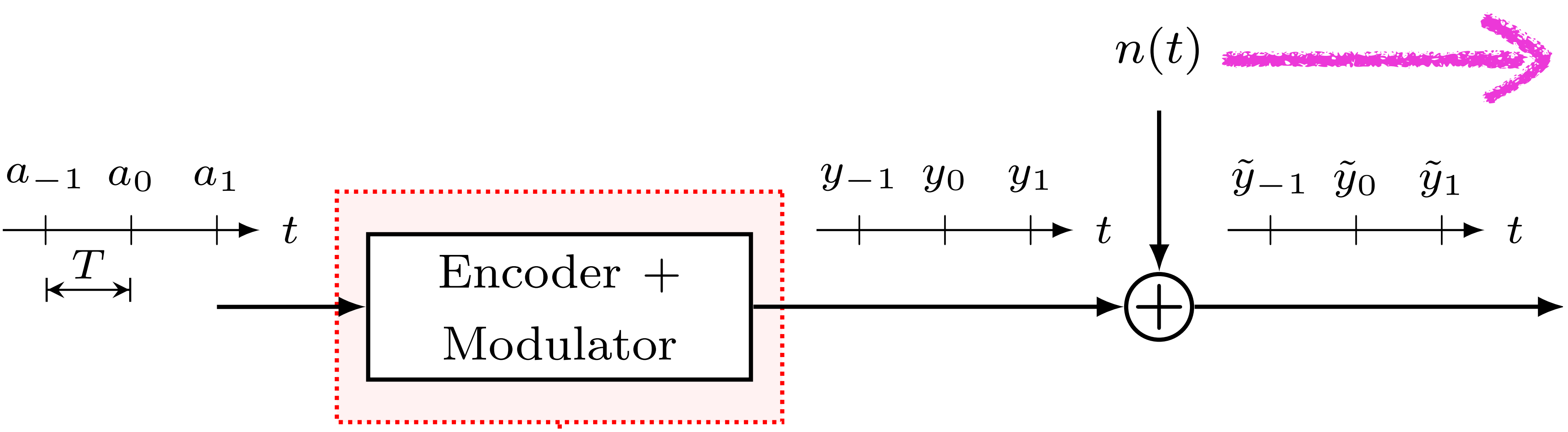
At time $t = 0$ with a symbol $a \in \mathcal{A}$ we associate an impulsive input

$$u(t) = u_0 \delta(t) \qquad \text{with } \|u_0\| \leq P$$



$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t)$$

$\Rightarrow$ $y_0(t) = Ce^{At}Bu_0$

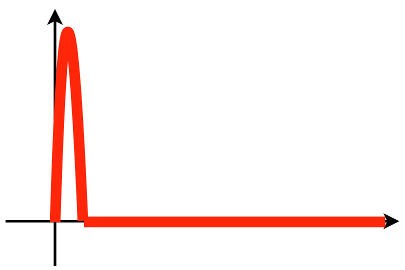


σ^2 is the noise power

Modulation by a system transient

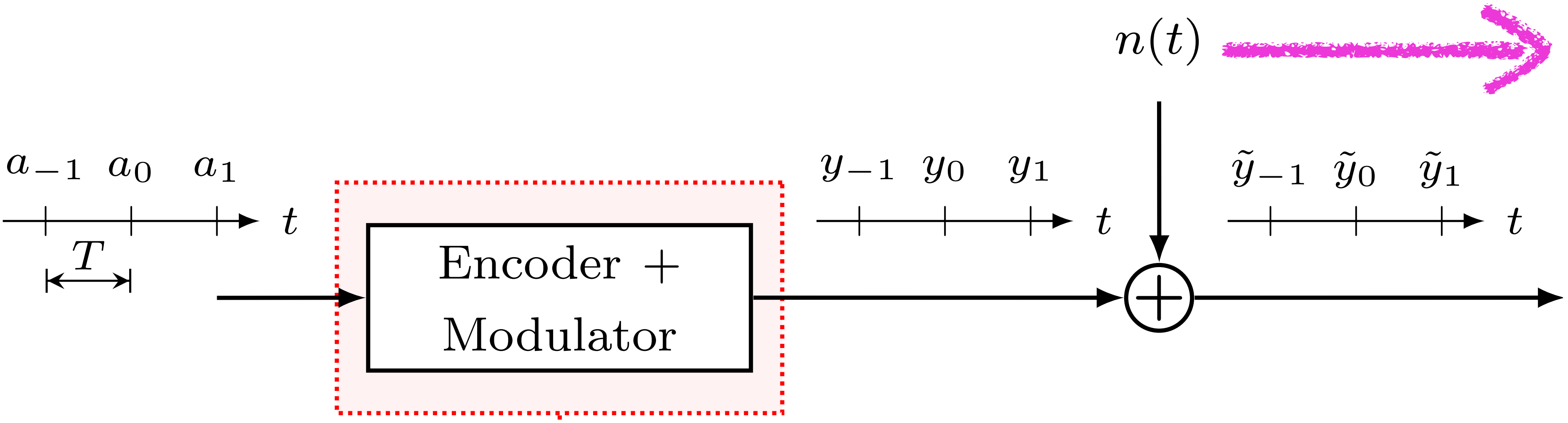
At time $t = 0$ with a symbol $a \in \mathcal{A}$ we associate an impulsive input

$$u(t) = u_0 \delta(t) \qquad \text{with } \|u_0\| \leq P$$



$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t)$$

$$\Rightarrow \quad y_0(t) = Ce^{At}Bu_0$$



σ^2 is the noise power

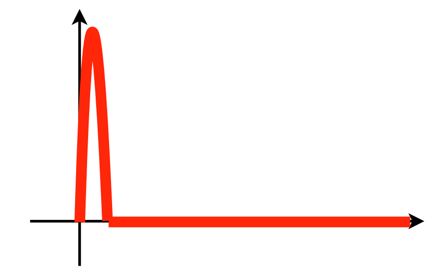
$SNR = P/\sigma^2$

signal to noise ratio

Modulation by a system transient

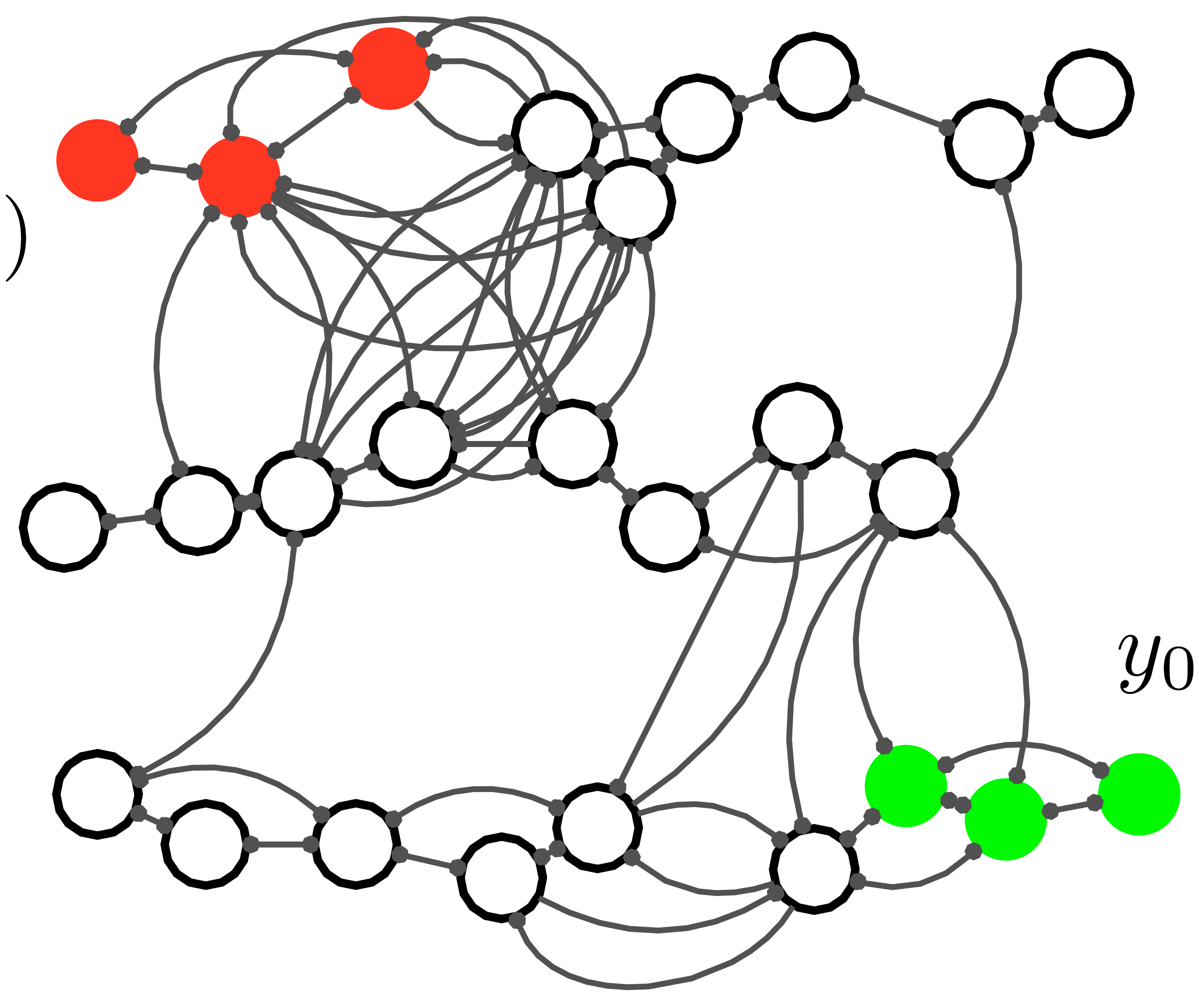
At time $t = 0$ with a symbol $a \in \mathcal{A}$ we associate an impulsive input

$$u(t) = u_0 \delta(t) \qquad \text{with } \|u_0\| \leq P$$

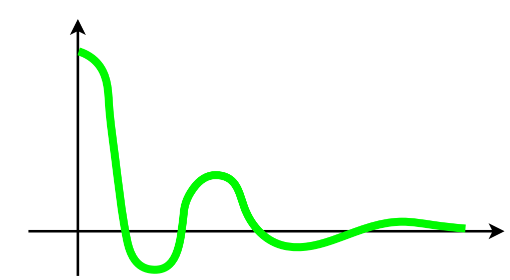


$a \in \mathcal{A}$

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t)$$



$$y_0(t) = Ce^{At}Bu_0$$



Example

$$A = \begin{bmatrix} -\delta & \alpha & 0 & & \\ 0 & -\delta & \alpha & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \\ & & \ddots & -\delta & \alpha \\ & & & 0 & -\delta \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, C = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

If the noise is small then

$$R_{\max} \simeq \frac{\delta n}{\ln 2} \longrightarrow \text{grows in } \delta$$

Notice moreover that $\omega(A) \simeq -\delta + \alpha$ and so $\alpha > \delta$ guarantees that $\omega(A) > 0$.

$$\ell(A, B, C) \geq \frac{1}{4(2n-1)\sqrt{\pi(n-1)}} \left(\frac{\alpha}{\delta}\right)^{2n-2} \longrightarrow \text{decreases in } \delta$$

Stratification

Set $d_h = \alpha^h$ for $i \in \mathcal{S}_h$. Then, for every edge $(i,j) \in \mathcal{E}$:

1. if $i, j \in \mathcal{S}_h$, then $\bar{A}_{ij} = A_{ij}$;
2. if $i \in \mathcal{S}_{h-1}$ and $j \in \mathcal{S}_h$, then $\bar{A}_{ij} = \frac{1}{\alpha} A_{ij}$;
3. if $i \in \mathcal{S}_h$ and $j \in\in \mathcal{S}_{h-1}$, then $\bar{A}_{ij} = \alpha A_{ij}$.

Then

$$\mathcal{H}(\bar{A}) \leq \alpha^2 \mathcal{H}(A)$$

while

$$\ell(\bar{A}) \geq \alpha^{2d} \ell(A)$$

where d is the number of layers in the stratification.

Generating non-normal networks

$A \in \mathbb{R}^{n \times n}$ and $D = \mathrm{diag}\{d_1, d_2, \ldots, d_n\}$

$$[DAD^{-1}]_{ij} = \frac{d_i}{d_j} A_{ij}$$

$$\ell(DAD^{-1}, B, C) \geq \max_{i \in \mathcal{B}, j \in \mathcal{C}} \frac{d_j^2}{d_i^2} \ell(A, e_i, e_j^T)$$

PROBLEM: How to grow ℓ while keeping the variability in A bounded

$$\mathcal{H}(A) := \frac{\max_{(i,j) \in \mathcal{E}} \{|A_{ij}|\}}{\min_{(i,j) \in \mathcal{E}} \{|A_{ij}|\}}$$

Generating non-normal networks: an example

$$\mathcal{H}(\bar{A}) = max_i\{\epsilon_i\}\mathcal{H}(A)$$

$$\ell(\bar{A}) = \left(\prod_{i=1}^{n-1} \epsilon_i\right)^2 \ell(A)$$

The best choice of ϵ_i is to take $\epsilon_i = \epsilon$ for all i

$$\ell(\bar{A}) = \epsilon^{2n-2}\ell(A), \quad \mathcal{H}(\bar{A}) = \epsilon\mathcal{H}(A)$$