Supplementary Information

Pseudo-Bifurcations in Stochastic Non-Normal Systems and the Limits of Early-Warning Signals

Supplementary Note 1. Definition of Linear Stochastic System

Throughout this Supplementary Information, we consider an N-dimensional linear stochastic system of the form

$$\dot{\mathbf{x}}_t \;=\; \mathbf{A}\,\mathbf{x}_t + \sqrt{2\delta}\,\boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \overset{i.i.d}{\sim} \mathcal{N}(0,1) \tag{S1}$$

The $N \times N$ matrix $\mathbf{A}$, which governs the interactions of the system, is assumed diagonalizable with eigenvalues $\lambda_1, \ldots, \lambda_N$ all having negative real parts, thus ensuring that the equilibrium point $\mathbf{x} = 0$ remains stable. The diagonal N-dimensional noise term $\boldsymbol{\eta}_t$ is assumed to be normally distributed, and the scalar $\delta > 0$ is a scaling factor.

Supplementary Note 2. Proof of Non-Normal Dimensional Reduction

In this section, we systematically derive the representation of the non-normal system (S1) as an effective two-dimensional dynamics along the non-normal and reaction mode. We first provide a short review of the singular value decomposition that our analysis relies upon. We then provide some intuition for the emergence of transients, followed by a detailed derivation of the two effective system components.

2.1 Singular Value Decomposition

Because $\mathbf{A}$ is diagonalizable, there exists a matrix $\mathbf{P}$ which is the eigenbasis transformation of $\mathbf{A}$ such that

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1} \tag{S2}$$

where $\boldsymbol{\Lambda} = \mathrm{Diag}\,(\lambda_1, \cdots, \lambda_N)$ is the diagonal matrix composed of the eigenvalues of $\mathbf{A}$. Here, we assume that $\lambda_i < 0\ \forall i$ and we explicitly set the minus sign to indicate that the eigenvalues are negative.

For $\mathbf{A}$ normal, that is, for $\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger\mathbf{A}$, then $\mathbf{P}$ is unitary and $\mathbf{A}$ can be diagonalized by means of a rotation. But in general, we assume here that $\mathbf{A}$ is non-normal, and hence $\mathbf{P}$ is not unitary. Quite generally, a matrix $\mathbf{P}$ can be represented as a sequence of a rotation, a rescaling, and another rotation, via the singular value decomposition (SVD) given by

$$\mathbf{P} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger \tag{S3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, $\mathbf{V}^\dagger$ is the Hermitian conjugate of $\mathbf{V}$ and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \cdots, \sigma_N)$ is a diagonal matrix of the singular values $\{\sigma_i\}$. Without loss of generality, we can assume that $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_N > 0$ because a permutation of a unitary matrix remains unitary.

Combining (S2) with (S3) we can write

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger\boldsymbol{\Lambda}\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^\dagger = \sum_{i,j=1}^{N} \frac{\sigma_i}{\sigma_j}\lambda_{ij}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_j^\dagger, \qquad \text{with } \lambda_{ij} = -\sum_{k=1}^{n} \lambda_k v_{ik} v_{jk}^*, \tag{S4}$$

and $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ denote the i^{th} column of the unitary matrix $\mathbf{U}$ and $\mathbf{V}$, respectively. Further, v_{ij} is the element of $\mathbf{V}$ at the i^{th} column and j^{th} row and v_{ij}^* is the complex conjugate of v_{ij}.

Expression (S4) is crucial for our subsequent derivations, since it shows how the matrix $\mathbf{A}$ can be decomposed into N^2 rank-1 projections $\hat{\mathbf{u}}_i\hat{\mathbf{u}}_j^\dagger$ modulated by a factor $\frac{\sigma_i}{\sigma_j}\lambda_{ij}$. As we will see below, the factor λ_{ij} is of order one, such that the dominant contributions to $\mathbf{A}$ will come from components where σ_i/σ_j is large.

2.2 Effective Two-Dimensional Dynamics

We now show how a system (S1) that is non-normal along one dominant component can be decomposed into an effective two-dimensional dynamics. To this end, recall that the SVD (S3) of $\mathbf{P}$ is composed of two rotations with a scaling along the singular values $\mathbf{\Sigma}$ in-between. The greater the deviation of these singular values from 1, the more pronounced the scaling effect. For the remainder of this Supplementary Information, we assume that only the final dimension has a singular value significantly smaller than 1, while all other dimensions are roughly equal to one, that is $1 = \sigma_1 \gtrsim \sigma_2 \gtrsim \ldots \gtrsim \sigma_{N-1} \gg \sigma_N$. This assumption essentially encapsulates the non-normal behavior along one component. We further define matrix $\mathbf{A}$'s condition number as $\kappa = \sigma_1/\sigma_N \gg 1$. For the remaining singular values, $i = 2, \ldots, N-1$, we write $\sigma_i = (1-\delta_i)\sigma_1$ with $\delta_i \ll 1$ such that σ_i is slightly less than but close to $\sigma_1 = 1$. To leading order, we can thus write

$$\frac{\sigma_i}{\sigma_j} = \begin{cases} 1 + (\delta_j - \delta_i) + O(\delta_i\delta_j, \delta_j^2) & \text{if } i,j \neq N \\ \kappa(1-\delta_i) & \text{if } i \neq N, j = N \\ \kappa^{-1} + O(\kappa^{-1}\delta_j) & \text{if } i = N, j \neq N \\ 1 & \text{if } i = j = N. \end{cases} \tag{S5}$$

Defining further $\delta^2 = \sum_i \delta_i^2$ and plugging these approximations into (S4) allows us to decompose the matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{A}' + (\kappa - 1)a_{nr}\hat{\mathbf{r}}\hat{\mathbf{n}}^\dagger + \kappa\delta\mathbf{e}\hat{\mathbf{n}}^\dagger + (\kappa^{-1} - 1)a_{nr}\hat{\mathbf{n}}\hat{\mathbf{r}}^\dagger + \delta\mathbf{E} + O(\delta^2, \kappa^{-1}\delta), \tag{S6}$$

where

$$\mathbf{A}' = \mathbf{U}\mathbf{V}^\dagger\mathbf{\Lambda}\mathbf{V}\mathbf{U}^\dagger, \tag{S7}$$

is the "normal component" of $\mathbf{A}$. Moreover, we have defined the non-normal mode $\hat{\mathbf{n}} = \hat{\mathbf{u}}_N$, and the reaction mode $\hat{\mathbf{r}}$ via

$$a_{nr}\hat{\mathbf{r}} = \sum_{i=1}^{N-1} \lambda_{iN}\hat{\mathbf{u}}_i \quad \text{where} \quad \|\hat{\mathbf{r}}\| = 1. \tag{S8}$$

The matrix $\mathbf{E}$ and vector $\mathbf{e}$ are given by

$$\mathbf{E} = \frac{1}{\delta}\sum_{i,j=1}^{N-1}(\delta_j - \delta_i)\lambda_{ij}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_j^\dagger, \quad \mathbf{e} = \frac{1}{\delta}\sum_{i=1}^{N-1}\delta_i\lambda_{iN}\hat{\mathbf{u}}_i, \tag{S9}$$

and encapsulate the comparatively small interactions resulting from the anisotropy of the small deviation δ_i from non-normality along the $N-1$ dimensions. We can see from equation (S6) that $\mathbf{A}$ is decomposed into a normal part as well as off-diagonal parts which are dominated by a term of order κ. Since these dominating interactions arise along $\hat{\mathbf{r}}$ and $\hat{\mathbf{n}}$, we can approximately restrict our study to these two dimensions, and project the matrix $\mathbf{A}$ into $(\hat{\mathbf{n}}, \hat{\mathbf{r}})$ to define

$$\mathbf{\Gamma}' \equiv \begin{pmatrix} a_n & \kappa^{-1}a_{nr} \\ \kappa a_{rn} & a_r \end{pmatrix} + O(\delta^2, \kappa^{-1}\delta) \tag{S10}$$

with the matrix coefficients of $\mathbf{\Gamma}'$ given by

$$a_n = \lambda_{NN} \qquad a_{nr} = \sqrt{\sum_{i=1}^{N-1}|\lambda_{iN}|^2} \tag{S11a}$$

$$a_r = a_r^0 + \delta a_r^1 \qquad a_{rn} = a_{nr} + \delta a_{rn}^1 \tag{S11b}$$

$$a_r^0 = \frac{1}{|a_{nr}|^2}\sum_{i,j=1}^{N-1}\lambda_{ij}\lambda_{iN}^*\lambda_{jN} \qquad a_r^1 = \frac{1}{|a_{nr}|^2}\sum_{i,j=1}^{N-1}\frac{\delta_j - \delta_j}{\delta}\lambda_{ij}\lambda_{iN}^*\lambda_{jN} \tag{S11c}$$

$$a_{rn}^1 = \frac{1}{a_{nr}}\sum_{i=1}^{N-1}\frac{\delta_i}{\delta}|\lambda_{iN}|^2 \qquad \lambda_{ij} = -\sum_{k=1}^{N}\lambda_k v_{ik}v_{jk}^*. \tag{S11d}$$

To summarize, we have shown that system (S1) can be reduced to the two-dimensional system

$$\frac{d}{dt}\begin{pmatrix} n_t \\ r_t \end{pmatrix} = \mathbf{\Gamma}'\begin{pmatrix} n_t \\ r_t \end{pmatrix} + \sqrt{2\delta}\begin{pmatrix} \eta_{n,t} \\ \eta_{r,t} \end{pmatrix} \tag{S12}$$

where the non-normal (n) and reaction (r) components are respectively the projection along $\hat{\mathbf{n}}$ and $\hat{\mathbf{r}}$.

2.3 Reduced 2×2 Non-Normal Form

We have demonstrated that any N-dimensional model can be reduced to two effective dimensions whose dynamics is governed by the reduced matrix $\mathbf{\Gamma}'$ (S10), which captures the essential non-normal behaviour. Assume $\mathbf{\Gamma}'$ is diagonalizable with right-eigenvectors $\mathbf{p}_\pm$, so that the eigenbasis matrix is $\mathbf{P} = (\mathbf{p}_+, \mathbf{p}_-)$ and

$$\mathbf{\Gamma}' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}, \qquad \mathbf{\Lambda} = \mathrm{diag}(\lambda_+, \lambda_-). \tag{S13}$$

As for the general N-dimensional case, write the SVD of $\mathbf{P}$ as $\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary and we can write the diagonal matrix $\mathbf{\Sigma}$ composed of the singular values of $\mathbf{P}$ as

$$\mathbf{\Sigma} = \sqrt{\sigma_+\sigma_-}\,\mathbf{K}^{1/2}, \qquad \mathbf{K} = \begin{pmatrix} \kappa & 0 \\ 0 & \kappa^{-1} \end{pmatrix}, \qquad \kappa = \frac{\sigma_+}{\sigma_-} \geq 1, \tag{S14}$$

by convention. Because the eigenvectors are normalized i.e. $\|\mathbf{p}_\pm\| = 1$,

$$\mathbf{P}^\dagger\mathbf{P} \;=\; \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^\dagger \;=\; \begin{pmatrix} 1 & c \\ c^* & 1 \end{pmatrix}, \qquad c \;=\; \langle \mathbf{p}_+, \mathbf{p}_- \rangle, \tag{S15}$$

so that

$$\sigma_\pm = \sqrt{1 \pm |c|}, \qquad \kappa = \sqrt{\frac{1+|c|}{1-|c|}}, \qquad \mathbf{V} = \frac{1}{\sqrt{2}}\begin{pmatrix} e^{i\phi} & -e^{i\phi} \\ 1 & 1 \end{pmatrix}, \quad \phi = \arg(c). \tag{S16}$$

Insert the decomposition into $\mathbf{\Gamma}'$:

$$\begin{aligned} \mathbf{\Gamma}' &= \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger\mathbf{\Lambda}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^\dagger \\ &= \frac{1}{2}\mathbf{U}\mathbf{K}^{1/2}\begin{pmatrix} 1 & e^{-i\phi} \\ 1 & -e^{-i\phi} \end{pmatrix}\mathbf{\Lambda}\begin{pmatrix} e^{i\phi} & -e^{i\phi} \\ 1 & 1 \end{pmatrix}\mathbf{K}^{-1/2}\mathbf{U}^\dagger = \mathbf{U}\mathbf{\Gamma}_u\mathbf{U}^\dagger, \end{aligned} \tag{S17}$$

with

$$\mathbf{\Gamma}_u = \begin{pmatrix} \alpha_+ & \alpha_-\kappa \\ \alpha_-\kappa^{-1} & \alpha_+ \end{pmatrix}, \qquad \alpha_\pm = \frac{1}{2}(\lambda_+ \pm \lambda_-). \tag{S18}$$

Provided the spectrum is real, choose the ordering $0 > \lambda_+ > \lambda_-$ so that $\alpha_- > 0$. Rescaling by α_- and applying $\mathbf{U}^\dagger$ yields the universal form

$$\mathbf{\Gamma} = \frac{1}{\alpha_-}\mathbf{U}^\dagger\mathbf{\Gamma}\mathbf{U} = \begin{pmatrix} -\alpha & \kappa \\ \kappa^{-1} & -\alpha \end{pmatrix}, \qquad \alpha = -\frac{\alpha_+}{\alpha_-}. \tag{S19}$$

Hence only two independent parameters remain:

- α measures the spectral distance from criticality ($\lambda_\pm/\alpha_- = -\alpha \pm 1$, so criticality corresponds to $\alpha = 1$);
- $\kappa \geq 1$ quantifies the degree of non-normality.

Using the time scaling $t \mapsto \alpha_- t$ and the unitary transformation $\mathbf{U}$, the reduced stochastic system becomes

$$\dot{n}_t \;=\; -\alpha n_t + \kappa^{-1} r_t + \sqrt{2\delta}\,\eta_{n,t}, \tag{S20a}$$

$$\dot{r}_t \;=\; -\alpha r_t + \kappa n_t + \sqrt{2\delta}\,\eta_{r,t}, \tag{S20b}$$

where $\eta_{n,t}, \eta_{r,t}$ are standard independent white noises.

Because r_t dominates the transient amplification when $\kappa \gg 1$, one may integrate (S20) to obtain

$$r_t \;=\; \sqrt{2\delta}\int_0^t e^{-\alpha(t-s)}\left[\kappa \sinh(t-s)\,\eta_{n,s} + \cosh(t-s)\,\eta_{r,s}\right] ds. \tag{S21}$$

Equivalently, introducing a single *i.i.d.* noise $\eta_t \sim \mathcal{N}(0,1)$ gives

$$r_t = \sqrt{2\delta}\int_0^t e^{-\alpha(t-s)}\sqrt{\kappa^2 \sinh^2(t-s) + \cosh^2(t-s)}\;\eta_s\, ds, \tag{S22}$$

which is the compact reduced description governing the leading-order non-normal dynamics.

2.4 Summary of results

We showed that any N-dimensional stochastic non-normal linear system (S1) with a diagonalizable interaction matrix $\mathbf{A}$ can be reduced to an effective 2–dimensional model when the following spectral structure holds:

- the eigenbasis matrix $\mathbf{P}$ has $N-1$ singular values that are nearly identical, $|\sigma_i - \sigma_j| \ll 1$ for $i, j = 1, \ldots, N-1$;
- the remaining singular value is much smaller, $\sigma_i \gg \sigma_N > 0$ for $i = 1, \ldots, N-1$.

Under these conditions, the full dynamics is captured by the 2×2 system

$$\dot{n}_t = -\alpha n_t + \kappa^{-1} r_t + \sqrt{2\delta}\, \eta_{n,t}, \tag{S23a}$$

$$\dot{r}_t = -\alpha r_t + \kappa n_t + \sqrt{2\delta}\, \eta_{r,t}, \tag{S23b}$$

where $\eta_{n,t}$ and $\eta_{r,t}$ are independent standard white noises.

The reaction coordinate r_t, which carries the leading $O(\kappa)$ amplification, can be written as

$$r_t = \sqrt{2\delta} \int_0^t e^{-\alpha(t-s)} \left[\kappa \sinh(t-s)\eta_{n,s} + \cosh(t-s)\eta_{r,s}\right] ds. \tag{S24}$$

Introducing a single *i.i.d.* noise $\eta_s \sim \mathcal{N}(0,1)$ gives the compact, equivalent representation

$$r_t = \sqrt{2\delta} \int_0^t e^{-\alpha(t-s)} \sqrt{\kappa^2 \sinh^2(t-s) + \cosh^2(t-s)};\eta_s \, ds, \tag{S25}$$

which fully characterises the dominant non-normal dynamics.

Supplementary Note 3. Early-Warning Signals in Stochastic Non-Normal Linear Systems

In a deterministic dynamical system $\dot{x} = f(x;\mu)$, a bifurcation occurs when a smooth variation of a control parameter μ drives an equilibrium x^* from stable to marginal stability, that is, when an eigenvalue of the Jacobian $J(\mu) = \partial_x f|_{x=x^*}(\mu)$ crosses the imaginary axis. Mathematically, if $\lambda(\mu)$ denotes the right–most eigenvalue of $J(\mu)$, the bifurcation point μ_c is defined by $\mathrm{Re}\,\lambda(\mu_c) = 0$. In stochastic or noisy data, one cannot observe the spectrum directly, so one monitors three early-warning indicators that follow from the linear Ornstein–Uhlenbeck approximation

$$\dot{x}_t = \lambda(\mu)\left(x_t - x^*\right) + \sqrt{2\delta}\,\eta_t, \qquad \eta_t \sim \mathcal{N}(0,1), \qquad \lambda(\mu) \le 0. \tag{S26}$$

E1 **Drift toward a unit root**: As $\mu \to \mu_c$, $\lambda(\mu) \to 0^-$; equivalently the discrete-time autoregressive coefficient $\phi = e^{\lambda(\mu)\Delta t}$ approaches the "unit root" 1, signalling loss of mean-reversion.

E2 **Diverging variance**: The stationary variance satisfies $\langle x^2 \rangle = \delta/|\lambda(\mu)|$, so $\langle x \rangle \to \infty$ as $\lambda(\mu) \to 0^-$.

E3 **Critical slowing down**: The lag-τ autocorrelation is $C(\tau) = e^{\lambda(\mu)\tau}$; its characteristic decay time $\tau_c = 1/|\lambda(\mu)|$ diverges, implying progressively slower recovery from perturbations.

Detecting these three trends $\lambda \to 0^-$, $\langle x^2 \rangle \uparrow$, and $\tau_c \uparrow$ in empirical data in real-time provides quantitative evidence that the system is approaching a bifurcation.

This section shows that a non-normal system, when examined along its reaction coordinate, displays exactly the same signatures as a system approaching a bifurcation, even though its spectrum remains fixed. These effects arise solely from the system's non-normality and are demonstrated using the reduced two-dimensional form (S20).

3.1 Proof: Drift Toward a Unit Root

Starting from the reduced reaction dynamics (S22), we rewrite r_t compactly as

$$r_t \;=\; \sqrt{2\delta}\int_0^t e^{-\Omega(t-s)}\eta_s\,ds, \qquad \Omega(t) \;=\; \alpha t - \frac{1}{2}\ln\Big[\kappa^2\sinh(t)^2+\cosh(t)^2\Big], \tag{S27}$$

so that r_t is an Ornstein-Uhlenbeck process whose mean-reversion rate $\theta(t) := \Omega'(t)$ is itself time-dependent. The system therefore imitates critical behaviour whenever $\theta(t) < 0$ for some t.

The minimum of $\Omega(t)$ occurs at the root of $\theta(t)$ where

$$\theta(t) \;=\; \alpha - \frac{(\kappa^2+1)\cosh t\;\sinh t}{\kappa^2\sinh(t)^2+\cosh(t)^2}\,. \tag{S28}$$

By writing (S28) as a quadratic polynomial in e^{2t}, the equation $\theta(t) = 0$ leads to

$$(\alpha-1)\frac{\kappa^2+1}{4}\,e^{-2t}\Big[e^{4t} - 2\,\frac{\kappa^2-1}{\kappa^2+1}\,\frac{\alpha}{\alpha-1}\,e^{2t} + \frac{\alpha+1}{\alpha-1}\Big] \;=\; 0, \tag{S29}$$

which yields the admissible (real and positive) solution

$$e^{2t^*} \;=\; \frac{\alpha}{\alpha-1}\,\frac{\kappa^2-1}{\kappa^2+1} + \sqrt{\Big(\frac{\alpha}{\alpha-1}\,\frac{\kappa^2-1}{\kappa^2+1}\Big)^2 - \frac{\alpha+1}{\alpha-1}}\,. \tag{S30}$$

A real root exists only when the discriminant is non-negative, which requires

$$\kappa \;\geq\; \kappa_c \;:=\; \alpha + \sqrt{\alpha^2-1} \;=\; \begin{cases} 2(\alpha-1) + O\big((\alpha-1)^{-1}\big), & \alpha \gg 1, \\ 1 + \sqrt{2(\alpha-1)} + O(\alpha-1), & \alpha \gtrsim 1. \end{cases} \tag{S31}$$

Far from criticality ($\alpha \gg 1$), when $\kappa \gg 2\alpha$, expanding (S30) gives

$$t^* \;=\; \frac{1}{2}\ln\Big(\frac{\alpha+1}{\alpha-1}\Big) - \frac{\alpha}{\kappa^2} + O(\kappa^{-4}), \tag{S32}$$

and the minimal mean-reversion rate becomes

$$\Omega_{\min} \;:=\; \Omega(t=t^*) \;=\; \frac{1}{2}\ln\Big[\frac{(\alpha+1)^{\alpha+1}}{(\alpha-1)^{\alpha-1}}\Big] - \ln\kappa + O\big((\alpha/\kappa)^2\big). \tag{S33}$$

Thus $\Omega_{\min} < 0$ (i.e. local loss of mean reversion) whenever

$$\kappa \;\gtrsim\; \frac{(\alpha+1)^{(\alpha+1)/2}}{(\alpha-1)^{(\alpha-1)/2}} \;\approx\; e(\alpha-1)\,, \qquad \alpha \gg 1. \tag{S34}$$

Near criticality ($\alpha \gtrsim 1$), the condition simplifies to

$$\kappa \;\gtrsim\; 2 + \big[1+\ln 2\big](\alpha-1) + O\big((\alpha-1)\ln(\alpha-1)\big). \tag{S35}$$

Hence, the reduced non-normal system (S20) locally experiences pseudo-critical intervals, characterised by $\Omega(t) < 0$, whenever κ exceeds the above bounds. This establishes that criterion **E1** ("approach to the unit root") is reproduced purely by non-normality, without any change in the spectrum, as illustrated in Figure S1.

3.2 Proof: Diverging Variance

The divergence of the variance in a non-normal system follows directly from the reduced reaction dynamics (S22). The time-dependent variance is

$$\big\langle r_t^2\big\rangle \;=\; 2\delta\int_0^t e^{-2\alpha s}\big[\kappa^2\sinh(s)^2+\cosh(s)^2\big]\,ds, \tag{S36}$$

so that the long–time (stationary) variance becomes

$$\langle r^2\rangle \;:=\; \lim_{t\to\infty}\langle r_t^2\rangle \;=\; \frac{\delta\alpha}{\alpha^2-1},\frac{\kappa^2+1}{2}\Big[1 - \frac{\kappa^2-1}{\kappa^2+1},\frac{\alpha^2-1}{\alpha^2}\Big]. \tag{S37}$$

The leading term grows as $O(\kappa^2)$, demonstrating that $\langle r^2\rangle$ increases quadratically with the degree of non-normality. Hence, a non-normal system naturally exhibits the early-warning signal *E2* of increasing variance when κ increases.

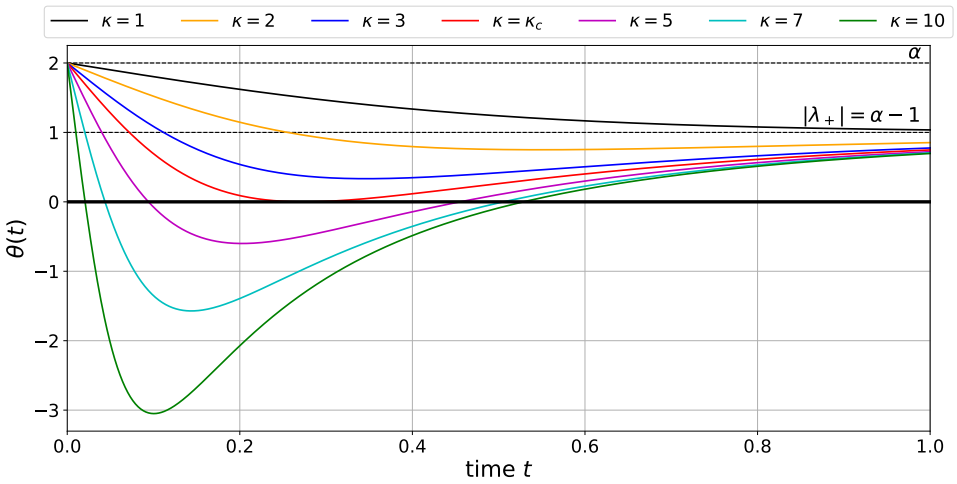

Figure S1: Mean-reversion rate θ_t of the reaction mode for different values of κ.

3.3 Proof: Critical Slowing Down

As with the variance, proving that non-normal systems exhibit critical slowing down is straightforward. Consider the reduced non-normal form along the reaction mode given by (S21), then the covariance of the reaction component of the system is

$$\langle r_t r_{t+\tau} \rangle \;=\; 2\delta \int_0^t e^{-2\alpha t - \alpha\tau} \left[\kappa^2 \sinh(t+\tau)\sinh(t) + \cosh(t+\tau)\cosh(t)\right] \, dt, \tag{S38}$$

from which the asymptotic covariance becomes

$$\mathrm{Cov}[r](\tau) = \lim_{t\to\infty} \langle r_t r_{t+\tau} \rangle = \frac{\delta\alpha}{\alpha^2-1}\frac{\kappa^2+1}{2}\left[1 - \frac{\kappa^2-1}{\kappa^2+1}\frac{\alpha^2-1}{\alpha^2}\right] e^{-\alpha\tau}\left[\cosh(\tau) + \frac{1}{\alpha}\frac{1}{1-\frac{\kappa^2-1}{\kappa^2+1}\frac{\alpha^2-1}{\alpha^2}}\sinh(\tau)\right], \tag{S39}$$

which yields the asymptotic correlation function

$$C(\tau) = \frac{\mathrm{Cov}[r](\tau)}{\langle r^2 \rangle} = e^{-\alpha\tau}\left[\cosh(\tau) + \frac{1}{\alpha}\frac{1}{1-\frac{\kappa^2-1}{\kappa^2+1}\frac{\alpha^2-1}{\alpha^2}}\sinh(\tau)\right]. \tag{S40}$$

We can define the characteristic decay time τ_0 as

$$\tau_0 \;=\; \left[-\frac{d}{d\tau}\ln C(\tau)\Big|_{\tau=0}\right]^{-1} \;=\; \left[\frac{1}{\alpha}\left(\alpha^2 - \frac{1}{1-\frac{\kappa^2-1}{\kappa^2+1}\frac{\alpha^2-1}{\alpha^2}}\right)\right]^{-1} \tag{S41a}$$

$$\Rightarrow \qquad \tau_0 \;=\; \frac{\langle r^2 \rangle}{\delta} \;=\; \frac{\alpha}{\alpha^2-1}\frac{\kappa^2+1}{2}\left[1 - \frac{\kappa^2-1}{\kappa^2+1}\frac{\alpha^2-1}{\alpha^2}\right], \tag{S41b}$$

so that the correlation time diverges for $\kappa \to \infty$, demonstrating that non-normal systems reproduce the early-warning signal *E3* (critical slowing down), as illustrated in Figure S2.

3.4 Summary of Results

In this section, we have shown that, along the reaction mode, non-normal dynamics exhibit the same early-warning signals as a system approaching a bifurcation. These findings can be summarised as follows:

E1 **Drift toward a unit root**: The reaction dynamics can be written as an effective Ornstein-Uhlenbeck process with a time-dependent mean-reverting rate:

$$\dot{r}_t \;=\; -\,\theta(t)\, r_t \;+\; \sqrt{2\delta}\,\eta_t, \qquad \eta_t \overset{i.i.d}{\sim} \mathcal{N}(0,1), \tag{S42}$$

such that for $\kappa \geq \kappa_c$ (S31), there exists a time interval during which $\theta_t < 0$, meaning the system becomes locally critical.

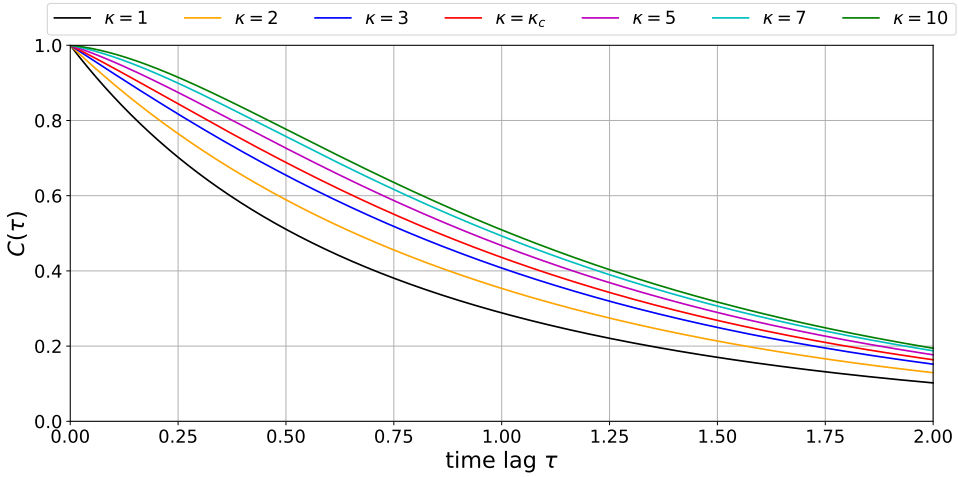

Figure S2: Autocorrelation $C(\tau)$ of the reaction mode r_t for different levels of κ.

E2 **Diverging variance**: The stationary variance of the reaction is given by $\langle r^2 \rangle \sim \delta\kappa^2\alpha/(\alpha^2 - 1)$, which increases with κ and diverges in the limit of large κ.

E3 **Critical slowing down**: The lag-τ autocorrelation function is $C(\tau) = e^{-\tau/\tau_c + O((\alpha\tau)^2)}$, where the characteristic decay time $\tau_0 \sim \kappa^2$ increases with κ, indicating progressively slower recovery from perturbations.

Supplementary Note 4. Degree of Non-Normality

A degree of non-normality quantifies how far a matrix is from being normal. Specifically, it increases as the eigenvectors become increasingly non-orthogonal, as the angle between them progressively narrows. Consequently, the condition number κ of the eigenbasis is a natural measure of non-normality. However, other metrics also exist, such as the Henrici departure from normality and the Kreiss constant.

In this section, we first provide an intuitive and didactic example to illustrate how transient deviations emerge in a non-normal system. In the second part, we demonstrate that the condition number κ serves as an equivalent metric to the Henrici departure from normality.

4.1 Intuition behind Transients in Two Dimensions

Here we analyze system (S1) for the special case where $\delta = 0$ (vanishing noise) in order to gain some intuition on the transient perturbations that push the system away from the equilibrium.

The deterministic solution of the reduced system (S21) is given by

$$\begin{pmatrix} n_t \\ r_t \end{pmatrix} = e^{\mathbf{\Gamma} t} \begin{pmatrix} n_0 \\ r_0 \end{pmatrix}, \quad \text{where } e^{\mathbf{\Gamma} t} = e^{-\alpha t} \begin{pmatrix} \cosh(t) & \kappa^{-1}\sinh(t) \\ \kappa \sinh(t) & \cosh(t) \end{pmatrix} \tag{S43}$$

where (n_0, r_0) defines the initial position. A transient emerges when the matrix norm $\|e^{\mathbf{\Gamma} t}\|$ increases in time. Using the L_2-norm and applying the triangle inequality, the following upper bound to the exponential matrix norm is obtained:

$$\|e^{\mathbf{\Gamma} t}\| \leq e^{-\alpha t}\left[\cosh(t) + \kappa \sinh(t)\right], \quad t > 0. \tag{S44}$$

When the system is normal, $\kappa = 1$, and the the matrix norm is decreasing according to the largest eigenvalue $\lambda_+ = -\alpha + 1$. When the system is non-normal i.e. $\kappa > 1$; a new term appears in the upper bound which involves e^{-t}, leading to a linear combination of two exponential $e^{-(\alpha+1)t}$ and $e^{-(\alpha-1)t}$. When $\kappa \gg 1$, the leading order term is given by $\kappa(e^{\lambda_+ t} - e^{\lambda_- t})$, where $\lambda_\pm = -\alpha \pm 1$. The function $e^{\lambda_+ t} - e^{\lambda_- t}$ is non-monotonic and exhibits a maximum at $t^* = \ln(\lambda_+/\lambda_-)/(\lambda_+ - \lambda_-)$ as shown in Figure S3. Therefore, the exponential matrix norm is not necessarily strictly decreasing as long as κ is large enough. In fact, it is this initial amplification that underlies the emergence of transient bursts in non-normal systems.

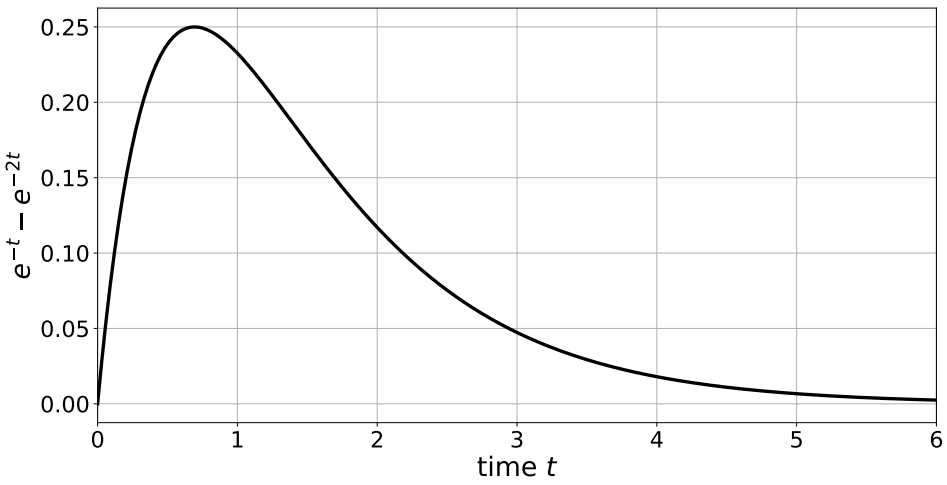

Figure S3: Function $e^{-t} - e^{-2t}$, characterising the essence of the non-monotonic behavior of a non-normal system relaxation to the equilibrium.

4.2 Asymptotic Behavior of the Henrici Departure from Normality

The degree of non-normality of a matrix $\mathbf{A}$ is typically measured by means of Henrici's departure from normality, defined as

$$d_F(\mathbf{A}) = \frac{\sqrt{\|\mathbf{A}\|_F^2 - \sum_i |\lambda_i|^2}}{\|\mathbf{A}\|_F} \tag{S45}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. We now re-write d_F for the case where $\mathbf{A}$ is approximately normal along all directions (that is, all singular values are approximately one) with the exception of the last component (see previous section for details about this assumption). Above, we have expressed the system's non-normality conveniently via its condition number κ. Here we further show that the behavior of d_F is also dominated by κ.

Assuming that the eigenbasis is nearly unitary in all directions except the last one, the transformation matrix $\mathbf{P}$ can be expressed as

$$\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger \quad \text{with} \quad \mathbf{\Sigma} = \mathbf{I} - (1-\kappa^{-1})\mathbf{E}_{NN} \tag{S46}$$

and $\mathbf{E}_{NN}$ is an $N \times N$ matrix that has all its elements equal to zero except for its (N,N)-th component which is equal to one.

We now expand the eigenbasis $\mathbf{P}$ and its inverse in terms of κ

$$\mathbf{P} = \mathbf{Q}_0 - (1-\kappa^{-1})\mathbf{Q}_N \quad \text{and} \quad \mathbf{P}^{-1} = \mathbf{Q}_0^\dagger - (1-\kappa)\mathbf{Q}_N^\dagger \tag{S47}$$

where $\mathbf{Q}_0 = \mathbf{U}\mathbf{V}^\dagger$ is a unitary matrix, and $\mathbf{Q}_N = \hat{\mathbf{u}}_N\hat{\mathbf{v}}_n^\dagger$. Therefore, $\mathbf{A}$ becomes

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} = \mathbf{Q}_0\mathbf{\Lambda}\mathbf{Q}_0^\dagger - (1-\kappa^{-1})\mathbf{Q}_N\mathbf{\Lambda}\mathbf{Q}_0^\dagger - (1-\kappa)\mathbf{Q}_0\mathbf{\Lambda}\mathbf{Q}_N^\dagger - (1-\kappa^{-1})(1-\kappa)\mathbf{Q}_N\mathbf{\Lambda}\mathbf{Q}_N^\dagger. \tag{S48}$$

Since the Frobenius norm is invariant under unitary transformations, we compute it in the basis $\mathbf{Q}_0$,

$$\mathbf{A}_Q = \mathbf{Q}_0^\dagger\mathbf{A}\mathbf{Q}_0 = \mathbf{\Lambda} - (1-\kappa^{-1})\mathbf{V}_N\mathbf{\Lambda} - (1-\kappa)\mathbf{\Lambda}\mathbf{V}_N - (1-\kappa^{-1})(1-\kappa)\mathbf{V}_N\mathbf{\Lambda}\mathbf{V}_N, \tag{S49}$$

where $\mathbf{V}_N = \mathbf{Q}_0^\dagger\mathbf{Q}_N = \hat{\mathbf{v}}_N\hat{\mathbf{v}}_N^\dagger$ is a projection matrix and we have used that $\mathbf{V}_N = \mathbf{V}_N^\dagger$ and $\mathbf{V}_N^2 = \mathbf{V}_N$. The conjugate transpose $\mathbf{A}_Q^\dagger$ reads

$$\mathbf{A}_Q^\dagger = \mathbf{\Lambda}^\dagger - (1-\kappa^{-1})\mathbf{\Lambda}^\dagger\mathbf{V}_N - (1-\kappa)\mathbf{V}_n\mathbf{\Lambda}^\dagger - (1-\kappa^{-1})(1-\kappa)\mathbf{V}_N\mathbf{\Lambda}^\dagger\mathbf{V}_N\,. \tag{S50}$$

The Frobenius norm $\|\mathbf{A}_Q\|_F^2$ is given by

$$\|\mathbf{A}_Q\|_F^2 = \mathrm{Tr}\left[\mathbf{A}_Q^\dagger\mathbf{A}_Q\right]. \tag{S51}$$

Using expressions (S49) and (S50), the Frobenius norm of $\mathbf{A}$ is equal to

$$\|\mathbf{A}\|_F^2 = \|\mathbf{A}_Q\|_F^2 = \|\mathbf{\Lambda}\|_F^2 + \mathrm{Tr}\left[\mathbf{\Lambda}\mathbf{\Lambda}^\dagger \mathbf{V}_N\right] f_1(\kappa) + \mathrm{Tr}\left[\mathbf{\Lambda}\mathbf{V}_N\mathbf{\Lambda}^\dagger \mathbf{V}_N\right] f_2(\kappa) \tag{S52}$$

where we used the cyclic property of the trace and defined

$$f_1(\kappa) = (\kappa - \kappa^{-1})^2 \quad \text{and} \quad f_2(\kappa) = 3\kappa^2(1-\kappa^{-1})^4. \tag{S53}$$

Expression (S52) can further be simplified by noting that

$$\mathrm{Tr}\left[\mathbf{\Lambda}\mathbf{\Lambda}^\dagger \mathbf{V}_N\right] = \sum_{i=1}^N |\lambda_i|^2 |v_{Ni}|^2 = \langle |\lambda|^2 \rangle_w, \tag{S54a}$$

$$\mathrm{Tr}\left[\mathbf{\Lambda}\mathbf{V}_N\mathbf{\Lambda}^\dagger \mathbf{V}_N\right] = \left|\sum_{i=1}^N \lambda_i |v_{Ni}|^2\right|^2 = |\langle \lambda \rangle_w|^2 . \tag{S54b}$$

Additionally, the Frobenius norm of $\mathbf{\Lambda}$ is simply given by

$$\|\mathbf{\Lambda}\|_F^2 = \sum_{i=1}^N |\lambda_i|^2 = N\langle |\lambda|^2 \rangle. \tag{S55}$$

Putting all the above together, we obtain the following expression for Henrici's departure from normality:

$$d_F(\mathbf{A}) = \sqrt{\frac{\langle |\lambda|^2 \rangle_w f_1(\kappa) + |\langle \lambda \rangle_w|^2 f_2(\kappa)}{N\langle |\lambda|^2 \rangle + \langle |\lambda|^2 \rangle_w f_1(\kappa) + |\langle \lambda \rangle_w|^2 f_2(\kappa)}}. \tag{S56}$$

For κ close to 1, that is for the system close to normal, we define $\epsilon = \kappa - 1 \gtrsim 0$ and obtain to first order

$$d_F(\mathbf{A}) = 2\sqrt{\frac{\langle |\lambda|^2 \rangle_w}{N\langle |\lambda|^2 \rangle}}\, \epsilon + O\left(\epsilon^3\right), \tag{S57}$$

showing that Henrici's departure from normality decays linearly to zero as a function of $\kappa - 1$ as anticipated. By contrast, when $\mathbf{A}$ is strongly non-normal, that is $\kappa \gg 1$, the Henrici departure from normality becomes

$$d_F(\mathbf{A}) = 1 - \kappa^{-2} \frac{N\langle |\lambda|^2 \rangle}{\langle |\lambda|^2 \rangle_w + 3|\langle \lambda \rangle_w|^2} + O(\kappa^{-3}) \tag{S58}$$

and hence approaches the maximum non-normality $d_F = 1$ as $1 - O(\kappa^{-2})$.

Supplementary Note 5. Empirical Study & Simulation

Here we describe how to measure the non-normal component of a matrix $\mathbf{A}$ governing system (S1).

5.1 Measuring Non-Normal and Reaction Mode From Observations

Let us consider the discretized version of (S1) given by

$$\mathbf{x}_{t+1} = \mathbf{A}'\mathbf{x}_t + \sqrt{2\delta\Delta t}\, \boldsymbol{\eta}_t \quad \text{where } \boldsymbol{\eta}_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{I}) \quad \text{and } \mathbf{A}' = \mathbf{I} + \Delta t \mathbf{A} \tag{S59}$$

and we assume that its states $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$ are observed at equidistant times. To estimate the matrix $\mathbf{A}'$, we stack these observations into matrices $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ and $\mathbf{Y} = (\mathbf{x}_2, \ldots, \mathbf{x}_T)$. The interaction matrix $\mathbf{A}'$ can be extracted by solving the least squares problem $\mathbf{Y} = \mathbf{A}'\mathbf{X}$ with solution $\widehat{\mathbf{A}'} = \mathbf{Y}\mathbf{X}^+$, where $\mathbf{X}^+$ is the pseudo-inverse of $\mathbf{X}$. In the limit where T is very large, this allows for a faithful reconstruction of the true matrix $\mathbf{A}$. However, in practice, especially when the ratio of N/T is comparatively high, the estimate $\hat{\mathbf{A}}$ is error prone.

In principle, once $\mathbf{A}$ is estimated, one can then estimate the non-normal component $\hat{\mathbf{n}}$ and reaction component $\hat{\mathbf{r}}$ via SVD of $\hat{\mathbf{A}}$ (see Section 2). However, estimation errors of $\mathbf{A}$ will further be amplified by the SVD, leading to imprecise estimations of the non-normal and reaction components. These errors are exacerbated when the

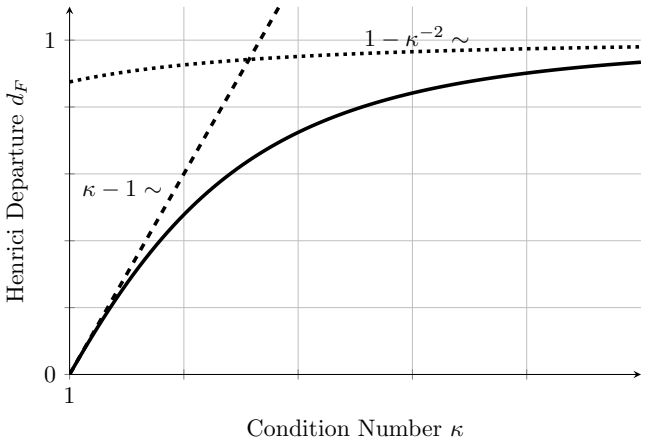

Figure S4: Henrici departure from normality as a function of the condition number κ, where the dashed straight line represent the asymptotic behavior close to normality ($0 < \kappa - 1 \ll 1$); the dotted line gives the asymptotic behavior when the system tends to be highly non-normal $\kappa \gg 1$.

condition number κ of the matrix is high, making the matrix ill-conditioned and sensitive to small perturbations in the input.

Therefore, we propose to estimate the non-normal and reaction component via an optimization-based approach that is more robust to numerical errors. Noting that $\mathbf{A}' \sim \kappa \hat{\mathbf{r}}\hat{\mathbf{n}}^\dagger + O(1)$ in highly non-normal systems, we reformulate the problem as

$$\hat{\mathbf{r}}, \hat{\mathbf{n}} = \underset{\mathbf{v},\mathbf{w}\in\mathbb{S}^{n-1}}{\arg\max} \langle \mathbf{v}, \hat{\mathbf{A}}'\mathbf{w} \rangle \;\text{ s.t. }\; \langle \mathbf{v}, \mathbf{w} \rangle = 0 \tag{S60}$$

where $\mathbb{S}^{n-1}$ is the unit sphere in n dimensions. Problem (S60) is a standard quadratic convex optimization problem and readily solvable with standard gradient descent methods.

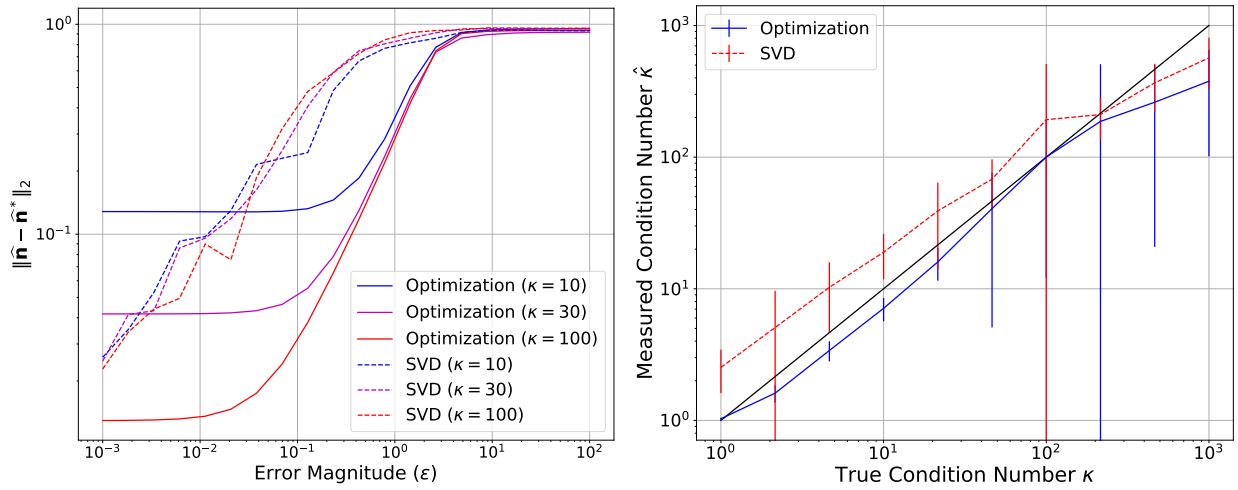

Figure S5: (Left) L2 norm between true non-normal mode $\hat{\mathbf{n}}$ and estimated non-normal mode $\hat{\mathbf{n}}^*$ for different levels of noise $\epsilon = \|\mathbf{E}\|_2/\|\mathbf{A}\|_2$. The solid lines represent errors from estimates via the optimization procedure (S60) while dashed lines indicate errors from estimates via SVD. The maximum error is one, which coincides with an estimate that is fully orthogonal to the ground truth one. (Right) Measured condition number $\hat{\kappa}$ as a function of the true condition number $\kappa := \kappa(\mathbf{P})$ of the matrix $\mathbf{P}$ diagonalizing the matrix $\mathbf{A}$. We compare our method (S61), which bypasses the need to directly measure κ via singular values extracted from the SVD, to the method using the SVD.

We now compare the performance of optimization (S60) with the direct calculation of $\hat{\mathbf{n}}$ via SVD. To this end, we generate noisy matrices $\mathbf{A} + \mathbf{E}$ where $\mathbf{A}$ is the known ground-truth and $\mathbf{E}$ is a Gaussian matrix. We scale the matrix such that $\epsilon = \|\mathbf{E}\|/\|\mathbf{A}\|$ is the amplitude of the perturbation, allowing us to control the relative level of noise. We then measure $\hat{\mathbf{n}}$ from $\mathbf{A} + \mathbf{E}$ both via (S60) and via SVD, and measure the difference between

true non-normal mode $\hat{\mathbf{n}}$ and estimated non-normal mode $\hat{\mathbf{n}}^*$ with the $L2$-norm $||\hat{\mathbf{n}} - \hat{\mathbf{n}}^*||$. Plotting this error as a function of the noise scale ϵ in Figure S5 (left), we notice that, especially for large noise and strongly non-normal systems κ, our optimization estimates are more robust, demonstrating resilience to noise levels up to an order of magnitude greater.

In principle, once the matrix $\hat{\mathbf{A}}$ has been measured by calibrating a VAR(1) process, one could proceed to estimate its non-normality by first determining the transition matrix $\mathbf{P}$ that diagonalizes $\hat{\mathbf{A}}$, and subsequently calculating the SVD of $\mathbf{P} = \mathbf{U}\Sigma\mathbf{V}^\dagger$ to identify the non-normal mode $\hat{\mathbf{n}} \sim \mathbf{u}_n$ and reaction mode $\hat{\mathbf{r}} \sim \sum_{i=1}^{n-1} \mathbf{u}_i$. However, as detailed above, this approach is highly sensitive to noise, especially when $\mathbf{A}$ is non-normal. Therefore, we note that, from (S20), the condition number κ can be estimated directly from $\hat{\mathbf{r}}$ and $\hat{\mathbf{n}}$ via

$$\kappa = \sqrt{\left| \frac{\langle \hat{\mathbf{r}}, \hat{\mathbf{A}}\hat{\mathbf{n}} \rangle}{\langle \hat{\mathbf{n}}, \hat{\mathbf{A}}\hat{\mathbf{r}} \rangle} \right|}. \tag{S61}$$

In Figure S5 (right), we compare the difference in the estimated κ when using (S61) versus directly calculating it from the SVD, and find that our approach is generally more accurate up to $\kappa \approx 200$. Both methods tend to underestimate κ as the system's non-normality becomes extremely large.

5.2 Measuring Quasi-Deterministic Loops

We identify quasi-deterministic cycles and their amplitude as follows.

First, we project the discretized dynamics $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_T$ onto the normal-mode $\hat{\mathbf{n}}$ and reaction-mode $\hat{\mathbf{r}}$, giving rise to a two-dimensional trajectory $(n_1, r_1), (n_2, r_2), \ldots, (n_T, r_T)$. For a given starting point (n_t, r_t) at time t, we then ask what is the first time for which the trajectory returns back to this point. To this end, we draw two circles C_I and C_O around (n_t, r_t), one of radius $5B$ (inner circle C_I) and one of radius $10B$ (outer circle C_O) where B is the (empirically measured) standard deviation of the fluctuations (providing the unit scale for these circles). We define the departure from (n_t, r_t) as the first time $\tau_1 > t$ at which the trajectory leaves C_O, and the return time as the first time $\tau_2 > \tau_1$ at which the trajectory returns back into circle C_I. The trajectory $\{(n_t, r_t)\}_{t=\tau_1}^{\tau_2}$ then defines one cycle, and we measure its amplitude as its diameter.

We then measure all cycles in terms of the above cycle trajectories by iterating over all possible starting points along a given trajectory.

5.3 Dynamical Interpretation of Quasi-Deterministic Loops

From system (S20) we deduce that $dr_t \approx (\kappa n_t - \alpha_r r_t)dt$ and distinguish between the expansion phase, $dr_t/r_t > 0$, during which the transient grows, and contraction phase $dr_t/r_t < 0$ when the system relaxes back to equilibrium. The non-normal mode, meanwhile, is governed by the mean-reversion process $dn_t \approx -\alpha_n n_t dt + \sigma dW_t^n$ where we do not ignore the stochastic component because it is not necessarily negligible. Taken together, this yields

$$r_t \approx \kappa \int_0^t e^{-\alpha_r(t-s)} n_s ds = \kappa\sigma \int_0^t e^{-\alpha_r(t-s)} \int_0^s e^{-\alpha_n(s-u)} dW_u^n \, ds \tag{S62}$$

and suggests that the Quasi-Deterministic loops arise from the exponential smoothing of the non-normal mode along the reaction mode, with the reaction dynamics driven by a mean-reverting OU process.

Supplementary Note 6. Relaxing the Assumptions Regarding Singular Values

Throughout the manuscript, we have assumed that $\sigma_1 \approx \sigma_2 \approx \ldots \approx \sigma_{N-1} \approx 1 \gg \sigma_N = \kappa^{-1}$. The low-rank hypothesis in complex networks supports this assumption, stating that many high-dimensional nonlinear systems can be reduced to lower-dimensional linear systems. In numerous sociotechnical network systems, only a "few" singular values from a Singular Value Decomposition (SVD) of the network are large, while most middle singular values are "almost" the same, and a final "few" are much smaller. In our framework, these smaller singular values correspond to multiple non-normal modes driving transient deviations.

To further clarify this point, we consider the generic case where the singular values are independently Gaussian distributed, i.e., $\sigma_i \overset{\text{i.i.d.}}{\sim} |\mathcal{N}(0,1)|$ where the absolute value ensures positivity. It is well known that the

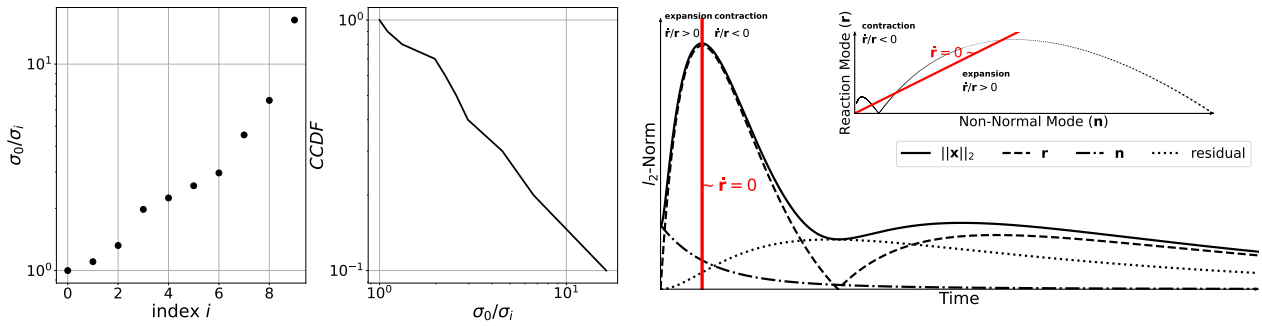

Figure S6: The plot on the right is the same as Figure 1 in the main manuscript, but generated for an $N = 10$-dimensional system in which all singular values are sampled from $|\mathcal{N}(0, 1)|$. As evidenced in the left and center plot, the singular value ratios are Cauchy-distributed, and the non-normality is chiefly driven by σ_9/σ_{10}, which plays a dominant role in shaping the transient response. as seen in the right most plot.

ratio between any two such singular values follows a Cauchy distribution with the probability density function (PDF)

$$p(x) \sim \frac{1}{x^{1+\alpha}}, \quad \text{with } \alpha = 1. \tag{S63}$$

This distribution implies undefined variance and mean, mathematically, and captures the divergence where the largest singular values dominate significantly.

For instance, without loss of generality, let us order the singular values as

$$\sigma_0 \geq \cdots \geq \sigma_n \geq \cdots \geq \sigma_N > 0,$$

then the ratio σ_0/σ_n grows linearly with n.

To illustrate that this generic setup gives rise to non-normal transients described by our formalism, we sample singular values according to the above procedure for $N = 10$, and assume the same eigenvalues as in the main manuscript ($\lambda_i = -i$). The left-most and center plot of Figure S6 confirm that the ratios of these singular values are indeed power law distributed according to (S63). The right-most plot shows the same graph as Figure 1 in our main article but with this adjusted set of singular values. Evidently, even with anisotropy in the singular values, highly non-normal and hierarchical networks do exhibit transients that arise within a reduced subspace of the system.

Therefore, while our study adopts a simplified scenario in which the first $N - 1$ singular values are approximately equal and the last is markedly smaller, the key result, transient dimensionality reduction, holds broadly across diverse non-normal linear systems.