Manifold learning for parameter reduction, Supplementary Materials

Alexander Holiday[a,*], Mahdi Kooshkbaghi[b,*], Juan M. Bello-Rivas[b], C. William Gear[a], Antonios Zagaris[c,**], Ioannis G. Kevrekidis[a,b,d,**]

[a]*Chemical and Biological Engineering, Princeton University, USA*
[b]*The Program in Applied and Computational Mathematics (PACM), Princeton University, USA*
[c]*Wageningen Bioveterinary Research, Wageningen UR, The Netherlands*
[d]*Department of Chemical and Biomolecular Engineering, John Hopkins University, USA*

1. Mathematical setting

We start our discussion with a simple framework covering all models encountered in this manuscript; a more general setting and some theoretical comments regarding sloppiness are presented later, in Section 10 herein. Specifically, we consider multivariable vector functions $\mathbf{x}(t|\mathbf{p}) \in \mathrm{R}^D$, with D finite but arbitrary. Typically, $\mathbf{x} = [x_1, \ldots, x_D]$ is only known implicitly as the solution to some problem, e.g. to an initial-value ODE problem with $t > 0$ representing time and $\mathbf{p} = (p_1, \ldots, p_M) \in \Theta \subset \mathrm{R}^M$ model parameters. Also typically, one monitors *not* the entire trajectory $\chi(\mathbf{p})$, for all $t > 0$, but merely a number of *functionals* on it, $f_1, \ldots, f_N$. We call the N−tuple $\mathbf{f}(\mathbf{p}) = [f_1(\chi(\mathbf{p})), \ldots, f_N(\chi(\mathbf{p}))] \in \mathrm{R}^N$ model *output* or *response*. In our discussion, these functionals concretely correspond to a state variable x_d observed at specific preset times, i.e. $\mathbf{f}(\mathbf{p}) = [x_d(t_1), \ldots, x_d(t_N)]$ for certain times $t_1, \ldots, t_N$. As $\mathbf{p}$ ranges over Θ, $\mathbf{f}(\mathbf{p})$ traces out a (generically) M−dimensional manifold called the *model manifold* $\mathcal{M}$. In general, we will understand that manifold as

$$\mathcal{M} = \mathrm{graph}(\mathbf{f}) = \{(\mathbf{p}, \mathbf{f}(\mathbf{p})) \,|\, \mathbf{p} \in \Theta\} \subset \mathrm{R}^{M+N}.$$

---

[*]These two authors contributed equally to this work.
[**]Corresponding author
Email addresses: `antonios.zagaris@asml.com` (Antonios Zagaris), `yannisk@jhu.edu` (Ioannis G. Kevrekidis)

The projection of $\mathcal{M}$ on $\mathrm{R}^M \times \{0\}^N$ is injective and yields the *input domain* Θ. The projection on $\{0\}^M \times \mathrm{R}^N$, on the other hand, is not guaranteed to be injective even in the important case $N > M$; we consider such a *non-invertible* model in the main text. During our discussion, $\mathcal{M}$ will be endowed with a metric suggested by the application (in a sense by us, the observers, and our measurement capabilities), which will turn it into a Riemannian manifold. For the time being, we postpone a discussion of this important subject to a later section, where we discuss diffusion maps (DMAPS) and their relation to the Laplace-Beltrami operator.

2. Toy example: singular perturbation

The system discussed in the main text is a supremely basic prototype for singularly perturbed dynamical systems,

$$\frac{dx}{dt} = 2 - x - y, \tag{1}$$

$$\frac{dy}{dt} = \frac{1}{\varepsilon}(x - y). \tag{2}$$

For $\varepsilon \ll 1$, trajectories are attracted to the slow manifold which, at leading order, has the form $y = x$. In our numerical experiments, we kept the initial condition for the slow variable at $x_0 = -1$, fixed $t_1 = 0.5$, $t_2 = 1.0$ and $t_3 = 1.5$ and sampled the input space (ε, y_0) for small parameter values ($10^{-3} \leq \varepsilon \leq 1$) and fast variable initializations ($3 \leq y_0 \leq 5$). The map from input to output space assumes the form

$$[3, 5] \times [10^{-3}, 1] \supset \Theta \ni (\varepsilon, y_0) \mapsto \mathbf{f}(\varepsilon, y_0) = [y(t_1), y(t_2), y(t_3)] \in \mathbb{R}^3, \tag{3}$$

with a closed-form expression easily derived by quadrature on (2).

The 3×2 Jacobian of the transformation is

$$\boldsymbol{Df} = \boldsymbol{J}(\varepsilon, y_0, t) = \begin{bmatrix} \frac{\partial y(t_1)}{\partial \varepsilon} & \frac{\partial y(t_1)}{\partial y_0} \\ \frac{\partial y(t_2)}{\partial \varepsilon} & \frac{\partial y(t_2)}{\partial y_0} \\ \frac{\partial y(t_3)}{\partial \varepsilon} & \frac{\partial y(t_3)}{\partial y_0} \end{bmatrix}, \tag{4}$$

and its singular value decomposition (SVD) yields the transformation rank and, eventually, the model manifold dimensionality. In Fig. 1, we have plotted the singular values of the transformation against ε for the trajectory

initialized at $(x_0, y_0) = (-1, 4)$. At larger values of ε, the transformation has rank two (evidenced by two $\mathcal{O}(1)$ singular values) and consequently the model manifold is, in principle, two-dimensional. As we decrease ε, however ($\varepsilon < 10^{-1}$), the smallest singular value clearly approaches zero. One may impose a threshold for that value, below which the matrix is effectively rank deficient and the model manifold one-dimensional. For even smaller $\varepsilon-$values, close to the boundary $\varepsilon = 0$, the model manifold becomes effectively zero-dimensional as also evidenced by our Diffusion Maps calculations.
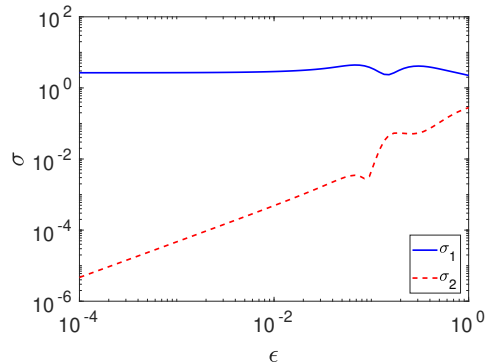

Figure 1: Singular values σ of the transformation map from the parameter space to the model response manifold for the singular perturbation prototype (2) and the trajectory starting at the initial condition $(x_0, y_0) = (-1, 4)$. Model output y is sampled at the time instants $t_1 = 0.5$, $t_2 = 1.0$ and $t_3 = 1.5$.

3. Spectral geometry, diffusion maps, and multi-criteria optimization

3.1. The Laplace–Beltrami operator and diffusion maps

A Riemannian manifold (Θ, g) is a smooth m-dimensional manifold $\Theta \subseteq \mathbb{R}^n$ endowed with a metric g. We can regard the metric as a device that allows us to measure distances and angles on Θ. At each point of Θ, the Riemannian metric g can be represented as a symmetric and positive definite matrix. If u is a smooth real-valued function on Θ, the Laplace–Beltrami operator Δ [8] is a linear operator given, in local coordinates $(z_1, \ldots, z_m)$, by the formula

$$\Delta u = \frac{1}{\sqrt{\det g}} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial}{\partial z_i} \left(\sqrt{\det g} \, (g^{-1})_{ij} \frac{\partial u}{\partial z_j} \right).$$

3

Here, $\det g$ is the determinant of the matrix associated with the Riemannian metric and g^{-1} is the corresponding inverse matrix. Intuitively, the eigenfunctions of the Laplace–Beltrami operator determine how heat propagates on the manifold Θ.

Let $S_N = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ be a set of points sampled from an arbitrary probability distribution on the manifold Θ. The computational complexity of traditional approximation schemes for the Laplace–Beltrami operator, such as finite differences or finite elements, scales exponentially in the dimension n. By contrast, the diffusion maps method (DMAPS; to be discussed below) is a data-driven approximation of Δ with a computational complexity that scales quadratically in the number of samples N.

To approximate Δ in the DMAPS sense, fix a scale $\epsilon > 0$ and consider the affinity matrix $A \in \mathbb{R}^{N \times N}$ with entries

$$A_{ij} = \exp\left\{ -\frac{\|x_i - x_j\|^2}{2\epsilon} \right\}.$$

For each $i = 1, \ldots, N$, set further $q_i = \sum_j A_{ij}$ and introduce the matrix $W \in \mathbb{R}^{N \times N}$ with entries

$$W_{ij} = \frac{A_{ij}}{q_i \, q_j}.$$

We define the (random walk) graph Laplacian L as [3]

$$L = I - D^{-1}W,$$

where $I \in \mathbb{R}^{N \times N}$ is the identity matrix and $D \in \mathbb{R}^{N \times N}$ is the diagonal matrix with diagonal entries

$$D_{ii} = \sum_j W_{ij}.$$

Note that L is determined by ϵ and the set of point-samples S_N. Let us now represent the smooth function $u \colon \Theta \to \mathbb{R}$ by a vector $U \in \mathbb{R}^N$ with components $U_i = u(x_i)$, for $i = 1, \ldots, N$. It is known [5] that

$$\lim_{\epsilon \to 0} \epsilon^{-1} \lim_{N \to \infty} \sum_j L_{ij} U_j = \tfrac{1}{2} \Delta u(x_i).$$

More precisely, it can be shown [9] that the choice

$$\epsilon = C \, N^{\frac{-2}{6+m}}, \tag{5}$$

4

with $C > 0$ a constant depending on the geometry of M but not on N, leads to the minimal error bound

$$\epsilon^{-1} \sum_j L_{ij} U_j = \tfrac{1}{2} \Delta u(x_i) + \mathcal{O}(N^{-\frac{2}{m+6}}) \quad \text{as } N \to \infty. \tag{6}$$

For example, for a 2D manifold ($m = 2$), (5)–(6) establish that the error decays *slowly* as $N^{-1/4}$.

Let $\psi_0, \psi_1, \psi_2, \dots$ be the eigenfunctions of Δ corresponding to the eigenvalues $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots < +\infty$ and let $d \geq m$. The diffusion map $\Psi_\epsilon \colon \mathbb{R}^n \to \mathbb{R}^d$, defined by $\Psi_\epsilon = (e^{\lambda_1 \epsilon} \psi_1, \dots, e^{\lambda_d \epsilon} \psi_d)$, is used for dimensionality reduction in manifold learning. The leading eigenfunction ψ_0 is not included in the definition of Ψ_ϵ, because it is always a constant function and carries no information on (Θ, g).

In practice, we calculate the spectral decomposition of the Laplace–Beltrami operator Δ by the eigenvalues and eigenvectors of the graph Laplacian L. However, in cases in which the dimensionality of the problem is sufficiently low, using the finite element method [2] is faster and more accurate. This additional accuracy is crucial, if we are interested in computing a large number of eigenfunctions. To demonstrate the method, we consider the simple example of a rectangle, $\Theta = [0, 1] \times [0, \ell] \subset \mathbb{R}^2$, endowed with the Euclidean metric. In Fig. 2, we show the first few eigenfunctions of the Laplace–Beltrami operator, computed using both diffusion maps and a high-accuracy finite element method for fixed aspect ratio $1/\ell = 2$. The exact spectral decomposition is given by the family of eigenfunctions $\psi_{ij}(x, y) = \cos(i\pi x) \cos(j\pi y/\ell)$ and eigenvalues $\lambda_{ij} = \pi^2(i^2 + j^2/\ell^2)$, for $i, j = 0, 1, 2, \dots$. On one hand, the spectral decomposition of the graph Laplacian was computed using a set of $N = 10^4$ points, sampled uniformly from Θ, with $\epsilon = 10^{-1}$. On the other hand, the corresponding spectral decomposition of the Laplace-Beltrami operator was computed using the FEniCS finite element library [6] with quadratic Lagrange elements on an adaptively refined mesh.

3.2. Interpretation in terms of continuum mechanics

Let Θ and $\mathcal{M}$ be two smooth manifolds. Given a smooth map $\mathbf{f} \colon \Theta \to \mathcal{M}$, we consider the deformation gradient, given by the Jacobian $D\mathbf{f}$, and the associated Green deformation tensor [7] (also known as the right Cauchy–Green tensor), defined by $D\mathbf{f}^T D\mathbf{f}$. The deformation tensor can be regarded as a tool quantifying the change in $\mathbf{f}(\Theta) \subseteq \mathcal{M}$ that results from a change in Θ. We elaborate on this idea in what follows.
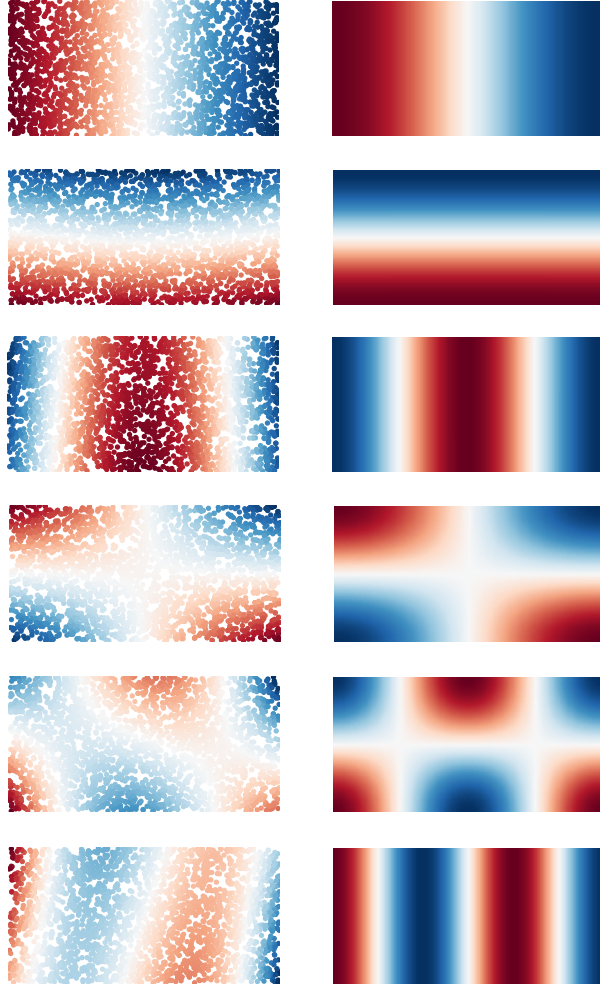
5

Figure 2: Comparison of some eigenfunctions of the Laplace–Beltrami operator for a rectangle with aspect ratio 2. First column shows eigenfunctions obtained by applying DMAPS on 10^4 points sampled uniformly and with $\epsilon = 10^{-1}$. Second Column shows the corresponding eigenfunctions obtained by the finite element method, using an adaptive scheme with Lagrange elements of degree 2.

For the sake of concreteness, we look at a particular example: the analytically solvable, linear, singular perturbation example studied in the paper,

$$\begin{cases} \dot{x} = 2 - x - y, \\ \varepsilon \dot{y} = x - y, \\ x(0) = -1, \\ y(0) = y_0. \end{cases} \tag{7}$$

Using the solution of (7) evaluated at $t_1 = 1/2$, $t_2 = 1$, and $t_3 = 3/2$, we obtain a diffeomorphism $\mathbf{f} \colon \Theta \to \mathcal{M}$ given by

$$(\varepsilon, y_0) \mapsto \mathbf{f}(\varepsilon, y_0) = [y(t_1|\varepsilon, y_0), y(t_2|\varepsilon, y_0), x(t_3|\varepsilon, y_0)].$$

Here, $\Theta = [3, 5] \times [0.00225, 0.1] \subset \mathbb{R}^2$ and $\mathcal{M} = \mathbf{f}(\Theta) \subset \mathbb{R}^3$. In our current context, the manifold Θ represents *inputs* for the transformation $\mathbf{f}$ and, correspondingly, $\mathcal{M}$ will be the manifold of *outputs* of $\mathbf{f}$. A perturbation of an input x in a neighborhood $V \subseteq \Theta$ by a tangent vector $\Delta x \in T_x V$, with $\|\Delta x\| = h > 0$, results in a change in the output of size

$$\|\mathbf{f}(x+\Delta x)-\mathbf{f}(x)\|^2 = \|D\mathbf{f}(x)\,\Delta x\|^2+o(\varepsilon) = (\Delta x)^T(D\mathbf{f}^T D\varphi)\,\Delta x+o(h), \quad \text{as } h \downarrow 0.$$

Therefore, we can endow Θ with the metric determined by the deformation tensor $g = D\mathbf{f}^T D\mathbf{f}$. Indeed, g measures how close to each other are the images, under $\mathbf{f}$, of the points $x = (\varepsilon, y_0)$ and $x + \Delta x = (\varepsilon + \Delta \varepsilon, y_0 + \Delta y_0)$. It is then natural to study the eigenfunctions of the Laplace–Beltrami operator on (Θ, g). These eigenfunctions yield another parametrization [1] of Θ that reflects the sensitivity of the outputs to changes in the inputs. Some of the relevant eigenfunctions are plotted in Fig. 3.

3.3. Thoughts on multi-criteria optimization

The curve $\gamma \colon [0, 1] \to \Theta$ depicted (dashed) in Fig. 4 is a level set of the first non-constant eigenfunction ψ_1 of the Laplace–Beltrami operator on (Θ, g), i.e.

$$\psi_1(\gamma(t)) = c, \text{ for some constant } c \in \mathbb{R} \text{ and all } t \in [0, 1].$$

The image of γ under ψ_{15}, which is not a harmonic of ψ_1 (cf. Fig. 3) is increasing monotonically function *along* the curve; indeed, $\mathrm{d}\psi_{15}(\gamma(t))/\mathrm{d}t$ neither changes sign nor vanishes for any $t \in [0, 1]$. As a result, we can parameterize
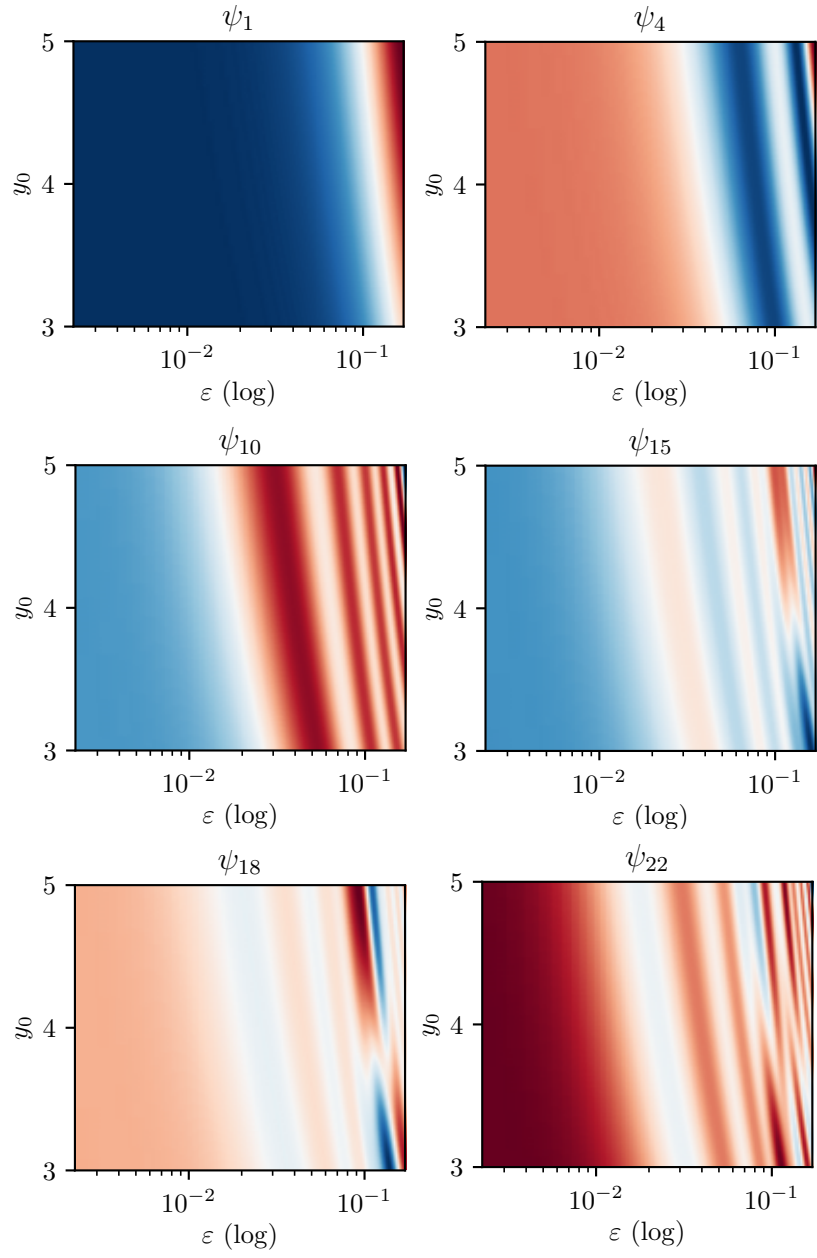
Figure 3: Some eigenfunctions of the singularly perturbed problem. The eigenfunctions ψ_4 and ψ_{10} (top) are harmonics of the first non-constant eigenfunction ψ_1 (also top), whereas the rest (bottom) are not.

γ using the values of ψ_{15}. The fact that level sets of ψ_1 can be parameterized using another eigenfunction ψ_k has obvious applications to multi-criteria optimization [4]. In particular, if our optimization criterion follows a lexicographic ordering, in which we first seek an optimal level set γ of ψ_1, the parametrization of γ in terms of ψ_k is a natural way to subsequently seek an optimal point *on* γ.
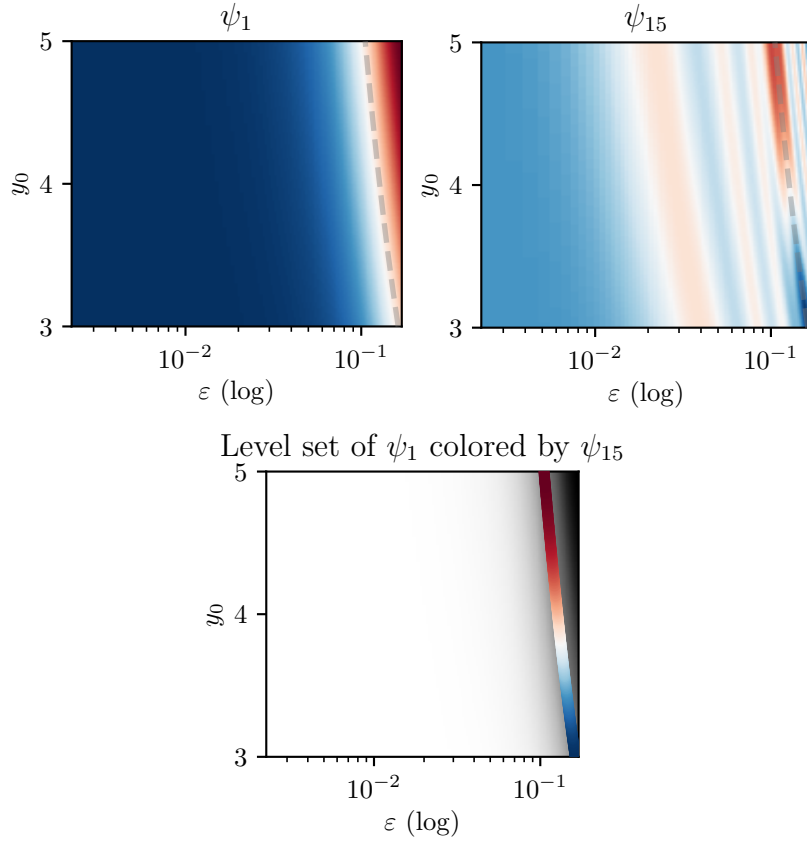

Figure 4: Parameterization of the first non-constant eigenfunction of the singularly perturbed problem using another eigenfunction. Left: a level set of ψ_1 (dashed) overimposed on a heatmap of ψ_1, the first non-constant eigenfunction. Middle: same but with a heatmap of ψ_{15}, the first eigenfunction that is not a higher harmonic of ψ_1. Plainly, ψ_{15} is monotonic on that level set. Right: a concatenation of the top panels. The gray heatmap covers the entire panel and represents ψ_1, whereas the colored one is confined in a neighborhood of the level set and represents ψ_{15}.

9

4. Regular perturbation

The system under investigation is $\dot{x} = \varepsilon x^3 - x$, with $\varepsilon \ll 1$ a small parameter and x the system state. Using that this differential equation is separable and restricting ourselves to $x \geq 0$, we can write the analytical solution

$$x(t|\varepsilon, x_0) = \left(\varepsilon + e^{2t} \left(\frac{1}{x_0^2} - \varepsilon \right) \right)^{-1/2}. \tag{8}$$

We are interested in the behavior of the system in the adiabatic limit $\varepsilon \to 0$. The limiting solution can be found simply by omitting the cubic term from the differential equation, giving us

$$\lim_{\varepsilon \to 0} x(t) = x_0 e^{-t}. \tag{9}$$

The key distinction between this system and the singularly perturbed model presented above is that the initial conditions *does* influence the trajectories, even in the limiting case of small ε. Varying both ε and x_0 for large ε−values, one obtains the model response 8 that depends non-trivially on *both* ε and x_0. For small values of ε, on the other hand, trajectories converge to the limiting solution (9), where ε does not affect the model response. This is precisely what we see in the model manifold depicted in Fig. 5 of the main text. For large values of ε, the model manifold is two dimensional and variations in both x_0 and ε affect the model response. At smaller values of ε, the model manifold converges to a 1D object parameterized by x_0. The exact form of that object is simply the ray $f(x_0) = x_0(e^{-t_1}, e^{-t_2}, e^{-t_3})$, with $x_0 > 0$, bearing no dependence on ε.

To create Fig. 5 in the paper, we fixed $t_1 = 0.25$, $t_2 = 1.0$ and $t_3 = 1.75$ and set the model response to $\mathbf{f}(x_0, \varepsilon) = [x(t_1|x_0, \varepsilon), x(t_2|x_0, \varepsilon), x(t_3|x_0, \varepsilon)] \in \mathbb{R}^3$, with $x(\cdot|x_0, \varepsilon)$ given by (8). We then drew 2500 points from the rectangle $x_0 \in [1.0, 2.5]$ and $\log(\varepsilon) \in [-3, -1]$, uniformly in these two parameters, and performed DMAPS on the input–output combinations with $\epsilon = 5.0$ as kernel scale. Coloring parameter space by the resulting ϕ_1 values gives Fig. 5 in the paper, showing that ϕ_1 "discovers" the regularly perturbed nature of the problem in a data-driven manner.

5. The ABC system

In this section, we identify the singularly perturbed regime for the dynamics associated with the reaction scheme

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} B \overset{k_2}{\rightarrow} C.$$
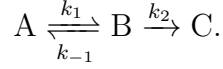
As we will see, the dynamics in that regime is driven by a *single* effective parameter k_{eff}.

5.1. Exact solution

The evolution of the molar concentrations is dictated by the linear ODEs

$$\frac{d}{dt} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} -k_1 & k_{-1} & 0 \\ k_1 & -(k_{-1}+k_2) & 0 \\ 0 & k_2 & 0 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix}, \text{ subject to } \begin{bmatrix} A(0) \\ B(0) \\ C(0) \end{bmatrix} = \begin{bmatrix} A_0 \\ B_0 \\ C_0 \end{bmatrix}.$$

This system has the explicit solution

$$
\begin{aligned}
\begin{bmatrix} A(t) \\ B(t) \\ C(t) \end{bmatrix} &= \frac{\lambda_+ A_0 + (\lambda_+ + k_2)B_0}{\Delta\lambda} \begin{bmatrix} -(1 + \lambda_-/k_2) \\ \lambda_-/k_2 \\ 1 \end{bmatrix} e^{\lambda_- t} \\
&+ \frac{\lambda_- A_0 + (\lambda_- + k_2)B_0}{\Delta\lambda} \begin{bmatrix} -(1 + \lambda_+/k_2) \\ \lambda_+/k_2 \\ 1 \end{bmatrix} e^{\lambda_+ t} \quad (10) \\
&+ (A_0 + B_0 + C_0) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},
\end{aligned}
$$

where $\Delta\lambda = \lambda_+ - \lambda_- > 0$ is the difference between the nonzero eigenvalues

$$\lambda_\pm = \frac{1}{2} \left[-(k_{-1} + k_1 + k_2) \pm \sqrt{(k_{-1} + k_1 + k_2)^2 - 4k_1 k_2} \right] < 0. \quad (11)$$

Among them, λ_- controls the fast time scale (*transient dynamics*) and λ_+ the rate of approach to the steady state $(A_*, B_*, C_*) = (0, 0, A_0 + B_0 + C_0)$ (*slow dynamics*). The third eigenvalue is identically zero and associated with the conservation law $A(t) + B(t) + C(t) = A_0 + B_0 + C_0$. For our purposes, we assume a fixed initial state $(A_0, B_0, C_0) = (1, 0, 0)$ (in arbitrary units) and variable reaction rates $\mathbf{p} = (k_{-1}, k_1, k_2)$. Observing the system takes the form of monitoring the product concentration at preset times, $\mathbf{f}(\mathbf{p}) = [C(t_1|\mathbf{p}), \ldots, C(t_5|\mathbf{p})]$. Here, the times $t_1, \ldots, t_5$ form a uniform grid on the interval $[-0.5/\lambda_+, -5/\lambda_+]$, chosen to resolve the slow dynamics.

5.2. Singularly perturbed regime

Rearranging terms from slowest to fastest in (10) yields

$$C(t|\mathbf{p}) = 1 + \frac{\lambda_-}{\Delta\lambda}e^{\lambda_+ t}\left(1 + \frac{\lambda_+}{\lambda_-}e^{r\lambda_+ t}\right), \quad \text{with } r = \frac{\Delta\lambda}{|\lambda_+|} > 0. \quad (12)$$

To resolve the slow dynamics, we select $t_1 = \alpha/|\lambda_+|$ for $\alpha \approx 0.5$. In the presence of time scale disparity, the fast component should be negligible by that time already, which necessitates that $|\lambda_+/\lambda_-|e^{-\alpha r} \ll 1$. This order relation determines the asymptotic regime in parameter space. Since the ratio $|\lambda_+/\lambda_-|$ depends only algebraically on $\mathbf{p}$, time scale separation must arise from the exponential term for all $r > r_*$ with $\alpha r_* \gg 1$. As an indication, the value $r_* = 6$ reduces the exponential term to approximately 0.05.

To identify the asymptotic regime explicitly, we work with the compactification

$$\varepsilon = \frac{1}{4}\left[1 - \left(\frac{r}{r+2}\right)^2\right] = \frac{k_1 k_2}{(k_{-1} + k_1 + k_2)^2} < \varepsilon_* = \frac{1}{4}\left[1 - \left(\frac{r_*}{r_* + 2}\right)^2\right],$$
$$(13)$$

where we have used (11) to express r in terms of the kinetic parameters. The composite parameter ε decreases with r from $1/4$ to zero, hence timescale disparity exists for $\varepsilon < \varepsilon_* \ll 1$: ε acts as a *small parameter*. As an indication, the value $r_* = 6$ above yields $\varepsilon_* \approx 0.11$. To understand the quadratic curve (13) bounding the singularly perturbed regime, we introduce the transformation

$$\begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 + \sqrt{1 - 4\varepsilon_*} & -1 + \sqrt{1 - 4\varepsilon_*} \\ -1 + \sqrt{1 - 4\varepsilon_*} & 1 + \sqrt{1 - 4\varepsilon_*} \end{bmatrix}\begin{bmatrix} k_1/k_{-1} - 2\varepsilon_*/(1 - 4\varepsilon_*) \\ k_2/k_{-1} - 2\varepsilon_*/(1 - 4\varepsilon_*) \end{bmatrix},$$

which factorizes (13) as $\kappa_1 \kappa_2 < 2\varepsilon_*/(1 - 4\varepsilon_*)$. In the (κ_1, κ_2)−plane, the asymptotic regime is bounded by two hyperbolas in the first and third quadrants. Since the κ_1− and κ_2−axes align with the axes in the $(k_1/k_{-1}, k_2/k_{-1})$ plane, at leading order, the regime $\varepsilon_* \downarrow 0$ becomes the narrow sliver shown in Fig. 5.

5.3. Effective parameter

By construction, data generated by parameter values in the asymptotic regime are well-described by the slow component alone, whose decay rate is

$$|\lambda_+| = \frac{1 - \sqrt{1 - 4\varepsilon}}{2}(k_{-1} + k_1 + k_2) \sim \frac{k_1 k_2}{k_{-1} + k_1 + k_2} = k_{\text{eff}}.$$
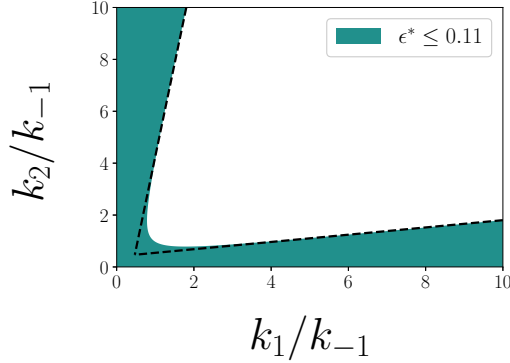
12

Figure 5: The region on the $(k_1/k_{-1}, k_2/k_{-1})$−plane within which there is timescale separation.

For parameter values in that regime, the observable effectively reads (cf. (12))

$$C(t|\mathbf{p}) = 1 - e^{-k_{\text{eff}}t}, \quad \text{since } \frac{\lambda_+}{\lambda_-}e^{-r\lambda_+t} \ll 1 \text{ and } \frac{\lambda_-}{\Delta\lambda} = -\frac{1 + \sqrt{1 - 4\varepsilon}}{2\sqrt{1 - 4\varepsilon}} \sim -1.$$

This is the formula we based our discussion in the main text on, together with its domain of applicability, cf. (13). This leading order result shows the observable to depend solely on the decay rate k_{eff} in the singularly perturbed regime. That regime is foliated by (subsets of) the level sets of k_{eff}, with the model response remaining largely constant on each such surface.

It is interesting to note that, next to sloppiness, the setup above also exhibits *structural non-identifiability*. Indeed, the observable only depends on the two parameter combinations $\lambda_\pm$, see the exact formula (12). It is important to understand that this effect is distinct from time scale disparity and thus not limited to the singularly perturbed regime. This further entails that the parameter space is foliated by curves along which $\lambda_\pm$ remain constant, with all points on any such curve yielding *identical* model responses. Since k_{eff} is merely another name for λ_+, each surface $k_{\text{eff}} = const.$ is itself foliated by these curves. In other words, the curves of *identical* model response (constant $\lambda_\pm$) foliate surfaces of *leading-order identical* model response (constant λ_+) which, in turn, foliate the singularly perturbed regime. We remark once again that sloppiness and non-identifialibity are pertinent to both the system and the monitoring protocol employed. Allowing $B_0 > 0$ in the initial

condition or observing $A(t)$ and/or $B(t)$, next to $C(t)$, suffices to lift the non-identifiability.

6. The Michaelis–Menten–Henri system

We consider the prototypical chemical pathway [14, 17]

$$\text{S} + \text{E} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{C} \overset{k_2}{\rightarrow} \text{P} + \text{E},$$
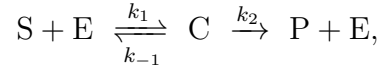
modeling the two-step conversion of a substrate S into product P through the mediation of an enzyme E. The constituent concentrations evolve under

$$
\begin{aligned}
S' &= -k_1 ES + k_{-1}C, \\
C' &= k_1 ES - (k_{-1} + k_2)C, \\
E' &= -k_1 ES + (k_{-1} + k_2)C, \\
P' &= k_2 C,
\end{aligned}
\tag{14}
$$

supplemented by the initial concentrations S_0, E_0, C_0 and P_0. This system has two exact conservation laws expressing mass balance,

$$S + C + P = S_0 + C_0 + P_0 =: S_T \quad \text{and} \quad C + E = C_0 + E_0 =: E_T.$$

Typically, one uses these to eliminate the last two ODEs, thus obtaining

$$
\begin{aligned}
S' &= -k_1(E_T - C)S + k_{-1}C, \\
C' &= k_1(E_T - C)S - (k_{-1} + k_2)C.
\end{aligned}
\tag{15}
$$

This is the classical Michaelis–Menten–Henri system in *dimensional* form [18]. In a typical experimental setting, $C_0 = P_0 = 0$ so that $S_T = S_0$. We follow this setup here and consider a problem with five parameters, the three kinetic constants $k_{\pm 1}$ and k_2 and the total concentrations S_T and E_T. To further emulate an experimental setting, we set our observable to be the product concentration, whose time course is determined by the IVP

$$P' = k_2 C, \quad \text{subject to } P_0 = 0. \tag{16}$$

Equations (15) and (16) represent the *original form* of the Michaelis–Menten–Henri system.

6.1. System dynamics

The multiscale dynamics of this system has been analyzed in a series of publications, among which the landmark articles [13, 18]. The salient features of these two articles is the system nondimensionalizations they propose and, in particular, the small parameter they use to define the singularly perturbed regime. Specifically, the authors of [13] work with the small parameter $\bar{\varepsilon} = E_T/S_T$, whereas those of [18] use $\varepsilon = E_T/(S_T + K_M)$; the asymptotic regime is defined as $\bar{\varepsilon} \ll 1$ and $\varepsilon \ll 1$, respectively.

To nondimensionalize the system, we draw inspiration from [18] and define the new parameter set

$$(\sigma, K_M, V_M, \kappa, \varepsilon) = \left(\frac{S_T}{K_M}, \frac{k_{-1} + k_2}{k_1}, k_2 E_T, \frac{k_{-1}}{k_2}, \frac{E_T}{S_T + K_M} \right). \qquad (17)$$

Here, S_T and K_M have units of concentration, V_M of reaction speed and κ and ε are non-dimensional. The inverse of this bijection is

$$S_T = \sigma K_M, \ E_T = \varepsilon K_M (\sigma + 1), \ (k_{-1}, k_1, k_2) = \frac{V_M}{\varepsilon K_M (\sigma + 1)} \left(\kappa, \frac{\kappa + 1}{K_M}, 1 \right).$$

Additionally, we nondimensionalize time and reactant concentrations through

$$\tau = \frac{t}{t_s} \quad \text{and} \quad (s, c, p) = \left(\frac{S}{S_T}, \frac{C}{\bar{C}}, \frac{P}{S_T} \right),$$

where the slow timescale t_S and complex concentration estimate $\bar{C}$ are [18]

$$t_s = \frac{S_T + K_M}{V_M} \quad \text{and} \quad \bar{C} = \frac{E_T S_T}{S_T + K_M}.$$

The new system of ODEs becomes

$$\begin{aligned} \dot{s} &= (\kappa + 1) \left[-(1 + \sigma)s + \sigma cs + \frac{\kappa}{\kappa + 1} c \right], \\ \varepsilon \dot{c} &= (\kappa + 1) \left[(1 + \sigma)s - \sigma cs - c \right], \end{aligned} \qquad (18)$$

with initial conditions $s_0 = 1$ and $c_0 = 0$. The observable p evolves under

$$\dot{p} = c, \quad \text{subject to } p_0 = 0. \qquad (19)$$

This is the *rescaled form* of the system, studied in the main text.

7. The catalyst pellet system

When a heterogeneous chemical reaction takes place in a porous catalyst pellet, transport of the reactants to the pellet surface and then through its pores to the catalyst sites plays an important role in determining the overall reaction rate. The Thiele modulus (Φ) is a dimensionless parameter determining the relative strengths of diffusive transport and reaction. For a spherical particle, Φ can be defined as

$$\Phi = R\sqrt{k/D}, \tag{20}$$

with R the particle radius, k the first order reaction rate constant and D the effective diffusivity.

The overall performance of the catalyst pellet is traditionally expressed in terms of an *effectiveness factor* (η), which compares the average reaction rate throughout the catalyst to the reaction rate had the conditions (concentration, temperature) in it been uniformly the same as on its surface. For an isothermal reaction, $\eta \downarrow 0$ in the presence of severe diffusion limitations ($\Phi \ll 1$), while $\eta \uparrow 1$ in the absence of transport limitations ($\Phi \approx 1$ and the conditions inside the pellet match those on its surface). For a spherical particle with first-order kinetics,

$$\eta = \frac{3}{\Phi^2} \frac{\delta C}{\delta r}\bigg|_{r=R}. \tag{21}$$

In the non-isothermal case where one allows for heat generation during the reaction, however, one may find that η exceeds unity. This is so because the hot pellet interior accelerates the reaction, relative to the cool region close to its surface. In this case, both mass and heat transfer play a role. Writing H for the molar heat of reaction, K for the heat conductivity, T_0 for the boundary temperature, and C_0 for the boundary concentration, one can define the parameter $\beta = C_0 H R/(K T_0)$ characterizing non-isothermal behavior. In the isothermal case ($\beta = 0$), η is in an one-to-one correspondence with Φ. In non-isothermal cases ($\beta \neq 0$), the relation between Φ and η loses injectivity and a single $\eta-$value can correspond to various $\Phi-$values. Another significant parameter when considering non-isothermal reactions is $\gamma = Q/(R T_0)$, with Q the activation energy of the reaction. This parameter can be interpreted as the sensitivity of the reaction to temperature changes, since its logarithm corresponds to the Arrhenius expression of the reaction

16

rate outside the pellet. For our discussion in the main text, we fixed the γ−value to 20 and varied the value of β to produce different model response curves. For detailed information regarding the numerical solution of η vs. Φ curves in non-isothermal catalysis, we refer the interested reader to [25].

With this data in hand, we can proceed to apply DMAPS as usual. To use "offset" data ($\eta_{i+\Delta}$, cf. main text) as part of our model response, we generated the (Φ, η)−curve using a regular grid of step Δ in $\log \Phi$. That way, each model response η_i corresponding to the grid point $\log \Phi_i$ could easily be combined with $\eta_{i+\Delta}$ corresponding to the adjacent grid point $\log \Phi_i + \Delta$.

8. Characterizing the "good parameter set"

Traditionally, parameter sensitivities have been analyzed by inspecting the eigenvalues of the Hessian of some objective function near a reported minimum. Vanishingly small eigenvalues suggest directions in parameter space in which the goodness of fit remains nearly invariant. Such directions provide us with a sense of the dimensionality of the "good parameter set" (or set of good fits) – the set of parameter values leading to an objective function value practically indistinguishable from its value at the reported minimum. For the ABC model presented in the main text, this "good set" is visibly 2D (see Fig. 6 of the manuscript), and one might be tempted to exploit this feature to determine the number of effective parameters (see also the recent work of [26] for a connection to ideas from statistical mechanics). Specifically, the ABC model has a total of three parameters and two neutral directions parameterizing the "good set," which suggests the existence of a *single* effective parameter. This is confirmed by our output-informed kernel DMAPS computations for that model in the main text.

This short section uses a somewhat contrived example to showcase a caveat: that nonlinearity *in the way the inputs enter the model* may obscure the true dimensionality of the "good set" and, through this, lead to an erroneous estimation of the degree of model sloppiness. We first introduce an ODE model followed by a transformation of the states which in total contains four parameters $(\lambda, \varepsilon, a, b)$. The two-dimensional ODE system

$$
\begin{aligned}
X' &= -\lambda X, \\
\varepsilon Y' &= -Y,
\end{aligned}
\tag{22}
$$

17

is followed by the transformation $(X, Y) \mapsto (x, y)$ given by

$$\begin{aligned}
x &= X + by^2, \\
y &= Y + ax^2.
\end{aligned} \tag{23}$$

where λ and a control slow contraction rate and slow manifold topology respectively (these two are effective parameters). The parameter ε dictates the fast transient and b controls fast fiber shape (ε and b are here sloppy parameters). When $\varepsilon \ll \lambda$, Eq. (22) becomes singularly perturbed and Y quickly decays to zero. The transformation in (23) serves to create *nonlinear* fast and slow manifolds in the (x, y) plane, $x = X(t_0) + by^2$ and $y = ax^2$, respectively. To make our point, we now transform the parameters a and λ to two other parameters u_2, w_2 that are invertible functions of them. This is accomplished through two iterations of the Hénon map (for $A = 1.4$ and $B = 0.3$) (that provide the nonlinear invertible transformation) as

$$\begin{aligned}
u_2 &= 1 - A(1 - A\lambda^2 + a)^2 + B(1 - A\lambda^2 + a), \\
w_2 &= b(1 - A\lambda^2 + a).
\end{aligned} \tag{24}$$

The new model (which can be thought of as an observation of (22) through a "curved mirror") has two parameters $\mathbf{p} = (u_2, w_2)$, which are in an one–to–one correspondence with the original parameters (λ, a). We now fix a base value of $\mathbf{p}^* = (0.7956, 1.8)$, corresponding to $(\lambda^*, a^*) = (1, 1)$ and compute the model response

$$\mathbf{f}(\mathbf{p}) = \begin{bmatrix} x(t_0|\mathbf{p}) & y(t_0|\mathbf{p}) \\ x(t_1|\mathbf{p}) & y(t_1|\mathbf{p}) \\ \vdots & \vdots \\ x(t_N|\mathbf{p}) & y(t_N|\mathbf{p}). \end{bmatrix} \tag{25}$$

Here, we fixed $\varepsilon = 10^{-3}$, $b = 10^{-2}$ and $t_1, \ldots, t_N$ to be $N = 10$ evenly spaced points in $[0.1, 1.0]$. To investigate which parameter values $\mathbf{p}$ generate points on the model manifold close to $\mathbf{f}^* = \mathbf{f}(\mathbf{p}^*)$, we first sampled $\mathbf{p}$ uniformly on the rectangle $u_2 \in (-2, 30)$, $w_2 \in (-1.5, 0.7)$; we then used each point as an initial value for a least squares minimization routine with objective function

$$c(\mathbf{p}) = \|\mathbf{f}(\mathbf{p}) - \mathbf{f}^*\|_F^2; \tag{26}$$

here, $\| \cdot \|_F$ denotes the Frobenius norm. Since noise is not part of our setup, our objective function has a unique minimum at $\mathbf{p}^*$ making it zero;
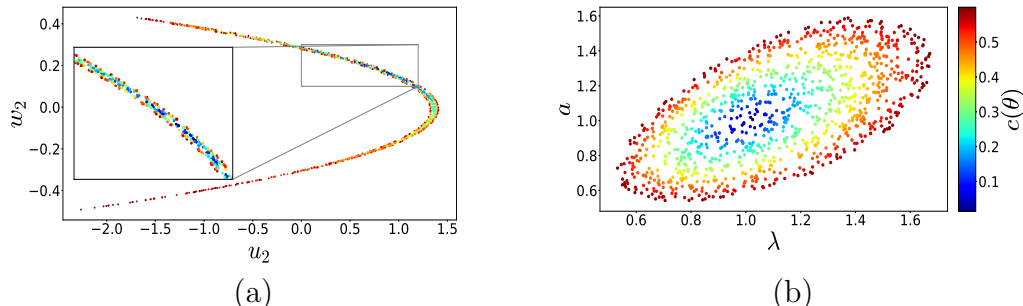
Figure 6: (a) Sample of the transformed (u_2, w_2) parameter space colored by c. Significant deviations from the expected ellipsoidal "good" parameter set are observed. (b) Original (a, λ) parameter space, in which the expected ellipse is recovered. Both figures share the color bar on the right.

to reflect that, we terminate our iterative minimization routine at any point $\theta = \mathbf{p}_i$ satisfying $c(\theta) < 0.8$. Since different initializations for our gradient descent algorithm terminate at different points, our procedure samples the "good set" of parameter settings giving predictions close to $\mathbf{f}^*$. This set of parameter combinations is shown in Fig. 6 (a), with each point $\mathbf{p}_i$ colored by its objective function value $c(\theta)$. Model nonlinearity is evident in that the set deviates markedly from the expected elliptical shape close to an isolated minimum. Transforming this set back to the original parameters (a, λ), as in Fig. 6 (b), we recover the typical, elliptical structure expected around the perfect fit.

A standard DMAPS analysis of that "good set" with the *input*-only informed kernel, i.e with the Euclidean norm in $\mathbf{p}$ (i.e. in u_2, w_2), suggests an apparent dimensionality of one. The first DMAPS eigenvector parameterizes the long, curved, thin direction "along" the cloud, see Fig. 7(a), while the second, thin dimension is "lost" in subsequent, higher-order eigenvectors.

As discussed in Eq. 8 of the main text, applying a more informative kernel, that includes both input and outputs, in our data-driven DMAPS analysis can give a more informative result. Figure 7 (b) shows the *original* parameter space (a, λ) parameterized by the first two non-trivial DMAPS eigenvectors (ϕ_1^*, ϕ_2^*) (top); and the converse (bottom) using the mixed kernel; as we expected the "good" parameter set now appears visibly two dimensional.

This phenomenon is the result of our (intentionally) poor choice of the new model parameters (u_2, w_2). Data-driven approaches, such as DMAPS, can
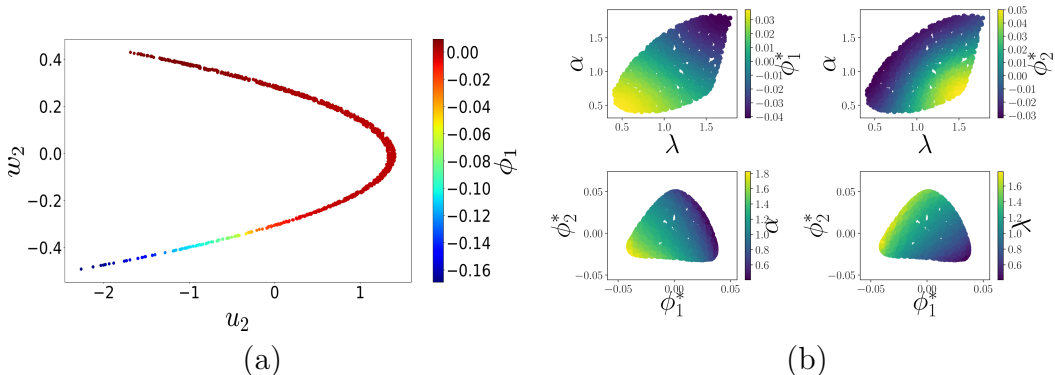
19

Figure 7: (a) Coloring the (u_2, w_2) plane by ϕ_1 from a DMAPS analysis based solely on input–only (parameter) informed kernel. The long curve is captured but the thin dimension (the width) is not resolved (the next thirty eigenvectors did not capture this new direction!). (b) Original parameters α and λ colored by the mixed, input–output kernel DMAPS coordinates (ϕ_1^*, ϕ_2^*) together with diffusion space colored by the original parameters. The two-dimensional effective nature of the "good" input set (and its correlation with the original inputs (a, λ) is clearly visible.

thus help us reparameterize (i.e. appropriately transform) parameter space (u_2, w_2) to a new one (ϕ_1^*, ϕ_2^*) that has a much better bi-Lipschitz relation with the original parameter set (a, λ) better resolving model variability.

9. A quick discussion of Active Subspaces

The Active Subspaces algorithm of P. Constantine and coworkers [23] has been developed based on the idea of finding the "important directions" in the space of all inputs of a nonlinear scalar function $f : \mathbb{R}^m \to \mathbb{R}$, a map from m-dimensional parameter space to the real line. The aforementioned directions are assumed to be weighted *linear* combinations of the input parameters. These directions are called "active subspaces" and point towards the direction of most intense change of an "observable."

1. We consider N sample points in parameter space. For each sample point $\mathbf{x_n} \in \mathbb{R}^m$, we observe $f(\mathbf{x_n})$ and the gradient $\nabla f(\mathbf{x_n})$.
2. The average of the outer product of the gradient with itself on the sample data is computed through

$$\hat{C} = \frac{1}{M} \Sigma_{n=1}^N \nabla f(\mathbf{x_n}) \nabla f(\mathbf{x_n})^T.$$

20

Here, ∇f is seen as an $m-$dimensional column vector, hence $\hat{C}$ is an $m \times m$ matrix.

3. We find the eigendecomposition of $\hat{C} = \hat{W}\hat{\Lambda}\hat{W}^T$.

4. The observable exhibits the greatest change in the space spanned by the leading eigenvectors of $\hat{C}$, which are stored column-wise in $\hat{W}$ [23].

To illustrate this procedure and its outcome on a simple example, we consider the map

$$f : \mathbb{R}^2 \to \mathbb{R} \quad \text{given by } (x_1, x_2) \mapsto f(x_1, x_2) = e^{x_1^\alpha + x_2}. \qquad (27)$$

where α can be fixed at an arbitrary value. Although the parameter space (x_1, x_2) is 2-D, the map effectively only depends on the single effective parameter $x_1^\alpha + x_2$. As a result, the parameter space can be re-parameterized by any one-to-one function of $x_1^\alpha + x_2$. To discover active subspaces for (27), we consider a uniform grid on $[-1, 1] \times [-1, 1]$ and evaluate the map f on each grid point. Since the parameter space is effectively 1-D (composed of level curves of the effective parameter), the active subspace parameterization is given by $\psi_1 = w_1^T \cdot [x_1, x_2]^T$, at each point. We compute the active subspaces for $\alpha = 1$ and $\alpha = 5$:

- For $\alpha = 1$: As expected, the effective parameter ψ_1 is the linear combination $x_1 + x_2$ of the input parameters. As shown in Fig. 8(a), the observable is in an one-to-one correspondence with ψ_1, meaning that active subspaces recover the effective parameter.

- For $\alpha = 5$: In this case, the effective parameter is not a linear combination of input parameters. As shown in Fig. 8(b), the basic algorithm does not discover the effective one-dimensional (but nonlinear) relation between the input parameters.

To conclude this discussion and compare methodologies, we applied the output-only DMAPS algorithm to each of these cases and plotted the observable against the first non-trivial eigenfunction of the graph Laplacian; see Fig. 9. Plainly, th DMAPS algorithm captures the effective parameter both when that is linear and when it is nonlinear in the input parameters.
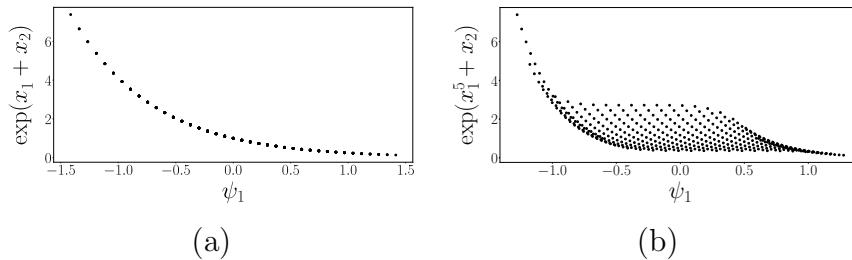
Figure 8: The observable f, plotted against the first active subspace coordinate for (a) $\alpha = 1$ and (b) $\alpha = 5$.
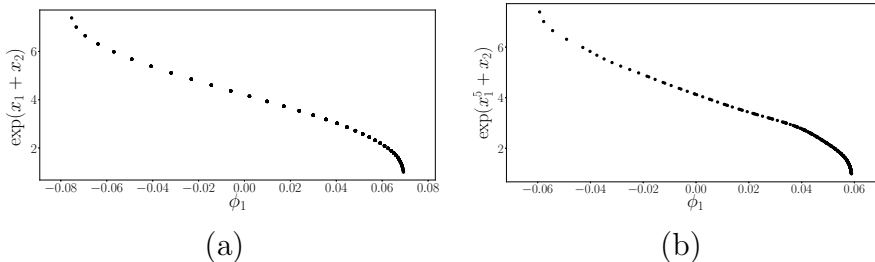

Figure 9: The observable f plotted against the first DMAPS coordinate for (a) $\alpha = 1$ and (b) $\alpha = 5$.

The idea of active subspaces has been employed by Constantine and coworkers to decompose input space using products of powers of the inputs, since these become linear combinations in a logarithmic scale ("ridge functions" [24]). It will be interesting to explore how more general nonlinear relations can be used in discovering effective nondimensionalizations.

10. On the origins of sloppiness

We now attempt to generalize the setting presented in Section 1 and explore the origins of parameter (input) sloppiness. Here also, we consider multivariable vector functions $\mathbf{x}(\mathbf{t}|\mathbf{p}) \in \mathrm{R}^D$, where D is arbitrary. The independent variables are partitioned in what one might call "bona fide variables" $\mathbf{t} = (t_1, \ldots, t_K) \in I \subset \mathrm{R}^K$ and *parameters* $\mathbf{p} = (p_1, \ldots, p_M) \in \Theta \subset \mathrm{R}^M$. For each $\mathbf{p} \in \Theta$, we term the function $\chi(\mathbf{p}) = \mathbf{x}(\,\cdot\,|\mathbf{p}) \in \mathrm{X}$ the *full model response* and the function space X containing it the *full model space*. To know the mapping $\chi : \Theta \to \mathrm{X}$ is to know fully the solution to the problem at hand, and such knowledge is typically unattainable. As $\mathbf{p}$ ranges over Θ, χ traces

out the *full model manifold* $\mathcal{X} = \{(\mathbf{p}, \chi(\mathbf{p}))\}_{\mathbf{p} \in \Theta} \subset \Theta \times \mathrm{X}$. That manifold is generically M−dimensional and, if $\mathbf{x}$ depends on $\mathbf{p}$ in a C^1 manner, also continuously differentiable. Here also, as in Section 1, we only monitor one or more *functionals*, $f_1, \ldots, f_N : \mathrm{X} \to \mathrm{R}$. We term each individual $f_n(\chi(\mathbf{p}))$ a *partial observation* and the N−tuple

$$\mathbf{f}(\mathbf{p}) = [f_1(\chi(\mathbf{p})), \ldots, f_N(\chi(\mathbf{p}))] \in \mathrm{R}^N$$

the *observed* model response. Under the action of $\mathbf{f}$, the full model manifold $\mathcal{X}$ is projected to the *(observed) model manifold* $\mathcal{M} = \{(\mathbf{p}, \mathbf{f}(\mathbf{p}))\}_{\mathbf{p} \in \Theta} \subset \Theta \times \mathrm{R}^N$. We will assume $\mathbf{f} : \Theta \to \mathbf{f}(\Theta)$ to be a homeomorphism, unless explicitly stated, and the linear map $\mathrm{D_p f}$ to have rank M. These assumptions make $\mathbf{f}$ an atlas for $\mathcal{M}$ and are satisfied by all models we consider in the main text, save for the non-invertible model therein.

To discuss sensitivity of the observed model response to parameter variations, one must assess shifts in that response relative to such variations. In [20], the authors defined distances on $\mathcal{M}$ by omitting the $\mathbf{p}$−component and working with the projected manifold $\pi \mathcal{M}$, where $(\mathbf{p}, \mathbf{f}(\mathbf{p})) \overset{\pi}{\mapsto} \mathbf{f}(\mathbf{p})$. Distances were measured using a Riemannian metric induced by a norm in the carrier space of $\pi \mathcal{M}$, i.e. R^N. Similarly, in this section we work with the standard Euclidean norm, reserving important questions on weighing and correlation of partial observations for the future. An infinitesimal displacement $\mathrm{d}\mathbf{p} = (\mathrm{d}p_1, \ldots, \mathrm{d}p_M)^\mathrm{T}$ in Θ yields the infinitesimal displacement $\mathrm{d}z = (\mathrm{d}z_1, \ldots, \mathrm{d}z_N)^\mathrm{T} = (\mathrm{D_p f}) \, \mathrm{d}\mathbf{p} \in \mathrm{T_p}(\pi \mathcal{M})$ with length $||\mathrm{d}z||^2 = (\mathrm{d}\mathbf{p})^\mathrm{T} g \, \mathrm{d}\mathbf{p}$. The $M \times M$ positive definite matrix $g = (\mathrm{D_p f})^\mathrm{T}(\mathrm{D_p f})$ is the *metric tensor* for $\pi \mathcal{M}$ for the specific atlas $\mathbf{f}$. Although the positive definiteness of g makes the system responsive to *all* parameter variations, the observed model response locally around some point $\mathbf{p} \in \Theta$ may vary greatly with the direction of $\mathrm{d}\mathbf{p}$ and be disproportionately small along certain directions. This is the phenomenon termed *sloppiness*, and it manifests itself in the spectrum of the metric. Small eigenvalues yield small observed responses, with sloppy directions in Θ being the pull-backs under μ of the associated eigendirections in $\mathrm{T_p}(\pi \mathcal{M})$.

In terms of the full model manifold, parameter values are first mapped to $\mathcal{X} \subset \Theta \times \mathrm{X}$ equipped with the Riemannian metric $h = (\mathrm{D_p}\chi)^\mathrm{T}(\mathrm{D_p}\chi)$, then projected to the model manifold $\mathcal{M} \subset \mathrm{R}^N$, equipped with the metric

$g = (\mathrm{D_p f})^{\mathrm{T}}(\mathrm{D_p f})$. We can quantify the relation between these two metrics and examine how that projection can generate non-identifiability as well as sloppiness. For clarity of presentation, we restrict our attention to linear functionals $f_1, \ldots, f_N$ on a Hilbert space X equipped with inner product $\langle \cdot, \cdot \rangle$ and induced norm $|| \cdot ||$. In that setting, X is isomorphic to its dual X^* and hence $f_n = \langle e_n, \cdot \rangle$, for some $e_n \in \mathrm{X}$ and all $n = 1, \ldots, N$. Therefore, $\mathbf{f(p)} = [\langle e_1, \chi(\mathbf{p}) \rangle, \ldots, \langle e_N, \chi(\mathbf{p}) \rangle]^{\mathrm{T}}$, with push-forward

$$
\mathrm{D_p f} = \begin{bmatrix} \langle e_1, \partial_{p_1}\chi \rangle & \cdots & \langle e_1, \partial_{p_N}\chi \rangle \\ \vdots & \ddots & \vdots \\ \langle e_N, \partial_{p_1}\chi \rangle & \cdots & \langle e_N, \partial_{p_N}\chi \rangle \end{bmatrix} = \begin{bmatrix} \langle e_1, \cdot \rangle \\ \vdots \\ \langle e_N, \cdot \rangle \end{bmatrix} \mathrm{D_p}\chi.
$$

Based on this, the metric on $\mathcal{M}$ is written as

$$
g = (\mathrm{D_p f})^{\mathrm{T}}(\mathrm{D_p f}) = (\mathrm{D_p}\chi)^{\mathrm{T}} F (\mathrm{D_p}\chi). \tag{28}
$$

The linear operator $F : \mathrm{X} \to \mathrm{X}^*$, here, is given for each $v \in \mathrm{R}^N$ by

$$
Fv = \sum_{n=1}^{N} \langle e_n, v \rangle \langle e_n, \cdot \rangle = \begin{bmatrix} \langle e_1, \cdot \rangle, \ldots, \langle e_N, \cdot \rangle \end{bmatrix} \begin{bmatrix} \langle e_1, v \rangle \\ \vdots \\ \langle e_N, v \rangle \end{bmatrix} \in \mathrm{X}^*.
$$

To understand (28) better, we use the isomorphism $\mathrm{X}^* \cong \mathrm{X}$ to interpret F as a symmetric endomorphism on X. Its spectrum consists of the zero eigenvalue, linked to the co-dimension N kernel $\mathrm{Ker}F = \bigcap_{n=1}^{N} e_n^{\perp}$, and of a nontrivial part linked to the invariant subspace $\mathrm{Im}F = \mathrm{span}(e_1, \ldots, e_N)$. In the basis $\{e_1, \ldots, e_N\}$ for $\mathrm{Im}F$, the restriction $F|_{\mathrm{Im}F}$ is represented by the matrix

$$
G = \begin{bmatrix} \langle e_1, e_1 \rangle & \cdots & \langle e_1, e_N \rangle \\ \vdots & \vdots & \vdots \\ \langle e_N, e_1 \rangle & \cdots & \langle e_N, e_N \rangle \end{bmatrix}. \tag{29}
$$

This proves that the nontrivial part of the spectrum consists of the eigenvalues of the $N \times N$ Gram matrix G. To rewrite (28) using this information, we decompose the columns of $\mathrm{D_p}\chi$ along the invariant subspaces $\mathrm{Ker}F$ and $\mathrm{Im}F$,

$$
\partial_{p_n}\chi = \mathrm{N}_n + \mathrm{I}_n = \mathrm{N}_n + [e_1, \ldots, e_N]\, C_n, \quad \text{with } \mathrm{N}_n \in \mathrm{Ker}F \text{ and } \mathrm{I}_n \in \mathrm{Im}F.
$$

The corresponding matrix decomposition is $D_{\mathbf{p}}\chi = N + [e_1, \ldots, e_N]\,C$, with C_{mn} the component of $\partial_{p_n}\chi$ along e_m, and thus

$$g = C^{\mathrm{T}} G^2 C. \tag{30}$$

This equation expresses the metric on $\mathcal{M}$ in terms of a Gram matrix, determined by the functionals $f_1, \ldots, f_N$, and of a matrix quantifying the projection of $\mathrm{T}_{\chi(\mathbf{p})}\mathcal{X}$ on the combined range of those functionals along their joint kernel.

Equation (30) shows that an ill-conditioned G or C leads to sloppiness, manifested in disparities within $\sigma(g)$. An ill-conditioned matrix G points to functionals that are either badly scaled or nearly dependent. This was the case in [22], where sloppiness was traced to a Vandermonde matrix specific to Taylor polynomials. Replacing those polynomials by an orthonormal set would have sufficed to remove sloppiness. An ill-conditioned matrix C, instead, is due to directions in $\mathrm{T}_{\mathbf{p}}\mathcal{X}$ that align well with $\bigcap_{n=1}^{N} \mathrm{Ker} f_n$. In that case, parameter variations can generate negligible model responses on $\pi\mathcal{M}$, although the observation functionals are proportionate and the *full* model response on $\mathcal{X}$ appreciable. This is the case with multiscale systems, in which certain parameters combinations affect behavior at unobserved scales (fast transients).

References

[1] P. Bérard, G Besson, and S Gallot, *Embedding Riemannian manifolds by their heat kernel*, Geometric and Functional Analysis **4** (1994), no. 4, 373–398.

[2] S. C. Brenner and L. R. Scott, *The mathematical theory of Finite Element Methods*, Texts in Applied Mathematics, vol. 15, Springer New York, New York, NY, 2008.

[3] F. R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, vol. 92, American Mathematical Society, Providence, Rhode Island, dec 1996.

[4] M. Ehrgott, *Multicriteria optimization*, second ed., Springer-Verlag, Berlin, 2005.

[5] S. Lafon, *Diffusion maps and geometric harmonics*, Yale University (2004), no. May, 97.

[6] A. Logg, K.-A. Mardal, and G. Wells (eds.), *Automated solution of differential equations by the Finite Element Method*, Lecture Notes in Computational Science and Engineering, vol. 84, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[7] J. E. Marsden and T. J. R. Hughes, *Mathematical foundations of elasticity*, Dover Publications, Inc., New York, 1994.

[8] S. Morita, *Geometry of differential forms*, Translations of Mathematical Monographs, vol. 201, American Mathematical Society, Providence, RI, 2001.

[9] A. Singer, *From graph to manifold Laplacian: The convergence rate*, Applied and Computational Harmonic Analysis **21** (2006), no. 1, 128–134.

[10] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (10th ed.), National Bureau of Standards, Washington, DC, 1972.

[11] P. Achard and E. De Schutter, Complex parameter landscape for a complex neuron model, *PLoS Comp. Biol.* **2(7)** 0794–0804, 2006.

[12] J. I. Ankenman, Geometry and Analysis of Dual Networks on Questionnaires, Ph.D. Thesis, Yale University, 2014.

[13] F. G. Heineken, H. M. Tsuchiya and R. Aris, On the mathematical status of the pseudo-steady state hypothesis of biochemical kinetics, *Math. Biosci.* **1** 95–113, 1967.

[14] K. A. Johnson and R. S. Goody, The original Michaelis constant: translation of the 1913 Michaelis–Menten paper, *Biochem.* **50** 8264–8269, 2011.

[15] H. Lineweaver and D. Burk, The determination of enzyme dissociation constants, *J. Am. Chem. Soc.* **56** 658–666, 1934.

[16] P. Marjoram, J. Molitor, V. Plagnol and S. Tavaré, Markov chain Monte Carlo without likelihoods, *Proc. Natl. Acad. Sci.* **100** 15324–15328, 2003.

[17] L. Michaelis and M. L. Menten, Die Kinetik der Invertinwirkung, *Biochem. Z.* **49** 333–369, 1913.

[18] L. A. Segel and M. Slemrod, The Quasi-Steady-State Assumption: A Case Study in Perturbation, *SIAM Review* **31(3)** 446–477, 1989.

[19] M. K. Transtrum, B. B. Machta and J. P. Sethna, Why are nonlinear fits to data so challenging?, *Phys. Rev. Lett.* **104** 060201, 2010.

[20] M. K. Transtrum, B. B. Machta and J. P. Sethna, Geometry of nonlinear least squares with applications to sloppy models and optimization, *Phys. Rev. E* **83** 036701, 2011.

[21] M. K. Transtrum and P. Qiu, Model reduction by manifold boundaries, *Phys. Rev. Lett.* **113** 098701, 2014.

[22] J. J. Waterfall, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser and J. P. Sethna, Sloppy-model universality class and the Vandermonde matrix, *Phys. Rev. Lett.* **97** 150601, 2006.

[23] Constantine, Paul G and Dow, Eric and Wang, Qiqi, Active subspace methods in theory and practice: applications to kriging surfaces *SIAM Journal on Scientific Computing*, **36** (4) A1500–A1524.

[24] Constantine, Paul G and Eftekhari, Armin and Hokanson, Jeffrey and Ward, Rachel A, A near-stationary subspace for ridge approximation, *Computer Methods in Applied Mechanics and Engineering*, **326** 402–421, 2017.

[25] Weisz, P.B. and Hicks, J.S., The behaviour of porous catalyst particles in view of internal mass and heat diffusion effects, *Chemical Engineering Science*, **17(4)** 265–275, 1962.

[26] LaMont, Colin H and Wiggins, Paul A, A correspondence between thermodynamics and inference, arXiv preprint arXiv:1706.01428,2017