Coarse-Graining Hamiltonian Systems Using WSINDy: Supplementary Information

Daniel A. Messenger*, **Joshua W. Burby, David M. Bortz**

***Corresponding author:** daniel.messenger@colorado.edu

A Robust cornerpoint method

In [1], the authors developed a method of choosing the test function hyperparameters using a cornerpoint algorithm applied to the Fourier spectrum of the data. This approach was taylored to the case of extrinsic noise, in which case the cumulative sum of the power spectrum of the data can be approximated be two line segments, the intersection of which (the cornerpoint) closely marks the changepoint from signal-dominated to noise-dominated Fourier modes. For nearly periodic systems without extrinsic noise, the two line segment assumption is not accurate, however, a feasible test functions support width T_ϕ can still be computed using a slightly altered cornerpoint approach, simply by taking the minimum of the rotated cumulative sum of the power spectrum. We also extend the method in [1] to work with any test function, not just those approximated by Gaussian distributions.

To this end, denote the discrete Fourier transform (DFT) of the nth component of the data $\mathbf{Z}$ by

$$\mathcal{F}[\mathbf{Z}_n](k) := \sum_{j=0}^{m-1} \exp\left(-\frac{2\pi i}{m} jk\right) \mathbf{Z}_n(t_j)$$

and let its cumulative sum (taken over the negative wavenumbers) be denoted by

$$\mathbf{H}_n(k) = \sum_{k'=-m/2}^{k} \left|\mathcal{F}[\mathbf{Z}_n](k')\right|, \quad k \in \{-m/2, \dots, 0\}.$$

For convenience we will assume m is even as the odd case is nearly identical. With high probability, $\mathbf{H}_n$ will lie below the line

$$k \to \mathbf{Y}_n(k) := \left(\frac{k_{\max} - k}{k_{\max} + (m/2)}\right) \mathbf{H}_n(-m/2) + \left(\frac{k + (m/2)}{k_{\max} + (m/2)}\right) \mathbf{H}_n(k_{\max})$$

where $k_{\max} = \operatorname{argmax}_{k \in \{-m/2, \dots, 0\}} |\mathcal{F}[\mathbf{Z}_n](k)|$ is the wavenumber of the largest Fourier coefficient. A tie is broken by the maximizing k with the largest magnitude. Moreover, $\mathbf{H}_n$ will be approximately convex for $k \in \{-m/2, \dots, k_{\max}\}$. For instance, convexity and $\mathbf{H}_n \leq \mathbf{Y}$ over this region are both true if $\mathcal{F}[\mathbf{Z}_n]$ exhibits any decay of the form $|\mathcal{F}[\mathbf{Z}_n](k)| \sim |k|^{-s}$ for $s > 0$. We define the cornerpoint of $\mathbf{H}_n$ as the point $(k_n^*, \mathbf{H}_n(k_n^*))$ that maximizes the distance between $\mathbf{H}_n(k)$ and $\mathbf{Y}_n(k)$ over $k \in \{-m/2, \dots, k_{\max}\}$ in the total least squares sense. A practical computation of this is the following. Let $R(\theta)$ be the rotation matrix in the $(k, \mathbf{H}_n)$ plane such that

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \mathbf{R}(\theta) \begin{pmatrix} k \\ \mathbf{Y}_n(k) \end{pmatrix} = 0.$$

That is, $\mathbf{R}(\theta)$ rotates $\mathbf{Y}_n$ parallel to the k-axis. Then

$$k_n^* := \underset{k \in \{-m/2, \dots, k_{\max}\}}{\operatorname{argmin}} \left[\mathbf{R}(\theta) \begin{pmatrix} k \\ \mathbf{H}_n(k) \end{pmatrix}\right]_2$$

where $[\cdot]_2$ indicates the second component of the vector. This process is visualized in Figure 1, and in this work, we let the final cornerpoint be the average $k^* = \frac{1}{2N} \sum_{n=1}^{2N} k_n^*$, used for all observed variables. Figure 1 displays that the previous method is more prone to labeling irrelevant modes as signal-dominated (2nd row), but overall the performance of both methods is comparable.

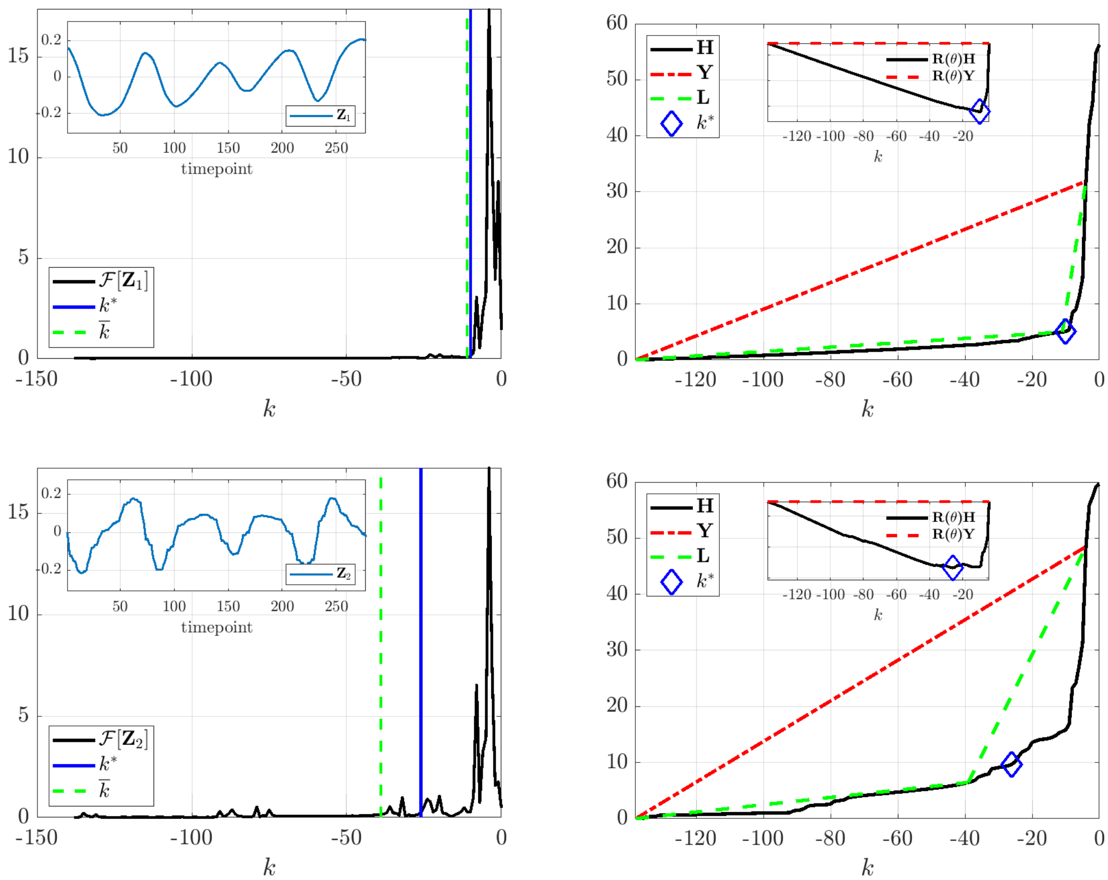

Figure 1. Visualization of cornerpoint detection and comparison to the method used in [1]. The data $\mathbf{Z}$ is taken from Example 2 with $\varepsilon = 0.05$ and $Q(0) = \frac{31\pi}{32}$. The top and bottom rows display coordinates 1 and 2, the configuration and momentum variables for the first oscillator. The data and DFTs are plotted in the left column, along with markers for the detected cornerpoints, while the right column shows the cumulative sums $\mathbf{H}_n$ together with the line $\mathbf{Y}$ and the piecewise linear approxmation $\mathbf{L}$ used in [1] (inset plots show the rotated $\mathbf{H}$ which is minimized at k^*). The value k^* is the cornerpoint detected by the current method, while $\overline{k}$ is the value detected by the method in [1]. In the top row, the two methods nearly agree, while the more perturbed dynamics in the bottom row cause the $\overline{k}$ to be significantly larger, leading to more irrelevant modes entering the recovered model.

B MSTLS algorithm

To solve for a sparse coefficient vector $\widehat{\mathbf{w}}$ such that $\mathbf{G}\widehat{\mathbf{w}} \approx \mathbf{b}$, we use the MSTLS algorithm proposed in [1] that uses the original STLS algorithm from [2] with variable thresholding and then performs a line search for the sparsity threshold λ. By heterogeneous thresholding, we mean that instead of employing the typical hard thresholding operator

$$H_\lambda(\mathbf{w})_j = \begin{cases} \mathbf{w}_j, & |\mathbf{w}_j| \geq \lambda \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

which treats all columns of $\mathbf{G}$ equally, we define the variable hard-thresholding operator

$$H_{\mathbf{L},\mathbf{U}}(\mathbf{w})_j = \begin{cases} \mathbf{w}_j, & L_j \leq |\mathbf{w}_j| \leq U_j \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

for specified upper and lower bound vectors $\mathbf{U}, \mathbf{L} \in \mathbb{R}^J$ (for a J-term library $\mathbb{H}$). This is particular useful for enforcing not only that the coefficients $\mathbf{w}_j$ stay within a certain range, but also that the term magnitudes $\|\mathbf{G}_j\mathbf{w}_j\|$ have a reasonable contribution to the dynamics given by $\mathbf{b}$.

In this work, we define $\mathbf{L}$ and $\mathbf{U}$ to enforce that $\widehat{\mathbf{w}}_j$ and $\widehat{\mathbf{w}}_j\mathbf{G}_j$ are comparable to the best 1-term solution. That is, denoting the projection operator by $\mathbf{P}$, we define

$$\hat{j} := \operatorname*{argmax}_j \|\mathbf{P}_{\mathbf{G}_j}\mathbf{b}\|_2, \qquad \hat{w} := \frac{|\mathbf{b}\cdot\mathbf{G}_{\hat{j}}|}{\|\mathbf{G}_{\hat{j}}\|_2^2}, \qquad \hat{G} := \frac{\hat{w}\|\mathbf{G}_{\hat{j}}\|_2}{\|\mathbf{b}\|_2}$$

where $\hat{w}$ and $\hat{G}$ are the coefficient and relative term magnitudes of the best 1-term projection of $\mathbf{b}$ onto the library $\mathbb{H}$. We then let

$$L_j(\lambda) = \lambda^{-1}\max\{\hat{w}, \hat{G}/\|\mathbf{G}_j\|_2\} \tag{3}$$
$$U_j(\lambda) = \lambda\min\{\hat{w}, \hat{G}/\|\mathbf{G}_j\|_2\} \tag{4}$$

which implies that the nonzero entries of $H_{\mathbf{L}(\lambda),\mathbf{U}(\lambda)}(\mathbf{w})$ are the entries of $\mathbf{w}$ satisfying

$$\lambda\hat{w} \leq |\mathbf{w}_j| \leq \lambda^{-1}\hat{w} \qquad \& \qquad \lambda\hat{G} \leq \frac{|\mathbf{w}_j|\|\mathbf{G}_j\|_2}{\|\mathbf{b}\|_2} \leq \lambda^{-1}\hat{G}.$$

The MSTLS algorithm then proceeds as follows. With initial guess $\mathbf{w}^{(0)} = \mathbf{G}^\dagger\mathbf{b}$ and support set $S^{(0)} = \{1, \ldots, J\}$ the inner STLS loop for fixed λ is defined by

$$(\text{STLS}) \qquad \begin{cases} S^{(\ell+1)} = \operatorname{supp}\left(H_{\mathbf{U}(\lambda),\mathbf{L}(\lambda)}(\mathbf{w}^{(\ell)})\right) \\ \mathbf{w}^{(\ell+1)} = \operatorname*{argmin}_{\operatorname{supp}(\mathbf{w})\subset S^{(\ell+1)}} \|\mathbf{G}\mathbf{w} - \mathbf{b}\|_2^2 \end{cases} \tag{5}$$

which is run until termination, which must occur in maximum J iterations [3]. Denoting the solution by $\text{STLS}(\mathbf{G}, \mathbf{b}; \lambda) = \mathbf{w}^\lambda$ (where $\mathbf{w}^0 = \mathbf{w}^{(0)}$), we then define the auxiliary loss function $\mathcal{L}$ by

$$\mathcal{L}(\lambda) = \frac{\left\|\mathbf{G}(\mathbf{w}^\lambda - \mathbf{w}^0)\right\|_2}{\|\mathbf{G}\mathbf{w}^0\|_2} + \frac{\left\|\mathbf{w}^\lambda\right\|_0}{\|\mathbf{w}^0\|_0}, \quad \mathbf{w}^\lambda = \text{STLS}(\mathbf{G}, \mathbf{b}; \lambda). \tag{6}$$

Finally, we define MSTLS with candidate sparsity thresholds $\boldsymbol{\lambda}$ by

$$(\text{MSTLS}) \qquad \begin{cases} \widehat{\lambda} = \min\left\{\lambda \in \boldsymbol{\lambda} \ : \ \mathcal{L}(\lambda) = \min_{\lambda\in\boldsymbol{\lambda}} \mathcal{L}(\lambda)\right\} \\ \widehat{\mathbf{w}} = \text{STLS}(\mathbf{G}, \mathbf{b}; \widehat{\lambda}). \end{cases} \tag{7}$$

We denote by $\text{MSTLS}(\mathbf{G}, \mathbf{b}, \boldsymbol{\lambda}) := \widehat{\mathbf{w}}$ the output weight vector in (7). The collection of thresholds $\boldsymbol{\lambda}$ can be chosen by the user. Following the strategy seen to be successful in [1], we let $\boldsymbol{\lambda}$ be 100 equally log-spaced values from 10^{-4} to 1. This process is visualized in Figure 2.

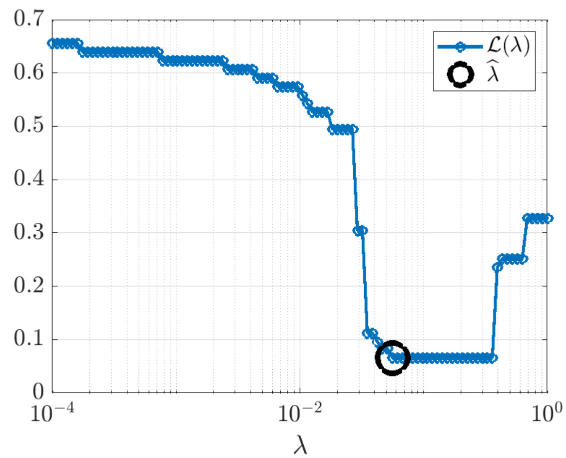

Figure 2. Example profile of the loss function $\mathcal{L}(\lambda)$ employed in MSTLS (7) for the data in Figure 16.

C Initial conditions for forward simulations

In order to measure agreement between forward simulations of $\mathcal{H}_0^\mu$ and $\widehat{\mathcal{H}}_0^\mu$, we first use a data-driven nonlinear least squares approach to find an adequate set of initial conditions. For all time intervals $I_k = [t_0, t_k]$, $k = 2, \ldots, 100$, we fit a polynomial of degree 2 through the data $\mathbf{Z}(I_k)$ and let $z(0)$ be the value of this polynomail at t_0. We then simulate the system of interest (defined by $\mathcal{H}_0^\mu$ or $\widehat{\mathcal{H}}_0^\mu$) starting from the selected $z(0)$ for 1/8 of the total time of the available data. The initial condition $z(0)$ that minimizes the error between these partial forward simulations and the data $\mathbf{Z}$ over the 99 time intervals I_k is chosen for forward simulations.

References

1. Messenger, D. A. & Bortz, D. M. Weak SINDy For Partial Differential Equations. J. Comput. Phys. **443**, 110525, DOI: 10.1016/j.jcp.2021.110525 (2021).

2. Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. **113**, 3932–3937, DOI: 10.1073/pnas.1517384113 (2016).

3. Zhang, L. & Schaeffer, H. On the Convergence of the SINDy Algorithm. Multiscale Model. Simul. **17**, 948–972, DOI: 10.1137/18M1189828 (2019).