**Figure 1.** Root Mean Square Error ($RMSE$) and Error Correlation ($EC$, the correlation coefficient between the error of the derivative estimate and the actual derivative) are two metrics for evaluating the quality of a derivative estimate which balance the accuracy, smoothness, and underestimation of the estimate.

A. The finite difference of noisy data produces excessively noisy derivatives.

B. $RMSE$ vs. $EC$ for derivative estimates from a Savitzky-Golay smoothing filter for various parameter choices (gray dots). Our proposed loss function [Eqn XX] produces estimates that lie along the violet path, approximately following the pareto front. Arrow indicates direction of increasing $\gamma$. Large circles indicate where examples shown in C-D lie in this space. Red star indicates the location of $\hat{\gamma}$, the “optimal $\gamma$”, which balances low $RMSE$ and low $EC$.

C. Derivative estimates from a Saviztky-Golay smoothing filter using five different parameter choices.

D. Error correlation for each of the examples from C. Note that smoother estimates produce higher error correlations.

Somewhere in the text:

We define the optimal choice of $\gamma$, denoted $\hat{\gamma}$, by the location of the

shoulder of the pareto curve that describes the relationship between $EC$ and $RMSE$ (e.g. the violet curve shown in Fig. 1B). Ideally, the shoulder can be found by locating the point of maximum curvature. In practice, the pareto curves often have complex curvatures, and we resort to defining $\hat{\gamma}$ as the point that minimizes $RMSE$ subject to the constraint that the $EC<0.25$.

**Figure 2.** Optimal value of $\gamma$, $\hat{\gamma}$, can automatically be chosen based on the frequency content of the data and its temporal resolution according to a simple empirical relationship.

A. (top row) Four examples of noisy sinusoidal position data with zero-mean gaussian noise, and sampling resolutions between 0.1 and 0.001. (middle row) Derivative estimate from a Savitzky-Golay filter using $\hat{\gamma}$ to select the parameters (colored line); actual derivative (black dashed line); and best possible estimate from all potential parameter choices (thick gray line). (bottom row) EC vs. RMSE for all potential parameter choices (gray), and the pareto curve defined by Eqn XX. Red star indicates the location of $\hat{\gamma}$.

B. Relationship between $\hat{\gamma}$ and frequency, for sinsusoidal curves with three choices of dt (0.1: light orange; 0.01: orange; 0.001: dark red), noise (zero-mean, standard deviation between 0.01 and 0.5), and timeseries lengths (between 1 and 128 sec). “+” symbols indicate scenarios where the timeseries length < 1/frequency, these points were omitted from subsequent analysis. Diagonal lines show the regression for each dt, constrained such that all three slopes are equal.

**Figure 3.** Empirical relationship for choosing $\hat{\gamma}$ based on signal frequency and timestep applies to a wide range of multi-frequency and nonlinear synthetic timeseries data. Each row shows the noisy data, EC vs. RMSE as plotted in Figs. 1-2, and the power spectra of the data. Based on the power spectra, we chose a cutoff frequency, which determined $\hat{\gamma}$ according to the relationship determined in Fig. 2. This choice of $\hat{\gamma}$ resulted in the position and velocity estimates shown in the next two columns, and the location of the red star on the EC vs. RMSE plot. We provide four examples from a Lorenz system with various levels of noise and timesteps, and four other datasets.

**Figure 4.** Proposed loss function, and empirical approach for choosing $\hat{\gamma}$, works equally well for four different differentiation methods applied to noisy Lorenz data, resulting in similar estimates from each method.

A. Noisy synthetic position data from a Lorenz system.

B. (first column) EC vs. RMSE for each of three methods, as plotted in Figs 1-3. Note the similar distribution of gray points in the first three rows. The method from the final row only has a single parameter, resulting in gray points that are hidden behind the colored curve. (2nd-3rd columns) Position and velocity estimates for each method (colored lines) and the actual values (black dashed line).

C. Data from B, overlaid on a single axis to highlight the similarity across the four methods.

**Figure 5.** Proposed loss function, and empirical approach for choosing $\hat{\gamma}$, works equally well for four different differentiation methods applied to a diversity of synthetic data. Each row corresponds to the same synthetic data shown in Fig. 3, and the results are plotted as in Fig. 4C.

**Figure 6.** Proposed loss function, and empirical approach for choosing $\hat{\gamma}$, produces expected results for real data.

A. Noisy timeseries data of wind speed from a 3D ultrasonic anemometer.

B. Power spectra of the data. Red line indicates our chosen cutoff-frequency, based on the location where the amplitude of the noise in the spectra subjectively increases.

C. Zoomed in portion of the data from A that was used in the proposed loss function to determine the optimal parameter choices.

D. (top) Smooth wind speed estimate (red) overlaid on the raw data. (bottom) Wind acceleration estimates. Both estimates were made using the Savitzky-Golay filter. Butterworth smoothing, constant acceleration Kalman smoothing, and Total Variation Regularized Jerk all produce indistinguishable results.

E-F. Zoomed in view of the results shown in D, from the gray-shaded section.