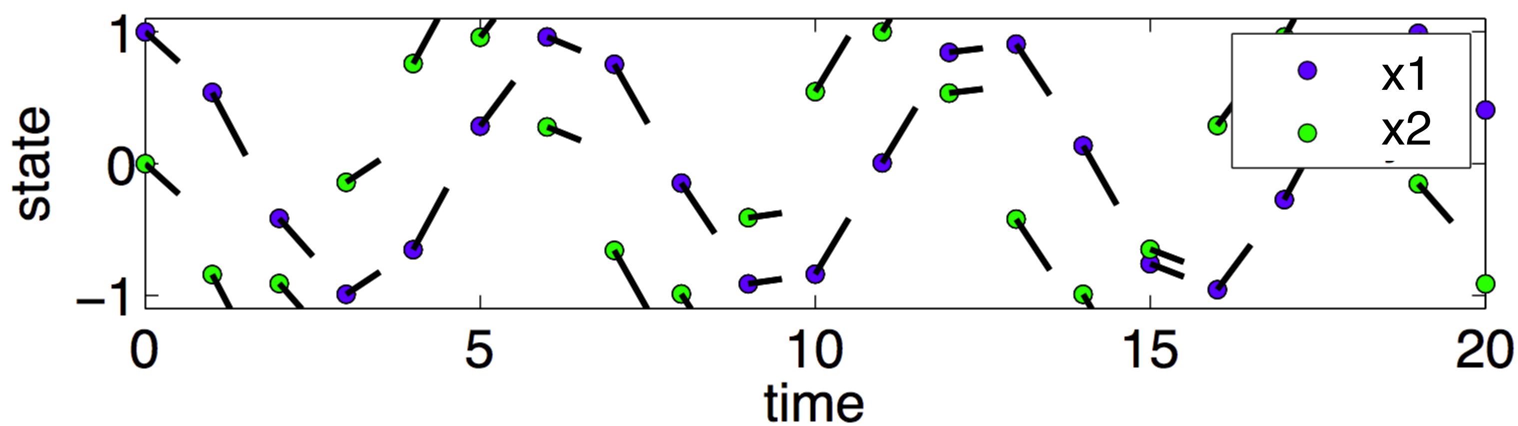
# Outline

- A "traditional" formulation of the parameter-fitting problem
  - A highly nonlinear problem / analytical intractability!
  - Numerical methods & examples
  - Comments on noise and observability
- "Nontraditional" regression formulation of the problem
  - Advantages and disadvantages
  - Brief discussion of examples

# A "nontraditional" formulation of the model fitting problem

- Suppose we observed not just $x(t)$ but also the time derivatives $\dot{x}(t)$
- Then we could ask the model to reproduce the observed time derivatives, rather than the trajectory:

$$E(\theta) = \frac{1}{TM} \sum_{i=1}^{T} \sum_{j=1}^{M} (\dot{x}_j(t_i) - f_j(x(t_i)|\theta))^2$$




- Of course, we don't usually observe the $\dot{x}(t)$, but we can estimate them!

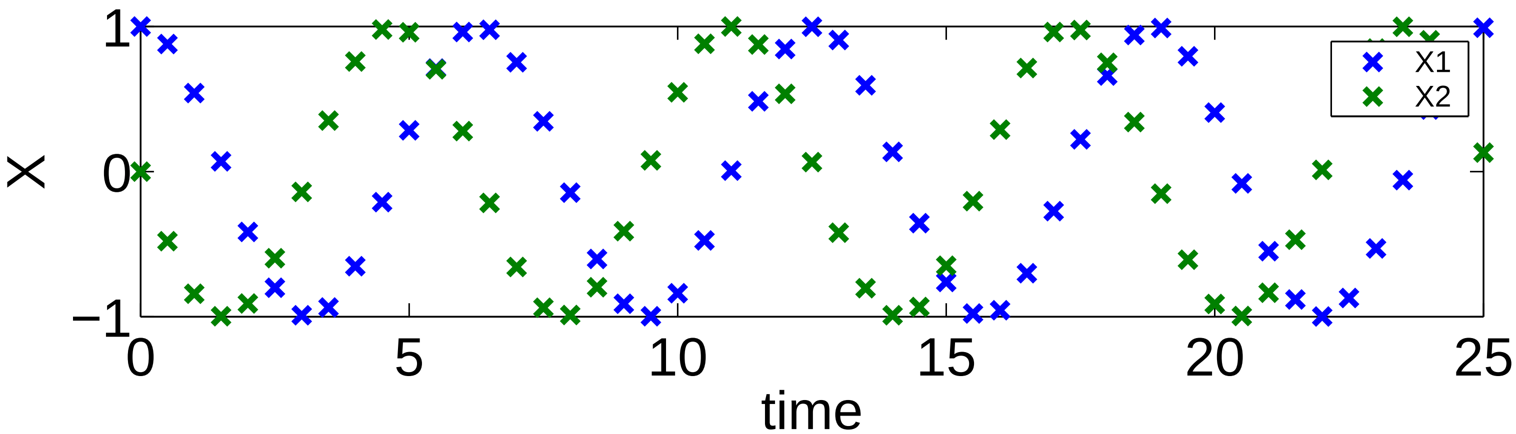
# Some advantages of derivative-based error

$$E(\theta) = \frac{1}{TM} \sum_{i=1}^{T} \sum_{j=1}^{M} (\dot{x}_j(t_i) - f_j(x(t_i)|\theta))^2$$

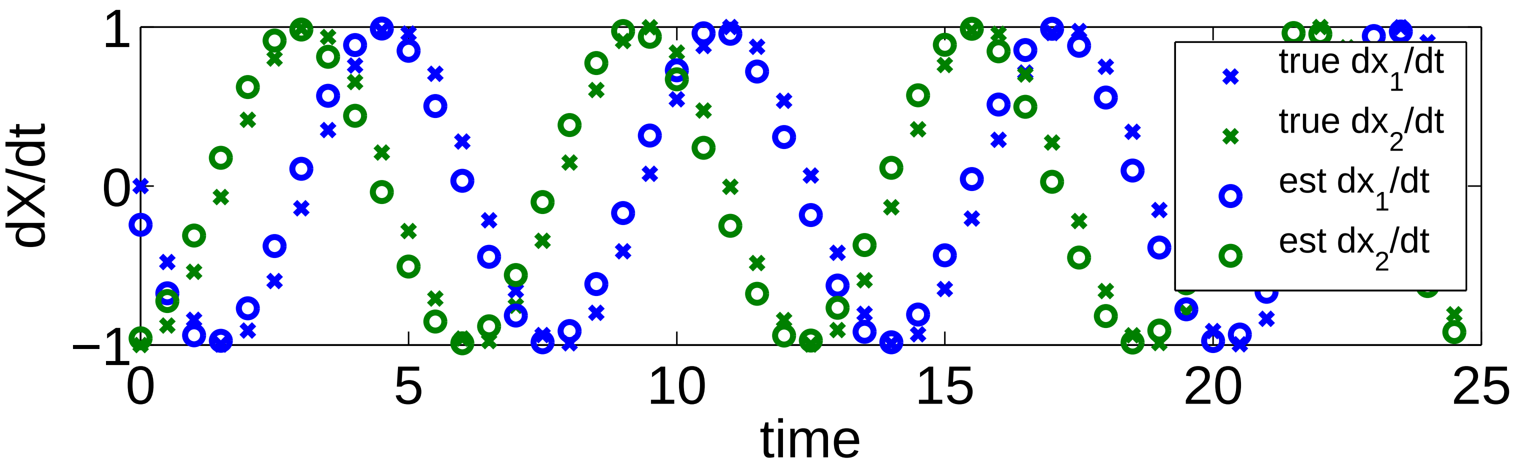
- It is readily evaluated, without solving a differential equation.
- If $f$ is differentiable, then we can differentiate $E$ w.r.t. $\theta$.
- We may even be able to minimize $E$ analytically.

## Finite difference estimates of derivatives

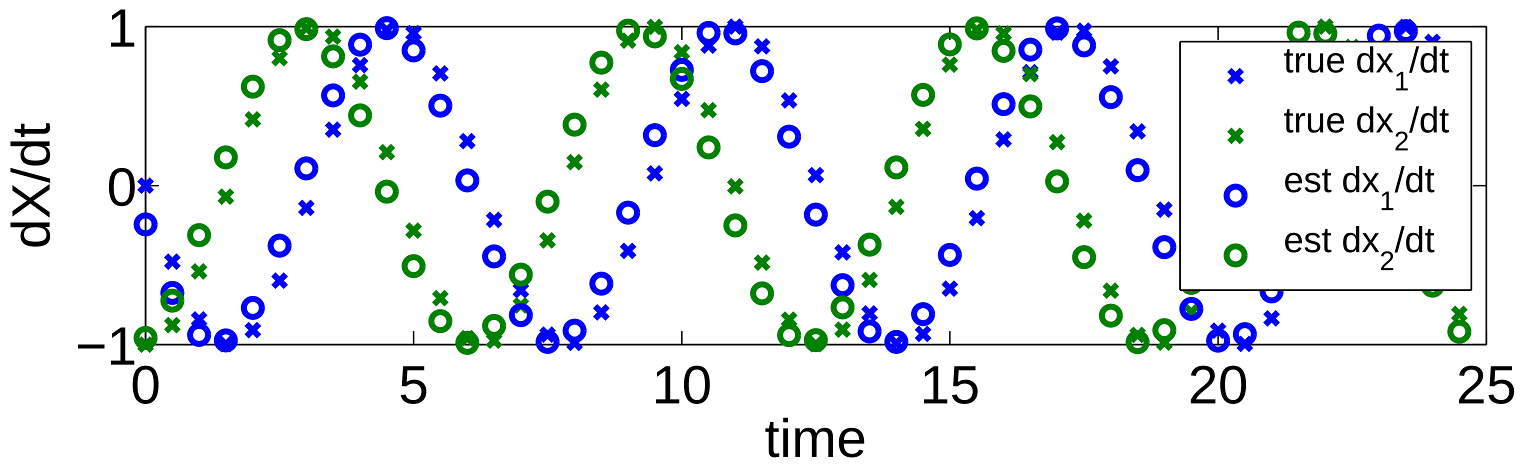
Original data:




Derivative estimates, $\dot{x}_i(t_j) = (x_i(t_{j+1}) - x_i(t_j))/(t_{j+1} - t_j)$:



# Finite difference estimates of derivatives

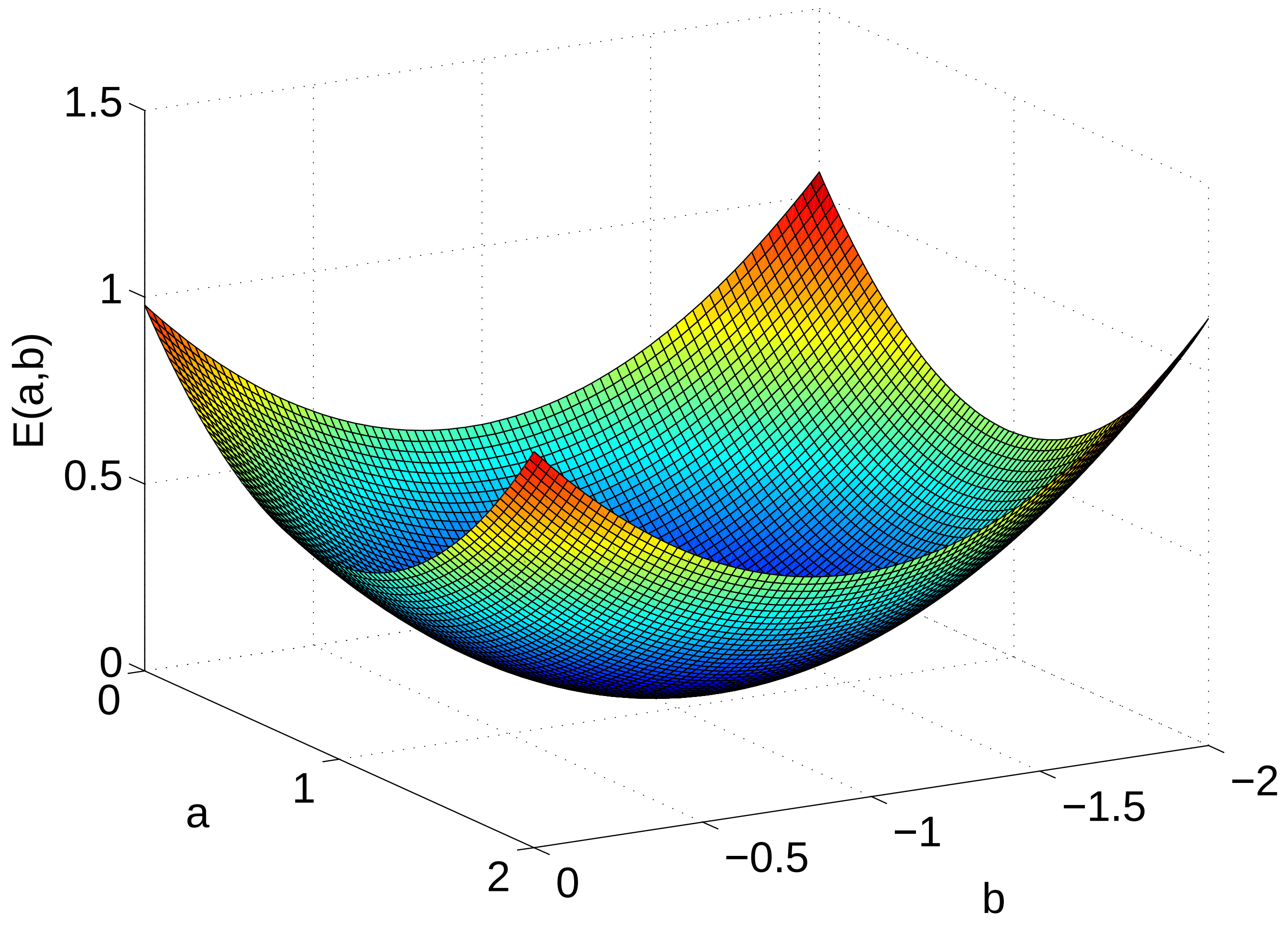



Minimizing

$$\frac{1}{TM}\sum_{i=1}^{T}\left\|\begin{bmatrix}\hat{\dot{x}}_1(t_i)\\ \hat{\dot{x}}_2(t_i)\end{bmatrix}-\begin{bmatrix}0 & a\\ b & 0\end{bmatrix}\begin{bmatrix}x_1(t_i)\\ x_2(t_i)\end{bmatrix}\right\|^2$$

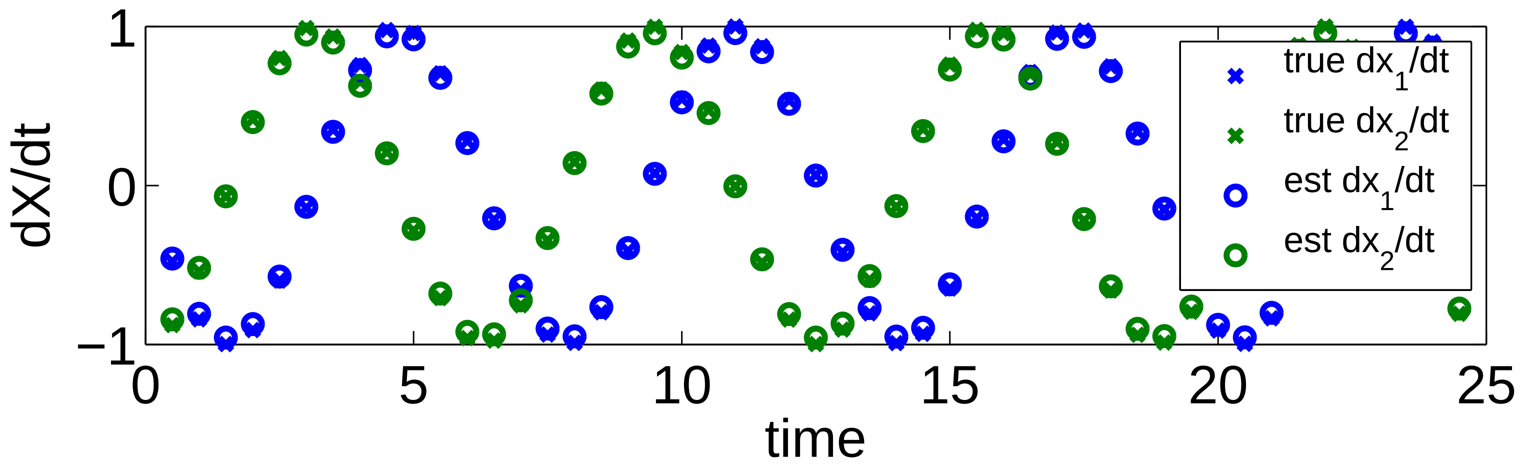
can be done analytically, yielding $a = 0.9597$ and $b = -0.9581$.

# The error surface



# Central difference estimates of derivatives

Estimating as $\dot{x}_i(t_j) = (x_i(t_{j+1}) - x_i(t_{j-1}))/(t_{j+1} - t_{j-1})$:

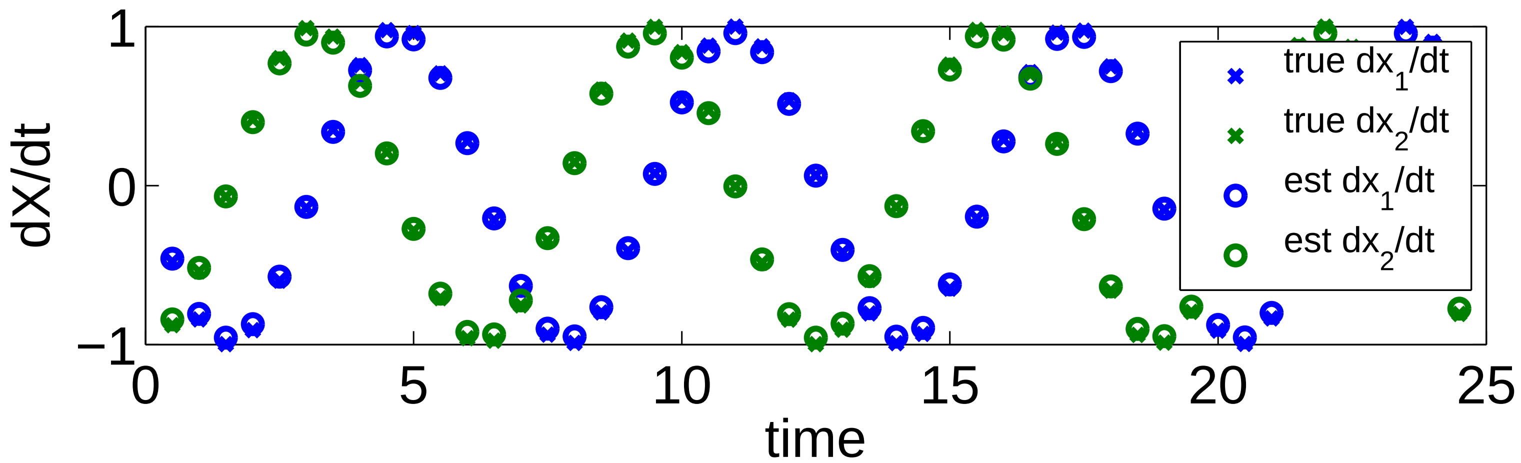



Minimizing

$$\frac{1}{TM}\sum_{i=1}^{T}\left\|\begin{bmatrix}\hat{\dot{x}}_1(t_i)\\ \hat{\dot{x}}_2(t_i)\end{bmatrix} - \begin{bmatrix}0 & a\\ b & 0\end{bmatrix}\begin{bmatrix}x_1(t_i)\\ x_2(t_i)\end{bmatrix}\right\|^2$$

yields $a = 0.9589$ and $b = -0.9589$.

# Fitting full interconnect matrix

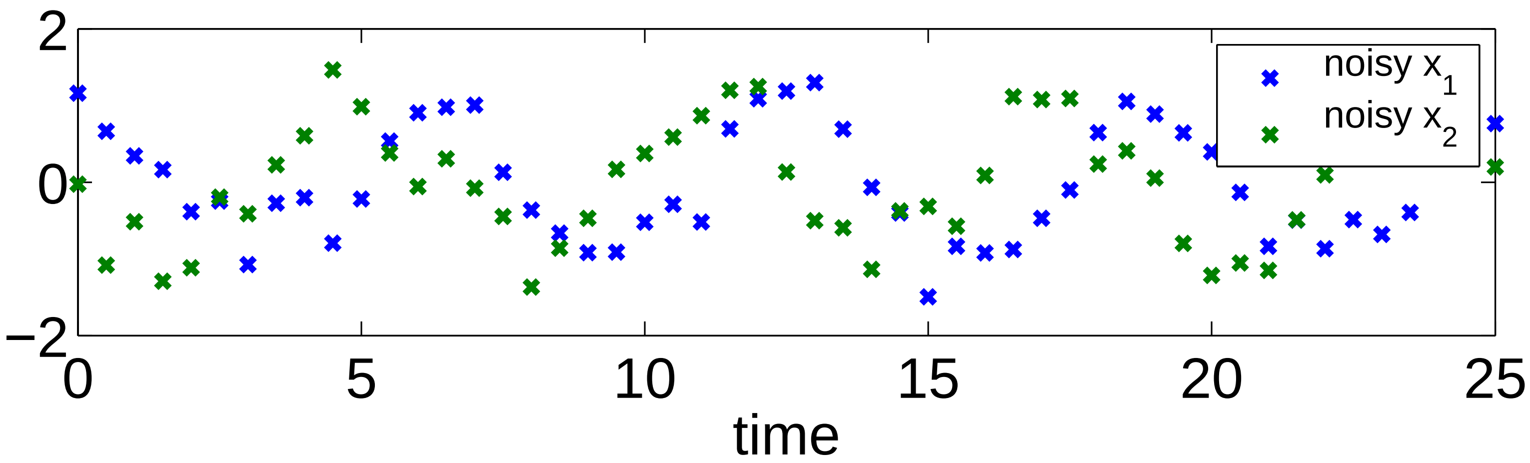


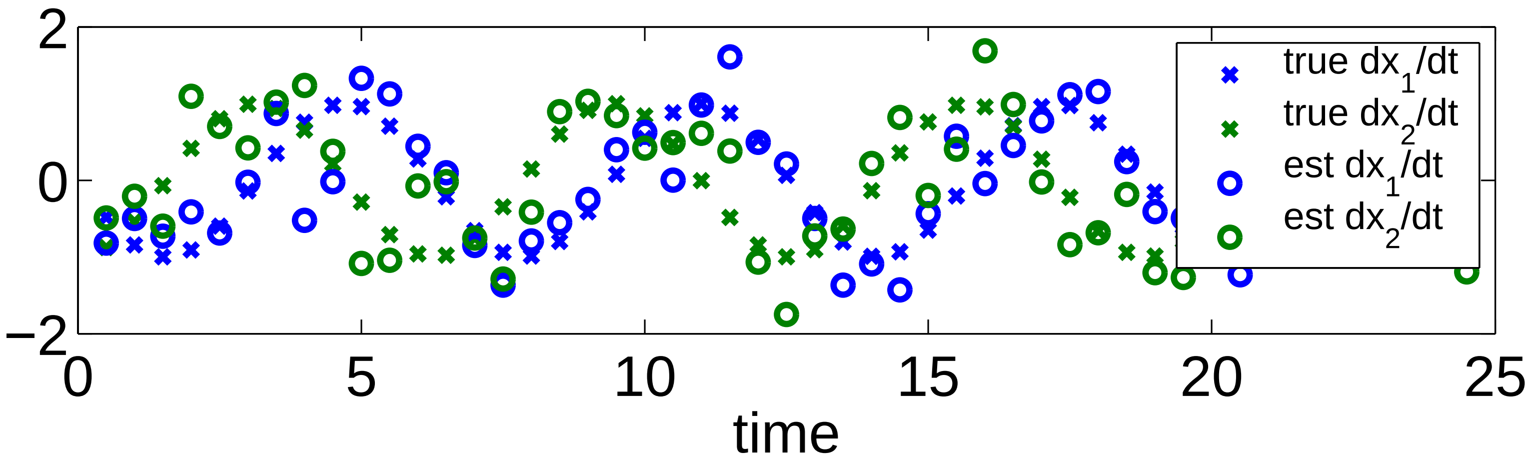

Minimizing

$$\frac{1}{TM}\sum_{i=1}^{T}\left\| \begin{bmatrix} \hat{\dot{x}}_1(t_i) \\ \hat{\dot{x}}_2(t_i) \end{bmatrix} - \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1(t_i) \\ x_2(t_i) \end{bmatrix} \right\|^2$$

yields $A = \begin{bmatrix} -0.00003 & 0.9589 \\ -0.9589 & -0.00002 \end{bmatrix}$.

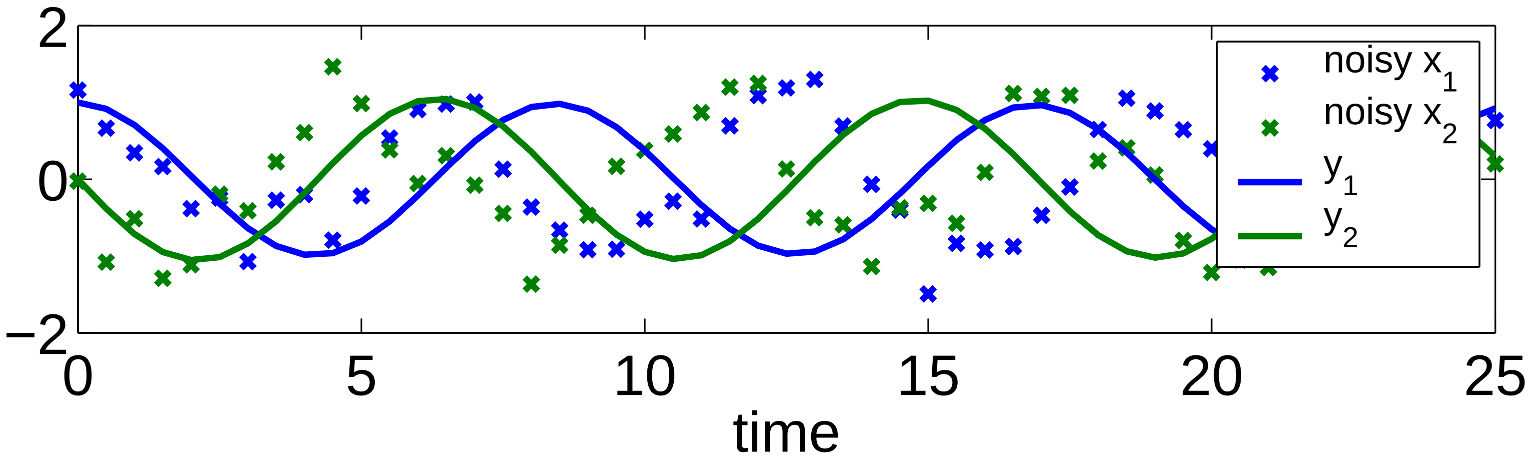
## What if the data are noisy?



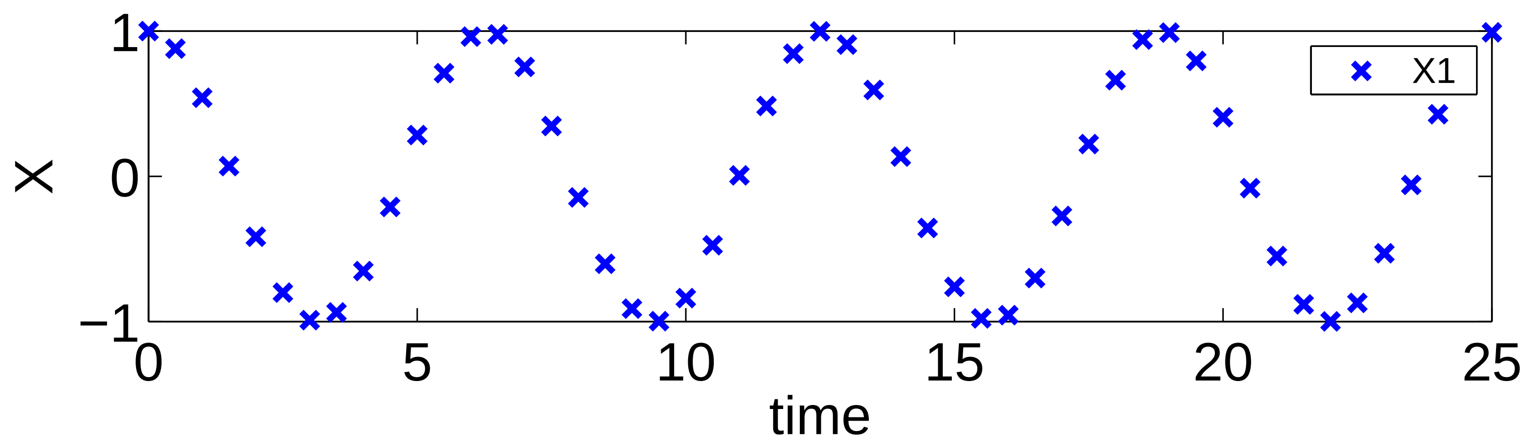
# Fitting a full interconnect matrix

Yields $A = \begin{bmatrix} -0.0337 & 0.6994 \\ -0.7886 & 0.0298 \end{bmatrix}$




A general caveat: Note that the simulated model diverges from the real trajectory.
(This doesn't have to do just with noise.)

# Caveat: What if we only observe $x_1$?




- We don't have $x_2$ to predict $\dot{x}_1$!
- We can't estimate $\dot{x}_2$ to resolve $x_1$'s influence!

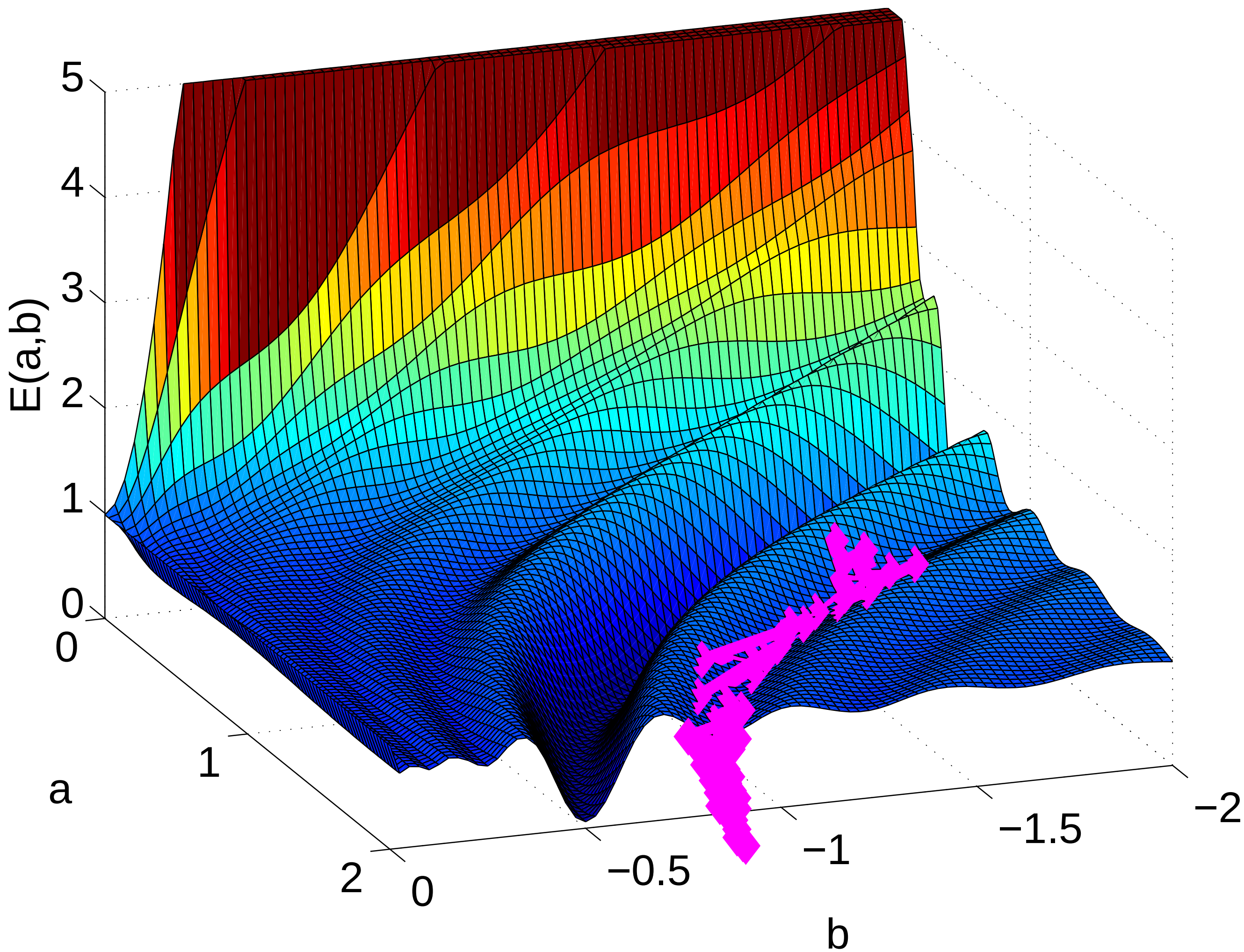
# Select literature examples

- Functional Data Analysis community in statistics (e.g., Ramsay & Silverman, 1997)
- D'Haeseleer *et al.* (PSB,1999) essentially used finite differences to fit a linear ODE model to central nervous system gene expression time series.
- Perkins *et al.* (PLoS Comp Bio, 2006) used a hybrid approach to fit the gap gene data. A regression approach was used to initialize parameter estimates; trajectory-based fitting was used to tune parameters and ensure good fit between simulated and observed data.
- Summer & Perkins (BMC Genetics, 2010) used spatio-temporal smoothing plus logistic regression to fit the gap gene data. Because it's fast, we explicitly evaluated all possible network structures.

## Conclusions

- "Traditional" trajectory-based formulation of model-fitting results in highly nonlinear, difficult optimization problems.
- Numerical methods for fitting to trajectories are subject to local optimality (grad. descent, Newton, fminsearch, local search) or are computationally intensive (simulated annealing).
- However, they apply most generally, including when some model variables are not observed and/or data are noisy.
- "Nontraditional" approaches based on derivative-fitting in a regression framework, including functional data analysis, are computationally efficient, and sometimes analytically solvable.
- However, they only apply (easily) when all model variables are observed. When the model is simulated, it may not match the observed trajectory well.

# `fminsearch` **from starting point** $a = 1, b = -1.5$



# `fminsearch` **from starting point** $a = 2, b = -0.5$