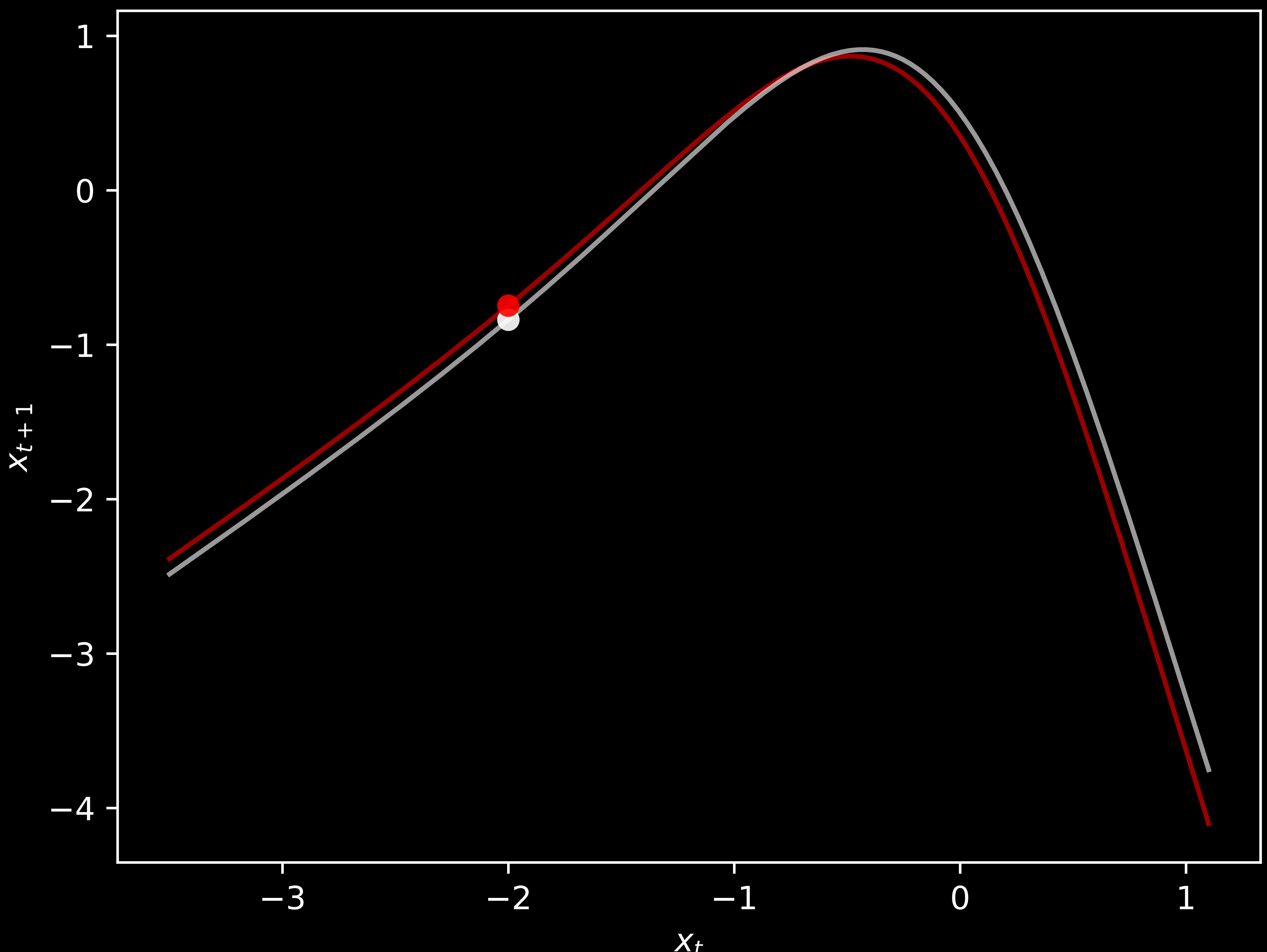
Overcoming Mean-Field Approximations in Recurrent Gaussian Process Models

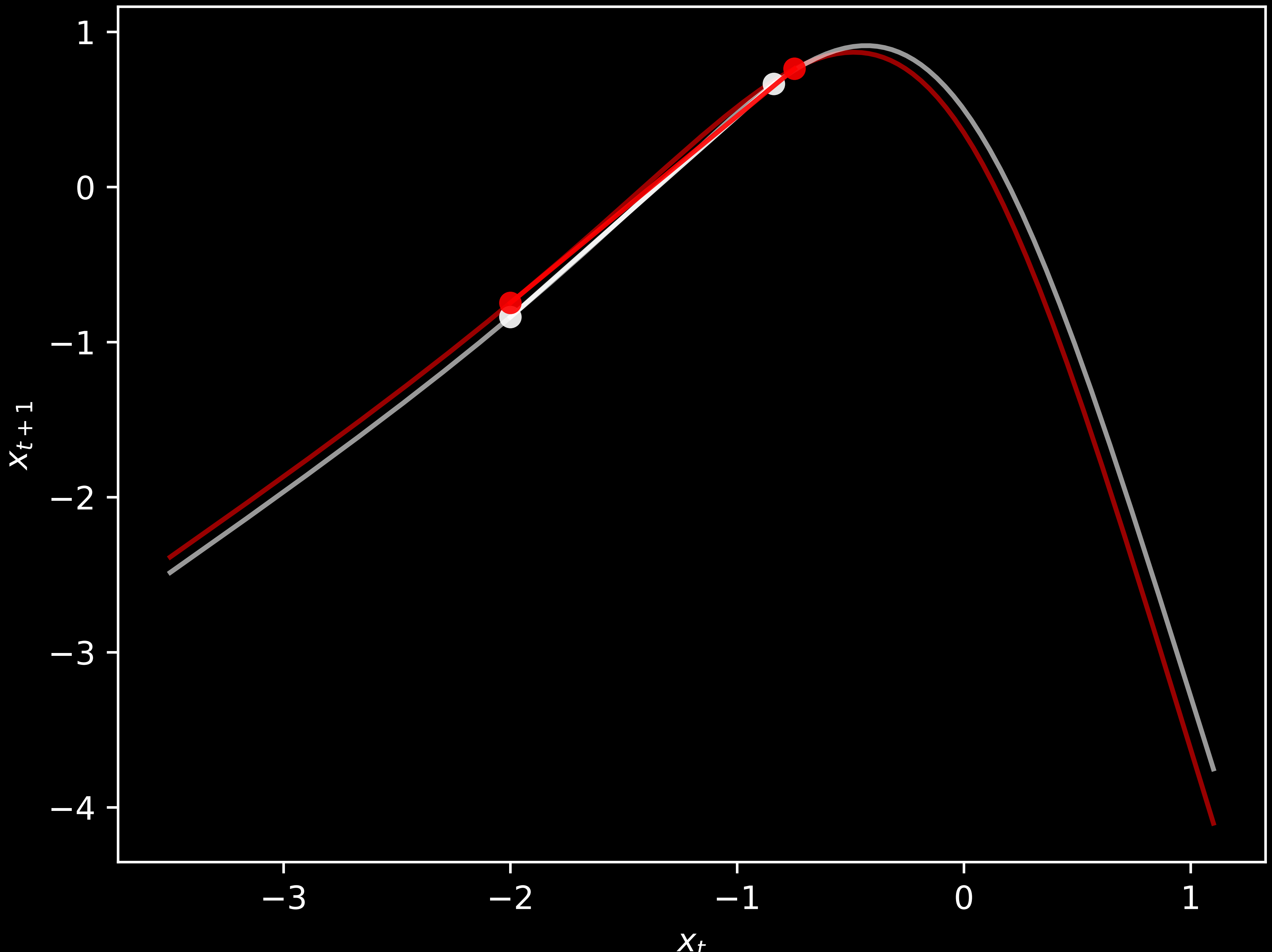
Alessandro Davide Ialongo
Mark van der Wilk
James Hensman
Carl Edward Rasmussen

UNIVERSITY OF CAMBRIDGE
MAX-PLANCK-GESELLSCHAFT
PROWLER.io the decision company

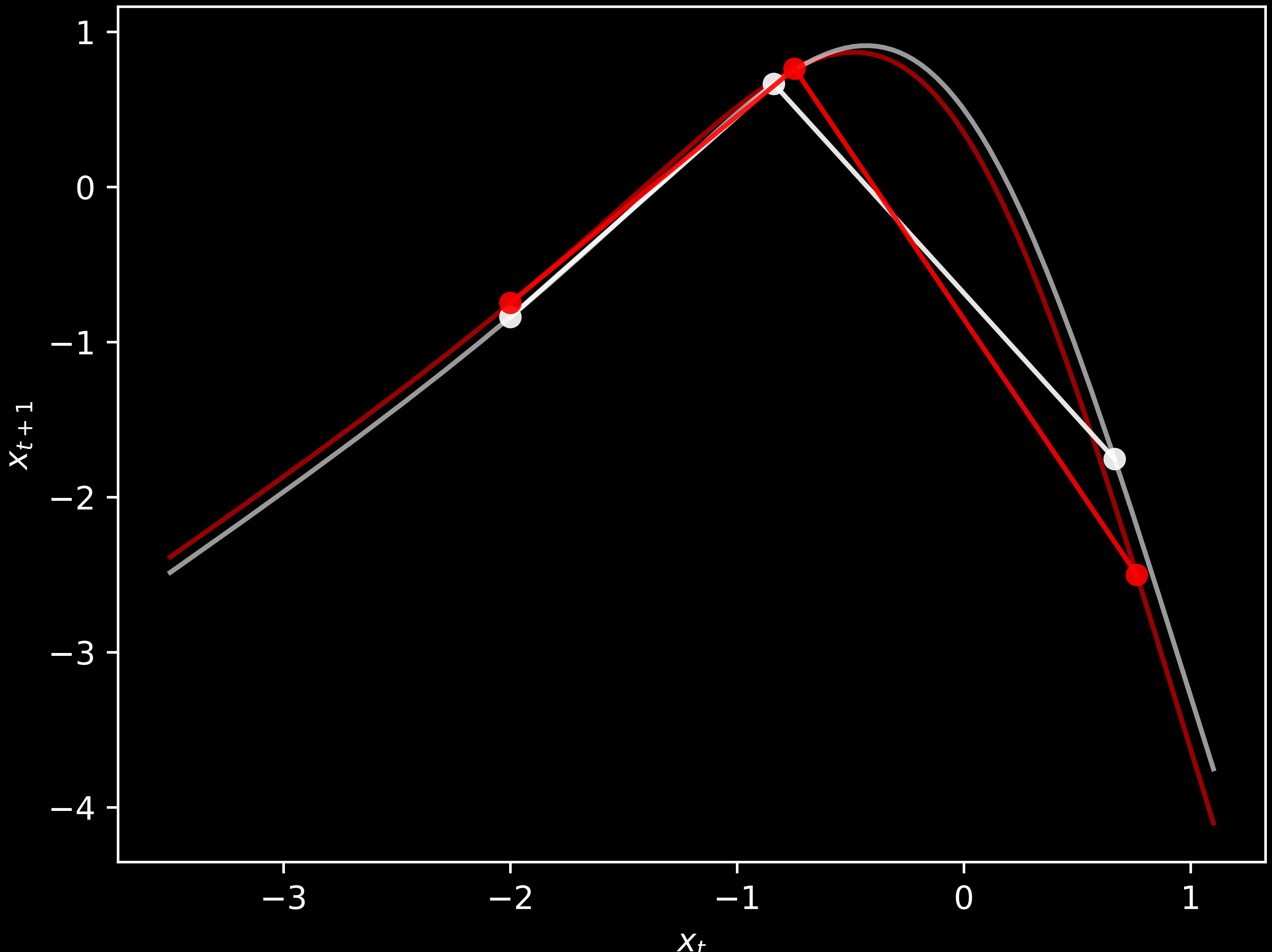
A motivating example



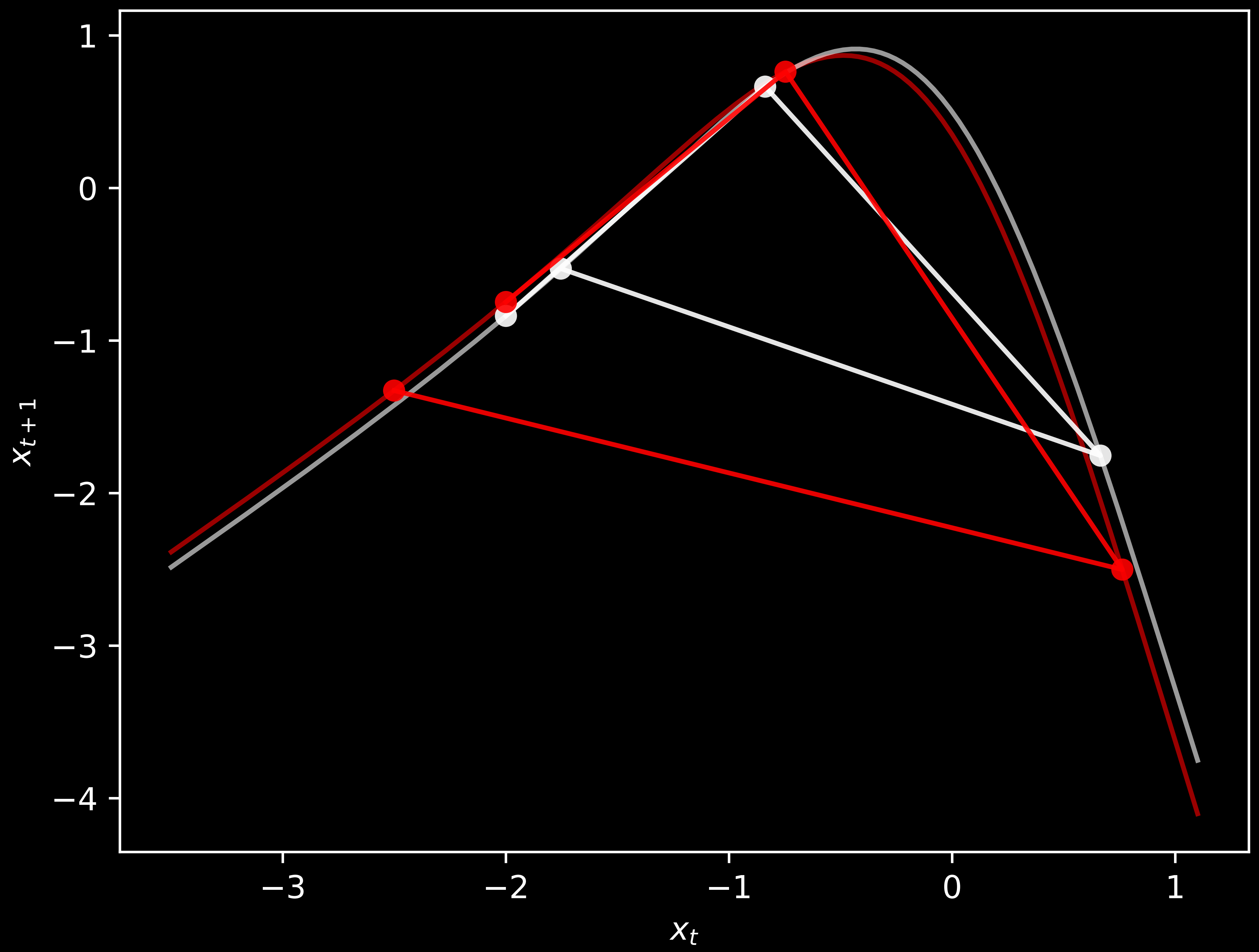
A motivating example



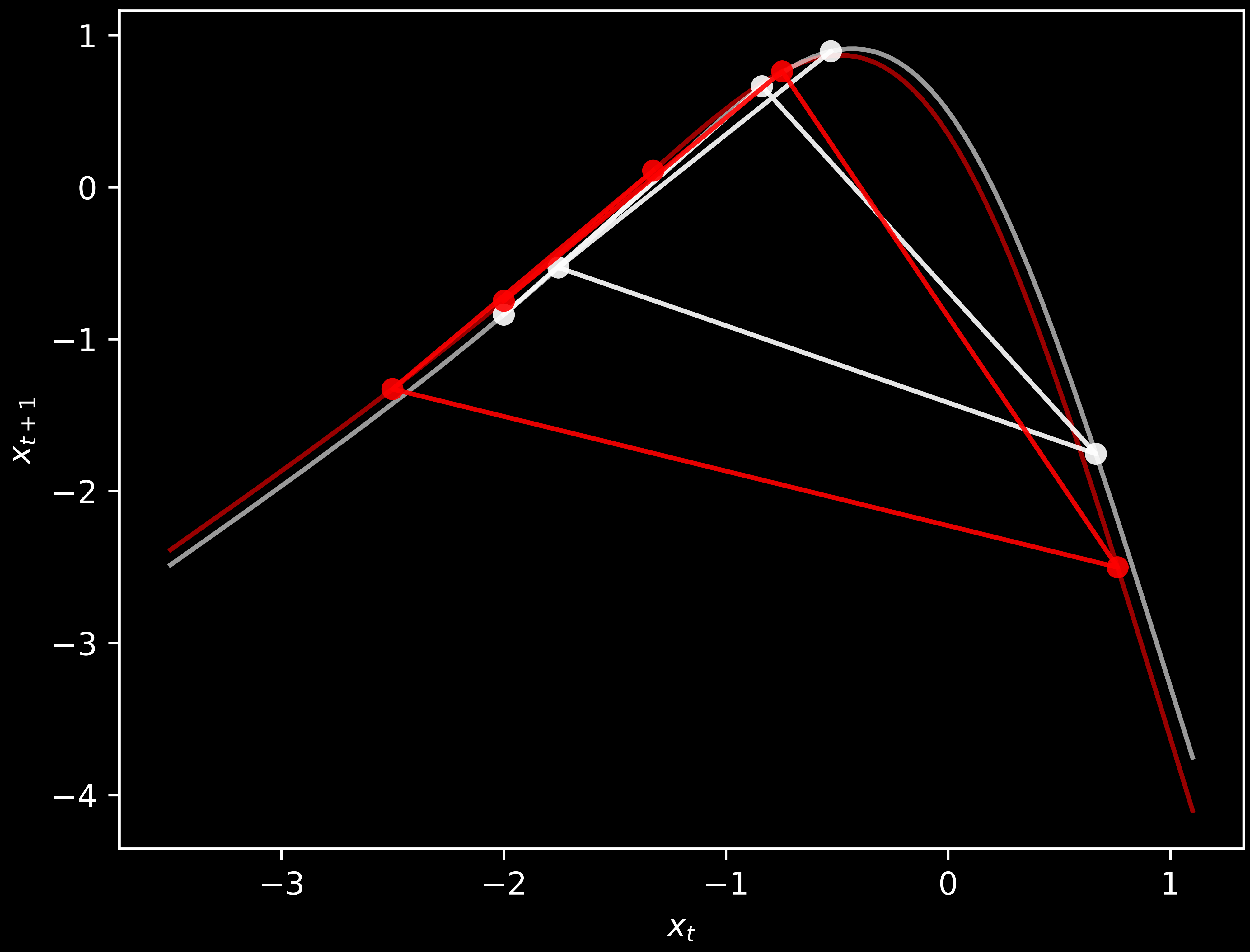
A motivating example



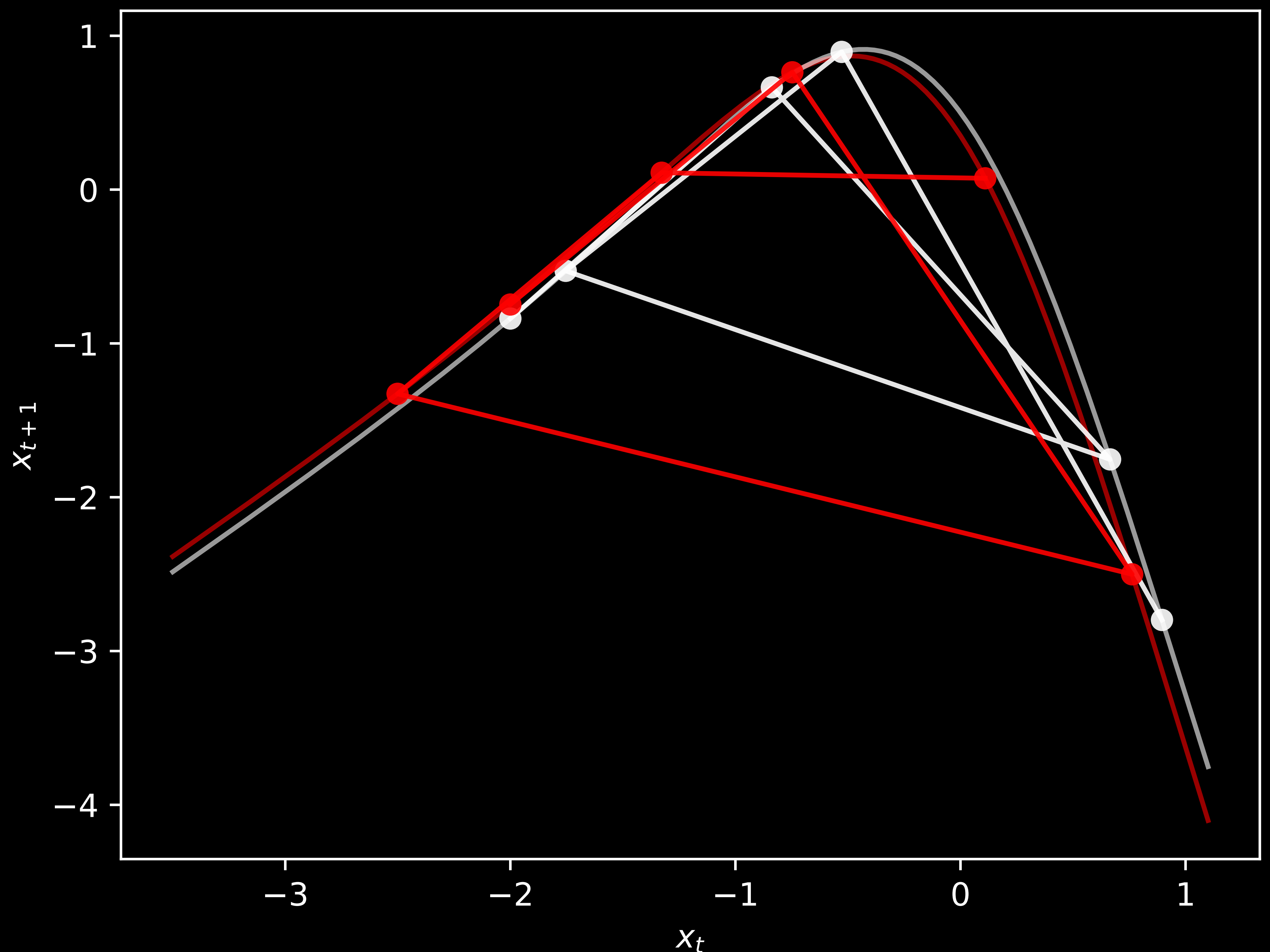
A motivating example



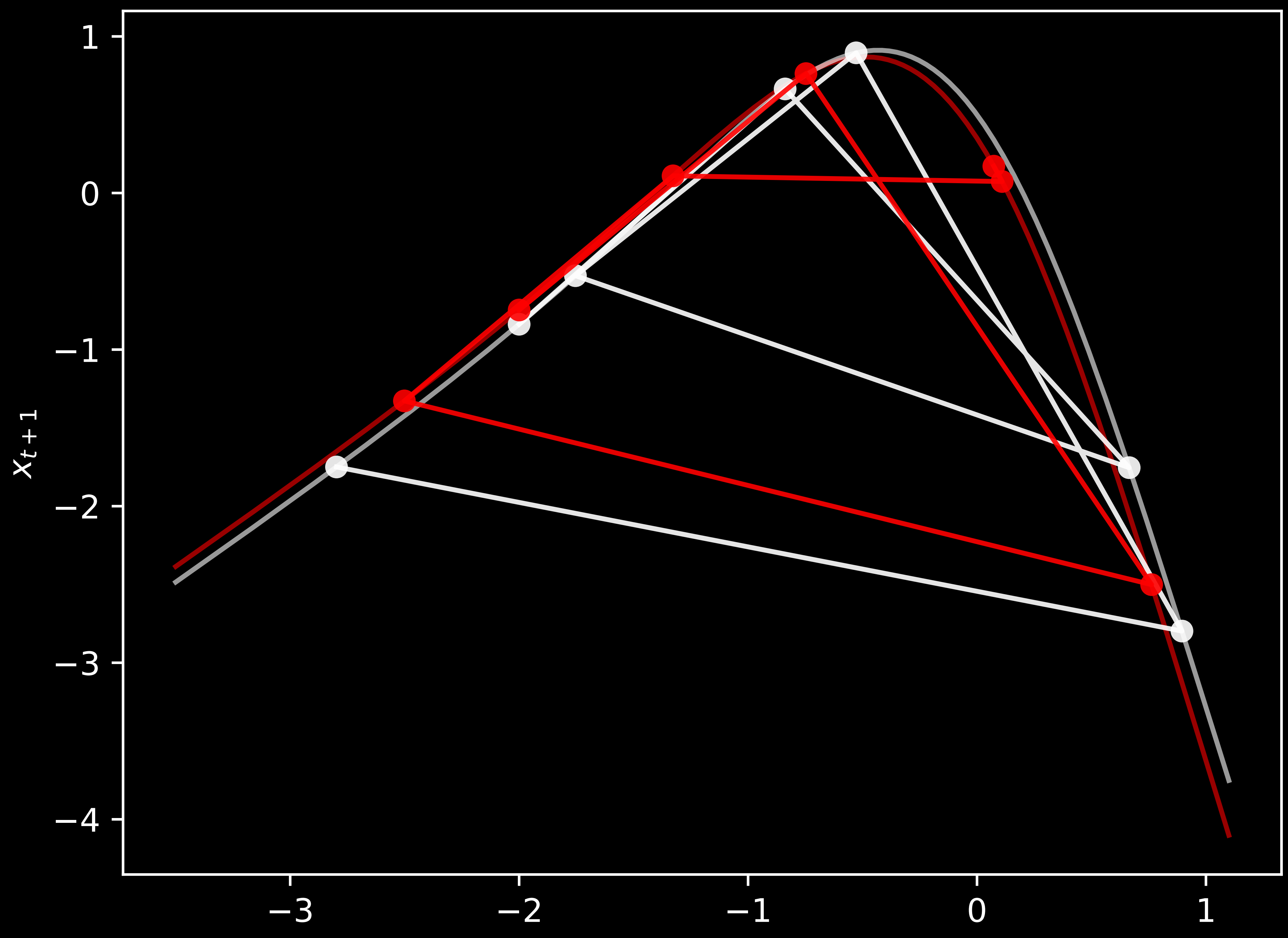
A motivating example



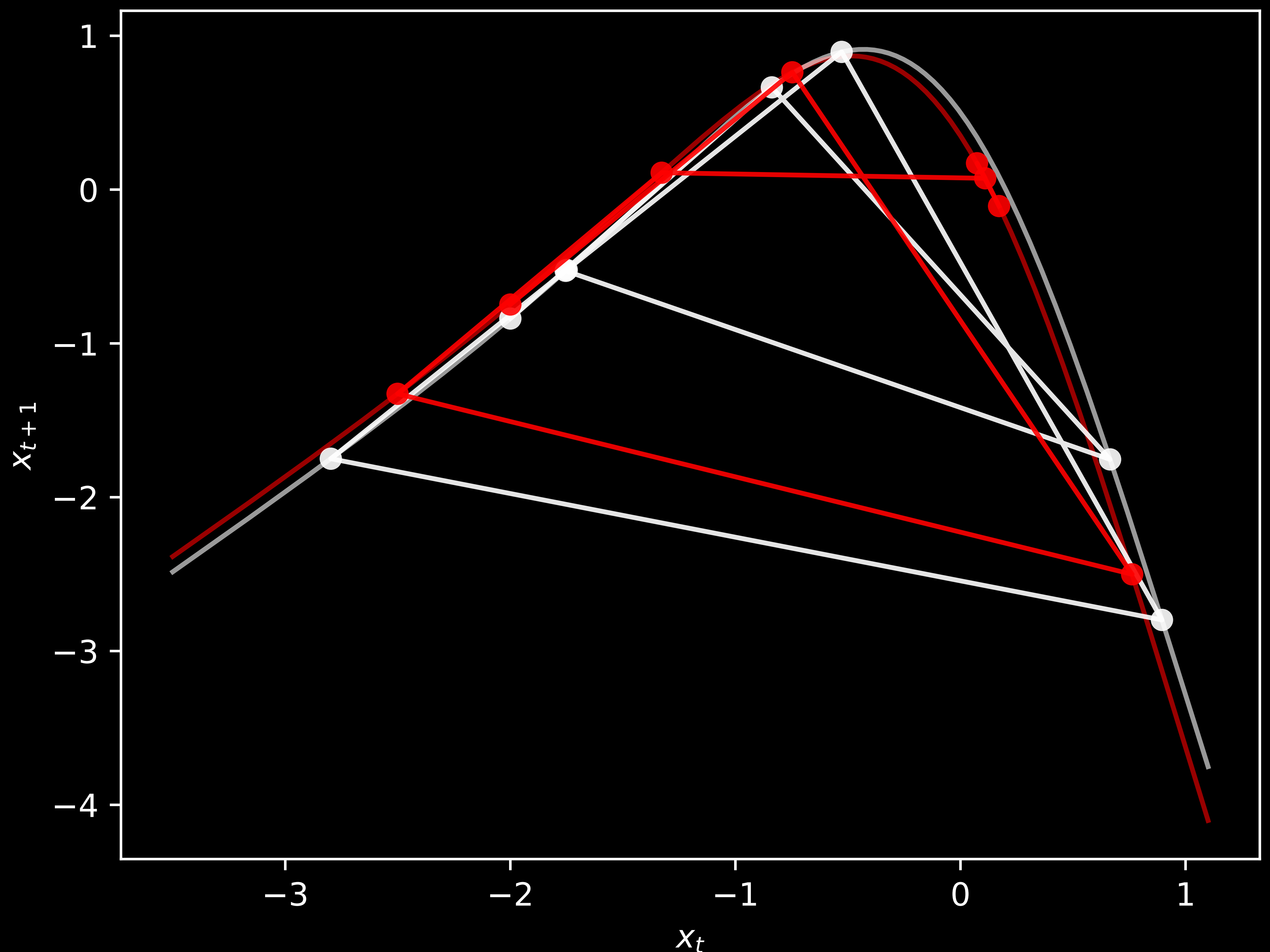
A motivating example



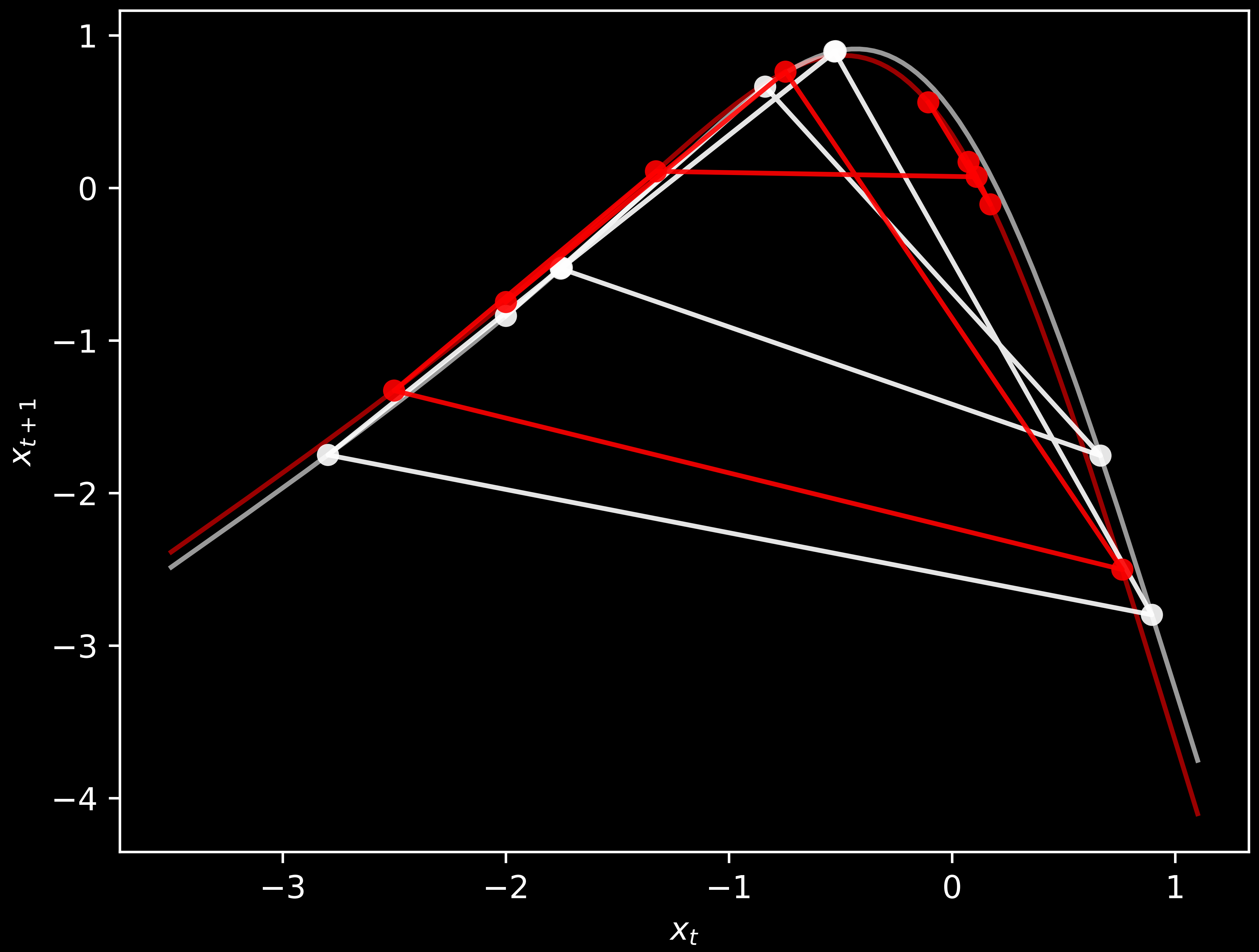
A motivating example

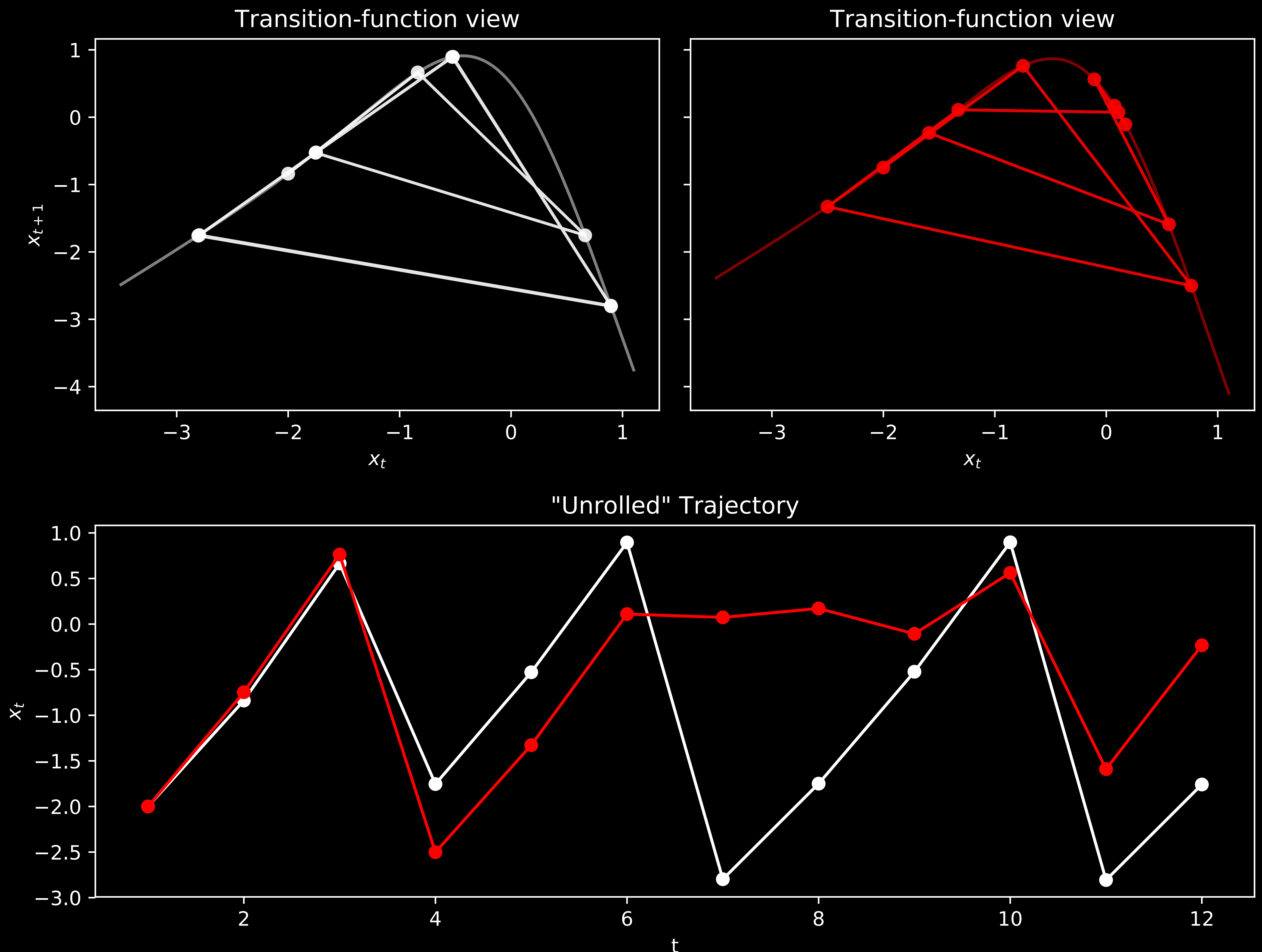


A motivating example



A motivating example





Transition-function view
Transition-function view
x_{t+1}
x_t
x_t
"Unrolled" Trajectory
x_t
t

GP State-Space Model

$$f \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot))$$

$$\mathbf{x}_1 \sim \mathcal{N}\left(\boldsymbol{\mu}_{p_1}, \boldsymbol{\Sigma}_{p_1}\right)$$

$$\mathbf{x}_{t+1} \,|\, f, \mathbf{x}_t \sim \mathcal{N}(f(\mathbf{x}_t), \mathbf{Q})$$

$$\mathbf{y}_t \,|\, \mathbf{x}_t \sim \mathcal{N}(\mathbf{C}\mathbf{x}_t + \mathbf{d}, \mathbf{R})$$

Variational Inference

$$p(X|f,Y) = p(\mathbf{x}_1|f,Y)\prod_{t=2}^{T} p(\mathbf{x}_t|f(\mathbf{x}_{t-1}),\mathbf{y}_{t:T})$$

$$q(X \mid f) = q(\mathbf{x}_1)\prod_{t=1}^{T-1} q(\mathbf{x}_{t+1} \mid f, \mathbf{x}_t)$$

A Natural Posterior Choice

$$q(\mathbf{x}_{t+1} \mid f, \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t+1}; \mathbf{A}_t \mathbf{f}_t + \mathbf{b}_t, \mathbf{S}_t^*)$$

$$\mathbf{S}_t^* = (\mathbf{Q}^{-1} + \mathbf{C}^\intercal \mathbf{R}^{-1} \mathbf{C})^{-1}$$

$$\mathbf{A}_t = \mathbf{S}_t^* \mathbf{Q}^{-1}$$

$$\mathbf{b}_t = \mathbf{S}_t^* \mathbf{C}^\intercal \mathbf{R}^{-1} (\mathbf{y}_{t+1} - \mathbf{d})$$

$$p(\mathbf{x}_{t+1} | f, \mathbf{x}_t, \mathbf{y}_{1:t+1})$$

Avoiding the Cubic Sampling Cost

If: $\mathbf{f}_t = f(\mathbf{x}_t)$

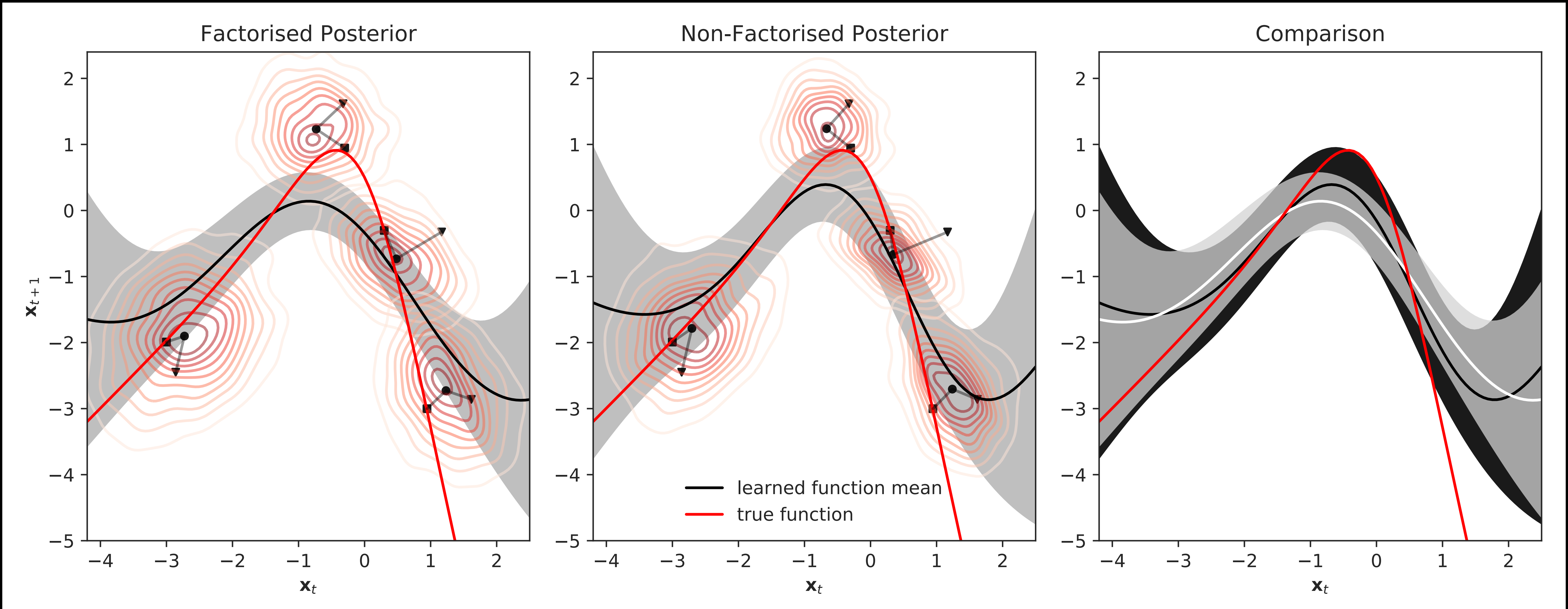
then, to sample, we need to condition on all previous f's.

So, take: $\mathbf{f}_t \approx K_{\mathbf{x}_t,Z} K_{Z,Z}^{-1} \mathbf{u}$

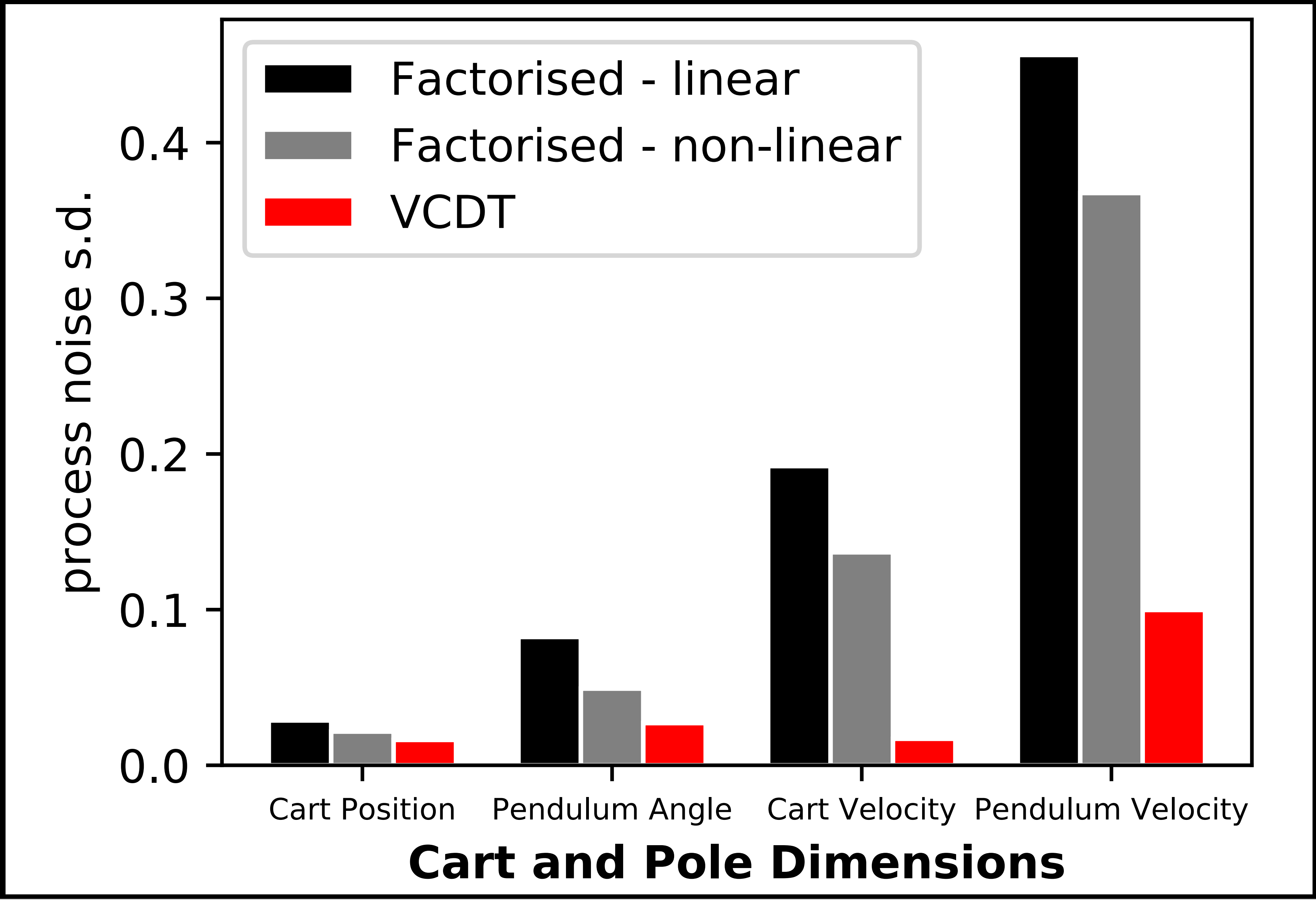
$$q(X \mid f) \approx q(X \mid \mathbf{u})$$

	$q(X \mid f)$	$\mathbf{f}_t$	$\mathbf{S}_t^*$
1) Factorised - linear	$q(X)$	$\mathbf{x}_t$	$\mathbf{S}_t$
2) Factorised - non-linear	$q(X)$	$K_{\mathbf{x}_t,Z} K_{Z,Z}^{-1} \boldsymbol{\mu}_{\mathbf{u}}$	$\mathbf{S}_t + \mathbf{A}_t \mathbf{C}_f(\mathbf{x}_t) \mathbf{A}_t^\top$
3) Non-Factorised - non-linear	$q(X \mid f)$	$f(\mathbf{x}_t)$	$\mathbf{S}_t$
4) VCDT	$q(X \mid \mathbf{u})$	$K_{\mathbf{x}_t,Z} K_{Z,Z}^{-1} \mathbf{u}$	$\mathbf{S}_t + \mathbf{A}_t \mathbf{C}_{f\mid\mathbf{u}}(\mathbf{x}_t) \mathbf{A}_t^\top$

Better Calibration



Better Predictive Performance



Model	NLPP	RMSE
Factorised - linear	2.268	3.548
Factorised - non-linear	1.847	2.974
VCDT	0.694	2.139