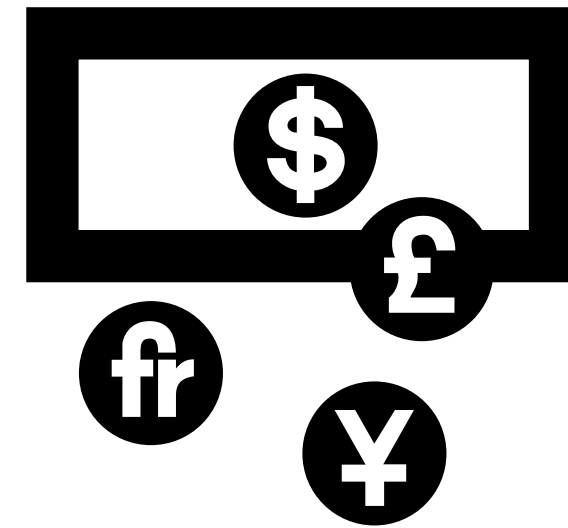
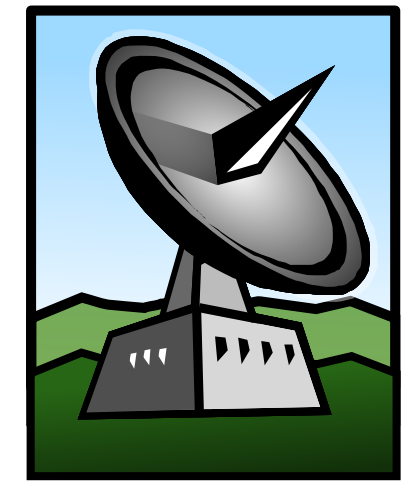
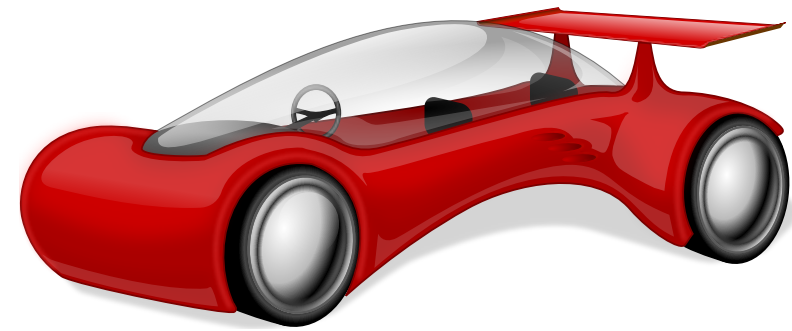
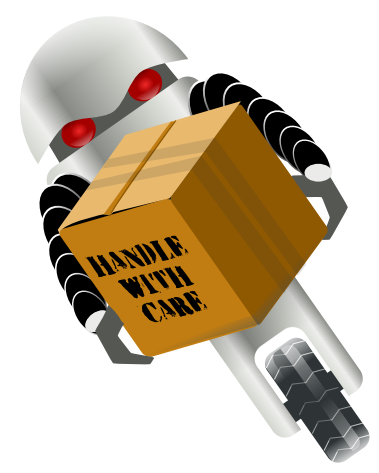


TIME SERIES EVERYWHERE



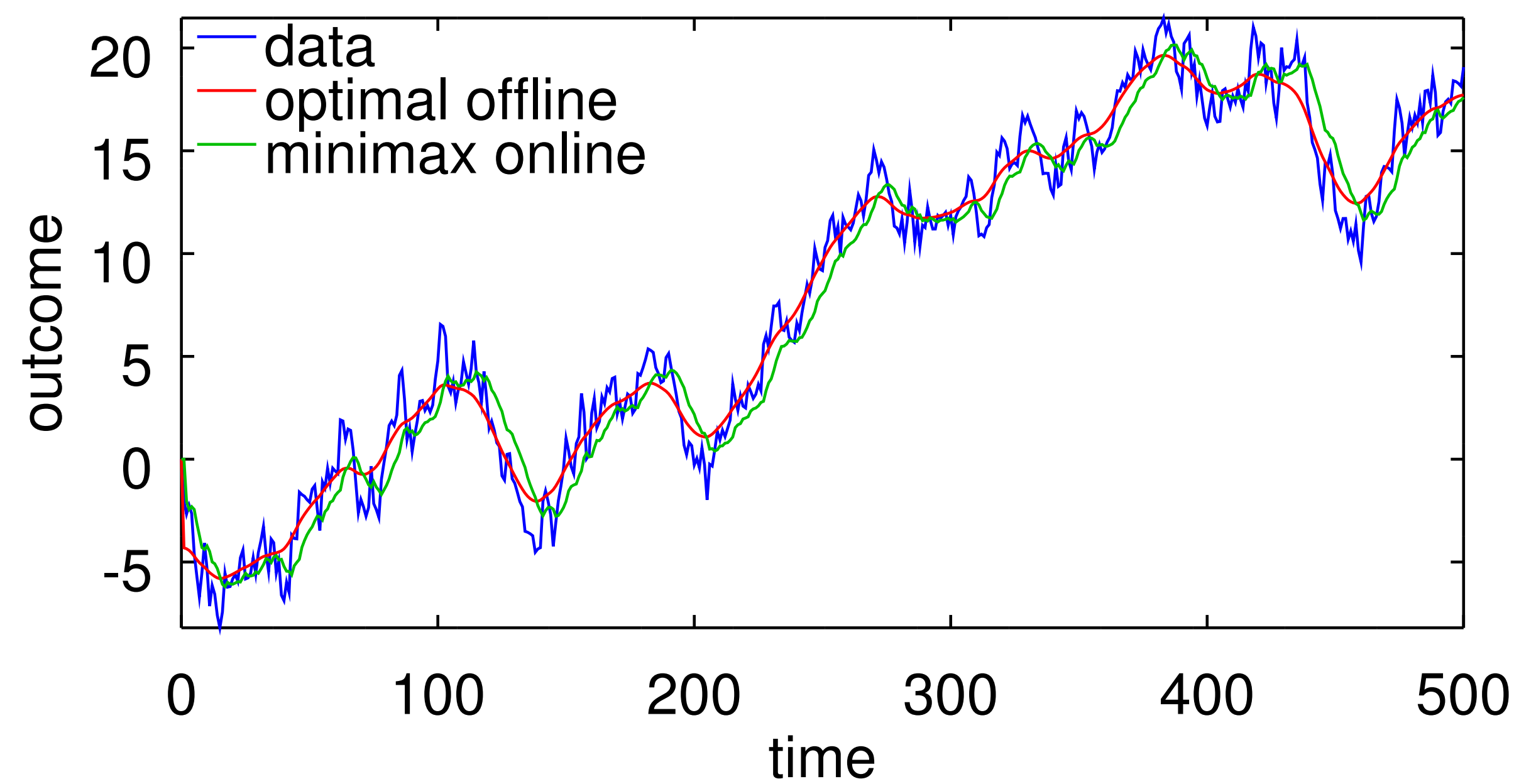
Pose estimation

Self-driving

Tracking

Finance

WHAT WE DO: ONLINE PREDICTION



MODEL: TIME SERIES GAME

For some convex set \mathcal{C} , each round $t = 1, \dots, T$

- We play $\mathbf{a}_t \in \mathcal{C}$
- Nature reveals $\mathbf{x}_t \in \mathcal{C}$
- We incur loss $\|\mathbf{a}_t - \mathbf{x}_t\|^2$

Fix horizon T and regularization scalar $\lambda_T > 0$. Regret is:

$$\underbrace{\sum_{t=1}^T \|\mathbf{a}_t - \mathbf{x}_t\|^2}_{\text{Our loss}} - \min_{\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_T} \left\{ \underbrace{\sum_{t=1}^T \|\hat{\mathbf{a}}_t - \mathbf{x}_t\|^2}_{\text{Loss of Comparator}} + \underbrace{\lambda_T \sum_{t=1}^{T+1} \|\hat{\mathbf{a}}_t - \hat{\mathbf{a}}_{t-1}\|^2}_{\text{Comparator Complexity}} \right\}$$

OBJECTIVE: MINIMAX REGRET

If we assume a perfect adversary, how well can we do? Value is

$$V := \min_{\mathbf{a}_1} \max_{\mathbf{x}_1: \|\mathbf{x}_1\| \leq 1} \dots \min_{\mathbf{a}_T} \max_{\mathbf{x}_T: \|\mathbf{x}_T\| \leq 1} \text{Regret}$$

We play to minimize the worst-case regret; e.g. find \mathbf{a}_t that guarantees we achieve the game's value.

IN GENERAL

Let $\mathbf{X}_t = [\mathbf{x}_1 \dots \mathbf{x}_t]$ and $\hat{\mathbf{A}} = [\hat{\mathbf{a}}_1 \dots \hat{\mathbf{a}}_T]$. For $\mathbf{v}_t \in \mathbb{R}^d$ and $\mathbf{K} \succeq \mathbf{0}$,

Data domain $\|\mathbf{X}_t \mathbf{v}_t\| \leq 1$ e.g. $\|\mathbf{x}_t\| \leq 1$

Complexity $\text{tr}(\mathbf{K} \hat{\mathbf{A}}^T \hat{\mathbf{A}})$ e.g. $\sum_{t=1}^{T+1} \|\hat{\mathbf{a}}_t - \hat{\mathbf{a}}_{t-1}\|^2$

Spectrum of games (in particular: higher order differences)

OFFLINE PROBLEM

Theorem 1 For any complexity matrix $\mathbf{K} \succeq \mathbf{0}$, regularization scalar $\lambda_T \geq 0$, and $d \times T$ data matrix $\mathbf{X}_T = [\mathbf{x}_1 \dots \mathbf{x}_T]$ the problem

$$L^* := \min_{\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_T} \sum_{t=1}^T \|\hat{\mathbf{a}}_t - \mathbf{x}_t\|^2 + \lambda_T \text{tr}(\mathbf{K} \hat{\mathbf{A}}^T \hat{\mathbf{A}})$$

has linear (in \mathbf{X}_T) minimizer and quadratic (in \mathbf{X}_T) value given by

$$\hat{\mathbf{A}} = \mathbf{X}_T (\mathbf{I} + \lambda_T \mathbf{K})^{-1} \quad \text{and} \quad L^* = \text{tr}(\mathbf{X}_T (\mathbf{I} - (\mathbf{I} + \lambda_T \mathbf{K})^{-1}) \mathbf{X}_T^T).$$

BACKWARD INDUCTION SOLUTION

We solve for the **value-to-go** V from each state $\mathbf{X}_t = [\mathbf{x}_1 \dots \mathbf{x}_t]$. We have $V(\mathbf{X}_T) := -L^*$ and

$$V(\mathbf{X}_{t-1}) := \min_{\mathbf{a}_t} \max_{\mathbf{x}_t: \|\mathbf{x}_t \mathbf{v}_t\| \leq 1} \|\mathbf{a}_t - \mathbf{x}_t\|^2 + V(\mathbf{X}_t).$$

The minimax regret V equals value-to-go $V(\epsilon)$ from empty history.

CRUX

Value-to-go V stays quadratic in \mathbf{X}_t for all $t \leq T$ and corresponding minimax strategy is linear in \mathbf{X}_{t-1} . Remains to compute coefficients.

SINGLE-SHOT SQUARED LOSS GAME

Theorem 2 If $\|\mathbf{b}\| \leq 1$, then the minimax problem

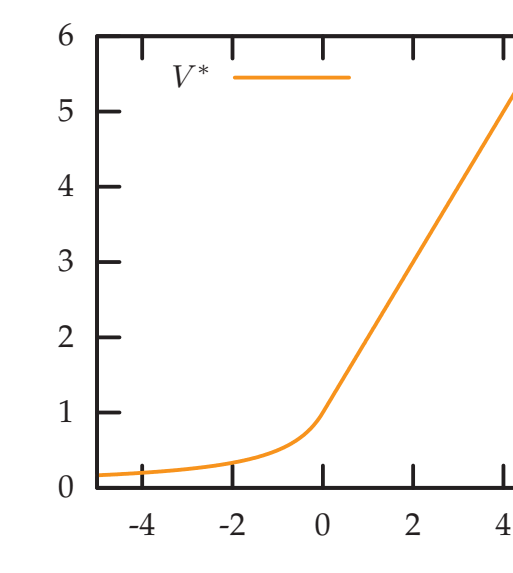
$$V^* := \min_{\mathbf{a}} \max_{\mathbf{x}: \|\mathbf{x}\| \leq 1} \|\mathbf{a} - \mathbf{x}\|^2 + (\alpha - 1) \|\mathbf{x}\|^2 + 2\mathbf{b}^T \mathbf{x}$$

has value and minimizer

$$V^* = \begin{cases} \frac{\|\mathbf{b}\|^2}{1-\alpha} & \text{if } \alpha \leq 0, \\ \|\mathbf{b}\|^2 + \alpha & \text{if } \alpha \geq 0, \end{cases} \quad \text{and} \quad \mathbf{a} = \begin{cases} \frac{\mathbf{b}}{1-\alpha} & \text{if } \alpha \leq 0, \\ \mathbf{b} & \text{if } \alpha \geq 0. \end{cases}$$

Non-trivial induction:

- Curvature of optimization can switch between rounds
- Yet can pre-compute beforehand



MINIMAX TIME SERIES PREDICTION

We precompute:

Input: $T, \mathbf{K}, \lambda_T, \mathbf{v}_1, \dots, \mathbf{v}_T$



Output: matrices $\mathbf{R}_t = \begin{pmatrix} \mathbf{A}_t & \mathbf{b}_t \\ \mathbf{b}_t^T & c_t \end{pmatrix}$

Theorem 3 Under a (typical) no clipping condition on \mathbf{X}_T ,

$$V(\mathbf{X}_t) = \text{tr}(\mathbf{X}_t (\mathbf{R}_t - \mathbf{I}) \mathbf{X}_t^T) + \sum_{s=t+1}^T \max\{c_s, 0\}$$

linear filter

$$\mathbf{a}_t = \mathbf{X}_{t-1} \begin{cases} \frac{\mathbf{b}_t}{1-c_t} & \text{if } c_t \leq 0, \\ \mathbf{b}_t - c_t \mathbf{v}_t^{<t} & \text{if } c_t \geq 0. \end{cases}$$

VANILLA CASE NORM-BOUNDED DATA WITH INCREMENT SQUARED REGULARIZATION

- Cheap scalar $O(T)$ preprocessing sweep.
- Predict in $O(d)$ time per round using $O(d)$ memory.
- Filter weights roughly decay exponentially backwards.
- Can upper and lower bound regret to get

$$V = \Theta\left(\frac{T}{\sqrt{1+\lambda_T}}\right).$$

constant λ_T
overfits

BUT WAIT, THERE'S MORE

- Computation: if \mathbf{K} and \mathbf{v}_t are banded then \mathbf{R}_t^{-1} is sparse
- Here we imposed data bound $\|\mathbf{X}_t \mathbf{v}_t\| \leq 1$. In the paper we show that the minimax strategy guarantees an *adaptive* bound scaling with $\|\mathbf{X}_t \mathbf{v}_t\|$.
- A second order smoothness version of \mathbf{K} gives complicated c_t

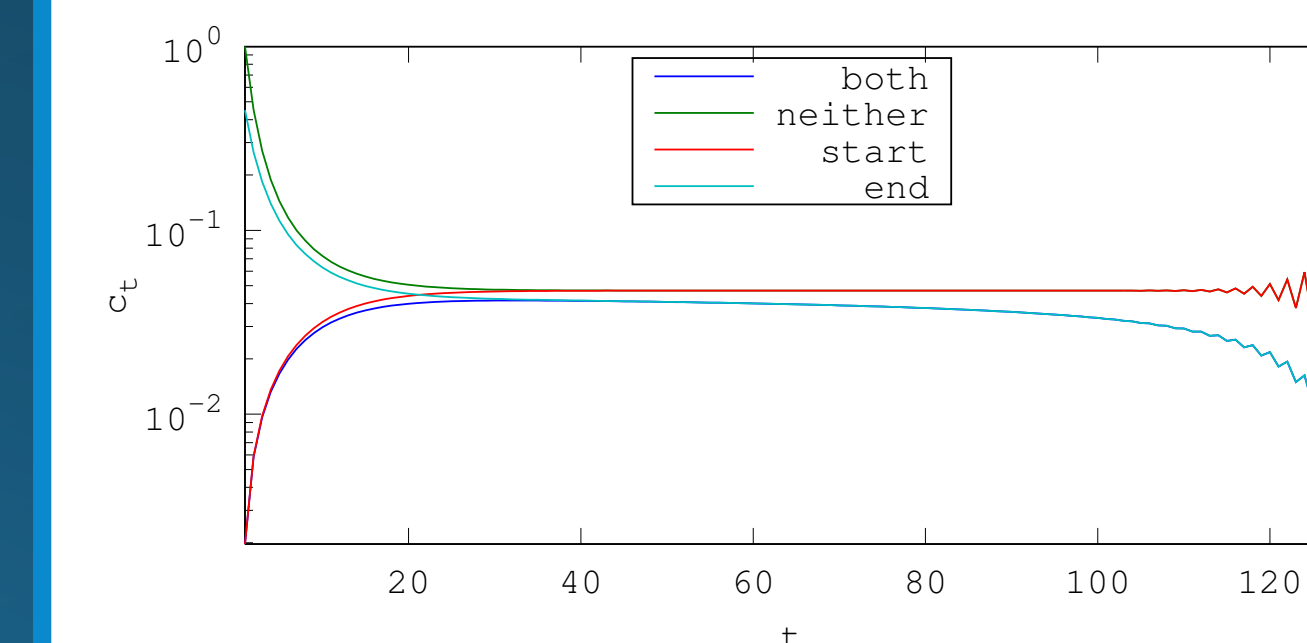


Figure 1: $\mathbf{v}_t = \mathbf{e}_t - \mathbf{e}_{t-1}$

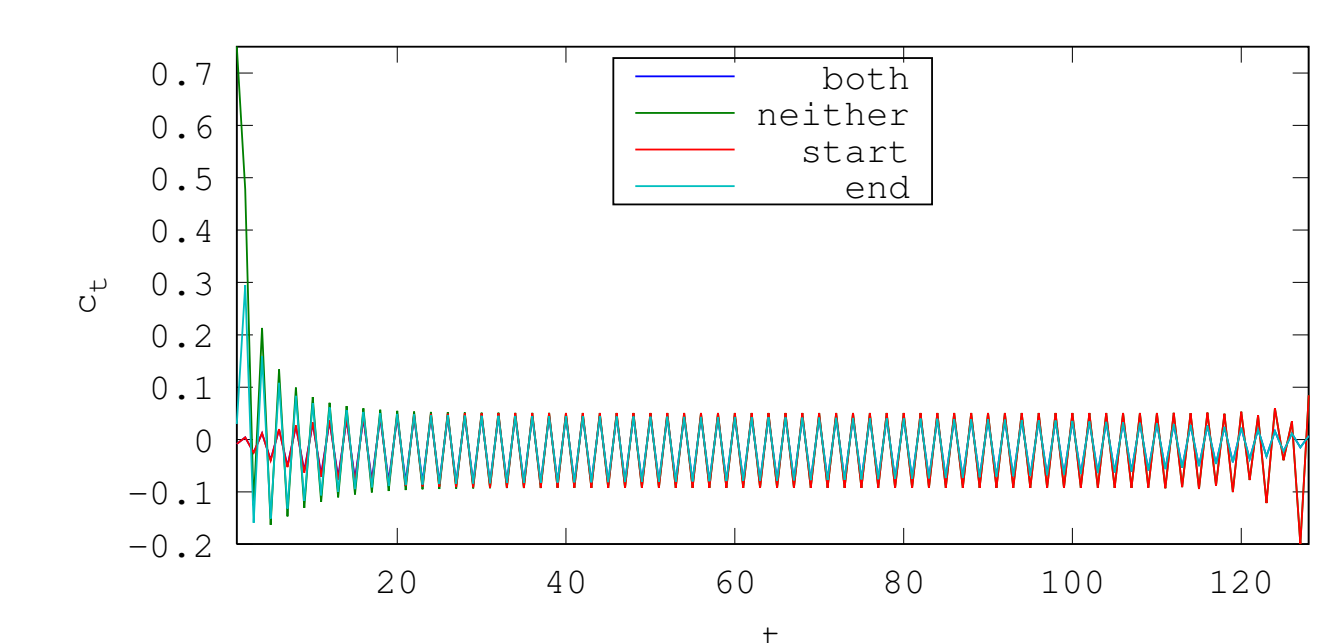


Figure 2: $\mathbf{v}_t = \mathbf{e}_t - 2\mathbf{e}_{t-1} + \mathbf{e}_{t-2}$