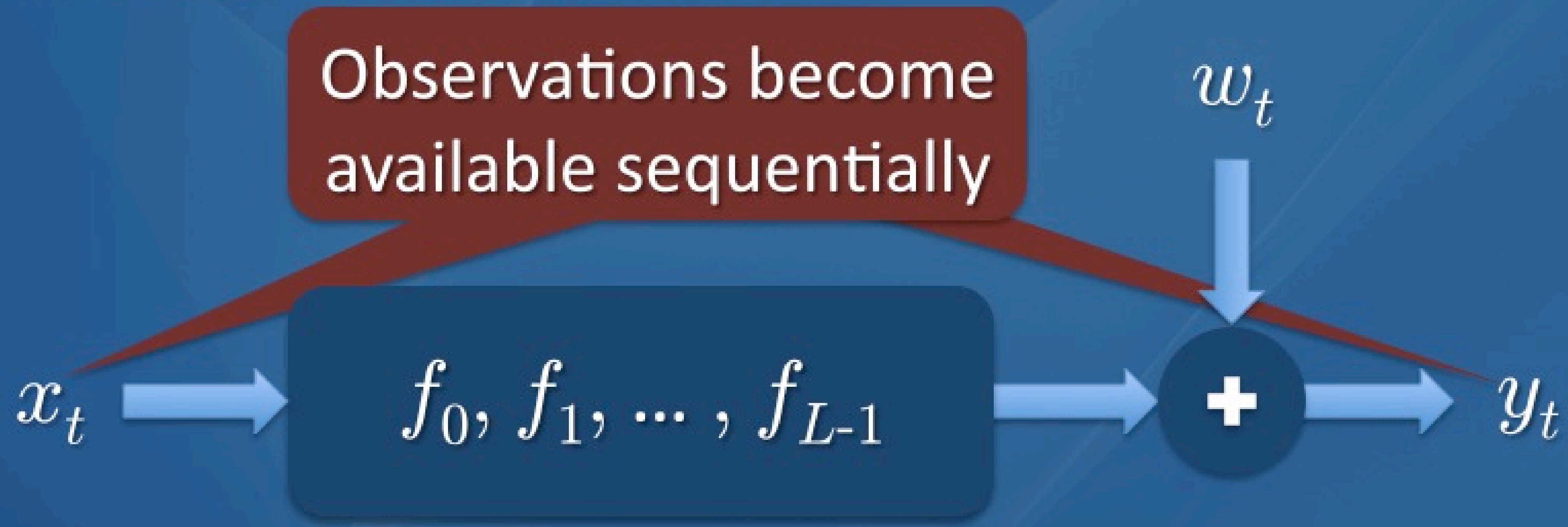
Sequential Detection & System Identification

George V. Moustakides,
University of Patras, Greece
Rutgers University, NJ, USA

Outline

- Problem definition
- Scalar case
  - Constrained stochastic optimization
  - Optimum estimation (system identification)
  - Optimum detection
  - Optimum stopping time
- Vector case
- Further analysis

Problem definition



$$y_t = f_0 x_t + \cdots + f_{L-1} x_{t-L+1} + w_t$$

$$= X_t' F + w_t;\ t = 1, 2, \ldots$$

$$F = [f_0, \ldots, f_{L-1}]'$$

$$X_t = [x_t, \ldots, x_{t-L+1}]'$$

System identification: Adaptively (sequentially) estimate F

We are given sequentially pairs $\{(y_t, x_t)\}$ that are related through

$$y_t = X_t' F + w_t, \quad t = 1, 2, \ldots$$

$\{w_t\}$ is a noise sequence and F satisfies the following two hypotheses:

$$\mathrm{H}_0 : \; F = F_0 \qquad (F_0 = 0)$$

$$\mathrm{H}_1 : \; F \sim \text{Random}$$

Goal: Sequentially decide between H_0,H_1. When decision is H_1 estimate F.

Applications include:

- Model validation of industrial systems
- Statistical analysis of structural changes
- Faulty sensor detection and identification
- Damage detection and isolation
- Water supply management

The scalar case

$$y_t = f x_t + w_t \qquad w_t \sim \mathcal{N}(0, \sigma^2)$$

$$\mathsf{H}_0: \ f = 0$$

$$\mathsf{P}\left(\sum_{n=1}^{t} x_n^2 \to \infty\right) = 1$$

$$\mathsf{H}_1: \ f \sim \mathcal{N}(\mu_f, \sigma_f^2)$$

$\{w_t\}, \{x_t\}, f$ indep.

T : Stopping time, to decide when to stop sampling

d_T : Decision function to decide between H_0 , H_1

$\hat{f}_T$: Estimator for the impulse response

$$\{\mathscr{F}_t\},\ \mathscr{F}_t = \sigma\{(y_1, x_1), \ldots, (y_t, x_t)\}$$

$$\{\mathscr{X}_t\},\ \mathscr{X}_t = \sigma\{x_1, \ldots, x_t\}$$

d_T : $\mathscr{F}_T -$ measurable

$\hat{f}_T$: $\mathscr{F}_T -$ measurable

T : $\{\mathscr{X}_t\} -$ adapted

Grambsh (1983)
Fellouris (2012)

T : $\{\mathscr{F}_t\} -$ adapted ?

Ghosh (1987),(1991)
unattractive!

Detection part: (under H_0 or H_1)

$$\text{Type} - \text{I}: \mathsf{P}_0(d_T = 1 | \mathscr{X}_T)$$

$$\text{Type} - \text{II}: \mathsf{P}_1(d_T = 0 | \mathscr{X}_T)$$

Estimation part: (under H_1)

When $d_T = 1$ then estimate: $\hat{f}_T \;\Rightarrow\; (\hat{f}_T - f)^2$

$$\mathsf{E}_1[(\hat{f}_T - f)^2 \mathbb{1}_{\{d_T=1\}} | \mathscr{X}_T]$$

When $d_T = 0$ then like: $\hat{f}_T = 0 \;\Rightarrow\; (0 - f)^2$

$$\mathsf{E}_1[f^2 \mathbb{1}_{\{d_T=0\}} | \mathscr{X}_T]$$

Constrained optimization

Select costs c_0, c_1, c_e and define the following combined cost function:

$$\mathscr{C}(T, d_T, \hat{f}_T) =$$

$$c_0 \mathsf{P}_0(d_T = 1 | \mathscr{X}_T) + c_1 \mathsf{P}_1(d_T = 0 | \mathscr{X}_T)$$

$$+ c_e \mathsf{E}_1 \left[(\hat{f}_T - f)^2 \mathbb{1}_{\{d_T = 1\}} + f^2 \mathbb{1}_{\{d_T = 0\}} | \mathscr{X}_T \right]$$

$$\inf_{T, d_T, \hat{f}_T} T; \quad \text{subject to} : \mathscr{C}(T, d_T, \hat{f}_T) \leq C$$

No expectation!!

$$\inf_{T, d_T, \hat{f}_T} \mathsf{E}[T] \geq \mathsf{E}[\inf_{T, d_T, \hat{f}_T} T]$$

Optimum estimation

Fix T and d_T. Consider the auxiliary minimization:

$$\inf_{\hat{f}_T} \mathsf{E}_1[(\hat{f}_T - f)^2 \mathbb{1}_{\{d_T=1\}} | \mathscr{X}_T]$$

Optimum estimator:

$$\hat{\mathsf{f}}_T = R_T^{-1} V_T$$

MMSE estimator!

$$V_t = V_{t-1} + y_t x_t, \quad R_t = R_{t-1} + x_t^2$$

$$V_0 = \mu_f \sigma^2 \sigma_f^{-2} \qquad R_0 = \sigma^2 \sigma_f^{-2}$$

$$\inf_{\hat{f}_T} \mathsf{E}_1[(\hat{f}_T - f)^2 \mathbb{1}_{\{d_T=1\}} | \mathscr{X}_T] =$$

$$\sigma^2 R_T^{-1} \mathsf{P}_1 \left(d_T = 1 | \mathscr{X}_T\right)$$

$$\mathsf{E}_1 \left[f^2 \mathbb{1}_{\{d_T=0\}} | \mathscr{X}_T\right] = \mathsf{E}_1 \left[\hat{\mathsf{f}}_T^2 \mathbb{1}_{\{d_T=0\}} | \mathscr{X}_T\right] + \sigma^2 R_T^{-1} \mathsf{P}_1 \left(d_T = 0 | \mathscr{X}_T\right)$$

$$\mathscr{C}(T, d_T, \hat{f}_T) \geq \mathscr{C}(T, d_T, \hat{\mathsf{f}}_T)$$

Optimum detection

$$\mathscr{C}(T, d_T, \hat{\mathsf{f}}_T) =$$
$$c_0 \mathsf{P}_0(d_T = 1 | \mathscr{X}_T) + c_1 \mathsf{P}_1(d_T = 0 | \mathscr{X}_T)$$
$$+ c_e \mathsf{E}_1 \left[\hat{\mathsf{f}}_T^2 \mathbb{1}_{\{d_T = 0\}} | \mathscr{X}_T \right] + c_e \sigma^2 R_T^{-1}.$$

Fix T and consider the auxiliary minimization:

$$\inf_{d_T} \Big\{ c_0 \mathsf{P}_0(d_T = 1 | \mathscr{X}_T) + c_1 \mathsf{P}_1(d_T = 0 | \mathscr{X}_T)$$
$$+ c_e \mathsf{E}_1 \left[\hat{\mathsf{f}}_T^2 \mathbb{1}_{\{d_T = 0\}} | \mathscr{X}_T \right] \Big\}$$

Optimum detector:

$$\mathsf{d}_T = \begin{cases} 1 & \text{if} \quad c_0 \le \mathsf{L}_T \left\{ c_1 + c_e \hat{\mathsf{f}}_T^2 \right\} \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathsf{L}_T = \frac{\sigma}{\sigma_f \sqrt{R_T}} e^{\frac{1}{2\sigma^2} R_T^{-1} V_T^2 - \frac{1}{2} \mu_f^2 \sigma_f^{-2}}$$

Using the optimum estimator and detector yields

$$\mathscr{C}(T, d_T, \hat{f}_T) \ge \mathscr{C}(T, d_T, \hat{\mathsf{f}}_T) \ge \mathscr{C}(T, \mathsf{d}_T, \hat{\mathsf{f}}_T)$$

Optimum stopping time

Apply optimum detector and estimator, then

$$\mathscr{C}(T,\mathsf{d}_T,\hat{\mathsf{f}}_T) = \mathsf{E}_0\left[\left(c_0 - \mathsf{L}_T\left\{c_1 + c_e\hat{\mathsf{f}}_T^2\right\}\right)^{-} |\mathscr{X}_T\right]$$

$$+ c_1 + c_e(\mu_f^2 + \sigma_f^2)$$

Define the function

$$\mathscr{G}(R) = \int_{-\infty}^{\infty}\left(c_0 - \frac{\frac{\sigma}{\sigma_f}e^{\frac{R^{-1}V^2}{2\sigma^2} - \mu_f^2\frac{1}{2\sigma_f^2}}}{\sqrt{R}}\left[c_1 + c_eR^{-2}V^2\right]\right)^{-}$$

$$\times\frac{e^{-\frac{1}{2\sigma^2}(R-\frac{\sigma^2}{\sigma_f^2})^{-1}(V-\mu_f\frac{\sigma^2}{\sigma_f^2})^2}}{\sqrt{2\pi\sigma^2(R-\frac{\sigma^2}{\sigma_f^2})}}dV + c_1 + c_e(\mu_f^2 + \sigma_f^2)$$

Theorem:

Function $\mathscr{G}(R)$ is strictly **decreasing** in R.

It is also true that

$$\mathscr{C}(T, \mathsf{d}_T, \hat{\mathsf{f}}_T) = \mathscr{G}(R_T)$$

Optimum stopping time:

$$\mathsf{T} = \inf\{t \geq 0 : \mathscr{G}(R_t) \leq C\}$$

$$R_t = R_{t-1} + x_t^2; \quad \mathsf{P}(R_t \to \infty) = 1$$

Summary: We sequentially observe $\{(y_t, x_t)\}$

$$y_t = f x_t + w_t$$

$$\mathrm{H}_0: \;\; f = 0$$

$$\mathrm{H}_1: \;\; f \sim \mathcal{N}(\mu_f, \sigma_f^2)$$

Sequentially decide between H_0,H_1. Whenever decision in favor of H_1 estimate f. Find triplet $T, \; d_T, \; \hat{f}_T$ that solves

$$\inf_{T, d_T, \hat{f}_T} T; \quad \text{subject to}: \mathscr{C}(T, d_T, \hat{f}_T) \leq C$$

$$R_t = x_t^2 + \cdots + x_1^2 + \sigma_f^{-2}\sigma^2$$

$$V_t = y_t x_t + \cdots + y_1 x_1 + \mu_f \sigma_f^{-2}\sigma^2$$

$$\mathsf{T} = \inf\{t \geq 0 : \mathscr{G}(R_t) \leq C\}$$

$$\hat{\mathsf{f}}_{\mathsf{T}} = R_{\mathsf{T}}^{-1} V_{\mathsf{T}}$$

$$\mathsf{L}_{\mathsf{T}} = \frac{\sigma}{\sigma_f \sqrt{R_{\mathsf{T}}}} e^{\frac{1}{2\sigma^2} R_{\mathsf{T}}^{-1} V_{\mathsf{T}}^2 - \frac{1}{2}\mu_f^2 \sigma_f^{-2}}$$

$$\mathsf{d}_{\mathsf{T}} = \begin{cases} 1 & \text{if } \; c_0 \leq \mathsf{L}_{\mathsf{T}} \left\{ c_1 + c_e \hat{\mathsf{f}}_{\mathsf{T}}^2 \right\} \\ 0 & \text{otherwise,} \end{cases}$$

Theorem:

The triplet $\mathsf{T}, \mathsf{d}_\mathsf{T}, \hat{\mathsf{f}}_\mathsf{T}$ is optimum in the sense that it solves the optimization problem:

$$\inf_{T, d_T, \hat{f}_T} T; \quad \text{subject to} : \mathscr{C}(T, d_T, \hat{f}_T) \leq C$$

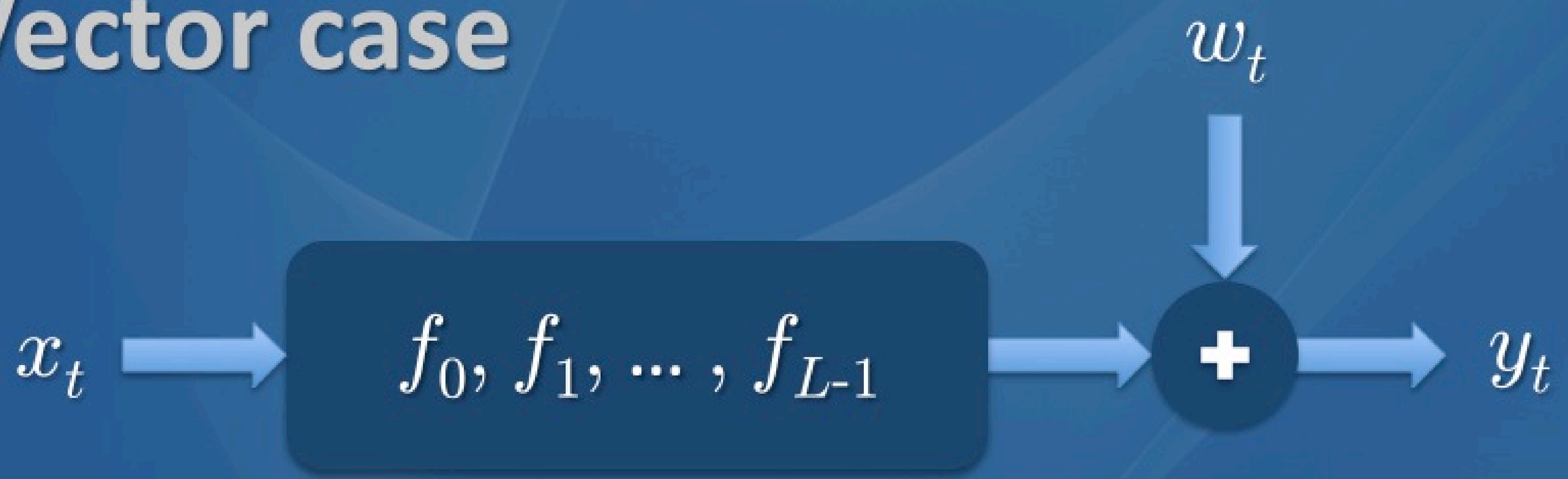
Proof:

For any other triplet satisfying the constraint we can write

$$C \geq \mathscr{C}(T, d_T, \hat{f}_T) \geq \mathscr{C}(T, \mathsf{d}_T, \hat{\mathsf{f}}_T) = \mathscr{G}(R_T)$$

$$\Rightarrow T \geq \mathsf{T}$$

Vector case



$$y_t = X_t' F + w_t$$
$$F = [f_0, \ldots, f_{L-1}]'$$
$$X_t = [x_t, \ldots, x_{t-L+1}]'$$

$\mathsf{H}_0: \; F = 0$

$\mathsf{H}_1: \; F \sim \mathcal{N}(\mu_F, \Sigma_F)$

$$\mathscr{C}(T, d_T, \hat{F}_T) =$$
$$c_0 \mathsf{P}_0(d_T = 1 | \mathscr{X}_T) + c_1 \mathsf{P}_1(d_T = 0 | \mathscr{X}_T)$$
$$+ c_e \mathsf{E}_1 \left[\|\hat{F}_T - F\|^2 \mathbb{1}_{\{d_T = 1\}} + \|F\|^2 \mathbb{1}_{\{d_T = 0\}} | \mathscr{X}_T \right]$$

$$\inf_{T, d_T, \hat{F}_T} T; \quad \text{subject to} : \mathscr{C}(T, d_T, \hat{F}_T) \leq C$$

MMSE estimates

Complexity $O(L^3)$

$$\hat{\mathsf{F}}_t = \mathbf{R}_t^{-1} V_t$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + X_t X_t', \quad \mathbf{R}_0 = \sigma^2 \Sigma_F^{-1}$$

$$\mathbf{P}_t = \mathbf{R}_t^{-1}$$

$$V_t = V_{t-1} + y_t X_t, \quad V_0 = \sigma^2 \Sigma_F^{-1} \mu_F$$

Recursive Least Squares (RLS)

Complexity $O(L^2)$

$$e_t = y_t - X_t' \hat{\mathsf{F}}_{t-1}$$

$$K_t = \mathbf{P}_{t-1} X_t$$

$$\gamma_t = 1/(1 + X_t' K_t)$$

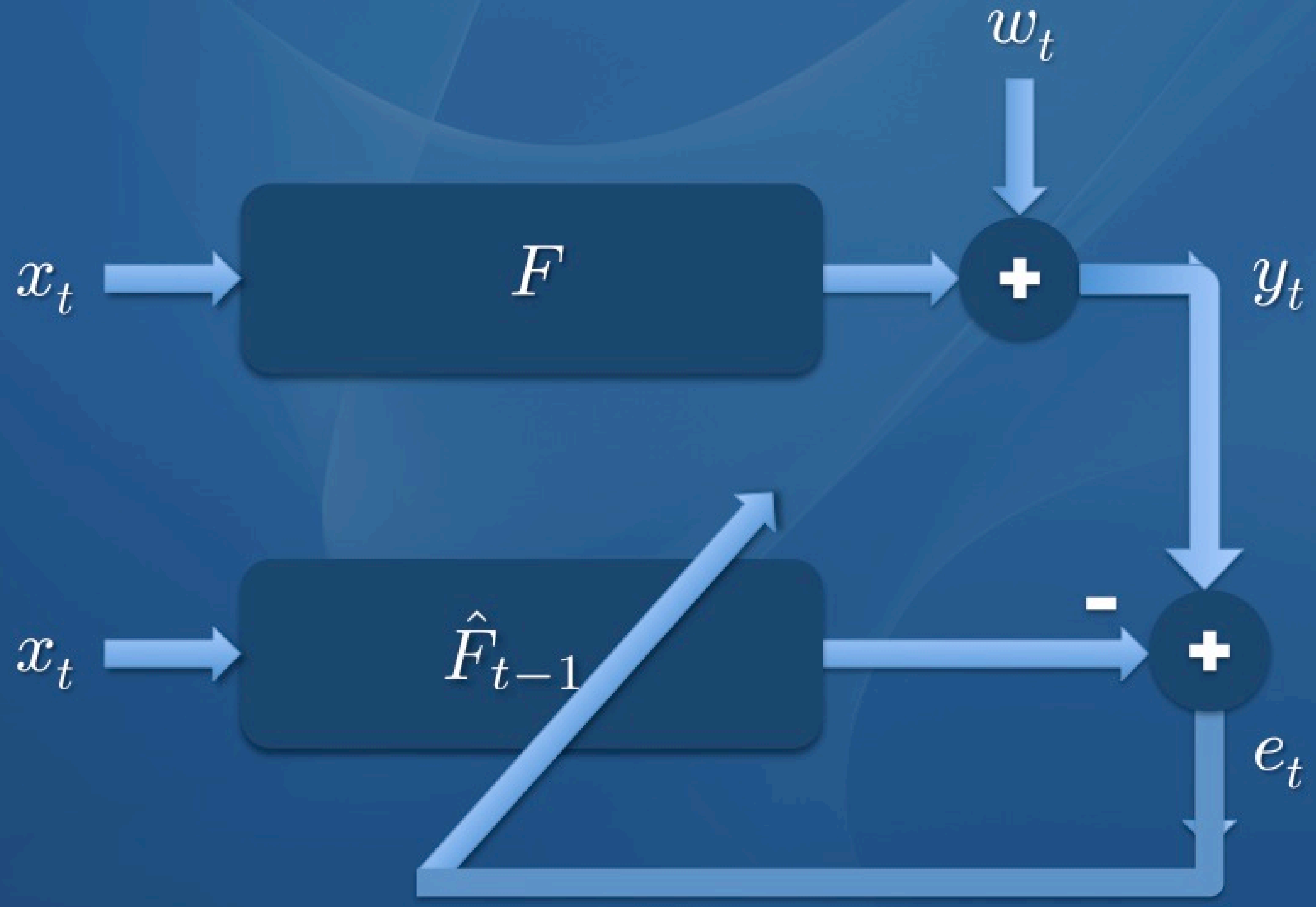
$$\hat{\mathsf{F}}_t = \hat{\mathsf{F}}_{t-1} + \gamma_t e_t K_t$$

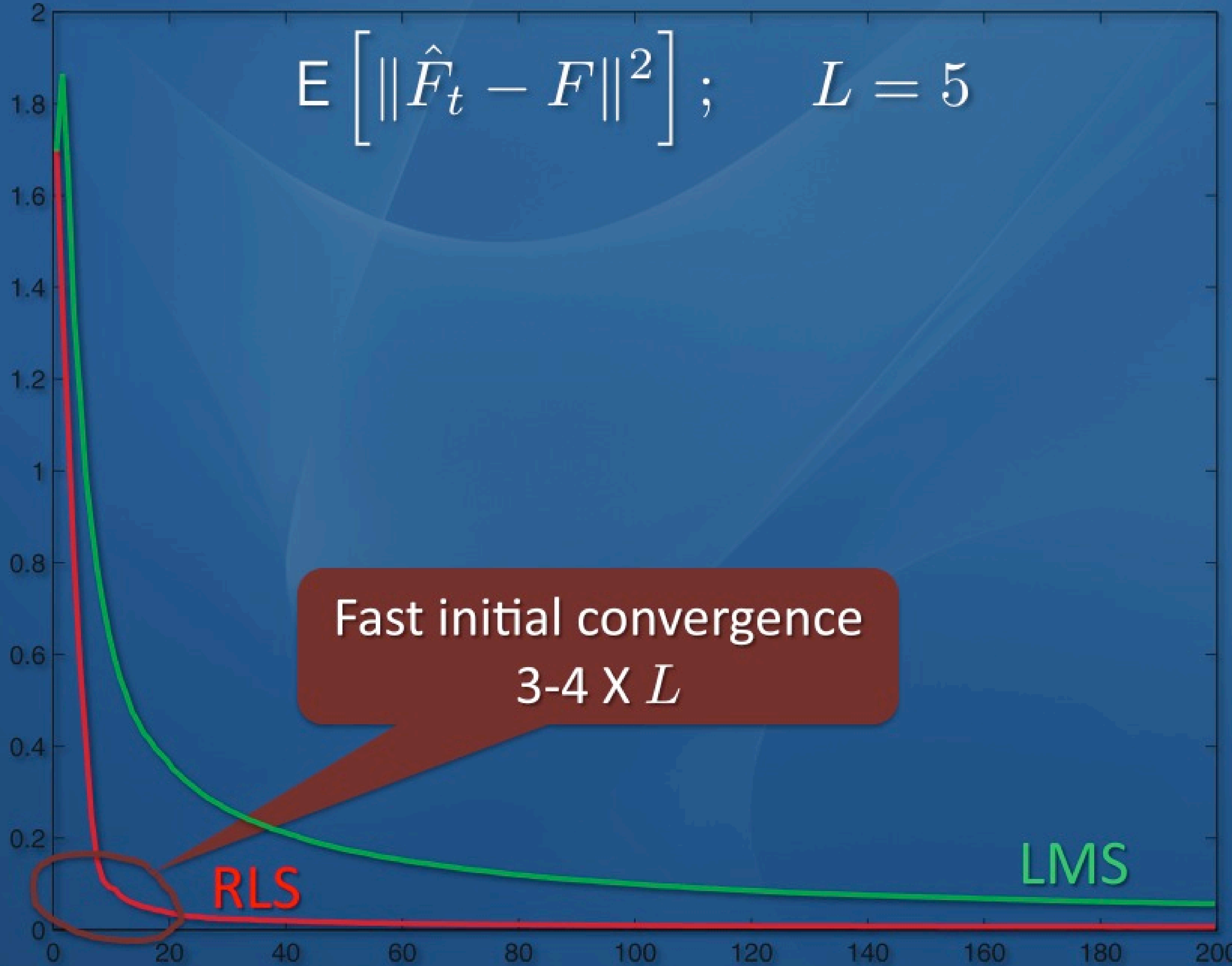
$$\mathbf{P}_t = \mathbf{P}_{t-1} - \gamma_t K_t K_t'$$

Optimum MMSE estimates (Implementation of Kalman Filter); Extremely fast convergence.

High Complexity, Non-robust to finite precisions

Adaptive System Identification





$E\left[\|\hat{F}_t - F\|^2\right]; \quad L = 5$
Fast initial convergence
3-4 X L
RLS
LMS
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
100
120
140
160
180
200

Least Mean Square (LMS) estimates

$$e_t = y_t - X_t'\hat{F}_{t-1}$$

$$\hat{F}_t = \hat{F}_{t-1} + \gamma_t e_t X_t$$

$$\gamma_t = 1/t$$

Amazing algorithm!

Very low complexity ($2L$).

Insensitive to finite precision and Robust to model mismatch.

It is **THE** algorithm used in practice for real-time applications.

Slow convergence; Non-optimum estimates.

$$\mathbf{R}_t = X_t X_t' + \cdots + X_1 X_1' + \sigma^2 \Sigma_F^{-1}$$

$$V_t = y_t X_t + \cdots + y_1 X_1 + \sigma^2 \Sigma_F^{-1} \mu_F$$

$$\mathsf{T} = \inf\{t \geq 0 : \mathscr{G}(\mathbf{R}_t) \leq C\}$$

$$\mathbf{A} > \mathbf{B} \Downarrow \mathscr{G}(\mathbf{A}) < \mathscr{G}(\mathbf{B})$$

$$\mathsf{P}(\lambda_{\min}(\mathbf{R}_t) \to \infty) = 1$$

$$\hat{\mathsf{F}}_\mathsf{T} = \mathbf{R}_\mathsf{T}^{-1} V_\mathsf{T}$$

$$\mathsf{L}_\mathsf{T} = \frac{\sigma^L}{\sqrt{|\Sigma_F||\mathbf{R}_\mathsf{T}|}} e^{\frac{1}{2\sigma^2} V_\mathsf{T}' \mathbf{R}_\mathsf{T}^{-1} V_\mathsf{T} - \frac{1}{2} \mu_F' \Sigma_F^{-1} \mu_F}$$

Update for determinant

$$\mathsf{d}_\mathsf{T} = \begin{cases} 1 & \text{if } c_0 \leq \mathsf{L}_\mathsf{T} \left\{ c_1 + c_e \|\hat{\mathsf{F}}_\mathsf{T}\|^2 \right\} \\ 0 & \text{otherwise,} \end{cases}$$

Further analysis

- The function $\mathscr{G}(\mathbf{R})$ can be computed only numerically. This is tractable solely for the scalar case. Obtain efficient approximations for small constraint cost values C (asymptotically optimal solutions for T).

- Instead of looking for the optimum triplet, fix a suboptimum (adaptive) estimator (e.g. LMS) and optimize only over the stopping time and the decision function.