Rare Events Modeling for Linear Estimation and Control

Kenji Kashima
Kyoto University

SSS’21 Tutorial Seminar

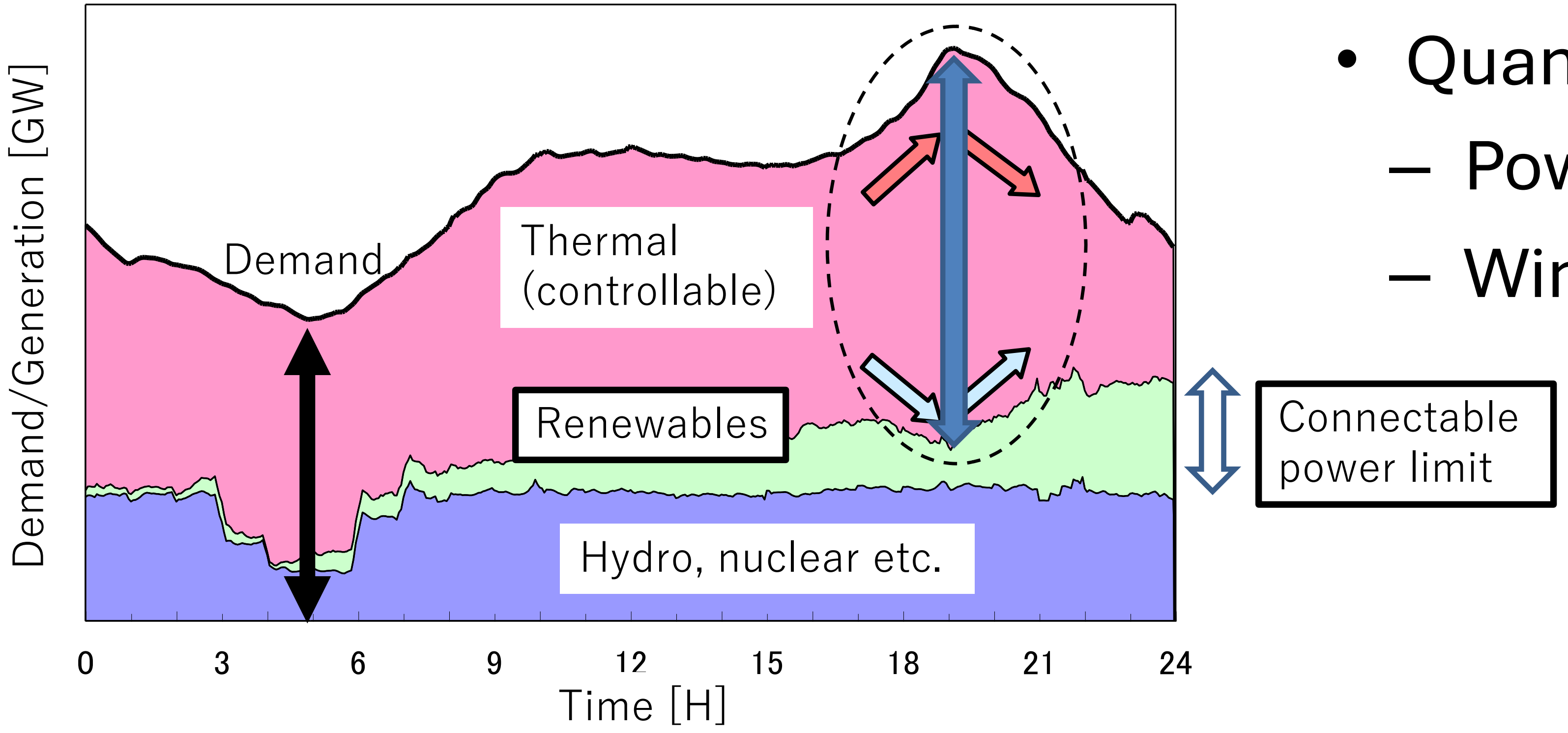
Motivation

- Stable power supply under the widely introduced renewable energies
  - Evaluating the impact of wind power fluctuation on power system quality

JST CREST
System Theory for Harmonized Power System Control Based on Photovoltaic Power Prediction

Motivation: Renewables are uncontrollable

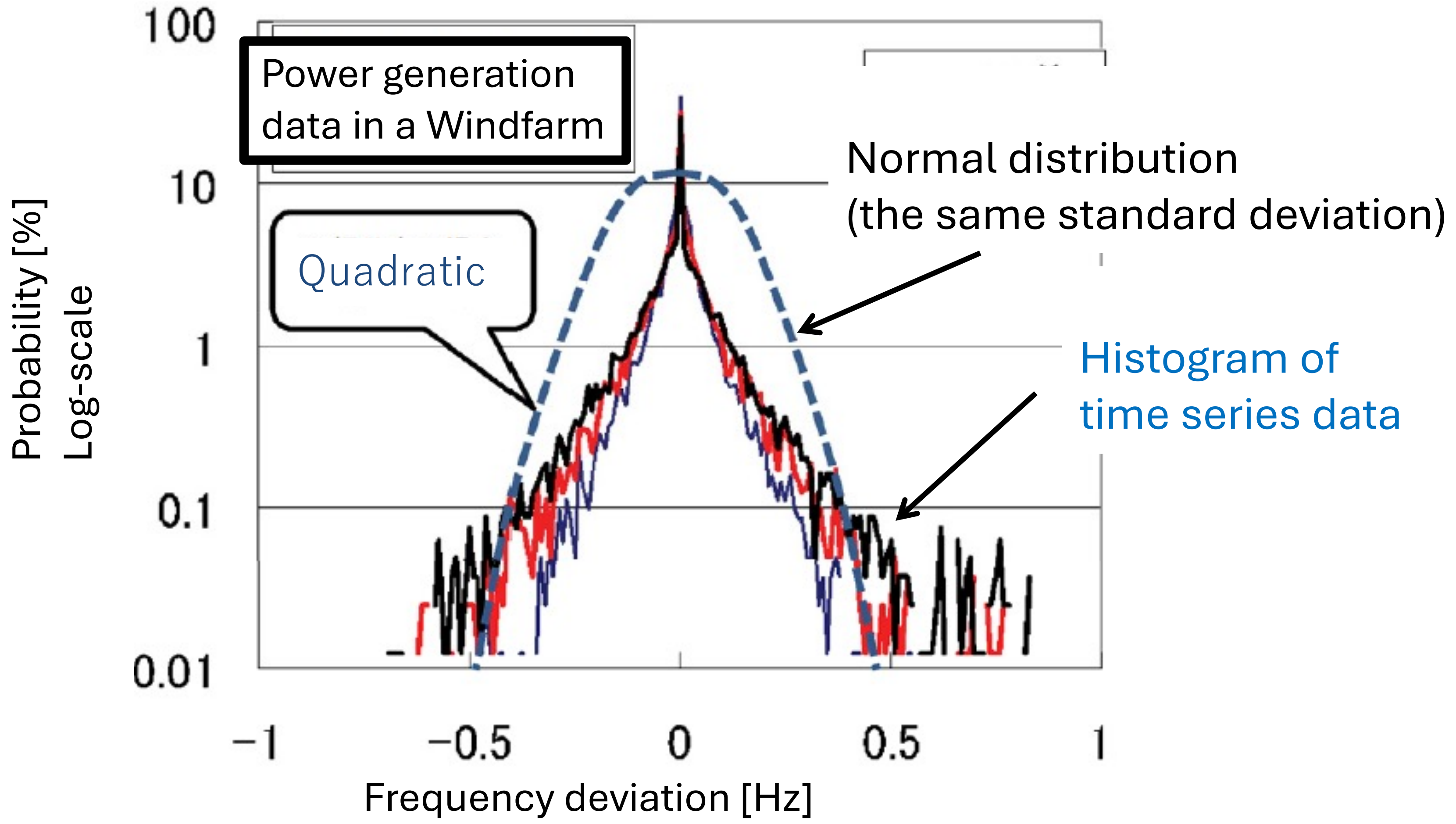
- Long-term (>20 [min])
  - Thermal unit output may reach its upper/lower limit
- Short-term (1〜20 [min])
  - Thermal output change speed may reach its limit.



- Quantification required
  - Power plant dynamics
  - Wind power uncertainty

Motivation: Extremum event

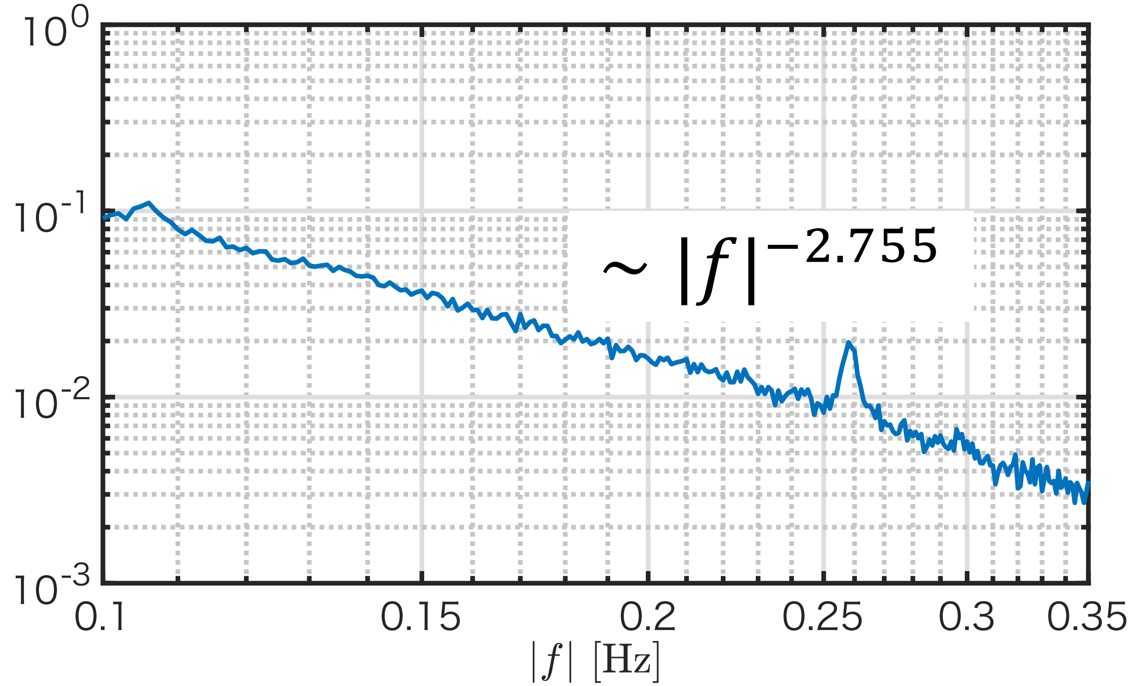
- The fluctuation of wind power generation is usually small, but it becomes extremely large due to the occurrence of gusts and turbulence



Motivation: Power law

- Power law (a.k.a. scale-free property)
  - linear in log-log scale

Frequency deviation histogram of PS interconnected with wind power



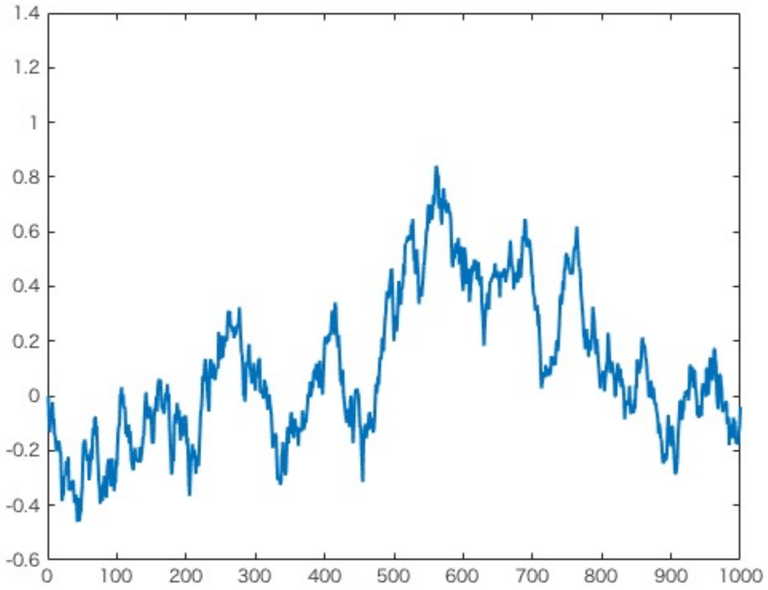
Outline

- **Gaussian distribution revisited**
  - **Affinity to linear systems, its limitation and generalization**
- Key theoretical results
  - Linear system analysis and equivalent linearization
- New application
  - Control systems privacy
- Sparsity VS rare events
  - Sparse optimal stochastic control

Rationale: Emergence of Gaussianity

- Central Limit Theorem
  - Average of independent random variables **having finite variance** converges to a Gaussian.

- Wiener process
  - If a stochastic process is **almost continuous i.i.d. increment**, then it is a Wiener process.



Mathematical property

- Simple density function characterized only by two parameters
  - $\varphi(x) \propto \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, σ: SD, μ: mean
- Simple expression for characteristic function
  - $\mathbb{E}[\exp(-\mathrm{i}\omega x)] = \exp(\mathrm{i}\mu\omega - \sigma^2\omega^2)$
- Superposition principle
  - Gaussian r.v. + Gaussian r.v. = Gaussian r.v.
- Conjugation (closed under Bayes estimation), to list a few

Superposition and linear systems

- If the input signal is i.i.d. Gaussian, the output signal of linear systems are Gaussian.

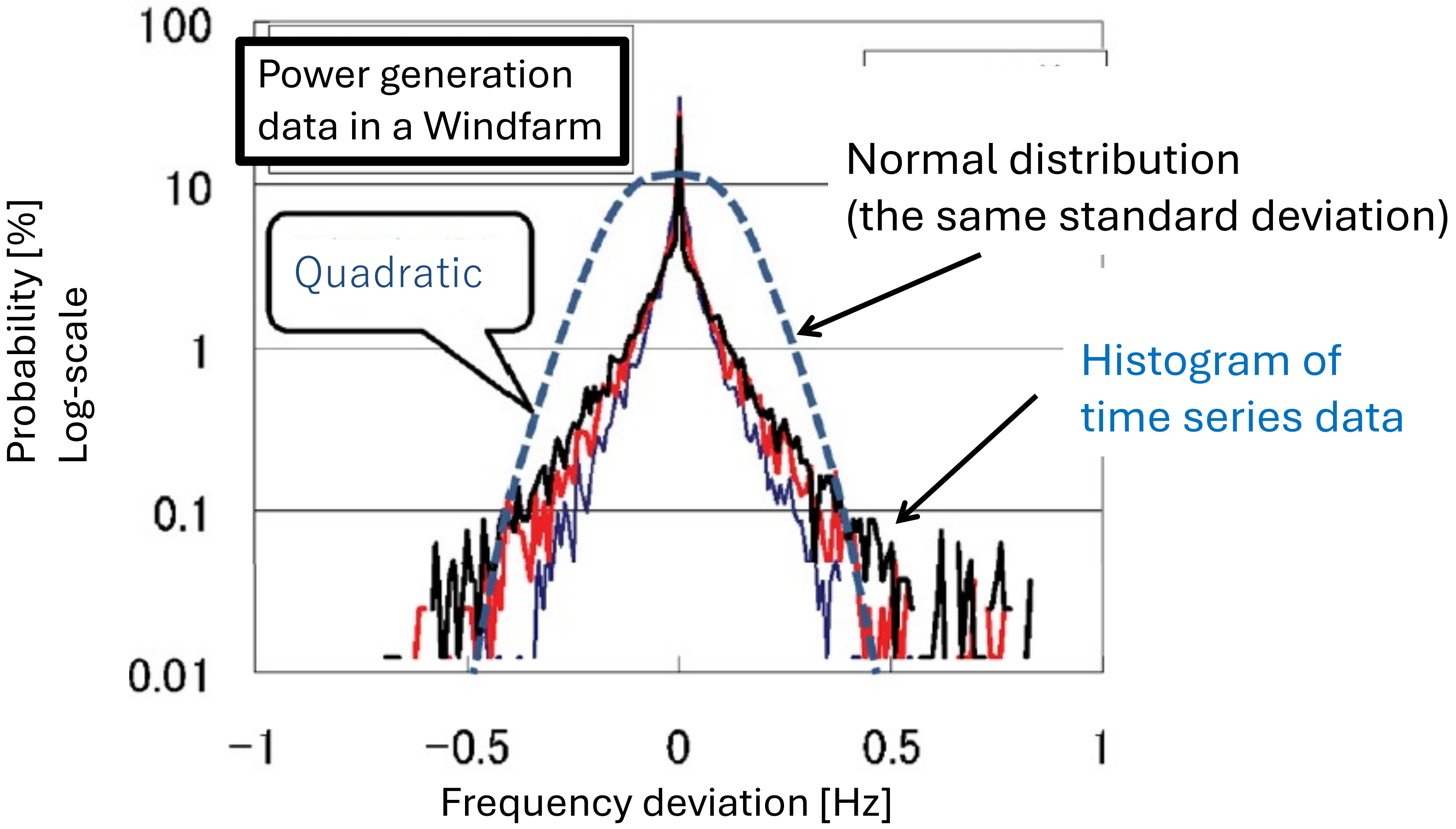
  - Discrete-time case
    - $x_{k+1} = Ax_k + Bv_k, \quad v_k \sim$ i.i.d. Gaussian
    - $x_k, \forall k$ and stationary distributions are Gaussian

  - Continuous-time case
    - $dx_t = Ax_t dt + BdW_t, \quad W_t$: Wiener process
    - $x_t, \forall t$ and stationary distributions are Gaussian

Limitation: Quickly decaying tails

- Density function decays in a square exponential manner.
  - Large variance does not imply heavy tail.

$$\varphi(x) \propto \exp\left(-\frac{x^2}{2\sigma^2}\right)$$



Generalization: α -Stable distribution $x \sim S_\alpha(\mu, \sigma)$

- ~~Simple~~ density function characterized only by ~~two~~ parameters (Gaussian if $\alpha = 2$)
  three
  - $\varphi(x) \propto \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, σ: SD, μ: mean

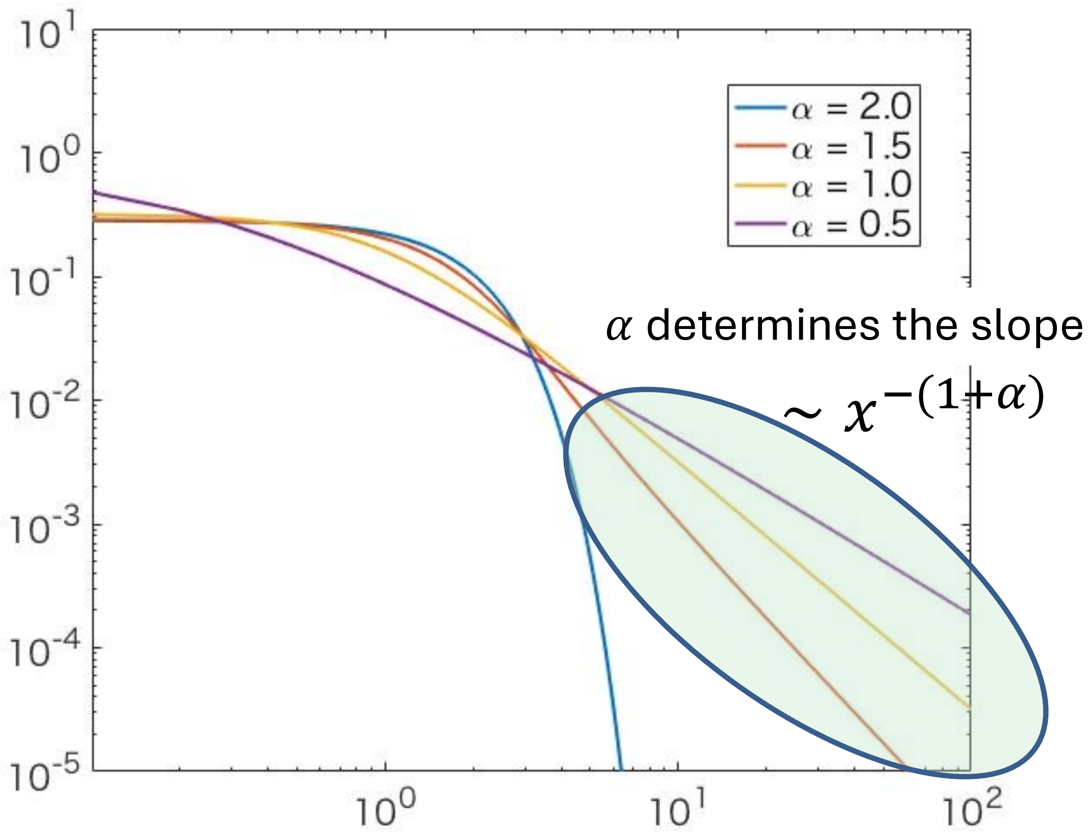
- Simple expression for characteristic function
  - $\mathrm{E}[\exp(-\mathrm{i}\omega x)] = \exp(\mathrm{i}\mu\omega - \sigma^\alpha|\omega|^\alpha)$

- Central Limit Theorem
  - Average of independent random variables ~~**having finite variance**~~ converges to a **stable** distribution.

- “Stable” has nothing to do with “dynamical stability”.

Stable distribution $x \sim S_\alpha(\mu, \sigma)$

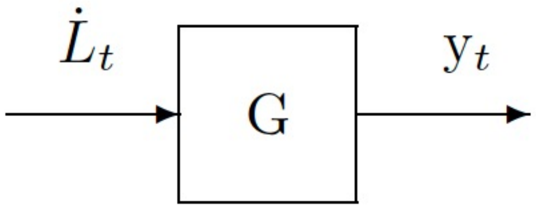
- Tails of density functions follow power law.
  - Suitable for dealing with rare events
- Superposition principle
  - Affinity to linear systems
- Careful mathematical treatment
  - Unbounded variance, etc



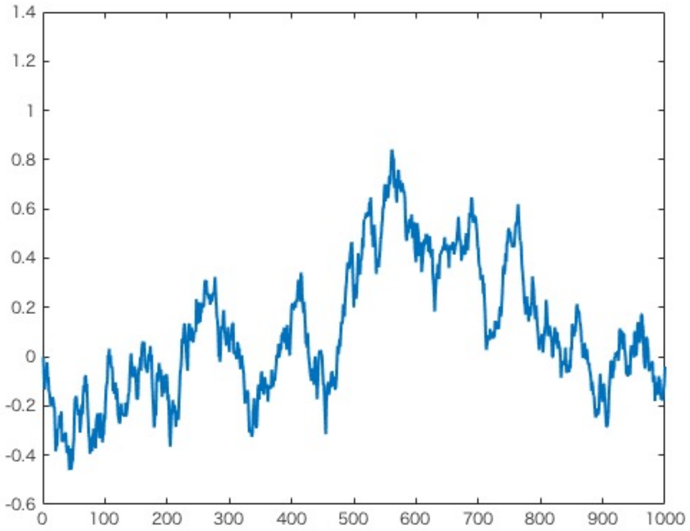
Log-log plot of pdf

Stable process

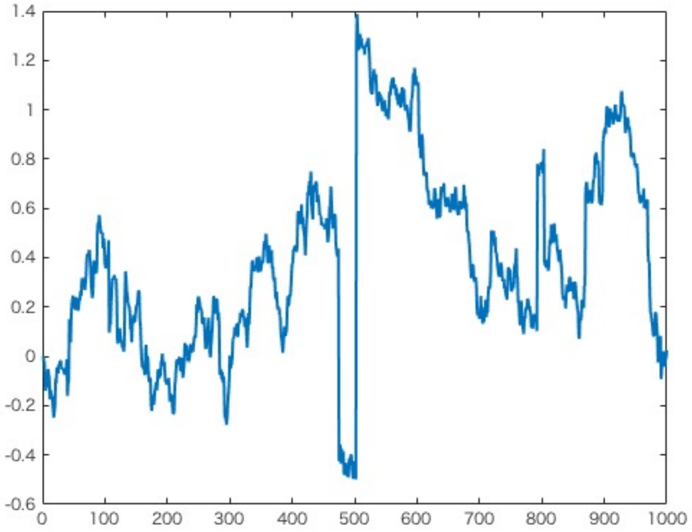
- Wiener process: $W_t \sim N(0, \sqrt{t})$
  - Scale (variance) = time t
- Stable process: $L_t \sim S_\alpha(0, t^{1/\alpha})$
  - Scale = $t^{2/\alpha}$



Sample paths



$\alpha = 2$

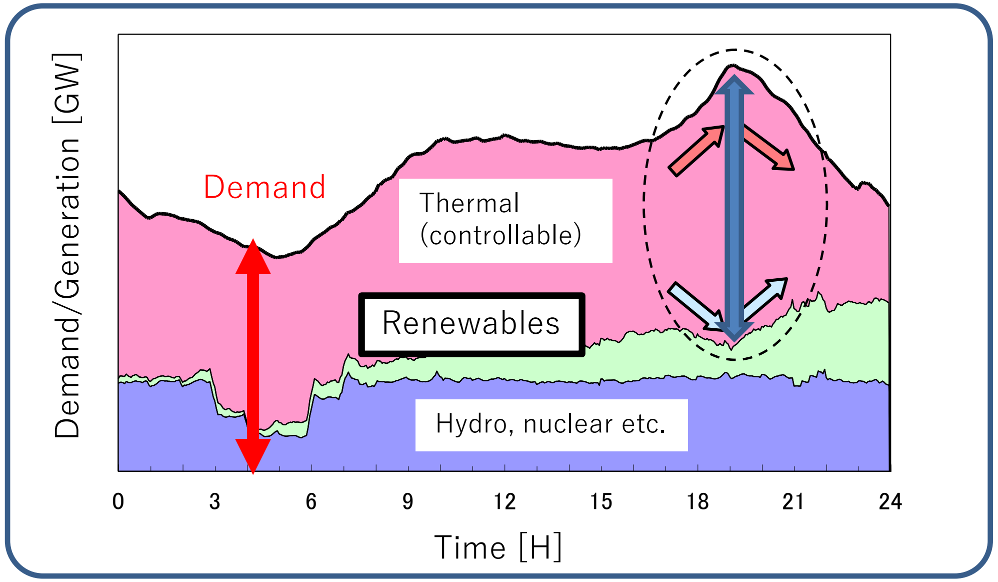


$\alpha = 1.9$

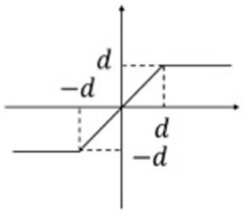
Outline

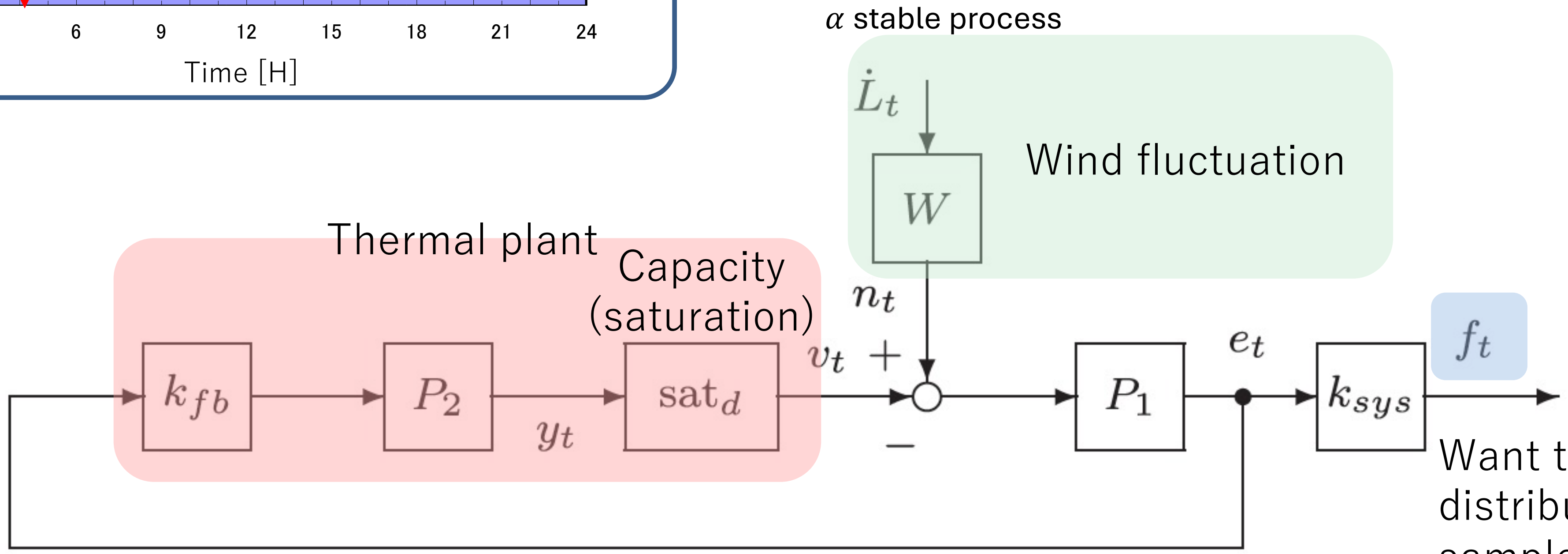
- Gaussian distribution revisited
  - Affinity to linear systems, its limitation and generalization
- **Key theoretical results**
  - **Linear system analysis and equivalent linearization**
- New application
  - Control systems privacy
- Sparsity VS rare events
  - Sparse optimal stochastic control

Modeling example revisited: Power network



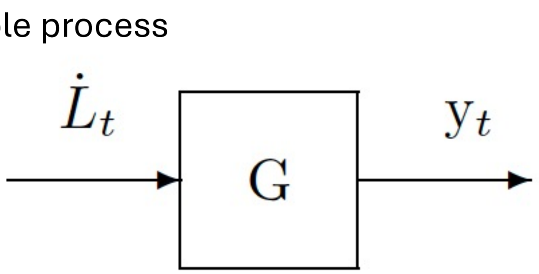
1. Wind fluctuation model
2. Calculate the stationary distribution
3. Effect of nonlinearity





Want to know the stationary distribution without sample path generation

Frequency domain model



$$\kappa_{(\alpha,p)} = \frac{(2\kappa_\alpha)^p \Gamma(1+\frac{p}{2})\Gamma(1-\frac{p}{\alpha})}{\Gamma(1-\frac{p}{2})}$$

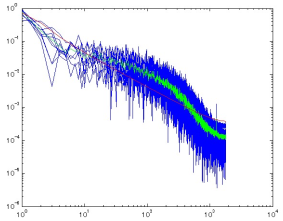
$$\kappa_\alpha := \left(\frac{1}{\pi}\int_0^\pi |\cos t|^\alpha dt\right)^{1/\alpha}.$$

- Theorem 1: For any $\alpha \in (1,2], p \in (-1,\alpha), \omega$,

$$\lim_{T\to\infty} \mathbb{E}\left[\left|\frac{1}{T^{1/\alpha}}\int_0^T e^{-j\omega t} y_t dt\right|^p\right] = \kappa_{(\alpha,p)} |G(j\omega)|^p$$

$\approx$ Power spectrum density

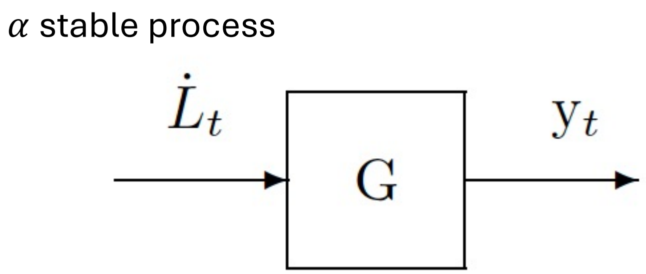
= Frequency gain of transfer function



Easily computable from data

Gain diagram fitting

Stationary distribution of linear systems driven by stable process



- Theorem 2: For any $\alpha \in (1,2]$ and stable $G(s) = c(sI - A)^{-1}b$, the stationary distribution of y_t is $S_\alpha(0, ||ce^{At}b||_\alpha)$.
  - $||f||_\alpha := \left(\int |f(t)|^\alpha dt\right)^{\frac{1}{\alpha}}$
  - Same α as input noise
  - Stationary variance is L^α-norm of impulse response.

Generalized plant representation

$$dx_t = Ax_t dt + Bu_t dt + b dL_t$$
$$f_t = c_z x_t$$
$$y_t = C_y x_t$$
$$u_t = \mathrm{sat}_d(y_t)$$

$x_t \in \mathbb{R}^n$: State, $b, c_z^\top \in \mathbb{R}^n$ A, B, C_y : Matrices

L_t : Stable process with parameter α

- If the nonlinearity is negligible, the stationary output distribution can be obtained analytically by Theorem 2.

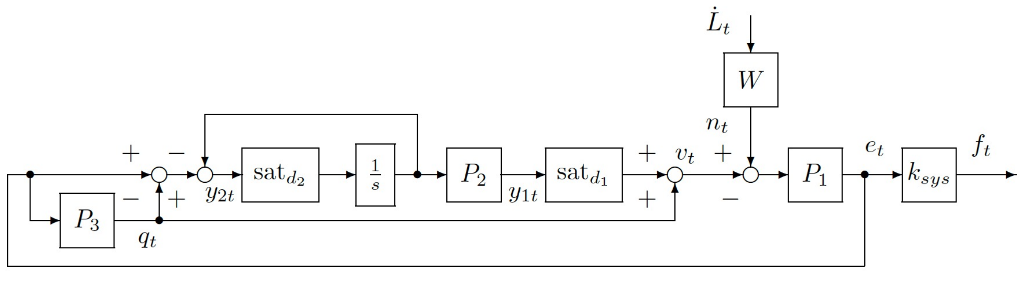
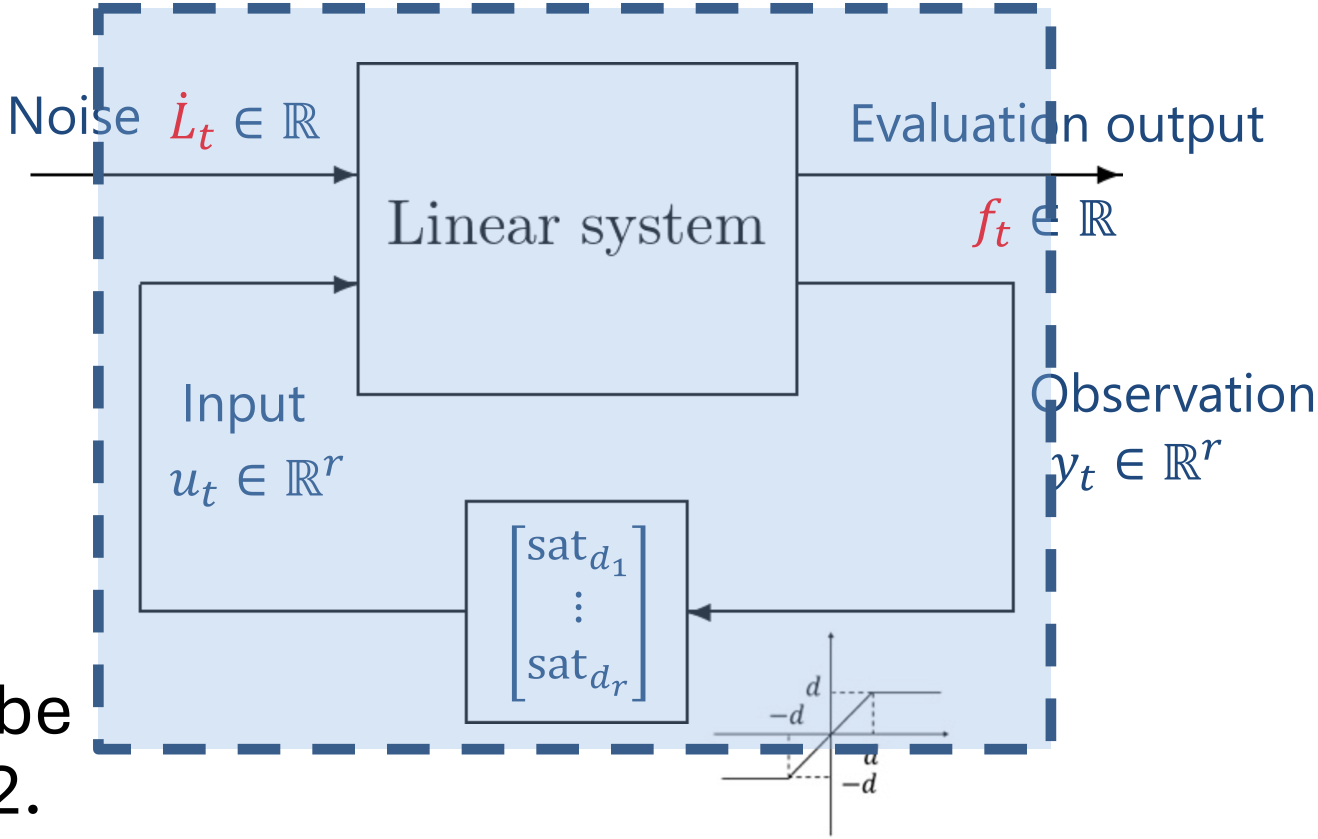
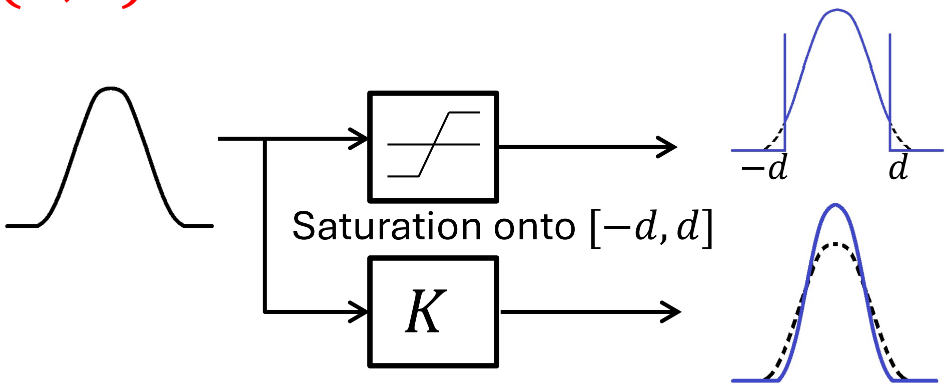
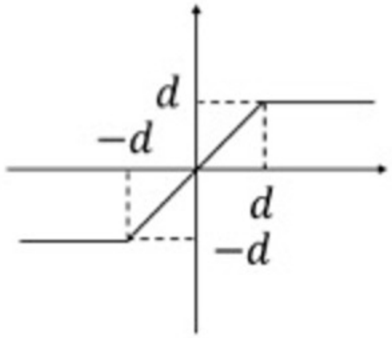


Fig. 12: Block diagram of the load frequency control system.



- Linear gain $K \in (0,1)$ that minimizes the error variance

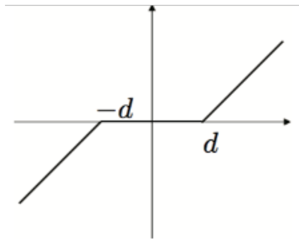




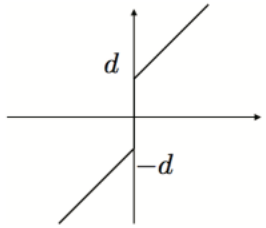
Saturation

- For $x \sim N(0, \sigma^2)$, the optimal $K = \mathrm{erf}\left(\frac{d}{\sqrt{2}\sigma}\right)$ with $\mathrm{erf}(x) := \frac{1}{\sqrt{\pi}} \int_{-x}^{+x} \mathrm{e}^{-z^2} dz.$

- Theorem 3:
  For $x \sim S_\alpha(0, \sigma^2)$, the optimal $K = \min(1, d/\sigma\gamma_\alpha)$



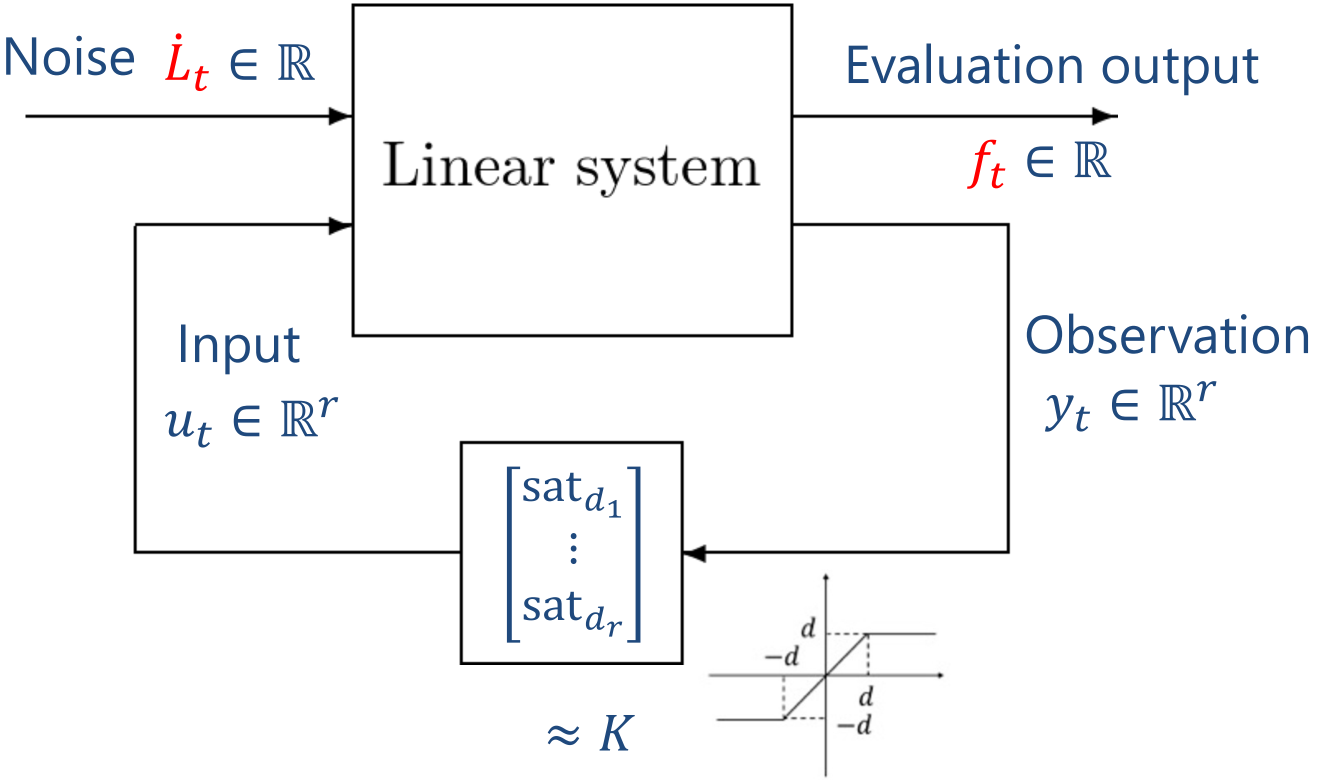
Deadzone



Friction

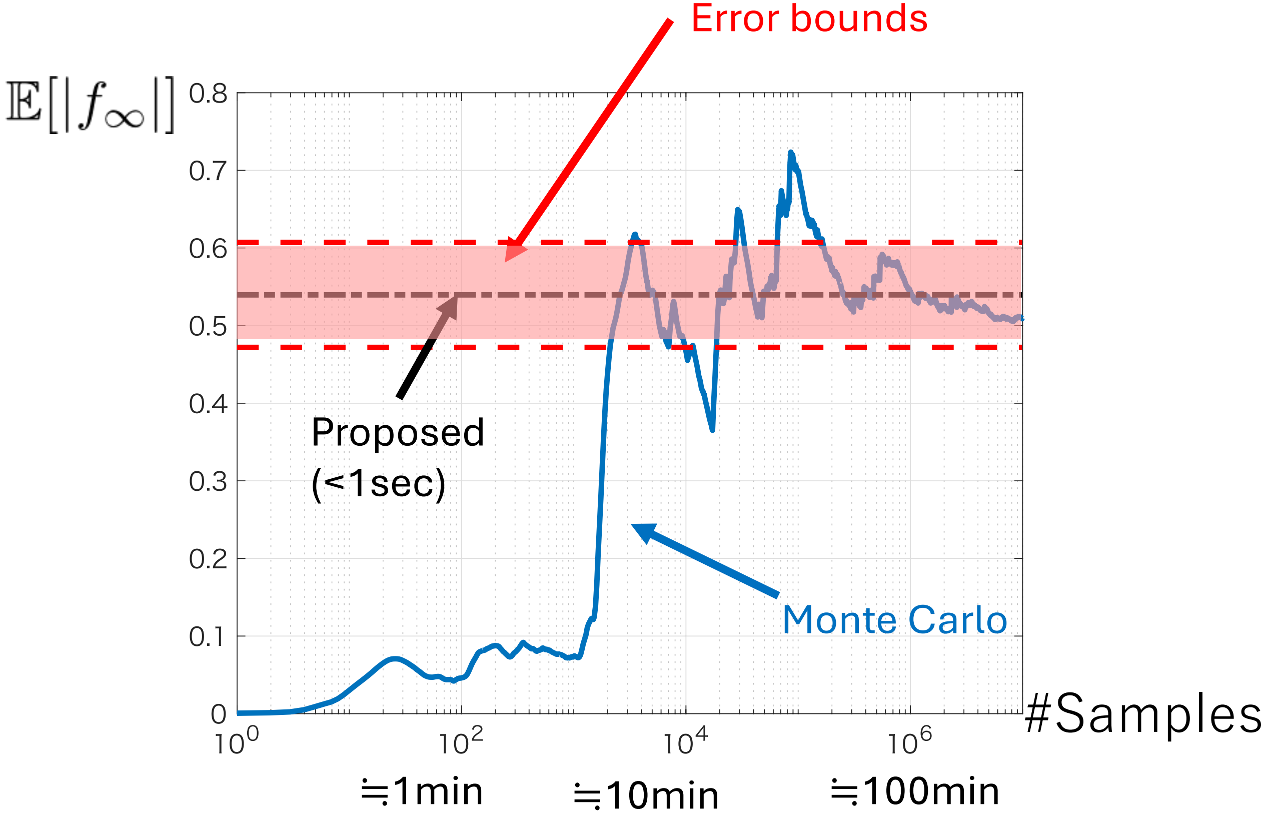
Equivalent linearization in the feedback loop

- Variance of y determines the approximated linear gain K .
  - $K = f(\sigma_y) = \min(1, d/\sigma_y \gamma_\alpha)$
- The gain K determines the stationary distribution of y.
  - $\sigma_y = g(K)$
- The solution to $K = f(g(K))$
  - Theoretical error bound



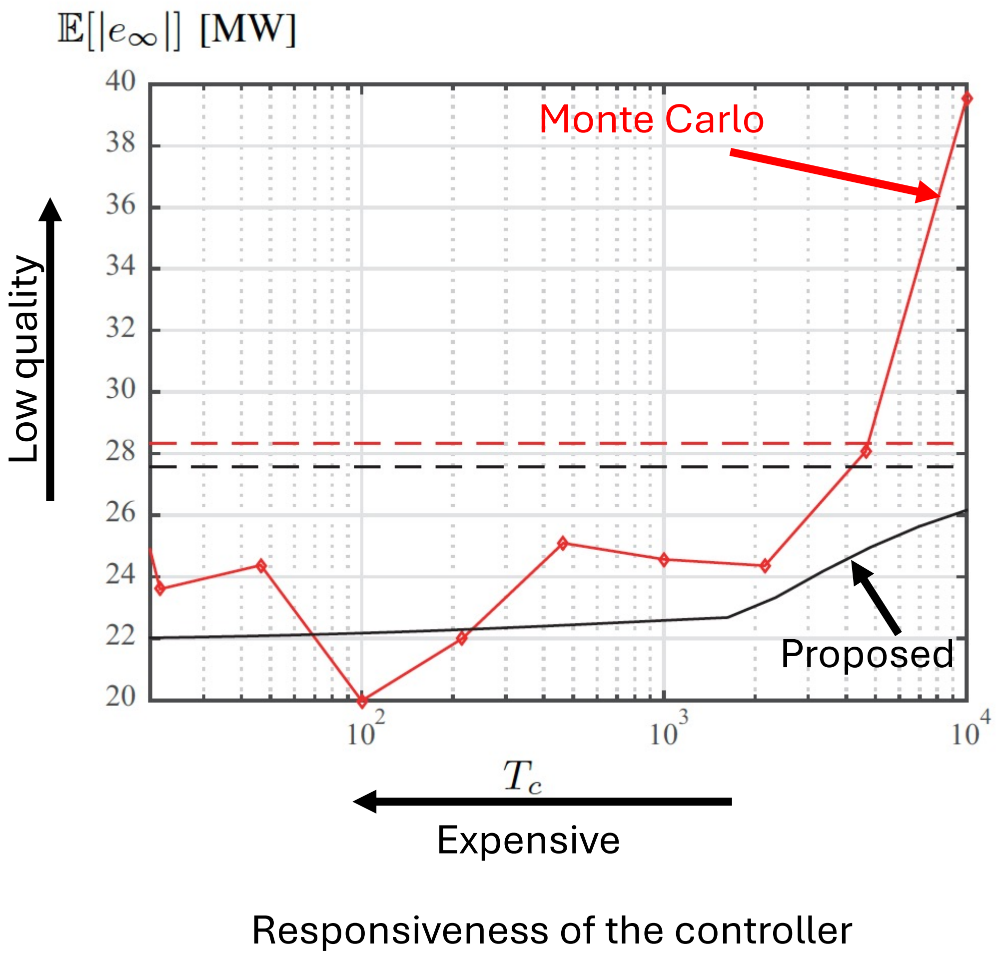
Example: Proposed method vs Monte Carlo

#Samples and MC average



Rare event simulation requires too many samples.

Physical parameter T_c and stationary error

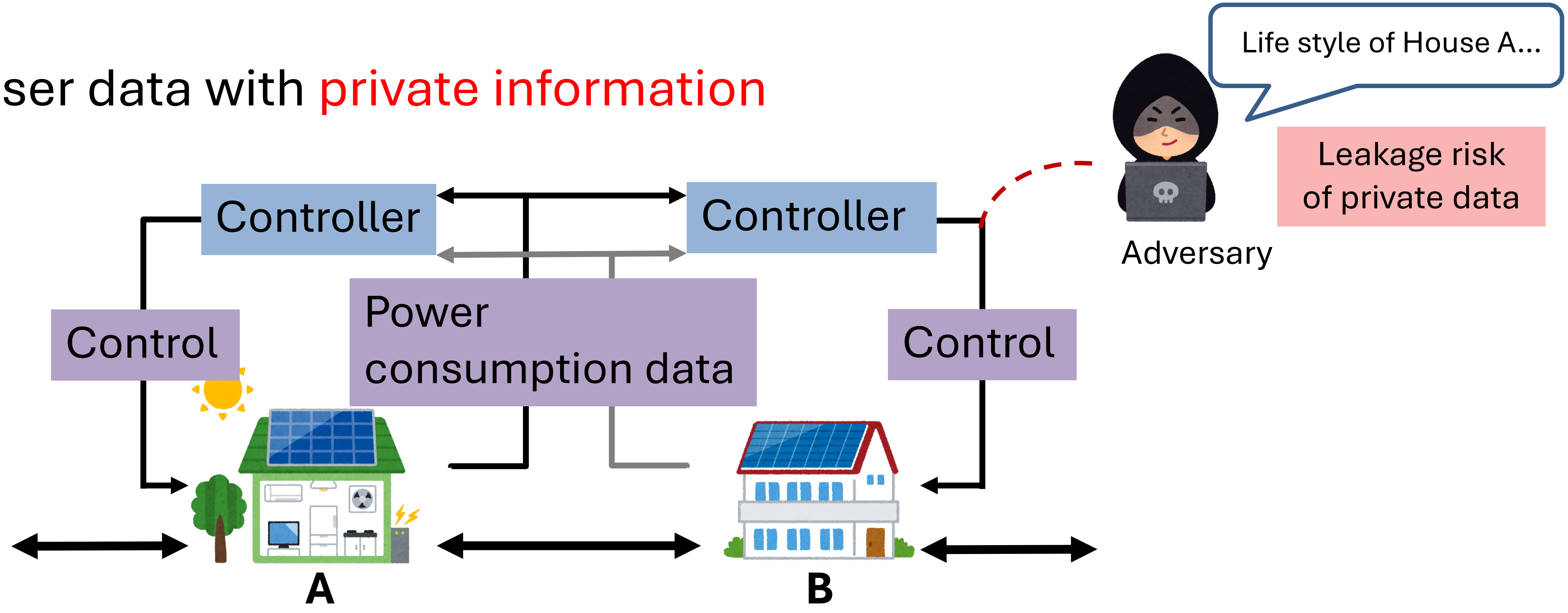


Outline

- Gaussian distribution revisited
  - Affinity to linear systems and its limit
- Key theoretical results
  - Linear system analysis and equivalent linearization
- **New application**
  - **Control systems privacy**
- Sparsity VS rare events
  - Sparse optimal stochastic control

Control systems security

- Utility of user data with private information

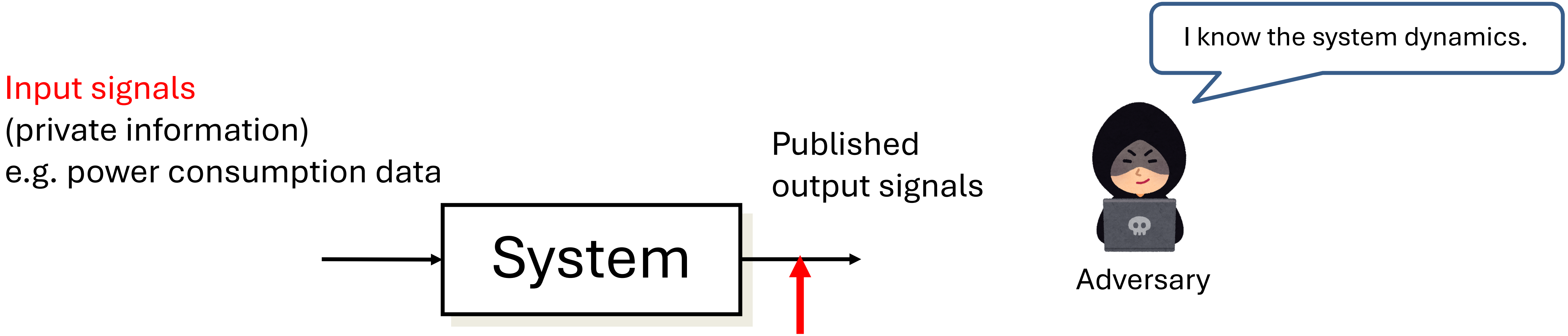


- Need for control method considering privacy protection
  - Trade-off “Privacy protection level vs. information usefulness”

Differential privacy

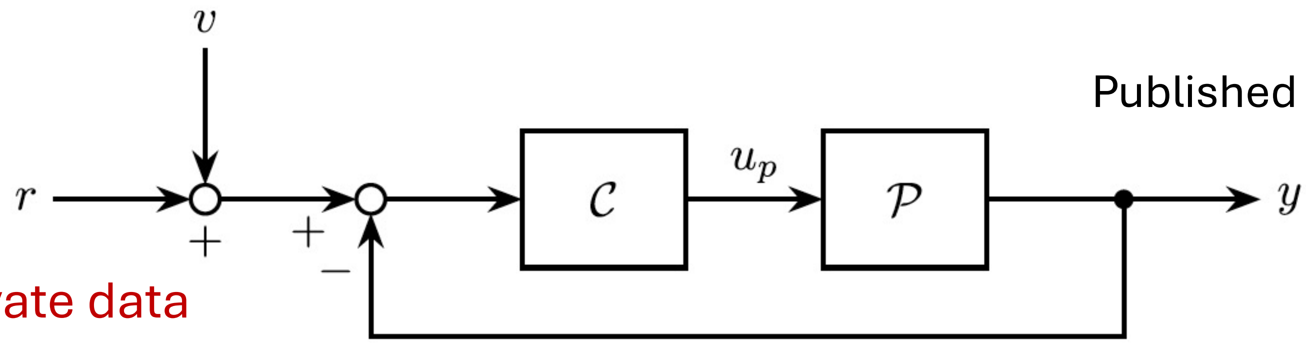
- Protection by adding noise
  - Large noise decreases the information usefulness

- Differential privacy : Difficulty of distinguishing input signals
  - can be viewed as the degree of unobservability.
  - Output noise statistics is crucial for the differential privacy calculation.



Bayesian differential privacy: Example

Intentional noise



Private data

Published output data

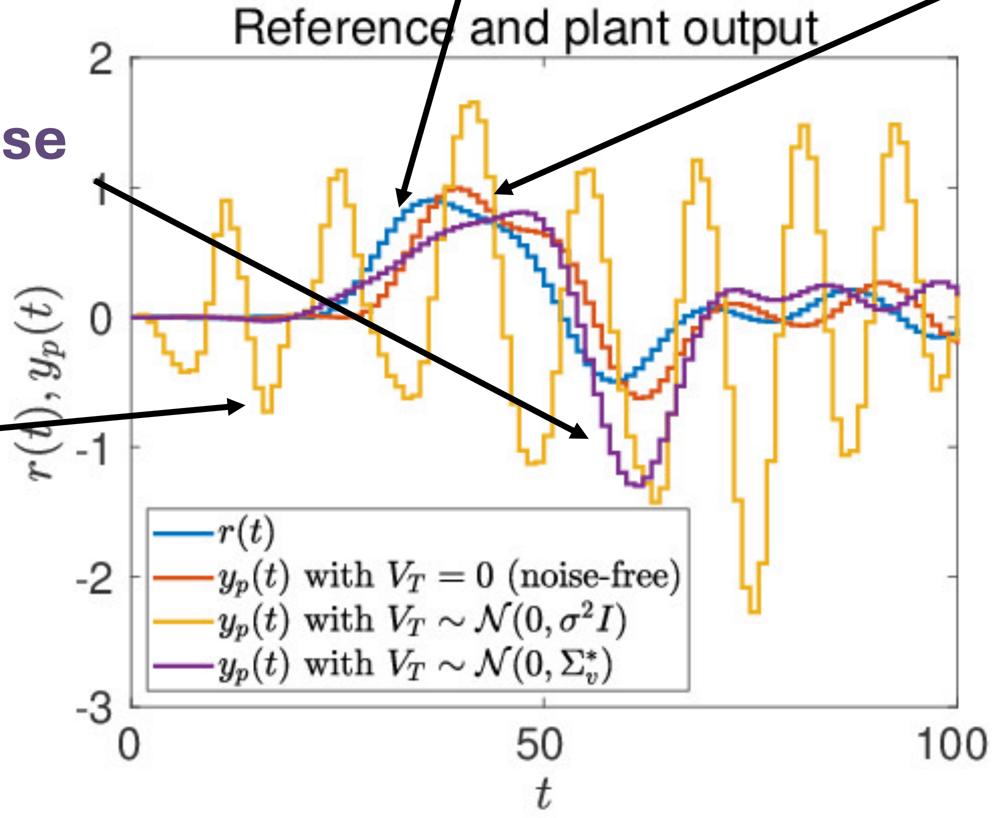
Reference

No noise (good tracking, no privacy)

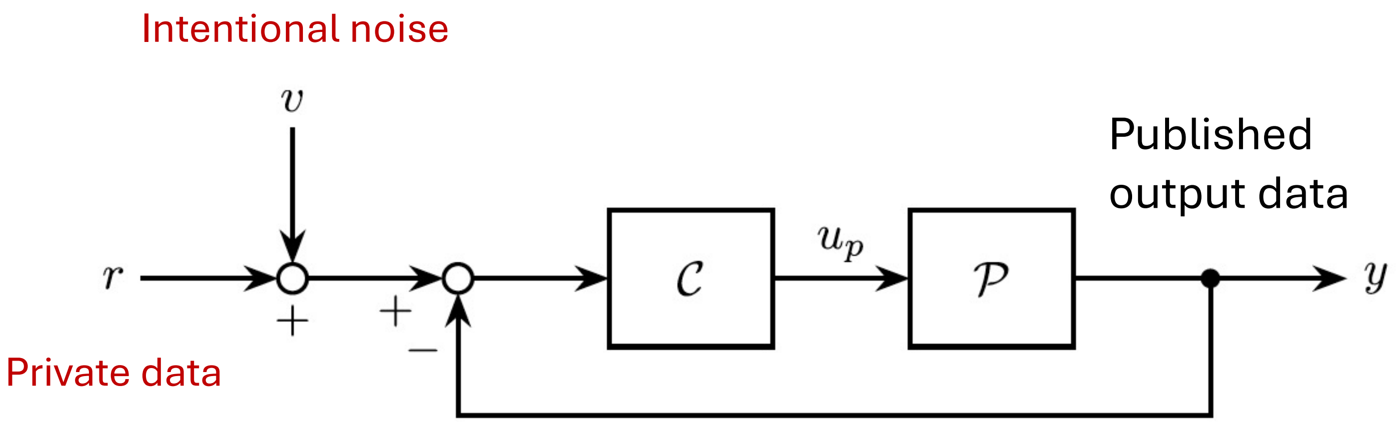
Same differential privacy level

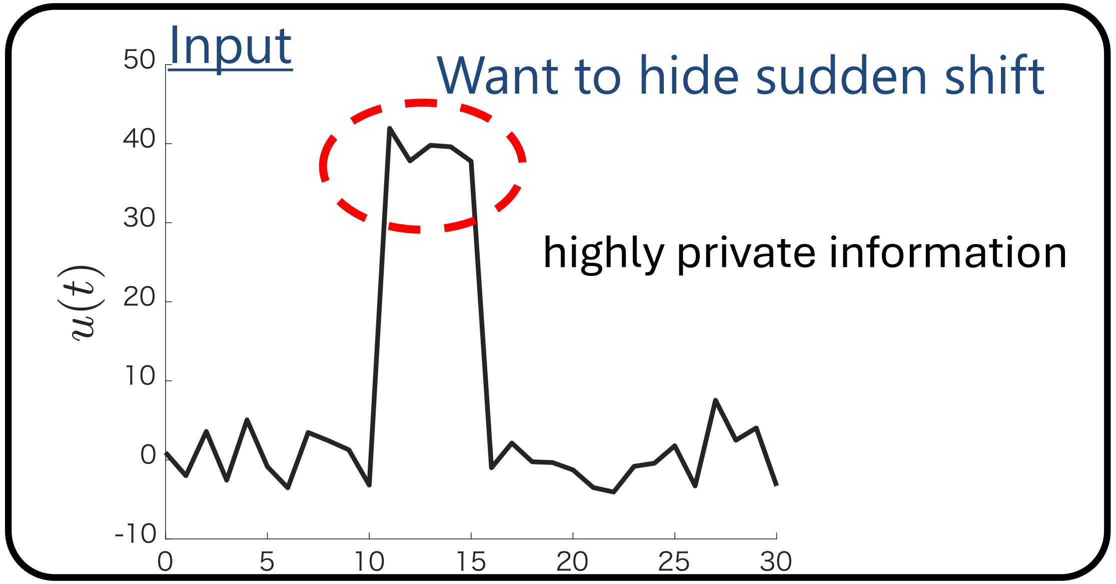
Bayesian optimal noise (minor deterioration)

Optimal i.i.d. noise (not tracking at all)

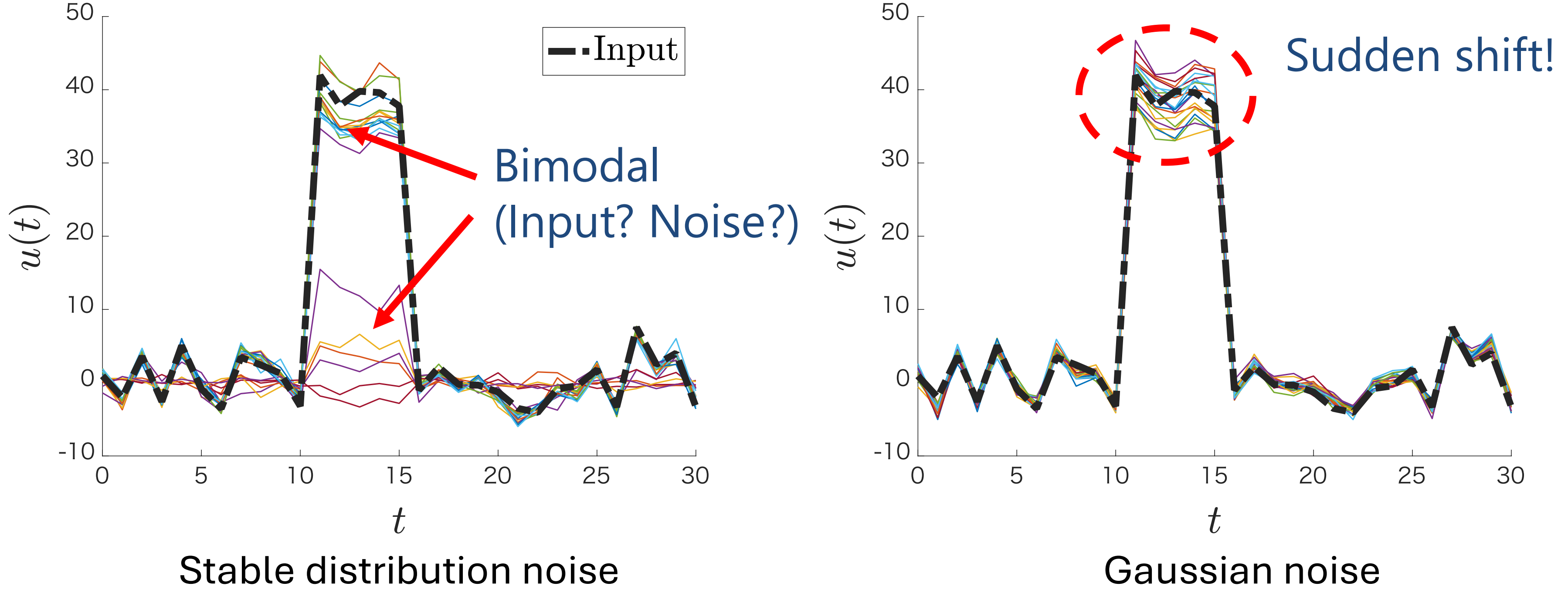


Stable distribution noise can hide outliers: Example





Adversary’s estimate

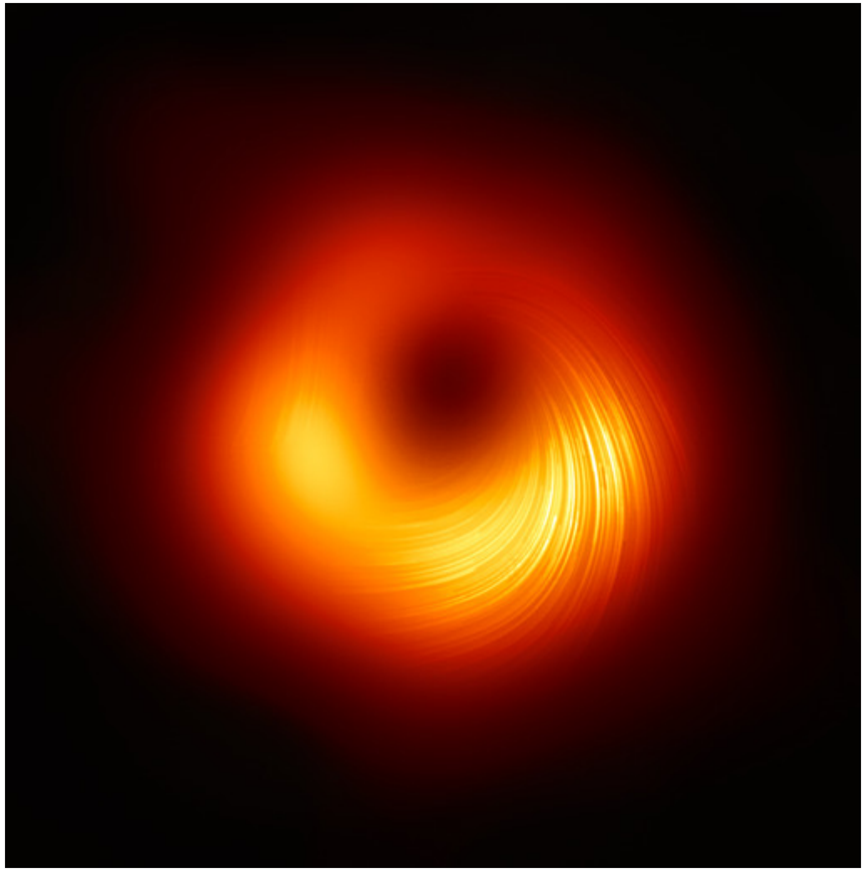


Outline

- Gaussian distribution revisited
  - Affinity to linear systems, its limitation and generalization
- Key theoretical results
  - Linear system analysis and equivalent linearization
- New application
  - Control systems privacy
- **Sparsity VS rare events**
  - **Sparse optimal stochastic control**

Sparsity

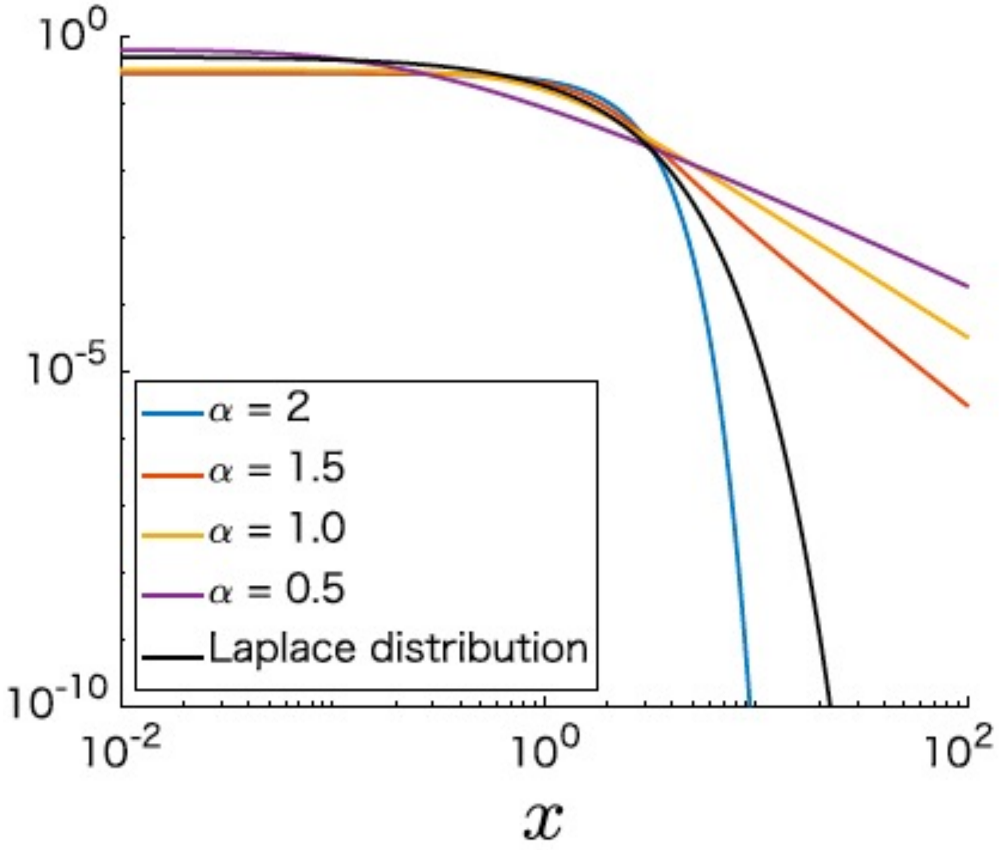
- Sparsity plays a key role of recent AI techniques.
  - Image processing (MRI, EHT), ML (Dropout, LASSO)



https://eventhorizontelescope.org/

Laplace distribution

- Popular model for heavy tail distribution
- Simple density function $\varphi(x) \propto \exp(-|x|)$
  - Slower decay than $\propto \exp(-x^2)$
- No superposition principle



Modal trajectory estimation

- Discrete-time system
  - $x_{k+1} = f(x_k) + Bv_k, y_k = h(x_k) + w_k$
  - For observed trajectory $\mathrm{y}_{0:k}$, estimate real trajectory $\mathrm{x}_{0:k}$

- MAP (Maximum a posteriori) estimation
  - $\varphi(\mathrm{x}_{0:k}|\mathrm{y}_{0:k})$

$$\varphi_{x_{0:k}}(\mathbf{x}_{0:k}|y_{0:k} = \mathbf{y}_{0:k}) = \frac{\varphi_{y_{0:k}}(\mathbf{y}_{0:k}|x_{0:k} = \mathbf{x}_{0:k})\,\varphi_{x_{0:k}}(\mathbf{x}_{0:k})}{\varphi_{y_{0:k}(\mathbf{y}_{0:k})}}$$

Equivalence between modal trajectory estimate and optimal control

$$\varphi_{x_{0:k}}(\mathbf{x}_{0:k}|y_{0:k}=\mathbf{y}_{0:k}) = \frac{\varphi_{y_{0:k}}(\mathbf{y}_{0:k}|x_{0:k}=\mathbf{x}_{0:k})\,\varphi_{x_{0:k}}(\mathbf{x}_{0:k})}{\varphi_{y_{0:k}}(\mathbf{y}_{0:k})}$$

- MAP estimate
  - $x_{k+1} = f(x_k) + Bv_k, y_k = h(x_k) + w_k$
  - For observed trajectory $\mathrm{y}_{0:k}$, maximiz $\varphi_{x_{0:k}}(\mathbf{x}_{0:k}|y_{0:k}=\mathbf{y}_{0:k})$
    $\mathrm{x}_{0:k}$

$$\boxed{v_k \Leftrightarrow u(k)}$$

- Optimal control
  - $x(k+1) = f(x(k)) + Bu(k)$
  - Minimize $\ell(0,x(0)) + \sum_{k=0}^{n-1} \ell(i,x(i),u(i))$

log of prior pdfs

$$\ell(i,\mathbf{x},\mathbf{u}) := \log\varphi_{w_i}(\mathbf{y}_i - h(\mathbf{x})) + \log\varphi_{v_i}(\mathbf{u}),\ \ell(0,\mathbf{x}) := \log\varphi_{x_0}(\mathbf{x})$$

Laplace prior $\varphi(x) \propto \exp(-|x|)$ leads to LASSO

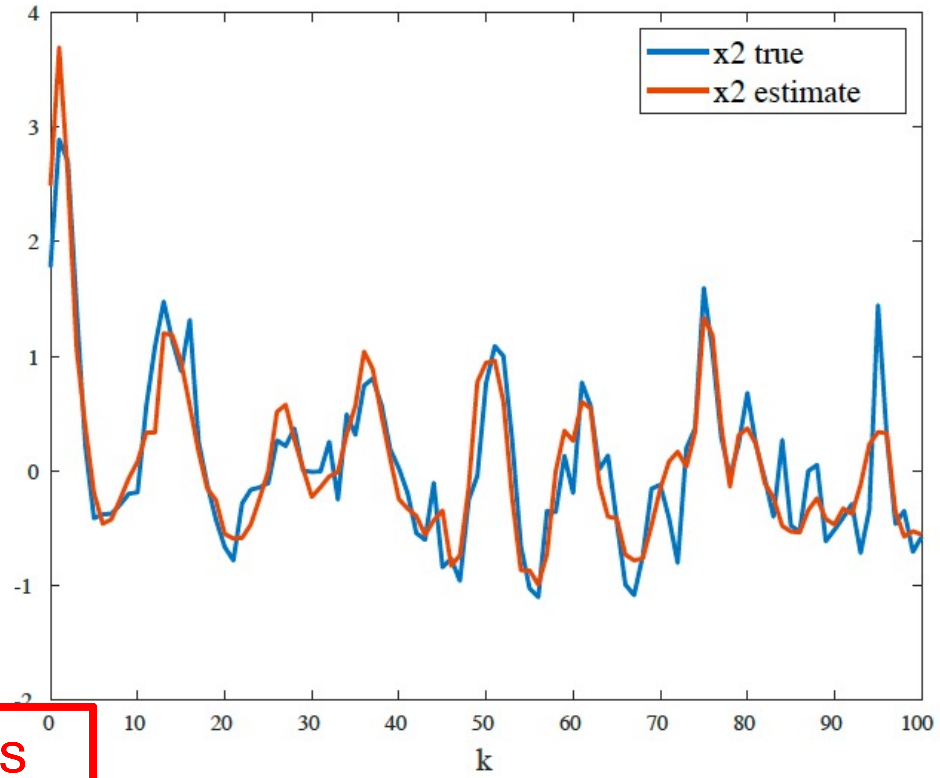
$$x_{k+1} = Ax_k + 0.1v_k,\ x_0 \sim \mathcal{N}(\mu_0, 0.4I)$$

$$y_k = Cx_k + 0.1w_k$$

$v_k, w_k \sim \mathrm{Lap}(0,1)$, i.i.d.

$$A := \begin{bmatrix} 1.12 & -.49 \\ 1 & 0 \end{bmatrix},\ C := \begin{bmatrix} 1 & 0 \end{bmatrix},\ \mu_0 := \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\boxed{v_k \Leftrightarrow u(k)}$$



log of prior pdfs

$$\begin{array}{ll} \text{minimize}_{\mathbf{x}_0,\mathbf{u}} & 2.5\|\mathbf{x}_0 - \mu_0\|^2 + 10\sum_{i}^{N-1} \|\mathbf{u}_i\|_1 + \|C\mathbf{x}_i - \mathbf{y}_i\|_1 \\ \text{subj. to} & \mathbf{x}_{k+1} = A\mathbf{x}_k + \mathbf{u}_k \end{array}$$

Minimization with ℓ^1 regularization

Sparsity of solutions of LASSO

- Ill-conditioned linear equation $Ax = b$
  - Feature extraction, Small-data ML

$$x = [x_1, x_2, \dots, x_n]'$$

- The solution having the minimum ℓ^0-norm
  - ℓ^0-norm: number of non-zero elements
  - Combinatorial optimization

$$\#\{i : x_i \neq 0\}$$

- The solution having the minimum ℓ^1-norm
  - ℓ^1-norm: sum of the absolute values of elements
  - Convex optimization
  - Guarantee for the sparsity under mild assumption

$$\sum_i |x_i|$$

Sparse optimal stochastic control

ℓ^1-regularization (LASSO) gives sparse solutions.
Can we relate the sparse optimal control with its L^1 counterpart?

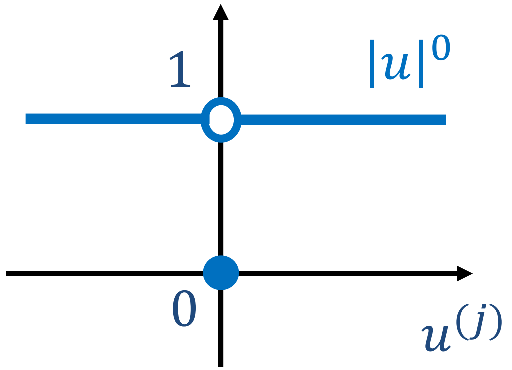
Sparse optimal stochastic control

Minimize $\underset{u}{}$ $\quad \mathbb{E}\left[\underbrace{\int_0^T \sum_{j=1}^m |u_t^{(j)}|^0 \, dt}_{L^0 \text{ norm}} + \underbrace{g(x_T)}_{\text{(Continuous) terminal cost}}\right]$, $\quad T$: final time

Subject to $\quad dx_t = f(x_t, u_t)dt + \sigma(x_t, u_t)dw_t$

$x_0 = x_{\mathrm{i}},\ \{u_t\} \in \mathcal{U} \qquad (x_{\mathrm{i}} \in \mathbb{R}^n)$

- $\mathcal{U} :=$ {Causal $\{u_t\}$ valued in $\mathbb{U}$}
- $\mathbb{U} \subset \mathbb{R}^m$: a compact set that contains 0
- $f(x,u)$ and $\sigma(x,u)$ are Lipschitz continuous in x uniformly in u.

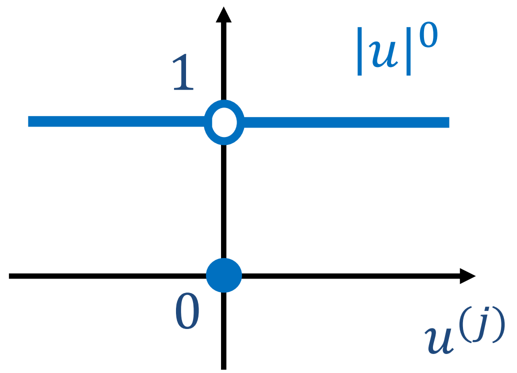


Result (continuity and HJB equation)

- The value function V is continuous on $\mathbb{R}^n \times [0, T]$.

$$V(x,t) := \inf_{u \in \mathcal{U}} \mathbb{E}\left[\sum_{j=1}^{m} \int_t^T |u_s^{(j)}|^0 ds + g(x_T) \,\middle|\, x_t = x\right]$$

Discontinuous cost



- V is a viscosity solution of the HJB equation:

$$\begin{cases} -v_t(x,t) + H(x, v_x(x,t), v_{xx}(x,t)) = 0, \ \mathbb{R}^n \times (0,T), \\ v(x,T) = g(x) \quad \text{in} \ \ \mathbb{R}^n \end{cases}$$

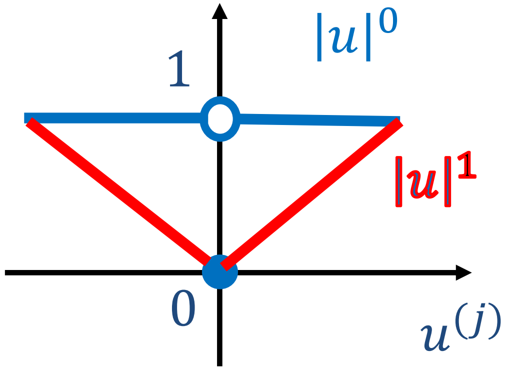
$$H(x,p,M) := \sup_{u \in \mathbb{U}} \left\{ -f(x,u)^\top p - \frac{1}{2}\mathrm{tr}(\sigma\sigma^\top(x,u)M) - \sum_{j=1}^{m} |u^{(j)}|^0 \right\}$$

Result (relationship with L^1 optimization, discreteness)

- Equivalence between L^0 and L^1 optimality

◆ L^1 optimal control problem

$$\underset{u}{\text{Minimize}} \quad \mathbb{E}\left[\int_0^T \sum_{j=1}^{m} |u_t^{(j)}|^1 dt + g(x_T)\right]$$



- For control-affine systems, the optimal control process is Bang-Off-Bang
  – takes only three values of $\{-1, 0, 1\}$

Example

Main problem

Minimize $\underset{u}{}$ $\quad \mathbb{E}\left[\int_0^T |u_t|^0 dt + x_T^2\right]$

Subject to $\quad dx_t = x_t dt + u_t dt + 0.1\, dw_t$

$x_0 = x, \quad u_t \in [-1,1] \quad \forall t \in [0,1]$

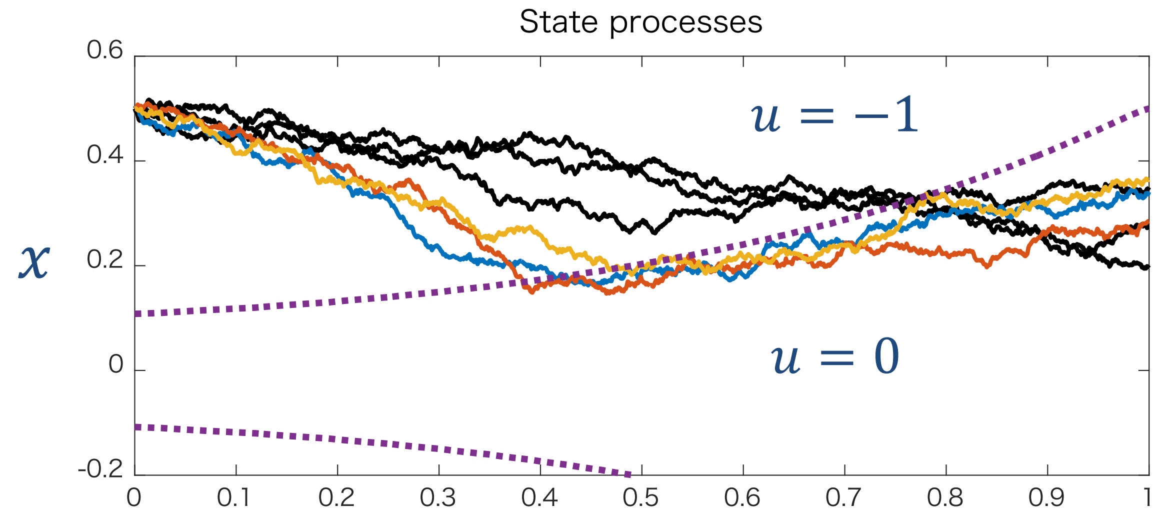
Equivalent relaxed problem

Minimize $\underset{u}{}$ $\quad \mathbb{E}\left[\int_0^T |u_t|^1 dt + x_T^2\right]$

Subject to $\quad dx_t = x_t dt + u_t dt + 0.1\, dw_t$

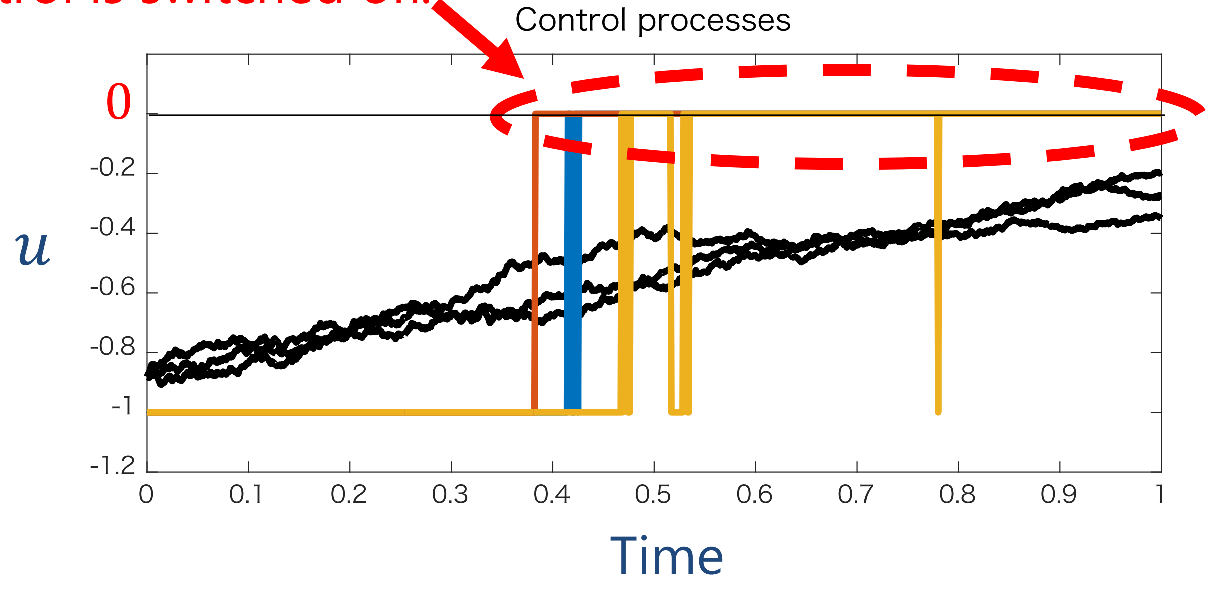
$x_0 = x, \quad u_t \in [-1,1] \quad \forall t \in [0,1]$

Example: $dx_t = x_t dt + u_t dt + 0.1\, dw_t, t \in [0,1]$

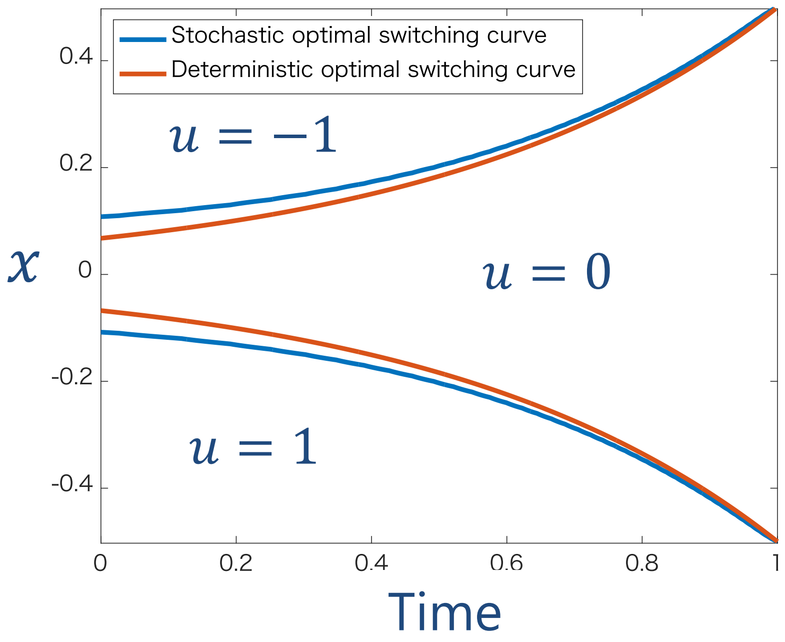


- Blue, red, yellow : L^0 optimal sample paths
- Black: LQG optimal sample paths

The control is switched off.



Switching curve ($\sigma = 0.1$)
" ($\sigma = 0$)



Rare Events Modeling for Linear Estimation and Control

- Gaussian distribution revisited
  - Affinity to linear systems, its limitation and generalization

- Key theoretical results
  - Linear system analysis and equivalent linearization

- New application
  - Control systems privacy

伊藤 海斗，加嶋 健司，
動的システムにおけるレアイベントモデリングとその応用 ー安定分布によるアプローチー，
電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, Vol. 14, Issue 4, pp. 269-278, 2021.

- Sparsity VS rare events
  - Sparse optimal stochastic control ← Ito, K, *Automatica*, 2021