Sequential Coding of Gauss–Markov Sources with Packet Erasures and Feedback

Anatoly Khina, Caltech

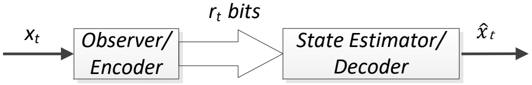
Joint work with
Victoria Kostina and Babak Hassibi, Caltech
Ashish Khisti, University of Toronto

ITW
Kaohsiung, Taiwan
November 10, 2017

Tracking Gauss–Markov Processes over a Noiseless Channel

Tracking Gauss–Markov Processes over a Noiseless Channel

- $x_{t+1} = \alpha x_t + w_t$
- w_t – White ~~Gaussian~~Kaohsiung noise $\mathcal{N}(0, W)$



- **Goal:** Minimize the distortions $D_t \triangleq \mathbb{E}\left[(x_t - \hat{x}_t)^2\right]$
- Might be a tension betweeen optimizing D_{t_1} and D_{t_2}

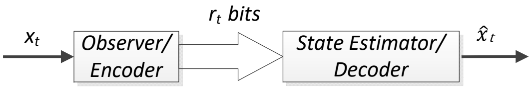
Noiseless finite-rate channel of rate r

Fixed rate: Exactly r bits are available at every time sample t

Variable rate: r bits are available **on average** at every t

Tracking Gauss–Markov Processes over a Noiseless Channel

- $x_{t+1} = \alpha x_t + w_t$
- w_t – White ~~Gaussian~~Kaohsiung noise $\mathcal{N}(0, W)$



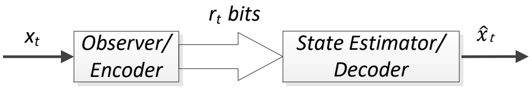
- **Goal:** Minimize the distortions $D_t \triangleq \mathbb{E}\left[(x_t - \hat{x}_t)^2\right]$
- Might be a tension betweeen optimizing D_{t_1} and D_{t_2}

Random-rate budget

At time t: Exactly r_t bits are given.

Tracking Gauss–Markov Processes over a Noiseless Channel

- $x_{t+1} = \alpha x_t + w_t$
- w_t – White ~~Gaussian~~Kaohsiung noise $\mathcal{N}(0, W)$



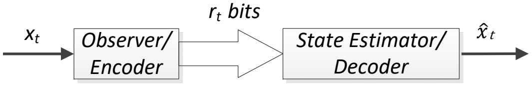
- **Goal:** Minimize the distortions $D_t \triangleq \mathbb{E}\left[(x_t - \hat{x}_t)^2\right]$
- Might be a tension betweeen optimizing D_{t_1} and D_{t_2}

Packet erasures with instantaneous acknowledgments (ACKs)

[Minero et al. AC'09]: Erasure + ACK $\iff$ $r_t = 0$

Tracking Gauss–Markov Processes over a Noiseless Channel

- $x_{t+1} = \alpha x_t + w_t$
- w_t – White ~~Gaussian~~Kaohsiung noise $\mathcal{N}(0, W)$



- **Goal:** Minimize the distortions $D_t \triangleq \mathbb{E}\left[(x_t - \hat{x}_t)^2\right]$
- Might be a tension betweeen optimizing D_{t_1} and D_{t_2}

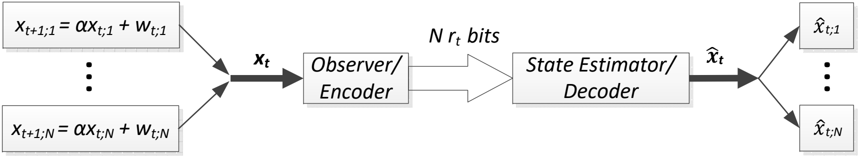
Packet erasures with delayed acknowledgments (ACKs)

More tricky... We'll get back to it later...

Multi-Tracking Gauss–Markov Processes over a Noiseless Channel

Multi-track

Several processes are tracked and controlled over a shared channel.



- N — Number of tracked processes

(Multi-)Track: Impossibility

Lower bound

- Given rates $R_1, \ldots, R_T$
- $D_t \geq D_t^*$ where

$$D_t^* = \left(\alpha^2 D_{t-1}^* + W\right) 2^{-2R_t}$$
$$D_0^* = 0$$

Previous results

- Sum-rate explicitly proved by [Ma-Ishwar IT'11]
- Steady-state can be deduced from [Gorbunov-Pinsker PPI'74] [Tatikonda-Sahai-Mitter AC'04][Kostina-Hassibi Allerton'16]

(Multi-)Track: Impossibility

Lower bound

- Given rates $R_1, \ldots, R_T$
- $D_t \geq D_t^*$ where

$$D_t^* = \left(\alpha^2 D_{t-1}^* + W\right) 2^{-2R_t}$$
$$D_0^* = 0$$

Inductive proof sketch

- Condition on previously received packets
- Shannon's lower bound $\Rightarrow$ Entropy-power calculations
- Entropy-power inequality $\Rightarrow$ Separates w_t from $\alpha(x_t - \hat{x}_t)$
- Jensen's inequality + simple IT inequalities

(Multi-)Track: Impossibility

Lower bound

- Random rates $r_1, \ldots, r_T$
- $D_t \geq D_t^*$ where

$$D_t^* = \left(\alpha^2 D_{t-1}^* + W\right) \mathbb{E}\left[2^{-2r_t}\right]$$

$$D_0^* = 0$$

Proof adjustment

- Condition on rates
- Take expectation at the end

(Multi-)Track: Impossibility

Lower bound

- Random rates $r_1, \ldots, r_T$
- $D_t \geq D_t^*$ where

$$D_t^* = \left(\alpha^2 D_{t-1}^* + W\right) \mathbb{E}\left[2^{-2r_t}\right]$$

$$D_0^* = 0$$

Track vs. Multi-track

- Holds for any number of tracked processes N
- Applies for single-process tracking

Multi-Track: Achievability

Upper bound

- Any $\epsilon > 0$ and large enough N
- Given rates $R_1, \ldots, R_T$
- $D_t \leq D_t^* + \epsilon$ where

$$D_t^* = \left(\alpha^2 D_{t-1}^* + W\right) 2^{-2R_t}$$

$$D_0^* = 0$$

Multi-Track: Achievability

Upper bound

- Any $\epsilon > 0$ and large enough N
- Random rates $r_1, \ldots, r_T$
- $D_t \leq D_t^*$ where

$$D_t^* = \left(\alpha^2 D_{t-1}^* + W\right) \mathbb{E}\left[2^{-2r_t}\right]$$
$$D_0^* = 0$$

Multi-Track: Achievability

Optimal multi-track for large N

- The upper and lower bounds coincide for large N
- Optimal distortions:

$$D_t^* = \left(\alpha^2 D_{t-1}^* + W\right) \mathbb{E}\left[2^{-2r_t}\right]$$
$$D_0^* = 0$$

- Optimal performance achieved by greedy quantization
- No tension between optimizing D_{t_1} and D_{t_2}

Multi-Track: Achievability

Optimal multi-track for large N

- The upper and lower bounds coincide for large N
- Optimal distortions:

$$D_t^* = \left(\alpha^2 D_{t-1}^* + W\right) \mathbb{E}\left[2^{-2r_t}\right]$$

$$D_0^* = 0$$

Steady-state distortion

$$D_\infty^* \triangleq \lim_{t \to \infty} D_t^* = \frac{W \mathbb{E}\left[2^{-2r_1}\right]}{1 - \alpha^2 \mathbb{E}\left[2^{-2r_1}\right]}$$

- Recovers the data–rate theorem of [Minero et al. AC'09]:

$$\alpha^2 \mathbb{E}\left[2^{-2r_1}\right] < 1$$

- Or in the deterministic case: $R > \log|\alpha|$

Multi-Track: Achievability

DPCM scheme

Observer at time t

- Generates the prediction error
$$\tilde{\boldsymbol{x}}_t \triangleq \boldsymbol{x}_t - \alpha \hat{\boldsymbol{x}}_{t-1}$$
- Quantizes the error:
$$\hat{\tilde{\boldsymbol{x}}}_t = Q_t(\tilde{\boldsymbol{x}}_t)$$
- Sends the quantization index
- Generates next estimate:
$$\hat{\boldsymbol{x}}_t = \alpha \hat{\boldsymbol{x}}_{t-1} + \hat{\tilde{\boldsymbol{x}}}_t$$

State estimator at time t

- Receives quantization index
- Recovers $\hat{\tilde{\boldsymbol{x}}}_t$
- Generates an estimate of $\boldsymbol{x}_t$:
$$\hat{\boldsymbol{x}}_t = \alpha \hat{\boldsymbol{x}}_{t-1} + \hat{\tilde{\boldsymbol{x}}}_t$$

Random rates

Algorithm is greedy $\Rightarrow$ Does not need to know rates in advance

Packet Erasures with Instantaneous Feedback

One packet per one state frame

- I.i.d. packet erasures: $b_t = \begin{cases} 1 & \text{w.p. } \beta & (\textit{arrived}) \\ 0 & \text{w.p. } 1-\beta & (\textit{erased}) \end{cases}$
- Instantaneous feedback $\Rightarrow$ Random-rate budget scenario
- I.i.d. rates $r_t = Rb_t = \begin{cases} R, & \text{w.p. } \beta \\ 0, & \text{w.p. } 1-\beta \end{cases}$

Packet Erasures with Instantaneous Feedback

Multiple packets per one state frame

- I.i.d. $\mathscr{Ber}(\beta)$ packet erasures
- K packets each of rate R/K
- $b_t \sim \mathscr{Bin}(K, \beta)$ — Number of **successful** packet arrivals
- Instantaneous feedback $\Rightarrow$ Random-rate budget scenario
- I.i.d. rates $r_t = R\, b_t / K$

Packet Erasures with Instantaneous Feedback

Steady-state distortion

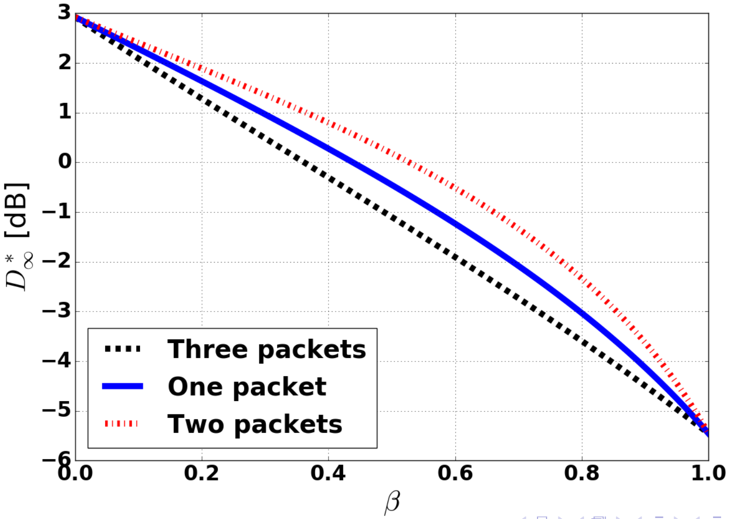
$$D_\infty^* \triangleq \lim_{t\to\infty} D_t^* = \frac{W\mathbb{E}\left[2^{-2r_1}\right]}{1-\alpha^2\mathbb{E}\left[2^{-2r_1}\right]}$$

Are more packets necessarily better?

- More packets $\Rightarrow$ Higher chance to convey with non-zero rate
- Less packets $\Rightarrow$ Higher chance to send full rate R
- How to choose the optimal K?
- Choose K that optimizes $\mathbb{E}\left[2^{-2r_t}\right]$ w.r.t. $b_t \sim \mathscr{Bin}(K,\beta)$

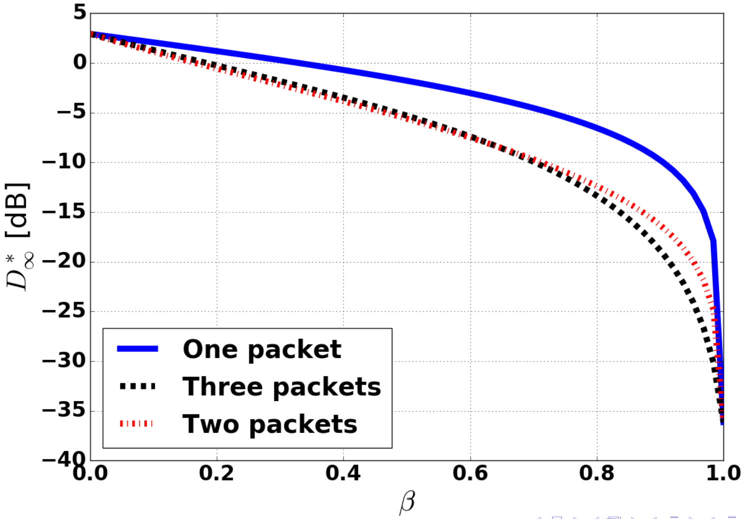
Packet Erasures with Instantaneous Feedback

- $\boldsymbol{x}_t = \alpha \boldsymbol{x}_t + \boldsymbol{w}_t$, with $\alpha = 0.7$, $R = 1$, $W = 1$



Packet Erasures with Instantaneous Feedback

- $\mathbf{x}_t = \alpha \mathbf{x}_t + \mathbf{w}_t$, with $\alpha = 0.7$, $R = 6$, $W = 1$



Packet Erasures with Delayed Feedback

Delay by one time unit

Observer at time t:

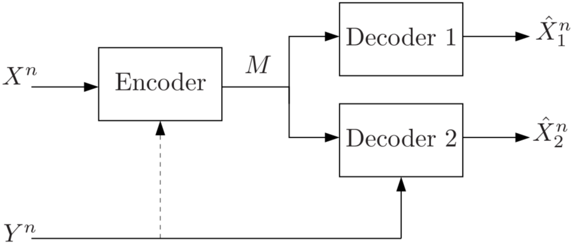
- Does not know whether the last packet arrived or not (b_{t-1})
- Knows whether all preceding packets arrived or not (b^{t-2})

Idea

At time t, treat packet of time $t-1$ as

- Side-information that may be known at the decoder
- Side-information that is known at the encoder

Packet Erasures with Delayed Feedback



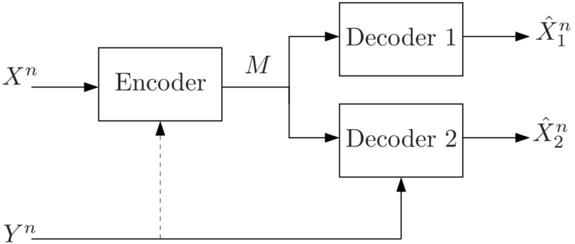
(Fig. from the lectures notes of El Gamal & Kim)

Idea

At time t, treat packet of time $t-1$ as side information

- Side-information that may be known at the decoder
- Side-information that is known at the encoder

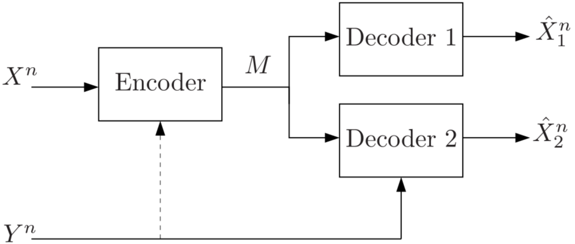
Packet Erasures with Delayed Feedback



Idea

- $X^n \to \boldsymbol{x}_t$
- $Y^n \to$ Packet sent at time $t-1$
- Decoder 1 $\to$ Previous packet did not arrive
- Decoder 2 $\to$ Previous packet arrived successfully

Packet Erasures with Delayed Feedback

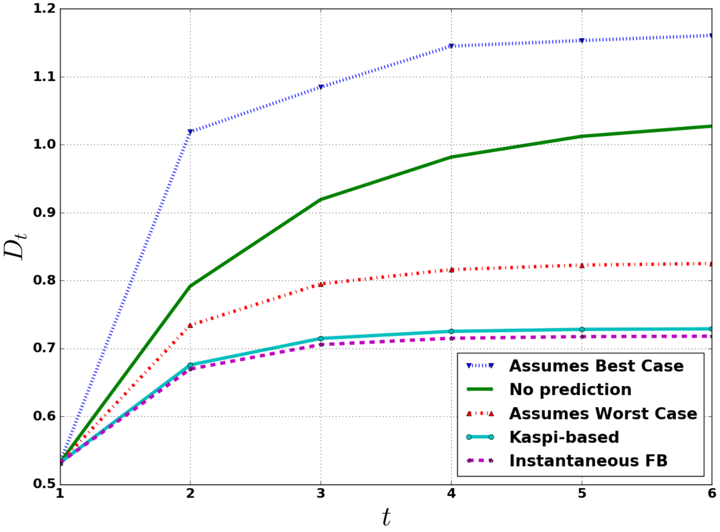


Rate–distortion region [Kaspi IT'94 ('80)]

- Gaussian case explicitly solved by [Perron et al. ISIT'06]
- Availability of SI @ Tx improves performance!
  - Does not help when SI always available @ Rx (Wyner–Ziv)
  - Does not help in when no SI available @ Rx
- SI@Tx ⇒ (anti-)correlate quantization noise and prediction error

Packet Erasures with Delayed Feedback

- $\mathbf{x}_t = \alpha \mathbf{x}_t + \mathbf{w}_t$, with $\beta = 0.5$, $\alpha = 0.7$, $R = 2$, $W = 1$



Variable-Length Coding

- Back to tracking a *single* process

Fixed-length coding

Exactly R bits are available at time t.

Variable-length coding

R bits are available *on average* at time t.

- Rx decides on bit allocation depending on the source value

Variable-Length Coding

Entropy-Coded Dithered Quantization (ECDQ) [Ziv'85][Zamir-Feder'96]

- Quantize x using an infinite grid of step size Δ
- Apply entropy-coding to resulting bits

Improvement 1: Add uniform random dither z modulo Δ:

$$[x + z] \mod \Delta$$

- Makes quantization error uniform and independent of x

Improvement 2: Linear pre- and post-processing

- Essentially applies MMSE estimation

Variable-Length Coding

ECDQ performance [Zamir-Feder IT'96]

- Source power X
- R bits on average

$$R \leq \frac{1}{2}\log\frac{X}{D} + \frac{1}{2}\log\frac{2\pi e}{12} \quad \Leftrightarrow \quad D \leq \frac{2\pi e}{12} X 2^{-2R}$$

- Applies for any distribution of x
- For higher dimensions shaping loss decreases
- For $N \to \infty$, loss goes to 0
- We assume one-to-one codes (not prefix free)

Variable-Length Coding

DPCM-ECDQ scheme

- Use ECDQ as the quantizer in the DPCM scheme
- Similar scheme proposed in [Silva-Derpich-Østergaard IT'11]
- Achieves the following distortions:

$$D_t = \frac{2\pi e}{12} \left(\alpha^2 D_{t-1} + W\right) \mathbb{E}\left[2^{-2r_t}\right]$$

- Steady-state: $D_\infty^* = \dfrac{\frac{2\pi e}{12} W \mathbb{E}\left[2^{-2r_1}\right]}{1 - \frac{2\pi e}{12} \alpha^2 \mathbb{E}\left[2^{-2r_1}\right]}$

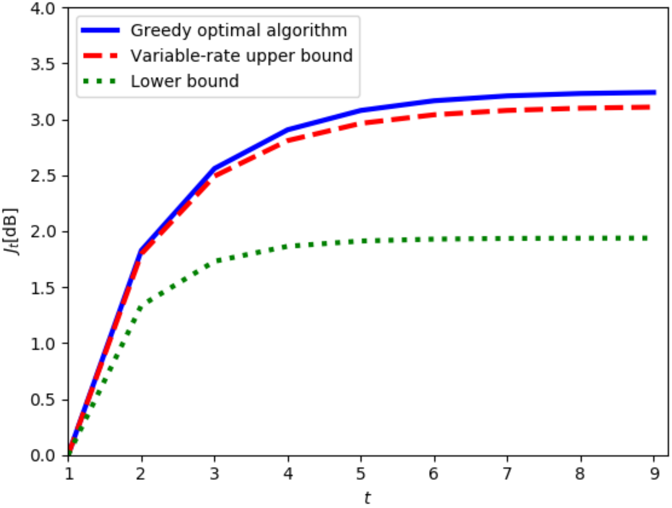
Lower bound

$$D_t = \left(\alpha^2 D_{t-1} + W\right) \mathbb{E}\left[2^{-2r_t}\right]$$

- Steady-state: $D_\infty^* = \dfrac{W \mathbb{E}\left[2^{-2r_1}\right]}{1 - \alpha^2 \mathbb{E}\left[2^{-2r_1}\right]}$

Variable-Length Coding for Control

- $A = 1.2, W = 1, \mathrm{Q}_t \equiv 1, \mathrm{R}_t = 0$, no erasures: $J_t = \alpha^2 D_t + W$



Multi-Track: Rate Allocation

Goal

Maximize average (across time) distortion: $\bar{D}_\infty = \frac{1}{T}\sum_{t=1}^{T} D_t$

- Total rate budget $\frac{1}{T}\sum_{t=1}^{T} R_t \leq R$
- How to allocate transmission rates $R_1, \ldots, R_T$?

No erasures

- Recall: $D_t = \left(\alpha^2 D_{t-1} + W\right) 2^{-2R_t}$
- Convexity arguments $\overset{T\to\infty}{\Longrightarrow}$ Uniform allocation $R_t \equiv R$ optimal

Multi-Track: Rate Allocation

Goal

Maximize average (across time) distortion: $\bar{D}_\infty = \frac{1}{T}\sum_{t=1}^{T} D_t$

- Total rate budget $\frac{1}{T}\sum_{t=1}^{T} R_t \leq R$
- How to allocate transmission rates $R_1, \ldots, R_T$?

With i.i.d. erasures

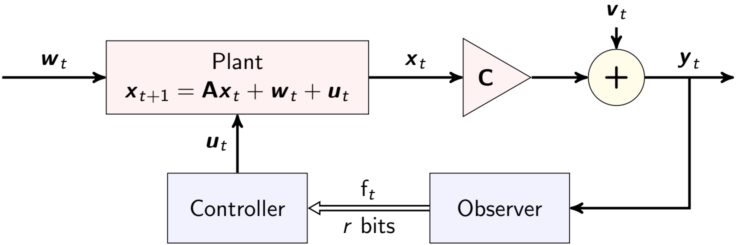
- Reminiscent of channel coding with fading that is known @ Tx
- Water-filling over current distortion level can do better(?)

Linear Quadratic Gaussian Control over Noiseless Channels

Linear quadratic Gaussian (LQG) system

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \text{ i.i.d. } \mathcal{N}(0, \mathbf{W})$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \text{ i.i.d. } \mathcal{N}(0, \mathbf{V})$$



Noiseless finite-rate channel of rate r

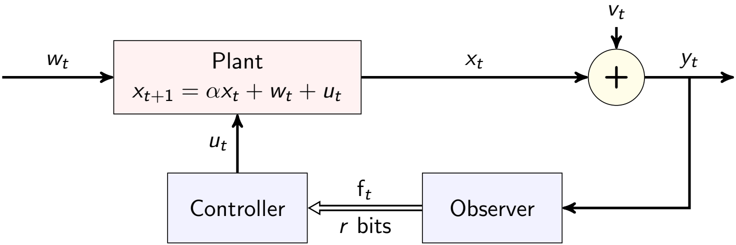
Fixed rate: Exactly r bits are available at every time sample t

Variable rate: r bits are available **on average** at every t

Linear Quadratic Gaussian Control over Noiseless Channels

Scalar linear quadratic Gaussian (LQG) system

$$x_{t+1} = \alpha x_t + u_t + w_t, \qquad w_t \sim \text{ i.i.d. } \mathcal{N}(0, W), \qquad |\alpha| > 1$$
$$y_t = x_t + v_t, \qquad v_t \sim \text{ i.i.d. } \mathcal{N}(0, V)$$



Noiseless finite-rate channel of rate r

Fixed rate: Exactly r bits are available at every time sample t

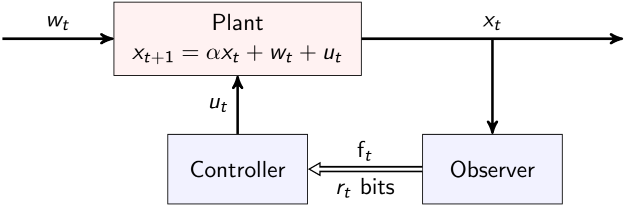
Variable rate: r bits are available **on average** at every t

Linear Quadratic Gaussian Control over Noiseless Channels

Scalar linear quadratic Gaussian (LQG) system

$$x_{t+1} = \alpha x_t + u_t + w_t, \qquad w_t \sim \text{ i.i.d. } \mathcal{N}(0, W), \qquad |\alpha| > 1$$

$$y_t = x_t + \cancel{v_t}, \qquad \cancel{v_t} \sim \cancel{\text{i.i.d. } \mathcal{N}(0, V)}$$



Random-rate budget

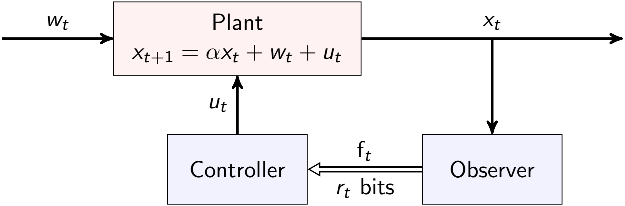
At time t: Exactly r_t bits are given.

Linear Quadratic Gaussian Control over Noiseless Channels

Scalar linear quadratic Gaussian (LQG) system

$$x_{t+1} = \alpha x_t + u_t + w_t, \qquad w_t \sim \text{ i.i.d. } \mathcal{N}(0, W), \qquad |\alpha| > 1$$

$$y_t = x_t + \cancel{v_t}, \qquad \cancel{v_t} \sim \cancel{\text{i.i.d. } \mathcal{N}(0, V)}$$



Packet erasures with instantaneous acknowledgments (ACKs)

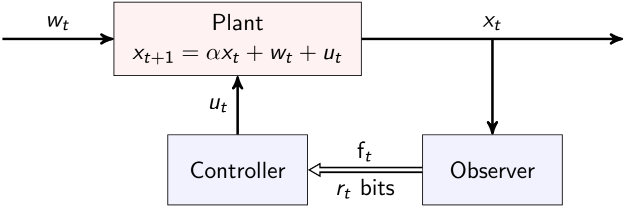
[Minero et al. AC'09]: Erasure + ACK $\Longleftrightarrow$ $r_t = 0$

Linear Quadratic Gaussian Control over Noiseless Channels

Scalar linear quadratic Gaussian (LQG) system

$$x_{t+1} = \alpha x_t + u_t + w_t, \qquad w_t \sim \text{ i.i.d. } \mathcal{N}\left(0, W\right), \qquad |\alpha| > 1$$

$$y_t = x_t + \cancel{v_t}, \qquad \cancel{v_t} \sim \cancel{\text{i.i.d. } \mathcal{N}\left(0, V\right)}$$



Packet erasures with delayed acknowledgments (ACKs)

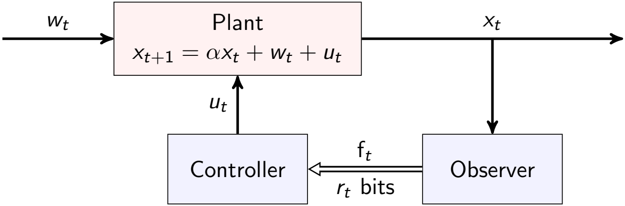
More tricky... We'll get back to it later...

Linear Quadratic Gaussian Control over Noiseless Channels

Scalar linear quadratic Gaussian (LQG) system

$$x_{t+1} = \alpha x_t + u_t + w_t, \qquad w_t \sim \text{ i.i.d. } \mathcal{N}(0, W), \qquad |\alpha| > 1$$

$$y_t = x_t + \cancel{v_t}, \qquad \cancel{v_t} \sim \cancel{\text{i.i.d. } \mathcal{N}(0, V)}$$



LQG cost

$$\bar{J}_T = \mathbb{E}\left[\sum_{t=1}^{T-1}\left(Q_t x_t^2 + R_t u_t^2\right) + Q_T x_T^2\right]$$

The Control–Estimation Separation Principle

Control–estimation separation for networked control systems [Fischer AC'82][Tatikonda-Sahai-Mitter AC'04]

- Optimal control action: $u_t = -K_t \hat{x}_t$
- LQR coefficients: $\begin{cases} K_t = \frac{\alpha L_{t+1}}{R_t + L_{t+1}}, & K_T = 0, \\ L_t = Q_t + \alpha R_t K_t, & L_{T+1} = 0 \end{cases}$
- MMSE estimate: $\hat{x}_t = \mathbb{E}\left[x_t | f^t\right]$
- Optimal cost: $\bar{J}_T^* = \frac{1}{T} \sum_{t=1}^{T} \left(W L_t + \alpha K_t L_{t+1} D_t^* \right)$
- $D_t = \mathbb{E}\left[(x_t - \hat{x}_t)^2 \right]$

- Past u_t known to all $\Rightarrow$ Same $\{D_t\}$ for all control actions $\{u_t\}$
- Control–estimation separation extends to networked control