Feedback Capacity of Gaussian channels and Regret-based Control

Oron Sabag, Caltech

Online Seminar on Control and Information
May 10, 2021

Outline

1. Regret-optimal control

 - This part is based on joint work w. Gautam Goel, Sahin Lale and Babak Hassibi

2. The feedback capacity of Gaussian channels

 - This part is based on joint work w. Victoria Kostina and Babak Hassibi

The LQR setting

- A time-invariant linear dynamical system is given by

$$x_{t+1} = Ax_t + B_u u_t + B_w w_t,$$

  where $x_t \in \mathbb{R}^n$ is state, $u_t \in \mathbb{R}^m$ is the control and $w_t \in \mathbb{R}^p$ is the disturbance vector

- The pair (A, B_u) is stabilizable

The LQR setting

- A time-invariant linear dynamical system is given by

$$x_{t+1} = Ax_t + B_u u_t + B_w w_t,$$

  where $x_t \in \mathbb{R}^n$ is state, $u_t \in \mathbb{R}^m$ is the control and $w_t \in \mathbb{R}^p$ is the disturbance vector

  - The pair (A, B_u) is stabilizable

The operation:

- A policy $\mathcal{K}$ is a linear operator from $w = \{w_t\}$ to $u = \{u_t\}$
  - A causal mapping is a sequence of mappings

$$K_t : (w_{-\infty}, \ldots, w_t) \to u_t$$

  - A strictly causal policy is

$$K_t : (w_{-\infty}, \ldots, w_{t-1}) \to u_t$$

The linear quadratic cost

- The LQR cost of a linear controller $\mathcal{K}$ is

$$\begin{aligned}\mathrm{cost}(\mathcal{K}; w) &= \sum_{t=-\infty}^{\infty} (x_t^* Q x_t + u_t^* R u_t) \\ &\triangleq w^* T_{\mathcal{K}}^* T_{\mathcal{K}} w\end{aligned}$$

where $Q, R \succ 0$ are weight matrices.

The linear quadratic cost

- The LQR cost of a linear controller $\mathcal{K}$ is

$$\begin{aligned}\text{cost}(\mathcal{K}; w) &= \sum_{t=-\infty}^{\infty} (x_t^* Q x_t + u_t^* R u_t) \\ &\triangleq w^* T_{\mathcal{K}}^* T_{\mathcal{K}} w\end{aligned}$$

  where $Q, R \succ 0$ are weight matrices.
- For a linear controller (policy) $\mathcal{K}$, we can always write

$$\begin{bmatrix} x \\ u \end{bmatrix} = \underbrace{\begin{bmatrix} \mathcal{F}\mathcal{K} + \mathcal{G} \\ \mathcal{K} \end{bmatrix}}_{T_{\mathcal{K}}} w. \tag{1}$$

Strategies to design a controller

- One aims minimize the cost

$$\mathrm{cost}(\mathcal{K}; w),$$

  but need to specify the disturbance w

Strategies to design a controller

- One aims minimize the cost

$$\mathrm{cost}(\mathcal{K}; w),$$

but need to specify the disturbance w

- If w_t is white with Gaussian distribution,

$$\min_{\mathcal{K}} \mathbb{E}[\mathrm{cost}(\mathcal{K}; w)],$$

the solution is LQR controller $u_t = -K_{lqr} x_t$ (H_2 control)

Strategies to design a controller

- One aims minimize the cost

$$\mathrm{cost}(\mathcal{K}; w),$$

but need to specify the disturbance w

- If w_t is white with Gaussian distribution,

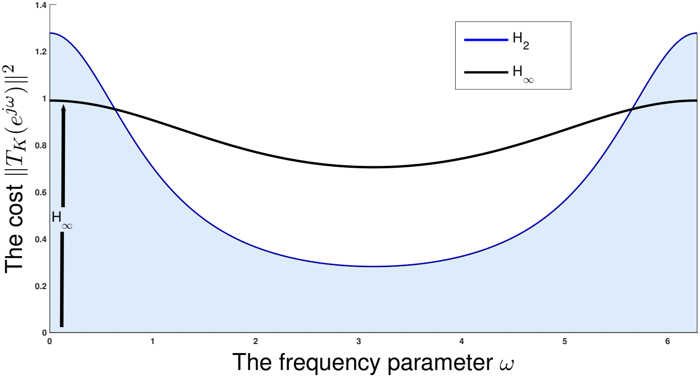
$$\min_{\mathcal{K}} \mathbb{E}[\mathrm{cost}(\mathcal{K}; w)],$$

the solution is LQR controller $u_t = -K_{lqr} x_t$ (H_2 control)

- The robust approach (H_∞ control):

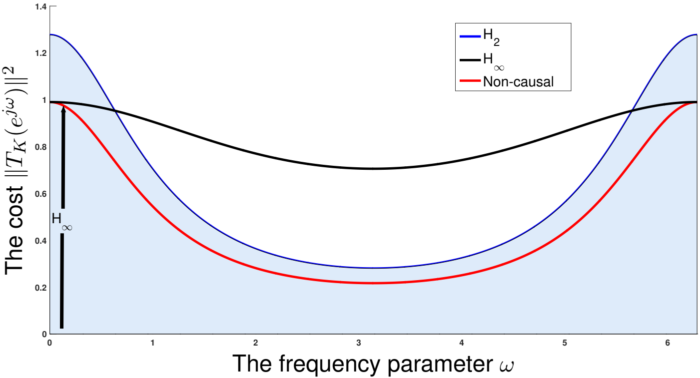
$$\min_{\mathcal{K}} \max_{w \in \ell_2} \frac{\mathrm{cost}(\mathcal{K}; w)}{\|w\|_2}$$

Example



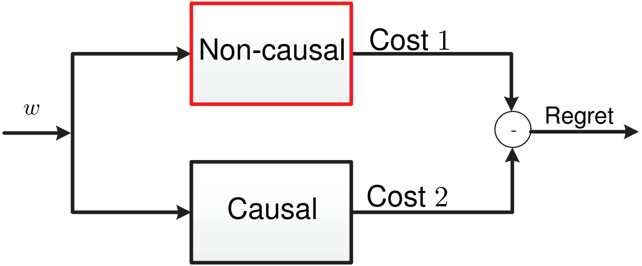
Scalar system with $A = 0.9$ and $B_u = B_w = Q = R = 1$.

Example

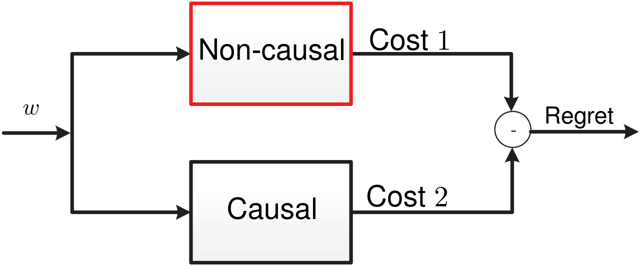


Scalar system with $A = 0.9$ and $B_u = B_w = Q = R = 1$.

The Regret-Optimal Controller



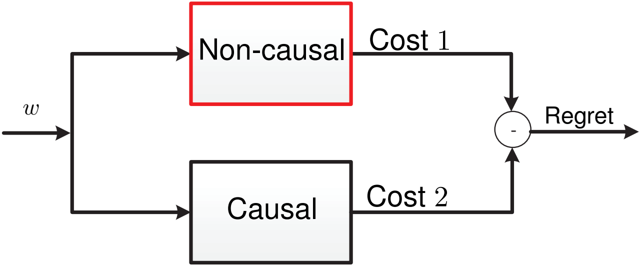
The Regret-Optimal Controller



- Our regret approach:

$$\text{Regret}(\mathcal{K}; w) = \left(\text{cost}(\mathcal{K}; w) - \inf_{\mathcal{K}' \text{ is non-causal}} \text{cost}(\mathcal{K}'; w) \right)$$

The Regret-Optimal Controller



- Our regret approach:

$$\text{Regret}(\mathcal{K};w) = \left(\text{cost}(\mathcal{K};w) - \inf_{\mathcal{K}' \text{ is non-causal}} \text{cost}(\mathcal{K}';w)\right)$$

- The design criterion is the worst-case regret:

$$\text{Regret}^* = \inf_{\mathcal{K} \text{ is causal}} \sup_{\|w\|_2 \leq 1} \text{Regret}(\mathcal{K};w).$$

Main results: the regret

Theorem (Sabag, Goel, Lale, Hassibi 21)

The optimal regret for the strictly-causal scenario is given by

$$\text{Regret}^* = \bar{\sigma}(Z\Pi), \tag{2}$$

where Z *and* Π *are the unique solutions for the Lyapunov equations*

$$\begin{aligned} Z &= A_K Z A_K^* + B_u(R + B_u^* P B_u)^{-1} B_u^* \\ \Pi &= A_K^* \Pi A_K + P B_w B_w^* P. \end{aligned} \tag{3}$$

Main results: the regret

Theorem (Sabag, Goel, Lale, Hassibi 21)

The optimal regret for the strictly-causal scenario is given by

$$\text{Regret}^* = \bar{\sigma}(Z\Pi), \tag{2}$$

where Z *and* Π *are the unique solutions for the Lyapunov equations*

$$\begin{aligned} Z &= A_K Z A_K^* + B_u(R + B_u^* P B_u)^{-1} B_u^* \\ \Pi &= A_K^* \Pi A_K + P B_w B_w^* P. \end{aligned} \tag{3}$$

where P solves the LQR Riccati equation

$$\begin{aligned} P &= Q + A^*PA - A^*PB_u(R + B_u^*PB_u)^{-1}B_u^*PA \\ K_{lqr} &= (R + B_u^*PB_u)^{-1}B_u^*PA \\ A_K &= A - B_u K_{lqr} \end{aligned}$$

Main results: strictly-causal controller

Theorem (Sabag, Goel, Lale, Hassibi)

A strictly causal regret-optimal controller is given by

$$u_t = \hat{u}_t - K_{lqr} x_t, \quad (4)$$

where $\hat{u}_t$ is given by

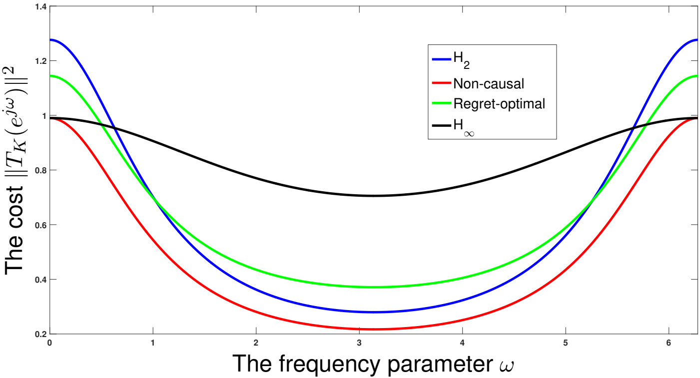
$$\begin{aligned} \xi_{t+1} &= F\xi_t + G w_t \\ \hat{u}_t &= -(R + B_u^* P B_u)^{-1} B_u^* \Pi \xi_t. \end{aligned} \quad (5)$$

and

$$\begin{aligned} G &= (I - A_K Z A_K^* \Pi)^{-1} A_K Z P B_w \\ F &= A_K - G B_w^* P, \end{aligned}$$

- Recall that $-K_{\text{lqr}} x_t$ is the standard LQR (H_2) controller

Example



Scalar system with $A = 0.9$ and $B_u = B_w = Q = R = 1$.

Comparison

	H_2 criterion (Frobenius)	H_∞ criterion (operator)
Noncausal	0.47	0.99
Regret-optimal	0.618	1.14
H_2 controller	0.598	1.28
H_∞ controller	0.84	0.99

Main ideas

- The regret can be reduced to a Nehari problem (1957)
  - Given an anticausal (upper triangular) operator $\mathcal{U}$,

$$\inf_{\mathcal{L} \text{ is causal}} \|\mathcal{L} - \mathcal{U}\|$$

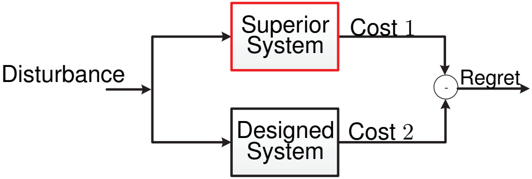
- Explicit regret and controller via frequency domain

Main ideas

- The regret can be reduced to a Nehari problem (1957)
  - Given an anticausal (upper triangular) operator $\mathcal{U}$,

$$\inf_{\mathcal{L} \text{ is causal}} \|\mathcal{L} - \mathcal{U}\|$$

- Explicit regret and controller via frequency domain
- The full-information control is just an example:

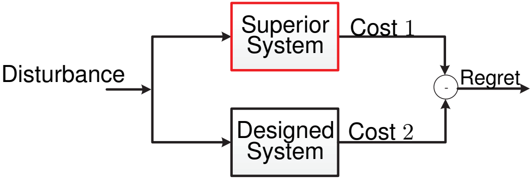


Main ideas

- The regret can be reduced to a Nehari problem (1957)
  - Given an anticausal (upper triangular) operator $\mathcal{U}$,

$$\inf_{\mathcal{L} \text{ is causal}} \|\mathcal{L} - \mathcal{U}\|$$

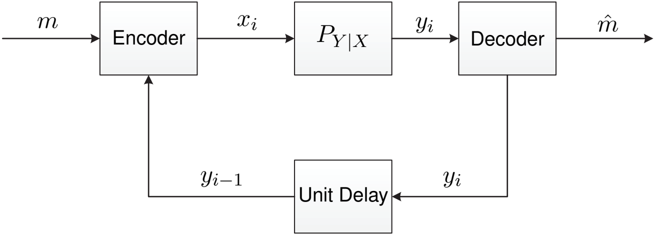
- Explicit regret and controller via frequency domain
- The full-information control is just an example:



  - The filtering problem (Kalman setting) in AISTATS 2021

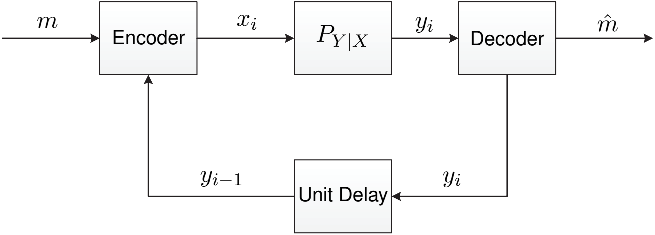
Part II: Feedback capacity of Gaussian channels

Channel with feedback



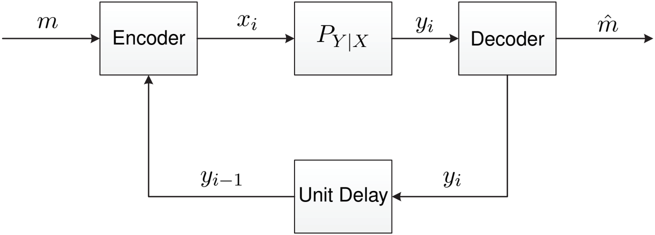
- A uniform message $m \in [1 : 2^{nR}]$
- At time i, encoding mapping is $e_i : [1 : 2^{nR}] \times \mathcal{Y}^{i-1}$
- Decoder mapping $\mathcal{Y}^n \to [1 : 2^{nR}]$

Channel with feedback



- A uniform message $m \in [1 : 2^{nR}]$
- At time i, encoding mapping is $e_i : [1 : 2^{nR}] \times \mathcal{Y}^{i-1}$
- Decoder mapping $\mathcal{Y}^n \to [1 : 2^{nR}]$
- Given a channel law, $P_{Y|X}$, the channel capacity is the maximal information rate R such that $\Pr(M \neq \hat{M}) \xrightarrow{n\to\infty} 0$

Channel with feedback



- A uniform message $m \in [1:2^{nR}]$
- At time i, encoding mapping is $e_i : [1:2^{nR}] \times \mathcal{Y}^{i-1}$
- Decoder mapping $\mathcal{Y}^n \to [1:2^{nR}]$
- Given a channel law, $P_{Y|X}$, the channel capacity is the maximal information rate R such that $\Pr(M \neq \hat{M}) \xrightarrow{n\to\infty} 0$
- *Feedback does not increase the capacity* (Shannon 56)

- But, feedback has other benefits...

The AWGN channel

- The channel is given by

$$y_i = x_i + z_i,$$

where $\{z_i\}_{i\geq 1}$ is a white process with $z_i \sim N(0, Z)$

- An average power constraint $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[x_i^2] \leq P$

The AWGN channel

- The channel is given by

$$y_i = x_i + z_i,$$

where $\{z_i\}_{i\geq 1}$ is a white process with $z_i \sim N(0, Z)$

- An average power constraint $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[x_i^2] \leq P$

1. Feedback does not increase the capacity

$$C_{fb}(P) = C(P) = \max I(X;Y) = 0.5 \log\left(1 + \frac{P}{Z}\right)$$

The AWGN channel

- The channel is given by

$$y_i = x_i + z_i,$$

where $\{z_i\}_{i\geq 1}$ is a white process with $z_i \sim N(0, Z)$

- An average power constraint $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[x_i^2] \leq P$

1. Feedback does not increase the capacity

$$C_{fb}(P) = C(P) = \max I(X;Y) = 0.5 \log\left(1 + \frac{P}{Z}\right)$$

2. Feedback improves the probability of error

- In part, the linear Schalkwijk-Kailath (1966) coding

$$x_i \propto (z_0 - \hat{z}_0(y^{i-1}))$$

achieves doubly-exponential decay (as n grows)

The additive Gaussian noise channel

- The channel is given by

$$Y_i = X_i + Z_i,$$

  where $\{z_i\}_{i\geq 1}$ is a colored Gaussian process

- Feedback capacity is defined similarly

The additive Gaussian noise channel

- The channel is given by

$$Y_i = X_i + Z_i,$$

where $\{z_i\}_{i\geq 1}$ is a colored Gaussian process

- Feedback capacity is defined similarly

- This is a channel with memory:

- The current noise Z_i is correlated with Z^{i-1}

- An optimal input should exploit this correlation via Z^{i-1}

- The optimal input distribution is not i.i.d.

The additive Gaussian noise channel

- The channel is given by

$$Y_i = X_i + Z_i,$$

  where $\{z_i\}_{i\geq 1}$ is a colored Gaussian process
  - Feedback capacity is defined similarly
- This is a channel with memory:
  - The current noise Z_i is correlated with Z^{i-1}
  - An optimal input should exploit this correlation via Z^{i-1}
- The optimal input distribution is not i.i.d.
- *Feedback can increase the channel capacity*
  - But, not too much (Pinsker 69) (Ebert 70) (Cover-Pombra 89)

The first works

- Motivated by the SK scheme, Butman (67,69,76) studied $\{Z_i\}$ an auto-regressive (AR) noise

$$Z_i = \sum_{i=1}^{k} \alpha_i Z_{i-k} + U_i, \tag{6}$$

where $U_i \sim N(0,1)$ is i.i.d.

The first works

- Motivated by the SK scheme, Butman (67,69,76) studied $\{Z_i\}$ an auto-regressive (AR) noise

$$Z_i = \sum_{i=1}^{k} \alpha_i Z_{i-k} + U_i, \tag{6}$$

  where $U_i \sim N(0,1)$ is i.i.d.

- Achievable rates using linear coding schemes
- Upper bounds on the feedback capacity of AR noise

- Schemes and bounds also in Tiernan and Schalkwijk (74,76)

General capacity expression

Theorem (Cover, Pombra 89)

The feedback capacity of Gaussian channels is

$$C_{fb}(P) = \lim_{n\to\infty} \frac{1}{2n} \max_{B,\Sigma_V} \log \frac{\det \Sigma_{X+Z}^{(n)}}{\det \Sigma_Z^{(n)}}, \quad (7)$$

where the n*th maximization is over*

$$X^n = BZ^n + V^n$$

with B *being a strictly causal operator,* V^n *is a Gaussian process and*

$$\frac{1}{n}\mathrm{Tr}(\Sigma_X^{(n)}) \leq P.$$

General capacity expression

Theorem (Cover, Pombra 89)

The feedback capacity of Gaussian channels is

$$C_{fb}(P) = \lim_{n\to\infty} \frac{1}{2n} \max_{B,\Sigma_V} \log \frac{\det \Sigma_{X+Z}^{(n)}}{\det \Sigma_Z^{(n)}}, \quad (7)$$

where the nth maximization is over

$$X^n = BZ^n + V^n$$

with B being a strictly causal operator, V^n is a Gaussian process and

$$\frac{1}{n}\mathrm{Tr}(\Sigma_X^{(n)}) \leq P.$$

- For a fixed n, it is a convex program (Ordentlich, Boyd 98)
- Non-trivial to compute the limit

Past literature - I

- A. Dembo, ”On Gaussian feedback capacity,” 1989
- S. Ihara, ”Capacity of discrete time Gaussian channel with and without feedback-I,” 1988
- S. Ihara, ”Capacity of mismatched Gaussian channels with and without feedback,” 1990
- E. Ordentlich, ”A class of optimal coding schemes for moving average additive Gaussian noise channels with feedback,” 1994
- L. H. Ozarow, ”Random coding for additive Gaussian channels with feedback,” 1990.
- L. H. Ozarow, ”Upper bounds on the capacity of Gaussian channels with feedback,” 1990
- J. Wolfowitz, ”Signalling over a Gaussian channel with feedback and autoregressive noise,” 1975.
- L. Vandenberghe, S. Boyd, and S.-P. Wu, ”Determinant maximization with linear matrix inequality constraints,” 1998

The control perspective

- Yang-Kavcic-Tatikonda (2007) derive an MDP formulation
  - The formulation holds for any n
  - The MDP state is a covariance matrix
- For first-order ARMA,

$$Z_i + \beta Z_{i-1} = U_i + \alpha U_{i-1}, \text{ with } U_i \sim N(0,1) \tag{8}$$

they demonstrated the lower bound

$$C_{fb}(P) \geq -\log x_0,$$

and conjectured it to be the feedback capacity where x_0 is the positive root of $\frac{Px^2}{1-x^2} = \frac{(1+\sigma\alpha x)^2}{(1+\sigma\beta x)^2}$ with $\sigma = \text{sign}(\beta - \alpha)$

The control perspective

- Yang-Kavcic-Tatikonda (2007) derive an MDP formulation
  - The formulation holds for any n
  - The MDP state is a covariance matrix
- For first-order ARMA,

$$Z_i + \beta Z_{i-1} = U_i + \alpha U_{i-1}, \text{ with } U_i \sim N(0,1) \tag{8}$$

  they demonstrated the lower bound

$$C_{fb}(P) \geq -\log x_0,$$

  and conjectured it to be the feedback capacity where x_0 is the positive root of $\frac{Px^2}{1-x^2} = \frac{(1+\sigma\alpha x)^2}{(1+\sigma\beta x)^2}$ with $\sigma = \text{sign}(\beta - \alpha)$
- Kim (2006) confirms their conjecture for $\beta = 0$

Variational formula

- Kim (2009) - variational formula for stationary noise:

$$C_{\mathrm{FB}} = \sup_{S_V, B} \int_{-\pi}^{\pi} \frac{1}{2} \log \frac{S_V(e^{i\theta}) + |1 + B(e^{i\theta})|^2 S_Z(e^{i\theta})}{S_Z(e^{i\theta})} \frac{d\theta}{2\pi} \tag{6}$$

where $S_Z(e^{i\theta})$ is the power spectral density of the noise process $\{Z_i\}_{i=1}^{\infty}$ and the supremum is taken over all power spectral densities $S_V(e^{i\theta}) \geq 0$ and all strictly causal filters $B(e^{i\theta}) = \sum_{k=1}^{\infty} b_k e^{ik\theta}$ satisfying the power constraint

$$\int_{-\pi}^{\pi} (S_V(e^{i\theta}) + |B(e^{i\theta})|^2 S_Z(e^{i\theta})) \frac{d\theta}{2\pi} \leq P.$$

Variational formula

- Kim (2009) - variational formula for stationary noise:

$$C_{\mathrm{FB}} = \sup_{S_V, B} \int_{-\pi}^{\pi} \frac{1}{2} \log \frac{S_V(e^{i\theta}) + |1 + B(e^{i\theta})|^2 S_Z(e^{i\theta})}{S_Z(e^{i\theta})} \frac{d\theta}{2\pi} \tag{6}$$

where $S_Z(e^{i\theta})$ is the power spectral density of the noise process $\{Z_i\}_{i=1}^{\infty}$ and the supremum is taken over all power spectral densities $S_V(e^{i\theta}) \geq 0$ and all strictly causal filters $B(e^{i\theta}) = \sum_{k=1}^{\infty} b_k e^{ik\theta}$ satisfying the power constraint

$$\int_{-\pi}^{\pi} (S_V(e^{i\theta}) + |B(e^{i\theta})|^2 S_Z(e^{i\theta})) \frac{d\theta}{2\pi} \leq P.$$

- Still, not computable
- Resembles entropy in robust control (Mustafa, Glover 90), (Doyle, Glover 88)
- Computation of optimal S_V, B for ARMA noise of first order
  - This confirms the conjecture in (Yang et al. 07)

Past literature - II

- C. Li and N. Elia, ”Youla coding and computation of Gaussian feedback capacity,” 2018
- T. Liu and G. Han, ”Feedback capacity of stationary Gaussian channels further examined,” 2019
- C. D. Charalambous, C. K. Kourtellaris and S. Loyka ”Capacity achieving distributions and separation principle for feedback Gaussian channels with memory: the LQG theory of directed information,” 2018
- A. Gattami, ”Feedback capacity of Gaussian channels revisited,” 2019
- C. D. Charalambous, C. K. Kourtellaris and S. Loyka, ”New formulas of ergodic feedback capacity of AGN channels driven by stable and unstable autoregressive noise,” 2020
- S. Fang and Q. Zhu, ”A connection between feedback capacity and Kalman filter for colored Gaussian noises,” 2020

Our setting

- The channel is MIMO

$$\mathbf{y}_i = \Lambda \mathbf{x}_i + \mathbf{z}_i,$$

where $\Lambda \in \mathbb{R}^{m \times p}$ is known.

Our setting

- The channel is MIMO

$$\mathbf{y}_i = \Lambda \mathbf{x}_i + \mathbf{z}_i,$$

  where $\Lambda \in \mathbb{R}^{m \times p}$ is known.

- The noise is generated by a state-space

$$\begin{aligned} \mathbf{s}_{i+1} &= F\mathbf{s}_i + G\mathbf{w}_i \\ \mathbf{z}_i &= H\mathbf{s}_i + \mathbf{v}_i, \end{aligned}$$

  where $(\mathbf{w}_i, \mathbf{v}_i) \sim N(0, \begin{pmatrix} W & L \\ L^T & V \end{pmatrix})$ is an i.i.d. sequence

- The initial state $s_1 \sim N(0, \Sigma_{1|0})$

Our setting

- The channel is MIMO

$$\mathbf{y}_i = \Lambda \mathbf{x}_i + \mathbf{z}_i,$$

  where $\Lambda \in \mathbb{R}^{m \times p}$ is known.
- The noise is generated by a state-space

$$\begin{aligned} \mathbf{s}_{i+1} &= F\mathbf{s}_i + G\mathbf{w}_i \\ \mathbf{z}_i &= H\mathbf{s}_i + \mathbf{v}_i, \end{aligned}$$

  where $(\mathbf{w}_i, \mathbf{v}_i) \sim N(0, \begin{pmatrix} W & L \\ L^T & V \end{pmatrix})$ is an i.i.d. sequence

- The initial state $s_1 \sim N(0, \Sigma_{1|0})$
- When F (and $L = 0$) is stable, it is the *stationary case*

Reminder: Kalman filter

- Define

$$\hat{\mathbf{s}}_i = \mathrm{E}[\mathbf{s}_i | \mathbf{z}^{i-1}]$$
$$\Sigma_i = \mathbf{cov}(\mathbf{s}_i - \hat{\mathbf{s}}_i).$$

- The Kalman filter is given by

$$\hat{\mathbf{s}}_{i+1} = F\,\hat{\mathbf{s}}_i + K_{p,i}(\mathbf{z}_i - H\,\hat{\mathbf{s}}_i), \qquad (9)$$

with

$$K_{p,i} = (F\Sigma_i H^T + GL)\Psi_i^{-1}, \quad \Psi_i = H\Sigma_i H^T + V,$$

and the covariance update is

$$\Sigma_{i+1} = F\Sigma_i F^T + GWG^T - K_p\Psi_i K_p^T. \qquad (10)$$

- The *innovations process* is $\mathbf{e}_i = \mathbf{z}_i - H\,\hat{\mathbf{s}}_i$ with $\mathbf{e}_i \sim N(0, \Psi_i)$

The Riccati equation

- The recursion converges to the stabilizing solution of

$$\Sigma = F\Sigma F^T + W - K_p \Psi K_p^T,$$

  where $K_p = (F\Sigma H^T + GL)\Psi^{-1}$ and $\Psi = H\Sigma H^T + V$.
- In the stationary case, no further assumptions
- In the non-stationary case, we assume detectability and stabilizability

Main result

Theorem (Sabag, Kostina, Hassibi 21)

The feedback capacity of the MIMO Gaussian channel is

$$C^{fb}(P) = \max_{\Pi,\hat{\Sigma},\Gamma} \frac{1}{2}\log\det(\Psi_Y) - \frac{1}{2}\log\det(\Psi)$$

$$\Psi_Y = \Lambda\Pi\Lambda^T + H\hat{\Sigma}H^T + \Lambda\Gamma H^T + H\Gamma^T\Lambda^T + \Psi$$

The channel:

$$\mathbf{y}_i = \Lambda\mathbf{x}_i + \mathbf{z}_i$$

The noise:

$$\mathbf{s}_{i+1} = F\mathbf{s}_i + G\mathbf{w}_i$$

$$\mathbf{z}_i = H\mathbf{s}_i + \mathbf{v}_i$$

Main result

Theorem (Sabag, Kostina, Hassibi 21)

The feedback capacity of the MIMO Gaussian channel is

$$C^{fb}(P) = \max_{\Pi,\hat{\Sigma},\Gamma} \frac{1}{2}\log\det(\Psi_Y) - \frac{1}{2}\log\det(\Psi)$$

$$\Psi_Y = \Lambda\Pi\Lambda^T + H\hat{\Sigma}H^T + \Lambda\Gamma H^T + H\Gamma^T\Lambda^T + \Psi$$

s.t. $\begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0, \quad \mathbf{Tr}(\Pi) \leq P,$

The channel:

$$\mathbf{y}_i = \Lambda\mathbf{x}_i + \mathbf{z}_i$$

The noise:

$$\mathbf{s}_{i+1} = F\mathbf{s}_i + G\mathbf{w}_i$$

$$\mathbf{z}_i = H\mathbf{s}_i + \mathbf{v}_i$$

Main result

Theorem (Sabag, Kostina, Hassibi 21)

The feedback capacity of the MIMO Gaussian channel is

$$C^{fb}(P) = \max_{\Pi,\hat{\Sigma},\Gamma} \frac{1}{2}\log\det(\Psi_Y) - \frac{1}{2}\log\det(\Psi)$$

$$\Psi_Y = \Lambda\Pi\Lambda^T + H\hat{\Sigma}H^T + \Lambda\Gamma H^T + H\Gamma^T\Lambda^T + \Psi$$

$$\textit{s.t.} \quad \begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0, \quad \mathbf{Tr}(\Pi) \le P,$$

$$\begin{pmatrix} F\hat{\Sigma}F^T + K_p\Psi K_p^T - \hat{\Sigma} & F\Gamma^T\Lambda^T + F\hat{\Sigma}H^T + K_p\Psi \\ (\cdot)^T & \Psi_Y \end{pmatrix} \succeq 0$$

The channel:

$$\mathbf{y}_i = \Lambda\mathbf{x}_i + \mathbf{z}_i$$

The noise:

$$\mathbf{s}_{i+1} = F\mathbf{s}_i + G\mathbf{w}_i$$

$$\mathbf{z}_i = H\mathbf{s}_i + \mathbf{v}_i$$

The linear matrix inequalities (LMIs)

- The decision variable Π is the inputs covariance:
- The constraint $\mathbf{Tr}(\Pi) \leq P$ is the power constraint
- The first LMI

$$\begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0$$

  is a verification that X_i forms a covariance matrix with a correlated signal

The linear matrix inequalities (LMIs)

- The decision variable Π is the inputs covariance:
  - The constraint $\mathbf{Tr}(\Pi) \leq P$ is the power constraint
  - The first LMI

$$\begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0$$

  is a verification that X_i forms a covariance matrix with a correlated signal

- The second LMI

$$\begin{pmatrix} F\hat{\Sigma}F^T + K_p\Psi K_p^T - \hat{\Sigma} & F\Gamma^T\Lambda^T + F\hat{\Sigma}H^T + K_p\Psi \\ (\cdot)^T & \Psi_Y \end{pmatrix} \succeq 0$$

  corresponds to a Riccati inequality

$$\hat{\Sigma} \preceq F\hat{\Sigma}F^T + K_p\Psi K_p^T - (F\Gamma^T\Lambda^T + F\hat{\Sigma}H^T + K_p\Psi)\Psi_Y^{-1}(F\Gamma^T\Lambda^T + F\hat{\Sigma}H^T + K_p\Psi)^T$$

Main results: a scalar channel

Theorem

The feedback capacity of the scalar Gaussian channel is

$$C^{fb}(P) = \max_{\hat{\Sigma},\Gamma} \frac{1}{2}\log\left(1 + \frac{P + H\hat{\Sigma}H^T + 2\Gamma H^T}{\Psi}\right)$$

s.t. $\begin{pmatrix} P & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0,$

$$\begin{pmatrix} F\hat{\Sigma}F^T + K_p\Psi K_p^T - \hat{\Sigma} & F\Gamma^T + F\hat{\Sigma}H^T + K_p\Psi \\ (F\Gamma^T + F\hat{\Sigma}H^T + K_p\Psi)^T & P + H\hat{\Sigma}H^T + 2\Gamma H^T + \Psi \end{pmatrix} \succeq 0,$$

where K_p *and* Ψ *are constants.*

- If $H = 0$, the capacity is $C(P) = \frac{1}{2}\log\left(1 + \frac{P}{V}\right)$.

Discussion

- This is the most general formulation with solution:
  1. General state-space
  2. Noise may be non-stationary
  3. MIMO channels
- The state-space structure is important
- The solution subsumes (Kim 06,09),

  and is *similar* to (Gattami 19) that studies a scalar channel with state-space that is stationary, controllable with fully-correlated disturbances

Can the capacity be simplified further?

The moving average noise

Consider $Z_i = U_i + \alpha U_{i-1}$ with $\alpha \in \mathbb{R}$ and $U_i \sim N(0,1)$

Theorem (Alternative expression for (Kim, 06))

The feedback capacity of first-order MA noise process is

$$C_{fb}(P) = \frac{1}{2}\log(1+\mathbf{SNR}), \tag{11}$$

where $\mathbf{SNR}$ *is the positive root of the polynomial* $\mathbf{SNR} = \left(\sqrt{P} + |\alpha|\sqrt{\frac{\mathbf{SNR}}{1+\mathbf{SNR}}}\right)^2$.

The moving average noise

Consider $Z_i = U_i + \alpha U_{i-1}$ with $\alpha \in \mathbb{R}$ and $U_i \sim N(0,1)$

Theorem (Alternative expression for (Kim, 06))

The feedback capacity of first-order MA noise process is

$$C_{fb}(P) = \frac{1}{2}\log(1+\mathbf{SNR}), \tag{11}$$

where $\mathbf{SNR}$ *is the positive root of the polynomial* $\mathbf{SNR} = \left(\sqrt{P} + |\alpha|\sqrt{\frac{\mathbf{SNR}}{1+\mathbf{SNR}}}\right)^2$.

- Proof: it is easy to show that the Schur complement of both LMIs equals zero. Substitute these equations into the objective.
- The fixed-point polynomial is different from (Kim 06)
  - However, their positive roots coincide

Main steps

Reminder: the capacity is given by

$$C_{fb}(P) = \lim_{n\to\infty} \frac{1}{2n} \max_{\{X_i = BZ^{i-1}+V_i\}_{i=1}^n} \log \frac{\det \Sigma_{X+Z}^{(n)}}{\det \Sigma_Z^{(n)}}$$

Road map:

1. *Sequentialize* the objective
2. *Sequentialize* the domain
3. Formulate a SCOP (sequential convex optimization problem)
4. A ”single-letter” upper bound
5. Show that the upper bound can be achieved

The directed information (DI)

1. The DI was defined in (Massey 90)

$$
\begin{aligned}
I(X^n \to Y^n) &= \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}) \\
&= \sum h(Y_i|Y^{i-1}) - h(X_i + Z_i | Y^{i-1}, X^i, Z^{i-1}) \\
&= \sum h(Y_i|Y^{i-1}) - h(Z_i|Z^{i-1}) \\
&= h(Y^n) - h(Z^n)
\end{aligned}
$$

2. For Gaussian inputs, the *Cover and Pombra objective* is DI

$$
I(X^n \to Y^n) = \log \frac{\det K_{X+Z}^{(n)}}{\det K_{Z}^{(n)}} \tag{12}
$$

3. Aligns with feedback capacity theorems (Tatikonda, Mitter 00,09) (Permuter, Weissman, Goldsmith 08)

The directed information (DI)

1. The DI was defined in (Massey 90)

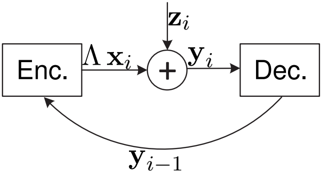
$$
\begin{aligned}
I(X^n \to Y^n) &= \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}) \\
&= \sum h(Y_i|Y^{i-1}) - h(X_i + Z_i | Y^{i-1}, X^i, Z^{i-1}) \\
&= \sum h(Y_i|Y^{i-1}) - h(Z_i|Z^{i-1}) \\
&= h(Y^n) - h(Z^n)
\end{aligned}
$$

2. For Gaussian inputs, the *Cover and Pombra objective* is DI

$$
I(X^n \to Y^n) = \log \frac{\det K_{X+Z}^{(n)}}{\det K_{Z}^{(n)}} \tag{12}
$$

3. Aligns with feedback capacity theorems (Tatikonda, Mitter 00,09) (Permuter, Weissman, Goldsmith 08)

The structure: cascaded filtering problem

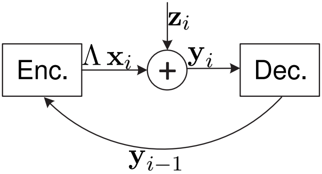


The encoder constructs
$\hat{\mathrm{s}}_i \triangleq \mathrm{E}[\mathbf{s}_i|\,\mathbf{z}^{i-1}]$ from

$$\mathbf{s}_{i+1} = F\mathbf{s}_i + G\mathbf{w}_i$$
$$\mathbf{z}_i = H\mathbf{s}_i + \mathbf{v}_i,$$

The innovation $\Psi_i = \text{cov}(\mathbf{z}_i - H\,\hat{\mathrm{s}}_i)$

The structure: cascaded filtering problem



The encoder constructs $\hat{\mathbf{s}}_i \triangleq \mathrm{E}[\mathbf{s}_i|\,\mathbf{z}^{i-1}]$ from

$$\mathbf{s}_{i+1} = F\mathbf{s}_i + G\mathbf{w}_i$$
$$\mathbf{z}_i = H\mathbf{s}_i + \mathbf{v}_i,$$

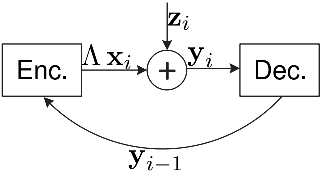
The innovation $\Psi_i = \text{cov}(\mathbf{z}_i - H\,\hat{\mathbf{s}}_i)$

The decoder constructs $\hat{\hat{\mathbf{s}}}_i \triangleq \mathrm{E}[\hat{\mathbf{s}}_i\,|\,\mathbf{y}^{i-1}]$from

$$\hat{\mathbf{s}}_{i+1} = F\,\hat{\mathbf{s}}_i + K_{p,i}\,\mathbf{e}_i,$$
$$\mathbf{y}_i = \mathbf{x}_i + H\,\hat{\mathbf{s}}_i + (\mathbf{z}_i - H\,\hat{\mathbf{s}}_i),$$

The innovation $\Psi_{Y,i} = \text{cov}(\mathbf{y}_i - \hat{\hat{\mathbf{y}}}_i)$

The structure: cascaded filtering problem



The encoder constructs $\hat{\mathbf{s}}_i \triangleq \mathrm{E}[\mathbf{s}_i|\,\mathbf{z}^{i-1}]$ from

$$\mathbf{s}_{i+1} = F\mathbf{s}_i + G\mathbf{w}_i$$
$$\mathbf{z}_i = H\mathbf{s}_i + \mathbf{v}_i,$$

The innovation $\Psi_i = \text{cov}(\mathbf{z}_i - H\,\hat{\mathbf{s}}_i)$

The decoder constructs $\hat{\hat{\mathbf{s}}}_i \triangleq \mathrm{E}[\hat{\mathbf{s}}_i\,|\,\mathbf{y}^{i-1}]$from

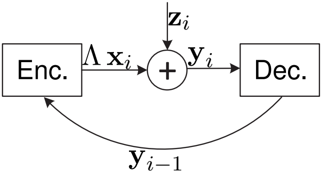
$$\hat{\mathbf{s}}_{i+1} = F\,\hat{\mathbf{s}}_i + K_{p,i}\,\mathbf{e}_i,$$
$$\mathbf{y}_i = \mathbf{x}_i + H\,\hat{\mathbf{s}}_i + (\mathbf{z}_i - H\,\hat{\mathbf{s}}_i),$$

The innovation $\Psi_{Y,i} = \text{cov}(\mathbf{y}_i - \hat{\hat{\mathbf{y}}}_i)$

- The objective reads

$$h(Y_i|Y^{i-1}) - h(Z_i|Z_{i-1}) = 0.5\log\det(\Psi_{Y,i}) - 0.5\log\det(\Psi_i)$$

The structure: cascaded filtering problem



The encoder constructs $\hat{\mathbf{s}}_i \triangleq \mathrm{E}[\mathbf{s}_i \,|\, \mathbf{z}^{i-1}]$ from

$$\mathbf{s}_{i+1} = F\mathbf{s}_i + G\mathbf{w}_i$$
$$\mathbf{z}_i = H\mathbf{s}_i + \mathbf{v}_i,$$

The innovation $\Psi_i = \mathrm{cov}(\mathbf{z}_i - H\,\hat{\mathbf{s}}_i)$

The decoder constructs $\hat{\hat{\mathbf{s}}}_i \triangleq \mathrm{E}[\hat{\mathbf{s}}_i \,|\, \mathbf{y}^{i-1}]$from

$$\hat{\mathbf{s}}_{i+1} = F\,\hat{\mathbf{s}}_i + K_{p,i}\,\mathbf{e}_i,$$
$$\mathbf{y}_i = \mathbf{x}_i + H\,\hat{\mathbf{s}}_i + (\mathbf{z}_i - H\,\hat{\mathbf{s}}_i),$$

The innovation $\Psi_{Y,i} = \mathrm{cov}(\mathbf{y}_i - \hat{\hat{\mathbf{y}}}_i)$

- The objective reads

$$h(Y_i|Y^{i-1}) - h(Z_i|Z_{i-1}) = 0.5\log\det(\Psi_{Y,i}) - 0.5\log\det(\Psi_i)$$

The optimal policy

Lemma

For each n, it is sufficient to optimize with inputs of the form

$$\mathbf{x}_i = \Gamma_i \hat{\Sigma}_i^{\dagger} (\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i) + \mathbf{m}_i, \quad i = 1, \ldots, n$$

where:

- Similar policy structures in (Yang et al. 07), (Kim 09), (Gattami 19), (Charalmbous et al. 18, 20)

The optimal policy

Lemma

For each n, it is sufficient to optimize with inputs of the form

$$\mathbf{x}_i = \Gamma_i \hat{\Sigma}_i^{\dagger}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i) + \mathbf{m}_i, \quad i = 1, \ldots, n$$

where:

- $\mathbf{m}_i \sim N(0, M_i)$ *is independent of* $(\mathbf{x}^{i-1}, \mathbf{y}^{i-1})$

- Similar policy structures in (Yang et al. 07), (Kim 09), (Gattami 19), (Charalmbous et al. 18, 20)

The optimal policy

Lemma

For each n, it is sufficient to optimize with inputs of the form

$$\mathbf{x}_i = \Gamma_i \hat{\Sigma}_i^{\dagger}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i) + \mathbf{m}_i, \quad i = 1, \ldots, n$$

where:

- $\mathbf{m}_i \sim N(0, M_i)$ *is independent of* $(\mathbf{x}^{i-1}, \mathbf{y}^{i-1})$
- $\hat{\Sigma}_i^{\dagger}$ *is the pseudo-inverse of* $\hat{\Sigma}_i = \mathbf{cov}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i)$

- Similar policy structures in (Yang et al. 07), (Kim 09), (Gattami 19), (Charalmbous et al. 18, 20)

The optimal policy

Lemma

For each n, it is sufficient to optimize with inputs of the form

$$\mathbf{x}_i = \Gamma_i \hat{\Sigma}_i^{\dagger} (\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i) + \mathbf{m}_i, \quad i = 1, \ldots, n$$

where:

- $\mathbf{m}_i \sim N(0, M_i)$ *is independent of* $(\mathbf{x}^{i-1}, \mathbf{y}^{i-1})$
- $\hat{\Sigma}_i^{\dagger}$ *is the pseudo-inverse of* $\hat{\Sigma}_i = \mathbf{cov}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i)$
- Γ_i *is a matrix that satisfies*

$$\Gamma_i (I - \hat{\Sigma}_i^{\dagger} \hat{\Sigma}_i) = 0$$

- Similar policy structures in (Yang et al. 07), (Kim 09), (Gattami 19), (Charalmbous et al. 18, 20)

The optimal policy

Lemma

For each n, it is sufficient to optimize with inputs of the form

$$\mathbf{x}_i = \Gamma_i \hat{\Sigma}_i^{\dagger}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i) + \mathbf{m}_i, \quad i = 1, \dots, n$$

where:

- $\mathbf{m}_i \sim N(0, M_i)$ *is independent of* $(\mathbf{x}^{i-1}, \mathbf{y}^{i-1})$
- $\hat{\Sigma}_i^{\dagger}$ *is the pseudo-inverse of* $\hat{\Sigma}_i = \mathbf{cov}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i)$
- Γ_i *is a matrix that satisfies*

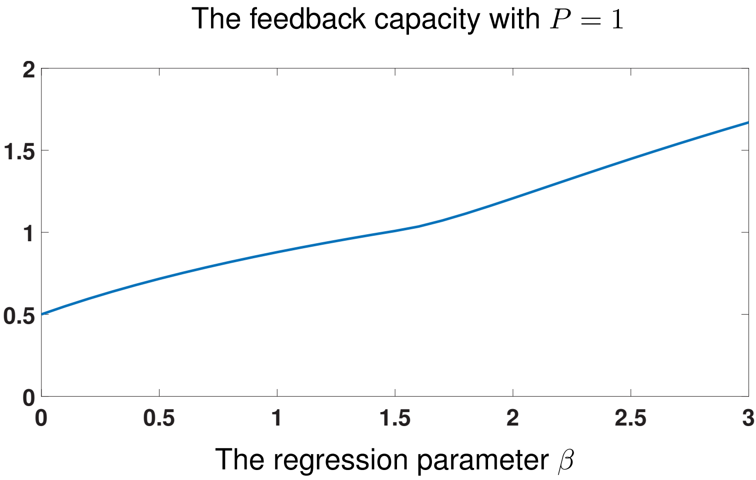
$$\Gamma_i(I - \hat{\Sigma}_i^{\dagger}\hat{\Sigma}_i) = 0$$

- *the input satisfies* $\sum_{i=1}^{n} \mathbf{Tr}(\Gamma_i \hat{\Sigma}_i^{\dagger} \Gamma_i^T + M_i) \leq nP$

- Similar policy structures in (Yang et al. 07), (Kim 09), (Gattami 19), (Charalmbous et al. 18, 20)

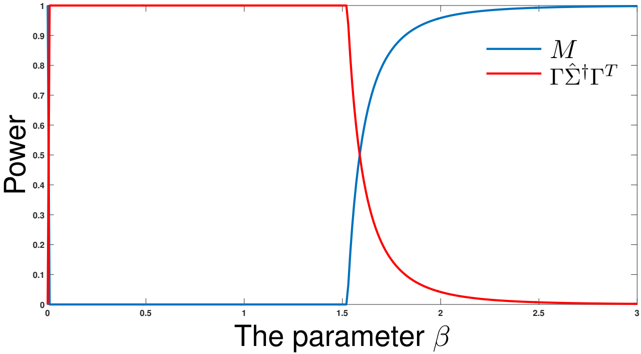
The AR noise

- Consider the AR noise $Z_i + \beta Z_{i-1} = U_i$ with $U_i \sim N(0,1)$



The AR noise - contd.

- The optimal inputs are $\mathbf{x}_i = \Gamma\hat{\Sigma}^{\dagger}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i) + \mathbf{m}_i$
  - The power of each component



- The range $\beta \in [0, 1.5]$ shows our disagreement with (Gattami 19)
- For large β, i.i.d. inputs become optimal

Main steps

Reminder: the capacity is given by

$$C_{fb}(P) = \lim_{n\to\infty} \frac{1}{2n} \max_{\{X_i = BZ^{i-1}+V_i\}_{i=1}^n} \log \frac{\det \Sigma_{X+Z}^{(n)}}{\det \Sigma_Z^{(n)}}$$

Road map:

- ✓ *Sequentialize* the objective
- ✓ *Sequentialize* the domain

3. Formulate a SCOP
4. A "single-letter" upper bound
5. Show that the upper bound can be achieved

The controlled state-space

Lemma

For a fixed policy $\{(\Gamma_i, M_i)\}_{i=1}^n$,

$$\hat{\mathbf{s}}_{i+1} = F\,\hat{\mathbf{s}}_i + K_{p,i}\,\mathbf{e}_i,$$
$$\mathbf{y}_i = (\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)\,\hat{\mathbf{s}}_i - \Lambda\Gamma_i\hat{\Sigma}_i^\dagger\,\hat{\hat{\mathbf{s}}}_i + \Lambda\mathbf{m}_i + \mathbf{e}_i,$$

Consequently, the error covariance $\hat{\Sigma}_i = \mathbf{cov}(\hat{\mathbf{s}}_i - \hat{\hat{\mathbf{s}}}_i)$ *satisfies*

$$\hat{\Sigma}_{i+1} = F\hat{\Sigma}_iF^T + K_{p,i}\Psi_iK_{p,i}^T - K_{Y,i}\Psi_{Y,i}K_{Y,i}^T$$

with $\hat{\Sigma}_1 = 0$, *and*

$$\Psi_{Y,i} = (\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)\hat{\Sigma}_i(\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)^T + \Lambda M_i\Lambda^T + \Psi_i$$
$$K_{Y,i} = (F\hat{\Sigma}_i(\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)^T + K_{p,i}\Psi_i)\Psi_{Y,i}^{-1}$$

- Similar state-space in (Kim 09), (Charalmbous et al. 20)

SCOP formulation

Lemma (Sequential convex-optimization problem)

The n-letter capacity can be bounded as

$$C_n(P) \leq \max_{\{\Gamma_i, \Pi_i, \hat{\Sigma}_{i+1}\}_{i=1}^n} \frac{1}{2n} \sum_{i=1}^{n} \log\det(\Psi_{Y,i}) - \log\det(\Psi_i)$$

$$s.t. \quad \begin{pmatrix} \Pi_t & \Gamma_t \\ \Gamma_t^T & \hat{\Sigma}_t \end{pmatrix} \succeq 0, \quad \frac{1}{n} \sum_{i=1}^{n} \mathbf{Tr}(\Pi_i) \leq P,$$

$$\begin{pmatrix} F\hat{\Sigma}_t F^T + K_{p,t}\Psi_t K_{p,t}^T - \hat{\Sigma}_{t+1} & K_{Y,t}\Psi_{Y,t} \\ \Psi_{Y,t}K_{Y,t}^T & \Psi_{Y,t} \end{pmatrix} \succeq 0,$$

where the LMIs hold for $t = 1, \ldots, n$ and $\hat{\Sigma}_1 = 0$.

Proof outline

- The argument of the objective is

$$\Psi_{Y,i} = (\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)\hat{\Sigma}_i(\Lambda\Gamma_i\hat{\Sigma}_i^\dagger + H)^T + \Lambda M_i\Lambda^T + \Psi_i$$

  - Define an auxiliary decision variable $\Pi_i \triangleq M_i + \Gamma_i\hat{\Sigma}_i^\dagger\Gamma^T$
  - Reduce the variable M_i
- The Schur complement transformation (e.g. Boyd 94)

$$\begin{matrix} \Pi_i \succeq \Gamma_i\hat{\Sigma}_i^\dagger\Gamma_i^T \\ \Gamma_i(I - \hat{\Sigma}_i^\dagger\hat{\Sigma}_i) = 0 \end{matrix} \iff \begin{pmatrix} \Pi_i & \Gamma_i \\ \Gamma_i^T & \hat{\Sigma}_i \end{pmatrix} \succeq 0.$$

- Relax Riccati recursion to a matrix inequality + Schur complement transformation

Single-letter Upper Bound

Lemma (The upper bound)

The feedback capacity is bounded by the convex optimization problem

$$C_{fb}(P) \leq \max_{\Pi,\hat{\Sigma},\Gamma} \frac{1}{2}\log\det(\Psi_Y) - \frac{1}{2}\log\det(\Psi)$$

$$\textit{s.t.} \quad \begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0, \quad \mathbf{Tr}(\Pi) \leq P,$$

$$\Psi_Y = \Lambda\Pi\Lambda^T + H\hat{\Sigma}H^T + \Lambda\Gamma H^T + H\Gamma^T\Lambda^T + \Psi$$

$$K_Y = (F\Gamma^T\Lambda^T + F\hat{\Sigma}H^T + K_p\Psi)\Psi_Y^{-1}$$

$$\begin{pmatrix} F\hat{\Sigma}F^T + K_p\Psi K_p^T - \hat{\Sigma} & K_Y\Psi_Y \\ \Psi_Y K_Y^T & \Psi_Y \end{pmatrix} \succeq 0.$$

Proof outline

- Define the uniform convex combinations

$$\bar{\Pi}_n = \frac{1}{n}\sum_{i=1}^{n}\Pi_i, \quad \bar{\Gamma}_n = \frac{1}{n}\sum_{i=1}^{n}\Gamma_i, \quad \bar{\hat{\Sigma}}_n = \frac{1}{n}\sum_{i=1}^{n}\hat{\Sigma}_i$$

- By the concavity of $\log\det(\cdot)$,

$$\frac{1}{n}\sum_{i=1}^{n}\log\det(\Psi_{Y,i}) \leq \log\det\left(\frac{1}{n}\sum_{i=1}^{n}\Psi_{Y,i}\right)$$

- Some of the constraints are satisfied for each n
- The Riccati LMI, however, is satisfied in the asymptotics only

Lower bound

Lemma (Lower bound)

The feedback capacity is lower bounded by the optimization problem

$$C_{fb}(P) \geq \max_{\Gamma,\Pi,\hat{\Sigma}} \log\det(\Psi_Y) - \log\det(\Psi)$$

$$\textit{s.t.} \quad \begin{pmatrix} \Pi & \Gamma \\ \Gamma^T & \hat{\Sigma} \end{pmatrix} \succeq 0, \quad \mathbf{Tr}(\Pi) \leq P$$

$$K_Y = (F\hat{\Sigma}H^T + F\Gamma^T\Lambda^T + K_p\Psi)\Psi_Y^{-1}$$

$$\Psi_Y = \Lambda\Pi\Lambda^T + \Lambda\Gamma H^T + H\Gamma^T\Lambda^T + \Psi$$

$$\hat{\Sigma} = F\hat{\Sigma}F^T + K_p\Psi K_p^T - K_Y\Psi_Y K_Y^T$$

$$\exists K : \rho(F - K(\Lambda\Gamma\hat{\Sigma}^\dagger + H)) < 1.$$

- Convergence of Riccati recursion (Nicolao, Gevers 92)

Conclusions

- A closed-form capacity expression as a finite-dimensional convex optimization problem
- The derivation relies on the noise state-space
- Sequential structures also exploited in (Tanaka, Kim, Parillo, Mitter 16) and its extension in (Sabag, Tian, Kostina, Hassibi 20)
- Ongoing work:
  - Optimal (and simple) coding scheme

Thank you for your attention!