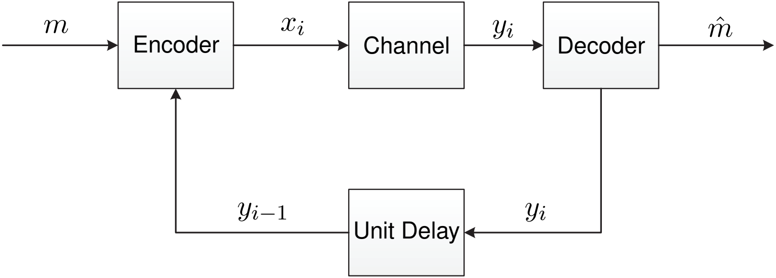
Feedback Capacity of MIMO Gaussian channels

Oron Sabag, Victoria Kostina and Babak Hassibi
Caltech

International Symposium on Information Theory
July 2021

Channel with feedback



- Encoder (at time i): $[1:2^{nR}]\times\mathcal{Y}^{i-1}\rightarrow\mathcal{X}$
- The channel is additive with Gaussian noise

$$y_i = x_i + z_i,$$

and an average power constraint $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[x_i^2]\leq P$

The Gaussian channel

- If the noise is i.i.d. (AWGN)

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

- If $\{z_i\}$ is not i.i.d.: this is a channel with memory:
  - The encoder knows z^{i-1} and can predict z_i
  - The optimal input distribution is not i.i.d.

1. *Feedback increases the channel capacity*
  - First works in (Butman 67,69,76) for AR noise

General capacity expression

Theorem (Cover, Pombra 89)

The feedback capacity of Gaussian channels is

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

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

$$\frac{1}{n}\operatorname{Tr}(\Sigma_X^{(n)}) \le P.$$

- For a fixed n, it is a convex program (Ordentlich, Boyd 94)
- 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
- 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
- A. Shahar-Doron, M. Feder ”On a capacity achieving scheme for the colored Gaussian channel with feedback,” 1994
- 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 approach

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

$$Z_i + \beta Z_{i-1} = U_i + \alpha U_{i-1}, \text{ with } U_i \sim N(0,1) \quad (2)$$

  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) proves their conjecture for $\beta = 0$
- Kim (2009) proves their conjecture for $|\beta| \leq 1, |\alpha| \leq 1$ via frequency domain formula of general stationary noise

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.

- The noise is generated by a state-space

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

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 $s_1 \sim N(0, \Sigma_{1|0})$

- If F is stable, it is the *stationary case* in (Kim 09)

Example: state space for ARMA(1)

- First-order ARMA noise

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

  can be represented as

$$\begin{aligned} S_{i+1} &= -\beta S_i + U_i \\ Z_i &= (\alpha - \beta) S_i + U_i, \end{aligned}$$

  - Can be verified via Z-transform $T(z) = 1 + (\alpha - \beta)(z + \beta)^{-1}$
- Similar representation for any ARMA process of order k
- The value of β determines the (asymptotic) stationarity

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 (time-invariant) Kalman filter is given by

$$\hat{\mathbf{s}}_{i+1} = F\,\hat{\mathbf{s}}_i + K_p(\mathbf{z}_i - H\,\hat{\mathbf{s}}_i), \tag{3}$$

where $K_p = (F\Sigma H^T + GL)\Psi^{-1}$ and $\Psi = H\Sigma H^T + V$.

- The error covariance is the solution to the Riccati equation

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

Main result

Theorem

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

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

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 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 $C(P) = \frac{1}{2}\log\left(1 + \frac{P}{V}\right)$.

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))

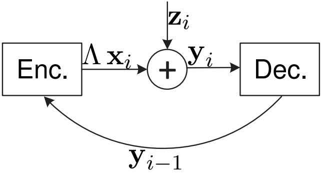
The feedback capacity of first-order MA noise process is

$$C_{fb}(P) = \frac{1}{2}\log(1+\mathbf{SNR}), \quad (4)$$

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: 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

Problem structure: cascaded filtering problem



	Encoder	Decoder
Information	$(\mathbf{x}^{i-1}, \mathbf{y}^{i-1}) \to \mathbf{z}^{i-1}$	$\mathbf{y}^{i-1}$
Estimation	$\hat{\mathbf{s}}_i \triangleq \mathrm{E}[\mathbf{s}_i \,\vert\, \mathbf{z}^{i-1}]$	$\hat{\hat{\mathbf{s}}}_i \triangleq \mathrm{E}[\hat{\mathbf{s}}_i \,\vert\, \mathbf{y}^{i-1}]$
State-space	$\mathbf{s}_{i+1} = F\mathbf{s}_i + G\mathbf{w}_i$ $\mathbf{z}_i = H\mathbf{s}_i + \mathbf{v}_i,$	$\hat{\mathbf{s}}_{i+1} = F\,\hat{\mathbf{s}}_i + K_{p,i}\,\mathbf{e}_i,$ $\mathbf{y}_i = \mathbf{x}_i + H\,\hat{\mathbf{s}}_i + (\mathbf{z}_i - H\,\hat{\mathbf{s}}_i),$
Innovation	$\Psi_i = \mathrm{cov}(\mathbf{z}_i - H\,\hat{\mathbf{s}}_i)$	$\Psi_{Y,i} = \mathrm{cov}(\mathbf{y}_i - H\,\hat{\hat{\mathbf{s}}}_i)$

Objective:

$$h(Y_i|Y^{i-1}) - h(Z_i|Z_{i-1}) = \frac{1}{2}(\log\det(\Psi_{Y,i}) - \log\det(\Psi_i))$$

The optimal policy

Lemma

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

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

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*

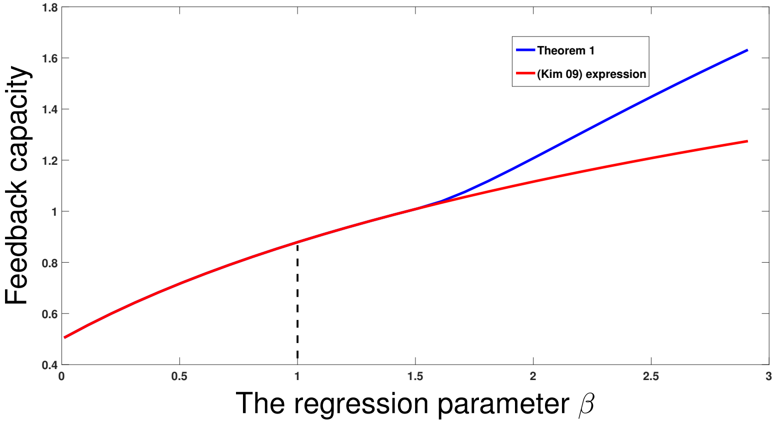
$$\Gamma_i(I - \hat{\Sigma}_i^{\dagger}\hat{\Sigma}_i) = 0$$

- *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., 20)

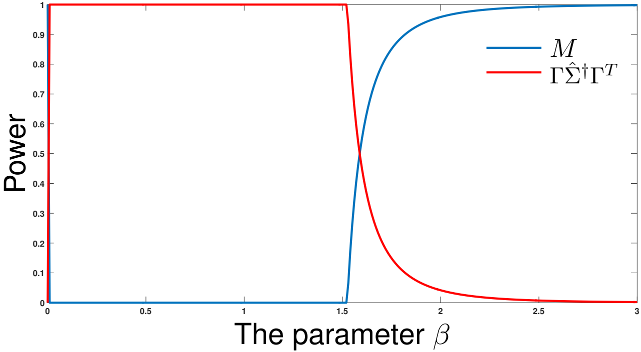
Auto regressive (AR) noise

- AR noise
$$Z_i + \beta Z_{i-1} = U_i, \quad U_i \sim N(0,1)$$
and power constraint $P = 1$



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 $\beta \in [0, 1]$ shows an error in (Gattami 19)
- For large β, i.i.d. inputs become optimal

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 $\Pi_i \triangleq M_i + \Gamma_i\hat{\Sigma}_i^{\dagger}\Gamma^T$
  - The objective $\Psi_{Y,i}$ is now a linear function
  - 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

Broader view - Directed information

The optimal directed information:

$$\begin{aligned} I(X^n \to Y^N) &= \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}) \\ &= \sum_{i=1}^{n} I(X^i, \hat{S}_i(X^{i-1}, Y^{i-1}); Y_i | Y^{i-1}) \\ &= \sum_{i=1}^{n} I(X_i, \hat{S}_i(X^{i-1}, Y^{i-1}); Y_i | Y^{i-1}) \\ &= \sum_{i=1}^{n} I(X_i, \hat{S}_i(X^{i-1}, Y^{i-1}); Y_i | \hat{\hat{S}}_i(Y^{i-1}), Y^{i-1}) \\ &\sim n I(X, \hat{S}; Y | \hat{\hat{S}}) \end{aligned}$$

- The variable $\hat{S}_i(X^{i-1}, Y^{i-1})$ serves as a state

Conclusions

- This is the most general formulation with computable solution:
  1. General state-space
  2. Noise may be non-stationary
  3. MIMO channels
- Sequential structures also exploited in (Tanaka, Kim, Parillo, Mitter 16), (Sabag, Tian, Kostina, Hassibi 20)
- Ongoing work:
  - Partial results on an optimal coding scheme (a la Schalkwijk-Kailath)

Thank you for your attention!