Sensor Scheduling

Jay Paek

Updated January 1, 2026

Contents

1 Nomenclature

For the equations below, let X, Y be random variables (r.v.s) with pdf/pmf $p(x), p(y)$ and supports $\mathcal{X} := \text{supp}\, X, \mathcal{Y} := \text{supp}\, Y$.

Notation	**Description**
$\mathcal{X}$	State space
$\mathcal{U}$	Control space
$q : \mathcal{X} \to [0, 1]$	pmf (belief) of the state estimate
$p(\cdot \mid s)$	Observation likelihood conditioned on state s

- Differential entropy of continuous r.v. X with pdf $p(\cdot)$:

$$H(X) := -\int_{\mathcal{X}} p(x) \log p(x)\, dx$$

- Discrete entropy of discrete r.v. X with pmf $p(\cdot)$:

$$H(X) := -\sum_{\mathcal{X}} p(x) \log p(x)$$

- Conditional entropy (continuous case):

$$H(Y \mid X) = -\iint_{\mathcal{X}\times\mathcal{Y}} p(x, y)\, \log p(y \mid x)\, dx\, dy$$

- Mutual information:

$$I(X; Y) = \iint_{\mathcal{X}\times\mathcal{Y}} p(x, y)\, \log\Big(\frac{p(x, y)}{p(x)p(y)}\Big)\, dx\, dy$$

2 Problem Formulation

Let the true robot state be $\mathbf{x} \in \mathcal{X}$. We represent our *estimate* of $\mathbf{x}$ as a random variable X with distribution q_t. Let the initial estimate be a uniform distribution $q_0(s) = 1/|\mathcal{X}|$ for all $s \in \mathcal{X}$ (if discrete) or uniform over the domain.

We have N sensors; sensor i returns an observation Z. For the toy problem, let the observation likelihood be Gaussian: $Z \mid \mathbf{x} \sim \mathcal{N}(\mathbf{x}, \sigma_i^2)$. Thus, the observation at time t using sensor i is modelled as:

$$Z_t := \mathbf{x} + N_i$$

where $\mathbf{x}$ is the true state and $N_i \sim \mathcal{N}(0, \sigma_i^2)$ is the noise from the ith sensor.

At each time t, we sample an observation and update the belief posterior via Bayes' rule:

$$f(q_t, z_t, u_t) := q_{t+1}(s) = \frac{p(z_t \mid s; u_t)\, q_t(s)}{p(z_t; u_t)},$$

which is the belief "motion model," where u_t selects a sensor. We formulate the finite-horizon optimal control problem:

$$\begin{aligned} \min_{\pi_{0:T-1}} \quad & \sum_{t=0}^{T-1} c(u_t) \;+\; \alpha\, H(q_T) \\ \text{s.t.} \quad & q_{t+1} = f(q_t, z_t, u_t), \\ & u_t = \pi_t(q_t). \end{aligned} \tag{1}$$

where the terminal cost $H(q_T)$ is the entropy of the final belief.

3 Commutativity of Observations

3.1 A Little Lemma

Lemma 1. *Let $\mathcal{V}$ be a vector space on a field F equipped with a linear functional $\phi : \mathcal{V} \to F$. Let L be the set of commutative linear maps on $\mathcal{V}$ such that $\phi(T\mathbf{v}) \neq 0$. Define the normalized composition $\boxtimes : L \times \mathcal{V} \to \mathcal{V}$ where $T \boxtimes \mathbf{v} \mapsto \dfrac{T\mathbf{v}}{\phi(T\mathbf{v})}$. Then,*

$$T_1 \boxtimes (T_2 \boxtimes \mathbf{v}) = T_2 \boxtimes (T_1 \boxtimes \mathbf{v})$$

Proof. By definition of the operator $\boxtimes$:

$$\begin{aligned} T_1 \boxtimes (T_2 \boxtimes \mathbf{v}) &= \frac{T_1 \frac{T_2\mathbf{v}}{\phi(T_2\mathbf{v})}}{\phi\left(T_1 \frac{T_2\mathbf{v}}{\phi(T_2\mathbf{v})}\right)} = \frac{\frac{1}{\phi(T_2\mathbf{v})} T_1 T_2 \mathbf{v}}{\frac{1}{\phi(T_2\mathbf{v})} \phi(T_1 T_2 \mathbf{v})} \\ &= \frac{T_1 T_2 \mathbf{v}}{\phi(T_1 T_2 \mathbf{v})} \end{aligned}$$

Since the operators T_1, T_2 commute, $T_1 T_2 = T_2 T_1$, the expression is symmetric in T_1 and T_2, proving the identity. □

3.2 Application to Bayesian Updates

The lemma shows that the posterior distribution dynamics commute as long as the selected sensor makes the same observation. This is because the Bayesian update is an instance of the normalized composition defined above, where the operator T_i is multiplication by the likelihood $p(z_i|s)$ and ϕ is the normalization constant (marginal likelihood).

Therefore, for two different processes with sensor-observation pairs $(u_1, \mathbf{z}_1)$ and $(u_2, \mathbf{z}_2)$, the final belief state is the same: $q_{t+2} = q'_{t+2}$. The policy space can thus be reduced to combinations of sensor uses, rather than permutations, which has size $\binom{T+N-1}{N-1}$.

4 Monotonicity

Proposition 1. ***(Expected posterior entropy is monotone in sensor variance)*** *Let X be any random variable, and let a measurement from sensor i be $Z_i = X + N_i$, where $N_i \sim \mathcal{N}(0, \sigma_i^2)$ is independent of X. Let q_t be the prior for X and q_{t+1} be the posterior of X given Z_i. Then*

$$\frac{d}{d\sigma_i^2} \mathbb{E}\big[H(q_{t+1})\big] \;\geq\; 0,$$

with equality iff X is almost surely constant.

Proof. Let σ_i^2 be denoted by τ. First, we show that the expected posterior entropy is the conditional entropy, i.e., $\mathbb{E}[H(q_{t+1})] = H(X \mid Z_i)$.

$$\begin{aligned} \mathbb{E}[H(q_{t+1})] &= -\mathbb{E}\Big[\int_{\mathcal{X}} \log(q_{t+1}(s \mid z_t)) q_{t+1}(s \mid z_t) ds\Big] \\ &= -\mathbb{E}\Big[\int_{\mathcal{X}} \log(q_{t+1}(s \mid z_t)) \frac{q_{t+1}(s, z_t)}{p(z_t)} ds\Big] \\ &= -\int_{\mathbb{R}} \Big[\int_{\mathcal{X}} \log(q_{t+1}(s \mid z_t)) \frac{q_{t+1}(s, z_t)}{p(z_t)} ds\Big] p(z_t) dz_t \\ &= -\int_{\mathbb{R}} \Big[\int_{\mathcal{X}} \log(q_{t+1}(s \mid z_t)) q_{t+1}(s, z_t) ds\Big] dz_t \\ &= H(X \mid Z_t) \end{aligned}$$

By the identity $H(X \mid Z_i) = H(X) - I(X;Z_i)$ [**?**], we have

$$\frac{d}{d\tau}\, H(X \mid Z_i) = -\frac{d}{d\tau} I(X;Z_i).$$

We know $I(X;Z_i) = H(Z_i) - H(Z_i \mid X) = H(X+N_i) - \frac{1}{2}\log(2\pi e\tau)$. By de Bruijn's identity [2],

$$\frac{d}{d\tau} H(X+N_i) \;=\; \tfrac{1}{2}\, J(X+N_i),$$

where $J(\cdot)$ is the Fisher information. Therefore

$$\frac{d}{d\tau} I(X;Z_i) \;=\; \tfrac{1}{2} J(X+N_i) \;-\; \frac{1}{2\tau}.$$

By Stam's inequality [2],

$$\frac{1}{J(X+N_i)} \geq \frac{1}{J(X)} + \frac{1}{J(N_i)} = \frac{1}{J(X)} + \tau \geq \tau$$

Taking reciprocals implies $J(X+N_i) \leq \frac{1}{\tau}$. Hence $\frac{d}{d\tau} I(X;Z_i) \leq 0$, which implies

$$\frac{d}{d\tau}\, \mathbb{E}\big[H(q_{t+1})\big] \;=\; -\frac{d}{d\tau} I(X;Z_i) \;\geq\; 0.$$

The continuous case follows similarly. □

5 Diminishing Return in Precision

Let X be any real-valued r.v. and for a precision parameter $\lambda = \sigma^{-2} \geq 0$, consider the AWGN observation model $Z_\lambda = \sqrt{\lambda}\, X + N$, where $N \sim \mathcal{N}(0,1)$ is independent of X. Let q_λ be the posterior of X given Z_λ. The expected posterior entropy is

$$\mathbb{E}\big[H(q_\lambda)\big] = H(X \mid Z_\lambda) = H(X) - I(X;Z_\lambda) \tag{2}$$

I–MMSE relation Define the minimum mean-squared error as $\mathrm{mmse}(\lambda) \;:=\; \mathbb{E}\big[(X - \mathbb{E}[X \mid Z_\lambda])^2\big]$. The I–MMSE identity [1] gives

$$\frac{d}{d\lambda}\, I(X;Z_\lambda) \;=\; \frac{1}{2}\, \mathrm{mmse}(\lambda), \quad \lambda \geq 0. \tag{3}$$

Moreover, $\mathrm{mmse}(\lambda)$ is nonincreasing in λ.

Theorem 1 (Monotonicity and convexity in precision)**.** *The map* $\lambda \mapsto \mathbb{E}[H(q_\lambda)]$ *is* monotone decreasing *and* convex *on* $[0,\infty)$.

Proof. Differentiate (2) with respect to λ and apply (3):

$$\frac{d}{d\lambda}\, \mathbb{E}[H(q_\lambda)] = -\frac{d}{d\lambda} I(X;Z_\lambda) = -\tfrac{1}{2}\, \mathrm{mmse}(\lambda) \leq 0,$$

which proves monotonicity. Differentiating again yields

$$\frac{d^2}{d\lambda^2}\, \mathbb{E}[H(q_\lambda)] = -\tfrac{1}{2}\, \frac{d}{d\lambda}\, \mathrm{mmse}(\lambda) \geq 0,$$

which proves convexity, as $\mathrm{mmse}(\lambda)$ is nonincreasing. □

This implies that the incremental entropy reduction per unit increase in precision exhibits diminishing marginal returns.

6 Information Additivity

Proposition 2 (Information depends only on total precision)**.** *Assume the AWGN model with independent measurements $Z_t = X + N_t$, where $N_t \sim \mathcal{N}(0, \sigma_t^2)$. Define precisions $\lambda_t := \sigma_t^{-2}$ and total precision $\Lambda := \sum_{t=0}^{T-1} \lambda_t$. Then*

$$I(X; Z_{0:T-1}) \;=\; I(X; \sqrt{\Lambda}\, X + N_0),$$

where $N_0 \sim \mathcal{N}(0,1)$ is independent of X. Hence

$$\mathbb{E}[H(q_T)] \;=\; H(q_0) - I(\Lambda).$$

Proof. The set of observations $\{Z_t\}_{t=0}^{T-1}$ can be shown to be equivalent to a single observation from a channel with precision equal to the sum of the individual precisions, Λ.

Stack $Z = (Z_1, \dots, Z_T)^\top$ and let $D = \mathrm{diag}(\sigma_1, \dots, \sigma_T)$. Define $Y := D^{-1} Z = v\, X + W$, where $v = (1/\sigma_1, \dots, 1/\sigma_T)^\top$ and $W \sim \mathcal{N}(0, I_T)$. Choose an orthogonal U with $Uv = \|v\| e_1$, where $\|v\| = \sqrt{\sum_t \lambda_t} = \sqrt{\Lambda}$. Then $\tilde{Y} := UY = \|v\| X\, e_1 + \tilde{W}$, with $\tilde{W} \sim \mathcal{N}(0, I_T)$. Thus $\tilde{Y}_2, \dots, \tilde{Y}_T$ are pure noise independent of X, so

$$I(X; Z) = I(X; Y) = I(X; \tilde{Y}) = I\big(X; \tilde{Y}_1\big) = I\big(X; \sqrt{\Lambda}\, X + N_0\big)$$

because mutual information is invariant under bijective transforms □

7 A Greedy Algorithm

We formulate the sensor selection as an optimal stopping problem. Let the objective function for a set of selected sensors S with total precision $\Lambda_S = \sum_{i \in S} \lambda_i$ and total cost $C_S = \sum_{i \in S} c_i$ be:

$$J(S) \;=\; \alpha\, I(\Lambda_S) \;-\; C_S.$$

We aim to find the subset $S^\star$ that maximizes $J(S)$.

Proposition 3 (Optimality of Greedy Selection)**.** *Let the marginal utility of sensor k given current precision Λ be $\Delta_k(\Lambda) = \alpha[I(\Lambda + \lambda_k) - I(\Lambda)] - c_k$. If the greedy algorithm selects sensors recursively such that $u_t = \arg\max_k \Delta_k(\Lambda_{t-1})$ and stops when $\max_k \Delta_k(\Lambda_t) \leq 0$, it yields the globally optimal set $S^\star$.*

Proof. From Theorem 1, $H(q_\lambda)$ is convex in λ, implying the mutual information $I(\lambda) = H(q_0) - \mathbb{E}[H(q_\lambda)]$ is concave and its discrete derivative (marginal gain) is monotonically decreasing. Consider the function $f(\Lambda) = \alpha I(\Lambda) - c_{eff} \Lambda$, where c_{eff} represents the cost per unit precision. Due to the concavity of I, $f(\Lambda)$ is unimodal. The greedy strategy in Algorithm 1 is equivalent to a coordinate ascent on $J(S)$.

1. **Commutativity:** From Lemma 1, the value of a set S depends only on the sum of precisions, not the order of selection. Thus, we cannot get ”stuck” in a suboptimal path due to ordering.

2. **Diminishing Returns:** Theorem 1 ensures that if a sensor k is not profitable to add at precision Λ (i.e., $\Delta_k(\Lambda) < 0$), it will never be profitable at any $\Lambda' > \Lambda$, as the marginal gain strictly decreases.

Therefore, the stopping condition $\Delta_k \leq 0$ is sufficient to guarantee that no remaining sensor can improve the objective. The greedy choice of the *maximum* Δ_k is not strictly required for optimality (any positive step suffices), but it maximizes the rate of convergence (steepest ascent). □

The greedy selection policy is detailed in Algorithm 1. Note that while Eq. (2) specifies a fixed horizon T, this algorithm implements the optimal stopping condition derived above.

The mutual information $I(\Lambda)$ for a discrete prior can be approximated using Gauss-Hermite quadrature. The resulting greedy selection policy is detailed in Algorithm 1. Note that while Eq. (2) specifies a fixed horizon T, the algorithm below implements a stopping condition based on nonnegative marginal gain, effectively optimizing the horizon length implicitly.

Algorithm 1 Greedy Sensor Selection with α-Tradeoff

Input: Precisions $(\lambda_i)_{i=1}^m$, costs $(c_i)_{i=1}^m$, horizon T, tradeoff $\alpha > 0$, oracle $I(\cdot)$, tolerance $\varepsilon \geq 0$.
Initialize: $\Lambda \leftarrow 0; \quad n_i \leftarrow 0$ for all $i; \quad C \leftarrow 0$.
for $t = 1$ **to** T **do**
  $b \leftarrow -\infty; \quad j^\star \leftarrow \varnothing$.
  for $i = 1$ **to** m **do**
    $\Delta I \leftarrow I(\Lambda + \lambda_i) - I(\Lambda)$
    score $\leftarrow \Delta I - c_i/\alpha$
    if score $> b$ **then**
      $b \leftarrow$ score; $\quad j^\star \leftarrow i$
    end if
  end for
  if $b \leq \varepsilon$ **then**
    break {Stop if marginal cost exceeds information gain}
  end if
  $n_{j^\star} \leftarrow n_{j^\star} + 1; \quad \Lambda \leftarrow \Lambda + \lambda_{j^\star}; \quad C \leftarrow C + c_{j^\star}$.
end for
Output: Counts $(n_i)_{i=1}^m$, total precision Λ, total cost C, term. entropy $H(q_T) = H(q_0) - I(\Lambda)$.

7.1 Optimality Guarantee

While the greedy strategy is optimal for the continuous relaxation of the problem (where fractional sensor usage is allowed), the restriction to integer sensor counts introduces a discretization error. We bound this error below.

Theorem 2 (Integrality Gap Bound)**.** *Let $J(\mathbf{n})$ be the net utility for a configuration of integer sensor counts $\mathbf{n} \in \mathbb{Z}^m_{\geq 0}$. Let $J^\star_{cont}$ be the maximum utility achievable if sensor counts were continuous variables $\mathbf{n} \in \mathbb{R}^m_{\geq 0}$. The utility J_{greedy} returned by Algorithm 1 satisfies:*

$$J^\star_{cont} - J_{greedy} \;\leq\; \max_k \left(\Delta J_k^+\right),$$

where ΔJ_k^+ is the maximum possible single-step net gain of sensor k.

Proof. Let $g(\Lambda) = \alpha I(\Lambda) - C(\Lambda)$ be the objective function projected onto the total precision axis Λ. Since $I(\cdot)$ is concave and cost is linear, $g(\Lambda)$ is strictly unimodal (concave). The greedy algorithm effectively climbs this discrete concave surface. The algorithm halts at a precision Λ_{stop} when the marginal gain of all available sensors becomes non-positive. In the worst-case scenario, the true continuous maximum $\Lambda^\star$ lies between Λ_{stop} and $\Lambda_{stop} + \lambda_{best}$, where λ_{best} is the precision of the sensor that would have been chosen next. Due to the concavity of g, the drop in utility from the continuous peak to the discrete stopping point is bounded by the vertical rise of the step that "overshoots" the peak. Thus, the regret is bounded by the maximum single atomic step size in the utility space. □

Remark. This implies that as the sensor precisions $\lambda_i \to 0$ (high-rate or low-noise limit), the gap vanishes and the greedy solution converges to the true global optimum.

8 Comparison Study

Recall the following

Predictive (evidence) density. Given belief q and sensor i, the predictive density is the Gaussian mixture

$$p(z \mid q, i) \;:=\; \sum_{s=1}^{M} q(s)\, p(z \mid s; i) \;=\; \sum_{s=1}^{M} q(s)\, \mathcal{N}(z;\, x_s, \sigma_i^2).$$

Belief update (Bayes filter). After selecting sensor i and observing z, the belief updates to $q' = f(q, z, i)$:

$$q'(s) \;=\; f(q,z,i)(s) \;:=\; \frac{p(z \mid s; i)\, q(s)}{\sum_{r=1}^{M} p(z \mid r; i)\, q(r)} \;=\; \frac{\mathcal{N}(z;\, x_s, \sigma_i^2)\, q(s)}{\sum_{r=1}^{M} \mathcal{N}(z;\, x_r, \sigma_i^2)\, q(r)}. \tag{4}$$

Use Shannon entropy for a discrete belief:

$$H(q) \;:=\; -\sum_{s=1}^{M} q(s) \log q(s).$$

Finite-horizon control problem (belief-MDP). For horizon T, find a policy sequence $\pi_{0:T-1}$ with $u_t = \pi_t(q_t)$ minimizing

$$\min_{\pi_{0:T-1}} \; \mathbb{E}\left[\sum_{t=0}^{T-1} c(u_t) \;+\; \alpha\, H(q_T)\right] \quad \text{s.t.} \quad q_{t+1} = f(q_t, Z_t, u_t), \;\; u_t = \pi_t(q_t). \tag{5}$$

Bellman value function. Define the cost-to-go from time t with belief q:

$$V_t(q) \;:=\; \inf_{\pi_{t:T-1}} \mathbb{E}\left[\sum_{k=t}^{T-1} c(u_k) \;+\; \alpha\, H(q_T) \;\middle|\; q_t = q\right].$$

The terminal condition is

$$V_T(q) = \alpha\, H(q).$$

For $t = T-1, T-2, \ldots, 0$, the Bellman equation is

$$V_t(q) \;=\; \min_{i \in \{1,\ldots,K\}} \left\{ c(i) \;+\; \int_{\mathbb{R}} V_{t+1}\big(f(q,z,i)\big)\; p(z \mid q, i)\; dz \right\}. \tag{6}$$

Optimal policy. An optimal policy at time t is any minimizer of (6):

$$\pi_t^{\star}(q) \in \arg \min_{i \in \{1,\ldots,K\}} \left\{ c(i) \;+\; \int_{\mathbb{R}} V_{t+1}\big(f(q,z,i)\big)\; p(z \mid q, i)\; dz \right\}. \tag{7}$$

One can notice that the greedy algorithm is essentially the Bellman Dynamic Programming policy but with look-ahead depth 1.

9 Results

We test the greedy algorithm with a horizon of $T = 20$ and a cost-entropy tradeoff of $\alpha = 1$. The results show the difference in cost and sub-costs across different algorithms

It can be seen that the greedy algorithm consistently outperforms the other algorithms.

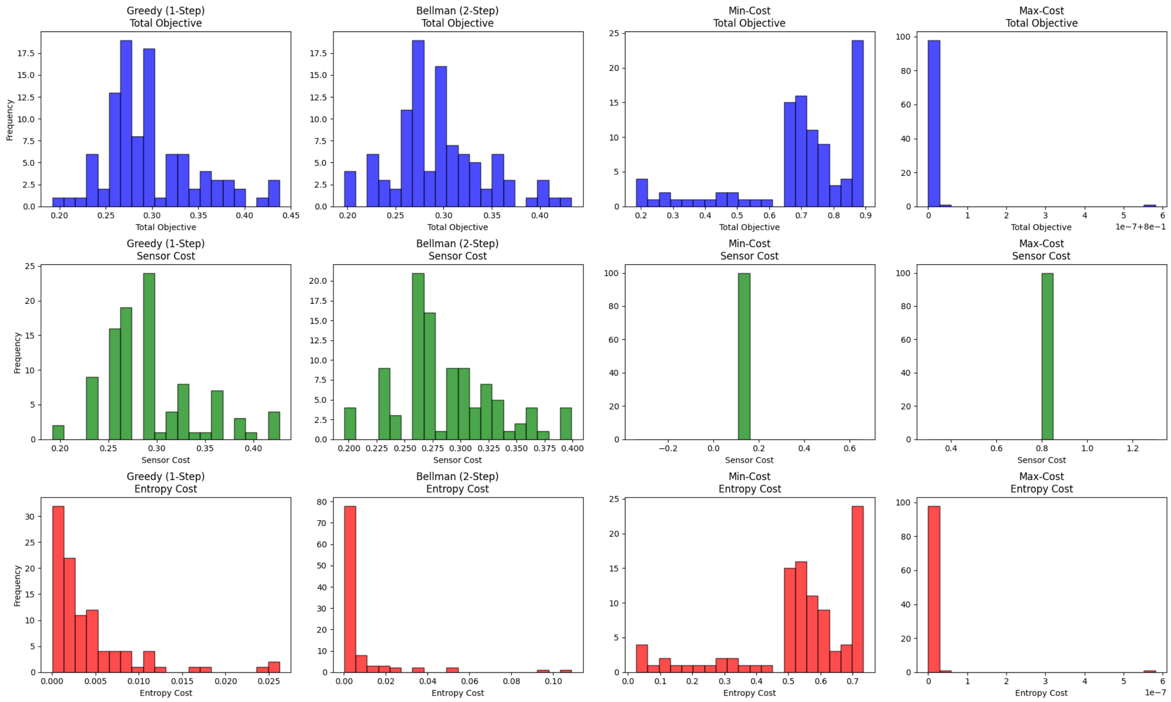

Figure 1: 4 sensors with costs [0.04, 0.01, 0.009, 0.008] and noise variances [0.5, 1.0, 1.5, 2.0]. The policy balances high precision with low cost.

10 Extension to General Observation Models

We now generalize the observation model to $Z = h(X) + N$, where h is a continuous function. The core result holds under the mild condition that $h(X)$ is not almost surely constant.

Let X be a real-valued random variable, and let N_1, N_2 be independent Gaussian noises

$$N_i \sim \mathcal{N}(0, \sigma_i^2), \quad i = 1, 2,$$

independent of X, with

$$0 < \sigma_1^2 < \sigma_2^2 < \infty.$$

Let $h : \mathbb{R} \to \mathbb{R}$ be any measurable function and define

$$Z_i := h(X) + N_i, \quad i = 1, 2.$$

We show that, except for the degenerate case where $h(X)$ is almost surely constant,

$$I(X; Z_1) > I(X; Z_2).$$

Proof. Since $\sigma_2^2 > \sigma_1^2$, we can represent the channel $X \to Z_2$ as a degraded version of $X \to Z_1$. Let $W \sim \mathcal{N}(0, \sigma_2^2 - \sigma_1^2)$ be independent of Z_1 and X. Construct a random variable Z_2' as:

$$Z_2' = Z_1 + W = h(X) + N_1 + W.$$

Since N_1 and W are independent Gaussians, their sum is Gaussian with variance $\sigma_1^2 + (\sigma_2^2 - \sigma_1^2) = \sigma_2^2$. Thus, the conditional distribution of Z_2' given X is identical to that of Z_2 given X (i.e., $Z_2' \overset{d}{=} Z_2 \mid X$). Consequently, $I(X; Z_2) = I(X; Z_2')$.

Observe that $X \to Z_1 \to Z_2'$ forms a Markov Chain. That is, given Z_1, Z_2' is independent of X because W is independent of X. By the Data Processing Inequality (DPI), we have:

$$I(X; Z_2') \le I(X; Z_1).$$

Equality in the DPI holds if and only if $X \to Z_2' \to Z_1$ also forms a Markov Chain (i.e., X and Z_1 are conditionally independent given Z_2'). In the additive Gaussian noise setting, this reverse Markov condition implies that Z_2' contains all the information Z_1 has about X. Since Z_2' is a strictly noisy version of Z_1 (added variance $\sigma_2^2 - \sigma_1^2 > 0$), this is impossible unless X (and thus $h(X)$) is deterministic (almost surely constant). For any non-degenerate $h(X)$, the inequality is strict:

$$I(X; Z_2) < I(X; Z_1).$$

□

Alternative Derivation via de Bruijn's Identity

We can alternatively prove the monotonicity of mutual information by examining its derivative with respect to the noise variance. Let $Z_\gamma = h(X) + N_\gamma$, where $N_\gamma \sim \mathcal{N}(0, \gamma)$. We aim to show that $\frac{d}{d\gamma} I(X; Z_\gamma) < 0$.

Recall that the mutual information for an additive noise channel is given by:

$$I(X; Z_\gamma) = h(Z_\gamma) - h(Z_\gamma \mid X) = h(Z_\gamma) - h(N_\gamma).$$

Substituting the entropy of the Gaussian noise $h(N_\gamma) = \frac{1}{2} \log(2\pi e \gamma)$, we differentiate with respect to γ:

$$\frac{d}{d\gamma} I(X; Z_\gamma) = \frac{d}{d\gamma} h(Z_\gamma) - \frac{d}{d\gamma} \left(\frac{1}{2} \log(2\pi e \gamma) \right).$$

The second term is simply $\frac{1}{2\gamma}$. For the first term, we invoke de Bruijn's identity, which relates the derivative of differential entropy with respect to noise variance to the Fisher information $J(\cdot)$:

$$\frac{d}{d\gamma} h(Z_\gamma) = \frac{1}{2} J(Z_\gamma).$$

Thus, the derivative of the mutual information becomes:

$$\frac{d}{d\gamma} I(X; Z_\gamma) = \frac{1}{2} J(Z_\gamma) - \frac{1}{2\gamma} = \frac{1}{2} \left(J(Z_\gamma) - J(N_\gamma) \right),$$

where we used the fact that the Fisher information of a Gaussian variable N_γ is $J(N_\gamma) = \frac{1}{\gamma}$.

To determine the sign of this derivative, we apply **Stam's inequality** (specifically the convolution inequality for Fisher information). For independent random variables A and B, the Fisher information of their sum satisfies:

$$\frac{1}{J(A+B)} \geq \frac{1}{J(A)} + \frac{1}{J(B)}.$$

Setting $A = h(X)$ and $B = N_\gamma$, we have:

$$\frac{1}{J(Z_\gamma)} \geq \frac{1}{J(h(X))} + \frac{1}{J(N_\gamma)} = \frac{1}{J(h(X))} + \gamma.$$

Assuming $h(X)$ is non-degenerate (i.e., not a constant), its Fisher information $J(h(X))$ is positive, implying $\frac{1}{J(h(X))} > 0$. Therefore:

$$\frac{1}{J(Z_\gamma)} > \gamma \implies J(Z_\gamma) < \frac{1}{\gamma} = J(N_\gamma).$$

Substituting this back into our derivative:

$$\frac{d}{d\gamma} I(X; Z_\gamma) < 0.$$

This proves that $I(X; Z_\gamma)$ is strictly decreasing with respect to the noise variance γ.

References

[1] Dongning Guo, Shlomo Shamai, and Sergio Verdu. Mutual information and minimum mean-square error in gaussian channels, 2004.

[2] A.J. Stam. Some inequalities satisfied by the quantities of information of fisher and shannon. *Information and Control*, 2(2):101–112, 1959.