Mismatched Quantum Filtering and Entropic Information

[arXiv:1310.0291]

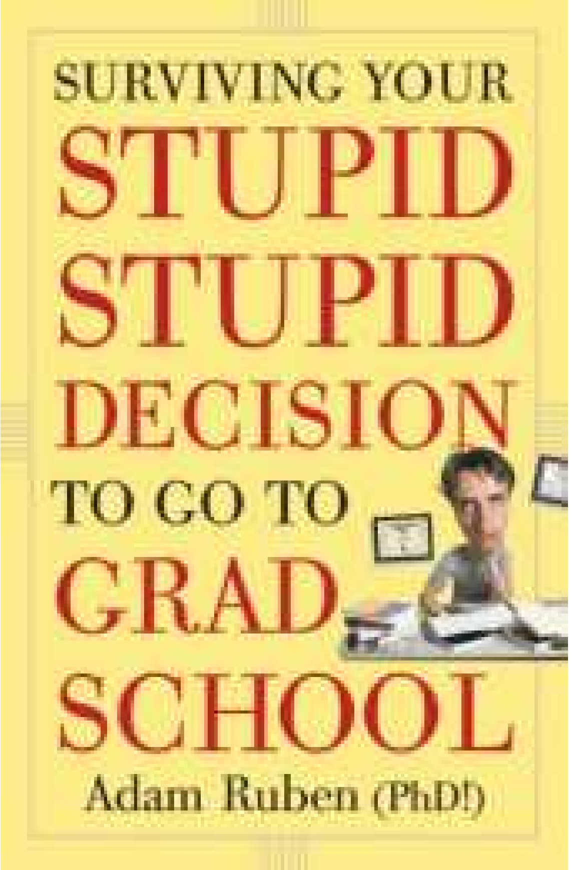
Mankei Tsang
Department of Electrical and Computer Engineering
Department of Physics
National University of Singapore

eletmk@nus.edu.sg
http://mankei.tsang.googlepages.com/

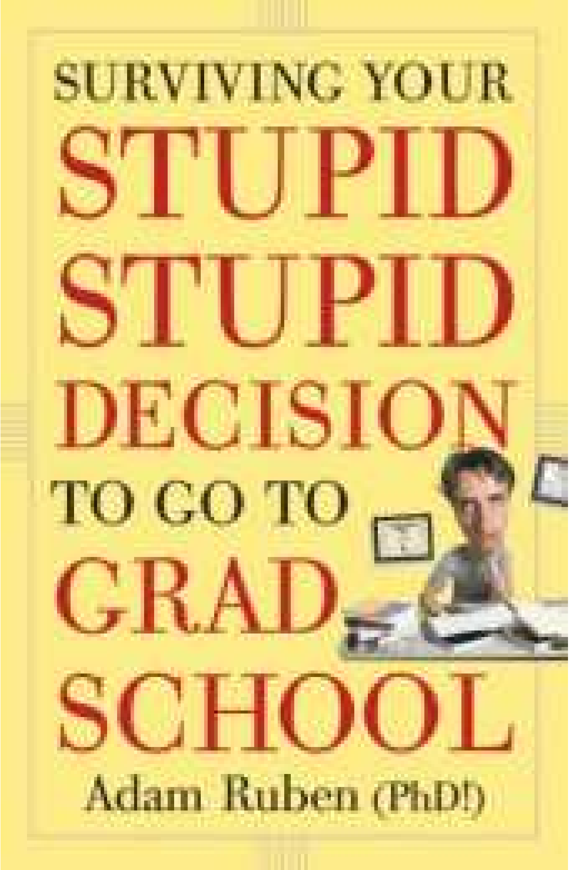
Feb 21, 2014

Regret

Regret

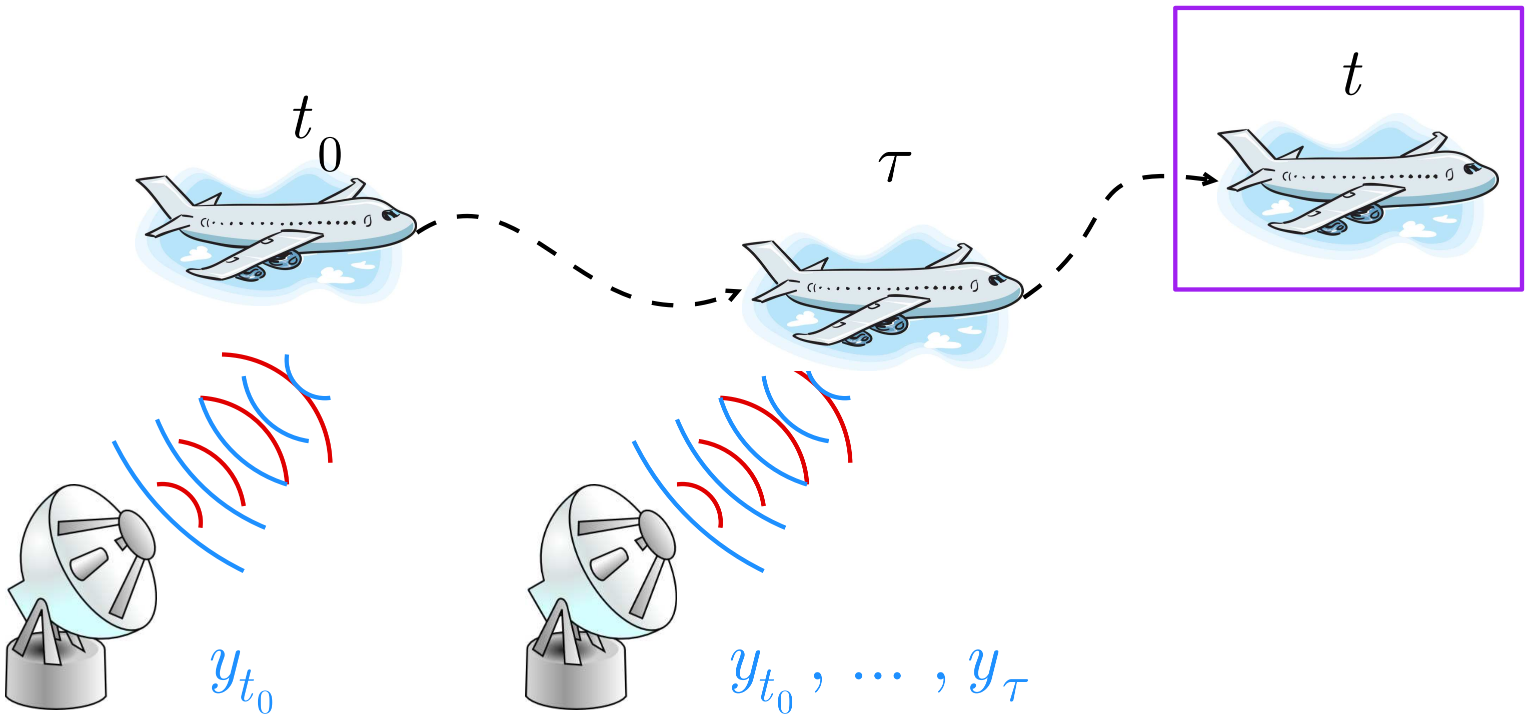

SURVIVING YOUR
STUPID
STUPID
DECISION
TO GO TO
GRAD
SCHOOL
Adam Ruben (PhD!)

Regret

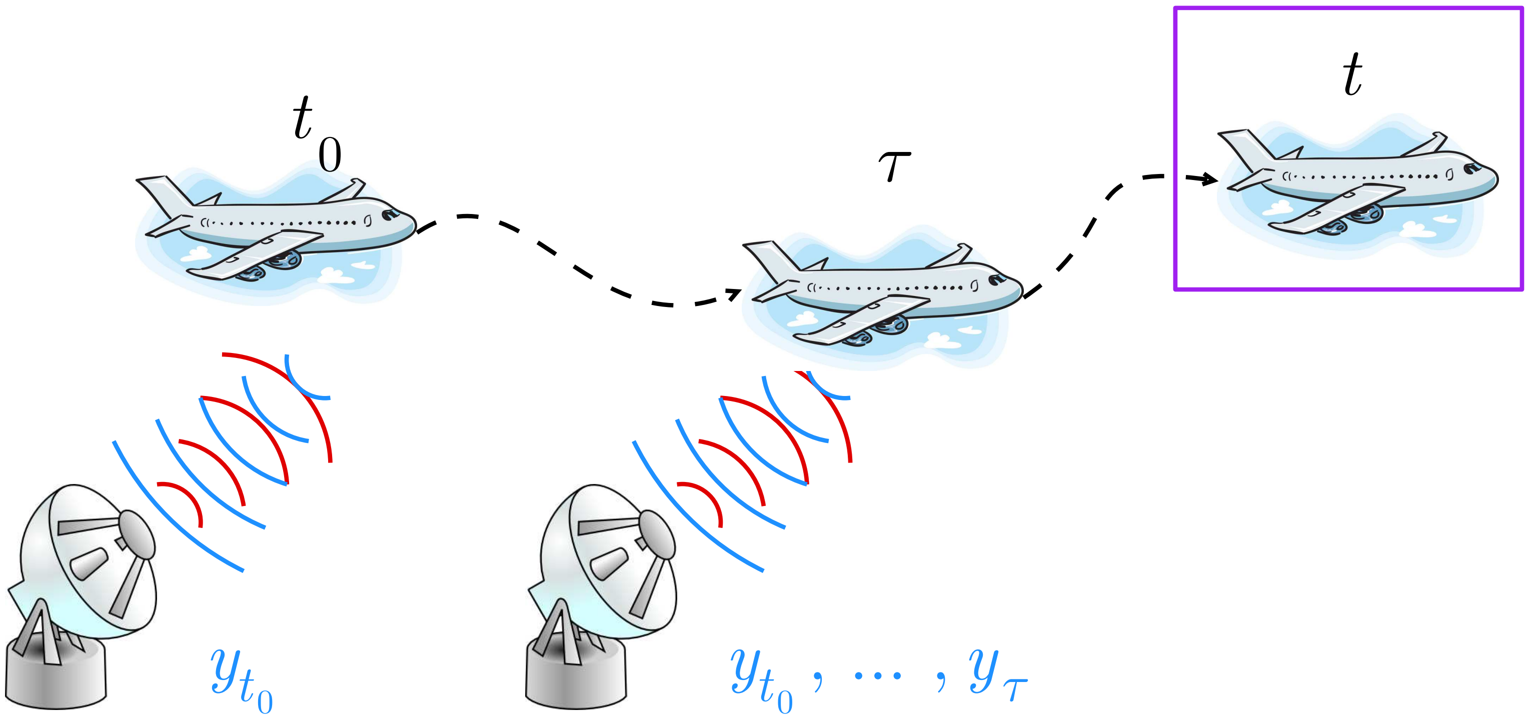


Cost of making a bad decision

Regret in Statistics

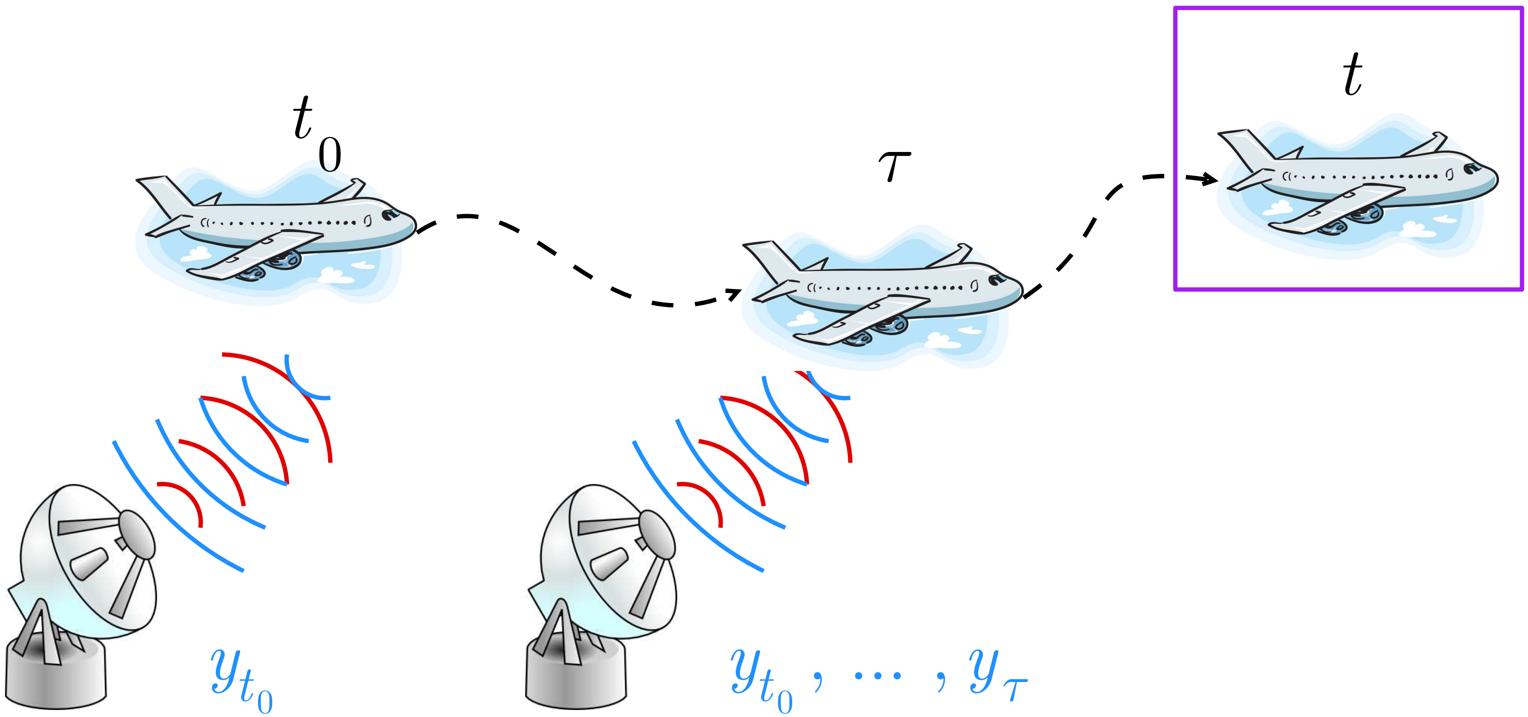


$$\text{mean-square error} = \mathbb{E}\left[X - \check{X}(Y)\right]^2$$



$$\text{mean-square error} = \mathbb{E}\left[X - \check{X}(Y)\right]^2$$

$$\text{minimum mean-square error} = \mathbb{E}\left[X - \mathbb{E}\left(X|Y\right)\right]^2$$

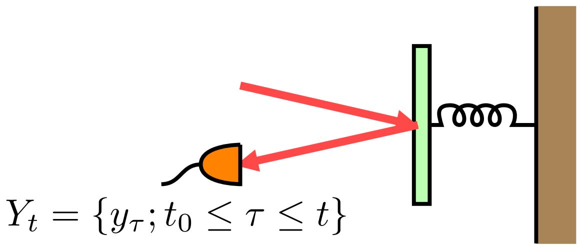


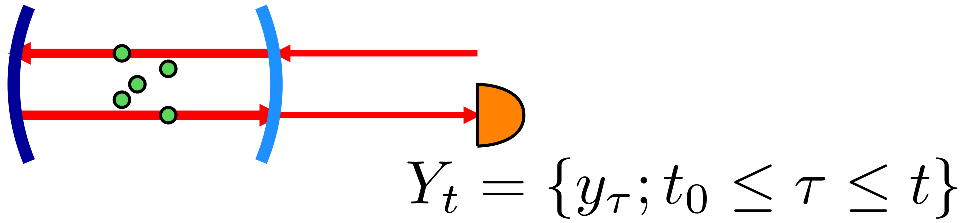
$$\text{mean-square error} = \mathbb{E}\left[X - \check{X}(Y)\right]^2$$

$$\text{minimum mean-square error} = \mathbb{E}\left[X - \mathbb{E}\left(X|Y\right)\right]^2$$

Regret = mse−mmse

Quantum Filtering



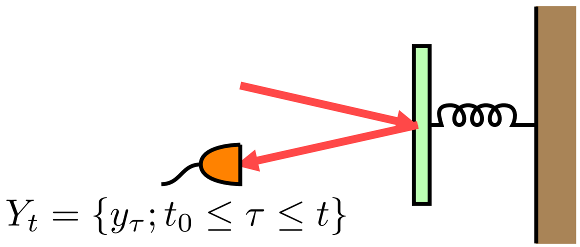


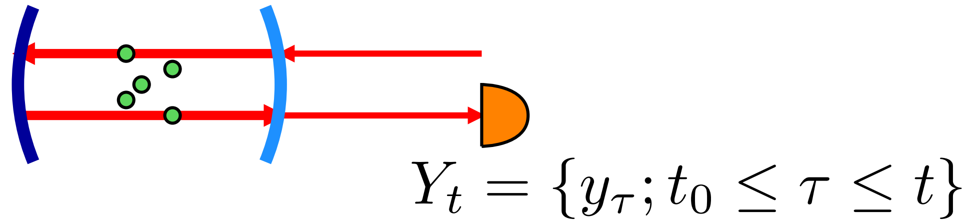
- Measurement-induced squeezing, cooling, control, etc.
- linear Belavkin equation (continuous Gaussian measurements):

$$df_t = \mathcal{L}_t f_t dt + \frac{1}{2}\left(a_t f_t + f_t a_t^\dagger\right) dy_t, \tag{1}$$

Solve for unnormalized posterior f_t from $(\rho_0, a_t, \mathcal{L}_t)$.

Quantum Filtering





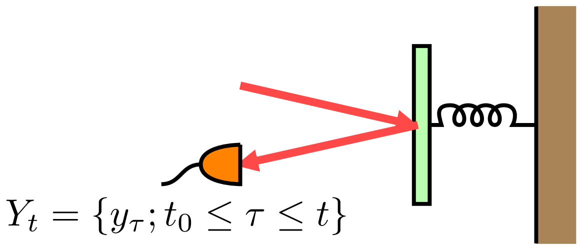
- Measurement-induced squeezing, cooling, control, etc.
- linear Belavkin equation (continuous Gaussian measurements):

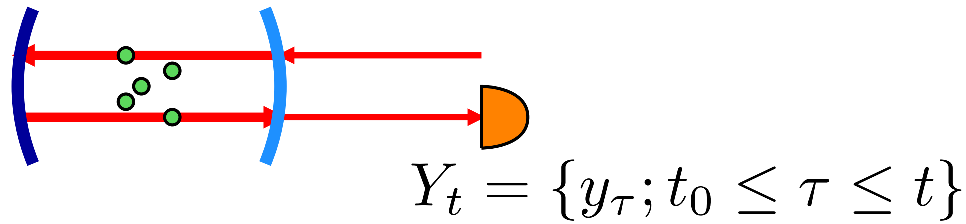
$$df_t = \mathcal{L}_t f_t dt + \frac{1}{2}\left(a_t f_t + f_t a_t^\dagger\right) dy_t, \tag{1}$$

  Solve for unnormalized posterior f_t from $(\rho_0, a_t, \mathcal{L}_t)$.
- Conditional expectation:

$$\mathbb{E}\left(q_t | Y_t\right) = \frac{\mathrm{tr}\, q_t f_t}{\mathrm{tr}\, f_t}. \tag{2}$$

Quantum Filtering





- Measurement-induced squeezing, cooling, control, etc.
- linear Belavkin equation (continuous Gaussian measurements):

$$df_t = \mathcal{L}_t f_t dt + \frac{1}{2}\left(a_t f_t + f_t a_t^\dagger\right) dy_t, \tag{1}$$

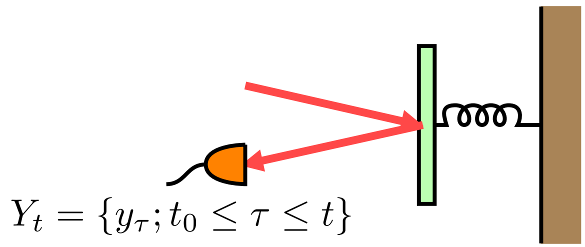
  Solve for unnormalized posterior f_t from $(\rho_0, a_t, \mathcal{L}_t)$.
- Conditional expectation:

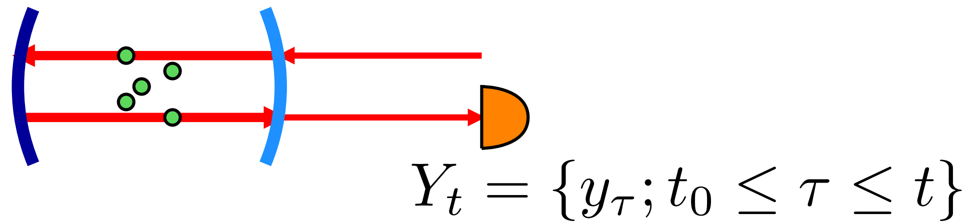
$$\mathbb{E}\left(q_t | Y_t\right) = \frac{\operatorname{tr} q_t f_t}{\operatorname{tr} f_t}. \tag{2}$$

- Minimum mean-square error:

$$\text{mmse}_t = \mathbb{E}\left[q_t - \mathbb{E}\left(q_t | Y_t\right)\right]^2. \tag{3}$$

Quantum Filtering





- Measurement-induced squeezing, cooling, control, etc.
- linear Belavkin equation (continuous Gaussian measurements):

$$df_t = \mathcal{L}_t f_t dt + \frac{1}{2}\left(a_t f_t + f_t a_t^\dagger\right) dy_t, \tag{1}$$

  Solve for unnormalized posterior f_t from $(\rho_0, a_t, \mathcal{L}_t)$.
- Conditional expectation:

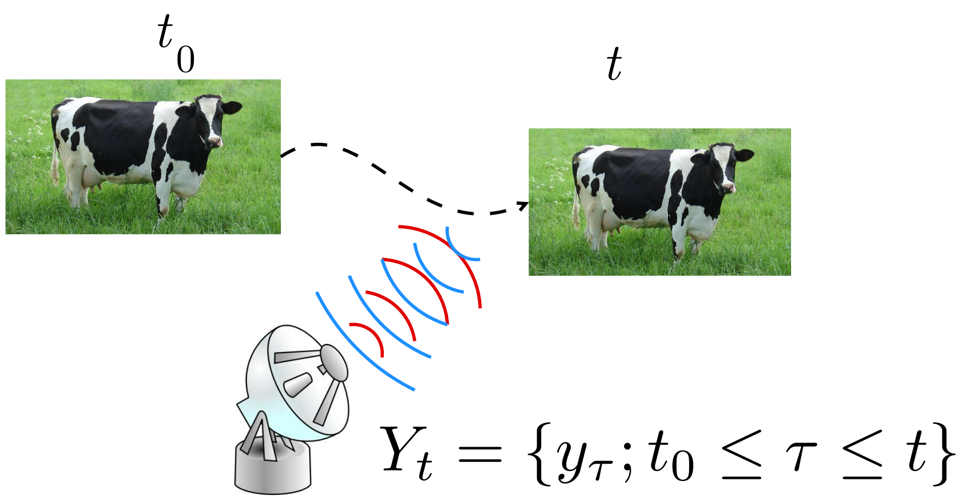
$$\mathbb{E}\left(q_t | Y_t\right) = \frac{\mathrm{tr}\, q_t f_t}{\mathrm{tr}\, f_t}. \tag{2}$$

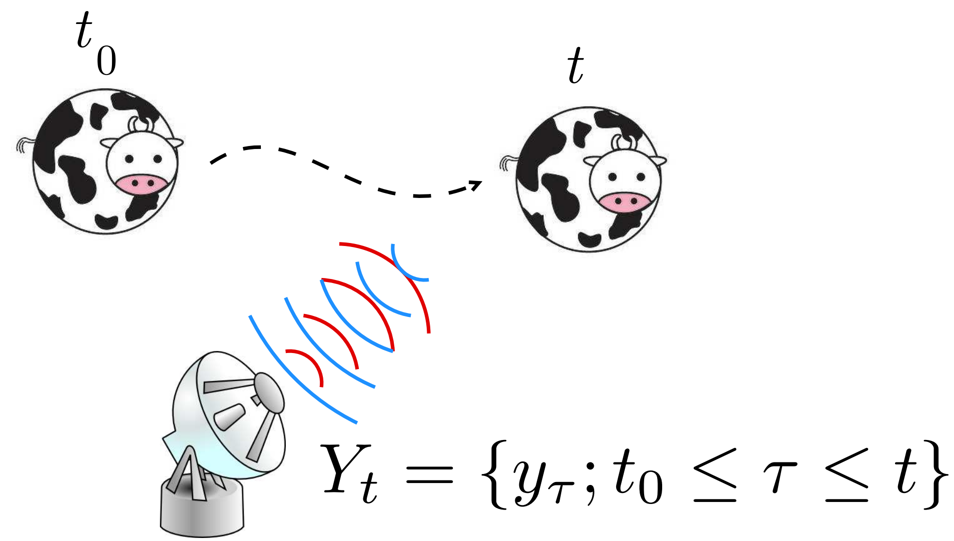
- Minimum mean-square error:

$$\mathrm{mmse}_t = \mathbb{E}\left[q_t - \mathbb{E}\left(q_t | Y_t\right)\right]^2. \tag{3}$$

- Suppose $q_t = (a_t + a_t^\dagger)/2$, the **directly measured** observable. $y_t - \int_0^t d\tau\, \mathbb{E}(q_\tau | Y_\tau)$ is a Wiener process, called the innovation process.

Filter Mismatch Regret





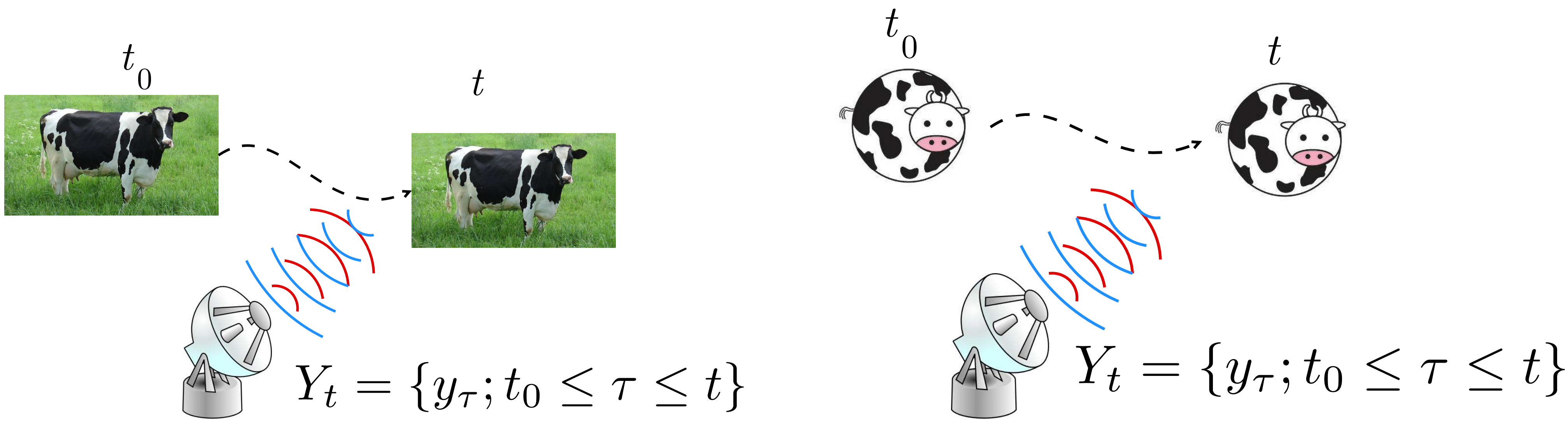
- Assume a wrong model $(\rho'_0, a'_t, \mathcal{L}'_t)$:

$$df'_t = \mathcal{L}'_t f'_t dt + \frac{1}{2}\left(a'_t f'_t + f'_t a'^{\dagger}_t\right) dy_t, \qquad \mathbb{E}'\left(q'_t|Y_t\right) = \frac{\mathrm{tr}\, q'_t f'_t}{\mathrm{tr}\, f'_t}. \tag{4}$$

- Regret due to filter mismatch:

$$R \equiv \frac{1}{2}\int_0^T dt \left\{\mathbb{E}\left[q_t - \mathbb{E}'\left(q'_t|Y_t\right)\right]^2 - \mathrm{mmse}_t\right\}. \tag{5}$$

Regret = Relative Entropy



$$\boxed{R = D(dP||dP') \equiv \int dP(Y_T) \ln \frac{dP(Y_T)}{dP'(Y_T)}.} \quad (6)$$

- **Relative entropy**, Kullback-Leibler divergence, etc.
- A measure of **distinguishability** between distributions

Proof [arXiv:1310.0291]

- $\operatorname{tr} f_t$ is probability density:

$$\operatorname{tr} f_t = \frac{dP(Y_t)}{dP_0(Y_t)}, \tag{7}$$

  $dP_0 =$ Wiener measure.
- From Belavkin,

$$d \operatorname{tr} f_t = \operatorname{tr} df_t = \operatorname{tr} \mathcal{L}_t f_t dt + \frac{1}{2} \operatorname{tr} \left(a_t f_t + f_t a_t^\dagger \right) dy_t = \mathbb{E} \left(q_t | Y_t \right) \left(\operatorname{tr} f_t \right) dy_t . \tag{8}$$

- Itō calculus ($dy_t^2 = dt$):

$$\ln \operatorname{tr} f_T = \int_0^T dy_t \, \mathbb{E} \left(q_t | Y_t \right) - \frac{1}{2} \int_0^T dt \, \mathbb{E}^2 \left(q_t | Y_t \right), \tag{9}$$

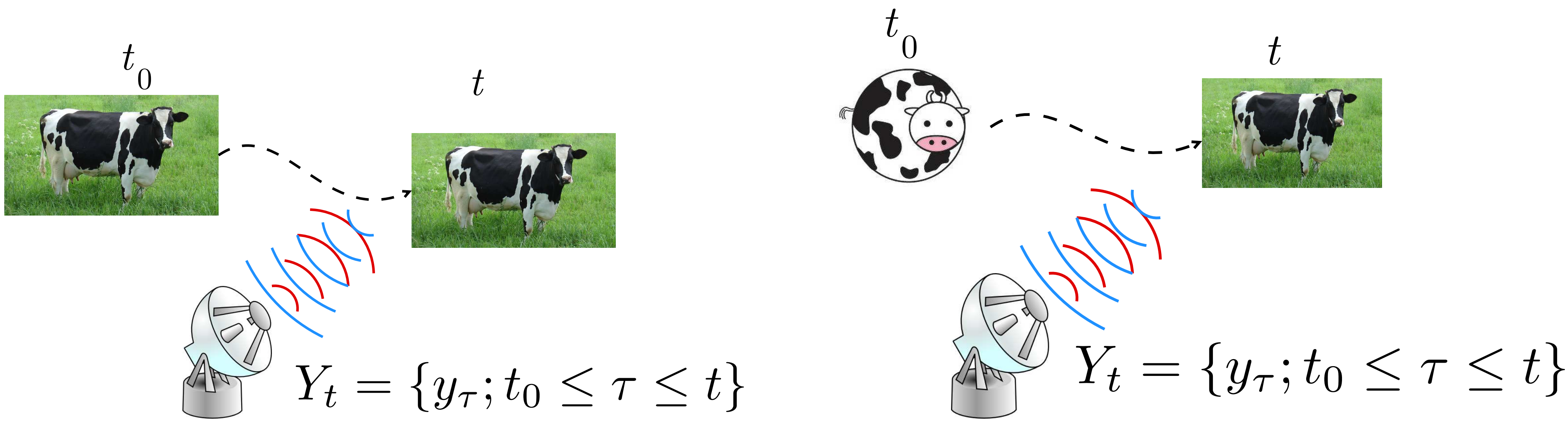
$$\tag{10}$$

  Similar for $\ln \operatorname{tr} f'_T$ [Tsang, PRL **108**, 170502 (2012)].
- Relative entropy is expected log-likelihood ratio:

$$D(dP||dP') = \mathbb{E} \ln \frac{dP(Y_T)}{dP'(Y_T)} = \mathbb{E} \left(\ln \operatorname{tr} f_T - \ln \operatorname{tr} f'_T \right) . \tag{11}$$

- Use basic properties of **conditional expectation** and **innovation process** $\mathbb{E}[dy_t g(Y_t)] = \mathbb{E}[\mathbb{E}(dy_t|Y_t) g(Y_t)] = \mathbb{E}[dt \, \mathbb{E}(q_t|Y_t) g(Y_t)] = dt \, \mathbb{E}[q_t g(Y_t)]$

Example: Mismatched Initial Condition



- Suppose dynamics and measurements are accurate ($\mathcal{L}_t = \mathcal{L}'_t$, $a_t = a'_t$), only $\rho'_0 \neq \rho_0$,

$$dP(Y_T) = \operatorname{tr} d\mu(Y_T)\rho_0, \qquad dP'(Y_T) = \operatorname{tr} d\mu(Y_T)\rho'_0. \tag{12}$$

- From quantum information,

$$D(dP||dP') \leq D(\rho_0||\rho'_0) = \operatorname{tr} \rho_0 \left(\ln \rho_0 - \ln \rho'_0\right). \tag{13}$$

- Regret is **upper-bounded**:

$$R \leq \operatorname{tr} \rho_0 \left(\ln \rho_0 - \ln \rho'_0\right). \tag{14}$$

Ignorance Regret

- True model is one of $\{\rho_0^\theta, a_t^\theta, \mathcal{L}_t^\theta\}$, chosen with a prior distribution of $d\pi(\theta)$.
- If I know θ, $\mathbb{E}(q_t|Y_t) = \mathbb{E}(q_t|Y_t, \theta)$.
- If I don't know θ, mse is minimized by Bayesian estimation of both q_t and θ from Y_t;

$$\mathbb{E}'(q_t'|Y_t) = \mathbb{E}_\theta \left[\mathbb{E}(q_t|Y_t, \theta)|Y_t\right]. \tag{15}$$

- Observation probability measures:

$$dP(Y_T) = dP_\theta(Y_T), \qquad dP'(Y_T) = \mathbb{E}_\theta\, dP_\theta(Y_T). \tag{16}$$

Ignorance Regret

- True model is one of $\{\rho_0^\theta, a_t^\theta, \mathcal{L}_t^\theta\}$, chosen with a prior distribution of $d\pi(\theta)$.
- If I know θ, $\mathbb{E}(q_t|Y_t) = \mathbb{E}(q_t|Y_t,\theta)$.
- If I don't know θ, mse is minimized by Bayesian estimation of both q_t and θ from Y_t;

$$\mathbb{E}'(q_t'|Y_t) = \mathbb{E}_\theta\left[\mathbb{E}(q_t|Y_t,\theta)|Y_t\right]. \tag{15}$$

- Observation probability measures:

$$dP(Y_T) = dP_\theta(Y_T), \qquad dP'(Y_T) = \mathbb{E}_\theta\, dP_\theta(Y_T). \tag{16}$$

- Regret due to **ignorance**:

$$\boxed{\min_{\{\rho_0', a_t', \mathcal{L}_t'\}} \mathbb{E}_\theta\, R = \mathbb{E}_\theta\, D(dP_\theta || \mathbb{E}_\theta\, dP_\theta) \equiv I(\theta; Y),} \tag{17}$$

Shannon mutual information.

Ignorance Regret

- True model is one of $\{\rho_0^\theta, a_t^\theta, \mathcal{L}_t^\theta\}$, chosen with a prior distribution of $d\pi(\theta)$.
- If I know θ, $\mathbb{E}(q_t|Y_t) = \mathbb{E}(q_t|Y_t, \theta)$.
- If I don't know θ, mse is minimized by Bayesian estimation of both q_t and θ from Y_t;

$$\mathbb{E}'(q_t'|Y_t) = \mathbb{E}_\theta\left[\mathbb{E}(q_t|Y_t, \theta)|Y_t\right]. \tag{15}$$

- Observation probability measures:

$$dP(Y_T) = dP_\theta(Y_T), \qquad dP'(Y_T) = \mathbb{E}_\theta\, dP_\theta(Y_T). \tag{16}$$

- Regret due to **ignorance**:

$$\boxed{\min_{\{\rho_0', a_t', \mathcal{L}_t'\}} \mathbb{E}_\theta\, R = \mathbb{E}_\theta\, D(dP_\theta || \mathbb{E}_\theta\, dP_\theta) \equiv I(\theta; Y),} \tag{17}$$

  Shannon mutual information.
- Holevo bound for mismatched initial condition:

$$\min_{\{\rho_0', a_t', \mathcal{L}_t'\}} \mathbb{E}_\theta\, R = I(\theta; Y) \le \mathbb{E}_\theta\, D(\rho_0^\theta || \mathbb{E}_\theta\, \rho_0^\theta). \tag{18}$$

- **measure of parameter importance**.

Maximin and Minimax Regret

- Worst-case Bayesian regret (maximin regret):

$$\max_{d\pi} \min_{\{\rho'_0, a'_t, \mathcal{L}'_t\}} \mathbb{E}_\theta\, R = \max_{d\pi} I(\theta; Y) \equiv C, \tag{19}$$

  which is **channel capacity**.
- maximin is equal to minimax (von Neumann minimax theorem):

$$\max_{d\pi} \min_{\{\rho'_0, a'_t, \mathcal{L}'_t\}} \mathbb{E}_\theta\, R = \min_{\{\rho'_0, a'_t, \mathcal{L}'_t\}} \max_{d\pi} \mathbb{E}_\theta\, R = C. \tag{20}$$

- Redundancy-capacity theorem:

$$C = \max_{d\pi} \min_{dP'} \mathbb{E}_\theta\, D(dP_\theta || dP') = \min_{dP'} \max_{d\pi} \mathbb{E}_\theta\, D(dP_\theta || dP'). \tag{21}$$

Poissonian Measurements

- Belavkin equation:

$$df_t = \mathcal{L}_t f_t dt + \left[a_t f_t a_t^\dagger - f_t\right](dy_t - dt), \qquad \mathrm{tr}\, f_t = \frac{dP(Y_t)}{dP_0(Y_t)}. \tag{22}$$

dP_0 =Poisson process measure. Directly measured observable is $q_t = a_t^\dagger a_t$.

- Define loss function

$$\boxed{l(q, \check{q}) \equiv q \ln \frac{q}{\check{q}} - q + \check{q}.} \tag{23}$$

- Define regret:

$$R_l \equiv \int dt \left[l\left(q_t, \mathbb{E}'(q_t'|Y_t)\right) - l\left(q_t, \mathbb{E}(q_t|Y_t)\right)\right]. \tag{24}$$

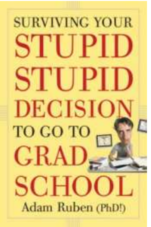
- Same identity:

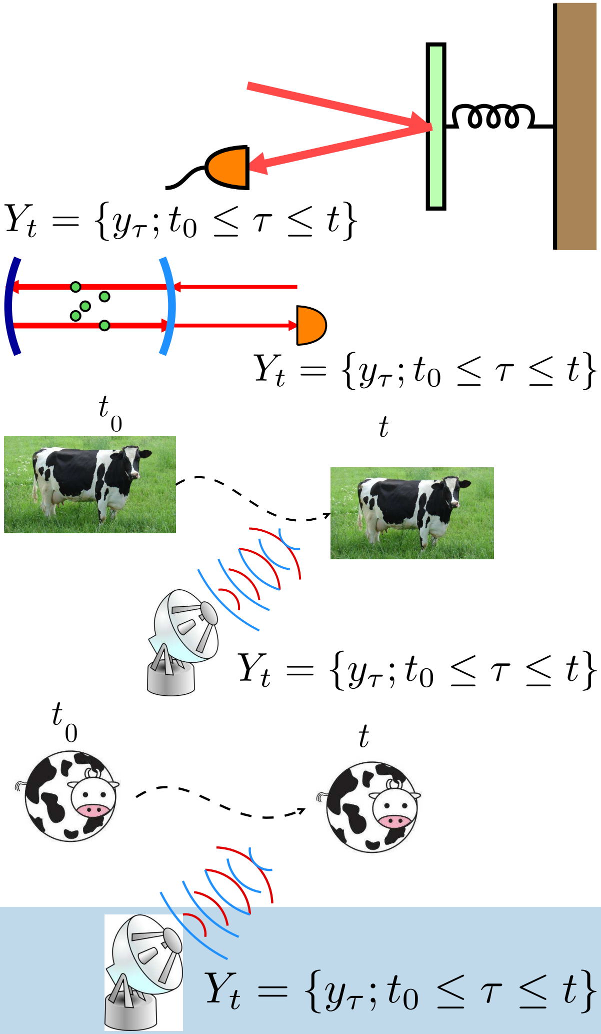
$$R_l = D(dP||dP'). \tag{25}$$

Classical Estimation and Information

- T. E. Duncan, "On the calculation of mutual information," SIAM Journal on Applied Mathematics 19, 215–220 (1970).
- D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," IEEE Transactions on Information Theory, vol. 51, no. 4, pp. 1261–1282, april 2005.
- —, "Mutual information and conditional mean estimation in Poisson channels," IEEE Transactions on Information Theory, vol. 54, no. 5, pp. 1837–1849, 2008.
- S. Verdu, "Mismatched estimation and relative entropy," IEEE Transactions on Information Theory, vol. 56, no. 8, pp. 3712–3720, 2010.
- T. Weissman, "The relationship between causal and noncausal mismatched estimation in continuous-time AWGN channels," IEEE Transactions on Information Theory, vol. 56, no. 9, pp. 4256–4273, 2010.
- R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," IEEE Transactions on Information Theory, vol. 58, no. 3, pp. 1302–1318, 2012.
- A. No and T. Weissman, "Minimax Filtering via Relations between Information and Estimation," arXiv:1301.5096, ArXiv e-prints, Jan. 2013.

Conclusion





- Regret = relative entropy
- Upper bound on regret
- Mutual information, channel capacity
- M. Tsang, arXiv:1310.0291.
- **Quantum information for dynamical systems**
- `http://mankei.tsang.googlepages.com/`
- `eletmk@nus.edu.sg`