relations between

information

and

estimation

in the presence of feedback

Tsachy Weissman

talk at workshop on:

Information and Control in Networks

LCCC - Lund Center for Control of Complex Engineering Systems

Monday, October 22, 12

information, control, networks

- real-time and limited delay communication
- **•** feedback communications
- "action" in information theory
- relations between information and estimation (w. feedback + networks)

Outline

Haves and Have-Nots (in this talk)

we'll have:

- some theorems
- cute (and meaningful) relations
- we won't have: • an algorithmic framework
- account of related literature
- stipulations
- proofs
- algorithms

"de Bruijn's identity" \overline{a} ["]de Bruijn's ident $\ddot{ }$ *by''* [A. J. Stam 1959]:

D (exp(x) = 12) = 12 = 12 = 12 = 12 = 12

Suppose that *X* is a non-negative random variable and the conditional law of a r.v. *Y*, given *X*, is Poisson(⇤*X*). If

X ⌅ *P*, denote expectation w.r.t. the corresponding joint law of *X* and *Y* by *E^P* , the distribution of *Y* by *P^Y* ,

the conditional expectation by *E^P* [*X|Y*], etc. We denote the mutual information by *I^P* (*X*; *Y*) or simply *I*(*X*; *Y*)

when there is no ambiguity. Let further mle*P,Q*(⇤) denote the mean loss under ⌥ in estimating *X* based on *Y* using

^N (*µ*1*,* ⌃²)⌥*^N* (*µ*2*,* ⌃²)

^X ⁺

the estimator that would have been optimal had *X* ⌅ *Q* when in fact *X* ⌅ *P*, i.e.,

X independent of $Z \sim \mathcal{N}(0, 1)$ \overline{a} λ ²⌃² *·* (*µ*¹ *^µ*2)

$$
\frac{d}{dt}h\left(X+\sqrt{t}Z\right) = \frac{1}{2}J(X+\sqrt{t}Z)
$$

Suppose that *X* is a non-negative random variable and the conditional law of a r.v. *Y*, given *X*, is Poisson(⇤*X*). If

X ⌅ *P*, denote expectation w.r.t. the corresponding joint law of *X* and *Y* by *E^P* , the distribution of *Y* by *P^Y* ,

the conditional expectation by *E^P* [*X|Y*], etc. We denote the mutual information by *I^P* (*X*; *Y*) or simply *I*(*X*; *Y*)

X independent of *Z* ⌅ *N* (0*,* 1)

6.2 Random Variables

GuoShamai Verdu setting $\sqrt{ }$ lu *·* cmmse() *I* erdu \textsf{S} **C** ² *·* cmmse()

2

E

^I() =

^I(*X^T* ; *^Y ^T*) = ¹

3 Introduction

^I() = ¹

d

d

$$
Y=\sqrt{\gamma}\cdot X+W
$$

W is a standard Gaussian, independent of *X* W i *X* \sim *x* \sim *i* \sim *x* $\$ *W* is a standard Gaussian, independent of *X*

$$
I(\gamma)=I(X;Y)
$$

^I() = ¹

$$
\mathsf{mmse}(\gamma) = E\left[(X - E[X|Y])^2 \right]
$$

· cmmse()

 $\mathbb{E}\left[\mathbf{X}^T\mathbf{X$

(*X^t ^E*[*Xt|^Y ^t*

(*^X ^E*[*X|^Y*])²⇥

mmse()

² *·* cmmse()

^Y ⁼ ⌅ *· ^X* ⁺ *^W*

])²*dt*⌅

I() = *I*(*X*; *Y*)

W is a standard Gaussian, independent of *X*

[Guo, Shamai and Verd´u 2005]:

[Guo, Shamai and Verd´u 2005]:

2 for GSV slide

 $\frac{1}{2}$

[Guo, Shamai and Verdú 2005]: mmse() = *E* (*^X ^E*[*X|^Y*])²⇥

I() = *I*(*X*; *Y*)

mmse() = *E*

d

d

the input and the output in a real-valued scalar Gaussian channel, with respect to the signal-to-noise ratio (SNR),

is equal to the minimum mean square error (MMSE) in estimating the input based on the output. This simple

relationship holds regardless of the input distribution, and carries over essentially verbatim to vectors, as well as the

continuous-time Additive White Gaussian Noise (AWGN) channel (cf. [34, 21] for even more general settings where

this relationship holds). When combined with Duncan's theorem [7], it was also shown to imply a remarkable rela-

tionship between the MMSEs in causal (filtering) and non-causal (smoothing) estimation of an arbitrarily distributed

continuous-time signal correction at SNR level is equal to the filtering MMSE at SNR level is equal to the me

 $\frac{d}{dx}I(\gamma)=\frac{1}{x}$ In the seminal paper (13) , Guo, Shamai and Verdiu discovered that the derivative of the mutual information between α *d* $d\gamma$ $I(\gamma) = \frac{1}{\gamma}$ 2 mm se (γ)

In the seminal paper [13], Guo, Shamai and Verd´u discovered that the derivative of the mutual information between

the input and the output in a real-valued scalar Gaussian channel, with respect to the signal-to-noise ratio (SNR),

GSV in continuous time ⇤ *^T*])²*dt*⇥ **I**CC ⇤ *^T* (*X^t ^E*[*Xt|^Y ^t* me

dY^t = *Xtdt* + *dWt,* 0 ⇥ *t* ⇥ *T*

^Y ⁼ ⇤ *· ^X* ⁺ *^W*

dY^t = *Xtdt* + *dWt,* 0 ⇥ *t* ⇥ *T*

$$
dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T
$$

$$
I(\gamma)=I(X^T;Y^T)
$$

2

E

22
22 March 2014
22 March 2014

$$
\text{mmse}(\gamma) = E\left[\int_0^T (X_t - E[X_t|Y^T])^2 dt\right]
$$

(*X^t ^E*[*Xt|^Y ^t*

])²*dt*⇥

W is white Gaussian noise, independent of *X*

[Guo, Shamai and Verd´u 2005]:

W is white Gaussian noise, independent of *X*

W is a standard Gaussian, independent of *X*

mai and Verdú 2005]
 [Zakai 2005]: [Guo, Shamai and Verdú 2005] [Zakai 2005]:

mmse() = *E*

W is a standard Gaussian, independent of *X*

$$
\frac{d}{d\gamma}I(\gamma)=\frac{1}{2}\mathsf{mmse}(\gamma)
$$

or in its integral version

is equal to the minimum mean square error (MMSE) in estimating the input based on the output. This simple

relationship holds regardless of the input distribution, and carries over essentially verbatim to vectors, as well as the

$$
I(\mathsf{snr}) = \frac{1}{2} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma
$$

(*^X ^E*[*X|^Y*])²⇥

^Y ⁼ ⇤ *· ^X* ⁺ *^W*

I() = *I*(*X*; *Y*)

⇧ snr

(*^X ^E*[*X|^Y*])²⇥

(*X^t ^E*[*Xt|^Y ^T*])²*dt*

[Guo, Shamai and Verd´u 2005]:

[Guo, Shamai and Verd´u 2005]:

Duncan Duncan slide 1 for Duncan slides in the Duncan slides in the Duncan slides in the Duncan slides in the Duncan slides in the
1 for Duncan slides in the Duncan slides in the Duncan slides in the Duncan slides in the Duncan slides in the

$$
dY_t = X_t dt + dW_t, \qquad 0 \le t \le T
$$

 W is standard white Gaussian noise, independent of X

^I(*X^T* ; *^Y ^T*) = ¹ *E* $\frac{1}{2}$ $\left[\frac{X}{X}\right]$: *[Duncan 1970]:*

$$
I(X^T; Y^T) = \frac{1}{2}E\left[\int_0^T (X_t - E[X_t|Y^t])^2 dt\right]
$$

(*X^t ^E*[*Xt|^Y ^t*

^I(*X^T* ; *^Y ^T*) = ¹

W is white Gaussian noise, independent of *X*

])²*dt*⇥

⇤ *^T*

^I(*X^T* ; *^Y ^T*) = ¹

(*X^t ^E*[*Xt|^Y ^t*

(*X^t ^E*[*Xt|^Y ^t*

(*X^t ^E*[*Xt|^Y ^t*

dY^t ⁼ ⇧*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T*

E

cmmse() = ¹

1 for Duncan slide

W is white Gaussian noise, independent of *X*

SNR in Duncan 2 *^I*(*X^T* ; *^Y ^T*) = ¹ *E* 0 \sum_{x} *I*^{*Z*} *Z IX I*^{*X*} *IX I*^{*X*} *IX I* 2 *E*_{*I*} \int *EDuncan* ⇤ *^T* 2 ⇤⇧ *^T* **can**

])²*dt*⇥

 $\mathcal{L} = \mathcal{L} \times \mathcal{L}$

^I(*X^T* ; *^Y ^T*) = ¹

])²*dt*⇥

dY^t ⁼ ⇧*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T*

⇤ *^T*

I() = *I*(*X^T* ; *Y ^T*)

dY^t ⁼ ⇧*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T* $I(\gamma) = I(X^T;Y^T)$ $\mathsf{cmmse}(\gamma)$ $= E$ $(3 \times t \times t^2)$ **E**[$2 \times t$] $\left\lfloor \frac{2}{t} \right\rfloor$ $\overline{I}(x)$ ² \overline{a} $\overline{}$ $I(\gamma) = \frac{\gamma}{\gamma} \cdot \text{cmms}(1)$ $dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T$ γ) = $I(\Lambda, T)$ $\text{cmmsc}(1) =$ $\frac{10}{20}$ $\int_0^1 \left| \int_0^1 \right|$ $\left[\frac{t}{\sqrt{2}}\right]$ $\frac{1}{2}$ $\frac{y}{-}$ 0 m se (γ) *[Duncan 1970]:* 2 Introduction $dY_t =$ $=$ $\sqrt{ }$ $\int \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T$])²*dt*⇥ *^I*(*X^T* ; *^Y ^T*) = ¹ $111.$ $\mathsf{se}(\mathsf{f})$ $T = T$ $\int_{0}^{T} (X_t - E[X_t|Y^t])^2 dt$ $I(\gamma) = \frac{\gamma}{\Omega}$ 2 \cdot cmmse (γ) $= I(X^T; Y^T)$ $\mathsf{cmmse}(\gamma) = E$ $\int f^T$ 0 $(X_t - E[X_t|Y^t])$ $])^2dt$ ^{$\Big]$} ⇤⇧ *^T* (*X^t ^E*[*Xt|^Y ^t* 2 ² *·* cmmse()

W is standard white Gaussian noise, independent of *X*

 $\overline{}$

 $\overline{}$

^I(*X^T* ; *^Y ^T*) = ¹ 2 *E* 0 *d ^I*() = ¹ 2 for GSV slide snr ⇧ snr Recap

mmse() = *E*

0

(*^X ^E*[*X|^Y*])²⇥

E

^I(*X^T* ; *^Y ^T*) = ¹

$$
I(\gamma) = \frac{\gamma}{2} \cdot \mathsf{cmmse}(\gamma)
$$

I() = *I*(*X^T* ; *Y ^T*)

])²*dt*⌅

(*X^t ^E*[*Xt|^Y ^t*

(*X^t ^E*[*Xt|^Y ^t*

(*X^t ^E*[*Xt|^Y ^t*

cmmse()

is equal to the minimum mean square error (MMSE) in estimating the input based on the output. This simple

continuous-time signal corrupted by Gaussian noise: the filtering MMSE at SNR level is equal to the mean value

of the smoothing MSE with SNR uniformly distributed between \mathcal{M} and . The relation of the mutual information of

 \Rightarrow $\hat{ }$

the mismatch is equal to the relative entropy between the true channel output distribution and the channel output

More recently, Verd´u has shown in [31] that when *X* ⌅ *P* is estimated based on *Y* by a mismatched estimator

This result was key in $[33]$, where it was shown that the relationship between the causal α

to both types of MMSE thus served as a bridge between the two quantities.

In the seminal paper [13], Guo, Shamai and Verd´u discovered that the derivative of the mutual information between [Guo, Shamai and Verdú 2005], [Zakai 2005]:

$$
I({\sf snr}) = \frac{1}{2} \int_0^{{\sf snr}} {\sf mmse}(\gamma) d\gamma
$$

is equal to the minimum mean square error (MMSE) in estimating the input based on the output. This simple

 r

 \Rightarrow ?

continuous-time Additive White Gaussian Noise (AWGN) channel (cf. [34, 21] for even more general settings where

this relationship holds). When combined with Duncan's theorem [7], it was also shown to imply a remarkable rela-Monday, October 22, 12

[Guo, Shamai and Verd´u 2005]:

Relationship between cmmse and mmse?

ationship h $\frac{1}{2}$ (*X^t ^E*[*Xt|^Y ^T*])²*dt*⌅ Relationship between cmmse and mmse?

2

mmse()

mmse() = *E*

 $\frac{1}{\sqrt{2}}$

d

d

In the seminal paper [13], Guo, Shamai and Verd´u discovered that the derivative of the mutual information between

the input and the output in a real-valued scalar Gaussian channel, with respect to the signal-to-noise ratio (SNR),

is equal to the minimum mean square error (MMSE) in estimating the input based on the output. This simple

relationship holds regardless of the input distribution, and carries over essentially verbatim to vectors, as well as the

^I() = ¹

d

^I(snr) = ¹ [Guo, Shamai and Verdú 2005]:

$$
\text{cmmse}(\text{snr}) = \frac{1}{\text{snr}} \int_0^{\text{snr}} \text{mmse}(\gamma) d\gamma
$$

Relationship between cmmse and mmse?

Mismatch communisms natch

⇧ snr

· cmmse()

^I() =

snr ⇧ snr

⇧ snr

mmse() = *E*

X ⇤ *P*

^Y ⁼ ⌅ *· ^X* ⁺ *^W*

I() = *I*(*X*; *Y*)

^I() = ¹

^I() =

^I(snr) = ¹

[Duncan 1970]:

or in its integral version

$$
Y=\sqrt{\gamma}\cdot X+W
$$

 $\frac{W}{S}$ *W* is a standard Gaussian, independent of *X*

> What if $X \sim P$ but the estimator thinks $X \sim Q$? $\frac{1}{2}$

$$
\text{mse}_{P,Q}(\gamma) = E_P\left[(X-E_Q[X|Y])^2\right]
$$

mmse ()
mmse ()
mmse ()

^I() = ¹

3 Introduction of the state of t
3 Introduction of the state of

What if *X* ⇤ *P* but the estimator thinks *X* ⇤ *Q* ?

[Guo, Shamai and Verd´u 2005]:

Relationship between cmmse and mmse?

A representation of relative entropy [Verdu 2010]: \parallel *A* representatic mse*P,Q*() = *E^P* (*^X ^EQ*[*X|^Y*])²⇥ $\overline{6}$ $\overline{ }$

+

$$
D(P\|Q)=\int_0^\infty [{\rm mse}_{P,Q}(\gamma)-{\rm mse}_{P,P}(\gamma)]d\gamma
$$

$$
D(P_{Y_{\text{snr}}} \| Q_{Y_{\text{snr}}}) = \int_0^{\text{snr}} [\text{mse}_{P,Q}(\gamma) - \text{mse}_{P,P}(\gamma)] d\gamma
$$

X ⇥ *P*

What is Cost of Mismatch?

What is Cost of Mismatch?

Relationship between cmmse and mmse?

What if *X* ⇥ *P* but the estimator thinks *X* ⇥ *Q* ?

3 Introduction

Causal vs. Non-causal Mismatched Estimation *dY^t* ⁼ ⇤*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T dY^t* ⁼ ⇤*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T* a **I** vs. Non-causa **T**
I Mis (*X^t ^E*[*Xt|^Y ^t* na<mark>tio</mark>r

$$
dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T
$$

W is standard white Gaussian noise, independent of *X*

² *·* cmmse()

² *·* cmmse()

(*X^t ^E*[*Xt|^Y ^t*

(*X^t ^E*[*Xt|^Y ^t*

I() = *I*(*X^T* ; *Y ^T*)

⇤ *^T*

^I(*X^T* ; *^Y ^T*) = ¹

$$
\text{cmse}_{P,Q}(\gamma) = E_P \left[\int_0^T (X_t - E_Q[X_t|Y^t])^2 dt \right]
$$

$$
\text{mse}_{P,Q}(\gamma) = E_P \left[\int_0^T (X_t - E_Q[X_t|Y^T])^2 dt \right]
$$

^I() =

^I() =

Causal vs. Non-causal Mismatched Estimation *dY^t* ⁼ ⇤*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T dY^t* ⁼ ⇤*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T I*() = *I*(*X^T* ; *Y ^T*) a **I** vs. Non-causa **T**
I Mis (*X^t ^E*[*Xt|^Y ^t* na<mark>tio</mark>r

$$
dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T
$$

W is standard white Gaussian noise, independent of *X* Frontenant Miss

(*X^t ^E*[*Xt|^Y ^t*

⇤⇧ *^T*

dY^t ⁼ ⇤*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T*

(*X^t ^E*[*Xt|^Y ^t*

⇤⇧ *^T*

I() = *I*(*X^T* ; *Y ^T*)

⇤ *^T*

E

2

^I(*X^T* ; *^Y ^T*) = ¹

^I(*X^T* ; *^Y ^T*) = ¹

$$
\text{cmse}_{P,Q}(\gamma) = E_P \left[\int_0^T (X_t - E_Q[X_t|Y^t])^2 dt \right]
$$

$$
\text{mse}_{P,Q}(\gamma) = E_P \left[\int_0^T (X_t - E_Q[X_t|Y^T])^2 dt \right]
$$

 $\text{and} \ \text{msc}_P Q$ *^I*(*X^T* ; *^Y ^T*) = ¹ \mathbf{r} \overline{C} $\overline{}$ (*X^t ^E*[*Xt|^Y ^t* $\frac{d}{dx}$ *^I*(*X^T* ; *^Y ^T*) = ¹ $\overline{}$ *E* x^2 $\sum_{i=1}^n x^i$])²*dt*⌅ Relationship between $\mathsf{cmse}_{P,Q}$ and $\mathsf{mse}_{P,Q}$? \overline{C} \tt{twoen} $\mathsf{cmse}_{P,Q}$ and $\mathsf{mse}_{P,Q}$ $\frac{2}{\pi}$

² *·* cmmse()

² *·* cmmse()

^I() =

^I() =

Relationship between cmse*P,Q* and mse*P,Q*

⇤⇧ *^T*

[Weissman 2010]: mse*P,Q*() = *E^P* $\overline{}$ (*X^t ^EQ*[*Xt|^Y ^T*])²*dt*⌅

^I(*X^T* ; *^Y ^T*) = ¹

])²*dt*⌅

])²*dt*⌅

^I() =

^Y ⁼ ⇤ *· ^X* ⁺ *^W*

$$
\text{cmse}_{P,Q}(\textsf{snr}) \quad = \quad \frac{1}{\textsf{snr}} \int_{0}^{\textsf{snr}} \textsf{mse}_{P,Q}(\gamma) d\gamma
$$

(*X^t ^E*[*Xt|^Y ^t*

E

Relationship between cmse*P,Q* and mse*P,Q* ?

Relationship between cmse*P,Q* and mse*P,Q*

Relationship between cmse*P,Q* and mse*P,Q* ationship betwe $\mathsf{mse}_{P, Q}$ and $\mathsf{ms}_{P, Q}$ $e_{P,Q}$

⇤⇧ *^T*

[Weissman 2010]: mse*P,Q*() = *E^P* sman 2 (*X^t ^EQ*[*Xt|^Y ^T*])²*dt*⌅ mse*P,Q*() = *E^P* (*X^t ^EQ*[*Xt|^Y ^T*])²*dt*⌅

$$
\begin{array}{lcl} \textsf{cmse}_{P,Q}(\textsf{snr}) & = & \displaystyle \frac{1}{\textsf{snr}} \int_{0}^{\textsf{snr}} \textsf{mse}_{P,Q}(\gamma) d\gamma \\ \\ & = & \displaystyle \frac{2}{\textsf{snr}} \left[I(\textsf{snr}) + D \left(P_{Y^T} \| Q_{Y^T} \right) \right] \end{array}
$$

⇤⌃ *^T*

(*X^t ^E*[*Xt|^Y ^t*

E

^I(*X^T* ; *^Y ^T*) = ¹

^I(*X^T* ; *^Y ^T*) = ¹

])²*dt*⌅

])²*dt*⌅

])²*dt*⌅

^I() =

^Y ⁼ ⇤ *· ^X* ⁺ *^W*

Relationship between cmse*P,Q* and mse*P,Q*

Relationship between cmse*P,Q* and mse*P,Q* ?

Relationship between cmse*P,Q* and mse*P,Q*

Implications and Applications

$$
\text{Minimax}(P, \text{snr}) \triangleq \underset{\{\hat{X}_t(\cdot)\}_{0 \leq t \leq T}}{\min} \max_{P \in \mathcal{P}} \left\{ E_P \left[\int_0^T \ell(X_t, \hat{X}_t(Y^t)) dt \right] - \text{cmse}_{P,P}(\text{snr}) \right\}
$$

)*,* (29)

^P ²*^P* [cmse*P,Q*(snr) cmse*P,P* (snr)] (30)

(31)

: ⇥ is a *^P*-valued RV (32)

Z

Furthermore, the 'strong redundancy-capacity' redundancy-capacity r redundancy-capacity applicable here and imply:

$$
\text{Minimax}(P, \text{snr}) \stackrel{\triangle}{=} \underset{\{\hat{X}_t(\cdot)\}_{0\leq t\leq T}}{\min} \underset{P \in \mathcal{P}}{\max} \left\{ E_P \left[\int_0^T \ell(X_t, \hat{X}_t(Y^t)) dt \right] - \text{cms} e_{P,P}(\text{snr}) \right\}
$$

)*,* (29)

^P ²*^P* [cmse*P,Q*(snr) cmse*P,P* (snr)] (30)

(31)

: ⇥ is a *^P*-valued RV (32)

Z

classical

Furthermore, the 'strong redundancy-capacity' redundancy-capacity r redundancy-capacity applicable here and imply:

$$
\text{Minimax}(P, \text{snr}) \stackrel{\triangle}{=} \underset{\{\hat{X}_t(\cdot)\}_{0\leq t\leq T}}{\min} \underset{P \in \mathcal{P}}{\max} \left\{ E_P \left[\int_0^T \ell(X_t, \hat{X}_t(Y^t)) dt \right] - \text{cms} e_{P,P}(\text{snr}) \right\}
$$

)*,* (29)

^P ²*^P* [cmse*P,Q*(snr) cmse*P,P* (snr)] (30)

(31)

: ⇥ is a *^P*-valued RV (32)

Z

classical

ours

Furthermore, the 'strong redundancy-capacity' redundancy-capacity r redundancy-capacity applicable here and imply:

$$
\text{Minimax}(P, \text{snr}) \stackrel{\triangle}{=} \underset{\{\hat{X}_t(\cdot)\}_{0\leq t\leq T}}{\min} \underset{P \in \mathcal{P}}{\max} \left\{ E_P \left[\int_0^T \ell(X_t, \hat{X}_t(Y^t)) dt \right] - \text{cms} e_{P,P}(\text{snr}) \right\}
$$

)*,* (29)

^P ²*^P* [cmse*P,Q*(snr) cmse*P,P* (snr)] (30)

(31)

: ⇥ is a *^P*-valued RV (32)

Z

classical

ours

Furthermore, the 'strong redundancy-capacity' redundancy-capacity r redundancy-capacity applicable here and imply:

Redundancy-Capacity theory

$$
\text{Minimax} \left(\text{causal} \right) \text{Estimation}
$$
\n
$$
\text{minimax}(\mathcal{P}, \text{snr}) \triangleq \min_{\{\hat{X}_t(\cdot)\}_{0 \leq t \leq T}} \max_{P \in \mathcal{P}} \left\{ E_P \left[\int_0^T \ell(X_t, \hat{X}_t(Y^t)) dt \right] - \text{cmse}_{P,P}(\text{snr}) \right\}
$$

)*,* (29)

^P ²*^P* [cmse*P,Q*(snr) cmse*P,P* (snr)] (30)

(31)

: ⇥ is a *^P*-valued RV (32)

Z

classical

ours

Redundancy-Capacity theory

Shannon

Furthermore, the 'strong redundancy-capacity' redundancy-capacity r redundancy-capacity applicable here and imply:

⁸" *>* 0 and *any* filter *{X*ˆ*t*(*·*)*}*⁰*t^T* , Monday, October 22, 12

$$
\text{Minimax} \text{ (causal) Estimation} \\ \text{minimax}(\mathcal{P}, \text{snr}) \triangleq \min_{\{\hat{X}_t(\cdot)\}_{0 \leq t \leq T}} \max_{P \in \mathcal{P}} \left\{ E_P \left[\int_0^T \ell(X_t, \hat{X}_t(Y^t)) dt \right] - \text{cmse}_{P,P}(\text{snr}) \right\}
$$

)*,* (29)

)

)*,* (29)

Z

cmle*P,Q*() cmle*P,P* () = ¹

$$
\begin{array}{rcl}\n\text{classical} \\
\text{minimax}(\mathcal{P}, \text{snr}) & = & \min_{Q} \max_{P \in \mathcal{P}} \left[\text{cmse}_{P,Q}(\text{snr}) - \text{cmse}_{P,P}(\text{snr}) \right] \\
& = & \frac{2}{\text{snr}} \min_{Q} \max_{P \in \mathcal{P}} D\left(P_{Y_{\text{snr}}^T} \middle| Q_{Y_{\text{snr}}^T} \right) \\
\text{Redundancy-Capacity theory} & = & \frac{2}{\text{snr}} \max \left\{ I\left(\Theta; Y_{\text{snr}}^T\right) : \Theta \text{ is a } \mathcal{P}\text{-valued RV} \right\} \\
& = & \frac{2}{\text{snr}} C\left(\left\{ P_{Y_{\text{snr}}^T} \right\}_{P \in \mathcal{P}} \right)\n\end{array}
$$

Strong Converse Furthermore, the 'strong redundancy-capacity' results are directly applicable here and imply: = 2 *C*(*P,*snr) (33) $\overline{\mathbf{G}}$ max *I* ⇥; *Y ^T* snr : ⇥ is a *^P*-valued RV (32) \mathbf{I}_q *C* ⇣ *P^Y ^T* **V**

P 2*P*

snrtide and the same of the same

Q^Y ^T

snrting and the control of

snr

Furthermore, the 'strong redundancy-capacity' redundancy-capacity r redundancy-capacity applicable here and imply:

Q

2

 $m_{\tilde{t}}$

snr

))*dt*⌥

^w⇥(*B*) ⇥ *^e ·* ²⇥*·C*(*P,*snr)

^w⇥(*B*) ⇥ *^e ·* ²⇥*·C*(*P,*snr)

, (35)

: is a *^P*-valued RV⌅ (32)

C(*P,*snr) (33)

cmle*P,P* (snr) ⇤ (1 ⇤) *·* minimax(*P,*snr) (34)

for all *P* ⌃ *P* with the possible exception of sources in a subset *B* ⌅ *P* where

for all *P* ⌃ *P* with the possible exception of sources in a subset *B* ⌅ *P* where

10

P \rightarrow *<i>P* \rightarrow *P* \rightarrow *P*

 \mathcal{L}

"strong redundancy-capacity" result of [Merhav and Feder 1995] applied here implies: \mathcal{F}_{rel} 6.5 strong red cap Furthermore, the 'strong redundancy-capacity' results are directly applicable here and imply:

$$
\forall \varepsilon > 0 \text{ and any filter } {\{\hat{X}_t(\cdot)\}_{0 \le t \le T}},
$$

$$
E_P\left[\int_0^T \ell(X_t, \hat{X}_t(Y^t))dt\right] - \text{cmse}_{P,P}(\text{snr}) \geq (1-\varepsilon) \cdot \text{minimax}(\mathcal{P}, \text{snr})
$$

for all $P\in\mathcal{P}$ with the possible exception of sources in a subset $\mathcal{B}\subset\mathcal{P}$ where $\frac{1}{2}$ as **p** $\frac{1}{2}$ as *P* where

$$
w^*(\mathcal{B}) \leq e \cdot 2^{-\varepsilon \cdot \min\{\mathcal{P}, \mathsf{snr}\}},
$$

 $\frac{1}{\sqrt{2}}$ *w*⇥ being the capacity achieving prior

1
11
11 -

w
<u>w</u>

w
<u>w</u>

6.5 strong red cap

6.5 strong red cap

⇤ *>* 0 and *any* filter *{X*

Example This is essentially the best you can do not be the best you can do not be the best you can do not be the best
This is estentially the best you can do not be the best you can do not be the best you can do not be the best *h*(*Q*(*x*)) = ⇧ *h*(*Q*(*x*)) = *h*(*Q*(*x*)) = *x* λ *h*(*Q*(*x*)) =

⇧

i=1

Given: This is essentially the best you can do not be the best you can do not be the best you can do. THE EXEMPLE SERVE HET SEE

 ${\phi_i}$ 0 *thonormal* signal set ${\phi_i}$

 $\{\phi_i(t), 0 \le t \le T\}_{i=1}^n$ *i*=1 set $\{\phi_i(t), 0 \leq$

laws *^P* on *^X^T* : *^E^P* ⌅*B*⌅² ⇤ *ⁿ*⇥ and *^E^P* ⌅*B*⌅⁰ ⇤ *ⁿ*

I(*X^T* ; *Y ^T*) = *I*(*Bn*; *Y ⁿ*)

h(*Q*(*x*)) =

^Q(*xi*+1*|Xⁱ*

n

i=1

$$
X_t = \sum_{i=1}^n B_i \cdot \phi_i(t)
$$

 $S = \{ \text{laws } P \text{ on } X^T : E_P ||B||^2 \leq n\beta \text{ and } E_P ||B||_0 \leq n\alpha \}$ *^P* ⁼ ⇤ laws *^P* on *^X^T* : *^E^P* ⌅*B*⌅² ⇤ *ⁿ*⇥ and *^E^P* ⌅*B*⌅⁰ ⇤ *ⁿ* $\mathcal{P} = \left\{ \text{laws } P \text{ on } X^T : E_P ||B||^2 \leq n\beta \text{ and } E_P ||B||_0 \leq n\alpha \right\}$ $x \int$

$$
\max I(X^T;Y^T) = ?
$$

Given:

since

orthonormal signal set

Example (cont.) *Bⁱ ·* ⇤*i*(*t*) laws *^P* on *^X^T* : *^E^P* ⌅*B*⌅² ⇤ *ⁿ*⇥ and *^E^P* ⌅*B*⌅⁰ ⇤ *ⁿ*

^X^t ⁼ ⇧

$$
Y_i = \int_0^T \phi_i(t) dY_t \quad 1 \le i \le n
$$

are sufficient statistics for Y^T , are sucient statistics for *Y ^T* , *I*(*X^T* ; *Y ^T*) = *I*(*Bn*; *Y ⁿ*)

I(*X^T* ; *Y ^T*) = *I*(*Bn*; *Y ⁿ*)

$$
I(X^T;Y^T) = I(B^n;Y^n)
$$

 $\mathcal{L}(X,Y) = \mathcal{L}(X,Y) =$ $=\frac{m}{2}$ $\max I(X^T; Y^T) = \max I(B^n; Y^n) = \max \{I(B; Y) : B^2 \le \beta, P(B = 0) \ge (1 - \alpha)\}$

max *^I*(*X^T* ; *^Y ^T*) = max *^I*(*Bn*; *^Y ⁿ*) = max*{I*(*B*; *^Y*) : *^B*² ⇥ ⇥*, P*(*^B* = 0) ⇤ (1)*}* max *^I*(*X^T* ; *^Y ^T*) = max *^I*(*Bn*; *^Y ⁿ*) = max*{I*(*B*; *^Y*) : *^B*² ⇥ ⇥*, P*(*^B* = 0) ⇤ (1)*} I*(*X*) = *Matter* considered and numerically solved in: latter considered and numerically solved in:

 \Longrightarrow

 \overline{a}

 \Longrightarrow

I(*X^T* ; *Y ^T*) = *I*(*Bn*; *Y ⁿ*)

latter solved by and Power Constraints", IEEE Int. Symposium on Information Theory 2011 Lei Zhang and Dongning Guo, "Capacity of Gaussian Channels with Duty Cycle

since

are sucient statistics for *Y ^T*

are sucient statistics for *Y ^T*

are sucient statistics for *Y ^T*

Example (cont.) \blacksquare are sucient statistics for *Y ^T* , *I*(*X^T* ; *Y ^T*) = *I*(*Bn*; *Y ⁿ*)

I(*X^T* ; *Y ^T*) = *I*(*Bn*; *Y ⁿ*)

thus the minimax filter here is the Bayes filter assuming:

$$
X_t = \sum_{i=1}^n B_i^* \cdot \phi_i(t)
$$

ⁱ are iid according to the capacity achieving distribution of [Zhang and Guo, 2011]

where B_i^* are iid according to the capacity achieving distribution of [Zhang and Guo, 2011]

cf. [Albert No + T.W., ISIT 2013]...

(well) beyond Gaussian noise

- Poisson channel
- Lévy-type channels:
	- Input-Output relationship expressed via Lévy-

type stochastic integral

• can obtain formulae via Lévy-Khintchine-type decompositions

\checkmark information

control

networks

What if *X* ⇠ *P* but the estimator thinks *X* ⇠ *Q* ?

The presence of Feedback

The presence of Feedback

• what of what we've seen carries over to presence of feedback?

Duncan Duncan slide 1 for Duncan slides in the Duncan slides in the Duncan slides in the Duncan slides in the Duncan slides in the
1 for Duncan slides in the Duncan slides in the Duncan slides in the Duncan slides in the Duncan slides in the

$$
dY_t = X_t dt + dW_t, \qquad 0 \le t \le T
$$

 W is standard white Gaussian noise, independent of X

^I(*X^T* ; *^Y ^T*) = ¹ *E* $\frac{1}{2}$ $\left[\frac{X}{X}\right]$: *[Duncan 1970]:*

$$
I(X^T; Y^T) = \frac{1}{2}E\left[\int_0^T (X_t - E[X_t|Y^t])^2 dt\right]
$$

(*X^t ^E*[*Xt|^Y ^t*

^I(*X^T* ; *^Y ^T*) = ¹

W is white Gaussian noise, independent of *X*

])²*dt*⇥

⇤ *^T*

^I(*X^T* ; *^Y ^T*) = ¹

(*X^t ^E*[*Xt|^Y ^t*

(*X^t ^E*[*Xt|^Y ^t*

(*X^t ^E*[*Xt|^Y ^t*

dY^t ⁼ ⇧*Xtdt* ⁺ *dWt,* ⁰ ⇥ *^t* ⇥ *^T*

⇤ *^T* Breaks down in presence of feedback! cmmse() = ¹

E

cmmse() = ¹

1 for Duncan slide

W is white Gaussian noise, independent of *X*

cont time directed info The following definition is now natural: and *pointwise universal* if

length 2 in this case) is upper bounded by the (conditional) mutual information [1, Thm 2].

[W., Permuter, Kim 2012] $\overline{1}$ *n* log *^PXⁿ* (*Xn*) *^QXⁿ* (*Xn*) ⁰ *^P a.s.*

I $(X_0^T \rightarrow Y_0^T$ \setminus $:=$ inf t *I*t $(X_0^T \rightarrow Y_0^T$ $\bigg)$ where the infimum is over all *n* and t as in (6). where $I(X_0^1 \rightarrow Y_0^1) := \inf$ $where$

⁰) be a pair of jointly distributed stochastic processes. The *Directed Information between*

1

, (17)

) = *^E* [log *^P*(*^Y ⁿ*k*Xn*)] (2)

$$
I(X^n \to Y^n) \triangleq \sum_{i=1}^n I(X^i; Y_i | Y^{i-1})
$$

where *^H*(*^Y ⁿ*k*Xn*) is the *causally conditional entropy*

^H(*^Y ⁿ*k*Xn*) , ^X

is alway well defined as an extended non-negative real number (i.e., as an element of [0*,* 1]). It is also worth noting,

by recalling (4), that each of the conditional mutual informations in (9), and hence the sum, is a supremum over

^H(*Yi|^Y ⁱ*1*, Xⁱ*

Duncan with feedback Definition 2.25]). Recall also that when the standard Brownian motion is adapted to *{Ft}* then, by definition, it is **B** DUILCALL WILL IEEUDACK Duncan's theorem. To state it formally we assume a probability space (⌦*, F, P*) with an associated filtration *{Ft}* satisfying the "usual conditions" (right-continuous and *F*⁰ contains all the *P*-negligible events in *F*, cf., e.g., [?, implied that, for any *s<t*, *^B^t ^B^s* is independent of *^F^s* (rather than merely of *^B^s* ⁰, cf., e.g., [?, Definition 1.1]).

Definition 2.25]). Recall also that when the standard Brownian motion is adapted to *{Ft}* then, by definition, it is

satisfying the "usual conditions" (right-continuous and *F*⁰ contains all the *P*-negligible events in *F*, cf., e.g., [?,

Note that unlike in Theorem 1, where the channel input process is independent of the channel noise process,

in Theorem 2 no such stipulation exists and thus the setting in the latter accommodates the presence of feedback.

⁰) is not invariant to the direction of the flow of time in general, Theorem 2 implies,

⁰ *is a signal of finite average power*

⁰ ! *^Y ^T*

⁰ *is a signal of finite average power*

⁰ *be the output of the AWGN channel whose*

Theorem 2. *sdjhd* implied that, for any *s<t*, *^B^t ^B^s* is independent of *^F^s* (rather than merely of *^B^s ^t*=0 *be adapted to the filtration {Ft}^T* [W., Permuter, Kim 2012]

 $L(X, B_1)$ ^T be adapted to the filtration $\int F \cdot T$ where X^T is a signal of finite gyerges power $\int_0^T E[X_t^2]dt < \infty$ and B_0^T is a standard Brownian motion. Let Y_0^T be the output of the AWGN channel whose ⁰ *be the output of the AWGN channel whose i*₁ is X_0 and whose noise is driven by B_0 , i.e., Let $\{(X_t,B_t)\}_{t=0}^T$ be adapted to the filtration $\{\mathcal{F}_t\}_{t=0}^T$, where X_0^T is a signal of finite average power *input is* X_0^T *and whose noise is driven by* B_0^T *, i.e.,* Let $\{(X_t, B_t)\}_{t=0}^T$ be adapted to the filtricity $\int_{0}^{T} E[X_t^2] dt < \infty$ and B_t^T is a standard Brow

$$
dY_t = X_t dt + dB_t.
$$

Suppose that the regularity assumptions of Proposition 2 are satisfied for all $0 < t < T$. Then

⁰ *is a standard Brownian motion. Let Y ^T*

$$
\frac{1}{2} \int_0^T E[(X_t - E[X_t|Y_0^t])^2] dt = I(X_0^T \to Y_0^T)
$$

$a₂$ \overline{B} $\overline{$ compare with $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ α s should be expected is the causal MMSE for processes α note that under the interest in the compare with IK adota 7 akai 7 iv 19711 compare with [Kadota, Zakai, Ziv 1971]

communication theoretic framework by Kadota, Zakai, and Ziv [22]. Indeed, in communication over the AWGN

in Theorem 2 no such stipulation exists and thus the setting in the latter accommodates the presence of feedback.

^t]*dt <* ¹ *and ^B^T*

Furthermore, since *I*(*X^T*

GSV in continuous time **I**CC ⇤ *^T* (*X^t ^E*[*Xt|^Y ^t* me

dY^t = *Xtdt* + *dWt,* 0 ⇥ *t* ⇥ *T*

(*^X ^E*[*X|^Y*])²⇥

mmse() = *E*

^Y ⁼ ⇤ *· ^X* ⁺ *^W*

I() = *I*(*X*; *Y*)

⇧ snr

(*^X ^E*[*X|^Y*])²⇥

(*X^t ^E*[*Xt|^Y ^T*])²*dt*

$$
dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T
$$

$$
\frac{d}{d\gamma}I(\gamma)=\frac{1}{2}\mathsf{mmse}(\gamma)
$$

 $\overline{\mathbf{c}}$ $\mathbf n$.
E
E \sim (*X^t ^E*[*Xt|^Y ^t* or in its integral version

22
22 March 2014
22 March 2014

relationship holds regardless of the input distribution, and carries over essentially verbatim to vectors, as well as the

is equal to the minimum mean square error (MMSE) in estimating the input based on the output. This simple

$$
I(\mathsf{snr}) = \frac{1}{2} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma
$$

W is white Gaussian noise, independent of *X*

[Guo, Shamai and Verd´u 2005]:

[Guo, Shamai and Verd´u 2005]:

GSV in continuous time **I**CC ⇤ *^T* (*X^t ^E*[*Xt|^Y ^t* me

dY^t = *Xtdt* + *dWt,* 0 ⇥ *t* ⇥ *T*

(*^X ^E*[*X|^Y*])²⇥

mmse() = *E*

^Y ⁼ ⇤ *· ^X* ⁺ *^W*

I() = *I*(*X*; *Y*)

⇧ snr

(*^X ^E*[*X|^Y*])²⇥

(*X^t ^E*[*Xt|^Y ^T*])²*dt*

$$
dY_t = \sqrt{\gamma} X_t dt + dW_t, \qquad 0 \le t \le T
$$

$$
\frac{d}{d\gamma}I(\gamma)=\frac{1}{2}\mathsf{mmse}(\gamma)
$$

 $\overline{\mathbf{c}}$ $\mathbf n$.
E
E \sim (*X^t ^E*[*Xt|^Y ^t* or in its integral version

$$
I(\mathsf{snr}) = \frac{1}{2} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma
$$

n in presence of is equal to the minimum mean space of feedback in each square error (MMSE) in each on the input based on the o relationship holds regardless of the input distribution, and carries over essentially verbatim to vectors, as well as the Breaks down in presence of feedback

22
22 March 2014
22 March 2014

W is white Gaussian noise, independent of *X*

[Guo, Shamai and Verd´u 2005]:

[Guo, Shamai and Verd´u 2005]:

GSV in continuous time with DI? X *i*=1 *i D*^{(*t*}) \overline{a} with DI? *I*() = *I*(*X*; *Y*) [Guo, Shamai and Verd´u 2005]:

^Y ⁼ ⌥ *· ^X* ⁺ *^W*

ⁱ are iid according to the capacity achieving distribution of [Zhang and Guo, 2011]

$$
I(X^T \to Y^T) \stackrel{?}{=} \frac{1}{2} \int_0^{\textsf{snr}} \textsf{mmse}(\gamma) \mathsf{d}\gamma
$$

No. In general \overline{a} $\overline{}$ eral

$$
I(X^T \to Y^T) \neq \frac{1}{2} \int_0^{\text{snr}} \text{mmse}(\gamma) d\gamma
$$

and so

$$
\mathsf{cmmse}(\mathsf{snr}) \neq \frac{1}{\mathsf{snr}} \int_0^{\mathsf{snr}} \mathsf{mmse}(\gamma) d\gamma
$$

l.e., breakdown in presence of feedback

Relationship between cmmse and mmse?

where *B*

or in its integral version

[Guo, Shamai and Verd´u 2008]

W is a standard Gaussian, independent of *X*

latter solved by

Mismatched setting Mismatched mse*P,Q*()*d*

(*X^t ^EQ*[*Xt|^Y ^T*])²*dt*⌅

mse*P,Q*() = *E^P*

a fortiori, in presence of feedback, in general snr [*I*(snr) + *^D* (*P^Y ^T* ⇧*Q^Y ^T*)]

$$
\text{cmse}_{P,Q}(\textsf{snr}) \quad \neq \quad \frac{1}{\textsf{snr}} \int_{0}^{\textsf{snr}} \textsf{mse}_{P,Q}(\gamma) d\gamma
$$

Relationship between cmse*P,Q* and mse*P,Q* ?

Relationship between cmse*P,Q* and mse*P,Q*

Mismatched setting Mismatched mse*P,Q*()*d*

(*X^t ^EQ*[*Xt|^Y ^T*])²*dt*⌅

mse*P,Q*() = *E^P*

a fortiori, in presence of feedback, in general snr [*I*(snr) + *^D* (*P^Y ^T* ⇧*Q^Y ^T*)]

$$
\text{cmse}_{P,Q}(\textsf{snr}) \quad \neq \quad \frac{1}{\textsf{snr}} \int_{0}^{\textsf{snr}} \textsf{mse}_{P,Q}(\gamma) d\gamma
$$

end of story?

Relationship between cmse*P,Q* and mse*P,Q* ?

Relationship between cmse*P,Q* and mse*P,Q*

Mismatched setting (cont.) *D*(*P^Y ^T Q^Y ^T*) ⌥ cmle*P,Q* cmle*P,P* (28)

$\mathsf{cmse}_{P,Q} - \mathsf{cmse}_{P,P} = D(P_{Y^T} || Q_{Y^T})$

*^d*law of homogenous Poisson as a filtering integral

*^d*law of homogenous Poisson

D(*P^Y ^T Q^Y ^T*) = *E^P* V
i \/ orksnop book chapter
enkat, W. 2012] *d* and **defining the set of** \mathbb{Z} **and** \mathbb{Z} **and** \mathbb{Z} **and** \mathbb{Z} **are** \mathbb{Z} **and** \mathbb{Z} **and** \mathbb{Z} **are** \mathbb{Z} **and** \mathbb{Z} **are** \mathbb{Z} **and** \mathbb{Z} **are** \mathbb{Z} **and** \mathbb{Z} **are** \mathbb{Z} **and** \mathbb{Z} **ar** holds with or without FB, appears in TW2010 implicitly and explicitly in workshop book chapter [Asnani, Venkat, W. 2012]

 (why?) log *dQ^Y ^T*

• Girsanov-type theory for expressing log *dQ^Y ^T*

Theorem 6.5 *(under mild conditions)*

implications and apps

- minimax estimation setting carries over
- directed info maximization instead of mutual info but same idea
- similar extensions to the more general channels

\checkmark information $\mathbf v$

v control

networks

Relationship between cmmse and mmse?

Relationship between cmmse and mmse?

What if *X* ⇠ *P* but the estimator thinks *X* ⇠ *Q* ?

Distributed estimation (known source) *X Y*1*, X*ˆ1(*Y*1)

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^I(*X^T* ; *^Y ^T*) = ^Z *^T*

^E[*Xt|^Y ^t*

*X*1*, X*2*, X*3*,...,Xi*1*, Xi,... X*1*, X*2*, X*3*,...,Xi*1*, Xi,... Y*1*, Y*2*, Y*3*,...,Yi*1*, Yi,... X*1*, X*2*, X*3*,...,Xi*1*, Xi,... X*1*, X*2*, X*3*,...,Xi*1*, Xi,... Yn, X*ˆ*n*(*Yn*) α *can* (and should) be greedy!

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^I(*X^T* ; *^Y ^T*) = ^Z *^T*

^E[*Xt|^Y ^t*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^I(*X^T* ; *^Y ^T*) = ^Z *^T*

^E[*Xt|^Y ^t*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^I(*X^T* ; *^Y ^T*) = ^Z *^T*

^E[*Xt|^Y ^t*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^Y ^T [|]X^T is non-homogenous Poisson of intensity *^X^T*

^I(*X^T* ; *^Y ^T*) = ^Z *^T*

^E[*Xt|^Y ^t*

*X*1*, X*2*, X*3*,...,Xi*1*, Xi,...* Monday, October 22, 12

\checkmark information \vee ? $\sqrt{ }$

v control $\overline{1}$ $\sqrt{ }$

v networks

=?

Relationship between cmmse and mmse?

Relationship between cmmse and mmse?

Relationship between cmmse and mmse?

What if *X* ⇠ *P* but the estimator thinks *X* ⇠ *Q* ?

conclusion

- relations between mutual information, relative entropy, and estimation
- findings of pure estimation theoretic significance
- allow the transfer of tools
- much carries over to presence of feedback
- implications for networks