Finding the best mismatched detector for channel coding and hypothesis testing

Sean Meyn

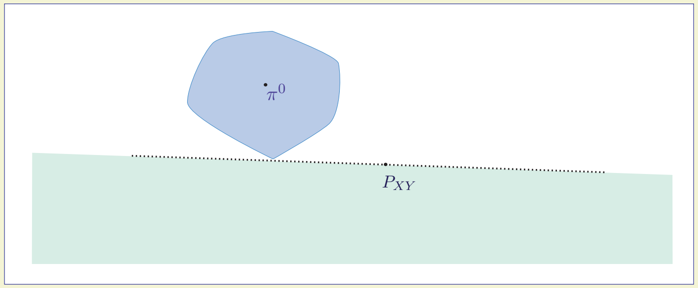
Department of Electrical and Computer Engineering
University of Illinois
and the Coordinated Science Laboratory

Joint work with Emmanuel Abbe, Muriel Me'dard, and Lizhong Zheng
Laboratory for Information and Decision Systems, MIT

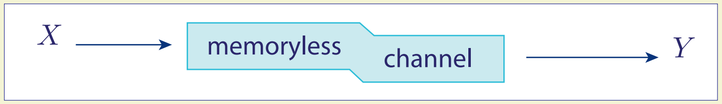
Supported in part by Motorola RPS #32 and ITMANET DARPA RK 2006-07284

Outline

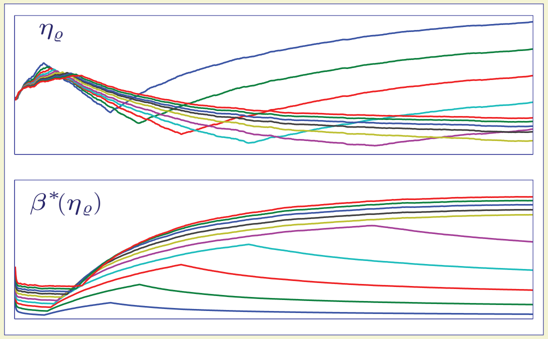
Introduction



Mismatched channel



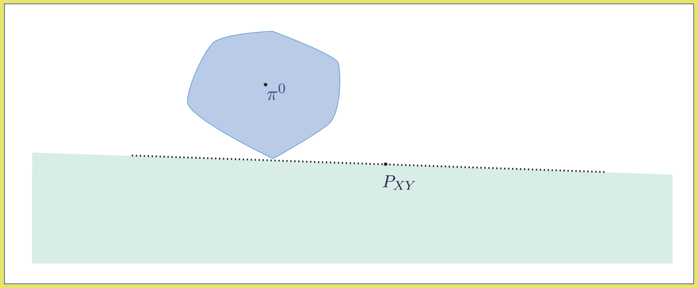
Learning the best detector



Conclusions

I

Introduction



Background: Mismatch Channel

$X \longrightarrow$ $\boxed{\text{memoryless channel}}$ $\longrightarrow Y$

Optimal detector ML: Based on the empirical distributions,

$$i^* = \arg\max_i \langle \Gamma_n^i, F \rangle$$

where, F is the log-likelihood ratio (real-valued function of x, y)

Γ_n^i is the empirical distribution of (X^i, Y)

Background: Mismatch Channel

$X \longrightarrow$ memoryless channel $\longrightarrow Y$

Optimal detector ML: Based on the empirical distributions,

$$i^* = \arg\max_i \langle \Gamma_n^i, F \rangle$$

where, F is the log-likelihood ratio

Γ_n^i is the empirical distribution of (X^i, Y)

Question considered by Lapidoth 1994, Csiszar et. al. 1981, 1995:

What if the channel model is not known exactly ?

Background: Robust Hypothesis Testing

Uncertainty classes defined by moment constraints $\mathbb{P}_0$, $\mathbb{P}_1$

For given false alarm exponent η,
min-max optimal missed detection exponent β^*

$$\beta^* = \inf_{\pi_1 \in \mathbb{P}_1} \inf_{\mu \in \mathcal{Q}_\eta(\mathbb{P}_0)} D(\mu \,\|\, \pi_1)$$

Meyn, Pandit, Veeravalli 200X

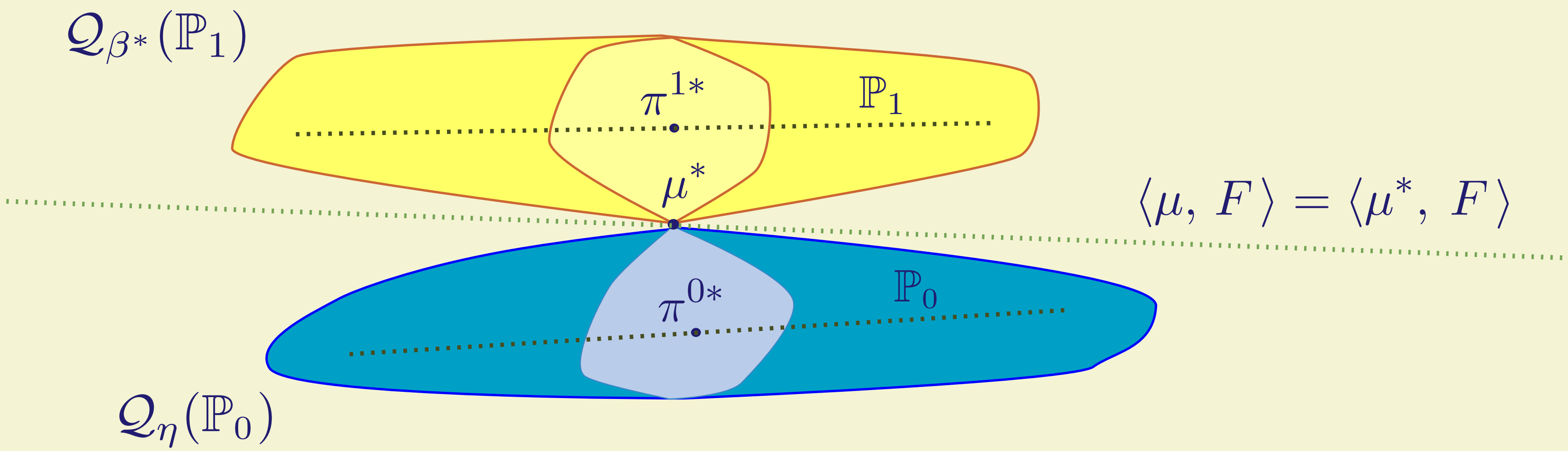
Background: Robust Hypothesis Testing

Uncertainty classes defined by moment constraints $\mathbb{P}_0\,,\ \mathbb{P}_1$

For given false alarm exponent η,

min-max optimal missed detection exponent β^*

$$\beta^* = \inf_{\pi_1 \in \mathbb{P}_1} \ \inf_{\mu \in \mathcal{Q}_\eta(\mathbb{P}_0)} D(\mu \,\|\, \pi_1)$$



F : Generalized Log Likelihood Ratio

Meyn, Pandit, Veeravalli 200X

Goal: Robustness

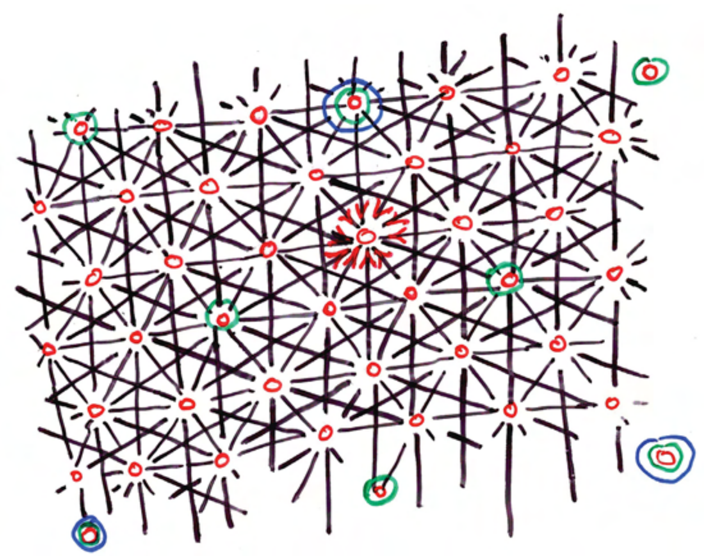
Original motivation for the research on mismatched detectors

What is the impact on capacity given a poor model?
How should codes be constructed
to take into account uncertainty?

Goal: Complexity

Detection and capacity - easy. *Its just mutual information*!!

Linear detectors suggest relaxation techniques for multiuser and network settings



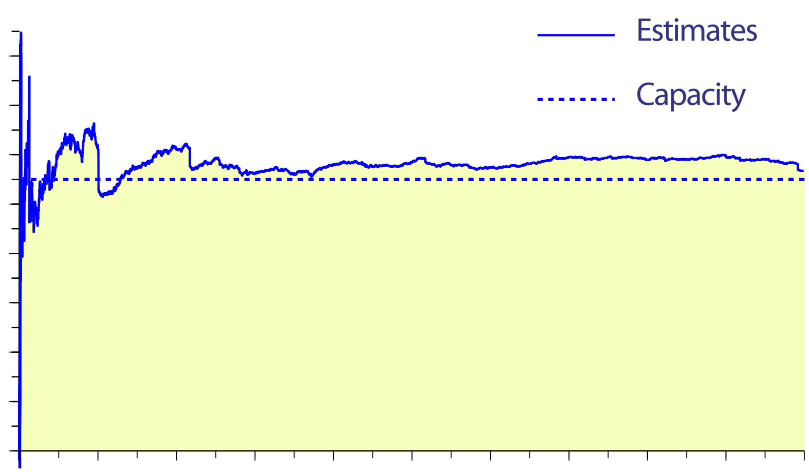
Goal: Adaptation

A new point of view!

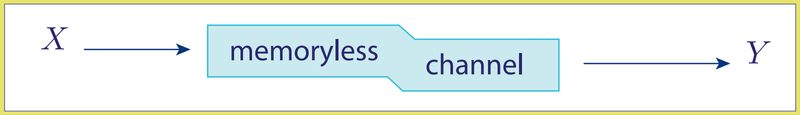
How can a detector be tuned on-line
in a dynamic environment?

$X \longrightarrow$ memoryless channel $\longrightarrow Y$

This will depend on SNR
Other users
Network conditions ...



II
Mismatched Channel



Mismatched Detector

A given function $F(x,y)$ defines a surrogate ML detector:

$$i^* = \arg\max_i \langle \Gamma_n^i, F \rangle = \arg\max_i \Big\{ \sum_{t=1}^{n} F(X_t^i, Y_t) \Big\}$$

Empirical distributions for ith codeword:

$$\Gamma_n^i = n^{-1} \sum_{t=1}^{n} \delta_{X_t^i, Y_t}$$

Reliably received rate (in general a lower bound):

Generalized Mutual Information (Lapidoth, Csiszar)

Mismatched Detector: Sanov's Theorem

Sanov's Theorem (in form of Chernoff's Bound) gives

$$\log P_e \leq -nI_{\mathrm{LDP}}(P_X; F)$$ (P_e *pair-wise prob. of error*)

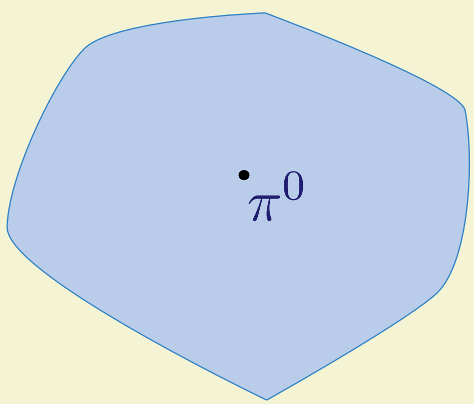
$$I_{\mathrm{LDP}}(P_X; F) := \inf\Big\{D(\Gamma \| P_X \otimes P_Y) : \langle \Gamma, F \rangle = \langle P_{XY}, F \rangle\Big\}$$

Mismatched Detector: Sanov's Theorem

Sanov's Theorem (in form of Chernoff's Bound) gives

$$\log P_e \leq -nI_{\mathrm{LDP}}(P_X; F) \qquad (P_e \textit{ pair-wise prob. of error})$$

$$I_{\mathrm{LDP}}(P_X; F) := \inf\Big\{ D(\Gamma \| P_X \otimes P_Y) : \langle \Gamma, F \rangle = \langle P_{XY}, F \rangle \Big\}$$



$$\mathcal{Q}_{\beta_0}(\pi^0) = \{ \Gamma : D(\Gamma \,\|\, \pi^0) \leq \beta_0 \}$$

$$\pi^0 = P_X \otimes P_Y$$

Mismatched Detector: Sanov's Theorem

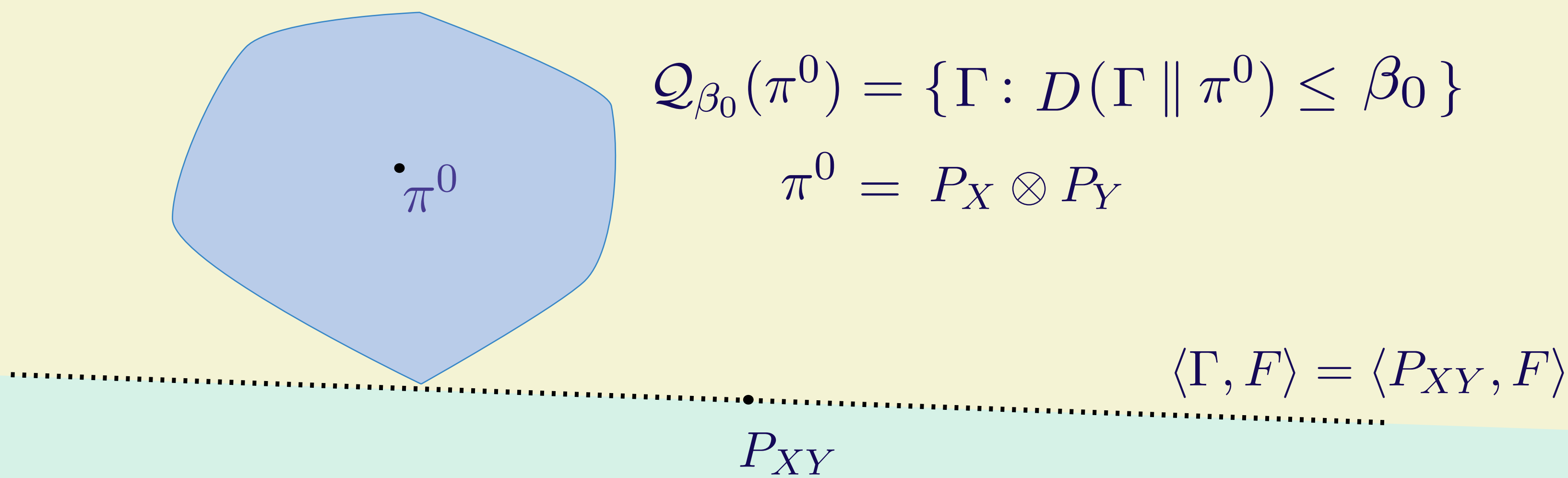
Sanov's Theorem (in form of Chernoff's Bound) gives

$$\log P_e \le -n I_{\mathrm{LDP}}(P_X; F)$$

$$I_{\mathrm{LDP}}(P_X; F) := \inf\Big\{ D(\Gamma \| P_X \otimes P_Y) : \langle \Gamma, F\rangle = \langle P_{XY}, F\rangle \Big\}$$

$$\mathcal{Q}_{\beta_0}(\pi^0) = \{\Gamma : D(\Gamma \,\|\, \pi^0) \le \beta_0\}$$

$$\pi^0 = P_X \otimes P_Y$$



$$\beta_0 = I_{\mathrm{LDP}}(P_X; F)$$

Mismatched Detector: Generalized Mutual Information

Generalized Mutual Information - introduces typicality constraint,

$$\boxed{\log P_e \leq -nI_{\mathrm{GMI}}(P_X;F)}$$

$$I_{\mathrm{GMI}}(P_X;F) := \inf\Big\{D(\Gamma\|P_X\otimes P_Y) : \langle\Gamma,F\rangle = \langle P_{XY},F\rangle\,,\,\Gamma_2 = P_Y\Big\}$$

Mismatched Detector: Generalized Mutual Information

Generalized Mutual Information

$$\log P_e \leq -nI_{\mathrm{GMI}}(P_X; F)$$

$$I_{\mathrm{GMI}}(P_X; F) := \inf\Big\{D(\Gamma \| P_X \otimes P_Y) : \langle \Gamma, F \rangle = \langle P_{XY}, F \rangle \,,\, \Gamma_2 = P_Y\Big\}$$

Derivation of GMI bound using Sanov:

Mismatched Detector: Generalized Mutual Information

Generalized Mutual Information

$$\log P_e \leq -nI_{\mathrm{GMI}}(P_X; F)$$

$$I_{\mathrm{GMI}}(P_X; F) := \inf\Big\{D(\Gamma \| P_X \otimes P_Y) : \langle \Gamma, F \rangle = \langle P_{XY}, F \rangle \,,\, \Gamma_2 = P_Y\Big\}$$

Derivation of GMI bound using Sanov: For any function $G(y)$,

$$i^* = \arg\max_i \langle \Gamma_n^i, F + G \rangle = \arg\max_i \Big\{ \sum_{t=1}^{n} [F(X_t^i, Y_t) + G(Y_t)] \Big\}$$

$$\log P_e \leq -nI_{\mathrm{LDP}}(P_X; F + G) \quad \textit{any } G(y)$$

Mismatched Detector: Generalized Mutual Information

Generalized Mutual Information

$$\log P_e \leq -n I_{\mathrm{GMI}}(P_X; F)$$

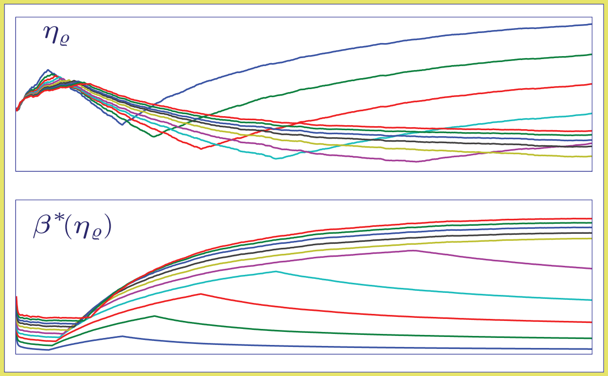
$$I_{\mathrm{GMI}}(P_X; F) := \inf\Big\{D(\Gamma \| P_X \otimes P_Y) : \langle \Gamma, F \rangle = \langle P_{XY}, F \rangle \,,\, \Gamma_2 = P_Y \Big\}$$

$$\log P_e \leq -n I_{\mathrm{LDP}}(P_X; F + G\,)$$ *any* $G(y)$

$$\sup_{G(y)} I_{\mathrm{LDP}}(P_X; F + G) = I_{\mathrm{GMI}}(P_X; F)$$

G^* : Lagrange multiplier for equality constraint $\Gamma_2 = P_Y$

III
Learning the Best Detector



Learning the Best Test

Given a parameterized class of detectors,

$$\{F_\alpha = \sum \alpha_i \psi_i : \alpha \in \mathbb{R}^d\}$$

What is the best value of α?

Maximize
MM capacity: $J(\alpha) := I_{\mathrm{LDP}}(P_X\,; F_\alpha)$

Learning the Best Test

Maximize

MM capacity: $J(\alpha) := I_{\mathrm{LDP}}(P_X; F_\alpha)$

Derivatives: $\nabla J = \bar{\psi} - \langle \Gamma^\alpha, \psi \rangle$

$$\nabla^2 J = \langle \Gamma^\alpha, \psi\psi^T \rangle - \langle \Gamma^\alpha, \psi \rangle \langle \Gamma^\alpha, \psi \rangle^T$$

where for any set $A \in \mathcal{B}(\mathsf{X} \times \mathsf{Y})$,

$$\Gamma^\alpha(A) := \frac{\langle P_X \otimes P_Y, e^{F_\alpha} \mathbb{I}_A \rangle}{\langle P_X \otimes P_Y, e^{F_\alpha} \rangle}$$

Learning the Best Test

Maximize
MM capacity: $J(\alpha) := I_{\mathrm{LDP}}(P_X; F_\alpha)$

Derivatives: $\nabla J = \bar{\psi} - \langle \Gamma^\alpha, \psi \rangle$

$$\nabla^2 J = \langle \Gamma^\alpha, \psi\psi^T \rangle - \langle \Gamma^\alpha, \psi \rangle \langle \Gamma^\alpha, \psi \rangle^T$$

where for any set $A \in \mathcal{B}(\mathsf{X} \times \mathsf{Y})$,

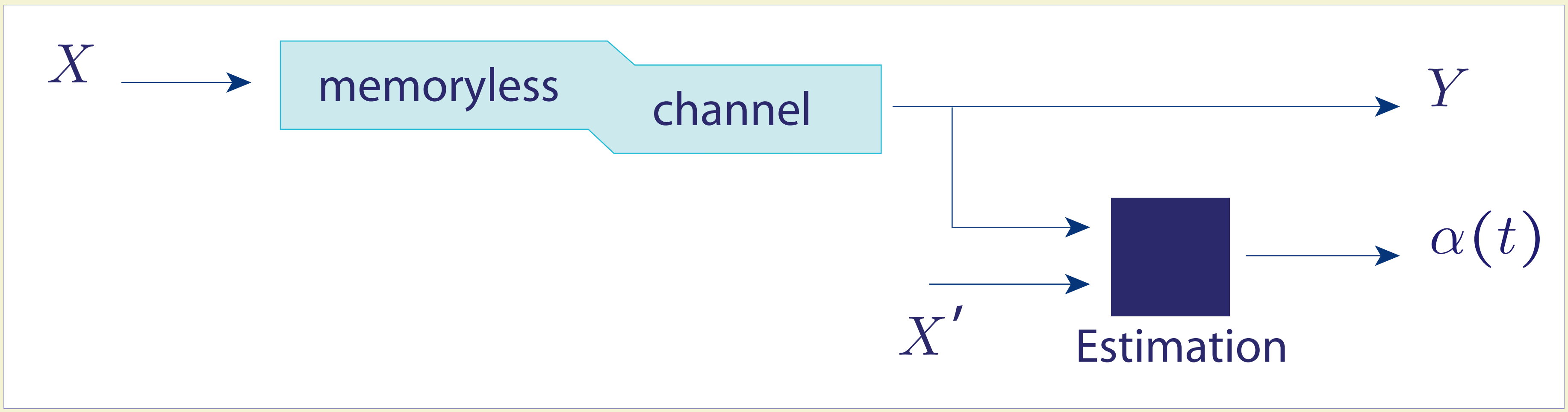
$$\Gamma^\alpha(A) := \frac{\langle P_X \otimes P_Y, e^{F_\alpha} \mathbb{I}_A \rangle}{\langle P_X \otimes P_Y, e^{F_\alpha} \rangle}$$

Newton Raphson based on channel model, or

Stochastic Newton Raphson based on measurements alone!

Learning the Best Test

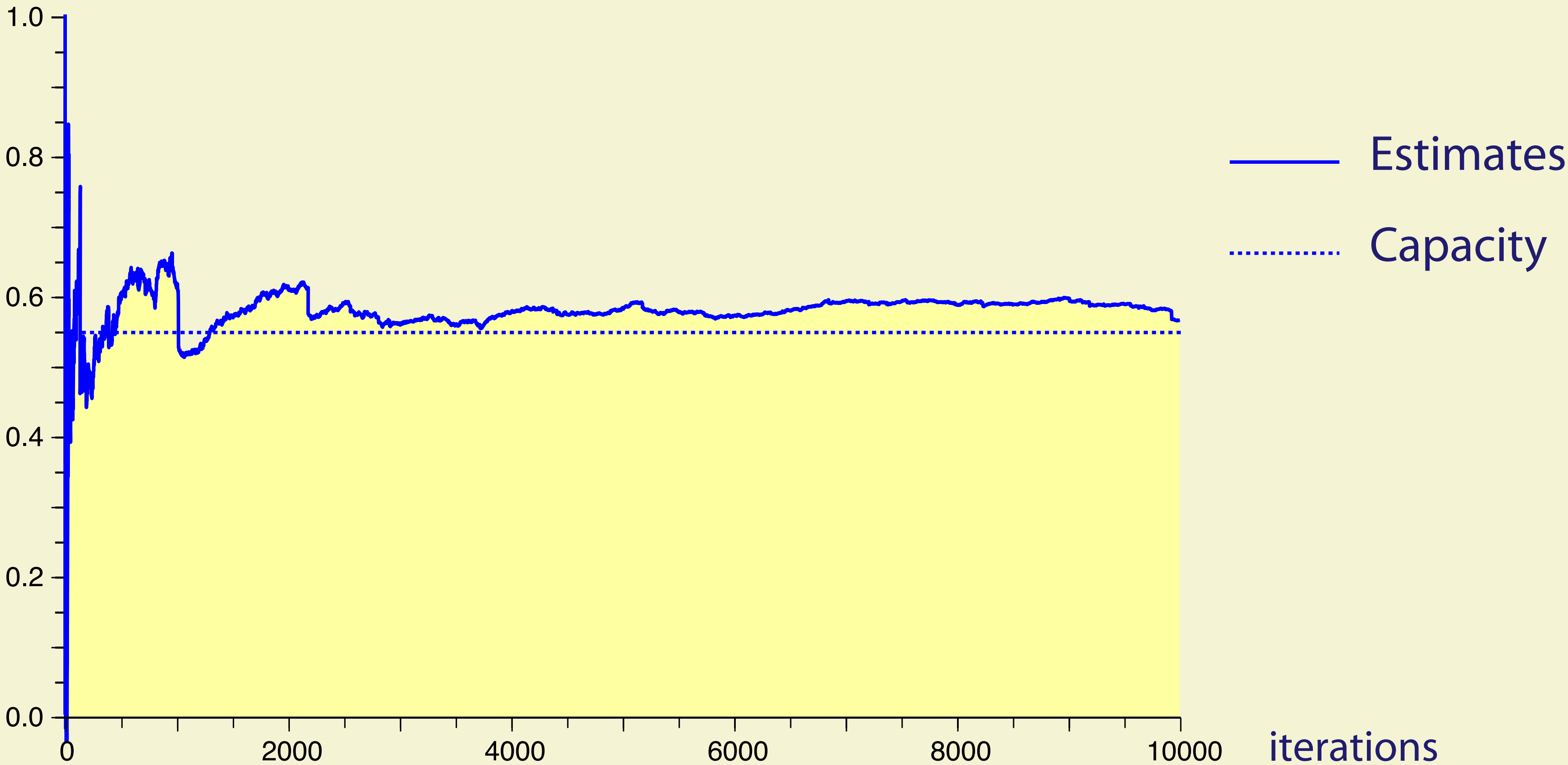
Estimation algorithm is *decentralized* and *model-free*



AWGN Channel

$Y = X + N$ SNR = 2

Basis: $\psi = (x^2, y^2, xy)^{\mathrm{T}}$



Mismatched Neyman Pearson Hypothesis Testing

Numerical experiment: X^1 uniform on $[0,1]$, $X^0 = \sqrt[5]{X^1}$

η *False Alarm Error Exponent*

$\beta^*(\eta)$ *Missed Detection Error Exponent*

Stochastic Newton-Raphson to find [η, $\beta^*(\eta)$] pairs

Parameterized class of detectors,

$$\{F_\alpha = \sum \alpha_i \psi_i : \alpha \in \mathbb{R}^d\}$$

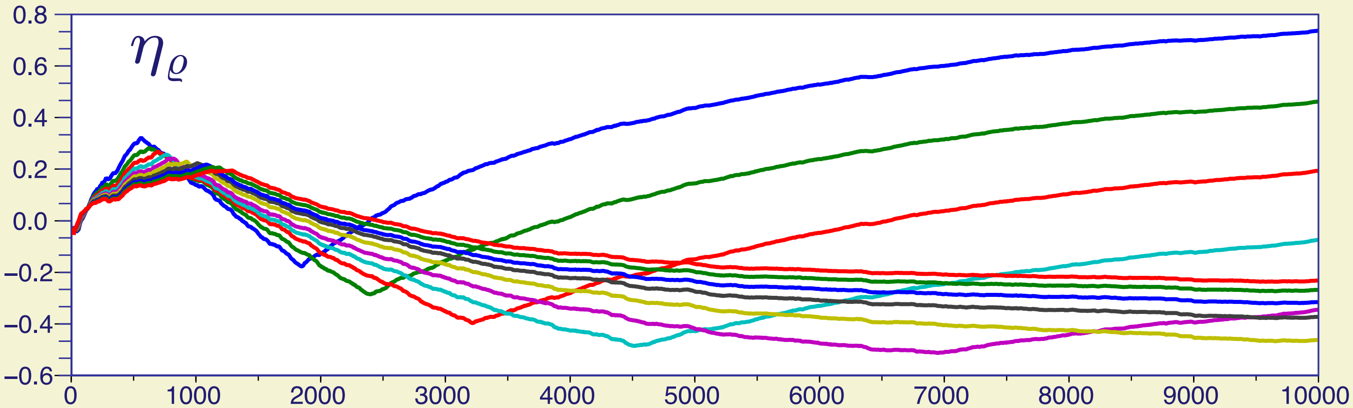
$$\psi(x) = [x, \log(x)/10]^T$$

Optimal detector

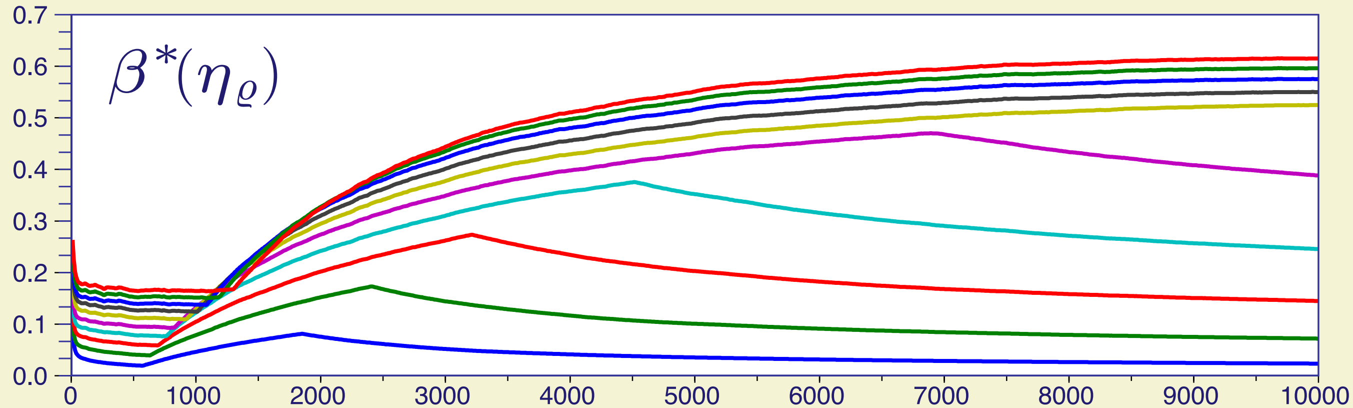
Mismatched Neyman Pearson Hypothesis Testing

X^1 uniform on $[0,1]$, $\quad X^0 = \sqrt[5]{X^1}$

Stochastic Newton-Raphson



False Alarm Error Exponent

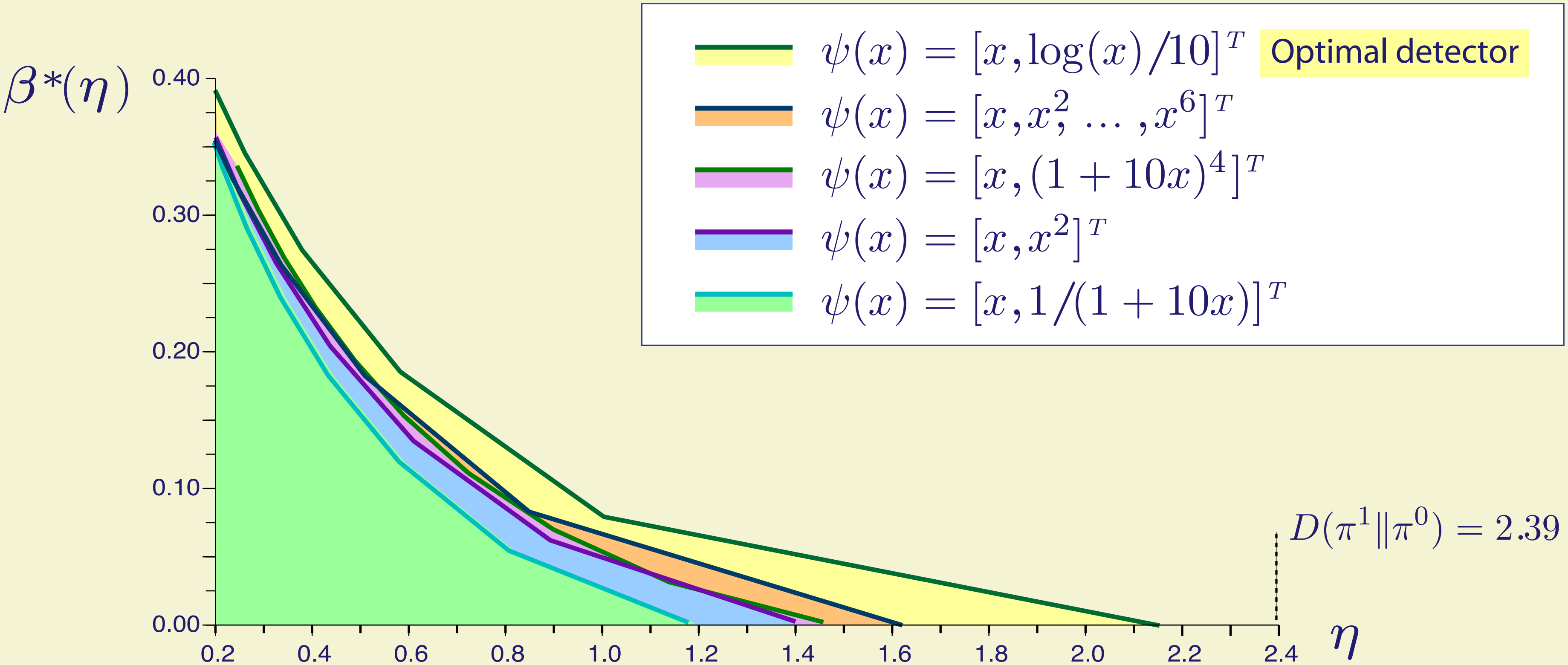


Missed Detection Error Exponent

Parallel simulation for a range of η

Mismatched Neyman Pearson Hypothesis Testing

X^1 uniform on $[0,1]$, $\quad X^0 = \sqrt[5]{X^1}$



Parallel simulation for a range of η

IV
Conclusions

Conclusions

Geometry based on Sanov’s Theorem combined with Stochastic Approximation provides powerful computational tools

Conclusions

Geometry based on Sanov's Theorem combined with Stochastic Approximation provides powerful computational tools

Future work:

Optimizing the input distribution

Application to MIMO will simultaneously resolve coding and resource allocation. *Extensions to network coding possible*?

Simulation algorithm exhibits high variance.
Application of importance sampling? *Revisit robust approach*?

References

[1] E. Abbe, M. Médard, S. Meyn, and L. Zheng. Geometry of mismatched decoders. Submitted to IEEE International Symposium on Information Theory, 2007.

[3] S. Borade and L. Zheng. I-projection and the geometry of error exponents. In *Proceedings of the Forty-Fourth Annual Allerton Conference on Communication, Control, and Computing, Sept 27-29, 2006*, UIUC, Illinois, USA, 2006.

[4] I. Csiszár. The method of types. *IEEE Trans. Inform. Theory*, 44(6):2505–2523, 1998. Information theory: 1948–1998.

[7] I. Csiszár and P. Narayan. Channel capacity for a given decoding metric. *IEEE Trans. Inform. Theory*, 41(1):35–43, 1995.

[12] G. Kaplan, A. Lapidoth, S. Shamai, and N. Merhav. On information rates for mismatched decoders. *IEEE Trans. Inform. Theory*, 40(6):1953–1967, 1994.

[14] C. Pandit and S. P. Meyn. Worst-case large-deviations with application to queueing and information theory. *Stoch. Proc. Applns.*, 116(5):724–756, May 2006.

[15] O. Zeitouni and M. Gutman. On universal hypotheses testing via large deviations. *IEEE Trans. Inform. Theory*, 37(2):285–290, 1991.