ISIT 2020

Approaching Capacity Without Pilots via Nonlinear Processing at the Edge

Guido Carlo Ferrante

Ericsson Research, Stockholm

Motivation and Goal

Design a communication scheme

- completely decentralized (indep. transmissions, no rx cooper.)
- optimal at high SNR and long coherence block
  - noise is negligible but interference may not be so
  - # active users << coh. block

K users in the entire network

$\equiv$ single user setup with
K indep. tx antennas
n rx antennas

What is possible and what is not in the single user system?

$$\underset{n\times 1}{\underline{y}} = H \underset{K\times 1}{\underline{x}} + \underline{z} \rightsquigarrow \underset{n\times T}{Y} = H \underset{K\times T}{X} + Z$$

$$C \triangleq \sup_{P_X} I(X;Y), \quad \mathbb{E}|x_{kt}|^2 \leq \rho$$

[Zheng, Tse] $\quad C \sim n^*(T-n^*)\log\rho$

$$n^* = n \wedge K \wedge \frac{T}{2}$$

achievable via training...

... but $\perp$ pilots are needed

Assume $n = K < \frac{T}{2} \rightsquigarrow n^* = n = K$

$\rightsquigarrow C \sim n(T-K)\log\rho$

Can we derive this result w/o pilots in a way that suggests an optimal algorithm?

Main result

$$\text{Assume } X \sim P_X \text{ s.t. } \overbrace{h(X) \sim KT\log p}^{\text{"maximally entropic"}} \text{ and } n = K < \frac{T}{2}.$$

"maximally entropic" ← e.g. iid $h(x) \sim \log p$

$$\text{Then } I(X;Y) \gtrsim \underbrace{h(Y|H)}_{=} - \underbrace{nK\log p}_{\text{"given"}} \sim \underbrace{n(T-K)\log p}_{\text{opt. scaling}}$$

$$\mathbb{E}\Big[\max_{H'} \Big\{ -D\big(f_{Y|H=H_0} \,\|\, f_{Y|H=H'} \,\big|\, H=H_0\big) - \mathbb{E}\big[\log f_{Y|H=H'} \,\big|\, H=H_0\big] \Big\}\Big] \qquad (*)$$

wrt $H_0 \sim \mathcal{CN}(0,\ldots)$

H': arg. of optimiz.

H_0: true channel

Comments:

1) since $D(\cdots\|\cdots) \geq 0$, then $H' = H_0$ removes the first term and what remains is $h(Y|H)$
2) the result is tight due to Zheng-Tse
3) $(*)$ suggests an algorithm ...

Pf.

1) $I(X;Y) = h(Y) - h(Y|X)$

2) $h(Y|X) \leq n\, \mathbb{E} \log\det(I + XDX^*)$ from indep. channels and their Gauss.

$\sim n\,(n \wedge T) \log \rho$ from $X \in \mathbb{C}^{n \times T}$ and max. entropy (at high SNR)

$= n^2 \log \rho$ from $n = K < \frac{T}{2}$

3) $h(Y) \geq h(Y|H)$ because conditioning reduces entropy

$\sim h(HX|H)$ ignoring noise

$= h(X) + \underbrace{\mathbb{E} \log\det H}_{O_p(1)}$ from properties of diff. entropy

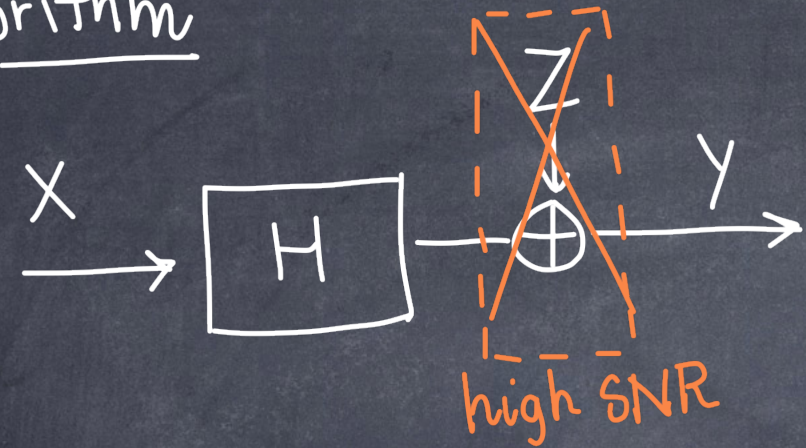
$\sim KT \log \rho$ from max. entropic inputs

4) $I(X;Y) \gtrsim (KT - n^2) \log \rho$ from 1), 2) and 3)

$= n\,(T-K) \log \rho$ from $n = K$

5) Just rewrite $h(Y|H)$ in terms of KL divergence with an unknown $f_{Y|H=H'}$ ∎

Algorithm



$Y \simeq HX$ given H

$\downarrow$

process data "block by block"

Recall the step: $h(Y) \geq h(Y|H)$

$$= \underbrace{-D(f_{Y|H} \,\|\, \tilde{f} \,|\, H)}_{\max} - \underbrace{\mathbb{E}[\log \tilde{f} \,|\, H]}_{\text{ignore}}$$

Also, minimizing K-L divergence = max log likelihood

$X \xrightarrow{H=H_0} Y_0 \sim f_{Y|H=H_0}$ ← actual

$X \xrightarrow{H=H_1} Y_1 \sim f_{Y|H=H_1}$ ← try

$\vdots$

$$\hat{H} \triangleq \arg\max_{H'} \log f_{Y|H=H'}(Y_0)$$

$\uparrow$ there are multiple solutions (among which H_0)

In summary: MLE on each block is asymptotically optimal when $n=K<\frac{T}{2}$.

N.B. H can be known up to symmetries of $f_{Y|H}$ (it is enough)

Important example: iid samples.

$y_1, y_2, \dots, y_T$ iid random vectors (given the channel)

$$f_{Y|H}(Y) = \prod_{t=1}^{T} f_{y|H}(y_t)$$

$$= \prod_{t=1}^{T} |\det B| \, f_x(B y_t)$$

by change of variable in $Y = HX$ and $B = H^{-1}$

MLE: $\hat{B} \triangleq \underset{B'}{\operatorname{argmax}} \left\{ T \log|\det B'| + \sum_{t=1}^{T} \log f_x(B' y_t) \right\}$

instead of H

for a fixed f_x we can solve this numerically

highly nonlinear and many variables: machine learning tools are perfect for this task

Numerical experiment

f_x : Laplacian density, i.i.d. samples

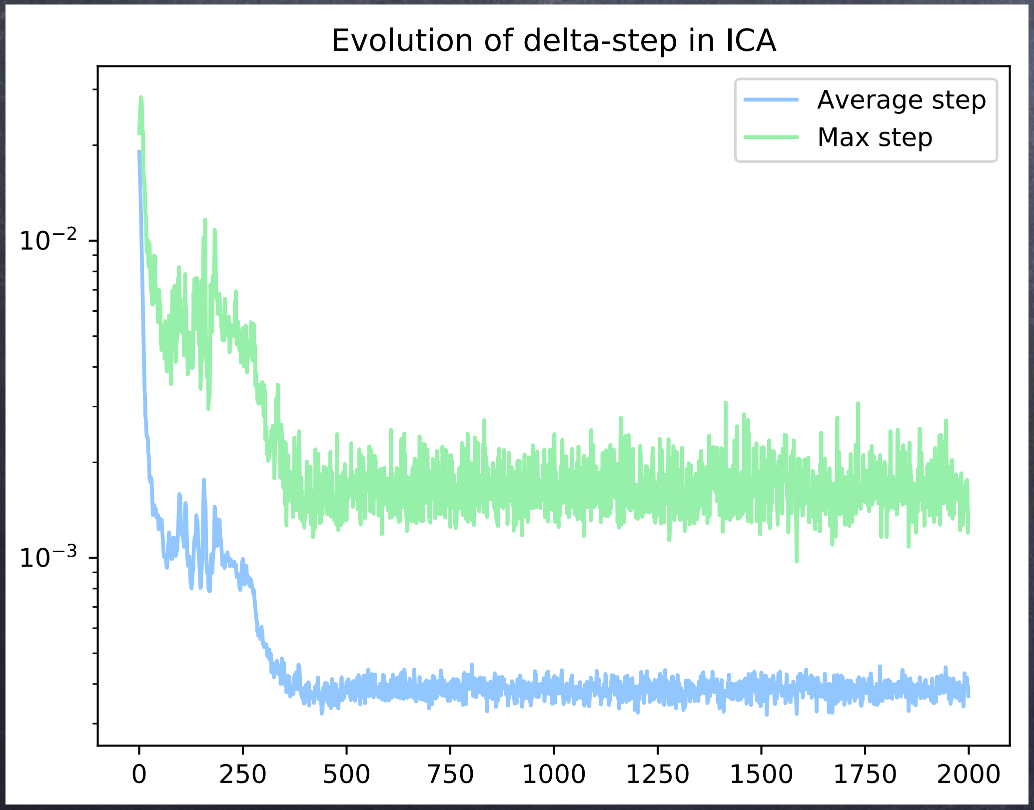
H : "not important" (Gaussian)
↖ max rank...

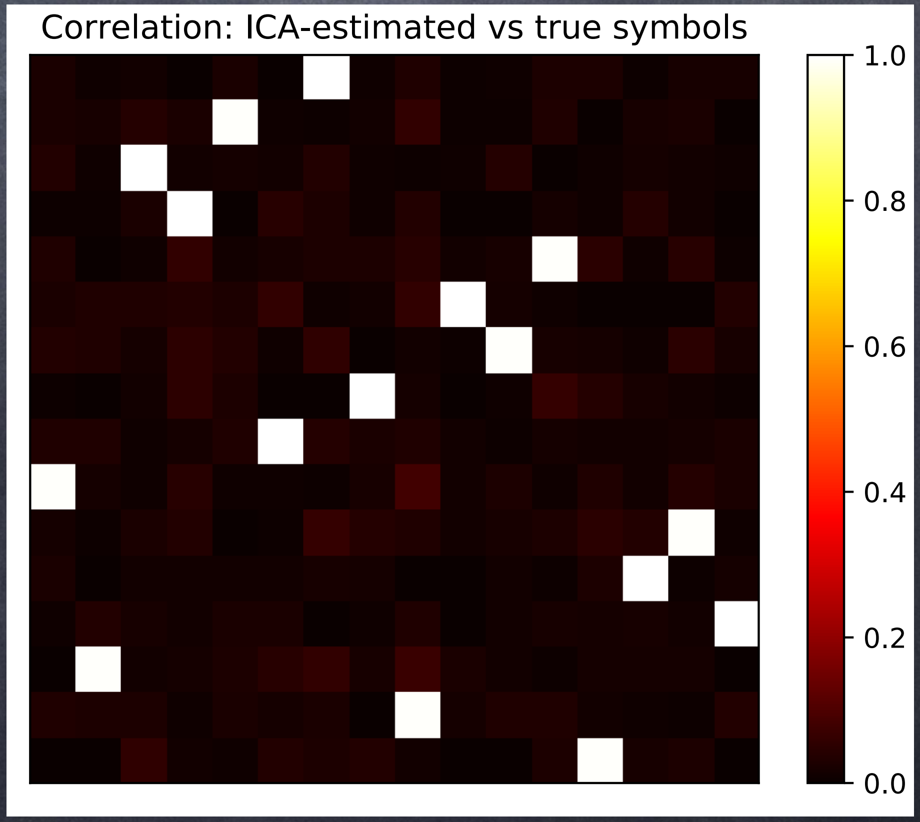
$\hat{B}$ from MLE and $\hat{X} \triangleq \hat{B} Y \leadsto$ if $\hat{B} = B$ then $\hat{X} = X \leadsto X^* X \propto I$

but in general $f_{Y|H}$ has symmetries, e.g. the "order" is lost

$\leadsto |X^* \hat{X}| \propto$ permutation matrix

Tensorflow is used to solve the numerical problem (SGD or Adam opt.)





Conclusion

- MLE on each block of the fading channel is asymptotically optimal
- The problem is numerically solvable with ML tools

Questions & open challanges

no compressed sensing...

1) Since we are estimating H ($= n^2$ samples) and H is "dense", do we need $\cancel{n}T \geq n^2$ for the algorithm to work? Would it work with less?

2) Do we really need to know f_X? What do we really need to know?

3) How Can we speed up the algorithm? Today's GPUs take $\sim 10\,s$

channel coherence time $\sim 10\,ms$

10^3

4) Would Markov models instead of block models speed up convergence?

5) How much Can a little bit of cooperation help?

guido.carlo.ferrante@ericsson.com
gmail.com