Identification of ARMA Models using Intermittent and Quantized Output Observations

Minyue Fu
Zhejiang University, China

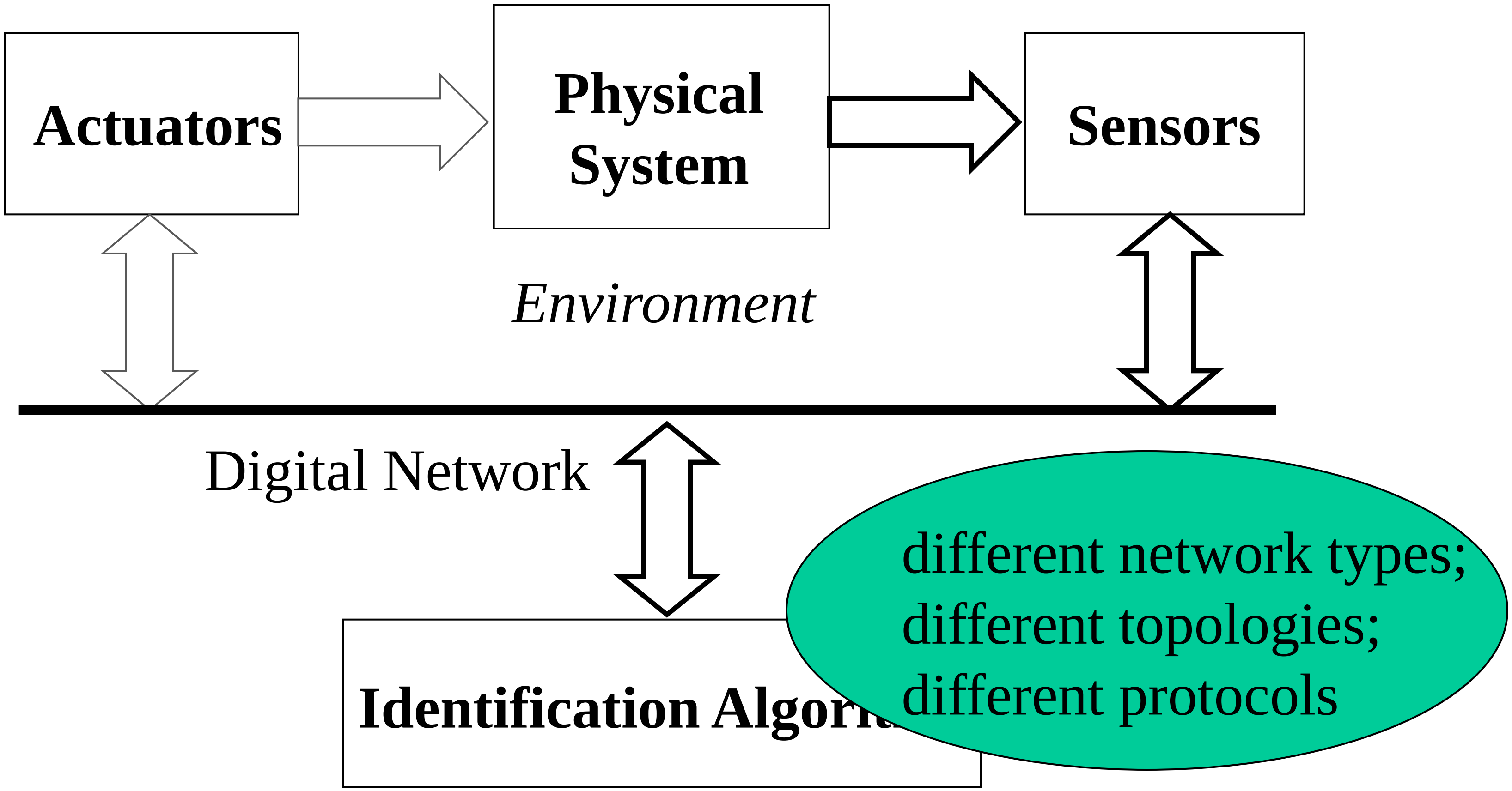
Joint work with
Keyou You at Nanyang Technological University, Singaprore
Damian Marelli at University of Newcastle, Australia

The 5th Swedish-Chinese Conference on Control

Outline

- Problem Formulation
- Maximum Likelihood Estimation
- Asymptotic Analysis
- Quantizer Design
- Simulation Examples
- Concluding Remarks

Networked System Identification



Our research problems:

- To study the joint effects of quantization and packet dropouts on system identification.
- To derive effective system identification algorithms to cope with both quantization and packet dropouts
- To jointly design quantizer and parameter estimator for system identification.

For simplicity, we consider i.i.d. packet dropouts, i.e., the probability that each measured output is lost is constant and is independent of other measurements.

Motivating Example on quantization

System:

$$y = \theta + v$$

where θ is the unknown parameter and v is a zero-mean noise.

Quantized measurement (1-bit quantizer):

$$z = Q(y) = \begin{cases} 1, & \text{if } y \geq a \\ 0, & \text{if } y < a \end{cases}$$

The key question is where to place the threshold a so that the MSE for the estimated parameter $\hat{\theta}$, based on the quantized measurement z, is minimized.

Answer: $a = \theta$. [Ribeiro and Giannakis, IEEE-TSP, 2006.]

But the trouble is: θ is the unknown!

How to design quantizers in a general setting is a difficult question.

ARMA Model:
$$\begin{aligned} x(t) &= \frac{B(q)}{A(q)}u(t) \\ y(t) &= x(t) + w(t) \qquad \text{(scalar output)} \\ z(t) &= \gamma_t \mathcal{Q}_t(y(t)), \end{aligned}$$

where $u(t)$ is a given deterministic signal
$w(t)$ is an i.i.d. measurement noise with distribution $N(0,\sigma^2)$
$y(t)$ is quantized by a K-level quantizer $\mathcal{Q}_t$
γ_t is a packet dropout parameter, a sequence of i.i.d. Bernolli random variables with $\mathrm{P}(\gamma_t = 1) = \lambda, \quad \mathrm{P}(\gamma_t = 0) = 1 - \lambda$

Quantizer: $\mathcal{Q}_t : \mathbb{R} \to \{v_{t,1}, \cdots, v_{t,K}\},\ t \in \mathbb{Z}$ (possibly time-varying)

Quantization intervals: $[b_{t,k-1}, b_{t,k}] = \mathcal{Q}^{-1}[v_{t,k}],\ k = 1, \cdots, K$
with $b_{t,0} = -\infty$ and $b_{t,K} = \infty$

Notation:

Set of received sample indices: $\mathcal{I} = \{1 \leq t \leq N : \gamma_t = 1\}$ (N is fixed)

Set of received quantized samples: $Z_{\mathcal{I}} = \{z(t) : t \in \mathcal{I}\}$

Plant denominator: $A(q) = 1 + a_1 q^{-1} + \cdots + a_m q^{-m}$

Plant numerator: $B(q) = b_0 + \cdots + b_n q^{-n}$

Parameter vector: $\theta_\star = [b_0, \cdots, b_n, a_1, \cdots, a_m]^T$ (subscript star: true)

Parameterized denominator and numerator: $A(q,\theta)$ and $B(q,\theta)$

Input sequence: $U_N = \{u(t) : t = 1, \cdots, N\}$

Research Problem:

Given *N*, U_N and $Z_{\mathcal{I}}$, compute the maximum likelihood estimate

$\hat{\theta}_N$ of $\theta_\star$.

Outline

- Problem Formulation
- $\boxed{\text{Maximum Likelihood Estimation}}$
- Asymptotic Analysis
- Quantizer Design
- Simulation Examples
- Concluding Remarks

Maximum Likelihood Estimation (MLE)

The MLE problem is to compute

$$\hat{\theta}_N \in \arg\max_{\theta} p\left(Z_{\mathcal{I}} | U_N, \theta\right)$$

$$= \arg\max_{\theta} l(\theta | U_N, Z_{\mathcal{I}}),$$

where

$$l(\theta | U_N, Z_{\mathcal{I}}) = \log p\left(Z_{\mathcal{I}} | U_N, \theta\right)$$ (log-likelihood function)

(Note: The term U_N will be suppressed for simplicity).

Direct solution to MLE is known to be difficult in general.

Proposed 2-Step Method:

Step 1: Expectation maximization (EM) algorithm

Advantages:

- No initialization needed;
- Quick descending

Disadvantage: Slow convergence

Step 2: Quasi-Newton gradient search algorithm

Advantage: Fast convergence

Expectation Maximization Method

If $Y_N = \{y(t) : t = 1, \cdots, N\}$ were available, it would be much easier to maximize $\log p(Y_N|\theta)$, which is the case in the traditional system identification problem.

Now, because Y_N is not available, we replace $\log p(Y_N|\theta)$ with the average of

$$\log p(Z_{\mathcal{I}}, Y_N|\theta) = \log p(Z_{\mathcal{I}}|Y_N) + \log p(Y_N|\theta)$$

over all possible values of Y_N.

This averaging is done using the conditional probability $p\left(Y_N|Z_{\mathcal{I}}, \hat{\theta}\right)$ of Y_N, given $Z_{\mathcal{I}}$ and some previous estimate $\hat{\theta}$.

Iterative Procedure:

$$\begin{aligned}\hat{\theta}_N^{(i)} &\in \arg\max_{\theta} Q(\theta, \hat{\theta}_N^{(i-1)}) \\ Q(\theta, \hat{\theta}) &= \int \log p\left(Z_{\mathcal{I}}, Y_N | \theta\right) p\left(Y_N | Z_{\mathcal{I}}, \hat{\theta}\right) dY_N\end{aligned}$$

In theory, for each N, we should iterate over $i = 1, 2, \ldots.,$ until convergence. But this is too much time consuming.

We propose: One iteration for each time step N, i.e.,

$$\hat{\theta}_N \in \arg\max_{\theta} Q(\theta, \hat{\theta}_{N-1})$$

Closed Form for $\max Q(\theta, \hat{\theta}_{N-1})$:

Lemma:

$$Q(\theta,\hat{\theta}) = -\frac{|\mathcal{I}|}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{N}(\bar{y}(t,\hat{\theta}) - x(t,\theta))^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{N}(\check{y}(t,\hat{\theta}) - \bar{y}^2(t,\hat{\theta}))$$

where

$$\bar{y}(t,\hat{\theta}) = \begin{cases} \mathcal{E}\left\{y(t)|z(t),\hat{\theta}\right\}, & t\in\mathcal{I} \\ x(t,\hat{\theta}) & t\in\bar{\mathcal{I}} \end{cases} \qquad \check{y}(t,\hat{\theta}) = \begin{cases} \mathcal{E}\left\{y^2(t)|z(t),\hat{\theta}\right\}, & t\in\mathcal{I} \\ x^2(t,\hat{\theta}) + \sigma^2 & t\in\bar{\mathcal{I}} \end{cases}$$

Using the above, we obtain

$$\hat{\theta}_N \in \arg\min_{\theta}\sum_{t=1}^{N}(\bar{y}(t,\hat{\theta}_{N-1}) - x(t,\theta))^2$$

Recursive Implementation of the EM Algorithm:

$$\hat{\theta}_N = \hat{\theta}_{N-1} + L_N \left(\tilde{y}(N, \hat{\theta}_{N-1}) - \tilde{\phi}^T(N, \hat{\theta}_{N-1})\hat{\theta}_{N-1} \right)$$

where

$$L_N = \frac{P_{N-1}\tilde{\phi}(N, \hat{\theta}_{N-1})}{1 + \tilde{\phi}^T(N, \hat{\theta}_{N-1})P_{N-1}\tilde{\phi}(N, \hat{\theta}_{N-1})}$$

$$P_N = P_{N-1} - \frac{P_{N-1}\tilde{\phi}(N, \hat{\theta}_{N-1})\tilde{\phi}^T(N, \hat{\theta}_{N-1})P_{N-1}}{1 + \tilde{\phi}^T(N, \hat{\theta}_{N-1})P_{N-1}\tilde{\phi}(N, \hat{\theta}_{N-1})}$$

$$\tilde{y}(t, \theta) = \frac{\bar{y}(t, \theta)}{A(q, \theta)}$$

$$\tilde{\phi}(t, \theta) = \frac{1}{A(q, \theta)}[u(t), \cdots, u(t-n), -\bar{y}(t-1, \theta), \cdots, -\bar{y}(t-m, \theta)]^T$$

Quasi-Network Search Method

Iterative Procedure:

$$\hat{\theta}_N = \hat{\theta}_{N-1} - \mu_N T_N g_N$$

where the matrix T_N denotes the approximate Hessian of $l(\theta|Z_{\mathcal{I}})$ at $\hat{\theta}_{N-1}$,g_N represents the gradient of $l(\theta|Z_{\mathcal{I}})$ at $\hat{\theta}_{N-1}$.

Computation of T_N (using BFGS Formula):

$$\begin{aligned} T_{N+1} &= T_N + \left(1 + \frac{q_N^T T_N q_N}{s_N^T q_N}\right) \frac{s_N s_N^T}{s_N^T q_N} \\ &\quad - \frac{s_N q_N^T T_N + T_N q_N s_N^T}{s_N^T q_N}, \end{aligned}$$

where

$$s_N = \theta_{N+1} - \theta_N \qquad q_N = g_{N+1} - g_N$$

Computation of g_N:

$$g_N = \frac{\partial}{\partial \tilde{\theta}} l(\tilde{\theta} | Z_{\mathcal{I}}) \Big|_{\theta} = \frac{1}{\sigma^2} \sum_{t \in \mathcal{I}} \left(\bar{y}(t,\theta) - x(t,\theta) \right) \psi(t,\theta),$$

where

$$\begin{aligned} \psi(t,\theta) &= \frac{\partial}{\partial \tilde{\theta}} x(t,\tilde{\theta}) \Big|_{\theta} = \phi(q,\theta) u(t), \\ \phi(q,\theta) &= \left[\frac{\Omega_n^T(q)}{A(q,\theta)}, \frac{q^{-1} B(q,\theta) \Omega_{m-1}^T(q)}{A^2(q,\theta)} \right]^T, \end{aligned}$$

$$\Omega_n(q) = [1, q^{-1}, \cdots, q^{-n}]^T$$

Outline

- Motivation
- Problem Formulation
- Maximum Likelihood Estimation
- Asymptotic Analysis
- Quantizer Design
- Simulation Examples
- Concluding Remarks

Strong Consistency

Theorem 1: Let $\mathcal{D} \subset \mathbb{R}$ be a compact set containing the true parameter vector $\theta_\star$, and such that, for all $\theta \in \mathcal{D}$, the roots of $A(q,\theta)$ have magnitudes smaller than or equal to $1-\epsilon$, for some $\epsilon > 0$. Let *u*(*t*) be bounded and such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \left(x(t,\theta) - x(t,\theta_\star)\right)^2 = 0$$

holds if and only if $\theta = \theta_\star$. If for each *N*,

$$\hat{\theta}_N \in \arg\max_{\theta \in \mathcal{D}} l(\theta | Z_{\mathcal{I}})$$

then

$$\lim_{N\to\infty} \hat{\theta}_N = \theta_\star \ \textit{w.p.1}$$

Asymptotic Normality

Theorem 2: Let $$\Phi_\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \mu(t)\psi(t,\theta_\star)\psi^T(t,\theta_\star)$$

where

$$\mu(t) = \frac{\bar{\sigma}^2(t)}{\sigma^2}$$

$$\bar{\sigma}^2(t) = \mathcal{E}\left\{\left(\bar{y}(t,\theta_\star) - x(t,\theta_\star)\right)^2\right\}$$

Suppose the conditions of Theorem 1 hold and that $\theta_\star$ is in the interior of $\mathcal{D}$. Then, Φ_μ is invertible and

$$\sqrt{N}\left(\hat{\theta}_N - \theta_\star\right) \to \mathcal{N}(0, C) \ \textit{in dist.}$$

where

$$\boxed{C = \frac{\sigma^2}{\lambda}\Phi_\mu^{-1}}$$

Outline

- Problem Formulation
- Maximum Likelihood Estimation
- Asymptotic Analysis
- $\boxed{\text{Quantizer Design}}$
- Simulation Examples
- Concluding Remarks

Result: The optimal choice for the quantizer design is

$$b_{t,k} = \tilde{b}_k + x(t, \theta_\star)$$

where $\tilde{b}_k,\ k = 0, \cdots, K$ are the boundaries of the Lloyd's quantizer for the process noise *w*. With the above quantizer, we have

$$C = \frac{\sigma^2}{\lambda\mu}\Phi^{-1}$$

where

$$\Phi = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} \psi(t, \theta_\star)\psi^T(t, \theta_\star)$$

$$\mu = \frac{\bar{\sigma}^2}{\sigma^2} \qquad \bar{\sigma}^2 = \mathcal{E}\left\{ \tilde{Q}^2\left[w(t)\right]\right\}$$

Implementation:

$$b_{t,k} = \tilde{b}_k + x(t, \hat{\theta}_{t-1}).$$

Outline

- Problem Formulation
- Maximum Likelihood Estimation
- Asymptotic Analysis
- Quantizer Design
- $\boxed{\text{Simulation Examples}}$
- Concluding Remarks

Simulation Examples

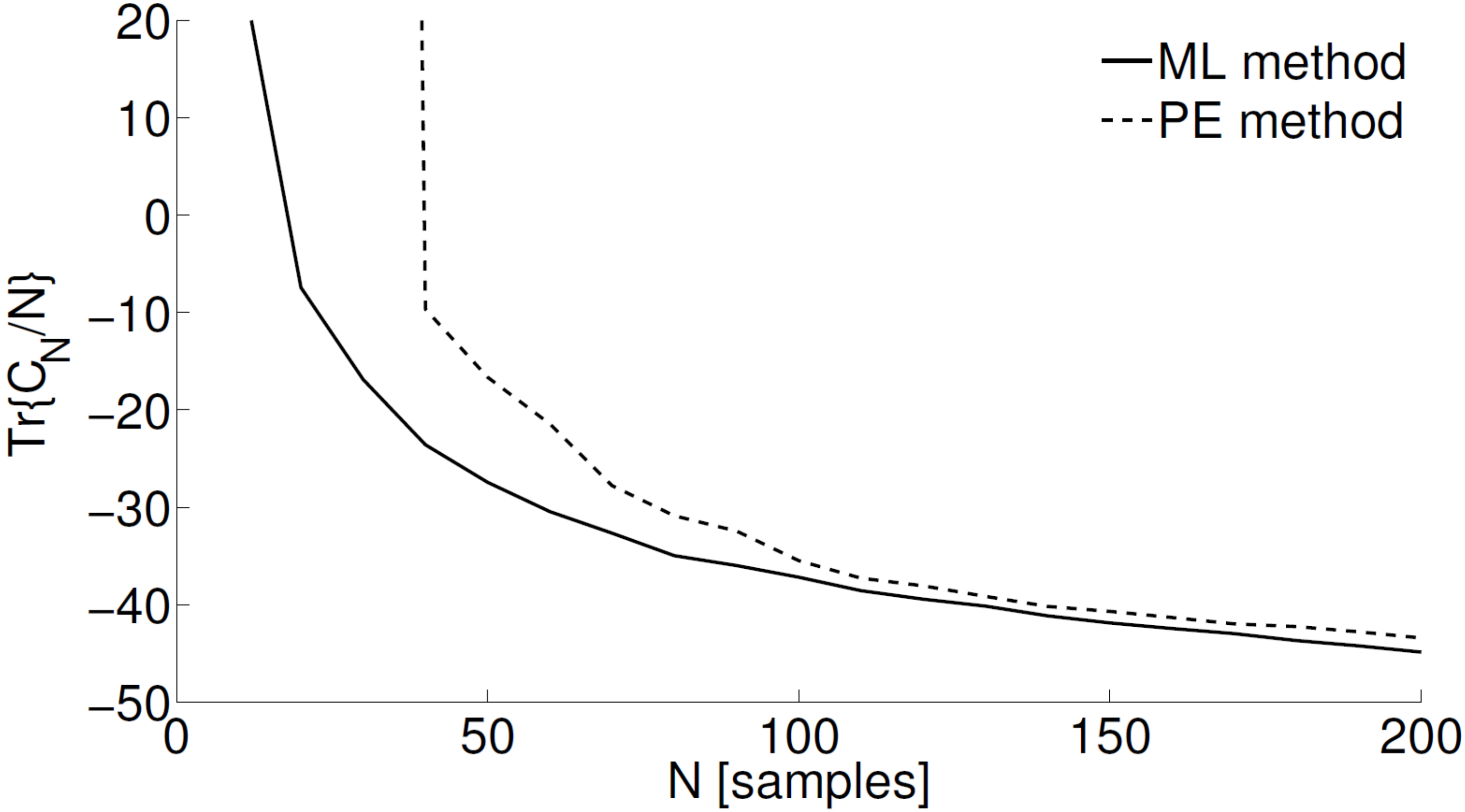
Example 1: Comparison with Prediction Error Criterion
(No packet loss is considered)

PE method: Ignore the presence of the quantizer and estimates the parameters to minimize the power of the difference between the quantized samples $z(t)$ and their value predicted using the input signal $u(t)$ and θ.

ARMA Model: $$\frac{B(q)}{A(q)} = \frac{1}{1 - 1.764q^{-1} + 0.81q^{-2}}.$$

Noise model for w: Truncated Gaussian with $\sigma^2 = 0.1$

Quantizer: 2-bit ($K = 4$)



ML method
PE method
Tr{C_N/N}
N [samples]
20
10
0
−10
−20
−30
−40
−50
0
50
100
150
200

Example 2: Convergence Comparison for Different Quantizers

Same ARMA model, noise model and data rate as in Example 1.
Packet arrival rate: $\lambda = 0.9$.

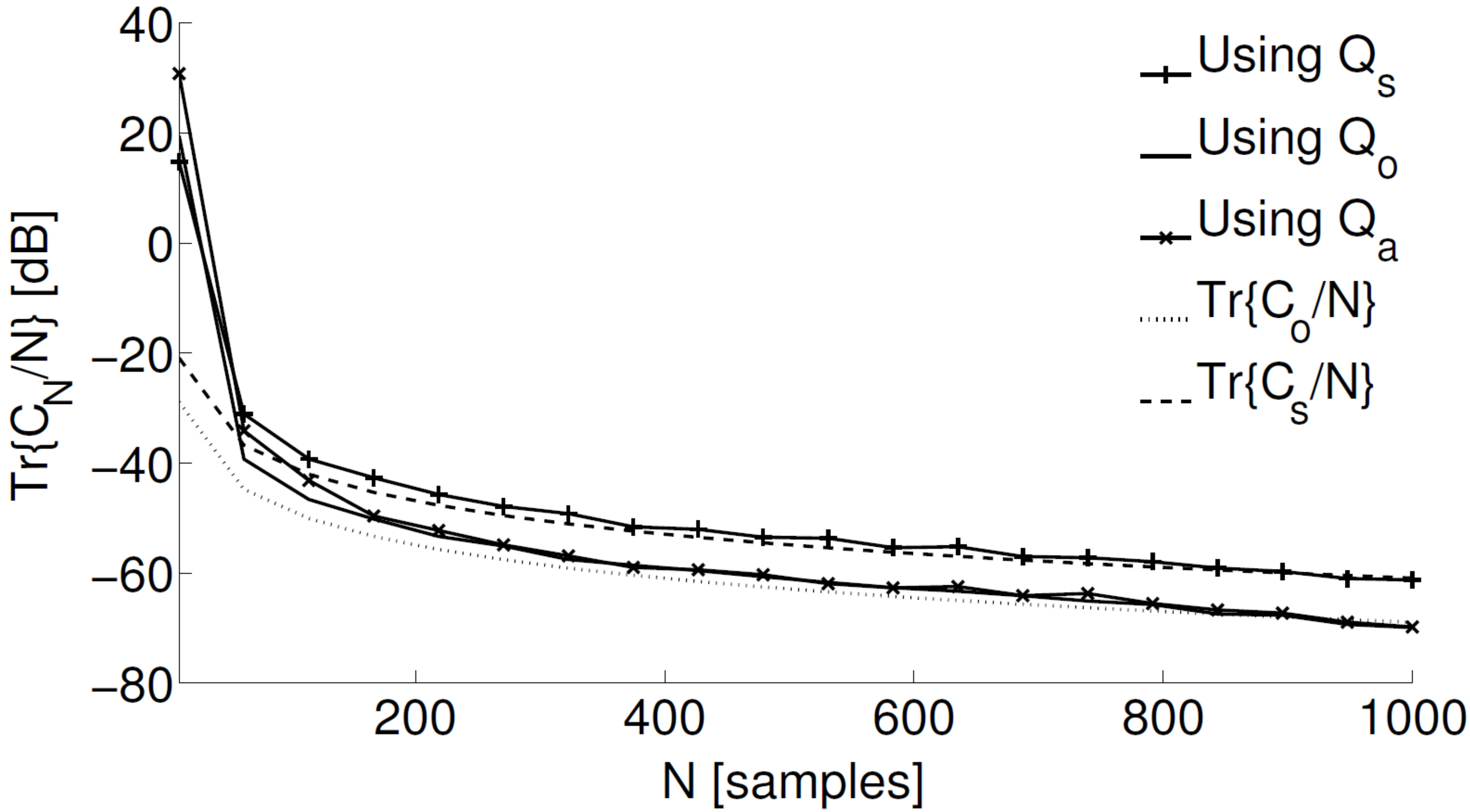
Quantizer 1: $\mathcal{Q}_s$ is stationary

Quantizer 2: $\mathcal{Q}_o$ is the optimal quantizer

$$b_{t,k} = \tilde{b}_k + x(t, \theta_\star)$$

Quantizer 3: $\mathcal{Q}_a$ is the adaptive quantizer

$$b_{t,k} = \tilde{b}_k + x(t, \hat{\theta}_{t-1}).$$



Using Q_s
Using Q_o
Using Q_a
$Tr\{C_o/N\}$
$Tr\{C_s/N\}$
$Tr\{C_N/N\}$ [dB]
N [samples]
40
20
0
−20
−40
−60
−80
200
400
600
800
1000

Concluding Remarks

- We have studied a system identification problem with quantization constraints and packet losses
- A new recursive algorithm has been proposed based on the EM method and Newton search.
- **Asymptotic analysis shows that the proposed algorithm has similar asymptotic properties as in the case with network constraints, but the convergence rate is affected by the quantizer and the packet dropout rate.**
- **A precise characterization of the convergence rate is provided.**
- Quantizer design is considered and an adaptive quantizer is suggested.