Particle-based simulation of the Gel'fand-Pinsker channel capacity and the Wyner-Ziv rate-distortion function

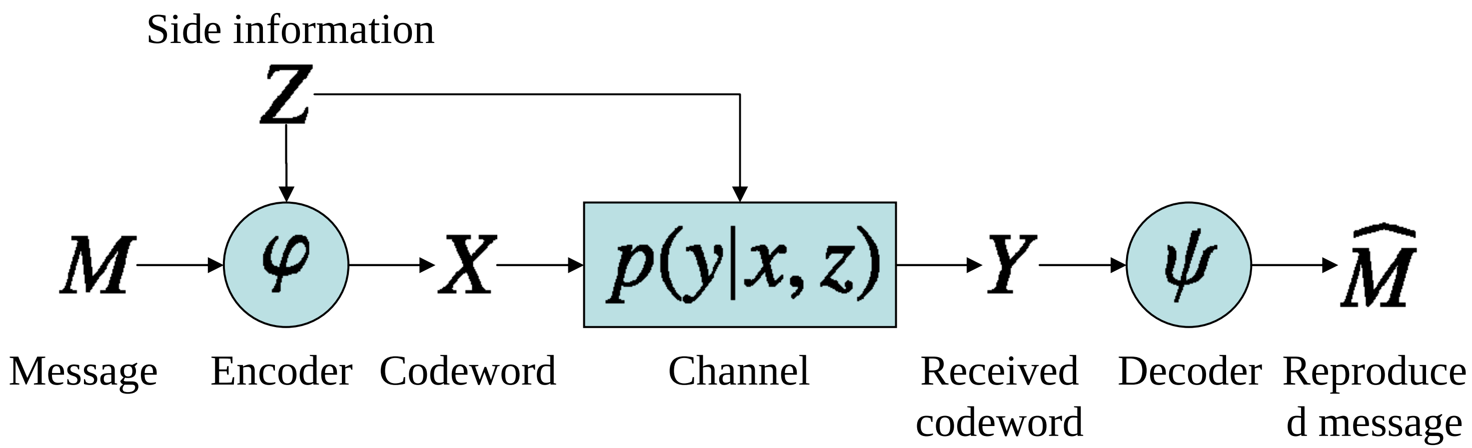
Akisato Kimura

Media Information Laboratory
NTT Communication Science Laboratories
NTT Corporation

Photo by Junko Kimura (in Seven Sisters, East Sussex, United Kingdom)

Abstract

- A new numerical algorithm for simulating
  - Gel'fand-Pinsker channel capacity for memoryless channels
  - Wyner-Ziv rate-distortion function for memoryless sources (skipped)



- Basic idea
  - Represent a density by a finite number of "particles" (= sample values + the associated weights)

Outline

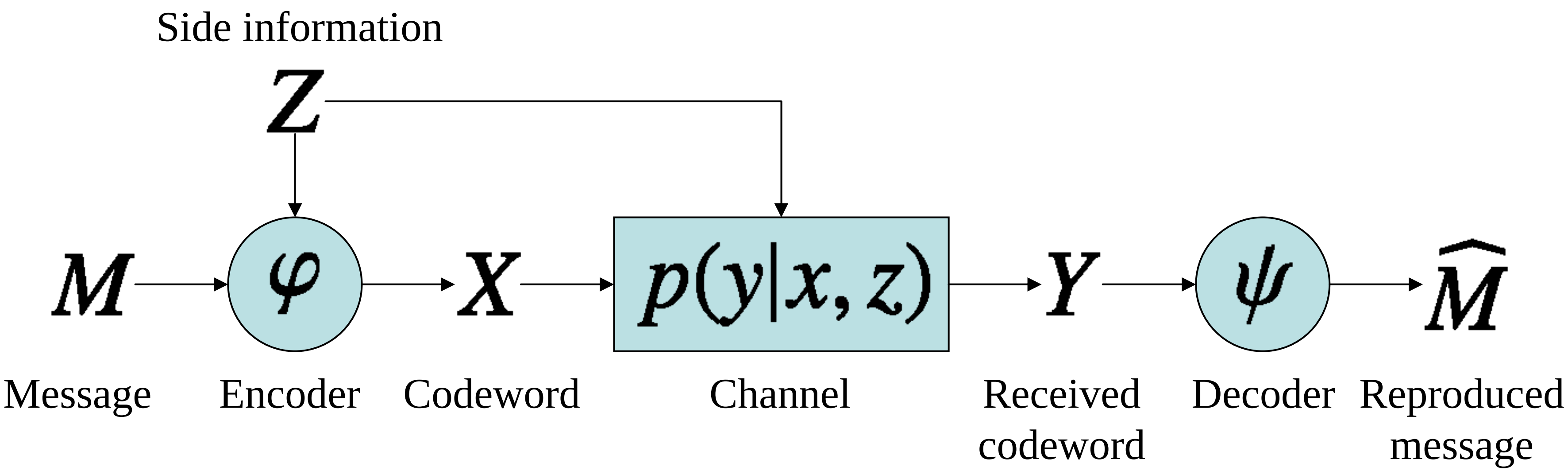
1. Introduction
2. Previous method
3. Proposed method
4. Summary

Introduction

1. Introduction
2. Previous method
3. Proposed method
4. Summary

Background

- Gel'fand and Pinsker (1980)
  - Investigated a channel with non-causal encoder side information



Capacity:

$$C_{GP}(X;Y|Z) = \max_{P_{U|Z}, P_{X|UZ}} \{I(U;Y) - I(U;Z)\}$$

Background

- Dupuis, Fu and Willems (2004)
  - Modified Arimoto-Blahut algorithm for computing Gel'fand-Pinsker channels
  - Applicable only for finite alphabets

  - GP capacity for memoryless channels with infinite alphabets

$$C_{GP}(X;Y|Z) = \sup_{P_{U|Z},P_{X|UZ}} \{I(U;Y) - I(U;Z)\}$$

$$= \sup_{P_{U|Z},P_{X|UZ}} \int_{(u,x,y,z)} P_{UXYZ}(u,x,y,z) \log \frac{P_{U|Y}(u|y)}{P_{U|Z}(u|z)} dudxdydz$$

→ Integration calculations are often intractable.

Approach

- Represent a density by a finite number of "particles" (= samples + associated weights)

$$P_X(x) \approx \widetilde{P}_X(x) = \sum_{i=1}^{N} w_i \cdot \delta(x - \widetilde{x}_i)$$

$\widetilde{x}_i$: sample
w_i: weight
$\delta(\cdot)$: delta function

(Note) Particle filter

One of the most popular Bayesian algorithms for estimating hidden states based only on observations.

The procedure:
1. Update weights
2. Re-sampling
3. Update samples

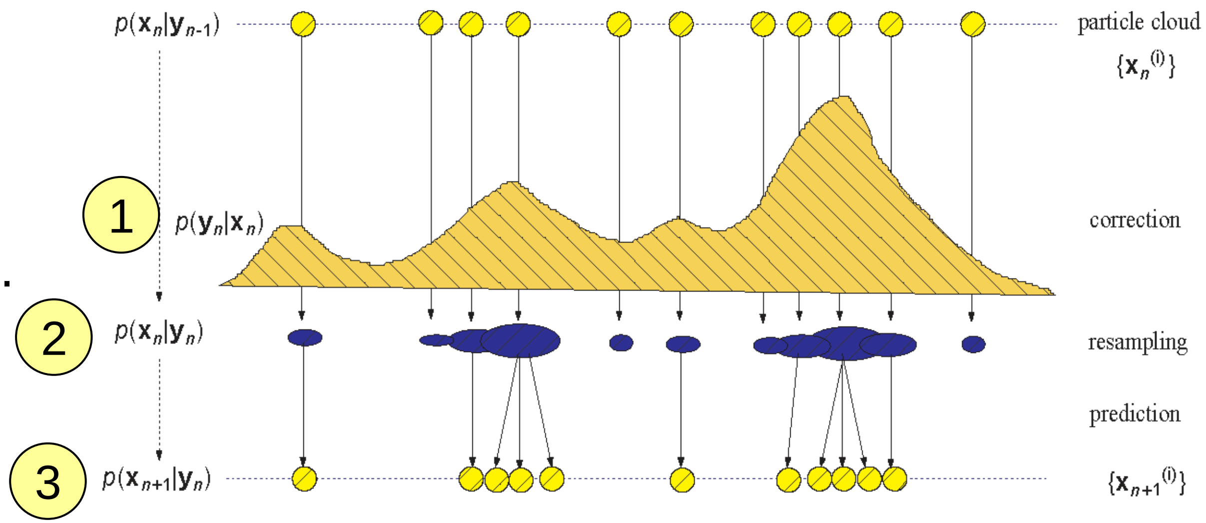


Fig. 9. An illustration of generic particle filter with importance sampling and resampling.

Previous method

1. Introduction
2. Previous method
3. Proposed method
4. Summary

Another form of GP capacity

- Original form of Gel'fand-Pinsker (GP) capacity

$$C_{GP}(X;Y|Z) = \sup_{P_{U|Z}, P_{X|UZ}} \int_{(u,x,y,z)} P_{UXYZ}(u,x,y,z) \log \frac{P_{U|Y}(u|y)}{P_{U|Z}(u|z)} du dx dy dz$$

  – Simultaneous optimization may be intractable.

- Dupuis, Fu and Willems (2004)
  – The capacity is given with a function $t : \mathcal{Z} \to \mathcal{X}$ by

$$\begin{aligned} C_{GP}(X;Y|Z) &= \sup_{P_{T|Z}} \{ I(T;Y) - I(T;Z) \} \\ &= \sup_{P_{T|Z}} \int_{(u,y,z)} P_{TYZ}(t,y,z) \log \frac{P_{T|Y}(t|y)}{P_{T|Z}(t|z)} dt dy dz \\ P_{TYZ} &= P_Z(z) P_{T|Z}(t|z) P_{Y|XZ}(y|t(z),z) \end{aligned}$$

Arimoto-Blahut for GP channels

- The GP capacity is rewritten as

$$C_{GP}(X;Y|Z) = \sup_{Q_{T|Z}, Q_{T|Y}} \int_{(t,y,z)} P_Z(z) Q_{T|Z}(t|z) P_{Y|XZ}(y|t(z), z) \log \frac{Q_{T|Y}(t|y)}{Q_{T|Z}(t|z)} dt dy dz$$

- $Q_{T|Z}$ and $Q_{T|Y}$ can be updated in each iteration as

$$Q^{(m)}_{T|Y}(t|y) \propto \int_z P_Z(z) Q^{(m-1)}_{T|Z}(t|z) P_{Y|XZ}(y|t(z), z) dz,$$

$$Q^{(m)}_{T|Z}(t|z) \propto \exp\left\{ \int_y P_{Y|XZ}(y|t(z), z) \log Q^{(m)}_{T|Y}(t|y) dy. \right\}$$

  - Integration calculations are often intractable.
  - Calculating $Q_{T|Z}$ and $Q_{T|Y}$ for all t is impossible

Proposed method

1. Introduction
2. Previous method
3. <u>Proposed method</u>
4. Summary

Framework (1 / 4)

- $P_{T|Z}$ is represented by a finite number of particles

$$P_{T|Z}(t|z) \approx \widetilde{P}_{T|Z}(t|\widetilde{z}_i) = \sum_{j=1}^{N} w_{i,j} \cdot \delta(t - \widetilde{t}_{i,j}),$$

$$\mathcal{S}_{T|Z} = \left\{\mathcal{S}_{T|Z}(\widetilde{z}_i)\right\}_{i=1}^{M_Z}, \quad \mathcal{S}_{T|Z}(\widetilde{z}_i) = \left\{\widetilde{t}_{i,j}\right\}_{j=1}^{N}, \quad \text{: samples}$$

$$\mathcal{W}_{T|Z} = \left\{\mathcal{W}_{T|Z}(\widetilde{z}_i)\right\}_{i=1}^{M_Z}, \quad \mathcal{W}_{T|Z}(\widetilde{z}_i) = \left\{w_{i,j}\right\}_{j=1}^{N} \text{: weights}$$

  – $\mathcal{W}_{T|Z}$ and $\mathcal{S}_{T|Z}$ determine the density $P_{T|Z}$

Framework (2 / 4)

- The original optimization problem

$$C_{GP}(X;Y|Z) = \sup_{P_{T|Z}} \int_{(u,y,z)} P_{TYZ}(t,y,z) \log \frac{P_{T|Y}(t|y)}{P_{T|Z}(t|z)} dt dy dz$$

can be replaced as

$$\underline{C}_{GP}(X;Y|Z) = \max_{\mathcal{S}_{T|Z}, \mathcal{W}_{T|Z}} I(\mathcal{S}_{T|Z}, \mathcal{W}_{T|Z})$$

$$I(\mathcal{S}_{T|Z}, \mathcal{W}_{T|Z})$$

$$= \int_y \sum_{i=1}^{M_Z} \sum_{j=1}^{N} \widetilde{P}_{TYZ}(\widetilde{t}_{i,j}, y, \widetilde{z}_i; w_{i,j}) \log \frac{\widetilde{P}_{T|Y}(\widetilde{t}_{i,j}|y)}{w_{i,j}} dy$$

$$\widetilde{P}_{TYZ}(\widetilde{t}_{i,j}, y, \widetilde{z}_i; w_{i,j}) = w_{i,j}\, P_Z(\widetilde{z}_i)\, P_{Y|XZ}(y|\widetilde{t}_{i,j}(\widetilde{z}_i), \widetilde{z}_i)$$

Framework (3 / 4)

- The maximization can be solved by the following alternative maximization:

$$\mathcal{W}_{T|Z}^{(n)} = \arg\max_{\mathcal{W}_{T|Z}} I(\mathcal{S}_{T|Z}^{(n-1)}, \mathcal{W}_{T|Z}),$$
$$\mathcal{S}_{T|Z}^{(n)} = \arg\max_{\mathcal{S}_{T|Z}} I(\mathcal{S}_{T|Z}, \mathcal{W}_{T|Z}^{(n)})$$

- The first maximization:
  almost the same as the previous method

$$Q_{T|Y}^{(n,m)}(\tilde{t}_{i,j}^{(n-1)}|y) \propto \sum_{i=i}^{M_Z} w_{i,j}^{(n,m-1)} P_Z(\tilde{z}_i) P_{Y|XZ}(y|\tilde{t}_{i,j}^{(n-1)}(\tilde{z}_i), \tilde{z}_i),$$
$$w_{i,j}^{(n,m)} \propto \exp\left\{ \int_y P_{Y|XZ}(y|\tilde{t}_{i,j}^{(n)}(\tilde{z}_i), \tilde{z}_i) \log Q_{T|Y}^{(n,m)}(\tilde{t}_{i,j}^{(n-1)}|y) dy \right\}.$$

Framework (4 / 4)

- The second maximization:
  difficult to maximize it directly
  → update samples to increase the capacity
  1. Re-sampling to avoid "degeneracy".
     New samples are drawn from

$$\widetilde{P}_{T|Z}(t|\widetilde{z}_i) = \sum_{j=1}^{N} w_{i,j} \cdot \delta(t - \widetilde{t}_{i,j})$$

  2. Update samples based on steepest descent method.

$$\widetilde{t}_{i,j}^{(n)} = \widetilde{t}_{i,j}^{(n-1)} + \left(\lambda_{i,j}^{(n)} + \eta_{i,j}^{(n)}\right) \frac{\partial}{\partial \widetilde{t}_{i,j}} I(\mathcal{S}_{T|Z}, \mathcal{W}_{T|Z}^{(n)}) \Big|_{\mathcal{S}_{T|Z} = \mathcal{S}_{T|Z}^{(n-1)}}$$

$\lambda_{i,j}^{(n)}$: step size, $\eta_{i,j}^{(n)}$: r.v. which depends on $\lambda_{i,j}^{(n)}$

Procedure: summary

1. Initialization
   i. Draw samples $\{\tilde{z}_i\}_i$
   ii. Initialize particles $\mathcal{S}^{(0)}_{T|Z}, \mathcal{W}^{(0)}_{T|Z}$
2. Update weights $\mathcal{W}^{(n)}_{T|Z}$ given samples $\mathcal{S}^{(n-1)}_{T|Z}$
   i. Update a density $Q^{(n,m)}_{T|Y}$
   ii. Update weights $\mathcal{W}^{(n,m)}_{T|Z}$
   iii. Repeat until i.-ii. convergence
3. Update samples $\mathcal{S}^{(n)}_{T|Z}$ given weights $\mathcal{W}^{(n)}_{T|Z}$
   i. (if needed) Re-sampling
   ii. Update samples $\mathcal{S}^{(n)}_{T|Z}$
4. Repeat 2.-3. until convergence

Propositions

Lemma . $\underline{C}_{GP}(X;Y|Z) \leq C_{GP}(X;Y|Z)$

Lemma . *With the proposed method (without the resampling step)*
$I(\mathcal{S}_{T|Z}^{(n-1)}, \mathcal{W}_{T|Z}^{(n-1)}) \leq I(\mathcal{S}_{T|Z}^{(n-1)}, \mathcal{W}_{T|Z}^{(n)}) \leq I(\mathcal{S}_{T|Z}^{(n)}, \mathcal{W}_{T|Z}^{(n)})$

Theorem . *With the proposed method (without resampling), $\underline{C}_{GP}(X;Y|Z)$ converges to a local extrema if the step sizes $\{\lambda_i\}_{i=1,2,\ldots}$ satisfy the Wolfe conditions.*

Proposition . *Especially, if all the alphabets are equal to $\mathcal{R}^d$ (d: dimension) and the density P_{TYZ} is continuous, then with the proposed method we have*

$$\underline{C}_{GP}(X;Y|Z) \quad \xrightarrow{p} \quad C_{GP}(X;Y|Z) \quad (M_Z, N \to \infty)$$

How do we check the continuity of P_{TYZ} ??

Problems

1. Complexity and memory

   [Complexity priority]

   complexity: $O(M_Z N M_Y)$, memory: $O(M_Z N M_Y)$

   [Memory priority]

   complexity: $O(M_Z^2 N M_Y)$, memory: $O(\max\{M_Z N, N M_Y\})$

2. General convergence
   - Not proved yet.
   - When estimating the rate-distortion function by empirical estimators, in certain cases it may fail to converge to the rate-distortion function. [Harrison and Kontoyiannis 2008]

Summary

1. Introduction
2. Previous method based on Arimoto-Blahut
3. Proposed method
4. <u>Summary</u>

Concluding remarks

- Present a new numerical method for simulating the Gel'fand-Pinsker channel capacity
  - By representing a density with particles
  - It can be applied to simulate the Wyner-Ziv rate-distortion function

- Future work
  - Prove the theoretical convergence
  - Reduce the computational complexity

Thank you for your kind attention.

Appendix

Note

- Sampling itself is often intractable.
  → Markov Chain Monte-Carlo (MCMC)

- The algorithm sometimes encounters underflow problems, especially when calculating

$$w_{i,j}^{(n,m)} \propto \exp\left\{ \int_y P_{Y|XZ}(y|\tilde{t}_{i,j}^{(n)}(\tilde{z}_i), \tilde{z}_i) \log Q_{T|Y}^{(n,m)}(\tilde{t}_{i,j}^{(n-1)}|y) dy \right\}.$$

- The step size should be controlled so as to satisfy the Wolfe conditions.
  - Criteria for the convergence of the steepest descent