UNIVERSITÀ di VERONA
Dipartimento di INFORMATICA

Greedy additive approximation algorithms for minimum-entropy coupling problem

Supported by “Scuola di Dottorato di Scienze Naturali e Ingegneristiche” of the University of Verona, and by Entropy Journal via ISIT student travel grant.

Massimiliano Rossi

Department of Computer Science – University of Verona

ISIT 2019

07-12 Jul 2019, Paris

entropy

an open access journal by MDPI

IMPACT FACTOR 2.305

Aims and Scope

Entropy (ISSN 1099-4300), an open access journal, deals with the development and/or application of entropy or information-theoretic concepts in a wide variety of applications.

Ten Thematic Sections:

- Information Theory, Probability and Statistics
- Thermodynamics
- Statistical Physics
- Quantum Information
- Complexity
- Astrophysics and Cosmology, and Black Holes
- Entropy Reviews
- Entropy and Biology
- Signal and Data Analysis
- Multidisciplinary Applications

Editor-in-Chief

Prof. Dr. Kevin H. Knuth

University at Albany, Albany, USA

2017 Impact Factor: **2.305**

5-year Impact Factor: 2.303

WoS Rank: **22/78**

Q2 (category "Multidisciplinary Physics")

6,500

Total cites in 2018

4,300+

articles published since 1999

120+

Editorial Board Members

44

days from submission to publication in 2018

The minimum-entropy coupling problem

Given two probability distributions $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_m)$ the **minimum-entropy coupling** is the joint distribution $M = \{m_{i,j}\}_{\substack{i=1..m \\ j=1..n}}$.

Objective: $minimize\ H(M)$

$q\backslash p$	p_1	p_2	$\cdots$	p_j	$\cdots$	p_n
q_1				$m_{1,j}$		
q_2				$m_{2,j}$		
$\vdots$				$\vdots$		
q_i	$m_{i,1}$	$m_{i,2}$	$\cdots$	$m_{i,j}$	$\cdots$	$m_{i,n}$
$\vdots$				$\vdots$		
q_m				$m_{m,j}$		

$$\sum_{i=1}^{m} \sum_{j=1}^{n} m_{i,j} = 1$$

$$\forall \boldsymbol{i} = \boldsymbol{1}, \dots, \boldsymbol{m}, \qquad \boldsymbol{q_i} = \sum_{k=1}^{n} m_{i,k}$$

$$\forall \boldsymbol{j} = \boldsymbol{1}, \dots, \boldsymbol{n}, \qquad \boldsymbol{p_j} = \sum_{k=1}^{m} m_{k,j}$$

This problem is NP-Hard

M. Kovačcević, I. Stanojević, V. Šenk, Information and Computation, 2015.

Motivations – Causality Discovery

Given two phenomena A and B, determine whether it is A that causes B or vice versa.

E.g., determine whether variations of altitude (A) causes variations of temperature (B) or vice versa.

Given two random variables X and Y , if X causes Y then there exists an exogenous random variable E independent of X and a function $f()$ such that $Y = f(X, E)$.

J. Pearl, Causality., 2009.

In general, it is possible to find pairs (f, E) and (g, E') such that both $Y = f(X, E)$ and $X = g(Y, E')$ hold.

Kocaoglu et al. postulated that in the true causal direction, the entropy of the exogenous random variable E is small

M. Kocaoglu, A. G. Dimakis, S. Vishwanath, B. Hassibi, AAAI 2017.

Motivations – Finding the maximal mutual information

Given two probability distributions X and Y,

$$H(XY) = H(X) + H(Y) - I(X;Y)$$

Minimum-entropy coupling $\equiv$ maximal mutual information

F. Cicalese, L. Gargano, U. Vaccaro, IEEE Transactions on Information Theory 2019

This problem is central in,

- Information theoretic clustering.
- Design of quantizers.

Related Work

Greedy heuristic guaranteed to be a local minimum.

M. Kocaoglu, A. G. Dimakis, S. Vishwanath, B. Hassibi, AAAI 2017.

Independently, *A. Painsky, S. Rosset, M. Feder, ISIT 2013.*

Greedy algorithm that guarantees an additive approximation factor that scale with $\log(n)$, where n is the dimension of the marginals.

M. Kocaoglu, A. G. Dimakis, S. Vishwanath, B. Hassibi, ISIT 2017.

Greedy algorithm that guarantees an additive approximation 1

F. Cicalese, L. Gargano, U. Vaccaro, ISIT 2017.

Our contribution

Greedy heuristic guaranteed to be a local minimum.

M. Kocaoglu, A. G. Dimakis, S. Vishwanath, B. Hassibi, AAAI 2017.

Greedy algorithm that guarantees an additive approximation 1

F. Cicalese, L. Gargano, U. Vaccaro, ISIT 2017.

Algorithm

Proof technique: majorization

Our contribution

1. The greedy approach proposed by *Kocaoglu et al.* guarantees an additive approximation 1.
2. A general criterion for guaranteeing an additive approximation factor of 1, that might be of independent interest.

Majorization theory

Let $\mathcal{P}_n = \{(p_1, p_2, \ldots, p_n) | \, p_1 \geq p_2 \geq \; \ldots \geq p_n \geq 0 \; and \; \sum_{i=1}^{n} p_i = 1\}$

Given $p, q \in \mathcal{P}_n$ we say that p is **majorized** by q ($p \preccurlyeq q$) if and only if

$$\bigvee k = 1, \ldots, n \qquad \sum_{i=1}^{k} p_i \leq \sum_{i=1}^{k} q_i$$

Recalling the Shur-concavity of the entropy function,

if $p \preccurlyeq q$ then $H(p) \geq \; H(q)$

A. W. Marshall, I. Olkin, B. C. Arnold, Inequalities: theory of majorization and its applications. 1979

The partially ordered set $(\mathcal{P}_n, \preccurlyeq)$ is a lattice.
For each pairs $p, q \in \mathcal{P}_n$ there exists a unique **greatest lower bound** $z = p \wedge q$ that can be found in linear time.

F. Cicalese, U. Vaccaro, IEEE Transactions on Information Theory, 2002

Majorization theory - II

For any M, coupling of p and q, $H(M) \geq H(p \wedge q)$

F. Cicalese, L. Gargano, U. Vaccaro, ISIT 2017

Let $s = (s_1, s_2, \ldots, s_{2n})$ be the non zero elements of M such that $s_1 \geq s_2 \geq \ldots \geq s_{2n} \geq 0$

Let $hz = (hz_1, hz_2, \ldots, hz_{2n})$ be the array of the two halves of each element of $z = p \wedge q$,
i.e. for all $i = 1, .., n$ $hz_{2i-1} = hz_{2i} = \frac{z_i}{2}$

Note that $\mathrm{H}(hz) = H(z) + 1 = H(p \wedge q) + 1$.

Thus, if for all $k = 1, .., 2n$ it holds that $\sum_{i=1}^{k} hz_i \leq \sum_{i=1}^{k} s_i$, we have that $hz \preccurlyeq s$ thus $H(s) \leq H(hz)$ then $H(M) \leq H(p \wedge q) + 1$

Kocaoglu et al. Algorithm

Given two probability distributions $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$ have each n elements,
Let $p^{(k)} = \left(p_1^{(k)}, p_2^{(k)}, \dots, p_n^{(k)}\right)$ be the array of the residual pieces of p after k iterations, and $p^{(0)} = p$.
The algorithm, at each iteration:

1. finds the largest residual mass probability of the two marginals p and q;
2. places the minimum of them in the position corresponding to their coordinates in the joint probability matrix M;
3. updates the residual mass probability of the two marginals, removing from both the maximal elements, the value of the minimum of them already placed in M.

$\boldsymbol{p_3^{(6)} = p_3^{(5)} - s_6}$

$\boldsymbol{q_5^{(6)} = q_5^{(5)} - s_6}$

$\boldsymbol{q \backslash p}$	$\boldsymbol{p_1^{(5)}}$	$\boldsymbol{p_2^{(5)}}$	$\boldsymbol{p_3^{(5)}}$	$\boldsymbol{p_4^{(5)}}$	$\boldsymbol{p_5^{(5)}}$	$\boldsymbol{p_6^{(5)}}$
$\boldsymbol{q_1^{(5)}}$	s_1					
$\boldsymbol{q_2^{(5)}}$		s_2	s_3			
$\boldsymbol{q_3^{(5)}}$				s_4		
$\boldsymbol{q_4^{(5)}}$				s_5		
$\boldsymbol{q_5^{(5)}}$			$s_6 = \min\left(p_3^{(5)}, q_5^{(5)}\right)$			
$\boldsymbol{q_6^{(5)}}$						

In this case the algorithm terminates in at most $2n - 1$ steps.

Guaranteeing additive approximation 1 - Scaffold of the proof

We will show that at each step $k = 1, .., 2n - 1$, the quantity of mass placed (i.e. $\sum_{i=1}^{k} s_i$) is enough to reconstruct half of the elements of either p or q considered so far (i.e. $\sum_{i=1}^{\lceil k/2 \rceil} p_i$ and $\sum_{i=1}^{\lceil k/2 \rceil} q_i$).

In particular, we will show that $\sum_{i=1}^{k} s_i \geq \min\left(\sum_{i=1}^{\lceil k/2 \rceil} p_i, \sum_{i=1}^{\lceil k/2 \rceil} q_i\right) \geq \sum_{i=1}^{k} hz_i$

Recall: For any M, coupling of p and q, $\quad H(s) \geq H(z)$

F. Cicalese, L. Gargano, U. Vaccaro, ISIT 2017

This implies that $\quad H(z) \leq H(s) \leq H(z) + 1$

Guaranteeing additive approximation 1 - Sketch of the proof

Let $p^{(k)} = \left(p_1^{(k)}, p_2^{(k)}, \dots, p_n^{(k)}\right)$ be the array of the residual pieces of p such that at least $\lceil k/2 \rceil$ elements are 0. We partition $p^{(k)}$ into four sets.

$i \le \lceil k/2 \rceil$ $\qquad$ $i > \lceil k/2 \rceil$

$$p^{(5)} = \left(p_1^{(5)}, p_2^{(5)}, p_3^{(5)}, p_4^{(5)}, p_5^{(5)}, p_6^{(5)}\right)$$

$p_1^{(5)}, p_2^{(5)}, p_3^{(5)}$	$p_1^{(5)}, p_2^{(5)}, p_3^{(5)}$	$p_4^{(5)}, p_5^{(5)}, p_6^{(5)}$	$p_4^{(5)}, p_5^{(5)}, p_6^{(5)}$
$p_i^{(k)} = 0$	$p_i^{(k)} > 0$	$p_i^{(k)} = 0$	$p_i^{(k)} > 0$
$U = \{1,3\}$	$W = \{2\}$	$R = \{4\}$	$V = \{5,6\}$

Sketch of the proof

$i \le \lceil k/2 \rceil$ $\quad$ $i > \lceil k/2 \rceil$

$$p^{(5)} = \left(p_1^{(5)}, p_2^{(5)}, p_3^{(5)}, p_4^{(5)}, p_5^{(5)}, p_6^{(5)}\right)$$

$U = \{1,3\}$ $\quad$ $W = \{2\}$ $\quad$ $R = \{4\}$ $\quad$ $V = \{5,6\}$

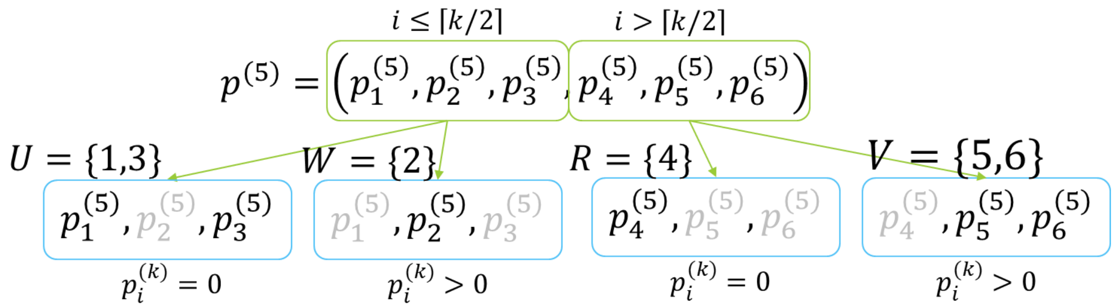
$p_1^{(5)}, p_2^{(5)}, p_3^{(5)}$ $\quad$ $p_1^{(5)}, p_2^{(5)}, p_3^{(5)}$ $\quad$ $p_4^{(5)}, p_5^{(5)}, p_6^{(5)}$ $\quad$ $p_4^{(5)}, p_5^{(5)}, p_6^{(5)}$

$p_i^{(k)} = 0$ $\quad$ $p_i^{(k)} > 0$ $\quad$ $p_i^{(k)} = 0$ $\quad$ $p_i^{(k)} > 0$

$$\sum_{i=1}^{k} s_i = p_1 + p_3 + p_2 - p_2^{(5)} + p_4 + p_5 - p_5^{(5)} + p_6 - p_6^{(5)}$$

$$\sum_{i=1}^{k} s_i = \sum_{i \in U} p_i + \sum_{i \in W} \left(p_i - p_i^{(k)}\right) + \sum_{i \in R} p_i + \sum_{i \in V} \left(p_i - p_i^{(k)}\right)$$

Sketch of the proof



$$\sum_{i=1}^{k} s_i = \sum_{i \in U} p_i + \sum_{i \in W} \left(p_i - p_i^{(k)}\right) + \sum_{i \in R} p_i + \sum_{i \in V} \left(p_i - p_i^{(k)}\right)$$

Observe that for all $i \in R$, it holds that for all $j \in W$, $p_i \ge p_j^{(k)}$. Hence, we have that

$$\sum_{i \in R} p_i + \sum_{i \in W} \left(p_i - p_i^{(k)}\right) \ge \sum_{i \in W} p_i$$

$$\sum_{i=1}^{k} s_i \ge \sum_{i \in U} p_i + \sum_{i \in W} p_i = \sum_{j \in \lceil k/2 \rceil} p_j \ge \sum_{i=1}^{k} h z_i$$

Then $$H(z) \le H(s) \le H(z) + 1$$

A general criterion for guaranteeing an additive approximation factor of 1

A general criterion for guaranteeing an additive approximation factor of 1

Given two marginal distributions $p = (p_1, p_2, \dots, p_n)$ and $q = (q_1, q_2, \dots, q_n)$, let M be a $n \times n$ coupling matrix such that:

$\forall i = 1, \dots, n$

$$\sum_{j=1}^{i} \sum_{k=1}^{i} m_{j,k} = \sum_{j=1}^{i} z_j$$

At most $2i - 1$ elements

$q \backslash p$	p_1	$\cdots$	p_i	p_{i+1}	$\cdots$	p_n
q_1	s_1					
$\vdots$		s_2	s_3			
q_i				s_4		
q_{i+1}		s_7	s_6	s_5		
$\vdots$					s_9	
q_n		s_8			s_{10}	s_{11}

If $t + 1$ elements there, then t elements in the matrix of size i are from z.

Then M is an additive approximation with factor 1 with respect to the optimal minimum-entropy coupling.

Thank you for your attention!