Generalizing Bottleneck Problems

Hsiang Hsu *, Shahab Asoodeh †, Salman Salamatian ‡, Flavio P. Calmon *

* SEAS, Harvard University, {hsianghsu, fcalmon}@g.harvard.edu, † University of Chicago, shahab@uchicago.edu,
‡ Massachusetts Institute of Technology, salmansa@mit.edu

Overview

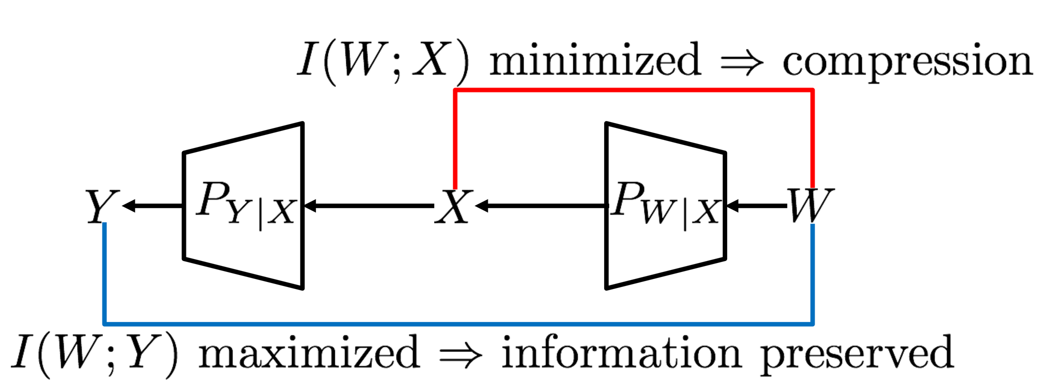
Given a pair of random variables $(X,Y)\sim P_{XY}$ and two convex functions f_1 and f_2, we introduce two bottleneck functionals as the lower and upper boundaries of the two-dimensional convex set that consists of the pairs $\left(I_{f_1}(W;X), I_{f_2}(W;Y)\right)$, where I_f denotes f-information and W varies over the set of all discrete random variables satisfying the Markov condition $W \to X \to Y$. Applying Witsenhausen and Wyner's approach, we provide an algorithm for computing boundaries of this set for f_1 and f_2, and discrete P_{XY}. In the binary symmetric case, we fully characterize the set when (i) $f_1(t)=f_2(t)=t\log t$, (ii) $f_1(t)=f_2(t)=t^2-1$, and (iii) f_1 and f_2 are both ℓ^β norm function for $\beta > 1$. We then argue that upper and lower boundaries in (i) correspond to Mrs. Gerber's Lemma and its inverse (which we call Mr. Gerber's Lemma), in (ii) correspond to estimation-theoretic variants of Information Bottleneck and Privacy Funnel, and in (iii) correspond to Arimoto Information Bottleneck and Privacy Funnel.

Keywords: Information bottleneck, privacy funnel, f-information, Mrs. Gerber's lemma, Arimoto's conditional entropy.

Two Special Cases of Bottleneck Problems

The Information Bottleneck (IB) [1]

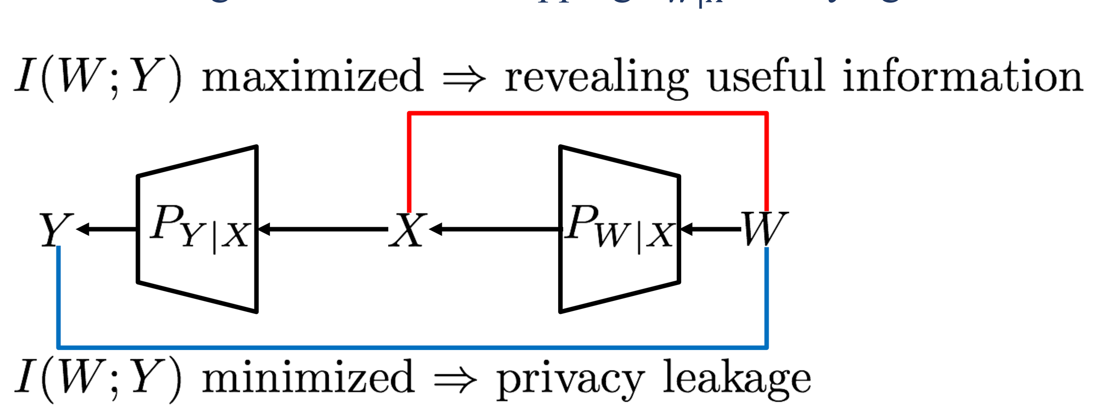
Given two correlated random variables X, Y and P_{XY}, the goal is to determine a mapping $P_{W|X}$



- Lagrangian functional:
  $$B(P_{XY},\lambda) = \max_{P_{W|X}} I(W;Y) - \lambda I(W;X)$$
- Clustering, natural language processing, analysis on the training process of deep neural nets [2, 3].
- E.g. X: MNIST handwritten digits, Y: labels, W: features

The Privacy Funnel (PR) [4]

The Privacy Funnel is a converse optimization problem comparing to the IB. The goal is to seek a mapping $P_{W|X}$ satisfying:



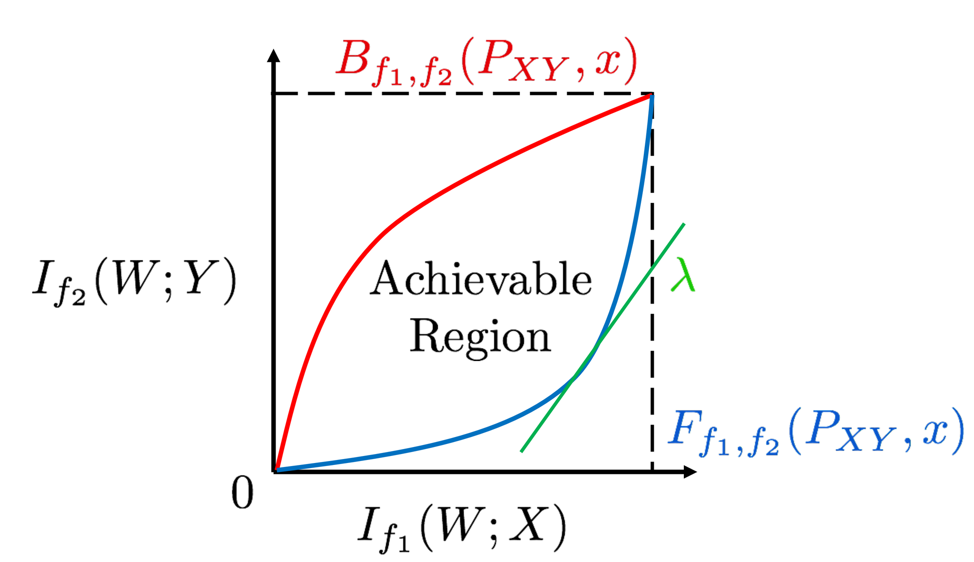
- Lagrangian functional:
  $$F(P_{XY},\lambda) = \min_{P_{W|X}} I(W;Y) - \lambda I(W;X)$$
- Useful in information-theoretic privacy
- E.g. X: Movie rating, Y: Political preference, W: Movie favor

Motivation

- No operational meanings
- f-divergence: $D_f(P||Q) = \mathbb{E}\left[f\left(\frac{dP}{dQ}\right)\right]$
- f-divergence. χ^2-divergence: MMSE, Total variation, Hellinger distance: hypothesis testing [5, 6].
- The method for the IB does not hold for f-divergence [7].

Goal

- f-information: $I_f(X;Y) = D_f(P_{XY}||P_XP_Y)$
- Bottleneck functional:
  $B_{f_1,f_2}(P_{XY},x) = \max_{W\to X\to Y} I_{f_2}(W;Y)$ such that $I_{f_1}(W;X) \le x$
- Funnel functional:
  $F_{f_1,f_2}(P_{XY},x) = \min_{W\to X\to Y} I_{f_2}(W;Y)$ such that $I_{f_1}(W;X) \ge x$
- Upper and lower boundaries: $\left\{\left(I_{f_1}(W;X), I_{f_2}(W;Y)\right)\right\}$



Geometric Properties of Bottleneck Problems

Generalizing Witsenhausen's and Wyner's results in 1975 [8]

- $W \to X \to Y, P_{Y|X} = T, p_w = P_{X|W}(X|W=w), \alpha_w = P_W(w)$

$$\begin{cases} C(T) = \left\{(q, \mathbb{E}[f(p_w)], \mathbb{E}[g(Tp_w)]): p_w \in \Delta_m, \sum_{w\in W} \alpha_w p_w = q\right\} \\ S(T) = \{(p, f(p), g(Tp)): p \in \Delta_m\} \end{cases}$$

- $C(T)$ = convex hull of $S(T)$
- Lower convex hull of $\phi(p,\lambda) = g(Tp) - \lambda f(p)$ equals to $\min\{y - \lambda x: (q,x,y) \in C(T)\}$

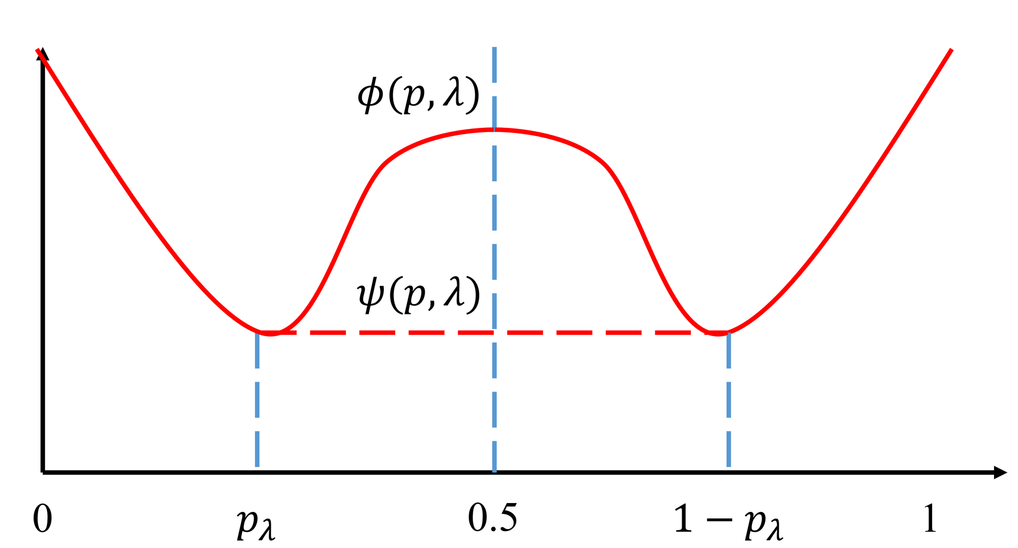


Figure 1: T follows BSC with crossover probability $\delta = 0.1$ and $P\{X=1\} = q = 0.5$.

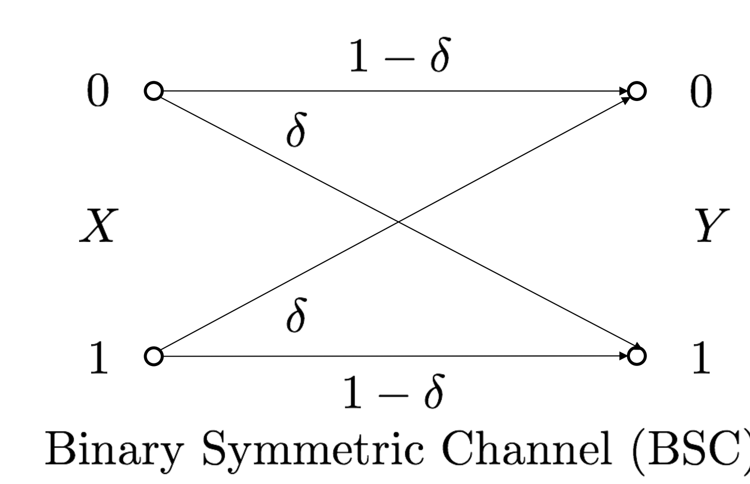
Main Results

Numerical Solutions for Generalized Bottleneck Problems

1. Compute $\phi(p,\lambda) = g(Tp) - \lambda f(p)$
2. Compute the lower convex envelope $\psi(p,\lambda)$
   1) If $\phi(q,\lambda) = \psi(q,\lambda)$, return (f(q), g(Tq))
   2) If not, find a convex combination for q, and apply the combination to (f(q), g(Tq))

Applications of the Generalized Bottleneck Problems

1. Estimation Bottleneck (EB) Problem
   - $f_1(t) = f_2(t) = t^2 - 1$.
   - $I_{f_1}(W;X) = \chi^2(W;X) = \mathbb{E}\left[\left(\frac{P_{WX}(W,X)}{P_W(W)P_X(X)}\right)\right] - 1$
   - $\chi^2(X;Y) = \sum_{i=1}^d \lambda_i(X;Y)$, the sum of Principal Inertia Components (PICs) [4], which is a direct bound of the largest MMSE of estimating X given Y
2. Estimation Privacy Funnel
   - $F_{\chi^2}(P_{XY},x) = \min_{W\to X\to Y} I_{f_2}(W;Y)$ such that $I_{f_1}(W;X) \ge x$
   - Privacy measured in MMSE



2. Mr. Gerber's Lemma [9]
   - $f_1(t) = f_2(t) = h_b$, the binary entropy function
   - Mrs. Gerber's Lemma
     $$L_T(\boldsymbol{q},x) = h_b\left(\delta * h_b^{-1}(x)\right), \forall x \in [0, h_b(q)]$$
   - Mr. Gerber's Lemma
     $$U_T(\boldsymbol{q},x) = a h_b(\delta * q/z) + \bar{a} h_b(\delta),$$
     $$x = a\, h_b(q/z)], z = \max(a, 2q), a \in [0,1]$$

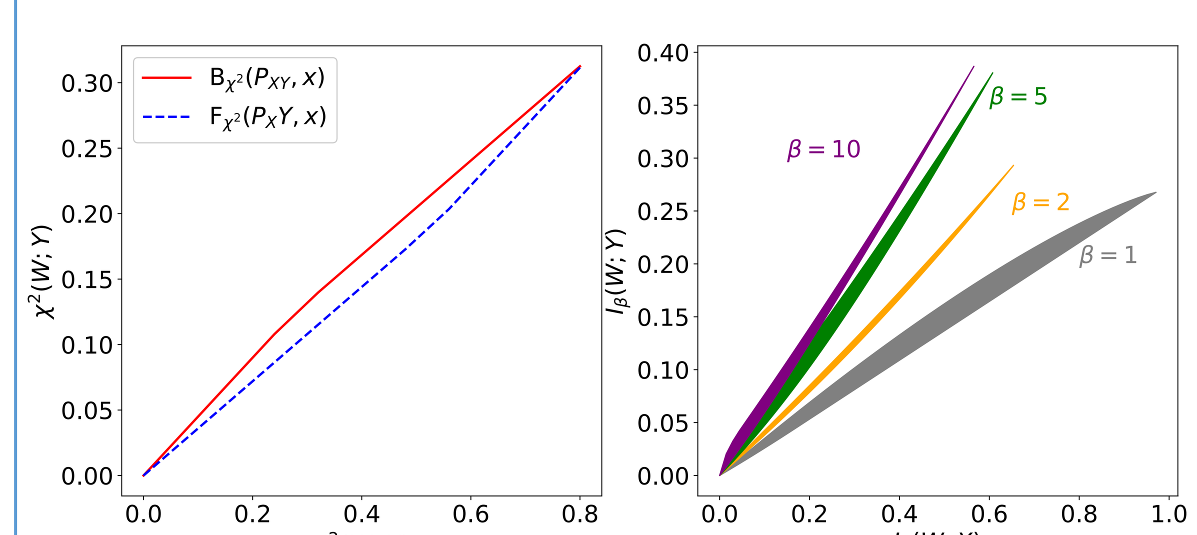


Figure 2: Left: The estimation bottleneck and privacy funnel. **Right**: Set of achievable pairs of Arimoto mutual information $\left(I_\beta(W;X), I_\beta(W;Y)\right)$

4. Arimoto's Mrs. And Mr. Gerber's Lemma
   - Ariosto's version of conditional Renyi entropy: for $\beta > 2$,
     $$H_\beta(X|W) = \frac{\beta}{1-\beta} log \sum_{w\in W} \alpha_w \|\boldsymbol{p}_w\|_\beta$$
   - $K_\beta(X) = exp\left\{\frac{1-\beta}{\beta} H_\beta(X)\right\}, \phi(p,\lambda) = K_\beta(\delta * p) - \lambda K_\beta(p)$
   - Arimoto's Mrs. Gerber's Lemma:
     $$\frac{\beta}{1-\beta} logy = \min_{W\to X\to Y} H_\beta(Y|W) \text{ such that } H_\beta(X|W) \ge \frac{\beta}{1-\beta}$$
     $$(x,y) \in \{\left(K_\beta(p), K_\beta(p*\delta)\right): 0 \le p \le q\}$$
   - Arimoto's Mr.s Gerber's Lemma:
     $$\frac{\beta}{1-\beta} logy = \max_{W\to X\to Y} H_\beta(Y|W) \text{ such that } H_\beta(X|W) \le \frac{\beta}{1-\beta}$$
     $$(x,y) \in \{\left(\bar{a} + aK_\beta\left(\frac{q}{z}\right), aK_\beta\left(\frac{q}{z} * \delta\right) + aK_\beta(\delta)\right): 0 \le p \le q, a \in [0,1], z = \max\{a, 2q\}\}$$

Remarks

1. Allows different f-divergence for X and Y
2. How to find the lower/ upper convex hull in high-dimensional space?
3. EB/ estimation privacy funnel: new clustering techniques
4. The Arimoto's Mrs. And Mr. Gerber's lemma for network information theory: a new form of the Entropy Power Inequality (EPI)

Selected References

[1] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in Proc. of IEEE Allerton, 2000.

[2] N. Tishby and N. Zaslavsky, "Deep learning and the information bottleneck principle," in Proc. of IEEE ITW , 2015.

[3] R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," arXiv preprint arXiv:1703.00810, 2017.

[4] F. P. Calmon, A. Makhdoumi, M. Medard, M. Varia, M. Christiansen, and K. R. Duffy, "Principal inertia components and applications," IEEE Trans. Inf. Theory , vol. 63, no. 9, pp. 5011–5038, 2017.

[5] C. Nair, "Upper concave envelopes and auxiliary random variables," Int. J. Adv. Eng. Sci. Appl. Math., vol. 5, no. 1, pp. 12–20, 2013.

[6] I. Sason and S. Verd´u, "Arimoto-R´enyi conditional entropy and bayesian m-ary hypothesis testing," arXiv preprint arXiv:1701.01974, 2017.

[7] P. Harremoes and N. Tishby, "The information bottleneck revisited or how to choose a good distortion measure," in Proc. of IEEE ISIT, 2007.

[8] H. Witsenhausen and A. Wyner, "A conditional entropy bound for a pair of discrete random variables," IEEE Trans. Inf. Theory, vol. 21, no. 5, pp. 493–501, 1975.

[9] A. El Gamal and Y.-H. Kim, Network information theory. Cambridge university press, 2011.

The materials in this poster have been accepted by 2018 IEEE International Symposium on Information Theory (ISIT). An extended version can be accessed at https://arxiv.org/abs/1802.05861