Characterizing the Region of Entropic Vectors via Information Geometry

John MacLaren Walsh
Department of Electrical and Computer Engineering
Drexel University
Philadelphia, PA
jwalsh@ece.drexel.edu

Thanks to NSF CCF-1016588 & NSF CCF-1053702.

Outline

- Entropic Vectors Review
  - What Are They?
  - Why are They Important?
    - Unconstrained Importance in Network Coding Capacity Regions
    - Constrained Importance in Multiterminal Information Theory
  - What do we know about them? Open Problems/Issues
- Information Geometry "Review"
  - What is it?
  - Places it has been shown to be useful
- Relating These Two Disciplines
  - A information projection construction of the set of entropic vectors
- Conclusions

Outline

- **Entropic Vectors Review**
  - What Are They?
  - Why are They Important?
    * Unconstrained Importance in Network Coding Capacity Regions
    * Constrained Importance in Multiterminal Information Theory
  - What do we know about them? Open Problems/Issues
- Information Geometry "Review"
  - What is it?
  - Places it has been shown to be useful
- Relating These Two Disciplines
  - A information projection construction of the set of entropic vectors
- Conclusions

Entropic Vectors – What are they?

1. Let $\mathbf{X} = (X_1, \ldots, X_N)$ be N discrete random variables with finite support.
2. Let $h(\mathbf{X}_\mathcal{A})$ be the entropy of the subset of rvs $\mathbf{X}_\mathcal{A} = (X_i, i \in \mathcal{A})$ for some non-empty subset $\mathcal{A} \subseteq \{1, \ldots, N\} \equiv [N]$.
3. Let $\mathbf{h} = (h(\mathbf{X}_\mathcal{A}),\ \mathcal{A} \subseteq [N])$ be the vector of entropies of each non-empty subset $\mathcal{A} \subseteq [N]$. Note $\mathbf{h}$ has $2^N - 1$ entries.
   - Example: for $N = 3$, $\mathbf{h} = (h_1, h_2, h_3, h_{12}, h_{13}, h_{23}, h_{123})$.
4. A vector $\mathbf{h} \in \mathbb{R}^{2^N-1}$ is called entropic if its elements are the entropies for some joint distribution $p_\mathbf{X}$ on the N rvs $\mathbf{X}$.
5. The entropy vector region (EVR) $\bar{\Gamma}^*_N$ is the closure of the set of all entropic vectors. It's a convex cone [1].
6. Normalize by the number of bits for the support m: $\tilde{\mathbf{h}} = \mathbf{h}/\log_2 m$, and define $\bar{\Omega}^*_N$ as the set of normalized entropy vectors [2, 3]

Entropic Vectors – Why are they Important?

- **Network Coding**: Capacity region of a multi-source network under network coding formed from a linear map of $\bar{\Gamma}_N^*$ intersected w/ a series of polyhedral constraints [4]. For every Non-Shannon face there is a network whose capacity region depends on this face [5, 6, 7].

- **Multiterminal Information Theory** More generally, if we allow extra constraints $\mathcal{C}$ on the random variables, then all multiterminal rate regions are expressible in terms of a linear map of $\bar{\Gamma}_N^*(\mathcal{C})$

Entropic Vectors – What do we know? – Outer Bounds

- **Yeung & Zhang Non-Shannon [8, 9]**
  - Showed that Shannon Outer bound

$$\Gamma_n := \left\{ \mathbf{h} \;\middle|\; \begin{array}{c} h_{\mathcal{A}} + h_{\mathcal{B}} \geq h_{\mathcal{A}\cap\mathcal{B}} + h_{\mathcal{A}\cup\mathcal{B}} \\ \mathcal{A} \subseteq \mathcal{B} \implies h_{\mathcal{A}} \leq h_{\mathcal{B}} \end{array} \right\} \tag{1}$$

    ($=$ matroid rank function cond. [1]) was not tight for $N \geq 4$ via new inequality

- **Dougherty, Freiling, & Zeger [10, 11]** & Others [12]
  - More Non-Shannon Information Inequalities
  - Construction of Codes via Representable Matroids

- **Matùš [13, 14]**
  - Showed that $\bar{\Gamma}_4^*$ is not polyhedral

- **Technique for creating all of these Non-Shannons [15]:**
  - Create one or more R.V.s in terms of the originals (d-copy over), and look at the implications of Shannon inequalities among this larger collection of variables on the subset of original variables.

Entropic Vectors – What do we know? – Inner Bounds

- **Matroid Representation Based [16]:**
  - Binary matroids: (convex hull of rank functions of) $\forall N$. Not tight $N \geq 4$.
  - Ternary Matroids: (convex hull of rank functions of) $\forall N$.
  - Regular Matroids: (both binary and ternary $=$ rep. over *any* field)

  Algorithm: check all possible rank functions for spec. forbidden minors, then take convex hull of remaining

- **convex hull of representable (over some field) matroids**
  - explicitly known only for $N \leq 6$. (4=Ingleton [17, 18, 19], 5,6 new inequalities [20, 21, 22])
  - Not a fully tight inner bound for $N \geq 4$.

- **Binary entropic vectors**
  - Membership test via a finite terminating numerical algorithm for any N [23, 24, 25, 26].
  - Contains points outside Ingleton (Representable matroids) at $N = 4$.
  - together w/ vertex enumeration can list extreme points of any outer bound which are extreme points of convex hull of binary entropic vectors.

Outline

- Entropic Vectors Review
  - What Are They?
  - Why are They Important?
    - Unconstrained Importance in Network Coding Capacity Regions
    - Constrained Importance in Multiterminal Information Theory
  - What do we know about them? Open Problems/Issues
- **Information Geometry "Review"**
  - What is it?
  - Places it has been shown to be useful
- Relating These Two Disciplines
  - A information projection construction of the set of entropic vectors
- Conclusions

Information Geometry [27] – What is it? - Notation

- Overall idea: treat family of probability distributions as a differentiable manifold: $p(x;\xi)$ is parameterized by ξ
- Endow w/ Riemannian metric (inner product between Tangent vectors) given by Fisher Information Matrix $g_{i,j}(\xi) = \mathbb{E}_\xi[\partial_i \ell_\xi \partial_j \ell_\xi]$ w/ $\ell_\xi = \log p(x;\xi)$, $\partial_i = \frac{\partial}{\partial \xi_i}$.
- Select α-affine connections $\nabla^{(\alpha)}$ such that $\left\langle \nabla^{(\alpha)}_{\partial_i} \partial_j, \partial_k \right\rangle = \Gamma^{(\alpha)}_{ij,k}$

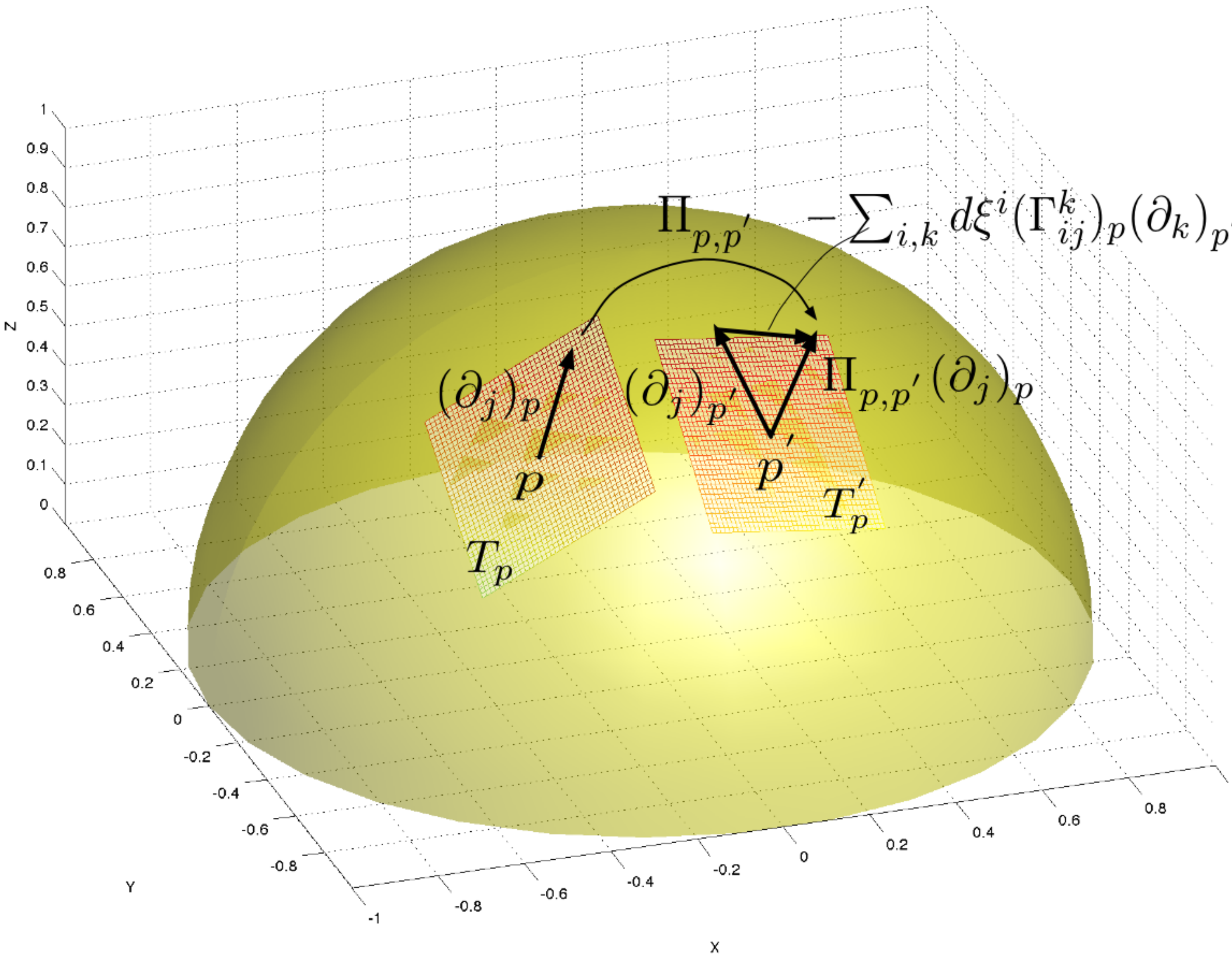
$$\Gamma^{(\alpha)}_{ij,k} = \mathbb{E}\left[\left(\partial_i \partial_j \ell_\xi + \frac{1-\alpha}{2} \partial_i \ell_\xi \partial_j \ell_\xi\right)(\partial_k \ell_\xi)\right] \tag{2}$$

- purpose of affine connection: define parallel translation $\Pi_{p,p'} : T_p \to T_{p'}$ to correspond tangent vectors along curves $\gamma : [a,b] \to \mathcal{P}$

$$\Pi_{\gamma(t),\gamma(t+dt)}(X(t)) = \sum_{ijk} \left\{ X^k(t) - dt \dot{\gamma}^i(t) X^j(t) \left(\Gamma_{ij,k}\right)_{\gamma(t)} \right\} (\partial_k)_{\gamma(t+dt)} \tag{3}$$

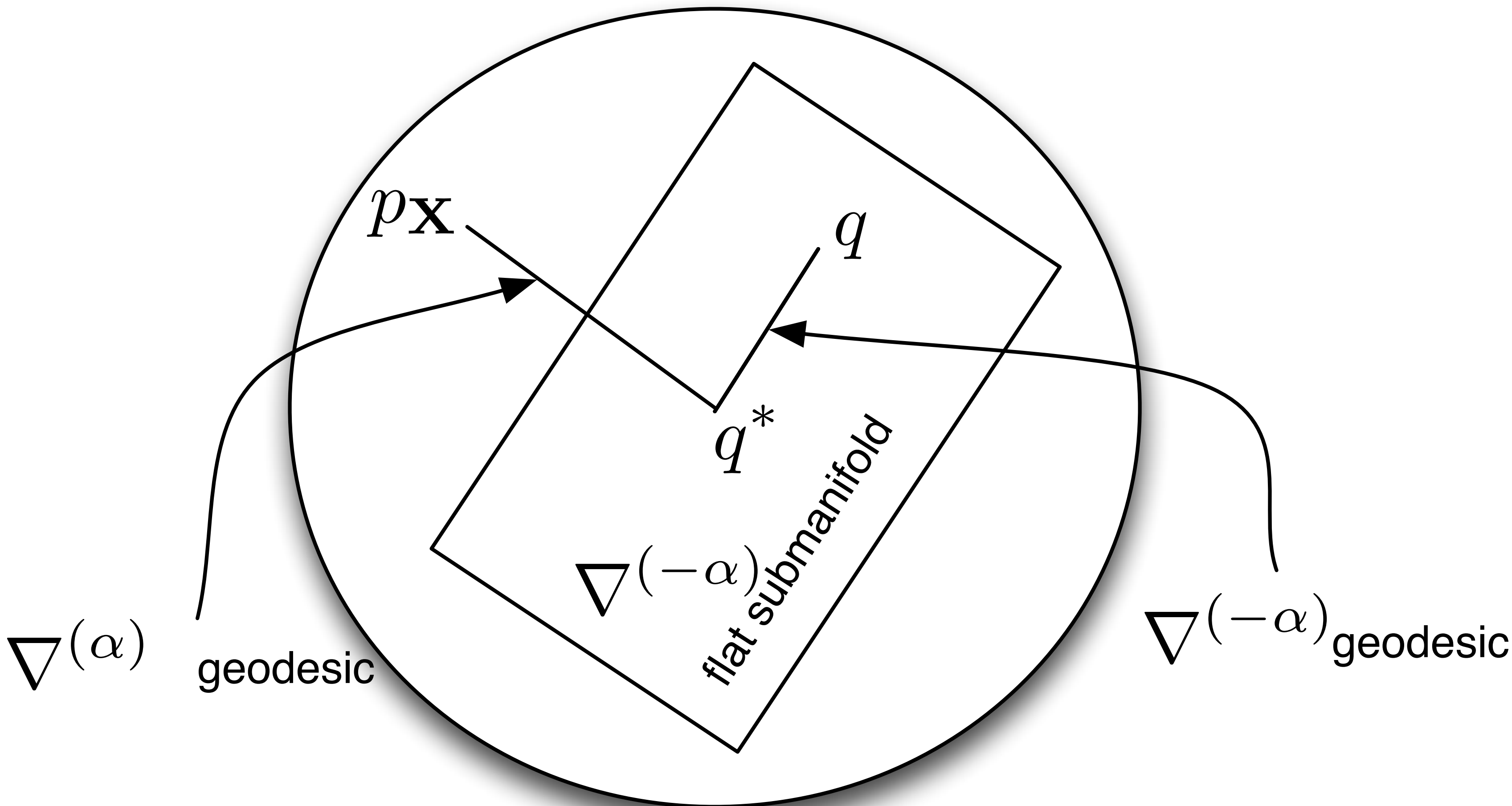
- Curve w/ tangent vector transported by parallel transl. w/ $\nabla^{(\alpha)}$ is $\nabla^{(\alpha)}$ *geodesic*
- If there is a coordinate system in which every parallel translation under $\nabla^{(\alpha)}$ leaves coefficients in Tangent vector unchanged, the manifold is said to be α-flat, and associated coordinate system is an affine coordinate system.
- $\nabla^{(\alpha)}$ has property $\langle X, Y \rangle_p = \langle \Pi^{(\alpha)}_{p,p'}(X), \Pi^{(-\alpha)}_{p,p'}(Y) \rangle_{p'}$

Information Geometry [27] – What is it? - Picture - Parallel Translation



$$\nabla_{\partial_i}\partial_j = \sum_k \Gamma_{ij,k}\partial_k \quad \Gamma_{ij,k} = 0 \text{ if “flat”} \tag{4}$$

Information Geometry [27] – What is it? - Picture - Information Projection



$$D^{(\alpha)}(p_{\mathbf{x}}||q) = D^{(\alpha)}(p_{\mathbf{x}}||q^*) + D^{(\alpha)}(q^*||q)$$

Information Geometry [27] – What is it? - Examples

- 2 flat coordinate systems (associated with $\alpha = -1, 1$) for finite discrete $\mathbf{X} \in \mathcal{X} = \{x_0, x_1, x_2, \ldots, x_N\}$
  - e-flat: exponential family: $q_1(x), q_2(x) \in \mathcal{E} \implies c(\lambda) q_1^{\lambda}(x) q_2^{1-\lambda}(x) \in \mathcal{E}$

$$p_X(x) = \exp\left(\boldsymbol{\theta}^T \mathbf{t}(x) - \psi(\boldsymbol{\theta})\right) \tag{5}$$

with $\theta_i = \log \frac{p_X(x_i)}{p_X(x_0)}, i \in \{1, \ldots, N\}$, $\psi(\boldsymbol{\theta}) = \log\left(1 + \|\exp(\boldsymbol{\theta})\|_1\right)$

  - m-flat: mixture family: $q_1(x), q_2(x) \in \mathcal{M} \implies \lambda q_1(x) + (1-\lambda) q_2(x) \in \mathcal{M}$

$$p_X(x) = \boldsymbol{\eta} \cdot \mathbf{t}(x) + (1 - \mathbf{1}^T \boldsymbol{\eta}) \mathbf{1}_{x = x_0}, \quad \eta_i = p_X(x_i) \tag{6}$$

- Legendre Transform Relationship
- KL Divergence (Relative Entropy)

Information Geometry [27] – Examples, Cont'd

- **e-flat submanifold:** set of all product distributions

$$\mathcal{E}_0 = \left\{ p_{\mathbf{X}} \,\middle|\, p_{\mathbf{X}}(x_1, \ldots, x_N) = \prod_{i=1}^{N} p_{X_i}(x_i) \right\} \tag{7}$$

- **m-flat submanifold:** set of joint distributions with given marginals

$$\mathcal{M}_0 = \left\{ p_{\mathbf{X}} \,\middle|\, \sum_{\mathbf{x}_{\setminus i}} p_{\mathbf{X}}(\mathbf{x}) = q_i(x_i) \quad \forall i \in \{1, \ldots, N\} \right\} \tag{8}$$

- **Information Projections & Pythagorean Relation:**

$$q^* = \arg\min_{q \in \mathcal{E}_0} D(p_{\mathbf{X}}||q), \quad D(p_{\mathbf{X}}||q) = D(p_{\mathbf{X}}||q^*) + D(q^*||q) \ \forall q \in \mathcal{E}_0 \tag{9}$$

$$q^* = \arg\min_{q \in \mathcal{M}_0} D(q||p_{\mathbf{X}}), \quad D(q||p_{\mathbf{X}}) = D(q^*||p_{\mathbf{X}}) + D(q||q^*) \ \forall q \in \mathcal{M}_0 \tag{10}$$

Information Geometry [27] – What has it been used for?

- re-interpretation of EM algorithm [27]
- acceleration of Blahut Arimoto algorithm [28]
- learning algorithms in Neural Networks [29]
- analysis of Belief propagation & Turbo Decoding [30, 31, 32, 33]

Outline

- Entropic Vectors Review
  - What Are They?
  - Why are They Important?
    * Unconstrained Importance in Network Coding Capacity Regions
    * Constrained Importance in Multiterminal Information Theory
  - What do we know about them? Open Problems/Issues
- Information Geometry "Review"
  - What is it?
  - Places it has been shown to be useful
- **Relating These Two Disciplines**
  - A information projection construction of the set of entropic vectors
- Conclusions

Relating These – Constructing Entropic Vectors via Information Geometry

Easy to relate Shannon entropy to rel. entropy/ KL Divergence:

$$
\begin{aligned}
D(p_{\mathbf{X}} || \mathcal{U}_{|\boldsymbol{\mathcal{X}}|}) &= \sum_{\mathbf{x} \in \boldsymbol{\mathcal{X}}} p_{\mathbf{X}}(\mathbf{x}) \log_2 \left(\frac{p_{\mathbf{X}}(\mathbf{x})}{1/|\boldsymbol{\mathcal{X}}|} \right) && (11) \\
&= \log_2(|\boldsymbol{\mathcal{X}}|) - H(p_{\mathbf{X}}) = H(\mathcal{U}_{\boldsymbol{\mathcal{X}}}) - H(p_{\mathbf{X}}) && (12)
\end{aligned}
$$

...perhaps put solution to minimization here as well ..

Relating These – Constructing Entropic Vectors via Information Geometry

Next consider the family of distributions

$$\mathcal{H}_i := \left\{ p_{\mathbf{X}} \middle| p(\mathbf{X}) = \frac{1}{|\mathcal{X}_i|} q(\mathbf{X}_{\setminus i}), \text{ some } q(\mathbf{X}_{\setminus i}) \right\} \tag{13}$$

Observe:

- $\mathcal{U}_{\boldsymbol{\mathcal{X}}} \in \mathcal{H}_i$
- $\mathcal{H}_i$ is *both* an e-flat and m-flat submanifold.
- Defining $q^*_{\mathcal{H}_i}(p_{\mathbf{X}}) = \arg\min_{q \in \mathcal{H}_i} D(p_{\mathbf{X}}||q)$, have Pythagorean relation:

$$D(p_{\mathbf{X}}||\mathcal{U}_{\boldsymbol{\mathcal{X}}}) = \underbrace{D(p_{\mathbf{X}}||q^*_{\mathcal{H}_i}(p_{\mathbf{X}}))}_{\log_2 |\mathcal{X}_i| - H(X_i|\mathbf{X}_{\setminus i})} + \underbrace{D(q^*_{\mathcal{H}_i}(p_{\mathbf{X}})||\mathcal{U}_{\mathcal{X}})}_{\log_2 |\boldsymbol{\mathcal{X}}| - \log_2 |\mathcal{X}_i| - H(\mathbf{X}_{\setminus i})} \tag{14}$$

(erm... $H(\mathbf{X}) = H(X_i) + H(\mathbf{X}_{\setminus i}|X_i)$ tyco)

Moving this around, we have

$$H(\mathbf{X}_{\setminus i}) = D(p_{\mathbf{X}}||q^*_{\mathcal{H}_i}(p_{\mathbf{X}})) - D(p_{\mathbf{X}}||\mathcal{U}_{\boldsymbol{\mathcal{X}}}) + \log_2 |\boldsymbol{\mathcal{X}}| - \log_2 |\mathcal{X}_i| \tag{15}$$

Relating These – Constructing Entropic Vectors via Information Geometry

Generalizing this idea, consider the family of distributions

$$\bigcap_{i \in \mathcal{A}^c} \mathcal{H}_i = \left\{ p_{\mathbf{X}} = \frac{q(\mathbf{X}_{\mathcal{A}})}{\prod_{i \in \mathcal{A}^c} |\mathcal{X}_i|} \right\} \tag{16}$$

Observe:

- $\mathcal{U}_{\boldsymbol{\mathcal{X}}} \in \bigcap_{i \in \mathcal{A}^c} \mathcal{H}_i$
- $\bigcap_{i \in \mathcal{A}^c} \mathcal{H}_i$ is *both* an e-flat and m-flat submanifold
- Defining $q^*_{\mathcal{A}}(p_{\mathbf{X}}) = \arg\min_{q \in \bigcap_{i \in \mathcal{A}^c} \mathcal{H}_i} D(p_{\mathbf{X}} || q)$, have Pythagorean relation:

$$D(p_{\mathbf{X}} || \mathcal{U}_{\boldsymbol{\mathcal{X}}}) = \underbrace{D(p_{\mathbf{X}} || q^*_{\mathcal{A}}(p_{\mathbf{X}}))}_{\sum_{i \in \mathcal{A}^c} \log_2 |\mathcal{X}_i| - H(\mathbf{X}_{\mathcal{A}^c} | \mathbf{X}_{\mathcal{A}})} + \underbrace{D(q^*_{\mathcal{A}}(p_{\mathbf{X}}) || \mathcal{U}_{\boldsymbol{\mathcal{X}}})}_{\log_2 |\boldsymbol{\mathcal{X}}| - \sum_{i \in \mathcal{A}^c} \log_2 |\mathcal{X}_i| - H(\mathbf{X}_{\mathcal{A}})} \tag{17}$$

(erm... $H(\mathbf{X}) = H(\mathbf{X}_{\mathcal{A}}) + H(\mathbf{X}_{\mathcal{A}^c} | \mathbf{X}_{\mathcal{A}})$ tyco)

From which we observe that

$$H(\mathbf{X}_{\mathcal{A}}) = D(p_{\mathbf{X}} || q^*_{\mathcal{A}}(p_{\mathbf{X}})) - D(p_{\mathbf{X}} || \mathcal{U}_{\boldsymbol{\mathcal{X}}}) - \sum_{i \in \mathcal{A}^c} \log_2 |\mathcal{X}_i| + \log_2 |\boldsymbol{\mathcal{X}}| \tag{18}$$

Relating These – Constructing Entropic Vectors via Information Geometry

Defining the set function (then stack into a vector $\mathbf{d}$)

$$d_{\mathcal{A}} := \min_{q \in \bigcap_{i \in \mathcal{A}^c} \mathcal{H}_i} D(p_{\mathbf{X}}||q) = D(p_{\mathbf{X}}||q_{\mathcal{A}}(p_{\mathbf{X}})) \tag{19}$$

It is evident from the relation we derived

$$H(\mathbf{X}_{\mathcal{A}}) = D(p_{\mathbf{X}}||q^*_{\mathcal{A}}(p_{\mathbf{X}})) - D(p_{\mathbf{X}}||\mathcal{U}_{\boldsymbol{\mathcal{X}}}) - \sum_{i \in \mathcal{A}^c} \log_2 |\mathcal{X}_i| + \log_2 |\boldsymbol{\mathcal{X}}| \tag{20}$$

that

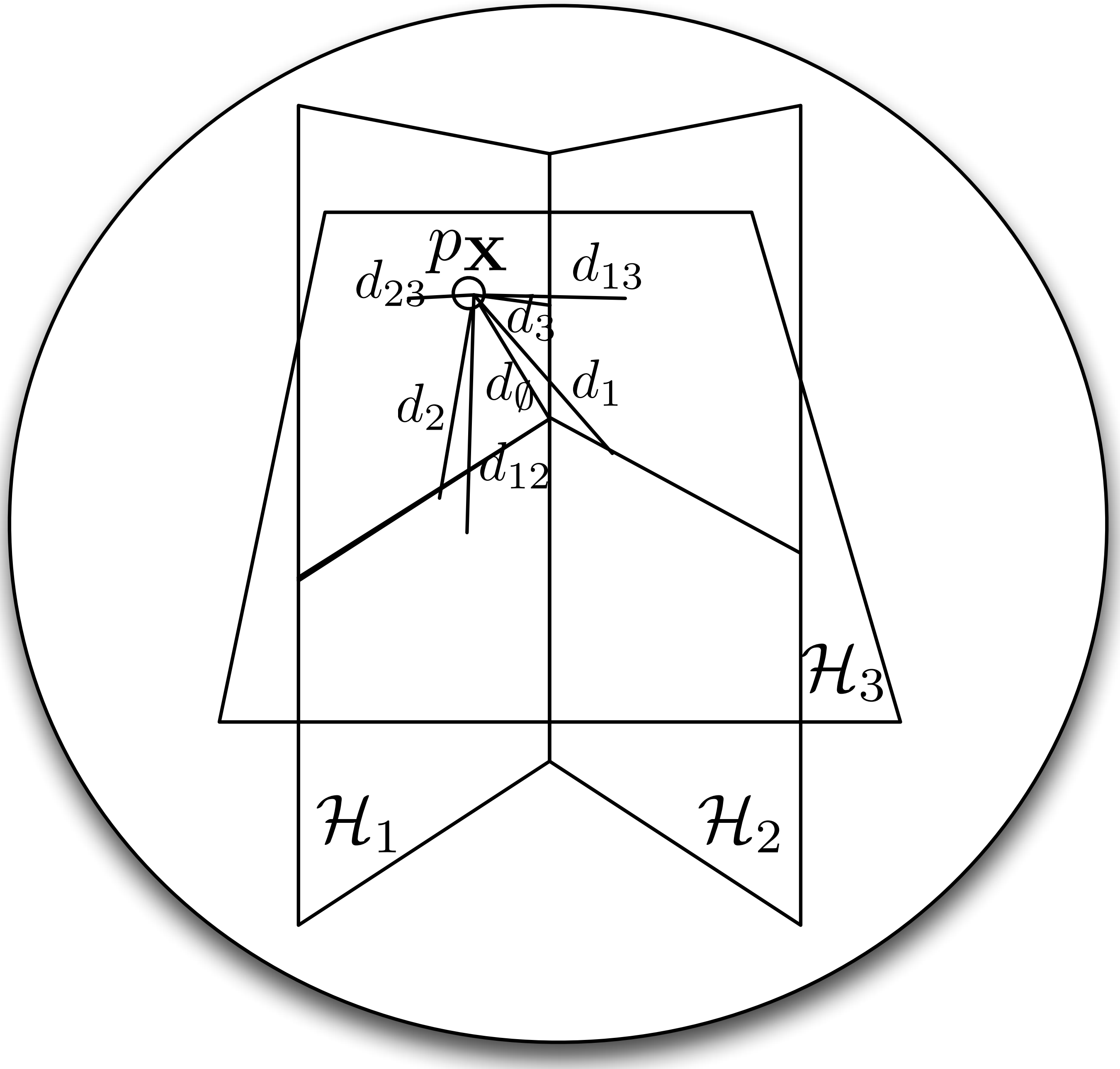
$$h_{\mathcal{A}} = d_{\mathcal{A}} - d_{[N]} - \sum_{i \in \mathcal{A}^c} \log_2 |\mathcal{X}_i| + \log_2 |\boldsymbol{\mathcal{X}}| \tag{21}$$

thus we can express entropic vector in terms of $\mathbf{d}$ via

$$\mathbf{h}(\mathbf{d}) = \mathbf{A}\mathbf{d} + \mathbf{b} \tag{22}$$

Region of entropic vectors is affine transformation of region of simultaneous divergences between submanifolds $\mathcal{H}_i$ and their intersections!

Constructing Entropic Vectors via Information Geometry



References

[1] Raymond W. Yeung, *Information Theory and Network Coding*. Springer, 2008.

[2] B. Hassibi and S. Shadbakht, "Normalized Entropy Vectors, Network Information Theory and Convex Optimization," in *IEEE Information Theory Workshop on Information Theory for Wireless Networks*, July 2007, pp. 1 – 5.

[3] ——, "On a Construction of Entropic Vectors Using Lattice-Generated Distributions," in *IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 501 – 505.

[4] Xijin Yan, Raymond W. Yeung, and Zhen Zhang, "The Capacity Region for Multi-source Multi-sink Network Coding," in *IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 116 – 120.

[5] T. Chan and A. Grant, "Entropy Vectors and Network Codes," in *IEEE International Symposium on Information Theory*, June 2007.

[6] T. Chan and A. Grant, "Dualities between entropy functions and network codes," in *Fourth Workshop on Network Coding, Theory and Applications (NetCod)*, January 2008.

[7] ——, "Mission impossible: computing the network coding capacity region," in *IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 320–324.

[8] Raymond W. Yeung, "A Framework for Linear Information Inequalities," *IEEE Transactions on Information Theory*, vol. 43, no. 6, Nov. 1997.

[9] Zhen Zhang and Raymond W. Yeung, "On Characterization of Entropy Function via Information Inequalities," *IEEE Transactions on Information Theory*, vol. 44, no. 4, July 1998.

[10] R. Dougherty, C. Freiling, and K. Zeger, "Six new non-Shannon information inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, July 2006, pp. 233–236.

[11] R. Dougherty, C. Freiling, and K. Zeger, "Networks, Matroids, and Non-Shannon Information Inequalities," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 1949–1969, June 2007.

[12] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-Shannon-type inequalities for entropies," *Communication in Information and Systems*, vol. 2, no. 2, pp. 147–166, December 2002.

[13] František Matúš, "Infinitely Many Information Inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, June 2007, pp. 41–44.

[14] ——, "Entropy Functions, information inequalities, and polymatroids," in *Applications of Matroid Theory and Combinatorial Optimization to Information and Coding Theory*, Banff International Research Station, Aug. 2009. [Online]. Available: http://temple.birs.ca/~09w5103/matus_09w5103_talk.pdf

[15] Weidong Xu, Jia Wang, Jun Sun, "A projection method for derivation of non-Shannon-type information inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, 2008, pp. 2116 – 2120.

[16] James Oxley, *Matroid Theory, 2nd. Ed.* Oxford University Press, 2011.

[17] A. W. Ingleton, "Representation of Matroids," in *Combinatorial Mathematics and its Applications*, D. J. A. Welsh, Ed. San Diego: Academic Press, 1971, pp. 149–167.

[18] F. Matúš and M. Studený, "Conditional Independences among Four Random Variables I," *Combinatorics, Probability and Computing*, no. 4, pp. 269–278, 1995.

[19] D. Hammer, A. Romashschenko, A. Shen, N. Vereshchagin, "Inequalities for Shannon Entropy and Kolmogorov Complexity," *Journal of Computer and System Sciences*, vol. 60, pp. 442–464, 2000.

[20] Ryan Kinser, "New Inequalities for Subspace Arrangements," *New Inequalities for Subspace Arrangements*, vol. 188, no. 1, pp. 152–161, Jan. 2011.

[21] Randall Dougherty, Chris Freiling, Kenneth Zeger, "Linear rank inequalities on five or more variables," submitted to SIAM J. Discrete Math. arXiv:0910.0284.

[22] Terence Chan, Alex Grant, Doris Kern, "Existence of new inequalities for representable polymatroids," in *IEEE International Symposium on Information Theory (ISIT)*, 2010, pp. 1364 – 1368.

[23] J. M. Walsh and S. Weber, "A Recursive Construction of the Set of Binary Entropy Vectors and Related Inner Bounds for the Entropy Region," *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6356–6363, Oct. 2011. [Online]. Available: http://dx.doi.org/10.1109/TIT.2011.2165817

[24] John MacLaren Walsh and Steven Weber, "Relationships Among Bounds for the Region of Entropic Vectors in Four Variables," in *2010 Allerton Conference on Communication, Control, and Computing*, Sept. 2010. [Online]. Available: http://www.ece.drexel.edu/walsh/allerton10.pdf

[25] ——, "Tunable Inner Bounds for the Region of Entropy Vectors," Feb. 2010, 2010 Information Theory and Applications Workshop, University of California San Diego. [Online]. Available: http://www.ece.drexel.edu/walsh/walshITA10.pdf

[26] ——, "A Recursive Construction of the Set of Binary Entropy Vectors," in *Forty-Seventh Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2009, pp. 545–552. [Online]. Available: http://www.ece.drexel.edu/walsh/binEntSetConf-v5-JW.pdf

[27] S. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Society Translations of Mathematical Monographs, 2004, vol. 191.

[28] G. Matz and P. Duhamel, "Information Geometric Formulation and Interpretation of Accelerated Blahut-Arimoto-Type Algorithms," in *IEEE Information Theory Workshop (ITW-2004)*, Oct. 2004, pp. 66 – 70.

[29] S. Amari, "Information geometry of the EM and em algorithms for neural networks," *Neural Networks*, vol. 8, no. 9, pp. 1379–1408, 1996.

[30] S. Ikeda, T. Tanaka, and S. Amari, "Stochastic reasoning, free energy and information geometry," *Neural Computation*, vol. 16, no. 9, pp. 1779–1810, Sept. 2004.

[31] ——, "Information geometry of turbo and low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1097 – 1114, June 2004.

[32] B. Muquet, P. Duhamel, and M. de Courville, "Geometrical interpretations of iterative 'turbo' decoding," in *Proceedings ISIT*, June 2002.

[33] A. J. Grant, "Information geometry and iterative decoding," in *Proceedings IEEE Communication Theory Workshop*, may 1999.