1/ Geometry of \(w\)-mixtures:

In probability, a statistical mixture:
\[
m(x; w) = m(x; \eta) = \sum_{i=1}^{k-1} \eta_i p_i(x) + \left(1 - \sum_{i=1}^{k-1} \eta_i\right) p_0(x)
\]

In information geometry, a mixture family:
\[
\mathcal{M} = \{m(x; \eta) = \sum_{i=1}^{k-1} \eta_i f_i(x) + c(x), \quad \eta \in \Delta_D^\circ\}
\]

\(f_i(x) = p_i(x) - p_0(x)\) for \(i \in [D]\), \(c(x) = p_0(x)\)

2/ Dually flat space from a strictly convex and smooth functional (here, statistical):

Shannon differential entropy of a mixture \(m(x)\) (concave):
\[
h(m) := -\int m(x) \log m(x) \, d\mu(x)
\]

Shannon information as a Bregman generator (convex):
\[
F^*(\eta) = \int m(x; \eta) \log m(x; \eta) \, d\mu(x)
\]

Dual Legendre convex conjugate (cross-entropy):
\[
F(\theta) = -\int p_0(x) \log m(x; \eta) \, d\mu(x)
\]

Dual parameterization of \(\eta\)-mixtures:
\[
\theta^*(\eta) = (\nabla_{\eta} F^*(\eta)), \quad \int (p_i(x) - p_0(x)) \log m(x; \eta) \, d\mu(x)
\]

3/ Applications:

- Optimal KL-averaging integration:

**Theorem:** The KL-averaging integration of \(w\)-mixtures performed optimally without information loss.
\[
\hat{\eta} = \text{argmin}_{\eta} \sum_{i=1}^{m} KL(m(\eta_i) : m(\eta)) \equiv \sum_{i=1}^{m} B_{F^*}(\eta_i : \eta)
\]
\[
\Rightarrow \hat{\eta} = \frac{1}{m} \sum_{i=1}^{m} \tilde{\eta}_i \quad \text{(Bregman right centroid indep. of } F^*)
\]

4/ Divergence inequalities and family closure:
\[
m^\epsilon(p,q) = (1-\epsilon)p + \epsilon q = p + \epsilon (q-p) = m^{1-\epsilon}(q:p) \quad \text{for } \epsilon \in [0,1], \quad I_f(p:q) := I_f(m^\epsilon(p,q) : m^\epsilon(q,p))
\]

The \(f\)-divergence \(I_f(m(x; w) : m(x; w'))\) between any two \(w\)-mixtures is upper bounded by
\[
I_f(w: w') = \sum_{i=0}^{k-1} w_i f\left(\frac{w_i}{w_i'}\right).
\]
\[
I_f(p:q) \leq (1-\epsilon)I_f(p:p) + \epsilon I_f(q:p),
\]
\[
I_f^\epsilon(p:q) \leq (1-\epsilon)f\left(\frac{\epsilon}{1-\epsilon}\right) + \epsilon f\left(\frac{1-\epsilon}{\epsilon}\right).
\]

**Fact:** Kullback-Leibler divergence between two \(\eta\)-mixtures (or \(w\)-mixtures) is equivalent to a Bregman divergence defined for the Shannon negentropy generator on the \(\eta\)-parameters.

**Corollary:** The KL between \(w\)-Gaussian mixture model is a Bregman divergence for the Shannon negentropy generator.

\[
KL(m_1 : m_2) = \int m(x; \eta_1) \log \frac{m(x; \eta_1)}{m(x; \eta_2)} \, d\mu(x)
\]
\[
= B_{F^*}(\eta_1 : \eta_2) = F_{\eta_1 : \eta_2} = DF_{\eta_1 : \eta_2} = DF_{\eta_2 : \eta_1}
\]

where \(DF_{\eta_1 : \eta_2} = F^*(\eta_1) + F(\eta_2) - \langle \eta_1, \eta_2 \rangle\)

- Skew \(\alpha\)-Jensen-Shannon divergence:
\[
JS_\alpha(p : q) := (1-\alpha)KL(p : m_\alpha) + \alpha KL(q : m_\alpha), \quad \text{for } \alpha \in [0,1], \quad m_\alpha = (1-\alpha)p + \alpha q.
\]

\(\alpha\)-Jensen divergences
\[
J_{\alpha}(\eta_1 : \eta_2) := (1-\alpha)F^*(\eta_1) + \alpha F^*(\eta_2) - F^*((1-\alpha)\eta_1 + \alpha \eta_2), \quad \text{for } F^*(\eta) = -h(m(x; \eta)).
\]

**Limit cases:**
\[
\lim_{\alpha \to 1^-} \frac{J_{\alpha}(\eta_1 : \eta_2)}{\alpha(1-\alpha)} = B_{F^*}(\eta_1 : \eta_2) = KL(m_1 : m_2)
\]
\[
\lim_{\alpha \to 0^+} \frac{J_{\alpha}(\eta_1 : \eta_2)}{\alpha(1-\alpha)} = B_{F^*}(\eta_1 : \eta_2) = KL(m_2 : m_1)
\]

**Theorem:** The \(\alpha\)-Jensen-Shannon statistical divergences between \(\eta\)-mixtures amount to \(\alpha\)-Jensen divergences between their corresponding \(\eta\)-mixtures parameter:
\[
JS_\alpha(m(x; \eta_1) : m(x; \eta_2)) = J_{\alpha}(\eta_1 : \eta_2).
\]

**References:**
- On \(w\)-mixtures: Finite convex combinations of prescribed component distributions, arxiv 1708.00568
- Monte Carlo Information Geometry: The dually flat case, arxiv 1803.07225