$$H(\vec{X_d}||\vec{X_c}) = \sum \int p(\vec{x_d}||\vec{x_c}) log(p(\vec{x_d}||\vec{x_c}))d\vec{x_c}$$

$$= \sum p(\vec{X_d} = \vec{x_d}) \int p(\vec{x_c}|\vec{X_d} = \vec{x_d}) log(p(\vec{X_d} = \vec{x_d})p(\vec{x_c}|\vec{X_d} = \vec{x_d}))d\vec{x_c}$$

$$p(\vec{x_c}|\vec{X_d} = \vec{x_d}) = \frac{1}{n_{\vec{x_d}}\sqrt{2^m \pi^m |\rho|}} \sum_{i=1}^{n_{\vec{x_d}}} e^{-\frac{1}{2}(\vec{x}-\vec{x_i})^T \rho^{-1}(\vec{x}-\vec{x_i})}$$

where $\rho = \rho(n_{\vec{x_d}}, m)\sigma$ *where* $c(n_{\vec{x_d}}, m) \in (0, 1]$ *and* σ *is the covariance matrix of* $\vec{X_c}|(\vec{X_d} = \vec{x_d})$

and m *is the dimensionality of* $\vec{x_c}$

$$H(\vec{X_d}||\vec{X_c}) = \sum \frac{p(\vec{X_d} = \vec{x_d})}{n\sqrt{2^m \pi^m |\rho|}} \sum_{i=1}^{n_{\vec{x_d}}} \int e^{-\frac{1}{2}(\vec{x}-\vec{x_i})^T \rho^{-1}(\vec{x}-\vec{x_i})} log(\frac{p(\vec{X_d} = \vec{x_d})}{n\sqrt{2^m \pi^m |\rho|}} \sum_{j=1}^{n} e^{-\frac{1}{2}(\vec{x}-\vec{x_j})^T \rho^{-1}(\vec{x}-\vec{x_j})})d\vec{x_c}$$

Let $\vec{u_i} = \vec{x_c} - \vec{x_i} \implies d\vec{u_i} = d\vec{x_i}$

$$H(\vec{X_d}||\vec{X_c}) = \sum \frac{p(\vec{X_d} = \vec{x_d})}{n\sqrt{2^k \pi^k |\rho|}} \sum_{i=1}^{n} \int e^{-\frac{1}{2}\vec{u_i}^T \rho^{-1}\vec{u_i}} log(\frac{p(\vec{X_d} = \vec{x_d})}{n\sqrt{2^k \pi^k |\rho|}} \sum_{j=1}^{n} e^{-\frac{1}{2}(\vec{u_i}+\vec{x_i}-\vec{x_j})^T \rho^{-1}(\vec{u_i}+\vec{x_i}-\vec{x_j})})d\vec{u_i}$$

Let θ *be the orthogonal svd of* ρ

Let $\vec{\mu_i} = \frac{1}{\sqrt{2}}\theta^{-1}\vec{u_i} \implies \vec{u_i} = \sqrt{2}\theta\vec{\mu_i} \implies \vec{u_i}^T = \sqrt{2}\vec{\mu_i}^T\theta^T \implies d\vec{u_i} = \sqrt{2}|\theta|d\vec{\mu_i}$

$$\implies e^{-\frac{1}{2}\vec{u_i}^T \rho^{-1}\vec{u_i}} = e^{-\frac{1}{2}\sqrt{2}\vec{\mu_i}^T\theta^T \rho^{-1}\sqrt{2}\theta\vec{\mu_i}} = e^{-\vec{\mu_i}^T\theta^T\rho^{-1}\theta\vec{\mu_i}} = e^{-\vec{\mu_i}^T D_\lambda \vec{\mu_i}} = e^{-\sum_{k=1}^{m}\lambda_k \mu_{i,k}^2} = \prod_{k=1}^{m} e^{-\lambda_k \mu_{i,k}^2}$$

$$\implies H(\vec{X_d}||\vec{X_c}) = \sum_{\vec{x_d}\in\vec{X_d}} \frac{\sqrt{2}|\theta_{\vec{x_d}}|p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}}\sqrt{2^k\pi^k|\rho_{\vec{x_d}}|}} \sum_{i=1}^{n_{\vec{x_d}}} \int_{R^m} \prod_{k=1}^{m} e^{-\lambda_k\mu_{i,k}^2} log(\frac{p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}}\sqrt{2^k\pi^k|\rho_{\vec{x_d}}|}} \sum_{j=1}^{n} e^{-\frac{1}{2}(\sqrt{2}\theta\vec{\mu_i}+\vec{x_i}-\vec{x_j})^T\rho^{-1}(\sqrt{2}\theta\vec{\mu_i}+\vec{x_i}-\vec{x_j})})d\vec{\mu_i}$$

$$= \sum_{\vec{x_d}\in\vec{X_d}} \frac{\sqrt{2}|\theta_{\vec{x_d}}|p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}}\sqrt{2^k\pi^k|\rho_{\vec{x_d}}|}} \sum_{i=1}^{n_{\vec{x_d}}} \int_{-\infty}^{\infty} e^{-\lambda_1\mu_{i,1}^2} \int_{-\infty}^{\infty} e^{-\lambda_2\mu_{i,2}^2} ... \int_{-\infty}^{\infty} e^{-\lambda_m\mu_{i,m}^2} log(\frac{p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}}\sqrt{2^k\pi^k|\rho_{\vec{x_d}}|}} \sum_{j=1}^{n} e^{-\frac{1}{2}(\sqrt{2}\theta\vec{\mu_i}+\vec{x_i}-\vec{x_j})^T\rho^{-1}(\sqrt{2}\theta\vec{\mu_i}+\vec{x_i}-\vec{x_j})})d\vec{\mu_{i,m}}...d\vec{\mu_{i,2}}d\vec{\mu_{i,1}}$$

Let $\tau_{i,k} = \sqrt{\lambda_k}\mu_{i,k} \implies \vec{\tau_i} = \vec{\mu_i} \odot \sqrt{\vec{\lambda}} \implies \vec{\mu_i} = (\frac{\vec{1}}{\sqrt{\lambda}}) \odot \vec{\tau_i} \implies d\vec{\mu_i} = \prod_{k=1}^{m} \frac{1}{\sqrt{\lambda_k}} d\vec{\tau_i} = \frac{1}{\sqrt{|\rho_{\vec{x_d}}^{-1}|}} d\vec{\tau_i} = \sqrt{|\rho_{\vec{x_d}}|} d\vec{\tau_i}$

$$\implies H(\vec{X_d}||\vec{X_c}) = \sum_{\vec{x_d}\in\vec{X_d}} \frac{\sqrt{2}|\theta_{\vec{x_d}}|p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}}\sqrt{2^m\pi^m}} \sum_{i=1}^{n_{\vec{x_d}}} \int_{-\infty}^{\infty} e^{-\tau_{i,1}^2} \int_{-\infty}^{\infty} e^{-\tau_{i,2}^2} ... \int_{-\infty}^{\infty} e^{-\tau_{i,m}^2} log(\frac{p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}}\sqrt{2^m\pi^m|\rho_{\vec{x_d}}|}} \sum_{j=1}^{n_{\vec{x_d}}} e^{-\frac{1}{2}(\sqrt{2}\theta((\frac{\vec{1}}{\sqrt{\lambda}})\odot\vec{\tau_i})+\vec{x_i}-\vec{x_j})^T\rho^{-1}(\sqrt{2}\theta((\frac{\vec{1}}{\sqrt{\lambda}})\odot\vec{\tau_i})+\vec{x_i}-\vec{x_j})})d\vec{\tau_{i,m}}...d\vec{\tau_{i,2}}d\vec{\tau_{i,1}}$$

U sing Gaussian $-$ *Hermite Quadrature where* H_l *is the set of Hermitian abscissas with* l *evaluation points*

$$H(\vec{X_d}||\vec{X_c}) \approx \sum_{\vec{x_d}\in\vec{X_d}} \frac{\sqrt{2}|\theta_{\vec{x_d}}|p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}}\sqrt{2^m\pi^m}} \sum_{i=1}^{n_{\vec{x_d}}} \sum_{\vec{\tau_i}\in H_l^m} \omega_{\tau_i} log(\frac{p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}}\sqrt{2^m\pi^m|\rho_{\vec{x_d}}|}} \sum_{j=1}^{n_{\vec{x_d}}} e^{-\frac{1}{2}(\sqrt{2}\theta((\frac{\vec{1}}{\sqrt{\lambda}})\odot\vec{\tau_i})+\vec{x_i}-\vec{x_j})^T\rho^{-1}(\sqrt{2}\theta((\frac{\vec{1}}{\sqrt{\lambda}})\odot\vec{\tau_i})+\vec{x_i}-\vec{x_j})})$$

$$= \sum_{\vec{x_d} \in \vec{X_d}} \frac{\sqrt{2}|\theta_{\vec{x_d}}| p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}} \sqrt{2^m \pi^m}} \sum_{\vec{\tau_i} \in H_l^m} \omega_{\tau_i} \sum_{i=1}^{n_{\vec{x_d}}} log\left(\frac{p(\vec{X_d} = \vec{x_d})}{n_{\vec{x_d}} \sqrt{2^m \pi^m |\rho_{\vec{x_d}}|}} \sum_{j=1}^{n_{\vec{x_d}}} e^{-\frac{1}{2}(\sqrt{2}\theta((\frac{\vec{1}}{\sqrt{\lambda}}) \odot \vec{\tau_i}) + \vec{x_i} - \vec{x_j})^T \rho^{-1} (\sqrt{2}\theta((\frac{\vec{1}}{\sqrt{\lambda}}) \odot \vec{\tau_i}) + \vec{x_i} - \vec{x_j})} \right)$$

$H(\vec{A_d}||\vec{A_c}, \vec{B_d}||\vec{B_c}) = \sum \int p(\vec{a_d}||\vec{a_c}) log(p(\vec{b_d}||\vec{b_c})) d\vec{x_c}$ --- cross entropy

$$= \sum p(\vec{A_d} = \vec{x_d}) \int p(\vec{x_c}|\vec{A_d} = \vec{x_d}) log(p(\vec{B_d} = \vec{x_d}) p(\vec{x_c}|\vec{B_d} = \vec{x_d})) d\vec{x_c}$$

$$p_A(\vec{x_c}|\vec{A_d} = \vec{x_d}) = \frac{1}{n_{\vec{A_d} = \vec{x_d}} \sqrt{2^m \pi^m |\rho_A|}} \sum_{i=1}^{n_{\vec{A_d} = \vec{x_d}}} e^{-\frac{1}{2}(\vec{x} - \vec{x_i})^T \rho_A^{-1} (\vec{x} - \vec{x_i})}$$

The maximum Entropy for a continuous sample distribution is when the sample distribution is Gaussian.

Using the Kernel Density Estimation

$$p(\vec{x_c}|\vec{X_d} = \vec{x_d}) = \frac{1}{n_{\vec{x_d}} \sqrt{2^m \pi^m |\rho|}} \sum_{i=1}^{n_{\vec{x_d}}} e^{-\frac{1}{2}(\vec{x} - \vec{x_i})^T \rho^{-1} (\vec{x} - \vec{x_i})}$$

the covariance