Pequeno Relatório a respeito da nossa última discussão. A relembrar, a nossa discussão levantou a seguinte questão, a operação efetuada pelo operador τ de tal modo que: $\tau(D) = -\frac{HD^2H}{2}$ é essencial para que o passo realizado pela Decomposição espectral não possua ambiguidades em sua solução? (Neste caso, a ambiguidade estaria relacionada ao conjunto de pontos de dimensão 'd' resultantes do produto dos autovetores por seus respectivos autovalores da matriz D). Tivemos a ideia de tratar a descoberta dos pontos que geram o manifold como um problema de otimização cuja função objetivo, ou de custo, a ser otimizada seria a norma euclidiana do vetor $\vec{w} = \vec{v} - D'_n$. Neste caso, $\vec{v}$ é um vetor em que cada elemento é a distância entre p_k e p_i, em que p_k, p_i são vetores de pontos gerados aleatoriamente de dimensão 'd' calculados através da norma euclidiana. D'_n é o vetor linha n transposto da matriz D contendo as distâncias/similaridades/dissimilaridades, gerada através das respectivas abordagens utilizadas pelos algoritmos: ISOMAP e LLE. Com isso foi feito o seguinte desenvolvimento com o intuito de demonstrar como calcular o gradiente através da álgebra linear e do cálculo que atualizará o ponto p_k. Para efeito de visualização no desenvolvimento, a dimensão 'd' escolhida é 2 para facilitar o processo matemático a seguir.

Temos que $\vec{v} = [v_{k1}\ v_{k2} \dots v_{kn}]^t$, em que cada $v_{ki} = \sqrt{P'_kP_k + P'_iP_i - 2P'_kP_i}$ é a distância de P_kà P_i. A função de custo escolhida foi $J(w) = \left|\left|w\right|\right|$, ou seja:

$$J(w) = \sqrt{< \vec{v} - D'_n, \vec{v} - D'_n >} = [v^tv + {D'_n}^tD'_n - 2v^tD'_n]^{1/2}$$

Uma vez definida a função de custo a sua derivada com relação a cada coordenada nos fornece um método iterativo de ajuste dos P_k ésimos pontos de tal forma que a distância entre os P-késimos pontos tendam a matriz D construída.

$$\frac{\partial J(w)}{\partial x_k} = \frac{\partial [v^tv + {D'_n}^tD'_n - 2v^tD'_n]^{\frac{1}{2}}}{\partial x_k} = \frac{1}{2\left|\left|w\right|\right|}\frac{\partial [v^tv + {D'_n}^tD'_n - 2v^tD'_n]}{\partial x_k}, \qquad (1)$$

Dando um zoom nas operações dentro dos colchetes, temos que:

$$\begin{aligned} v^tv = \sum_i v_{ki}^2 &= \sum_i [P'_kP_k + P'_iP_i - 2P'_kP_i] \\ &= \sum_i [(x_k^2 + y_k^2) + (x_i^2 + y_i^2) - 2(x_kx_i + y_ky_i)], \qquad (2) \end{aligned}$$

$$v^tD'_n = \sum_i v_{ki}\delta_{ni} = \sum_i \left(\sqrt{P'_kP_k + P'_iP_i - 2P'_kP_i}\right)\delta_{ni}, \qquad (3)$$

Como essas duas equações podemos aplicar o operador derivada parcial nelas para compor a equação 1.

Aqui o símbolo v^t, t está sendo usado como o operador de transposição.

$$\frac{\partial v^t v}{\partial x_k} = \sum_i \frac{\partial v_i^2}{\partial x_k} = \sum_i 2v_i \frac{\partial\sqrt{P_k'P_k + P_i'P_i - 2P_k'P_i}}{\partial x_k}$$
$$= \sum_i \left[\frac{\partial(x_k^2 + y_k^2)}{\partial x_k} + \frac{\partial(x_i^2 + y_i^2)}{\partial x_k} - 2\frac{\partial(x_k x_i + y_k y_i)}{\partial x_k}\right]$$
$$= \sum_i [2x_k - 2x_i] = 2\sum_i [x_k - x_i], \qquad (4)$$

$$\frac{\partial v^t D_n'}{\partial x_k} = \sum_i d_{ki}\delta_{ni} = \sum_i [\delta_{ni}\frac{\partial d_{ki}}{\partial x_k}] = \sum_i \left[\delta_{ni}\frac{\partial\sqrt{P_k'P_k + P_i'P_i - 2P_k'P_i}}{\partial x_k}\right]$$
$$= \sum_i \left[\delta_{ni}\frac{1}{d_{ki}}\ (x_k - x_i)\right], (5)$$

Como o segundo termo dentro do Operador derivada parcial na equação 1 não depende de x_k, sua derivada em relação ao próprio x_k vai à zero. Substituindo os resultados obtidos em (4) e (5), em (1), temos:

$$\frac{\partial J(w)}{\partial x_k} = \frac{1}{2||w||}\sum_i (x_k - x_i)(1 - \delta_{ni}\frac{1}{d_{ki}})$$

Para calcular $\frac{\partial J(w)}{\partial y_k}$, o padrão é repetido, mudando somente as coordenadas resultando em:

$$\frac{\partial J(w)}{\partial y_k} = \frac{1}{2||w||}\sum_i (y_k - y_i)(1 - \delta_{ni}\frac{1}{d_{ki}})$$

Com isso pode-se formar um vetor gradiente ∇J, em que:

$$\nabla J = \frac{1}{||w||}\begin{bmatrix} \sum_i (x_k - x_i)(1 - \delta_{ni}\frac{1}{d_{ki}}) \\ \vdots \\ \sum_i (x_d - x_i)(1 - \delta_{ni}\frac{1}{d_{ki}}) \end{bmatrix}$$

Cuja a dimensão do vetor gradiente é igual a ‘d’.

Portanto a regra de atualização dos P_k pontos se torna:

$$p_{k-novo} \to p_{k-antigo} + \alpha\nabla J$$

Aqui o símbolo v^t, t está sendo usado como o operador de transposição.