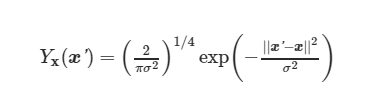
<https://academic.oup.com/bioinformatics/article/31/18/2989/241305>

# Diffusion maps for high-dimensional single-cell analysis of differentiation data

## Diffusion maps

n为细胞数量，G为每个细胞测量的基因数。表示测量的细胞集合为Ω

允许每个细胞X 在其被测量的位置范围x∈RG通过各方向相等的高斯波动函数扩散。

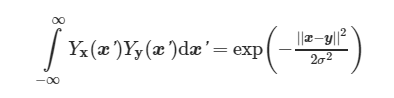


Yx(x’)的标准化为

高斯宽σ2确定每个细胞可以随机扩散的长度尺度。

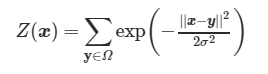
从细胞x到y的转换概率由连个波动函数Yx和Yy定义。

这种干扰产物是另一种高斯（所有前因都抵消了）



因此，可以构建表示在Ω中所有细胞对的n × n Markovian transition probability matrix P

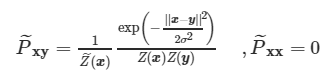


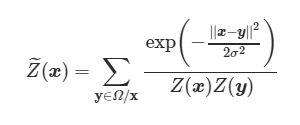


在每一个细胞位置上，Z(x)是分区函数，提供由σ定义的特定体积中x的邻居的数量的估计。

可以被解释为附近的细胞密度。

重新定义密度归一化转换概率矩阵：



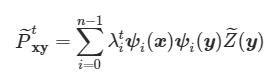


由于仅关系细胞之前转换概率，而不是细胞内局部密度，我们将的对角线设置为0，并从分割函数的和中排除y=x 。

对于足够大的*σ， 矩阵*定义一个在数据集上的遍历性的马尔科夫扩散过程（Markovian diffusion process），具有n个有序的特征值λ0=1>λ1≥...≥λn−1

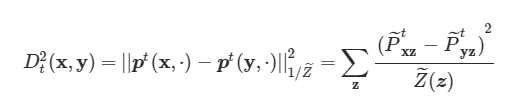
的t次方将呈现在扩散过程长度为t的细胞间的转移概率。

与具有相同的特征向量。转移概率可以表示如下：

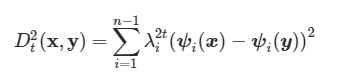


每一个可以视为一个向量，pt(x,.)表示每个细胞x的代表特征。

通过计算特征空间中带权重的L2距离，细胞x与y之间的距离定义如下：



该扩散距离可以用的特征向量表示：



最大特征值λ0对应的特征向量ψ0是一个常量1。因此他仅对Dt2贡献0，并且从等式中的Dt2谱分解中排除。

这意味着，第一个特征向量空间中的单元欧几里得距离表示他们的扩散距离Dt2 的近似值

此外，的特征值确定相应特征向量方向上的扩散系数。

选择diffusion components(DCs)表示ψ

然后将马尔可夫过渡概率矩阵（DC1和DC2）的本征函数（Eigenfunction）用于低维表示和数据可视化

