

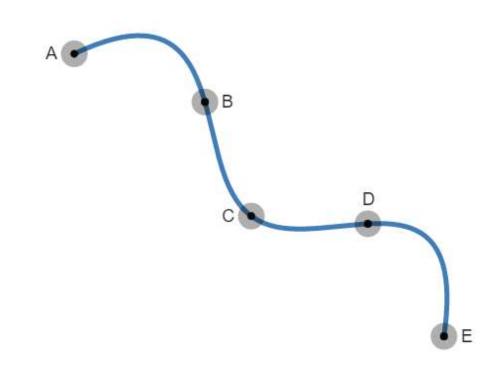
插值,拟合 Interpolation, curve fitting

阮良旺

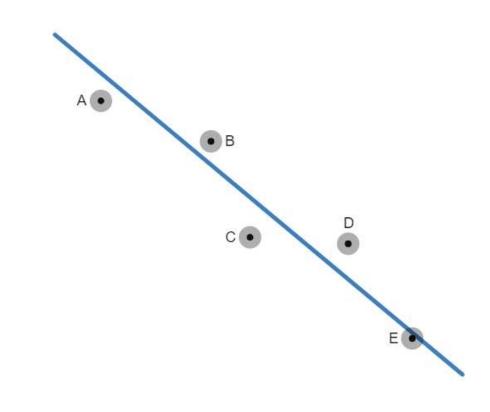
2024.4.8



插值 vs 拟合



插值 经过采样点



拟合 不要求经过采样点

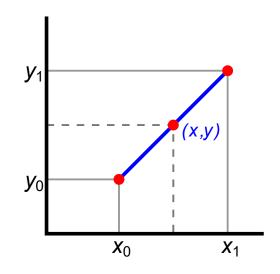
插值 Interpolation

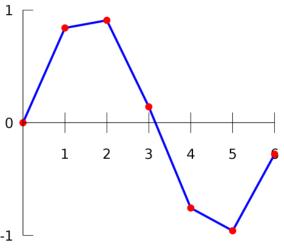
线性插值 (Linear Interpolation)

$$y = y_0 + \frac{x - x_0}{x_1 - x_0} (y_1 - y_0)$$

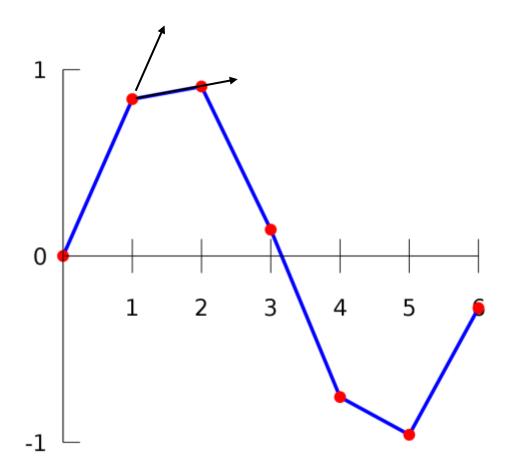
$$y = \frac{x_1 - x}{x_1 - x_0} y_0 + \frac{x - x_0}{x_1 - x_0} y_1$$

$$y = (1 - t)y_0 + ty_1, t = \frac{x - x_0}{x_1 - x_0}$$





连续性



无断点,但是导数有跳变 CO连续 (CO Continuity)

二阶插值

$$y = ax^2 + bx + c$$

 $y(x_i) = y_i, i = 0, 1, 2$

$$\begin{pmatrix} x_0^2 & x_0 & 1 \\ x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix}$$

 x_0 x_1 x_2

范德蒙矩阵 (Vandermonde Matrix)

$$y = y_0 \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + y_1 \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + y_2 \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

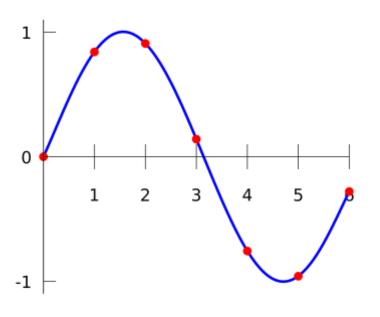
多项式插值 (Polynomial Interpolation)

一共
$$n+1$$
个数据点 $(x_0,y_0), \dots, (x_n,y_n),$
寻找多项式 $y(x)=a_0+a_1x+\dots+a_nx^n$ 经
过所有数据点 $y(x_i)=y_i, i=0,\dots,n$

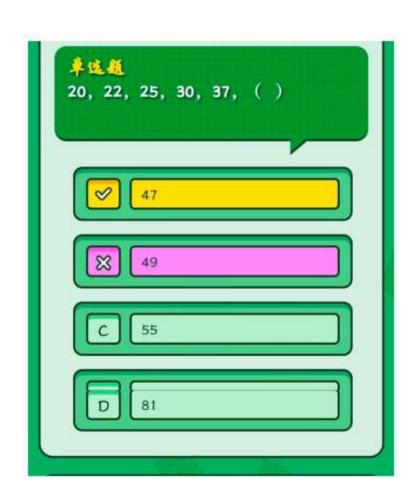
$$y = \sum_{j=0}^{n} y_j \prod_{i,i \neq j} \frac{x - x_i}{x_j - x_i}$$

拉格朗日多项式 $L_i(x)$

$$L_{j}(x_{k}) = \prod_{i,i \neq j} \frac{x_{k} - x_{i}}{x_{j} - x_{i}} = \delta_{jk} = \begin{cases} 1, k = j \\ 0, k \neq j \end{cases}$$



如何寻找数列下一项...?





940 人赞同了语回答

料个机类

考虑函数

$$f(x) = 15 + \frac{539}{60}x - \frac{137}{24}x^2 + 2x^3 - \frac{7}{24}x^4 + \frac{1}{60}x^5$$

则有

$$f(1) = a_1 = 20, f(2) = a_2 = 22, f(3) = a_3 = 25$$

 $f(4) = a_4 = 30, f(5) = a_5 = 37$

所以

$$a_6 = f(6) = 47$$

考虑函数

$$f(x) = 13 + \frac{271}{20}x - \frac{227}{24}x^2 + \frac{41}{12}x^3 - \frac{13}{24}x^4 + \frac{1}{30}x^5$$

则有

$$f(1) = a_1 = 20, f(2) = a_2 = 22, f(3) = a_3 = 25$$

$$f(4) = a_4 = 30, f(5) = a_5 = 37$$

所以

$$a_6 = f(6) = 49$$

考虑函数

$$f(x) = 7 + \frac{109}{4}x - \frac{497}{24}x^2 + \frac{23}{3}x^3 - \frac{31}{24}x^4 + \frac{1}{12}x^5$$

则有

$$f(1) = a_1 = 20, f(2) = a_2 = 22, f(3) = a_3 = 25$$

$$f(4) = a_4 = 30, f(5) = a_5 = 37$$

所以

$$a_6 = f(6) = 55$$

考虑函数

$$f(x) = -19 + \frac{5197}{60}x - \frac{1667}{24}x^2 + \frac{313}{12}x^3 - \frac{109}{24}x^4 + \frac{3}{10}x^5$$

则有

$$f(1) = a_1 = 20, f(2) = a_2 = 22, f(3) = a_3 = 25$$

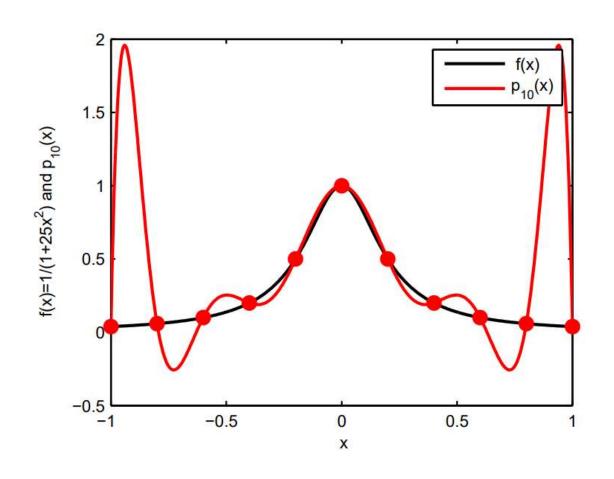
$$f(4) = a_4 = 30, f(5) = a_5 = 37$$

所以

$$a_6 = f(6) = 81$$

所以ABCD都是下确的

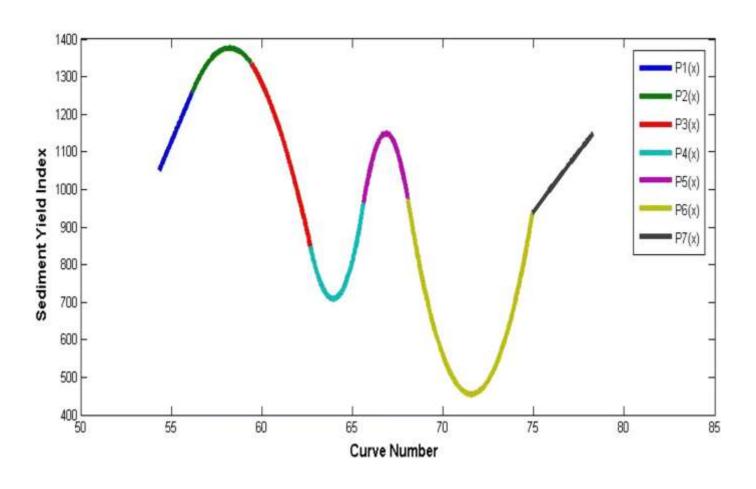
问题: 龙格现象 (Runge's phenomenon)



尽管曲线是处处光滑的, 但是可能出现非常振荡的结果

解决方法: 样条插值 (Spline Interpolation)

通过拼接多段低阶多项式进行插值

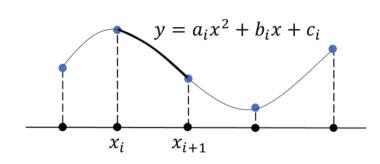


二阶样条插值 (Quadratic Spline Interpolation)

n+1个数据点,一共n个区间,对应 3n 个未知数 a_i, b_i, c_i

• 插值条件: 区间i的二次函数经过 $(x_i, y_i), (x_{i+1}, y_{i+1})$

$$\begin{cases} a_i x_i^2 + b_i x_i + c_i = y_i \\ a_i x_{i+1}^2 + b_i x_{i+1} + c_i = y_{i+1} \end{cases}$$



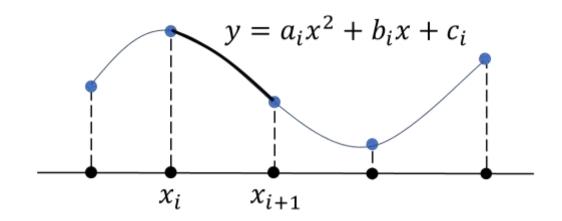
一共 2n 个方程

• 导数光滑条件:数据点i左右区间在该点的导数相同 $2a_{i-1}x_i + b_{i-1}x_i = 2a_ix_i + b_i$

一共n-1个方程

• 3n-2n-(n-1)=1,剩下一个方程可指定第一段为线性函数: $a_0=0$

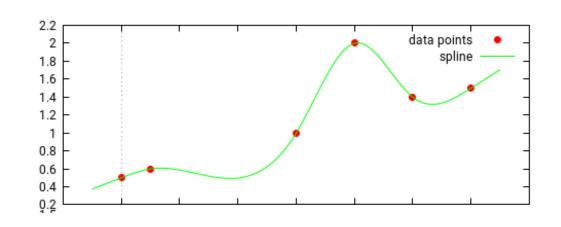
二阶样条插值 (Quadratic Spline Interpolation)

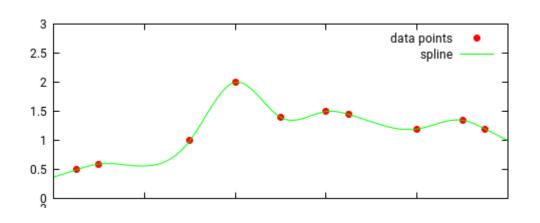


- 插值结果是一阶导连续的 (C1 Continuity), 但是二阶导不连续
- · 需要联立求解大小为 3n 的系数矩阵方程, 没有显式解, 速度慢
- 每个数据点都会影响全局的插值结果

三阶样条插值 (Cubic Spline Interpolation)

- 可以通过类似二阶样条的方式定义,建立系数矩阵方程
- 4n个未知数, 2n个值连续条件, n-1个导数连续条件, n-1个二个字数连续条件, n-1个二个二个字数连续条件, p-1个指定常数条件
- 保证插值结果是二阶导数连续的 (C2 Continuity)
- 问题与二阶情况类似: 需要求解系数矩阵方程, 单个数据点影响全局插值结果





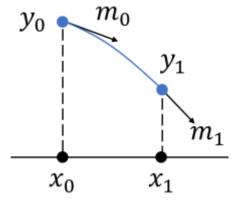
关于解方程

$$\begin{pmatrix} u_1 & h_1 \\ h_1 & u_2 & h_2 \\ & h_2 & u_3 & h_3 \\ & \ddots & \ddots & \ddots \\ & & h_{n-3} & u_{n-2} & h_{n-2} \\ & & & h_{n-2} & u_{n-1} \end{pmatrix} \begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_{n-2} \\ v_{n-1} \end{pmatrix}$$

对称三对角的对角占优矩阵方程,有快速求解方法 GAMES102 学习材料2

三阶厄米插值 (Cubic Hermite Interpolation)

我们可以通过放弃二阶导数连续的要求, 让插值获得更好的局部性



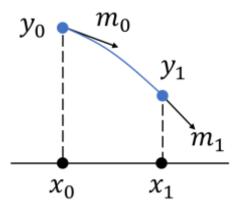
$$y(x) = ax^3 + bx^2 + cx + d$$

$$y(x_0) = y_0, y(x_1) = y_1, y'(x_0) = m_0, y'(x_1) = m_1$$

$$y = (2t^{3} - 3t^{2} + 1)y_{0} + (t^{3} - 2t^{2} + t)m_{0}(x_{1} - x_{0}) + (-2t^{3} + 3t^{2})y_{1} + (t^{3} - t^{2})m_{1}(x_{1} - x_{0})$$

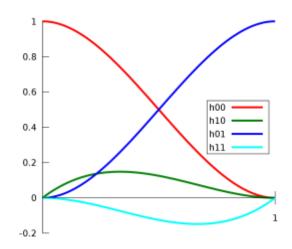
$$t = \frac{x - x_{0}}{x_{1} - x_{0}}$$

三阶厄米插值 (Cubic Hermite Interpolation)



$$y_1 = \frac{h_{00}}{y_1} \qquad y = \frac{(2t^3 - 3t^2 + 1)y_0 + (t^3 - 2t^2 + t)m_0(x_1 - x_0)}{+(-2t^3 + 3t^2)y_1 + (t^3 - t^2)m_1(x_1 - x_0)}$$

$$h_{01} \qquad h_{11}$$



厄米基函数 $h_{00}(0) = 1, h'_{00}(0) = 0, h_{00}(1) = 0, h'_{00}(1) = 0$ $h_{10}(0) = 0, h'_{10}(0) = 1, h_{10}(1) = 0, h'_{10}(1) = 0$

$$h_{10}(0) = 0, h_{10}(0) = 1, h_{10}(1) = 0, h_{10}(1) = 0$$
 $h_{01}(0) = 0, h'_{01}(0) = 0, h_{01}(1) = 1, h'_{01}(1) = 0$
 $h_{11}(0) = 0, h'_{11}(0) = 0, h_{11}(1) = 0, h'_{11}(1) = 1$

三阶厄米插值 (Cubic Hermite Interpolation)

如何选择每个采样点的导数mi?

• 有限差分:

$$m_i = \frac{1}{2} \left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i} + \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \right)$$

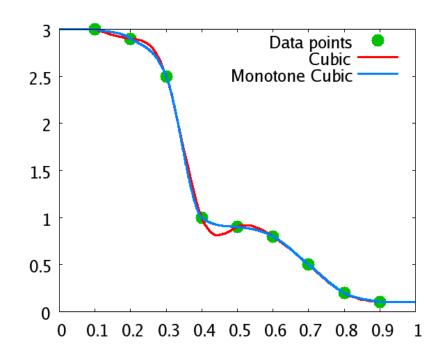
• Cardinal样条:

$$m_i = (1-c)\frac{y_{i+1} - y_i}{x_{i+1} - x_i}$$

• Catmull-Rom样条(假设 x_i 均匀划分): $m_i = \frac{y_{i+1} - y_{i-1}}{2}$

• . . .

单调三阶样条插值 (Monotone Cubic Interpolation)



Piecewise Cubic Hermite Interpolating Polynomial (PCHIP)

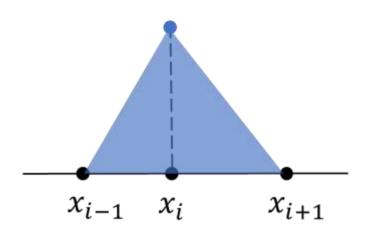
F. N. Fritsch and J. Butland, A method for constructing local monotone piecewise cubic interpolants, SIAM J. Sci. Comput., 5(2), 300-304 (1984). DOI:10.1137/0905021.

通过调整每个顶点的导数mi,使得插值结果C1连续,并且区间内单调

总结

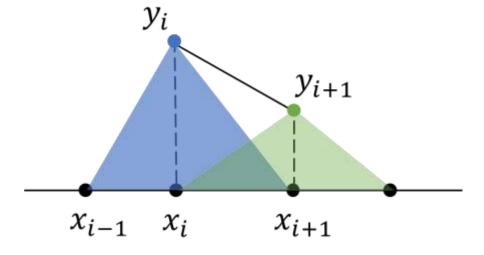
插值方法	连续性	复杂性	局部性
线性插值	CO		
多项式插值	Cn (取决于点的个数)		◎(龙格现象)
二阶样条插值	C1	② (解方程)	
三阶样条插值	C2	② (解方程)	
三阶厄米样条插值	C1		
单调三阶样条插值	C1		

基函数 (Basis Function)



定义基函数
$$B_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, x \in [x_{i-1}, x_i] \\ \frac{x_{i+1} - x}{x_{i+1} - x_i}, x \in [x_i, x_{i+1}] \\ 0, else \end{cases}$$

满足
$$B_i(x_j) = \delta_{ij} = \begin{cases} 1, i = j \\ 0, i \neq j \end{cases}$$

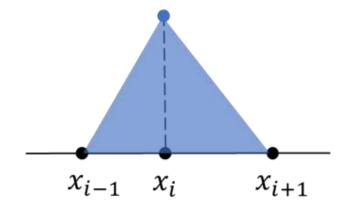


$$y = \frac{x_{i+1} - x}{x_{i+1} - x_i} y_0 + \frac{x - x_i}{x_{i+1} - x_i} y_1$$
$$y = B_i(x) y_i + B_{i+1}(x) y_{i+1}$$

$$y(x) = \sum_{i} y_i B_i(x)$$

基函数与插值

$$y(x) = \sum_{i} y_i B_i(x)$$

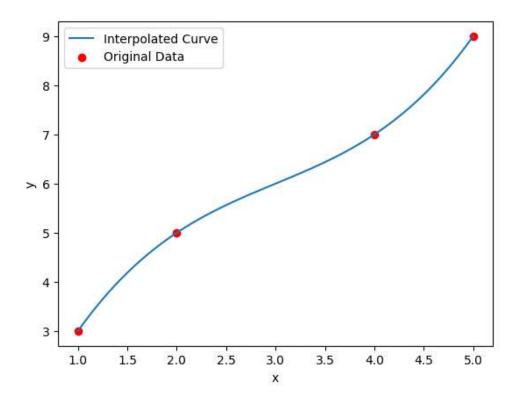


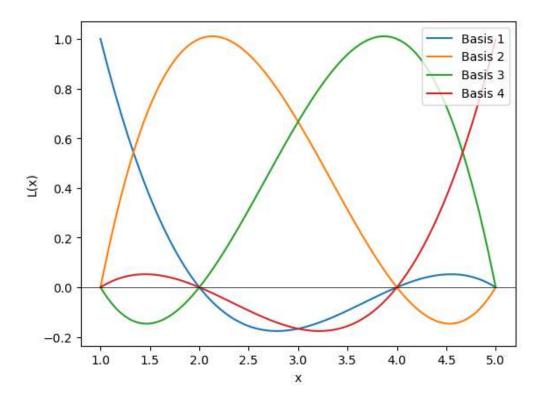
- 基函数为插值提供了新的观察视角
- 可以通过改变基函数定义新的插值方式
- 只要满足 $B_i(x_j) = \delta_{ij}$ 就能保证插值结果经过采样点:

$$y(x_i) = \sum_j y_j B_j(x_i) = \sum_j y_j \delta_{ij} = y_i$$

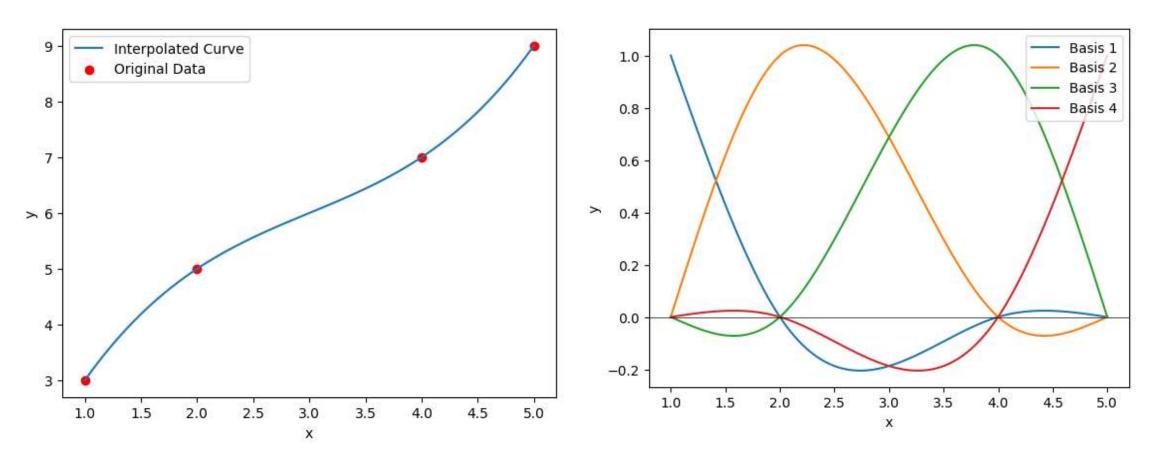
多项式插值的基函数

$$y = \sum_{j=0}^{n} y_j L_j(x) = \sum_{j=0}^{n} y_j \prod_{i,i \neq j} \frac{x - x_i}{x_j - x_i}$$



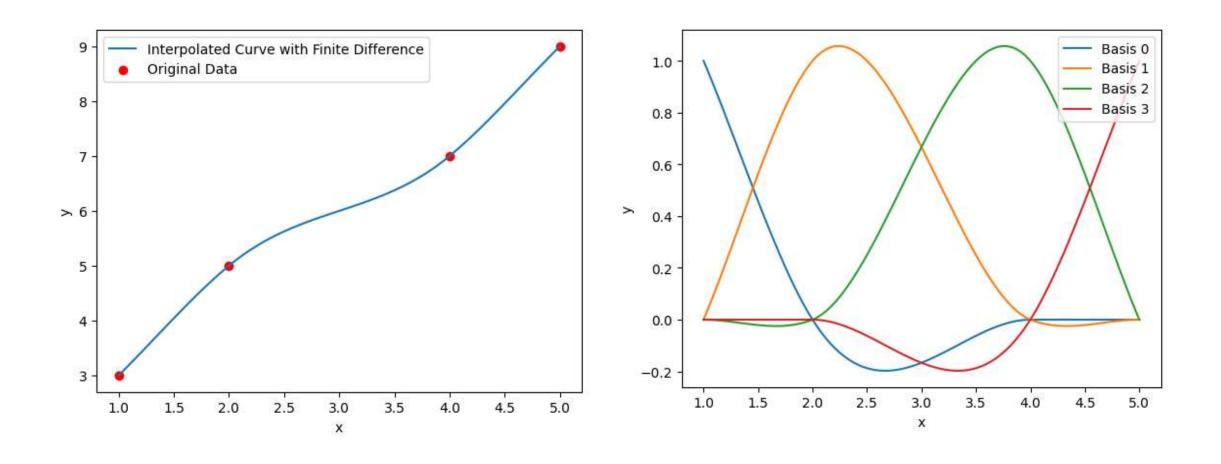


三阶样条插值的基函数

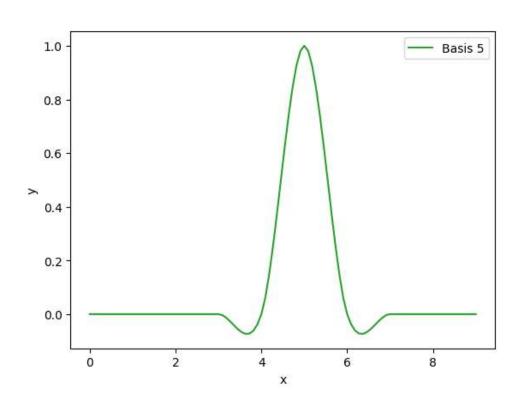


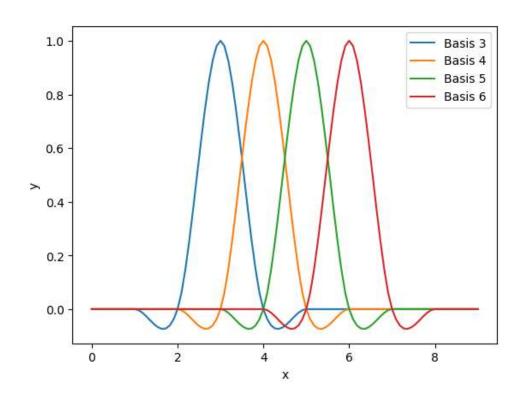
通过求解方程得到

三阶厄米样条插值的基函数



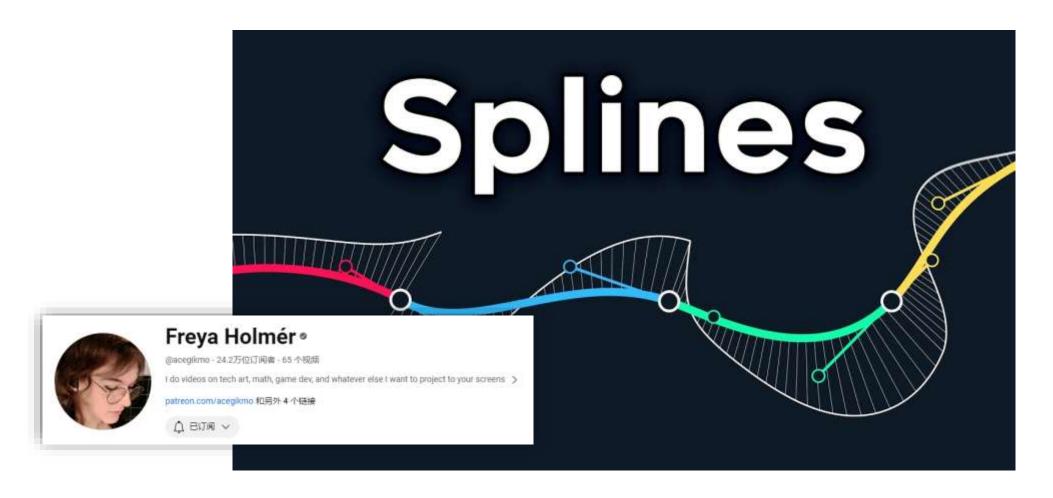
三阶厄米样条插值的基函数





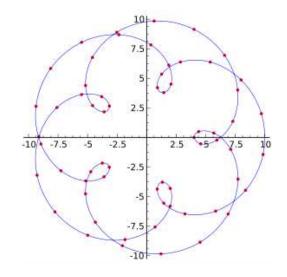
三阶厄米插值的基函数具有局部性,并且当 x_i 间隔相同时, B_i 形状相同

推荐补充资料



https://www.youtube.com/@acegikmo

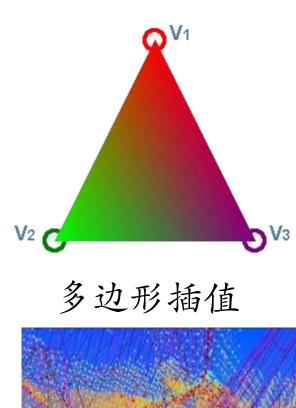
高维插值

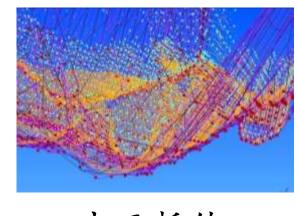


曲线插值



像素插值

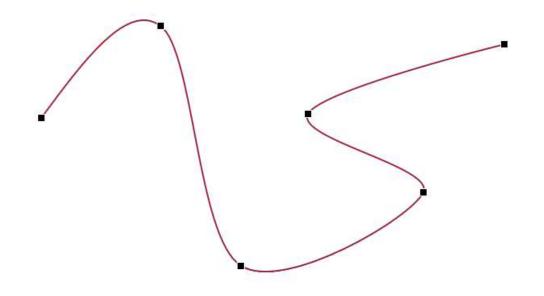




点云插值

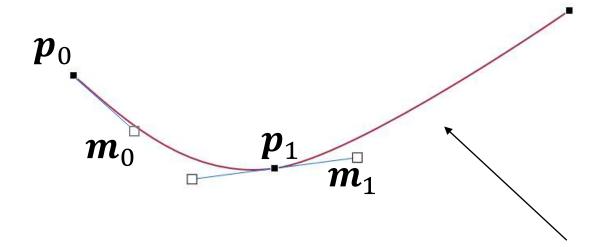
曲线插值

按顺序给定 p_0, \dots, p_n 一系列点,插值出中间的曲线p(t) 可以认为在相邻点之间 $t \in [0,1]$



三阶厄米样条曲线

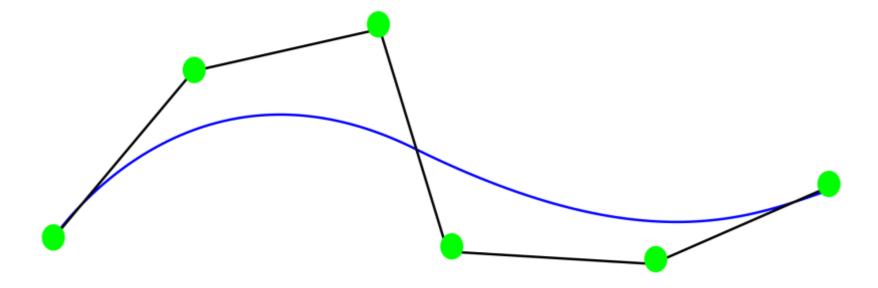
$$p(t) = (2t^3 - 3t^2 + 1)p_0 + (t^3 - 2t^2 + t)m_0 + (-2t^3 + 3t^2)p_1 + (t^3 - t^2)m_1$$



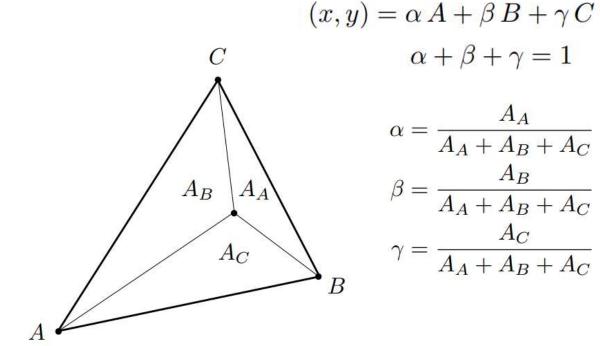
Power Point用的就是这条曲线

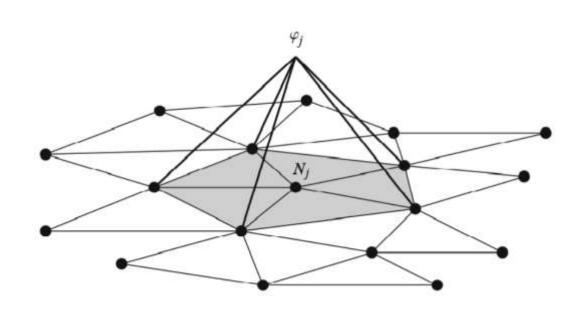
插值与曲线

- 图形学中的曲线并不一定要求经过控制点,通过控制点的曲线称为插值曲线 (Interpolation Curve)
- 比如最常用的贝塞尔曲线 (Bezier Curve)和B样条曲线 (B-Spline Curve)就不经过控制点,因此不是插值曲线
- 关于曲线更多的知识可以学习 GAMES102



三角形插值



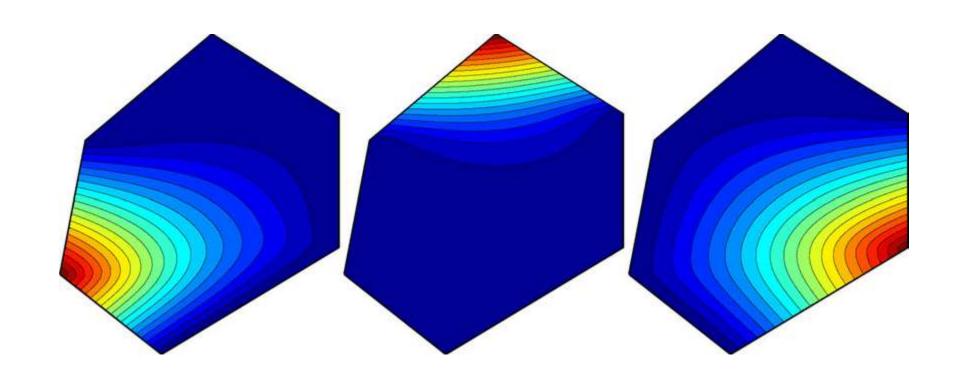


重心坐标 (Barycentric Interpolation)

重心坐标基函数

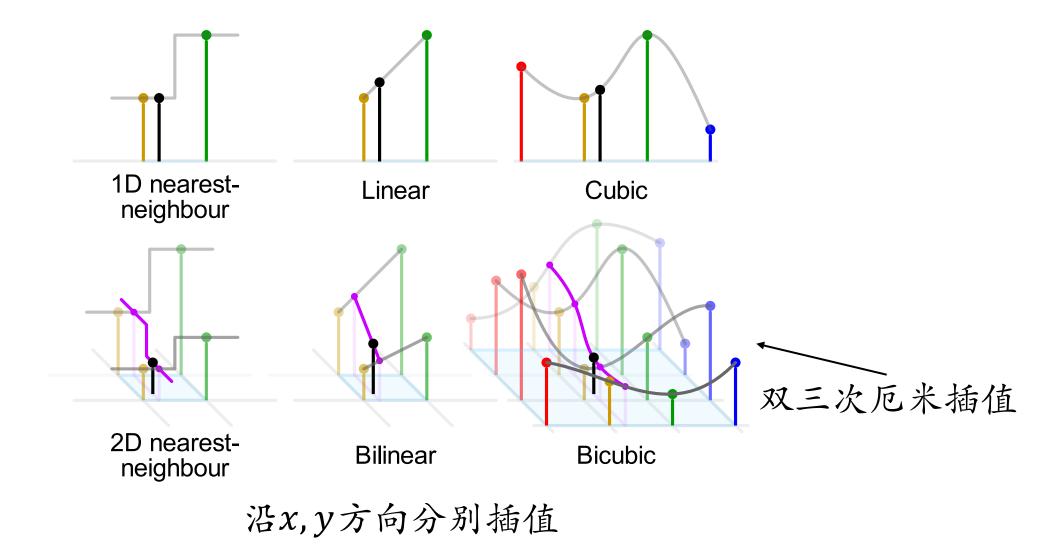
GAMES101: 插值、高级纹理映射31

多边形插值

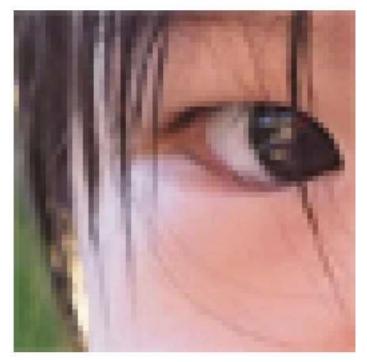


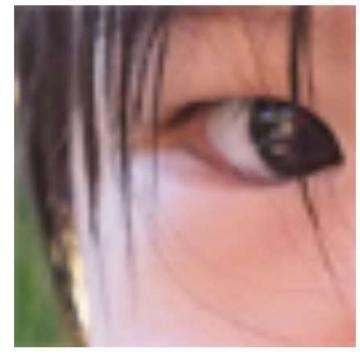
均值坐标 (Mean Value Coordinate)

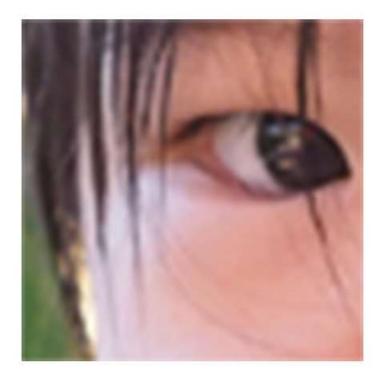
像素插值



像素插值







Nearest

Bilinear

Bicubic

GAMES101: 插值、高级纹理映射34

点云插值

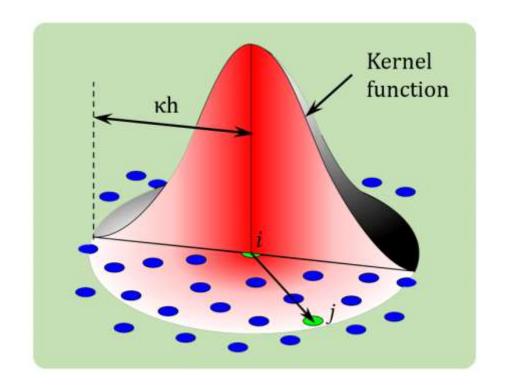
径向基函数 (Radial Basis Function,RBF),核函数 (Kernel Function) $\varphi(\|x-x_i\|)$

高斯核: $\exp(-(\varepsilon r)^2)$

拉普拉斯核: $\exp(-\varepsilon r)$

二次有理核: $\frac{1}{1+(\varepsilon r)^2}$

...



点云插值

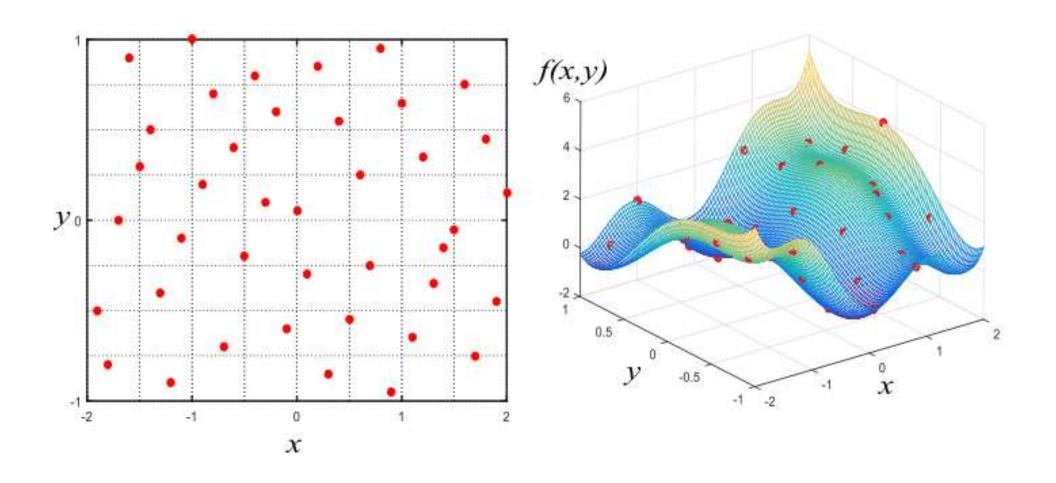
$$f(x) = \sum_{i} y_{i} \varphi(\|x - x_{i}\|)$$

求解 y_i 使得 $f(x_i) = f_i$, 以满足插值条件:

$$\begin{pmatrix} \varphi(\|x_0 - x_0\|) & \cdots & \varphi(\|x_0 - x_{n-1}\|) \\ \vdots & \ddots & \vdots \\ \varphi(\|x_{n-1} - x_0\|) & \cdots & \varphi(\|x_{n-1} - x_{n-1}\|) \end{pmatrix} \begin{pmatrix} y_0 \\ \vdots \\ y_{n-1} \end{pmatrix} = \begin{pmatrix} f_0 \\ \vdots \\ f_{n-1} \end{pmatrix}$$

 $n \times n$ 的对称矩阵方程,n为顶点个数

点云插值



点云插值

跟在点云上插值相比,更常见的情况是我们使用点云近似表达一个连续场,此时可以放弃插值条件 $f(x_i) = f_i$,而是认为一个顶点表达的就是空间中由核函数定义的一定范围内分布,直接给出

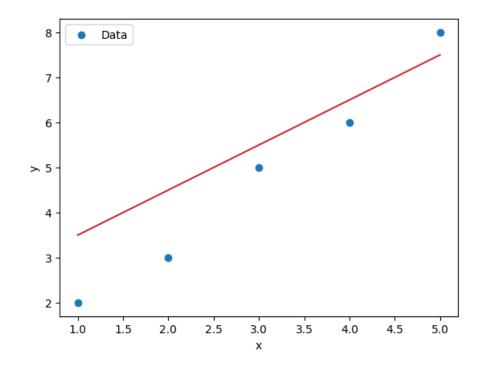
$$f(x) = \sum_{i} f_i \varphi(\|x - x_i\|)$$

此时虽然 $f(x_i) \neq f_i$,但一般还需假定核函数满足归一化条件: $\int_{\Omega} \varphi(\|x - x_i\|) dV = 1$

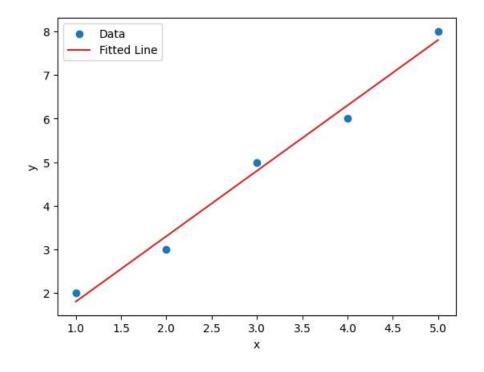
例如: SPH流体, 3D Gaussian Splatting ...

拟合 Fitting

拟合 (Fitting)



假定数据分布满足某种规律 $y = f_{\theta}(x)$



寻找最佳的参数最小化误差 $\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{i} \operatorname{Dist}(f_{\theta}(x_i), y_i)$

假定数据分布满足:

$$y = a_{1}x + a_{0} = (1 \quad x) \begin{pmatrix} a_{0} \\ a_{1} \end{pmatrix}$$

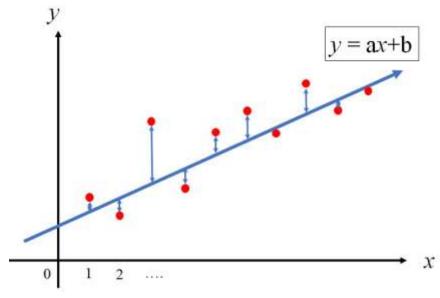
$$y = a_{2}x^{2} + a_{1}x + a_{0} = (1 \quad x \quad x^{2}) \begin{pmatrix} a_{0} \\ a_{1} \\ a_{2} \end{pmatrix}$$

$$y = a_{1}\sqrt{x} + a_{0} = (1 \quad \sqrt{x}) \begin{pmatrix} a_{0} \\ a_{1} \end{pmatrix}$$
...
$$y = (g_{0}(x) \quad \cdots \quad g_{k-1}(x)) \begin{pmatrix} a_{0} \\ \vdots \\ a_{k-1} \end{pmatrix} = \mathbf{g}^{T}(x)\mathbf{a}$$

$$y = (g_0(x) \quad \cdots \quad g_{k-1}(x)) \begin{pmatrix} a_0 \\ \vdots \\ a_{k-1} \end{pmatrix} = \boldsymbol{g}^{\mathrm{T}}(x) \boldsymbol{a}$$

给定数据集 $(x_0, y_0), \cdots, (x_{n-1}, y_{n-1}),$ 最小化

$$\sum_{i} (y_i - \boldsymbol{g}^{\mathrm{T}}(x_i)\boldsymbol{a})^2$$

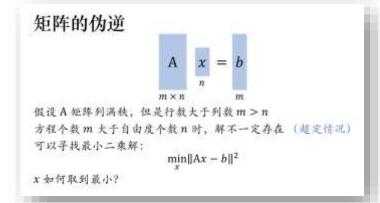


$$\mathbf{G} = \begin{pmatrix} g_0(x_0) & \cdots & g_{k-1}(x_0) \\ g_0(x_1) & \cdots & g_{k-1}(x_1) \\ \vdots & & & \vdots \\ g_0(x_{n-1}) & \cdots & g_{k-1}(x_{n-1}) \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{pmatrix}$$

$$\sum_{i} (y_i - \boldsymbol{g}^{\mathrm{T}}(x_i)\boldsymbol{a})^2 = \|\boldsymbol{y} - \mathbf{G}\boldsymbol{a}\|^2$$

$$\min_{\boldsymbol{a}} \|\boldsymbol{y} - \mathbf{G}\boldsymbol{a}\|^2$$
^{矩阵的伪逆}

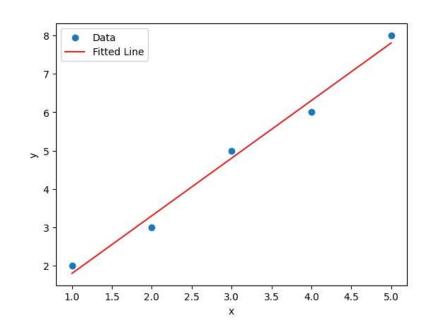
$$\boldsymbol{a} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\boldsymbol{y}$$



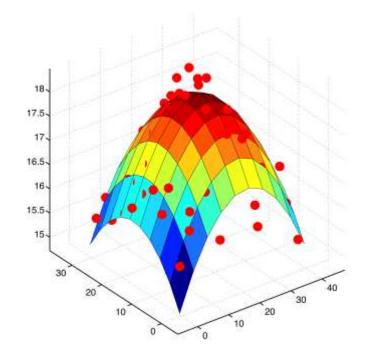
SVD&PCA课件第10页

算法流程:

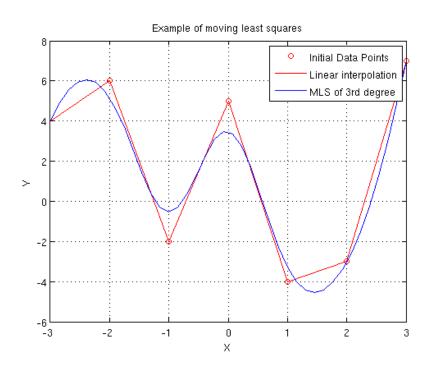
- 构造矩阵 $\mathbf{G} \in \mathbb{R}^{n \times k}$, $\mathbf{G}_{ij} = g_j(x_i)$
- 求解 $\boldsymbol{a} = (\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\boldsymbol{y}$
- 拟合函数 $y = g^{T}(x)a$



$$\mathbf{G} = \begin{pmatrix} g_0(x_0) & \cdots & g_{k-1}(x_0) \\ g_0(x_1) & \cdots & g_{k-1}(x_1) \\ & \vdots & & \\ g_0(x_{n-1}) & \cdots & g_{k-1}(x_{n-1}) \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{pmatrix}$$



- 只在局部进行多项式的最小二乘拟合
- 对局部点施加权重, 距离进的权重大, 距离远的权重小
- · 每当需要点x处的值时,就以x点为中心进行拟合



- 在x点的邻域内有 x_0 ,…, x_{n-1} 共n个点,对应的函数值是 f_0 ,…, f_{n-1}
- 计算相对点x的偏移 $r_i = x_i x$

• 认为
$$x$$
邻域内数据分布满足 p 阶多项式:
$$f(r) = \begin{pmatrix} 1 & r & \cdots & r^p \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = p^T(r)a$$

- 注意此时 $f(0) = a_0$ 就是x点处的拟合值,而不是f(x)
- 并且 $f'(0) = a_1$ 对应x点的斜率, $\frac{1}{2}f''(0) = a_2$ 与x点的曲率相关

•如果是对高维数据进行拟合,比如二维数据 $\mathbf{r} = (r_x, r_y)$,则可能出现混合项 $r_x r_y, r_x^2 r_y, \cdots$:

$$- \text{ \begin{aligned} \begin{aligned} -\beta : \mathbf{p}^{T}(\mathbf{r}) &= (1 & r_{x} & r_{y} & r_{x}^{2} & r_{x}r_{y} & r_{y}^{2} \end{aligned} \]
二 \begin{aligned} \begin{aligned} -\beta : \begin{aligned} \mathbf{r} & r_{x} & r_{x} & r_{x}r_{y} & r_{y}^{2} & r_{x}r_{y} & r_{y}^{2} & r_{y} & r_{y}^{2} & r_{y} & r_{y}^{2} &$$

• 此时 $f(\mathbf{0}) = \mathbf{p}^{T}(\mathbf{0})\mathbf{a} = a_{0}$ 是x处的拟合值, $\nabla f(\mathbf{0}) = (a_{1}, a_{2})$ 是x处的梯度…

$$f(r) = (1 \quad r \quad \cdots \quad r^p) \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = \mathbf{p}^{\mathrm{T}}(r) \mathbf{a}$$

• 最小化目标:

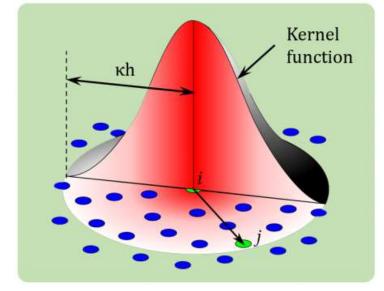
$$J = \sum_{i} (\boldsymbol{p}^{\mathrm{T}}(r_i)\boldsymbol{a} - f_i)^2 w(r_i)$$

其中w(r)是权重函数,在点云插值中提到的核函数都可以使用,比如:

高斯核: $\exp(-(\varepsilon r)^2)$

拉普拉斯核: $\exp(-\varepsilon r)$

二次有理核: $\frac{1}{1+(\varepsilon r)^2}$



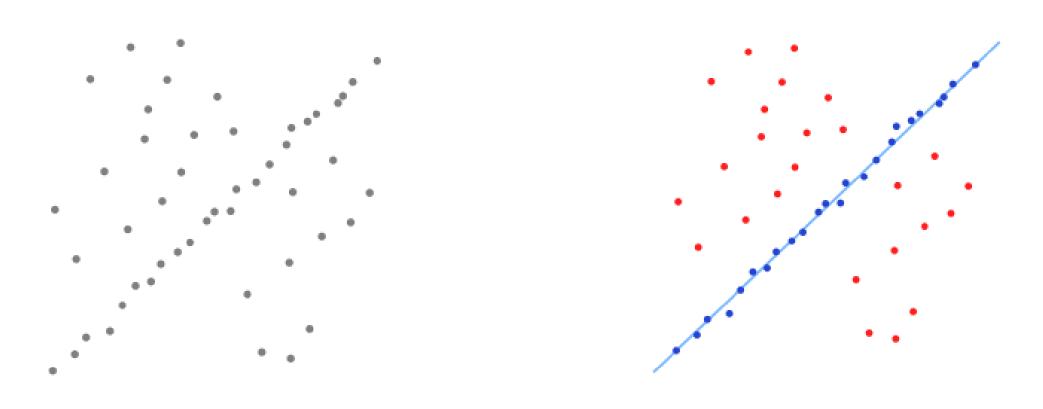
$$J = \sum_{i} (\boldsymbol{p}^{\mathrm{T}}(r_i)\boldsymbol{a} - f_i)^2 w(r_i)$$

$$\frac{\partial J}{\partial \boldsymbol{a}} = 0$$

$$\mathbf{M} = \sum_{i} w(r_i) \boldsymbol{p}(r_i) \boldsymbol{p}^{\mathrm{T}}(r_i), \boldsymbol{b} = \sum_{i} w(r_i) f_i \boldsymbol{p}(r_i)$$

$$\boldsymbol{a} = \mathbf{M}^{-1} \boldsymbol{b}$$

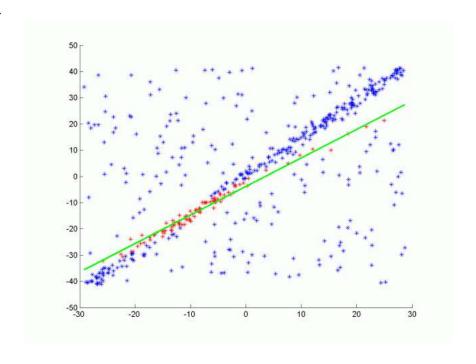
RANSAC (Random sample consensus)



适用于数据集中有大量噪声的情况

RANSAC (Random sample consensus)

- 随机从数据集中选择 n 个点拟合直线
- 将剩下的点分类成两类: 距离拟合直线距离小于 t 的为内点, 大于 t 的是外点
- •如果内点数量大于等于 d,使用所有内点 重新拟合并终止算法
- 否则重新选点重复上面过程





谢谢

