

适用于复杂力学问题的分区数据驱动量纲分析方法及应用

安翼^{1, 3*}, 王驰^{1, 2}, 庞嘉顺¹

1. 中国科学院力学研究所
2. 中国矿业大学（北京）
3. 中国科学院大学工学院

2022-12-23



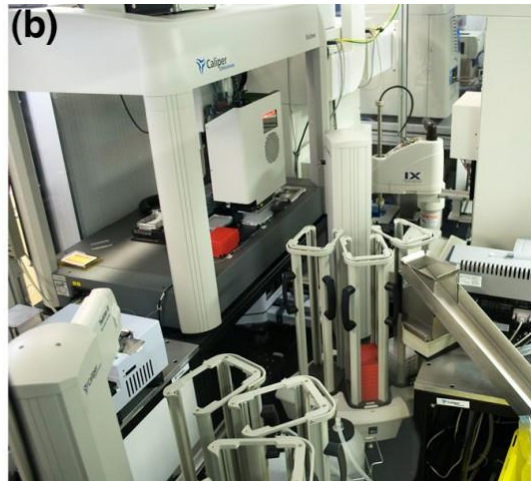
1. AI科学家和数据驱动量纲分析

2. 方法和流程

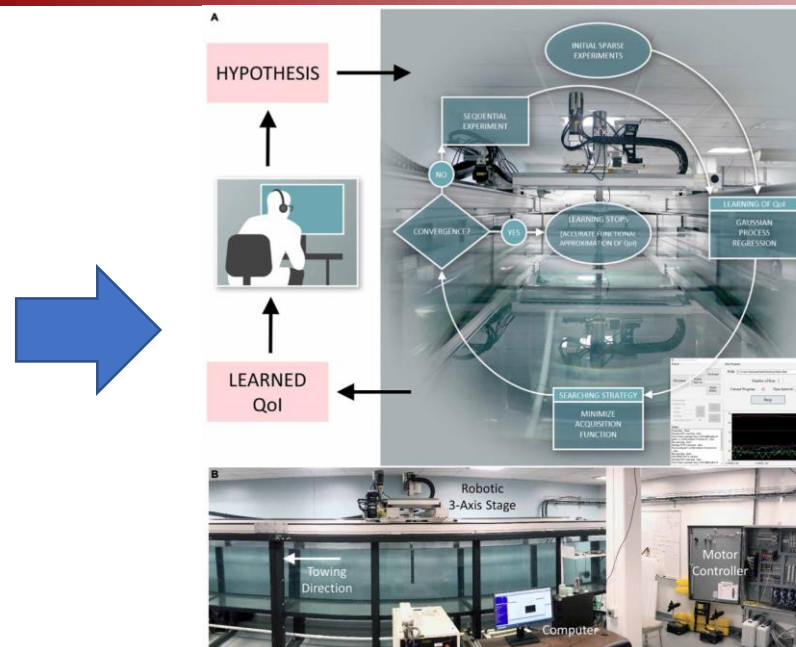
3. 具体案例

4. 鲁棒性及人为参数的影响

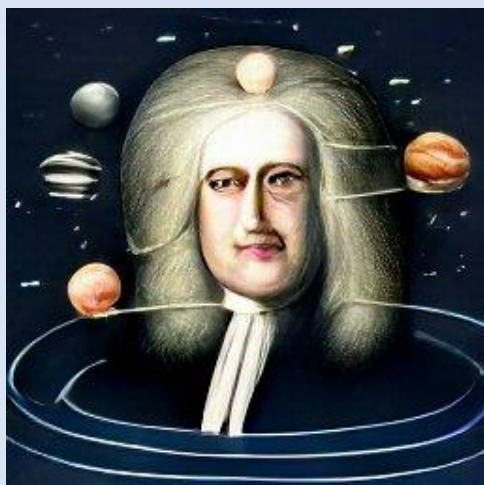
1.0 力学领域的机器科学家——趋势



化学领域：Adam's laboratory robotic system, *Science*, 2009



力学领域：
A robotic Intelligent Towing Tank for learning complex fluid-structure dynamics, *Science Robotics*, 2019



DALL·E生成的牛顿和太阳系 ^_^

$$g = f(t)$$

动力系统

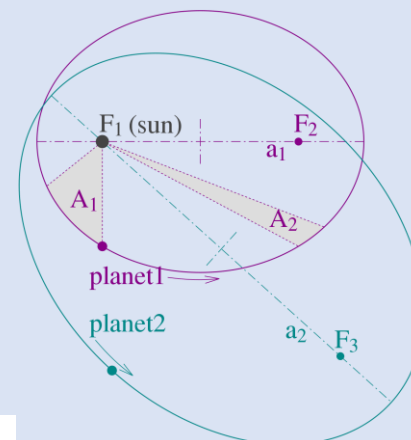
$$\ddot{r} - r\dot{\theta}^2 = -\frac{GM}{r^2}$$

$$r\ddot{\theta} + 2\dot{r}\dot{\theta} = 0$$

$$\mathbf{F} = -G\frac{mM}{r^2}\hat{\mathbf{r}}$$



AI生成的开普勒画像



开普勒定律

$$G = F(m, M, r)$$

经验公式

适合于“复杂”问题

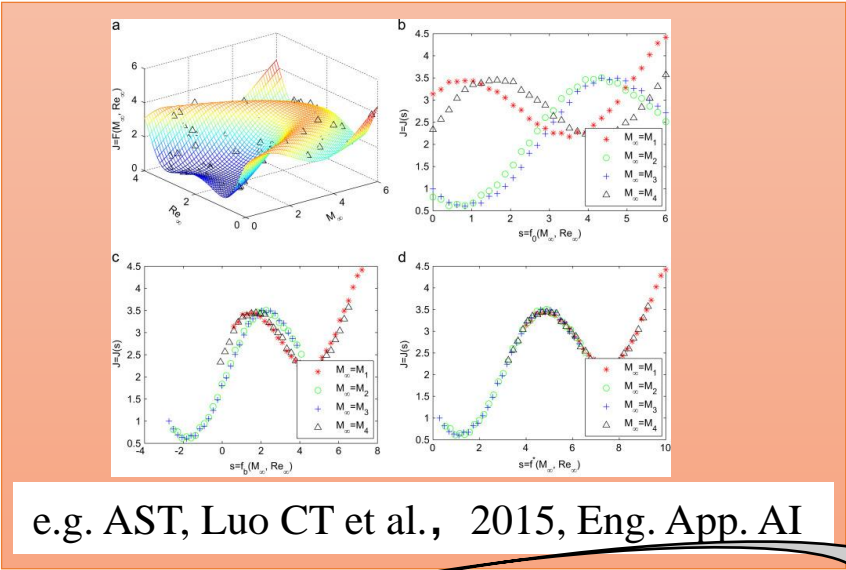
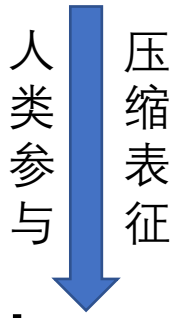
高维, 耦合



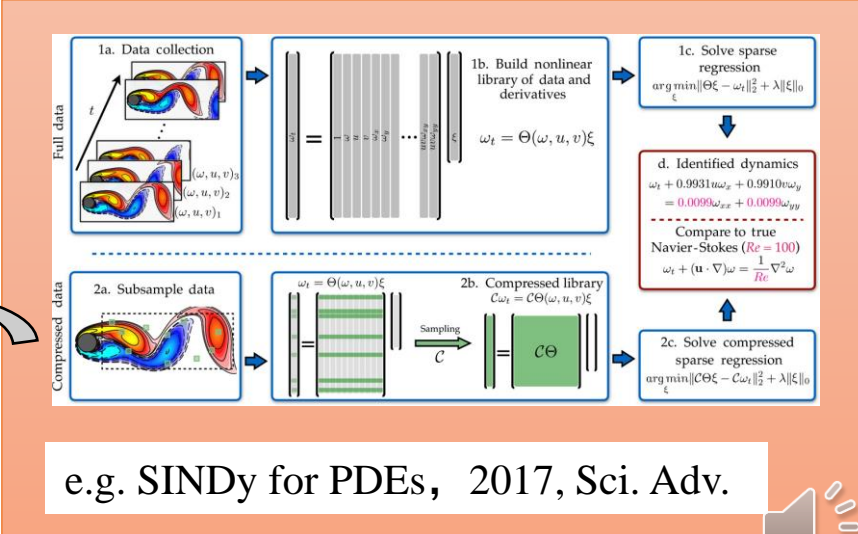
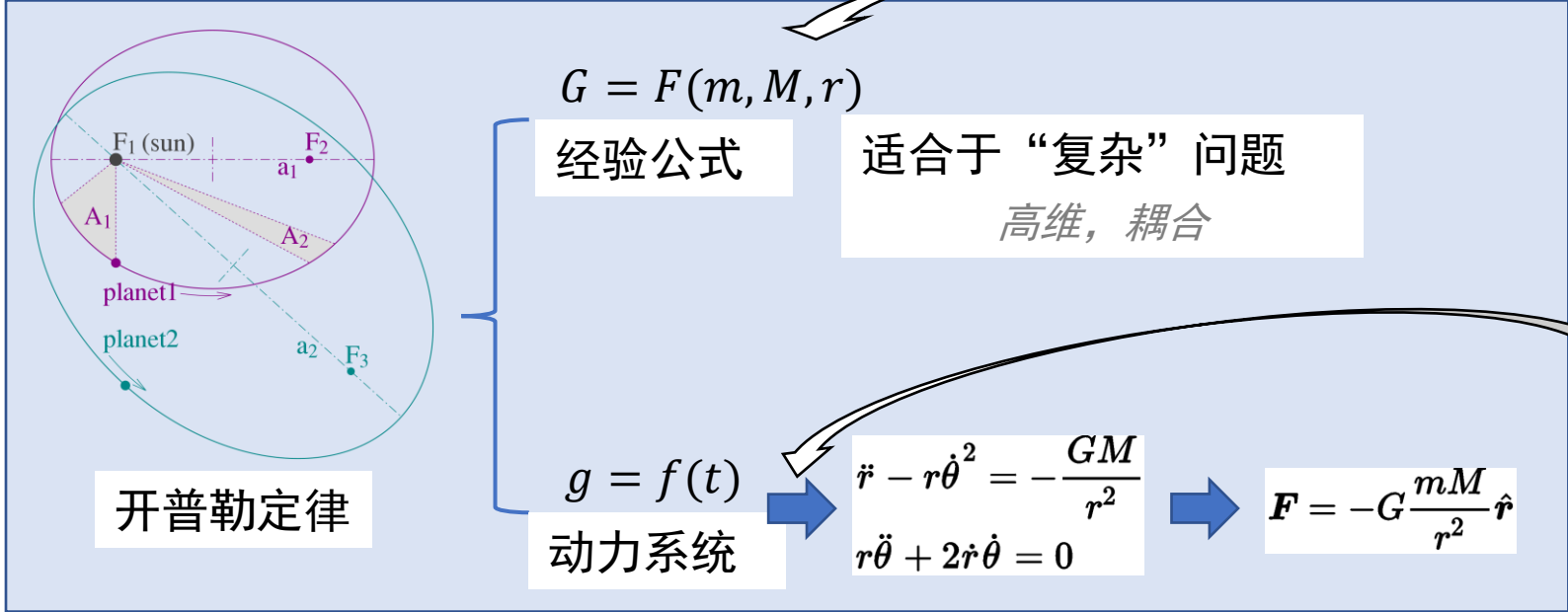
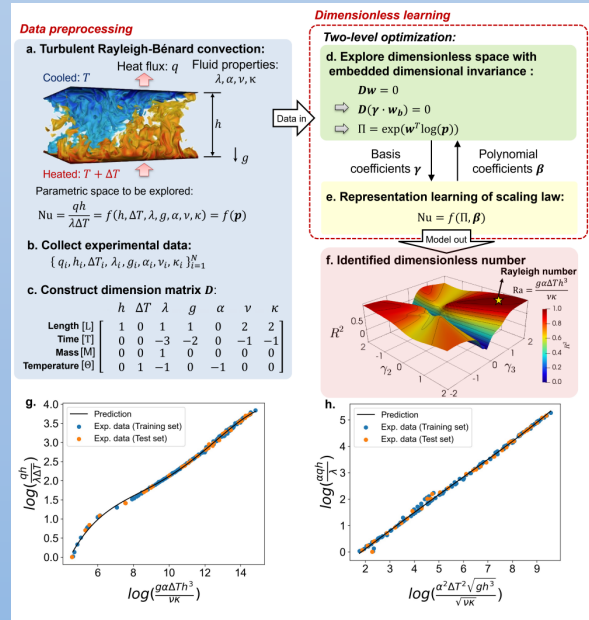
1.0 力学领域的机器科学家——现状

科学问题的一般流程：

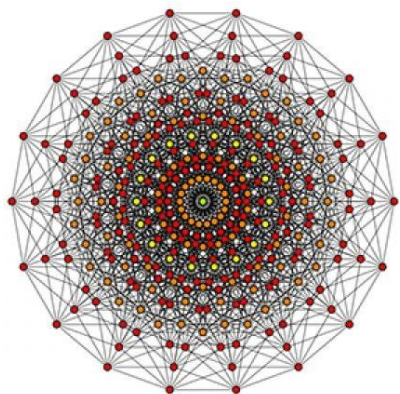
- 数据收集；
- 规律发现；
- 机理解释；
- 新理论和新方程；
- 新理论检验和拓展。



e.g.
Gan ZT et al.,
D.D. Disc dimensionless numbers
2021
arXiv.



1.1 数据驱动量纲分析研究背景——数据压缩之路



环境力学问题的主要特点？

1. 多相多介质；
2. 复杂和高维度；
3. 耦合性；
4. 工程需求。



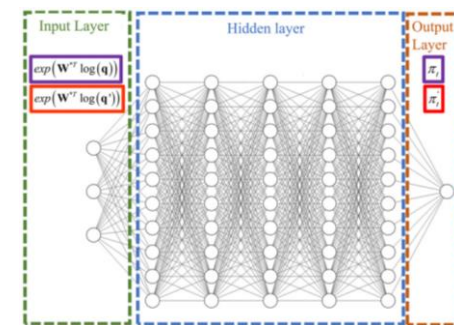
显式表达，如公式

$$\mathbf{F} = -G \frac{mM}{r^2} \hat{\mathbf{r}}$$

或分区公式

$$\begin{aligned} \mathcal{G} &= \mathcal{F}_a(\mathbb{T}_{a1}, \mathbb{T}_{a2}) \text{ when } A \\ \mathcal{G} &= \mathcal{F}_b(\mathbb{T}_{b1}, \mathbb{T}_{b2}) \text{ when } B \\ \mathcal{G} &= \mathcal{F}_c(\mathbb{T}_{c1}, \mathbb{T}_{c2}) \text{ when } C \end{aligned}$$

或



黑箱表达，如NN
Gan, 2021

$$G = \mathbf{F}(m, M, r, Q, d, D, \omega, \dots)$$

“复杂/高维”经验公式， \mathbb{R}^n

Constantine; Luo; Xu; Gan

基于自变量特点的压缩

基于函数特点的压缩

$$GT^2/MD = \mathbf{F}(M/m, r/D, Q, d/D, \omega T/D, \dots)$$

“无量纲”经验公式， \mathbb{R}^{n-m}

$$GT^2/MD = \mathcal{F}(\mathbb{T}1, \mathbb{T}2)$$

紧凑“无量纲”经验公式， \mathbb{R}^{n-m-1}

量纲分析：白金汉定理（ Π 定理）

数据驱动量纲分析

减少问题维度
反映物理本质
模型化（相似律）

无量纲数组形式不唯一
相对重要程度无法识别
面对复杂问题需要先验分区

给出 \mathcal{F} 的表达

1.1 数据驱动量纲分析研究背景——数据驱动量纲分析



Constantine et al. (2017)

认为经验数学模型都可以写成脊函数形式

根据脊函数偏微分获取参数空间特征矩阵

识别脊函数明显变化方向S

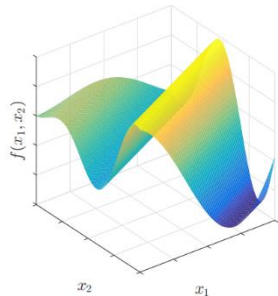
按照S方向旋转零空间矩阵W

需先验分区!

$$C = \int \nabla f(X) \nabla f(X)^T \sigma(X) dX$$

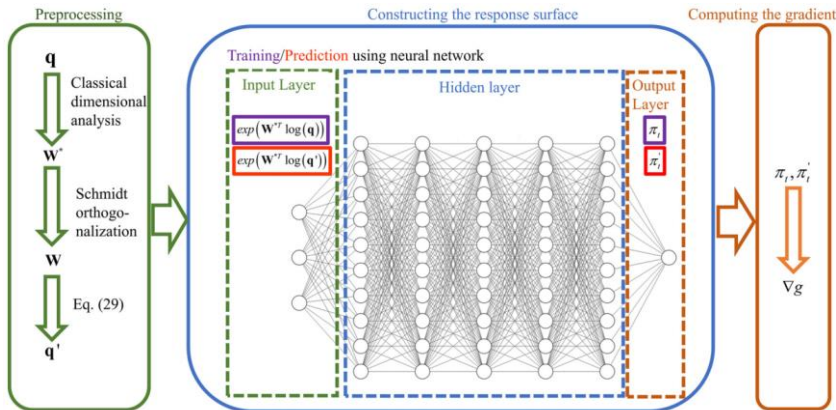
$$C = S\varphi S^T$$

$$Z = WS$$



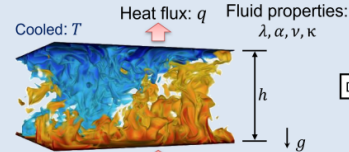
Xu et al. (2022)

机器学习替代高斯回归等求取脊函数偏微分



Data preprocessing

a. Turbulent Rayleigh-Bénard convection:



Parametric space to be explored:

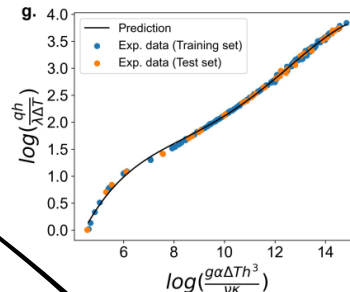
$$Nu = \frac{qh}{\lambda \Delta T} = f(h, \Delta T, \lambda, g, \alpha, \nu, \kappa) = f(p)$$

b. Collect experimental data:

$$\{q_i, h_i, \Delta T_i, \lambda_i, g_i, \alpha_i, \nu_i, \kappa_i\}_{i=1}^N$$

c. Construct dimension matrix D:

	h	ΔT	λ	g	α	ν	κ
Length [L]	1	0	1	1	0	2	2
Time [T]	0	0	-3	-2	0	-1	-1
Mass [M]	0	0	1	0	0	0	0
Temperature [Θ]	0	1	-1	0	-1	0	0



Dimensionless learning

Two-level optimization:

d. Explore dimensionless space with embedded dimensional invariance:

$$Dw = 0$$

$$D(\gamma \cdot w_b) = 0$$

$$\Pi = \exp(w^T \log(p))$$

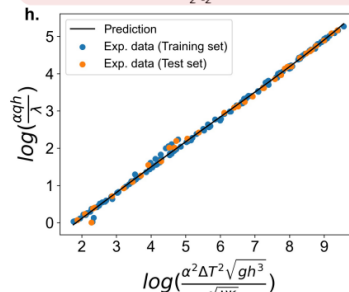
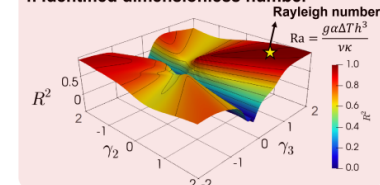
Basis coefficients γ

Polynomial coefficients β

e. Representation learning of scaling law:

$$Nu = f(\Pi, \beta)$$

f. Identified dimensionless number



Xie et al. (2021)

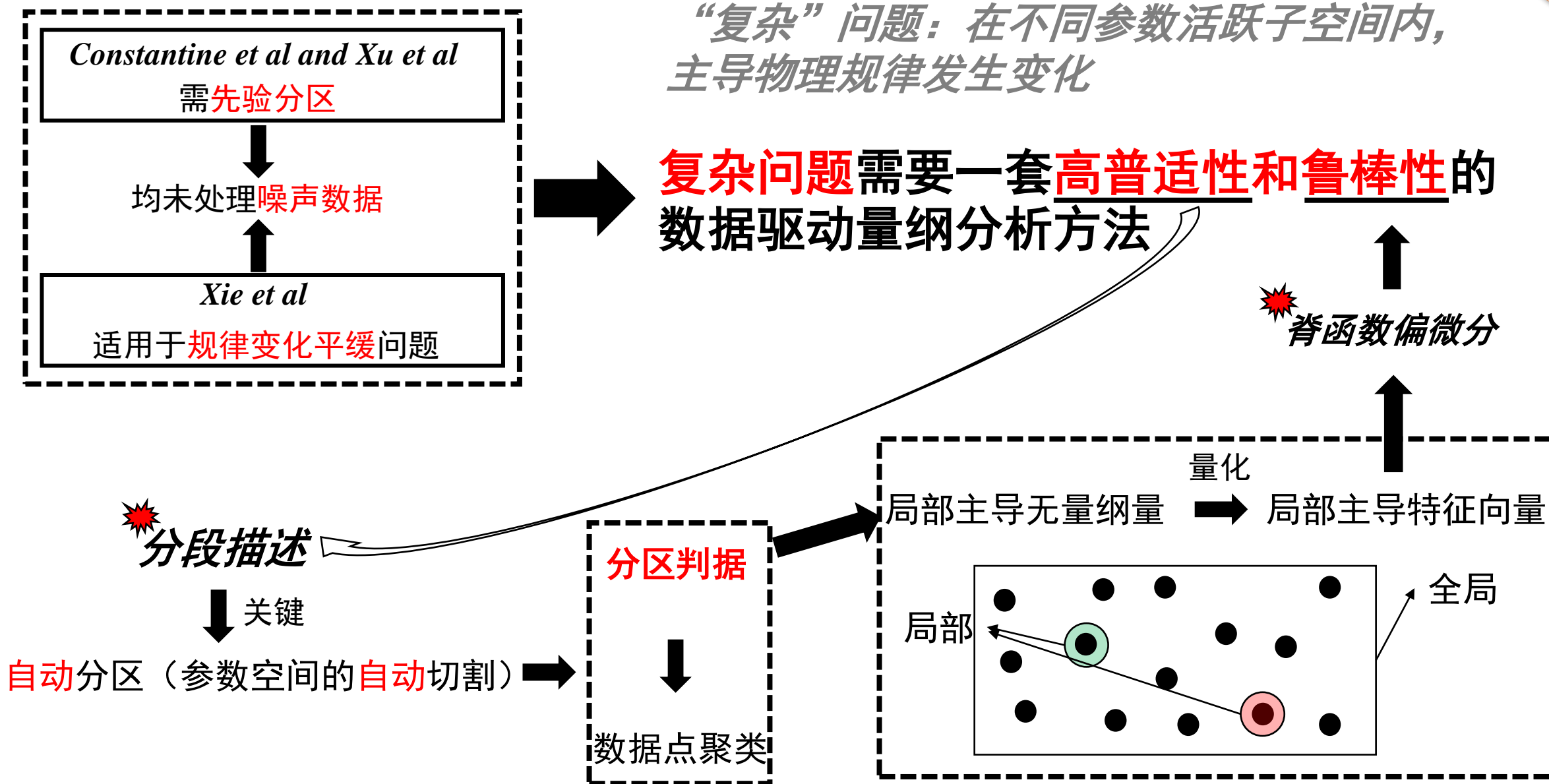
$$w = \gamma_1 w_{b1} + \gamma_2 w_{b2} + \gamma_3 w_{b3}$$

$$Nu = \beta_0 + \beta_1 \Pi + \beta_2 \Pi^2 + \dots + \beta_5 \Pi^5$$

实现连续性描述



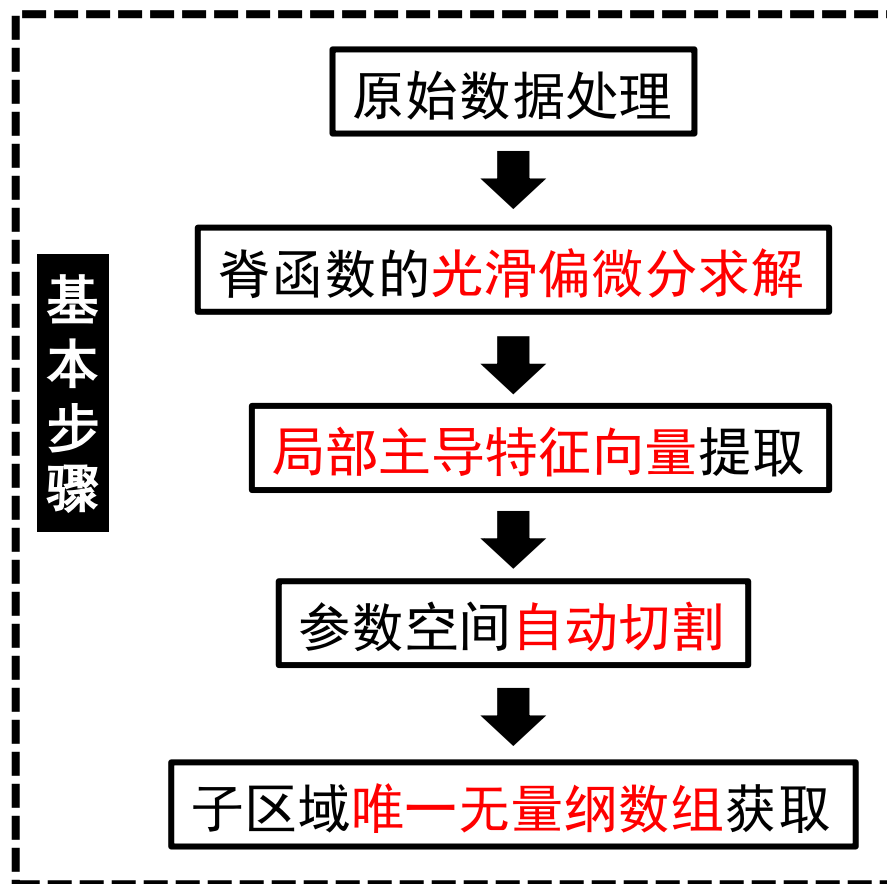
1.2 数据驱动量纲分析研究背景——存在问题、解决思路



2 适用于复杂力学问题的分区数据驱动量纲分析方法



适用于复杂力学问题的数据驱动量纲分析方法



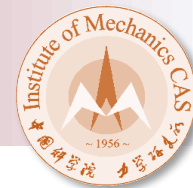
自动分区实现分段规律描述

$$\begin{aligned} \mathcal{G} &= \mathcal{F}_a(\mathbb{T}_{a1}, \mathbb{T}_{a2}) \text{ when } A \\ \mathcal{G} &= \mathcal{F}_b(\mathbb{T}_{b1}, \mathbb{T}_{b2}) \text{ when } B \\ \mathcal{G} &= \mathcal{F}_c(\mathbb{T}_{c1}, \mathbb{T}_{c2}) \text{ when } C \end{aligned}$$

适应高噪声



2.1 方法详解——原始数据处理



传统量纲分析

获得自变量维数矩阵 D 和初始零空间矩阵 W :

$$DW = [0]_{v \times n} \quad (1)$$

v 为基础量纲（长度，时间，质量等）数量， n 为无量纲自变量个数



获得 n 个无量纲自变量 γ （ $n = m - v$ ， m 为初始自变量个数）；
假设问题只包含一个因变量，将其转化为无量纲因变量 δ ，他们之间的关系可表示为：

$$\delta = g(\gamma) \quad (2)$$



将式（2）转化为连续函数形式：

$$\delta = f(\pi) \quad (3)$$

假设问题包含 d 组数据，经上述操作后，最终获得： $[\delta_o]_{d \times 1}$ $[\pi_o]_{d \times n}$



2.1 方法详解——脊函数的光滑偏微分求解

光滑偏微分求解

$$\langle \nabla f(x) \rangle = \sum_{j=1}^N \frac{m_j}{\rho_j} f(x_j) \cdot \nabla_i W_{ij} \quad (4)$$

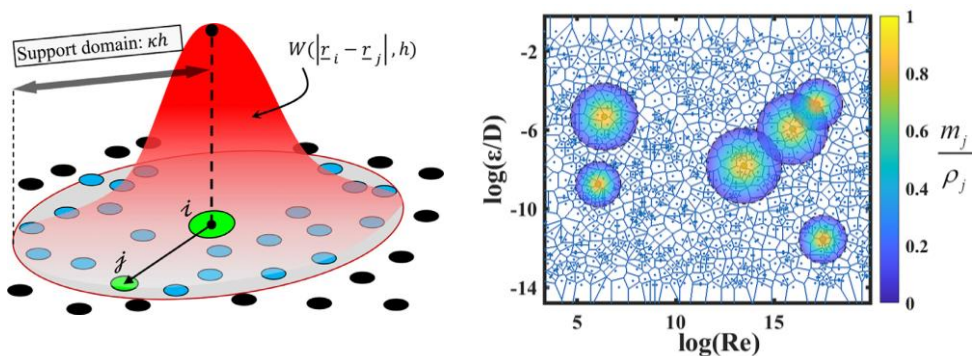
$\frac{m_j}{\rho_j}$: 数据点控制区域大小, 称为权重

$$W(R, h) = \alpha_d e^{-R^2}$$

$$R = \frac{r}{h} = \frac{|x - x_j|}{h}$$

α_d 在一维二维, 三维空间中分别为 $\frac{1}{\pi^{0.5}h}$ $\frac{1}{\pi h^2}$ $\frac{1}{\pi^{1.5}h^3}$

$$\nabla_i W_{ij} = \frac{x_i - x_j}{r_{ij}} \cdot \frac{\partial W_{ij}}{\partial r_{ij}} = \frac{x_{ij}}{r_{ij}} \cdot \frac{\partial W_{ij}}{\partial r_{ij}}$$



1. 无需先插值后求导, 直接计算偏微分
2. QHull算法计算权重, 目前支持9维以下

可变光滑长度 \rightarrow 核函数支持域内迭代收敛

将每个所求点的最优光滑长度 h_{best}^j 带入式 (4)

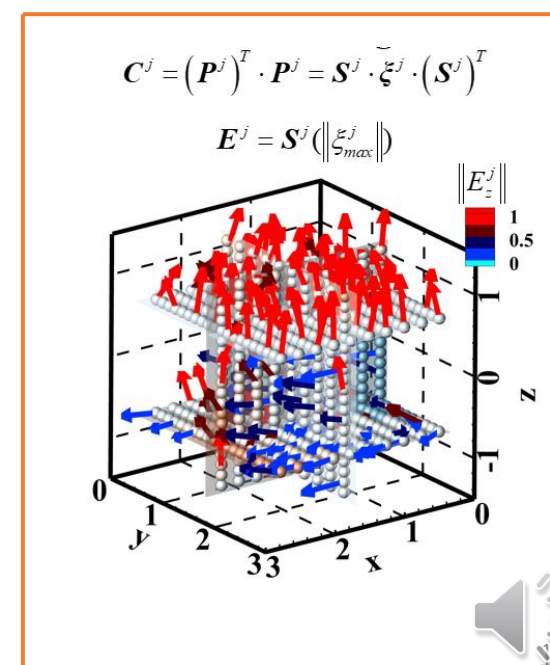
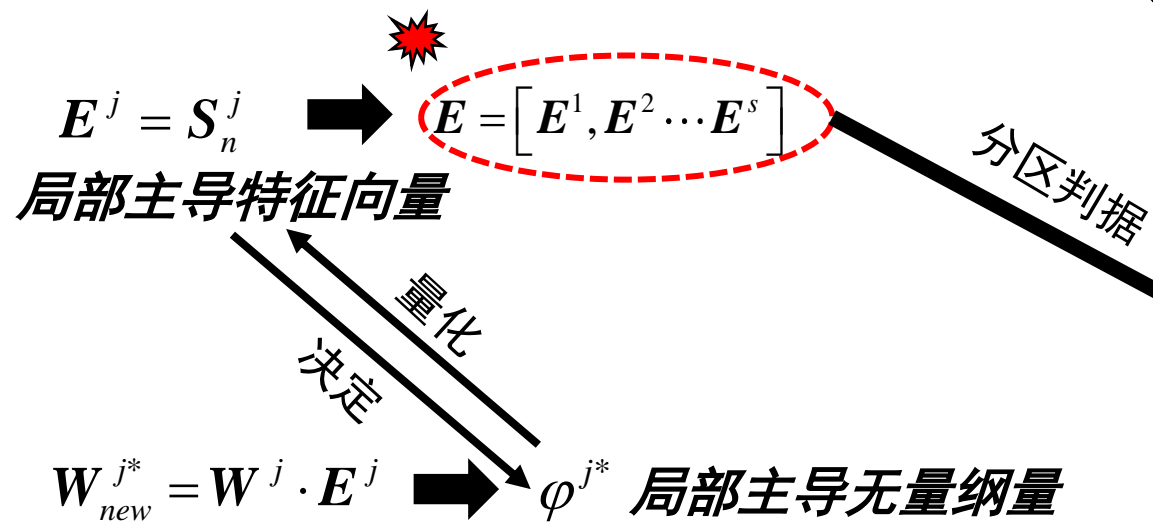
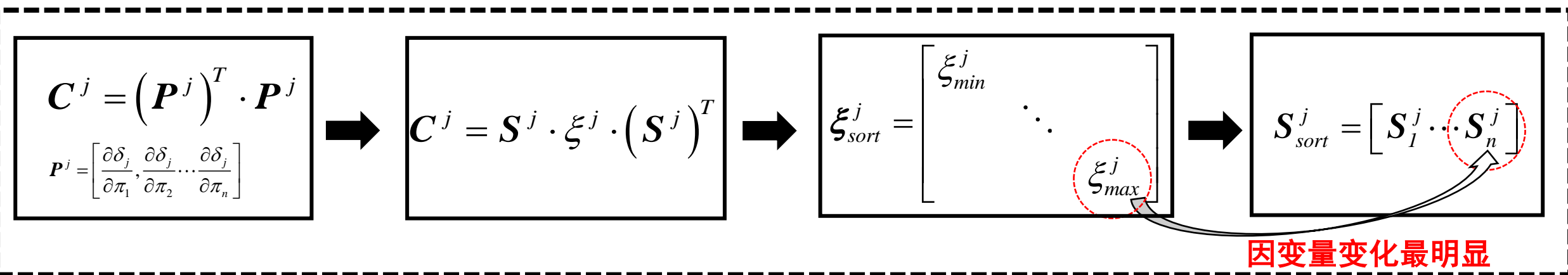
$$[P]_{s \times n} = \begin{bmatrix} \frac{\partial \delta_1}{\partial \pi_1} & \dots & \frac{\partial \delta_1}{\partial \pi_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \delta_s}{\partial \pi_1} & \dots & \frac{\partial \delta_s}{\partial \pi_n} \end{bmatrix}$$

$[\pi]_{s \times n}$: 所求数据点, 个数为 s

$[P]_{s \times n}$: 每点的离散偏微分值



2.1 方法详解——局部主导特征向量提取



2.1 方法详解——参数空间自动切割

参数空间自动切割——局部主导特征向量聚类

模糊聚类 (Fuzzy C-Means (FCM))

数据对中心点的隶属度在0到1之间

更客观

硬聚类 (K-means)

数据对中心点的隶属度只有0或1



FCM聚类

$$J = \sum_{i=1}^s \sum_{j=1}^c (U_j^i)^m \|E^i - V^j\|^2$$

中心点个数

$$V^j = \frac{\sum_{i=1}^s (U_j^i)^m E^i}{\sum_{i=1}^s (U_j^i)^m}$$

中心点坐标

$$U_j^i = \frac{1}{\sum_{l=1}^c \left(\frac{\|E^i - V^j\|}{\|E^i - V^l\|} \right)^{\frac{2}{m-1}}}$$

隶属度矩阵

中心点信息直接影响聚类结果

预测中心点信息

Subclust算法

$$p_i = \sum_{j=1}^s e^{-\alpha \|E^i - E^j\|^2}$$

$$\alpha = \frac{4}{r^a}$$

$$E^{1*} = E^i (P_{i-max})$$

revise

$$p_i \leftarrow p_i - p_1^* e^{-\beta \|E^i - E^{1*}\|^2}$$

$$\beta = \frac{4}{r^b}$$

$$r^b = O^q \times r^a$$

[1.25, 1.5]

选取候选点

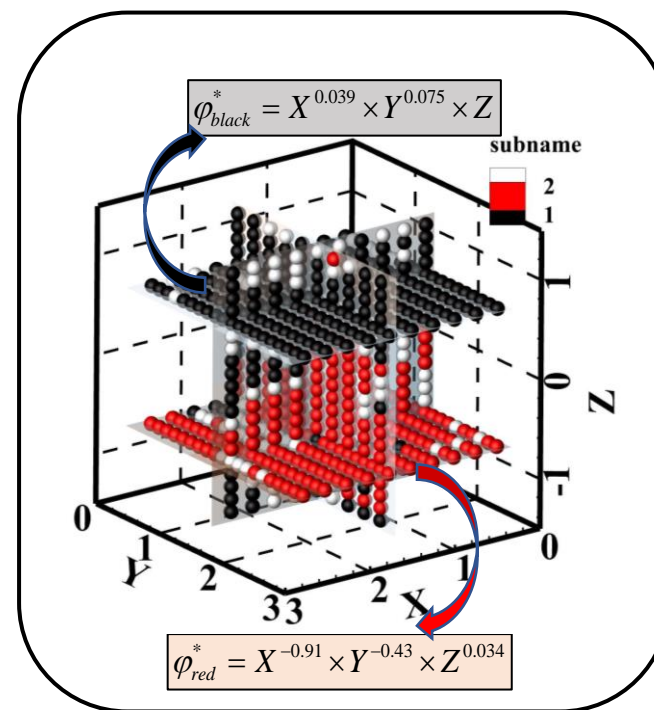
The second cluster center can be chosen by the following criteria:^[4]
 $\bar{\zeta}$ is the accept factor, $\underline{\zeta}$ is the reject factor;^[4]
 If $\frac{p_i}{p_1} > \bar{\zeta}$, E^i is accepted as the candidate point of the second cluster center;^[4]
 Else if $\underline{\zeta} < \frac{p_i}{p_1} < \bar{\zeta}$, appended judgment is performed;^[4]
 Making d_m is the shortest distance between E^i and former cluster points;^[4]
 If $\frac{d_m}{r^a} + \frac{p_i}{p_1} \geq 1$, E^i is accepted as the candidate point of the second cluster center;
 Else, E^i is removed;^[4]
 Else, E^i is removed.^[4]

重复上述操作获得后续中心点, 直到候选点的个数为0

2.1 方法详解——子区域唯一无量纲数组获取

$$C_k = \int \nabla f_k(\pi_k) \cdot \nabla f_k(\pi_k)^T d\pi_k \approx \sum_{j=1}^{s_k} (P_k^j)^T P_k^j$$
$$\downarrow$$
$$C_k = S_k \cdot \xi_k \cdot S_k^T$$
$$\downarrow$$
$$W_k = W \cdot S_k$$
$$\downarrow$$
$$\delta_k = g_k(\varphi_k)$$

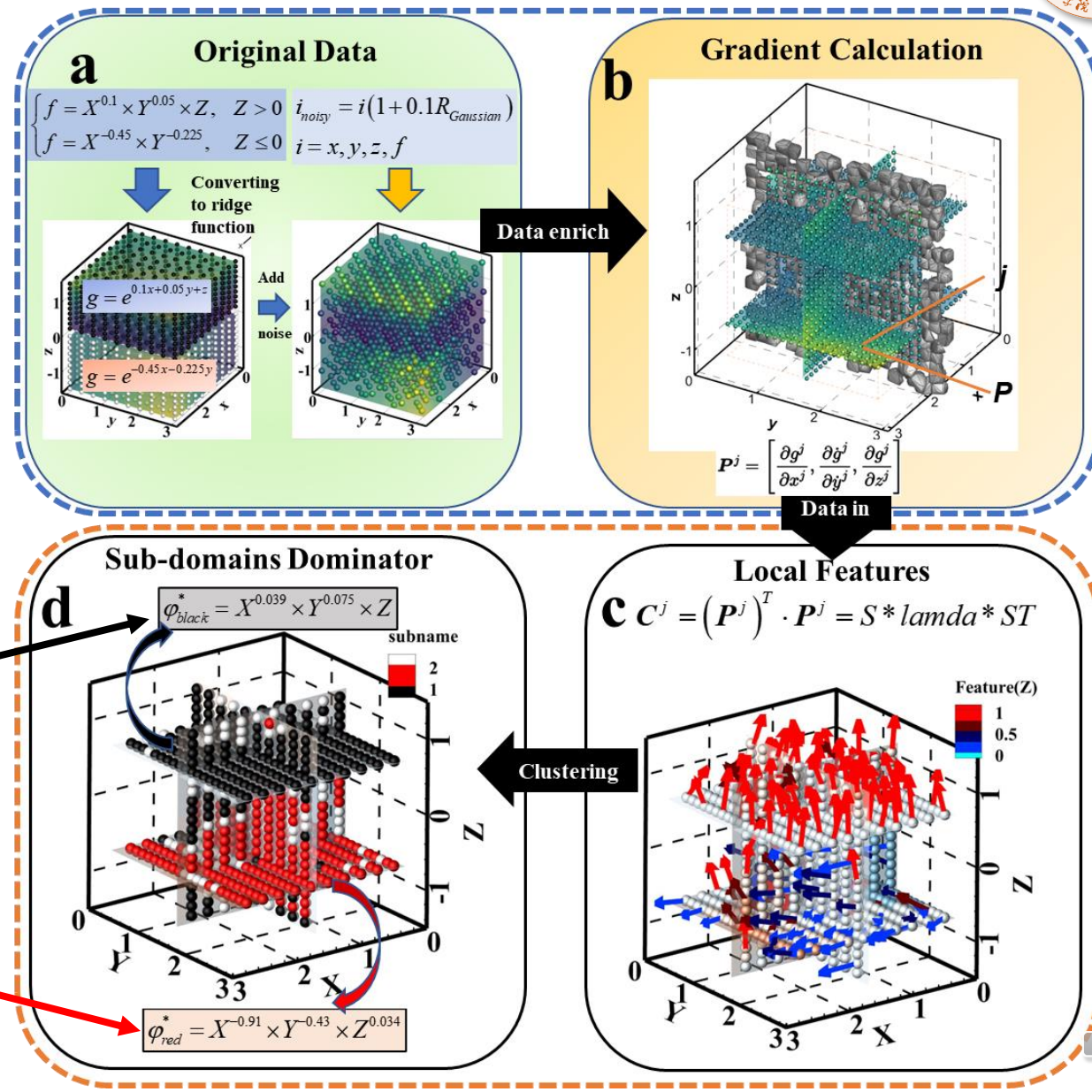
- 有效性?
- 鲁棒性?
- 人为参数的影响?



2.2 方法可行性验证

参数空间实现**自动切割**

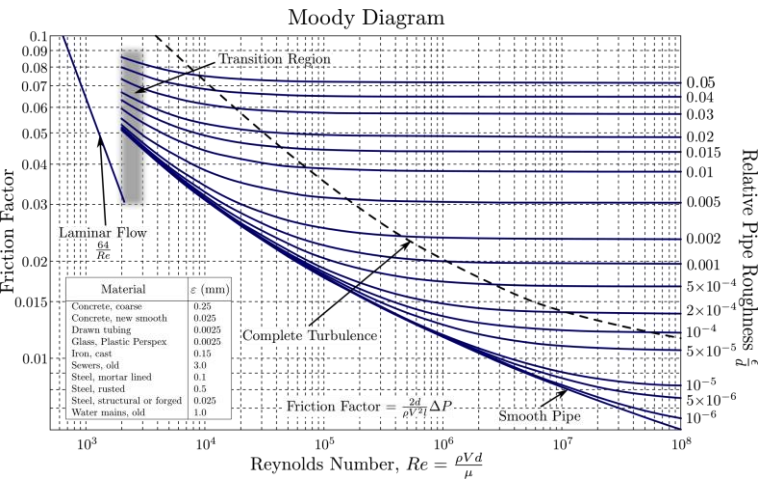
子空间主导无量纲量**获取**



$$r_a = 0.9, o^q = 1.25, \bar{\zeta} = 0.5, \underline{\zeta} = 0.15, m = 2$$

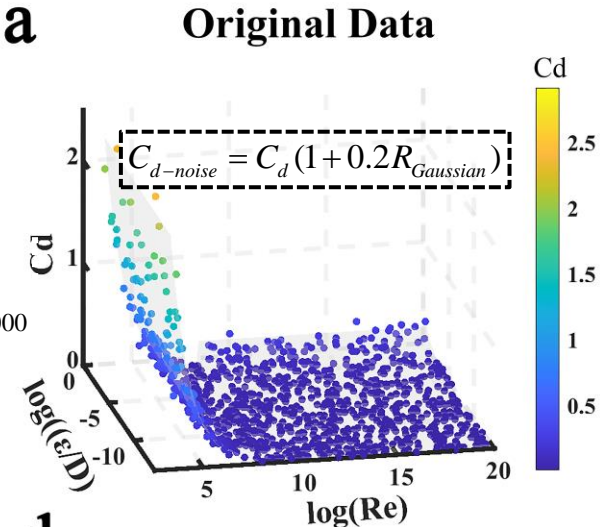
$S_{black} =$	$\begin{bmatrix} 5.7 \times 10^{-1} & 8.2 \times 10^{-1} & 3.9 \times 10^{-2} \\ -8.2 \times 10^{-1} & 5.7 \times 10^{-1} & 7.5 \times 10^{-2} \\ 3.9 \times 10^{-2} & 7.5 \times 10^{-2} & 1.0 \end{bmatrix}$
$S_{red} =$	$\begin{bmatrix} -2.9 \times 10^{-1} & -3.0 \times 10^{-1} & -9.1 \times 10^{-1} \\ 6.8 \times 10^{-1} & 6.0 \times 10^{-1} & -4.3 \times 10^{-1} \\ -6.7 \times 10^{-1} & 3.3 \times 10^{-2} & 3.4 \times 10^{-2} \end{bmatrix}$

3 案例应用：管流阻力问题

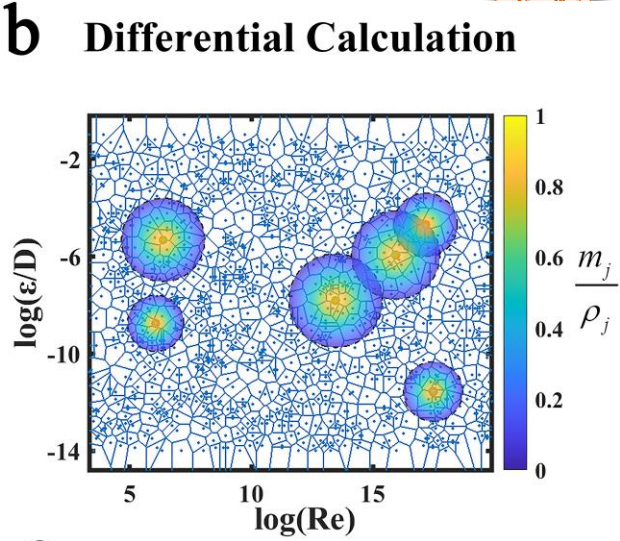


Moody图：管流D-W阻力系数

$$\begin{cases} C_d = \frac{64}{Re}, Re \leq 3000 \\ \frac{1}{\sqrt{C_d}} = -2\log_{10}\left(\frac{1}{3.7} \frac{\varepsilon}{D} + \frac{2.51}{Re\sqrt{C_d}}\right), Re > 3000 \end{cases}$$

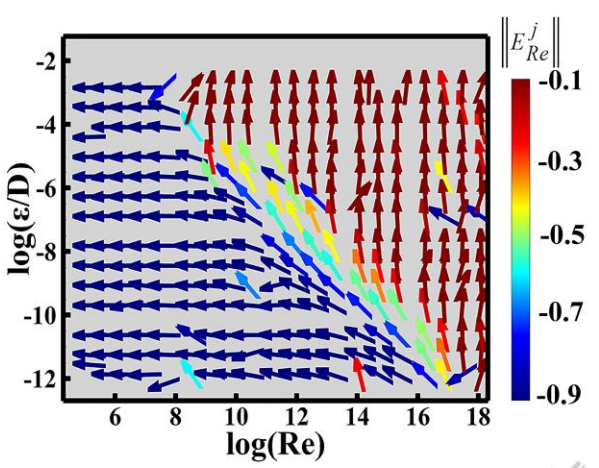
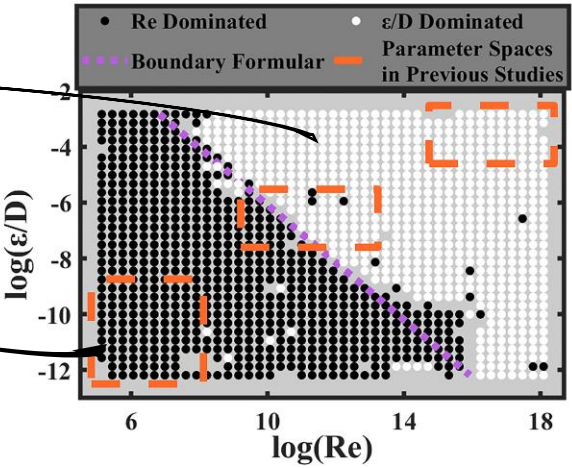


Parameter Space Partition



Local Dominated Eigenvectors

	$Re - dominated$	$\frac{\varepsilon}{D} - dominated$
S_k	$\begin{bmatrix} -2.0 \times 10^{-3} & -1.0 \\ -1.0 & 2.0 \times 10^{-3} \end{bmatrix}$	$\begin{bmatrix} -1.0 & -1.7 \times 10^{-3} \\ -1.7 \times 10^{-3} & 1.0 \end{bmatrix}$
φ_k^*	$Re^{-1} \times \frac{\varepsilon}{D}^{-2.0 \times 10^{-3}}$	$Re^{-1.7 \times 10^{-3}} \times \frac{\varepsilon}{D}^{1.0}$
f_k	$24.24 \times 10^{-3} + 78.21 \varphi_k^*$	$12.13 \times 10^{-3} + 0.30 (\varphi_k^*)^{0.5}$
$f_k - R^2$	0.9802	0.9699
$f_k - boundary$	$\log(\frac{\varepsilon}{D}) = -1.03 \log(Re) + 4.17$	



真实案例依旧可行

$$r_a = 0.5, o^q = 1.25, \bar{\zeta} = 0.5, \underline{\zeta} = 0.15, m = 2$$



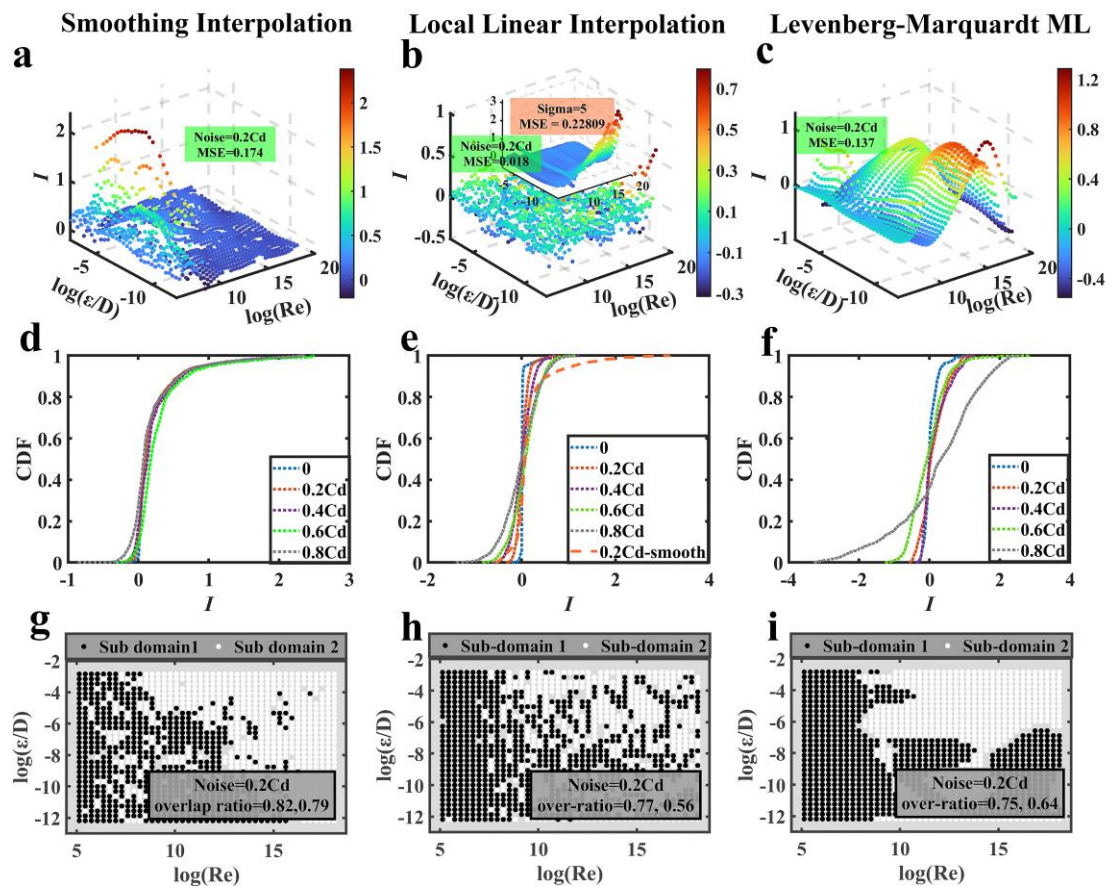
4.1 方法优缺点：鲁棒性

脊函数偏微分求解决定方法鲁棒性

预测规则点函数值+公式法偏微分求解

Function value prediction and Parameter Space Partition

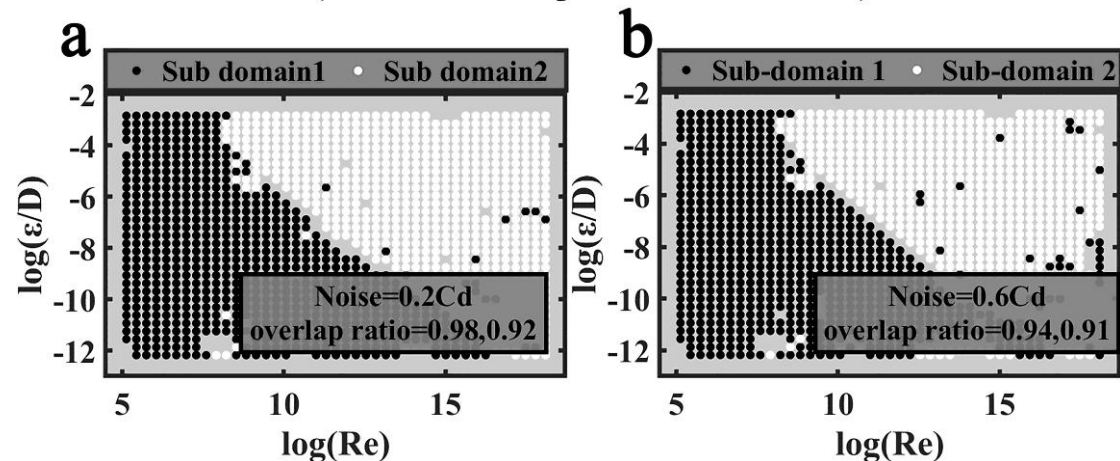
(Based on Three Kinds of Two Steps Differential Calculation)



光滑法偏微分求解

Parameter Space Partition

(Based on Smoothing Differential Calculation)



光滑法偏微分求解在相同噪声下，分区结果明显优于前者，且在较高噪声下仍能实现准确分区

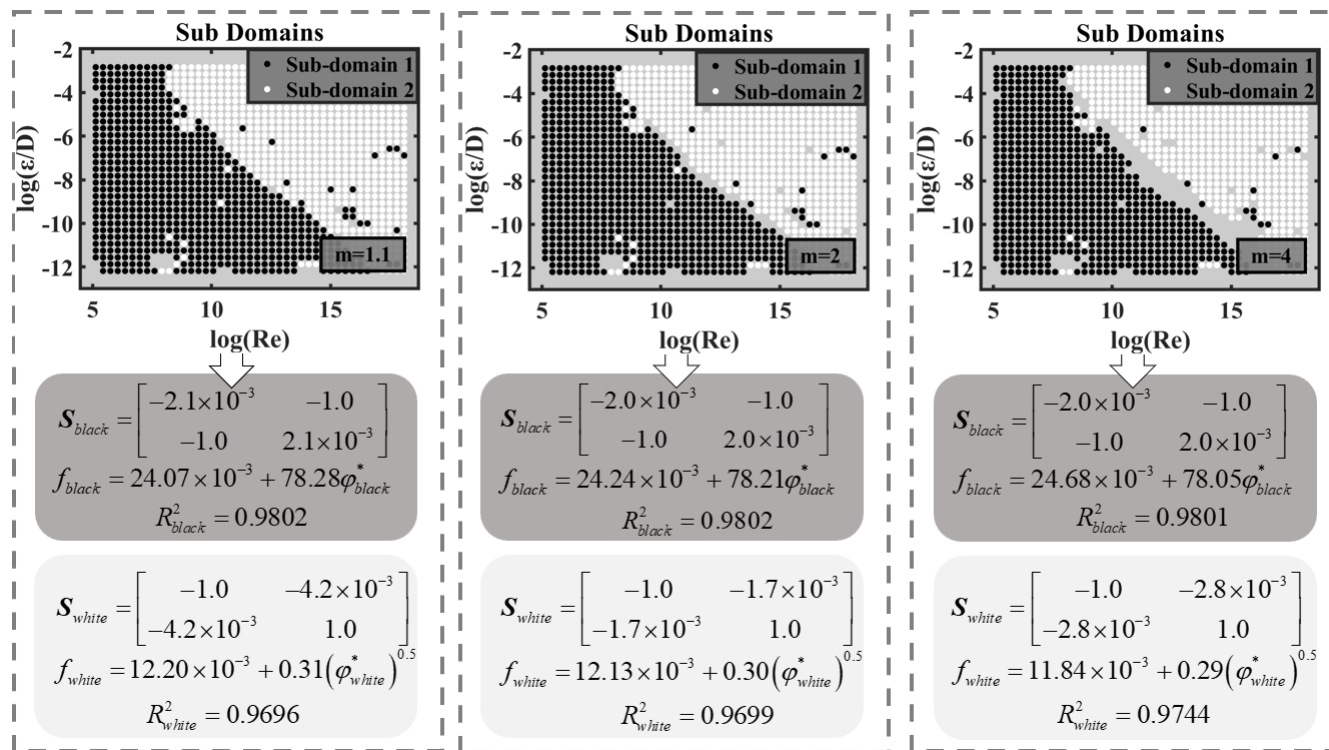
该方法具有强鲁棒性

4.2 方法优缺点：人为参数影响

隶属度矩阵

$$U_j^i = \frac{1}{\sum_{l=1}^c \left(\frac{\|E^i - V^j\|}{\|E^i - V^l\|} \right)^{\frac{2}{m-1}}}$$

$$r^a = 0.5$$

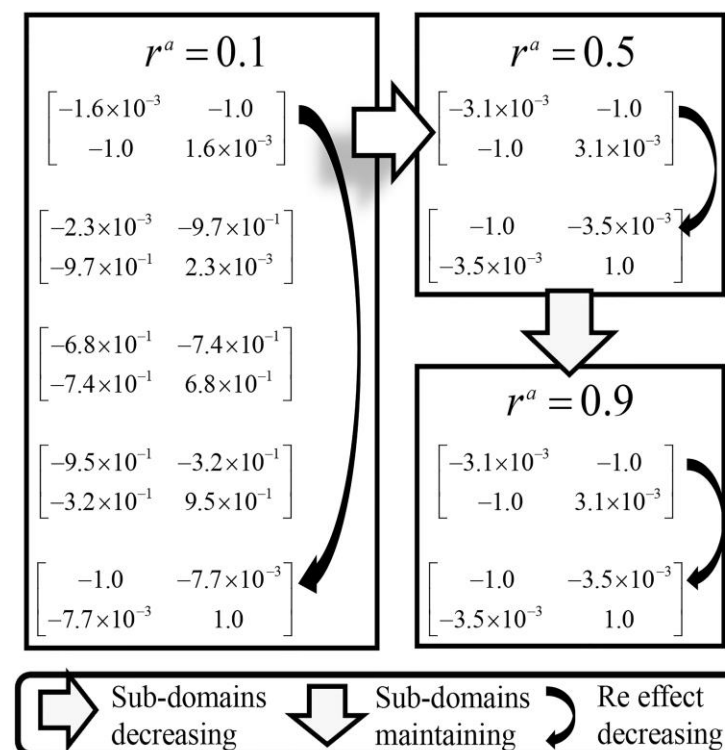


m的增大导致点的隶属度下降，**模糊点数目增多**，能获得较为纯净的子区域

$$p_i = \sum_{j=1}^s e^{-\alpha \|E^i - E^j\|^2}$$

$$\alpha = \frac{4}{r^a}$$

$$m = 2$$



r^a 的增大会导致子区域个数的减少，直至所剩子区域存在本质差异

发展了适用于复杂力学问题的数据驱动量纲分析方法

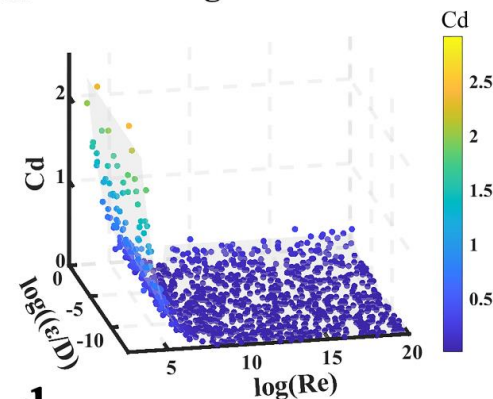
1. 方法适应性强，噪声忍耐度高；
2. 光滑偏微分求解保证了方法的鲁棒性；
3. 局部主导特征向量作为分区判据实现具有不同主导无量纲量的子区域的自动识别、分割；
4. 在一些复杂力学问题中可以较好的实现规律提取和发现。

感谢聆听！ 望各位专家同学批评指正

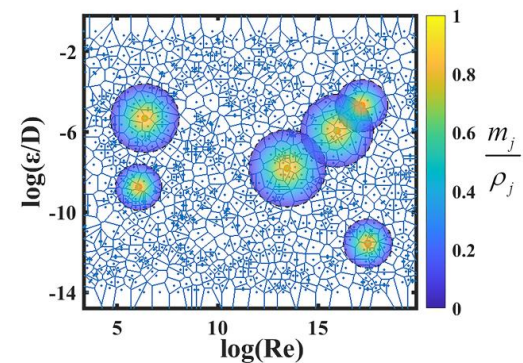


代码公开，欢迎索取&使用&合作

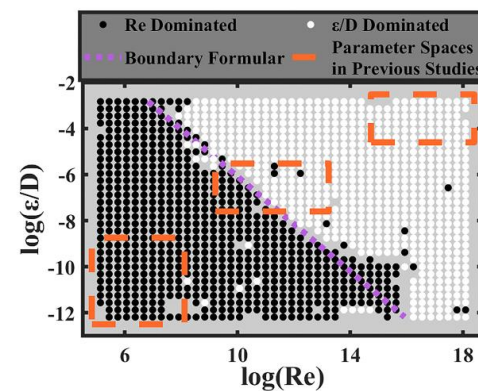
a Original Data



b Differential Calculation



d Parameter Space Partition



c Local Dominated Eigenvectors

