

基于粒子群采样的双精度数据融合建模



海春龙¹, 梅立泉^{1,*}, 钱炜祺²

¹ 西安交通大学数学与统计学院, 陕西西安 710049

² 中国空气动力研究与发展中心计算空气动力研究所, 四川绵阳 621050

摘要: 计算机模拟和试验测试是工业设计中的常用的数据获取手段, 因此综合使用以廉价计算数据为代表的低精度数据和以昂贵试验数据为代表的高精度数据进行融合建模是实际应用中的重要问题。为了提高模型的质量并控制高精度数据的使用量, 本文提出了基于粒子群采样的双精度融合建模方法。算法使用 Kriging 模型对高低精度数据建模, 并使用 PSO (Particle swarm optimization) 取低精度模型信息用于高精度建模。结合低精度数据数量充足, 高精度数据昂贵、不易获取的特点, 使用充分采样后建模的低精度模型指导高精度模型的初始采样, 使用高低精度两个模型训练得到融合模型, 并使用两个加点点准则进行序贯建模, 对融合模型进行持续优化。算法在一个标准函数和一个应用问题上进行了测试, 实验结果显示在不提高高精度数据用量的前提下, 算法改善了建模的质量, 验证了提出算法的有效性。

关键词: 双精度数据; 融合建模; 粒子群算法; 序贯试验设计; 克里金模型

DOI: [10.57237/j.wjms.2023.01.002](https://doi.org/10.57237/j.wjms.2023.01.002)

Two-fidelity Data Fusion Modeling with Particle Swarm Sampling

Hai Chunlong¹, Mei Liquan^{1,*}, Qian Weiqi²

¹School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China

²Computational Aerodynamics Institute, China Aerodynamics Research and Development Center, Mianyang 621050, China

Abstract: Computer simulation and experimental testing are commonly used means of data acquisition in industrial design, so the combination of low-precision data represented by cheap calculation data and high-precision data represented by expensive test data for fusion modeling is a hot topic in practical applications question. In order to improve the quality of the model and control the usage of high-precision data, this paper proposes a double-precision fusion modeling method based on particle swarm sampling. The algorithm uses the Kriging model to model high- and low-precision data, and uses PSO (Particle swarm optimization) to obtain low-precision model information for high-precision modeling. Combined with the characteristics of sufficient amount of low-precision data, high-precision data is expensive and difficult to obtain, the low-precision model modeled after sufficient sampling is used to guide the initial sampling of the high-precision model, and the fusion model is obtained by training with two high-precision and low-precision models, and two The point-adding criterion is used for sequential modeling, and the fusion model is continuously optimized. The algorithm is tested on a standard function and an application problem. The experimental

基金项目: 国家自然科学基金项目 (12171385).

*通信作者: 梅立泉, lqmei@mail.xjtu.edu.cn

收稿日期: 2023-03-27; 接受日期: 2023-05-17; 在线出版日期: 2023-05-25

<http://www.wjoms.com>

results show that the algorithm improves the quality of the modeling without increasing the amount of high-precision data, which verifies the effectiveness of the proposed algorithm.

Keywords: Two-fidelity Data; Fusion Modeling; Particle Swarm Optimization; Sequential Design of Experiments; Kriging Model

1 引言

在复杂工程系统中, 对设计变量和目标变量间关系的研究有丰富的技术手段, 最为常见的是以有限元方法 (FEM, finite element method) 为代表的计算技术和科学试验技术[1-2]。因此这两项技术获取的数据是工程系统建模的重要数据来源。有限元计算往往使用抽象简化的方程进行以方便进行大规模计算并节省计算时长, 因此精度普遍较低, 但数据量十分丰富。而科学试验从准备阶段直到数据采集都伴随着较高的成本, 尤其是涉及多学科、多参数的试验, 例如飞行器布局的设计中往往需要确定大量的参数[3-4]。试验获得的数据精度高, 但花费巨大, 因此从控制成本的角度考虑, 需要减少试验次数。为了达到降低成本, 但仍能保证建模精度的目的, 需要充分挖掘计算数据的可用信息, 然后结合试验数据建模。

在此背景下, 引出一个双精度数据优化建模问题[5]。目前有很多学者对双精度建模问题进行了研究。例如[6-7]中, 首先使用大量的低精度数据建立低精度神经网络模型, 并将低精度神经网络模型和少量高精度数据一起训练得到高精度建模, 核心观点是建立低精度数据和高精度数据的对应关系, 以此达到预测的目的。

考虑到计算所用的简化模型和实际场景间关系的复杂性, 利用少量高精度数据难以捕捉低精度数据和高精度数据的对应关系。本文从信息挖掘的角度利用低精度数据, 使用从低精度数据获取的信息指导建模的选点, 并且在模型构建中赋予低精度模型一定的权重, 以起到对模型的调整作用。

文章接下来的内容分别为: 第二节简要介绍了算法使用的相关技术; 第三节对算法做了详细阐述; 第四节给出了实验结果展示; 最后对全文做出了总结。

2 优化及建模技术

算法的基础框架是双精度的序贯加点建模过程, 使用 Kriging 模型[8-9]进行高低精度的建模, 序贯加点[10]是由 PSO 优化算法[11]搜索产生的, 该部分对

Kriging 模型和 PSO 算法进行简要介绍。

2.1 Kriging 模型

设数据 $\{(\mathbf{x}_1, y(\mathbf{x}_1)), (\mathbf{x}_2, y(\mathbf{x}_2)), \dots, (\mathbf{x}_N, y(\mathbf{x}_N))\}$,

其中 $\mathbf{x} \in \mathbf{R}^n$, Kriging 模型利用公式(1)来学习构造出 $f(\mathbf{x})$

$$y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon \quad (1)$$

其中 $\varepsilon \sim N(0, \sigma^2)$, 是一组独立且均匀分布的噪声。

对于未知点 \mathbf{x}_* , 其预可表示为 $f(\mathbf{x}_*)$ 。

Kriging 模型使用一个核函数 k 来估计未知点 \mathbf{x}_* 与已知建模数据点 \mathbf{x} 的相关性, 在本文中选取平方指数函数为核函数(2):

$$k(x, x') = \theta \exp\left(-\frac{1}{2l^2} (x - x')^T (x - x')\right) \quad (2)$$

其中 θ, l 为超参数, 本文所展示的算法中关于 Kriging 模型的部分调用 MATLAB dace 工具包完成, 其中的参数设置为 $\theta = 0.01, l \in [0.0001, 20]$ 。

2.2 粒子群优化算法

美国学者 Eberhart 和 Kennedy [11]在 1995 年提出了粒子群优化算法 (PSO, particles swarm optimization), 现在已经广泛的应用于各种工程优化问题之中。

粒子群优化算法首先在建模空间中初始化一组自变量点, 这些点跟据它们自身的函数值情况更新自身位置, 具体计算公式如下:

$$\mathbf{x}^{t+1} = \mathbf{x}^t + \mathbf{v}^{t+1} \quad (3)$$

$$\mathbf{v}^{t+1} = \omega \mathbf{v}^t + a_1 r_1 (\mathbf{x}_p - \mathbf{x}^t) + a_2 r_2 (\mathbf{x}_g - \mathbf{x}^t) \quad (4)$$

其中 \mathbf{x}_p 是个体的自身历史最优位置, \mathbf{x}_g 是粒子群

的历史最优位置。 r_1, r_2 均是 $[0,1]$ 间的随机数。粒子群每次更新的速度包含三个部分, 由公式(3)可以将三部分描述为惯性分量、向自身最优位置学习的分量和向群体最优学习的分量。而三个分量的系数的大小也决定了粒子群搜索的偏向性: ω 逐渐变大, 则搜索的随机性越小; a_1 越大, 越偏向于局部寻优; 而 a_2 越大, 粒子群越快收敛, 迅速降低多样性向群体最优靠拢。一般这些参数的值会根据问题的需要调整。由于本文所提出的算法不需要特殊的偏向性, 根据参考文献[12], 本文将 PSO 的参数 ω 设为 0.729, a_1, a_2 均为 1.49445。

3 基于粒子群采样的双精度数据融合建模

为了充分利用低精度数据提供的信息, 并将获得的信息用于高精度数据的建模中, 本文提出了一种基于粒子群采样的双精度数据融合建模方法 (TDF-PSS, Two-fidelity data fusion modeling with particle swarm sampling)。首先, 现代实验设计 (MDOE, Modern Design of Experiments) [13-14] 为基于粒子群采样的双精度数据融合建模提供了基本框架, 如图 1 所示。

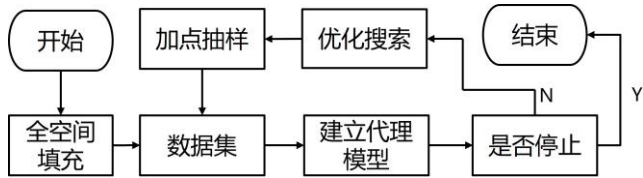


图 1 MDOE 过程

首先是进行全空间填充并计算数据点的函数值, 获得建模数据集进而建立代理模型, 然后使用一个搜索方式产生加点点抽样, 并计算新抽样点的函数值, 丰富数据集, 循环迭代直至达到停止条件, 随着数据集的不断丰富提高建模质量。而在框架内的具体内容上有如下几点不同。

在双精度数据融合建模中, 数据集部分包含两个精度级别数据, 因此也需要建立两个代理模型, 并完成模型的融合。融合模型 F_D 的建立如下:

$$F_D = \lambda_1 F_L + \lambda_2 F_H \quad (5)$$

$$MSE(\lambda_1, \lambda_2) = \frac{1}{N_T} \sum_{i=1}^{N_T} (\lambda_1 F_L(\mathbf{x}_i) + \lambda_2 F_H(\mathbf{x}_i) - y_i)^2 \quad (6)$$

其中 F_L, F_H 分别为低精度模型和高精度模型, λ_1, λ_2 是两者在最终的融合模型 F_D 中所占的权重。 \mathbf{x}_i 为训练集 T 中的点, 使用训练集计算 $MSE(\lambda_1, \lambda_2)$, 并选取使其达到最小值的 λ_1, λ_2 , 这样就完成了融合模型 F_D 的建立。关于训练集的描述在下文中算法的详细介绍里。

然后是 PSO 在代理模型上搜索加点, 需要注意的是本文使用了两种加点抽样来提高样本集的多样性, 两种加点准则由公式(7), (8)所描述, 最终达到停止条件时结束。图 2 给出了 TDF-PSS 的流程图。

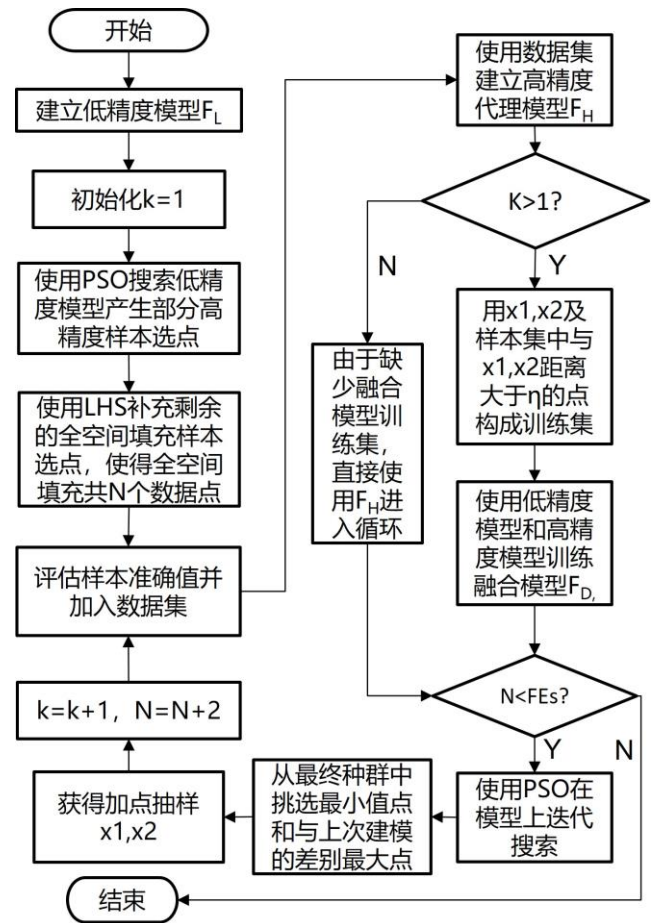


图 2 TDF-PSS 流程图

算法的具体实现步骤如下:

步骤 1 首先建立低精度数据的 Kriging 模型。并初始化 $k=1$ 。需要说明的是, 由于低精度数据十分丰富, 因此获取的 Kriging 模型对低精度流行具有很好的近似。

步骤 2 使用 PSO 算法对低精度模型进行反复搜索寻优, 需要注意的是, 这里要搜索最小值和最大值, 以搜索最小值为例, 具体操作过程如下:

1) 初始化粒子群, 并进行 MG 次迭代;

- 2) 挑选最终粒子群 S 中值最小的点加入 Min_Set ;
- 3) 循环 1, 2 步 SG 次;
- 4) 剔除掉 Min_Set 中距离小于 η 的点;
- 5) 将 Min_Set 和 Max_Set 加入样本集 D 中。

步骤 3 初始化选点:使用拉丁超立方采样[15]将样本集 D 补充至包含 N 个点;

步骤 4 获取样本集的高精度值;

步骤 5 使用样本集建立高精度模型; 当 $t=1$ 时高精度模型直接进入下一次循环, 否则使用步骤 8 中的训练集 T 依据公式 (5-6) 训练融合模型;

步骤 6 在融合模型上使用步骤 2;

步骤 7 挑选最终粒子群中满足下边公式的两点;

$$\mathbf{x}_p = \arg \min_{\mathbf{x} \in S \setminus D} F_D(\mathbf{x}) \quad (7)$$

$$\mathbf{x}_u = \arg \min_{\mathbf{x} \in S \setminus D} |F_{H1}(\mathbf{x}) - F_{H2}(\mathbf{x})| \quad (8)$$

其中 F_{H1}, F_{H2} 分别是当前高精度模型, 和上一次高精度模型;

步骤 8 将 $\mathbf{x}_p, \mathbf{x}_u$ 加入样本集 D 。 $k=k+1, N=N+2$;

步骤 9 由 $\mathbf{x}_p, \mathbf{x}_u$ 获得训练集 T , 具体过程如下:

- 1) 先将 $\mathbf{x}_p, \mathbf{x}_u$ 放入 T , 然后不放回地抽样 T 中的点 \mathbf{x}_t
- 2) 如果 \mathbf{x}_t 与 $\mathbf{x}_p, \mathbf{x}_u$ 的距离大于 η , 将该点放入 T ;
- 3) 重复 1), 2) 步直至 T 包含 N_T 个点, 或抽样结束。

步骤 10 返回步骤 4 直至 F_{es} 用完, 其中 F_{es} 是建模全过程所使用高精度数据个数的上线。

在本文的算法设计中, 对参数做如下设置: MG 设为 30, S 包含 $N=5d$ 个点, SG 设为 $2d$, N 的初始值与 S 相等, $\eta = m \cdot s \cdot i^2 \cdot q \cdot n \cdot r - [t]$, 其中 d 为维度, 而 T 中元素的数量保持为 D 中元素的一半。而 F_{es} 将由不同算例分别给出。TDF-PSS 使用了明显少于一般建模时使用的高精度数据点, 然后在由大量低精度数据建立的低精度模型的参与下建模, 以达到融合建模的目的。

4 实验结果

双精度数据建模的目的是减少高精度数据的使用量, 并提高建模精度, 在这一节的实验中, 均规定了高精度点的使用个数。在统一高精度点的个数的情况

下, 用测试建模的精度来验证算法的性能。首先使用标准测试函数在 10 维决策空间上进行了试验, 然后给出了算法在一个钝锥模型的升降比系数上的实验。

4.1 标准函数测试

一个包含准确 Rastrigin 函数和近似 Rastrigin 函数的双精度函数[16]被用于测试, 函数的具体形式为:

$$f(\mathbf{x}) = \sum_{i=1}^5 [x_i^2 + 1 - \cos(10\pi x_i)] \quad (9)$$

$$f_l(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^5 [x_i^2 + 1 - \cos(10\pi x_i)] \quad (10)$$

其中 $x \in \Omega = [-1, 1]^{10}$ 。TDF-PSS 可延伸出不使用低精度指导选点的融合模型算法 (TDF-SS, Two-fidelity data fusion modeling with sequential sampling) 和只用高精度建模的算法 (MHD, modeling with high-fidelity data)。实验选取这两个算法作为对比设置。根据文献[17], 本例中 F_{es} 设为 1000, 每组设置独立运行 20 次后求取平均值, 三个算法的收敛效果如图 3 所示。其中横轴代表高精度数据的选点过程, 从初始的 50 个选点直至 1000 个真实值全部用完, 而纵轴代表所建立的融合模型 (MHD 中为高精度模型) 在决策空间中的误差的积分, 计算公式如下:

$$\text{error} = \int_{\Omega} |F_H(\mathbf{x}) - f(\mathbf{x})| d\Omega \quad (11)$$

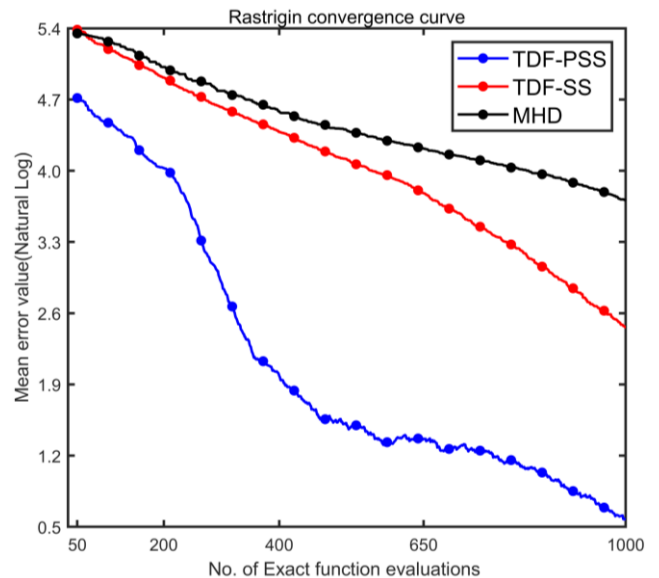


图 3 误差收敛图

图中蓝色曲线为 TDF-PSS 的收敛曲线,可以看出,使用低精度模型指导高精度建模的初始选点有效降低了初始阶段的误差,提高了初始建模的精度。在搜索的全过程中 TDF-PSS 都处于领先位置,这是由于融合建模具有对模型进行微调优化的作用。红色曲线也证明了这一点,相比于没有使用融合建模的 MHD 方法(黑色曲线),TDF-SS 具有更快的误差收敛速度。在搜索的全过程中,TDF-PSS 的误差和误差收敛速度均优于 MHD,这说明了双精度建模的有效性。在与 TDF-SS 的对比中,TDF-PSS 误差全程占优,但在搜索的一段过程中收敛速度不如 TDF-SS,一种可能的原因是在搜索的过程中,两种加点准则选取的抽样点开始集中,样本集 D 的多样性并没有明显提升。但在这一阶段过后,TDF-PSS 的收敛速度又恢复到了与 TDF-SS 相同的水平。表 1 给出了三种算法在初始阶段和最后的误差数据。

表 1 建模误差表

算法名称	不同 FEs 下的建模误差的均值		
	d	Fes=50	Fes=1000
TDF-PSS	10	1.11E+02	1.83E+00
TDF-SS	10	2.19E+02	1.23E+01
MHD	10	2.15E+02	2.15E+02

4.2 钝锥应用算例

这一小节使用 TDF-PSS 对一个实际应用算例进行建模。以飞行器试验为背景设计了一个双精度数据融合问题:为了计算某一型号钝锥的升阻比,将其放置在流场中,来流与钝锥的迎角随时间不断变化,由此产生较为复杂的升阻比系数变化过程。使用有限元求解该时变流场,首先设置较为粗糙的网格求解获得低精度数据,再设置较为细化的网格求解获得高精度数据。计算成本与网格密度成正比,所以该双精度问题确实具有低精度数据成本低,高精度数据成本高的特点。首先对时间参数进行扫描,获取全部的高低精度数据,数据直观展示在图 4 中。由于问题背景较为复杂,且只是用来提供双精度数据,因此不再详细描述。

由图 4 可以看出,高精度数据和低精度数据又类似的趋势,但仍有明显不同。在开始较为尖锐的漏斗区域,低精度数据与高精度数据的值较为接近,但在极小值处并不相等,高精度数值比低精度数值小,再之后高精度数值很快大幅超过低精度数据,具有更大的梯度,且两条曲线的极值点并不相同,且低精度数据在第 650 到第 850 个数据点之间具有较为平直的趋势。总体看来,该问题的数据有一定复杂性,其走

势对双精度融合建模具有误导性。因此,选用该组双精度问题进行测试,具有一定挑战性。

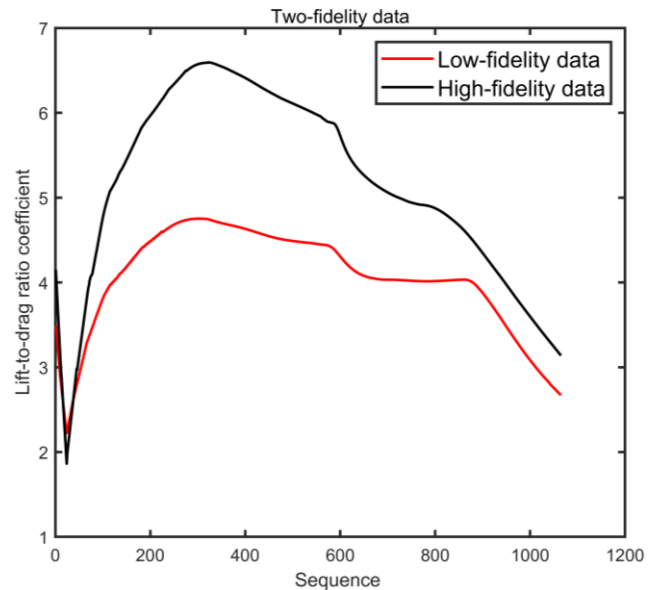


图 4 双精度数据

实验条件的设置与标准函数算例不同,在该算例中 $S=10$,并限制 $Fes=50$ 。为了方便使用离散的数据点,在算法中对每个搜索到的抽样点进行取整操作。图 5 展示了最终建立的融合模型的效果和与高精度数据的对比。

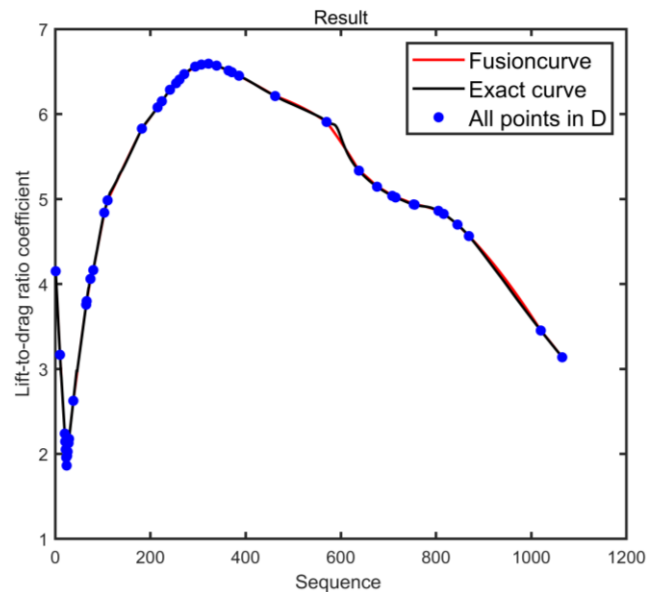


图 5 融合模型效果图

5 结论

本文首先对双精度问题的应用背景做了描述。本文

提出的 TDF-PSS 算法使用 PSO 算法对低精度模型的最值信息进行搜索, 并把这些信息加入到高精度建模的初始选点中, 这一策略有效的利用了低精度数据与高精度数据间类似的趋势。在序贯建模中采用的两个加点准则加速了建模过程的收敛。融合模型的训练过程利用低精度模型对高精度建模进行微调, 进一步挖掘了低精度数据的作用。在该算法中, 有一些参数的确定仍然需要进一步研究, 在对低精度采样的 PSO 算法中, 选择合适的参数使得算法提高对局部最优的选取能力, 而提高采样效率。这也是接下来研究需要考虑的问题之一。

参考文献

- [1] Jin Y C, W H D, Tinkle C, et al. Data-Driven Evolutionary Optimization: An Overview and Case Studies [J]. IEEE transactions on evolutionary computation, 2019, 23 (3): 442-458.
- [2] 刘深深, 陈江涛, 桂业伟, 等. 基于数据挖掘的飞行器气动布局设计知识提取[J]. 航空学报, 2021, 42 (4): 524708.
- [3] 鞠胜军, 阎超, 叶志飞. 吸气式高超声速飞行器多参数灵敏度分析 [J]. 北京航空航天大学学报, 2017, 43 (5): 927-934.
- [4] LI Y H, HUANG Y, WEI Z. A case study of application of modern design of experiment methods in high speed wind tunnel test [J]. Applied Mechanics and Materials, 2013, 444: 1229-33.
- [5] Wang H D, Jin Y C, Yang C, et al. Transfer stacking from low-to high-fidelity: A surrogate-assisted bi-fidelity evolutionary algorithm., Applied Soft Computing, 2020, 92: 106276.
- [6] MENG X H, KARNIADAKIS G E. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems [J]. Journal of Computational Physics, 2020, 401: 109020.
- [7] MENG X H, BABAE H, KARNIADAKIS G E. Multi-fidelity Bayesian neural networks: Algorithms and applications [J]. Journal of Computational Physics, 2021, 438: 110361.
- [8] C. K. Williams, C. E. Rasmussen. Gaussian Processes for Machine Learning [M]. Cambridge, MA, USA: MIT Press, 2006: 11-89.
- [9] 韩忠华. Kriging 模型及代理优化算法研究进展 [J]. 航空学报, 2016, 37 (11): 3197-3225.
- [10] 姚雯, 陈小前, 罗文彩, 等. 基于部分交叉验证的多准则序贯近似建模方法 [J]. 系统工程与电子技术, 2010, 32 (7): 1462-1467.
- [11] J Kennedy, R Eberhart. Particle swarm optimization [J]. Proceedings of IEEE International Conference on Neural Networks, 1995, 4 (1): 1942-1948.
- [12] Y Shi, R Eberhart. A modified particle swarm optimizer [J]. IEEE Congress on Evolutionary Computation, 1998, 3 (1): 69-73.
- [13] DeLOACH R. The modern design of experiments — A technical and marketing framework [R]. AIAA 2000-2691, 2000.
- [14] 海春龙, 何磊, 梅立泉, 等. 现代试验设计及其在空气动力学中的应用进展 [J]. 实验流体力学, 2022, 36 (3): 1-10.
- [15] 杜丽, 吕利叶, 孙伟, 等. 一种适用于约束空间的拉丁超立方取点策略 [J]. 机械设计与制造, 2021, 366 (8): 43-47.
- [16] Wang H D, Jin Y C, J Doherty. A generic test suite for evolutionary multifidelity optimization [J]. IEEE transactions on evolutionary computation, 2018, 22 (6): 836-850.
- [17] Li F, Cai X, Gao L, et al. A surrogate-assisted multiswarm optimization algorithm for high-dimensional computationally expensive problems [J]. IEEE Transactions on cybernetics, 51 (3): 1390-1402.