**基于模型降阶的三维模型与一维**

**模型融合建模技术研究报告**

**目 录**

[1. 引言 2](#_Toc132120812)

[2. 国内外研究的现状及发展趋势 3](#_Toc132120813)

[2.1 国外研究现状 3](#_Toc132120814)

[2.1.1 理论性研究 3](#_Toc132120815)

[2.1.2 工具及应用研究 12](#_Toc132120816)

[2.2 国内研究的现状 22](#_Toc132120817)

[2.3 发展趋势 24](#_Toc132120818)

[2.3.1 模型降阶——数字孪生的关键技术 24](#_Toc132120819)

[2.3.2 使用模型降阶技术理解Modelica中实现的模型 29](#_Toc132120820)

[3. 模型降阶理论与方法研究 33](#_Toc132120821)

[3.1 总体研究 33](#_Toc132120822)

[3.2 模型降阶方法研究 34](#_Toc132120823)

[3.2.1 基于投影的模型降阶方法 36](#_Toc132120824)

[3.2.2 基于数据拟合的模型降阶方法 43](#_Toc132120825)

[4. 总结 53](#_Toc132120826)

# 引言

现代大型数值计算体系由理论分析方法、CAE（Computer Aided Engineering）方法和实验测量三部分组成。对于理论分析方法，只有少数方程的少数工况下才存在解析解。实验测量方法对于现实情况的刻画更为精准，但是成本十分高昂。

随着计算机硬件水平的不断发展，CAE的理论及工程应用都得到了十分广泛的发展。最新流行的数字孪生的重要理论基础之一便是CAE方法。CAE方法对于复杂领域、复杂场没有很好的求解方案，三维模型计算时间长，远远达不到数字孪生要求的实时仿真的速度。基于这个问题，人们发现模型降阶可以较好的解决CAE方法在复杂场工况下无法实时仿真的问题。

对于大型动态复杂系统，在诸如控制器设计、参数优化、不确定性条件下的设计评估和对系统行为的洞察能力方面，找到一个简化模型通常是有用的。在解决控制工程问题中，模型降阶技术具有两个主要功能：在使用现代控制综合方法去解决控制设计问题方面，模型降阶具备高效性；在获得较简单的硬件或软件控制器方面，模型降阶可简化控制器阶数。此外，在研究复杂物理现象方面，动态系统的直接数字仿真是极为有效的工具，然而，当更多的细节被包括在内时，这种仿真的维数将增加到不可管理的地步，致使超出了要求的存储和计算的水平。通过模型降阶可产生一个与原系统具有相同响应特性的低阶系统，代替在大规模仿真中的原系统，或在实时应用中构建低阶控制器，从而达到解决上述问题的目的。例如，设计制造飞船、各类电子产品等复杂系统往往经济成本极为昂贵，为保证所设计的产品能够满足一定的性能指标，需要在产品投入生产之前对其进行计算模拟。由于其规模较大或复杂性较高，往往耗时过长，不利于产品的研发。

模型降阶可以通过大量三维数据（可以是CAE计算结果也可以是实验结果），采用代理模型，重新构建了一个近似的一维映射。这个一维映射可以在保证一定精度的同时，能够以接近实时仿真的速度快速给出计算结果。虽然训练的时间会很长，但是训练的过程是在事前发生的，而且一旦训练完成，在其使用范围内随时使用都可以得到几乎实时仿真的数据结果。在训练过程中，大多数降阶模型是基于数据的，因此可以用到所有数值计算领域。由于其通用性、经济性和快速响应的特点，模型降阶功能被各大商业软件重视并投入研发。

# 国内外研究的现状及发展趋势

## 国外研究现状

### 理论性研究

模型降阶技术是优化设计、优化控制和反问题应用中常见的方法。其降阶本质是将随时间变化的多维物理过程进行低维的近似描述，在捕捉系统能量的意义上达到最优化，从而达到降低计算维数、减少计算量、节省计算时间和CPU负荷的效果。系统的模型降阶问题，是对系统模型的高维性提出的。模型降阶囊括了各种不同的概念和技术，但共同的目标是一致的，即降低大系统的数学模型维数以简化控制和预估方案的设计。

模型降阶思想早在上世纪70年代就散见于一些工程界的文献中，主要用于电路系统领域及自动控制领域。几十年来，经过模型降阶领域中众多学者的共同努力，现已经成为较为系统的数学理论方法和应用工具。降阶方法主要分为两大类：时域降阶方法和频域降阶方法。时域降阶法是在状态空间上进行降阶处理，例如聚集法和摄动法等；频域降阶法是在传递函数模型上进行降阶处理，例如连分式法等。模型降阶方法众多，比较经典的降阶方法有：集结法、奇异摄动法、模态近似法、Padé逼近法、Routh逼近法、误差极小化法等。这些方法的共同特点是通过纯粹的数学计算来降阶，导致计算量很大，其中有的不能保证降阶模型的稳定性，有的不适于多输入多输出（MIMO，Multi-Input Multi-Output）系统。

**平衡截断方法**

1981年，Moore从系统内平衡实现（IBR，Internal Balance Realization）的角度出发，提出了一种渐进稳定的平衡截断模型降阶方法，成为国际上评价新的模型降阶算法的近似精度和模拟速度的标杆之一，是国际主流的模型降阶方法。通过该方法可计算出降阶系统与原始系统之间的误差关系，并且所得降阶系统能保持原始系统的稳定性、可控性及可观性。平衡截断模型降阶方法主要包括两个过程：平衡变换过程和平衡截断过程。首先，通过平衡变换矩阵将渐近稳定的最小实现系统转化为平衡系统， 然后对平衡系统进行截断得到渐近稳定的降阶系统。在这过程中，进行截断的目的在于利用Hankel 奇异值（Hankel singular values）将系统中不易观测且不易控制的状态变量去掉。但是，由于在构造平衡变换矩阵时需要求解两个大规模Lyapunov方程，因而平衡截断模型降阶方法的计算复杂度比较高且计算量比较大，无法用于大规模动力系统的降阶过程。为此，模型降阶领域中的许多学者提出了求解大型Lyapunov方程的近似方法，大大地减少了计算量，扩大了平衡截断模型降阶方法的应用范围，包括对不稳定、不可控、不可观线性定常系统，及离散、随机、双线性系统分别研究了有效的平衡截断模型降阶方法，形成了一套完整的理论。此外，平衡降阶方法与奇异摄动法、奇异值分解法等结合而形成的联合降阶法，完善了降阶方法本身的理论，使其实际应用成为可能。

**Krylov子空间方法**

模型降阶的另一类方法是基于所谓矩匹配的Krylov子空间理论的方法。该类方法是一种基于投影的模型降阶方法，利用系统的系数矩阵构造Krylov子空间的一组标准正交基，然后用该标准正交基得到的变换矩阵对原始系统进行降阶，由此得到的降阶系统的传递函数在展开点处能够匹配原始系统的传递函数一定数量的矩。在Krylov子空间模型降阶方法中，常采用Arnoldi算法和Lanczos算法来计算Krylov子空间的一组标准正交基，由此形成了Arnoldi降阶方法和Lanczos降阶方法。Arnoldi算法和Lanczos算法是科学计算领域中重要的基础理论，常用来求解大规模线性方程组及大规模特征值问题。根据所采用的变换矩阵的不同, Krylov子空间模型降阶方法又分为单侧Krylov子空间方法和双侧Krylov子空间方法。相较于单侧Krylov子空间方法，在降阶系统的阶数相等的前提下， 双侧Krylov子空间方法能匹配原始系统更多的矩，从而具有更高的降阶精度。该类方法简单易实现而且快速高效，适用于大规模系统，已经成为国际上评价新的模型降阶算法的计算速度和近似精度的标杆之一。但是，由于矩匹配具有局部插值的特点，所以在早期的研究中很难得到该方法的全局误差界。此外，该类方法无法保持原始系统的无源性及稳定性。为了解决以上问题, Bai等学者对Krylov子空间模型降阶方法进行了改进，并探讨了基于Krylov子空间的保持系统动力属性的模型降阶方法。对于二阶系统，通过定义二重Krylov子空间，提出了二重Krylov子空间模型降阶方法，保持了二阶系统的结构。由于该方法无需将二阶系统转化为一阶系统进行降阶，大大减少了计算量，提高了计算效率。目前，Krylov子空间模型降阶方法已经被应用于分数阶系统、二阶时滞系统等。

**正交多项式方法**

在工程领域中，正交多项式常用来逼近已知的函数。正交多项式模型降阶方法受到了人们的广泛关注，这其中包括Chebyshev多项式算法、Hermite多项式算法及Laguerre-SVD算法。这些现有的基于正交多项式的模型降阶方法均可归为一般正交多项式模型降阶方法的特殊情况。为了便于一般化处理，针对大规模线性输入输出系统，首次将一般正交多项式技术应用到模型降阶中，提出了时间域上的一般正交多项式模型降阶方法。首先将线性系统的状态变量及输出变量在正交多项式基底构成的空间中展开，然后利用正交多项式的基本性质求得状态变量一定数目的展开系数，进而构造出变换矩阵对原始系统进行降阶。由此得到的降阶系统不仅保持了原始系统的输出变量一定数量的多项式展开系数，而且能保持原始系统的结构性、稳定性及无源性。积分微分方程系统具有特殊的结构，在实际工程应用中，保持系统的结构具有重要的实际意义。为此，提出了积分微分方程系统的一般正交多项式模型降阶方法，由于所得降阶系统保持了原始系统系数矩阵的块对角形式，因此保持了积分微分系统的结构特征。此外，一般正交多项式技术已经被广泛用于求解耦合系统、二阶系统、双线性系统及K-power双线性系统的模型降阶问题。该类方法保持了原始系统的结构性、稳定性及无源性等动力学属性，具有计算复杂度低、计算性能好、降阶精度高、CPU 耗时少、应用范围广等优点。

**线性系统最优化方法**

线性系统的最优化模型降阶方法也是行之有效的降阶方法，该类方法的基本过程是给定一定的衡量标准，利用原始系统的传递函数求得降阶系统的传递函数，使得该传递函数在当前衡量标准下（例如, H∞范数和H2范数）是原始系统传递函数的最佳近似，然后进一步确定降阶系统的状态空间。在众多的最优化模型降阶方法中，最常见的有Hankel 最优化模型降阶方法，H∞最优化模型降阶方法及H2最优化模型降阶方法。对于线性时不变系统，H2范数是衡量原始系统与降阶系统误差的最常用范数。早期，Halevi和Wilson等人对线性时不变系统的H2最优化模型降阶问题进行了研究。特别需要指出的是，基于单输入单输出（SISO，Single-Input Single-Output）线性时不变系统的可控Gram矩阵和可观Gram矩阵，Wilson提出了H2最优一阶必要条件，即Wilson’s条件。Dooren等人利用矩阵的迹的性质将Wilson’s条件推广到了多输入多输出情形。此外，通过分析线性时不变系统的传递函数，也可从插值的角度对该优化问题进行求解。但是由于线性时不变系统的H2最优化模型降阶问题是非凸的，我们很难求得整体最小值解。考虑到由该优化问题的局部最优解构造的降阶系统具有较高的近似精度，因此现有的许多方法都基于H2最优一阶必要条件来求解局部最小值解。

不同于以上求解方法, Lam等人将流形（manifold）应用到模型降阶中，充分运用黎曼几何的相关知识求解了线性时不变系统的H2最优化模型降阶问题。该方法基于单侧变换矩阵，在流形上求解H2最优化模型降阶问题，保证了算法的严格收敛性也确保了整体解的存在性。流形是控制领域中重要的理论工具，在流形上研究光滑代价函数的优化问题引起了人们的热切关注，产生了许多丰富的理论成果。近年来，黎曼优化技术也被应用到不同的领域中。随着黎曼优化技术的不断发展，基于黎曼优化的模型降阶方法受到了研究者们的关注。但是基于黎曼流形的模型降阶方法的相关研究较少，仍待进一步完善。另外，也产生了频域加权意义下的Hankel最优化模型降阶方法、频域加权H∞最优化模型降阶方法和频域加权H2最优化模型降阶方法。

工程领域中，参数线性系统也是一种非常常见的系统。在参数空间中，参数线性系统的系数矩阵会随着参数的变化而变化。对于参数线性系统，有理插值模型降阶方法是一类应用较为普遍的参数模型降阶方法，其中比较常见的方法有矩匹配方法、切线插值方法等。由矩匹配方法得到的降阶系统在插值点处具有较好的近似精度，但是在远离插值点处近似精度较低。为解决此问题，可以在多个插值点处进行矩匹配，即多点矩匹配方法。对于参数线性系统，切线插值方法的目的在于在给定的插值点处，构造参数降阶系统的传递函数使其沿着适当的左插值方向（或右插值方向或双向插值方向） 满足左切线插值条件（或右切线插值条件或双向插值条件）。注意到，相较于线性时不变系统的切线插值方法，由参数线性系统的双向切线插值方法得到的参数降阶系统，其传递函数不仅是原始参数系统传递函数的双向插值函数，同时它们能够匹配关于参数变量的一阶梯度。由上述几种方法，在参数样本点处可以得到参数线性系统的局部降阶系统或构成这些局部降阶系统的局部基底，通过构造“局部”变换矩阵来得到参数降阶系统。 在参数空间中新的参数点处对局部基底矩阵直接进行插值得到“局部”变换矩阵，这是构造“局部”变换矩阵的一种方法。然而，由于这些局部基底矩阵所属的空间并不是向量空间，所以直接对其进行插值得到的局部变换矩阵可能不属于这个空间。考虑到Grassmann流形的切空间是向量空间，可将各个参数样本点处的局部基底矩阵映射到该流形的切空间中，进而在流形的切空间中进行插值，最后通过逆映射得到“局部”变换矩阵。该方法充分运用了Grassmann流形的测地线、切空间等几何性质，已经被应用于空气动力学、弹性动力学等领域。另一种基于流形的模型降阶方法是对各个参数样本点处局部降阶系统的系数矩阵进行插值。特别地，可采用合同变换将局部降阶系统的系数矩阵转化为广义坐标系下的局部降阶模型，然后在此广义坐标系中进行插值，由此得到在新的参数点处的降阶系统。

**非线性系统降阶方法**

由于非线性系统自身结构较为复杂，因此其模型降阶方法发展得相对缓慢，非线性模型降阶理论仍待进一步完善。可以采用本征正交分解方法（POD，Proper Orthongonal Decomposition）直接对非线性系统进行降阶，由此得到的降阶系统保持了原始系统的非线性结构，但是计算量比较大。充分运用经验插值方法（EIM，Empirical Interpolation Method）、离散经验插值方法（DEIM，Discrete EIM）等可以提高计算精度减少计算量。据我们所知，处理非线性系统模型降阶问题的其它方法主要有三种：一种是近似化方法，一种是双线性化方法，另外一种是将非线性系统转化为等价的二次—线性微分代数（QLDAE，Quadratic-Linear Differential Algebra Equation）系统。采用近似化方法处理非线性系统中的非线性项时，通常将非线性项进行Taylor展开只保留线性项或二次项，然后利用现有的模型降阶方法对该近似系统进行降阶，从而得到线性化模型降阶方法或者二次化模型降阶方法。以上方法产生的近似系统在局部近似精度较高，但是整体近似效果不好；由于Taylor近似的局部性，以上方法仅仅适用于弱非线性系统或者具有“small”输入的系统。此外，采用Taylor展开可能导致系统稳定性及无源性丢失。White 和Rewienski等提出的分段线性化（TPWL，trajectory piece-wise linearization）方法有效地消除了“弱非线性”条件的约束，并且提高了整体近似精度。进一步，将该方法与一般正交多项式技术相结合，产生了基于分段线性化技术的一般正交多项式模型降阶方法。此外，一些学者也研究了二次化及更高次情形下相应的模型降阶方法。只是，采用二次化近似技术对非线性系统进行处理时往往需要涉及到张量理论。

在许多实际应用中，线性系统无法充分描述某些物理过程的动态行为，然而由于非线性系统结构复杂，对其进行分析或者合成它们的控制器往往需要花费昂贵的计算代价。双线性系统是介于线性系统及一般非线性系统之间的一类特殊系统，因其具有丰富的理论基础而备受青睐。通过双线性化过程，一些非线性系统可近似为双线性系统，相应地产生了一系列模型降阶方法。早期，Hsu及Al-Baiyat等人研究了基于平衡的双线性模型降阶方法。2001年，基于Volterra理论，Lam和Zhang首次定义了双线性时不变系统的H2范数。利用双线性系统的可控Gram矩阵和可观Gram矩阵给出了计算系统H2范数的具体表达式，从而将线性时不变系统的相关理论推广到了双线性情形。Benner等从插值的角度进一步求解了双线性时不变系统的H2最优化模型降阶问题。详细分析了基于矩阵方程的H2最优一阶必要条件，并提出了相应的迭代降阶算法。理论分析表明，由此得到的降阶系统满足所导出的一阶必要条件。鉴于该迭代算法需要求解广义的Sylvester方程，Benner等进一步提出了基于有理插值的模型降阶方法，从而将线性时不变系统的有理插值方法推广到了双线性时不变系统中。相关文献也研究了截断情形下的双线性时不变系统的H2最优化模型降阶问题，以及双线性时不变系统的平衡截断模型降阶方法和Krylov子空间方法等。

采用以上方法对非线性系统进行降阶时，大都需要对非线性函数进行近似处理，得到近似系统或者双线性系统。在不涉及近似过程的前提下，Gu在2011年将某些非线性系统转化为了等价的QLDAE系统。例如，我们常见的Burgers方程和RC电路系统，通过双线性化过程可将其近似为双线性系统进行处理，然而利用Gu提出的方法得到的则是它的等价系统。该方法主要通过增加新的状态变量，在此基础上引入关于新变量的代数方程或者通过李导数（Lie derivate）引入新的微分方程，由此得到QLDAE系统。该系统的优点在于：它消除了“弱”非线性约束条件，不仅适用于弱非线性系统也适用于强非线性系统；另外，该系统的最高次项是二次的，不含三次及其以上高阶项，但是却保留了原始非线性系统的“强”非线性函数。自该系统被提出以来，便受到了研究者们的广泛关注并产生了许多行之有效的方法。

#### 模型降阶方法分类

关于模型降阶方法大致可以分为两大类：(1)保留系统中的主要状态，例如Aggregation方法，Model方法，Lyapunov函数方法和Perturbation方法。(2)对输入-输出数据的辨识，这些数据是通过用一特定的输入信号驱动原系统而获得的。

在模型降阶中．一般要重点考虑降阶系统逼近原始系统的程度如何，通常，可以分时间域和频率域两种情形讨论降阶系统与原始系统之间的误差。基于系统模型描述所用的域，模型降阶方法可归结为两大类：当把模型降阶法应用于系统的状态空间模型时，称为“时域降阶法”，简称，“时域法”；而当应用于系统的传递函数模型时，称为“频域降阶法”，简称。“频域法”。与时域法比较，频域法更趋成熟。在时间域上，可以直接比较降阶系统与原始系统的输出函数在某种范数下的大小，这种刻画误差的方式较为麻烦，需要用到丰富的微分方程理论，甚至还需要定义一些特殊的范数。在频率域上，需要对原始系统和降阶系统的状态方程和输出方程做Laplace变换以获得各自的系统传递函数，然后通过比较传递函数的矩来衡量降阶系统逼近原始系统的程度。为衡量降阶系统与原始系统之间的近似程度，可以比较相应的传递函数。具体地，可以比较降阶系统与原始系统传递函数相同阶的矩。这种误差刻画方式由于可借助的数学工具较多，其理论研究结果往往很丰富也很深刻。但无论是从时间域还是频率域出发，由于模型降阶研究的对象是系统的整体近似问题。不是以往单纯的函数计算或方程的求解等，这当中需要更多和更深奥的数学理论是自然的，其结果是相关的研究工作无疑会大大丰富和扩充现有的数学理论。

目前所提到的模型降阶方法大都基于线性系统，结合本文所提到的各种方法，对模型降阶方法的大致分类和特性总结。给出如下图表：

**表 2‑1 模型降阶方法分类与特性表**

|  |  |  |  |
| --- | --- | --- | --- |
| 模型降 特性  阶方法  分类 | 稳定性 | 精确度 | 简易性 |
| 频域降阶方法 | Routh近似法、Hurwitz近似法、Schwarz近似法、稳定方程法、主导极点法、FF-pade近似法、连分式法、积分全等变换和最优化模型降阶方法、Arnoldi和Lanczos模型降阶方法、Krylov子空间模型降阶方法、平衡截断模型降阶方法 | Pade近似法、Routh近似法、Hurwitz近似法、Arnoldi和Lanczos模型降阶方法、积分全等变换和最优化模型降阶方法、Krylov子空间模型降阶方法、渐近波形估计模型降阶方法 | Pade近似法、Routh近似法、Hurwitz近似法、Schwarz近似法、FF-pade近似法、Arnoldi和Lanczos模型降阶方法、Krylov子空间模型降阶方法 |
| 适用于SISO和MIMO | Routh近似法、Hurwitz近似法、Schwarz近似法、稳定方程法、主导极点法、FF-pade近似法（仅适用于SISO系统）、连分式法、Krylov子空间模型降阶方法、平衡截断模型降阶方法、积分全等变换和最优化模型降阶方法、Arnoldi和Lanczos模型降阶方法 | Pade近似法、Routh近似法、Hurwitz近似法、Krylov子空间模型降阶方法、积分全等变换和最优化模型降阶方法、Arnoldi和Lanczos模型降阶方法、渐近波形估计模型降阶方法、多点拟合模型降阶方法 | Pade近似法、Routh近似法、Hurwitz近似法、Schwarz近似法、FF-pade近似法（仅适用于SISO）、Krylov子空间模型降阶方法、Arnoldi和Lanczos模型降阶方法 |

#### 多种模型降阶方法对比研究

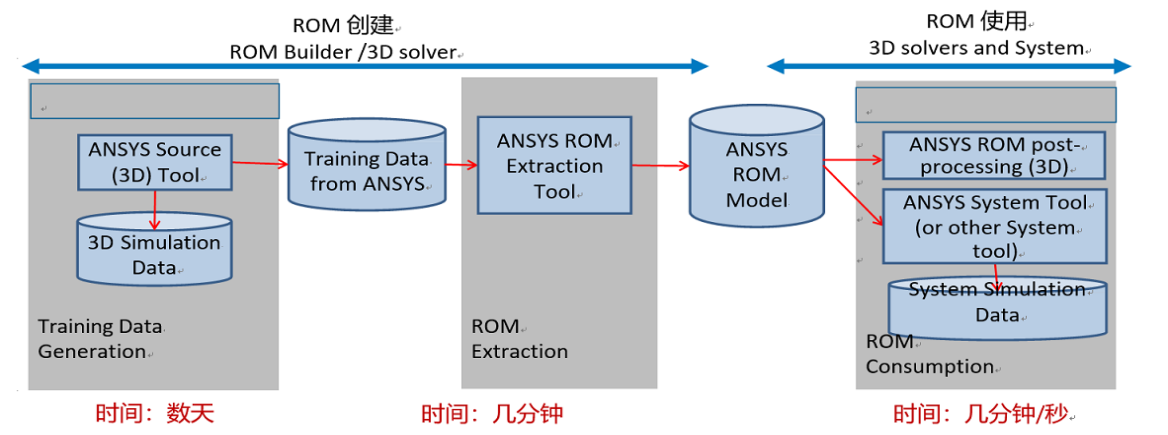
模型降阶方法目前已广泛应用于各种工程设计中，传统的模型降阶方法主要有基于奇异值分解的平衡法和基于Krylov子空间理论的矩匹配法，它们在理论上各有优缺点，如平衡法，它的优点在于能够保持降阶模型的结构特性，可根据系统奇异值的大小进行截断降阶，并能提供降阶误差界。但此类方法也有其缺点，对于大型的动态系统，由于在降阶过程中要求解两个高阶次的Lyapunov方程，因而其计算量相当大，对于计算机的存储容量和计算时间耗费较大，这使得平衡法难以应用于现代大规模系统的降阶。模型降阶的另一类方法是基于所谓矩匹配的Krylov子空间理论的方法。此类方法的优点在对大型动态系统的降阶过程中，其计算 量小、速度快，能够匹配原系统传递函数在某点处的若干矩。但此类方法的缺点也是显而易见的，那就是它不能保证降阶系统的稳定性，也很难得到降阶系统与原系统之间的误差范围。

目前，在模型降阶领域的一个新的研究方向是将上述两类方法结合起来应用于工程系统中。这便是基于SVD—Krylov模型降阶理论,此方法一方面由于降阶过程中涉及了系统的Gramian矩阵，因此它反映了平衡法的降阶特性。另一方面，由于Krylov子空间引入，因此它又使得降阶过程计算量减小。总体言之，基于SVD—Krylov的模型降阶方法既保持了原系统的结构特性，又降低了计算开销，所得降阶系统传递函数能与原系统传递函数在某点处的若干矩相匹配。基于SVD—Krylov的模型降阶方法虽然较传统的方法有了很大的进步，能匹配原系统若干矩，但它只是单纯的匹配，将那些未能匹配的矩进行了单纯地舍弃处理，这就使得降阶误差不能进一步地缩小。为了使降阶误差能够进一步减小，在SVD—Krylov降阶方法基础之上提出的最小二乘模型降阶正是基于此目的的。它不仅能匹配原系统的若干矩，也能将那些未能匹配的矩以最小二乘的方式进行逼近，使降阶误差得到有效的降低。

最小二乘模型降阶法是基于SVD—Krylov降阶算法的，事实上它也是一种投影法，依赖Krylov子空间和系统的Gramian矩阵。对于离散型系统，该方法可直接应用于原系统，而对连续系统，则要进行双线性变换使之成为离散系统后再降阶，最后再对所得降阶系统进行逆双线性变换得到连续型降阶系统。

### 工具及应用研究

构造Reduced Order Model (ROM)的目标主要有两点：一是以远少于原数值模型的阶数和计算耗费提供系统主要时空特征较精确的数学描述；二是为研究者解释系统特征提供解释性工具。ROM是由数值模拟模型(通常在几十万、几百万阶或更高)近似投影获得的低阶数学模型。它能以相对较少的自由度(通常在几十阶或几百阶)来描述原系统的主要系统特性，在保留全阶高精度CFD模型的可信度和高保真度的同时，计算量又不太大(几乎可以实时获得结果)，且能够方便地与其他物理场模型进行耦合用于多学科耦合分析与系统优化设计。通过ROM就在单点仿真的高可信度CFD数值模型和复杂多学科耦合系统仿真与设计之间架起了一座桥梁。ROM的思想提出后迅速得到了学术界和工程界的广泛关注。ROM可以分为两个步骤：ROM创建与ROM使用。其主要步骤如下图所示:



**图2‑1 ROM创建与使用示意图**

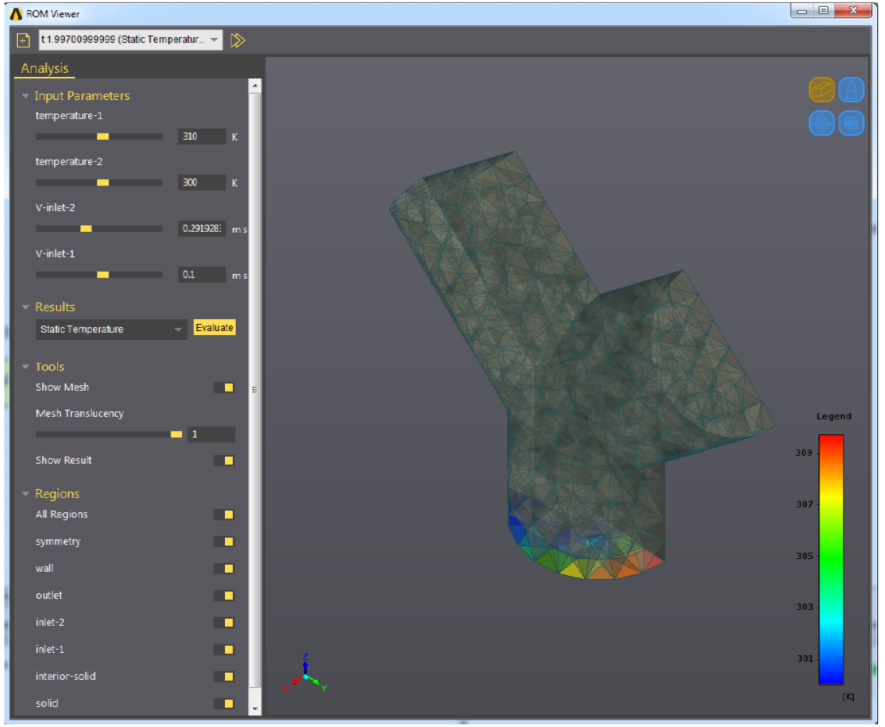
目前市场上主要的模型降阶的大型商业软件有ANSYS、AMESim、Matlab等。

**工具对比**

**组件配置**

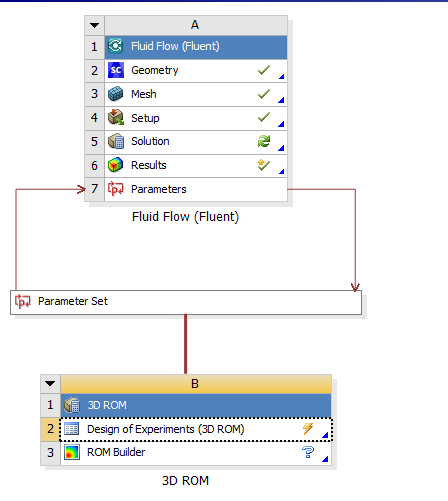
1. ANSYS

ANSYS的模型降阶功能内置在其系统仿真软件Twin Builder内，并设有专门的3D ROM后处理技术。其界面上可以改变参数，并可以实时显示计算结果。结果表示形式上，其可以显示三维场的渲染结果以及网格。



**图 2-2 ROM Viwer界面**

同时，ANSYS 的Fluent也可以生成 ROM模型，并显示通过ROM计算的三维仿真结果。如下图所示：



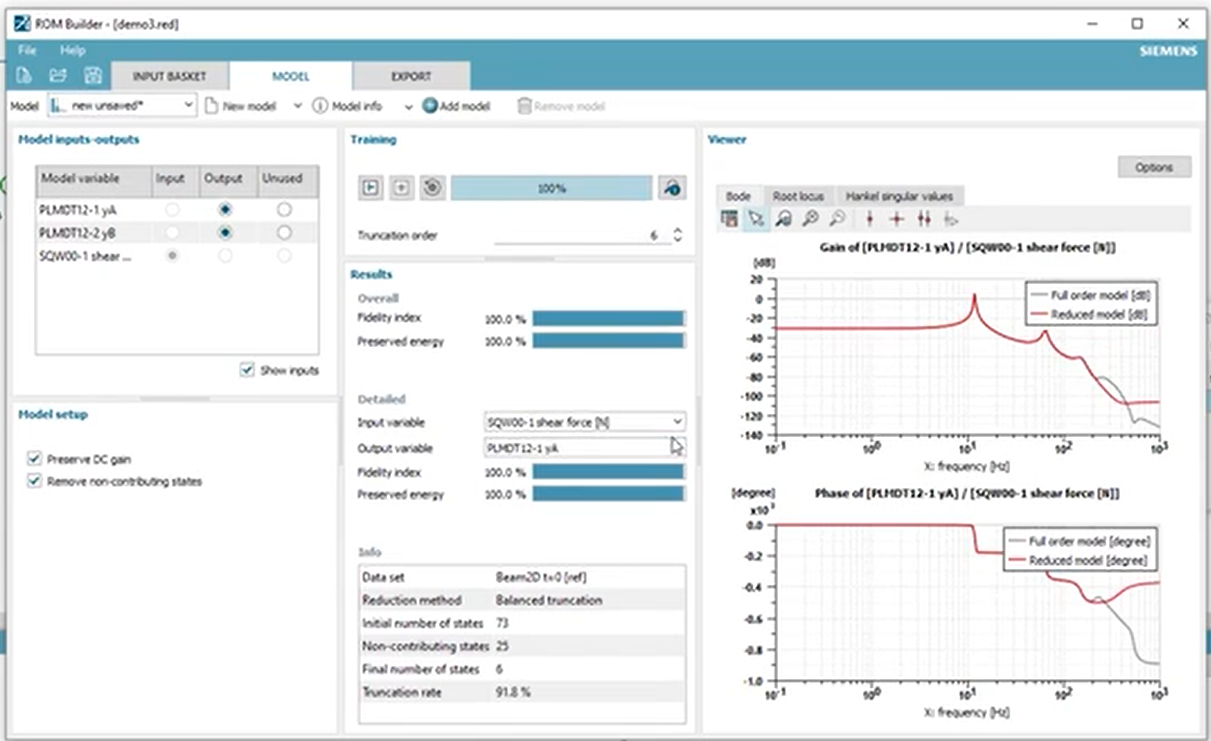
**图 2-3 ANSYS ROM模型流程**

ANSYS Workbench内含有Parameter Set组件和3D ROM组件。Parameter Set可以对求解过程中设置的输入参数进行样本抽样设计，并将数据传递给求解器（如：Fluent），求解器求解样本集中的所有工况，然后将所有计算数据传递给3D ROM模块。3D ROM 模块主要功能是针对数据结果进行ROM构建。构建完的结果可以用fluent打开，也可以用ROM Viewer打开。

同时，对于生成的ROM，可以放置到Twin Builder作为一个组件进行实时仿真。

1. AMESim

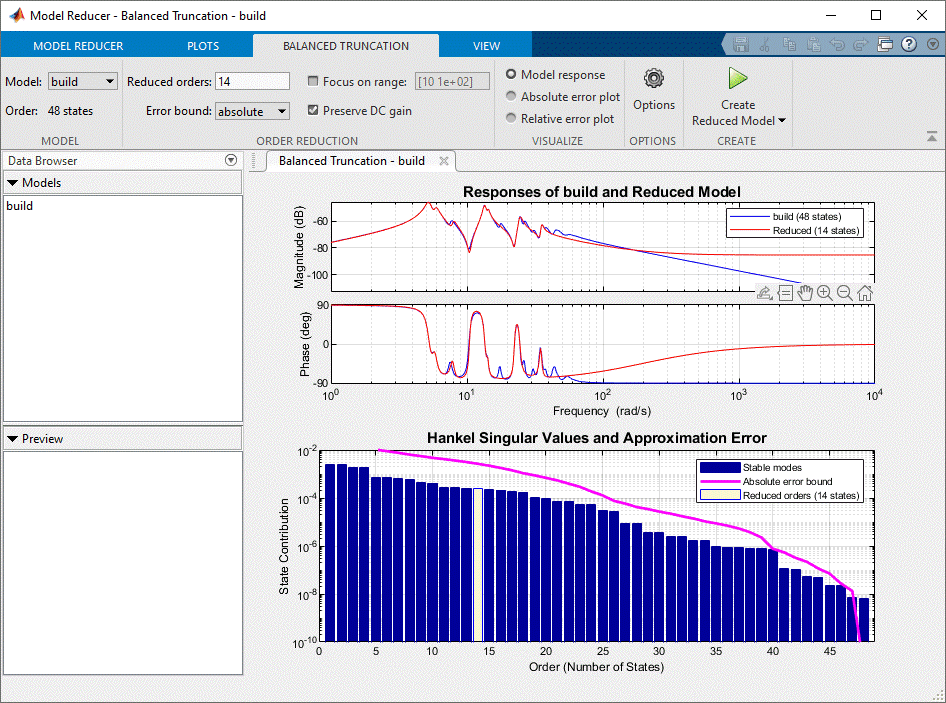
AMESim的模型降阶功能全部集中在AMESim ROM Builder软件上。AMESim ROM Builder内置数据导入、模型建立、模型导出功能。AMESim ROM Builder建立ROM后，可导出到Simcenter AMESim软件中进行系统仿真。



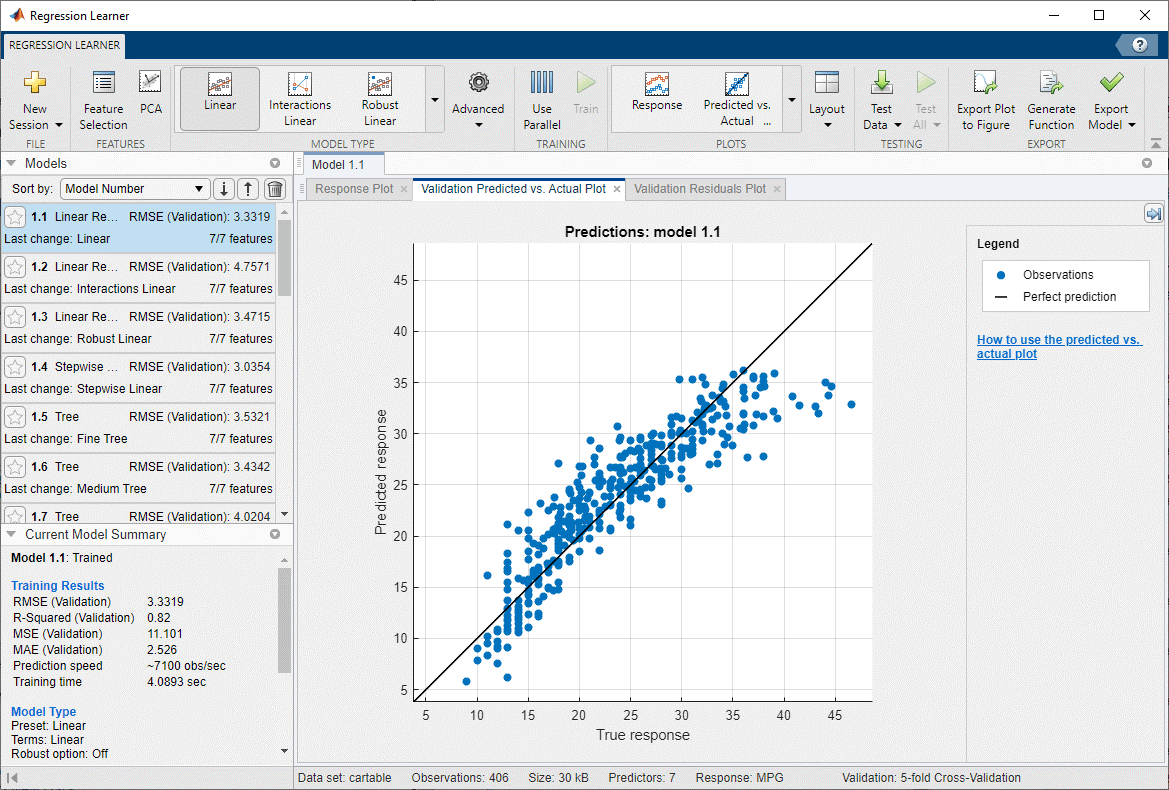
**图 2-4 ROM Builder界面**

1. Matlab

Matlab的模型降阶功能集中在Statistics and Machine Learning Toolbox工具箱内。其中内置了大量从试验样本设计、降维处理、数据还原、系统仿真等模型降阶段各个步骤的算法函数。并含有Moder Reducer模型降阶APP，以及Regression Learner 回归器APP。



**图 2-5 Model Reducer界面**



**图 2-6 Regression Learner界面**

**类型与方法**

本节主要介绍ANSYS、AMESim、Matlab的降阶模型的类型以及采用的降阶段模型方法。

1. ANSYS

ANSYS Twin Builder内置了静态与瞬态降阶模型。

其中静态降阶模型采用SVD分解和遗传聚合算法。它是由以下插值方法建立的模型: 多项式回归，克里格，支持向量回归，移动最小二乘。这个通用算法自动返回基于几个互补方法的几个插值的最佳组合。

对于瞬态降阶模型，ANSYS从2019至2020R2的版本一直在迭代降阶方法。采用的方法有一阶显式向量拟合、高阶显式向量拟合、机器学习等。

1. AMESim

AMESim ROM Builder中根据方程特点，分为静态样本降阶模型、瞬态降阶模型以及状态空间降阶模型。

对于静态样本降阶模型与瞬态降阶模型，AMESim ROM Builder均采用响应面模型以及机器学习方法进行模型降阶段。而对于状态空间降阶模型，则采用平衡截断法与Krylov子空间法。

1. Matlab

Matlab中模型降阶方法可支持静态降阶模型、瞬态降阶模型以及状态空间降阶模型。

Matlab中可采用的降阶方法十分广泛，但是集成至Model Reducer中的降阶方法不多，主要为PCA、SVD分解、傅里叶分析。

**优点缺总结**

对于ANSYS而言，其最大的优势在于内置了众多CAE求解器，因此可以将DOE模块直接附加到求解器中，然后让求解器按照DOE设计的样本批量计算模型降阶所需数据。ANSYS的模型降阶是过程最为完整的，所有步骤都可以在ANSYS中完成。其强大的后处理显示功能也是ANSYS独有的优势。

Amesim对比ANSYS而言，最大的优势在于状态空间降阶模型以及良好的神经网络设计界面。用户可以很好的去使用神经网络模型去调试模型。Amesim另一大优势在于Amesim中包含了众多系统仿真的模型库积累，ROM加入到多领域联合仿真更为方便。

相比前两个，Matlab更像为专业人员提供的数据处理工具箱，用户可以用工具箱的函数搭建各式各样的降阶模型。但是，这点对于用户的专业水平要求过高，没有搭建的比较好的降阶模型APP供用户使用。Matlab的另一个优势在于其拥有三者中最为强大的机器学习工具箱，可供专业机器学习用户调试降阶模型。

三者的功能对比总结如下表所示：

**表 2‑2 模型降阶方法分类与特性表**

|  |  |  |  |
| --- | --- | --- | --- |
| 功能模块 | ANSYS | AMESim | MATLAB |
| 求解器 | 热、流、结构、电磁以及Twin Builder系统仿真各领域求解器 | 系统仿真求解器 | Simulink系统仿真求解器 |
| 试验设计 | Parameter Set，可与CAE求解器相关联 | 有DOE模块，但模型降阶模块不包含样本设计。 | 有DOE模块，但模型降阶模块不包含样本设计。 |
| 支持的模型降阶类型及所用方法 | 静态ROM：SVD分解和遗传聚合算法  瞬态ROM：一阶显式向量拟合、高阶显式向量拟合、机器学习等 | 静态ROM：响应面模型以及机器学习方法  瞬态ROM：响应面模型以及机器学习方法  状态空间ROM：平衡截断法与Krylov子空间法 | 静态ROM：PCA、SVD分解、机器学习（未集成至Model Reducer）  瞬态ROM：PCA、SVD分解、傅里叶分析、机器学习（未集成至Model Reducer）  状态空间ROM：平衡截断法 |
| ROM使用 | 1.ROM Viewer中使用  2.Fluent中使用  3.集成到Twin Builder进行系统仿真 | 集成到AMESim中进行系统仿真 | 集成到Simulink进行系统仿真 |
| 结果显示 | 1、ROM Viewer支持三维显示  2、Twin Builder支持图表显示 | 只支持图表显示，无三维显示功能 | 只支持图表显示，无三维显示功能 |

**典型应用**

1. 反应堆堆芯

随着计算机软、硬件技术的巨大进步，计算流体动力学（Computational fluid dynamics, CFD）已在核工程中的流动和传热问题研究中获得了广泛的应用。但由于需要采用精细的计算网格，CFD模拟通常面临计算量大、计算效率低等瓶颈。早在2012年，美国的CASL项目即尝试耦合商业CFD软件STAR-CCM+和中子学软件，对1/4堆芯的流动和传热行为进行了模拟。由于计算规模庞大，整个计算共花费了10000 CPU核时，占用内存1 TB。近期在利用CFD方法研究反应堆压力容器承压热冲击现象时，300s的瞬态过程模拟共耗时200小时（计算硬件：64核、64GB内存；计算软件：STAR-CCM+）。显然，利用这些模型开展数值实验或者进行优化设计在工程上基本不具备可行性。尽管近年来计算能力获得了巨大的进步，但受限于计算节点之间的通讯开销，仅通过并行计算技术也难以获得显著的加速效果。

开展数值实验是建立数值模型的主要目的。在进行参数化数值试验时，待研究问题的控制方程不变，但其物理参数（如粘度系数）、边界条件（如入口速度和温度）甚至求解区域（如注水口方向）等却不断变化。若将这些变化的系数或条件视为参数，则可用参数化的控制方程描述这类问题。加速求解参数化问题的最直接手段是应用各种降阶法（model order reduction, MOR）降低模型的阶次，大幅度减少需要求解的未知量个数。模型降阶的基本思想是将包含多个自由度的全阶模型（Full order model, FOM）投影到低阶子空间，将复杂的全阶模型转化为仅包含少量自由度的降阶模型（Reduced order model, ROM）。对线性模型进行降阶时，应用最为广泛的是缩减基有限元方法。但利用缩减基有限元方法对Navier-Stokes（N-S）方程描述的非线性模型进行降阶时，输运项和缩减基函数的内积会导致三线性（Trilinear）项的出现，这给推导模型降阶方法和编制计算程序带来了较大的困难。和有限元方法不同，有限体积法在控制体积内直接对控制方程进行积分，不需要定义测试函数以及局部的稳定项。因此，基于有限体积法的模型降阶方法更适合流动问题的求解。

本征正交分解(Proper orthogonal decomposition, POD)和贪婪算法（Greedy algorithm，GA）是构建低阶子空间时最常用的两类方法。前者直接对典型的解样本进行分析，而后者需要进行后验误差估算。对非线性方程构造严格的后验误差估算理论具有相当的挑战性，因此POD方法在构建复杂流动问题的降阶模型时获得了广泛的应用。POD方法的基本原理是利用奇异值分解（Singular value decomposition，SVD）从不同时刻、不同参数的诸多解瞬像(Solution snapshots)中筛选出最具代表性的模式，然后通过Gram-Schmidt正交化过程构造作用在整个求解域上的基函数，最后将近似解分解为全局作用的基函数的线性组合,通过采用全局作用的基函数，模型的自由度大大降低。

利用POD方法构造降阶模型大致可分为非侵入式（Non-intrusive）和侵入式(Intrusive)两种方法。非侵入式的降阶模型不依赖于控制方程，本质上是在输入数据和输出数据之间建立一个代理模型（Surrogate model），利用代理模型获得新参数对应的分解系数。径向基函数（Radial basis function，RBF）和Kriging方法是常见的代理模型构建方法。非侵入式的降阶模型虽然构造简单、计算速度快，但其本质是对分解系数进行拟合，求解瞬态问题时会产生较大的误差。侵入式的降阶模型仍需求解控制方程，最常见的做法是将控制方程投影到POD基张成的子空间上。相比非侵入式降阶模型，侵入式降阶模型仍保持原问题的物理特性，因此其数学形式更加严谨，泛化能力更强。在有限体积求解框架内利用POD-Galerkin投影构建流动问题的降阶模型已成为近几年的研究热点。

1. 高超声速飞行器

美国俄亥俄州立大学的David O. Sigthorsson与A. Serrani根据空军实验室所建立典型高超声速飞行器纵向弹性模型first-principle model。其纵向动力学的稳定轴运动方程是通过拉格朗日方程和可压缩流理论推导得到的。该模型是包含11个状态的微分方程，高阶数的特性使得无论是模型分析还是控制器设计都变得十分困难。

以往的线性模型降阶方法只能应用于稳定的系统，然而，高超声速飞行器具有静不稳定特性，无法保证模型运行于有效工作范围，这给模型降阶带来了极大的困难。因此，需要对模型进行处理，保证模型状态信号运行至有效范围。针对该问题，本文采用系统的右互质分解方法。系统的右互质分解方法可以将不稳定的系统表示成两个稳定系统的组合形式，为模型降阶方法应用于不稳定模型降阶提供了条件。

通过平衡截断模型降阶方法，首先通过计算系统的能控和能观的格莱姆矩阵，得到系统的Hankel奇异值，确定对系统输入输出影响较小的状态。然后构造平衡变换矩阵，并对系统进行平衡变换，使系统的能控和能观的格莱姆矩阵相等，都等于由系统奇异值构成的对角阵，处于平衡状态。然后对系统进行截断处理，将对系统输入输出影响较小的状态截断，得到降阶模型。最后，利用灰色关联一致性检验方法对降阶模型进行了模型验证。

1. 飞机数字孪生体

2011年，美同空军研究实验室提出将数字孪生技术应用于飞机结构的寿命管理，产生了飞机数字孪生体(Airframe Digital Twin)的概念，以解决未来复杂服役环境下的飞机运行维护的问题。其提出，飞机的数字孪生体应具有高保真性，包含实际飞机制造过程的公差和材料特性；借助高性能计算，数字孪生机体能在飞机研制阶段进行大量虚拟飞行，发现非预期失效模式以修正设计；通过在实际飞机上布置传感器，可采集飞机飞行过程中的多种数据(六自由度加速度、表面压力等)，并输入数字孪生机体中更新其结构响应，进而预测机体的实际寿命消耗。ZONA公司正基于Volterra级数方法研究一种降阶模型，用于预测机体所受的气动载荷和应力。将降阶模型集成到结构寿命预测模型中，能够进行高保真应力历史预测、结构可靠性分析和结构寿命监测，以提升飞机机体的运维管理。上述技术突破后，就能形成初始（低保真度）的数字孪生体。加拿大皇家空军也开始评估机体数字孪生在机群管理中的应用，并以一架CF118的全尺寸部件测试进行了验证。

## 国内研究的现状

#### 理论、方法研究

近些年来，随着模型与数据驱动的数字化工程的发展，国内对大系统的模型降阶研究也日益重视，仿真工作者在模型降阶方面进行了许多有益的探索，基于各领域和各系统的方法研究与应用层出不穷，其中包括非定常流场降阶模型及其应用的研究、跨音速气动弹性问题的降阶气动模型构建、车辆侧倾动力学模型降阶方法研究、电力系统模型降阶及最优控制等。

西安交通大学的蒋耀林教授在其著作《模型降阶方法》中详细介绍了模型降阶方法的理论和应用，并将模型降阶推广到大型系统近似过程中；南京航空航天大学与中国运载火箭技术研究院对大长细比导弹的气动弹性模型进行了可信性和适用性研究；天津大学结合美国空军实验室提出的弹性纵向模型进行了高超声速飞行器模型降阶研究，证实了该方法的有效性；北京理工大学基于代理模型研究高超声速气动热模型降阶，提高了降阶模型的精度；国防科技大学和上海微小卫星工程中心运用模型降阶对柔性航天器动力学建模进行了研究；北京空天技术研究所公开发表了一种适用于飞行器的模型降阶分析方法，建立飞行器全量耦合动力学模型；对所述全量耦合动力学模型的相关数据进行解析拟合，得到与相关数据对应的解析拟合表达式；在飞行器平衡点处对非线性动力学模型进行线性化，得到面向控制的线性动力学模型；对所述面向控制的线性动力学模型进行右互质分解得到分解后的状态空间模型；使用基于兰克泽斯过程的模型降阶方法对所述分解后的状态空间模型进行降阶，得到降阶后的状态空间模型。北京航空航天大学公开发表了一种基于结构降阶模型的非线性气动弹性动响应分析方法，采用对输入测试载荷及其对应结构变形进行回归分析的方法求解给定形式的结构非线性刚度系数，利用非线性刚度系数构建大变形结构动力学方程，在此基础上利用模态组合乘比例系数的方式构建测试载荷，通过构建位移残量基函数准确恢复机翼展向位移；最后结合结构降阶模型给出的动力学方程及非定常曲面涡格法搭建几何非线性气动弹性阵风响应求解流程。另外，还有很多相关单位、组织也正在积极参加模型降阶技术研究，并在各方面取得了很多成果。

#### 工具及应用研究

进入21世纪以来，数字化技术成为国家战略的重要一环，建模技术从传统的由一维系统建模、三维有限元建模发展为一三维融合建模。一三维融合建模综合了一维模型的效率优势和三维模型的精度优势，核心在于实现仿真精度与计算速度之间的平衡。目前，国内对于模型降阶技术仍在持续研究中，尚未形成成熟的商用模型降阶工具。

1. 中山大学禹力软件降阶模型构造器ylROMBuilder

禹力软件ylROMBuilder工具主要包含POD投影、缩减基有限元、(D)EIM、HyperReduction等基于方程的降阶方法，并将其应用于固体力学、流动传热、中子输运等领域。该工具支持线性和非线性模型的降阶，对三维有限元计算效率有显著提高。目前该工具还处于科研阶段，与国际领先的商业软件还存在较大差距。

2. 苏州同元软控MWORKS.Sysplorer模型降阶工具箱

同元软控Sysplorer在物理建模方面积累了充分的技术能力，拥有自主开发的一维系统仿真编译器和求解器，参与发展维护了“面向对象的系统建模统一语言Modelica国际规范”和“功能样机接口FMI国际规范”，使一维模型和三维模型可在同一平台MWORKS.Sysplorer上联合仿真，确保一三维模型融合的无缝衔接。基于MWORKS SDK的模型降阶工具箱，选择先跳过传统的基于控制方程的模型降阶方法，开展机器学习方法研究，结合基于数据的奇异值分解、主成分分析等模型降阶方法，自主研发了基于神经网络、响应面等代理模型的模型降阶工具，具备适用于数据驱动的多领域多系统虚拟实验系统降阶模型构建能力。针对特定领域的模型降阶工具开发，未来将开展基于数据和方程联合的模型降阶方法研究，提高降阶模型的置信度和适配度。

## 发展趋势

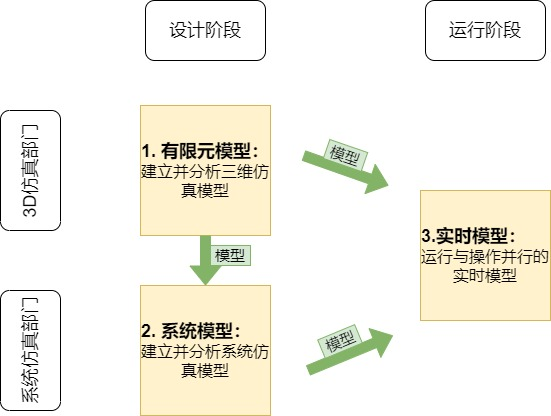
### 模型降阶——数字孪生的关键技术

越来越多具有高度经济和社会影响的颠覆性创新塑造了我们的数字化世界。这些开发的速度和扩展范围受到可用工具和范例的限制。仿真技术是实现数字化的关键技术，通过数字孪生，可以将产品和系统的镜像映射在数字世界。数字孪生需要一种范式转变，工程和操作需要的不是以各专业为中心的工具，而是通过后台模拟与物理和数字环境不断交互的自主辅助系统。模型降阶是将高度详细和复杂的仿真模型转移到其他领域和生命周期阶段的关键技术，减少自由度，即提高模型执行速度，同时保持所需的准确性和可预测性。

我们正处在产品和复杂系统的工程和操作的转折点。日益增长的复杂性（复杂系统、异构宏观系统或新型制造技术的集成度更高）和不断增加的业务压力（加快创新步伐）导致工程和运营中对基于模型的软件工具的需求增加，带来了新的挑战。同时，现有的专业知识（如训练有素的专家）限制了基于模型的方法的扩展。因此，需要计算机辅助工程和操作的新模式。在过去的几年里，数字孪生已经成为一种新的范式，带来了技术进步，成为未来十大数字化战略技术趋势之一。数字孪生系统是一个系统中所有可操作的系统相关信息和数据的集合，该系统在其生命周期的各个阶段（从构思到运行结束）都进行了更新和升级。对于配备嵌入式功能的产品，可以为日常决策提供帮助，以提高系统的可靠性、质量和效率。

模型降阶是数字孪生的核心技术之一，因为它可以压缩仿真模型进行实时仿真，并且可以在没有知识产权风险的情况下进行模型交换。最重要地，它能够在产品生命周期的后期，特别是在产品运行阶段，重用产品开发早期阶段的仿真模型。

经典仿真应用于产品和生产开发的早期阶段的设计和工程问题，以支持设计规划和尺寸标注。然而，在新兴的数字孪生系统中，需要对整个生命周期进行仿真。仿真将不仅支持早期开发阶段或后期测试阶段的工程师，还支持运营阶段的操作员和用户，例如，基于模型的状态监测和优化控制，最终实现基于新仿真的维护服务。这就需要将新的动态软件模型集成到现有的仿真工作流中，以便在可执行程序中存储和访问现有的专业工程知识。正是这种对现有专业知识的映射和存储使得数字孪生视觉具有价值。非专业和产品生命周期其他阶段的知识是可重用的，如图2-1所示。



**图2‑7 数字孪生中的模型转换**

#### 从三维仿真到系统仿真的模型交换

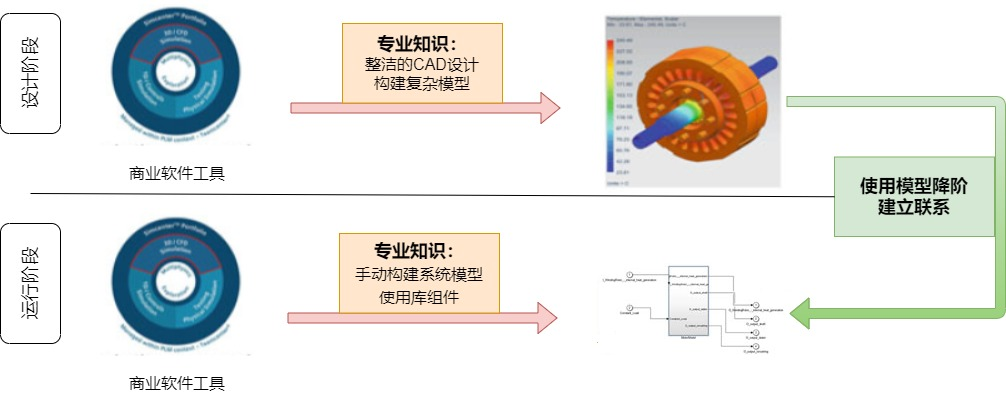
如今，仿真几乎在所有工程部门都是一项成熟的技术。然而，仿真通常只在Silo（用于读/写各种科学数据至二进制磁盘文件的API函数库）格式使用。仿真的一个关键挑战是仿真在其他领域的重用，是更广泛采用基于模型的系统工程（MBSE，Model-Based System Engineering）的主要拦路虎。MBSE使用数字系统模型，在系统级捕获所有子系统的互联行为。虚拟规程特定模型和系统模型是验证所需产品行为的基础。利用这些增量式、连续式的适应性模型，在整个研发过程中模拟了由需求派生的虚拟测试场景。因此，系统需求、功能和行为之间的早期关联是可能的。

MBSE区分了不同层次的仿真，如：

系统仿真：模拟主要行为（如Mworks、Matlab），涵盖软件和控制等多个学科，是系统级系统研究已开发解决方案的基础。

专业特定的建模和仿真：专业特定的三维几何或三维计算机辅助工程（CAE）模型（如西门子Simcenter、Comsol或ANSYS multi-physics）具有高度复杂性和准确性，用于详细方面的验证和优化。

然而，在当今的仿真实践中，三维仿真部门和系统仿真部门是分开的，因为每个部门都使用各自领域中可用的商业或内部软件工具在内部构建仿真模型。然而，一旦建立并校准了详细的三维仿真模型，就需要压缩所包含的信息，并将其作为系统仿真模型中组件的自定义模型重用。今天的工业实践仍然是手工地重建和更新系统模型。这正是模型降阶发挥作用的地方，因为它允许实现压缩步骤。与供应商中立的功能模型接口（FMI）标准一起（截至目前已在100多种工具中实施），模型降阶算法可以将详细的三维仿真模型作为输入，并以标准格式为系统仿真软件提供压缩模型作为输出，如图2.2所示。值得注意的是，模型降阶技术不仅可以实现模型转移，还可以实现模型交换的自动化。

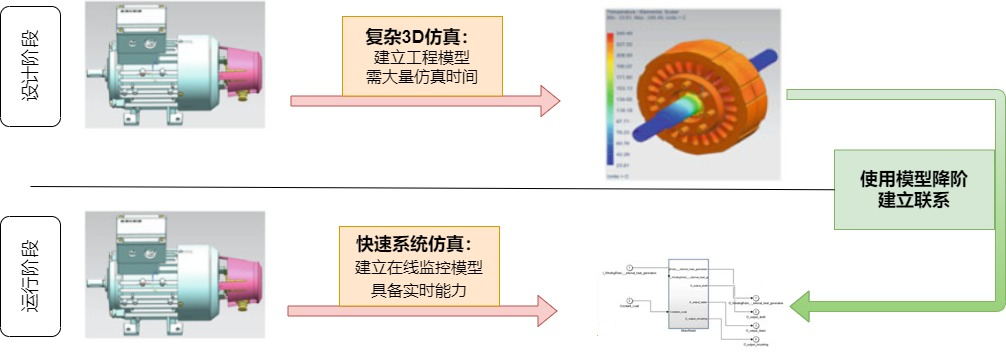


**图2-8从三维仿真到系统仿真的模型交换**

#### 从设计阶段到运行阶段的模型交换

随着可用计算能力的不断增长以及数学算法的发展，计算机辅助范例也被确立为操作的关键技术。比较潜在收益与开发和安装成本的比率，计算机辅助方法远远优于经典方法。例如，通过先进的控制解决方案来提高卡车的能源效率比空气动力措施有显著更好的成本比改善。

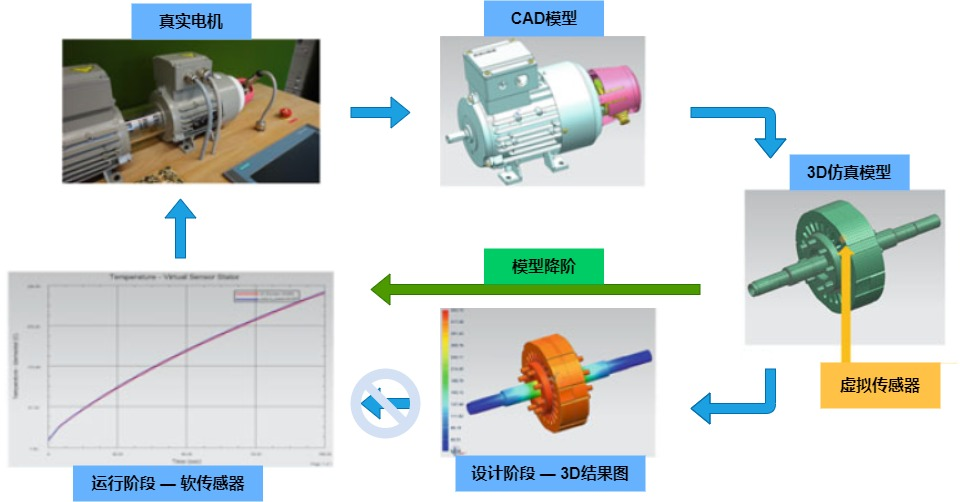
一方面，机器学习方法正在通过从越来越多的传感器数据中学习来取代目前基于专业知识和启发式的概念。而在工业环境中，机器学习常常受到数据可用性的限制。例如，许多客户的数据本身携带知识产权信息，或处理过去没有收集到足够的数据的特定故障。另一方面，模型预测方法可以显著改善操作。然而，由于并行运行的仿真模型必须比实时运行的要快，这就要求专家付出很高的努力。在产品设计过程中使用的仿真模型在计算上通常很昂贵，需要大量的解决时间。因此，相应的模型需要额外实现，相应的方法被限制在高价值或高批量的应用程序，以确保业务进展。模型降阶算法甚至可以压缩巨大的三维仿真模型，这样得到的降阶模型就足够小，可以比实时评估更快，如图2.3。因此，模型降阶可以处理有限的数据，并避免使用基于数据的方法在操作阶段重新生成信息。这使得可以提供新的服务来预测故障、提高运营效率或进行业务规划。



**图2-9从设计阶段到运行阶段的模型交换**

#### 模型降阶

数字孪生技术的核心是连续的背景模拟，使新型辅助系统能够在所有生命周期阶段实现。如上所述，模型降阶是实现实时功能的关键技术。其中，重点研究（线性）三维偏微分方程的模型降阶，这在施工设计中被广泛使用。经过空间离散化，得到了巨大的线性稀疏方程组（未知量级通常由106）。通过将计算拆分为离线阶段和在线阶段，计算工作量转移到离线阶段，允许在线阶段进行交互式和实时仿真。在过去的几十年里，各种概念和方法被引入，主要是使用基于投影的方法，如本征正交分解、平衡截断、缩减基或Krylov子空间方法。大多数方法的关键思想是通过适当的低维基来减少需考虑函数的空间。对于(接近)线性模型，模型降阶是计算工程和科学中最先进的技术，如图2.4所示。然而，对于非线性模型来说，这是一个非常活跃的研究领域。



**图2-10产品生命周期仿真链中的模型降阶（绿）超越目前的方法（蓝）**

#### 小结

模型降阶可能是在产品生命周期中实现数字孪生的最重要的数学技术。它使仿真模型的交互性、可靠性、连续性、可访问性和可分布性达到了新的水平：

* 交互性：有许多方法可以加速3D模拟，例如，利用计算机架构或先进的解决方案技术。模型降阶直接减少了自由度的数量，因此能够快速生成任意模型。对于实时应用，可以实现精度和速度之间的适当折中，例如，通过自适应方法。
* 可靠性：在大多数情况下，模拟工具的加速会导致精度的下降。然而，这种精度的下降应该被量化，因为用户想知道简化模型的预期准确性。对于许多模型降阶方案，详细的误差分析是已知的。因此，可以使用简化模型提供误差范围和置信区间(通过数据校准时)。
* 连续性：多年来，仿真工具在生命周期内的无缝工作流仍然是一个主要的愿景。对于设计来说，为了获得精确的模型，需要花费大量的手动工作。在系统级仿真和实时应用程序中，完全相同的组件需要非常快速的模型，这同样需要大量的手动工作。模型降阶技术恰好连接了不同的应用，即模型降阶技术旨在从综合设计模型中自动生成快速模型。
* 可访问性：为设计建立复杂的模型仍然是专家的任务。然而，通过模型降阶生成的快速模型可以在标准化格式后进行分解，因此也可以被非专家集成到系统仿真中。这一点是数字孪生的一个核心方面。
* 可分布性：求解由模型降阶生成的快速模型只需要矩阵向量乘法。因此，也可以在非常小型的计算机上运行(而且不仅仅是在专门的工作站上)。此外，模型不允许重建组件，因为这可能来自于计算机辅助设计（CAD）或网格计算机辅助工程（CAE）模型，即保护知识产权，这是另一个主要问题。每一个复杂的部件都将与相应的数字孪生部件一起交付。

### 使用模型降阶技术理解Modelica中实现的模型

Modelica能够快速开发异构和复杂系统的详细模型。然而，结果模型与现实本身一样复杂，因此可能很难确定模型行为的原因或验证模型行为正确。传统的工程方法是使用直觉和经验来识别特定场景中对模型行为影响最大的模型的重要部分。大量的模型降阶和简化技术(例如这些方法使用的度量标准)已经被开发出来，以自动评估特定场景的模型的重要部分，从而减轻对主观因素的依赖，例如直觉和过去的经验。

本章节讨论了模型降阶和简化技术(例如这些技术用于元素排序的度量)，这些技术适用于从现有库构建的Modelica模型。Modelica模型被转换为微分代数方程，对于后者，已经有许多工具用于模型阶数的降低。然而，这些工具并不是为了帮助用户理解模型的行为而设计的，而且简化后的模型可能很难被用户理解，因为原始模型的结构会丢失。必须保留模型的层次分解，如果模型是用图形原理图开发的，则必须对原理图的元素(节点)进行排序。因此，将用于键合图元素排序的基于能量的度量标准调整为定义更松散的Modelica原理图，可使它们可以与处理方程的排序方法互补使用。

#### 简介

Modelica语言设计的一个重要方面是用于大型、复杂和异构物理系统高效建模的用户交互。模型通常被分解为几个层次。层次结构底部是基本物理现象的子模型，通常表示为一组微分方程，这些方程可以直接输入(不需要任何形式的操纵,甚至转换其他形式描述)。在更高的层次上，模型用原理图(即对象图)进行图形化描述，得到的方案通常反映了系统的拓扑结构。因此，Modelica中的模型表示对于不熟悉动态系统计算机仿真的领域专家也是可以理解的。

Modelica是面向对象的建模语言，因此包括继承和可替换模型等特性。这种语言特性对于模型库的有效实现是必要的，但是它们增加了实现的复杂性，并且使从库中浏览组件的源代码更加困难。例如，DynamicPipe组件是一个来自标准Modelica库的具有分布式质量、能量和动量平衡的直管模型，它由四个基本模型和三个可替换的元素组成，这些元素也是具有复杂继承层次的模型。对管道的动力学、平衡方程和管道中介质的热力学状态的描述，被分成十多个(部分)模型，以实现高效的组件重用和防止代码重复。这种模型分解可能没有物理意义，它只处理实现问题。

在实践中，领域专家通常已经有了一些计算(例如在Excel中)，他们想用这些计算来验证在Modelica中实现的模型，也可以用来厘清Modelica中更为详细和复杂的模型的意外行为(通常)。因此，需要在他们的计算和Modelica中的模型之间进行匹配。然而，支持Modelica的建模环境目前并没有提供许多工具来帮助研究和探索模型。最复杂的模型的建立与使用不同的模型库，当这些库的文档不够了，特别是当子模型是高度定制的组件(可替换的子组件和修改),有必要看看下面调用的组件使用。但是由于库组件的实现很复杂，所以这是令大多数领域专家都措手无策的问题。

工程师利用经验和直觉来确定模型的重要部分，在特定的场景下，对系统的主导动态或模型的仿真响应有最高的影响。为了减少对主观因素(如经验)的依赖，已经开发了许多建模度量和方法。它们通常需要严格的建模形式，因此不需要花太多精力将它们集成到Modelica环境中。

#### 模型降阶和简化技术

复杂系统的详细模型也和它们所建模的现实一样复杂和难以理解。即使对于相对较小的系统，对基本方程的解释或对深入系统理解的提取也有几率是不可能的。因此，符号分析方法，特别是在电路中，包含了各种符号近似技术，这些技术用于简化符号表达式或示意图，并降低模型的阶数(状态空间维数)。

在系统分析或结构设计中使用的模型阶数降阶和简化方法的一个重要类别是当它们生成适当的模型时，即，具有满足性能规范所需的最小复杂度和具有物理意义的参数和状的降阶模型。到目前为止，已经开发并成功应用了多种数字-符号混合模型的降阶和简化技术。它们通常包括运行一系列模拟，根据适当的(定量)指标对单个坐标或元素进行排序，并删除低于特定阈值的元素。

##### 基于方程的简化

所有分析模型都可以用一个方程组来描述，即使使用了一些其他的建模形式(例如方框图、键合图等)，也可以将模型导出为一个方程组。然而，由符号(代数)表达式组成的模型表示只有在某些情况下对用户才有意义，例如，在控制设计中使用传递函数。对于基于方程的简化，必须选择特定的变量，然后选择用于表达式项排序的指标作为有关特定变量给出的目标函数的数值误差。简化策略包括各种代数操作(例如变量的替换)，其中没有错误被引入到简化的方程式，和方程式的修改导致近似系统(例如项删除、方程式的线性化等)，这需要一个数值模拟来确定由修改引起的错误。简化可以产生全局效应，即对系统的某些变量进行操作时影响整个方程组，或对一个方程的单个项进行操作时产生局部效应。

##### 基于结构的简化

大多数现代建模工具都提供了图形界面，其中模型由原理图表示。模型的图形化描述基于各种建模形式，原理图可以仅仅是代数表达式的图形化表示(例如块图)，组成原理图的符号表示单个或一组代数操作，或者它们可以提供关于系统的附加信息(例如系统拓扑信息)。在后一种情况下，选择自定义简化技术是明智的，尽管可以将基于方程的阶数减少和简化技术简化的模型映射为原始模型的图形表示。因为所有的物理系统都有共同的质量和能量守恒，所以在物理系统建模中，一个广泛使用的用于适当模型降阶的度量类与能量或功率有关。基于能量的度量需要一种建模形式，其中模型组件的能量很容易提取。键图建模是面向对象的物理系统建模的一种形式:元素可以看作是相互作用的对象，相互作用被描述为因果键。成功应用的基于能量的键合图简化技术有基于键合RMS功率的元素排序，在给定的时间内，对元素进出的能量活动量进行排名，并将每个键的能量与相邻键的能量进行比较，消除相对能量最小的。这些技术简化模型的结果也是一个键合图描述的模型。此外，所有基于能量的指标都有一些物理意义，因此可以帮助理解和解决建模问题。

#### Modelica模型的简化

目前还没有建模环境可以为Modelica中直接实现的模型的简化和次序缩减提供工具。一个Modelica模型必须被平坦化，然后将生成的DAE系统导出到指定的工具中，在那里模型的降阶和简化将被执行。

这种方法适用于某些应用，例如需要简化模型进行控制设计时。在这种情况下，扁平化造成的信息损失是没有问题的，因为只需要简化模型和原始模型的行为紧密匹配。然而，在模型验证和调试等应用中，或者当模型被用于洞察系统性能改进时，希望简化模型也是一个有效的Modelica模型，具有与原始模型相同的结构(具有相同的层次分解和原理图拓扑)。

#### 小结

模型的展示对用户来说是建模环境的一个重要方面，它有助于理解和维护模型。然而，许多Modelica的语言特性(例如继承)对于有效实现和防止代码重复是很重要的，但是可能会使模型实现的清晰度变差。此外，复杂系统的详细模型和现实模型一样难以理解。因此，应该在建模环境中集成一个工具，以帮助非建模专家用户更好地理解模型，并为解释模型行为和模型验证(和调试)提供有效的方法。模型降阶和简化技术(例如这些技术所使用的排名指标)可以用于此目的。重要的是，结果以与原始模型相同的形式呈现。Modelica模型用图形表示，通过对象图，或作为一组因果微分代数方程。因此，必须对这两种表示法使用模型阶数的降阶和简化技术。目前已经有许多方法可以简化DAE系统表示的(排序)模型。简化键合图的方法(图形建模形式)可以适用于Modelica的对象图。

# 模型降阶理论与方法研究

## 总体研究

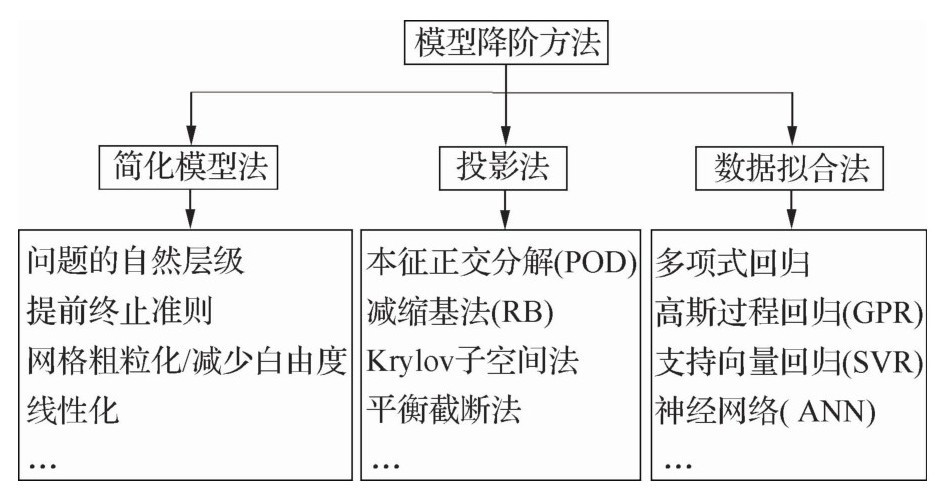
模型降阶技术是打通CAE到数字孪生必不可少的一环。成熟的模型降阶工具箱可以用于市场上绝大多数CAE的结果处理，可以为庞大的数字孪生市场提供方案。传统CAE/实验方法对于复杂领域、复杂场没有很好的求解方案，计算时间长，远远达不到数字孪生要求的实时仿真的速度。即使经过传统方法计算得到了结果，但是对于其中任意参数改变，传统方法需要重新花费大量的时间计算。而模型降阶模型一旦建立完成，对于给定范围内的任意工况均可以在一定工况下以实时计算的速度计算给出结果。

对于一个数字孪生体，其几何特征、材料属性、载荷信息、检查方法和引发的损伤等都是有差别的。通过对物理实体进行实时监控并将数据传输到构建的虚拟孪生模型中进行仿真，是数字孪生技术的基本模式。下面将围绕数字孪生的四大特点分别阐述基于模型降阶的一三维融合建模技术在其中体现的作用：

1. 虚实结合：旨在建立真实模型与虚拟模型之问的双向映射，需要实现虚实深度融合，基于真实环境传递而来的参数等历史数据、经验与知识数据等进行智能分析与决策。基于模型降阶技术构建的一维代理模型，一方面由于对于大量数据的降阶处理，可通过更为简洁的形式管理不同工况的模型和数据，另一方面，保证了系统级虚拟模型向真实模型映射过程的效率和精度。
2. 时效性：真实模型与虚拟模型的数据通信、监测与检测数据的获取与模型的预测都对时效性提出了很高的要求。传统的三维有限元模型与一维系统模型的融合无法解决计算效率问题，模型降阶技术通过对数据的模态分析及模型的简化处理，可以完美解决三维设备级模型和系统级模型的耗时问题。以大型CAE工程计算为例，对于大型复杂流体、结构、耦合场需要近千万级的网格去计算，常用的100核工作站大概需要计算上7-30天，而用模型降阶计算，在普通个人笔记本电脑计算只需要0.1s甚至更短。
3. 多学科/多物理性：武器装备数字孪生体是基于物理特性的实体产品数字化映射模型，不仅需要描述武器装备的几何特性，还需要描述实体产品的多种物理特性，其可以包括并耦合飞行力学、空气动力学、结构力学、疲劳扩展、流场分析等多种物理模型，以及刚度、强度、疲劳等多方面的材料特性。基于数据的模型降阶方法本身由于不受各学科控制方程的影响，降阶模型的创建只需要仿真/实验数据作为输入，因此，这一类降阶方法可完美覆盖多学科/多物理性的需求。
4. 多尺度/多保真性：武器装备结构典型部件的尺寸在10 m量级，而初始裂纹尺寸一般在10-4 m量级，模型尺寸跨越了部件-零件-元件等多个尺度。数字孪生体应可描述武器装备系统级、设备级、组件级等多个尺度的行为和特性，也应该使用一系列不同保真度的全阶／降阶模型，从而稳健、高效、准确地对武器装备的状态和行为进行高性能仿真。基于数据的模型降阶方法可根据不同尺度和仿真精度的要求灵活调整，保证求解速度与精度达到平衡。

## 模型降阶方法研究

从降阶的实现方法分类，可以将现有的降阶模型分为3种，简化模型法、投影法和数据拟合（代理模型）法，如下图所示：



**图3-1常用模型降阶方法**

模型简化法结合领域的专业知识，对模型细节进行适当简化以降低模型的复杂度，是基于物理方程的模型降阶。因此对于相同的全阶模型，根据对细节的简化程度，可以给出不同保真度的模型。如对湍流的分析，从直接数值模拟、大涡模拟、雷诺时均NS方程等，模型的保真度依次递减。此外还可以简化有限元模型的网格，忽略模型中的非线性项等。

基于投影的方法主要基于数学推导，而不是专业知识。这类方法一般通过构造一个子空间，将控制方程投影到该子空间，以实现模型空问的降维。常用的有本征正交分解方法、平衡截断法、Krylov子空间法、动态模式分解法、缩减基法等。

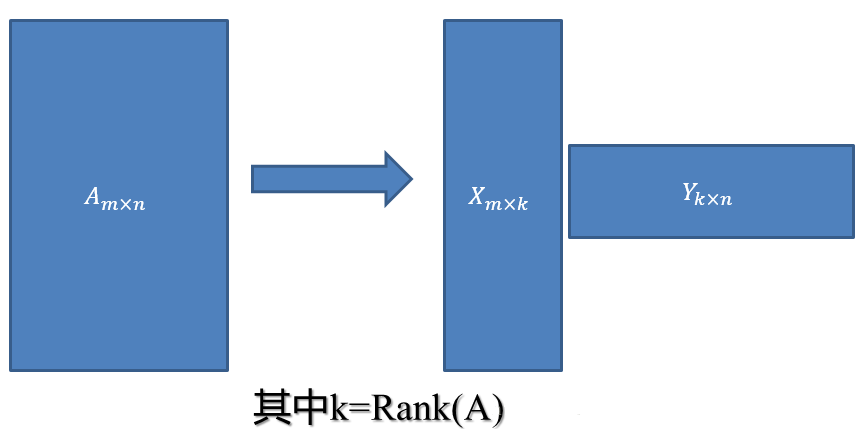
数据拟合方法，也称代理模型方法，旨在建立模型输入输出参数之问黑箱式的映射关系，以替代精细化仿真。是实验设计、数理统计和优化技术的综合应用，可以减少复杂、耗时的分析过程。在离线阶段使用昂贵的仿真或试验手段建立代理模型，并在在线阶段使用该模型进行快速预测，日前在结构分析中已经得到广泛使用。因为结构分析输出大部分是一个连续值，所以代理模型一般是回归模型，常用方法响应面、高斯过程回归、支持向量回归、人工神经网络等。随着近些年深度学习的逐步发展，各种深度学习方法也逐渐应用于武器装备的维护和管理，如长短期记忆网络、深度自编码器、卷积神经网络、深度置信网络等，展现出了较大的应用潜力。

下面主要介绍基于投影的模型降阶方法和数据拟合方法的研究内容。

### 基于投影的模型降阶方法

#### 奇异值分解（SVD）

SVD不仅是一个数学问题，在工程应用中的很多地方都有它的身影。用SVD可以很容易得到任意矩阵的满秩分解，用满秩分解可以对数据做压缩。SVD这个可以应用在数据降维压缩上，在数据相关性特别大的情况下存储和矩阵比存储矩阵占用空间更小（如下图所示）。



**图3-2 SVD图例**

奇异值分解是一个能适用于任意的矩阵的一种分解的方法，对于任意矩阵总是存在一个奇异值分解：

其中，是一个的方阵，称为左奇异矩阵，该方阵中的正交向量称为左奇异向量。是一个的对角阵，对角线上的元素称为奇异值，由大到小排列。是一个的方阵，是方阵的转置矩阵，方阵V又称右奇异矩阵，该方阵中的正交向量称为右奇异向量。为得到这两种奇异向量的表达式，首先需要将矩阵通过其转置矩阵转化为一个方阵，然后进行EVD处理。设为左奇异向量，为右奇异向量，则有

的特征向量组成的矩阵即为SVD的矩阵，如下所示

其中，，，式3.6变为

同理可知，的特征向量组成的矩阵即为SVD的矩阵。

奇异值的求解方法有两种：

1. 第一种：
2. 第二种：

由式3.7可知，特征值矩阵等于奇异值矩阵的平方，即

在奇异值分解矩阵中，矩阵里的奇异值按从大到小的顺序排列，在很多情况下，前10 %甚至1 %的奇异值的和就占了全部的奇异值之和的99 %以上。也就是说，剩下的90 %甚至99 %的奇异值几乎没有什么作用。因此，可以用前r个奇异值来近似描述矩阵，因此，奇异值分解公式可以写成如下：

r的取值越接近于n，则式3.10右边的结果越接近于。如果r的取值远远小于n，从数值计算的角度来说，右边三个矩阵的存储内存要远远小于原矩阵。所以在奇异值分解中，r的取值就是在计算精度和时间空间之间做选择。

#### 主成分分析（PCA）

主成分分析方法（PCA，Principal Component Analysis）是一种使用最广泛的数据降维算法。PCA的主要思想是将n维特征映射到k维上，这k维是全新的正交特征，或称为主成分，是在原有n维特征的基础上重新构造出来的k维特征。PCA的工作就是从原始的空间中顺序地找一组相互正交的坐标轴，新的坐标轴的选择与数据本身是密切相关的。其中，第一个新坐标轴选择是原始数据中方差最大的方向，第二个新坐标轴选取是与第一个坐标轴正交的平面中使得方差最大的，第三个轴是与第一、二个轴正交的平面中方差最大的。依次类推，可以得到n个这样的坐标轴。通过这种方式获得的新的坐标轴，大部分方差都包含在前面k个坐标轴中，之后的坐标轴所含的方差几乎为0。于是，余下的坐标轴可以忽略，只保留前面k个含有绝大部分方差的坐标轴。事实上，这相当于只保留包含绝大部分方差的维度特征，而忽略包含方差几乎为0的特征维度，实现对数据特征的降维处理。

通过计算数据矩阵的协方差矩阵，得到协方差矩阵的特征值和特征向量，然后选择特征值最大(即方差最大)的k个特征值所对应的特征向量组成的矩阵。这样就可以将数据矩阵转换到新的空间当中，实现数据特征的降维。得到协方差矩阵的特征值特征向量有两种方法：特征值分解协方差矩阵和奇异值分解协方差矩阵，因此，PCA算法有两种实现方法：基于EVD的协方差矩阵实现PCA算法以及基于SVD的协方差矩阵实现PCA算法。下面简单介绍基于SVD的协方差矩阵实现PCA算法。

SVD方法在实现PCA的步骤上与EVD一致，因此，二者在降维结果上也是完全一致的。对于协方差矩阵，若使用EVD，优化过程如下：

令

式3.11变为

由上可知，特征向量矩阵即为特征向量矩阵的转置，映射后数据的每个特征的方差信息在中。

同理，若使用SVD，可得

也就是说，用SVD方法来代替EVD方法能保持结果不变。同时，使用SVD方法的优势在于当进行奇异值矩阵按数值从大到小排序时，前10 %甚至1 %的奇异值的和就占了全部的奇异值之和的99 %以上的比例，因而在计算机矩阵存储时往往非常有效，大大减少了矩阵存储所消耗的储存空间。

#### 中心流形方法

中心流形模型降阶技术是基于中心流形定理提出的，是一种局部降阶方法。该方法通过寻找平衡点附近的中心子空间与稳定子空间的微分同胚映射，从而将高维系统投影到低维的中心子空间上，但保留了平衡点附近高维系统的全部拓扑性质。因此高维非线性动力系统在局部分岔点邻域的渐近行为与原动力系统的中心子空间上的渐近行为拓扑等价。

中心流形模型降阶技术的基本思想是通过寻找平衡点附近的有限维中心子空间，然后将方程向中心子空间投影，进而得到低维模型。

考虑微分方程

其中，，。

设平衡点为，如果平衡点不在点，可以利用平移变换和将其移至平衡点。

当时，式3.15可以写为

其中为非线性函数，而为线性算子。算子的谱由两部分组成，为对应实部为负的特征值对应的谱，为对应实部为0特征值对应的谱

式3.16可以分解为

根据中心流形定理，存在微分同胚，使得，且满足，将其带入式3.17的第一式，则可求出中心流形上流的方程为

这样一来，式3.18可以代替式3.15描述在附近的动力学行为。

为计算**，**将和式3.17的第一式带入第二式得

在中心流形的求解过程中，通常需要将展成幂级数，然后带入式5.30，通过系数之间的比较后可求出，之后将带入式3.19，便可得到中心流形上的流的方程，从而达到降维目的。

中心流形方法具有如下优点：

1. 在适当的初始条件下，所有的渐进解都可以被中心流形吸引；
2. 有严格的数学基础，而且在降维时，不需要积分计算；
3. 由于中心子空间对于每个确定系统来说是固定的，因而利用中心流形方法降维免去了讨论降维后系统的维数问题。

中心流形方法的局限性：

1. 局部理论决定了只能讨论平衡点附近的行为；
2. 求解过程中可能存在收敛方面的问题。

#### Lyapunov-Schmidt约化

L-S约化方法全名又叫 Lyapunov-Schmidt 方法，该方法与中心流形方法一样，也是一种局部降维方法，不同的是：中心流形理论是将高维非线性系统缩减为一个低维的动力系统，且保持与原系统动力学拓扑等价；而 L-S 约化方法直接处理高维非线性系统定常解的静态分岔问题。

L-S 约化方法的基本思想是把高维非线性代数方程组投影到其值空间的两个相互正交的子空间上，从而得到两组方程，由隐函数（或广义隐函数）定理可知，其中一组方程有唯一解，通过求解该方程，并将其解代入另一组方程中，便可将高维静态分岔问题的求解约化为一个低维方程的求解。与式3.15对应，考虑由非线性映射给出的方程

其中，，是巴拿赫空间，是参数集。

考虑远点附近的分岔问题，且为弗雷德霍姆算子。

记为的值空间，为的零空间，且（分岔的必要条件），则存在直和分解为：，。

定义投影算子：和其补算子：，为恒同算子，记，式3.20等价化为

若是上的算子，那么可逆，由隐函数定理，式3.21*a*存在唯一解（且）。将代入3.21*b*，设映射，由

确定，从而

此时得到低维方程，达到降维目的，以上就是L-S约化技术的过程。

式3.22可以写为

此时只含有零空间坐标。

L-S优化方法具有如下优点：

1. 可以同时讨论多个参数的问题；
2. 应用范围不局限于自治系统。

L-S优化方法的局限性：

1. 对于实际系统，空间分解存在困难。

#### 模态综合法

模态综合（CMS）法也是一种有效的降阶方法，其本质上是一种经典的Galerkin方法，该方法被广泛应用于大型复杂结构系统的模型降阶。模态综合法的基本思想是将大型复杂结构系统划分为若干子结构，忽略各子结构高阶模态的影响，综合各子结构低阶模态所构成的系统，从而达到缩减系统自由度的目的。该方法按子结构的不同联接方式，可分为固定界面模态综合法、自由界面模态综合法以及混合界面模态综合法。

模态综合法的实现可以分以下几个基本步骤：

(1) 总结构分割为若干子结构. 将整个结构按照工程结构部件的自然组织情况和几何轮廓分割成若干子结构，分割的子结构要便于进行独立试验和计算；

(2) 各子结构的模态分析.通过数值分析、试验方法等对各子结构独立地进行模态分析并获得各子结构的模态信息，保留其主要模态信息，略去高阶模态，从而达到模型降阶的目标；

(3) 各子结构运动方程的集成。在各个子结构模态分析的基础上，将各个结构的模态信息按照交界面的协调条件装配，获得自由度大大降低的整个结构运动方程.。求解此低阶方程便可获得结构的各阶固有频率和模态坐标下的主振型；

(4) 子结构再现。实际问题中学者感兴趣的是物理坐标下的振动特性，因此必须完成由模态坐标返回到各子结构物理坐标的转换，进一步得到实际结构的固有振型和位移等动态响应。

#### 本征正交分解（POD）

本征正交分解方法本质上是将高维系统投影到一个低维子空间上，不同的是该方法是一种统计方法。POD 降阶模态或降阶基函数是通过求自相关矩阵的特征向量获得，而自相关矩阵是具有统计意义的 2 阶中心距。POD 方法的基本思想是：通过复杂系统的数值仿真信号或实验数据信号构造自相关矩阵，再求解自相关矩阵的特征向量获得 POD 降阶模态或降阶基函数，然后将高维系统投影到前几个最大特征值对应的降阶模态张成的子空间上，从而达到降阶的目的。

随着计算机技术蓬勃发展，各类CFD求解技术层出不穷，计算效率与计算规模飞速提升。然而，现有全模态CFD的计算效率基本无法满足实时系统仿真计算需求。Reduce Order Method（ROM），即模型降阶方法，应运而生，其主要思想为：基于一定的准测对流场数据进行模态分解并提取其中的重要成分，然后在根据相应的动力学原理将其还原为原流场的近似解。该解可在保流较高精度的条件下实现在对原流场的快速复现，并允许在一定条件下改变流场的边界条件，以获得其他工况下的估计解。

尽管在数学上已经证明在条件合适且不断提高阶数的条件下POD-Galerkin模型可收敛至全模态N-S方程，然而在实际应用中基于能量解析的POD展开式可能导致降阶模型的计算结果脱离设计要求。N-S方程对边界条件与初值条件较为敏感，理论上较为稳健的流场方可得到较为稳健的降阶结果。POD分解关注与流场的主模态即惯性尺度，忽略了流场的小尺度运动，然而流动中小尺度运动往往对流场的稳定与性质起到了关键作用，高精度POD-Galerkin模型应对小尺度运动进行合理建模。边界条件或初值条件的变动可能彻底改变流场的模态，致使POD-Galerkin模型无法准确捕获变条件后的流场。因而使用参数化边界条件的POD-Galerkin模型时，应当注意保持求解基底流场同目标流场间的模态同质性。

### 基于数据拟合的模型降阶方法

#### 响应面方法

20世纪90年代开始响应面模型开始广泛的应用于结构优化设计领域中。响应面模型具有模型简单、构造容易、具有显式表达式、优化时收敛速度快的优点，二阶响应面模型如下：

其中 为试验输出量，为系数，为输入变量。分析发现，最大到一次项太粗略，往往和实际值差别较大。实践证明，最大到二次项一般情况下可以给出较为满意的答案，即

多次观测后写成矩阵的形式为：

其中

方差为

当最小方差最小时，显然拟合曲面和实际值最接近。当 对的偏导数为零时，最小方差取极小值。此时

可得

可以获得拟合的响应面为：

这种通过一系列多个变量、确定性的“试验”，来模拟真实极限状态曲面的方法称为响应面法。同理，高阶也是类似求法。

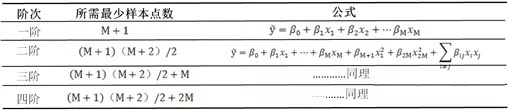


图3‑3响应面公式

#### Kriging插值法

面对大多数复杂的工程结构设计问题，为获取设计参数与结构响应之间的关系，工程技术人员往往采用代理模型。实践表明，代理模型具有精度高、计算量小，低误差等特点，可广泛应用于航空航天等领域。由于Kriging模型具有良好处理非线性预测能力，Kriging模型已成为最具代表的代理模型之一。

Kriging插值方法可以接受任意数目的设计量,保证足够的插值精度而不会对计算时间造成影响，可以应用于任意维度的插值过程。考虑模型

只要得到了的表达式，就可以得到参数空间中的任意设计点的估计值。Kriging模型的重要假设是将未知函数视为一个高斯随机过程,该随机过程可以表示为

其中，为参数变量，称为全局趋势模型，为随机分布的误差。将随机过程的方差记为，可用协方差描述不同位置点这些随机变量的相关性,其协方差可描述为

其中，为相关函数(只与空间距离相关)，相关性会随距离增大而不断减小。相关函数通常有指数函数、高斯函数、三次样条函数等多种形式。

基于上述假设，Kriging模型寻找最优加权系数ω,使得均方差最小,并且满足无偏差条件。通常利用拉格朗日乘子法可以推导得出一个线性方程组,求解该方程组便可得到Kriging模型预估值为

其中，只与已有样本点相关，，为样本数据，为一维单位向量，和则是前述的相关函数矩阵及其中的相关向量，故便可利用样本数据得到模型的估计值。

构建Kriging模型时,还可以将对其模型参数(或超参数)也作为未知量来加入训练，可以进一步增强代理模型的灵活性。利用已有样本点和响应值可将问题转化为一个非线性无约束优化问题，一般采用“最大似然估计”或“交叉验证”得到超参数。对于最大似然估计而言，就是选择超参数使下式函数值达到最大：

其中，和的最优值可以给出解析解为

将和代入后,取对数忽略常数项，优化问题转化为

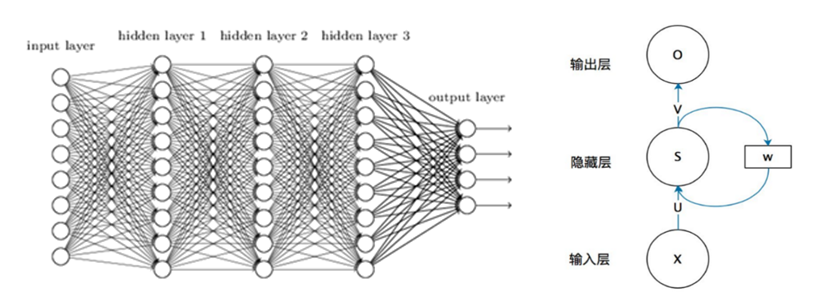
利用数值优化算法如遗传算法、单纯形法、模式搜索法等训练超参数，就可以构建最终的Kriging模型。

#### 传统神经网络方法

神经网络具有初步的自适应与自组织能力。在学习或训练过程中改变突触权重值，以适应周围环境的要求。同一网络因学习方式及内容不同可具有不同的功能。人工神经网络是一个具有学习能力的系统，可以发展知识，以致超过设计者原有的知识水平。通常，它的学习训练方式可分为两种，一种是有监督或称有导师的学习，这时利用给定的样本标准进行分类或模仿；另一种是无监督学习或称无为导师学习，这时，只规定学习方式或某些规则，则具体的学习内容随系统所处环境 （即输入信号情况）而异，系统可以自动发现环境特征和规律性，具有更近似人脑的功能。

我们把学习集（Learning Set）中的每个输入加到神经网络中，并告诉神经网络输出应该是什么分类。在全部学习集都运行完成之后，神经网络就根据这些例子总结出她自己的想法，到底她是怎么归纳的就是一个黑盒了。之后我们就可以把测试集（Testing Set）中的测试例子用神经网络来分别作测试，如果测试通过（比如80%或90%的正确率），那么神经网络就构建成功了。

神经网络是通过对人脑的基本单元——神经元的建模和联接，探索模拟人脑神经系统功能的模型，并研制一种具有学习、联想、记忆和模式识别等智能信息处理功能的人工系统。神经网络的一个重要特性是它能够从环境中学习，并把学习的结果分布存储于网络的突触连接中。神经网络的学习是一个过程，在其所处环境的激励下，相继给网络输入一些样本模式，并按照一定的规则（学习算法）调整网络各层的权值矩阵，待网络各层权值都收敛到一定值，学习过程结束。然后我们就可以用生成的神经网络来对真实数据做分类。下面将介绍两种最常见的神经网络，如下图所示：

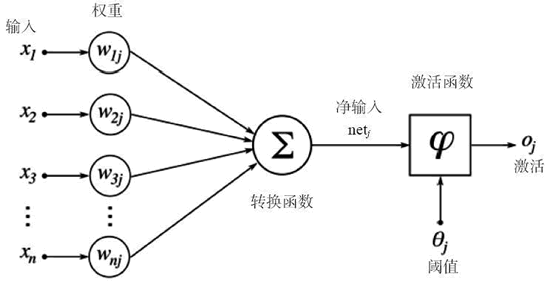


**图3-4 深度神经网络（左）和循环神经网络（右）**

建立神经网络模型的第一步是定义瞬态模拟模型（几何、网格、材质属性等）和设置（如时间步长和边界条件）。神经网络模型是通过数据训练确定的，仿真数据可由一系列数值模拟产生。将每个仿真算例对应的初始条件与降阶后的矩阵作为神经网络的输入与输出，将输入与输出集合分为训练集和测试集。使用训练集对神经网络模型进行训练，使用测试集对训练后的神经网络模型进行评估，不断重复此步骤，直至模型收敛，可得到训练后的神经网络模型，至此神经网络模型的建立阶段完成。

神经网络是基于感知机的扩展，而DNN可以理解为有很多隐藏层的神经网络。多层神经网络和深度神经网络DNN其实也是指的一个东西，DNN有时也叫做多层感知机（Multi-Layer perceptron,MLP）。

感知机（图3-7）是一种线性分类模型，由两个主要部分构成：一个加法器，将所有输入加权求和到神经元上；一个处理单元，根据预定义函数产生一个输出，这个函数被称为激活函数。每个神经元都有自己的一组权重和阈值（偏置），它通过不同的学习算法学习这些权重和阈值。



**图3-5感知机**

输入层被称为第零层，因为它只是缓冲输入，存在的唯一一层神经元形成输出层。输出层的每个神经元都有自己的权重和阈值。感知机算法是利用那些分类错误的实例点不断的调整其中的权重，最后达到没有错误分类的效果即可。感知机有一个致命的缺点，那就是只能对线性的样本进行分类，所以对于异或这样的简单非线性问题，感知机是解决不了的。

将多个感知机连接起来，使得所有的感知机输入都相同，但调节后每次只有一个感知机输出，就形成了一个多元的分类器。这种结构称为人工神经网络（ANN，Artificial Neural Networks）。在ANN中，感知机被重新命名为“神经元”。尽管将感知机组成了单层的网络，但其本质上仍然是线性变换，处理能力很有限。但是神经网络有一个好处，就是很容易增加层数(即神经网络的深度)。多层感知机可以摆脱早期离散传输函数的束缚，使用sigmoid或tanh等连续函数模拟神经元对激励的响应，在训练算法上则使用Werbos发明的反向传播BP算法，这也就是我们所说的神经网络。

从DNN按不同层的位置划分，DNN内部的神经网络层可以分为三类，输入层，隐藏层和输出层,如下图示例，一般来说第一层是输入层，最后一层是输出层，而中间的层数都是隐藏层。层与层之间是全连接的，也就是说，第i层的任意一个神经元一定与第i+1层的任意一个神经元相连。虽然DNN看起来很复杂，但是从小的局部模型来说，还是和感知机一样，即一个线性关系

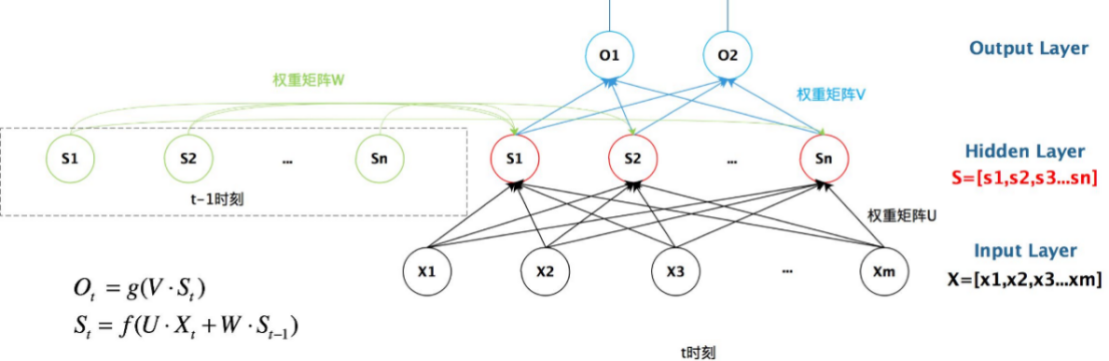
加上一个激活函数，隐藏层和输出层的输出值为，利用和感知机一样的思路，我们可以利用上一层的输出计算下一层的输出，也就是所谓的DNN前向传播算法。

传统神经网络对于时间序列信号，由于网络的单向无反馈连接方式使得网络只能处理输入信号所包含时间段的信号，对于该信号所包含时间段以外的信号在处理本段信号时常没有任何参考，而常见时间序列信号都和它所在时间段前后时间区间的背景信号有着密切的联系。

循环神经网络（RNN）存在神经元反馈连接，这种形式的连接使得网络能够以一种激励的形式存储最近时间段的输入数据信息(短时记忆)，而网络的这种潜在意义在实际应用方面着广泛的意义。

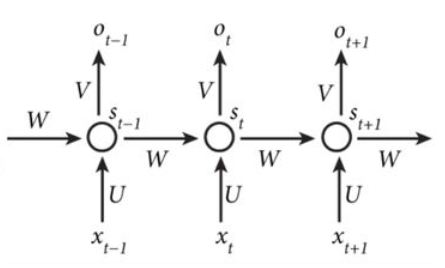
RNN由输入层、一个隐藏层和一个输出层组成（图3-6右）。其中，*X*是一个向量，它表示输入层的值；*S*是一个向量，它表示隐藏层的值；*O*也是一个向量，它表示输出层的值。*U*是输入层到隐藏层的权重矩阵，*V*是隐藏层到输出层的权重矩阵。

循环神经网络的隐藏层的值*S*不仅仅取决于当前这次的输入*X*，还取决于上一次隐藏层的值*S*。如图所示，权重矩阵*W*就是隐藏层上一次的值作为这一次的输入的权重。



**图3-6循环神经网络具体图**

其中，循环神经网络可按时间线展开，如图所示：



**图3-7按时间线展开的循环神经网络**

这个网络在*t*时刻接收到输入*Xt*之后，隐藏层的值是*St*，输出值是*Ot*。*St*的值不仅仅取决于*Xt*，还取决于*St-1*。循环神经网络中各节点参数可通过下列公式计算：

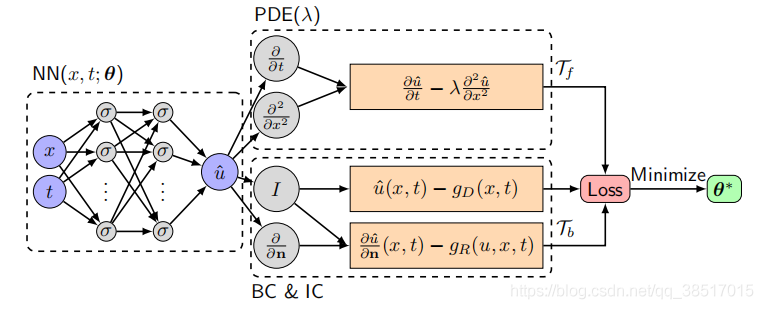
#### 内嵌物理知识神经网络（PINN）方法

大多数物理规律都可以表述为偏微分方程(PDE)的形式。偏微分方程，尤其是高阶偏微分方程难以求解析解，通常是采用各种方式逼近从而获得近似解。而神经网络的强大之处就在于其是万能近似器：如果神经网络具有至少一层非线性隐藏层，只要网络有足够数量的神经元，那么就可以充分地逼近任意一个在 的紧子集上定义的连续函数。

物理神经网络（PINN）本质是将物理方程作为限制加入神经网络中使训练的结果满足物理规律。物理神经网络通过把物理方程的迭代前后的差值加到神经网络的损失函数里面去，让物理方程也“参与”到了训练过程。这样，神经网络在训练迭代时候优化的不仅仅的网络自己的损失函数，还有物理方程每次迭代的差，使得最后训练出来的结果就满足物理规律了。

与传统数值求解算法区别：解PDEs的传统方法是基于网格的，如有限差分法、有限单元法等；神经网络使用自动微分方法，不需要网格。对于PINNs可以求解不同类型的微分方程，且对于求解逆问题只需改变极少的代码。

带有约束条件的偏微分方程如下：

对于初始条件，可以看成狄利克雷条件的特殊情况。 

**图3-8 PINN求解PDE的原理图**

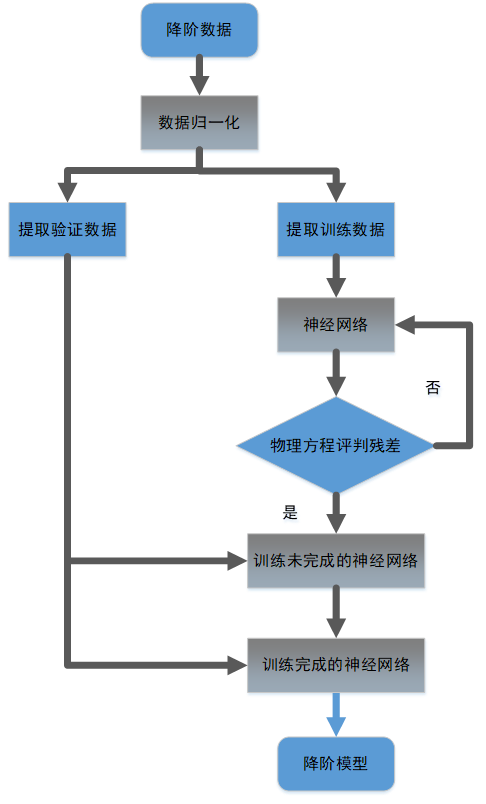
在PINN中，构建一个神经网络作为解决方案的替代，接受输入和输出一个向量，得约束神经网络使之满足PDE和边界施加的条件。在整个域上约束很难，所以在一些离散点（训练数据）上约束

、被称为残差点(residual points)；

 损失函数：

其中

物理神经网络的训练过程，也即通过最小化损失函数找到一个好的参数的过程，得到的参数就是满足物理约束的一组方程的解，在训练过程中，通常用基于梯度的优化器最小化损失函数。物理神经网络时变仿真技术路线如下：



**图3-9 PINN时变仿真技术流程图**

PINN时变仿真技术基于物理方程，部分物理场场景在神经网络训练之后，未必满足物理特性，因此需根据物理方程对神经网络的预测结果进行修正，神经网络迭代评判标准变为物理方程残差。PINN时变仿真技术在神经网络模型未完全训练结束之前，便可使用未训练数据进行验证及预测。

# 总结

本文主要研究了模型降阶技术应用于一维系统模型和三维CAE模型的融合问题。文中简要介绍了模型降阶的概况，总结了国内外对模型降阶技术在各领域的研究现状以及相关的工具、应用情况以及未来的发展趋势。基于上述信息，描述了模型降阶方法在数字化的大背景下，尤其是数字孪生应用相关的理论与方法进行研究，包括基于投影的模型降阶方法（奇异值分解、主成分分析、中心流形、L-S约化、模态综合、本征正交分解）和基于数据拟合的模型降阶方法（响应面、Kriging插值、神经网络等），从机理上系统性的对模型降阶理论进行了展开说明，为后续工程应用打下坚实的基础。