# MSET算法调研

### 介绍：

多元状态估计技术(Multivariate State Estimation Technique，MSET)是美国阿尔贡国立实验室(Argonne National Laborary)Kenny C. Gross等工程师开发的一种基于软件的过程与设备预警系统[]。早期MSET融合了模式识别技术和序贯概率比检验方法[]，获得了一系列美国开发专利[]和1998年R&D top100的奖项。2000年，阿尔贡的工程师进一步将MSET算法与支持向量机(SVM)方法融合[]。

田纳西大学核能实验室将MSET推广为非线性状态估计方法(Nonlinear State Estimation Technique，NSET)的方法[]。

### 基本思想：

Step 1：利用设备健康状态下各个参数的历史观测向量，设法构建出尽量覆盖设备正常工作空间的“过程记忆矩阵”；

Step 2：得到实际的运行观测数据后，MSET通过生成一个权重向量来衡量实际状态与正常状态的相似性，对当前实际运行状态做出最优估计；

Step 3：得到实际运行观测数据与状态估计的残差，利用序贯概率比检验等残差分析手段进行预测诊断。

### 模型推导：

MSET本质上是多元回归参数的最小二乘法的推广。

记表示第个参数第个正常运行状态下的观测值，

则过程记忆矩阵可表示为：

 （1）

记，即中的每一列代表一个历史观测向量。

一个新的观测向量输入MSET时，MSET生成一个权值向量，输出当前运行的状态估计向量为：

 （2）

对比MSET的输入与输出，得到观测残差为：

 （3）

如何得到权重？利用最小二乘法使残差向量的平方和最小化，即可得到权重的最小二乘估计：

 （4）

于是当前运行状态估计为：

 （5）

若要得到当前某一参数的状态估计，则：

 （6）

观察（5）式，当前运行状态的估计向量通过对历史的观测向量加权求和得到，而权重公式（4）中中的元素为历史观测向量间的点积运算，而为历史观测向量与当前观测的点积运算，这表明权重由当前观测向量与历史观测的相似程度决定。考虑到非线性状态估计方法具有更直观的物理意义，并且避免变量之间的相关性造成不可逆的情况，MSET欲选择不同的算子衡量向量间的相似性。

备选的相似性算子可以选取欧氏距离、相对熵、线性相关系数等。以欧氏距离为例，对向量，记其欧式距离运算符为：。即令（4）式变为：

 （7）

于是对于新的观测变量（5）式变为，

 （8）

当前某一参数的状态估计（6）式变为：

 （9）

式（8）的物理意义为：新的观测向量与过程记忆矩阵中的个历史观测向量两两之间计算欧氏距离来反应它们之间的相似性，假设与过程记忆中的历史观测向量最相似，则两者的欧式距离最小，其对应的 权值最大，对预测结果的贡献也最大。

MSET模型的精度可以根据观测残差（3）来衡量。设备正常工作时，新的观测向量应与过程记忆矩阵中的历史观测向量距离较近，MSET将凭借“记忆”肯定此时的工作状态；而当设备工作或过程运行出现异常时，观测向量将偏离过程记忆矩阵所覆盖的正常工作空间，此时MSET将无法准确预测状态，产生较大预测残差，辅助残差分析手段在必要时给出预警。

### 残差分析与预警

记某段时间内，MSET模型的预测残差序列为：

 (11)

其中。

1. 滑动窗口残差统计方法

对于该序列，取一个宽度为的滑动窗口。对窗口内连续个残差计算其均值和标准差：

 （12）

通过合理选择滑动窗口的宽度，既能及时迅速地反映出残差统计特性的连续变化，又能有效消除随机因素的影响，降低误报警的几率。设备正常工作时，残差均值接近零，标准差较小。出现故障隐患时，残差统计特性发生变化，故可通过对两个统计量的监控来实现报警。报警阈值可由运行人员由经验确定。

在监测过程中，可使用参数估计方法给出残差均值和标准差的区间估计，以降低误报率。假设残差服从均值和方差未知的正态分布，则根据正态分布置信区间估计的方法，可得出残差均值和标准差的置信度为的置信区间分别为：

 （13）

 （14）

其中：表示自由度为的分布的分位数。若滑动窗口残差均值或标准差的置信区间超过阈值时，则发出设备故障报警信息。

2）序贯概率比检验

序贯概率比检验(Sequential probability ratio test，SPRT)是由著名统计学家Wald提出的一种统计决策方法。不同于固定样本检验方法，即用于检验的样本大小是一个固定的常数，序贯检验不一劳永逸的决定检验样本大小，而是进行阶段性的抽样、检验，直到足以得出满意的检验结果为止。的基本框架是通过比较监控过程的真实传感器数据与估计传感器数据的误差来分析对象过程中的序列观测值。

给定原假设

### 过程说明

构建MSET算法模型过程需注意以下几点：

1. 用来生成历史观测向量集合K的历史数据应满足
2. 涵盖了一段足够长的运行时间；
3. 每组数据都表达了设备对象的一个正常状态；
4. 满足每一组采样值中各个变量的同时性，必须是同一时刻的采样值；
5. 涵盖了机组在不同季节、不同负荷下的运行情况。

 （10）

其中表示不同运行工况下历史观测向量总数，表示设备变量数。

1. 数据归一化处理

利用去均值，标准化方法将实际测量值映射到[0,1]区间，并消除变量量纲。

1. 过程记忆矩阵的构造

通过历史数据矩阵K得到矩阵的过程中，为使中选取的观测向量尽量覆盖设备正常工作空间，对归一化后的矩阵K的每一个变量，将[0,1]区间等分为100份，以0.01为步长从集合K的所有观测向量中查找出若干个观测向量加入矩阵。对于某些重要的变量，可取更小的步长。使用此方法构造过程记忆矩阵，能够将组成观测变量的个变量的不同测量值对应的历史记录选入矩阵中，从而使其较好的覆盖设备正常工作空间。