统计推断在数模转化系统中的应用

组号：62　　姓名：何流(组长) 学号：5110309595

摘要：本报告基于对科创3A的实验数据的研究。采用多种拟合并结合统计数学的方法，对实验课程“科创3A”中制作的DC-DC电子装置的D-U非线性特性进行了研究。使用Matlab，通过模拟退火算法，发现可有效地取七个点来达到较高的拟合度程度。从而有效的降低计算的复杂度，求出满足评价函数的局部最优解，并且对定标方法的准确性用度量函数进行评价。

关键词： 模拟退火算法 ; 定标

Application of Statistic Inference in A/D & D/A Conversion System

ABSTRACT: The paper uses the Simulation Annealing algorithm and statistic mathematics to research on the characteristic of D-U in the circuit of the course Sience and technology Innovation part 3A.In this way, we found 7 data and fit a cubic polynomial using Matlab software to predict other 46 data effectively. As a result, we can reduce the complexity of the calculations, get the optional result partal of the data,and evaluate the accuracy of the calibration.

Key words: Simulation Annealing Algorithm ; Calibration

1. 问题的提出

统计推断是一类重要的统计数学方法，在各种自然科学和社会科学研究、工程技术等领域都有较广泛的应用。

在本次课程中，将使用统计推断方法，研究一个模数、数模转换系统的特性，进而对系统性能进行优化设计。实验中，原始数据来源于科创 3A 学生实验数据上传汇总。

1.1 任务概述

科创 3A 中，系统输入为 PWM 信号占空比 D，输出为开关电源输出电压 U，D-U 函数关系可以表示为论域内的连续变化曲线。在存在一定误差的条件下，通过实际测量 7 个数据，能够有效地推断出 D-U 函数关系。

1.2 设计思路

老师提供的数据总表中共有 469 个小组数据，我们需要根据每组数据中的 7 个特征点，推断出函数关系。值得强调的是，事先并没有筛选或者剔除“不符合要求的样本”，尽管从数据查看工具中我们发现很多线性的，甚至是看来不合理的样本，而且这样的情况并不在少数，在某种条件的影响下确实出现了类似线性的分布。为了更加客观真实反映这种结果，我们并没有筛选数据，尽管这种做法会使得最终的结果误差偏大，但是牺牲了一部分准确性，换取了更加广泛的适用范围。

2. 研究目的和背景

单片机开环控制的 DC-DC 开关电源系统由单片机输出占空比(D)可调的 PWM 信号，此信号经过中间模块的处理得以控制开关电源的输出电压(U)，占空比的改变对应于输出电压的变化（输出电压范围 5.0—10.0V，步进 0.1V），存在占空比(D)—输出电压(U)特性关系，即为研究对象。

考虑到工业化生产下的成本、效率因素，针对该系统的某一实例，实测其 51 个实用设定点输出电压对应的占空比是不经济的。因而考虑跳点测定，在保证所要求精度的前提下，

通过测定 7 组电压值（即特征点）对应的占空比数据来推断整体的占空比(D)—输出电压(U)特性关系[1]。

研究给定的样本的这一特性关系，课题最终目标是找到七个特征点 ,使得根据这七个点拟合的数据按照提供的评价函数计算后,所有样本的花费尽可能低。

研究方法可以总结为四步:

1.给定原件实测的数据；

2.确定合理的拟合方法；

3.选取合理的最优算法；

4.得到适用度最高样本。

3. 拟合方法的讨论

所谓拟合，就是用最接近的函数的关系去表征数据点所反映的 D—U 关系。拟合方法选取是否得当直接影响了实验结果能够具有良好的适应度。不同的拟合方法有不同优劣。下面初步就三次多项式拟合法和三次样条插值法的实现过程作介绍。

3.1 三次多项式拟合

（1）使用三次多项式对选取的七个特征点组合即可能解进行拟合。样本为 $S=\{S_1, S_2, \cdots S_7\}$，如图 1 所示。

$$u = a_1 d^3 + a_2 d^2 + a_3 d + a_4 \tag{1}$$

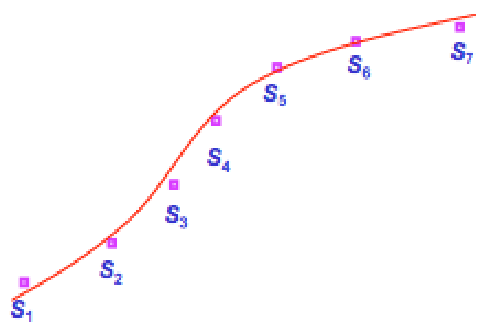

图 1 三次多项式曲线拟合[2]

（2）曲线的拟合度。将去掉 1 和 53 号点的其余 51 个点的占空比 D 代入上述三次多项式计算出相应 51 个电压值，通过评价函数得到评价分值。

（3）现最优化过程，即不断的寻找存在的可能解，继续前两步骤。每次得到可能解，通过评价，做出一定概率的接受或者舍弃，实现当前解的优化，直至达到终止条件。

3.2 三次样条插值拟合法

（1）三次样条插值对确定一个可能解，即 7 个特征点组合 $S=\{s_1, s_2, \cdots, s_7\}$ 进行拟合。具体方法为：首先对于非两端点，以四个连续点确定一条三次曲线，但仅在中间两点之间用该三次曲线表示，以此类推，所有非两端点之间均有三次曲线。两端点由端点处三个点用二次曲线拟合，如图 2 所示。

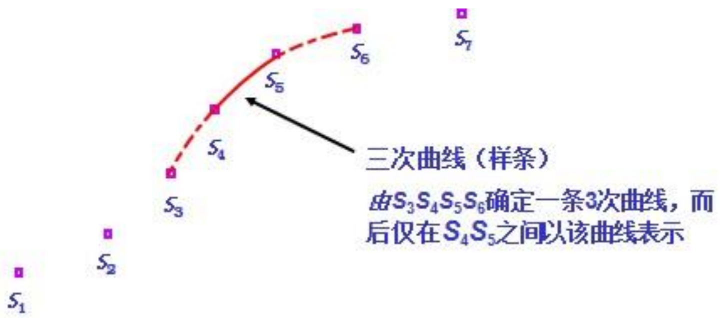

图 2 三次样条插值曲线拟合[2]

（2） 评估该曲线的拟合度。将去掉 1 和 53 号点的其余 51 个点的占空比 D 代入得到三次样条插值后相应的 51 个电压值，通过评价函数得到评价分值。

（3） 现最优化过程，即不断的寻找存在的可能解，继续前两步骤。每次得到可能解，通过评价，做出一定概率的接受或者舍弃，实现当前解的优化，直至达到终止条件。

4. 解决问题

4.1 确定拟合方法

通过分析比对，我们确定对每组 51 个数据，只选取其中的 7 个点进行拟合并计算适用度。每组数据选取的七个点的下标一致。考虑到难易程度，时间和效率的问题，我们最终确定使用模拟退火算法，拟合过程同时使用三次样条插值法和三次多项式拟合法。

4.2 拟合方法实现[3]

可以写出伪代码：

（1）读入老师提供的 469 组数据。

（2）随机函数随机排序并选取前七个点，再次排序，得到随机的几个点。

（3）用模拟退火算法作为大循环。（注：关于模拟退火算法见特别说明）

（4）在循环内前部分加入随机变化一个点生成新的七个点组合。并用拟合法得到每个$U_{i,j}-U'_{i,j}$的值。

（5）循环中间部分加入计算 51 个点的评价函数的花费（cost）。

（6）循环的后半部分决定是否接受这七个点。若分值小于最小花费，则接受；若大于，则以 exp((cost-cost_save)/Tk)的概率接受。其中 Tk 为温度。

（7）输出最优解。

4.3 算法框图

算法框图如图 3 所示。

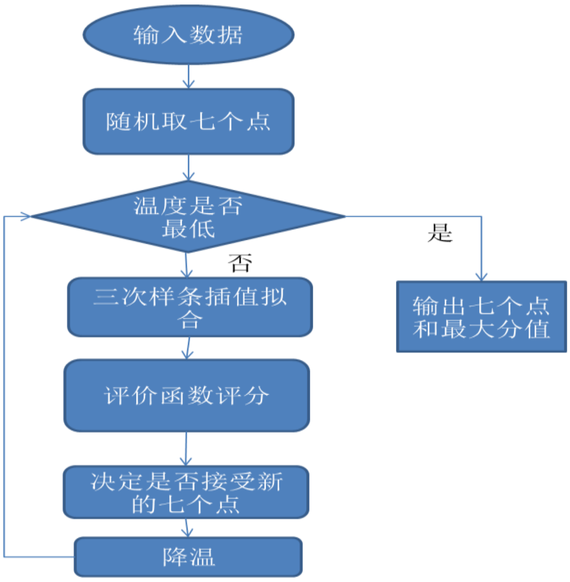

图 3 模拟退火算法框图

4.4 特别说明

（1）三次拟合与三次样条插值拟合均用了 matlab 内置函数[4]。分别是三次拟合 p=polyfit(X,Y,3)；三次样条插值拟合:interp1(X,Y,D(i,1:51),'spline')。使用两种拟合方法的目的是探究确定那种方法有更好的适用效果。

（2）模拟退火算法的实现过程中仅运用了“降温”过程就得到了花费小于 90 的较优解，故没有再加入“恒温”的“如果多次无最优解，则降温”的过程。但为何我们可以运行出较好的结果引发了我们的思考。我们发现，寻找可行解过程中，使用传统的“临近代换”，也就是爬山算法，易导致“局部最优解”出现，以图 4 为例，C 临近代换到 A 得到局部最优解，E 劣于 A，所以 A 为结果，而我们代换一个点采用的是随机代换，这就使得“局部最优解”的出现的几率大大降低。

（3） 对于模拟退火算法及其随机代换，为了让读者能够形象理解，我们做一个特别说明。模拟退火其实是一种贪心算法，但是它的搜索过程引入了随机因素。模拟退火算法以一定的概率来接受一个比当前解要差的解，因此有可能会跳出这个局部的最优解，达到全局的最优解。以图 4 为例，A、B 均为局部最优解，B 要优于 A，模拟退火算法在搜索到局部最优解 A 后，会以一定的概率接受到 E 的移动。也许经过几次这样的不是局部最优的移动后会到达 D 点，于是就跳出了局部最大值 A，达到 B 甚至更优的解。打个比方来说，兔子喝醉了，它随机跳了很长时间。期间可能跳向高处，可能跳入平地，但它渐渐清醒了并朝最高方向跳去，这就是模拟退火。

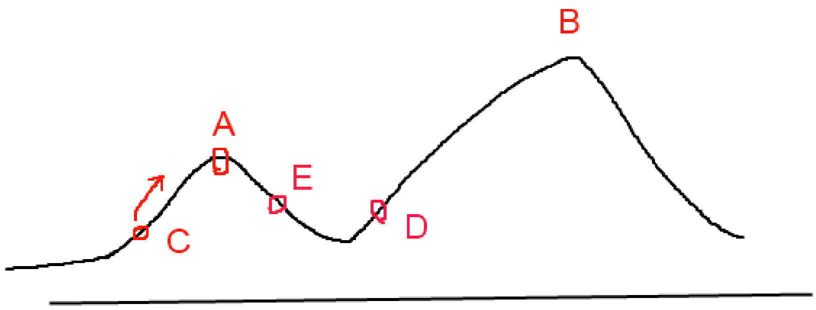

图 4 模拟退火算法说明

5. 结果分析

5.1 程序运行结果图示

表 1 三次样条插值拟合法

编号	运行结果							cost	时间
1	5	7	24	29	43	50	51	91	20.24563s
2	1	8	26	31	34	42	48	87	18.07662s
3	3	7	14	20	26	47	48	87	20.52947s
4	1	17	24	32	35	47	50	86.5	25.43329s
5	6	8	27	31	32	36	46	85	18.97072s

表 2 三次拟合法

编号	运行结果							cost	时间
1	3	10	22	32	38	39	48	126	85.403314s
2	12	16	25	28	39	47	49	124.5	87.065907s
3	5	15	16	34	44	45	49	123	86.223599s
4	3	14	17	25	27	37	47	122.5	87.576696s
5	2	9	12	16	21	34	47	119.5	84.776218s

5.2 两种方法有效性对比

对两种拟合方法分别运行了 5 次。由运行结果示例可以很清楚地看出来，使用三次样条插值法相比于三次多项式拟合法有两处优势。1、使用前者，花费明显要低于后者，后者花费至少在 115 以上。2、使用前者 5 次运行所花费的时间少于使用后者，可能是提供的 matlab 自带拟合函数内部运行时间问题有关，因为程序的其他部分基本上都是一样的。我认为，D-U 关系可以使用更高次数的多项式得到更加精确地拟合，而三次样条插值一定程度上可以理解为更高次数的多项式，拟合起来会更加细腻，结果也就更加准确 。

6. 拓展探究

6.1 三次拟合和四次拟合的对比分析

6.1.1 分析原因

我们小组在基础探究部分发现，三次拟合相比较于三次样条插值来说带入评价函数得分普遍较低。于是，我们猜想四次拟合相对于三次拟合会有较高的拟合度。

6.1.2 程序核心部分

三次拟合部分变为：

```
[p,v]=polyfit(X,Y,4);
```

由于 matlab 含有拟合函数 polyfit，故改变函数的参数即可。

6.1.3 程序运行结果表

表 3　四次拟合法

编号	运行结果							分数	时间
1	3	12	21	29	35	45	52	103	108.350684
2	4	11	23	30	41	50	52	113	145.702875
3	3	12	22	32	40	49	51	108	106.381317
4	2	11	22	30	40	48	52	107	131.848665
5	3	13	21	31	41	48	52	111	126.338227

6.1.4 结果分析

由表 5.2 和表 6.1 对比可知，四次拟合的拟合度要高于三次拟合的拟合度。假设成立。但四次拟合用时远远大于三次拟合用时，故比较之下均不如三次样条插值拟合法拟合效果好。

7. 课题结论

由程序运行结果以及对结果的分析，三次插条拟合比三次拟合，或是四次拟合效果要好，考虑到效率、效果等因素，我们组得出结论。我们推荐的方案为选取[6　8　27　31　32　36　46] 七点，运用三次样条插值拟合，能够在兼顾效率效果的同时，满足课题要求，有很好的适用效果。

8. 参考文献

[1]袁炎.　“统计推断”课程设计的要求 V2.1 2014-11-22updated　ftp://202.120.39.248

[2]袁炎.　讲义和参考资料 2014-10-12updated　ftp://202.120.39.248

[3]网络资源.　http://baike.baidu.com/view/18185.htm

[4]网络资源.　China-pub.com　《matlab 教程》

9. 附录（程序代码）

三次拟合

```
KCdata=xlsread('20141010dataform.csv');%¶ÁÈ¡Êý¾Ý
D=KCdata(1:2:end,1:end);
U=KCdata(2:2:end,1:end);%X¶ÁÈëD£¬Y¶ÁÈëU

A=randperm(51);
B=sort(A(1:7)); %Ëæ»úÔÚ1-51ÖÐÑ¡È¡7¸öµã²¢ÅÅÐò

B_min=B;
```

```
cost_min=1000;
cost_save=1000;
cost_process = zeros(1,303);
ii=1;
Tf=0.01;
Tk=100;

tic;
while Tk>Tf
    remain=setdiff(A,B);%51¸öµãÖÐÃ»±»Ñ¡È¡µÄÊ£ÏÂ44¸öµã
    E=remain(randperm(44));
    F=randperm(7);
    S=B;
    S(1,F(1))=E(1,F(1)+1);
    S=sort(S);%Ëæ»úÌæ»»ÔÀ´7¸öµãÖÐµÄÒ»¸ö,²¢ÅÅÐò
    M=zeros(469,51);
    for i=1:469
        X=D(i,S);
        Y=U(i,S);
        p=polyfit(X,Y,3);

M(i,:)=U(i,1:51)-p(1)*D(i,1:51).^3-p(2)*D(i,1:51).^2-p(3)*D(i,1:51).^
1 -p(4);%3´ÎÏßÐÔ·½³Ì£¬mÊÇÎó²îÖµ
        Q=12;
        errabs=abs(M(i,:));
        le0_5=(errabs<=0.5);
        le1_0=(errabs<=1);
        le2_0=(errabs<=2);
        le3_0=(errabs<=3);
        le5_0=(errabs<=5);
        g5_0=(errabs>5);

sij=0.5*(le1_0-le0_5)+1.5*(le2_0-le1_0)+6*(le3_0-le2_0)+12*(le5_0-le3
_0)+25*g5_0;
        si=sum(sij,2)+Q*ones(469,1)*7;
        cost=sum(si)/469;
         %¼ÆËãÆ½¾ùµÃ·Ö
    end

    if cost<cost_min
        cost_min=cost;

cost_save=cost;%Èô¼ÆËã³öµÄÆ½¾ùÏûºÄµÍÓÚ×îµÍÏûºÄ£¬Ôò½«´ËÏûºÄ¼Çª×îµÍÏûº
ÄµÃ·Ö£¬±£´æ´ËÏûºÄ
```

```
        B_min=S;
        B=S;
    elseif
rand>exp((cost-cost_save)/Tk)%Èô´óÓÚ×îµÍÏû°Ä£¬ÔòÒÔ´Ë¸ÅÂÊ½ÓÊÜ
        cost_save=cost;
        B=S;
    end
    cost_min;
    cost;
    Tk=Tk*0.97;%½µÎÂ
    cost_process(1,ii)=cost;
    ii=ii+1;
end
cost_min;
B_min;
toc;
plot(cost_process);
```

四次拟合

```
KCdata=xlsread('20141010dataform.csv');%¶ÁÈ¡Êý¾Ý
D=KCdata(1:2:end,1:end);
U=KCdata(2:2:end,1:end);%X¶ÁÈëD£¬Y¶ÁÈëU

A=randperm(51);
B=sort(A(1:7)); %Ëæ»úÔÚ1-51ÖÐÑ¡È¡7¸öµã²¢ÅÅÐò

B_min=B;
cost_min=1000;
cost_save=1000;
cost_process = zeros(1,303);
ii=1;
Tf=0.01;
Tk=100;

tic;
while Tk>Tf
    remain=setdiff(A,B);%51¸öµãÖÐÃ»±»Ñ¡È¡µÄÊ£ÏÂ44¸öµã
    E=remain(randperm(44));
    F=randperm(7);
    S=B;
    S(1,F(1))=E(1,F(1)+1);
    S=sort(S);%Ëæ»úÌæ»»ÔÀ´7¸öµãÖÐµÄÒ»¸ö,²¢ÅÅÐò
    M=zeros(469,51);
    for i=1:469
```

```
        X=D(i,S);
        Y=U(i,S);
        p=polyfit(X,Y,4);

M(i,:)=U(i,2:52)-p(1)*D(i,2:52).^4-p(2)*D(i,2:52).^3-p(3)*D(i,2:52).^
2-p(4)*D(i,2:52).^1-p(5);%3´ÎÏßÐÔ·½³Ì£¬mÊÇÎó²îÖµ
        Q=12;
        errabs=abs(M(i,:));
        le0_5=(errabs<=0.5);
        le1_0=(errabs<=1);
        le2_0=(errabs<=2);
        le3_0=(errabs<=3);
        le5_0=(errabs<=5);
        g5_0=(errabs>5);

sij=0.5*(le1_0-le0_5)+1.5*(le2_0-le1_0)+6*(le3_0-le2_0)+12*(le5_0-le3
_0)+25*g5_0;
        si=sum(sij,2)+Q*ones(469,1)*7;
        cost=sum(si)/469;
         %¼ÆËãÆ½¾ùµÃ·Ö
    end

    if cost<cost_min
        cost_min=cost;

cost_save=cost;%Èô¼ÆËã³öµÄÆ½¾ùÏûºÄµÍÓÚ×îµÍÏûºÄ£¬Ôò½«´ËÏûºÄ¼Ç×÷×îµÍÏûº
ÄµÃ·Ö£¬±£´æ´ËÏûºÄ
        B_min=S;
        B=S;
    elseif
rand>exp((cost-cost_save)/Tk)%Èô´óÓÚ×îµÍÏûºÄ£¬ÔòÒÔÒ»¶¨µÄ¸ÅÂÊ½ÓÊÜ
        cost_save=cost;
        B=S;
    end
    cost_min;
    cost;
    Tk=Tk*0.97;%½µÎÂ
    cost_process(1,ii)=cost;
    ii=ii+1;
end
cost_min;
B_min;
toc;
plot(cost_process);
```

三次样条插值

```
KCdata=xlsread('20141010dataform.csv');%¶ÁÈ¡Êý¾Ý
D=KCdata(1:2:end,1:end);
U=KCdata(2:2:end,1:end);%X¶ÁÈëD£¬Y¶ÁÈëU

A=randperm(51);
B=sort(A(1:7)); %Ëæ»úÔÚ1-51ÖÐÑ¡È¡7¸öµã²¢ÅÅÐò

B_min=B;
cost_min=1000;
cost_save=1000;
cost_process = zeros(1,303);
S_process = zeros(303,7);
ii=1;
Tf=0.01;
Tk=100;

tic;
while Tk>Tf
    remain=setdiff(A,B);%51¸öµãÖÐÃ»±»Ñ¡È¡µÄÊ£ÏÂ44¸öµã
    E=remain(randperm(44));
    F=randperm(7);
    S=B;
    S(1,F(1))=E(1,F(1)+1);
    S=sort(S);%Ëæ»úÌæ»»ÔÀ´7¸öµãÖÐµÄÒ»¸ö,²¢ÅÅÐò
    M=zeros(469,51);
    for i=1:469
        X=D(i,S);
        Y=U(i,S);
        M(i,:)=U(i,1:51)-interp1(X,Y,D(i,1:51),'spline');%Èý´ÎÑùÌõ²åÖµ
        Q=12;
        errabs=abs(M(i,:));
        le0_5=(errabs<=0.5);
        le1_0=(errabs<=1);
        le2_0=(errabs<=2);
        le3_0=(errabs<=3);
        le5_0=(errabs<=5);
        g5_0=(errabs>5);

sij=0.5*(le1_0-le0_5)+1.5*(le2_0-le1_0)+6*(le3_0-le2_0)+12*(le5_0-le3
_0)+25*g5_0;
        si=sum(sij,2)+Q*ones(469,1)*7;
        cost=sum(si)/469;
         %¼ÆËã¼Æ½¾ùµÃ·Ö
```

```
    end

    if cost<cost_min
        cost_min=cost;

cost_save=cost;%Èô¼ÆËã³öµÄÆ½¾ùÏû°ÄµÍÓÚ×îµÍÏû°Ä£¬Ôò½«´ËÏû°Ä¼ÇÎª×îµÍÏû°
ÄµÃ·Ö£¬±£´æ´ËÏû°Ä
        B_min=S;
        B=S;
    elseif
rand>exp((cost-cost_save)/Tk)%Èô´óÓÚ×îµÍÏû°Ä£¬ÔòÒÔ´Ë¸ÅÂÊ½ÓÊÜ
        cost_save=cost;
        B=S;
    end
    Tk=Tk*0.97;%½µÎÂ
    cost_process(1,ii)=cost;
    S_process(ii,:)=S;
    ii=ii+1;
end
toc;
B_min;
plot(cost_process);
```