中 国 科 学 院 上 海 天 文 台 年 刊
1998 年第 19 期 ANNALS OF SHANGHAI OBSERVATORY ACADEMIA SINICA No. 19, 1998

SVD 算法简介与模拟数据检验*

李 金 岭
(中国科学院上海天文台,上海 200030)

提 要

简单介绍 SVD(奇异值分解)算法的数学基础、背景和解算过程,并基于模拟数据检验结果分析,说明了 SVD 算法的特性和功能。SVD 算法不但能够解决最小二乘方法所能解决的各种问题,而且还能够对后者所不能解决的部分问题给出判据,因而有利于问题的深入分析和解决。推荐对 SVD 算法,予以关注和应用。

主题词:数据处理 — 参数 — 最小二乘法 — SVD 算法
分类号:O241.2

1 引 言

在数据分析建模中,如果独立方程的个数小于未知数个数,或模型中某一参数是其他一个或几个参数的函数等等,都会导致法方程系数矩阵出现奇异现象,即矩阵是非满秩的。此时,系数矩阵的逆是不存在的,利用通常的最小二乘法不能获得稳定解。SVD(Singular Value Decomposition)算法能够对系数矩阵是否奇异作出判断,并基于物理背景的考虑对正解进行估计[1]。如果矩阵是满秩的,则 SVD 算法将能够给出与通常的最小二乘方法一致的解。本文介绍 SVD 算法的原理,并通过模拟数据检验,对 SVD 算法的特点和功能予以说明,以期对该算法予以关注和运用。

2 SVD 算法原理

SVD 算法的数学基础为,对于任何行数 M 大于或等于列数 N 的 $M \times N$ 矩阵 A,均可分解为 $M \times N$ 的列正交矩阵 U、元素大于或等于零的 $N \times N$ 的对角矩阵 W 和 $N \times N$ 正交矩阵 V 的转置的乘积,形式地表示为

$$(A) = (U) \cdot \begin{bmatrix} W_1 & & & 0 \\ & W_2 & & \\ & & \cdots & \\ 0 & & & W_N \end{bmatrix} \cdot (V^T) = UWV^T \tag{1}$$

1997 年 1 月 28 日收到。
* 上海市科技发展基金资助项目

无论矩阵 A 是否奇异，这样的分解总能实现，而且具备唯一性。(1)式的数学证明此处从略，其各种实现程序中最具代表性的一个为 SVDCMP，对于任意的输入矩阵 A ，可以返回矩阵 U，W 和 V 。该程序主要应用 Householder 变换（通过 $N-1$ 次正交变换将 $N \times N$ 的对称矩阵化为三角矩阵）和 QR 算法（将实数矩阵分解为正交矩阵和三角矩阵的乘积），性能十分稳定[1~4]。

如果 A 为 $N \times N$ 方阵，则 U，W 和 V 亦为方阵。由于 U 和 V 为正交矩阵，W 是对角矩阵，则 $U^{-1}=U^{\mathrm{T}}$，$V^{-1}=V^{\mathrm{T}}$ 和 $[W_j]^{-1}=\left[\dfrac{1}{W_j}\right]$，进而，

$$A^{-1}=V\cdot\left[\frac{1}{W_j}\right]\cdot U^{\mathrm{T}} \tag{2}$$

此时，若 W 的某元素 W_j 为 0 或很小，则 A 为奇异的或近于奇异，$\dfrac{1}{W_j}$ 成为不确定量。可见，通过 W_j 的取值可以判断矩阵 A 是否奇异。

一般地，设方阵 A 为矢量 $\vec{x}$ 至矢量 $\vec{b}$ 的映射，即

$$A\cdot\vec{x}\Rightarrow\vec{b} \tag{3}$$

若矩阵 A 是满秩的，则通过(2)式可唯一确定(3)式的解矢量 $\vec{x}$；若 A 为奇异或降秩阵，则(3)式存在无穷解或无解。事实上，(1)式对 A 构造了相互正交的两组基矢，一组为 A 的最大满秩子阵 B 的基矢，由 $W_j \neq 0$ 所对应的矩阵 U 的所有列矢量构成，另一组为 A 的最大降秩子阵 C 的基矢，由 $W_j=0$ 所对应的矩阵 V 的所有列矢量组成。显然，$B\cdot\vec{x}$ 为矢量 $\vec{b}$ 或其部分分量，而 $C\cdot\vec{x}$ 在矢量 $\vec{b}$ 的空间为零。此时，

$$\text{若 } B\cdot\vec{x}=\vec{b}\text{，则(3) 式有无穷解} \tag{4}$$

$$\text{若 } B\cdot\vec{x}\neq\vec{b}\text{，则(3) 式无解} \tag{5}$$

所谓 SVD 算法，即是经(1)式分解之后，若 $W_j=0$ ，则令 $\dfrac{1}{W_j}=0$ ，亦即，

$$\begin{cases}\vec{x}=V\cdot\left[\dfrac{1}{W_j}\right]\cdot(U^{\mathrm{T}}\cdot\vec{b})\\[2ex] \text{若 } W_j=0\text{，则令}\dfrac{1}{W_j}=0\end{cases} \tag{6}$$

这样处理的背景为如果 W_j 很小，以致于被计算机的截断误差所掩盖，W_j 将成为不确定量而失去意义，其倒数非常之大，进而将影响甚至破坏参数解的稳定性。在这种情况下，即使对观测数据是绝对拟合，参数解亦将失去物理意义。通过令 $\dfrac{1}{W_j}=0$ ，排除了对应参数不确定性的影响，进而确保了其它参数解的稳定性。可以证明[1,5,6]，在(4)式情况下，(6)式为(3)式的无穷解中模最小的解矢量，即 $\vec{x}\cdot\vec{x}=\min$；在(5)式情况下，(6)式为(3)式的所有可能解矢量中最能减小 $|A\cdot\vec{x}-\vec{b}|$ 的矢量，因而是一种广义的最小二乘解。

若用 U_j 和 V_j($j=1,\cdots,N$)分别表示矩阵 U 和 V 的列矢量，则(6)式可以表示为

$$\begin{cases}\vec{x}=\displaystyle\sum_{j=1}^{N}\left[\frac{U_j\cdot\vec{b}}{W_j}\right]\cdot V_j\\[2ex] W_j=0\Rightarrow\dfrac{1}{W_j}=0\end{cases} \tag{7}$$

(7)式说明，矢量 $\vec{x}$ 是矩阵 V 的列矢量的线性组合，系数则为矩阵 U 的列矢量与矢量 $\vec{b}$ 的点积。实际上，V_j 是矢量 $\vec{x}$ 的误差椭圆的主轴，进而可得解矢量 $\vec{x}$ 的方差 $\sigma^2(x_i)$ 和协方差

$Cov(x_i, x_k)$ 为[1,6,7]

$$\sigma^2(x_i) = \sum_{j=1}^{N}\left[\frac{U_{ij}}{W_j}\right]^2 \tag{8}$$

$$Cov(x_i, x_k) = \sum_{j=1}^{N}\left[\frac{V_{ij}V_{kj}}{W_j^2}\right] \tag{9}$$

在方程个数 M 大于未知数个数 N 的情况下，仍以(1)式为数学基础，并可直接应用(6)至(9)式。若 M 小于 N，则用零元素将矩阵 A 补齐为 $N \times N$ 的方阵，并将 $\vec{b}$ 作对应处理，即增加 $N-M$ 个观测方程，然后应用(6)至(9)式求解。若拟合变量中存在相关性，即矩阵 A 是列奇异的，SVD 算法通过将相应的 $\frac{1}{W_j}$ 置为零进而获得对正解的估计。如果利用通常的最小二乘法处理后两种情况，则很难得到稳定解。总之，在对观测数据建模的过程中，若系数矩阵是满秩的，SVD 算法可以得到通常的最小二乘意义上的解矢量。若 A 为奇异或降秩阵，SVD 算法则可以通过(6)式“组建”一个矢量 $\vec{x}$。此时，$\vec{x}$ 并非完全满足 $A \cdot \vec{x} \Rightarrow \vec{b}$，但它却是在可利用条件下最接近正解的一个矢量。可见，SVD 算法能够完全代替通常的最小二乘(LS)解法，并能在系数矩阵接近或出现奇异的情况下对解矢量进行估计。SVD 算法的缺点是所需内存较大和计算速度较慢，对现代计算设备而言，其缺点与其优点相比是不显著的。

3 模拟数据检验

对于

$$y = ax + bx^2 + cx^3 \tag{10}$$

取 $a=1, b=3, c=5$。如表1所示，对各 x_i 由(10)式计算其理论值 y_{ci}，并附加正态白噪声 $\varepsilon_i \sim (0,1)$(以示观测误差)，得模拟观测数据 y_i。以下分三种情况进行讨论。

表 1 模拟数据

Table 1 Analog data

i	x_i	y_{ci}	ε_i	y_i
1	1.0	9.00	−0.69	8.31
2	3.0	165.00	1.49	166.49
3	5.0	705.00	−1.83	703.17
4	7.0	1869.00	1.19	1870.19
5	9.0	3897.00	0.47	3897.47
6	11.0	7029.00	0.85	7029.85
7	13.0	11505.00	0.73	11505.73
8	15.0	17565.00	1.00	17566.00
9	17.0	25449.00	−0.55	25448.45
10	19.0	35397.00	1.19	35398.19
11	21.0	47649.00	−1.01	47647.99
12	23.0	62445.00	−0.89	62444.11
13	25.0	80025.00	−0.21	80024.79
14	27.0	100629.00	0.33	100629.33
15	29.0	124497.00	−0.53	124496.47
16	31.0	151869.00	−0.73	151868.27
17	33.0	182985.00	−2.01	182982.99

(1) 模型合适且方程个数大于未知数个数的情况

直接利用(10)式分别由 SVD 算法和 LS 方法对表 1 中模拟数据进行拟合,结果如下:

SVD 算法:$a = 1.08 \pm 0.12$, $b = 3.00 \pm 0.01$, $c = 5.00 \pm 2 \times 10^{-4}$,

SL 方法:$a = 1.08 \pm 0.12$, $b = 3.00 \pm 0.01$, $c = 5.00 \pm 2 \times 10^{-8}$。

可见,二者所得结果一致,且与理论值($a = 1$, $b = 3$, $c = 5$)相符。

(2)模型参数不独立的情况(对应于(4)式)

用 $y = ax + bx^2 + cx^3 + d(x + x^2)$ 分别由 SVD 算法和 LS 方法对表 1 模拟数据进行拟合,结果如下:

SVD 算法:$a = -0.28 \pm 0.09$, $b = 1.64 \pm 0.05$, $c = 5.00 \pm 0.00$, $d = 1.36 \pm 0.04$,

SL 方法:计算失败

与理论值相比,SVD 算法仅提供了对正解的估计,但是,($a + d$)和($b + d$)分别与 a 和 b 的理论值非常接近。这并非巧合,而是可以从理论上作出证明的。此例中,W 的四个对角元素依次为 61140.3, 522.0, 9.4 和 0.0。这已经明确显示了模型中第四个拟合参量($x + x^2$)存在问题。可见,通过考察 SVD 算法中 W_j 的取值,可以发现问题所在,进而有利于对模型的深入分析。

(3) 模型合适但方程个数小于未知数个数的情况(对应于(5)式)

直接利用(10)式分别由 SVD 算法和 LS 方法对表 1 中第 i 组和 $i+1$ 组($i = 1, 2, \cdots, 16$)模拟数据进行拟合,结果如下:

SVD 算法:

$i = 1,2$ $a = 0.896 \pm 0.751$, $b = 2.021 \pm 0.500$, $c = 5.393 \pm 0.254$

$i = 2,3$ $a = 1.430 \pm 0.121$, $b = 3.295 \pm 0.220$, $c = 4.909 \pm 0.051$

$i = 3,4$ $a = 0.799 \pm 0.046$, $b = 2.752 \pm 0.133$, $c = 5.043 \pm 0.021$

$i = 4,5$ $a = 0.723 \pm 0.024$, $b = 3.159 \pm 0.095$, $c = 4.986 \pm 0.012$

$i = 5,6$ $a = 0.573 \pm 0.015$, $b = 3.087 \pm 0.074$, $c = 4.996 \pm 0.007$

$i = 6,7$ $a = 0.487 \pm 0.010$, $b = 3.108 \pm 0.061$, $c = 4.995 \pm 0.005$

$i = 7,8$ $a = 0.418 \pm 0.008$, $b = 3.087 \pm 0.052$, $c = 4.997 \pm 0.004$

$i = 8,9$ $a = 0.373 \pm 0.006$, $b = 3.131 \pm 0.045$, $c = 4.994 \pm 0.003$

$i = 9,10$ $a = 0.322 \pm 0.005$, $b = 3.029 \pm 0.040$, $c = 5.001 \pm 0.002$

$i = 10,11$ $a = 0.301 \pm 0.004$, $b = 3.127 \pm 0.036$, $c = 4.996 \pm 0.002$

$i = 11,12$ $a = 0.268 \pm 0.003$, $b = 3.058 \pm 0.033$, $c = 4.999 \pm 0.002$

$i = 12,13$ $a = 0.246 \pm 0.003$, $b = 3.046 \pm 0.030$, $c = 4.999 \pm 0.001$

$i = 13,14$ $a = 0.228 \pm 0.002$, $b = 3.049 \pm 0.027$, $c = 4.999 \pm 0.001$

$i = 14,15$ $a = 0.213 \pm 0.002$, $b = 3.071 \pm 0.025$, $c = 4.999 \pm 0.001$

$i = 15,16$ $a = 0.198 \pm 0.002$, $b = 3.055 \pm 0.024$, $c = 4.999 \pm 0.001$

$i = 16,17$ $a = 0.187 \pm 0.001$, $b = 3.067 \pm 0.022$, $c = 4.999 \pm 0.001$

SL 方法:所有计算均失败

可见,当系数矩阵存在行奇异现象时,LS 方法不能提供稳定解,而 SVD 算法仍可对正解作出估计,即通过(6)式对(5)式提供了一种广义最小二乘意义上的解,且具有一定的可靠性。显然,实际工作中应尽量避免系数矩阵发生奇异现象,但偶尔出现却是无法回避的。

模拟数据检验结果表明，通常情况下 SVD 算法和 LS 方法能够提供彼此相符的参数估计。当系数矩阵接近或出现奇异现象时，LS 方法失败，而 SVD 算法仍能提供近似解。亦即 SVD 算法完全能够取代 LS 方法，并能解决更为广泛的问题。

4 天文应用举例

在研究射电与光学天球参考架的连接时，射电星与河外射电源光学对应体是两类非常重要的天体。然而，由于射电星的射电流量非常弱和射电源光学对应体的星等非常暗，以致于相应的观测资料非常有限，因而也特别宝贵。将不同时期、相对于各种星表和带有各种误差的少量观测资料转换为统一系统，这是研究参考架连接问题的重要工作之一。一般而言，参考历元和参考星表的统一化是比较容易实现的，但对于如何将资料进一步统一，这是一个十分棘手的问题。在处理恒星观测时，由于星多、天区覆盖面广且密度高，可以采用统计手段通过比较共同星的坐标而将观测资料进一步统一化。但是，由于射电星与河外射电源光学对应体观测资料的数量非常有限，无法采取类似于恒星观测处理中的统计学方法。此时，采用小角旋转的处理方法[8,9]是一项有益的尝试。

文[10]中，通过对 28 个河外射电源光学对应体观测表的参考历元和参考星表进行归一化处理之后，基于小角旋转模型和采用 SVD 算法，得到了基于 FK5/J2000.0 系统和由 510 颗源组成的综合表，其基本源、二级源和补充源的赤经和赤纬坐标内符精度分别为$0''.088$，$0''.256$，$0''.377$ 和 $0''.093$，$0''.232$，$0''.374$。进一步的比较表明，该综合表与国际地球自转服务(IERS)射电天球参考架(ICRF)的定向差异为

$$A_1 = -0''.013 \pm 0''.012, \quad A_2 = 0''.077 \pm 0''.012, \quad A_3 = 0''.005 \pm 0''.009 \tag{11}$$

这与 Arias 等人关于 FK5/J2000.0 系统与 ICRF 的定向差异研究结果非常吻合[11]，因而间接地证明了该综合表与 FK5/J2000.0 系统是一致的，同时也说明采用小角旋转模型是合理的、SVD 算法是有效的。

事实上，在该综合表的组建过程中需要基于共同源确定不同观测源表之间的小角旋转。由于共同源很少，限制了观测方程的个数，再加上观测误差和参考星表系统化算误差等因素的影响，致使一些运算中的法方程接近或出现奇异现象，LS 方法不能够提供稳定解，只能运用 SVD 算法。可见，(11)式是将 SVD 算法成功运用于天文学研究的很好例证。

另外，在利用国际纬度服务(ILS)资料研究长期极移的工作中，Gross 采用了一种有别于传统的研究方法，亦即在利用纬度观测确定极移(x，y 分量)时，同时确定各观测台站的改正量[12]。如果 t 时刻有 N 个台站观测，则需要确定 N 个台站的各自改正和两个极移分量，即未知数比方程多两个。Gross[13]采用了迭代方法进行求解，但迭代方程的推导是不严格的，而且迭代过程也比较烦琐。对于此类方程的求解不妨采用 SVD 算法，一方面，通过考察 W_j 的取值，可以判断各拟合参量在模型中的作用，另一方面，也可以将 SVD 算法的解算结果与 Gross 的迭代结果进行比较，进而对极移中 10 年尺度分量的存在与否作出判断。

总之，SVD 算法的数学依据是可靠的，背景分析是合理的，实现程序是稳定的，模拟数据检验结果表明其优点是明显的，天文学研究中也存在应用实例。建议对 SVD 算法予以关注和运用。

参 考 文 献

[1] Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., Numerical Recipes: the Art of Scientific Computing, the Press Syndicate of the Univ. of combridge, New York, 1986

[2] Golub, G. H., and Van Loan, C. F., Matrix computations, Baltimore: Johns Hopkins University Press, 1983

[3] Forsythe, G. E., Malcolm, M. A., and Moler, C. B., Computer Methods for Mathematical Computations, englewood Cliffs, N. J.: Prentice－Hall, 1977

[4] Stoer, J., and Bulirsch, R., Introduction to Numerical Analysis, New York: Springer－Verlag, 1980

[5] Bevington, Philip R., Data Reduction and Error Analysis for the Physical Sciences, New York: McGraw－Hill, 1969

[6] Forsythe, George E., Malcolm, Michael A., and Moler, Cleve B., Computer Methods for Mathematical Computations, Endlewood Cliffs, N. J.: Prentice－Hall, 1977

[7] Arni, Y., AJ 210, 1976, 642

[8] Lampton, M., Margon, M., and Bowyet, S., AJ 208, 1976, 177

[9] 李金岭,金文敬,李金增,云南天文台台刊,1994(3), 1

[10] 李金岭,金文敬,天文学进展,1995,13(2):93－106

[11] Li Jinling, Jin Wenjing, A&As 120, 1996,1

[12] Arias, E. F., Charlot, P., Feissel, M., Lestrade, J.—F., A&A 303, 1995, 604

[13] Gross, R. S., The Secular Drift of the Rotation Pole, in: Earth Rotation and Coordinate Reference Frames, eds. C. Boucher & G. A. Wilkins, Springer－Verlag, 1990, 146

AN INTRODUCTION OF THE SVD ALGORITHM AND ITS TEST OF ARTIFICIAL DATA

Li Jinling

(Shanghai Observatory, The Chinese Academy of Sciences, Shanghai 200030)

Abstract

The singular value decomposition (SVD) algorithm is briefly introduced concerning its mathematic basis of theorem, consideration of physical background and the process of data reduction. The advantage and function of the SVD algorithm are demonstrated by basing on tests of artificial data. It is highlighted that SVD algorithm can be applied to solve not only the problems which the least squares method can be, but also some of the problems which normal equations will be fail. It is therefore recommended that to use SVD techniques instead of using the normal equations.

Key words data procesing — parameter — least square method — singular value decomposition (SVD) algorithm