总体最小二乘估计

1 问题描述

给定一组长度为的二维坐标序列，假设有元参数的线性模型用于描述该组二维坐标序列，即，其中。

由于观测数据长度远大于未知数个数，因此，该参数估计方程是超定的。如果自变量中不包含噪声，模型数据和测量数据之间的唯一区别一定是由于的测量误差和模型无法反映的数据的自然可变性。在上述情况下，最佳参数估计要求尽可能减小观测数据与模型数据之间的度量。参数估计问题则可以表示为



其中，向量表示测量值与模型之间的差异。能最小化二范数的模型参数称为一般最小二乘解（OLS, ordinary least squares）。

如果自变量同因变量均有测量误差，那么最优估计模型则需要减小自变量与因变量两者测量值与模型之间度量。因此，需要在式中加入自变量测量误差矩阵来表示新的参数估计问题



在参数估计中，将因变量和自变量误差同时最小化的算法称为总体最小二乘法（TLS, total least squares）。由此，问题转化为寻找满足方程的自变量与因变量的最小测量误差，即：



其中，表示与的组合，表示矩阵的*Frobenius*范数，



其中，表示矩阵的第个奇异值。因此，



上式的假设了与的误差量级接近。

2 奇异值分解（SVD）

对于矩阵的奇异值分解可以表示为以下三个矩阵的乘积：



其中，矩阵与为单位正交阵，即。矩阵的前行为对角阵，对角元素为奇异值。

矩阵与向量的乘积为向量，可解释为矩阵的乘积，上述三个矩阵分别将向量旋转、拉伸再旋转到维的数据空间中。

由于矩阵单位正交，因此其每列的长度为1。的每一列的外积长度也为1。也可以表示为



上式中，每一次累加都将使矩阵的秩增加1。由于奇异值随的增大而减小，因此第一项贡献最大，其余每一项的贡献随着的增大而减小。在一般参数估计问题中，奇异值都会出现“断崖”式下降，即



上述情况中，前项的贡献量级较大。

**Eckart-Young定理** 秩为n的矩阵的奇异值分解前项构成的矩阵满足使极小。

SVD可用于OLS中超定方程组的求解，利用的单位正交性与的对角性，



其中，称为Moore-Penrose逆。

3 SVD与TLS

SVD也可用于寻找TLS的特解。满足问题的方程可以改写为：



其中，向量是矩阵的零空间中的向量。

由于数据向量不会位于矩阵张成的子空间中（），矩阵的秩为。即，有列线性不相关。

为使式所表示的超定方程组的解唯一，矩阵的秩需等于（矩阵的秩=矩阵列空间长度+矩阵零空间长度）。因此，问题转化为寻找最小矩阵使得秩为的矩阵转化为秩为的矩阵。引入Eckart-Young定理的意义即在于此，定义为的最优秩估计。丢弃的最后一个（最小的）奇异值可以消除数据中最少的信息，并确保唯一解。

的SVD可以表示为



其中，有列，的对角元素为最大的个奇异值。



现定义为Frobenius准则下的最优秩估计。



将式与式代入中，



因此，









对比式与式，参数估计问题的总体最小二乘解为



4 SVD与TLS