Quadratic Programming

张忠桢 著

数学★软件★应用系列教材

二次规划

——非线性规划与投资组合的算法

WUHAN UNIVERSITY PRESS
武汉大学出版社

数学★软件★应用系列教材

二次规划——非线性规划与投资组合的算法

张忠桢　著

WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

二次规划:非线性规划与投资组合的算法/张忠桢著.—武汉:武汉大学出版社,2006.4
数学·软件·应用系列教材
ISBN 7-307-04964-3

Ⅰ.二… Ⅱ.张… Ⅲ.凸规划—研究 Ⅳ.O221.2

中国版本图书馆 CIP 数据核字(2006)第 019505 号

责任编辑:杨 华　　责任校对:刘 欣　　版式设计:支笛

出版发行:**武汉大学出版社** (430072 武昌 珞珈山)
(电子邮件:wdp4@whu.edu.cn 网址:www.wdp.com.cn)
印刷:湖北省通山县九宫印务有限公司
开本:787×980 1/16 印张:14.75 字数:269 千字
版次:2006 年 4 月第 1 版 2006 年 4 月第 1 次印刷
ISBN 7-307-04964-3/O·337 定价:25.00 元

前　言

二次规划是最简单的非线性规划,通常通过解其库恩-塔克条件(KT 条件)获取一个解. KT 条件的解称为 KT 对,其中与原问题的变量对应的部分称为 KT 点. 二次规划分为凸二次规划和非凸二次规划,前者的 KT 点便是其全局极小点,而后者的 KT 点可能连局部极小点都不是.

有很多算法可用于解凸二次规划,如 Wolfe, Dantzig, Lemke, van de Panne and Whinston, Fletcher, Keller, Best and Ritter, Gill and Murray, von Hohenbalken, Beale, Goldfarb, Benveniste, Sacher, Jensen and King 等人的方法. 还有不少内点算法可用于解凸二次规划. 应用最广泛的是 Fletcher 的有效集法. 这种方法可用于非凸二次规划,但是也只能求出 KT 点,不能保证它是局部极小点. 为了判断 KT 点是不是局部极小点,通常的做法是随机地产生许多距离 KT 点很近的可行解,看是否有一个解对应的目标函数值更小,称为摄动(perturbation)法. 这种方法有时如同大海捞针,当 KT 点处可行下降方向的集合是一个十分狭窄的区域时,很难发现一个更好的解,而且可能不存在可行下降方向. 例如 $x_1=0$, $x_2=0$ 是二次规划

$$\text{minimize } f(\boldsymbol{x})=x_1^2-20x_1x_2+(100-\delta)x_2^2 \text{ subject to } x_1+x_2\leqslant 1,\ x_1,\ x_2\geqslant 0$$

的 KT 点. 当 δ 是很小正数时,目标函数在 KT 点处下降方向的区域很狭窄;当 δ 小于或等于零时没有下降方向.

本书重点介绍非凸二次规划的旋转算法,由简单到复杂逐步讨论以下几个问题:

(i) 求二次齐次函数的非负下降方向;

(ii) 求二次齐次函数在锥内的下降方向;

(iii) 求非凸二次规划的局部极小点;

(iv) 求非凸二次规划的全局极小点.

问题(i)是在 n 维空间直角坐标系的第一象限找一个点使一个不定二次型小于零,可用数学规划模型表示为

$$\text{minimize } f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} \text{ subject to } \boldsymbol{x}\geqslant \boldsymbol{0},$$

其中 $\boldsymbol{H}$ 是 n 阶不定对称矩阵,$\boldsymbol{x}$ 是 n 维未知列向量. 这是一个 NP hard 问题,见

Murty 和 Kabadi 的著述(1987). 如果 $\boldsymbol{H}$ 有一个负的对角元素,或者有一个 2 阶主子矩阵,其两个对角元素为 0,非对角元素是负数,很容易确定一个非负的 $\boldsymbol{x}$ 使得 $f(\boldsymbol{x})<0$. 问题难在 $\boldsymbol{H}$ 的所有对角元素大于零且每行(列)都有负的非对角元素的情形. 本书利用旋转运算,结合一种参数化技术,证明了此问题等价于在一个有限集中寻找 I 型或 II 型非负下降方向. 问题(ii)与问题(i)类似,不过形式要复杂一些. 问题(iii)的关键在于判断 KT 点处是否有可行下降方向,有赖于问题(ii)的解决. 本书以较大篇幅介绍非凸二次规划局部极小点,即问题(iii)的计算方法. 至于问题(iv),只有当变量数目较少时才可能以较小的计算量获得全局极小点. 随着问题规模不断增加,计算量成指数膨胀. Hansen 等人(1993)、Vandenbussche 和 Nemhauser(2005)提出用分支定界或分支切割算法解一个简单的非凸二次规划,其约束条件仅仅是每个变量在 0 和 1 之间. 也就是说,用类似于解混合整数规划的方法解一个简单的非凸二次规划,其计算量之大可想而知. 本书简单介绍如何从一个局部极小点出发寻找其他局部极小点,为进一步研究非凸二次规划的全局极小点提供一种思路.

序列二次规划(SQP)法是求解复杂非线性规划最有效的方法. SQP 法将一个复杂的非线性规划在某一点处用一个二次规划近似,前者称为原问题,后者称为子问题. 通过求解二次规划子问题获取一个新的解(点),然后再将原问题在新的点用一个二次规划近似并求解,如此等等,以获取原问题的一个解. 用 SQP 法解非线性规划关键在于二次规划模型的形式及其计算方法. 序列二次规划法分为牛顿法和拟牛顿法. 在无约束极值问题的牛顿法中,子问题目标函数的海色(Hesse)矩阵是原问题目标函数 $f(\boldsymbol{x})$ 在某点 $\boldsymbol{x}^{(k)}$ 处的海色矩阵 $\nabla^2 f(\boldsymbol{x}^{(k)})$. 对于约束极值问题,子问题目标函数的海色矩阵通常是拉格朗日函数在 $(\boldsymbol{x}^{(k)}, \boldsymbol{\lambda}^{(k)})$ 处的海色矩阵 $\nabla_x^2 L(\boldsymbol{x}^{(k)}, \boldsymbol{\lambda}^{(k)})$. 如果原问题的目标函数和约束条件较为复杂,计算二阶偏导数是一项烦琐的工作,而且得到的海色矩阵可能不是正定的,导致一个非凸的二次规划子问题. 于是,人们常用一个正定矩阵 $\boldsymbol{B}_k$ 代替目标函数或拉格朗日函数的海色矩阵,并且在迭代过程中不断校正 $\boldsymbol{B}_k$,这种方法便是拟牛顿法. 近几十年来,拟牛顿法的理论研究成果很多,实际计算效果也较好. 校正 $\boldsymbol{B}_k$ 的公式有 DFP 公式、BFGS 公式、Broyden 族、Huang 族等,其中最有名的是 BFGS 公式. 这些公式力图使 $\boldsymbol{B}_k$ 正定,或者说使子问题成为凸二次规划. 可以证明,当 $\boldsymbol{B}_k$ 正定且充分接近 $\nabla^2 f(\boldsymbol{x}^{(k)})$ 或 $\nabla_x^2 L(\boldsymbol{x}^{(k)}, \boldsymbol{\lambda}^{(k)})$ 时,算法的收敛性很好. 对于无约束极值问题,容易做到所有 $\boldsymbol{B}_k$ 正定,例如使用精确线搜索,但是这一点并不能保证算法对于所有问题收敛. Dai Yuhong(2002)和 Mascarenhas W. F.(2004)就给出了 BFGS 法用于非凸函数失效的一些例子. 至于某些约束极值问题,想在迭代过程中让每个 $\boldsymbol{B}_k$ 正定都很困难. 早在 1987 年,

Murty 和 Kabadi 就指出了这些算法的局限性，他们说："许多关于非线性规划的教科书留给读者一种印象：其算法收敛于凸非线性规划的全局极小点，收敛于非凸非线性规划的局部极小点，许多专业非线性规划的软件包文件也制造同样的印象，这是十分错误的."

本书也用序列二次规划法解更为复杂的非线性规划. 但与拟牛顿法不同，二次规划子问题目标函数的海色矩阵就是原问题目标函数的海色矩阵，或者是拉格朗日函数的海色矩阵. 当约束条件较为复杂时，还可能使用增广拉格朗日函数的海色矩阵. 这时的子问题可能是凸二次规划也可能是非凸的. 非线性规划的形式多种多样，不同的算法有各自的适用范围. 有的问题用序列线性规划法计算效果也很好. 希望本书介绍的方法能弥补传统算法的不足.

本书还介绍若干种资产组合选择模型的计算方法. 这些模型都以方差作为风险的度量，其中一种模型考虑凹交易成本，另有一种考虑 V 形交易成本，用序列二次规划法计算. 据我们所知，现有文献都是用线性规划解这两种问题，用绝对偏差等指标作为风险的度量，而不是实际中通常使用的方差风险.

全书分为六章. 第一章介绍线性代数和微分学的一些基础知识以及最优化的有关定理. 第二章介绍线性不等式组的解法及参数化技术. 第三章介绍二次规划旋转算法的基本概念和凸二次规划的解法. 第四章由简单到复杂逐步介绍非凸二次规划的计算方法. 第五章介绍约束非线性规划的序列二次规划法. 第六章介绍若干均值方差资产组合选择问题的解法.

全书结构严谨，深入浅出，所有算法都有相应的定理为基础. 为便于初学者掌握，仅用到一些较基本的数学知识，并且大多数问题有简单的例子说明算法的计算步骤. 阅读本书只需具备线性代数、概率论和微分学的一些基础知识. 本书可作为应用数学、经济学、管理科学以及许多工程技术学科研究生的教材或教学参考书. 涉足运筹学优化的教师和研究人员更应该知道非线性规划特别是一些较为简单的非线性规划如何计算.

张忠桢
2006 年 3 月于武汉

目　录

第一章　基础知识

1.1　矩阵的旋转运算及分块矩阵公式

1.1.1　矩阵的一些概念

设 $A=(a_{ij})$ 是 $m\times n$ 矩阵，其中 a_{ij} 是第 i 行第 j 列的元素，$i=1,2,\cdots,m$，$j=1,2,\cdots,n$. A 的 m 个行向量的秩称为 A 的**行秩**，A 的 n 个列向量的秩称为 A 的**列秩**. 若 A 的 m 行线性无关，则称 A **行满秩**. 同样，若 A 的 n 列线性无关，则称 A **列满秩**. 可以证明，A 的行秩等于 A 的列秩. A 的行秩或列秩称为 A 的**秩**. 秩等于行数或列数的矩阵称为**满秩矩阵**.

若 $A=(a_{ij})$ 的行数等于列数等于 n，称 A 是一个 n 阶方阵，其中 a_{11}，a_{22}，$\cdots$，a_{nn} 称为对角元素，其他元素称为非对角线元素. 一个满秩方阵称为**非奇异矩阵**. 若存在方阵 $\boldsymbol{B}$，使得 $\boldsymbol{AB}=\boldsymbol{BA}=\boldsymbol{I}$，则称 A 可逆，并称 $\boldsymbol{B}$ 为 A 的**逆矩阵**，记为 A^{-1}，这里的 $\boldsymbol{I}$ 是单位矩阵. 可以证明，A 为非奇异矩阵的必要充分条件是 A 可逆.

n 阶方阵 $A=(a_{ij})$ 的元素如果满足 $a_{ij}=a_{ji}(i\neq j)$ 或者说 $A^{\mathrm{T}}=A$，则称 A 为对称矩阵，这里的 T 是转置符. 设 A 为 n 阶对称矩阵. 如果对于任何 n 维向量 $\boldsymbol{x}$(列向量) 均有 $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}\geqslant 0$，则称 A **正半定**. 若等号仅当 $\boldsymbol{x}=\boldsymbol{0}$ 时成立，则称 A **正定**.

可以证明以下结论成立.

结论 1.1.1　设 A 为对称矩阵，则以下诸命题等价：

(i) A 为正半定矩阵；

(ii) 有矩阵 $\boldsymbol{B}$ 使 $A=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$；

(iii) A 的所有(顺序) 主子式非负；

(iv) A 的所有特征值非负.

结论 1.1.2　设 A 为对称矩阵，则以下诸命题等价：

(i) A 为正定矩阵；

(ii) 有列满秩矩阵 B 使 $A = B^{\mathrm{T}}B$；

(iii) A 的所有(顺序) 主子式为正；

(iv) A 的所有特征值大于 0.

1.1.2 矩阵的旋转运算

设 $A = (a_{ij})$ 是 $m \times n$ 矩阵. 用 $\boldsymbol{a}_i$ 表示 A 的第 i 行，$\boldsymbol{e}_i$ 表示 n 阶单位矩阵的第 i 行，则

$$\boldsymbol{a}_i = a_{i1}\boldsymbol{e}_1 + a_{i2}\boldsymbol{e}_2 + \cdots + a_{in}\boldsymbol{e}_n, \quad i = 1,2,\cdots,m. \tag{1.1.1}$$

对于某个 $r \in \{1,2,\cdots,m\}, s \in \{1,2,\cdots,n\}$，如果 $a_{rs} \neq 0$，则可由式(1.1.1) 的第 r 式解得

$$\boldsymbol{e}_s = \frac{1}{a_{rs}}\boldsymbol{a}_r + \sum_{j=1,\, j\neq s}^{n}\left(-\frac{a_{rj}}{a_{rs}}\right)\boldsymbol{e}_j. \tag{1.1.2}$$

将其代入式(1.1.1) 的其余各式整理后可得

$$\boldsymbol{a}_i = \frac{a_{is}}{a_{rs}}\boldsymbol{a}_r + \sum_{j=1,j\neq s}^{n}\left(a_{ij} - \frac{a_{is}}{a_{rs}}a_{rj}\right)\boldsymbol{e}_j, \quad i = 1,\cdots,r-1,r+1,\cdots,m. \tag{1.1.3}$$

以上运算过程称为旋转运算，运算前后的数据可用表 1.1、表 1.2 表示.

表 1.1 初 始 表

	$\boldsymbol{e}_s$	$\boldsymbol{e}_j$
$\boldsymbol{a}_r$	a_{rs}	a_{rj}
$\boldsymbol{a}_i$	a_{is}	a_{ij}

表 1.2 旋转运算的结果

	$\boldsymbol{a}_r$	$\boldsymbol{e}_j$
$\boldsymbol{e}_s$	$\dfrac{1}{a_{rs}}$	$-\dfrac{a_{rj}}{a_{rs}}$
$\boldsymbol{a}_i$	$\dfrac{a_{is}}{a_{rs}}$	a_{ij}'

其中 $a_{ij}' = a_{ij} - \dfrac{a_{is}}{a_{rs}}a_{rj}$, $i = 1,\cdots,r-1,r+1,\cdots,m$, $j = 1,\cdots,s-1,s+1,\cdots,n$.

如果将向量组 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m,\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$ 的一个最大线性无关组称为**基**，那么 $\boldsymbol{e}_1, \boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$ 是其一个基. 由于 $a_{rs} \neq 0$，向量组 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_{s-1},\boldsymbol{a}_r,\boldsymbol{e}_{s+1},\cdots,\boldsymbol{e}_n$ 也线性无关，因而也是一个基. 在以上运算中，我们说 $\boldsymbol{a}_r$ 入基 $\boldsymbol{e}_s$ 出基，两者位置的交换记为 $\boldsymbol{a}_r \leftrightarrow \boldsymbol{e}_s$, a_{rs} 称为**枢轴**(元素). 表 1.1 中(即矩阵 A 中) 枢轴元素 a_{rs} 所在行称为**枢轴行**，其所在列称为**枢轴列**. 去掉表 1.2 中 $\dfrac{1}{a_{rs}}$ 所在行和列后，剩下的元素

构成的矩阵(a'_{ij})称为$\boldsymbol{A}$关于a_{rs}的**残余矩阵**.

给定一组向量$\boldsymbol{a}_i=(a_{i1},a_{i2},\cdots,a_{in})$，$i=1,2,\cdots,m$，其中有多少向量线性无关，或者一个向量能否表示为其余若干向量的线性组合，这个问题在许多问题的讨论中具有十分重要的作用，下面用旋转运算研究这个问题.

为叙述方便起见，将向量组$\boldsymbol{a}_1,\cdots,\boldsymbol{a}_m,\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$的一个基中的向量称为**基向量**，将其余向量称为**非基向量**，并将$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m$称为**原向量**，$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$称为**外来向量**. 首先以$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$作为基向量，并将所有原向量表示为这些向量的线性组合，如表1.1所示. 然后尽可能地将非基向量中的原向量与基中的外来向量交换，能够进行这种交换的必要充分条件是残余矩阵中至少有一个非零元素. 最后将出现以下两种情况：

(i) 所有原向量$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m$进入基中，因此这些向量线性无关；

(ii) $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m$中的部分向量进入基中，而所有非基向量中的原向量已表示为这些向量的线性组合.

由于每进行一次旋转运算有一个原向量入基，所以计算结束时，旋转运算的次数等于向量组$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m$的秩. 实际计算时，某个外来向量一旦出基便可删掉. 例如可不要表1.2中的$\boldsymbol{e}_s$行. 如果目的仅在于求向量组$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m$的秩或其中的一个最大无关组，还可不要表1.2中的$\boldsymbol{a}_r$列，因为原向量进入基中后，其所在列各元素不能再作枢轴元素.

现设$\boldsymbol{A}=(a_{ij})$是个n阶方阵. 同样将其n行$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n$表示为$n$个单位向量$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$的线性组合，如表1.1所示. 如果$a_{11}\neq 0$，则可进行一次旋转运算. 在得到的表中，如果主对角线上第二个元素也不为零，则可进行第二次旋转运算，依此类推. 假设共进行了n次旋转运算，得到的结果用式子表示即为

$$\begin{pmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\\ \vdots\\ \boldsymbol{e}_n\end{pmatrix}=\begin{pmatrix}u_{11} & u_{12} & \cdots & u_{1n}\\ u_{21} & u_{22} & \cdots & u_{2n}\\ \vdots & \vdots & & \vdots\\ u_{n1} & u_{n2} & \cdots & u_{nn}\end{pmatrix}\begin{pmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\\ \vdots\\ \boldsymbol{a}_n\end{pmatrix},$$

可见得到的矩阵(u_{ij})是$\boldsymbol{A}$的逆矩阵.

注　如果用计算机求一个n阶方阵$\boldsymbol{A}$的逆阵，为减少计算误差，最好在残余矩阵(开始时为$\boldsymbol{A}$)中选一个绝对值最大的元素. 如果大于一个充分小的正数ε(一般为$10^{-6}\sim10^{-4}$)便以其为枢轴作一次旋转运算. 只要能进行n次旋转运算，$\boldsymbol{A}$可逆. 这时，初始的基向量$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$成为非基向量，非基向量$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m$成为基向量，但它们的次序往往会发生变化. 为得到$\boldsymbol{A}$的逆，只需将

e_i 行和 a_j 列上的元素放在第 i 行和第 j 列，$i,j=1,2,\cdots,n$. 挑选绝对值最大的元素作枢轴相当于线性方程组的全主元高斯消元法，计算量大但解的精度高. 当 n 很大时，可使用部分选主元的方法. 如果 A 正定，沿其主对角线进行旋转运算便可，数值稳定性非常好(见曹志浩、张玉得和李瑞遐(1979)). 这意味着，当 A 是一个具有正对角元素的对称矩阵时，仅以对角线上的正元素为枢轴作旋转运算，数值稳定性将会非常好.

定理 1.1.1(旋转运算公式)　设一个 $m\times n$ 矩阵的分块矩阵为$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$,其中 A 为可逆方阵. 那么，用旋转运算在此分块矩阵中求 A 的逆，其结果为 $\begin{pmatrix} A^{-1} & -A^{-1}B \\ CA^{-1} & D-CA^{-1}B \end{pmatrix}$.

证　用 a_i 表示此 $m\times n$ 矩阵的第 i 行，e_i 表示 n 阶单位矩阵的第 i 行，设 A 的阶数为 k，那么

$$\begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} = A\begin{pmatrix} e_1 \\ \vdots \\ e_k \end{pmatrix} + B\begin{pmatrix} e_{k+1} \\ \vdots \\ e_n \end{pmatrix},$$

两边左乘 A^{-1} 并移项得

$$\begin{pmatrix} e_1 \\ \vdots \\ e_k \end{pmatrix} = A^{-1}\begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} - A^{-1}B\begin{pmatrix} e_{k+1} \\ \vdots \\ e_n \end{pmatrix},$$

所以对 A 求逆后上方的两个子块变为 A^{-1} 和 $-A^{-1}B$. 另外

$$\begin{aligned}\begin{pmatrix} a_{k+1} \\ \vdots \\ a_m \end{pmatrix} &= C\begin{pmatrix} e_1 \\ \vdots \\ e_k \end{pmatrix} + D\begin{pmatrix} e_{k+1} \\ \vdots \\ e_n \end{pmatrix} = C\left[A^{-1}\begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} - A^{-1}B\begin{pmatrix} e_{k+1} \\ \vdots \\ e_n \end{pmatrix}\right] + D\begin{pmatrix} e_{k+1} \\ \vdots \\ e_n \end{pmatrix} \\ &= CA^{-1}\begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} + (D-CA^{-1}B)\begin{pmatrix} e_{k+1} \\ \vdots \\ e_n \end{pmatrix},\end{aligned}$$

所以下方的两个子块变为CA^{-1} 和 $D-CA^{-1}B$.

1.1.3　若干分块矩阵公式

以下是一些较为常用的分块矩阵公式.

(i)若 A_{11} 可逆，则

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} I_1 & O \\ A_{21}A_{11}^{-1} & I_2 \end{pmatrix}\begin{pmatrix} A_{11} & O \\ O & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix}\begin{pmatrix} I_1 & A_{11}^{-1}A_{12} \\ O & I_2 \end{pmatrix}, \tag{1.1.4a}$$

其中 I_1 和 I_2 为单位矩阵；若 A_{22} 可逆，则

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} I_1 & A_{12}A_{22}^{-1} \\ O & I_2 \end{pmatrix}\begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & O \\ O & A_{22} \end{pmatrix}\begin{pmatrix} I_1 & O \\ A_{22}^{-1}A_{21} & I_2 \end{pmatrix}. \tag{1.1.4b}$$

(ii) 若 A_{11} 可逆，则

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \text{可逆} \Leftrightarrow (A_{22} - A_{21}A_{11}^{-1}A_{12}) \text{可逆},$$

其中“$\Leftrightarrow$”表示必要充分条件，并且

$$\begin{aligned}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} &= \begin{pmatrix} I_1 & -A_{11}^{-1}A_{12} \\ O & I_2 \end{pmatrix}\begin{pmatrix} A_{11}^{-1} & O \\ O & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{pmatrix}\begin{pmatrix} I_1 & O \\ -A_{21}A_{11}^{-1} & I_2 \end{pmatrix} \\ &= \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \end{aligned} \tag{1.1.5a}$$

其中

$$\begin{aligned} B_{11} &= A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}, \\ B_{12} &= -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}, \\ B_{21} &= -(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}, \\ B_{22} &= (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}. \end{aligned}$$

若 A_{22} 可逆，则

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \text{可逆} \Leftrightarrow (A_{11} - A_{12}A_{22}^{-1}A_{21}) \text{可逆},$$

并且

$$\begin{aligned}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1} &= \begin{pmatrix} I_1 & O \\ -A_{12}A_{22}^{-1} & I_2 \end{pmatrix}\begin{pmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & O \\ O & A_{22}^{-1} \end{pmatrix}\begin{pmatrix} I_1 & -A_{12}A_{22}^{-1} \\ O & I_2 \end{pmatrix} \\ &= \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \end{aligned} \tag{1.1.5b}$$

其中

$$B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1},$$

$$B_{12} = -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1},$$

$$B_{21} = -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1},$$

$$B_{22} = A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1}.$$

(iii)若 $A_{11}, A_{22}, \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$均可逆，则

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1},$$

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} = A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1},$$

$$A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1},$$

$$A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}. \quad (1.1.6)$$

(iv)若 A_{12}, A_{21} 可逆，则

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & O \end{pmatrix}^{-1} = \begin{pmatrix} O & A_{21}^{-1} \\ A_{12}^{-1} & -A_{12}^{-1}A_{11}A_{21}^{-1} \end{pmatrix}. \quad (1.1.7)$$

以上公式选自须田信英等的相关著述(1979).

1.1.4 其他旋转运算公式

定理 1.1.2 设$\begin{pmatrix} A & B \\ B^{\mathrm{T}} & D \end{pmatrix}$是秩为 r 的 n 阶正半定矩阵,其中 A 为 k 阶可逆方阵, $k<n$. 则 $D-B^{\mathrm{T}}A^{-1}B$ 是秩为 $r-k$ 的正半定矩阵.

证 在下式两边右乘$(0,x)$左乘$(0,x)^{\mathrm{T}}$，其中 0 是 k 维零向量，x 是任意 $n-k$维向量(都是行向量)：

$$\begin{pmatrix} I & O \\ -B^{\mathrm{T}}A^{-1} & I \end{pmatrix}\begin{pmatrix} A & B \\ B^{\mathrm{T}} & D \end{pmatrix}\begin{pmatrix} I & -A^{-1}B \\ O & I \end{pmatrix} = \begin{pmatrix} A & O \\ O & D-B^{\mathrm{T}}A^{-1}B \end{pmatrix}.$$

由于$\begin{pmatrix} A & B \\ B^{\mathrm{T}} & D \end{pmatrix}$正半定，$x(D-B^{\mathrm{T}}A^{-1}B)x^{\mathrm{T}}$ 非负，因而 $D-B^{\mathrm{T}}A^{-1}B$ 正半定，且秩为 $r-k$. ▌

定理 1.1.1 和定理 1.1.2 说明，如果矩阵 A 正半定，则以主对角线上的正元素为枢轴进行旋转运算得到的残余矩阵也正半定.

定理 1.1.3 设 A 是一个 n 阶对称矩阵，则 A 正定的必要充分条件是可沿其主对角线进行 n 次旋转运算并且所有枢轴元素都是正数.

证 考虑 A 的 $k+1$ 阶主子矩阵并将其表示为分块矩阵 $\begin{pmatrix} A_k & b \\ b^{\mathrm{T}} & d \end{pmatrix}$，其中 A_k 是 A 中左上角的 k 阶矩阵，$1 \leqslant k < n$，并设 A_k 的行列式大于 0. 由于 $\begin{vmatrix} A_k & b \\ b^{\mathrm{T}} & d \end{vmatrix} = \left| A_k \right| (d - b^{\mathrm{T}} A_k^{-1} b)$ 并且 $d - b^{\mathrm{T}} A_k^{-1} b$ 是在 A 中沿 A_k 的主对角线进行 k 次旋转运算所得矩阵主对角线上第 $k+1$ 个元素，由归纳不难看出此定理成立. ▌

定理 1.1.4 在分块矩阵 $\begin{pmatrix} B & -A^{\mathrm{T}} \\ A & D \end{pmatrix}$ 中 B 正定，D 正半定，那么沿 B 的对角线进行旋转运算得到的矩阵 $\begin{pmatrix} B^{-1} & B^{-1}A^{\mathrm{T}} \\ AB^{-1} & D + AB^{-1}A^{\mathrm{T}} \end{pmatrix}$ 正半定并且其秩等于 B 与 D 的秩之和.

证 因为

$$\begin{pmatrix} B^{-1} & B^{-1}A^{\mathrm{T}} \\ AB^{-1} & D + AB^{-1}A^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} I_1 & O \\ A & I_2 \end{pmatrix} \begin{pmatrix} B^{-1} & O \\ O & D \end{pmatrix} \begin{pmatrix} I_1 & A^{\mathrm{T}} \\ O & I_2 \end{pmatrix}. \quad ▌$$

设 A 和 B 是同阶方阵. 如果存在非奇异矩阵 P，使得 $B = P^{\mathrm{T}} A P$，称 A 合同于 B，或者说，按这种方式由 A 得到 B 的运算是**合同变换**. 合同变换不改变实对称矩阵的正、负和零特征值的数目.

定理 1.1.5 设 $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{12}^{\mathrm{T}} & B_{22} \end{pmatrix}$，$A_{11}$ 是非奇异方阵，那么在分块矩阵

$$\begin{pmatrix} B_{11} & B_{12} & -A_{11}^{\mathrm{T}} \\ B_{12}^{\mathrm{T}} & B_{22} & -A_{12}^{\mathrm{T}} \\ A_{11} & A_{12} & \mathbf{0} \end{pmatrix}$$

中先后求 A_{11} 和 $-A_{11}^{\mathrm{T}}$ 的逆矩阵是对 B 作合同变换.

证 按照定理 1.1.1，在以上 3×3 分块矩阵中先后求 A_{11} 和 $-A_{11}^{\mathrm{T}}$ 的逆矩阵，结果为

$$\begin{pmatrix} A_{11}^{-\mathrm{T}} B_{11} A_{11}^{-1} & A_{11}^{-\mathrm{T}} B_{12} - A_{11}^{-\mathrm{T}} B_{11} A_{11}^{-1} A_{12} & -A_{11}^{-\mathrm{T}} \\ B_{12}^{\mathrm{T}} A_{11}^{-1} - A_{12}^{\mathrm{T}} A_{11}^{-\mathrm{T}} B_{11} A_{11}^{-1} & B_{22}' & A_{12}^{\mathrm{T}} A_{11}^{-\mathrm{T}} \\ A_{11}^{-1} & -A_{11}^{-1} A_{12} & O \end{pmatrix},$$

其中

$$B_{22}' = B_{22} - B_{12}^{\mathrm{T}} A_{11}^{-1} A_{12} - A_{12}^{\mathrm{T}} A_{11}^{-\mathrm{T}} (B_{12} - B_{11} A_{11}^{-1} A_{12}),$$

$A_{11}^{-\mathrm{T}}$是A_{11}^{-1}的转置;其左上角的 2×2 分块矩阵可表示为

$$\begin{pmatrix} A_{11}^{-\mathrm{T}} & O \\ -A_{12}^{\mathrm{T}} A_{11}^{-\mathrm{T}} & I \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{12}^{\mathrm{T}} & B_{22} \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & -A_{11}^{-1} A_{12} \\ O & I \end{pmatrix},$$

所以是对 B 作的一种合同变换. ∎

1.2 向量与矩阵的范数及其性质

1.2.1 向量与矩阵的范数

设 x 是 n 维向量空间 $\mathbf{R}^n$ 的一个向量. x 的**范数**是一个非负实数, 记为 $\|x\|$, 它满足以下三个条件:

(i) $\|x\| \geqslant 0$,等号仅当 $x=0$ 时成立;

(ii) 对于任意实数 $\alpha \in \mathbf{R}$, $\|\alpha x\| = |\alpha| \, \|x\|$;

(iii) 对于任意向量 $x, y \in \mathbf{R}^n$,有 $\|x+y\| \leqslant \|x\| + \|y\|$.

容易证明

$$\|x\| - \|y\| \leqslant \|x+y\|, \quad \|y\| - \|x\| \leqslant \|x+y\|.$$

所以

$$\big|\, \|x\| - \|y\| \,\big| \leqslant \|x+y\| \leqslant \|x\| + \|y\|,$$

此式称为**三角不等式**.

设 $x = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$,常用的向量范数有以下几种:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|,$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2},$$

$$\|x\|_\infty = \max_i |x_i|.$$

设 A 是 n 阶方阵. A 的范数是一个非负实数, 记为 $\|A\|$, 它满足以下四个条件:

(i) $\|A\| \geqslant 0$,等号仅当 $A=O$ 时成立;

(ii) 对于任意实数 $\alpha \in \mathbf{R}$, $\|\alpha A\| = |\alpha| \, \|A\|$;

(iii) 对于任意 n 阶方阵 A 和 B,有 $\|A+B\| \leqslant \|A\| + \|B\|$;

(iv) $\|AB\| \leqslant \|A\| \|B\|$.

设 $A=(a_{ij})_{n\times n}$,常用的矩阵范数有以下三种:

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|,$$

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^TA)},$$

其中 $\lambda_{\max}(A^TA)$ 表示 A^TA 的最大特征值,

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|.$$

可以证明,当 A 是实对称矩阵时,$\|A\|_2=\max\limits_i|\lambda_i(A)|$,其中 $\lambda_i(A)$ 表示 A 的第 i 个特征值.

对于任意 n 维向量 x 和 n 阶方阵 A,如果

$$\|Ax\| \leqslant \|A\| \|x\|$$

成立,称矩阵范数 $\|A\|$ 与向量范数 $\|x\|$ 相容.

可以证明,$\|A\|_1,\|A\|_2,\|A\|_\infty$ 分别与 $\|x\|_1,\|x\|_2,\|x\|_\infty$ 相容.

1.2.2 有关公式

向量 $x=(x_1,x_2,\cdots,x_n)$ 与 $y=(y_1,y_2,\cdots,y_n)$ 的**内积**是一个实数,记为 $x\cdot y$,规定

$$x\cdot y=x_1y_1+x_2y_2+\cdots+x_ny_n.$$

定义了内积的向量空间 $\mathbf{R}^n$ 称为 n 维欧氏(向量)空间. 欧氏空间中向量 x 的长度为

$$\|x\|_2=(x_1^2+x_2^2+\cdots+x_n^2)^{\frac{1}{2}}.$$

两向量 x 与 y 的距离为

$$\|x-y\|_2=[(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2]^{\frac{1}{2}}.$$

如果 $x\cdot y=0$,称 x 与 y **直交**,记为 $x\perp y$.

可证明以下公式成立:

公式 1.2.1(平行四边形公式) $\|x+y\|_2^2+\|x-y\|_2^2=2\|x\|_2^2+2\|y\|_2^2$;

公式 1.2.2(勾股定理) 当 $x\perp y$ 时,$\|x+y\|_2^2=\|x\|_2^2+\|y\|_2^2$;

公式 1.2.3(柯西-许瓦茨不等式) $|x\cdot y|\leqslant\|x\|_2\|y\|_2$.

为了简单起见,以后我们将范数符号 $\|\cdot\|_2$ 记为 $\|\cdot\|$.