Operations Research and Fuzziolgy 运筹与模糊学, 2013, 3, 35-39
http://dx.doi.org/10.12677/orf.2013.34006 Published Online November 2013 (http://www.hanspub.org/journal/orf.html)

An Extension of Kantorovich Inequality with an Application in the Analysis of Steepest Decent Method*

Wentao Cao, Yong Xia

LMIB of the Ministry of Education, School of Mathematics and System Sciences, Beihang University, Beijing
Email: dearyxia@gmail.com

Received: Apr. 29th, 2013; revised: Oct. 15th, 2013; accepted: Oct. 20th, 2013

Abstract: Based on Kuhn-Tucker condition in optimization theory, we extend the canonical Kantorovich inequality. As an application, the analysis on the convergence rate of steepest descent for minimizing a positive definite quadratic function is extended for the positive semi-definite case.

Keywords: Kantorovich Inequality; Kuhn-Tucker Condition; Steepest Descent; Convergence Rate

Kantorovich 不等式的推广及其在最速下降法分析中的应用*

曹文涛，夏　勇

北京航空航天大学数学与系统科学学院，数学、信息与行为教育部重点试验室，北京
Email: dearyxia@gmail.com

收稿日期：2013 年 4 月 29 日；修回日期：2013 年 10 月 15 日；录用日期：2013 年 10 月 20 日

摘　要: 本文利用最优化理论中经典的 Kuhn-Tucker 条件证明并推广了 Kantorovich 不等式。作为应用，将极小化正定二次函数的最速下降法的收敛速度分析推广到半正定情形。

关键词：Kantorovich 不等式；Kuhn-Tucker 条件；最速下降法；收敛速度

1. 引言

Kantorovich 不等式以前苏联著名的经济学家、数学家、诺贝尔奖得主 Leonid Kantorovich 的名字命名，在最优化理论、统计分析中有重要应用[1,2]。

定理 1：(Kantorovich)设 A 为 $n\times n$ 正定阵，λ_1 和 λ_n 分别为其最大和最小特征值，则对任意非零向量 x，有

$$\frac{\left(x^{\mathrm{T}}Ax\right)\left(x^{\mathrm{T}}A^{-1}x\right)}{\left(x^{\mathrm{T}}x\right)^2}\le\frac{\left(\lambda_1+\lambda_n\right)^2}{4\lambda_1\lambda_n}$$

有趣的是，Kantorovich 不等式是如下 Cauchy-Schwarz 不等式[1,2]

$$\left(x^{\mathrm{T}}x\right)^2\le x^{\mathrm{T}}Axx^{T}A^{-1}x$$

*资助信息：国家自然科学基金(11001006)。

的“逆”形式，即对任意 $x \neq 0$，Kantorovich 不等式与 Cauchy-Schwarz 不等式可以整合成

$$1 \le \frac{\left(x^{\mathrm{T}} A x\right)\left(x^{\mathrm{T}} A^{-1} x\right)}{\left(x^{T} x\right)^{2}} \le \frac{\left(\lambda_1 + \lambda_n\right)^2}{4\lambda_1\lambda_n}$$

Kantorovich 不等式有很多推广[2,3]，比如将 $n\times 1$ 向量 x 替换成 $n\times p$ 矩阵 x 等。其中一个推广如下：

定理 2：(Greub-Rheinboldt)设 A 和 B 为两个正定阵，且 $AB = BA$，记 $\lambda_1 \ge \cdots \ge \lambda_n$ 和 $\mu_1 \ge \cdots \ge \mu_n$ 分别为 A 和 B 的特征值，则对任意非零向量 x，有

$$\frac{\left(x^{\mathrm{T}} A^2 x\right)\left(x^{\mathrm{T}} B^2 x\right)}{\left(x^{\mathrm{T}} ABx\right)^{2}} \le \frac{\left(\lambda_1\mu_1 + \lambda_n\mu_n\right)^2}{4\lambda_1\mu_1\lambda_n\mu_n}$$

需要指出的是，定理 2 的估计并不紧。Kantorovich 不等式中令 x 为 $A^{1/2}B^{1/2}y$，令 A 为 AB^{-1}，则有

$$\frac{\left(x^{\mathrm{T}} A^2 x\right)\left(x^{\mathrm{T}} B^2 x\right)}{\left(x^{\mathrm{T}} ABx\right)^{2}} \le \frac{\left(\lambda_1\left(AB^{-1}\right) + \lambda_n\left(AB^{-1}\right)\right)^2}{4\lambda_1\left(AB^{-1}\right)\lambda_n\left(AB^{-1}\right)} = \frac{1}{2} + \frac{\lambda_1\left(AB^{-1}\right)}{\lambda_n\left(AB^{-1}\right)} + \frac{\lambda_n\left(AB^{-1}\right)}{\lambda_1\left(AB^{-1}\right)} \tag{1}$$

其中 $\lambda_1\left(AB^{-1}\right)$，$\lambda_n\left(AB^{-1}\right)$ 分别为 AB^{-1} 的最大和最小特征值。注意到式(1)上界分别是 $\lambda_1\left(AB^{-1}\right)$ 和 $\lambda_n\left(AB^{-1}\right)$ 的递增函数和递减函数，利用 $\frac{\lambda_n}{\mu_1} \le \lambda_n\left(AB^{-1}\right) \le \lambda_1\left(AB^{-1}\right) \le \frac{\lambda_1}{\mu_n}$ 放大替换便得到 Greub-Rheinboldt 不等式。

Kantorovich 不等式一个著名的应用是刻画了最速下降法的收敛性速度。

本文我们利用最优化中的 Kuhn-Tucker 条件给出了 Kantorovich 不等式一个新的证明，该证明方法同时推广了 Kantorovich 不等式，最后，作为推广的 Kantorovich 不等式的应用，我们将目前对极小化正定二次函数的最速下降法收敛性分析推广到了半正定情形。

2. Kantorovich 不等式的推广

本节中我们推广 Kantorovich 不等式，得到如下主要结果：

定理 3：设 A 和 B 为 n 阶半正定阵，且 $AB = BA$，记 $a_1, a_2, \cdots, a_n$ 和 $b_1, b_2, \cdots, b_n$ 分别为 A 和 B 的特征值，则有

$$\frac{\left(x^{\mathrm{T}} A x\right)\left(x^{\mathrm{T}} B x\right)}{\left(x^{\mathrm{T}} x\right)^{2}} \le \max\left\{-\frac{\left(b_i a_j - a_i b_j\right)^2}{4\left(a_i - a_j\right)\left(b_i - b_j\right)}\left(\text{其中} a_i \ne a_j, b_i \ne b_j, \text{且} b_i a_j - a_i b_j \ne 0\right),\ a_i b_i\right\}$$

证明：我们考虑最优化问题

$$\max \frac{\left(x^{\mathrm{T}} A x\right)\left(x^{\mathrm{T}} B x\right)}{\left(x^{\mathrm{T}} x\right)^{2}}$$

由于分子和分母是 x 的齐次函数，所以可以归一化，同时由于 $AB = BA$，矩阵 A，B 有相同的特征向量，因此，上述问题等价为

$$\max\left(\sum\nolimits_{i=1}^{n} a_i x_i^2\right)\left(\sum\nolimits_{i=1}^{n} b_j x_i^2\right) \ \text{s.t.}\ x^{\mathrm{T}} x = 1 \tag{2}$$

其 Kuhn-Tucker 条件为

$$\left(\sum_{i=1}^{n} a_i x_i^2\right) b_j x_i + \left(\sum_{i=1}^{n} b_j x_i^2\right) a_i x_i = \lambda x_i,\ i = 1, 2, \cdots, n,\ x^{\mathrm{T}} x = 1$$

据此可得

$$\lambda = 2\left(\sum_{i=1}^{n} a_i x_i^2\right)\left(\sum_{i=1}^{n} b_j x_i^2\right)$$

设

$$I^* = \{i : x_i \neq 0\}，\ y = \left(\sum_{i=1}^{n} a_i x_i^2\right),\ \ z = \left(\sum_{i=1}^{n} b_j x_i^2\right).$$

其中 $\frac{\lambda}{2}$ 恰为最优目标函数值。进一步，Kuhn-Tucker 条件可化简为如下必要条件：

$$b_j y + a_i z = 2yz，\ \forall i \in I^*. \tag{3}$$

记 $\mathrm{Card}\{I^*\}$ 为 I^* 中元素的个数，下面分别根据 $\mathrm{Card}\{I^*\}$ 进行讨论。

1) $\mathrm{Card}\{I^*\} = 1.$ 假设 $i_0 \in I^*$，则任意 $i \notin I^*$ 有 $i \neq i_0$。因此

$$x_{i_0} = 1,\ \ y = a_{i_0},\ \ z = b_{i_0}.$$

此时，问题(2)的最优值为

$$\max_{i_0 = 1,2,\cdots,n} a_{i_0} b_{i_0}$$

2) $\mathrm{Card}\{I^*\} = 2.$ 设 $i_1 \in I^*$ 和 $i_2 \in I^*$. (3)等价为

$$b_{i_1} y + a_{i_1} z = b_{i_2} y + a_{i_2} z = 2yz$$

下面分如下几种情况讨论。

a) 若 $b_{i_1} = b_{i_2}$，$a_{i_1} = a_{i_2}$，

则由 $x_{i_1}^2 + x_{i_2}^2 = 1$ 知 $y = a_{i_1}, \mathrm{z} = b_{i_1}$，问题(2)的最优解为

$$\max_{i_0 = 1,2,\cdots,n} a_{i_0} b_{i_0}$$

b) 若 $b_{i_1} \neq b_{i_2}, a_{i_1} = a_{i_2}$，由(3)知道 $y = 0$，即 $\sum_{i=1}^{n} a_i x_i^2 = a_{i_1} x_{i_1}^2 + a_{i_2} x_{i_2}^2 = 0$，

由于 $i_1 \in I^*$ 和 $i_2 \in I^*$，我们有 $a_{i_1} = a_{i_2} = 0$，此时最优值为 0.

c) 若 $b_{i_1} = b_{i_2}, a_{i_1} \neq a_{i_2}$，类似可知最优值为 0.

d) 若 $b_{i_1} \neq b_{i_2}, a_{i_1} \neq a_{i_2}$ 且 $\begin{bmatrix} b_{i_1} & a_{i_1} \\ b_{i_2} & a_{i_2} \end{bmatrix}$ 奇异，此时(3)无解。

e) 若 $b_{i_1} \neq b_{i_2}, a_{i_1} \neq a_{i_2}$，$\begin{bmatrix} b_{i_1} & a_{i_1} \\ b_{i_2} & a_{i_2} \end{bmatrix}$ 非奇异。解 Kuhn-Tucker 条件方程组(3)，得

$$\begin{cases} y = \dfrac{b_{i_1} a_{i_2} - a_{i_1} b_{i_2}}{2\left(b_{i_1} - b_{i_2}\right)} \\ z = -\dfrac{b_{i_1} a_{i_2} - a_{i_1} b_{i_2}}{2\left(a_{i_1} - a_{i_2}\right)} \end{cases}$$

此时，问题(2)的最优解为

$$\max_{i_1, i_2} -\frac{\left(b_{i_1} a_{i_2} - a_{i_1} b_{i_2}\right)^2}{4\left(a_{i_1} - a_{i_2}\right)\left(b_{i_1} - b_{i_2}\right)}$$

3) $\mathrm{Card}\{I^*\}\geq 3$ 注意到(3)左边的系数矩阵的阶为 $\mathrm{Card}\{I^*\}\times 2$，而右边项均相等。因此该系数矩阵的任何一行均为其它两行(权和为 1)的加权组合，否则无解。对该系数矩阵分下列情况讨论：

a) 该系数矩阵存有两行线性无关。如果满足 2)中情形 d)，那么无解；如果满足 2)中情形 e)，那么这两行已然确定了唯一解。

b) 该系数矩阵任意两行线性相关，且满足 2)中情形 a)。则系数矩阵的所有行均相等，易知(3)的解同于 2)中情形 a)。

c) 该系数矩阵任意两行线性相关，且满足 2)中情形 b)。则系数矩阵的所有行的 a_i 列均相等，进一步类比 2)中情形 b)可知所有的 a_i 列均为 0，同时(3)的最优值为 0.

d) 该系数矩阵任意两行线性相关，且满足 2)中情形 c)。此时同上讨论，系数矩阵的所有行的 b_i 列均相等且为 0，同时(3)的最优值为 0.

综合 1)、2)和 3)，问题(2)的目标最优值为

$$\max\left\{-\frac{\left(b_ia_j-a_ib_j\right)^2}{4\left(a_i-a_j\right)\left(b_i-b_j\right)}\left(\text{其中}a_i\neq a_j,b_i\neq b_j,\text{且}b_ia_j-a_ib_j\neq 0\right),\ a_ib_i\right\}$$

定理证毕。

作为定理 3 的推论，Kantorovich 不等式立得证明：

$$\frac{\left(x^{\mathrm{T}}Ax\right)\left(x^{\mathrm{T}}A^{-1}x\right)}{\left(x^{\mathrm{T}}x\right)^2}\leq\max\left\{-\frac{\left(\lambda_i\dfrac{1}{\lambda_j}-\lambda_j\dfrac{1}{\lambda_i}\right)^2}{4\left(\lambda_i-\lambda_j\right)\left(\dfrac{1}{\lambda_i}-\dfrac{1}{\lambda_j}\right)},1\right\}=\max\left\{\frac{\left(\lambda_i+\lambda_j\right)^2}{4\lambda_i\lambda_j},1\right\}=\frac{\left(\lambda_1+\lambda_2\right)^2}{4\lambda_1\lambda_2}$$

3. 推广 Kantorovich 不等式的在最速下降法分析中的应用

记对称矩阵 A 的 Moore-Penrose 广义逆为 A^+，对 A 的特征值 λ， A^+ 的相应特征值为 $\lambda^+=\begin{cases}\lambda^{-1}, & \lambda\neq 0\\ 0, & \lambda\neq 0\end{cases}$ 作为定理 3 的推论，我们有

定理 4：设 A 为 n 阶半正定阵，记 $\lambda_1\geq\cdots\geq\lambda_r$ 为 A 的正特征值，则对任意非零向量 x，

$$\frac{\left(x^{\mathrm{T}}Ax\right)\left(x^{\mathrm{T}}A^+x\right)}{\left(x^{\mathrm{T}}x\right)^2}\leq\frac{\left(\lambda_1+\lambda_r\right)^2}{4\lambda_1\lambda_r}$$

本节应用定理 4 推广最速下降法收敛速度分析。设 $f(x)$ 为半正定二次函数，Hessen 阵为 G，极小化 $f(x)$ 的使用精确步长的最速下降法迭代公式为

$$x_{(k+1)}=\begin{cases}x_{(k)}=x^*, & \text{如果}g_k=0\\ -\infty, & \text{如果}g_k\neq 0,g_k^{\mathrm{T}}Gg_k=0\\ x_{(k)}-\dfrac{g_k^{\mathrm{T}}g_k}{g_k^{\mathrm{T}}Gg_k}g_k, & \text{如果}g_k^{\mathrm{T}}Gg_k\neq 0\end{cases}$$

其中 g_k 为 $f(x)$ 在 $x_{(k)}$ 处的梯度，x^* 为最优解。上述迭代或者有限步内发散到负无穷，或者有限步中止并得到最优解 x^*，或者收敛到最优解 x^*，且 $g\left(x^*\right)=0$ 由 Taylor 展开，有

$$f(x)=f\left(x^*\right)+\frac{1}{2}\left(x-x^*\right)G\left(x-x^*\right)$$

则

$$
\begin{aligned}
& f\left(x_{(k+1)}\right)-f\left(x^*\right) \\
& =\frac{1}{2}\left(x_{(k+1)}-x^*\right)G\left(x_{(k+1)}-x^*\right) \\
& =\frac{1}{2}\left(x_{(k)}-\frac{g_k^{\mathrm{T}}g_k}{g_k^{\mathrm{T}}Gg_k}g_k-x^*\right)G\left(x_{(k)}-\frac{g_k^{\mathrm{T}}g_k}{g_k^{\mathrm{T}}Gg_k}g_k-x^*\right) \\
& =\frac{1}{2}\left(x_{(k)}-x^*\right)G\left(x_{(k)}-x^*\right)-\frac{g_k^{\mathrm{T}}g_k}{g_k^{\mathrm{T}}Gg_k}g_k^{\mathrm{T}}G\left(x_{(k)}-x^*\right)+\frac{1}{2}\left(\frac{g_k^{\mathrm{T}}g_k}{g_k^{\mathrm{T}}Gg_k}\right)^2 g_k^{\mathrm{T}}Gg_k
\end{aligned}
$$

由于

$$G\left(x_{(k)}-x^*\right)=g\left(x_{(k)}\right)-g\left(x^*\right)=g\left(x_{(k)}\right)=g_k$$

所以

$$f\left(x_{(k+1)}\right)-f\left(x^*\right)=f\left(x_{(k)}\right)-f\left(x^*\right)-\frac{1}{2}\frac{\left(g_k^{\mathrm{T}}g_k\right)^2}{g_k^{\mathrm{T}}Gg_k}$$

又因为

$$g_k^{\mathrm{T}}G^+g_k=\left(x_{(k)}-x^*\right)^{\mathrm{T}}GG^+G\left(x_{(k)}-x^*\right)$$

所以

$$f\left(x_{(k+1)}\right)-f\left(x^*\right)=\left(f\left(x_{(k)}\right)-f\left(x^*\right)\right)\left[1-\frac{\left(g_k^{\mathrm{T}}g_k\right)^2}{g_k^{\mathrm{T}}Gg_k g_k^{\mathrm{T}}G^+g_k}\right]$$

即有

$$f\left(x_{(k+1)}\right)-f\left(x^*\right)\le\left(\frac{\lambda_1-\lambda_r}{\lambda_1+\lambda_r}\right)^2\left(f\left(x_{(k)}\right)-f\left(x^*\right)\right)$$

其中 $\lambda_1\ge\cdots\ge\lambda_r>0$ 为 G 的所有正特征值。上述不等式刻画了函数值 $f\left(x_{(k)}\right)$ 收敛到最优值 $f\left(x^*\right)$ 的线性收敛速度，教科书中的类似结果只是针对 G 正定的二次函数情形，参见[4]。从该推广的结论我们看到，收敛速度与 G 的零特征值个数似乎没有关系，不过如果只有一个非零特征值，那么仅一步迭代即可收敛。

4. 结论

本文探讨 Kantorovich 不等式。指出推广之一的 Greub-Rheinboldt 不等式比 Kantorovich 不等式本身来得更弱。本文利用最优化理论中经典的 Kuhn-Tucker 条件证明并实质推广了 Kantorovich 不等式。作为应用，将极小化正定二次函数的最速下降法的收敛速度分析推广到了半正定情形。

参考文献 (References)

[1] 匡继昌 (2010) 常用不等式. 山东科学技术出版社, 济南.

[2] 王松桂, 吴密霞, 贾忠贞(2006) 矩阵不等式. 科学出版社, 北京.

[3] Huang, J. and Zhou, J. (2005) A direct proof and a generalization for a Kantorovich type inequality. *Linear Algebra and its Applications*, **397**, 185-192.

[4] 刘红英, 夏勇, 周水生 (2012) 数学规划基础. 北京航空航天大学出版社, 北京.