多变量选优方法的比较理论

赖 万 才
（华 侨 大 学）

如文献[1]中所指出的：到目前为止，对于两个变量或两个以上的多变量的选优问题，还没有一个如处理单变量问题的分数法那样研究得比较成熟的最优的选优方法，甚至对两个不同的方法如何比较其优劣的问题，在理论上也还没有解决．一般只是通过各种实际例子的比较，相对地来比较它们的效果．在本文中，我们严格地建立了一种比较任何两个方法优劣的理论．这里对于 Wilde 和 Beightler[2] 所说的"不能找到一种与试验的偶然性无关的效率测度"的问题，或者 Beveridge 和 Schechter[3] 综合许多优选学家评论的"不存在一个判别法来定义各个数值最优化方法的效果，因而不可能找到一个比其它方法都好的方法"的问题，我们做了这样的回答：首先，我们以最简单的例子来否定了被很多著名的优选学家以为真理的"采用区间缩减率的测度对多维区域往往只能消去一段段的直线[2]"的论断，并且对于一般情形证明每一种选优方法都能达到缩减试验区域一部分的目的（目前尚未见到有人指出过一个不依赖于导数或者差商就能达到这一目的的选优方法．而当导数无从求得或者计算差商无足够接近的试验点时，已有的缩减方法不能用）；其次，我们克服了对于试验区域经过几次缩减后，留下的不定性区域的形状不规则，在高维空间中难于摹想和数学描述的困难[2]；最后，我们找到了一种与试验的偶然性无关的、能够估计最优的已试点到最优点的距离的效率测度，在这种测度下，可以比较任何两个选优方法的优劣，也可以定义 n 个变量时与单变量的分数法，分批试验法和黄金分割法相对应的方法，我们虽然还无法对于比较一般的情形具体找出最优的方法，但在理论上已经相当好地回答了文献[4]中论及的最基本的问题，在应用上已经能够用来指出目前的各个方法在选点上的不合理之处，并使 Newman[2] 的新点取在留下区域的中心处的方法得以一般化．

§1. 把 n 维欧氏空间 E^n 中的点 $(x_1, x_2, \cdots, x_n)$ 记作X．试验区域规范化为 E^n 中的单位闭室 I^n：$0 \leqslant x_i \leqslant 1, i = 1, 2, \cdots, n$．因为直接的数值方法要求保证区间缩减永远可行，所以假设目标函数 $f(X)$ 在 I^n 中的每一直线段上为单峰是必要的．这样的函数恒可保持最优点（极大点）不变地延拓为在 E^n 中的每一直线上为单峰．后者全体所成的类记为 Ω．

设 Ω 中的函数在各点的值均可由试验测得．

对于 E^n 中的任意区域（可以是 E^n 自身，也可以是开的、闭的或包含部分边界）Δ 和 $\dot{X} \in \bar{\Delta}$，所谓 Δ 为一关于 $\dot{X}$ 成星形的区域，是指 Δ 上异于 $\dot{X}$ 的任何点均可用全在 Δ 中的直线段与 $\dot{X}$ 连接．

下例说明了"采用区间缩减率的测度对多维区域往往只能消去一段段的直线"的论断是错误的．

设 $f \in \Omega$．对于任意给定的不在同一直线上的三点 X_0，X_1，X_2，若成立 $f(X_0) \leqslant f(X_i)$

本文 1979 年 3 月 30 日收到．

$(i=1,2)$，则 $f(X)$ 的最优点必不在半直线 l_1: $X=X_0+\lambda(X_1-X_0)$, $\lambda\leqslant 0$ 和半直线 l_2: $X=X_0+\lambda(X_2-X_0)$, $\lambda\leqslant 0$ 所夹的平面闭角域 D 上。

证. 由所设的单峰性知道，$f(X)$ 的最优点不可能在 l_1 和 l_2 上，如果 $f(X)$ 的最优点落在 D 的内部 X^* 处，那末连接 X_2 和 X^* 的直线段必与 l_1 有交点，设为 $\tilde{X}$（显然 $\tilde{X}\neq X_0$）. 于是，一方面由所设的单峰性知道

$$f(X_2)<f(\tilde{X})<f(X^*). \qquad (*)$$

另一方面，据 $f(X)$ 在连接 X_1 和 $\tilde{X}$ 的直线段上为单峰的假设，因为 $f(X_0)\leqslant f(X_1)$ 且 $\tilde{X}\neq X_0$，所以 $f(\tilde{X})<f(X_0)$. 由此联合 $f(X_0)\leqslant f(X_2)$ 得到 $f(\tilde{X})<f(X_2)$. 这与 $(*)$ 矛盾. 证毕.

下面的定理叙述了 n 个变量时与单变量单峰函数所具有的基本特性相对应的特性：

定理 1 设 $f\in\Omega$，对于任意给定的 s 个点 $X_1, X_2,\cdots X_s$，用 P_k 表示 E^n 中形如

$$X=X_k+\sum_{\substack{1\leqslant l\leqslant s\\ f(X_l)\geqslant f(X_k)}}\lambda_l(X_l-X_k),\ \lambda_l\leqslant 0 \qquad (1)$$

的一切点所成的集合，并置

$$P=P(X_1,X_2,\cdots,X_s;f)=\bigcup_{k=1}^{s}{}' P_k,$$

其中和号带撇表示不对退化成一点 X_k 的 P_k 求和*. 那末 P 上决不会有 $f(X)$ 的最优点.

§2 设 r 是 $X_1, X_2,\cdots, X_s$ 中适合 $f(X_l)\geqslant f(X_k)$ 的非零向量 X_l-X_k 的全体 $\{X_{l_i}-X_k, i=1,2,\cdots,t\}$ 的极大线性无关组所含的向量个数. 设 h 是对向量组 $\{X_{l_i}-X_k,\ i=1,2,\cdots,t\}$ 中的每一个向量 $X_{l_i}-X_k$，施行正数倍的伸缩，使其终点落在同一个 $r-1$ 维的超平面上时，包含这 t 个终点的最小凸多面形的顶点个数. 数组 $\{r,h\}$ 称为 P_k 的特征.

定理 2 $G=G(X_1,X_2,\cdots,X_s;\ f)=I^n-P$ 是一个由 I^n 去掉一些超棱锥的和集之后的区域. 更精确地说，特征为 $\{r,h\}$ 的 P_k 由 $h-r+1$ 个两两无公共内点的、以 X_k 为顶点的、r 维 r 棱的无限延伸的超棱锥构成（r 维 r 棱的无限延伸的超棱锥当 $r=1,2,3$ 和 n 时，分别是半直线、角平面、通常的无限延伸的三棱锥和 E^n 在斜角坐标系 $\{o;e_1,e_2,\cdots,e_n\}$ 下各坐标均取正值的那个挂限. G 关于 $X_1,X_2,\cdots,X_s$ 中的每一个极大点成星形.

§3 对于任意给定的正整数的无穷序列 $k_1,k_2,\cdots k_i,\cdots$，所谓在 I^n 上给定了一个试验策略 φ，就是给定了一列映照 $\{\varphi_i\}$: φ_i $(i=1,2,3\cdots)$ 对于 I^n 上带有至多一个例外点的任意星形区域有定义，且把每一个这样的区域映成自身（保持例外点和边界点不动）带上另 k_i 个例外点，φ 与 I^n、$f(X)$ 有如下的结合：

φ_1 把 I^n 映成自身带上 k_1 个例外点 $X_1^{(1)}, X_2^{(1)},\cdots, X_{k_1}^{(1)}$. 在这些点上做第一批试验，并比较其上的试验结果（$f(X)$ 的函数值）. 根据定理 1，我们把 $G(X_1^{(1)}, X_2^{(1)},\cdots, X_{k_1}^{(1)}; f)$ 称为 f 在 φ 下做了第一批试验后的留下区域，记作 $\Delta_1(\varphi,f)$. 由定理 1 中关于 P 的定义知道，它至多只包含一个例外点——唯一的最优的已试点. 由定理 2，它是一个星形区域.

依次对于 $i=2,3,4,\cdots$，φ_i 把 $\Delta_{i-1}(\varphi,f)$ 映成自身带上另 k_i 个例外点 $X_1^{(i)}, X_2^{(i)},\cdots, X_{k_i}^{(i)}$，在这些点上做第 i 批试验，并比较头 i 批的试验结果. 易见 $\Delta_{i-1}(\varphi,f)-P=I^n-P$，我们把 $G(X_1^{(1)}, X_2^{(1)},\cdots, X_{k_1}^{(1)},\cdots, X_1^{(i)}, X_2^{(i)},\cdots, X_{k_i}^{(i)};f)$ 称为 f 在 φ 下做了第 i 批试验后的留下区域、记作 $\Delta_i(\varphi,f)$. 它至多只包含一个例外点——唯一的最优的已试点，且是一个星形区域.

我们把 $\Delta_i(\varphi,f)$ 上的一个其内部不包含已试点的 n 维最大开球称为 $\Delta_i(\varphi,f)$ 的一个核，

---

* 显然，当且仅当在 f 下 X_k 是 $X_1, X_2,\cdots, X_s$ 中的唯一极大点时，P_k 退化成一点 X_k.

它的直径记为 $d_i(\varphi, f)$。称 $d_i(\varphi)=\sup\limits_{f\in\Omega} d_i(\varphi, f)$ 为试验策略 φ 在第 i 批试验后的精度。

定理 3 设 φ 是 I^n 上的一个试验策略，当 $i\to\infty$ 时，具有 $d_i(\varphi)\to 0$。那末对于任何 $f\in\Omega$，f 在 φ 下做了第 i 批试验后的最优的已试点，必趋于 f 的最优点。

§ 4 上述定理 3 虽然说了对任何 $f\in\Omega$，它在 φ 下做了第 i 批试验后的最优的已试点一定随 $d_i(\varphi)\to 0$ 而趋于 f 的最优点。但对于某些 f，收敛的速度可以是很慢的。收敛的快慢取决于 f 在 I^n 上的正规程度。后者被如下定义的 f 在 I^n 上的偏差系数刻划着。

对于每一个 $X_0\in I^n$，用 $U_f(X_0)$ 和 $V_f(X_0)$ 分别表示 f 的最优点到 f 过 X_0 的等高面与 I^n 的交集的最大距离和最小距离。我们称 $\tau_f=\sup\limits_{X_0\in I^n}\dfrac{U_f(X_0)}{V_f(X_0)}$ 为 f 在 I^n 上的偏差系数。

下面的定理述出了最优的已试点收敛于最优点的速度。

定理 4 对于 I^n 上的任何试验策略 φ 和任何 $f\in\Omega$，设 X^* 是 f 的最优点。$X^{(i)}$ 是 f 在 φ 下做了第 i 批试验后的一个最优的已试点，那末

$$|X^{(i)}-X^*|\leqslant\frac{\tau_f}{2}d_i(\varphi, f)\leqslant\frac{\tau_f}{2}d_i(\varphi).$$

易知，在考虑用不同的试验策略对 f 做试验时，并不需要知道 f 的偏差系数 τ_f。

§ 5 我们在对两个试验策略的第 i 步进行比较时，总是假定它们头 i 批的试验个数的总和 $k_1+k_2+\cdots+k_i$ 是相同的。

对于试验策略 φ 和 ψ，若成立 $d_i(\varphi)\leqslant d_i(\psi)$，则说 φ 在第 i 步不劣于 ψ；若成立 $d_i(\varphi)<d_i(\psi)$，则说 φ 在第 i 步优于 ψ；若成立 $d_i(\varphi)=d_i(\psi)$，则说 φ 和 ψ 在第 i 步等价。

n 个变量时与单变量的分批最优法相对应的方法，就是在某一步不劣于任何试验策略的方法，这样的策略特别当 $k_i=1, i=1,2,\cdots,m$ 时，就是 n 个变量时与单变量的 m 次试验的分数法相对应的方法。n 个变量时与单变量的黄金分割法相对应的方法，可以定义为在任何相邻的两步中，至少有一步不劣于任何试验策略的方法（这定义基于当 $n=1$，$k_i=1$ 时，在任何相邻的两步中至少有一步不劣于任何试验策略的方法，有且只有黄金分割法）。

§ 6 对于目前的各个方法，可以逐一用本文的理论来检验其在选点上的不合理之处。如熟知的正规单纯形法，会出现步长本来已经不大，可是留下区域的核却还很大，就要进一步缩短步长的不合理现象。

单变量的分批最优法，实质上是将试验区间均分，从分点中去找出最优者来近似代替最优点。对于多变量，将试验区域均分，从分点中去找出最优者来近似代替最优点的情况，无疑具有重要的实用意义。由定理 1 知道，经过一些试验后，留下点组的确定和留下点组的核心点的确定，都归结为纯分析的计算问题了。于是，在无从得知更好方法的情况下，我们把新点取在留下点组的核心点处是可行的。

参 考 文 献

[1] 北京市优选法应用推广小组，优选法及其应用，人民出版社，1972，138.
[2] D. J. 华尔德，C. S. 皮特勒著(尤云程译)，优选法基础，科学出版社，1978，162—167、176.
[3] G. S. G. Beveridge, & Schechter R. S., *Optimization: Theory and Practice*, McGraw-Hill, New-York, 1970, 502.
[4] 中国科学院数学研究所运筹室优选法小组，优选法，科学出版社，1978，144—147.