

论优化方法的可视化及教学

谢益辉 *

摘要

知识的真正获得往往要建立在理解的基础上，而几何方式一般比代数方式更容易让人理解，因为我们可以通过对几何图形的观察更加清楚代数推导以及方法实施的过程。本文利用 R 语言结合若干优化方法给出这些方法的动态几何图形展示（动画），并通过对动态图形的观察提出优化方法在应用以及教学上的建议。

关键词：优化方法；可视化；动画；梯度；单纯形；数据包络分析；R 语言

§1 引言

一般说来，优化方法常常有其自身的几何解释，例如常见的牛顿法、梯度下降法以及二分法，但我们的教科书由于种种限制不能充分展示出这些几何形象，这些限制包括：版面的限制（无法排出太多图形）、纸张的限制（无法打印出动态图形以及彩纸的昂贵）、学术规则的限制（以数学推导为主）等等。然而，伴随着现在计算机技术的高速发展，数值计算在计算机上的实现变得越来越简单，而且各种可视化软件也不断涌现，为我们理解数学模型和展示数学方法提供了很大的便利。这些软件中，有早期出现的 Java 语言，也有近期出现的 Processing 语言¹等。本文要介绍的语言是一门统计语言，即：R 语言 (R Development Core Team, 2008)。

“R”是一款优秀的统计软件，同时也是一门统计计算与作图的语言，它最初由奥克兰大学统计学系的 Ross Ihaka 和 Robert Gentleman 编写 (Ihaka and Gentleman, 1996)；自1997年 R 开始由一个核心团队 (R Core Team) 开发，这个团队的成员大部分来自大学机构（统计及相关院系），包括牛津大学、华盛顿大学、威斯康星大学、爱荷华大学、奥克兰大学等，除了这些作者之外，R 还拥有一大批贡献者（来自哈佛大学、加州大学洛杉矶分校、麻省理工大学

^{*}中国人民大学统计学院明德主楼 1037, 北京, 100872; Email: xieyihui@gmail.com

¹作者：Casey Reas 和 Benjamin Fry；网址：<http://www.processing.org>

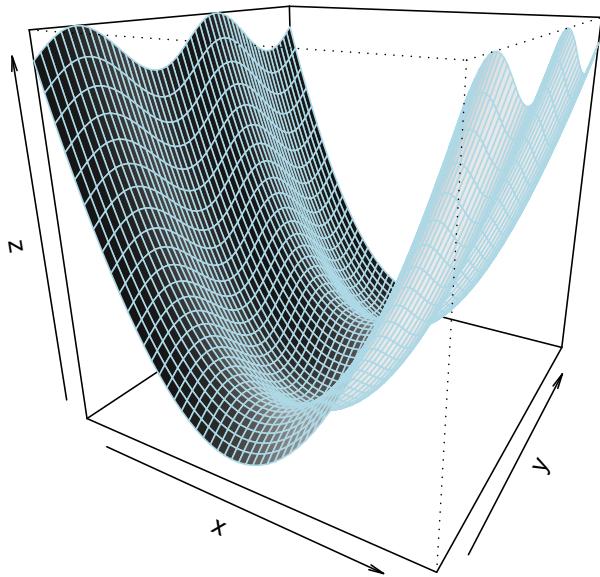


图 1 函数 $f(x, y) = x^2 + 3\sin(y)$ 的三维透视图

等)，他们为 R 编写代码、修正程序缺陷和撰写文档。迄今为止，R 中的程序包 (Package) 已经是数以千计，各种统计前沿理论方法的相应计算机程序都会在短时间内以软件包的形式得以实现，这种速度是其它统计软件无法比拟的。除此之外，R 还有一个重要的特点，那就是它是免费、开源的。

作者从 2007 年 11 月开始关注各种统计计算、数值计算以及优化方法的可视化问题，并利用 R 语言编写了一个程序包 **animation** (Xie, 2008)，该程序包致力于用动画的方式将数学方法的主要思想表达出来，以使得初学者更快理解方法的本质。

§2 优化方法的可视化

首先我们需要寻找动画与优化方法的共同点，然后才能进行具体的可视化工作。显然，这二者都有一个很明显的特点，那就是它们一般都由多个步骤构成：动画由一幅一幅静态图形构成，而优化方法也往往由多次迭代才能找到最优值。基于这个特点，动画的思路则非常明确。我们只需要在一个固定的图形框架下将优化方法的每一步计算结果以图形的方式表达出来即可。

由于 R 在数值计算以及统计作图上的巨大优势，我们可以很快将上述想法付诸实践。在 R 程序代码中，优化算法可以表达为一个循环结构，循环的每一步我们生成一幅新的图形，并且稍作停顿（例如 0.5 秒），这样下来，随

着程序的执行，整个画面就会动起来。综上，动画的基本原理很简单，关键是以何种方式表达计算的思想和细节问题。下面我们以两种常见的优化方法梯度下降法和单纯形法为例，说明动画设计的过程和结果。

§2.1 梯度下降

梯度下降算法的基本思想是从一个随机起点沿着负梯度的方向前进，这样一直下去就会走到函数的局部最小值处。对于一个实值函数 $F(\mathbf{x})$ ，假设它在点 \mathbf{a} 附近一阶可导，那么 $F(\mathbf{x})$ 的值沿其负梯度方向 $-\nabla F(\mathbf{a})$ 下降最快，用数学式子表达出来就是：

如果

$$\mathbf{b} = \mathbf{a} - \gamma \nabla F(\mathbf{a}) \quad (1)$$

(其中 $\gamma > 0$ 是一个足够小的值) 那么将有 $F(\mathbf{a}) \geq F(\mathbf{b})$ 。基于这个事实，我们可以从任意一个值 \mathbf{x}_0 开始，生成一个序列 $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$ 使得：

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \gamma_n \nabla F(\mathbf{x}_n), n \geq 0. \quad (2)$$

那么我们可以得到：

$$F(\mathbf{x}_0) \geq F(\mathbf{x}_1) \geq F(\mathbf{x}_2) \geq \dots, \quad (3)$$

序列 $\{\mathbf{x}_n\}$ 将会收敛到一个局部最小值。

该算法除了求梯度以外几乎都是简单的四则运算，而 R 本身提供了对函数求导的功能²，因此梯度下降算法用 R 处理起来很方便；在 **animation** 程序包中作者提供了 *grad.desc()* 函数用来生成任意可导的二维函数的优化过程动画，此处不详细讨论程序问题，仅仅给出两个示例，用以说明数值计算的过程。

例 1 我们先来看一个简单的例子：目标函数为 $f(x, y) = x^2 + 3\sin(y)$ ，其三维透视图如图 1 所示，为了观察梯度下降的过程，我们并不直接使用三维图形，而是用等高线图，所谓梯度，就是与等高线垂直的方向。

图 2 的一系列小图充分展示了梯度下降的迭代路径。我们从点 $[-2\pi, 2]$ 出发，沿着梯度的方向以箭头的形式表达迭代计算的过程，经过 43 次迭代，程序最终到点 $[-0.06770298, 4.70141282]$ ，相应的函数值为 -2.995236 。

例 2 这里我们取一个复杂的目标函数 $f(x, y) = \sin(1/2 * x^2 - 1/4 * y^2 + 3) * \cos(2 * x + 1 - e^y)$ ，并取不同的迭代起点，从图 3 不难看出，不同的起点最终迭代的结果并不完全一样，大多数点都迭代到了一个局部最小值处，但是有些点并不能迭代到最小值，甚至还有些点到最后并不一定收敛。这个例子告

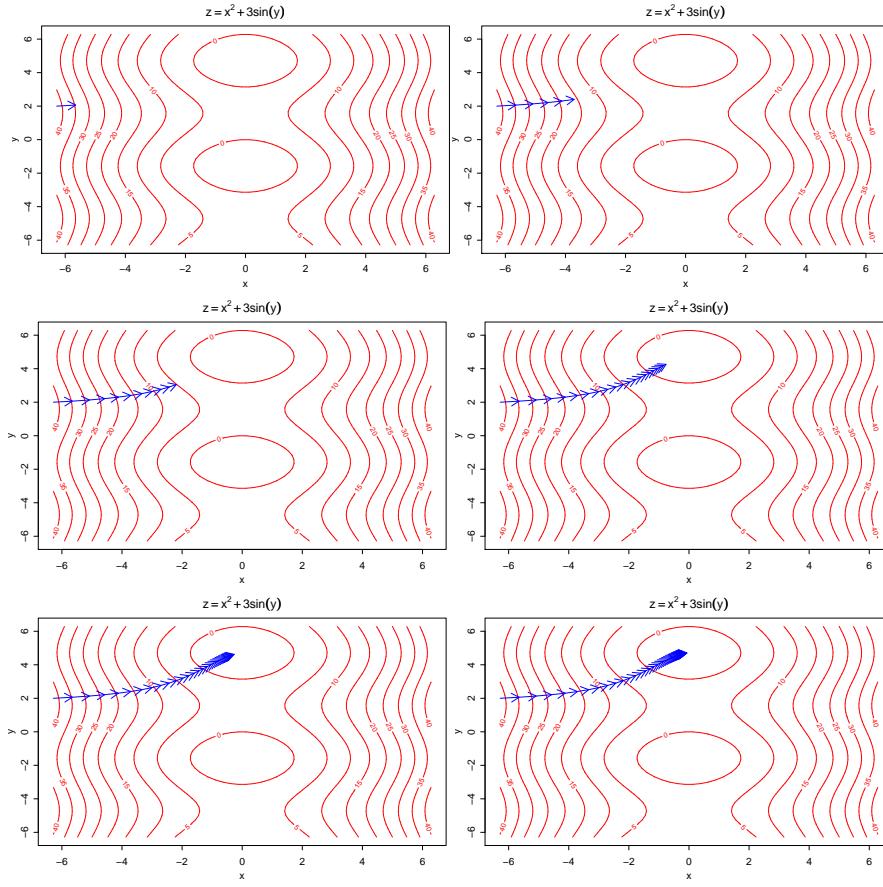


图 2 梯度下降算法的迭代路径

诉我们，优化方法中初始值的选择也是很重要的，并非任给一个初始值都能保证收敛到极值处。

本小节的动画可以在这个网页上观看：http://animation.yihui.name/compstat:gradient_descent_algorithm

§2.2 单纯形法

单纯形是在 N 维空间中具有 $N + 1$ 个顶点的形状。例如一维直线中的单纯形为线段，二维平面上的单纯形为三角形，三维空间中的单纯形为四面体。单纯形方法名称的来历与单纯形的形状本身密切相关，因为单纯形法的迭代方向恰好是沿着单纯形的某一条使得函数值上升（或下降）最快的边。

这里作者不再赘述单纯形法的数学理论，只是给出一个简单的示例，说明单纯形法的几何意义。

²参见 **stats** 包中的 `deriv()` 函数。

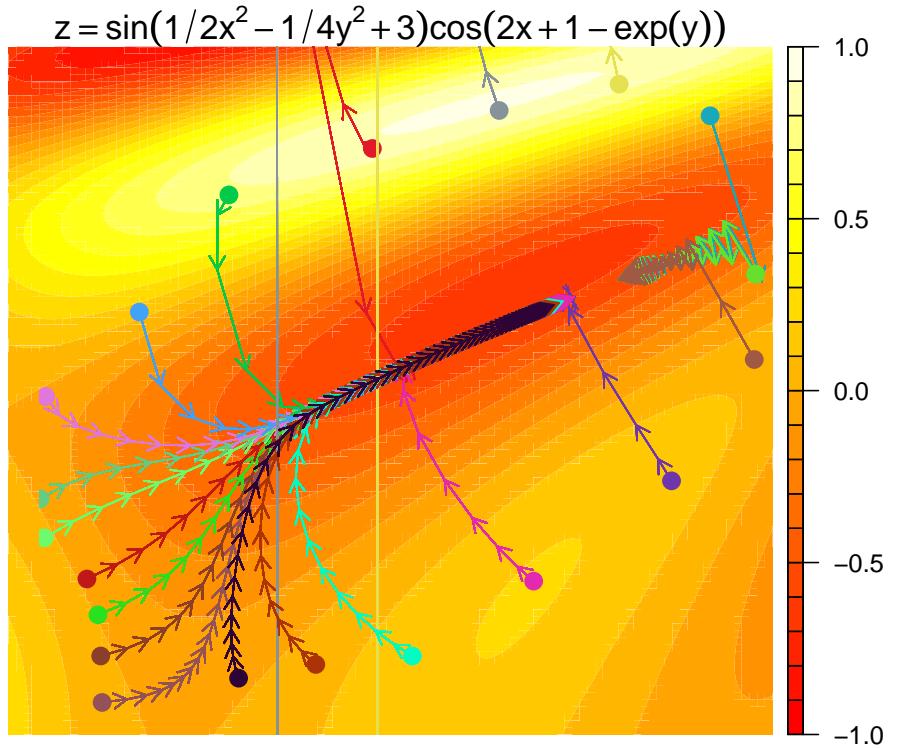


图 3 函数 $f(x, y) = \sin(1/2 * x^2 - 1/4 * y^2 + 3) * \cos(2 * x + 1 - e^y)$ 从不同起始点的梯度下降优化过程

例 3 考虑线性规划问题：

$$\begin{aligned} \max f &= 20x_1 + 15x_2 \\ \left\{ \begin{array}{l} 5x_1 + 2x_2 \leq 180 \\ 3x_1 + 3x_2 \leq 135 \\ x_1 \geq 0, x_2 \geq 0 \end{array} \right. \end{aligned} \quad (4)$$

图 4 给出了该规划的约束条件的几何表示以及单纯形法计算的中间步骤。左上图为线性规划的几何图示；右上图为第一个基础可行解 $(0, 0)$ ；左下图从 $(0, 0)$ 点迭代到了 $(36, 0)$ ，注意该步有两种可能的前进方式，即以 $(0, 0)$ 为顶点的三角形（单纯形）的两条边，之所以迭代到点 $(36, 0)$ 而不到 $(0, 45)$ 的原因从图中的等高线（虚线）可以清楚看出，因为后者使目标函数值增大的幅度不如前者；右下图表示第二步迭代，仍然是沿着三角形的边前进，同样的道理我们可以知道下一步不到 $(0, 0)$ 点的原因。该规划比较简单，两步之后就已经达到最大值。

当线性规划问题的约束条件较多时，平面上就会出现若干个三角形，单

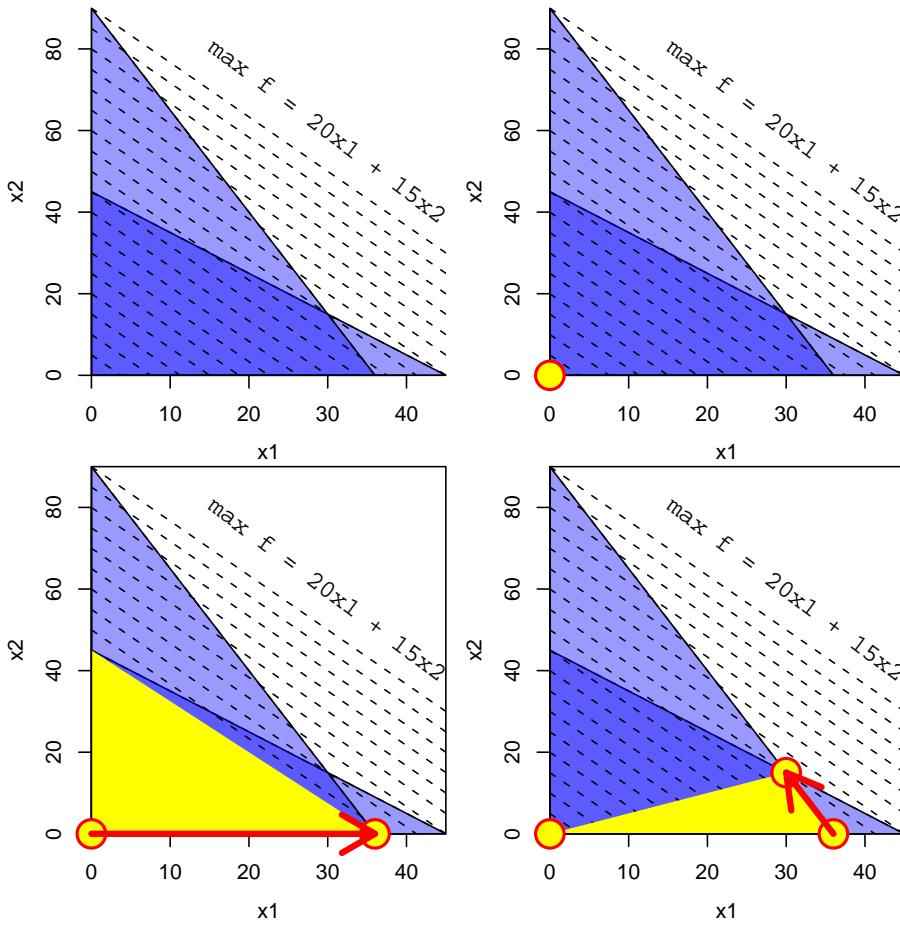


图 4 单纯形法迭代的路径

单纯形法的迭代会在众多三角形中挑选使目标函数增加（或减小）最多的那条边进行计算。这种情况下，动画尤其适合展示单纯形法的真正含义。

单纯形法的计算主要涉及矩阵的乘法和求逆，这样的运算在 R 语言中实现起来也非常容易。例如乘法可以用运算符 `%*%`，求逆可以用函数 `solve()`。

§3 优化方法的应用及含义讨论

本学期课程重点之一是数据包络分析 (DEA)，因此本文也对该方法从可视化以及统计学的角度分别作出讨论。

§3.1 数据包络分析

数据包络分析 (魏权龄, 胡显佑, 严颖, 2001) 一般用来评价多输入多输出

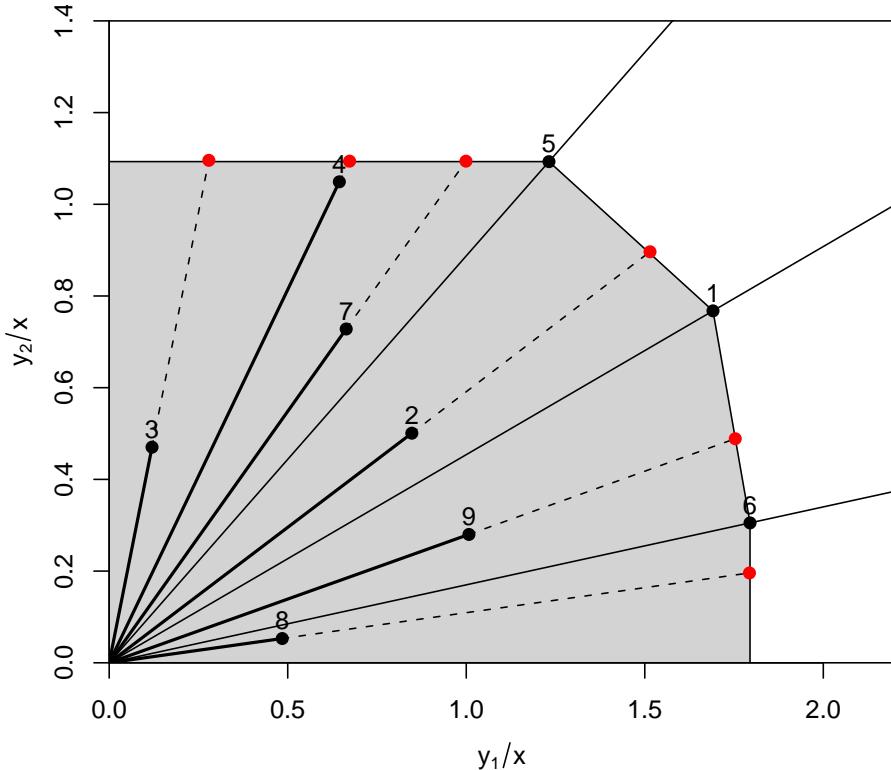


图 5 低维情况下的数据包络分析图示

出的决策单元的相对有效性，关于它的图形展示仅仅在低维条件下（即输入和输出变量都比较少）比较容易完成且有助于理解概念。我们不妨想象最简单的情形：若所有的决策单元只有一个输入变量和一个输出变量，那么相对有效性的评价就很容易了，因为此时“效率”就是简单的输出和输入的比率，比率最高（数值为 1）的决策单元为 DEA 有效，其它决策单元视比率的相对大小可以确定相对有效性。这种情况下的几何图形仅仅是一条直线或者数轴，有效的单元在 1 处，非有效的单元在 [0, 1] 之间。

对 DEA 方法最好的图形展示是在输出为二维变量而输入为一维变量的情况下，此时我们可以计算两个输出变量分别对输入变量的比率，然后根据两个比率得到一幅散点图，图中所有数据点的“外壳”（凸包）与坐标轴围成的多边形便形成了所谓的生产前沿面，处于生产前沿面上的决策单元都为 DEA 有效，而凸包内的决策单元不是 DEA 有效。进一步，我们不难通过数学推导得出，此时相对有效性的大小正是每个决策单元的点和原点连线长度与该连线从原点延伸到生产前沿面的长度之比。通过这样的图示，DEA 的若干概念如“包络”、“效率”、“生产前沿面”等都会变得非常直观。

如图 5 所示，决策单元 1、5、6 都处在生产前沿面上，相对有效率为 1；

而其它决策单元的相对有效性则为粗实线长度占其延伸到前沿面的线段长度的比例。从图中我们即可直观判断出各决策单元相对有效性的高低。

DEA 从一个角度解决了投入产出的评价问题，它将多变量问题综合为一个判断指标，为决策者节省了时间和精力，然而，一般来说压缩信息必然伴随着损失信息，综合评价的缺点就是不能照顾细节问题。例如，即使我们知道某决策单元非 DEA 有效，那么我们具体应该怎样改进该决策单元的效率呢³？如果仅仅停留在评价的层面上而没有进一步的指导，那么对于实际工作来说并没有太大的意义。另外，从图 5 引发的思考可以有：提高决策单元生产效率的路径必须是从原点出发的射线吗？可否保持一种产出不变、而通过改变另一种产出到达前沿面？提出这个问题的背景是：很多情况下，企业都有自身的生产优势和劣势，同时提高多种产出可能并不现实。

§3.2 关于数据的讨论

关于 DEA 的数据本文也略作两点讨论：

一、DEA 要求投入产出数据都为正数，这样就限制了负指标或零指标的使用，而实际工作中，并不能保证产出非零或非负，也许某工厂在某段时间并没有生产，也许某企业生产利润为负值。在零值或负值的情况下，以比率作为目标函数可能不合适，我们为何不考虑将其换为产出和投入减法呢？这样就可以针对所有范围内的数据做线性规划。

二、DEA 和大多数数学方法一样，并不考虑样本数据的随机性，把收集到的数据视为确定性的数据。表面上 DEA 不假定数据的分布，但作者认为这仅仅是数学设定的问题，并不代表样本没有随机性。如果真的将样本的统计分布考虑进来，那么评价的结果必然也有变化，例如相对效率的估计值可能不再是一个确定的点值，而是成为一个置信区间。

§4 结论

很多统计系的学生念了几年统计，但当他们被问到“你是否见过似然函数的样子”之类的问题时，回答通常是“没有”。类似地，作者猜测很多学完优化方法或运筹学的学生并不知道什么叫做“单纯形”或者“包络”。另一方面，通过作者的接触发现，也有相当一部分学生的数学功底很好，所有的数学公式都能推导，但就是不知道公式是什么意思。本文从可视化的角度对优化方法的可视化作出了一些初步探讨，力求将抽象的概念具体化、并以一种“看得见摸得着”的方式介绍出来，期望对优化方法等数学意味较重的课程能有所帮助。

³ 我们不能笼统地说减少投入、增大产出。

参考文献

魏权龄, 胡显佑, 严颖(2001). 运筹学通论 (修订本). 北京: 中国人民大学出版社.

Ihaka R, Gentleman R (1996). “R: A Language for Data Analysis and Graphics.” *Journal of Computational and Graphical Statistics*, 5(3), 299–314. ISSN 10618600.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Xie Y (2008). *animation: Demonstrate Animations in Statistics*. R package version 1.0-1, URL <http://animation.yihui.name>.