SIMULATION,OPTIMIZATION

and Application for Management system

管理系统仿真、优化及应用

陈智民◎著

中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

管理系统仿真、优化及应用

陈智民 著

中国经济出版社

图书在版编目(CIP)数据

管理系统仿真、优化及应用/陈智民 著. —北京:中国经济出版社,2005.6
ISBN 7-5017-6950-8

I. 管… II. 陈.. III. 管理信息系统 IV. C931.6

中国版本图书馆 CIP 数据核字(2005)第 026691 号

出版发行: 中国经济出版社 (100037·北京百万庄北街 3 号)
网　　址: www.economyph.com
责任编辑: 邓媛媛 (89809929; editordeng@163.com)
责任印刷: 石星岳
封面设计: 中子画工作室
经　　销: 各地新华书店
承　　印: 北京金瀑印刷有限责任公司
开　　本: 787×960 毫米 1/16　　**印　张:** 13.5　　**字　数:** 208 千字
版　　次: 2005 年 6 月第 1 版　　**印　次:** 2005 年 6 月第 1 次印刷
书　　号: ISBN 7-5017-6950-8/F·5565　　**定　价:** 26.00 元

服务热线: 68344225　68353507　68341876　68341879　68353624

中国经济书店: 66162744　　地　址: 西四北大街 233 号

前　言

系统仿真作为一种求解复杂系统的分析技术，已经得到了广泛的应用。现时情况下，人们在对许多复杂问题做决策时，最优化作为一个基本原则已牢固地树立起来，它能够使对问题的分析论证得到无可争议的完善化。根据最优化原理，人们把实际问题抽象为一定类型的分析模型，如线性规划模型、网格模型、存贮模型等，以便使用相应的解析算法对系统的目标进行优化。但是在处理一个复杂系统的决策问题时，要把所有参数之间的量及其逻辑关系约束条件以及所求目标之间的复杂情况描述出来，那几乎是不可能的。因此现存的最优化模型只能是一种对实际问题的近似描述。当人们面对一个复杂的决策问题时，一方面希望建立一个复杂的分析模型，以准确地把实际问题描述出来；另一方面又希望建立一个容易处理的分析模型，以便于问题的求解，实际上这是两个相互矛盾的目标。对于一个复杂的决策问题，人们如果找不到一种合适的分析模型来描述它，就不得不对问题本身进行简化，以适应相应的分析模型的各种假设条件。这种简化往往使得问题变得面目全非. 这种情况下，系统仿真技术不失为一种可行的分析方法。理论上讲，仿真模型可以描述任何复杂程度的实际问题。但仿真技术只是一种分析问题的手段，它没有建立起最优化的概念，也就是说，仿真模型的一次运行，只能给出一组决策变量下的系统输出(目标函数值)，它不能给决策者提供系统输出最优的决策变量的组合。目前比较著名并且应用较为广泛的几种仿真语言如 SLAM，ARENA，SIMAN 等也都没有提供优化手段，这不能不说是仿真技术的遗憾，也是影响其应用的原因之一。自六十年代中期提出仿真最优化概念以来，许多人从不同的方面做了不少的工作，但由于问题本身的

难度而发展缓慢。随着计算机硬件的飞速发展，仿真的成本越来越低，系统仿真应用越来越广泛，仿真环境下的最优化成为迫切需要解决的问题。

本书共分六章。第一章首先阐述了管理系统仿真技术研究的范围和步骤以及有关概念，然后定义了仿真环境下最优化问题的具体形式，最后对当前仿真优化问题的发展状况作了全面的综述，并且给出了两个仿真优化例子。

第二章主要讨论仿真优化环境下响应曲面法的有效实施，以改进现有响应曲面法采用重复运行估计曲面参数时效率不高的状况。同时把响应曲面法推广到稳态仿真模式下的仿真优化问题。依据试验设计的D-优良性和G-优良性讨论响应曲面参数单运行估计的可行性。采用通用随机数技术使得不同试验点的仿真目标响应产生正相关、以获得更准确的单运行响应曲面参数估计，在稳态仿真模式下，应用了终态仿真模式下曲面多数单运行估计的思想，减小了每次仿真运行的长度，达到了提高优化效率的目的，解决了稳态仿真模式下响应曲面法应用的具体技术问题。给出了一种在一阶响应曲面上寻优时确定最佳搜索步长的方法。

第三章讨论随机逼近算法在仿真环境下的最优化中的应用问题，给出了无约束和有约束仿真优化问题的随机逼近型最优化算法，并且对算法中的有关参数的选择做具体的讨论，给出我们的建议。差分梯度估计作为一种随机梯度估计方法，在本章中也将进行讨论，着重分析单边差分和对称差分梯度估计在不同的随机数安排策略下所得到的梯度估计的统计特性，并且以M/M/1排队系统和存储模型为例，做实际的试验比较。

第四章主要讨论梯度估计方法中的无穷小扰动分析(IPA)方法。首先阐述离散事件动态系统(DEDS)的无穷小扰动分析的基本思想和利用IPA在稳态仿真模式下的梯度估计问题，在此基础上再将无穷小扰动分析推广到终态仿真问题的梯度估计，并且证明IPA梯度估计的无偏性和强一致性。同时给出基于仿真响应的IPA梯度估计的随机逼近型仿真优化算法。

第五章基于第二、第三、第四章仿真环境下的最优化理论，讨论其程序实现问题。以GASP IV仿真语言为基础仿真语言，实现了面向优化的通用仿真系统GASPOPT。

第六章对单运行曲面参数估计的响应曲面法，对称差分梯度估计下的随机逼近过程和IPA梯度估计下的随机逼近过程三种仿真优化方法进行试验比较。给出仿真优化方法有效性的评价准则。设计了五个试验函数和五个试验仿真模型。对上述三种仿真优化方法做了定性和定量的分析比较，最后给出我们的评价结论和建议。

第七章应用系统仿真技术，对一厂多工地混凝土供应系统进行研究，建立了该供应系统的SLAM仿真模型，对仿真模型的输入变量的概率分布进行拟合，仿真输出结果给出了该混凝土供应系统的行为度量，对合理调度混凝土搅拌车给出来了建议。

第八章是全文总结。

在仿真技术应用愈来愈广泛的今天，寻求有效的仿真优化方法十分迫切并具有很大的实用价值。我们对仿真环境下的最优化问题做了较系统的探讨，有许多东西还需进一步完善，仍然有许多工作有待于进行。由于笔者水平有限，书中错误在所难免，望读者不吝批评指正。

陈智民

2005年4月于深圳大学

目　录

第一章　仿真环境下最优化的基本问题

§1.1　系统仿真

系统仿真是指建立一个系统的数学逻辑模型，并且对该模型在计算机上进行试验处理，通过对系统动态特性的观测，以研究系统行为的过程。

这个定义包含三方面的含义：(1)系统仿真的目的是观测系统的动态特性，以研究系统的行为。(2)模型的建立过程，也即系统的数学逻辑模型的表示，计算机程序的准备等。(3)系统试验，也即进行试验设计，系统的仿真运行及输出分析等。一般情况下，仿真研究的步骤和内容如下，在 *Banks* (1984)中，也有类似的论述。

1. 问题说明：对所研究的系统的一般描述，系统边界的基本定义等。
2. 研究目的：明确仿真研究应解决的问题。
3. 模型建造：确定系统中的实体，每个实体的属性以及从事的活动，系统的可能状态等，明确系统的基本特性，选择和修正系统的基本假设，最终建立一个能刻画被研究系统行为，满足研究目标的模型。
4. 数据收集：收集对系统的每一个实体进行描述时所需要的信息，确定系统参数的概率分布等。
5. 程序编码：转换系统模型成为可执行的计算机程序。
6. 程序验证：确定一个仿真模型能否按预期的要求执行，也即验证程序与模型的一致性。
7. 模型确认：比较模型所表现的系统行为和实际观测到的系统行为之间

的差异，缩小这些差异以提高模型表现被研究系统的能力。程序验证和模型确认的不断重复，最终建立一个能刻画实际系统、并且能正确执行的仿真模型。

8. 试验设计：确定各种可选方案，选择主要的输入变量及其试验水平，给定仿真运行的长度和重复次数。

9. 运行与分析：评价特定输入变量对系统性能度量的影响，仿真数据的输出分析以及决策参数的优化。

10. 仿真报告：编写仿真输出综合报告。

以上十个步骤概括了仿真研究的基本步骤和内容。由于终态仿真问题和稳态仿真问题的性能测度有着不同的统计特性，输出分析时需要加以区分，因此给出终态仿真和稳态仿真的定义。

终态仿真是一种系统性能测度与仿真的时间区［0，T_F］有关的仿真模式。这里 T_F 是指在仿真中特定的仿真终止事件发生的时间。

稳态仿真是在仿真终止时间趋于无限时，定义系统性能测度极限的一种仿真模式。终态仿真与稳态仿真之间的主要差异在于，终态仿真主要研究在规定的时间区间内的系统行为，仿真终止事件在仿真开始前已被确定，系统初始状态的选择对系统的性能测度有着显著的影响；而稳态仿真主要研究系统的稳态行为、仿真运行的长度要足够的长，以便得到好的系统性能测度的估计，系统的初始状态对系统性能测度的影响是无关紧要的，但合适的初始状态的选择可以缩短仿真运行的长度。由于这些差异，导致了在仿真输出分析时，终态仿真与稳态仿真在统计方法上的不同。终态仿真和稳态仿真的区分也不是绝对的。对于某些系统，究竟采用何种模式进行分析，取决于分析者所要达到的目的。

§1.2 仿真优化问题的一般描述

最优化的概念作为决策时的一种依据已普遍被人们接受。对不同的问题，存在着不同的最优化结构，如线性规划、动态规划、网络技术等，都有

它们自身的结构特点，相应地也存在着适应这种结构特点的优化算法。把仿真问题作为最优化的对象，ω 其最优化结构自然有其自身的特点。

系统仿真技术主要用来解决哪些难以用现有的分析模型来求解的复杂系统的决策问题，如果能用分析模型以解析方法求得问题的最优解，那么绝对不用仿真技术。仿真问题的目标函数是一非线性随机函数的数学期望，与分析模型不同的是这种函数关系是通过仿真模型来描述的，在系统的输入变量和输出变量之间没有明确的数学表达式。因此只有通过运行仿真模型来求目标函数值，这导致了不能象分析模型那样，用解析方法求最优解。特别是仿真技术处理的大多都是随机性问题，使得优化更加复杂和困难。

仿真模型表达了系统内部的量及逻辑关系，通过仿真优化处理，可以找出一组输入变量，使得系统的输出达到最优。

令：$X=(x_1,\ x_2,\ \cdots,\ x_K)$ 是仿真模型的可控因素，也可以称作仿真模型的决策参数、决策变量或输入变量，ω 是仿真模型中的不可控因素。定义：

$$f(X,\omega)=(f_1(X,\omega),f_2(X,\omega),\cdots,f_P(X,\omega)) \qquad (1.2.1)$$

是仿真模型的输出，即仿真响应，该仿真模型共有 P 个仿真响应。

通常情况下，我们把仿真模型看作是实值函数 $f(X,\ \omega)$。函数 $f(X,\ \omega)$ 的具体形式是不知道的，一次仿真运行即是对 $f(X,\ \omega)$ 的一次计算，或仿真响应的一次估计。若仿真模型中没有不可控因素 ω，那么就是确定性仿真问题，否则是随机性仿真问题。确定性仿真问题是随机性仿真问题的一个特例，因此以后我们只讨论随机性仿真问题。

定义 $Y(X)=E\ [\gamma(X,f(X,\ \omega))\mid X]$ 是仿真问题的目标函数。目标函数不仅依赖于决策参数 X，也依赖于仿真响应 $f(X,\ \omega)$。$\gamma(X,f(X,\ \omega)$ 是非线性随机函数，一次仿真运行得到一个 $\gamma(X,f(X,\ \omega))$ 的估计值，我们也称 $\gamma(X,f(X,\ \omega))$ 为仿真模型的目标响应。一般我们假设 $\gamma(X,f(X,\ \omega))$ 可以表示为 X 和 $f(X,\ \omega)$ 的显式函数，但由于仿真响应 $f(X,\ \omega)$ 没有明确的数学表达式，并且其概率密度函数一般也是不知道的，因此 $Y(X)$、$\gamma(X,f(X,\ \omega))$ 并不能表示成 X 的显式函数。

仿真环境下的最优化，即确定决策参数 X，使得 $Y(X)$ 达到最优。具体地，我们给出以下的定义：仿真问题的最优化即求解

$$Min\ Y(X) = E[y(X, f(X,\omega)) \mid X]$$
$$s.t. \qquad G_j(X) \leqslant 0, \qquad j = 1,2,\cdots,M \qquad (1.2.2)$$

这里 $G_j(X) \leqslant 0$ 是 M 个约束条件。求目标函数的最大值即求 $-Y(X)$ 的最小值，因此以后仅讨论求最小值问题。

(1.2.2)式定义的仿真优化问题是单目标的，根据目标函数的多少，仿真优化问题可以被分为单目标和多目标问题。多目标问题即求解：

$$Min\{Y_1(X), Y_2(X), \cdots, Y_q(X)\}$$
$$s.t. \qquad G_j(X) \leqslant 0, \qquad j = 1,2,\cdots,M \qquad (1.2.3)$$

基于约束条件的存在性，仿真优化问题可以被分为约束问题和无约束问题；基于约束条件的特性，仿真优化问题可以被分为线性约束、非线性约束和随机约束问题。根据可控因素 $X = (x_1, x_1, \cdots, x_K)$ 的数量，可以把仿真优化问题分为三种类型：具有大约 10 个或不到 10 个可控因素的小型问题和具有大约 10 个到 30 个可控因素的中型问题，具有从大约 30 个以上可控因素的大型问题。根据可控因素取值性质，仿真优化问题可以分为连续变量和离散变量问题。

下面我们给出两个仿真优化问题的例子，在以后各章中，这两个例子将作为我们试验研究的对象，同时进一步说明终态仿真和稳态仿真的概念。

例 1 单队单服务员排队系统

排队系统一般由三个部分组成，也即输入过程、排队规则和服务机构。对于单队单服务员排队系统，假设顾客的到达时间间隔是独立同分布的随机变量。当顾客到达发现服务员闲时就立即进入服务台服务，否则就排在队的末尾。服务员服务时间也是独立同分布的随机变量。当服务员结束一名顾客的服务后，若队列不空，则按照先到先服务的原则从队列中选出第一个顾客进行服务。

对于单队单服务员排队系统，仿真的目的是要估计顾客在系统中的平均滞留时间以及对系统运行的经济性进行评价。假设第 i 个顾客在系统中的滞

留时间(等待时间加上服务时间)是 D_i，可以按照终态仿真或者稳态仿真两种模式来研究。在终态仿真下，可以是估计前 N 个顾客的平均滞留时间的期望值，那么仿真终止事件 $E=\{$第 N 个顾客结束服务$\}$。给定初始条件 $L(0)=d$，也即在仿真时钟 $t=0$ 时，系统中的顾客数 $L(0)$ 为 d，$d=0$，1，…，服务时间的均值是 X，那么终态仿真下系统的性能测度(仿真响应的期望值)定义为：

$$F(X)=E[f(X,\omega)\,|\,L(0)=d]=E\left[\frac{1}{N}\sum_{i=1}^{N}D_i\,|\,L(0)=d\right] \tag{1.2.4}$$

在稳态仿真下，系统的性能测度定义为：

$$F(X)=E[f(X,\omega)]=\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N}D_i\,|\,L(0)=d\right]\qquad d=0,1,\cdots \tag{1.2.5}$$

也即求稳态的平均滞留时间。

一般情况下，若 $d_1\neq d_2$，则 $L(0)=d_1\neq L(0)=d_2$，因此在终态仿真下系统的性能测度与初始状态有关，而稳态仿真下 d 可以取任何有意义的值。

当顾客到达间隔时间和服务时间均服从负指数分布时，单队单服务员系统即 $M/M/1$。为了对 $M/M/1$ 排队系统的经济性进行评估，以系统运行的总成本作为仿真的目标函数［李维铮等(1982) *Hillier*(2004)］即：

$$\begin{aligned} &Min\ Y(X)=C_cF(X)+C_s/X\\ &s.t.\qquad 0\leqslant X\leqslant R \end{aligned} \tag{1.2.6}$$

这里 C_c 是顾客的单位时间成本；C_s 是服务员成本；$F(X)$是顾客在系统中的平均滞留时间；R 是顾客到达间隔时间的均值。对于 $M/M/1$ 队，在稳态意义下，$F(X)=E[f(X,\omega)]=RX/(R-X)$是已知的，那么(1.2.6)的最优解是：

$$X^*=\frac{R\sqrt{C_S}}{R\sqrt{C_C}+\sqrt{C_S}} \tag{1.2.7}$$

当 $F(X)$未知时，可以通过仿真来求 $f(X,\omega)$，以估计出 $F(X)$，但 Y

(X)并不能表达成 X 的显式函数，因此(1.2.6)是一个典型的仿真最优化问题。

例2　(s, S) 存储系统［*Pritsker*(1973), *Fu*, *Healy*(1992)(1997)］

一个零售商店销售一种产品，顾客的需求是均值为0.2周的泊松分布。假设订货费用是每次50元；每个产品采购价格和销售价格分别是40元和65元；订货周期为3周；存储费用率为0.003836元/每元每周。当需求发生变化而不能满足时，产生缺货损失，每次20元。商店每2周进行一次盘点，查看是否需要订货，每次盘点的费用是30元。基于以上的情况，商店经营者要确定一种(s, S)存储控制策略，即：当每次盘点时，若发现库存水平 *POS*(初始值为36)小于订货点 s，则订货，订货量为了 $S-s$，否则不订货。

对(s, S)存储系统研究的目标主要是分析费用的发生情况，确定最优的(s, S)存储策略使系统盈利最大。库存控制问题中的费用一般包括定货费用，存储费用，缺货费用和盘点费用。可以用仿真方法来完成对(s, S)存储系统的研究。令 $X=(x_1, x_2)=(s, S)$，通过仿真运行，可以确定在 T_F周内，商店平均每周获得的利润 $y(X, \omega)$，总销售量 $f_1(X, \omega)$，平均存储量 $f_2(X, \omega)$，定货次数 $f_3(X, \omega)$，缺货次数 $f_4(X, \omega)$和盘点次数 $f_5(X, \omega)$。那么 T_F 周内的平均利润为：

$$\begin{aligned} y(X, \omega) = [&25f_1(X, \omega) - 0.003836f_2(X, \omega) - 50f_3(X, \omega) \\ &- 20f_4(X, \omega) - 30f_5(X, \omega)] / T_F \end{aligned} \tag{1.2.8}$$

此仿真问题是终态仿真问题，它研究的是在 $[0, T_F]$ 周内(s, S)存储系统行为。仿真优化问题为：

$$\underset{X}{Max}Y(X) = E[y(X, \omega)] \tag{1.2.9}$$

即求最优的(s, S)库存控制策略，使每周的平均利润的期望值为最大。

§1.3　仿真优化技术综述

本节给出仿真与优化的结合结构，然后综述仿真优化技术目前的发展状况。

§1.3.1　仿真优化的结合结构

式(1.2.2)的仿真优化问题初看起来无异于一般的非线性规划问题，它们之间的区别在于目标函数的表达式不同。在非线性规划问题中，目标函数是显式，并且是确定的。而仿真优化问题中，目标函数是隐式，且存在着不可控因素的干扰，给定一组输入变量 X，目标函数值只有通过运行仿真模型来求得。仿真优化问题的这些特性，使得依赖于导数的最优化技术的应用变得非常困难。因此应用不使用导数信息的直接搜索技术是一个可行的选择。

直接搜索技术一般过程是：给定初始点 X_0，决定合适的搜索方向 d，沿这个方向 d 确定一可能的改进解，再重新确定新的搜索方向，直至满足规定的最优性条件。在这种情况下，仿真优化的结合结构如图 1.3.1 所示。

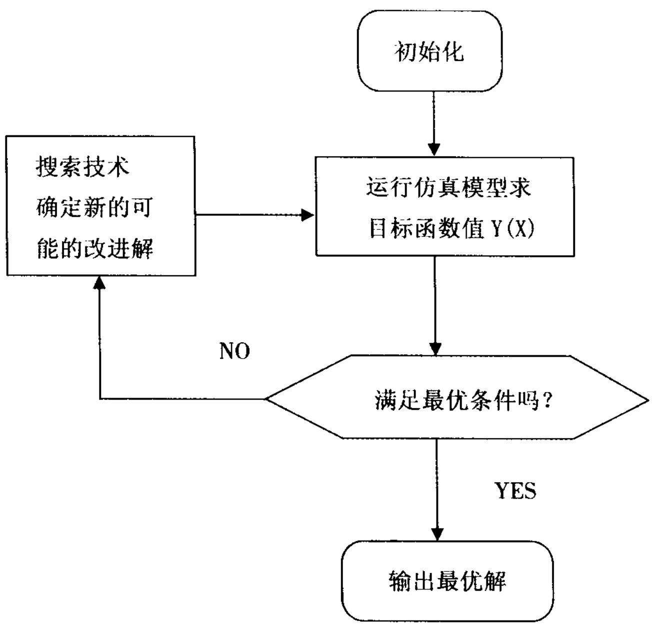

图 1.3.1　仿真优化的结合结构 I

由于 $y(X, f(X, \omega))$ 的随机性，因此对仿真目标响应的处理也是一个必

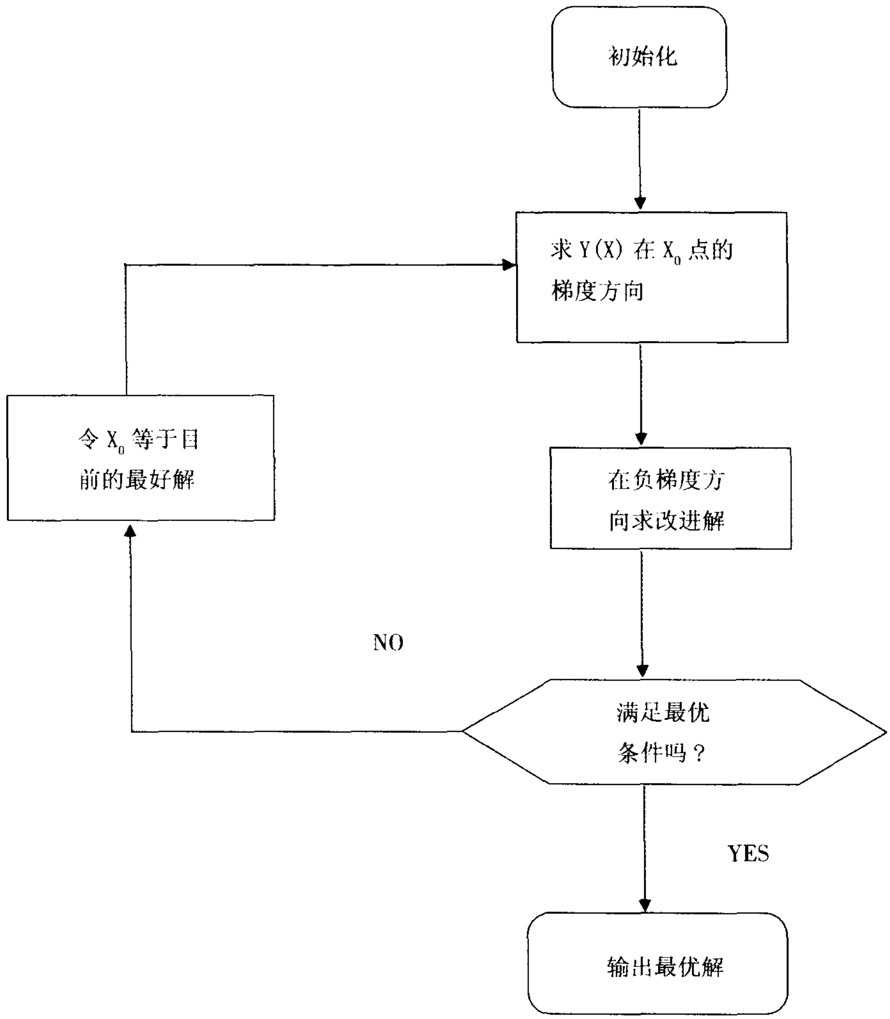

图 1.3.2　仿真优化的结合结构 Ⅱ

须慎重考虑的问题，一般考虑是，目标函数的优劣不仅应根据 $Y(X)$ 估计值的大小来判断，而且应该考虑估计值方差的影响。

尽管在仿真环境下采用使用梯度的最优化方法是困难的，困难的所在是梯度的不易求得，但许多人在如何求得目标函数的梯度上做了不少的工作。

在结构 Ⅰ 中的直接搜索技术，是在不知道目标函数改进方向的情况下，

靠一定的程式去逐步搜索最优解，效率较低。利用使用梯度的优化搜索技术，将有利于算法效率的提高。图 1.3.2 是仿真与使用梯度搜索技术的结合结构。

在结构Ⅱ下，目标函数被看作是 K+1 维欧氏空间的曲面，这里 K 是仿真模型可控因素的数量。一般过程是：给定初始点 X_0，确定 X_0 点目标函数的改进方向，即梯度方向，在此梯度方向上求得最好点，再以此最好点作为初始点，重复，直至满足规定的最优性条件为止。

以上描述的两种结构，特别是第二种结构，我们在以后将要做详细的讨论。

§1.3.2 直接搜索技术

不使用梯度的直接搜索技术在非线性规划中已讨论得很多，如 Hooke - Jeeves 模式搜索技术，Rosenbrock 旋转坐标法等。这些方法在仿真环境下是容易实现的，并且在仿真优化技术发展的初始阶段，直接搜索技术是主要的仿真优化技术。

直接搜索技术把仿真的目标函数作为 K+1 维欧氏空间中的一个曲面。在搜索过程中企图沿着一个可能的改进方向寻优而不需要梯度信息。直接搜索技术一般先做局部探测搜索以确定一个可能使目标函数能够减小的方向，沿着这个方向再进行模式搜索，即以一定的步长确定可能的改进解。不同的直接搜索方法之间差别只是局部探测的方式和计算搜索步长的方法不同。这里我们并不给出具体的迭代过程，请参看李维铮等(1982)，希梅尔布劳(1980)。

Smith(1973)；Farrel(1977)；Pegden，Gately(1977)；Schweful(1979)；Pegden，Gatel y(1980)；魏建平(1985)；Bengu，Haddock(1986)；Haddock，Bengu(1987)等都对直接搜索技术在仿真优化中的应用做了大量的研究。直接搜索技术简单、实用，程序实现非常方便，这也可能是其被大量应用的原因，但直接搜索技术与一些使用梯度信息的优化技术相比，其效率较低。Smith(1973)基于 64 种试验函数对有些直接搜索方法和重复运行估计曲面参