5.4 单纯形法的矩阵表示

线性规划问题：

$$\max\ z = \mathbf{c}^T\mathbf{x}$$

$$s.t. \begin{cases} \boldsymbol{A}\mathbf{x} \le \mathbf{b} \\ \mathbf{x} \ge \mathbf{0} \end{cases}$$

标准化：

$$\max\ z = \mathbf{c}^T\mathbf{x}$$

$$s.t. \begin{cases} \boldsymbol{A}\mathbf{x} + \boldsymbol{I}\mathbf{x}_s = \mathbf{b} \\ \mathbf{x}, \mathbf{x}_s \ge \mathbf{0} \end{cases}$$

松弛变量

5.4 单纯形法的矩阵表示

初始单纯形表：

项目			非基变量		基变量
			$\mathbf{x}_B$	$\mathbf{x}_N$	$\mathbf{x}_s$
0	$\mathbf{x}_s$	b	B	N	I
$c_j - z_j$			$\mathbf{c}_B^T$	$\mathbf{c}_N^T$	0

当迭代后基变量为 $\mathbf{x}_B$ 时：

项目			基变量	非基变量	
			$\mathbf{x}_B$	$\mathbf{x}_N$	$\mathbf{x}_s$
$\mathbf{c}_B$	$\mathbf{x}_B$	$B^{-1}b$	I	$B^{-1}N$	B^{-1}
$c_j - z_j$			0	$\mathbf{c}_N^T - \mathbf{c}_B^T B^{-1} N$	$-\mathbf{c}_B^T B^{-1}$

$\mathbf{c}_B^T - \mathbf{c}_B^T B^{-1} B = 0$

当 B 为最优基时：≤ 0　　≤ 0

5.4 单纯形法的矩阵表示

$$\max\ z = 5x_1 + 6x_2$$

$$s.t. \begin{cases} 3x_1 + x_2 \le 48 \\ 3x_1 + 4x_2 \le 120 \\ x_1, x_2 \ge 0 \end{cases}$$

$$\Longrightarrow \begin{cases} 3x_1 + x_2 + x_3 = 48 \\ 3x_1 + 4x_2 + x_4 = 120 \\ x_1, x_2, x_3, x_4 \ge 0 \end{cases}$$

5.4 单纯形法的矩阵表示

			5	6	0	0
	x_B	b	x_1	x_2	x_3	x_4
0	x_3	48	3	1	1	0
0	x_4	120	3	(4)	0	1
		0	5	6	0	0
0	x_3	18	(9/4)	0	1	-1/4
6	x_2	30	3/4	1	0	1/4
		180	1/2	0	0	-3/2
5	x_1	8	1	0	4/9	-1/9
6	x_2	24	0	1	-1/3	1/3
		184	0	0	-2/9	-13/9

B I I B^{-1}

5.5 单纯形法的最坏情形

基本解的个数：$\binom{n}{m}$

- 用单纯形法求解线性规划问题有没有可能检验到所有的基，直到最后一个才找到最优解？
- 从理论上讲是有可能的，但直到上世纪70年代，还没有人对任意个变量的问题找到这样的例子。

m	$\binom{2m}{m}$	m	$\binom{2m}{m}$
1	2	100	9×10^{58}
10	184756	200	1×10^{119}
20	1×10^{11}	300	1×10^{179}
50	1×10^{29}	400	2×10^{239}

3, 199, 243, 548, 502 years!

每秒运行10亿次单纯形运算

5.5 单纯形法的最坏情形

- **Bad counter example:** **Klee and Minty (1972)**

$$\max \quad z = \sum_{j=1}^{m} 10^{m-j} x_j$$

$$s.t. \quad 2\sum_{j=1}^{i-1} 10^{i-j} x_j + x_i \leq 100^{i-1}, \forall i = 1, \dots, m$$

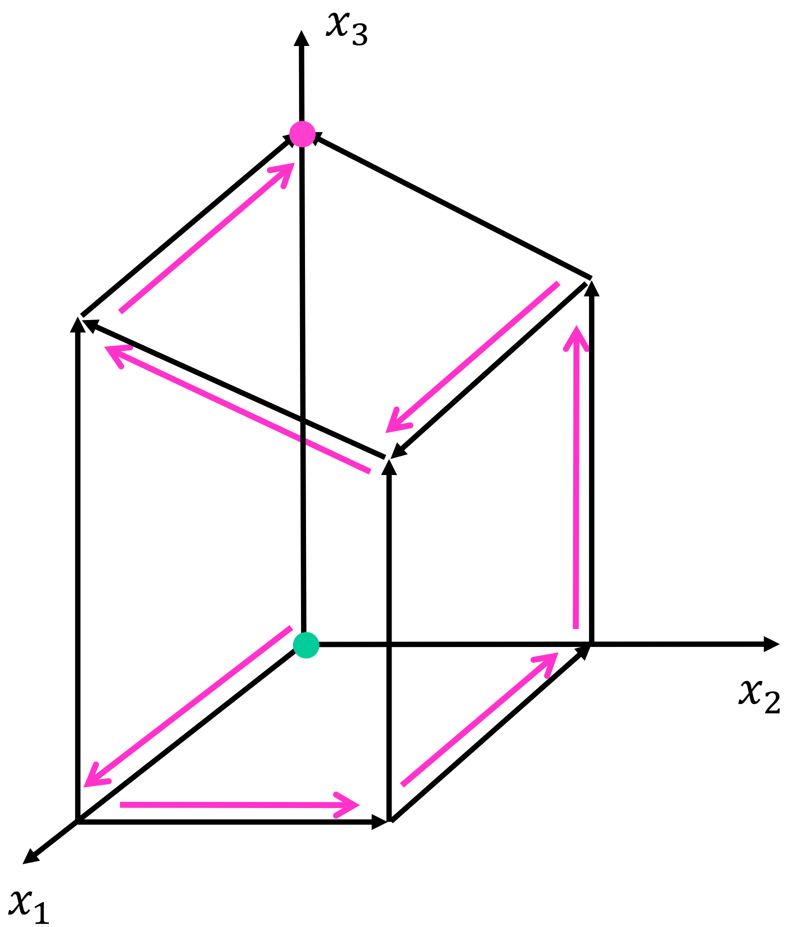
- 通过添加松弛变量，该问题可以变成一个有$2m$个决策变量和m个约束条件的标准问题。有2^m个可行基。
- 如果使用单纯形法求解，在每一步都选择最大的检验数作为进基变量，则其每一个可行基都会被检验到。

m = 3

$$\max\ z = 100x_1 + 10x_2 + x_3$$

$$s.t.\begin{cases} x_1 \le 1 \\ 20x_1 + x_2 \le 100 \\ 200x_1 + 20x_2 + x_3 \le 10000 \\ x_1, x_2, x_3 \ge 0 \end{cases}$$

Basis			z
$s_1 = 1$	$s_2 = 100$	$s_3 = 10000$	0
$x_1 = 1$	$s_2 = 80$	$s_3 = 9800$	100
$x_1 = 1$	$x_2 = 80$	$s_3 = 8200$	900
$s_1 = 1$	$x_2 = 100$	$s_3 = 8000$	1000
$s_1 = 1$	$x_2 = 100$	$x_3 = 8000$	9000
$x_1 = 1$	$x_2 = 80$	$x_3 = 8200$	9100
$x_1 = 1$	$s_2 = 80$	$x_3 = 9800$	9900
$s_1 = 1$	$s_2 = 100$	$x_3 = 10000$	10000



5.5 单纯形法的最坏情形

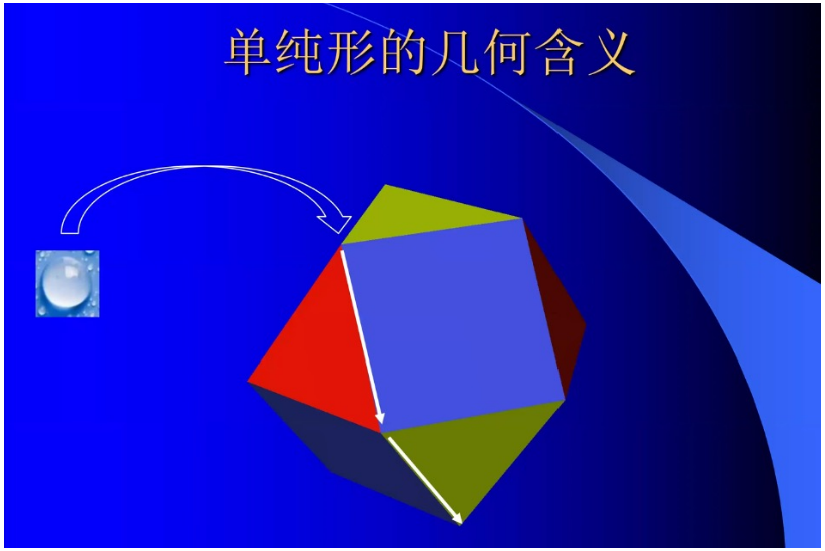
- 上述例子说明单纯形算法的时间复杂性有可能是指数型的。**从计算复杂性的标准来看，单纯形算法不是好算法**。
- 线性规划是否存在多项式算法便成了计算机科学家和数学家十分感兴趣的问题。
- 1979年，苏联数学家哈奇扬(L. G. Khachian)给出了一个求解线性规划的多项式算法：椭球算法(ellipsoid method)，其计算复杂性为$O(n^6L^2)$。
- 1984年，印度数学家卡玛卡(N. Karmarkar)给出了求解线性规划的一个新的多项式算法：内点算法(interior point method)，其计算复杂性为$O(n^{3.5}L^2)$。
- 1989年，Renato D.C. Monteiro和Ilan Adler给出了求解线性规划的一个“原—对偶”内点算法，其计算复杂性为$O(n^{3.5}L)$。

L = bitlength of data, n = dim(x)

单纯形法 VS 内点法

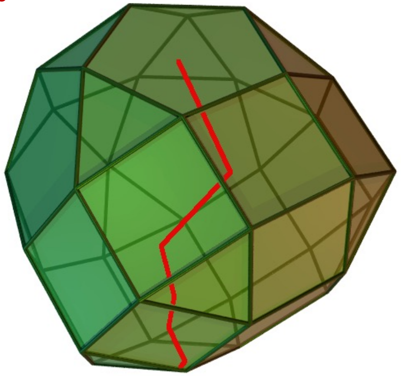
- **单纯形法（Danzig，1947）**

单纯形法的基本思想是从可行域的一个顶点出发，寻找一个比之更“好”的顶点；由此不断改进，最后得到最优解。



- **内点法（Karmarkar，1984）**

多项式时间算法，从可行域内部逐渐逼近最优解。



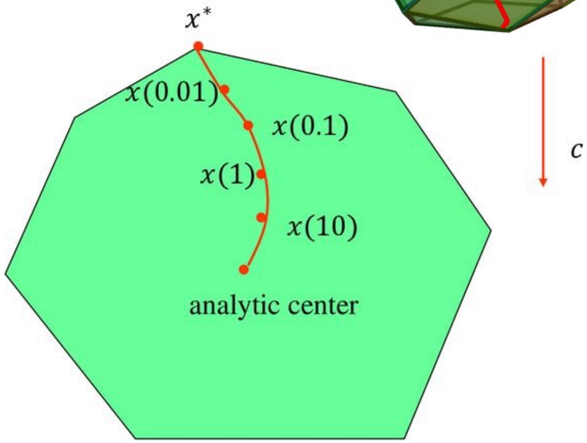


Figure 9.4: The central path and the analytic center

第一章 线性规划

6. 线性规划求解工具

求解线性规划的常用插件和软件

➢ Excel Solver

- 以Excel2016为例，文件 → 选项 → 加载项 → 管理（Excel加载项）→ 转到... → 规划求解加载项 Solver add-in（打勾并确定）。
- 在“数据”标签下即出现“规划求解”功能。

➢ LINDO/LINGO　http://www.lindo.com/

- LINDO和LINGO是美国LINDO系统公司(LINDO SYSTEM INC.)开发的一套专门用于求解最优化问题的软件包。
- LINDO用于求解线性规划和二次规划，LINGO除了具有LINDO的全部功能外，还可以用于求解非线性规划，也可以用于一些线性和非线性方程组的求解以及代数方程求根等。

LINDO: Linear INteractive and Discrete Optimizer
LINGO: Linear INteractive General Optimizer

➢ 用MATLAB函数linprog求解线性规划

$$
\begin{aligned}
&\min \quad c^T x \\
&s.t. \begin{cases} Ax \le b \\ Aeqx = beq \\ lb \le x \le ub \end{cases}
\end{aligned}
$$

基本调用格式为：

```
[x,fval]=linprog(c,A,b,Aeq,beq,lb,ub)
```

➢ 用MATLAB函数linprog求解线性规划

例：使用linprog求解以下线性规划：

$$\min \quad x_1 + 2x_2 + 3x_3$$
$$s.t. \begin{cases} 2x_1 - 3x_2 + 4x_3 = -2 \\ x_1, x_2, x_3 \geq 0 \end{cases}$$

输入：c=[1 2 3]’, A=[ ], b=[ ], Aeq=[2 -3 4], beq=-2, lb=[0 0 0]’, ub=[inf inf inf]’

执行：

```
[x,fval]=linprog(c,A,b,Aeq,beq,lb,ub)
```

得到最优解：x=(0,0.6667,0)^T，最优解为1.3333.

大规模优化求解器

大规模优化求解器一般针对问题规模比较庞大，变量和约束数量达到几万或者以上级别的问题。运筹学领域公认的大规模优化器中，商业优化器包括IBM Cplex, Gurobi, Fico Xpress等；免费优化器包括SCIP, CBC, GLPK等。

➢ **CPLEX** https://www.ibm.com/analytics/cplex-optimizer

- IBM ILOG CPLEX是IBM公司的一个优化引擎，该优化引擎可以用来求解线性规划 (Linear Programming)、二次规划 (Quadratic Programming)、二次约束规划 (Quadratically Constrained Programming) 和混合整数规划 (Mixed Integer Programming) 问题。
- CPLEX Studio IDE中自带该优化引擎，这是IBM官方出的CPLEX工具平台。

大规模优化求解器

➢ Gurobi   http://www.gurobi.com/, http://www.gurobi.cn/

- Gurobi 是由美国 Gurobi Optimization 公司开发新一代大规模优化求解器（优化引擎）。
- Gurobi 可以解决的数学问题包括线性问题 (Linear Problems)、二次型目标问题(Quadratic Problems)、混合整数线性和二次型问题 (Mixed Integer Linear and Quadratic problems) 。
- Gurobi 支持 Python, R, Matlab, C, C++, Java, .Net 等。

GUROBI OPTIMIZATION

大规模优化求解器

- 优化求解器是数学规划建模后求解这些实际问题的核心，这个核心长期被国外软件公司所垄断，华人在此领域在2017年前几乎是空白。
- 中国被“卡脖子”之处

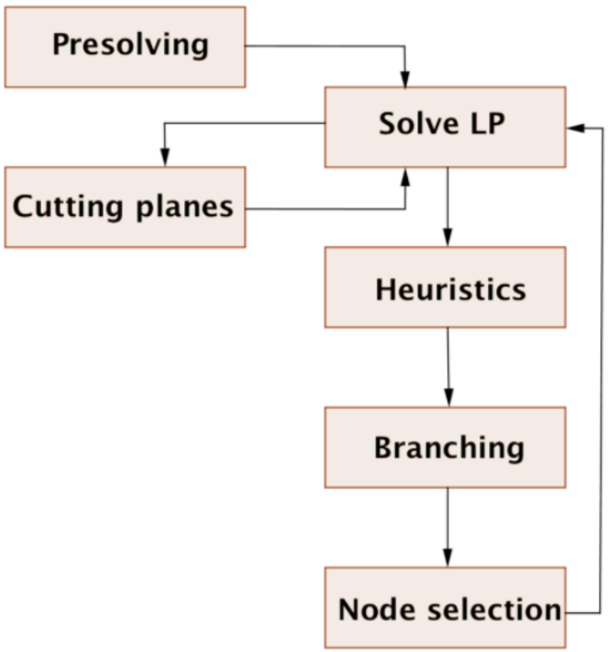
求解器	国家	发布年份	求解器	国家	发布年份
XPRESS	英国	1983	CPLEX	美国	1988
GLPK	俄国	2000	MOSEK	丹麦	2000
SYMPHONY	美国	2000	SCIP	德国	2001
LPSOLVE	芬兰	2004	SAS	美国	2004
CBC	美国	2005	GUROBI	美国	2009
MATLAB	美国	2014	MIPCL	俄国	2015
CMIP	中国	2018	COPT	中国	2019

中国的优化求解器

➢ **CMIP (C++ Mixed Integer Programming)**

http://copt.amss.ac.cn/

- 中国科学院数学与系统科学院戴彧虹研究员带领CMIP团队从2015年开始，历经30个月，终于自主研发了我国第一个具有国际水平的整数规划求解器CMIP，并于2018年3月确定版本为CMIP 1.0版本。



大规模优化求解器

中国的优化求解器

➢ **COPT (Cardinal Optimizer)**

https://www.shanshu.ai/copt

- COPT 是杉数科技自主研发的一款针对大规模优化问题的高效数学规划求解器套件。目前已经可以为客户提供线性规划问题、整数规划问题、非线性问题等多种数学规划求解方案。
- 提供公共下载测试的COPT是杉数求解器的线性规划求解器部分。其高效地实现了单纯形算法，可用于快速求解线性规划问题。
- 支持所有主流操作系统（均为64位系统），包括：Windows、Linux和MacOS，并提供以下接口：Python、PuLP、Pyomo、C、C++、C#、Java、AMPL和GAMS。

中国的优化求解器

➢ MDOPT (阿里巴巴达摩院研发的数学规划求解器)

- 2020年8月，达摩院决策智能实验室推出的MindOpt是一款具备线性规划等多种功能的求解器，其中单纯形法模块首先发布。
- 应用于阿里集团多项业务，包括云计算资源调度、金融资金分配、新零售智能营销等，每年可为阿里云节约数亿元成本。

The simplex methods were tested of the codes:

```
MOSEK-9.2.10      www.mosek.com
CLP-1.17.6        projects.coin-or.org/Clp
Google-GLOP       LP with Glop
SOPLEX-5.0.0      soplex.zib.de/
Gurobi-9.0.2      Gurobi
GLPK-4.65         www.gnu.org/software/glpk/glpk.html
MATLAB-R2020a     mathworks.com (dual-simplex)
COPT-1.2          COPT
MindOpt-0.9.0     MindOpt
HiGHS-1.0.0       HiGHS
SAS-OR-15.1       SAS (dual-simplex)
```

达摩院自研线性规划求解器MindOpt

Scaled shifted (by 10 sec) geometric mean of runtimes

40 probs	MSK	CLP	GLOP	SPLX	Gurob	GLPK	MATL	COPT	MDOPT	HiGHS	SAS
	6.00	3.01	13.8	16.1	1.10	53.9	15.9	1.41	1	10.4	6.12
solved	38	40	35	39	40	31	33	40	40	37	37

大规模优化求解器

- 2020年8月28日，杉数科技发布了数学规划求解器COPT的单纯形法模块的第三次更新，并在国际权威第三方测评Mittelmann的线性规划单纯形法榜单中重回排名第一。

```
27 Aug 2020     ====================================
                 Benchmark of Simplex LP solvers
                ====================================
                 H. Mittelmann (mittelmann@asu.edu)
```

Logfiles of these runs at: plato.asu.edu/ftp/lp_logs/

This benchmark was run on a Linux-PC (i7-7700K, 4.2GHz, 32GB).
The MPS-datafiles for all testcases are in one of (see column "s")

```
 miplib2010.zib.de/ [1]
 plato.asu.edu/ftp/lptestset/ [2]
 www.netlib.org/lp/data/ [3,7]
 www.sztaki.hu/~meszaros/public_ftp/lptestset/
(MISC[4], PROBLEMATIC[5], STOCHLP[6], INFEAS[8])
```

NOTE: some files in [2-8] need to be expanded with emps in same directory!

The simplex methods were tested of the codes:

```
 MOSEK-9.2.10      www.mosek.com
 CLP-1.17.6        projects.coin-or.org/Clp
 Google-GLOP       LP with Glop
 SOPLEX-5.0.0      soplex.zib.de/
 Gurobi-9.0.2      Gurobi
 GLPK-4.65         www.gnu.org/software/glpk/glpk.html
 MATLAB-R2020a     mathworks.com (dual-simplex)
 COPT-1.4          COPT
 MindOpt-0.9.0     MindOpt
 HiGHS-1.0.0       HiGHS
 SAS-OR-15.1       SAS (dual-simplex)
```

Scaled shifted (by 10 sec) geometric mean of runtimes

40 probs	MSK	CLP	GLOP	SPLX	Gurob	GLPK	MATL	COPT	MDOPT	HiGHS	SAS
	6.82	3.47	15.7	18.2	1.25	61.2	18.1	1	1.14	11.8	6.96
solved	38	40	35	39	40	31	33	40	40	37	37

大规模优化求解器

- 2020年10月26日，杉数优化求解器COPT公布国内首个线性规划内点法求解模块，其性能与单纯形、大规模网络规划均在国际权威第三方测评Mittelmann平台稳居前三甲，标志着杉数优化求解器COPT已跻身世界一流水准。
- 2020年10月18日**杉数科技凭借在数学规划求解器方面的突出贡献，荣获2020年度中国运筹学会运筹应用奖。**

线性规划内点法测评

```
22 Oct 2020   ==================================
               Benchmark of Barrier LP solvers
              ==================================
               H. Mittelmann (mittelmann@asu.edu)
```

Logfiles of these runs at: plato.asu.edu/ftp/lp_logs/

This benchmark was run on a Linux-PC (i7-7700K, 4.2GHz, 32GB).
The MPS-datafiles for all testcases are in one of (see column "s")

```
 miplib2010.zib.de/ [1]
 plato.asu.edu/ftp/lptestset/ [2]
 www.netlib.org/lp/data/ [3,7]
 www.sztaki.hu/~meszaros/public_ftp/lptestset/
(MISC[4], PROBLEMATIC[5], STOCHLP[6], INFEAS[8], NEW[9])
```

NOTE: Most files in [2-9] need to be expanded with emps in same directory!

The barrier methods were tested of:

```
 MOSEK-9.2.21     www.mosek.com
 MATLAB-R2020a    mathworks.com (interior-point, NO CROSSOVER!)
 CLP-1.17.6       projects.coin-or.org/Clp
 SAS-OR-15.1:     SAS
 Tulip-0.5.1:     Tulip (NO CROSSOVER!)
 COPT-1.4.3       COPT
 More codes to be added
```

Scaled shifted (by 10 sec) geometric mean of runtimes

45 probs	1.69	22.0	1	34.0	3.77	25.9	1.71
solved	44	34	45	37	43	35	45

problem	MOSEK	MATLAB	Gurobi	CLP	SAS	TULIP	COPT
L1_six1000	1997	f	106	f	196	f	224
degme	260	7211	44	3944	62	5610	135
karted	451	132	46	1227	15	1660	243
tp-6	151	2299	32	1218	63	2499	79
ts-palko	23	71	33	380	12	424	158
L1_sixm250	39	f	6	1985	38	f	3
Linf_520c	2767	f	13	f	19	t	13
buildingen	6	410	7	110	18	26	30

第一章 线性规划

7. 线性规划模型的应用

7.1 混合配料问题

某糖果厂用原料A，B，C加工成三种不同牌号的糖果甲、乙、丙。已知各种牌号糖果中A，B，C含量，原料成本，各种原料的每月限制用量，三种牌号糖果的单位加工费及售价如下表。问该厂每月生产这三种牌号糖果各多少kg，使其获利最大。

原料	甲	乙	丙	原料成本(元/kg)	每月限制用量(kg)
A	≥60%	≥30%		2.00	2000
B				1.50	2500
C	≤20%	≤50%	≤60%	1.00	1200
加工费(元/kg)	0.50	0.40	0.30		
售　价(元/kg)	3.40	2.85	2.25		

确定变量：

用$i=1,2,3$分别代表原料 A, B, C, 用$j=1,2,3$分别代表甲、乙、丙三种糖果；x_{ij} 为生产第 j 种糖果耗用的第 i 种原料的kg数，则

三种糖果的产量分别为：	三种原料的消耗量分别为：
甲： $x_{11}+x_{21}+x_{31}$	A： $x_{11}+x_{12}+x_{13}$
乙： $x_{12}+x_{22}+x_{32}$	B： $x_{21}+x_{22}+x_{32}$
丙： $x_{13}+x_{23}+x_{33}$	C： $x_{31}+x_{32}+x_{33}$

确定目标函数：

获利 = 三种糖果的售价 – 加工费 – 原料成本

$$\max z$$
$$= 3.40\,(x_{11}+x_{21}+x_{31})+2.85\,(x_{12}+x_{22}+x_{32}) + 2.25\,(x_{13}+x_{23}+x_{33})$$
$$-0.50\,(x_{11}+x_{21}+x_{31})-0.40\,(x_{12}+x_{22} + x_{32})-0.30\,(x_{13}+x_{23}+x_{33})$$
$$-2.0\,(x_{11}+x_{12}+x_{13})-1.50\,(x_{21}+x_{22} + x_{23})-1.0\,(x_{31}+x_{32}+x_{33})$$

确定约束条件：

原料月供应量限制：

$$x_{11} + x_{12} + x_{13} \leq 2000$$
$$x_{21} + x_{22} + x_{23} \leq 2500$$
$$x_{31} + x_{32} + x_{33} \leq 1200$$

含量成分限制：

$$x_{11} \geq 0.6\,(x_{11} + x_{21} + x_{31})$$
$$x_{31} \leq 0.2\,(x_{11} + x_{21} + x_{31})$$
$$x_{12} \geq 0.3\,(x_{12} + x_{22} + x_{32})$$
$$x_{32} \leq 0.5\,(x_{12} + x_{22} + x_{32})$$
$$x_{33} \leq 0.6\,(x_{13} + x_{23} + x_{33})$$

非负限制：

$$x_{ij} \geq 0$$
$$(i = 1,2,3; j = 1,2,3)$$

7.2 连续投资问题

某部门在五年内考虑给下列项目投资：

项目A：从第一年到第四年每年年初需要投资，并于次年末回收本利115%。

项目B：从第三年初需投资，到第五年末能回收本利125%，但规定最大投资额不超过4万元。

项目C：第二年初需要投资，到第五年末能回收本利140%，但规定最大投资额不超过3万元。

项目D：五年内每年初可购买公债于当年末归还并加利息6%。

该部门现有资金10万元，问应如何确定给这些项目每年的投资额，使到第五年末拥有的资金本利总额最大。

确定变量：

以 $X_{iA}, X_{iB}, X_{iC}, X_{iD}$ (i=1, 2, 3, 4, 5) 分别表示第 i 年年初给项目A, B, C, D 的投资额。具体列于下表：

	第1年	第2年	第3年	第4年	第5年
A	X_{1A}	X_{2A}	X_{3A}	X_{4A}	
B			X_{3B}		
C		X_{2C}			
D	X_{1D}	X_{2D}	X_{3D}	X_{4D}	X_{5D}

确定目标函数： 第五年末资金最大

$$\max\ z = 1.15X_{4A} + 1.40X_{2C} + 1.25X_{3B} + 1.06X_{5D}$$

确定约束条件：

投资额应等于手中拥有的资金额，不应有剩余的呆滞资金。

第一年： $X_{1A} + X_{1D} = 100000$

第二年： 年初资金仅为 D 项目第一年的本息。

$$X_{2A} + X_{2C} + X_{2D} = X_{1D}(1 + 6\%)$$

第三年： 年初资金为 A 第一次回收本利 + D 第二次本息

$$X_{3A} + X_{3B} + X_{3D} = X_{1A}(1 + 15\%) + X_{2D}(1 + 6\%)$$

第四年： $X_{4A} + X_{4D} = X_{2A}(1 + 15\%) + X_{3D}(1 + 6\%)$

第五年： $X_{5D} = X_{3A}(1 + 15\%) + X_{4D}(1 + 6\%)$

B、C投资额限制：$X_{3B} \le 40000$; $X_{2C} \le 30000$

完整的模型为：

$$\max\ z = 1.15x_{4A} + 1.25x_{3B} + 1.4x_{2C} + 1.06x_{5D}$$

$$s.t.\begin{cases} x_{1A} + x_{1D} = 100000 \\ x_{2A} + x_{2C} + x_{2D} - 1.06x_{1D} = 0 \\ x_{3A} + x_{3B} + x_{3D} - 1.15x_{1A} - 1.06x_{2D} = 0 \\ x_{4A} + x_{4D} - 1.15x_{2A} - 1.06x_{3D} = 0 \\ x_{5D} - 1.15x_{3A} - 1.06x_{4D} = 0 \\ x_{3B} \le 40000 \\ x_{2C} \le 30000 \\ x_{iA}, x_{iB}, x_{iC}, x_{iD} \ge 0, \quad i = 1,2,\cdots,5 \end{cases}$$

求解结果（单位：元）：

第一年：$X_{1A}=34783, X_{1D}=65217$

第二年：$X_{2A}=39130, X_{2C}=30000, X_{2D}=0$

第三年：$X_{3A}=0, X_{3B}=40000, X_{3D}=0$

第四年：$X_{4A}=45000, X_{4D}=0$

第五年：$X_{5D}=0$

第五年末总资金：$z=143750$

7.3 生产贮存问题

某厂签订了5种产品（$i=1,\ldots,5$）上半年的交货合同。已知各产品在第j月（$j=1,\ldots,6$）的合同交货量D_{ij}，该月售价s_{ij}，成本价c_{ij}及生产1件时所需工时a_{ij}。该厂第j月的正常生产工时为t_j，但必要时可加班生产，第j月允许的最大加班工时不超过t_j'，并且加班时间生产出来的产品每件成本增加额外费用c_{ij}'元。若生产出来的产品如当月不交货，每件库存一个月交存贮费p_i元。

试为该工厂设计一个保证完成交货合同，又使上半年预期盈利总额为最大的生产计划安排。

解：设x_{ij}为 i 种产品 j 月份在正常时间内生产的数量，x'_{ij}为第 i 种产品 j 月份在加班时间内生产的数量。

<u>目标函数</u>：

该厂盈利总额 = 5种产品销售价 – 成本 – 库存费用

$$\max z = \sum_{i=1}^{5}\sum_{j=1}^{6} s_{ij}D_{ij} - \sum_{i=1}^{5}\sum_{j=1}^{6}\left[c_{ij}x_{ij} + (c_{ij} + c'_{ij})x'_{ij}\right] - \sum_{i=1}^{5}\sum_{j=1}^{6}\left[p_i \sum_{k=1}^{j}(x_{ik} + x'_{ik} - D_{ik})\right]$$

约束条件：

各个月的正常和加班的允许工时：

$$\sum_{i=1}^{5} a_{ij}x_{ij} \le t_j \quad (j = 1, \cdots, 6)$$

$$\sum_{i=1}^{5} a_{ij}x'_{ij} \le t'_j \quad (j = 1, \cdots, 6)$$

各个月的交货要求：

$$\sum_{k=1}^{j} (x_{ik} + x'_{ik}) \ge \sum_{k=1}^{j} D_{ik} \quad (j = 1, \cdots, 6)$$

非负要求： $x_{ij} \ge 0,\ x'_{ij} \ge 0$

7.4 人力资源问题

某昼夜服务的公交线路每天各时间段内所需司机和乘务人员数如下表所示。设司机和乘务人员分别在各时间段一开始时上班，并连续工作八小时，问该公交线路怎样安排司机和乘务人员，既能满足工作需要，又配备最少司机和乘务人员?

□□	蜉□	枦□尘竦
1	06▭00 》 10▭00	60
2	10▭00 》 14▭00	70
3	14▭00 》 18▭00	60
4	18▭00 》 22▭00	50
5	22▭00 》 02▭00	20
6	02▭00 》 06▭00	30

解：设 x_i 表示第 i 班次时开始上班的司机和乘务人员数，这样我们建立如下的数学模型。

<u>目标函数</u>： $\min \; x_1 + x_2 + x_3 + x_4 + x_5 + x_6$

<u>约束条件</u>：

$$
\begin{aligned}
x_1 + x_6 &\geq 60 \\
x_1 + x_2 &\geq 70 \\
x_2 + x_3 &\geq 60 \\
x_3 + x_4 &\geq 50 \\
x_4 + x_5 &\geq 20 \\
x_5 + x_6 &\geq 30 \\
x_1, x_2, x_3, x_4, x_5, x_6 &\geq 0
\end{aligned}
$$

7.6 Business or Personal Spend?

南京大學 NANJING UNIVERSITY 工程管理学院 SCHOOL OF MANAGEMENT & ENGINEERING

- *A Case of Optimization for Data Analytics (from Mark Peters and Yinyu Ye, Stanford University)*
  - For better risk-management of the business or corporate credit card financial service, banks or credit agents would like to know probabilities for each type of transactions indicating whether it is a business-spend or personal usage.
  - We're attempting to build such a model to achieve this objective based on historical business data.
  - Since there is no training data where particular transactions are identified as being personal, we use total personal remittances as the best proxy.
  - On the transaction side, the industry code of each transaction is a key initial differentiator between transactions.

7.6 Business or Personal Spend?

南京大学 NANJING UNIVERSITY 工程管理学院 SCHOOL OF MANAGEMENT & ENGINEERING

Transaction Types by Industrial Codes

Industry Code	Description
995	CLUB - WAREHOUSE
25	DEPARTMENT STORE - MASS MERCHANDISER
728	GASOLINE/OIL COMPANY - NATIONAL DEALER
729	GASOLINE/OIL COMPANY - INDEPENDENT DEALER
429	SHOP - HOME IMPROVEMENT
415	DEPARTMENT STORE - FULL SERVICE
87	INTERNET TRAVEL
504	SHOP - ELECTRONIC GOODS
616	COMMUNICATION - CABLE & BROADCAST SERVICES
215	AUTO SERVICES - MOTOR RELATED SERVICES/DEALER
404	AUTO SERVICES - AUTO SALES & SERVICE
443	SHOP - SPORTING GOODS
457	SHOP - CHEMIST/PHARMACY
522	SHOP - FURNITURE
463	SHOP - JEWELRY
757	ENTERTAINMENT - TICKET AGENT - COMPANY
407	SHOP - CLOTHING - FAMILY
680	SHOP - COMPUTER HARDWARE
465	SHOP - LIQUOR STORE
400	AUTO SERVICES - VEHICLE ACCESSORIES
416	DEPARTMENT STORE - SPECIALITY
428	SHOP - HOME FURNISHINGS
414	SHOP - CLOTHING - WOMEN'S
793	TRAVEL - TOUR OPERATOR GENERAL
412	SHOP - CLOTHING - MEN'S & WOMEN'S
787	TRAVEL - NON - AGENT RETAILER
447	SHOP - SHOES - MEN'S ONLY
427	SHOP - HARDWARE/DO IT YOURSELF
554	MAIL ORDER SELF IMPROVEMENT/BUSINESS SEMINARS
603	SERVICES - BEAUTY SHOPS/BEAUTICIAN

7.6 Business or Personal Spend?

Personal Remittances

A typical transaction data in a period

Each Column represents an Industry Code

Personal Remittances

Account	1	2	3	…	n			Actual	
1	$156	$0	$87		$25			$200	
2	$200	$25	$0		$0			$195	
…	$0	$134	$35		$60			$210	

Value of a transaction in the period

A Model to Compute Probabilities

For each of the industry codes, the model will determine a probability (in red) which indicates the likelihood that a transaction was personal. The goal is to minimize the sum of the squares of the differences (in blue).

Probability Personal

Each Column represents an Industry Code

Personal Remittances

	25%	10%	0%	…	5%				
Account	**1**	**2**	**3**	**…**	**n**		**Predicted**	**Actual**	**Difference**
1	$156	$0	$87		$25		$244	$200	$44
2	$200	$25	$0		$0		$200	$195	$5
…	$0	$134	$35		$60		$230	$210	$20

Value of transactions in the period

Mathematical Model Formulation

Our model will determine the probability that a transaction from each industry code is personal in such a manner which will minimize the sum of the squared errors (between predicted personal remittances and actual personal remittances).

$$\min \sum_i \left| \sum_j a_{ij} x_j - r_i \right| + \lambda \|x\|_1$$

$$\text{s.t.} \quad 0 \le x_j \le 1$$

add a L_1 norm regularization to reduce the number of non-zero entries in x

In our model, we will use the following variables and parameters:

- Let x_j be such a probability that a transaction is personal for industry code j
- a_{ij} – transaction amount for account i and industry code j
- r_i – amount paid by personal remit for account i
- $\sum_j a_{ij} x_j$ – the expected personal expenses for account i

We'd like to choose x_j such that $\sum_j a_{ij} x_j$ matches b_i for ALL i

7.7 数据包络分析(DEA)

Data Envelopment Analysis

— A.Charnes,W.W.Cooper & E.Rhodes (1978)

是一种基于线性规划的用于评价同类型组织工作绩效相对有效性的工具手段。

如：学校，银行，超市，医院……

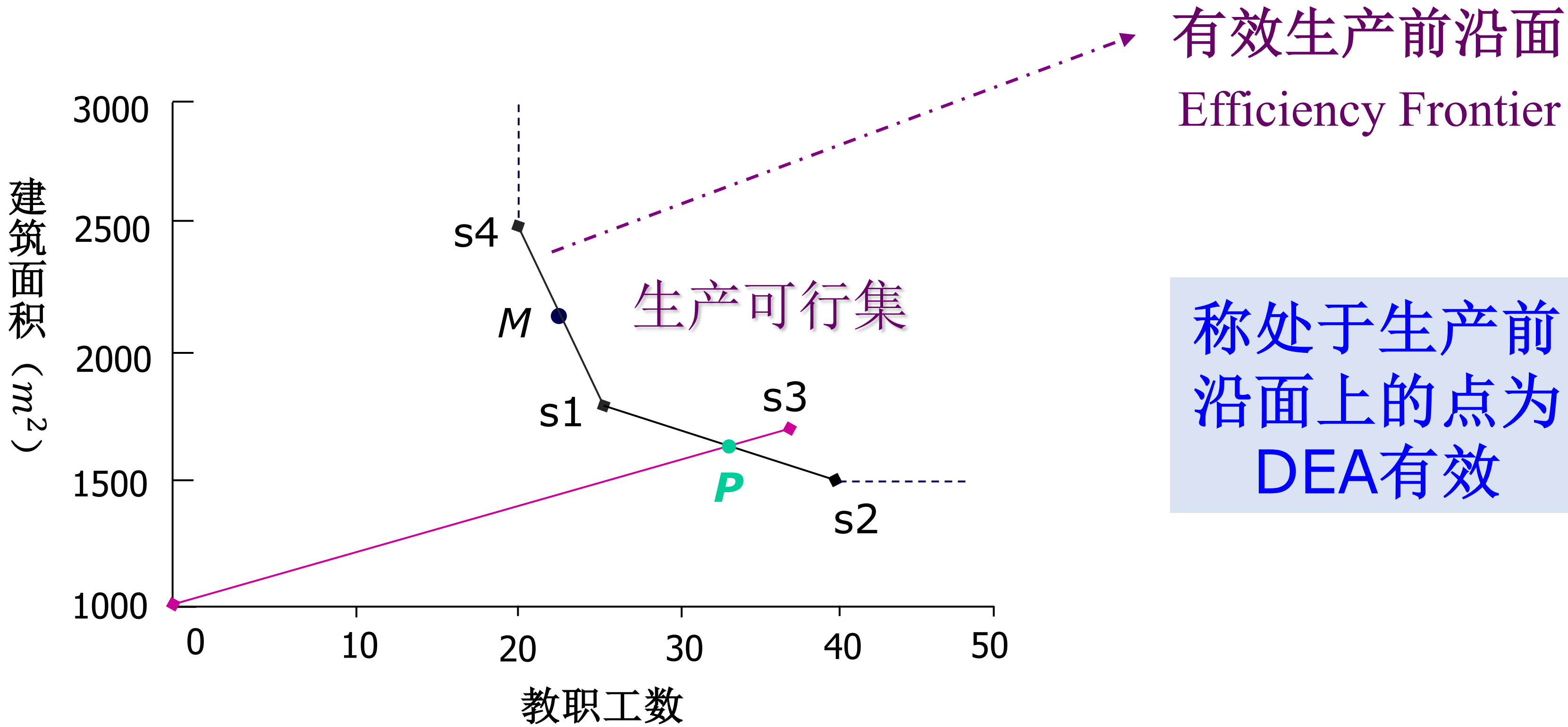
衡量这类组织之间的绩效高低，通常采用**“投入产出”**这个指标。

- 当投入和产出均可折算成同一单位时，容易计算比较。
- 当有多项投入和多项产出，且不能统一折算时，就无法计算投入产出比。

例如：有 4 所小学s1，s2，s3，s4，在校学生分别为1200，1000，1600，1400人，首先按800名学生的规模折算各校教职工数和建筑面积的投入。

学校 投入	s1	s2	s3	s4
教职工数	25	40	35	20
建筑面积/m^2	1800	1500	1700	2500

7.7 数据包络分析(DEA)



称处于生产前沿面上的点为DEA有效

就培养800名学生来看，s1，s2，s4 三所学校的投入处于Pareto最优状态，即不可能保持其中一项投入不变的情况下，减少另一项投入的水平。

7.7 数据包络分析(DEA)

当投入与产出变量数总和超过3个时，就需要借助线性规划的方法。

在DEA中通常称被衡量绩效的组织为决策单元（decision making unit，缩写DMU）

设有n个决策单元$(j=1,\dots,n)$，每个决策单元有相同的m项投入$(i=1,\dots,m)$和相同的 s 项产出$(r=1,\dots,s)$)。

用 x_{ij} 表示第 j 决策单元的第 i 项投入，用 y_{rj} 表示第 j 决策单元的第 r 项产出。

决策单元

投入

$$
\begin{array}{c|cccc}
 & 1 & 2 & \dots & n \\
\hline
1 \rightarrow & x_{11} & x_{12} & \cdots & x_{1n} \\
2 \rightarrow & x_{21} & x_{22} & \cdots & x_{2n} \\
\mathrm{M} & \vdots & \vdots & x_{ij} & \vdots \\
m \rightarrow & x_{m1} & x_{m2} & \cdots & x_{mn}
\end{array}
$$

$$
\begin{array}{cccc|c}
y_{11} & y_{12} & \cdots & y_{1n} & \rightarrow 1 \\
y_{21} & y_{22} & \cdots & y_{2n} & \rightarrow 2 \\
\vdots & \vdots & y_{rj} & \vdots & \mathrm{M} \\
y_{s1} & y_{s2} & \cdots & y_{sn} & \rightarrow s
\end{array}
$$

产出

构造一个假想决策单元，其

第 i 项投入为：

$$\sum_{j=1}^{n} \lambda_j x_{ij}\ (i = 1, \dots, m) \text{ 且 } \sum_{j=1}^{n} \lambda_j = 1\ (\lambda_j \geq 0)$$

第 r 项产出为：

$$\sum_{j=1}^{n} \lambda_j y_{rj}\ (r = 1, \dots, s) \text{ 且 } \sum_{j=1}^{n} \lambda_j = 1\ (\lambda_j \geq 0)$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rj_0}\ (r = 1, \dots, s)$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq E \cdot x_{ij_0}\ (i = 1, \dots, m, E < 1)$$

说明 j_0 决策单元不处在生产前沿面上

7.7 数据包络分析(DEA)

基于上述可写出如下的线性规划模型：

$$\min \ E$$

$$s.t.\begin{cases} \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rj_0} \ (r = 1, \dots, s) \\ \sum_{j=1}^{n} \lambda_j y_{rj} \leq E \cdot x_{ij_0} \ (r = 1, \dots, m) \\ \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0 \ (j = 1, \dots, n) \end{cases}$$

当求解结果有$E < 1$时，则 j_0 决策单元非DEA有效，否则 j_0 决策单元DEA 有效。

例：振华银行的4个分理处的投入产出情况如表所示。要求分别确定各分理处的运行是否DEA有效。

分理处	投入		产出（处理笔数/月）		
	职员数	营业面积 (m^2)	储蓄存取	贷款	中间业务
分理处1	15	140	1800	200	1600
分理处2	20	130	1000	350	1000
分理处3	21	120	800	450	1300
分理处4	20	135	900	420	1500

✓ 确定分理处1的运行是否DEA有效？

$$\min \quad E$$

$$s.t.\begin{cases} 1800\lambda_1 + 1000\lambda_2 + 800\lambda_3 + 900\lambda_4 \geq 1800 \\ 200\lambda_1 + 350\lambda_2 + 450\lambda_3 + 420\lambda_4 \geq 200 \\ 1600\lambda_1 + 1000\lambda_2 + 1300\lambda_3 + 1500\lambda_4 \geq 1600 \\ 15\lambda_1 + 20\lambda_2 + 21\lambda_3 + 20\lambda_4 \leq 15E \\ 140\lambda_1 + 130\lambda_2 + 120\lambda_3 + 135\lambda_4 \leq 140E \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1 \\ \lambda_j \geq 1 \quad (j = 1, \dots, 4) \end{cases}$$

求解结果为 $E = 1$，说明分理处1的运行为DEA有效。