from *SIAM News*, Volume 33, Number 4

The Best of the 20th Century: Editors Name Top 10 Algorithms

By Barry A. Cipra

Algos is the Greek word for pain. *Algor* is Latin, to be cold. Neither is the root for *algorithm*, which stems instead from al-Khwarizmi, the name of the ninth-century Arab scholar whose book *al-jabr wa'l muqabalah* devolved into today's high school algebra textbooks. Al-Khwarizmi stressed the importance of methodical procedures for solving problems. Were he around today, he'd no doubt be impressed by the advances in his eponymous approach.

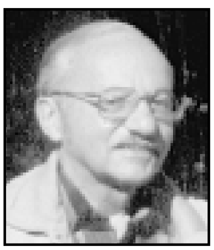
Some of the very best algorithms of the computer age are highlighted in the January/February 2000 issue of *Computing in Science & Engineering*, a joint publication of the American Institute of Physics and the IEEE Computer Society. Guest editors Jack Don-garra of the University of Tennessee and Oak Ridge National Laboratory and Fran-cis Sullivan of the Center for Comput-ing Sciences at the Institute for Defense Analyses put togeth-er a list they call the "Top Ten Algorithms of the Century."

"We tried to assemble the 10 al-gorithms with the greatest influence on the development and practice of science and engineering in the 20th century," Dongarra and Sullivan write. As with any top-10 list, their selections—and non-selections—are bound to be controversial, they acknowledge. When it comes to picking the algorithmic best, there seems to be no best algorithm.

Without further ado, here's the CiSE top-10 list, in chronological order. (Dates and names associated with the algorithms should be read as first-order approximations. Most algorithms take shape over time, with many contributors.)

1946: John von Neumann, Stan Ulam, and Nick Metropolis, all at the Los Alamos Scientific Laboratory, cook up the Metropolis algorithm, also known as the **Monte Carlo method**.

The Metropolis algorithm aims to obtain approximate solutions to numerical problems with unmanageably many degrees of freedom and to combinatorial problems of factorial size, by mimicking a random process. Given the digital computer's reputation for deterministic calculation, it's fitting that one of its earliest applications was the generation of random numbers.



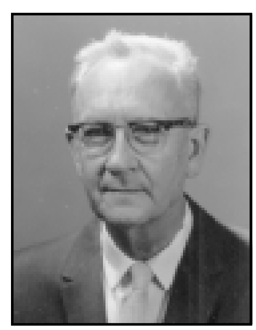
In terms of widespread use, George Dantzig's simplex method is among the most successful algorithms of all time.

1947: George Dantzig, at the RAND Corporation, creates the **simplex method for linear programming**.

In terms of widespread application, Dantzig's algorithm is one of the most successful of all time: Linear programming dominates the world of industry, where economic survival depends on the ability to optimize within budgetary and other constraints. (Of course, the "real" problems of industry are often nonlinear; the use of linear programming is sometimes dictated by the computational budget.) The simplex method is an elegant way of arriving at optimal answers. Although theoretically susceptible to exponential delays, the algorithm in practice is highly efficient—which in itself says something interesting about the nature of computation.

1950: Magnus Hestenes, Eduard Stiefel, and Cornelius Lanczos, all from the Institute for Numerical Analysis at the National Bureau of Standards, initiate the development of **Krylov subspace iteration methods**.

These algorithms address the seemingly simple task of solving equations of the form $Ax = b$. The catch, of course, is that A is a huge $n \times n$ matrix, so that the algebraic answer $x = b/A$ is not so easy to compute. (Indeed, matrix "division" is not a particularly useful concept.) Iterative methods—such as solving equations of the form $Kx_{i+1} = Kx_i + b - Ax_i$ with a simpler matrix K that's ideally "close" to A—lead to the study of Krylov subspaces. Named for the Russian mathematician Nikolai Krylov, Krylov subspaces are spanned by powers of a matrix applied to an initial "remainder" vector $r_0 = b - Ax_0$. Lanczos found a nifty way to generate an orthogonal basis for such a subspace when the matrix is symmetric. Hestenes and Stiefel proposed an even niftier method, known as the conjugate gradient method, for systems that are both symmetric and positive definite. Over the last 50 years, numerous researchers have improved and extended these algorithms. The current suite includes techniques for non-symmetric systems, with acronyms like GMRES and Bi-CGSTAB. (GMRES and Bi-CGSTAB premiered in *SIAM Journal on Scientific and Statistical Computing*, in 1986 and 1992, respectively.)



Alston Householder

1951: Alston Householder of Oak Ridge National Laboratory formalizes the **decompositional approach to matrix computations**.

The ability to factor matrices into triangular, diagonal, orthogonal, and other special forms has turned out to be extremely useful. The decompositional approach has enabled software developers to produce flexible and efficient matrix packages. It also facilitates the analysis of rounding errors, one of the big bugbears of numerical linear algebra. (In 1961, James Wilkinson of the National Physical Laboratory in London published a seminal paper in the *Journal of the ACM*, titled "Error Analysis of Direct Methods of Matrix Inversion," based on the LU decomposition of a matrix as a product of lower and upper triangular factors.)

1957: John Backus leads a team at IBM in developing the **Fortran optimizing compiler**.

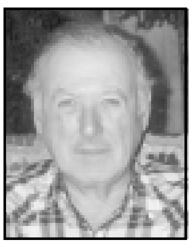
The creation of Fortran may rank as the single most important event in the history of computer programming: Finally, scientists (and others) could tell the computer what they wanted it to do, without having to descend into the netherworld of machine code. Although modest by modern compiler standards—Fortran I consisted of a mere 23,500 assembly-language instructions—the early compiler was nonetheless capable of surprisingly sophisticated computations. As Backus himself recalls in a recent history of Fortran I, II, and III, published in 1998 in the *IEEE Annals of the History of Computing*, the compiler "produced code of such efficiency that its output would startle the programmers who studied it."

1959–61: J.G.F. Francis of Ferranti Ltd., London, finds a stable method for computing eigenvalues, known as the **QR algorithm**.

Eigenvalues are arguably the most important numbers associated with matrices—and they can be the trickiest to compute. It's relatively easy to transform a square matrix into a matrix that's "almost" upper triangular, meaning one with a single extra set of nonzero entries just below the main diagonal. But chipping away those final nonzeros, without launching an avalanche of error, is nontrivial. The QR algorithm is just the ticket. Based on the QR decomposition, which writes A as the product of an orthogonal matrix Q and an upper triangular matrix R, this approach iteratively changes $A_i = QR$ into $A_{i+1} = RQ$, with a few bells and whistles for accelerating convergence to upper triangular form. By the mid-1960s, the QR algorithm had turned once-formidable eigenvalue problems into routine calculations.

1962: Tony Hoare of Elliott Brothers, Ltd., London, presents **Quicksort**.

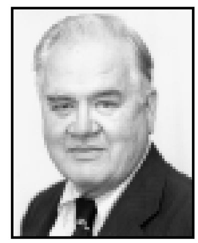
Putting N things in numerical or alphabetical order is mind-numbingly mundane. The intellectual challenge lies in devising ways of doing so quickly. Hoare's algorithm uses the age-old recursive strategy of divide and conquer to solve the problem: Pick one element as a "pivot," separate the rest into piles of "big" and "small" elements (as compared with the pivot), and then repeat this procedure on each pile. Although it's possible to get stuck doing all $N(N-1)/2$ comparisons (especially if you use as your pivot the first item on a list that's already sorted!), Quicksort runs on average with $O(N \log N)$ efficiency. Its elegant simplicity has made Quicksort the pos-terchild of computational complexity.



James Cooley

1965: James Cooley of the IBM T.J. Watson Research Center and John Tukey of Princeton University and AT&T Bell Laboratories unveil the **fast Fourier transform**.

Easily the most far-reaching algo-rithm in applied mathematics, the FFT revolutionized signal processing. The underlying idea goes back to Gauss (who needed to calculate orbits of asteroids), but it was the Cooley–Tukey paper that made it clear how easily Fourier transforms can be computed. Like Quicksort, the FFT relies on a divide-and-conquer strategy to reduce an ostensibly $O(N^2)$ chore to an $O(N \log N)$ frolic. But unlike Quick- sort, the implementation is (at first sight) nonintuitive and less than straightforward. This in itself gave computer science an impetus to investigate the inherent complexity of computational problems and algorithms.



John Tukey

1977: Helaman Ferguson and Rodney Forcade of Brigham Young University advance an **integer relation detection algorithm**.

The problem is an old one: Given a bunch of real numbers, say $x_1, x_2, \ldots, x_n$, are there integers $a_1, a_2, \ldots, a_n$ (not all 0) for which $a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$? For $n = 2$, the venerable Euclidean algorithm does the job, computing terms in the continued-fraction expansion of x_1/x_2. If x_1/x_2 is rational, the expansion terminates and, with proper unraveling, gives the "smallest" integers a_1 and a_2. If the Euclidean algorithm doesn't terminate—or if you simply get tired of computing it—then the unraveling procedure at least provides lower bounds on the size of the smallest integer relation. Ferguson and Forcade's generalization, although much more difficult to implement (and to understand), is also more powerful. Their detection algorithm, for example, has been used to find the precise coefficients of the polynomials satisfied by the third and fourth bifurcation points, $B_3 = 3.544090$ and $B_4 = 3.564407$, of the logistic map. (The latter polynomial is of degree 120; its largest coefficient is 257^{30}.) It has also proved useful in simplifying calculations with Feynman diagrams in quantum field theory.

1987: Leslie Greengard and Vladimir Rokhlin of Yale University invent the **fast multipole algorithm**.

This algorithm overcomes one of the biggest headaches of N-body simulations: the fact that accurate calculations of the motions of N particles interacting via gravitational or electrostatic forces (think stars in a galaxy, or atoms in a protein) would seem to require $O(N^2)$ computations—one for each pair of particles. The fast multipole algorithm gets by with $O(N)$ computations. It does so by using multipole expansions (net charge or mass, dipole moment, quadrupole, and so forth) to approximate the effects of a distant group of particles on a local group. A hierarchical decomposition of space is used to define ever-larger groups as distances increase. One of the distinct advantages of the fast multipole algorithm is that it comes equipped with rigorous error estimates, a feature that many methods lack.

What new insights and algorithms will the 21st century bring? The complete answer obviously won't be known for another hundred years. One thing seems certain, however. As Sullivan writes in the introduction to the top-10 list, "The new century is not going to be very restful for us, but it is not going to be dull either!"

Barry A. Cipra is a mathematician and writer based in Northfield, Minnesota.

"An intuitive algebraic approach for solving Linear Programming problems"

Source: Zionts [1974] (or many others).

$$\begin{array}{llllll} [\max] z = & 0{,}56x_1 & +0{,}42x_2 & & \\ \text{s. to} & x_1 & +2x_2 & \leq & 240 \\ & 1{,}5x_1 & +x_2 & \leq & 180 \\ & x_1 & & \leq & 110 \end{array} \quad \{1\}$$

A[1]

$$\begin{array}{llllllll} [\max] z = & 0{,}56x_1 & +0{,}42x_2 & & & & & \\ & x_1 & +2x_2 & +\{x_3\} & & & = & 240 \\ & 1{,}5x_1 & +x_2 & & +\{x_4\} & & = & 180 \\ & x_1 & & & & +\{x_5\} & = & 110 \end{array} \quad \{2\}$$

This has (always) an obvious, sure solution. Let

$$x_1, x_2 = 0 \quad \{3\}$$

Then

$$\begin{bmatrix} x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 240 \\ 180 \\ 110 \end{bmatrix} \quad \{4\}$$

$$z = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 240 \\ 180 \\ 110 \end{bmatrix} = 0 \quad \{5\}$$

Is this optimal ? How to improve ?

There does not appear (Dantzig) to be a systematic way of setting *all* the nonbasic variables simultaneously to optimal values —hence, an *iterative*[2] method.

Choose the variable that increases the objective function *most* per unit (this choice is arbitrary), in the example, x_1, because its coefficient (0,56) is the largest.

According to the constraints, x_1 can be increased till:

B

$$\begin{array}{ll} x_1 & = 240 \\ 1{,}5x_1 & = 180 \\ x_1 & = 110 \end{array} \quad \rightarrow \quad \begin{array}{ll} x_1 & = 240 \\ x_1 & = 120 \\ x_1 & = 110 \end{array} \quad \{6\}$$

The *third* equation (why ?) in {2} leads to $x_1 = 110$ and $x_5 = 0$. The variable x_1 will be the *entering* variable and x_5 the *leaving* variable:

[1] A, B, C identify the iteration, as summarized below.

[2] *Iterative:* involving repetition; relating to *iteration*. *Iterate* (from Latin *iterare*), to say or do again (and again). Not to be confused with *interactive*.

$$\boxed{\mathbf{C}} \qquad x_1 = 110 - x_5 \qquad \{7\}$$

Substituting for x_1 everywhere (except in its own constraint), we have

$$\begin{array}{lllllllll} [\max] z = & 0{,}56(110 - x_5) & +0{,}42x_2 & & & & & & \\ & (110 - x_5) & +2x_2 & +x_3 & & & = & 240 & \\ & 1{,}5(110 - x_5) & +x_2 & & +x_4 & & = & 180 & \{8\} \\ & x_1 & & & & +x_5 & = & 110 & \end{array}$$

$$\boxed{\mathbf{A}} \qquad \begin{array}{llllllll} [\max] z = & & 0{,}42x_2 & & & -0{,}56x_5 & & +61{,}6 \\ & & +2x_2 & +\{x_3\} & & -x_5 & = & 130 \\ & & x_2 & & +\{x_4\} & -1{,}5x_5 & = & 15 \\ & \{x_1\} & & & & +x_5 & = & 110 \end{array} \qquad \{9\}$$

which is of course equivalent to Eq. {2}.

We now have a **new** (equivalent) LP problem, **to be treated as the original was.** The process can continue *iteratively*.

$$\begin{bmatrix} x_1 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 110 \\ 130 \\ 15 \end{bmatrix} \qquad \{10\}$$

From Eq. {2} or Eq. {9}, respectively,

$$z = \begin{bmatrix} 0{,}56 & 0 & 0 \end{bmatrix} \begin{bmatrix} 110 \\ 130 \\ 15 \end{bmatrix} = 61{,}6 \qquad \{11\}$$

$$z = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 110 \\ 130 \\ 15 \end{bmatrix} + 61{,}6 = 61{,}6 \qquad \{12\}$$

Now, x_2 is the new entering variable. According to the constraints, it can be increased till:

$$\boxed{\mathbf{B}} \qquad \begin{array}{ll} 2x_2 & = 130 \\ x_2 & = 15 \\ 0x_2 & = 110 \end{array} \quad \rightarrow \quad \begin{array}{ll} x_2 & = 65 \\ x_2 & = 15 \\ x_2 & = \infty \end{array} \qquad \{13\}$$

$$\boxed{\mathbf{C}} \qquad x_2 = 15 - x_4 + 1{,}5x_5 \qquad \{14\}$$

Substituting for x_2 everywhere (except its own constraint), we have

$$\begin{array}{lllllll}
[\max] z = & & 0{,}42(15 - x_4 + 1{,}5x_5) & & -0{,}56x_5 & & +61{,}6 \\
 & & +2(15 - x_4 + 1{,}5x_5) & +x_3 & -x_5 & = & 130 \\
 & & x_2 & +x_4 & -1{,}5x_5 & = & 15 \\
 & x_1 & & & +x_5 & = & 110
\end{array} \quad \{15\}$$

A

$$\begin{array}{llllllll}
[\max] z = & & & & -0{,}42x_4 & +0{,}07x_5 & & +67{,}9 \\
 & & & \{x_3\} & -2x_4 & +2x_5 & = & 100 \\
 & & \{x_2\} & & +x_4 & -1{,}5x_5 & = & 15 \\
 & \{x_1\} & & & & +x_5 & = & 110
\end{array} \quad \{16\}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 110 \\ 15 \\ 100 \end{bmatrix} \quad \{17\}$$

Now, x_5 is the new entering variable. According to the constraints, it can be increased till:

B

$$\begin{array}{llcll}
2x_5 & = 100 & & x_5 & = 50 \\
-1{,}5x_5 & = 15 & \rightarrow & x_5 & = \ldots \\
x_5 & = 110 & & x_5 & = 110
\end{array} \quad \{18\}$$

C

$$x_5 = 50 - \frac{1}{2}x_3 + x_4 \quad \{19\}$$

Substituting for x_5 everywhere (except its own constraint), we have

$$\begin{array}{llllllll}
[\max] z = & & & & -0{,}42x_4 & +0{,}07\left(50 - \frac{1}{2}x_3 + x_4\right) & & +67{,}9 \\
 & & & x_3 & -x_4 & +x_5 & = & 50 \\
 & & x_2 & & +x_4 & -1{,}5\left(50 - \frac{1}{2}x_3 + x_4\right) & = & 15 \\
 & x_1 & & & & +\left(50 - \frac{1}{2}x_3 + x_4\right) & = & 110
\end{array} \quad \{20\}$$

A

$$\begin{array}{llllllll}
[\max] z = & & & -0{,}035x_3 & -0{,}35x_4 & & & +71{,}4 \\
 & & & x_3 & -x_4 & +\{x_5\} & = & 50 \\
 & & \{x_2\} & +0{,}75x_3 & -0{,}5x_4 & & = & 90 \\
 & \{x_1\} & & -0{,}5x_3 & +x_4 & & = & 60
\end{array}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_5 \end{bmatrix} = \begin{bmatrix} 60 \\ 50 \\ 50 \end{bmatrix} \quad \{21\}$$

Now, no variable produces an increase. So, this is a *maximum.*

In sum:

A In the system of equations, find the identity matrix (immediate solution).
B search for an *entering* variable (or finish)
C consequently, find a *leaving* variable (if wrongly chosen, negative values will appear).

References:

– ZIONTS, Stanley, 1974, "Linear and integer programming", Prentice-Hall, Englewood Cliffs, NJ (USA), p 5. (IST Library.) ISBN 0-13-536763-8.

– See others on the course webpage *(http://web.ist.utl.pt/mcasquilho)*.

❖

Understand Customer Behavior And Complaints

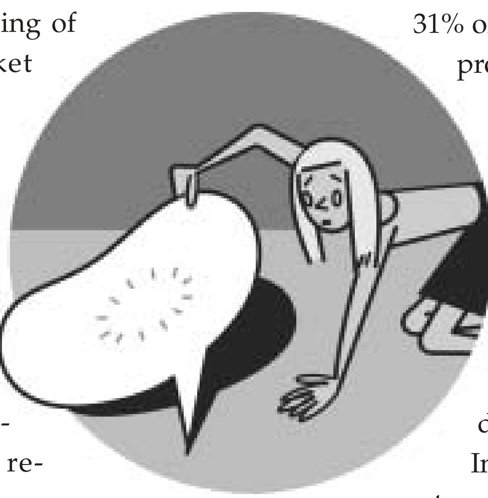
Eight areas of quantifiable data can be integrated into quality assurance decisions

by
John Goodman and Steve Newman

CUSTOMER COMPLAINTS PROVIDE valuable quality assurance, service and marketing data. But the challenge is to use the data to make decisions that result in substantive action.

To use complaint data to solve problems in design, marketing, installation, distribution and after sale use and maintenance, you should have a basic understanding of customer complaint and market behavior.

This understanding will provide a framework for interpreting the data and extrapolating it to the entire customer base. The framework will allow organizations not only to quantify the implications of the data but also to set priorities and allocate scarce quality assurance resources to mitigate problems.



In fact, unsolicited complaints submitted at the time a problem occurs are less costly than systematic sampling and inspection and provide more timely information than is typically available from warranty data.

Eight factors about customer behavior are key to understanding the implications of complaint data.

1. Dissatisfied individual and business customers tend not to complain.

Research by TARP[1,2] indicates most customers do not complain when they encounter a problem. In one case that could have resulted in an average loss of $142 to the customer, TARP found about 31% of individuals who encountered the problem did not complain.

We also found for small problems that resulted in either a loss of a few dollars or a minor inconvenience, only 3% of consumers complained and 30% returned the product. The balance of consumers encountering this problem either did nothing or discarded the product.

In a survey of 600 business software customers conducted by TARP,[3] results indicated 37% of the companies that encountered problems did not complain to anyone, even to the software support center. In several business to business studies, an average score of 25% of business customers made no contact with the vendor.

Finally, a 2001 TARP survey of purchasing agents for companies using electronic broadcast equipment found more than 50% who had encountered problems took immediate punitive action against a company without complaining to either the salesperson or sales manager. Companies indicated it was easier to switch vendors than complain.

2. Complaints often do not directly identify the source or cause of the problem.

The causes of customer dissatisfaction and questions can be grouped into three major categories: individual employee caused, company or retailer product or process caused and customer caused.

Our experience is that the distribution of problems across these three major cause categories is about 20%, 40% and 40%, respectively. By reviewing case closing information, analysts are in a position to differentiate among and identify key company and customer based causes.

It also should be noted there are several possible solutions to a particular problem. For example, an automobile company could either modify the normal operation of a vehicle or make customers aware at the outset that the vehicle will operate a certain way.

A major problem in the collection of customer problem data is a lack of differentiation between the reason for the complaint and the cause of the complaint. Customers usually discuss symptoms that are evident to them rather than the underlying cause.

An organization must classify customer contacts using either three or four categorization schemes:

1. Reason for contact (symptom).
2. General cause (employee error, company caused or customer caused).
3. Root cause (specific detail).
4. Reason for escalation of the complaint to a manager or headquarters unit (usually an exacerbating factor different from the original problem).

An example will illustrate the use of the four schemes.

> A consumer complains about a cancellation notice on his auto insurance policy. The company representative explains, "You failed to pay your premium." The consumer retorts, "I never got the premium notice." The representative says, "We sent it to 123 Main St." The consumer replies, "But I live at 127 Main St."

The reason for the call is a cancellation carried out in error. The general cause is a bad address. The root cause is the source of the bad address, which might be a keying error or illegible information on the application sent in by the agent.

If the company representative is not authorized to override the cancellation and the consumer goes to an executive or regulator, the reason for such an escalation would be lack of frontline authority.

Frontline representatives will almost always be able to identify the reason for a complaint call and the general cause. Root cause usually requires investigation unless the consumers indicate their own mistakes or abuse caused the problem (as is the case 30% to 40% of the time).

Unless these several types of data are collected in significant detail, the data cannot be analyzed to produce actionable results. We usually find at least 100 complaint reason for call categories are needed to provide sufficient detail.

Broad categories may appear to be easier to use and just as effective when, in fact, valuable detail is lost. Airlines formerly used the category "smoking complaint" that included "wanted to smoke but couldn't" as well as "being seated in a smoking rather than a nonsmoking section." Putting the detail in the verbatim text was not useful because text cannot easily be cross tabulated and analyzed by computers, and manual case analysis is not practical for large volumes of customer contacts.

3. Retail, field sales and service systems filter and discourage complaints.

Several recent TARP studies determined that for package goods (small ticket items sold in a supermarket, for example), only one person in 50 who encounters a problem writes a letter to the manufacturer and only two use the toll-free number.

Therefore, in a letter based environment, a package goods manufacturer at best hears only about one out of 50 problem experiences at the headquarters level unless the difficulty is severe (such as loss of a substantial amount of money, a threat to the consumer's good name or a life threatening result of use).

Our survey found fewer than half who complained at the retail level were ultimately satisfied. Furthermore, fewer than half who were dissatisfied bothered to escalate their complaint to the retailer's headquarters or to the manufacturer. The retailer or field service outlet may handle or mishandle the complaint but, in any case, may stop it from going further.

Thus, complaint data must be extrapolated to the customer base to determine the potential severity of the problem. The absolute number of articulated complaints in a particular area cannot be considered in isolation. A key factor is the potential extent to which the field or retail service systems have reduced the signal received by headquarters.

For example:

- After inadvertent production of a defective ladies' garment that cost $20 and tore during its first use, either the customer or the retailer returned only one in 2,000 of the defective garments.
- Fewer than half of the residential customers who experienced a billing problem with a telecommunications supplier articulated it to the company. Additionally, corporate clients have been found to complain to service technicians rather than account executives because of perceptions that marketing staff is powerless to solve technical problems.
- A business customer of a major computer company was told his staff was the cause of system failures. Company headquarters did not realize there was a problem until the dissatisfied consumer placed an ad in the *Wall Street Journal* and was joined by 300 other companies in the action.[4] Company regional sales representatives and management had decided the problem was customer incompetence and not product related, because each had heard only one or two complaints.
- The average customer who complained to the headquarters of a major credit card company had previously tried to use routine channels an average of six times.
- Both medical product manufacturers and insurance companies found sales representatives tended to forward complaints only when it would ingratiate them with an important customer, or when the product was of such low margin the sales staff would rather see it discontinued. (Complaints provide a good rationale for discontinuing a product.).

The ratio of complaints heard at headquarters to the instances of occurrence in the marketplace (whether articulated or not) is called the multiplier. Based on a review of more than 500 studies with individual companies, multipliers can be characterized as follows:

- A 6-1 ratio for serious problems, when there is no visible field or retail contact organization.
- A 2,000-1 ratio for less serious problems, when there is an extensive field service organization to receive and absorb problems.

This multiplier can be used to extrapolate to the marketplace.

4. Brand loyalty can be retained by merely getting customers to articulate their problems.

The primary interest of any organization is to maximize sales and market share in the most profitable way. Customer satisfaction, therefore, is a means to an end—it is the way to retain customers. Getting customers to articulate their problems provides an effective mechanism to increase satisfaction and brand loyalty.

Eight Facts About Customer Behavior

1. Dissatisfied individual and business customers tend not to complain.
2. Complaints often do not directly identify the source or cause of the problem.
3. Retail, field sales and service systems filter and discourage complaints.
4. Brand loyalty can be retained by merely getting customers to articulate their problems.
5. Increasing the ease of access to the provider can reduce the complaint ratio (multiplier).
6. The propensity to complain is directly proportional to the perceived severity of the problem and damage to the respondent.
7. Complainers tend to be the heaviest users of the product or service.
8. Problem experience, especially in the case of those consumers who remain unsatisfied after complaining, results in substantial amounts of negative word of mouth.

Original research executed by TARP projectable to the U.S. population shows the following for consumers who experienced a problem with a potential financial loss of less than $5:

- 37% of those who did not articulate the problem stated they would continue to buy the product.
- 46% of those who did complain but were not satisfied by the company remained brand loyal.
- There were several cases in which articulated complaints did not lead to increased loyalty; in fact, if a complaint handling system is poor, it will further alienate the customer, resulting in lower repurchase rates.
- 70% of those who articulated the problem and were satisfied remained brand loyal, and more than 95% of complainants who were satisfied quickly remained brand loyal.

For consumers who experienced a problem with a potential financial loss of more than $100, our surveys show the following:

- 9% of those who did not articulate the problem remained brand loyal.
- 19% of those who articulated the problem but were not satisfied remained brand loyal.
- 54% of those who articulated the problem and were satisfied remained brand loyal.

The research has since been confirmed in over 500 separate surveys of at least 700 customers from both business and consumer markets. Thus, brand loyalty can be retained by encouraging consumers to complain. Encouragement can include posting a number in a store or on an invoice. Employees can simply make eye contact and ask, "Is there anything else I can do for you?"

Even if the complaint handling mechanism is not able to satisfy the consumer, incremental brand loyalty can be achieved. Of course, if the complainant is satisfied, substantial amounts of brand loyalty can be obtained. In fact, loyalty can actually become up to 8% higher than loyalty when no problem has occurred.

5. Increasing the ease of access to the provider can reduce the complaint ratio (also known as the multiplier).

Research by TARP across both manufacturing and service industries shows consumers don't complain because of the following:

- It isn't worth the time and trouble.
- They don't know how or where to complain.
- They don't believe the company will do anything.
- They fear retribution in medical, financial, governmental and some auto environments.

By breaking down these perceived barriers to complaining, an organization can successfully increase the percentage of customers who articulate their problems. Barriers can be broken down by making it easy to complain via toll-free numbers or through "contact us" or feedback buttons on a website or invoice that are accompanied by a message that says, "We can only solve problems we know about."

The market implications of this type of aggressive complaint solicitation are shown by our research:

- In the telecommunications industry, seven of 10 respondents who encountered a problem and did not articulate it would have complained had the company maintained a toll-free number. Overall, this aggressive solicitation strategy would reduce unarticulated dissatisfaction by more than half.
- For a manufacturer of household products, the establishment of a toll-free telephone system for consumer contact led to a doubling of complaints to the manufacturer. Additionally, the mix of complaints was different. Many were received that would have otherwise been handled and filtered by the retailer.

By breaking down these perceived barriers to complaining, an organization can successfully increase the percentage of customers who articulate their problems.

6. The propensity to complain is directly proportional to the perceived severity of the problem and damage to the respondent.

Consumers tend not to complain about things they consider minor inconveniences. Think about your own experience as a consumer—how many times have you complained about a mediocre meal in a restaurant or slow service in a department store?

If, however, the problem will cause a major financial loss or damage to a consumer's reputation, the tendency to complain is much greater:

- Significantly more (70%) purchasers of high priced telecom equipment articulated their problems than did purchasers of low or moderately priced equipment. Still, 30% of those with inoperable equipment never complained but simply discarded the $100 item.
- Six out of 10 respondents who encountered a billing problem by a residential telecom service provider never complained. It was easier to pay the small amount in dispute than to voice the problem. That was due, at least in part, to the difficulty customers

encountered in dealing with the company.

- For major problems with an average loss of $142, 69% of the households complained, and half of those not satisfied complained a second time; for package goods, only one-third returned the item and only one in 50 wrote to headquarters.

7. Complainers tend to be the heaviest users of the product or service.

Consumers who are heavy users of a product or service are those who have made a commitment. Thus, in a sense, they have a vested interest in having the company improve its offerings.

These are the consumers who represent the potential for the most market damage if their loyalty is compromised. In fact, our research indicates the following:

- 40% of those who escalated their problems to the headquarters of a provider of credit card services charged more than $1,000 per month and represented a potential annual loss of profits of more than $500. This is in contrast to the average potential loss of $50 to $150 experienced by those who complained at the initial point of service for the same company.
- Complainants to a major Midwest bank, a medical products manufacturer and a car rental company on average had been loyal customers for longer periods and had purchased in heavier volumes than had an average customer.

8. Problem experience, especially in the case of those consumers who remain unsatisfied after complaining, results in substantial amounts of negative word of mouth.

Consumers typically tell others about their positive and negative experiences with a product or service. Positive communication can effectively serve to increase market share and revenue because those who hear it try the product or service.

Conversely, negative word of mouth can result in market damage and revenue loss. Additionally, dissatisfied complainants generate twice the negative word of mouth as do satisfied complainants generate positive word of mouth.

Some word of mouth research conducted by TARP showed the following:

- Satisfied Coca-Cola complainants told an average of four to five people about their positive experience, while dissatisfied complainants told an average of nine to 10 people about their negative experience.[5]
- In the automotive industry, one TARP study found an average of eight positive word of mouth communications resulted from each satisfied complainant and 16 negative word of mouth communications from each dissatisfied one.
- Word of mouth from unarticulated dissatisfaction can also result in market damage. In this instance also, a 2 to 1 ratio is seen.
- Consumers who experience a problem and don't articulate it to the provider tell twice as many people as satisfied consumers who do not experience a problem.
- For a residential telecom service provider, there were an average of 1.5 positive word of mouth communications from satisfied consumers and 3.7 negative word of mouth communications from consumers who experienced a problem and did not articulate it to the provider.

A Harvard study found that negative word of mouth had twice the market damage as positive word of mouth had a positive impact.[6]

All this information about consumer behavior provides a framework for integrating complaint data into quality assurance decisions, a topic we will discuss in the February 2003 issue of *Quality Progress*.

REFERENCES

1. TARP was originally founded as Technical Assistance Research Programs at Harvard University in 1971 to study customer service in the public sector.

2. TARP, *Consumer Complaint Handling in America: Final Report*, sponsored by the U.S. Office of Consumer Affairs, 1979.

3. TARP, "Basic Facts on Customer Complaint Behavior and the Impact of Service on the Bottom Line," *Competitive Advantage* (newsletter of the ASQ Service Quality Division), June 1999, pp. 1, 4.

4. *Wall Street Journal*, July 19, 1980.

5. TARP, *Measuring the Grapevine—Consumer Response and Word of Mouth*, Coca-Cola U.S.A., 1981.

6. Johan Arndt, "Role of Product Related Conversations in the Diffusions of a New Product," *Journal of Marketing Research*, August 1967, pp. 291-295.

JOHN A. GOODMAN *is president of TARP in Arlington, VA. He earned a master's degree in business administration from Harvard University. Goodman is a member of ASQ.*

STEVE NEWMAN *is director of research for TARP. He earned a bachelor's degrees in mathematics and political science from Vanderbilt University.*

IF YOU WOULD LIKE to comment on this article, please post your remarks on the *Quality Progress* Discussion Board at www.asqnet.org, or e-mail them to editor@asq.org.

(Blank page)

5S for Suppliers

How this technique can help you maintain a lean material supply chain

by **Kimball E. Bullington**

WHAT DOES SUPPLY management for lean production look like? In some companies it looks like lean production because these lean leaders use the 5S—*seiri*, *seiton*, *seiso*, *seiketsu* and *shitsuke*—technique to ensure the proper maintenance of a lean material supply.

The term "lean supply" implies that the supply chain is appropriate for lean production. This concept of waste elimination in processes has been popular at some manufacturing companies. Its basic tenets are to:[1]

- Specify value.
- Identify the value stream.
- Organize the value stream to promote flow.
- Communicate demand through pull.
- Strive for perfection.

What is 5S?

The 5S's are key lean concepts derived from the Japanese words *seiri* (sort), *seiton* (set in order), *seiso* (shine), *seiketsu* (standardize) and *shitsuke* (sustain).[2] Companies adopting the lean production philosophy often implement the 5S process to bring order to the workplace and support lean production. One important aspect of the lean production philosophy is its emphasis on value. In the United States, the cost of purchased materials accounts for approximately 35 cents of every sales dollar.[3] Any comprehensive effort to focus on the value a customer receives must include the supply perspective.

Why use 5S?

The 5S program is a proven model for organizing and maintaining a production operation. It is frequently used in manufacturing operations, particularly progressive ones.

The supply function, especially the purchasing department, often reports to the manager of operations, and that person is usually not a purchasing expert. The relationship between purchasing personnel and the general management of operations can be improved if they use a common vocabulary built around concepts familiar to the head of the organization and the heads of the other departments. The 5S program provides that common vocabulary.

The program also supports the visual workplace. Hiroyuki Hirano referred to the 5Ss as "pillars of the visual workplace" in the title of his book on 5S.[4] In a 5S environment there is a place for everything, and everything is in its place. Time spent searching for items is essentially eliminated, and out of place or missing items are immediately obvious in a properly functioning 5S facility.

The 5S program performs a similar function for supplier maintenance. The users of the system immediately know where to find information about their supply base, and missing or out-of-date information is instantly apparent. To implement a 5S program for supplier maintenance, you should abide by the following five guidelines.

1. Sort

The first step in implementing 5S for lean manufacturing is to take a tour of the target area and mark with red tags those items that appear out of place or unnecessary. After each item is reviewed, it is either put in its proper place or removed if it is unnecessary or redundant. The sort process is essential to organizing the workspace needed for lean production.

When applying the sort method to the supply base, you select suppliers to add to the system and eliminate from the system. In purchasing jargon, this is known as supply base consolidation or rationalization. Sorting the supplier base through consolidation:

- Reduces the waste of inefficient work methods by reducing the number of suppliers that must be managed by the procurement staff.

- Reduces the waste of supplier selection quality by focusing efforts of selection, evaluation and improvement on a few select suppliers. It also improves the quality (conformance to specifications and delivery) of the products received from these suppliers by focusing quality assurance, control and improvement activities on a smaller number of suppliers.
- Reduces processing waste as fewer purchase orders are necessary and fewer selection audits are needed.
- Increases the opportunity for supply chain partnering when suppliers are aware of their sole-source status.

So for the management of supply, the primary focus of the sort step is to select suppliers. It cannot be done by physically attaching red tags to suppliers, so how can it be applied? Several criteria can be used to identify candidates for elimination in the sorting process:

- Isolate some candidates for elimination by conducting a performance review—a review of their quality, delivery and price performance. A lean producer may be much more interested in delivery performance or inventory levels than the actual purchase price. But even a lean manufacturer's emphasis on delivery, price and inventory will shift as conditions dictate.
- Perform a review of redundant suppliers. Ask how many suppliers have identical or overlapping capabilities. The effort to consolidate suppliers this way rewards the best suppliers with additional business.
- Determine how many parts are purchased from each supplier, which will often result in identification of a large number of suppliers providing only one or a few parts. Some suppliers may not have been used for some time and may be candidates for sorting or consolidation.

After completing all the sorting, you will likely end up with an optimum number of suppliers. W. Edwards Deming said, "No manufacturer I know of possesses enough knowledge and manpower to work effectively with more than one vendor for any item."[5] Many companies choose to use more than one source for a single item for a variety of reasons, including risk, capacity and price, but Deming's point should be heeded. Multiplying suppliers increases variation and overhead. The practice of using multiple suppliers for a single part in order to reduce risk often increases risk, just as increasing the number of components in an assembly usually increases the probability of failure. Veteran supply managers often build their improvement efforts on an initial sorting of the supply base.

2. Set in order

In a manufacturing implementation of 5S, "set in order" means to arrange products and equipment so they are easy to find and use. Equipment and storage locations are labeled so the tools will be easy to identify when they are needed and put away when they are no longer in use. Labeling storage locations with tape on the floor or the workstation facilitates visual management. One glance is all that's needed to identify missing or improperly stored tools.

TABLE 1 Supplier Segmentation Value Matrix

	Low value potential	High value potential
High risk	Risk value	Partnership value
Low risk	Transaction value	Price value

Arranging suppliers so they are easy to use involves the concept of segmentation. To complete the initial segmentation, you need to perform a portfolio analysis. This analysis provides a place for everything and allocates everything in its place. It sorts the supply base by value potential and risk, strategic value and opportunity for cost improvement, and value potential and criticality.

The proper place for a supplier is identified in a segmentation matrix. To create one, you need to clearly determine how each supplier will be treated based on identifiable criteria. See Table 1 for a sample supply base segmentation by annual expenditures and risk. It yields four segments of suppliers with different opportunities for value contribution.

The key suppliers for lean production companies tend to fall into the high risk/high value partnership segment of the matrix. Partnership suppliers represent a higher risk to the company in terms of design complexity, start-up communication, custom tooling, overall higher demand for buyer input and schedule pressures. You can also think of risk as the level of opportunity for adverse effects on value, such as deterioration in delivery, lead time, price or quality.

The other supplier segments have different needs. The low risk/high value potential segment includes commodity items. The value contribution of these suppliers is primarily driven by price. A partnership is not the ideal model for these relationships.

Low risk/high value potential purchases are characterized by intense negotiations, competitive bidding, online auctions and long-term contracts. If the risk can be reduced for low risk/high value potential items, significant savings can be realized by some form of competitive bidding. For example, automobile manufacturers willing to redefine their tire requirements so they can purchase fewer types of tires could realize cost savings in the tens of millions of dollars.

The high risk/low value potential suppliers affect value by the nature of the factors that make them high risk, including demanding delivery requirements and advanced technology. The factors that drive risk are often the factors that contribute to the value of the product. Temporary situations, such as cash flow problems or capacity limitations, could be the major risk factors.

A minor supplier of plastic injection moldings, for example, lost a major account, which significantly changed its cash flow situation. Normally this supplier would have been considered moderately high risk because it required customized tooling, but for a time it moved near the top of the high risk category. In an undesirable situation such as this, the supplier has the potential to reduce the value, but little or no potential for adding value. Suppliers in this category are usually overlooked by supply management until some sort of failure occurs, but segmentation can help the supply manager identify these potential problem suppliers before something bad happens.

The low risk/low value potential suppliers typically have high transaction costs compared to the value of the product. The opportunity for adding value comes by consolidating these purchases and reducing transaction costs.

Several different segmentations may be conducted to properly categorize your suppliers and can include an evaluation of the supplier's quality. Performance measures may be helpful in segmenting the remaining supply base.

Deterioration in the segmentation step can arise when new suppliers are added. It can only be prevented by making classification a step in the process of adding a supplier. Another source of deterioration comes when a supplier's risk factor changes. Changes such as financial difficulties or capacity problems should be evaluated on an ongoing basis for key suppliers, but the problem can also arise at a relatively minor supplier.

To address the location aspect of the set in order step, identify the location of each supplier on a large map. It will help you further consolidate by forcing you to group suppliers locally, in targeted areas or along trucking routes so more than one supplier can be visited on a single trip.

3. Shine: Keep everything swept and clean

Cleaning implies system maintenance and inspection. As a work area is cleaned, problems such as oil leaks or other maintenance issues become apparent before they have a chance to affect performance.

To maintain and inspect suppliers, you can conduct surveys or audits. The audits can include site surveys, supplier self-assessments, remote surveys, third-party certification type surveys such as ISO 9000 or QS-9000, or third-party quality awards such as the Malcolm Baldrige National Quality Award.

When auditing suppliers, you should try to obtain objective evidence that supports your sort and segmentation decisions or evidence that supports action of a different type, such as risk reduction and continuous improvement. Audits enable the supply manager to detect problems early so they can be corrected before further deterioration occurs. Auditing provides an opportunity for reviewing supplier performance. It lets you ask, are the right measures being used? Is the supplier performing adequately? What are the reasons behind a particular supplier's less than desirable performance?

Third-party audits such as those conducted during the ISO 9000 certification process provide limited insight under these circumstances. First-person audits or customized surveys are much better, but you must remember audits are not a panacea. For example, a team audit of one new supplier revealed a low level of business activity, which should have been a warning of future new product introduction problems. Unfortunately, the warning signs were ignored.

You should schedule regular on-site visits with the key suppliers identified in the set in order or segmentation stage. High risk/high value potential suppliers usually receive the most visits, followed by high risk/low value potential suppliers and low risk/high value potential suppliers. Low risk/low value potential suppliers are generally not surveyed except through mail surveys of regulatory compliance issues.

Partnership maintenance is a key element in preserving your supply base. A sound preventive maintenance program is as essential to supplier maintenance as it is to manufacturing. The basic elements of a preventive maintenance program include:

- A schedule of the maintenance, such as a team meeting schedule with expectations defined for each meeting.
- A log of the maintenance, which could include meeting minutes.
- An audit of the maintenance, such as an audit of the team's processes and performance.

Partnership maintenance is not complicated, but it is necessary.

4. Standardize: Integrate sort, set in order and shine

This step should follow your successful implementation of the sort, set in order and shine steps. Standardizing or integrating all three steps ensures your implementation of the first three pillars won't deteriorate over time. It formalizes the procedures, schedules and practices that sustain the system and drive future improvements.

Several problems can be avoided through standardization, including:

- The number of suppliers growing unchecked.
- The number of suppliers becoming unknown, which can lead to supply base growth and segmentation deterioration.
- The segmentation deteriorating and the classification of the suppliers becoming unknown.
- Suppliers not being visited on a regular basis.
- The conducting of surveys informally or with renegade processes.

What's the easiest way to standardize? By assigning 3S duties. Make sure the personal plans or objectives of the supply management personnel cover the necessary sort (supplier consolidation), set in order (segmentation) and shine (audit) issues. For instance, some purchasing agents maintain bar charts showing the number of suppliers over time. They also keep a segmentation chart showing their key suppliers in the relevant segments and use different colors to indicate when the supplier was last audited.

Strategic buyers, commodity managers or the purchasing manager are responsible for surveying the charts in each buyer's area to guarantee they are current. The results of these surveys can be displayed on checklists that show the level of implementation.

The motivation for consolidating suppliers often comes from outside the purchasing department, but do these other departments understand why consolidation is valuable? One advantage of the 5S approach is that it forces purchasing and manufacturing to use a common language. This facilitates the communication between these two groups, but what about interactions between design engineering and purchasing?

This is a critical interface for two reasons. First, engineering is the source of many requests for new suppliers. Second, engineering, particularly design engineering, may harbor a creative environment that feels constrained by programs that promote rigid discipline. Engineers have been known to say they see no reason to limit their supplier selections just so the buyers can play more solitaire on their computers. Fortunately, the 5S program provides the reasoning behind the consolidation efforts.

5. Sustain: Discipline starts with the leadership

Do you care enough to be consistent with your message and vision? Are you communicating the program, including the reasons for your actions, outside the procurement department? Are you properly training new employees? Does the appropriate structure exist to support this program? Is supply base maintenance a significant part of the employee performance appraisal process?

These are issues for leadership to conquer. No 5S process for supply management will be effective without vigilant leadership. Lean producers have used this process effectively, and consistent leadership over time is necessary to prevent system deterioration.

ACKNOWLEDGMENT

Thanks to Kent Johnson, industrial engineer with Chromalox; Jerry Benefield, former CEO of Nissan Motors, USA; and Kevin Stone, director of tactical operations and planning for Comcast Communications, for their comments, suggestions and examples.

REFERENCES:

1. James P. Womack and Daniel T. Jones, *Lean Thinking*, Simon and Schuster, 1996.
2. Hiroyuki Hirano, *5S for Operators: 5 Pillars of the Visual Workplace*, Productivity Press, 1996.
3. *Annual Survey of Manufacturers*, U.S. Census Bureau, 2000.
4. Hiroyuki Hirano, *5 Pillars of the Visual Workplace*, Productivity Press, 1990.
5. W. Edwards Deming, *Out of the Crisis*, Massachusetts Institute of Technology, 1986.

KIMBALL BULLINGTON *is an associate professor of operations management at Middle Tennessee State University in Murfreesboro, TN, and a supply chain management consultant. He earned a doctorate in industrial engineering from Mississippi State University in Starkville, MS. Bullington is a Senior Member of ASQ.*

IF YOU WOULD LIKE to comment on this article, please post your remarks on the *Quality Progress* Discussion Board at www.asqnet.org, or e-mail them to editor@asq.org.

Scientific computing over the Web

CPQ Centro de Processos Químicos
Centre for Chemical Processes

CERENA

DEQ
DEPARTAMENTO
DE ENGENHARIA QUÍMICA
TÉCNICO LISBOA

***Miguel* CASQUILHO**

Assist. Prof. (retired)

CERENA "Centro de Recursos Naturais e Ambiente" (*Centre for Natural Resources and the Environment*) & CPQ

DEQ Department of Chemical Eng.ing

IST *Instituto Superior Técnico*

UL *University of Lisbon*

http://web.tecnico.ulisboa.pt/mcasquilho/

U LISBOA | UNIVERSIDADE DE LISBOA

CERENA's Seminars

Lisboa, IST, 19 March 2014

Scientific computing over the Web

ABSTRACT ▪ In our technological era, **scientific computing over the Web** (SCW) has been overlooked: by Academia, by Industry. Besides its value to both, it can provide an easy link between them.

SCW: the user (in a website) supplies his data, executes an available program, and gets his results. ***(No software installation.)***

Based on cases that I use in teaching (***via the Web***), the talk will follow these points:

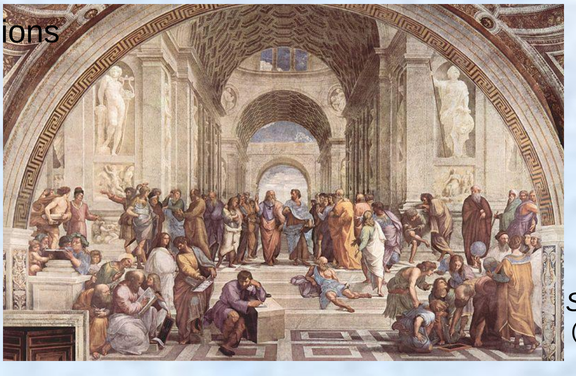
1. Antecedents;
2. Examples;
3. *Producer-consumer* communication (of SCW); and
4. Conclusions — and the Web offers a link between Academia and Industry.

- ***Cogito, ergo sum*.**

- ***In Interrete non existo, ergo non sum.***

Scientific computing over the Web

1. Antecedents
2. Examples
3. Producer-consumer communication
4. Conclusions



Raphael
(1483–1520)
Scuola di Atene
(School of Athens)

1. 2. 3. 4. Antecedents

- Scientific computing is laboriously taken to Industry…

- …But Industry often does not take advantage from SC:
  - Lack of *structure* — no “surplus” technical personnel for it
  - Lack of *stimulus* — no disposition; fear

We don’t use *those things [*] **that you taught us…** (***sic***)

[*] One of my former students, about Operational Research.

1. 2. 3. 4. Antecedents

- We talk to: **Industry**, but also other **academic institutions**.
- Yet:
- The Web[*] itself **has not paid attention** to **SC** (either with free or paid software).
- **Many** webpages and applications deal with **science**, but tend to the novelty of the medium (images, video, etc.).
- Among innumerable webpages, there are:
  - ***many*** (didactic) for consultation, study
  - ***few*** for dynamic use, *i.e.*: data $\rightarrow$ results
- Scarce ***applicable*** **material** is found on
  - Mathematics, Physics, Engineering, industrial applications
- ***What about solving problems (academic, industrial) ?***

[*] http://en.wikipedia.org/wiki/Internet#Services (about: Web $\in$ Internet)

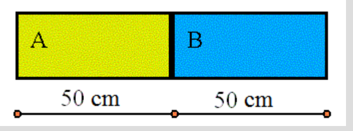
Software that exists in our organization

(A parenthesis)

- ***One of us* (technical staff, professor, etc.)**
  - ***can use*** the software (Matlab ⇨), **to which he can access remotely**,
  - to solve this simple problem (via Monte Carlo simulation):

What is the total length of these 2 (similar) bars, each with ~50 cm ? (Perhaps 1 m !)

Suppose *uniform* distributions in 49–51 cm.



- On the Web: sum of uniforms ⇨ ***You also can do it !***

1. 2. 3. 4. Antecedents

Webpages on **science** and…

- Wolfram *Mathworld*, by Eric Weisstein
  http://mathworld.wolfram.com/
- Springer ***E**ncyclopaedia **of** **M**athematics*
  http://eom.springer.de/
- Wikipedia (part)
  http://en.wikipedia.org/
- NIST/SEMATECH e-Handbook of Statistical Methods
  (Statistical Eng.ing Division ← **I**nformation **T**echnology **L**aboratory ← **N**at'l **I**nst. of **S**tandards and **T**echnology, USA)
  http://www.itl.nist.gov/div898/handbook/
- Etc., etc.

1. 2. 3. 4. Antecedents

…webpages on **scientific computing**

- Emphasis on images, interactivity
  - **“Atractor”** (.pt)   http://atractor.pt/   (but needs “our” Java !)
    (in WebMathematica) Example (polyhedra, etc.):

    http://atractor.pt/webM/exemplos/poliedros.htm
- **Academic SCW:** scarce ***working*** material on Mathematics, Physics, Industry. **Exception:**

  - **Prof. V. M. Ponce,** San Diego St. U. (in Fortran: remarkable, although without default data)

    http://ponce.sdsu.edu/onlinewaterbalance.php
- **“scientific computing on the internet”** ← ← ←
  - **~6** results (sources of software, university courses)

1. 2. 3. 4. Antecedents

FROM MY EXPERIENCE

- **Computer Science engineers** (informatics engineers), generally:
  - Are not enthusiastic about (this) Scientific Computing
  - Are not inclined to Mathematics

- The **other engineers**, generally:
  - Leave Information Technology (Informatics) to the former
  - Fear (justifiably…) the Internet environment

1. 2. 3. 4. Antecedents

- In my academic work:
  - *"Computing"* ◆ *"Operational Research"* ◆ *"Quality Control"*
  - I have adopted SC **over the Web**
  - since 1998
- 1.st example:
  - Area of a triangle by Heron's* formula (year AD 60)

$$A = \sqrt{s(s-a)(s-b)(s-c)}$$

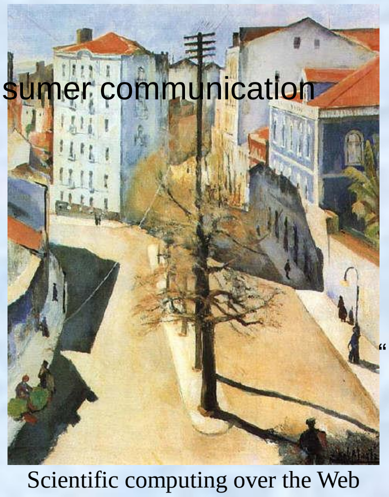
  - Program in Fortran (5 lines) (*a*, *b*, *c*, sides; *s*, semiperimeter)
  - About 1 year to put it on the Internet

http://web.tecnico.ulisboa.pt/mcasquilho/compute/misc/F-triang.html

* Heron of Alexandria (AD 10 ?–75 ?)

Scientific computing over the Web

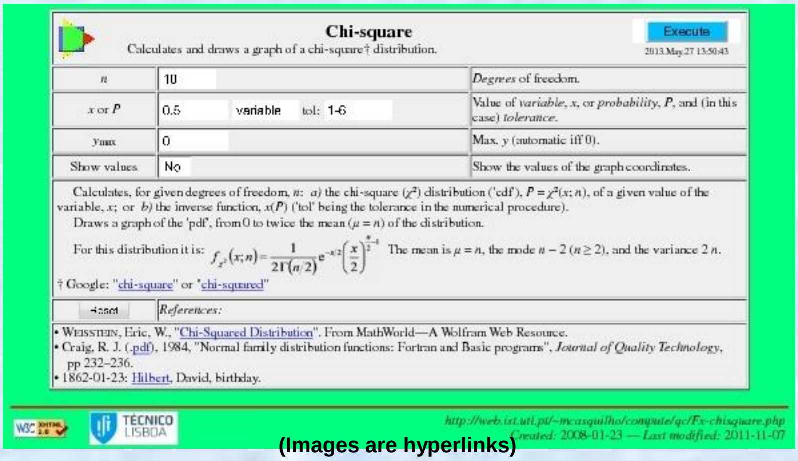
1. Antecedents
2. Examples
3. Producer-consumer communication
4. Conclusions



Abel Manta
(1888-1982)
"Rua de São Bernardo"
(*St. Bernard Str.*)

1. 2. 3. 4. Examples

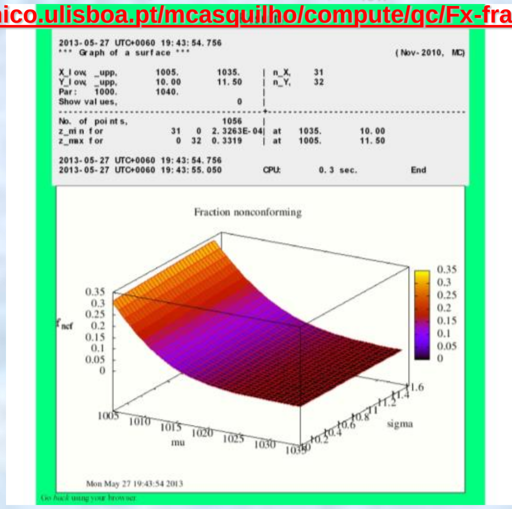
- **Chi-square** ➲ Try $\nu = 3$, $\nu = 50$ (nearly Gaussian)
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/qc/Fx-chisquare.php



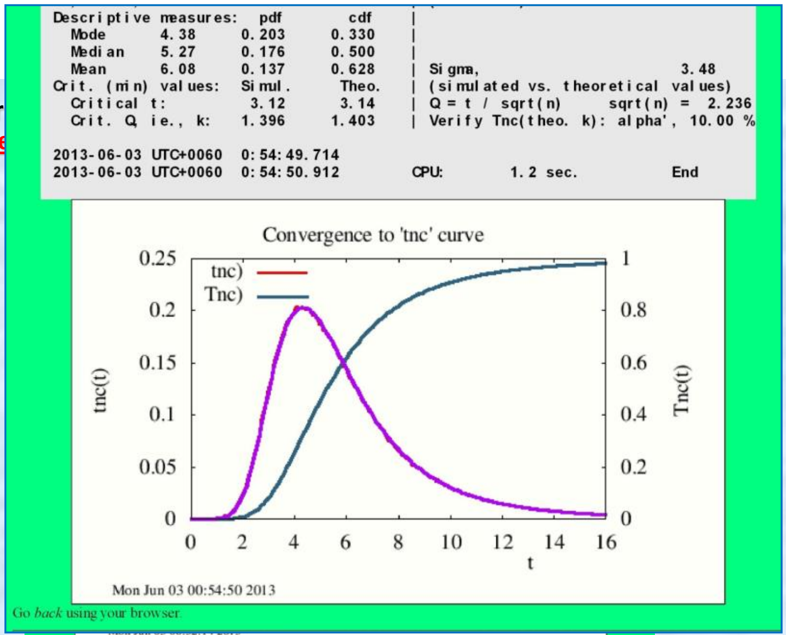
(Images are hyperlinks)

1. 2. 3. 4. Examples

- **Fraction defective** (Quality Control) ➲ Try σ_{low} = 10, 1.
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/qc/Fx-fracdefective.php

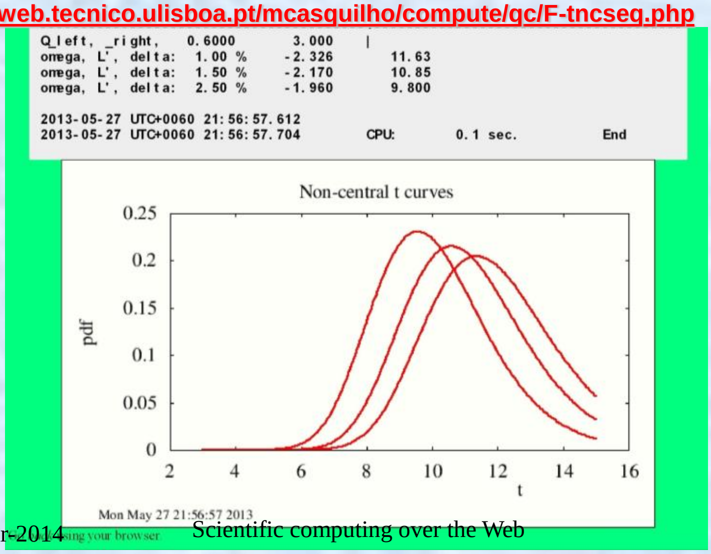


- **Conver**
  http://we



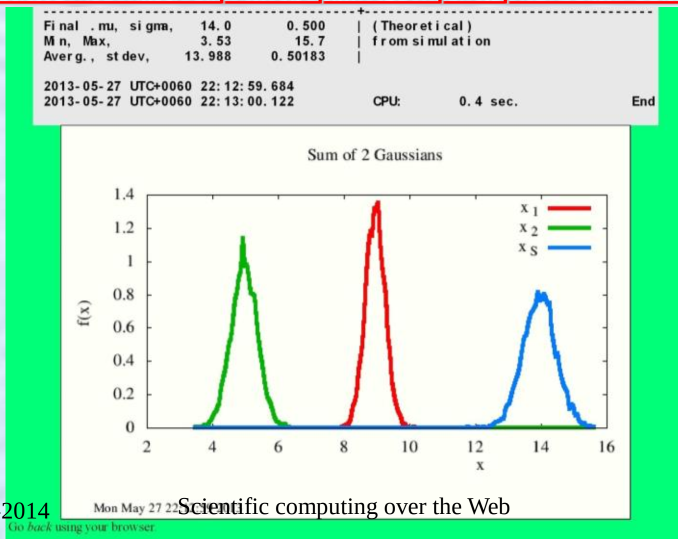
1. **2.** 3. 4. Examples

- **Sequence of non-central *t* curves** ➲ Try AQL = 1,1.5, 6.5 (%).
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/qc/F-tncseq.php



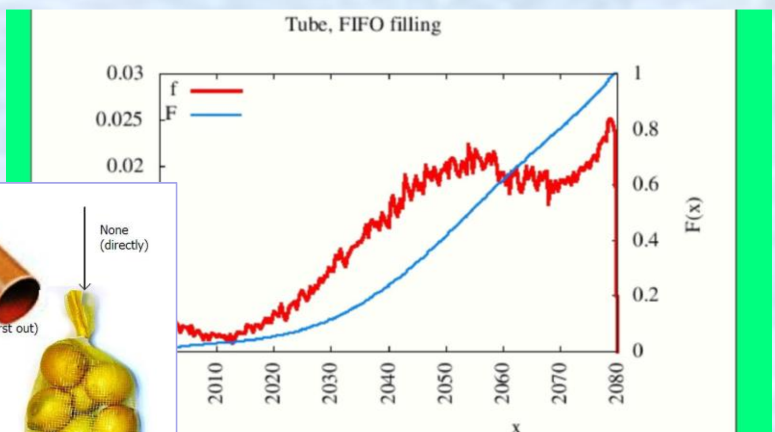
1. **2.** 3. 4. Examples

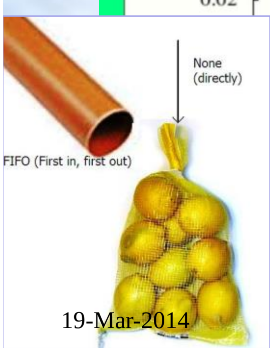
- **Sum of 2 Gaussians** ➲ Try (μ_1, σ_1) = (5, 0.1). $\left(5+9\overset{?}{=}14\right)$
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/qc/Fx-sum-2Gauss.php



1. **2.** 3. 4. Examples

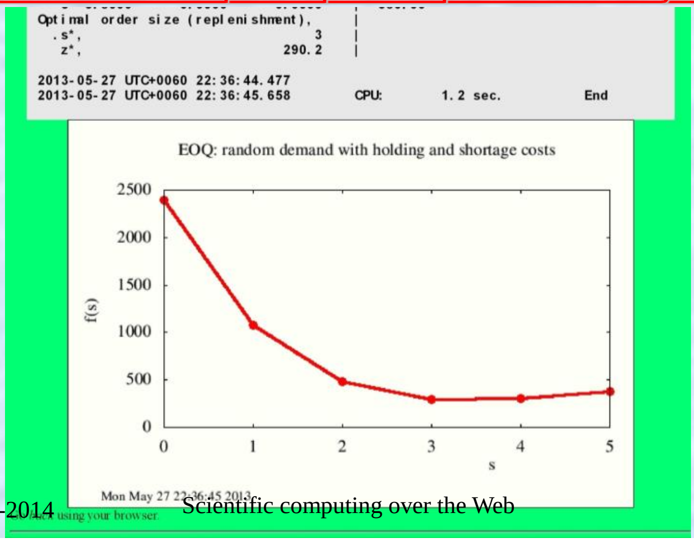
- **Packet filling, as tubes (FIFO)** ➲ Result FIFO. Try: none, FIFO.
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/qc/Fx-3tubefill.php





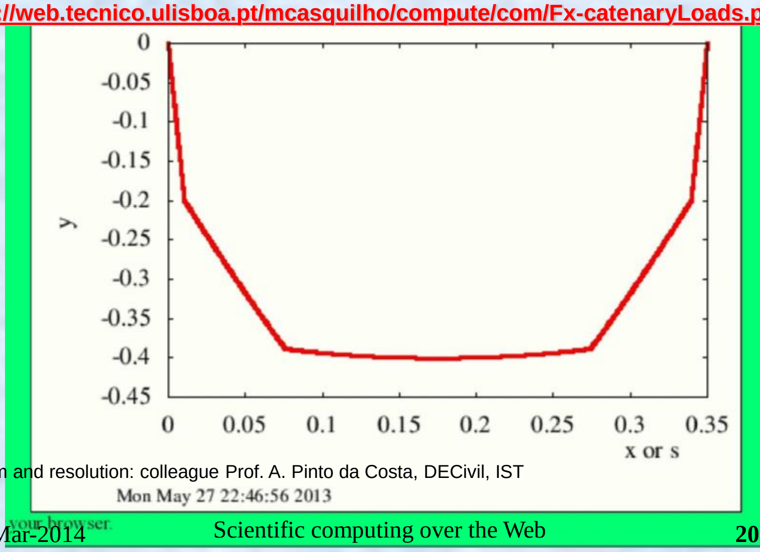
1. 2. 3. 4. Examples

- **EOQ (economic order q.) with random demand** ➲ Try c_p = 2000, 150.
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/or/Fx-inventoryRand.php



1. 2. 3. 4. Examples

- **Catenary with concentrated loads*** ➲ Try 3.rd load -1, -1.9.
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/com/Fx-catenaryLoads.php

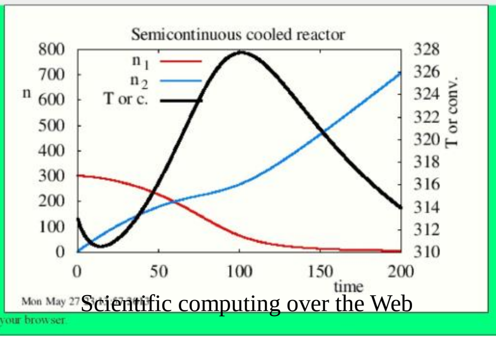


* Problem and resolution: colleague Prof. A. Pinto da Costa, DECivil, IST

- **Semicontinuous cooled reactor** ➲ Try T = 298, 348 K (25, 75 °C).
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/com/Fx-scoolreactor.php

Results

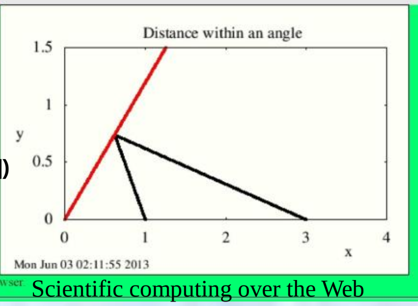
```
2013-05-27 UTC+0060 23:13:57.870
*** Semicontinuous cooled reactor ***                    (May-2005, MC)

U, A,                30.0      0.624  | kJ/m2-min-K, m2
.n_A0, .n_B0,        300.      0.00   |
T_0,                           313.   | K
Plot {1|2} ?                   1      | Temperature or conversion
Show values (0|1) ?            0      |
-----------------------------------------------------------
Final T,                       313.9  |
Final .n,            4.553     704.6  |
Conversion of A,               0.9848 |

2013-05-27 UTC+0060 23:13:57.870
2013-05-27 UTC+0060 23:13:58.238        CPU:   0.4 sec.      End
```



1. 2. 3. 4. Examples

- **Minimum distance "within an angle"** ➲ Try γ = 50°, 5°
  http://web.tecnico.ulisboa.pt/mcasquilho/compute/com/Fx-angdistImg.php

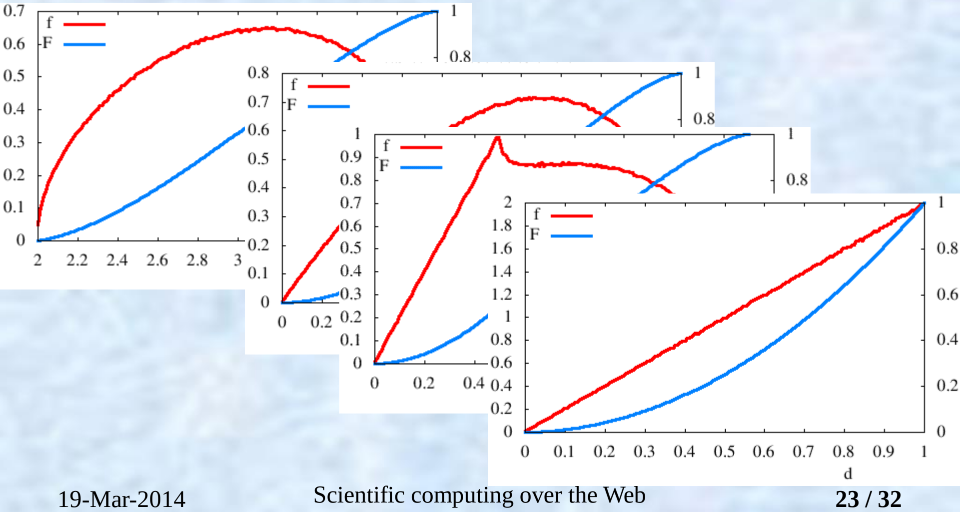
```
If x_2 = x_1:                          | (absolute minimum d0)
  x_P, y_P,      0.41318    0.49240    | d0 = x_1 cos^2 = x_P =  0.41318
If x2 = infinity:                      |
  x_P, y_P,      0.82635    0.98481    | x_P = 2 d0

MINIMUM DISTANCE for given x_2
x_P, y_P,        0.61976    0.73861    | Cartesian coordinates
rho_P,                      0.96418    | (Polar) radial coordinate
d_1, d_2,        0.8307     2.492      | A--P, P--B
Min. distance,              3.3229     | A--P--B

2013-06-03 UTC+0060  2:11:55.207
2013-06-03 UTC+0060  2:11:55.729        CPU:   0.5 sec.      End
```



(Casquilho & Buescu [2011])

1. 2. 3. 4. Examples

- **Distance to points in a circle** ➲ Try x_0 = 3, 1, 0.5, 0 (R = 1)
  http://web.tecnico.ulisboa.pt/~mcasquilho/compute/or/Fx-distInCircle.php



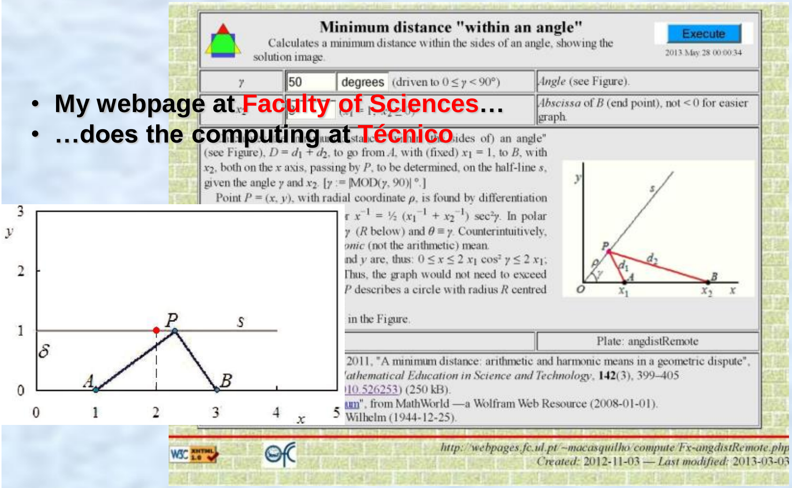
Scientific computing over the Web

1. Antecedents
2. Examples
3. Producer-consumer communicaation
4. Conclusions

Henrique Medina
(1901–1988)
"A rapariga da
galinha branca"
(*The girl with the white hen*)

1. 2. **3.** 4. Producer-consumer communication

- **Minimum distance “within an angle”** ➲ Compute: here at Técnico
  http://webpages.fc.ul.pt/~macasquilho/compute/Fx-angdistRemote.php



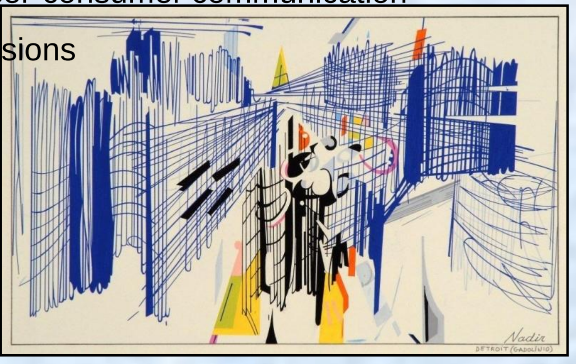
- **My webpage at Faculty of Sciences…**
- **…does the computing at Técnico**

1. 2. **3.** 4. Producer-consumer communication

- **For a company, the computing can be:**
  - Started on its website;
  - Executed on another site (University ?); and
  - Presented on the company’s website.
- **The user does not even notice the circuit.**

Scientific computing over the Web

1. Antecedents
2. Examples
3. Producer-consumer communication
4. Conclusions



Nadir Afonso
(1920–2013)
“Detroit”

1. 2. 3. **4.** Conclusions…

- **Scientific computing over the World Wide Web** has been overlooked.
- It can provide an easy link — namely, with Industry.
- The “start” can be difficult, perhaps justifying the scarcity.
- Working on the Web
  - avoids platform incompatibilities (Windows, Mac, Linux)
  - obviates limitations of the user’s terminal (PC, phone)
  - avoids software installation (uninstallation)
- Modern (scientific) languages (Mathematica, Matlab, etc.) favour productivity, but appear obsolete (!), as they aren’t (yet) Web-friendly (licences, permissions).

1. 2. 3. **4.** …Conclusions

- Scientific computing over the Web benefits from
  - sharing — in team work
  - visibility — attracting connections, *e.g.*, with Industry
  - network computing — “grid”, “cloud”, parallel
  - easy access, just using a *browser*

- Problems: welcome !

■

Acknowledgements

- **CERENA:** “Centro de Recursos Naturais e Ambiente” (*Centre for Natural Resources and the Environment*), IST, ULisboa
- **CPQ:** “Centro de Processos Químicos” (*Centre for Chemical Processes*), IST, ULisboa (CPQ ∈ CERENA)
- **DEQ:** Department of Chemical Engineering, IST, ULisboa
- **CIIST:** “Centre of Informatics of IST”, Instituto Superior Técnico, University of Lisbon
- **Milipeia** (Lab. of Advanced Computation), University of Coimbra
- Profs.(*DEngCivil*) A. Pinto da **Costa**, M. Azevedo **Coutinho** (⇦ Ponce), (*DMat*, FacC) Jorge **Buescu**, (*FacLetras*, UL) Arnaldo **Espírito Santo**

Keywords

- Scientific computing; Web; Internet; academia, industry

Bibliography

- **Casquilho,** M. [2012] “Computação científica na Internet, um campo pouco explorado”, Universidade Aberta, Porto Salvo.
- **Casquilho,** M., J. **Buescu** [2011] "A minimum distance: arithmetic and harmonic means in a geometric dispute", *International Journal of Mathematical Education in Science and Technology*, **142**(3), 399–405 (doi:10.1080/0020739X.2010.526253).
- **Ferreira,** M., M. **Casquilho** [2013] “Scientific computing over the Internet: an example in Geometry”, WorldCIST'13, World Conference on Information Systems and Technologies, 27 Mar., Olhão (Portugal)
- **Ferreira,** M., M. **Casquilho** [2013] “A web application for distributed computing — solving a typical statistical problem in Industry”, CISTI‘2013, 8.ª Conf. Ibérica de Sistemas e Tecnologias de Informação, 19 Jun., Lisboa (Portugal)
- **Franco,** B., M. **Casquilho** [2011] “A Web application for scientific computing: combining several tools and languages to solve a statistical problem”, CISTI'2011, 6.ª Conferência Ibérica de Sistemas e Tecnologias de Informação, Chaves (Portugal)
- **Hughes,** Alan [2006] “University-industry linkages and UK science and innovation policy”, Working Paper No. 326, Centre for Business Research, University of Cambridge, Cambridge (UK)
- **[“Internet” or “the Internet” ?]** in *http://microsites.oii.ox.ac.uk/ics2011/*. (2011-Sep)

Scientific computing over the Web

Centro de Processos Químicos
Centre for Chemical Processes

CERENA

DEQ
DEPARTAMENTO
DE ENGENHARIA QUÍMICA
TÉCNICO LISBOA

Miguel Casquilho

Assist. Prof. (retired)

CERENA “Centro de Recursos Naturais e Ambiente” (*Centre for Natural Resources and the Environment*) & CPQ

DEQ Department of Chemical Eng.ing

IST *Instituto Superior Técnico*

UL *University of Lisbon*

http://web.tecnico.ulisboa.pt/mcasquilho/

U LISBOA | UNIVERSIDADE DE LISBOA

CERENA’s Seminars

Lisboa, IST, 19 March 2014

Table 12.7

	Operating level	Cost per hour	Reservoir depth reduction per hour	Start-up cost
Hydro A	900 MW	£90	0·31 metres	£1500
Hydro B	1400 MW	£150	0·47 metres	£1200

generators: one of type A and one of type B. When a hydro generator is running, it operates at a fixed level and the depth of the reservoir decreases. The costs associated with each hydro generator are a fixed start-up cost and a running cost per hour. The characteristics of each type of generator are shown in Table 12.7.

For environmental reasons, the reservoir must be maintained at a depth of between 15 and 20 metres. Also, at midnight each night, the reservoir must be 16 metres deep. Thermal generators can be used to pump water into the reservoir. To increase the level of the reservoir by 1 metre requires 3000 MWh of electricity. You may assume that rainfall does not affect the reservoir level.

At any time it must be possible to meet an increase in demand for electricity of up to 15%. This can be achieved by any combination of the following: switching on a hydro generator (even if this would cause the reservoir depth to fall below 15 metres); using the output of a thermal generator which is used for pumping water into the reservoir; and increasing the operating level of a thermal generator to its maximum. Thermal generators cannot be switched on instantaneously to meet increased demand (although hydro generators can be).

Which generators should be working in which periods of the day, and how should the reservoir be maintained to minimize the total cost?

12.17 Three-dimensional Noughts and Crosses

Twenty-seven cells are arranged $3 \times 3 \times 3$ in a three-dimensional array as shown in Figure 12.5.

Three cells are regarded as lying in the same line if they are on the same horizontal or vertical line or the same diagonal. Diagonals exist on each

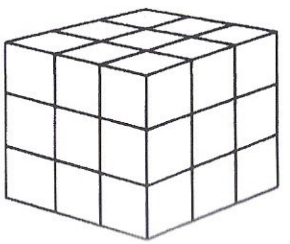


Figure 12.5

(Blank page)

horizontal and vertical section and connecting opposite vertices of the cube. (There are 49 lines altogether.)

Given 13 white balls (noughts) and 14 black balls (crosses), arrange them, one to a cell, so as to minimize the number of lines with balls all of one colour.

12.18 Optimizing a Constraint

In an integer programming problem the following constraint occurs:

$$9x_1 + 13x_2 - 14x_3 + 17x_4 + 13x_5 - 19x_6 + 23x_7 + 21x_8 \leqslant 37.$$

All the variables occurring in this constraint are 0–1 variables, i.e. they can only take the value of 0 or 1.

Find the 'simplest' version of this constraint. The objective is to find another constraint involving these variables which is logically equivalent to the original constraint but which has the smallest possible absolute value of the right-hand side (with all coefficients of similar signs to the original coefficients).

If the objective were to find an equivalent constraint where the sum of the absolute values of the coefficients (apart from the right-hand side coefficient) were a minimum what would be the result?

12.19 Distribution 1

A company has two factories, one at Liverpool and one at Brighton. In addition it has four depots with storage facilities at Newcastle, Birmingham, London and Exeter. The company sells its product to six customers C1, C2, . . . , C6. Customers can be supplied either from a depot or from the factory direct (see Figure 12.6).

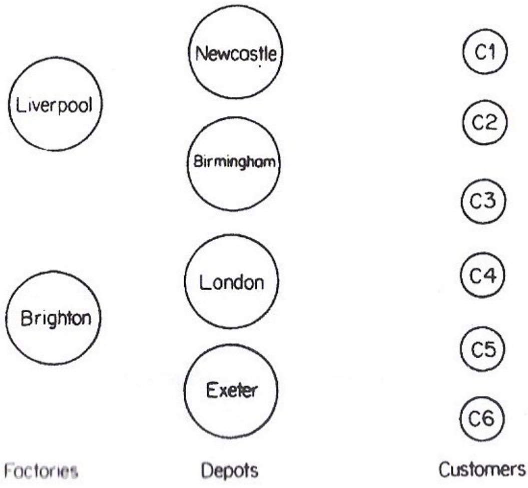


Figure 12.6

Table 12.7[a]

Supplied to	Supplier					
	Liverpool factory	Brighton factory	Newcastle depot	Birmingham depot	London depot	Exeter depot
Depots						
Newcastle	0·5	—				
Birmingham	0·5	0·3				
London	1·0	0·5				
Exeter	0·2	0·2				
Customers						
C1	1·0	2·0	—	1·0	—	—
C2	—	—	1·5	0·5	1·5	—
C3	1·5	—	0·5	0·5	2·0	0·2
C4	2·0	—	1·5	1·0	—	1·5
C5	—	—	—	0·5	0·5	0·5
C6	1·0	—	1·0	—	1·5	1·5

[a] A dash indicates the impossibility of certain suppliers for certain depots or customers.

The distribution costs (which are borne by the company) are known; they are given in Table 12.7 (in £ per ton delivered).

Certain customers have expressed preferences for being supplied from factories or depots which they are used to. The preferred suppliers are

C1	Liverpool (factory)
C2	Newcastle (depot)
C3	No preferences
C4	No preferences
C5	Birmingham (depot)
C6	Exeter or London (depots)

Each factory has a monthly capacity given below which cannot be exceeded:

Liverpool	150 000 tons
Brighton	200 000 tons

Each depot has a maximum monthly throughput given below which cannot be exceeded:

Newcastle	70 000 tons
Birmingham	50 000 tons
London	100 000 tons
Exeter	40 000 tons

Each customer has a monthly requirement given below which must be met:

C1	50 000 tons
C2	10 000 tons
C3	40 000 tons
C4	35 000 tons
CS	60 000 tons
C6	20 000 tons

The company would like to determine:

(1) What distribution pattern would minimize overall cost?
(2) What the effect of increasing factory and depot capacities would be on distribution costs?
(3) What the effects of small changes in costs, capacities and requirements would be on the distribution pattern?
(4) Would it be possible to meet all customers preferences regarding suppliers and if so what would the extra cost of doing this be?

12.20 Depot Location (Distribution 2)

In the distribution problem there is a possibility of opening new depots at Bristol and Northampton as well as of enlarging the Birmingham depot.

It is not considered desirable to have more than four depots and if necessary Newcastle or Exeter (or both) can be closed down.

The monthly costs (in interest charges) of the possible new depots and expansion at Birmingham are given in Table 12.8 together with the potential monthly throughputs.

The monthly savings of closing down the Newcastle and Exeter depots are given in Table 12.9.

Table 12.8

	Cost (£1000)	Throughput (1000 tons)
Bristol	12	30
Northampton	4	25
Birmingham (expansion)	3	20

Table 12.9

	Saving (£1000)
Newcastle	10
Exeter	5

Table 12.10

Supplied to	Supplier: Liverpool factory	Brighton factory	Bristol depot	Northampton depot
New depots				
Bristol	0·6	0·4		
Northampton	0·4	0·3		
Customers				
C1	As given for distribution problem		1·2	—
C2			0·6	0·4
C3			0·5	—
C4			—	0·5
C5			0·3	0·6
C6			0·8	0·9

The distribution costs involving the new depots are given in Table 12.10 (in £ per ton delivered).

Which new depots should be built? Should Birmingham be expanded? Should Exeter or Newcastle be closed down? What would be the best resultant distribution pattern to minimize overall costs?

12.21 Agricultural Pricing

The government of a country wants to decide what prices should be charged for its dairy products, milk, butter and cheese. All these products arise directly or indirectly from the country's raw milk production. This raw milk is usefully divided into the two components of fat and dry matter. After subtracting the quantities of fat and dry matter which are used for making products for export or consumption on the farms there is a total yearly availability of 600 000 tons of fat and 750 000 tons of dry matter. This is all available for producing milk, butter and two kinds of cheese for domestic consumption.

The percentage compositions of the products are given in Table 12.11.

For the previous year the domestic consumption and prices for the products are given in Table 12.12.

Table 12.11

	Fat	Dry matter	Water
Milk	4	9	87
Butter	80	2	18
Cheese 1	35	30	35
Cheese 2	25	40	35

Table 12.12

	Milk	Butter	Cheese 1	Cheese 2
Domestic consumption (1000 tons)	4820	320	210	70
Price (£/ton)	297	720	1050	815

Price elasticities of demand, relating consumer demand to the prices of each product, have been calculated on the basis of past statistics. The price elasticity E of a product is defined by

$$E = \frac{\text{percentage decrease in demand}}{\text{percentage increase in price}}.$$

For the two makes of cheese there will be some degree of substitution in consumer demand depending on relative prices. This is measured by cross-elasticity of demand with respect to price. The cross-elasticity E_{AB} from a product A to a product B is defined by

$$E_{\mathrm{AB}} = \frac{\text{percentage increase in demand for A}}{\text{percentage increase in price of B}}.$$

The elasticities and cross-elasticities are given in Table 12.13.

The objective is to determine what prices and resultant demand will maximize total revenue.

It is, however, politically unacceptable to allow a certain price index to rise. As a result of the way this index is calculated this limitation simply demands that the new prices must be such that the total cost of last year's consumption would not be increased. A particularly important additional requirement is to quantify the economic cost of this political limitation.

Table 12.13

Milk	Butter	Cheese 1	Cheese 2	Cheese 1 to Cheese 2	Cheese 2 to Cheese 1
0·4	2·7	1·1	0·4	0·1	0·4

12.22 Efficiency Analysis

A car manufacturer wants to evaluate the efficiencies of different garages who have received a franchise to sell its cars. The method to be used is Data Envelopment Analysis (DEA). References to this technique are given in Section 3.2. Each garage has a certain number of measurable 'inputs'. These are taken to be: *Staff*, *Showroom Space*, *Catchment Population* in different coefficients does not exceed the right-hand side coefficient. Such a subset is maximal in the sense that no subset properly containing it, or to the left in the implied lexicographical ordering, can also be a ceiling. For example the subset {1, 2, 4, 8} is a ceiling, $23 + 21 + 17 + 9 \leqslant 70$, but any subset properly containing it (e.g. {1, 2, 4, 7, 8}) or to the 'left' of it (e.g. {1, 2, 4, 7}) is not a ceiling. 'Roofs' are 'minimal' subsets of the indices for which the sum of the corresponding coefficients exceeds the right-hand side coefficient. Such a subset is 'minimal' in the same sense as a subset is 'maximal'. For example {2, 3, 4, 5} is a roof, $21 + 19 + 17 + 14 > 70$, but any subset properly contained in it (e.g. {3, 4, 5}) or to the 'right' of it (e.g. {2, 3, 4, 6}) is not a roof.

If $\{i_1, i_2, \ldots, i_r\}$ is a 'ceiling' the following condition among the new coefficients a_i is implied:

$$a_{i1} + a_{i2} + \cdots a_{ir} \leqslant a_0$$

If $\{i_1, i_2, \ldots, i_r\}$ is a 'roof' the following condition among the new coefficients a_i is implied:

$$a_{i1} + a_{i2} + \cdots a_{ir} \geqslant a_0 + 1$$

It is also necessary to guarantee the ordering of the coefficients. This can be done by the series of constraints:

$$a_1 \geqslant a_2 \geqslant a_3 \geqslant \cdots \geqslant a_8$$

If these constraints are given together with each constraint corresponding to a roof or ceiling then this is a sufficient set of conditions to guarantee that the new 0–1 constraint has exactly the same set of feasible 0–1 solutions as the original 0–1 constraint.

In order to pursue the first objective we minimize $a_0 - a_3 - a_5$ subject to these constraints.

For the second objective we minimize $\sum_{i=1}^{8} a_i$.

For this example the set of ceilings is

$$\{1,2,3\},\ \{1,2,4,8\},\ \{1,2,6,7\},\ \{1,3,5,6\},\ \{2,3,4,6\},\ \{2,5,6,7,8\}$$

The set of roofs is

$$\{1,2,3,8\},\ \{1,2,5,7\},\ \{1,3,4,7\},\ \{1,5,6,7,8\},\ \{2,3,4,5\},\ \{3,4,6,7,8\}$$

The resultant model has 19 constraints and nine variables.

If the constraint were to involve *general integer* rather than 0–1 variables, then we could still formulate the simplification problem in a similar manner after first converting the constraint to one involving 0–1 variables in the way described in Section 10.1. It is, however, necessary to ensure, by extra constraints in our LP model, the correct relationship between the coefficients in the simplified 0–1 form. How this may be done is described in Section 10.2.

13.19 Distribution 1

This problem can be regarded as one of finding the minimum cost flow through a network. Such network flow problems have been extensively treated in the mathematical programming literature. A standard reference is Ford and Fulkerson (1962). Specialized algorithms exist for solving such problems and are described in Ford and Fulkerson (1962), Jensen and Barnes (1980), Glover and Klingman (1977), and Bradley (1975).

It is, however, always possible to formulate such problems as ordinary linear programming models. Such models have the total unimodularity property described in Section 10.1. This property guarantees that the optimal solution to the LP problem will be integer as long as the right-hand side coefficients are integer.

We choose to formulate this problem as an ordinary LP model in order that we may use the standard revised simplex algorithm. There would be virtue in using a specialized algorithm. The special features of this sort of problem which make the use of a specialized algorithm worthwhile also, fortunately, make the problem fairly easy to solve as an ordinary LP problem. Sometimes, however, when formulated in this way the resultant model is very large. The use of a specialized algorithm then also becomes desirable as it results in a compact representation of the problem. As the example presented is very small, such considerations do not arise here.

The factories, depots and customers will be numbered as below:

Factories	1	Liverpool
	2	Brighton
Depots	1	Newcastle
	2	Birmingham
	3	London
	4	Exeter
Customers	C1	to C6

Variables

$$x_{ij} = \text{quantity sent from factory } i \text{ to depot } j,$$

$$i = 1, 2, \quad j = 1, 2, 3, 4;$$

$$y_{ik} = \text{quantity sent from factory } i \text{ to customer } k,$$

$$i = 1, 2, \quad k = 1, 2, \ldots, 6;$$

$$z_{jk} = \text{quantity sent from depot } j \text{ to customer } k,$$

$$j = 1, 2, 3, 4, \quad k = 1, 2, \ldots, 6.$$

There are 44 such variables.

Constraints

Factory Capacities

$$\sum_{j=1}^{2} x_{ij} + \sum_{k=1}^{6} y_{ik} \leqslant \text{capacity}, \qquad i = 1, 2.$$

Quantity into Depots

$$\sum_{i=1}^{2} x_{ij} \leqslant \text{capacity}, \qquad j = 1, 2, 3, 4.$$

Quantity out of Depots

$$\sum_{k=1}^{6} z_{jk} = \sum_{i=1}^{2} x_{ij} \qquad j = 1, 2, 3, 4$$

Customer Requirements

$$\sum_{i=1}^{2} y_{ik} + \sum_{j=1}^{4} z_{jk} = \text{requirement}, \qquad k = 1, 2, \ldots, 6.$$

The capacity, quantity and requirement figures are given with the statement of the problem in Part 2.

There are 16 such constraints.

Objectives

The first objective is to minimize cost. This is given by

$$\sum_{\substack{i=1\\j=1}}^{\substack{i=2\\j=4}} c_{ij}x_{ij} + \sum_{\substack{i=1\\k=1}}^{\substack{i=2\\k=6}} d_{ik}y_{ik} + \sum_{\substack{j=1\\k=1}}^{\substack{j=2\\k=6}} e_{jk}z_{jk},$$

where the coefficients c_{ij}, d_{ik}, and e_{jk} are given with the problem in Part 2.

The second objective will take the same form as that above, but this time the c_{ij}, d_{ik}, and e_{jk} will be defined as below:

$$d_{ik} = \begin{cases} 0 & \text{if customer } k \text{ prefers factory } i, \\ 1 & \text{otherwise} \end{cases}$$

$$e_{ik} = \begin{cases} 0 & \text{if customer } k \text{ prefers depot } j, \\ 1 & \text{otherwise} \end{cases}$$

$$c_{ij} = 0 \qquad \text{for all } i, j.$$

This objective is to be minimized.

13.20 Depot Location (Distribution 2)

The linear programming formulation of the distribution problem can be extended to a *mixed integer* model to deal with the extra decisions of whether to build or close down depots. Extra 0–1 integer variables are introduced with the following interpretations:

$$\delta_1 = \begin{cases} 1 & \text{if the Newcastle depot is retained,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\delta_2 = \begin{cases} 1 & \text{if the Birmingham depot is expanded,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\delta_4 = \begin{cases} 1 & \text{if the Exeter depot is retained,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\delta_5 = \begin{cases} 1 & \text{if a depot is built at Bristol,} \\ 0 & \text{otherwise;} \end{cases}$$

$$\delta_6 = \begin{cases} 1 & \text{if a depot is built at Northampton,} \\ 0 & \text{otherwise.} \end{cases}$$

In addition extra continuous variables x_{i5}, x_{i6}, z_{5k}, and z_{6k} are introduced to represent quantities sent to and from the new depots.

The following constraints are added to the model.

If a depot is closed down or not built then nothing can be supplied to it or from it:

$$\sum_{i=1}^{2} x_{ij} \leqslant T_j \delta_j,$$

where T_j is the capacity of depot j.

From Birmingham the quantity supplied to and from the depot must lie within the extension:

$$\sum_{i=1}^{2} x_{12} \leqslant 50 + 20\delta_2.$$

There can be no more than four depots (including Birmingham and London):

$$\delta_1 + \delta_4 + \delta_5 + \delta_6 \leqslant 2.$$

In the objective function the new x_{ij} and z_{jk} variables are given their appropriate costs. The additional expression involving the δ_j variables is added to the objective function:

$$10\delta_1 + 3\delta_2 + 5\delta_4 + 12\delta_5 + 4\delta_6 - 15.$$

This model has 21 constraints and 65 variables (five are integer and 0–1).

Pumping should take place in the following periods at the given levels:

Period 1	815 MW
Period 3	950 MW
Period 5	350 MW

Although it may seem paradoxical to both pump and run Hydro B in period 5, this is necessary to meet the requirement of the reservoir being at 16 metres at the beginning of period 1, given that the Hydro can work only at a fixed level. It would be possible to use the model to cost this environmental requirement.

The height of the reservoir at the beginning of each period should be:

Period 1	12 metres
Period 2	17.63 metres
Period 3	17.63 metres
Period 4	19.53 metres
Period 5	18.12 metres

The cost of these operations is £986 630.

In solving this model it is valuable to exploit the fact that the optimal objective value must be less than or equal to that reported in Section 14.15. This is for two reasons. Firstly the optimal solution in Section 14.15 using only thermal generators is a feasible solution to this model. Secondly the 15% extra output guarantee can be met at no start-up cost using the hydro generators. When solving this model by the branch and bound method the optimal objective value in Section 14.15 could be used as an 'objective cut-off' to prune the tree search.

Planning the use of hydro power by means of Stochastic Programming in order to model uncertainty is described by Archibald, Buchanan, McKinnon and Thomas (1999).

14.17 Three-dimensional Noughts and Crosses

The minimum number of lines of the same colour is four. There are many alternative solutions, one of which is given in Figure 14.5, where the top, middle and bottom sections of the cube are given. Cells with black balls are shaded.

This solution was obtained in 15 nodes. A total of 1367 nodes were needed to prove optimality.

14.18 Optimizing a Constraint

The 'simplest' version of this constraint (with minimum right-hand side coefficient) is

$$6x_1 + 9x_2 - 10x_3 + 12x_4 + 9x_5 - 13x_6 + 16x_7 + 14x_8 \leqslant 25.$$

This is also the equivalent constraint with the minimum sum of absolute values of the coefficients.

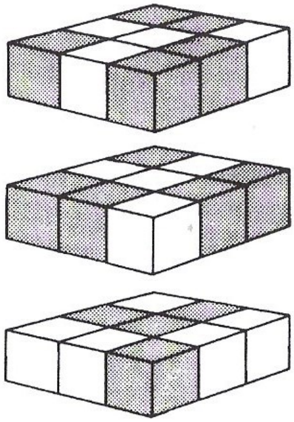


Figure 14.5

14.19 Distribution 1

The minimum cost distribution pattern is shown in Figure 14.6 (with quantities in thousands of tons).

There is an alternative optimal solution in which the 40 000 tons from Brighton to Exeter come from Liverpool instead.

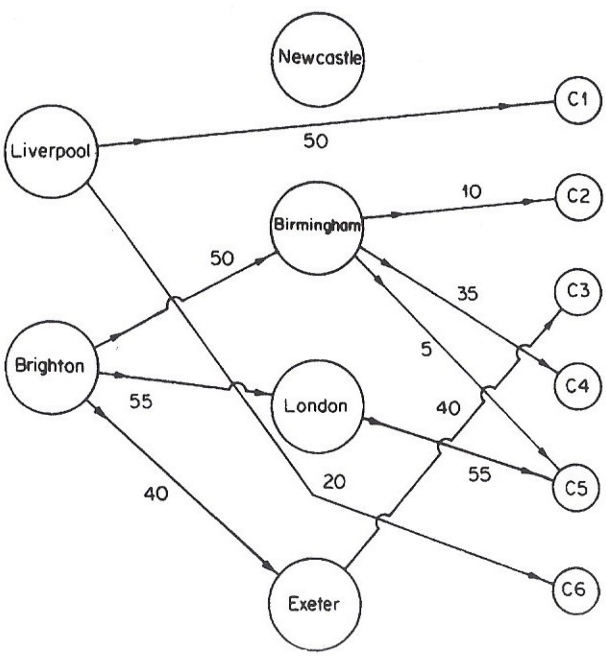


Figure 14.6

This distribution pattern costs £198 500 per month.

Depot capacity is exhausted at Birmingham and Exeter. The value (in reducing distribution costs) of an extra ton per month capacity in these depots is £0·20 and £0·30 respectively.

This distribution pattern will remain the same as long as the unit distribution costs remain within certain ranges. These are given below (for routes which are to be used):

Route	Cost range
Liverpool to C1	$-\infty$ to 1·5
Liverpool to C6	$-\infty$ to 1·2
Brighton to Birmingham	$-\infty$ to 0·5
Brighton to London	0·3 to 0·8
Brighton to Exeter	$-\infty$ to 0·2
Birmingham to C2	$-\infty$ to 1·2
Birmingham to C4	$-\infty$ to 1·2
Birmingham to C5	0·3 to 0·7
London to C5	0·3 to 0·8
Exeter to C3	0 to 0·5

Depot capacities can be altered within certain limits. For the not fully utilized depots of Newcastle and London changing capacity within these limits has no effect on the optimal distribution pattern. For Birmingham and Exeter the effect on total cost will be £0·2 and £0·3 per ton per month within the limits. Outside certain limits the prediction of the effect requires resolving the problem. The limits are:

Depot	Capacity range
Birmingham	45 000 to 105 000 tons
Exeter	40 000 to 95 000 tons

N.B. All the above effects of changes are only valid if *one* thing is changed at a time within the permitted ranges. Clearly the above solution does not satisfy the customer preferences for suppliers.

By minimizing the second objective it is possible to reduce the number of goods sent by non-preferred suppliers to a customer to a minimum. This was done and revealed that it is impossible to satisfy all preferences. The best that could be done resulted in the distribution pattern shown in Figure 14.7, where customer C5 receives 10 000 tons from his non-preferred depot of London. This is the minimum cost such distribution pattern. (There are alternative patterns which also minimize the number of non-preferences but which cost more.) The minimum cost here is £246 000, showing that the extra cost of satisfying more customers preferences is £47 500.

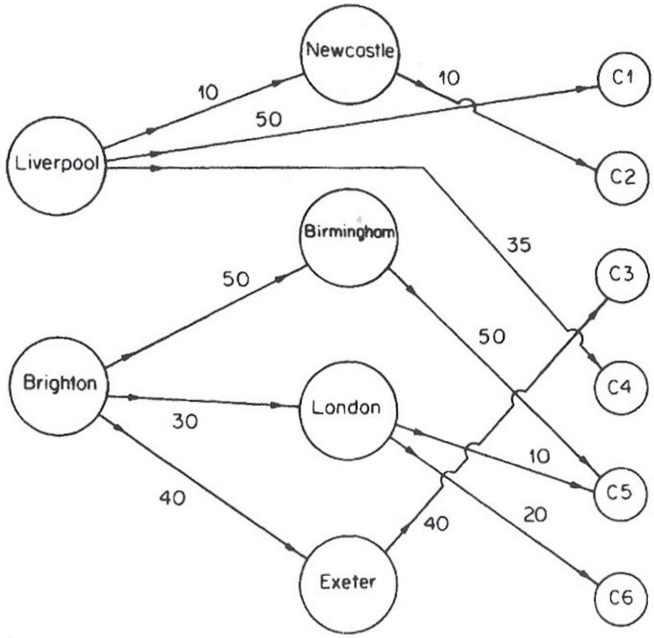


Figure 14.7

14.20 Depot Location (Distribution 2)

The minimum cost solution is to close down the Newcastle depot and open a depot in Northampton. The Birmingham depot should be expanded. The total monthly cost (taking account of the saving from closing down Newcastle) resulting from these changes and the new distribution pattern is £174 000. Figure 14.8 shows the new distribution pattern (with quantities in thousands of tons).

This solution was obtained in 40 iterations. The continuous optimal solution was integer. Therefore no tree search was necessary.

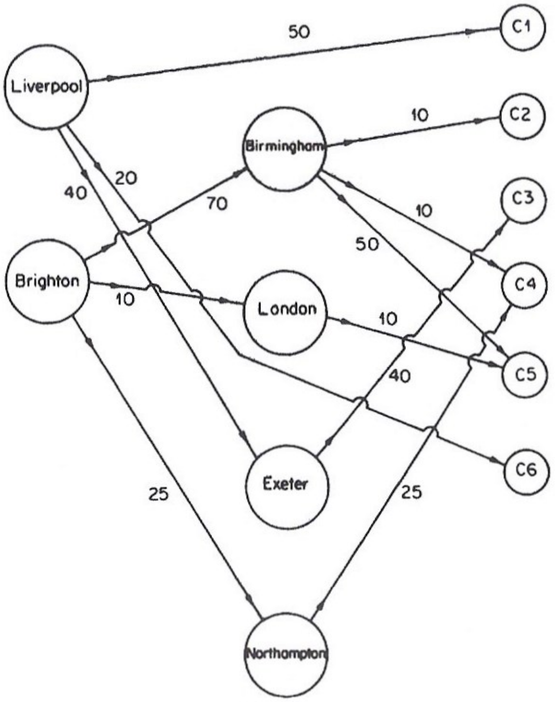


Figure 14.8

14.21 Agricultural Pricing

The optimal prices are

Milk	£303 per ton
Butter	£667 per ton
Cheese 1	£900 per ton
Cheese 2	£1085 per ton

The resultant yearly revenue will be £1992m. It is straightforward to calculate the yearly demands which will result from these prices. They are

Milk	4781000 tons
Butter	384000 tons
Cheese 1	250000 tons
Cheese 2	57000 tons

The economic cost of imposing a constraint on the price index can be obtained from the shadow price on the constraint. For this example this shadow price in the optimal solution indicates that each £1 by which the new prices are allowed to increase the cost of last year's consumption would result in an increased revenue of £0·61.

14.22 Efficiency Analysis

The efficient garages turn out to be: 3 (Basingstoke), 6 (Newbury), 7 (Portsmouth), 8 (Alresford), 9 (Salisbury), 11 (Alton), 15 (Weymouth), 16 (Portland), 18 (Petersfield), 22 (Southampton), 23 (Bournemouth), 24 (Henley), 25 (Maidenhead), 26 (Fareham) and 27 (Romsey).

It should be observed that these garages may be efficient for different reasons. For example, Newbury has 12 times the staff of Basingstoke but only five times as much showroom space. It sells 10 times as many Alphas, 10.4 times as many Betas and makes nine times as much profit. This suggests it makes more efficient use of showroom space, but less of staff.

The other garages are deemed inefficient. They are listed in Table 14.8 in decreasing order of efficiency together with the multiples of the efficient garages which demonstrate them to be inefficient.

For example, the comparators to Petworth taken in the multiples given below use inputs of:

Staff	5.02
Showroom space	550 square metres
Category 1 population	2 (1000s)
Category 2 population	2 (1000s)
Alpha enquiries	7·35 (100s)
Beta enquiries	3·98 (100s)

to produce outputs of

1·518 (1000s)	Alpha sales
0·568 (1000s)	Beta sales
1·568 (£million)	Profit

Table 14.8

Garage	Efficiency number	Multiples of efficient garages
19 Petworth	0·988	0·066(6) + 0·015(18) + 0·034(25) + 0·675(26)
21 Reading	0·982	1·269(3) + 0·544(15) + 1·199(16) + 2·86(24) + 1·37(25)
14 Bridport	0·971	0·033(3) + 0.470(16) + 0.783(24) + 0·195(25)
2 Andover	0·917	0·857(15) + 0·214(25)
28 Ringwood	0·876	0·008(3) + 0·320(16) + 0·146(24)
5 Woking	0·867	0·952(8) + 0·021(11) + 0·009(22) + 0·148(25)
4 Poole	0·862	0·329(3) + 0·757(16) + 0·434(24) + 0·345(25)
12 Weybridge	0·854	0·797(15) + 0·145(25) + 0·018(26)
1 Winchester	0·840	0·005(7) + 0·416(8) + 0·403(9) + 0·333(15) + 0·096(16)
13 Dorchester	0·839	0·134(3) + 0·104(8) + 0·119(15) + 0·752(16) + 0·035(24) + 0·479(26)
20 Midhurst	0·829	0·059(9) + 0·066(15) + 0·472(16) + 0·043(18) + 0·009(25)
17 Chichester	0·824	0·058(3) + 0·097(8) + (0·335(15) + 0·166(16) + 0·236(24) + 0·154(26)
10 Guildford	0·814	0·425(3) + 0·150(7) + 0·623(8) + 0·192(15) + 0·168(16)

```
! HPWilliams "Distribution 1"
min
 0.5 xline +0.5 xlibi + 1 xlilo + 0.2 xliex
 + 1 xli1 + 999 xli2 + 1.5 xli3 + 2 xli4 + 999 xli5 + 1 xli6
 + 999 xbrne + 0.3 xbrbi + 0.5 xbrlo + 0.2 xbrex
 + 2xbr1 + 999 xbr2 + 999xbr3 + 999 xbr4 + 999 xbr5 + 999 xbr6
 + 999 xne1 + 1.5 xne2 + 0.5 xne3 + 1.5 xne4 + 999 xne5 + 1 xne6
 + 1 xbi1 + 0.5 xbi2 + 0.5 xbi3 + 1 xbi4 + 0.5 xbi5 + 999 xbi6
 + 999 xlo1 + 1.5 xlo2 + 2 xlo3 + 999 xlo4 + 0.5 xlo5 + 1.5 xlo6
 + 999 xex1 + 999 xex2 + 0.2 xex3 + 1.5 xex4 + 0.5 xex5 + 1.5 xex6
st
fcap1) xline + xlibi + xlilo + xliex + xli1 + xli2 + xli3 + xli4 +
xli5 + xli6 <= 150
fcap2) xbrne + xbrbi + xbrlo + xbrex + xbr1 + xbr2 + xbr3 + xbr4 +
xbr5 + xbr6 <= 200
qine) xline + xbrne <= 70
qibi) xlibi + xbrbi <= 50
qilo) xlilo + xbrlo <= 100
qiex) xliex + xbrex <= 40
qone) xne1 + xne2 + xne3 + xne4 + xne5 + xne6 - xline - xbrne = 0
qobi) xbi1 + xbi2 + xbi3 + xbi4 + xbi5 + xbi6 - xlibi - xbrbi = 0
qolo) xlo1 + xlo2 + xlo3 + xlo4 + xlo5 + xlo6 - xlilo - xbrlo = 0
qoex) xex1 + xex2 + xex3 + xex4 + xex5 + xex6 - xliex - xbrex = 0
cust1) xli1 + xbr1 + xne1 + xbi1 + xlo1 + xex1 = 50
cust2) xli2 + xbr2 + xne2 + xbi2 + xlo2 + xex2 = 10
cust3) xli3 + xbr3 + xne3 + xbi3 + xlo3 + xex3 = 40
cust4) xli4 + xbr4 + xne4 + xbi4 + xlo4 + xex4 = 35
cust5) xli5 + xbr5 + xne5 + xbi5 + xlo5 + xex5 = 60
cust6) xli6 + xbr6 + xne6 + xbi6 + xlo6 + xex6 = 20
xbrex >= 20 ! "<" to force the "other" solution
end
```

JORGE BUESCU
Professor na Faculdade de Ciências da Universidade de Lisboa

O mistério do armazém absorvido

Imagine o leitor que pretende construir um armazém para distribuição de bens por um certo número de cidades. Qual é a localização óptima do armazém, no sentido de minimizar os custos totais de distribuição?

O problema não tem (apenas) um interesse académico. Se o leitor trabalhar numa empresa de distribuição, quererá ter armazéns localizados de forma a minimizar os custos de transporte dos bens aos clientes. Se trabalhar em energia ou telecomunicações, quererá ter uma rede de cabos eléctricos ou fibra óptica com o menor comprimento total de rede, mas que sirva as necessidades de tráfego (que são diferentes para pontos diferentes: Lisboa não terá a mesma intensidade de tráfego do que, digamos, as Berlengas). Se for responsável pela construção de uma unidade fabril, quererá localizá-la de forma a minimizar a distância total aos clientes. Se quisermos localizar uma central de transportes (um aeroporto, por exemplo), um critério importante é minimizar a distância total aos centros populacionais servidos, ponderando cada um pela sua população.

Os exemplos multiplicam-se: saber resolver este problema pode, dependendo das circunstâncias, poupar milhões de euros. O seu interesse económico é claro.

Tentemos formalizar o problema. Temos um conjunto de n pontos no plano, $X_1,\ldots, X_n$, com pesos $p_1,\ldots, p_n$. Em geral, interessa permitir que os pesos associados de pontos distintos sejam diferentes: significa que o número de clientes em cada ponto de distribuição x_i é variável. Se eu estiver a construir um armazém para distribuição de leite na região de Lisboa e Vale do Tejo, a importância económica do ponto Lisboa é muito maior do que, digamos, do ponto Bombarral: vou fazer muito mais vezes a viagem para Lisboa e, portanto, interessa-me reflectir a cidade de Lisboa com um peso maior (talvez proporcional ao número de habitantes).

Se eu situar o armazém num dado ponto (a,b) do plano, o custo associado $C(a,b)$ será a soma, ponderada pelos pesos p_i, das distâncias dos pontos fixos $X_i = (x_i, y_i)$ ao ponto (a,b):

$$C(a,b)=\sum_{i=1}^{n} p_i\sqrt{(x_i-a)^2+(y_i-b)^2}. \qquad (1)$$

O problema é, portanto, de optimização: o objectivo é escolher o ponto (a,b) de forma a minimizar o custo total $C(a,b)$ dado pela equação (1).

Em primeiro lugar, é fácil concluir que existe um ponto de mínimo: quando o ponto (a,b) se afasta do domínio convexo definido pelos X_i, a função custo C cresce (naturalmente!) sem limite. Por continuidade, tem de ter um mínimo nesse domínio convexo. Esse mínimo é, portanto, uma "média espacial" e, ingenuamente, poderíamos supor que essa média é um centro de massa.

Não é! De facto, esse ponto tem propriedades muito diferentes e contra-intuitivas. Para ilustrar a situação, suponhamos, como na figura 1, que temos cinco cidades diferentes, dispostas num pentágono regular. Comecemos com uma situação em que todas as cidades têm igual peso (20%); é óbvio, por simetria, que o ponto de mínimo, representado a vermelho, está no centro do pentágono.

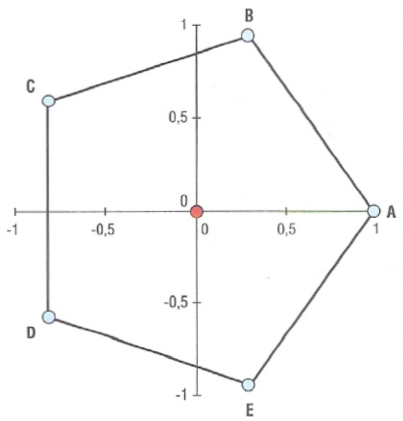


Figura 1

Façamos agora aumentar o peso relativo do ponto A, mantendo as outras quatro com peso igual entre si. Por simetria, o ponto óptimo estará sobre o eixo vertical. No entanto, desloca-se muito rapidamente à medida que o ponto A aumenta de peso. Na figura 2 está representada a sua localização quando o peso relativo de A é 43%: o ponto óptimo quase coincide com A!

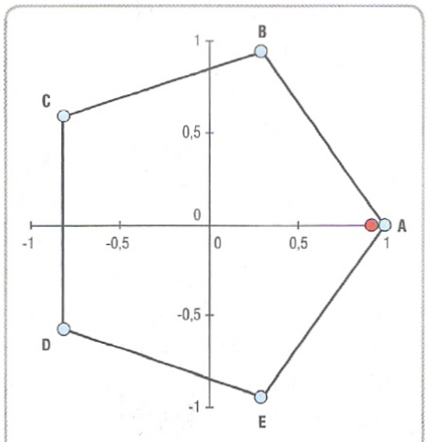


Figura 2

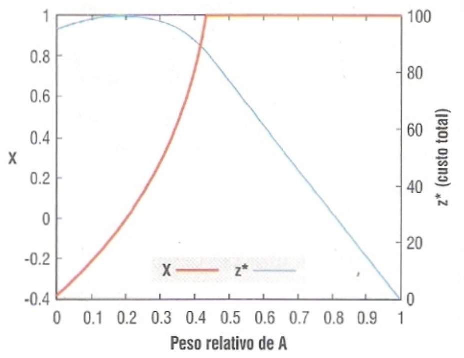


Figura 3

Na figura 3 representa-se, a vermelho, a forma como evolui a posição do ponto óptimo com o aumento do peso relativo de A: pouco antes de este atingir 45%, o ponto óptimo foi "absorvido" pelo ponto A. Tudo se passa como se os outros pontos simplesmente não existissem: o ponto óptimo já foi absorvido por A. Antes mesmo de o peso relativo de A ser 50%, o ponto óptimo já coincide com A.

Pensando no significado desta situação em termos do problema original, o que se está a passar é extremamente contra-intuitivo: se o peso relativo de uma das cidades é suficientemente grande (e nem sequer esmagadoramente grande!), a localização óptima do armazém (ou aeroporto, ou fábrica) colapsa automaticamente nela.

Tudo se passa como se uma cidade, atingida uma dimensão crítica, funcionasse de forma "absorvente" e fizesse com que a solução para optimização de localização de recursos fosse... concentrar ainda mais recursos nela própria.

O problema matemático em questão tem uma longa e ilustre história, tendo sido redescoberto várias vezes, e é conhecido como problema de Fermat-Weber, ou problema da mediana geométrica (nome que deriva de ser o análogo, no plano, da mediana numa distribuição unidimensional).

A essência do fenómeno contra-intuitivo de "absorção" por pontos suficientemente pesados provém do seguinte facto. Ao contrário do que uma consideração leviana poderia fazer pensar, a mediana geométrica (ou ponto de Fermat-Weber) *não é* o centro de massa do conjunto de pontos! De facto, o centro de massa C é o ponto que minimiza a soma (ponderada) *dos quadrados* das distâncias dos pontos X_i a C. O ponto de Fermat-Weber, pelo contrário, minimiza a soma ponderada *dos módulos* das distâncias, como se pode ver pela equação (1).

Este (aparentemente pequeno) pormenor faz toda a diferença. Para determinar o centro de massa, a aplicação do Cálculo Diferencial é imediata e as equações do centro de massa são lineares, tendo o comportamento "bonito" a que estamos habituados. As equações para o ponto de Fermat-Weber são, contudo, não-lineares: daí o comportamento distorcido das soluções. →

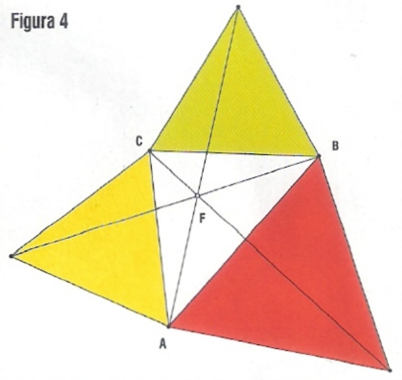
JORGE BUESCU
Professor na Faculdade de Ciências da Universidade de Lisboa

Pior: a função $C(a,b)$ não é diferenciável no ponto de Fermat-Weber (pois a função módulo não possui derivada na origem), pelo que estamos perante um problema de optimização que *não* se resolve por aplicação cega do cálculo diferencial.

A distinção entre mediana geométrica (como é conhecida nos meios mais ligados à estatística) e centro de massa é subtil. No início do século XX o *Census Bureau* dos Estados Unidos cometeu o erro de os identificar, afirmando que o "ponto de distância mínima agregada" era o centro de gravidade das populações. Resultado: durante quase duas décadas o erro propagou-se a livros de demografia e estatística, sendo corrigido por volta de 1930 pelos matemáticos Eells e Gini.

O problema de Fermat-Weber originou-se, como o nome indica, com Pierre de Fermat no século XVII. Dado um triângulo, Fermat pede o ponto que minimiza a soma das distâncias aos vértices do triângulo. O problema foi resolvido pelo próprio Fermat, por Evangelista Torricelli e por vários outros matemáticos. O ponto óptimo (de Fermat) está no cruzamento das medianas dos triângulos equiláteros apoiados nas arestas do triângulo (figura 4).

Figura 4



Esse ponto de Fermat tem a propriedade de "ver" todos os vértices do triângulo original a 120º. Note-se, em particular, o seguinte: se o triângulo possui um ângulo maior ou igual a 120º, então não existem pontos nestas condições, e o ponto de Fermat coincide com o vértice correspondente ao maior ângulo. Portanto, até no problema de Fermat surge o fenómeno de "absorção" do ponto óptimo!

O economista alemão Alfred Weber, irmão de Max Weber, reformulou, no início do século XX, o problema de Fermat como um problema de localização óptima, utilizando n pontos e pesos distintos, como interessa na interpretação económica do problema. Foi o primeiro problema da classe a que se chama hoje *Facility location*.

O problema de Fermat-Weber permite uma solução mecânica interessante, dada originalmente pelo matemático francês Varignon (sécs. XVII-XVIII) e representada na figura 5. Num plano fazem-se n furos correspondentes às localizações das cidades $X_1, \dots X_n$. Atam-se n fios num único nó, passando o fio i pelo furo correspondente a X_i e suspendendo-se dele o peso correspondente p_i. Então, atingido o equilíbrio, uma análise de forças mostra que a posição do nó estará no ponto de Fermat-Weber (fig. 5). Esta *armação de Varignon* chegou mesmo a ser utilizada, nalgumas circunstâncias, como computador analógico para resolver o problema de Fermat-Weber.

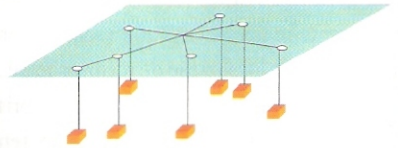


Figura 5

Se tiver inclinação para isso, o leitor pode observar todos os fenómenos contra-intuitivos do problema de Fermat-Weber construindo uma armação de Varignon. Por exemplo, se um dos pesos for maior do que a soma de todos os outros, o nó da armação de Varignon vai ser "sugado" – ou, na terminologia anterior, "absorvido" – pelo furo correspondente a esse peso.

Na verdade, como o nosso exemplo para $n=5$ mostra, esta condição nem sequer é necessária para que um ponto seja "absorvente". Por análise de forças, basta que a resultante das forças dos outros pesos seja menor ou igual do que o peso suspenso do furo em questão. Esta condição é suficiente para garantir que o ponto de Fermat-Weber é absorvido pelo furo. Isto pode mesmo acontecer se o peso correspondente for pequeno!

É possível mostrar que o problema geral de Fermat-Weber, ao contrário do problema dos centros de massa, não possui uma solução analítica (em radicais). Assim, os métodos aproximados revelam-se cruciais. Em particular, com o advento dos computadores digitais, os métodos iterativos para solução permitem hoje em dia calcular soluções aproximadas com um grau arbitrariamente elevado de precisão. Existem vários algoritmos para o fazer, sendo o mais utilizado o algoritmo de Weiszfeld.

São possíveis muitas generalizações úteis do problema de Fermat-Weber. Por exemplo, podemos querer não trabalhar no plano, mas no espaço (ou numa superfície esférica, se estivermos a pensar numa rede sismológica de detecção de *tsunamis*) ou mesmo em dimensão n. Podemos querer ter vários armazéns a servir os pontos de procura (problema que tem o nome de *Multifacility location*). Podemos querer usar outras noções de distância que não a euclidiana (por exemplo, que leve em conta a rede de vias de transporte). E assim por diante.

Cada contexto poderá, naturalmente, dar origem a novos fenómenos possivelmente inesperados. No entanto, é interessante observar que o misterioso fenómeno da "absorção" do ponto óptimo por um dos pontos não está relacionado com este tipo de complexidade, surgindo no contexto mais simples possível: três pontos num plano, o humilde triângulo de Fermat. Em contextos mais complexos podem surgir mais paradoxos – mas este estará obrigatoriamente entre eles. Do ponto de vista económico, o fenómeno da absorção do ponto de Fermat-Weber mostra que, em problemas de localização de pontos de distribuição, existe uma dimensão crítica acima da qual a decisão mais racional de um agente económico é fazer coincidir o ponto de distribuição com o maior ponto de procura. Parece pouco democrático, mas nada tem a ver com questões ideológicas: é um resultado matemático.

Um pouco especulativamente, talvez seja esta a lógica do aparecimento de megametrópoles no Terceiro Mundo: enquanto nos países "ricos" existe um excedente para pagar o "prémio" de deslocalizar os armazéns para zonas industriais periféricas, preservando a qualidade de vida nas cidades, no Terceiro Mundo não existe esse excedente de recursos. A opção é, assim, pela solução mais barata: os pontos de distribuição concentram-se, de acordo com a absorção de Fermat-Weber, na maior das cidades, que em consequência se torna cada vez maior. Talvez até aos 20 milhões de habitantes de São Paulo ou Cidade do México. ■

O autor agradece a colaboração com o Prof. Miguel Casquilho, do IST, que lhe deu a conhecer o problema e autorizou a reprodução de alguns dos gráficos apresentados. A sua página http://web.ist.utl.pt/mcasquilho/compute/_location/ possui referências bibliográficas e recursos para realizar cálculos numéricos de Fermat-Weber em tempo real.

(Blank page)

Travelling Salesman Problem

MIGUEL A. S. CASQUILHO
Technical University of Lisbon, Ave. Rovisco Pais, 1049-001 Lisboa, Portugal

The "Travelling Salesman Problem" is briefly presented, with reference to problems that can be assimilated to it and solved by the same technique. Examples are shown and solved.

Key words: *travelling salesman problem; supply; demand; optimization.*

1. Fundamentals and scope

With optimization in mind, the "travelling salesman problem", frequently denoted by the initials TSP[1], is a fundamental subject related to travelling and transportation, with several generalizations and with insertion in more complex situations, and also akin to others apparently unrelated, resoluble by the techniques used for the typical case. The TSP is known for the striking contrast between the simplicity of its formulation and the difficulty of its resolution, some even saying that it still does not have a solution. It is a so-called NP-hard[2] problem (its difficulty increasing more than polynomially with its size). Anyway, something substantial can be presented about the problem.

The problem arises from the typical situation of a salesman who wants to visit his clients in a given set of cities and return to his own city, thus performing a cycle. The problem can be envisaged in this large scale, but also exists in any other scales, such as within a factory or on a microchip. An asymmetrical TSP can also be the search for the optimum ordering of paint manufacturing or the preparation of fruit juices in a common plant, because, in these cases, setup costs (washing, etc.) depends significantly on the "vicinity" of the colours or of the flavours.

The mathematical formulation of the problem can be as in Eq. {1}.

$$[\min] z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

subject to

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} &= 1 \qquad i = 1..m \\ \sum_{i=1}^{m} x_{ij} &= 1 \qquad j = 1..n \\ x_{ij} &= 0 \text{ or } 1 \qquad \forall i, j \end{aligned} \qquad \{1\}$$

Solution must be a cycle.

As in the Assignment Problem (AP), c_{ij}, with $i, j = 1..n$, is the cost (or distance, etc.) to go from city i to city j, and x_{ij} will be 1 if the arc from i to j is used and 0 otherwise. The problem would be an AP if the last condition (one and only one

[1] Also "traveling salesman problem" (American English) or sometimes "salesperson".
[2] See, *e.g.*, http://en.wikipedia.org/wiki/NP-hard .

Miguel Casquilho is Assistant Professor (retired) in the Departament of Chemical Engineering, Instituto Superior Técnico, Universidade Técnica de Lisboa (Technical University of Lisbon), Lisbon, Portugal. E-mail address: *mcasquilho@ist.utl.pt*.

cycle) —a fundamental difference— did not exist. In order to use the above formulation the costs c_{ii}, to go from one city to itself must be made prohibitive, otherwise the (useless) solution would be $x_{ii} = 1$, others zero (not a cycle).

A branch-and-bound (B&B) exact algorithm based on AP relaxation will be presented with some examples. A language such as Mathematica has functions *TravelingSalesman* and *FindShortestTour* that solve the problem, even in 3D, but the objective here is to clarify the method itself. An advanced algorithm by Carpaneto *et al.* [1995CAR] is used and made available at the author's website [2012CAS]. Several other exact algorithms, all detailed by Lawler *et al.* [1995LAW], have been constructed, as well as heuristic ones (i.e., approximate), namely for large size problems or more complex problems.

2. Examples

EXAMPLE: 5 CITIES (WINSTON)

A travelling salesman has to cover a set of 5 cities (his own included) periodically (say, once per week) and return home. The distances between the cities are given in Table 1, as could have been read on a map. Determine the most economical cycle, i.e., with minimum length (example from Winston [2003WIN], p 530 ff).

Table 1 — Cost matrix (distances in km) from one city to another

	1	2	3	4	5
1	—	132	217	164	58
2	132	—	290	201	79
3	217	290	—	113	303
4	164	201	113	—	196
5	58	79	303	196	—

In this problem the cost matrix is symmetric, which appears obvious. The cost matrix can, however, be asymmetric, as in the case of air travel because of predominant wind or in one-way urban streets. No advantage is taken from symmetry in the present text, but symmetry may be important in many variants of the solution methods.

RESOLUTION

The strategy of AP relaxation is to solve the problem as an AP, leaving aside the condition of "one cycle". The solution to the AP can be obtained, *e.g.*, with Excel or Excel/Cplex, and is the one in Table 2, with total cost $z = 495$ km.

Table 2 — Solution to the AP relaxation

	1	2	3	4	5
1	—	0	0	0	1
2	1	—	0	0	0
3	0	0	—	1	0
4	0	0	1	—	0
5	0	1	0	0	—

This is, however, not a solution to the TSP, because there are subtours: $x_{15} = x_{21} = x_{34} = x_{43} = x_{52} = 1$, i.e., two subtours, 1–5–2–1 and 3–4–3. The B&B technique will now be used, as follows.

The former problem, say, Problem 1, is replaced by others, considering the shortest subtour (the one with least arcs) to try to save computation effort. So, 3–4–3 will be chosen. Now, two problems replace the previous one: Problem 2, from Problem 1 but prohibiting 3–4; and Problem 2, from Problem 1 but prohibiting 4–3. The two new problems have the cost matrices in Table 3 and the solutions in Table 4

Table 3 — Cost matrices for Problem 2 (l.-h.) and Problem 3 (r.-h.)

	1	2	3	4	5
1	—	132	217	164	58
2	132	—	290	201	79
3	217	290	—	—	303
4	164	201	113	—	196
5	58	79	303	196	—

	1	2	3	4	5
1	—	132	217	164	58
2	132	—	290	201	79
3	217	290	—	113	303
4	164	201	—	—	196
5	58	79	303	196	—

Table 4 — Solutions to the AP relaxations

	1	2	3	4	5
1	—	0	0	1	0
2	0	—	0	0	1
3	1	0	—	0	0
4	0	0	1	—	0
5	0	1	0	0	—

	1	2	3	4	5
1	—	0	1	0	0
2	0	—	0	0	1
3	0	0	—	1	0
4	1	0	0	—	0
5	0	1	0	0	—

Again, these are not solutions to the TSP, because there are subtours: for Problem 2, 1–4–3–1, 2–5–2; and for Problem 3, 1–3–4–1, 2–5–2. (Symmetry causes some redundancy.) In file *TSP_Wins530.xls*, the complete procedure is shown until the optimum is found.

The solution procedure is usually shown as a tree, constructed progressively as the problem is solved. The one for this problem is (from *TSP_Raffensperger.ppt*) given in Figure 1, with solution 1–5–2–4–3–1 and $z^* = 668$ km.

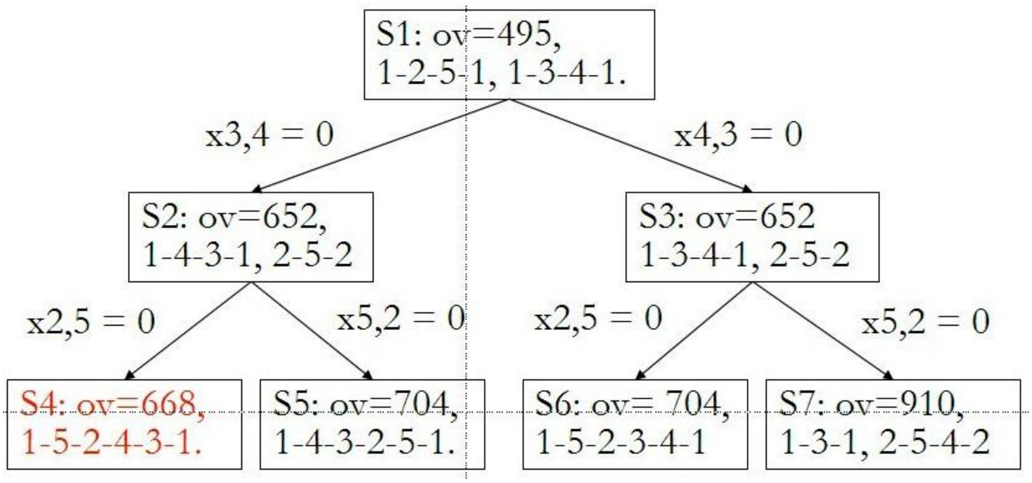


Figure 1 — Tree for the Winston TSP problem.

■

About the B&B, notice that:

(a) (disadvantage) Every replacement of a problem gives rise to two or more "children" and can never improve the value of the objective function (495, 652, 668, increasing, while a minimum was sought); but

(b) (advantage) It is, hopefully, not necessary to investigate all the combinations, this being the real merit of the B&B.

The systematic procedure of B&B permits to avoid the investigation of all the combinations. For a mere $n = 18$ cities —the number of Portugal's main cities ("district capitals"[3])—, these combinations would be $17\,!$, which is $3{,}5 \times 10^{14}$ (11 years of computing at one combination per microsecond).

The motive it is considered that the TSP still "has no solution" is that the size of the tree is not predictable. So, the amount of memory or disk storage occupied can increase beyond the availability, making this not only a matter of time, but often more so one of space.

EXAMPLE: GRID 2

In a manufacture of grids having 4 points in a square arrangement of 2×2, as in Figure 2, all the points have to be treated (*e.g.*, welded, connected, painted) in a certain order. Determine the most economic order.

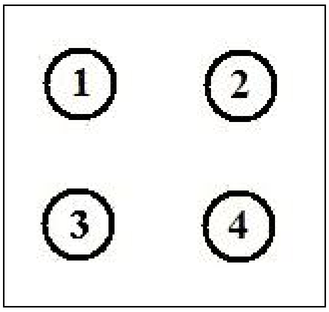


Figure 2 — Grid with 2 by 2 points in a square arrangement.

RESOLUTION

It is obvious that the answer is 1–2–4–3, with $z^* = 4$, if, as is implicit, the distances between the adjacent points (horizontally or vertically) is one. If, however, the problem is to be solved by a convenient algorithm[4], the cost matrix has to be provided and is given in Table 5. In order to use the web site, supply –1 as infinity (diagonal entries).

Table 5 — Cost matrix for the "Grid 2"

	1	2	3	4
1	—	1	1	2
2	1	—	2	1
3	1	2	—	1
4	2	1	1	—

[3] The *district* ("distrito"), of which there are 18 (average 5 500 km^2) in continental Portugal, is roughly equivalent in area to one French *département* (totalling 96) or two Italian *province* (110).

[4] The one on the web, based on Carpaneto *et al.*.

Example: Grid 3

Now suppose that (previous example) it is the manufacture of grids having 9 points in a square arrangement of 3×3, as in Figure 3. Determine the most economic order.

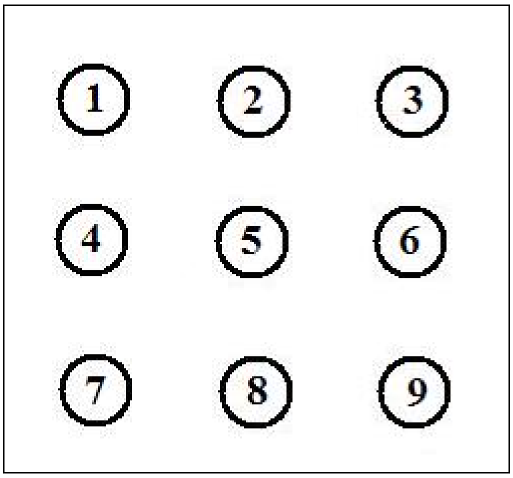


Figure 3 — Grid with 3 by 3 points in a square arrangement.

Resolution

Now the solution is by no means obvious. See *TSP_grids.xls*. The problem is solved assuming both a *taxicab*[5] *geometry*[6] and a *Euclidean geometry*. The taxicab geometry assumes a grid layout, such as the arrangement of streets in certain zones of the cities, hence the mention to the taxicab, so (as in the previous example) only "vertical" and "horizontal" movements are possible. The cost matrix is given in Table 6.

Table 6 — Taxicab cost matrix for the "Grid 3"

	1	2	3	4	5	6	7	8	9
1	—	1	2	1	2	3	2	3	4
2	1	—	1	2	1	2	3	2	3
3	2	1	—	3	2	1	4	3	2
4	1	2	3	—	1	2	1	2	3
5	2	1	2	1	—	1	2	1	2
6	3	2	1	2	1	—	3	2	1
7	2	3	4	1	2	3	—	1	2
8	3	2	3	2	1	2	1	—	1
9	4	3	2	3	2	1	2	1	—

The solution is 1-2-5-4-7-8-9-6-3-1, with $z^* = 10$. Looking at these results, one may wonder if allowing other movements may result in a better value. This leads to the common, Euclidean geometry. Now the cost matrix is given in Table 7.

The solution to the Euclidean geometry Grid 3 problem, as shown in Table 7, is 1-4-7-8-9-6-3-5-2-1, with $z^* = 9{,}41$. As expected, allowing a less constrained geometry has led to a better result.

---

[5] From *taximeter cabriolet*.
[6] See, *e.g.*, http://en.wikipedia.org/wiki/Taxicab_geometry .

Table 7 — Solution for the Euclidean cost matrix for the “Grid 3”

	1	2	3	4	5	6	7	8	9
1	—	1	2	1	1,41421	2,23607	2	2,23607	2,82843
2	1	—	1	1,41421	1	1,41421	2,23607	2	2,23607
3	2	1	—	2,23607	1,41421	1	2,82843	2,23607	2
4	1	1,41421	2,23607	—	1	2	1	1,41421	2,23607
5	1,41421	1	1,41421	1	—	1	1,41421	1	1,41421
6	2,23607	1,41421	1	2	1	—	2,23607	1,41421	1
7	2	2,23607	2,82843	1	1,41421	2,23607	—	1	2
8	2,23607	2	2,23607	1,41421	1	1,41421	1	—	1
9	2,82843	2,23607	2	2,23607	1,41421	1	2	1	—

3. Conclusions

The TSP is a remarkable problem both for the contrast between the simplicity of its formulation and the complexity of its resolution, and the variety of its applications. The exact resolution was presented using the branch-and-bound technique applied to Assignment Problem relaxations. A more advanced algorithm by Carpaneto *et al.* is made available for the solution of this type of problems.

Examples of typical situations were presented, namely, a problem with taxicab and Euclidean distances as costs.

Acknowledgements

This work was done at “Centro de Processos Químicos do IST” (Chemical Process Research Centre of IST), Department of Chemical Engineering, Technical University of Lisbon. Computations were done on the central system of CIIST (Computing Centre of IST).

Bibliography

– [1995CAR] CARPANETO, G., M. DELL’AMICO and P. TOTH, 1995, “Exact solution of large-scale asymmetric traveling salesman problems”, *ACM Transactions on Mathematical Software*, 21:394–409.
– [2012CAS] CASQUILHO, M., http://web.ist.utl.pt/~mcasquilho/compute/or/Fx-TSP.php, accessed May 2012..
– [1985LAW] LAWLER, E. L., J. K. LENSTRA, A. H. G. RINNOOY KAN, D. B. Shmoys, “The travelling salesman problem”, editors, Chichester (UK). ISBN 0-471-90413-9.
– [2003WIN] WINSTON, Wayne L., and M. VENKATARAMANAN, 2003, “Introduction to Mathematical Programming”, Thomson, New York, NY (USA). ISBN 0-534-35964-7

Distance to points in rectangle

Computes the distance from a given point to the others, in a rectangle.

Execute

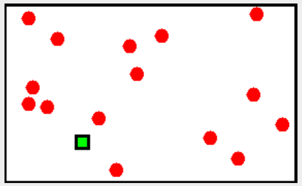
2014.Jul.06 01:42:13

x, *y*	0.6 0.2	*Location* of given (source) point. •
b, *h*	2 1	*Base* and *height* of rectangle. •
N	1e+7 (≤ 1e+8)	No. of random destination *points*. •
.seed, klass	0 200	*Seed* for random numbers, and no. of histogram *classes*. •
Show values	No	Shows the coordinates of the graph. •

Simulates, via Monte Carlo, the distances from the source point to *N* random points in a given rectangle, with base *b* and height *h*, in order to find the distribution of the distance, *d*. In the Figure, is shown a rectangle with the source point (square) and *N* (here, a few) random points.

Plots the density function (pdf), *f*(*d*), and the probability function (cdf), *F*(*d*), for the distance, *d*, and computes its mean and standard-deviation. (The user given point can be *out* of the rectangle.)

Other suggested data for (*x*, *y*): (0.2, 0.2), (1, 0.2), (2.6, 0.2)

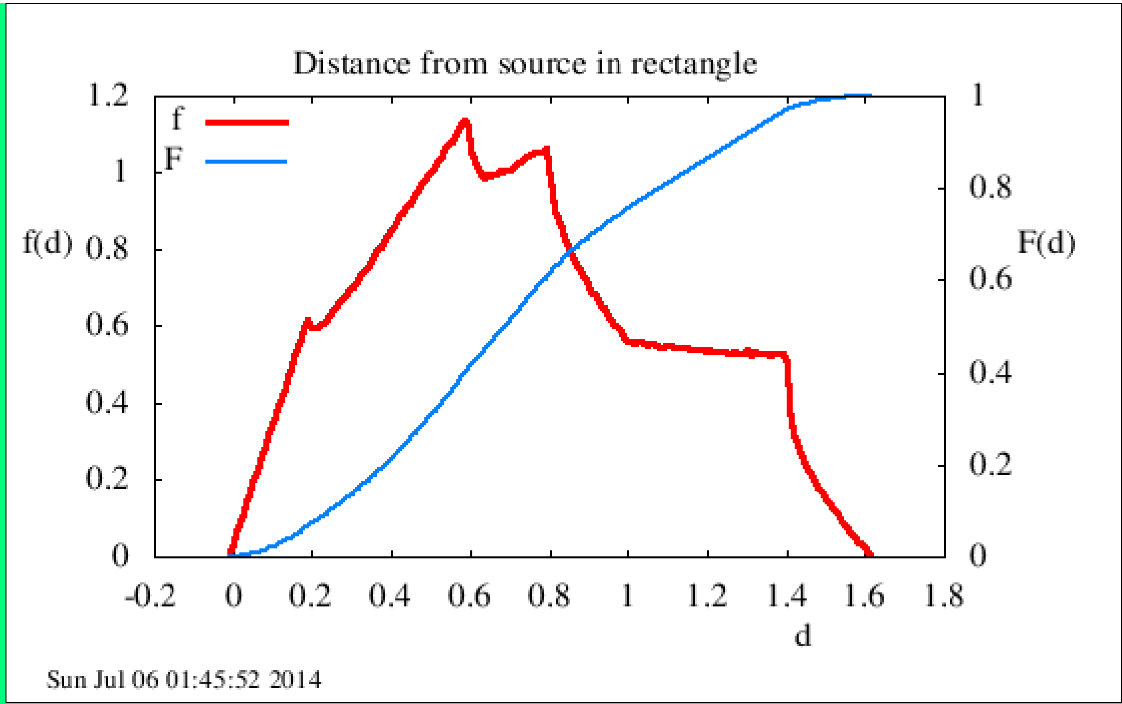


Reset	*References:*	Plate: distInRectang

• Dartmouth College, 2014, Chapter 2.*pdf*, "Continuous probability densities"
• Weisstein, Eric W., "Disk Point Picking", *MathWorld*—A Wolfram Web Resource
• 1838-01-05: JORDAN, Marie Ennemonde Camille (1922-01-22) (not Wilhelm Jordan, 1842–1899).

Results

```
2014-07-06 UTC+0060  1:45:52.078
*** Distance from point in a rectangle ***                          (Jan-2014, MC)
                                                                       [149914306]
Source point,      0.6000      0.2000    | (x, y)
Rectangle,          2.000       1.000    | base, height
No. of random points, N,    10000000    | = 10 ^  7.00
.Seed (repeatability),             0    | (0|>=1: no|yes)
No. of histogram classes,        200    |
Show coord.s ?                     0    | (0|1: no|yes)
----------------------------------------+----------------------------------------
Rectangle diagonal,           2.2361    |
Distances of source to vertices:        |
                    1.414       1.612    |
                   0.6325       1.000    |
Distance is in     0.0000      1.6125    |
Distance interval width,      1.6125    |
Dist. min, max,  3.50862E-04  1.6117    |
Aver., stdev,    0.73129     0.36291    |

2014-07-06 UTC+0060  1:45:52.078
2014-07-06 UTC+0060  1:45:52.911      "CPU":       0.8 sec.          End
                                       0d 00h 00m  0.88s
```



Go *back* using your browser.

(Blank page)

Packet filling, as tubes

Fills (via Monte Carlo) packets, as tubes, with items. **SPE Nazaré/pt**

Execute

2014.Jul.06 01:53:43

.*L*, *U*	2000 2080 g (specification limits)	*Lower*, *upper* specs on packet weight (ΣX). •
μ, σ	100 3 g	*Mean* and *st. dev.* for the weight of each item, *X*.
X_a, X_b	92 108 g ($0 < X_a < x < X_b$)	Truncation X_a (min) and X_b (max) for item. •
Strategy	none FIFO	Strategy: *none* or 'FIFO *rectify*'. •
N, .seed	1×10^6 0	No. of *items* ("lot size"), random no. gener. *seed*. •
tol, klass, y_{max}	1.-6 200 0 g^{-1} ['0' (¬ '.0'), auto.]	*Tolerance*, no. of histo. *classes*, max. *y* for graph. •
Show values	No	Shows the coordinates of the graph. •

Simulates, via Monte Carlo, the filling of a packet ("bag") of items, in a tube-style, FIFO (first in, first out) strategy (as in a doubly open tube). The objective is a final filling with weight (mass) in (*L*, *U*). An example might be to fill a bag of 2 kg of oranges, namely (as in the base data), weighing between *L* = 2000 and *U* = 2080 g.

The weight of each item is considered *truncated Gaussian*. (*Tolerance* is for the inversion of the Gaussian distribution.)

Terminology for *fractions* (costs incurred): 'frTr', fr. truncated from the original product; 'frWa' (waste), fr. wasted as giveaway (weight above *L*); 'frRe' (rejected), fr. of unused items, possibly recyclable.

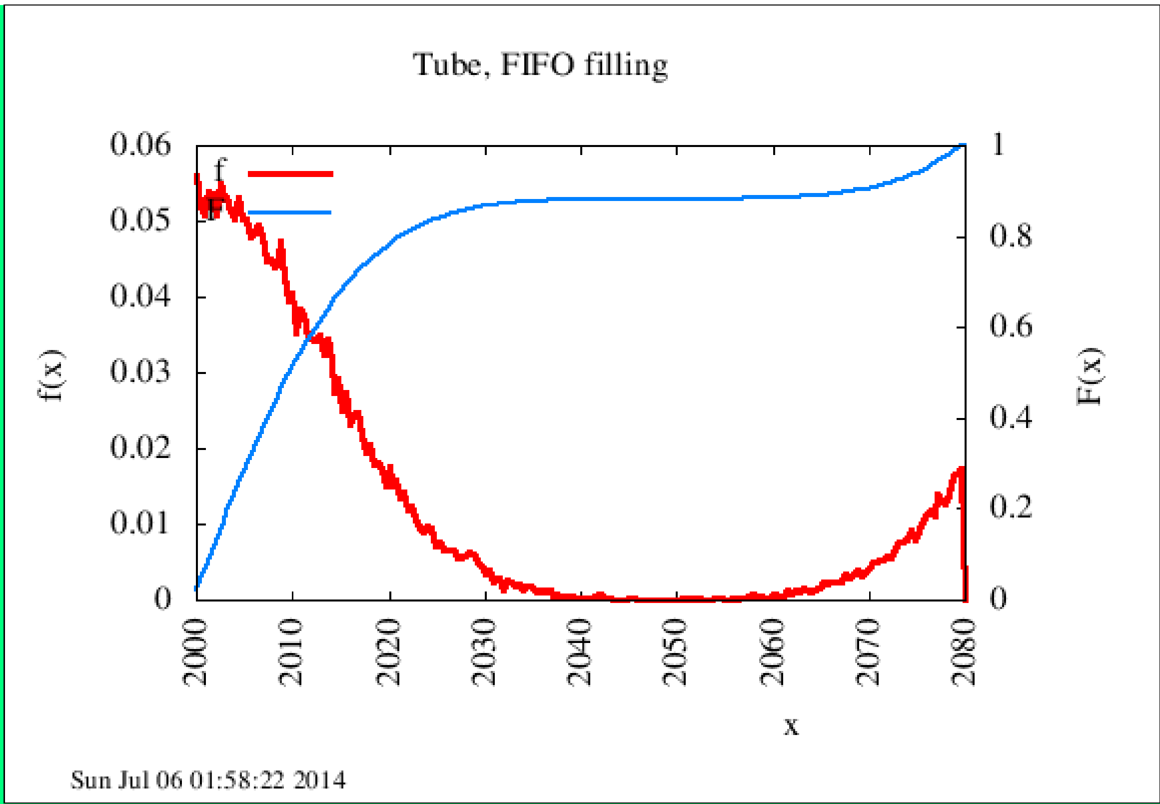
Several other, better strategies (worthy of research) would be convenient to try to reach the final sum within the limits.

Reset	*References:*	Plate: TubePacketFill3

• "Calibrafruta", Leiria (Portugal).
• 1900-04-12: YOUDEN, William John (1971-03-31).

Results

```
2014-07-06 UTC+0060  1:58:11.237
*** Tube-style packet filling ***                                 (Apr-2011, MC)

.L, U,              2000.       2080.    | specs for packet (total) weight
.mu, sigma,         100.0       3.000    | for 'X'
x_a, x_b (tr.),     92.00       108.0    | truncated Gaussian, X in (x_a, x_b)
Rectifying strategy,                0    | (0: none)
No. of trials, N,             10^ 6.0    | =   1000000       (max int, 10^ 9.33)
.Seed (repeatability),              0    | (0|>=1: no|yes)
No. of histogram classes,         200    | tol,  1.0E-06 (for Gaussian inversion)
Show coord.s ?                      0    | (0|1: no|yes)
-----------------------------------------+---------------------------------------
mid_value, occup., 2040.         20.4    | (occupancy ~= .mid_value / .mu)
                                         |
*** THEORETICAL *** truncated Gaussian   | (_t, truncated)
a', b',            -2.667       2.667    | [a' = (x_a - mu) / sigma, b' idem b]
.mu_t, sigma_t,    100.00      2.9067    | for truncated X
D_mu, rho_sigma,    0.00        0.969    | (mu_t-mu, sigma_t/sigma: from trunc.)
a', b',             -2.67        2.67    | a' = (a - mu) / sigma, idem b'
Phi(a'), Phi(b'),   3.83E-03  0.9962     |
  deltaPhi,                   99.234 %   | (fraction retained)
Fraction truncated, frTr,      0.766 %   | (= 1 - deltaPhi)
Packet occupancy:                        | from   18.52 to   22.61, i.e.,
  items min, max,            19    22    | [=(L/b)+, (U/a)-]
L/max, U/min,       90.91       109.5    | extreme equal X's ('x_a', 'x_b')
                                         |
*** SIMULATED ***                        |
 No. of data     average       stdev     | (moments)
     1000000      100.00      2.9098     | all items
      556344      100.30      2.9128     | accepted items
       27655      2017.8      21.883     | packets
sigma_acc / sigma_all,        1.0010     | (accepted vs. all)
Occupancy, frequency: --  20  0.8827     |
                      --  21  0.1173     |
                     min         max     | average      stdev
Occupancy,            20          21     |  20.117     0.32179
Weight,            2000.0      2080.0    |  2017.8      21.883
Rejected,                     443656     | (all,   1000000)
Fraction rejected, frRe,      44.366 %   | (rejected / all)
SUMMARY, fractions:                      |
            frTr      frWa      frRe     |
            0.77 %    0.89 %   44.37 %   |

2014-07-06 UTC+0060  1:58:11.237
2014-07-06 UTC+0060  1:58:22.929        CPU:      11.7 sec.          End
```



Go *back* using your browser.

Packet filling, as tubes

Fills (via Monte Carlo) packets, as tubes, with items. **SPE Nazaré/pt**

Execute

2014.Jul.06 01:53:43

.L, U	2000 2080 g (specification limits)	*Lower*, *upper* specs on packet weight (ΣX). •
μ, σ	100 3 g	*Mean* and *st. dev.* for the weight of each item, X.
X_a, X_b	95 105 g ($0 < X_a < x < X_b$)	Truncation X_a (min) and X_b (max) for item. •
Strategy	none FIFO	Strategy: *none* or 'FIFO *rectify*'. •
N, .seed	1×10^6 0	No. of *items* ("lot size"), random no. gener. *seed*. •
tol, klass, y_{max}	1.-6 200 0 g^{-1} ['0' (¬ '.0'), auto.]	*Tolerance*, no. of histo. *classes*, max. y for graph. •
Show values	No	Shows the coordinates of the graph. •

Simulates, via Monte Carlo, the filling of a packet ("bag") of items, in a tube-style, FIFO (first in, first out) strategy (as in a doubly open tube). The objective is a final filling with weight (mass) in (L, U). An example might be to fill a bag of 2 kg of oranges, namely (as in the base data), weighing between L = 2000 and U = 2080 g.

The weight of each item is considered *truncated Gaussian*. (*Tolerance* is for the inversion of the Gaussian distribution.)

Terminology for *fractions* (costs incurred): 'frTr', fr. truncated from the original product; 'frWa' (waste), fr. wasted as giveaway (weight above L); 'frRe' (rejected), fr. of unused items, possibly recyclable.

Several other, better strategies (worthy of research) would be convenient to try to reach the final sum within the limits.

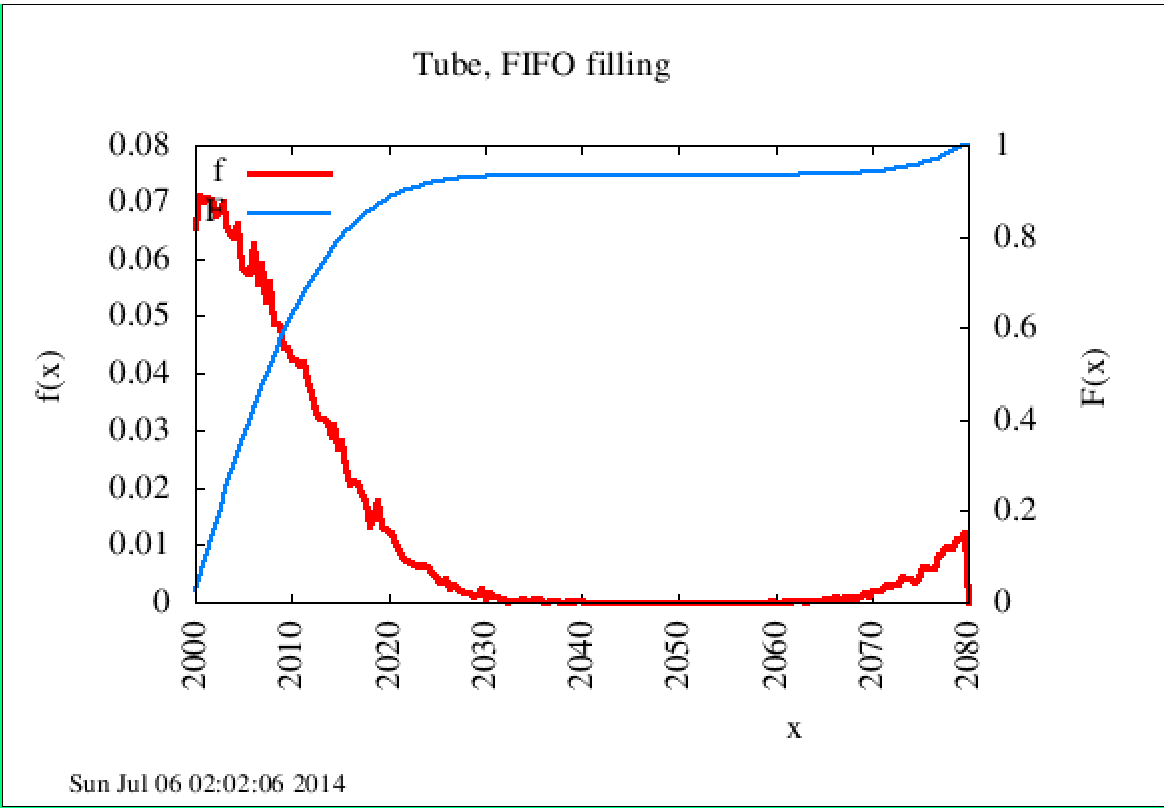
Reset	*References:*	Plate: TubePacketFill3

• "Calibrafruta", Leiria (Portugal).
• 1900-04-12: YOUDEN, William John (1971-03-31).

http://web.tecnico.ulisboa.pt/~mcasquilho/compute/qc/Fx-3tubefill.php
Created: 2011-04-12 — *Last modified:* 2011-10-01

Results

```
2014-07-06 UTC+0060  2:01:17.493
*** Tube-style packet filling ***                                (Apr-2011, MC)

.L, U,              2000.       2080.    | specs for packet (total) weight
.mu, sigma,         100.0       3.000    | for 'X'
x_a, x_b (tr.),     95.00       105.0    | truncated Gaussian, X in (x_a, x_b)
Rectifying strategy,                0    | (0: none)
No. of trials, N,              10^ 6.0   | =   1000000        (max int, 10^ 9.33)
.Seed (repeatability),              0    | (0|>=1: no|yes)
No. of histogram classes,         200    | tol,  1.0E-06 (for Gaussian inversion)
Show coord.s ?                      0    | (0|1: no|yes)
----------------------------------------+---------------------------------------
mid_value, occup., 2040.         20.4    | (occupancy ~= .mid_value / .mu)
                                         |
*** THEORETICAL *** truncated Gaussian   | (_t, truncated)
a', b',            -1.667       1.667    | [a' = (x_a - mu) / sigma, b' idem b]
.mu_t, sigma_t,    100.00       2.3875   | for truncated X
D_mu, rho_sigma,     0.00       0.796    | (mu_t-mu, sigma_t/sigma: from trunc.)
a', b',             -1.67        1.67    | a' = (a - mu) / sigma, idem b'
Phi(a'), Phi(b'),   4.78E-02  0.9522     |
  deltaPhi,                   90.442 %   | (fraction retained)
Fraction truncated, frTr,      9.558 %   | (= 1 - deltaPhi)
Packet occupancy:                        | from   19.05 to   21.89, i.e.,
  items min, max,             20    21   | [=(L/b)+, (U/a)-]
L/max, U/min,       95.24       104.0    | extreme equal X's ('x_a', 'x_b')
                                         |
*** SIMULATED ***                        |
 No. of data     average       stdev     | (moments)
     1000000      99.998      2.3850     | all items
      520047      100.32      2.3735     | accepted items
       25919      2012.9      17.629     | packets
sigma_acc / sigma_all,        0.9952     | (accepted vs. all)
Occupancy, frequency: --   20  0.9357    |
                      --   21  0.0643    |
                     min          max    | average       stdev
Occupancy,            20           21    |  20.064     0.24532
Weight,            2000.0      2080.0    |  2012.9      17.629
Rejected,                      479953    | (all,   1000000)
Fraction rejected, frRe,      47.995 %   | (rejected / all)
SUMMARY, fractions:                      |
            frTr      frWa      frRe     |
            9.56 %    0.64 %   48.00 %   |

2014-07-06 UTC+0060  2:01:17.493
2014-07-06 UTC+0060  2:02:06.842       CPU:      49.3 sec.          End
```



Go *back* using your browser.