E

EIGENVALUE ENCLOSURES FOR ORDINARY DIFFERENTIAL EQUATIONS

Selfadjoint eigenvalue problems for ordinary differential equations are very important in the sciences and in engineering. The characterization of eigenvalues by a minimum-maximum principle for the *Rayleigh quotient* forms the basis for the famous *Rayleigh–Ritz method*. This method allows for an efficient computation of nonincreasing upper eigenvalue bounds. N.J. Lehmann and H.J. Maehly [6], [7], [8] independently developed complementary characterizations that can be used to compute lower bounds. These methods are based on extremal principles for the *Temple quotient*. In general, however, an application of the *Lehmann–Maehly method* requires that certain quantities can be determined explicitly. This may be difficult or even impossible when dealing with partial differential equations. Of great importance is therefore a generalization, the *Goerisch method* [3], [4], [5], that may be used to overcome these problems. Nevertheless, the original Lehmann–Maehly method can easily be applied to a large class of ordinary differential equations; in [10] it is shown, that the method can be interpreted as a special application of the Rayleigh–Ritz method.

Inclusion Method. Let $(H, (\cdot|\cdot))$ be an infinite-dimensional Hilbert space with the inner product $(\cdot|\cdot)$ and the norm $\|\cdot\|$. Suppose that V is a dense subspace of H and that one has the inner product $[\cdot|\cdot]$ in V such that $(V, [\cdot|\cdot])$ is a Hilbert space (the norm in V is denoted by $\|\cdot\|_V$). The embedding $V \hookrightarrow H$ is assumed to be compact.

One can consider the right-definite eigenvalue problem

$$\begin{cases} \text{Find} & \lambda \in \mathbf{R} \text{ and } \varphi \in V,\ \varphi \neq 0, \\ \text{s.t.} & [\varphi|v] = \lambda(\varphi|v) \text{ for all } v \in V. \end{cases} \tag{1}$$

Problem (1) has a countable spectrum of eigenvalues, and the eigenvalues can be ordered by magnitude:

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots, \quad \lim_{j\to\infty} \lambda_j = \infty.$$

The Rayleigh–Ritz procedure for calculating upper bounds is a discretization of the Poincaré principle (cf. [9, Chapt. 22])

$$\lambda_j = \min_{\substack{E\subset V \\ \dim E = j}} \max_{\substack{u\in E \\ u\neq 0}} \frac{[u|u]}{(u|u)}, \quad j \in \mathbf{N}. \tag{2}$$

If the linearly independent trial functions

$$u_1, \ldots, u_n \in V, \quad n \in \mathbf{N},$$

are chosen, one can reduce (2) to the n-dimensional subspace V_n (the span of the chosen functions $\{u_1, \ldots, u_n\}$) and obtains the values

$$\Lambda_1^{[n]} \leq \cdots \leq \Lambda_n^{[n]},$$

which are upper bounds to the following λ_j:

$$\lambda_j \leq \Lambda_j^{[n]}, \quad j = 1, \ldots, n.$$

$\Lambda_j^{[n]}$ is called a *Rayleigh–Ritz bound* for λ_j. Now one forms the real $n \times n$-matrices

$$\begin{cases} A_0 := ((u_i|u_k))_{i,k=1,\ldots,n}, \\ A_1 := ([u_i|u_k])_{i,k=1,\ldots,n}, \end{cases} \tag{3}$$

the Rayleigh–Ritz bounds are the eigenvalues of the matrix eigenvalue problem

$$A_1 x = \Lambda^{[n]} A_0 x, \quad (\Lambda^{[n]}, x) \in \mathbf{R} \times \mathbf{R}^n. \tag{4}$$

The Rayleigh–Ritz bounds are monotonically decreasing in $n \in \mathbf{N}$.

The Lehmann–Goerisch procedure (see [6], [7], [4], [5], [3]) for calculating lower bounds can be understood as the discretization of a variational principle for characterizing the eigenvalues as well. This principle and a proof of the method is due to S. Zimmermann and U. Mertins [10].

Let $\rho \in \mathbf{R}$ be a spectral parameter such that for an $N \in \mathbf{N}$ the inequality

$$\lambda_N < \rho < \lambda_{N+1} \tag{5}$$

holds true. One expresses the first N eigenvalues in the form

$$\lambda_{N+1-i} = \rho + \frac{1}{\sigma_i}, \quad i = 1, \ldots, N$$

(assuming $\sigma_i < 0$). For $u \in V$, $w_u \in H$ denotes the uniquely determined solution of the equation

$$[u|v] = (w_u|v) \quad \text{for all } v \in V,$$

the following σ_i therefore are characterized by

$$\sigma_i = \inf_{\substack{E \subset V \\ \dim E = i}} \max_{\substack{u \in E \\ u \neq 0}} \frac{[u|u] - \rho(u|u)}{(w_u|w_u) - 2\rho[u|u] + \rho^2(u,u)}, \tag{6}$$

$i = 1, \ldots, N$. A negative upper bound for σ_i results in a lower bound for λ_{N+1-i}. In order to discretize (6), one determines $w_1, \ldots, w_n \in H$ such that

$$[u_i|v] = (w_i|v) \quad \text{for all } v \in V, \tag{7}$$

then one defines the matrix

$$A_2 := ((w_i|w_k))_{i,k=1,\ldots,n}, \tag{8}$$

and solves the matrix eigenvalue problem

$$(A_1 - \rho A_0)\, x = \tau \left(A_2 - 2\rho A_1 + \rho^2 A_0\right) x, \quad (\tau,\, x) \in \mathbf{R} \times \mathbf{R}^n. \tag{9}$$

If for $n \in \mathbf{N}$ the condition $\Lambda_N^{[n]} < \rho$ is fulfilled, then (9) has exactly N negative eigenvalues $\tau_1 \leq \cdots \leq \tau_N < 0 \leq \cdots \leq \tau_n$. These τ_i are upper bounds for our σ_i ($\sigma_i \leq \tau_i$, $i = 1, \ldots, N$). One obtains the lower bounds

$$\Lambda_j^{\rho[n]} := \rho + \frac{1}{\tau_{N+1-j}} \leq \lambda_j, \quad j = 1, \ldots, N. \tag{10}$$

This discretization (9), (10) is the Lehmann–Goerisch procedure. $\Lambda_j^{\rho[n]}$ is called a *Lehmann–Goerisch bound* for λ_j.

Numerical Example. The numerical example is the well known Mathieu equation. This equation has been considered by several authors, bounds for eigenvalues of the Mathieu equation can be found in [1], [9] and [3]. The eigenvalue problem reads as follows

$$-\Phi''(x) + s\cos^2(x)\Phi(x) = \lambda\Phi(x), \quad x \in \left[0, \frac{\pi}{2}\right],$$
$$\Phi'(0) = \Phi'\left(\frac{\pi}{2}\right) = 0,$$

where $s \in \mathbf{R}$, $s > 0$, is a parameter.

In order to treat this problem, the required quantities can be defined as follows: $I := (0, \pi/2)$,

$$H := L_2(I), \quad V := H^1(I).$$

The inner products $(\cdot,\cdot)$ and $[\cdot,\cdot]$ are given by

$$(f,g) := \int_0^{\pi/2} f(x)\, g(x)\, dx \quad \text{for all } f, g \in H,$$

$$[f,g] := \int_0^{\pi/2} \left(f'(x)g'(x) + s\cos^2(x) f(x) g(x)\right)\, dx$$

for all $f, g \in V$.

With this definition the inner product $[\cdot,\cdot]$ and the usual H^1 inner product are equivalent; the embedding $(V, [\cdot,\cdot]) \hookrightarrow (H, (\cdot,\cdot))$ is compact.

Now the eigenvalue problem

$$\begin{cases} \text{Find} & \lambda \in \mathbf{R} \text{ and } \varphi \in V,\ \varphi \neq 0 \\ \text{s.t.} & [\varphi|v] = \lambda(\varphi|v) \text{ for all } v \in V. \end{cases}$$

is equivalent to the Mathieu equation. The trial functions $v_k \in V$ are defined by

$$v_1(x) := 1, \qquad v_k(x) := \cos(2(k-1)x) \quad \text{for } x \in I, \quad k = 2, \ldots, n. \tag{11}$$

With these trial functions the Rayleigh–Ritz upper bounds $\Lambda_i^{[n]}$ (cf. (3), (4)) can be computed. For $n = 5$ one obtains

i	$\Lambda_i^{[5]}$
1	2.28404873592
2	8.4560567005
3	19.606719005
4	39.5439779
5	67.609198

The quality of these upper bounds can be increased by increasing n.

An application of the Lehmann–Goerisch procedure requires a spectral parameter ρ which is a rough eigenvalue bound (cf. (5)). For this aim the Mathieu equation is considered for $s = 0$. This is a second order problem with constant coefficients and can be solved in closed form. Its eigenvalues are $\widetilde{\lambda}_i = 4(i-1)^2$, $i \in \mathbf{N}$. From the comparison theorem (see [3]) one can see that the $\widetilde{\lambda}_i$ are lower bounds for the eigenvalues of the Mathieu equation with $s > 0$; this can be used to verify the left hand side inequality of (5), the right-hand side inequality can be examined by means of the Rayleigh–Ritz bounds. For $N = 4$ one obtains

$$\lambda_3 \leq \Lambda_3^{[n]} \leq 19.607 < \rho := \widetilde{\lambda}_4 = 36 < \lambda_4.$$

If s is increased dramatically, it may be impossible to satisfy (5). If this happens, one can link the eigenvalue problem under consideration and the comparison problem by a homotopy method (cf. [3]).

The next task is the determination of $w_i \in H$ such that (7) holds true. In general this is a problem, but for differential equations, where the right-hand side is the identity, one can proceed as follows: The operator on the left-hand side of the differential equation is denoted by M; then the trial functions v_i are chosen from $\mathcal{D}(M)$ (that means sufficiently smooth) such that all essential and natural boundary conditions are satisfied. Now $w_i := M\,v_i$ fulfills (7). For the Mathieu equation one can define

$$(Mf)(x) := -f''(x) + s\cos^2(x)f(x)$$

and

$$\widetilde{V} := \left\{ f \in H^2(I) \colon f'(0) = f'\left(\frac{\pi}{2}\right) = 0 \right\};$$

now it is easy to see that the v_i from (11) fulfill $v_i \in \widetilde{V}$ and $w_i := Mv_i$ can be used in (7), (8).

From the eigenvalues of the matrix eigenvalue problem (9) one obtains the following bounds:

i	$\Lambda_i^{\rho[5]}$	$\Lambda_i^{[5]}$
1	2.28404873561	2.28404873592
2	8.4560566942	8.4560567005
3	19.6067171	19.6067191

For an example with a system of ordinary differential equations see [2].

See also: **Hemivariational inequalities: Eigenvalue problems; Interval analysis: Eigenvalue bounds of interval matrices; Semidefinite programming and determinant maximization; αBB algorithm.**

References

[1] ALBRECHT, J.: 'Iterationsverfahren zur Berechnung der Eigenwerte der Mathieuschen Differentialgleichung', *Z. Angew. Math. Mechanics* **44** (1964), 453–458.

[2] BEHNKE, H.: 'A numerically rigorous proof of curve veering in an eigenvalue problem for differential equations', *Z. Anal. Anwend.* **15** (1996), 181–200.

[3] BEHNKE, H., AND GOERISCH, F.: 'Inclusions for eigenvalues of selfadjoint problems', in J. HERZBERGER (ed.): *Topics in validated computations*, Elsevier, 1994, pp. 277–322.

[4] GOERISCH, F.: 'Eine Verallgemeinerung eines Verfahrens von N.J. Lehmann zur Einschließung von Eigenwerten', *Wiss. Z. Techn. Univ. Dresden* **29** (1980), 429 - 431.

[5] GOERISCH, F., AND HAUNHORST, H.: 'Eigenwertschranken für Eigenwertaufgaben mit partiellen Differentialgleichungen', *Z. Angew. Math. Mechanics* **65**, no. 3 (1985), 129–135.

[6] LEHMANN, N.J.: 'Beiträge zur Lösung linearer Eigenwertprobleme I', *Z. Angew. Math. Mechanics* **29** (1949), 341–356.

[7] LEHMANN, N.J.: 'Beiträge zur Lösung linearer Eigenwertprobleme II', *Z. Angew. Math. Mechanics* **30** (1950), 1–16.

[8] MAEHLY, H.J.: 'Ein neues Verfahren zur genäherten Berechnung der Eigenwerte hermitescher Operatoren', *Helv. Phys. Acta* **25** (1952), 547–568.

[9] WEINSTEIN, A., AND STENGER, W.: *Methods of intermediate problems for eigenvalues*, Acad. Press, 1972.

[10] ZIMMERMANN, S., AND MERTINS, U.: 'Variational bounds to eigenvalues of self-adjoint problems with arbitrary spectrum', *Z. Anal. Anwend.* **14** (1995), 327–345.

H. Behnke
Inst. Math. TU Clausthal
Erzstr. 1, 38678 Clausthal, Germany
E-mail address: behnke@math.tu-clausthal.de

MSC2000: 49R50, 65L15, 65L60, 65G20, 65G30, 65G40

Key words and phrases: upper and lower bounds to eigenvalues, Rayleigh–Ritz method, Lehmann–Maehly method.

ENTROPY OPTIMIZATION: INTERIOR POINT METHODS, *interior point algorithms for entropy optimization, interior point methods for entropy optimization*

This section introduces the interior point approach to solving entropy optimization problems with linear constraints. In particular, we consider the following problem:

Program EL:

$$\begin{cases} \min & f(\mathbf{x}) \equiv \mathbf{c}^\top \mathbf{x} + \sum_{j=1}^{n} d_j x_j \ln x_j \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{O}, \end{cases} \tag{1}$$

where $\mathbf{c} \in \mathbf{R}^n$, $\mathbf{d} \in \mathbf{R}^n$, $\mathbf{d} > \mathbf{O}$, $\mathbf{b} \in \mathbf{R}^m$, $\mathbf{A}$ is an $(m \times n)$-matrix, $\mathbf{O}$ is an n-dimensional zero vector, and $0 \ln 0 \equiv 0$. When $\mathbf{c} = \mathbf{O}$ and $d_j = 1$, $j = 1, \ldots, n$, Program EL becomes a pure entropy optimization problem.

Denote the feasible region of Program EL by $F_p \equiv \{\mathbf{x} \in \mathbf{R}^n \colon \mathbf{A}\mathbf{x} = \mathbf{b}; \quad \mathbf{x} \geq \mathbf{O}\}$ and the (relative) interior of F_p by $F_p^0 \equiv \{\mathbf{x} \in \mathbf{R}^n \colon \mathbf{A}\mathbf{x} = \mathbf{b}; \mathbf{x} > \mathbf{O}\}$. An n-vector $\mathbf{x}$ is called an *interior solution* of Program EL if $\mathbf{x} \in F_p^0$. With these definitions, we have the following verifiable result:

LEMMA 1 If F_p is nonempty, then Program EL has a unique optimal solution. Moreover, if F_p has a nonempty interior, then the unique optimal solution is strictly positive. □

All *interior point methods*, including those to be discussed in this section, require the fundamental assumption that F_p has a nonempty interior, i.e., $F_p^0 \neq \emptyset$. A Lagrangian dual can be derived in the following manner. For all $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{y} \in \mathbf{R}^m$, and $\mathbf{z} \in \mathbf{R}_+^n \equiv \{\mathbf{x} \colon \mathbf{x} \in \mathbf{R}^n, \quad \mathbf{x} \geq \mathbf{O}\}$, define the following Lagrangian function:

$$L(\mathbf{x}, \mathbf{y}, \mathbf{z}) \equiv \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} d_j e(x_j) \tag{2}$$
$$- \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right) y_i - \sum_{j=1}^{n} z_j x_j,$$

where

$$e(x) \equiv \begin{cases} x \ln x & \text{if} \quad x \geq 0, \\ \infty & \text{if} \quad x < 0, \end{cases}$$

is a proper convex function with the set $\{x \colon x \in \mathbf{R}, x \geq 0\}$ being its effective domain [6]. The concept of proper convex function has often been used to simplify convex analysis. For details about the theory of using Lagrange multipliers for solving constrained optimization problems defined in terms of proper convex functions, the reader is referred to [6, Chap. 28].

Rearranging terms in (2) results in

$$L(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} d_j e(x_j)$$
$$+ \sum_{i=1}^{m} b_i y_i - \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i + z_j \right) x_j.$$

Considering the fact that $d_j > 0$ and the shape of the entropic function $x \ln x$, we know that, for any given $\mathbf{y} \in \mathbf{R}^m$ and $\mathbf{z} \in \mathbf{R}_+^n$, $L(\mathbf{x}, \mathbf{y}, \mathbf{z})$ achieves its unique minimum at $\mathbf{x}^* > \mathbf{O}$. Also, its first derivative at $\mathbf{x}^*$ vanishes. This implies

$$d_j \ln x_j^* - \sum_{i=1}^{m} a_{ij} y_i + c_j + d_j = z_j \geq 0. \tag{3}$$

Multiplying both sides of (3) by x_j^* and summing over j produces

$$\sum_{j=1}^{n} c_j x_j^* + \sum_{j=1}^{n} d_j x_j^* \ln x_j^*$$
$$- \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} y_i + z_j \right) x_j^*$$
$$= - \sum_{j=1}^{n} d_j x_j^*.$$

Consequently, for any $\mathbf{y} \in \mathbf{R}^m$ and $\mathbf{z} \in \mathbf{R}_+^n$,

$$L(\mathbf{x}^*, \mathbf{y}, \mathbf{z}) = \sum_{i=1}^{m} b_i y_i - \sum_{j=1}^{n} d_j x_j^*,$$

where $\mathbf{x}^*$ satisfies (3). Therefore, a Lagrangian *dual* of Program EL becomes

$$\begin{cases} \max\limits_{\substack{\mathbf{y}\in\mathbf{R}^m \\ \mathbf{z}\in\mathbf{R}^n_+}} & L(\mathbf{y},\mathbf{z}) \equiv \sum\limits_{i=1}^{m} b_i y_i - \sum\limits_{j=1}^{n} d_j x_j^* \\ \text{s.t.} & d_j \ln x_j^* - \sum\limits_{i=1}^{m} a_{ij} y_i + c_j + d_j = z_j, \\ & j = 1,\ldots,n. \end{cases}$$

This dual is equivalent to

Program DEL:

$$\begin{cases} \max\limits_{\substack{\mathbf{y}\in\mathbf{R}^m \\ \mathbf{O}<\mathbf{x}\in\mathbf{R}^n}} & L(\mathbf{x},\mathbf{y}) \equiv \sum\limits_{i=1}^{m} b_i y_i - \sum\limits_{j=1}^{n} d_j x_j \\ \text{s.t.} & d_j \ln x_j + c_j + d_j - \sum\limits_{i=1}^{m} a_{ij} y_i \geq 0, \\ & j = 1,\ldots,n. \end{cases} \tag{4}$$

Note that $\mathbf{x}$ is strictly positive because $\ln 0$ is not well-defined. However, if we define $\ln 0 = -\infty$, the domain of $\mathbf{x}$ in Program DEL can be replaced by $\{\mathbf{x}\colon \mathbf{x} \in \mathbf{R}^n, \mathbf{x} \geq \mathbf{O}\}$. Denote the excess vector $\nabla f(\mathbf{x}) - \mathbf{A}^\top \mathbf{y}$ by $\mathbf{s}$. The jth component of $\mathbf{s}$ is simply $d_j \ln x_j + c_j + d_j - \sum_{i=1}^m a_{ij} y_i$, which is the left-hand side of (4). Denote the feasible region of Program DEL by $F_d \equiv \{(\mathbf{x},\mathbf{y})\colon \nabla f(\mathbf{x}^*) - \mathbf{A}^\top \mathbf{y}^* \geq \mathbf{O}\}$, and assume that F_d has a nonempty interior.

We now derive the Karush–Kuhn–Tucker conditions for Program DEL. First, define, for all $\mathbf{u} \geq \mathbf{O}$, the following Lagrangian:

$$L'(\mathbf{x},\mathbf{y},\mathbf{u}) \equiv \sum_{i=1}^{m} b_i y_i - \sum_{j=1}^{n} d_j x_j + \sum_{j=1}^{n} u_j \left(d_j \ln x_j + c_j + d_j - \sum_{i=1}^{m} a_{ij} y_i \right).$$

Setting the partial derivatives with respect to y_i and x_j to zero gives

$$b_i - \sum_{j=1}^{n} a_{ij} u_j = 0, \quad i = 1,\ldots,m,$$

$$-d_j + \frac{u_j d_j}{x_j} = 0, \quad j = 1,\ldots,n. \tag{5}$$

Note that (5) is equivalent to $u_j = x_j$. Therefore, the KKT conditions for Program DEL become

i) There exists $\mathbf{x} \in \mathbf{R}^n$ such that $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{O}$. This can be viewed as the 'primal feasibility condition'.

ii) There exists $\mathbf{y} \in \mathbf{R}^m$ such that, together with $\mathbf{x}$, $d_j \ln x_j + c_j + d_j - \sum_{i=1}^m a_{ij} y_i \geq 0$ or $\nabla f(\mathbf{x}) - \mathbf{A}^\top \mathbf{y} \geq \mathbf{O}$. Similarly, this can be viewed as the 'dual feasibility condition'.

iii) For all $j = 1,\ldots,n$, $(d_j \ln x_j + c_j + d_j - \sum_{i=1}^m a_{ij} y_i) x_j = 0$. This can be viewed as the 'complementary slackness condition'.

Note that, by (5), the Lagrange multipliers associated with the constraints of Program DEL at its optimal solution happen to coincide with the $\mathbf{x}$-component of the optimal solution of Program DEL. This, together with the fact that the dual of Program DEL is Program EL, imply that the optimal solution of Program DEL contains the optimal solution of Program EL.

Also note that an alternative *dual* program can be defined by considering the following Lagrangian:

$$L''(\mathbf{x},\mathbf{y}) \equiv \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} d_j x_j \ln x_j - \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right) y_i,$$

for $\mathbf{x} \geq \mathbf{O}$ and $\mathbf{y} \in \mathbf{R}^m$. In this expression, no Lagrange multipliers are defined for the constraints $\mathbf{x} \geq \mathbf{O}$, and it leads to the following dual program:

$$\max_{\mathbf{y}\in\mathbf{R}^m} \sum_{i=1}^{m} b_i y_i - \sum_{j=1}^{n} d_j \exp\left\{ \frac{\sum_{i=1}^m a_{ij} y_i - c_j}{d_j} - 1 \right\}.$$

Since this dual program is *unconstrained*, any solution algorithm can be viewed as an interior point algorithm. For details about this approach and companion efficient solution *algorithms*, see [2].

In the rest of this section, we focus on the development of a *primal-dual* interior point algorithm [5]. Note that, to obtain the algorithm, Program DEL, rather than the unconstrained dual program, was used in [5]. The primal-dual interior point algorithm starts with an initial primal feasible solution $\mathbf{x}^0$ and an initial dual feasible solution $\mathbf{y}^0$. While the algorithm iterates, it maintains the primal and dual feasibility conditions and reduces the complementary slackness. In other words, the algorithm iterates from a pair of inte-

rior solutions $(\mathbf{x}^k, \mathbf{y}^k)$, with $\mathbf{A}\mathbf{x}^k = \mathbf{b}, \mathbf{x}^k > \mathbf{O}$ and $\mathbf{s}^k = \nabla f(\mathbf{x}^k) - \mathbf{A}^\top \mathbf{y}^k > \mathbf{O}$, to a new interior solution pair $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1})$ such that the complementary slackness is reduced from $\delta_k \equiv (\mathbf{x}^k)^\top \mathbf{s}^k$ to $\delta_{k+1} \equiv (\mathbf{x}^{k+1})^\top \mathbf{s}^{k+1}$. The algorithm terminates when $\delta_k \leq \epsilon$, for some given $\epsilon > 0$ (or when the difference between $f(\mathbf{x}^k)$ and the optimum is sufficiently small).

To describe the algorithm, we use the boldface upper-case letters $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{W}$ to denote the diagonal matrices formed by the components of vectors $\mathbf{x}$, $\mathbf{s}$, and $\mathbf{w}$, respectively. We also denote the vectors of all ones of appropriate dimensions by $\mathbf{e}$, the l_2 norm by $\|\cdot\|$, and the vector whose components are $\ln(x_j)$'s, $j = 1, \ldots, n$, by $\ln \mathbf{x}$.

Rather than dealing with the complementary slackness δ_k directly, the following *primal-dual potential function* [8]

$$\psi(\mathbf{x}, \mathbf{s}) = \rho \ln(\mathbf{x}^\top \mathbf{s}) - \sum_{j=1}^{n} \ln(x_j s_j),$$

where $\rho \geq n + \sqrt{n}$, can be used as a surrogate measure [5].

Given the initial solution pair, the potential of the associated complementary slackness can be calculated. Given the inaccuracy tolerance ϵ, a target potential can be calculated. Therefore, the amount of required potential reduction can be calculated. The primal-dual interior point algorithm, under proper conditions, will reduce the potential by a constant amount in each iteration.

Note that two different pairs of $(\mathbf{x}, \mathbf{s})$ that have the same complementary slackness measure may have different potentials. Therefore, to ensure that the target potential is sufficiently small, we need to find the minimum potential among all those $(\mathbf{x}, \mathbf{s})$ pairs such that $\mathbf{x}^\top \mathbf{s} = \epsilon$, or a lower bound of this minimum potential.

Rewrite the potential function as

$$\psi(\mathbf{x}, \mathbf{s}) = (\rho - n) \ln(\mathbf{x}^\top \mathbf{s}) - \sum_{j=1}^{n} \ln\left(\frac{x_j s_j}{\mathbf{x}^\top \mathbf{s}}\right).$$

Applying the geometric-arithmetic inequality results in

$$\prod_{j=1}^{n} \left(\frac{x_j s_j}{\mathbf{x}^\top \mathbf{s}}\right)^{1/n} \leq \frac{1}{n} \sum_{j=1}^{n} \left(\frac{x_j s_j}{\mathbf{x}^\top \mathbf{s}}\right) = \frac{1}{n}.$$

Taking the natural logarithm leads to

$$\frac{1}{n} \sum_{j=1}^{n} \ln\left(\frac{x_j s_j}{\mathbf{x}^\top \mathbf{s}}\right) \leq \ln\left(\frac{1}{n}\right) = -\ln n.$$

Consequently,

$$\sum_{j=1}^{n} \ln\left(\frac{x_j s_j}{\mathbf{x}^\top \mathbf{s}}\right) \leq -n \ln n.$$

Therefore, the target potential should be $(\rho - n) \ln \epsilon + n \ln n$. Given the potential associated with the initial solution, the exact amount of potential reduction is $\psi(\mathbf{x}^0, \mathbf{s}^0) - (\rho - n) \ln \epsilon - n \ln n$. Note that for a given inaccuracy tolerance ϵ, the target potential is indeed the minimum of all the potentials associated with all $(\mathbf{x}, \mathbf{s})$ pairs such that $\mathbf{x}^\top \mathbf{s} = \epsilon$. This is indicated by the tight geometric-arithmetic inequality.

Given the knowledge of how much potential reduction needs to be, if an algorithm reduces the potential by a constant amount in each iteration, then the complexity of the algorithm is $O(\psi(\mathbf{x}^0, \mathbf{s}^0) - (\rho - n) \ln \epsilon - n \ln n)$.

Assume that, in iteration k, we have a primal-dual feasible solution pair $(\mathbf{x}^k, \mathbf{y}^k)$ and the slack vector $\mathbf{s}^k \equiv \nabla f(\mathbf{x}^k) - \mathbf{A}^\top \mathbf{y}^k > \mathbf{O}$. Ideally, one would like to find $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1})$ such that the KKT conditions are met, i.e.,

$$\mathbf{A}\mathbf{x}^{k+1} = \mathbf{b}, \quad \mathbf{x}^{k+1} \geq \mathbf{O},$$
$$\nabla f(\mathbf{x}^{k+1}) - \mathbf{A}^\top \mathbf{y}^{k+1} \geq \mathbf{O},$$
$$\mathbf{X}^{k+1}(\nabla f(\mathbf{x}^{k+1}) - \mathbf{A}^\top \mathbf{y}^{k+1}) = \mathbf{O}.$$

Define

$$\triangle \mathbf{x} \equiv \mathbf{x}^{k+1} - \mathbf{x}^k,$$
$$\triangle \mathbf{y} \equiv \mathbf{y}^{k+1} - \mathbf{y}^k,$$
$$\triangle \mathbf{s} \equiv \mathbf{s}^{k+1} - \mathbf{s}^k,$$
$$\triangle \mathbf{X} \equiv \mathbf{X}^{k+1} - \mathbf{X}^k.$$

With these definitions, the conditions stated above become

$$\begin{aligned} &\mathbf{A}(\mathbf{x}^k + \triangle \mathbf{x}) = \mathbf{b} \quad , \quad \mathbf{x}^k + \triangle \mathbf{x} \geq \mathbf{O}, \\ &\nabla f(\mathbf{x}^k + \triangle \mathbf{x}) - \mathbf{A}^\top (\mathbf{y}^k + \triangle \mathbf{y}) \geq \mathbf{O}, \\ &(\mathbf{X}^k + \triangle \mathbf{X}) \\ &\times [\nabla f(\mathbf{x}^k + \triangle \mathbf{x}) - \mathbf{A}^\top (\mathbf{y}^k + \triangle \mathbf{y})] = \mathbf{O}. \end{aligned} \tag{6}$$

Note that quantity in the bracket of (6) is simply $\mathbf{s}^{k+1} = \mathbf{s}^k + \triangle \mathbf{s}$, where

$$\triangle \mathbf{s} \equiv \nabla f(\mathbf{x}^k + \triangle \mathbf{x}) - \nabla f(\mathbf{x}^k) - \mathbf{A}^\top \triangle \mathbf{y}.$$

Therefore, we have

$$(\mathbf{X}^k + \triangle\mathbf{X})(\mathbf{s}^k + \triangle\mathbf{s}) = \mathbf{O},$$

or

$$\mathbf{X}^k\mathbf{s}^k + \mathbf{X}^k\triangle\mathbf{s} + \triangle\mathbf{X}\mathbf{s}^k + \triangle\mathbf{X}\triangle\mathbf{s} = \mathbf{O},$$

or

$$\mathbf{X}^k\triangle\mathbf{s} + \mathbf{S}^k\triangle\mathbf{x} = -\triangle\mathbf{X}\triangle\mathbf{s} - \mathbf{X}^k\mathbf{s}^k.$$

Solving the equations

$$\mathbf{A}(\mathbf{x}^k + \triangle\mathbf{x}) = \mathbf{b},$$
$$\mathbf{X}^k\triangle\mathbf{s} + \mathbf{S}^k\triangle\mathbf{x} = -\triangle\mathbf{X}\triangle\mathbf{s} - \mathbf{X}^k\mathbf{s}^k,$$

subject to the condition

$$\nabla f(\mathbf{x}^k + \triangle\mathbf{x}) - \mathbf{A}^\top(\mathbf{y}^k + \triangle\mathbf{y}) \geq \mathbf{O}$$

is in general difficult.

Given $\mathbf{O} < \mathbf{x}^k \in F_p$, $\mathbf{s}^k = \nabla f(\mathbf{x}^k) - \mathbf{A}^\top\mathbf{y}^k > 0$, and $\delta^k = (\mathbf{x}^k)^\top\mathbf{s}^k$, the algorithm proposed in [4] solves the following system of nonlinear equations for $\triangle\mathbf{x}$ and $\triangle\mathbf{y}$:

$$\mathbf{X}^k\triangle\mathbf{s} + \mathbf{S}^k\triangle\mathbf{x} = \theta\mathbf{p}^k, \tag{7}$$
$$\mathbf{A}\triangle\mathbf{x} = \mathbf{O}, \tag{8}$$

where $\theta > 0$ is a constant to be specified later and

$$\mathbf{p}^k \equiv \frac{\delta^k}{\rho}\mathbf{e} - \mathbf{X}^k\mathbf{S}^k\mathbf{e},$$

and $n + \sqrt{n} \leq \rho < 2n$.

By choosing

$$\theta = \frac{\beta \min_j(\sqrt{x_j^k s_j^k})}{\|(\mathbf{X}^k\mathbf{S}^k)^{-0.5}\mathbf{p}^k\|}$$

for some $0 < \beta < 1$ yet to be determined, we obtain

$$\psi(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) \leq \psi(\mathbf{x}^k, \mathbf{s}^k) - \gamma$$

for a constant $\gamma > 0$. Let $C > 0$ be a real number. Choose β such that

$$0 < \beta < 1, \tag{9}$$
$$\beta(1 + C\beta) \leq \frac{1}{2}, \qquad 1 - C\beta \geq 0. \tag{10}$$

It can be shown [5] that, to reduce the potential by a constant amount in each iteration, solving a linear approximation of equations (7) and (8) can achieve the required accuracy.

Suppose that $n + \sqrt{n} \leq \rho < 2n$ and that $\triangle\mathbf{x}$ and $\triangle\mathbf{y}$ satisfy

$$\mathbf{A}\triangle\mathbf{x} = \mathbf{O}, \qquad \mathbf{X}^k\triangle\mathbf{s} + \mathbf{S}^k\triangle\mathbf{x} = \theta\mathbf{p}^k + \mathbf{z}^k,$$

$$\left\|\mathbf{z}^k\right\| \leq C\beta^2 \min(x_j^k s_j^k), \tag{11}$$

then

$$\psi(\mathbf{x}^k, \mathbf{s}^k) - \psi(\mathbf{x}^{k+1}, \mathbf{s}^{k+1}) > \gamma,$$

where $\gamma = (\sqrt{3}/2)\beta(1 - C\beta) - \beta^2(1 + C\beta)^2$.

Condition (11) can be achieved by solving the following set of linear equations:

$$\mathbf{X}^k(\nabla^2 f(\mathbf{x}^k)\triangle\mathbf{x} - \mathbf{A}^\top\triangle\mathbf{y}) + \mathbf{S}^k\triangle x = \theta\mathbf{p}^k, \tag{12}$$
$$\mathbf{A}\triangle\mathbf{x} = \mathbf{O}. \tag{13}$$

Note that the vector $\nabla^2 f(\mathbf{x}^k)\triangle\mathbf{x}$ replaces $\nabla f(\mathbf{x}^k + \triangle\mathbf{x}) - \nabla f(\mathbf{x}^k)$ of (7) and serves as a simple linear approximation. Equations (12) and (13) are key to the 'potential-reduction' primal-dual interior point algorithm.

Given an initial interior point solution, an interior point algorithm can be stated as follows.

Initialization:
Given an initial primal interior point solution $\mathbf{x}^0$ and an initial dual solution $\mathbf{y}^0$ such that $\mathbf{A}\mathbf{x}^0 = \mathbf{b}$, $\mathbf{x}^0 > 0$, and $\mathbf{s}^0 = \nabla f(\mathbf{x}^0) - \mathbf{A}^\top\mathbf{y}^0 > 0$,
calculate $\delta^0 = (\mathbf{x}^0)^\top\mathbf{s}^0$;
set $k \leftarrow 0$.
Iteration:
IF $\delta^k < \epsilon$, THEN STOP
ELSE
solve (12), (13) for $\triangle\mathbf{x}$ and $\triangle\mathbf{y}$;
set
$\mathbf{x}^{k+1} \equiv \mathbf{x}^k + \triangle\mathbf{x}$;
$\mathbf{y}^{k+1} \equiv \mathbf{y}^k + \triangle\mathbf{y}$;
$\mathbf{s}^{k+1} \equiv \nabla f(\mathbf{x}^{k+1}) - \mathbf{A}^\top\mathbf{y}^{k+1}$;
$\delta^{k+1} \equiv (\mathbf{x}^{k+1})^\top\mathbf{s}^{k+1}$;
reset $k \leftarrow k + 1$ for the next iteration.
END IF

With a standard procedure for obtaining an initial solution [1], the following theorem of *polynomial time convergence* was shown in [5].

THEOREM 2 Suppose that $\epsilon > 0$ and $2n \geq \rho \geq n + \sqrt{n}$. Then, in the kth iteration, $\mathbf{x}^k > \mathbf{O}$, $\mathbf{s}^k > \mathbf{O}$, and $\mathbf{x}^k$ and $\mathbf{y}^k$ are feasible for Programs EL and DEL. Moreover, the interior point algorithm terminates in at most $O(\psi(\mathbf{x}^0, \mathbf{s}^0) - (\rho - n)\ln\epsilon - n\ln n)$ iterations. □

It was also suggested that, in practical implementation, the stepsize can be set to $\overline{\eta}$ based on a line search such that $\overline{\eta} \equiv \arg\min_{\eta \geq 0} \psi(\mathbf{x}^k + \eta\triangle\mathbf{x}, \mathbf{s}^k + \eta\triangle\mathbf{s})$. With this stepsize, one can set $\mathbf{x}^{k+1} \equiv \mathbf{x}^k + \overline{\eta}\triangle\mathbf{x}$, $\mathbf{y}^{k+1} \equiv \mathbf{y}^k + \overline{\eta}\triangle\mathbf{y}$.

The search direction is a combination of a decent direction and a centering direction. To enable local quadratic convergence, a computable criterion was developed under which a pure Newton method for solving $\nabla f(\mathbf{x}) - \mathbf{A}^\top\mathbf{y} = \mathbf{O}$, $\mathbf{A}\mathbf{x} = \mathbf{b}$ (by solving the linear system of $\nabla^2 f(\mathbf{x}^k)\triangle\mathbf{x} - \mathbf{A}^\top\triangle\mathbf{y} = -\mathbf{s}^k$ and $\mathbf{A}\triangle\mathbf{x} = \mathbf{O}$) can be applied for the rest of the search process. Note that when $\mathbf{x}^k$ is close to the optimal solution, we have $\mathbf{x}^k$ being strictly positive, and therefore $\nabla f(\mathbf{x}) - \mathbf{A}^\top\mathbf{y}$ should be close to $\mathbf{O}$. Implementation of primal-dual interior point algorithms proposed in [5] is discussed in [3].

In addition to the 'potential-reduction' interior point method described above, the *'path following' interior point method*, which follows an ideal interior trajectory to reach an optimal solution, was proposed in [9], [7]. The convergence of the path following interior point method has been established. However, to the best of our knowledge, possible polynomial time convergence behavior remains an open issue.

See also: **Entropy optimization: Shannon measure of entropy and its properties; Jaynes' maximum entropy principle; Maximum entropy principle: Image reconstruction; Entropy optimization: Parameter estimation; Homogeneous selfdual methods for linear programming; Linear programming: Interior point methods; Linear programming: Karmarkar projective algorithm; Potential reduction methods for linear programming; Successive quadratic programming: Solution by active sets and interior point methods; Sequential quadratic programming: Interior point methods for distributed optimal control problems; Interior point methods for semidefinite programming.**

References

[1] FANG, S.-C., AND PUTHENPURA, S.: *Linear optimization and extensions: theory and algorithms*, Prentice-Hall, 1993.

[2] FANG, S.-C., RAJASEKERA, J.R., AND TSAO, H.-S.J.: *Entropy optimization and mathematical programming*, Kluwer Acad. Publ., 1997.

[3] HAN, C.-G., PARDALOS, P.M., AND YE, Y.: 'Implementation of interior-point algorithms for some entropy optimization problems', *Optim. and Software* **1** (1992), 71–80.

[4] KORTANEK, K.O., POTRA, F., AND YE, Y.: 'On some efficient interior point methods for nonlinear convex programming', *Linear Alg. & Its Appl.* **152** (1991), 169–189.

[5] POTRA, F., AND YE, Y.: 'A quadratically convergent polynomial algorithm for solving entropy optimization problems', *SIAM J. Optim.* **3** (1993), 843–860.

[6] ROCKAFELLAR, R.T.: *Convex analysis*, Princeton Univ. Press, 1970.

[7] SHEU, R.L., AND FANG, S.-C.: 'On the generalized path-following methods for linear programming', *Optim.* **30** (1994), 235–249.

[8] TODD, M.J., AND YE, Y.: 'A centered projective algorithm for linear programming', *Math. Oper. Res.* **15** (1990), 508–529.

[9] ZHU, J., AND YE, Y.: 'A path-following algorithm for a class of convex programming problems', *Working Paper College of Business Administration, Univ. Iowa*, no. 90-14 (1990).

Shu-Cherng Fang
North Carolina State Univ.
North Carolina, USA
E-mail address: fang@eos.ncsu.edu
H.-S. Jacob Tsao
San Jose State Univ.
San Jose, California, USA
E-mail address: jtsao@email.sjsu.edu

MSC2000: 94A17, 90C51, 90C25
Key words and phrases: entropy optimization, interior point methods, primal-dual algorithm, polynomial time convergence.

ENTROPY OPTIMIZATION: PARAMETER ESTIMATION

Introduction. Entropy optimization has been applied to problems in various fields of interest from thermodynamics to financial planning. In this context 'entropy' refers to the amount of uncertainty in a system, rather than the amount of disorder. A detailed definition of entropy can be found in [4].

One area of application, which has not received much attention in recent years, is that of param-

eter estimation. The estimation of parameters in semi-empirical mathematical models is a process which is important in many disciplines in the sciences and engineering. This article will focus on a few different areas of the parameter estimation problem which have been approached from an entropy perspective. *Jaynes' maximum entropy* principle allows for the estimation of parameters in a statistical distribution function by specification of the characteristic moments. This method can also be used to derive the principle of *maximum likelihood*, one of the most widely used parameter estimation approaches. Entropy principles have also been used to derive theoretical 'best estimators' for recursive parameter estimation schemes. These results can then be used to gauge the performance of various nonoptimal approaches. A final application involves the development of a measure which not only allows for the estimation of model parameters, but also simultaneously choosing the best mathematical form of the model.

Entropy Measures. In order to optimize entropy, one must possess some quantitative measure of the entropy of a given distribution. One such measure was developed by C.E. Shannon [8]. Shannon arrived at the function by postulating a set of properties which the measure should have, and then deriving a form which possesses those properties. For a probability distribution $\mathbf{p} = (p_1, \ldots, p_n)$, the function takes the form of:

$$S = -\sum_{i=1}^{n} p_i \ln p_i. \tag{1}$$

Shannon also proved that this function was unique for the postulated set of properties. Other researchers have postulated different sets of properties, but arrived at the same result [4].

Another measure of entropy, in this case the cross entropy or distance between two distributions, was presented by S. Kullback and R.A. Leibler [5]. For two given distributions, $\mathbf{p} = (p_1, \ldots, p_n)$, and $\mathbf{q} = (q_1, \ldots, q_n)$, the function takes the form:

$$I = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}. \tag{2}$$

It is assumed that when $q_i = 0$, the associated p_i also is zero and $0 \ln \frac{0}{0} \equiv 0$. This function is referred to as the *Kullback–Leibler measure of cross-entropy*.

Jaynes' Maximum Entropy for Continuous Distributions. Since most distributions encountered in practice are continuous in nature, Jaynes' principle of maximum entropy (*MaxEnt*), must first be extended to continuous distributions. This extension is straight forward and results in:

$$\begin{cases} \max & -\displaystyle\int_a^b f(x) \ln f(x)\, dx \\ \text{s.t.} & \displaystyle\int_a^b f(x)\, dx = 1 \\ & \displaystyle\int_a^b f(x) g_r(x)\, dx = a_r, \\ & r = 1, \ldots, m, \end{cases} \tag{3}$$

where $f(x)$ is a continuous probability density function from a to b. The Lagrange function takes the form of:

$$\begin{aligned} L \equiv & -\int_a^b f(x) \ln f(x)\, dx \\ & -(\lambda_0 - 1)\left[\int_a^b f(x) dx - 1\right] \\ & -\sum_{r=1}^{m} \lambda_r \left[\int_a^b f(x) g_r(x)\, dx - a_r\right]. \end{aligned} \tag{4}$$

Using the Euler–Lagrange equation the following expression results:

$$f(x) = \exp\left[-\lambda_0 - \lambda_1 g_1(x) \cdots - \lambda_m g_m(x)\right]. \tag{5}$$

A detailed discussion can be found in [4].

MaxEnt Estimation Method. The estimation of parameters in a statistical distribution using MaxEnt follows these steps:

1) Specify m characterizing functions, $g_1(x), \ldots, g_m(x)$.
2) Use MaxEnt to find $f(x)$, which is given by (5).

3) Find estimates of the values of the moment equations from the observed data set $\mathbf{x} = \{x_1, \ldots, x_n\}$ through the relationship:

$$\widehat{a}_r = \frac{1}{n}\left[g_r(x_1) + \cdots + g_r(x_n)\right]. \quad (6)$$

4) Determine estimates of the Lagrange multipliers, $\widehat{\lambda}_0, \ldots, \widehat{\lambda}_m$, from:

$$a_r = \frac{\int_a^b g_r(x) e^{-\widehat{\lambda}_1 g_1(x) - \cdots - \widehat{\lambda}_m g_m(x)}\, dx}{\int_a^b e^{-\widehat{\lambda}_1 g_1(x) - \cdots - \widehat{\lambda}_m g_m(x)}\, dx} \quad (7)$$

and

$$e^{\widehat{\lambda}_0} = \int_a^b e^{-\widehat{\lambda}_1 g_1(x) - \cdots - \widehat{\lambda}_m g_m(x)}\, dx. \quad (8)$$

5) The estimated function then takes the form:

$$f(x) = \exp\left[-\widehat{\lambda}_0 - \cdots - \widehat{\lambda}_m g_m(x)\right]. \quad (9)$$

Maximum Likelihood from MaxEnt. The principle of maximum likelihood has been widely used to estimate the parameters of both statistical distributions and semi-empirical models. Maximum likelihood assumes that information exists about a random variable in the form of an observation, $x_1, \ldots, x_n$, and a density function, $f(x; \theta_1, \ldots, \theta_m)$, unlike in MaxEnt where the forms of the characterizing moments are known. The approach seeks to maximize the likelihood that the given observations will occur given a set of parameters. If each observation is independent, then this 'likelihood' is defined as:

$$L(\mathbf{X}; \Theta) = \prod_{i=1}^{n} f(x_i|\Theta). \quad (10)$$

The log likelihood function is most often used:

$$\ln L(\mathbf{X}; \Theta) = \sum_{i=1}^{n} \ln f(x_i|\Theta). \quad (11)$$

The $\ln L$ is maximized to determine the optimal parameter estimates $\widehat{\Theta}$.

The same objective can also be derived using the concept of MaxEnt, even though the former predates the latter. The parameters need to be chosen such that the entropy which remains after the observed values are known is large as possible. This implies that the entropy of the observation itself has to be a minimum. The entropy is given by:

$$-\int_a^b f(x,\Theta) \ln f(x,\Theta)\, dx = \int_a^b \ln f(x,\Theta)\, dF. \quad (12)$$

The knowledge which is given by the observation is:

$$\begin{cases} F(x,\Theta) = 0 & \text{when } x < x_1 \\ F(x,\Theta) = \dfrac{1}{n} & \text{when } x_1 \le x < x_2 \\ \vdots & \vdots \\ F(x,\Theta) = \dfrac{r}{n} & \text{when } x_r \le x < x_{r+1} \\ \vdots & \vdots \\ F(x,\Theta) = 1 & \text{when } x_n \le x \end{cases}$$

where $F(x,\Theta)$ is the cumulative density. Thus the entropy of the sample is then written as:

$$-\frac{1}{n}\left[\ln f(x_1,\Theta) + \cdots + \ln f(x_n,\Theta)\right], \quad (13)$$

which is equal to :

$$-\frac{1}{n}\left[L(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)\right], \quad (14)$$

where L is the same as described by (11). Therefore to minimize the entropy of the sample the likelihood function must be maximized.

Recursive Parameter Estimation. The determination of the parameters of a dynamical system on-line is a key step in the implementation of a wide range of control schemes. The estimation procedure is conducted in a recursive fashion in which the estimates from the previous time step are combined with the current state observations to calculate a new set of parameter estimates. The analysis of the estimation process used is typically approached from a mean square error criterion. This method requires some assumptions about the error to be made and the form of the data processor to be restricted. H.L. Weidemann and E.B. Stear [9] presented an approach based on entropy concepts which has various benefits over the mean squared error method:

- The form of the optimal data processor is not constrained nor does it have to be known.
- Errors are not restricted to have a normal probability distribution.

- None of the operators in the system are required to be linear.

Before continuing with the analysis, various measures need to be defined. The entropy of a K-dimensional random vector $\mathbf{X}$ with the joint probability density function, $p_x(x_1, \ldots, x_k)$ is defined as:

$$H(\mathbf{X}) = -\int_{-\infty}^{\infty} p_x(\mathbf{X}) \ln p_x(\mathbf{X})\, d\mathbf{X}. \tag{15}$$

If R_x is the covariance matrix of the vector $\mathbf{X}$ then the following holds:

$$H(\mathbf{X}) \leq \frac{1}{2} \ln \left\{ (2\pi e)^K \det[R_x] \right\}. \tag{16}$$

When $\mathbf{X}$ is a Gaussian random vector then (16) holds as an equality. Another quantity which will be used in the analysis is referred to as the *mutual information* between $\mathbf{X}$ and $\mathbf{Y}$.

$$I(\mathbf{X};\mathbf{Y}) = \int_{-\infty}^{\infty} \int p_{xy}(\mathbf{X},\mathbf{Y}) \ln \frac{p_{xy}(\mathbf{X},\mathbf{Y})}{p_y(\mathbf{Y}) p_y(\mathbf{Y})}\, d\mathbf{X}\, d\mathbf{Y}. \tag{17}$$

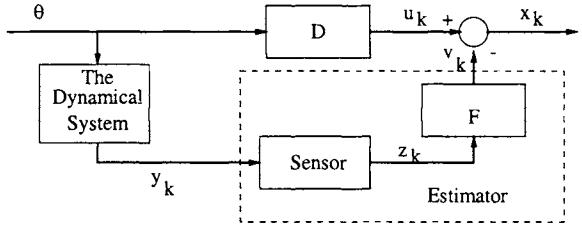

Fig.: Typical parameter estimator.

The object is to estimate a vector Θ of unknown parameters with the joint probability density function, $p_\theta(\theta_1, \ldots, \theta_m)$. The output of the dynamical model as a function of these parameters is expressed as $y_k(\theta_1, \ldots, \theta_m, k)$. These outputs are then measured by a sensor to produce $\{z_k\}$. These measurements are then used by the data processor F to produce an r-dimensional vector $\mathbf{V}$ which is an estimate of $D(\Theta)$. The estimation error is given by:

$$X = D(\Theta) - \mathbf{V} = D(\Theta) - F(\mathbf{Z}) = \mathbf{U} - \mathbf{V}. \tag{18}$$

Also, under certain conditions, the transform $D(\Theta)$ will possess a property that for any given random vector ξ, the following holds:

$$I(\xi;\Theta) = I(\xi; D(\Theta)), \tag{19}$$

or, in words, that $D(\Theta)$ preserves energy. When this does not hold, $I(\xi;\Theta) > I(\xi; D(\Theta))$.

The problem now is to determine the function $\widehat{F}$ which will produce an optimal estimator. The theoretically best function results in a minimum of the error entropy, defined to be $\widehat{H}_0$. The only constraint on the approach is that the mutual information, $I(\Theta;\mathbf{Z})$, must be known. With that the following can be stated:

- The minimum entropy of the error vector is given by:

$$H_0 = H(\mathbf{U}) - I(\mathbf{U};\mathbf{Z}). \tag{20}$$

- Minimizing the mutual information, $I(\mathbf{X};\mathbf{Z})$, is equivalent to the minimization of the error vector. This is achieved be choosing $F(\mathbf{Z})$ such that $\mathbf{Z}$ and $\mathbf{X}$ are independent.
- Whether or not $D(\Theta)$ preserves energy, the reduction in the processed parameter entropy, $H(\mathbf{U})$, is bounded above by $I(\Theta;\mathbf{Z})$, that is,

$$H(D(\Theta)) - H(\mathbf{X}) \leq I(\Theta;\mathbf{Z}), \tag{21}$$

  and the equality holds when $D(\Theta)$ preserves energy and the optimal processor, $\widehat{F}$, is used.

These three statements now make it possible to determine the best possible performance an estimator can achieve for a given system. The proofs of these statements and a simple example can be found in [9]. The extension of the theorems to the continuous time case is given in [6], and to the similar problem of state estimation in [7].

Parameter Estimation and Model Selection. For most problems of any physical significance the form of the model equations are not known with absolute certainty. In this lies the problem of not only estimated unknown parameters, but also determining the best fitting model. Given a set of N independent observations, $x_1, \ldots, x_N$, of a random variable from an unknown true distribution $g(x)$, the objective is to estimate this true distribution by choosing a member of a family of distributions given by $f(x|\Theta)$ where Θ is a vector of parameters. In order to accomplish this, the distance between the two distributions needs to be minimized. The

entropy of the true distribution is given by:

$$S(g;g) = \int g(x) \ln g(x)\, dx \tag{22}$$

while a measure of the *cross-entropy* is given by:

$$S(g; f(x|\Theta) = \int g(x) \ln f(x|\Theta)\, dx. \tag{23}$$

The Kullback–Leibler (K-L) measure is defined as:

$$\begin{aligned} I &= S(g;g) - S(g; f(x|\Theta) \\ &= \int g(x) \ln \frac{g(x)}{f(x|\Theta)}\, dx. \end{aligned} \tag{24}$$

Therefore the solution involves the minimization of the K-L measure [3].

Take the example of a family of possible distributions each one having a different number, k, of unknown parameters, Θ_k. These are denoted by $f(x|\Theta_k)$. The resulting form of the measure to choose the correct distribution is referred to as *Akaike's information criterion* (AIC) [1]:

$$\mathrm{AIC}(k) = -2 \ln L(\widehat{\Theta}_k) + 2k, \tag{25}$$

where $\ln L(\widehat{\Theta}_k)$ is the value of the log likelihood function with optimally determined parameters $\widehat{\Theta}_k$. It is proven in [3] that this result is obtained by the minimization of the K-L measure given by (24).

A secondary problem in the area of model selection, is sequential design of experiments. The concept of entropy has been applied to this problem in [2]. A total entropy criterion is developed which includes the uncertainty in the model selected as well as the uncertainty in the parameter values in each model. The use of this measure leads to a choice of an experiment for which the outcome is the most uncertain.

See also: **Entropy optimization: Shannon measure of entropy and its properties**; **Jaynes' maximum entropy principle**; **Maximum entropy principle: Image reconstruction**; **Entropy optimization: Interior point methods**.

References

[1] AKAIKE, H.: 'A new look at the statistical model identification', *IEEE Trans. Autom. Control* **19**, no. 6 (1974), 716–723.

[2] BORTH, D.M.: 'A total entropy criterion for the dual problem of model discrimination and parameter estimation', *J. Royal Statist. Soc. B* **37** (1975), 77–87.

[3] BOZDOGAN, H.: 'Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions', *Psychometrika* **52**, no. 3 (1987), 345–370.

[4] KAPUR, J.N., AND KESAVAN, H.K.: *Entropy optimization principles and applications*, Acad. Press, 1992.

[5] KULLBACK, S., AND LEIBLER, R.A.: 'On information and sufficiency', *Ann. Math. Statist.* **22** (1951), 79–86.

[6] MINAMIDE, N.: 'An extension of the entropy theorem for parameter estimation', *Inform. and Control* **53** (1982), 81–90.

[7] MINAMIDE, N., AND NIKIFORUK, P.N.: 'Conditional entropy theorem for recursive parameter estimation and its application to state estimation problems', *Internat. J. Syst. Sci.* **24**, no. 1 (1993), 53–63.

[8] SHANNON, C.E.: 'A mathematical theory of communication', *Bell System Techn. J.* **27** (1948), 379–423,623–659.

[9] WEIDEMANN, H.L., AND STEAR, E.B.: 'Entropy analysis of parameter estimation', *Inform. and Control* **14** (1969), 493–506.

William R. Esposito
Dept. Chemical Engin. Princeton Univ.
Princeton, NJ 08544-5263, USA
E-mail address: randy@titan.princeton.edu
Christodoulos A. Floudas
Dept. Chemical Engin. Princeton Univ.
Princeton, NJ 08544-5263, USA
E-mail address: floudas@titan.princeton.edu

MSC2000: 94A17, 62F10
Key words and phrases: maximum entropy, parameter estimation, model identification.

ENTROPY OPTIMIZATION: SHANNON MEASURE OF ENTROPY AND ITS PROPERTIES

The word *entropy* originated in the literature on thermodynamics around 1865 in Germany and was coined by R. Clausius [4] to represent a measure of the amount of energy in a thermodynamic system as a function of the temperature of the system and the heat that enters the system. Clausius wanted a word similar to the German word *energie* (i.e., energy) and found it in the Greek word $\eta\tau\rho o\pi\eta$, which means transformation [1]. The word entropy had belonged to the domain of physics until 1948 when C.E. Shannon, while developing his theory of communication at Bell Laboratories, used the term to represent a measure of information after a suggestion made by J. von Neumann. Shannon wanted a word to describe his newly found *measure of un-*

certainty and sought Von Neumann's advice. Von Neumann's reasoning to Shannon [25] was that: 'No one really understands entropy. Therefore, if you know what you mean by it and you use it when you are in an argument, you will win every time.'

Whatever the reason for the name is, the concept of Shannon's entropy has penetrated a wide range of disciplines, including statistical mechanics [12], thermodynamics [12], statistical inference [24], business and finance [5], nonlinear spectral analysis [21], image reconstruction [3], transportation and regional planning [26], queueing theory [10], information theory [20], [9], statistics [17], econometrics [8], and linear and nonlinear programming [6], [7].

The concept of entropy is closely tied to the concept of *uncertainty embedded in a probability distribution*. In fact, entropy can be defined as a measure of probabilistic uncertainty. For example, suppose the probability distribution for the outcome of a coin-toss experiment is $(0.0001, 0.9999)$, with 0.0001 being the probability of having a tail. One is likely to notice that there is much more 'certainty' than 'uncertainty' about the outcome of this experiment and hence about the probability distribution. In fact, one is almost certain that the outcome will be a head. If, on the other hand, the probability distribution governing that same experiment were $(0.5, 0.5)$, one would realize that there is much less 'certainty' and much more 'uncertainty,' when compared to the previous distribution. Generalizing this observation to the case of n possible outcomes, we conclude that the uniform distribution has the highest uncertainty out of all possible probability distributions. This implies that, if one had to choose a probability distribution for a chance experiment without any prior knowledge about that distribution, it would seem reasonable to pick the uniform distribution. This is because one would have no reason to choose any other and because that distribution maximizes the 'uncertainty' of the outcome. This is called *Laplace's principle of insufficient reasoning* [15]. Note that we are able to justify this principle without resorting to a rigorous definition of 'uncertainty.' However, this principle is inadequate when one has some prior knowledge about the distribution. Suppose, for example, that one knows some particular moments of the distribution, e.g., the expected value. In this case, a mathematical definition of 'uncertainty' is crucial. This is the case where Shannon's measure of uncertainty, or Shannon's entropy, plays an indispensable role [20].

To define *entropy*, Shannon proposed some axioms that he thought any measure of uncertainty should satisfy and deduced a unique function, up to a multiplicative constant, that satisfies them. It turned out that this function actually possesses many more desirable properties. In later years, many researchers modified and replaced some of his axioms in an attempt to simplify the reasoning. However, they all deduced that same function.

We first focus on finite-dimensional entropy, i.e., Shannon's entropy defined on discrete probability distributions that have a finite number of outcomes (or states). Let $\mathbf{p} \equiv (p_1, \ldots, p_n)^\top$ be a probability distribution associated with n possible outcomes, denoted by $\mathbf{x} \equiv (x_1, \ldots, x_n)^\top$, of an experiment. Denote its entropy by $S_n(\mathbf{p})$. Among those defining axioms, J.N. Kapur and H.K. Kesavan stated the following [15]:

1) $S_n(\mathbf{p})$ should depend on all the p_j's, $j = 1, \ldots, n$.

2) $S_n(\mathbf{p})$ should be a continuous function of p_j, $j = 1, \ldots, n$.

3) $S_n(\mathbf{p})$ should be permutationally symmetric. In other words, if the p_j's are merely permuted, then $S_n(\mathbf{p})$ should remain the same.

4) $S_n(1/n, \ldots, 1/n)$ should be a monotonically increasing function of n.

5) $S_n(p_1, \ldots, p_n) = S_{n-1}(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2)\, S_2(p_1/(p_1 + p_2), p_2/(p_1 + p_2))$.

Properties 1, 2 and 3 are obvious. Property 4 states that the maximum uncertainty of a probability distribution should increase as the number of possible outcomes increases. Property 5 is the least obvious but states that the uncertainty of a probability distribution is the sum of the uncertainty of the probability distribution that combines two of the outcomes and the uncertainty of the probability distribution consisting of only those two outcomes adjusted by the combined probabilities of the two outcomes.

It turns out that the unique family of functions that satisfy the defining axioms has the form $S_n(\mathbf{p}) = -k\sum_{j=1}^{n} p_j \ln p_j$, where k is a positive constant, ln represents the natural logarithmic function, and $0 \ln 0 \equiv 0$ [15]. Shannon chose $-\sum_{j=1}^{n} p_j \ln p_j$ to represent his concept of entropy [20]. Among its many other desirable properties, we state the following:

6) Shannon's measure is nonnegative and concave in $p_1, \ldots, p_n$.

7) The measure does not change with the inclusion of a zero-probability outcome.

8) The entropy of a probability distribution representing a completely certain outcome is 0, and the entropy of any probability distribution representing uncertain outcomes is positive.

9) Given any fixed number of outcomes, the maximum possible entropy is that of the uniform distribution.

10) The entropy of the joint distribution of two independent distributions is the sum of the individual entropies.

11) The entropy of the joint distribution of two dependent distributions is no greater than the sum of the two individual entropies.

Property 6 is desirable because it is much easier to maximize a concave function than a nonconcave one. Properties 7 and 8 are appealing because a zero-probability outcome contributes nothing to uncertainty, and neither does a completely certain outcome. Property 9 was discussed earlier. Properties 10 and 11 state that joining two distributions does not affect the entropy, if they are independent, and may actually reduce the entropy, if they are dependent.

Shannon's entropy was originally defined for a probability distribution over a finite sample space, i.e., a finite number of possible outcomes, and can be interpreted as a measure of uncertainty of the probability distribution. It has subsequently been defined for general discrete and continuous random vectors. It has been rigorously proved that Shannon's entropy is the unique measure of uncertainty (up to a multiplicative constant) of a finite probability distribution that satisfies a set of axioms considered necessary for any reasonable measure of uncertainty [19], [20], [16]. The concept of entropy, when extended for probability distributions defined on a countably infinite sample space, takes the form of $-\sum_{j=1}^{\infty} p_j \ln p_j$. It can still be viewed as a measure of uncertainty but such an interpretation does not enjoy the same degree of mathematical rigor as its finite-sample-space counterpart. When the concept is extended for continuous probability distributions, it is defined to be $-\int p(x) \ln p(x)\, dx$. However, it can no longer be interpreted as a measure of uncertainty at all [9], [11]. Rather, it can only be viewed as a measure of relative uncertainty [15].

Note that, with Shannon's entropy as the measure of uncertainty, in the absence of any prior information about the underlying probability distribution, the best course of action suggested by the principle of insufficient reasoning is to choose the uniform distribution because it possesses maximum uncertainty. Given the knowledge of some moments of the underlying distribution, the same reasoning leads to the following principle:

- Out of all possible distributions that are consistent with the moment constraints, choose the one that has maximum entropy.

This principle was proposed by E.T. Jaynes ([15, Chapter 2]), and has been known as the *principle of maximum entropy* or Jaynes' *maximum entropy principle*. It has often been abbreviated as *MaxEnt* in literature.

Let X be a random variable with n possible outcomes $\{x_1, \ldots, x_n\}$ and $\mathbf{p} \equiv (p_1, \ldots, p_n)^\top$ be a vector consisting of corresponding probabilities. Suppose that $g_1(X), \ldots, g_m(X)$ are m functions of X with known expected values $a_1, \ldots, a_m$, respectively. The principle of maximum entropy leads to the following mathematical *optimization* problem:

$$\begin{cases} \max & H_1(\mathbf{p}) = -\displaystyle\sum_{j=1}^{n} p_j \ln p_j \\ \text{s.t.} & \displaystyle\sum_{j=1}^{n} p_j\, g_i(x_j) = a_i, \quad i = 1, \ldots, m, \\ & \displaystyle\sum_{j=1}^{n} p_j = 1, \\ & p_j \geq 0, \quad j = 1, \ldots, n. \end{cases}$$

This is a convex programming problem with linear constraints. The nonnegativity constraints are not binding for the optimal solution $\mathbf{p}^*$ because each p_j^* can be expressed as an exponential function in terms of the Lagrange multipliers associated with the equality constraints. Note that, in the absence of the moment constraints, the solution to the problem is the uniform probability distribution, whose entropy is $\ln n$. As such, the maximum entropy principle can be viewed as an extension of the Laplace's principle of insufficient reasoning. The distribution selected under the maximum entropy principle has also been interpreted as one that is the 'most probable' in the sense that the maximum entropy distribution coincides with the frequency distribution that can be realized in the greatest number of ways [13]. An explanation of this linkage in the context of the well-known application of entropy maximization in transportation planning can be found in [7].

Recall that the above discussion was originally motivated by the task of choosing a probability distribution among those that are consistent with some given moments. Now, in addition to the moment constraints, suppose that we have an *a priori* probability distribution $\mathbf{p}^0$ that we think our probability distribution $\mathbf{p}$ should be close to. In fact, in the absence of the moment constraints, we would like to choose $\mathbf{p}^0$ for $\mathbf{p}$ because it is clearly the closest to $\mathbf{p}^0$. However, in the presence of some moment constraints which $\mathbf{p}^0$ does not satisfy, we need a precise definition of 'closeness' or 'deviation'. In other words, we need to define some sort of deviation or, more precisely, *'directed divergence'* [15] on the space of discrete probability distributions where the distribution is chosen from. Note that we deliberately avoid calling this measure a 'distance'. This is because a distance measure should be symmetric and should satisfy the triangular inequality, but these two properties are not important in this context. In fact, we can be content with a 'one-way (asymmetric) deviation measure', $D(\mathbf{p}, \mathbf{p}^0)$, from $\mathbf{p}$ to $\mathbf{p}^0$. If a 'one-way deviation measure' from $\mathbf{p}$ to $\mathbf{p}^0$ is not satisfactory, one can consider using a symmetric measure defined as the sum of $D(\mathbf{p}, \mathbf{p}^0)$ and $D(\mathbf{p}^0, \mathbf{p})$. What is desirable for this 'directed divergence' measure includes the following properties:

1) $D(\mathbf{p}, \mathbf{p}^0)$ should be nonnegative for all $\mathbf{p}$ and $\mathbf{p}^0$.
2) $D(\mathbf{p}, \mathbf{p}^0) = 0$ if and only if $\mathbf{p} = \mathbf{p}^0$.
3) $D(\mathbf{p}, \mathbf{p}^0)$ should be a convex function of $p_1, \ldots, p_n$.
4) When $D(\mathbf{p}, \mathbf{p}^0)$ is minimized subject to moment constraints but without the explicit presence of the nonnegativity constraints, the resulting p_j's should be nonnegative.

Property 1 is desirable for any such measure of deviation. If property 2 were not satisfied, then it would be possible to choose a vector $\mathbf{p}$ that has a zero directed divergence from $\mathbf{p}^0$, i.e., one that is as 'close' to $\mathbf{p}^0$ as $\mathbf{p}^0$ itself, but differs from $\mathbf{p}^0$. Property 3 makes minimizing the measure much simpler, and property 4 spares us from explicitly considering n nonnegativity constraints. Fortunately, there are many measures that satisfy these properties. We may even be able to find one that satisfies the triangular inequality. But, simplicity of the measure is also desirable. The simplest and most important of those measures is the Kullback–Leibler measure ([15, Chapt. 4]), defined as $D(\mathbf{p}, \mathbf{p}^0) = \sum_{j=1}^n p_j \ln(p_j/p_j^0)$, with the convention that, whenever p_j^0 is 0, p_j is set to 0 and $0 \ln(0/0)$ is defined to be 0. This measure is also known as the *cross-entropy*, *relative entropy*, *directed divergence* or *expected weight of evidence* of $\mathbf{p}$ with respect to $\mathbf{p}^0$. A. Hobson [11] provided an axiomatic characterization of *cross-entropy*. He interpreted $D(\mathbf{p}, \mathbf{p}^0)$ as the 'information in $\mathbf{p}$ relative to $\mathbf{p}^0$', and showed that the only function $I(\mathbf{p}, \mathbf{p}^0)$ satisfying the following five properties has the form of $k \sum_{j=1}^n p_j \ln(p_j/p_j^0)$, where k is a positive constant:

5) $I(\mathbf{p}, \mathbf{p}^0)$ is a continuous function of $\mathbf{p}$ and $\mathbf{p}^0$.
6) $I(\mathbf{p}, \mathbf{p}^0)$ is permutationally symmetric, i.e., the measure does not change if the pairs of (p_j, p_j^0) are permuted among themselves.
7) $I(\mathbf{p}, \mathbf{p}) = 0$.
8) For any pair of integers n and n_0 such that $n_0 \geq n > 0$, $I(1/n, \ldots, 1/n, 0, \ldots, 0; 1/n_0, \ldots, 1/n_0)$ is an increasing function of n_0 and a decreasing function of n, where $I(1/n, \ldots, 1/n, 0, \ldots, 0; 1/n_0, \ldots, 1/n_0)$ de-

notes the information obtained when the number of equally likely possibilities is reduced from n_0 to n.

9)

$$I(p_1,\ldots,p_n;p_1^0,\ldots,p_n^0) = I(q_1,q_2;q_1^0,q_2^0) + q_1 I\left(\frac{p_1}{q_1},\ldots,\frac{p_r}{q_1};\frac{p_1^0}{q_1^0},\ldots,\frac{p_r^0}{q_1^0}\right) + q_2 I\left(\frac{p_{r+1}}{q_2},\ldots,\frac{p_n}{q_2};\frac{p_{r+1}^0}{q_2^0},\ldots,\frac{p_n^0}{q_2^0}\right),$$

where $1 \le r \le n$, $q_1 \equiv p_1 + \cdots + p_r$, $q_2 \equiv p_{r+1} + \cdots + p_n$, $q_1^0 \equiv p_1^0 + \cdots + p_r^0$, $q_2^0 \equiv p_{r+1}^0 + \cdots + p_n^0$.

Property 8 says, for example, that the information obtained upon reducing the number of equally likely sides on a die from 6 to 3 is greater than the information obtained upon reducing the number from 6 to 4. Property 9 says that one may give information about the outcome associated with the random event either by specifying the probabilities $p_1,\ldots,p_n$ directly, or by specifying the probabilities q_1 and q_2 first and then specifying the conditional probabilities p_i/q_1 and p_i/q_2.

In addition to the nine properties discussed above, we state the following desirable properties for cross-entropy:

10) $D(\mathbf{p},\mathbf{p}^0)$ is convex in both $\mathbf{p}$ and $\mathbf{p}^0$.

11) $D(\mathbf{p},\mathbf{p}^0)$ is not symmetric.

12) If $\mathbf{p}$ and $\mathbf{q}$ are independent and $\mathbf{r}$ and $\mathbf{s}$ are also independent, then $D(\mathbf{p}*\mathbf{q},\mathbf{r}*\mathbf{s}) = D(\mathbf{p},\mathbf{r}) + D(\mathbf{q},\mathbf{s})$, where $*$ denotes the convolution operation between two independent distributions.

13) In general, the triangular inequality does not hold. But, if distribution $\mathbf{p}$ minimizes $D(\mathbf{p},\mathbf{p}^0)$ subject to some moment constraints and $\mathbf{q}$ is any other distribution that satisfies those same constraints, then $D(\mathbf{q},\mathbf{p}^0) = D(\mathbf{q},\mathbf{p}) + D(\mathbf{p},\mathbf{p}^0)$. Thus, in this special case, the triangular inequality holds, but as an equality.

Kullback and Leibler's cross-entropy was also originally defined for probability distributions with a finite sample space and can be interpreted as a measure of deviation of one probability distribution from another. It has been extended subsequently for distributions defined on countably infinite and continuous sample spaces. The corresponding forms become $\sum_{j=1}^{\infty} p_j \ln(p_j/p_j^0)$ and $\int p(x)\ln(p(x)/p^0(x))\,dx$, respectively. It has also been derived rigorously as the unique measure of deviation of one probability distribution from another that satisfies a set of axioms considered as necessity for any reasonable measure of deviation, for both finite probability distributions [11] and continuous distributions [14]. Cross-entropy for probability distributions with countably infinite sample space can be viewed and has been used as a measure of deviation, although the justification is not as strong as their finite-sample-space and continuous counterparts.

With cross-entropy interpreted as a measure of 'deviation', the Kullback–Leibler's *principle of minimum cross-entropy*, or *MinxEnt*, can be stated as follows [15]:

> Out of all possible distributions that are consistent with the moment constraints, choose the one that minimizes the cross-entropy with respect to the given *a priori* distribution.

Mathematically, we consider the following *optimization* problem:

$$\begin{cases} \min & H_2(\mathbf{p}) = \displaystyle\sum_{j=1}^{n} p_j \ln \frac{p_j}{p_j^0} \\ \text{s.t.} & \displaystyle\sum_{j=1}^{n} p_j\, g_i(x_j) = a_i, \quad i = 1,\ldots,m, \\ & \displaystyle\sum_{j=1}^{n} p_j = 1, \\ & p_j \ge 0, \quad j = 1,\ldots,n. \end{cases}$$

Note that the nonnegativity constraints are not binding, for the same reason as in the MaxEnt problem. For a detailed discussion of the properties of MinxEnt, the reader is referred to [23].

Note that, if there is no a priori information, then one may use the uniform distribution, denoted by $\mathbf{u}$, as the a priori distribution. In this case, $D(\mathbf{p},\mathbf{p}^0) = D(\mathbf{p},\mathbf{u}) = \sum_{j=1}^{n} p_j \ln(p_j/(1/n)) = \ln n + \sum_{j=1}^{n} p_j \ln p_j$. Since minimizing $\sum_{j=1}^{n} p_j \ln p_j$ is equivalent to maximizing $-\sum_{j=1}^{n} p_j \ln p_j$, minimizing the cross-entropy with respect to the uniform distribution is equiv-

alent to maximizing entropy and, therefore, MaxEnt is a special case of MinxEnt. These two principles can now be combined into a general principle:

> Out of all probability distributions satisfying the given moment constraints, choose the distribution that minimizes the cross-entropy with respect to the given a priori distribution and, in the absence of it, choose the distribution that minimizes the cross-entropy with respect to the uniform distribution.

Both the MaxEnt and MinxEnt principles for selecting finite-sample-space probability distributions and the MinxEnt principle for selecting continuous probability distributions can be *axiomatically derived* [22]. Under four consistency axioms, it was shown that the two principles are uniquely correct methods for inductive inference when new information is given in the form of expected values. Many well-known and widely used distributions, including the normal, gamma and geometric distributions, can actually be derived as solutions to some MaxEnt or MinxEnt problems [15].

The maximum entropy principle has also been shown to be a dual principle of the *maximum likelihood principle* for the exponential family of probability distributions in the sense that a dual problem to the linearly constrained entropy maximization problem is equivalent to the problem of maximizing a likelihood function with respect to the parameters of an exponential family [2]. This principle has also been shown to be related to the *Bayesian parameter estimation* problem [7]. *Duality theory* and major mathematical *algorithms* for solving finite-dimensional MaxEnt or MinxEnt problems can be found in [7] and the references therein.

See also: **Jaynes' maximum entropy principle**; **Maximum entropy principle: Image reconstruction**; **Entropy optimization: Parameter estimation**; **Entropy optimization: Interior point methods**; **Optimization in medical imaging**.

References

[1] BAIERLEIN, R.: 'How entropy got its name', *Amer. J. Phys.* **60** (1992), 1151.

[2] BEN-TAL, A., TEBOULLE, M., AND CHARNES, A.: 'The role of duality in optimization problems involving entropy functionals with applications to information theory', *J. Optim. Th. Appl.* **58** (1988), 209–223.

[3] BURCH, S.F., GULL, S.F., AND SKILLING, J.K.: 'Image restoration by a powerful maximum entropy method', *Computer Vision, Graphics, and Image Processing* **23** (1983), 113–128.

[4] CLAUSIUS, R.: 'Ueber Verschiedene fur die Anwendung Bequeme Formen der Hauptgleichungen der Mechanischen Warmetheorie', *Ann. Physik und Chemie* **125** (1865), 353–400.

[5] COZZOLINO, J.M., AND ZAHNER, M.J.: 'The maximum entropy distribution of the future market price of a stock', *Oper. Res.* **21** (1973), 1200–1211.

[6] ERLANDER, S.: 'Entropy in linear programming', *Math. Program.* **21** (1981), 137–151.

[7] FANG, S.-C., RAJASEKERA, J.R., AND TSAO, H.-S.J.: *Entropy optimization and mathematical programming*, Kluwer Acad. Publ., 1997.

[8] GOLAN, A., JUDGE, G., AND MILLER, D.: *Maximum entropy econometrics: robust estimation with limited data*, Wiley, 1996.

[9] GUIASU, S.: *Information theory with applications*, McGraw-Hill, 1977.

[10] GUIASU, S.: 'Maximum entropy condition in queueing theory', *J. Oper. Res. Soc.* **37** (1986), 293–301.

[11] HOBSON, A.: *Concepts in statistical mechanics*, Gordon and Breach, 1987.

[12] JAYNES, E.T.: 'Information theory and statistical mechanics II', *Phys. Rev.* **108** (1957), 171–190.

[13] JAYNES, E.T.: 'Prior probabilities', *IEEE Trans. Syst., Sci. Cybern.* **SSC-4** (1968), 227–241.

[14] JOHNSON, R.W.: 'Axiomatic characterization of the directed divergence and their linear combinations', *IEEE Trans. Inform. Theory* **25** (1979), 709–716.

[15] KAPUR, J.N., AND KESAVAN, H.K.: *Entropy optimization principles with applications*, Acad. Press, 1992.

[16] KHINCHIN, A.I.: *Mathematical foundations of information theory*, Dover, 1957.

[17] KULLBACK, S.: *Information theory and statistics*, Dover, 1968.

[18] SCOTT, C.H., AND JEFFERSON, T.R.: 'Entropy maximizing models of residential location via geometric programming', *Geographical Anal.* **9** (1977), 181–187.

[19] SHANNON, C.E.: 'A mathematical theory of communication', *Bell System Techn. J.* **27** (1948), 379–423; 623–656.

[20] SHANNON, C.E., AND WEAVER, W.: *The mathematical theory of communication*, Univ. Illinois Press, 1962.

[21] SHORE, J.E.: 'Minimum cross-entropy spectral analysis', *IEEE Trans. Acoustics, Speech and Signal Processing* **29** (1981), 230–237.

[22] SHORE, J.E., AND JOHNSON, R.W.: 'Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy', *IEEE Trans. Inform.*

Theory **26** (1980), 26–37.

[23] SHORE, J.E., AND JOHNSON, R.W.: 'Properties of cross-entropy minimization', *IEEE Trans. Inform. Theory* **27** (1981), 472–482.

[24] TRIBUS, M.: *Rational descriptions, decisions, and designs*, Pergamon, 1969.

[25] TRIBUS, M.: 'An engineer looks at Bayes', in G.J. ERICKSON AND C.R. SMITH (eds.): *Maximum-Entropy and Bayesian Methods in Sci. and Engineering: Foundations*, Vol. 1, Kluwer Acad. Publ., 1988, pp. 31–52.

[26] WILSON, A.G.: *Entropy in urban and regional modeling*, Pion, 1970.

Shu-Cherng Fang
North Carolina State Univ.
North Carolina, USA
E-mail address: fang@eos.ncsu.edu

H.-S. Jacob Tsao
San Jose State Univ.
San Jose, California, USA
E-mail address: jtsao@email.sjsu.edu

MSC2000: 94A17, 90C25

Key words and phrases: entropy, cross-entropy, maximum entropy principle, minimum cross-entropy principle.

EQUALITY-CONSTRAINED NONLINEAR PROGRAMMING: KKT NECESSARY OPTIMALITY CONDITIONS, *EQNLP*

An equality-constrained nonlinear programming problem may be posed in the form

$$\begin{cases} \min\limits_{x \in \mathbf{R}^n} & f(x) \\ \text{subject to} & c(x) = 0, \end{cases} \tag{1}$$

where f is a real-valued nonlinear function and c is an m-vector of real-valued nonlinear functions with ith component $c_i(x)$, $i = 1, \ldots, m$. Normally, with the term *equality-constrained nonlinear programming problem* is meant a problem of the form (1) where f and c are sufficiently smooth, at least continuously differentiable. This will be assumed throughout this discussion, with the gradient of $f(x)$ denoted by $g(x)$ and the $m \times n$ Jacobian of $c(x)$ denoted by $J(x)$.

Of fundamental importance for equality-constrained optimization problems are the *first order necessary optimality conditions*. These conditions are often referred to as the *KKT necessary optimality conditions*, or more briefly, the *KKT conditions*. The KKT conditions state that if x^* is a local minimizer to (1) that satisfies a certain *constraint qualification*, then there exists an m-dimensional vector λ^* such that

$$g(x^*) - J(x^*)^\top \lambda^* = 0,$$
$$c(x^*) = 0.$$

The vector λ^* is usually referred to as the vector of *Lagrange multipliers*. For equality-constrained problems, the KKT conditions are attributed to J.L. Lagrange, and hence 'classical'. The acronym KKT arises from the more general results on inequality-constrained problems provided by W. Karush [3], H.W. Kuhn and A.W. Tucker [4], [5].

For an equality-constrained problem, the KKT conditions state that x^* must be feasible, i.e., $c(x^*) = 0$; and that the gradient must have zero projection onto the null space of the constraint gradients, i.e., there exists a λ^* such that $g(x^*) = J(x^*)^\top \lambda^*$. In the case of linear equality constraints, i.e., $c(x) = Ax - b$ for some $(m \times n)$-matrix A and m-vector b, it follows that if x^* is feasible, then $x^* + p$ is feasible if and only if $Ap = 0$. Hence, in this situation, if x^* is a local minimizer, it must hold that $g(x^*)^\top p = 0$ for all p such that $Ap = 0$. But this is equivalent to the existence of a λ^* such that $g(x^*) = A^\top \lambda^*$. Consequently, in the case of linear constraints, the KKT conditions are necessary for x^* to be a local minimizer to problem (1). Constraint qualifications essentially ensure that the linearization of c at x^* provided by $J(x^*)$ adequately describes c in a neighborhood of x^*. A constraint qualification which is frequently used is that $J(x^*)$ has rank m, i.e., that the gradients of the constraints are linearly independent at x^*. The related *Fritz John necessary optimality conditions* are valid without any constraint qualification.

The KKT conditions are of fundamental importance, not only from a theoretical point of view, but also algorithms for solving equality-constrained nonlinear programming problems are often based on finding a solution to the KKT conditions. In general, the KKT conditions are not sufficient for x^* to be a local minimizer, but second order optimality conditions need be considered. However, if c is affine and f is a convex function on the feasible region, then the KKT conditions are sufficient for x^* to be a global minimizer. Detailed discussions on optimality conditions can

be found in textbooks on nonlinear programming, e.g., [1], [2], [6].

As a simple example, consider the two-dimensional problem where $f(x) = x_1$ and $c(x) = (x_1^2 + x_2^2 - 1)/2$. Then, the KKT conditions have two solutions: $\widetilde{x} = (1,0)^\top$ together with $\widetilde{\lambda} = 1$, and $\widehat{x} = (-1,0)^\top$ together with $\widehat{\lambda} = -1$. However, only $\widehat{x}$ is a local minimizer (and in fact also a global minimizer).

See also: **Inequality-constrained nonlinear optimization; Second order optimality conditions for nonlinear optimization; Lagrangian duality: Basics; Saddle point theory and optimality conditions; First order constraint qualifications; Second order constraint qualifications; Kuhn–Tucker optimality conditions; Rosen's method, global convergence, and Powell's conjecture; Relaxation in projection methods; SSC minimization algorithms; SSC minimization algorithms for nonsmooth and stochastic optimization.**

References

[1] BAZARAA, M.S., SHERALI, H.D., AND SHETTY, C.M.: *Nonlinear programming: Theory and algorithms*, second ed., Wiley, 1993.

[2] BERTSEKAS, D.P.: *Nonlinear programming*, Athena Sci., 1995.

[3] KARUSH, W.: 'Minima of functions of several variables with inequalities as side constraints', *Master's Thesis Dept. Math. Univ. Chicago* (1939).

[4] KUHN, H.W.: 'Nonlinear programming: A historical note', in J.K. LENSTRA, A.H.G. RINNOOY KAN, AND A. SCHRIJVER (eds.): *History of Mathematical Programming: A Collection of Personal Reminiscences*, Elsevier, 1991, pp. 82–96.

[5] KUHN, H.W., AND TUCKER, A.W.: 'Nonlinear programming', in J. NEYMAN (ed.): *Proc. Second Berkeley Symp. Math. Stat. Probab.*, Univ. Calif. Press, 1951, pp. 481–492.

[6] NASH, S.G., AND SOFER, A.: *Linear and nonlinear programming*, McGraw–Hill, 1996.

Anders Forsgren
Royal Inst. Technol. (KTH)
Stockholm, Sweden
E-mail address: andersf@math.kth.se

MSC2000: 49M37, 65K05, 90C30

Key words and phrases: equality-constrained optimization, KKT necessary optimality conditions.

EQUILIBRIUM NETWORKS

Many complex systems in which agents compete for scarce resources on a network, be it a physical one, as in the case of congested urban transportation systems, or an abstract one, as in the case of certain economic and financial problems, can be formulated and studied as *network equilibrium* problems. Applications of network equilibrium problems are common in many disciplines, in particular, in operations research and management science and in economics and engineering (cf. [17], [10]).

Network equilibrium problems as opposed to network optimization problems involve competition among the agents or users of the network system. Moreover, network equilibrium problems are governed by an underlying behavioral principle as to the behavior of the agents as well as the equilibrium conditions. For example, in congested urban transportation systems in which users seek to determine their cost minimizing routes of travel, the equilibrium conditions, due to J.G. Wardrop [23] (see also [2] and [8]), state that, in equilibrium all used paths connecting an origin/destination pair will have minimal and equal user travel costs. On the other hand, in the case of spatial price equilibrium patterns one seeks to determine the commodity production, trade, and consumption pattern satisfying the equilibrium condition, due to S. Enke [9] and P.A. Samuelson [20], that expresses that there will be trade between a pair of spatially separated supply and demand markets provided the supply price of the commodity at the supply market plus the unit cost of transportation associated with shipping the commodity is equal to the demand price of the commodity at the demand market; if the supply price plus the transportation cost exceed the demand price, then there will be no trade between this pair of supply and demand markets.

M.J. Beckmann, C.B. McGuire, and C.B. Winsten [2] initiated the systematic study of network equilibrium problems in the general setting of traffic networks and demonstrated that the equilibrium flow pattern satisfying the traffic network equilibrium conditions (see also [23]), under certain symmetry assumptions on the underlying

functions, could be reformulated as the solution to an optimization problem. Samuelson [20], following [9], had made a similar connection but in the more specialized context of spatial price equilibrium problems on networks that were bipartite.

M.J. Smith [22] later proposed an alternative formulation of traffic network equilibrium conditions which were then identified by S.C. Dafermos [3] to satisfy a finite-dimensional variational inequality problem. This connection allowed for the relaxation of the symmetry assumption and, consequently, for the construction of more realistic models (cf. [17], [21], and the references therein).

Other network equilibrium applications whose study and understanding have benefited from this methodology (cf. [10], [14], [17], [19]), include: spatial price equilibrium problems (see, e.g., [11], [15]), oligopolistic market equilibrium problems ([7], [12], [13]), migration equilibrium problems (cf. [16], [18]), and general economic equilibrium problems (cf. [5]).

Here we present two examples of network equilibrium problems for illustrative purposes with the first example being a multimodal/multiclass transportation network equilibrium problem in which the network is a physical one whereas the second problem is a multiclass migration equilibrium problem which is isomorphic to a specially structure multiclass traffic network equilibrium problem.

Additional background, models and applications, qualitative results, as well as computational procedures and references can be found in [17] and [10].

A Multimodal Traffic Network Equilibrium Model. We now present a multimodal traffic network equilibrium model (cf. [3], [4], [6]). The model is a fixed demand model in that the demands associated with traveling between the origin/destination pairs are assumed known. See [17] for additional background, as well as elastic demand traffic network equilibrium models and other network equilibrium problems.

Consider a general network $N = [G, A]$, where N denotes the set of nodes and A the set of directed links. Let $a, b, c, \ldots$ denote the links, $p, q, \ldots$ the paths. Assume that there are J origin/destination (O/D) pairs, with a typical O/D pair denoted by w, and n modes of transportation on the network with typical modes denoted by $i, j, \ldots$.

The flow on a link a generated by mode i is denoted by f_a^i, and the user cost associated with traveling by mode i on link a is denoted by c_a^i. Group the link flows into a column vector $f \in \mathbf{R}^{nL}$, where L is the number of links in the network. Group the link costs into a row vector $c \in \mathbf{R}^{nL}$. Assume that the user cost on a link and a particular mode may, in general, depend upon the flows of every mode on every link in the network, that is,

$$c = c(f),$$

where c is a known smooth function.

The travel demand of users of mode i traveling between O/D pair w is denoted by d_w^i and the travel disutility associated with traveling between this O/D pair using the mode is denoted by λ_w^i. Group the demands into a vector $d \in \mathbf{R}^{nJ}$.

The flow on path p due to mode i is denoted by x_p^i. Group the path flows into a column vector $x \in \mathbf{R}^{nQ}$, where Q denotes the number of paths in the network.

The conservation of flow equations are as follows. The demand for a mode and O/D pair must be equal to the sum of the flows of the mode on the paths joining the O/D pair, that is,

$$d_w^i = \sum_{p \in P_w} x_p^i, \qquad \forall i, \quad \forall w,$$

where P_w denotes the set of paths connecting w.

A nonnegative path flow vector x which satisfies the demand constraint is termed feasible. Moreover, we must have that

$$f_a^i = \sum_p x_p^i \delta_{ap},$$

that is, for each mode, the link load associated with a mode is equal to the sum of the path flows of that mode on paths that utilize that link.

A user traveling on path p using mode i incurs a user (or personal) travel cost C_p^i satisfying

$$C_p^i = \sum_a c_a^i \delta_{ap},$$

in other words, the cost on a path p due to mode i is equal to the sum of the link costs of links comprising that path and using that mode.

The traffic network equilibrium conditions are given below.

DEFINITION 1 (*multimodal traffic network equilibrium*) ([2], [3], [4]) A link load pattern f^* satisfying the feasibility conditions is an equilibrium pattern, if, once established, no user has any incentive to alter his travel arrangements. This state is characterized by the following equilibrium conditions, which must hold for every mode i, every O/D pair w, and every path $p \in P_w$:

$$C_p^i \begin{cases} = \lambda_w^i & \text{if} \quad x_p^{i\,*} > 0, \\ \geq \lambda_w^i & \text{if} \quad x_p^{i\,*} = 0, \end{cases}$$

where λ_w^i is the equilibrium travel disutility associated with the O/D pair w and mode i. □

We now define the feasible set K as

$$K \equiv \left\{ f: \begin{array}{c} \exists x \geq 0, \\ \text{the demand constraints and} \\ \text{the link load constraints hold} \end{array} \right\}.$$

One can verify (see [3]) that the variational inequality governing equilibrium conditions for this model would be given as in the subsequent theorem.

THEOREM 2 (*variational inequality formulation*) A vector $f^* \in K$ is an equilibrium pattern, if and only if, it satisfies the variational inequality problem

$$\langle c(f^*), f - f^* \rangle \geq 0, \qquad \forall f \in K.$$

□

Note that this variational inequality is in link loads. One can also derive a variational inequality problem in path flows (see also [1], [4], [17]). Existence of an equilibrium f^* follows from the standard theory of variational inequalities (cf. [14]) solely from the assumption that c is continuous, since the feasible set K is now compact.

In the special case where the symmetry condition

$$\left[\frac{\partial c_a^i}{\partial f_b^j} = \frac{\partial c_b^j}{\partial f_a^i} \right], \qquad \forall i, j; a, b,$$

holds, then the variational inequality problem can be reformulated as the solution to an optimization problem. This symmetry assumption, however, is not expected to hold in most applications. Consequently, the variational inequality problem which is the more general problem formulation is needed. For example, the symmetry condition essentially says that the flow on link b due to mode j should affect the cost of mode i on link a in the same manner that the flow of mode i on link a affects the cost on link b and mode j. In the case of a single mode problem, the symmetry condition would imply that the cost on link a is affected by the flow on link b in the same manner as the cost on link b is affected by the flow on link a.

A Migration Network Equilibrium Model. Human migration is a topic that has been studied not only by economists, but also by demographers, sociologists, and geographers. Here a model of human migration is described, which is shown to have a simple, abstract network structure in which the links correspond to locations and the flows on the links to populations of a particular class at the particular location. Hence, the model is isomorphic to the traffic network equilibrium problem just described on a network with special structure. For additional details, see [16], [17], [18].

Assume a closed economy in which there are n locations, typically denoted by i, and J classes, typically denoted by k. Assume further that the attractiveness of any location i as perceived by class k is represented by a utility u_i^k. Let $\overline{p}^k$ denote the fixed and known population of class k in the economy, and let p_i^k denote the population of class k at location i. Group the utilities into a row vector $u \in \mathbf{R}^{Jn}$ and the populations into a column vector $p \in \mathbf{R}^{Jn}$. Assume no births and no deaths in the economy.

The conservation of flow equation for each class k is given by

$$\overline{p}^k = \sum_{i=1}^{n} p_i^k,$$

where $p_i^k \geq 0$, $k = 1, \ldots, J$; $i = 1, \ldots, n$. Let

$$K \equiv \left\{ p: \begin{array}{c} p \geq 0 \text{ and satisfy the} \\ \text{conservation of flow equation} \end{array} \right\}.$$

The *conservation of flow equation* expresses that the population of each class k must be conserved in the economy.

DEFINITION 3 (*migration equilibrium*) Assume that the migrants are rational and that migration will continue until no individual of any class has any incentive to move since a unilateral decision will no longer yield an increase in the utility. Mathematically, hence, a multiclass population vector $p^* \in K$ is said to be in equilibrium if for each class k, $k = 1, \ldots, J$:

$$u_i^k \begin{cases} = \lambda^k & \text{if } p_i^{k*} > 0 \\ \leq \lambda^k & \text{if } p_i^{k*} = 0. \end{cases}$$

□

The equilibrium conditions express that for a given class k only those locations i with maximal utility will have a positive population volume of the class. Moreover, the utilities for a given class are equilibrated across the locations.

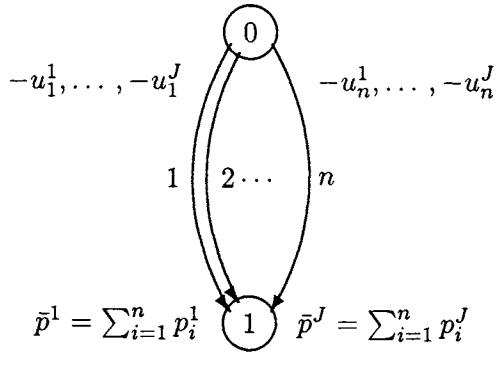

Fig. 1: Network equilibrium formulation of a multiclass migration equilibrium model.

We now discuss the utility functions. Assume that, in general, the utility associated with a particular location as perceived by a particular class, may depend upon the population associated with every class and every location, that is, assume that

$$u = u(p).$$

Note that in allowing the utility to depend upon the populations of the classes, we are using populations as a proxy for amenities associated with a particular location. Such a utility function can also model the negative externalities associated with overpopulation, such as congestion, increased crime, competition for scarce resources, etc.

As illustrated in [17], the above migration model is equivalent to a network equilibrium model with a single origin/destination pair and fixed demands. Indeed, one can make the identification as follows. Construct a network consisting of two nodes, an origin node 0 and a destination node 1, and n links connecting the origin node to the destination node. Associate with each link i, J costs: $-u_i^1, \ldots, u_i^J$, and link flows represented by $p_i^1, \ldots, p_i^J$. This model is, hence, equivalent to a multimodal traffic network equilibrium model with fixed demand for each mode, consisting of a single origin/destination pair, and J paths connecting the O/D pair. Note that one can make J copies of the network, in which case, each ith network will correspond to class i with the cost functions on the links defined accordingly. This identification enables us to immediately write down the following:

THEOREM 4 (*variational inequality formulation*) A population pattern $p^* \in K$ is in equilibrium, if and only if it satisfies the variational inequality problem:

$$\langle -u(p^*), p - p^* \rangle \geq 0, \qquad \forall p \in K.$$

□

Existence of an equilibrium then follows from the standard theory of variational inequalities, since the feasible set K is compact, assuming that the utility functions are continuous. Uniqueness of the equilibrium population pattern also follows from the standard theory provided that the $-u$ function is strictly monotone. The interpretation of this monotonicity condition in the context of applications is that condition implies that the utility associated with a given class and location is expected to be a decreasing function of the population of that class at that location.

See also: **Spatial price equilibrium**; **Traffic network equilibrium**; **Oligopolistic market equilibrium**; **Walrasian price equilibrium**; **Financial equilibrium**; **Generalized monotonicity: Applications to variational inequalities and equilibrium problems**; **Minimum cost flow problem**; **Nonconvex network**

flow problems; Network location: Covering problems; Maximum flow problem; Shortest path tree algorithms; Steiner tree problems; Survivable networks; Directed tree networks; Dynamic traffic networks; Auction algorithms; Piecewise linear network flow problems; Communication network assignment problem; Generalized networks; Evacuation networks; Network design problems; Stochastic network problems: Massively parallel solution.

References

[1] AASHTIANI, H.Z., AND MAGNANTI, T.L.: 'Equilibria on a congested transportation network', *SIAM J. Alg. Discrete Meth.* **2** (1981), 213–226.

[2] BECKMANN, M.J., MCGUIRE, C.B., AND WINSTEN, C.B.: *Studies in the economics of transportation*, Yale Univ. Press, 1956.

[3] DAFERMOS, S.: 'Traffic equilibrium and variational inequalities', *Transport. Sci.* **14** (1980), 43–54.

[4] DAFERMOS, S.: 'The general multimodal network equilibrium problem with elastic demand', *Networks* **14** (1982), 43–54.

[5] DAFERMOS, S.: 'Exchange price equilibria and variational inequalities', *Math. Program.* **46** (1990), 391–402.

[6] DAFERMOS, S., AND NAGURNEY, A.: 'Stability and sensitivity analysis for the general network equilibrium-travel choice model', in J. VOLMULLER AND R. HAMERSLAG (eds.): *Proc. 9th Internat. Symp. Transportation and Traffic Theory*, VNU Sci. Press, 1984, pp. 217–234.

[7] DAFERMOS, S., AND NAGURNEY, A.: 'Oligopolistic and competitive behavior of spatially separated markets', *Regional Sci. and Urban Economics* **17** (1987), 245–254.

[8] DAFERMOS, S., AND SPARROW, F.T.: 'The traffic assignment problem for a general network', *J. Res. Nat. Bureau Standards* **73B** (1969), 91–118.

[9] ENKE, S.: 'Equilibrium among spatially separated markets: solution by electronic analogue', *Econometrica* **10** (1951), 40–47.

[10] FLORIAN, M., AND HEARN, D.: 'Network equilibrium models and algorithms', in M.O. BALL, T.L. MAGNANTI, C.L. MONMA, AND G.L. NEMHAUSER (eds.): *Network Routing*, Vol. 8 of *Handbook Oper. Res. and Management Sci.*, Elsevier, 1995, pp. 485–550.

[11] FLORIAN, M., AND LOS, M.: 'A new look at static spatial price equilibrium models', *Regional Sci. and Urban Economics* **12** (1982), 579–597.

[12] GABAY, D., AND MOULIN, H.: 'On the uniqueness and stability of Nash-equilibria in noncooperative games', in A. BENSOUSSAN, P. KLEINDORFER, AND C.S. TAPIERO (eds.): *Applied Stochastic Control in Econometrics and Management Sci.*, North-Holland, 1980, pp. 271–294.

[13] HAURIE, A., AND MARCOTTE, P.: 'On the relationship between Nash–Cournot and Wardrop equilibria', *Networks* **15** (1985), 295–308.

[14] KINDERLEHER, D., AND STAMPACCHIA, G.: *An introduction to variational inequalities and their applications*, Acad. Press, 1980.

[15] NAGURNEY, A.: 'Computational comparisons of spatial price equilibrium methods', *J. Reg. Sci.* **27** (1987), 55–76.

[16] NAGURNEY, A.: 'Migration equilibrium and variational inequalities', *Economics Lett.* **31** (1989), 109–112.

[17] NAGURNEY, A.: *Network economics: A variational inequality approach*, second ed., Kluwer Acad. Publ., 1999.

[18] NAGURNEY, A., PAN, J., AND ZHAO, L.: 'Human migration networks', *Europ. J. Oper. Res.* (1991).

[19] PATRIKSSON, M.: *The traffic assignment problem*, VSP, 1994.

[20] SAMUELSON, P.A.: 'A spatial price equilibrium and linear programming', *Amer. Economic Rev.* **42** (1952), 283–303.

[21] SHEFFI, Y.: *Urban transportation networks*, Prentice-Hall, 1985.

[22] SMITH, M.J.: 'The existence, uniqueness, and stability of traffic equilibria', *Transport. Res.* **13B** (1979), 259–304.

[23] WARDROP, J.G.: 'Some theoretical aspects of road traffic research', *Proc. Inst. Civil Engineers* **II** (1952), 325–378.

Anna Nagurney
Univ. Massachusetts
Amherst, Massachusetts 01003, USA
E-mail address: nagurney@gbfin.umass.edu

MSC2000: 90C30
Key words and phrases: traffic network equilibrium, spatial price equilibrium, migration equilibrium, multimodal networks, multiclass migration.

EQUIVALENCE BETWEEN NONLINEAR COMPLEMENTARITY PROBLEM AND FIXED POINT PROBLEM

Complementarity theory is a new domain of applied mathematics strongly related to Linear Analysis, Nonlinear Analysis, Topology, Variational Inequalities Theory, Ordered Topological Vector Spaces, Numerical Analysis etc. The main goal in this theory is the study of complementarity problems. It is well known that complementarity problems encompass a variety of practical problems arising in: Optimization, Structural Mechan-

ics, Elasticity, Economics etc. [8]. The relation between the general nonlinear complementarity problem and the fixed point problem it seems to be remarkable. The main aim of this article is the study of this relation.

Preliminaries. Let E, E^* be a pair of real locally convex spaces. The space E^* can be the topological dual of E. Let $\langle \cdot, \cdot \rangle$ be a bilinear form on $E \times E^*$ satisfying the separation axioms:

$s_1)$ $\langle x_0, y \rangle = 0$ for all $y \in E^*$ implies $x_0 = 0$;

$s_2)$ $\langle x, y_0 \rangle = 0$ for all $x \in E$ implies $y_0 = 0$.

The triplet $(E, E^*, \langle \cdot, \cdot \rangle)$ is called a *dual system* or a *duality* (denoted by $\langle E, E^* \rangle$). In practical problems, the space E can be a Banach space and E^* its topological dual and $\langle x, y \rangle = y(x)$ for all $x \in E$ and $y \in E^*$. When E is a Hilbert space $(H, \langle \cdot, \cdot \rangle)$ or the Euclidean space $(\mathbf{R}^n, \langle \cdot, \cdot \rangle)$ we have that H^* (respectively, $(\mathbf{R}^n)^*$) is isomorphic to H (respectively, to $\mathbf{R}^n$). Let $\langle E, E^* \rangle$ be a dual system of locally convex spaces. Denote by $\mathbf{K}$ a *pointed convex cone* in E, i.e., a subset of E satisfying the following properties:

1) $\mathbf{K} + \mathbf{K} \subseteq \mathbf{K}$;

2) $\lambda \mathbf{K} \subseteq \mathbf{K}$ for all $\lambda \in \mathbf{R}_+$ (the set of nonnegative real numbers); and

3) $\mathbf{K} \cap (-\mathbf{K}) = \{0\}$.

The closed convex cone

$$K^* = \{y \in E^* \colon \langle x, y \rangle \geq 0 \text{ for all } x \in \mathbf{K}\}$$

is called the *dual* of $\mathbf{K}$. The polar of $\mathbf{K}$ is $\mathbf{K}^0 = -\mathbf{K}^*$. Given the pointed convex cone $\mathbf{K} \subset E$ we denote by $\leq$ the ordering defined on E by $\mathbf{K}$, i.e., $x \leq y$ if and only if $y - x \mathbf{K}$. In some situations, E is a *vector lattice* with respect to this ordering, i.e., for every pair $x, y \in E$ there exist $\inf(x, y)$ (denoted by $x \wedge y$) and $\sup(x, y)$ (denoted by $x \vee y$). We say that the bilinear form $\langle \cdot, \cdot \rangle$ is $\mathbf{K}$-local if $\langle x, y \rangle = 0$, whenever $x, y \in \mathbf{K}$ and $x \wedge y = 0$.

Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space and $\mathbf{K} \subset H$ a closed pointed convex cone. It is known that the projection operator onto $\mathbf{K}$, denoted by $P_{\mathbf{K}}$ is well defined [20] and for every $x \in H$, $P_{\mathbf{K}}(x)$ is the unique element of $\mathbf{K}$ satisfying $\|x - P_{\mathbf{K}}(x)\| = \min_{y \in \mathbf{K}} \|x - y\|$.

THEOREM 1 For every $x \in H$, $P_{\mathbf{K}}(x)$ is characterized by the following property:

1) $\langle P_{\mathbf{K}}(x) - x, y \rangle \geq 0$ for all $y \in \mathbf{K}$;

2) $\langle P_{\mathbf{K}}(x) - x, x \rangle = 0$.

□

PROOF. A proof of this theorem is in [20]. □

Very useful is also the following classical *Moreau's theorem*:

THEOREM 2 If $\mathbf{K} \subset H$ is a closed convex cone and $x, y, z \in H$, then the following statements are equivalent:

i) $z = x + y$, $x \in \mathbf{K}$, $y \in \mathbf{K}^0$ and $\langle x, y \rangle = 0$;

ii) $x = P_{\mathbf{K}}(z)$ and $y = P_{\mathbf{K}^0}(z)$.

□

PROOF. For the proof the reader is referred to [16]. □

We say that the closed pointed convex cone $\mathbf{K} \subset H$ is *isotone projection* if and only if, for every $x, y \in H$ such that $y - x \in \mathbf{K}$ we have $P_{\mathbf{K}}(y) - P_{\mathbf{K}}(x) \in \mathbf{K}$. This remarkable class of cones has been studied in several papers (see for example [13]). We say that a closed pointed convex cone $\mathbf{K} \subset H$ is a *Galerkin cone* if there exists a family of convex subcones $\{\mathbf{K}_n\}_{n \in \mathbf{N}}$ of $\mathbf{K}$ such that:

1) $\mathbf{K}_n$ is a locally compact cone, for every $n \in \mathbf{N}$;

2) if $n \leq m$, then $\mathbf{K}_n \subseteq \mathbf{K}_m$;

3) $\mathbf{K} = \overline{\cup_{n \in \mathbf{N}} \mathbf{K}_n}$.

We denote a Galerkin cone by $\mathbf{K}(\mathbf{K}_n)_{n \in \mathbf{N}}$. For more information about the application of Galerkin cones in complementarity theory, we indicate the papers [7], [8], [10], [11], [12], [13] and [14].

Nonlinear Complementarity Problem. Let $\langle E, E^* \rangle$ be a dual system of locally convex spaces and $\mathbf{K} \subset E$ a pointed convex cone. Given the mapping $f \colon \mathbf{K} \to E^*$, the *nonlinear complementarity problem* associated to f and $\mathbf{K}$ is:

$$\text{NLCP}(f, \mathbf{K}) \quad \begin{cases} \text{find} & x_0 \in \mathbf{K} \\ \text{s.t.} & f(x_0) \in \mathbf{K}^* \\ & \text{and } \langle x_0, f(x_0) \rangle = 0. \end{cases}$$

Given two mappings $f\colon \mathbf{K} \to E^*$ and $g\colon \mathbf{K} \to E$ the *implicit complementarity problem* is:

$$\mathrm{ICP}(f,g,\mathbf{K}) \quad \begin{cases} \text{find} & x_0 \in \mathbf{K} \\ \text{s.t.} & g(x_0) \in \mathbf{K}, \quad f(x_0) \in \mathbf{K}^* \\ & \text{and } \langle g(x_0), f(x_0)\rangle = 0. \end{cases}$$

The problem $\mathrm{NLCP}(f,\mathbf{K})$ is important in optimization, Economics, mechanics, engineering, game theory, etc. [8]. The problem $\mathrm{ICP}(f,g,\mathbf{K})$ was defined in relation with the study of some problems in stochastic optimal control [8]. The problems $\mathrm{NLCP}(f,\mathbf{K})$, $\mathrm{ICP}(f,g,\mathbf{K})$ can be solvable or unsolvable.

Solvability By Fixed Points Theorems. Given a topological space X and a mapping $f\colon X \to X$, the *fixed point problem* is to know under what conditions there exists a point $x_* \in X$ such that $f(x_*) = x_*$. This problem is studied in the Fixed Point Theory, which is a very popular domain in Nonlinear Analysis. In particular the Fixed Point Theory has been used by several authors in the study of solvability of the problem $\mathrm{NLCP}(f,\mathbf{K})$. The results obtained in this sense, are based on some equivalences between $\mathrm{NLCP}(f,\mathbf{K})$ and the fixed point problem. Let $(H,\langle\cdot,\cdot\rangle)$ be a Hilbert space, $\mathbf{K} \subset H$ a pointed closed convex cone and $f\colon \mathbf{K} \to H$ a mapping.

THEOREM 3 The element $x_* \in \mathbf{K}$ is a solution of the problem $\mathrm{NLCP}(f,\mathbf{K})$ if and only if x_* is a fixed point in $\mathbf{K}$ for the mapping $T(x) = P_{\mathbf{K}}(x - f(x))$. □

PROOF. Suppose that $x_* \in \mathbf{K}$ is a solution of the problem $\mathrm{NLCP}(f,\mathbf{K})$. We can show that x_* satisfies properties 1), 2), of Theorem 1 for $x = x_* - f(x_*)$.

Conversely, if $x_* \in \mathbf{K}$ and $x_* = P_{\mathbf{K}}(x_* - f(x_*))$, then since $P_{\mathbf{K}}(x_* - f(x_*))$ satisfies properties 1), 2) of Theorem 1 we deduce that $x_*$is a solution of the problem $\mathrm{NLCP}(f,\mathbf{K})$. □

THEOREM 4 The problem $\mathrm{NLCP}(f,\mathbf{K})$ has a solution if and only if the mapping $\Phi(x) = P_{\mathbf{K}}(x) - f(P_{\mathbf{K}}(x))$, defined for every $x \in H$, has a fixed point in H. Moreover, if x_0 is a fixed point of Φ, then $x_* = P_{\mathbf{K}}(x_0)$ is a solution of the problem $\mathrm{NLCP}(f,\mathbf{K})$. □

PROOF. Suppose that x_0 is a fixed point for the mapping Φ, i.e.,

$$x_0 = P_{\mathbf{K}}(x_0) - f\left(P_{\mathbf{K}}(x_0)\right).$$

If we denote by $x_* = P_{\mathbf{K}}(x_0)$, we have that $x_* \in \mathbf{K}$ and $x_0 = x_* - f(x_*)$, or $x_* - x_0 = f(x_*)$. Applying Theorem 1 we can show that $f(x_*) \in \mathbf{K}^*$ and $\langle x_*, f(x_*)\rangle = 0$, i.e., x_* is a solution of the problem $\mathrm{NLCP}(f,\mathbf{K})$.

Conversely, if $x_* \in \mathbf{K}$ is a solution of the problem $\mathrm{NLCP}(f,\mathbf{K})$, then denoting by $x_0 = x_* - f(x_*)$ and applying Theorem 2 we deduce that $P_{\mathbf{K}}(x_0) = x_*$ and finally,

$$\begin{aligned} \Phi(x_0) &= P_{\mathbf{K}}(x_0) - f(P_{\mathbf{K}}(x_0)) \\ &= x_* - f(x_*) = x_0, \end{aligned}$$

i.e., x_0 is a fixed point of Φ. □

The mapping, Φ defined in Theorem 4 was applied in complementarity theory in 1988, [7], while the mapping $\Psi(x) = x - \Phi(x)$ was used in 1992 [19]. The mapping Ψ is known as the *normal map*. By Theorem 3 the $\mathrm{NLCP}(f,\mathbf{K})$ is transformed in a fixed point problem for the mapping T with respect to the cone $\mathbf{K}$ while, by Theorem 4 the problem $\mathrm{NLCP}(f,\mathbf{K})$ is transformed in a fixed point problem with respect to the whole space H. Several existence results for the problem $\mathrm{NLCP}(f,\mathbf{K})$ have been obtained by several authors using the fixed point theory and the mappings T and Φ, [6], [7], [8], [10], [13], [3]. The fixed point problem associated to the mappings T and Φ has been also used in several iterative methods for solving numerically the problem $\mathrm{NLCP}(f,\mathbf{K})$ [1], [8], [13], [17], [18] etc.

In [15] and also in [2] it is shown that the problem $\mathrm{NLCP}(f,\mathbf{K})$ is equivalent to the following variational inequality

$$\mathrm{VI}(f,\mathbf{K}) \quad \begin{cases} \text{find} & x \in \mathbf{K} \\ \text{s.t.} & \langle f(x), y - x\rangle \geq 0 \\ & \text{for all } y \in \mathbf{K}. \end{cases}$$

Because, the fixed point theory is systematically applied to the study of variational inequalities, we have by this way another possibility to use the fixed point theory in the study of the problem $\mathrm{NLCP}(f,\mathbf{K})$. In this sense are relevant the results obtained in [5], [7], [8], [12] and in many other papers dedicated to the study of variational inequal-

ities. In the study of some economical problems, we are interested to find a solution of the problem NLCP$(f, \mathbf{K})$ which is also the least element of the feasible set

$$F = \{x \in \mathbf{K}\colon f(x) \in \mathbf{K}^*\}.$$

This particular problem can be also studied by the fixed point theory [5], [8]. If the cone $\mathbf{K}$ is an *isotone projection cone* in a Hilbert space H and if the mapping $f\colon H \to H$ satisfies some properties with respect to the ordering defined by $\mathbf{K}$, we obtain that the mappings T and Φ are monotone increasing or the difference of two monotone increasing mappings. In this case, we can apply some fixed point theorems based on the ordering, to study of the problem NLCP$(f, \mathbf{K})$. Several results in this sense are presented in [13].

The Nonlinear Complementarity Problem As a Mathematical Tool In Fixed Point Theory. The fixed point theorems on cones attracted the attention of many mathematicians. The applications of such kind of fixed point theorems are very important. We will show now how the problem NLCP$(f, \mathbf{K})$ can be used to obtain new fixed point theorems on cones.

Let H be a Hilbert space, $\mathbf{K} \subset H$ a closed pointed convex cone and $h\colon \mathbf{K} \to \mathbf{K}$ a mapping. The fixed point problem associated to h and $\mathbf{K}$ is:

$$\text{FP}(h, \mathbf{K}) \quad \begin{cases} \text{find} & x_0 \in \mathbf{K} \\ \text{s.t.} & h(x_0) = x_0. \end{cases}$$

Consider the mapping $f\colon \mathbf{K} \to H$ defined by $f(x) = x - h(x)$ for all $x \in \mathbf{K}$.

THEOREM 5 The problems NLCP$(f, \mathbf{K})$ and FP$(h, \mathbf{K})$ are equivalent. □

PROOF. Suppose that x_* is a solution of the problem FP$(h, \mathbf{K})$. In this case we have $h(x_*) = x_*$, which implies that $f(x_*) = 0$. It is evident that x_* is a solution of the problem NLCP$(f, \mathbf{K})$. Conversely, if x_* is a solution of the problem NLCP$(f, \mathbf{K})$ we have that x_* is a solution of the problem VI$(f, \mathbf{K})$, i.e., $x_* \in \mathbf{K}$ and $\langle f(x_*), y - x_* \rangle \geq 0$ for all $y \in \mathbf{K}$. But $f(x_*) = x_* - h(x_*)$ and $h(x_*) \in \mathbf{K}$ (by hypothesis). This means that

$$0 \leq \langle x_* - h(x_*), x_* - h(x_*) \rangle \leq 0,$$

which implies that $h(x_*) = x_*$. □

We note that Theorem 5 was applied to obtain new fixed point theorems [7], [10], [11]. We cite only the following two fixed point theorems.

THEOREM 6 Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space ordered by a Galerkin cone $\mathbf{K}(\mathbf{K})_{n \in \mathbf{N}}$. Let $T\colon \mathbf{K} \to \mathbf{K}$ be a mapping satisfying the following assumptions:

1) $T(0) \neq 0$;
2) T is a (ws)-compact operator;
3) T is ϕ-asymptotically bounded, with $\lim_{t \to \infty} \phi(t) \neq +\infty$.

Then, T has a fixed point $x_* \in \mathbf{K} \setminus \{0\}$. Moreover, x_* is the limit of a sequence $\{x_m\}_{m \in \mathbf{N}}$ where for every $m \in \mathbf{N}$, x_m is a solution of the problem NLCP$(T, \mathbf{K}_m)$. □

PROOF. The terminology and the proof is in [7]. □

Recently, a new proof for this theorem was proposed in [14].

THEOREM 7 Let $(H, \langle \cdot, \cdot \rangle)$ be a Hilbert space ordered by a Galerkin cone $\mathbf{K}(\mathbf{K})_{n \in \mathbf{N}} \subset H$. Suppose, given two continuous operators $S, T\colon \mathbf{K} \to H$ such that S is bounded, T is compact and $(S+T)(\mathbf{K}) \subseteq \mathbf{K}$. If the following assumptions are satisfied:

1) $I - S$ satisfies condition $(S)_+$;
2) $I - S - T$ satisfies condition (GM),

then $S + T$ has a fixed point in $\mathbf{K}$. □

PROOF. The terminology and the proof is in [11]. □

We note that Theorem 7 has several interesting corollaries. In [10] the reader can find other fixed point theorems for set-valued operators.

Conclusions. This interesting double relation between the *nonlinear complementarity problem* and the *fixed point theory*, can be exploited to obtain new results in complementarity theory and also in fixed point theory.

See also: **Principal pivoting methods for linear complementarity problems; Linear complementarity problem; Convex-simplex algorithm; Sequential simplex method; Parametric linear programming: Cost sim-**

plex algorithm; Linear programming; Lemke method; Integer linear complementary problem; LCP: Pardalos–Rosen mixed integer formulation; Order complementarity; Generalized nonlinear complementarity problem; Topological methods in complementarity theory.

References

[1] AHN, B.H.: 'Solution of nonsymmetric linear complementarity problems by iterative methods', *J. Optim. Th. Appl.* **33**, no. 2 (1981), 175–185.

[2] COTTLE, R.W.: *Complementarity and variational problems*, Vol. 19, Amer. Math. Soc., 1976, pp. 177–208.

[3] HYERS, D.H., ISAC, G., AND RASSIAS, T.M.: *Topics in non-linear analysis and applications*, World Sci., 1997.

[4] ISAC, G.: 'On the implicit complementarity problem in Hilbert spaces', *Bull. Austral. Math. Soc.* **32**, no. 2 (1985), 251–260.

[5] ISAC, G.: 'Complementarity problem and coincidence equations o convex cones', *Boll. Unione Mat. Ital. Ser. B* **6** (1986), 925–943.

[6] ISAC, G.: 'Fixed point theory and complementarity problems in Hilbert spaces', *Bull. Austral. Math. Soc.* **36**, no. 2 (1987), 295–310.

[7] ISAC, G.: 'Fixed point theory, coincidence equations on convex cones and complementarity problem', *Contemp. Math.* **72** (1988), 139–155.

[8] ISAC, G.: *Complementarity problems*, Vol. 1528 of *Lecture Notes Math.*, Springer, 1992.

[9] ISAC, G.: 'Tihonov's regularization and the complementarity problem in Hilbert spaces', *J. Math. Anal. Appl.* **174**, no. 1 (1993), 53–66.

[10] ISAC, G.: 'Fixed point theorems on convex cones, generalized pseudo-contractive mappings and the complementarity problem', *Bull. Inst. Math. Acad. Sinica* **23**, no. 1 (1995), 21–35.

[11] ISAC, G.: 'On an Altman type fixed point theorem on convex cones', *Rocky Mountain J. Math.* **25**, no. 2 (1995), 701–714.

[12] ISAC, G., AND GOELEVEN, D.: 'Existence theorems for the implicit complementarity problem', *Internat. J. Math. and Math. Sci.* **16**, no. 1 (1993), 67–74.

[13] ISAC, G., AND NEMÉTH, A.B.: 'Projection methods, isotone projection cones and the complementarity problem', *J. Math. Anal. Appl.* **153**, no. 1 (1990), 258–275.

[14] JACHYMSKI, J.: 'On Isac's fixed point theorem for self-maps of a Galerkin cone', *Ann. Sci. Math. Québec* **18**, no. 2 (1994), 169–171.

[15] KARAMARDIAN, S.: 'Generalized complementarity problem', *J. Optim. Th. Appl.* **8** (1971), 161–168.

[16] MOREAU, J.: 'Décomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires', *C.R. Acad. Sci. Paris* **225** (1962), 238–240.

[17] NOOR, M.A.: 'Fixed point approach for complementarity problems', *J. Math. Anal. Appl.* **133** (1988), 437–448.

[18] NOOR, M.A.: 'Iterative methods for a class of complementarity problems', *J. Math. Anal. Appl.* **133** (1988), 366–382.

[19] ROBINSON, S.M.: 'Normal maps induced by linear transformations', *Math. Oper. Res.* **17**, no. 3 (1992), 691–714.

[20] ZARANTONELLO, E.H.: 'Projection on convex sets in Hilbert space and spectral theory', in E.H. ZARANTONELLO (ed.): *Contributions to Nonlinear Functional Analysis*, Acad. Press, 1971, pp. 237–424.

George Isac
Royal Military College of Canada
Kingston, Ontario, Canada
E-mail address: isac-g@rmc.ca

MSC2000: 90C33
Key words and phrases: nonlinear complementarity problem, fixed point problem.

ESTIMATING DATA FOR MULTICRITERIA DECISION MAKING PROBLEMS: OPTIMIZATION TECHNIQUES

One of the most crucial steps in many multicriteria decision making methods (MCDM) is the accurate estimation of the pertinent data [18]. Very often these data cannot be known in terms of absolute values. For instance, what is the worth of the ith alternative in terms of a political impact criterion? Although information about questions like the previous one is vital in making the correct decision, it is very difficult, if not impossible, to quantify it correctly. Therefore, many decision making methods attempt to determine the *relative* importance, or weight, of the alternatives in terms of each criterion involved in a given decision making problem.

Consider the case of having a single decision criterion and a set of n alternatives, denoted as A_i (for $i = 1, \ldots, n$). The decision maker wants to determine the relative performance of these alternatives in terms of a single criterion. An approach based on *pairwise comparisons* which was proposed by T.L. Saaty [11], and [12] has long attracted the interest of many researchers, because both of its easy applicability and interesting mathematical properties. Pairwise comparisons are used

to determine the relative importance of each alternative in terms of each criterion.

In that approach the decision maker has to express his/her opinion about the value of one single pairwise comparison at a time. Usually, the decision maker has to choose his/her answer among 10–17 discrete choices. Each choice is a linguistic phrase. Some examples of such linguistic phrases when two concepts, **A** and **B** are considered might be: '**A** is more important than **B**', or '**A** is of the same importance as **B**', or '**A** is a little more important than **B**', and so on. When one focuses directly on the *data elicitation* issue one may use linguistic statements such as 'How much more does alternative **A** belong to the set **S** than alternative **B**'?

The main problem with the pairwise comparisons is how to quantify the *linguistic choices* selected by the decision maker during the evaluation of the pairwise comparisons. All the methods which use the pairwise comparisons approach eventually express the qualitative answers of a decision maker into some numbers.

Pairwise comparisons are quantified by using a *scale*. Such a scale is nothing but an one-to-one mapping between the set of discrete linguistic choices available to the decision maker and a discrete set of numbers which represent the importance, or weight, of the previous linguistic choices. There are two major approaches in developing such scales. The first approach is based on the *linear scale* proposed by Saaty [12] as part of the *analytic hierarchy process* (AHP). The second approach was proposed by F. Lootsma [8], [9], [10] and determines *exponential scales*. Both approaches depart from some psychological theories and develop the numbers to be used based on these psychological theories. For an extensive study of the scale issue, see [18] and [19].

In this article we examine three problems related to the use of pairwise comparisons for data elicitation in MCDM. The first problem is how to combine the $n(n-1)/2$ comparisons needed to compare n entities (alternatives or *criteria*) under a given goal and extract their relative preferences. This subject was extensively studied in [21] and it is briefly discussed in the second section. The second problem in this article is how to estimate *missing comparisons*. The third problem is how to select the order for eliciting the comparisons and determine whether all comparisons are needed. These problems are examined in detail in the following sections.

Extraction of Relative Priorities from Complete Pairwise Matrices. Let $A_1, \ldots, A_n$ be n alternatives (or criteria or, in general, concepts) to be compared. We are interested in evaluating the relative preference values of the above concepts. Saaty [11], [12], [14] proposed to use a matrix A of rational numbers taken from the set $\{1/9, 1/8, 1/7, \ldots, 1, \ldots, 9\}$. Each entry of the above matrix A represents a *pairwise judgment*. Specifically, the entry a_{ij} denotes the number that estimates the relative preference of element A_i when it is compared with element A_j. Obviously, $a_{ij} = 1/a_{ji}$ and $a_{ii} = 1$. That is, the matrix is reciprocal.

The Eigenvalue Approach. Let us first examine the case in which it is possible to have perfect values a_{ij}. In this case it is $a_{ij} = W_i/W_j$ (W_s denotes the actual value of element s) and the previous reciprocal matrix A is *consistent*. That is:

$$a_{ij} = a_{ik} \times a_{kj} \quad \text{for } i, j, k = 1, \ldots, n, \tag{1}$$

where n is the number of elements in the comparison set. It can be proved [12] that the matrix A has rank 1 with n to be its nonzero eigenvalue. Thus, we have:

$$Ax = nx, \tag{2}$$

where x is an eigenvector. From the fact that $a_{ij} = W_i/W_j$, the following are obtained:

$$\sum_{j=1}^{n} a_{ij} W_j = \sum_{j=1}^{n} W_i = nW_i, \quad i = 1, \ldots, n, \tag{3}$$

or

$$AW = nW. \tag{4}$$

Equation (4) states that n is an eigenvalue of A with W being a corresponding eigenvector. The same equation also states that in the *perfectly consistent case* (i.e., when $a_{ij} = a_{ik} \times a_{kj}$ for all possible triplets), the vector W, with the relative pref-

erences of the elements $A_1, \ldots, A_n$, is the principal right eigenvector (after normalization) of A.

In the nonconsistent case (which is the most common) the pairwise comparisons are not perfect, that is, the entry a_{ij} might deviate from the real ratio W_i/W_j (i.e., from the ratio of the real relative preference values W_i and W_j). In this case, the previous expression (1) does not hold for all possible combinations. Now the new matrix A can be considered as a perturbation of the previous consistent case. When the entries a_{ij} change slightly, then the eigenvalues change in a similar fashion [12]. Moreover, the maximum eigenvalue is close to n (actually greater than n) while the remaining eigenvalues are close to zero. Thus, in order to find the relative preferences in the nonconsistent cases, one should find an eigenvector that corresponds to the maximum eigenvalue $\lambda_{\max}$. That is to say, to find the principal right eigenvector W that satisfies:

$$AW = \lambda_{\max} W \quad \text{where } \lambda_{\max} = n.$$

Saaty estimates the principal right eigenvector W by multiplying the entries in each row of A together and taking the nth root (n being the number of the elements in the comparison set). Since we desire to have values that add up to 1, we normalize the previously found vector by the sum of the above values. If we want to have the element with the highest value to have a relative preference value equal to 1, we divide the previously found vector by the highest value.

Under the assumption of *total consistency*, if the judgments are gamma distributed (something that Saaty claims to be the case), the principal right eigenvector of the resultant reciprocal matrix A is Dirichlet distributed. If the assumption of total consistency is relaxed, then L.G. Vargas [23] proved that the hypothesis that the principal right eigenvector follows a Dirichlet distribution is accepted if the consistency ratio is 10% or less.

The *consistency ratio* (CR) is obtained by first estimating $\lambda_{\max}$. Saaty estimates $\lambda_{\max}$ by adding the columns of matrix A and then multiplying the resulting vector with the vector W. Then, he uses what he calls the *consistency index* (CI) of the matrix A. He defined CI as follows:

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n - 1}.$$

Then, the consistency ratio CR is obtained by dividing the CI by the random consistency index (RCI) as given in table 1. Each RCI is an average random consistency index derived from a sample of size 500 of randomly generated reciprocal matrices with entries from the set $\{1/9, 1/8, 1/7, \ldots, 1, \ldots, 9\}$ to see if its CI is 10% or less. If the previous approach yields a CR greater than 10%, then a reexamination of the pairwise judgments is recommended until a CR less than or equal to 10% is achieved.

Optimization Approaches. A.T.W. Chu, R.E. Kalaba and K. Spingarn [2] claimed that given the data a_{ij}, the values W_i to be estimated are desired to have the property:

$$a_{ij} \approx \frac{W_i}{W_j}. \tag{5}$$

This is reasonable since a_{ij} is meant to be the estimation of the ratio W_i/W_j. Then, in order to get the estimates for the W_i given the data a_{ij}, they proposed the following constrained optimization problem:

$$\begin{cases} \min & S = \displaystyle\sum_{i=j}^{n} \sum_{j=i}^{n} (a_{ij} w_j - w_i)^2, \\ \text{s.t.} & \displaystyle\sum_{i=j}^{n} W_i = 1, \\ & W_i > 0 \quad \text{for } i = 1, \ldots, n. \end{cases} \tag{6}$$

They also provide an alternative expression S_1 that is more difficult to solve numerically. That is,

$$S_1 = \sum_{i=j}^{n} \sum_{j=i}^{n} (a_{ij} - W_j/W_i)^2. \tag{7}$$

In [3] a variation of the above *least squares* formulation is proposed. For the case of only one decision maker it recommends the following models:

$$\log a_{ij} = \log W_i - \log W_j + \psi_2(W_i, W_j)\varepsilon_{ij}, \tag{8}$$

$$a_{ij} = \frac{W_i}{W_j} + \psi_2(W_i, W_j)\varepsilon_{ij}, \tag{9}$$

where W_i and W_j are the true (and hence unknown) relative preferences; $\psi_1(X, Z)$ and $\psi_2(X, Z)$ are given positive functions (where

n	1	2	3	4	5	6	7	8	9
RCI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Table 1: RCI values for sets of different order n [12].

$X, Z > 0$). The random errors ε_{ij} are assumed independent with zero mean and unit variance. Using these two assumptions one is able to calculate the variance of each individual estimated relative preference. However, is fails to give a way of selecting the appropriate positive functions. In the second example, presented later, a sample problem which originates in [11] and later in [3] is solved for different functions ψ_1, ψ_2 using this method.

Considering the Human Rationality Factor. According to the *human rationality assumption* [21] the decision maker is a rational person. Rational persons are defined here as individuals who try to minimize their regret [15], to minimize losses, or to maximize profit [24]. In the relative preference evaluation problem, *minimization of regret*, losses, or maximization of profit could be interpreted as the effort of the decision maker to minimize the errors involved in the pairwise comparisons.

As it is stated in previous paragraphs, in the inconsistent case the entry a_{ij} of the matrix A is an estimation of the real ratio W_i/W_j. Since it is an estimation, the following is true:

$$a_{ij} = \left(\frac{W_i}{W_j}\right) d_{ij}, \quad i, j = 1, \ldots, n. \tag{10}$$

In the above relation d_{ij} denotes the deviation of a_{ij} from being an accurate judgment. Obviously, if $d_{ij} = 1$, then the a_{ij} was perfectly estimated. From the previous formulation we conclude that the errors involved in these pairwise comparisons are given by:

$$\varepsilon_{ij} = d_{ij} - 1.00,$$

or after using (10), above:

$$\varepsilon_{ij} = a_{ij}\left(\frac{W_j}{W_i}\right) - 1.00. \tag{11}$$

When a comparison set contains n elements, then Saaty's method requires the estimation of the following $n(n-1)/2$ pairwise comparisons:

$$\frac{W_2}{W_1}, \cdots, \frac{W_n}{W_1}, \tag{12}$$

$$\frac{W_3}{W_2}, \cdots, \frac{W_n}{W_2},$$

$$\vdots$$

$$\frac{W_{n-1}}{W_n}.$$

The corresponding $n(n-1)/2$ errors are (after using relations (11) and (12)):

$$\varepsilon_{ij} = a_{ij}\left(\frac{W_j}{W_i}\right) - 1.00, \tag{13}$$

$$i, j = 1, \ldots, n, \text{ and } j > 1.$$

Since the W_i are relative preferences that add up to 1, the following relation (14) should also be satisfied:

$$\sum_{i=1}^{n} W_i = 1.00. \tag{14}$$

Apparently, since the W_i represent relative preferences we also have:

$$W_i > 0, \quad i = 1, \ldots, n. \tag{15}$$

Relations (13) and (14), when the data are consistent (i.e., all the errors are equal to zero), can be written as follows:

$$BW = b. \tag{16}$$

The vector b has zero entries everywhere except the last one that is equal to 1, and the matrix B has the following form (blank entries represent zeros):

$$B = \begin{bmatrix} 1 & 2 & 3 & \cdots & n & \\ -1 & a_{1,2} & & & & 1 \\ -1 & & a_{1,3} & & & 2 \\ \vdots & & & \ddots & & \vdots \\ -1 & & & & a_{1,n} & n-1 \\ & -1 & a_{2,3} & & & 1 \\ & \vdots & & \ddots & & \vdots \\ & -1 & & & a_{2,n} & n-2 \\ & & \ddots & & & \vdots \\ & & & & a_{n-1,n} & 1 \\ 1 & 1 & 1 & \cdots & 1 & \end{bmatrix}.$$

The *error minimization* issue is interpreted in many cases (regression analysis, linear least squares problem) as the minimization of the *sum of squares* of the residual vector: $r = b - BW$ [16]. In terms of formulation (15) this means that in a real life situation (i.e., when errors are not zero any more) the real intention of the decision maker is to minimize the expression:

$$f^2(x) = \|b - BW\|, \tag{17}$$

which, apparently, expresses a typical linear least squares problem.

If we use the notation described previously, then the quantity (6) which is minimized in [2] becomes:

$$S = \sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij}W_j - W_i)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}(\varepsilon_{ij}W_i)^2$$

and the alternative expression (7) becomes:

$$S_1 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(a_{ij}\frac{W_j}{W_i}\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\varepsilon_{ij}\frac{W_i}{W_j}\right)^2.$$

Clearly, both expressions are too complicated to reflect, in a reasonable way, the intentions of the decision maker.

The models proposed in [3] are closer to the one developed under the human rationality assumption. The only difference is that instead of the relations:

$$\log a_{ij} = \log w_i - \log W_j + \psi_1(W_i, W_j)\varepsilon_{ij}$$

and

$$a_{ij} = \frac{W_i}{W_j} + \psi_2(W_i, W_j)\varepsilon_{ij},$$

the following simpler expression is used:

$$a_{ij} = \frac{W_i}{W_j}d_{ij}, \tag{18}$$

or

$$a_{ij} = \frac{W_i}{W_j} \times (\varepsilon_{ij} + 1.00).$$

However, as the second example illustrates, the performance of this method is greatly dependent on the selection of the $\psi_1(X, Z)$ or $\psi_2(X, Z)$ functions. Now, however, these functions are further modified by (17).

EXAMPLE 1 Let us assume that the following is the matrix with the pairwise comparisons for a set of four elements:

$$A = \begin{bmatrix} 1 & 2/1 & 1/5 & 1/9 \\ 1/2 & 1 & 1/8 & 1/9 \\ 5/1 & 8/1 & 1 & 1/4 \\ 9/1 & 9/1 & 4/1 & 1 \end{bmatrix}.$$

Using the methods presented in previous sections we can see that

$$\lambda_{\max} = 4.226;$$
$$\mathrm{CI} = \frac{4.226 - 4}{4 - 1} = 0.053,$$
$$\mathrm{CR} = \frac{\mathrm{CI}}{0.90} = 0.0837 < 0.10.$$

The formulation (15) that corresponds to this example is as follows:

$$\begin{bmatrix} -1 & 2/1 & 0.0 & 0 \\ -1 & 0.0 & 1/5 & 0 \\ 1 & 0.0 & 0 & 1/9 \\ 0.0 & -1 & 1/8 & 0 \\ 0.0 & -1 & 0 & 1/9 \\ 0.0 & 0.0 & -1 & 1/4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1.0 \end{bmatrix}.$$

The vector V that solves the above least squares problem is calculated to be:

$$V = (0.065841\ 0.039398\ 0.186926\ 0.704808).$$

Hence, the sum of squares of the residual vector components is 0.003030. The average squared residual for this problem is $0.003030/((4(4 - 1)/2) + 1 = 0.000433$; that is, the average residual is $\sqrt{0.000433} = 0.020806$. □

EXAMPLE 2 The second example uses the same data used originally in [11], and later in [2] and [3]. These data are presented in Table 2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)	1	4	9	6	6	5	5
(2)	1/4	1	7	5	5	3	5
(3)	1/9	1/7	1	1/5	1/5	1/7	1/5
(4)	1/6	1/5	5	1	1	1/3	1/3
(5)	1/6	1/5	5	1	1	1/3	1/3
(6)	1/5	1/3	7	3	3	1	2
(7)	1/5	1/4	5	3	3	1/2	1

Table 2: Data for the second example.

Table 3 presents a summary of the results (as found in the corresponding references) when the methods described in the subsections above are used. The *power method* for deriving the eigenvector was applied as presented in [7]. In the last row of Table 2 are the results obtained by using the

least square method under the human rationality assumption (HR).

As it is shown in the last column of Table 3, the performance of each method is very different as far the mean residual is concerned. The results also illustrate how critical is the role of the functions $\psi_1(X,Z)$ and $\psi_2(X,Z)$ in the method of [3]. The mean residual obtained by using the least squares method under the human rationality assumption is the smallest one by 16%. □

Matrices with Missing Comparisons. For one to evaluate n concepts, normally all the required $n(n-1)/2$ pairwise comparisons are needed. However, for large numbers of concepts to be compared, the decision maker may become quite bored, tired and inattentive with assigning the values to the comparisons as time is going on, which may easily lead to erroneous judgments. Moreover, the time spent to elicit all the comparisons for a judgment matrix may be unaffordable. Also the decision maker may not be sure about the values of some comparisons and thus may not want to make a direct evaluation of them. In cases like the previous ones, the decision maker may wish to stop the process and then try to derive the relative preferences from an incomplete pairwise comparison (judgment) matrix.

Given an incomplete pairwise comparison matrix, there are two central and closely interrelated problems. The first problem is how to estimate the missing comparisons. The second problem is which comparison to evaluate next. In other words, if the decision maker wishes to estimate a few extra comparisons (from the remaining undetermined ones) how should the next comparison be selected? Should it be selected randomly or according to some rule (to be determined)? Next, we study the first of these two closely related problems.

Estimating Missing Comparisons.

Using Connecting Paths. Suppose that $X_{i,j}$ is a missing comparison to be estimated. Next, also assume that there are two known comparisons $a_{i,k}$ and $a_{k,j}$ for some index k. In the perfectly consistent case the following relationship should be true:

$$X_{i,j} = a_{i,k} \times a_{k,j}.$$

In the more general inconsistent case, the $X_{i,j}$ value can be approximated by the product $a_{i,k} \times a_{k,j}$. In [5], and [6] the pair $a_{i,k}$ and $a_{k,j}$ is called an *elementary connecting path* connecting the missing comparison $X_{i,j}$. Obviously, given a missing comparison, more than one such connecting path may exist (i.e., if there are more than one k indexes which satisfy the above relationship). Moreover, it is also possible to have connecting paths comprised by more than two known comparisons (i.e., paths of size larger than 2). The general structure of a connecting path of size r, denoted as CP_r, has the following form:

$$\mathrm{CP}_r: \quad X_{i,j} = a_{i,k1} \times a_{k1,k2} \times \cdots \times a_{kr,j},$$

for $i, j, k1, \ldots, kr = 1, \ldots, n$, $1 \leq r \leq n-2$.

According to P.T. Harker [5], [6] the value of the missing comparison $X_{i,j}$ should be equal to the geometric mean of all connecting paths related to this missing comparison. That is, the following should be true:

$$X_{ij} = \sqrt[q]{\prod_{r=1}^{q} \mathrm{CP}_r}.$$

In the previous expression it is assumed that there are q such connecting paths. For the above reasons, this method is known as the *geometric mean method* for estimating missing comparisons.

A method alternative to the geometric means method is to express the missing comparisons in terms of the arithmetic averages of all related connecting paths and some error terms. In this way, one can also introduce error terms on consistency relations which are defined on pairs of missing comparisons (for more details, please see [1]). A natural objective then, could be to minimize the sum of the absolute terms of all these error terms (which can be of any sign). That is, the above consideration leads to the formulation of a linear programming (LP) problem. A similar approach is presented in [17] (in which the path problem does not occur).

However, there is a serious drawback with any method which attempts to use connecting paths. The number of connecting paths may be astronomically large, rendering any such method computa-

	elements in set							
method used	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Ave. residual
Saaty eigenvector method	0.429	0.231	0.021	0.053	0.053	0.119	0.095	0.134
Power method eigenvector	0.427	0.230	0.021	0.052	0.052	0.123	0.094	0.135
Chu's method	0.487	0.175	0.030	0.059	0.059	0.104	0.085	0.097
Federov model 1 with $\psi_1 = 1$	0.422	0.232	0.021	0.052	0.052	0.127	0.094	0.138
Federov Model 2 with $\psi_2 = 1$	0.386	0.287	0.042	0.061	0.061	0.088	0.075	0.161
Federov Model 2 with $\psi_2 = \lvert W_i - W_j \rvert$	0.383	0.262	0.032	0.059	0.059	0.122	0.083	0.152
Federov Model 2 with $\psi_2 = W_i/W_j$	0.047	0.229	0.021	0.051	0.051	0.120	0.081	0.130
Least squares method under the HR assumption	0.408	0.147	0.037	0.054	0.054	0.080	0.066	0.082

Table 3: Comparison of the relative preferences for the data in Table 2.

tionally intractable. For instance, for a comparison matrix of dimension of six, the number of possible connecting paths to be considered might be equal to 64, while in a case of dimension equal to ten, the number of paths may become equal to 109,600. As a result, some alternative approaches have been developed. The revised geometric means method (or RGM) method and a least squares formulation are two such methods and are discussed next.

Revised Geometric Mean Method (RGM). An alternative approach to the use of connecting paths, is to convert the incomplete judgement matrix into a transformed matrix and then determine its principal right eigenvector. This was proposed by Harker [4] and it is best illustrated by means of an example.

Suppose that the following is an incomplete judgement matrix of order 3 (taken from [4]).

$$A_0 = \begin{bmatrix} 1 & 2 & - \\ 1/2 & 1 & 2 \\ - & 1/2 & 1 \end{bmatrix}.$$

One can replace the missing elements (denoted by $-$) by the corresponding ratios of weights. Therefore, the previous matrix becomes:

$$A_1 = \begin{bmatrix} 1 & 2 & w_1/w_3 \\ 1/2 & 1 & 2 \\ w_3/w_1 & 1/2 & 1 \end{bmatrix}.$$

That is, the missing comparison $X_{1,3}$ was replaced by the ratio w_1/w_3 (similar for the reciprocal entry $X_{3,1}$). Next observe that the product A_1W is equal to:

$$A_1W = \begin{bmatrix} 1 & 2 & w_1/w_3 \\ 1/2 & 1 & 2 \\ w_3/w_1 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 2w_1 + 2w_2 \\ w_1/2 + w_2 + 2w_3 \\ w_2/2 + 2w_3 \end{bmatrix}$$

The same result can also be obtained if one considers the matrix C, given as follows:

$$C = \begin{bmatrix} 2 & 2 & 0 \\ 1/2 & 1 & 2 \\ 0 & 1/2 & 2 \end{bmatrix},$$

that is, matrix C satisfies the relationship

$$A_1W = CW.$$

Therefore, the desired relative preferences (i.e., the entries of vector W) can be determined as the principal right eigenvector of the new matrix C. This is true because:

$$A_1W = CW = \lambda W.$$

In general, the entries of matrix C can be determined from the entries of an incomplete judgement matrix A_0 as follows (where $c_{i,j}$ and $a_{i,j}$ are the elements of the matrices C and A_0, respectively):

$$c_{i,i} = 1 + m_i$$

and for $i \neq j$:

$$c_{i,j} = \begin{cases} a_{i,j} & \text{if } a_{i,j} \text{ is a positive number,} \\ 0 & \text{otherwise,} \end{cases}$$

where m_i is the number of unanswered questions in the ith row of the incomplete comparison matrix.

Next, the elements of the W vector can be determined by using one of the methods presented in the second section.

Least Squares Formulation. This formulation is a natural extension of the formulation discussed earlier in the section on the HR factor. The only difference is that in relations (12) one should only consider known comparisons. This, as a result, implies that the new matrix B (as defined earlier) should not have rows which would correspond to missing comparisons. Finally, observe that in order to solve the least squares problem given as (16), one has to calculate the vector W as follows:

$$W = (B^\top B)^{-1} B^\top b,$$

where $B^\top$ stands for the transpose of B.

In [1] the revised geometric means and the previous least squares method were tested on random problems. First, a complete judgment matrix was determined. These matrices, in general, were slightly inconsistent. They were derived according to the procedures used in [22], [20], and [19]. Then, some comparisons were randomly removed and set as missing. Then, the previous two methods were applied on the incomplete judgment matrix and the missing comparisons were estimated. The estimated matrix was used to derive a ranking of the compared entities. This ranking was compared with the ranking derived when the original complete judgment matrix is used. In these computational experiments it was found that the two estimation methods for missing comparisons performed almost in a similar manner. This manner was different for matrices of different order and various percentages of missing comparisons. More details on these issues can be found in [1].

Determining the Comparison to Elicit Next. Suppose that the decision maker has determined some of the $n(n-1)/2$ comparisons when a set of n entities is considered for extracting relative preferences. Next assume that the decision maker wishes to proceed with only a few additional comparisons and not determine the entire judgment matrix. The question we examine at this point is which ones the additional comparisons should be. To be more specific, the question we consider is best stated as follows: Given an incomplete judgment matrix, and the option to elicit just some additional comparisons, then which one should be the comparison to elicit next?

One obvious approach is to select the next comparison just randomly among the missing ones. This problem was examined by Harker in [5] and [6]. Harker focused his attention on how to determine which comparison, among the missing ones, is the most *critical* one. He determined as the most critical one, to be the comparison which would have the largest impact (when the appropriate derivatives are considered) on the vector W.

He observed that the largest absolute gradient (i.e., the largest partial derivative) means that a unit change of the specific missing comparison brings out the biggest change on the vector W. Therefore, he asserted, that the missing comparison related to the largest absolute gradient should be the most critical one and therefore, the one to evaluate next. Then, the following formula calculating the largest absolute gradient can be used to choose the most critical comparison index (i,j):

$$(i,j) = \arg \max_{(k,l)\epsilon Q} \left\| \frac{\partial x(A)}{\partial_{k,l}} \right\|_\infty ,$$

where Q is the set of missing comparisons and $\|\cdot\|_\infty$ is the Tchebyshev norm. The most critical comparison index (i,j) is determined by the maximum norm of the vector of $\partial x(A)/\partial_{k,l}$ which corresponds to all missing comparisons.

The previous approach is intuitively plausible but computationally non trivial. Moreover, its effectiveness had not been addressed until recently. In [1] Harker's derivatives approach was tested versus a method which randomly selects the next comparison to elicit. The test problems were generated similarly to the ones described at the end of the previous section. The two methods were also tested in a similar manner as before. To our surprise, the two methods performed in a similar manner. Therefore, the obvious conclusion is that one does not have to implement the more complex derivatives method. It is sufficient to select the next comparison just randomly. Of course, the more comparisons are selected, the better is for the accuracy of the final results. Since the order of comparisons seems not to have an impact, the

best strategy is to select as the next comparison the one which is easier for the decision maker to elicit.

Conclusions. Deriving the data for MCDM problems is an approach which requires trade-offs. Thus, it should not come as a surprise that optimization can be used at various stages of this crucial phase in solving many MCDM problems. The previous analysis of some key problems signifies that optimization becomes more critical as the size of the decision problem increases.

Finally, it should be stated here that an in depth analysis of many key issues in multicriteria decision making theory and practice is provided in [18].

See also: **Multi-objective optimization: Pareto optimal solutions, properties; Multi-objective optimization: Interactive methods for preference value functions; Multi-objective optimization: Lagrange duality; Multi-objective optimization: Interaction of design and control; Outranking methods; Preference disaggregation; Fuzzy multi-objective linear programming; Multi-objective optimization and decision support systems; Preference disaggregation approach: Basic features, examples from financial decision making; Preference modeling; Multiple objective programming support; Multi-objective integer linear programming; Multi-objective combinatorial optimization; Bi-objective assignment problem; Multicriteria sorting methods; Financial applications of multicriteria analysis; Portfolio selection and multicriteria analysis; Decision support systems with multiple criteria.**

References

[1] CHEN, Q., TRIANTAPHYLLOU, E., AND ZANAKIS, S.: 'Estimating missing comparisons and selecting the next comparison to elicit in MCDM', *Working Paper Dept. Industrial Engin. Louisiana State Univ.* (2001), http://www.imse.lsu.edu/vangelis.

[2] CHU, A.T.W., KALABA, R.E., AND SPINGARN, K.: 'A comparison of two methods for determining the weights of belonging to fuzzy sets', *J. Optim. Th. Appl.* **27**, no. 4 (1979), 321–538.

[3] FEDEROV, V.V., KUZMIN, V.B., AND VERESKOV, A.I.: 'Membership degrees determination from Saaty matrix totalities', in M.M. GUPTA AND E. SANCHEZ (eds.): *Approximate Reasoning in Decision Analysis*, North-Holland, 1982, pp. 23–30.

[4] HARKER, P.T.: 'Alternative modes of questioning in the analytic hierarchy process', *Math. Model.* **9**, no. 3–5 (1987), 353–360.

[5] HARKER, P.T.: 'Derivatives of the Perron root of a positive reciprocal matrix: With application to the analytic hierarchy process', *Appl. Math. Comput.* **22** (1987), 217–232.

[6] HARKER, P.T.: 'Incomplete pairwise comparisons in the analytic hierarchy process', *Math. Model.* **9**, no. 11 (1987), 837–848.

[7] KALABA, R., AND SPINGARN, K.: 'Numerical approaches to the eigenvalues of Saaty's matrices for fuzzy sets', *Comput. Math. Appl.* **4** (1979).

[8] LOOTSMA, F.A.: 'Numerical scaling of human judgment in pairwise-comparison methods for fuzzy multi-criteria decision analysis': *Mathematical Models for Decision Support*, Vol. 48 of *NATO ASI F: Computer and System Sci.*, Springer, 1988, pp. 57–88.

[9] LOOTSMA, F.A.: 'The French and the American school in multi-criteria decision analysis', *Rech. Oper./Operat. Res.* **24**, no. 3 (1990), 263–285.

[10] LOOTSMA, F.A.: 'Scale sensitivity and rank preservation in a multiplicative variant of the AHP and SMART', *Techn. Report Fac. Techn. Math. and Informatics Delft Univ. Techn.*, no. 91–67 (1991).

[11] SAATY, T.L.: 'A scaling method for priorities in hierarchical structures', *J. Math. Psych.* **15**, no. 3 (1977), 234–281.

[12] SAATY, T.L.: *The analytic hierarchy process*, McGraw–Hill, 1980.

[13] SAATY, T.L.: 'Priority setting in complex problems', *IEEE Trans. Engin. Management* **EM–30**, no. 3 (1983), 140–155.

[14] SAATY, T.L.: *Fundamentals of decision making and priority theory with the analytic hierarchy process*, Vol. VI, RWS Publ., 1994.

[15] SIMON, H.A.: *Models of man*, 2 ed., Wiley, 1961.

[16] STEWART, S.M.: *Introduction to matrix computations*, Acad. Press, 1973.

[17] TRIANTAPHYLLOU, E.: 'Linear programming based decomposition approach in evaluating priorities from pairwise comparisons and error analysis', *J. Optim. Th. Appl.* **84**, no. 1 (1995), 207–234.

[18] TRIANTAPHYLLOU, E.: *Multi-criteria decision making methods: A comparative study*, Kluwer Acad. Publ., 2000.

[19] TRIANTAPHYLLOU, E., LOOTSMA, F.A., PARDALOS, P.M., AND MANN, S.H.: 'On the evaluation and application of different scales for quantifying pairwise comparisons in fuzzy sets', *J. Multi-Criteria Decision Anal.* **3** (1994), 133–155.

[20] TRIANTAPHYLLOU, E., AND MANN, S.H.: 'A computational evaluation of the AHP and the revised AHP

when the eigenvalue method is used under a continuity assumption', *Computers and Industrial Engin.* **26**, no. 3 (1994), 609–618.

[21] TRIANTAPHYLLOU, E., PARDALOS, P.M., AND MANN, S.H.: 'A minimization approach to membership evaluation in fuzzy sets and error analysis', *J. Optim. Th. Appl.* **66**, no. 2 (1990), 275–287.

[22] TRIANTAPHYLLOU, E., AND SANCHEZ, A.: 'A sensitivity analysis approach for some deterministic multicriteria decision-making methods', *Decision Sci.* **28**, no. 1 (1997), 151–194.

[23] VARGAS, L.G.: 'Reciprocal matrices with random coefficients', *Math. Model.* **3** (1982), 69–81.

[24] WRITE, C., AND TATE, M.D.: *Economics and systems analysis: Introduction for public managers*, Addison-Wesley, 1973.

Qing Chen
Dept. Industrial and Manufacturing Systems Engin.
3128 CEBA Building
Louisiana State Univ.
Baton Rouge, LA 70803–6409, USA
Evangelos Triantaphyllou
Dept. Industrial and Manufacturing Systems Engin.
3128 CEBA Building
Louisiana State Univ.
Baton Rouge, LA 70803–6409, USA
E-mail address: trianta@lsu.edu
Web address: www.imse.lsu.edu/vangelis

MSC2000: 90C29

Key words and phrases: pairwise comparisons, data elicitation, multicriteria decision making, MCDM, scale, analytic hierarchy process, AHP, consistent judgment matrix, eigenvalue, eigenvector, least squares problem, incomplete judgments.

EVACUATION NETWORKS

Planning and design of evacuation networks is both a complex and critically important optimization problem for a number of emergency situations. One particularly critical class of examples concerns the *emergency* evacuation of *chemical plants*, *high-rise buildings*, and *naval vessels* due to fire, *explosion* or other emergencies. The problem is compounded because the solution must take into account the fact that human occupants may *panic* during the evacuation, therefore, there must be a well-defined set of *evacuation routes* in order to minimize the sense of panic and at the same time create safe, effective routes for evacuation. The problem is a highly *transient*, *stochastic*, *nonlinear*, *combinatorial optimization programming problem*. We focus on evacuation networks where congestion is a significant problem.

Introduction. *Evacuation* is one of the most perilous, pernicious, and persistent problems faced by humanity. *Hurricanes*, *fires*, *earthquakes*, explosions and other natural and man-made disasters happen on almost a daily basis throughout the world. How can we safely evacuate a collection of occupants within an affected region or facility is the fundamental problem faced in evacuation.

Purpose. The purpose of this article is to both introduce to the reader the problem of evacuation and its manifest nature, and also suggest some alternative approaches to optimize this process. That life-threatening evacuations happen as often as they do is somewhat surprising. That people often do not know how to safely evacuate in time of need is a sad reality. That people must help people plan for evacuation is one of the most important activities of a research scientist.

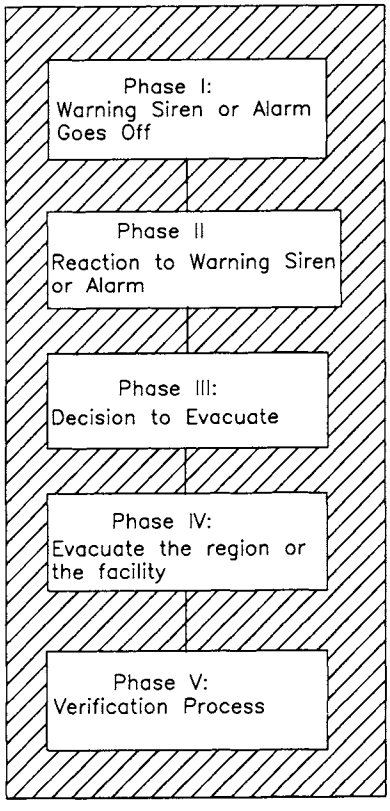

Fig. 1: Processes for an evacuation.

Outline. In this article we first introduce the problem in Section 1 and then describe our fundamental modeling 3-step methodology in Section 2.

In Section 3, we array the number of different of static and dynamic approaches to this problem and present our general approach which has guided our research on the problem. Finally, in Section 4 we discuss the algorithmic approaches to the problem where we capture the congested flow of occupants in the network and attempt to define the safest evacuation routes trading off the different objective performance measures in the network.

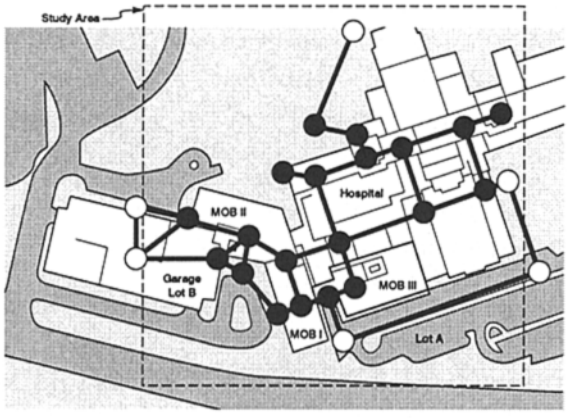

Fig. 2: Evacuation plan for a hospital complex.

Modeling Fundamentals. The process of an evacuation is captured in the simplified flow chart of Fig. 1. There are essentially five phases which underly the evacuation process. The first and foremost is a warning bell or siren signaling the occupant population to leave. Unfortunately, one must react to the warning and recognize the problem at hand, so there is often a great deal of uncertainty associated with the second phase. Thirdly, after the warning is taken seriously, the occupants must decide to evacuate. The first three phases are highly uncertain and transient. Once the occupants decide to evacuate, the general evacuation process gets underway and this is where the evacuation plans should be followed. Finally, there is a verification phase, were one must account for all the occupants to ensure their safe arrival at the destination. As a constructive framework for this Chapter on evacuation networks, we establish that the modeling of evacuation problems has three fundamental steps:

1) Representation: How should a region, e.g. Fig. 2 or facility be represented or modeled?
2) Analysis: Given the model, how should analyze the evacuation of the occupants, i.e. a deterministic or stochastic evacuation process? What performance measures are crucial to measuring performance of the evacuation? and
3) Synthesis: How should one synthesize the results of the analysis step so as to best evacuate the occupants in light of the performance measures?

Representation Stage. Fig. 2 depicts a large *hospital* campus with many inter-connected buildings, many different levels, and a complex array of *circulation* passages, and illustrates that the evacuation problem is a difficult one to represent. However, one can begin to accurately model the evacuation process through a network as depicted in Fig. 3. By definition, an *evacuation network* (graph) $G(V,E)^{\ell}$ is comprised of a finite set V of nodes (vertices) of size N, where $V = \{V_1, \ldots, V_n\}$ together with a finite set E of arcs $e_k = (v_i, v_j)$, $\forall(i,j)$, nodal pairs and an indication of the level at which the network is defined ℓ. The levels actually correspond to the degree of aggregation inherent in modeling large complex networks. V can further be partitioned into three sets of nodes:

V_1) representing the occupant source nodes during the evacuation,

V_2) representing the intermediate nodes during the evacuation;

V_3) representing the sink or destination nodes of the occupants.

The set of arcs represent the different *streets*, passageways, or routes from V_1 to V_3. Associated with each node $\ell \in V$ and each arc $(v_i, v_j) \in E$ are variables and parameters which represent node and arc processing times, node and arc capacities, arrival times to the network, distances, and occupant population sizes at the source nodes.

Fig. 3 illustrates the example evacuation network with the key congested routes in the evacuation planning problem embedded in the network model.

The Representation Step is often defined in terms of the size and composition of the customer population: infinite, finite, or mixed and how the facility under study should be decomposed by V, E, and ℓ. The crucial link between the Represen-

tation and Analysis Steps is the complexity (i.e., number of nodes and arcs) of G^{ℓ}, which governs the number of equations used in the mathematical model in the Analysis Step. The Representation Step presents an interesting and challenging problem because of the many possible ways of representing regions, facilities, ships, vehicles, and building components.

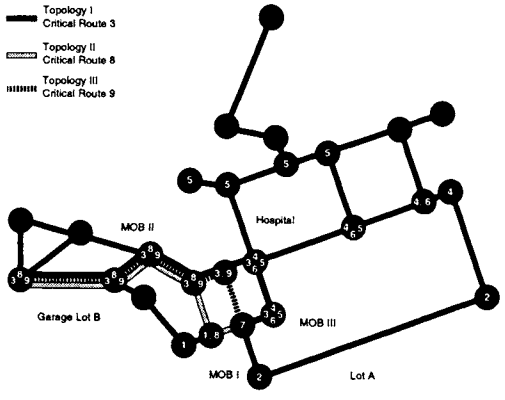

Fig. 3: Route site plan.

Analysis Step. The Analysis Step is the point at which the methodology and mathematical models underlying the flow processes, and the algorithmic structure for computing the performance characteristics of $G^{\ell}(Z, E)$ come together. Mathematically, we have a network $G(V, E)$, with a finite set of nodes V and edges(arcs) E over which multiple classes of customers (occupants) flow from source(s) to sink(s) while a vector of objective functions $\Omega = \{f_1(\overline{x}), \ldots, f_p(\overline{x})\}$ is simultaneously extremized subject to a set of constraints on the occupants flowing through the network. Fig. 4 captures many of the recognized criteria appropriate in analyzing a network evacuation problem. In our studies, we have often used Minimum Total Evacuation Time and Minimum Total Distance Travelled to capture the evacuation problem. The Total Distance travelled is a suitable surrogate objective for approaching the route complexity, since reducing the evacuation path length will often begin to capture the path complexity and, hopefully, minimizing this measure will abate the occupants sense of panic. Other objectives might be appropriate given the particular context or decision situation.

Synthesis Step. Given the performance characteristics determined during the Analysis Step, we can begin to optimize the network topology itself, routing and resource allocation problems within:

- *Topological Network Design* (TND): Determination of the number, type, and subset of nodes and arcs as well as the particular node and arc topology to be used for the evacuation.
- *Routing Network Design* (RND): Determination of the routing scheme in both steady-state and real time.
- *Capacitated Network Design* (CND): Determination of the Network Resources: Number of highway lanes, corridor length, widths, areas, landing shape, reception center capacity, configuration etc.

Mathematical Models. There are many possible mathematical modeling approaches once our network is constructed and Fig. 5 represents the range of approaches many research scientists have followed. References are provided for further details. The *boldface* text along the morphological tree represents the approach suggested in this article which we have applied in many different contexts.

Many mathematical models which have appeared in the literature for generating and evaluating evacuation paths for an occupant population [5], [2], [8].

Set Partitioning Model. The model which is presented below is a variation of one model appearing in [8]. It was one of the first to account for the critical features of the stochastic evacuation problem. Another class of models that one might utilize to formulate the problem are those of the class of multicommodity flow models. Unfortunately, these models will not control the Bernoulli splitting of the occupant population along the different evacuation paths which is problematic since splitting the different source populations will engender confusion and create a potential sense of panic among the evacuating occupants. The integer set partitioning programming model presented below has the desired property to control splitting of the flows.

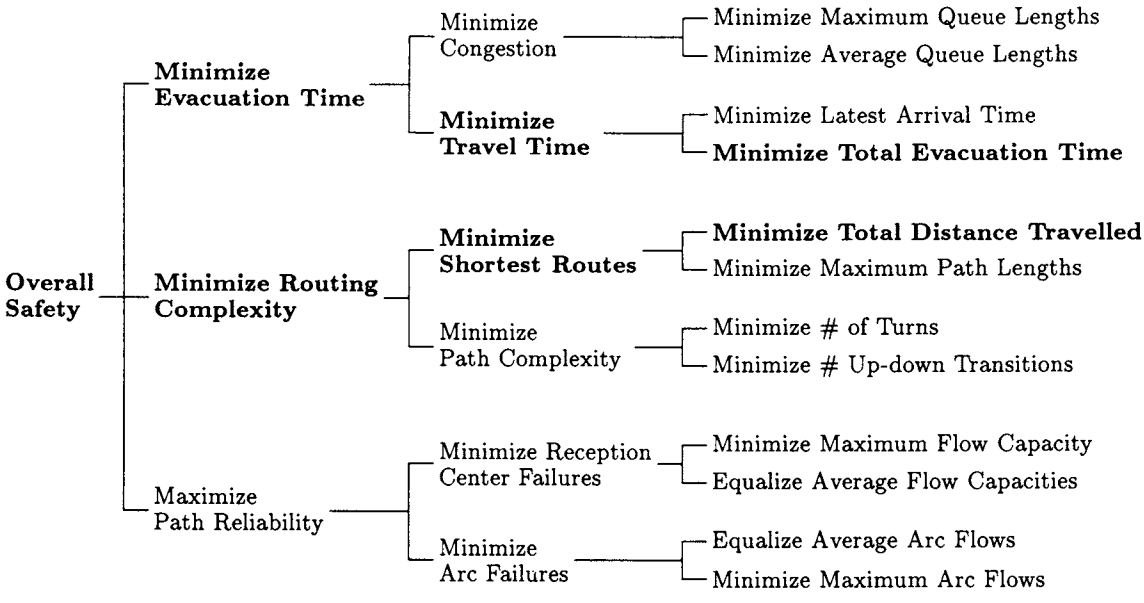

Fig. 4: Morphological diagram of multi-objective approaches.

The *multi-objective model* of our routing problem is:

$$\text{minimize } \{f_1(\overline{x}); f_2(\overline{x})\}$$

where the *Evacuation Time*, respectively the *Distance Travelled* are:

$$f_1(\overline{x}) = \sum_i \sum_j \sum_k q_{ijk}\lambda_{ijk}x_{ijk},$$

$$f_2(\overline{x}) = \sum_i \sum_j \sum_k d_{ijk}\lambda_{ijk}x_{ijk},$$

subject to:

- V_2 Arcs:

$$\sum_i \sum_j \sum_k \alpha_{\ell ijk}\lambda_{ijk}x_{ijk} \leq \rho_\ell, \quad \forall \ell,$$

- V_3 Sinks:

$$\sum_i \sum_j \sum_k p_{ijk}x_{ijk} \leq C_q, \quad \forall q,$$

- Occupant Classes:

$$\sum_k x_{ijk} = 1, \quad \forall i, j,$$

- Routes:

$$x_{ijk} = 0, 1, \quad \forall i, j, k,$$

and where:

- $x_{ijk} = 1$ if the ith occupant class from the jth source is assigned the kth route alternative.
- $\alpha_{\ell ijk}$ is a data coefficient which equals 1 if the ℓth arc is included in the ijkth route assignment and equals 0 otherwise.
- ρ_ℓ is the maximum allowable traffic along arc ℓ.
- C_q is the capacity of sink (destination) node q.
- p_{ijk} is the occupant population of source ij on the kth route alternative.
- q_{ijk} is the expected evacuation (sojourn) time of the ijkth occupant class. These values must be calculated from the particular stochastic model used in the evacuation study, see the discussion below.
- d_{ijk} is the average distance travelled for the ijkth occupant class.

Since we have two objective in our model, it makes sense to talk of the *NonInferior* (ni) set of route alternatives, since the *trade-offs* between f_1 and f_2 naturally underlie the optimal set of solutions we seek. Because of the complexity of solving this model directly, an alternative approach which systematically generates feasible routing alternatives to a relaxed version of our mathematical model but at the same time measures the critical objectives of evacuation time and distance trav-

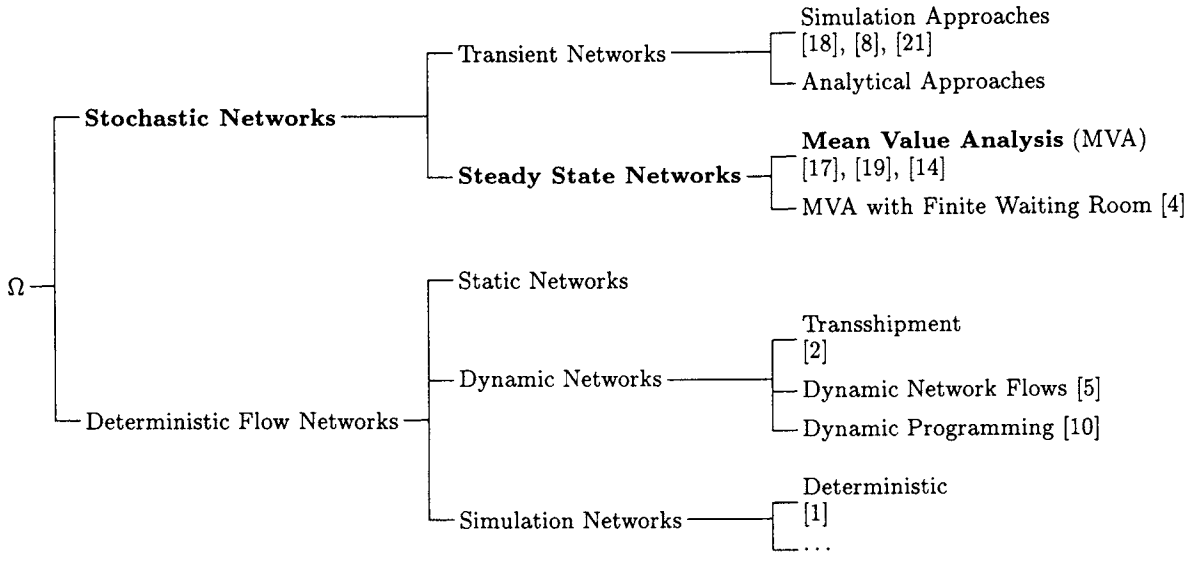

Fig. 5: Morphological diagram of EEP approaches.

elled is proposed and demonstrated in the next two sections.

Congestion Models. The real crux of the evacuation problem is to capture the *congestion* that naturally occurs when occupants choose the shortest routes to evacuate. There are some deterministic measures possible for measuring congestion, yet stochastic ones are the most accurate, because queueing is a nonlinear complex phenomenon.

Erlang Loss/Delay Networks. Fundamentally, each S_j node in the circulation network is an *M/G/C/C queue*, i.e. there is no waiting room and C depends on the square footage area of the circulation segment or the number of vehicles which can maximally occupy a highway segment [23]. Let's for the sake of the argument, focus on pedestrian evacuation. Later on we will show how our model extends to vehicular congestion. Each occupant in the circulation system consumes approximately $0.2m^2$ of floorspace, and, therefore, the capacity of a circulation system element is:

$$C = 5LW,$$

where L (length) and W (width) are given in meters.

Each circulation segment is a representative 'building block' for modeling pedestrian movements through the facility. Corridor segments, intersections, landings, stairwells, ramps, and so on represent a network of interconnected M/G/C/C queues. The separations of the circulation blocks are due to changes in flow direction, level, or merging and splitting decisions. Further, the cardinality of **S** depends on the configuration and complexity of movement patterns within the facility.

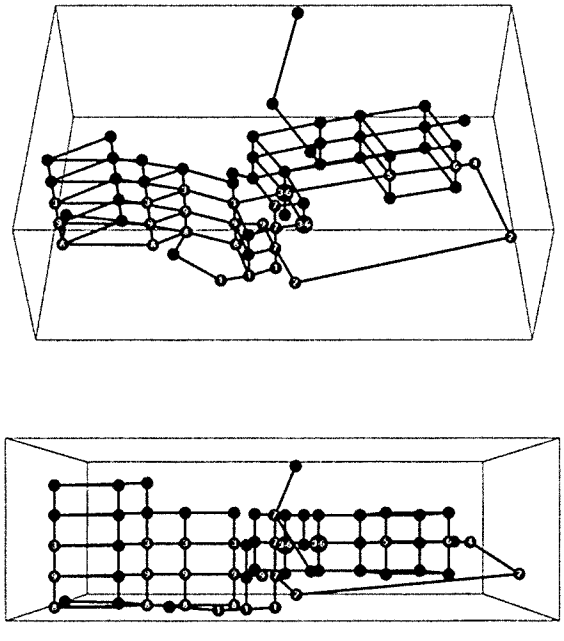

Fig. 6: Three-dimensional network models.

Flows through the nodes of **S**, the circulation system of a building are largely *state dependent*, in that a customer receives service in the circu-

lation node S_j and this service rate decays with increasing amounts of customer traffic.

Fig. 7 shows a family of curves which represent the variety of empirical studies (the curves in Fig. 7) that document the decay rate of the customer service rate as a function of population density in a corridor. Empirical models are also available showing distributions for stairs and other circulation elements with bi-, and multidirectional pedestrian flows [6], [20].

Finally, there are a set of classical linear and exponential curves which relate vehicle speed and vehicle density captured in Fig. 8. We have utilized these type of vehicular *speed/density relations* to develop state dependent models for vehicular traffic analysis [7].

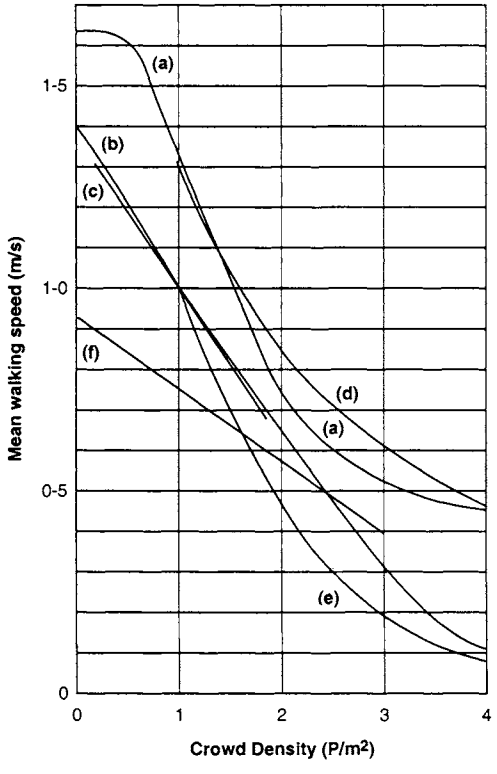

Fig. 7: Empirical distributions of pedestrian traffic flows.

In general, the service rate μ is a function of velocity v_i, which is a constant for each individual in the corridor. Thus, it takes t_i (seconds)

$$t_i = \frac{L}{v_i}$$

for each person to traverse the corridor, where i is the number of occupants in the circulation system when an individual enters.

Because of the complexity of dynamically updating the service rate as a function of the number of customers within a corridor segment, it becomes extremely difficult to utilize digital simulation models in the design of circulation systems within buildings. Our computational experience in digital simulation of access and egress networks underscores this defect in simulation models. We must, therefore, look to analytical models to aid the network design process if state dependent models are to be effectively utilized. Also, since we are examining the pedestrian/vehicular network as a design problem rather than as a control problem, it makes most sense to look at steady state measures rather than transient ones.

We have recently developed a generalized model of the M/G/C/C *Erlang loss queueing model* for service rate decay which can model any service rate distribution (linear, exponential, etc.) [3], [4], [15]. It is a special case of an Erlang loss model. F.P. Kelly [9] has treated M/G/C/C state dependent models in his book, but only ones with a linear, increasing function of the number of customers in the queue, whereas, we treat the queue with an nonlinear, decreasing service rate, see Fig. 3.

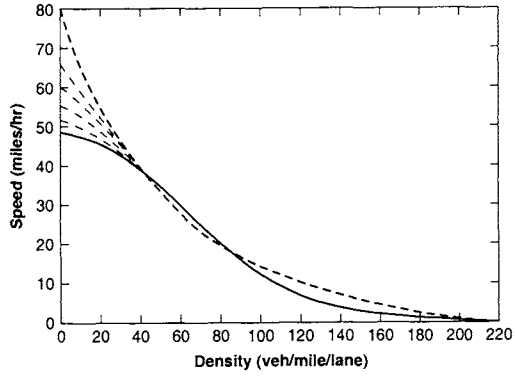

Fig. 8: Empirical distributions of pedestrian traffic flows.

Our M/G/C/C state dependent model dynamically models the flow rate of *pedestrians* within a corridor as a function of the population within the corridor. Suppose that G is a continuous distribution having density g and failure rate $\mu(t) = g(t)/G(t)$. Loosely speaking, $\mu(t)$ is the instantaneous probability intensity that a service t units old will end. The service rate depends on the number of customers in the system: given that there are n people in the system, each server processes work at rate $f(n)$. In other words, if there is an ar-

rival, the service rate will change to $f(n+1)$ and if there is a departure, the service rate will change to $f(n-1)$.

In particular, the probability distribution of the number of occupants in the corridor is given by:

$$\mathsf{P}(n \text{ in system}) = \frac{[\lambda \mathsf{E}(S)]^n \mathsf{P}_0}{n! f(n) \cdots f(1)}, \quad n = 1, \ldots, C,$$

where

$$\mathsf{P}_0 = \frac{1}{1 + \sum_{i=1}^{C} \frac{[\lambda \mathsf{E}(S)]^i}{i! f(i) \cdots f(1)}}$$

$$\mathsf{E}(S) = \frac{L}{1.5}, \quad f(n) = \frac{v_n}{v_1},$$

and $\mathsf{E}(S)$ is the mean service time of a lone occupant flowing through a corridor of length L, with service rate 1.5m/sec (see Fig. 3). The term v_n is defined as the average walking speed when n people are in the corridor.

For the M/G/C/C state dependent model, we have also shown that the departure process (including customers completing service and those that are lost) is a Poisson process with rate λ [3], [4].

Algorithms. The problem we face in our evacuation planning problem is that we do not know a priori which paths are NI without assessing the congestion in $G(V, E)$. We must iteratively generate candidate paths, assess the congestion in $G(V, E)$, and then iterate again until the desired trade-offs between distance travelled and evacuation time is acceptable to the planner. This iterative process leads to the algorithm described below. For product form networks where the estimate of time delays in the *Expected Savings* calculation for re-routing among the alternative Noninferior paths can be computed exactly, then the algorithm will guarantee finding a Noninferior path for re-routing the occupant classes. For nonproduct form networks, which are typically the case, we can only approximate these time delays, therefore, the algorithm can only guarantee an approximate Noninferior solution. Considering the complexity of the underlying stochastic-integer programming problem, this is a reasonable and practical strategy.

k-Shortest Paths. The algorithm to facilitate the design methodology can be incorporated into any simulation e.g. Q-GERT or analytical model e.g. QNET-C to estimate f_1, f_2, and carry out the evacuation planning/routing analysis. To summarize and focus the efforts in this article, an algorithmic description of Steps 1–3 and it substeps are presented.

1) Representation Step: Represent the underlying facility or region as a network $G(V, E)$ where V is a finite set of nodes and E is a finite set of arcs or nodal pairs.
2) Analysis Step: Analyze $G(V, E)$ as a *queueing network* either with a transient or *steady-state model* and compute the total evacuation time of the occupant population along with total distance travelled to evacuate given a set of evacuation paths.
3) Synthesis Step:
   3.1) Analyze the queueing output from the evacuation model and compute the set of NonInferior evacuation paths which simultaneously minimize time and distance travelled in $G(V, E)$ for each occupant population.
      3.1.1) If the set on NI paths are uniquely optimal, then

$$\mathsf{E}_{ij}^k = q_{ijk} - \left[\frac{d_{ij}^k}{\omega} + q_{ij}^k \right] \leq 0, \quad \forall i, j, k,$$

      go to Step 3.2, where:
      a) E_{ijk} is the net increase or decrease in the average egress time per person caused by re-routing occupants to the $(k+1)$st NonInferior route.
      b) q_{ijk} is the sum of the average queue times per person on the original route.
      c) d_{ij}^k is the increased distance travelled on the $(k+1)$st NonInferior route (e.g. if the kth NonInferior route is 100 feet and the $(k+1)$st NonInferior route is 120 feet, d_{ij}^k is equal to 20 feet, i.e. 120 − 100).

d) ω is the average travel speed for d_{ij}^k.

e) q_{ij}^k is the sum of the expected queue times per person on the $(k+1)$st NonInferior route.

otherwise:

3.1.2) Significant queueing (congestion) exists on one or more routes then go to Step 3.3.

3.2) STOP! The NI shortest time/distance routes are optimal and identical and total evacuation time, distance and congestion are minimized.

3.3) Determine the total number of occupants who pass through the queueing area(s) and trace them back to their origins.

3.4) Select the total number of occupants to be re-routed from each source node. The total number of occupants re-routed is correlated to both the size of the queues and the number of occupants on each route. In selecting the population, the analyst should strive to achieve uniformity of occupants and queues on each egress route.

3.5) Re-route the population to the kth route of the NI set of paths where k is selected by employing the following formula:

$$\mathsf{E}_{ij}^k = q_{ijk} - \left\lceil \frac{d_{ij}^k}{\omega} + q_{ij}^k \right\rceil, \quad \forall i,j,k.$$

3.6) Select the largest positive E^* for each set of populations to be re-routed, where:

$$\mathsf{E}^* = \max_{\forall i,j \text{ sources}} \{E_{11}, \ldots, E_{IJ}\}$$

for all possible savings, and then re-run the computer evacuation planning model with the new set of routes, by returning to Step 2.0 of the General Algorithm. If all $E_i's$ are negative, stop! The current set of NI shortest routes used on the previous iteration are selected.

Other Algorithms. Besides the *k-shortest path* approach, one might utilize a turn-penalty algorithm to guide the process of determining the evacuation paths. This is probably very appropriate in *vehicular* evacuation schemes. Also, another approach which seems quite viable, would be to define the set of arc disjoint paths, since this would tend to completely separate the occupant congestion along the paths. We have not experimented with these approaches to define the evacuation routes, but their use might be quite appropriate in the future.

Summary and Conclusion. We have given some insights into the performance modeling and optimization problems associated with evacuation networks. As the maturity of this application area grows, and more research is devoted to the area, then more theoretical and algorithmic issues and progress that emerge.

See also: **Minimum cost flow problem; Nonconvex network flow problems; Traffic network equilibrium; Network location: Covering problems; Maximum flow problem; Shortest path tree algorithms; Steiner tree problems; Equilibrium networks; Survivable networks; Directed tree networks; Dynamic traffic networks; Auction algorithms; Piecewise linear network flow problems; Communication network assignment problem; Generalized networks; Network design problems; Stochastic network problems: Massively parallel solution.**

References

[1] BERLIN, G.N.: 'A simulation model for assessing building firesafety', *Fire Techn.* **18**, no. 1 (1982), 66–76.

[2] CHALMET, L.G., FRANCIS, R.L., AND SAUNDERS, P.B.: 'Network models for building evacuation', *Managem. Sci.* **28**, no. 1 (1982), 86–105.

[3] CHEAH, JENYENG: 'State dependent queueing models', *Master's Thesis Dept. Industr. Engin. and Oper. Res. Univ. Massachusetts, Amherst MA 01003* (1990).

[4] CHEAH, JENYENG, AND MACGREGOR SMITH, J.: 'Generalized $M/G/C/C$ state dependent queueing models and pedestrian traffic flows', *Queueing Systems and Their Applications* **15** (1994), 365–386.

[5] FRANCIS, R.L., AND CHALMET, L.G.: 'Network models for building evacuation: A prototype primer', *Unpublished Paper Dept. Industr. Systems Engin. Univ. Florida, Gainesville, Florida* (1980).

[6] FRUIN, J.J.: *Pedestrian planning and design*, Metropolitan Assoc. Urban Designers and Environmental Planners, 1971.

[7] JAIN, R., AND MACGREGOR SMITH, J.: 'Modeling vehicular traffic flow using M/G/C/C state dependent queueing models', *Tansportation Sci.* **31**, no. 4 (1997), 324–336.

[8] KARBOWICZ, C.J., AND MACGREGOR SMITH, J.: 'A k-shortest path routing heuristic for stochastic evacuation networks', *Engin. Optim.* **7** (1984), 253–280.

[9] KELLY, F.P.: *Reversibility and stochastic networks*, Wiley, 1979.

[10] KOSTREVA, M., AND WIECEK, M.W.: 'Time dependency in multiple objective dynamic programming', *J. Math. Anal. Appl.* **173** (1993), 289–307.

[11] MACGREGOR SMITH, J.: 'The use of queueing networks and mixed integer programming to optimally allocate resources within a library layout', *JASIS* **32**, no. 1 (1981), 33–42.

[12] MACGREGOR SMITH, J.: 'An analytical queueing network computer program for the optimal egress problem', *Fire Techn.* **18**, no. 1 (1982), 18–37.

[13] MACGREGOR SMITH, J.: 'Queueing networks and facility planning', *Building and Environment* **17**, no. 1 (1982), 33–45.

[14] MACGREGOR SMITH, J.: 'QNET-C: An interactive graphics computer program for evacuation planning', in R. NEWKIRK (ed.): *Proc. Soc. for Computer Simulation Emergency Planning Session*, 1987, pp. 19–24.

[15] MACGREGOR SMITH, J.: 'State dependent queueing models in emergency evacuation networks', *Transport. Sci. B* **25B**, no. 6 (1991), 373–389.

[16] MACGREGOR SMITH, J., AND ROUSE, W.B.: 'Application of queueing network models to optimization of resource allocation within libraries', *JASIS* **30**, no. 5 (1979), 250–263.

[17] MACGREGOR SMITH, J., AND TOWSLEY, S.: 'The use of queueing networks in the evaluation of egress from buildings', *Environment & Planning B* **8** (1981), 125–139.

[18] STAHL, FRED I.: 'BFIRES-II: A behavior based computer simulation of emergency egress during fires', *Fire Techn.* **18**, no. 1 (1982), 49–65.

[19] TALEBI, K., AND MACGREGOR SMITH, J.: 'Stochastic network evacuation models', *Comput. Oper. Res.* **12**, no. 6 (1985), 559–577.

[20] TREGENZA, P.: *The design of interior circulation*, v. Nostrand Reinhold, 1976.

[21] WATTS, J.M.: 'Computer models for evacuation analysis. Paper presented at the SFPE Symposium': *Quantitative Methods for Life Safety Analysis*, College Park Maryland, 1986, Available from the Fire Safety Inst., Middlebury Vermont.

[22] WOODSIDE, C.M., AND HUNT, R.E.: 'Medical facilities planning using general queueing network analysis', *IEEE Trans.* **SMC-7**, no. 11 (1977), 793–799.

[23] YUHASKI, S., AND MACGREGOR SMITH, J.: 'Modeling circulation systems in buildings using state dependent queueing models', *Queueing Systems and Their Applications* **4** (1989), 319–338.

J. MacGregor Smith
Dept. Mechanical and Industrial Engin. Univ. Massachusetts
Amherst, Massachusetts 01003, USA
E-mail address: jmsmith@ecs.umass.edu

MSC2000: 90-XX
Key words and phrases: combinatorial optimization, evacuation network, congestion.

EVOLUTIONARY ALGORITHMS IN COMBINATORIAL OPTIMIZATION, *EACO*

Most of the *NP*-hard combinatorial optimization problems cannot be solved to optimality in practice. Therefore heuristic techniques have to be used to obtain solutions of high quality. There exists different approaches to design a heuristic algorithm, such as tabu search and genetic algorithm for example. The latter solution method belongs to a wider class of algorithms, called *evolutionary algorithms*, that handle a set of several solutions. Within this class, the best known algorithms that are applied to combinatorial optimization problems are genetic algorithms (cf. **Genetic algorithms**) and ant systems. For a general presentation, one can mention [22], [72] for genetic algorithms and [12], [23] for ant systems.

In this article, a review of the evolutionary algorithms used up to 1998 in combinatorial optimization is being made. For a certain number of combinatorial problems, the main papers that present an evolutionary algorithm for that problem are referenced, and some short remarks are given. While it is difficult to provide a very precise definition of an evolutionary algorithm, this term will be used here as a synonym of population-based algorithm: an algorithm that makes evolve several solutions, in particular by exchanging some kind of information between them. Algorithms that iteratively modify a solution in order to obtain a good one (like tabu search or genetic algorithms with a 'population' of size 1) will not be considered as evolutionary algorithms.

The Traveling Salesman Problem. The traveling salesman problem (or TSP) is probably the problem on which the largest number of evolutionary algorithms have been applied. It consists in determining a shortest tour visiting all of the given cities exactly once. A very complete survey of local search approaches to this problem has been

provided by D.S. Johnson and L.A. McGeoch [51], while J.-Y. Potvin [70] compared several genetic algorithms for TSP. In [51], the authors recommend different solving techniques depending on the quality of the solution desired and the time available. Genetic algorithms or ant systems are a good choice if enough running time is allowed and good solutions are needed. With similar running times, the iterated Lin–Kernighan algorithm (or ILK) yields better results but is more complex to implement. In ILK, a single solution instead of a population of individuals is considered and this method will therefore not be referred to as an evolutionary algorithm. If there is no restriction on the running time, the best results can be obtained by genetic algorithms based on ILK.

An important breakthrough in the field of evolutionary algorithms for the TSP was the paper [67] by H. Mühlenbein, M. Gorges-Schleuter and O. Krämer. In their algorithm, implemented on a parallel machine, a solution was allowed to mate only with certain other solutions and some optimization technique was applied to the offsprings. Indeed, the use of a local search algorithm to improve created offsprings is a necessary condition for an evolutionary algorithm to be efficient. Moreover, they designed a crossover specific to the TSP, called MPX (maximum preservative crossover). It consists in copying a segment of a certain length from a first parent into the offspring and adding cities consecutively from the second parent according to some rules. This crossover is very suitable for the TSP, as shown in [66]. Further researches studied the impact of the different elements on the results and improved the quality of the solutions obtained [44], [7], [89]. Several other crossovers, most of them using two parents, have been suggested by various authors. In particular, B. Freisleben and P. Merz proposed [37], [38] the distance preserving crossover (or DPX): An offspring is created by keeping the edges that are found in both parents, and greedily reconnecting the different pieces without using the edges contained in only one parent. They obtain a very efficient algorithm, that won both the ATSP (asymmetric TSP) and the TSP competitions at the First International Contest in Evolutionary Optimization [6]. They further improved their algorithm, in terms of speed and quality of solutions, in [39]. Their use of an edge-preserving crossover and of a hill-climbing algorithm illustrates important elements necessary to obtain an efficient genetic algorithm for TSP. These elements have been put forward in different comparisons between various genetic algorithms for TSP [78], [70], together with the necessity to split the population into several subpopulations for solving large instances (more than a few hundred cities).

The first presentation of ant colony optimization (ACO) [12] was made with the TSP as illustration and this problem remains the most often used application problem of works on ant colony optimization. The initial ACO system, named ant system, has been extended to what is called ant colony system (ACS). A description of this algorithm can be found in [23] by M. Dorigo and L.M. Gambardella. In the same paper, local search has been added to ACS and the resulting algorithm has been applied to ATSP and TSP. The results reported are better in [39] for TSP, but are better in [23] for ATSP. Another proposed extension of ant system, called *MAX-MIN ant system* [79], consists in introducing explicit maximum and minimum values for the trail factors on the arcs. Good results are obtained with such an algorithm when local search is added.

The Vehicle Routing Problem. The most studied extension of the *vehicle routing problem* (VRP) is the one with time windows (VRPTW). In order to solve this problem, a two-phase heuristic, called *GIDEON*, has been proposed in [84]. The first phase uses a genetic algorithm to cluster the customers, and the solutions obtained are improved by local optimization techniques in the second phase. This procedure has first been improved in [83], and then extended in [85]. In this last paper, S.R. Thangiah, I.H. Osman and T. Sun present several metaheuristics, all having a first phase similar to the one in GIDEON. These algorithms have been compared to several other heuristics and showed very good results on test problems taken from the literature. Some improvements have still to be brought for solving problems with large time windows. For such problems, a heuristic based on *simulated annealing* and a population-based algo-

rithm called *GENEROUS* [71] are shown to be a little more efficient. The latter is not a standard genetic algorithm since it does not represent solutions by chromosomes, but it nevertheless handles several solutions and uses a recombination operator. An *adaptive memory* procedure, in conjunction with tabu search, has also been applied to this problem [75].

Improvements of the GIDEON approach with local post-optimization procedures have also been used for the VRP with time deadlines. A comparison done in [87], [86] with two other heuristics shows that the cluster-first route-second algorithm with a genetic algorithm in the first phase performs well for problems in which the customers are distributed uniformly and/or with short time deadlines.

The Quadratic Assignment Problem. The *quadratic assignment problem* (or QAP) allows the modelization of many practical problems in location science, but can be solved optimally only for very small instances. Therefore different heuristics have been proposed for this problem. Several of them are compared in [13], [81]. For real-world problems (irregular and structured), the genetic hybrid by C. Fleurent and J.A. Ferland in [33] appears to be one of the most efficient algorithms [81]. Based on a standard genetic algorithm with solutions encoded as permutations [82], this genetic hybrid applies a robust tabu search on the offsprings and was able to find several new best solutions on some benchmark problems.

The ant colony optimization approach has also been considered, first in [64]. This ant system algorithm, hybridized with a local search, has been improved in [63, 62] and provides very good results. A different ACO approach, where at each iteration the solutions are modified instead of newly constructed, has been proposed in [40]. This algorithm, also hybridized with a local search procedure, yields better results on real-world problems than the genetic hybrid of [33], but is not competitive on random problems. A further promising method, based on *scatter search*, has been presented in [19].

The Satisfiability Problem (SAT). The problem of finding a truth assignment for variables to make a propositional formula true is probably the best known, and historically the first, *NP*-complete problem. But only few evolutionary algorithms for SAT can be found in the literature. After a straightforward approach in [52], a rather different solution representation has been proposed in [45]. But the drawback of this method, despite adapted operators, is that it increases the size of the individuals in an important way, compared to the coding 'one gene for one variable'. This last coding has been used in [35], together with a SAT-adapted crossover (the objective function being simply the number of satisfied clauses). But the evolutionary algorithm thus obtained was not able to compete with a tabu search (also presented in [35]). The tabu search-genetic hybrid (where some iterations of tabu search is used for mutation) is computationally expensive, but is able to solve large instances that a tabu search alone cannot solve. For smaller instances, the hybridization is not useful.

Another heuristic approach to SAT consists in assigning weights to the different clauses and minimizing the sum of the weights of the unsatisfied clauses. These weights are adapted during the algorithm depending on the 'difficulty' of each constraint. This mechanism has been used in evolutionary algorithms in [25] and [90], but in both cases it came out that the best results are obtained with a 'population' of size 1. Such an algorithm is therefore no longer considered as an evolutionary algorithm.

The Set Covering and Set Partitioning Problems. The *set covering problem* (SCP) is a zero-one integer programming problem where the constraints are all of the type $\sum_j a_{ij}x_j \geq 1$ with zero-one coefficients. It is a well-known problem, that has also been used to study penalty functions in genetic algorithms [74], [3].

Different genetic algorithms approaches have been proposed in the literature (see for example [60], [61], [50]), and a very efficient one has been presented by J.E. Beasley and P.C. Chu in [5]. This algorithm uses binary representation of the solutions, and a repair operator to preserve the feasibility of the individuals and to improve the

solutions. Moreover, a variable mutation rate has been introduced. Results on standard test problems up to 1000 constraints and 10,000 variables show the efficiency of this algorithm that was able to improve the best-known result on some of the larger instances. The same paper shows no significant difference between various crossovers.

The *set partitioning problem* (SPP) is also a zero-one integer programming problem, the difference with SCP being that the constraints are equalities instead of inequalities. Relatively few heuristics have been developed for this problem. D. Levine investigated sequential and parallel genetic algorithms for SPP [59]. His best algorithm was a genetic algorithm in an island model, hybridized with a local search heuristic. But this algorithm remained less efficient, both in terms of quality of the solutions and in terms of running time, than the branch and cut approach of [49]. Some problems met by his algorithm were due to the penalty term for infeasible solutions in the fitness function. In order to overcome these problems, other authors decomposed the single fitness measure in two distinct parts (the objective function and a measure of 'infeasibility') [10]. Adapting the parent selection method to this modification, and also using an improvement operator, they obtained a better genetic algorithm, but that is still not able, for the problems they considered, to compete with a commercial mixed integer solver.

The Knapsack Problem. The multidimensional (zero-one) *knapsack problem* is equivalent to the zero-one integer programming problem with non-negative coefficients. Only few papers tried to solve this problem with evolutionary algorithms. While the first such algorithms did not give high-quality results and were not competitive with other heuristics [56], [88], the quality has improved. Genetic algorithms as presented in [11], [48], both working only with feasible solutions, are able to obtain optimal solutions on standard test problems (instances with at most 105 variables and 30 constraints). In [11], Chu and Beasley proposed some larger test problems (up to 500 variables and 30 constraints), without known optimal solution, and used them for a comparison with other heuristics. Their genetic algorithm uses a 'repair' operator specific to this problem to ensure good feasible offsprings and obtained high-quality results, but needed also more computation time (on a same machine, about one hour for the genetic algorithm against a few seconds for the other heuristics).

The Bin Packing Problem. The standard one-dimensional *bin packing problem* consists in putting items of given sizes in bins of given capacity. Many evolutionary algorithms proposed for this problem (genetic algorithms and evolution strategy, see for example [77], [16], [57]) performed worse than a simple heuristic like first fit decreasing. E. Falkenauer and A. Delchambre then suggested in [30] a genetic algorithm designed for grouping problems: the *grouping genetic algorithm* (GGA). In this algorithm, solutions are represented by chromosomes having two parts: the item part encodes for each item its bin and the group part, of variable length, encodes the bin identifiers used. The crossover, mutation and inversion operators have been adapted to this encoding. Instead of simply using the number of bins, the authors designed a fitness function that also takes into account the proportion to which each bin is filled. With this approach, they obtained very satisfactory results. The arguments presented for this new encoding are discussed by C. Reeves in [73]. In the same paper, a hybrid genetic algorithm is presented, where solutions are represented by permutations and decoded using heuristics like first fit and best fit. The results obtained are more or less similar to those in [30]. A problem size reduction heuristic, similar to the reduction process used in [16], has also been introduced in this genetic algorithm. According to Falkenauer [29], this reduction violates the search strategy of the genetic algorithm and he therefore prefers the GGA's crossover, that has the same goal of propagating promising bins. In the same paper, the GGA is improved by the introduction of local optimization inspired by the dominance criterion of [65]. The new algorithm is compared with an efficient branch and bound algorithm and gives excellent results.

Extensions of the standard bin packing problem, like the two-dimensional bin packing problem, have also been considered with evolutionary algo-

rithms [77], [15], [69]. An overview of these variations is presented in [43].

Graph Coloring. The *graph coloring problem* is a well-known problem in graph theory; it consists in determining the smallest number of colors that must be used to color the vertices of a graph such that two adjacent vertices do not have the same color. L. Davis is the first author who proposed an evolutionary algorithm for this problem [22]. In fact, he considered a graph with weights on the vertices and an integer k. He then designed a hybrid genetic algorithm for finding a partial k-coloring such that the colored vertices have maximum total weight. In this algorithm, individuals are represented as permutations of the vertices of the graph. This order-based encoding is not very efficient, as shown by Fleurent and Ferland in [34]. In this paper, they also present hybrid genetic algorithms that use string-based encodings of the solutions for finding a coloring in k colors with as few conflicting edges (edges with both ends of the same color) as possible. They consider different crossovers, including a graph-adapted one, and hybridize the genetic algorithm with a simple local search or with tabu search (a modified version of [46]). The results on random graphs $G_{n,0.5}$ improve the previous best results. For graphs up to 300 vertices, their tabu search-genetic hybrid and their tabu search give similar results, but in much less time for the latter. For larger graphs (500 or 1000 vertices), the running time becomes prohibitive, and both the evolutionary algorithm and the tabu search must be used within a different approach (determining large stable sets and coloring the residual graph). The tests on 450-vertices Leighton graphs (with known chromatic numbers) showed that the tabu search-genetic hybrid outperforms the tabu search on about half of the instances, while the opposite is true for the remaining instances. The hybrid algorithm was able to find an optimal solution for two instances (out of twelve) that could not be solved by the tabu search alone.

Another evolutionary algorithm has been proposed in [18], with a graph-adapted crossover that takes into account how 'close' a vertex is to conflicting edges. The improving algorithm applied to offsprings is a steepest descent method, instead of a tabu search like in [34]. Despite this less sophisticated method, their algorithm gives similar results to those obtained by the hybrid algorithm in [34]. Moreover, the latter gives worser results when the tabu search is replaced by a simple descent method.

Concerning ant colony optimization, a first approach to graph coloring has been proposed in [17], but the results obtained need improvements.

Other Graph Problems.

Maximum Clique. The problem of determining the *maximum clique* (complete subgraph) in a graph is equivalent to the problem of determining the minimum vertex cover or the maximum stable set in the complementary graph. A first genetic algorithm, hybridized with a tabu search, has been proposed by Fleurent and Ferland in [35], but they show that their tabu search alone gives similar results in a shorter time. In these algorithms, a solution is a set of vertices of given size and the objective function measures how many edges are missing for a set to be a clique. Improving an algorithm of [2], E. Balas and W. Niehaus [4] proposed a genetic algorithm (without improving algorithm applied to the offsprings) for both the maximum cardinality and maximum weight clique problems where an individual is a clique. In this algorithm, the recombination operation ('crossover') used is designed specifically for this problem and taken from another heuristic. The results obtained on the DIMACS benchmark graphs are very good, similar to those obtained in [35] from the point of view of the solutions' quality. A different fitness function has been suggested in [8] and included in a hybrid genetic algorithm using a local optimization step. The fitness value associated to a set of vertices is a weighted combination of the size of the set and the number of edges missing to have a clique, but the weights are modified during the run of the algorithm according to a simple rule. Despite the introduction of a preprocessing step that determines the order of the vertices on the chromosome, this algorithm is less efficient (but this may be due to the use of the 2-point crossover).

Graph Partitioning. Evolutionary algorithms are rather seldom used to tackle the *k-way graph partitioning problem* (partitioning a (weighted) graph in k equal-sized parts), even if the graph bisectioning problem (the case $k = 2$) is sometimes taken to illustrate various ingredients in genetic algorithms ([9], [54]). For the general k-way graph partitioning problem, different problem-oriented operators are introduced and studied in a parallel genetic algorithm in [58]. In this algorithm, the population is only composed of feasible solutions. Another approach has been proposed in [76] where the population is split in two halves: one containing only feasible solutions and the other only infeasible ones. This algorithm uses the same encoding scheme and crossover operator as [58], but has not been applied on similar instances of the problem. In a general way, genetic algorithms give good results on partitioning problems, but at a very high computational cost.

Miscellaneous.

Sequencing and Scheduling. The best-known *sequencing and scheduling problems* are the flow-shop, job-shop and open shop problems. The first paper applying an evolutionary algorithm to such a problem is [21]. Later, several other genetic algorithms have been proposed ([36], [80] for example). One of the first efficient evolutionary algorithm for *job-shop* problems has been presented in [68] and improved in [91], [20]. Comparisons done with other heuristics on benchmark problems show that sophisticated genetic algorithms (with the use of problem-adapted crossovers and hybridization) yield the best results for *flow-shop* and job-shop problems [1], [24], [42]. The *open shop problems* have less attracted researchers of the evolutionary algorithms' field, but a genetic algorithm has been proposed in [32], [31]. An ant colony approach of job-shop problems has also been tested, in [14], but gave worse results than known genetic algorithms.

Steiner Trees. Only very few works deal with *Steiner trees* and evolutionary algorithms. Moreover, they consider different variants of this problem. The first paper [47] proposes a genetic algorithm with local optimization for determining minimum Steiner trees in the Euclidean plane. A solution is represented by the coordinates of the Steiner points. A comparison with *simulated annealing* and the Rayward-Smith–Care algorithm shows no significant differences. The problem of the rectilinear Steiner problem has been addressed in [53] with a specific coding and an adapted crossover. The minimal Steiner tree problem in graphs has attracted a little more interest. A standard genetic algorithm (with bit strings as chromosomes) that gave good results on the sparse graphs tested has been proposed in [55]. Later, H. Esbensen and P. Mazumder [28] designed a genetic algorithm in which the encoding method is based on the distance network heuristic. Improvements have been brought in [26] and [27], where there is also a comparison between different algorithms. But this genetic algorithm is not competitive with an efficient tabu search as presented in [41].

Conclusion. In this paper, some references on the evolutionary approaches that have been proposed up to 1998 for different combinatorial problems have been given. A general remark that can be made on these solution methods is that evolutionary algorithms in general, and genetic algorithms in particular, are not efficient for such problems if implemented too naively. To obtain an algorithm with good performances, it is necessary to make adjustments of the basic method. Moreover, knowledge about the problem considered is very often also needed, in order to design adapted operators.

Another remark concerns their competitivity compared to other heuristic methods. While evolutionary algorithms can quite easily be adapted to (almost) any problem, their running time is often quite high. Local search algorithms, like tabu search or simulated annealing, can also be adapted to the different combinatorial problems quite easily. If they are designed in an intelligent way, they are very often able to obtain better results than evolutionary algorithms. Moreover, they are usually faster. For some problems, specifically designed heuristics can use theoretical results about this problem, allowing them to obtain good results. In general, evolutionary algorithms are not competitive against (extended) local search or specific

algorithms for small to medium size instances of combinatorial problems.

But this does not mean that population-based algorithms are not useful. In fact, the different approaches have various (dis)advantages, and the efficient algorithms that will be developed in the future will probably mix these different approaches. Such algorithms are usually called *'hybrid algorithms'* and have already been proposed for example for the traveling salesman problem [39] or the quadratic assignment problem [33], demonstrating their potentials.

See also: **Fractional combinatorial optimization**; **Replicator dynamics in combinatorial optimization**; **Neural networks for combinatorial optimization**; **Combinatorial matrix analysis**; **Multi-objective combinatorial optimization**; **Combinatorial optimization games**.

References

[1] AARTS, E.H.L., LAARHOVEN, P.J.M. VAN, LENSTRA, J.K., AND ULDER, N.L.J.: 'A computational study of local search algorithms for job shop scheduling', *ORSA J. Comput.* **6** (1994), 118–125.

[2] AGGARWAL, C.C., ORLIN, J.B., AND TAI, R.P.: 'An optimized crossover for maximum independent set', *Oper. Res.* **45** (1995), 226–234.

[3] BÄCK, T., SCHÜTZ, M., AND KHURI, S.: 'A comparative study of a penalty function, a repair heuristic, and stochastic operators with the set-covering problem', in J.M. ALLIOT, E. LUTTON, E. RONALD, M. SCHOENHAUER, AND D. SNYERS (eds.): *Artificial Evolution: European Conf.*, Vol. 1063 of *Lecture Notes Computer Sci.*, Springer, 1996, pp. 3–20.

[4] BALAS, E., AND NIEHAUS, W.: 'Optimized crossover-based genetic algorithms for the maximum cardinality and maximum weight clique problems', *J. Heuristics* **4** (1998), 107–122.

[5] BEASLEY, J., AND CHU, P.: 'A genetic algorithm for the set covering problem', *Europ. J. Oper. Res.* **94** (1996), 392–404.

[6] BERSINI, H., DORIGO, M., LANGERMAN, S., SERONT, G., AND GAMBARDELLA, L.M.: 'Results of the first international contest on evolutionary optimisation (1st ICEO)': *Proc. 1996 IEEE Internat. Conf. Evolutionary Computation*, IEEE Press, 1996, pp. 611–615.

[7] BRAUN, H.: 'On solving travelling salesman problems by genetic algorithms', in H.-P. SCHWEFEL AND R. MÄNNER (eds.): *Parallel Problem Solving from Nature*, Vol. 496 of *Lecture Notes Computer Sci.*, Springer, 1991, pp. 129–133.

[8] BUI, T.N., AND EPPLEY, P.H.: 'A hybrid genetic algorithm for the maximum clique problem', in L.J. ESHELMAN (ed.): *Proc. 6th Internat. Conf. Genetic Algorithms*, Morgan Kaufmann, 1995, pp. 478–484.

[9] BUI, T.N., AND MOON, B.R.: 'On multi-dimensional encoding/crossover', in L.J. ESHELMAN (ed.): *Proc. 6th Internat. Conf. Genetic Algorithms*, Morgan Kaufmann, 1995, pp. 49–56.

[10] CHU, P.C., AND BEASLEY, J.E.: 'Constraint handling in genetic algorithms: the set partitioning problem', *J. Heuristics* **4** (1998), 323–357.

[11] CHU, P.C., AND BEASLEY, J.E.: 'A genetic algorithm for the multidimensional knapsack problem', *J. Heuristics* **4** (1998), 63–86.

[12] COLORNI, A., DORIGO, M., AND MANIEZZO, V.: 'Distributed optimization by ant coloniess', in F. VARELA AND P. BOURGINE (eds.): *Proc. ECAL91 - European Conf. Artificial Life*, Elsevier, 1991, pp. 134–142.

[13] COLORNI, A., DORIGO, M., AND MANIEZZO, V.: 'Algodesk: An experimental comparison of eight evolutionary heuristics applied to the quadratic assignment problem', *Europ. J. Oper. Res.* **81** (1995), 188–205.

[14] COLORNI, A., DORIGO, M., MANIEZZO, V., AND TRUBIAN, M.: 'Ant system for job-shop scheduling', *JORBEL - Belgian J. Oper. Res., Statist. and Computer Sci.* **34**, no. 1 (1994), 39–53.

[15] CORCORAN, A.L., AND WAINWRIGHT, R.L.: 'A genetic algorithm for packing in three dimensions': *Proc. 1992 ACM/SIGAPP Symposium on Applied Computing SAC'92*, ACM, 1992, pp. 1021–1030.

[16] CORCORAN, A.L., AND WAINWRIGHT, R.L.: 'A heuristic for improved genetic bin packing', *Techn. Report UTULSA-MCS-93-08, Univ. Tulsa, USA* (1993).

[17] COSTA, D., AND HERTZ, A.: 'Ants can color graphs', *J. Oper. Res. Soc.* **48** (1997), 295–305.

[18] COSTA, D., HERTZ, A., AND DUBUIS, O.: 'Embedding a sequential procedure within an evolutionary algorithm for coloring problems in graphs', *J. Heuristics* **1** (1995), 105–128.

[19] CUNG, V.-D., MAUTOR, TH., MICHELON, PH., AND TAVARES, A.: 'A scatter search based approach for the quadratic assignment problem': *Proc. 1997 IEEE Internat. Conf. Evolutionary Computation*, IEEE Press, 1997, pp. 190–206.

[20] DAVIDOR, Y., YAMADA, T., AND NAKANO, R.: 'The ecological framework II: Improving GA performance with virtually zero cost', in S. FORREST (ed.): *Proc. 5th Internat. Conf. Genetic Algorithms*, Morgan Kaufmann, 1993, pp. 171–176.

[21] DAVIS, L.: 'Job shop scheduling with genetic algorithms', in J.J. GREFENSTETTE (ed.): *Proc. 1st Internat. Conf. on Genetic Algorithms*, Lawrence Erlbaum Ass., 1985, pp. 136–140.

[22] DAVIS, L.: *Handbook of genetic algorithms*, v. Nostrand Reinhold, 1991.

[23] DORIGO, M., AND GAMBARDELLA, L.M.: 'Ant colony

system: A cooperative learning approach to the traveling salesman problem', *IEEE Trans. Evolutionary Computation* **1** (1997), 53–66.

[24] DUVIVIER, D., PREUX, PH., AND TALBI, E.-G.: 'Stochastic algorithms for optimization and application to job-shop-scheduling', *Techn. Report LIL-95-5, Univ. du Littoral, France* (1995).

[25] EIBEN, A.E., AND HAUW, J.K. VAN DER: 'Solving 3-SAT with adaptive genetic algorithms': *Proc. 4th IEEE Conf. Evolutionary Computation*, IEEE Press, 1997, pp. 81–86.

[26] ESBENSEN, H.: 'Computing near-optimal solutions to the Steiner problem in a graph using a genetic algorithm', *Networks* **26** (1995), 173–185.

[27] ESBENSEN, H.: 'Finding (near-)optimal Steiner trees in large graphs', in L.J. ESHELMAN (ed.): *Proc. 6th Internat. Conf. on Genetic Algorithms*, Morgan Kaufmann, 1995, pp. 485–491.

[28] ESBENSEN, H., AND MAZUMDER, P.: 'A genetic algorithm for the Steiner problem in a graph', *Techn. Report Univ. Michigan, Ann Arbor* (1993).

[29] FALKENAUER, E.: 'A hybrid grouping genetic algorithm for bin packing', *J. Heuristics* **2** (1996), 5–30.

[30] FALKENAUER, E., AND DELCHAMBRE, A.: 'A genetic algorithm for bin packing and line balancing': *Proc. 1992 IEEE Internat. Conf. on Robotics and Automation*, IEEE Computer Soc. Press, 1992, pp. 1186–1192.

[31] FANG, H.-L.: 'Genetic algorithms in timetabling and scheduling', *PhD Thesis, Univ. Edinburgh* (1994).

[32] FANG, H.-L., ROSS, P., AND CORNE, D.: 'A promising genetic algorithm approach to job-shop scheduling, re-scheduling, and open-shop scheduling problems', in S. FORREST (ed.): *Proc. 5th Internat. Conf. Genetic Algorithms*, Morgan Kaufmann, 1993, pp. 375–382.

[33] FLEURENT, C., AND FERLAND, J.A.: 'Genetic hybrids for the quadratic assignement problem', in P.M. PARDALOS AND H. WOLKOWICZ (eds.): *Quadratic assignment and related problems*, DIMACS 16, Amer. Math. Soc., 1994, pp. 190–206.

[34] FLEURENT, C., AND FERLAND, J.A.: 'Genetic and hybrid algorithms for graph coloring', in G. LAPORTE AND I.H. OSMAN (eds.): *Metaheuristics in combinatorial optimization*, Vol. 63 of *Ann. Oper. Res.*, Baltzer, 1996, pp. 437–461.

[35] FLEURENT, C., AND FERLAND, J.A.: 'Object-oriented implementation of heuristic search methods for graph coloring, maximum clique, and satisfiability', in D.S. JOHNSON AND M.A. TRICK (eds.): *Cliques, coloring, and satisfiability*, Amer. Math. Soc., 1996, p. 619.

[36] FOX, B.R., AND MCMAHON, M.B.: 'Genetic operators for sequencing problems', in G.J.E. RAWLINS (ed.): *Foundations of Genetic Algorithms*, Morgan Kaufmann, 1991, pp. 284–300.

[37] FREISLEBEN, B., AND MERZ, P.: 'A genetic local search algorithm for solving symmetric and asymmetric traveling salesman problems': *Proc. 1996 IEEE Internat. Conf. on Evolutionary Computation*, IEEE Press, 1996, pp. 616–621.

[38] FREISLEBEN, B., AND MERZ, P.: 'New genetic local search operators for the traveling salesman problem', in H.-M. VOIGT, W. EBELING, I. RECHENBERG, AND H.-P. SCHWEFEL (eds.): *Proc. 4th Conf. on Parallel Problem Solving from Nature*, Vol. 1141 of *Lecture Notes Computer Sci.*, Springer, 1996, pp. 890–899.

[39] FREISLEBEN, B., AND MERZ, P.: 'Genetic local search for the TSP: new results': *Proc. 1997 IEEE Internat. Conf. on Evolutionary Computation*, IEEE Press, 1997, pp. 159–164.

[40] GAMBARDELLA, L.-M., TAILLARD, E.D., AND DORIGO, M.: 'Ant colonies for the quadratic assignment problems', *J. Oper. Res. Soc.* **50** (1999), 167–176.

[41] GENDREAU, M., LAROCHELLE, J.-F., AND SANSÓ, B.: 'A tabu search heuristic for the Steiner tree problem', *GERAD G-98-01, Univ. Montréal, Canada* (1998).

[42] GLASS, C.A., AND POTTS, C.N.: 'A comparison of local search methods for flow shop', in G. LAPORTE AND I.H. OSMAN (eds.): *Metaheuristics in combinatorial optimization*, Vol. 63 of *Ann. Oper. Res.*, Baltzer, 1996, pp. 489–509.

[43] GOODMAN, E.D., TETELBAUM, A.Y., AND KURECHIK, V.M.: 'A genetic algorithm approach to compaction, bin packing and nesting problems', *Techn. Report GARAGe94-4, Michigan State Univ.* (1994).

[44] GORGES-SCHLEUTER, M.: 'Asparagos: An asynchronous parallel genetic optimization strategy', in J.D. SCHAFFER (ed.): *Proc. 3rd Internat. Conf. on Genetic Algorithms*, Morgan Kaufmann, 1989, pp. 422–427.

[45] HAO, J.K.: 'A clausal genetic representation and its related evolutionary procedures for satisfiability problems', in D.W. PEARSON, N.C. STEELE, AND R.F. ALBRECHT (eds.): *Proc. 2nd Internat. Conf. on Artificial Neural Networks and Genetic Algorithms*, Springer, 1995, pp. 289–292.

[46] HERTZ, A., AND WERRA, D. DE: 'Using tabu search techniques for graph coloring', *Computing* **39** (1987), 345–351.

[47] HESSER, J., MÄNNER, R., AND STUCKY, O.: 'On Steiner trees and genetic algorithms', in J.D. BECKER, I. EISELE, AND F.W. MÜNDEMANN (eds.): *Parallelism, Learning, Evolution*, Vol. 565 of *Lecture Notes Artificial Intelligence*, Springer, 1991, pp. 509–525.

[48] HOFF, A., LÖKKETANGEN, A., AND MITTET, I.: 'Genetic algorithms for 0/1 multidimensional knapsack problems', *Proc. Norsk Informatik Konferanse, NIK '96* (1996).

[49] HOFFMAN, K., AND PADBERG, M.: 'Solving airline crew-scheduling problems by branch-and-cut', *Managem. Sci.* **39** (1993), 657–682.

[50] HUANG, W.-C., KAO, C.-Y., AND HORNG, J.-T.: 'A genetic algorithm approach for set covering problems': *Proc. First IEEE Internat. Conf. on Evolution-*

ary Computation, IEEE Press, 1994, pp. 569–574.

[51] JOHNSON, D.S., AND MCGEOCH, L.A.: 'The traveling salesman problem: A case study in local optimization', in E.H.L. AARTS AND J.K. LENSTRA (eds.): *Local Search in Combinatorial Optimization*, Wiley, 1997, pp. 215–310.

[52] JONG, K.A. DE, AND SPEARS, W.M.: 'Using genetic algorithms to solve NP-complete problems', in J.D. SCHAFFER (ed.): *Proc. 3rd Internat. Conf. on Genetic Algorithms*, Morgan Kaufmann, 1989, pp. 123–132.

[53] JULSTROM, B.A.: 'A genetic algorithm for the rectilinear Steiner problem', in S. FORREST (ed.): *Proc. 5th Internat. Conf. Genetic Algorithms*, Morgan Kaufmann, 1993, pp. 474–480.

[54] KAHNG, A.B., AND MOON, B.R.: 'Toward more powerful recombinations', in L.J. ESHELMAN (ed.): *Proc. 6th Internat. Conf. on Genetic Algorithms*, Morgan Kaufmann, 1995, pp. 96–103.

[55] KAPSALIS, A., RAYWARD-SMITH, V.J., AND SMITH, G.D.: 'Solving the graphical Steiner tree problem using genetic algorithms', *J. Oper. Res. Soc.* **44** (1993), 397–406.

[56] KHURI, S., BÄCK, T., AND HEITKÖTTER, J.: 'The zero/one multiple knapsack problem and genetic algorithms': *Proc. 1994 ACM Symposium on Applied Computing*, ACM, 1994, pp. 188–193.

[57] KHURI, S., SCHÜTZ, M., AND HEITKÖTTER, J.: 'Evolutionary heuristics for the bin packing problem', in D.W. PEARSON, N.C. STEELE, AND R.F. ALBRECHT (eds.): *Proc. 2nd Internat. Conf. on Artificial Neural Networks and Genetic Algorithms*, Springer, 1995, pp. 285–288.

[58] LASZEWSKI, G. VON: 'Intelligent structural operators for the k-way graph partitioning problem', in R. BELEW AND L. BOOKER (eds.): *Proc. 4th Internat. Conf. on Genetic Algorithms*, Morgan Kaufmann, 1991, pp. 45–52.

[59] LEVINE, D.: 'A parallel genetic algorithm for the set partitioning problem', *PhD Thesis Illinois Inst. Techn.* (1994).

[60] LIEPINS, G.E., HILLIARD, M.R., PALMER, M.R., AND MORROW, M.: 'Greedy genetics', in J.J. GREFENSTETTE (ed.): *Proc. 2nd Internat. Conf. on Genetic Algorithms*, Lawrence Erlbaum Ass., 1987.

[61] LIEPINS, G.E., HILLIARD, M.R., RICHARDSON, J.T., AND PALMER, M.: 'Genetic algorithms applications to set covering and traveling salesman problems', in D.E. BROWN AND C.C. WHITE (eds.): *Oper. Res. and Artificial Intelligence: The Integration of Problem-Solving Strategies*, Kluwer Acad. Publ., 1990, pp. 29–57.

[62] MANIEZZO, V.: 'Exact and approximate nondeterministic tree-search procedures for the quadratic assignment problem', *Techn. Report Univ. Bologna* **CSR 98-1** (1998).

[63] MANIEZZO, V., AND COLORNI, A.: 'The ant system applied to the quadratic assignment problem', *IEEE Trans. Knowledge and Data Engin.* (1998).

[64] MANIEZZO, V., COLORNI, A., AND DORIGO, M.: 'The ant system applied to the quadratic assignment problem', *Techn. Report IRIDIA/94-28, Univ. Libre de Bruxelles, Belgium* (1994).

[65] MARTELLO, S., AND TOTH, P.: 'Lower bounds and reduction procedures for the bin packing problem', *Discrete Appl. Math.* **22** (1990), 59–70.

[66] MATHIAS, K., AND WHITLEY, D.: 'Genetic operators, the fitness landscape and the traveling salesman problem', in R. MÄNNER AND B. MANDERICK (eds.): *Parallel Problem Solving from Nature*, Elsevier, 1992, pp. 219–228.

[67] MÜHLENBEIN, H., GORGES-SCHLEUTER, M., AND KRÄMER, O.: 'Evolution algorithms in combinatorial optimization', *Parallel Comput.* **7** (1988), 65–85.

[68] NAKANO, R., AND YAMADA, T.: 'Conventional genetic algorithm for job shop problems', in R. BELEW AND L. BOOKER (eds.): *Proc. 4th Internat. Conf. on Genetic Algorithms*, Morgan Kaufmann, 1991, pp. 474–479.

[69] PARGAS, R.P., AND JAIN, R.: 'A parallel stochastic optimization algorithm for solving 2D bin packing problems': *Proc. 9th Conf. on Artificial Intelligence for Applications*, 1993, pp. 18–25.

[70] POTVIN, J.-Y.: 'Genetic algorithms for the traveling salesman problem', in G. LAPORTE AND I.H. OSMAN (eds.): *Metaheuristics in combinatorial optimization*, Vol. 63 of *Ann. Oper. Res.*, Baltzer, 1996, pp. 339–370.

[71] POTVIN, J.-Y., AND BENGIO, S.: 'A genetic approach to the vehicle routing problem with time windows', *Techn. Report CRT-953, Univ. Montréal* (1993).

[72] REEVES, C.R. (ed.): *Modern heuristic techniques for combinatorial problems*, Blackwell, 1993.

[73] REEVES, C.: 'Hybrid genetic algorithms for bin-packing and related problems', in G. LAPORTE AND I.H. OSMAN (eds.): *Metaheuristics in combinatorial optimization*, Vol. 63 of *Ann. Oper. Res.*, Baltzer, 1996, pp. 371–396.

[74] RICHARDSON, J.T., PALMER, M.R., LIEPINS, G.E., AND HILLIARD, M.: 'Some guidelines for genetic algorithms with penalty functions', in J.D. SCHAFFER (ed.): *Proc. 3rd Internat. Conf. on Genetic Algorithms*, Morgan Kaufmann, 1989, pp. 191–197.

[75] ROCHAT, Y., AND TAILLARD, E.D.: 'Probabilistic diversification and intensification in local search for vehicle routing', *J. Heuristics* **1** (1995), 147–167.

[76] SEKHARAN, D.A., AND WAINWRIGHT, R.L.: 'Manipulating subpopulations in genetic algorithms for solving the k-way graph partitioning problem': *Proc. 7th Oklahoma Symposium on Artificial Intelligence*, 1993, pp. 215–225.

[77] SMITH, D.: 'Bin packing with adaptive search', in J.J. GREFENSTETTE (ed.): *Proc. 1st Internat. Conf. on Genetic Algorithms*, Lawrence Erlbaum Ass., 1985, pp. 202–207.

[78] STARKWEATHER, T., MCDANIEL, S., MATHIAS, K., WHITLEY, D., AND WHITLEY, C.: 'A comparison of genetic sequencing operators', in R. BELEW AND L. BOOKER (eds.): *Proc. 4th Internat. Conf. on Genetic Algorithms*, Morgan Kaufmann, 1991, pp. 69–76.

[79] STÜTZLE, T., AND HOOS, H.: 'The MAX-MIN ant system and local search for the traveling salesman problem': *Proc. 1997 IEEE Internat. Conf. on Evolutionary Computation*, IEEE Press, 1997, pp. 308–313.

[80] SYSWERDA, G.: 'Schedule optimization using genetic algorithms', in L. DAVIS (ed.): *Handbook Genetic Algorithms*, v. Nostrand Reinhold, 1991, p. 333.

[81] TAILLARD, E.: 'Comparison of iterative searches for the quadratic assignment problem', *Location Sci.* **3** (1995), 87–105.

[82] TATE, D.M, AND SMITH, A.E.: 'A genetic approach to the quadratic assignment problem', *Computers Oper. Res.* **22** (1995), 73–83.

[83] THANGIAH, S.R.: 'Vehicle routing with time windows using genetic algorithms', in L. CHAMBERS (ed.): *Applications handbook of genetic algorithms: new frontiers*, CRC Press, 1995.

[84] THANGIAH, S.R., NYGARD, K.E., AND JUELL, P.L.: 'GIDEON: A genetic algorithm system for vehicle routing with time windows': *Proc. 7th IEEE Conf. Artificial Intelligence Applications*, IEEE Computer Soc. Press, 1991, pp. 322–328.

[85] THANGIAH, S.R., OSMAN, I.H., AND SUN, T.: 'Metaheuristics for vehicle routing problems with time windows', *Techn. Report Slippery Rock Univ.* (1995).

[86] THANGIAH, S.R., OSMAN, I.H., VINAYAGAMOORTHY, R., AND SUN, T.: 'Algorithms for the vehicule routing problems with time deadlines', *American J. Math. Management Sci.* **13** (1994), 323–355.

[87] THANGIAH, S.R., VINAYAGAMOORTHY, R., AND GUBBI, A.: 'Vehicle routing with time deadlines using genetic and local algorithms', in S. FORREST (ed.): *Proc. 5th Internat. Conf. Genetic Algorithms*, Morgan Kaufmann, 1993, pp. 506–513.

[88] THIEL, J., AND VOSS, S.: 'Some experiences on solving multiconstraint zero-one knapsack problems with genetic algorithm', *INFOR special issue: Knapsack, packing and cutting, Part II* **32** (1994), 226–242.

[89] ULDER, N.L.J., AARTS, E.H.L, BANDELT, H.J., LAARHOVEN, P.J.M. VAN, AND PESCH, E.: 'Genetic local search algorithms for the traveling salesman problems', in H.-P. SCHWEFEL AND R. MÄNNER (eds.): *Parallel Problem Solving from Nature*, Vol. 496 of *Lecture Notes Computer Sci.*, Springer, 1991, pp. 109–116.

[90] VINK, M.: 'Solving combinatorial problems using evolutionary algorithms', *Techn. Report Leiden Univ., Netherlands* (1997).

[91] YAMADA, T., AND NAKANO, R.: 'A genetic algorithm applicable to large-scale job-shop problems', in R. MÄNNER AND B. MANDERICK (eds.): *Parallel Problem Solving from Nature, 2*, Elsevier, 1992, pp. 281–290.

Daniel Kobler
Dept. Math. Swiss Federal Inst. Technol.
CH-1015 Lausanne, Switzerland
E-mail address: Daniel.Kobler@epfl.ch

MSC2000: 90C27, 05-04
Key words and phrases: evolutionary algorithm, combinatorial optimization, heuristics.

EXTENDED CUTTING PLANE ALGORITHM

The α-ECP (*extended cutting plane*) algorithm ([12], [14]) is an algorithm for solving quasiconvex MINLP (*mixed integer nonlinear programming*) problems. The algorithm approximates the feasible region with linear approximations and solves a sequence of MILP problems based on these approximations. There are several other similar methods, for instance the generalized Benders decomposition method ([6]), the outer approximation method ([3]), the generalized outer approximation method ([15]), the LP/NLP based branch and bound method ([8]) and the linear outer approximation method ([4]). A good overview of MINLP algorithms and applications is given in [5]. All other methods iteratively solve both NLP and MILP problems, while the α-ECP method only solves MILP problems. The size of the MILP problems grow in each iteration, so efficient algorithms of this type require efficient MILP solvers.

Most of the MINLP methods can only ensure global convergence for convex MINLP problems. The α-ECP method can also solve quasiconvex problems. Different heuristic procedures for some of the above algorithms have been introduced for the nonconvex case, e.g., [10], [13]. Although these methods perform quite well in different applications, convergence towards the optimal solution cannot generally be ensured by these algorithms for nonconvex problems.

There are also some recent MINLP global optimization methods ([1], [2], [9], [11]). In these algorithms the function space is separated for the continuous and discrete variables and the discrete variables can only occur in linear space. The α-ECP method can solve quasiconvex problems

where the discrete variables are involved in nonlinear equations as well.

Although a valid optimal solution is ensured only for quasiconvex problems, the algorithm also provides good approximations for the global optimal solution of general MINLP global optimization problems.

Formulation of the MINLP Problem. The α-ECP algorithm can be used to solve problems of the form

$$\begin{cases} \min & c^\top z \\ \text{s.t.} & g(z) \leq 0 \\ & Az \leq a \\ & Bz = b \\ & z \in X \times Y \end{cases} \tag{1}$$

where c is a vector of constants, $z = (x, y)$ consists of a vector x of continuous variables in $\mathbf{R}^n$ and a vector y of integer variables in $\mathbf{Z}^m$ and $g(z)$: $\mathbf{R}^n \times \mathbf{Z}^m \rightarrow \mathbf{R}^p$ is a vector of continuous differentiable *quasiconvex functions* defined on the set $X \times Y$ having nonzero gradients in the infeasible region of (P). The feasible region of (P) is assumed to be nonempty. Furthermore X is a compact convex set $X \subset \mathbf{R}^n$ and Y is a finite discrete set $Y \subset \mathbf{Z}^m$.

The matrices A and B and vectors a and b are used to define the linear constraints of the problem and are of suitable dimensions.

The α-ECP method guarantees global optimal solutions for MINLP problems having a linear objective function and differentiable quasiconvex constraints. The linear objective function is not too restrictive since most optimization problems having a nonlinear objective $f(z)$ can be rewritten as a problem involving an additional variable u and an additional constraint

$$f(z) - u \leq 0. \tag{2}$$

The new problem, then, will be to minimize u subject to the original constraints and the additional constraint (2). Note, however, that this is not, in general, possible for quasiconvex objectives since $f(z) - u$ is not necessarily quasiconvex when $f(z)$ is quasiconvex.

Definition of the Algorithm. The algorithm solves the problem (1) by approximating the maximal violated nonlinear function with a linear function

$$l(z) = g_i(z^k) + \alpha \cdot \nabla g_i(z^k)^\top (z - z^k) \tag{3}$$

in the current iterate z^k, where $i = \arg\max_i \{g_i(z^k)\}$. To simplify notation, let $g_k = g_i(z^k)$. Furthermore, if the linearization added to the MILP problem is the jth linearization, let $\bar{g}_j = g_i(z^k)$, $\bar{g}_j(z) = g_i(z)$, $\nabla \bar{g}_j = \nabla g_i(z^k)$ and $\bar{z}^j = z^k$ where i is defined as above. The α values change from iteration to iteration so to be able to reference the value of the jth constant in iteration k the α constants are replaced with $\alpha_j^{(k)}$. Thus the linearization (3) is redefined so that in iteration k the jth linear approximation $l_j^{(k)}$ will be

$$l_j^{(k)}(z) = \bar{g}_j + \alpha_j^{(k)} \cdot (\nabla \bar{g}_j)^\top (z - \bar{z}^j)$$

and the algorithm adds the linear constraint

$$l_j^{(k)}(z) \leq 0 \tag{4}$$

to the MILP problem. The α constants initially have the value $\alpha_j^{(k)} = 1$ and they are either left unchanged or increased by a factor in each iteration. The algorithm then iteratively adds more and more constraints to a MILP problem originally consisting of only the linear constraints $Az \leq a$ and $Bz = b$ from (1). In iteration k it thus solves the MILP problem

$$\begin{cases} \min & c^\top z \\ \text{s.t.} & l_j^{(k)} \leq 0, \quad j = 1, \ldots, L_k \\ & Az \leq a \\ & Bz = b \\ & z \in X \times Y \end{cases} \tag{5}$$

where L_k is the number of linearizations in iteration k. The solution to this MILP problem will be the new iteration point. Using this point a new linearization is added to the MILP problem or one or several of the α constants are updated. The procedure is then repeated until a feasible point of (1) is found. A point is considered feasible if

$$g_i(z) \leq \epsilon_g, \quad i = 1, \ldots, p, \tag{6}$$

for some prespecified tolerance ϵ_g. Note that the constraints $Az \leq a$ and $Bz = b$ are automatically satisfied since the current iteration point is

the solution to (5). The idea of finding a feasible and optimal point by solving a sequence of MILP problems is the same as in the classical Kelley's cutting plane method for NLP problems. However, J.E. Kelley [7] considered only the continuous case using LP subsolutions. Furthermore, Kelley's cutting plane algorithm assumes that the linearizations will always be valid underestimators of the corresponding nonlinear functions. This is true if the functions are convex, since for convex functions it holds that

$$g_i(z^k) + \nabla g_i(z^k)^\top (z - z^k) \leq g_i(z) \qquad (7)$$

for all $z, z^k \in X \times Y$. Thus $l_j^{(k)}(z) \leq 0$ whenever $\overline{g}_j(z) \leq 0$ even when $\alpha_j^{(k)} = 1$.

Unfortunately (7) does not generally hold for quasiconvex functions. It is possible that the linear approximations are not valid underestimators of the corresponding nonlinear function and thus the constraint $l_j^{(k)} \leq 0$ may cut away parts of the feasible region. To avoid this problem the α constants have been introduced. By using sufficiently large α values it is ensured that $l_j^{(k)} \leq 0$ whenever $\overline{g}_j(z) \leq 0$ holds, the linearizations will then be valid outer approximations of the feasible region of (1).

Generally it is not known how large the α constants should be. Instead an updating strategy is used. The α values are checked and updated in each iteration if they turn out to be too small. The updated value is obtained by multiplying the current value with a constant greater than one. When the current MILP solution is a feasible solution in (1) and all α constants are large enough, the optimal solution to (1) has been found and the algorithm terminates.

Calculating Sufficiently Large α-Values. Since it is not known beforehand how large α values to use, it is shown below how to obtain large enough values to ensure ϵ-optimality. As previously mentioned, parts of the feasible region may be cut out when linearizing the quasiconvex functions, if the value of the α constant is not increased.

If a sufficiently large α value can be found so that the linearization is a *global underestimator* of the corresponding nonlinear function in the entire feasible region, the linearization should satisfy

$$\overline{g}_j + \alpha_j^{(k)} \cdot (\nabla \overline{g}_j)^\top (z - \overline{z}^j) \leq \overline{g}_j(z), \qquad (8)$$
$$\forall z \in \left\{ z \in X \times Y : \overline{g}_j(z) \leq 0 \right\}.$$

A weaker condition is that the inequality (8) is satisfied only for all current iteration points. If this condition is satisfied, the linearization is called a *local underestimator*. Thus the linearization is a local underestimator if it satisfies the following inequality in iteration k

$$\overline{g}_j + \alpha_j^{(k)} \cdot (\nabla \overline{g}_j)^\top (z^k - \overline{z}^j) \leq \overline{g}_j(z^k), \qquad (9)$$
$$j = 1, \ldots, L_k. \qquad (10)$$

This inequality is easy to check in each iteration. If there is some α constant $\alpha_j^{(k)}$ that does not satisfy (9) then it is updated by multiplying the constant with β. The update formula is thus

$$\alpha_j^{(k+1)} = \begin{cases} \beta \cdot \alpha_j^{(k)}, & l_j^{(k)}(z^k) > \overline{g}_j(z^k), \\ \alpha_j^{(k)} & \text{otherwise.} \end{cases} \qquad (11)$$

The β constant is a prespecified constant ($\beta > 1$). The concept of local underestimators is now extended to *feasible underestimators*. A linearization is called a feasible underestimator if it approximates the entire feasible region. Thus, for such linearizations, it holds that

$$\overline{g}_j + \alpha_j^{(k)} \cdot (\nabla \overline{g}_j)^\top (z - \overline{z}^j) \leq 0, \qquad (12)$$
$$\forall z \in \left\{ z \in X \times Y : \overline{g}_j(z) \leq 0 \right\}. \qquad (13)$$

This is a much more strict requirement since a local underestimator need only underestimate the nonlinear function in a finite set of infeasible points. But condition (12) is weaker than the condition for global underestimators (8) since a feasible underestimator does not necessarily have to underestimate all points in the feasible region of the corresponding nonlinear function. It is only required that $l_j^{(k)} \leq 0$ in this region. In practice, a feasible underestimator needs to underestimate the entire boundary or, more precisely, the convex hull of the feasible region.

To see how to get a feasible underestimator, a new constant $h_j^{(k)}$ is introduced where, as previously with the α constants, the constant will be used in the jth linearization and k stands for the value of the constant in the kth iteration. The constant is defined as

$$h_j^{(k)} = \frac{\overline{g}_j}{\alpha_j^{(k)}}. \tag{14}$$

Since (12) can be divided by $\alpha_j^{(k)}$ the inequality becomes

$$h_j^{(k)} + (\nabla \overline{g}_j)^\top (z - \overline{z}^j) \leq 0 \tag{15}$$

and moreover, because $\alpha_j^{(k)} \geq 1$, it holds that

$$h_j^{(k)} \leq \overline{g}_j. \tag{16}$$

The level sets of quasiconvex functions are convex, which means that if the constant parameter $h_j^{(k)}$ is replaced with zero, then the linearization (15) is always an outer approximation of the feasible region. In fact the linearization is then an approximation of an even larger region

$$\left\{ z \in X \times Y : \overline{g}_j(z) \leq \overline{g}_j(z^k) \right\}$$

containing the feasible region. Thus, if $h_j^{(k)}$ is sufficiently small, (15) is an approximation of the feasible region. In practice the h constants should satisfy

$$h_j^{(k)} \leq \epsilon_h, \quad \forall j = 1, \ldots, L_k.$$

This is the same as requiring that

$$\alpha_j^{(k)} \geq \frac{\overline{g}_j}{\epsilon_h}, \quad \forall j = 1, \ldots, L_k, \tag{17}$$

which can easily be seen from (14). Equation (17) shows that there is an important connection between *sufficiently large* α values and the value of the nonlinear function in the linearization point $(\overline{g}_j)$. The larger the term $\overline{g}_j$ is, the larger the constant α has to be, to be sufficiently large. One could use the same updating scheme (11) as was used for obtaining a local underestimator, but to speed up the process a new updating factor $\gamma > 1$ (and $\gamma \geq \beta$) is introduced. This constant is used to update the α values if the corresponding linearizations are not feasible underestimators.

Whenever the algorithm finds a feasible point it checks that all linearizations are feasible underestimators, i.e. that (17) holds. If there is some $\alpha_j^{(k)}$ constant that violates this inequality, the value of that constant is updated by multiplying it with γ. Thus the α constants will be updated according to

$$\alpha_j^{(k+1)} = \begin{cases} \gamma \cdot \alpha_j^{(k)}, & \alpha_j^{(k)} < \overline{g}_j / \epsilon_h, \\ \alpha_j^{(k)} & \text{otherwise.} \end{cases} \tag{18}$$

In fact, it would be sufficient to require that the linear underestimators should not cut away the optimal point z^*, i.e. that $l_j^{(k)}(z^*) \leq 0$. The algorithm would then terminate in considerably fewer iterations, but since the optimal solution z^* is not known it is very difficult to check this requirement. The same difficulty also appears if the algorithm would be based on global underestimators of the type (8). However, as will follow, global convergence of the algorithm towards the optimal solution can be guaranteed by using local and feasible underestimators. That is why the concepts of local and feasible underestimators have been introduced.

Handling Infeasible MILP Problems. It is possible that the linearizations cut out enough of the feasible region of (P) to make the corresponding MILP problem infeasible. Then there would be no new iteration point and the algorithm would not be able to continue. The solution to this problem is to update all α values and solve the MILP problem again, after updating the values. If there is still no feasible point, this process is repeated until a feasible point is obtained. There exist large enough α values to make the MILP problem feasible, since the nonlinear problem (1) was assumed to be feasible. Thus, if the MILP problem is infeasible, the α update will be

$$\alpha_j^{(k+1)} = \beta \cdot \alpha_j^{(k)}, \quad j = 1, \ldots, L_k. \tag{19}$$

To illustrate the algorithm, a flowsheet of the algorithm is given below.

Convergence of the proposed method. Convergence properties of the algorithm are now studied. Below it is proven that the algorithm converges towards the optimal solution for the quasiconvex problem (1). There are three important properties which are needed to prove convergence. First, the algorithm will never return to the same point if it is infeasible, secondly the generated points will converge to a feasible solution and finally this feasible solution will be the global optimal solution to the original quasiconvex problem (1).

Cycling. First it is shown that the algorithm never returns to the same point if it is infeasible, i.e.,

that cycling is not possible. Note that compactness or quasiconvexity of the constraint functions are unnecessary to prove this theorem.

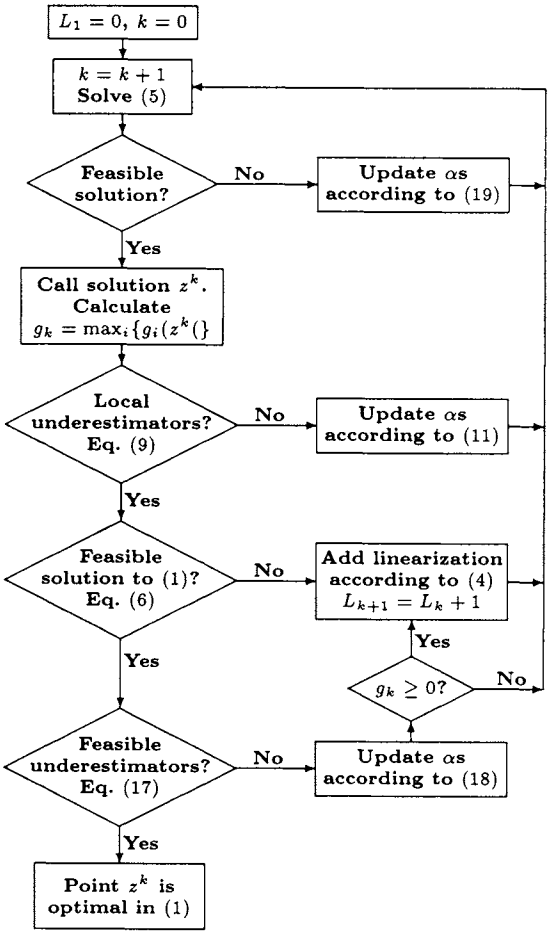

Fig. 1.

THEOREM 1 If, in iteration k, the current point z^k is not feasible, then all new points generated by the algorithm will be different from z^k. □

PROOF. If z^k is infeasible, then $g_k > 0$ and a linearization is added to the MILP problem. If this linearization was the jth one added, then all new points z^l generated by the algorithm will satisfy

$$\overline{g}_j + \alpha_j^{(l)} \cdot (\nabla \overline{g}_j)^\top (z^l - \overline{z}^j) \leq 0, \quad l > k. \tag{20}$$

Since $z^l = z^k$ $(= \overline{z}^j)$ does not satisfy the inequality (20), all new points will be different from z^k. □

It immediately follows that all previous points generated by the algorithm are different from z^k as well.

COROLLARY 2 If the current point z^k is infeasible, then z^k is different from all previous points. □

PROOF. If there is a z^j, $j < k$, such that $z^j = z^k$, then z^j would be a point not satisfying the previous theorem. □

Convergence To a Feasible Point. Convergence to a feasible point for discrete problems is directly ensured by the above cycling theorem. By assumption, there are only a finite number of points in Y, and there is at least one feasible point. Consequently, if the algorithm does not find any of the feasible points in finite time, it would have to repeat an infeasible point after generating at most $|Y|$ iteration points, which is not possible under the cycling theorem.

Convergence in the mixed integer case can be proven by utilizing the fact that the points x^k are taken on a compact set X, and the set Y is finite. This implies that any infinite sequence of points $\{z^k = (x^k, y^k) : k \in \mathcal{K}\}$ taken on the set $X \times Y$ has a subsequence with a limit point. The following theorem shows that any limit point will be a feasible point which is a property required for convergence. Note that the quasiconvex property of the nonlinear functions is not required to prove convergence of the algorithm. Quasiconvexity is only required to ensure a global optimal solution.

The algorithm ensures that $\alpha_j^{(k)} \geq \overline{g}_j/\epsilon_h$, but for simplicity assume that equality holds for those j where $\overline{g}_j \geq \epsilon_h$. Then the constant $h_j^{(k)}$ satisfies

$$\min(\epsilon_h, \overline{g}_j) \leq h_j^{(k)} \leq \overline{g}_j. \tag{21}$$

This follows directly from (16) and the fact that (17) is already satisfied for $\alpha_j^{(k)} = 1$ if $\overline{g}_j < \epsilon_h$.

Below it is proven that any limit point is a feasible point.

THEOREM 3 Suppose that the α-ECP algorithm generates an infinite sequence of points $\{z^k : k \in \mathcal{K}\}$. Then the limit point of any convergent subsequence $\overline{\mathcal{K}} \subset \mathcal{K}$ is feasible. □

PROOF. Assume there is a convergent subsequence $\{z^k : k \in \overline{\mathcal{K}}\}$ with a limit point that is not feasible. Then $\lim_{k \in \overline{\mathcal{K}}} g_k = \epsilon > 0$ and one can find a constant M such that

$$h_j^{(k)} \geq \min\left(\epsilon_h, \frac{\epsilon}{2}\right), \quad \forall j > L_M, \forall k > M,$$

by (21). Since subsequent points z^k are solutions to a linear program containing the linearization (15) it holds for all k that

$$\begin{aligned} 0 &\geq h_j^{(k)} + (\nabla \overline{g}_j)^\top (z^k - \overline{z}^j) \\ &\geq h_j^{(k)} - \|\nabla \overline{g}_j\| \cdot \left\| z^k - \overline{z}^j \right\| \end{aligned}$$

when $j = 1, \ldots, L_k$. Define G as the maximal norm of the gradient of $g(z)$ in $X \times Y$. That is, $G = \max\{\|\nabla g_i(z)\| : z \in X \times Y, i = 1, \ldots, p\}$. Then

$$\left\| z^k - \overline{z}^j \right\| \geq \frac{h_j^{(k)}}{\|\nabla \overline{g}_j\|} \geq \frac{\min(\epsilon_h, \epsilon/2)}{G} > 0$$

when $k > M$ and $j > L_M$. This implies that the sequence is not a Cauchy sequence and thus not convergent, which is a contradiction since it was assumed that the sequence $\{z^k : k \in \overline{\mathcal{K}}\}$ was convergent. □

Convergence To the Optimal Solution. Finally, convergence of the algorithm to the global optimal solution of (1) is shown.

First note that the algorithm will terminate in finite time at a point where all underestimators are ϵ-feasible underestimators, i.e. equation (17) is satisfied. This follows from the convergence theorem. Since any convergent subsequence has a limit point that is feasible, it means that the entire sequence of points will also converge to a feasible point. Thus there is a tail of the sequence, say $\{\overline{z}^j : j = M, \ldots\}$, where the initial α values of the corresponding linearizations directly satisfy (17). This is true for those M values that satisfy $\overline{g}_j \leq \epsilon_h$, $\forall j > M$. These α values will remain constant in subsequent iterations. On the other hand, after reaching a feasible point ($\overline{g}_j \leq \epsilon_g$), the old constants $\alpha_j^{(k)}$, $j = 1, \ldots, M$, can only be updated a finite number of times before being sufficiently large to satisfy (17). Therefore the algorithm will eventually reach a feasible point where all linearizations are ϵ-feasible underestimators and the algorithm terminates. It remains to see if this point is also the optimal solution.

To prove that the obtained solution is the optimal solution one needs to assume that all linear constraints are feasible underestimators according to (12). This is in general true if $h_j^{(k)} = 0$. However, in the actual algorithm it was only required that $h_j^{(k)} \leq \epsilon_h$. Thus the actual solution obtained by the algorithm can only be ensured to be ϵ-optimal.

THEOREM 4 Assume that the α-ECP algorithm converges to a feasible solution z^∞ and that all linearizations are feasible underestimators according to (12). Then z^∞ is an optimal point in (P) and $Z(z^\infty)$, where $Z(z) = c^\top z$, is the optimal solution of (1). □

PROOF. Denote the feasible region of (1) with Ω, the feasible region of the MILP problem that was solved to obtain z^∞ with Ω^∞ and an optimal point of (1) with z^*. By (12) it holds that $\Omega \subset \Omega^\infty$ and thus

$$Z(z^\infty) \leq Z(z^*). \tag{22}$$

On the other hand z^∞ was feasible in (1) and thus

$$Z(z^*) \leq Z(z^\infty). \tag{23}$$

From (22) and (23) one gets that $Z(z^*) = Z(z^\infty)$ and thus $Z(z^\infty)$ is the optimal solution to (1) and z^∞ is an optimal point in (1). □

EXAMPLE 5 The algorithm is demonstrated on a quasiconvex integer problem. In these, as well as in other test runs, it has turned out that a suitable choice of β and γ is $\beta = 1.3$ and $\gamma = 10$. The ϵ-tolerances in these examples are $\epsilon_g = \epsilon_h = 0.001$.

Consider the problem

$$\begin{cases} \min & 3y_1 + 2y_2 \\ \text{s.t.} & 3.5 - y_1 y_2 \leq 0 \\ & y \in \{1, \ldots, 5\}^2. \end{cases} \tag{24}$$

The optimal solution to this problem is $y = (2, 2)$, which can be seen from the figure below.

The steps executed by the α-ECP algorithm are:

Iteration 1. Solve problem

$$\begin{cases} \min & 3y_1 + 2y_2 \\ \text{s.t.} & y \in \{1, \ldots, 5\}^2. \end{cases}$$

The solution is $y^1 = (1, 1)$. A linearization in this point

$$2.5 + \alpha_1^{(1)} \begin{pmatrix} -1 & -1 \end{pmatrix} \begin{pmatrix} y_1 - 1 \\ y_2 - 1 \end{pmatrix} \leq 0$$

is added to the MILP problem according to (4). Set $\alpha_1^{(1)} = 1$. The linearization $l_1^{(1)}$ is shown in Fig. 2.

As can be seen from this figure, the linearization cuts away the optimal solution to the problem.

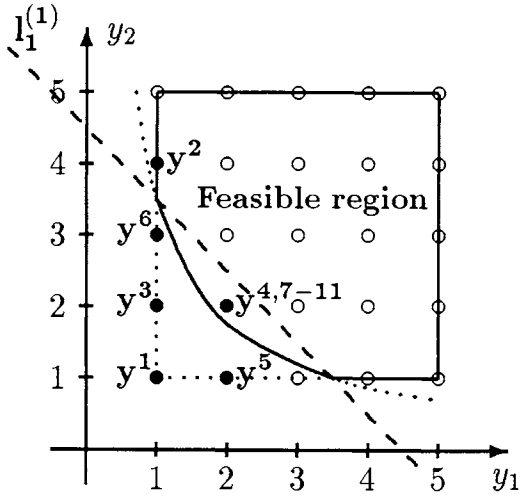

Fig. 2: Feasible region of (24).

Iteration 2. The solution to the new MILP problem is $y^2 = (1, 4)$. This point is a feasible solution to the INLP problem. The linearization satisfies the requirements of a local underestimator but is not a feasible underestimator. Observe, that without the concept of feasible underestimators the algorithm would stop here at a nonoptimal point. However, in order to ensure the linear function be a feasible underestimator, the α constant is updated according to (18) and $\alpha_1^{(3)} = 10$. Since $\max_i\{g_i(z^k)\} < 0$, no additional linearization is added.

Iteration 3. The solution to the new MILP problem is $y^3 = (1, 2)$. A new linearization at this point is added to the MILP problem ($\alpha_2^{(3)} = 1$)

$$1.5 + \alpha_2^{(3)} \begin{pmatrix} -2 & -1 \end{pmatrix} \begin{pmatrix} y_1 - 1 \\ y_2 - 2 \end{pmatrix} \leq 0$$

Iteration 4. The MILP solution is $y^4 = (2, 2)$ which is feasible, however, neither linearization is a feasible underestimator, so the α values are updated using (18). The new values are $\alpha_1^{(5)} = 100$ and $\alpha_2^{(5)} = 10$.

Iteration 5. The solution of the modified MILP problem is $y^5 = (2, 1)$. Since it is infeasible, a new linearization

$$1.5 + \alpha_3^{(5)} \begin{pmatrix} -1 & -2 \end{pmatrix} \begin{pmatrix} y_1 - 2 \\ y_2 - 1 \end{pmatrix} \leq 0$$

is added, where $\alpha_3^{(5)} = 1$.

Iteration 6. The MILP solution is $y^6 = (1, 3)$ which is also infeasible. A new linearization

$$0.5 + \alpha_4^{(6)} \begin{pmatrix} -3 & -1 \end{pmatrix} \begin{pmatrix} y_1 - 1 \\ y_2 - 3 \end{pmatrix} \leq 0$$

is added ($\alpha_4^{(6)} = 1$).

Iteration 7. The MILP solution is again the feasible solution $y^7 = (2, 2)$. The linearizations are not feasible underestimators and thus the α values are updated. The new α values are $\alpha_1^{(8)} = 1000$, $\alpha_2^{(8)} = 100$ and $\alpha_3^{(8)} = \alpha_4^{(8)} = 10$.

Iterations 8–10. The new solutions to the MILP problems are still $y^{8,9,10} = (2, 2)$ but the α values are not large enough to guarantee that the linearizations are feasible underestimators. Therefore the α constants are updated.

Iteration 11. The solution is $y^{11} = (2, 2)$ and all linearizations are feasible underestimators. The algorithm terminates with $y^* = (2, 2)$.

Result. The algorithm thus returns the global solution $y^* = (2, 2)$ to (24) with the optimal value $Z(2, 2) = 10$. The final MILP problem solved in iteration 11 is

$$\begin{cases} \min & 3y_1 + 2y_2 \\ \text{s.t.} & 2.5 + 10000(2 - y_1 - y_2) \leq 0 \\ & 1.5 + 10000(4 - 2y_1 - y_2) \leq 0 \\ & 1.5 + 10000(4 - y_1 - 2y_2) \leq 0 \\ & 0.5 + 1000(6 - 3y_1 - y_2) \leq 0 \\ & y \in \{1, \ldots, 5\}^2. \end{cases}$$

□

Conclusions. The above algorithm has several advantages when compared to other similar algorithms for solving MINLP problems. At each iteration, the procedure only solves MILP subproblems and is thus a competitive alternative to algorithms where only NLP problems or both NLP and MILP problems are solved in each iteration.

One consequence is that since only MILP problems are solved in each iteration, the nonlinear constraints need not be calculated at relaxed values of the integer variables. It can be very diffi-

cult to calculate the value in a relaxed point if, for instance, there are binary variables that represent the existence of units in a process and the constraints are evaluated by simulating the result of having those units present or not. Then it may sometimes be impossible to evaluate the constraints if the integer variables are relaxed.

The α-ECP algorithm also solves MINLP problems that have general integer variables, not only binary variables. Also, no integer cuts are needed to ensure convergence. This is not the case with all outer approximation MINLP methods. In addition, the proposed algorithm ensures global convergence for quasiconvex MINLP problems.

Cutting plane methods are claimed to have slow convergence. This, generally, is not the case if the convergence rate is measured as the number of nonlinear function evaluations. Numerical experience with the algorithm indicates that there are many cases where the number of function evaluations are even magnitudes lower than for competing algorithms that solve both MINLP and NLP subproblems. This is a significant advantage if evaluation of the constraints is the most time-consuming part of the problem.

The algorithm is especially suitable for solving INLP problems.

See also: **Chemical process planning**; **Mixed integer linear programming: Mass and heat exchanger networks**; **Mixed integer nonlinear programming**; **MINLP: Outer approximation algorithm**; **Generalized outer approximation**; **MINLP: Generalized cross decomposition**; **Generalized Benders decomposition**; **MINLP: Logic-based methods**; **MINLP: Branch and bound methods**; **MINLP: Branch and bound global optimization algorithm**; **MINLP: Global optimization with αBB**; **MINLP: Heat exchanger network synthesis**; **MINLP: Reactive distillation column synthesis**; **MINLP: Design and scheduling of batch processes**; **MINLP: Applications in the interaction of design and control**; **MINLP: Application in facility location-allocation**; **MINLP: Applications in blending and pooling problems**.

References

[1] ADJIMAN, C.S., ANDROULAKIS, I.P., AND FLOUDAS, C.A.: 'Global optimization of MINLP problems in process synthesis and design', *Computers Chem. Engin.* **21** (1997), 445–450.

[2] ANDROULAKIS, I.P., MARANAS, C.D., AND FLOUDAS, C.A.: 'α-BB: A global optimization method for general constrained nonconvex problems', *J. Global Optim.* **7** (1995), 337–363.

[3] DURAN, M.A., AND GROSSMANN, I.E.: 'An outer approximation algorithm for a class of mixed-integer nonlinear programs', *Math. Program.* **36** (1986), 307–339.

[4] FLETCHER, R., AND LEYFFER, S.: 'Solving mixed-integer nonlinear programs by outer approximation', *Math. Program.* **66** (1994), 327–349.

[5] FLOUDAS, C.A.: *Nonlinear and mixed-integer optimization, fundamentals and applications*, Oxford Univ. Press, 1995.

[6] GEOFFRION, A.M.: 'Generalized Benders decomposition', *J. Optim. Th. Appl.* **10** (1972), 237–260.

[7] KELLEY, J.E.: 'The cutting plane method for solving convex programs', *J. SIAM* **VIII**, no. 4 (1960), 703–712.

[8] QUESADA, I., AND GROSSMANN, I.E.: 'An LP/NLP based branch-and-bound algorithm for convex MINLP optimization problems', *Computers Chem. Engin.* **16** (1992), 937–947.

[9] RYOO, H.S., AND SAHINIDIS, N.V.: 'Global optimization of nonconvex NLPs and MINLPs with applications in process design', *Computers Chem. Engin.* **19** (1995), 551–566.

[10] VISWANATHAN, J., AND GROSSMANN, I.E.: 'A combined penalty function and outer approximation method for MINLP optimization', *Computers Chem. Engin.* **14** (1990), 769–782.

[11] VISWESWARAN, V., AND FLOUDAS, C.A.: 'New formulations and branching strategies for the GOP algorithm', in I.E. GROSSMANN (ed.): *Global Optimization in Engineering Design*, Kluwer Acad. Publ., 1996, pp. 75–109.

[12] WESTERLUND, T., AND PETTERSSON, F.: 'An extended cutting plane method for solving convex MINLP problems', *Computers Chem. Engin. Suppl.* **19** (1995), S131–136.

[13] WESTERLUND, T., PETTERSSON, F., AND GROSSMANN, I.E.: 'Optimization of pump configurations as a MINLP problem', *Computers Chem. Engin.* **18** (1994), 845–858.

[14] WESTERLUND, T., SKRIFVARS, H., HARJUNKOSKI, I., AND PÖRN, R.: 'An extended cutting plane method for a class of non-convex MINLP problems', *Computers Chem. Engin.* **22** (1998), 357–365.

[15] YUAN, X., PIBOULEAN, L., AND DOMENECH, S.: 'Experiments in process synthesis via mixed-integer programming', *Chem. Engin. and Processing* **25** (1989), 99–116.

Claus Still
Dept. Math. Åbo Akademi Univ.
Fänriksgatan 3
FIN-20500 Åbo, Finland
E-mail address: cstill@abo.fi
Tapio Westerlund
Process Design Lab. Åbo Akad. Univ.
Biskopsgatan 8
FIN-20500 Åbo, Finland
E-mail address: twesterl@abo.fi

MSC2000: 90C11, 90C26
Key words and phrases: mixed integer nonlinear programming, extended cutting plane, quasiconvex function, feasible underestimators.

EXTREMUM PROBLEMS WITH PROBABILITY FUNCTIONS: KERNEL TYPE SOLUTION METHODS, *KSM*

Two types of stochastic programs are widely known: two-stage and chance constrained problems. The last ones were introduced to stochastic programming by A. Charnes and W.W. Cooper in the 1950s [1] and are formally described defining a nonlinear probability function $v(x,t)$ of the form:

$$v(x,t) = \mathsf{P}\{\xi\colon f(x,\xi) \le t\}. \qquad (1)$$

Here $f(x,\xi)$ is a real valued function, defined on $\mathbf{R}^r \times \mathbf{R}^v$, t is a fixed level of reliability, $\xi = \xi(\omega)$ is a random parameter and P denotes probability. Note that for a fixed x the function $v(x,t)$ as a function of t is the distribution function of the random variable $f(x,s)$.

Various examples of extremum problems with probability function $v(x,t)$ can be found in [3, Chap. 1], where among others also the so-called 'stock exchange' paradox is analyzed. To overcome a paradoxical situation being caused by an unsuccessful choice of the objective expected return, the strategy which maximizes the expected growth of return (Kelly strategy), was applied in [2]. In [3] it was demonstrated that a risky (i.e. probabilistic) strategy is better than the Kelly one.

In the approximate maximization of $v(x,t)$ over the constraint set $X \subset \mathbf{R}^r$ we should apply some (quasi-) gradient type method. This in turn needs the presentation of $v(x,t)$ as an integral, which we can realize via the Heaviside zero-one function $\chi(\cdot)$:

$$\chi(t - f(x,\xi)) = \begin{cases} 1, & \text{if } f(x,\xi) \le t, \\ 0, & \text{if } f(x,\xi) > t. \end{cases}$$

Then

$$v(x,t) = \int_S \chi(t - f(x,\xi))\sigma(\,d\xi), \qquad (2)$$

where $\sigma(\cdot)$ is the distribution function of a random vector ξ and the integral in (2) is understood in the Lebesgue–Stieltjes sense.

Integral representation (2) of the probability function $v(x,t)$ demonstrates us expressively difficulties which arise in approximate maximization of its value: integrand $\chi(\cdot)$ itself is a discontinuous zero-one function and integral (2) over $\chi(\cdot)$ is never convex. Only in some cases, e.g., if function $f(x,\xi)$ is jointly convex and continuous in (x,ξ) and $\sigma(\cdot)$ as a measure is quasiconcave, then function $v(x,t)$ is quasiconcave in x, see [12].

In this survey we at first will solve iteratively, using stochastic analogues of linearization and gradient projection methods, the following probability maximization problem:

$$\max_{x\in X} v(x,t) = \max_{x\in X} \mathsf{P}\{\xi\colon f(x,\xi) \le t\}, \qquad (3)$$

where the constraint set X is assumed to be simple, i.e. on X we can effectively solve auxiliary problems of maximization of linear or quadratic functions. At second, we will exploit the introduced technique for minimization of a smooth function over probabilistic equality-inequality type constraints, using a stochastic analogue of the modified Lagrange method.

Gradient type methods require differentiability of a cost function. A lot of papers have been devoted to differentiability conditions of $v(x,t)$ in x, starting from [13] where $v'_x(x,t)$ was presented via surface integral. The gradient of $v(x,t)$ via volume integral was presented in [16]; see also the survey paper [4]. All these formulas are quite uncomfortable to handle, especially for numerical methods. In the following we will assume differentiability of $v(x,t)$ in x and in (x,t), i.e. there exist $v'_x(x,t)$ and $v''_{xt}(x,t)$.

Define solution sets X^* for the problem (3) as follows:

$$X^* = \{x^*\colon (v'_x(x^*,t), x - x^*) \le 0, \quad \forall x \in X\}, \qquad (4)$$

or

$$X^* = \{x^* : x^* = \pi[x^* + \rho v'_x(x,t)], \quad \forall \rho > 0\}, \tag{5}$$

where $\pi[y]$ means the projection of a vector y to the set X. Then we can interpret linearization and gradient projection methods as iteration ways for testing conditions (4) and (5), respectively.

Following [17, Chap. IV], method for solution of a problem is said to be convergent, if limit points of the sequence $\{x_n\}$, generated by the algorithm, belong to the solution set X^*.

Denote n independent realizations $\xi_1, \ldots, \xi_n$ of a random vector ξ by ξ^n, i.e., $\xi^n = (\xi_1, \ldots, \xi_n)$. Then, following [14] and [10], the smoothed approximation of $v(x,t)$ looks as follows:

$$v_n(x,t,\xi^n) = v_n(x,t,\xi_1 ..., \xi_n) \tag{6}$$
$$= \frac{1}{nh_n} \sum_{i=1}^{n} \int_{-\infty}^{t} K\left(\frac{\tau - f(x,\xi_i)}{h_n}\right) d\tau,$$

where the sequence $\{h_n\}$ is connected with the sequence $\mathbf{N} = \{1, 2, \ldots\}$ as

$$\lim h_n = 0, \quad \lim nh_n = \infty, \quad n \in \mathbf{N}, \tag{7}$$

and the continuous kernel function $K(y)$ satisfies conditions [14]:

$$\int_{-\infty}^{\infty} K(y)\, dy = 1, \quad \sup_{-\infty<y<\infty} |K(y)| < \infty, \tag{8}$$

$$\int_{-\infty}^{\infty} yK(y)\, dy = 0, \quad \int_{-\infty}^{\infty} |yK(y)|\, dy < \infty. \tag{9}$$

Gradient of the smoothed approximate probability function $v_n(x,t,\xi^n)$ from (6) looks now as follows:

$$v'_{nx}(x,t,\xi^n) \tag{10}$$
$$= -\frac{1}{nh_n} \sum_{i=1}^{n} f'_x(x,\xi_i) K\left(\frac{t - f(x,\xi_i)}{h_n}\right).$$

Even estimates (6) and (10) are biased, i.e.,

$$\mathsf{E} v'_{nx}(x,t,\xi^n) \neq v'_x(x,t),$$

we still have

$$\mathsf{E} v'_{nx}(x,t,\xi^n)$$
$$= v'_x(x,t) - h_n \int_{-\infty}^{\infty} yK(y) v'_{x,t}(x, t - \theta h_n y)\, dy,$$

where $0 \leq \theta \leq 1$, see [15], and consequently,

$$\lim_{n\to\infty} \sup_{x\in X} \left|\mathsf{E} v'_{nx}(x,t,\xi^n) - v'_x(x,t)\right| = 0.$$

For approximate solution of (3) consider the stochastic analogue of the linearization method:

$$x_{n+1} = x_n + \gamma_n(\overline{x}_n - x_n), \tag{11}$$

where $\overline{x}_n$ is a solution of the linear problem:

$$\max_{x\in X}(v'_{nx}(x_n,t,\xi^n), x) = v'_{nx}(x_n,t,\xi^n,\overline{x}_n)$$

and $x_0 \in X$.

Explain the stochastic nature of the sequence $\{x_n\}$, generated by the algorithm (11). For each n the random vector x_n is defined on the sigma-algebra F_{n-1}, generated by random vectors $\xi_1, \ldots, \xi_{n-1}$. Union of the sequence of sub-sigma-algebras $\cup_{i=1}^{\infty} F_i$ is equal to the sigma-algebra F of the initial probability space (Ω, F, P), where the random vector ξ was defined. Note that in each iteration step we should generate new (independent) realizations of the random vector ξ.

Assume that function $f(x,\xi)$ is differentiable in x and that for all $t \in \mathbf{R}^1$ and all $x \in X$ its gradient is bounded with a σ-integrable function $K(\xi)$:

$$|f_x(x,\xi)| \leq K(\xi), \quad \int_{\mathbf{R}^v} K(\xi)\sigma(\, d\xi) < \infty. \tag{12}$$

Let the sequence $\{\gamma_n\}$ of steplength satisfy conditions:

$$0 \leq \gamma_n \leq 1, \quad \gamma_n \to 0, \quad \sum_{n=1}^{\infty} \gamma_n = \infty. \tag{13}$$

Then the following convergence theorem holds [5]

THEOREM 1 Let differentiable in x function $f(x,\xi)$ satisfy conditions (12), smoothing continuous kernel $K(y)$ conditions (8), (9), sequence $\{\gamma_n\}$ of steplength conditions (13), and let the solution set X^* be finite. Then all limit points of the sequence $\{x_n\}$, generated by the algorithm (11), belong almost surely to the solution set X^*. □

REMARK 2 Proof of the theorem relies on the stochastic analogue of [17, Thm. A], see [9, Chap. II, Thm. 8], and was verified in [5]. □

REMARK 3 Statements of the theorem are valid also for the stochastic analogue of the gradient projection method, see [5]:

$$x_{n+1} = \pi[x_n + \gamma_n v'_{nx}(x_n, t, \xi_n)], \quad x_0 \in X. \quad (14)$$

□

As it was described earlier, algorithms (11) and (14) need in nth iteration step n independent realizations of the random vector ξ. In [11] it was verified that in asymptotic sense statistical estimation type methods, as algorithms (11) and (14) are, have no advantages compared with methods of random search, but need more calculating efforts.

As an example of the last statement consider the free maximum problem:

$$\max_{x \in \mathbf{R}^r} = \max_{x \in \mathbf{R}^r} \mathsf{P}\{\xi \colon f(x, \xi) \le t\}. \quad (15)$$

Let ξ_n be the nth realization of the random variable ξ. Consider the algorithm:

$$x_{n+1} = x_n - \frac{\gamma_n}{h_n} f'_x(x_n, \xi_n) K\left(\frac{t - f(x_n, \xi_n)}{h_n}\right). \quad (16)$$

Assume, in addition to assumptions (7)–(9) and (12), (13) to $\{h_n\}$, $\{\gamma_n\}$, $K(y)$ and $f(x, \xi)$, that

$$\sum_{n=1}^{\infty} \gamma_n^2 < \infty; \quad \sum_{n=1}^{\infty} \gamma_n h_n < \infty; \quad \sum_{n=1}^{\infty} \frac{\gamma_n^2}{h_n^2} < \infty. \quad (17)$$

Then, if $\int_{\mathbf{R}^v} |f'_x(x, \xi)| \, \sigma(d\xi)$ is bounded for bounded x, the limit points of the sequence $\{x_n\}$ belong almost surely to the set X^* of stationary points,

$$X^* = \{x^* \colon v'_x(x^*, t) = 0\},$$

see [7].

REMARK 4 Even algorithms (11) and (14) take more calculating efforts compared with random search method (16), the last one is very unstable, and converges only 'in probability' sense. □

Consider the following nonlinear programming problem with a smooth cost function $f(x)$ and with probabilistic constraints of inequality type with a fixed level of reliability α, $0 < \alpha < 1$, i.e.,

$$\min_{x \in \mathbf{R}^r} \{f(x) \colon v(x, t) \ge \alpha\} \quad (18)$$

(for sake of simplicity consider only the case with one inequality constraint).

Define the solution set X^* for the problem (18) as follows [8]:

$$X^* = \{x^* \colon F \cap G\}, \quad (19)$$

where

$$F = \left\{x^* \colon \left|f'_x(x^*) + v'_x(x^*, t)\lambda^*\right|^2 = 0\right\}, \quad (20)$$

with

$$\lambda^* = \arg\min_{\lambda \ge 0} \left|f'_x(x^*) + v'_x(x^*, t)\lambda\right|^2, \quad (21)$$

and

$$G = \{x^* \colon v(x^*, t) \ge \alpha\}, \quad (22)$$

where λ^* is the optimal Lagrange multiplier of the Lagrangian.

Replacing $v(x, t)$ and $v'_x(x, t)$ with their estimates (6) and (10), we should regularize the estimated analogue of (21) since the approximated subproblem (21) could be ill-posed.

Denote by

$$w_n(x, t, \xi_n) = \min\{0, v_n(x, t, \xi^n) - \alpha\}.$$

Then the stochastic analogue of modified Lagrange method looks as follows:

$$\begin{aligned} x_{n+1} &= x_n \\ &-\gamma_n[f'_x(x_n) + v'_{nx}(x_n, t, \xi^n)\lambda_n(\xi^n) \\ &+Mv'_{nx}(x_n, t, \xi^n)w_n(x_n, t, \xi_n)], \end{aligned} \quad (23)$$

where $\lambda_n(\xi_n)$ is a solution of the regularized auxiliary subproblem of quadratic programming

$$\min_{\lambda \ge 0} \left[\left|f'_x(x_n) + v'_{nx}(x_n, t, \xi^n)\lambda\right|^2 + \alpha_n |\lambda|^2\right]$$

with $\alpha_n > 0$, $\alpha_n \to 0$, $n \to \infty$ and $M > 0$. The following convergence theorem is valid, see [6]:

THEOREM 5 Let conditions of the previous theorem be satisfied, let the cost function $f(x)$ be continuously differentiable and let

$$\sum_{n=1}^{\infty} \alpha_n \gamma_n < \infty.$$

Then limit points of the sequence, generated by the algorithm (23), belong almost surely to the solution set X^*, defined by (19). □

See also: **Stochastic programming with simple integer recourse; Two-stage stochastic programs with recourse; Stochastic vehicle routing problems; Stochastic integer programming: Continuity, stability, rates of convergence; Logconcave measures, logconvexity; Logconcavity of discrete distributions; General moment optimization problems; Approximation of multivariate probability integrals; Discretely distributed stochastic programs: Descent directions and efficient points; Static stochastic programming models; Static stochastic programming models: Conditional expectations; Stochastic programming models: Random objective; Stochastic programming: Minimax approach; Simple recourse problem: Primal method; Simple recourse problem: Dual method; Probabilistic constrained linear programming: Duality theory; Probabilistic constrained problems: Convexity theory; Approximation of extremum problems with probability functionals; Multistage stochastic programming: Barycentric approximation; Stochastic linear programs with recourse and arbitrary multivariate distributions; Stochastic programs with recourse: Upper bounds; Stochastic integer programs; L-shaped method for two-stage stochastic programs with recourse; Stochastic linear programming: Decomposition and cutting planes; Stabilization of cutting plane algorithms for stochastic linear programming problems; Two-stage stochastic programming: Quasigradient method; Stochastic quasigradient methods in minimax problems; Stochastic programming: Nonanticipativity and Lagrange multipliers; Preprocessing in stochastic programming; Stochastic network problems: Massively parallel solution.**

References

[1] CHARNES, A., AND COOPER, W.W.: 'Chance-constrained programming', *Managem. Sci.* **5** (1959), 73–79.

[2] KELLY, J.: 'A new interpretation of information rate', *Bell System Techn. J.* **35** (1956), 917–926.

[3] KIBZUN, A.I., AND KAN, Y.S.: *Stochastic programming problems with probability and quantile functions*, Wiley, 1995.

[4] KIBZUN, A., AND URYASEV, S.: 'Differentiability of probability functions', *Stochastic Anal. Appl.* **16** (1998), 1101–1128.

[5] LEPP, R.: 'Maximization of a probability function over simple sets (in Russian)', *Proc. Acad. Sci. Estonian SSR. Phys. Math.* **28** (1979), 303–308.

[6] LEPP, R.: 'Minimization of a smooth function over probabilistic constraints (in Russian)', *Proc. Acad. Sci. Estonian SSR. Phys. Math.* **29** (1980), 140–144.

[7] LEPP, R.: 'Stochastic approximation type algorithm for the maximization of the probability function', *Proc. Acad. Sci. Estonian SSR. Phys. Math.* **32** (1983), 150–156.

[8] MIELE, A., CRAGG, E.G., IVER, R.R., AND LEVY, A.V.: 'Use of the augmented penalty functions in mathematical programming problems. Part I', *J. Optim. Th. Appl.* **8** (1971), 115–130.

[9] NURMINSKII, E.A.: *Numerical methods for solution of deterministic and stochastic Minimax Problems*, Nauk. Dumka, 1979. (In Russian.)

[10] PARZEN, E.: 'On the estimation of a probability density and the mode', *Ann. Math. Statist.* **33** (1962), 1065–1076.

[11] POLYAK, B.T., AND TSYPKIN, Y.Z.: 'Adaptive algorithms of estimation (convergence, optimality, stability) (in Russian)', *Avtomatika i Telemekhanika (Automatics and Remote Control)* (1979), 74–84.

[12] PRÉKOPA, A.: 'Logarithmic concave measures and related topics', in M.A.H. DEMPSTER (ed.): *Stochastic Programming*, Acad. Press, 1980.

[13] RAIK, E.: 'Differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method', *Proc. Acad. Sci. Estonian SSR. Phys. Math.* **24** (1975), 3–6. (In Russian.)

[14] ROSENBLATT, M.: 'Remarks on some nonparametric estimates of a density function', *Ann. Math. Statist.* **27** (1957), 832–837.

[15] TAMM, E.: 'On the minimization of the probability function (in Russian)', *Proc. Acad. Sci. Estonian SSR. Phys. Math.* **28** (1979), 17–24.

[16] URYASEV, S.: 'A differentiation formula for integrals over sets given by inclusion', *Numer. Funct. Anal. Optim.* **10** (1989), 827–841.

[17] ZANGWILL, W.I.: *Nonlinear programming. A unified approach*, Prentice-Hall, 1969.

Riho Lepp
Tallinn Technical Univ.
Tallinn, Estonia
E-mail address: lprh@ioc.ee

MSC2000: 90C15

Key words and phrases: probability function, kernel estimates, stochastic approximation.