European Journal of Operational Research 112 (1999) 432–459

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

Gradient projection and local region search for multiobjective optimisation

Jian-Bo Yang

Manchester School of Management, University of Manchester Institute of Science and Technology, P.O. Box 88, Manchester M60 1QD, UK

Received 7 May 1997; accepted 17 November 1997

Abstract

This paper presents a new method for multiobjective optimisation based on gradient projection and local region search. The gradient projection is conducted through the identification of normal vectors of an efficient frontier. The projection of the gradient of a nonlinear utility function onto the tangent plane of the efficient frontier at a given efficient solution leads to the definition of a feasible local region in a neighbourhood of the solution. Within this local region, a better efficient solution may be sought. To implement such a gradient-based local region search scheme, a new auxiliary problem is developed. If the utility function is given explicitly, this search scheme results in an iterative optimisation algorithm capable of general nonseparable multiobjective optimisation. Otherwise, an interactive decision making algorithm is developed where the decision maker (DM) is expected to provide local preference information in order to determine trade-off directions and step sizes. Optimality conditions for the algorithms are established and the convergence of the algorithms is proven. A multiobjective linear programming (MOLP) problem is taken for example to demonstrate this method both graphically and analytically. A nonlinear multiobjective water quality management problem is finally examined to show the potential application of the method to real world decision problems.

Keywords: Multiobjective optimisation; Utility functions; Interactive methods; Marginal rates of substitution

1. Introduction

Many activities in management and engineering design consist of procedures which essentially involve dealing with optimisation problems with multiple, potentially conflicting requirements (objectives) reflecting technical and economical performance. The aim of multiobjective optimisation is to develop a compromise solution which attains all the objectives as greatly as possible. Many real-world optimisation problems are nonlinear and their decision spaces are often nonconvex. A few methods may be applied to deal with such nonlinear and nonconvex problems, for example the surrogate worth trade-off methods [2], the minimax solution approach [23] and its variations [26], and the ISTM method [27,28].

In certain decision situations such as engineering design synthesis, interactive multiple objective decision making (MODM) methods are desirable.

PII: S 0 3 7 7 - 2 2 1 7 (9 7) 0 0 4 5 1 - 7

This is because such techniques allow the solution to progress towards a preferred solution through an adaptive approach [8,23,24,26,28]. This mirrors the common adaptive engineering design processes [20,21,32]. As a learning-oriented interactive technique, the ISTM methods supports the DM to search for favourable efficient solutions, which leads to a progressive, implicit articulation of DM priorities. The DM can discover his true preferences by such implicit trade-off studies as is often done in real-life nontechnical decision situations.

In the original version of ISTM, the DM is required to terminate the interactive process by judging intuitively that his best solution is generated [28]. Such a nondirected termination strategy requires too much guesswork on the part of the DM [1,31]. A search process was incorporated into ISTM based on the assessment and optimisation of local additive utility functions [31,32]. This revised version of ISTM can support the DM to search for the best compromise solution in a natural manner without bearing of too much guess work. It can also help to check and eliminate any inconsistency possibly involved in the preferences provided by the DM. Furthermore, the DM can explicitly modify the preferences he has already provided whenever necessary. This allows the DM to carry out sensitivity analysis to examine the robustness of the best compromise solution.

However, employing an additive (separable) utility function requires that the objectives satisfy the mutual preference independence condition. In some decision situations, this condition may not always be satisfied. In such circumstances, it is usually difficult, if not impossible, to assess general nonseparable utility functions. On the other hand, it is relatively easy to estimate local information of a utility function such as its gradients by articulating the DM's local preferences such as marginal rates of substitution (MRS) [4–6,8,9,11]. Even if a nonseparable utility function could eventually be represented in an explicit form, the nonseparability will still make it extremely inefficient to solve the optimisation problem in certain decision situations, for example, where the original problem is a large optimisation problem [12] or a multiple linear quadratic control problem [13].

Several methods have been developed to support decision analysis using local preference information such as MRS. One of the early examples is Geoffrion's method [8], where MRS are used to estimate utility gradients. In Geoffrion's method, however, the linear approximation of a utility function is optimised in the original decision space, which leads to a cumbersome procedure for estimating step sizes. In the cutting plan methods [22], MRS are used to construct a trade-off cut at each efficient solution generated. The cutting plane algorithms normally converge to the best compromise solutions if the marginal rates of substitution provided are consistent. The construction of a cutting plane at a current solution is based on the concept that the linear approximation of a utility function at a next solution should be better than at the current solution. Such construction adds a new constraint to the existing set of constraints but does not indicate which objectives should be sacrificed in order to improve utility until the next solution is actually found. In an interactive decision process, the failure of indicating the consequences (sacrifice) of a trade-off analysis may not help the decision maker to provide consistent preference information.

This paper is devoted to developing a gradient-based method on the basis of ISTM. This new method relies on identifying normal vectors at efficient solutions. A new technique is suggested in this paper in order to facilitate the identification in light of the author's other work as reported in [15,30,33]. Given a normal vector of an efficient frontier at an efficient solution, the gradient of a utility function can be projected onto the tangent plane of the efficient frontier at the efficient solution. This projection gives an ascent direction of the utility function although an infinitesimal move along this direction will generally lead out of the feasible objective space. From the projection, however, a local region can be constructed in the feasible objective space, where objectives for sacrifice from the current efficient solution are identified for improving utility.

Based upon an auxiliary model developed in ISTM [28], a new auxiliary problem is proposed for a MODM problem to implement the above gradient-based local region search scheme. This auxiliary problem is designed in a way such that its optimal solution is an efficient solution of the original MODM problem, maximising the utility function in the established local region. At the optimal solution of the auxiliary problem, a normal vector and the gradient of the utility function can be identified. A new local region and a new auxiliary problem can then be constructed, resulting in another efficient solution with probably even better utility. This process is repeated and not terminated until the best compromise solution is generated maximising the utility function in the feasible decision space. To terminate the process, optimality conditions are established in terms of normal vectors and utility gradients. The convergence of this process is also proven. This approach can lead to an iterative optimisation algorithm or an interactive decision making algorithm, depending upon whether the utility function is known explicitly or implicitly.

A new auxiliary problem is defined first and its features are explored. A new technique is suggested to identify normal vectors by solving a set of linear equations. Then, a gradient-based local region search scheme is proposed by constructing a local region in a neighbourhood of an efficient solution. Given a utility function explicitly, an iterative algorithm summarising this search scheme is developed and the convergence of the algorithm is proven. In the following section, a gradient-based interactive step trade-off (simply referred to as GRIST) algorithm is proposed for problems with a utility function only known implicitly. Techniques for assessing trade-off directions and step sizes are discussed in this section. A numerical example is then presented to illustrate the new gradient-based local region search algorithms. In the last section, a nonlinear water quality management problem is examined to demonstrate the potential of the new method for dealing with real world decision problems.

2. The auxiliary problem and normal vectors

2.1. The auxiliary problem for local region search

A general multiobjective optimisation problem (MOP) may be represented by

$$\begin{aligned} \text{MOP} \quad & \max \quad F(X) = \{f_1(X) \cdots f_i(X) \cdots f_k(X)\} \\ & \quad \text{s.t.} \quad X \in \Omega, \end{aligned} \tag{1}$$

$$\Omega = \{X | g_j(X) \leqslant 0,\ \ h_l(X) = 0;\ \ j = 1, \ldots, m_1,\ \ l = 1, \ldots, m_2\},$$

where $f_i(X) (i = 1, \ldots, k)$ are continuously differentiable objective functions which are usually conflicting with one another, $X = [x_1 \cdots x_n]^{\mathrm{T}}$ denotes a solution (decision) and Ω represents a feasible decision space. Suppose Ω is compact, closed and bounded but not necessarily convex. $g_j(X)$ and $h_l(X)$ are continuously differentiable constraint functions.

Let u be the DM's overall utility function, aggregating all objectives into one criterion, or $u = u(f_1(X), \ldots, f_k(X))$. Suppose u is a continuously differentiable nonlinear function, defined as a strictly increasing function of $f_i(X)$ as follows:

$$\partial u / \partial f_i > 0, \quad i = 1, \ldots, k. \tag{2}$$

The problem is then to search for an efficient solution of problem (1) which maximises the utility function u. Such an efficient solution may generally be referred to as the best compromise solution of problem (1). If u can only be expressed implicitly, the best compromise solution may be sought in an interactive manner with u progressively assessed using the DM's local preferences. If u is expressed explicitly, the best compromise solution could be obtained by solving the following scalar problem:

$$\begin{aligned} \text{SOP} \quad & \max \quad u = u(f_1(X) \cdots f_i(X) \cdots f_k(X)) \\ & \quad \text{s.t.} \quad X \in \Omega. \end{aligned} \tag{3}$$

In certain decision situations, however, solving problem (3) directly may be extremely inefficient as u is often a nonlinear (nonseparable) function of the objective functions. If the original MOP defined in Eq. (1) is a dynamic problem or a large optimisation problem [12,13], for example, the nonlinear (especially nonseparable) features of u may make it difficult to adopt traditional methods such as dynamic programming or decomposition methods to generate optimal solutions of problem

(3). It is desirable to develop a search scheme to improve u in an iterative manner. Such an iterative search scheme is especially useful when u is not known explicitly and has to be assessed using local preference information. To develop such a scheme, it is fundamental to design an auxiliary model.

From an efficient solution, an objective can only be improved from its current level by sacrificing some other objectives. The set of objectives can therefore be classified into three subsets at an efficient solution, the first subset consisting of objectives which need to be improved from their current levels, the second of objectives which have to be kept at least at their current levels, and the third of objectives which may be sacrificed to some extent [28]. As a result of such classification, an auxiliary model was developed to search for a new solution which embodies the above trade-off analysis [28]. A new auxiliary model is defined below, which is denoted by EAP^t.

$$\text{EAP}^t \qquad \max \quad y = \sum_{i=1}^{k} \sigma_i^{t-1} y_i$$
$$\text{s.t.} \quad X_a \in \Omega_a, \; X_a = [X^{\mathrm{T}} y_1 \dots y_k]^{\mathrm{T}}, \tag{4}$$

$$\Omega_a = \{X_a | f_i(X) \geqslant f_i(X^{t-1}) - \Delta f_i^{t-1} + y_i, \; y_i \geqslant 0, \; i = 1, \dots, k; \; X \in \Omega\}.$$

This model may be regarded as an extension of the original auxiliary model as reported in [28].

In the above model, X is the original variable vector, y_i an auxiliary variable, X^{t-1} a given (efficient) solution, $\sigma_i^{t-1} \geqslant 0$ and $\Delta f_i^{t-1} \geqslant 0$. Ω_a defines a local region around $F(X^{t-1})$ within the feasible objective space. σ_i^{t-1} is referred to as a normalising and weighting coefficient for objective i. Δf_i^{t-1} is a nonnegative decrement of objective i at the current stage. It can be proven that EAP^t possesses the following features.

Lemma 1. *(i) The local region Ω_a defined for EAP^t is always nonempty if X^{t-1} is a feasible solution of MOP. (ii) If X_a^t is an optimal solution of EAP^t with $\sigma_i^{t-1} > 0$ $(i = 1, \dots, k)$, where $X_a^t = [(X^t)^{\mathrm{T}} y_1^t \dots y_k^t]^{\mathrm{T}}$, then*

$$f_i(X^t) = f_i(X^{t-1}) - \Delta f_i^{t-1} + y_i^t, \quad i = 1, \dots, k. \tag{5}$$

Proof. (i) Let $y_i^0 = \Delta f_i^{t-1}$ and $X_a^0 = [(X^{t-1})^{\mathrm{T}} y_1^0 \dots y_k^0]^{\mathrm{T}}$. Since $X^{t-1} \in \Omega$ and $\Delta f_i^{t-1} \geqslant 0$, we have $X_a^0 \in \Omega_a$.

(ii) Suppose conclusion (5) is not true for some $i \in \{1, \dots, k\}$, say i_1. Then, we have

$$f_{i_1}(X^t) > f_{i_1}(X^{t-1}) - \Delta f_{i_1}^{t-1} + y_{i_1}^t.$$

Let's define a solution $X_a' = [(X^t)^{\mathrm{T}} y_i^t \dots y_{i_1}' \dots y_k^t]^{\mathrm{T}}$ where y_{i_1}' is determined by

$$f_{i_1}(X^t) = f_{i_1}(X^{t-1}) - \Delta f_{i_1}^{t-1} + y_{i_1}'.$$

Obviously, $X_a' \in \Omega_a$. However, from the above equations we have $y_{i_1}' > y_{i_1}^t$. As $\sigma_i^{t-1} > 0$, we then get

$$\sum_{\substack{i=1 \\ i \neq i_1}}^{k} \sigma_i^{t-1} y_i^t + \sigma_{i_1}^{t-1} y_{i_1}' > \sum_{i=1}^{k} \sigma_i^{t-1} y_i^t,$$

which contradicts our assumption that X_a^t is an optimal solution of EAP^t.

As $\sigma_i^{t-1} > 0$ and $y_i \geqslant 0$ for all $i = 1, \dots, k$, such contradiction can also be detected if conclusion (5) is assumed to be false for more than one i. □

Theorem 1. *Suppose X_a^t is an optimal solution of EAP^t with $X_a^t = [(X^t)^{\mathrm{T}} y_1^t \dots y_k^t]^{\mathrm{T}}$.*

(i) If $\sigma_i^{t-1} \geqslant 0$ for all $i = 1, \dots, k$, then X^t is an efficient or a weakly efficient solution of MOP, dependent upon whether X_a^t is unique or not.

(ii) Any efficient solution of MOP can be generated by solving EAP^t with regulating $f_i(X^{t-1}), \sigma_i^{t-1}$ and $\Delta f_i^{t-1} (i = 1, \dots, k)$.

(iii) If $\sigma_i^{t-1} > 0$ for all $i = 1, \dots, k$, then X^t is an efficient solution of MOP.

Proof. (i) and (ii) As the auxiliary model developed in [28] is a special case of EAP^t, these two conclusions can be proven in the same way as for Theorem 1 in [28].

(iii) Let $F(X) = [f_1(X) \dots f_k(X)]^{\mathrm{T}}$, $Y = [y_1 \dots y_k]^{\mathrm{T}}$, $F^{t-1} = [f_1^{t-1} \dots f_k^{t-1}]^{\mathrm{T}}$ with $f_i^{t-1} = f_i(X^{t-1})$ and $\Delta F^{t-1} = [\Delta f_1^{t-1} \dots \Delta f_k^{t-1}]^{\mathrm{T}}$. Define the following two subspaces, Ω_1 and Ω_2

$$\Omega_1 = \{X_a | F(X) \geqslant F^{t-1} - \Delta F^{t-1} + Y, \; Y \geqslant 0\},$$

$$\Omega_2 = \{X_a | F(X) \ngeqslant F^{t-1} - \Delta F^{t-1} + Y, \; Y \geqslant 0\},$$

where the symbol "$\ngeqslant$" means that there exists at least one "<" relation between the elements of a vector inequality. Note that $\Omega_1 \cup \Omega_2 = \mathbb{R}^{n+k}$.

For convenience, Ω of problem (1) is redefined as follows. Let $X_e = [X^{\mathrm{T}}\ 0\dots0]^{\mathrm{T}}$ so that X_e has the same dimension as X_a. Then, Ω is redefined as the following equivalent

$$\Omega = \{X_e | g_i(X_e) \leqslant 0;\ h_j(X_e) = 0;\ i = 1,\dots,m_1;\ j = 1,\dots,m_2\},$$

where g_i and h_j are inequality and equality constraint functions. As $\Omega \subseteq \mathbb{R}^{n+k}$, we have

$$\Omega = \Omega \cap \mathbb{R}^{n+k} = \Omega \cap (\Omega_1 \cup \Omega_2) = (\Omega \cap \Omega_1) \cup (\Omega \cap \Omega_2) = \Omega_a \cup \Omega_b,$$

where $\Omega_b = \Omega \cap \Omega_2$.

For any $X_a \in \Omega$, we have either $X_a \in \Omega_b$ or $X_a \in \Omega_a$. Therefore, we only need to consider the following two cases:

(a) For any $X_a \in \Omega_b$, there must be at least one $l(l \in \{1,\dots,k\})$, so that

$$f_l(X) < f_l^{t-1} - \Delta f_l^{t-1} + y_l \text{ for any } y_l \geqslant 0.$$

Let $y_l = 0$. We then have $f_l(X) < f_l^{t-1} - \Delta f_l^{t-1}$. Considering Lemma 1 and $y_l^t \geqslant 0$, we can thus get for any $X_a \in \Omega_b$

$$f_l(X^t) = f_l^{t-1} - \Delta f_l^{t-1} + y_l^t \geqslant f_l^{t-1} - \Delta f_l^{t-1} > f_l(X),$$

which means that $F(X) \ngeqslant F(X^t)$ for any $X_a \in \Omega_b$.

(b) Suppose there exists some $\bar{X}_a \in \Omega_a$ with $\bar{X}_a = [\bar{X}^{\mathrm{T}}\ \bar{y}_1 \dots \bar{y}_k]^{\mathrm{T}}$, such that

$$F(\bar{X}) \geqslant F(X^t) \quad \text{and} \quad f_l(\bar{X}) > f_l(X^t)$$
for at least one $l \in \{1,\dots,k\}$.

Define $\bar{Y} = [\bar{y}_1 \dots \bar{y}_k]^{\mathrm{T}}$ where $\bar{Y}$ is determined so that

$$F(\bar{X}) = F(X^{t-1}) - \Delta F(X^{t-1}) + \bar{Y}.$$

From the above assumption and Lemma 1, we have

$$F(\bar{X}) \geqslant F(X^t) = F(X^{t-1}) - \Delta F(X^{t-1}) + Y^t,$$

$$f_l(\bar{X}) > f_l(X^t) = f_l(X^{t-1}) - \Delta f_l(X^{t-1}) + y_l^t,$$

where $Y^t \geqslant 0$. Hence, we have

$$\bar{Y} \geqslant Y^t \quad \text{and} \quad \bar{y}_l > y_l^t$$

or equivalently

$$\bar{y}_i \geqslant y_i^t, \quad i = 1,\dots,k \text{ and } \bar{y}_l > y_l^t.$$

As $\sigma_i^{t-1} > 0$ and $y_i \geqslant 0$ for all $i = 1,\dots,k$, it is concluded that

$$\sum_{i=1}^{k} \sigma_i^{t-1} y_i^t < \sum_{i=1}^{k} \sigma_i^{t-1} \bar{y}_i$$

which, however, contradicts our assumption that X_a^t is an optimal solution of EAP^t.

As a result of the above discussion, we finally conclude that there exists no $X_a \in \Omega = \Omega_a \cup \Omega_b$ such that $F(X) \geqslant F(X^t)$ and $f_l(X) > f_l(X^t)$ for at least one $l \in \{1,\dots,k\}$. □

Remark 1. The above proof does not necessarily require that Ω be convex or X^{t-1} be feasible although Ω_a should be assigned to be nonempty [28]. This may facilitate the selection of an initial point X^0. Lemma 1 and Theorem 1 show that a new efficient solution X^t can be generated from a current efficient solution X^{t-1} by solving problem (4) given $\sigma_i^{t-1} > 0$ and $\Delta f_i^{t-1} \geqslant 0$, regardless of the convexity of Ω. Note that X^{t-1} is a feasible solution of EAP^t.

Whether X^t is better than X^{t-1} or not depends upon how σ_i^{t-1} and Δf_i^{t-1} are assigned. The assignment of σ_i^{t-1} and Δf_i^{t-1} should therefore be related to the improvement of a utility function. Such a relation could be established as explained in Fig. 1.

In Fig. 1(a), for example, $F(\Omega)$ is the projection of Ω into the objective space, point A represents an efficient solution $F(X^{t-1})$, $\nabla u(X^{t-1})$ is the gradient of u at $F(X^{t-1})$, S_u^{t-1} is the tangent plane of the contour of u at $F(X^{t-1})$ and N^{t-1} and S_e^{t-1} are the normal vector and the tangent plane (line in this example) of the efficient frontier at $F(X^{t-1})$. If $\nabla u(X^{t-1})$ is projected onto S_e^{t-1}, one may find that the utility function will increase in a neighbourhood of $F(X^{t-1})$ by improving f_1 at the expense of f_2. In the case of Fig. 1(b), improving f_2 at the expense of f_1 will increase u in a neighbourhood of $F(X^{t-1})$.

Hence, σ_i^{t-1} and Δf_i^{t-1} could be assigned based on $\nabla u(X^{t-1})$ and its projection onto S_e^{t-1}. $\nabla u(X^{t-1})$

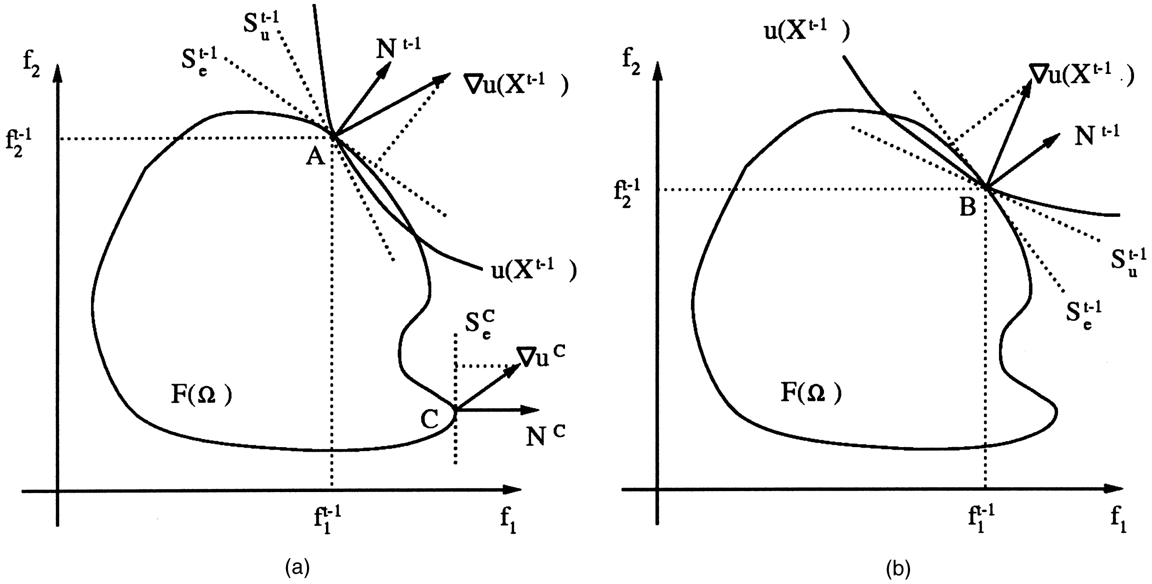

Fig. 1. (a) Trade-off analysis at point A. (b) Trade-off analysis at point B.

may be calculated or estimated using local iterative or preference information, dependent upon whether u is given explicitly or implicitly. If N^{t-1} can be identified, it is then possible to calculate the projection of $\nabla u(X^{t-1})$ onto S_e^{t-1}.

2.2. The normal vectors and optimality conditions

A normal vector at an efficient point can be identified using the minimax solution scheme, as proposed in [15,33]. The following results extend Theorem 1 of [33] so that a normal vector at an efficient solution can be obtained by solving a set of linear equations, whatever approaches may be used to generate this efficient solution.

Theorem 2. *Suppose X^t is an efficient solution and f_i^* is the best feasible value of objective i. Let $\omega_i^t = 1/(f_i^* - f_i^t)$ $(i = 1, \ldots, k)$. If $(f_i^* - f_i^t) > 0$ for all $i = 1, \ldots, k$, then X^t is an optimal solution of the following minimax problem:*

$$\min_{X\in\Omega} \max_i \{\omega_i^t(f_i^* - f_i(X))\}. \tag{6}$$

Proof. Suppose X^t is not an optimal solution of formulation (6). Note that problem (6) has the following equivalent [23]:

$$\begin{aligned} \min \quad & r \\ \text{s.t.} \quad & \omega_i^t(f_i^* - f_i(X)) \leqslant r, \quad i = 1, \ldots, k, \quad r \geqslant 0, \\ & X \in \Omega. \end{aligned} \tag{7}$$

Suppose $\bar{X}$ is an optimal solution of Eq. (6). Let $\bar{r}$ be the optimal value with respect to $\bar{X}$ and r^t with respect to X^t. Obviously, $r^t = 1$ at X^t. Then, there should be $\bar{r} < r^t$. We thus deduce

$$\omega_i^t(f_i^* - f_i(\bar{X})) \leqslant \bar{r} < r^t = 1 \quad \text{for all } i = 1, \ldots, k.$$

As $\omega_i^t > 0$ for all $i = 1, \ldots, k$, we get

$$f_i^* - f_i(\bar{X}) < \frac{1}{\omega_i^t} = f_i^* - f_i(X^t) \quad \text{for all } i = 1, \ldots, k.$$

In other words

$$f_i(\bar{X}) > f_i(X^t) \quad \text{for all } i = 1, \ldots, k.$$

Thus, X^t is dominated by $\bar{X}$, which contradicts our assumption. □

A similar conclusion was also suggested using the concept of canonical weights [16]. It should be noted that an optimal solution of formulation (6) may not necessarily be an efficient solution of problem (1) if $\omega_i^t \geqslant 0$ [23]. For identifying a normal vector, we have the following results.

Theorem 3. *Let X^t be an efficient solution, f_i^* the best feasible value of objective i and $\omega_i^t = 1/(f_i^* - f_i^t)$ with $(f_i^* - f_i^t) > 0$ $(i = 1, \ldots, k)$. Suppose at X^t all $f_i(X)$, $g_j(X)$ and $h_p(X)$ are continuously differentiable and the constraint qualification condition holds [18]. Suppose λ_i^t is equivalent to the Kuhn–Tucker multiplier with respect to an objective constraint $\omega_i^t(f_i^* - f_i(X)) \leqslant r$ of problem (7) at X^t. Then,*

(i) The normal vector of the efficient frontier of problem (1) at $F(X^t)$ in the objective space, denoted by N^t, is given by

$$N^t = [\omega_1^t \lambda_1^t \ldots \omega_i^t \lambda_i^t \ldots \omega_k^t \lambda_k^t]^{\mathrm{T}}. \tag{8}$$

(ii) $\lambda_i^t (i = 1, \ldots, k)$ are given by solving the following linear equations:

$$\sum_{i=1}^{k} \lambda_i = 1 - \sum_{i=1}^{k} \omega_i^t \frac{\partial f_i(X^t)}{\partial x_l} \lambda_i + \sum_{j \notin J_g} \frac{\partial g_j(X^t)}{\partial x_l} \beta_j + \sum_{p=1}^{m_2} \frac{\partial h_p(X^t)}{\partial x_l} \gamma_p = 0, \quad l = 1, \ldots, n, \tag{9}$$

with $\lambda_i, \beta_j, \gamma_p \geqslant 0$, $i = 1, \ldots, k$, $j = 1, \ldots, m_1$, $p = 1, \ldots, m_2$ and $\beta_j = 0$ for $j \notin J_g$, where $J_g = \{j | g_j(X^t) = 0, \ j \in \{1, \ldots, m_1\}\}$.

Proof. (i) With Theorem 2 in hand, the proof of conclusion (i) is the same as that of Theorem 1 in [33].

(ii) The Lagrangian formulation of problem (7) can be constructed as follows:

$$\min L(X, r, \lambda, \beta, \gamma) = r + \sum_{i=1}^{k} \lambda_i [\omega_i^t (f_i^* - f_i(X)) - r] + \sum_{j=1}^{m_1} \beta_j g_j(X) + \sum_{p=1}^{m_2} \gamma_p h_p(X),$$

where $\lambda = [\lambda_1 \cdots \lambda_k]^{\mathrm{T}}$. Theorem 2 shows that X^t is an optimal solution of problem (7) and $r^t = 1$ at X^t. The Kuhn–Tucker optimality condition at X^t reads

$$\sum_{i=1}^{k} \lambda_i = 1 - \sum_{i=1}^{k} \lambda_i \left(\omega_i^t \frac{\partial f_i(X^t)}{\partial x_l} \right) + \sum_{j=1}^{m_1} \beta_j \frac{\partial g_j(X^t)}{\partial x_l} + \sum_{p=1}^{m_2} \gamma_p \frac{\partial h_p(X^t)}{\partial x_l} = 0, \quad l = 1, \ldots, n,$$

$$\lambda_i(\omega_i^t (f_i^* - f_i(X^t)) - r^t) = 0, \quad i = 1, \ldots, k, \tag{10}$$

$$\beta_j g_j(X^t) = 0, \quad j = 1, \ldots, m_1,$$

$$h_p(X^t) = 0, \quad p = 1, \ldots, m_2.$$

From Theorem 2 $r^t = 1$. So $\omega_i^t(f_i^* - f_i(X^t)) - r^t = 1 - 1 = 0$. We then have $\lambda_i \geqslant 0$, $i = 1, \ldots, k$. When $g_j(X^t) = 0$, $\beta_j \geqslant 0$; when $g_j(X^t) < 0$, $\beta_j = 0$. Given X^t, equation (9) are thus equivalent to equations (10). □

Remark 2. If the gradients of the saturated objective and constraint functions at an efficient solution X^t are linearly independent, then the constraint qualification condition holds and X^t is referred to as a regular efficient solution [17,18]. At such a solution, the Kuhn–Tucker vector λ is unique and so is the normal vector.

Remark 3. If λ is not unique at X^t, X^t is referred to as an irregular solution. At such a solution, we could have multiple normal vectors.

It is easy to show by contradiction that the best compromise solution of problem (1) or the optimal solution of problem (3) is attained at an efficient solution [13,15]. The optimality of an efficient solution could be tested using the following theorem [33].

Theorem 4. *Suppose X^t is an efficient solution of problem (1) and the gradient of u at X^t is given by $\nabla u(X^t) = [\partial u(X^t)/\partial f_1 \ldots \partial u(X^t)/\partial f_k]^{\mathrm{T}}$. If X^t is an optimal solution of problem (3), then there exists a normal vector N^t at X^t so that the following condition is satisfied:*

$$\nabla u(X^t) = cN^t, \tag{11}$$

where N^t *is given by* Eq. (8) *and* c *is a constant.*

Remark 4. Condition (11) indicates that at an optimal solution the gradient of the utility function is proportional to the normal vector in the objective space. This condition is necessary and sufficient for X^t to be a stationary solution (optimum, local optimum or inflexion) [18]. In cases where u is concave and differentiable and $F(\Omega)$ is convex, then condition (11) is both necessary and sufficient for X^t to be an optimal solution [18].

Remark 5. If X^t is an irregular efficient solution, there will be multiple normal vectors at X^t. If X^t is also an optimal solution, then one of these normal vectors satisfies condition (11).

3. Gradient-based local region search algorithm

3.1. Gradient projection and local region search scheme

Graphically, condition (11) means that at an optimal solution the utility function must be tangent to the efficient frontier. In other words, the gradient $\nabla u(X^t)$ is not orthogonal to any tangent plane of the efficient frontier at an efficient solution X^t if condition (11) is not satisfied at X^t. In this case, we can always project $\nabla u(X^t)$ onto a tangent plant at X^t. This projection as given below provides an ascent direction of u from X^t.

Lemma 2. *Suppose* X^t *is an efficient solution of problem (1),* $\nabla u(X^t)$ *the gradient of* u *and* N^t *a normal vector at* X^t*. Let* $\mathrm{d}F^t$ *be the projection of* $\nabla u(X^t)$ *onto a tangent plane of the efficient frontier at* X^t *given by*

$$\mathrm{d}F^t = \nabla u(X^t) - \frac{[(\nabla u(X^t))^{\mathrm{T}} N^t]}{[(N^t)^{\mathrm{T}} N^t]} N^t. \tag{12}$$

Then, $\mathrm{d}F^t$ *is an ascent direction of* u *at* X^t *in the objective space.*

Proof. Multiplying the two sides of Eq. (12) by $(\nabla u(X^t))^{\mathrm{T}}$, we get

$$(\nabla u(X^t))^{\mathrm{T}} \mathrm{d}F^t = (\nabla u(X^t))^{\mathrm{T}} \nabla u(X^t) - \frac{[(\nabla u(X^t))^{\mathrm{T}} N^t]^2}{[(N^t)^{\mathrm{T}} N^t]}.$$

As $(\nabla u(X^t))^{\mathrm{T}} \nabla u(X^t) = |\nabla u(X^t)|^2, (N^t)^{\mathrm{T}} N^t = |N^t|^2$ and $(\nabla u(X^t))^{\mathrm{T}} N^t = |\nabla u(X^t)||N^t \mid \cos\theta$ while θ is the minimum angle between the two vectors $\nabla u(X^t)$ and N^t, we get

$$(\nabla u(X^t))^{\mathrm{T}} \mathrm{d}F^t = |\nabla u(X^t)|^2 (1 - \cos^2\theta) = |\nabla u(X^t)|^2 \sin^2\theta \geqslant 0. \qquad \square$$

Remark 6. Let $\mathrm{d}F^1 = [\mathrm{d}f_1^t \ldots \mathrm{d}f_i^t \ldots \mathrm{d}f_k^t]^{\mathrm{T}}$. If $\mathrm{d}F^t \neq 0$, it provides local trade-off information on whether an objective $f_i(X)$ should be improved, or must be kept at least unchanged or may be sacrificed in a neighbourhood of F^{t-1}, dependent upon whether $\mathrm{d}f_i^t$ is larger than, or equal to or smaller than zero. If $\mathrm{d}F^t = 0$, then we have the following conclusion.

Corollary 1. *If an efficient solution* X^t *is also an optimal solution of Eq. (3), then there exists a normal vector* N^t *at* X^t *so that* $\mathrm{d}F^t$ *is a zero vector, or*

$$\mathrm{d}F^t = 0. \tag{13}$$

Proof. From Theorem 4, we have $\nabla u(X^t) = cN^t$ if X^t ifs an optimal solution of Eq. (3). So

$$\mathrm{d}F^t = \nabla u(X^t) - \frac{[(\nabla u(X^t))^{\mathrm{T}} N^t]}{[(N^t)^{\mathrm{T}} N^t]} N^t = cN^t - cN^t = 0.$$

$\square$

Remark 7. Condition (13) provides an alternative way to test the optimality of an efficient solution and it is easy to check in an iterative optimisation process. From Theorem 4 and Corollary 1, it is straightforward to obtain the following condition [13,33].

Corollary 2. *If an efficient solution* X^t *is an optimal solution of problem (3) and if* $\omega_i^t \lambda_i^t > 0$ *for* $i = 1, \ldots, k$*, then there exists a normal vector* N^t at X^t *so that the following condition is satisfied:*

$$\frac{\partial u/\partial f_1}{\omega_1^t\lambda_1^t} = \cdots = \frac{\partial u/\partial f_i}{\omega_i^t\lambda_i^t} = \cdots = \frac{\partial u/\partial f_k}{\omega_k^t\lambda_k^t}. \tag{14}$$

If $\omega_i^t\lambda_i^t = 0$ *for any* i, *then* $\partial u/\partial f_i = 0$.

Remark 8. When the utility function is not known explicitly, condition (14) is useful to test the optimality of an efficient solution. This is because in this case the gradients of the utility function may only be assessed interactively using the DM's local preferences, such as marginal rates of substitution. Condition (14) can then be used to facilitate the assessment by defining the so called optimal indifference trade-offs, as discussed later.

From Corollary 1, X^{t-1} is not an optimal solution if $\mathrm{d}F^{t-1} \neq 0$. In this case, we have $\mathrm{d}f_i^{t-1} \neq 0$ for at least one i ($i \in \{1,\ldots,k\}$). From the definition of efficiency [2,11], it is clear that an objective can only be improved from an efficient point at the expense of at least one of the other objectives. This implies that if $\mathrm{d}f_i^{t-1} > 0$ there should be $\mathrm{d}f_j^{t-1} < 0$ for at least one j ($i,j \in \{1,\ldots,k\}$ $i \neq j$). Otherwise, no compromise could be accommodated. It is easy to show that $\mathrm{d}F^{t-1} \ngeqslant 0$. This is because $(\mathrm{d}F^{t-1})^{\mathrm{T}} N^{t-1} = 0$ while $N^{t-1} \geqslant 0$ but $N^{t-1} \not\equiv 0$.

At certain points, it is possible that $\mathrm{d}F^{t-1} \geqslant 0$, or $\mathrm{d}f_i^{t-1} \geqslant 0$ for all $i = 1,\ldots,k$ and $\mathrm{d}f_j^{t-1} = 0$ for at least one j ($j \in \{1,\ldots,k\}$). For instance, point C in Fig. 1(a) is such a point where the normal vector N^C is orthogonal to the ordinate f_2. When the gradient of u at point C, or ∇u^C, is projected onto the tangent line S_e^C (parallel to the ordinate f_2), we have $\mathrm{d}f_1^C = 0$ and $\mathrm{d}f_2^C > 0$.

When $\mathrm{d}F^{t-1} \geqslant 0$, we may modify $\mathrm{d}F^{t-1}$ using a perturbation technique as discussed below so that $\mathrm{d}F^{t-1} \ngeqslant 0$. Suppose

$$J_1 = \{i \mid \mathrm{d}f_i^{t-1} > 0,\ i \in \{1,\ldots,k\}\}$$
$$\text{and} \tag{15}$$
$$J_2 = \{j \mid \mathrm{d}f_j^{t-1} = 0,\ j \in \{1,\ldots,k\}\}.$$

Let μ be a small positive constant. To ensure that the perturbed $\mathrm{d}F^{t-1}$ is still an ascent direction of u, it is necessary that $(\nabla u(X^{t-1}))^{\mathrm{T}}\mathrm{d}F^{t-1} \geqslant 0$. For $j \in J_2, f_j$ may thus be perturbed as follows.

$$\mathrm{d}f_j^{t-1} = -\mu|f_j^{t-1}| \quad \text{for all } j \in J_2$$

where

$$0 \leqslant \mu \leqslant \sum_{i\in J_1} \frac{\partial u(X^{t-1})}{\partial f_i}\mathrm{d}f_i^{t-1} \Bigg/ \sum_{j\in J_2} \frac{\partial u(X^{t-1})}{\partial f_j}|f_j^{t-1}|. \tag{16}$$

Hence, we can always obtain $\mathrm{d}F^{t-1} \ngeqslant 0$ with $(\nabla u(X^{t-1}))^{\mathrm{T}}\mathrm{d}F^{t-1} \geqslant 0$ if $\mathrm{d}F^{t-1} \neq 0$. As $\mathrm{d}F^{t-1}$ lies on the tangent plane at $F(X^{t-1})$, an infinitesimal movement along $\mathrm{d}F^{t-1}$ generally leads out of the feasible space immediately. However, $\mathrm{d}F^{t-1}$ provides local trade-off information on the objectives, which could be used to define a local region in which to search for a new (hopefully better) solution, instead of along the line $\mathrm{d}F^{t-1}$.

Note that the utility function will increase from $u(X^{t-1})$ by improving $f_i(X)$ at the expense of $f_j(X)$ in a neighbourhood of X^{t-1} if $\mathrm{d}f_i^{t-1} \geqslant 0$ and $\mathrm{d}f_j^{t-1} < 0$. We could then define the permissible sacrifice of objective i at the current stage as follows

$$\Delta f_i^{t-1} = \frac{\alpha}{2}\left(|\mathrm{d}f_i^{t-1}| - \mathrm{d}f_i^{t-1}\right), \quad i = 1,\ldots,k, \tag{17}$$

where α is a positive constant which can be regulated to ensure that u increases in a neighbourhood of X^{t-1}. Thus, $\Delta f_i^{t-1} = 0$ if $\mathrm{d}f_i^{t-1} \geqslant 0$ and $\Delta f_i^{t-1} > 0$ if $\mathrm{d}f_i^{t-1} < 0$. A local region within the feasible space around $F(X^{t-1})$, denoted by Ω_a, may then be defined as in problem (4).

In Ω_a of problem (4), y_i is an auxiliary variable and $(-\Delta f_i^{t-1} + y_i)$ represents an increment (or a decrement) of the ith objective. As $\mathrm{d}F^{t-1} \neq 0$, it is then possible to search for another solution improving u from $u(X^{t-1})$. The first order Taylor expansion of u around X^{t-1} reads

$$u(X) \approx u(X^{t-1}) + \nabla u(X^{t-1}))^{\mathrm{T}}(F(X) - F(X^{t-1})). \tag{18}$$

Within the local region defined by Ω_a, we then have

$$\begin{aligned}\Delta u(X) &= u(X) - u(X^{t-1}) \\ &= \sum_{i=1}^{k} \frac{\partial u(X^{t-1})}{\partial f_i}\left(f_i(X) - f_i(X^{t-1})\right) \\ &\geqslant \sum_{i=1}^{k} \frac{\partial u(X^{t-1})}{\partial f_i}\left(-\Delta f_i^{t-1} + y_i\right).\end{aligned} \tag{19}$$

In view of the fact that the term $\sum_{i=1}^{k} -\Delta f_i^{t-1} \partial u(X^{t-1})/\partial f_i$ is a constant given X^{t-1}, we can construct the following auxiliary problem:

$$\begin{aligned} \max \quad & \sum_{i=1}^{k} \frac{\partial u(X^{t-1})}{\partial f_i} y_i \\ \text{s.t.} \quad & X_a \in \Omega_a, \quad X_a = [X^{\mathrm{T}} y_1 \ldots y_k]^{\mathrm{T}}, \end{aligned} \tag{20}$$

the solution of which maximises the linear increment $\Delta u(X)$ of u within the local region.

Obviously, problem (20) is a special case of problem (4) as $\partial u(X^{t-1})/\partial f_i > 0$ and $\Delta f_i^{t-1} \geqslant 0$ $(i = 1, \ldots, k)$, as defined by Eqs. (2) and (17). From Lemma 1, the local region defined by Ω_a is never empty if X^{t-1} is a feasible solution of Ω, that is problem (20) has at least one solution (i.e. $X^0 = [(X^{t-1})^{\mathrm{T}} \Delta f_1^{t-1} \cdots \Delta f_k^{t-1}]^{\mathrm{T}}$).

3.2. Convergence analysis

Suppose X^t is an optimal solution of problem (20). Theorem 1 shows that X^t is an efficient solution of problem (1). Furthermore, if X^{t-1} is not the best compromise solution, X^t can always become more favourable than X^{t-1} by regulating α, as shown by the following results.

Theorem 5. *If X^{t-1} is an efficient solution of problem (1) but not an optimal solution of problem (3), then a sufficiently small (non-negative) α defined in Eq. (17) can be assigned, such that a new efficient solution X^t $(X^t \neq X^{t-1})$ can be obtained by solving problem (20) with $u(X^t) \geqslant u(X^{t-1})$.*

Proof. As discussed above, an ascent direction of u from $u(X^{t-1})$, denoted by $\mathrm{d}F^{t-1}$, can be obtained so that $\mathrm{d}F^{t-1} \ngeqslant 0$. Let

$$\begin{aligned} & I_1 = \left\{ i | \mathrm{d}f_i^{t-1} \geqslant 0, \quad i \in \{1, \ldots, k\} \right\} \\ & \text{and} \\ & I_2 = \left\{ i | \mathrm{d}f_i^{t-1} < 0, \quad i \in \{1, \ldots, k\} \right\}. \end{aligned} \tag{21}$$

Thus, $I_1 \cup I_2 = \{1, \ldots, k\}$, $I_1 \cap I_2 = \emptyset, I_1 \neq \emptyset$ and $I_2 \neq \emptyset$.

The current solution X^{t-1} is denoted in the objective space by $F(X^{t-1}) = [f_1^{t-1} \cdots f_k^{t-1}]^{\mathrm{T}}$. A new solution $\overline{X}$ in a neighbourhood of X^{t-1} is denoted by $F(\overline{X}) = [\bar{f}_1 \cdots \bar{f}_k]^{\mathrm{T}}$ in the objective space, where $\bar{f}_i = f_i^{t-1} - \Delta f_i^{t-1} + \bar{y}_i$. Suppose δ_i $(i = 1, \ldots, k)$ are small nonnegative constants. Let $\bar{y}_i = \delta_i$ for $i \in I_1$ and $\bar{y}_j = \alpha |\mathrm{d}f_j^{t-1}| - \delta_j$ with $\delta_j \leqslant \alpha |\mathrm{d}f_j^{t-1}|$ for $j \in I_2$. Then

$$F(\overline{X}) = [\bar{f}_1 \cdots \bar{f}_k]^{\mathrm{T}} \quad \text{and} \\ \bar{f}_i = \begin{cases} f_i^{t-1} + \delta_i & \text{for } i \in I_1, \\ f_i^{t-1} - \delta_i & \text{for } i \in I_2. \end{cases} \tag{22}$$

A vector pointing from $F(X^{t-1})$ to $F(\overline{X})$ can thus be denoted by δ:

$$\delta = [\tau_1 \cdots \tau_i \cdots \tau_k]^{\mathrm{T}}, \ \tau_i = \begin{cases} \delta_i & \text{for } i \in I_1, \\ -\delta_i & \text{for } i \in I_2. \end{cases} \tag{23}$$

If X^{t-1} is not an optimal (or a local optimal or an inflexion) solution of problem (3), then from Theorem 4 we have $\nabla u(X^{t-1}) \neq cN^{t-1}$. As $\nabla u(X^{t-1}), cN^{t-1} \geqslant 0$, we have

$$\theta = \cos^{-1}\Big(\big(\nabla u\big(X^{t-1}\big)\big)^{\mathrm{T}} N^{t-1} / |\nabla u\big(X^{t-1}\big)\|N^{t-1}| \Big) \\ > 0, \tag{24}$$

where θ is the minimum angle between the k-dimensional gradient vector $\nabla u(X^{t-1})$ and the k-dimensional normal vector N^{t-1}. Since the two vectors $\nabla u(X^{t-1})$ and N^{t-1} pass the same point $F(X^{t-1})$, then $\theta > 0$ means that they are not on the same line. In other words, θ is the minimum angle between the plane orthogonal to $\nabla u(X^{t-1})$ and the tangent plane of the efficient frontier at $F(X^{t-1})$ which is orthogonal to N^{t-1}.

At $F(X^{t-1})$, the tangent planes of the efficient frontier and the utility contour at $u(X^{t-1})$ are denoted by S_e^{t-1} and S_u^{t-1}, respectively. Beneath S_e^{t-1} (in the direction of $-N^{t-1}$), an acute cone can be constructed between S_e^{t-1} and S_u^{t-1} with $F(X^{t-1})$ being at the top of the cone. This cone is spanned by all the vectors δ emerging from $F(X^{t-1})$ which satisfy the following conditions:

$$\big(N^{t-1}\big)^{\mathrm{T}} \delta \leqslant 0 \quad \text{and} \quad \big(\nabla u\big(X^{t-1}\big)\big)^{\mathrm{T}} \delta \geqslant 0. \tag{25}$$

$\theta > 0$ means that all the vectors δ satisfying Eq. (25) must not be on the same plane as the tangent plane of the efficient frontier. In other words, in the cone defined by Eq. (25) there exists at least a vector δ such that $(N^{t-1})^{\mathrm{T}} \delta < 0$ and

$(\nabla u(X^{t-1}))^{\mathrm{T}}\delta \geqslant 0$. A solution $\overline{X}$, defined by Eq. (22) with such a vector, will become a feasible solution of problem (1) with $u(\overline{X}) \geqslant u(X^{t-1})$ if $|\delta|$ is sufficiently small. Let's construct the following feasible subspace Ω_δ

$$\Omega_\delta = \left\{ X_a \left| \begin{array}{l} f_i(X) \geqslant f_i^{t-1} + \delta_i, \quad y_i = \delta_i \qquad \text{for } i \in I_1, \\ f_j(X) \geqslant f_j^{t-1} - \delta_j, \quad y_j = \alpha|\mathrm{d}f_j^{t-1}| - \delta_j \\ \qquad \text{for } j \in I_2, \\ \left(N^{t-1}\right)^{\mathrm{T}}\delta \leqslant 0, \qquad \left(\nabla u\left(X^{t-1}\right)\right)^{\mathrm{T}}\delta \geqslant 0, \\ X_a = \left[X^{\mathrm{T}} y_1 \cdots y_k\right]^{\mathrm{T}}, \quad X \in \Omega, \ , X \neq X^{t-1}. \end{array} \right. \right\} \tag{26}$$

As $\theta > 0, \Omega_\delta \neq \emptyset$ if $\alpha > 0$. Moreover, $\Omega_\delta \subseteq \Omega_a$ if $\delta_j \leqslant \alpha|\mathrm{d}f_j^{t-1}|$ for all $j \in I_2$. This indicates that if $\theta > 0$ then there exists a feasible solution in $\Omega_\delta \subseteq \Omega_a$ for a sufficiently small α, at which the utility is better than that at X^{t-1}. Solving Eq. (20) will then find a feasible solution X^t which maximises the linear increment of u, or the first order term of the Taylor expansion of u denoted by $\Delta u(X^t)$. When α approaches zero, this linear increment of u will become equivalent to the real increment of u, that is

$$u(X^t) \approx u(X^{t-1}) + \Delta u(X^t) \geqslant u(X^{t-1}).$$

From Theorem 1, X^t must be an efficient solution of problem (1). $\square$

The step size α may be taken as a small constant, or more elaborately determined as follows. If $\mathrm{d}F^{t-1} \neq 0$, a better (usually infeasible) solution along $\mathrm{d}F^{t-1}$ can be sought. As u is a nonlinear function of $f_i(X) (i = 1, \ldots, k)$, u will not always increase along $\mathrm{d}F^{t-1}$. So, let us define a maximum step size α_1^{t-1} along $\mathrm{d}F^{t-1}$ as an optimal solution of the following one-dimension search problem:

$$\max_{\alpha_1} u(X) = u\left(F\left(X^{t-1}\right) + \alpha_1 \mathrm{d}F^{t-1}\right). \tag{27}$$

As $\mathrm{d}F^{t-1}$ is generally an infeasible direction, the designated decrement Δf_j^{t-1} in $f_j(X)$ for $j \in I_2$, as defined by Eq. (17) with $\alpha = \alpha_1^{t-1}$, may not result in the expected increase $\alpha_1^{t-1}\mathrm{d}f_i^{t-1}$ in $f_i(X)$ for $i \in I_1$ by solving auxiliary problem (20). In convex cases, $f_i(X^t) \leqslant f_i(X^{t-1}) + \alpha_1^{t-1}\mathrm{d}f_i^{t-1}$ for $i \in I_1$. If the feasible increases in $f_i(X)$ for $i \in I_1$ are too small, this may cause the decrease of $u(X)$ rather than increase. In such cases, α needs to be modified as follows:

$$\alpha = \alpha_2 \alpha_1^{t-1}, \tag{28}$$

where α_2 is a nonnegative coefficient with $0 \leqslant \alpha_2 \leqslant 1$. Thus, α_1^{t-1} is used as the largest step size at stage $t-1$. Regulating α_2 may lead to a different solution of problem (20).

α_2 may be regulated using one-dimension search techniques, such as the dichotomy method [18]. As this may require too much calculation effort, we simply halve at each step the length of the feasible interval for α_2, which is initially [0 1]. We start to assign $\alpha_2 = 1$. If at $\alpha_2 = 1$ the solution X^t of problem (20) is not better than X^{t-1} in terms of utility, then let α_2 be equal to the midpoint of [0 1], or $\alpha_2 = 0.5$. If at $\alpha_2 = 0.5$ X^t is not better either, take the midpoint of [0 0.5], or $\alpha_2 = 0.25$ and so on. Theorem 5 ensures that if $\mathrm{d}F^{t-1} \neq 0$ we can always find X^t with $u(X^t) \geqslant u(X^{t-1})$ as α_2 approaches zero.

Theorem 5 could be explained using Fig. 2. In Fig. 2(a), point A denotes the current solution $F(X^{t-1})$ in the objective space. N^{t-1} and S_e^{t-1} are the normal vector and the tangent plane of the efficient frontier at $F(X^{t-1})$. Curve (1) is the utility contour at $u(X^{t-1})$. $\nabla u(X^{t-1})$ and S_u^{t-1} are the gradient and the tangent plane of $u(X)$ at point A.

If $\nabla u(X^{t-1})$ is projected onto S_e^{t-1} we obtain the trade-off direction $\mathrm{d}F^{t-1}$, indicating that in order to increase $u(X)$ from $u(X^{t-1})$ f_2 needs to be increased at the expense of f_1. By maximising $u(X) = u(F(X^{t-1}) + \alpha_1 \mathrm{d}F^{t-1})$ along $\mathrm{d}F^{t-1}$, we obtain α_1^{t-1} while $(F(X^{t-1}) + \alpha_1^{t-1}\mathrm{d}F^{t-1})$ is denoted by point B. The utility at point B is $u(F(X^{t-1}) + \alpha_1^{t-1}\mathrm{d}F^{t-1})$. Generally, $u(F(X^{t-1}) + \alpha_1^{t-1}\mathrm{d}F^{t-1}) \geqslant u(X^{t-1})$ and point B is infeasible. Point C is the intersection of the efficient frontier and line $\overline{BD}$ parallel to f_2 and passing point B.

Suppose $\alpha_2 = 1$. Then, the area enclosed by curve $\widetilde{AC}$, line $\overline{CD}$ and line $\overline{DA}$ is the local region Ω_a as defined in problem (4). The area enclosed by curve $\widetilde{AC}$ and lines $\overline{Cd}$ and $\overline{dA}$ without point A is the feasible subspace Ω_δ as defined by Eq. (26)

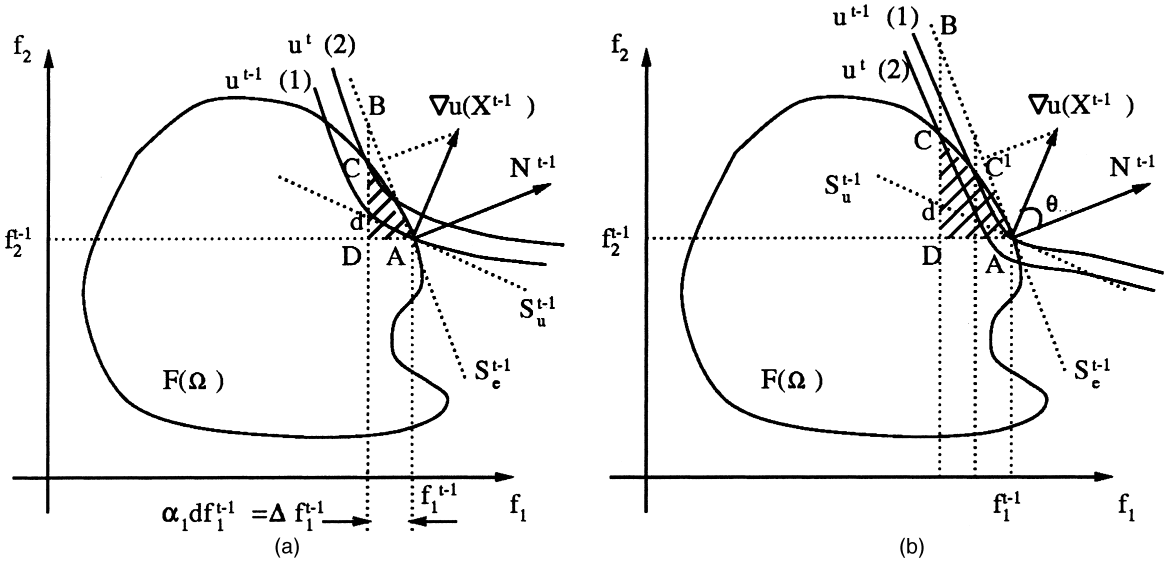

Fig. 2. (a) Graphic interpretation of theorem 5. (b) Modification of step sizes.

where point d is the intersection of lines $\overline{BD}$ and S_u^{t-1}. Obviously, $\Omega_\delta \subseteq \Omega_a$. Solving problem (20) is thus equivalent to moving line S_u^{t-1} in parallel in the direction $\nabla u(X^{t-1})$ as far as possible so that the intersection of Ω_a and S_u^{t-1} is not empty. This leads to a solution X^t while $F(X^t)$ is denoted by point C. It is desirable that $u(X^t) \geqslant u(X^{t-1})$. This is the case in Fig. 2(a), where curve (2) is the utility contour at $u(X^t)$.

If $u(X^t) < u(X^{t-1})$, as shown in Fig. 2(b), then α_2 must be reduced. This means that point B moves closer to point A along $\mathrm{d}F^{t-1}$ on S_e^{t-1}, or line $\overline{BD}$ moves closer to point A in parallel. If α_2 is reduced so that line $\overline{BD}$ has passed point C^1, the intersection of the efficient frontier and the utility contour at $u(X^{t-1})$ (curve (1)), then $u(X^t) \geqslant u(X^{t-1})$. Thus, a sufficiently small α_2 can be assigned so that $u(X^t) \geqslant u(X^{t-1})$ if $\theta > 0$ or $\nabla u(X^{t-1})$ is not proportional to N^{t-1}.

3.3. Iterative algorithm and its convergence properties

Summarising the above discussions, we can develop the following iterative algorithm for multiobjective optimisation with a utility function known explicitly.

Step 1. Define a multiobjective optimisation problem as shown in Eq. (1) and assign the best and worst values for each objective by constructing for example the pay-off table [9,28].

Step 2. Select an initial point X^0 which may not necessarily be feasible. For instance, at $t = 0$, $(f_i^{-1} - \Delta f_i^{-1})$ in problem (4) may be assigned to the worst value of objective i in the pay-off table and $\partial u(X^{-1})/\partial f_i = 1$ for $i = 1, \ldots, k$. Then, solving Eq. (20) will lead to an efficient solution, denoted by X^0.

Step 3. At X^{t-1}, calculate the normal vector N^{t-1} using Eqs. (8) and (9), and the gradient $\nabla u(X^{t-1})$ of the utility function.

Step 4. Calculate the projection $\mathrm{d}F^{t-1}$ of $\nabla u(X^{t-1})$ using Eq. (12). If $\mathrm{d}F^{t-1} = 0$, the optimal solution is obtained and stop. Otherwise, if $\mathrm{d}F^{t-1} \ngeqslant 0$, then continue; if $\mathrm{d}F^{t-1} \geqslant 0$, perturb the objective $f_j(X)$ for all $j \in J_2$ as shown by Eqs. (15) and (16) and then continue.

Step 5. Obtain the maximum step size α_1^{t-1} by solving Eq. (27) and let $\alpha_2 = 1$.

Step 6. Let $\alpha = \alpha_2 \alpha_1^{t-1}$ and define the local region Ω_a of problem (4) using Eq. (17). Then, construct and solve the auxiliary problem as defined by Eq. (20), yielding a new efficient solution X^t. If X^t is an irregular point and if $F(X^t) = F(X^{t-1})$, X^t is an optimal solution and stop. Note that if original MOP problem (1) is linear then the auxiliary problem (20) is linear as well and can be solved using simplex method. Otherwise, problem (20) is a general nonlinear scalar problem and can be solved by for example sequential linear programming [18,29].

Step 7. If $u(X^t) \geqslant u(X^{t-1})$, let $t = t + 1$ and then go to Step 3. Otherwise, let $\alpha_2 = \alpha_2/2$ and then go to Step 6.

Regarding the convergence of the above algorithm, it is straightforward to draw the following conclusions.

Corollary 3. *For a MOP problem as defined by Eqs. (1)–(3), the above algorithm composed of Steps 1 – 7 generates a series of efficient solutions* $\{X^0, X^1, \ldots, X^{t-1}, X^t, \ldots\}$ *with* $u(X^t) \geqslant u(X^{t-1})$. *The solution series converges to a stationary solution of problem (3).*

The efficiency of the solutions is guaranteed by Theorem 1. The convergence of the solution series is secured by Theorem 4 and Theorem 5, though the algorithm may end up with an inflexion solution if such a solution does exist at all. The convergence of the algorithm will be demonstrated in the following numerical studies.

If problem (1) is linear and a utility function u is a linear function of objectives as well, then from any starting efficient solution X^0 the algorithm can arrive at the optimal solution of problem (3) in one iteration. This is because in this case $u(X)$ defined in problem (27) always increases with α_1 along the direction $\mathrm{d}F^{t-1}$. Select a sufficiently large α_1 so that

$$f_i(X^{t-1}) - \Delta f_i^{t-1} \leqslant f_i^- \quad \text{for all } \Delta f_i^{t-1} > 0, \qquad (29)$$

where f_i^- is the worst value of $f_i(X)$ and Δf_i^{t-1} is defined by Eq. (17). Then the best compromise solution must lie within Ω_a as defined by problem (4) and Eqs. (17) and (29). Solving problem (20) can then identify the best solution.

If problem (1) is linear but u is nonlinear, then in the worst case the algorithm converges within a maximum number of $M \times \tau$ iterations where M is the total number of constraints and τ is the smallest integer satisfying the following inequality:

$$\left(\frac{1}{2}\right)^\tau \leqslant \zeta, \qquad (30)$$

with ζ being a given small constant defining the convergence precision. Given $\zeta = 10^{-6}$, for example, $\tau = 20$.

The above conclusion relies on the fact that in this second case the efficient solution frontier is composed of hyperplanes. The maximum number of efficient hyperplanes is bounded by the total number M of constraints including any constraints on the lower and upper bounds of variables. Therefore the normal vector only needs to be calculated M times at most, and so does the step size α_1. In the worst case, α_2 might need to be regulated until $\alpha_2 \leqslant \zeta$ for every α_1 obtained, which is normally unlikely.

If both problem (1) and u are nonlinear, then the algorithm may converge at a speed similar to the steepest ascent method as the search direction is determined by the first order derivatives of the utility function and the step size α_1 is set using one dimensional search. However, α_2 may need to be regulated from time to time, which will slow down the convergence process.

4. Gradient-based interactive step trade-off algorithm

4.1. Gradient-based trade-off analysis

In most MODM problems, utility functions are not known explicitly. In such circumstances, the gradients of utility functions need to be assessed using local preference information so that the auxiliary problem defined by Eq. (20) can be constructed. This section is aimed to develop a gradient-based interactive step trade-off (GRIST) algorithm using the DM's indifference trade-offs or marginal rates of substitution. This algorithm adopts an explicit trade-off analysis, complementary to the implicit trade-off analysis proposed in

ISTM [28]. Suppose $f_l(X)$ is arbitrarily chosen as the reference function with $\partial u/\partial f_l > 0$. Let's define $\sigma^{t-1} = [\sigma_1^{t-1} \cdots \sigma_k^{t-1}]^{\mathrm{T}}$ as a weighting vector or a normalised gradient of u with regards to the reference function $f_l(X)$ and $\mathrm{d}\overline{F}^{t-1}$ as the projection of σ^{t-1} onto the tangent plane at X^{t-1} as follows:

$$\sigma_i^{t-1} = \frac{\partial u(X^{t-1})/\partial f_i}{\partial u(X^{t-1})/\partial f_l},$$

$$\mathrm{d}\overline{F}^{t-1} = \sigma^{t-1} - \frac{[(\sigma^{t-1})^{\mathrm{T}} N^{t-1}]}{[(N^{t-1})^{\mathrm{T}} N^{t-1}]} N^{t-1}. \tag{31}$$

Then, we have

$$\sigma^{t-1} = \frac{\nabla u(X^{t-1})}{\partial u(X^{t-1})/\partial f_l},$$

$$\mathrm{d}\overline{F}^{t-1} = \frac{\mathrm{d}F^{t-1}}{\partial u(X^{t-1})/\partial f_l}. \tag{32}$$

The local region could thus be defined as follo ws:

$$\Omega_a = \{X_a | f_i(X) \geqslant f_i(X^{t-1}) - \Delta f_i^{t-1} + y_i,\ y_i \geqslant 0,\ i = 1,\ldots,k;\ X \in \Omega\}, \tag{33}$$

$$\Delta f_i^{t-1} = \frac{\overline{\alpha}}{2}\left(\left|\mathrm{d}\bar{f}_i^{t-1}\right| - \mathrm{d}\bar{f}_i^{t-1}\right),\quad i = 1,\ldots,k,\ \overline{\alpha} = \frac{\partial u(X^{t-1})}{\partial f_l}\alpha. \tag{34}$$

As dividing an objective function by a positive constant does not change its optimum, we can define the following problem equivalent to problem (20):

$$\begin{aligned} \max \quad & \sum_{i=1}^{k} \sigma_i^{t-1} y_i \\ \text{s.t.} \quad & X_a \in \Omega_a, \quad X_a = [X^{\mathrm{T}} y_1 \cdots y_k]^{\mathrm{T}}. \end{aligned} \tag{35}$$

σ_i^{t-1} reflects the DM's trade-off between f_i and f_l at X^{t-1} and σ^{t-1} denotes the trade-off direction from X^{t-1}, which has to be determined before problem (35) can be solved. σ_i^{t-1} can be estimated using the DM's indifference trade-off or marginal rate of substitution as follows [8,11]. Suppose a small change Δf_l in f_l is exactly offset by a change Δf_i in f_i (i.e. the utility function is kept constant) while all other objectives remain unchanged. σ_i^{t-1} is then approximated by

$$\sigma_l^{t-1} = 1, \qquad \sigma_i^{t-1} \approx -\frac{\Delta f_l^{t-1}}{\Delta f_i^{t-1}}, \quad i = 1,\ldots,k;\ i \neq l. \tag{36}$$

The step size $\overline{\alpha}$ can be assigned by the DM. As the explicit form of $u(X)$ is not known, the following technique is suggested to facilitate the assignment. Similar to definition (21), we can define $\bar{I}_1$ and $\bar{I}_2$ as follows:

$$\bar{I}_1 = \left\{ i \middle| \mathrm{d}\bar{f}_i^{t-1} \geqslant 0,\ i \in \{1,\ldots,k\} \right\} \quad \text{and} \quad \bar{I}_2 = \left\{ i \middle| \mathrm{d}\bar{f}_i^{t-1} < 0,\ i \in \{1,\ldots,k\} \right\} \tag{37}$$

Note that $\Delta f_i^{t-1} = 0$ for $i \in \bar{I}_1$ and $\Delta f_i^{t-1} = \overline{\alpha}|\mathrm{d}\bar{f}_i^{t-1}|$ for $i \in \bar{I}_2$. Then, the maximum step size could be determined as the minimum of the largest permissible relative sacrifices of f_i $(i \in \bar{I}_2)$ from their current values f_i^{t-1}, or

$$\overline{\alpha}_{\max}^{t-1} = \min_{i \in \bar{I}_2} \{\overline{\alpha}_i^{t-1}\}, \qquad \overline{\alpha}_i^{t-1} = \frac{f_i^{t-1} - f_i^{-}}{\left|\mathrm{d}\bar{f}_i^{t-1}\right|}, \quad i \in \bar{I}_2, \tag{38}$$

where f_i^{-} is the lowest permissible value of objective i. The step size $\overline{\alpha}^{t-1}$ may thus be taken as a value of the interval $[0,\ \overline{\alpha}_{\max}^{t-1}]$ by the DM.

Suppose $\hat{\alpha}_l^{t-1} = \overline{\alpha}_{\max}^{t-1} l / C_\alpha$ $(l = 1,\ldots,C_\alpha)$, where C_α is an integer which could be regulated to change the accuracy of $\overline{\alpha}^{t-1}$. For instance, C_α could be initially taken as 10 and may be increased if the accuracy of $\overline{\alpha}^{t-1}$ needs to be improved. The step size may then be assigned as shown in Table 1 where the DM needs to identify the indifference trade-offs between f_i for $i \in \bar{I}_1$ and f_j for $j \in \bar{I}_2$.

In Table 1, f_j for all $j \in \bar{I}_2$ monotonically decrease and f_i for all $i \in \bar{I}_1$ monotonically increase with the increase of $\hat{\alpha}_l^{t-1}$ as the direction $\mathrm{d}\overline{F}^{t-1}$ is determined in the objective space. Such features

Table 1
Trade-off table for assignment of step size $\bar{\alpha}^{t-1}$

$\hat{\alpha}_l^{t-1}$	f_j $(j \in \bar{I}_2)$			f_i $(i \in \bar{I}_1)$		
	...	$f_j - \hat{\alpha}_l^{t-1}\lvert d\bar{f}_j^{t-1}\rvert$	...	...	$f_i^{t-1} + \hat{\alpha}_l^{t-1}\lvert d\bar{f}_i^{t-1}\rvert$	...
$\hat{\alpha}_0^{t-1}$	...	$f_j^{t-1}(\hat{\alpha}_0^{t-1})$	...	...	$f_i\ (\hat{\alpha}_0^{t-1})$	...
$\hat{\alpha}_1^{t-1}$	...	$f_j\ (\hat{\alpha}_1^{t-1})$	...	...	$f_i\ (\hat{\alpha}_1^{t-1})$	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$\hat{\alpha}_{C_\alpha}^{t-1}$	...	$f_j\ (\hat{\alpha}_{C_\alpha}^{t-1})$	...	...	$f_i\ (\hat{\alpha}_{C_\alpha}^{t-1})$	...

can help the DM to conduct an explicit trade-off analysis along $d\bar{F}^{t-1}$. If for a $\hat{\alpha}_l^{t-1}$ the sacrifices of f_j for all $j \in \bar{I}_2$ can be offset by the improvement of f_i for all $i \in \bar{I}_1$ and if any larger sacrifices can no longer be offset, then $\hat{\alpha}_l^{t-1}$ may be taken as the step size, or $\bar{\alpha}^{t-1} = \hat{\alpha}_l^{t-1}$.

However, f_j for any $j \in \bar{I}_2$ should not be reduced to be smaller than f_i^-. It should also be kept in mind that the trade-offs are made along the tangent plane. This means that the feasible increments of f_i for all $i \in \bar{I}_1$ will generally be different from those listed in Table 1. In convex cases, for example, the feasible increments will not be larger than those listed in Table 1. A big step size may lead to a larger difference of this type. Therefore, $\bar{\alpha}^{t-1}$ may need to be modified using the technique suggested by Eq. (28), or $\bar{\alpha}^{t-1} = \alpha_2\hat{\alpha}_l^{t-1}$ while α_2 can be regulated with $0 \leqslant \alpha_2 \leqslant 1$.

4.2. Optimality test and interactive algorithm

If σ_i^{t-1} $(i = 1, \ldots, k)$ obtained by Eq. (36) satisfy the following condition

$$\frac{\sigma_1^{t-1}}{N_1^{t-1}} = \cdots = \frac{\sigma_i^{t-1}}{N_i^{t-1}} = \cdots = \frac{\sigma_l^{t-1}}{N_l^{t-1}} = \cdots = \frac{\sigma_k^{t-1}}{N_k^{t-1}}, \tag{39}$$

then from Corollary 2 and definition (31) the best compromise solution is attained at X^{t-1}. On the other hand, this condition can be used to help the DM to articulate his local preferences. Suppose f_l is taken as the reference objective. From Eqs. (36) and (39), we have

$$\Delta f_i^{t-1} = -\Delta f_l^{t-1}\frac{N_l^{t-1}}{N_i^{t-1}}, \quad i = 1, \ldots, k, \ i \neq l. \tag{40}$$

Given a unit change in f_l, or $|\Delta f_i^{t-1}| = 1$, Δf_l^{t-1} defined by Eq. (40) is referred to as an optimal indifference trade-off between f_i and f_l.

Thus, indifference trade-off questions may be put forward using Eq. (40) as follows. Given a unit change $\Delta f_l^{t-1} = 1$ in the reference objective f_l, for example, the DM is asked whether this change can be exactly offset by a change $(-N_l^{t-1}/N_i^{t-1})$ in f_i with other objectives unchanged. If the answers to such questions for all $i = 1, \ldots, k$ are "yes", then condition (39) is satisfied and X^{t-1} is the best compromise solution. If any of the answers is "no", or $\Delta f_i^{t-1} \neq -N_l^{t-1}/N_i^{t-1}$, then the DM must assign another change in f_i, indifferent to $\Delta f_l^{t-1} = 1$. In the latter case, the utility function can be further improved from $u(X^{t-1})$. A new solution with a better utility may then be generated and this process continues until condition (39) is satisfied.

The above decision making process can be summarised by the following interactive decision making algorithm. This interactive algorithm is composed of seven steps and its first two steps are the same as those of the previous iterative algorithm.

Step 3. At X^{t-1}, calculate the normal vector N^{t-1} using Eqs. (8) and (9). Then, use Eq. (40) as heuristics to articulate the DM's indifference trade-offs σ_i^{t-1} for $i = 1, \ldots, k$ Eq. (36).

Step 4. Calculate the projection $d\bar{F}^{t-1}$ of σ^{t-1} using Eq. (31). If condition (39) is satisfied or $d\bar{F}^{t-1} = 0$, the best compromise solution is obtained and stop. Otherwise, if $d\bar{F}^{t-1} \ngeqslant 0$, continue; if $d\bar{F}^{t-1} \geqslant 0$, perturb f_j using Eq. (16) for all

$j \in \{j|\mathrm{d}\bar{f}_j^{t-1} = 0,\ j \in \{1, \ldots, k\}\}$ and then continue.

Step 5. Assign the step size $\hat{\alpha}_t^{t-1}$ using Eq. (38) and Table 1 and let $\alpha_2 = 1$.

Step 6. Let $\overline{\alpha}^{t-1} = \alpha_2 \hat{\alpha}_l^{t-1}$ and define the local region using Eqs. (33) and (34). Then, construct and solve the auxiliary problem (35), yielding a new efficient solution X^t.

Step 7. If X^t is preferred to X^{t-1} (thus $u(X^t) \geqslant u(X^{t-1})$), let $t = t + 1$ and go to Step 3. Otherwise, let $\alpha_2 = \alpha_2/2$ and go to Step 6.

Note that the above interactive process is not necessarily irreversible as the optimality condition could be tested separately at each generated efficient solution, independent of other solutions generated before. In other words, the process does not necessarily require that the utility function always increase as the process proceeds. Rather, the process can be terminated at any individual point by assessing the optimal indifference trade-offs defined by Eq. (40). Such flexibility preserves the favourable features of ISTM and leads to a progressive and explicit articulation of DM priorities. However, assuming that the utility function increases monotonically in the interactive process guarantees the convergence of the process.

5. An illustrative example

5.1. Problem description and analytical solution

In this section, a small example with two objectives and two variables is analysed. This problem is assumed to have a quadratic overall utility function so that its objective space can be illustrated explicitly. This allows us to demonstrate the basic principles of the new approach analytically, iteratively, graphically and interactively. The problem is defined as follows

$$\begin{aligned} \max \quad & F(X) = \{f_1(X) = 5x_1 - 2x_2, \\ & f_2(X) f_2(X) = -x_1 + 4x_2\} \\ \text{s.t.} \quad & X \in \Omega, \quad X = [x_1\ x_2]^{\mathrm{T}}, \\ & \Omega = \left\{ X \,\middle|\, \begin{array}{l} g_1(X) = -x_1 + x_2 - 3 \leqslant 0, \\ g_2(X) = x_1 + x_2 - 8 \leqslant 0 \\ g_3(X) = x_1 - 6 \leqslant 0, \\ g_4(X) = x_2 - 4 \leqslant 0,\ x_1, x_2 \geqslant 0 \end{array} \right\}. \end{aligned} \tag{41}$$

The feasible objective space of the problem is shown in Fig. 3 as enclosed by lines $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, $\overline{DE}$, $\overline{EF}$ and $\overline{FA}$. The single objective optimal

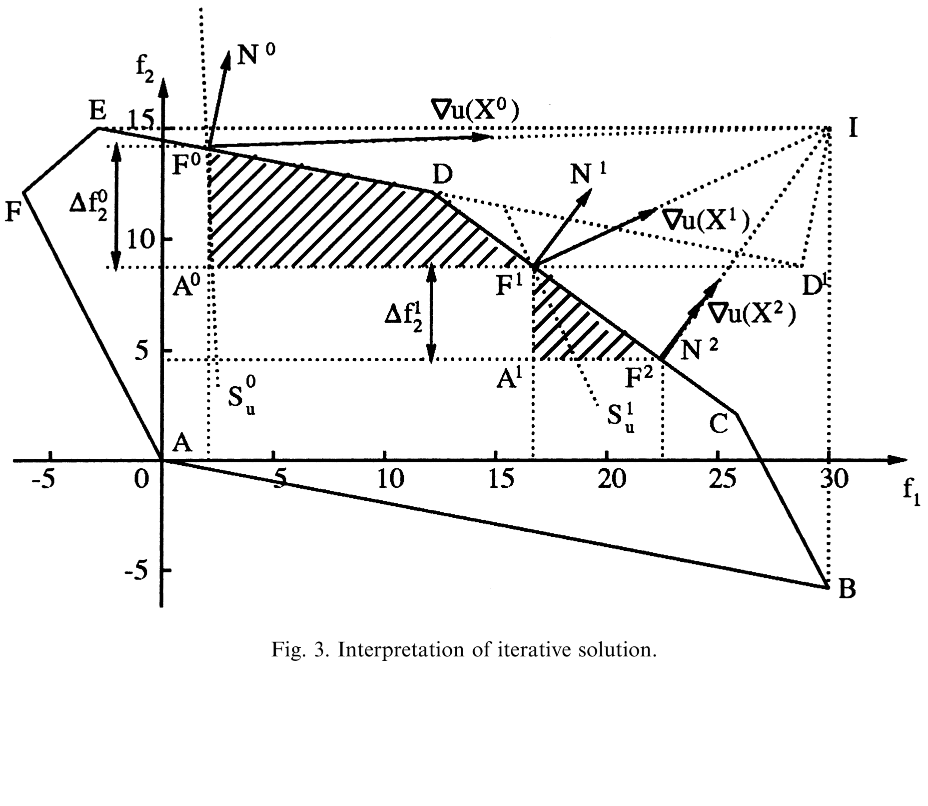

Fig. 3. Interpretation of iterative solution.

solutions, denoted by $\hat{X}^1$ and $\hat{X}^2$ and obtained by maximising objectives $f_1(X)$ and $f_2(X)$ in Ω, respectively, are obtained as follows:

$$\begin{aligned}\hat{X}^1 &= \left[\hat{x}_1^1\ \hat{x}_2^1\right]^{\mathrm{T}} = [6\ 0]^{\mathrm{T}},\\ \hat{F}^1 &= \left[\hat{f}_1^1\ \hat{f}_2^1\right]^{\mathrm{T}} = [30\ \ -6]^{\mathrm{T}},\\ \hat{X}^2 &= \left[\hat{x}_1^2\ \hat{x}_2^2\right]^{\mathrm{T}} = [1\ 4]^{\mathrm{T}},\\ \hat{F}^2 &= \left[\hat{f}_1^2\ \hat{f}_2^2\right]^{\mathrm{T}} = [-3\ 15]^{\mathrm{T}}.\end{aligned} \tag{42}$$

An overall utility function is assumed to be a quadratic function defined as follows

$$u(X) = 1800 - \left[(30 - f_1(X))^2 + (15 - f_2(X))^2\right]. \tag{43}$$

In the objective space, the contour of the above utility function for a given utility value is a circle with point (30 15) being its centre. The advantage of assuming this simple utility function is that all its gradients point to the centre. At points $\hat{X}^1$ and $\hat{X}^2$, we have $u(\hat{X}^1) = 1359$ and $u(\hat{X}^2) = 711$.

It is clear from Fig. 3 that the efficient frontier of the problem, denoted by Ω_E, is composed of line segments $\overline{ED}$, $\overline{DC}$ and $\overline{CB}$, or

$$\Omega_E = \left\{ (f_1\ f_2) \left| \begin{array}{l} (\overline{ED})\colon f_1 + 5f_2 = 72,\\ (\overline{DC})\colon f_1 + 1.4f_2 = 28.8\\ (\overline{CB})\colon f_1 + 0.5f_2 = 27,\\ (f_1\ f_2) \in F(\Omega) \end{array} \right. \right\}. \tag{44}$$

Problem (41) together with the utility function $u(X)$ given by Eq. (43) can be rewritten as the following quadratic programming problem

$$\begin{aligned} \max \quad & u(X) = 180 - (30 - 5x_1 + 2x_2)^2 - (15 \\ & \qquad + x_1 - 4x_2)^2 \\ \text{s.t.} \quad & X \in \Omega. \end{aligned} \tag{45}$$

The optimal solution of problem (45) can be readily generated by

$$\begin{aligned} X^* &= \left[x_1^*\ x_2^*\right]^{\mathrm{T}} = [5.5\ 2.5]^{\mathrm{T}},\\ F^* &= \left[f_1^*\ f_2^*\right]^{\mathrm{T}} = [22.5\ 4.5]^{\mathrm{T}}, \qquad u(X^*) = 1633.5.\end{aligned}$$

In Fig. 3, F^* is denoted by point F^2 which is a feasible solution closest to point I, the centre of the utility contours. Thus, $u(X^*)$ is the maximum feasible utility. At F^* the utility contour is tangent to the efficient frontier (line $\overline{DC}$).

5.2. Iterative optimisation

For many MODM problems with nonseparable utility functions, solving problem (3) directly may become extremely inefficient while it is much easier to deal with problem (20). This is the case especially when problem (1) is a large MODM problem with several sub-objectives and a nonseparable overall objective [12] or a nonlinear dynamic MODM problem [13]. In this small MODM example, we used a simple separable utility function. The main purpose of this study is to demonstrate the iterative process of the new approach developed in the previous sections. In the next section, a more complex decision making problem is simulated using the new approach.

In Appendix A, the detailed calculations for the first iteration is demonstrated. The iterative optimisation procedure is summarised as in Table 2. From an arbitrarily chosen initial solution

Table 2
Iterative optimisation procedure for the illustrative example

t	X^t	F^t	$u(X^t)$	W^t	$\nabla u(X^t)$	N^t	$\mathrm{d}F^t$	α_1^t	α_2
0	$\begin{bmatrix}2.0\\4.0\end{bmatrix}$	$\begin{bmatrix}2.0\\14.0\end{bmatrix}$	1019.0	$\begin{bmatrix}0.0357\\1.0\end{bmatrix}$	$\begin{bmatrix}56.0\\2.0\end{bmatrix}$	$\begin{bmatrix}0.0303\\0.1515\end{bmatrix}$	$\begin{bmatrix}53.462\\-10.69\end{bmatrix}$	0.5	1.0
1	$\begin{bmatrix}4.669\\3.331\end{bmatrix}$	$\begin{bmatrix}16.683\\8.655\end{bmatrix}$	1582.4	$\begin{bmatrix}0.0751\\0.1576\end{bmatrix}$	$\begin{bmatrix}26.634\\12.692\end{bmatrix}$	$\begin{bmatrix}0.045\\0.063\end{bmatrix}$	$\begin{bmatrix}11.634\\-8.31\end{bmatrix}$	0.5	1.0
2	$\begin{bmatrix}5.5\\2.5\end{bmatrix}$	$\begin{bmatrix}22.5\\4.5\end{bmatrix}$	1633.5	$\begin{bmatrix}0.1333\\0.0952\end{bmatrix}$	$\begin{bmatrix}15.0\\21.0\end{bmatrix}$	$\begin{bmatrix}0.045\\0.063\end{bmatrix}$	$\begin{bmatrix}0.0\\0.0\end{bmatrix}$		

$X^0 = [2\ 4]^T$, the optimal solution is achieved only after two iterations, that is $X^* = X^2 = [5.5\ 2.5]^T$. Different utility functions were also tested for the example problem. The algorithm always converged within a few iterations.

The iterative optimisation process can be illustrated by Fig. 3. In Fig. 3, the starting point and the ideal point are denoted by point F^0 and point I, respectively. The utility function defined by Eq. (43) shows that a solution is better if it is closer to point I in Euclidean distance. The whole iterative optimisation process is then to search for a feasible solution which is nearest point I.

At point F^0, the normal vector is N^0, orthogonal to line $\overline{ED}$ (the tangent plane of the efficient frontier at F^0), and the gradient of utility function is $\nabla u(X^0)$, pointing from point F^0 to point I. Obviously, $\nabla u(X^0)$ is not proportional to N^0. By projecting $\nabla u(X^0)$ onto line $\overline{ED}$, we can reach point D^1 by solving the one-dimension search problem (27). This is because point D^1 is the point on line $\overline{ED}$ which is closest to point I. Although point D^1 is infeasible, this one-dimension search results in the identification of the decrement Δf_2^0 in f_2.

Let $\alpha_2 = 1$. The local region is then constructed as the area enclosed by lines $\overline{F^0D}$, $\overline{DF^1}$, $\overline{F^1A^0}$ and $\overline{A^0F^0}$. Solving the auxiliary problem (20) is thus equivalent to moving in parallel line S_u^0 (the tangent line of u at F^0) to point I as far as possible so that the intersection of S_u^0 with the constructed local region is not empty. This leads to point F^1.

At F^1, $\nabla u(X^1)$ can be projected onto line $\overline{DC}$ as $\nabla u(X^1)$ is not proportional to N^1. Thus, f_1 still needs to be increased at the expense of f_2 in order to improve $u(X)$. The one-dimension search from F^1 along line $\overline{DC}$ results in point F^2, yielding a new decrement ∇f_2^1 in f_2. Given $\alpha_2 = 1$, we can then establish the new local region enclosed by lines $\overline{F^1F^2}$, $\overline{F^2A^1}$ and $\overline{A^1F^1}$. Solving the new auxiliary problem is equivalent to moving in parallel line S_u^1 to point I as far as possible so that the intersection between S_u^1 and the new local region is not empty. This leads to point F^2.

Point F^2 is the optimal solution as the gradient $\nabla u(X^2)$ is proportional to the normal vector N^2 at F^2. In fact, point F^2 is the feasible point closest to point I as the contour of the utility function is tangent to the feasible objective space at F^2.

So far, we haven't touched an irregular efficient solution. Such points may be non-smooth points on the efficient frontier, such as the extreme efficient points in MOLP problems. In Appendix A, an irregular efficient solution is examined.

5.3. Interactive decision analysis

5.3.1. First interaction

To demonstrate the interactive algorithm, we use the same example as defined by Eq. (41) without defining an explicit utility function. Suppose the starting point is also given by

$$X^0 = [2\ 4]^T, \qquad F^0 = [2\ 14]^T, \qquad N^0 = \frac{1}{33}[1\ 15]^T.$$

Suppose f_2 is treated as the reference objective. If the following indifference trade-off is provided: "A unit change in f_2 is exactly offset by a change of 1/20 in f_1 at X^0", or

$$[2\ \ 14]^T \iff [2 - \frac{1}{20}\ \ 14 + 1]^T$$

(" $\iff$ " reads "is indifferent to"),

then the normalised gradient of the utility function at X^0 can be estimated by

$$\sigma^0 = [\sigma_1^0\ \ \sigma_2^0]^T \approx \left[-\frac{1}{-1/20}\ \ 1\right]^T = [20\ \ 1]^T.$$

The projection of σ^0 onto the tangent plane at $F(X^0)$ is then given by

$$\mathrm{d}\bar{F}^0 = \sigma^0 - \frac{[(\sigma^0)^T N^0]}{[(N^0)^T N^0]} N^0 = [19.038\ \ -3.808]^T.$$

Thus, the utility function could be improved from $u(X^0)$ by increasing f_1 at the expense of f_2, or $\bar{I}_1 = \{1\}$ and $\bar{I}_2 = \{2\}$. The amount in which f_2 could be sacrificed is determined as follows. The maximum permissible amount of sacrifice in f_2 and the maximum step size are given by

Table 3
Assignment of step size $\bar{\alpha}^0$ given $C_\alpha = 10$

l	$\hat{\alpha}_l^{t-1}$	f_2	f_1
		$f_2^{t-1} - \hat{\alpha}_l^{t-1}\lvert d\bar{f}_2^{t-1}\rvert$	$f_1^{t-1} + \hat{\alpha}_l^{t-1}\lvert d\bar{f}_1^{t-1}\rvert$
0	0	14	2
1	0.5252	12	12
2	1.0504	10	22
3	1.5756	8	32
4	2.1008	6	42
5	2.6260	4	52
6	3.1512	2	62
7	3.6764	0	72
8	4.2016	−2	82
9	4.7268	−4	92
10	5.2520	−6	102

$$\Delta f_2^0 = f_2^0 - f_2^- = 14 - (-6) = 20,$$
$$\bar{\alpha}^0_{\max} = \frac{\Delta f_2^0}{|d\bar{f}_2^0|} = \frac{20}{3.808} = 5.252.$$

Let $C_\alpha = 10$. $\bar{\alpha}^0$ can then be assigned using Table 3. In Table 3, one can find that the increase of f_1 along the tangent plane at $F(X^0)$ is much quicker than along the efficient frontier. This is because the maximum feasible value of f_1 (i.e. 30) has already been exceeded at $l = 3$ for $C_\alpha = 10$ while f_2 is only reduced to 8. In this case, $\hat{\alpha}_2^0 = 1.0504$ may be used as the step size. If the DM wishes to find a step size such that f_1 is nearer its maximum feasible value 30, C_α could be increased to 100 and Table 4 is thus constructed. In Table 4, $\bar{\alpha}^0 = \hat{\alpha}_{28}^0 = 1.47056$ is taken as the current step size.

Table 4
Assignment of step size $\bar{\alpha}^0$ given $C_\alpha = 100$

l	$\hat{\alpha}_l^{t-1}$	f_2	f_1
		$f_2^{t-1} - \hat{\alpha}_l^{t-1}\lvert d\bar{f}_2^{t-1}\rvert$	$f_1^{t-1} + \hat{\alpha}_l^{t-1}\lvert d\bar{f}_1^{t-1}\rvert$
20	1.05040	10	22
21	1.10292	9.8	23
22	1.15544	9.6	24
23	1.20796	9.4	25
24	1.26048	9.2	26
25	1.31300	9.0	27
26	1.36552	8.8	28
27	1.41804	8.6	29
28	1.47056	8.4	30
29	1.52308	8.2	31
30	1.57560	8.0	32

Given $\bar{\alpha}^0 = \hat{\alpha}_{28}^0$, an auxiliary problem can be constructed as follows

$$\begin{aligned} \max \quad & (\sigma_1^0 y_1 + \sigma_2^0 y_2) \\ \text{s.t.} \quad & X_a \in \Omega_a, \end{aligned}$$

$$\Omega_a = \left\{ X_a \left| \begin{array}{l} f_1(X) \geqslant f_1^0 + y_1, \\ f_2(X) \geqslant f_2^0 - \alpha_2 \hat{\alpha}_{28}^0 |d\bar{f}_2^0| + y_2 \\ X \in \Omega, \quad y_1, y_2 \geqslant 0 \end{array} \right. \right\}.$$

Let $\alpha_2 = 1$. Then, we have

$$\begin{aligned} \max \quad & (20 y_1 + y_2) \\ \text{s.t.} \quad & X_a \in \Omega_a, \end{aligned}$$

$$\Omega_a = \left\{ X_a \left| \begin{array}{ll} f_1(X) \geqslant 2 + y_1, & f_2(X) \geqslant 8.4 + y_2 \\ X \in \Omega, & y_1, y_2 \geqslant 0 \end{array} \right. \right\}.$$

The optimal solution of the above problem is given by

$$X^1 = [x_1^1 \;\; x_2^1]^{\mathrm{T}} = [4.72 \;\; 3.28]^{\mathrm{T}},$$
$$F^1 = [f_1^1 \;\; f_2^1]^{\mathrm{T}} = [17.04 \;\; 8.4]^{\mathrm{T}},$$

which is an efficient solution of problem (41). Suppose the DM still prefers X^1 to X^0 (thus $u(X^1) \geqslant u(X^0)$) although f_1^1 is smaller than expected. In this case, α_2 need not be reduced. If this is not the case, α_2 needs to be reduced and the process is then repeated resulting in another efficient solution.

5.3.2. Second interaction

The normal vector at F^1 can be obtained by $N^1 = 0.063[1/1.4 \;\; 1]^{\mathrm{T}}$. Suppose the DM provides the following indifference trade-off at F^1:

$$[17.04 \;\; 8.4]^{\mathrm{T}} \iff [17.04 - 1 \;\; 8.4 + 1]^{\mathrm{T}}.$$

Then, it is easy to show that the optimal condition is not satisfied. In fact,

$$\sigma^1 = [\sigma_1^1 \;\; \sigma_2^1]^{\mathrm{T}} = [1 \;\; 1]^{\mathrm{T}}, \quad d\bar{F}^1 = 0.135[1.4 \;\; -1]^{\mathrm{T}}.$$

Hence, the utility function can still be improved from $u(X^1)$ by increasing f_1 at the expense of f_2, or $\bar{I}_1 = \{1\}$ and $\bar{I}_2 = \{2\}$. The new maximum permissible amount of sacrifice in f_2 and the new maximum step size are given by

Table 5
Assignment of step size $\bar{\alpha}^1$ given $C_\alpha = 10$

l	$\hat{\alpha}_l^{t-1}$	f_2	f_1
		$f_2^{t-1} - \hat{\alpha}_l^{t-1}\lvert d\bar{f}_2^{t-1}\rvert$	$f_1^{t-1} + \hat{\alpha}_l^{t-1}\lvert d\bar{f}_1^{t-1}\rvert$
0	0	8.40	17.040
1	10.6667	6.96	19.056
2	21.3334	5.52	21.072
3	32.0001	4.08	23.088
4	42.6668	2.64	25.104
5	53.3335	1.20	27.120
6	64.0002	−0.24	29.136
7	74.6669	−1.68	31.152
8	85.3336	−3.12	33.168
9	96.0003	−4.56	35.184
10	106.667	−6.00	37.200

$$\Delta f_2^1 = 14, \qquad \bar{\alpha}^1_{\max} = 106.667.$$

The step size may then be determined using Tables 5 and 6. Table 6 is constructed because the decrease of f_2 will no longer be offset by the increase of f_1 when $l \geqslant 3$ for $C_\alpha = 10$ but f_2 can still be decreased from the value of 5.52. $\hat{\alpha}^1_{27} = 28.8$ is taken as the current step size as 4.5 is regarded as the acceptable lower bound of f_2.

Then, an auxiliary problem can be constructed as follows, given $\alpha_2 = 1$:

$$\begin{aligned} \max \quad & (y_1 + y_2) \\ \text{s.t.} \quad & X_a \in \Omega_a, \end{aligned}$$

Table 6
Assignment of step size $\bar{\alpha}^1$ given $C_\alpha = 100$

l	$\hat{\alpha}_l^{t-1}$	f_2	f_1
		$f_2^{t-1} - \hat{\alpha}_l^{t-1}\lvert d\bar{f}_2^{t-1}\rvert$	$f_1^{t-1} + \hat{\alpha}_l^{t-1}\lvert d\bar{f}_1^{t-1}\rvert$
20	21.33340	5.520	21.072
21	22.40007	5.376	21.274
22	23.46674	5.232	21.475
23	24.53341	5.088	21.677
24	25.60008	4.944	21.878
25	26.66675	4.800	22.107
26	27.73342	4.656	22.282
27	28.80009	4.512	22.483
28	29.86667	4.368	22.685
29	30.93343	4.224	22.886
30	32.00010	4.080	23.088

$$\Omega_a = \left\{ X_a \left| \begin{array}{ll} f_1(X) \geqslant 17.04 + y_1, & f_2(X) \geqslant 4.512 + y_2 \\ X \in \Omega, & y_1, y_2 \geqslant 0 \end{array} \right. \right\}.$$

The optimal solution of the above problem is given by

$$X^2 = [x_1^2 \;\; x_2^2]^{\mathrm{T}} = [5.498 \;\; 2.502]^{\mathrm{T}},$$

$$F^2 = [f_1^2 \;\; f_2^2]^{\mathrm{T}} = [22.283 \;\; 4.512]^{\mathrm{T}}$$

By examining F^2, it is clear that the actual achievement levels of f_1 and f_2 are both the same as expected in Table 6. This is because $d\bar{F}^1$ is on the efficient frontier, as shown by line $\overline{DC}$ in Fig. 3. Thus, α_2 need not be reduced.

5.3.3. Third interaction

The normal vector at F^2 is given by $N^2 = 0.063[1/1.4 \;\; 1]^{\mathrm{T}}$. The DM is expected to provide the following optimal indifference trade-off at F^2

$$[22.483 \;\; 4.512]^{\mathrm{T}} \iff [22.483 - 1.4 \;\; 4.512 + 1]^{\mathrm{T}}.$$

If the DM accepts this trade-off, then the optimal condition is satisfied as

$$\sigma^2 = [\sigma_1^2 \;\; \sigma_2^2]^{\mathrm{T}} = \left[-\frac{1}{-1.4} \;\; 1 \right]^{\mathrm{T}} = \frac{1}{0.063} N^2.$$

Otherwise, he should provide another indifference trade-off and the interactive process will then continue until the optimality condition is satisfied.

The above interactive decision making process can also be interpreted graphically using Fig. 3 as shown before.

6. A simulation study

6.1. Description of the software system and a decision problem

The above simple example was only used for illustration purpose. To demonstrate the potential of this new approach to deal with real-world decision problems. A software subsystem has been developed on the basis of a large multiple criterion decision support system (MC-DSS) which the author has developed since 1991 as a main part of a long term research project funded by the UK Engineering and Physical Sciences Research Council. The MC-DSS currently consists of more than 25,000 lines of C and FORTRAN code and is composed of a general purpose linear/nonlinear optimiser, several MODM methods including goal programming, the ideal point method, Geoffrion's method and the ISTM method, and several other methods for multiple attribute decision making [21,29].

The new software for the proposed approach has been developed using several modules, including numerical differentiation using the central difference method, automatic formulation of Kuhn–Tucker conditions, solution of linear equations defined in equations (9) using the LU decomposition [19], one-dimensional optimisation using the golden section search [19], the gradient projection as given by Eq. (12) and the solution of the auxiliary problem defined as problem (20) using simplex method for linear problems and for general nonlinear problems the specially designed sequential linear programming software with mechanisms for detecting oscillations and regulating step sizes automatically [29].

A decision problem chosen for this simulation study is the modified Bow River Valley water quality management problem, which was modelled as a nonlinear three objective optimisation problem [9,31]. The first objective $f_1(X)$ represents DO level at Robin State Park, $f_2(X)$ the percentage return on equity at Pierce-Cannery and $f_3(X)$ the addition to the tax rate at Bowville. There are three decision variables and they are the treatment levels of waste discharges at the Pierce-Cannery, Bowville, and Plympto, denoted by x_1, x_2 and x_3, respectively. The mathematical formulation of the problem is defined as follows

$$\begin{aligned}
\max \quad & f_1(X) = 2.0 + 0.524(x_1 - 0.3) \\
& \qquad + 2.79(x_2 - 0.3) + 0.882(w_1 - 0.3) \\
& \qquad + 2.65(w_2 - 0.3) \\
\max \quad & f_2(X) = 7.5 - 0.012\left(\frac{59}{1.0 - x_1^2} - 59\right), \\
\min \quad & f_3(X) = 1.8 \times 10^{-3}\left(\frac{532}{1.09 - x_2^2} - 532\right), \\
\text{s.t.} \quad & X \in \Omega,\ X = [x_1, x_2, x_3]^{\mathrm{T}},
\end{aligned}$$

where the decision space Ω is defined by

$$\Omega = \left\{ X \,\middle|\, \begin{array}{l}
g_1(X):\quad 4.75 + 2.27(x_1 - 0.3) \geqslant 6.0, \\
g_2(X):\quad 5.1 + 0.177(x_1 - 0.3) \\
+0.978(x_2 - 0.3) + 0.216(w_1 - 0.3) \\
+0.768(w_2 - 0.3) \geqslant 6.0, \\
g_3(X):\quad 2.50 \times 10^{-3}\left(\frac{450}{1.09 - x_3^2} - 450\right) \leqslant 1.5, \\
g_4(X):\quad 1.0 + 0.0332(x_1 - 0.3) \\
+0.0186(x_2 - 0.3) + 3.34(x_3 - 0.3) \\
+0.0204(w_1 - 0.3) + 0.78(w_2 - 0.3) \\
+2.62(w_3 - 0.3) \geqslant 3.5, \\
g_5(X):\quad x_1 \geqslant 0.3, \qquad g_6(X):\ x_2 \geqslant 0.3, \\
g_7(X): x_3 \geqslant 0.3, \\
g_8(X):\quad x_1 \leqslant 1.0, \qquad g_9(X):\ x_2 \leqslant 1.0, \\
g_{10}(X): x_3 \leqslant 1.0, \\
w_i = \frac{0.39}{1.39 - x_i^2}, \quad i = 1, 2, 3
\end{array} \right\}.$$

By optimising each of the three objective functions within the decision space, we can obtain the three single-objective optimal solutions, denoted by $\hat{X}^1, \hat{X}^2, \hat{X}^3$ maximising $f_1(X)$ and $f_2(X)$ and minimising $f_3(X)$, respectively. The payoff table is listed as in Table 7.

Table 7
Payoff table for the water quality management problem

	$f_1(\hat{X}^i)$	$f_2(\hat{X}^i)$	$f_3(\hat{X}^i)$
$\hat{X}^1$	6.79	0.34	9.68
$\hat{X}^2$	6.35	6.28	9.68
$\hat{X}^3$	4.86	0.34	1.04

Table 8
Iterative optimisation procedure I for the water quality management problem

t	X^t	F^t	$u(X^t)$	W^t	$\nabla u(X^t)$	N^t	$\mathrm{d}F^t$	α_1^t	α_2
0	$\begin{bmatrix}0.9617\\0.9558\\0.8133\end{bmatrix}$	$\begin{bmatrix}6.0253\\3.9215\\4.4687\end{bmatrix}$	82.1	$\begin{bmatrix}1.3039\\0.4248\\0.2917\end{bmatrix}$	$\begin{bmatrix}0.0186\\0.0575\\0.0835\end{bmatrix}$	$\begin{bmatrix}0.6235\\0.0446\\0.1216\end{bmatrix}$	$\begin{bmatrix}-0.0188\\0.0548\\0.0762\end{bmatrix}$	41.05	1.0
1	$\begin{bmatrix}0.8507\\0.9063\\0.8133\end{bmatrix}$	$\begin{bmatrix}5.2542\\6.2756\\2.6071\end{bmatrix}$	95.2	$\begin{bmatrix}0.6502\\\infty\\0.6384\end{bmatrix}$	$\begin{bmatrix}0.0323\\0.0001\\0.0329\end{bmatrix}$	$\begin{bmatrix}0.4770\\0.0000\\0.1701\end{bmatrix}$	$\begin{bmatrix}-0.0068\\0.0001\\0.0190\end{bmatrix}$	47.59	1.0
2	$\begin{bmatrix}0.8507\\0.8650\\0.8133\end{bmatrix}$	$\begin{bmatrix}4.9319\\6.2756\\1.8443\end{bmatrix}$	95.9	$\begin{bmatrix}0.5376\\\infty\\1.2442\end{bmatrix}$	$\begin{bmatrix}0.0388\\0.0001\\0.0168\end{bmatrix}$	$\begin{bmatrix}0.4417\\0.0000\\0.2220\end{bmatrix}$	$\begin{bmatrix}0.0011\\0.0001\\-0.0022\end{bmatrix}$		

6.2. Decision analysis assuming a separable utility function

We first assume that the decision maker wishes to find the right treatment levels of waste discharges so that both the DO level and the percentage return to equity could be as large as their ideal levels of 6.79 and 6.28, respectively, but the addition to the tax rate as small as 1.04. Such preferences could be modelled by a separable utility function defined as follow:

$$u = 100.0 - [(6.79 - f_1(X))^2 + (6.28 - f_2(X))^2 + (f_3(X) - 1.04)^2].$$

Obviously u increases as $f_1(X), f_2(X)$ and $f_3(X)$ approach their ideal values.

Given the above utility function to simulate the decision maker, we could apply the iterative algorithm to find the best compromise solution maximising u. To start the iterative solution, we could generate an ideal feasible solution as X^0 using the ISTM method [28,31] assuming that all the objectives are of equal importance. We thus obtain $X^0 = [0.9617,\ 0.9558,\ 0.8133]^T$. Using the MC-DSS, X^0 can be obtained by using the ISTM method. From X^0, the iterative procedure is initialised and can be summarised as in Table 8.

The above procedure converges very fast to the maximum utility as shown by column 4 of Table 8. The termination condition used was $\mathrm{d}Ft' < [0.01,\ 0.01,\ 0.01]^T$. More iterations would be needed to achieve more precise maximum solution although u could only be improved marginally. $X^2 = [0.8507,\ 0.8650,\ 0.8133]^T$ could thus be used to approximate the best compromise solution. Note that all the three generated solutions X^0, X^1 and X^2 are efficient solutions.

6.3. Decision analysis assuming a nonseparable utility function

To test the algorithm for more complex preference information, a nonseparable utility function was formulated. It was assumed that the decision maker changed the ideal level of the percentage return on equity ($f_2(X)$) from 6.28 to 6.0. It was also assumed that the decision maker would be satisfied with an efficient solution at which any two of the three objectives could be as close to their ideal levels as possible. A function incorporating the above preferences could be defined as follows:

$$u = 100.0 - [(6.79 - f_1(X))^2(6.0 - f_2(X))^2 + (6.79 - f_1(X))^2(f_3(X) - 1.04)^2 + (6.0 - f_2(X))^2(f_3(X) - 1.04)^2],$$

which is a nonseparable function.

From the initial solution X^0, we search for the best compromise solution that could best satisfy the decision maker's preferences. The iterative calculation procedure is as shown in Table 9.

Note that the above "utility" function does not increase with $f_2(X)$ when $f_2(X)$ is above the level

Table 9
Iterative optimisation procedure II for the water quality management problem

t	X^t	F^t	$u(X^t)$	W^t	$\nabla u(X^t)$	N^t	$\mathrm{d}F^t$	α_1^t	α_2
0	$\begin{bmatrix}0.9617\\0.9558\\0.8133\end{bmatrix}$	$\begin{bmatrix}6.0253\\3.9215\\4.4687\end{bmatrix}$	39.81	$\begin{bmatrix}1.3039\\0.4248\\0.2917\end{bmatrix}$	$\begin{bmatrix}0.6176\\1.2886\\0.8448\end{bmatrix}$	$\begin{bmatrix}0.6235\\0.0446\\0.1216\end{bmatrix}$	$\begin{bmatrix}-0.2208\\1.2286\\0.6814\end{bmatrix}$	1.774	1.0
1	$\begin{bmatrix}0.8507\\0.9459\\0.8133\end{bmatrix}$	$\begin{bmatrix}5.6337\\6.2756\\3.9452\end{bmatrix}$	87.97	$\begin{bmatrix}0.8632\\\infty\\0.3443\end{bmatrix}$	$\begin{bmatrix}0.2239\\-0.0613\\0.0933\end{bmatrix}$	$\begin{bmatrix}0.5509\\0.0000\\0.1246\end{bmatrix}$	$\begin{bmatrix}-0.0092\\-0.0613\\0.0406\end{bmatrix}$	7.21	1.0
2	$\begin{bmatrix}0.8905\\0.9317\\0.8133\end{bmatrix}$	$\begin{bmatrix}5.5680\\5.8245\\3.3553\end{bmatrix}$	91.78	$\begin{bmatrix}0.8169\\2.2170\\0.4320\end{bmatrix}$	$\begin{bmatrix}0.1436\\0.0262\\0.0769\end{bmatrix}$	$\begin{bmatrix}0.5197\\0.0815\\0.1413\end{bmatrix}$	$\begin{bmatrix}-0.0099\\0.0021\\0.0352\end{bmatrix}$	65.9	1.0
3	$\begin{bmatrix}0.8809\\0.8544\\0.8133\end{bmatrix}$	$\begin{bmatrix}4.9176\\5.9533\\1.7022\end{bmatrix}$	98.45	$\begin{bmatrix}0.5334\\3.1031\\1.5114\end{bmatrix}$	$\begin{bmatrix}0.0168\\0.0037\\0.0472\end{bmatrix}$	$\begin{bmatrix}0.4372\\0.0736\\0.2369\end{bmatrix}$	$\begin{bmatrix}-0.0157\\-0.0017\\0.0296\end{bmatrix}$	22.16	1.0
4	$\begin{bmatrix}0.8839\\0.8340\\0.8133\end{bmatrix}$	$\begin{bmatrix}4.7892\\5.9149\\1.4697\end{bmatrix}$	99.23	$\begin{bmatrix}0.4993\\2.7730\\2.3305\end{bmatrix}$	$\begin{bmatrix}0.0077\\0.0072\\0.0347\end{bmatrix}$	$\begin{bmatrix}0.4295\\0.0708\\0.2662\end{bmatrix}$	$\begin{bmatrix}-0.0138\\0.0036\\0.0214\end{bmatrix}$		

of 6.0. The construction of this special "utility" function is to avoid the solutions converging to an extreme efficient solution as is the case in Table 8. We would have the maximum utility of $u = 100$ if any two of the objectives could achieve their exact ideal levels. From the payoff table, we know that this is not possible.

The above procedure quickly converges to the maximum utility as shown by column 4 of Table 9. The termination condition was $\mathrm{d}Ft' < [0.025, 0.025, 0.025]^{\mathrm{T}}$. X^1, X^2, X^3 and X^4 of Table 9 are all efficient solutions. The utility at $X^4 = [0.8839, 0.8340, 0.8133]^{\mathrm{T}}$ is approximately the maximum and X^4 can thus be suggested as the best compromise solution.

7. Concluding remarks

The new gradient-based local region search method proposed in this paper is developed to extend the ISTM method [28] and its enhanced version [31], so that a general nonlinear (nonconvex) multiobjective optimisation problem with a nonseparable explicit or implicit utility function could be dealt with [20,21,32]. This method is based on the identification of normal vectors at efficient solutions. It provides an iterative or interactive manner to generate the best compromise solution with its efficiency and optimality in terms of utility being guaranteed.

Given nonseparable utility functions explicitly, the proposed iterative algorithm provides an alternative way to solve multiobjective optimisation problems, which may be desirable in certain decision situations where the nonseparability of a utility function causes serious problems in terms of computation efficiency [12,13]. The optimality conditions for this iterative process are established and the convergence of the process is proven. A simple numerical example is presented to illustrate the implementation of this algorithm and a decision problem is examined to demonstrate the potential applications of the new method.

In cases where utility functions can only be known implicitly, the proposed interactive decision making algorithm facilitates an explicit trade-off analysis using the DM's local preference information. This trade-off analysis is conducted in the objective space. The proposed optimality conditions can therefore be used as a guideline to help assess marginal rates of substitution, rather than as a trick to terminate the interactive process when the DM does not consciously know why this happens and therefore may not be able to control the interactive process properly. Furthermore, the ISTM type of trade-off analysis can be implemented in an explicit manner to facilitate the estimation of step sizes as the DM clearly knows the current achievement levels of the objectives and the possible consequences caused by his trade-offs. The above features may be favourable as they lead to a progressive and explicit exploration of DM priorities. On the whole, the new interactive algorithm may basically be viewed as a search-oriented process while the DM can also learn how his preferences should be articulated to attain desirable trade-offs. The same small example problem is used to illustrate the interactive decision making algorithm.

8. For further readings

[3,7,10,14,25]

Acknowledgements

The author is in debt to the two anonymous referees for their constructive and valuable comments. He is also grateful to the editor for passionately encouraging him to submit the revised paper.

Appendix A. Some detailed results for the illustrative example

A.1. The calculation procedure for the first iteration

To initiate the iterative optimisation process, an initial solution is required. An ISTM type of trade-off analysis can be conducted to generate an initial solution. Suppose from $\hat{X}^2$ $f_1(X)$ is increased by sacrificing one unit of $f_2(X)$, or $\Delta f_2 = 1$. Let $h_1 = \hat{f}_1^1 - \hat{f}_1^2 = 33$. Then the following auxiliary problem can be constructed [28]

$$\begin{aligned} \max \quad & y_1/33 \\ \text{s.t.} \quad & X \in \Omega, \end{aligned} \tag{A.1}$$

$$f_1(X) - y_1 \geqslant -3, \; f_2(X) \geqslant 14.$$

The solution of problem Eq. (A.1), denoted by X^0, is given by $X^0 = [x_1^0 \; x_2^0]^{\mathrm{T}} = [2 \; 4]^{\mathrm{T}}$, $F^0 = [f_1^0 \; f_2^0]^{\mathrm{T}} = [2 \; 14]^{\mathrm{T}}$ and $u(X^0) = 1019$. It is easy to show that X^0 is an efficient solution. Let

$$\omega_1^0 = \frac{1}{\hat{f}_1^1 - f_1^0} = \frac{1}{28}, \qquad \omega_2^0 = \frac{1}{\hat{f}_2^2 - f_2^0} = 1. \tag{A.2}$$

At X^0, only the 4th constraint is saturated, or $g_4(X^0) = 0$. Then, the Kuhn–Tucker multipliers λ_1, λ_2 and μ_4 are obtained by solving the following linear equations (Theorem 3)

$$\begin{aligned} &\lambda_1 + \lambda_2 = 1, \\ &-\frac{1}{28} \times 5 \times \lambda_1 - 1 \times (-1) \times \lambda_2 + 0 \times \mu_4 = 0, \\ &-\frac{1}{28} \times (-2) \times \lambda_1 - 1 \times 4 \times \lambda_2 + 1 \times \mu_4 = 0. \end{aligned} \tag{A.3}$$

Solving Eq. (A.3), we have $\lambda_1 = 28/33$ and $\lambda_2 = 5/33$. Thus, the normal vector at $F(X^0)$ is given by

$$N^0 = [\omega_1^0 \lambda_1^0 \; \omega_2^0 \lambda_2^0]^{\mathrm{T}} = \frac{1}{33}[1 \;\; 5]^{\mathrm{T}}.$$

The gradient of the utility function and its projection onto the tangent plane of the efficient frontier at $F(X^0)$ can be calculated as follows:

$$\nabla u(X^0) = \left[\frac{\partial u(X^0)}{\partial f_1} \;\; \frac{\partial u(X^0)}{\partial f_2} \right]^{\mathrm{T}} = 2[28 \;\; 1]^{\mathrm{T}},$$

$$\begin{aligned} \mathrm{d}F^0 &= \nabla u(X^0) - \frac{[(\nabla u(X^0))^{\mathrm{T}} N^0]}{[(N^0)^{\mathrm{T}} N^0]} N^0 \\ &= [53.462 \;\; -10.69]^{\mathrm{T}}. \end{aligned}$$

Along $\mathrm{d}F^0$, the maximum step size α_1^0 can be found by solving the following one dimension search problem:

$$\max u(\alpha_1) = 1800 - [(30 - (f_1^0 + \alpha_1 \mathrm{d}f_1^0))^2 + (15 - (f_2^0 + \alpha_1 \mathrm{d}f_2^0))^2].$$

This gives $\alpha_1^0 = 0.5$. Let $\alpha_2 = 1$. An auxiliary problem can then be constructed as follows:

$$\begin{aligned} \max \quad & \left(\frac{\partial u(X^0)}{\partial f_1} y_1 + \frac{\partial u(X^0)}{\partial f_2} y_2\right) \\ \text{s.t.} \quad & X_a \in \Omega_a, \ X_a = [X^{\mathrm{T}} \ y_1 \ y_2]^{\mathrm{T}}, \end{aligned}$$

$$\Omega_a = \left\{ X_a \left| \begin{array}{l} f_1(X) \geqslant f_1^0 + y_1, \\ f_2(X) \geqslant f_2^0 - \alpha_2 \alpha_1^0 |\mathrm{d}f_2^0| + y_2 \\ X \in \Omega, \ y_1, y_2 \geqslant 0 \end{array} \right. \right\},$$

or

$$\begin{aligned} \max \quad & (56 y_1 + 2 y_2) \\ \text{s.t.} \quad & X_a \in \Omega_a, \end{aligned}$$

$$\Omega_a = \left\{ X_a \left| \begin{array}{c} f_1(X) \geqslant 2 + y_1, \\ f_2(X) \geqslant 8.655 + y_2 \\ X \in \Omega, \ y_1, y_2 \geqslant 0 \end{array} \right. \right\}.$$

The optimal solution of the above problem is given by

$$\begin{aligned} X^1 &= [x_1^1 \ x_2^1]^{\mathrm{T}} \\ &= [4.669 \ 3.331]^{\mathrm{T}}, \qquad F^1 = [f_1^1 \ f_2^1]^{\mathrm{T}} \\ &= [16.683 \ 8.655]^{\mathrm{T}}, \qquad u(X^1) = 1582.399. \end{aligned}$$

A.2. Irregular points and multiple normal vectors

To examine an irregular efficient solution, let's take for example point D (see Fig. A1) which is given by

$$X^D = [4 \ 4]^{\mathrm{T}},$$
$$F^D = [12 \ 12]^{\mathrm{T}}, \qquad u(X^D) = 1467.$$

Let

$$\omega_1^D = \frac{1}{\hat{f}_1^1 - f_1^D} = \frac{1}{18}, \qquad \omega_2^D = \frac{1}{\hat{f}_2^2 - f_2^D} = \frac{1}{3}.$$

At X^D, both the second and fourth constraints are saturated, or $g_2(X^D) = 0$ and $g_4(X^D) = 0$. Thus, we can obtain the following linear equations from Theorem 3:

$$\lambda_1 + \lambda_2 = 1, \qquad -\frac{5}{8}\lambda_1 + \frac{1}{3}\lambda_2 + \mu_2 = 0,$$
$$\frac{2}{18}\lambda_1 - \frac{4}{3}\lambda_2 + \mu_2 + \mu_4 = 0.$$

Note that $\lambda_1, \lambda_2, \ \mu_2, \ \mu_4 \geqslant 0$.

The above equations have multiple solutions and thus X^D is an irregular efficient solution. This means that there are multiple normal vectors at X^D. In this case, condition (11) is tested by examining all normal vectors at this point. Let's take μ_4 as a free parameter. Then, the family of all possible normal vectors at X^D can be obtained as follows:

$$N^D = \frac{1}{37}\left[\frac{1}{18}(30 - 18\mu_4) \ \ \frac{1}{3}(7 + 18\mu_4)\right]^{\mathrm{T}},$$
$$0 \leqslant \mu_4 \leqslant \frac{108}{198}. \tag{A.4}$$

X^D would be an optimal solution if there existed a normal vector so that condition (11) could be satisfied at X^D, or $\nabla u(X^D) = cN^D$. As $\nabla u(X^D) = 2[18 \ 3]^{\mathrm{T}}$, however, we would have

$$2[18 \ 3]^{\mathrm{T}} = c\frac{1}{37}\left[\frac{1}{18}(30 - 18\mu_4) \ \ \frac{1}{3}(7 + 18\mu_4)\right]^{\mathrm{T}}.$$

The solution of the above equations is $\mu_4 = -0.33 < 0$ which is not permitted as μ_4 must not be negative. Therefore, $\nabla u(X^D)$ is not proportional to any normal vector, or X^D is not optimal.

Hence, the utility function can still be improved from X^D. To obtain an ascent direction, $\nabla u(X^D)$ can in principle be projected onto any permissible tangent planes. Given $\mu_4 = 0$, for example, we have

$$N^{D0} = 0.0451[1 \ 1.4]^{\mathrm{T}},$$
$$\nabla u^{D0} = [36 \ 6]^{\mathrm{T}},$$
$$\mathrm{d}F^{D0} = [21 \ -15]^{\mathrm{T}}, \qquad \alpha^{D0} = 0.5.$$

The auxiliary problem at X^D can then be constructed as follows:

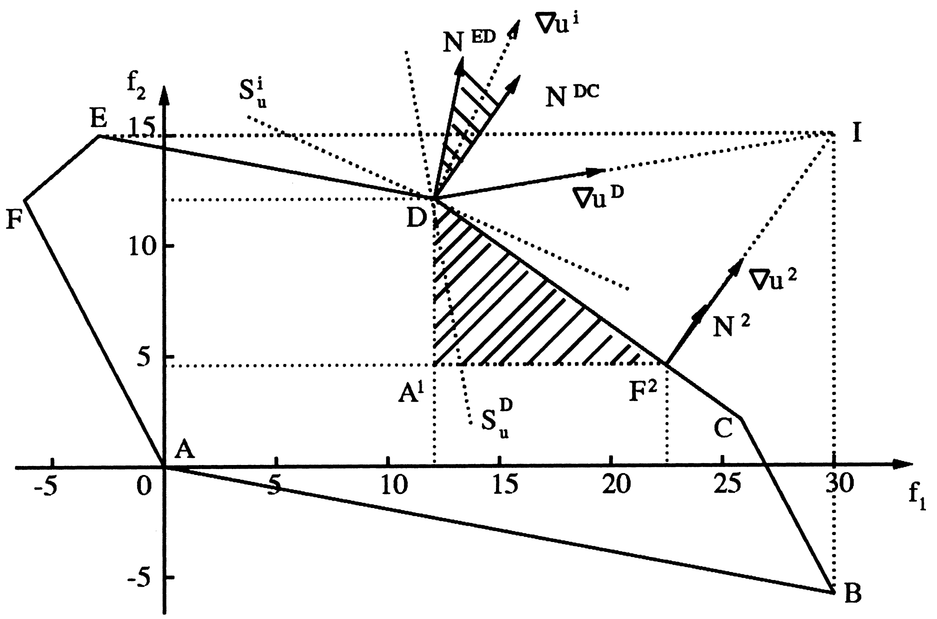

Fig. 4. Irregular points and multiple normal vectors

$$\max \quad (36y_1 + 6y_2)$$
$$\text{s.t.} \quad X_a \in \Omega_a,$$

$$\Omega_a = \left\{ X_a \middle| \begin{array}{l} f_1(X) \geqslant 12 + y_1, \quad f_2(X) \geqslant 4.5 + y_2 \\ X \in \Omega, \ y_1, y_2 \geqslant 0 \end{array} \right\}.$$

The optimal solution of the above problem is given by

$$X^* = [5.5 \ \ 2.5]^{\mathrm{T}},$$
$$F^* = [22.5 \ \ 4.5]^{\mathrm{T}}, \qquad u(X^*) = 1633.5,$$

which is exactly the best compromise solution of the original problem.

At point D, as shown in Fig. 4, there are multiple (infinite number of) normal vectors. N^{DC} and N^{ED} are two normal vectors at this point, orthogonal to line $\overline{DC}$ and line $\overline{ED}$ respectively, which can be obtained by setting $\mu_4 = 0$ and 108/198 in Eq. (A.4). Any vector emerging from point D and between vectors N^{DC} and N^{ED} could be viewed as a normal vector at point D. ∇u^D is the gradient of u at point D. As ∇u^D is not proportional to any normal vector, D is not an optimal solution. Projecting ∇u^D onto line $\overline{DC}$ will then result in point F^2 in one step.

Finally, if the utility gradient at point D were between normal vectors N^{ED} and N^{DC}, such as ∇u^i, then the projection of ∇u^i onto any tangent plane except that one orthogonal to ∇u^i would not be zero. However, solving any auxiliary problem constructed as a result of such projection will lead to the same solution as denoted by point D, whatever local regions may be constructed within the feasible space. This is because an infinitesimal move of the tangent plane S_u^i of the utility contour in the direction of ∇u^i will lead out of the feasible space immediately.

Thus, instead of identifying all normal vectors at an irregular solution as given by Eq. (A.4), we could choose any one of the normal vectors to construct an auxiliary problem. If the solution of such a problem is the same as the irregular solution, it is an optimal solution of the original problem. Otherwise, a new and better solution is found and the iterative process continues. It should be noted, however, that the above argument about the optimality test for irregular solutions may only be valid for testing global optimality in non-convex cases, where condition (11) is not sufficient for global optimality. In such cases, it is suggested to project ∇u onto two or more tangent planes to test the optimality of an efficient solution or to move away from the solution.

References

[1] J. Buchanan, A two-phase interactive solution method for multiple-objective programming problems, IEEE Transactions on Systems, Man, and Cybernetics 21 (4) (1991) 743–749.

[2] V. Chankong, Y.Y. Haimes, Multiobjective Decision Making: Theory and Methodology, North-Holland, New York, 1991.

[3] A. Charnes, W.W. Cooper, Goal programming and multiple objective optimisation – Part I, European Journal of Operational Research 1 (1977) 39–54.

[4] J.S. Dyer, Interactive goal programming, Management Science 19 (1) (1972) 62–70.

[5] J.S. Dyer, P.C. Fishburn, R.E. Steuer, J. Wallenius, S. Zionts, Multiple criteria decision making, multiattribute utility theory: The next ten years, Management Science 38 (1992) 645–654.

[6] P.A.V. Ferreira, J.C. Geromel, An interactive projection method for multicriteria optimisation problems, IEEE Transactions on Systems, Man, and Cybernetics 20 (3) (1990) 596–605.

[7] P.C. Fishburn, Nonlinear Preference and Utility Theory, AT & T Bell Laboratories, 1988.

[8] A.M. Geoffrion, J.S. Dyer, A. Feinberg, An interactive approach for multicriteria optimisation, with an application to the operation of an academic department, Management Science 19 (4) (1972) 357–368.

[9] C.L. Huang, A.S. Masnd, Multiple Objective Decision Making Methods and Applications, A State-of-Art Survey, Springer, Berlin, 1979.

[10] E. Jacquet-Lagreze, J. Siskos, Assessing a set of additive utility functions for multicriteria decision making, the UTA method, European Journal of Operational Research 10 (1982) 151–164.

[11] R.L. Keeney, H. Raiffa, Decisions with Multiple Objectives: Preferences and Value Trade-Offs, Wiley, New York, 1976.

[12] D. Li, A decomposition method for optimisation of large-system reliability, IEEE Transactions on Reliability 35 (1990) 183–188.

[13] D. Li, On general multiple linear-quadratic control problems, IEEE Transactions on Automatic Control 38 (1993) 1722–1727.

[14] D. Li, Convexification of noninferior frontier, Journal of Optimisation Theory and Applications 88 (1) (1996) 177–196.

[15] D. Li, J.B. Yang, Iterative parametric minimax method for a class of composite optimisation problems, Journal of Mathematical Analysis and Applications 198 (1996) 64–83.

[16] M.R. Lightner, S.W. Director, Multiple criterion optimisation for the design of electronic circuits, IEEE Transactions on Circuits and Systems 28 (3) (1981) 169–179.

[17] D.G. Luenberger, Introduction to Linear and Nonlinear Programming, Addison–Wesley, Reading, MA, 1973.

[18] M. Minoux, Mathematical Programming – Theory and Algorithms, Wiley, Chichester, 1986.

[19] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B. Flannery, Numerical Recipes in C: The Art of Scientific Computing, Cambridge University Press, 1992.

[20] P. Sen, J.B. Yang, P. Meldrum, Interactive trade-off analysis in multiple criteria preliminary design of a semi-submersible, in: Proceedings of the Sixth International Conference on Marine Engineering Systems, University of Hamburg, Germany, 1993.

[21] P. Sen, J.B. Yang, Multiple criteria decision making in design selection and synthesis, Journal of Engineering Design 6 (3) (1995) 207–230.

[22] W.S. Shin, A. Ravindran, A comparative study of interactive tradeoff cutting plane methods for MOMP, European Journal of operational Research 56 (1992) 380–393.

[23] R.E. Steuer, E.U. Choo, An interactive weighted Tchebycheff procedure for multiple objective programming, Mathematical Programming 26 (1983) 326–344.

[24] T.J. Stewart, An interactive multiple objective linear programming method based on piecewise linear additive value functions, IEEE Transactions on Systems, Man, and Cybernetics 17 (1987) 799–805.

[25] T.J. Stewart, A critical survey on the status of multiple criteria decision making theory and practice, OMEGA International Journal of Management Science 20 (5/6) (1992) 569–586.

[26] D. Vanderpooten, P. Vincke, Description and analysis of some representative interactive multicriteria procedures, Mathematical and Computer Modelling 12 (1989) 1221–1238.

[27] J.B. Yang, C. Chen, Z.J. Zhang, The interactive decomposition method for multiobjective linear programming and its applications, Information and Decision Technologies 14 (3) (1988) 275–288.

[28] J.B. Yang, C. Chen, Z.J. Zhang, The interactive step trade-off method (ISTM) for multiobjective optimisation, IEEE Transactions on Systems, Man, and Cybernetics 20 (3) (1990) 688–695.

[29] J.B. Yang, A decision support system for multiple criteria decision making: The C and FORTRAN source code and application examples, Research Report, Engineering Design Centre, University of Newcastle upon Tyne, UK, 1995, 325 p.

[30] J.B. Yang, P. Sen, An interactive gradient projection approach for multiobjective optimisation, in Proceedings of IFAC/IFORS/IMACS Symposium on Large Scale Systems: Theory and Applications, London, UK, 11–13 July, 1995, pp. 463–468.

[31] J.B. Yang, P. Sen, Preference modelling by estimating local utility functions for multiobjective optimisation, European Journal of Operational Research 95 (1996) 115–138.

[32] J.B. Yang, P. Sen, Interactive tradeoff analysis and preference modelling for preliminary multiobjective ship design, Systems Analysis, Modelling, and Simulation 26 (1996) 25–55.

[33] J.B. Yang, D. Li, Optimisation scheme using iterative parametric minimax method, submitted to Operations Research, 1998.