A reference point-based evolutionary algorithm for many-objective fuzzy portfolio selection*

Jian Chen[1], Xiaoliang Ma[1], Yiwen Sun[2], and Zexuan Zhu[1]**

[1] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China; Guangdong Laboratory of Artificial Intelligence and Digital Economy(SZ), Shenzhen University, Shenzhen 518060, China

[2] School of Medicine, Shenzhen University, Shenzhen 518060, China

zhuzx@szu.edu.cn

Abstract. Portfolio selection is an important problem in the practice and theory of finance. This paper builds a five-objective (including mean, variance, skewness, kurtosis, and entropy) model to replace the classical Markowitz mean-variance model for finding better portfolio selection. To obtain a more accurate estimation of risk asset returns, a fuzzy number variable, instead of a random variable, based on the acknowledge of experts is used to estimate the return of a risk asset. A new reference point-based evolutionary algorithm (NRPEA) is proposed to obtain well-convergence and well-distributed solutions for the many-objective optimization problems. In NRPEA, the auxiliary reference points are generated and selected to guide the population evolution. Experiment results on six well-known data sets demonstrate the effectiveness and efficiency of NRPEA in the comparison with other three state-of-the-art many-objective optimization algorithms.

Keywords: Fuzzy portfolio selection · Many-objective optimization · High-order moment · Reference point.

1 Introduction

Portfolio selection problem is one hot topic in economic theory research. Modern portfolio selection problem stems from the Markowitz mean-variance model [1],

---

* This work was supported in part by the National Natural Science Foundation of China, under Grants 61976143, 61471246, 61603259, and 61871272, Guangdong Special Support Program of Top-notch Young Professionals, under Grant 2014TQ01X273, Shenzhen Fundamental Research Program, under Grant JCYJ20170302154328155, Scientific Research Foundation of Shenzhen University for Newly-introduced Teachers, under Grant 2019048, and Zhejiang Lab's International Talent Fund for Young Professionals. This work was supported by the National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen 518060, China and Guangdong Laboratory of Artificial Intelligence and Digital Economy(SZ), Shenzhen University, Shenzhen 518060, China.

** Corresponding author.

which is a bi-objective optimization problem via maximizing the return and minimizing the risk simultaneously. The mean-variance portfolio model assumes an underlying normal probability distribution or quadratic utility function on the asset return [2–5]. However, many studies show that returns on assets tend to have a heavy-tailed and asymmetric leptokurtic distribution [2], i.e., not a normal distribution in statistics [2]. This implies that the higher-order moments should be considered. Recently, 3-order moment or 4-order moment of asset return have been considered in [3–5] and shown to be helpful for portfolio selection problem. Shannon's entropy can also be used to measure the dispersion degree of assets for effective risk diversification [6, 7].

Besides high-order moments, uncertainty is another key factor in portfolio selection problem. The majority of the current portfolio selection problems over the decades are based on probability theory. However, the uncertain returns of risk assets have many influence factors, such as politics, social conditions, economic change, and the status of the related company. Compared with a random variable, a fuzzy number variable based on the acknowledge of experts is more suitable for estimating the return of a risk asset [4]. Thus, a few researchers have extended the probabilistic portfolio model to the fuzzy environment [4–6].

To find a better portfolio selection, this paper constructs a five-objective fuzzy portfolio selection model. The portfolio selection problem is formulated as a five-objective optimization problem. On top of the traditional 1-order and 2-order moments (mean and variance), the proposed model also considers objectives including 3-order and 4-order moments (skewness and kurtosis) and the mean proportional entropy. The mean proportional entropy based on the sum of Minkowski distance between weights of the invested assets is used to acquire a well-diversified portfolio. The fuzzy variable based on the acknowledge of experts is introduced to handle the uncertain return of risk asset in the proposed portfolio selection problem. Pareto dominance-based multi-objective evolutionary algorithms (MOEAs) have been widely used to solve multi-objective optimization problems. However, traditional MOEAs might have difficulty in dealing with many-objective optimization problems (with number of objectives greater than three) [8–14]. This paper introduces a new reference point-based evolutionary algorithm (NRPEA) to deal with the proposed five-objective portfolio selection problem. The auxiliary reference points in NRPEA are generated and selected to guide the population evolution for five-objective optimization. Experimental results on real-world data sets show that the proposed NRPEA can obtain superior or comparable performance to the other state-of-the-art many-objective evolutionary algorithms.

In the rest of this paper, the related background acknowledge is introduced in Section II. Section III presents the proposed five-objective fuzzy portfolio selection problem. The details of the proposed NRPEA are described in Section IV. Section V presents the experimental studies. Finally, Section VII concludes this paper.

2 Background

This section introduces the background of fuzzy theory, portfolio selection model, and many-objective portfolio selection model.

2.1 Fuzzy numbers and weighted possibilistic moments

Definition 1. A fuzzy number $\tilde{A}$ is defined as any fuzzy subset of R, whose membership function $\mu_{\tilde{A}}(x)$ meets the following four conditions: i) $\tilde{A}$ is normal, i.e., $\exists x \in R$, s.t. $\mu_{\tilde{A}}(x) = 1$; ii) $\mu_{\tilde{A}}(x)$ is quasi-concave, i.e., $\mu_{\tilde{A}}(\lambda x + (1 - \lambda)y) \leq \min\{\mu_{\tilde{A}}(x), \mu_{\tilde{A}}(y)\}$ for $\forall \lambda \in [0,1]$; iii) $\mu_{\tilde{A}}(x)$ is upper semicontinuous, i.e., $\{x \in R|\mu_{\tilde{A}}(x) \leq \epsilon\}$ is a closed set for any $\epsilon \in [0,1]$; and iv) the closure of set $\{x \in R|\mu_{\tilde{A}}(x) > 0\}$ is compact.

Definition 2. Given the left width θ, the right width δ and the core $[c,d]$, a fuzzy number $\tilde{A}$ is a trapezoidal fuzzy number if its membership function is of the following form:

$$\mu_{\tilde{A}}(x) = \begin{cases} 1 - \frac{c-x}{\delta}, & \text{if } c - \delta \leq x \leq c \\ 1, & \text{if } c \leq x \leq d \\ 1 - \frac{x-d}{\theta}, & \text{if } d \leq x \leq d + \theta \\ 0, & \text{if otherwise} \end{cases} \tag{1}$$

The trapezoidal fuzzy number $\tilde{A}$ can be denoted by $\tilde{A} = (c, d, \sigma, \theta)$.

Definition 3. γ-level set of the trapezoidal fuzzy number $\tilde{A}$ is defined as $[\underline{a}(\gamma), \overline{a}(\gamma)] = \{x \in R|\mu_{\tilde{A}}(x) \geq \gamma\}$, where $\gamma \in [0,1]$.

Definition 4. Suppose $\tilde{A}$ be a fuzzy number having $[\underline{a}(\gamma), \overline{a}(\gamma)], \gamma \in [0,1]$. The weighted possibilistic moment (WPM) and the weighted possibilistic variance (WPV) of the fuzzy number $\tilde{A}$ are defined as follows:

$$E_f(\widetilde{A}) = \int_0^1 f(\gamma)((\underline{a}(\gamma) + \overline{a}(\gamma))/2)d\gamma, \tag{2}$$

$$Var_f(\widetilde{A}) = \frac{1}{2}\int_0^1 f(\gamma) \cdot \left[(\underline{a}(\gamma) - E)^2 + (\overline{a}(\gamma) - E)^2\right] d\gamma, \tag{3}$$

where $f(\gamma) = (n+1)\gamma^n$ is a probability density function such that $\int_0^1 f(\gamma)d\gamma = 1$. Specifically, WPM (2) is the first-order moment of fuzzy number $\tilde{A}$. *Definition5.* Let $E_l(\tilde{A}), l = 1, ..., L$ be the l-order weighted possibilistic moments of fuzzy number $\tilde{A}$. The weight possibilistic skewness (WPS) and weight possibilistic kurtosis (WPK) are defined as follows

$$Skew_f(\widetilde{A}) = \text{WPS}(\widetilde{A}) = E_3(\widetilde{A})/E_2(\widetilde{A})^{1/3} \tag{4}$$

$$Kur_f(\widetilde{A}) = \text{WPK}(\widetilde{A}) = E_4(\widetilde{A})/\sqrt{E_2(\widetilde{A})} \tag{5}$$

2.2 The classical portfolio selection model

The basic assumptions of Markowitz's portfolio theory include that 1) investors are risk-averse and pursue maximum expected utility, 2) investors choose their portfolios according to the expected return and variance of return, and 3) all investors are in the same investment period. Markowitz proposed to find the effective portfolios that maximize the return and minimizing the risk simultaneously.

The expected return E is used to measure the return of securities, and the variance σ^2 of return is used to represent the investment risk. The Markowitz optimization model is defined as follows:

$$\begin{aligned} &\max E\left(r_p\right)=\sum w_i r_i \\ &\min \sigma^2\left(r_p\right)=\sum\sum w_i w_j \operatorname{cov}\left(r_i, r_j\right), \end{aligned} \tag{6}$$

where r_p is the portfolio return, r_i, r_j denote the return of assets i and j, respectively, w_i, w_j are the weight of assets i and j in the portfolio, respectively, $\delta^2(r_p)$ denotes the total risk variance of portfolio return, and $cov(r_i, r_j)$ is the covariance between two assets.

2.3 Multi-objective optimization problem

Without loss of generality, a multi-objective optimization problem can be defined as follows:

$$\min_{\mathbf{x}\in\Omega} \mathbf{F}(\mathbf{x})=\left[f_1(\mathbf{x}),\cdots,f_M(\mathbf{x})\right]^T,$$

where Ω is an n-deminsional feasible variable space, $\mathbf{x}$ is an n-deminsional decision vector, and $\mathbf{F}(\mathbf{x}): \Omega \rightarrow R^m$ consists of m objective functions.

1) ***Pareto dominance:*** Solution $x_1 \in \Omega$ is said to dominate solution $x_2 \in \Omega$ (denoted as $x_1 \succ x_2$) if $\forall m = 1,2,\cdots,M, f_m(x_1) \leq f_m(x_2)$, and $\exists i = 1,2,\cdots,M, f_i(x_1) < f_i(x_2)$.
2) ***Pareto optimal set:*** For solution $x^\star \in S$, if there is no $x' \in S$ satisfying $x' \succ x^\star$, then $x^\star$ is called the Pareto-optimal solution. All Pareto-optimal solutions are called Pareto-optimal set, i.e., $PS = \{x^* | \neg\exists x \in \Omega : x \succ x^*\}$.
3) ***Pareto front:*** The mapping of the PS on the objective space is known as the Pareto-optimal front (PF), i.e., $PF = \{\mathbf{F}(\mathbf{x}) | \mathbf{x} \in PS\}$.

3 The proposed five-objective portfolio selection model and individual encoding

For the convenience of introduction, this paper defines the following relevant variables for the portfolio selection problem with n risk assets:

x_i: The rate of the total investment assigned to the risk asset $i, i = 1,2,\cdots,n$;
k_i: The rate of transaction cost on the risk asset $i, i = 1,2,\cdots,n$;

$\tilde{r}_i$: The fuzzy return rate on the risk asset $i, i = 1, 2, \cdots, n$;
$\mathbf{x}^0 = (x_1^0, ..., x_n^0)$: The previous portfolio;

The proposed fuzzy portfolio selection problem in this paper is a five-objective optimization model.

$$\begin{cases} \max & f_1(\mathbf{x}) = E_f\left(\sum_{i=1}^{n} \tilde{r}_i x_i\right) - \sum_{i=1}^{n} k_i \left|x_i - x_i^0\right| \\ \min & f_2(\mathbf{x}) = Var_f\left(\sum_{i=1}^{n} \tilde{r}_i x_i\right) \\ \max & f_3(\mathbf{x}) = Skew_f\left(\sum_{i=1}^{n} \tilde{r}_i x_i\right) \\ \min & f_4(\mathbf{x}) = Kur_f\left(\sum_{i=1}^{n} \tilde{r}_i x_i\right) \\ \max & f_5(\mathbf{x}) = 2 - \sum_{i=1}^{n-1} |x_i - x_{i+1}| \\ s.t & K_{min} \leq ||\mathbf{x}||_0 \leq K_{\max}, \\ & \sum_{i=1}^{n} x_i = 1, \quad x_i \geq 0, i = 1, 2, \ldots, n \end{cases} \tag{7}$$

In problem (7), $E_f(\cdot), V_f(\cdot), Skewness_f(\cdot)$, and $Kur_f(\cdot)$ are defined in Eqs. (2-5). In $f_1(\mathbf{x})$, the total transaction cost from the previous portfolio $\mathbf{x}^0$ to the current portfolio $\mathbf{x}$ is $\sum_{i=1}^{n} k_i |x_i - x_i^0|$.

The fifth objective function $f_5(\mathbf{x})$ in Eq. (7), i.e., the mean proportional entropy $2 - \sum_{i=1}^{n-1} |x_i - x_{i+1}|$, is introduced in the model to diversify the investment. This entropy is defined based on the Minkowski distance. Compared with Shannon entropy, it has higher diversity and matches a better market asset allocation.

To decentralize the investment risk and control the asset management complexity, the amount of assets usually is subject to a lower limit and an upper limit, i.e., $K_{min} \leq ||\mathbf{x}||_0 \leq K_{max}$ in the constraint of the problem (7). Here, K_{min} and K_{max} are the minimum and maximum amount of the selected assets, respectively. In the optimization of NRPEA, all maximization objectives are transferred to minimization ones by multiplying -1.

4 A new reference point-based evolutionary algorithm

This section introduces the proposed new reference point-based evolutionary algorithm (NRPEA) for portfolio selection problem, including the general framework, individual encoding scheme, evolutionary operators, reference point generation, Tchebychev selection, and environmental selection.

4.1 The general framework

The general framework of the proposed NRPEA is outlined in Algorithm 1. The procedure is similar to most of MOEAs. Firstly, a population P_t is randomly initialized. The non-dominated individuals of P_t is used to set the reference points R_t and calculate the ideal point $\mathbf{z}^*$. Secondly, an offspring population O_t is generated by genetic operators, where the Tchebychev distance $d_{i,j}^{tch}$ of a population individual to a reference point is defined as follows:

$$d_{i,j}^{tch} = \max_{m=1,2,\ldots,M} \omega_m \left(\frac{f_m\left(x_i\right) - r_j^m}{f_m^{\max} - f_m^{\min}} \right), \tag{8}$$

where $i = 1, ..., N, j = 1, ..., |R_t|$. r_j is the m-th objective value of the j-th reference point. $f_j max$ and $f_j min$ are respectively the maximum and minimum values of the j-th objective in the population. Thirdly, a clone population P_c is generated by immune clone operator on non-dominated individuals of $P_t \cup O_t$ based on their normalized crowding distance. The number of clones q_i for the i-th individual is denoted as follows:

$$q_i = \lceil N \times \frac{CD(P_t(i))}{\sum_{j=1}^{N} CD(P_t(j))} \rceil.$$

Fourthly, a set of reference points R_t is selected from $P_t \cup P_c \cup O_t$ each generation introduced in Algorithm 2. Finally, N best solutions are chosen based on the reference points as the next parent population introduced in Algorithm 3.

4.2 Individual encoding scheme

To deal with the constraints of problem (7), we use an individual encoding scheme including asset weight vector and asset selection vector. Let n be the number of risk assets. The following example assumes $K_{max} = 10$, i.e., the maximum amount of the selected assets is set as 10.

A asset weight vector $\mathbf{w} = (w_1, ..., w_n)$, $w_i \in [0.01, 0.99), \forall i \in \{1, ..., n\}$. *For example,*

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}

A asset selection vector $\mathbf{z} = (z_1, ..., z_n), z_i \in \{0, 1\}$

1	0	0	0	0	1	1	0	1	1

$\Downarrow$ *normalized*

A investment rate $\mathbf{x} = (x_1, ..., x_n), x_i \in [0.01, 0.99)$

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}

The investment rate $\mathbf{x} = (x_1, ..., x_n)$ on each asset is a normalization of the asset weight vector, i.e.,

$$x_i = l_i z_i + \frac{w_i z_i}{\sum\limits_{i=1}^{m} w_i z_i} \left(u_i z_i - \sum_{i=1}^{m} l_i z_i \right), \tag{9}$$

Algorithm 1 New reference point-based archiving evolutionary algorithm (NRPEA).

Input: Many-objective portfolio selection model, a stopping condition, the population size N, the number of evaluations.
Output: The final external archive A;
Set $t = 0$, initialize a population P_t of N individuals, set reference points R_t as non-dominated individuals of P_t, and calculate the ideal point $\mathbf{z}^*$ based on P_t;
While the stopping condition is not met **do**
  Compute the Tchebychev distance $d_{i,j}^{tch}$ of each population individual to each reference point by Eq. (8);
  Generate mating population P_m by tournament selection based on Tchebychev distance $d_i^{tch} = \min_{1\leq j\leq |R|} d_{i,j}^{tch}$;
  Apply genetic operators on P_m to generate an offspring population O_t;
  Perform non-dominated sorting for $P_t \cup O_t$;
  Generate a clone population P_c by an immune clone operator on non-dominated individuals of $P_t \cup O_t$;
  Use $P_t \cup P_c \cup O_t$ to update reference points R based on Algorithm 2;
  Conduct Environment selection based on reference points R; //see Algorithm 3;
  $t = t + 1$;
End While
Return P_t;

There is a special case $x_i = 0$ if $z_i = 0$. When $z_i = 1$, the investment rate x_i usually has the upper limit u_i and lower limit l_i due to the actual trading volume.

4.3 Evolutionary operator with a heuristic repair strategy

The individual expression includes real-valued asset weight vector and binary asset selection vector. The proposed evolutionary operator consists of crossover and mutation operators. The simulated binary crossover (SBX) is used in this paper for the real-valued asset weight vector, while the single-point crossover is used for the binary asset selection vector. The detail is showed in the following example, where $K_{max} = 10$, i.e., the maximum amount of the selected assets is set as 10.

Select variable $p_5^{(1)}$ *to crossover* *Select variable* $p_5^{(2)}$ *to crossover*

$$\begin{array}{|c|c|c|c|c|c|c|c|c|}\hline p_1^{(1)} & p_2^{(1)} & p_3^{(1)} & p_4^{(1)} & p_5^{(1)} & p_6^{(1)} & \cdots & p_9^{(1)} & p_{10}^{(1)} \\ \hline \end{array} \quad \begin{array}{|c|c|c|c|c|c|c|c|c|}\hline p_1^{(2)} & p_2^{(2)} & p_3^{(2)} & p_4^{(2)} & p_5^{(2)} & p_6^{(2)} & \cdots & p_9^{(2)} & p_{10}^{(2)} \\ \hline \end{array}$$

$$p_5^{(1)} = w_5^{(1)} + a_{73}^1 \qquad p_5^{(2)} = w_5^{(2)} + a_{48}^2$$

$\Downarrow$ *Crossover*

Variable $o_5^{(1)}$ *in the 1st offspring* $\downarrow$ *Variable* $o_5^{(2)}$ *in the 2nd offspring* $\downarrow$

$$\underbrace{\begin{array}{|c|c|c|c|c|c|c|c|c|}\hline o_1^{(1)} & o_2^{(1)} & o_3^{(1)} & o_4^{(1)} & o_5^{(1)} & o_6^{(1)} & \cdots & o_9^{(1)} & o_{10}^{(1)} \\ \hline\end{array}}_{o_5^{(1)} = \bar{w}_5^{(1)} + \bar{a}^1_{(73||random)}} \quad \underbrace{\begin{array}{|c|c|c|c|c|c|c|c|c|}\hline o_1^{(2)} & o_2^{(2)} & o_3^{(2)} & o_4^{(2)} & o_5^{(2)} & o_6^{(2)} & \cdots & o_9^{(2)} & o_{10}^{(2)} \\ \hline\end{array}}_{o_5^{(2)} = \bar{w}_5^{(2)} + \bar{a}^2_{(48||random)}}$$

In the above crossover operator, $w_5^{(1)}$ is the 5th asset weight in the 1st parent individual, and a_{73}^1 is the 73-th asset in the 1st parent individual. $\bar{a}^1_{(73||random)}$ indicates that the index of the select asset is 73 mainly and a random integer partly.

The polynomial mutation is used for the real-valued asset weight vector, while the bit-flip mutation is used for the binary asset selection vector.

After crossover operator and/or mutation operator, the same asset may appear in different variables of offspring. Thus, the repeatedly selected asset need to be replaced. To deal with this issue, this paper proposed a heuristic repair strategy based on the occurrence frequency of each asset in the non-dominated archive A, i.e.,$c_i = \frac{\sum_{j=1}^{|A|} s_{i,j}}{|A|}, i = 1, ..., n,$.

where $s_{i,j} \in \{0, 1\}$ is the number of the i-th asset appearing in the j-th archived individual. The higher the asset c_i score is, the more likely it will appear in the next generation. According to the occurrence frequency in the non-dominated archive, roulette wheel select is used to select an asset to replace the repeated selected asset in the offspring.

4.4 Constructing reference points

In the proposed NRPEA, a set of reference points with good convergence and diversity is constructed to guide the population evolution. Algorithm 2 presents the constructing process of reference points. First, we set Q_t as non-dominated individuals of $P_t \cup P_c \cup O_t$. Second, given a tolerant vector δ, each individual $F(\boldsymbol{x}^i) = [f_1(\boldsymbol{x}^i), ..., f_M(\boldsymbol{x}^i)]$ in set Q_t generates M auxiliary reference points as follows:

$$\mathbf{r}_m^{(i)} = \left[f_1(\boldsymbol{x}^i), \ldots, f_m(\boldsymbol{x}^i) - \varepsilon_m, f_{m+1}(\boldsymbol{x}^i), \ldots, f_M(\boldsymbol{x}^i)\right] \qquad (10)$$

Thus, $M|Q_t|$ auxiliary reference points are generated as shown in Fig. 1. Finally, N best individuals are selected from the auxiliary reference points as the reference points based on non-dominated sorting for the subsequent environment selection.

4.5 Environment selection based on reference points

Algorithm 3 gives a detailed procedure of the environment selection based on reference points. More specifically, the fitness assignment of an individual is based on the distance to reference points R_t. First, the Tchebychev distances (8) between the individuals and the reference points are calculated to evaluate the individuals in lines 2-6. Second, for each reference point, the individual with

Algorithm 2 Constructing reference points

Input: Learning parameter δ, α, and the combined population $P_t \cup P_c \cup O_t$.
Output: The reference points R_t;
Set the non-dominated individuals of $P_t \cup P_c \cup O_t$ as Q_t;
Generate an auxiliary reference point set R base on Q_t, δ, and Eq. (10);
Select the non-dominated reference points of R as R_t;
If $|R_t| > N$, choose N best individuals with the largest crowding distances as R_t;
ReturnR_t

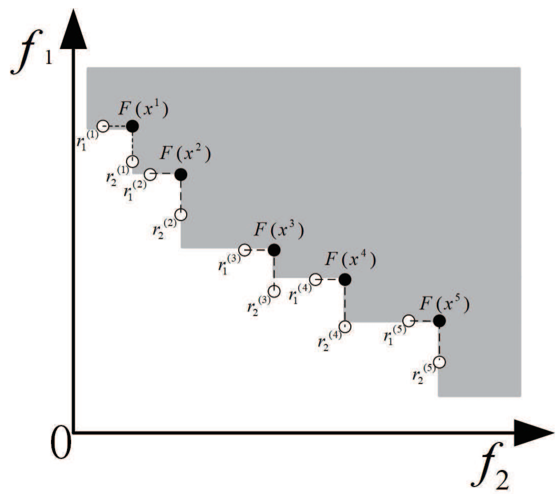


Fig. 1: Generate the auxiliary reference points.

the smallest Tchebychev distance to the reference point is chosen to constitute the next parent population in lines 9-11. It is noted that the number of reference points R_t may be less than the population size N. To deal with this issue, all reference points are available again to choose the rest individuals in line 8.

5 Experimental study

In this section, the proposed NRPEA is compared with three state-of-the-art many-objective evolutionary algorithms. All compared algorithms are implemented in MATLAB. Thirty independent runs are performed for each algorithm on each test problem.

5.1 Dataset and data processing

Six datasets collected from well-known international stock markets are used to test the performance of NRPEA. They are authoritative test cases for portfo-

Algorithm 3 Environment selection based on reference points

Input: The reference points $R_t = \{r^1, \cdots, r^{|R_t|}\}$, the population size N, and the combined population $S_t = P_t \cup P_c \cup O_t$;
Output: The next parent population P_{t+1};
Set $P_{t+1} = \emptyset$ and $R' = R_t$;
For each $\mathbf{x}^i \in S_t$ **do**
    For each $r^j \in R_t$ **do**
        Calculate Tchebychev distance d_{ij}^{tch} based on Eq. (8).
    End For
End For
While $|P_{t+1}| < N$ **do**
    If $R_t = \emptyset$ **then**
        $R_t = R'$;
    End If
    For each $r^j \in R_t$ **do**
        Find the individual $\mathbf{x}^{j_m in}$ with the smallest Tchebychev distance to r^j;
    End For
    $R_t = R_t \setminus \{r^j\}$, $P_{t+1} = P_{t+1} \cup \{\mathbf{x}^{j_{min}}\}$, $S_t = S_t \setminus \{\mathbf{x}^{j_{min}}\}$;
End While
Return P_{t+1};

lio selection optimization and are publicly available from Pubmed central (PMC) at:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4959918/. These datasets provide weekly return time series for assets and stock indexes of several major markets across the world. The details of datasets are shown as in Table I.

For the fuzzy return rates of the assets, we use the statistical indicator of sample percentage to approximate the trapezoidal fuzzy returns distribution [15]. In numercial statistics, the core $[a_i, b_i]$ of the fuzzy return $\tilde{r}_i$ is set to the interval $[40th, 60th]$, and the quantities $40th - 5th$ and $95th - 60th$ as the left α_i and right β_i spreads, respectively, where $k-th$ is the $k-th$ percentile of the sample. The corresponding membership function is defined in *Definition 1*, i.e.,

$$\mu_{\tilde{r}_i}(x) = \begin{cases} 1 - \frac{a_i - x}{\alpha_i}, & \text{if } a_i - \alpha_i \le x \le a_i \\ 1, & \text{if } a_i \le x \le b_i \\ 1 - \frac{x-d}{\beta_i}, & \text{if } b_i \le x \le b_i + \beta_i \\ 0, & \text{if otherwise} \end{cases} \tag{11}$$

At the same time, we calculate the correlation coefficient matrix $r_{i,j}$ among these assets, the average value of weekly return, and the expected return $\bar{r}$ on each asset. The relevant data is used in the learning mechanism to guide the generation of candidate good assets.

Table 1: This benchmark instances datasets from PMC

	Dataset Name	N	|T|	Time interval	Country	Description	# of rebalancing
1	DowJones	28	1363	Feb 1990-Apr 2016	USA	Dow Jones Industrial Average	110
2	NASDAQ100	82	596	Nov 2004-Apr 2016	USA	NASDAQ 100	46
3	FTSE100	83	717	Jul 2002-Apr 2016	UK	FTSE 100	56
4	SP500	442	595	Nov 2004-Apr 2016	USA	S&P 500	46
5	NASDAQComp	1203	685	Feb 2003-Apr 2016	USA	NASDAQ Composite	53
6	FF49Industries	49	2325	Jul 1969-Jul 2015	USA	Fama and French 49 Industry	190

5.2 Performance metric and the experimental settings

In this paper, hyper-volume (HV) metric is used as the performance measure to assess the convergence and diversity of the obtained solution set [16], Generally speaking, a larger value of HV metric indicates a better approximation to the true PF. The HV indicator is defined as follows:

$$HV(S) = \text{Volume}\left(\bigcup_{x^i \in S} \left[F(x^i), \mathbf{r}\right]\right), \tag{12}$$

where $\left[F(x^i), \mathbf{r}\right]$ indicates a hyper-rectangle formed by the reference point $\mathbf{r}$ and the i-th non-dominated individual $F(x^i)$. To compute the HV metric, the reference point $\mathbf{r}$ is calculated as $1.1 \times (f_1^{max}, f_2^{max}, \cdots, f_m^{max})$, where $f_k^{max} \quad (k = 1, 2, \cdots, m)$ is the maximum value of the kth objective in the true PF.

NRPEA is compared with the other three state-of-the-art MOEAs, i.e., NSGA-III [17], RSEA [18], and SPEAR [19,20]. The population size of the four MOEAs is set to 100, and the maximal function evaluation times is set to 100,000. η_c and η_m indicate the distribution indexes of SBX and PM, respectively. p_c and p_m represent the probabilities of SBX and PM, respectively. More details of the parameter settings for the MOEAs are summarized in Table II. where CR and F in DE operator use the recommended setting in the proposed paper.

5.3 Experimental results

Table III shows the mean and standard deviation of the HV-metric values by each algorithm in solving the six portfolio problems. Wilcoxon's rank-sum test is performed to show the statistically significant differences between the results obtained by NRPEA and another algorithm with a 5% significance level. The results show that the proposed NRPEA outperforms the other algorithms in most of the benchmark problems in terms of the HV metric.

Fig. 2 reports the evolution process of the mean HV-metric values. It can be seen that the proposed NRPEA obtains the best performance on most of the benchmark problems based on the HV metric .

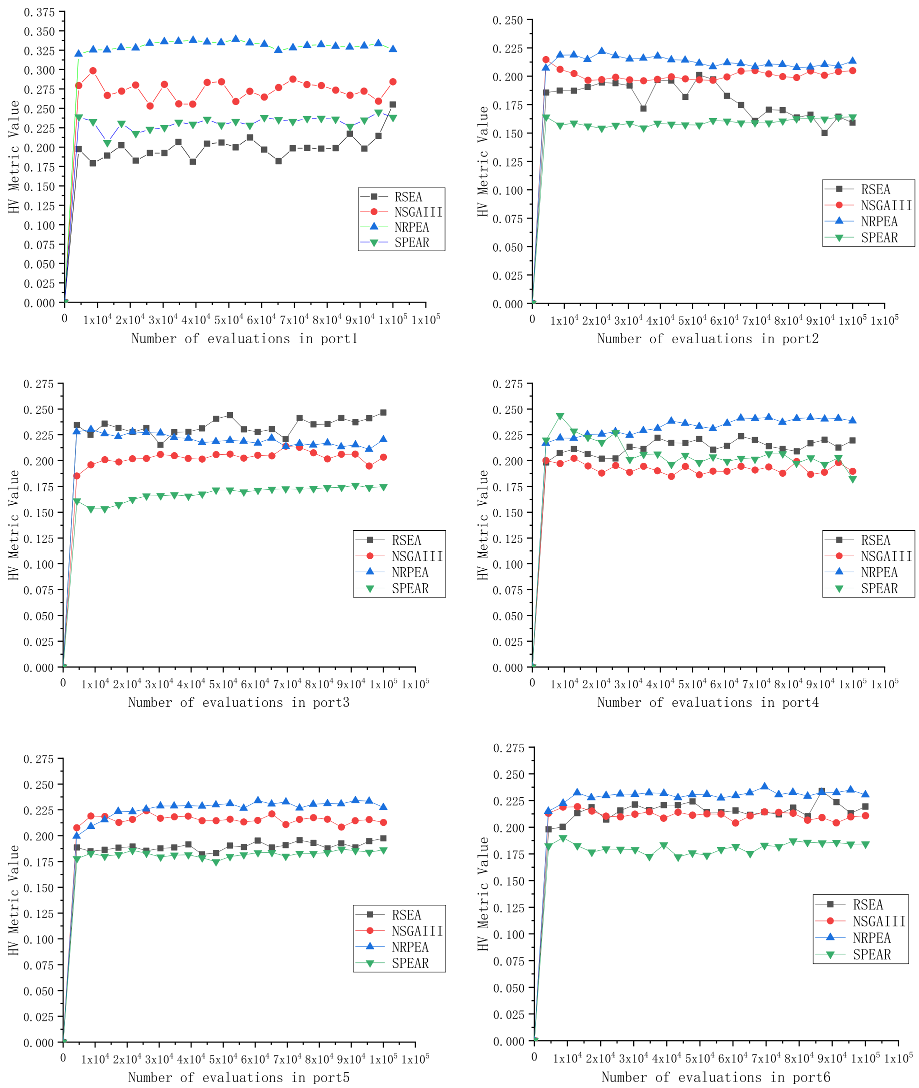


Fig. 2: The evolution process of the mean HV-metric values.

Table 2: Parameter settings of the compared algorithms

Algorithm	Parameter settings
RSEA	$CR = 1.0, F = 0.5, T = 1/N$
	$p_c = 1, p_m = 1/n, \eta_c = 20, \eta_m = 20$
NSGA-III	$p_c = 1, p_m = 1/n, \eta_c = 20, \eta_m = 20$
SPEAR	$p_c = 1, p_m = 1/n, \eta_c = 20, \eta_m = 20$
NRPEA	$\alpha = 0.4, \beta = 0.1$
	$p_c = 1.0, p_m = 1/n, \eta_c = 20, \eta_m = 20$

Table 3: Mean and standard deviation values of the HV metric obtained by the compared algorithm

Problem	NSGAIII	SPEAR	RSEA	NRPEA
TR_port1	2.8517e-1 (4.28e-2) -	2.3560e-1 (3.29e-2) -	2.3790e-1 (4.30e-2) -	**3.3139e-1 (1.39e-2)**
TR_port2	2.0261e-1 (1.46e-2) -	1.6440e-1 (1.07e-2) -	1.5152e-1 (4.04e-2) -	**2.1360e-1 (1.51e-2)**
TR_port3	2.0583e-1 (1.84e-2) -	1.7636e-1 (1.89e-2) -	**2.4176e-1 (3.06e-2)** =	2.2059e-1 (2.31e-2)
TR_port4	1.8954e-1 (2.41e-2) -	1.8246e-1 (3.38e-2) -	1.8869e-1 (3.31e-2) -	**2.3850e-1 (1.58e-2)**
TR_port5	2.1240e-1 (2.62e-2) -	1.8481e-1 (2.15e-2) -	1.9716e-1 (3.92e-2) -	**2.3395e-1 (2.66e-2)**
TR_port6	2.1585e-1 (2.47e-2) -	1.8538e-1 (2.93e-2) -	2.1990e-1 (1.55e-2) =	**2.3297e-1 (2.34e-2)**
+/-/=	**0/6/0**	**0/6/0**	**1/4/1**	**-**

5.4 Portfolio performance evaluation based on the improved Sharpe ratio

Sharpe rate (SR) is usually used to evaluate the performance of the portfolio. It measures the asset return under different risk situations and is calculated by the following form:

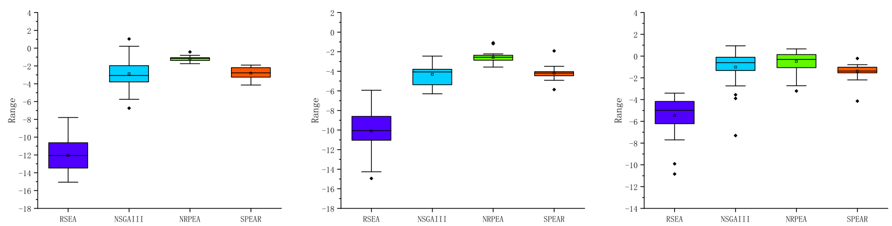
$$\mathrm{SR}(\mathrm{r}(\mathbf{x})) = E(r(\mathbf{x}))/\sqrt{Var(r(\mathbf{x}))}, \tag{13}$$

where $r_{\mathbf{x}}$ is the return of a portfolio $\mathbf{x}$. However, the Sharpe ratio based on the mean-variance theory is not accurate when the return distribution appears skew or kurtosis. To deal with this issue, we use an adjusted Sharpe ratio (ASR) [21], which takes into account portfolio skewness and kurtosis:

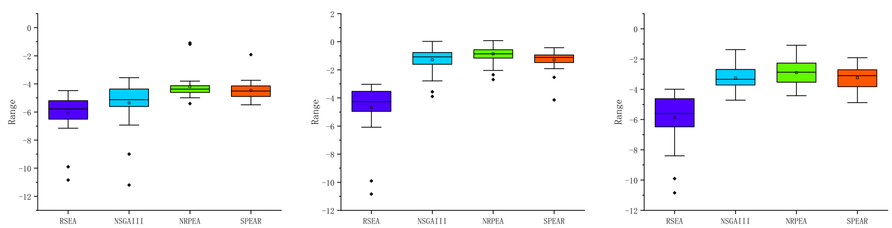
$$\begin{aligned} ASR(r(\mathbf{x})) = SR(r(\mathbf{x}))[1 + (Skew(r(\mathbf{x}))/6)SR(r(\mathbf{x})) \\ - (Kur(r(\mathbf{x}))/24)SR(r(\mathbf{x}))^2] \end{aligned} \tag{14}$$

Skewness $Skew(r(x))$ and kurtosis $Kur(r(x))$ are two important high-order moments. Different from SR (13), ASR (14) reflects the average expected utility of the portfolio $\mathbf{x}$. The bigger ASR is, the better the performance of the corresponding portfolio $\mathbf{x}$ is.

Fig. 3 shows boxplots of the ASR values attained by four MOEAs on six benchmark problems. It can be seen that the proposed NRPEA obtains superior or comparable performance to other MOEAs on most of the benchmark problems based on the ASR values.



(a) Boxplots of ASR attained by five MOPSOs in prort1 and port2 and port3



(b) Boxplots of ASR attained by five MOPSOs in prort4 and port5 and port6

Fig. 3: Boxplots of the ASR values attained by four MOEAs on six test problems

6 Conclusion

Compared with a random variable, fuzzy number variable based on the knowledge of experts is more suitable to estimate the risk asset due to politics, social conditions, economic change, and the status of the related company. To deal with complex portfolio selection problem, this paper builds five-objective fuzzy portfolio selection problem enhancing the classical Markowitz mean-variance problem by introducing skewness, kurtosis, and the mean proportional entropy. A new reference point-based evolutionary algorithm (NRPEA) to deal with this five-objective problem. The auxiliary reference points are generated and selected to guide the population evolution. Finally, experimental results show that the proposed NRPEA has obtained better performance than other compared algorithm.

For future work, the proposed NRPEA can be extended considering multi-period portfolio and constraints, such as assets liquidity measurement and short selling.

References

1. Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.
2. Paresh Kumar Narayan and Huson Ali Ahmed. Importance of skewness in decision making: evidence from the indian stock exchange. *Global Finance Journal*, 25(3):260–269, 2014.
3. Rui Pedro Brito, Hélder Sebastião, and Pedro Godinho. Efficient skewness/semivariance portfolios. *Journal of Asset Management*, 17(5):331–346, 2016.

4. Rubén Saborido, Ana B Ruiz, José D Bermúdez, Enriqueta Vercher, and Mariano Luque. Evolutionary multi-objective optimization algorithms for fuzzy portfolio selection. *Applied Soft Computing*, 39:48–63, 2016.
5. Zahra Mashayekhi and Hashem Omrani. An integrated multi-objective markowitz–dea cross-efficiency model with fuzzy returns for portfolio selection problem. *Applied Soft Computing*, 38:1–9, 2016.
6. Xiaoxia Huang. Mean-entropy models for fuzzy portfolio selection. *IEEE Transactions on Fuzzy Systems*, 16(4):1096–1101, 2008.
7. Zhongfeng Qin, Xiang Li, and Xiaoyu Ji. Portfolio selection based on fuzzy cross-entropy. *Journal of Computational and Applied mathematics*, 228(1):139–149, 2009.
8. Hisao Ishibuchi, Yu Setoguchi, Hiroyuki Masuda, and Yusuke Nojima. Performance of decomposition-based many-objective algorithms strongly depends on pareto front shapes. *IEEE Transactions on Evolutionary Computation*, 21(2):169–190, 2016.
9. Yuan Yuan, Hua Xu, Bo Wang, Bo Zhang, and Xin Yao. Balancing convergence and diversity in decomposition-based many-objective optimizers. *IEEE Transactions on Evolutionary Computation*, 20(2):180–198, 2015.
10. Yiping Liu, Dunwei Gong, Xiaoyan Sun, and Yong Zhang. Many-objective evolutionary optimization based on reference points. *Applied Soft Computing*, 50:344–355, 2017.
11. L. Pan, C. He, Y. Tian, Y. Su, and X. Zhang. A region division based diversity maintaining approach for many-objective optimization. *Integrated Computer-Aided Engineering*, 24(3):279–296, 2017.
12. C. He, Y. Tian, Y. Jin, X. Zhang, and L. Pan. A radial space division based evolutionary algorithm for many-objective optimization. *Applied Soft Computing*, 61:603–621, 2017.
13. L. Pan, C. He, Y. Tian, H. Wang, X. Zhang, and Y. Jin. A classification-based surrogate-assisted evolutionary algorithm for expensive many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 23(1):74–88, 2018.
14. L. Pan, L. Li, C. He, and K. C. Tan. A subregion division-based evolutionary algorithm with effective mating selection for many-objective optimization. *IEEE transactions on cybernetics*, 2019. DOI: 10.1109/TCYB.2019.2906679.
15. Wei Yue, Yuping Wang, and Cai Dai. An evolutionary algorithm for multiobjective fuzzy portfolio selection models with transaction cost and liquidity. *Mathematical Problems in Engineering*, 2015, 2015.
16. Kalyanmoy Deb, Ankur Sinha, and Saku Kukkonen. Multi-objective test problems, linkages, and evolutionary methodologies. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pages 1141–1148. ACM, 2006.
17. Wiem Mkaouer, Marouane Kessentini, Adnan Shaout, Patrice Koligheu, Slim Bechikh, Kalyanmoy Deb, and Ali Ouni. Many-objective software remodularization using nsga-iii. *ACM Transactions on Software Engineering and Methodology*, 24(3):17–27, 2015.
18. Wei Yue and Yuping Wang. A new fuzzy multi-objective higher order moment portfolio selection model for diversified portfolios. *Physica A: Statistical Mechanics and its Applications*, 465:124–140, 2017.
19. Shouyong Jiang and Shengxiang Yang. A strength pareto evolutionary algorithm based on reference direction for multiobjective and many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 21(3):329–346, 2017.
20. Cheng He, Ye Tian, Yaochu Jin, Xingyi Zhang, and Linqiang Pan. A radial space division based evolutionary algorithm for many-objective optimization. *Applied Soft Computing*, 61:603–621, 2017.

21. Valeri Zakamouline and Steen Koekebakker. Portfolio performance evaluation with generalized sharpe ratios: Beyond the mean and variance. *Journal of Banking & Finance*, 33(7):1242–1254, 2009.