A Property Preserving Method for Extending a Single-Objective Problem Instance to Multiple Objectives with Specific Correlations

Ruby L.V. Moritz[1(✉)], Enrico Reich[2], Matthias Bernt[2], and Martin Middendorf[2]

[1] Intelligent Systems Group, Otto-von-Guericke University Magdeburg, Magdeburg, Germany
`ruby.moritz@ovgu.de`
[2] Parallel Computing and Complex Systems Group, Institute of Computer Science, University of Leipzig, Leipzig, Germany
`enrico.reich@gmx.de, {bernt,middendorf}@informatik.uni-leipzig.de`

Abstract. A method is proposed to generate multi-objective optimization problem instances from a corresponding single-objective instance. The user of the method can specify the correlations between the generated the objectives. Different from existing instance generation methods the new method allows to keep certain properties of the original single-objective instance. In particular, we consider optimization problems where the objective is defined by a matrix, e.g., a distance matrix for the Traveling Salesperson problem (TSP) or a flow matrix for the Quadratic Assignment problem. It is shown that the method creates new distance matrices with specific correlations between each other and also have the same average distance and variance of distances as the distance matrix of the original instance. This property is important, e.g., when the influence of correlations between the objectives on the behavior of metaheuristics for the multi-objective TSP are investigated. Some properties of the new method are shown theoretically. In an empirical analysis the new method is compared with instance generation methods from the literature.

Keywords: Multi-objective optimization · Problem instance generation · Traveling salesperson problem

1 Introduction

The empirical analysis of a metaheuristic or other types of algorithms for an optimization problem is usually done on a set of (benchmark) test instances. These have various property values to study how their properties influence the optimization behavior. Considering the *Traveling Salesperson Problem* (TSP) an increased standard deviation of the distances between the cities of a TSP instance increases its difficulty for *Ant Colony Optimization* (ACO) [22].

Other relevant properties that might influence the difficulty of a TSP instance for an optimization algorithm are the distribution of the distances between the cities, if the cities occur clustered or not, and the size of the convex hull of the cities [16]. The state of such properties might also influence the best choice of parameter values for a metaheuristic (e.g. shown in [19] for ACO).

For applications it is important to have real world test instances or at least test instances that reflect the properties of real world instances. However, for many optimization problems only a relatively small number of real world test instances is available. In those cases it is difficult to find real world instances with specific and varying properties and a method for generating problem instances with such properties and similarities to given real world instances is helpful.

In this paper such a problem instance generation method for *multi-objective optimization problems* (MOOPs) is proposed. An important and characteristic property of MOOPs is if and how the objectives are correlated. Many real world MOOP instances have inter-dependencies between their objectives that influence the characteristics of their fitness landscape [25]. It was shown that correlations between the objectives can influence the correlation between the Pareto optimal solutions [13]. In general, it holds that objectives which are not positively correlated lead to a diverse and large set of Pareto optimal solutions [9]. In contrast, positively correlated objectives decrease the number of Pareto optimal solutions, e.g. for NK-landscapes [23].

It has been demonstrated for various MOOPs that the correlations between objectives can influence the performance of optimization algorithms. If some objectives have a strong positive correlation, a dimensionality reduction can have positive effects on the performance of a metaheuristic [1,5]. To this end, groups of positively correlated objectives can often be aggregated into a single objective [18]. Several studies have demonstrated the influence of the correlation between objectives on the performance of metaheuristics, e.g. [4]. A co-influence of the correlation between objectives, the dimension of the objective space, and the degree of non-linearity on the size of the Pareto set was shown in [24].

Often, a multi-objective problem is an extension of a corresponding single-objective problem. One example is the *multi-objective TSP* (MO-TSP) which is to find a round trip through n given cities that minimizes the traveled distance with respect to multiple $n \times n$ distance matrices. For the MO-TSP it has been demonstrated that correlated objectives influence the optimization behavior of population based ACO (P-ACO) algorithms [17]. Another example is the *multi-objective 0/1 Knapsack problem* which is to find an assignment of a subset of n items to k knapsacks, where each knapsack has a weight capacity limit and the items have knapsack specific weights and profits that are defined by $n \times k$ matrices, such that the total profit is maximized and the weights in the knapsacks satisfy the capacity constraints. The influence of correlated objectives on evolutionary multi-objective algorithms for this problem has been investigated in [6–8]. The performance of MOEA/D is severely degraded by an increase in the number of objectives when they are strongly correlated [7]. Algorithms NSGA-II and SPEA2, on the other hand, perform well on multi-objective problems with strongly correlated objectives [6]. Also, the search behavior of the hypervolume-

based SMS-EMOA algorithm is biased toward the region of the Pareto front for duplicated (i.e. strongly correlated) objectives [8]. The last example is the *Quadratic Assignment Problem* (QAP) where given are n facilities, n locations, a flow matrix $F = [f_{ij}]$ where f_{ij} is the flow from facility i to j, and a distance matrix $D = [d_{ij}]$ where d_{ij} is the distance between locations i and j. The problem is to find an assignment π of the facilities to the locations such that $\sum_{i,j \in [1:n]} f_{ij} d_{\pi(i)\pi(j)}$ is minimized. The *multi-objective QAP* considers multiple flows, where each flow defines one objective. It was shown in [20] that the performance of local search operators (which are a central component of many metaheuristic algorithms) for the multi-objective QAP are strongly influenced by the strength of the correlation between the objectives. ACO algorithms for multi-objective QAPs perform better with 'less aggressive' search strategies (e.g. with iteration-best pheromone update instead of best-so-far pheromone update) when the objectives have a strong positive correlation [14]. On instances with weak or negative correlation between the objectives this does not hold.

As shown by the above examples, various combinatorial MOOP instances are defined mainly by a set of matrices where each objective is represented by one matrix. In the examples there is one distance matrix for each objective of the multi-objective TSP, one flow matrix for each objective of the multi-objective QAP, and one profit matrix for each objective of the multi-objective 0/1 Knapsack problem. In all these cases, the values of the objective functions depend directly on the values of the matrix. Moreover, correlations between the matrices result in a correlation between the respective objectives.

The discussion shows that it would be very helpful for analyzing the optimization behavior of metaheuristics for MOOPs to have the following type of methods for generating MOOP instances. Starting from a given real world single-objective problem instance a method generates new MOOP instances such that:

1. the correlation between the objectives of a MOOP instance can be controlled by the user and
2. some important properties of the single-objective problem instance also hold for a newly generated MOOP instances.

In this paper such an instance generation method for MOOPs where each objective is defined by a matrix is presented. In particular, the method generates new matrices from a given matrix such that the following two properties hold: (i) the correlation between the generated matrices (and the given matrix) can be controlled by the user and (ii) the mean value and the variance of the values of each generated matrix are equal to the corresponding values for the given matrix. Thus, the presented instance generation method is suitable to generate correlated cost matrices for different MOOPs. The new method improves a method that we have presented in [17] which can generate correlated matrices from a given matrix such that property (i) is guaranteed (to a certain extend) whereas property (ii) does not hold.

In the following section related work is described. Our new instance generation method and its properties are described in Sect. 3. Experimental results for the new instance generation method when applied to the TSP are presented in Sect. 4. Conclusions are given in Sect. 5.

2   Related Work

Two common approaches to generate instances of combinatorial MOOPs are to generate (usually uniformly distributed) random cost matrices (e.g. [10]) and to combine single-objective test instances of the same size (as done, e.g. in [9] by combining instances of the TSPLIB [21]). The latter approach creates test instances with a potentially larger practical relevance, but the number of generatable instances is very limited. Both approaches are certainly a useful practice but do not allow to control the correlation between the objectives. The creation of correlated objective functions by using different linear combinations of two random single objective functions using two parameters $\alpha, \beta \in (-1, 1)$ has been suggested in [7,8] and applied to the multi-objective 0/1 Knapsack problem. The practical relevance of this approach is limited due to the use of strict linear (cor)relations. Related problems also occur in multi objectivization approaches which try to avoid local optima by adding objectives to single objective problems such that the global optimum is not affected. Typically subproblems are used as additional objectives, e.g. [11] suggested to use the length of a sub-tour as second objective for the TSP. Also for MOOPs subproblems could define large sets of objectives, but the correlation can not be controlled easily.

The first problem instance generator for MOOPs with correlated objectives was proposed for the multi-objective QAP with k flow matrices that define the different objectives [2,13]. Entries of the flow matrices are defined by a (exponential) function of a random variable. For the first matrix a uniformly random variable X is used and for jth matrix, $j \in [2 : k]$, some entries are generated using a random variable X_j that is correlated with X and the remaining with an independent random variable X_j'. The correlation and the random fraction of the matrix is set with parameters. Several authors have used the QAP generator to study the influence of correlation – mostly for bi-objective problems – on the optimization behavior of metaheuristics and local search operators [4,14,20].

The generation of instances of the multi-objective TSP with correlated objectives was covered by [12]. For each pair of cities (i, j) k distance values $d_h(i, j)$, $h \in [2 : k]$ were created with

$$d_h(i, j) = \alpha \cdot d_{h-1}(i, j) + (1 - \alpha) \cdot rand \qquad (1)$$

where the values $d_1(i, j)$ are chosen uniformly at random from $[0, 1]$, $\alpha \in [-1, 1]$ is a "correlation parameter", and $rand$ is a uniform random number from $[0, 1]$. Let us observe here, that (1) has the following potential problems: (i) since the distance values can become > 1 for $\alpha < 0$ even for distance values from $[0, 1]$ for the original matrix it might lead to an uneven influence of different objectives on the behavior of metaheuristics, (ii) the distance values are randomized to a different extent for α and $-\alpha$.

All methods mentioned above generate problem instances that have an inhomogeneous correlation structure, i.e. different strengths of pairwise correlations occur between the objectives. In the method of [2,13] each of the objectives 2 to k has a defined (possibly identical) correlation with the first objective.

But the correlation between pairs of the objectives 2 to k might be different. In the method from [12] each objective h, $h \in [2 : k]$ has a defined correlation with objective $h - 1$. But there are many different pairwise correlations between objectives i and j with $|i - j| \geq 2$. Also, for each of the problem instances created by the methods of [7] different pairwise correlations occur.

A simple method to create multi-objective TSP test instances with a homogeneous correlation structure was presented in [17]. The method solves some of the possible disadvantages of [12] by using the following equation that differentiates between the cases of positive and negative correlation:

$$d_h(i,j) = \begin{cases} \alpha \cdot d(i,j) & +(1 - \alpha) \cdot rand \text{ for } \quad 0 \leq \alpha \leq 1 \\ |\alpha| \cdot (1 - d(i,j)) & +(1 - |\alpha|) \cdot rand \text{ for } -1 \leq \alpha < 0. \end{cases} \tag{2}$$

The authors suggested to use normalized real world instances for the original matrix, e.g. from the TSPLIB. The homogeneity is achieved by creating all k objectives on the same original distance matrix which is then discarded. It was shown that the Pearson correlation coefficient of the pairs of matrices depends on the parameter α and the variance of the original distance matrix.

A method to design MOOP instances where the correlation between the objectives is defined by a correlation matrix has been presented in [23]. In particular, NK-landscapes have been investigated, but according to [23] the method can also be applied more generally to other MOOPs. For their empirical investigations the same correlation strength was used for each pair of objectives.

A topic that is related to the generation of MOOP instances is the generation of test instances for dynamic optimization problems. These problems are often single-objective problems, but the objective function changes over time. Here it is interesting to study how the strength of the modification of the objective function, e.g. measured by the correlation between the new function and the old function, influences the optimization behavior of metaheuristics. For the dynamic TSP it has been suggested to create the dynamics by renaming a subset of the cities [15]. The size of the renamed subset influences the extent of the change.

3   Method

In the following we present our method to generate a multi-objective TSP instance from a given distance matrix such that the generated distance matrices are expected to (i) correlated to each other (and to the original matrix) in a user defined way and (ii) have the similar statistical properties as the original matrix, in particular, they have the same expected mean value and variance.

From now, we consider only distance matrices $D = (d_{ij})$ that are symmetric, i.e. $d_{ij} = d_{ji}$ and where all diagonal values are zero, i.e. $d_{ii} = 0$, for $i, j \in [1 : n]$. Let D be such a distance matrix. Then, $\bar{d} = 2/(n^2 - n) \sum_{i<j} d_{ij}$ is the mean value of all elements in the upper triangular submatrix and $s^2 = 2/(n^2 - n - 1) \sum_{i \neq j} (d_{ij} - \bar{d})^2$ is the sample variance of all elements in the upper triangular submatrix.

The instance generation method uses matrix D to create k symmetric distance matrices $D_1, \ldots, D_k$ that define the k objectives of a multi-objective TSP instance. Using the parameters $q \in [1 : k]$ and $\alpha \in [0,1]$ to determine the correlation structure the method creates a set of distance matrices $\{D_1, \ldots, D_k\}$ such that each matrix D_i with $i \in [1 : q]$ is positively correlated with D and each matrix in D_i with $i \in [q + 1 : k]$ is negatively correlated with D. For $D^+ \in \{D_1, \ldots, D_q\}$ or $D^- \in \{D_{q+1}, \ldots, D_k\}$ the method works as follows:

Step 1: Two random symmetric distance matrices R_1 and R_2 are created by randomly sampling with uniform probability and with replacement from the elements within D. In order to create a symmetric matrix with all diagonal values being zero only values in the corresponding upper triangular matrices are considered and the values in the other half are set accordingly.

Step 2: Matrix D^+ (respectively D^-) is created by:

$$D^+ = \alpha \qquad D + (1 - \alpha)R_1 + \sqrt{2\alpha(1 - \alpha)}(R_2 - \overline{d}) \qquad (3)$$

$$D^- = \alpha(2\overline{d} - D) + (1 - \alpha)R_1 + \sqrt{2\alpha(1 - \alpha)}(R_2 - \overline{d}). \qquad (4)$$

Note that different randomized matrices R_1, R_2 need to be used for each D_i, $i \in [1 : k]$.

Similar to the methods proposed in [12, 17], see also (1) and (2), the first (resp. second) summand in (3) and (4) represent the non-random influence (resp. the random influence) which is controlled by α. The main differences are (i) that the random influence by R_1 is realized via sampling from D (which has the effect that the mean is preserved) and (ii) that a third summand is added (which has the effect that the variance is preserved). Furthermore, the mean $\overline{d}$ of D is used for determining the influencing of the non-random component in (4).

Observe, that the newly generated matrices can contain negative distance values. Therefore, we propose to add the absolute value of the smallest negative value that is contained in the generated matrices (let this value be d_{min}) to all (nondiagonal) values in all distance matrices. Adding d_{min} does not change the correlation and variance and it increases the mean value for all matrices by exactly d_{min}. Hence after such an operation, all matrices are still expected to share the same variance and mean value. In case zero values are undesired a larger constant value can be added, e.g. the sum d_{min} and the minimum of the original matrix which would yield a matrix with the same minimum.

For the following theoretical results we specify the method in terms of random variables. Analogous to the creation of the matrices a set of random variables $\{X_1, \ldots, X_k\}$ can be created such that each $X^+ \in \{X_1, \ldots, X_q\}$ is positively correlated with X and each $X^- \in \{X_{q+1}, \ldots, X_k\}$ is negatively correlated with X. For a given random variable X and independent random variables Z_1, Z_2 that have the same distribution as X we define

$$X^+ = \alpha \qquad X + (1 - \alpha)Z_1 + \sqrt{2\alpha(1 - \alpha)}(Z_2 - E[X]) \qquad (5)$$

$$X^- = \alpha(2E[X] - X) + (1 - \alpha)Z_1 + \sqrt{2\alpha(1 - \alpha)}(Z_2 - E[X]). \qquad (6)$$

The relation to our method is as follows. The distance values in the upper triangular submatrix of D can be considered to represent the distribution of a random variable X with $E[X] = \bar{d}$ and $V[X] = s^2$. By sampling R_1 and R_2 from D with replacement the corresponding random variables Z_1, Z_2 are independent to X and from the same distribution as X. Correlation between two random variables X and Y is measured in this paper by the correlation coefficient $\rho(X,Y) = \sigma(X,Y)/\sqrt{V[X]V[Y]}$ where $\sigma(X,Y) = E[XY] - E[X]E[Y]$ is the covariance. The following theorem shows how X^+ and X^- are related to X.

Theorem 1. *For random variables X^+ and X^- as given in* (5) *and* (6), *respectively, it holds that: (i) $E[X] = E[X^+] = E[X^-]$, (ii) $V[X] = V[X^+] = V[X^-]$, and (iii) $\rho(X^+, X) = \alpha$ and $\rho(X^-, X) = -\alpha$.*

Proof. By definition $E[Z_i] = E[X]$ and $V[Z_i] = V[X]$, $i \in \{1,2\}$. Due to space limitations the proof is shown only for $\rho(X^+, X)$. For ease of readability set $\beta := \sqrt{2\alpha(1-\alpha)}$. Since Z_1, Z_2, and X are independent random variables it holds that

$$
\begin{aligned}
\rho(X, X^-) &= \frac{E[XX^-] - E[X]E[X^-]}{\sqrt{V[X]V[X^-]}} = \frac{E[XX^-] - E[X]^2}{E[X^2] - E[X]^2} \\
&= \frac{E\left[X\left(\alpha(2E[X] - X) + (1-\alpha)Z_1 + \beta(Z_2 - E[X])\right)\right] - E[X]^2}{X^2 - E[X]^2} \\
&= \frac{\alpha(2E[X]^2 - E[X^2]) + (1-\alpha)E[X]^2 - E[X]^2}{E[X^2] - E[X]^2} = -\alpha.
\end{aligned}
$$

The theorem shows that the generated distance matrices are expected to (i) maintain basic characteristics (mean and variance) of a specific initial single objective benchmark instance and (ii) have all the same strength of correlation (potentially with different sign) to the given matrix. Mean and variance of a distance matrix are certainly not the only properties that characterize a TSP instance, however, they are basic properties that are well defined, easily computable, and (potentially) have a strong influence on the behavior of optimization algorithms. The correlation between the generated matrices is covered by the following theorem.

Theorem 2. *For random variables X^+, Y^+ and X^-, Y^- that are generated according to* (5) *and* (6), *respectively, it holds that: (i) $\rho(X^+, Y^+) = \rho(X^-, Y^-) = \alpha^2$ and (ii) $\rho(X^+, X^-) = \rho(X^-, X^+) = -\alpha^2$.*

Proof. Note, that for each random variable different random variables Z_i, $i \in \mathbb{N}^+$ are used. These do not appear explicitly in the proof since $E[Z_i] = E[X]$. Because $E[Z_i - E[X]] = 0$ we omit all terms involving the last summands of (5) and (6), respectively. For readability we abbreviate $2E[X] - X$ with $\hat{X}$. Due to space limitations cases $\rho(X^+, Y^+)$ and $\rho(X^+, Y^+)$ are omitted. The following holds

$$\rho(X^+, X^-) = \frac{E[X^+X^-] - E[X^+]E[X^-]}{\sqrt{V[X^+]V[X^-]}} = \frac{E[X^+X^-] - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{\alpha^2 E[X\hat{X}] + \alpha(1-\alpha)E[X](E[X] + E[\hat{X}]) + (1-\alpha)^2 E[X]^2 - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{\alpha^2(2E[X]^2 - E[X^2]) + 2\alpha(1-\alpha)E[X]^2 + (1-\alpha)^2 E[X]^2 - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{\alpha^2(E[X]^2 - E[X^2])}{E[X^2] - E[X]^2} = -\alpha^2.$$

The theorem shows that for the partition $\{X_1, \dots, X_q\}$ and $\{X_{q+1}, \dots, X_k\}$ it holds that two random variables from the same set are positively correlated and two random variables from different sets are negatively correlated. By choosing $q = k$ the special case of homogeneous correlations, which was advocated in [17], is covered. Clearly, also the iterated application of the procedure as suggested in [12] is possible. The corresponding correlations are as follows.

Theorem 3. *Let* $X^{+0} := X$ *and* X^{+i} *be the result of applying* (5) *on* X^{+i-1}. *Then* $\rho(X, X^{+i}) = \alpha^i$ *holds.*

Proof. Theorem 1 implies $E[X^{+i}] = E[X]$ and $V[X^{+i}] = V[X]$. We set $\gamma_i = (1-\alpha)Z_{2i-1} + \sqrt{2\alpha(1-\alpha)}(Z_{2i} - E[X])$. Then

$$\rho(X, X^i) = \frac{E[XX^i] - E[X]E[X^i]}{\sqrt{V[X]V[X^i]}} = \frac{E[XX^i] - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{E\left[X\left(\alpha(\dots\alpha(\alpha X + \gamma_i) + \gamma_{i-1})\dots + \gamma_1\right)\right] - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{E\left[X(\alpha^i X + \sum_{k=1}^i \alpha^{k-1}\gamma_k)\right] - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{\alpha^i E[X^2] + E[X \sum_{k=1}^i \alpha^{k-1}\gamma_k] - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{\alpha^i E[X^2] + \sum_{k=1}^i \alpha^{k-1} E[X\gamma_k] - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{\alpha^i E[X^2] + \sum_{k=1}^i \alpha^{k-1}(1-\alpha)E[XZ_{2k-1}] - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{\alpha^i E[X^2] + (1-\alpha^i)E[X]^2 - E[X]^2}{E[X^2] - E[X]^2}$$

$$= \frac{\alpha^i E[X^2] - \alpha^i E[X]^2}{E[X^2] - E[X]^2} = \alpha^i.$$

Theorem 3 also holds analogously if (6) is applied in some or all recursive steps (instead of (5)), however, with a sign switch for each such step.
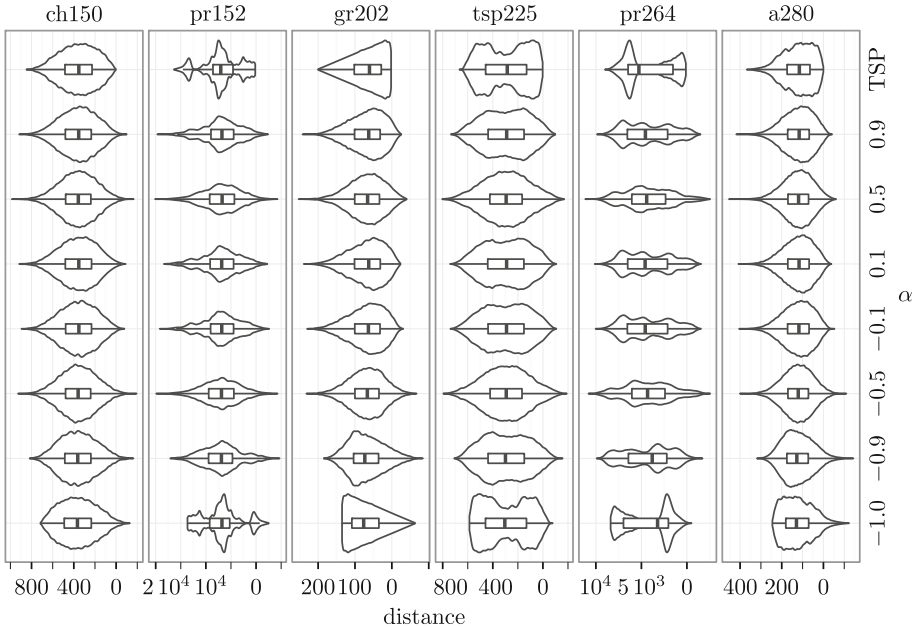
Fig. 1. Violin plot and boxplot of the distance values of six TSPLIB instances (top row) and of the distance matrices that are generated with the new method from these instances with different $\alpha \in \{0.1, 0.5, 0.9\}$ using (3) and (4), the latter indicated by negative values of α.

4   Empirical Analysis

To visualize and complement our theoretical results we performed a series of experiments. For the experiments we used all 85 benchmark instances from the TSPLIB that have at most 7400 cities. A more detailed analysis has been done for the following six of these instances which have been chosen to represent different types of distributions of the distance values that can be found within the instances in the TSPLIB (see topmost row in Fig. 1):

(i) Instances gr202 and a280 represent a negatively skewed distance distribution. For gr202 the number of city pairs decreases almost linearly with an increasing distances, whereas, for a280 this relation is more irregular.

(ii) Instance ch150 represents a distance distribution with a single peak that is similar to a normal distribution (it is only slightly negatively skewed).

(iii) Instances pr152, tsp225, and pr264 represent distance distributions that have more than one pronounced peak. Instance pr264 and pr152 have a non-symmetric distribution with two, respectively three, peaks of different height. Instance tsp225 has two peaks but is more symmetrical and the peaks are less pronounced.
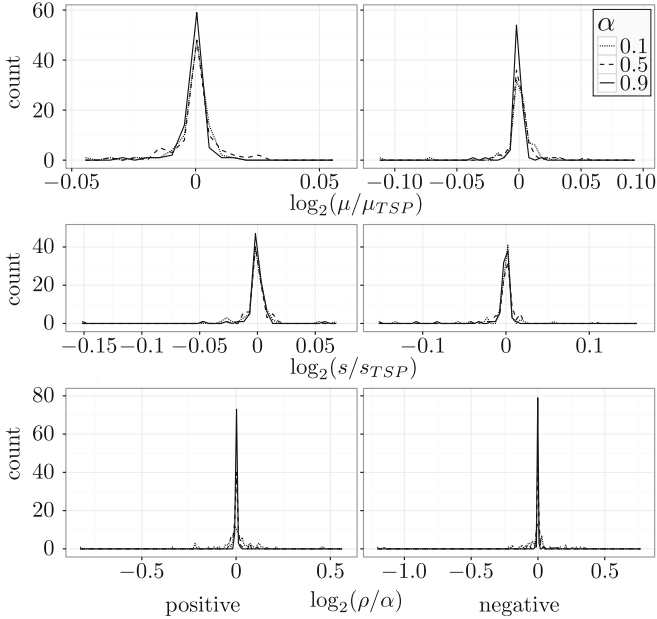
Fig. 2. Distribution of relative mean value (top) and relative standard deviation (middle) of generated matrix with respect to the corresponding value of the original TSPLIB instance over all 85 test instances; also shown is the distribution of the correlation between the generated matrix with the original TSPLIB instance divided by α (bottom for all 85 test instances; shown are the results for the new method; shown are the values for positively (left column) and negatively (right column); $\alpha \in \{0.1, 0.5, 0.9\}$.

Note that for the results that involve a comparison with the results of the methods of [12,17], i.e. (1) and (2), the distance matrices have been normalized.

The result of applying the instance generation method on the selected 85 TSPLIB instances is shown in Fig. 2. For each of the TSPLIB instances and combinations of $\alpha \in \{0.1, 0.5, 0.9\}$ and $q \in \{0, 1\}$ one new matrix was generated, i.e. $k = 1$. The shown distributions of the relative mean values and the relative standard deviations shows that the generated matrices have (with small random fluctuations) the same mean value and the same standard deviation as the corresponding original TSPLIB instances. Note, that the scale of the horizontal axis is logarithmic to base 2 such that 0 corresponds to the fact that the values of the new matrix and the old matrix are equal. The figure also shows that the correlation between the new matrix and the original TSPLIB matrix equals α (with minor random influence). This confirms our theoretical results for the mean, variance, and correlation.

Figure 3 shows the corresponding results when the method of [17], i.e. (2) is applied. The results show a potential disadvantage of this method. The distribution of the distance values in the matrices that is generated with this method is
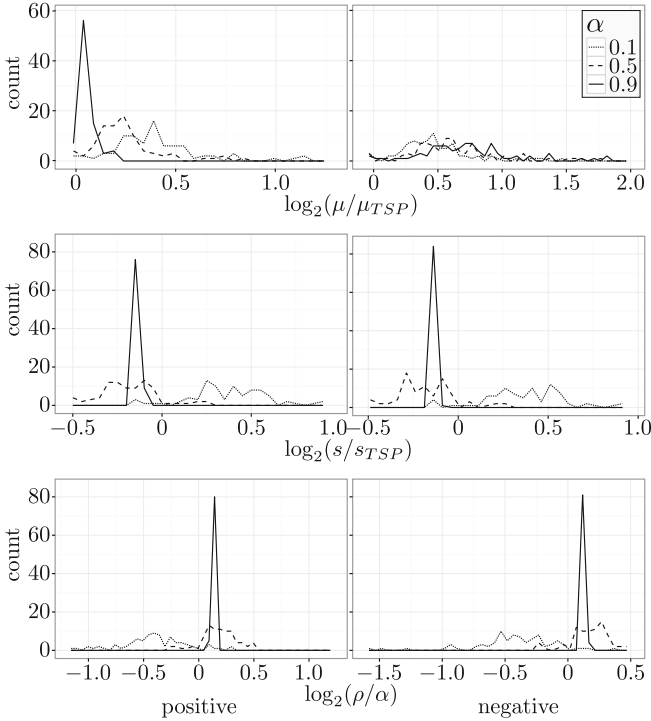
Fig. 3. Distribution of relative mean value (top) and relative standard deviation (middle) relative correlation (bottom) for the 85 test instances from the TSPLIB; shown are the results for the method of [12]; see caption of Fig. 2.

very different compared to the distribution of the values of the original matrices. Their mean values for smaller α approach 0.5 which is the expected value of the random part from (2). Thus, the lower the correlation parameter α is set, the higher the similarity of the resulting distribution to a uniform distribution from 0 to 1. Since most TSPLIB instances have a normalized mean below 0.5 this behavior results in an increase of the mean value for positive correlations and a decrease for negative correlations.

Equation (2) has a notable influence on the standard deviation of the distances in the generated matrices which depends on the values chosen for α. The standard deviation is on average lower for $\alpha \in \{0.5, 0.9\}$ and higher for $\alpha = 0.1$ compared to the original TSPLIB matrix. Moreover, the resulting correlations are virtually never equal to α and they vary strongly, in particular for smaller α. This has been predicted by [17] where it was shown that the observed correlation depends on the variance of the original TSPLIB instance.

As shown, our instance generation method is able to maintain the mean value and variance of the distance distribution of the original matrix. But clearly, other statistical measures might not be preserved by the method. For instance, the
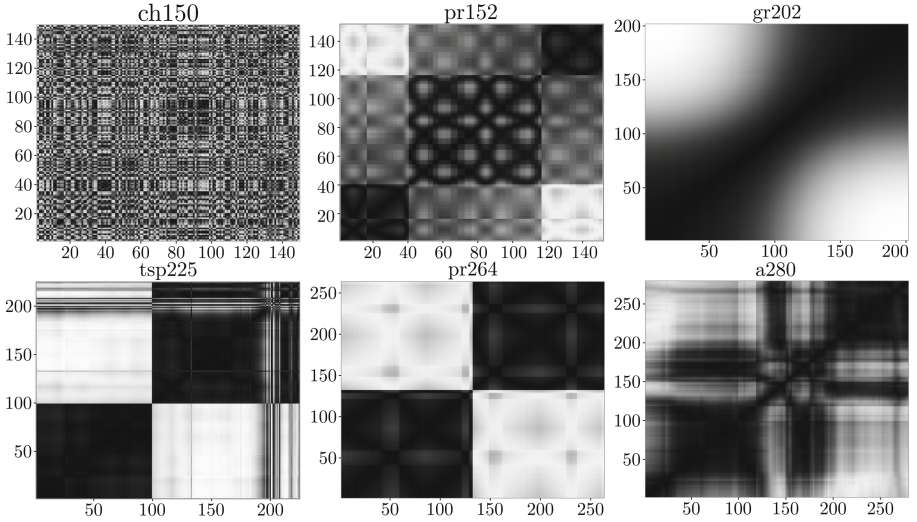
Fig. 4. Correlation $\rho(i, j)$ between all pairs of rows i and j of six TSPLIB distance matrices ; black: $\rho = 1$, white: $\rho = -1$.

median value of the distance distribution might change in the newly generated matrices. This is illustrated in the boxplots in Fig. 1 and can be seen in particular for the asymmetrically clustered instance pr264. With correlation values closer to zero, bulges in the distribution are flattened out. Obviously, this is an effect of the increased random influence on the values in the newly generated matrices. Note, that our instance generation method produces negative distance values, particularly for negative correlations. Since negative distance or cost values could be unrealistic (or impossible) in some application we suggest to deal with them in such cases by a post processing step as described in Sect. 3.

In the following we consider the correlations between the different rows of the distance matrix. More exactly, for an $n \times n$ matrix $M = [m_{ij}]$ the correlation between rows i and j - denoted by $\rho(i, j)$ - is computed over all pairs of values (m_{ih}, m_{jh}) for all $h \in [1 : n] \backslash \{i, j\}$. Note, that the values in row i of the distance matrix are the distances from city i to all other cities. Hence, for a Euclidean TSP instance two cities which are very close to each other have similar distances to all other cities and thus have a high positive correlation between their corresponding rows in the distance matrix. Figure 4 shows the correlations $\rho(i, j)$ for all pairs of rows i and j for each of the six test matrices. Off course, care has to be taken when interpreting the figure because a renumbering of the cities would lead to a reordering of the rows and columns. However, the different characteristics of the six TSPLIB instances can be seen. Instances tsp225, pr264, and pr152 show a clustering of the cities into two (tsp225, pr264) or three (pr152) clusters. Instance gr202 (and to a much lesser extend also instance a280) shows a gradient away from the antidiagonal from positive to negative correlations. Only instance ch150 does not show a clear structure within the correlation plot.
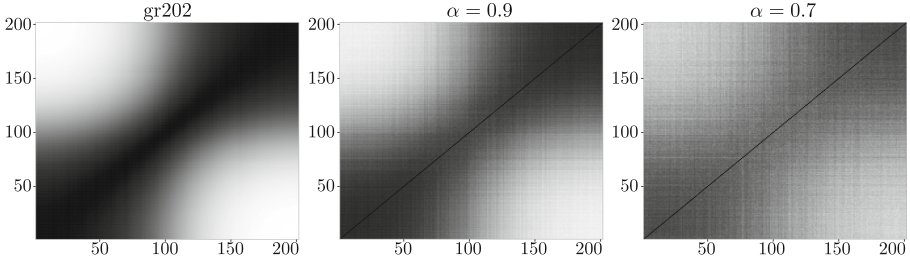
Fig. 5. Correlation $\rho(i,j)$ for all pairs of rows i and j: gr202 (left) and matrices generated from gr202 (middle: $\alpha = 0.9$, right: $\alpha = 0.7$); black: $\rho = 1$, white: $\rho = -1$.
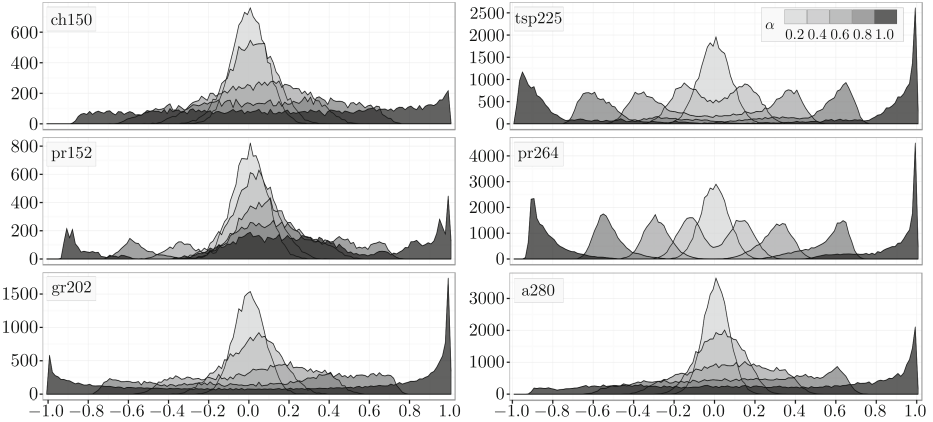

Fig. 6. Distribution of correlations $\rho(i,j)$ between each two rows i and j of matrices generated with $\alpha \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$ for six instances of the TSPLIB; $\alpha = 1.0$ corresponds to the original TSPLIB instance.

In Fig. 5 the correlations between rows in the original TSPLIB matrix and the matrices generated with $\alpha \in \{0.7, 0.9\}$ are shown for gr202. Only results for positively correlated matrices are shown because the results for negative correlation are similar. The general structure of the original matrix is still visible in the generated matrices. Clearly, the random influence introduces considerable noise and reduces the correlation between the rows. This effect is stronger for smaller α.

Indeed, in Fig. 6 it can be seen that the row correlations approach zero for increased random influence, i.e. smaller α. For $0 < \alpha < 1$ the distribution of the row correlations is a mixture of the original TSPLIB instance distribution and the distribution of correlation coefficients of a random sample with correlation zero. The latter distribution was described by Fisher in [3].

Our instance generation method can also provide problem instances where the objectives exhibit different correlations. This can be achieved, for example, by a recursive application of the method. Clearly, each recursive application decreases
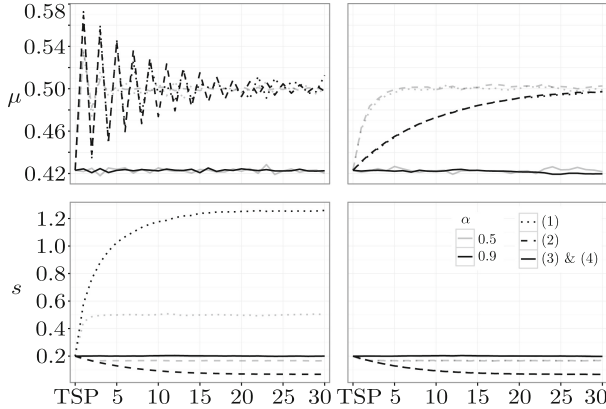
Fig. 7. Mean value and standard deviation (SD) of the distance matrices generated by up to 30 recursive applications of (1), (2), (3) and (4), start matrix is the normalized distance matrix of TSPLIB instance ch150; $\alpha \in \{0.5, 0.9\}$; left: positive correlation; right: negative correlation

the correlation to the original matrix. Other instance generation methods also provide such an option [12, 17]. However, a potential disadvantage of these methods is that they change the mean and standard deviation of the generated matrices and therefore important characteristics of the corresponding optimization problems. This effect becomes stronger the more iterations are applied, such that the derived matrices are very different from the original benchmark matrices. This is shown in Fig. 7 where all three methods (i.e. the method from [12] using (1), the method from [17] using (2), and the newly proposed method using (3) and (4)) are compared. It can be seen that for the methods from [12, 17] the mean value approaches 0.5 and either increases the standard deviation vastly [12] or reduces it [17]. When we compare the results for positive and negative correlations the compared methods behave asymmetrically. In contrast our method maintains the mean value and standard deviation of the original matrix for positively as well as for negatively for correlated matrices (up to small random influences).

5   Conclusion

We have proposed a new method to generate multi-objective optimization problem instances from a corresponding single-objective instance. The method provides the option to choose specific correlations between the objectives, while maintaining important characteristics of the original single-objective instance. In particular, the proposed method can be applied to optimization problems were the objective is defined by a matrix. Examples are the Traveling Salesperson problem (TSP) which uses a distance matrix and the Quadratic Assignment problem which uses a flow matrix. It was shown that the new method provides the option to choose specific correlations between the generated objectives, while

maintaining the mean value and variance of the given single-objective problem instance. This is not guaranteed by existing instance generation methods that have been proposed in the literature. However, maintaining mean value and variance can be important since these characteristics been shown to be of great influence on the difficulty of TSP instances [22]. Our method was analyzed theoretically and it was experimentally compared for the TSP to other instance generation methods from the literature.

It should be noted the application of the proposed instance generation method to problems other than the TSP but where each objective is controlled by a matrix, e.g. the QAP and the multi objective 0/1 Knapsack problem, requires some minor problem specific modifications. For example, the independent generation of both triangular matrices for the flow matrices of the QAP and the generation of all element of the profit matrices of the Knapsack problem.

References

1. Brockhoff, D., Saxena, D., Deb, K., Zitzler, E.: On handling a large number of objectives a posteriori and during optimization. In: Knowles, J., Corne, D., Deb, K., Chair, D. (eds.) Multiobjective Problem Solving from Nature. Natural Computing Series, pp. 377–403 (2008)
2. Corne, D.W., Knowles, J.D.: Techniques for highly multiobjective optimisation: some nondominated points are better than others. In: Genetic and Evolutionary Computation Conference, pp. 773–780 (2007)
3. Fisher, R.A.: Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika **10**, 507–521 (1915)
4. Garrett, D., Dasgupta, D., Vannucci, J., Simien, J.: Applying hybrid multiobjective evolutionary algorithms to the sailor assignment problem. In: Jain, L.C., Palade, V., Srinivasan, D. (eds.) Advances in Evolutionary Computing for System Design. SCI, vol. 66, pp. 269–301. Springer, Heidelberg (2007)
5. Goel, T., Vaidyanathan, R., Haftka, R.T., Shyy, W., Queipo, N.V., Tucker, K.: Response surface approximation of Pareto optimal front in multi-objective optimization. Comput. Methods Appl. Mech. Eng. **196**(4), 879–893 (2007)
6. Ishibuchi, H., Akedo, N., Nojima, Y.: A study on the specification of a scalarizing function in MOEA/D for many-objective Knapsack problems. In: Nicosia, G., Pardalos, P. (eds.) LION 7. LNCS, vol. 7997, pp. 231–246. Springer, Heidelberg (2013)
7. Ishibuchi, H., Akedo, N., Ohyanagi, H., Nojima, Y.: Behavior of EMO algorithms on many-objective optimization problems with correlated objectives. In: 2011 IEEE Congress on Evolutionary Computation (CEC), pp. 1465–1472 (2011)
8. Ishibuchi, H., Yamane, M., Nojima, Y.: Effects of duplicated objectives in many-objective optimization problems on the search behavior of hypervolume-based evolutionary algorithms. In: 2013 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (MCDM), pp. 25–32 (2013)
9. Jaszkiewicz, A.: Genetic local search for multi-objective combinatorial optimization. Eur. J. Oper. Res. **137**(1), 50–71 (2002)
10. Jozefowiez, N., Glover, F., Laguna, M.: Multi-objective meta-heuristics for the traveling salesman problem with profits. J. Math. Modell. Algor. **7**(2), 177–195 (2008)

11. Knowles, J.D., Watson, R.A., Corne, D.W.: Reducing local optima in single-objective problems by multi-objectivization. In: Zitzler, E., Deb, K., Thiele, L., Coello Coello, C.A., Corne, D.W. (eds.) EMO 2001. LNCS, vol. 1993, pp. 268–282. Springer, Heidelberg (2001)

12. Knowles, J.D., Corne, D.W.: Instance generators and test suites for the multiobjective quadratic assignment problem. In: Fonseca, C.M., Fleming, P.J., Zitzler, E., Deb, K., Thiele, L. (eds.) EMO 2003. LNCS, vol. 2632, pp. 295–310. Springer, Heidelberg (2003)

13. Knowles, J.D., Corne, D.: Towards landscape analyses to inform the design of hybrid local search for the multiobjective quadratic assignment problem. HIS **87**, 271–279 (2002)

14. López-Ibáñez, M., Paquete, L., Stützle, T.: On the design of ACO for the biobjective quadratic assignment problem. In: Dorigo, M., Birattari, M., Blum, C., Gambardella, L.M., Mondada, F., Stützle, T. (eds.) ANTS 2004. LNCS, vol. 3172, pp. 214–225. Springer, Heidelberg (2004)

15. Mavrovouniotis, M., Yang, S., Yao, X.: A benchmark generator for dynamic permutation-encoded problems. In: Coello, C.A.C., Cutello, V., Deb, K., Forrest, S., Nicosia, G., Pavone, M. (eds.) PPSN 2012, Part II. LNCS, vol. 7492, pp. 508–517. Springer, Heidelberg (2012)

16. Mersmann, O., Bischl, B., Trautmann, H., Wagner, M., Bossek, J., Neumann, F.: A novel feature-based approach to characterize algorithm performance for the traveling salesperson problem. Ann. Math. Artif. Intell. **69**(2), 151–182 (2013)

17. Moritz, R.L.V., Reich, E., Bernt, M., Middendorf, M.: The influence of correlated objectives on different types of P-ACO algorithms. In: Blum, C., Ochoa, G. (eds.) EvoCOP 2014. LNCS, vol. 8600, pp. 230–241. Springer, Heidelberg (2014)

18. Murata, T., Taki, A.: Examination of the performance of objective reduction using correlation-based weighted-sum for many objective knapsack problems. In: 10th International Conference on Hybrid Intelligent Systems (HIS), pp. 175–180 (2010)

19. Nallaperuma, S., Wagner, M., Neumann, F.: Analyzing the effects of instance features and algorithm parameters for max-min ant system and the traveling salesperson problem. Frontiers in Robotics and AI **2**(18) (2015)

20. Paquete, L., Stützle, T.: A study of stochastic local search algorithms for the biobjective QAP with correlated flow matrices. Eur. J. Oper. Res. **169**(3), 943–959 (2006)

21. Reinelt, G.: TSPLIB-A traveling salesman problem library. ORSA J. Comput. **3**(4), 376–384 (1991)

22. Ridge, E., Kudenko, D.: Determining whether a problem characteristic affects heuristic performance. In: Cotta, C., van Hemert, J. (eds.) Recent Advances in Evol. Comp. SCI, vol. 153, pp. 21–35. Springer, Heidelberg (2008)

23. Verel, S., Liefooghe, A., Jourdan, L., Dhaenens, C.: Analyzing the effect of objective correlation on the efficient set of MNK-landscapes. In: Coello, C.A.C. (ed.) LION 2011. LNCS, vol. 6683, pp. 116–130. Springer, Heidelberg (2011)

24. Verel, S., Liefooghe, A., Jourdan, L., Dhaenens, C.: Pareto local optima of multi-objective NK-landscapes with correlated objectives. In: Merz, P., Hao, J.-K. (eds.) EvoCOP 2011. LNCS, vol. 6622, pp. 226–237. Springer, Heidelberg (2011)

25. Xu, Y., Qu, R., Li, R.: A simulated annealing based genetic local search algorithm for multi-objective multicast routing problems. Ann. Oper. Res. **206**(1), 527–555 (2013)