A hybrid approach of optimization and sampling for robust portfolio selection

Omar Rifki and Hirotaka Ono
Department of Economic Engineering
Kyushu University, Fukuoka, Japan
omar.rifki.910@s.kyushu-u.ac.jp, hirotaka@econ.kyushu-u.ac.jp

***Abstract*—Dealing with ill-defined problems, where the actual values of input parameters are unknown or not directly measurable, is generally not an easy task. In this paper, we propose a hybrid metaheuristic approach, incorporating a sampling-based simulation module, in order to enhance the robustness of the final solutions. Empirical application to the classical mean-variance portfolio optimization problem, which is known to be extremely sensitive to noises in asset means, is provided through a genetic algorithm solver. Results of the proposed approach are compared with that specified by the worst-case scenario.**

***Keywords*-robustness; simulation model; hybridization; evolutionary algorithm; portfolio optimization;**

I. INTRODUCTION

Metaheuristics (MH) are a general class of approximate algorithms, particularly useful for solving difficult optimization problems. Usually simpler to implement compared with gradient-based techniques, they selectively guide the search in succeeding iterations to produce high quality solution(s). The prominent classical families of metaheuristics, that have been used through the years, are tabu search, evolutionary algorithms such as genetic algorithms and evolutionary strategies, simulated annealing, iterated local search and ant colony optimization among others. However, in recent years metaheuristic algorithms combining algorithmic ideas from diverse metaheuristics and a variety of different disciplines like computer science, operations research and artificial intelligence, have been widely and frequently reported. Such algorithms, so-called hybrid metaheuristics, do not restrict their attention solely to the classical metaheuristic families, but extend the scope of its applicability to wide algorithmic areas. The goal is typically to make meaningful improvements either in solution quality or in terms of running time. According to the famous no free lunch theorem (NFL), roughly stated as follows "averaged over the space of all the possible problems all search algorithms perform equally", no single algorithm will outperform random search in average. That is to say, a standard metaheuristic will eventually gain in performance when combined with efficient heuristics and handy problem-specific knowledge. Thereby, hybridization could be a promising tool in order to improve the relative efficiency of the MH in hand for the target optimization problem. The following figure[1] depicts this view for unmixed, hybrid MHs and problem-oriented methods. The last ones have the best efficiency since they are specifically designed for the purpose of solving the target problem. The first ones are instead quite monotonous involving little problem-specific knowledge.

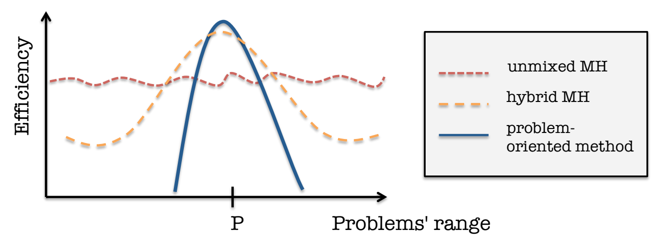


Figure 1. Current view of hybrid MHs

The main focus of MH hybridization process is basically turned to finding efficient good approximate solutions. However, finding robust solutions, which are less sensitive to small changes in the problem variables, can be highly important for effective reasons. In fact, from a practical point of view, solutions of real-world optimization problems are expected to be, not only optimal, but also insensitive to small changes that affect problem variables, whether endogenous or exogenous. Otherwise, a sensitive solution may not be attainable in practice, due to the difficulty of meeting the theoretical assumptions. In this paper, we propose a hybrid metaheuristic approach for solving optimization problems where some exogenous parameters are unknown or unkownable. The intent of hybridization is to enhance the robustness of the final solutions, i.e. the design variables, to changes in those parameters. In other terms, solutions that do not exhibit larger departures when slight changes affect exogenous parameters will be favoured.

The conception of our hybrid MH embeds two phases, in addition to the MH itself, i.e. the algorithmic part, a simulation procedure is performed. Theoretically, any MH, independently of its structure, e.g. population-based or not, related to local search or not, can be utilized within the

[1]This figure is redrawn from [1] and [2] who addressed the case of evolutionary algorithms.

proposed approach. The combination between the MH and the simulation is actually of high level nature, allowing each part to retain its own identity. The aim of the simulation procedure is to identify quality robust solutions among a set of solutions. The overall process comprises, first heuristically solving the problem for several instances of the uncertain parameters, thereafter, evaluating the solutions' performance under a large enough number of samples of the uncertain parameters, slightly and randomly derived from a nominal case. Percentage of being top-ranked solution across the scenarios, plus the across-scenario average of *performance ratio*, i.e. ratio between the solution evaluation and the top-ranked evaluation for the related scenario, are taken as measures of robustness.

The simulation model is described in the next section, while the process of hybridization is presented in Section III. A practical application involving a hybrid genetic algorithm to the financial problem of mean-variance portfolio optimization is provided in Section IV. Section V concludes this study.

II. Simulation Module

This section briefly introduces the notation used in the paper, and subsequently describes the simulation module to be integrated later in the hybrid MH. Special emphasis is given to finding a minimum sample size to guarantee reliable estimates, in subsection B.

A. Notation and model specification

Consider the following constrained optimization problem,

$$\max_x f(\tilde{\theta}, x) \quad \text{subject to} \quad x \in \mathcal{C} \subseteq \mathbb{R}^n,\ \tilde{\theta} \in \Delta \subseteq \mathbb{R}^l, \quad (1)$$

where x denotes a vector of design variables constrained to a set $\mathcal{C} \subseteq \mathbb{R}^n$ and $\tilde{\theta} \in \Delta \subseteq \mathbb{R}^l$ is a vector of random variables representing uncertain exogenous parameters, while $f(\tilde{\theta}, x) : \Delta \times \mathcal{C} \to \mathbb{R}$ is an objective function assumed to be scalar-valued. This problem is not well-defined, since it involves the random parameters $\tilde{\theta}$ that lead to an ambiguous function $f(\tilde{\theta}, x)$. We make it well-defined by assuming that the parameters $\tilde{\theta}$ vary within an uncertainty set Θ, defined as a norm-bounded ball with a center θ_0,

$$\Theta \equiv \{\theta \in \Delta : ||\theta - \theta_0||_p \leq \xi\} = B_\xi(\theta_0) \subseteq \mathbb{R}^l,$$

where $||.||_p$ denotes the l_p norm in $\mathbb{R}^l$, e.g. $p = 1, 2, \infty$. We shall call θ_0 the *nominal value* of $\tilde{\theta}$.

The focus of our simulation is to detect optimal design variables, that have desirable robustness properties within the uncertainty set Θ. To this purpose, given a set of optimal solutions $\mathcal{X} = \{x_1, x_2, x_3, ..., x_k\}$, we proceed by studying their behaviour in terms of performance throughout a number of randomly generated samples of $\tilde{\theta} \in \Theta$, that we shall call *scenarios*. N independent and identically distributed (i.i.d.) scenarios of $\tilde{\theta}$ are drawn, $\theta_1, \theta_2, ..., \theta_N \in \Theta$. To measure robustness of an instance $x \in \mathcal{X}$, we have chosen two measures $R_1^{\mathcal{X}}(x)$ and $R_2^{\mathcal{X}}(x)$. The first one reports the percentage across scenarios of being *top-ranked solution*, i.e. having the best f value compared with the other solutions, for the corresponding scenario. Using the following indicator function,

$$I_{\mathcal{X}}(\theta, x) = \begin{cases} 1, & \text{if } f(\theta, x) = \max_{y \in \mathcal{X}} f(\theta, y) \\ 0, & \text{otherwise} \end{cases}$$

$$\forall \theta \in \{\theta_1, \theta_2, ..., \theta_N\},\ \forall x \in \mathcal{X},$$

we can obtain an estimate of $R_1^{\mathcal{X}}(x)$,

$$\widehat{R_1^{\mathcal{X}}}(x) = \frac{1}{N} \sum_{i=1}^{N} I_{\mathcal{X}}(\theta_i, x) \quad \forall x \in \mathcal{X}, \quad (2)$$

which is equivalent to,

$$\widehat{R_1^{\mathcal{X}}}(x) = \frac{M_{\mathcal{X}}(x)}{N} \quad \forall x \in \mathcal{X}, \quad (3)$$

with $M_{\mathcal{X}}(x)$ is the number of scenarios θ, such that $x \in \arg\max_{y \in \mathcal{X}} f(\theta, y)$. The second measure $R_2^{\mathcal{X}}(x)$ starts by computing for each scenario θ_i the ratio: $f(\theta_i, x)/f_{max}^{\mathcal{X}}(\theta_i)$. $f(\theta_i, x)$ is the evaluation of x correspondingly to θ_i, and $f_{max}^{\mathcal{X}}(\theta_i) = \max_{y \in \mathcal{X}} f(\theta_i, y)$ is the maximum function value reached for θ_i. Afterwards, these ratios are averaged over all scenarios. An estimate can be written as,

$$\widehat{R_2^{\mathcal{X}}}(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{f(\theta_i, x)}{f_{max}^{\mathcal{X}}(\theta_i)}.$$

$f_{max}^{\mathcal{X}}(\theta_i)$ can be also expressed as,

$$f_{max}^{\mathcal{X}}(\theta_i) = \sum_{y \in \mathcal{X}} I_{\mathcal{X}}(\theta_i, y) f(\theta_i, y).$$

The reason of using the first measure is intuitive. It relates the robustness of a solution to the number of times it is optimum (according to $\mathcal{X}$) for a number of random samples $\tilde{\theta}$. The advantage of the second measure is however to take into account the relative span of each solution x to the top-ranked solution across scenarios. Indeed this ratio may be desirable to know if a solution is rarely ranked best, but frequently exhibits a small gap to the top-ranked solutions.

	$\mathcal{X}$				
	x_1	x_2	...	x_k	
(I_1) scenario θ_1	$f(\theta_1, x_1)$	$f(\theta_1, x_2)$	...	$f(\theta_1, x_k)$	$\to f_{max}^{\mathcal{X}}(\theta_1)$
(I_2) scenario θ_2	$f(\theta_2, x_1)$	$f(\theta_2, x_2)$	...	$f(\theta_2, x_k)$	$\to f_{max}^{\mathcal{X}}(\theta_2)$
...	...	...	...	...	
(I_N) scenario θ_N	$f(\theta_N, x_1)$	$f(\theta_N, x_2)$	...	$f(\theta_N, x_k)$	$\to f_{max}^{\mathcal{X}}(\theta_N)$
$R_1^{\mathcal{X}}(x)$	-	-	...	-	
$R_2^{\mathcal{X}}(x)$	-	-	...	-	

Table I. Description of the simulation module

Table I illustrates the overall simulation module. In performing the simulation, three elements need to be set:

- $\mathcal{X}$, the set of optimal solutions to be compared. This set will be generated according to the hybridization process described in Section 3,
- N, the number of randomly generated samples of $\tilde{\theta}$, which will be discussed in the following section,
- and the distribution of the random draw of $\tilde{\theta} \in \Theta$. This includes the shape, e.g. Gaussian, uniform, etc, and the magnitude, i.e. the parameter ξ of $\Theta = B_\xi(\theta_0)$, of the noise distribution. The related choice will be mentioned in the application example.

B. Minimum sample size

The estimate (3) of $R_1^{\mathcal{X}}$, formally known as *relative frequency*, is empirically derived from the actual data, i.e. the used scenarios. A crucial question is to know how many samples of scenarios have to be drawn in our simulation in order to obtain reliable estimates of $R_1^{\mathcal{X}}$ with a high probability. In other words, we are seeking a minimum number of N such that the probability value of the difference $|\widehat{R_1^{\mathcal{X}}}(x) - R_1^{\mathcal{X}}(x)| \leq \epsilon$ with $\epsilon \in (0,1)$, is sufficiently high. Given a *margin error* $\epsilon \in (0,1)$ and a *confidence level* $\delta \in (0,1)$, we will make use of this inequality,

$$\mathbb{P}(|\hat{p}_{\mathcal{X}}(x) - p_{\mathcal{X}}(x)| \leq \epsilon) \geq 1 - \delta, \tag{4}$$

with $\hat{p}_{\mathcal{X}}(x) = \widehat{R_1^{\mathcal{X}}}(x)$ and $p_{\mathcal{X}}(x) = R_1^{\mathcal{X}}(x)$. Since for all $i \in [|1,N|]$ the θ_i's are drawn independently, the event whether a solution x is top-ranked solution or not for a scenario θ_i, which can be expressed as $I_{\mathcal{X}}(\theta_i, x)$, is viewed as an independent Bernoulli trial. Thus, the estimated probability of the global process, which is a binomial process, i.e. sum of independent Bernoulli trials, can be expressed in accordance with (2) as $\hat{p}_{\mathcal{X}}(x) = \frac{1}{N}\sum_{i=1}^{N} I_{\mathcal{X}}(\theta_i, x)$. The expectation and the variance of the general process are,

$$\mathbb{E}(\hat{p}_{\mathcal{X}}(x)) = \frac{\mathbb{E}(\sum_{i=1}^{N} I_{\mathcal{X}}(\theta_i, x))}{N} = p_{\mathcal{X}}(x),$$

$$Var(\hat{p}_{\mathcal{X}}(x)) = \frac{Var(\sum_{i=1}^{N} I_{\mathcal{X}}(\theta_i, x))}{N^2} = \frac{\sigma^2_{\mathcal{X}}(x)}{N},$$

where $\sigma^2_{\mathcal{X}}(x)$ is the variance of the parameter $I_{\mathcal{X}}(\theta_i, x)$. Note that, $p_{\mathcal{X}}(x)$ is actually the probability of success in each trial $I_{\mathcal{X}}(\theta_i, x)$. By directly applying Chebyshev's inequality, which is formally stated as follows, given a random variable Y,

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq \epsilon) \leq \frac{Var(Y)}{\epsilon^2}, \ \forall \epsilon \in (0,1),$$

we obtain the following inequality,

$$\mathbb{P}(|\hat{p}_{\mathcal{X}}(x) - p_{\mathcal{X}}(x)| \leq \epsilon) \geq 1 - \frac{\sigma^2_{\mathcal{X}}(x)}{N\ \epsilon^2} \geq 1 - \frac{p_{\mathcal{X}}(x)\ (1 - p_{\mathcal{X}}(x))}{N\ \epsilon^2}.$$

The corresponding confidence level is then,

$$\delta = \frac{p_{\mathcal{X}}(x)\ (1 - p_{\mathcal{X}}(x))}{N\ \epsilon^2}.$$

Taking into account, $p_{\mathcal{X}}(x)\ (1 - p_{\mathcal{X}}(x)) \leq \frac{1}{4}$, the sample size is bounded, as follows,

$$N \geq \frac{1}{4\epsilon^2\delta}. \tag{5}$$

This bound is known as *Bernoulli bound*. Its expression is actually given in the well-known Bernoulli's theorem of weak law of large number. A tighter bound derived from Hoeffding's inequality[2], called *Chernoff bound*, improves the minimum sample size. Hoeffding's inequality in our case is written as,

$$\mathbb{P}(\Big|\frac{1}{N}\sum_{i=1}^{N} I_{\mathcal{X}}(\theta_i, x) - \mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} I_{\mathcal{X}}(\theta_i, x))\Big| \leq \epsilon) \geq 1 - 2e^{-2\epsilon^2 N}.$$

The confidence level here is,

$$\delta = 2\exp(-2\epsilon^2 N),$$

which subsequently leads to the expression of the Chernoff bound,

$$N \geq \frac{ln(2/\delta)}{2\epsilon^2}. \tag{6}$$

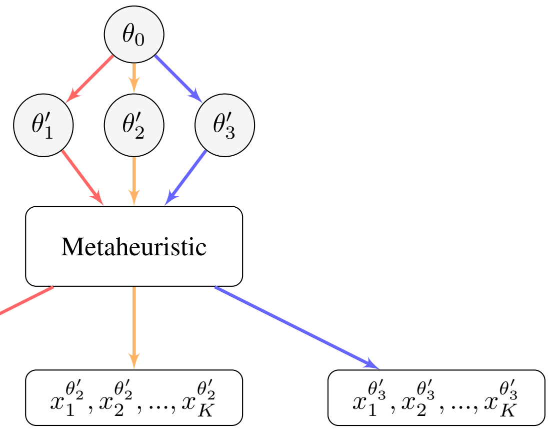
The bound (5) is specified for any random variable, while the bound (6) requires a sum of random variables. The main disadvantage of (5) is that, it requires a large minimum sample size. For instance, Bernoulli and Chernoff bound for $\epsilon = \delta = 0.1\%$ are respectively $N = 2.5\ 10^8$ and $N \approx 3.80\ 10^6$. Tight and rigorous bounds as Chernoff bound are more desirable, since they allow for exact approximation of the minimum value of N necessary for reliable estimates.

III. Hybridization

Hybrid MHs are classified according to several aspects. The *level of hybridization*: deep combination or high level cooperation, the *order of execution*: batch, interleaved or parallel and the *control strategy*: integrative or collaborative, are among others classification criteria of heuristic systems[3]. Control strategy concerns the nature of relationship between the hybrid MH parts. It is integrative if one part is subordinate to the other, as in memetic algorithms where the local search improvement is embedded into genetic algorithm procedures. In the collaborative way however, there is an exchange of information in a synergetic manner without any subordination. Our approach falls into the second class. On one side the simulation technique presented in the previous section, and separately on the other side, the MH algorithm. Before running the simulation, the MH will be performed several times for several instances of the uncertain parameters. Let V denote the number of those runs. For

[2]The formal proof of this inequality can be found in [3].

[3]This taxonomy of classification is mentioned with respect to [4].



	$x_1^{\theta'_1}$	...	$x_K^{\theta'_1}$	$x_1^{\theta'_2}$	...	$x_K^{\theta'_2}$	$x_1^{\theta'_3}$	...	$x_K^{\theta'_3}$
(I_1) scenario θ_1	$f(\theta_1, x_1^{\theta'_1})$	...	$f(\theta_1, x_K^{\theta'_1})$	$f(\theta_1, x_1^{\theta'_2})$	...	$f(\theta_1, x_K^{\theta'_2})$	$f(\theta_1, x_1^{\theta'_3})$	...	$f(\theta_1, x_K^{\theta'_3})$
(I_1) scenario θ_2	$f(\theta_2, x_1^{\theta'_1})$	...	$f(\theta_2, x_K^{\theta'_1})$	$f(\theta_2, x_1^{\theta'_2})$	...	$f(\theta_2, x_K^{\theta'_2})$	$f(\theta_2, x_1^{\theta'_3})$	...	$f(\theta_2, x_K^{\theta'_3})$
...	...	...	...	...	...	...	...	...	...
(I_N) scenario θ_N	$f(\theta_N, x_1^{\theta'_1})$	...	$f(\theta_N, x_K^{\theta'_1})$	$f(\theta_N, x_1^{\theta'_2})$	...	$f(\theta_N, x_K^{\theta'_2})$	$f(\theta_N, x_1^{\theta'_3})$	...	$f(\theta_N, x_K^{\theta'_3})$
$R_1^{\mathcal{X}}(x)$	-	...	-	-	...	-	-	...	-
$R_2^{\mathcal{X}}(x)$	-	...	-	-	...	-	-	...	-

Choosing the K best solutions $x_1, x_2, ..., x_K \in \mathcal{X}$ according to $R_1^{\mathcal{X}}$ or $R_2^{\mathcal{X}}$

Figure 2. General scheme of the proposed hybrid algorithm

example, in Figure 2 $V = 3$, the MH is run three times for three randomly generated variables $\theta'_1, \theta'_2, \theta'_3 \in \Theta$ with θ_0 the nominal value. The overall solutions outputted by the MH runs are grouped to be considered as inputs for the simulation module. The set of input solutions $\mathcal{X}$ considered here according to Figure 2 is,

$$\mathcal{X} = \{x_1^{\theta'_1}, x_2^{\theta'_1}, ..., x_K^{\theta'_1}, x_1^{\theta'_2}, x_2^{\theta'_2}, ..., x_K^{\theta'_2}, x_1^{\theta'_3}, x_2^{\theta'_3}, ..., x_K^{\theta'_3}\}.$$

If we suppose that the MH generates $K \geq 1$ solutions, the final output of the hybrid MH will be also of size K, more specifically the K best solutions sorted according to $R_1^{\mathcal{X}}$ or $R_2^{\mathcal{X}}$. Notice that, it is obviously possible to consider the nominal value θ_0 side by side with $\theta'_1, \theta'_2, \theta'_3$, and the simulation can be computationally costly if the number of required scenarios is quite large. Each new scenario adds at least a cost of $V \times K$ evaluations. Thus, finding a mechanism that may reduce the number of sampled scenarios is of high interest, even if it is not considered in this paper.

Facing optimization under uncertainty, one well-applied approach is the worst-case analysis, known also as max-min, and extensively used in robust optimization. It is based on optimizing the original problem under the worst realization of the uncertain parameters contained within a specially constructed set, called the *uncertainty set*. This approach is preferred to other optimization approaches under uncertainty, as dynamic and stochastic methodologies, since it is more tractable computationally, i.e. the optimization is performed only one time. We shall compare our simulation result: solutions ranked according to $R_1^{\mathcal{X}}$ or $R_2^{\mathcal{X}}$, to the results indicated by the the worst-case scenario θ_{wc} among the set $\{\theta_1, \theta_2, ..., \theta_N\}$ seen here as an uncertainty set. To spot θ_{wc}, we seek given the set of the θ_i's, a lower value of the following expression[4],

$$\sum_{x \in \mathcal{X}} f(x, \theta_i), \ \ \forall i \in [|1, N|]. \tag{7}$$

This comparison will allow us to evaluate further the differences between the results of our approach and those suggested by the worst-case view.

IV. Application

In this section, we apply our hybrid metaheuristic construction to the problem of portfolio optimization. Originally proposed by Markowitz in 1952 [5], the problem of finding the best financial investment strategy can be pinned down mathematically, according to Markowitz' model, as an optimization problem based on two criteria, minimizing the risk while maximizing the expected return of an investment which consists of several financial assets, e.g. stock securities, bonds, real estate investment, options, etc. It is called the mean-variance model (MV), as the

[4]If our problem is a minimization problem, we will seek a higher value.

investment risk is measured by variances and covariances of asset returns. The model considered in this section is the standard single-period MV problem, stated for n risky assets as,

$$\begin{aligned}
\underset{w}{\text{maximize}}\ \{EU(w)\} &= \{\mu' w - \lambda w' \Sigma w\} \qquad (8)\\
&= \{\sum_{i=1}^{n} w_i \mu_i - \lambda \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij}\}\\
\text{subject to}\ \sum_{i=1}^{n} w_i &= 1, \quad \text{and} \quad \forall i \in [|1, n|],\ 0 \le w_i \le 1
\end{aligned}$$

where $w \in \mathbb{R}^n$ is the n-vector of weights corresponding to the decision variable, each w_i is the fraction held in the i-th asset. The expected return (mean) of the i-th asset and the covariance between the return of the i-th and j-th assets are respectively denoted by μ_i and σ_{ij}, such that $\sigma_{ii} = \sigma_i^2$ is the variance of the i-th asset. Accordingly, $\mu = (\mu_i)_i$ and $\Sigma = (\sigma_{ij})_{i,j}$ are the n-vector of means and the $n \times n$ covariance matrix in the same order. The objective function is designated by the expected utility function $EU(w)$, and the parameter λ indicates the degree of the investor's risk aversion. The higher λ the more the investor is risk-averse. Nevertheless, the MV analysis suffers from the drawback of extreme sensitivity to input's errors, especially concerning asset means μ. This is an issue because the model inputs are predominantly based on statistical estimations from historical data, which may induce potential and significant errors in asset means and return covariances. Actually, this sensitivity issue is considered a major barrier for POP to be effective in real-life situations. Several studies have addressed this problem: Jobson and Korkie [6], Michaud [7], Best and Grauer [8], Chopra and Ziemba [9] and Broadie [10].

A. Genetic Algorithm solver

Concerning the MH choice, we apply Genetic Algorithm (GA), which is a well-known population-based metaheuristic grounded on the darwinian natural selection principle – *survival of the fittest*. In GA, the population of individuals evolves through three main operators to fit a goal formulated by the fitness function, a function that evaluates the accuracy of the evolved individuals to the problem in hand. The first operator is *selection*, which selects parent individuals intended to generate the next generation via two evolution operators which are *crossover* that merges features of the parents and *mutation* that slightly changes them. GA is typically applied to global optimization problems, especially solving complex nonlinear problems where exact solutions are difficult to obtain. The choice of GA, in the current application, is due to its superior performance reported for solving MV models in comparison with other MHs. For instance, Chang *et al.* [11] compared three MHs, namely Tabu search, Simulated Annealing and GA, for solving a cardinality constrained MV optimization problem. They found that in the unconstrained case, i.e. without cardinality constraints, which corresponds to our target problem (8), GA gives the best approximate solutions with an almost zero mean percentage error. For more information about GA based applications to portfolio optimization, the reader is referred to the surveys [12] [13] [14] which cover single and multi-objective GA-based applications.

Since the optimization problem (8) is continuous, we adopt a real-valued representation rather than the original binary-string representation that is well suited for discrete problems. Simulated binary crossover (SBX) and polynomial mutation (PM) are equally employed. These two operators, proposed respectively by Deb and Agrawal [15] and Deb [16] are specifically designed for real-valued encodings. SBX is the real domain analog of single-crossover operator of binary GA in terms of search power. It uses for offspring production a probability distribution similar to that of binary single-crossover. PM, on the other hand, is a nonuniform mutation where the underlying distribution is similar to that of SBX, i.e, polynomial. Both operators have been shown to yield improved results over real-valued encodings, according to [15] and [16]. For the selection procedure, we used a binary tournament operator with elitism. The binary tournament compares the fitness of a set of randomly chosen individuals, it retains after the best ranked solution, i.e. the one with the highest fitness. This process is repeated until the mating pool, from which the next generation will be drawn, is complete. The elitism is applied by keeping the two fittest solutions of the population for the next generation. The following table summarizes the GA operators and the parameter values used in the current application.

GA Parameter	Value
Population size	100
Generations	300
Selection	tournament (size = 2)
Crossover	SBX
Mutation	polynomial
Crossover probability	0.25
Mutation probability	0.01

Table II. GA parameters

B. Data

We use the following dataset, taken from Michaud [17, p. 16]. It describes monthly sampled expected returns and covariances for six developed countries. The data is given over a period of 18 years ($T = 216$), from January 1978 to December 1995.

		Asset means	Standard deviation
1	Canada Equity	.39	5.50
2	France Equity	.88	7.029
3	Germany Equity	.53	6.220
4	Japan Equity	.88	7.039
5	U.K. Equity	.79	6.010
6	U.S. Equity	.71	4.300
7	U.S. Bonds	.25	2.010
8	Europe Bonds	.27	1.558

	Correlation matrix							
	1	2	3	4	5	6	7	8
1	1							
2	.4099	1						
3	.2999	.6199	1					
4	.2500	.4200	.3500	1				
5	.5799	.5401	.4798	.3999	1			
6	.7099	.4399	.3402	.2200	.5599	1		
7	.2596	.2200	.2703	.1399	.2500	.3598	1	
8	.3300	.2600	.2805	.1603	.2903	..4207	.9191	1

Table III. Historical means and return correlation matrix

C. Methodology and results

We focus on errors in asset means, since their impact can be 10 times or more important than those in variances and covariances, in terms of cash equivalent loss [9]. Consequently, the uncertain parameter here is the n-vector of asset means $\tilde{\theta} = \tilde{\mu}$. By taking a margin error $\epsilon = 0.5\%$ and a confidence level $\sigma = 0.5\%$, the corresponding Chernoff bound, according to (6), is $1.198\,10^5$. For the entire experiment, we fix the sample size at $N = 1.5\ 10^5$, and the risk-aversion parameter at $\lambda = 2$. About the scenarios' random draw, we opt for a uniformly iid sampling from the set,

$$\left\{\mu \in \mathbb{R}^n : |\mu_i - \mu_{0\,i}| \leq \xi,\ \forall i \in [|1, n|]\right\} = B_\xi(\mu_0),$$

where μ_0 represents the nominal vector of asset means given by Table III. Two noise magnitudes ξ are considered, a low magnitude $\xi = 0.01$ and a large one $\xi = 0.3$. To investigate the impact of the initial choice of asset means[5], ten instances of μ are considered (from μ_0 to μ_9); nine are randomly sampled according to the same sampling process described above, plus the nominal value μ_0. For each case of ξ, 100 runs of the hybrid GA are performed, the average value are reported thereafter. For each run of the hybrid GA, the following solution set $\mathcal{X}$ is constituted,

$$\mathcal{X} = \{w_i^{\mu_j}, \forall i \in [|1, 100|], \forall j \in [|0, 9|]\},$$

with $w_i^{\mu_j}$ is the portfolio number $1 \leq i \leq K = 100$, generated using the asset mean μ_j, $0 \leq j \leq 9$. K refers to the GA population size. The set $\mathcal{X}$ is slightly changed in our program by removing identical portfolios $w_i^{\mu_j}$ that are found using different μ_j. Only the first occurrence of these duplicates are retained. It is attributed to the initial μ_j that finds them.

[5]The corresponding initial choice of Figure 2 is θ'_1, θ'_2 and θ'_3.

In our setting, we compare three populations induced from the final output, namely the first K portfolios sorted according to $R_1^{\mathcal{X}}$ and to $R_2^{\mathcal{X}}$ measures and to the worst-case scenario (wc), identified by expression (7). Following the results obtained for $\xi = 0.01$.

1) Low noise magnitude: Figure 3 compares the three populations (according to wc, $R_1^{\mathcal{X}}$ and $R_2^{\mathcal{X}}$) averaged over 100 runs of the hybrid GA. The μ_j in the figure's legend represents the instance of asset means used in the portfolios generation. In general, the three populations do not differ much. However, the ones according to the worst-case scenario and $R_1^{\mathcal{X}}$ are quite similar, expect very small differences between slots which does not exceed 0.4% (larger difference for μ_0). On the other hand, the $R_2^{\mathcal{X}}$ population, even similar, allows more space for the nominal portfolios (generated using μ_0), for almost 21.35%. Table IV gives an example of the best ten rankings across the populations, for one run among the 100 used runs. It shows that, in this particular case, portfolios generated using μ_1 are more adapted to the worst-case scenario, which corresponds to the scenarios number $38\,886$ of this instance run. Also, portfolios generated using μ_4 have better $R_2^{\mathcal{X}}$ measure, averagely performing better across scenarios, i.e. higher performance ratio. The first rankings by $R_1^{\mathcal{X}}$ are however more diverse, including portfolios yielded by μ_6, μ_5, μ_4 and μ_1 among others. In summary, while the 10^{th} first rankings of the populations can be completely different, the overall repartition according to wc scenario and $R_1^{\mathcal{X}}$ looks much alike.

	wc population		$R_1^{\mathcal{X}}$ population		$R_2^{\mathcal{X}}$ population	
ranking	portfolio	μ	portfolio	μ	portfolio	μ
1^{st}	$w_1^{\mu_1}$	μ_1	$w_1^{\mu_6}$	μ_6	$w_1^{\mu_4}$	μ_4
2^{nd}	$w_2^{\mu_1}$	μ_1	$w_1^{\mu_5}$	μ_5	$w_2^{\mu_4}$	μ_4
3^{rd}	$w_{15}^{\mu_1}$	μ_1	$w_1^{\mu_4}$	μ_4	$w_{10}^{\mu_4}$	μ_4
4^{th}	$w_{19}^{\mu_1}$	μ_1	$w_1^{\mu_1}$	μ_1	$w_{12}^{\mu_4}$	μ_4
5^{th}	$w_{20}^{\mu_1}$	μ_1	$w_1^{\mu_8}$	μ_8	$w_{13}^{\mu_4}$	μ_4
6^{th}	$w_{21}^{\mu_1}$	μ_1	$w_1^{\mu_0}$	μ_0	$w_{14}^{\mu_4}$	μ_4
7^{th}	$w_{55}^{\mu_1}$	μ_1	$w_1^{\mu_7}$	μ_7	$w_{15}^{\mu_4}$	μ_4
8^{th}	$w_{58}^{\mu_1}$	μ_1	$w_1^{\mu_9}$	μ_9	$w_{18}^{\mu_4}$	μ_4
9^{th}	$w_{59}^{\mu_1}$	μ_1	$w_{95}^{\mu_9}$	μ_9	$w_{19}^{\mu_4}$	μ_4
10^{th}	$w_{78}^{\mu_8}$	μ_8	$w_1^{\mu_3}$	μ_3	$w_{21}^{\mu_4}$	μ_4

Table IV. An instance of best portfolios by population for $\xi = 0.01$

Figure 4 shows the distribution of rankings across 100 runs of the hybrid MH. For each top ranked portfolio, which has a rank 1, of each population (according to wc, $R_1^{\mathcal{X}}$ or $R_2^{\mathcal{X}}$), we examine its analogous ranking in the other wc, $R_1^{\mathcal{X}}$ and $R_2^{\mathcal{X}}$ populations. This is done 100 times (100 runs of hybrid MH) in order to construct the box plots of Figure 4. Rankings

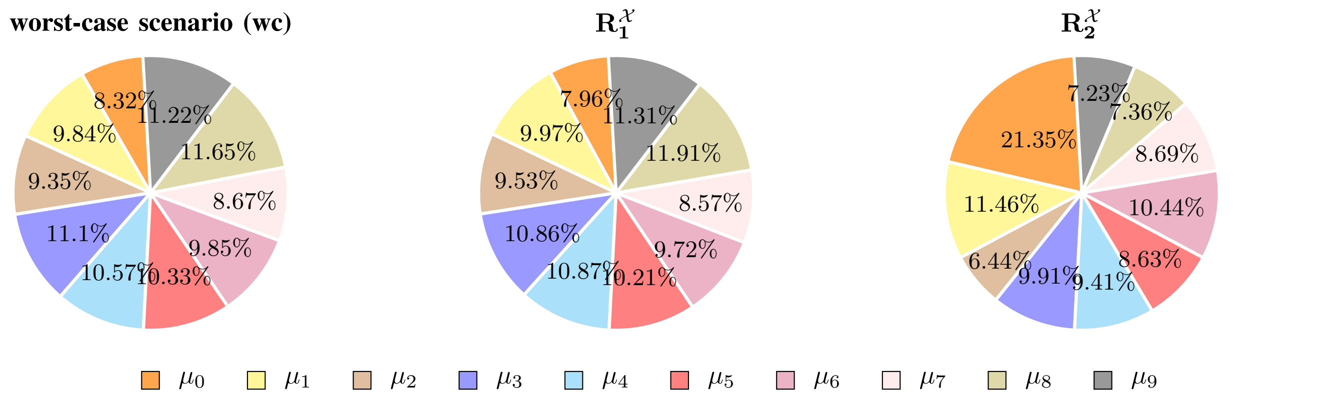


Figure 3. Repartition of final populations, according to wc, $R_1^{\mathcal{X}}$ and $R_2^{\mathcal{X}}$, in terms of asset means instances μ_i (for $\xi = 0.01$)

below the red line designate portfolios that can be part of the final population, i.e. having a ranking less or equal than $K = 100$. From the figure, we observe that $R_1^{\mathcal{X}}$ measure assigns high rankings to the top portfolios according to wc (the first box plot from the left), which suggests a similarity between rankings in wc and $R_1^{\mathcal{X}}$ populations. However, this is not reciprocal, the best portfolios according to $R_1^{\mathcal{X}}$ have spread ranks in wc population (the third box plot). We can similarly notice that $R_1^{\mathcal{X}}$ measure also allocates high ranks to the best $R_2^{\mathcal{X}}$ individuals (the sixth box plot). First ranked portfolios by $R_1^{\mathcal{X}}$ cover large area of rankings in populations of both measures $R_2^{\mathcal{X}}$ and wc, according to the two blue box plots of Figure 4. All the box plot means are above the red line, except the first and sixth ones.

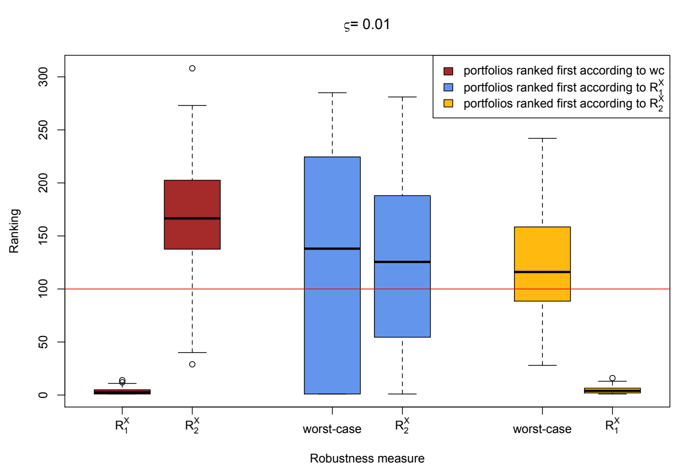


Figure 4. Distribution of the analogous rankings of the best portfolios according to wc, $R_1^{\mathcal{X}}$ and $R_2^{\mathcal{X}}$ (for $\xi = 0.01$)

2) Large noise magnitude: The second part of the experimentation concerns the magnitude of $\xi = 0.3$. Figures 5 and 6 are respectively the equivalent results of Figures 3 and 4. Same observations concerning the rankings distributions of law magnitude can be made here also. The overall populations' repartitions follow likewise, although the populations according to $R_2^{\mathcal{X}}$ gives more room to the nominal portfolios, for around 28.31%. Notice that the nominal asset means μ_0 based portfolios take, according to Figure 5, negligible proportions in wc and $R_1^{\mathcal{X}}$ populations, for around 4.1%. This is consistent with the high risk of having unlimited belief in nominal values of uncertain parameters.

The implementation of the overall hybrid MH, GA and simulation module, is made in the Java environment[6]. The various tests are done on a machine with an Intel Core i7 CPU at 3GHz and 8GB DDR3 RAM, using JDK 1.7.0.

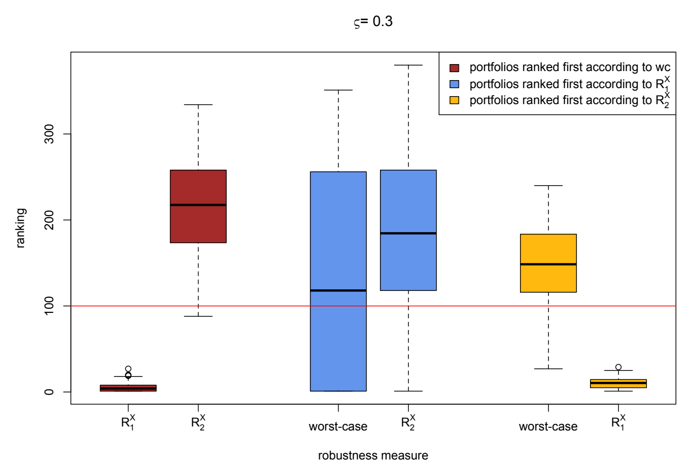


Figure 6. Distribution of the analogous rankings of the best portfolios according to wc, $R_1^{\mathcal{X}}$ and $R_2^{\mathcal{X}}$ (for $\xi = 0.3$)

V. Conclusion

In order to capture the randomness of input parameters for ill-defined problems, we proposed a hybrid metaheuristic (MH) approach incorporating a simulation module. This

[6]The code is available upon request by email.

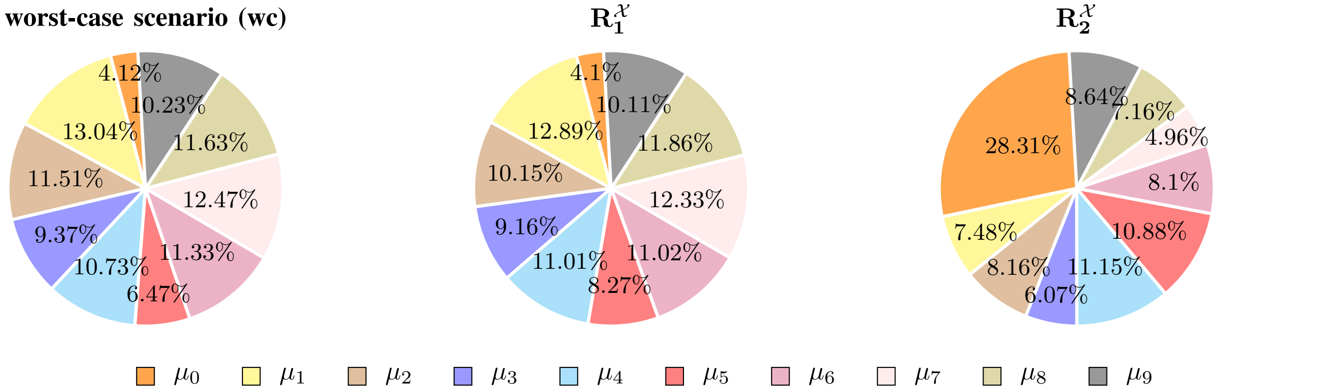


Figure 5. Repartition of final populations, according to wc, $R_1^{\mathcal{X}}$ and $R_2^{\mathcal{X}}$, in terms of asset means instances (for $\xi = 0.3$)

module is created for the purpose of finding less sensitive solutions to perturbations affecting some uncertain input parameters. Two measures of robustness are used within the simulation module :

1) ($\mathbf{R_1^{\mathcal{X}}}$) the percentage of being top-ranked solution throughout randomly sampled scenarios of the uncertain parameters,
2) ($\mathbf{R_2^{\mathcal{X}}}$) the average across scenarios of the solution performance ratio, which corresponds to the solution evaluation over top-ranked evaluation for the related scenario.

Empirical experimentations are conducted for the problem of portfolio optimization against noises in the expected returns of assets, which are inputs of the model. The emphasis of the application is on comparing ranks of the final solutions induced by three criteria: both robustness measures ($R_1^{\mathcal{X}}$ and $R_2^{\mathcal{X}}$) plus ranks implied by the worst-case scenario (wc).

Results have shown, for both cases of high and low magnitude perturbations, that the final population to the $R_1^{\mathcal{X}}$ measure is highly similar to that suggested by the $\mathbf{wc}$ scenario, meanwhile $R_1^{\mathcal{X}}$ population performs better when it comes to ranking the top-ranked portfolios according to the other measures $R_2^{\mathcal{X}}$ and wc.

In future work, we intend to extend the hybrid approach by embedding in it a selection mechanism of the sampled scenarios. Instead of treating scenarios equally, the user can have more control over the process.

REFERENCES

[1] A. Eiben and J. Smith, *Introduction to evolutionary computing*. springer, 2008.

[2] Z. Michalewicz, *Genetic algorithms+ data structures*. Springer, 1996.

[3] K. Sridharan, "A gentle introduction to concentration inequalities," 2009, available: http://ttic.uchicago.edu/ karthik/ concentration.pdf.

[4] G. R. Raidl, "A unified view on hybrid metaheuristics," in *Hybrid Metaheuristics*. Springer, 2006, pp. 1–12.

[5] H. Markowitz, "Portfolio selection," *The journal of finance*, vol. 7, no. 1, pp. 77–91, 1952.

[6] J. D. Jobson and B. Korkie, "Estimation for markowitz efficient portfolios," *Journal of the American Statistical Association*, vol. 75, no. 371, pp. 544–554, 1980.

[7] R. Michaud, "The markowitz optimization enigma: is 'optimized' optimal?" *Financial Analysts Journal*, pp. 31–42, 1989.

[8] M. Best and R. Grauer, "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results," *Review of Financial Studies*, vol. 4, no. 2, pp. 315–342, 1991.

[9] V. Chopra and W. Ziemba, "The effect of errors in means, variances, and covariances on optimal portfolio choice," *The journal of portfolio management*, vol. 19, no. 2, pp. 6–11, 1993.

[10] M. Broadie, "Computing efficient frontiers using estimated parameters," *Annals of Operations Research*, vol. 45, no. 1, pp. 21–58, 1993.

[11] T. Chang, N. Meade, J. Beasley, and Y. Sharaiha, "Heuristics for cardinality constrained portfolio optimisation," *Computers and Operations Research*, vol. 27, no. 13, pp. 1271–1302, 2000.

[12] M. Castillo Tapia and C. Coello, "Applications of multi-objective evolutionary algorithms in economics and finance: A survey," in *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*. IEEE, 2007, pp. 532–539.

[13] F. Schlottmann and D. Seese, "Financial applications of multi-objective evolutionary algorithms: Recent developments and future research directions,"," *Applications of Multi-Objective Evolutionary Algorithms*, pp. 627–652, 2004.

[14] O. Rifki and H. Ono, "A survey of computational approaches to portfolio optimization by genetic algorithms," 2012, available: http://hdl.handle.net/ 2324/ 25317.

[15] R. B. Agrawal, K. Deb, and R. B. Agrawal, "Simulated binary crossover for continuous search space," 1994.

[16] K. Deb *et al.*, *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons Chichester, 2001, vol. 2012.

[17] R. O. Michaud and R. O. Michaud, "Efficient asset management: a practical guide to stock portfolio optimization and asset allocation," *OUP Catalogue*, 2008.