Bilevel Optimization based on Kriging Approximations of Lower Level Optimal Value Function

Ankur Sinha[1], Samish Bedi[2] and Kalyanmoy Deb[3]

[1]Production and Quantitative Methods
Indian Institute of Management, Ahmedabad 380015 India
asinha@iima.ac.in

[2]Department of Computer Science
Birla Institute of Technology and Science
Pilani 333031 India
f2012735@pilani.bits-pilani.ac.in

[3]Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI, USA
kdeb@egr.msu.edu

COIN Report 2018003

***Abstract*— A large number of application problems involve two levels of optimization, where one optimization task is nested inside the other. These problems are known as bilevel optimization problems and have been widely studied by researchers in the area of mathematical optimization. Bilevel optimization problems are known to be difficult and computationally demanding. Most of the solution procedures proposed until now are either computationally very expensive or applicable to only a narrow class of bilevel optimization problems involving small number of variables. In this paper, we propose a global optimization algorithm for bilevel optimization using Kriging approximation based model that tries to reduce the computational expense by iteratively approximating an important mapping in bilevel optimization; namely, the lower level optimal value function mapping. The lower level optimal value function is useful in reducing the two level optimization task to one; however, identifying this function is not straightforward. Our approach aims at meta-modeling this mapping and solving a number of auxiliary single level problems to arrive at the bilevel optimum. In our study, we test the methodology on a number of test problems. The preliminary results are quite promising which suggest the viability of the approach in solving more complicated bilevel test problem. To the best knowledge of the authors, such kind of a solution procedure based on iterative approximation of the optimal lower level value function using a stochastic process has not been widely used in bilevel optimization.**

I. Introduction

Bilevel optimization has recently gained significant interest in the area of optimization because it allows modeling of integrated optimization problems with hierarchy. For instance, consider a toll setting problem consisting of a network of highways operated by an agency, say government, and a large number of users of the network. If the government decides to write its own optimization problem to optimally set the tolls in order to maximize its revenues, it has to integrate the highway users' optimization problem as well. The integration of the network users' optimization problem within the government's optimization problem leads to a hierarchical optimization problem, commonly referred to as bilevel optimization problem. The users' optimization problem attempts to minimize the generalized cost (travel time and travel cost) for the users, which appears as a constraint (nested optimization task) in the government's problem.

The government's problem is commonly referred to as the upper level problem and the users' problem is referred to as the lower level problem. Such problems have been widely studied in the area of transportation policy formulation [29], [11], [36]. In the domain of engineering such integrated optimization problems commonly arise in structural optimization. At the upper level the objective involves cost or weight minimization subject to bounds on state variables like displacements, stresses and contact forces and with the decision variables like shape, material etc. For all possible choices of decision variables, the model might not be stable. The stability is ensured by solving an equilibrium satisfaction problem that appears as a lower level optimization task. Therefore, such integrated optimization problem are ubiquitous and are also known as Stackelberg games [43], whenever it involves two players with sequential decisions and a hierarchy in terms of decision powers. In such cases, we refer to the upper level as the leader's problem and the lower level as a follower's problem. The toll setting problem is a good example of a Stackelberg game. There has been a growing interest in these problems ever since these problems were first realized by the mathematical optimization community in 1973 [10]. Many new applications of such problems have come up in the past few decades, for example, such problems are found in the areas of defense [12], [48], investigation of strategic behavior in deregulated markets [19], model production processes [30], chemical engineering [42], [13], and optimal tax policies [23], [40]. A large body of literature already exists on bilevel optimization, for instance [39], [15], [8].

Bilevel optimization problems are hard to an extent such that mere proving whether a solution is optimal or not is an NP-hard [45] task. A number of attempts have been made to solve simpler versions of these problems by exploiting the mathematical properties, for example, linear [49], [9], [6], [17], [44] and quadratic [7], [16], [3] bilevel problem have been solved using Karush-Kuhn-Tucker (KKT) based reduction of the two level problem into a single level. Such approaches first replace the lower level optimization problem with the KKT constraints resulting in a mixed integer program, which is then solved using branch and bound, and vertex enumeration ideas. Gradient-based techniques have also been employed to determine the descent direction in bilevel problems. However, such descent methods often require solving an auxiliary problem to determine the descent direction [32], [45]. Other ideas are based on penalties [1], [2], [20], [50] or trust-regions [26], [27], [14]. The mentioned methods mostly targeted problems adhering to assumptions like linearity, convexity, differentiability etc. Early 1990s saw the advent of evolutionary algorithms being used for bilevel problems [28]. Many of the evolutionary algorithms proposed during the decade and after were approaches that solved the upper level problem using evolutionary algorithms and the lower level problem using a classical techniques [28], [52], [53]. Later some studies also used evolutionary algorithms at both level, if the lower level problem was difficult to solve using a classical technique [25], [41], [4]. Researchers in the evolutionary community have also used KKT conditions of the lower level bilevel to convert the bilevel task into a single level. For instance, see [18], [46], [47], [21], [24]. However, most of the evolutionary techniques have been computationally extensive, and researchers have also looked into exploiting certain mappings of the bilevel problems to solve it more efficiently [34], [38], [35], [5], [37], [33].

In this paper, we plan to exploit one such mapping of the bilevel problem. The mapping is referred to as the lower level optimal value function mapping. This mapping relates the upper level variables with the corresponding optimal objective function value of the lower level problem. The method discussed in this paper creates a meta-model of the optimal value function mapping using Kriging approximations. This is a preliminary study in which we have tested the idea on a small set of test problems. Given the promising results the aim is to extend this study and show its viability in solving larger bilevel instances.

This paper is organized as follows. In Section II we discuss the bilevel formulations, and then in Section III provide a brief introduction of Kriging technique and its application in modeling the optimal value function mapping. This is followed by a step-by-step procedure for an algorithm that has been used to solve a set of bilevel problems in Section IV. In Section V we provide the computational results on set of problems chosen from the literature. The paper concludes with the conclusions in Section VI. The aim of this study is to primarily test the feasibility of the approach; therefore, we have not provided any comparison results with other approaches. The notations used in this paper have been summarized in Table I.

II. Bilevel Formulation

Bilevel optimization problems contain two levels of optimization task, where one optimization problem is nested within the other. The inner or nested optimization problem is often referred to as the lower level problem and the outer optimization problem is often referred to as the upper level problem. Correspondingly, there are two types of variables; namely, upper and lower level variables. Each level has its own objective function and set of constraints. The lower level optimization problem is a parameterized optimization problem that acts as a constraint to the upper level problem. A feasible solution to the bilevel problem is a vector of upper and lower level variables, such that, the vector satisfies all the constraints in the problem, and the lower level variables are optimal to the lower level problem for the given upper level variables as parameters. It is important to note that in the situations where the lower level problem has multiple optimal solution for a given upper level vector, it is necessary to define, which solution out of the multiple lower level solutions should be considered. There are two positions commonly analyzed in the literature, the *optimistic* and the *pessimistic* positions. In the optimistic position the follower is assumed to cooperate with the leader and choose the solution that is best for the leader; while in the pessimistic position the follower is assumed to not cooperate with the leader and choose the solution that is worst for the leader. There are, of course, intermediate positions possible which can be defined using a selection function. In this paper, we focus on solving optimistic bilevel problems. A general formulation for the bilevel optimization problem is provided below.

Definition 1: For the upper-level objective function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and lower-level objective function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, the bilevel optimization problem is given by

$$\min_{x,y} F(x,y) \tag{1}$$

subject to

$$y \in \operatorname*{argmin}_y \{f(x,y) : g_j(x,y) \leq 0, j = 1,\ldots,J\} \tag{2}$$

$$G_k(x,y) \leq 0, k = 1,\ldots,K \tag{3}$$

where $G_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $k = 1,\ldots,K$ denotes the upper level constraints, and $g_j : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $j = 1,\ldots,J$ represents the lower level constraints, respectively. Variables x and y are n and m dimensional vectors, respectively.

TABLE I

NOTATIONS USED IN THE PAPER.

Category	Notation(s)	Description
Decision vectors	x y	Upper level decision variable. Lower level decision variable.
Objectives	F f	Upper level objective function. Lower level objective function.
Constraints	$G_k,\ k = 1, \ldots, K$ $g_j,\ j = 1, \ldots, J$	Upper level constraint functions. Lower level constraint functions.
Lower level feasible region	Ω	$\Omega(x) = \{y : g_j(x, y) \leq 0\ \forall\ j\}$, represents the lower level feasible region for any given upper level decision vector
Relaxed feasible set	$\Phi = \text{gph}\,\Omega$	$\Phi = \{(x, y) : G_k(x, y) \leq 0\ \forall\ k,\ g_j(x, y) \leq 0\ \forall\ j\}$, represents the region that satisfies both upper and lower level constraints
Lower level reaction set	Ψ	$\Psi(x) = \{y : y \in \underset{y}{\text{argmin}}\{f(x, y) : y \in \Omega(x)\}\}$, represents the lower level optimal solution(s) for an upper level decision vector
Inducible region (Feasible set)	$I = \text{gph}\,\Psi$	$I = \{(x, y) : (x, y) \in \Phi, y \in \Psi(x)\}$, represents the set of upper level decision vectors and corresponding lower level optimal solution(s) belonging to feasible constraint region
Optimal value function	φ	$\varphi(x) = \min_y\{f(x, y) : y \in \Omega(x)\}$ represents the minimum lower level function value corresponding to a given upper level decision vector.

A. *Lower Level Optimal Value Function*

An equivalent definition of the bilevel optimization problem can be given in terms of the lower level optimal value function [51]. We define the lower level optimal value function below before providing the formulation for bilevel optimization.

Definition 2: Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be the lower level optimal value function mapping,

$$\varphi(x) = \min_y\{f(x, y) : g_j(x, y) \leq 0, j = 1, \ldots, J\}, \tag{4}$$

represents the minimum lower level function value corresponding to any upper level decision vector. This is called as the lower level optimal value function or the φ-mapping. The bilevel optimization problem can be expressed as follows in terms of the φ-mapping:

$$\min_{x,y} \quad F(x, y) \tag{5}$$

$$\text{subject to}$$

$$f(x, y) \leq \varphi(x) \tag{6}$$

$$g_j(x, y) \leq 0, j = 1, \ldots, J \tag{7}$$

$$G_k(x, y) \leq 0, k = 1, \ldots, K. \tag{8}$$

In this paper, we approximate the φ-mapping using Kriging, and solve the *reduced bilevel optimization problem* in Definition 2 using a standard single level optimization algorithm. The process is carried out iteratively to converge toward the bilevel optimal solution. It is important to note at this point that the optimal value function mapping is a standard function that takes an upper level vector as input and returns a scalar as output. However, this mapping is not readily available.

III. KRIGING

Kriging is a well-known method for interpolation of a function when some of the data points for the function are available. The interpolation process is governed by a Gaussian process, which makes it different from other interpolation techniques like piecewise-polynomial spline. In this study, we utilize Kriging in a similar manner proposed by Jones et al. [22] for solving global optimization problems. In [22], the authors have used Kriging to approximate the objective function to be optimized. A meta-model is updated iteratively by addition of new points in the sample and re-approximation of the original objective function. However, in our study we use Kriging to meta-model the φ-mapping using a sample of points instead of the objective function. As discussed in the section above, a bilevel problem can be reduced to a single level if the φ-mapping is known. However, the mapping is not available, hence we generate a set of data points to approximate the mapping using a response surface. The generation of the data points to approximate the φ-mapping is very costly, as generation of each point requires us to solve the lower level optimization problem. Below we provide the working of the Kriging procedure in the context of the φ-mapping.

The φ-mapping maps a given vector x to a scalar. The dimensionality of the vector is assumed to be n. Assume that a random sample $\mathcal{S}$ of size p is generated in the relaxed feasible region Φ. The sample points are denoted as $x^{(i)} : i \in \{1, \ldots, p\}$. A lower level optimization problem is solved for each $x^{(i)}$ leading to an optimal lower level function value of $f^{(i)}$. If the sample satisfies all the constraints at the upper level, then along with lower level optimality these p points represent feasible bilevel solutions or belong to the set I. Though finding such points itself can be a challenge in a bilevel optimization task, at this point we disregard this issue and assume that such a sample can be generated.

Let $\mathcal{H}$ be the hypothesis space of functions, then one can

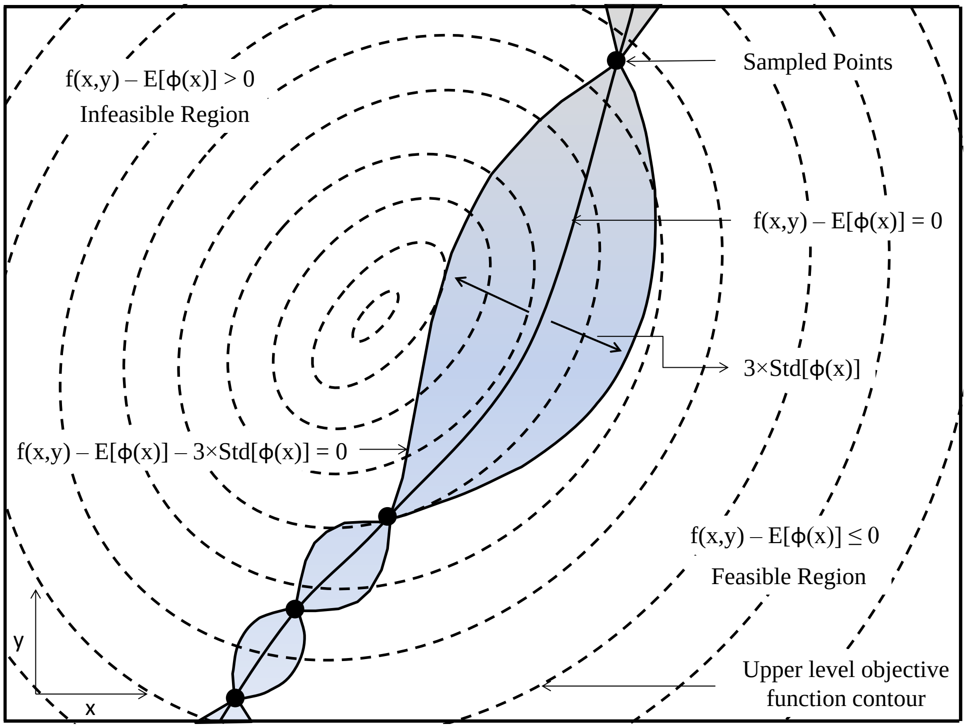

Fig. 1. Expected φ-mapping and its standard deviation obtained from a sample of points.

identify an approximate model $\hat{\varphi} \in \mathcal{H}$ that minimizes the empirical error on the sample, i.e.

$$\hat{\varphi} = \underset{u \in \mathcal{H}}{\operatorname{argmin}} \sum_{i \in \mathcal{I}} L(u(x^{(i)}), f^{(i)}), \tag{9}$$

where $L : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ denotes the prediction error. The prediction error is commonly computed by assuming a quadratic loss function. The above idea works well if one assumes a flexible functional form with unknown parameters that can be identified by minimizing the loss function. Such an approach has been used earlier in a couple of recent studies [37], [33]. In both these studies, the hypothesis space $\mathcal{H}$ is restricted to consist of second-order polynomials. However, in the context of this paper, we consider a slightly different approach. The model is supposed to follow a stochastic process and the parameters are identified using the maximum likelihood estimation.

In a stochastic process approach the model that is used is given as follows:

$$\varphi(x^{(i)}) = \mu + \epsilon(x^{(i)}) \quad i = 1, \ldots, n, \tag{10}$$

where μ is the mean of the stochastic process and $\epsilon(x^{(i)})$ is a normally distributed random error with mean 0 and variance σ^2. The errors at two points are correlated and the correlation is assumed to be decreasing with increasing distance between two points. The correlation between two points is assumed to be given as:

$$\text{Corr}[\epsilon(x^{(i)}), \epsilon(x^{(j)})] = e^{-d(x^{(i)}, x^{(j)})}, \tag{11}$$

where $d(x^{(i)}, x^{(j)}) = \Sigma_{k=1}^{n} \theta_k |x_k^{(i)} - x_k^{(j)}|^{p_k}$ represents weighted distance between two points $x^{(i)}$ and $x^{(j)}$. The above model has $2n + 2$ parameters, namely, $\mu, \sigma, \theta_1, \ldots, \theta_n$ and $p_1, \ldots, p_n$. The optimal value of the parameters can be figured out by maximizing the likelihood of the sample. For brevity, we skip the further details in this paper and refer the readers to the original contributions on design and analysis of computer experiments (DACE) stochastic process model [31] and its application to optimization [22].

A. Sampling Points

Let $\hat{\varphi}(x)$ be the approximated mapping from the sample $\mathcal{S}$ containing p points. The estimation also provides us the standard error for $\varphi(x)$ that we denote as $\hat{s}(x)$. For representation, consider a simple bilevel problem with one upper level variable (x) and one lower level variable (y) and no constraints at the upper level apart from the constrained lower level optimization task (Equation 6). The contour of the upper level objective function for such an example is shown in Figure 1 along with the estimation of the constraint $f(x, y) \leq \varphi(x)$. The possible range for the constraint, $f(x, y) \leq \varphi(x)$, with a 99.74% probability is also plotted in the figure. The use of Kriging provides us the estimates $\hat{\varphi}(x)$ and $\hat{s}(x)$ in one procedure. For any bilevel problem, in order to find out a new point to be sampled, we propose to solve the following auxiliary problem:

$$\min_{x,y} \quad F(x,y) \tag{12}$$

subject to

$$f(x,y) \leq \hat{\varphi}(x) + 3\hat{s}(x) \tag{13}$$

$$g_j(x,y) \leq 0, j = 1,\ldots,J \tag{14}$$

$$G_k(x,y) \leq 0, k = 1,\ldots,K. \tag{15}$$

Note that the inequality 13 in the above formulation contains a factor 3 before the standard error. This has been done to ensure that there is less than 0.13% of a probability of eliminating the actual bilevel optimum by constraint 13. One can choose to further relax the constraint by choosing a larger factor; however, we have chosen a value of 3 for the experiments done in this paper. Let the optimal solution to the above auxiliary problem be (x^{new}, y^{new}). A lower level optimization is performed at x^{new} in order to find out the lower level optimal function value f^{new}. The new point is appended to the sample $\mathcal{S}$, and $\hat{\varphi}(x)$ and $\hat{s}(x)$ are re-estimated. The process is repeated until convergence. In the next section, we provide a step-by-step procedure for the algorithm that we use to solve the bilevel problems based on the principles discussed in this section.

IV. Working of the Algorithm

In this section, we provide a step-by-step procedure for the proposed algorithm. Let the dimensions of the upper level variables be n and the lower level variables be m.

1) The algorithm begins by initializing a sample population $\mathcal{S}$ of size $p = 10n$ within the relaxed feasible region Φ of the bilevel problem. This is achieved by solving the following problem p times with random starting points. Assume that an algorithm $\mathcal{A}$ is used to solve the problem. Since the objective function is a fixed number, the algorithm terminates when a feasible solution is found.

$$\min_{x,y} \quad 0 \tag{16}$$

subject to

$$G_k(x,y) \leq 0, k = 1,\ldots,K, \tag{17}$$

$$g_j(x,y) \leq 0, j = 1,\ldots,J. \tag{18}$$

Let the population generated after this step be $x^{(1)},\ldots,x^{(p)}$.

2) Solve the lower level optimization problem for each point in the sample $x^{(1)},\ldots,x^{(p)}$ using algorithm $\mathcal{A}$ and record the corresponding upper and lower level function values $F^{(1)},\ldots,F^{(p)}$ and $f^{(1)},\ldots,f^{(p)}$. Initialize iteration counter $c \leftarrow 0$. Find the best point in the sample with the minimum upper level function value and denote it as $F_c^{(best)}$.
3) Compute the covariance matrix for the sample $\mathcal{S}$ and denote it by Σ_c. Update $c \leftarrow c+1$.
4) Estimate $\hat{\varphi}(x)$ and $\hat{s}(x)$ and solve the auxiliary problem (12-15) to identify a new point x^{p+1} and add it to sample population $\mathcal{S}$.
5) Solve the lower level optimization problem for this point using algorithm $\mathcal{A}$ to obtain f^{p+1}, and compute F^{p+1}.
6) Create additional random sample of n points (drawn from a normal random distribution with covariance $0.1\times\Sigma_c$) with mean x^{p+1} and solve lower level optimization problem for each point. Add these additional sample points to S after computing the corresponding upper and lower level objective function values.
7) Update $p \leftarrow p+1+n$.
8) For the current iteration find the point in the sample of p points with the minimum upper level function value and denote it as $F_c^{(best)}$.
9) Terminate, if the improvement over multiple iterations is small, otherwise go to step 3.

V. Results

In this section, we provide results for the proposed algorithm on a small set of bilevel test problems. The definition of the problems are provided in Table II. Since the bilevel optimal solution for the test problems are known, the iterations in the problems were repeated until an accuracy of $\epsilon = 0.001$ was achieved for the upper level function values.

The work is in a preliminary stage, where we tested the solution method proposed in this paper on a small set of 5 test problems. It is noteworthy at this point that the challenge with the proposed methodology is not only the approximation of the φ-mapping, but also solving the auxiliary problem (12-15), which tends to be non-convex because of the nature of constraint 13. The nature of the constraint can be realized in Figure 1, where with a small set of sample points, the constraint is already highly non-linear because of the standard error term. Therefore, the algorithm $\mathcal{A}$ required to solve the problem needs to be a global optimization algorithm, where evolutionary algorithms can play an important role. However, in this paper we have relied on sequential quadratic programming (SQP), which is capable of solving convex problems. In order to take care of the non-convexity, we execute SQP multiple times (3) with random starting points on the auxiliary problem and consider the best solution from multiple runs. Finding a local optimum while solving the auxiliary problem will anyway not lead to errors in the proposed bilevel solution method, as a lower level optimization is separately performed for every new point that is generated.

The proposed bilevel solution method was able to successfully solve the chosen 5 test problem. The results on these test problems are provided in Table III from 31 runs on each test problem. Figures 2 and 3 provide additional insights about test problem 4 and the working of the algorithm on this test problem. Figure 2 shows the actual φ-mapping for problem 4 and the upper level objective function plot, when the lower level variables are at its optimum. Note that the optimal lower level variables are a function of the upper level variables; therefore, the plot of the upper level objective function is with respect to the upper level variables alone. From the figure it is clear that the objective function takes the minimum value at the highlighted point. Figure 3 shows the approximated

TABLE II

STANDARD TEST PROBLEMS.

Problem	Formulation	Best Known Sol.
Problem 1		
$n=2$, $m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=(x_1-30)^2+(x_2-20)^2-20y_1+20y_2,$ s.t. $y \in \underset{(y)}{\text{argmin}} \left\{ \begin{array}{l} f(x,y)=(x_1-y_1)^2+(x_2-y_2)^2 \\ 0 \le y_i \le 10, \quad i=1,2 \end{array} \right\},$ $x_1+2x_2 \ge 30, x_1+x_2 \le 25, x_2 \le 15$	$F=225.0$ $f=100.0$
Problem 2		
$n=2$, $m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=2x_1+2x_2-3y_1-3y_2-60,$ s.t. $y \in \underset{(y)}{\text{argmin}} \left\{ \begin{array}{l} f(x,y)=(y_1-x_1+20)^2+(y_2-x_2+20)^2 \\ x_1-2y_1 \ge 10, x_2-2y_2 \ge 10 \\ -10 \ge y_i \ge 20, \quad i=1,2 \end{array} \right\},$ $x_1+x_2+y_1-2y_2 \le 40,$ $0 \le x_i \le 50, \quad i=1,2.$	$F=0.0$ $f=100.0$
Problem 3	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=-(x_1)^2-3(x_2)^2-4y_1+(y_2)^2,$ s.t.	
$n=2$, $m=2$	$y \in \underset{(y)}{\text{argmin}} \left\{ \begin{array}{l} f(x,y)=2(x_1)^2+(y_1)^2-5y_2 \\ (x_1)^2-2x_1+(x_2)^2-2y_1+y_2 \ge -3 \\ x_2+3y_1-4y_2 \ge 4 \\ 0 \le y_i, \quad i=1,2 \end{array} \right\},$ $(x_1)^2+2x_2 \le 4,$ $0 \le x_i, \quad i=1,2$	$F=-18.6787$ $f=-1.0156$
Problem 4	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=rt(x)x-3y_1-4y_2+0.5t(y)y,$ s.t.	
$n=2$, $m=2$	$y \in \underset{(y)}{\text{argmin}} \left\{ \begin{array}{l} f(x,y)=0.5t(y)hy-t(b(x))y \\ -0.333y_1+y_2-2 \le 0 \\ y_1-0.333y_2-2 \le 0 \\ 0 \le y_i, \quad i=1,2 \end{array} \right\},$ where $h=\begin{pmatrix} 1 & 3 \\ 3 & 10 \end{pmatrix}, b(x)=\begin{pmatrix} -1 & 2 \\ 3 & -3 \end{pmatrix} x, r=0.1$ $t(\cdot)$ denotes transpose of a vector	$F=-3.6$ $f=-2.0$
Problem 5		
$n=2$, $m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=-\frac{(x_1+y_1)(x_2+y_2)}{1+x_1y_1+x_2y_2},$ s.t. $y \in \underset{(y)}{\text{argmin}} \left\{ \begin{array}{l} f(x,y)=\frac{(x_1+y_1)(x_2+y_2)}{1+x_1y_1+x_2y_2} \\ 0 \le y_i \le x_i, \quad i=1,2 \end{array} \right\},$ $(x_1)^2+(x_2)^2 \le 100$ $x_1-x_2 \le 0$ $0 \le x_i, \quad i=1,2$	$F=-1.96$ $f=1.96$

$\hat{\varphi}$-mapping at the end of the bilevel solution method. The approximate mapping nicely approximates the true mapping in the region where sample points are present. In the regions where sample points are not present, the approximation errors are high. The standard error corresponding to the $\hat{\varphi}$-mapping (not shown) provides us information about the errors in different regions and where to sample the next points. There could be regions with high approximation errors, but still we need not sample points in that region as the upper level objective function is unlikely to be minimized in that region.

Overall, the appeal of the approach comes from the fact that it does not require assumption of any functional form. Such ideas have been used in the context of costly objective functions to be optimized. Through this study, we demonstrate its utility on bilevel optimization, where the cost arises from solving the lower level optimization problem repeatedly. Interestingly, instead of approximating the objective function, in this study the stochastic process plays a role in approximating a constraint.

TABLE III

MEAN OF THE SUM OF UPPER LEVEL (UL) AND LOWER LEVEL (LL) FUNCTION EVALUATIONS.

	Median Func. Evals.		Standard Deviation	
	Upper Level	Lower Level	Upper Level	Lower Level
Problem 1	522	1872	102.34	308.32
Problem 2	684	2648	134.56	442.34
Problem 3	77	324	21.44	86.57
Problem 4	724	2202	112.84	266.34
Problem 5	910	2211	166.87	299.75

VI. CONCLUSIONS

In this paper, we have proposed to meta-model the lower level optimal value function as a function of upper level variables, so that the overall computational time to solve a bilevel problem is reduced. The well known Kriging method has been used for approximation, mainly due to its ability to provide the mean and standard error for the φ-mapping from a sample of points. The approximate model for the lower level optimal value function is used to transform the original bilevel problem into a single level optimization task. The algorithm starts with a sample of points for approximating the φ-mapping. The reduced bilevel problem is then solved frequently and new points produced by the model are added into the sample of points to improve the approximation of the φ-mapping. The approach has shown promising results on some simple bilevel optimization problems but requires a further study and improvements to make it viable for solving difficult bilevel problems. To the best of our knowledge, in the context of bilevel optimization Kriging has not been used in such a manner, i.e., for iterative approximation of the φ-mapping, to solve the bilevel optimization problems. The proposed approach aims to bring down the computational expense as compared to the contemporary approaches (nested approaches in particular). The efficacy of the proposed methodology is now ready to be tested on higher dimensional problems. The concept can be extended to multi-objective bilevel optimization problems, as well. Despite their practical significance, bilevel optimization problems have been avoided in practice for a long time due to their computational burden. However, we believe that the mathematical, statistical and meta-heuristics based approaches can be integrated to develop approximate solution methodologies for bilevel optimization problems.

REFERENCES

[1] E. Aiyoshi and K. Shimizu. Hierarchical decentralized systems and its new solution by a barrier method. *IEEE Transactions on Systems, Man, and Cybernetics*, 6:444–449, 1981.

[2] E. Aiyoshi and K. Shimizu. A solution method for the static constrained Stackelberg problem via penalty method. *IEEE Transactions on Automatic Control*, 29:1111–1114, 1984.

[3] F. Al-Khayyal, R. Horst, and P. Pardalos. Global optimization of concave functions subject to quadratic constraints: an application in nonlinear bilevel programming. *Annals of Operations Research*, 34:125–147, 1992.

[4] J. Angelo, E. Krempser, and H. Barbosa. Differential evolution for bilevel programming. In *Proceedings of the 2013 Congress on Evolutionary Computation (CEC-2013)*. IEEE Press, 2013.

[5] Jaqueline S Angelo, Eduardo Krempser, and Helio JC Barbosa. Differential evolution assisted by a surrogate model for bilevel programming problems. In *Evolutionary Computation (CEC), 2014 IEEE Congress on*, pages 1784–1791. IEEE, 2014.

[6] J. Bard and J. Falk. An explicit solution to the multi-level programming problem. *Computers and Operations Research*, 9:77–100, 1982.

[7] J. Bard and J. Moore. A branch and bound algorithm for the bilevel programming problem. *SIAM Journal on Scientific and Statistical Computing*, 11:281–292, 1990.

[8] J. F. Bard. *Practical Bilevel Optimization: Algorithms and Applications*. The Netherlands: Kluwer, 1998.

[9] O. Ben-Ayed. Bilevel linear programming. *Computers and Operations Research*, 20:485–501, 1993.

[10] J. Bracken and J. McGill. Mathematical programs with optimization problems in the constraints. *Operations Research*, 21:37–44, 1973.

[11] Luce Brotcorne, Martine Labbe, Patrice Marcotte, and Gilles Savard. A bilevel model for toll optimization on a multicommodity transportation network. *Transportation Science*, 35(4):345–358, 2001.

[12] G. Brown, M. Carlyle, D. Diehl, J. Kline, and K. Wood. A Two-Sided Optimization for Theater Ballistic Missile Defense. *Operations Research*, 53(5):745–763, 2005.

[13] Peter A. Clark and Arthur W. Westerberg. Bilevel programming for steady-state chemical process design-i. fundamentals and algorithms. *Computers & Chemical Engineering*, 14(1):87–97, 1990.

[14] Benoît Colson, Patrice Marcotte, and Gilles Savard. A trust-region method for nonlinear bilevel programming: algorithm and computational experience. *Computational Optimization and Applications*, 30(3):211–227, 2005.

[15] Stephan Dempe. *Foundations of Bilevel Programming*. Kluwer Academic Publishers, Secaucus, NJ, USA, 2002.

[16] T. Edmunds and J. Bard. Algorithms for nonlinear bilevel mathematical programming. *IEEE Transactions on Systems, Man, and Cybernetics*, 21:83–89, 1991.

[17] J. Fortuny-Amat and B. McCarl. A representation and economic interpretation of a two-level programming problem. *Journal of the Operational Research Society*, 32:783–792, 1981.

[18] S. Reza Hejazi, Azizollah Memariani, G. Jahanshahloo, and Mohammad Mehdi Sepehri. Linear bilevel programming solution by genetic algorithm. *Computers & Operations Research*, 29(13):1913–1925, 2002.

[19] X. Hu and D. Ralph. Using EPECs to Model Bilevel Games in Restructured Electricity Markets with Locational Prices. *Operations Research*, 55(5):809–827, 2007.

[20] Y. Ishizuka and E. Aiyoshi. Double penalty method for bilevel optimization problems. *Annals of Operations Research*, 34:73–88, 1992.

[21] Yan Jiang, Xuyong Li, Chongchao Huang, and Xianing Wu. Application of particle swarm optimization based on chks smoothing function for solving nonlinear bilevel programming problem. *Applied Mathematics and Computation*, 219(9):4332–4339, 2013.

[22] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[23] M. Labbé, P. Marcotte, and G. Savard. A Bilevel Model of Taxation and Its Application to Optimal Highway Pricing. *Management Science*, 44(12):1608–1622, 1998.

[24] Hecheng Li. A genetic algorithm using a finite search space for solving nonlinear/linear fractional bilevel programming problems. *Annals of Operations Research*, pages 1–16, 2015.

[25] Xiangyong Li, Peng Tian, and Xiaoping Min. A hierarchical particle swarm optimization for solving bilevel programming problems. In Leszek Rutkowski, Ryszard Tadeusiewicz, Lotfi A. Zadeh, and Jacek M. Zurada, editors, *Artificial Intelligence and Soft Computing - ICAISC 2006*, volume 4029 of *Lecture Notes in Computer Science*, pages 1169–1178. Springer Berlin Heidelberg, 2006.

[26] Guoshan Liu, Jiye Han, and Shouyang Wang. A trust region algorithm for bilevel programing problems. *Chinese science bulletin*, 43(10):820–824, 1998.

[27] Patrice Marcotte, Gilles Savard, and D. L. Zhu. A trust region algorithm for nonlinear bilevel programming. *Operations research letters*, 29(4):171–179, 2001.

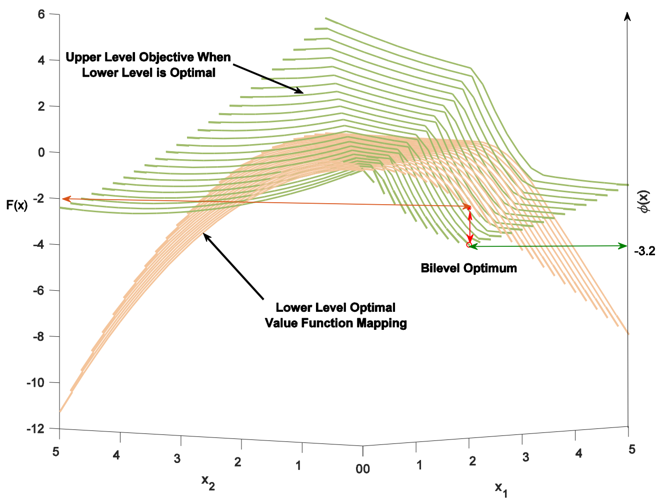

Fig. 2. Problem 4: Bilevel optimum is shown with the actual φ-mapping and the upper level objective as a function of upper level variables and optimal lower level variables.

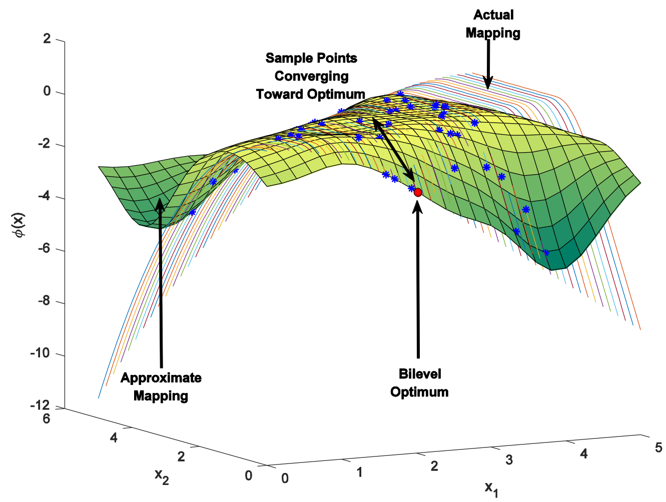

Fig. 3. Problem 4: Approximated $\bar{\varphi}$-mapping plotted over the actual φ-mapping with sample points at the end of the algorithm. The algorithm is able to converge towards the true bilevel optimum.

[28] R. Mathieu, L. Pittard, and G. Anandalingam. Genetic algorithm based approach to bi-level linear programming. *Operations Research*, 28(1):1–21, 1994.

[29] Athanasios Migdalas. Bilevel programming in traffic planning: Models, methods and challenge. *Journal of Global Optimization*, 7(4):381–405, 1995.

[30] M.G. Nicholls. Aluminium Production Modeling - A Nonlinear Bilevel Programming Approach. *Operations Research*, 43(2):208–218, 1995.

[31] Jerome Sacks, William J Welch, Toby J Mitchell, and Henry P Wynn. Design and analysis of computer experiments. *Statistical science*, pages 409–423, 1989.

[32] G. Savard and J. Gauvin. The steepest descent direction for the nonlinear bilevel programming problem. *Operations Research Letters*, 15:275–282, 1994.

[33] A. Sinha, Z. Lu, K. Deb, and P. Malo. Bilevel optimization based on iterative approximation of multiple mappings. *arXiv preprint arXiv:1702.03394*, 2017.

[34] A. Sinha, P. Malo, and K. Deb. Efficient evolutionary algorithm for single-objective bilevel optimization. *arXiv preprint arXiv:1303.3901*, 2013.

[35] A. Sinha, P. Malo, and K. Deb. An improved bilevel evolutionary algorithm based on quadratic approximations. In *2014 IEEE Congress on Evolutionary Computation (CEC-2014)*, pages 1870–1877. IEEE Press, 2014.

[36] A. Sinha, P. Malo, and K. Deb. Transportation policy formulation as a multi-objective bilevel optimization problem. In *2015 IEEE Congress on Evolutionary Computation (CEC-2015)*. IEEE Press, 2015.

[37] A. Sinha, P. Malo, and K. Deb. Solving optimistic bilevel programs by iteratively approximating lower level optimal value function. In *2016 IEEE Congress on Evolutionary Computation (CEC-2016)*. IEEE Press, 2016.

[38] A. Sinha, P. Malo, and K. Deb. Evolutionary algorithm for bilevel optimization using approximations of the lower level optimal solution mapping. *European Journal of Operational Research*, 257(2):395–411, 2017.

[39] A. Sinha, P. Malo, and K. Deb. A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 2017.

[40] A. Sinha, P. Malo, A. Frantsev, and K. Deb. Multi-objective stackelberg game between a regulating authority and a mining company: A case study in environmental economics. In *2013 IEEE Congress on Evolutionary Computation (CEC-2013)*. IEEE Press, 2013.

[41] A. Sinha, P. Malo, A. Frantsev, and K. Deb. Finding optimal strategies in a multi-period multi-leader-follower stackelberg game using an evolutionary algorithm. *Computers & Operations Research*, 41:374–385, 2014.

[42] W. R. Smith and R. W. Missen. *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*. John Wiley & Sons, New York, 1982.

[43] H. Stackelberg. *The theory of the market economy*. Oxford University Press, New York, Oxford, 1952.

[44] H. Tuy, A. Migdalas, and P. Värbrand. A global optimization approach for the linear two-level program. *Journal of Global Optimization*, 3:1–23, 1993.

[45] L. Vicente, G. Savard, and J. Júdice. Descent approaches for quadratic bilevel programming. *Journal of Optimization Theory and Applications*, 81:379–399, 1994.

[46] Y. Wang, Y. C. Jiao, and H. Li. An evolutionary algorithm for solving nonlinear bilevel programming based on a new constraint-handling scheme. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 32(2):221–232, 2005.

[47] Yuping Wang, Hong Li, and Chuangyin Dang. A new evolutionary algorithm for a class of nonlinear bilevel programming problems and its global convergence. *INFORMS Journal on Computing*, 23(4):618–629, 2011.

[48] L. Wein. Homeland Security: From Mathematical Models to Policy Implementation: The 2008 Philip McCord Morse Lecture. *Operations Research*, 57(4):801–811, 2009.

[49] U. Wen and S. Hsu. Linear bi-level programming problems - a review. *Journal of the Operational Research Society*, 42:125–133, 1991.

[50] D. White and G. Anandalingam. A penalty function approach for solving bi-level linear programs. *Journal of Global Optimization*, 3:397–419, 1993.

[51] Jane J. Ye and Daoli Zhu. New necessary optimality conditions for bilevel programs by combining the mpec and value function approaches. *SIAM Journal on Optimization*, 20(4):1885–1905, 2010.

[52] Y. Yin. Genetic algorithm based approach for bilevel programming models. *Journal of Transportation Engineering*, 126(2):115–120, 2000.

[53] Xiaobo Zhu, Qian Yu, and Xianjia Wang. A hybrid differential evolution algorithm for solving nonlinear bilevel programming with linear constraints. In *Cognitive Informatics, 2006. ICCI 2006. 5th IEEE International Conference on*, volume 1, pages 126–131. IEEE, 2006.