Real-Time Optimal Parametric Design of a Simple Infiltration-Evaporation Model using the Assess-Predict-Optimize(APO) Strategy

S. Ali*, M. Damodaran†, A. T. Patera‡

Abstract— **Optimal parametric design of a system must be able to respond quickly to short term needs as well as long term conditions. To this end, we present an Assess-Predict-Optimize (APO) strategy which allows for easy modification of a system's characteristics and constraints, enabling quick design adaptation. There are three components to the APO strategy: Assess - extract necessary information from given data; Predict - predict future behavior of system; and Optimize - obtain optimal system configuration based on information from the other components. The APO strategy utilizes three key mathematical ingredients to yield real-time results which would certainly conform to given constraints: dimension reduction of the model, *a posteriori* error estimation, and optimization methods. The resulting formulation resembles a bilevel optimization problem with an inherent nonconvexity in the inner level. Using a simple infiltration-evaporation model to simulate an irrigation system, we demonstrate the APO strategy's ability to yield real-time optimal results. The linearized model, described by a coercive elliptic partial differential equation, is discretized by the reduced-basis output bounds method. A primal-dual interior point method is then chosen to solve the resulting APO problem.**

Keywords—**reduced–basis, *a posteriori* error estimation, design optimization, nonlinear optimization, bilevel optimization, inverse problems**

I. Introduction

Optimal design of a system usually involves three different aspects: the identification of the system's parameters; the ability to do failure analysis so as to be able to avoid these faults in the design; and to find the optimal design variables. In non-destructive testing, experimental data is collected for parameter identification and for failure analysis. In most cases, especially in geophysics, the parameters are determined through the methods of inverse problems ([1], [2], [3]). However, none of the works cited above is able to determine the accuracy of the results obtained.

The optimization problems for parametric design models are often multiobjective and multilevel, depending on the constraints and needs of the models. They are usually solved using global optimization approaches, favoring genetic algorithms and simulated annealing techniques ([4], [5], [6], [7]). While these methods assures global convergence, they require a lot of computational effort and time to solve for the output of interests of the system.

To add to the complexity, the system is described by partial differential equations. The discretization techniques employed would yield variables of very large dimensions, costing more time and effort to solve.

The Assess-Predict-Optimize (APO) strategy presented here overcomes some of the issues above. Using the data gathered as constraints of the optimization problem directly, it is able to give the assurance that the true parameters are encompassed within the constraints, and quantify the error gap of the predicted solutions. It utilizes the reduced order model approach to reduce the dimension of the system to be solved, and the *a posteriori* error estimates to ensure that the correct criteria are predicted. Using a primal-dual interior point method, it is able to find a local optimal solution in a much faster response time than using semi-random methods like genetic algorithms.

For now, we have not tested the APO strategy on multiobjective systems, though the problem presented in this paper translates into a min-max problem. The problem, an infiltration-evaporation model, attempts to simulate an irrigation system, where we would like to optimize the intake of water. The model presented is a very basic model that doesn't take into consideration many of the physical processes involved. Nor does it describe the way in which the water could be distributed. It is, however, a useful model to illustrate how the APO strategy could be employed to optimize the system.

II. Infiltration-Evaporation model

The two dimensional infiltration-evaporation model presented here is an adaptation of Buchwald's and Viera's model [8]. It assumes a land divided into two regions; a wetted area and a dry area. The wetted surface represents areas where irrigation occurs; either a plantation or cropfield, and the dry region represents areas which need not be irrigated. Underneath the two regions, there is a water table, the layer where the ground is saturated with water. Figure 1 shows the model. In this model we normalized the distance with respect to the depth of the soil to the water

* Graduate Student, Nanyang Technological University, Singapore-Massachusetts Institute of Technology Alliance (SMA), School of Mechanical and Production Engineering, Nanyang Avenue, Singapore 639798

† Associate Professor, Nanyang Technological University, Singapore-Massachusetts Institute of Technology Alliance (SMA), School of Mechanical and Production Engineering, Nanyang Avenue, Singapore 639798

‡ Professor, Massachusetts Institute of Technology, Department of Mechanical Engineering, Bldg. 3-264, Cambridge, MA 02139-4307 USA

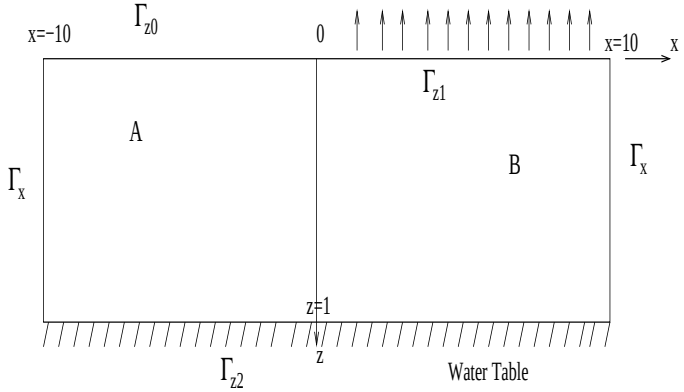

Fig. 1. 2D model of Infiltration and Evaporation system: A wetted layer at $z = 0, x < 0$ with a horizontal water-table at z = 1. Linear evaporation is assumed at $z = 0, x > 0$.

table.

In the real world the depth to the water table is between 10-50 m whereas plantations and crop-fields can spread up to several km. As such, Buchwald and Viera assumed semi-infinite wet and dry regions. In this model we limit the regions to 10 times the depth. At x = 0, we have the border between these two regions, a critical region to ensure there is enough water in the ground for the crops growing in the area.

The ground above the water table, be it region A or B, is unsaturated and thus obeys Darcy's law:

$$F = -K(\Psi)\nabla(\Psi - z) \tag{1}$$

where the volumetric flux,F, is equal to Θu, Θ is the volumetric water content, and u is the fluid velocity. $K(\Psi)$ is the hydraulic conductivity and Ψ is the potential for the local forces arising from the interaction of the water, soil, and air.

A. Linearized Model

The equation for the conservation of water is

$$-\nabla(K(\Psi)\nabla\Psi) + \frac{\partial K(\Psi)}{\partial z} = 0 \tag{2}$$

Eq. 2 is nonlinear and in order to linearize it we introduce the relative permeability, κ, where

$$\kappa = \mu K/\rho g k_0, \qquad 0 \leq \kappa \leq 1 \tag{3}$$

Here μ is the dynamic viscosity, k_0 is the saturation permeability, and ρ is the density of water. The relative permeability, κ, measures the wetness of the soil. Philip [9] proposed another formula for κ which can be stated as

$$\kappa = e^{\alpha\Psi}, \qquad -\infty < \Psi < 0 \tag{4}$$

where α is a constant related to the capillary forces within the soil. It depends on the type of soil within the area. We note that in Waetcher and Philip [10], α ranges from 0.2 in the case of fine soil to 5 in the case of rough soil. In our model, we assume that both regions have the same type of soil, which means that α is a uniform constant throughout the model.

Substituting eq. 3 and 4 into eq. 2 the linearized equation becomes

$$-\nabla^2\kappa + \alpha\frac{\partial\kappa}{\partial z} = 0 \tag{5}$$

We also let $\kappa = \phi(x,z)e^{\beta z}$ where $\beta = .5\alpha$ to make the problem symmetric. As in Buchwald and Viera, we assume that infiltration occurs at a constant rate in region A and evaporation occurs in region B. We then derive the following governing equation and boundary conditions:

$$-\nabla^2\phi + \beta^2\phi = 0 \tag{6}$$

i. A wetted region described by $\phi = \kappa_0$, at Γ_{z0}
ii. The water table described by $\phi = e^{-\beta}$, at Γ_{z2}
iii. In the soil layer, denoted in the model as Γ_x, the condition is described by $-\nabla\phi.\hat{n} = 0$
iv. In the dry region denoted by Γ_{z1}, the occurring evaporation is described by $-\nabla\phi.\hat{n} = (h+\beta)\phi$, where $\beta = .5\alpha$. Here the constant h is a measure of the evaporation.

The strong form represented by Eq. 6 can be rewritten in terms of a minimization statement or variational formulation. Let us first introduce X, and X^1, subsets of the Hilbert space such that $X = \{w \in H^1 | w|_{\Gamma_{z0}} = 0, w|_{\Gamma_{z2}} = 0\}$, and $X^1 = \{w \in H^1 | w|_{\Gamma_{z0}} = \kappa_0, w|_{\Gamma_{z2}} = e^{-\beta}\}$. The variational formulation can be stated as

$$\int_\Omega \nabla v\nabla\phi dA + (h+\beta)\int_{\Gamma_{z1}} v\phi dS + \beta^2\int_\Omega v\phi dA = 0 \tag{7}$$

$\forall v \in X$, and $\phi \in X^1$.

We introduce two new subsets: $X^2 = \{w \in H^1 | w|_{\Gamma_{z0}} = 1, w|_{\Gamma_{z2}} = 0\}$; and $X^3 = \{w \in H^1 | w|_{\Gamma_{z0}} = 0, w|_{\Gamma_{z2}} = 1\}$. We can then separate the problem into two sub-problems:

$$\phi = \kappa_0\phi_1 + e^{-\beta}\phi_2 \tag{8}$$

where ϕ_1 is described by

$$\int_\Omega \nabla v\nabla\phi_1 dA + (h+\beta)\int_{\Gamma_{z1}} v\phi_1 dS + \beta^2\int_\Omega v\phi_1 dA = 0 \tag{9}$$

$\forall v \in X$, and $\phi_1 \in X^2$. Similarly, ϕ_2 is described by

$$\int_\Omega \nabla v\nabla\phi_2 dA + (h+\beta)\int_{\Gamma_{z1}} v\phi_2 dS + \beta^2\int_\Omega v\phi_2 dA = 0 \tag{10}$$

$\forall v \in X$, and $\phi_2 \in X^3$.

B. Optimal Design of the Irrigation System

We are interested in two outputs: the average evaporation rate at Γ_{z1}, E, and the average relative permeability at x = 0, κ_{border}. E can be evaluated from

$$\begin{aligned} E &= h\int_{\Gamma_{z1}} \phi dS \\ &= h\kappa_0 \int_{\Gamma_{z1}} \phi_1 dS + he^{-\beta}\int_{\Gamma_{z1}} \phi_2 dS \\ &= E_1 + E_2 \end{aligned} \tag{11}$$

and similarly κ_{border} is evaluated from

$$\begin{aligned} \kappa_{border} &= \int_{x=0} \phi e^{\beta z} dS \\ &= \kappa_0 \int_{x=0} \phi_1 e^{\beta z} dS + e^{-\beta}\int_{x=0} \phi_2 e^{\beta z} dS \\ &= \kappa^1_{border} + \kappa^2_{border} \end{aligned} \tag{12}$$

In order to maintain a certain amount of moisture in region A (wet region), a certain minimum average relative permeability at x = 0, κ^c, needs to be satisfied. This is the main constraint for the model's design.

The amount of water loss in the ground is related to evaporation, which is dependent on the evaporation rate, barring all other influences. Therefore we wish to retain water in the soil by minimizing the evaporation rate, E.

In our model, E and κ_{border} depend on three parameters; h, which is dependent on the relative humidity, β, related to the ability of the soil to retain water, and κ_0, which measures the amount of water being supplied to region A.

Assuming that α is uniform, β is also uniform and cannot be controlled. The parameter h, being dependent on the relative humidity, cannot be controlled either.

Let us assume that the h is an average over the course of a time period. It is not easy to determine the values of β and h since their values are derived from the physical processes and quantities related to the soil. Let us denote the true values of β and h as β^* and h^*. We are only able to predict these values using experimental data.

Our design variable is κ_0 and our design problem is now

$$\begin{aligned} &\min_{\kappa_0} E(\kappa_0, h^*, \beta^*) \\ &\text{s.t. } \kappa_{border}(\kappa_0, h^*, \beta^*) \geq \kappa^c \\ &0 \leq \kappa_0 \leq 1 \end{aligned} \tag{13}$$

where β^* and h^* are found from the experimental data.

The advantage of using the APO strategy in solving this design problem is in its ability to overcome the difficulty in finding β^* and h^* as will be shown in the next section. It also allows the design problem to respond quickly to the changes in κ^c and in the experimental data sets to find β^* and h^*.

III. Assess-Predict-Optimize (APO) strategy

The APO strategy is applied to the design optimization problem, Eq. 13, in order to find the optimal value for κ_0. The APO strategy has three components, elements separately found in various design optimization models: the assess component which extracts useful and necessary information from the available data; the predict component which is used in failure prevention; and the optimize component which tries to find the optimal control variable or output of interest. The strategy allows for an iterative process in which all three elements can be evaluated dynamically.

A. APO formulation

Let us define some general terms related to the APO strategy. Let μ be the parameter set to be determined, and ρ be the control variable used to find the experimental data set. For each ρ^l, $l = 1, ..., L$, we define the output of interest related to the experimental data set, $s(\rho^l, \mu)$, as

$$s(\rho^l, \mu) \in I^l \tag{14}$$

where I^l is defined to be

$$I^l = [s_{min}(\rho^l), s_{max}(\rho^l)], \tag{15}$$

and, $s_{min}(\rho^l)$ and $s_{min}(\rho^l)$ are found by manipulating the original experimental data set using statistical and error analysis.

The space spanned by μ is denoted as $D_\mu = \{\mu \in \Re^n | \mu_l \leq \mu \leq \mu_u\}$. Using Eq. 14, we can define the set of values of μ that yield $s(\rho^l, \mu)) \in I^l$ as

$$B = \{\mu \in D_\mu | s(\rho^l, \mu)) \in I^l, l = 1, ..., L\} \tag{16}$$

We call B as the feasible domain of μ.

We assume that in the absence of systematic error,

$$s(\rho^l, \mu^*) \in I^l \tag{17}$$

where μ^* is the true value of the parameter set, μ. This indicates that $\mu^* \in B$.

We define a new control variable θ, which differs from ρ, the control variable of the experimental data set. The space spanned by θ is denoted as $D_\theta = \{\theta \in \Re^m | \theta_l \leq \theta \leq \theta_u\}$. We also define two new terms which are the results of a minimization and a maximization process.

DEFINITION 1: Define the set of optimal μ, $\quad \forall\, \theta \in D_\theta$ *as*

$$\bar{\Upsilon}^y_{max}(\theta) = \arg\max_{\mu \in B} y(\theta, \mu)$$

and

$$\bar{\Upsilon}^{y}_{min}(\theta) = \arg\min_{\mu \in B} y(\theta, \mu)$$

for a certain output of interest, y.

We can now present the following result. Given that $\mu^* \in B$,

$$y(\theta, \bar{\Upsilon}^{y}_{min}(\theta)) \leq y(\theta, \mu^*) \leq y(\theta, \bar{\Upsilon}^{y}_{max}(\theta)) \tag{18}$$

The proof of Eq. 18 is embedded in the definitions of the minimization and maximization statements themselves.

Let $f(\theta, \mu)$ and $g(\theta, \mu)$ be the outputs of interest controlled by θ. We generalize the optimization problem given by Eq. 13 with the following formulation

$$\min_{\theta \in D_\theta} f(\theta, \mu^*) \tag{19}$$
$$\text{s.t. } g(\theta, \mu^*) \geq C$$

where C is a constant.

From Eq. 18, it is easy to see that $f(\theta, \mu^*) \leq f(\theta, \bar{\Upsilon}^{f}_{max}(\theta))$ and $g(\theta, \mu^*) \geq g(\theta, \bar{\Upsilon}^{g}_{min}(\theta))$. We can easily replace μ^* in Eq. 19 with $\bar{\Upsilon}^{f}_{max}(\theta)$ and $\bar{\Upsilon}^{g}_{min}(\theta)$. The replacements ensure that the results of Eq. 19 will always be feasible, and the result of $f(\theta, \bar{\Upsilon}^{f}_{max}(\theta))$ approaches the true optimal f from above as proven in [11].

We can then write our general APO formulation to be

$$\min_{\theta \in D_\theta} f(\theta, \bar{\Upsilon}^{f}_{max}(\theta)) \tag{20}$$
$$\text{s.t. } g(\theta, \bar{\Upsilon}^{g}_{min}(\theta)) \geq C$$

This formulation embodies the three components of the APO strategy to be described in the following subsections.

B. Assess

In the assess component, μ^* is usually obtained by applying methods of inverse problems to the experimental data ([1], [2], [3]). We present here a different approach to parametric determination.

By using statistical analysis, the experimental data set can be reformulated to find I^l given in Eq. 15. The statistical analysis tools takes into account the systematic and random errors, thus assuring that μ^* is within the feasible domain B. From its definition, the feasible domain, B, can be found regardless whether the given problem is ill-posed or well-posed. This indicates that we are assured that the true solution is always bounded and feasible.

C. Predict

The APO strategy adapts the idea by Haber et. al. [12] where they used the experimental data as a constraint of the inverse optimization. Referring back to the Assess component, it uses the formulation of B as the constraints to find $\bar{\Upsilon}^{y}_{min}(\theta)$ and $\bar{\Upsilon}^{y}_{max}(\theta)$, the minimum and maximum values of the output of interest, y, respectively, for a particular θ.

The result from Eq. 18 yields the uncertainty gap. We define the uncertainty gap as follows; for a particular control variable set, θ, the uncertainty gap for a particular output of interest y, $U_y(\theta)$ is evaluated to be

$$U_y(\theta) = y(\theta, \bar{\Upsilon}^{y}_{max}(\theta)) - y(\theta, \bar{\Upsilon}^{y}_{min}(\theta)) \tag{21}$$

In the case of well-posedness, as B converges to μ^*, the uncertainty gap, U_y converges to zero. Therefore, without even evaluating the domain B, the APO strategy is able to quantify the accuracy of the experimental data, and their predictions of y through the uncertainty gap U.

D. Optimize

With the ability to quantify the accuracy of the output of interest, y, and given the assurance that the true output of interest is always feasible, the APO strategy seeks to replace μ^* with $\bar{\Upsilon}^{g}_{min}(\theta)$ and $\bar{\Upsilon}^{f}_{max}(\theta)$ in Eq. 19 to form Eq. 20. The assurance of feasibility allows the APO strategy to find an optimal value of θ that guarantees the value of $f(\theta)$ approaches the true minimum from above, while still satisfying the constraints on g.

However, the introduction of $\bar{\Upsilon}^{g}_{min}(\theta)$ and $\bar{\Upsilon}^{f}_{max}(\theta)$ changes the complexity of the optimization problem. Instead of a single level optimization problem, the APO problem is now a bilevel optimization problem due to the inner level optimization problems: $\bar{\Upsilon}^{f}_{max}(\theta)$ and $\bar{\Upsilon}^{g}_{min}(\theta)$. The inner level problems are nonconvex due to the formulation of their constraints represented by B. We will discuss the method we use to solve for the APO problem in later sections.

E. Applying APO to the model problem

In our problem, we assume we have the experimental values for the evaporation rate for several specific values of κ_0. Therefore, κ_0 acts as the control variable, ρ. In reality, there could be different experimental sets to solve for h^* and β^* and so ρ need not be κ_0. For example h^* can be evaluated from the relative humidity which is controlled by the temperature on the surface or air pressure.

Our parameter set μ is the set $[h, \beta]$. Therefore, $n = 2$ and D_μ is actually $D_{h,\beta}$. In Eq. 13, θ is κ_0, f is E, g is κ_{border}, and C is κ^c. We can then write the APO problem for our model as

$$\min_{\kappa_0} E(\kappa_0, \bar{\Upsilon}^{E}_{max}(\kappa_0)) \tag{22}$$
$$\text{s.t. } \kappa_{border}(\kappa_0, \bar{\Upsilon}^{\kappa_{border}}_{min}(\kappa_0)) \geq \kappa^c$$
$$0 \leq \kappa_0 \leq 1$$

IV. Reduced-basis Output-bound method

In order to solve the APO formulation given by Eq. 22, E and κ_{border} need to be evaluated repeatedly. A reduced order method is used to ensure fast computation of the outputs. We chose the reduced-basis output-bound method as it also incorporates the *a posteriori* error estimation needed to ensure feasibility of the results.

A. *Reduced-basis and* A Posteriori *Error Bounds formulation*

The reduced-basis technique was introduced first in the 70s and later named by Noor and Peters [13]. Current reduced-basis technique ([14], [15], [16], [17], [18]) uses a two stage *offline/online* method which solves the finite element approximation of Eqs. 9 and 7 in the offline stage and the 'reduced' problem in the online stage.

In general, Eqs. 9 and 10 can be written as

$$a(u, v; \mu) = l(v), \forall v \in Y \tag{23}$$

where $\mu \in D_\mu$ is the parameter set of the problem, Y is the appropriate Hilbert space, $u \in Y$, $a(u, v; \mu)$ is the bilinear form, affine with respect to μ, and $l(v)$ is the linear functional. We denote $s(\mu) = l^o(u(\mu))$ as the output of interest.

We introduce a sample set in the parameter space spanned by D_μ, $S^N = \{\mu_1, ..., \mu_N\}$, and its associated reduced-basis space $W^N = span\{\zeta_n \equiv u(\mu_n), n = 1, ..., N\}$ where $u(\mu_n)$ is the solution of Eq. 23 for $\mu_n \in D_\mu$. Then our reduced-basis approximation to $u(\mu)$ for a given μ, $u_N(\mu) \in W^N$ satisfies

$$a(u_N(\mu), v; \mu) = l(v), \forall v \in W^N \tag{24}$$

and we can evaluate the approximation to the output of interest as $s^N(\mu) = l^o(u_N(\mu))$. It has been shown that the reduced-basis approximation is optimal in the Y norm and so is the solution, s^N.

We also introduce a simple error bound which is described in [18] in further detail. A short description is shown here.

We choose two integer values which when added will be less than the number of bases, N. Let these two values be I and J. We need the ratio of $J/I^{crit} > 0.7$, for some I^{crit}. In this problem, we chose $J = I$, for $I > I^{crit}$. So we can evaluate s^I and s^{I+J} to obtain the average $\bar{s}_I$ and its error bound gap. The average output, $\bar{s}_I$, and its error gap is given by

$$\bar{s}_I = s^I + 1/(2\gamma)|s^{I+J} - s^I| \tag{25}$$

and the error bound

$$\triangle_I = 1/\gamma|s^{I+J} - s^I| \tag{26}$$

for some constant $\gamma \in (0, 1)$.

It is shown in [18] that

$$\bar{s}_I^- \leq s(\mu) \leq \bar{s}_I^+$$

where $\bar{s}_I^- = \bar{s}_I - \triangle_I$ and $\bar{s}_I^+ = \bar{s}_I + \triangle_I$.

In our application, both outputs of interest, E and κ_{border}, can be evaluated in the same way as s^N. Here, $l^o(u_N(\mu))$ is actually represented by Eqs. 11 and 12. Therefore, according to the reduced-basis output bounds, we are able to find

$$\bar{E_i}_I^- \leq E_i(h, \beta) \leq \bar{E_i}_I^+$$

and

$$\bar{\kappa}_{border,I}^{i-} \leq \kappa_{border}^i(h, \beta) \leq \bar{\kappa}_{border,I}^{i+}$$

.

for $i = 1, 2$. Before we discuss the significance of these bounds to the APO problem described by Eq. 22, let us describe the two stage *Offline/Online* approach to the reduced-basis output bounds formulation.

B. Offline/Online *approach*

We first assume that for some finite integer Q, there exist an affine decomposition of $a(w, v; \mu)$ such that

$$a(w, v; \mu) = \sum_{q=1}^{Q} \sigma^q(\mu) a^q(w, v), \forall w, v \in Y, \mu \in D_\mu \tag{27}$$

Here, $a^q(w, v)$, for each q, is no longer dependent on μ. This allows us to separate the problem into two stages: *offline* and *online*.

In the *offline* stage, we calculate $u(\mu_n)$, $n = 1, ..., N$ to form W^N. Defining the sets $\mathcal{N}$ as $\{1, ..., N\}$, and $\mathcal{Q}$ as $\{1, ..., Q\}$, we can compute $A^q \in \Re^{N \times N}$ as

$$A_{i,j}^q = a^q(\zeta_i, \zeta_j), \quad \forall\, i, j \in \mathcal{N} \tag{28}$$

and $q \in \mathcal{Q}$. Similarly we compute F_N as $F_N = l(\zeta_i)$ and L_N as $L_N = l^o(\zeta_i)$, $\forall i \in \mathcal{N}$, where $F_N, L_N \in \Re^N$.

The computational effort to construct the reduced-basis space, W^N, is at most of the order of $O(n^3)$ where n is the dimension of the finite element matrices. In our case, the computational cost might be of the order of $O(n^2)$ since the matrices are sparse and symmetric, and can be solved using iterative methods like the conjugate gradient methods.

In the online stage, we can simply reassemble, for a given μ, our bilinear form as $A_N = \sum_{q=1}^{Q} \sigma^q(\mu) A^q$. We then solve for $u_N(\mu)$ from $A_N u_N = F_N$, and evaluate our output of interest as $s_N(\mu) = (L_N)^T u_N$. Following Eqs. 25 and 26, we obtain our bounds.

The computational effort for the *online* stage is of the order of $O(N^3 + QN^2)$.

C. Application to our problem

As mentioned, the reduced-basis method gives much faster results for the solution of our outputs of interest, E and κ_{border}, helping the APO strategy achieve its solution in real-time. This is due to the two stage *offline/online* approach. With this approach, the evaluations of E and κ_{border} need only invoke the *online* stage which saves much computational effort.

In order to solve the APO problem stated in Eq. 22, it would require solving first order and second order derivatives of our outputs of interest. The two stage *offline/online* approach also enable us to compute the derivatives by invoking the *online* stage alone. This, once again, saves us much valued computational resources.

Furthermore, the error bounds developed for the reduced-basis output bounds method play an important role in ensuring that the true solutions always satisfy the constraints of the APO formulation. In general, we can denote the error bound gap for some $\mu \in D_\mu$ as

$$H(\rho^l) = \bar{s}_I^+(\rho^l, \mu) - \bar{s}_I^-(\rho^l, \mu) \tag{29}$$

for $l = 1, ..., L$ and where $\bar{s}_I$ is the reduced-basis approximation of $s(\rho^l, \mu)$.

We can now write our reduced-basis feasible domain as

$$B' = \{\mu \in D_\mu | H(\rho^l) \cap I^l \neq \emptyset, l = 1, ..., L\} \tag{30}$$

We have proven that $B \subseteq B'$ and $\mu^* \in B'$ [11]. We can thus use B' to approximate B, and can define the new $\bar{\Upsilon}^y_{max}(\theta)$ and $\bar{\Upsilon}^y_{min}(\theta)$ for a certain output y as

$$\bar{\Upsilon}'^y_{max}(\theta) = \arg\max_{\mu \in B'} y(\theta, \mu)$$

and

$$\bar{\Upsilon}'^y_{min}(\theta) = \arg\min_{\mu \in B'} y(\theta, \mu)$$

$\forall \theta \in D_\theta$. As $\mu^* \in B'$, these new sets of arguments still yield results which bounds $h(\theta, \mu^*)$.

Applying these results to the evaluation of g and f in Eq. 20, we can see that

$$g(\theta, \bar{\Upsilon}'^g_{min}(\theta)) \leq g(\theta, \mu^*)$$

and

$$f(\theta, \bar{\Upsilon}'^f_{max}(\theta)) \geq f(\theta, \mu^*)$$

.

Therefore, the feasibility of the constraint, $g \geq C$, is not compromised. The evaluation of f ensures that the minimum value of f obtained will always approach the true minimum from above.

V. Optimization method

The APO formulation described in Eq. 20 can be looked at as a two level problem; an inner optimization problem to solve for $\bar{\Upsilon}^f_{max}(\theta)$ or $\bar{\Upsilon}^g_{min}(\theta)$, and an outer problem to solve for the minimum of $f(\theta)$. The inner problem is nonconvex due to the constraints described by the feasible domain B. Following the idea by Braken and McGill [19], we present a two level iterative approach to solve the APO formulation.

The outer level could be written as

$$\min_{\theta \in D_\theta} f(\theta, x) \tag{31}$$

$$\text{s.t. } g(\theta, z) \geq C$$

where $x \in \bar{\Upsilon}^f_{max}(\theta)$ and $z \in \bar{\Upsilon}^g_{min}(\theta)$ are treated as constants in the outer level.

This formulation easily fits the classical logarithmic barrier method [20] used to solve it with one difference. At each iteration, it will call upon the inner level as stated in Eq. 32 to obtain the values of x and z. Similarly to find the correct step length, the method needs to find the suitable values of x and z.

The inner level could be written in general as

$$\min_{\mu \in D_\mu} y(\theta, \mu) \tag{32}$$

$$\text{s.t. } s^l_{min} \leq s(\rho^l, \mu) \leq s^l_{max}, \, l = 1, ..., L$$

where y is the general output of interest. Here y represents both f and g.

Eq. 32 could easily be manipulated into a similar structure as Eq. 31 and we use the logarithmic barrier method to solve for it. However, the inner level is inherently nonconvex even if the functions y and s are convex and the feasible region might be disconnected or disjointed. A way to overcome this is to solve Eq. 32 for several different points. We call this way the multistart method. It is hoped that the method would help find the best local minimum.

The barrier method needs the algorithm to start from a feasible point. In order to find a feasible point we design another algorithm which tries to find a point in the center of B.

A. Feasible Point Algorithm

In the feasible point algorithm, we reformulate the inequality constraints into equality constraints. The inequality constraints represented by B can be written as

$$s(\rho^l, \mu) = s^l_{eq}$$

where $s^l_{eq} = (s^l_{max} + s^l_{min})/2$. The new objective function is

$$J(x) = \sum_{j=1}^{M} \frac{1}{2\nu} (s(\rho^l, \mu) - s^l_{eq})^2$$

.

This is an unconstraint minimization which can be solved using a primal-dual interior point method.

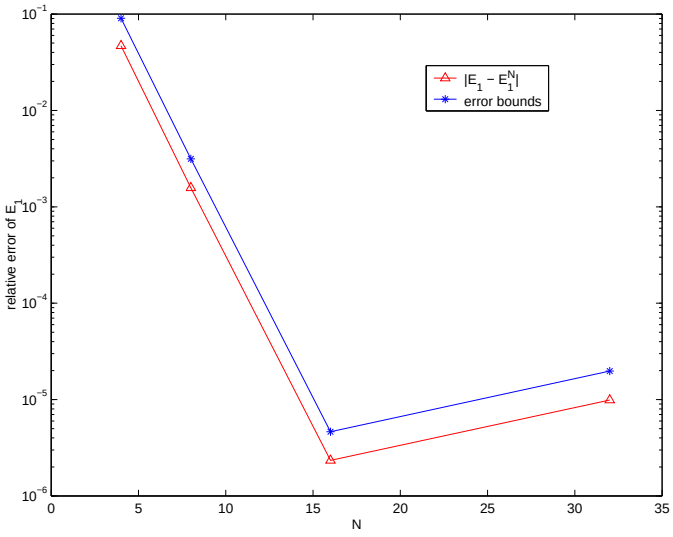

Fig. 2. The relative error convergence plot of E_1 vs N. Here $\gamma = 0.5$

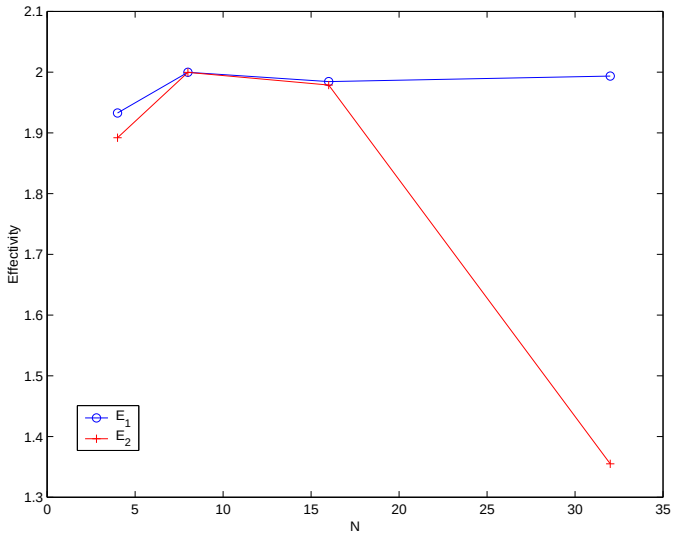

Fig. 3. The plot of effectivity, η, vs N.

B. General Description of the Overall Algorithm

The overall algorithm has three components: the algorithm to find feasible point, the inner level algorithm, and the outer level algorithm. The general algorithm solving Eq. 20 is detailed below.

1. Multistart: For a first guess θ, several $\mu \in D_\mu$ are supplied to the feasible point algorithm, and then to the inner level barrier algorithm to find their minimum points, $\bar{\mu}$. Selecting $\bar{\mu}$ which yields the minimum value of all $g(\theta, \bar{\mu})$, we obtain our $\bar{\Upsilon}^g_{min}(\theta)$. We do the same for f to find $\bar{\Upsilon}^f_{max}(\theta)$.
2. Outer level: The set of values; θ, $\bar{\Upsilon}^g_{min}(\theta)$, and $\bar{\Upsilon}^f_{max}(\theta)$ are supplied to the outer level which will find the next set of values by invoking the feasible point and inner level algorithms when finding its step length.
3. Step 2 is repeated until θ^* is found.

VI. Results and Discussion

We first demonstrate the efficiency and accuracy of the reduced-basis method. Choosing S^N randomly over $D_{h,\beta}$, we carry out the reduced-basis approximation for 50 different combinations of h, β and κ_0. We define the effectivity to be

$$\eta = \frac{\triangle}{|s - s_N|}$$

where $\triangle$ is the error gap, s is the true solution and s_N is the reduced-basis solution. Figure 2 shows the maximum error bounds and their respective true error,$|E_1 - E_1^N|$ relative to E_1 for the set of 40 parameter combinations. Here E_1 is as indicated in Eq. 11.

We see that the error bounds computed are larger than the true error. In fig. 3, η is roughly between 1 and 2 for both E_1 and E_2. The error bounds and the true error generally decrease as N increases. When N doubles from 16 to 32, the error increases due to round off error.

We compare the time taken to solve the finite element

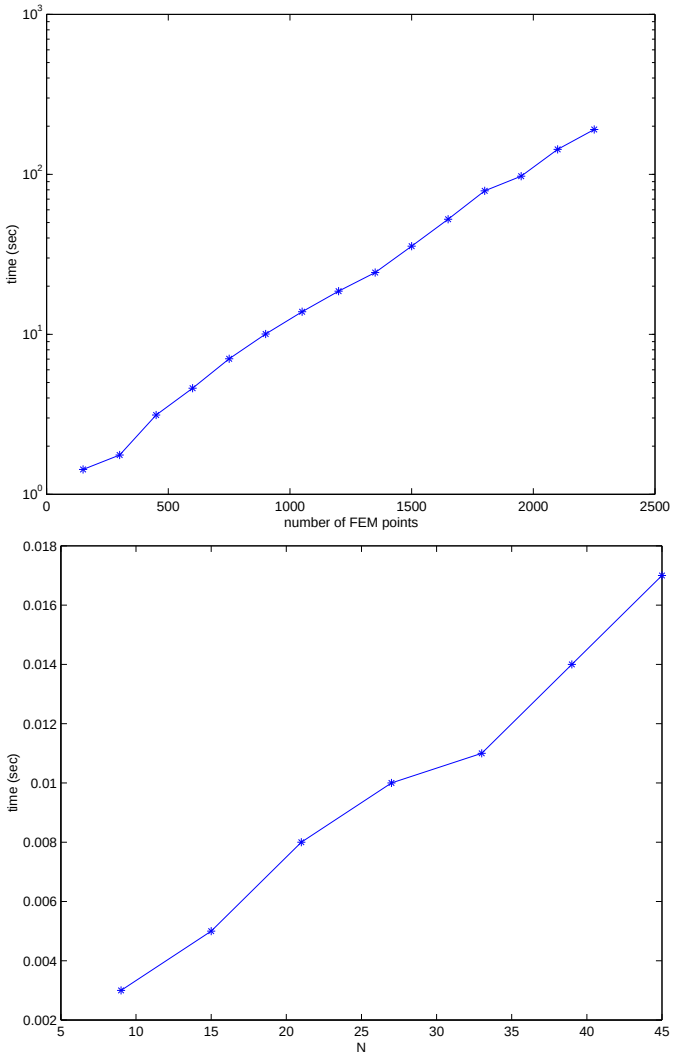

Fig. 4. Plot of the time taken with (a) increasing FEM dimension and (b) increasing number of basis N

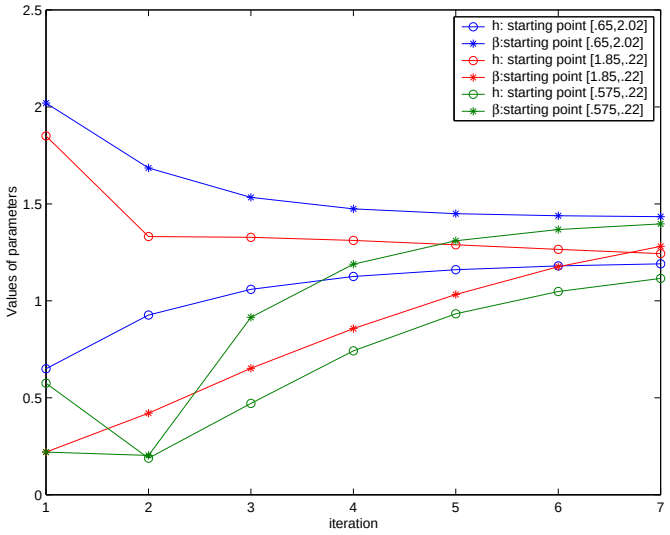

Fig. 5. Plot of the convergence of the feasible point algorithm for three different starting points of h and β

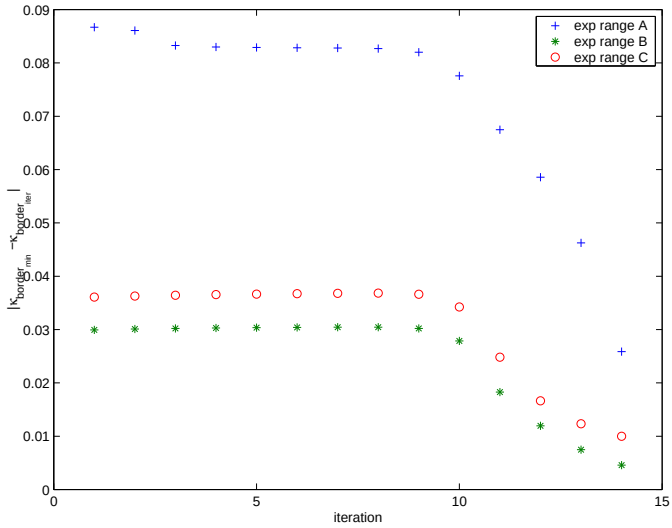

Fig. 6. Plot of the convergence to the minimum of κ_{border} in inner level algorithm

approximation and the reduced-basis approximation of Eq. 7. The theory of finite element indicates that as the dimension of finite element increases, the more accurate the approximation, and similarly for the reduced-basis theory. In fig. 4, we note that the reduced-basis method takes less than 1% of the time taken for the FEM solution. This increases the efficiency of the APO strategy.

We then look at the results from the feasible point algorithm. In fig. 5 we see that all three different starting points converge to similar values of h and β which were verified to be in the vicinity of the central point of the feasible region. Assuming that the central point is near the central path, the barrier method employed will have good convergence properties and the Hessian will be well-conditioned.

Figure 6 shows the convergence history to the minimum point for one instant of the inner level optimization for

κ^c	Number of online calls	Time (sec)
.3	72842	323
.35	71570	307
.4	69902	302
.45	71150	308
.5	75368	326

TABLE I

THE TOTAL NUMBER OF ONLINE CALLS OF THE APO ALGORITHM AND THE TOTAL TIME TAKEN FOR DIFFERENT κ^c BASED ON 5 MULTISTART POINTS.

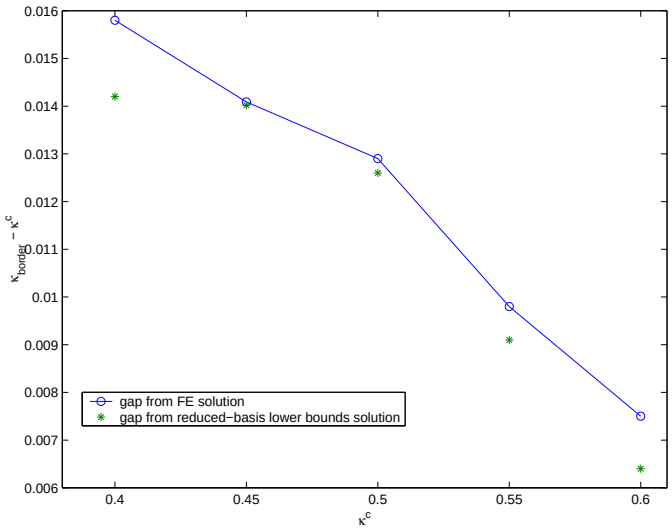

Fig. 7. Plots of $\kappa_{border} - \kappa^c$ for both the true κ_{border} and the κ^-_{border}.

three different feasible domain, B. Here the starting points of the inner level parameters are the points obtained from the feasible point algorithm. Yet, there has been unsuccessful attempts to find the optimal point, where the algorithm converges to stationary points. This is due to the general nonconvexity of the inner level problem. It is noted that about 70% of the attempts were successful in converging.

Table I shows the average total number of online calls to solve for E and κ_{border}, their first, and second order derivatives for use in the algorithm, as well as the time taken for the overall algorithm to converge to a solution. It takes roughly five minutes to compute 70000 reduced-basis approximations and solve for the Newton directions in both inner and outer levels. Given that the reduced-basis solution takes 1% of the time that FEM takes, the APO strategy would have taken several hours to converge using the FEM technique. We would like to decrease the time even further by finding ways to lessen the feedback process between the outer and inner levels.

From fig. 7, we see that the true solution of the optimal κ_0 of the APO strategy is always larger than the lower bounds of its reduced-basis solution and therefore never violates the outer level constraint given by κ^c. Thus the solution obtained by the APO formulation is always feasi-

ble.

VII. Conclusion and Future Work

We have shown that the reduced-basis output-bound method has allowed the APO formulation to be solved very fast and without violating the constraints. However, the method of solving the APO still needs improvement in terms of efficiency and accuracy.

Current algorithms to solve bilevel problems are still very focused on convex and linear problems, where the reformulation of the problems into a one level optimization problem with complimentarity constraints works quite well. We have tried such reformulation techniques to our problem but find that the values of $\bar{\Upsilon}_{min}(\theta)$ and $\bar{\Upsilon}_{max}(\theta)$ converged to are not the true minimum or maximum over B. We hope with advancement in this research area, the APO formulation can be solved using such techniques.

Currently we are finding ways to improve the feedback between the outer and inner levels to increase efficiency and accuracy. There is a subset of the problems we are considering in applying the APO strategy that allows the inner and outer levels to be decoupled[1]. These problems have monotonic functions with respect to their parameters. For example $s(\rho, \mu)$ can be monotonically increasing with respect to μ and decreasing with respect to ρ. Similarly with $g(\theta, \mu)$ and $f(\theta, \mu)$. The monotonicity of these functions gives rise to a single value of $\bar{\Upsilon}^f_{max}(\theta)$ and $\bar{\Upsilon}^g_{min}(\theta)$, which in turns allows the problem to be decoupled. The feedback is no longer necessary, and thus, we save on the computational effort to do the feedback loop.

We also hope to test the robustness of our APO strategy on higher dimensions of ρ and μ, and perhaps on more complex models with multiple objectives.

Acknowledgments

We would like to thank Dr. Ivan Oliveira for sharing his ideas on the optimization aspects of the APO; thanks also to Professor Robert Freund of MIT for his time and effort; and to Associate Professor Sun Jie and Associate Professor Teo Chung Piaw of NUS for their valuable advice. We would also like to acknowledge Dr. Christophe Prudhomme, Dr. Dimitrios Rovas, and Miss Karen L. Veroy for their valuable contributions in the Reduced-Basis Output-Bound Methods. This work was supported by the Singapore-MIT Alliance.

References

[1] M. Hanke. A regularizing levenberg-marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Problems*, 13:79–95, 1997.

[2] A. Neubauer. Tikhonov regularization of ill-posed linear operator equations on closed-convex sets. *J. Approx. Theory*, 53:304–320, 1988.

[3] M. Sambridge. Exploring multidimensional landscapes without a map. *Inverse Problems*, 14:427–440, 1998.

[4] R.B. Statnikov. *Multicriteria Design: Optimization and Identification.* Kluwer Academic Publishers, 1999.

[5] N. Marco, J.A. Desideri, and S. Lanteri. Multi-objective optimization in cfd by genetic algorithms. Technical report, Institut National De Recherie En Informatic Et En Automatique, 1999.

[6] C.M. Fonseca and P.J. Fleming. Multiobjective optimization and multiple constraint handling with evolutionary algorithms i: Unified formulation. Technical report, Dept. Automatic Control and Systems Eng., University of Sheffield, 1995.

[7] C.A.C. Coello and A.D. Christiansen. Two new approaches to multiobjective optimization using genetic algorithms. Preprint.

[8] V.T. Buchwald and F. Viera. Linearised evaporation from a soil of finite depth above a water table. *J. Austral. Math. Soc. Ser. B*, 39:557–576, 1998.

[9] J.R. Philip. Flow in porous media. *Annual Rev. Fluid Mech.*, 2:177–204, 1970.

[10] R. T. Waetcher and J. R. Philip. Steady two- and three-dimensional flows in unsaturated soil: The scattering analog. *Water Resources Research*, 21(12):1875–1887, Dec 1985.

[11] S. Ali. *Real-Time Optimal Parametric Design: The Assess-Predict-Optimize strategy.* PhD thesis, Nanyang Technological University, 2003. In progress.

[12] E. Haber, U.M. Ascher, and D. Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse Problems*, 16:1263–1280, 2000.

[13] A.K. Noor and J.M. Peters. Reduced basis technique for nonlinear analysis of structures. *AIAA Journal*, 18(4):455–462, April 1980.

[14] L. Machiels, Y. Maday, I. B. Oliveira, A.T. Patera, and D.V. Rovas. Output bounds for reduced-basis approximations of symmetric positive definite eigenvalue problems. *C. R. Acad. Sci. Paris, Série I*, 331(2):153–158, July 2000.

[15] Y. Maday, L. Machiels, A. T. Patera, and D. V. Rovas. Blackbox reduced-basis output bound methods for shape optimization. In *Proceedings* 12^{th} *International Domain Decomposition Conference*, pages 429–436, Chiba, Japan, 2000.

[16] A. T. Patera, D. Rovas, and L. Machiels. Reduced–basis output–bound methods for elliptic partial differential equations. *SIAG/OPT Views-and-News*, 11(1), April 2000.

[17] C. Prud'homme, D.V. Rovas, K. Veroy, L. Machiels, Y. Maday, A.T. Patera, and G. Turnici. Reduced–basis output bound methods for parametrized partial differential equations. In *Proceedings SMA Symposium*, January 2002.

[18] C. Prud'homme, D. Rovas, K. Veroy, Y. Maday, A.T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bounds methods. *Journal of Fluids Engineering*, 2002. To appear.

[19] J. Bracken and J.T. McGill. Defence applications, of mathematical programs with optimization problems in the constraints. *Operations Research*, 21:37–44, 1973.

[20] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer-Verlag New York, Inc., 1999.

[21] K. Veroy. *Reduced Basis Methods Applied to Problems in Elasticity: Analysis and Applications.* PhD thesis, Massachusetts Institute of Technology, 2003. In progress.

[1] Refer to [21]