Design Methodology of Modifier Adaptation for On-Line Optimization of Uncertain Processes

Eric A. Rodger ∗ Benoît Chachuat ∗

* Department of Chemical Engineering, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L7, Canada
(e-mail: benoit@mcmaster.ca).

Abstract: This paper is concerned with the on-line optimization of uncertain processes using the modifier-adaptation technology. In this approach, an optimization model is solved repeatedly and the available measurements are used to correct the values and gradients of the predicted outputs in that model. Following the paradigm of dual control, a successful modifier-adaptation scheme must reconcile two conflicting objectives, namely optimizing the process and obtaining accurate gradient information. During the on-line execution phase, a certain quality of the gradient estimates is enforced through additional constraints in the optimization problem. Then, a systematic off-line procedure is developed for the design of these constraints.

1. INTRODUCTION

The optimization of uncertain processes is traditionally based on a first-principles model that is updated in real-time using the available measurements. For those problems in which the optimization execution period is much longer than the closed-loop process dynamics, steady-state models are adequate to conduct the optimization (Darby and White, 1988; Marlin and Hrymak, 1997). Such on-line or real-time optimization (RTO) systems are well-accepted by industrial practitioners, and numerous successful applications of RTO in industrial practice have been reported.

The performance of an RTO system is influenced by two factors (Forbes and Marlin, 1996; Yip and Marlin, 2003): (i) the offset between the actual process optimum and the (expected) RTO performance, which is due to plant-model mismatch; and (ii) the variability of the RTO performance, caused by high frequency stationary disturbances propagating through the feedback RTO loop. With the usual two-step RTO systems, it is well-known that the lack of integration between the model-update and optimization steps can lead to large optimality losses. This has motivated the development of alternative RTO paradigms; see the recent reviews by Engell (2007); Chachuat et al. (2009).

The modifier-adaptation methodology (Marchetti et al., 2009) has been recently formulated by unifying and harmonizing various RTO technologies, including ISOPE (Roberts, 1995), constraint adaptation (Chachuat et al., 2008) and gradient adaptation (Gao and Engell, 2005). Similar to two-step RTO, the available process model is used by an online optimizer; the key difference being that the process measurements are now used to update so-called modifiers that are added to the predicted outputs in the optimization model, rather than (a subset of) the model parameters. This methodology greatly alleviates the offset with the actual process optimum. On the other hand, a modifier-adaptation RTO system requires that the plant output gradients be estimated from process measurements, which makes it more sensitive to noise than its two-step RTO analog.

In optimizing an uncertain process, a successful modifier-adaptation system must therefore accommodate two conflicting objectives: First, it must optimize the system as well as possible; second, it must ensure that enough information is known about the system to allow accurate and reliable gradient estimates. In essence, this problem is similar to the dual control problem (Wittenmark, 1995), which has been studied extensively since the early 1960s. It has been well-known that the exact dual-control problem for nonlinear systems is computationally intractable, in response to which a variety of practical, yet suboptimal, dual controllers have been developed.

In this work, an RTO algorithm that combines modifier adaptation with a gradient-update scheme based on Broyden’s method is considered. Following the same ideas as in dual control, a dual modifier-adaptation system is developed which reconciles the conflicting goals of optimization and gradient estimation. Our methodology consists of an off-line design phase, followed by the on-line execution phase:

- During the on-line phase, a certain quality of the gradient estimates is enforced through additional constraints in the optimization model. Such constraints are used to keep both the gradient offset and gradient variance small.
- The off-line phase is concerned with the shaping of the gradient quality constraint. Our approach builds upon the so-called design cost procedure by Forbes and Marlin (1996), and extends it to dual modifier adaptation.

The paper is organized as follows. Background on modifier adaptation and a formulation of the combined modifier-adaptation-with-Broyden-update scheme are given in §2. The dual modifier-adaptation problem is presented in §3 from an on-line execution perspective. Then, the off-line design of dual modifier adaptation is investigated in §4. Finally, conclusions are presented in §5.

2. MODIFIER ADAPTATION WITH BROYDEN UPDATES

Consider an optimization problem of the form:
\[ \min_{u} \phi(u,y^p) \quad (P^p) \]
\[ \text{s.t.} \quad y^p = F(u) \]
\[ g(u,y^p) \leq 0 \]
\[ u^\min \leq u \leq u^\max, \]

where \( u \in \mathbb{R}^{n_u} \) denotes the decision (or input) variables; \( y^p \in \mathbb{R}^{n_y} \), the process outputs; \( F : \mathbb{R}^{n_u} \to \mathbb{R}^{n_y} \), the process input-output map; \( \phi : \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R} \), the scalar objective function; and \( g : \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \to \mathbb{R}^{n_g} \), the vector of inequality constraints.

For uncertain processes, the exact map \( F \) is unknown, and only an approximate model is available such as:
\[ y = f(u, \theta), \]

where \( \theta \in \mathbb{R}^{n_{\theta}} \) is a set of adjustable model parameters. An approximate solution to the original problem \((P^p)\) can then be obtained by solving the model-based optimization problem:
\[ \min_{u} \phi(u,y) \quad (P) \]
\[ \text{s.t.} \quad y = f(u, \theta) \]
\[ g(u,y) \leq 0 \]
\[ u^\min \leq u \leq u^\max, \]

In the presence of model mismatch and process disturbances, however, the optimal solution value of \((P)\) may be quite different from that of \((P^p)\). On-line optimization takes advantage of the available measurements in order to compensate for such uncertainty and adapt the model-based problem in order to get closer to the actual process optimum. In a modifier-adaptation system (Chachuat et al., 2009; Marchetti et al., 2009), the measurements are used to correct the values and the first-order derivatives of the outputs in the model-based optimization problem \((P)\). At a given operating point \(u^0\), the idea is to determine an improved operating point by solving the following modified optimization problem:
\[ \min_{u} \phi(u,y^\delta) \quad (\tilde{P}) \]
\[ \text{s.t.} \quad y^\delta = f(u, \theta) + \varepsilon + \lambda^T(u - u^0) \]
\[ g(u,y^\delta) \leq 0 \]
\[ u^\min \leq u \leq u^\max, \]

where \( \varepsilon \in \mathbb{R}^{n_y} \) is the vector of output modifiers, and \( \lambda \in \mathbb{R}^{n_u \times n_y} \) the matrix of output-gradient modifiers.

The use of modifiers is appealing in that a KKT point \( u^\infty \) for the corrected model-based problem \((\tilde{P})\) is also a KKT point for the original problem \((P^p)\), provided that the modifiers for each output \( i = 1, \ldots, n_y \) satisfy (Marchetti et al., 2009):
\[ \varepsilon_i = f_i(u^\infty) - f_i(u^\infty, \theta) \]
\[ \lambda_i = \partial f_i(u^\infty, \theta)/\partial u_i \]

An iterative scheme that adapts the modifiers in such a way that the foregoing conditions are satisfied on convergence can be devised as follows:
\[ \Lambda^{k+1} = (I - K)\Lambda^k + K \]
where \( \Lambda = [\varepsilon_1 \lambda_1^T \cdots \varepsilon_{n_y} \lambda_{n_y}^T] \in \mathbb{R}^{n_A} \), with \( n_A = n_y(n_y + 1) \), is the vector of modifiers, and the gain matrix \( K \in \mathbb{R}^{n_A \times n_A} \) can be conveniently chosen as:
\[ K = \text{diag}(k_{\varepsilon_1}, k_{\lambda_1} I_{n_u}, \ldots, k_{\varepsilon_n}, k_{\lambda_n} I_{n_u}), \]

with entries \( k_{\varepsilon_i}, k_{\lambda_i} \in (0, 1], i = 1, \ldots, n_y \). The next operating point \( u^{k+1} \) is obtained from the current point \( u^k \) and modifiers \( \Lambda^k \) by solving the modified optimization problem,
\[ u^{k+1} \in \arg \min_{u} \phi(u,y^\delta) \quad (1) \]
\[ \text{s.t.} \quad y = f(u, \theta) + \varepsilon^k + \lambda^T(u - u^k) \]
\[ g(u,y) \leq 0 \]
\[ u^\min \leq u \leq u^\max. \]

Perhaps the biggest burden with this approach is estimating the gradients of the plant outputs with respect to the plant inputs, \( \partial y_i/\partial u_i, i = 1, \ldots, n_y \). Several methods have been developed over the last 30 years to conduct this task. Finite difference methods and linear system identification methods can be applied, yet they have the attendant disadvantage of requiring additional perturbations. A method that relies on past operating points instead of extra points was proposed by Brydys and Tatjewski (1994), and further developed by Gao and Engell (2005) and Marchetti et al. (2010). Concurrently, the use of Brydys’ method, a recursive rank-1 update, was investigated by Roberts (2000). In a comparison study conducted on a chemical reactor system, Mansour and Ellis (2003) found Brydys’ method to perform well. This method is the one used subsequently to get estimates \( \Delta_i \approx \partial y_i/\partial u_i \) of the plant output gradients in the following recursive way:
\[ \Delta_i^{k+1} = B(\Delta_i^k, u^{k+1}, y_i, p^{k+1}) = \Delta_i^k + \partial f_i(u^k, \theta)/\partial u_i \]
\[ \hat{y}_i^{k+1} - y_i^{k+1} - \Delta_i^T [u^{k+1} - u^k] / [u^{k+1} - u^k]. \]

Overall, the modifier-adaptation scheme is given by (1) and (2), along with the modifier update law:
\[ \Delta_i^{k+1} = (I - K)\Delta_i^k + K \]
\[ \Delta_i^k = \partial f_i(u^k, \theta)/\partial u_i. \]

This scheme can be initialized by supplying an initial point \( u^0 \), initial modifiers \( \Delta^0 \), typically \( \Delta^0 = 0 \), and initial gradient estimates \( \Delta^0 \), typically \( \Delta^0 = \partial f_i(u^0, \theta)/\partial u_i \).

3. DUAL MODIFIER ADAPTATION: ON-LINE SCHEME

Unlike the ideal modifier-adaptation algorithm with exact output gradients (Marchetti et al., 2009), the modifier-adaptation algorithm (1)-(3) with recursive gradient updates suffers from two deficiencies: (i) the converged operating point, in a noiseless environment, may not correspond to a KKT point of \((P^p)\); and (ii) the estimated output gradients, in the presence of noise, can take arbitrarily large values thereby leading to a peaking phenomenon. The following example illustrates these deficiencies.
Example 1. Consider the optimization problem

\[
\begin{align*}
\min_u & \quad y_1 \\
n & \quad y_1 = u_1^2 + u_2^2 + \theta_1 u_1 u_2 \\
n & \quad y_2 = 1 - u_1 + u_2^2 + \theta_2 u_2 + \theta_3 \\
n & \quad y_2 \leq 0,
\end{align*}
\]

with the (unknown) parameter values \( \theta^p = [1 \ 2 \ -2] \). The optimum for this problem (and the only KKT point here) is \( u^* = 0 \) and the inequality constraint is inactive.

The parameter values for the model are chosen as \( \theta = [0 \ 0 \ 0] \), and the modifier-adaptation scheme (1)-(3) is applied with the initial values \( u^0 = [\frac{1}{2} \ \frac{1}{2}]^T \), \( \lambda^0 = 0 \), \( \Delta_1^0 = [1 \ 1] \), \( \Delta_2^0 = [-1 \ 1] \), and the gain matrix \( K = \frac{1}{2} I_6 \). The iterates drawn in magenta on Fig. 1 are obtained with noiseless outputs. They converge to the suboptimal point \( u_\infty \approx [-0.0512 \ 0.1978]^T \), with \( \Delta_1^\infty = 0 \) albeit \( \frac{\partial \Gamma}{\partial \nu}(u_\infty) = [0.0954 \ 0.3444] \). Also represented on this figure in blue are the iterates obtained when \( y_1^p = 0 \) and \( y_2^p = 0 \) are corrupted with a white noise of variance \( \sigma_{y_1}^2 = \sigma_{y_2}^2 = 10^{-3} \). It is seen that the first few iterates are close to those obtained in the noiseless scenario, before suddenly deviating from this path due to gradient peaking.

Failure of the scheme (1)-(3) to reach a KKT point upon convergence can be attributed to inaccurate output-gradient estimates. This may only happen for multiple-input problems, namely when the estimates \( \Delta_i \) do not get repeatedly updated in all directions. Consider the gradient offset \( e_i \) for the \( i \)th output defined as \( \Delta_i = \frac{\partial \Gamma}{\partial u_i} + e_i \). At iteration \( k + 1 \), the component of this offset in the unit direction \( \delta^{k+1} \Delta_i \) is given by:

\[
e_i^{k+1} \delta^{k+1} = -\frac{1}{2} (u^{k+1} - u^k)^T \frac{\partial^2 \Gamma}{\partial u^2} (u^{k+1}) \delta^{k+1} + O(\|u^{k+1} - u^k\|^2).
\]

On the other hand, the component of the gradient offset in any direction \( n \) orthogonal to \( [u^{k+1} - u^k] \) is:

\[
e_i^{k+1} n = e_i^k n - (u^{k+1} - u^k)^T \frac{\partial^2 \Gamma}{\partial u^2} (u^{k+1}) n + O(\|u^{k+1} - u^k\|^2).
\]

The following recursive formula is obtained by combining these two components:

\[
e_i^{k+1} = \begin{bmatrix} I - \delta^{k+1} \delta^{k+1} \end{bmatrix} e_i^k - (u^{k+1} - u^k)^T \frac{\partial^2 \Gamma}{\partial u^2} (u^{k+1}) \left[ I - \frac{1}{2} \delta^{k+1} \delta^{k+1} \right] + O(\|u^{k+1} - u^k\|^2).
\]

Quite expectedly, the offset growth appears to be linear in the step \( [u^{k+1} - u^k] \), so taking small steps is a prerequisite for accurate gradient estimation. The formula (5) also confirms that the gradient offset component may not vanish in directions orthogonal to the step direction upon convergence. For example, when the iterates in a two-input problem converge along a given line, the gradient component in the direction orthogonal to that line is not updated and is therefore biased; cf. Example 1.

The second deficiency arises when the distance between two successive points \( u^k \) and \( u^{k+1} \) becomes small, but because of the measurement noise \( \nu_i \) the corresponding distance between \( y_i^{p,k} = F_i(u^k) + \nu^k \) and \( y_i^{p,k+1} = F_i(u^{k+1}) + \nu^{k+1} \) does not vanish. It follows that the gradient estimate in the direction \( \delta^{k+1} \) can take on arbitrarily large values as \( \|u^{k+1} - u^k\| \).

\[
\Delta_i^{k+1} \delta^{k+1} = \frac{y_i^{p,k+1} - y_i^{p,k}}{\|u^{k+1} - u^k\|}.
\]

To analyze the estimated gradient covariance, white noise (not necessarily Gaussian) is assumed subsequently, with variance \( \sigma_{y_i}^2 \) for the \( i \)th output. It is not hard to establish that a first-order approximation of the covariance matrix of the output-gradient estimate \( \Delta_i \) is given by the following recursive formula:

\[
\begin{align*}
\forall(\Delta_i^{k+1}) & \approx \begin{bmatrix} I - \delta^{k+1} \delta^{k+1} \end{bmatrix} \forall(\Delta_i^k) \begin{bmatrix} I - \delta^{k+1} \delta^{k+1} \end{bmatrix}^T + \sigma_{y_i}^2 \left( \frac{2\delta^{k+1} \delta^{k+1} \Gamma_i \delta^{k+1} \Gamma_i \delta^{k+1} \delta^{k+1} \Gamma_i}{\|u^{k+1} - u_i^k\|^2} \right) \\
& \approx \begin{bmatrix} I - \delta^{k+1} \delta^{k+1} \end{bmatrix} \delta^{k+1} \delta^{k+1} + \delta^{k+1} \delta^{k+1} \left[ I - \delta^{k+1} \delta^{k+1} \right] \\
& \approx \left[ \begin{array}{rr} \delta^{k+1} \delta^{k+1} & \delta^{k+1} \delta^{k+1} \\
\delta^{k+1} \delta^{k+1} & \delta^{k+1} \delta^{k+1} \end{array} \right].
\end{align*}
\]

Briefly, the gradient covariances at iterations \( k \) and \( k + 1 \) are equal at first order, except in the direction \( \delta^{k+1} \) where the latest rank-one update has been performed. Apart from the choice of the direction \( \delta^{k+1} \), the magnitudes of the current step \( \|u^{k+1} - u^k\| \) and of the previous step \( \|u^{k+1} - u^{k-1}\| \) appear to have a large influence on this covariance. In particular, \( \forall(\Delta_i^{k+1}) \) is inversely proportional to the steps \( \|u^{k+1} - u^k\| \) and \( \|u^{k+1} - u^{k-1}\| \), which supports the intuition that large enough steps are required to keep the gradient variance small.

It is paramount to account for the aforementioned deficiencies in devising a successful modifier-adaptation system. In this work, a dual approach is considered, whereby a certain quality of the gradient estimates is enforced through extra constraints in the optimization problem (1). By quality, it is understood that both the offset and the variance of the gradient estimates should remain small simultaneously.

- To maintain a small gradient variance, the following constraint is considered:
Fig. 2. Geometrical interpretation of the dual constraints.

\[ [u - u^k]^T \Sigma^T \Sigma^{-1} [u - u^k] \geq 1, \] (6)

which defines an ellipsoidal exclusion region around the current iterate \( u^k \). Observe that the size of the ellipse is controlled by the entries of \( \Sigma \) and the constraint (6) is always nonconvex.

- To maintain a small gradient offset, the step length should be not too large and the Broyden update should be made repeatedly in every direction. The first condition is easily enforced, e.g., by defining a trust region as:

\[ [u - u^k]^T [\Gamma^T \Gamma]^{-1} [u - u^k] \leq 1. \] (7)

In so doing, special care must be taken that the chosen matrices \( \Sigma \) and \( \Gamma \) in (6) and (7), respectively, do not make the optimization problem infeasible. In particular, for those outputs corrupted by a high level of noise \( \sigma^2_{yi} \), it may not always be possible to find an acceptable compromise leading to both a small offset error \( e_\ell \) and a small gradient variance \( \nabla \phi(\Delta_i) \).

Meeting the second condition is more involved. Let the number of past iterates \( k \) be greater than or equal to \( n_u \), and suppose that the \( (n_u - 1) \) previous search directions \( [u^k - u^{k-1}], \ldots, [u^{k-n_u+2} - u^{k-n_u+1}] \) are linearly independent. Then, it can be shown that a sufficient condition for the next direction \( [u - u^k] \) to not be in the hyperplane

\[ H_k \triangleq \text{span}([u^k - u^{k-1}], \ldots, [u^{k-n_u+2} - u^{k-n_u+1}]) \]

and, at the same time, the variance condition (6) to be satisfied at \( u \) is:

\[ \left\{ \begin{array}{l}
\pm \alpha^k [u - u^k] \geq \sqrt{\alpha^k T \Sigma^T \Sigma \alpha^k} \\
-\alpha^k [u - u^k] \geq \sqrt{\alpha^k T \Sigma^T \Sigma \alpha^k} \end{array} \right\}, \] (8)

where \( \alpha^k \) is any nonzero vector orthogonal to those \( (n_u - 1) \) directions; for example, \( \alpha^k \) can be taken as

\[ [\text{adj} U^k]_{(1,1)} \]

the first row of the adjugate of

\[ U^k \triangleq [u - u^k, u^k - u^{k-1}, \ldots, u^{k-n_u+2} - u^{k-n_u+1}] \]

— which is independent of \( u \).

A geometrical interpretation of the constraints (6), (7) and (8) in the 2-d case is given in Fig. 2. Due to the disjunctive nature of (8), two modified optimization problems are solved at each execution of dual modifier adaptation, so that no extra nonconvexity is introduced by the dual constraints:

\[ u^{k+1}_\pm \in \arg \min_u \phi(u, \hat{y}) \]

s.t. \[ \hat{y} = f(u, \theta) + e^k + \lambda^k [u - u^k) \]

\[ g(u, \hat{y}) \leq 0 \]

\[ \pm \alpha^k [u - u^k] \geq \sqrt{\alpha^k T \Sigma^T \Sigma \alpha^k} \]

\[ [u - u^k]^T [\Gamma^T \Gamma]^{-1} [u - u^k] \leq 1 \]

then, the next point \( u^{k+1}_\pm \) is simply chosen as \( u^{k+1}_+ \) or \( u^{k+1}_- \), whichever one yields the best performance. Overall, the dual modifier-adaptation scheme consists of the gradient, modifier and input adaptation laws (2), (3) and (9), respectively.

**Example 1 (continued).** Two scenarios are investigated, where the unknown parameter values are chosen as \( \theta^p = [1 2 -2] \) (same as before) and \( \theta^p = [1 2 0] \); the latter set of parameters gives rise to a constrained optimum. Moreover, the outputs \( y^p_1 \) and \( y^p_2 \) are corrupted with a white noise of variance \( 10^{-3} \).

The dual modifier-adaptation scheme is applied to Problem (4), with \( \Sigma = \text{diag}(0.1, 0.1) \) and \( \Gamma = \text{diag}(0.5, 0.5) \). The resulting iterates are displayed in Fig. 3 for both scenarios. It is seen that the iterates are distributed around the actual optima. Moreover, the variance around the optima results from the combination of two effects, namely the output noise and the dual constraints (8). The choice of \( \Sigma \) is critical in this respect.

4. DUAL MODIFIER ADAPTATION: OFF-LINE DESIGN

The biggest issue in devising an efficient dual modifier-adaptation scheme remains selecting good values for its design parameters. Regarding the choice of \( \Sigma \) in particular, there appears to be an optimal trade-off. Roughly speaking, small entries in \( \Sigma \) limit the disturbances imposed by the dual constraints (8), but make the gradient estimates—and therefore the adaptation scheme—more sensitive to noise. On the other hand, large entries reduce the variance of the gradient estimates, but then (8) produces large input variations at each iteration. Clearly, such a trade-off is problem dependent and some kind of systematic procedure is needed to guide this selection.

This problem can be cast as an optimization problem, whereby the matrix \( \Sigma \) that minimizes the expected cost of the dual modifier-adaptation system is sought. In general, getting the minimum possible cost requires varying \( \Sigma \) from one iteration to the next, especially during major transient phases (e.g., after a large disturbance). Unfortunately, such an optimization problem—in essence a stochastic dynamic program—is seldom tractable from an on-line perspective.

The main focus in this section is on steady-state performance; that is, transient phases are ignored for simplicity. This simplification makes it possible to address the problem from an off-line perspective, by solving the following optimization problem:

\[ \min_{\Sigma} \mathbb{E}[\Phi^p(u^\infty)], \] (10)

where \( u^\infty \) are the dual modifier-adaptation iterates around the steady-state optimum, and \( \Phi^p(\cdot) \triangleq \phi(\cdot, \mathcal{F}(\cdot)) \). Recall that the variations in \( u^\infty \) are caused by both the output noise and the dual excitation.

Using a Taylor expansion of \( \Phi \) around \( \mathbb{E}[u^\infty] \) and neglecting terms higher than second order, (10) can be approximated by:
\[ \min_{\Sigma} \Phi_p(\mathbb{E}[u^\infty]) + 1^T \left[ \frac{1}{2} \partial^2 \Phi_p(\mathbb{E}[u^\infty]) \circ \mathbb{V}(u^\infty) \right] 1, \quad (11) \]

where \( 1 \) is a vector of ones, and \( \circ \) stands for the Hadamard matrix product. The first and second terms in this expression represent the effect of, respectively, the expected value (offset) and of the variance the iterates on the cost.

Solving (11) requires that approximate expressions be available for \( \mathbb{E}[u^\infty] \) and \( \mathbb{V}(u^\infty) \). Because the dual constraints (8) fail to be differentiable when the entries in \( \Sigma \) vanish, direct linearization of the dual modifier-adaptation map to derive such approximations is not possible here. A simplification is adopted subsequently, whereby the effects of the output and gradient noise, on the one hand, and of the dual excitation, on the other hand, are treated separately and then simply added up, even though these effects are intrinsically coupled. This simplification turns out to be accurate when the dual excitation is large compared to the output noise, but progressively breaks down as the dual excitation is reduced.

An important assumption is that the ideal modifier-adaptation scheme—with exact gradient information and no noise—converges for a given gain matrix \( K \). This assumption is not very restrictive because a modifier-adaptation scheme that would fail to converge under such ideal conditions would also typically fail to converge with approximate gradient estimates and noise. In this case, the entries in the gain matrix \( K \) can always be reduced to promote convergence. Moreover, the convergence point is assumed to be both regular and a strict local minimum of (1); this point is denoted by \( u^* \) hereafter.

- The ideal modifier-adaptation scheme with dual excitation and no noise is considered first. The dual constraints (8) act by distributing the iterates \( \{u_k\} \) around \( u^* \). A further effect of these constraints is to shift the expected value \( \mathbb{E}[u^\infty] \) of the iterates away from \( u^* \) as \( \Sigma \) increases. This latter effect is especially sensitive when constraints are active at \( u^* \) since the optimizer tries to reduce infeasibility by shifting \( u^* \) inside the feasible region.

To conduct the analysis, it is assumed as a first approximation that the iterates are equiprobably distributed on the ellipsoid \( (u^\infty - \mathbb{E}[u^\infty])^T \Sigma (u^\infty - \mathbb{E}[u^\infty]) = 1 \). It follows directly from this assumption that the iterate variance is given by:

\[ \mathbb{V}(u^\infty) \approx \frac{1}{2} \Sigma^T \Sigma. \]

An approximation of \( \mathbb{E}[u^\infty] \) can also be obtained by noting that the largest violation for the \( j \)th inequality constraint occurs when the iterates are excited in the direction \( \frac{\partial g_j}{\partial u} (u^*) \). In this worst-case scenario, preserving feasibility requires that each inequality constraint be backed-off as follows:

\[ g_j(u) + \frac{\partial g_j}{\partial u} (u^*)^T \frac{\partial g_j}{\partial y} (u^*) \leq 0. \]

The resulting change in optimum corresponds to \( \Delta u^* \triangleq \mathbb{E}[u^\infty] - u^* \). Using NLP sensitivity theory (Fiacco, 1983), this change can be easily approximated by a linear relationship of the form \( \Delta u = C^* \gamma \), so that:

\[ \mathbb{E}[u^\infty] \approx u^* + C^* \gamma(\Sigma). \]

- Next, the ideal modifier-adaptation scheme without dual excitation is considered. The effect of white measurement noise of zero mean and variance \( \sigma^2_n \) is taken into account for each output \( y_i \). Moreover, the analysis assumes that measurements are also available for the corresponding output gradient \( \Delta_i \), with white noise of zero mean and variance \( \mathbb{V}(\Delta_i) = \sigma^2_{\Delta_i} \).

A first-order approximation of the iterates around \( u^* \) is given by:

\[ \begin{bmatrix} \delta u_{k+1}^1 \\ \delta \lambda_{k+1}^1 \\ \vdots \\ \delta u_{n_y}^k \\ \delta \lambda_{n_y}^k \end{bmatrix} \approx A^* \begin{bmatrix} \delta u_k^1 \\ \delta \lambda_k^1 \\ \vdots \\ \delta u_{n_y}^k \\ \delta \lambda_{n_y}^k \end{bmatrix} + B^* \begin{bmatrix} \delta y_{k+1}^1 \\ \delta \Delta_{k+1}^1 \\ \vdots \\ \delta y_{n_y}^{k+1} \\ \delta \Delta_{n_y}^{k+1} \end{bmatrix}. \quad (12) \]

Using (12) and starting with the values \( \delta u^0 = 0 \) and \( \delta \lambda_i^0 = \cdots = \delta \lambda_{n_y}^0 = 0 \), the following approximation can be obtained for the iterates covariance:
\[ V(u^*) \approx P_u \left[ \sum_{i=0}^{\infty} (A^*)^i B^* S^*(\Sigma) B^* T(A^T)^i \right] P_u^T, \quad (13) \]

with: \( S^*(\Sigma) = \text{diag}(\ldots, \sigma^2_{y_1}, 2\sigma^2_{y_1}[\Sigma\Sigma^{-1}] \ldots) \).

Note that the infinite sum in (13) converges when the ideal modifier-adaptation scheme converges.

In sum, an approximation of the design problem (10) for dual modifier adaptation is given by:

\[ \min_{\Sigma} \Phi_p(u^* + C^* \gamma(\Sigma)) + 1^T \left[ \frac{1}{2} \partial^2 \Phi_p}{\partial u^2} (u^* + C^* \gamma(\Sigma)) \right] \]

\[ + \left( P_u^T \left[ \sum_{i=0}^{\infty} (A^*)^i B^* S^*(\Sigma) B^* T(A^T)^i \right] P_u + \frac{1}{2} \Sigma \Sigma \right) \].

Ideally, the optimum point \( u^* \) (and the Hessian matrix \( \partial^2 \Phi_p}{\partial u^2} \) at \( u^* \) should be estimated through plant experimentation. When no such experimentation is possible, an alternative is to use detailed models, for example models which are too complex for on-line implementation. In this case, the effect of model mismatch on the design problem can be alleviated by solving a multi-scenario optimization problem, as discussed in Forbes and Marlin (1996).

Example 1 (continued). The design procedure for dual modifier adaptation is applied to Problem (4), in order to choose \( \Sigma \) in the dual excitation constraint (8). For simplicity, the case of a diagonal matrix \( \Sigma = \sigma^2 I_q \) is considered. The two scenarios where \( \theta^p = [1, 2, -2] \) (unconstrained optimum) and \( \theta^p = [1, 2, 0] \) (constrained optimum) are investigated, and both outputs are corrupted with a white noise of variance \( 10^{-3} \).

The objective functions in the design problem (10) and in the approximate problem (14) are plotted versus the parameter \( \sigma_u \) in Fig. 4. In both scenarios, it is found that the approximate problem (14) accurately predicts the expected performance and the value \( \sigma^2_u \), at which the minimum expected cost is attained. This case study also confirms that the design of the dual excitation constraint is critical and that inappropriate choices for \( \Sigma \) can lead to poor performance.

5. CONCLUSIONS

A design methodology for dual modifier adaptation has been presented in this paper. It has been shown how a certain quality of the gradient estimates can be enforced during the on-line execution through additional constraints in the optimization problem. It has also been shown how the design of the dual excitation constraints can be taken care of off-line, via the solution of a simpler optimization problem. This methodology applies to both unconstrained and constrained problems and has been demonstrated on a simple case study.

REFERENCES