

Implementation techniques for the SCFO experimental optimization framework

Gene A. Bunin, Grégory François, Dominique Bonvin

Submitted: June 17, 2014

Abstract The material presented in this document is intended as a comprehensive, implementation-oriented supplement to the experimental optimization framework presented in [10]. The issues of physical degradation, unknown Lipschitz constants, measurement/estimation noise, gradient estimation, sufficient excitation, and the handling of soft constraints and/or a numerical cost function are all addressed, and a robust, implementable version of the sufficient conditions for feasible-side global convergence is proposed.

Keywords experimental optimization · SCFO · estimation of Lipschitz constants · gradient estimation · optimization of degrading processes · sufficient excitation for optimization · optimization of noisy functions

1 Overview

Following the conventions that were already established in [10], let us define an *experimental optimization* problem as

$$\begin{aligned} & \underset{\mathbf{u}}{\text{minimize}} \quad \phi_p(\mathbf{u}) \\ & \text{subject to} \quad g_{p,j}(\mathbf{u}) \leq 0, \quad j = 1, \dots, n_{g_p} \\ & \quad \quad \quad g_j(\mathbf{u}) \leq 0, \quad j = 1, \dots, n_g \\ & \quad \quad \quad \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U, \end{aligned} \quad (1)$$

with the subscript p used to denote *experimental functions* that can only be evaluated for a given set of decision variables, $\mathbf{u} \in \mathbb{R}^{n_u}$, by carrying out a (presumably expensive) experiment. Apart from the standard task of finding a set \mathbf{u}^* that is a local minimum of (1), experimental optimization problems come with the additional requirement that

Laboratoire d'Automatique, Ecole Polytechnique Fédérale de Lausanne, Lausanne, CH-1015
E-mail: gene.a.bunin@ccapprox.info

the constraints be satisfied *for all* experimental iterates $\mathbf{u}_0, \mathbf{u}_1, \mathbf{u}_2, \dots$ that are tested in the optimization process.

In the companion work of [10], a theoretical framework – referred to hereafter as the SCFO (“sufficient conditions for feasibility and optimality”) framework – was proposed, and it was proven that enforcing the SCFO

$$g_{p,j}(\mathbf{u}_k) + \sum_{i=1}^{n_u} \kappa_{p,ji} |u_{k+1,i} - u_{k,i}| \leq 0, \quad \forall j = 1, \dots, n_{g_p}, \quad (2)$$

$$g_j(\mathbf{u}_{k+1}) \leq 0, \quad \forall j = 1, \dots, n_g, \quad (3)$$

$$\mathbf{u}^L \preceq \mathbf{u}_{k+1} \preceq \mathbf{u}^U, \quad (4)$$

$$\nabla g_{p,j}(\mathbf{u}_k)^T (\mathbf{u}_{k+1} - \mathbf{u}_k) \leq -\delta_{g_p,j}, \quad \forall j : g_{p,j}(\mathbf{u}_k) \geq -\varepsilon_{p,j}, \quad (5)$$

$$\nabla g_j(\mathbf{u}_k)^T (\mathbf{u}_{k+1} - \mathbf{u}_k) \leq -\delta_{g,j}, \quad \forall j : g_j(\mathbf{u}_k) \geq -\varepsilon_j, \quad (6)$$

$$\nabla \phi_p(\mathbf{u}_k)^T (\mathbf{u}_{k+1} - \mathbf{u}_k) \leq -\delta_\phi, \quad (7)$$

$$\nabla \phi_p(\mathbf{u}_k)^T (\mathbf{u}_{k+1} - \mathbf{u}_k) + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} M_{\phi,i_1 i_2} |(u_{k+1,i_1} - u_{k,i_1})(u_{k+1,i_2} - u_{k,i_2})| \leq 0 \quad (8)$$

for the future experimental iterate \mathbf{u}_{k+1} at every experimental iteration k would generate a chain of iterates that were all feasible and converged to a point \mathbf{u}_∞ in a finite number of experiments, with $\mathbf{u}_\infty \rightarrow \mathbf{u}^*$ in the Fritz-John-error sense as the positive projection parameters $\varepsilon_p, \varepsilon, \delta_{g_p}, \delta_g, \delta_\phi \downarrow \mathbf{0}$. Here, the constants $\kappa_{p,ji}$ and $M_{\phi,i_1 i_2}$ are used to denote Lipschitz constants for the experimental

constraints and the Lipschitz constants for the cost function derivatives, respectively, with

$$-\kappa_{p,ji} < \left. \frac{\partial g_{p,j}}{\partial u_i} \right|_{\mathbf{u}} < \kappa_{p,ji}, \quad \forall \mathbf{u} \in \mathcal{S}, \quad (9)$$

$$-M_{\phi,i_1i_2} < \left. \frac{\partial^2 \phi_p}{\partial u_{i_2} \partial u_{i_1}} \right|_{\mathbf{u}} < M_{\phi,i_1i_2}, \quad \forall \mathbf{u} \in \mathcal{S}, \quad (10)$$

and with the experimental space \mathcal{S} defined as

$$\mathcal{S} = \{\mathbf{u} : \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U\}. \quad (11)$$

The two assumptions necessary to prove the guarantee of feasible-side global convergence are that:

A1: The functions ϕ_p , g_p , and g are twice continuously differentiable (C^2) on an open set containing \mathcal{S} .

A2: The initial experimental iterate, \mathbf{u}_0 , is strictly feasible with respect to the experimental constraints ($g_{p,j}(\mathbf{u}_0) < 0$, $\forall j = 1, \dots, n_{g_p}$), feasible with respect to the numerical constraints ($g_j(\mathbf{u}_0) \leq 0$, $\forall j = 1, \dots, n_g$), and lies in the experimental space ($\mathbf{u}_0 \in \mathcal{S}$).

The actual scheme proposed to apply the SCFO (2)-(8) is a project-and-filter approach, where one first projects some suggested solution \mathbf{u}_{k+1}^* , obtained by whatever means, onto the local descent space of the cost and ε -active constraints:

$$\begin{aligned} \bar{\mathbf{u}}_{k+1}^* := \arg \underset{\mathbf{u}}{\text{minimize}} \quad & \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ \text{subject to} \quad & \nabla g_{p,j}(\mathbf{u}_k)^T (\mathbf{u} - \mathbf{u}_k) \leq -\delta_{g_{p,j}} \\ & \forall j : g_{p,j}(\mathbf{u}_k) \geq -\varepsilon_{p,j} \\ & \nabla g_j(\mathbf{u}_k)^T (\mathbf{u} - \mathbf{u}_k) \leq -\delta_{g,j} \\ & \forall j : g_j(\mathbf{u}_k) \geq -\varepsilon_j \\ & \nabla \phi_p(\mathbf{u}_k)^T (\mathbf{u} - \mathbf{u}_k) \leq -\delta_\phi \\ & \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U, \end{aligned} \quad (12)$$

and then filters this solution by applying a numerical line search to determine the filter gain value $K_k \in [0, 1]$, with the next iterate defined as

$$\mathbf{u}_{k+1} := \mathbf{u}_k + K_k (\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_k), \quad (13)$$

where the goal is to maximize K_k while satisfying the conditions

$$g_{p,j}(\mathbf{u}_k) + K_k \sum_{i=1}^{n_u} \kappa_{p,ji} |\bar{u}_{k+1,i}^* - u_{k,i}| \leq 0, \quad \forall j = 1, \dots, n_{g_p}, \quad (14)$$

$$g_j(\mathbf{u}_k + K_k (\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_k)) \leq 0, \quad \forall j = 1, \dots, n_g, \quad (15)$$

$$\begin{aligned} & \nabla \phi_p(\mathbf{u}_k)^T (\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_k) + \\ & \frac{K_k}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} M_{\phi,i_1i_2} \left| \begin{array}{c} (\bar{u}_{k+1,i_1}^* - u_{k,i_1}) \\ (\bar{u}_{k+1,i_2}^* - u_{k,i_2}) \end{array} \right| \leq 0. \end{aligned} \quad (16)$$

In numerical implementation, one may start with an initial choice of projection parameters that correspond to the expected sizes of the different function ranges and then proceed to decrease these values once the projection (12) becomes infeasible. If this infeasibility persists for very small values of the projection parameters, then the user has a theoretical guarantee that the Fritz John error at the current experimental iterate is very small and that the iterate is therefore close to \mathbf{u}^* in the Fritz-John-error sense, believed to be sufficient for most practical applications.

While the theoretical SCFO framework is very powerful, it is, as presented in [10], only so *in theory* and as such must be modified for real applications where different implementation issues may arise and make the nominal formulation of the SCFO (2)-(8) either difficult or impossible to satisfy. Here, we list the key issues that we have encountered or expect to encounter:

- *Physical degradation:* When the experimental optimization problem involves a physical system, there is a high likelihood that the system at hand will not function identically on the day that it is first used and on some day a year or two later. This essentially means that the underlying experimental functions ϕ_p and $g_{p,j}$ will evolve and change over time, thereby changing the optimization problem.
- *Unknown Lipschitz constants:* The Lipschitz constants $\kappa_{p,ji}$ and M_{ϕ,i_1i_2} will rarely be known in practice. It follows that methods that either bypass using these constants or can reliably estimate their values are required.
- *Noise:* The function values $\phi_p(\mathbf{u}_k)$ and $g_{p,j}(\mathbf{u}_k)$ will often be corrupted by noise or estimation error in many experimental settings, thereby corrupting Condition (2), which relies on the exact value of $g_{p,j}(\mathbf{u}_k)$, and making the verification of cost decrease, $\phi_p(\mathbf{u}_{k+1}) - \phi_p(\mathbf{u}_k) < 0$, difficult.
- *Gradient estimation:* Conditions (5), (7), and (8) all rely on using gradients of experimental functions, which will be unavailable. In practice, one must work with estimates instead.
- *Sufficient excitation:* As a key requirement for estimating the gradient of an experimental function at \mathbf{u}_k is being able to perturb the decision variables \mathbf{u} locally around \mathbf{u}_k , there needs to be some guarantee of this being possible without violating the safety constraints.
- *Conservatism and slow convergence:* The SCFO are sufficient for feasible-side global convergence and, not surprisingly, conservative by necessity so as to enforce the guarantees that they bring. In some applications, it may be desired to trade these guarantees for faster convergence, which may involve relaxing certain Lipschitz constants or allowing certain experimental constraints to be violated temporarily.

The goal of this document is to address all of these issues and to show how the SCFO may be modified to accommodate each of these challenges. The remainder of this work will treat these issues one-by-one as follows:

- Section 2: Physical Degradation,
- Section 3: Estimation of Lipschitz Constants and Potential Relaxations in the Lipschitz Bounds,
- Section 4: Measurement/Estimation Noise in the Experimental Function Values,
- Section 5: Working with Bounded Gradient Estimates,
- Section 6: Preserving Feasibility While Allowing for Sufficient Excitation,
- Section 7: Incorporating Soft Constraints,
- Section 8: Incorporating a Numerical Cost Function.

Because all of these issues are expected to arise concurrently in real application, we have decided that the presentation not be modular – i.e., that each section should build on the results of its predecessor. While the individual steps to modify the SCFO are, for the most part, simple and intuitive, their final accumulation yields some rather convoluted formulations, which may appear incomprehensible unless the sections are read and understood in the order that they are presented.

The final two sections then seek to consolidate the preceding derivations by

- providing the full implementable SCFO procedure in Section 9, and
- proving the theoretical safety properties of the procedure in Section 10.

We confess that we have chosen, given the complexity of the full modified SCFO, not to derive a modified proof of feasible-side global convergence in the current version of the document and have limited the theoretical results to feasible-side iterates only. A future version will seek to prove global convergence as well – for the time being, we have been unable to find a satisfactory set of assumptions that allow us to prove this result.

2 Physical Degradation

It is almost inevitable that any experimental optimization problem involving some sort of physical system will be subject to degradation effects. Here, we will use the term “degradation” – also known in both the literature and industry as “process drift” [14, 20] – in a fairly general sense to denote changes that occur in the system over time. While such changes can often be unwanted, such as damage due to repeated loading in a structure [23] or the growth of oxide scales in a solid oxide fuel cell [4], they need not be. For example, in steady-state optimization problems where

there are both fast and slow time-scale effects but where the influence of the fast-scale effects is greater, one may successfully optimize at a higher frequency while treating the minor effects that appear gradually on the slower time scale as a sort of “degradation” [11].

To express this phenomenon in mathematical terms, we may consider the experimental functions¹ as being functions not only of the decision variables \mathbf{u} but also of the time. Denoting this time by τ , we restate the experimental optimization problem (1) as

$$\begin{aligned} & \underset{\mathbf{u}}{\text{minimize}} \phi_p(\mathbf{u}, \tau) \\ & \text{subject to } g_{p,j}(\mathbf{u}, \tau) \leq 0, \quad j = 1, \dots, n_{g_p} \\ & \quad g_j(\mathbf{u}) \leq 0, \quad j = 1, \dots, n_g \\ & \quad \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U. \end{aligned} \quad (17)$$

This is basically tantamount to assuming that there remains a deterministic functional relationship between the decision variables and the cost and constraints for the “degraded” problem at a given instant in time τ , which seems reasonable as degradation is generally a deterministic process.

Note that (17) is equivalent in form to the nominal version (1), save that the variable τ is not a decision variable but one that grows freely of any influence. We may, however, take advantage of this similarity to rederive the SCFO for (17) in a manner identical to how they were derived for (1).

2.1 Feasibility in the Presence of Degradation

Denote by τ_k the time of the experiment at \mathbf{u}_k . To guarantee feasibility at the experimental iteration $k + 1$ for a given experimental constraint $g_{p,j}$, it is necessary to guarantee

$$g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \leq 0. \quad (18)$$

As in the nominal case, this condition is intractable since the function $g_{p,j}$ is unknown. In the same manner as before [10], one may avoid this issue by exploiting the Lipschitz bound²

$$\begin{aligned} g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) & \leq g_{p,j}(\mathbf{u}_k, \tau_k) \\ & \quad + \kappa_{p,j\tau} (\tau_{k+1} - \tau_k) + \sum_{i=1}^{n_u} \kappa_{p,ji} |u_{k+1,i} - u_{k,i}|, \end{aligned} \quad (19)$$

where $\kappa_{p,j\tau}$ may be seen as the extension of the Lipschitz constant to the degradation effects over time, and is implicitly defined as

¹ While one could generalize and let the numerical constraints g_j be subject to degradation as well, in our experience these are usually simple mathematical relationships that remain the same all throughout operation.

² Throughout the document, we will call upon Lipschitz bounds several times without giving their derivations. The interested reader is referred to [8] to see how they are derived.

$$-\kappa_{p,j\tau} \leq \left. \frac{\partial g_{p,j}}{\partial \tau} \right|_{(\mathbf{u}, \tau)} \leq \kappa_{p,j\tau}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{I}_\tau, \quad (20)$$

with \mathcal{I}_τ the *temporal experimental space* defined as

$$\mathcal{I}_\tau = \mathcal{I} \times \{\tau : \tau_0 \leq \tau \leq \bar{\tau}\}, \quad (21)$$

i.e., as the Cartesian product of the standard experimental space with time. To maintain the boundedness of \mathcal{I}_τ , we will suppose an upper bound on the time considered, i.e., that $\bar{\tau} < \infty$. Naturally, the definitions of the other Lipschitz constants also need to be modified for (19) to be valid:

$$-\kappa_{p,ji} < \left. \frac{\partial g_{p,j}}{\partial u_i} \right|_{(\mathbf{u}, \tau)} < \kappa_{p,ji}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{I}_\tau. \quad (22)$$

The Lipschitz constants for the cost function derivatives will also need to be modified accordingly:

$$-M_{\phi,i_1i_2} < \left. \frac{\partial^2 \phi_p}{\partial u_{i_2} \partial u_{i_1}} \right|_{(\mathbf{u}, \tau)} < M_{\phi,i_1i_2}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{I}_\tau. \quad (23)$$

Note that we will not require the Lipschitz constants for the degradation in (20) to be strict like the others, as this will allow us to return to the no-degradation case easily by setting $\kappa_{p,j\tau} := 0$.

Prior to advancing the analysis further, let us reflect on the meaning of the Lipschitz constant defined in (20). Basically, this is a bound on the ‘‘slope’’ of the drift or degradation that can occur in the experimental function values at any time and for any choice of decision variables. This value will, of course, never be known exactly, but reasonable guesses should be possible in many of the engineering contexts where degradation would occur, especially for those applications where some sort of effort has already been put into modeling the nature of the degradation. As degradation effects also tend to be relatively slow [14], it is expected that $\kappa_{p,j\tau}$ be relatively small for most applications.

Clearly, the version of (2) that holds in the degradation case may be derived by simply forcing the upper bound of (19) to be nonpositive:

$$g_{p,j}(\mathbf{u}_k, \tau_k) + \kappa_{p,j\tau} (\tau_{k+1} - \tau_k) + \sum_{i=1}^{n_u} \kappa_{p,ji} |u_{k+1,i} - u_{k,i}| \leq 0, \quad (24)$$

as this implies (18), with strict inequality holding whenever $g_{p,j}(\mathbf{u}_k, \tau_k) < 0$.

Note, however, that while one could always preserve feasibility for the nominal case by choosing \mathbf{u}_{k+1} to be sufficiently close to \mathbf{u}_k , this is no longer the case here – (24) is clearly impossible to satisfy for any choice of \mathbf{u}_{k+1} if $-g_{p,j}(\mathbf{u}_k, \tau_k) < \kappa_{p,j\tau} (\tau_{k+1} - \tau_k)$. This motivates the need for a sort of backtracking, or choosing a reference point other than \mathbf{u}_k to use in the SCFO.

2.2 Choice of Reference Point

Consider modifying (24) by using the reference iteration k^* instead of k :

$$g_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \kappa_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) + \sum_{i=1}^{n_u} \kappa_{p,ji} |u_{k+1,i} - u_{k^*,i}| \leq 0, \quad (25)$$

with the experimental iteration $k^* \in [0, k]$ chosen as

$$k^* := \arg \underset{\bar{k} \in [0, k]}{\text{maximize}} \bar{k} \quad \text{subject to} \quad g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \kappa_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0 \quad \forall j = 1, \dots, n_{g_p}, \quad (26)$$

i.e., as the most recent experimental iteration for which feasibility for the experiment at $k+1$ may still be guaranteed.

The motivation for choosing the most recent iteration comes from the innate properties of the SCFO – since the cost is lowered monotonically from one experiment to the next, the most recent experiment should, in the nominal case, be the most optimal one and thus the best to use as a reference for further improvement. The motivation for wanting to satisfy the constraint $g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \kappa_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0$ should be clear – it is impossible to guarantee that $g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \leq 0$ if this is not the case.

However, there is still no guarantee that this constraint will be satisfied for any choice of k^* , and for problems where degradation is particularly fast it may occur that one has no choice but to go ahead with the next experiment without a guarantee that the experimental iterate will be feasible. If this is preferable to shutting down the physical system out of safety concerns, then one may attempt to choose k^* as

$$k^* := \arg \underset{\bar{k} \in [0, k]}{\text{minimize}} \max_{j=1, \dots, n_{g_p}} \left[g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \kappa_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \right], \quad (27)$$

by essentially choosing the point that has the lowest worst-case constraint value. Because (27) is sensitive to scaling, it is recommended that the different constraint functions be scaled beforehand. Following this step, the most sensible action, from the point of view of safety, would be to set $\mathbf{u}_{k+1} := \mathbf{u}_{k^*}$, as one cannot guarantee feasibility at $(\mathbf{u}_{k+1}, \tau_{k+1})$ and changing the decision variables could potentially make things even worse. If, then, the observed value $-g_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1})$ is greater than $\kappa_{p,j\tau} (\tau_{k+2} - \tau_{k+1})$, it will be possible for the SCFO to once more guarantee feasibility at $k+2$.

Yet another alternative for choosing k^* is to define, based on *a priori* knowledge, some ‘‘safe point’’ and use this point as \mathbf{u}_{k^*} when (26) fails and when (27) is deemed too risky.

Not surprisingly, some additional assumptions on the degradation, or on the existence of a safe point, will be necessary to make any rigorous theoretical claims regarding feasibility. These will be addressed in Section 10. As a final remark regarding the use of the reference \mathbf{u}_k^* , note that Conditions (5)-(8) must be modified accordingly:

$$\begin{aligned} \nabla g_{p,j}(\mathbf{u}_k^*)^T (\mathbf{u}_{k+1} - \mathbf{u}_k^*) &\leq -\delta_{g_{p,j}} \\ \forall j : g_{p,j}(\mathbf{u}_k^*) &\geq -\varepsilon_{p,j}, \end{aligned} \quad (28)$$

$$\nabla g_j(\mathbf{u}_k^*)^T (\mathbf{u}_{k+1} - \mathbf{u}_k^*) \leq -\delta_{g,j}, \quad \forall j : g_j(\mathbf{u}_k^*) \geq -\varepsilon_j, \quad (29)$$

$$\nabla \phi_p(\mathbf{u}_k^*)^T (\mathbf{u}_{k+1} - \mathbf{u}_k^*) \leq -\delta_\phi, \quad (30)$$

$$\begin{aligned} \nabla \phi_p(\mathbf{u}_k^*)^T (\mathbf{u}_{k+1} - \mathbf{u}_k^*) + \\ \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} M_{\phi,i_1 i_2} \begin{vmatrix} (u_{k+1,i_1} - u_{k^*,i_1}) \\ (u_{k+1,i_2} - u_{k^*,i_2}) \end{vmatrix} &\leq 0. \end{aligned} \quad (31)$$

2.3 Projecting in the Presence of Degradation

We now turn to modifying (28) and (30) to account for degradation. Following the same philosophy as before – i.e., treating τ as an additional, albeit not controlled, variable – we propose:

$$\begin{aligned} \nabla g_{p,j}(\mathbf{u}_k^*, \tau_{k+1})^T \begin{bmatrix} \mathbf{u}_{k+1} - \mathbf{u}_k^* \\ 0 \end{bmatrix} &\leq -\delta_{g_{p,j}}, \\ \forall j : g_{p,j}(\mathbf{u}_k^*, \tau_{k+1}) + \kappa_{p,j\tau}(\tau_{k+1} - \tau_k^*) &\geq -\varepsilon_{p,j}, \end{aligned} \quad (32)$$

$$\nabla \phi_p(\mathbf{u}_k^*, \tau_{k+1})^T \begin{bmatrix} \mathbf{u}_{k+1} - \mathbf{u}_k^* \\ 0 \end{bmatrix} \leq -\delta_\phi, \quad (33)$$

where the degrees of freedom, \mathbf{u}_{k+1} , are used to both decrease the cost and stay away from the ε -active constraints locally for the functions in their one-step ahead degraded state.

A couple of remarks are necessary:

- The gradients $\nabla g_{p,j}(\mathbf{u}_k^*, \tau_{k+1})$ and $\nabla \phi_p(\mathbf{u}_k^*, \tau_{k+1})$ are unknown and must be estimated based on the available data. Our current algorithm of choice for carrying out this task is that of [5]. The incorporation of gradient estimates into the SCFO is discussed fully in Section 5.
- We choose to use τ_{k+1} instead of τ_k^* in the projection so as to enforce local descent for the functions as they are expected to be at the future iteration $k+1$. For the same reasons, we would prefer to use the most up-to-date ε -activity condition $g_{p,j}(\mathbf{u}_k^*, \tau_{k+1}) \geq -\varepsilon_{p,j}$. However, since $g_{p,j}(\mathbf{u}_k^*, \tau_{k+1})$ cannot be measured, we work instead with its upper Lipschitz bound $g_{p,j}(\mathbf{u}_k^*, \tau_k^*) + \kappa_{p,j\tau}(\tau_{k+1} - \tau_k^*)$.

2.4 Monotonic Cost Improvement with Degradation

Let us modify Condition (31), recalling first the Lipschitz bound

$$\begin{aligned} \phi_p(\mathbf{u}_{k+1}, \tau_{k+1}) - \phi_p(\mathbf{u}_k^*, \tau_{k+1}) &\leq \\ \nabla \phi_p(\mathbf{u}_k^*, \tau_{k+1})^T \begin{bmatrix} \mathbf{u}_{k+1} - \mathbf{u}_k^* \\ 0 \end{bmatrix} & \\ + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} M_{\phi,i_1 i_2} \begin{vmatrix} (u_{k+1,i_1} - u_{k^*,i_1}) \\ (u_{k+1,i_2} - u_{k^*,i_2}) \end{vmatrix}, & \end{aligned} \quad (34)$$

where we ultimately aim to achieve a decrease between $\phi_p(\mathbf{u}_k^*, \tau_{k+1})$ and $\phi_p(\mathbf{u}_{k+1}, \tau_{k+1})$, i.e., to change \mathbf{u} in a manner that leads to an improvement in the cost value at the future instance τ_{k+1} .

To force this decrease, it is then sufficient to enforce

$$\begin{aligned} \nabla \phi_p(\mathbf{u}_k^*, \tau_{k+1})^T \begin{bmatrix} \mathbf{u}_{k+1} - \mathbf{u}_k^* \\ 0 \end{bmatrix} & \\ + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} M_{\phi,i_1 i_2} \begin{vmatrix} (u_{k+1,i_1} - u_{k^*,i_1}) \\ (u_{k+1,i_2} - u_{k^*,i_2}) \end{vmatrix} &\leq 0, \end{aligned} \quad (35)$$

as this would imply $\phi_p(\mathbf{u}_{k+1}, \tau_{k+1}) - \phi_p(\mathbf{u}_k^*, \tau_{k+1}) < 0$ when $\mathbf{u}_{k+1} \neq \mathbf{u}_k^*$. Note that this is nearly identical to the degradation-free case, with the only difference being in the gradient.

2.5 The Project-and-Filter Approach in the Presence of Degradation

The project-and-filter approach may now be modified accordingly. First, the conditions (32) and (33) are used in place of (5) and (7) in the projection, with (29) also replacing (6) to account for the choice of reference:

$$\begin{aligned} \bar{\mathbf{u}}_{k+1}^* &:= \\ \arg \min_{\mathbf{u}} &\|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ \text{subject to } &\nabla g_{p,j}(\mathbf{u}_k^*, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_k^* \\ 0 \end{bmatrix} \leq -\delta_{g_{p,j}} \\ &\forall j : g_{p,j}(\mathbf{u}_k^*, \tau_{k+1}) + \kappa_{p,j\tau}(\tau_{k+1} - \tau_k^*) \geq -\varepsilon_{p,j} \\ &\nabla g_j(\mathbf{u}_k^*)^T (\mathbf{u} - \mathbf{u}_k^*) \leq -\delta_{g,j} \\ &\forall j : g_j(\mathbf{u}_k^*) \geq -\varepsilon_j \\ &\nabla \phi_p(\mathbf{u}_k^*, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_k^* \\ 0 \end{bmatrix} \leq -\delta_\phi \\ &\mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U. \end{aligned} \quad (36)$$

Noting that the future iterate is now defined as

$$\mathbf{u}_{k+1} := \mathbf{u}_k^* + K_k (\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_k^*), \quad (37)$$

substitution of this law into (25), (3), and (35) then yields the following constraints on $K_k \in [0, 1]$:

$$g_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \kappa_{p,j}\tau(\tau_{k+1} - \tau_{k^*}) + K_k \sum_{i=1}^{n_u} \kappa_{p,ji} |\bar{u}_{k+1,i}^* - u_{k^*,i}| \leq 0, \quad \forall j = 1, \dots, n_{g_p}, \quad (38)$$

$$g_j(\mathbf{u}_{k^*} + K_k(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*})) \leq 0, \quad \forall j = 1, \dots, n_g, \quad (39)$$

$$\nabla \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} + \frac{K_k}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} M_{\phi,i_1 i_2} \begin{vmatrix} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}) \end{vmatrix} \leq 0. \quad (40)$$

2.6 Example

Consider the constructed example used in [10]:

$$\begin{aligned} & \text{minimize}_{u_1, u_2} \phi_p(\mathbf{u}) := (u_1 - 0.5)^2 + (u_2 - 0.4)^2 \\ & \text{subject to } g_{p,1}(\mathbf{u}) := -6u_1^2 - 3.5u_1 + u_2 - 0.6 \leq 0 \\ & \quad g_{p,2}(\mathbf{u}) := 2u_1^2 + 0.5u_1 + u_2 - 0.75 \leq 0 \\ & \quad g_1(\mathbf{u}) := -u_1^2 - (u_2 - 0.15)^2 + 0.01 \leq 0 \\ & \quad -0.5 \leq u_1 \leq 0.5 \\ & \quad 0 \leq u_2 \leq 0.8, \end{aligned} \quad (41)$$

for which we reproduce the solution obtained by the nominal SCFO methodology in Fig. 1, referring the reader to [10] for all details regarding choice of initial experiment, Lipschitz constants, etc.

To study the effects of degradation, let us suppose, for simplicity, that $\tau = k$ and that the experiments are carried out at a constant frequency. A version of (41) that degrades with time is constructed:

$$\begin{aligned} & \text{minimize}_{u_1, u_2} (u_1 - 0.5)^2 + (u_2 - 0.4 - \frac{\tau}{500})^2 \\ & \text{subject to } -6u_1^2 - (3.5 + \frac{\tau}{500})u_1 + u_2 - 0.6 \leq 0 \\ & \quad 2u_1^2 + 0.5u_1 + u_2 - 0.75 \pm \frac{\tau}{500} \leq 0 \\ & \quad -u_1^2 - (u_2 - 0.15)^2 + 0.01 \leq 0 \\ & \quad -0.5 \leq u_1 \leq 0.5 \\ & \quad 0 \leq u_2 \leq 0.8, \end{aligned} \quad (42)$$

with \pm denoting two different scenarios – one where the feasible space with respect to $g_{p,2}$ shrinks due to degradation (+), and one where it expands (–). With the time considered limited to $\bar{\tau} = 200$ (i.e., to 200 experimental iterations), the Lipschitz constants over the resulting \mathcal{S}_τ are appropriately chosen as

$$\begin{aligned} \kappa_{p,11} &= 10, \quad \kappa_{p,12} = 2, \quad \kappa_{p,1\tau} = \frac{1}{1000}, \\ \kappa_{p,21} &= 3, \quad \kappa_{p,22} = 2, \quad \kappa_{p,2\tau} = \frac{1}{500}, \\ M_{\phi,11} &= 3, \quad M_{\phi,12} = 1, \\ M_{\phi,21} &= 1, \quad M_{\phi,22} = 3. \end{aligned} \quad (43)$$

Using these constants and the modifications described in (36)-(40), with k^* chosen via (26) and, if this is not possible, via (27), the results obtained for the (+) and (–) scenarios are illustrated geometrically in Figs. 2 and 3, with the corresponding cost function values given in Fig. 4. Though it may not be obvious from the figures, we note that the scheme is able to satisfy the constraints at all iterations for this problem, sometimes doing so at the cost of optimality. In the (+) case, one sees that the scheme often does not have enough time to track the changing optimum due to the continuing need to return to the shrinking feasible region – this becomes particularly pertinent after Experiment 150.

While this approximate satisfaction of the SCFO is largely satisfactory – certainly for the (–) case – it should again be emphasized that *no guarantees are possible in the general degradation case without additional assumptions on the nature of the degradation*. That we cannot track the changing optimum perfectly should be evident, as being able to do so would be tantamount to solving Problem (42), which changes at every experiment, in only a single experimental iteration. That we cannot guarantee feasibility throughout should be evident as well – if we ran the optimization for longer (to, say, $k = 1000$), then the feasible space of the problem would simply shrink and vanish.

3 Estimating and Relaxing the Lipschitz Bounds

The Lipschitz bounds and the constants that define them play a fundamental role in the SCFO, and as such are a double-edged sword. While it is precisely because of these bounds that it is possible for the SCFO to ensure feasible-side convergence and safety throughout the experimental optimization, one must nevertheless do some work to obtain proper estimates of the Lipschitz constants, which cannot be known with certainty in many practical applications. Traditionally, the problem of Lipschitz constant estimation has been largely restricted to the domain of global optimization [32, 16, 35], where these constants are typically used to build lower bounds in branch-and-bound algorithms [17]. Some of the philosophy in these methods will be employed here – notably via the Lipschitz consistency check of Section 3.5.3. However, we will largely tackle the estimation problem from scratch in a more engineering-based, holistic manner – combining the physical meaning of the Lipschitz constant that is largely nonexistent in the mathematical context with additional

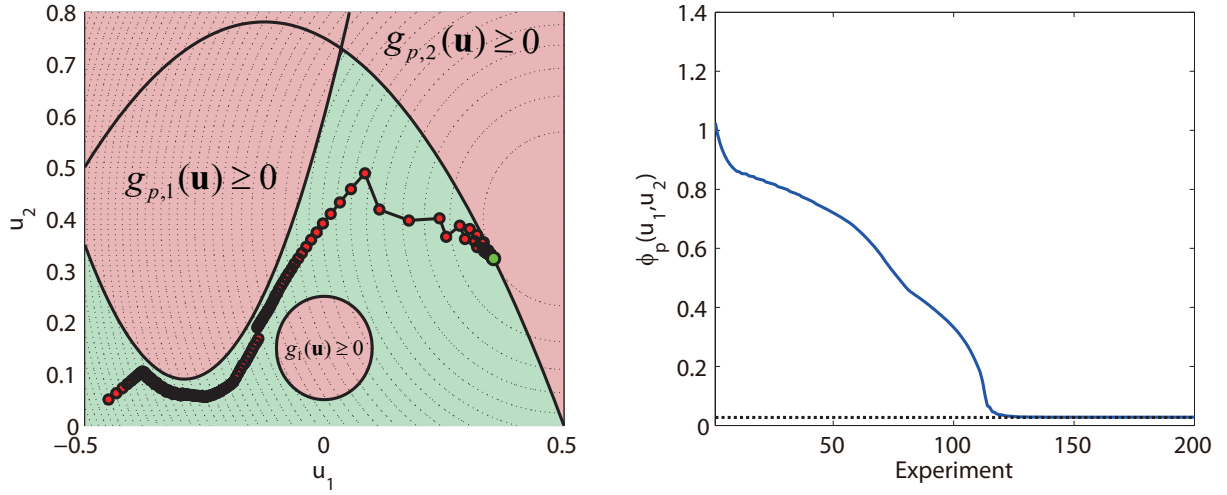


Fig. 1 Chain of experiments (red points) generated by applying the nominal SCFO methodology to Problem (41). The green point denotes the only local minimum. The dotted lines on the left plot denote the contours of the cost function, while the constant dotted line on the right denotes the cost value at the minimum.

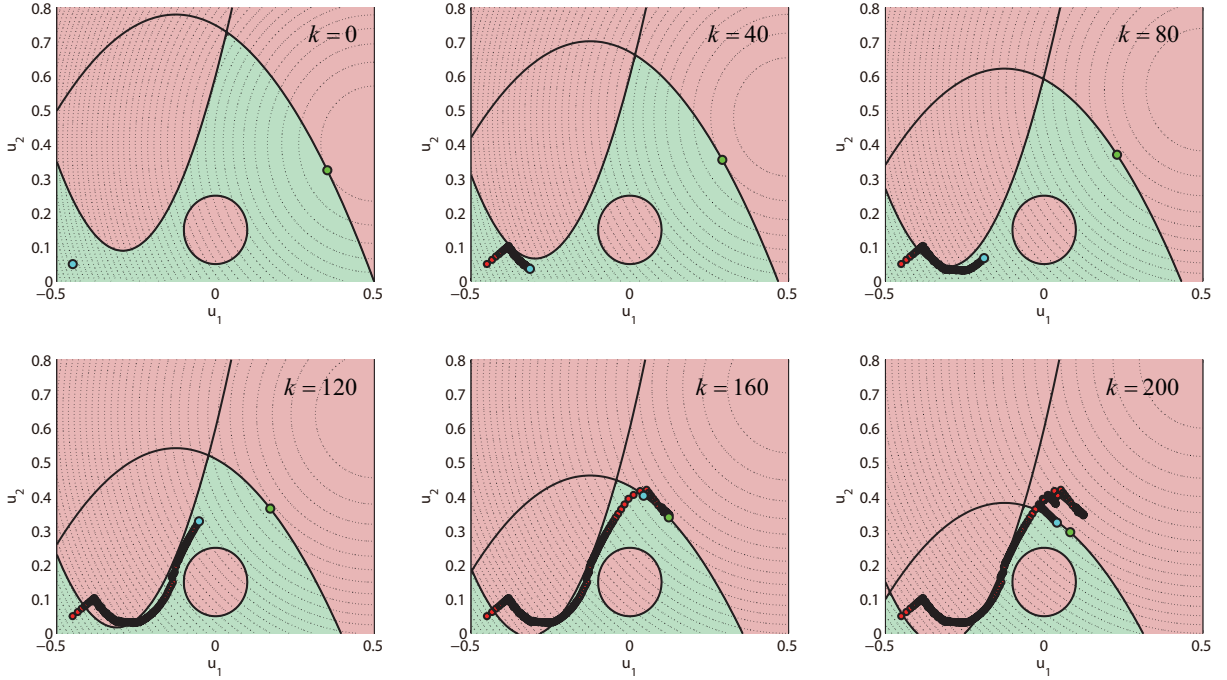


Fig. 2 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (+) scenario. The blue point is used to highlight the current experimental iterate.

mathematical “tricks” wherever possible so as to shed some light on the possible ways to provide the SCFO the estimates it needs. In addition to looking for ways to provide valid estimates, much of our discussion will focus on how to make these valid estimates as non-conservative as possible – clearly, one could almost always guarantee validity by setting the constants to some enormous number, but this would lead to very slow convergence.

3.1 Lower and Upper Lipschitz Constants

Let us start by seemingly doubling the complexity of the estimation task and distinguishing between lower and upper Lipschitz constants as follows:

$$\underline{\kappa}_{p,ji} < \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}, \tau)} < \bar{\kappa}_{p,ji}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{I}_\tau, \quad (44)$$

$$\underline{M}_{\phi, i_1 i_2} < \frac{\partial^2 \phi_p}{\partial u_{i_2} \partial u_{i_1}} \Big|_{(\mathbf{u}, \tau)} < \bar{M}_{\phi, i_1 i_2}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{I}_\tau, \quad (45)$$

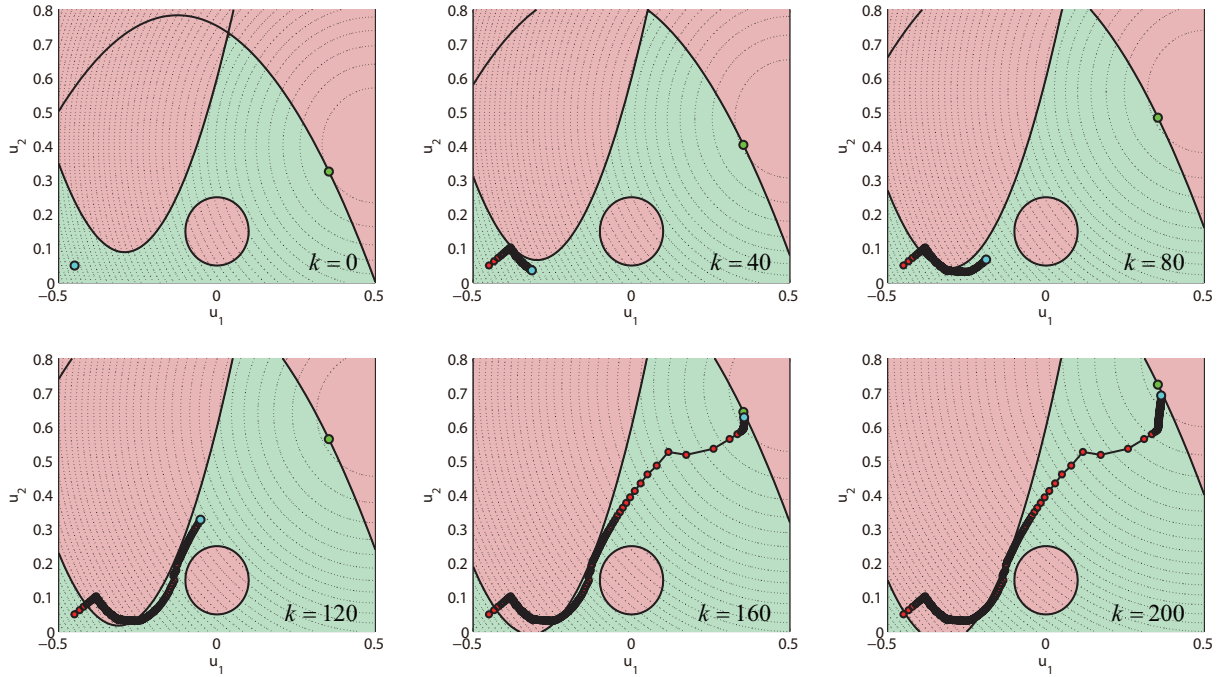


Fig. 3 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (–) scenario.

$$\underline{\kappa}_{p,j\tau} \leq \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u},\tau)} \leq \overline{\kappa}_{p,j\tau}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{I}_\tau, \quad (46)$$

which, despite doubling the number of constants that require estimates, nevertheless both allows more flexibility and relaxes the original bounds – note that one could always remove this flexibility by setting, e.g., $-\underline{\kappa}_{p,ji} := \overline{\kappa}_{p,ji} := \kappa_{p,ji}$, as had been done previously.

An additional motivation for working with both lower and upper constants is that it often allows us to exploit engineering/physical knowledge with regard to the sign of the derivative. As should already be apparent, the physical world is full of well-documented proportional or inversely proportional relationships for which the derivatives must have positive and negative signs, respectively. Consider, as some simple examples, the following:

- An athlete interested in controlling their weight knows that they cannot start eating more of a certain food and lose weight by doing so. The derivative between the quantity of food consumed and the athlete’s weight is thus usually positive. On the contrary, it is very unlikely that the same athlete will gain weight by increasing their daily running time. Here, the derivative is negative.
- In most power systems, increasing the amount of fuel fed while keeping everything else constant should increase the power produced. The derivative between power and fuel is thus positive.

- In a jacketed reactor, it will often be the case – though not always, depending on reaction – that increasing (resp., decreasing) the temperature of the jacket will increase (resp., decrease) the temperature of the reactor. The derivative between the two temperatures is usually positive.

When such relationships are present, one may translate them into appropriate Lipschitz constants by setting either the lower or upper constant equal to 0, or to a value arbitrarily close to 0 if strict inequality is needed. For the Lipschitz constants corresponding to the derivatives, M , such physical links are harder to make, as they relate to the second derivatives and to the curvature of the function in question. However, even here we have encountered examples where *a priori* knowledge regarding the sign of M may be used – for a simple solid oxide fuel cell stack [25, 11], the relationship between the current and the power is almost always concave, which implies negative curvature and a negative M ($\overline{M} \approx 0$). Outside of engineering applications, convex and concave relationships have been noted in fields like R&D management [19, 21] and investment [33], among others.

The algorithmic usefulness of differentiating between lower and upper constants manifests itself by allowing for the SCFO framework to identify directions in which certain functions simply cannot increase, which cannot be done with the bounds employed so far in (19) and (34). Following the substitution $k \rightarrow k^*$ in (19), we use the results of [8] to formulate the relaxed bounds

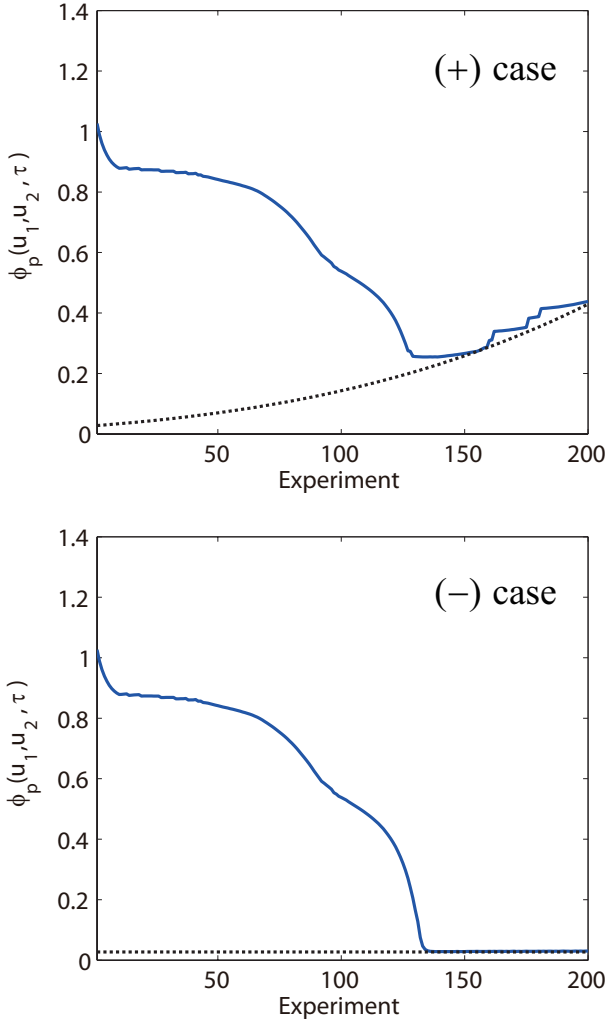


Fig. 4 Cost function values obtained by the modified SCFO methodology for Problem (42).

$$g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \leq g_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) + \sum_{i=1}^{n_u} \max \left[\frac{\underline{\kappa}_{p,ji}(u_{k+1,i} - u_{k^*,i})}{\bar{\kappa}_{p,ji}(u_{k+1,i} - u_{k^*,i})} \right], \quad (47)$$

$$\begin{aligned} & \phi_p(\mathbf{u}_{k+1}, \tau_{k+1}) - \phi_p(\mathbf{u}_{k^*}, \tau_{k+1}) \leq \\ & \nabla \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u}_{k+1} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \\ & + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \begin{bmatrix} \underline{M}_{\phi,i_1 i_2}(u_{k+1,i_1} - u_{k^*,i_1}) \\ (u_{k+1,i_2} - u_{k^*,i_2}), \\ \bar{M}_{\phi,i_1 i_2}(u_{k+1,i_1} - u_{k^*,i_1}) \\ (u_{k+1,i_2} - u_{k^*,i_2}) \end{bmatrix}. \end{aligned} \quad (48)$$

Clearly, we have opened the door for the possibility of the worst-case evolutions being *negative*, as this is allowed by the maximum operator when the lower and upper constants do not have opposite signs.

Forcing the right-hand sides of (47) and (48) to be negative then leads to the modified SCFO:

$$g_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) + \sum_{i=1}^{n_u} \max \left[\frac{\underline{\kappa}_{p,ji}(u_{k+1,i} - u_{k^*,i})}{\bar{\kappa}_{p,ji}(u_{k+1,i} - u_{k^*,i})} \right] \leq 0, \quad (49)$$

$$\begin{aligned} & \nabla \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u}_{k+1} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \\ & + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \begin{bmatrix} \underline{M}_{\phi,i_1 i_2}(u_{k+1,i_1} - u_{k^*,i_1}) \\ (u_{k+1,i_2} - u_{k^*,i_2}), \\ \bar{M}_{\phi,i_1 i_2}(u_{k+1,i_1} - u_{k^*,i_1}) \\ (u_{k+1,i_2} - u_{k^*,i_2}) \end{bmatrix} \leq 0. \end{aligned} \quad (50)$$

Slight modifications to (26) and (27) are also in order, and simply consist in the substitution $\kappa_{p,j\tau} \rightarrow \bar{\kappa}_{p,j\tau}$:

$$\begin{aligned} k^* & := \arg \underset{\bar{k} \in [0, k]}{\text{maximize}} \bar{k} \\ & \text{subject to } g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & \quad + \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0 \\ & \quad \forall j = 1, \dots, n_{g_p}, \end{aligned} \quad (51)$$

$$k^* := \arg \underset{\bar{k} \in [0, k]}{\text{minimize}} \max_{j=1, \dots, n_{g_p}} \left[g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \right]. \quad (52)$$

The project-and-filter approach may benefit from the following (small) modification in the Boolean condition on $g_{p,j}$:

$$\begin{aligned} & \bar{\mathbf{u}}_{k+1}^* := \\ & \arg \underset{\mathbf{u}}{\text{minimize}} \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ & \text{subject to } \nabla g_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_{g_{p,j}} \\ & \quad \forall j: g_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ & \quad \quad + \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) \geq -\varepsilon_{p,j} \\ & \quad \nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j} \\ & \quad \forall j: g_j(\mathbf{u}_{k^*}) \geq -\varepsilon_j \\ & \quad \nabla \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_\phi \\ & \quad \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U, \end{aligned} \quad (53)$$

with the constraints (38) and (40) in the line search modified as

$$\begin{aligned} & g_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) \\ & + K_k \sum_{i=1}^{n_u} \max \left[\frac{\underline{\kappa}_{p,ji}(\bar{u}_{k+1,i}^* - u_{k^*,i})}{\bar{\kappa}_{p,ji}(\bar{u}_{k+1,i}^* - u_{k^*,i})} \right] \leq 0, \\ & \quad \forall j = 1, \dots, n_{g_p}, \end{aligned} \quad (54)$$

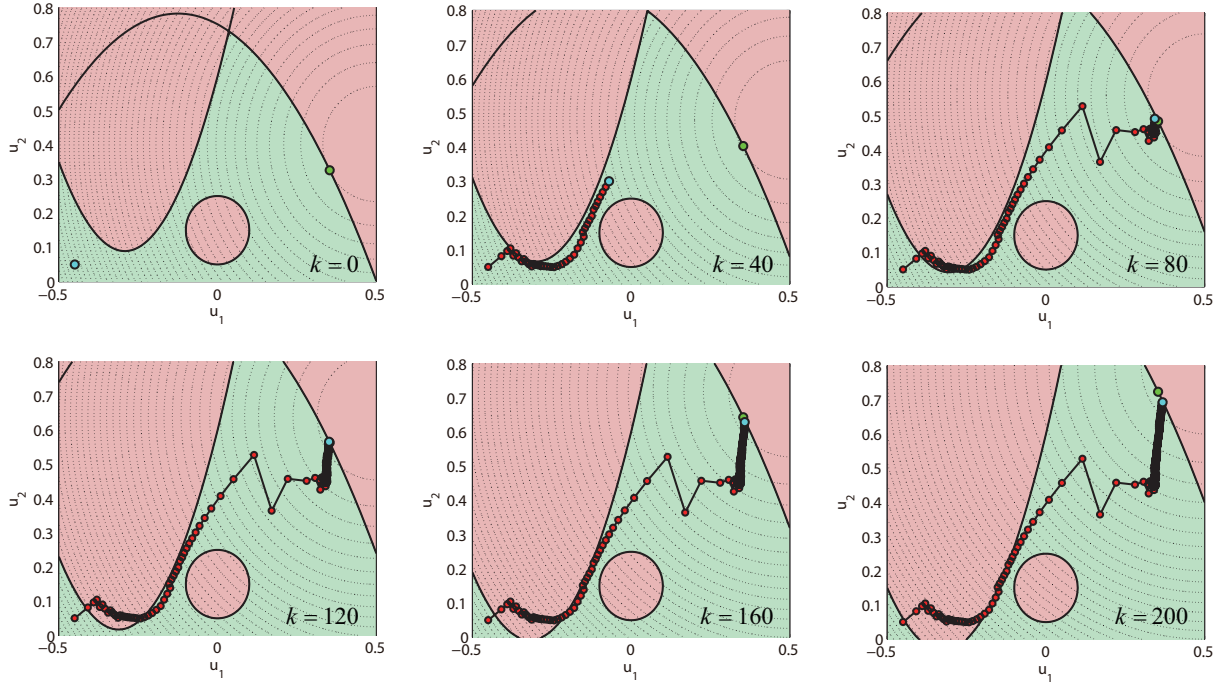


Fig. 5 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (–) scenario with differentiation between lower and upper Lipschitz constants as given in (56).

$$\begin{aligned} & \nabla \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \\ & + \frac{K_k}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \begin{bmatrix} \underline{M}_{\phi, i_1 i_2} (\bar{u}_{k+1, i_1}^* - u_{k^*, i_1}) \\ (\bar{u}_{k+1, i_2}^* - u_{k^*, i_2}), \\ \bar{M}_{\phi, i_1 i_2} (\bar{u}_{k+1, i_1}^* - u_{k^*, i_1}) \\ (\bar{u}_{k+1, i_2}^* - u_{k^*, i_2}) \end{bmatrix} \leq 0. \end{aligned} \quad (55)$$

To illustrate the effect of using both lower and upper Lipschitz constants, let us consider the (–) case of Problem (42) and relax the choice of Lipschitz constants for this problem to

$$\begin{aligned} \underline{\kappa}_{p,11} &= -10, \quad \underline{\kappa}_{p,12} = 0, \quad \underline{\kappa}_{p,1\tau} = -\frac{1}{1000}, \\ \bar{\kappa}_{p,11} &= 3, \quad \bar{\kappa}_{p,12} = 2, \quad \bar{\kappa}_{p,1\tau} = \frac{1}{1000}, \\ \underline{\kappa}_{p,21} &= -2, \quad \underline{\kappa}_{p,22} = 0, \quad \underline{\kappa}_{p,2\tau} = -\frac{1}{500}, \\ \bar{\kappa}_{p,21} &= 3, \quad \bar{\kappa}_{p,22} = 2, \quad \bar{\kappa}_{p,2\tau} = -\frac{1}{500}, \\ \underline{M}_{\phi,11} &= 1, \quad \underline{M}_{\phi,12} = -1, \\ \bar{M}_{\phi,11} &= 3, \quad \bar{M}_{\phi,12} = 1, \\ \underline{M}_{\phi,21} &= -1, \quad \underline{M}_{\phi,22} = 1, \\ \bar{M}_{\phi,21} &= 1, \quad \bar{M}_{\phi,22} = 3. \end{aligned} \quad (56)$$

Applying the modified project-and-filter approach, we obtain the results given in Figs. 5 and 6. Comparing these to those of Figs. 3 and 4, we note that there is a significant reduction in the number of experiments needed to reach the neighborhood of the optimum, which this is mainly due to

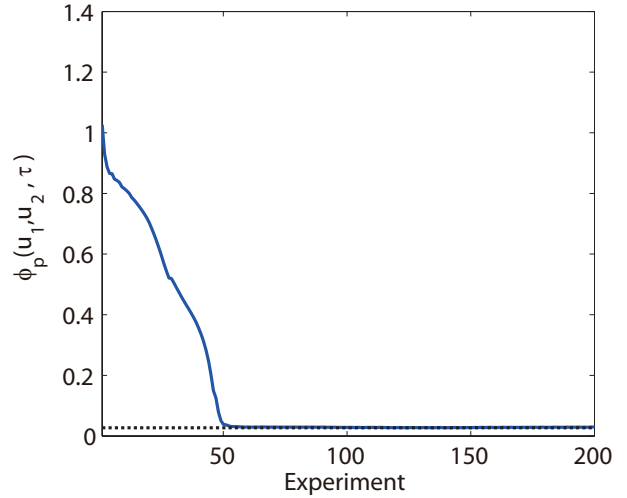


Fig. 6 Cost function values obtained by the modified SCFO methodology for Problem (42) for the (–) scenario with differentiation between lower and upper Lipschitz constants as given in (56).

relaxing the Lipschitz constant $\kappa_{p,11} = 10$ to the lower and upper constants $\underline{\kappa}_{p,11} = -10$ and $\bar{\kappa}_{p,11} = 3$, as the latter is what appears in the maximum term as the algorithm increases u_1 . Indeed, we note that this ~ 3 -fold reduction leads to a similar reduction in the number of experiments.

3.2 Concavity Relaxations of the SCFO

As mentioned previously, there may be applications where the relationships between certain decision variables and experimental functions are *concave* in nature. When this is the case, the Lipschitz bounds on the experimental constraint functions may be made tighter, which in turn will make the SCFO less conservative. So as to maintain a fair degree of generality, we choose to work with *partial concavity*, where the function in question only needs to be concave in certain variables [26].

Definition 1 (Partial concavity of constraint function)

Let the decision variables \mathbf{u} be partitioned into subvectors \mathbf{v} and \mathbf{z} , so that $g_{p,j}(\mathbf{u}, \tau) = g_{p,j}(\mathbf{v}, \mathbf{z}, \tau)$. The function $g_{p,j}$ will be said to be *globally partially concave* in \mathbf{z} if it is concave everywhere on \mathcal{S}_τ for \mathbf{v} and τ fixed. Likewise, $g_{p,j}$ will be said to be globally partially concave in both \mathbf{z} and τ if it is concave everywhere on \mathcal{S}_τ for \mathbf{v} fixed.

Lemma 1 (Concave tightening of upper Lipschitz bound)

Let $I_{c,j}$ denote the set of indices defining the variables \mathbf{z} , with respect to which the experimental constraint function $g_{p,j}$ is globally partially concave. Furthermore, let $\eta_{c,j}$ denote a Boolean indicator that is equal to 1 if $g_{p,j}$ is globally partially concave in both \mathbf{z} and τ , and is equal to 0 otherwise. It follows that the following tightening of (47) is valid:

$$\begin{aligned} g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) &\leq g_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ &+ \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ &+ (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) \\ &+ \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_{k+1,i} - u_{k^*,i}) \\ &+ \sum_{i \notin I_{c,j}} \max \left[\frac{\underline{\kappa}_{p,ji} (u_{k+1,i} - u_{k^*,i})}{\bar{\kappa}_{p,ji} (u_{k+1,i} - u_{k^*,i})} \right], \end{aligned} \quad (57)$$

where $i \notin I_{c,j}$ implicitly denotes $\{1, \dots, n_u\} \setminus I_{c,j}$.

Proof The constraint function value at the next experimental iterate may be written as

$$\begin{aligned} g_{p,j}(\mathbf{v}_{k+1}, \mathbf{z}_{k+1}, \tau_{k+1}) &= g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*}) \\ &+ g_{p,j}(\mathbf{v}_{k+1}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*}) \\ &+ g_{p,j}(\mathbf{v}_{k+1}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1}). \end{aligned} \quad (58)$$

We will proceed to obtain the desired result by bounding the two addends, $g_{p,j}(\mathbf{v}_{k+1}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*})$ and $g_{p,j}(\mathbf{v}_{k+1}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1})$, separately.

First, suppose that $\eta_{c,j} = 1$. If this is so, then we have, by the first-order condition of concavity:

$$\begin{aligned} &g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*}) \\ &\leq \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ &+ \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_{k+1,i} - u_{k^*,i}). \end{aligned} \quad (59)$$

In the case that $\eta_{c,j} = 0$, consider the decomposition

$$\begin{aligned} &g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*}) = \\ &g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k^*}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*}) \\ &+ g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k^*}), \end{aligned} \quad (60)$$

where the first addend may be bounded, again, by exploiting concavity:

$$\begin{aligned} &g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k^*}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*}) \\ &\leq \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_{k+1,i} - u_{k^*,i}), \end{aligned} \quad (61)$$

while the second may be bounded by using the Lipschitz bound

$$\begin{aligned} &g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k^*}) \\ &\leq \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}). \end{aligned} \quad (62)$$

Adding (61) and (62) thus yields

$$\begin{aligned} &g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*}) \\ &\leq \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) \\ &+ \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_{k+1,i} - u_{k^*,i}) \end{aligned} \quad (63)$$

for the case when $\eta_{c,j} = 0$.

To account for both $\eta_{c,j} = 1$ and $\eta_{c,j} = 0$ with a single bound, it is clear that we can simply combine (59) and (63) as

$$\begin{aligned} &g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k^*}, \tau_{k^*}) \\ &\leq \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ &+ (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) \\ &+ \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_{k+1,i} - u_{k^*,i}). \end{aligned} \quad (64)$$

To bound the second addend in (58), we revert to the Lipschitz bound

$$\begin{aligned} &g_{p,j}(\mathbf{v}_{k+1}, \mathbf{z}_{k+1}, \tau_{k+1}) - g_{p,j}(\mathbf{v}_{k^*}, \mathbf{z}_{k+1}, \tau_{k+1}) \\ &\leq \sum_{i \notin I_{c,j}} \max \left[\frac{\underline{\kappa}_{p,ji} (u_{k+1,i} - u_{k^*,i})}{\bar{\kappa}_{p,ji} (u_{k+1,i} - u_{k^*,i})} \right]. \end{aligned} \quad (65)$$

Applying (64) and (65) to (58), or simply adding (64) to (65) and rearranging, then yields (57). \square

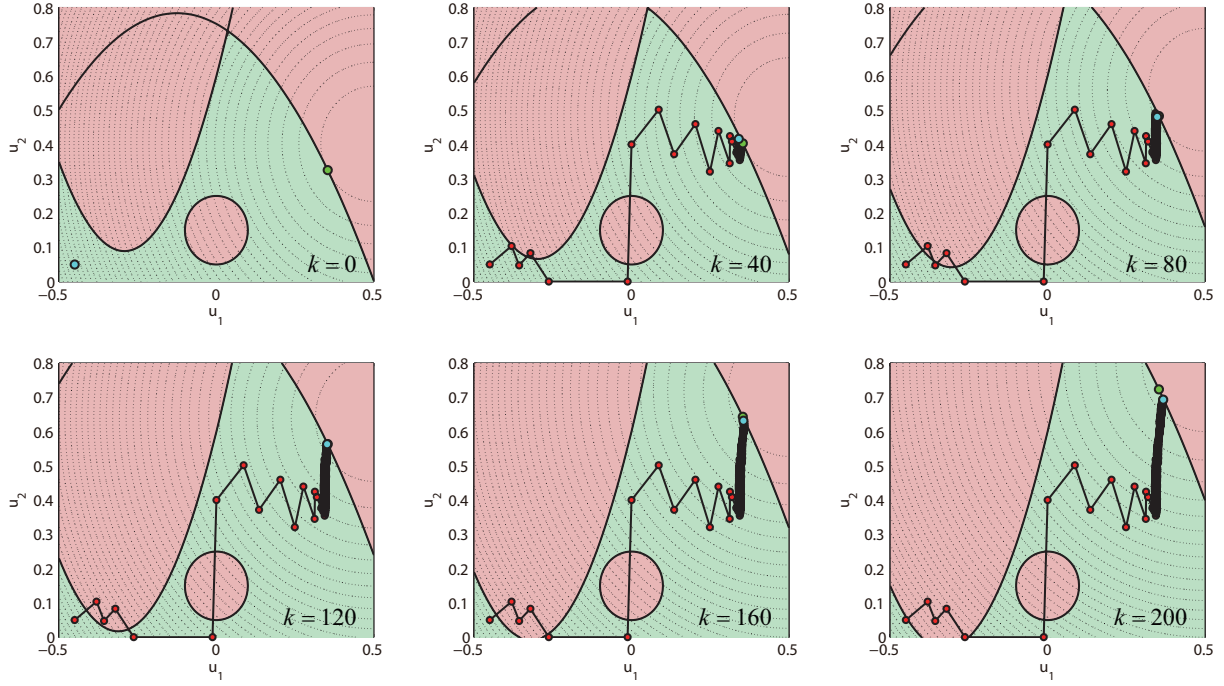


Fig. 7 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (-) scenario with concavity relaxations included.

Because

$$\begin{aligned} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_k^*, \tau_k^*)} (u_{k+1,i} - u_{k^*,i}) \\ \leq \max \left[\frac{\underline{\kappa}_{p,ji}(u_{k+1,i} - u_{k^*,i})}{\bar{\kappa}_{p,ji}(u_{k+1,i} - u_{k^*,i})} \right] \end{aligned} \quad (66)$$

and

$$\frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_k^*, \tau_k^*)} (\tau_{k+1} - \tau_k^*) \leq \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_k^*), \quad (67)$$

one sees that (57) is indeed a tighter bound than (47).

The modified SCFO (49) can thus be relaxed by forcing the right-hand side of (57) to be nonpositive:

$$\begin{aligned} g_{p,j}(\mathbf{u}_k^*, \tau_k^*) + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_k^*, \tau_k^*)} (\tau_{k+1} - \tau_k^*) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_k^*) \\ + \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_k^*, \tau_k^*)} (u_{k+1,i} - u_{k^*,i}) \\ + \sum_{i \notin I_{c,j}} \max \left[\frac{\underline{\kappa}_{p,ji}(u_{k+1,i} - u_{k^*,i})}{\bar{\kappa}_{p,ji}(u_{k+1,i} - u_{k^*,i})} \right] \leq 0. \end{aligned} \quad (68)$$

In applying the project-and-filter approach, we may, using the same arguments as above, relax the Boolean condition in the projection:

$$\begin{aligned} \bar{\mathbf{u}}_{k+1}^* := \\ \arg \underset{\mathbf{u}}{\text{minimize}} \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ \text{subject to } \nabla g_{p,j}(\mathbf{u}_k^*, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_k^* \\ 0 \end{bmatrix} \leq -\delta_{g_{p,j}} \\ g_{p,j}(\mathbf{u}_k^*, \tau_k^*) \\ \forall j : +\eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_k^*, \tau_k^*)} (\tau_{k+1} - \tau_k^*) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_k^*) \geq -\varepsilon_{p,j} \\ \nabla g_j(\mathbf{u}_k^*)^T (\mathbf{u} - \mathbf{u}_k^*) \leq -\delta_{g_j} \\ \forall j : g_j(\mathbf{u}_k^*) \geq -\varepsilon_j \\ \nabla \phi_p(\mathbf{u}_k^*, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_k^* \\ 0 \end{bmatrix} \leq -\delta_\phi \\ \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U, \end{aligned} \quad (69)$$

The feasibility condition (54) is then be modified as

$$\begin{aligned} g_{p,j}(\mathbf{u}_k^*, \tau_k^*) + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_k^*, \tau_k^*)} (\tau_{k+1} - \tau_k^*) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_k^*) \\ + K_k \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_k^*, \tau_k^*)} (\bar{u}_{k+1,i}^* - u_{k^*,i}) \\ + K_k \sum_{i \notin I_{c,j}} \max \left[\frac{\underline{\kappa}_{p,ji}(\bar{u}_{k+1,i}^* - u_{k^*,i})}{\bar{\kappa}_{p,ji}(\bar{u}_{k+1,i}^* - u_{k^*,i})} \right] \leq 0, \\ \forall j = 1, \dots, n_{g_p}. \end{aligned} \quad (70)$$

Further modifications to (51) and (52) are also necessary to account for the possible relaxations:

$$\begin{aligned}
k^* &:= \arg \max_{\bar{k} \in [0, k]} \bar{k} \\
&\text{subject to } g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
&\quad + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\
&\quad + (1 - \eta_{c,j}) \bar{\kappa}_{p,j} \tau (\tau_{k+1} - \tau_{\bar{k}}) \leq 0, \\
&\quad \forall j = 1, \dots, n_{g_p},
\end{aligned} \quad (71)$$

$$\begin{aligned}
k^* &:= \\
&\arg \min_{\bar{k} \in [0, k]} \max_{j=1, \dots, n_{g_p}} \begin{bmatrix} g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j} \tau (\tau_{k+1} - \tau_{\bar{k}}) \end{bmatrix}. \quad (72)
\end{aligned}$$

To demonstrate the potential benefits of these relaxations, let us again consider the (–) case of Problem (42) and apply the modified SCFO. For $g_{p,2}$, it is clear that we can assume global partial concavity – in fact, linearity – in both u_2 and τ , with $I_{c,2} := \{2\}$ and $\eta_{c,2} := 1$. For $g_{p,1}$, there is actually a choice due to the bilinear term τu_1 , which is not concave. More specifically, we can either set $I_{c,1} := \{1, 2\}$, $\eta_{c,1} := 0$ or $I_{c,1} := \{2\}$, $\eta_{c,1} := 1$, as partial concavity in τ and u_1 is mutually exclusive. Since the decision variable u_1 has a much greater influence on the function value than τ , we choose the former as this is expected to lead to more significant relaxations in the SCFO. The results are given in Figs. 7 and 8, and show that the number of experiments needed to arrive in the neighborhood of the optimum is reduced drastically (compare with Fig. 5), as the relaxations in the feasibility condition allow much bigger steps and thus faster progress.

While impressive, such relaxations are of course of limited use in practice, as in many experimental problems one may not know about the existence of concave relationships even when they are present. A natural case where one might try to apply these relaxations would be when a sufficiently accurate surrogate model of the experimental function $g_{p,j}$ is available and exhibits obvious concave behavior in certain variables (e.g., the power-current relationship in the fuel cell system [25]). When such a model is not available, an alternative may be to regress the available experimental data with concave models as a means of “discovering” or validating the concavity assumption [3, 34, 5]. However, as falsely assuming concavity could lead to constraint violations due to unjustified relaxations in the Lipschitz bounds, one should always remain careful when making this assumption in practice.

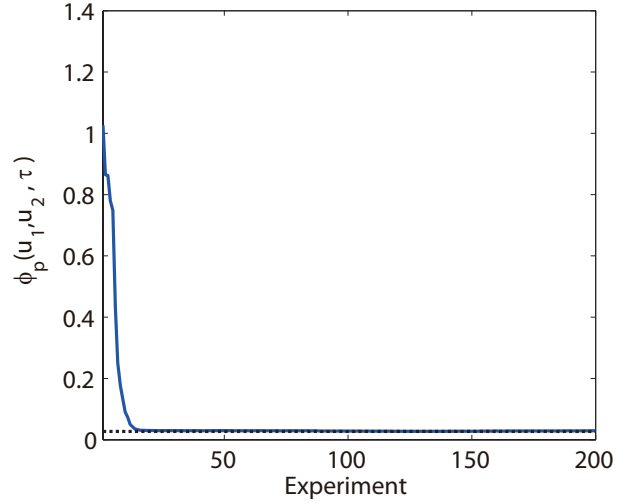


Fig. 8 Cost function values obtained by the modified SCFO methodology for Problem (42) for the (–) scenario with concavity relaxations included.

3.3 Using All Available Data

So far we have only considered stating the feasibility condition (68) with respect to the future experimental iteration $k+1$ and the reference iteration k^* . However, as the bound (57) applies everywhere on \mathcal{S}_τ , it follows that any other iteration from 0 to k may be used in place of k^* without endangering the validity of the bound. Because every one of the $k+1$ bounds is valid, it follows that their minimum is valid as well, and so we may tighten (57) further:

$$\begin{aligned}
g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) &\leq \\
\min_{\bar{k}=0, \dots, k} &\begin{bmatrix} g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j} \tau (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \\ + \sum_{i \notin I_{c,j}} \max \left[\frac{\underline{\kappa}_{p,ji} (u_{k+1,i} - u_{\bar{k},i})}{\bar{\kappa}_{p,ji} (u_{k+1,i} - u_{\bar{k},i})} \right] \end{bmatrix}. \quad (73)
\end{aligned}$$

Forcing the right-hand side to be negative then leads to the relaxation of Condition (68):

$$\begin{aligned}
\min_{\bar{k}=0, \dots, k} &\begin{bmatrix} g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j} \tau (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \\ + \sum_{i \notin I_{c,j}} \max \left[\frac{\underline{\kappa}_{p,ji} (u_{k+1,i} - u_{\bar{k},i})}{\bar{\kappa}_{p,ji} (u_{k+1,i} - u_{\bar{k},i})} \right] \end{bmatrix} \leq 0, \quad (74)
\end{aligned}$$

with the corresponding modification in the project-and-filter approach:

$$\min_{\bar{k}=0,\dots,k} \left[\begin{array}{l} g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,j}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \\ + \sum_{i \notin I_{c,j}} \max \left[\begin{array}{l} \underline{\kappa}_{p,ji}(u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \bar{\kappa}_{p,ji}(u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \leq 0. \quad (75)$$

In general, one would expect the relaxation to be useful when an algorithm revisits parts of the experimental space where it has been before, as the Lipschitz bounds essentially construct small feasibility-guaranteeing regions around each $\mathbf{u}_{\bar{k}}$ [9]. We list some scenarios where an algorithm might visit the same region twice:

- Errors in the gradients (Section 5) lead to haphazard behavior until better gradient estimates are obtained.
- The algorithm converges in a zig-zag manner.
- After the algorithm converges to the optimum of (1), the cost function changes – due to, e.g., a change in market demand – but the feasible region remains the same. As such, the algorithm must now navigate to a new optimum, possibly revisiting regions of the experimental space where it has been already.
- Additional experiments have been carried out prior to \mathbf{u}_0 , and may be included in the data set.

We illustrate the potential usefulness of this relaxation by considering an example of the third point. Let us imagine an experimental optimization where the problem to be solved is originally (41) – we will, for simplicity, make degradation effects negligible – but which then changes to

$$\begin{aligned} & \underset{u_1, u_2}{\text{minimize}} && (u_1 + 0.25)^2 + (u_2 - 0.6)^2 \\ & \text{subject to} && -6u_1^2 - 3.5u_1 + u_2 - 0.6 \leq 0 \\ & && 2u_1^2 + 0.5u_1 + u_2 - 0.75 \leq 0 \\ & && -u_1^2 - (u_2 - 0.15)^2 + 0.01 \leq 0 \\ & && -0.5 \leq u_1 \leq 0.5 \\ & && 0 \leq u_2 \leq 0.8 \end{aligned} \quad (76)$$

for $k > 50$ (i.e., the cost function changes after 50 experimental iterations). We will assume that the new cost function has not been evaluated during the first 51 experiments, but that the constraint functions, which remain the same, have and may be used in the relaxation (75).

The results for the case when this relaxation is not employed are given first (Fig. 9), and show that the algorithm successfully converges to the first optimum and then to the second. However, because the initial point for the second problem is very close to an experimental constraint, the initial steps during the second phase are forced to be small so as to not violate $g_{p,2}(\mathbf{u}) \leq 0$. We now compare this to the case where the relaxation is used (Fig. 10) and see that, while there is no change in convergence behavior with respect to the first optimum, the second phase is radically different, the algorithm being allowed to take a very big step away from the constraint and to the neighborhood of one of the previous experimental iterates. The result is that the optimum is approached much quicker for the second cost function when this relaxation is used.

3.4 Additional Safeguards to Guarantee Cost Decrease

The Lipschitz constants on the cost function derivatives, M_ϕ , are expected to be more difficult to estimate than the Lipschitz constants on the function κ , as the former correspond to lower and upper bounds on the second, rather than first, derivatives. As such, it may occur that the choice of M_ϕ is not sufficiently conservative to guarantee (50), which could then lead to the scheme not converging due to a lack of cost decrease from one experiment to the next. We present here one additional condition that may act as a support to (55) and possibly help enforce convergence when (55) does not due to a poor choice of Lipschitz constants or due to errors in the cost function gradient.

We start by defining additional Lipschitz constants for the cost function:

$$\underline{\kappa}_{\phi,i} \leq \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}, \tau)} \leq \bar{\kappa}_{\phi,i}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{J}_\tau, \quad (77)$$

$$\underline{\kappa}_{\phi,\tau} \leq \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}, \tau)} \leq \bar{\kappa}_{\phi,\tau}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{J}_\tau, \quad (78)$$

noting that we do not require these to be strict. Since the SCFO aim to decrease the cost function value between $(\mathbf{u}_{k^*}, \tau_{k+1})$ and $(\mathbf{u}_{k+1}, \tau_{k+1})$, it follows that we should cross out of consideration all \mathbf{u}_{k+1} where it is guaranteed that $\phi_p(\mathbf{u}_{k+1}, \tau_{k+1}) \geq \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})$. Since the value of $\phi_p(\mathbf{u}_{k+1}, \tau_{k+1})$ cannot be known without testing the experimental point \mathbf{u}_{k+1} , one may proceed by working with its *lower bound*

$$\begin{aligned} \phi_p(\mathbf{u}_{k+1}, \tau_{k+1}) & \geq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \underline{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) + \sum_{i=1}^{n_u} \min \left[\frac{\underline{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i})}{\bar{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i})} \right], \end{aligned} \quad (79)$$

which, however, may be tightened further by supposing the cost function to be globally partially *convex* in certain

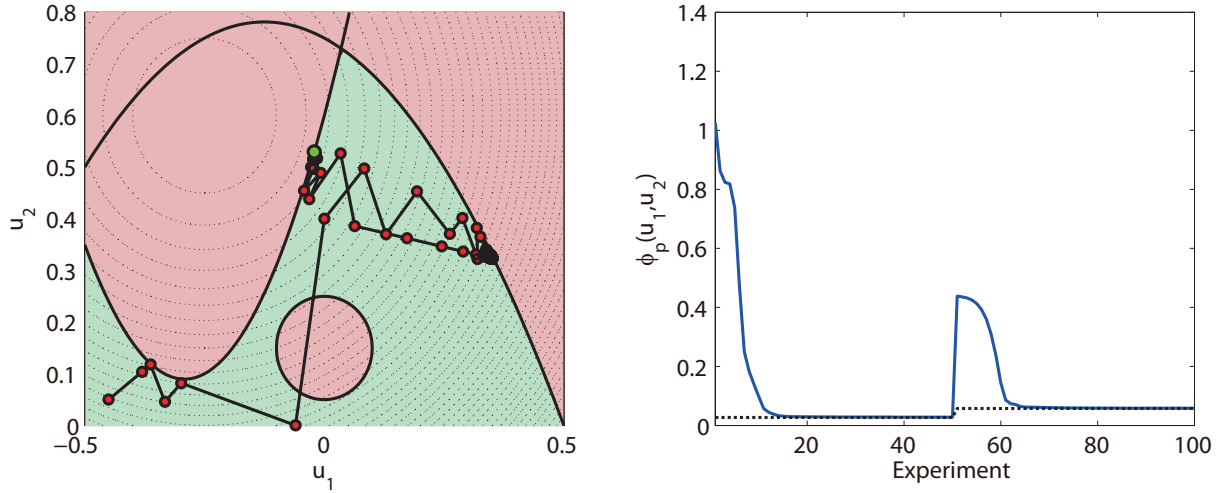


Fig. 9 Solution generated by the modified SCFO methodology for the switching cost Problem (41)/(76) when the relaxed condition (75) is not used. The cost contours and minimum given in the left-hand-side plot are only those of the second cost function, while the minimum cost value in the right-hand-side plot switches accordingly after the 50th experiment.

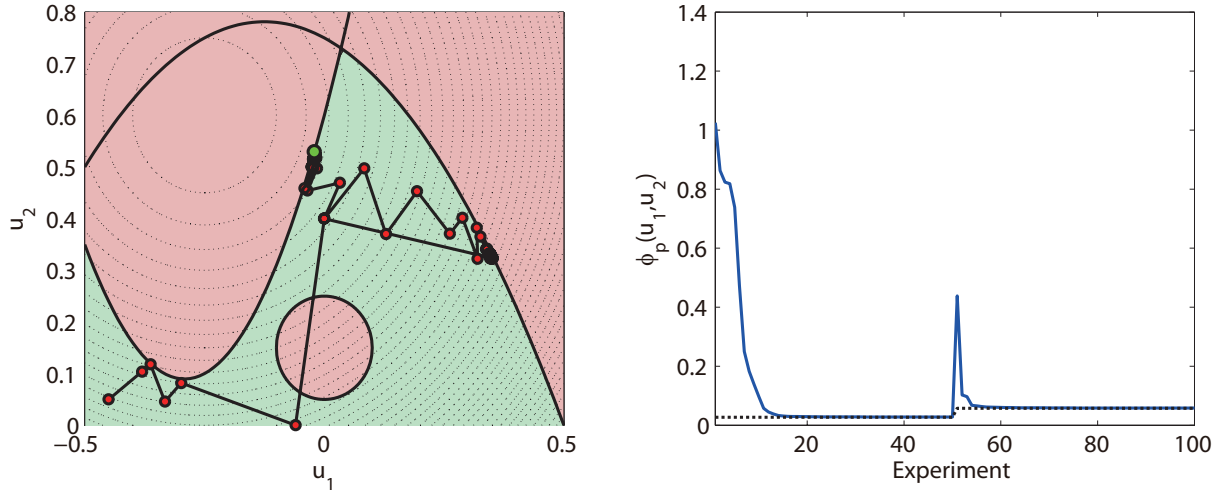


Fig. 10 Solution generated by the modified SCFO methodology for the switching cost Problem (41)/(76) when the relaxed condition (75) is used.

variables, in a manner analogous to what was done for the constraints in Section 3.2. For the sake of completeness, we state here the analogous definition and lemma.

Definition 2 (Partial convexity of cost function) Let the decision variables \mathbf{u} be partitioned into subvectors \mathbf{v} and \mathbf{z} , so that $\phi_p(\mathbf{u}, \tau) = \phi_p(\mathbf{v}, \mathbf{z}, \tau)$. The function ϕ_p will be said to be globally *partially convex* in \mathbf{z} if it is convex everywhere on \mathcal{S}_τ for \mathbf{v} and τ fixed. Likewise, ϕ_p will be said to be globally partially convex in both \mathbf{z} and τ if it is convex everywhere on \mathcal{S}_τ for \mathbf{v} fixed.

Lemma 2 (Convex tightening of lower Lipschitz bound) Let $I_{v,\phi}$ denote the set of indices defining the variables \mathbf{z} , with respect to which the experimental cost function ϕ_p is globally partially convex. Furthermore, let $\eta_{v,\phi}$ denote a

Boolean indicator that is equal to 1 if ϕ_p is globally partially convex in both \mathbf{z} and τ , and is equal to 0 otherwise. It follows that the following tightening of (79) is valid for any $\bar{k} = 0, \dots, k$:

$$\begin{aligned} \phi_p(\mathbf{u}_{k+1}, \tau_{k+1}) &\geq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ &+ \eta_{v,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ &+ (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ &+ \sum_{i \in I_{v,\phi}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \\ &+ \sum_{i \notin I_{v,\phi}} \min \begin{bmatrix} \underline{\kappa}_{\phi,i} (u_{k+1,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i} (u_{k+1,i} - u_{\bar{k},i}) \end{bmatrix}. \end{aligned} \quad (80)$$

where $i \notin I_{v,\phi}$ implicitly denotes $\{1, \dots, n_u\} \setminus I_{v,\phi}$.

Proof The proof is identical to that of Lemma 1, save that many things are reversed to account for lower bounding instead of upper bounding and for convexity instead of concavity. We do not consider it necessary to repeat the steps here and leave this to the reader. \square

It follows from (80) that any choice of \mathbf{u}_{k+1} for which

$$\begin{aligned} & \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{v,\phi}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{v,\phi}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i}) \end{array} \right] \geq \phi_p(\mathbf{u}_{k^*}, \tau_{k+1}) \end{aligned} \quad (81)$$

is invalid, as this guarantees that $\phi_p(\mathbf{u}_{k+1}, \tau_{k+1}) \geq \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})$.

Likewise, as the actual value of $\phi_p(\mathbf{u}_{k^*}, \tau_{k+1})$ is not available, we must instead work with its upper bound. Denoting by $I_{c,\phi}$ and $\eta_{c,\phi}$ the cost *concavity* indicators, we may employ a bound analogous to (57):

$$\begin{aligned} \phi_p(\mathbf{u}_{k^*}, \tau_{k+1}) & \leq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \eta_{c,\phi} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,\phi}) \bar{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,\phi}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{c,\phi}} \max \left[\begin{array}{l} \underline{\kappa}_{\phi,i}(u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}(u_{k^*,i} - u_{\bar{k},i}) \end{array} \right], \end{aligned} \quad (82)$$

which is valid for any $\bar{k} = 0, \dots, k$.

This bound may then be used in an implementable version of (81):

$$\begin{aligned} & \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{v,\phi}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{v,\phi}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i}) \end{array} \right] \\ & \geq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \eta_{c,\phi} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,\phi}) \bar{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,\phi}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{c,\phi}} \max \left[\begin{array}{l} \underline{\kappa}_{\phi,i}(u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}(u_{k^*,i} - u_{\bar{k},i}) \end{array} \right], \end{aligned} \quad (83)$$

as any \mathbf{u}_{k+1} that satisfies (83) must also satisfy (81), and thus may be considered as invalid.

As in the previous subsection, we may take \bar{k} to be any of the past experimental iterations. Since satisfying (83) with respect to any single iteration is sufficient to fathom a given \mathbf{u}_{k+1} from consideration, one may take the maximum of the left-hand side of (83) over all experimental iterations:

$$\begin{aligned} & \max_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{v,\phi}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \\ + \sum_{i \notin I_{v,\phi}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \\ & \geq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \eta_{c,\phi} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,\phi}) \bar{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,\phi}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{c,\phi}} \max \left[\begin{array}{l} \underline{\kappa}_{\phi,i}(u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}(u_{k^*,i} - u_{\bar{k},i}) \end{array} \right]. \end{aligned} \quad (84)$$

The same may be done for the right-hand side, save that one chooses the \bar{k} that *minimizes* the upper bound:

$$\begin{aligned} & \max_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{v,\phi}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \\ + \sum_{i \notin I_{v,\phi}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}(u_{k+1,i} - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \\ & \geq \min_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,\phi} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}) \bar{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,\phi}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \\ + \sum_{i \notin I_{c,\phi}} \max \left[\begin{array}{l} \underline{\kappa}_{\phi,i}(u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}(u_{k^*,i} - u_{\bar{k},i}) \end{array} \right] \end{array} \right]. \end{aligned} \quad (85)$$

Since $\phi_p(\mathbf{u}_{k+1}, \tau_{k+1}) \geq \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})$ is guaranteed for any \mathbf{u}_{k+1} that satisfies (85), we obtain a necessary condition for cost decrease by reversing the inequality sign:

$$\begin{aligned}
& \max_{\bar{k}=0,\dots,k} \left[\begin{aligned} & \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{v,\phi}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{v,\phi}} \min \left[\begin{aligned} & \underline{\kappa}_{\phi,i} (u_{k+1,i} - u_{\bar{k},i}), \\ & \bar{\kappa}_{\phi,i} (u_{k+1,i} - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right] \\
& \leq \min_{\bar{k}=0,\dots,k} \left[\begin{aligned} & \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,\phi}) \bar{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,\phi}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{c,\phi}} \max \left[\begin{aligned} & \underline{\kappa}_{\phi,i} (u_{k^*,i} - u_{\bar{k},i}), \\ & \bar{\kappa}_{\phi,i} (u_{k^*,i} - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right]. \quad (86)
\end{aligned}$$

Substituting the filtering law into (86) then leads to the following condition on K_k in the project-and-filter approach:

$$\begin{aligned}
& \max_{\bar{k}=0,\dots,k} \left[\begin{aligned} & \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{v,\phi}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & \quad K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{v,\phi}} \min \left[\begin{aligned} & \underline{\kappa}_{\phi,i} (u_{k^*,i} + \\ & \quad K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{\kappa}_{\phi,i} (u_{k^*,i} + \\ & \quad K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right] \\
& \leq \min_{\bar{k}=0,\dots,k} \left[\begin{aligned} & \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,\phi}) \bar{\kappa}_{\phi,\tau} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,\phi}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \\ & + \sum_{i \notin I_{c,\phi}} \max \left[\begin{aligned} & \underline{\kappa}_{\phi,i} (u_{k^*,i} - u_{\bar{k},i}), \\ & \bar{\kappa}_{\phi,i} (u_{k^*,i} - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right]. \quad (87)
\end{aligned}$$

Since we are now supposing the very practical possibility of the scheme failing to satisfy the sufficient conditions for cost decrease – due to, for example, an improper choice of Lipschitz constants – it makes sense to modify the way in which we choose the reference point as well, since the last experimental iterate for which feasibility may be guaranteed is no longer guaranteed to be the best as the cost is not guaranteed to decrease from experiment to experiment. What we propose is to take as a reference the latest experimental iterate for which feasibility may be

guaranteed and which *cannot be proven to have a cost function value greater to that at another provably feasible point*.

The issue that arises here is that it is not sufficient to simply compare the cost function values for the different experiments, since $\phi_p(\mathbf{u}_{k-1}, \tau_{k-1}) < \phi_p(\mathbf{u}_k, \tau_k)$ need not imply $\phi_p(\mathbf{u}_{k-1}, \tau_k) < \phi_p(\mathbf{u}_k, \tau_k)$ due to degradation effects. As such, one could not say that \mathbf{u}_{k-1} is a better reference point than \mathbf{u}_k by simply comparing their measured values at different time instants. A robust approach would be to use an upper bound on $\phi_p(\mathbf{u}_{k-1}, \tau_k)$ in the comparison instead. As one example, consider the bound

$$\phi_p(\mathbf{u}_{k-1}, \tau_k) \leq \phi_p(\mathbf{u}_{k-1}, \tau_{k-1}) + \bar{\kappa}_{\phi,\tau} (\tau_k - \tau_{k-1}). \quad (88)$$

Clearly, if the right-hand side is inferior to $\phi_p(\mathbf{u}_k, \tau_k)$, then this implies $\phi_p(\mathbf{u}_{k-1}, \tau_k) < \phi_p(\mathbf{u}_k, \tau_k)$ and we can safely say that \mathbf{u}_{k-1} is a better reference than \mathbf{u}_k because it has a lower cost function value at time instant τ_k .

The above example is special in that we do not need a lower or upper bound on $\phi_p(\mathbf{u}_k, \tau_k)$, as this value is current and does not require accounting for degradation. In general, however, we may want to be able to compare the *current* cost function values for any two past experimental iterates. For example, in choosing a more appropriate reference between \mathbf{u}_{k-1} and \mathbf{u}_{k-2} , we will give our preference to \mathbf{u}_{k-1} , since it comes later, provided that it cannot be proven to have a cost function value greater than that at \mathbf{u}_{k-2} , i.e., provided that we cannot prove $\phi_p(\mathbf{u}_{k-1}, \tau_k) > \phi_p(\mathbf{u}_{k-2}, \tau_k)$.

To prove this would entail using *two* bounds and showing that the lower bound on $\phi_p(\mathbf{u}_{k-1}, \tau_k)$ is strictly greater than the upper bound on $\phi_p(\mathbf{u}_{k-2}, \tau_k)$. For a general \bar{k} , such bounds are obtained as

$$\phi_p(\mathbf{u}_{\bar{k}}, \tau_k) \leq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) + (1 - \eta_{c,\phi}) \bar{\kappa}_{\phi,\tau} (\tau_k - \tau_{\bar{k}}), \quad (89)$$

$$\phi_p(\mathbf{u}_{\bar{k}}, \tau_k) \geq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) + (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_k - \tau_{\bar{k}}). \quad (90)$$

Using these results, we proceed to replace (71) by

$$\begin{aligned}
& k^* := \\
& \arg \text{maximize}_{\bar{k} \in [0,k]} \bar{k} \\
& \text{subject to } g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
& \quad + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\
& \quad + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0, \\
& \quad \forall j = 1, \dots, n_{g_p}
\end{aligned} \quad (91)$$

$$\begin{aligned} & \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ & + (1 - \eta_{v,\phi}) \underline{\kappa}_{\phi,\tau} (\tau_k - \tau_{\bar{k}}) \leq \\ \min_{\bar{k} \in \mathbf{k}_f} & \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \\ \eta_{c,\phi} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}) \bar{\kappa}_{\phi,\tau} (\tau_k - \tau_{\bar{k}}) \end{array} \right], \\ \mathbf{k}_f = & \left\{ \bar{k} : \begin{array}{l} g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0, \\ \forall j = 1, \dots, n_{g_p} \end{array} \right\}, \end{aligned} \quad (92)$$

which chooses the most recent point while guaranteeing that feasibility may be maintained for this point and that the corresponding cost function value at time instant τ_k will not be greater than the value for any other points for which feasibility is guaranteed. In the case where this problem has no solution, a possible implementation, apart from simply terminating the procedure, is to fall back on (72) in hopes of finding a feasible point.

To illustrate the potential usefulness of (87), let us consider the following modified version of (41):

$$\begin{aligned} & \underset{u_1, u_2}{\text{minimize}} \quad \phi_p(\mathbf{u}) := (u_1 - 0.2)^2 + (u_2 - 0.4)^2 \\ & \text{subject to} \quad g_{p,1}(\mathbf{u}) := -6u_1^2 - 3.5u_1 + u_2 - 0.6 \leq 0 \\ & \quad \quad \quad g_{p,2}(\mathbf{u}) := 2u_1^2 + 0.5u_1 + u_2 - 0.75 \leq 0 \\ & \quad \quad \quad g_1(\mathbf{u}) := -u_1^2 - (u_2 - 0.15)^2 + 0.01 \leq 0 \\ & \quad \quad \quad -0.5 \leq u_1 \leq 0.5 \\ & \quad \quad \quad 0 \leq u_2 \leq 0.8, \end{aligned} \quad (93)$$

where we again, for simplicity, ignore degradation effects. Because this problem has an unconstrained optimum, it follows that the Lipschitz constants on the cost derivatives will play a more crucial role – when an optimum is constrained, errors in the sufficient condition (55) for cost decrease are often compensated for by the conservatism introduced by the feasibility-guaranteeing conditions, which lower K_k sufficiently even when the cost decrease condition does not.

We will suppose that an improper choice of M_ϕ has been made, with

$$\begin{aligned} \underline{M}_{\phi,11} &= 0.1, & \underline{M}_{\phi,12} &= -2, \\ \underline{M}_{\phi,11} &= 0.5, & \underline{M}_{\phi,12} &= -1.5, \\ \underline{M}_{\phi,21} &= -2, & \underline{M}_{\phi,22} &= 0.1, \\ \underline{M}_{\phi,21} &= -1.5, & \underline{M}_{\phi,22} &= 0.5, \end{aligned} \quad (94)$$

which clearly does not satisfy (45) for the ϕ_p given in (93). Employing (91) to select k^* but not employing (87), we see

that the scheme fails to converge to the optimum (Fig. 11). What essentially happens here is that the scheme finds a good reference at approximately $\mathbf{u} = (0.15, 0.4)$ and then proceeds to take overly big steps in the positive u_1 direction. As no sort of feedback is applied to tell the algorithm that smaller steps are needed, it simply continues to take the same large steps due to the improper choice of M_ϕ and then going back to the best point as its reference.

By contrast, enforcing (87) with the Lipschitz constants

$$\begin{aligned} \underline{\kappa}_{\phi,1} &= -1.4, & \underline{\kappa}_{\phi,2} &= -0.8, \\ \bar{\kappa}_{\phi,1} &= 0.6, & \bar{\kappa}_{\phi,2} &= 0.8 \end{aligned} \quad (95)$$

and the convexity relaxations defined by $I_{v,\phi} = \{1, 2\}$ actually allows convergence to the optimum despite the poor choice of M_ϕ (Fig. 12). The reason is simple: although the algorithm plans to take an overly large step, it does not because doing so would lead to an experimental iterate for which the cost is guaranteed to be higher than that at the reference. As such, the steps are gradually shortened as the experimental iterates are driven to the optimum.

We finish by reflecting on the implementability of this technique. Clearly, the price to pay for employing it comes via the need to provide *even more* Lipschitz constants – those for the cost function – which may be seen as undesirable. However, as the constants κ_ϕ are bounds on the *first-order* derivatives, they are expected to be much easier to obtain. Additionally, choosing very conservative values for κ_ϕ would not be detrimental to the performance of the scheme and would only reduce the usefulness of (87). As before, we give the standard warnings associated with concavity and convexity relaxations – i.e., that one should only use these when they are well justified and the concave/convex relationships are well documented. While it might even be possible to prove that the necessary condition for cost decrease (87) becomes *sufficient* asymptotically as more and more regions are explored and fathomed from consideration until only the cost descent regions are left, we do not attempt to do so here and only propose (87) as a useful supplement to (55).

3.5 Methods of Estimating Lipschitz Constants

Here, we discuss the possible means of estimating the Lipschitz constants for the experimental functions (κ) and the cost function derivatives (M_ϕ). Three approaches are proposed and may be classified as being based on:

- physical laws,
- conservative model-based estimates,
- consistency checks.

We note that these three are not mutually exclusive and may often be complementary.

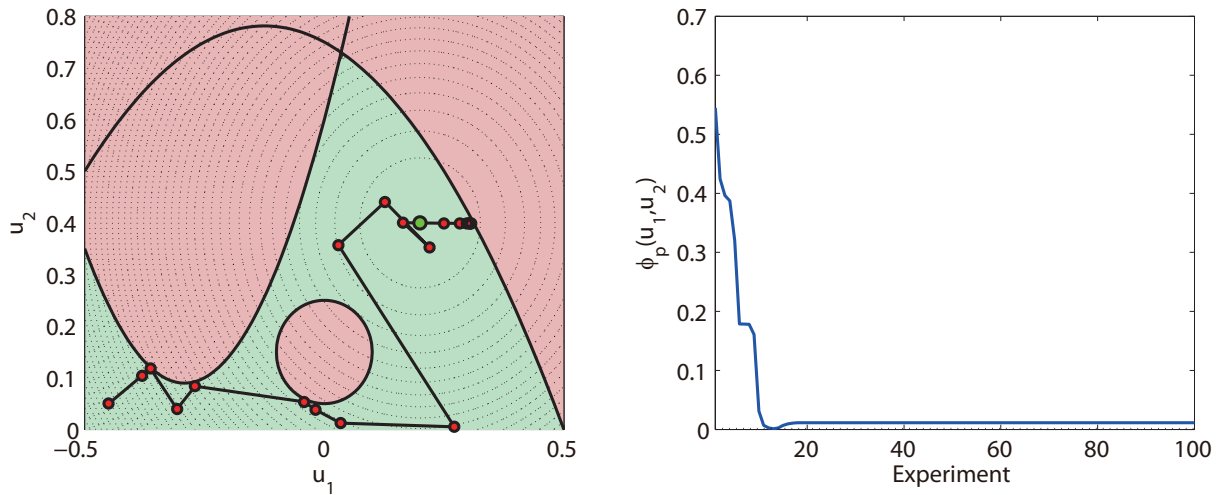


Fig. 11 Solution generated by the modified SCFO methodology for the unconstrained-optimum problem (93) when the necessary condition (87) is not enforced and the Lipschitz constants M_ϕ are improperly chosen.

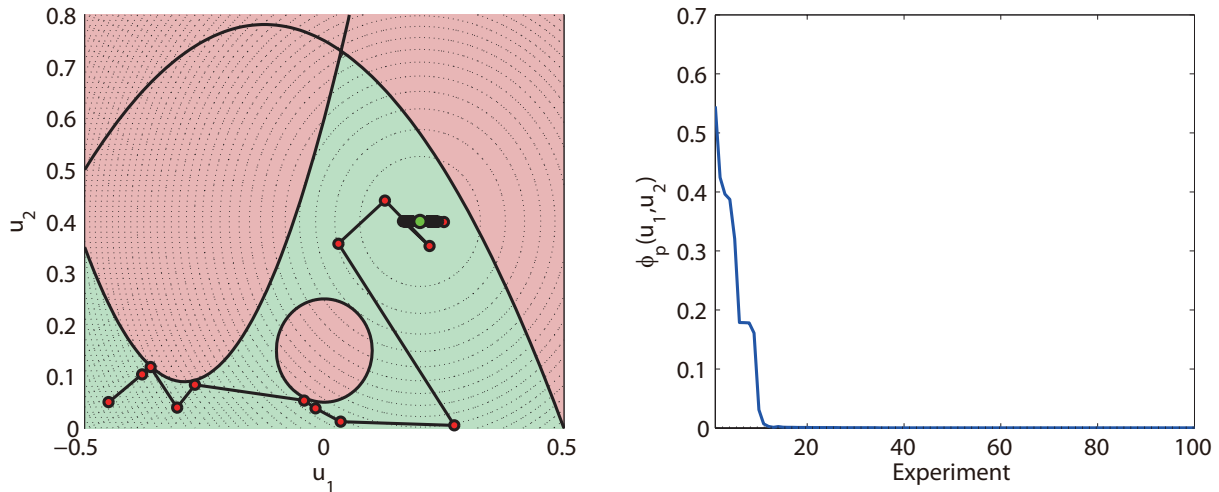


Fig. 12 Solution generated by the modified SCFO methodology for the unconstrained-optimum problem (93) when the necessary condition (87) is enforced.

3.5.1 The Physical Meaning of a Lipschitz Constant

As already illustrated with some examples in Section 3.1, the sign of the Lipschitz constants κ may be known with certainty in some applications due to underlying physical principles or simply due to engineering experience. As each Lipschitz constant is essentially a lower or upper bound on a given derivative, or sensitivity, of an experimental function, such knowledge allows either the lower or upper constant to be set to 0.

In other cases one may be able to limit not only the sign but also the magnitude – consider, for example, a heat exchanger where a hot element comes into contact with a colder one, thereby heating it. In many such set-ups, one may raise or lower the temperature of the heating element to raise or lower the temperature of the heated element,

but usually by no more than the change in temperature in the heating element – i.e., one cannot raise the temperature of the heating element by 5°C and expect the temperature of the heated element to go up by *more than* 5°C . If we now imagine the temperature of the heated element to be an experimental function and the temperature of the heating element to be a decision variable, then it is clear that the upper Lipschitz constant that relates the two cannot exceed 1.

Finally, if a certain decision variable is known to have a negligible effect on the function, its corresponding Lipschitz constants may be set to very small negative and positive values (or to 0, if strict inequality is not needed).

For the cost function derivative Lipschitz constants M_ϕ the physical significance may be less obvious, and there

will likely be less *a priori* knowledge to exploit. As already mentioned, the curvature of the cost function with respect to certain decision variables may be known for certain applications where the function's convexity or concavity is well documented. If this is so, then the signs of the different M_ϕ may be restricted. Furthermore, if the function is known to be linear in certain variables, then all of the second derivatives that include this variable are 0, and so the corresponding Lipschitz constants may be set to very small negative and positive values.

3.5.2 Model-Based Estimation

In many engineering applications, one will often have a model of the process or system being optimized, which, though not a perfect description, will nevertheless capture the basic trends and behaviors. Such models will often be parameterized by some finite set of parameters θ and will approximate the true experimental functions as

$$\hat{\phi}_p(\mathbf{u}, \tau, \theta) \approx \phi_p(\mathbf{u}, \tau), \quad (96)$$

with $\hat{\phi}$ used to denote the model (we use the cost function as the example).

As the model is a known numerical function, one may compute its Lipschitz constants directly. In the general case, we may suppose the parameters θ to belong to an uncertain set Θ , and may compute the model constants as

$$\underline{\kappa}_{\hat{\phi},i} = \min_{\substack{\mathbf{u}, \tau \in \mathcal{I}_\tau \\ \theta \in \Theta}} \left. \frac{\partial \hat{\phi}_p}{\partial u_i} \right|_{(\mathbf{u}, \tau, \theta)}, \quad (97)$$

$$\bar{\kappa}_{\hat{\phi},i} = \max_{\substack{\mathbf{u}, \tau \in \mathcal{I}_\tau \\ \theta \in \Theta}} \left. \frac{\partial \hat{\phi}_p}{\partial u_i} \right|_{(\mathbf{u}, \tau, \theta)},$$

$$\underline{\kappa}_{\hat{\phi},\tau} = \min_{\substack{\mathbf{u}, \tau \in \mathcal{I}_\tau \\ \theta \in \Theta}} \left. \frac{\partial \hat{\phi}_p}{\partial \tau} \right|_{(\mathbf{u}, \tau, \theta)}, \quad (98)$$

$$\bar{\kappa}_{\hat{\phi},\tau} = \max_{\substack{\mathbf{u}, \tau \in \mathcal{I}_\tau \\ \theta \in \Theta}} \left. \frac{\partial \hat{\phi}_p}{\partial \tau} \right|_{(\mathbf{u}, \tau, \theta)},$$

with the $M_{\hat{\phi}}$ values computed in an analogous manner. Note that such computations may not be easy – if the model used is highly nonlinear/nonconvex in \mathbf{u} and τ , robust global optimization techniques may be required. However, this is likely to be a secondary concern as all the relevant computations may be done offline prior to experimental optimization.

To account for potential model inaccuracies, one may add an additional layer of safety to the estimates by setting the actual Lipschitz constant estimates as

$$\begin{aligned} \underline{\kappa}_{\phi,i} &:= \underline{\kappa}_{\hat{\phi},i} - 0.5(\bar{\kappa}_{\hat{\phi},i} - \underline{\kappa}_{\hat{\phi},i}), \\ \bar{\kappa}_{\phi,i} &:= \bar{\kappa}_{\hat{\phi},i} + 0.5(\bar{\kappa}_{\hat{\phi},i} - \underline{\kappa}_{\hat{\phi},i}), \\ \underline{\kappa}_{\phi,\tau} &:= \underline{\kappa}_{\hat{\phi},\tau} - 0.5(\bar{\kappa}_{\hat{\phi},\tau} - \underline{\kappa}_{\hat{\phi},\tau}), \\ \bar{\kappa}_{\phi,\tau} &:= \bar{\kappa}_{\hat{\phi},\tau} + 0.5(\bar{\kappa}_{\hat{\phi},\tau} - \underline{\kappa}_{\hat{\phi},\tau}), \end{aligned} \quad (99)$$

with analogous settings for M_ϕ .

Note, as well, that the model need not be available *a priori* and could also be constructed from a set of initial experiments, as is standardly done in the response-surface methodology [28]. Such models are typically linear or quadratic and are easier to work with, although they may not capture all of the involved characteristics of the process or system in consideration. This data-driven approach is particularly well-suited for experimental functions that are very difficult to model and for which the underlying physics are not well documented or well explained.

3.5.3 Consistency Checks

We will consider a given choice of Lipschitz constants *consistent* if they satisfy the Lipschitz bounds for all pairs of available experimental iterates. In particular, letting k_1 and k_2 denote any two iterations from the interval $[0, k]$, we know that for a consistent choice of Lipschitz constants the following relations must hold:

$$\begin{aligned} g_{p,j}(\mathbf{u}_{k_2}, \tau_{k_2}) &\leq g_{p,j}(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &\quad + \max \left[\frac{\underline{\kappa}_{p,j\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{p,j\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &\quad + \sum_{i=1}^{n_u} \max \left[\frac{\underline{\kappa}_{p,ji}(u_{k_2,i} - u_{k_1,i})}{\bar{\kappa}_{p,ji}(u_{k_2,i} - u_{k_1,i})} \right], \end{aligned} \quad (100)$$

$$\begin{aligned} g_{p,j}(\mathbf{u}_{k_2}, \tau_{k_2}) &\geq g_{p,j}(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &\quad + \min \left[\frac{\underline{\kappa}_{p,j\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{p,j\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &\quad + \sum_{i=1}^{n_u} \min \left[\frac{\underline{\kappa}_{p,ji}(u_{k_2,i} - u_{k_1,i})}{\bar{\kappa}_{p,ji}(u_{k_2,i} - u_{k_1,i})} \right], \end{aligned} \quad (101)$$

$$\begin{aligned} \phi_p(\mathbf{u}_{k_2}, \tau_{k_2}) &\leq \phi_p(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &\quad + \max \left[\frac{\underline{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &\quad + \sum_{i=1}^{n_u} \max \left[\frac{\underline{\kappa}_{\phi,i}(u_{k_2,i} - u_{k_1,i})}{\bar{\kappa}_{\phi,i}(u_{k_2,i} - u_{k_1,i})} \right], \end{aligned} \quad (102)$$

$$\begin{aligned} \phi_p(\mathbf{u}_{k_2}, \tau_{k_2}) &\geq \phi_p(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &\quad + \min \left[\frac{\underline{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &\quad + \sum_{i=1}^{n_u} \min \left[\frac{\underline{\kappa}_{\phi,i}(u_{k_2,i} - u_{k_1,i})}{\bar{\kappa}_{\phi,i}(u_{k_2,i} - u_{k_1,i})} \right], \end{aligned} \quad (103)$$

$$\begin{aligned}
& \phi_p(\mathbf{u}_{k_2}, \tau_{k_2}) \leq \\
& \phi_p(\mathbf{u}_{k_1}, \tau_{k_1}) + \nabla \phi_p(\mathbf{u}_{k_1}, \tau_{k_1})^T \begin{bmatrix} \mathbf{u}_{k_2} - \mathbf{u}_{k_1} \\ 0 \end{bmatrix} \\
& + \max \left[\frac{\underline{\kappa}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1})}{\overline{\kappa}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1})}, \right. \\
& \left. + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \begin{bmatrix} \underline{M}_{\phi, i_1 i_2}(u_{k_2, i_1} - u_{k_1, i_1}) \\ (u_{k_2, i_2} - u_{k_1, i_2}), \\ \overline{M}_{\phi, i_1 i_2}(u_{k_2, i_1} - u_{k_1, i_1}) \\ (u_{k_2, i_2} - u_{k_1, i_2}) \end{bmatrix}, \right]
\end{aligned} \tag{104}$$

$$\begin{aligned}
& \phi_p(\mathbf{u}_{k_2}, \tau_{k_2}) \geq \\
& \phi_p(\mathbf{u}_{k_1}, \tau_{k_1}) + \nabla \phi_p(\mathbf{u}_{k_1}, \tau_{k_1})^T \begin{bmatrix} \mathbf{u}_{k_2} - \mathbf{u}_{k_1} \\ 0 \end{bmatrix} \\
& + \min \left[\frac{\underline{\kappa}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1})}{\overline{\kappa}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1})}, \right. \\
& \left. + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \min \begin{bmatrix} \underline{M}_{\phi, i_1 i_2}(u_{k_2, i_1} - u_{k_1, i_1}) \\ (u_{k_2, i_2} - u_{k_1, i_2}), \\ \overline{M}_{\phi, i_1 i_2}(u_{k_2, i_1} - u_{k_1, i_1}) \\ (u_{k_2, i_2} - u_{k_1, i_2}) \end{bmatrix}, \right]
\end{aligned} \tag{105}$$

where the last two bounds may be derived by considering the decomposition

$$\begin{aligned}
& \phi_p(\mathbf{u}_{k_2}, \tau_{k_2}) - \phi_p(\mathbf{u}_{k_1}, \tau_{k_1}) = \\
& \phi_p(\mathbf{u}_{k_2}, \tau_{k_2}) - \phi_p(\mathbf{u}_{k_2}, \tau_{k_1}) \\
& + \phi_p(\mathbf{u}_{k_2}, \tau_{k_1}) - \phi_p(\mathbf{u}_{k_1}, \tau_{k_1}),
\end{aligned} \tag{106}$$

with the linear Lipschitz bound then applied to the first addend, $\phi_p(\mathbf{u}_{k_2}, \tau_{k_2}) - \phi_p(\mathbf{u}_{k_2}, \tau_{k_1})$, and the quadratic Lipschitz bound applied to the second, $\phi_p(\mathbf{u}_{k_2}, \tau_{k_1}) - \phi_p(\mathbf{u}_{k_1}, \tau_{k_1})$.

It then follows that if there exist k_1 and k_2 such that at least one of the relations above does not hold, the Lipschitz constants are not consistent with the data and as such should be modified. One possible algorithm to carry out this operation is now provided.

Algorithm 1 – Lipschitz Consistency Check

1. Set $a := 1$.
2. Check the validity of (102) and (103) for every combination $(k_1, k_2) \in \{0, \dots, k\} \times \{0, \dots, k\}$. If these inequalities are satisfied for every (k_1, k_2) , then terminate. Otherwise, proceed to Step 3.
3. If $a \leq 5$, then for each $i = 1, \dots, n_u$ set

$$\begin{aligned}
& \underline{\kappa}_{\phi, i} := 2^{-\text{sign } \underline{\kappa}_{\phi, i}} \underline{\kappa}_{\phi, i}, \\
& \overline{\kappa}_{\phi, i} := 2^{\text{sign } \overline{\kappa}_{\phi, i}} \overline{\kappa}_{\phi, i}, \\
& \underline{\kappa}_{\phi, \tau} := 2^{-\text{sign } \underline{\kappa}_{\phi, \tau}} \underline{\kappa}_{\phi, \tau}, \\
& \overline{\kappa}_{\phi, \tau} := 2^{\text{sign } \overline{\kappa}_{\phi, \tau}} \overline{\kappa}_{\phi, \tau}.
\end{aligned}$$

If $5 < a \leq 10$, set

$$\begin{aligned}
& \underline{\kappa}_{\phi, i} := -2 \max [|\underline{\kappa}_{\phi, i}|, |\overline{\kappa}_{\phi, i}|], \\
& \overline{\kappa}_{\phi, i} := 2 \max [|\underline{\kappa}_{\phi, i}|, |\overline{\kappa}_{\phi, i}|], \\
& \underline{\kappa}_{\phi, \tau} := -2 \max [|\underline{\kappa}_{\phi, \tau}|, |\overline{\kappa}_{\phi, \tau}|], \\
& \overline{\kappa}_{\phi, \tau} := 2 \max [|\underline{\kappa}_{\phi, \tau}|, |\overline{\kappa}_{\phi, \tau}|].
\end{aligned}$$

If $10 < a$, set

$$\begin{aligned}
& \underline{\kappa}_{\phi, i} := 2^{a-10} \underline{\kappa}_{\phi, i}, \\
& \overline{\kappa}_{\phi, i} := 2^{a-10} \overline{\kappa}_{\phi, i}, \\
& \underline{\kappa}_{\phi, \tau} := 2^{a-10} \underline{\kappa}_{\phi, \tau}, \\
& \overline{\kappa}_{\phi, \tau} := 2^{a-10} \overline{\kappa}_{\phi, \tau}.
\end{aligned}$$

4. Set $a := a + 1$ and return to Step 2.

Identical algorithms may be used to verify (100)-(101) for κ_{p, j_i} (for each $j = 1, \dots, n_{g_p}$) and to verify (104)-(105) for $M_{\phi, i_1 i_2}$, with the latter using the values of $\underline{\kappa}_{\phi, \tau}$ and $\overline{\kappa}_{\phi, \tau}$ that have been fixed by Algorithm 1.

Clearly, the essence of the algorithm is to make the Lipschitz constant estimates more conservative until the corresponding Lipschitz bounds are consistent with the data. In the version proposed above, this is done in three phases. The first ($a \leq 5$) allows all constants to retain their signs while making their values lower or higher, as appropriate. As mentioned earlier, information regarding the sign of a derivative can be very useful and so it is for this reason that we attempt to retain it here. If this is not sufficient due to an error in the sign for some constant, the second phase ($5 < a \leq 10$) makes the lower and upper bounds equal in magnitude as this magnitude continues to be increased. If after 10 augmentations the new constants are still not consistent, which would very likely be due to scaling errors – the derivative assumed to vary on a magnitude smaller than the one on which it actually does – the final phase ($10 < a$) proceeds to augment the estimates at a much quicker (geometric) rate until they become consistent.

While this algorithm is not perfect and comes with several fundamental weaknesses, such as augmenting all the constants for a given function via a single factor of 2, it nevertheless provides a basic and useful tool for the pure knowledge-free case, until other methods of estimating the Lipschitz constants become possible. In the case where the available estimates are already reasonably close to valid, we would expect Algorithm 1 to only refine these estimates slightly without needing to proceed to the second phase.

3.6 Exploiting Locality

We have, up to now, only employed global definitions of the Lipschitz constants and the concavity/convexity properties of the experimental functions. However, as the SCFO only depend on these being valid on a given subregion of the experimental space, it follows that one can relax the

resulting bounds so that they only hold over the subregion in question.

Let us start by defining the *local temporal experimental space with respect to \bar{k} , $\mathcal{S}_\tau^{\bar{k}}$* , as

$$\mathcal{S}_\tau^{\bar{k}} = \text{bbox}(\mathbf{u}_{\bar{k}}, \mathbf{u}_{k^*}, \bar{\mathbf{u}}_{k+1}^*) \times \left\{ \tau : \begin{array}{l} \min[\tau_{\bar{k}}, \tau_{k^*}] \\ \leq \tau \leq \tau_{k+1} \end{array} \right\}, \quad (107)$$

i.e., as the Cartesian product of the (minimal) bounding box of the points $\mathbf{u}_{\bar{k}}$, \mathbf{u}_{k^*} , and $\bar{\mathbf{u}}_{k+1}^*$ with the shortest time interval that includes $\tau_{\bar{k}}$, τ_{k^*} , and τ_{k+1} . Clearly, $\mathcal{S}_\tau^{\bar{k}} \subseteq \mathcal{S}_\tau$. Furthermore, $(\mathbf{u}_{k+1}, \tau_{k+1}) \in \mathcal{S}_\tau^{\bar{k}}$ follows from the convexity of $\mathcal{S}_\tau^{\bar{k}}$ and the fact that $(\mathbf{u}_{k^*}, \tau_{k+1})$ and $(\bar{\mathbf{u}}_{k+1}^*, \tau_{k+1})$, the convex combination of which defines $(\mathbf{u}_{k+1}, \tau_{k+1})$, are both in $\mathcal{S}_\tau^{\bar{k}}$.

The motivation for defining $\mathcal{S}_\tau^{\bar{k}}$ this way stems from the fact that the upper and lower bounds on the evolution of an experimental function between any two points in $\mathcal{S}_\tau^{\bar{k}}$ only need to be written using the Lipschitz constants and convexity/concavity properties relevant to $\mathcal{S}_\tau^{\bar{k}}$. Since these bounds will generally involve the past experimental iterate $\mathbf{u}_{\bar{k}}$, the reference iterate \mathbf{u}_{k^*} , the target iterate $\bar{\mathbf{u}}_{k+1}^*$, and the actual future iterate \mathbf{u}_{k+1} , together with their relevant time values τ , the definition of $\mathcal{S}_\tau^{\bar{k}}$ essentially considers the smallest box space over which such bounds can be applied³.

The local Lipschitz constants with respect to \bar{k} are then defined as

$$\underline{\kappa}_{p,ji} \leq \underline{\kappa}_{p,ji}^{\bar{k}} < \left. \frac{\partial g_{p,j}}{\partial u_i} \right|_{(\mathbf{u},\tau)} < \bar{\kappa}_{p,ji}^{\bar{k}} \leq \bar{\kappa}_{p,ji}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{\bar{k}}, \quad (108)$$

$$\underline{M}_{\phi,i_1i_2} \leq \underline{M}_{\phi,i_1i_2}^{\bar{k}} < \left. \frac{\partial^2 \phi_p}{\partial u_{i_2} \partial u_{i_1}} \right|_{(\mathbf{u},\tau)} < \bar{M}_{\phi,i_1i_2}^{\bar{k}} \leq \bar{M}_{\phi,i_1i_2}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{\bar{k}}, \quad (109)$$

$$\underline{\kappa}_{p,j\tau} \leq \underline{\kappa}_{p,j\tau}^{\bar{k}} \leq \left. \frac{\partial g_{p,j}}{\partial \tau} \right|_{(\mathbf{u},\tau)} \leq \bar{\kappa}_{p,j\tau}^{\bar{k}} \leq \bar{\kappa}_{p,j\tau}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{\bar{k}}, \quad (110)$$

$$\underline{\kappa}_{\phi,i} \leq \underline{\kappa}_{\phi,i}^{\bar{k}} \leq \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u},\tau)} \leq \bar{\kappa}_{\phi,i}^{\bar{k}} \leq \bar{\kappa}_{\phi,i}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{\bar{k}}, \quad (111)$$

$$\underline{\kappa}_{\phi,\tau} \leq \underline{\kappa}_{\phi,\tau}^{\bar{k}} \leq \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u},\tau)} \leq \bar{\kappa}_{\phi,\tau}^{\bar{k}} \leq \bar{\kappa}_{\phi,\tau}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{\bar{k}}. \quad (112)$$

Extending this locality to the concavity/convexity assumptions, let us now state the local versions of Definitions 1 and 2.

³ One could be even more minimalist and use something other than a box (e.g., a convex hull), which in some cases would simplify down to a line [8]. However, to discuss such possibilities rigorously would require additional development that we believe would complicate the message without adding anything extremely useful from the implementation perspective.

Definition 3 (Local partial concavity) Let the decision variables \mathbf{u} be partitioned into subvectors \mathbf{v} and \mathbf{z} , so that $g_{p,j}(\mathbf{u}, \tau) = g_{p,j}(\mathbf{v}, \mathbf{z}, \tau)$. The function $g_{p,j}$ will be said to be *locally partially concave with respect to \bar{k} in \mathbf{z}* if it is concave everywhere on $\mathcal{S}_\tau^{\bar{k}}$ for \mathbf{v} and τ fixed. Likewise, $g_{p,j}$ will be said to be locally partially concave with respect to \bar{k} in both \mathbf{z} and τ if it is concave everywhere on $\mathcal{S}_\tau^{\bar{k}}$ for \mathbf{v} fixed.

Definition 4 (Local partial convexity) Let the decision variables \mathbf{u} be partitioned into subvectors \mathbf{v} and \mathbf{z} , so that $\phi_p(\mathbf{u}, \tau) = \phi_p(\mathbf{v}, \mathbf{z}, \tau)$. The function ϕ_p will be said to be *locally partially convex with respect to \bar{k} in \mathbf{z}* if it is convex everywhere on $\mathcal{S}_\tau^{\bar{k}}$ for \mathbf{v} and τ fixed. Likewise, ϕ_p will be said to be locally partially convex with respect to \bar{k} in both \mathbf{z} and τ if it is convex everywhere on $\mathcal{S}_\tau^{\bar{k}}$ for \mathbf{v} fixed.

Accordingly, let $I_{c,j}^{\bar{k}}$ denote the indices of the variables \mathbf{u} in which $g_{p,j}$ is locally partially concave with respect to \bar{k} , and let $\eta_{c,j}^{\bar{k}}$ denote a Boolean indicator that is equal to 1 if $g_{p,j}$ is locally partially concave with respect to \bar{k} in both the variables indexed by $I_{c,j}^{\bar{k}}$ and in τ , and is equal to 0 otherwise. Using analogous extensions for convexity, let $I_{v,\phi}^{\bar{k}}$ denote the indices of the variables \mathbf{u} in which ϕ_p is locally partially convex with respect to \bar{k} , and let $\eta_{v,\phi}^{\bar{k}}$ denote a Boolean indicator that is equal to 1 if ϕ_p is locally partially convex with respect to \bar{k} in both the variables indexed by $I_{v,\phi}^{\bar{k}}$ and in τ , and is equal to 0 otherwise. While the language here may appear complicated and verbose, we note that these definitions are just trivial localizations of what has already been presented in the previous sections.

In some cases – in particular, those related only to degradation – it will also be of interest to take a local space defined by only two iterations, $\bar{k}_1, \bar{k}_2 \in [0, k+1]$. Denoting this space by $\mathcal{S}_\tau^{\bar{k}_1, \bar{k}_2}$ and assuming, without loss of generality, that $\bar{k}_1 \leq \bar{k}_2$, we define it as

$$\mathcal{S}_\tau^{\bar{k}_1, \bar{k}_2} = \text{bbox}(\mathbf{u}_{\bar{k}_1}, \mathbf{u}_{\bar{k}_2}) \times \left\{ \tau : \tau_{\bar{k}_1} \leq \tau \leq \tau_{\bar{k}_2} \right\}. \quad (113)$$

As this will be particularly relevant with respect to degradation, we define the appropriate degradation-related Lipschitz constants as

$$\underline{\kappa}_{p,j\tau} \leq \underline{\kappa}_{p,j\tau}^{\bar{k}_1, \bar{k}_2} \leq \left. \frac{\partial g_{p,j}}{\partial \tau} \right|_{(\mathbf{u},\tau)} \leq \bar{\kappa}_{p,j\tau}^{\bar{k}_1, \bar{k}_2} \leq \bar{\kappa}_{p,j\tau}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{\bar{k}_1, \bar{k}_2}, \quad (114)$$

$$\underline{\kappa}_{\phi,\tau} \leq \underline{\kappa}_{\phi,\tau}^{\bar{k}_1, \bar{k}_2} \leq \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u},\tau)} \leq \bar{\kappa}_{\phi,\tau}^{\bar{k}_1, \bar{k}_2} \leq \bar{\kappa}_{\phi,\tau}, \quad \forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{\bar{k}_1, \bar{k}_2}. \quad (115)$$

We are now equipped with all of the necessary notation to state the local versions of the SCFO. Although one could go through and rederive all of the previous results with the

local constants and local assumptions, such a task would be rather space-consuming and redundant, as it would not introduce anything insightful or unexpected – put otherwise, we find it sufficient to let the reader accept on trust that all of the following bounds and inequalities, which are obtained by simply replacing global constants and assumptions by the local ones, will hold.

Considering first Condition (75), we state its local version:

$$\min_{\bar{k}=0,\dots,k} \left[\begin{array}{l} g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j}^{\bar{k}} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,j}^{\bar{k}}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \\ + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{array}{l} \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \leq 0. \quad (116)$$

With respect to the cost descent conditions, we have the following local version of (55):

$$\nabla \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \left[\begin{array}{l} \bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*} \\ 0 \end{array} \right] + \frac{K_k}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \left[\begin{array}{l} \underline{M}_{\phi,i_1 i_2}^{k^*} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ \quad (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}), \\ \bar{M}_{\phi,i_1 i_2}^{k^*} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ \quad (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}) \end{array} \right] \leq 0. \quad (117)$$

For the supplementary condition to help enforce cost decrease, given by (87), the local version is as follows:

$$\max_{\bar{k}=0,\dots,k} \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{v,\phi}^{\bar{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{v,\phi}^{\bar{k}}) \underline{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{v,\phi}^{\bar{k}}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \\ + \sum_{i \notin I_{v,\phi}^{\bar{k}}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \quad (118)$$

$$\leq \min_{\bar{k}=0,\dots,k} \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,\phi}^{\bar{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}^{\bar{k}}) \bar{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,\phi}^{\bar{k}}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \\ + \sum_{i \notin I_{c,\phi}^{\bar{k}}} \max \left[\begin{array}{l} \underline{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}) \end{array} \right] \end{array} \right].$$

For the choice of reference point, given in (91) and (72), note that one cannot make relaxations over $\mathcal{S}_{\tau}^{\bar{k}}$ since this space is defined, in part, by k^* , which here would be undetermined. However, we may instead work with $\mathcal{S}_{\tau}^{\bar{k}_1, \bar{k}_2}$, as defined in (113), to refine (91) and (72) as

$$\begin{aligned} k^* := & \arg \text{maximize } \bar{k} \\ & \bar{k} \in [0, k] \\ & \text{subject to } g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & \quad + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & \quad + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0, \\ & \quad \forall j = 1, \dots, n_{g_p} \\ & \quad \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k},k} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ & \quad + (1 - \eta_{v,\phi}^{\bar{k},k}) \underline{\kappa}_{\phi,\tau}^{\bar{k},k} (\tau_k - \tau_{\bar{k}}) \leq \\ & \quad \min_{\bar{k} \in \mathbf{k}_f} \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \\ \eta_{c,\phi}^{\bar{k},k} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}^{\bar{k},k}) \bar{\kappa}_{\phi,\tau}^{\bar{k},k} (\tau_k - \tau_{\bar{k}}) \end{array} \right], \end{aligned} \quad (119)$$

$$\mathbf{k}_f = \left\{ \bar{k} : \left[\begin{array}{l} g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0, \\ \forall j = 1, \dots, n_{g_p} \end{array} \right] \right\}, \quad (120)$$

$$\begin{aligned} k^* := & \arg \text{minimize } \max_{\bar{k} \in [0, k]} \max_{j=1,\dots,n_{g_p}} \left[\begin{array}{l} g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \end{array} \right]. \end{aligned} \quad (121)$$

At this point, it may be fair to ask if these refinements are actually of practical interest, as very specific knowledge is

not assumed to be available for the experimental functions, and both the local Lipschitz constants and the local structural assumptions are examples of specific knowledge as they characterize how the function behaves in different portions of the experimental space. At the time of writing, it is not clear if one could routinely and reliably obtain these local relaxations for most experimental optimization problems. However, it is possible to envision a scenario where the function is particularly steep in one region of the experimental space, where it has large Lipschitz constants, and somewhat flat in another, where its Lipschitz constants are small. Likewise, one could also imagine a function that is very nonlinear in some regions (i.e., has large M values) and close to linear otherwise (i.e., has M values close to 0). If a model of the function is available and can predict such trends approximately, then one may obtain the local constants for the model. Returning to the example given in (97) and (98), the appropriate local analogues would be

$$\begin{aligned}\underline{\kappa}_{\hat{\phi},i}^{\bar{k}} &= \min_{\substack{\mathbf{u}, \tau \in \mathcal{J}_{\tau}^{\bar{k}} \\ \theta \in \Theta}} \left. \frac{\partial \phi_{\hat{p}}}{\partial u_i} \right|_{(\mathbf{u}, \tau, \theta)}, \\ \bar{\kappa}_{\hat{\phi},i}^{\bar{k}} &= \max_{\substack{\mathbf{u}, \tau \in \mathcal{J}_{\tau}^{\bar{k}} \\ \theta \in \Theta}} \left. \frac{\partial \phi_{\hat{p}}}{\partial u_i} \right|_{(\mathbf{u}, \tau, \theta)},\end{aligned}\quad (122)$$

$$\begin{aligned}\underline{\kappa}_{\hat{\phi},\tau}^{\bar{k}} &= \min_{\substack{\mathbf{u}, \tau \in \mathcal{J}_{\tau}^{\bar{k}} \\ \theta \in \Theta}} \left. \frac{\partial \phi_{\hat{p}}}{\partial \tau} \right|_{(\mathbf{u}, \tau, \theta)}, \\ \bar{\kappa}_{\hat{\phi},\tau}^{\bar{k}} &= \max_{\substack{\mathbf{u}, \tau \in \mathcal{J}_{\tau}^{\bar{k}} \\ \theta \in \Theta}} \left. \frac{\partial \phi_{\hat{p}}}{\partial \tau} \right|_{(\mathbf{u}, \tau, \theta)},\end{aligned}\quad (123)$$

followed by the local analogue to (99)

$$\begin{aligned}\underline{\kappa}_{\hat{\phi},i}^{\bar{k}} &:= \underline{\kappa}_{\hat{\phi},i}^{\bar{k}} - 0.5(\bar{\kappa}_{\hat{\phi},i}^{\bar{k}} - \underline{\kappa}_{\hat{\phi},i}^{\bar{k}}), \\ \bar{\kappa}_{\hat{\phi},i}^{\bar{k}} &:= \bar{\kappa}_{\hat{\phi},i}^{\bar{k}} + 0.5(\bar{\kappa}_{\hat{\phi},i}^{\bar{k}} - \underline{\kappa}_{\hat{\phi},i}^{\bar{k}}), \\ \underline{\kappa}_{\hat{\phi},\tau}^{\bar{k}} &:= \underline{\kappa}_{\hat{\phi},\tau}^{\bar{k}} - 0.5(\bar{\kappa}_{\hat{\phi},\tau}^{\bar{k}} - \underline{\kappa}_{\hat{\phi},\tau}^{\bar{k}}), \\ \bar{\kappa}_{\hat{\phi},\tau}^{\bar{k}} &:= \bar{\kappa}_{\hat{\phi},\tau}^{\bar{k}} + 0.5(\bar{\kappa}_{\hat{\phi},\tau}^{\bar{k}} - \underline{\kappa}_{\hat{\phi},\tau}^{\bar{k}}).\end{aligned}\quad (124)$$

Such relaxations could be particularly useful when a certain portion of the experimental space possesses large Lipschitz constants – thus requiring small steps of the SCFO, and thereby slow progress – but is not actually the part of the experimental space that is explored during optimization.

To illustrate what a difference locality can make, we again consider the (–) case of the problem with degradation (42). To ease comparison, we remove the concavity assumptions on the constraints and set $I_{c,1}^{\bar{k}} := I_{c,2}^{\bar{k}} := \emptyset$, since not doing this would lead to very fast progress even without the locality relaxations. The Lipschitz

constants are chosen correctly, as in (56), and the same convexity assumptions as before are made on the cost, with $I_{v,\phi}^{\bar{k}} := \{1, 2\}$ for any \bar{k} . $\eta_{c,1}^{\bar{k}} := 0$, $\eta_{c,2}^{\bar{k}} := 1$, and $\eta_{v,\phi}^{\bar{k}} := 1$ are used for all \bar{k} . The upper Lipschitz constant on the cost with respect to degradation is taken as $\bar{\kappa}_{\phi,\tau} = 2/500$. The results for this scenario are given in Fig. 13, and are essentially comparable to what was already obtained in Fig. 5, the only difference between the two being certain concavity/convexity assumptions with respect to degradation, which contribute little in this case.

Consider now applying local relaxations to the relevant Lipschitz constants for the two experimental constraints, for which we will, for simplicity, minimize and maximize the derivatives of the true functions over the relevant local spaces (as mentioned, in practice one would likely perform such minimizations/maximizations for a model):

$$\underline{\kappa}_{p,11}^{\bar{k}} := \arg \underset{u_1, u_2, \tau \in \mathcal{J}_{\tau}^{\bar{k}}}{\text{minimize}} \quad -12u_1 - 3.5 - \frac{\tau}{500} - 10^{-3}, \quad (125)$$

$$\bar{\kappa}_{p,11}^{\bar{k}} := \arg \underset{u_1, u_2, \tau \in \mathcal{J}_{\tau}^{\bar{k}}}{\text{maximize}} \quad -12u_1 - 3.5 - \frac{\tau}{500} + 10^{-3}, \quad (126)$$

$$\underline{\kappa}_{p,21}^{\bar{k}} := \arg \underset{u_1, u_2, \tau \in \mathcal{J}_{\tau}^{\bar{k}}}{\text{minimize}} \quad 4u_1 + 0.5 - 10^{-3}, \quad (127)$$

$$\bar{\kappa}_{p,21}^{\bar{k}} := \arg \underset{u_1, u_2, \tau \in \mathcal{J}_{\tau}^{\bar{k}}}{\text{maximize}} \quad 4u_1 + 0.5 + 10^{-3}, \quad (128)$$

with 10^{-3} added or subtracted to enforce strict inequality. While other relaxations are also possible, we do not make them here as their effect on performance is likely to be negligible.

The results for the implementation using these local versions are given in Fig. 14, with the cost values corresponding to the two versions given in Fig. 15. One notes a drastic improvement in convergence speed.

4 Lower and Upper Bounds on the Cost and Constraint Values

In most experimental settings, what one observes after applying a given $\mathbf{u}_{\bar{k}}$ at time $\tau_{\bar{k}}$ are not the exact experimental function values $\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ but rather some noise-corrupted or estimated $\hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$. The error in the obtained measurement or estimate is commonly expressed with the aid of the additive noise model (see, e.g., [18, 28, 24, 27, 22]):

$$\hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) = \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + w_{\phi,\bar{k}}, \quad (129)$$

$$\hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) = g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + w_{j,\bar{k}}, \quad (130)$$

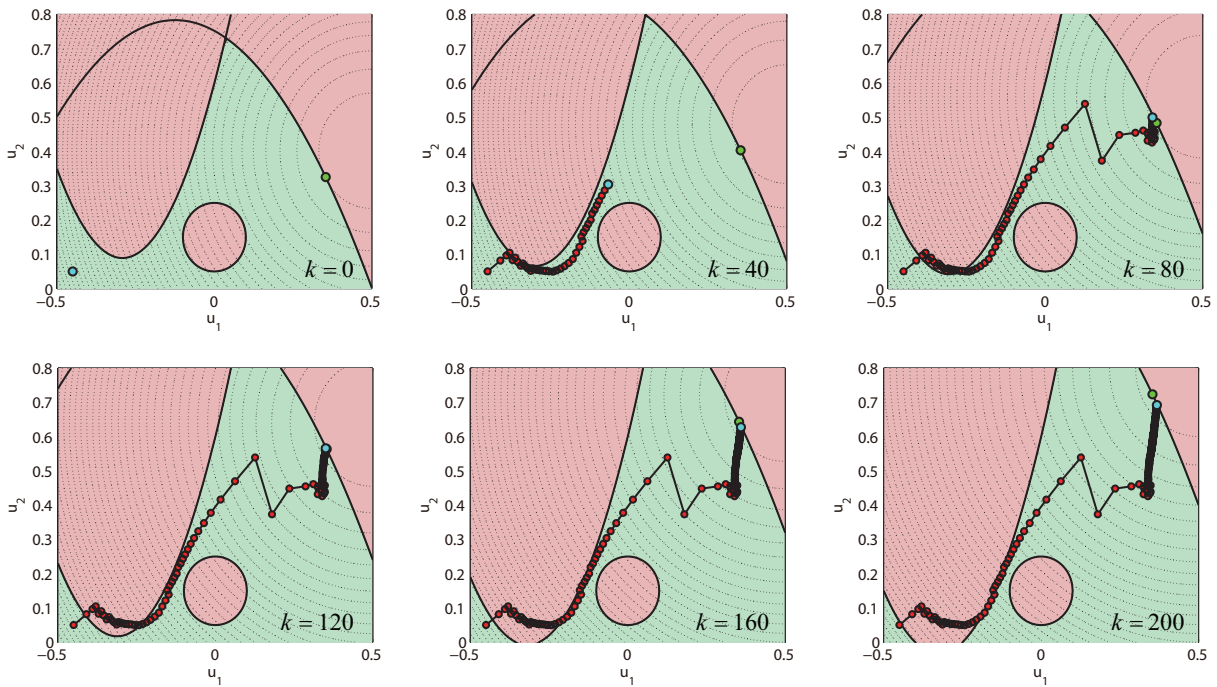


Fig. 13 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (-) scenario without local relaxations included.

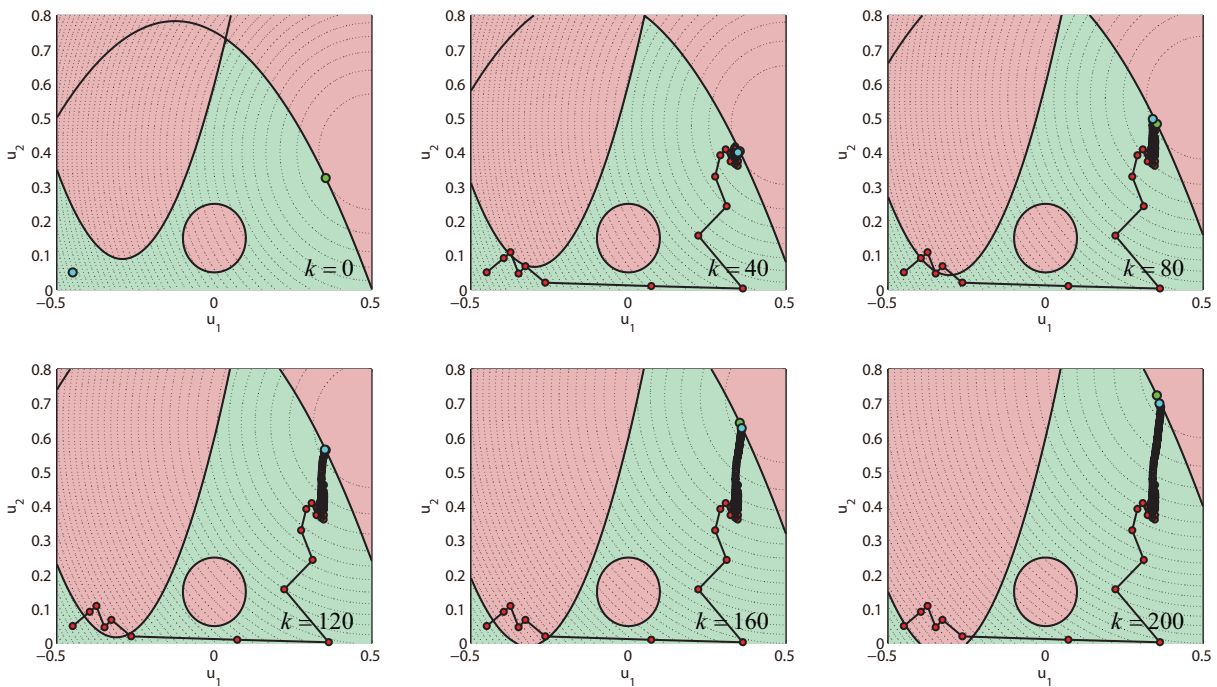


Fig. 14 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (-) scenario with local relaxations of certain Lipschitz constants included.

where the noise components, w_ϕ and w_j , will be assumed to be independent.

In the general case where w_ϕ and w_j may change sign, it should be clear that applying the SCFO with the estimated values $\hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ does not lead to a robust implementation, as this can falsely under or overpredict the true function values. To make these conditions robust against noise, it is then necessary to somehow derive lower and upper bounds on the true values of $\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ with respect to the observed $\hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$.

We will now discuss three possible ways to obtain the needed robust bounds, using the cost function ϕ_p as the example throughout.

4.1 Bounding the True Function Value by Bounding the Noise

The natural requirement for deriving lower and upper bounds on $\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ is that the stochastic terms w_ϕ be bounded with a certain probability. Noise bounds that are expected to hold with, for example, a probability of 99% would need to satisfy

$$\mathbb{P}(w_{\phi,\bar{k}} < w_{\phi,\bar{k}} < \bar{w}_{\phi,\bar{k}}) = 0.99. \quad (131)$$

To obtain the bounds $w_{\phi,\bar{k}}$ and $\bar{w}_{\phi,\bar{k}}$, one should at least know the mean ($w_{ave,\bar{k}}$) and variance ($\sigma_{\bar{k}}^2$) of w_ϕ . If the probability density function (pdf) of $w_{\phi,\bar{k}}$ is not known, the most ‘‘brute’’ way to determine a sufficiently low (resp., high) w_ϕ (resp., \bar{w}_ϕ) is to use Chebyshev’s inequality [15, 27]:

$$\mathbb{P}(w_{\phi,\bar{k}} < w_{\phi,\bar{k}} < \bar{w}_{\phi,\bar{k}}) \geq \frac{4[(w_{ave,\bar{k}} - w_{\phi,\bar{k}})(\bar{w}_{\phi,\bar{k}} - w_{ave,\bar{k}}) - \sigma_{\bar{k}}^2]}{(\bar{w}_{\phi,\bar{k}} - w_{\phi,\bar{k}})^2} = 0.99, \quad (132)$$

where one may solve the algebraic equation so as to obtain the appropriate $w_{\phi,\bar{k}}$ and $\bar{w}_{\phi,\bar{k}}$.

Given the generality of Chebyshev’s inequality, such bounds will tend to be fairly conservative, and tighter bounds can naturally be obtained if the pdf is known. Taking the well-known example of zero-mean Gaussian noise, one may simply set the bounds as

$$w_{\phi,\bar{k}} := -3\sigma_{\bar{k}}, \quad \bar{w}_{\phi,\bar{k}} = 3\sigma_{\bar{k}}, \quad (133)$$

which yield a probability slightly higher than 99% in (131). For the general pdf, one may solve (131) numerically to obtain an appropriate pair of $w_{\phi,\bar{k}}$ and $\bar{w}_{\phi,\bar{k}}$.

With these bounds available, we may then rearrange (129) and apply the bounds to obtain

$$\hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \bar{w}_{\phi,\bar{k}} < \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) < \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - w_{\phi,\bar{k}}, \quad (134)$$

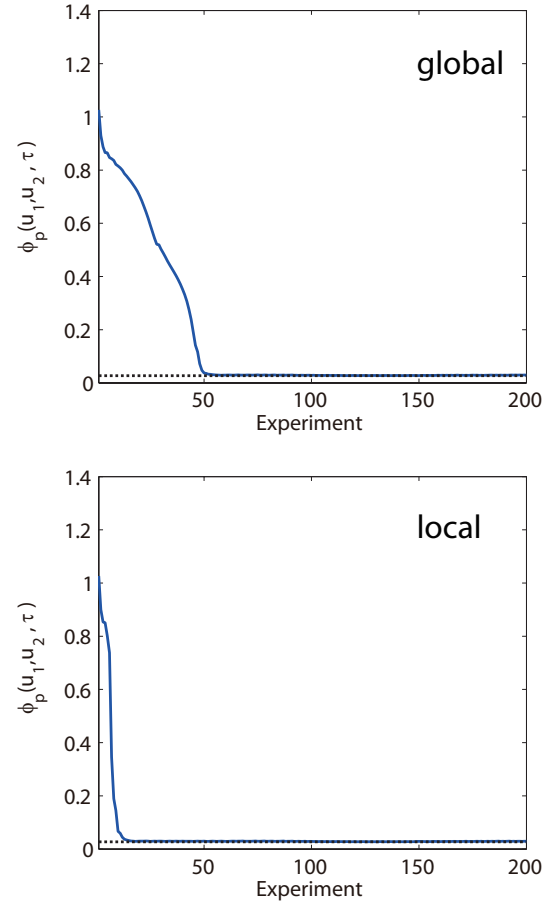


Fig. 15 Cost function values obtained by the modified SCFO methodology for Problem (42) for the (–) scenario when global Lipschitz constants (top) and local Lipschitz constants (bottom) are used.

which effectively bounds the true value to lie in the interval $(\hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \bar{w}_{\phi,\bar{k}}, \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - w_{\phi,\bar{k}})$ with some desired probability.

4.2 Reduced-Variance Bounds in the Case of Repeated Measurements

One may attempt to tighten the bounds of (134) in the case when the same variables $\mathbf{u}_{\bar{k}}$ have been applied for multiple iterations – this may occur, for example, if one chooses to temporarily stop experimental optimization or, in cases where the adaptation must be fast, one is forced to ‘‘skip’’ an iteration to allow for the algorithm to have enough time to carry out its computations (for an example of the latter, we refer the reader to Case Study 4.2 in [6]). Let us denote by $\bar{\mathbf{k}}$ a set of iterations, with cardinality N , where the decision variables are equal to $\mathbf{u}_{\bar{k}}$, i.e.:

$$\bar{\mathbf{k}} \subseteq \{\tilde{k} : \mathbf{u}_{\tilde{k}} = \mathbf{u}_{\bar{k}}\}, \quad (135)$$

where the cardinality of the right-hand side will be denoted by \bar{N} , with $N \leq \bar{N}$.

The corresponding N measurement identities (129) then become

$$\hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) = \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + w_{\phi, \bar{k}}, \quad \forall \bar{k} \in \bar{\mathbf{k}}. \quad (136)$$

To proceed further, let us recall the bounds (89)-(90) and state their more general (local) versions for the two iterations $\bar{k}, \tilde{k} \in [0, k]$ over the space $\mathcal{S}_{\tau}^{\bar{k}, \tilde{k}}$ as defined in (113):

$$\begin{aligned} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &\leq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ &\quad + \eta_{c, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &\quad + (1 - \eta_{c, \phi}^{\bar{k}, \tilde{k}}) \max \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right], \end{aligned} \quad (137)$$

$$\begin{aligned} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &\geq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ &\quad + \eta_{v, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &\quad + (1 - \eta_{v, \phi}^{\bar{k}, \tilde{k}}) \min \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right]. \end{aligned} \quad (138)$$

Applying these bounds to (136) then yields

$$\begin{aligned} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &+ \eta_{v, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &+ (1 - \eta_{v, \phi}^{\bar{k}, \tilde{k}}) \min \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right] + w_{\phi, \bar{k}} \\ &\leq \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \leq \\ &\quad \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &\quad + (1 - \eta_{c, \phi}^{\bar{k}, \tilde{k}}) \max \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right] + w_{\phi, \bar{k}}, \end{aligned} \quad (139)$$

which, upon rearrangement with respect to $\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, yields

$$\begin{aligned} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &- \eta_{c, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &- (1 - \eta_{c, \phi}^{\bar{k}, \tilde{k}}) \max \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right] - w_{\phi, \bar{k}} \\ &\leq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \leq \\ &\quad \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \eta_{v, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &\quad - (1 - \eta_{v, \phi}^{\bar{k}, \tilde{k}}) \min \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right] - w_{\phi, \bar{k}}. \end{aligned} \quad (140)$$

Noting that (140) represents N pairs of inequalities, we may sum them up to obtain

$$\begin{aligned} \sum_{\bar{k} \in \bar{\mathbf{k}}} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &- \sum_{\bar{k} \in \bar{\mathbf{k}}} \eta_{c, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &- \sum_{\bar{k} \in \bar{\mathbf{k}}} (1 - \eta_{c, \phi}^{\bar{k}, \tilde{k}}) \max \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right] - \sum_{\bar{k} \in \bar{\mathbf{k}}} w_{\phi, \bar{k}} \\ &\leq N \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \leq \\ &\sum_{\bar{k} \in \bar{\mathbf{k}}} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \sum_{\bar{k} \in \bar{\mathbf{k}}} \eta_{v, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &- \sum_{\bar{k} \in \bar{\mathbf{k}}} (1 - \eta_{v, \phi}^{\bar{k}, \tilde{k}}) \min \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right] - \sum_{\bar{k} \in \bar{\mathbf{k}}} w_{\phi, \bar{k}}, \end{aligned} \quad (141)$$

which then yields

$$\begin{aligned} \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &- \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \eta_{c, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &- \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} (1 - \eta_{c, \phi}^{\bar{k}, \tilde{k}}) \max \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right] \\ &- \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} w_{\phi, \bar{k}} \\ &\leq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \leq \\ &\quad \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ &\quad - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \eta_{v, \phi}^{\bar{k}, \tilde{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\tilde{k}}) \\ &\quad - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} (1 - \eta_{v, \phi}^{\bar{k}, \tilde{k}}) \min \left[\frac{\underline{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}{\bar{\mathbf{K}}_{\phi, \tau}^{\bar{k}, \tilde{k}} (\tau_{\bar{k}} - \tau_{\tilde{k}})}, \right] \\ &\quad - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} w_{\phi, \bar{k}}. \end{aligned} \quad (142)$$

In a manner analogous to (131), we may define bounds, $\underline{W}_{\phi, \bar{k}}$ and $\overline{W}_{\phi, \bar{k}}$, satisfying the condition

$$\mathbb{P} \left(\underline{W}_{\phi, \bar{k}} < \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} w_{\phi, \bar{k}} < \overline{W}_{\phi, \bar{k}} \right) = 0.99, \quad (143)$$

which finally allows us to state:

$$\begin{aligned}
& \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \eta_{c,\phi}^{\bar{k},\bar{k}} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \\
& - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} (1 - \eta_{c,\phi}^{\bar{k},\bar{k}}) \max \left[\frac{\underline{\mathbf{K}}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}})}{\bar{\mathbf{K}}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}})}, \right] - \bar{W}_{\phi,\bar{k}} \\
& \leq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \leq \\
& \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \eta_{v,\phi}^{\bar{k},\bar{k}} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \\
& - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} (1 - \eta_{v,\phi}^{\bar{k},\bar{k}}) \min \left[\frac{\underline{\mathbf{K}}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}})}{\bar{\mathbf{K}}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}})}, \right] - \underline{W}_{\phi,\bar{k}}.
\end{aligned} \tag{144}$$

As before, $\underline{W}_{\phi,\bar{k}}$ and $\bar{W}_{\phi,\bar{k}}$ may be obtained in different ways depending on what is known about the noise. For the case when only the mean and variance are known, one may turn to generalized multivariate versions of Chebyshev's inequality [1]. For the known general pdf, Monte Carlo methods may be used to compute these bounds by numerically generating the pdf of the summation. For the oft-assumed case of white Gaussian noise, we have the well known variance reduction of σ^2/N and the high-probability bounds of $\underline{W}_{\phi,\bar{k}} := -3\sigma/\sqrt{N}$ and $\bar{W}_{\phi,\bar{k}} := 3\sigma/\sqrt{N}$.

For the particular case when $\bar{N} = 1$ (and $\bar{\mathbf{k}} = \bar{k}$), we see that (144) simply yields (134). Otherwise, different bounds are obtained. The motivation for using (144) comes from its ability to filter out the variance of the noise as $N \rightarrow \infty$, although it is possible that this offers no advantage in the presence of significant degradation, and it is not difficult to show that the lower and upper bounds of (144) may actually *diverge* as $N \rightarrow \infty$ when degradation effects are present due to the corresponding summations being potentially unbounded. When degradation is negligible, one does expect the lower and upper bounds to converge to the true value, however. The white Gaussian noise case is a simple illustration as both $\underline{W}_{\phi,\bar{k}}$ and $\bar{W}_{\phi,\bar{k}}$ would tend to 0 as $N \rightarrow \infty$ and the mean of the measured values would tend to the true value.

It is also interesting to note that one may essentially obtain $\sum_{N=1}^{\bar{N}} \binom{\bar{N}}{N}$ different bounds by playing with the definition of $\bar{\mathbf{k}}$ and considering all possible sets of the repeated measurements, and while taking the largest set possible – with $N := \bar{N}$ – may seem like the intuitive best option, this need not be the case. As such, one may enumerate all the possible $\bar{\mathbf{k}}$ and take the tightest bounds that result.

4.3 Refinement Using Lipschitz Bounds

Let us now denote the best lower and upper bounds obtained by (144) – considering all possible $\bar{\mathbf{k}}$, as these also include (134) – as

$$\underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) < \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) < \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \tag{145}$$

where $\underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ is simply the maximum of the left-hand sides of (144) for different $\bar{\mathbf{k}}$, while $\bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ is the minimum of the right-hand sides.

At this point, we may attempt to tighten these bounds further by coupling them with the local concavity/convexity-refined Lipschitz bounds that are valid for a given two iterations \bar{k} and \tilde{k} :

$$\begin{aligned}
\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) & \leq \phi_p(\mathbf{u}_{\tilde{k}}, \tau_{\tilde{k}}) \\
& + \eta_{c,\phi}^{\bar{k},\tilde{k}} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\tilde{k}}, \tau_{\tilde{k}})} (\tau_{\tilde{k}} - \tau_{\bar{k}}) \\
& + (1 - \eta_{c,\phi}^{\bar{k},\tilde{k}}) \max \left[\frac{\underline{\mathbf{K}}_{\phi,\tau}^{\bar{k},\tilde{k}} (\tau_{\tilde{k}} - \tau_{\bar{k}})}{\bar{\mathbf{K}}_{\phi,\tau}^{\bar{k},\tilde{k}} (\tau_{\tilde{k}} - \tau_{\bar{k}})}, \right] \\
& + \sum_{i \in \bar{I}_{c,\phi}^{\bar{k},\tilde{k}}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\tilde{k}}, \tau_{\tilde{k}})} (u_{\tilde{k},i} - u_{\bar{k},i}) \\
& + \sum_{i \notin \bar{I}_{c,\phi}^{\bar{k},\tilde{k}}} \max \left[\frac{\underline{\mathbf{K}}_{\phi,i}^{\bar{k},\tilde{k}} (u_{\tilde{k},i} - u_{\bar{k},i})}{\bar{\mathbf{K}}_{\phi,i}^{\bar{k},\tilde{k}} (u_{\tilde{k},i} - u_{\bar{k},i})}, \right],
\end{aligned} \tag{146}$$

$$\begin{aligned}
\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) & \geq \phi_p(\mathbf{u}_{\tilde{k}}, \tau_{\tilde{k}}) \\
& + \eta_{v,\phi}^{\bar{k},\tilde{k}} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\tilde{k}}, \tau_{\tilde{k}})} (\tau_{\tilde{k}} - \tau_{\bar{k}}) \\
& + (1 - \eta_{v,\phi}^{\bar{k},\tilde{k}}) \min \left[\frac{\underline{\mathbf{K}}_{\phi,\tau}^{\bar{k},\tilde{k}} (\tau_{\tilde{k}} - \tau_{\bar{k}})}{\bar{\mathbf{K}}_{\phi,\tau}^{\bar{k},\tilde{k}} (\tau_{\tilde{k}} - \tau_{\bar{k}})}, \right] \\
& + \sum_{i \in \bar{I}_{v,\phi}^{\bar{k},\tilde{k}}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\tilde{k}}, \tau_{\tilde{k}})} (u_{\tilde{k},i} - u_{\bar{k},i}) \\
& + \sum_{i \notin \bar{I}_{v,\phi}^{\bar{k},\tilde{k}}} \min \left[\frac{\underline{\mathbf{K}}_{\phi,i}^{\bar{k},\tilde{k}} (u_{\tilde{k},i} - u_{\bar{k},i})}{\bar{\mathbf{K}}_{\phi,i}^{\bar{k},\tilde{k}} (u_{\tilde{k},i} - u_{\bar{k},i})}, \right].
\end{aligned} \tag{147}$$

From (145), it should be clear that

$$\begin{aligned}
\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) & < \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
& + \eta_{c,\phi}^{\bar{k},\tilde{k}} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\tilde{k}}, \tau_{\tilde{k}})} (\tau_{\tilde{k}} - \tau_{\bar{k}}) \\
& + (1 - \eta_{c,\phi}^{\bar{k},\tilde{k}}) \max \left[\frac{\underline{\mathbf{K}}_{\phi,\tau}^{\bar{k},\tilde{k}} (\tau_{\tilde{k}} - \tau_{\bar{k}})}{\bar{\mathbf{K}}_{\phi,\tau}^{\bar{k},\tilde{k}} (\tau_{\tilde{k}} - \tau_{\bar{k}})}, \right] \\
& + \sum_{i \in \bar{I}_{c,\phi}^{\bar{k},\tilde{k}}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\tilde{k}}, \tau_{\tilde{k}})} (u_{\tilde{k},i} - u_{\bar{k},i}) \\
& + \sum_{i \notin \bar{I}_{c,\phi}^{\bar{k},\tilde{k}}} \max \left[\frac{\underline{\mathbf{K}}_{\phi,i}^{\bar{k},\tilde{k}} (u_{\tilde{k},i} - u_{\bar{k},i})}{\bar{\mathbf{K}}_{\phi,i}^{\bar{k},\tilde{k}} (u_{\tilde{k},i} - u_{\bar{k},i})}, \right],
\end{aligned} \tag{148}$$

$$\begin{aligned}
\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &> \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
&+ \eta_{v,\phi}^{\bar{k},\bar{k}} \left. \frac{\partial \phi_p}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \\
&+ (1 - \eta_{v,\phi}^{\bar{k},\bar{k}}) \min \left[\begin{array}{l} \underline{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ \bar{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{array} \right] \\
&+ \sum_{i \in I_{v,\phi}^{\bar{k},\bar{k}}} \left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}) \\
&+ \sum_{i \notin I_{v,\phi}^{\bar{k},\bar{k}}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}) \end{array} \right]
\end{aligned} \tag{149}$$

hold as well, which yields an additional set of k pairs of lower/upper bounds that are themselves a function of the lower/upper bounds obtained for the other k measurements.

As illustrated in Fig. 16, these bounds have the potential to carry a very nice ‘‘chain effect’’, in that a tight bound for one measurement can be effectively propagated to nearby measurements for which the initial bounds may not be as tight. This depends, of course, on using valid Lipschitz constants, although we note that one is not obliged to use the same constants here as for the SCFO in general, and more conservative values may be chosen specifically for this refinement without having the same detrimental effects on performance as they might if this were done for the SCFO. Additionally, as refining certain bounds can lead to better refinement in others, this bounding procedure should be iterated until no more improvement is obtained.

4.4 An Algorithm to Compute Lower and Upper Bounds for Experimental Function Values

To consolidate the results presented so far, we now provide an algorithm that uses all of the techniques presented to obtain the best possible lower and upper bounds on the true values $\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ for all $\bar{k} \in [0, k]$.

Algorithm 2 – Lower and Upper Bounds for Experimental Function Values

1. Set $\bar{k} := 0$. Choose $\Delta_r, \underline{\Delta}_r \in \mathbb{R}_{++}$ such that $\Delta_r > \underline{\Delta}_r$.
2. If $\bar{k} > k$, go to Step 6. Otherwise, proceed to Step 3.
3. Let \bar{N} denote the number of iterates with the decision variable values $\mathbf{u}_{\bar{k}}$. Set $\underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := -\infty$ and $\bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := \infty$.
4. For $N = 1, \dots, \bar{N}$:
 - (a) Generate the $\binom{\bar{N}}{N}$ index sets that correspond to different combinations of iterations with the same decision variable values. For each set:

- i. Define $\bar{\mathbf{k}}$ as the corresponding index set and compute the lower and upper bounds as given by (144). Set these bounds as the candidate values $\underline{\phi}_{p,test}$ and $\bar{\phi}_{p,test}$.

- ii. If $\underline{\phi}_{p,test} > \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, set $\underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := \underline{\phi}_{p,test}$.
If $\bar{\phi}_{p,test} < \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, set $\bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := \bar{\phi}_{p,test}$.

5. Set $\bar{k} := \bar{k} + 1$ and return to Step 2.
6. If $\Delta_r \leq \underline{\Delta}_r$, terminate. Otherwise, set $\Delta_r := 0$, $\bar{k} := 0$ and go to Step 7.
7. If $\bar{k} > k$, return to Step 6. Otherwise, proceed to Step 8.
8. For $\bar{k} := \{0, \dots, k\} \setminus \bar{k}$:

- (a) Compute the lower and upper bounds as given by (148) and (149). Set these bounds as the candidate values $\underline{\phi}_{p,test}$ and $\bar{\phi}_{p,test}$.
- (b) Set

$$\Delta_r := \max \left[\begin{array}{l} \Delta_r, \underline{\phi}_{p,test} - \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \\ \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \bar{\phi}_{p,test} \end{array} \right]. \tag{150}$$

- (c) If $\underline{\phi}_{p,test} > \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, set $\underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := \underline{\phi}_{p,test}$. If $\bar{\phi}_{p,test} < \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, set $\bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := \bar{\phi}_{p,test}$.

9. Set $\bar{k} := \bar{k} + 1$ and return to Step 7.

Steps 2-5 essentially generate the initial lower and upper bounds, while Steps 6-9 carry out the Lipschitz refinements. For each refinement cycle, Δ_r represents the best refinement obtained. When this best refinement becomes inferior to some (preferably small) tolerance $\underline{\Delta}_r$, the algorithm terminates and the resulting lower and upper bounds are outputted. In the case when \bar{N} is large, an ‘‘incomplete’’ version of the algorithm that only considers a limited number of subsets in the definition of $\bar{\mathbf{k}}$ may be used so as to avoid computational issues due to the combinatorial nature of $\bar{\mathbf{k}}$.

We finish this discussion on obtaining lower/upper bounds by noting that identical procedures may be followed to obtain the corresponding bounds for the constraints, $g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$.

4.5 Accounting for Measurement and Estimation Error in the SCFO

We now go through the modifications necessary to make the SCFO robust against measurement/estimation error.

4.5.1 Accounting for Error in the Feasibility Conditions

From $g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) < \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, it follows that enforcing

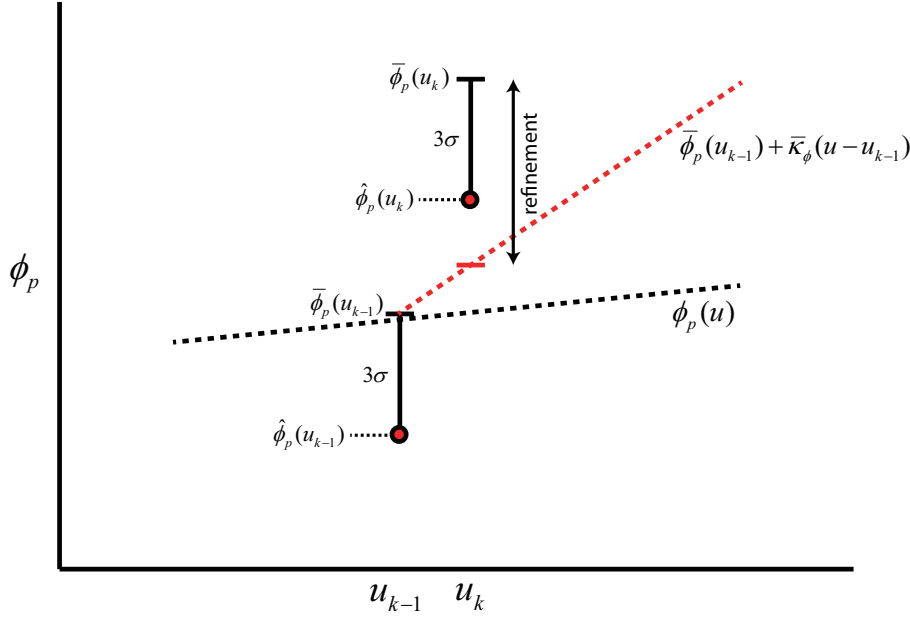


Fig. 16 An example of how a bound on the true experimental function value may be refined by exploiting the Lipschitz bound. A simple one-dimensional case with neither degradation effects nor convexity/concavity refinements is shown here, with a strong negative corruption due to noise in the measurement at u_{k-1} and a strong positive corruption due to noise at u_k . White zero-mean Gaussian noise is assumed, and only the upper bound is considered. If the bound (134) is used independently for both u_{k-1} and u_k , one sees that the upper bound for $\phi_p(u_k)$ ends up being very high. However, as the upper bound on $\phi_p(u_{k-1})$ is relatively tight, this tightness may be propagated to the next point over with the help of the Lipschitz bound, and a much better upper bound may be obtained.

$$\min_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \boldsymbol{\tau}_{\bar{k}}) \\ + \eta_{c,j}^{\bar{k}} \frac{\partial g_{p,j}}{\partial \boldsymbol{\tau}} \Big|_{(\mathbf{u}_{\bar{k}}, \boldsymbol{\tau}_{\bar{k}})} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{\bar{k}}) \\ + (1 - \eta_{c,j}^{\bar{k}}) \bar{K}_{p,j\boldsymbol{\tau}}^{\bar{k}} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{\bar{k}}) \\ + \sum_{i \in I_{c,j}^{\bar{k}}} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \boldsymbol{\tau}_{\bar{k}})} (u_{k^*,i} + \\ \quad K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \\ + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{array}{l} \underline{K}_{p,ji}^{\bar{k}} (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}, \\ \bar{K}_{p,ji}^{\bar{k}} (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i} \end{array} \right] \end{array} \right] \leq 0 \quad (151)$$

enforces (116). Naturally, since $g_{p,j}(\mathbf{u}_{\bar{k}}, \boldsymbol{\tau}_{\bar{k}}) < \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \boldsymbol{\tau}_{\bar{k}})$, the price one pays for this form is more conservative steps, as K_k will naturally need to be smaller to satisfy (151).

As before, some additional assumptions are needed on the past measurements for (151) to be enforceable. Namely, in the general case where the lower and upper Lipschitz constants may have opposite signs and the summation terms are always positive for any positive value of K_k , it is sufficient that

$$\bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) + \eta_{c,j}^{k^*} \frac{\partial g_{p,j}}{\partial \boldsymbol{\tau}} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\ + (1 - \eta_{c,j}^{k^*}) \bar{K}_{p,j\boldsymbol{\tau}}^{k^*} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \leq 0 \quad (152)$$

to guarantee that one can choose K_k in a manner so as to guarantee (151) and thus $g_{p,j}(\mathbf{u}_{k+1}, \boldsymbol{\tau}_{k+1}) \leq 0$ – in the worst case, one can always choose $K_k := 0$.

To avoid premature convergence, the following modified version of the projection (69) is also proposed:

$$\begin{aligned} \bar{\mathbf{u}}_{k+1}^* &:= \\ \arg \underset{\mathbf{u}}{\text{minimize}} & \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ \text{subject to} & \nabla g_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_{g_{p,j}} \\ & \bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) \\ \forall j : & + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \boldsymbol{\tau}} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\ & + (1 - \eta_{c,j}) \bar{K}_{p,j\boldsymbol{\tau}} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \geq -\varepsilon_{p,j} \\ & \nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j} \\ & \forall j : g_j(\mathbf{u}_{k^*}) \geq -\varepsilon_j \\ & \nabla \phi_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_\phi \\ & \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U, \end{aligned} \quad (153)$$

where we have simply substituted $\bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})$ for $g_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})$ in the Boolean condition. Alternatively, one could propose to use the estimate $\hat{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})$, but it is easy to show that this choice may lead to failure for certain noise distributions. For example, the distribution with a

mean smaller than $-\varepsilon_{p,j}$ and a very small variance would essentially lead to the projection never being applied with respect to the constraint $g_{p,j}$, which could in turn lead to premature convergence if this constraint were to become active.

4.5.2 Accounting for Error in the Necessary Conditions for Cost Decrease

To make the condition (118) implementable in the presence of errors in the measurement of $\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, we propose

$$\begin{aligned} & \max_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{v,\phi}^{\bar{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{v,\phi}^{\bar{k}}) \underline{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{v,\phi}^{\bar{k}}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \\ + \sum_{i \notin I_{v,\phi}^{\bar{k}}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + \\ \quad K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \\ & \leq \min_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,\phi}^{\bar{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}^{\bar{k}}) \bar{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,\phi}^{\bar{k}}} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \\ + \sum_{i \notin I_{c,\phi}^{\bar{k}}} \max \left[\begin{array}{l} \underline{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \end{aligned} \quad (154)$$

as its robust version, with the necessity of (118) implying the necessity of (154) as the left-hand side of the inequality is made lower and the right-hand side higher.

The result, as should be expected, is that less of the decision variable space can be ruled out in the robust version as larger values of K_k are permitted – i.e., less of the decision space can be robustly proven to have cost function values superior to the cost at the reference point.

4.5.3 Accounting for Error in the Choice of Reference Point

In choosing the reference point via the optimizations in (119) or (121), we propose the following robust versions:

$k^* :=$

arg maximize \bar{k}
 $\bar{k} \in [0, k]$

subject to $\bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$

$$\begin{aligned} & + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0 \\ & \forall j = 1, \dots, n_{g_p} \end{aligned} \quad (155)$$

$$\begin{aligned} & \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ & + (1 - \eta_{v,\phi}^{\bar{k}}) \underline{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_k - \tau_{\bar{k}}) \leq \end{aligned}$$

$$\min_{\bar{k} \in \mathbf{k}_f} \left[\begin{array}{l} \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \\ \eta_{c,\phi}^{\bar{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}^{\bar{k}}) \bar{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_k - \tau_{\bar{k}}) \end{array} \right],$$

$$\mathbf{k}_f = \left\{ \begin{array}{l} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0, \\ \forall j = 1, \dots, n_{g_p} \end{array} \right\}, \quad (156)$$

$k^* :=$

$$\text{arg minimize } \max_{\bar{k} \in [0, k]} \max_{j=1, \dots, n_{g_p}} \left[\begin{array}{l} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j} \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \end{array} \right] \quad (157)$$

where (155) chooses the most recent iterate that satisfies (152) and is robustly guaranteed to not have a cost value greater than the value at any of the other feasible iterates.

4.5.4 Modifying the Lipschitz Consistency Check

If Algorithm 1 is to be run in order to verify the consistency of the Lipschitz constants, then one must modify the bounds (100)-(105) to make them implementable in the presence of measurement or estimation error. The modifications we propose here are the following:

$$\begin{aligned} \underline{g}_{p,j}(\mathbf{u}_{k_2}, \tau_{k_2}) & \leq \bar{g}_{p,j}(\mathbf{u}_{k_1}, \tau_{k_1}) \\ & + \max \left[\begin{array}{l} \underline{\kappa}_{p,j\tau} (\tau_{k_2} - \tau_{k_1}) \\ \bar{\kappa}_{p,j\tau} (\tau_{k_2} - \tau_{k_1}) \end{array} \right] \\ & + \sum_{i=1}^{n_u} \max \left[\begin{array}{l} \underline{\kappa}_{p,ji} (u_{k_2,i} - u_{k_1,i}), \\ \bar{\kappa}_{p,ji} (u_{k_2,i} - u_{k_1,i}) \end{array} \right], \end{aligned} \quad (158)$$

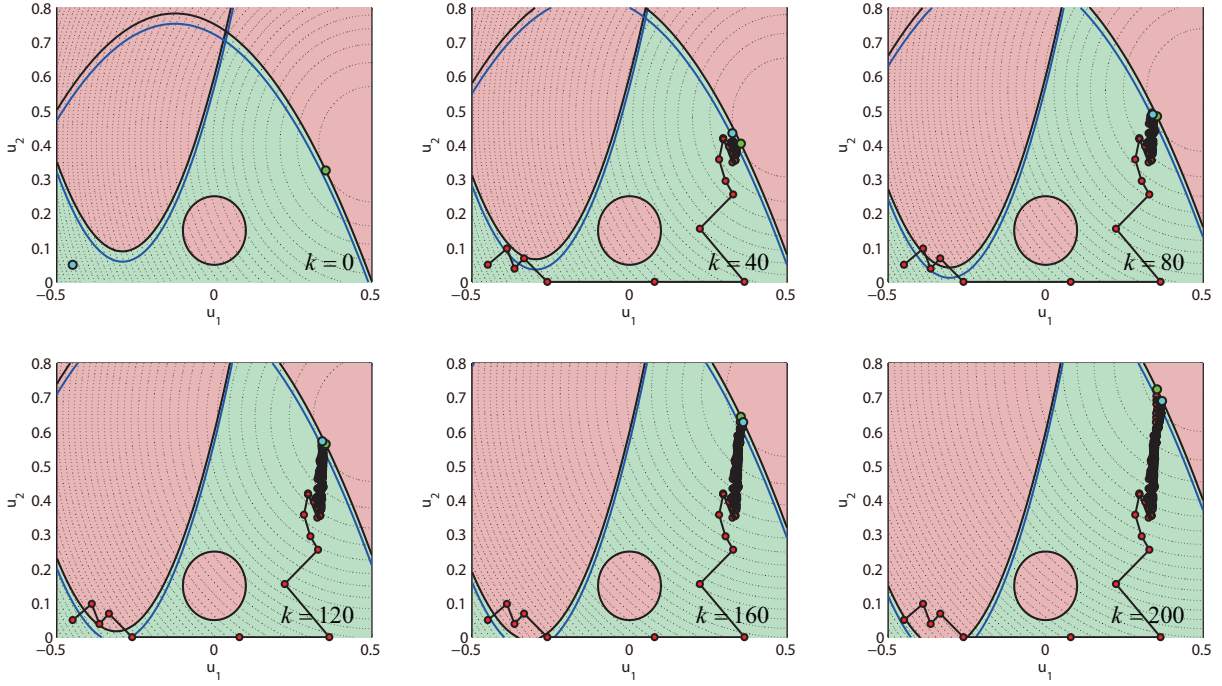


Fig. 17 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (–) scenario with measurement/estimation error in the experimental functions accounted for. The effect of the noise in the constraints is illustrated by showing the 3σ back-offs in blue.

$$\begin{aligned} \bar{g}_{p,j}(\mathbf{u}_{k_2}, \tau_{k_2}) &\geq \underline{g}_{p,j}(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &+ \min \left[\frac{\underline{\kappa}_{p,j\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{p,j\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &+ \sum_{i=1}^{n_u} \min \left[\frac{\underline{\kappa}_{p,ji}(u_{k_2,i} - u_{k_1,i})}{\bar{\kappa}_{p,ji}(u_{k_2,i} - u_{k_1,i})} \right], \end{aligned} \quad (159)$$

$$\begin{aligned} \underline{\phi}_p(\mathbf{u}_{k_2}, \tau_{k_2}) &\leq \bar{\phi}_p(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &+ \max \left[\frac{\underline{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &+ \sum_{i=1}^{n_u} \max \left[\frac{\underline{\kappa}_{\phi,i}(u_{k_2,i} - u_{k_1,i})}{\bar{\kappa}_{\phi,i}(u_{k_2,i} - u_{k_1,i})} \right], \end{aligned} \quad (160)$$

$$\begin{aligned} \bar{\phi}_p(\mathbf{u}_{k_2}, \tau_{k_2}) &\geq \underline{\phi}_p(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &+ \min \left[\frac{\underline{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &+ \sum_{i=1}^{n_u} \min \left[\frac{\underline{\kappa}_{\phi,i}(u_{k_2,i} - u_{k_1,i})}{\bar{\kappa}_{\phi,i}(u_{k_2,i} - u_{k_1,i})} \right], \end{aligned} \quad (161)$$

$$\begin{aligned} \underline{\phi}_p(\mathbf{u}_{k_2}, \tau_{k_2}) &\leq \\ &\bar{\phi}_p(\mathbf{u}_{k_1}, \tau_{k_1}) + \nabla \phi_p(\mathbf{u}_{k_1}, \tau_{k_1})^T \begin{bmatrix} \mathbf{u}_{k_2} - \mathbf{u}_{k_1} \\ 0 \end{bmatrix} \\ &+ \max \left[\frac{\underline{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &+ \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \left[\frac{\underline{M}_{\phi,i_1 i_2}(u_{k_2,i_1} - u_{k_1,i_1})}{\bar{M}_{\phi,i_1 i_2}(u_{k_2,i_1} - u_{k_1,i_1})} \right. \\ &\quad \left. \frac{(u_{k_2,i_2} - u_{k_1,i_2})}{(u_{k_2,i_2} - u_{k_1,i_2})} \right], \end{aligned} \quad (162)$$

$$\begin{aligned} \bar{\phi}_p(\mathbf{u}_{k_2}, \tau_{k_2}) &\geq \\ &\underline{\phi}_p(\mathbf{u}_{k_1}, \tau_{k_1}) + \nabla \phi_p(\mathbf{u}_{k_1}, \tau_{k_1})^T \begin{bmatrix} \mathbf{u}_{k_2} - \mathbf{u}_{k_1} \\ 0 \end{bmatrix} \\ &+ \min \left[\frac{\underline{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{\phi,\tau}(\tau_{k_2} - \tau_{k_1})} \right] \\ &+ \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \min \left[\frac{\underline{M}_{\phi,i_1 i_2}(u_{k_2,i_1} - u_{k_1,i_1})}{\bar{M}_{\phi,i_1 i_2}(u_{k_2,i_1} - u_{k_1,i_1})} \right. \\ &\quad \left. \frac{(u_{k_2,i_2} - u_{k_1,i_2})}{(u_{k_2,i_2} - u_{k_1,i_2})} \right], \end{aligned} \quad (163)$$

as inconsistency in (158)-(163) implies inconsistency in (100)-(105). The natural effect of the noise in the consistency check is that less conservative Lipschitz constants suffice for the robust version given by (158)-(163), as some of the change in function values across the decision variable space are “absorbed” by the noise and not attributed

to the deterministic changes in the function, the latter being linked to the Lipschitz constants.

It is, however, important to note that the two tasks of checking the consistency of the Lipschitz constants and computing lower and upper bounds become convoluted with these modifications, as the Lipschitz constants will affect the bounds (by Algorithm 2) and the bounds, in turn, may affect the Lipschitz constants (by Algorithm 1). To avoid computing improper bounds when a poor choice of Lipschitz constants is used, it is proposed that the bounds used in (158)-(163) are those computed by (134), as this bound is independent of the Lipschitz constants. Afterwards, Lipschitz constants that are consistent with (158)-(163) may be used in Algorithm 2.

4.6 Example

Considering again the (–) case of Problem (42), we corrupt the measurements of the cost and constraints with the additive noise elements

$$w_\phi, w_1, w_2 \sim \mathcal{N}(0, 10^{-4}), \quad (164)$$

and apply a modified version of Algorithm 2 where only the full sets (of cardinality \bar{N}) and the single-element sets (of cardinality 1) are considered in the definition of $\bar{\mathbf{k}}$ – this is done to avoid complexity issues. Using the computed bounds, we then use the modified SCFO implementation as given by (151), (153), (154), and (155). Convexity and concavity refinements with $I_{c,1} := \{1, 2\}$, $\eta_{c,1} := 0$, $I_{c,2} := \{2\}$, and $\eta_{c,2} := 1$ are used for the constraints, with the cost refined via $I_{v,\phi} := \{1, 2\}$ and $\eta_{v,\phi} := 1$. The same local refinements as given in (125)-(128) are used. The noise bounds $\underline{W}_{\phi,\bar{k}} := -3\sigma_{\bar{k}}/\sqrt{\bar{N}}$ and $\overline{W}_{\phi,\bar{k}} := 3\sigma_{\bar{k}}/\sqrt{\bar{N}}$ are employed, with $\underline{\Delta}_r := 10^{-6}$ in Algorithm 2.

We give the results in Figs. 17 and 18, from which we can see that the modified SCFO are able to obtain good performance without violating the constraints even in the presence of measurement/estimation error.

5 Gradient Estimation and Robust Projection

As an extension to the issues discussed in the previous section, let us now discuss the handling of errors in the *gradient estimates* $\nabla \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\nabla \hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, which is what one will have to work with in practice since the true gradients $\nabla \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\nabla g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ will never be available exactly⁴. An additional implementation difficulty, and one that we admittedly will not treat here in detail, is that

⁴ We will use the notations $\nabla \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\nabla \hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ to refer to the estimates of $\nabla \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\nabla g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, and *not* to the gradients of $\hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $\hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ as given in (129) and (130).

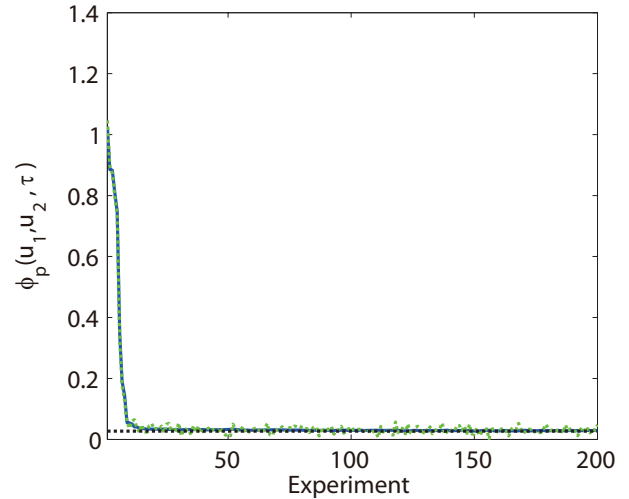


Fig. 18 Cost function values obtained by the modified SCFO methodology for Problem (42) for the (–) scenario with measurement/estimation error in the experimental functions accounted for. The effect of the noise in the cost is illustrated via the dotted green line, which represents the corrupted function values.

the additive noise model (129)-(130) for the function values does not extend to the gradient case in the measurement sense – i.e., one does not simply “measure” a gradient with additive noise. One may, however, consider obtaining a gradient estimate with some uncertainty bounds that are guaranteed to contain the true gradient. At the time of writing, methods for doing this are not very developed – we refer the reader to our previous work [5] for one possible approach, as well as a regularization-based algorithm that provides bounds on a gradient estimate and that is being employed in the latest version of the SCFO solver [7].

Throughout this discussion, we will assume to have at our disposal estimation algorithms $\Omega_{\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}}$, $\Omega_{\mathbf{u}_{\bar{k}}, \tau_{k+1}}$ that take all of the available measured/estimated data and return estimates of the gradient at $(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ and $(\mathbf{u}_{\bar{k}}, \tau_{k+1})$, together with the bounds on these estimates:

$$\begin{aligned} & \left[\nabla \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \nabla \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \nabla \overline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \right] \\ & = \Omega_{\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}} \left(\begin{array}{c} \mathbf{u}_0, \dots, \mathbf{u}_k, \tau_0, \dots, \tau_k, \\ \hat{\phi}_p(\mathbf{u}_0, \tau_0), \dots, \hat{\phi}_p(\mathbf{u}_k, \tau_k) \end{array} \right), \end{aligned} \quad (165)$$

$$\begin{aligned} & \left[\nabla \hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \nabla \underline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \nabla \overline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \right] \\ & = \Omega_{\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}} \left(\begin{array}{c} \mathbf{u}_0, \dots, \mathbf{u}_k, \tau_0, \dots, \tau_k, \\ \hat{g}_{p,j}(\mathbf{u}_0, \tau_0), \dots, \hat{g}_{p,j}(\mathbf{u}_k, \tau_k) \end{array} \right), \end{aligned} \quad (166)$$

$$\begin{aligned} & \left[\nabla \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{k+1}), \nabla \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{k+1}), \nabla \overline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{k+1}) \right] \\ & = \Omega_{\mathbf{u}_{\bar{k}}, \tau_{k+1}} \left(\begin{array}{c} \mathbf{u}_0, \dots, \mathbf{u}_k, \tau_0, \dots, \tau_{k+1}, \\ \hat{\phi}_p(\mathbf{u}_0, \tau_0), \dots, \hat{\phi}_p(\mathbf{u}_k, \tau_k) \end{array} \right), \end{aligned} \quad (167)$$

$$\begin{aligned} & \left[\nabla \hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{k+1}), \nabla \underline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{k+1}), \nabla \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{k+1}) \right] \\ & = \Omega_{\mathbf{u}_{\bar{k}}, \tau_{k+1}} \begin{pmatrix} \mathbf{u}_0, \dots, \mathbf{u}_k, \tau_0, \dots, \tau_{k+1}, \\ \hat{g}_{p,j}(\mathbf{u}_0, \tau_0), \dots, \hat{g}_{p,j}(\mathbf{u}_k, \tau_k) \end{pmatrix}. \end{aligned} \quad (168)$$

Given the current state of the art with regard to the gradient estimation problem, we will make no strong statistical assumptions on the estimates, apart from the general statement that $\nabla \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, $\nabla \hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, $\nabla \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{k+1})$, and $\nabla \hat{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{k+1})$ are, in some statistical sense, “best” estimates. We will also assume that

$$\nabla \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \in \left[\nabla \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \nabla \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \right], \quad (169)$$

$$\nabla g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \in \left[\nabla \underline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \nabla \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \right], \quad (170)$$

$$\nabla \phi_p(\mathbf{u}_{\bar{k}}, \tau_{k+1}) \in \left[\nabla \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{k+1}), \nabla \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{k+1}) \right], \quad (171)$$

$$\nabla g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{k+1}) \in \left[\nabla \underline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{k+1}), \nabla \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{k+1}) \right] \quad (172)$$

all hold with sufficiently high probability. Finally, we will assume the estimates and their bounds to be consistent with the Lipschitz constants, i.e., that

$$\begin{aligned} \underline{\kappa}_{p,ji} & < \frac{\partial \underline{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} \leq \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} < \bar{\kappa}_{p,ji}, \\ \underline{\kappa}_{p,j\tau} & \leq \frac{\partial \underline{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} \leq \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} \leq \bar{\kappa}_{p,j\tau}, \\ \underline{\kappa}_{\phi,i} & \leq \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} \leq \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} \leq \bar{\kappa}_{\phi,i}, \\ \underline{\kappa}_{\phi,\tau} & \leq \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} \leq \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} \leq \bar{\kappa}_{\phi,\tau}, \\ \underline{\kappa}_{p,ji} & < \frac{\partial \underline{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} \leq \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} < \bar{\kappa}_{p,ji}, \\ \underline{\kappa}_{p,j\tau} & \leq \frac{\partial \underline{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} \leq \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} \leq \bar{\kappa}_{p,j\tau}, \\ \underline{\kappa}_{\phi,i} & \leq \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} \leq \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} \leq \bar{\kappa}_{\phi,i}, \\ \underline{\kappa}_{\phi,\tau} & \leq \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} \leq \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} \leq \bar{\kappa}_{\phi,\tau}. \end{aligned} \quad (173)$$

Given only the estimates and bounds provided by (165)-(168), we are now interested in formulating a robust version of the SCFO that has, until now, only utilized the true (unavailable) gradients.

5.1 A Robust Version of the Projection

In proposing an implementable version of (153), one could, of course, simply replace the true experimental gradients by their estimated versions:

$$\begin{aligned} \bar{\mathbf{u}}_{k+1}^* & := \\ \arg \min_{\mathbf{u}} & \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ \text{subject to} & \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_{g_{p,j}}, \\ & \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ \forall j : & +\eta_{c,j} \frac{\partial \hat{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ & + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) \geq -\varepsilon_{p,j} \\ & \nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j}, \\ \forall j : & g_j(\mathbf{u}_{k^*}) \geq -\varepsilon_j \\ & \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_{\phi} \\ & \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U. \end{aligned} \quad (174)$$

While such an implementation may yield desired results in the long run – if, for example, the estimates are unbiased – there may be significant short-term losses that would result from using particularly bad estimates without accounting for their uncertainty or, in statistical terms, their *variance*, which is in some sense represented by the size of the box corresponding to the lower and upper bounds.

Here, we propose to use the standard robust approach of guaranteeing that the SCFO hold for the true gradients by ensuring that they hold for *all of the gradients in the uncertainty set*. In other words, we propose to replace (174) by the more robust

$$\begin{aligned} \bar{\mathbf{u}}_{k+1}^* & := \\ \arg \min_{\mathbf{u}} & \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ \text{subject to} & \nabla \tilde{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_{g_{p,j}}, \\ & \forall \nabla \tilde{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) \in \\ & \left[\nabla \underline{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}), \nabla \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) \right], \\ & \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ \forall j : & +\eta_{c,j} \frac{\partial \tilde{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ & + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) \geq -\varepsilon_{p,j} \\ & \nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j}, \\ \forall j : & g_j(\mathbf{u}_{k^*}) \geq -\varepsilon_j \\ & \nabla \tilde{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_{\phi}, \\ & \forall \nabla \tilde{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) \in \\ & \left[\nabla \underline{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}), \nabla \bar{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) \right] \\ & \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U, \end{aligned} \quad (175)$$

as any $\bar{\mathbf{u}}_{k+1}^*$ that satisfies the constraints of (175) is guaranteed to satisfy the constraints of (153) as well.

It should be clear that (175) is not numerically implementable as given since it has a semi-infinite number of constraints – i.e., the gradient uncertainty set, while bounded, contains an infinite number of gradients. However, due to the box nature of the uncertainty set, one may easily reformulate (175) by adding auxiliary slack variables s (contained in the vector \mathbf{s}_ϕ and matrix \mathbf{S}):

$$\begin{aligned}
& \bar{\mathbf{u}}_{k+1}^* := \\
& \arg \underset{\mathbf{u}, \mathbf{s}_\phi, \mathbf{S}}{\text{minimize}} \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\
& \text{subject to } \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_{p,j}} \\
& \quad \frac{\partial \underline{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji} \\
& \quad \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji}, \\
& \quad \forall i = 1, \dots, n_u, \\
& \quad \bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) \\
& \quad \forall j : +\eta_{c,j} \frac{\partial \bar{g}_{p,j}}{\partial \boldsymbol{\tau}} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\
& \quad \quad + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\boldsymbol{\tau}} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \geq -\varepsilon_{p,j} \\
& \quad \nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j}, \\
& \quad \forall j : g_j(\mathbf{u}_{k^*}) \geq -\varepsilon_j \\
& \quad \sum_{i=1}^{n_u} s_{\phi,i} \leq -\delta_\phi \\
& \quad \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\
& \quad \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\
& \quad \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U.
\end{aligned} \tag{176}$$

Clearly, (176) is a standard quadratic program with no more than $2n_{g_p}(1+n_u) + n_g + 4n_u + 2$ linear constraints. We prove the equivalence of (176) and (175) in the following lemma.

Lemma 3 (Equivalence of Projections (176) and (175))
Problems (176) and (175) are equivalent in the following sense:

$$\mathbf{u} \text{ feasible for (175)} \Leftrightarrow \exists \mathbf{s}_\phi, \mathbf{S} : \mathbf{u}, \mathbf{s}_\phi, \mathbf{S} \text{ feasible for (176)}. \tag{177}$$

Proof Let us first prove the forward implication

$$\mathbf{u} \text{ feasible for (175)} \Rightarrow \exists \mathbf{s}_\phi, \mathbf{S} : \mathbf{u}, \mathbf{s}_\phi, \mathbf{S} \text{ feasible for (176)}. \tag{178}$$

To do this, consider the inequality constraints – on the cost only, for brevity's sake – in summation form:

$$\begin{aligned}
& \sum_{i=1}^{n_u} \frac{\partial \tilde{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq -\delta_\phi, \\
& \forall \nabla \tilde{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \in \left[\nabla \underline{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}), \nabla \bar{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \right].
\end{aligned} \tag{179}$$

Since the terms inside the summation are linear in u_i , the following bound must hold:

$$\begin{aligned}
& \frac{\partial \tilde{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq \\
& \max \left[\begin{array}{l} \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \end{array} \right],
\end{aligned} \tag{180}$$

with the maximum term in (180) determined by the sign of $u_i - u_{k^*,i}$.

We may thus replace the semi-infinite family of constraints in (179) by their worst-case maximum:

$$\sum_{i=1}^{n_u} \max \left[\begin{array}{l} \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \end{array} \right] \leq -\delta_\phi. \tag{181}$$

By (180), it is clear that any \mathbf{u} that satisfies (181) satisfies (179) as well. Likewise, since the gradient realization of (181) belongs to the gradient uncertainty set in (179), it follows that any \mathbf{u} that satisfies (179) naturally satisfies (181). (179) and (181) are thus equivalent. The feasibility of \mathbf{u} for (175) then implies the feasibility of \mathbf{u} for (179) and, consequently, for (181).

Having established this, let us choose the slack variables \mathbf{s}_ϕ as

$$s_{\phi,i} := \max \left[\begin{array}{l} \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \end{array} \right]. \tag{182}$$

Substituting this choice into (181) clearly shows that the constraint

$$\sum_{i=1}^{n_u} s_{\phi,i} \leq -\delta_\phi \tag{183}$$

is satisfied. Furthermore, it is easily seen that this choice satisfies the constraints

$$\begin{aligned}
& \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\
& \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i}
\end{aligned} \tag{184}$$

as well.

An entirely analogous procedure for the experimental constraint functions and the analogous choice of slack variables \mathbf{S} ,

$$s_{ji} := \max \left[\begin{array}{l} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}), \\ \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \end{array} \right], \quad (185)$$

leads to the constraints

$$\sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_p,j}, \quad (186)$$

$$\begin{aligned} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) &\leq s_{ji} \\ \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) &\leq s_{ji} \end{aligned} \quad (187)$$

being satisfied.

As such, we have proven that for any \mathbf{u} that is feasible for (175) there exist $\mathbf{s}_\phi, \mathbf{S}$ – namely, those given by (182) and (185) – so that the extended set $\mathbf{u}, \mathbf{s}_\phi, \mathbf{S}$ is feasible for (176).

We will now prove the backward implication

$$\mathbf{u} \text{ feasible for (175)} \Leftrightarrow \exists \mathbf{s}_\phi, \mathbf{S} : \mathbf{u}, \mathbf{s}_\phi, \mathbf{S} \text{ feasible for (176)}. \quad (188)$$

In this case, the constraints (183) and (184) are known to hold, which clearly implies that (181) holds as well. By the equivalence of (179) and (181), it follows that semi-infinite set of constraints (179) is also satisfied. An identical analysis for the experimental constraint functions shows that the analogous semi-infinite sets of constraints for these functions hold as well by virtue of (186) and (187). Since these semi-infinite sets are the constraints of (175), we have thereby proven (188) and, consequently, (177). \square

The practical implication of Lemma 3 is that all \mathbf{u} that are feasible solutions for the tractable convex problem (176) are also feasible solutions for the semi-infinite and intractable problem (175). As such, no \mathbf{u} is lost by solving the reformulated (176) – additionally, if a feasible \mathbf{u} exists for (175), then we are guaranteed to find a feasible \mathbf{u} for (176) provided the availability of a reliable convex programming solver [3].

5.2 Infeasibility Due to Robustness

As discussed in [10], the infeasibility of the standard projection for a sufficiently low choice of projection parameters essentially implies that the algorithm has converged very close to a Fritz John (FJ) point.

Unfortunately, this is no longer the case for the robust projection where lower and upper bounds on the gradients are used instead. In fact, it is easily shown that the projection may be infeasible at any experimental iteration regardless of how far one is from an FJ point. For example, suppose that the gradient bounds are sufficiently loose and one has the following

$$\nabla \underline{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) = -\nabla \bar{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}). \quad (189)$$

It is readily seen that any \mathbf{u} that satisfies the constraint (179) for $\nabla \tilde{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) = \nabla \underline{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1})$ cannot possibly satisfy it for $\nabla \tilde{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) = \nabla \bar{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1})$ as well, as the two constraint values will have opposite signs. As both need to be satisfied in the robust implementation, the constraint set will thus be infeasible and, consequently, Projection (176) will admit no solution. Setting $\mathbf{u}_{k+1} := \mathbf{u}_{k^*}$ in this case may not be a good option since the gradient bounds may simply be bad due to lack of data and would not be an indicator of proximity to an FJ point.

The trade-off that comes up in such scenarios is the desire to continue to move and to adapt the decision variables so as to optimize the system while maintaining some notion of robustness. What is proposed here is to artificially tighten the gradient uncertainty bounds until the projection becomes feasible. Note, however, that, as infeasibility in the projection may be caused both by large projection parameters and conservative gradient bounds, a systematic way to reduce both in tandem is needed. The projection algorithm provided in [10] is thus modified accordingly.

Algorithm 3 – Projection with Automatic Choice of Projection Parameters and Gradient Robustness

1. Set $\varepsilon_p := \bar{\varepsilon}_p$, $\varepsilon := \bar{\varepsilon}$, $\delta_{g_p} := \bar{\delta}_{g_p}$, $\delta_g := \bar{\delta}_g$, and $\delta_\phi := \bar{\delta}_\phi$, where $\bar{\varepsilon}_{p,j} := \bar{\delta}_{g_p,j} \approx -\min_{(\mathbf{u}, \tau) \in \mathcal{J}_\tau} g_{p,j}(\mathbf{u}, \tau)$, $\bar{\varepsilon}_j := \bar{\delta}_{g,j} \approx -\min_{\mathbf{u} \in \mathcal{J}} g_j(\mathbf{u})$, and $\bar{\delta}_\phi \approx \phi_p(\mathbf{u}_0, \tau_0) - \min_{(\mathbf{u}, \tau) \in \mathcal{J}_\tau} \phi_p(\mathbf{u}, \tau)$. Set $\underline{P} := 0$, $\bar{P} := 1$.
2. Check the feasibility of (174) for the given choice of $\varepsilon_p, \varepsilon, \delta_{g_p}, \delta_g, \delta_\phi$ by solving a linear programming feasibility problem. If no solution exists and $\delta_\phi \geq \bar{\delta}_\phi/2^{10}$, set $\varepsilon_p := \varepsilon_p/2$, $\varepsilon := \varepsilon/2$, $\delta_{g_p} := \delta_{g_p}/2$, $\delta_g := \delta_g/2$, $\delta_\phi := \delta_\phi/2$, and repeat this step. Otherwise, proceed to Step 3.
3. If $\delta_\phi < \bar{\delta}_\phi/2^{10}$, terminate with $\bar{\mathbf{u}}_{k+1}^* := \mathbf{u}_{k^*}$. Otherwise, set $P := 0.5\underline{P} + 0.5\bar{P}$ and define the tightened bounds as

$$\begin{aligned} \nabla \underline{\phi}'_p(\mathbf{u}_{k^*}, \tau_{k+1}) &:= \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) \\ &+ P \left[\nabla \underline{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) - \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) \right], \\ \nabla \bar{\phi}'_p(\mathbf{u}_{k^*}, \tau_{k+1}) &:= \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) \\ &+ P \left[\nabla \bar{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) - \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1}) \right], \end{aligned} \quad (190)$$

$$\begin{aligned}\nabla g_{p,j}^t(\mathbf{u}_{k^*}, \tau_{k+1}) &:= \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) \\ &+ P \left[\nabla \underline{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) - \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) \right], \\ \nabla \bar{g}_{p,j}^t(\mathbf{u}_{k^*}, \tau_{k+1}) &:= \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) \\ &+ P \left[\nabla \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) - \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) \right].\end{aligned}$$

4. Check the feasibility of the robust projection with the tightened bounds for the obtained $\varepsilon_p, \varepsilon, \delta_{g_p}, \delta_g, \delta_\phi$:

$$\begin{aligned}\bar{\mathbf{u}}_{k+1}^* &:= \\ \arg \min_{\mathbf{u}, s_\phi, S} &\| \mathbf{u} - \mathbf{u}_{k+1}^* \|_2^2 \\ \text{subject to} & \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_p,j} \\ & \frac{\partial g_{p,j}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji} \\ & \frac{\partial \bar{g}_{p,j}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji}, \\ & \forall i = 1, \dots, n_u, \\ & \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ \forall j: & +\eta_{c,j} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ & + (1 - \eta_{c,j}) \bar{K}_{p,j\tau} (\tau_{k+1} - \tau_{k^*}) \geq -\varepsilon_{p,j} \\ & \nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j}, \\ \forall j: & g_j(\mathbf{u}_{k^*}) \geq -\varepsilon_j \\ & \sum_{i=1}^{n_u} s_{\phi,i} \leq -\delta_\phi \\ & \frac{\partial \phi^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\ & \frac{\partial \bar{\phi}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\ & \mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U.\end{aligned} \quad (191)$$

If (191) is infeasible, set $\bar{P} := P$. Otherwise, set $\underline{P} := P$.

If $\bar{P} - \underline{P} < 0.01$, proceed to Step 5. Otherwise, return to Step 3.

5. Set $P := 0.5\underline{P}$, define the tightened bounds as in (190), and solve (191) to obtain the projected target $\bar{\mathbf{u}}_{k+1}^*$. Terminate.

Algorithm 3 can be seen as consisting of two parts. In the first, appropriate projection parameters are chosen based on the non-robust projection carried out with the estimated gradients. Once a sufficiently low set of projection parameters is found, the second part consists in adding as much robustness as possible for this particular set by finding the largest possible gradient uncertainty set that still maintains the feasibility of the projection. Since this way of “pushing robustness to its limits” can have performance

drawbacks (as will be shown later), we take the heuristic approach of setting $P := 0.5\underline{P}$ to remove some of the robustness.

5.3 Accounting for Gradient Uncertainty in the Feasibility Conditions

Consider the feasibility-guaranteeing condition (151). From the inequalities

$$\frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \leq \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}), \quad (192)$$

$$\begin{aligned}& \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \\ & \leq \max \left[\begin{aligned} & \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right], \quad (193)\end{aligned}$$

we may propose a version of (151) that accounts for gradient uncertainty:

$$\begin{aligned}\min_{\bar{k}=0, \dots, k} & \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{K}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \bar{K}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{K}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right] \leq 0, \quad (194)\end{aligned}$$

which is implementable and robust since the satisfaction of (194) implies the satisfaction of (151).

5.4 Accounting for Gradient Uncertainty in the Cost Decrease Conditions

For the sufficient condition on cost decrease as given by (117), we consider first the inner product of the gradient with the change in decision variables in summation form:

$$\nabla \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} = \sum_{i=1}^{n_u} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}). \quad (195)$$

Using bounds analogous to that of (193), we may thus upper bound this term as

$$\nabla \phi_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq \sum_{i=1}^{n_u} \max \begin{bmatrix} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}) \end{bmatrix}, \quad (196)$$

which allows for the robust version of (117):

$$\sum_{i=1}^{n_u} \max \begin{bmatrix} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}) \end{bmatrix} + \frac{K_k}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \begin{bmatrix} \frac{M_{\phi,i_1 i_2}^{k^*}}{\bar{M}_{\phi,i_1 i_2}^{k^*}} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}), \\ \frac{M_{\phi,i_1 i_2}^{k^*}}{\bar{M}_{\phi,i_1 i_2}^{k^*}} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}) \end{bmatrix} \leq 0. \quad (197)$$

Satisfying (197) implies satisfying (117). However, it may occur that very conservative gradient bounds – as those, for example, in (189) – would lead to the terms in (197) being positive for any $K_k \in [0, 1]$, thereby making the satisfaction of this condition impossible. To avoid this, we propose to use the tightened gradient bounds instead:

$$\sum_{i=1}^{n_u} \max \begin{bmatrix} \frac{\partial \phi^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}), \\ \frac{\partial \bar{\phi}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}) \end{bmatrix} + \frac{K_k}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \begin{bmatrix} \frac{M_{\phi,i_1 i_2}^{k^*}}{\bar{M}_{\phi,i_1 i_2}^{k^*}} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}), \\ \frac{M_{\phi,i_1 i_2}^{k^*}}{\bar{M}_{\phi,i_1 i_2}^{k^*}} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}) \end{bmatrix} \leq 0. \quad (198)$$

As using more conservative gradient bounds will almost always lead to smaller acceptable K_k , it is also for this reason that we employ the heuristic $P := 0.5\bar{P}$ in Algorithm 3, as this allows for larger steps while compromising only part of the robustness.

For the necessary conditions for cost decrease, we are forced to modify (154) to account for gradient uncertainty.

Here, one must make modifications with respect to both the lower and upper bounds on the cost at the future and reference (respectively) experimental iterates, as we are interested in excluding those points for which the lower bound at the former is superior to the upper bound at the latter. Let us consider the following lower bound analogues to (192) and (193):

$$\frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \geq \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}), \quad (199)$$

$$\frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \geq \min \begin{bmatrix} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{bmatrix}, \quad (200)$$

which lead to the robust version of (154):

$$\max_{\bar{k}=0, \dots, k} \begin{bmatrix} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{v,\phi}^{\bar{k}}) \underline{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{v,\phi}^{\bar{k}}} \min \begin{bmatrix} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{bmatrix} \\ + \sum_{i \notin I_{v,\phi}^{\bar{k}}} \min \begin{bmatrix} \underline{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{bmatrix} \end{bmatrix} \quad (201)$$

$$\leq \min_{\bar{k}=0, \dots, k} \begin{bmatrix} \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,\phi}^{\bar{k}} \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}^{\bar{k}}) \bar{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,\phi}^{\bar{k}}} \max \begin{bmatrix} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \end{bmatrix} \\ + \sum_{i \notin I_{c,\phi}^{\bar{k}}} \max \begin{bmatrix} \underline{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}) \end{bmatrix} \end{bmatrix}.$$

Clearly, any point that would be excluded for failing to satisfy (201) would also be excluded for failing to satisfy (154).

5.5 Choosing a Reference Point While Accounting for Gradient Uncertainty

In choosing a reference point, one must further modify the subproblems of (155) and (157). Using bounds analogous to (192) and (199), one obtains

$$\begin{aligned}
 k^* &:= \\
 \arg \text{maximize } & \bar{k} \\
 & \bar{k} \in [0, k] \\
 \text{subject to } & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
 & + \eta_{c,j} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\
 & + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0, \\
 & \forall j = 1, \dots, n_{g_p}
 \end{aligned} \tag{202}$$

$$\begin{aligned}
 & \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k},k} \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\
 & + (1 - \eta_{v,\phi}^{\bar{k},k}) \underline{\kappa}_{\phi,\tau}^{\bar{k},k} (\tau_k - \tau_{\bar{k}}) \leq \\
 & \min_{\bar{k} \in \mathbf{k}_f} \left[\begin{aligned} & \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \\ & \eta_{c,\phi}^{\bar{k},k} \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,\phi}^{\bar{k},k}) \bar{\kappa}_{\phi,\tau}^{\bar{k},k} (\tau_k - \tau_{\bar{k}}) \end{aligned} \right], \\
 \mathbf{k}_f = & \left\{ \bar{k} : \begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \eta_{c,j} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \leq 0, \\ & \forall j = 1, \dots, n_{g_p} \end{aligned} \right\},
 \end{aligned} \tag{203}$$

$$\begin{aligned}
 k^* &:= \\
 \arg \text{minimize } & \max_{\bar{k} \in [0, k]} \max_{j=1, \dots, n_{g_p}} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \eta_{c,j} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}) \bar{\kappa}_{p,j\tau} (\tau_{k+1} - \tau_{\bar{k}}) \end{aligned} \right].
 \end{aligned} \tag{204}$$

Here, (202) selects the most recent point that is guaranteed to be robustly feasible and that cannot be guaranteed to have a cost function value superior to that at any other feasible point. (204) is a possible alternative for when no robustly feasible point is available, and simply selects the point with the lowest upper constraint value.

5.6 Accounting for Gradient Uncertainty in the Computation of Lower and Upper Bounds on the Experimental Function Values

As certain methods of calculating the lower and upper bounds on the experimental function values (discussed in

Section 4) use the function derivatives, these too must be modified to account for gradient uncertainty. Consider first the robust version of (144):

$$\begin{aligned}
 & \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
 & - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \eta_{c,\phi}^{\bar{k},\bar{k}} \max \left[\begin{aligned} & \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ & \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{aligned} \right] \\
 & - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} (1 - \eta_{c,\phi}^{\bar{k},\bar{k}}) \max \left[\begin{aligned} & \underline{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ & \bar{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{aligned} \right] - \bar{W}_{\phi,\bar{k}} \\
 & \leq \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \leq \\
 & \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
 & - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} \eta_{v,\phi}^{\bar{k},\bar{k}} \min \left[\begin{aligned} & \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ & \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{aligned} \right] \\
 & - \frac{1}{N} \sum_{\bar{k} \in \bar{\mathbf{k}}} (1 - \eta_{v,\phi}^{\bar{k},\bar{k}}) \min \left[\begin{aligned} & \underline{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ & \bar{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{aligned} \right] - \underline{W}_{\phi,\bar{k}},
 \end{aligned} \tag{205}$$

the robustness of which is easily seen by the fact that the resulting bounds must be more conservative than the bounds of the version with the exact derivatives (144).

Consider as well the bounds (148) and (149), the robust versions of which are given as

$$\begin{aligned}
 \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) & < \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
 & + \eta_{c,\phi}^{\bar{k},\bar{k}} \max \left[\begin{aligned} & \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ & \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{aligned} \right] \\
 & + (1 - \eta_{c,\phi}^{\bar{k},\bar{k}}) \max \left[\begin{aligned} & \underline{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ & \bar{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{aligned} \right] \\
 & + \sum_{i \in \bar{J}_{c,\phi}^{\bar{k},\bar{k}}} \max \left[\begin{aligned} & \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}), \\ & \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}) \end{aligned} \right] \\
 & + \sum_{i \notin \bar{J}_{c,\phi}^{\bar{k},\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}), \\ & \bar{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}) \end{aligned} \right],
 \end{aligned} \tag{206}$$

$$\begin{aligned}
\phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &> \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
&+ \eta_{v,\phi}^{\bar{k},\bar{k}} \min \left[\begin{array}{l} \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{array} \right] \\
&+ (1 - \eta_{v,\phi}^{\bar{k},\bar{k}}) \min \left[\begin{array}{l} \underline{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ \bar{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{array} \right] \\
&+ \sum_{i \in \bar{I}_{v,\phi}^{\bar{k},\bar{k}}} \min \left[\begin{array}{l} \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}) \end{array} \right] \\
&+ \sum_{i \notin \bar{I}_{v,\phi}^{\bar{k},\bar{k}}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}) \end{array} \right].
\end{aligned} \tag{207}$$

These follow from (148) and (149) by the same reasoning as before.

5.7 Gradient Uncertainty in the Lipschitz Consistency Check

Of the inequalities considered in the robust version of Algorithm 1, only (162) and (163) use derivatives and thus require modification. Using the same concepts as before, we modify them as follows:

$$\begin{aligned}
\underline{\phi}_p(\mathbf{u}_{k_2}, \tau_{k_2}) &\leq \\
\bar{\phi}_p(\mathbf{u}_{k_1}, \tau_{k_1}) &+ \max \left[\begin{array}{l} \underline{\kappa}_{\phi,\tau} (\tau_{k_2} - \tau_{k_1}), \\ \bar{\kappa}_{\phi,\tau} (\tau_{k_2} - \tau_{k_1}) \end{array} \right] \\
&+ \sum_{i=1}^{n_u} \max \left[\begin{array}{l} \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k_1}, \tau_{k_1})} (u_{k_2,i} - u_{k_1,i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k_1}, \tau_{k_1})} (u_{k_2,i} - u_{k_1,i}) \end{array} \right] \\
&+ \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \left[\begin{array}{l} \underline{M}_{\phi,i_1 i_2} (u_{k_2,i_1} - u_{k_1,i_1}) \\ \bar{M}_{\phi,i_1 i_2} (u_{k_2,i_1} - u_{k_1,i_1}) \\ (u_{k_2,i_2} - u_{k_1,i_2}), \end{array} \right],
\end{aligned} \tag{208}$$

$$\begin{aligned}
\bar{\phi}_p(\mathbf{u}_{k_2}, \tau_{k_2}) &\geq \\
\underline{\phi}_p(\mathbf{u}_{k_1}, \tau_{k_1}) &+ \min \left[\begin{array}{l} \underline{\kappa}_{\phi,\tau} (\tau_{k_2} - \tau_{k_1}), \\ \bar{\kappa}_{\phi,\tau} (\tau_{k_2} - \tau_{k_1}) \end{array} \right] \\
&+ \sum_{i=1}^{n_u} \min \left[\begin{array}{l} \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k_1}, \tau_{k_1})} (u_{k_2,i} - u_{k_1,i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k_1}, \tau_{k_1})} (u_{k_2,i} - u_{k_1,i}) \end{array} \right]
\end{aligned} \tag{209}$$

$$+ \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \min \left[\begin{array}{l} \underline{M}_{\phi,i_1 i_2} (u_{k_2,i_1} - u_{k_1,i_1}) \\ \bar{M}_{\phi,i_1 i_2} (u_{k_2,i_1} - u_{k_1,i_1}) \\ (u_{k_2,i_2} - u_{k_1,i_2}), \end{array} \right].$$

It follows that any Lipschitz constants that are inconsistent for (208) and (209) will be inconsistent for (162) and (163) as well.

5.8 Example

Considering the example of Section 4.6, we suppose that the gradient estimation algorithm is able to obtain derivative estimates that are the true derivatives corrupted by additive uniform noise, i.e.:

$$\begin{aligned}
\frac{\partial \hat{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} &:= \\
&\frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} + \alpha_\sigma (\bar{\kappa}_{\phi,i} - \underline{\kappa}_{\phi,i}) \mathcal{U}[-1, 1], \\
\frac{\partial \hat{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} &:= \\
&\frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} + \alpha_\sigma (\bar{\kappa}_{\phi,\tau} - \underline{\kappa}_{\phi,\tau}) \mathcal{U}[-1, 1], \\
\frac{\partial \hat{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} &:= \\
&\frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} + \alpha_\sigma (\bar{\kappa}_{\phi,i} - \underline{\kappa}_{\phi,i}) \mathcal{U}[-1, 1], \\
\frac{\partial \hat{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} &:= \\
&\frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{k+1})} + \alpha_\sigma (\bar{\kappa}_{\phi,\tau} - \underline{\kappa}_{\phi,\tau}) \mathcal{U}[-1, 1],
\end{aligned} \tag{210}$$

with analogous formulas for the constraints. Here, $\alpha_\sigma > 0$ is used to control the size of the uncertainty, with the scaling taken care of by multiplying with the range that the derivatives may take, which is simply the difference between the upper and lower global Lipschitz constants.

Valid bounds are then computed as

$$\begin{aligned}
\frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} &:= \frac{\partial \hat{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} - \alpha_\sigma (\bar{\kappa}_{\phi,i} - \underline{\kappa}_{\phi,i}), \\
\frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} &:= \frac{\partial \hat{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} + \alpha_\sigma (\bar{\kappa}_{\phi,i} - \underline{\kappa}_{\phi,i}), \\
\frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} &:= \frac{\partial \hat{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} - \alpha_\sigma (\bar{\kappa}_{\phi,\tau} - \underline{\kappa}_{\phi,\tau}), \\
\frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} &:= \frac{\partial \hat{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} + \alpha_\sigma (\bar{\kappa}_{\phi,\tau} - \underline{\kappa}_{\phi,\tau}),
\end{aligned} \tag{211}$$

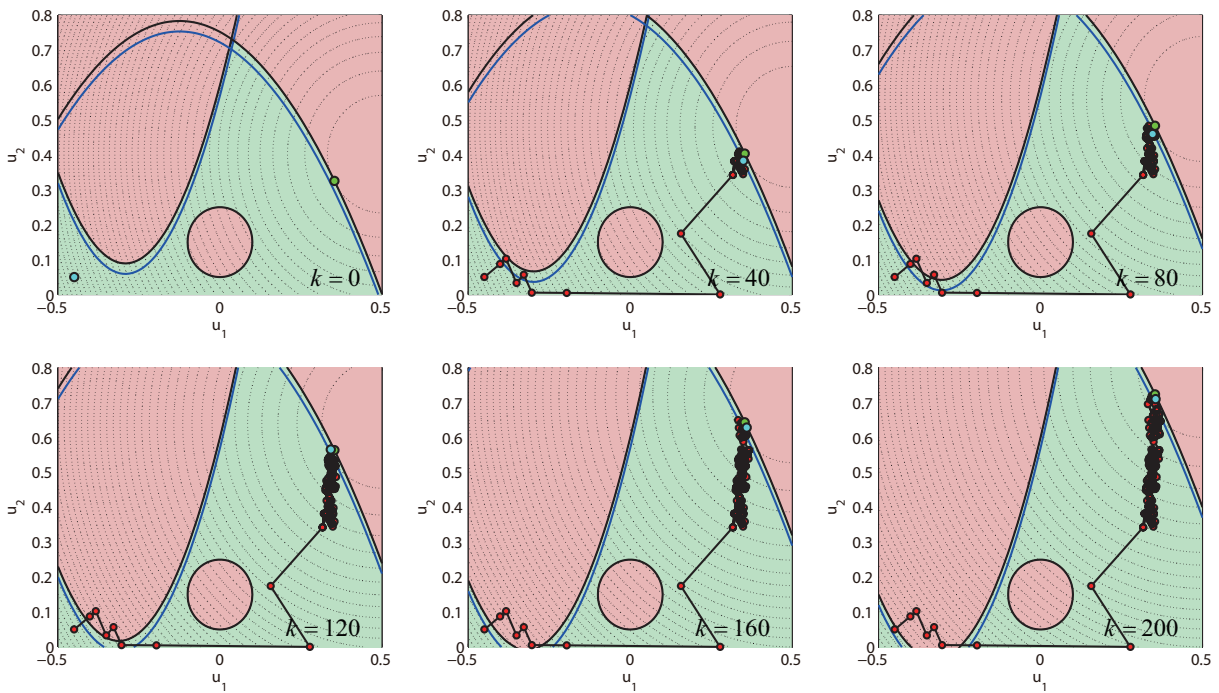


Fig. 19 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (–) scenario with estimation error in the experimental function gradients accounted for. $\alpha_\sigma = 0.05$ is used here.

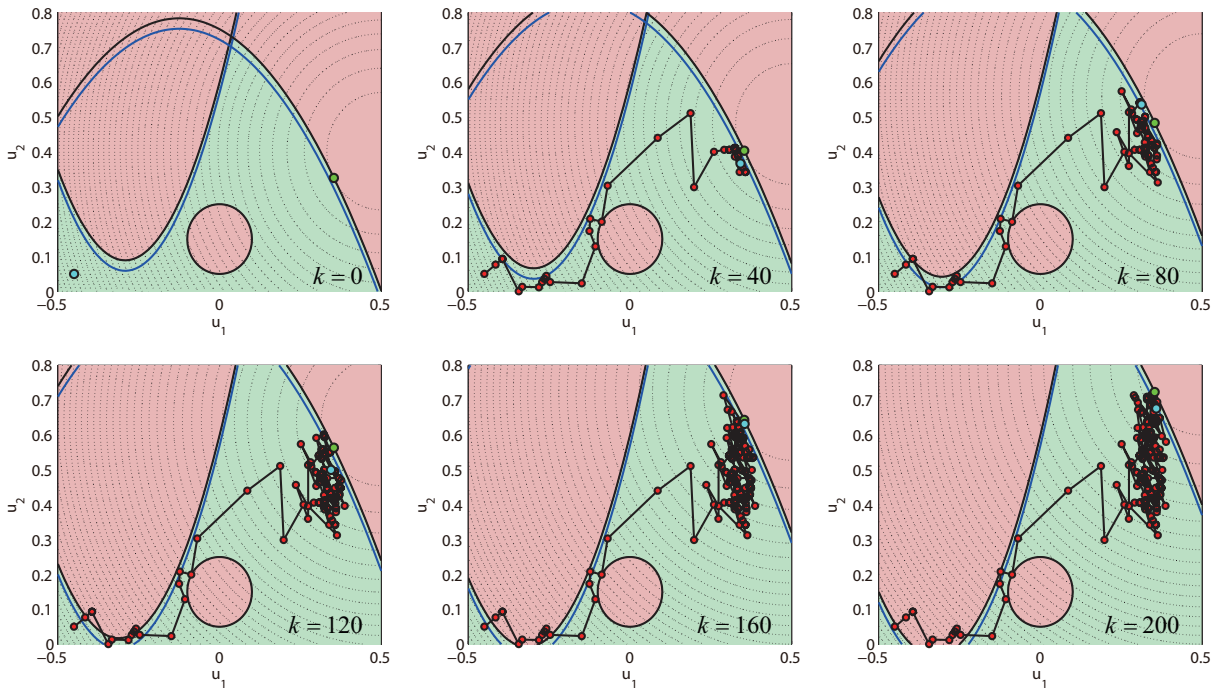


Fig. 20 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (–) scenario with estimation error in the experimental function gradients accounted for. $\alpha_\sigma = 0.15$ is used here.

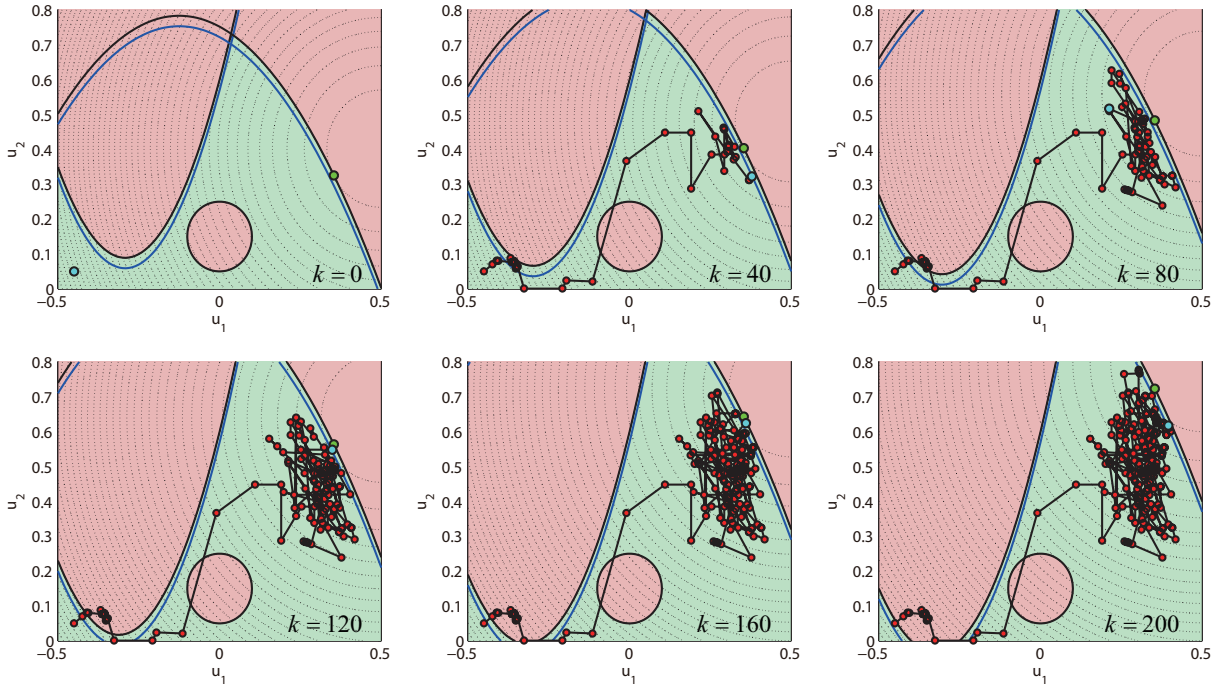


Fig. 21 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (–) scenario with estimation error in the experimental function gradients accounted for. $\alpha_\sigma = 0.25$ is used here.

$$\begin{aligned} \frac{\partial \underline{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_k, \tau_{k+1})} &:= \frac{\partial \hat{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_k, \tau_{k+1})} - \alpha_\sigma (\bar{\mathbf{K}}_{\phi,i} - \underline{\mathbf{K}}_{\phi,i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_k, \tau_{k+1})} &:= \frac{\partial \hat{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_k, \tau_{k+1})} + \alpha_\sigma (\bar{\mathbf{K}}_{\phi,i} - \underline{\mathbf{K}}_{\phi,i}), \\ \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_k, \tau_{k+1})} &:= \frac{\partial \hat{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_k, \tau_{k+1})} - \alpha_\sigma (\bar{\mathbf{K}}_{\phi,\tau} - \underline{\mathbf{K}}_{\phi,\tau}), \\ \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_k, \tau_{k+1})} &:= \frac{\partial \hat{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_k, \tau_{k+1})} + \alpha_\sigma (\bar{\mathbf{K}}_{\phi,\tau} - \underline{\mathbf{K}}_{\phi,\tau}). \end{aligned}$$

Results are provided for increasing values of α_σ (0.05, 0.15, and 0.25) in Figs. 19-21, with the corresponding cost values shown in Fig. 22. As expected, uncertainty in the gradient estimates does not affect the feasibility of the experimental iterates, as this is guaranteed by the robust modifications. However, performance with regard to convergence speed does degrade with increasing uncertainty, in that the general trend remains the same but with the variance of the iterates increasing with α_σ . Additionally, it is more difficult for the implementation to “maneuver” around the experimental concave constraint as the uncertainty increases, which leads to an increase in the number of experiments needed to get to the neighborhood of the optimum.

We note that the gradient estimation algorithm supposed for this example is artificial in nature and used only for its simplicity – for a look at how real estimation algorithms

would perform in the SCFO context, the interested reader is referred to the examples in [7].

6 Guarantee of Sufficient Excitation

The proper management of the information-optimization tradeoff is a common requirement in both experimental and derivative-free optimization frameworks. Stated in simple terms, this basically means that one cannot *only* optimize since to optimize well one needs to sample the experimental space in a manner that is adequate for getting the information necessary for optimization. In the SCFO framework, this means sufficiently perturbing the decision variables in all directions so as to be able to allow the gradient estimation algorithm to properly estimate the full gradients of the experimental functions in question [31, 5]. Similarly, derivative-free optimization algorithms often need to ensure the more general property of well-posedness so as to be able to build proper interpolation or regression models from the obtained data [12, 13]. Finally, methods employing parametric models of the experimental functions may also want to perturb the decision variables in a way that makes identification of these parameters possible [30, 29].

Regardless of the context, the basic requirement is mathematically the same – one should be able, when needed, to sufficiently excite the decision variables without incurring any significant drawbacks. As temporary economic losses reflected by increase in the cost function are, in the general

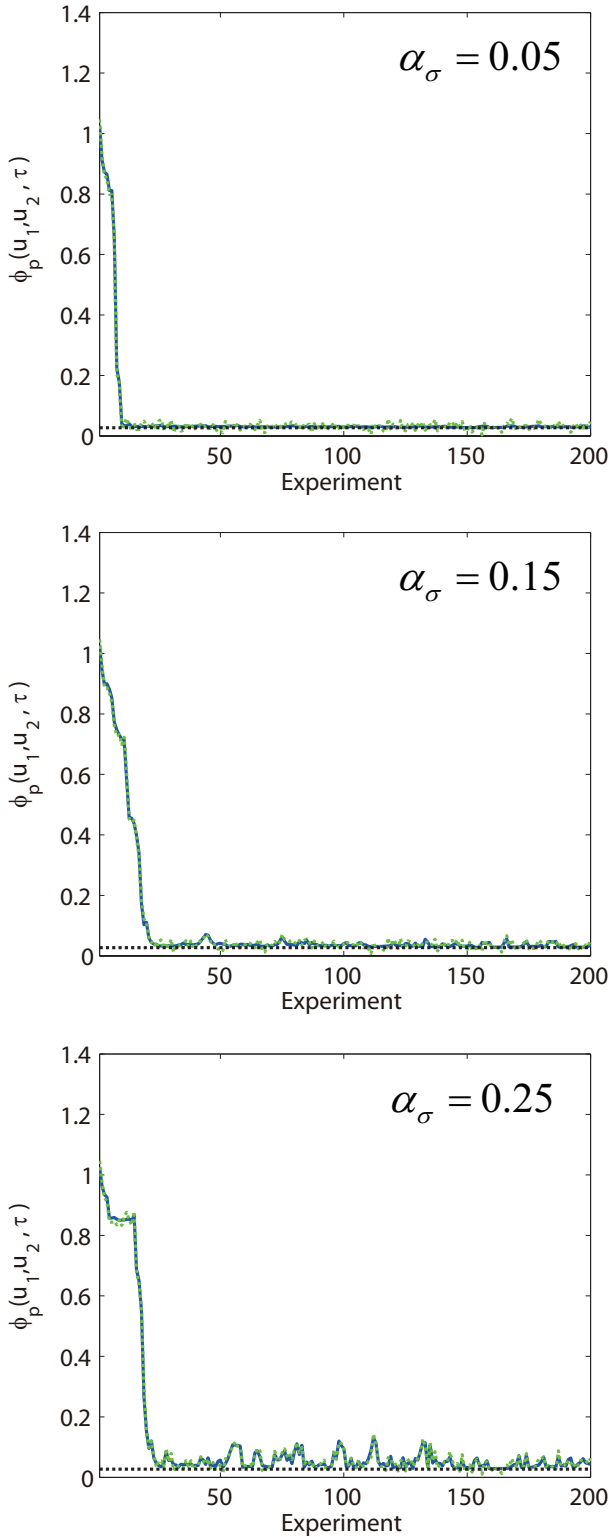


Fig. 22 Cost function values obtained by the modified SCFO methodology for Problem (42) for the (–) scenario with estimation error in the experimental function gradients accounted for.

case, inevitable once one overrides optimization to gather information, we will not consider these as “significant”. What is significant, however, are constraint violations, which are not admissible and must be avoided even while perturbing the system for information. This issue becomes particularly relevant when one approaches a constrained optimum.

In this section, we will propose to solve this problem by ensuring that there always exists a ball of radius $\delta_e > 0$ around the reference iterate \mathbf{u}_{k^*} :

$$\mathcal{B}_{e,k^*} = \{\mathbf{u} : \|\mathbf{u} - \mathbf{u}_{k^*}\|_2 \leq \delta_e\}, \quad (212)$$

such that all $\mathbf{u} \in \mathcal{B}_{e,k^*}$ are guaranteed to be feasible. The following two points justify this choice of approach:

- By ensuring the existence of a feasible *ball*, one is able to perturb in all directions and to generate a sample set with any geometry.
- Enforcing that the ball be of a certain radius δ_e allows for one to tune the size of the excitations with respect to the noise or estimation error, as perturbations that are too small may, for example, lead to overly corrupted gradient estimates [24, 31].

It should be clear that guaranteeing the existence of such a ball will be impossible if a constraint is approached too closely – one cannot sit on an active constraint and perturb in absolutely any direction without losing feasibility. From this, it follows that a natural way to guarantee that an entire ball is feasible is by adding constraint *back-offs*, denoted here by $b > 0$, to the constraints of the original problem (17):

$$\begin{aligned} & \underset{\mathbf{u}}{\text{minimize}} \quad \phi_p(\mathbf{u}, \tau) \\ & \text{subject to} \quad g_{p,j}(\mathbf{u}, \tau) + b_{p,j}^{k^*,k} \leq 0, \quad j = 1, \dots, n_{g_p} \\ & \quad \quad \quad g_j(\mathbf{u}) + b_j^{k^*} \leq 0, \quad j = 1, \dots, n_g \\ & \quad \quad \quad \mathbf{u}^L + \mathbf{b}_u \preceq \mathbf{u} \preceq \mathbf{u}^U - \mathbf{b}_u, \end{aligned} \quad (213)$$

which will allow us to achieve our stated goal of guaranteeing a feasible sufficient-excitation ball at the cost of potential suboptimality, which will naturally result due to the feasible space being tightened. So as to work with a very general case, we make the back-offs for the experimental constraints *iteration-dependent* with respect to both k^* and k so as to properly handle degradation effects and to exploit some of the relaxation theory developed in the earlier sections. For the numerical constraints, it will be shown that iteration dependence with respect to k^* only is needed since degradation is absent in these constraints. As the bound constraints are not subject to degradation and cannot be relaxed, we make their back-offs constant and independent of iteration. The basic idea of adding a back-off to guarantee the existence of a feasible ball is illustrated geometrically in Fig. 23.

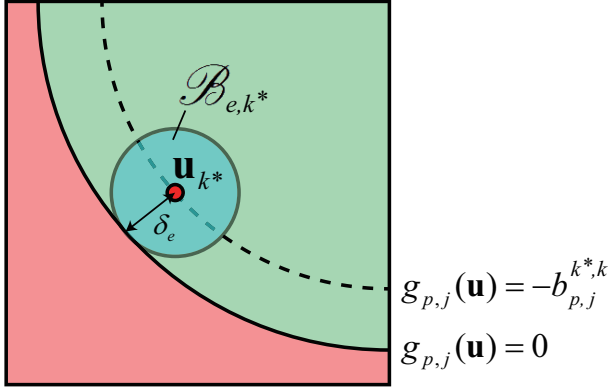


Fig. 23 Guaranteeing the existence of a sufficient-excitation ball by adding a constraint back-off. A single experimental constraint is considered here.

In what follows, we will derive the back-off values sufficient to guarantee the feasibility of \mathcal{B}_{e,k^*} , and then proceed to describe how the SCFO should be modified further to account for solving (213) instead of (17).

6.1 Determining the Sufficient Back-offs

Although we seek to guarantee

$$\begin{aligned} g_{p,j}(\mathbf{u}, \tau_{k+1}) &\leq 0, \quad \forall j = 1, \dots, n_g, \\ g_j(\mathbf{u}) &\leq 0, \quad \forall j = 1, \dots, n_g, \\ \mathbf{u}^U &\preceq \mathbf{u} \preceq \mathbf{u}^L \end{aligned} \quad (214)$$

$\forall \mathbf{u} \in \mathcal{B}_{e,k^*}$, we may proceed by deriving the appropriate back-off for each constraint separately as the back-offs do not “interact” – i.e., a back-off that guarantees the existence of a ball that is feasible with respect to one constraint in no way influences the value of the back-off needed to guarantee the existence of such a ball with respect to another.

We will first carry out the theoretical analysis for a given experimental constraint $g_{p,j}$. So as to derive back-offs that reduce conservatism by exploiting local relaxations, we define the following local temporal experimental space:

$$\mathcal{S}_\tau^{e,k^*} = \mathcal{B}_{e,k^*} \times \{\tau : \tau_{k^*} \leq \tau \leq \tau_{k+1}\}, \quad (215)$$

and the corresponding local Lipschitz constants

$$\underline{\kappa}_{p,ji} \leq \underline{\kappa}_{p,ji}^{e,k^*} < \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}, \tau)} < \overline{\kappa}_{p,ji}^{e,k^*} \leq \overline{\kappa}_{p,ji}, \quad (216)$$

$$\forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{e,k^*},$$

$$\underline{\kappa}_{p,j\tau} \leq \underline{\kappa}_{p,j\tau}^{e,k^*} \leq \frac{\partial g_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}, \tau)} \leq \overline{\kappa}_{p,j\tau}^{e,k^*} \leq \overline{\kappa}_{p,j\tau}, \quad (217)$$

$$\forall (\mathbf{u}, \tau) \in \mathcal{S}_\tau^{e,k^*}.$$

Likewise, we will define the concavity index sets $\eta_{c,j}^{e,k^*}$ and $I_{c,j}^{e,k^*}$ in the same manner as $\eta_{c,j}^{\bar{k}}$ and $I_{c,j}^{\bar{k}}$ but with respect to \mathcal{S}_τ^{e,k^*} rather than $\mathcal{S}_\tau^{\bar{k}}$. By virtue of all of the same principles as before, this allows us to state the bound

$$\begin{aligned} g_{p,j}(\mathbf{u}, \tau_{k+1}) &\leq \overline{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ &+ \eta_{c,j}^{e,k^*} \frac{\partial \overline{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ &+ (1 - \eta_{c,j}^{e,k^*}) \overline{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ &+ \sum_{i \in I_{c,j}^{e,k^*}} \max \left[\frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_i - u_{k^*,i}), \right. \\ &\quad \left. \frac{\partial \overline{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_i - u_{k^*,i}) \right] \\ &+ \sum_{i \notin I_{c,j}^{e,k^*}} \max \left[\frac{\underline{\kappa}_{p,ji}^{e,k^*}}{\overline{\kappa}_{p,ji}^{e,k^*}} (u_i - u_{k^*,i}), \right] \end{aligned} \quad (218)$$

for all $\mathbf{u} \in \mathcal{B}_{e,k^*}$.

The major theoretical result of this section is given in the following theorem.

Theorem 1 (Sufficient back-off for an experimental constraint function) *Let the reference point \mathbf{u}_{k^*} satisfy the backed off robust constraint $\overline{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + b_{p,j}^{k^*,k} \leq 0$. Defining*

$$\begin{aligned} \kappa_{p,ji}^{m,k^*} &= \begin{cases} \max \left[\left| \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} \right|, \right. \\ \quad \left. \left| \frac{\partial \overline{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} \right| \right], & i \in I_{c,j}^{e,k^*} \\ \max \left[|\underline{\kappa}_{p,ji}^{e,k^*}|, |\overline{\kappa}_{p,ji}^{e,k^*}| \right], & i \notin I_{c,j}^{e,k^*} \end{cases} \\ \kappa_{p,j}^{m,k^*} &= \left[\kappa_{p,j1}^{m,k^*} \quad \dots \quad \kappa_{p,jn_u}^{m,k^*} \right]^T \end{aligned} \quad (219)$$

and setting the back-off as

$$\begin{aligned} b_{p,j}^{k^*,k} &= \eta_{c,j}^{e,k^*} \frac{\partial \overline{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ &+ (1 - \eta_{c,j}^{e,k^*}) \overline{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) + \delta_e \|\kappa_{p,j}^{m,k^*}\|_2 \end{aligned} \quad (220)$$

guarantees that

$$g_{p,j}(\mathbf{u}, \tau_{k+1}) \leq 0, \quad \forall \mathbf{u} \in \mathcal{B}_{e,k^*}. \quad (221)$$

Proof We start with the following general analysis for any given $x, \bar{x}, y \in \mathbb{R}$:

$$\begin{aligned} xy &\leq |x||y|, \quad \bar{x}y \leq |\bar{x}||y| \\ \Rightarrow xy &\leq \max[|x|, |\bar{x}|]|y|, \quad \bar{x}y \leq \max[|x|, |\bar{x}|]|y| \\ \Leftrightarrow \max[xy, \bar{x}y] &\leq \max[|x|, |\bar{x}|]|y|, \end{aligned} \quad (222)$$

from which we have

$$\begin{aligned} \max \left[\begin{array}{l} \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_i - u_{k^*,i}), \\ \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (u_i - u_{k^*,i}) \end{array} \right] \\ \leq \kappa_{p,ji}^{m,k^*} |u_i - u_{k^*,i}|, \quad i \in I_{c,j}^{e,k^*} \\ \max \left[\begin{array}{l} \underline{\kappa}_{p,ji} (u_i - u_{k^*,i}), \\ \bar{\kappa}_{p,ji} (u_i - u_{k^*,i}) \end{array} \right] \leq \kappa_{p,ji}^{m,k^*} |u_i - u_{k^*,i}|, \quad i \notin I_{c,j}^{e,k^*}. \end{aligned} \quad (223)$$

This then allows us to extend (218) as

$$\begin{aligned} g_{p,j}(\mathbf{u}, \tau_{k+1}) &\leq \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ &+ \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ &+ (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ &+ \sum_{i=1}^{n_u} \kappa_{p,ji}^{m,k^*} |u_i - u_{k^*,i}|. \end{aligned} \quad (224)$$

Let us denote by \mathcal{K}_k^* the set of decision variables for which the right-hand side of (224) is nonpositive:

$$\mathcal{K}_k^* = \left\{ \mathbf{u} : \begin{array}{l} \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ + (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ + \sum_{i=1}^{n_u} \kappa_{p,ji}^{m,k^*} |u_i - u_{k^*,i}| \leq 0 \end{array} \right\}. \quad (225)$$

It follows that $g_{p,j}(\mathbf{u}, \tau_{k+1}) \leq 0, \forall \mathbf{u} \in \mathcal{K}_k^* \cap \mathcal{B}_{e,k^*}$.

We will complete the proof by essentially inscribing the ball \mathcal{B}_{e,k^*} inside \mathcal{K}_k^* and showing that $\mathcal{B}_{e,k^*} \subseteq \mathcal{K}_k^*$ for the back-off proposed in (220) – this, of course, implies that $g_{p,j}(\mathbf{u}, \tau_{k+1}) \leq 0, \forall \mathbf{u} \in \mathcal{B}_{e,k^*}$ as well, which is our desired result. However, since \mathcal{B}_{e,k^*} is a hypersphere and \mathcal{K}_k^* is a convex set, it is sufficient⁵ to show that $\text{bd}(\mathcal{B}_{e,k^*}) \subseteq \mathcal{K}_k^*$ since $\text{bd}(\mathcal{B}_{e,k^*}) \subseteq \mathcal{K}_k^* \Leftrightarrow \mathcal{B}_{e,k^*} \subseteq \mathcal{K}_k^*$.

Let us parameterize the boundary points of \mathcal{B}_{e,k^*} as

$$\mathbf{u}_{\text{bd}} = \mathbf{u}_{k^*} + \delta_e \delta \mathbf{u}, \quad (226)$$

where $\delta \mathbf{u} \in \mathbb{R}^{n_u}$ is a unit vector. Since the mapping between $\delta \mathbf{u}$ and \mathbf{u}_{bd} is bijective, the desired result may be obtained by proving that for any choice of $\delta \mathbf{u}$ the resulting boundary point belongs to \mathcal{K}_k^* . Substituting \mathbf{u}_{bd} for \mathbf{u} in (225) yields

⁵ This is easily proven by considering the fact that any point in the interior of the sphere may be described as a convex combination of two boundary points. If all boundary points of \mathcal{B}_{e,k^*} belong to the convex set \mathcal{K}_k^* , their convex combinations (the interior points) must as well.

$$\begin{aligned} \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ + (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ + \delta_e \sum_{i=1}^{n_u} \kappa_{p,ji}^{m,k^*} |\delta u_i| \leq 0. \end{aligned} \quad (227)$$

To prove that (227) holds for all $\delta \mathbf{u}$, it is sufficient to only consider the direction that maximizes the value of the right-hand side or, more specifically, the direction that maximizes the summation term. In other words, we need to prove that

$$\begin{aligned} \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ + (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ + \delta_e \max_{\|\delta \mathbf{u}\|_2=1} \sum_{i=1}^{n_u} \kappa_{p,ji}^{m,k^*} |\delta u_i| \leq 0. \end{aligned} \quad (228)$$

To make the analysis easier, we will first show that we may restrict our consideration to $\delta \mathbf{u} \in \mathbb{R}_+^{n_u}$ since the point where the maximum is attained is nonunique and since there will always be one in the positive orthant. To prove this, suppose that there exists a direction $\delta \tilde{\mathbf{u}} \notin \mathbb{R}_+^{n_u}, \|\delta \tilde{\mathbf{u}}\|_2 = 1$ such that

$$\sum_{i=1}^{n_u} \kappa_{p,ji}^{m,k^*} |\delta \tilde{u}_i| > \max_{\substack{\|\delta \mathbf{u}\|_2=1 \\ \delta \mathbf{u} \in \mathbb{R}_+^{n_u}}} \sum_{i=1}^{n_u} \kappa_{p,ji}^{m,k^*} |\delta u_i|. \quad (229)$$

Clearly, the choice $\delta u_i := |\delta \tilde{u}_i|, i = 1, \dots, n_u$ shows that there exists a unit vector $\delta \mathbf{u} \in \mathbb{R}_+^{n_u}$ with a value equal to the left-hand side of (229), thereby proving that (229) cannot hold for any choice of $\delta \tilde{\mathbf{u}}$. As such, we may simplify (228) further:

$$\begin{aligned} \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ + (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ + \delta_e \max_{\substack{\|\delta \mathbf{u}\|_2=1 \\ \delta \mathbf{u} \in \mathbb{R}_+^{n_u}}} \left(\kappa_{p,j}^{m,k^*} \right)^T \delta \mathbf{u} \leq 0, \end{aligned} \quad (230)$$

where we have removed the absolute value due to the positive-orthant restriction and have written the resulting summation in vector form.

To evaluate the maximum term analytically, we may consider a simpler case by taking the following steps:

$$\begin{aligned} \max_{\substack{\|\delta \mathbf{u}\|_2=1 \\ \delta \mathbf{u} \in \mathbb{R}_+^{n_u}}} \left(\kappa_{p,j}^{m,k^*} \right)^T \delta \mathbf{u} &= \max_{\substack{\|\delta \mathbf{u}\|_2=1 \\ \delta \mathbf{u} \in \mathbb{R}_+^{n_u}}} \left(\kappa_{p,j}^{m,k^*} \right)^T \delta \mathbf{u} \\ &\leq \max_{\|\delta \mathbf{u}\|_2 \leq 1} \left(\kappa_{p,j}^{m,k^*} \right)^T \delta \mathbf{u}, \end{aligned} \quad (231)$$

as squaring the unit-norm constraint does not affect the directions considered and relaxing the restrictions by relaxing the equality and removing the positivity constraint can only increase the maximum value. As the last term is equivalent to maximizing a linear function over a unit ball, writing out its stationarity conditions readily yields the following maximum point:

$$\frac{\boldsymbol{\kappa}_{p,j}^{m,k^*}}{\|\boldsymbol{\kappa}_{p,j}^{m,k^*}\|_2} = \arg \max_{\boldsymbol{\delta}\mathbf{u}} \left(\boldsymbol{\kappa}_{p,j}^{m,k^*} \right)^T \boldsymbol{\delta}\mathbf{u} \quad (232)$$

subject to $\boldsymbol{\delta}\mathbf{u}^T \boldsymbol{\delta}\mathbf{u} \leq 1$,

provided that $\boldsymbol{\kappa}_{p,j}^{m,k^*} \neq \mathbf{0}$. This allows us to advance (231) to state:

$$\max_{\substack{\|\boldsymbol{\delta}\mathbf{u}\|_2=1 \\ \boldsymbol{\delta}\mathbf{u} \in \mathbb{R}^{n_u}}} \left(\boldsymbol{\kappa}_{p,j}^{m,k^*} \right)^T \boldsymbol{\delta}\mathbf{u} \leq \|\boldsymbol{\kappa}_{p,j}^{m,k^*}\|_2, \quad (233)$$

which we note is valid for any $\boldsymbol{\kappa}_{p,j}^{m,k^*}$ (the bound holding trivially in the somewhat pathological case of $\boldsymbol{\kappa}_{p,j}^{m,k^*} = \mathbf{0}$).

Using this result to push (230) further, we obtain

$$\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \boldsymbol{\tau}} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\ & + (1 - \eta_{c,j}^{e,k^*}) \bar{\boldsymbol{\kappa}}_{p,j\boldsymbol{\tau}}^{e,k^*} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) + \delta_e \|\boldsymbol{\kappa}_{p,j}^{m,k^*}\|_2 \leq 0, \end{aligned} \quad (234)$$

which must clearly hold given the choice of back-off in (220) and the assumed condition $\bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) + b_{p,j}^{k^*,k} \leq 0$. Since (234) \Rightarrow (230) \Rightarrow (228) \Rightarrow (227), we have proven that $\mathbf{u}_{bd} \in \mathcal{K}_{k^*}$ for all possible \mathbf{u}_{bd} and thus that $\mathcal{B}_{e,k^*} \subseteq \mathcal{K}_{k^*}$, which, since $g_{p,j}(\mathbf{u}, \boldsymbol{\tau}_{k+1}) \leq 0, \forall \mathbf{u} \in \mathcal{K}_{k^*} \cap \mathcal{B}_{e,k^*}$, finally proves that $g_{p,j}(\mathbf{u}, \boldsymbol{\tau}_{k+1}) \leq 0, \forall \mathbf{u} \in \mathcal{B}_{e,k^*}$. \square

We now proceed with the simpler derivations for the numerical and bound constraints.

Theorem 2 (Sufficient back-off for a numerical constraint function) *Let the reference point \mathbf{u}_{k^*} satisfy the backed off numerical constraint $g_j(\mathbf{u}_{k^*}) + b_j^{k^*} \leq 0$. Setting the back-off as*

$$b_j^{k^*} = \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} g_j(\mathbf{u}) - g_j(\mathbf{u}_{k^*}) \quad (235)$$

guarantees that

$$g_j(\mathbf{u}) \leq 0, \quad \forall \mathbf{u} \in \mathcal{B}_{e,k^*}. \quad (236)$$

Proof Substituting (235) into the assumed condition $g_j(\mathbf{u}_{k^*}) + b_j^{k^*} \leq 0$ yields

$$\max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} g_j(\mathbf{u}) \leq 0, \quad (237)$$

which is equivalent to (236). \square

Corollary 1 (Sufficient back-off for a bound constraint)

Let the reference point \mathbf{u}_{k^} satisfy the backed off bound constraints $u_i^L + b_{u,i} \leq u_{k^*,i} \leq u_i^U - b_{u,i}$. Setting the back-off as*

$$b_{u,i} = \delta_e \quad (238)$$

guarantees that

$$\mathbf{u}^L \preceq \mathbf{u} \preceq \mathbf{u}^U, \quad \forall \mathbf{u} \in \mathcal{B}_{e,k^*}. \quad (239)$$

Proof As the bound constraints are just special cases of the general numerical constraints, we will place them into the canonical form $g_j(\mathbf{u}) \leq 0$:

$$u_i^L - u_i \leq 0, \quad u_i - u_i^U \leq 0, \quad (240)$$

and then apply the result of Theorem 2:

$$\begin{aligned} b_{u,i} &= \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} (u_i^L - u_i) - (u_i^L - u_{k^*,i}) \\ &= \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} (u_i - u_i^U) - (u_{k^*,i} - u_i^U) = \delta_e. \quad \square \end{aligned} \quad (241)$$

6.2 Choosing the Reference Point to Satisfy the Back-offs

The back-offs derived in (220), (235), and (238) guarantee that both the reference point \mathbf{u}_{k^*} and all the points in the ball around it, \mathcal{B}_{e,k^*} , satisfy the problem constraints provided that

$$\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) + b_{p,j}^{k^*,k} \leq 0, \quad \forall j = 1, \dots, n_{g_p} \\ & g_j(\mathbf{u}_{k^*}) + b_j^{k^*} \leq 0, \quad \forall j = 1, \dots, n_g \\ & \mathbf{u}^L + \mathbf{b}_u \preceq \mathbf{u}_{k^*} \preceq \mathbf{u}^U - \mathbf{b}_u, \end{aligned} \quad (242)$$

or, if we substitute in the results of (220), (235), and (238) directly,

$$\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \boldsymbol{\tau}} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\ & + (1 - \eta_{c,j}^{e,k^*}) \bar{\boldsymbol{\kappa}}_{p,j\boldsymbol{\tau}}^{e,k^*} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) + \delta_e \|\boldsymbol{\kappa}_{p,j}^{m,k^*}\|_2 \leq 0, \\ & \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} g_j(\mathbf{u}) \leq 0, \quad \forall j = 1, \dots, n_g \\ & \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{k^*} \preceq \mathbf{u}^U - \delta_e \mathbf{1}. \end{aligned} \quad (243)$$

We may thus use these conditions as constraints in the search for a reference point and modify the search of (202) as

$$\begin{aligned}
k^* &:= \\
\arg \text{maximize } \bar{k} \\
&\quad \bar{k} \in [0, k] \\
&\quad \text{subject to } \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
&\quad \quad + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\
&\quad \quad + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\
&\quad \quad + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 \leq 0, \quad \forall j = 1, \dots, n_{g_p} \\
&\quad \quad \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \leq 0, \quad \forall j = 1, \dots, n_g \\
&\quad \quad \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{\bar{k}} \preceq \mathbf{u}^U - \delta_e \mathbf{1} \\
&\quad \quad \phi_{-p}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k},k} \frac{\partial \phi_{-p}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\
&\quad \quad + (1 - \eta_{v,\phi}^{\bar{k},k}) \bar{\kappa}_{\phi,\tau}^{\bar{k},k} (\tau_k - \tau_{\bar{k}}) \leq \\
&\quad \quad \min_{\bar{k} \in \mathbf{k}_f} \left[\begin{array}{l} \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \\ \eta_{c,\phi}^{\bar{k},k} \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}^{\bar{k},k}) \bar{\kappa}_{\phi,\tau}^{\bar{k},k} (\tau_k - \tau_{\bar{k}}) \end{array} \right], \quad (244)
\end{aligned}$$

$$\mathbf{k}_f = \left\{ \bar{k} : \left[\begin{array}{l} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 \leq 0, \quad \forall j = 1, \dots, n_{g_p}; \\ \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \leq 0, \quad \forall j = 1, \dots, n_g; \\ \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{\bar{k}} \preceq \mathbf{u}^U - \delta_e \mathbf{1} \end{array} \right] \right\}. \quad (245)$$

It may of course occur that no experimental iterate satisfy the restrictions of (244). Naturally, this is more likely in the back-off case since the restrictions are made tighter, and choosing both overly conservative Lipschitz constants or a δ_e value that is too large can lead to infeasibility in the reference point search. If one cannot obtain a \mathbf{u}_{k^*} by reducing the conservatism of the Lipschitz constants or choosing to work with a smaller δ_e , the natural option, if one insists on continuing to optimize, would be to choose as the reference that iteration at which the maximum violation is minimized:

$$k^* := \arg \min_{\bar{k} \in [0, k]} \max \left[g_{p,m}^{\bar{k}}, g_m^{\bar{k}}, u_{L,m}^{\bar{k}}, u_{U,m}^{\bar{k}} \right], \quad (246)$$

where

$$\begin{aligned}
g_{p,m}^{\bar{k}} &= \max_{j=1, \dots, n_{g_p}} \left[\begin{array}{l} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 \end{array} \right] \quad (247) \\
g_m^{\bar{k}} &= \max_{j=1, \dots, n_g} \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \\
u_{L,m}^{\bar{k}} &= \max_{i=1, \dots, n_u} \left(u_i^L + \delta_e - u_{\bar{k},i} \right) \\
u_{U,m}^{\bar{k}} &= \max_{i=1, \dots, n_u} \left(u_{\bar{k},i} + \delta_e - u_i^U \right).
\end{aligned}$$

To close this subsection, we note that computing the maximum of g_j over $\mathcal{B}_{e,\bar{k}}$ may not be trivial if the numerical constraint g_j is not concave, and numerical global optimization methods may be required to obtain this value. Alternatively, one may also take an upper concave relaxation of the function and maximize this instead, which may solve numerical issues but may introduce more conservative back-offs.

6.3 Adding Back-offs to the Feasibility Conditions

In principle, the condition (194), together with (15), will guarantee feasibility without any additional modifications. However, the goal of the back-offs is to leave some feasible space specifically intended for perturbations – as already mentioned, the problem being solved once these are added is no longer (17) but (213). As such, we are not really interested in generating a chain of experiments that satisfy the constraints of (17) but violate the constraints of (213).

In the ideal case, we would expect the future experimental iterate \mathbf{u}_{k+1} to become the next reference \mathbf{u}_{k^*} , and so would like to be able to ensure that the appropriate back-offs are ensured for it in advance. In other words, we would like to be able to ensure, *a priori*, that

$$\begin{aligned}
\bar{g}_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) + b_{p,j}^{k+1,k+1} &\leq 0, \quad \forall j = 1, \dots, n_{g_p} \\
g_j(\mathbf{u}_{k+1}) + b_j^{k+1} &\leq 0, \quad \forall j = 1, \dots, n_g \\
\mathbf{u}^L + \mathbf{b}_u &\preceq \mathbf{u}_{k+1} \preceq \mathbf{u}^U - \mathbf{b}_u, \quad (248)
\end{aligned}$$

where we have simply made the substitutions $k^* \rightarrow k+1$ (assuming that $k+1$ becomes the next reference) and $k \rightarrow k+1$ (indicating a shift in indices) in (242). Substituting in the back-off expressions yields the one-iteration-ahead analogue of (243):

$$\begin{aligned}
& \bar{g}_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \\
& + \eta_{c,j}^{e,k+1} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k+1}, \tau_{k+1})} (\tau_{k+2} - \tau_{k+1}) \\
& + (1 - \eta_{c,j}^{e,k+1}) \bar{\kappa}_{p,j\tau}^{e,k+1} (\tau_{k+2} - \tau_{k+1}) \\
& + \delta_e \|\kappa_{p,j}^{m,k+1}\|_2 \leq 0, \quad \forall j = 1, \dots, n_{g_p},
\end{aligned} \tag{249}$$

$$\max_{\mathbf{u} \in \mathcal{B}_{e,k+1}} g_j(\mathbf{u}) \leq 0, \quad \forall j = 1, \dots, n_g, \tag{250}$$

$$\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{k+1} \preceq \mathbf{u}^U - \delta_e \mathbf{1}. \tag{251}$$

Let us consider (250) and (251), as these are relatively simple. For (251), we note that one may ensure these constraints by simply modifying the constraints in the projection (shown later). For (250), one may modify Condition (15) to

$$\max_{\mathbf{u} \in \mathcal{B}_K} g_j(\mathbf{u}) \leq 0, \quad \forall j = 1, \dots, n_g, \tag{252}$$

where

$$\mathcal{B}_K = \{\mathbf{u} : \|\mathbf{u} - \mathbf{u}_{k^*} - K_k(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*})\|_2 \leq \delta_e\}. \tag{253}$$

It is important to note that the computational burden of the line search, considered up to now to be negligible, may increase significantly depending on the computational burden of evaluating $\max_{\mathbf{u} \in \mathcal{B}_K} g_j(\mathbf{u})$, as this computation would now have to be done for every K_k considered in the line search. Again, it may be of practical interest to upper bound g_j by a simpler function for which the computation is cheap if this issue arises.

Condition (249) is a bit more difficult to grasp. First, note that while enforcing (252) automatically enforces (15), enforcing (249) *does not imply* that the standard feasibility condition $g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \leq 0$ is satisfied. In fact, all that it implies is that $g_{p,j}(\mathbf{u}, \tau_{k+2}) \leq 0$, $\forall \mathbf{u} \in \mathcal{B}_K$ and, in turn, that $g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+2}) \leq 0$, which may be possible to satisfy while allowing a violation at $(\mathbf{u}_{k+1}, \tau_{k+1})$ due to degradation effects. As such, (249) is an additional condition that should be added to the line search, and not a replacement.

In handling (249), note that the bound $\bar{g}_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1})$ may be obtained as the left-hand side of (194), with

$$\begin{aligned}
& \bar{g}_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) = \\
& \left[\begin{aligned}
& \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\
& + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\
& + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned}
& \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\
& K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\
& \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\
& K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i})
\end{aligned} \right] \\
& + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned}
& \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\
& K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\
& \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\
& K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i})
\end{aligned} \right]
\end{aligned} \right]. \tag{254}
\end{aligned}$$

The portion of (249) that corresponds to the back-off is more troublesome, however, as $\eta_{c,j}^{e,k+1}$, $\frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k+1}, \tau_{k+1})}$, $\bar{\kappa}_{p,j\tau}^{e,k+1}$, and $\kappa_{p,j}^{m,k+1}$ are all functions of K_k and must be evaluated at each point in the line search. While the difficulty is not conceptual – in theory, one could obtain all of these values given a certain K_k – it is relevant in practice, since certain components require calling a gradient estimation algorithm and thus could lead to large computational burdens. Removing the concavity and locality relaxations would solve this issue, however, as the need to estimate gradients would be removed and the above-stated components would simply be constants.

The addition to (194) is then given as

$$\begin{aligned}
& \min_{\bar{k}=0, \dots, k} \left[\begin{aligned}
& \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\
& + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\
& + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned}
& \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\
& K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\
& \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\
& K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i})
\end{aligned} \right] \\
& + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned}
& \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\
& K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\
& \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\
& K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i})
\end{aligned} \right]
\end{aligned} \right] \tag{255} \\
& + \eta_{c,j}^K \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*+K_k(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*}), \tau_{k+1})} (\tau_{k+2} - \tau_{k+1}) \\
& + (1 - \eta_{c,j}^K) \bar{\kappa}_{p,j\tau}^K (\tau_{k+2} - \tau_{k+1}) + \delta_e \|\kappa_{p,j}^{m,K}\|_2 \leq 0, \\
& \quad \forall j = 1, \dots, n_{g_p},
\end{aligned}$$

where, defining

$$\mathcal{I}_\tau^K = \mathcal{B}_K \times \{\tau : \tau_{k+1} \leq \tau \leq \tau_{k+2}\}, \quad (256)$$

we define $\eta_{c,j}^K$ and $\bar{\kappa}_{p,j\tau}^K$ as the analogues of $\eta_{c,j}^{e,k^*}$ and $\bar{\kappa}_{p,j\tau}^{e,k^*}$ over \mathcal{I}_τ^K (as opposed to \mathcal{I}_τ^{e,k^*}). Likewise, $\kappa_{p,j}^{m,K}$ is defined as

$$\kappa_{p,ji}^{m,K} = \begin{cases} \max \left[\left| \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*+K_k(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*}), \tau_{k+1})} \right|, \right. \\ \left. \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*+K_k(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*}), \tau_{k+1})} \right], & i \in I_{c,j}^K \\ \max \left[|\underline{\kappa}_{p,ji}^K|, |\bar{\kappa}_{p,ji}^K| \right], & i \notin I_{c,j}^K \end{cases} \quad (257)$$

$$\kappa_{p,j}^{m,K} = \left[\kappa_{p,j1}^{m,K} \dots \kappa_{p,jn_u}^{m,K} \right]^T,$$

with $I_{c,j}^K$ denoting the concavity indices over \mathcal{I}_τ^K and $\underline{\kappa}_{p,ji}^K, \bar{\kappa}_{p,ji}^K$ defined as

$$\underline{\kappa}_{p,ji}^K \leq \kappa_{p,ji}^K < \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}, \tau)} < \bar{\kappa}_{p,ji}^K \leq \bar{\kappa}_{p,ji}^K, \quad \forall (\mathbf{u}, \tau) \in \mathcal{I}_\tau^K. \quad (258)$$

Note that an additional bit of information, in the form of τ_{k+2} , is required in this implementation, as the user should specify not only the time of the upcoming experiment at τ_{k+1} but also the time of the one after.

Practically, it may occur that the degradation term

$$\eta_{c,j}^K \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*+K_k(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*}), \tau_{k+1})} (\tau_{k+2} - \tau_{k+1}) \\ + (1 - \eta_{c,j}^K) \bar{\kappa}_{p,j\tau}^K (\tau_{k+2} - \tau_{k+1})$$

render the condition (255) infeasible. In this case, we suggest using an alternative to (255) that ignores this term:

$$\min_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k(\bar{\mathbf{u}}_{k+1}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k(\bar{\mathbf{u}}_{k+1}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K_k(\bar{\mathbf{u}}_{k+1}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K_k(\bar{\mathbf{u}}_{k+1}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \delta_e \|\kappa_{p,j}^{m,K}\|_2 \leq 0, \quad \forall j = 1, \dots, n_{g_p}. \end{aligned} \right] \quad (259)$$

6.4 Adding Back-offs in the Projection

Since the feasibility conditions have essentially been offset by the back-offs, it makes sense that the projection conditions of the SCFO are offset as well to avoid premature convergence to a backed off constraint. As already mentioned, the bound constraints are also offset so as to ensure satisfaction of (251):

$$\begin{aligned} & \bar{\mathbf{u}}_{k+1}^* := \\ & \arg \underset{\mathbf{u}, \mathbf{s}_\phi, \mathbf{S}}{\text{minimize}} \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ & \text{subject to} \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_{p,j}} \\ & \quad \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji} \\ & \quad \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji}, \\ & \quad \forall i = 1, \dots, n_u, \\ & \quad \forall j : \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) + b_{p,j}^{k^*,k} \geq -\varepsilon_{p,j} \\ & \quad \nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j}, \\ & \quad \forall j : g_j(\mathbf{u}_{k^*}) + b_j^{k^*,k} \geq -\varepsilon_j \\ & \quad \sum_{i=1}^{n_u} s_{\phi,i} \leq -\delta_\phi \\ & \quad \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\ & \quad \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\ & \quad \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u} \preceq \mathbf{u}^U - \delta_e \mathbf{1}, \end{aligned} \quad (260)$$

or, with the expressions (220) and (235) substituted in for the back-offs:

$$\begin{aligned} & \bar{\mathbf{u}}_{k+1}^* := \\ & \arg \underset{\mathbf{u}, \mathbf{s}_\phi, \mathbf{S}}{\text{minimize}} \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\ & \text{subject to} \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_{p,j}} \\ & \quad \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji} \\ & \quad \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji}, \\ & \quad \forall i = 1, \dots, n_u, \\ & \quad \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1}) \\ & \quad + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\tau_{k+1} - \tau_{k^*}) \\ & \quad \forall j : \\ & \quad + (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ & \quad + \delta_e \|\kappa_{p,j}^{m,k^*}\|_2 \geq -\varepsilon_{p,j} \end{aligned} \quad (261)$$

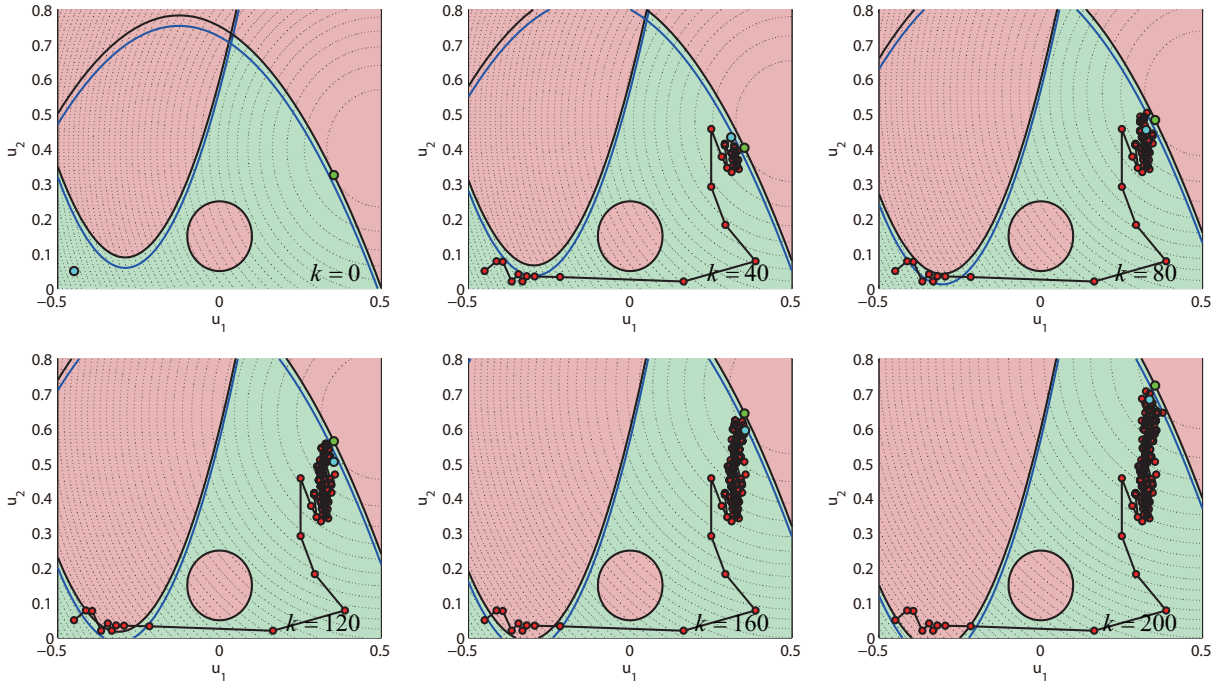


Fig. 24 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (–) scenario with a sufficient excitation requirement of $\delta_e = 0.02$ added.

$$\nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j},$$

$$\forall j: \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} g_j(\mathbf{u}) \geq -\varepsilon_j$$

$$\sum_{i=1}^{n_u} s_{\phi,i} \leq -\delta_\phi$$

$$\left. \frac{\partial \phi_p}{\partial u_i} \right|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i}$$

$$\left. \frac{\partial \bar{\phi}_p}{\partial u_i} \right|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i}$$

$$\mathbf{u}^L + \delta_e \mathbf{1} \leq \mathbf{u} \leq \mathbf{u}^U - \delta_e \mathbf{1}.$$

$$\begin{aligned} & \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} (-u_1^2 - (u_2 - 0.15)^2 + 0.01) \\ & \leq \max_{\mathbf{u} \in \text{bbox}(\mathbf{u}_{k^*} \pm \delta_e \mathbf{1})} (-u_1^2 - (u_2 - 0.15)^2 + 0.01) \\ & = - \min_{u_1 \in [u_1 - \delta_e, u_1 + \delta_e]} u_1^2 \\ & \quad - \min_{u_2 \in [u_2 - \delta_e, u_2 + \delta_e]} (u_2 - 0.15)^2 + 0.01, \end{aligned} \quad (262)$$

where the separate minimum terms may be easily solved analytically, the minimum occurring at either $u_i - \delta_e$ or $u_i + \delta_e$ or at the value that brings the term to 0, if this value belongs to $\text{bbox}(\mathbf{u}_{k^*} \pm \delta_e \mathbf{1})$. The same is done to easily compute the maximum over \mathcal{B}_K in the line search.

Sufficient excitation is enforced whenever the $\|\mathbf{u}_{k+1} - \mathbf{u}_{k^*}\|_2$ value for the \mathbf{u}_{k+1} found by the standard implementation is inferior to δ_e . In this example, we do this in a somewhat brute manner by simply picking a random unit vector $\delta \mathbf{u}_r$ and redefining \mathbf{u}_{k+1} as

$$\mathbf{u}_{k+1} := \mathbf{u}_{k^*} + \delta_e \delta \mathbf{u}_r, \quad (263)$$

more intelligent perturbation schemes of course being possible.

Results for $\delta_e = 0.02$ are provided in Fig. 24, which show that feasibility is maintained throughout despite the sufficient excitation requirement, which is shown to be enforced in Fig. 25. The cost function values are given in Fig. 26.

6.5 Example

We build on the $\alpha_\sigma = 0.05$ example of Section 5.8 by adding the requirement that feasible perturbations of norm δ_e always be possible, and incorporate the theory discussed in this section into the implementation.

Due to the concavity (or absence thereof) being global in the constraint functions, we simply set $I_{c,j}^{e,k^*} := I_{c,j}$, but set $I_{c,j}^K := \emptyset$ to simplify the line search and avoid gradient bound computations during the search. The local Lipschitz constants are computed in a manner analogous to (125)-(128). The maximum of g_1 over a ball may be upper bounded by its maximum over an inscribing box, for which the maximum may be easily evaluated due to the separability of g_1 :

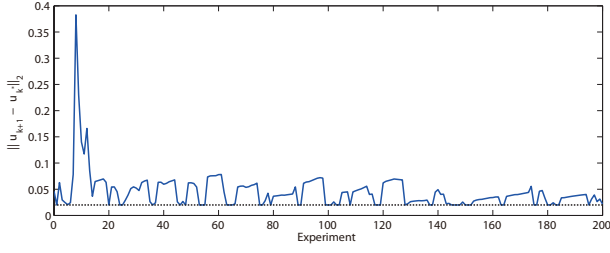


Fig. 25 Size of the steps taken between experimental iterations. The dashed black line gives the value of δ_ϵ .

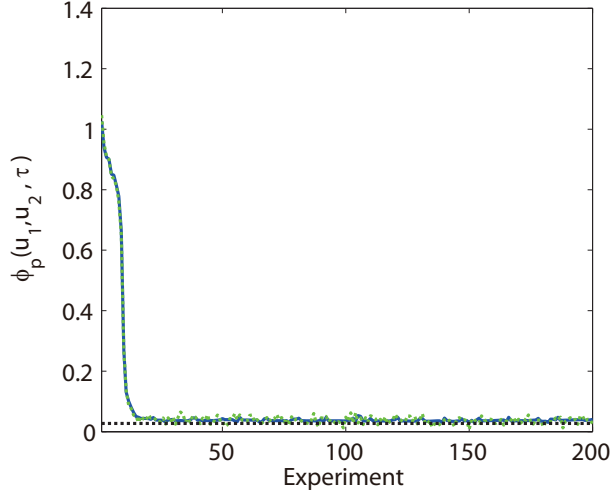


Fig. 26 Cost function values obtained by the modified SCFO methodology for Problem (42) for the (–) scenario with a sufficient excitation requirement of $\delta_\epsilon = 0.02$ enforced.

7 Accomodating Soft Constraints

Up to now, we have developed all of our theory for the case where the experimental and numerical constraints were considered inviolable, in that we had the requirement

$$\begin{aligned} g_{p,j}(\mathbf{u}_k, \tau_k) &\leq 0, \quad \forall j = 1, \dots, n_{g_p} \\ g_j(\mathbf{u}_k) &\leq 0, \quad \forall j = 1, \dots, n_g \end{aligned} \quad (264)$$

for all experimental iterations k . However, it may often occur that one can violate certain constraints temporarily during the convergence process. Supposing an allowable violation for each constraint and denoting this by $d \geq 0$, this is tantamount to saying that one may allow

$$\begin{aligned} g_{p,j}(\mathbf{u}_k, \tau_k) &\leq d_{p,j}, \quad \forall j = 1, \dots, n_{g_p} \\ g_j(\mathbf{u}_k) &\leq d_j, \quad \forall j = 1, \dots, n_g \end{aligned} \quad (265)$$

for some iterations.

Such a relaxation is of practical interest as it generally leads to faster convergence speed, the SCFO being allowed to take greater steps due to less stringent feasibility conditions. However, some care should be taken to manage

the slacks d , as one needs to ensure that the violations are, in some sense, bounded and do not last for an arbitrarily large number of experiments since they are constrained to be temporary by definition.

Here, we propose to ensure these properties by making the slacks iteration-dependent

$$\begin{aligned} g_{p,j}(\mathbf{u}_k, \tau_k) &\leq d_{p,j}^k, \quad \forall j = 1, \dots, n_{g_p} \\ g_j(\mathbf{u}_k) &\leq d_j^k, \quad \forall j = 1, \dots, n_g \end{aligned} \quad (266)$$

and managing them as follows:

$$\begin{aligned} d_{p,j}^0 &:= \bar{d}_{p,j}, \quad d_j^0 := \bar{d}_j \\ d_{p,j}^{k+1} &:= \begin{cases} \beta_{p,j} d_{p,j}^k, & \bar{g}_{p,j}(\mathbf{u}_k, \tau_k) > 0 \\ d_{p,j}^k, & \bar{g}_{p,j}(\mathbf{u}_k, \tau_k) \leq 0 \end{cases} \\ d_j^{k+1} &:= \begin{cases} \beta_j d_j^k, & g_j(\mathbf{u}_k) > 0 \\ d_j^k, & g_j(\mathbf{u}_k) \leq 0, \end{cases} \end{aligned} \quad (267)$$

where $\bar{d}_{p,j}$ and \bar{d}_j are the maximum allowable violations while $\beta_{p,j}, \beta_j \in [0, 1)$ are slack reduction constants. The temporary nature of the violations is then quantified by

$$\begin{aligned} \sum_{\bar{k}=0}^{\infty} \max[0, g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})] &\leq d_{p,j}^S \\ \sum_{\bar{k}=0}^{\infty} \max[0, g_j(\mathbf{u}_{\bar{k}})] &\leq d_j^S, \end{aligned} \quad (268)$$

where $d_{p,j}^S, d_j^S \geq 0$ are some user-specified constants and may be thought of as upper bounds on the “discrete violation integrals”.

We now derive sufficiently low values of $\beta_{p,j}, \beta_j$ to guarantee that the bounds (268) are satisfied.

Theorem 3 (Slack reduction constant for a given experimental constraint) *Let I_β denote the set of experimental iteration indices, ordered from smallest to largest, where the constraint cannot be proven to be satisfied:*

$$I_\beta = \{\bar{k} : \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) > 0\}. \quad (269)$$

If the update strategy (267) is applied with

$$\beta_{p,j} \leq \frac{d_{p,j}^S - \bar{d}_{p,j}}{d_{p,j}^S}, \quad (270)$$

then the bound for $g_{p,j}$ in (268) is satisfied.

Proof Let us rewrite the violation integral as

$$\begin{aligned} \sum_{\bar{k}=0}^{\infty} \max[0, g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})] \\ = \sum_{i=0}^{\#I_\beta-1} \max[0, g_{p,j}(\mathbf{u}_{I_\beta(i)}, \tau_{I_\beta(i)})], \end{aligned} \quad (271)$$

$$\begin{aligned}
\min_{\bar{k}=0,\dots,k} & \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \delta_e \|\kappa_{p,j}^{m,K}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p}. \end{aligned} \right] \quad (279)
\end{aligned}$$

7.2 Accounting for Slack in the Choice of Reference Point

Similar modifications are made to (244) and (245):

$$\begin{aligned}
& k^* := \\
& \arg \text{maximize } \bar{k} \\
& \quad \bar{k} \in [0, k] \\
& \text{subject to } \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\
& \quad + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\
& \quad + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\
& \quad + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p} \\
& \quad \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \leq d_j^{k+1}, \quad \forall j = 1, \dots, n_g \quad (280)
\end{aligned}$$

$$\begin{aligned}
& \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{\bar{k}} \preceq \mathbf{u}^U - \delta_e \mathbf{1} \\
& \phi_{-p}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k},k} \frac{\partial \phi_{-p}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\
& \quad + (1 - \eta_{v,\phi}^{\bar{k},k}) \bar{\kappa}_{\phi,\tau}^{\bar{k},k} (\tau_k - \tau_{\bar{k}}) \leq \\
& \quad \min_{\bar{k} \in \mathbf{k}_f} \left[\begin{aligned} & \bar{\phi}_{-p}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \\ & \eta_{c,\phi}^{\bar{k},k} \frac{\partial \bar{\phi}_{-p}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,\phi}^{\bar{k},k}) \bar{\kappa}_{\phi,\tau}^{\bar{k},k} (\tau_k - \tau_{\bar{k}}) \end{aligned} \right], \quad (281) \\
& \mathbf{k}_f = \left\{ \bar{k} : \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p}; \\ & \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \leq d_j^{k+1}, \quad \forall j = 1, \dots, n_g; \\ & \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{\bar{k}} \preceq \mathbf{u}^U - \delta_e \mathbf{1} \end{aligned} \right] \right\}, \quad (281)
\end{aligned}$$

as this would yield a \mathbf{u}_{k^*} that would guarantee the existence of a $K_k \geq 0$ that satisfies, at least, the relaxed condition of (279).

In the case that no K_k satisfying these restrictions can be found, one may proceed with (246) with the following modifications made to (247):

$$\begin{aligned}
g_{p,m}^{\bar{k}} &= \max_{j=1,\dots,n_{g_p}} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ & + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 - d_{p,j}^{k+1} \end{aligned} \right] \quad (282) \\
g_m^{\bar{k}} &= \max_{j=1,\dots,n_g} \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} (g_j(\mathbf{u}) - d_j^{k+1}) \\
u_{L,m}^{\bar{k}} &= \max_{i=1,\dots,n_u} (u_i^L + \delta_e - u_{\bar{k},i}) \\
u_{U,m}^{\bar{k}} &= \max_{i=1,\dots,n_u} (u_{\bar{k},i} + \delta_e - u_i^U).
\end{aligned}$$

7.3 Accounting for Slack in the Projection

Finally, we make the same modification to the Boolean triggers in the projection (261):

$$\begin{aligned}
& \bar{\mathbf{u}}_{k+1}^* := \\
& \arg \text{minimize } \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\
& \quad \mathbf{u}, s_{\phi}, \mathbf{S} \\
& \text{subject to } \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_p,j} \\
& \quad \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji} \\
& \quad \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji}, \\
& \quad \forall i = 1, \dots, n_u, \\
& \quad \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\
& \quad + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\
& \quad + (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\
& \quad + \delta_e \|\kappa_{p,j}^{m,k^*}\|_2 \geq -\varepsilon_{p,j} + d_{p,j}^{k+1} \\
& \quad \nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j}, \\
& \quad \forall j : \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} g_j(\mathbf{u}) \geq -\varepsilon_j + d_j^{k+1} \\
& \quad \sum_{i=1}^{n_u} s_{\phi,i} \leq -\delta_{\phi} \\
& \quad \frac{\partial \phi_{-p}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\
& \quad \frac{\partial \bar{\phi}_{-p}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{\phi,i} \\
& \quad \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u} \preceq \mathbf{u}^U - \delta_e \mathbf{1}, \quad (283)
\end{aligned}$$

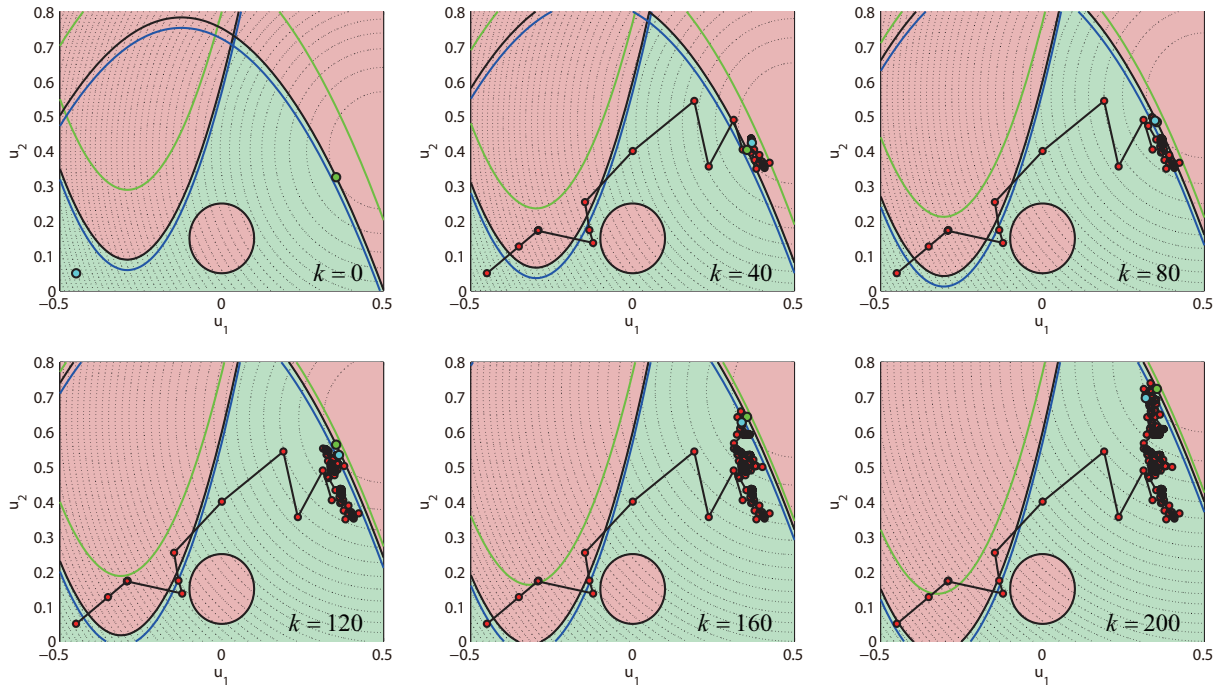


Fig. 27 Chain of experiments generated by applying the modified SCFO methodology to Problem (42) for the (-) scenario with slacks added to the experimental constraints (shown in green).

which simply “delays” the projection into the locally feasible space for those constraints that are relaxed.

7.4 Example

We build further on the previous example of Section 6.5 by adding the slacks

$$\begin{aligned} \bar{d}_{p,1} &= \bar{d}_{p,2} = 0.2 \\ d_{p,1}^S &= 5, \quad d_{p,2}^S = 10, \end{aligned} \quad (284)$$

to the experimental constraints only, and apply the proposed modified SCFO with $\beta_{p,j}$ taken as the maximum value given by the bound (270).

The results are shown in Fig. 27, with the corresponding cost function values given in Fig. 28. As expected, we see two key benefits of allowing temporary violations – the first being that it is possible to converge to the optimum faster, and the second that it is also possible to temporarily obtain a cost that is better than the one at the optimum, thus leading to greater overall gains. Fig. 29 shows that the sums of the violations are well below⁶ the limits set by $d_{p,j}^S$.

⁶ While we do not explore this option here, it should be clear that one could propose more elegant slack reduction techniques that would likely lead to even better performance.

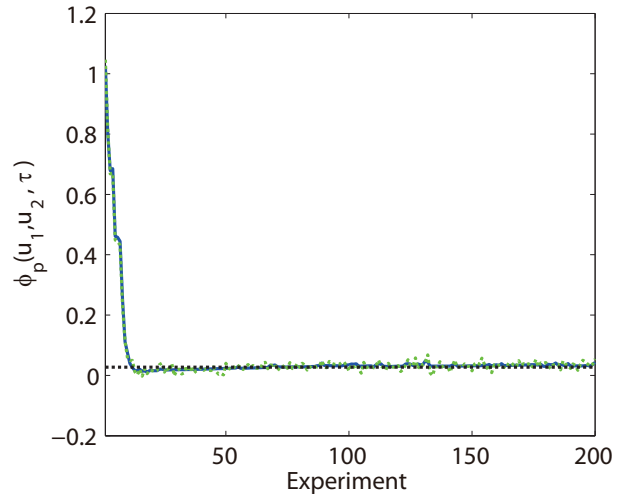


Fig. 28 Cost function values obtained by the modified SCFO methodology for Problem (42) for the (-) scenario when slacks are added to the experimental constraints.

8 Accomodating a Numerical Cost Function

In some experimental optimization problems it may occur that the cost function is numerical and can be easily evaluated without requiring experiments – see, e.g., the second application example in [7]. When this is so, we may simplify the SCFO considerably by treating ϕ (a known function) instead of ϕ_p (an unknown function),

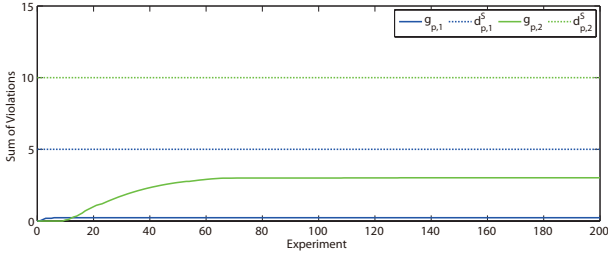


Fig. 29 The values of the sums of experimental constraint violations over the course of optimization.

and exploiting the fact that derivative information on the function is available, that there is no degradation, and that the Lipschitz constants for the cost are no longer required for the desired properties to be enforced.

Most notably, the line search may be modified from finding the largest K_k that satisfies the SCFO to finding the K_k that minimizes the cost while satisfying (276)-(278):

$$K_k := \arg \underset{K \in [0,1]}{\text{minimize}} \phi(\mathbf{u}_{k^*} + K(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*})) \quad (285)$$

subject to (276)–(278).

Because this line search is guaranteed to yield a cost decrease if one exists – unlike the line search incorporating (197), which can fail to yield an actual decrease in the case of erroneous Lipschitz constants – the necessary condition for cost decrease (201) is not needed.

In choosing the reference point as in (280) and (281), one no longer needs to use the heuristic of taking the latest feasible experimental iterate that is guaranteed to not have a cost value that is worse than that of another feasible experimental iterate. Instead, one may simply choose the feasible iterate with the lowest cost value, i.e.:

$$k^* := \arg \underset{\bar{k} \in [0,k]}{\text{minimize}} \phi(\mathbf{u}_{\bar{k}})$$

subject to $\bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$

$$+ \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \quad (286)$$

$$+ (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}})$$

$$+ \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p}$$

$$\max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \leq d_j^{k+1}, \quad \forall j = 1, \dots, n_g$$

$$\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{\bar{k}} \preceq \mathbf{u}^U - \delta_e \mathbf{1}.$$

Finally, the projection (283) may be simplified to

$$\bar{\mathbf{u}}_{k+1}^* := \arg \underset{\mathbf{u}, \mathbf{S}}{\text{minimize}} \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2$$

subject to $\sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_p,j}$

$$\frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji}$$

$$\frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji},$$

$$\forall i = 1, \dots, n_u,$$

$$\bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \quad (287)$$

$$+ \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*})$$

$$\forall j: \quad + (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*})$$

$$+ \delta_e \|\kappa_{p,j}^{m,k^*}\|_2 \geq -\varepsilon_{p,j} + d_{p,j}^{k+1}$$

$$\nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j},$$

$$\forall j: \quad \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} g_j(\mathbf{u}) \geq -\varepsilon_j + d_j^{k+1}$$

$$\nabla \phi(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_\phi$$

$$\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u} \preceq \mathbf{u}^U - \delta_e \mathbf{1}.$$

9 Complete Implementable Form of the SCFO

In this section, we will simply summarize the results derived so far to give the SCFO in their ready-to-code and fully implementable form, where by “fully implementable” we mean that all of the different issues raised in this document are simultaneously accounted for. Additionally, we will restate all presented algorithms fully so as to include the proposed modifications – some of which came several sections later in the text after the algorithm was first stated – directly inside the algorithm statement. This is done to ease the task of coding for those readers who are intent on writing their own SCFO implementation or modifying the one currently available [7].

We start by noting that the practical SCFO implementation may be seen as consisting of an initialization step followed by six iterative steps, the latter being carried out at every experimental iteration:

1. (Initialization – carried out once prior to any optimization) One defines all the elements of the experimental optimization problem that one expects to remain constant (i.e., $\bar{\varepsilon}_{p,j}$, $\bar{\varepsilon}_j$, $\bar{\delta}_{g_p,j}$, $\bar{\delta}_{g,j}$, $\bar{\delta}_\phi$, $\bar{\kappa}$, \bar{M} , \bar{M} , I_c , I_v , η , \bar{W} , \bar{W} , δ_e , \bar{d} , d^S).
2. (Pretreatment of collected data) Using the measurements obtained up until the current iteration, one runs the Lipschitz consistency check algorithm and then uses the consistent Lipschitz constants to compute lower and

upper bounds on the measured experimental function values.

3. (Choice of reference point) The reference point \mathbf{u}_{k^*} is chosen as the point that meets the feasibility requirements while being likely to have the best cost function value obtained in all experiments carried out so far.
4. (Computation of the optimal target) The target \mathbf{u}_{k+1}^* is found by some prescribed optimization method – see [10] for a review of such methods.
5. (Projection of optimal target) The target \mathbf{u}_{k+1}^* is projected to obtain the projected target $\bar{\mathbf{u}}_{k+1}^*$ satisfying a portion of the SCFO in some robust sense.
6. (Filtering of adaptation step) The filter gain K_k is found by a line search to define \mathbf{u}_{k+1} as a function of \mathbf{u}_{k^*} , $\bar{\mathbf{u}}_{k+1}^*$, and K_k . If the excitation obtained by changing the decision variables is insufficient, this choice may be overridden by a sufficiently exciting \mathbf{u}_{k+1} .
7. (Experimental application) An experiment defined by the computed \mathbf{u}_{k+1} is carried out, new measurements are obtained, and the cycle restarts with Step 2.

Note that this scheme does not have a termination criterion. This is intentional and reflects the philosophy discussed by, for example, Box and Draper in [2], in that in experimental optimization there is little reason to stop continuous improvement, especially when effects like degradation may continually force the algorithm to pursue a moving optimum. From an algorithmic viewpoint, it also does not make sense to terminate the SCFO adaptations since these adaptations are designed to obtain better and better performance until this becomes impossible. Of course, one may nevertheless propose certain termination criteria when the current level of optimality is known to be sufficient, and the most recent version of the SCFO solver accomodates this [7] – we do not, however, address this here.

The rest of this section is devoted to going through Steps 2, 3, 5, and 6 and giving the relevant formulas used in the different subroutines of each in their finalized, complete forms.

9.1 Pretreatment of Measured Data

In the pretreatment of the data collected, one has the tasks of (a) verifying that the Lipschitz constants provided are consistent with the data and, if not, making the constants consistent, and (b) computing lower and upper bounds on the true experimental function values. As mentioned previously, the two tasks are convoluted since the Lipschitz consistency check uses the computed lower and upper bounds and the lower and upper bounds may in turn be a function of the Lipschitz constants. To deconvolute the two, we make it so that the bounds used

in the Lipschitz consistency check are independent of the Lipschitz constants, i.e., we use the bounds given in (134):

$$\begin{aligned}\hat{\phi}_{\bar{p}}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &= \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \bar{w}_{\phi, \bar{k}} \\ \bar{\hat{\phi}}_{\bar{p}}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) &= \hat{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \underline{w}_{\phi, \bar{k}},\end{aligned}\quad (288)$$

where the double lines denote that these are more conservative bounds than what would be obtained if we were to run Algorithm 2 with the correct Lipschitz constants.

We now state the finalized version of Algorithm 1.

Algorithm 1F – Lipschitz Consistency Check (Finalized)

1. Set $a := 1$.
2. Check the validity of

$$\begin{aligned}\hat{\phi}_{\bar{p}}(\mathbf{u}_{k_2}, \tau_{k_2}) &\leq \bar{\hat{\phi}}_{\bar{p}}(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &+ \max \begin{bmatrix} \underline{K}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1}) \\ \bar{K}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1}) \end{bmatrix} \\ &+ \sum_{i=1}^{n_u} \max \begin{bmatrix} \underline{K}_{\phi, i}(u_{k_2, i} - u_{k_1, i}) \\ \bar{K}_{\phi, i}(u_{k_2, i} - u_{k_1, i}) \end{bmatrix}\end{aligned}\quad (289)$$

and

$$\begin{aligned}\bar{\hat{\phi}}_{\bar{p}}(\mathbf{u}_{k_2}, \tau_{k_2}) &\geq \hat{\phi}_{\bar{p}}(\mathbf{u}_{k_1}, \tau_{k_1}) \\ &+ \min \begin{bmatrix} \underline{K}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1}) \\ \bar{K}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1}) \end{bmatrix} \\ &+ \sum_{i=1}^{n_u} \min \begin{bmatrix} \underline{K}_{\phi, i}(u_{k_2, i} - u_{k_1, i}) \\ \bar{K}_{\phi, i}(u_{k_2, i} - u_{k_1, i}) \end{bmatrix}\end{aligned}\quad (290)$$

for every combination $(k_1, k_2) \in \{0, \dots, k\} \times \{0, \dots, k\}$. If these inequalities are satisfied for every (k_1, k_2) , then terminate. Otherwise, proceed to Step 3.

3. If $a \leq 5$, then for each $i = 1, \dots, n_u$ set

$$\begin{aligned}\underline{K}_{\phi, i} &:= 2^{-\text{sign } \underline{K}_{\phi, i}} \underline{K}_{\phi, i}, \\ \bar{K}_{\phi, i} &:= 2^{\text{sign } \bar{K}_{\phi, i}} \bar{K}_{\phi, i}, \\ \underline{K}_{\phi, \tau} &:= 2^{-\text{sign } \underline{K}_{\phi, \tau}} \underline{K}_{\phi, \tau}, \\ \bar{K}_{\phi, \tau} &:= 2^{\text{sign } \bar{K}_{\phi, \tau}} \bar{K}_{\phi, \tau}.\end{aligned}$$

If $5 < a \leq 10$, set

$$\begin{aligned}\underline{K}_{\phi, i} &:= -2 \max [|\underline{K}_{\phi, i}|, |\bar{K}_{\phi, i}|], \\ \bar{K}_{\phi, i} &:= 2 \max [|\underline{K}_{\phi, i}|, |\bar{K}_{\phi, i}|], \\ \underline{K}_{\phi, \tau} &:= -2 \max [|\underline{K}_{\phi, \tau}|, |\bar{K}_{\phi, \tau}|], \\ \bar{K}_{\phi, \tau} &:= 2 \max [|\underline{K}_{\phi, \tau}|, |\bar{K}_{\phi, \tau}|].\end{aligned}$$

If $10 < a$, set

$$\begin{aligned}\underline{K}_{\phi, i} &:= 2^{a-10} \underline{K}_{\phi, i}, \\ \bar{K}_{\phi, i} &:= 2^{a-10} \bar{K}_{\phi, i}, \\ \underline{K}_{\phi, \tau} &:= 2^{a-10} \underline{K}_{\phi, \tau}, \\ \bar{K}_{\phi, \tau} &:= 2^{a-10} \bar{K}_{\phi, \tau}.\end{aligned}$$

4. Set $a := a + 1$ and return to Step 2.

For the sake of completeness, let us also state the version for the higher-order Lipschitz constants M .

Algorithm 1FA – Higher-Order Lipschitz Consistency Check (Finalized)

1. Set $a := 1$.
2. Check the validity of

$$\begin{aligned} \hat{\phi}_{\underline{p}}(\mathbf{u}_{k_2}, \tau_{k_2}) \leq & \bar{\phi}_{\underline{p}}(\mathbf{u}_{k_1}, \tau_{k_1}) + \max \left[\frac{\underline{\kappa}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1})}, \right. \\ & \left. + \sum_{i=1}^{n_u} \max \left[\frac{\partial \hat{\phi}_{\underline{p}}}{\partial u_i} \Big|_{(\mathbf{u}_{k_1}, \tau_{k_1})} (u_{k_2, i} - u_{k_1, i}), \right. \right. \\ & \left. \left. \frac{\partial \bar{\phi}_{\underline{p}}}{\partial u_i} \Big|_{(\mathbf{u}_{k_1}, \tau_{k_1})} (u_{k_2, i} - u_{k_1, i}) \right] \right. \\ & \left. + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \left[\frac{\underline{M}_{\phi, i_1 i_2} (u_{k_2, i_1} - u_{k_1, i_1})}{\bar{M}_{\phi, i_1 i_2} (u_{k_2, i_1} - u_{k_1, i_1})}, \right. \right. \\ & \left. \left. \frac{(u_{k_2, i_2} - u_{k_1, i_2})}{(u_{k_2, i_2} - u_{k_1, i_2})} \right] \right] \end{aligned} \quad (291)$$

and

$$\begin{aligned} \hat{\phi}_{\bar{p}}(\mathbf{u}_{k_2}, \tau_{k_2}) \geq & \bar{\phi}_{\bar{p}}(\mathbf{u}_{k_1}, \tau_{k_1}) + \min \left[\frac{\underline{\kappa}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1})}{\bar{\kappa}_{\phi, \tau}(\tau_{k_2} - \tau_{k_1})}, \right. \\ & \left. + \sum_{i=1}^{n_u} \min \left[\frac{\partial \hat{\phi}_{\bar{p}}}{\partial u_i} \Big|_{(\mathbf{u}_{k_1}, \tau_{k_1})} (u_{k_2, i} - u_{k_1, i}), \right. \right. \\ & \left. \left. \frac{\partial \bar{\phi}_{\bar{p}}}{\partial u_i} \Big|_{(\mathbf{u}_{k_1}, \tau_{k_1})} (u_{k_2, i} - u_{k_1, i}) \right] \right. \\ & \left. + \frac{1}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \min \left[\frac{\underline{M}_{\phi, i_1 i_2} (u_{k_2, i_1} - u_{k_1, i_1})}{\bar{M}_{\phi, i_1 i_2} (u_{k_2, i_1} - u_{k_1, i_1})}, \right. \right. \\ & \left. \left. \frac{(u_{k_2, i_2} - u_{k_1, i_2})}{(u_{k_2, i_2} - u_{k_1, i_2})} \right] \right] \end{aligned} \quad (292)$$

for every combination $(k_1, k_2) \in \{0, \dots, k\} \times \{0, \dots, k\}$. If these inequalities are satisfied for every (k_1, k_2) , then terminate. Otherwise, proceed to Step 3.

3. If $a \leq 5$, then, for each $i_1 = 1, \dots, n_u$, $i_2 = 1, \dots, n_u$, set

$$\begin{aligned} \underline{M}_{\phi, i_1 i_2} &:= 2^{-\text{sign} \underline{M}_{\phi, i_1 i_2}} \underline{M}_{\phi, i_1 i_2}, \\ \bar{M}_{\phi, i_1 i_2} &:= 2^{\text{sign} \bar{M}_{\phi, i_1 i_2}} \bar{M}_{\phi, i_1 i_2}. \end{aligned}$$

If $5 < a \leq 10$, set

$$\begin{aligned} \underline{M}_{\phi, i_1 i_2} &:= -2 \max [|\underline{M}_{\phi, i_1 i_2}|, |\bar{M}_{\phi, i_1 i_2}|], \\ \bar{M}_{\phi, i_1 i_2} &:= 2 \max [|\underline{M}_{\phi, i_1 i_2}|, |\bar{M}_{\phi, i_1 i_2}|]. \end{aligned}$$

If $10 < a$, set

$$\begin{aligned} \underline{M}_{\phi, i_1 i_2} &:= 2^{a-10} \underline{M}_{\phi, i_1 i_2}, \\ \bar{M}_{\phi, i_1 i_2} &:= 2^{a-10} \bar{M}_{\phi, i_1 i_2}. \end{aligned}$$

4. Set $a := a + 1$ and return to Step 2.

To obtain consistent Lipschitz constants for the experimental constraints, one applies Algorithm 1F to the functions $g_{p,j}$ as well.

Following this treatment of the Lipschitz constants, we may now apply the finalized version of Algorithm 2.

Algorithm 2F – Lower and Upper Bounds for Experimental Function Values (Finalized)

1. Set $\bar{k} := 0$. Choose $\Delta_r, \underline{\Delta}_r \in \mathbb{R}_{++}$ such that $\Delta_r > \underline{\Delta}_r$.
2. If $\bar{k} > k$, go to Step 6. Otherwise, proceed to Step 3.
3. Let \bar{N} denote the number of iterates with the decision variable values $\mathbf{u}_{\bar{k}}$. Set $\hat{\phi}_{\underline{p}}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := -\infty$ and $\hat{\phi}_{\bar{p}}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := \infty$.
4. For $N = 1, \dots, \bar{N}$:

- (a) Generate the $\binom{\bar{N}}{N}$ index sets that correspond to different combinations of iterations with the same decision variable values. For each set:
 - i. Define $\bar{\mathbf{k}}$ as the corresponding index set and compute the candidate lower and upper bounds as

$$\begin{aligned} \hat{\phi}_{\underline{p}, test} &:= \frac{1}{N} \sum_{\bar{\mathbf{k}} \in \bar{\mathbf{k}}} \hat{\phi}_{\underline{p}}(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}}) \\ &\quad - \frac{1}{N} \sum_{\bar{\mathbf{k}} \in \bar{\mathbf{k}}} \eta_{c, \phi}^{\bar{\mathbf{k}}, \bar{\mathbf{k}}} \max \left[\frac{\partial \hat{\phi}_{\underline{p}}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}})} (\tau_{\bar{\mathbf{k}}} - \tau_{\bar{\mathbf{k}}}), \right. \\ &\quad \left. \frac{\partial \bar{\phi}_{\underline{p}}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}})} (\tau_{\bar{\mathbf{k}}} - \tau_{\bar{\mathbf{k}}}) \right] \\ &\quad - \frac{1}{N} \sum_{\bar{\mathbf{k}} \in \bar{\mathbf{k}}} (1 - \eta_{c, \phi}^{\bar{\mathbf{k}}, \bar{\mathbf{k}}}) \max \left[\frac{\underline{\kappa}_{\phi, \tau}^{\bar{\mathbf{k}}, \bar{\mathbf{k}}} (\tau_{\bar{\mathbf{k}}} - \tau_{\bar{\mathbf{k}}})}{\bar{\kappa}_{\phi, \tau}^{\bar{\mathbf{k}}, \bar{\mathbf{k}}} (\tau_{\bar{\mathbf{k}}} - \tau_{\bar{\mathbf{k}}})}, \right. \\ &\quad \left. - \bar{W}_{\phi, \bar{\mathbf{k}}}, \right] \end{aligned} \quad (293)$$

$$\begin{aligned} \hat{\phi}_{\bar{p}, test} &:= \frac{1}{N} \sum_{\bar{\mathbf{k}} \in \bar{\mathbf{k}}} \hat{\phi}_{\bar{p}}(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}}) \\ &\quad - \frac{1}{N} \sum_{\bar{\mathbf{k}} \in \bar{\mathbf{k}}} \eta_{v, \phi}^{\bar{\mathbf{k}}, \bar{\mathbf{k}}} \min \left[\frac{\partial \hat{\phi}_{\bar{p}}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}})} (\tau_{\bar{\mathbf{k}}} - \tau_{\bar{\mathbf{k}}}), \right. \\ &\quad \left. \frac{\partial \bar{\phi}_{\bar{p}}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}})} (\tau_{\bar{\mathbf{k}}} - \tau_{\bar{\mathbf{k}}}) \right] \\ &\quad - \frac{1}{N} \sum_{\bar{\mathbf{k}} \in \bar{\mathbf{k}}} (1 - \eta_{v, \phi}^{\bar{\mathbf{k}}, \bar{\mathbf{k}}}) \min \left[\frac{\underline{\kappa}_{\phi, \tau}^{\bar{\mathbf{k}}, \bar{\mathbf{k}}} (\tau_{\bar{\mathbf{k}}} - \tau_{\bar{\mathbf{k}}})}{\bar{\kappa}_{\phi, \tau}^{\bar{\mathbf{k}}, \bar{\mathbf{k}}} (\tau_{\bar{\mathbf{k}}} - \tau_{\bar{\mathbf{k}}})}, \right. \\ &\quad \left. - \underline{W}_{\phi, \bar{\mathbf{k}}}. \right] \end{aligned} \quad (294)$$

- ii. If $\hat{\phi}_{\underline{p}, test} > \hat{\phi}_{\underline{p}}(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}})$, set $\hat{\phi}_{\underline{p}}(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}}) := \hat{\phi}_{\underline{p}, test}$.
If $\hat{\phi}_{\bar{p}, test} < \hat{\phi}_{\bar{p}}(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}})$, set $\hat{\phi}_{\bar{p}}(\mathbf{u}_{\bar{\mathbf{k}}}, \tau_{\bar{\mathbf{k}}}) := \hat{\phi}_{\bar{p}, test}$.
5. Set $\bar{k} := \bar{k} + 1$ and return to Step 2.
6. If $\Delta_r \leq \underline{\Delta}_r$, terminate. Otherwise, set $\Delta_r := 0$, $\bar{k} := 0$ and go to Step 7.

7. If $\bar{k} > k$, return to Step 6. Otherwise, proceed to Step 8.
 8. For $\bar{k} := \{0, \dots, k\} \setminus \bar{k}$:

- (a) Compute the lower and upper bound candidate values as

$$\begin{aligned} \bar{\phi}_{p,test} &:= \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ &+ \eta_{c,\phi}^{\bar{k},\bar{k}} \max \left[\begin{array}{l} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{array} \right] \\ &+ (1 - \eta_{c,\phi}^{\bar{k},\bar{k}}) \max \left[\begin{array}{l} \underline{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ \bar{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{array} \right] \\ &+ \sum_{i \in I_{c,\phi}^{\bar{k},\bar{k}}} \max \left[\begin{array}{l} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}) \end{array} \right] \\ &+ \sum_{i \notin I_{c,\phi}^{\bar{k},\bar{k}}} \max \left[\begin{array}{l} \underline{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}) \end{array} \right], \end{aligned} \quad (295)$$

$$\begin{aligned} \underline{\phi}_{p,test} &:= \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ &+ \eta_{v,\phi}^{\bar{k},\bar{k}} \min \left[\begin{array}{l} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{array} \right] \\ &+ (1 - \eta_{v,\phi}^{\bar{k},\bar{k}}) \min \left[\begin{array}{l} \underline{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}), \\ \bar{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_{\bar{k}} - \tau_{\bar{k}}) \end{array} \right] \\ &+ \sum_{i \in I_{v,\phi}^{\bar{k},\bar{k}}} \min \left[\begin{array}{l} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{\bar{k},i} - u_{\bar{k},i}) \end{array} \right] \\ &+ \sum_{i \notin I_{v,\phi}^{\bar{k},\bar{k}}} \min \left[\begin{array}{l} \underline{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k},\bar{k}} (u_{\bar{k},i} - u_{\bar{k},i}) \end{array} \right]. \end{aligned} \quad (296)$$

- (b) Set

$$\Delta_r := \max \left[\begin{array}{l} \Delta_r, \underline{\phi}_{p,test} - \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \\ \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) - \bar{\phi}_{p,test} \end{array} \right]. \quad (297)$$

- (c) If $\underline{\phi}_{p,test} > \underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, set $\underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := \underline{\phi}_{p,test}$. If $\bar{\phi}_{p,test} < \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$, set $\bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) := \bar{\phi}_{p,test}$.

9. Set $\bar{k} := \bar{k} + 1$ and return to Step 7.

The same algorithm is then carried out for the experimental constraint functions $g_{p,j}$.

It is important to note the slight discrepancy between Algorithms 1F and 2F, in that the latter uses local Lipschitz

constants while the former uses global ones. To have compatibility between the two, it would be necessary either for both to use local or for both to use global. How this compatibility is established is left up to the user. For example, one may simply choose to work with global constants throughout since local relaxations may not be reliable in many cases, in which case one would use, e.g., $\underline{\kappa}_{\phi,i}^{\bar{k},\bar{k}} = \underline{\kappa}_{\phi,i}$ in Algorithm 2F. Alternatively, one may modify the consistency check of Algorithm 1F to be local in nature, although this may be a bit complex and so we do not treat the modification here. Finally, one may dispense with compatibility and simply use the global constants in Algorithm 1F and the local ones in Algorithm 2F – in this case, no sort of consistency check would be carried out for the local constants, which would most likely be obtained from a model.

9.2 Choosing the Reference Point

With consistent Lipschitz constants obtained and the experimental function values bounded, we now proceed to choose the point to use as the reference for optimization. In the case where the cost function is experimental in nature, this is taken as the decision variable vector at the most recent iterate that is robustly guaranteed to satisfy the problem constraints with some back-off to allow for excitation if needed:

$$\begin{aligned} k^* &:= \\ \arg \max_{\bar{k} \in [0,k]} &\bar{k} \\ \text{subject to } &\bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ &+ \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ &+ (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j,\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ &+ \delta_e \|\bar{\kappa}_{p,j}^{m,\bar{k}}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p} \\ &\max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \leq d_j^{k+1}, \quad \forall j = 1, \dots, n_g \\ &\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{\bar{k}} \preceq \mathbf{u}^U - \delta_e \mathbf{1} \\ &\underline{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k},\bar{k}} \frac{\partial \underline{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ &+ (1 - \eta_{v,\phi}^{\bar{k},\bar{k}}) \underline{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_k - \tau_{\bar{k}}) \leq \\ &\min_{\bar{k} \in \mathcal{K}_f} \left[\begin{array}{l} \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \\ \eta_{c,\phi}^{\bar{k},\bar{k}} \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_k - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}^{\bar{k},\bar{k}}) \bar{\kappa}_{\phi,\tau}^{\bar{k},\bar{k}} (\tau_k - \tau_{\bar{k}}) \end{array} \right], \end{aligned} \quad (298)$$

$$\mathbf{k}_f = \left\{ \bar{k} : \begin{cases} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p}; \\ \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \leq d_j^{k+1}, \quad \forall j = 1, \dots, n_g; \\ \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{\bar{k}} \preceq \mathbf{u}^U - \delta_e \mathbf{1} \end{cases} \right\}. \quad (299)$$

If no point satisfying these restrictions can be found, then one should either cease running experiments, as feasibility cannot be guaranteed, or choose a point that violates the constraints the least in the worst case:

$$k^* := \arg \min_{\bar{k} \in [0, k]} \max \left[g_{p,m}^{\bar{k}}, g_m^{\bar{k}}, u_{L,m}^{\bar{k}}, u_{U,m}^{\bar{k}} \right], \quad (300)$$

where

$$g_{p,m}^{\bar{k}} = \max_{j=1, \dots, n_{g_p}} \begin{bmatrix} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 - d_{p,j}^{k+1} \end{bmatrix} \quad (301)$$

$$g_m^{\bar{k}} = \max_{j=1, \dots, n_g} \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} (g_j(\mathbf{u}) - d_j^{k+1})$$

$$u_{L,m}^{\bar{k}} = \max_{i=1, \dots, n_u} (u_i^L + \delta_e - u_{\bar{k},i})$$

$$u_{U,m}^{\bar{k}} = \max_{i=1, \dots, n_u} (u_{\bar{k},i} + \delta_e - u_i^U).$$

If the cost function is numerical, then one may choose k^* as

$$\begin{aligned} k^* &:= \\ &\arg \min_{\bar{k} \in [0, k]} \phi(\mathbf{u}_{\bar{k}}) \\ &\text{subject to } \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \\ &\quad + \eta_{c,j}^{e,\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ &\quad + (1 - \eta_{c,j}^{e,\bar{k}}) \bar{\kappa}_{p,j\tau}^{e,\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ &\quad + \delta_e \|\kappa_{p,j}^{m,\bar{k}}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p} \\ &\quad \max_{\mathbf{u} \in \mathcal{B}_{e,\bar{k}}} g_j(\mathbf{u}) \leq d_j^{k+1}, \quad \forall j = 1, \dots, n_g \\ &\quad \mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{\bar{k}} \preceq \mathbf{u}^U - \delta_e \mathbf{1}, \end{aligned} \quad (302)$$

and resort to (300) if no point satisfying the restrictions can be found.

9.3 Projection of Optimal Target

Having chosen a reference \mathbf{u}_{k^*} and with some optimal target \mathbf{u}_{k+1}^* supplied by an external optimization method, we now project this point to satisfy the SCFO in some robust sense. To this end, let us state the following finalized version of Algorithm 3.

Algorithm 3F – Projection with Automatic Choice of Projection Parameters and Gradient Robustness (Finalized)

1. Set $\varepsilon_p := \bar{\varepsilon}_p$, $\varepsilon := \bar{\varepsilon}$, $\delta_{g_p} := \bar{\delta}_{g_p}$, $\delta_g := \bar{\delta}_g$, and $\delta_\phi := \bar{\delta}_\phi$, where $\bar{\varepsilon}_{p,j} := \bar{\delta}_{g_{p,j}} \approx -\min_{(\mathbf{u}, \tau) \in \mathcal{J}_\tau} g_{p,j}(\mathbf{u}, \tau)$, $\bar{\varepsilon}_j := \bar{\delta}_{g_j} \approx -\min_{\mathbf{u} \in \mathcal{J}} g_j(\mathbf{u})$, and $\bar{\delta}_\phi \approx \phi_p(\mathbf{u}_0, \tau_0) - \min_{(\mathbf{u}, \tau) \in \mathcal{J}_\tau} \phi_p(\mathbf{u}, \tau)$. Set $\underline{P} := 0, \bar{P} := 1$.

2. Check the feasibility of

$$\min_{\mathbf{u}} \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2$$

$$\text{subject to } \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_{p,j}}$$

$$\nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_{g_{p,j}},$$

$$\begin{aligned} &\bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ &+ \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ &+ (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ &+ \delta_e \|\kappa_{p,j}^{m,k^*}\|_2 \geq -\varepsilon_{p,j} + d_{p,j}^{k+1} \\ &\nabla g_j(\mathbf{u}_{k^*})^T (\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g_j}, \\ &\forall j : \max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} g_j(\mathbf{u}) \geq -\varepsilon_j + d_j^{k+1} \end{aligned} \quad (303)$$

$$\nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_\phi,$$

$$\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u} \preceq \mathbf{u}^U - \delta_e \mathbf{1}$$

for the given choice of $\varepsilon_p, \varepsilon, \delta_{g_p}, \delta_g, \delta_\phi$ by solving a linear programming feasibility problem if the cost function is experimental. If the cost function is numerical, check the feasibility of

$$\min_{\mathbf{u}} \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2$$

$$\text{subject to } \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_{p,j}}$$

$$\nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k+1})^T \begin{bmatrix} \mathbf{u} - \mathbf{u}_{k^*} \\ 0 \end{bmatrix} \leq -\delta_{g_{p,j}},$$

$$\begin{aligned} &\bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) \\ &+ \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ &+ (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ &+ \delta_e \|\kappa_{p,j}^{m,k^*}\|_2 \geq -\varepsilon_{p,j} + d_{p,j}^{k+1} \end{aligned} \quad (304)$$

$$\begin{aligned}
\nabla g_j(\mathbf{u}_{k^*})^T(\mathbf{u} - \mathbf{u}_{k^*}) &\leq -\delta_{g,j}, \\
\forall j: \max_{\mathbf{u} \in \mathcal{B}_{\varepsilon, k^*}} g_j(\mathbf{u}) &\geq -\varepsilon_j + d_j^{k+1} \\
\nabla \phi_p(\mathbf{u}_{k^*})^T(\mathbf{u} - \mathbf{u}_{k^*}) &\leq -\delta_\phi, \\
\mathbf{u}^L + \delta_e \mathbf{1} &\leq \mathbf{u} \leq \mathbf{u}^U - \delta_e \mathbf{1}
\end{aligned}$$

instead. If no solution exists and $\delta_\phi \geq \bar{\delta}_\phi/2^{10}$, set $\varepsilon_p := \varepsilon_p/2$, $\varepsilon := \varepsilon/2$, $\delta_{g_p} := \delta_{g_p}/2$, $\delta_g := \delta_g/2$, $\delta_\phi := \delta_\phi/2$, and repeat this step. Otherwise, proceed to Step 3.

3. If $\delta_\phi < \bar{\delta}_\phi/2^{10}$, terminate with $\bar{\mathbf{u}}_{k+1}^* := \mathbf{u}_{k^*}$. Otherwise, set $P := 0.5\underline{P} + 0.5\bar{P}$ and define the tightened bounds as

$$\begin{aligned}
\nabla \underline{\phi}_p^t(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) &:= \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \\
&+ P \left[\nabla \underline{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) - \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \right], \\
\nabla \bar{\phi}_p^t(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) &:= \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \\
&+ P \left[\nabla \bar{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) - \nabla \hat{\phi}_p(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \right], \\
\nabla \underline{g}_{p,j}^t(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) &:= \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \\
&+ P \left[\nabla \underline{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) - \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \right], \\
\nabla \bar{g}_{p,j}^t(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) &:= \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \\
&+ P \left[\nabla \bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) - \nabla \hat{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1}) \right].
\end{aligned} \tag{305}$$

4. Check the feasibility of the robust projection with the tightened bounds for the obtained $\varepsilon_p, \varepsilon, \delta_{g_p}, \delta_g, \delta_\phi$:

$$\begin{aligned}
\bar{\mathbf{u}}_{k+1}^* &:= \\
\arg \text{minimize } \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\
&\text{subject to } \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_p,j} \\
&\frac{\partial \underline{g}_{p,j}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji} \\
&\frac{\partial \bar{g}_{p,j}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji}, \\
&\forall i = 1, \dots, n_u, \\
&\bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) \\
&+ \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \boldsymbol{\tau}} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\
&+ (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\boldsymbol{\tau}}^{e,k^*} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\
&+ \delta_e \|\boldsymbol{\kappa}_{p,j}^{m,k^*}\|_2 \geq -\varepsilon_{p,j} + d_{p,j}^{k+1} \\
&\nabla g_j(\mathbf{u}_{k^*})^T(\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j}, \\
&\forall j: \max_{\mathbf{u} \in \mathcal{B}_{\varepsilon, k^*}} g_j(\mathbf{u}) \geq -\varepsilon_j + d_j^{k+1}
\end{aligned} \tag{306}$$

$$\begin{aligned}
\sum_{i=1}^{n_u} s_{\phi,i} &\leq -\delta_\phi \\
\frac{\partial \phi^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) &\leq s_{\phi,i} \\
\frac{\partial \bar{\phi}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) &\leq s_{\phi,i} \\
\mathbf{u}^L + \delta_e \mathbf{1} &\leq \mathbf{u} \leq \mathbf{u}^U - \delta_e \mathbf{1},
\end{aligned}$$

or, if the cost is numerical, check the feasibility of

$$\begin{aligned}
\bar{\mathbf{u}}_{k+1}^* &:= \\
\arg \text{minimize } \|\mathbf{u} - \mathbf{u}_{k+1}^*\|_2^2 \\
&\text{subject to } \sum_{i=1}^{n_u} s_{ji} \leq -\delta_{g_p,j} \\
&\frac{\partial \underline{g}_{p,j}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji} \\
&\frac{\partial \bar{g}_{p,j}^t}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k+1})} (u_i - u_{k^*,i}) \leq s_{ji}, \\
&\forall i = 1, \dots, n_u, \\
&\bar{g}_{p,j}(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*}) \\
&+ \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \boldsymbol{\tau}} \Big|_{(\mathbf{u}_{k^*}, \boldsymbol{\tau}_{k^*})} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\
&+ (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\boldsymbol{\tau}}^{e,k^*} (\boldsymbol{\tau}_{k+1} - \boldsymbol{\tau}_{k^*}) \\
&+ \delta_e \|\boldsymbol{\kappa}_{p,j}^{m,k^*}\|_2 \geq -\varepsilon_{p,j} + d_{p,j}^{k+1} \\
&\nabla g_j(\mathbf{u}_{k^*})^T(\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_{g,j}, \\
&\forall j: \max_{\mathbf{u} \in \mathcal{B}_{\varepsilon, k^*}} g_j(\mathbf{u}) \geq -\varepsilon_j + d_j^{k+1} \\
&\nabla \phi(\mathbf{u}_{k^*})^T(\mathbf{u} - \mathbf{u}_{k^*}) \leq -\delta_\phi \\
&\mathbf{u}^L + \delta_e \mathbf{1} \leq \mathbf{u} \leq \mathbf{u}^U - \delta_e \mathbf{1}.
\end{aligned} \tag{307}$$

If (306) – or, in the case of the numerical cost, (307) – is infeasible, set $\bar{P} := P$. Otherwise, set $\underline{P} := P$. If $\bar{P} - \underline{P} < 0.01$, proceed to Step 5. Otherwise, return to Step 3.

5. Set $P := 0.5\underline{P}$, define the tightened bounds as in (305), and solve (306) – or, in the case of a numerical cost, (307) – to obtain the projected target $\bar{\mathbf{u}}_{k+1}^*$. Terminate.

9.4 The Line Search in K_k

Having computed $\bar{\mathbf{u}}_{k+1}^*$, we now define the next experimental decision variable set \mathbf{u}_{k+1} via the filter law

$$\mathbf{u}_{k+1} := \mathbf{u}_{k^*} + K_k(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*}), \tag{308}$$

with $K_k \in [0, 1]$ kept sufficiently small so as to ensure the desired feasibility and cost decrease properties.

In the case of an experimental cost function, we may write the appropriate K_k as the solution to the following univariate line search problem:

$$\begin{aligned}
& K_k := \arg \max_{K \in [0,1]} K \\
& \text{subject to} \\
& \min_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right] \\
& \leq d_{p,j}^{k+1}, \forall j = 1, \dots, n_{gp} \\
& \min_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right] \\
& + \eta_{c,j}^K \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*+K(\bar{u}_{k+1,i}^* - \mathbf{u}_{k^*}), \tau_{k+1}})} (\tau_{k+2} - \tau_{k+1}) \\
& + (1 - \eta_{c,j}^K) \bar{\kappa}_{p,j\tau}^K (\tau_{k+2} - \tau_{k+1}) + \delta_e \|\kappa_{p,j}^{m,K}\|_2 \leq d_{p,j}^{k+1}, \\
& \quad \forall j = 1, \dots, n_{gp} \\
& \max_{\mathbf{u} \in \mathcal{B}_K} g_j(\mathbf{u}) \leq d_j^{k+1}, \forall j = 1, \dots, n_g \\
& \sum_{i=1}^{n_u} \max \left[\begin{aligned} & \frac{\partial \phi_{-p}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}), \\ & \frac{\partial \phi_{-p}}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}) \end{aligned} \right] \\
& + \frac{K}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \left[\begin{aligned} & \underline{M}_{\phi,i_1 i_2}^{k^*} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ & \quad (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}), \\ & \bar{M}_{\phi,i_1 i_2}^{k^*} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ & \quad (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}) \end{aligned} \right] \leq 0
\end{aligned}
\tag{309}$$

$$\begin{aligned}
& \max_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \frac{\phi_{-p}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k}} \frac{\partial \phi_{-p}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}})} \\ & + (1 - \eta_{v,\phi}^{\bar{k}}) \underline{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{v,\phi}^{\bar{k}}} \min \left[\begin{aligned} & \frac{\partial \phi_{-p}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \phi_{-p}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{v,\phi}^{\bar{k}}} \min \left[\begin{aligned} & \underline{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right] \\
& \leq \min_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,\phi}^{\bar{k}} \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,\phi}^{\bar{k}}) \bar{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,\phi}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}), \\ & \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,\phi}^{\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}), \\ & \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right].
\end{aligned}$$

However, in the case that this does not yield a feasible solution due to degradation effects, one may try the relaxed alternative

$$\begin{aligned}
& K_k := \arg \max_{K \in [0,1]} K \\
& \text{subject to} \\
& \min_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right] \\
& + \delta_e \|\kappa_{p,j}^{m,K}\|_2 \leq d_{p,j}^{k+1}, \forall j = 1, \dots, n_{gp} \\
& \max_{\mathbf{u} \in \mathcal{B}_K} g_j(\mathbf{u}) \leq d_j^{k+1}, \forall j = 1, \dots, n_g
\end{aligned}
\tag{310}$$

$$\begin{aligned}
& \sum_{i=1}^{n_u} \max \left[\begin{array}{l} \frac{\partial \phi'_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}), \\ \frac{\partial \bar{\phi}'_p}{\partial u_i} \Big|_{(\mathbf{u}_{k^*}, \tau_{k+1})} (\bar{u}_{k+1,i}^* - u_{k^*,i}) \end{array} \right] \\
& + \frac{K}{2} \sum_{i_1=1}^{n_u} \sum_{i_2=1}^{n_u} \max \left[\begin{array}{l} M_{\phi,i_1 i_2}^{k^*} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}), \\ M_{\bar{\phi},i_1 i_2}^{k^*} (\bar{u}_{k+1,i_1}^* - u_{k^*,i_1}) \\ (\bar{u}_{k+1,i_2}^* - u_{k^*,i_2}) \end{array} \right] \leq 0 \\
& \max_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \phi_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{v,\phi}^{\bar{k}} \frac{\partial \phi_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{v,\phi}^{\bar{k}}) \bar{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,\phi}^{\bar{k}}} \min \left[\begin{array}{l} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \\ + \sum_{i \notin I_{c,\phi}^{\bar{k}}} \min \left[\begin{array}{l} \kappa_{\phi,i}^{\bar{k}} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \\ + \sum_{i \in I_{c,\phi}^{\bar{k}}} \max \left[\begin{array}{l} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \end{array} \right] \\ + \sum_{i \notin I_{c,\phi}^{\bar{k}}} \max \left[\begin{array}{l} \kappa_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \\
& \leq \min_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \bar{\phi}_p(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,\phi}^{\bar{k}} \frac{\partial \bar{\phi}_p}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,\phi}^{\bar{k}}) \bar{\kappa}_{\phi,\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,\phi}^{\bar{k}}} \max \left[\begin{array}{l} \frac{\partial \phi_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}), \\ \frac{\partial \bar{\phi}_p}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} - u_{\bar{k},i}) \end{array} \right] \\ + \sum_{i \notin I_{c,\phi}^{\bar{k}}} \max \left[\begin{array}{l} \kappa_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}), \\ \bar{\kappa}_{\phi,i}^{\bar{k}} (u_{k^*,i} - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \\
& K_k := \arg \underset{K \in [0,1]}{\text{minimize}} \phi(\mathbf{u}_{k^*} + K(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*})) \\
& \text{subject to} \\
& \left[\begin{array}{l} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{array}{l} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \\ + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{array}{l} \kappa_{p,j,i}^{\bar{k}} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \bar{\kappa}_{p,j,i}^{\bar{k}} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \\
& \leq d_{p,j}^{k+1}, \forall j = 1, \dots, n_{gp} \\
& \min_{\bar{k}=0, \dots, k} \left[\begin{array}{l} \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{array}{l} \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \\ + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{array}{l} \kappa_{p,j,i}^{\bar{k}} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ \bar{\kappa}_{p,j,i}^{\bar{k}} (u_{k^*,i} + \\ K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{array} \right] \end{array} \right] \\
& + \eta_{c,j}^K \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*} + K(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*}), \tau_{k+1})} (\tau_{k+2} - \tau_{k+1}) \\
& + (1 - \eta_{c,j}^K) \bar{\kappa}_{p,j\tau}^K (\tau_{k+2} - \tau_{k+1}) + \delta_e \|\kappa_{p,j}^{m,K}\|_2 \leq d_{p,j}^{k+1}, \\
& \quad \forall j = 1, \dots, n_{gp} \\
& \max_{\mathbf{u} \in \mathcal{B}_K} g_j(\mathbf{u}) \leq d_j^{k+1}, \forall j = 1, \dots, n_g,
\end{aligned} \tag{311}$$

or, if this is infeasible, by

If a feasible K_k is still not found, one may simply set $K_k := 0$.

In the case that the cost is numerical in nature, the line search of (309) is replaced by

$$\begin{aligned}
& K_k := \arg \underset{K \in [0,1]}{\text{minimize}} \phi(\mathbf{u}_{k^*} + K(\bar{\mathbf{u}}_{k+1}^* - \mathbf{u}_{k^*})) \\
& \text{subject to}
\end{aligned} \tag{312}$$

$$\begin{aligned}
& \min_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \left. \frac{\partial \bar{g}_{p,j}}{\partial \tau} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in \mathcal{I}_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \left. \frac{\partial \bar{g}_{p,j}}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \left. \frac{\partial \bar{g}_{p,j}}{\partial u_i} \right|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin \mathcal{I}_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \bar{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right] \\ & + \delta_e \|\kappa_{p,j}^{m,K}\|_2 \leq d_{p,j}^{k+1}, \forall j = 1, \dots, n_{g_p} \\ & \max_{\mathbf{u} \in \mathcal{B}_K} g_j(\mathbf{u}) \leq d_j^{k+1}, \forall j = 1, \dots, n_g,
\end{aligned}$$

with $K_k := 0$ as a last resort.

Following this step, one obtains the next set of decision variables \mathbf{u}_{k+1} . Ideally, one would apply these directly in the next experiment, but it may occur that the \mathbf{u}_{k+1} generated, even if satisfying the SCFO, is not sufficiently exciting in some sense – e.g., the step $\mathbf{u}_{k+1} - \mathbf{u}_{k^*}$ is not sufficiently large enough to reject the data corruption due to measurement noise, or the local data set that \mathbf{u}_{k+1} becomes a member of is not well-poised for regression and a better poised set is needed. If the user deems the generated \mathbf{u}_{k+1} as not sufficiently exciting, they then have the freedom to override it and replace it by any other $\mathbf{u} \in \mathcal{B}_{e,k^*}$, as any such \mathbf{u} is guaranteed to be feasible and may be better from the information-gathering point of view.

10 Theoretical Properties

We conclude this document by examining the very important theoretical questions that the development in the previous sections has given rise to – namely, what can we prove about the implementable version of the SCFO? As should be expected given the number of additional complexities that were introduced, we require stronger assumptions than those in [10] in order to prove anything substantial.

As before, we will be able to show that feasibility – or relaxed feasibility requirements, when soft constraints are used – may be guaranteed throughout the optimization process. For convergence to an FJ point, the analysis is complicated significantly by the presence of degradation, the errors in both the experimental function values and the experimental function gradients, and the addition of iteration-dependent back-offs in the optimization problem.

While we could work around all of these issues by making degradation negligible, making the function value/gradient errors arbitrarily small as $k \rightarrow \infty$, and making the back-offs constant – all of which would essentially allow us to recycle the analysis of [10] – we are not, at the present moment, satisfied with this approach as it requires a number of strong assumptions. As such, global convergence to an optimum will not be examined in the present document, but will instead be addressed in a more adequate manner in a future version.

10.1 Proof of Feasible-Side Iterates

Let us first make the following assumption, which is simply the modified version of Assumption A1:

A1F: The functions ϕ_p (or ϕ , if the cost is numerical), g_p , and g are C^2 on an open set containing \mathcal{I}_τ .

As before, Assumption A1F is needed to guarantee the existence of the necessary Lipschitz constants over \mathcal{I}_τ .

To ensure that one can always maintain feasibility, one needs an analogue to Assumption A2 that guarantees that a certain input point, taken here to be \mathbf{u}_0 , is known to be feasible regardless of the time τ . In doing so, it will always be possible to choose this point as the reference \mathbf{u}_{k^*} when a feasibility-guaranteeing reference cannot be found by standard means. However, so that sufficient excitation may be accounted for, one needs to expand this assumption to include the ball $\mathcal{B}_{e,0}$.

A2F: The initial experimental iterate, \mathbf{u}_0 , and the ball of radius δ_e around it, $\mathcal{B}_{e,0}$, are strictly feasible with respect to the experimental constraints for all τ ($g_{p,j}(\mathbf{u}, \tau) < 0$, $\forall \mathbf{u} \in \mathcal{B}_{e,0}$, $\forall j = 1, \dots, n_{g_p}$), feasible with respect to the numerical constraints ($g_j(\mathbf{u}) \leq 0$, $\forall \mathbf{u} \in \mathcal{B}_{e,0}$, $\forall j = 1, \dots, n_g$), and lies in the compressed experimental space ($\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_0 \preceq \mathbf{u}^U - \delta_e \mathbf{1}$).

The practical significance of this assumption is the guarantee that setting $\mathbf{u}_{k^*} := \mathbf{u}_0$ will always result in a reference point that both is feasible and guarantees the existence of a feasible \mathcal{B}_{e,k^*} . In applications, it may be seen as a sort of “safe point” – a set of decision variables that one knows is suboptimal but which one also knows will always meet safety specifications with some tolerance.

Additional assumptions are required on the correctness of the Lipschitz constants and the concavity relaxations, as well as on the probabilistic validity of the gradient-estimate bounds and the upper bounds on the experimental constraint function values.

A3: The supplied Lipschitz constants $\underline{\kappa}_{p,ji}$, $\bar{\kappa}_{p,ji}$, $\underline{\kappa}_{p,j\tau}$, $\bar{\kappa}_{p,j\tau}$, $\underline{\kappa}_{p,ji}^{\bar{k}}$, $\bar{\kappa}_{p,ji}^{\bar{k}}$, $\underline{\kappa}_{p,j\tau}^{\bar{k}}$, $\bar{\kappa}_{p,j\tau}^{\bar{k}}$, $\underline{\kappa}_{p,j\tau}^{\bar{k}_1, \bar{k}_2}$, $\bar{\kappa}_{p,j\tau}^{\bar{k}_1, \bar{k}_2}$, $\underline{\kappa}_{p,ji}^{e,k^*}$, $\bar{\kappa}_{p,ji}^{e,k^*}$

$\underline{\kappa}_{p,j\tau}^{e,k^*}$, $\overline{\kappa}_{p,j\tau}^{e,k^*}$, $\underline{\kappa}_{p,ji}^K$, $\overline{\kappa}_{p,ji}^K$ are *correct* in the sense of satisfying (44), (46), (108), (110), (114), (216), (217), and (258).

A4: The supplied degradation concavity indicators $\eta_{c,j}$, $\eta_{c,j}^{e,\bar{k}}$, $\eta_{c,j}^{\bar{k}}$, $\eta_{c,j}^K$ and the concavity index sets $I_{c,j}$, $I_{c,j}^{\bar{k}}$ are *correct* in the sense that the concave relationships that they denote exist on the relevant domains.

A5: $g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \leq \overline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})$ almost surely for all $j = 1, \dots, n_{g_p}$ and for all $\bar{k} \in [0, k]$.

A6: $\nabla g_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \in \left[\nabla \underline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}), \nabla \overline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) \right]$ almost surely for all $j = 1, \dots, n_{g_p}$ and for all $\bar{k} \in [0, k]$.

We may now state the main feasibility result.

Theorem 4 (SCFO feasibility in the implementable case)

Let Assumptions A1F, A2F, A3, A4, A5, A6 hold and let the reference point be chosen as follows:

- \mathbf{u}_{k^*} is set as the solution to (298) – or as the solution to (302) when the cost is numerical – if such a solution exists.
- If (298) – or, for the case of a numerical cost, (302) – is infeasible, \mathbf{u}_{k^*} is set as \mathbf{u}_0 .

Let \mathbf{u}_{k+1} then be defined by (308), with $\bar{\mathbf{u}}_{k+1}^*$ computed by Algorithm 3F and K_k chosen as follows:

- K_k is set as the solution to (309) (resp., to (311), for a numerical cost) if such a solution exists.
- If (309) (resp., (311)) is infeasible, K_k is set as the solution to (310) (resp., to (312)) if such a solution exists.
- If both (309) and (310) (resp., (311) and (312)) are infeasible, K_k is set as 0.

In the case that the resulting \mathbf{u}_{k+1} is not sufficiently exciting with respect to user specifications, it is replaced with some other $\mathbf{u}_{k+1} \in \mathcal{B}_{e,k^*}$.

It follows that if the slack update law (267) is applied, with the slack-reduction constants satisfying (270) and (275), the bounds (268) hold almost surely for all j , with $\mathbf{u}^L \preceq \mathbf{u}_k \preceq \mathbf{u}^U$ for all k .

Proof Considering the case where the cost function is experimental and enumerating the algorithmic possibilities, one has:

- (i) \mathbf{u}_{k^*} is the solution to (298) and \mathbf{u}_{k+1} is defined by (308), with K_k the solution to (309),
- (ii) \mathbf{u}_{k^*} is the solution to (298) and \mathbf{u}_{k+1} is defined by (308), with K_k the solution to (310),
- (iii) \mathbf{u}_{k^*} is the solution to (298) and \mathbf{u}_{k+1} is defined by (308), with $K_k := 0$,
- (iv) \mathbf{u}_{k^*} is the solution to (298) and \mathbf{u}_{k+1} is chosen as some sufficiently exciting point belonging to \mathcal{B}_{e,k^*} ,

- (v) $\mathbf{u}_{k^*} := \mathbf{u}_0$ and \mathbf{u}_{k+1} is defined by (308), with K_k the solution to (309),
- (vi) $\mathbf{u}_{k^*} := \mathbf{u}_0$ and \mathbf{u}_{k+1} is defined by (308), with K_k the solution to (310),
- (vii) $\mathbf{u}_{k^*} := \mathbf{u}_0$ and \mathbf{u}_{k+1} is defined by (308), with $K_k := 0$,
- (viii) $\mathbf{u}_{k^*} := \mathbf{u}_0$ and \mathbf{u}_{k+1} is chosen as some sufficiently exciting point belonging to \mathcal{B}_{e,k^*} .

We will proceed to prove the desired result by showing that the conditions needed to employ Theorem 3 – i.e., those given by (266) – hold for all k . That $\mathbf{u}^L \preceq \mathbf{u}_k \preceq \mathbf{u}^U$ holds for all k will be verified in a much simpler manner.

That these conditions are met for $k = 0$ follows directly from Assumption A2F. For the general k , it is required that we consider each of the eight algorithmic scenarios listed above.

Considering Scenario (i), where we assume that K_k satisfies the constraints of (309), we have that

$$\begin{aligned} & \left[\overline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \overline{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \right. \\ & \left. + (1 - \eta_{c,j}^{\bar{k}}) \overline{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \right] \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \overline{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \overline{\kappa}_{p,ji}^{\bar{k}} (u_{k^*,i} + \\ & K_k (\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & \leq d_{p,j}^{k+1}, \forall j = 1, \dots, n_{g_p}, \end{aligned} \quad (313)$$

which, with the backwards substitution of the filter law (308), becomes

$$\begin{aligned} & \left[\overline{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \overline{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \right. \\ & \left. + (1 - \eta_{c,j}^{\bar{k}}) \overline{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \right] \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial g_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}), \\ & \frac{\partial \overline{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \underline{\kappa}_{p,ji}^{\bar{k}} (u_{k+1,i} - u_{\bar{k},i}), \\ & \overline{\kappa}_{p,ji}^{\bar{k}} (u_{k+1,i} - u_{\bar{k},i}) \end{aligned} \right] \\ & \leq d_{p,j}^{k+1}, \forall j = 1, \dots, n_{g_p}. \end{aligned} \quad (314)$$

Since

$$g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \leq \min_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k+1,i} - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\bar{\kappa}_{p,ji}^{\bar{k}}}{\bar{\kappa}_{p,ji}^{\bar{k}}} (u_{k+1,i} - u_{\bar{k},i}), \\ & \frac{\bar{\kappa}_{p,ji}^{\bar{k}}}{\bar{\kappa}_{p,ji}^{\bar{k}}} (u_{k+1,i} - u_{\bar{k},i}) \end{aligned} \right] \end{aligned} \right], \quad (315)$$

it follows that $g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \leq d_{p,j}^{k+1}$, $\forall j = 1, \dots, n_{g_p}$, which thus meets the first requirement of (266) with an index shift. Also from (309), we have that

$$\max_{\mathbf{u} \in \mathcal{B}_{e,k+1}} g_j(\mathbf{u}) \leq d_j^{k+1}, \quad \forall j = 1, \dots, n_g \quad (316)$$

is satisfied, which in turn implies $g_j(\mathbf{u}_{k+1}) \leq d_j^{k+1}$, $\forall j = 1, \dots, n_g$ and satisfies the second requirement of (266) with an index shift. To show that $\mathbf{u}^L \preceq \mathbf{u}_{k+1} \preceq \mathbf{u}^U$, it is sufficient to note that since both $\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{k+1} \preceq \mathbf{u}^U - \delta_e \mathbf{1}$ and $\mathbf{u}^L + \delta_e \mathbf{1} \preceq \bar{\mathbf{u}}_{k+1}^* \preceq \mathbf{u}^U - \delta_e \mathbf{1}$, the latter following from the constraints in the projection of Algorithm 3F, it must be that $\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{k+1} \preceq \mathbf{u}^U - \delta_e \mathbf{1}$ as $K_k \in [0, 1]$.

Scenario (ii) follows an identical analysis, save that instead of (313) we have the more restricting condition

$$\min_{\bar{k}=0, \dots, k} \left[\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}}) + \eta_{c,j}^{\bar{k}} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + (1 - \eta_{c,j}^{\bar{k}}) \bar{\kappa}_{p,j\tau}^{\bar{k}} (\tau_{k+1} - \tau_{\bar{k}}) \\ & + \sum_{i \in I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\partial \bar{g}_{p,j}}{\partial u_i} \Big|_{(\mathbf{u}_{\bar{k}}, \tau_{\bar{k}})} (u_{k^*,i} + \\ & K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \sum_{i \notin I_{c,j}^{\bar{k}}} \max \left[\begin{aligned} & \frac{\bar{\kappa}_{p,ji}^{\bar{k}}}{\bar{\kappa}_{p,ji}^{\bar{k}}} (u_{k^*,i} + \\ & K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}), \\ & \frac{\bar{\kappa}_{p,ji}^{\bar{k}}}{\bar{\kappa}_{p,ji}^{\bar{k}}} (u_{k^*,i} + \\ & K_k(\bar{u}_{k+1,i}^* - u_{k^*,i}) - u_{\bar{k},i}) \end{aligned} \right] \\ & + \delta_e \|\kappa_{p,j}^{m,K}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p}, \end{aligned} \right] \quad (317)$$

which implies (313) since $\delta_e \|\kappa_{p,j}^{m,K}\|_2 \geq 0$, thus implying that $g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \leq d_{p,j}^{k+1}$ for this case as well.

Let us consider Scenario (iv) next, since Scenario (iii) is just a special case of (iv). Since \mathbf{u}_{k^*} is assumed to solve (298), we know that

$$\begin{aligned} & \bar{g}_{p,j}(\mathbf{u}_{k^*}, \tau_{k^*}) + \eta_{c,j}^{e,k^*} \frac{\partial \bar{g}_{p,j}}{\partial \tau} \Big|_{(\mathbf{u}_{k^*}, \tau_{k^*})} (\tau_{k+1} - \tau_{k^*}) \\ & + (1 - \eta_{c,j}^{e,k^*}) \bar{\kappa}_{p,j\tau}^{e,k^*} (\tau_{k+1} - \tau_{k^*}) \\ & + \delta_e \|\kappa_{p,j}^{m,k^*}\|_2 \leq d_{p,j}^{k+1}, \quad \forall j = 1, \dots, n_{g_p}. \end{aligned} \quad (318)$$

From Theorem 1, we know that (318) implies that $g_{p,j}(\mathbf{u}, \tau_{k+1}) \leq d_{p,j}^{k+1}$, $\forall \mathbf{u} \in \mathcal{B}_{e,k^*}$. Since it is assumed that $\mathbf{u}_{k+1} \in \mathcal{B}_{e,k^*}$, it then follows that $g_{p,j}(\mathbf{u}_{k+1}, \tau_{k+1}) \leq d_{p,j}^{k+1}$, $\forall j = 1, \dots, n_{g_p}$, which satisfies the first requirement of (266) with an index shift. Since

$$\max_{\mathbf{u} \in \mathcal{B}_{e,k^*}} g_j(\mathbf{u}) \leq d_j^{k+1}, \quad \forall j = 1, \dots, n_g \quad (319)$$

must also hold, it follows that $g_j(\mathbf{u}_{k+1}) \leq d_j^{k+1}$, $\forall j = 1, \dots, n_g$ by the same logic, which meets the second requirement of (266) with an index shift. Finally, since $\mathbf{u}^L + \delta_e \mathbf{1} \preceq \mathbf{u}_{k+1} \preceq \mathbf{u}^U - \delta_e \mathbf{1}$, it is clear that $\mathbf{u}^L \preceq \mathbf{u}_{k+1} \preceq \mathbf{u}^U$ by Corollary 1.

For Scenario (iii), note that $K_k := 0$ results in $\mathbf{u}_{k+1} := \mathbf{u}_{k^*} \in \mathcal{B}_{e,k^*}$, and so the results derived just above for Scenario (iv) may be applied.

The analyses for Scenarios (v) and (vi) are identical to those of Scenarios (i) and (ii), as all are independent of how \mathbf{u}_{k^*} is chosen since both assume the feasibility of (309) (or (310)).

The proof for Scenarios (vii) and (viii) follows directly from Assumption A2F.

Having proven that (266) must hold for all k regardless of the implementation scenario, the desired result follows from Theorem 3. We can only say that this result holds *almost surely* since (315) is a probabilistic bound that is assumed to hold almost surely here. For the case where the cost is numerical, an identical analysis may be repeated to yield the same result. \square

References

1. Birnbaum, Z.W., Raymond, J., Zuckerman, H.S.: A generalization of Tshebyshev's inequality to two dimensions. *Ann. Math. Statist.* **18**(1), 70–79 (1947)
2. Box, G., Draper, N.: *Evolutionary Operation: A Statistical Method for Process Improvement*. John Wiley & Sons (1969)
3. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press (2008)
4. Brylewski, T., Nanko, M., Maruyama, T., Przybylski, K.: Application of Fe-16Cr ferritic alloy to interconnector for a solid oxide fuel cell. *Solid State Ionics* **143**, 131–150 (2001)

5. Bunin, G.A., François, G., Bonvin, D.: From discrete measurements to bounded gradient estimates: A look at some regularizing structures. *Ind. Eng. Chem. Res.* **52**(35), 12500–12513 (2013)
6. Bunin, G.A., François, G., Bonvin, D.: A real-time optimization framework for the iterative controller tuning problem. *Processes* **1**(2), 203–237 (2013)
7. Bunin, G.A., François, G., Bonvin, D.: The SCFO Real-Time Optimization Solver: Users' Guide (version 0.9.4). Ecole Polytechnique Fédérale de Lausanne (2013). <http://infoscience.epfl.ch/record/186672>
8. Bunin, G.A., François, G., Bonvin, D.: On linear and quadratic Lipschitz bounds for twice continuously differentiable functions. *arXiv [math.OC]* (2014)
9. Bunin, G.A., François, G., Bonvin, D.: Sufficient conditions for feasibility and optimality of real-time optimization schemes - II. Implementation issues. *arXiv:1308.2625v1 [math.OC]* (2013)
10. Bunin, G.A., François, G., Bonvin, D.: Feasible-side global convergence in experimental optimization. *SIAM J. Optim.* (submitted) (2014)
11. Bunin, G.A., Wuillemin, Z., François, G., Nakajo, A., Tsikonis, L., Bonvin, D.: Experimental real-time optimization of a solid oxide fuel cell stack via constraint adaptation. *Energy* **39**, 54–62 (2012)
12. Conn, A.R., Scheinberg, K., Vicente, L.N.: Geometry of sample sets in derivative-free optimization: polynomial regression and underdetermined interpolation. *IMA J. Numer. Anal.* **28**, 721–748 (2008)
13. Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. Cambridge University Press (2009)
14. Divoky, J.J., Taylor, R.W.: Detecting process drift with combinations of trend and zonal supplementary runs rules. *Int. J. Qual. Reliab. Manag.* **12**(2), 60–71 (1995)
15. Ferentinos, K.: On Tchebycheff's type inequalities. *Trabajos Estadíst. Investigación Oper.* **33**(1), 125–132 (1982)
16. Hansen, P., Jaumard, B., Lu, S.: On using estimates of Lipschitz constants in global optimization. *J. Optim. Theory Appl.* **75**(1), 195–200 (1992)
17. Horst, R., Pardalos, P., Thoai, N.: Introduction to Global Optimization, *Nonconvex Optimization and its Applications*, vol. 3. Kluwer Academic Publishers (1995)
18. Hotelling, H.: Experimental determination of the maximum of a function. *Ann. Math. Statist.* **12**(1), 20–45 (1941)
19. Katz, R., Allen, T.J.: Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R & D project groups. *R&D Manage.* **12**(1), 7–20 (1982)
20. Kettaneh-Wold, N., Michel, B., Chalmers, J., Ho, T.T., Schimunek, B.: Process drift: recognition and solutions. In: 9th European Advanced Equipment Control / Advanced Process Control Conference (Tel Aviv) (2008)
21. Kuemmerle, W.: Optimal scale for research and development in foreign environments - an investigation into size and performance of research and development laboratories abroad. *Res. Policy* **27**, 111–126 (1998)
22. Larson, J.M.: Derivative-free optimization of noisy functions. Ph.D. thesis, University of Colorado (2012)
23. Lee, J., Fenves, G.L.: Plastic-damage model for cyclic loading of concrete structures. *J. Eng. Mech.* **124**, 892–900 (1998)
24. Marchetti, A., Chachuat, B., Bonvin, D.: A dual modifier-adaptation approach for real-time optimization. *J. Process Control* **20**(9), 1027–1037 (2010)
25. Marchetti, A., Gopalakrishnan, A., Chachuat, B., Bonvin, D., Tsikonis, L., Nakajo, A., Wuillemin, Z., van Herle, J.: Robust real-time optimization of a solid oxide fuel cell stack. *J. Fuel Cell Sci. Tech.* **8**(5) (2011)
26. Metelskii, N.N., Martynchik, V.N.: Partial convexity. *Math. Notes* **60**(3), 300–305 (1996)
27. Moré, J.J., Wild, S.M.: Estimating computational noise. *SIAM J. Sci. Comput.* **33**(3), 1292–1314 (2011)
28. Myers, R., Montgomery, D., Anderson-Cook, C.: Response Surface Methodology. John Wiley & Sons (2009)
29. Pfaff, G., Forbes, J.F., McLellan, P.J.: Generating information for real-time optimization. *Asia Pac. J. Chem. Eng.* **1**(1-2), 32–43 (2006)
30. Pfaff, G.C.: Generating information for real-time optimization. Master's thesis, University of Alberta (2001)
31. Rodger, E.A., Chachuat, B.: Design methodology of modifier adaptation for on-line optimization of uncertain processes. In: Proceedings of the 18th IFAC World Congress (Milano, Italy), pp. 4113–4118 (2011)
32. Strongin, R.G.: On the convergence of an algorithm for finding a global extremum. *Engin. Cybern.* **11**, 549–555 (1973)
33. Teo, M.: Does size matter in the hedge fund industry? Working Paper, BNP Paribas Hedge Fund Centre, Singapore Management University (2009)
34. Ubhaya, V.: Regression by special functions: Algorithms and complexity. In: C. Floudas, P. Pardalos (eds.) *Encyclopedia of Optimization*, pp. 3268–3273. Springer (2009)
35. Wood, G., Zhang, B.: Estimation of the Lipschitz constant of a function. *J. Global Optim.* **8**, 91–103 (1996)