Optimal measurement-based cost gradient estimate for real-time optimization

Lucas Ferreira Bernardino[a], Sigurd Skogestad[a]

[a]*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway*

Abstract

This work presents a simple and efficient way of estimating the steady-state cost gradient J_u based on available uncertain measurements y. The main motivation is to control J_u to zero in order to minimize the economic cost J. For this purpose, it is shown that the optimal cost gradient estimate for unconstrained operation is simply $\hat{J}_u = H(y_m - y^*)$ where H is a constant matrix, y_m is the vector of measurements and y^* is their nominally unconstrained optimal value. The derivation of the optimal H-matrix is based on existing methods for self-optimizing control and therefore the result is exact for a convex quadratic economic cost J with linear constraints and measurements. The optimality holds locally in other cases. For the constrained case, the unconstrained gradient J_u should be multiplied by the nullspace of the active constraints and the resulting "reduced gradient" controlled to zero. The process and cost model, including a model of the expected disturbances, is used only offline, so in effect, the method provides a data-based alternative to conventional RTO which uses online models both for optimization and estimation. In summary, the proposed gradient estimate is simple to implement and may form the basis for solving industrial RTO problems in an efficient manner.

Keywords: self-optimizing control, optimal operation, controlled variable design, gradient estimation

1. Introduction

When the aim is to implement a control strategy to achieve optimal steady-state operation, one usually thinks of adding a conventional real-time optimization (RTO) layer [1] which adjusts the setpoints to the control layer. However, conventional RTO requires online optimization of a detailed nonlinear model which is usually expensive to obtain and maintain. In addition, the success of RTO relies on estimating the disturbances, which is also model-based and is usually slow [2]. These limitations can be circumvented by feedback-optimizing control where the aim is to move the optimization into the control layer [3] so that the RTO layer may be eliminated or at least less frequent RTO updates are needed. In this respect, the most important decision is the proper selection of the controlled variables (CVs) which is the idea of self-optimizing control [4]. According to the first-order optimality conditions [5], the ideal self-optimizing CV would be the gradient of the economic cost function (reduced gradient for the constrained case), which when driven to zero achieves optimal operation without the need for an RTO layer. In this paper, the cost gradient is denoted J_u, but it is also sometimes denoted $\nabla_u J$ in other works. The goal of this paper is then to derive a simple measurement-based estimate of the gradient J_u for use in a feedback-optimizing control scheme.

The most common industrial approach is to combine the available measurements with a steady-state plant model and from this derive an estimate of the states (including disturbances) and subsequently estimate the gradient. This gradient (or reduced gradient) can be driven to zero by feedback control, or most commonly with RTO updates. This approach is preferred when model-free approaches cannot be applied, especially in large-scale processes [6]. To avoid the steady-state wait time for the estimation, Krishnamoorthy et al. [7] propose a Kalman filter for dynamic state and disturbance estimation, and the model is then linearized around the expected steady-state operating point to give an estimate for the gradient J_u. Both these approaches are referred to as "conventional RTO" in this paper.

An alternative model-free approach is to directly estimate the cost gradient J_u from plant data by input excitation, which is done for example in extremum-seeking control strategies [8, 9]. These methods classically rely on sinusoidal excitation of the inputs [10], but one may also perform least-squares regression on other kinds of perturbation data [11]. In theory, these approaches are ideal and avoid the problem of model-plant mismatch [12], and they can even deal with constraints with specific formulations [13]. Unfortunately, they are of little practical significance for most chemical processes. The reason is that the convergence time is typically at least 100 times larger than the time constant of the process. And since the time constant for a typical full process (for which we want to minimize the economic cost J) is typically several hours, the convergence time may be in the order of days or even months, which is of course unsuitable for real applications with frequent disturbances. Similar problems arise with other data-based methods for gradient estimation, for example in modifier adaptation schemes [14, 15]. There are also model-free methods that directly deal with dynamic data [16], but they must be continuously excited to ensure convergence.

A more realistic approach for a typical full process is to combine model-free approaches, like extremum-seeking control, with faster self-optimizing solutions, such as the one presented in this paper. The reason is that these two approaches are complementary and not competing [17]. The self-optimizing

layer has the advantage of both giving faster control of the process and requiring less frequent updates of the setpoints. Of course, there do exist cases where model-free approaches, like extremum seeking control, may be used alone, and this is when the dynamics (of the controlled plant) are fast and a fast cost measurement J is available (which is not the case for a full process where J has to be computed because it is a combination of many process streams and utilities). Typically, this may apply to local optimization of a part of the process.

The existing model- and data-based estimation approaches have in common that they are rather complex and that the gradient estimation and the use of the gradient for control are divided into separate tasks. However, this separation between estimation and control is not generally optimal. In other words, since it is not clearly defined upfront what the gradient J_u will be used for, we cannot expect that the estimated gradient will be optimal for minimizing the cost J. In this sense, a recent work [18] has attempted to bridge estimation and control into the same extremum-seeking control scheme, which highlights the problem's relevance.

The present work, which may be viewed as a third approach for estimating the gradient, also aims at combining estimation and control. The proposed approach for estimating the cost gradient uses self-optimizing control methods to design controlled variables (CVs) that directly minimize the cost, that is, there is no intermediate step to estimate disturbances or gradients. A further advantage of the proposed approach is simplicity. The resulting CVs are static linear combinations of the available measurements, which greatly simplifies implementation. It has been known that self-optimizing CVs are linked to the cost gradient for the simple case with a sufficient number of noise-free measurements [17]. An equivalent result is also obtained with a neighboring-extremal scheme for gradient estimation assuming output feedback [19]. The main contribution of the present work is to extend this link to more general cases and derive a simple static linear measurement-based expression for the gradient J_u for any number of noisy measurements. The basis is the exact local method [20] of self-optimizing control. We also show in this paper how this gradient estimate is useful when dealing with constraints, both to set the unconstrained degrees of freedom and to identify constraint switching.

In summary, model-free approaches, like extremum-seeking control, should be viewed as complementary to the approach presented in this paper. Also, note that although the gradient estimation presented is based on the system model, this model is used only offline, so when implemented, the proposed approach is measurement-based and provides an efficient and realistic approach to conventional RTO.

The paper is organized as follows. Section 2 presents the mathematical problem considered in this work. Section 3 describes how this problem is related to self-optimizing control. In Section 4 we present the main result of this work based on the analysis of the unconstrained problem, which is complemented by the analysis of the constrained problem in Section 5. An example of the application of these results to a decentralized control framework is shown in Section 6 with a numerical example, showing its use with changing active constraints. Some remarks about the presented results are made in Section 8, and the paper is concluded in Section 9.

2. Problem statement

The steady-state optimization problem considered in this work is of the form:

$$\begin{aligned} \min_{u} \quad & J(u,d) \\ \text{s.t.} \quad & g(u,d) \le 0 \end{aligned} \tag{1}$$

Here, J: $\mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \to \mathbb{R}$ denotes the objective function, g: $\mathbb{R}^{n_u} \times \mathbb{R}^{n_d} \to \mathbb{R}^{n_g}$ the inequality constraints, $u \in \mathbb{R}^{n_u}$ the decision variables (inputs; manipulated variables for steady-state control), and $d \in \mathbb{R}^{n_d}$ the disturbance variables (including model parameters) which are assumed varying and generally unknown in this paper. The available online information about the system is assumed to be the measured variables $y \in \mathbb{R}^{n_y}$ (which usually include u and may include measured disturbances). Any internal states have been formally eliminated from the mathematical formulation in (1).

The optimal input, which is the solution to the problem in Equation (1), is in the paper denoted $u^{opt}(d)$. It satisfies the following first-order KKT conditions [5]:

$$J_u(u^{opt},d) + g_u(u^{opt},d)^T \lambda^{opt} = 0 \tag{2a}$$
$$g(u^{opt},d) \le 0 \tag{2b}$$
$$\lambda^{opt} \ge 0 \tag{2c}$$
$$g(u^{opt},d)^T \lambda^{opt} = 0 \tag{2d}$$

Here, $J_u(u,d) \in \mathbb{R}^{n_u}$ denotes the gradient of J with respect to u, $g_u(u,d) \in \mathbb{R}^{n_g \times n_u}$ denotes the gradient of g with respect to u, and $\lambda^{opt} \in \mathbb{R}^{n_g}$ denotes the Lagrange multipliers at the optimum. Note that it is the *unconstrained* cost gradient J_u that enters into the first-order optimality conditions.

The cost $J(u,d)$ and the constraints $g(u,d)$ in Equation (1) can be approximated locally by the following Taylor expansions centered at the nominal point (u^*, d^*):

$$\begin{aligned} J(u,d) = J^* + \begin{bmatrix} J_u^{*T} & J_d^{*T} \end{bmatrix} \begin{bmatrix} (u-u^*) \\ (d-d^*) \end{bmatrix} \\ + \frac{1}{2} \begin{bmatrix} (u-u^*)^T & (d-d^*)^T \end{bmatrix} \underbrace{\begin{bmatrix} J_{uu} & J_{ud} \\ J_{ud}^T & J_{dd} \end{bmatrix}}_{\mathcal{H}} \begin{bmatrix} (u-u^*) \\ (d-d^*) \end{bmatrix} \end{aligned} \tag{3}$$

$$g(u,d) = g^* + \begin{bmatrix} g_u^* & g_d^* \end{bmatrix} \begin{bmatrix} (u-u^*) \\ (d-d^*) \end{bmatrix} \tag{4}$$

where $(u-u^*)$ and $(d-d^*)$ denote, respectively, the inputs and disturbances as their deviation from the nominal point.

The cost expression in Equation (3) is exact for quadratic problems where the Hessian $\mathcal{H}$ (including J_{uu}) is independent of the operating point. In general, there will be an approximation error if the actual operation moves away from the nominal

point. Strictly speaking, the elements in the Hessian matrix $\mathcal{H}$ should have a superscript $*$ (e.g. J^*_{uu}), but this is omitted to simplify notation, and also because it assumed that they remain approximately constant.

The objective of this paper is to find from the available measurements y (which are subject to noise n^y) an optimal estimate of the gradient J_u (which will vary as a function of u and d) for use in real-time optimization. The expected magnitudes of the disturbances and measurement errors are quantified by diagonal weight matrices W_d and W_{n^y}. That is, we assume that:

$$\begin{aligned}(d - d^*) &= W_d d' \\ n^y &= W_{n^y} n^{y\prime}\end{aligned} \tag{5}$$

where the combined generating set of possible d' and $n^{y\prime}$ is unit two-norm bounded, i.e.:

$$\left\| \begin{bmatrix} d' \\ n^{y\prime} \end{bmatrix} \right\|_2 \leq 1 \tag{6}$$

Note that we are considering steady-state operation, so n^y represents the static measurement error, that is, the measurement bias. Often, n^y is called measurement noise, but this may be a bit misleading because the average (steady-state) value is not zero, as is usually assumed in stochastic optimal control. For example, $n^y = 0.15$ means that if the actual value is $y = 2.7$, then the measured value is $y_m = y + n^y = 2.85$. Finally, note that the objective of this paper is not to find the "optimal" gradient J_u in itself, but the optimal estimate $\hat{J}_u$ to be used in the first-order optimality condition (2a) to solve the problem in (1).

3. Optimal operation for the unconstrained case: Self-optimizing control

In the following consider the case with no constraints g and assume that the nominal operating point is optimal, that is,

$$u^* = u^{opt}(d^*)$$

It then follows from the first-order KKT condition (2a) that:

$$J^*_u = 0$$

This assumption is made to simplify the expressions for the loss, and the controlled variables derived here do not depend on this assumption (see chapter 6 in Alstad [21]).

Following Halvorsen et al. [22], we can derive from Equation (3) the economic loss encountered by applying an input u, compared to using the optimal input $u^{opt}(d)$:

$$L = J(u,d) - J^{opt}(d) = \frac{1}{2}(u - u^{opt})^T J_{uu} (u - u^{opt}) = \frac{1}{2}\|z\|_2^2 \tag{7}$$

where $J^{opt}(d) = J(u^{opt}(d), d)$ is the optimal cost for a given d and the loss variable z is defined as:

$$z \triangleq J_{uu}^{1/2}(u - u^{opt}) \tag{8}$$

The idea of self-optimizing control is to achieve optimal operation using feedback control. In this paper, the controlled variables (CVs) c are assumed to be linear combinations of the measured variables, $c = Hy$, and we use a linear steady-state measurement model:

$$y = G^y u + G^y_d d \tag{9}$$

Note that the actual measured value is $y_m = y + n^y$. The setpoints c_s are assumed to be constant; see Figure 1. To be nominally optimal (with no disturbances or measurement noise), we must choose $c_s = c^* = Hy^*$ where $y^* = y^{opt}(d^*)$. The controller K has integral action, which means that at steady state the control error

$$(c_m - c^*) = H(y_m - y^*)$$

is controlled to a constant value of zero. The controlled variables c should use up all the available degrees of freedom, and therefore $n_c = n_u$. In this paper, H is allowed to be a full matrix, that is, there are no structural limitations on H.

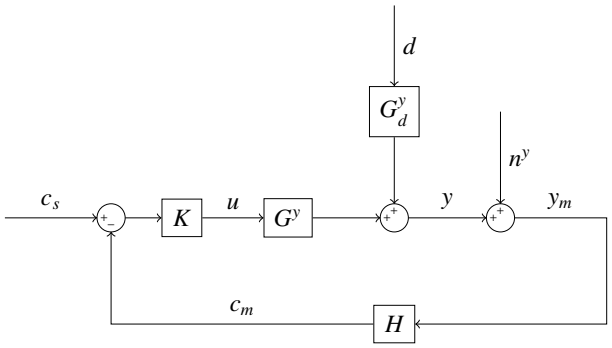


Figure 1: Block diagram of closed-loop system. When H is selected as proposed in this paper, the input to the controller K is the negative cost gradient, that is, $c_s - Hy_m = -\hat{J}_u$ see eq. (21). This achieves optimal steady-state operation if in addition any active constraints are controlled.

For the expected disturbances and noise in Equation (6), Alstad et al. [20] derived the following analytical expression for the optimal H, known as the "exact local method", which minimizes both the worst-case and average loss L in Equation (7):

$$H = M_n^{-1} J_{uu}^{1/2} \left[G^{yT} \left(\tilde{F}\tilde{F}^T \right)^{-1} G^y \right]^{-1} G^{yT} \left(\tilde{F}\tilde{F}^T \right)^{-1} \tag{10}$$

where

$$\begin{aligned}\tilde{F} &= [F W_d \quad W_{n^y}] \\ F &= \frac{dy^{opt}}{dd} = G^y_d - G^y J_{uu}^{-1} J_{ud}\end{aligned} \tag{11}$$

The solution for H is not unique as the matrix $M_n = J_{uu}^{1/2}(HG^y)^{-1}$ can be freely chosen. The non-uniqueness comes because if $c - c^* = 0$ then so is $D(c - c^*) = 0$ for any non-singular D. In the solution derived in Alstad et al. [20], the choice is $M_n = I$. The simplest expression for the optimal H results if we select M_n such that $H = G^{yT} \left(\tilde{F}\tilde{F}^T \right)^{-1}$ [23].

However, in the next section, we want to find an estimate for J_u (also when $J_u \neq 0$), and in this case directions matter. For this reason, we will choose:

$$M_n = J_{uu}^{-1/2} \tag{12}$$

and we show below that the optimal estimate for the gradient J_u is then equal to $H^J(y - y^*)$, where according to the exact local method:

$$H^J = J_{uu}\left[G^{yT}\left(\tilde{F}\tilde{F}^T\right)^{-1}G^y\right]^{-1}G^{yT}\left(\tilde{F}\tilde{F}^T\right)^{-1} \tag{13}$$

With different assumptions, other expressions for H may be derived. For the case with a sufficient number of independent measurements ($n_y \geq n_u + n_d$) it is possible to achieve zero disturbance loss for the case with no measurement noise by choosing H such that $HF = 0$ (nullspace method). For the case $n_y = n_u + n_d$, we have the following explicit expression for the nullspace method:

$$H = M_n^{-1}\tilde{J}(\tilde{G}^y)^{-1} \tag{14}$$

where $\tilde{G}^y = [G^y \quad G_d^y]$ and $\tilde{J} = J_{uu}^{1/2}[I \quad J_{uu}^{-1}J_{ud}]$. The generalization to use all measurements ($n_y \geq n_u + n_d$) in a way that also minimizes the effect of measurement noise is known as the extended nullspace method [20] for which we have:

$$H = M_n^{-1}\tilde{J}(W_{n^y}^{-1}\tilde{G}^y)^{\dagger}W_{n^y}^{-1} \tag{15}$$

All these expressions for H can be used for gradient estimation, provided that we choose $M_n = J_{uu}^{-1/2}$, or equivalently $HG^y = J_{uu}$.

4. Optimal gradient estimate for the unconstrained case

We will now use the results from self-optimizing control to derive the optimal gradient estimate, where by "optimal" we mean that controlling the gradient estimate to zero achieves optimal steady-state operation, that is, it minimizes the loss L in Equation (7) (worst-case or average value) for the expected disturbances and noise as in Equation (6).

To do this, we want to express the loss variable z from (8) in terms of the gradient J_u. First, note that (Figure 1):

$$(c - c^{opt}(d)) = HG^y(u - u^{opt})$$

Second, a first-order Taylor expansion of the gradient around the optimal operating point gives:

$$J_u(u,d) = \underbrace{J_u(u^{opt},d)}_{J_u^{opt}(d)} + J_{uu}(u - u^{opt}(d))$$

Inserting the above two expressions into the definition of the loss variable z in (8) gives:

$$z \triangleq J_{uu}^{1/2}(u - u^{opt}) = \underbrace{J_{uu}^{1/2}(HG^y)^{-1}}_{M_n}(c - c^{opt}(d)) = J_{uu}^{-1/2}(J_u - J_u^{opt}(d)) \tag{16}$$

For the unconstrained case, we have $J_u^{opt}(d) = 0$, and this is assumed in the following. We then get $z = J_{uu}^{-1/2}J_u$ and to minimize the norm of z, and thereby the loss in (7), we conclude that we ideally want $J_u = 0$ at steady state. However, as we will see, it is not possible to achieve $J_u = 0$ in practice because of measurement error.

For the choice $M_n = J_{uu}^{-1/2}$ (which we will use in the following), we derive from (16) the following expression for the gradient:

$$J_u = c - c^{opt}(d) = Hy - Hy^{opt}(d)$$

which may be rewritten as:

$$J_u = H(y_m - y^*) - H\underbrace{(y_m - y)}_{n^y} - H(y^{opt}(d) - y^*) \tag{17}$$

where we choose $y^* = y^{opt}(d^*)$ because the nominal point is assumed optimal. Note from (11) that $(y^{opt}(d) - y^*) = F(d - d^*)$ for the unconstrained case. We then have:

$$J_u = H(y_m - y^*) - Hn^y - HF(d - d^*) \tag{18}$$

Note that with a fixed matrix H, the last two terms are unaffected by the input u, that is, unaffected by control.

With no measurement error ($n^y = 0$), the second term in Equation (18) is zero. If we use the nullspace method to choose H, then $HF = 0$, and also the third term is zero. The optimal control policy, according to self-optimizing control, is then to adjust u such that the first term is zero, for example, to use feedback control to keep the measurement combinations keep $c_m = Hy_m$ at a constant setpoint $c^* = Hy^*$. This gives $J_u = 0$ and the loss is zero.

More generally, with measurement noise and disturbances, we can use the exact local method to choose the H that minimizes the combined effect of the second and third terms in (18). The optimal control policy, similarly to the case without noise, is then to adjust u such that the first term in Equation (18) is zero. This minimizes the expected norm of z as in (16), and consequently the economic loss L in (7). More importantly, and this is the main result of the paper, the optimal gradient estimate for unconstrained operation, which should be kept at zero at steady state, is simply the first term in (18), that is:

$$\hat{J}_u = H(y_m - y^*) \tag{19}$$

where y_m is the measurement vector, $y^* = y^{opt}(d^*)$ is the nominal optimal value of the measurement y, and H is given by H^J in Equation (13) (exact local method). This follows from self-optimizing control theory, because choosing $H = H^J$ minimizes the effect of the second and third terms in Equation (18) (it minimizes both the expected and worst-case loss when d and n^y vary as given in (6)).

Interestingly, since the second and third terms in (18) are generally nonzero (due to measurement noise and disturbances), it follows that optimal operation (in terms of minimizing the economic loss) does not give $J_u = 0$ at steady state. This may seem surprising, but it is expected because one cannot achieve truly optimal steady-state operation (with $J_u = 0$ and zero loss) with unknown disturbances and static measurement bias (nonzero n^y).

In summary, the steady-state loss L in Equation (7) is minimized when we keep $\hat{J}_u = H^J(y_m - y^*) = 0$, and we have proven the following theorem:

Theorem 1. Optimal unconstrained gradient estimate. *Consider the static optimization problem in* (1) *with no active constraints, where the quadratic approximation* (3) *holds. The available measurements are* $y_m = G^y u + G_d^y d + n^y$ *(linear approximation) where the unknown disturbances* d *and static measurement errors* n^y *are bounded as given in* (5) *and* (6). *Consider further that the point* (u^*, d^*) *is an optimal unconstrained point, such that* $J_u(u^*, d^*) = 0$, $u^* = u^{opt}(d^*)$ *and* $y^* = y^{opt}(d^*)$. *The cost gradient* J_u *is then given in* (18) *and the estimate* $\hat{J}_u = H^J(y_m - y^*)$ *with* H^J *in* (13) *is an optimal estimate in the sense that adjusting the inputs* u *to make* $\hat{J}_u = 0$ *(e.g., by feedback control, see Figure 1) minimizes both the average and the worst-case value of the economic loss* (7).

If there is no measurement error ($n^y = 0$, that is, $W_{n^y} = 0$) and we have a sufficient number of measurement ($n_y = n_u + n_d$) then instead of using $H = H^J$ from the exact local method, we may use H from the nullspace method (equation (14) with $M_n = J_{uu}^{-1/2}$). This gives H in the nullspace of F ($HF = 0$) and achieves zero loss for disturbances (with no measurement error), that is, the last term in (18) is zero. If we have additional measurements ($n_y > n_u + n_d$) then we may use H from the "extended nullpace method" (equation (15) with $M_n = J_{uu}^{-1/2}$) which uses the extra measurements to minimize also the second term in (18). However, in general we recommend using $H = H^J$ from the exact local method. It gives the optimal balance between disturbances and measurement error (as it minimizes both the average and worst-case sum of last two terms in (18)) and importantly applies also to the case with fewer measurements ($n_y < n_u + n_d$).

5. Optimal gradient estimate for the constrained case

Now, we focus on the use of this result in the operation with changing active constraints. For that, we state the following:

Theorem 2. Optimal gradient estimate in constrained case. *The optimal unconstrained gradient estimate* $\hat{J}_u = H^J(y_m - y^*)$ *(Theorem 1) is optimal also in the constrained case when used in the first-order KKT conditions* (2). *This also means that the optimal estimate of the reduced gradient (which should be zero at the optimal point) is* $N_{\mathcal{A}}{}^T \hat{J}_u = N_{\mathcal{A}}{}^T H^J(y_m - y^*)$ *where* $N_{\mathcal{A}}$ *is a basis for the nullspace of* $g_{u,\mathcal{A}}$, *that is,* $g_{u,\mathcal{A}} N_{\mathcal{A}} = 0$, *and* $\mathcal{A}$ *represents the set of active constraints.*

The theorem may seem straightforward and require no further proof since J_u in (2a) is the unconstrained gradient, and the gradient estimate $\hat{J}_u$ in (19) is the one that minimizes the loss in the unconstrained case for a given measurement set y. Furthermore, the idea of reduced gradient is well-established, being used in optimization methods [24, 25] and to solve control problems [26, 27]. Nevertheless, in Appendix A, we provide a detailed proof that controlling the reduced gradient estimate $N_{\mathcal{A}}{}^T \hat{J}_u$ minimizes the loss for the constrained case.

It is important to note that Equation (19) is valid when the nominal point (u^*, d^*) is an *optimal unconstrained* reference point. If the reference point has a non-zero gradient, the optimal gradient estimate takes the form (the reader is referred to Appendix B for a derivation of this expression):

$$\hat{J}_u = H(y_m - y^*) + J_u^* \tag{20}$$

where $J_u^* = J_u(u^*, d^*)$ (obtained from the nonlinear model). Note here that both (19) and (20) can be written in the form:

$$\hat{J}_u = H y_m - c_s \tag{21}$$

where c_s is a constant (see Figure 1).

The simple gradient estimate in (19) and (20) avoids implementing a model-based estimator, for example, a dynamic Kalman filter, and thus greatly simplifies the practical use of feedback-based real-time optimization, which is based on the first-order KKT condition (2a).

The gradient estimate can be used in a wide array of RTO control applications. In particular, it may be used in the following approaches for optimal steady-state operation with changes in active constraints:

1. *Primal-dual approaches* [28] based directly on the optimality condition (2a) with a (slow) update of the Lagrange multiplier λ. This may be done using a slow controller K_{dual} which controls the measured constraints by manipulating the dual variables (λ) and with max-selectors for switching active constraints, see Figure 2 [29, 30].

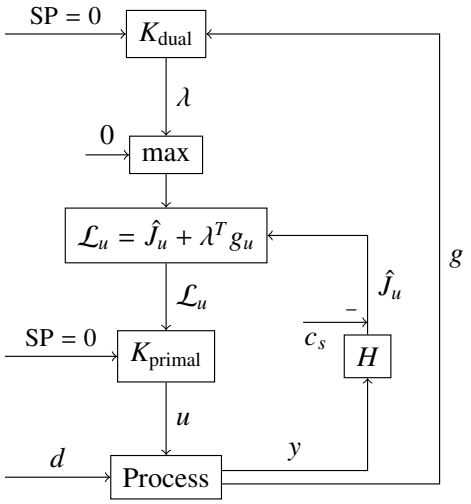


Figure 2: Primal-dual optimizing control structure using the proposed gradient estimate. The controller K_{dual} is always diagonal (decentralized), whereas the controller K_{primal} may be multivariable or diagonal.

2. *Region-based control* [26, 2] where we in each region i control the active constraints and the associated reduced gradient $N_{\mathcal{A},i}^T \hat{J}_u$ to zero, see Figure 3.

2A. Region-based control may be applied to multivariable control, for example, model predictive control, by changing the cost function for designing the controller for each region, according to Bernardino and Skogestad [31]. There, the gradient estimate is also used for constraint switching.

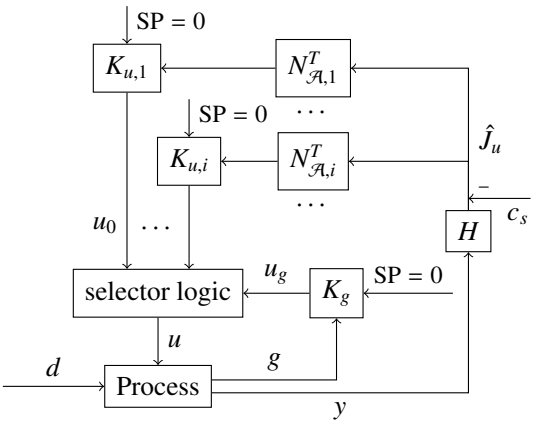


Figure 3: Region-based optimizing control structure using the proposed gradient estimate. In this scheme, each projection matrix $N_{\mathcal{A},i}$ is linked to a different set of active constraints $\mathcal{A}_i$, and the resulting gradient projection $N^T_{\mathcal{A}_i}\hat{J}_u$ is controlled by a different controller $K_{u,i}$ (which in general is multivariable). If $n_u \geq n_g$, a fixed projection matrix can be used for all $\mathcal{A}_i$, and simple max/min-selectors can be used (see Figure 4).

2B. Decentralized region-based control with constraint switching using selectors [32, 33] (Figure 4). This approach requires at least as many inputs (degrees of freedom) as constraints, that is, $n_u \geq n_g$. An example of its application is given next.

In summary, the cost gradient estimate presented in Equation (20) can be used in a wide array of control applications focused on optimal operation, eliminating the need for a dynamic state estimator and thus greatly simplifying implementation.

6. Example 1: Decentralized region-based control

Here, we consider a system with more inputs than constraints ($n_u \geq n_g$) and design a region-based decentralized control structure with simple min/max-selectors (Figure 4) that minimizes the loss in all active constraint regions [33]. In order to use simple switching, the nullspace associated with the unconstrained gradients (Theorem 2) needs to be selected in accordance with the constraint directions. This is done using the following steps [33]:

- Define N^0 as an orthonormal basis for the nullspace of g_u, such that $g_u N^0 = 0$;
- Find $W = \begin{bmatrix} g_u \\ N^{0^T} \end{bmatrix}^{-1}$, and define the vectors N_i, $i = 1, \cdots, n_g$ as the first n_g normalized columns of W.

Then, controlling the active constraints g_i, for $i \in \mathcal{A}$ and the remaining unconstrained degrees of freedom $N_i^T J_u$, for $i \notin \mathcal{A}$, and $N^{0^T} J_u$ will lead to optimal operation [33]. The final simple decentralized control system with min or max selectors can be implemented as shown in Figure 4 where all controllers (K) are single-input single-output (SISO), for example, PID controllers. The controllers linked to selectors must have anti-windup action, to cancel the integral action when the controllers are inactive.

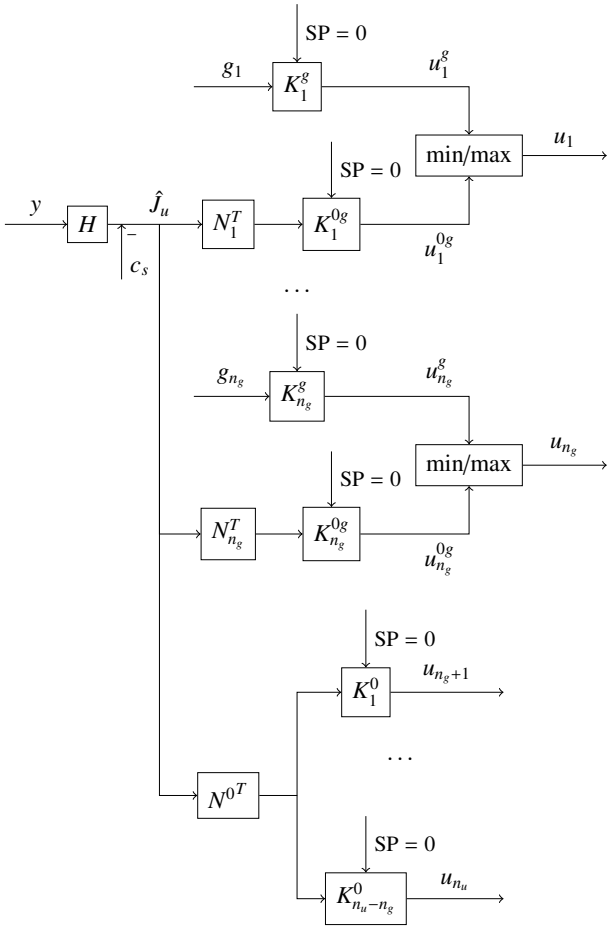


Figure 4: Decentralized region-based optimizing control structure using SISO controllers and selectors.

As a case study, we consider a linear dynamic system with a quadratic cost function given by:

$$\min_u \quad \frac{1}{2}x^T \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix} x + \frac{1}{2}u^T \begin{bmatrix} 1 & -0.1 & -0.2 \\ -0.1 & 0.8 & -0.1 \\ -0.2 & -0.1 & 0.3 \end{bmatrix} u \qquad \text{s.t.} \quad \begin{cases} g_1 = x_1 - 0.8x_2 \leq 0 \\ g_2 = u_1 + u_2 + u_3 \leq 0 \end{cases} \tag{22}$$

$$\dot{x} = \begin{bmatrix} -\frac{1}{\tau_1} & 0 \\ 0 & -\frac{1}{\tau_2} \end{bmatrix} x + \begin{bmatrix} \frac{0.2}{\tau_1} & 0 & 0 \\ 0 & \frac{0.2}{\tau_2} & 0 \end{bmatrix} u + \begin{bmatrix} \frac{1}{\tau_1} & 0 \\ 0 & \frac{1}{\tau_2} \end{bmatrix} d \tag{23}$$

with $\tau_1 = 1$ and $\tau_2 = 2$. The set of optimal active constraint regions can be visualized as a function of the two disturbances as shown in Figure 5. Here, the upper left green region is unconstrained and the lower middle grey region is with all constraints being active (and one unconstrained degree of freedom).

For estimating the cost gradient, the following measurements

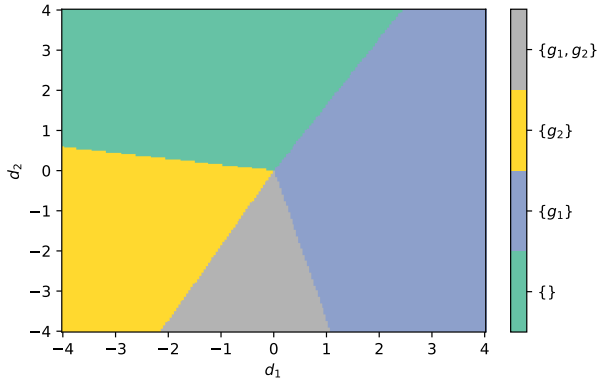


Figure 5: Optimality regions for Example 1

are available:

$$y = \begin{bmatrix} g_1 \\ g_2 \\ x_1 \\ x_2 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 1 & -0.8 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} u \tag{24}$$

Note that both constraints and both states are measured. In addition, we choose to include two of the three inputs. The expected static disturbance and noise magnitudes are $W_d = \text{diag}([4,\ 4])$ and $W_{n^y} = \text{diag}([0,\ 0,\ 1,\ 2,\ 1.5,\ 5])$. The two first zeros in W_{n^y} imply that the constraints have no static measurement error, that is, the constraints can be perfectly controlled. In general, static measurement error for a constraint may be counteracted by using back-off for its setpoint, but this issue is not explored in the case study.

To find the optimal cost gradient estimate using the formulation proposed in this work, we first use (23) with $\dot{x} = 0$ to derive the steady-state relationship:

$$x = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & 0.2 & 0 \end{bmatrix} u + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} d \tag{25}$$

This is used to eliminate the states x from the problem (22), resulting in the following steady-state optimization problem:

$$\begin{aligned} \min_u \quad & J = \frac{1}{2} u^T \underbrace{\begin{bmatrix} 1.04 & -0.1 & -0.2 \\ -0.1 & 1.2 & -0.1 \\ -0.2 & -0.1 & 0.3 \end{bmatrix}}_{J_{uu}} u + u^T \underbrace{\begin{bmatrix} 0.2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}}_{J_{ud}} d \\ \text{s.t.} \quad & g = \underbrace{\begin{bmatrix} 0.2 & -0.16 & 0 \\ 1 & 1 & 1 \end{bmatrix}}_{g_u} u + \begin{bmatrix} 1 & -0.8 \\ 0 & 0 \end{bmatrix} d \le 0 \end{aligned} \tag{26}$$

From the matrix g_u, we can find the projections N_i and N^0 to be multiplied with the unconstrained gradient J_u. N^0 is the nullspace of g_u given by:

$$N^0 = \begin{bmatrix} -0.36214 & -0.45268 & 0.81482 \end{bmatrix}^T \tag{27}$$

The vectors N_i are the first n_g normalized columns of $W = \begin{bmatrix} g_u \\ N^{0T} \end{bmatrix}^{-1}$, calculated as:

$$W = \begin{bmatrix} 2.8689 & 0.29508 & -0.36214 \\ -2.6639 & 0.36885 & -0.45267 \\ -0.20491 & 0.33607 & 0.81482 \end{bmatrix} \tag{28}$$

$$N_1 = \begin{bmatrix} 0.73179 & -0.67952 & -0.052271 \end{bmatrix}^T \tag{29}$$

$$N_2 = \begin{bmatrix} 0.50902 & 0.63627 & 0.57971 \end{bmatrix}^T \tag{30}$$

To estimate the gradient from the measurements, we also need their corresponding steady-state model. Plugging the steady-state expression for the states into (24) leads to:

$$y = \underbrace{\begin{bmatrix} 0.2 & -0.16 & 0 \\ 1 & 1 & 1 \\ 0.2 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{G^y} u + \underbrace{\begin{bmatrix} 1 & -0.8 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}}_{G^y_d} d \tag{31}$$

The optimal sensitivity is then:

$$F = \frac{dy^{opt}}{dd} = G^y_d - G^y J_{uu}^{-1} J_{ud} = \begin{bmatrix} 0.9599 & -0.5830 \\ -0.4207 & -2.8867 \\ -0.0065 & 0.6479 \\ -0.0324 & -1.7605 \\ -0.1618 & -0.8026 \\ 0.9547 & -0.0647 \end{bmatrix} \tag{32}$$

With this information and the matrices from Equation (26), we can calculate the measurement combinations H^J from Equation (13), which gives:

$$H^J = \begin{bmatrix} 0.2741 & 0.9842 & 0.1560 & -1.0715 & -1.1842 & 0.0050 \\ -0.1897 & -0.0735 & 1.7813 & 0.8869 & -0.0265 & 0.0570 \\ -0.0180 & -0.1964 & -0.0091 & 0.0953 & 0.4964 & -0.0003 \end{bmatrix} \tag{33}$$

and the estimated gradient is $\hat{J}_u = H^J(y - y^*) = H^J - c_s$. Here, we note that the approximations in (3) and (4) are exact for this example, and therefore H^J does not depend on the nominal point to be considered. However, we still need a reference point to calculate the constant $c_s = H^J y^*$, and for that, we choose an optimal point with $d^* = [0,\ 0]^T$. This gives $c_s = [0,\ 0,\ 0]^T$.

Dynamic simulation results for the closed-loop system with the proposed control structure in Figure 4 with $H = H^J$ are shown in Figure 6. The PI controllers tuning are given in Table 1. The simulated disturbances cover all four active constraint regions but we did not include measurement noise. The responses are fairly smooth (see the three input profiles) and there are as expected three changes in active constraints. The gradient estimate with $H = H^J$ is optimal in terms of minimizing the average loss with the expected (assumed) disturbances and noise. However, this means that the gradient estimates (and

resulting CVs) are not designed to reject the disturbances completely, as they simultaneously try to reduce the effect of measurement noise. This is the reason why the resulting steady-state inputs u_i (blue lines) do not match exactly the corresponding optimal values (magenta dashed lines). At steady state, the economic loss L resulting from this input mismatch is, however, very small.

Controller	Parameter	Value
K_1^g	K_c	50
	K_I	50
K_2^g	K_I	100
K_1^{0g}	K_I	-1.191
K_2^{0g}	K_I	1.528
K^0	K_I	2.761

Table 1: Proportional and integral gains of controllers for Example 1. All controllers have anti-windup with tracking time $\tau_T = 0.01$.

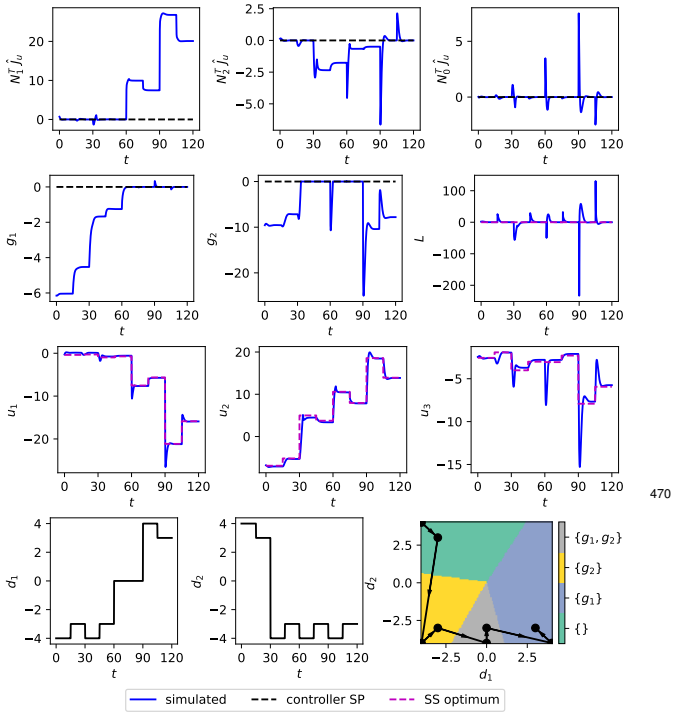
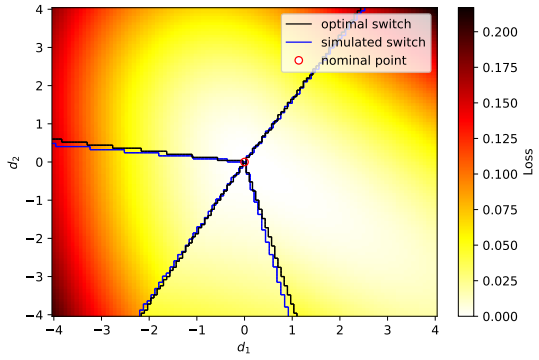
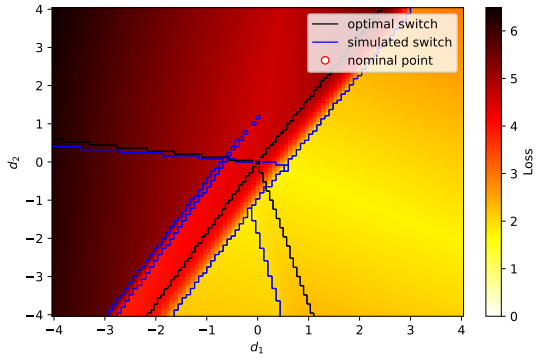


Figure 6: Dynamic simulation over all active constraint regions using the proposed control structure with $H = H^J$ (exact local method) (Example 1).

In Figure 7, we present the steady-state loss obtained in closed loop both without and with static measurement noise (bias). The loss is shown as a heatmap for each disturbance combination. The much larger loss (note the difference in scale) with measurement noise in Figure 7b is for the worst-case measurement error satisfying $n^y = W_{n^y} n^{y\prime}$ with $\|n^{y\prime}\|_2 \leq 1$. The optimal active constraint regions (same as Figure 5) are shown by black lines whereas the actual operating regions resulting from using the control structure are shown by blue lines. Note that the constraint switching is moved away from the optimal, which is not surprising (see discussion).



(a) Loss without measurement error ($n^y = 0$).



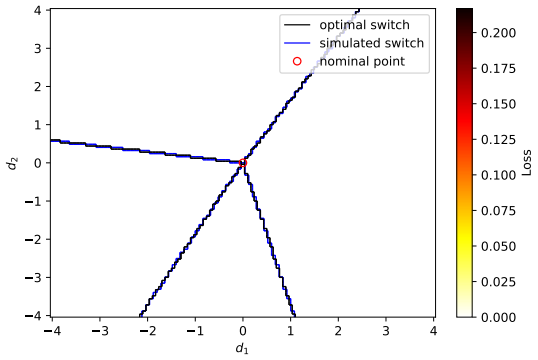
(b) Worst-case loss with measurement error. The new narrow operating region (which starts from point $d = [-2.8,\ -4]^T$) has both constraints g_1 and g_2 active.

Figure 7: Steady-state loss for closed-loop operation with $H = H^J$ from the exact local method (Example 1).

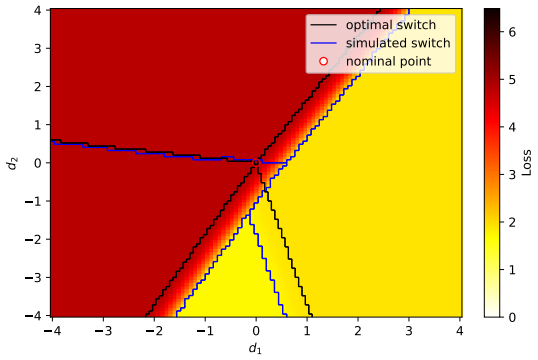
Figure 7a shows that the measurement combination $H = H^J$ (which is based on the exact local method of self-optimizing control) does not perfectly reject disturbances, even without measurement error. To achieve zero loss for disturbances, H must be in the nullspace of F. For instance, if we apply the extended nullspace method (15) to this problem (with $M_n = J_{uu}^{-1/2}$), we get:

$$H = \begin{bmatrix} 0.195 & 1 & 0.156 & -1.1 & -1.2 & 0.005 \\ -0.0624 & -0.1 & 1.95 & 0.9 & 0 & 0.0624 \\ 0 & -0.2 & 0 & 0.1 & 0.5 & 0 \end{bmatrix} \quad (34)$$

With the resulting gradient estimate (and set of CVs), the steady-state closed loop loss for the extended nullspace method (without noise and with the worst-case noise) are presented in Figure 8. We see that in the case without noise (Figure 8a), the economic loss is exactly zero in all constraint regions. This is expected since the original problem is linear with a quadratic cost. However, we see that the exact local method (H^J) is better at locally rejecting noise (note that the worst-case loss in Figure 7b is smaller around the nominal point), but the extended nullspace method (Figure 8b) handles large disturbances better, as expected.



(a) Zero loss without measurement error ($n^y = 0$).



(b) Worst-case loss with measurement error.

Figure 8: Steady-state loss for closed-loop operation with H from the extended nullspace method (Example 1)

7. Example 2: Williams-Otto reactor

This case study is a well-known benchmark for process control introduced by Williams and Otto [34]. The reactor is illustrated in Figure 9.

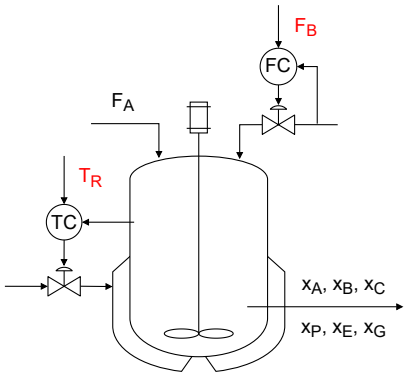


Figure 9: Schematic representation of Williams-Otto reactor with control inputs in red (Example 2).

The following chemical reactions take place in the system:

$$A + B \xrightarrow{k_1} C \qquad k_1 = k_{0,1} e^{-E_1/T_r}$$
$$C + B \xrightarrow{k_2} P + E \qquad k_2 = k_{0,2} e^{-E_2/T_r}$$
$$P + C \xrightarrow{k_3} G \qquad k_3 = k_{0,3} e^{-E_3/T_r}$$

The component mass balances for the six components give the following set of ODEs:

$$\frac{dx_A}{dt} = \frac{F_A}{W} - \frac{(F_A + F_B)x_A}{W} - k_1 x_A x_B \tag{35a}$$
$$\frac{dx_B}{dt} = \frac{F_B}{W} - \frac{(F_A + F_B)x_B}{W} - k_1 x_A x_B - k_2 x_C x_B \tag{35b}$$
$$\frac{dx_C}{dt} = -\frac{(F_A + F_B)x_C}{W} + 2k_1 x_A x_B - 2k_2 x_C x_B - k_3 x_P x_C \tag{35c}$$
$$\frac{dx_P}{dt} = -\frac{(F_A + F_B)x_P}{W} + k_2 x_C x_B - 0.5 k_3 x_P x_C \tag{35d}$$
$$\frac{dx_E}{dt} = -\frac{(F_A + F_B)x_E}{W} + 2k_2 x_C x_B \tag{35e}$$
$$\frac{dx_G}{dt} = -\frac{(F_A + F_B)x_G}{W} + 1.5 k_3 x_P x_C \tag{35f}$$

The steady-state optimization problem to be considered for this system is:

$$\begin{aligned} \min_u \; & J = p_A F_A + p_B F_B - (F_A + F_B)\left[p_P(1 + \Delta p_P)x_P + p_E x_E\right] \\ \text{s.t.} \; & g_1 = x_E - 0.30 \le 0 \\ & g_2 = x_A - 0.12 \le 0 \end{aligned} \tag{36}$$

The inputs considered for this example are $u = \begin{bmatrix} F_B & T_r \end{bmatrix}^T$. The active constraint regions of the problem are shown in Figure 10 as a function of the disturbances $d = \begin{bmatrix} F_A & \Delta p_P \end{bmatrix}^T$. The vector of available measurements is:

$$y = \begin{bmatrix} g_1 & g_2 & x_B & x_C & x_P & x_G & \Delta p_P \end{bmatrix}^T$$

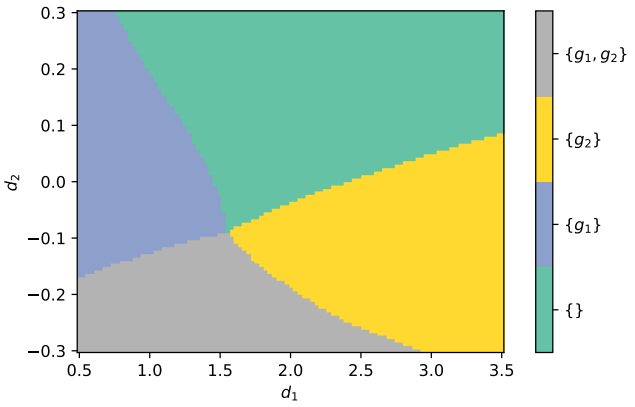


Figure 10: Active constraint regions as a function of disturbances for example 2.

Similarly to the first example, we calculate the optimal measurement combinations to be used for gradient estimation. For that, we approximate Equation (36) as a quadratic programming (QP) problem around a nominal point, here considered as the optimal point for $d^* = \begin{bmatrix} 2 & 0 \end{bmatrix}^T$. To scale the disturbances and measurement error, W_d was chosen as $W_d = \text{diag}\left(\begin{bmatrix} 1.5 & 0.3 \end{bmatrix}\right)$, and W_{n^y} was chosen as the maximum deviation between the

approximate and the true model predictions, which resulted in $W_{n^y} = \mathrm{diag}\left(\left[0, 0, 0.076, 0.0089, 0.0056, 0.038, 0\right]\right)$.

The optimal measurement combinations for the exact local method and the extended nullspace method are, respectively:

$$H^J = \begin{bmatrix} -1388 & 136.5 \\ -508 & -5.5 \\ 6.57153 & -1.71026 \\ 143.648 & -37.3849 \\ 786.6 & -204.715 \\ -51.2488 & 13.3377 \\ -116 & 0.6875 \end{bmatrix}^T \tag{37}$$

$$H = \begin{bmatrix} -1363.26 & 129.003 \\ -511.492 & -4.98053 \\ 8.00163 & -2.08245 \\ 174.909 & -45.5206 \\ 957.78 & -249.265 \\ -62.4016 & 16.2402 \\ -115.267 & 0.428895 \end{bmatrix}^T \tag{38}$$

The steady-state loss obtained in closed loop with the control structure proposed in Figure 4 and using the exact local method ($H = H^J$) is shown in Figure 11 as a function of the disturbances. Similarly, the results for using the extended nullspace method are shown in Figure 12. The economic performance of both methods is reasonable (for comparison, the optimal cost at the nominal point is $J^* = -88.24$). For this example, choosing the exact local method results in a smaller maximum loss, which comes at the expense of having a higher loss than the extended nullspace method for certain disturbance realizations.

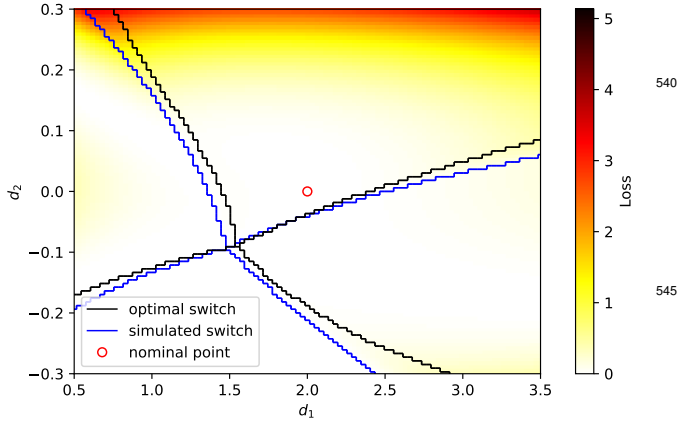


Figure 11: Steady-state loss for closed-loop operation with $H = H^J$ from the exact local method for example 2.

8. Discussion

8.1. Local gradient estimation (block-diagonal H)

The matrix H^J in (14) for the optimal gradient estimate is a full matrix. This means that control systems in Figures 2 and 4 may not be decentralized, even if the controllers K themselves are decentralized. To obtain a decentralized control system, the relationship from y (measurements) to u (inputs) needs to be decoupled. For example, if we have a complex process with many units, then decentralized control implies that only measurements from unit k should be used by the control system to compute the inputs for unit k. To accomplish this, the matrix H needs to be block-diagonal. There exists no analytical solution in this case so the optimal block-diagonal H must be obtained numerically. Depending on the case study, there may be a small or large performance loss compared to using a full H. This problem has been studied in detail by [23] using mixed-integer quadratic programming (MIQP). However, their objective was to find self-optimizing controlled variables $c = Hy_m + c_s$, so their results need to be modified to estimate instead the gradient, $\hat{J}_u = Hy_m + c_s$.

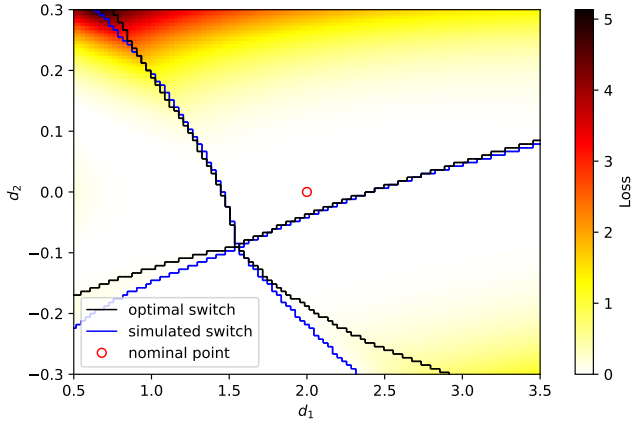


Figure 12: Steady-state loss for closed-loop operation with H from the extended nullspace method for example 2.

Finally, note that for the primal-dual optimizing control structure in Figure 2, the Lagrange multiplier (λ) may introduce a coupling from the measured constraint (g) to the inputs (u) even for cases where the gradient controller (K_{primal}) is diagonal and the gradient estimator (matrix H) is block-diagonal.

8.2. Addition of RTO layer and model mismatch

The optimality of the static gradient estimate is based on a quadratic approximation (3) of the cost, and a linear approximation of the constraints (4) and of the measurement model (9). In general, these assumptions are not satisfied, and in this case, a static real-time optimization layer may be used to provide updates of the constants presented in this work, namely the controller setpoints c_s, the measurement combinations H, and the projection matrices N_i and N^0 (or $N_{\mathcal{A}}$ when generalizing to centralized approaches).

Using the RTO layer to update the setpoints c_s is the simplest and most important, being sufficient to drive the system to optimality in a new operating condition. That is, c_s is optimally updated, while the matrix H and the projection matrices constant can be kept constant. The use of constant matrices implies the self-optimizing properties (related to optimality on a shorter time scale) may degrade somewhat in a new operating point. On the other hand, changing these matrices will affect the control problem and, consequently, the controllers' tuning

that should be used. Thus, updating only c_s is recommended in most practical applications.

As an alternative to a model-based RTO layer, an upper data-based layer based on perturbing the process, for example, extremum-seeking control, may be used to update c_s. However, data-based methods are not realistic for most process control applications because the convergence of these methods is too slow to track changing disturbances. Regardless, these methods are complementary to the method discussed in this work, as they are applied on an upper layer.

We remark that these RTO updates address the mismatch between the model used for the design of the gradient estimate and the plant behavior, but the improvement obtained is often small. In fact, for the unconstrained part of the optimization problem, model uncertainty is usually not critical as long as we are reasonably close to the optimum where the cost function is flat. Regardless of this, even though model mismatch is an important challenge for RTO problems to ensure exact plant optimality, it is not the main challenge that hinders its implementation. The main challenges for practical RTO implementation are slow convergence and numerical problems in performing disturbance estimation and optimization for given disturbances and high costs of implementation and maintenance.

8.3. Required model information

The methods for self-optimizing control used to obtain the matrix H for the gradient estimate use model information only offline. Furthermore, they only need model information in the form of the sensitivity matrix F and the gain matrix G^y, both of which can be estimated from plant data or steady-state simulations with relative ease. For estimating the gradient J_u, we additionally require knowledge of the Hessian matrix J_{uu} so that the directions of the unconstrained gradient are retrieved. The Hessian is harder to estimate from measurement information, as it requires more data. In addition, the constraint gradient g_u is needed to find the nullspace matrix for the reduced gradient $N_{\mathcal{A}}^T J_u$, but g_u is easy to estimate from data. However, if a steady-state model is available for control structure design, all of these matrices can easily be obtained

8.4. Discussion of example 1

In this work, we illustrate the method with a case study where the formulation is exact, that is, Equations (3), (4) and (9) hold. It was shown that the exact local method (13) is not designed to perfectly reject disturbances, that is $\Delta c_{opt} = HF\Delta d \neq 0$, which results in non-zero loss as shown in Figure 7a. Therefore, if a new estimate of the disturbances is available, an update of c_s will lead to improved performance around the new operating point, even if the optimal H is unchanged. This is not the case for the extended nullspace method, where we see in Figure 8a that the obtained loss is zero for all disturbance values, which means that the optimal setpoint value is constant, i.e. $\Delta c_{opt} = HF\Delta d = 0$.

We see from Figure 7 that measurement bias has a comparatively bigger effect on the economic loss than the disturbances in this numerical example, which is worsened the further the disturbances are from their design value. We also see that the measurement bias may trigger control of constraints that are not optimally active, which could be a problem if there were no constraint controllers. This is the reason why the pattern of the operating regions is so different from the optimal in Figure 7b. Overall, we see that for the nominal case (Figure 7a), optimal behavior is well captured, with the closed-loop operating regions closely resembling the optimal active constraint regions. For the worst-case loss (Figure 7b), the resulting economic loss is still small when compared to the values attained dynamically in Figure 6.

8.5. Discussion of example 2

This example illustrates the application of the proposed method to nonlinear systems. If a nonlinear model is available, it must be first locally approximated as a QP problem with linear measurements, see Equations (3), (4) and (9). With this, the method is applied in the same way as described for example 1, and the results illustrate the obtained steady-state performance. It is no longer possible to attain perfect disturbance rejection as there is model mismatch, but the performance can still be deemed acceptable, with small losses in the neighborhood of the design point. The use of a gradient estimate based on the exact local method led to a lower maximum loss in the domain of interest when compared to the extended nullspace method, which favors the effect of the modeled disturbances over the possible measurement errors.

The values chosen for W_d and W_{n^y} reflect the expected behavior of the real system when compared to that of the approximate model used for design. This was possible because the nonlinear model was available for evaluation. In practice, these are tuning parameters that weight the importance of disturbance rejection versus the presence of measurement error, and they can therefore be chosen based on process knowledge.

9. Conclusion

The optimal local gradient estimate for use in steady-state real-time optimization is simply $\hat{J}_u = H^J(y_m - y^*) + J_u^*$ with H^J as in Equation (13) (Theorem 1). This gradient estimate is optimal also in the constrained case when used with the KKT optimality conditions (2) (Theorem 2). The gradient estimate $\hat{J}_u$ may be used in a multitude of control applications (Figures 1 to 4) where it is desired to include the optimality conditions (2) directly into the feedback control layer.

Funding

This work was funded by the Research Council of Norway through the IKTPLUSS programme (project number 299585).

Appendix A Proof: Optimal gradient estimate for the constrained case

We begin by describing the loss function for the constrained optimization problem, resulting in a simple form. Then, we

show that the ideal variables for a given set of active constraints are the projection of the unconstrained gradient estimate onto the nullspace of the gradient of the active constraints, in the sense that they minimize the expected loss.

A.1 Loss for constrained optimization problem

From Equation (3), we have:

$$\begin{aligned} L &= J(u,d) - J^{opt}(d) = J_u^{*T}(u-u^{opt}) \\ &\quad + \frac{1}{2}(u-u^*)^T J_{uu}(u-u^*) + (d-d^*)^T J_{ud}^T(u-u^{opt}) \\ &\quad - \frac{1}{2}(u^{opt}-u^*)^T J_{uu}(u^{opt}-u^*) \end{aligned}$$

$$\begin{aligned} L = &\left(J_u^* + J_{ud}(d-d^*)\right)^T (u-u^{opt}) + \frac{1}{2}(u-u^*)^T J_{uu}(u-u^*) \\ &- \frac{1}{2}(u^{opt}-u^*)^T J_{uu}(u^{opt}-u^*) \end{aligned} \tag{39}$$

The optimality conditions state that:

$$\begin{aligned} &\mathcal{L}_u(u^{opt}, d, \lambda^{opt}) = J_u(u^{opt}, d) + g_u^{*T}\lambda^{opt} = 0 \\ &\implies J_u^* + J_{uu}(u^{opt}-u^*) + J_{ud}(d-d^*) + g_u^{*T}\lambda^{opt} = 0 \\ &\implies J_u^* + J_{ud}(d-d^*) = -(J_{uu}(u^{opt}-u^*) + g_u^{*T}\lambda^{opt}) \end{aligned} \tag{40}$$

We can therefore rewrite Equation (39) as:

$$\begin{aligned} L = &-\left(J_{uu}(u^{opt}-u^*) + g_u^{*T}\lambda^{opt}\right)^T (u-u^{opt}) \\ &+ \frac{1}{2}(u-u^*)^T J_{uu}(u-u^*) - \frac{1}{2}(u^{opt}-u^*)^T J_{uu}(u^{opt}-u^*) \\ = &-\lambda^{opt T} g_u^*(u-u^{opt}) + \frac{1}{2}(u-u^*)^T J_{uu}(u-u^*) \\ &- \frac{1}{2}(u^{opt}-u^*)^T J_{uu}(u^{opt}-u^*) - (u^{opt}-u^*)^T J_{uu}(u-u^{opt}) \\ = &-\lambda^{opt T} g_u^*(u-u^{opt}) + \frac{1}{2}(u-u^*)^T J_{uu}(u-u^*) \\ &- \frac{1}{2}(u^{opt}-u^*)^T J_{uu}(u^{opt}-u^*) - (u^{opt}-u^*)^T J_{uu}(u-u^*) \\ &+ (u^{opt}-u^*)^T J_{uu}(u^{opt}-u^*) \\ = &-\lambda^{opt T} g_u^*(u-u^{opt}) + \frac{1}{2}(u-u^*)^T J_{uu}(u-u^*) \\ &- (u^{opt}-u^*)^T J_{uu}(u-u^*) + \frac{1}{2}(u^{opt}-u^*)^T J_{uu}(u^{opt}-u^*) \end{aligned}$$

From this, we conclude that:

$$L = \frac{1}{2}(u-u^{opt})^T J_{uu}(u-u^{opt}) - \lambda^{opt T} g_u^*(u-u^{opt}) \tag{41}$$

This expression is very similar to Equation (7), the difference being the linear term $\lambda^{opt T} g_u^*(u-u^{opt})$, which is related to constraint control. Because the optimal Lagrange multipliers for the inactive constraints are zero, we have that $\lambda^{opt T} g_u^*(u-u^{opt}) = \lambda_{\mathcal{A}}^{opt T} g_{u,\mathcal{A}}(u-u^{opt})$, with $g_{u,\mathcal{A}}$ defined as the gradient of the active constraints with respect to the inputs. If the optimal active constraint set $\mathcal{A}$ is perfectly controlled, we have:

$$\begin{aligned} &\begin{cases} g_{\mathcal{A}}(u^{opt}, d) = g_{\mathcal{A}}^* + g_{u,\mathcal{A}}(u^{opt}-u^*) + g_{d,\mathcal{A}}(d-d^*) = 0 \\ g_{\mathcal{A}}(u, d) = g_{\mathcal{A}}^* + g_{u,\mathcal{A}}(u-u^*) + g_{d,\mathcal{A}}(d-d^*) = 0 \end{cases} \\ &\implies g_{u,\mathcal{A}}(u-u^{opt}) = 0 \end{aligned} \tag{42}$$

This means that only the quadratic term on Equation (41) is relevant when the correct constraints are controlled, with the additional restriction on the allowed directions of $(u-u^{opt})$, which are in the nullspace of $g_{u,\mathcal{A}}$. Define $N_{\mathcal{A}}$ as a basis for the nullspace of $g_{u,\mathcal{A}}$. This means that the loss from Equation (41) is further simplified when the correct constraints are controlled to give:

$$L = \frac{1}{2}(u-u^{opt})^T J_{uu}(u-u^{opt}) = \frac{1}{2} w^T N_{\mathcal{A}}{}^T J_{uu} N_{\mathcal{A}} w \tag{43}$$

Here, w is an appropriately sized vector that represents the unconstrained degrees of freedom.

A.2 Connection with the unconstrained problem

We now show that the ideal controlled variables for this problem are directly linked to the ones from the unconstrained problem. First, note that the matrix $J_{ww} = N_{\mathcal{A}}{}^T J_{uu} N_{\mathcal{A}}$ is invertible by definition, and therefore we can write:

$$L = \frac{1}{2} w^T J_{ww} J_{ww}^{-1} J_{ww} w \tag{44}$$

From this, we can see that the loss variable z_w for this problem can be represented by:

$$z_w = J_{ww}^{-1/2} N_{\mathcal{A}}{}^T J_{uu} N_{\mathcal{A}} w = J_{ww}^{-1/2} N_{\mathcal{A}}{}^T J_{uu}(u-u^{opt}) \tag{45}$$

Similarly to Equation (8), we can write z_w in terms of the unconstrained CVs c as:

$$\begin{aligned} z_w &= J_{ww}^{-1/2} N_{\mathcal{A}}{}^T J_{uu}(HG^y)^{-1}(c-c^{opt}) \\ &= J_{ww}^{-1/2} N_{\mathcal{A}}{}^T J_{uu}^{1/2} M_n (c-c^{opt}) \end{aligned}$$

We can similarly write z_w in terms of the unconstrained gradient:

$$\begin{aligned} &J_u = J_u(u^{opt}, d) + J_{uu}(u-u^{opt}) \\ &\implies z_w = J_{ww}^{-1/2} N_{\mathcal{A}}{}^T (J_u - J_u(u^{opt}, d)) \end{aligned}$$

Note that, because of the optimality conditions, we have that:

$$J_u(u^{opt}, d) + g_u^{*T}\lambda^{opt} = 0 \implies N_{\mathcal{A}}{}^T J_u(u^{opt}, d) = 0$$

and with the choice of $M_n = J_{uu}^{-1/2}$, we compare both expressions for z_w and we see that:

$$
\begin{aligned}
{N_{\mathcal{A}}}^T J_u &= {N_{\mathcal{A}}}^T (c - c^{opt}) \\
&= {N_{\mathcal{A}}}^T \left(H(y_m - y^*) - Hn^y - H(y^{opt}(d) - y^*) \right)
\end{aligned} \tag{46}
$$

This formulation is similar to that of Equation (18), with the exception that now $u^{opt}(d)$ and $y^{opt}(d)$ represent a constrained optimal point, and therefore are a different function of the disturbances, $(y^{opt}(d) - y^*) = F_{\mathcal{A}}(d - d^*)$. We can determine $F_{\mathcal{A}}$ from the constrained optimization problem as follows:

$$
\begin{bmatrix} J_{uu} & g_{u,\mathcal{A}}^T \\ g_{u,\mathcal{A}} & 0 \end{bmatrix} \begin{bmatrix} \Delta u^{opt} \\ \Delta \lambda_{\mathcal{A}}^{opt} \end{bmatrix} = \begin{bmatrix} -J_{ud} \\ -g_{d,\mathcal{A}} \end{bmatrix} \Delta d \tag{47}
$$

First we eliminate Δu^{opt} by premultiplying both sides by $\begin{bmatrix} g_{u,\mathcal{A}} J_{uu}^{-1} & -I \end{bmatrix}$, leading to the solution $\Delta \lambda_{\mathcal{A}}^{opt} = W_{\mathcal{A}} \Delta d$, where

$$
W_{\mathcal{A}} = \left(g_{u,\mathcal{A}} J_{uu}^{-1} g_{u,\mathcal{A}}^T \right)^{-1} \left(g_{d,\mathcal{A}} - g_{u,\mathcal{A}} J_{uu}^{-1} J_{ud} \right)
$$

The solution for the new optimal inputs follows as $\Delta u^{opt} = -J_{uu}^{-1} \left((g_{u,\mathcal{A}})^T W_{\mathcal{A}} + J_{ud} \right) \Delta d$, and the optimal sensitivity matrix $F_{\mathcal{A}}$ can be obtained as:

$$
F_{\mathcal{A}} = F - G^y J_{uu}^{-1} g_{u,\mathcal{A}}^T W_{\mathcal{A}} \tag{48}
$$

with F being the unconstrained optimal sensitivity matrix. The second term of $F_{\mathcal{A}}$ is related to constraint control, and we can see that, with $M_n = J_{uu}^{-1/2}$:

$$
\begin{aligned}
{N_{\mathcal{A}}}^T H F_{\mathcal{A}} &= {N_{\mathcal{A}}}^T H F - {N_{\mathcal{A}}}^T \overbrace{H G^y J_{uu}^{-1}}^{=I} g_{u,\mathcal{A}}^T W_{\mathcal{A}} \\
&= {N_{\mathcal{A}}}^T H F - \overbrace{{N_{\mathcal{A}}}^T g_{u,\mathcal{A}}^T}^{=0} W_{\mathcal{A}} = {N_{\mathcal{A}}}^T H F
\end{aligned}
$$

This means that the last two terms in Equation (46) are minimized by the unconstrained self-optimizing control solution for $H = H^J$ (13), and therefore the reduced gradient estimate

$$
{N_{\mathcal{A}}}^T \hat{J}_u = {N_{\mathcal{A}}}^T H^J (y_m - y^*) \tag{49}
$$

is the unconstrained CV that should be kept at zero to minimize the expected norm of z_w.

Appendix B Effect of nominal setpoint

Here, we evaluate the effect of having a non-optimal reference point. From Equation (16) and choosing $M_n = J_{uu}^{-1/2}$, we have:

$$
c(u, d) - c(u^{opt}(d), d) = J_u(u, d) - J_u(u^{opt}(d), d)
$$

The same expression is valid for the nominal point, according to:

$$
c(u^*, d^*) - c(u^{opt}(d^*), d^*) = J_u(u^*, d^*) - J_u(u^{opt}(d^*), d^*)
$$

Here, we assume that $u^* \neq u^{opt}(d^*)$, that is, the nominal point is not optimal. For the unconstrained problem, $J_u(u^{opt}(d), d) = J_u(u^{opt}(d^*), d^* = 0$, and we subtract the two equations to give:

$$
\begin{aligned}
J_u(u, d) = J_u(u^*, d^*) &+ c(u, d) - c(u^*, d^*) \\
&- \left(c(u^{opt}(d), d) - c(u^{opt}(d^*), d^*) \right)
\end{aligned}
$$

or

$$
J_u(u, d) = J_u(u^*, d^*) + H(y^m - y^*) - H \underbrace{(y_m - y)}_{n^y} - HF(d - d^*) \tag{50}
$$

Choosing the exact local method solution for H from (13), we minimize the last two terms from the previous equation, and the optimal gradient estimate to be controlled is given by:

$$
\hat{J}_u(u, d) = H^J (y^m - y^*) + J_u(u^*, d^*)
$$

as stated in Equation (20). As previously shown, this gradient estimate is also valid for the constrained region, with the corresponding reduced gradient estimate being the optimal variable to be controlled.

References

[1] S. Engell, Feedback control for optimal process operation, Journal of process control 17 (2007) 203–219.

[2] D. Krishnamoorthy, S. Skogestad, Real-time optimization as a feedback control problem - a review, Computers & Chemical Engineering (2022) 107723.

[3] M. Morari, Y. Arkun, G. Stephanopoulos, Studies in the synthesis of control structures for chemical processes: Part I: Formulation of the problem. Process decomposition and the classification of the control tasks. Analysis of the optimizing control structures, AIChE Journal 26 (1980) 220–232.

[4] S. Skogestad, Plantwide control: The search for the self-optimizing control structure, Journal of process control 10 (2000) 487–507.

[5] J. Nocedal, S. J. Wright, Numerical Optimization, second ed., Springer New York, NY, 2006.

[6] M. M. Câmara, A. D. Quelhas, J. C. Pinto, Performance evaluation of real industrial RTO systems, Processes 4 (2016) 44.

[7] D. Krishnamoorthy, E. Jahanshahi, S. Skogestad, Feedback real-time optimization strategy using a novel steady-state gradient estimate and transient measurements, Industrial & Engineering Chemistry Research 58 (2018) 207–216.

[8] Y. Tan, W. H. Moase, C. Manzie, D. Nešić, I. M. Mareels, Extremum seeking from 1922 to 2010, in: Proceedings of the 29th Chinese control conference, IEEE, 2010, pp. 14–26.

[9] A. Scheinker, 100 years of extremum seeking: a survey, Automatica 161 (2024) 111481.

[10] M. Krstić, H.-H. Wang, Stability of extremum seeking feedback for general nonlinear dynamic systems, Automatica 36 (2000) 595–601.

[11] B. Hunnekens, M. A. Haring, N. van de Wouw, H. Nijmeijer, A dither-free extremum-seeking control approach using 1st-order least-squares fits for gradient estimation, in: 53rd IEEE conference on decision and control, IEEE, 2014, pp. 2679–2684.

[12] Y. Tan, D. Nešić, I. Mareels, On non-local stability properties of extremum seeking control, Automatica 42 (2006) 889–903.

[13] K. T. Atta, M. Guay, R. Lucchese, A geometric phasor extremum seeking control approach with measured constraints, in: 2019 IEEE 58th Conference on Decision and Control (CDC), IEEE, 2019, pp. 1494–1500.

[14] A. G. Marchetti, G. François, T. Faulwasser, D. Bonvin, Modifier adaptation for real-time optimization—methods and applications, Processes 4 (2016) 55.

[15] J. Matias, J. Jäschke, Using a neural network for estimating plant gradients in real-time optimization with modifier adaptation, IFAC-PapersOnLine 52 (2019) 808–813.

[16] M. P. Golden, B. E. Ydstie, Adaptive extremum control using approximate process models, AIChE journal 35 (1989) 1157–1169.

[17] J. Jäschke, S. Skogestad, NCO tracking and self-optimizing control in the context of real-time optimization, Journal of Process Control 21 (2011) 1407–1416.

[18] A. Kashani, S. Panahi, A. Chakrabarty, C. Danielson, Robust data-driven dynamic optimization using a set-based gradient estimator, Optimal Control Applications and Methods (2024).

[19] S. Gros, B. Srinivasan, D. Bonvin, Optimizing control based on output feedback, Computers & Chemical Engineering 33 (2009) 191–198.

[20] V. Alstad, S. Skogestad, E. S. Hori, Optimal measurement combinations as controlled variables, Journal of Process Control 19 (2009) 138–148.

[21] V. Alstad, Studies on selection of controlled variables, Ph.D. thesis, Fakultet for naturvitenskap og teknologi, 2005.

[22] I. J. Halvorsen, S. Skogestad, J. C. Morud, V. Alstad, Optimal selection of controlled variables, Industrial & Engineering Chemistry Research 42 (2003) 3273–3284.

[23] R. Yelchuru, S. Skogestad, Convex formulations for optimal selection of controlled variables and measurements using mixed integer quadratic programming, Journal of Process Control 22 (2012) 995–1007.

[24] J. B. Rosen, The gradient projection method for nonlinear programming. Part I. Linear constraints, Journal of the Society for Industrial and Applied Mathematics 8 (1960) 181–217.

[25] L. S. Lasdon, R. L. Fox, M. W. Ratner, Nonlinear optimization using the generalized reduced gradient method, Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle 8 (1974) 73–103.

[26] J. Jäschke, S. Skogestad, Optimal controlled variables for polynomial systems, Journal of Process Control 22 (2012) 167–179.

[27] G. Torrisi, S. Grammatico, R. S. Smith, M. Morari, A projected gradient and constraint linearization method for nonlinear model predictive control, SIAM Journal on Control and Optimization 56 (2018) 1968–1999.

[28] D. Krishnamoorthy, A distributed feedback-based online process optimization framework for optimal resource sharing, Journal of Process Control 97 (2021) 72–83.

[29] R. Dirza, S. Skogestad, D. Krishnamoorthy, Optimal resource allocation using distributed feedback-based real-time optimization, IFAC-PapersOnLine 54 (2021) 706–711.

[30] R. Dirza, S. Skogestad, Primal-dual feedback-optimizing control with override for real-time optimization, Journal of Process Control 138 (2024) 103208.

[31] L. F. Bernardino, S. Skogestad, Optimal switching of controlled variables for MPC with changes in active constraints, 2024. *In preparation*.

[32] L. F. Bernardino, D. Krishnamoorthy, S. Skogestad, Comparison of simple feedback control structures for constrained optimal operation, IFAC-PapersOnLine 55 (2022) 883–888.

[33] L. F. Bernardino, S. Skogestad, Decentralized control using selectors for optimal steady-state operation with changing active constraints, Journal of Process Control 137 (2024) 103194.

[34] T. J. Williams, R. E. Otto, A generalized chemical processing model for the investigation of computer control, Transactions of the American Institute of Electrical Engineers, Part I: Communication and Electronics 79 (1960) 458–473.