

 Open access • Journal Article • DOI:10.1007/S10107-005-0685-1

An approximation technique for robust nonlinear optimization — [Source link](#)

[Moritz Diehl](#), [Hans Georg Bock](#), [Ekaterina Kostina](#)

Institutions: [Interdisciplinary Center for Scientific Computing](#)

Published on: 15 Jun 2006 - [Mathematical Programming](#) (Springer-Verlag)

Topics: [Robust optimization](#), [Penalty method](#), [Nonlinear programming](#), [Constrained optimization](#) and [Dual norm](#)

Related papers:

- [Open-loop and closed-loop robust optimal control of batch processes using distributional and worst-case analysis](#)
- [Robust Convex Optimization](#)
- [Dynamic optimization of batch processes: II. Role of measurements in handling uncertainty](#)
- [On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming](#)
- [Distributional uncertainty analysis using power series and polynomial chaos expansions](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/an-approximation-technique-for-robust-nonlinear-optimization-5avkc5l1ge>

Moritz Diehl · Hans Georg Bock · Ekaterina Kostina

An approximation technique for robust nonlinear optimization

Received: July 22, 2004 / Accepted: April 20, 2005

Published online: December 30, 2005 – © Springer-Verlag 2005

Abstract. Nonlinear equality and inequality constrained optimization problems with uncertain parameters can be addressed by a robust worst-case formulation that is, however, difficult to treat computationally. In this paper we propose and investigate an approximate robust formulation that employs a linearization of the uncertainty set. In case of any norm bounded parameter uncertainty, this formulation leads to penalty terms employing the respective dual norm of first order derivatives of the constraints. The main advance of the paper is to present two sparsity preserving ways for efficient computation of these derivatives in the case of large scale problems, one similar to the forward mode, the other similar to the reverse mode of automatic differentiation. We show how to generalize the techniques to optimal control problems, and discuss how even infinite dimensional uncertainties can be treated efficiently. Finally, we present optimization results for an example from process engineering, a batch distillation.

Key words. Nonlinear Programming – Robust Optimization – Optimal Control

1. Introduction

In this paper we consider uncertain nonlinear programming problems of the form

$$\min_{x \in \mathbb{R}^{n_x}, u \in \mathbb{R}^{n_u}} f_0(x, u) \quad \text{s.t.} \quad \begin{cases} f_i(x, u) \leq 0 & \text{for } i = 1, \dots, n_f, \\ g_j(x, u, p) = 0 & \text{for } j = 1, \dots, n_x. \end{cases} \quad (1)$$

with an uncertain parameter vector $p \in \mathbb{R}^{n_p}$, and smooth real valued functions $f_0, f_1, \dots, f_{n_f}, g_1, \dots, g_{n_x}$. We partition the optimization variables into the vectors x and u , and the number of equality constraints equals the dimension of x . We assume the jacobian $\frac{\partial g}{\partial x}$ to be invertible everywhere, so that we can regard the variables x as an implicit function of u and p . This division into “states” x and “controls” u arises naturally in model based optimization, where the equalities $g(x, u, p) = 0$ often contain discretized ordinary or partial differential equations, such that we are in particular interested in the case $n_x \gg 1$.

Following a classical approach to address uncertainty, we assume we have some knowledge about the parameter p in the sense that it is restricted to be in a generalized ball

$$\mathbb{P} := \{p \in \mathbb{R}^{n_p} \mid \|p - \bar{p}\| \leq 1\} \quad (2)$$

defined by using a suitable norm $\|\cdot\|$ in \mathbb{R}^{n_p} . This norm may e.g. be chosen to be a scaled Hölder q -norm ($1 \leq q \leq \infty$), $\|p\| = \|A^{-1}p\|_q$, with an invertible $n_p \times n_p$ -matrix A .

This definition of the uncertainty set includes the case of a confidence ellipsoid for a Gaussian random variable p with the expectation value \bar{p} , the variance-covariance matrix Σ , and a scalar $\gamma > 0$ depending on the desired confidence level:

$$\mathbb{P}_{\text{ellipsoid}} = \left\{ p \in \mathbb{R}^{n_p} \mid \left\| \frac{1}{\gamma} \Sigma^{-\frac{1}{2}} (p - \bar{p}) \right\|_2 \leq 1 \right\}.$$

It also includes the case of known upper and lower bounds p_l, p_u for the parameters, that leads to the box

$$\mathbb{P}_{\text{box}} = \left\{ p \mid p_l \leq p \leq p_u \right\} = \left\{ p \mid \left\| \text{diag} \left(\frac{p_u - p_l}{2} \right)^{-1} \left(p - \frac{p_l + p_u}{2} \right) \right\|_\infty \leq 1 \right\}.$$

Note that any product $\mathbb{P} = \mathbb{P}_a \times \mathbb{P}_b$ of two such sets can be expressed by employing the maximum $\|p\| = \|(p_a^T, p_b^T)^T\| = \max\{\|p_a\|_a, \|p_b\|_b\}$ of the two respective norms as a norm.

1.1. Robust counterpart formulation

In order to incorporate the uncertainty in the optimization problem formulation, we choose the classical worst-case formulation of (1). For this aim we assume that the optimizer, who chooses u first, has “nature” as an adverse player who chooses afterwards p and x . Whatever u the optimizer chooses, for each of the functions $f_i(x, u)$, $i = 0, 1, \dots, n_f$, the worst possible value, $\phi_i(u)$, is chosen by the adverse player:

$$\phi_i(u) := \max_{x \in \mathbb{R}^{n_x}, p \in \mathbb{R}^{n_p}} f_i(x, u) \quad \text{s.t.} \quad \begin{cases} g(x, u, p) = 0, \\ \|p - \bar{p}\| \leq 1. \end{cases} \quad (3)$$

Note that the adverse player “nature” is restricted by both the model equations $g(x, u, p) = 0$ and the norm bound on p . Employing the functions $\phi_i(u)$ we arrive at the following worst-case formulation that is often referred to as the “robust counterpart” of (1) (cf. Ben-Tal and Nemirovskii [1]):

$$\text{(RC)} \quad \min_{u \in \mathbb{R}^{n_u}} \phi_0(u) \quad \text{s.t.} \quad \phi_i(u) \leq 0 \quad \text{for } i = 1, \dots, n_f. \quad (4)$$

Due to its bi-level structure, this optimization problem is difficult to solve numerically for general nonlinear functions.

1.2. Expectation value optimization

Alternatively to the worst-case assumption we may also assume that the adverse player “nature” chooses p (and therefore also x) according to an equal probability distribution on \mathbb{P} , i.e. for any function $c(x, u)$ and fixed u we have

$$\mathbb{E}_x\{c(x, u)\} = \frac{\int_{p \in \mathbb{P}} c(g^{-1}(u, p), u) dp}{\int_{p \in \mathbb{P}} dp},$$

where g^{-1} is the hypothetical inverse function of g with respect to x that solves uniquely $g(g^{-1}(u, p), u, p) = 0$ for all u, p .

We may then aim to minimize the expectation value $\mathbb{E}_x\{c(x, u)\}$ of the generalized cost function

$$c(x, u) := f_0(x, u) + \begin{cases} 0 & \text{if } f_i(x, u) \leq 0 \text{ for } i = 1, \dots, n_f, \\ \infty & \text{else.} \end{cases}$$

that penalizes constraint violations by infinite costs. It is straightforward to see that the expectation value $\mathbb{E}_x\{c(x, u)\}$ is infinity for any u that does not satisfy $\phi_i(u) \leq 0$ for all $i = 1, \dots, n_f$. Therefore, the problem $\min_{u \in \mathbb{R}^{n_u}} \mathbb{E}_x\{c(x, u)\}$ can equivalently be written as

$$\min_{u \in \mathbb{R}^{n_u}} \mathbb{E}_x\{f_0(x, u)\} \quad \text{s.t.} \quad \phi_i(u) \leq 0 \quad \text{for } i = 1, \dots, n_f. \quad (5)$$

We refer to this optimization problem – that differs from the robust counterpart (4) only in the objective – as the “expectation value robust counterpart” of (1). It is equally difficult to solve this problem numerically for general nonlinear problems.

2. Approximation of the robust counterparts

In order to avoid the bi-level structure of the robust counterparts (4) (resp. (5)), we propose in this paper to replace the functions $\phi_i, i = 0, \dots, n_f$, by approximations $\tilde{\phi}_i$ that can much easier be evaluated and that will be defined below. The idea is to replace the robust counterpart (4) by the approximation

$$\text{(ARC)} \quad \min_{u \in \mathbb{R}^{n_u}} \tilde{\phi}_0(u) \quad \text{s.t.} \quad \tilde{\phi}_i(u) \leq 0 \quad \text{for } i = 1, \dots, n_f. \quad (6)$$

Likewise, the “expectation value robust counterpart” (5) will be approximated by

$$\min_{u \in \mathbb{R}^{n_u}} \tilde{f}_0(u) \quad \text{s.t.} \quad \tilde{\phi}_i(u) \leq 0 \quad \text{for } i = 1, \dots, n_f, \quad (7)$$

with a suitable approximation $\tilde{f}_0(u)$ of $\mathbb{E}_x\{f_0(x, u)\}$.

2.1. Worst-case approximation by linearization

Our approximation of the worst-case functions $\phi_i(u)$ is based on a linearization idea that has been developed within the authors’ workgroup [2] and has already been applied for design of robust optimal experiments, see K orkel et al. [10]. Independently, the idea had also been proposed by Ma and Braatz [15] and was applied for robust optimal control of batch processes by Nagy and Braatz [16]. By inspecting the expression (3) for the functions $\phi_i(u)$, we linearize both the model equations $g(x, u, p) = 0$ and the function $f_i(x, u)$, for given u , at a point (\bar{x}, u, \bar{p}) satisfying $g(\bar{x}, u, \bar{p}) = 0$, where \bar{p} is the nominal parameter.

We then approximate $\phi_i(u)$ by $\tilde{\phi}_i(u)$, defined by the convex optimization problem

$$\begin{aligned} \tilde{\phi}_i(u) &:= \max_{\Delta x \in \mathbb{R}^{n_x}, \Delta p \in \mathbb{R}^{n_p}} f_i(\bar{x}, u) + \frac{\partial f_i}{\partial x}(\bar{x}, u) \Delta x, \\ \text{s.t.} \quad &\frac{\partial g}{\partial x}(\bar{x}, u, \bar{p}) \Delta x + \frac{\partial g}{\partial p}(\bar{x}, u, \bar{p}) \Delta p = 0, \\ &\|\Delta p\| \leq 1. \end{aligned} \tag{8}$$

It is well known that this optimization problem can be solved analytically:

Lemma 1. *The convex optimization problem (8) has the optimal value*

$$\tilde{\phi}_i(u) = f_i(\bar{x}, u) + \left\| - \left(\frac{\partial g}{\partial p}(\bar{x}, u, \bar{p}) \right)^T \left(\frac{\partial g}{\partial x}(\bar{x}, u, \bar{p}) \right)^{-T} \left(\frac{\partial f_i}{\partial x}(\bar{x}, u) \right)^T \right\|_*. \tag{9}$$

Here we use the dual norm to $\|\cdot\|$, namely the mapping

$$\begin{aligned} \|\cdot\|_* : \mathbb{R}^{n_p} &\rightarrow \mathbb{R}, \\ a &\mapsto \|a\|_* := \max_{p \in \mathbb{R}^{n_p}} a^T p \quad \text{s.t.} \quad \|p\| \leq 1. \end{aligned} \tag{10}$$

Proof. We solve (8) by eliminating Δx , which yields the expression

$$\begin{aligned} \tilde{\phi}_i(u) = f_i(\bar{x}, u) + \max_{\Delta p \in \mathbb{R}^{n_p}} \frac{\partial f_i}{\partial x}(\bar{x}, u) \left(- \left(\frac{\partial g}{\partial x}(\bar{x}, u, \bar{p}) \right)^{-1} \frac{\partial g}{\partial p}(\bar{x}, u, \bar{p}) \right) \Delta p, \\ \text{s.t.} \quad \|\Delta p\| \leq 1. \end{aligned}$$

Remark. It is well known that for any scaled Hölder q -norm $\|p\| = \|Ap\|_q$ (A invertible, $1 \leq q \leq \infty$), the dual norm is $\|a\|_* = \|A^{-1}a\|_{\frac{q}{q-1}}$ (for $q = 1$ we define $\frac{q}{q-1} := \infty$ and for $q = \infty$, $\frac{q}{q-1} := 1$), as was observed in the context of worst-case analysis by Ma and Braatz [15] and independently within the authors’ workgroup [2]. For $\|\cdot\|$ being the maximum of two norms $\|\cdot\|_a$ and $\|\cdot\|_b$, i.e. $\|p\| = \max\{\|p\|_a, \|p\|_b\}$, the dual norm is the sum of the dual norms, $\|p\|_* = \|p\|_{a,*} + \|p\|_{b,*}$, and vice versa. Therefore, the dual norm can explicitly be given for those norms that we typically use to define our uncertainty sets.

2.2. The approximated robust counterpart

The optimization problem (6), which we call the “approximated robust counterpart” of (1), can explicitly be formulated as follows:

$$\min_{u \in \mathbb{R}^{n_u}, \bar{x} \in \mathbb{R}^{n_x}} f_0(\bar{x}, u) + \left\| \left(\frac{\partial g}{\partial p} \right)^T \left(\frac{\partial g}{\partial x} \right)^{-T} \left(\frac{\partial f_0}{\partial x} \right)^T \right\|_* \tag{11}$$

$$\text{(ARC) \quad s.t.} \quad f_i(\bar{x}, u) + \left\| \left(\frac{\partial g}{\partial p} \right)^T \left(\frac{\partial g}{\partial x} \right)^{-T} \left(\frac{\partial f_i}{\partial x} \right)^T \right\|_* \leq 0 \tag{12}$$

$$\begin{aligned} & i = 1, \dots, n_f, \\ & \text{and} \quad g(\bar{x}, u, \bar{p}) = 0. \end{aligned} \tag{13}$$

Here, all derivatives are evaluated at (\bar{x}, u, \bar{p}) . Note that they depend on the controls u and are thus subject to optimization.

2.2.1. Remark on higher order approximations: Instead of using only linear terms in the approximation (8), one might try to employ higher order Taylor series expansions of the mapping $f_i(g^{-1}(u, p), u)$ with respect to the uncertain parameters p . This has been suggested for robustness analysis by Ma and Braatz [15] who also present ways for computing polynomial time lower and upper bounds for the solution of the nonconvex worst-case problems in the case of second and third order expansions. However, these computations are considerably more expensive than the simple analytic expression (9) to the linearized worst-case problem (8), and they do not allow for similar smooth structure exploiting NLP formulations as they will be presented in Section 3 of this paper for the first order case.

2.3. Expectation value approximation by linearization

For the approximation (7) of the “expectation value robust counterpart” (5) we use the same linearization at the point (\bar{x}, u, \bar{p}) satisfying $g(\bar{x}, u, \bar{p}) = 0$, and approximate $\mathbb{E}_x\{f_0(x, u)\}$ by

$$\tilde{f}_0(u) := \mathbb{E}_p \left\{ f_0(\bar{x}, u) + \frac{\partial f_0}{\partial x}(\bar{x}, u) \left(- \left(\frac{\partial g}{\partial x}(\bar{x}, u, \bar{p}) \right)^{-1} \frac{\partial g}{\partial p}(\bar{x}, u, \bar{p}) \right) (p - \bar{p}) \right\}.$$

Because of symmetry of the probability distribution of p around \bar{p} , this evaluates to

$$\tilde{f}_0(u) = f_0(\bar{x}, u).$$

Therefore, the “expectation value robust counterpart approximation” (7) looks the same as the robust counterpart approximation (11)–(13) apart from the fact that the second term in the objective (11) – the dual norm penalty – is dropped.

3. Numerically efficient problem formulations

The above formulation (11)–(13) of the approximated robust counterpart suffers the drawback that problem structure inherent in the model equations $g(x, u, p) = 0$ like sparsity is likely to get lost, due to the explicit appearance of the inverse. We propose two equivalent formulations, that are computationally more convenient than (11)–(13), and that differ in the way they represent the required derivatives. Both formulations retain sparsity, which is very important for application to large scale systems. They bear some similarity to the “forward” and “reverse” mode of automatic differentiation, respectively, cf. Griewank [9].

3.1. Formulation 1: direct sensitivity equations

The first variant uses a straightforward approach to define the required derivative matrix

$$D(\bar{x}, u, \bar{p}) := \left(- \left(\frac{\partial g}{\partial x}(\bar{x}, u, \bar{p}) \right)^{-1} \frac{\partial g}{\partial p}(\bar{x}, u, \bar{p}) \right).$$

Note that this “sensitivity” matrix $D(\bar{x}, u, \bar{p})$ can be interpreted as the jacobian $\frac{\partial g^{-1}}{\partial p}(u, \bar{p})$ of the inverse function of g . Introducing $D \in \mathbb{R}^{n_x \times n_p}$ as an additional variable into the problem (11)–(13), we obtain:¹

$$\min_{u \in \mathbb{R}^{n_u}, x \in \mathbb{R}^{n_x}, D \in \mathbb{R}^{n_x \times n_p}} f_0(x, u) + \left\| D^T \left(\frac{\partial f_0}{\partial x}(x, u) \right)^T \right\|_* \quad (14)$$

$$\text{(ARC-D)} \quad \text{s.t.} \quad f_i(x, u) + \left\| D^T \left(\frac{\partial f_i}{\partial x}(x, u) \right)^T \right\|_* \leq 0, \quad (15)$$

$$i = 1, \dots, n_f,$$

$$g(x, u, \bar{p}) = 0, \quad (16)$$

$$\text{and} \quad \frac{\partial g}{\partial x}(x, u, \bar{p}) D + \frac{\partial g}{\partial p}(x, u, \bar{p}) = 0. \quad (17)$$

Note that the last equation (17) is a matrix equation in $\mathbb{R}^{n_x \times n_p}$.

3.2. Formulation 2: adjoint sensitivities

The second formulation does not use the sensitivity matrix $D \in \mathbb{R}^{n_x \times n_p}$, but instead an “adjoint sensitivity” matrix $\Lambda \in \mathbb{R}^{n_x \times (1+n_f)}$, that is defined as

$$\Lambda(\bar{x}, u, \bar{p}) := \left(-\frac{\partial f}{\partial x}(\bar{x}, u) \left(\frac{\partial g}{\partial x}(\bar{x}, u, \bar{p}) \right)^{-1} \right)^T.$$

Here, $f(x, u) = (f_0(x, u), \dots, f_{n_f}(x, u))$. Introducing $\Lambda \in \mathbb{R}^{n_x \times (1+n_f)}$ as an additional variable into the problem, and denoting the column vectors of Λ by $\lambda_0, \dots, \lambda_{n_f}$ we obtain

$$\min_{u \in \mathbb{R}^{n_u}, x \in \mathbb{R}^{n_x}, \Lambda \in \mathbb{R}^{n_x \times (1+n_f)}} f_0(x, u) + \left\| \left(\frac{\partial g}{\partial p}(x, u, \bar{p}) \right)^T \lambda_0 \right\|_* \quad (18)$$

$$\text{(ARC-A)} \quad \text{s.t.} \quad f_i(x, u) + \left\| \left(\frac{\partial g}{\partial p}(x, u, \bar{p}) \right)^T \lambda_i \right\|_* \leq 0 \quad (19)$$

$$i = 1, \dots, n_f,$$

$$g(x, u, \bar{p}) = 0, \quad (20)$$

$$\text{and} \quad \left(\frac{\partial g}{\partial x}(x, u, \bar{p}) \right)^T \Lambda + \left(\frac{\partial f}{\partial x}(x, u) \right)^T = 0. \quad (21)$$

Note that the last equation (21) is a matrix equation in $\mathbb{R}^{n_x \times (1+n_f)}$. This adjoint formulation may be preferred if the number n_p of uncertain parameters significantly exceeds the number n_f of inequality constraints.

¹ For notational convenience we drop the bar over the variable x from now on.

3.3. Addressing non-smoothness of the dual norms

In expressing the approximated robust counterparts via the direct or adjoint sensitivity formulations we avoided the explicit use of matrix inversions in order to preserve sparsity in the problem. However, the formulated problems have non-smooth objective resp. inequality constraints, due to the appearance of the dual norm. Therefore, if they shall be addressed by standard Newton-Lagrange algorithms for the solution of smooth nonlinear programming problems (cf. Nocedal and Wright [17]), further precaution needs to be taken.

3.3.1. Slack formulation for ℓ_1 or ℓ_∞ norms: Fortunately, the non-smoothness can be removed by introduction of slack variables, for scaled ℓ_1 or ℓ_∞ norms, or mixtures thereof. As an example let us assume the primal norm was a scaled maximum norm $\|p\| = \|A^{-1}p\|_\infty$ (expressing box uncertainty), such that the dual norm is given by

$$\|p\|_* = \|Ap\|_1 \tag{22}$$

By introducing slack vectors $s_i \in \mathbb{R}^{n_p}$, $i = 0, \dots, n_f$ and defining $e := (1, \dots, 1)^T \in \mathbb{R}^{n_p}$ we can formulate the approximated robust counterpart (11)–(13) as the following nonlinear program

$$\min_{u, x, s_0, \dots, s_{n_f}} f_0(x, u) + e^T s_0 \tag{23}$$

$$\text{s.t. } f_i(x, u) + e^T s_i \leq 0, \quad i = 1, \dots, n_f, \tag{24}$$

$$-s_i \leq A \left(\frac{\partial g}{\partial p}(x, u, \bar{p}) \right)^T \left(\frac{\partial g}{\partial x}(x, u, \bar{p}) \right)^{-T} \left(\frac{\partial f}{\partial x}(x, u) \right)^T \leq s_i, \quad i = 0, \dots, n_f, \tag{25}$$

$$\text{and } g(x, u, \bar{p}) = 0. \tag{26}$$

A similar formulation via slacks also exists for the maximum norm, and of course for the direct or adjoint sensitivity formulations. The resulting nonlinear programs have only smooth problem functions and can be solved by standard methods of nonlinear programming.

3.3.2. Euclidean norm: For all scaled Hölder norms $\|p\|_* = \|Ap\|_r$ with $1 < r < \infty$, there is no slack formulation to transform the problem into a smooth NLP. This is in particular true for the important case $r = 2$, the self-dual Euclidean norm, which is used to express ellipsoidal uncertainty. However, for this norm there is only one non-differentiable point, namely the zero vector $p = 0$. Therefore one might hope that a standard Newton-Lagrange NLP algorithm is able to solve (ARC), the approximated robust counterpart problem (11)–(13), without any modification, despite the non-differentiability. This approach was successfully applied for robust optimal design of experiments by Körkel et al. [10], and it is also employed in the application example at the end of this paper. However, while neglecting the non-differentiability works well for many practically relevant problems, it is important to note that there also exist interesting problems where the optimal solution lies exactly at the non-differentiable point. This occurs when it is both possible and optimal to choose the controls u such as to avoid the (linearized)

dependence of the objective or some inequality constraint function on the uncertain parameters altogether, i.e., by making one or more of the “total derivative” terms

$$\frac{df_i}{dp} := \frac{\partial f_i}{\partial x} \left(\frac{\partial g}{\partial x} \right)^{-1} \frac{\partial g}{\partial p} = \frac{\partial f_i}{\partial x} D = \lambda_i^T \frac{\partial g}{\partial p}$$

zero in the optimal solution.

This situation frequently occurs e.g. in robust *linear* programming, where f_0, \dots, f_{n_f} are linear functions and g is without loss of generality of the form $g(x, u, p) = -x + A(p)u + b(p)$, with $A(\cdot)$ and $b(\cdot)$ affine in p . Fortunately, for robust linear programming there exist well developed solution methods from the field of conic programming, see e.g. Ben-Tal and Nemirovskii [1] or Boyd and Vandenberghe [4]. It is easy to show that for uncertain robust *linear* programs the linearized formulation (11)–(13) coincides with the exact robust counterpart (4), and that a robust linear program with ellipsoidal uncertainty set can be cast as a second order cone program.

For general nonlinear problems with ellipsoidal uncertainty, however, the approximated robust counterpart (11)–(13) falls into a subclass of nonlinear semidefinite programming. To be specific, when the dual norm is a scaled Euclidean norm $\|p\|_* = \|Ap\|_2$, it can be formulated as a smooth nonlinear problem with (nonsmooth) second order cone constraints.

It can be hoped that the ongoing development of algorithms for the general class of nonlinear semidefinite programming (cf. [5, 7, 8, 11, 12]), will facilitate the reliable numerical solution of this type of large-scale nonlinear conic optimization problem.

4. Generalization to optimal control of uncertain systems

We will briefly describe how the direct and adjoint approximated robust counterpart formulations can be generalized to the context of optimal control problems. For this aim we regard a simplified optimal control problem in ordinary differential equations on the (variable) time horizon $[0, T]$ of the following form:

$$\min_{T, u(\cdot), x(\cdot)} f_0(x(T), T) \tag{27}$$

$$\text{s.t.} \quad \frac{dx}{dt}(t) = G(x(t), u(t)), \quad \forall t \in [0, T], \tag{28}$$

$$x(0) = x_0(p), \tag{29}$$

$$f_i(x(T), T) \leq 0, \quad i = 1, \dots, n_f. \tag{30}$$

We regard the horizon length T and the control profile $u : [0, T] \rightarrow \mathbb{R}^{n_u}$ as decision variables and the trajectory $x : [0, T] \rightarrow \mathbb{R}^{n_x}$ as the dependent variables. The latter are chosen by “nature”, which is only restricted by the model equations (28), the initial value constraint (29), and the norm bound $\|p - \bar{p}\| \leq 1$ on the parameters $p \in \mathbb{R}^{n_p}$.

The above problem class is fairly general and includes in particular uncertain parameter dependent ODE $\frac{dx}{dt}(t) = G(x(t), u(t), \tilde{p})$ (add some trivial differential equations $\frac{d\tilde{p}}{dt} = 0$ and let the initial value $\tilde{p}(0) = p$ be chosen by nature).

We show two equivalent ways to approximate the robust counterpart of the above optimal control problem, first a direct and second an adjoint sensitivity formulation.

4.1. Direct approximate robust counterpart formulation

Analogous to the direct sensitivity formulation (14)–(17) of the approximated robust counterpart, we obtain the following optimal control problem formulation, where the time dependent system state $x(t) \in \mathbb{R}^{n_x}$ is augmented by a time dependent sensitivity matrix $D(t) \in \mathbb{R}^{n_x \times n_p}$.

$$\min_{T, u(\cdot), x(\cdot), D(\cdot)} \quad f_0(x(T), T) + \left\| D(T)^T \left(\frac{\partial f_0}{\partial x}(x(T), T) \right)^T \right\|_* \quad (31)$$

$$\text{s.t.} \quad f_i(x(T), T) + \left\| D(T)^T \left(\frac{\partial f_i}{\partial x}(x(T), T) \right)^T \right\|_* \leq 0, \quad (32)$$

$$i = 1, \dots, n_f,$$

$$\frac{dx}{dt}(t) = G(x(t), u(t)), \quad \forall t \in [0, T], \quad (33)$$

$$x(0) = x_0(p), \quad (34)$$

$$\frac{dD}{dt}(t) = \frac{\partial G}{\partial x}(x(t), u(t))D(t), \quad \forall t \in [0, T], \quad (35)$$

$$D(0) = \frac{\partial x_0}{\partial p}(p). \quad (36)$$

Note that the matrix $D(t) \in \mathbb{R}^{n_x \times n_p}$ can be regarded as the derivative

$$D(t) = \frac{dx(t)}{dx(0)} \frac{\partial x_0}{\partial p}(p),$$

where we denote by $\frac{dx(t)}{dx(0)}$ the linearized dependence of the state $x(t)$ at time t on the initial state $x(0)$. This type of direct sensitivity formulation was e.g. suggested by Nagy and Braatz [16], who applied it to robust optimization of batch processes.

4.2. Adjoint approximate robust counterpart formulation

Analogous to the adjoint sensitivity formulation (18)–(21) we can alternatively formulate an optimal control problem where the time dependent system state $x(t) \in \mathbb{R}^{n_x}$ is augmented by $(n_f + 1)$ time dependent adjoint sensitivities $\lambda_i(t) \in \mathbb{R}^{n_x}$, that we combine in the matrix $\Lambda(t) = (\lambda_0(t) | \dots | \lambda_{n_f}(t)) \in \mathbb{R}^{n_x \times (n_f + 1)}$:

$$\min_{T, u(\cdot), x(\cdot), \Lambda(\cdot)} \quad f_0(x(T), T) + \left\| \left(\frac{\partial x_0}{\partial p}(p) \right)^T \lambda_0(0) \right\|_* \quad (37)$$

$$\text{s.t.} \quad f_i(x(T), T) + \left\| \left(\frac{\partial x_0}{\partial p}(p) \right)^T \lambda_i(0) \right\|_* \leq 0, \quad i=1, \dots, n_f, \quad (38)$$

$$\frac{dx}{dt}(t) = G(x(t), u(t)), \quad \forall t \in [0, T], \quad (39)$$

$$x(0) = x_0(p), \quad (40)$$

$$\frac{d\Lambda}{dt}(t) = - \left(\frac{\partial G}{\partial x}(x(t), u(t)) \right)^T \Lambda(t), \quad \forall t \in [0, T], \quad (41)$$

$$\Lambda(T) = \left(\frac{\partial f}{\partial x}(x(T), T) \right)^T. \quad (42)$$

The adjoint differential equations (41) with the terminal boundary condition (42) are well known in optimal control theory, albeit in a different context (for the formulation of necessary optimality conditions). In our context, the matrix variable $\Lambda(t) \in \mathbb{R}^{n_x \times (n_f + 1)}$ can be interpreted as an adjoint sensitivity:

$$\Lambda(t)^T = \frac{\partial f}{\partial x}(x(T), T) \frac{dx(T)}{dx(t)},$$

where we define $\frac{dx(T)}{dx(t)}$ to be the linearized dependence of the final state $x(T)$ on the intermediate state $x(t)$ at time t . Using this definition, we can derive the terminal condition (42) by noting that $\frac{dx(T)}{dx(T)} = \mathbb{I}$. Likewise, the adjoint differential equation (41) can be deduced as follows. For the fixed matrix $\Lambda(T)^T \frac{dx(T)}{dx(0)}$ holds

$$0 = \frac{d}{dt} \left(\Lambda(T)^T \frac{dx(T)}{dx(0)} \right) = \frac{d}{dt} \left(\Lambda(T)^T \frac{dx(T)}{dx(t)} \frac{dx(t)}{dx(0)} \right) \quad (43)$$

$$= \frac{d}{dt} \left(\Lambda(T)^T \frac{dx(T)}{dx(t)} \right) \frac{dx(t)}{dx(0)} + \Lambda(T)^T \frac{dx(T)}{dx(t)} \frac{d}{dt} \left(\frac{dx(t)}{dx(0)} \right) \quad (44)$$

$$= \frac{d\Lambda(t)^T}{dt} \frac{dx(t)}{dx(0)} + \Lambda(t)^T \frac{d}{dt} \left(\frac{dx(t)}{dx(0)} \right) \quad (45)$$

$$= \frac{d\Lambda(t)^T}{dt} \frac{dx(t)}{dx(0)} + \Lambda(t)^T \frac{\partial G}{\partial x}(x(t), u(t)) \frac{dx(t)}{dx(0)} \quad (46)$$

\Leftrightarrow

$$0 = \frac{d\Lambda(t)^T}{dt} + \Lambda(t)^T \frac{\partial G}{\partial x}(x(t), u(t)). \quad (47)$$

This adjoint formulation may be preferred to the direct one if the number n_p of uncertain parameters is considerably larger than the number $n_f + 1$ of objective and inequality constraint functions.

4.2.1. Remark on time dependent disturbances: The adjoint method can even be used to approximate the case of time dependent (i.e., infinite dimensional) disturbances $w(t)$, where the differential equation (28) is replaced by

$$\frac{dx}{dt}(t) = G(x(t), u(t)) + w(t), \quad \forall t \in [0, T], \quad (48)$$

and we have a norm bound on the function $w : [0, T] \rightarrow \mathbb{R}^{n_x}$, e.g. of the form

$$\max_{t \in [0, T]} \|w(t)\|_w \leq 1, \quad (49)$$

with an arbitrary norm $\|\cdot\|_w$ in \mathbb{R}^{n_x} . Such a problem formulation can be used to express uncertainty in the model equations. While it is difficult to address this problem class with the direct sensitivity formulation, a generalization of the adjoint formulation is straightforward. To incorporate the additional uncertainty, we simply have to add the additional penalty terms

$$\int_0^T \|\lambda_i(t)\|_{w,*} dt, \quad i = 0, \dots, n_f, \quad (50)$$

to the existing dual norm penalties in Eqs. (37) and (38). Here, $\|\cdot\|_{w,*}$ is the dual norm to $\|\cdot\|_w$. To justify this formulation, we regard the linearized dependence of f_i on variations in p and $w(t)$, $t \in [0, T]$:

$$\Delta f_i(\Delta p, \Delta w(\cdot)) = \lambda_i(0)^T \frac{\partial x_0}{\partial p}(p) \Delta p + \underbrace{\int_0^T \lambda_i(t)^T \Delta w(t) dt}_{=:\langle \lambda_i, \Delta w \rangle}. \quad (51)$$

Now, as before in Eq. (8) for the finite dimensional setting, we consider the worst-case

$$\max_{\Delta p, \Delta w(\cdot)} \Delta f_i(\Delta p, \Delta w(\cdot)) \quad \text{s.t.} \quad \max \left(\|\Delta p\|, \max_{t \in [0, T]} \|\Delta w(t)\|_w \right) \leq 1,$$

which evaluates indeed to the sum of the two dual norms:

$$= \left\| \lambda_i(0)^T \frac{\partial x_0}{\partial p}(p) \right\|_* + \int_0^T \|\lambda_i(t)\|_{w,*} dt,$$

where the second dual norm is dual to the maximum norm employed in Eq. (49) with respect to the L_2 -scalar product used in Eq. (51).

Equally, if the time dependent disturbances are bounded directly by an L_2 -norm (that corresponds to a confidence region for uncorrelated disturbances $w(t)$), i.e., if

$$\sqrt{\frac{1}{T} \int_0^T \|w(t)\|_w^2 dt} \leq 1, \quad (52)$$

the additional norm penalties in Eqs. (37) and (38) become

$$\sqrt{T \int_0^T \|\lambda_i(t)\|_{w,*}^2 dt}, \quad i = 0, \dots, n_f, \quad (53)$$

because this is the dual of the norm (52) with respect to the L_2 -scalar product used in Eq. (51).

5. Application example: batch distillation

In order to illustrate the proposed approximation technique we consider an example from chemical engineering, a batch distillation process. Aim of the process is to separate a mixture of two components and to produce distillate with a purity of at least 99% and to maximize the produced quantity of distillate.

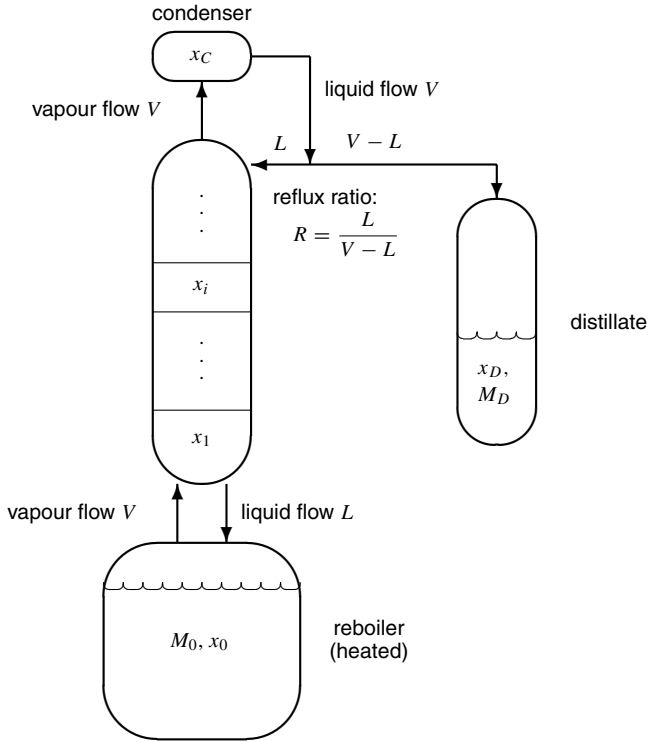


Fig. 1. Sketch of the batch distillation column

5.1. Process model

The batch distillation, that is sketched in Figure 1, is modelled as follows: the reboiler content M_0 with composition x_0 (molar percentage of the lighter component) is heated, such that vapour with equilibrium composition $y(x_0)$ is produced, with a constant molar flux $V = 100$. We assume a simple expression for the vapour equilibrium in the form $y(x) := \frac{x(1+\alpha)}{x+\alpha}$, cf. Figure 2. The vapour flux V is assumed to be constant throughout all trays of the column. It is totally condensed at the top, and a liquid molar flux $L = \frac{R}{1+R} V$ is fed back into the column, where the “reflux ratio” R is controlled. L is also assumed to be constant throughout the column. The remainder of the condensed liquid flux, $V - L$ with composition x_C , is collected within the distillate container with molar holdup M_D and composition x_D . We collect the following ordinary differential equations:

$$\dot{M}_0 = -V + L, \quad \dot{x}_0 = \frac{1}{M_0} (Lx_1 - Vy(x_0) + (V - L)x_0) \quad (54)$$

for the reboiler, and

$$\dot{x}_i = \frac{1}{m} (Lx_{i+1} - Vy(x_i) + Vy(x_{i-1}) - Lx_i) \quad i = 1, \dots, N, \quad (55)$$

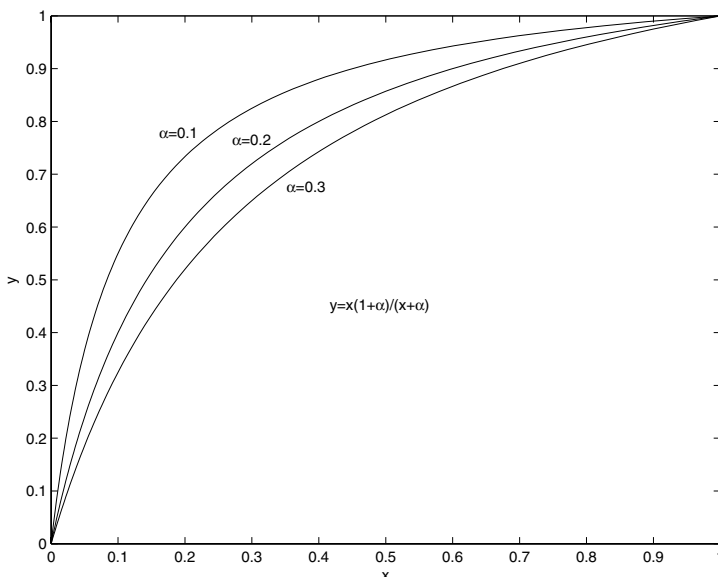


Fig. 2. The dependence of the equilibrium vapour concentration y on the liquid concentration x , for different values of α

for the tray concentrations (where $m = 0.1$ is the molar holdup of each tray assumed to be constant, and $N = 5$). For the condenser concentration x_C , that we identify with x_{N+1} , we obtain

$$\dot{x}_{N+1} = \frac{V}{m_C} (y(x_N) - x_{N+1}), \tag{56}$$

where $m_C = 0.1$ is the constant molar holdup of the condenser, and for the distillate container we obtain

$$\dot{M}_D = V - L, \quad \dot{x}_D = \frac{V - L}{M_D} (x_{N+1} - x_D). \tag{57}$$

As we want to include the uncertainty in α in formally the same way as described above, we add the trivial equation $\dot{\alpha} = 0$.

5.2. Uncertain optimization problem

Summarizing the states in $x = (M_0, x_0, x_1, \dots, x_{N+1}, M_D, x_D, \alpha)^T$ and setting $u = R$, we obtain equations $\dot{x} = G(x, u)$ in the form above. The objective is given by $f_0(x(T), T) := T - M_D(T)$, and the only inequality constraint by $f_1(x(T), T) := 0.99 - x_D(T) \leq 0$. Nominal initial values are $M_0 = 100, x_0 = 0.5, x_1 = \dots = x_{N+1} = x_D = 1, M_D = 0.1$, and $\alpha = 0.2$. Uncertainty is present in the parameter α as well as in the initial feed composition $x_0(0)$, and we assume

$$\sqrt{\left(\frac{\alpha - 0.2}{0.1}\right)^2 + \left(\frac{x_0(0) - 0.5}{0.05}\right)^2} \leq 1. \tag{58}$$

This corresponds to independent relative errors of a size of 50% for α and 10% for $x_0(0)$. In Figure 2 the effect of the considered variations in α on the vapour equilibrium curve $y(x)$ is illustrated. We consider the expectation value robust counterpart formulation, i.e., the optimization is performed for the nominal objective and the linearized worst-case constraint.

In addition to the setup presented above, there are bounds $0 \leq R \leq 15$ on the controlled reflux ratio $u = R$.

5.3. Optimization results

Since the number of uncertain parameters $n_p = 2$ is not considerably greater than the number $n_f = 1$ of uncertain constraints, the approximated robust counterpart is formulated in form of the direct sensitivity equations, (31)–(36). For the solution of this boundary constrained optimal control problem we employ the software package MUSCOD-II [13, 14] that is based on the direct multiple shooting method [3]. This method is particularly suitable for nonlinear optimal control problems with boundary and state constraints and has already been successfully applied to similar (but non-robust) batch distillation problems [6]. The control profile is approximated by continuous piecewise linear interpolation using 15 equally spaced intervals on the variable time horizon $[0, T]$.

For comparison, first the nominal solution is computed and shown in Figure 3. The employed time is $T = 1.3$ while the produced distillate is $M_D(T) = 51.0$, leading to an objective of -49.7 . In the middle row on the right it can be seen that the lower bound of 99% on $x_D(T)$ is active. At the bottom row the linearized dependence of $x_D(t)$ on a unit variation in α respectively in $x_0(0)$ is shown. Inspecting the last graph one can conclude that for an initial composition $x_0(0)$ that is reduced by 0.05, the final distillate composition $x_D(T)$ may be reduced by approximately $1.5 \times 0.05 = 0.075 = 7.5\%$. This means that in the linearized worst-case a purity of only $99.0\% - 7.5\% = 91.5\%$ may occur.

This is in contrast to the approximated robust solution shown in Figure 4. Here, the sensitivities on the uncertain parameters are considerably reduced (by 50% respectively 90%, as seen in the middle row), at the cost of a slightly increased objective of -49.0 , with longer duration $T = 1.4$ and less produced distillate $M_D(T) = 50.4$. Note that the nominal distillate composition in the middle row at the right is increased to 99.8%, to allow for a safety back-off with respect to the remaining uncertainty. Thus, the approximated robust optimization has two effects: it reduces the dependence of the constraint on the uncertain parameters, and provides some slack for safety. Both are achieved at little extra cost in the objective.

6. Conclusions

We have presented techniques for the numerically efficient formulation of approximated robust counterparts of nonlinear optimization problems. We assume that the optimizer chooses the controls u first, before an adverse player chooses the uncertain parameters

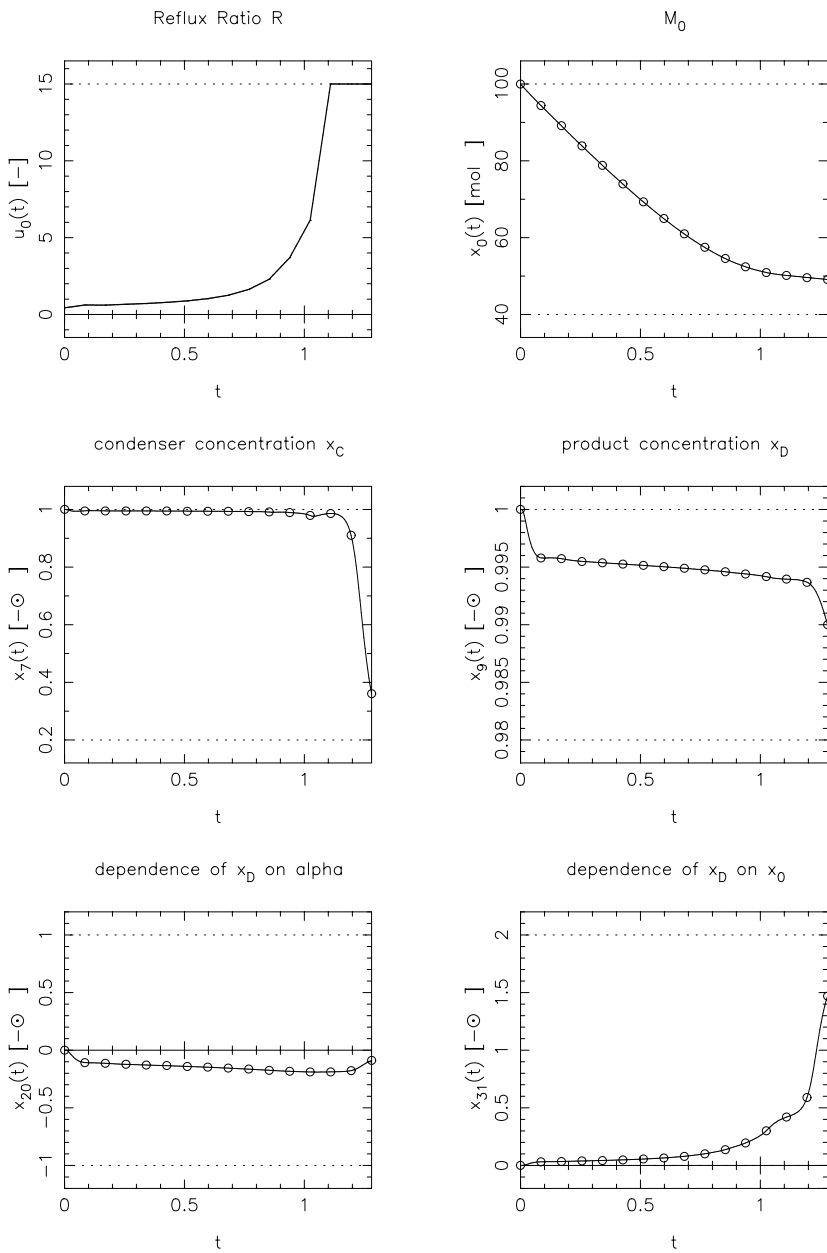


Fig. 3. The nonrobust solution of the batch distillation problem

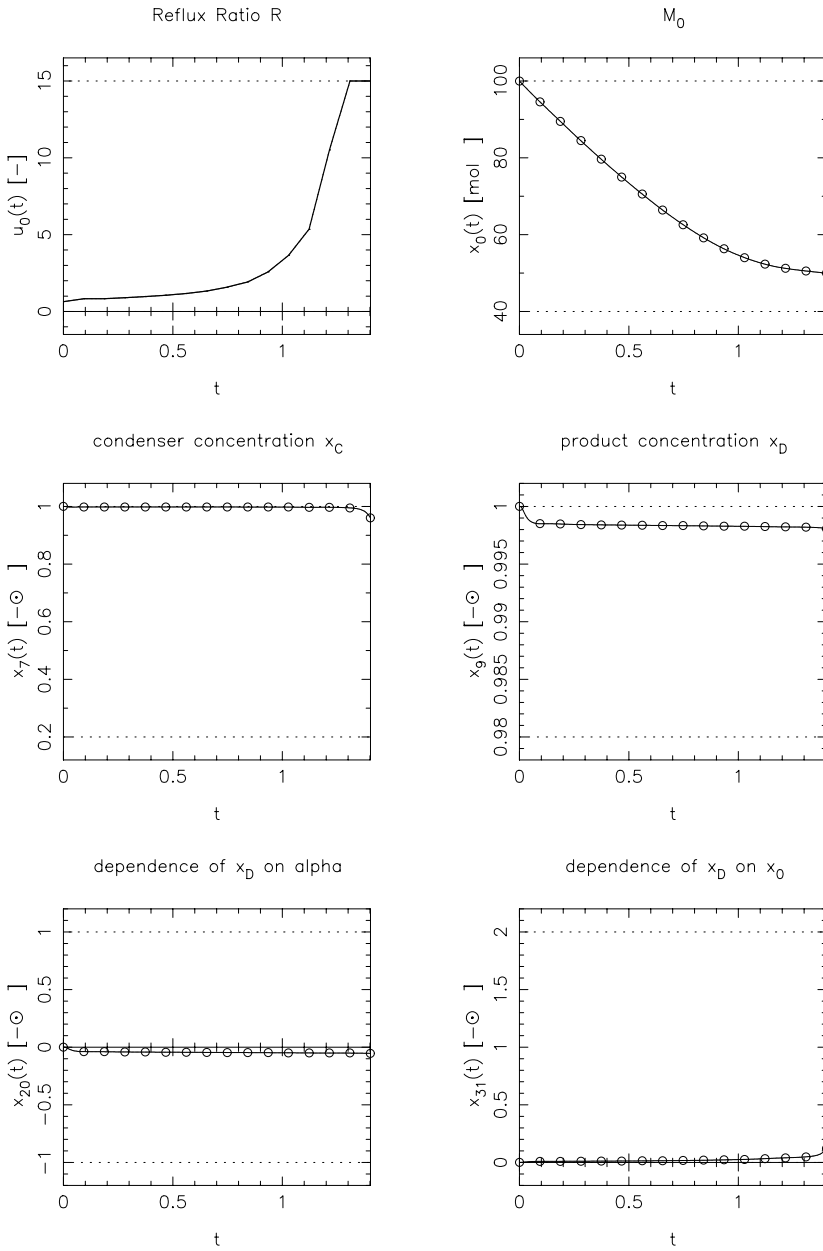


Fig. 4. The approximate robust solution of the batch distillation problem

p and states x , restricted by a norm bound $\|p - \bar{p}\| \leq 1$ on disturbances on p and by the model equations $g(x, u, p) = 0$. By linearizing the model equations at a point corresponding to the nominal parameter value \bar{p} and the chosen controls u , the worst-case can approximately be computed, and a numerically tractable approximation of the robust counterpart is obtained. The resulting optimization problem contains norms on the first order derivatives in its objective and constraints [15, 16, 2, 10].

The main advance of the paper are two equivalent formulations for the efficient numerical solution of this problem, the *direct* and the *adjoint* approximated robust counterpart. They have similarity to the forward mode respectively to the backward mode of automatic differentiation. The second, adjoint formulation is advantageous in the presence of many uncertain parameters and few inequality constraints. We also explored the consequences of the appearance of the dual norms in objective and constraints that leads to nonlinear conic programming problems.

We have shown how to generalize the two methods to optimal control problems in ODE, and discussed how the adjoint formulation could even be used to address infinite dimensional uncertainty in the form of time dependent disturbances. Finally, we presented a batch distillation process as an application example. It is illustrated that the approximated robust optimization leads to a solution profile that shows both, a reduced dependence of the constraints on the uncertain parameters, and a safety back-off to guarantee (linearized) worst-case feasibility.

Acknowledgements. Financial support by the *Deutsche Forschungsgemeinschaft (DFG)* within the SFB 359 “Reactive Flows, Diffusion and Transport” and within the project BO864/10-1 “Optimization Based Control of Chemical Processes” is gratefully acknowledged. The first author also thanks Z. Nagy for inspiring discussions.

References

1. Ben-Tal, A., Nemirovskii, A.: Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. MPS-SIAM Series on Optimization, MPS-SIAM, Philadelphia 2001
2. Bock, H., E.Kostina: Robust experimental design. In: H.G.Bock, Project A4, Optimization methods for reactive flows of Sonderforschungsbereich 359 Reactive Flows, Diffusion and Transport, Report 1999-2001, University of Heidelberg, 2001, pp. 133–135
3. Bock, H., Plitt, K.: A multiple shooting algorithm for direct solution of optimal control problems. In: Proc. 9th IFAC World Congress Budapest, Pergamon Press, 1984, pp. 243–247
4. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, 2004
5. Correa, R., Ramírez, C.: A global algorithm for nonlinear semidefinite programming. *SIAM Journal on Optimization* **15** (1), 303–318 (2002)
6. Diehl, M., Leineweber, D., Schäfer, A., Bock, H., Schlöder, J.: Optimization of multiple-fraction batch distillation with recycled waste cuts. *AIChE Journal* **48** (12), 2869–2874 (2002)
7. Fares, B., Noll, D., Apkarian, P.: Robust control via sequential semidefinite programming. *SIAM Journal on Control and Optimization* **40** (6), 1791–1820 (2002)
8. Freund, R.W., Jarre, F., Vogelbusch, C.: A sequential semidefinite programming method and an application in passive reduced-order modeling, 2005, (submitted)
9. Griewank, A.: Evaluating derivatives: principles and techniques of algorithmic differentiation. SIAM, Philadelphia 2000
10. Körkel, S., Kostina, E., Bock, H., Schlöder, J.: Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optimization Methods and Software* **19**, 327–338 (2004)
11. Kočvara, M., Stingl, M.: Pennon - a generalized augmented lagrangian method for semidefinite programming. In: G.D. Pillo, A. Murli (eds.) High Performance Algorithms and Software for Nonlinear Optimization, Kluwer Academic Publishers, Dordrecht, 2003, pp. 297–315

12. Leibfritz, F., Mostafa, E.M.E.: An interior point constrained trust region method for a special class of nonlinear semidefinite programming problems. *SIAM Journal on Optimization* **12** (4), 1048–1074 (2002)
13. Leineweber, D.: Efficient reduced SQP methods for the optimization of chemical processes described by large sparse DAE models, *Fortschr.-Ber. VDI Reihe 3, Verfahrenstechnik*, vol. 613. VDI Verlag, Düsseldorf, 1999
14. Leineweber, D., Schäfer, A., Bock, H., Schlöder, J.: An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization. Part II: Software aspects and applications. *Comp. & Chem. Eng.* **27**, 167–174 (2003)
15. Ma, D.L., Braatz, R.D.: Worst-case analysis of finite-time control policies. *IEEE Transactions on Control Systems Technology* **9** (5), 766–774 (2001)
16. Nagy, Z., Braatz, R.: Open-loop and closed-loop robust optimal control of batch processes using distributional and worst-case analysis. *Journal of Process Control* **14**, 411–422 (2004)
17. Nocedal, J., Wright, S.J.: *Numerical Optimization*. Springer, 1999