A General Robust-Optimization Formulation
for Nonlinear Programming

Yin Zhang[*]

July, 2004 (Revise June 2005)

Abstract

Most research in robust optimization has so far been focused on inequality-only, convex conic programming with simple linear models for uncertain parameters. Many practical optimization problems, however, are nonlinear and non-convex. Even in linear programming, coefficients may still be nonlinear functions of uncertain parameters. In this paper, we propose robust formulations (see (1) versus (5), and (2) versus (3), for example) that extend the robust-optimization approach to a general nonlinear programming setting with parameter uncertainty involving both equality and inequality constraints. The proposed robust formulations are valid in a neighborhood of a given nominal parameter value and are robust to the first-order, thus suitable for applications where reasonable parameter estimations are available and uncertain variations are moderate.

Keywords: Parameterized nonlinear program, Robust optimization formulation.

1 Introduction

Most optimal design or control problems in practice involve system parameters. These parameters must be assigned values before optimization techniques can be applied to obtain numerical solutions. More often than not, precise values of the involved system parameters are impossible to determine due to either measurement errors, random noises or other technical difficulties. On the other hand, some estimated values for the system parameters are often available. In many applications, good estimates for system parameter values are all that is needed, while variations in solutions induced by uncertain system parameters are either within the margin of errors or otherwise tolerable.

However, there are some critical applications where significant solution variations in certain directions are not allowable, while variations in other directions are. For example, the stress in a vital part of a structure may not be allowed to exceed a certain threshold value in order to guarantee safety, but less stress is certainly allowable.

1.1 An Example

The following model involving ordinary differential equations comes from an optimal control setting:

$$\min_{y,u} \quad \left\| y(T) - y^{end} \right\|_2^2$$

$$\text{s.t.} \quad \dot{y} = f(y, u, s), \quad y(0) = y^{ini}, \tag{1a}$$

$$y_i(t) \le \gamma, \qquad i = 1 : m, \tag{1b}$$

$$t \in (0, T),$$

where $y(t) : [0, T] \to \Re^d$ represents a vector-valued state function (or variable), $u = u(t)$ is a vector-valued control function, s is a vector-valued system parameter, and $i = 1 : m$ means $i = 1, 2, \cdots, m$. In this model, it is required that the first m ($m < d$) components of the state variable $y(t)$ not take values above a given safety threshold γ at any time (as such, the initial values should also satisfy $y_i^{ini} \le \gamma$ for $i = 1 : m$).

Suppose that the system parameter s and the initial value y^{ini} are roughly known, but still subject to yet unknown variations. Under this uncertainty, the question is how should one choose the control function $u(t)$ to ensure that (i) the safety requirement $y_i(t) \le \gamma$, $i = 1 : m$, be met *in the worst case*, and (ii) the final state $y(T)$ is as close as possible to a desired state y^{end}. This question may be addressed by solving a robust counterpart, or formulation, of (1). A specific robust formulation of (1) is given in (5).

It is worth emphasizing that optimization problem (1) contains equality constraints, or *state equations*, representing physical laws that should always be satisfied. It also contains critical inequality constraints that are considered "*safety constraints*". In addition, we observe that the state equations implicitly define the state variable y as a function of the control u and the parameter s.

1.2 A Brief Overview

Recently, robust optimization has become a very active research area for convex conic programming; specifically for linear programming (LP), second-order cone programming (SOCP), and semi-definite programming (SDP). In particular, we refer to the works of Ben-Tal, Nemirovski and their co-authors [2]-[9], and the works of El Ghaoui and his co-authors [10]-[14]. Other recent works on this subject include Mangasarian and Musicant [16], and Goldfarb and Iyengar [15], for example. When data uncertainty appears only in strictly satisfiable inequalities, and is restricted to ellipsoidal sets, the main results obtained so far can be briefly summarized as follows. The robust counterpart of LP is SOCP, of SOCP is SDP (though some remain SOCP), and of SDP can be approximated by SDP. These results represent significant advances in the development of rigorous approaches to treating data uncertainties in optimization. In addition,

a number of works have already appeared (see [17] or [18], for example) that apply the robust optimization methodology to various problems.

Other approaches, such as stochastic or sampling methods, or semi-infinite programming techniques, exist for treating parameter uncertainties (some of which have unfortunately also been termed robust optimization). These different approaches offer different applicabilities and advantages, but are in general more computationally demanding. For example, the so-called reliability-based design optimization (RBDO) approach has been actively studied in the field of engineering design (see [1] for a recent survey and a list of references). RBDO requires that probability distributions be given for uncertain parameters, and takes the form of either two-level optimization or mathematical programming with equilibrial constraints.

The robust optimization approach taken in this paper is suitable for applications where (i) reasonable parameter estimates (nominal values) exist for the uncertain parameters, and (ii) the magnitudes of uncertain variations are moderate. The resulting robust optimization formulations are locally valid around nominal values, but are generally more computationally tractable than those derived from more global approaches.

So far, existing robust optimization results have a number of limitations in their applicabilities. First of all, they apply only to linear constraints. Even for linear constraints, they are applicable only to cases where uncertain data appear exclusively in strictly satisfiable linear inequalities. In other words, data uncertainties in equality constraints cannot be readily handled by existing formulations. Other restrictive assumptions in existing results include that data elements be uncertain parameters themselves and be independent of each other.

In this paper we will derive robust optimization formulations for general nonlinear programming, using straightforward mathematical tools like linearization and the implicit function theorem. The resulting formulations should help extend the applicability of the robust optimization methodology to a wide range of scientific and engineering applications.

2 Nonlinear Optimization with Uncertainty

2.1 Nonlinear Programing Model

In this paper, we will focus on finite-dimensional spaces, though a similar approach may be developed in an infinite-dimensional setting. We consider the following general form of parameterized nonlinear optimization problem:

$$\min_{y,\, u \,\in\, \mathcal{U}} \quad \phi(y, u, s) \tag{2a}$$

$$\text{s.t.} \quad F(y, u, s) = 0 \tag{2b}$$

$$G(y, u, s) \leq 0 \tag{2c}$$

where $s \in \Re^{N_s}$ is the vector of uncertain system parameters and, for convenience, we call $y \in \Re^{N_y}$ the state variable, $u \in \Re^{N_u}$ the design variable constrained in a set $\mathcal{U} \subset \Re^{N_u}$,

$F(y, u, s) = 0$ the state equation for $F \in \Re^{N_y}$, and finally we call $G(y, u, s) \leq 0$ the safety constraints for $G \in \Re^m$.

Without loss of generality, we have tacitly assumed that all inequality constraints are safety constraints since non-safety inequality constraints will have no bearing in our derivation of robust formulations. It is worth noting that the sizes of y and F are equal. We will assume that the state equation $F(y, u, s) = 0$ implicitly defines a function $y = y(u, s)$ for relevant u and s. Throughout the paper, we will assume that F and G are continuously differentiable. Unlike in convex programming, we are concerned with only local optimization in the general nonlinear program (2).

2.2 A Robust Counterpart

To have an advanced look, we present a specific robust counterpart of (2) below without a derivation (though it will be verified later). For the function F, we define F_y to be the Jacobian of F with respect to y, and similarly for other functions and variables. Let $\hat{s}$ be our best estimate on the values of the parameter s, and τ our best estimate on the maximum deviation of s from $\hat{s}$ measured by the ℓ_1-norm. Under mild assumptions, a robust counterpart of (2) is

$$
\begin{align}
\min_{y, u \in \mathcal{U}} \quad & \phi(y, u, \hat{s}) \nonumber \\
\text{s.t.} \quad & F(y, u, \hat{s}) = 0 \tag{3a} \\
& \tau(F_y y_s + F_s) = 0 \tag{3b} \\
& \mathrm{diag}(G)E \pm \tau(G_y y_s + G_s) \leq 0 \tag{3c}
\end{align}
$$

where $y_s \in \Re^{N_y \times N_s}$ is a new variable representing the unknown Jacobian of y with respect to s, $E \in \Re^{m \times N_s}$ is the matrix of all ones, and the matrix inequalities are element-wise. The functions F_y, F_s, G_y, G_s in (3b)-(3c) are all evaluated at $(y, u, \hat{s})$. Several observations are now in order.

- When the deviation magnitude $\tau = 0$, the robust optimization problem (3) reduces to the original problem (2) (with multiple copies of the inequality constraint $G \leq 0$ present).

- In addition to the N_y equations, the robust problem (3) has $N_y N_s$ extra equations in (3b) for the new unknown $y_s \in \Re^{N_y \times N_s}$.

- The robust problem (3) has $2mN_s$ inequality constraints in (3c), which are equivalent to the following element-wise inequality

$$
\mathrm{diag}(G)E + \tau|(G_y y_s + G_s)| \leq 0, \tag{4}
$$

and are obviously more stringent than the original constraints $G \leq 0$. The newly added nonnegative terms in (4) play the important role of *safety margins*.

In a nutshell, the robust optimization problem (3) is a larger nonlinear program with safety margins added to the safety constraints, and the safety margins require the calculation of y_s – the sensitivity of the state with respect to the parameters. The amount of increase in problem size is proportional to the number of uncertain parameters in s. The robust problem (3) needs to be solved only at the parameter value $\hat{s}$.

2.3 The Example in (1)

As an illustration, we give a robust counterpart for the example (1). Assume that the dimension of the parameter vector s is N_s; hence $y_s(t) \in \Re^{d \times N_s}$ is a matrix-valued function for $t \in (0, T)$.

Although the robust formulation (3) is given in a finite-dimensional form, it can in fact be readily extended to infinite-dimensional problems like (1), resulting in a robust counterpart of the form:

$$
\min_{y,u} \quad \left\| y(T) - y^{end} \right\|_2^2
$$

$$
\text{s.t.} \quad \dot{y} = f(y, u, \hat{s}), \qquad\qquad y(0) = y^{ini}, \tag{5a}
$$

$$
\dot{y}_s = f_y(y, u, \hat{s})y_s + f_s(y, u, \hat{s}), \quad y_s(0) = 0, \tag{5b}
$$

$$
y_i(t) \pm \tau[y_s(t)]_{ij} \leq \gamma, \qquad i = 1:m, \ j = 1:N_s \tag{5c}
$$

$$
t \in (0, T).
$$

We observe that in addition to the original nonlinear ODE system in (5a), we now have N_s extra sets of linear ODE systems in (5b), one for each column of y_s. Moreover, the number of safety constraints (which are semi-infinite inequalities) has increased from m to $2mN_s$ in (5c). Moreover, it is not difficult to verify that (5c) is equivalent to

$$
y_i(t) + \tau \| e_i^T y_s(t) \|_\infty \leq \gamma, \ \ i = 1:m, \ t \in (0, T), \tag{6}
$$

where $e_i \in \Re^d$ is the i-th column of the identity matrix. In this case, the safety margin is simply τ times the infinity-norm of the gradient of y_i with respect to s evaluated at $s = \hat{s}$.

3 Robust-Optimization Formulation

To aid the development of robust formulations, we first consider using the state equation to eliminate the state variable y, even though such an action might not be advisable from the viewpoint of solving the problem. The elimination of the state variable y would result in an inequality-only optimization problem of the form:

$$
\min_{u \in \mathcal{U}} \quad \phi(u, s)
$$

$$
\text{s.t.} \quad G(u, s) \leq 0, \tag{7}
$$

where $u \in \Re^{N_u}$, $s \in \Re^{N_s}$, $G = (g_1, \cdots, g_m)^T \in \Re^m$, and we have slightly abused notation by using the same letter G in the constraints. It will be necessary to assume that the inequality constraints in (7) are strictly satisfiable; i.e., there exists a pair $(u, s) \in \Re^{N_u} \times \Re^{N_s}$ such that $G(u, s) < 0$.

We emphasize that (i) the parameter s is not precisely known, but for it we do have a reasonable estimate, $\hat{s}$, that is called the nominal value of s; (ii) our goal is to satisfy, as much as possible, the *safety constraints*, $G(u, s) \leq 0$, no matter what value the parameter s takes in a given set $\mathcal{S}$ containing $\hat{s}$ which we will define next.

3.1 Inequality-only case

Following the standard approach in robust optimization, we insist on the satisfaction of the inequality constraints for all parameter values in a given set $\mathcal{S}$, and obtain the following equivalence:

$$G(u, s) \leq 0, \ \forall s \in \mathcal{S} \iff \max_{s \in \mathcal{S}} g_i(u, s) \leq 0, \ i = 1 : m.$$

In order to solve, at least approximately but explicitly, the maximization problems on the right, it is necessary, in general, to restrict the set $\mathcal{S}$ into a simple form and to linearize the functions $g_i(u, s)$ with respect to s at a given nominal parameter value $\hat{s}$ (estimate of s). Towards this end, we first define, for $\tau > 0$ and $p \geq 1$,

$$\mathcal{S}_\tau := \{\hat{s} + \tau D\delta : \|\delta\|_p \leq 1\} \tag{8}$$

where $\delta \in \Re^{N_d}$, $N_d \leq N_s$, is the parameter variation within the unit ball in p-norm centered at $\hat{s}$, $\tau > 0$ is the magnitude of the variations which ideally should come from prior knowledge or sampling, and D is an $N_s \times N_d$ matrix.

The choice of the matrix D will have little relevance to the further development of our formulation, but it does play an important role in defining the set of variations. When the parameters have variations of different scales, one can choose D as a positive diagonal scaling matrix. If the parameters tend to have significant variations only in a certain subspace of dimension $N_d < N_s$, then one can define D as an $N_s \times N_d$ basis matrix for that subspace. In short, one can use the flexibility in D to better match the pattern of parameter variations and avoid being overly conservative. For the purpose of following the further development, however, the reader may choose to simply regard D as the $N_s \times N_s$ identity matrix.

By the first-order Taylor approximation, for $i = 1 : m$ and τ sufficiently small,

$$g_i(u, \hat{s} + \tau D\delta) \approx g_i(u, \hat{s}) + \tau \langle \nabla_s g_i(u, \hat{s}), D\delta \rangle,$$

where we use $\nabla_s g_i$ to denote the (partial) gradient of g_i with respect to s. Hence, using $\mathcal{S}_\tau$ in places of $\mathcal{S}$, we have

$$\begin{aligned} \max_{s \in \mathcal{S}_\tau} g_i(u, s) &\approx g_i(u, \hat{s}) + \tau \max_{\|\delta\|_p = 1} \langle D^T \nabla_s g_i(u, \hat{s}), \delta \rangle \\ &= g_i(u, \hat{s}) + \tau \|D^T \nabla_s g_i(u, \hat{s})\|_q, \end{aligned}$$

where $q \geq 1$ satisfies $\frac{1}{p} + \frac{1}{q} = 1$, D^T is the adjoint of D, and we have explicitly solved the linearized maximization problem on the right. The solution to this maximization problem follows from the well-known Holder's inequality

$$|\langle c, x \rangle| \leq \|x\|_p \|c\|_q \quad \text{for} \quad \frac{1}{p} + \frac{1}{q} = 1, \quad 1 \leq p, q \leq +\infty,$$

where the equality can be achieved in $\{x : \|x\|_p \leq 1\}$; that is,

$$\max_{\|x\|_p=1} \langle c, x \rangle = \|c\|_q.$$

Now replacing $\max g_i(u, s) \leq 0$ by $g_i(u, \hat{s}) + \tau \|D^T \nabla_s g_i(u, \hat{s})\|_q \leq 0$, we arrive at a robust version of the inequality constrained problem:

$$\min_{u \in \mathcal{U}} \quad \phi(u, \hat{s})$$
$$\text{s.t.} \quad g_i(u, \hat{s}) + \tau \|D^T \nabla_s g_i(u, \hat{s})\|_q \leq 0, \quad i = 1 : m. \tag{9}$$

We will call the added nonnegative terms the *"safety margins"*, which are proportional to the magnitude τ of the variations, and to the sensitivity of the constraints with respect to the parameter s at $\hat{s}$. It is worth pointing out that only the nominal value, $\hat{s}$, of the parameter s is involved in the robust version of the problem. In general, the robust formulation is considered only for local optimization due to absence of convexity.

Obviously, in general the *robust feasibility set*, defined by the inequalities in (9), is smaller than the original feasibility set defined by the inequalities without the added safety margins. The actual size of the robust feasibility set depends on the values of τ and q. More importantly, it depends on the *sensitivity* of the functions $g_i(u, s)$ with respect to the uncertain parameters s at the nominal value $\hat{s}$.

For an illustration, let us define $u = (x, y)$ and $s = (a, b)$, both in $\Re^2$. Consider the feasibility set

$$\{(x, y) : |x| \leq 2, (ax^2 + b) \leq y \leq 4(ax^2 + b)\}$$

with nominal value $\hat{s} = (\hat{a}, \hat{b}) = (0.25, 1)$. In this example, $g_1(u, s) = (ax^2 + b) - y$ and $g_2(u, s) = -4(ax^2 + b) + y$. The sensitivity vectors of $G = (g_1, g_2)$, with respect to the parameter s at $\hat{s}$, are $\nabla_s g_1(u, s) = (x^2, 1)$ and $\nabla_s g_2(u, s) = -4(x^2, 1)$. In Figure 1, we plot the original and robust feasibility sets for $q = 1, 2, \infty$, with four different τ values and with D equal to the identity.

It is worth noting that (i) the larger the τ value, the smaller the robust feasibility set is (and it will eventually become infeasible if τ is too large); (ii) at any fixed x value the upper curve g_2 is four times more sensitive to parameter changes than the lower curve g_1 is. These properties are clearly reflected in Figure 1. Indeed the safety margins for the upper curve are four times wider than those for the lower curve. This self-adaptivity is a very attractive feature.
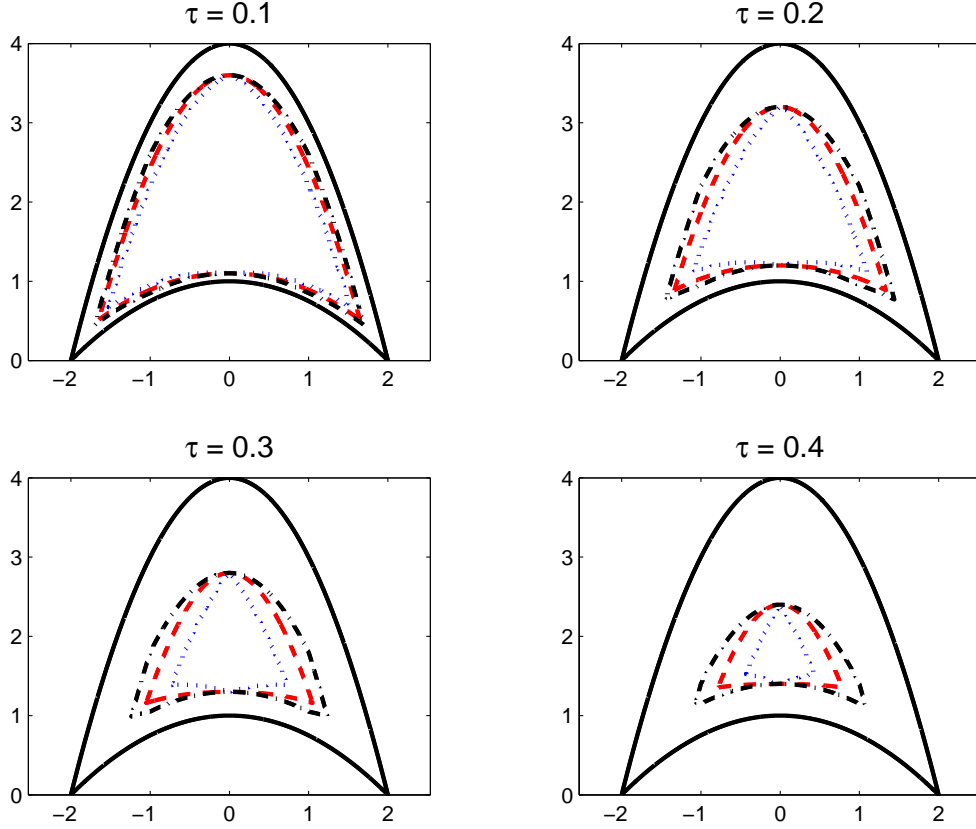
Figure 1: The original and the robust feasibility sets. The solid line depicts the original feasibility set, and the three smaller sets are robust feasibility sets for $q = 1, 2, \infty$. For each fixed τ, the largest robust feasibility set is for $q = \infty$ and the smallest for $q = 1$.

3.2 General Case

Let us define $F_y(y, u, s)$ to be the partial Jacobian of $F(y, u, s)$ with respect to y, i.e.,

$$[F_y(y, u, s)]_{ij} = \frac{\partial F_i(y, u, s)}{\partial y_j}, \quad i, j = 1, 2, \cdots, N_y.$$

Analogous notations will be used for other functions and variables as well. We also define

$$\hat{G}(u, s) = G(y(u, s), u, s), \tag{10}$$

where $y(u, s)$ is implicitly defined by the state equation $F(y, u, s) = 0$. By the Implicit Function Theorem, the function $y(u, s)$ is well defined in a neighborhood of (u_0, s_0) if there exists y_0 such that $F(y_0, u_0, s_0) = 0$ and $\nabla_y F(y_0, u_0, s_0)$ is nonsingular. In the sequel, we will assume that F_y is nonsingular in a region of interest.

Differentiating both sides of the equation $F(y(u, s), u, s) = 0$ with respect to s leads to the

matrix equation

$$F_y(y, u, s)y_s + F_s(y, u, s) = 0. \tag{11}$$

Now differentiating $\hat{G}(u, s)$ defined in (10) with respect to s, we obain

$$\hat{G}_s(u, s) = G_y(y, u, s)y_s + G_s(y, u, s), \tag{12}$$

where $y = y(u, s)$. Therefore,

$$\nabla_s \hat{g}_i(u, s)^T = e_i^T \hat{G}_s(u, s) \equiv e_i^T [G_y(y, u, s)y_s + G_s(y, u, s)], \tag{13}$$

where again $y = y(u, s)$ and $e_i \in \Re^m$ is the i-th column of the identity matrix. We note that even if g_i does not directly depend on s (i.e., $\nabla_s g_i = 0$), it may still be sensitive to changes in s because of its dependence on $y = y(u, s)$ that is defined via the state equation $F(y, u, s) = 0$.

It follows from (9), with G replaced by $\hat{G}$ (and a similar adjustment made in the objective function ϕ), a robust counterpart of (2) has the following inequality-only form:

$$\min_{u \in \mathcal{U}} \quad \phi(y(u, s), u, \hat{s})$$
$$\text{s.t.} \quad \hat{g}_i(u, \hat{s}) + \tau \|D^T \nabla_s \hat{g}_i(u, \hat{s})\|_q \leq 0, \quad i = 1 : m,$$

where the terms $\nabla_s \hat{g}_i(u, \hat{s})$, $i = 1 : m$, are defined by (13). In view of the definition for $y(u, s)$ and (10), (11) and (13), the above problem is clearly equivalent to

$$\min_{y,\, u \in \mathcal{U}} \quad \phi(y, u, \hat{s})$$
$$\text{s.t.} \quad F(y, u, \hat{s}) = 0, \tag{14a}$$
$$\tau(F_y y_s + F_s) = 0, \tag{14b}$$
$$g_i(y, u, \hat{s}) + \tau \|e_i^T(G_y y_s + G_s)D\|_q \leq 0, \quad i = 1 : m, \tag{14c}$$

where F_y, F_s, G_y and G_s are evaluated at $(y, u, \hat{s})$. The equation (14b) involves a new unknown $y_s \in \Re^{N_y \times N_s}$ and is linear in y_s. We choose to multiply (14b) by the magnitude τ in order to make it redundant whenever $\tau = 0$. Now one can easily verify that when $q = \infty$ (i.e., $p = 1$), the inequality constraints in (14c) reduce to those in (3c).

To allow the maximum flexibility, we can use different values for τ, D and q in different constraints, which leads to a more general form of the robust-optimization formulation for problem (2) with *robust safety constraints*:

$$g_i(y, u, \hat{s}) + \tau_i \|e_i^T(G_y y_s + G_s)D_i\|_{q_i} \leq 0, \quad i = 1 : m. \tag{15}$$

Without loss of generality, in the sequel we will only consider the simpler form of robust safety constraints as in (14c) instead of in (15), and in addition, we will treat D as the identity matrix.

3.3 First-order Robustness

We observe that if the original problem (2) is strictly feasible at $s = \hat{s}$, then so is the robust version (14) as long as τ, the magnitude of maximum parameter variations, is sufficiently small. The theorem below characterizes the robustness of feasible points for the robust optimization problem (14). A function $f : C \subset \Re^n \to \Re^m$, where $\Re^n$ and $\Re^m$ are endowed with a conjugate pair of p-norm and q-norm respectively, is said to be (p,q)-Lipschitz continuous in C modulo to L if

$$\|f(x) - f(y)\|_q \leq L\|x - y\|_p, \ \forall x, y \in C.$$

Recall that the set $\mathcal{S}_\tau$ is defined in (8), and we have assumed that the functions F and G are continuously differentiable.

Theorem 1. *Let $(\hat{y}, \hat{u})$ be strictly feasible to (14) corresponding to $\hat{s}$ and $\tau > 0$. Assume that in the set $\mathcal{S}_\tau$, (i) $y(\hat{u}, s)$ is implicitly defined as a differentiable function of s via the equation $F(y, \hat{u}, s) = 0$, and (ii) every row of $[G_y y_s + G_s](y(s), \hat{u}, s)$ is (p, q)-Lipschitz continuous modulo to L. Then*

$$G(y(\hat{u}, s), \hat{u}, s) \leq \frac{L}{2}\tau^2, \ \forall s \in \mathcal{S}_\tau.$$

Proof. Consider any element, $\hat{g}_i(\hat{u}, s)$, of the function $\hat{G}(u, s)$ defined in (10) with $u = \hat{u}$. It follows from calculus and the Lipcshitz continuity assumption on $\nabla_s \hat{g}_i(\hat{u}, s)$ (which is the i-th row of $G_y y_s + G_s$ evaluated at $(y(\hat{u}, s), \hat{u}, s)$) that

$$\hat{g}_i(\hat{u}, \hat{s} + h) - g_i(\hat{u}, \hat{s}) - \nabla_s \hat{g}_i(\hat{u}, \hat{s})^T h$$
$$= \int_0^1 (\nabla_s \hat{g}_i(\hat{u}, \hat{s} + th) - \nabla_s \hat{g}_i(\hat{u}, \hat{s}))^T h \, dt$$
$$\leq \int_0^1 \|\nabla_s \hat{g}_i(\hat{u}, \hat{s} + th) - \nabla_s \hat{g}_i(\hat{u}, \hat{s})\|_q \|h\|_p \, dt$$
$$\leq \int_0^1 L\|th\|_p \|h\|_p \, dt = L\|h\|_p^2 \int_0^1 t \, dt$$
$$= \frac{L}{2}\|h\|_p^2.$$

Therefore, for $s = \hat{s} + \tau\delta \in \mathcal{S}_\tau$ (i.e., $h = \tau\delta$ with $\|\delta\|_p \leq 1$)

$$g_i(y(\hat{u}, s), \hat{u}, s) \equiv \hat{g}_i(\hat{u}, \hat{s} + h)$$
$$\leq \hat{g}_i(\hat{u}, \hat{s}) + \tau \nabla_s \hat{g}_i(\hat{u}, \hat{s})^T \delta + \frac{L}{2}\|\tau\delta\|_p^2$$
$$\leq \hat{g}_i(\hat{u}, \hat{s}) + \tau\|\nabla_s \hat{g}_i(\hat{u}, \hat{s})\|_q + \frac{L}{2}\tau^2$$
$$\leq \frac{L}{2}\tau^2,$$

where the last inequality follows from the feasibility of $(\hat{y}, \hat{u})$ in (14). $\qquad\square$

We remark that if a safety constraint has a nonzero safety margin at $\hat{s}$, then by continuity there exists a neighborhood of $\hat{s}$ throughout which the safety constraint will be strictly satisfied.

It should be pointed out that solving the robust version (14) will give a design variable $\hat{u}$ that is robust for all $s \in \mathcal{S}_\tau$ in the sense given by Theorem 1, while the state variable $\hat{y}$ is good only at the nominal parameter value $\hat{s}$. This limitation is unavoidable whenever parameter uncertainty is present in the state equation, but inconsequential nevertheless in optimal control or design applications where the purpose of robust optimization is precisely to find robust control or design variables.

The robustness result given in Theorem 1 can be called the first-order robustness. While it does not guarantee the satisfaction of the safety constraint $G(y(\hat{u}, s), \hat{u}, s) \leq 0$ for all $s \in \mathcal{S}_\tau$, it does ensure that the worst-case violation is bounded by a second-order term in τ. This first-order robustness should be sufficient in applications where uncertain parameter variations are known to be relatively small.

3.4   General parameterized linear programming

It is not difficult to verify that when (i) all functions are linear, (ii) there are no uncertain equality constraints, and (iii) the parameter s is the matrix A in inequality constraints $Ax \leq b$, then our robust-optimization formulation (14) reduces to the existing robust linear programming model. Therefore, our robust-optimization formulation is indeed a generalization of the existing model. However, there is more to our formulation even in the case of linear programming.

Let us consider the following general, parameterized linear program

$$
\begin{aligned}
\min_{y,u} \quad & \langle c_0(s), y \rangle + \langle d_0(s), u \rangle + \gamma_0(s) \\
\text{s.t.} \quad & A(s)y + B(s)u - h(s) = 0, \\
& \langle c_i(s), y \rangle + \langle d_i(s), u \rangle + \gamma_i(s) \leq 0, \ i = 1 : m,
\end{aligned}
\tag{16}
$$

where $y \in \Re^{N_y}$, $u \in \Re^{N_u}$, $s \in \Re^{N_s}$, $A(s) \in \Re^{N_y \times N_y}$, and the dimensions of other coefficients all follow from the context. We will assume that all the coefficients are differentiable functions of s and, in particular, the matrix $A(s)$ is invertible in a neighborhood of some nominal value $\hat{s}$.

Since the coefficients can be linear or nonlinear functions of the uncertain parameter s, the above parameterized linear program is beyond the scope treated by the existing formulations. Besides the exclusion of uncertain equations, the existing robust optimization formulations usually assume that the uncertain parameters are the coefficients themselves.

From our general result for parameterized nonlinear programming, we can readily derive the following result.

Corollary 1. *The robust version of the parameterized linear program (16) can be written as a linear program for $p = 1$ or $p = \infty$. Moreover, it is a second-order cone program for $p = 2$.*

In the parameterized linear program (16), y is considered to be the state variable and u the design variable. Such a distinction is natural in optimal control or design problems. In other applications, however, there may not be such a clear distinction between two groups of variables. Nevertheless, in our formulation we can still split variables into two groups, say, primary and secondary variables based on some considerations. As long as we ensure that the coefficient submatrix corresponding to the secondary variables is square and nonsingular, we can still obtain robust solutions for the primary variables.

4   Concluding Remarks

Real-world optimization problems, such as those arising from optimal design of physical, medical or engineering systems, often contain parameters whose values cannot be exactly determined because of various technical difficulties. In critical and sensitive applications, such parameter uncertainties could negatively affect the quality of solutions. In particular, whenever safety is at stake, a central issue is how to ensure system safety, even in the worst case, while still achieving as good a performance as possible.

The robust optimization formulation (14), or in particular (3), extends the existing robust optimization methodology from inequality-only, convex conic programming to general nonlinear programming. Even in the heavily studied subject of robust linear programming, the extension still adds new applicability to the robust optimization approach. These robust formulations remain one-level optimization problems that need to be solved only once at nominal values; though the numbers of variables and constraints have been increased.

From a practical point of view, robust satisfaction of safety constraints boils down to the selection of safety margins. The proposed approach provides a mathematically rigorous, yet remarkably simple, mechanism to automatically and adaptively choose safety margins based on constraint sensitivity towards uncertain parameters. We believe that safety margins of this kind can strike a better balance between the safety and performance than can those based purely on heuristics.

On the other hand, we should caution that the proposed robust formulations are theoretically valid only in a neighborhood of the nominal value and robust only to the first-order. Therefore, their degrees of success will likely be dependent on the quality of parameter estimations and on the magnitude of parameter variations. Numerical experiments will be conducted to assess the behavior and performance of the proposed approach.

References

[1] Harish Agarwal. Reliability based design optimization: formulations and methodologies. Ph.D. Dissertation, University of Notre Dame, 2004.

[2] A. Ben-Tal, A. Nemirovski. Robust convex optimization. Mathematics of Operations Research, v. 23, 1998.

[3] A. Ben-Tal, A. Nemirovski. Robust solutions of uncertain linear programs. OR Letters v. 25, pp. 1-13, 1999.

[4] A. Ben-Tal, L. ElGhaoui, A. Nemirovski. Robust Semidefinite Programming. R. Saigal, H. Wolkowitcz, L. Vandenberghe, Eds. *Handbook on Semidefinite Programming*. Kluwer Academic Publishers, pp. 139-162. 2000.

[5] A. Ben-Tal, A. Nemirovski. Robust solutions of Linear Programming problems contaminated with uncertain data. Mathematical Programming (A), vol. 88, pp. 411-424, 2000.

[6] A. Ben-Tal, A. Nemirovski, C. Roos. Robust Solutions of Uncertain Quadratic and Conic-Quadratic Problems. SIAM Journal on Optimization, pp. 535-560, 2003.

[7] A. Ben-Tal, T. Margalit, A. Nemirovski. On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. SIAM Journal on Optimization. Vol. 12, pp. 811-833, 2002.

[8] A. Ben-Tal, A. Nemirovski, Robust Optimization - Methodology and Applications Mathematical Programming (Series B), 92: 453480 2002.

[9] A. Ben-Tal, A. Nemirovski. On approximate robust counterparts of uncertain semidefinite and conic quadratic programs. System Modeling and Optimization XX: Ifip Tc7 20th Conference on System Modeling and Optimization, July 23-27, 2001, Trier, Germany. Kluwer Academic Pub., 2003.

[10] G. Calafiore, L. El Ghaoui. Ellipsoidal bounds for uncertain linear equations and dynamical systems. Automatica, vol. 40, pp. 773-787, 2004.

[11] L. El Ghaoui and H. Lebret. Robust solutions to least squares problems with uncertain data. In "Recent advances in total least squares techniques and errors-in-variables modeling (Leuven, 1996)", SIAM, Philadelphia, PA, pp. 161–170, 1997.

[12] L. El Ghaoui, F. Oustry and H. Lebret. Robust Solutions to Uncertain Semidefinite Programs. SIAM J. Optimization, Vol. 9, No. 1, 1998.

[13] L. El Ghaoui and G. Calafiore. Worst-Case Simulation of Uncertain Systems. *Robustness in Identification and Control*, A. Garulli, A. Tesi and A. Vicino, Eds., Springer, 1999.

[14] L. El Ghaoui, M. Oks and F. Oustry. Worst-Case Value-at-Risk and Robust Portfolio Optimization: A Conic Programming Approach. Operations Research, vol. 51, no. 4, 2003.

[15] D. Goldfarb y G. Iyengar. Robust convex quadratically constrained programs. Mathematical Programming (A), 97(3):495-515, 2003.

[16] O. L. Mangasarian and D. R. Musicant. Robust linear and support vector regression. IEEE Trans. PAMI, 22, 9 2000.

[17] M. C. Pinar. Linear Huber M-Estimator Under Ellipsoidal Data Uncertainty Bit Numerical Mathematics, 42(4): 856-866, 2002.

[18] G. A. Watson. Robust solutions to a general class of approximation problems. SIAM J. Sci. Comp., Vol. 25, No. 4, pp. 1448-1460, 2004.