# A ROBUSTIFICATION APPROACH IN UNCONSTRAINED QUADRATIC OPTIMIZATION

MARTIN K. BERNAUER AND ROLAND GRIESSE

Abstract. Unconstrained convex quadratic optimization problems subject to parameter perturbations are considered. A robustification approach is proposed and analyzed which reduces the sensitivity of the optimal function value with respect to the parameter. Since reducing the sensitivity and maintaining a small objective value are competing goals, strategies for balancing these two objectives are discussed. Numerical examples illustrate the approach.

## 1 Introduction

We consider unconstrained quadratic optimization problems

$$\min_{x\in\mathbb{R}^n} f(x,p) = \frac{1}{2}x^\top Qx + b(p)^\top x + c(p) \tag{1.1}$$

where $x \in \mathbb{R}^n$ is the optimization variable and $p \in \mathbb{R}^m$ is a parameter. Throughout, $Q \in \mathbb{R}^{n\times n}$ is symmetric and positive definite, and $b : \mathbb{R}^m \to \mathbb{R}^n$ and $c : \mathbb{R}^m \to \mathbb{R}$ are twice continuously differentiable.

Let $x(p)$ be the unique solution of (1.1), then

$$F(p) = f(x(p), p)$$

denotes the value function, and $\|\nabla F(p)\| = \|f_p(x(p), p)\|$ is an indicator of the robustness of the optimal function value with respect to perturbations in $p$.

We propose and analyze the following modification of (1.1):

$$\min_{x\in\mathbb{R}^n} f(x,p) + \frac{\beta}{2}\|f_p(x,p_0)\|^2, \qquad \beta \geq 0. \tag{1.2}$$

Here $p_0 \in \mathbb{R}^m$ denotes some reference or nominal parameter. As noted above, the second term is related to a sensitivity measure for the unmodified value function $F(p)$. Note that the modification term depends only on problem data at the reference parameter $p_0$.

Let $\tilde{x}(p)$ be the unique solution of (1.2). For the modified function

$$\widehat{F}(p) = f(\tilde{x}(p), p),$$

we prove the result

$$\nabla\widehat{F}(p_0) = (I + 2\beta\bar{B})(I + \beta\bar{B})^{-2}\,\nabla F(p_0),$$

where $\bar{B} = b_p(p_0)^\top Q^{-1} b_p(p_0)$. This allows us to conclude that

$$\|\nabla\widehat{F}(p_0)\| \leq \|\nabla F(p_0)\|,$$

---

*Date*: June 18, 2009.

2000 *Mathematics Subject Classification.* Primary 90C31; Secondary 93B35.

*Key words and phrases.* robust optimization, robustification, parametric optimization.

i.e., the original objective function evaluated at the solution of the modified problem is more robust to first order with respect to changes in the parameter $p$. Moreover, we obtain

$$\|\nabla \widehat{F}(p_0)\| \to 0 \text{ monotonically as } \beta \to \infty,$$

if $b_p(p_0)$ has rank $m$. The modification, however, comes at the cost of an overall increased function value $\widehat{F}(p) \geq F(p)$. Thus, reducing the sensitivity and maintaining a small objective value are competing goals. To balance these goals, we discuss two strategies for the choice of the robustification parameter $\beta$. The first strategy aims at keeping the increase of the objective value $\widehat{F}(p_0) - F(p_0)$ as small as possible while reducing the sensitivity $\|\nabla \widehat{F}(p_0)\|$ below a given upper bound. The second approach minimizes the sensitivity $\|\nabla \widehat{F}(p_0)\|$ under the condition that $\widehat{F}(p_0) - F(p_0)$ does not exceed a given upper bound.

To put our work into perspective, we briefly recall two antithetic paradigms of dealing with uncertain parameters in optimization problems. On the one hand, the *robust optimization* approach aims to achieve the best possible value of the objective function under all possible circumstances (choices of the parameter), either in a worst-case [1, 4] or in an average sense [2, Chapter 6.4] and [3] and the references therein. The solution is thus designed to be acceptable for all conceivable values of the parameter, and it is not altered even if the actual value of the parameter became known. This paradigm is reflected by the bi-level structure of worst-case robust optimization problems.

On the other hand, in an *online* or *parametric optimization* approach, the solution follows the actual choice of the parameter. This can be achieved by either resolving the problem from scratch whenever a change in the parameter occurs, or by expressing the solution in terms of a feedback law $x(p)$ [7–9], or by updating the solution of a nearby problem using parametric sensitivity derivatives [5].

Our approach can be interpreted as a mixture of the two. We allow the solution to follow the parameter, but we also modify the objective in order to reduce the sensitivity of the optimal function value. The modification term depends on the choice of the reference parameter $p_0$ and of the robustification parameter $\beta \geq 0$.

While our motivation to study (1.2) is mainly intrinsic, we mention a potential application. Suppose that $f$ is a model for the quality of the product of a certain production process, which is to be maximized. The process is subject to a parameter $p$ which may involve, e.g., environmental conditions or properties of the raw materials. The parameter $p$ varies throughout the production process, and a controller $x(p)$ continuously solves (1.1) to ensure optimal product quality at all times. As a result, the quality indicator $F(p) = f(x(p), p)$ may vary strongly, which is undesirable. The modified problem (1.2), which gives rise to the modified controller $\tilde{x}(p)$, reduces the influence of parameter perturbations on the final product quality $\widehat{F}(p) = f(\tilde{x}(p), p)$ in a neighborhood of the reference parameter $p_0$.

The material is organized as follows. In Section 2 we derive representations for the function values and their derivatives of the original and modified problems, and Section 3 is devoted to their comparison. The dependence on the robustification parameter $\beta$ of the function value $\widehat{F}(p_0)$ and its sensitivity w.r.t. $p$ are analyzed in Section 4. This leads to several strategies for the selection of $\beta$ in Section 5. Numerical examples are presented in Section 6, which include a discretized optimal control (LQR) problem.

We make frequent use of the Sherman-Morrison-Woodbury formula

$$(A + UV^\top)^{-1} = A^{-1} - A^{-1}U(I + V^\top A^{-1}U)^{-1}V^\top A^{-1}, \tag{1.3}$$

whenever all inverses exist. All norms are either the Euclidean vector norm or its associated matrix norm. We use $b_p$ to denote the Jacobian of a vector valued function $b : \mathbb{R}^m \to \mathbb{R}^n$. Consequently, $f_x^\top$ or $\nabla f$ denote the gradient of a scalar valued function $f : \mathbb{R}^n \to \mathbb{R}$.

## 2 Preliminaries

In this section we derive representations for optimal function values and their derivatives. We begin with the original problem

$$\min_{x\in\mathbb{R}^n} f(x,p) = \frac{1}{2} x^\top Q x + b(p)^\top x + c(p) \tag{2.1}$$

and recall our standing assumptions:

**Assumption 2.1.** (1) *$Q \in \mathbb{R}^{n\times n}$ is symmetric and positive definite.*
(2) *$b : \mathbb{R}^m \to \mathbb{R}^n$ and $c : \mathbb{R}^m \to \mathbb{R}$ are twice continuously differentiable.*

The unique minimizer of (2.1) is given by

$$x(p) = -Q^{-1} b(p).$$

For the value function, we thus conclude

$$F(p) = f(x(p), p) = -\frac{1}{2} b(p)^\top Q^{-1} b(p) + c(p).$$

Taking derivatives leads to the following proposition. Here and in the sequel, we abbreviate

$$B_0 := b_p(p_0).$$

**Proposition 2.2.** *The value function $F$, its gradient and its Hessian matrix at $p = p_0$ are given by*

$$\begin{aligned} F(p_0) &= -\frac{1}{2} b(p_0)^\top Q^{-1} b(p_0) + c(p_0), \\ \nabla F(p_0) &= -B_0^\top Q^{-1} b(p_0) + c_p^\top(p_0), \\ \left(\nabla^2 F(p_0)\right)_{ij} &= -b_{p_i}^\top(p_0)\, Q^{-1} b_{p_j}(p_0) - b_{p_i p_j}(p_0)^\top\, Q^{-1} b(p_0) + c_{p_i p_j}(p_0), \end{aligned}$$

*where $1 \le i, j \le m$.*

### 2.1. The Modified Problem—Discussion of $\tilde{F}$.

We now turn to the proposed modification of (2.1):

$$\min_{x\in\mathbb{R}^n} \tilde{f}(x,p) = f(x,p) + \frac{\beta}{2} \left\| f_p(x, p_0) \right\|^2. \tag{2.2}$$

The gradient of $\tilde{f}$ w.r.t. $x$ is given by

$$\begin{aligned} \tilde{f}_x^\top(x,p) &= Qx + b(p) + \beta \sum_{i=1}^m \left( b_{p_i}^\top(p_0)\, x + c_{p_i}(p_0) \right) b_{p_i}(p_0) \\ &= (Q + \beta B_0 B_0^\top)\, x + b(p) + \beta B_0\, c_p^\top(p_0), \end{aligned}$$

and the matrix $Q + \beta B_0 B_0^\top$ is symmetric and positive definite. Thus, the unique minimizer of (2.2) is characterized by $\tilde{f}_x(\tilde{x}(p), p) = 0$, or equivalently,

$$\tilde{x}(p) = -(Q + \beta B_0 B_0^\top)^{-1} \left( b(p) + \beta B_0\, c_p^\top(p_0) \right). \tag{2.3}$$

The Sherman-Morrison-Woodbury formula (1.3) implies that

$$(Q + \beta B_0 B_0^\top)^{-1} = Q^{-1} - \beta\, Q^{-1} B_0 (I + \beta \bar{B})^{-1} B_0^\top Q^{-1}. \tag{2.4}$$

Here and in the sequel, we use the abbreviation

$$\bar{B} := B_0^\top Q^{-1} B_0.$$

In order to calculate the value function associated to (2.2)

$$\tilde{F}(p) = \tilde{f}(\tilde{x}(p), p)$$

and its first and second derivatives, we state the following lemma whose proof follows by direct calculation.

**Lemma 2.3.** *Let $M \in \mathbb{R}^{n\times n}$ such that the matrix $I + M$ is non-singular. Then*

$$I - (I + M)^{-1} M = (I + M)^{-1}, \tag{2.5}$$

$$I - M(I + M)^{-1} = (I + M)^{-1}. \tag{2.6}$$

**Proposition 2.4.** *The value function $\tilde{F}$, its gradient and its Hessian matrix at $p = p_0$ are given by*

$$\begin{aligned}
\tilde{F}(p_0) &= f(\tilde{x}(p_0), p_0) + \frac{\beta}{2}\big\|\nabla \tilde{F}(p_0)\big\|^2,\\
\nabla \tilde{F}(p_0) &= (I + \beta\bar{B})^{-1} \nabla F(p_0),\\
\big(\nabla^2 \tilde{F}(p_0)\big)_{ij} &= \big(\nabla^2 F(p_0)\big)_{ij} + b_{p_j}^\top(p_0)\,\beta\, Q^{-1} B_0 (I + \beta\bar{B})^{-1} B_0^\top Q^{-1} b_{p_i}(p_0)\\
&\quad - b_{p_i p_j}(p_0)^\top \beta\, Q^{-1} B_0 (I + \beta\bar{B})^{-1} \nabla F(p_0),
\end{aligned}$$

*where $1 \le i, j \le m$.*

The proof of this proposition is given in Appendix A.1.

### 2.2. Discussion of $\widehat{F}$.

We now turn to the discussion of the function

$$\widehat{F}(p) = f(\tilde{x}(p), p),$$

i.e., the objective of the original problem (2.1) evaluated at the solution $\tilde{x}(p)$ of the modified problem (2.2). For convenience, we define

$$\mathcal{I} := (I + \beta\bar{B})^{-1} \beta\bar{B} (I + \beta\bar{B})^{-1}$$

and recall that $I + \beta\bar{B}$ is invertible for all $\beta \ge 0$ since $\bar{B}$ is symmetric and positive semidefinite. Moreover, $\mathcal{I}$ is symmetric and positive semi-definite.

**Proposition 2.5.** *The value function $\widehat{F}$, its gradient and the entries of its Hessian matrix at $p = p_0$ are given by*

$$\begin{aligned}
\widehat{F}(p_0) &= F(p_0) + \frac{\beta}{2} \nabla F(p_0)^\top\, \mathcal{I}\, \nabla F(p_0),\\
\nabla \widehat{F}(p_0) &= (I + 2\beta\bar{B})(I + \beta\bar{B})^{-2} \nabla F(p_0),\\
\big(\nabla^2 \widehat{F}(p_0)\big)_{ij} &= \big(\nabla^2 F(p_0)\big)_{ij} + \beta^2 b_{p_j}^\top(p_0)\, Q^{-1} B_0 (I + \beta\bar{B})^{-2} \bar{B} B_0^\top Q^{-1} b_{p_i}(p_0)\\
&\quad - \beta^2 \nabla F(p_0)^\top \bar{B} (I + \beta\bar{B})^{-2} B_0^\top Q^{-1} b_{p_i p_j}(p_0),
\end{aligned}$$

*where $1 \le i, j \le m$.*

We refer to Appendix A.2 for the proof.

## 3 Main Results

We recall that the motivation of introducing the modified problem (2.2) lies in the desired reduction of the sensitivity of the objective. To verify that this goal is met, we investigate in this section the relations between the values of $F$, $\tilde{F}$, and $\widehat{F}$ at $p_0$, and their first and second derivatives.

The discussion is based on a diagonalization of the matrix $\bar{B} = B_0^\top Q^{-1} B_0$ of the form

$$\bar{B} = UDU^\top, \quad U^\top U = I, \quad D = \operatorname{diag}(\lambda_1, \ldots, \lambda_m), \quad \lambda_i \geq 0, \quad 1 \leq i \leq m. \tag{3.1}$$

The existence of such a diagonalization follows from the symmetry and positive semidefiniteness of $\bar{B}$. We note that (3.1) allows us to write $I + 2\beta\bar{B} = U(I + 2\beta D)U^\top$ and $(I + \beta\bar{B})^{-1} = U(I + \beta D)^{-1}U^\top$ etc.

**Remark 3.1.** (a) *In case $B_0 = b_p(p_0)$ has rank $m$, $\bar{B} = B_0^\top Q^{-1} B_0$ is positive definite and $\lambda_i > 0$ holds for all $1 \leq i \leq m$.*

(b) *Typically the number of parameters $m$ is smaller than the number of optimization variables $n$. In this case, the rank of $B_0$ is smaller than $m$ if and only if its columns are linearly dependent. That is, the influence of one or several of the parameters on the solution of (2.1) can be expressed by linear combinations of other parameters. Therefore, one may eliminate those parameters from (2.1) to achieve that the rank of $B_0$ equals $m$.*

We are now in the position to formulate and prove the main result. It states that the proposed modification (2.2) reduces the sensitivity of the optimal function value with respect to perturbations of the uncertain parameter.

**Theorem 3.2.** *The gradients of $F$, $\tilde{F}$ and $\widehat{F}$ satisfy the following relation:*

$$\begin{aligned} \|\nabla\tilde{F}(p_0)\| &= \|(I + \beta\bar{B})(I + 2\beta\bar{B})^{-1}\nabla\widehat{F}(p_0)\| &\leq \quad \|\nabla\widehat{F}(p_0)\|, \\ \|\nabla\widehat{F}(p_0)\| &= \|(I + 2\beta\bar{B})(I + \beta\bar{B})^{-2}\nabla F(p_0)\| &\leq \quad \|\nabla F(p_0)\| \end{aligned}$$

*for all $\beta \geq 0$. Moreover, if $B_0$ has rank $m$, the inequalities are strict for $\beta > 0$.*

*Proof.* We only show $\|\nabla\widehat{F}(p_0)\| \leq \|\nabla F(p_0)\|$, the proof of $\|\nabla\tilde{F}(p_0)\| \leq \|\nabla\widehat{F}(p_0)\|$ is analogous. Proposition 2.5 implies that

$$\nabla\widehat{F}(p_0) = (I + 2\beta\bar{B})(I + \beta\bar{B})^{-2}\nabla F(p_0)$$

holds. Using (3.1), we can write

$$\begin{aligned} (I + 2\beta\bar{B})(I + \beta\bar{B})^{-2} &= U(I + 2\beta D)(I + \beta D)^{-2}U^\top \\ &= U(I + \beta D)^{-1}(I + 2\beta D)(I + \beta D)^{-1}U^\top. \end{aligned}$$

The last equality is a consequence of Lemma 2.3 and it shows that this matrix is symmetric. Its eigenvalues are

$$0 < \frac{1 + 2\beta\lambda_i}{(1 + \beta\lambda_i)^2} \leq 1, \quad 1 \leq i \leq m.$$

Consequently, its spectral norm is bounded by 1 from above. And hence we get

$$\begin{aligned} \|\nabla\widehat{F}(p_0)\| &= \|(I + 2\beta\bar{B})(I + \beta\bar{B})^{-2}\nabla F(p_0)\| \\ &\leq \|(I + 2\beta\bar{B})(I + \beta\bar{B})^{-2}\| \cdot \|\nabla F(p_0)\| \leq \|\nabla F(p_0)\|. \end{aligned}$$

Moreover, if $B_0$ has rank $m$ and $\beta > 0$, the spectral norm is bounded away from 1 and strict inequality holds. ☐

We now turn to the discussion of the function values. By construction, the monotonicity property

$$\tilde{F}(p) \geq \widehat{F}(p) \geq F(p)$$

holds for all values of $p$. The following theorem states the exact relations between the value functions.

**Theorem 3.3.** *The function values of* $F$, $\tilde{F}$ *and* $\widehat{F}$ *satisfy*

$$\begin{aligned} \tilde{F}(p) &= \widehat{F}(p) + \frac{\beta}{2} \left\| f_p(\tilde{x}(p), p_0) \right\|^2 && \geq \quad \widehat{F}(p), \\ \widehat{F}(p) &= F(p) + \frac{\beta}{2} \big( B_0^\top Q^{-1} b(p) - c_p^\top(p_0) \big)^\top \, \mathcal{I} \, \big( B_0^\top Q^{-1} b(p) - c_p^\top(p_0) \big) && \geq \quad F(p) \end{aligned}$$

*for all* $\beta \geq 0$ *and all* $p \in \mathbb{R}^m$*.*

*Proof.* The claim follows directly from the proofs of Propositions 2.4 and 2.5, see Appendix A. ☐

Theorems 3.2 and 3.3 show that the reduced sensitivity of $\widehat{F}$ compared to $F$ comes at the cost of an increased function value. Thus, reducing the sensitivity and maintaining a small objective value are competing goals. Therefore, we further analyze the dependence on $\beta$ in Section 4, which leads to various possible selection strategies, see Section 5.

Under an additional assumption we also obtain a comparison result for the second derivatives.

**Theorem 3.4.** *If* $b$ *is an affine function of* $p$*, then*

$$\nabla^2 \tilde{F}(p_0) \succeq \nabla^2 \widehat{F}(p_0) \succeq \nabla^2 F(p_0)$$

*holds for all* $\beta \geq 0$*, where* $A \succeq B$ *indicates that* $A - B$ *is positive semidefinite.*

*Proof.* In case that $b$ is affine we have $b_{p_i p_j}(p) = 0$ for all $1 \leq i, j \leq m$, which implies

$$\begin{aligned} \big(\nabla^2 \tilde{F}(p_0)\big)_{ij} &= \big(\nabla^2 F(p_0)\big)_{ij} + \beta b_{p_j}^\top(p_0) \, Q^{-1} B_0 (I + \beta \bar{B})^{-1} B_0^\top Q^{-1} b_{p_i}(p_0), \\ \big(\nabla^2 \widehat{F}(p_0)\big)_{ij} &= \big(\nabla^2 F(p_0)\big)_{ij} + \beta^2 b_{p_j}^\top(p_0) \, Q^{-1} B_0 (I + \beta \bar{B})^{-2} \bar{B} B_0^\top Q^{-1} b_{p_i}(p_0), \end{aligned}$$

see Propositions 2.4 and 2.5. Consequently, we obtain by Lemma 2.3

$$\nabla^2 \tilde{F}(p_0) - \nabla^2 \widehat{F}(p_0) = \beta \bar{B} (I + \beta \bar{B})^{-2} \bar{B}.$$

For each $v \in \mathbb{R}^m$ we have

$$\big( \beta \bar{B} (I + \beta \bar{B})^{-2} \bar{B} \, v, v \big) = \beta \left( (I + \beta \bar{B})^{-1} \bar{B} v, (I + \beta \bar{B})^{-1} \bar{B} \, v \right) \geq 0,$$

which shows $\nabla^2 \tilde{F}(p_0) \succeq \nabla^2 \widehat{F}(p_0)$. Similarly, we can show $\nabla^2 \widehat{F}(p_0) \succeq \nabla^2 F(p_0)$. ☐

We conclude this section by showing that the proposed modification also improves the stability properties of the solution.

**Theorem 3.5.** *The Jacobian of* $\tilde{x}(p)$ *satisfies the following estimate for all* $\beta \geq 0$*:*

$$\begin{aligned} \|\tilde{x}_p(p_0)\| &= \|(I + \beta \bar{B})^{-1} x_p(p_0)\| \\ &\leq \|(I + \beta \bar{B})^{-1}\| \cdot \|x_p(p_0)\| \leq \|x_p(p_0)\|. \end{aligned}$$

*Proof.* From the proof of Proposition 2.5 (Appendix A.2) we recall that

$$\tilde{x}(p) = x(p) - \beta\, Q^{-1} B_0 (I + \beta \bar{B})^{-1} \big( -B_0^\top Q^{-1} b(p) + c_p^\top (p_0) \big). \tag{3.2}$$

Differentiation with respect to $p$ shows that

$$\tilde{x}_p(p) = x_p(p) - \beta\, Q^{-1} B_0 (I + \beta \bar{B})^{-1} \big( -B_0^\top Q^{-1} b_p(p) \big).$$

As $x_p(p) = -Q^{-1} b_p(p)$, we infer from Lemma 2.3

$$\begin{aligned} \tilde{x}_p(p_0) &= x_p(p_0) + \beta x_p(p_0)(I + \beta \bar{B})^{-1}(-B_0^\top Q^{-1} B_0) \\ &= x_p(p_0)\big(I - (I + \beta \bar{B})^{-1} \beta \bar{B}\big) \quad = \quad x_p(p_0)(I + \beta \bar{B})^{-1}. \end{aligned}$$

Taking norms concludes the proof in view of $\|(I + \beta \bar{B})^{-1}\| \leq 1$ for all $\beta \geq 0$. ☐

## 4 Dependence on $\beta$

It stands out as a result of the previous section that reducing the sensitivity and maintaining a small objective value are competing goals. Therefore, we discuss here the dependence of $\widehat{F}$, of its derivative and of $\tilde{x}$ on the parameter $\beta$. This will allow us to formulate strategies for the optimal selection of $\beta$.

To simplify the notation, we define the functions

$$l_0(\beta) := \widehat{F}(p_0) - F(p_0), \qquad l_1(\beta) := \|\nabla \widehat{F}(p_0)\|^2.$$

**Theorem 4.1.** *The function $l_0$ is continuous, monotonically increasing, and bounded above. If $B_0$ has rank $m$, then $l_0$ is strictly increasing and it converges to $\frac{1}{2}\nabla F(p_0)^\top \bar{B}^{-1} \nabla F(p_0)$ as $\beta \to \infty$.*

*Proof.* Using Proposition 2.5 and the decomposition (3.1), we may write

$$\begin{aligned} l_0(\beta) &= \frac{\beta}{2} \nabla F(p_0)^\top\, \mathcal{I}\, \nabla F(p_0) \\ &= \frac{1}{2} \nabla F(p_0)^\top U \operatorname{diag}\left( \frac{\beta^2 \lambda_1}{(1 + \beta\lambda_1)^2}, \dots, \frac{\beta^2 \lambda_m}{(1 + \beta\lambda_m)^2} \right) U^\top \nabla F(p_0), \end{aligned}$$

which shows the continuity of $l_0(\beta)$. The entries in the diagonal matrix are either zero (if $\lambda_i = 0$) or they are strictly increasing with $\beta$ and converge to $\lambda_i^{-1}$. This implies that $l_0(\beta)$ is bounded and increasing, or strictly increasing if all $\lambda_i \neq 0$. If $B_0$ has rank $m$, then $\bar{B}^{-1} = U \operatorname{diag}(\lambda_1^{-1}, \dots, \lambda_m^{-1}) U^\top$ holds, which concludes the proof. ☐

**Corollary 4.2.** *Differentiating the diagonal matrix in the proof above, we infer the following representation of the derivative:*

$$l_0'(\beta) = \beta\, \nabla F(p_0)^\top \bar{B} (I + \beta \bar{B})^{-3}\, \nabla F(p_0). \tag{4.1}$$

**Theorem 4.3.** *The function $l_1$ is continuous, monotonically decreasing, and bounded below. If $B_0$ has rank $m$, then $l_1$ is strictly decreasing and it converges to 0 as $\beta \to \infty$.*

*Proof.* Using Proposition 2.5 and the decomposition (3.1), we infer that

$$\begin{aligned} l_1(\beta) &= \nabla F(p_0)^\top U \big( (I + \beta D)^{-2} (I + 2\beta D)^2 (I + \beta D)^{-2} \big) U^\top \nabla F(p_0) \\ &= \nabla F(p_0)^\top U \operatorname{diag}\left( \frac{(1 + 2\beta\lambda_1)^2}{(1 + \beta\lambda_1)^4}, \dots, \frac{(1 + 2\beta\lambda_m)^2}{(1 + \beta\lambda_m)^4} \right) U^\top \nabla F(p_0). \end{aligned}$$

The rest of the proof follows similarly as the proof of Theorem 4.1. ☐

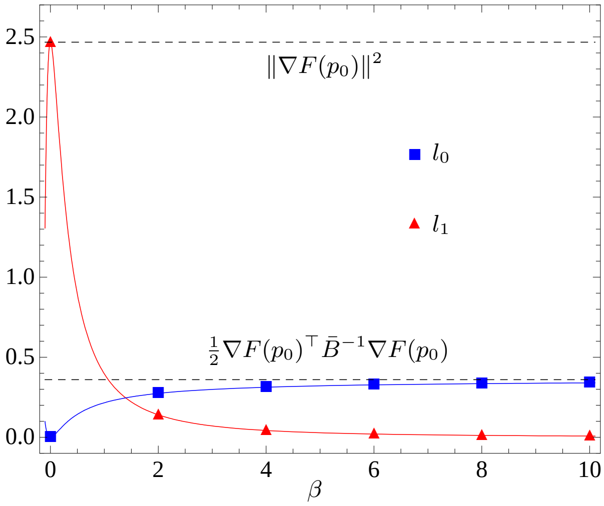


FIGURE 4.1. Typical behavior of $l_0(\beta)$ and $l_1(\beta)$.

**Corollary 4.4.** *The derivative can be written as*

$$l_1'(\beta) = -4\beta\nabla F(p_0)^\top (I+\beta\bar{B})^{-2}(I+2\beta\bar{B})(I+\beta\bar{B})^{-1}\bar{B}^2(I+\beta\bar{B})^{-2}\nabla F(p_0). \quad (4.2)$$

A typical plot for $l_0(\beta)$ and $l_1(\beta)$ is shown in Figure 4.1.

For completeness, we also state a result about the second derivative.

**Theorem 4.5.** *If $B_0$ has rank $m$, then the second derivative of $\widehat{F}$ satisfies*

$$\begin{aligned}\big(\nabla^2 \widehat{F}(p_0)\big)_{ij} \to \big(\nabla^2 F(p_0)\big)_{ij} &+ b_{p_j}^\top(p_0)\, Q^{-1}B_0\bar{B}^{-1}B_0^\top Q^{-1}b_{p_i}(p_0)\\ &- b_{p_ip_j}(p_0)^\top Q^{-1}B_0\bar{B}^{-1}\nabla F(p_0) \quad \text{as } \beta\to\infty.\end{aligned}$$

*Proof.* From Proposition 2.5 we deduce that it suffices to consider the expression $\beta^2(I+\beta\bar{B})^{-2}\bar{B}$. By (3.1) we conclude

$$\beta^2(I+\beta\bar{B})^{-2}\bar{B} = U\operatorname{diag}\left(\frac{\beta^2\lambda_1}{(1+\beta\lambda_1)^2},\ldots,\frac{\beta^2\lambda_m}{(1+\beta\lambda_m)^2}\right)U^\top \to \bar{B}^{-1} \text{ as } \beta\to\infty.$$

□

It remains to discuss the dependence of $\tilde{x}$ on $\beta$. To simplify notation, we introduce the function

$$l_2(\beta) = \frac{1}{2}\,\|\tilde{x}(p_0) - x(p_0)\|^2.$$

**Theorem 4.6.** *The function $l_2$ is continuous. If $B_0$ has rank $m$, then $l_2$ is bounded above and it converges to*

$$\frac{1}{2}\,\nabla F(p_0)^\top \bar{B}^{-1}B_0^\top Q^{-2}B_0\,\bar{B}^{-1}\,\nabla F(p_0) \quad \text{as } \beta\to\infty.$$

*Proof.* We recall from (3.2)

$$\frac{1}{2}\|\tilde{x}(p)-x(p)\|^2 = \frac{1}{2}\big\|\beta\, Q^{-1}B_0(I+\beta\bar{B})^{-1}\big(-B_0^\top Q^{-1}b(p) + c_p^\top(p_0)\big)\big\|^2.$$

For the particular case $p = p_0$, and using Proposition 2.2, we infer

$$l_2(\beta) = \frac{1}{2}\,\nabla F(p_0)^\top\beta\,(I+\beta\bar{B})^{-1}B_0^\top Q^{-2}B_0\,\beta\,(I+\beta\bar{B})^{-1}\,\nabla F(p_0),$$

which shows the continuity of $l_2(\beta)$. If $B_0$ has rank $m$, then $\beta\,(I+\beta\bar{B})^{-1}$ converges to $\bar{B}$ as $\beta\to\infty$. □

In contrast to $l_0$ and $l_1$, $l_2$ does not in general verify monotonicity with respect to $\beta$.

**Corollary 4.7.** *Differentiating the formula for $l_2$ from the proof above, we infer the following representation of the derivative:*

$$l_2'(\beta) = \nabla F(p_0)^\top (I + \beta\bar{B})^{-1}\, \beta\, (I + \beta\bar{B})^{-1} B_0^\top Q^{-2} B_0\, (I + \beta\bar{B})^{-1}\, \nabla F(p_0).$$

As $l_2'(\beta)$ involves the product of a positive definite and a positive semidefinite matrix which is in general not positive semidefinite, we can not expect $l_2$ to be monotone.

**Remark 4.8.** *We remark that*

$$\frac{1}{2}\|\tilde{x}(p_0) - x(p_0)\|_Q^2 = l_0(\beta)$$

*holds. And thus the solution deviation depends monotonically on $\beta$ when measured in the norm induced by $Q$, see Theorem 4.1. Indeed, we have*

$$\begin{aligned} \frac{1}{2}\|\tilde{x}(p_0) - x(p_0)\|_Q^2 &= \frac{1}{2}\,\nabla F(p_0)^\top \beta\, (I + \beta\bar{B})^{-1} B_0^\top Q^{-1} B_0\, \beta\, (I + \beta\bar{B})^{-1}\, \nabla F(p_0) \\ &= \frac{1}{2}\,\nabla F(p_0)^\top \beta\, (I + \beta\bar{B})^{-1}\, \bar{B}\, \beta\, (I + \beta\bar{B})^{-1}\, \nabla F(p_0) = l_0(\beta). \end{aligned}$$

## 5 Algorithms for the Choice of $\beta$

In this section, we propose two strategies concerning the choice of the robustification parameter $\beta$. We recall that reducing the sensitivity and maintaining a small objective value are competing goals, see Theorems 3.2 and 3.3. The proposed strategies differ in which goal takes precedence over the other.

The first strategy aims at keeping the increase of the objective value $l_0(\beta) = \widehat{F}(p_0) - F(p_0)$ as small as possible while reducing the sensitivity $\|\nabla\widehat{F}(p_0)\|$ below a given upper bound. By Theorem 3.2, $\alpha_0\|\nabla F(p_0)\|$ is a reasonable upper bound for some $\alpha_0 \leq 1$. The monotonicity of $l_0$ (see Theorem 4.1) then leads to the following problem:

$$\min_{\beta\geq 0}\ \beta \quad \text{s.t.} \quad \|\nabla\widehat{F}(p_0)\| \leq \alpha_0\|\nabla F(p_0)\|. \tag{S0}$$

The second approach minimizes the sensitivity $l_1(\beta) = \|\nabla\widehat{F}(p_0)\|^2$ under the condition that $l_0(\beta)$ does not exceed a given upper bound. By Theorem 3.3, valid upper bounds are $\alpha_1 \geq 0$. In view of the monotonicity of $l_1(\beta)$ (see Theorem 4.3), we arrive at:

$$\max_{\beta\geq 0}\ \beta \quad \text{s.t.} \quad \widehat{F}(p_0) - F(p_0) \leq \alpha_1. \tag{S1}$$

We now confirm that these two strategies are well-posed.

**Theorem 5.1.** (a) *There exists $\underline{\alpha} < 1$ such that problem* (S0) *is uniquely solvable for all $\alpha_0 \in (\underline{\alpha}, 1]$.*
(b) *If $B_0$ has rank $m$, then problem* (S0) *is uniquely solvable for all $\alpha_0 \in (0, 1]$.*

*Proof.* Theorem 4.3 shows that $l_1(\beta)$ is decreasing and bounded below as $\beta \to \infty$. If $\nabla F(p_0) = 0$ holds, then $\nabla\widehat{F}(p_0) = 0$ for all $\beta \geq 0$ and nothing is to be proved. Otherwise we choose

$$\underline{\alpha} = \frac{\lim\limits_{\beta\to\infty} \left\|\nabla\widehat{F}(p_0)\right\|}{\|\nabla F(p_0)\|}.$$

By monotonicity and continuity of $l_1(\beta)$, there exists a minimal $\beta \geq 0$ such that $\|\nabla \widehat{F}(p_0)\| = \alpha_0 \|\nabla F(p_0)\|$ holds, which shows (a). If $B_0$ has rank $m$, Theorem 4.3 implies $\underline{\alpha} = 0$, and (b) is proved. ☐

**Remark 5.2.** *If $B_0$ has rank $m$, $l_1(\beta)$ is strictly decreasing and the solution of* (S0) *reduces to the solution of the nonlinear equation $l_1(\beta) = \alpha_0^2 \, \|\nabla F(p_0)\|^2$ for $\beta$. This can be carried out, e.g., by bisection or Newton's method. The derivative of $l_1$ is given in* (4.2)*.*

**Theorem 5.3.** (a) *There exists $\overline{\alpha} > 0$ such that problem* (S1) *is uniquely solvable for all $\alpha_1 \in [0, \overline{\alpha})$.*
(b) *If $B_0$ has rank $m$, then* (a) *holds for $\overline{\alpha} = \frac{1}{2} \nabla F(p_0)^\top \bar{B}^{-1} \nabla F(p_0)$.*

*Proof.* We set

$$\overline{\alpha} = \lim_{\beta \to \infty} \widehat{F}(p_0) - F(p_0).$$

The argument is similar to the proof of Theorem 5.1 and we leave out the details. For the properties of $l_0(\beta)$, see Theorem 4.1. ☐

**Remark 5.4.** *If $B_0$ has rank $m$, $l_0(\beta)$ is strictly increasing and the solution of* (S1) *reduces to the solution of the nonlinear equation $l_0(\beta) = \alpha_1$ for $\beta$. This can be carried out, e.g., by bisection or Newton's method. The derivative of $l_0$ is given in* (4.1)*.*

We finally recall that strategy (S1) is equivalent to

$$\max_{\beta \geq 0} \; \beta \quad \text{s.t.} \quad \|\tilde{x}(p_0) - x(p_0)\|_Q^2 \leq 2\,\alpha_1,$$

which imposes a bound on the deviation of the solution, rather than on the function values, see Remark 4.8.

## 6 Examples

In this section, we consider three examples in order to verify the properties of the proposed modification (2.2). We also discuss its applicability to unconstrained *nonlinear* problems and a linear-quadratic optimal control (LQR) problem.

**6.1. A Quadratic Example.** The quadratic objective we consider in this section is

$$f(x,p) = \frac{1}{2} x^\top \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} x + \begin{pmatrix} -p \\ -2\,p \end{pmatrix}^\top x + p + p^2.$$

Its unique minimizer

$$x(p) = \left( \frac{p}{2}, p \right)^\top \tag{6.1}$$

generates the value function

$$F(p) = p - \frac{p^2}{4}.$$

The unique minimizer of the modified problem is given by

$$\tilde{x}(p) = \left( \frac{\beta + p}{2 + 5\,\beta}, \frac{2\,(\beta + p)}{2 + 5\,\beta} \right)^\top, \tag{6.2}$$

yielding the modified value function

$$\widehat{F}(p) = \frac{5\,\beta^2 + 4\,p\,(1 + 5\,\beta) + p^2\,(25\,\beta^2 - 5\,\beta - 1)}{(2 + 5\,\beta)^2}.$$

Table 6.1. Results of applying strategy (S0) for Example 6.1 and various values of $\alpha_0$.

| $\alpha_0$ | $\beta$ | $F(p_0)$ | $\nabla F(p_0)$ | $\widehat{F}(p_0)$ | $\nabla \widehat{F}(p_0)$ | $\tilde{F}(p_0)$ | $\nabla \tilde{F}(p_0)$ |
|---|---|---|---|---|---|---|---|
| 0.9000 | 0.1850 | 0.0000 | 1.0000 | 0.0200 | 0.9000 | 0.0632 | 0.6838 |
| 0.7000 | 0.4844 | 0.0000 | 1.0000 | 0.0600 | 0.7000 | 0.1095 | 0.4523 |
| 0.5000 | 0.9657 | 0.0000 | 1.0000 | 0.1000 | 0.5000 | 0.1414 | 0.2929 |
| 0.3000 | 2.0489 | 0.0000 | 1.0000 | 0.1400 | 0.3000 | 0.1673 | 0.1633 |
| 0.1000 | 7.3947 | 0.0000 | 1.0000 | 0.1800 | 0.1000 | 0.1897 | 0.0513 |

Table 6.2. Results of applying strategy (S1) for Example 6.1 and various values of $\alpha_1$.

| $\alpha_1$ | $\beta$ | $F(p_0)$ | $\nabla F(p_0)$ | $\widehat{F}(p_0)$ | $\nabla \widehat{F}(p_0)$ | $\tilde{F}(p_0)$ | $\nabla \tilde{F}(p_0)$ |
|---|---|---|---|---|---|---|---|
| 0.0100 | 0.0701 | 0.0000 | 1.0000 | 0.0044 | 0.9778 | 0.0298 | 0.8509 |
| 0.0500 | 0.2000 | 0.0000 | 1.0000 | 0.0222 | 0.8889 | 0.0667 | 0.6667 |
| 0.1000 | 0.3567 | 0.0000 | 1.0000 | 0.0444 | 0.7778 | 0.0943 | 0.5286 |
| 0.2000 | 0.8000 | 0.0000 | 1.0000 | 0.0889 | 0.5556 | 0.1333 | 0.3333 |
| 0.3000 | 1.7798 | 0.0000 | 1.0000 | 0.1333 | 0.3333 | 0.1633 | 0.1835 |

For the nominal value $p_0 = 0$, we obtain $x_0 = x(p_0) = (0, 0)^\top$ with an optimal function value $F(p_0) = f(x_0, p_0) = 0$ and the sensitivity $\nabla F(p_0) = 1$. Based on this sensitivity information, we apply (S0) and (S1) to find an appropriate damping parameter $\beta$ for different values of $\alpha_0$ and $\alpha_1$. The results are reported in Tables 6.1 and 6.2 and they clearly reflect the monotonicity properties established in Theorems 3.2, 3.3, 4.1 and 4.3.

In addition, we plot the three value functions for two different values of $\beta$ and $p \in [-1, 1]$ in Figure 6.1. In both cases, the ordering of the second derivatives (Theorem 3.4) can be observed. To illustrate that the robustification damps the sensitivity of the solutions as a side effect (see Theorem 3.5), we plot the optimal solutions $x(p)$ against the robustified solutions $\tilde{x}(p)$ for $p \in [-1, 1]$, see Figure 6.2. As our objective is affine in $p$, the feedback $\tilde{x}(p)$ is affine as well and both solutions are distributed along a line.

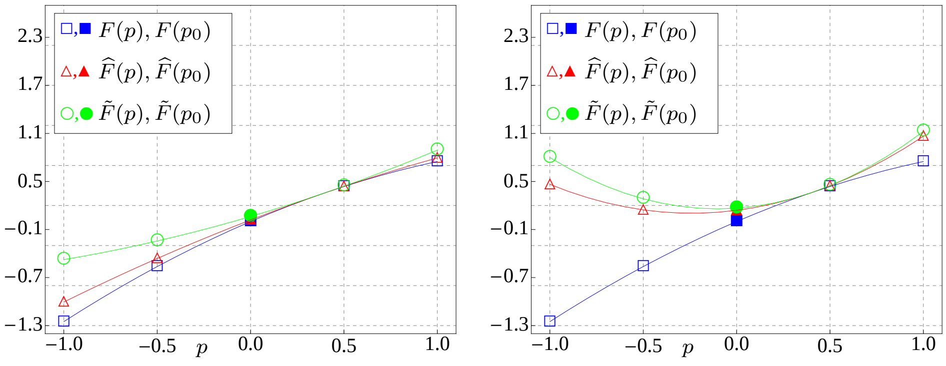
**6.2. A Nonlinear Example.** So far we have only discussed the case that the objective function $f$ is quadratic in the optimization variables with constant Hessian $Q = f_{xx}$. A possible extension of the proposed robustification to general nonlinear problems

$$\min_{x \in \mathbb{R}^n} g(x, p), \tag{NLP($p$)}$$

with $g \in C^2(\mathbb{R}^n \times \mathbb{R}^m)$ is the following. We approximate $g$ near a local optimum $(x_0, p_0)$ satisfying second order sufficient conditions by its second order Taylor polynomial:

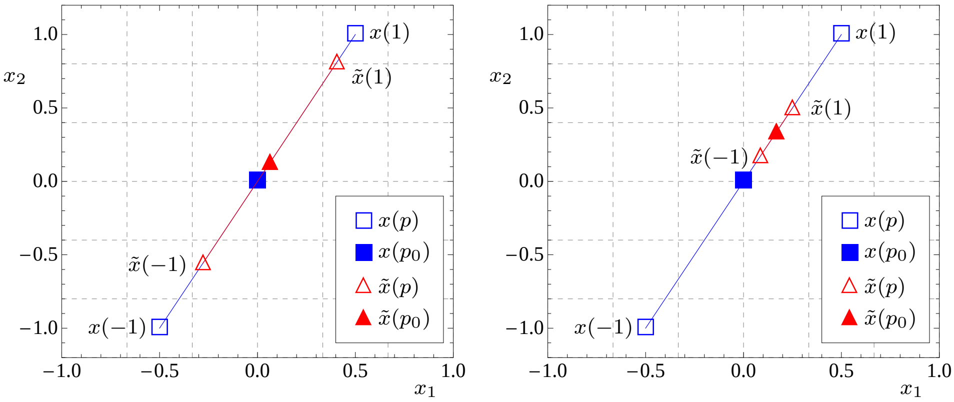
$$f(x, p) = \bar{g} + \begin{pmatrix} \bar{g}_x, & \bar{g}_p \end{pmatrix} \begin{pmatrix} x - x_0 \\ p - p_0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} x - x_0 \\ p - p_0 \end{pmatrix}^\top \begin{pmatrix} \bar{g}_{xx} & \bar{g}_{px} \\ \bar{g}_{xp} & \bar{g}_{pp} \end{pmatrix} \begin{pmatrix} x - x_0 \\ p - p_0 \end{pmatrix},$$

where $\bar{g} = g(x_0, p_0)$, $\bar{g}_x = g_x(x_0, p_0)$, etc. We then apply the robustification to this quadratic function for which all our theoretical results from Sections 2–5 hold. The



(a) $\beta = 0.1850$ (b) $\beta = 2.0489$

FIGURE 6.1. Value functions for two different values of $\beta$ and $p \in [-1,1]$ in the quadratic case.



(a) $\beta = 0.1850$ (b) $\beta = 2.0489$

FIGURE 6.2. Scatter plot of solutions for two different values of $\beta$ and $p \in [-1,1]$ in the quadratic case.

resulting robustified solution $\tilde{x}(p)$ is finally used as an affine feedback to approximate a solution of (NLP($p$)) if $p$ varies in a given neighborhood of the nominal parameter value $p_0$.

Unfortunately, this procedure does not in general result in a robustification in the sense of Theorem 3.2. In fact, we construct an example that shows that using this affine feedback might even increase the sensitivity of the value function in $p_0$. Consider the objective function

$$g(x_1, x_2, p) = x_1^2 + (x_2 - p)^2 - p\,(x_1 - 1) + a\,x_1^3.$$

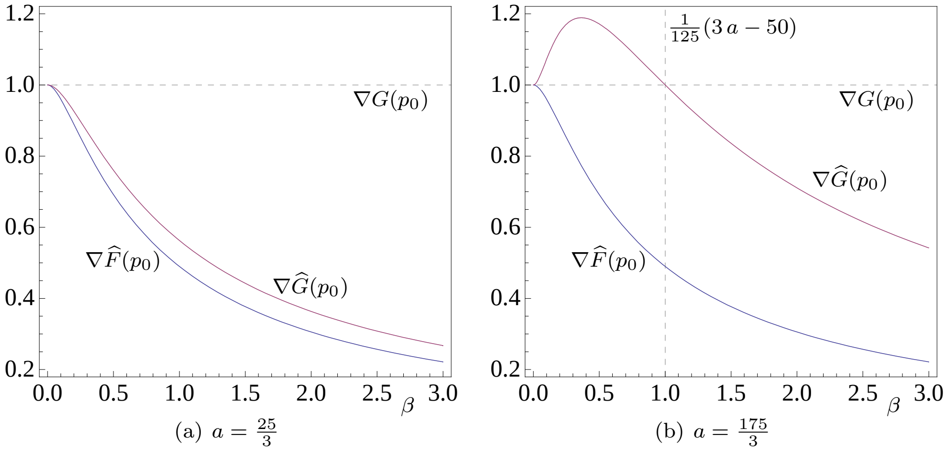
For each value of $a \geq 0$, $x_0 = (0,0)^\top$ is a local minimizer of $g$ for $p_0 = 0$. The second order Taylor polynomial for $g$ at $(x_0, p_0)$ is given by the function $f$ from Section 6.1. Plugging (6.1) and (6.2) into the objective $g$, we obtain, in addition to $F(p)$ and $\widehat{F}(p)$, the two functions $G(p)$ and $\widehat{G}(p)$. At $p_0$, the gradients of all these functions are given by

$$\nabla F(p_0) = 1, \quad \nabla \widehat{F}(p_0) = \frac{4\,(1+5\,\beta)}{(2+5\,\beta)^2},$$
$$\nabla G(p_0) = 1, \quad \nabla \widehat{G}(p_0) = \frac{8+\beta\,\big(60+\beta\,(100+3\,a)\big)}{(2+5\,\beta)^3}.$$

While the first derivative of $\nabla \widehat{G}(p_0)$ with respect to $\beta$ vanishes at $\beta = 0$, its second derivative is given by

$$\frac{d^2}{d\,\beta^2}\nabla \widehat{G}(p_0)\Big|_{\beta=0} = -\frac{75}{2} + \frac{100+3\,a}{4}.$$

It is non-positive for $a \le \frac{50}{3}$. In fact, $\nabla \widehat{G}(p_0)$ is monotonically decreasing with $\beta$ in that case. Conversely, if $a$ is greater than this threshold, $\nabla \widehat{G}(p_0) \ge \nabla G(p_0)$ holds for $\beta \le \frac{1}{125}(3\,a - 50)$. For larger values of $\beta$, we again have $\nabla \widehat{G}(p_0) < \nabla G(p_0)$ and $\nabla \widehat{G}(p_0)$ is monotonically decreasing. These two situations are illustrated in Figure 6.3. We conclude that the results of Theorem 3.2 do not carry over to the case of nonlinear unconstrained optimization. That is, a robustification of the objective value w.r.t. to parameter changes may not be achieved if the modified feedback is based on the quadratic expansion as in this example.



(a) $a = \frac{25}{3}$ (b) $a = \frac{175}{3}$

FIGURE 6.3. Application to nonlinear problems based on their second order Taylor expansion.

### 6.3. A Discretized Optimal Control Problem.

We consider the linear-quadratic regulator (LQR) problem

$$\min_{x,u} J(x,u,p) = \frac{1}{2}\int_0^3 (x(t)-2)^2 + (u(t)-1-\sin(t))^2\,dt \tag{6.3}$$

subject to

$$\dot{x}(t) = u(t) - 1 - \sin(t) \text{ for } t \in [0,3], \tag{6.4a}$$
$$x(0) = p > 0 \tag{6.4b}$$

with a nominal value of $p_0 = 1$. We discretize the problem using piecewise constant controls and eliminate the state equation using the classical fourth-order Runge-Kutta scheme. The integral in the objective is discretized using Simpson's rule. All derivatives are computed using the INTLAB toolbox for MATLAB, see [6, 10]. We obtain an optimal function value of $F(p_0) = 0.5030$.

Table 6.3. Results of applying strategy (S0) for Example 6.3 and various values of $\alpha_0$.

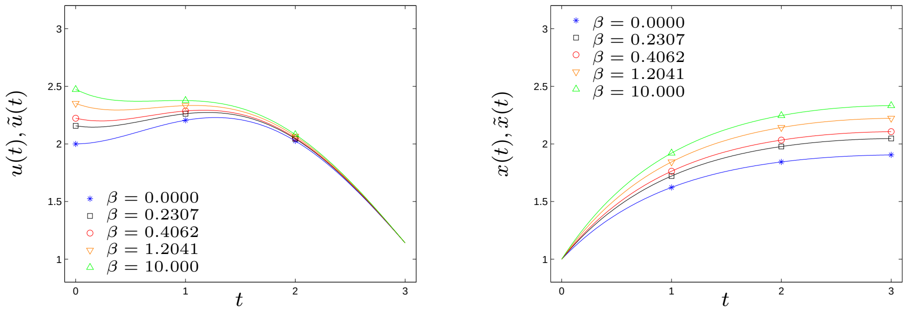
| $\alpha_0$ | $\beta$ | $F(p_0)$ | $\nabla F(p_0)$ | $\widehat{F}(p_0)$ | $\nabla\widehat{F}(p_0)$ | $\tilde{F}(p_0)$ | $\nabla\tilde{F}(p_0)$ |
|---|---|---|---|---|---|---|---|
| 0.9 | 0.2307 | 0.5030 | -1.0005 | 0.5280 | -0.9005 | 0.5820 | -0.6841 |
| 0.7 | 0.6040 | 0.5030 | -1.0005 | 0.5779 | -0.7004 | 0.6398 | -0.4525 |
| 0.5 | 1.2041 | 0.5030 | -1.0005 | 0.6278 | -0.5003 | 0.6795 | -0.2931 |
| 0.3 | 2.5548 | 0.5030 | -1.0005 | 0.6778 | -0.3002 | 0.7119 | -0.1634 |
| 0.1 | 9.2206 | 0.5030 | -1.0005 | 0.7277 | -0.1001 | 0.7399 | -0.0513 |

Table 6.4. Results of applying strategy (S1) for Example 6.3 and various values of $\alpha_1$.

| $\alpha_1$ | $\beta$ | $F(p_0)$ | $\nabla F(p_0)$ | $\widehat{F}(p_0)$ | $\nabla\widehat{F}(p_0)$ | $\tilde{F}(p_0)$ | $\nabla\tilde{F}(p_0)$ |
|---|---|---|---|---|---|---|---|
| $0.1 \cdot 0.5030$ | 0.4062 | 0.5030 | -1.0005 | 0.5533 | -0.7989 | 0.6151 | -0.5514 |
| $0.3 \cdot 0.5030$ | 1.7427 | 0.5030 | -1.0005 | 0.6539 | -0.3957 | 0.6971 | -0.2226 |
| $0.5 \cdot 0.5030$ | 10.000 | 0.5030 | -1.0005 | 0.7295 | -0.0928 | 0.7408 | -0.0475 |

Due to the linear-quadratic structure of (6.3)–(6.4), the discretized problem is of the form (2.1). We apply strategies (S0) and (S1) to find suitable robustification parameters $\beta$. The results are reported in Tables 6.3 and 6.4. In the case of (S1), we obtain from Theorem 5.3 (b) the value of $\overline{\alpha} = 0.2496$. For $\alpha \geq \overline{\alpha}$, any $\beta \geq 0$ is feasible for (S1), and we choose $\beta = 10$ arbitrarily.

We let the uncertain parameter $p$ vary in the interval $[0.8, 1.2]$ and compute the corresponding optimal controls for the original and robustified problems. The controls are shown in Figure 6.4, together with the corresponding optimal states. We infer that the control variables located at the beginning of the time interval are most sensitive with respect to the choice of the robustification parameter $\beta$ in this example. Figure 6.5 shows the associated function values. As for the previous example, these results confirm our analysis.



(a) Controls (b) States

Figure 6.4. Optimal ($u(t)$, $\beta = 0$) and robustified controls ($\tilde{u}(t)$) (left) and corresponding optimal ($x(t)$, $\beta = 0$) and robustified states ($\tilde{x}(t)$) (right) for $p = p_0 = 1$ and different values of $\beta$.

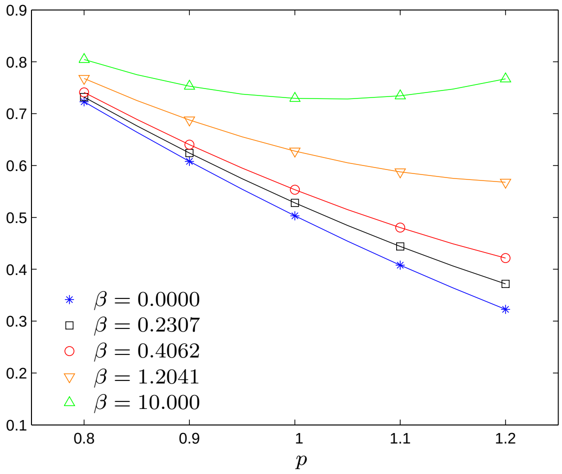


FIGURE 6.5. Plots of $F(p)$ and $\widehat{F}(p)$ for $p \in [0.8, 1.2]$ and different values of $\beta$ for Example 6.3.

**Acknowledgment.** The authors would like to thank two anonymous referees and the associate editor for their valuable suggestions which helped improve the presentation.

## A Proofs

**A.1. Proof of Proposition 2.4.** By definition of the value function $\tilde{F}$, it follows that

$$\tilde{F}(p_0) = f(\tilde{x}(p_0), p_0) + \frac{\beta}{2} \big\| f_p(\tilde{x}(p_0), p_0) \big\|^2.$$

We note that by construction, $f_p \equiv \tilde{f}_p$ holds. Moreover, we have $\tilde{f}_x(\tilde{x}(p_0), p_0) = 0$, and hence

$$\tilde{F}(p_0) = f(\tilde{x}(p_0), p_0) + \frac{\beta}{2} \big\| \tilde{f}_p(\tilde{x}(p_0), p_0) \big\|^2 = f(\tilde{x}(p_0), p_0) + \frac{\beta}{2} \big\| \nabla \tilde{F}(p_0) \big\|^2.$$

From

$$\tilde{x}(p) = -(Q + \beta B_0 B_0^\top)^{-1} \big( b(p) + \beta B_0\, c_p^\top(p_0) \big) \tag{2.3}$$

and

$$(Q + \beta B_0 B_0^\top)^{-1} = Q^{-1} - \beta\, Q^{-1} B_0 (I + \beta \bar{B})^{-1} B_0^\top Q^{-1}, \tag{2.4}$$

we infer that

$$\begin{aligned} \tilde{x}(p) = &- Q^{-1} \big( b(p) + \beta B_0\, c_p^\top(p_0) \big) \\ &+ \beta\, Q^{-1} B_0 (I + \beta \bar{B})^{-1} B_0^\top Q^{-1} \big( b(p) + \beta B_0\, c_p^\top(p_0) \big). \end{aligned} \tag{A.1}$$

Consequently,

$$\begin{aligned}
\nabla\tilde{F}(p_0) = \tilde{f}_p^\top(\tilde{x}(p_0),p_0) &= b_p(p_0)^\top\tilde{x}(p_0) + c_p^\top(p_0) = B_0^\top\tilde{x}(p_0) + c_p^\top(p_0)\\
&= -\,B_0^\top Q^{-1}b(p_0) - \beta\bar{B}c_p^\top(p_0) + \beta\bar{B}(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b(p_0)\\
&\quad + \beta\bar{B}(I+\beta\bar{B})^{-1}\beta\bar{B}c_p^\top(p_0) + c_p^\top(p_0) && \text{by (A.1)}\\
&= \nabla F(p_0) - \beta\bar{B}c_p^\top(p_0) + \beta\bar{B}(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b(p_0)\\
&\quad + \beta\bar{B}(I+\beta\bar{B})^{-1}\beta\bar{B}c_p^\top(p_0)\\
&= \nabla F(p_0) - \beta\bar{B}(I-(I+\beta\bar{B})^{-1}\beta\bar{B})\,c_p^\top(p_0)\\
&\quad + \beta\bar{B}(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b(p_0)\\
&= \nabla F(p_0) - \beta\bar{B}(I+\beta\bar{B})^{-1}\,c_p^\top(p_0) + \beta\bar{B}(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b(p_0) && \text{by (2.5)}\\
&= \nabla F(p_0) - \beta\bar{B}(I+\beta\bar{B})^{-1}\,\nabla F(p_0)\\
&= (I+\beta\bar{B})^{-1}\nabla F(p_0) && \text{by (2.6).}
\end{aligned}$$

For the second derivative, we note that

$$\begin{aligned}
(\nabla^2\tilde{F}(p))_{ij} &= \frac{d}{dp_j}\Big(\frac{d}{dp_i}\tilde{f}(\tilde{x}(p),p)\Big) = \frac{d}{dp_j}\,\Big(\tilde{f}_x(\tilde{x}(p),p)\,\tilde{x}_{p_i}(p) + \tilde{f}_{p_i}(\tilde{x}(p),p)\Big)\\
&= \Big[\tilde{f}_{xx}(\tilde{x}(p),p)\,\tilde{x}_{p_j}(p) + \tilde{f}_{x,p_j}^\top(\tilde{x}(p),p)\Big]^\top\,\tilde{x}_{p_i}(p)\\
&\qquad + \tilde{f}_{p_i,x}(\tilde{x}(p),p)\,\tilde{x}_{p_j}(p) + \tilde{f}_{p_ip_j}(\tilde{x}(p),p).
\end{aligned}$$

A tedious but straightforward calculation shows that

$$\begin{aligned}
&(\nabla^2\tilde{F}(p_0))_{ij}\\
&\quad = \big[(Q+\beta B_0B_0^\top)(Q+\beta B_0B_0^\top)^{-1}b_{p_j}(p_0) - b_{p_j}(p_0)\big]^\top(Q+\beta B_0B_0^\top)^{-1}b_{p_i}(p_0)\\
&\qquad - b_{p_i}(p_0)^\top(Q+\beta B_0B_0^\top)^{-1}b_{p_j}(p_0) + c_{p_ip_j}(p_0)\\
&\qquad - b_{p_ip_j}(p_0)^\top(Q+\beta B_0B_0^\top)^{-1}\left[b(p_0)+\beta B_0c_p^\top(p_0)\right]\\
&\quad = -b_{p_j}(p_0)^\top Q^{-1}b_{p_i}(p_0) - b_{p_ip_j}(p_0)^\top Q^{-1}b(p_0) + c_{p_ip_j}(p_0)\\
&\qquad + b_{p_j}(p_0)^\top\beta\,Q^{-1}B_0(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b_{p_i}(p_0)\\
&\qquad - b_{p_ip_j}(p_0)^\top Q^{-1}\beta B_0c_p^\top(p_0)\\
&\qquad + b_{p_ip_j}(p_0)^\top\beta\,Q^{-1}B_0(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}\left[b(p_0)+\beta B_0c_p^\top(p_0)\right] && \text{by (2.4)}\\
&\quad = \big(\nabla^2F(p_0)\big)_{ij} + b_{p_j}(p_0)^\top\beta\,Q^{-1}B_0(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b_{p_i}(p_0)\\
&\qquad - b_{p_ip_j}(p_0)^\top\beta\,Q^{-1}B_0\left[I-(I+\beta\bar{B})^{-1}\beta\bar{B}\right]c_p^\top(p_0)\\
&\qquad + b_{p_ip_j}(p_0)^\top\beta\,Q^{-1}B_0(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b(p_0)\\
&\quad = \big(\nabla^2F(p_0)\big)_{ij} + b_{p_j}(p_0)^\top\beta\,Q^{-1}B_0(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b_{p_i}(p_0)\\
&\qquad - b_{p_ip_j}(p_0)^\top\beta\,Q^{-1}B_0(I+\beta\bar{B})^{-1}\left[c_p^\top(p_0)-B_0^\top Q^{-1}b(p_0)\right] && \text{by (2.5).}
\end{aligned}$$

A comparison with Proposition 2.2 completes the proof.

**A.2. Proof of Proposition 2.5.** We begin by expanding (A.1) to obtain

$$\begin{aligned}\tilde{x}(p) &= -Q^{-1}b(p) + \beta\, Q^{-1}B_0(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b(p)\\ &\quad - \beta\, Q^{-1}B_0 c_p^\top(p_0) + \beta\, Q^{-1}B_0(I+\beta\bar{B})^{-1}\beta\bar{B}c_p^\top(p_0) && \text{by (2.4)}\\ &= -Q^{-1}b(p) + \beta\, Q^{-1}B_0(I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b(p)\\ &\quad - \beta\, Q^{-1}B_0 c_p^\top(p_0) + \beta\, Q^{-1}B_0 c_p^\top(p_0) - \beta\, Q^{-1}B_0(I+\beta\bar{B})^{-1}c_p^\top(p_0)\\ &&& \text{by (2.5)}\\ &= -Q^{-1}b(p) + \beta\, Q^{-1}B_0(I+\beta\bar{B})^{-1}\big(B_0^\top Q^{-1}b(p) - c_p^\top(p_0)\big).\end{aligned}$$

Plugging (2.3) into the objective $f$ leads to

$$\begin{aligned}\widehat{F}(p) =& c(p) + \left[\frac{1}{2}\tilde{x}(p)^\top Q + b(p)^\top\right]\tilde{x}(p)\\ =& c(p) + \left[-\frac{1}{2}b(p)^\top - \frac{\beta}{2}\big(B_0^\top Q^{-1}b(p) - c_p^\top(p_0)\big)^\top (I+\beta\bar{B})^{-1}B_0^\top\right]\\ &\cdot \left[Q^{-1}b(p) - \beta\, Q^{-1}B_0(I+\beta\bar{B})^{-1}\big(B_0^\top Q^{-1}b(p) - c_p^\top(p_0)\big)\right]\\ =& F(p) + \frac{\beta}{2}b(p)^\top Q^{-1}B_0(I+\beta\bar{B})^{-1}\big(B_0^\top Q^{-1}b(p) - c_p^\top(p_0)\big)\\ &- \frac{\beta}{2}\big(B_0^\top Q^{-1}b(p) - c_p^\top(p_0)\big)^\top (I+\beta\bar{B})^{-1}B_0^\top Q^{-1}b(p)\\ &+ \frac{\beta}{2}\big(B_0^\top Q^{-1}b(p) - c_p^\top(p_0)\big)^\top \mathcal{I}\,\big(B_0^\top Q^{-1}b(p) - c_p^\top(p_0)\big).\end{aligned}$$

Evaluating this expression at $p = p_0$ shows the first relation in Proposition 2.5. The computation of the gradient is now straightforward:

$$\nabla\widehat{F}(p) = \nabla F(p) + \beta B_0^\top Q^{-1}b_p(p)\,\mathcal{I}\,\big(B_0^\top Q^{-1}b(p) - c_p(p_0)\big),$$

which implies $\nabla\widehat{F}(p_0) = \nabla F(p_0) - \beta\bar{B}\,\mathcal{I}\,\nabla F(p_0)$. It follows from Lemma 2.3 with $M = \beta\bar{B}$ that

$$I - \beta\bar{B}\mathcal{I} = I - \beta\bar{B}(I+\beta\bar{B})^{-1}\beta\bar{B}(I+\beta\bar{B})^{-1} = (I+2\beta\bar{B})(I+\beta\bar{B})^{-2}$$

holds, which proves the second claim. The computation of the entries of the Hessian matrix is similar to the proof of Proposition 2.4 and is omitted.

## References

[1] A. Ben-Tal and A. Nemirovski. Robust optimization - methodology and applications. *Mathematical Programming*, 92(3, Ser. B):453–480, 2002.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, Cambridge, 2004.

[3] G. Calafiore and F. Dabbene. Near optimal solutions to least-squares problems with stochastic uncertainty. *Systems & Control Letters*, 54:1219–1232, 2005.

[4] M. Diehl, H. G. Bock, and E. Kostina. An approximation technique for robust nonlinear optimization. *Mathematical Programming*, 107(1–2, Ser. B):213–230, 2006.

[5] A. Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming.* Academic Press, New York, 1983.

[6] A. Griewank. *Evaluating Derivatives.* SIAM, Philadelphia, 2000.

[7] I. Kanellakopoulos. Robustification tools for nonlinear control design. In *Proceedings of the 33rd Conference on Decision and Control*, volume 4, pages 3464–3468. IEEE, 1994.

[8] D. E. Quevedo and M. E. Salgado. Robustification of model predictive control. In *Proceedings of the 39th Conference on Decision and Control*, volume 1, pages 178–180. IEEE, 2000.

[9] P. Rodriguez and D. Dumur. Generalized predictive control robustification under frequency and time-domain constraints. *IEEE Transactions on Control Systems Technology*, 13(4):577–587, 2005.

[10] S. Rump. INTLAB - INTerval LABoratory. In T. Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999. http://www.ti3.tu-harburg.de/rump/.

Chemnitz University of Technology, Faculty of Mathematics, D–09107 Chemnitz, Germany

*E-mail address*: `martin.bernauer@mathematik.tu-chemnitz.de`

*URL*: `http://www.tu-chemnitz.de/~bernm`

Chemnitz University of Technology, Faculty of Mathematics, D–09107 Chemnitz, Germany

*E-mail address*: `roland.griesse@mathematik.tu-chemnitz.de`

*URL*: `http://www.tu-chemnitz.de/~griesse`