A note on sensitivity analysis in nonlinear programming

François Pacaud

March 30, 2023

In this note, we are interested in the sensitivity analysis of optimization problem. We aim at summarizing results old and new, inside an unified framework.

1. Introduction

Sensitivity analysis is a long standing research thread in nonlinear programming. Its main motivation is to evaluate the sensitivity of the solution of a nonlinear problem with respect to problem changes. The sensitivity is here to be interpreted as the derivatives of the optimal value function and the primal-dual Karush-Kuhn-Tucker (KKT) solution with relation to parameter changes. If the problem is regular enough, the sensitivities can be evaluated at a KKT solution by applying the implicit function theorem on the KKT conditions of the problem. In nonlinear programming, the regularity of a given problem is related to its non-degeneracy: if we perturb slightly the parameters, can we guarantee that the primal solution exists and is unique? Depending on the structure of the problem, the perturbed solution may have the following demonstrable properties, as illustrated by Fiacco [FL93] by order of difficulty:

- Existence of the primal solution.
- Existence and continuity of the primal solution.
- (Upper) Lipschitz continuity of the primal solution, nonuniqueness of the primal-dual solution.
- (Upper) Lipschitz continuity of the primal-dual solution, nonuniqueness of the primal-dual solution.
- Directional differentiability and uniqueness of the primal solution, nonuniqueness of the dual solution.

In this note, we aim at summarizing the conditions under which the sensitivity analysis is well-posed. We start by a brief review of literature to remind of the main developments in the theory of sensitivity analysis for nonlinear programming.

The 1970s: ensuring regularity and differentiability. The regularity of the KKT solution has been studied extensively in the 1970s by Robinson [Rob80, Rob82], by interpreting the nonlinear problem as a generalized equation. It is now established that the regularity of a nonlinear problem usually follows from the constraint qualification conditions that hold at the KKT solution. Robinson proved in [Rob82] that the primal-dual solution is *strongly regular* (hence, locally unique and Lipschitz continuous) if LICQ and SSOSC holds at the original solution. The (Fréchet) differentiability of the primal-dual solution is established if we suppose that in addition SCS hold, as established by Fiacco in his seminal result [Fia76].

The 1980s and 1990s: tackling degeneracy. The previous results have been extended in the 1980s, with a primary focus on dropping the SCS conditions required in [Fia76]. Kojima [Koj80] proved that MFCQ and GSSOSC are sufficient to guarantee that the primal-dual solution is strongly stable. However, the primal-dual solution is no longer differentiable as soon as we drop the SCS condition. The idea is then to look at conditions under which the solution is *directionally differentiable* in a given direction in the parameter space. Jittorntrum [Jit84] proved that if LICQ and SSOSC hold, then the directional derivative of the primal-dual solution exists and is given as the solution of a quadratic problem (QP). One can go one step further by dropping LICQ, at the price of losing the directional differentiability of the dual solution. Shapiro [Sha85] showed that on its end the primal solution remains directionally differentiable if we suppose that SMCFQ and SSOSC hold. This result was extended by Kyparisis in [Kyp90], who proved that the primal solution is directionally differentiable if we assume that MFCQ and CRCQ and GSSOSC hold, that is, without unicity of the dual solution. This thread of research culminated in 1995 with the publication of the seminal result of Ralph and Dempe [RD95], giving a practical method to evaluate the directional derivative of the primal solution under MFCQ.

The 2000s and 2010s: extending the scope of applications. The theory underlying sensitivity analysis for nonlinear program has witnessed a significant consolidation after the 1990s, with the publications of three books summarizing the different approaches to the problem [BS00, FP03, DR09]. Notably, the domain of sensitivity analysis has been extended to the more general setting offered by variational inequalities, which covers nonlinear programs and complementarity problems. In parallel, numerous applications have flourished out of the sensitivity analysis theory, one of the most significant being bilevel programming [Dem02]. Furthermore, sensitivity analysis has proven to be a fertile ground for the analysis of stochastic optimization problems [Sha90, Sha91], parametric decomposition [Gau78, DN08], parameter estimation problems [LNB12] and last but not least, model predictive control [ZB09, JYB14, ZA10]. Recently, the field has witnessed a renewed interest with the emergence of the differentiable programming paradigm in the machine learning community. The idea there is to embed a (convex) optimization solver inside a neural network, whose training requires the evaluation of the sensitivity in the backward pass. In this emerging field, most recent results have focused on the convex programming case [AK17, AAB$^+$19].

2. Framework

The scope of this note is limited to finite-dimensional optimization problems. We refer to [BS00] for a broader description covering the finite and the infinite cases.

2.1. Nonlinear problem

For a real-valued parameter $p \in \mathbb{R}^{n_p}$, we are interested in solving the parametric nonlinear optimization problem:

$$\mathrm{NLP}(p) := \min_{x \in \mathbb{R}^{n_x}} f(x,p) \quad \text{subject to} \quad \begin{cases} g(x,p) = 0 \,, \\ h(x,p) \leq 0 \,, \end{cases} \tag{1}$$

where $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}, g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{m_e}, h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{m_i}$ are C^2 functions depending jointly on the optimization variable $x \in \mathbb{R}^{n_x}$ and a parameter $p \in \mathbb{R}^{n_p}$. The admissible set of (1) is

$$X(p) = \{x \in \mathbb{R}^{n_x} \mid g(x,p) = 0 \,,\, h(x,p) \leq 0\} \,. \tag{2}$$

Value function. We define the value function $\varphi : \mathbb{R}^{n_p} \to \mathbb{R} \cup \{+\infty\}$ as the solution of $\mathrm{NLP}(p)$ for a given parameter $p \in \mathbb{R}^{n_p}$:

$$\varphi(p) = \min_{x \in \mathbb{R}^{n_x}} f(x,p) \quad \text{subject to} \quad x \in X(p) \,. \tag{3}$$

By convention $\varphi(p) = +\infty$ if the problem is infeasible ($X(p) = \varnothing$). The solution set associated to (3) is noted

$$\Sigma(p) := \{x \in X(p) \;:\; f(x,p) = \varphi(p)\} \,. \tag{4}$$

Lagrangian. We introduce the Lagrangian associated to (1) as

$$\mathcal{L}(x,y,z,p) = f(x,p) + y^\top g(x,p) + z^\top h(x,p) \,, \tag{5}$$

where $y \in \mathbb{R}^{m_e}$ (resp. $z \in \mathbb{R}^{m_i}$) is the multiplier associated to the equality (resp. inequality) constraints. The Karush-Kuhn-Tucker (KKT) conditions of (1) writes out

$$\begin{aligned} &\nabla_x f(x,p) + \sum_{i=1}^{m_e} y_i \nabla_x g_i(x,p) + \sum_{i=1}^{m_i} z_i \nabla_x h_i(x,p) = 0 \,, \\ &g(x,p) = 0 \,, \quad h(x,p) \leq 0 \,, \quad z \geq 0 \,, \quad z^\top h(x,p) = 0 \,. \end{aligned} \tag{6}$$

We define a primal-dual stationary solution (or KKT stationary solution)

$$w(p) = \big(x(p), y(p), z(p)\big) \in \mathbb{R}^{n_x} \times \mathbb{R}^{m_e} \times \mathbb{R}^{m_i} \,, \tag{7}$$

any point satisfying the KKT conditions (6). To alleviate the notations, we often denote by $w^\star$ a given primal-dual stationary solution. The problem (1) is not assumed to be convex, hence a KKT solution is not necessary an optimal solution of (1).

We note

$H = [H_x \; H_p] = \partial_{(x,p)} h(x,p) \in \mathbb{R}^{m_i \times (n_x + n_p)}$	Jacobian of the inequality cons.
$G = [G_x \; G_p] = \partial_{(x,p)} g(x,p) \in \mathbb{R}^{m_e \times (n_x + n_p)}$	Jacobian of the equality cons.
$\nabla^2_{xx}\mathcal{L} = \nabla^2_{xx}\mathcal{L}(x,y,z,p)$	Hessian of Lagrangian w.r.t x.

Multiplier set. The set of multipliers $(y, z) \in \mathbb{R}^{m_e+m_i}$ satisfying the KKT conditions (6) is denoted by $\mathcal{M}_p(x)$. We notice that the set $\mathcal{M}_p(x)$ is polyhedral (hence closed and convex). The set $\mathcal{M}_p(x)$ being polyhedral, we note by $\mathcal{E}_p(x)$ the set of extreme points. We say that the problem is *degenerate* if $\mathcal{M}_p(x)$ has more than one element.

Active set. For a given $x \in \mathbb{R}^{n_x}$, we define the *active set* at x as

$$\mathcal{I}_p(x) = \{i \in [m_i] \mid h_i(x,p) = 0\} . \tag{8}$$

For given multipliers $(y, z) \in \mathcal{M}_p(x)$, we define the *strongly* and *weakly active constraints* as

$$\mathcal{I}_p^+(x;z) = \{i \in \mathcal{I}_p(x) \mid z_i > 0\} , \quad \mathcal{I}_p^0(x;z) = \{i \in \mathcal{I}_p(x) \mid z_i = 0\} . \tag{9}$$

By definition, $\mathcal{I}_p(x) = \mathcal{I}_p^+(x;z) \cup \mathcal{I}_p^0(x;z)$.

Degeneracy arises from the set of the undetermined weakly active constraints $\mathcal{I}_z^0(x)$: a small violation is likely to render them inactive.

Critical cone. The critical cone is a key notion to characterize the second-order optimality conditions of a nonlinear program.

Definition 2.1 (Critical cone)**.** *Let $x \in X(p)$. The critical cone of the feasible set $X(p)$ at x is defined as the set*

$$\begin{aligned} \mathcal{C}_p(x) = \{d \in \mathbb{R}^{n_x} \ : \ &\nabla_x f(x,p)^\top d \leq 0 , \\ &\nabla_x g_i(x,p)^\top d = 0 , \forall i \in [m_e] , \\ &\nabla_x h_j(x,p)^\top d \leq 0 , \forall j \in \mathcal{I}_p(x)\} . \end{aligned} \tag{10}$$

We note that at a KKT solution $(x^\star, y^\star, z^\star)$, the critical cone satisfies

$$\begin{aligned} \mathcal{C}_p(x^\star) = \{d \in \mathbb{R}^{n_x} \ : \ &\nabla_x g_i(x^\star,p)^\top d = 0 , \forall i \in [m_e] , \\ &\nabla_x h_j(x^\star,p)^\top d = 0 , \forall j \in \mathcal{I}_p^+(x^\star;z^\star) , \\ &\nabla_x h_j(x^\star,p)^\top d \leq 0 , \forall j \in \mathcal{I}_p^0(x^\star;z^\star)\} . \end{aligned} \tag{11}$$

Under strict complementarity condition, the critical cone is the null space of the active constraints' Jacobian: $\mathcal{C}_p(x^\star) = N(J_{act}(x^\star)$, as in that case the critical cone reduces to

$$\begin{aligned} \mathcal{C}_p(x^\star) = \{d \in \mathbb{R}^{n_x} \ : \ &\nabla_x g_i(x^\star,p)^\top d = 0 , \forall i \in [m_e] , \\ &\nabla_x h_j(x^\star,p)^\top d = 0 , \forall j \in \mathcal{I}(x^\star)\} . \end{aligned} \tag{12}$$

In parametric optimization, it is common to introduce the critical set associated to the critical cone at the solution [FP03, RD95].

Definition 2.2 (Critical set)**.** *Let $(x^\star, y^\star, z^\star)$ be a KKT stationary solution. The* critical set *at the solution is the polyhedral cone defined as*

$$\begin{aligned} \mathcal{K}_p(x^\star, z^\star) = \{w \in \mathbb{R}^{n_x+n_p} \ : \ &\nabla_{(x,p)} g_i(x^\star,p)^\top w = 0 , \forall i \in [m_e] , \\ &\nabla_{(x,p)} h_j(x^\star,p)^\top w = 0 , \forall j \in \mathcal{I}_p^+(x^\star;z^\star) , \\ &\nabla_{(x,p)} h_j(x^\star,p)^\top w \leq 0 , \forall j \in \mathcal{I}_z^0(x^\star;z^\star)\} . \end{aligned} \tag{13}$$

Elements of $\mathcal{K}_p(x^\star, z^\star)$ are called critical directions*. The* directional critical set *in the direction $h \in \mathbb{R}^{n_p}$ is defined as*

$$\mathcal{K}_p(x^\star, z^\star; h) = \{d \in \mathbb{R}^{n_x} \ : \ (d,h) \in \mathcal{K}_p(x^\star, z^\star)\} . \tag{14}$$

2.2. Constraints qualification

Constraint qualifications give conditions under which a linear approximation of the feasible set near an optimal solution $x^\star$ is capturing the geometry of the feasible set. As the regularity of a nonlinear program is given by the constraint qualification that holds at a stationary point, we recall in this section the main constraint qualifications we can encounter in practice, and their impacts on the topology of the problem.

Definition 2.3 (Linear-independence constraint qualification (LICQ)). *LICQ holds if the gradients of active constraints*

$$\{\nabla_x g_i(x,p) \;:\; i \in [m_e]\} \;\cup\; \{\nabla_x h_i(x,p) \;:\; i \in \mathcal{I}(x)\} \tag{15}$$

are linearly independent.

Theorem 2.4. *Suppose LICQ is satisfied. Then*

$$d^\top \nabla^2_{xx} L(x,y,z,p) d \geq 0 \quad \forall d \in \mathcal{C}_p(x) \;, \tag{16}$$

where $\mathcal{C}_p(x)$ is the critical cone of the feasible set $X(p)$ at x.

Definition 2.5 (Mangasarian-Fromovitz constraint qualification (MFCQ)). *MFCQ holds if the gradients of the equality constraints*

$$\{\nabla_x g_i(x,p) \;:\; i \in [m_e]\} \;, \tag{17}$$

are linearly independent and if there exists a vector $d \in \mathbb{R}^{n_x}$ such that

$$\begin{aligned} &\nabla_x g_i(x,p)^\top d = 0 \;, \quad \forall i \in [m_e] \;, \\ &\nabla_x h_i(x,p)^\top d < 0 \;, \quad \forall i \in \mathcal{I}_p(x) \;. \end{aligned} \tag{18}$$

It is well known that MFCQ is a generalization of the Slater's condition in the nonlinear case.

Theorem 2.6 ([Gau77]). *The set $\mathcal{M}(x,p)$ is bounded if and only if MFCQ holds.*

Proposition 2.7 (Dual MFCQ). *Using the theorem of alternatives, MFCQ is equivalent to 0 being the only solution of the linear system*

$$\begin{aligned} &\sum_{i=1}^{m_e} y_i \nabla_x g_i(x,p) + \sum_{i \in \mathcal{I}(x)} z_i \nabla_x h_i(x,p) = 0 \;, \\ &z_i \geq 0 \quad \forall i \in \mathcal{I}(x) \;. \end{aligned} \tag{19}$$

Definition 2.8 (Strict Mangasarian-Fromovitz constraint qualification (SMFCQ)). *Let $w = (x,y,z)$ be a KKT stationary point. SMFCQ holds at w if*

$$\{\nabla_x g_i(x,p) \;:\; i \in [m_e]\} \;\cup\; \{\nabla_x h_i(x,p) \;:\; i \in \mathcal{I}_p^+(x;z)\} \;, \tag{20}$$

are linearly independent and if there exists a vector $d \in \mathbb{R}^{n_x}$ such that

$$\begin{aligned} &\nabla_x g_i(x,p)^\top d = 0 \;, \forall i \in [m_e] \;, \\ &\nabla_x h_i(x,p)^\top d = 0 \;, \forall i \in \mathcal{I}_p^+(x;z) \;, \\ &\nabla_x h_i(x,p)^\top d < 0 \;, \forall i \in \mathcal{I}_p^0(x;z) \;. \end{aligned} \tag{21}$$

Proposition 2.9 (Theorem 2.1, [Kyp85]). *Let $w = (x, y, z)$ be a primal-dual solution of* (1)*, such that w satisfies SMFCQ. Then*

$$d^\top \nabla^2_{xx} L(x, y, z, p) d \geq 0 \quad \forall d \in \mathcal{C}_p(x) \ . \tag{22}$$

Proposition 2.10 (Proposition 1.1, [Kyp85]). *Let $w = (x, y, z)$ be a primal-dual solution. Then SMFCQ holds at x if and only if the multiplier vector (y, z) is unique.*

Hence, SMFCQ ensures that the multiplier set $\mathcal{M}_p(x)$ reduces to a singleton.

Definition 2.11 (Constant-rank constraint qualification (CRCQ)). *CRCQ holds if there exists a neighborhood W of $x^\star$ such that for any subsets I of $\mathcal{I}_p(x^\star)$ and J of $[m_e]$ the family of gradient vectors*

$$\{\nabla_x g_j(x, p) \ : \ j \in J\} \ \cup \ \{\nabla_x h_i(x, p) \ : \ i \in I\} \ , \tag{23}$$

has the same rank for all vectors $x \in W$.

Definition 2.12 (Second-order sufficiency condition (SOSC)). *SOSC holds if there exists $(y, z) \in \mathcal{M}_p(x)$ such that*

$$d^\top \nabla^2_{xx} L(x, y, z, p) d > 0 \quad \forall d \in \mathcal{C}_p(x) \ , \ d \neq 0 \ . \tag{24}$$

Proposition 2.13. *Suppose SOSC holds at a solution $x^\star$. Then, there exists $\sigma > 0$ and a neighborhood V of $x^\star$ such that for all $x \in X(p) \cap V$ such that for $x \neq x^\star$,*

$$f(x) > f(x^\star) + \frac{\sigma}{2} \|x - x^\star\|^2 \ . \tag{25}$$

Definition 2.14 (Strong second-order sufficiency condition (SSOSC)). *SSOSC holds if there exists $(y, z) \in \mathcal{M}_p(x)$ such that*

$$d^\top \nabla^2_{xx} L(x, y, z, p) d > 0 \quad \forall d \in \mathcal{D}_p(x; z) \ , d \neq 0 \ , \tag{26}$$

with

$$\begin{aligned} \mathcal{D}_p(x; z) = \{d \in \mathbb{R}^{n_x} \ : \nabla_x g_i(x, p) d = 0 \quad & \forall i \in [m_e] \ , \\ \nabla_x h_j(x, p) d = 0 \quad & \forall j \in \mathcal{I}^+_p(x; z)\} \ . \end{aligned} \tag{27}$$

Note that at a solution x, $\mathcal{C}_p(x; z) \subset \mathcal{D}_p(x; z)$ with both sets being equal if SCS holds. In other words, the strong second-order sufficiency condition and the second-order sufficiency condition are equivalent if SCS holds.

3. Sensitivity of the primal-dual solution

Let $p^\star \in \mathbb{R}^{n_p}$ be a parameter, and $w^\star = (x^\star, y^\star, z^\star)$ a KKT stationary solution of NLP(p). We aim at studying the behavior of the primal-dual solution $w(p) := [x(p), y(p), z(p)]$ near p. Notably we are interested in the local continuity of the solution $w(p)$ (i.e., *is $w(p)$ Lipschitz-continuous?*) as well as its local differentiability (i.e., *under which conditions $w(p)$ is differentiable?*).

3.1. Fiacco and the implicit function theorem

If the problem is regular, the local continuity is usually given by the implicit function theorem, the core result lying behind almost all results in sensitivity analysis.

Theorem 3.1 (Implicit function theorem). *Let $F : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$ be a C^1 function and $(x^\star, p^\star) \in \mathbb{R}^{n_x} \times \mathbb{R}^{n_p}$ such that $F(x^\star, p^\star) = 0$. If $\nabla_x F(x^\star, p^\star)$ is* invertible*, then there exists a neighborhood $U \subset \mathbb{R}^{n_p}$ of $p^\star$ and a differentiable function $x(p)$ such that $F(x(p), p) = 0$ for all $p \in U$, and $x(p)$ is the unique solution in a neighborhood of $x^\star$. In addition, $x(\cdot)$ is C^1 differentiable on U, with*

$$\nabla_p x(p) = -\big(\nabla_x F(x,p)\big)^{-1} \nabla_p F(x,p) \; . \tag{28}$$

In [Fia76], Fiacco investigated under which the KKT equations are regular, in order to apply the implicit function theorem afterward. To do so, we should ensure (i) the primal solution mapping $x(p)$ is single-valued (i.e. the problem (1) has an unique solution) (ii) the dual solution $(y(p), z(p))$ is unique and (iii) the active set is locally stable. Fiacco showed that these three conditions are satisfied if resp. (i) SOSC (ii) LICQ and (iii) strict complementarity slackness hold. In that particular case, the KKT equations (6) rewrites as a smooth system of nonlinear equations depending on $w = (x, y, z)$,

$$\nabla_w \mathcal{L}(w, p) = 0 \; , \tag{29}$$

a setting favorable to apply the implicit function theorem to find the sensitivity $\nabla_p w(p)$. Applying Theorem 3.1 on (29) requires the evaluation of the Hessian of the Lagrangian

$$\nabla^2_{ww}\mathcal{L} = \begin{bmatrix} \nabla^2_{xx}\mathcal{L} & G_x^\top & H_x^\top \\ G_x & 0 & 0 \\ H_x & 0 & 0 \end{bmatrix} \; , \tag{30}$$

often associated to the KKT system of the original problem.

Theorem 3.2 ([Fia76]). *Suppose that at a KKT stationary solution $x^\star$ of NLP($p^\star$), SOSC, LICQ and strict complementarity slackness hold. Then,*

1. *the primal-dual solution $w^\star = (x^\star, y^\star, z^\star)$ is unique;*
2. *there exists a unique continuously differentiable function $w(\cdot)$ defined in a neighborhood U of $p^\star$ such that $w(p^\star) = (x^\star, y^\star, z^\star)$, with*
$$\nabla_p w(p) = -(\nabla^2_{ww}\mathcal{L})^{-1} \, \nabla^2_{wp}\mathcal{L} \; ; \tag{31}$$
3. *both strict complementarity slackness and LICQ hold locally near $p^\star$.*

In detail, equation (31) gives the sensitivity of the primal-dual solution as

$$\begin{bmatrix} \nabla_p x \\ \nabla_p y \\ \nabla_p z \end{bmatrix} = - \begin{bmatrix} \nabla^2_{xx}\mathcal{L} & G_x^\top & H_x^\top \\ G_x & 0 & 0 \\ H_x & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} \nabla^2_{xp}\mathcal{L} \\ G_p \\ H_p \end{bmatrix} \; . \tag{32}$$

The operation translates as the solution of a (sparse) linear system with p right-hand-side. If the original problem NLP(p) is solved with a Newton algorithm, the factorization of the

KKT matrix $\nabla^2_{ww}\mathcal{L}$ is already available at the solution and only the backsolves have to be computed, saving a significant amount of computation.

The strict complementarity condition is unlikely to hold on practical engineering problems, where constraint can change locally from active to inactive. For that reason, it has been investigated how to generalize Fiacco's result without the strict complementarity assumption.

3.2. Robinson and generalized equations

In [Rob79], Robinson proposed a powerful framework to characterize the solution of nonlinear program (1), by interpreting the KKT conditions (6) as a *generalized equation* $F(w,p)+N_C(w) \ni 0$ (see Appendix C for a brief recall on generalized equations).

Proposition 3.3 (KKT system as generalized equation). *Let $w = (x, y, z) \in \mathbb{R}^{n_x} \times \mathbb{R}^{m_e} \times \mathbb{R}^{m_i}$ be a primal-dual variable. The KKT system (6) is equivalent to the generalized equation*

$$F(w,p) + N_C(w) \ni 0 \ , \tag{33}$$

with $F(w,p) := \begin{bmatrix} \nabla_x \mathcal{L}(x,y,z,p) \\ -\nabla_y \mathcal{L}(x,y,z,p) \\ -\nabla_z \mathcal{L}(x,y,z,p) \end{bmatrix}$ *and* $C = \mathbb{R}^{n_x} \times \mathbb{R}^{m_e} \times \mathbb{R}^{m_i}_{+}$.

When writing the KKT system as a generalized equation, the nonsmoothness arising from the active set changes is encoded inside the normal cone $N_C(x)$, F being a smooth functional. If the generalized equation (33) is strongly regular, then the sensitivities follow from Robinson's Theorem (Theorem C.4 in the appendix). Strong regularity is given if we suppose SSOSC and LICQ hold locally, allowing to drop SCS. Here, SSOSC is necessary to guarantee the unicity perturbed solution $x(p)$ remains without SCS.

Proposition 3.4 (Theorem 4.1, [Rob80]). *Let $w^\star = (x^\star, y^\star, z^\star)$ a primal-dual solution of the generalized equation (33). If SSOSC and LICQ hold at $w^\star$, then (33) is strongly regular at $w^\star$.*

The strong regularity given by Proposition 3.4 implies

(a) The solution $x^\star$ is an isolated local minimum of NLP(p) and the associated Lagrange multipliers $(y^\star, z^\star)$ are unique;

(b) For p near $p^\star$, there exists a unique Lipschitz continuous function $w(p) = (x(p), y(p), z(p))$ satisfying KKT and SOSC for NLP(p), with $w(p^\star) = w^\star$, and $x(p)$ being a unique solution of NLP(p);

(c) LICQ holds at $x(p)$ for p near $p^\star$.

Hence, strict complementarity slackness is not required to ensure the primal-dual solution $w(p)$ is locally Lipschitz continuous. In addition, one can go one step further and prove that the optimal solution is also directionally differentiable in a given direction $h \in \mathbb{R}^{n_p}$.

Proposition 3.5 (Theorem 2, [Jit84]). *Suppose that at a KKT stationary solution $x^\star$ of NLP($p^\star$), LICQ and SSOSC hold. Then*

- *there exists a local unique continuous primal-dual solution $w(\cdot)$, and is directionally differentiable in every direction $h \in \mathbb{R}^{n_p}$;*

- *for a given direction $h \in \mathbb{R}^{n_p}$, the directional derivative $w'(p;h)$ is the primal-dual solution of the QP problem*

$$QP_{y,z}(p;h) := \begin{cases} \min_d \frac{1}{2} d^\top \nabla^2_{xx}\mathcal{L}\, d + h^\top \nabla^2_{xp}\mathcal{L}\, d \\ s.t. \quad \nabla_x g_i(x,p)^\top d + \nabla_p g_i(x,p)^\top h = 0 & (\forall i \in [m_e]) \,, \\ \qquad h_j(x,p) + \nabla_x h_j(x,p)^\top d + \nabla_p h_j(x,p)^\top h \leq 0 & (\forall j \in [m_i]) \,, \end{cases} \tag{34}$$

with $\nabla^2_{xx}\mathcal{L} = \nabla^2_{xx}\mathcal{L}(x,y,z,p)$, $\nabla^2_{xp}\mathcal{L} = \nabla^2_{xp}\mathcal{L}(x,y,z,p)$.

The QP is parameterized implicitly by the dual multipliers (y,z) (appearing in the derivatives of the Lagrangian). LICQ guarantees that the multipliers are unique at the solution (leading to a non-ambiguous definition of the directional derivative (34)). Under LICQ and SSOSC, the directional derivative exists for every direction $h \in \mathbb{R}^{n_p}$. However, we cannot guarantee that the function $w'(p;\cdot)$ is continuous without assuming SCS.

3.3. Towards degeneracy: dropping the LICQ condition

The question now becomes: can we evaluate the sensitivities when the multipliers (y,z) are non unique? It turns out that the directional derivative of the primal solution $x(p)$ exists in that degenerate case, but special care should be given as the unicity of the multipliers is not guaranteed. Hence, the definition of QP (34) becomes ambiguous: which multipliers should we consider when evaluating the Hessian of the Lagrangian?

Kojima was the first to obtain the continuity of the optimal solution assuming only MFCQ and GSSOSC, and proved in addition that MFCQ and GSSOSC are the weakest conditions under which the perturbed solution is locally unique [Koj80].

Theorem 3.6 (Theorem 7.2, [Koj80])**.** *Suppose that at a KKT stationary solution $x^\star$ of NLP($p^\star$), MFCQ and GSSOSC hold. Then, there are open neighborhoods U of $p^\star$ and V of $x^\star$ and a function $x : U \to V$ such that $x(\cdot)$ is continuous. The function $x(p)$ is the unique local solution of NLP(p) in V, and MFCQ holds at $x(p)$.*

The LICQ condition in Proposition 3.5 can be relaxed to SMFCQ, which also implies the multipliers $(y^\star, z^\star)$ at the solution are unique (Proposition 2.10).

Proposition 3.7 (Theorem 4.2, [Sha85])**.** *Suppose that at a solution $x^\star$ of NLP($p^\star$), SMFCQ and SSOSC hold. Then, $x(\cdot)$ is directionally differentiable at p for every $h \in \mathbb{R}^{n_p}$, and the directional derivative $x'(p^\star, h)$ is the unique solution of the QP problem*

$$\min_d \frac{1}{2} d^\top \nabla^2_{xx}\mathcal{L}\, d + h^\top \nabla^2_{xp}\mathcal{L}\, d \quad s.t. \quad d \in \mathcal{K}_{p^\star}(x^\star, z^\star; h) \,, \tag{35}$$

for $\nabla^2_{xx}\mathcal{L} = \nabla^2_{xx}\mathcal{L}(x^\star, y^\star, z^\star, p^\star)$, $\nabla^2_{xp}\mathcal{L} = \nabla^2_{xp}\mathcal{L}(x^\star, y^\star, z^\star, p^\star)$.

Note that we are no longer able to evaluate the directional derivative of the dual solution in (35), on the contrary to (34). Kyparisis proved that we can relax SMFCQ by CRCQ to get the existence of the directional derivative $x'(p,h)$ in the degenerate case where the multipliers (y,z) are non-uniques: it suffices in that case to look at the multipliers in the extreme points $\mathcal{E}_{p^\star}(x^\star)$.

Proposition 3.8 (Theorem 2.2, [Kyp90]). *Suppose that at a KKT stationary solution $x^\star$ of NLP($p^\star$), MFCQ, CRCQ and GSSOSC hold. Then, then $x(\cdot)$ is directionally differentiable at $p^\star$ for every $h \in \mathbb{R}^{n_p}$, and the directional derivative $x'(p^\star, h)$ is the unique solution of the QP problem, for a given $(y, z) \in \mathcal{E}_{p^\star}(x^\star)$,*

$$\min_d \; \frac{1}{2} d^\top \nabla^2_{xx}\mathcal{L}\, d + h^\top \nabla^2_{xp}\mathcal{L} d \quad s.t. \quad d \in \mathcal{K}_{p^\star}(x^\star, z; h) \;, \tag{36}$$

with $\nabla^2_{xx}\mathcal{L} = \nabla^2_{xx}\mathcal{L}(x^\star, y, z, p^\star)$, $\nabla^2_{xp}\mathcal{L} = \nabla^2_{xp}\mathcal{L}(x^\star, y, z, p^\star)$.

Finally, Proposition 3.8 has been extended by [RD95], who gave a practical way to evaluate the sensitivities for a degenerate nonlinear program by looking at multipliers solution of the linear program

$$S_{p^\star}(x^\star; h) := \begin{cases} \arg\max\limits_{y,z} \; y^\top \nabla_p g(x^\star, p^\star) h + z^\top \nabla_p h(x^\star, p^\star) h \\ \quad \text{s.t.} \quad (y, z) \in \mathcal{M}_{p^\star}(x^\star) \;. \end{cases} \tag{37}$$

The set $S_{p^\star}(x^\star; h)$ is intimely related to the critical set $\mathcal{K}_p(x^\star, z; h)$. Indeed, for given $h \in \mathbb{R}^{n_p}$ and $(y, z) \in \mathcal{M}_p(x)$, the set $\mathcal{K}_p(x, z; h)$ is nonempty if there exists $d \in \mathbb{R}^{n_x}$ such that

$$\begin{aligned} &\nabla_x g_i(x^\star, p^\star)^\top d + \nabla_p g_i(x^\star, p^\star)^\top h = 0 && \forall i \in [m_e] \;, \\ &\nabla_x h_i(x^\star, p^\star)^\top d + \nabla_p h_i(x^\star, p^\star)^\top h \leq 0 && \forall i \in \mathcal{I}^0_p(x^\star; z) \;, \\ &\nabla_x h_i(x^\star, p^\star)^\top d + \nabla_p h_i(x^\star, p^\star)^\top h = 0 && \forall i \in \mathcal{I}^+_p(x^\star; z) \;. \end{aligned} \tag{38}$$

In addition, $(y, z) \in \mathcal{M}_p(x)$ gives

$$\nabla_x f(x^\star, p^\star) + \sum_{i=1}^{m_e} y_i \nabla_x g_i(x^\star, p^\star) + \sum_{i \in \mathcal{I}_p(x^\star)} z_i \nabla_x h_i(x^\star, p^\star) = 0 \;, \quad z_i \geq 0 \;. \tag{39}$$

Combining (38) and (39) together, they rewrite equivalently

$$\begin{aligned} &\nabla_x f(x^\star, p^\star) + \sum_{i=1}^{m_e} y_i \nabla_x g_i(x^\star, p^\star) + \sum_{i \in \mathcal{I}(x^\star)} z_i \nabla_x h_i(x^\star, p^\star) = 0 \\ &\nabla_x g_i(x^\star, p^\star)^\top d + \nabla_p g_i(x^\star, p^\star)^\top h = 0 && i \in [m_e] \;, \\ &\big[\nabla_x h_i(x^\star, p^\star)^\top d + \nabla_p h_i(x^\star, p^\star)^\top h\big] z_i = 0 && i \in \mathcal{I}_p(x^\star) \;, \\ &z_i \geq 0 \;, \quad \nabla_x h_i(x^\star, p^\star)^\top d + \nabla_p h_i(x^\star, p^\star)^\top h \leq 0 && i \in \mathcal{I}_p(x^\star) \;. \end{aligned} \tag{40}$$

The equations (40) are exactly the KKT optimality conditions of (37), with d a multiplier associated to the KKT stationary condition. Hence, by theory of linear programming, the problem (37) has an optimal solution if and only if the previous KKT system is consistent, meaning that $\mathcal{K}_p(x^\star, z; h)$ is nonempty. This is formalized in the following lemma.

Lemma 3.9 (Lemma 2.2, [Dem93]). *The critical cone $\mathcal{K}_p(x^\star, z; h)$ is nonempty if and only if $(y, z) \in S_p(x^\star; h)$.*

Theorem 3.10 (Theorem 2, [RD95]). *Suppose that at a KKT stationary solution $x^\star$ of NLP($p^\star$), MFCQ, CRCQ and GSSOSC hold. Then,*

- *There is open neighborhoods U of $p^\star$ and V of $x^\star$ and a function $x : U \to V$ such that $x(p)$ is the unique local solution of NLP(p) in V and MFCQ holds at $x(p)$;*
- *$x(\cdot)$ is a PC^1 function (hence locally Lipschitz and B-differentiable);*
- *For each $p \in U$, the directional derivative $x'(p; \cdot)$ is a piecewise linear function, and for $h \in \mathbb{R}^{n_p}$ and $(y, z) \in S_p(x^\star; h)$, $x'(p; h)$ is the unique solution of* (36).

Interestingly, the previous theorem shows that directional derivative is unique even when the solution of the LP (37) is not. It also proves that the local solution $x(\cdot)$ is piecewise differential in the sense of Scholtes [Sch12]. It turns out that the notion of PC^1 functions is adapted to characterize the solution of a parametric nonlinear program.

3.4. Differentiation with lexicographic derivative

The lexicographic derivative has been introduced recently has a practical differentiation method for nonsmooth functions, such as PC^1 functions [BKSW18]. It has been recently applied to the differentiation of parametric nonlinear program. Indeed, one can rewrite the KKT conditions (6) as a system of nonsmooth equations:

$$\Phi(x, y, z, p) := \begin{bmatrix} \nabla_x L(x, y, z, p) \\ g(x, p) \\ \min\{-h(x, p), z\} \end{bmatrix} = 0 . \tag{41}$$

It is well known that the nonsmooth KKT system (41) can be solved using a semi-smooth Newton method. At the solution, the sensitivity analysis of the nonsmooth system can be carried out using lexicographic derivatives. Interestingly, this method has the advantage of returning the sensitivity of the dual variables, which is missing in the results of Shapiro [Sha85] and Ralph and Dempe [RD95].

The function min is PC^1 in the sense of Scholtes [Sch12], so we deduce that the mapping Φ is itself PC^1 on its domain. The notion of coherently oriented structure allows to extend the implicit function theorem to such PC^1 functions. The following theorem is based on an extension of the implicit function theorem 3.1 to nonsmooth functions.

Proposition 3.11 (Theorem 5.1 [SKB18])**.** *Let $(x^\star, y^\star, z^\star)$ be a solution of KKT. If Φ is* coherently oriented *with respect to (x, y, z) at $(x^\star, y^\star, z^\star)$ then there exists a neighborhood U of p and V of $x^\star$ and a PC^1 mapping $w : U \to V$ such that for each $p \in U$ $w(p)$ is the unique solution of* (41)*. Moreover, for any $k \in \mathbb{N}$ and any $R \in \mathbb{R}^{p \times k}$ the LD-derivatives $x'(p; R)$, $y'(p, R)$ and $z'(p, R)$ are the unique solution X, Y and $Z = [Z^+ \; Z^0 \; Z^-]$ of the following nonsmooth equation system*

$$\begin{aligned} \begin{bmatrix} \nabla^2_{xx}\mathcal{L} & G_x^\top & (H_x^+)^\top \\ G_x & 0 & 0 \\ H_x^+ & 0 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z^+ \end{bmatrix} &= - \begin{bmatrix} \nabla^2_{xp}\mathcal{L} \\ G_p \\ H_p^+ \end{bmatrix} R , \\ \boldsymbol{LMmin}(-H_p^0 R - H_x^0 X, Z^0) &= 0 , \\ Z^- &= 0 , \end{aligned} \tag{42}$$

*where **LMmin** is the lexicographic matrix minimum function defined in* (106).

Without surprise, ensuring complete coherently orientation is equivalent to classical constraints qualification assumptions.

Proposition 3.12. *Let $(x^\star, y^\star, z^\star)$ be a KKT stationary point. Suppose that LICQ and SSOSC holds at $(x^\star, y^\star, z^\star)$. Then Φ is coherently oriented with respect to (x, y, z).*

Solving the nonsmooth system (42) is generally non trivial, and falls back to a nonsmooth Newton method or an active set algorithm (cycling through linear equation system solves parameterized by the current working active set). However, it has been proved in [SJB19] that evaluating the LD derivatives can be rewritten as the solution a hierarchy of k QP problems $QP_{(1)}, QP_{(2)}, \cdots, QP_{(k)}$, defined for $j = 1, \cdots, k$ as

$$\begin{aligned} QP_{(j)}(r_{(j)}) \quad : \quad & \min_{d \in \mathbb{R}^{n_x}} \frac{1}{2} d^\top \nabla^2_{xx}\mathcal{L}\, d + r_{(j)}^\top \nabla^2_{px}\mathcal{L}\, d \\ \text{subject to} \quad & \nabla_x g_i(x,p)^\top d + \nabla_p g_i(x,p)^\top r_{(j)} = 0 \qquad \forall i \in [m_e] \quad \leftarrow \text{multiplier } \alpha \\ & \nabla_x h_i(x,p)^\top d + \nabla_p h_i(x,p)^\top r_{(j)} = 0 \quad \forall i \in \mathcal{A}^+_{(j-1)} \quad \leftarrow \text{multiplier } \beta \\ & \nabla_x h_i(x,p)^\top d + \nabla_p h_i(x,p)^\top r_{(j)} \leq 0 \quad \forall i \in \mathcal{A}^0_{(j-1)} \quad \leftarrow \text{multiplier } \gamma \end{aligned}$$

We note $\alpha_{(j)} \in \mathbb{R}^{m_e}$, $\beta_{(j)} \in \mathbb{R}^{m_i^+}$, $\gamma_{(j)} \in \mathbb{R}^{m_i^0}$ the unique multipliers associated resp. to the equality constraints, the strongly active constraints and the weakly active constraints. The sets of strongly and weakly active constraints change between two consecutive indexes j using the information brought by the latest multiplier $\gamma_{(i)}$:

$$\begin{aligned} \mathcal{A}^0_{(j)} &= \{ i \in \mathcal{A}^0_{(j-1)} \ : \ \nabla_x h_i(x,p)^\top d + \nabla_p h_i(x,p)^\top p = 0 \text{ and } \gamma_i = 0 \} \ , \\ \mathcal{A}^+_{(j)} &= \{ i \in \mathcal{A}^0_{(j-1)} \ : \ \nabla_x h_i(x,p)^\top d + \nabla_p h_i(x,p)^\top p = 0 \text{ and } \gamma_i > 0 \} \cup \mathcal{A}^+_{(j-1)} \ , \end{aligned} \tag{43}$$

with $\mathcal{A}^0_{(0)} = \mathcal{I}^0_p(x; z)$ and $\mathcal{A}^+_{(0)} = \mathcal{I}^+_p(x; z)$.

Proposition 3.13 (Theorem 3.1 [SJB19])**.** *Let $(x^\star, y^\star, z^\star)$ be a solution of KKT. Suppose that LICQ and SSOSC hold. Then, for any $k \in \mathbb{N}$ and $R \in \mathbb{R}^{n_p \times k}$, the LD-derivatives $x'(p, R)$, $y'(p, R)$ and $z'(p, R)$ in the direction R are given as*

$$\begin{aligned} x'(p,R) &= \left[d_{(1)}(r_{(1)}) \ d_{(2)}(r_{(2)}) \ \cdots \ d_{(k)}(r_{(k)}) \right] \ , \\ y'(p,R) &= \left[\alpha_{(1)}(r_{(1)}) \ \alpha_{(2)}(r_{(2)}) \ \cdots \ \alpha_{(k)}(r_{(k)}) \right] \ , \\ z'(p,R) &= \left[\mu_{(1)}(r_{(1)}) \ \mu_{(2)}(r_{(2)}) \ \cdots \ \mu_{(k)}(r_{(k)}) \right] \ , \end{aligned} \tag{44}$$

with, for $j = 1, \cdots, k$,

$$\begin{aligned} \mu_{(j)}(r_{(j)})_{\mathcal{A}^+_{(j)}} &= \beta_{(j)}(r_{(j)}) \ , \\ \mu_{(j)}(r_{(j)})_{\mathcal{A}^0_{(j)}} &= \gamma_{(j)}(r_{(j)}) \ , \\ \mu_{(j)}(r_{(j)})_{\mathcal{A}^-_{(j)}} &= 0 \ . \end{aligned} \tag{45}$$

4. Sensitivity of the optimal value function

We now focus on the differentiability of the value function $\varphi(p)$.

4.1. Danskin lemma and its extensions

The study of sensitivity analysis of the optimal value function traces back to the seminal work of Danskin, who studied the sensitivity for a simplified problem where the admissible set $X(p)$ is a fixed subset $X \subset \mathbb{R}^{n_x}$, that is,

$$\varphi(p) = \min_{x \in X} \; f(x, p) \; . \tag{46}$$

Theorem 4.1 (Danskin's Lemma [Dan67]). *Suppose X is nonempty and compact, and f is continuously differentiable at $p \in \mathbb{R}^{n_p}$. Then φ is locally Lipschitz near p, and directionally differentiable in every direction $h \in \mathbb{R}^{n_p}$, with*

$$\varphi'(p; h) = \min_{x \in \Sigma(p)} \; \nabla_p f(x, p)^\top h \; . \tag{47}$$

Danskin's theorem has been extended by Hogan to the case where $X(p)$ is defined as $X(p) = \{x \in \mathbb{R}^{n_x} \; : \; x \in M \; , \; h(x, p) \leq 0\}$ for a fixed set M.

Theorem 4.2 ([Hog73]). *Suppose M is a closed convex set, f and h_i are convex on M for all p, and continuously differentiable on $M \times N$, with N a neighborhood of p. If (i) $\Sigma(p)$ is nonempty and bounded, (ii) $\varphi(p)$ is finite, (iii) there is a point $\hat{x} \in M$ such that $h(\hat{x}, p) < 0$ (Slater), then φ is directionally differentiable, and for all $h \in \mathbb{R}^{n_p}$,*

$$\varphi'(p; h) = \min_{x \in \Sigma(p)} \max_{z \in \mathcal{M}_p(x)} \nabla_p \mathcal{L}(x, z, p)^\top h \; . \tag{48}$$

The extension to the generic case associated to $X(p) = \{x \in \mathbb{R}^{n_x} \mid g(x, p) = 0 \, , \, h(x, p) \leq 0\}$ was given by Gauvin and Dubeau [GD82] and Fiacco [Fia82]. The following theorem bounds the lower and upper Dini derivatives of the value function (in general the bounds are sharp).

Theorem 4.3 (Theorem 2.3.4 [Fia83]). *Suppose that $X(p)$ is nonempty and uniformly compact near p and MFCQ holds at each $x \in \Sigma(p)$. Then φ is locally Lipschitz near p, and for any $h \in \mathbb{R}^{n_p}$,*

$$\begin{aligned} \inf_{x \in \Sigma(p)} \min_{(y,z) \in \mathcal{M}_p(x)} \nabla_p \mathcal{L}(x, y, z, p)^\top h &\leq \liminf_{t \downarrow 0} \frac{1}{t} \big(\varphi(p + th) - \varphi(p) \big) \\ &\leq \limsup_{t \downarrow 0} \frac{1}{t} \big(\varphi(p + th) - \varphi(p) \big) \\ &\leq \inf_{x \in \Sigma(p)} \max_{(y,z) \in \mathcal{M}_p(x)} \nabla_p \mathcal{L}(x, y, z, p)^\top h \; . \end{aligned} \tag{49}$$

Proposition 4.4 (Corollary 2.3.8 [Fia83]). *Let f and g be convex functions in x, h be affine in x, such that f, g, h are all jointly C^1 in (x, p). Then, if $X(p)$ is nonempty and uniformly compact near p and MFCQ holds at each $x \in \Sigma(p)$, then φ is directionally differentiable at p and*

$$\varphi'(p; h) = \min_{x \in \Sigma(p)} \max_{(y,z) \in \mathcal{M}_p(x)} \nabla_p \mathcal{L}(x, y, z, p)^\top h \; . \tag{50}$$

Note that if in addition SMCFQ holds, then $\mathcal{M}_p(x)$ reduces to a singleton $(y^\star, z^\star)$ and the directional derivative simplifies to

$$\varphi'(p; h) = \min_{x \in \Sigma(p)} \nabla_p \mathcal{L}(x, y^\star, z^\star, p)^\top h \; . \tag{51}$$

If instead we suppose GSSOSC holds, the local solution $x(p)$ is a singleton and the computation of the directional derivative also simplifies.

Theorem 4.5 (Theorem 7.5 [FK84])**.** *Suppose GSSOSC and MFCQ hold at a stationary solution $x^\star$. Then near p φ has a finite one-sided directional derivative in the Hadamard sense, and for every $h \in \mathbb{R}^{n_p}$,*

$$d_H\varphi(p;h) = \max_{(y,z)\in\mathcal{M}_p(x^\star)} \nabla_p\mathcal{L}(x^\star,y,z,p)^\top h \ . \tag{52}$$

4.2. The differentiable case

Now, we impose more stringent constraint qualifications where the implicit function theorem applies, the value function φ becomes C^2 differentiable. For a given $p \in \mathbb{R}^{n_p}$, the following results hold for any KKT stationary point $w(p) = (x(p), y(p), z(p))$ of NLP(p). As such, we define the *local value function* as

$$\varphi_\ell(p) = f(x(p), p) \ . \tag{53}$$

Theorem 4.6 (Theorem 3.4.1 [Fia83])**.** *Suppose SOSC, SCS and LICQ hold at a stationary KKT solution $(x^\star, y^\star, z^\star)$. Then the local value function φ_ℓ is C^2 near p, and*

- *$\varphi_\ell(p) = \mathcal{L}(x^\star, y^\star, z^\star, p)$.*
- *$\nabla_p\varphi_\ell(p) = \nabla_p\mathcal{L}(x^\star, y^\star, z^\star, p)$.*
- *and also*

$$\begin{aligned}\nabla^2_{pp}\varphi_\ell(p) = \nabla_p\big(\nabla_p\mathcal{L}(w^\star,p)\big) &= \nabla^2_{pp}\mathcal{L} + \nabla^2_{wp}\mathcal{L}\ \nabla_p w(p) \ ,\\ &= \nabla^2_{pp}\mathcal{L} - \nabla^2_{wp}\mathcal{L}\,(\nabla^2_{ww}\mathcal{L})^{-1}\nabla^2_{pw}\mathcal{L} \ .\end{aligned} \tag{54}$$

The previous theorem gives expressions both for the gradient and the Hessian of the optimal value function. It applies to generic nonlinear problems of the form NLP(p). It yields interesting extension when applied to problem with specific structures.

Case 1: the constraints are independent of p. First, we can prove that if the constraints of the nonlinear problem do not depend on the parameter p,

$$\text{NLP}_1(p) := \min_{x\in\mathbb{R}^{n_x}} f(x,p) \quad \text{subject to} \quad \begin{cases} g(x) = 0 \ , \\ h(x) \le 0 \ , \end{cases} \tag{55}$$

the gradient and the Hessian depend only on the derivatives of the objective function f.

Corollary 4.7. *Suppose the conditions of Theorem 4.6 hold at a stationary solution $(x^\star, y^\star, z^\star)$ of problem (55). Then, in a neighborhood of p the local value function φ_ℓ is C^2, with*

- *$\varphi_\ell(p) = f(x^\star, p)$.*
- *$\nabla_p\varphi_\ell(p) = \nabla_p f(x^\star)$.*
- *$\nabla^2_{pp}\varphi_\ell(p) = \nabla^2_{pp}f + \nabla^2_{xp}f\ \nabla_p x(p)$.*

Case 2: RHS perturbations and shadow prices. The other corollary draws a link between Theorem 4.6 and the shadow price of sensitivity: if the parameters appear only in the right-hand-side of the constraint,

$$\mathrm{NLP}_2(p) := \min_{x\in\mathbb{R}^{n_x}} f(x) \quad \text{subject to} \quad \begin{cases} g(x) = p \ , \\ h(x) \leq p \ , \end{cases} \tag{56}$$

then the gradient of the value function is equal to the Lagrange multipliers found at the solution. This is well-known result in convex analysis, the problem (56) being subject to what is called *canonical perturbations*.

Corollary 4.8 (Optimal value function derivatives for RHS perturbations)**.** *Suppose the conditions of Theorem 4.6 hold at a stationary solution $(x^\star, y^\star, z^\star)$ of problem* (56)*. Then, in a neighborhood of p the local value function φ_ℓ is C^2, with*

$$\nabla_p \varphi_\ell(p^\star) = \begin{bmatrix} y^\star \\ z^\star \end{bmatrix} \quad \textit{and} \quad \nabla^2_{pp}\varphi_\ell(p^\star) = \begin{bmatrix} \nabla_p y(p^\star) \\ \nabla_p z(p^\star) \end{bmatrix} \ . \tag{57}$$

5. Numerical algorithms for sensitivity analysis

Now the theory has been laid out, we give an overview of the existing numerical methods for the sensitivity analysis of nonlinear optimization problem.

5.1. Sensitivity with penalty functions

Numerous algorithms are using penalty or barrier functions to solve (approximately) the nonlinear problem NLP(p). The idea is then to evaluate the sensitivity at the approximate solution returned by the algorithm, and give error bounds w.r.t. the sensitivity at the optimal solution.

5.1.1. Fiacco and SENSUMT

In their seminal work [FM68], Fiacco and McCormick derived a practical approach to solve NLP(p). The idea is to find the minimum of the following penalty function

$$W(x, r, p) = f(x, p) - r \sum_{i=1}^{m_i} \log(-h_i(x, p)) + \frac{1}{2r} \sum_{j=1}^{m_i} g_j^2(x, p) \ , \tag{58}$$

where $r > 0$ is a penalty parameter appearing in the barrier associated to the inequality constraints and the quadratic penalty associated to the equality equations. For a fixed (r, p), a stationary point $\nabla_x W\big(x(p, r), r, p\big) = 0$ is found using the traditional Newton method. SENSUMT [FG80] was developed as an extension of SUMT using the penalty function W to estimate the sensitivity of the primal solution $x(r, p)$, but also the dual solution $y(r, p), z(r, p)$ (here estimated directly from the primal solution).

Note that

$$\begin{aligned} \nabla_x W &= \nabla_x f + \sum_{i=1}^{m_i} (r/h_i) \nabla_x h_i + \sum_{j=1}^{m_e} (g_j/r) \nabla_x g_j \ , \\ \nabla_x L &= \nabla_x f + \sum_{i=1}^{m_i} z_i \nabla_x h_i + \sum_{j=1}^{m_e} y_i \nabla_x g_i \ . \end{aligned} \tag{59}$$

Hence, any solution $x^\star$ of (58) will satisfy $\nabla_x L(x^\star, u^\star, w^\star, p) = 0$, with $u^\star = r/h(x^\star, p)$ and $w^\star = g(x^\star, p)/r$. The variables $(u^\star, w^\star)$ are an approximation of the Lagrange multipliers that depends on the primal solution $x^\star$.

Proposition 5.1 (Approximation of first-order sensitivity)**.** *Let $p^\star \in \mathbb{R}^{n_p}$. Suppose that a stationary solution $x^\star$ of NLP(p) LICQ, SOSC and SCS hold. Then in a neighborhood of $(0, p^\star)$ there exists a unique differentiable function $w(r,p) = (x(r,p), y(r,p), z(r,p))$ satisfying*

$$
\begin{aligned}
&\nabla_x \mathcal{L}(x, y, z, p) = 0 \; , && \\
&u_i h_i(x, p) = r \; , && \forall i \in [m_i] \; , \\
&g_j(x, p) = w_j r && \forall i \in [m_e],
\end{aligned}
\tag{60}
$$

and satisfying $w(0, p^\star) = (x^\star, y^\star, z^\star)$. In addition, for any (r, p) near $(0, p^\star)$ $x(r, p)$ is locally unique unconstrained local minimizing point of $W(x, r, p)$ with $h_i(x(r, p), p) < 0$ and $\nabla^2_{xx} W(x(r, p), r, p)$ positive definite.

At the solution of (58), the sensitivity of the primal $x(r, p)$ is given by the implicit function theorem as

$$
\nabla_p x(r, p) = -(\nabla^2_{xx} W)^{-1} \nabla_{pw} W \; , \tag{61}
$$

involving only the inverse of the $n_x \times n_x$ matrix $\nabla^2_{xx} W$. Using Proposition 5.1, we get that $\lim_{r \to 0} \nabla_p x(r, p) = \nabla_p x^\star(p)$. The dual sensitivities are recovered with (60) as $u_i(r, p) = r/h_i(x(r, p), p)$ and $w_j = g_j(x(r, p), p)/r$, yielding

$$
\begin{aligned}
\nabla_p u(r, p) &= -\frac{r}{h(x(r, p), p)} \big(\nabla_x h(x(r, p), p)^\top \nabla_p x(r, p) + \nabla_p h(x(r, p), p) \big) \; , \\
\nabla_p w(r, p) &= \frac{1}{r} \big(\nabla_x g(x(r, p), p)^\top \nabla_p x(r, p) + \nabla_p g(x(r, p), p) \big) \; .
\end{aligned}
\tag{62}
$$

5.1.2. Sensitivity in interior-point

The traditional approach adopted in SENSUMT has been refreshed to compute the sensitivity of a stationary solution found by a modern interior-point algorithm. Notably, sIpopt [PLNB12] is a package for the sensitivity analysis of NLP developed as an extension of the primal-dual interior-point solver Ipopt. Ipopt rewrites the inequalities in NLP(p) as equality constraints by introducing additional slack variables, yielding the reformulated NLP problem

$$
\begin{aligned}
&\min_x \; f(x, p) \\
&\text{s.t.} \quad g(x, p) = 0 \; , \quad x \geq 0 \; .
\end{aligned}
\tag{63}
$$

For a given barrier parameter $\mu > 0$, the bound constraints $x \geq 0$ are penalized using a barrier term in the objective function, giving a subproblem with only equality constraints:

$$
\begin{aligned}
&\min_x \; f(x, p) - \mu \sum_{i=1}^{n_x} \log(x_i) \\
&\text{s.t.} \quad g(x, p) = 0 \; .
\end{aligned}
\tag{64}
$$

Ipopt solves a sequence of subproblems (64) for given barrier parameters $\{\mu_k\}_k$, with $\mu_k \to 0$. For a fixed μ, solving (64) amounts to solve the system of nonlinear equations $F_\mu(x, y, z, p) =$

0, with

$$F_\mu(x, y, z, p) = \begin{bmatrix} \nabla_x f(x, p) + y^\top \nabla_x g(x, p) - z \\ g(x, p) \\ Xz - \mu e \end{bmatrix} . \tag{65}$$

The following proposition is a corollary of Proposition 5.1.

Proposition 5.2 (Convergence of interior-point)**.** *Let $w^\star = (x^\star, y^\star, z^\star)$ be a local primal-dual stationary point of* (63)*. Suppose that LICQ, SCS and SOSC hold at $w^\star$. We denote by $x(\mu_k, p)$ the solution of the barrier problem* (64) *at iteration k. Then*

- *There is at least one subsequence of $x(\mu_k, p)$ converging to $x^\star$ and for every convergent subsequence, the corresponding barrier multiplier approximations is bounded and converges to $(y^\star, z^\star)$.*
- *There exists a unique continuously differentiable function $w(\cdot, p)$ existing in a neighborhood of* 0 *for $\mu > 0$ and solution of* (64)*, and such that $\lim_{\mu \downarrow 0} w(\mu, p) = w^\star$.*
- *For μ small enough,*
$$\nabla_p w(\mu, p) = -(\nabla_w F_\mu)^{-1} \nabla_p F_\mu \, . \tag{66}$$

To solve the sparse linear system (66), sIpopt is reusing the factorization of the KKT system computed by Ipopt at the local solution $w(\mu, p)$. Computing the full Jacobian $\nabla_p w(\mu, p)$ can be performed with p backsolves. To handle change in the active set, sIpopt uses a fix-and-relax strategy, without resorting to the resolution of the QP (36). The new constraints activated are incorporated in the linear system (66), and the system is solved using a Schur-complement approach.

Variant 1: CasADi. The sensitivity analysis algorithm implemented in CasADi [AR18, AGH+19] is slightly different. Once a solution returned by Ipopt, it applies a primal-dual active set method to determine exactly the active set at the solution (an information that interior-point does not return by default, unless a crossover is used). This translates to the solution of a QP, CasADi uses a custom sparse QR solver to solve a non-symmetric formulation of the KKT condition. Such QR factorization allows to quickly detect and resolve the singularities associated to potential degeneracy.

Variant 2: QPTH. The method has been recently specialized in [AK17] to the parallel extraction of the sensitivity from a QP, for a fixed number of parameters $(p_1, \cdots, p_N)$. The solver QPTH solves N QP problems in parallel using a specialized interior-point method leveraging batched dense linear algebra on the GPU. Once a solution found, the sensitivity are evaluated using a formula analogous to (66). This has been proven relevant for machine learning application, where neural network are trained with batched stochastic gradient descent. To the best of our knowledge, the method has not been extended to sparse problems, limiting its adoption beyond machine learning. Also, little emphasis is put on the degenerate case.

5.2. Path following method

The method of Ralph and Dempe [RD95] gives a practical way to evaluate the directional derivative in the absence of LICQ. However, the method is valid only for small perturbations around $p^\star$. This can lead to poor approximation if the new parameter p is not close to $p^\star$. In that case, it is more appropriate to use a path following method to track the path of optimal solution between $p^\star$ and the new parameter p. To do so, we define the affine interpolation $p(t) = p^\star + t(p - p^\star)$ for $t \in [0, 1]$ and track the solution of NLP(p) from $t = 0$ to $t = 1$. This approach gives better performance, notably when the perturbation is causing active set changes. The method has been studied in different articles [JYB14, KJ17].

5.2.1. Canonical method

For a sequence of scalars $t_{(0)} < t_{(1)} < \cdots < t_{(N)}$ such that $t_{(0)} = 0$ and $t_{(N)} = 1$, the perturbation between $t_{(m)}$ and $t_{(m+1)}$ is defined as

$$h_{(m)} = (t_{(m+1)} - t_{(m)})(p - p^\star) \ , \quad \forall m = 0, \cdots , N - 1 \ . \tag{67}$$

Then, we apply $N - 1$ sensitivity updates to follow the path of the primal-dual solution $w(t)$ between $t_{(0)}$ and $t_{(N)}$. As noted in [JYB14], this approach is akin to an explicit Euler integration method, with multiple sensitivity updates performed in such a way to reduce the approximation error between $t_{(0)}$ and $t_{(N)}$.

Denote by $w^\star = (x^\star, y^\star, z^\star)$ the optimal solution at $p^\star$, and define $x_{(0)} = x^\star, y_{(0)} = y^\star, z_{(0)} = z^\star$. The algorithm proceeds iteratively, for $t = t_{(0)}, \cdots , t_{(N)}$.

Step 1: Updating the primal sensitivity. We update the primal variable as: $x_{(m+1)} = x_{(m)} + \Delta x_{(m)}$, with $\Delta x_{(m)}$ solution of the QP problem

$$\begin{aligned}
\min_d \ & \frac{1}{2} d^\top \nabla^2_{xx}\mathcal{L}_{(m)}\, d + h_{(m)}^\top \nabla^2_{xp}\mathcal{L}_{(m)} d \\
\text{s.t.} \ & \nabla_x g_i(x_{(m)}, p_{(m)})^\top d + \nabla_p g_i(x_{(m)}, p_{(m)})^\top h_{(m)} = 0 \quad \forall i \in [m_e] \ , \\
& \nabla_x h_i(x_{(m)}, p_{(m)})^\top d + \nabla_p h_i(x_{(m)}, p_{(m)})^\top h_{(m)} = 0 \quad \forall i \in \mathcal{A}^+_{(m)} \ , \\
h_i(x_{(m)}, p_{(m)}) + & \nabla_x h_i(x_{(m)}, p_{(m)})^\top d + \nabla_p h_i(x_{(m)}, p_{(m)})^\top h_{(m)} \leq 0 \quad \forall i \in \mathcal{A}^0_{(m)} \ ,
\end{aligned} \tag{68}$$

with $\mathcal{L}_{(m)} = \mathcal{L}(x_{(m)}, y_{(m)}, z_{(m)}, p_{(m)})$ We note $\mathcal{A}^+_{(m)} = \{i \ : \ (z_{(m)})_i > 0\}$ and $\mathcal{A}^0_{(m)} = \{i \ : \ (z_{(m)})_i = 0\}$. Note that QP (68) is almost similar to (36), apart from the last constraints representing the linearized inactive inequality constraints. That means that the QP (68) has exactly the same solution as (36) as long as there is no previously inactive constraint becoming active. In case one constraint becomes active, QP (68) will attempt to satisfy this constraint to first order; if it fails the QP becomes infeasible and the step $h_{(m)}$ has to be reduced by decreasing $t_{(m+1)}$.

Step 2: Updating the dual sensitivity. Updating the dual sensitivity is more demanding, as the multipliers of (68) do not correspond to the sensitivity of the original NLP's multipliers (unless LICQ and SCS hold). In addition, LP (37) is infeasible unless evaluated at an optimal point $x^\star$. Hence, we have to adapt the LP (37), this time using the optimality conditions associated to QP (68).

In that case, the multipliers are updated as $y_{(m+1)} = y_{(m)} + \Delta y_{(m)}$ and $z_{(m+1)} = z_{(m)} + \Delta z_{(m)}$, with $(\Delta y_{(m)}, \Delta z_{(m)})$ solution of

$$\begin{aligned}
\max_{\delta y, \delta z} \ & \delta y^\top \nabla_p g(x_{(m)}, p_{(m)}) h_{(m)} + \delta z^\top \nabla_p h(x_{(m)}, p_{(m)}) h_{(m)} \\
\text{s.t.} \ & \nabla^2_{xp} \mathcal{L}_{(m)} h_{(m)} + \nabla^2_{xx} \mathcal{L}_{(m)} \Delta x_{(m)} \\
& + \sum_{i=1}^{m_e} \nabla_x g_i(x_{(m)}, p_{(m)}) \delta y_i + \sum_{i=1}^{m_i} \nabla_x h_i(x_{(m)}, p_{(m)}) \delta z_i = 0 \ , \\
& \delta z_i \geq 0 \ , \ \forall i \in \mathcal{A}^0_{(m)} \setminus \hat{A}^0_{(m)} \ , \quad \delta z_i = 0 \ , \ \forall i \in \hat{A}^0_{(m)} \ , \\
& z_{(m)} + \delta z \geq 0 \ ,
\end{aligned} \tag{69}$$

where $\hat{A}^0_{(m)} = \{i \ : \ i \in \mathcal{A}^0_{(m)} \ , \ h_i(x_{(m)}, p_{(m)}) + \nabla_x h_i(x_{(m)}, p_{(m)})^\top \Delta x_{(m)} + \nabla_p h_i(x_{(m)}, p_{(m)})^\top h_{(m)} < 0\}$.

5.3. Sensitivity as a system of inequalities

We present in this subsection an alternative approach, differing slightly from Fiacco's methodology. In [CCC+06], the authors differentiate the KKT system at the solution, and pre-process the second-order constraints associated to the weakly active constraints

$$dz_i \big(\nabla_x h_i(x,p)^\top dx + \nabla_p h_i(x,p)^\top dp \big) \ , \quad \forall i \in \mathcal{I}^0_p(x; z) \ , \tag{70}$$

to classify manually the weakly active constraints in two sets: the one that should remain active, and the others than could become inactive (this gives $2^{m_i^0}$ possibilities). Note that in the classical framework, (70) is considered implicitly in the KKT conditions of QP (36). Once the constraints classified, the admissible perturbations are solution of the system of linear inequality equations

$$U \Delta w = S \Delta p \ , \quad V \Delta w \leq T \Delta p \ , \tag{71}$$

with

$$\begin{aligned}
U &= \begin{bmatrix} \nabla^2_{xx} \mathcal{L} & G_x^\top & H_x^\top \\ G_x & 0 & 0 \\ H_x^+ & 0 & 0 \end{bmatrix} \ , & S &= - \begin{bmatrix} \nabla^2_{xp} \mathcal{L} \\ G_p \\ H_p^+ \end{bmatrix} \ , \\
V &= \begin{bmatrix} H_x^0 & 0 & 0 \\ 0 & 0 & -I^0_{m_j} \end{bmatrix} \ , & T &= - \begin{bmatrix} H_p^0 \\ 0 \end{bmatrix} \ ,
\end{aligned}$$

where H^1 refers to the submatrix of H associated to the non-null multipliers, and H^0 the submatrix of H associated to the null multipliers of the active constraints. This system of inequalities defines the set of all feasible perturbations Δp at a given primal-dual solution $(x^\star, y^\star, z^\star)$. It is well known that the solution of a system of linear inequalities (71) is a polyhedral cone. Hence, using the vertex cone representation (obtained e.g. using the Γ algorithm), a feasible perturbation $(\Delta w, \Delta p)$ solution of (71) is given by

$$\begin{bmatrix} \Delta w \\ \Delta p \end{bmatrix} = \sum_{i=1}^{t} \rho_i v_i + \sum_{j=1}^{q} \pi_j w_j \ , \tag{72}$$

with $\rho_i \in \mathbb{R}$, $\pi_j \in \mathbb{R}_+$ and v_i, w_j vectors generating the polyhedral cone.

5.4. Sensitivity for conic problems

Convex programming has become an important subfield of optimization. In machine learning, a significant number of problems are conic, motivating a renewed interest for the sensitivity analysis of conic-structured problems, formulated as

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} \; c(p)^\top x \\ &\text{s.t.} \quad A(p)x + s = b(p) \; , \quad s \in K \; . \end{aligned} \tag{73}$$

where K is a convex cone. The KKT conditions write out

$$A(p)x + s = b(p) \; , \quad A(p)^\top y + c(p) = 0 \; , \quad s \in K \; , \quad y \in K^* \; , \quad s^\top y = 0 \; . \tag{74}$$

Interestingly, the sensitivity analysis of (73) does not differentiate the KKT conditions (74) at the solution (as done with the classical theory Fiacco) but instead differentiate an equivalent form of (74) called the *homogeneous self-dual embedding.*

Homogeneous self-dual embedding. We follow the description of the homogeneous self-dual embedding introduced in [BMB19, BLPSF21]. We first recall a classical result in convex analysis.

Proposition 5.3 (Moreau decomposition)**.** *Let $(s, y, v) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$. The following propositions are equivalent:*

1. *$v = y - s$ and $s \in K$, $y \in K^*$, $s^\top y = 0$.*
2. *$s = P_{K^*}(v) - v$ and $y = P_{K^*}(v)$.*

Using Proposition 5.3 and setting $v = y - s$ and $u = x$, the KKT conditions (74) rewrite equivalently as

$$\begin{aligned} A^\top P_{K^*}(v) + c &= 0 \; , \\ -Au + b &= P_{K^*}(v) - v \; , \end{aligned} \tag{75}$$

or, more compactly,

$$\begin{bmatrix} 0 & A^\top & c \\ -A & 0 & b \end{bmatrix} \begin{bmatrix} u \\ P_{K^*}(v) \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ v - P_{K^*}(v) \end{bmatrix} = 0 \; . \tag{76}$$

This form is equivalent to the original KKT conditions (74). If in addition we want to tackle the infeasible and unbounded case in (73), we extend (76) to the larger system:

$$\begin{bmatrix} 0 & A^\top & c \\ -A & 0 & b \\ -c^\top & -b^\top & 0 \end{bmatrix} \begin{bmatrix} u \\ P_{K^*}(v) \\ \tau \end{bmatrix} = \begin{bmatrix} 0 \\ P_{K^*}(v) - v \\ \kappa \end{bmatrix} \; , \tag{77}$$

for $\tau > 0$ and $\kappa > 0$. The parameter τ and κ characterize the primal and dual infeasibility of the conic problem (73). For $\tau = 1$ and $\kappa = 0$ we recover the original compact KKT system (76). If we introduce the skew symmetric matrix

$$Q(p) = \begin{bmatrix} 0 & A(p)^\top & c(p) \\ -A(p)^\top & 0 & b(p) \\ -c(p)^\top & -b(p)^\top & 0 \end{bmatrix} \; , \tag{78}$$

then (77) is equivalent as finding two vectors $r, t \in \mathbb{R}^{n+m+1}$ such as

$$Q(p)r = t \ , \quad r \in \mathcal{C} \ , \quad t \in \mathcal{C}^* \ , \tag{79}$$

with $\mathcal{C} = \mathbb{R}^n \times K^* \times \mathbb{R}_+$, $\mathcal{C}^* = \{0\}^n \times K \times \mathbb{R}_+$

If $(x^\star, y^\star, s^\star) \in \mathbb{R}^{n+2m}$ satisfies the KKT conditions (74), then $r^\star = (x^\star, y^\star, 1)$ and $t^\star = (0, s^\star, 0)$ are solutions of (79).

Residual map. Suppose now we have $z \in \mathbb{R}^{n+m+1}$. We decompose z as $z = r - t$, with $r = P_{\mathcal{C}}(z)$ and $t = P_{\mathcal{C}^*}(z)$. The residual map $R(z, p)$ returns the residual of the homogeneous self-dual embedding $Q(p)r - t$ for a particular candidate z, as

$$R(z, p) = Q(p)P_{\mathcal{C}}(z) + P_{\mathcal{C}^*}(z) = ((Q(p) - I)P_{\mathcal{C}} + I)z \ . \tag{80}$$

If $z^\star$ is such that $R(z^\star, p) = 0$, then $(r^\star, t^\star) = (P_{\mathcal{C}}(z^\star), P_{\mathcal{C}^*}(z^\star))$ is solution of the homogeneous self-dual embedding (79). If we now look at $z^\star$ component by component as $z^\star = (u^\star, v^\star, w^\star)$, with $(u^\star, v^\star, w^\star) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$, and $w^\star \neq 0$, the KKT solutions $(x^\star, y^\star, s^\star)$ of (73) is recovered as

$$(x^\star, y^\star, s^\star) = \phi(z^\star) \quad \text{with} \quad \phi(z^\star) := \frac{1}{w^\star}\big(u^\star, P_{K^*}(v^\star), P_{K^*}(v^\star) - v^\star\big) \ . \tag{81}$$

Sensitivity analysis. We now have all the elements to evaluate the sensitivity at a given primal-dual solution $w^\star = (x^\star, s^\star, y^\star)$ of (73). By defining $r^\star = (x^\star, y^\star, 1)$, $t^\star = (0, s^\star, 0)$ and $z^\star = u^\star - v^\star$, the vector $z^\star$ is solution of the residual map and $R(z^\star, p^\star) = 0$. In addition, if the projection operator $P_{\mathcal{C}}$ is differentiable, then R is itself differentiable with $\nabla_z R(z, p) = (Q - I)\nabla_z P_{\mathcal{K}}(z) + I$. In that case, by applying the implicit function theorem, we get

$$\nabla_p z^\star = -(\nabla_z R(z^\star))^{-1} \nabla_p R(z^\star) \ . \tag{82}$$

We recover the sensitivity of the primal-dual solution by noting that $w^\star = (x^\star, s^\star, y^\star) = \phi(z^\star)$. By applying the chain-rule, we get $\nabla_p w^\star = \nabla_z \phi \nabla_p z^\star$.

Note that in general the projection operator P_K is not differentiable in general. However, as recalled in [AWK17], if K is the nonnegative orthant, the second-order cone or the positive definite cone then P_K is strongly semismooth (see Appendix A.3), hence differentiable almost everywhere. As a result, the residual mapping (80) is also a strongly semismooth function: the equation (82) is well-defined if we use a version of the implicit function theorem adapted for the strongly semismooth case (and involving Clarke generalized Jacobians instead of the classical Jacobian we encounter in the regular case). Another rigorous treatment has been made in [BLPSF21], based on the newly introduced concept of *conservative Jacobians.*

Implementation. The approach has been implemented in cvxpylayer [AAB+19] and DiffOpt.jl [SBGL22]. Factorizing the matrix $\nabla_z R$ in (82) can be challenging, as on the contrary to the KKT matrix the Jacobian $\nabla_z R$ is non symmetric. The authors have investigated the resolution of (82) with the iterative linear solver LSQR, a variant of the conjugate gradient algorithm which cover also the case where $\nabla_z R$ is non-invertible. Regarding that last point, the original paper gives little detail about that particular degenerate case.

References

[AAB+19] Akshay Agrawal, Brandon Amos, Shane Barratt, Stephen Boyd, Steven Diamond, and J Zico Kolter. Differentiable convex optimization layers. *Advances in neural information processing systems*, 32, 2019.

[AGH+19] Joel AE Andersson, Joris Gillis, Greg Horn, James B Rawlings, and Moritz Diehl. CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11:1–36, 2019.

[AK17] Brandon Amos and J Zico Kolter. Optnet: Differentiable optimization as a layer in neural networks. In *International Conference on Machine Learning*, pages 136–145. PMLR, 2017.

[AR18] Joel AE Andersson and James B Rawlings. Sensitivity analysis for nonlinear programming in CasADi. *IFAC-PapersOnLine*, 51(20):331–336, 2018.

[AWK17] Alnur Ali, Eric Wong, and J Zico Kolter. A semismooth newton method for fast, generic convex programming. In *International Conference on Machine Learning*, pages 70–79. PMLR, 2017.

[BKSW18] Paul I Barton, Kamil A Khan, Peter Stechlinski, and Harry AJ Watson. Computationally relevant generalized derivatives: theory, evaluation and applications. *Optimization Methods and Software*, 33(4-6):1030–1072, 2018.

[BLPSF21] Jérôme Bolte, Tam Le, Edouard Pauwels, and Tony Silveti-Falls. Nonsmooth implicit differentiation for machine-learning and optimization. *Advances in neural information processing systems*, 34:13537–13549, 2021.

[BMB19] Enzo Busseti, Walaa M Moursi, and Stephen Boyd. Solution refinement at regular points of conic problems. *Computational Optimization and Applications*, 74:627–643, 2019.

[BS00] J Frédéric Bonnans and Alexander Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2000.

[CCC+06] E Castillo, AJ Conejo, C Castillo, R Mínguez, and D Ortigosa. Perturbation approach to sensitivity analysis in mathematical programming. *Journal of Optimization Theory and Applications*, 128:49–74, 2006.

[Dan67] John M Danskin. *The theory of max-min and its application to weapons allocation problems*, volume 5. Springer Science & Business Media, 1967.

[Dem93] Stephan Dempe. Directional differentiability of optimal solutions under Slater's condition. *Mathematical programming*, 59:49–69, 1993.

[Dem02] Stephan Dempe. *Foundations of bilevel programming*. Springer Science & Business Media, 2002.

[DN08] Victor DeMiguel and Francisco J Nogales. On decomposition methods for a class of partially separable nonlinear programs. *Mathematics of Operations Research*, 33(1):119–139, 2008.

[DR09] Asen L Dontchev and R Tyrrell Rockafellar. *Implicit functions and solution mappings: A view from variational analysis*, volume 11. Springer, 2009.

[FG80] Anthony V Fiacco and Abolfazl Ghaemi. A user's manual for SENSUMT: A penalty function computer program for solution, sensitivity analysis, and optimal value bound calculation in parametric nonlinear programs. Technical report, 1980.

[Fia76] Anthony V Fiacco. Sensitivity analysis for nonlinear programming using penalty methods. *Mathematical programming*, 10(1):287–311, 1976.

[Fia82] Anthony V Fiacco. Optimal value continuity and differential stability bounds under the mangasarian-fromovitz constraint qualification. *Mathematical programming with data perturbations II*, pages 65–90, 1982.

[Fia83] Anthony V Fiacco. *Introduction to sensitivity and stability analysis in non linear programming*. New York: Academic Press, 1983.

[FK84] Anthony V Fiacco and Jerzy Kyparisis. Sensitivity analysis in nonlinear programming under second order assumptions. In *Systems and Optimization: Proceedings of the Twente Workshop Enschede, The Netherlands, April 16–18, 1984*, pages 74–97. Springer, 1984.

[FL93] Anthony V Fiacco and Jiming Liu. Degeneracy in nlp and the development of results motivated by its presence. *Annals of Operations Research*, 46(1):61–80, 1993.

[FM68] Anthony V Fiacco and Garth P McCormick. *Nonlinear programming: sequential unconstrained minimization techniques*. SIAM, 1968.

[FP03] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

[Gau77] Jacques Gauvin. A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming. *Mathematical Programming*, 12(1):136–138, 1977.

[Gau78] Jacques Gauvin. The method of parametric decomposition in mathematical programming: the nonconvex case. In *IIASA PROCEEDINGS SERIES*, page 131, 1978.

[GD82] Jacques Gauvin and François Dubeau. Differential properties of the marginal function in mathematical programming. *Optimality and Stability in Mathematical Programming*, pages 101–119, 1982.

[Hog73] William W Hogan. Point-to-set maps in mathematical programming. *SIAM review*, 15(3):591–603, 1973.

[Jit84] Krisorn Jittorntrum. Solution point differentiability without strict complementarity in nonlinear programming. *Sensitivity, Stability and Parametric Analysis*, pages 127–138, 1984.

[JYB14] Johannes Jäschke, Xue Yang, and Lorenz T Biegler. Fast economic model predictive control based on NLP-sensitivities. *Journal of Process Control*, 24(8):1260–1272, 2014.

[KJ17] Vyacheslav Kungurtsev and Johannes Jaschke. A predictor-corrector path-following algorithm for dual-degenerate parametric optimization problems. *SIAM Journal on Optimization*, 27(1):538–564, 2017.

[Koj80] Masakazu Kojima. Strongly stable stationary solutions in nonlinear programs. In *Analysis and computation of fixed points*, pages 93–138. Elsevier, 1980.

[Kyp85] Jerzy Kyparisis. On uniqueness of Kuhn-Tucker multipliers in nonlinear programming. *Mathematical Programming*, 32:242–246, 1985.

[Kyp90] Jerzy Kyparisis. Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers. *Mathematics of operations research*, 15(2):286–298, 1990.

[LNB12] Rodrigo López-Negrete and Lorenz T Biegler. A moving horizon estimator for processes with multi-rate measurements: A nonlinear programming sensitivity approach. *Journal of Process Control*, 22(4):677–688, 2012.

[Nes05] Yu Nesterov. Lexicographic differentiation of nonsmooth functions. *Mathematical programming*, 104:669–700, 2005.

[PLNB12] Hans Pirnay, Rodrigo López-Negrete, and Lorenz T Biegler. Optimal sensitivity based on IPOPT. *Mathematical Programming Computation*, 4:307–331, 2012.

[RD95] Daniel Ralph and Stephan Dempe. Directional derivatives of the solution of a parametric nonlinear program. *Mathematical programming*, 70(1-3):159–172, 1995.

[Rob79] Stephen M Robinson. *Generalized equations and their solutions, Part I: Basic theory*. Springer, 1979.

[Rob80] Stephen M Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5(1):43–62, 1980.

[Rob82] Stephen M Robinson. *Generalized equations and their solutions, part II: applications to nonlinear programming*. Springer, 1982.

[SBGL22] Akshay Sharma, Mathieu Besançon, Joaquim Dias Garcia, and Benoît Legat. Flexible differentiable optimization via model transformations. *arXiv preprint arXiv:2206.06135*, 2022.

[Sch12] Stefan Scholtes. *Introduction to piecewise differentiable equations*. Springer Science & Business Media, 2012.

[Sha85] Alexander Shapiro. Second order sensitivity analysis and asymptotic theory of parametrized nonlinear programs. *Mathematical Programming*, 33:280–299, 1985.

[Sha90] Alexander Shapiro. On differential stability in stochastic programming. *Mathematical Programming*, 47:107–116, 1990.

[Sha91] Alexander Shapiro. Asymptotic analysis of stochastic programs. *Annals of Operations Research*, 30:169–186, 1991.

[SJB19] Peter Stechlinski, Johannes Jäschke, and Paul I Barton. Generalized sensitivity analysis of nonlinear programs using a sequence of quadratic programs. *Optimization*, 68(2-3):485–508, 2019.

[SKB18] Peter Stechlinski, Kamil A Khan, and Paul I Barton. Generalized sensitivity analysis of nonlinear programs. *SIAM Journal on Optimization*, 28(1):272–301, 2018.

[ZA10] Victor M Zavala and Mihai Anitescu. Real-time nonlinear optimization as a generalized equation. *SIAM Journal on Control and Optimization*, 48(8):5444–5467, 2010.

[ZB09] Victor M Zavala and Lorenz T Biegler. The advanced-step nmpc controller: Optimality, stability and robustness. *Automatica*, 45(1):86–93, 2009.

A. Derivatives

Let $f : \Omega \subset E \to F$ a function, with E and F two normed spaces. We start to recall the usual derivatives, before introducing the Clarke Jacobian more adapted in the nonsmooth context.

A.1. Differentiability

Definition A.1 (Directional derivatives)**.** *Let $x \in \Omega$. The function f has a* directional derivative *at x in the direction $h \in E$ if $x + th \in \Omega$ for t small enough, and if the following limit exists*

$$f'(x; h) = \lim_{t \downarrow 0} \frac{f(x + th) - f(x)}{t} \; . \tag{83}$$

We note that a function G-differentiable is not necessarily continuous.

Definition A.2 (Gâteaux-differentiability)**.** *The function f is Gâteaux-differentiable (or G-differentiable) in $x \in \Omega$ if there exists a directional derivatives for every direction $h \in E$, and if the application*

$$h \to f'(x; h) \; , \tag{84}$$

is continuous *and* linear*. We note $f'(x)$ the operator such that $f'(x; h) = f'(x) \cdot h$.*

Definition A.3 (Fréchet-differentiability)**.** *The function f is Fréchet-differentiable (F-differentiable, or simply differentiable) at $x \in \Omega$ if there exists a bounded linear operator L such that*

$$\lim_{\|h\| \downarrow 0} \frac{1}{\|h\|} \Big(f(x + h) - f(x) - Lh \Big) = 0 \; . \tag{85}$$

F-differentiability implies that f is continuous. The operator L is the *derivative* of f in x. The condition (85) rewrites equivalently

$$f(x+h) = f(x) + Lh + o(\|h\|) \; . \tag{86}$$

The Hadamard derivative gives the directional derivative along a curve tangential to h. It is required to study the sensitivity of infinite-dimensional problems, as often encounted in statistics or stochastic optimization.

Definition A.4 (Hadamard-differentiability)**.** *The function f is Hadamard-differentiable at x if for any mapping $\varphi : \mathbb{R}_+ \to X$ such that $\varphi(0) = x$ and $\frac{1}{t}(\varphi(t) - \varphi(0))$ converges to a vector h as $t \downarrow 0$, the limit*

$$d_H f(x;h) = \lim_{t \downarrow 0} \frac{1}{t}(f(\varphi(t)) - f(x)) \; , \tag{87}$$

does exist.

For any sequences $\{h_n\}_n$ and $\{t_n\}_n$ such that $h_n \to h$ and $t_n \to 0^+$, the Hadamard directional derivative can also be written in the form

$$d_H f(x;h) = \lim_{n \to \infty} \frac{1}{t_n}(f(x + t_n h_n) - f(x)) \; . \tag{88}$$

The Hadamard derivative has the advantage of being compatible with the chain-rule with minimal assumptions.

Proposition A.5 (Chain-rule)**.** *Let f be Hadamard directionally differentiable at x, and g be Hadamard directionally differentiable at $f(x)$. Then the composite mapping $g \circ f$ is Hadamard directionally differentiable at x, and satisfies the chain-rule $d_H(g \circ f)(x;h) = d_H g(f(x); d_H f(x;y))$.*

Whereas the Hadamard derivative allows for different rates for every directions, the Fréchet derivative imposes the rate is the same for each direction h. For finite-dimensional space E, if the directional derivative is continuous then the Fréchet and the Hadamard derivatives coincide.

A.2. Generalized differentiability

The concept of Dini-derivatives generalizes the notion of directional differentiability for non-smooth functions.

Definition A.6 (Dini-derivatives)**.** *Let $x \in \Omega$. The upper Dini derivative in the direction h is defined as*

$$D^+ f(x;h) = \lim_{t \downarrow 0} \sup \{ \frac{f(x+th) - f(x)}{t} \} \; . \tag{89}$$

Accordingly, the lower Dini derivative is defined as

$$D^- f(x;h) = \lim_{t \downarrow 0} \inf \{ \frac{f(x+th) - f(x)}{t} \} \; . \tag{90}$$

Theorem A.7 (Rademacher Theorem)**.** *Let $f : \Omega \to F$ be a locally Lipschitz continuous function. Then f is differentiable almost everywhere in the sense of the Lebesgue measure.*

Definition A.8 (Bouligand-differential). *Let $\mathcal{D}_f$ be the set of points at which f is (Fréchet) differentiable. The B-differential of f at $x \in \Omega$ is the set defined as*

$$\partial_B f(x) = \big\{ J \in \mathcal{L}(E,F) \ : \ \exists \{x_k\}_k \in \mathcal{D}_f \text{ such that } x_k \to x \ , f'(x_k) \to J \big\} \ . \tag{91}$$

Definition A.9 (Clarke-differential). *The C-differential of f at $x \in \Omega$ is the convex hull of the Bouligand-differential, that is*

$$\partial_C f(x) = conv\, \partial_B f(x) \ . \tag{92}$$

Definition A.10 (Clarke-generalized directional derivative). *The Clarke-generalized directional derivative of f in the direction h is defined by*

$$d_C f(x;h) = \lim_{y \to x} \sup_{t \downarrow 0} \frac{1}{t} \big[f(y+th) - f(y) \big] \ . \tag{93}$$

Proposition A.11. *Suppose that f is L-Lipschitz near $x \in \Omega$. Then*

1. *$\partial_C f(x)$ is nonempty compact and convex,*
2. *$\partial_C f(x)$ is locally bounded and upper semi-continuous at x.*
3. *$d_C f(x; \cdot)$ is the support function of $\partial_C f(x)$, that is,*

$$d_C f(x;h) = \max\{ s^\top h \ : \ \forall s \in \partial_C f(x) \} \ . \tag{94}$$

Proposition A.12. *If f is G-differentiable at x, then $f'(x) \in \partial_C f(x)$. If in addition f is C^1 at x, then $\partial_C f(x) = \{f'(x)\}$.*

Proposition A.13 (Chain-rule). *Suppose that f is L-Lipschitz continuous on Ω. If the function $g : F \to G$ is Lipschitz-continuous near $f(x)$, then*

$$\partial_C (g \circ f)(x) \subset conv\big\{ GF \ : \ G \in \partial_C g(f(x)) \ , \ F \in \partial_C f(x) \big\} \ . \tag{95}$$

A.3. Semi-smooth functions

Definition A.14 (Semi-smoothness). *The function $f : \Omega \to F$ is semi-smooth in $x \in \Omega$ if:*

1. *f is L-Lipschitz near x;*
2. *f has directional derivatives at x in all directions;*
3. *When $h \to 0$, $\sup_{J \in \partial_C f} \|f(x+h) - f(x) - Jh\| = o(\|h\|)$.*

Proposition A.15. *Suppose f is L-Lipschitz near x, and admits directional derivatives at x in all directions. Then the following statements are equivalent:*

1. *f is semi-smooth ;*
2. *for $h \to 0$, $\sup_{J \in \partial_C f} \|Jh - f'(x;h)\| = o(\|h\|)$.*
3. *for $h \to 0$ such that $x+h \in \mathcal{D}_f$, $f'(x+h)h - f'(x;h) = o(\|h\|)$.*

A.4. PC^1 functions

The class of PC^1 functions is a specific tool in nonsmooth analysis, often employed for the sensitivity analysis of nonlinear programs.

Definition A.16. *The function $f : X \to F$ is piecewise-differentiable (PC^1) at $x \in X$ if there exists a neighborhood $N \subset X$ and a finite collection of C^1 selection functions $\mathcal{F}_f(x) = \{f_{(1)}, \cdots, f_{(k)}\}$ defined on N such that f continuous on N and for all $y \in N$, $f(y) \in \{f_{(i)}(y)\}_{i\in[k]}$.*

If f is PC^1, we define the active set associated at $x \in X$ as

$$\mathcal{I}_f^{ess}(x) := \{i \in [k] \;:\; \forall y \in N,\; f(y) = f_{(i)}(y)\}\,. \tag{96}$$

Proposition A.17. *Let $f : X \to F$ a PC^1 function at $x \in X$. Then*

- *$\mathcal{I}_f^{ess}(x)$ is a non-empty set;*
- *f is Lipschitz continuous near x and directionally differentiable at x, with*

$$\partial_B f(x) = \{\nabla f_{(i)}(x) \;:\; i \in \mathcal{I}_f^{ess}(x)\}. \tag{97}$$

In addition, the directional derivative $f'(x;\cdot)$ is piecewise linear.

The notion of *coherent orientation* plays a central role in obtaining the invertibility of a PC^1 function in order to apply the implicit function theorem.

Definition A.18 (Coherent orientation)**.** *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open. A PC^1 function $F : U \times W \to \mathbb{R}^n$ is said* coherently oriented *with respect to x at $(x^\star, y^\star) \in U \times V$ if all matrices in $\partial_{B,x} f(x^\star, y^\star)$ have the same non-vanishing determinant sign.*

A.5. Lexicographic derivatives

Lexicographic derivatives have been introduced by Nesterov in [Nes05] as the derivatives of *lexicographically smooth* (L-smooth) functions.

Definition A.19 (Lexicographically smooth function)**.** *Let $f : X \to \mathbb{R}^m$ a locally Lipschitz continuous function on X. f is* lexicographically smooth *at x if for any $k \in \mathbb{N}$ and any $M = [m_{(1)}, \cdots, m_{(k)}] \in \mathbb{R}^{n\times k}$ the following higher-order derivatives are well defined*

$$\begin{aligned} &f^{(0)}_{x,M} : \mathbb{R}^n \to \mathbb{R}^m : d \to f'(x;d) \\ &f^{(j)}_{x,M} : \mathbb{R}^n \to \mathbb{R}^m : d \to [f^{(j-1)}_{x,M}]'(m_{(j)};d) \quad \forall j = 1, \cdots, k \end{aligned} \tag{98}$$

The class of L-smooth functions is closed under composition. C^1 functions, convex functions and PC^1 functions are all L-smooth functions.

Definition A.20 (Lexicographic derivative)**.** *Let $f : X \to \mathbb{R}^m$ a L-smooth function at x, and $M \in \mathbb{R}^{n\times n}$ a nonsingular matrix. The* lexicographic derivative *of f at x in the direction M is defined as*

$$d_L f(x; M) = \nabla f^{(n)}_{x,M}(0) \in \mathbb{R}^{m\times n}\,. \tag{99}$$

The lexicographic subdifferential *of f at x is defined as*

$$\partial_L f(x) = \{d_L f(x;N) \ : \ N \in \mathbb{R}^{n\times n} \ , \ det(N) \neq 0\} \ . \tag{100}$$

For any $k \in \mathbb{N}$, $M = [m_{(1)}, \cdots, m_{(k)}] \in \mathbb{R}^{n\times k}$ the LD-derivative *of f in the directions M is defined as*

$$f'(x;M) = \left[f^{(0)}_{x,M}(m_{(1)}), \ f^{(1)}_{x,M}(m_{(2)}), \ \cdots, f^{(k-1)}_{x,M}(m_{(k)})\right] \ . \tag{101}$$

The LD-derivative is uniquely defined for $M \in \mathbb{R}^{n\times k}$. If in addition $M \in \mathbb{R}^{n\times n}$ is square and nonsingular, it satisfies

$$f'(x;M) = d_L f(x;M)M \ . \tag{102}$$

Interestingly, the LD-derivative obeys a sharp chain-rule.

Proposition A.21 (Chain-rule)**.** *Let $h : X \to Y$ and $g : Y \to \mathbb{R}^q$ be L-smooth at $x \in X$ and $h(x) \in Y$, respectively. Then the function $g \circ h$ is L-smooth at $x \in X$, and for any $k \in \mathbb{N}$ and $M \in \mathbb{R}^{n\times k}$,*

$$[g \circ h]'(x;M) = g'(h(x); h'(x;M)) \ . \tag{103}$$

LD-derivatives give convenient calculus rules for the derivatives of min, max, abs, and other nonsmooth functions widely encountered in practice. These rules are based on *lexicographic ordering*, explaining why LD-derivatives are called "lexicographic". For given vectors $x, y \in \mathbb{R}^n$, we say that x is lexicographically lower than y if

$$\begin{aligned} &x \prec y \text{ if and only if } \exists j \in [n] \text{ such that } x_i = y_i \ \forall i < j \text{ and } x_j < y_j \ , \\ &x \preceq y \text{ if and only if } x = y \text{ or } x \prec y \ . \end{aligned} \tag{104}$$

The *lexicographic minimum function* returns the lexicographically ordered minimum of two vectors:

$$\mathbf{Lmin} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n : (x,y) \mapsto \begin{cases} x & \text{if } x \preceq y \ , \\ y & \text{if } x \succ y \ . \end{cases} \tag{105}$$

Similarly, the *lexicographic matrix minimum function* compares two matrices row by row, and is defined as

$$\mathbf{LMmin} : \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : (X,Y) \mapsto \begin{bmatrix} \mathbf{Lmin}(X_1^\top, Y_1^\top) \\ \mathbf{Lmin}(X_2^\top, Y_2^\top) \\ \cdots \\ \mathbf{Lmin}(X_m^\top, Y_m^\top) \end{bmatrix} \ . \tag{106}$$

B. Multivalued mapping

A multivalued mapping (or point-to-set map) is a mapping $\Gamma : E \rightrightarrows F$ where for each $x \in E$, $\Gamma(x)$ is a subset of F. They have been used extensively in mathematical programming to study the solution maps of optimization problems.

Definition B.1. *Let $\Gamma : E \rightrightarrows F$ a point-to-set map.*

- Γ *is* upper semicontinuous *at $x \in E$ if for each open set $\Omega \subset F$ satisfying $\Gamma(x) \subset \Omega$ there exists a neighborhood $N(x)$ of x such that for all $y \in N(x)$, $\Gamma(y) \subset \Omega$.*

- Γ *is* lower semicontinuous *at $x \in E$ if for each open set $\Omega \subset F$ satisfying $\Gamma(x) \cap \Omega \neq \varnothing$ there exists a neighborhood $N(x)$ of x such that for all $y \in N(x)$, $\Gamma(y) \cap \Omega \neq \varnothing$.*

- Γ *is* continuous *at x if it is lower semicontinuous and upper semicontinuous at x.*

Definition B.2. *The point-to-set map $\Gamma : E \rightrightarrows F$ is* closed *at x if there exists a sequence $x_n \in E$, $x_n \to x$ such that $y_n \in \Gamma(x_n)$ and $y_n \to y$ imply $y \in \Gamma(x)$.*

Definition B.3. *The point-to-set map $\Gamma : E \rightrightarrows F$ is* open *at x if $y \in \Gamma(x)$ implies there exists a sequence $x_n \in E$, $x_n \to x$ such that there exists m and $\{y_n\}$ such that $y_n \in \Gamma(x_n)$ for all $n \geq m$ and $y_n \to y$.*

Definition B.4. *The point-to-set map $\Gamma : E \rightrightarrows F$ is* uniformly compact *near $\hat{x}$ if the set $\cup_{x \in N(\hat{x})} \Gamma(x)$ is bounded for some neighborhood $N(\hat{x})$ of $\hat{x}$.*

If Γ is uniformly compact near $\hat{x}$, then Γ is closed if and onl if $\Gamma(\hat{x})$ is a compact set and Γ is upper semicontinuous at $\hat{x}$.

Definition B.5. *The point-to-set map $\Gamma : E \rightrightarrows F$ is* upper Lipschitzian *with modulus L at $\hat{x} \in E$ if there is a neighborhood N of $\hat{x}$ such that for each $x \in N$, $\Gamma(x) \subset \Gamma(\hat{x}) + L\|x - \hat{x}\|B$, with B the unit-ball in F.*

C. A primer on generalized equations

Let X be a linear space, and C be a non-empty closed convex set in X.

Definition C.1 (Normal cone)**.** *We define the* normal cone *operator associated to the set C as, for $x \in X$*

$$N_C(x) = \begin{cases} \{y \in X^\star \ : \ \langle y, v - x \rangle \leq 0 \ , \ \forall v \in C\} & \textit{if } x \in C \ , \\ \varnothing \ , & \textit{if } x \notin C \ . \end{cases} \tag{107}$$

With the normal cone, we can define formally the notion of *variational inequality*, generalizing optimization problem.

Definition C.2 (Variational inequality)**.** *Let $p \in P$ be a parameter. For $f : X \times P \to X^\star$, we define the* paramaterized variational inequality *(also called* generalized equation*) the problem of finding $x \in X$ such that*

$$f(x,p) + N_C(x) \ni 0 \ . \tag{108}$$

The equation (108) is a compact formulation stating that $-f(x,p) \in N_C(x)$, or, equivalently

$$\langle f(x,p), v - x \rangle \geq 0 \ , \quad \forall v \in C \ . \tag{109}$$

The solution mapping associated to (108) is defined as

$$S(p) = \{x \in X \ : \ f(x,p) + N_C(x) \ni 0\} \ . \tag{110}$$

Robinson extended the Implicit function theorem to the nonsmooth variational inequality (108), by extending the notion of *regular solution* to the variational inequality setting, where the solution mapping $S(p)$ is not necessarily single-valued. Robinson introduced the notion of *strong regularity*, which extends the non-singularity condition we impose in the implicit function theorem. If we suppose that f is (Fréchet) differentiable, the idea is to look at the linearized generalized equation around a solution x_0

$$T(x,p) \ni 0 \,, \tag{111}$$

where we have defined the linearized operator at x_0: $T(x,p) := f(x_0,p) + \nabla_x f(x_0,p)(x - x_0) + N_C(x)$.

Definition C.3 (Strong regularity [Rob80]). *We say that* (108) *is* strongly regular *at a solution x_0 with Lipschitz constant L if there exists a neighborhood $U \subset X^\star$ of the origin and $V \subset X$ of x_0 such that the restriction to U of $T^{-1} \cap V$ is a* single-valued *function from U to V and is L-Lipschitzian on U.*

Putting it more simply, strong regularity implies that the inverse of the linearized operator T^{-1} has a Lipschitz continuous *single-valued* localization at 0 for x_0. If C is the whole-space X, then strong regularity implies that the Jacobian $\nabla_x f(x_0,p)$ is non-singular. The following theorem extends the implicit function theorem to generalized equations (108) satisfying the strong regularity condition C.3.

Theorem C.4 (Theorem 2.1, [Rob80]). *Let $p \in P$ be a given parameter, x_0 solution of the generalized equation* (108)*. Suppose that f is Fréchet-differentiable w.r.t. x exists at p, and that both $f(\cdot,\cdot)$ and $\nabla_x f(\cdot,\cdot)$ are continuous at (x_0,p). If* (108) *is* strongly regular *at x_0, then for any $\varepsilon > 0$ there exists neighborhood N of p and W of x_0 as well as a single-valued function $x : N \to W$ sucht that for any $p \in N$, $x(p)$ is the unique solution of the inclusion*

$$f(x,p) + N_C(x) \ni 0 \,, \tag{112}$$

and, for all $p, q \in N$, one has

$$\|x(p) - x(q)\| \leq (L + \varepsilon)\|f(x(p),p) - f(x(q),q)\| \,. \tag{113}$$

Theorem C.4 states that the strong regularity of (108) implies that the solution mapping $S(p)$ of (108) has a *Lipschitz-continuous* single-valued localization at p for x_0.

Corollary C.5 (Locally-Lipschitz condition). *Suppose the hypotheses of Theorem C.4 hold. Suppose in addition there exists a constant κ such that for each $(p,q) \in N$, for each $x \in W$, one has*

$$\|f(x,p) - f(x,q)\| \leq \kappa\|p - q\| \,. \tag{114}$$

Then $x(\cdot)$ is Lipschitzian on N with modulus $\kappa(L + \varepsilon)$.