Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

ANTON SCHIELA

A continuity result for Nemyckii Operators and some applications in PDE constrained optimal control. [1]

[1]Supported by the DFG Research Center MATHEON "Mathematics for key technologies"

ZIB-Report 06-41 (August 2006)

A continuity result for Nemyckii Operators and some applications in PDE constrained optimal control. †

Anton Schiela

August 3, 2006

Abstract

This work explores two applications of a classical result on the continuity of Nemyckii operators to optimal control with PDEs. First, we present an alternative approach to the analysis of Newton's method for function space problems involving semi-smooth Nemyckii operators. A concise proof for superlinear convergence is presented, and sharpened bounds on the rate of convergence are derived. Second, we derive second order sufficient conditions for problems, where the underlying PDE has poor regularity properties. We point out that the analytical structure in both topics is essentially the same.

AMS MSC 2000: 49M15, 49K20, 46N40

Keywords:continuity of Nemyckii Operators, Newton methods in function space, optimal control, second order sufficient conditions

1 Introduction

Many nonlinear problems in function space are formulated in terms of Nemyckii operators. This class of operators plays a role, if the nonlinearity is defined by a pointwise nonlinear function. Due to their practical importance Nemyckii operators have been analysed thoroughly, and many standard results have been established (for a thorough exposition see [1]). The purpose of this note is to emphasize two applications of one of these results to PDE constrained optimal control: we revisit the convergence theory of semi-smooth Newton methods and derive a variant of second order sufficient conditions for a local minimizer subject to a nonlinear optimal control problem.

Newton's method is a standard method for solving nonlinear systems of equations, both in finite dimensional and infinite dimensional normed spaces. Classical assumptions in the analysis of Newton's method are Fréchet differentiability of F and Lipschitz continuity of F' in a neighbourhood of a solution x_*, together with

†Supported by the DFG Research Center MATHEON "Mathematics for key technologies"

invertibility of F'. Then local quadratic convergence of Newton's method can be shown.

In infinite dimensional function spaces the requirement of Fréchet differentiability is often too strong, and in the last few years many efforts have been undertaken to show superlinear convergence of Newton's method under weakened smoothness assumptions (cf. e.g. [6, 10, 4, 9]). As a result the concept of semi-smoothness was carried over to infinite dimensional function spaces, which still guarantees superlinear convergence. In [9] the semi-smoothness of Nemyckii operators was analysed, and convergence rates for Newton methods were derived in the presence of some smoothing operator. The corresponding proofs (cf. also [4]) rely on sophisticated splittings of the domain of definition into several subdomains and different estimates there.

In this note we present an alternative approach to the semi-smoothness of Nemyckii operators. The idea is to apply a standard result on the continuity of Nemyckii operators to appropriately chosen remainder terms. This, and simple application of the Hölder inequality already yield the main theorems on semi-smooth Newton methods. Our intention is to clarify the underlying analytical structure, and to make the theory of semi-smooth Newton methods accesible more easily.

Especially from an algorithmic point of view on Newton's method it is often favourable to state invertibility and smoothness assumptions as one affine invariant assumption (cf. [2]). An affine invariant analysis of semi-smooth Newton methods in function space is possible and has been performed in [7]. Since we concentrate on the analysis of nonlinear Nemyckii operators, we will not use an affine invariant framework here.

Closely related to the convergence theory of Newton's method as an optimization algorithm is the analysis of second order sufficient conditions for a minimizer. In essence, a nonlinear nonconvex problem behaves locally like a convex problem near a point, where those conditions hold. This justifies the use of Newton's method in a neighbourhood of this point. In the context of optimal control problems a so called two-norm discrepancy is observed. This adds some additional complexity to the analysis. Our aim in this note is to point out the connection between the two-norm discrepancy, the continuity of Nemyckii operators, and a smoothing property of the underlying PDE. For this purpose we derive a prototype theorem on minimizers with respect to an L_2-neighbourhood.

Our paper is structured as follows. In Section 2 we consider the continuity of Nemyckii operators and apply these results to derive estimates on remainder terms. In Section 3 we consider semi-smooth Newton methods for a class of nonlinear problems in function space. There we give a simple proof for superlinear convergence, alternative to the ones in [4, 9] and derive sharpened estimates for convergence rates. For illustration we consider in Section 3.2 the direct solution of an optimality system in function space by a semi-smooth Newton method. In Section 4 we study second order sufficient conditions for the case that the state equation only admits poor regularity estimates.

2 Nemyckii-Operators and Remainder Terms in $L_q(\Omega)$

Consider a bounded domain $\Omega \subset \mathbb{R}^d$, two Banach spaces $\mathcal{X}, \mathcal{Y}$, and a function

$$\begin{aligned} f : \mathcal{X} \times \Omega &\to \mathcal{Y} \\ (x,t) &\mapsto f(x,t). \end{aligned}$$

Let $L_p(\Omega, \mathcal{X})$ be the space of p-integrable functions $x : \Omega \to \mathcal{X}$, and $L_s(\Omega, \mathcal{Y})$ be defined analogously. For $\mathcal{Y} = \mathbb{R}$ we will write $L_s(\Omega)$, and if it is clear from the context, which domain, and image spaces are meant, we will simply use the notation L_s.

Assume $D \subset L_p(\Omega, \mathcal{X})$. If it is well defined, then the operator

$$\begin{aligned} F : D &\to L_s(\Omega, \mathcal{Y}) \\ x &\mapsto F(x) : F(x)(t) = f(x(t), t) \qquad \text{a.e.} \end{aligned}$$

is called a *Nemyckii* operator from D to $L_s(\Omega, \mathcal{Y})$ corresponding to f. To be well defined, F must necessarily map measurable functions to measurable functions. Usually, one assumes that f is a *Carathéodory* function, i.e., f is continuous in x and measurable in t. However, since pointwise limits of measurable functions are measurable, one can extend this class substantially by requiring that f is a *pointwise limit* of a sequence of Carathéodory functions. This is the class of *Baire-Carathéodory*-functions (cf. [1]).

In addition one has to require $F(x) \in L_s(\Omega, \mathcal{Y})$ for $x \in D$. This is usually achieved by imposing a *growth condition* on f, which reads (cf. [12, Section 26.3]):

$$f(x,t) \le a(t) + b\,|x|_{\mathcal{X}}^{p/s} \qquad a \in L_s(\Omega, \mathcal{Y}), b \in \mathcal{Y}. \tag{1}$$

Hence, the behaviour of f for large x restricts the choice of spaces, where a corresponding Nemyckii operator F can be defined.

The following well-known lemma on the continuity of Nemyckii operators is in the center of our considerations. Usually, this lemma is formulated in terms of Carathéodory functions. Its extension to Baire-Carathéodory functions is straightforward and very useful in the analysis that follows.

Lemma 2.1 (Continuity of Nemyckii Operators). *Let $\mathcal{X}$, $\mathcal{Y}$ be Banach spaces, Ω a Borel subset of $\mathbb{R}^d$, and $f : \Omega \times \mathcal{X} \to \mathcal{Y}$ a Baire-Carathéodory function. For each measurable function $x : \Omega \to \mathcal{X}$ let $F(x)$ be the measurable function $t \to f(x(t), t) \in \mathcal{Y}$. Let $x_* \in L_p(\Omega, \mathcal{X})$ be given. Then the following assertion holds:*

If f is continuous at $(x_*(t), t)$ for almost all $t \in \Omega$, and F maps $L_p(\Omega, \mathcal{X})$ into $L_s(\Omega, \mathcal{Y})$ for $1 \le p, s < \infty$, then F is continuous at x_* in the norm topology.

Proof. This is a slight generalization of a well known lemma of Krasnoselski, that can be found for example in [12, Proposition 26.7(a)]. The proof of this generalization is literaly the same. For convenience we give a simplified proof for the special case of f being uniformly bounded, and Ω being a bounded region in $\mathbb{R}^d$.

By construction of Baire-Carathéodory functions, F maps measurable functions to measurable functions. To show continuity at x_* for $s<\infty$ we consider an arbitrary sequence $\|x_n - x_*\|_{L_p(\Omega,\mathcal{X})} \to 0$, which converges pointwise almost everywhere. Define the function

$$w(x,t) := |f(x(t),t) - f(x_*(t),t)|_{\mathcal{Y}}^s .$$

Since $F : L_p(\Omega,\mathcal{X}) \to L_s(\Omega,\mathcal{Y})$ we have $W(x) : t \mapsto w(x(t),t) \in L_1(\Omega)$ for all $x \in L_p(\Omega,\mathcal{X})$. Inserting the sequence x_n, we conclude that $W(x_n) \to 0$ pointwise a.e. due to the continuity of $f(x,t)$ at $(x_*(t),t)$. Moreover, since $|f| < C$ and Ω is bounded, the sequence $W(x_n)$ is majorized by an integrable (constant) function. Thus we can apply the convergence theorem of Lebesgue to obtain

$$\lim_{n\to\infty} \int_\Omega w(x_n(t),t)\,dt = 0,$$

and thus continuity of the operator $F : L_p(\Omega,\mathcal{X}) \to L_s(\Omega,\mathcal{Y})$ at x_*. □

Remark 2.2. It is easy to see that Lemma 2.1 cannot hold for $p < s = \infty$, except for the case of constant f. The case $p = s = \infty$ is very different in character. Then one needs *uniform* continuity of f in a neighbourhood of x_* to show continuity of F at x_*. Moreover, no growth condition on f is required.

We turn now to Nemyckii operators with a special structure. For given $\beta > 0$ we consider functions $r(\delta x, t)$ of the form

$$r : \mathcal{X} \times \Omega \to \mathbb{R}$$
$$r(\delta x, t) = \psi(\delta x, t; \beta)\, |\delta x|_{\mathcal{X}}^\beta ,$$

with $\psi(\delta x, t; \beta)$ being a Baire-Carathéodory function. Usually, such functions and their corresponding Nemyckii operators R appear in approximation results as remainder terms. The following simple lemma is central for our considerations. We denote by Ψ the Nemyckii operator corresponding to ψ.

Lemma 2.3. *Assume that $\Psi(\delta x;\beta) \in L_s(\Omega)$. Then for $\beta \geq 0$ the following estimate holds for $q^{-1} = s^{-1} + \beta p^{-1}$:*

$$\|R(\delta x)\|_{L_q(\Omega)} \leq \|\Psi(\delta x;\beta)\|_{L_s(\Omega)}\, \|\delta x\|_{L_p(\Omega,\mathcal{X})}^\beta \tag{2}$$

Proof. This is a simple application of the Hölder inequality. □

If ψ is continuous at $\delta x = 0$, then we can improve our estimates, using Lemma 2.1. Here we denote by $\omega(\tau)$ a function with $\lim_{\tau\to 0}\omega(\tau) = 0$.

Lemma 2.4. *Assume that almost everywhere in Ω $\psi(\delta x, t; \beta)$ is continuous at $x = 0$ with $\psi(0,t;\beta) = 0$. Further, assume that $\Psi(\cdot;\beta) : L_{\tilde{p}}(\Omega,\mathcal{X}) \to L_s(\Omega)$ with $s < \infty$. Then for $q^{-1} = s^{-1} + \beta p^{-1}$:*

$$\|R(\delta x)\|_{L_q(\Omega)} = \omega(\|\delta x\|_{L_{\tilde{p}}(\Omega,\mathcal{X})})\, \|\delta x\|_{L_p(\Omega,\mathcal{X})}^\beta \tag{3}$$

Proof. By continuity of ψ at 0, and $\Psi(\cdot;\beta) : L_{\tilde{p}}(\Omega,\mathcal{X}) \to L_s(\Omega)$ we deduce that Ψ is continuous at 0 as a Nemyckii operator. Application of (2) yields the result. □

3 Semi-Smooth Newton Methods

Let X be a reflexive Banach space that is continously embedded in $L_2(\Omega)$. Consider the equation

$$G(x) = 0, \text{ where } G(x) := \mathcal{D}\, x + F(x) \tag{4}$$

with $\mathcal{D} : X \to X^*$ being a linear differential operator, and F a nonlinear Nemyckii operator, corresponding to a nonlinear function $f(x,t)$ that satisfies suitable growth conditions. This formulation also includes systems of partial differential equations and algebraic equations, which often arise in optimal control.

We assume that we can construct a family of operators $F'(x)$, depending on x only, such that the "linearized" equation $G'(x)\delta x := (\mathcal{D} + F'(x))\delta x = l$ is uniquely solvable for $l \in X^*$, and has a *smoothing property*:

$$\left\| G'(x)^{-1} \right\|_{L_q(\Omega,\mathcal{Y}) \to L_p(\Omega,\mathcal{X})} \leq C \tag{5}$$

uniformly in a neighbourhood of the solution x_*. This is often the case, if there is a continuous embedding $X \hookrightarrow L_p(\Omega)$ for $p > 2$ and $p^{-1} + q^{-1} = 1$, such as a Sobolev-type embedding.

If f is differentiable pointwise, then the canonical choice for $F'(x)\delta x$ is the Nemyckii operator corresponding to $f'(x(t),t)\,\delta x$. However, in the context of semi-smoothness, f does not necessarily have to be differentiable pointwise in order to construct a useful Newton method. Then F' is set to a Nemyckii operator corresponding to some function $\tilde{f}'$. Following the discussion in [4] ambiguities are resolved by setting $\tilde{f}'$ to some fixed value in regions, where the pointwise derivative f' is not well defined. An illustration for this is given in Section 3.2.

For the analysis of Newton's method we observe that due to the linearity of $\mathcal{D}$

$$G'(x+\delta x)\delta x - (G(x+\delta x) - G(x)) = F'(x+\delta x)\delta x - (F(x+\delta x) - F(x)) \tag{6}$$

is a Nemyckii operator. Thus we may consider the following pointwise remainder term at some point $x \in \mathcal{X}$:

$$\begin{aligned} r_x(\delta x, t) &:= \left| f'(x+\delta x,t)\delta x - (f(x+\delta x,t) - f(x,t)) \right|_{\mathcal{Y}} \\ &= \underbrace{\left(\frac{|f'(x+\delta x,t)\delta x - (f(x+\delta x,t) - f(x,t))|_{\mathcal{Y}}}{|\delta x|_{\mathcal{X}}^{\beta}} \right)}_{\psi_x(\delta x,t;\beta)} |\delta x|_{\mathcal{X}}^{\beta} \end{aligned} \tag{7}$$

for $1 \leq \beta < \infty$, and define for some given $x \in L_p(\Omega,\mathcal{X})$ the Nemyckii operator $\Psi_x(\cdot;\beta) : L_p(\Omega,\mathcal{X}) \to L_s(\Omega)$ by

$$\Psi_x(\delta x;\beta)(t) = \psi_{x(t)}(\delta x(t),t;\beta),$$

and analogously the Nemyckii operator $R_x(\delta x) : L_p(\Omega,\mathcal{X}) \to L_q(\Omega)$. Again, we assume throughout this section that $\psi_{x(t)}(\cdot,\cdot;\beta)$ is a Baire-Carathéodory function and thus also r_x.

Remark 3.1. The remainder term (6) is used in [4] to define *slant* or *Newton* differentiability of G at x. In contrast to remainder terms in the definition of Fréchet differentiability, a derivative G' is considered at $x+\delta x$, rather than at x and may be non-unique. As already pointed out in [6, 4], (6) is well suited for the analysis of Newton's method.

Application of Lemma 2.4 and Lemma 2.3 yields the following convergence theorem for Newton's method. We stress that the existence of a solution x_* is assumed. Showing existence of a solution x_* via Newton's method would require stronger smoothness assumptions.

Theorem 3.2 (Superlinear convergence of Newton's method). *Assume that $x_* \in X$ solves the equation $G(x)=0$, and suppose that (5) holds for some $1 \leq q < p \leq \infty$ uniformly in a neighbourhood of x_*. Let in the following $\psi_{x_*(t)}$ defined as in (7) be a Baire-Carathéodory function. Then the following two assertions are valid:*

(i) Assume that almost everywhere in Ω $\psi_{x_*(t)}(\delta x, t; 1)$ is continuous at $\delta x = 0$ with $\psi_{x_*(t)}(0, t; 1) = 0$. Further, assume that $\Psi_{x_*}(\cdot; 1) : L_p(\Omega, \mathcal{X}) \to L_s(\Omega)$ with $s < \infty$. Then for $q^{-1} = s^{-1} + p^{-1}$ Newton's method converges locally superlinearly to x_* in the normed space $(X, L_p(\Omega, \mathcal{X}))$.

(ii) If $\Psi_{x_*}(\cdot; \beta) : L_p(\Omega, \mathcal{X}) \to L_s(\Omega)$ for $\beta > 1$ is uniformly bounded as an operator, then Newton's method converges locally superlinearly to x_* in the normed space $(X, L_p(\Omega, \mathcal{X}))$ with the rate β, as long as $q^{-1} = s^{-1} + \beta p^{-1}$.

Proof. Consider a starting value $x := x_* + \delta x$, then one Newton step $x \to x_+$ yields

$$\begin{aligned} x_+ - x_* &= -G'(x)^{-1}G(x) + (x - x_*) \\ &= G'(x)^{-1}(G'(x)\delta x - (G(x) - G(x_*))) \\ &= G'(x)^{-1}(F'(x_* + \delta x)\delta x - (F(x_* + \delta x) - F(x_*))). \end{aligned}$$

We denote by R_{x_*} the Nemyckii-operator corresponding to

$$r_{x_*}(\delta x, t) = \left| f'(x_* + \delta x, t)\delta x - (f(x_* + \delta x, t) - f(x_*, t)) \right|_{\mathcal{Y}}$$

and we obtain by (5)

$$\|x_+ - x_*\|_{L_p} \leq \left\| G'(x)^{-1} \right\|_{L_q \to L_p} \|R_{x_*}(\delta x)\|_{L_q}.$$

Consider now part (i) of our assertions. Recalling the definition of $\psi_{x_*}(\delta x, t; 1)$ in (7) and using its pointwise continuity, Lemma 2.4 yields

$$\|x_+ - x_*\|_{L_p} \leq C \|R_{x_*}(\delta x)\|_{L_q} = \omega(\|x - x_*\|_{L_p}) \|x - x_*\|_{L_p}.$$

This is just the definition of local superlinear convergence.

Assertion (ii) follows by Lemma 2.3 applied to $\psi_{x_*}(\delta x, t; \beta)$, which shows

$$\|x_+ - x_*\|_{L_p} \leq C \|x - x_*\|_{L_p}^{\beta}.$$

□

Remark 3.3. Obviously, pointwise continuity of ψ_{x_*} at x_* is the essential requirement for Newton's method to converge superlinearly. This requirement can be checked easily for a given nonlinear function f. As an example the $\max(0, x)$-function is considered in Section 3.2.

To conclude the continuity of Ψ_{x_*} from the continuity of ψ_{x_*} the well-known $L_p - L_q$ *norm-gap* appears. This happens, even if f is pointwise differentiable and is a consequence of the *failure* of Lemma 2.1 for $p < \infty$ and $s = \infty$. To overcome this difficulty we have to assume the *smoothing property* (5).

3.1 Small Sets of High Nonlinearity

In many cases the boundedness condition $\Psi_{x_*}(\delta x; \alpha) \in L_s(\Omega)$ is formulated in terms of subsets of Ω where $\Psi_{x_*}(\delta x; \alpha)$ is large. Similarly to [9] we can define the set

$$\Omega_\varepsilon = \Omega_\varepsilon(\delta x) := \left\{ t \in \Omega : \Psi_{x_*}(\delta x; \alpha)(t) > \varepsilon^{1-\alpha} \right\}$$

for $\alpha > 1$, together with the assumption that there is $\rho > 0$ and $C < \infty$, such that for the Lebesque measure $|\Omega_\varepsilon|$ of Ω_ε the following bounds hold:

$$\sup_{\|\delta x\|_{L_p} \leq \rho} |\Omega_\varepsilon(\delta x)| \leq C\varepsilon^\gamma. \tag{8}$$

The qualitative meaning of this assumption is that the sets of high nonlinearity are small near x_*. We will give an illustration of this assumption in Section 3.2. For the sake of mathematical exactness we remark that since $\delta x \in L_p(\Omega, \mathcal{X})$, $\Omega_\varepsilon(\delta x)$ is only well defined up to a set of zero measure, but this does clearly not affect the assumption (8).

Relations like (8) can be described conveniently in terms of the *distribution function* $\mathcal{S}_f$ of a positive function f, defined by

$$\mathcal{S}_f(e) := |\{t \in \Omega : f(t) > e\}| \, .$$

The distribution function measures the size of the sets, where f is large (cf. Figure 1). Obviously, $\mathcal{S}$ is positive, monotonically decreasing, bounded on bounded domains, and has bounded support if f is bounded.

Lemma 3.4 (The Distribution Function). *Let Ω be a σ-finite measure space and f: $\Omega \to [0, \infty]$ measurable. Then*

$$\int_\Omega f(t)dt = \int_0^\infty \mathcal{S}_f(e)\, de. \tag{9}$$

Let $\varphi : [0, \infty] \to [0, \infty]$ be locally absolutely continuous, strictly monotone (increasing or decreasing), and bijective. Then

$$\int_\Omega f(t)dt = \int_0^\infty \mathcal{S}_f(\varphi(e))|\varphi'(e)|\, de. \tag{10}$$

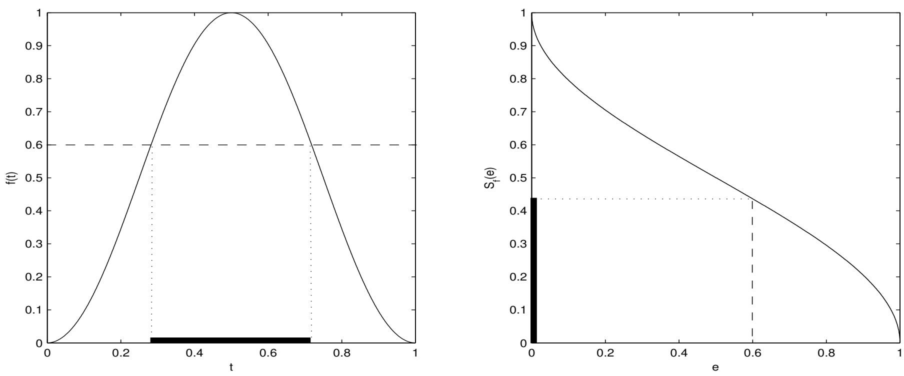

Figure 1: A function $f(t)$ (left) and its distribution function $\mathcal{S}_f(e)$ (right). The value of the distribution function at e is given by the area on which f is above e.

Proof. Equation (9) is a special case of [11, Theorem 8.8]. Equation (10) follows from the substitution rule (cf. [11, Theorem 8.1]), which shows that with $\tilde{e} = \varphi(e)$:

$$\int_0^\infty \mathcal{S}_f(\tilde{e})\, d\tilde{e} = \int_0^\infty \mathcal{S}_f(\varphi(e))|\varphi'(e)|\, de.$$

□

In terms of the distribution function (8) reads

$$\mathcal{S}_{\Psi_{x_*}(\delta x;\alpha)}\left(\varepsilon^{1-\alpha}\right) \leq C \min\left\{|\Omega|, \varepsilon^\gamma\right\}, \tag{11}$$

with a constant C that does not depend on δx, as long as $\|\delta x\|_{L_p} \leq \rho$ for some $\rho > 0$.

Lemma 3.5. *If* (8) *holds, then* $\Psi_{x_*}(\delta x;\alpha) \in L_s(\Omega)$ *for all* $1 \leq \alpha < 1 + \gamma s^{-1}$.

Proof. Setting $\psi := \Psi_{x_*}(\delta x;\alpha)$ we deduce

$$\|\psi\|_{L_s}^s = \|\psi^s\|_{L_1} = \left\|\mathcal{S}_{\psi^s}\left(\varepsilon^{1-\alpha}\right)\right\|_{L_1} = \left\|\mathcal{S}_\psi\left(\varepsilon^{(1-\alpha)s}\right)\right\|_{L_1}.$$

Inserting (11) and using (10) we obtain

$$\begin{aligned}\left\|\mathcal{S}_\psi\left(\varepsilon^{(1-\alpha)s}\right)\right\|_{L_1} &\leq C \int_0^\infty \min\left\{|\Omega|, \varepsilon^\gamma\right\} \varepsilon^{(1-\alpha)s-1} d\varepsilon \qquad (12)\\ &\leq C \min\left\{|\Omega|, \varepsilon^\gamma\right\} \varepsilon^{(1-\alpha)s}\Big|_0^\infty,\end{aligned}$$

which is bounded, if $\gamma + (1-\alpha)s > 0$ and $\alpha \geq 1$. □

We observe that the limiting case $\alpha = 1 + \gamma s^{-1}$ cannot be achieved, since $1/x \notin L_1([0,c])$. Combining Lemma 3.5 and Theorem 3.2 we obtain:

Theorem 3.6. *Let the general assumptions of Theorem 3.2 hold. Assume that* (8) *holds for some*

$$1 < \alpha < \frac{1+\gamma q^{-1}}{1+\gamma p^{-1}}. \tag{13}$$

Then Newton's method converges locally superlinearly in $(X, \|\cdot\|_{L_p(\Omega,\mathcal{X})})$ *with the rate* $\beta = \alpha$.

Proof. To be able to apply Theorem 3.2(ii) we have to assert that $\Psi_{x_*}(\delta x;\beta) \in L_s(\Omega)$ for $q^{-1} = s^{-1} + \beta p^{-1}$. By Lemma 3.5 this is true for $\beta = \alpha < 1 + \gamma s^{-1}$. Elimination of s^{-1} yields the restriction (13). □

Under the given restrictions on α Theorem 3.6 gives us a result as good as we can expect. If we additionally assume that r_{x_*} is globally Lipschitz continuous, then we may also consider the case $\alpha > \frac{1+\gamma q^{-1}}{1+\gamma p^{-1}}$. For this we have to split Ω into two parts, a technique that is simliar to the one used in [9], or [4]. However, as it will turn out, inspite of a considerably increased technical effort, there seems no substantial improvement on β possible: we merely obtain the closure of the open interval in (13) as a restriction on the rate of convergence.

Theorem 3.7. *Let the general assumptions of Theorem 3.2 hold. Assume that* (8) *holds for some*

$$\frac{1+\gamma q^{-1}}{1+\gamma p^{-1}} < \alpha \le \frac{p}{q}. \tag{14}$$

If additionally $\Psi_{x_*}(\delta x;1)$ *is uniformly bounded in* L_∞ *(i.e.* $r_{x_*}(\delta x,t)$ *is uniformly Lipschitz continuous), then we obtain superlinear convergence with the rate*

$$\beta = \frac{1+\gamma q^{-1}}{1+\gamma p^{-1}}. \tag{15}$$

Proof. Depending on a parameter κ we divide Ω into two parts. In view of (8) we call Ω_κ the set, where $\Psi_{x_*}(\delta x;\alpha) > \kappa^{1-\alpha}$ and obtain $|\Omega_\kappa| \le C\kappa^\gamma$. On Ω_κ we use the uniform boundedness of $\Psi_{x_*}(\delta x;1)$ in L_∞ to estimate

$$\|R_{x_*}(\delta x)\|_{L_q(\Omega_\kappa)} = \|\Psi_{x_*}(\delta x;1)\|_{L_\infty(\Omega)} |\Omega_\kappa|^{(q^{-1}-p^{-1})} \|\delta x\|_{L_p} \le C\kappa^{\gamma(q^{-1}-p^{-1})} \|\delta x\|_{L_p}.$$

Setting $\kappa := \|\delta x\|_{L_p}^{\Theta}$ (with Θ to be chosen later) this yields:

$$\|R_{x_*}(\delta x)\|_{L_q(\Omega_\kappa)} \le C \|\delta x\|_{L_p}^{\Theta\gamma(q^{-1}-p^{-1})+1}. \tag{16}$$

On the remaining set $\Omega \setminus \Omega_\kappa$ we apply, as before, the Hölder inequality

$$\|R_{x_*}(\delta x)\|_{L_q(\Omega\setminus\Omega_\kappa)} \le \|\Psi_{x_*}(\delta x;\alpha)\|_{L_s(\Omega\setminus\Omega_\kappa)} \|\delta x\|_{L_p}^{\alpha}$$

with $s^{-1} := q^{-1} - \alpha p^{-1}$. Since $\Psi_{x_*}(\delta x;\alpha)$ is bounded on $\Omega \setminus \Omega_\kappa$ by construction, we show just as in the proof of Lemma 3.5 (use κ instead of 0 as the lower bound in the integral (12)) that for $0 \le s^{-1} \le 1$, and $\gamma s^{-1} + 1 - \alpha < 0$

$$\|\Psi_{x_*}(\delta x;\alpha)\|_{L_s(\Omega\setminus\Omega_\kappa)} \le C\kappa^{\gamma s^{-1}+1-\alpha}.$$

Here hypothesis (14) ensures that the above inequalities for s^{-1} are valid. Hence, inserting the proper values for κ and s^{-1} we obtain

$$\|R_{x_*}(\delta x)\|_{L_q(\Omega\setminus\Omega_\kappa)} \leq C\,\|\delta x\|_{L_p}^{\Theta(\gamma s^{-1}+1-\alpha)+\alpha} \leq C\,\|\delta x\|_{L_p}^{\Theta\gamma(q^{-1}-\alpha p^{-1})-\Theta(\alpha-1)+\alpha}\,. \quad (17)$$

Adding both parts we have

$$\|R_{x_*}(\delta x)\|_{L_q(\Omega)}^q = \|R_{x_*}(\delta x)\|_{L_q(\Omega_\kappa)}^q + \|R_{x_*}(\delta x)\|_{L_q(\Omega\setminus\Omega_\kappa)}^q\,,$$

and a choice of Θ that balances (16) and (17) will provide the sharpest results. Thus, comparing the exponents in (16) and (17), we choose Θ such that

$$\Theta\gamma(q^{-1}-p^{-1})+1 = \Theta\gamma(q^{-1}-\alpha p^{-1}) - \Theta(\alpha-1)+\alpha \quad (18)$$

which yields $\Theta = (\gamma p^{-1}+1)^{-1}$. Note that α cancels out in solving (18) for Θ. Inserting Θ into (16) we finally obtain (15). $\square$

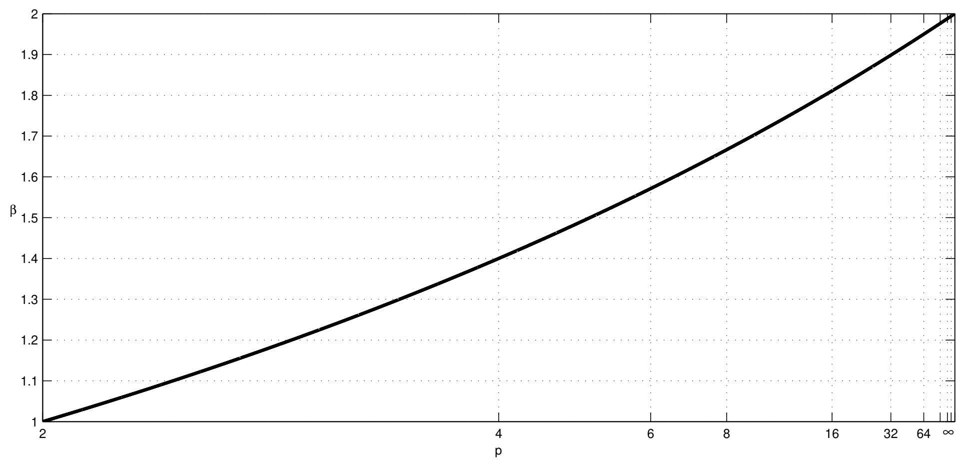

Figure 2: Convergence rates β according to Theorem 3.7 plotted for $p \in [2,\infty]$, $q^{-1}+p^{-1}=1$, and $\gamma=1$.

In [9, Example 3.53] an example was constructed that shows that the rate in Theorem 3.7 can be considered sharp. For completeness we note that in the remaining case $\alpha = \frac{1+\gamma q^{-1}}{1+\gamma p^{-1}}$ the same techniques as in the proof of Theorem 3.7 lead to the estimate

$$\|R_{x_*}(\delta x)\|_{L_q} \leq C\,|\ln(\|\delta x\|_{L_p})|\,\|\delta x\|_{L_p}^\alpha\,.$$

Thus, we obtain convergence that lags behind the rate α by a logarithmic factor.

3.2 Application to an Optimal Control Problem

As an illustration we consider the simple optimal control problem:

$$\begin{aligned}\min \frac{1}{2}\|y-y_d\|_{L_2(\Omega)}^2 &+ \frac{1}{2}\|u\|_{L_2(\Omega)}^2\\ \text{s.t.}\ -\Delta y - u = 0, &\quad y\big|_{\partial\Omega=0}, \qquad u \geq 0.\end{aligned}$$

To compute (u, y) one usually introduces an adjoint state λ and considers the solution of the first order necessary conditions:

$$\begin{aligned} y - y_d - \Delta\lambda &= 0, \quad \lambda\big|_{\partial\Omega=0}, \\ -\Delta y - \max(\lambda, 0) &= 0, \quad y\big|_{\partial\Omega=0} \end{aligned} \tag{19}$$

with $u = \max(\lambda, 0)$. This is a nonlinear system of equations $G(x) = 0$ of the form (4) with $x = (y, \lambda)$. Its direct solution in function space by Newton's method has been considered in [4, 9]. We revisit this example using the results gained in the previous sections.

Choosing a Linearization. The nonlinearity in this system is given by the function $m(\lambda) := \max(\lambda, 0)$, for which we define a pointwise linearization

$$m'(\lambda) := \begin{cases} 1 & : \quad \lambda > 0 \\ 0 & : \quad \lambda \leq 0. \end{cases}$$

At the nondifferentiable point $\lambda = 0$ the definition $m'(0) := 0$ resolves the non-uniqueness of the linearization there. Any other fixed value would also have been possible. With this definition we can construct a Jacobian matrix $G'(x)$ to (19), given by

$$G'(x) = \begin{pmatrix} I & -\Delta \\ -\Delta & -m'(\lambda) \end{pmatrix}.$$

Analysis of the Remainder Terms. Let $x_* = (y_*, \lambda_*)$ be the solution of (19). Computation of the remainder term yields

$$G'(x_* + \delta x)\delta x - (G(x_* + \delta x) - G(x_*)) = \begin{pmatrix} 0 \\ R_{\lambda_*}(\delta\lambda). \end{pmatrix}$$

Here $R_{\lambda_*}(\delta\lambda)$ is a Nemyckii operator corresponding to

$$r_{\lambda_*}(\delta\lambda) := m'(\lambda_* + \delta\lambda)\delta\lambda - (m(\lambda_* + \delta\lambda) - m(\lambda_*)).$$

We observe that for each $\lambda_* \in \mathbb{R}$ there is an $\varepsilon(\lambda_*) > 0$, such that r_{x_*} is identically 0 in $[\lambda_* - \varepsilon, \lambda_* + \varepsilon]$. Indeed, in the case $\lambda_* \neq 0$ there is ε, such that $\lambda \cdot \lambda_* > 0$ for all $\lambda - \lambda_* < \varepsilon$. This leads to a choice of m' such that r_{λ_*} vanishes. If $\lambda_* = 0$, then r_{λ_*} vanishes, too, as can be verified easily. Clearly, the same holds for

$$\psi_{\lambda_*}(\delta\lambda; 1) := \frac{|r_{\lambda_*}(\delta\lambda)|}{|\delta\lambda|},$$

which is a Baire-Carathéodory function and *pointwise continuous* at $\delta\lambda = 0$. Note that this result is independent of the definition of $m'(0)$. Moreover, $\psi_{\lambda_*}(\cdot; 1)$ is uniformly bounded, since r_{λ_*} is Lipschitz continuous. Hence, setting $\beta = 1$ in Theorem 3.2, the well known result of local superlinear convergence of Newton's method follows, if a *smoothing property* (5) holds for some $q < p$.

Smoothing Property. Since the nonlinear terms in our system depend on the adjoint state λ only, it is sufficient to consider a smoothing property for this component. It can be shown (cf. e.g. [7, Chapter 4]) that such a property holds for $q^{-1}+p^{-1}=1$, as long as for $-\Delta y = u$ we have $\|y\|_{L_p} \le c\,\|u\|_{L_2}$. We then obtain

$$\begin{pmatrix} \delta y_+ \\ \delta\lambda_+ \end{pmatrix} = G'(x)^{-1} \begin{pmatrix} 0 \\ R_{\lambda_*}(\lambda-\lambda_*) \end{pmatrix} \Longrightarrow \|\delta\lambda_+\|_{L_p} \le C\,\|R_{\lambda_*}(\lambda-\lambda_*)\|_{L_q}.$$

Especially, on regular domains $p=\infty$ and $q=1$ is usually obtained by H^2-regularity results for solutions of the state equation.

Rates of Convergence. To obtain rates of convergence we follow [9] and assume that

$$|\{t\in\Omega : 0 < |\lambda_*(t)| < \varepsilon\}| \le C\varepsilon^\gamma, \tag{20}$$

a condition that resembles a strengthened strict complementarity condition. If $|\lambda-\lambda_*| < |\lambda_*|$, or $\lambda_*=0$, then $r_{\lambda_*}=0$. Otherwise, we have $r_{\lambda_*} < |\lambda_*|$. Hence, for any $1<\alpha<\infty$

$$\psi_{\lambda_*}(\lambda-\lambda_*;\alpha) \le \frac{|\lambda_*|}{|\lambda-\lambda_*|^\alpha} \le |\lambda_*|^{1-\alpha},$$

and thus finally (20) becomes (8):

$$|\Omega_\varepsilon| := \left|\left\{t\in\Omega : \Psi_{\lambda_*}(\delta\lambda;\alpha)(t) < \varepsilon^{1-\alpha}\right\}\right| \le C\varepsilon^\gamma.$$

By Theorem 3.7 we obtain the rate of convergence β as in (15). Especially, for the very common case $\gamma=1$, $q=1$, $p=\infty$ we obtain $\beta=2$. Thus, Newton's method converges locally quadratically in function space. This observation coincides nicely with numerical results, as for example in [9, 5].

4 Second Order Sufficient Conditions for a Class of Optimal Control Problems

Our second application deals with second order sufficient conditions in optimal control. Again, we denote by y the state, and by u the control, and we set $x := (y,u)$. For simplicity we will only consider the case without inequality constraints and a quadratic functional for u. The extension to control constraints is straightforward. Moreover, we will assume that the control enters the state equation linearly. Thus, our model problem reads:

$$\begin{aligned} \min J(x) := \int_\Omega j(y) + \frac{1}{2}u^2\,dt \qquad s.t. \\ \mathcal{D}\,y + c(y) - u = 0. \end{aligned} \tag{21}$$

Here $j:\mathbb{R}\to\mathbb{R}$ and $c:\mathbb{R}\to\mathbb{R}$ are assumed to be twice differentiable nonlinear functions, and $\mathcal{D}$ is a linear differential operator.

Often, due to the so called *two-norm discrepancy*, one restricts the analysis of second order sufficient conditions to an L_∞-neighbourhood of the optimal solution (cf. e.g. [8] and references given there). As a consequence, the analysis of second order sufficient conditions is restricted to problems, where the state equation maps $L_q(\Omega)$ into $L_\infty(\Omega)$ for some $1 \le q \le \infty$. Many partial differential equations exhibit such a property, but often the needed regularity results require further analysis and additional regularity assumptions on the domain of definition and the data. We will point out in the following that the two-norm discrepancy is a consequence of the failure of Lemma 2.1 for $s = \infty$, and that in some cases a smoothing property of the underlying PDE (cf. (22) below) can be employed to overcome its effects.

Control-to-State Mapping. For our considerations we assume that the control-to-state mapping $S : u \to y$ is well defined and Lipschitz continuous as a mapping $L_2(\Omega) \to Y \hookrightarrow L_p(\Omega)$ with $p > 2$:

$$\|S(u_1) - S(u_2)\|_{L_p} \le \|S(u_1) - S(u_2)\|_Y \le C\,\|u_1 - u_2\|_{L_2} \tag{22}$$

for some reflexive Banach space $Y \hookrightarrow L_p(\Omega)$. Such a setting is usually satisfied, if the differential operator $\mathcal{D}$ is an isomorphism between a Sobolev space Y and its dual Y^*, and c satisfies suitable monotonicity and growth conditions.

Let $x_* = (y_*, u_*)$ be a point that satisfies the first order necessary conditions. Assume further that S is Fréchet differentiable at u_* with respect to u in the sense that there is a continuous linear operator $S'(u_*) : L_2(\Omega) \to Y$ with

$$\left\|S'(u_*)\delta u - (S(u_* + \delta u) - S(u_*))\right\|_Y = o\left(\|\delta u\|_{L_2}\right) \qquad \forall \delta u \in L_2(\Omega). \tag{23}$$

Such a result can usually be shown via an implicit function theorem.

Lagrangian Functions. To derive second order sufficient conditions we introduce the Lagrangian function

$$L(x;\lambda) := J(x) + \langle \lambda, \mathcal{D}\,y + c(y) - u\rangle_{Y,Y^*}\,. \tag{24}$$

The first order necessary conditions state the existence of a Lagrange multiplier $\lambda_* \in Y^{**} = Y$, such that $L_x(x_*, \lambda_*) = 0$. Since j and c are twice pointwise differentiable at y_* we can formulate the following second order condition at x_*:

$$L_{xx}(x_*;\lambda_*)\delta x^2 \ge \alpha\,\|\delta u\|_{L_2}^2 \quad \forall \delta y = S'(u_*)\delta u \tag{25}$$

with L_{xx} given by

$$L_{xx}(x_*;\lambda_*)\delta x^2 = \left\langle \delta y, j''(y_*)\delta y\right\rangle_{Y,Y^*} + \left\langle \lambda_*, c''(y_*)\delta y^2\right\rangle_{Y,Y^*} + \langle \delta u, \delta u\rangle_{L_2}\,.$$

It remains to study in which sense (25) is an essential part of second order sufficient conditions for x_* being a minimizer of (21). The results that we will obtain are related in character to the second order sufficient conditions from [3] considered also in [10], which rely on a similar structure in a quite different framework.

We start by showing that (25) carries over locally from the linearized problem to the nonlinear problem.

Lemma 4.1. *Assume that* $L_{yy}(x_*;\lambda_*)(\cdot,\cdot) : Y \times Y \to \mathbb{R}$ *is a continuous bilinear form, and suppose that* (22), (23), *and* (25) *hold. Then*

$$L_{xx}(x_*;\lambda_*)\delta x^2 \geq \tilde{\alpha}\,\|\delta u\|_{L_2}^2 \quad \forall \delta y = S(u_* + \delta u) - S(u_*) \tag{26}$$

for some $0 < \tilde{\alpha} < \alpha$ *and sufficiently small* δu.

Proof. Let $\delta\tilde{y} := S'(u_*)\delta u$, and $\delta y := S(u_* + \delta u) - S(u_*)$. By (22) and (23) we have

$$\|\delta y\|_Y = O(\|\delta u\|_{L_2}), \quad \|\delta\tilde{y}\|_Y = O(\|\delta u\|_{L_2}), \quad \|\delta y - \delta\tilde{y}\|_Y = o(\|\delta u\|_{L_2}).$$

Thus for $L_{yy}(x_*;\lambda_*)(\cdot,\cdot)$ being a continous bilinear form:

$$L_{yy}(\delta y, \delta y) - L_{yy}(\delta\tilde{y}, \delta\tilde{y}) = L_{yy}(\delta y - \delta\tilde{y}, \delta y + \delta\tilde{y}) \leq \|L_{yy}\|\, o(\|\delta u\|_{L_2}^2).$$

This implies $L_{xx}(x_*;\lambda_*)\delta x^2 \geq L_{xx}(x_*;\lambda_*)\delta\tilde{x}^2 - o(\|\delta u\|_{L_2})^2$. □

Analysis of the Remainder Terms. Let now $\rho(\delta y)$ be defined by

$$\rho(\delta x) := L(x_* + \delta x;\lambda_*) - \left(L(x_*;\lambda_*) + L_x(x_*;\lambda_*)\delta x + \frac{1}{2}L_{xx}(x_*;\lambda_*)\delta x^2\right), \tag{27}$$

which is the second order remainder term of a formal Taylor expansion of L at x_*. We observe that ρ does not depend on u due to the particular choice of (21). Moreover, due to the linearity of $\mathcal{D}$ we can write $\rho(\delta x)$ as follows:

$$\rho(\delta x) = \int_\Omega R(\delta y)\,dt.$$

Here R is a Nemyckii operator, corresponding to

$$r(\delta y, t) := r^j(\delta y) + \lambda_*(t) r^c(\delta y),$$

with r^j and r^c defined as the pointwise second order remainder terms for j and c:

$$\begin{aligned} r^j(\delta y) &:= j(y_* + \delta y) - \left(j(y_*) + j'(y_*)\delta y + j''(y_*)\delta y^2\right), \\ r^c(\delta y) &:= c(y_* + \delta y) - \left(c(y_*) + c'(y_*)\delta y + c''(y_*)\delta y^2\right). \end{aligned}$$

Since the first order necessary conditions hold, i.e., $L_x(x_*;\lambda_*) = 0$, the first order term in (27) drops out, and inserting (26) we derive

$$\begin{aligned} L(x_* + \delta x;\lambda_*) &= L(x_*;\lambda_*) + \frac{1}{2}L_{xx}(x_*;\lambda_*)\delta x^2 + \int_\Omega R(\delta y)\,dt \\ &\geq L(x_*,\lambda_*) + \frac{1}{2}\tilde{\alpha}\,\|\delta u\|_{L_2}^2 - \|R(\delta y)\|_{L_1}. \end{aligned} \tag{28}$$

Thus, we have to assert $\|R(\delta y)\|_{L_1} = o(\|\delta u\|_{L_2}^2)$ in order to show that x_* is a local minimizer of (21). For this purpose we use the framework developed in Section 2 and define $\Psi(\delta y; 2)$ as the Nemyckii operator corresponding to

$$\psi(\delta y, t; 2) := \frac{|r(\delta y, t)|}{|\delta y|^2}. \tag{29}$$

Since j and c are pointwise twice differentiable almost everywhere, $\psi(\delta y, t; 2)$ is pointwise continuous almost everywhere with $\psi(0, t; 2) = 0$. If $\Psi(\cdot; 2) : L_p(\Omega) \to L_s(\Omega)$ for suitable $s < \infty$, then we obtain by Lemma 2.4 with $\beta = 2$ and $1 = s^{-1} + 2p^{-1}$

$$\|R(\delta y)\|_{L_1} = \omega(\|\delta y\|_{L_p}) \|\delta y\|_{L_p}^2 .$$

Now we have arrived at the crucial point of our derivation. The *failure* of Lemma 2.1 for $p < \infty$ and $s = \infty$ implies that we have to choose $p > 2$, which is the cause for a *norm-gap*. Obviously we face the same situation as in the case of Newton methods (cf. Remark 3.3). To overcome this difficulty we need the *smoothing property* (22) to estimate

$$\|R(\delta y)\|_{L_1} = o(\|\delta y\|_{L_p}^2) = o\left(\|\delta u\|_{L_2}^2\right) . \tag{30}$$

Inserting our estimate (30) into (28) we obtain

$$L(x_* + \delta x; \lambda_*) \geq L(x_*, \lambda_*) + \frac{1}{2}\tilde{\alpha} \|\delta u\|_{L_2}^2 - o(\|\delta u\|_{L_2}^2) \geq L(x_*, \lambda_*) + \frac{1}{2}\hat{\alpha} \|\delta u\|_{L_2}^2$$

for all sufficiently small $\|\delta u\|_{L_2}$, $\delta y := S(u_* + \delta u) - S(u_*)$, and some $0 < \hat{\alpha} < \tilde{\alpha}$. Since both x_* and $x_* + \delta x$ satisfy the equality constraints it follows that

$$J(x_* + \delta x) \geq J(x_*) + \frac{1}{2}\hat{\alpha} \|\delta u\|_{L_2}^2 . \tag{31}$$

Consequently, x_* is a local minimizer of (21) among all feasible pairs (u, y) with sufficiently small $\|u - u_*\|_{L_2}$.

We thus have derived the following prototype result:

Theorem 4.2 (Second Order Sufficient Conditions). *Consider the optimal control problem* (21) *and a feasible solution* $x_* = (y_*, u_*)$*. Let* $L(x_*; \lambda_*)$ *be the Lagrangian function* (24) *and* $\Psi(\delta y; 2)$ *the Nemyckii operator corresponding to* $\psi(\delta y, t; 2)$*, defined by* (29)*. Let* Y *be a reflexive Banach space* $Y \hookrightarrow L_p(\Omega)$ *with* $p > 2$*. Suppose that the following conditions hold:*

(i) $j : \mathbb{R} \to \mathbb{R}$ *and* $c : \mathbb{R} \to \mathbb{R}$ *are twice differentiable functions.*

(ii) The equation $\mathcal{D}\, y + c(y) - u = 0$ *admits a solution operator* $S : L_2(\Omega) \to Y$*,* $y = S(u)$*, that satisfies* (22) *and* (23)*.* $S'(u_*) : L_2(\Omega) \to Y$ *is continuous as a linear operator.*

(iii) $L_{yy}(x_*; \lambda_*)(\cdot, \cdot) : Y \times Y \to \mathbb{R}$ *is a continuous bilinear form.*

(iv) $\Psi(\delta y; 2)$ *maps* $L_p(\Omega)$ *into* $L_s(\Omega)$ *for* $s^{-1} + 2p^{-1} = 1$*.*

(v) $x_* = (y_*, u_*)$ *satisfies the first order necessary conditions* $L_x(x_*; \lambda_*) = 0$ *and the second order condition* (25)*.*

Then $x_* = (y_*, u_*)$ *is a local minimizer of* (21) *among all feasible pairs* (y, u)*, for which* $\|u - u_*\|_{L_2}$ *is sufficiently small, and* (31) *holds.*

Condition (iii) implies the requirement that $j''(y_*), \lambda_* c''(y_*) \in L_q(\Omega)$, such that $q^{-1} + 2p^{-1} = 1$. In view of (1) condition (iv) implies that the remainder term r has to satisfy the following growth condition

$$r(\delta y, t) \le a(t) \, |\delta y|^2 + b \, |\delta y|^p \,, \quad a(t) \in L_s(\Omega).$$

Remark 4.3. If no smoothing property is available, for example, if we have to deal with a remainder term $R(\delta u)$ depending on the control, then we can still resort to a uniform continuity result for $\psi(\delta u, t; 2)$ with $p = s = \infty$, as indicated in Remark 2.2. Application of the Hölder-inequality then gives us the following estimate in place of (30)

$$\|R(\delta u)\|_{L_1} = \omega(\|\delta u\|_{L_\infty}) \, \|\delta u\|_{L_2}^2 \,.$$

This leads to the classical two-norm discrepancy: (y_*, u_*) is a minimizer of J among all feasible (y, u) with sufficiently small $\|u - u_*\|_{L_\infty(\Omega)}$, and (31) holds.

References

[1] J. Appell and P.P. Zabrejko. *Nonlinear superposition operators.* Cambridge University Press, 1990.

[2] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*, volume 35 of *Series Computational Mathematics.* Springer, 2004.

[3] J.C. Dunn and T. Tian. Variants of the Kuhn-Tucker sufficient conditions in cones of nonnegative functions. *SIAM J. Control Optim.*, 30(6):1361–1384, 2002.

[4] M. Hintermüller, K. Ito, and K Kunisch. The primal-dual active set strategy as a semi-smooth Newton method. *SIAM J. Optim.*, 13:865–888, 2002.

[5] M. Hintermüller and M. Ulbrich. A mesh-independence result for semismooth Newton methods. *Math. Programming*, 101:151–184, 2004.

[6] B. Kummer. Generalized Newton and NCP-methods: convergence, regularity and actions. *Discussiones Mathematicae - Differential Inclusions*, 20(2):209–244, 2000.

[7] A. Schiela. *The Control Reduced Interior Point Method - A Function Space Oriented Algorithmic Approach.* PhD thesis, Free University of Berlin, Dept. Math. and Comp. Sci., 2006.

[8] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen. Theorie, Verfahren und Anwendungen.* Vieweg, 2005.

[9] M. Ulbrich. Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. Habilitation thesis, Fakultät für Mathematik, Technische Universität München, 2002.

[10] M. Ulbrich and S. Ulbrich. Superlinear convergence of affine-scaling interior-point Newton methods for infinite-dimensional nonlinear problems with pointwise bounds. *SIAM J. Control Optim.*, 38(6):1938–1984, 2000.

[11] Martin Väth. *Integration Theory. A second course*. World Scientific Publishing, 2002.

[12] E. Zeidler. *Nonlinear Functional Analysis and its Applications*, volume II/B. Springer, New York, 1990.