3 Inexact Newton Continuation

A program may fail, but it must not lie.
— Beresford Parlett —

3.1 Predictor-Corrector Methods

Interior point methods are a means of defining a homotopy for the difficult to solve KKT systems, with a homotopy parameter μ such that the problem $F(v;\mu) = 0$ is far easier to solve for $\mu \approx 1$ than for $\mu = 0$. The path $v(\mu)$ defined implicitly by $F(v(\mu);\mu) = 0$ converges to the solution v^*.

The homotopy structure is then exploited by some continuation or path-following method, most often used in predictor-corrector form. Starting from $v(\mu_n)$, first a predictor is applied for the reduced homotopy parameter $\mu_{n+1} < \mu_n$ to generate a point $v_{n+1} \approx v(\mu_{n+1})$. A very popular and simple predictor is the tangential Euler predictor

$$v_{n+1} := v(\mu_n) + (\mu_{n+1} - \mu_n)v'(\mu_n)$$

with the derivative of the path given by $v'(\mu) = F_v(v(\mu);\mu)^{-1}F_\mu(v(\mu);\mu)$. Subsequently, a corrector is applied to compute the solution $v(\mu_{n+1})$ of $F(v;\mu_{n+1}) = 0$, starting from the point v_{n+1}. In general, a Newton type corrector is used.

Many different continuation algorithms have been suggested, using different predictors and correctors as well as different adaptive step size selection methods. For an overview, we refer to the textbook by Allgower and Georg [1].

In the following, we will develop a continuation method that is especially well suited for following the central path of the complementarity formulation of optimal control problems in function space. First we will consider the Newton corrector step, and its convergence and behavior under linear transformations are investigated. Then we will construct a predictor-corrector continuation scheme and, finally theoretically near-optimal parameter settings for controlling the algorithm are derived.

3.2 Affine Invariance

In the following we will state the standard assumptions that are assumed to hold throughout the section.

Assumption 3.2.1. Suppose V and Z are Banach spaces, $D \subset V$ is open and $F : D \to Z$ is Gâteaux-differentiable. Additionally, assume $F_v(v)$ is invertible for all $v \in D$.

For the solution of nonlinear equations of the form

$$F(v) = 0, \tag{3.1}$$

Newton's method

$$\begin{aligned} F_v(v^k)\Delta v^k &= -F(v^k) \\ v^{k+1} &= v^k + \Delta v^k \end{aligned}$$

in one of it's many variants is a useful tool. One interesting property of Newton's method is its invariance with respect to linear transformations. Let $A : Z \to Z$ and $B : V \to V$ be isomorphisms and consider the transformed problem

$$\hat{F}(y) := AF(By) = 0 \tag{3.2}$$

with the domain transformation $v = By$. Clearly, the solution is just transformed as the whole space: $v^* = By^*$. Using equivalent initial values $v^0 = By^0$ for problems (3.1) and (3.2), respectively, Newton's method generates equivalent sequences of iterates:

$$\begin{aligned} \Delta v^k = -F_v(v^k)^{-1}F(v^k) &= -B(AF_v(v^k)B)^{-1}AF(v^k) \\ &= -B\hat{F}(y^k)^{-1}\hat{F}(y^k) = B\Delta y^k \end{aligned}$$

implies $v^k = By^k$ by induction. This is a very desirable property, meaning that Newton's method does indeed work simultaneously on the whole class of problems that differ only by a linear transformation. Of course, we would like to have a convergence theory and actual implementations inheriting this invariance property.

Unfortunately, this is not possible to full extent, since for convergence to be measured there has to be a norm applied on the function values (or the Newton corrections, respectively). Therefore one can only hope to formulate a convergence theory that is invariant under a certain subset of all transformations of the general type (3.2). With no actual problem at hand, we just state the existence of a set of meaningful transformations.

Conjecture 3.2.2. *For each problem of type* (3.1) *there exists a subgroup* $T(F) \subset \mathcal{L}(Z) \times \mathcal{L}(V)$ *of* meaningful transformations*, under which convergence theory and actual implementation should be invariant.*

What transformations are meaningful is specific for the problem type and reflects its structure.

Affine invariant Newton theory and algorithms have been developed for several types of transformations (cf. [21, 28, 22, 23, 51]). For an overview we refer to the forthcoming textbook by DEUFLHARD [19].

Since an affine invariant convergence theory must be formulated in terms of an invariant norm, we have to define what makes a norm an affine invariant one.

Definition 3.2.3. Suppose the Assumptions 3.2.1 are satisfied. A family of norms $\|\cdot\|_v : Z \to \mathbb{R}$ with $v \in D$ is called *affine invariant* for problem (3.1), if for all pairs $(A,B) \in T(F)$ of meaningful transformations

$$\|\hat{F}(\xi)\|_y = \|F(B\xi)\|_{By}$$

for all $v, \xi \in D$, where $\hat{F}(\xi) := AF(B\xi)$. It is said to form a γ-continuous family of norms, if

$$\|r\|_{v+\xi} \leq (1 + \gamma\|F_v(v)\xi\|_v)\|r\|_v \tag{3.3}$$

for all $\xi \in V$.

Convergence theories and algorithms which are based solely on invariant norm values are then invariant under transformations from $T(F)$.

Assumption 3.2.4. Suppose the Assumptions 3.2.1 are satisfied. Additionally, let $\|\cdot\|_v : Z \to \mathbb{R}$ form a continuous family of norms.

3.3 Inexact Newton Corrector

Since we have to deal with function space problems, we cannot compute the Newton correction exactly. Discretization errors and possibly truncation errors from iterativly solving linear systems have to be taken into account. Therefore, we consider *inexact Newton methods*, where an *inner residual* remains:

$$F_v(v^k)\delta v^k = -F(v^k) + r^k$$
$$v^{k+1} = v^k + \delta v^k$$

Inexact Newton methods have been studied by BANK and ROSE [3] and DEMBO, EISENSTAT, and STEIHAUG [16]. YPMA [53] formulated an affine invariant theory, and adaptive affine invariant algorithms were designed by DEUFLHARD [18].

The relative accuracy of the *inexact Newton correction* δv^k, given by

$$\delta_k := \frac{\|r^k\|_{v^k}}{\|F(v^k)\|_{v^k}}, \tag{3.4}$$

will play a crucial role in the convergence analysis as well as in the implementation of inexact Newton methods.

Theorem 3.3.1. *Assume the standard assumptions 3.2.4 are satisfied. Let* γ *and* ω *be constants such that the family of norms is* γ*-continuous and the affine invariant Lipschitz condition*

$$\|(F_v(v+t\xi)-F_v(v))\xi\|_{v+\xi} \le t\omega\|F_v(v)\xi\|_v^2 \tag{3.5}$$

holds for all $v, \xi \in V$ *such that* $\mathrm{co}\{v, v+\xi\} \subset D$. *Let* $\Theta < 1$ *and*

$$\mathcal{L}(v) := \left\{\xi \in D : \|F(\xi)\|_\xi \le \left(1+\frac{\gamma}{\omega}\right)\|F(v)\|_v\right\}.$$

Assume that $v^0 \in D$*, the level set* $\mathcal{L}(v^0)$ *is closed, and define* $h_k := \omega\|F(v^k)\|_{v^k}$. *If* $h_0 < 2$ *and the inner iteration is controlled such that*

$$\frac{1}{2}(1+\delta_k)^2 h_k + (1+(1+\delta_k)\gamma\|F(v^k)\|_{v^k})\delta_k \le \Theta\,, \tag{3.6}$$

then the iterates are well defined for all $k \in \mathbb{N}$*, stay in* $\mathcal{L}(v^0)$*, and the residuals converge to zero at a rate of*

$$h_{k+1} \le \Theta h_k\,.$$

Furthermore,

$$\|F(v^{k+1})\|_{v^k} \le \left(\delta_k + \frac{1}{2}(1+\delta_k)^2 h_k\right)\|F(v^k)\|_{v^k}\,. \tag{3.7}$$

Proof. By induction, let $\mathcal{L}(v^k)$ be closed and $\omega\|F(v^k)\|_{v^k} < 2$. Then

$$F(v^k + s\delta v^k) = F(v^k) + \int_0^s F_v(v^k + t\delta v^k)\delta v^k\,dt \tag{3.8}$$

$$= (1-s)F(v^k) + sr^k + \int_0^s (F_v(v^k+t\delta v^k) - F_v(v^k))\delta v^k\,dt \tag{3.9}$$

for all $s \in [0,1]$ with $\mathrm{co}\{v^k, v^k + s\delta v^k\} \subset D$. From (3.4) we have

$$\|F_v(v^k)\delta v^k\|_{v^k} = \|F(v^k) - r^k\|_{v^k} \le (1+\delta_k)\|F(v^k)\|_{v^k}\,.$$

Using the Lipschitz continuity (3.5), the continuity of the family of norms (3.3), and the accuracy requirement (3.6), we have

$$\begin{aligned}
&\|F(v^k+s\delta v^k)\|_{v^k+s\delta v^k}\\
&\quad\le (1-s)\|F(v^k)\|_{v^k+s\delta v^k} + s\|r^k\|_{v^k+s\delta v^k}\\
&\qquad + \int_0^s \|(F_v(v^k+t\delta v^k)-F_v(v^k))\delta v^k\|_{v^k+s\delta v^k}\,dt\\
&\quad\le (1-s)(1+s\gamma\|F_v(v^k)\delta v^k\|_{v^k})\|F(v^k)\|_{v^k}\\
&\qquad + s(1+s\gamma\|F_v(v^k)\delta v^k\|_{v^k})\|r^k\|_{v^k} + \int_0^s t\omega\|F_v(v^k)\delta v^k\|_{v^k}^2\,dt\\
&\quad= (1+s\gamma(1+\delta_k)\|F(v^k)\|_{v^k})((1-s)\|F(v^k)\|_{v^k} + s\delta_k\|F(v^k)\|_{v^k})\\
&\qquad + \frac{\omega}{2}s^2(1+\delta_k)^2\|F(v^k)\|_{v^k}^2\\
&\quad\le \Big((1+s\gamma(1+\delta_k)\|F(v^k)\|_{v^k})(1-s+s\delta_k)\\
&\qquad\quad + \frac{1}{2}s^2(1+\delta_k)^2h_k\Big)\|F(v^k)\|_{v^k}\,.
\end{aligned}$$

Defining $\chi := \gamma\|F(v^k)\|_{v^k}$ and using $s \le 1$, $\delta_k < 1$, and (3.6), we have

$$\begin{aligned}
&\frac{\|F(v^k+s\delta v^k)\|_{v^k+s\delta v^k}}{\|F(v^k)\|_{v^k}}\\
&\qquad\le (1+s(1+\delta_k)\chi)(1-s+s\delta_k) + \frac{1}{2}s^2(1+\delta_k)^2h_k\\
&\qquad\le (1+s(1+\delta_k)\chi)(1-s) + s\Theta\\
&\qquad\le (1+2s\chi)(1-s) + s \le 1+\frac{\chi}{2}\\
&\qquad= 1+\frac{\gamma}{2}\|F(v^k)\|_{v^k} \le 1+\frac{\gamma}{\omega}
\end{aligned}$$

and thus

$$\|F(v^k+s\delta v^k)\|_{v^k+s\delta v^k} \le \left(1+\frac{\gamma}{\omega}\right)\|F(v^k)\|_{v^k}\,.$$

If $v^k+\delta v^k \notin D$, then there is some $s^* \in [0,1)$ with $\operatorname{co}\{v^k, v^k+s^*\delta v^k\} \subset D$ but $v^k+s^*\delta v^k \notin \mathcal{L}(v^k)$, i.e. $\|F(v^k+s^*\delta v^k)\|_{v^k+s^*\delta v^k} > (1+\gamma/\omega)\|F(v^k)\|_{v^k}$, which is a contradiction. Thus, $v^{k+1} \in D$. Furthermore, setting $s=1$ we have

$$\|F(v^{k+1})\|_{v^{k+1}} \le \Theta_k\|F(v^k)\|_{v^k} \tag{3.10}$$

and therefore $\mathcal{L}(v^{k+1}) \subset \mathcal{L}(v^k)$. Since $\mathcal{L}(v^k)$ is closed, every Cauchy sequence in $\mathcal{L}(v^{k+1})$ converges to a limit point in $\mathcal{L}(v^k)$, which is, by continuity of the norm, also contained in $\mathcal{L}(v^{k+1})$. Hence, $\mathcal{L}(v^{k+1})$ is closed.

From (3.9) we conclude in a similar way

$$\begin{aligned}
\|F(v^{k+1})\|_{v^k} &\le \|r_k\|_{v^k} + \frac{1}{2}(1+\delta_k)^2h_k\|F(v^k)\|_{v^k}\\
&\le \left(\delta_k + \frac{1}{2}(1+\delta_k)^2h_k\right)\|F(v^k)\|_{v^k}\,.
\end{aligned}$$

□

Remark 3.3.2. The linear convergence result (3.10) of inexact Newton methods can be strengthend to quadratic convergence if the accuracy matching is strengthened to $\delta_k = \mathcal{O}(\|F(v^k)\|^2_{v^k})$. ◁

Remark 3.3.3. For the application to optimal control problems, the Lipschitz condition (3.5) implies the use of a norm that is at least as fine as L_∞ (cf. Section 2.3). The consequence is, that possible discontinuities in the Newton corrector must be located exactly at discretization grid points. ◁

3.3.1 Computable Estimates

For actual implementation of the inexact Newton method analyzed in Theorem 3.3.1 we need easily computable estimates of the unknown constants ω and γ in order to satisfy the accuracy matching (3.6). From (3.7) we gain the a posteriori estimate

$$[\omega] := \frac{2}{(1+\delta_k)^2\|F(v^k)\|_{v^k}} \left(\frac{\|F(v^{k+1})\|_{v^k}}{\|F(v^k)\|_{v^k}} - \delta_k \right) \leq \omega \tag{3.11}$$

and $[h_k] := [\omega]\|F(v^k)\|_{v_k}$. The estimates can be expected to be reliable only if

$$\|F(v^{k+1})\|_{v^k} > (\rho_\omega + 1)\delta_k \|F(v^k)\|_{v^k} \tag{3.12}$$

with some suitable safety factor $\rho_\omega > 0$. The requirement (3.12) of being able to reliably estimate the Lipschitz constant ω motivates an additional accuracy requirement for δ_k. Because of

$$\begin{aligned} \|F(v^{k+1})\|_{v^k} &\approx \big(\delta_k + \frac{h_k}{2}(1+\delta_k)^2\big)\|F(v^k)\|_{v^k} \\ &\geq \big(\delta_k + \frac{h_k}{2}\big)\|F(v^k)\|_{v^k} \\ &\geq (\rho_\omega + 1)\|F(v^k)\|_{v^k} \end{aligned}$$

we end up with the requirement

$$\delta_k \leq \rho_\omega \frac{[h_k]}{2}\,.$$

This requirement implies quadratic convergence and is therefore too restrictive. In actual computation, linearly convergent Newton methods are to be preferred because of the more fine-grained control of the iteration's termination they provide. Hence the conclusion is that the estimate (3.11) has to be used with care in case (3.12) is violated.

From (3.3) we get the estimate

$$[\gamma] := \|F_v(v)\delta v\|_v^{-1} \left(\frac{\|F(v+\delta v)\|_{v+\delta v}}{\|F(v+\delta v)\|_v} - 1 \right) \tag{3.13}$$

in case $\|F(v+\delta v)\|_{v+\delta v} \geq \|F(v+\delta v)\|_v$ or the approximate estimate

$$[\gamma] := \|F_v(v)\delta v\|_v^{-1} \left| \frac{\|F(v+\delta v)\|_{v+\delta v}}{\|F(v+\delta v)\|_v} - 1 \right| ,$$

otherwise. In general, most of the values that must be computed for evaluating $[\gamma]$ are needed during the course of the Newton iteration anyway, such that the estimate comes at little additional cost.

Because invariant norms are frequently defined in terms of F itself, ω and γ are theoretically coupled since both describe aspects of the nonlinearity of F. Nevertheless, computation of the estimates $[\omega]$ and $[\gamma]$ is completely decoupled. In particular, $[\gamma]$ does not depend on δ. The independence of the estimates increases the reliability of the computed values.

Having estimates for ω and γ at hand, the termination of the inner iteration of the inexact Newton method can be controlled by

$$\frac{[h_k]}{2}(1+\delta_k)^2 + (1+(1+\delta_k)[\gamma]\|F(v^k)\|_{v^k})\delta_k \leq \Theta$$

with some contraction factor Θ between $[h_k]/2$ and 1.

Remark 3.3.4. For the implementation of the estimators in the context of adaptively refined discretizations it is of vital importance that the norms are evaluated using the same discretization, unless the discretization error is unreasonably small. Otherwise, the norm values can be noncontinuous for $\|\delta v^k\| \to 0$, rendering the computed estimators unusably wrong. ◁

3.4 Inexact Tangential Predictor

At the topmost level, interior point type methods use a continuation scheme to follow the central path $v(\mu)$ to the solution point $v(0)$. Several efficient and sophisticated algorithms have been proposed, ranging from general purpose methods like [17, 20] to highly specialized higher order predictors for IP methods [41].

Most of them use a tangential or higher order predictor and propose a step length based on $F(v - \Delta\mu p; \mu - \Delta\mu) = \mathcal{O}(\Delta\mu^s)$, where p is the predictor of order s. Accordingly, those methods implicitly assume that $\|F(v;\mu)\| \ll \|F(v - \Delta\mu p; \mu - \Delta\mu)\|$ for every reasonable μ. This is justified for finite dimensional problems, where Newton's method or quasi-Newton methods with quadratic or at least superlinear convergence rate are used. But for function space problems, a reduction of $\|F(v;\mu)\|$ is in general quite expensive, since it requires a reduction of the discretization error. Fortunately, it is not necessary to obtain a highly accurate corrector solution in the course of the continuation, since we are only interested in $v^* = v(0)$, as opposed to $v(\mu)$ for $\mu > 0$.

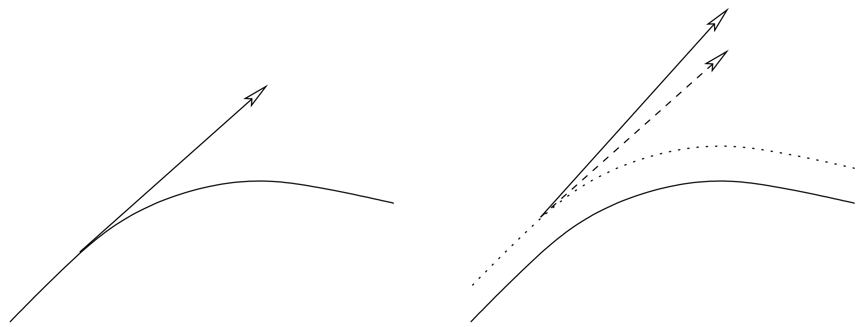

Figure 3.1: Exact tangential predictor (left) and its inexact counterpart (right).

Hence, an efficient continuation algorithm for function space problems will follow the path in an inexact way: The termination criterion for the corrector will be $\|F(v;\mu)\| \le a\delta_{\mathrm{TOL}}$ with $a\delta_{\mathrm{TOL}}$ relatively large. The predictor will be computed only up to a suitably chosen accuracy, thus accepting smaller continuation step sizes in favor of a cheaper discretization. The consequence of this tradeoff is, that the deviation from the central path has to be taken into account when computing a new step size suggestion. Moreover, the higher accuracy of higher order predictors will only pay off if the predictor is computed with high accuracy, too. Therefore, we will use a tangential predictor with a step size suggestion based on

$$F(v - \Delta\mu p; \mu - \Delta\mu) = F(v;\mu) - \Delta\mu F_v(v;\mu)p + \mathcal{O}(\Delta\mu^2)\,.$$

In the following we will establish a continuation method that takes this into account.

The step size selection mechanism is based on the assumption of a slowly varying curvature of the path, thus suggesting a step size that would have been optimal for the last step. For function space complementarity methods, though, the curvature of the central path can be expected to constantly increase for $\mu \to 0$, at least in the presence of bang-bang control. Without step size control, a reduction of μ by a constant factor instead of a constant difference would be appropriate. Therefore, we will use a scaled homotopy parameter $\tau := -\log\mu$, such that a slowly varying step size $\Delta\tau$ translates into a slowly varying reduction factor for μ.

Lemma 3.4.1. *Suppose the Assumptions 3.2.4 are satisfied. Let γ and β be constants such that the pair of norms is γ-continuous and that*

$$\begin{aligned}\|F(v + \Delta\tau\delta v; \tau + \Delta\tau)\|_v \\ \le \|F(v;\tau)\|_v + \Delta\tau\,\|F_v(v;\tau)\delta v + F_\tau(v;\tau)\|_v + \beta\Delta\tau^2 \quad (3.14)\end{aligned}$$

holds for $v, \delta v \in V$ *and* $\Delta\tau \in \mathbb{R}$ *such that* $v + \delta\tau\delta v \in D$. *If for some tolerance* $\delta_{\mathrm{TOL}} \geq \|F(v;\tau)\|_v$ *the inexact tangential predictor* p *defined by* $F_v(v;\tau)p = -F_\tau(v;\tau) + r$ *is used together with a step size* $\Delta\tau$ *satisfying*

$$(1 + \gamma\Delta\tau \|F_v(v;\tau)p\|_v)(\|F(v;\tau)\|_v + \Delta\tau \|r\|_v + \beta\Delta\tau^2) \leq \delta_{\mathrm{TOL}} , \qquad (3.15)$$

then

$$\|F(v + \Delta\tau p; \tau + \Delta\tau)\|_{v+\delta\tau p} \leq \delta_{\mathrm{TOL}} . \qquad (3.16)$$

Additionally,

$$\|F(v + \Delta\tau p; \tau + \Delta\tau)\|_v \leq \frac{\delta_{\mathrm{TOL}}}{1 + \gamma\Delta\tau \|F_v(v;\tau)p\|_v} . \qquad (3.17)$$

Proof. Inequality (3.17) is easily verified. The main statement (3.16) is then a direct consequence of the γ-continuity of the local norms. □

3.4.1 Computable Estimate

For actual implementation we need an easily computable estimate for the unknown constant β. From (3.14) we get the estimate

$$[\beta] := \frac{1}{\Delta\tau^2} (\|F(v + \Delta\tau p; \tau + \Delta\tau)\|_v - \|F(v;\tau)\|_v - \Delta\tau \|r\|_v) \leq \beta , \qquad (3.18)$$

which is reliable if $\|F(v + \Delta\tau p; \tau + \Delta\tau)\|_v - \|F(v;\tau)\|_v > (\rho_\beta + 1)\Delta\tau\|r\|_v$ with some safety factor $\rho_\beta > 0$. In order to obtain reliable estimates we derive the accuracy requirement $\|r\|_v \leq \frac{\Delta\tau}{\rho_\beta}[\beta]$ from

$$\|F(v + \Delta\tau p; \tau + \Delta\tau)\|_v - \|F(v;\tau)\|_v \approx \Delta\tau \|r\|_v + \beta\Delta\tau^2 .$$

Because of the high computational effort that must be spent on reducing the discretization error, we will impose a lower bound on the requirement:

$$\|r\|_v \leq \max\left\{\frac{\Delta\tau}{\rho_\beta}[\beta], r_{\min}\right\} . \qquad (3.19)$$

Note that in contrast to the computation of the Newton correction there is no definite upper bound on the tolerance for computing the inexact tangential predictor if only $[\beta]$ is sufficiently large. However, too large deviations from the exact tangential predictor lead to unreasonable small step sizes and hence inefficiency.

3.4.2 Step Size Selection

In order to be able to continue the homotopy method along the path, we will require the corrector to converge, starting from the predicted point $(v + \Delta\tau p, \tau + \Delta\tau)$. Clearly this imposes an upper bound on how far from the path the predicted point may be:

$$\delta_{\mathrm{TOL}} < \frac{2}{\omega}\,.$$

Then we can substitute the unknown quantities β, γ, and ω by their estimates $[\omega]$, $[\beta]$, and $[\gamma]$, and suggest a suitable stepsize via (3.15). Since the estimates may be too small, the proposed step size can be too large, such that we need to

- select $\delta_{\mathrm{TOL}} < {}^{2}\!/_{[\omega]}$.
- introduce a safety factor $0 < \rho < 1$ and compute a step size $\Delta\tau$ satisfying

$$(1 + [\gamma]\Delta\tau\,\|F_v(v;\tau)p\|_v)(\|F(v;\tau)\|_v + \Delta\tau\,\|r\|_v + [\beta]\Delta\tau^2) \le \rho^3\delta_{\mathrm{TOL}}\,. \tag{3.20}$$

- do a step size reduction if the corrector does not converge.

For efficiency reasons, we propose a three level step size reduction scheme:

1. Test for $\|F(v + \Delta\tau p; \tau + \Delta\tau)\|_v \le \rho^2\delta_{\mathrm{TOL}}$. If the test fails, update the estimate $[\beta]$ and compute a smaller step size based on the new estimate. In general, this will require one function evaluation.
2. Test for

$$\begin{aligned}\|F(v + \Delta\tau p; \tau + \Delta\tau)\|_{v+\Delta\tau p}&\\ \le \rho^{-1}(1 + [\gamma]\Delta\tau\|F_v(v;\tau)p\|_v)&\|F(v + \Delta\tau p; \tau + \Delta\tau)\|_v\,.\end{aligned}$$

   If the test fails, update $[\gamma]$ and $[\beta]$ according to the new information and compute a smaller step size based on the new estimate. In general, this will require one derivative evaluation and one system solve.
3. Test for corrector convergence. If the corrector does not converge, update the estimates $[\omega]$, $[\gamma]$, and $[\beta]$ according to the new information and compute a smaller step size based on the updated δ_{TOL}. In general, this will require one function evaluation and one system solve.

Theorem 3.4.2. *Under the assumptions of Lemma 3.4.1, the step size reduction scheme defined above terminates after finitely many reductions.*

First we prove that the tests 1 to 3 are passed if only the estimates are sufficiently accurate.

Lemma 3.4.3. *Under the assumptions of Lemma 3.4.1, the step size suggested in* (3.20) *will pass the step size reduction tests if* $\rho\beta \leq [\beta]$, $\rho\gamma \leq [\gamma]$, *and* $\rho\omega < [\omega]$.

Proof. By (3.14) and (3.20) we have

$$\begin{aligned} &\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_v \\ &\qquad \leq \|F(v;\tau)\|_v + \Delta\tau\,\|F_v(v;\tau)p + F_\tau(v;\tau)\|_v + \beta\Delta\tau^2 \\ &\qquad \leq \frac{1}{\rho}(\|F(v;\tau)\|_v + \Delta\tau\,\|F_v(v;\tau)p + F_\tau(v;\tau)\|_v + \rho\beta\Delta\tau^2) \\ &\qquad \leq \frac{1}{\rho}(\|F(v;\tau)\|_v + \Delta\tau\,\|F_v(v;\tau)p + F_\tau(v;\tau)\|_v + [\beta]\Delta\tau^2) \\ &\qquad \leq \frac{1}{\rho}\frac{\rho^3\delta_{\mathrm{TOL}}}{1+[\gamma]\Delta\tau\,\|F_v(v;\tau)p\|_v} \\ &\qquad \leq \frac{\rho^2\delta_{\mathrm{TOL}}}{1+[\gamma]\Delta\tau\,\|F_v(v;\tau)p\|_v}\,. \end{aligned} \tag{3.21}$$

Thus, the reduction test 1 is satisfied. Because of

$$\begin{aligned} &\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p} \\ &\qquad \leq (1+\gamma\Delta\tau\,\|F_v(v;\tau)p\|_v)\,\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_v \\ &\qquad \leq \rho^{-1}(1+[\gamma]\Delta\tau\,\|F_v(v;\tau)p\|_v)\,\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_v \end{aligned} \tag{3.22}$$

the reduction test 2 is passed, too. Combining (3.21) and (1) we have

$$\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p} \leq \rho\delta_{\mathrm{TOL}} \leq \frac{2\rho}{[\omega]} < \frac{2}{\omega}\,.$$

The predicted point is inside the local convergence domain of the Newton corrector, such that the final reduction test 3 is satisfied. □

Now we can prove Theorem 3.4.2.

Proof. First we show that whenever the reduction test 1 fails, a finite number of step size reductions suffices to pass the test. If the test fails for the step size defined by (3.15), i.e.

$$\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_v > \rho^2\delta_{\mathrm{TOL}}\,,$$

the new estimate $[\beta]_{\text{new}}$ for β computed from (3.18) is

$$\begin{aligned}
[\beta]_{\text{new}} &> \frac{1}{\Delta\tau^2}(\rho^2\delta_{\text{TOL}} - \|F(v;\tau)\|_v - \Delta\tau\|r\|_v) \\
&\geq \frac{1}{\Delta\tau^2}\left(\frac{\rho^3\delta_{\text{TOL}}}{1+[\gamma]\Delta\tau\|F_v(v;\tau)p\|_v} - \|F(v;\tau)\|_v - \Delta\tau\|r\|_v\right) \\
&\quad + (1-\rho)\frac{\rho^2\delta_{\text{TOL}}}{\Delta\tau^2} \\
&= [\beta] + (1-\rho)\frac{\rho^2\delta_{\text{TOL}}}{\Delta\tau^2}\,.
\end{aligned}$$

Thus, $[\beta]$ is increased by an amount that is bounded from below, which implies that a finite number of step size reductions suffices to satisfy the passing condition $\rho\beta \leq [\beta]$.

If, on the other hand, the first test is passed but the second test fails, i.e.

$$\begin{aligned}
&\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p} \\
&\qquad > \rho^{-1}(1+[\gamma]\Delta\tau\|F_v(v;\tau)p\|_v)\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_v\,,
\end{aligned}$$

the new estimate $[\gamma]_{\text{new}}$ for γ computed from (3.13) is

$$\begin{aligned}
[\gamma]_{\text{new}} &> (\Delta\tau\|F_v(v;\tau)p\|_v)^{-1} \\
&\quad \cdot \Big(\frac{\rho^{-1}(1+[\gamma]\Delta\tau\|F_v(v;\tau)p\|_v)\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_v}{\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_v} - 1\Big) \\
&= \frac{\rho^{-1}-1}{\Delta\tau\|F_v(v;\tau)p\|_v} + \rho^{-1}[\gamma]\,,
\end{aligned}$$

which is implies that $[\gamma]$ is geometrically increasing. Actually, the updated estimate $[\beta]$ may be smaller than before, but the reduction test 1 is guaranteed to pass again after a finite number of reductions.

Finally, if the tests 1 and 2 are satisfied but test 3 fails, i.e.

$$\|F(v+\Delta\tau p+\delta v;\tau+\Delta\tau)\|_{v+\Delta\tau p} \geq \|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p}\,,$$

the new estimate $[\omega]_{\text{new}}$ for ω computed from (3.11) is

$$\begin{aligned}
[\omega]_{\text{new}} &\geq \frac{2(1-\delta)}{(1+\delta)^2\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p}} \\
&= \frac{2(\Theta-\delta+(1-\Theta))}{(1+\delta)^2\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p}} \\
&\geq \frac{2((1+\delta)^2\frac{[\omega]}{2}\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p}+\delta-\delta+(1-\Theta))}{(1+\delta)^2\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p}} \\
&= [\omega] + \frac{2(1-\Theta)}{(1+\delta)^2\|F(v+\Delta\tau p;\tau+\Delta\tau)\|_{v+\Delta\tau p}}\,,
\end{aligned}$$

which is greater than $[\omega]$ by an amount that is bounded from below independently of $[\omega]$, $[\gamma]$, and $[\beta]$. Actually, the updated estimates $[\gamma]$ and $[\beta]$ may be smaller than before, but the reduction tests 1 and 2 are guaranteed to pass again after a finite number of reductions. □

So far the step size selection relies on the quality of the path curvature estimate (3.14). When applied to the complementarity formulation (2.9), the estimate is good only as long as the predictor does not cross the bend of the complementarity function ψ near the origin, which is especially sharp for small μ. Therefore it is highly advisable to restrict the predictor step in order to stay in the feasible region if possible. Although this additional restriction is not required for the algorithm to work, it helps to stabilize the step size selection mechanism a lot.

3.5 Accuracy Matching

Within the computationally available bounds (3.19) and (3.20), the predictor step size and accuracy and the corrector termination criterion can be chosen arbitrarily. In this section we will develop a simplified model of the computation process such that we are able to determine algorithmic parameters maximizing the overall efficiency, i.e. the information gain per unit work:

$$\text{maximize } Q := \frac{\Delta\tau}{\text{work}}$$

In order to achieve the maximal step size per unit work, the whole cycle of prediction and correction has to be taken into account. This cycle is characterized by three quantities:

- the norm $n_P := \|F(v_P;\tau)\|_v$ at the predicted point v_P, corresponding to the predictor target $\rho\delta_{\text{TOL}}$
- the norm $n_C := n_{C_j} := \|F(v_{C_j};\tau)\|_v$ at the subsequent corrector result v_{C_j} obtained by j Newton steps, corresponding to the corrector termination criterion
- the norm $n_R := \|F(v_{C_{j-1}};\tau) + F_v(v_{C_{j-1}};\tau)\delta v_{C_{j-1}}\|_v = \delta_{C_{j-1}} n_{C_{j-1}}$ of the last inner residual, corresponding to the relative accuracy $\delta_{C_{j-1}}$ imposed on the solution of the last Newton corrector

To ease the derivation, we will neglect the difference of the local norms and use a fixed norm $\|\cdot\|$ instead, thus assuming $\gamma = 0$.

In the context of adaptive methods for differential equations, the inner residual is essentially the discretization error, and the amount of work can be assumed to be proportional to the mesh size when using optimal solvers. We make the simplistic assumption that both predictor and corrector are computed using the same discretization. In contrast to pure inexact Newton methods, all corrector steps use the discretization that stems from the previous continuation cycle and that is assumed to be sufficiently fine to represent the Newton corrector steps δv_{C_i}, $i = 0, \ldots, j-1$, with the required accuracy δ_{C_i}. The size of the common mesh and hence the work required for one Newton step is asymptotically proportional to $n_R^{-1/q}$, where q denotes the approximation order of the discretization scheme. The convergence of the Newton corrector is quadratic until the accuracy gain is limited by the discretization error, at which time the corrector iteration reaches the requested accuracy and is terminated. Therefore, the number of corrector steps can be estimated as

$$j \approx \left\lceil \frac{\log\log\frac{\omega n_C}{2} - \log\log\frac{\omega n_P}{2}}{\log 2} \right\rceil .$$

Remark 3.5.1. Of course it is possible to start every corrector iteration on a coarse discretization and save some work during the first few corrector steps, but this imposes the difficulty that an error estimator may fail to detect the need for refinement when the error is visible only on a much finer grid than the coarse grid, terminating the corrector iteration on prematurely coarse grids. ◁

Since in addition to the corrector one predictor step is performed, the complete work for one cycle is

$$\text{work} \sim n_R^{-\frac{1}{q}} \left\lceil 1 + \frac{\log\log\frac{\omega n_C}{2} - \log\log\frac{\omega n_P}{2}}{\log 2} \right\rceil .$$

The largest possible predictor step size is

$$\Delta\tau = \frac{1}{2\beta}\left(\sqrt{\|r\|^2 + 4\beta(n_P - n_C)} - \|r\|\right) . \tag{3.23}$$

Since both computations are done on the same discretization, we assume that $\|r\| \approx n_R$. Thus, we end up with the theoretical efficiency indicator

$$Q = \frac{\sqrt{n_R^2 + 4\beta(n_P - n_C)} - n_R}{\left\lceil 1 + \frac{\log\log\frac{\omega n_C}{2} - \log\log\frac{\omega n_P}{2}}{\log 2} \right\rceil} n_R^{\frac{1}{q}} .$$

A short visual inspection shows that the choice of $j = 1$, i.e. one Newton corrector step, is optimal over a reasonably wide range of the involved parameters n_P, n_C, n_R, β, ω, and q, except for nearly straight paths ($\beta \ll 1$) and predictors very close to the boundary of the corrector's domain of convergence. In these cases, the choice of more Newton corrector steps may be more efficient, although the difference is in general not significant. Furthermore, small

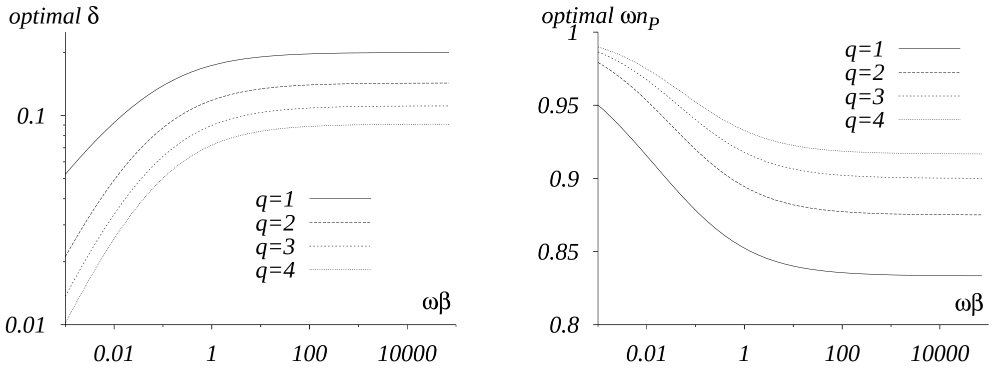

Figure 3.2: Theoretically optimal values of δ (left) and ωn_P (right) for $q = 1, \dots, 4$ plotted versus $\omega\beta$.

β are more difficult to estimate correctly and predictors near the boundary of the domain of convergence are unfavorable because of robustness reasons. The standard choice will therefore be $j = 1$, leading to the efficiency indicator

$$Q = n_R^{\frac{1}{q}} \left(\sqrt{n_R^2 + 4\beta(n_P - n_C)} - n_R \right) .$$

With $j = 1$ we have $n_R = \delta n_P$ and, assuming equality in (3.6), $n_C = ((1+\delta)^2 \frac{\omega}{2} n_P + \delta) n_P$:

$$Q = (\delta n_P)^{\frac{1}{q}} \left(\sqrt{(\delta n_P)^2 + 4\beta n_P \Big(1 - \frac{1}{2}(1+\delta)^2 \omega n_P - \delta\Big)} - \delta n_P \right)$$

This theoretical efficiency indicator may now easily be maximized numerically with respect to δ and n_P for relevant ranges of β and ω and different q. As it turns out, the optimal parameters δ and n_P do depend only on q and the product $\omega\beta$, but not on the quotient $\frac{\beta}{\omega}$. The optimal values of δ and ωn_P are plotted versus $\omega\beta$ in Figure 3.2 for different values of q. In the numerical examples in Section 4 we use the even simpler constant choice $\delta = 0.1$ and $n_P = 0.9$.

3.6 Affine Invariant Norm

In the light of Section 2.3 we have to work out two different norms, a coarse norm in which convergence of the path is measured, and a fine norm required for continuous differentiability.

Solving optimization problems, the actual error $\|x - x^*\|$ of an approximate solution x is in general of little interest. Instead, solution points are naturally characterized by feasibility ($c(x) = 0$ and $g(x) \geq 0$) and optimality ($J(x) - J(x^*) = 0$). The meaningful transformations under which the algorithm is expected to be invariant must therefore leave the constraints untouched, such that an appropriate norm of their violation can be taken. Since the dual space is directly coupled to the constraints space, the Lagrange multipliers will not be transformed. Furthermore, the cost functional values must not be transformed, such that cost improvements can be appropriately measured.

The remaining transformations are just the transformations of the domain space X, and, correspondingly, the dual space X^*. Transforming the domain space as $B\xi = x$

$$\min J(B\xi) \quad \text{subject to} \quad \begin{array}{c} c(B\xi) = 0 \\ g(B\xi) \geq 0 \end{array}$$

transforms F by the chain rule as $B_4^* F(B_4 \phi)$, where

$$B_4 := \begin{bmatrix} B & & & \\ & I & & \\ & & I & \\ & & & I \end{bmatrix} \text{ and } B_4 \phi = v\,.$$

The derivative of F is then transformed as $B_4^* F(B_4\phi) B_4$. Since the adjoint equation

$$J'(x) - c'(x)^* \lambda - g'(x)^* \eta = 0\,.$$

is transformed, the canonical norm of F in Z is not invariant. Instead, an invariant norm has to be constructed explicitly.

3.6.1 A Norm for Equality Constrained Problems

The elimination of the inequality constraints from the complementarity system as in the proof of Theorem 2.5.4 yields a pure saddle point operator

$$S(v) := \begin{bmatrix} H(v) & -c'(x)^* \\ -c'(x) & \end{bmatrix}$$

with the same structure as a Kuhn-Tucker system for an equality constrained optimization problem. For those problems an affine invariant norm has been developed recently in the context of augmented Lagrangian SQP methods [51].

If we assume that the modified Hessian

$$H(v) := L_{xx}(x, \lambda, \eta) + g'(x)^* \Psi_\eta(w, \eta; \mu)^{-1} \Psi_w(w, \eta; \mu) g'(x)$$

is positive definite on the nullspace of the equality constraints $c'(x)$ as in Theorem 2.5.4, we can utilize the induced norm on $\ker c'$ to construct an affine invariant norm on $\widehat{X^*}$.

Let us introduce the bounded linear operator $T(v) : \ker c'(v) \times \Lambda_2 \to X_2^*$ by

$$T(v) := \begin{bmatrix} H(v) & -c'(x)^* \end{bmatrix} .$$

Lemma 3.6.1. *Under the assumptions of Theorem 2.5.4, $T(v)$ is an isomorphism.*

Proof. For arbitrary $r \in X_2^*$ the equation $T(v)\xi = r$ for $\xi = (\xi_1, \xi_2)^T \in \ker c'(v) \times \Lambda_2$ is equivalent to

$$S \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} = \begin{bmatrix} r_1 \\ 0 \end{bmatrix} .$$

Because of the assumptions 1 and 3 of Theorem 2.5.4, c' satisfies the inf-sup-condition and H is positive definite on $\ker c'$, such that S is an isomorphism (cf. [8]). Thus, ξ is uniquely determined and $\|\xi\|_{X\times\Lambda_2} \le \text{const}\,\|r\|_{X_2^*}$. □

Additionally, we define the bounded, linear, symmetric positive definite operator $R(v) : \ker c'(v) \times \Lambda_2 \to X_2^* \times \Lambda_2$ by

$$R(v) := \begin{bmatrix} H(v) & \\ & I_R \end{bmatrix} .$$

Here, $I_R : \Lambda_2 \to \Lambda_2^*$ denotes the Riesz isomorphism.

Using R and T, we can now define an affine invariant norm.

Lemma 3.6.2. *The mapping $|||\cdot|||_v : X_2^* \to \mathbb{R}$ defined by*

$$|||r|||_v^2 := \langle R(v)T(v)^{-1}r, T(v)^{-1}r\rangle \tag{3.24}$$

is an affine invariant norm, which is equivalent to the canonical norm on X_2^*.

Proof. Since $R(v)$ is coercive and $T(v)$ an isomorphism, $|||\cdot|||_v$ defines indeed a norm which is equivalent to the canonical norm on X_2^*:

$$\begin{aligned} \text{const}\,\|r\|_{X_2^*}^2 \le \text{const}\,\|T(v)^{-1}r\|_{X_2^*\times\Lambda_2}^2 & \\ \le \langle R(v)T(v)^{-1}r, T(v)^{-1}r\rangle & \\ \le \text{const}\,\|T(v)^{-1}r\|_{X_2^*\times\Lambda_2}^2 \le \text{const}\,\|r\|_{X_2^*}^2 & \end{aligned}$$

for all $r \in X_2^*$.

Note that H transforms as

$$\begin{aligned} B^*L_{xx}(B\xi,\lambda,\eta)B + (g'(B\xi)B)^*\Psi_\eta(w,\eta;\mu)^{-1}\Psi_w(w,\eta;\mu)g'(B\xi)B & \\ = B^*H(B_4\phi)B & \end{aligned}$$

and $c'(x)$ as $c'(B\xi)B$. Thus, $R(v)$ and $T(v)$ transform as

$$B_2^* R(B_4\phi)B_2 = B_2^* R(v)B_2$$

and

$$B^* T(B_4\phi)B_2 = B^* T(v)B_2\,,$$

respectively, where

$$B_2 := \begin{bmatrix} B & \\ & I \end{bmatrix}.$$

The norm defined above is invariant, since the transformations cancel out:

$$\begin{aligned}
|||B^*r|||^2_{B_4\phi} &= \langle B_2^* R(B_4\phi)B_2(B^*T(B_4\phi)B_2)^{-1}B^*r, (B^*T(B_4\phi)B_2)^{-1}B^*r\rangle \\
&= \langle B_2^* R(v)B_2 B_2^{-1}T(v)^{-1}B^{-*}B^*r, B_2^{-1}T(v)^{-1}B^{-*}B^*r\rangle \\
&= \langle B_2^* R(v)T(v)^{-1}r, B_2^{-1}T(v)^{-1}r\rangle \\
&= \langle R(v)T(v)^{-1}r, T(v)^{-1}r\rangle \\
&= |||r|||^2_v\,. \qquad \square
\end{aligned}$$

Note that the equivalence between the invariant norm and the canonical norm is not uniform for $v \in D$.

In order to define a norm on the whole image space Z_2 we take additionally the norms of the equality constraints, the slacks, and the complementarity equations, since these are not transformed. Together with the norm (3.24) on $X_2^* \times \Lambda_2^*$ they define the following affine invariant norm on Z:

$$|||(z_1, z_2, z_3, z_4)^T|||^2_v := |||z_1|||^2_v + \|z_2\|^2_{\Lambda_2} + \|z_3\|^2_{W_2} + \|z_4\|^2_{W_2}\,.$$

Remark 3.6.3. Note that albeit its nontrivial definition, the norm (3.24) is comparatively easy to compute. By definition of $R(v)$ and $T(v)$ it is clear that

$$|||z_1|||^2_v = \langle H(v)\xi, \xi\rangle + \|l\|^2_{\Lambda_2}$$

where

$$S(v)\begin{bmatrix} \xi \\ l \end{bmatrix} = \begin{bmatrix} z_1 \\ 0 \end{bmatrix}. \tag{3.25}$$

By Step 1 in the proof of Theorem 2.5.4, ξ and l can be computed as solution of

$$F_v(v)(\xi, l, s_1, s_2)^T = (z_1, 0, 0, 0)^T\,,$$

where the s_1 and s_2 components are ignored.

Thus, computation of the affine invariant norm requires the solution of one more linear system with the same operator and a different right hand side. In actual computation, the required discretization $F_v(v)$ will already be assembled for computation of δv. In case a factorization of $F_v(v)$ is available, it can be reused to compute $|||\cdot|||_v$ with negligible cost. ◁

Concerning the finer norm $\|\cdot\|_v$ that is required for the Newton theory, we note that under the assumptions of Theorem 2.5.4

$$\langle H\xi, \xi\rangle = \int_\Omega \xi(t)^T H(t)\xi(t)\,dt$$

and

$$\xi(t)^T H(t)\xi(t) \geq \text{const}\,|\xi(t)|^2 .$$

Thus the norm

$$\|(z_1, z_2, z_3, z_4)^T\|_v := \|z_1\|_v^2 + \|z_2\|_{\Lambda_\infty}^2 + \|z_3\|_{W_\infty}^2 + \|z_4\|_{W_\infty}$$

where

$$\|z_1\|_v^2 := \|\phi\|_{L_\infty} + \|l\|_{\Lambda_\infty}^2 \quad \text{and} \quad \phi(t) := \xi(t)^T H(t)\xi(t)$$

is equivalent to $\|\cdot\|_{V_\infty}$. Moreover, $\|\cdot\|_v$ forms a γ-continuous family of invariant norms due to the Lipschitz continuity (2.19) of F_v.

Corollary 3.6.4. *Under the assumptions of Theorem 2.5.4, the complementarity formulation* (2.9) *satisfies the affine invariant Lipschitz condition* (3.5)*:*

$$\|(F_v(v+\delta v;\mu) - F_v(v;\mu))\delta v\|_v \leq \omega \|F_v(v;\mu)\delta v\|_v^2 .$$

Proof. The corollary is a direct consequence of the Lipschitz continuity of F_v asserted by Theorem 2.5.2 and the equivalence of the norms $\|\cdot\|_v$ and $\|\cdot\|_{V_\infty}$. □

3.6.2 A Norm for Inequality Constrained Problems

The norm defined in (3.24) suffers from the increasing ill-conditioning of the reduced complementarity saddle point operator $H(v(\mu))$ for $\mu \to 0$. This ill-conditioning has been introduced by the elimination step from (2.21) to (2.22), after it had been avoided by the introduction of Lagrange multipliers for the barrier formulation. To overcome this difficulty, we can resort to the splitting of the inequality constraints into nearly active and nearly inactive constraints introduced in Definition 2.6.1. Thus we can exploit a more careful elimination as has been used in Theorem 2.6.5.

To this extent we combine equality constraints and nearly active inequality constraints by setting $\Lambda_\rho := \Lambda_2^* \times W_A$ and defining $C^\rho(x) : X_2 \to \Lambda_\rho^*$, $D^\rho(v) : \Lambda_\rho \to \Lambda_\rho^*$, and $H^\rho(v) : X_2 \to X_2^*$ as

$$C^\rho(x) := -\begin{bmatrix} c'(x) \\ g_A'(x) \end{bmatrix}, \quad D^\rho(v) := \begin{bmatrix} 0 & 0 \\ 0 & \Psi_w^A(w^A, \eta^A;\mu)^{-1}\Psi_\eta^A(w^A, \eta^A;\mu) \end{bmatrix},$$

and

$$H^\rho(v) := L_{xx}(v) + g_I'(x)^* \Psi_\eta^I(w^I, \eta^I; \mu)^{-1} \Psi_w^I(w^I, \eta^I; \mu) g_I'(x)$$

which yields the perturbed saddle point operator

$$S^\rho(v) := \begin{bmatrix} H^\rho(v) & C^\rho(x)^* \\ C^\rho(x) & -D^\rho(v) \end{bmatrix} .$$

Then we can substitute $T(v)$ and $R(v)$ by their counterparts $T^\rho(v) : X_\rho(v) \times \Lambda_\rho \to X_2^*$,

$$T^\rho(v) := \begin{bmatrix} H^\rho(v) & C^\rho(x)^* \end{bmatrix} ,$$

and $R^\rho(v) : X_\rho(v) \times \Lambda_\rho \to X_2^* \times \Lambda_\rho^*$,

$$R^\rho(v) := \begin{bmatrix} H^\rho(v) & \\ & I_R \end{bmatrix} .$$

Because of the perturbation $D^\rho(v)$ in $S^\rho(v)$ we have to resort to

$$X_\rho(v) := S^\rho(v)^{-1}(X_2^* \times \{0_{\Lambda_\rho}\})$$

instead of $\ker c'(x)$ as before. Note that $X_\rho(v)$ is almost $\ker c'(x)$, especially for $\mu \to 0$, since $(\Psi_w^A)^{-1}\Psi_\eta^A$ tends to be small. This justifies the additional assumption of the following theorem, the proof of which is completely analogous to the proofs in the previous section.

Theorem 3.6.5. *In addition to the assumptions of Theorem 2.6.5, suppose that $H^\rho(v)$ is positive definite on X_ρ. Then the mapping $|||\cdot|||_v : X_2^* \to \mathbb{R}$ defined by*

$$|||r|||_v^2 := \langle R^\rho(v) T^\rho(v)^{-1} r, T^\rho(v)^{-1} r \rangle$$

is an affine invariant norm, which is equivalent to the canonical norm on X_2^*.

Again, efficiently computable norms on Z_2 and Z_∞ can be defined by

$$\begin{aligned} |||(z_1, z_2, z_3, z_4)^T|||_v^2 &:= |||z_1|||_v^2 + +\|z_2\|_{\Lambda_2}^2 + \|z_3\|_{W_2}^2 + \|z_4\|_{W_2}^2 \\ &= \langle H^\rho(v)\xi, \xi\rangle + \|l\|_{\Lambda_\rho}^2 + \|z_2\|_{\Lambda_2}^2 + \|z_3\|_{W_2}^2 + \|z_4\|_{W_2}^2 \end{aligned} \tag{3.26}$$

and

$$\|(z_1, z_2, z_3, z_4)^T\|_v^2 := \|\phi\|_{L_\infty} + \|l\|_{\Lambda_\rho}^2 + \|z_2\|_{\Lambda_2}^2 + \|z_3\|_{W_2}^2 + \|z_4\|_{W_2}^2 , \tag{3.27}$$

respectively, where

$$S^\rho(v) \begin{bmatrix} \xi \\ l \end{bmatrix} = \begin{bmatrix} z_1 \\ 0 \end{bmatrix} \quad \text{and} \quad \phi(t) := \xi(t)^T H^\rho(v)(t)\xi(t) .$$

Note that in general the norms $\|\cdot\|_v$ do not form a γ-continuous family, except for fixed nearly active sets. This must be taken into account when using the norm (3.27) for the continuation procedure.

For the special choice of $\rho = 1$ in the case of control constraints only, however, a slightly weaker continuity property of the coarser norm $|\!|\!|\cdot|\!|\!|_v$ can be shown.

Lemma 3.6.6. *Using 1-nearly active sets and the Fischer-Burmeister complementarity function* (2.7)*, there exists a constant γ, such that*

$$|\!|\!|z|\!|\!|_{v_1} \le (1 + \gamma \|F_v(v_1;\mu)(v_1 - v_2)\|_{v_1}) |\!|\!|z|\!|\!|_{v_2}$$

holds.

Proof. For $z = (z_1, z_2, z_3, z_4)^T \in Z_2$,

$$F_v(v_i;\mu)\begin{bmatrix}\xi_{v_i}\\ \lambda_{v_i}\\ \eta_{v_i}\\ w_{v_i}\end{bmatrix} = \begin{bmatrix}z_1\\0\\0\\0\end{bmatrix}, \quad \text{and} \quad l_{v_i} = \begin{bmatrix}\lambda_{v_i}\\ \eta^A_{v_i}\end{bmatrix} \quad \text{for } i = 1, 2 \tag{3.28}$$

consider

$$\begin{aligned}
&|\!|\!|z|\!|\!|^2_{v_1} - |\!|\!|z|\!|\!|^2_{v_2}\\
&\qquad = \langle H^\rho(v_1)\xi_{v_1}, \xi_{v_1}\rangle - \langle H^\rho(v_2)\xi_{v_2}, \xi_{v_2}\rangle + \|l_{v_1}\|^2_{\Lambda_\rho(v_1)} - \|l_{v_2}\|^2_{\Lambda_\rho(v_2)}\\
&\qquad = \langle L_{xx}(v_1)\xi_{v_1}, \xi_{v_1}\rangle - \langle L_{xx}(v_2)\xi_{v_2}, \xi_{v_2}\rangle\\
&\qquad\quad + \langle \mathcal{D}_I(v_1)\xi_{v_1}, \xi_{v_2}\rangle - \langle \mathcal{D}_I(v_2)\xi_{v_2}, \xi_{v_2}\rangle\\
&\qquad\quad + \|\lambda_{v_1}\|^2_{\Lambda_2} - \|\lambda_{v_2}\|^2_{\Lambda_2}\\
&\qquad\quad + \|\eta^A_{v_1}\|^2_{W_{2,A(v_1)}} - \|\eta^A_{v_2}\|^2_{W_{2,A(v_2)}},
\end{aligned}$$

where

$$\mathcal{D}(v) := g'_I(x)^* \Psi^I_\eta(v)^{-1}\Psi^I_w(v) g'_I(x).$$

The first and third difference in the right hand side sum do not depend on the partitioning of Ω in active and inactive sets. Therefore, they can be handled by Lipschitz continuity of F_v as in Corollary 3.6.4. The remaining terms are investigated below. Let $\Omega^i_I := \Omega^\rho_I(v_i)$ for $i = 1, 2$. First consider

$$\begin{aligned}
&\langle \mathcal{D}_I(v_1)\xi_{v_1}, \xi_{v_1}\rangle - \langle \mathcal{D}_I(v_2)\xi_{v_2}, \xi_{v_2}\rangle\\
&= \int_{\Omega^1_I} \xi^T_{v_1}\mathfrak{g}'(x_1)^T \frac{\psi_w(v_1)}{\psi_\eta(v_1)}\mathfrak{g}'(x_1)\xi_{v_1}\,dt - \int_{\Omega^2_I} \xi^T_{v_2}\mathfrak{g}'(x_2)^T \frac{\psi_w(v_2)}{\psi_\eta(v_2)}\mathfrak{g}(v_2)'\xi_{v_2}\,dt\\
&= \int_{\Omega^1_I\setminus\Omega^2_I} \xi^T_{v_1}\mathfrak{g}'(x_1)^T \frac{\psi_w(v_1)}{\psi_\eta(v_1)}\mathfrak{g}'(x_1)\xi_{v_1}\,dt - \int_{\Omega^2_I\setminus\Omega^1_I} \xi^T_{v_2}\mathfrak{g}'(x_2)^T \frac{\psi_w(v_2)}{\psi_\eta(v_2)}\mathfrak{g}(v_2)'\xi_{v_2}\,dt\\
&\quad + \int_{\Omega^1_I\cap\Omega^2_I} \left(\xi^T_{v_1}\mathfrak{g}'(x_1)^T \frac{\psi_w(v_1)}{\psi_\eta(v_1)}\mathfrak{g}'(x_1)\xi_{v_1} - \xi^T_{v_2}\mathfrak{g}'(x_2)^T \frac{\psi_w(v_2)}{\psi_\eta(v_2)}\mathfrak{g}(v_2)'\xi_{v_2}\right)dt\,.
\end{aligned}$$

The third integral can again be handled by the Lipschitz continuity of the involved functions. From (3.28) and (2.25) we infer

$$\psi_w(v_i)\mathfrak{g}'(x_i)\xi_{v_i} + \psi_\eta(v_i)\eta_{v_i} = 0$$

and thus

$$\begin{aligned}\int_{\Omega_I^1\setminus\Omega_I^2} &\xi_{v_1}^T \mathfrak{g}'(x_1)^T \frac{\psi_w(v_1)}{\psi_\eta(v_1)} \mathfrak{g}'(x_1)\xi_{v_1}\, dt \\ &= \int_{\Omega_I^1\setminus\Omega_I^2} \eta_{v_1}^T \frac{\psi_\eta(v_1)}{\psi_w(v_1)} \frac{\psi_w(v_1)}{\psi_\eta(v_1)} \frac{\psi_\eta(v_1)}{\psi_w(v_1)} \eta_{v_1}\, dt = \int_{\Omega_I^1\setminus\Omega_I^2} \eta_{v_1}^T \frac{\psi_\eta(v_1)}{\psi_w(v_1)} \eta_{v_1}\, dt\,.\end{aligned}$$

Let $r := \|z\|_{v_1}^2 \|F_v(v_1;\mu)(v_1-v_2)\|_{v_1}$. On $\Omega_I^1\setminus\Omega_I^2$ we have $w_1 \geq \eta_1$ and $w_2 \leq \eta_2$, such that there is some $\bar v$ with $\bar w = \bar\eta$ on $\Omega_I^1\setminus\Omega_I^2$ and

$$\|\bar v - v_1\|_{V_\infty} \leq \|v_1 - v_2\|_{V_\infty} \leq \text{const}\,\|F_v(v_1;\mu)(v_1-v_2)\|_{v_1}\,.$$

Then,

$$\begin{aligned}\int_{\Omega_I^1\setminus\Omega_I^2} \xi_{v_1}^T \mathfrak{g}'(x_1)^T \frac{\psi_w(v_1)}{\psi_\eta(v_1)} \mathfrak{g}'(x_1)\xi_{v_1}\, dt &= \int_{\Omega_I^1\setminus\Omega_I^2} \eta_{v_1}^T \frac{\psi_w(\bar v)}{\psi_\eta(\bar v)} \eta_{v_1}\, dt + \mathcal{O}(r) \\ &= \int_{\Omega_I^1\setminus\Omega_I^2} \eta_{v_1}^T \eta_{v_1}\, dt + \mathcal{O}(r)\end{aligned}$$

because of

$$\frac{\psi_w(\bar v)}{\psi_\eta(\bar v)} = \frac{\sqrt{\bar w^2 + \bar\eta^2 + 2\mu} - \bar w}{\sqrt{\bar w^2 + \bar\eta^2 + 2\mu} - \bar\eta} = 1\,.$$

Consequently we have

$$\langle \mathcal{D}_I(v_1)\xi_{v_1}, \xi_{v_1}\rangle - \langle \mathcal{D}_I(v_2)\xi_{v_2}, \xi_{v_2}\rangle = \int_{\Omega_I^1\setminus\Omega_I^2} \eta_{v_1}^T\eta_{v_1}\, dt - \int_{\Omega_I^2\setminus\Omega_I^1} \eta_{v_2}^T\eta_{v_2}\, dt + \mathcal{O}(r)\,.$$

On the other hand,

$$\begin{aligned}\|\eta_{v_1}^A\|_{W_{2,A}(v_1)}^2 - \|\eta_{v_2}^A\|_{W_{2,A}(v_2)}^2 &= \int_{\Omega\setminus\Omega_I^1} \eta_{v_1}^T\eta_{v_1}\, dt - \int_{\Omega\setminus\Omega_I^2} \eta_{v_2}^T\eta_{v_2}\, dt \\ &= \int_{\Omega_I^2\setminus\Omega_I^1} \eta_{v_1}^T\eta_{v_1}\, dt - \int_{\Omega_I^1\setminus\Omega_I^2} \eta_{v_2}^T\eta_{v_2}\, dt \\ &\quad + \int_{\Omega\setminus(\Omega_I^1\cap\Omega_I^2)} (\eta_{v_1}^T\eta_{v_1} - \eta_{v_2}^T\eta_{v_2})\, dt\,,\end{aligned}$$

such that

$$\begin{aligned}\langle \mathcal{D}_I(v_1)\xi_{v_1}, \xi_{v_1}\rangle - \langle \mathcal{D}_I(v_2)\xi_{v_2}, \xi_{v_2}\rangle + \|\eta^A_{v_1}\|^2_{W_{2,A(v_1)}} - \|\eta^A_{v_2}\|^2_{W_{2,A(v_2)}} \\ = \int_{(\Omega_I^1\setminus\Omega_I^2)\cup(\Omega_I^2\setminus\Omega_I^1)} (\eta_{v_1}^T\eta_{v_1} - \eta_{v_2}^T\eta_{v_2})\, dt + \mathcal{O}(r) \\ = \mathcal{O}(r)\end{aligned}$$

and

$$\|z\|^2_{v_1} - \|z\|^2_{v_2} = \mathcal{O}(r)\,.$$

Finally we introduce a constant γ and conclude

$$\begin{aligned} & \|z\|^2_{v_1} - \|z\|^2_{v_2} \le \gamma\|z\|^2_{v_1}\|F_v(v_1;\mu)(v_1 - v_2)\|_{v_1} \\ \Rightarrow\quad & \|z\|^2_{v_1} \le (1 + \gamma\|F_v(v_1;\mu)(v_1 - v_2)\|_{v_1})^2\|z\|^2_{v_2} \\ \Rightarrow\quad & \|z\|_{v_1} \le (1 + \gamma\|F_v(v_1;\mu)(v_1 - v_2)\|_{v_1})\|z\|_{v_2}\,. \end{aligned}$$

□