Chapter 5
Variational Problems: Globalization of Convergence

Newtonian methods for variational problems discussed in Chap. 3 have attractive local convergence and rate of convergence properties, but they are local by nature: for guaranteed convergence, they all require a starting point close enough to a solution. Therefore, to arrive to practical algorithms based on these methods, the process of computing and automatically accepting an appropriate "starting point" must be incorporated into the overall iterative scheme. Such extensions of locally convergent methods are referred to as globalization of convergence. This chapter discusses some strategies for globalization of convergence of Newtonian methods for variational problems.

5.1 Linesearch Methods

As seen in Sect. 2.2, descent methods for unconstrained optimization possess natural global convergence properties. Moreover, linesearch quasi-Newton methods combine global convergence of descent methods with high convergence rate of the Newton method. It is then natural to extend this idea to other problem settings. This requires relating the problem in question to unconstrained optimization, in some meaningful way. This task is usually approached by constructing a function measuring the quality of approximations to a solution of the original problem, called a merit function. Then, given a direction produced by a Newtonian method, we can perform linesearch in this direction, evaluating the candidate points using the chosen merit function. In the case of unconstrained optimization, the natural merit function is the objective function of the problem. For constrained optimization and variational problems, the choice of a merit function is not evident and is certainly not unique. Some possibilities will be discussed in this and the next chapters.

Essentially, the idea of linesearch globalization is to reduce the size of the Newtonian step when the full step does not provide a sufficient decrease for

A.F. Izmailov and M.V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-3-319-04247-3_5,

values of the chosen merit function. If the resulting algorithm turns out to be a descent method for the merit function, one can expect global convergence (in some sense). If, in addition, the stepsize is reduced only far from solutions (solutions satisfying certain properties, of course), this would imply that the algorithm asymptotically turns into the full-step Newtonian method possessing high convergence rate. For linesearch quasi-Newton methods for unconstrained optimization, this ideal combination of convergence properties is achieved in Theorems 2.24–2.26. However, as naturally expected and seen below, for more general and different problem settings the situation is more complex.

5.1.1 Globalized Newton Method for Equations

We start with considering linesearch methods for the usual equation

$$\Phi(x) = 0, \tag{5.1}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping. As discussed in Sect. 2.2.3, the natural choice of a merit function $\varphi : \mathbf{R}^n \to \mathbf{R}$ for (5.1) is the squared Euclidean residual

$$\varphi(x) = \frac{1}{2}\|\Phi(x)\|^2. \tag{5.2}$$

Moreover, for any $x \in \mathbf{R}^n$ which is not a solution of (5.1), if $\Phi'(x)$ is a nonsingular matrix, then the well-defined Newtonian direction

$$p = -(\Phi'(x))^{-1}\Phi(x) \tag{5.3}$$

is a direction of descent for this merit function at x, i.e., $p \in \mathcal{D}_\varphi(x)$. In Sect. 2.2.3 also some perturbed counterparts of the Newtonian direction p are discussed, which are relevant in the context of linesearch methods.

Here we only consider the basic choice (5.3). One special feature of this direction is that its quality as a descent direction can be readily estimated via the residual of (5.1):

$$\langle \varphi'(x),\, p\rangle = \langle (\Phi'(x))^{\mathrm{T}}\Phi(x),\, p\rangle = \langle \Phi(x),\, \Phi'(x)p\rangle = -\|\Phi(x)\|^2. \tag{5.4}$$

In particular, $\langle \varphi'(x),\, p\rangle$ may become close to zero (so that p is not a "good" descent direction) only when $\Phi(x)$ is close to zero. One possible way to deal with the latter situation is to employ the Levenberg–Marquardt regularization, blending the Newton direction with the steepest descent direction. However, the drawback of using the Levenberg–Marquardt directions in linesearch methods is the difficulty in choosing the regularization parameters; see Sect. 2.2.3.

Alternatively, the following algorithm implementing the Newton method equipped with the Armijo linesearch rule has the option of resorting directly to the steepest descent step as a safeguard, but only in those cases when the Newtonian direction does not exist or is too large.

Algorithm 5.1 Choose parameters $C > 0$, $\tau > 0$, $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $x^0 \in \mathbf{R}^n$ and set $k = 0$.

1. If $\Phi(x^k) = 0$, stop.
2. Compute $p^k \in \mathbf{R}^n$ as a solution of the linear equation
$$\Phi(x^k) + \Phi'(x^k)p = 0. \tag{5.5}$$
If such p^k exists and
$$\|p^k\| \leq \max\{C,\ 1/\|\Phi(x^k)\|^\tau\}, \tag{5.6}$$
go to step 4.
3. Set $p^k = -\varphi'(x^k) = -(\Phi'(x^k))^{\mathrm{T}}\Phi(x^k)$, with $\varphi : \mathbf{R}^n \to \mathbf{R}$ defined by (5.2). If $p^k = 0$, stop.
4. Set $\alpha = 1$. If the inequality
$$\varphi(x^k + \alpha p^k) \leq \varphi(x^k) + \sigma\alpha\langle\varphi'(x^k),\ p^k\rangle \tag{5.7}$$
is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (5.7) again, etc., until (5.7) becomes valid.
5. Set $x^{k+1} = x^k + \alpha_k p^k$.
6. Increase k by 1 and go to step 1.

Remark 5.2. According to (5.2) and (5.4), the condition (5.7) can be written as follows:

$$\frac{1}{2}\|\Phi(x^k + \alpha p^k)\|^2 \leq \frac{1}{2}\|\Phi(x^k)\|^2 - \sigma\alpha\|\Phi(x^k)\|^2 = \frac{1 - 2\sigma\alpha}{2}\|\Phi(x^k)\|^2,$$

or equivalently,

$$\|\Phi(x^k + \alpha p^k)\| \leq \sqrt{1 - 2\sigma\alpha}\|\Phi(x^k)\| = (1 - \sigma\alpha + o(\alpha))\|\Phi(x^k)\|$$

as $\alpha \to 0$. Therefore, in Algorithm 5.1, the condition (5.7) can be replaced by

$$\|\Phi(x^k + \alpha p^k)\| \leq (1 - \sigma\alpha)\|\Phi(x^k)\|. \tag{5.8}$$

Observe that, by the differentiability of Φ at x^k, and by (5.3),

$$\Phi(x^k + \alpha p^k) = \Phi(x^k) + \alpha\Phi'(x^k)p^k + o(\alpha) = (1 - \alpha)\Phi(x^k) + o(\alpha)$$

as $\alpha \to 0$, and hence, (5.8) is satisfied for all $\alpha > 0$ small enough.

Our global convergence result for Algorithm 5.1 is along the lines of Theorem 2.25.

Theorem 5.3. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be continuously differentiable on $\mathbf{R}^n$.*

Then for any starting point $x^0 \in \mathbf{R}^n$ Algorithm 5.1 generates the iterative sequence $\{x^k\}$ such that each of its accumulation points $\bar{x} \in \mathbf{R}^n$ satisfies

$$(\Phi'(\bar{x}))^{\mathrm{T}}\Phi(\bar{x}) = 0. \tag{5.9}$$

Proof. The fact that Algorithm 5.1 is well defined follows from Lemma 2.19. Indeed, for each k the corresponding direction p^k either satisfies (5.4) or equals $-\varphi'(x^k)$. Therefore, if $\varphi'(x^k) = (\Phi'(x^k))^{\mathrm{T}}\Phi(x^k) \neq 0$, then in either case

$$\langle \varphi'(x^k),\, p^k \rangle < 0. \tag{5.10}$$

Moreover, assuming that $\varphi'(x^k) \neq 0$ for all k, the sequence $\{\varphi(x^k)\}$ is monotonically decreasing. Since this sequence is bounded below (by zero), it converges, and hence, (5.7) implies the equality

$$\lim_{k \to \infty} \alpha_k \langle \varphi'(x^k),\, p^k \rangle = 0. \tag{5.11}$$

Let $\bar{x}$ be an accumulation point of the sequence $\{x^k\}$, and let $\{x^{k_j}\}$ be a subsequence convergent to $\bar{x}$ as $j \to \infty$. Consider the two possible cases:

$$\limsup_{j \to \infty} \alpha_{k_j} > 0 \quad \text{or} \quad \lim_{j \to \infty} \alpha_{k_j} = 0. \tag{5.12}$$

In the first case, passing onto a further subsequence if necessary, we can assume that the entire $\{\alpha_{k_j}\}$ is separated away from zero:

$$\liminf_{j \to \infty} \alpha_{k_j} > 0.$$

Then (5.11) implies that

$$\lim_{j \to \infty} \langle \varphi'(x^{k_j}),\, p^{k_j} \rangle = 0. \tag{5.13}$$

If p^{k_j} is defined by the Newton iteration system (5.5) for infinitely many indices j, by (5.4) we have that

$$\langle \varphi'(x^{k_j}),\, p^{k_j} \rangle = -\|\Phi(x^{k_j})\|^2$$

for these j, and then (5.13) implies that $\Phi(\bar{x}) = 0$, which certainly implies (5.9). On the other hand, if Newton directions are used only for finitely many indices j, then

$$\langle \varphi'(x^{k_j}),\, p^{k_j} \rangle = -\langle \varphi'(x^{k_j}),\, \varphi'(x^{k_j}) \rangle = -\|\varphi'(x^{k_j})\|^2$$

for all j large enough. Hence, by (5.13), $(\Phi'(\bar{x}))^{\mathrm{T}}\Phi(\bar{x}) = \varphi'(\bar{x}) = 0$, i.e., (5.9) holds in this case as well.

It remains to consider the second case in (5.12). Suppose first that the sequence $\{p^{k_j}\}$ is unbounded. Note that this can only happen when the Newton directions are used infinitely often, because otherwise $p^{k_j} = -\varphi'(x^{k_j})$ for all j large enough, and hence, $\{p^{k_j}\}$ converges to $-\varphi'(\bar{x})$. But then the condition (5.6) implies that

$$\liminf_{j\to\infty} \|\Phi(x^{k_j})\| = 0,$$

so that $\Phi(\bar{x}) = 0$, and hence, (5.9) is again valid.

Let finally $\{p^{k_j}\}$ be bounded. Taking a further subsequence, if necessary, assume that $\{p^{k_j}\}$ converges to some $\tilde{p}$. Since in the second case in (5.12) for each j large enough the initial stepsize value had been reduced at least once, the value $\alpha_{k_j}/\theta > \alpha_{k_j}$ does not satisfy (5.7), i.e.,

$$\frac{\varphi(x^{k_j} + \alpha_{k_j} p^{k_j}/\theta) - \varphi(x^{k_j})}{\alpha_{k_j}/\theta} > \sigma\langle\varphi'(x^{k_j}),\, p^{k_j}\rangle.$$

Employing the mean-value theorem (Theorem A.10, (a)) and the fact that $\alpha_{k_j} \to 0$ as $j \to \infty$, and passing onto the limit as $j \to \infty$, we obtain that

$$\langle\varphi'(\bar{x}),\, \tilde{p}\rangle \geq \sigma\langle\varphi'(\bar{x}),\, \tilde{p}\rangle,$$

which may only hold when $\langle\varphi'(\bar{x}),\, \tilde{p}\rangle \geq 0$. Combining this with (5.10), we obtain that

$$\langle\varphi'(\bar{x}),\, \tilde{p}\rangle = 0.$$

Considering, as above, the two cases when the number of times the Newton direction had been used is infinite or finite, the latter relation implies that (5.9) holds. □

According to the proof of Theorem 5.3, if along a subsequence convergent to $\bar{x}$ the Newton direction had been used infinitely many times, then $\bar{x}$ is a solution of (5.1). Convergence to a point $\bar{x}$ satisfying (5.9) which is not a solution of (5.1) can only happen when the Newton directions are not used along the corresponding subsequence from some point on at all, and when in addition the Jacobian $\Phi'(\bar{x})$ is singular.

Another important issue is the existence of accumulation points of iterative sequences generated by Algorithm 5.1. This is guaranteed when the residual $\|\Phi(\cdot)\|$ is coercive.

Finally, we show that Algorithm 5.1 preserves fast local convergence of the basic Newton method under natural assumptions.

Theorem 5.4. *Let $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ be continuously differentiable on $\mathbf{R}^n$. Let a sequence $\{x^k\} \subset \mathbf{R}^n$ be generated by Algorithm 5.1 with $\sigma \in (0,\, 1/2)$, and assume that this sequence has an accumulation point $\bar{x}$ such that $\Phi'(\bar{x})$ is nonsingular.*

Then the entire sequence $\{x^k\}$ *converges to* $\bar{x}$*, and the rate of convergence is superlinear. Moreover, the rate of convergence is quadratic provided the derivative of* Φ *is locally Lipschitz-continuous with respect to* $\bar{x}$*.*

Proof. If $x^k \in \mathbf{R}^n$ is close enough to $\bar{x}$, then, according to Theorem 2.2, there exists the unique $p^k \in \mathbf{R}^n$ satisfying (5.5), and

$$x^k + p^k - \bar{x} = o(\|x^k - \bar{x}\|) \tag{5.14}$$

as $x^k \to \bar{x}$. As a consequence, p^k would be accepted by the test (5.6).

Furthermore, Proposition 1.32 implies the estimate

$$x^k - \bar{x} = O(\|\Phi(x^k)\|)$$

as $x^k \to \bar{x}$. Employing this estimate and (5.2), (5.14), and also taking into account the local Lipschitz-continuity of Φ with respect to $\bar{x}$ (following from the differentiability of Φ at $\bar{x}$), we obtain that

$$\begin{aligned}
\varphi(x^k + p^k) &= \frac{1}{2}\|\Phi(x^k + p^k) - \Phi(\bar{x})\|^2 \\
&= O(\|x^k + p^k - \bar{x}\|^2) \\
&= o(\|x^k - \bar{x}\|^2) \\
&= o(\|\Phi(x^k)\|^2)
\end{aligned}$$

as $x^k \to \bar{x}$. The above relation implies that if x^k is close enough to $\bar{x}$, then

$$\begin{aligned}
\varphi(x^k + p^k) &\leq \frac{1 - 2\sigma}{2}\|\Phi(x^k)\|^2 \\
&= \varphi(x^k) - \sigma\|\Phi(x^k)\|^2 \\
&= \varphi(x^k) + \sigma\langle \varphi'(x^k),\, p^k\rangle,
\end{aligned}$$

where the last equality is by (5.4) (recall also that $\sigma \in (0,\, 1/2)$). Therefore, $\alpha_k = 1$ is accepted by step 4 of the algorithm: inequality (5.7) holds with $\alpha = 1$. This shows that the iteration of Algorithm 5.1 reduces to that of Algorithm 2.1. The assertions now follow from Theorem 2.2. □

5.1.2 Globalized Semismooth Newton Methods for Complementarity Problems

Consider now the nonlinear complementarity problem (NCP)

$$x \geq 0, \quad \Phi(x) \geq 0, \quad \langle x,\, \Phi(x)\rangle = 0, \tag{5.15}$$

where $\Phi : \mathbf{R}^n \to \mathbf{R}^n$ is a smooth mapping.