

# NEWTON'S METHOD IN THE CONTEXT OF GRADIENTS

J. KARÁTSON, J. W. NEUBERGER

ABSTRACT. The paper gives a common theoretical treatment for gradient and Newton type methods for general classes of problems. First, for Euler-Lagrange equations Newton's method is characterized as an (asymptotically) optimal variable steepest descent method. Second, Sobolev gradient type minimization is developed for general problems using a continuous Newton method which takes into account a 'boundary condition' operator.

## 1. INTRODUCTION

Gradient and Newton type methods are among the most important approaches for the solution of nonlinear equations, both in  $\mathbf{R}^n$  and in abstract spaces. The latter are often connected to PDE applications, and here the involvement of Sobolev spaces has proved an efficient strategy, see e.g. [8, 12] on the Sobolev gradient approach and [1, 5] on Newton type methods. Further applications of Sobolev space iterations are found in [4].

The two types of methods (gradient and Newton) are generally considered as two different approaches, although their connection has been studied in some papers, see e.g. [3] in the context of continuous steepest-descent, [7] on variable preconditioning and quasi-Newton methods, and [8, Chapter 7] on Newton's method and constrained optimization.

The goal of this paper is to establish a common theoretical framework in which gradient and Newton type methods can be treated, and thereby to clarify the relation of the two types of methods for general classes of problems.

Note that there are two distinct ways systems of differential equations may be placed into an optimization setting. Sometimes it is possible to show that a given system of PDEs are Euler-Lagrange equations for some functional  $\phi$ . In the more general case one looks for the critical points of a least-squares functional associated with the given system. Furthermore, one can approach Newton type methods also in two different ways: from numerical aspect it is the study of the discrete (i.e. iterative) solution method that is mostly relevant, whereas continuous Newton methods can lead to attractive theoretical results.

The first part of this paper characterizes Newton's method in the Euler-Lagrange case as an (asymptotically) optimal variable steepest descent method for the iterative minimization of the corresponding functional. The second part treats the more general (either Euler-Lagrange or least squares) case and develops Sobolev gradient type minimization using a continuous Newton method which takes into account a 'boundary condition' operator.

2. UNCONSTRAINED OPTIMIZATION: NEWTON'S METHOD AS A VARIABLE  
STEEPEST DESCENT

Let  $H$  be a real Hilbert space and  $F : H \rightarrow H$  an operator which has a potential  $\phi : H \rightarrow \mathbf{R}$ , i.e.

$$\phi'(u)h = \langle h, F(u) \rangle \quad (u, h \in H) \quad (1)$$

in Gateaux sense. Using the standard identification of  $H$  with its dual, in this section we simply write (1) as

$$\phi'(u) = F(u) \quad (u \in H). \quad (2)$$

We consider the operator equation

$$F(u) = 0 \quad (3)$$

and study the relationship between steepest descent and Newton methods.

We will observe that Newton's method can be regarded as a special variable steepest descent iteration, where the latter means that the gradients of  $\phi$  are taken w.r. to stepwise redefined inner products. Then our main result states the following principle: whereas the descents in the ordinary gradient method are steepest w.r. to different directions, in Newton's method they are steepest w.r. to both different directions and inner products. This optimality is understood in a (second order) asymptotic sense in the neighbourhood of the solution.

**2.1. Fixed and variable steepest descent iterations.** A steepest descent iteration corresponding to the gradient  $\phi'$  in (2) is

$$u_{n+1} = u_n - \alpha_n F(u_n) \quad (4)$$

with some constants  $\alpha_n > 0$ . Our aim is to modify this sequence by varying the inner product of the space  $H$ .

**2.1.1. Steepest descent under a fixed inner product.** First we modify the sequence (4) by introducing another fixed inner product. For this purpose let  $B : H \rightarrow H$  be a bounded self-adjoint linear operator which is *strongly positive* (i.e. it has a positive lower bound  $p > 0$ ), and let

$$\langle u, v \rangle_B \equiv \langle Bu, v \rangle \quad (u, v \in H).$$

Denote by  $\phi'_B$  the gradient of  $\phi$  w.r. to the energy inner product  $\langle \cdot, \cdot \rangle_B$ . Then

$$\langle \phi'_B(u), v \rangle_B = \frac{\partial \phi}{\partial v}(u) = \langle \phi'(u), v \rangle = \langle B^{-1}F(u), v \rangle_B \quad (u, v \in H),$$

which implies

$$\phi'_B(u) = B^{-1}F(u) \quad (u \in H). \quad (5)$$

That is, the change of the inner product yields the change of the gradient of  $\phi$ , namely, the modified gradient is expressed as the preconditioned version of the original one. Consequently, a steepest descent iteration corresponding to the gradient  $\phi'_B$  is the preconditioned sequence

$$u_{n+1} = u_n - \alpha_n B^{-1}F(u_n) \quad (6)$$

with some constants  $\alpha_n > 0$ .

Convergence results for such sequences are well-known if  $\phi$  is strongly convex, which can be formulated in terms of the operator  $F$  (see e.g. the monographs [4, 5]). For instance, if the spectral bounds of the operators  $F'(u)$  are between uniform constants  $M \geq m > 0$  (in the original resp. the energy inner product), then the constant stepsize  $\alpha_n \equiv 2/(M + m)$  yields convergence with ratio

$$q = \frac{M - m}{M + m}$$

for the sequences (4) and (6), resp. Clearly, the aim of the change of the inner product is to achieve better spectral bounds in the new inner product. For instance, for PDEs a sometimes dramatic improvement can be achieved by using the Sobolev inner product instead of the original  $L^2$  one (see the monograph [8] on Sobolev gradients).

*2.1.2. Steepest descent under a variable inner product.* Assume that the  $n$ th term of an iterative sequence is constructed and let  $B_n : H \rightarrow H$  be a strongly positive bounded self-adjoint linear operator. It follows similarly to (5) that the gradient of  $\phi$  w.r. to the inner product  $\langle \cdot, \cdot \rangle_{B_n}$  is

$$\phi'_{B_n}(u) = B_n^{-1}F(u) \quad (u \in H). \quad (7)$$

The relation (7) means that a one-step iterative sequence

$$u_{n+1} = u_n - \alpha_n B_n^{-1}F(u_n) \quad (8)$$

(with some constants  $\alpha_n > 0$ ) is a variable steepest descent iteration corresponding to  $\phi$  such that in the  $n$ th step the gradient of  $\phi$  is taken w.r. to the inner product  $\langle \cdot, \cdot \rangle_{B_n}$ .

Several such types of iterative methods are known including variable metric methods (see e.g. the monograph [13]). In this context 'variable' is understood as depending on the step  $n$ . We note that Sobolev gradients under variable inner product can also be defined in the context of continuous steepest descent, and the inner product may depend continuously on each element of the Sobolev space (see [11, 12]).

Convergence results for sequences of the form (8) are given in [2, 7], formulated again for convex functionals in terms of spectral bounds. Namely, under the stepwise spectral equivalence relation

$$m_n \langle B_n h, h \rangle \leq \langle F'(u_n)h, h \rangle \leq M_n \langle B_n h, h \rangle \quad (n \in \mathbf{N}, h \in H) \quad (9)$$

(with some constants  $M_n \geq m_n > 0$ ) and assuming the Lipschitz continuity of  $F'$ , one can achieve convergence with ratio

$$q = \limsup \frac{M_n - m_n}{M_n + m_n}.$$

(This convergence is global if  $\alpha_n$  includes damping.) In particular, superlinear convergence can also be obtained when  $q = 0$ , and its rate is characterized by the speed as  $M_n/m_n \rightarrow 1$ .

Clearly, the variable steepest descent iteration (8) can also be regarded as a quasi-Newton method, since the relation (9) provides the operators  $B_n$  as approximations of  $F'(u_n)$ . Moreover, the choice  $B_n = F'(u_n)$  yields optimal spectral bounds  $m_n = M_n = 1$  in (9), and the corresponding variable steepest descent iteration (8) becomes Newton method with quadratic convergence speed.

**2.1.3. Conclusion.** Altogether, we may observe the following relationship between steepest descent and Newton methods. A usual steepest descent method defines an optimal descent direction under a fixed inner product, but the search for an optimal descent may also include the stepwise change of inner product. If these inner products are looked for among energy inner products  $\langle \cdot, \cdot \rangle_{B_n}$  corresponding to (9), then a resulting variable steepest descent iteration coincides with a quasi-Newton method. Under the special choice  $B_n = F'(u_n)$  we obtain Newton's method itself in this way, and the convergence results suggest that the optimal convergence is obtained with this choice.

Roughly speaking, this means the following principle: whereas the descents in the gradient method are steepest w.r. to different directions, in Newton's method they are steepest w.r. to both different directions and inner products.

However, the above principle is not proved by the quoted convergence results themselves. Namely, in their proof they a priori compare the rate of quasi-Newton methods to the exact Newton's method, hence the obtained convergence estimates are obviously not better than those for the exact Newton's method. Therefore our goal in the next section is to verify the above stated principle in a proper sense.

**2.2. Newton's method as an optimal variable steepest descent.** We consider the operator equation (3) and the corresponding potential  $\phi : H \rightarrow \mathbf{R}$ .

In this subsection we assume that  $\phi$  is uniformly convex and  $\phi''$  is locally Lipschitz continuous. More exactly, formulated in terms of the operator  $F$  in (2), we impose the following conditions:

- (i)  $F$  is Gateaux differentiable;
- (ii) for every  $R > 0$  there exist constants  $P \geq p > 0$  such that

$$p\|h\|^2 \leq \langle F'(u)h, h \rangle \leq P\|h\|^2 \quad (\|u\| \leq R, h \in H); \quad (10)$$

- (iii) for every  $R > 0$  there exists a constant  $L > 0$  such that

$$\|F'(u) - F'(v)\| \leq L\|u - v\| \quad (\|u\|, \|v\| \leq R).$$

These conditions themselves do not ensure that equation (3) has a solution, hence we impose condition

- (iv) equation (3) has a solution  $u^* \in H$ .

Then the solution  $u^*$  is unique and also minimizes  $\phi$ . We note that the existence of  $u^*$  is already ensured if the lower bound  $p = p(R)$  in condition (ii) satisfies  $\lim_{R \rightarrow \infty} R p(R) = +\infty$ , or if  $p$  does not depend on  $R$  at all (see e.g. [4, 5])

Let  $u_0 \in H$  and let a variable steepest descent iteration be constructed in the form (8):

$$u_{k+1} = u_k - \alpha_k B_k^{-1} F(u_k) \quad (11)$$

with suitable constants  $\alpha_k > 0$  and strongly positive self-adjoint operators  $B_k$ .

Let  $n \in \mathbf{N}$  and assume that the  $n$ th term of the sequence (11) is constructed. The stepsize  $\alpha_n$  yields steepest descent w.r.t.  $B_n$  if  $\phi(u_{n+1})$  coincides with the number

$$\mu(B_n) \equiv \min_{\alpha > 0} \phi(u_n - \alpha B_n^{-1} F(u_n)).$$

We wish to choose  $B_n$  such that this value is the smallest possible within the class of strongly positive operators

$$\mathcal{B} \equiv \{B \in L(H) \text{ self-adjoint} : \exists p > 0 \quad \langle Bh, h \rangle \geq p \|h\|^2 \quad (h \in H)\} \quad (12)$$

where  $L(H)$  denotes the set of bounded linear operators on  $H$ . (The strong positivity is needed to yield  $R(B_n) = H$ , by which the existence of  $B_n^{-1} F(u_n)$  is ensured in the iteration.) Moreover, when  $B_n \in \mathcal{B}$  is varied then one can incorporate the number  $\alpha$  in  $B_n$ , since  $\alpha B_n \in \mathcal{B}$  as well for any  $\alpha > 0$ . That is, it suffices to replace  $\mu(B_n)$  by

$$m(B_n) \equiv \phi(u_n - B_n^{-1} F(u_n)), \quad (13)$$

and to look for

$$\min_{B_n \in \mathcal{B}} m(B_n).$$

Our aim is to verify that

$$\min_{B_n \in \mathcal{B}} m(B_n) = m(F'(u_n)) \quad \text{up to second order} \quad (14)$$

as  $u_n \rightarrow u^*$ , i.e. the Newton iteration realizes asymptotically the stepwise optimal steepest descent among different inner products in the neighbourhood of  $u^*$ . (That is, the descents in Newton's method are asymptotically steepest w.r. to both different directions and inner products.) We note that, clearly, the asymptotic result cannot be replaced by an exact one, this can be seen for fixed  $u_n$  by an arbitrary nonlocal change of  $\phi$  along the descent direction.

The result (14) can be given an exact formulation in the following way. First we define for any  $\nu_1 > 0$  the set

$$\mathcal{B}(\nu_1) \equiv \{B \in L(H) \text{ self-adjoint} : \langle Bh, h \rangle \geq \nu_1 \|h\|^2 \quad (h \in H)\}, \quad (15)$$

i.e. the subset of  $\mathcal{B}$  with operators having the common lower bound  $\nu_1 > 0$ .

**Theorem 1.** *Let conditions (i)-(iv) be satisfied. Let  $u_0 \in H$  and let the sequence  $(u_k)$  be given by (11) with some constants  $\alpha_k > 0$  and operators  $B_k \in \mathcal{B}$ , with  $\mathcal{B}$  defined in (12).*

*Let  $n \in \mathbf{N}$  be fixed,  $m(B_n)$  defined by (13) and let*

$$\hat{m}(B_n) \equiv \beta + \frac{1}{2} \langle H_n(B_n^{-1} g_n - H_n^{-1} g_n), B_n^{-1} g_n - H_n^{-1} g_n \rangle, \quad (16)$$

where

$$\beta = \phi(u^*), \quad g_n = F(u_n), \quad H_n = F'(u_n). \quad (17)$$

Then

(1) there holds

$$\min_{B_n \in \mathcal{B}} \hat{m}(B_n) = \hat{m}(F'(u_n));$$

(2)  $\hat{m}(B_n)$  is the second order approximation of  $m(B_n)$ , i.e., for any  $\nu_1 > 0$  and  $B_n \in \mathcal{B}(\nu_1)$

$$|m(B_n) - \hat{m}(B_n)| \leq C \|u_n - u^*\|^3 \quad (18)$$

(with  $\mathcal{B}(\nu_1)$  defined by (15)), where  $C > 0$  depends on  $u_0$  and  $\nu_1$ , but does not depend on  $B_n$  or  $u_n$ .

**Proof.** (1) This part of the theorem is obvious since, using that  $H_n = F'(u_n)$  is positive definite by assumption (ii), we obtain

$$\hat{m}(B_n) \geq \beta = \hat{m}(H_n) = \hat{m}(F'(u_n)).$$

(2) We verify the required estimate in four steps.

(i) First we prove that

$$\|u_n - u^*\| \leq R_0 \quad (19)$$

where  $R_0$  depends on  $u_0$ , that is, the initial guess determines an a priori bound for a ball  $B(u^*, R_0)$  around  $u^*$  containing the sequence (11). For this it suffices to prove that the level set corresponding to  $\phi(u_0)$  is contained in such a ball, i.e.,

$$\{u \in H : \phi(u) \leq \phi(u_0)\} \subset B(u^*, R_0), \quad (20)$$

since  $u_n$  is a descent sequence w.r.t.  $\phi$ .

Let  $u \in H$  be fixed and consider the real function

$$f(t) := \phi\left(u^* + t \frac{u - u^*}{\|u - u^*\|}\right) \quad (t \in \mathbf{R}),$$

which is  $C^2$ , convex and has its minimum at 0. Assumption (ii) implies that there exists  $p_1 > 0$  such that

$$\langle \phi''(v)h, h \rangle \geq p_1 \|h\|^2 \quad (\|v - u^*\| \leq 1, h \in H),$$

and hence

$$f''(t) \geq p_1 \quad (|t| \leq 1).$$

Then elementary calculus yields that  $f'(1) \geq p_1$  and  $f(1) - f(0) \geq p_1/2$ , hence

$$\begin{aligned} \phi(u) - \phi(u^*) &= f(\|u - u^*\|) - f(1) + f(1) - f(0) \\ &\geq f'(1)(\|u - u^*\| - 1) + f(1) - f(0) \geq p_1 \left(\|u - u^*\| - \frac{1}{2}\right). \end{aligned}$$

This implies that if

$$\|u - u^*\| \geq \frac{1}{p_1} \left(\phi(u_0) - \phi(u^*)\right) + \frac{1}{2} \equiv R_0$$

then  $\phi(u) \geq \phi(u_0)$ , that is, (20) holds with this  $R_0$ .

(ii) In the sequel we omit the index  $n$  for notational simplicity, and let

$$u = u_n, \quad g = g_n, \quad H = H_n, \quad B = B_n,$$

where  $g_n = F(u_n)$  and  $H_n = F'(u_n)$  were defined in (17). Using these notations, (13) turns into

$$m(B) = \phi(u - B^{-1}g). \quad (21)$$

Further, we fix  $\nu_1 > 0$  and assume that  $B \in \mathcal{B}(\nu_1)$  as defined by (15).

Now we verify that

$$m(B) = \phi(u) - \langle B^{-1}g, g \rangle + \frac{1}{2} \langle HB^{-1}g, B^{-1}g \rangle + R_1 \quad (22)$$

where

$$|R_1| \leq C_1 \|u - u^*\|^3 \quad (23)$$

with  $C_1 > 0$  depending only on  $u_0$  and  $\nu_1$ . Let  $z = B^{-1}g$ . Then the Taylor expansion yields

$$m(B) = \phi(u - z) = \phi(u) - \langle \phi'(u), z \rangle + \frac{1}{2} \langle \phi''(u)z, z \rangle + R_1, \quad (24)$$

here the Lipschitz continuity of  $\phi''$  implies

$$|R_1| \leq \frac{L_0}{6} \|z\|^3 \quad (25)$$

where  $L_0$  is the Lipschitz constant corresponding to the ball  $B(u^*, R_0)$  according to assumption (iii). Here

$$\phi'(u) = F(u) = g \quad \text{and} \quad \phi''(u) = F'(u) = H, \quad (26)$$

hence the definition of  $z$  and the symmetry of  $B$  yield

$$\langle \phi'(u), z \rangle = \langle B^{-1}g, g \rangle, \quad \langle \phi''(u)z, z \rangle = \langle HB^{-1}g, B^{-1}g \rangle$$

and in order to verify (23) it suffices to prove that

$$\|z\| \leq K_1 \|u - u^*\| \quad (27)$$

with  $K_1 > 0$  depending on  $u_0$  and  $\nu_1$ .

The Taylor expansion for  $\phi'$  yields

$$g = \phi'(u) = \phi'(u^*) + \phi''(u^*)(u - u^*) + \varrho_1, \quad (28)$$

where

$$|\varrho_1| \leq \frac{L_0}{2} \|u - u^*\|^2$$

with  $L_0$  as above. Here  $\phi'(u^*) = 0$ . Let  $P_0$  be the upper spectral bound of  $\phi''$  on the ball  $B(u^*, R_0)$ , obtained from assumption (ii). Then, also using (19), we have

$$\|g\| \leq P_0 \|u - u^*\| + \frac{L_0}{2} \|u - u^*\|^2 \leq \left( P_0 + \frac{L_0 R_0}{2} \right) \|u - u^*\| = K_0 \|u - u^*\|. \quad (29)$$

From this the assumption  $B \in \mathcal{B}(\nu_1)$  yields

$$\|z\| = \|B^{-1}g\| \leq (K_0/\nu_1) \|u - u^*\|,$$

hence (27) holds with  $K_1 = K_0/\nu_1$  and thus (22)-(23) are verified.

(iii) Now we prove that

$$\phi(u) = \beta + \frac{1}{2}\langle H^{-1}g, -^1g \rangle + R_2 \quad (30)$$

where

$$|R_2| \leq C_2 \|u - u^*\|^3 \quad (31)$$

with  $C_2 > 0$  depending only on  $u_0$  and  $\nu_1$ . Similarly to (24)-(25), we have

$$\phi(u) = \phi(u^*) + \langle \phi'(u^*), u - u^* \rangle + \frac{1}{2}\langle \phi''(u^*)(u - u^*), u - u^* \rangle + \varrho_2,$$

where

$$|\varrho_2| \leq \frac{L_0}{6} \|u - u^*\|^3.$$

Here  $\phi(u^*) = \beta$ ,  $\phi'(u^*) = 0$  and

$$|\langle \phi''(u^*)(u - u^*), u - u^* \rangle - \langle H(u - u^*), u - u^* \rangle| \leq L_0 \|u - u^*\|^3$$

from  $H = \phi''(u)$  and the Lipschitz condition. Hence

$$\phi(u) = \beta + \frac{1}{2}\langle H(u - u^*), u - u^* \rangle + \varrho_3,$$

where

$$|\varrho_3| \leq \frac{2L_0}{3} \|u - u^*\|^3.$$

Therefore it remains to prove that

$$|\langle H(u - u^*), u - u^* \rangle - \langle H^{-1}g, g \rangle| \leq C_3 \|u - u^*\|^3. \quad (32)$$

Here (28) implies

$$g = \phi'(u) = \phi''(u^*)(u - u^*) + \varrho_1 = H(u - u^*) + (\phi''(u^*) - H)(u - u^*) + \varrho_1.$$

Using again the Lipschitz condition for  $\phi''$ , we have

$$\|(\phi''(u^*) - H)(u - u^*)\| \leq L_0 \|u - u^*\|^2,$$

hence

$$g = H(u - u^*) + \varrho_4 \quad (33)$$

with

$$|\varrho_4| \leq C_4 \|u - u^*\|^2. \quad (34)$$

Setting (33) into the left-hand side expression in (32) and using the symmetry of  $H$ , we obtain

$$\begin{aligned} |\langle H(u - u^*), u - u^* \rangle - \langle H^{-1}g, g \rangle| &= |\langle g - \varrho_4, H^{-1}(g - \varrho_4) \rangle - \langle H^{-1}g, g \rangle| \\ &= |-2\langle H^{-1}g, \varrho_4 \rangle + \langle H^{-1}\varrho_4, \varrho_4 \rangle| \leq 2|\langle H^{-1}g, \varrho_4 \rangle| + |\langle H^{-1}\varrho_4, \varrho_4 \rangle|. \end{aligned}$$

Let  $p_0$  be the lower spectral bound of  $\phi''$  on the ball  $B(u^*, R_0)$ , obtained from assumption (ii). Then  $\|H^{-1}\| \leq 1/p_0$ . Hence, using (29), (34) and (19), we have

$$|\langle H(u - u^*), u - u^* \rangle - \langle H^{-1}g, g \rangle| \leq \frac{1}{p_0} \left( 2\|g\|\|\varrho_4\| + \|\varrho_4\|^2 \right)$$

$$\leq \frac{1}{p_0} \left( 2K_0 C_4 \|u - u^*\|^3 + C_4^2 \|u - u^*\|^4 \right) \leq \frac{1}{p_0} \left( 2K_0 C_4 + R_0 C_4^2 \right) \|u - u^*\|^3,$$

that is, (32) holds and thus (30)-(31) are verified.

(iv) Let us set (30) into (22) and use notation  $R_3 = R_1 + R_2$  :

$$\begin{aligned} m(B) &= \beta + \frac{1}{2} \langle H^{-1}g, -g \rangle - \langle B^{-1}g, g \rangle + \frac{1}{2} \langle HB^{-1}g, B^{-1}g \rangle + R_3 \\ &= \beta + \frac{1}{2} \langle H(B^{-1}g - H^{-1}g), B^{-1}g - H^{-1}g \rangle + R_3 = \hat{m}(B) + R_3, \end{aligned}$$

where by (23) and (31)

$$|R_3| \leq C \|u - u^*\|^3$$

with  $C = C_1 + C_2$ . Therefore (18) is true and the proof is complete.  $\blacksquare$

**Remark 1.** A main application of the above theorem arises for second order nonlinear elliptic problems. Then one can define various Sobolev gradients using different weight functions in the Sobolev inner product. For instance, in the case of Dirichlet problems one can use weighted Sobolev norms  $\langle h, h \rangle_w = \int_{\Omega} w(x) |\nabla h|^2 dx$  where  $w$  is a positive bounded function, or more generally  $\langle h, h \rangle_W = \int_{\Omega} W(x) \nabla h \cdot \nabla h dx$  where  $W$  is a bounded uniformly positive definite matrix function. Such weighted norms can be written as  $\langle Bh, h \rangle_{H_0^1}$  with some operator  $B$  as in (15) on the space  $H = H_0^1(\Omega)$ , where  $\langle \cdot, \cdot \rangle_{H_0^1}$  denotes the standard Sobolev inner product, hence the optimality result of Theorem 1 covers such Sobolev gradient preconditioners.

### 3. CONSTRAINED OPTIMIZATION FOR NEWTON'S METHOD AND SOBOLEV GRADIENTS

A different interpretation of Newton's method in Sobolev gradient context uses minimization subject to constraints, which we build up using a continuous Newton method. Suppose that  $\phi$  is a  $C^3$  function from  $R^n$  into  $R$ . What philosophy might guide a choice of a numerically efficient gradient for  $\phi$ ? We first give a quick development for the unconstrained case which gives a somewhat different point of view to the previous section. We then pass to the constrained case.

If  $\phi$  arises from a discretization of a system of differential equations then the ordinary gradient, a list of partial derivatives of  $\phi$  is a very poor choice for numerical purposes. We illustrate this by a simple example in which the underlying equation is  $u' - u = 0$  on  $[0, 1]$ . For  $n$  a positive integer, a finite dimensional least-squares formulation is, with  $\delta = 1/n$ ,

$$\phi(u_0, u_1, \dots, u_n) = \frac{1}{2} \sum_{k=1}^n \left( \frac{u_k - u_{k-1}}{\delta} - \frac{u_k + u_{k-1}}{2} \right)^2, \quad (35)$$

where  $(u_0, u_1, \dots, u_n) \in R^{n+1}$ . It may be seen that if  $(u_0, u_1, \dots, u_n)$  is a critical point of  $\phi$  then  $\phi(u_0, u_1, \dots, u_n) = 0$  and so

$$\frac{u_k - u_{k-1}}{\delta} - \frac{u_k + u_{k-1}}{2} = 0, \quad k = 1, \dots, n,$$

which are precisely the equations to be satisfied by the Crank-Nicholson method for this problem. It is widely understood that the ordinary gradient of  $\phi$  is a disaster numerically using steepest descent. By contrast, consider the gradient of  $\phi$  taken with respect the following finite dimensional emulation of of the Sobolev space  $H^{1,2}([0, 1])$ :

$$\alpha(u_0, u_1, \dots, u_n) = \|u\|_S^2 = \sum_{k=1}^n \left( \left( \frac{u_k - u_{k-1}}{\delta} \right)^2 + \left( \frac{u_k + u_{k-1}}{2} \right)^2 \right), \quad (36)$$

$u = (u_0, u_1, \dots, u_n) \in R^{n+1}$ . The Sobolev gradient of  $\phi$  at such a  $u$  is the element  $(\nabla_S \phi)(u)$  so that

$$\phi'(u)h = \langle h, (\nabla_S \phi)(u) \rangle_S, \quad h \in R^{n+1},$$

where  $\langle \cdot, \cdot \rangle_S$  denotes the inner product associated with (36).

In [8], it is indicated about seven steepest descent iterations suffices using the Sobolev gradient whereas for steepest descent with the ordinary gradient a large number of iterations is required (on the order of 30, 5000, 500000 iterations required for  $n=10, 20, 40$  respectively).

In the above example we might have been guided in our choice of metric by the fact that the Sobolev space  $H^{1,2}([0, 1])$  is a good choice of a metric for the underlying continuous least squares problem

$$\Phi(u) = \frac{1}{2} \int_0^1 (u' - u)^2, \quad u \in H^{1,2}([0, 1]).$$

That this Sobolev metric renders  $\Phi$  differentiable (in contrast with trying to define  $\Phi$  as a densely defined everywhere discontinuous function on  $L_2([0, 1])$ ) is a good indication that its finite dimensional emulation should provide a good numerical gradient.

Examining (35), (36) together we see that elements  $(u_0, u_1, \dots, u_n)$  have similar sensitivity (i.e., similar sized partial derivatives) in both expressions. Note that the first and last components of such a vector have sensitivity quite different from the other  $n - 1$  components. Roughly, when various components of the argument of  $\phi$  have widely different sensitivity, the resulting gradient is very likely to have poor numerical properties. As explained in [4, 8], the Sobolev gradient compensates, yielding an organized way to define a preconditioned version of the original gradient. This phenomena is pervasive for functionals which arise from discretizations of systems of differential equations. In what follows, we see how to achieve this benefit when a natural norm is not available. Essentially we see how Newton's method fits into the family of Sobolev gradients.

Suppose  $\phi$  is a  $C^3$  real-valued function on  $R^n$  and that a more or less obvious norm as in (36) has not presented itself. Following the opening remarks in [8], if  $u \in R^n$  define  $\beta : R^n \rightarrow R$  by

$$\beta(h) = \phi(h + u), \quad h \in R^n.$$

For  $h$  close to zero, one might expect the sensitivity in  $\beta$  of various components of  $h$  to somewhat match their sensitivity in  $\phi'(u)h$ . Now

$$\phi'(u)h = \langle h, (\nabla\phi)(u) \rangle_{R^n},$$

using the ordinary gradient of  $\phi$  and

$$\beta'(u)h = \langle h, (\nabla\phi)(u + h) \rangle_{R^n}.$$

For sensitivities of  $h$  in both of  $\beta'(u)h$  and  $\phi'(u)h$  to approximately match, one might ask that  $(\nabla\phi)(u)$  and  $\nabla\beta(u)$  (ordinary gradients) be dependent. The following result indicates conditions under which this dependency can be found.

**Theorem 2.** *Suppose  $u \in R^n$  and  $\phi$  is a  $C^3$  function from  $R^n$  to  $R$  so that  $((\nabla\phi)'(u))^{-1}$  exists. Then there is an open interval  $J$  containing 1 and a function  $z : J \rightarrow R^n$  so that*

$$t(\nabla\phi)(u) = (\nabla\phi)(z(t)), \quad t \in J.$$

*Proof.* Denote by  $\gamma$  a positive number so that if  $\|y - u\| \leq \gamma$ , then  $((\nabla\phi)(y))^{-1}$  exists. By basic existence and uniqueness theory for ODE, there is an open interval  $J$  containing 1 and  $z : J \rightarrow R^n$  so that  $z(1) = u$  and

$$z'(t) = ((\nabla\phi)'(z(t)))^{-1}(\nabla\phi)(u), \quad t \in J \tag{37}$$

and hence

$$((\nabla\phi)(z))'(t) = (\nabla\phi)(u), \quad t \in J. \tag{38}$$

Consequently,

$$\begin{aligned} (\nabla\phi)(z(t)) - (\nabla\phi)(z(1)) &= (t - 1)(\nabla\phi)(u), \quad t \in J, \\ (\nabla\phi)(z(t)) &= t(\nabla\phi)(u), \quad t \in J \end{aligned} \tag{39}$$

since  $z(1) = u$ . □

Thus starting at  $z(1) = u$ , the path followed by the solution  $z$  to (37) is a trajectory under a version of continuous Newton's method since  $(\nabla\phi)(u)$  in (39) may be replaced by  $(\nabla\phi)(z(t))$ ,  $t \in J$  with just a change of scalar multiples due to the fact that the vector field directions are not altered. Hence (37) traces out, in a sense, a path of equi-sensitivity. If the interval  $J$  can be chosen to include 0, then  $z(0)$  will be a sought after zero of  $\nabla\phi$ .

By [10] one may substantially reduce the  $C^3$  differentiability in the preceding. This reference also indicates how some of the above considerations apply to systems of PDE in which indicated inverses do not exist.

We now turn to a constrained optimization setting motivated in part by the above. Two versions are indicated, one for Sobolev gradient steepest descent and the other for continuous Newton's method.

First recall that there are two distinct ways systems of differential equations may be placed into an optimization setting. Sometimes a given system of PDE are Euler-Lagrange equations for some functional  $\Phi$ . In this case critical points of  $\Phi$  are precisely solutions to the given system of PDE. In the second case for  $F : X \rightarrow Y$  a  $C^2$  function from a Hilbert space  $X$  into a Hilbert space  $Y$ , think of

$$F(u) = 0$$

as representing a system of differential equations. Such a system may often be placed in an optimization setting by defining

$$\Phi(u) = \frac{1}{2} \|F(u)\|_X^2, \quad u \in X. \quad (40)$$

It is common that, for  $u \in X$ , the range of  $F'(u)$  is dense in  $X$ . In this case it follows that  $u \in X$  is a zero of  $F$  if and only if it is a critical point of  $\Phi$  (see [8]).

In either the Euler-Lagrange or the least squares cases one might want a critical point of  $\Phi$  which lies in some manifold contained in  $X$ . A convenient way that such a manifold might be specified is by means of a function  $B$  from  $X$  into a third Hilbert space  $S$ . In effect one can specify ‘boundary conditions’ or, more accurately, supplementary conditions on a given system by requiring that

$$B(u) = 0 \quad (41)$$

in addition to (40). For each  $u \in X$ , denote by  $P_B(u)$  the orthogonal projection of  $X$  onto  $N(B'(u))$ . For  $X$  a finite dimensional space assume that  $B'(u)B'(u)^*$  has an inverse for all  $u \in X$  where  $B'(u)^*$  is the adjoint of  $B'(u)$  considered as a member of  $L(X, S)$ . This is a natural assumption in that  $S$  would generally have smaller dimension than  $X$ .

With this assumption it may be seen that

$$P_B(u) = I - B'(u)^*(B'(u)B'(u)^*)^{-1}B'(u), \quad u \in X$$

since  $P_B(u)$  is idempotent, symmetric and has range  $N(B'(u))$ . We make the additional assumption that  $P_B$  is  $C^1$ . For  $\phi$  as in (40) and  $(\phi'(x)h = \langle h, \nabla\phi(u) \rangle_X, x, h \in X$ , define

$$(\nabla_B\phi)(x) = P_B(u)(\nabla\phi(x)), x \in X.$$

Then if

$$z(0) = x \in X, \quad z'(t) = -(\nabla_B\phi)(z(t)), \quad t \geq 0, \quad (42)$$

we have the following:

**Theorem 3.** *For  $z$  as in (42),*

$$B(z)'(t) = 0, \quad t \geq 0.$$

This follows since

$$B(z)'(t) = -B'(z(t))P_B(z(t))(\nabla\phi)(z(t)) = 0, \quad t \geq 0. \quad (43)$$

Thus if in (42),  $B(x) = 0$  it follows that  $B(z(t)) = 0, t \geq 0$  and hence if

$$u = \lim_{t \rightarrow \infty} z(t),$$

then  $B(u) = 0$  as well as  $(\nabla\phi)(u) = 0$ .

We now give a similar development for continuous Newton's method by means of the following result. Denote by each of  $X, Y, S$  a Banach space. For  $x \in X$ ,  $r > 0$ ,  $X_r(x)$  denotes the ball in  $X$  of radius  $r$  centered at  $x$ .

**Theorem 4.** *Suppose  $r > 0$ ,  $x_0 \in X$ ,  $F : X_r(x_0) \rightarrow Y$ ,  $B : X_r(x_0) \rightarrow S$  are each  $C^1$ ,  $B(x_0) = 0$ . Suppose also that  $h : X_r(x_0) \rightarrow H$  is a locally Lipschitzian function so that if  $x \in B_r(x_0)$  then*

$$F'(x)(h(x)) = -F(x_0) \text{ and } h(x) \in N(B'(x)), \|h(x)\|_X \leq r. \quad (44)$$

Denote by  $z : [0, 1] \rightarrow X_r(x_0)$  so that

$$z(0) = x_0, z'(t) = h(z(t)), t \in [0, 1]. \quad (45)$$

Then

$$F(z(1)) = 0, B(z(1)) = 0.$$

*Proof.* Note that  $z(t) \in B_r(x_0)$  since  $h(z(t)) \in X_r(0)$ ,  $t \in [0, 1]$ . Also note that

$$(Bz)'(t) = B'(z(t))z'(t) = B'(z(t))h(t) = 0, t \in [0, 1]$$

and so  $B(z(t)) = 0, t \in [0, 1]$  since  $B_r(x_0) = 0$ . Hence  $B(z(1)) = 0$ . But also,

$$F(z)'(t) = F'(z(t))z'(t) = F'(z(t))h(z(t)) = -F(x_0), t \in [0, 1]$$

and so

$$F(z(t)) - F(x_0) = -tF(x_0)$$

that is,

$$F(z(t)) = (1 - t)F(x_0), t \in [0, 1].$$

Thus  $F(z(1)) = 0$ . □

In case  $F'(x)$  has an inverse, continuous and defined on all of  $X$ , one may take in place of (45) the following:

$$h(x) = -F'(x)^{-1}F(x_0), x \in X, \quad (46)$$

more likely recognizable as a Newton vector field or else the conventional field:

$$h(x) = -F'(x)^{-1}F(x), x \in X. \quad (47)$$

With (46) continuous Newton's method is on  $[0, 1]$  and with (47) continuous Newton's method is on  $[0, \infty)$ . In these last two cases, there is no possibility of imposing further boundary conditions using a function  $B$ .

## REFERENCES

- [1] Axelsson, O., *On global convergence of iterative methods*, in: Iterative solution of nonlinear systems of equations, pp. 1-19, Lecture Notes in Math. 953, Springer, 1982.
- [2] Axelsson, O., Faragó I., Karátson J., On the application of preconditioning operators for nonlinear elliptic problems, in: *Conjugate Gradient Algorithms and Finite Element Methods*, pp. 247-261, Springer, 2004.
- [3] Castro, A., Neuberger, J. W., An inverse function theorem via continuous Newton's method. Proc. WCNA, Part 5 (Catania, 2000), *Nonlinear Anal.* 47 (2001), no. 5, 3223–3229.
- [4] Faragó, I., Karátson, J., *Numerical solution of nonlinear elliptic problems via preconditioning operators . Theory and applications*. Advances in Computation, Volume 11, NOVA Science Publishers, New York, 2002.
- [5] Gajewski, H., Gröger, K., Zacharias, K., *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen*, Akademie-Verlag, Berlin, 1974
- [6] Kantorovich, L.V. and Akilov, G.P., *Functional Analysis*, Pergamon Press, 1982.
- [7] Karátson J., Faragó I., Variable preconditioning via quasi-Newton methods for nonlinear problems in Hilbert space, *SIAM J. Numer. Anal.* 41 (2003), No. 4, 1242-1262.
- [8] J.W. Neuberger, *Sobolev Gradients and Differential Equations*, Springer Lecture Notes in Mathematics 1670, 1997.
- [9] Neuberger, J. W. Integrated form of continuous Newton's method, Evolution equations, 331–336, *Lecture Notes in Pure and Appl. Math.*, 234, Dekker, New York, 2003.
- [10] J.W. Neuberger, A near minimal hypothesis Nash-Moser Theorem, *Int. J. Pure. Appl. Math.*, 4 (2003), 269-280.
- [11] Neuberger, J. W., Renka, R. J., Minimal surfaces and Sobolev gradients. *SIAM J. Sci. Comput.* 16 (1995), no. 6, 1412–1427.
- [12] Neuberger, J. W., Renka, R. J., Sobolev gradients: introduction, applications, problems, to appear in: *Contemporary Mathematics* (AMS, Northern Arizona)
- [13] Rheinboldt, W.C., *Methods for solving systems of nonlinear equations* (second edition), CBMS-NSF Regional Conference Series in Applied Mathematics, 70, SIAM, Philadelphia, PA, 1998.

J. Karátson, Dept. of Applied Analysis, ELTE Univ., Budapest, H-1518 Pf. 120, Hungary; [karatson.cs.elte.hu](mailto:karatson.cs.elte.hu)

J. W. Neuberger, Dept. of Mathematics, Univ. of North Texas, Denton, TX 70203-1430, USA; [jwn.unt.edu](mailto:jwn.unt.edu)