Nonlinear Analysis 75 (2012) 6170–6179

Contents lists available at SciVerse ScienceDirect

Nonlinear Analysis

journal homepage: www.elsevier.com/locate/na

A Levenberg–Marquardt method based on Sobolev gradients

Parimah Kazemi [a,b], Robert J. Renka [c,*]

[a] Lindauerstrasse 100, 89079 Ulm, DE, United States

[b] Department of Mathematics and Computer Science, 300 Seward Street, PO Box 248, Ripon, WI 54971-0248, United States

[c] Department of Computer Science & Engineering, University of North Texas, Denton, TX 76203-1366, United States

ARTICLE INFO

Article history:
Received 17 January 2012
Accepted 19 June 2012
Communicated by S. Carl

Keywords:
Gradient system
Least squares
Levenberg–Marquardt
Sobolev gradient
Variable metric

ABSTRACT

We extend the theory of Sobolev gradients to include variable metric methods, such as Newton's method and the Levenberg–Marquardt method, as gradient descent iterations associated with stepwise variable inner products. In particular, we obtain existence, uniqueness, and asymptotic convergence results for a gradient flow based on a variable inner product.

1. Introduction

A nonlinear partial differential equation may be formulated as the problem of minimizing a sum of squared residuals. Consider the numerical solution of the corresponding discretized nonlinear least squares problem. Methods for treating this problem include steepest descent, Newton, Gauss–Newton, and Levenberg–Marquardt which combines a Newton or Gauss–Newton iteration with the method of steepest descent. In order to be effective, the numerical method should emulate an iteration in the infinite-dimensional Sobolev space in which the PDE is formulated. The standard steepest descent and Levenberg–Marquardt methods use a discretized L^2 gradient rather than a Sobolev gradient, and thus lack integrity because they approximate iterations that fail to maintain the smoothness required by the solution. A generalization of the Levenberg–Marquardt method, along with an equivalent trust-region method, is described in [1]. The purpose of this work is to provide a theoretical basis for that method.

Methods for analyzing partial differential equations may be more or less strongly connected to numerical algorithms. The fixed point theorems of Schauder and Leray–Schauder [2], for example, give conditions under which a function defined from a Banach space into itself has a fixed point, but do not provide a method for computing it. Existence proofs based on continuous Newton's method on the other hand, are obviously constructive; see, e.g., [3–5] for zero-finding results of Nash–Moser type [6]. In [7] Newton's method is discussed in relation to gradient descent methods. It is shown that, while the method of steepest descent is locally optimal in terms of the descent direction for a fixed metric, Newton's method is optimal (in a sense which is made precise) in terms of both the direction and the inner product in a variable metric method. In [8,9], nonlinear elliptic problems are treated by quasi-Newton methods with a class of preconditioners chosen via spectral equivalence to produce mesh-independent convergence rates.

In this paper, we describe conditions under which the trajectory of a gradient flow converges to a solution of a system of PDEs. Our results are similar in nature to results obtained in [4,5]. Starting with a system of partial differential equations, we

* Corresponding author.
E-mail addresses: parimah.kazemi@gmail.com (P. Kazemi), renka@cs.unt.edu (R.J. Renka).

0362-546X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.na.2012.06.022

form a variational problem from the sum of squared residuals. We then define a gradient system [10] using a variable metric, and give conditions under which a global solution exists and under which the asymptotic limit exists and is a zero of the least squares functional. When the gradient system is discretized in time and space, the resulting algorithm is a generalized Levenberg–Marquardt method. Our method is thus constructive in nature.

The paper is organized as follows. In Section 2, we review basic results from the theory of Sobolev gradients. In Section 3, we define the problem. In Section 4, we present results regarding existence, uniqueness, and convergence of the gradient flow, and in Section 5, we conclude the paper.

2. Sobolev gradients

Consider the problem of finding critical points of a C^1 energy functional ϕ defined on a Hilbert space H. The Fréchet derivative $\phi'(u)$ is a bounded linear functional on H, and is therefore represented by an element of H. This element is the Sobolev gradient of ϕ at u and is denoted by $\nabla_H\phi(u)$:

$$\phi'(u)h = \langle h, \nabla_H\phi(u)\rangle_H, \quad h \in H.$$

Note that the gradient depends on the inner product attached to H. Now consider the evolution equation

$$z(0) = z_0 \in H \quad \text{and} \quad z'(t) = -\nabla_H\phi(z(t)), \quad t \geq 0. \tag{1}$$

The energy ϕ is non-increasing on the trajectory z. Existence, uniqueness, and asymptotic convergence to a critical point are established by the following two theorems taken from [11, Chapter 4].

Theorem 1. *Suppose that ϕ is a non-negative C^1 real-valued function on a Hilbert space H with a locally Lipschitz continuous Sobolev gradient. Then for each $z_0 \in H$ there is a unique global solution of (1).*

Definition 1. The energy functional ϕ satisfies a *gradient inequality* on $K \subseteq H$ if there exists $\theta \in (0, 1)$ and $m > 0$ so that for all $x \in K$

$$\|\nabla_H\phi(x)\|_H \geq m\phi(x)^\theta.$$

Theorem 2. *Suppose that ϕ is a non-negative C^1 functional on H with a locally Lipschitz continuous gradient, z is the unique global solution of (1), and ϕ satisfies a gradient inequality on the range of z. Then $\lim_{t\to\infty} z(t)$ exists and is a zero of the gradient, where the limit is defined by the H-norm. By the gradient inequality, the limit is also a zero of ϕ.*

The above theorems provide a firm theoretical basis for the numerical treatment of a system of nonlinear PDEs by a gradient descent method that emulates (1); i.e., discretization in time and space results in the method of steepest descent with a discretized Sobolev gradient. Note that the Sobolev gradient method differs from methods based on calculus of variations in which the Euler–Lagrange equation is solved. Forming the Euler–Lagrange equation requires integration by parts to obtain the element that represents $\phi'(u)$ in the L^2 inner product. This L^2 gradient is usually only defined on a Sobolev space of higher order than that of H. Hence, unlike the Sobolev gradient, the L^2 gradient is only densely defined on the domain of ϕ. For gradient flows involving the L^2 gradient, existence and uniqueness results similar to those of Theorem 1 and Theorem 2 may be proved, but only under stricter assumptions.

In most applications of the Sobolev gradient method to date the underlying Hilbert space structure is equipped with a fixed metric, typically the metric associated with the Sobolev space $H^k(\Omega)$ for some positive integer k. While substituting the discretized H^k gradient for the L^2 gradient can result in a dramatic improvement in numerical performance, the method is still a gradient descent algorithm and has at best a linear rate of convergence. In [12,8,9,1,13], a quasi-Newton or variable metric method is used in place of a fixed metric to obtain superlinear rates of convergence. In this work, we establish a theoretical basis for the method of Sobolev gradients with a variable metric.

3. Defining the problem in a Hilbert space setting

Our focus in this section is on setting up a theoretical framework for the treatment of a least squares variational formulation of a system of nonlinear partial differential equations by a variable metric method.

3.1. Preliminaries

In order to simplify the notation we restrict consideration to a single first-order partial differential equation. Extension to higher-order derivatives and Cartesian-product spaces is straightforward. For an appropriate domain Ω in $\mathbb{R}^d$, denote a Sobolev space by $H = H^{1,2}(\Omega)$, and let $L = L^2(\Omega)$ so that H is densely and continuously embedded in L. Also, let $L_1 = L^n$ for $n = d + 1$, and define $D : H \to L_1$ by $Du = \{\partial_\alpha u : |\alpha| \leq 1\}$ for d-dimensional multi-index α. We can view H as the space of first terms of the closure of $\{Du : u \in C^1(\Omega)\}$ in L_1 equipped with the norm $\|u\|_H = \|Du\|_{L_1}$ ([14, Section 3.5]).

For a positive integer m define $L_2 = L^m$, and let $r : \mathbb{R}^n \to \mathbb{R}^m$ be a C^1 function so that the composition $r \circ D : H \to L_2$ defined by $((r \circ D)(u))(x) = r(Du(x))$ for almost every $x \in \Omega$ is well defined and C^1 with $((r \circ D)'(u)v)(x) = r'(Du(x))Dv(x)$ for almost every x. The mapping r is a Nemytskii operator.