NEWTON'S METHOD IN HIGHER DIMENSIONS

MARKUS GRASMAIR

In this note we will briefly discuss the application of Newton's method for the solution of systems of equations in several variables. Here we consider systems of n equations in n variables; that is, the number of equations is the same as the number of unknowns. For simplicity, we assume that the right hand side of all the equations is equal to zero; the system thus has the form

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0, \\ f_2(x_1, x_2, \dots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0, \end{aligned}$$

with functions $f_j \colon \mathbb{R}^n \to \mathbb{R}$. Usually, we will collect the variables $x_1,\dots,x_n$ in a vector $\vec{x}$ and the functions $f_1,\dots,f_n$ in a vector of functions:

$$F(\vec{x}) := \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_n(\vec{x}) \end{pmatrix}.$$

We are thus interested in solving the system of equations

$$F(\vec{x}) = \vec{0},$$

given a function $F \colon \mathbb{R}^n \to \mathbb{R}^n$.

Assuming that the functions f_j are differentiable, we now define the *Jacobian* of F at a point $\vec{x} \in \mathbb{R}^n$ as

$$JF(\vec{x}) = \begin{pmatrix} \partial_1 f_1(\vec{x}) & \partial_2 f_1(\vec{x}) & \dots & \partial_n f_1(\vec{x}) \\ \partial_1 f_2(\vec{x}) & \partial_2 f_2(\vec{x}) & \dots & \partial_n f_2(\vec{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_n(\vec{x}) & \partial_2 f_n(\vec{x}) & \dots & \partial_n f_n(\vec{x}) \end{pmatrix}.$$

That is, $JF(\vec{x})$ is an $n \times n$ matrix, where the j-th row contains the partial derivatives of the function f_j. Or, the j-th row of $JF(\vec{x})$ equals $\nabla f_j(\vec{x})^T$.

We will mainly need the Jacobian of a vector valued function F, because it appears in the Taylor series expansion of F:

Lemma 1. *(First order Taylor series) If the function $F \colon \mathbb{R}^n \to \mathbb{R}^n$ is twice continuously differentiable, we have*

$$F(\vec{x} + \vec{h}) = F(\vec{x}) + JF(\vec{x})\vec{h} + o(\|h\|^2)$$

for every $\vec{x}$, $\vec{h} \in \mathbb{R}^n$.

The idea of Newton's method in higher dimensions is now very similar to the one-dimensional idea: Given an iterate $\vec{x}_k$, we define the next iterate $\vec{x}_{k+1}$ by linearizing

Date: October 2015.

the equation $F(\vec{x}) = \vec{0}$ around $\vec{x}_k$ and solving the linearized equation. That is, we define $\vec{x}_{k+1}$ as the solution of the linear equation

$$F(\vec{x}_k) + JF(\vec{x}_k)(\vec{x} - \vec{x}_k) = \vec{0}.$$

Formulated slightly differently, this yields:

(1) Start with some initial guess $\vec{x}_0$ of the solution of the equation $F(\vec{x}) = \vec{0}$.
(2) Set $k = 0$.
(3) Iterate until the result is satisfactory:
   (a) Compute the solution $\vec{h}_k \in \mathbb{R}^n$ of the linear equation
   $$JF(\vec{x}_k)\vec{h} = -F(\vec{x}_k).$$
   (b) Define
   $$\vec{x}_{k+1} = \vec{x}_k + \vec{h}_k.$$
   (c) Increase the counter k by 1.

Alternatively, we can write the update rule as

$$\vec{x}_{k+1} = \vec{x}_k - JF(\vec{x}_k)^{-1} F(\vec{x}_k).$$

Example 2. We consider the system

$$\begin{aligned} xe^y &= 1, \\ -x^2 + y &= 1. \end{aligned}$$

First we rewrite this as

$$\begin{aligned} xe^y - 1 &= 0, \\ -x^2 + y - 1 &= 0. \end{aligned}$$

That is,

$$F(x,y) = \begin{pmatrix} xe^y - 1 \\ -x^2 + y - 1 \end{pmatrix}$$

and

$$JF(x,y) = \begin{pmatrix} e^y & xe^y \\ -2x & 1 \end{pmatrix}.$$

In this two-dimensional case we can also compute the inverse of $JF(x,y)$ as

$$JF(x,y)^{-1} = \frac{1}{e^y + 2x^2 e^y} \begin{pmatrix} 1 & -xe^y \\ 2x & e^y \end{pmatrix} = \frac{1}{1 + 2x^2} \begin{pmatrix} e^{-y} & -x \\ 2xe^{-y} & 1 \end{pmatrix}.$$

We therefore obtain the iteration

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \frac{1}{1 + 2x_k^2} \begin{pmatrix} e^{-y_k} & -x_k \\ 2x_k e^{-y_k} & 1 \end{pmatrix} \begin{pmatrix} x_k e^{y_k} - 1 \\ -x_k^2 + y_k - 1 \end{pmatrix}.$$

Starting with $x_0 = 0$ and $y_0 = 0$ (for instance) we thus obtain

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \frac{1}{1 + 0^2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Then

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{1 + 2 \cdot 1^2} \begin{pmatrix} e^{-1} & -1 \\ 2e^{-1} & 1 \end{pmatrix} \begin{pmatrix} e^1 - 1 \\ -1 \end{pmatrix} \approx \begin{pmatrix} 0.45596 \\ 0.91192 \end{pmatrix}.$$

Further iterations yield

$$\begin{pmatrix} x_3 \\ y_3 \end{pmatrix} \approx \begin{pmatrix} 0.32235 \\ 1.08606 \end{pmatrix}$$

and

$$\begin{pmatrix} x_4 \\ y_4 \end{pmatrix} \approx \begin{pmatrix} 0.33017 \\ 1.10895 \end{pmatrix}.$$

Newton's method is probably the most important method for the numerical solution of systems of non-linear equations, because of its (usually) fast convergence:

Theorem 3. *Assume that the function $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is three times differentiable, that $\vec{s} \in \mathbb{R}^n$ is a solution of $F(\vec{s}) = \vec{0}$, and that $JF(\vec{s})$ is an invertible matrix. Then Newton's method converges quadratically to $\vec{s}$ whenever $\vec{x}_0$ is chosen sufficiently close to $\vec{s}$.*

The main limitation of Newton's method is that it requires the solution of a linear system involving the Jacobian of F in each step. This is obviously problematic if the function F is not differentiable. However, it can also happen that the function F theoretically is differentiable, but the Jacobian is for some reason not available for the user. Other problems arise if the dimension n becomes very large, in which case the numerical solution of the linear system (which has to be performed in each Newton step) may take too much time, or the Jabion JF is either non-invertible or close to being non-invertiable. Then the solution of the linear system is either impossible or unstable.

Department of Mathematics, Norwegian University of Science and Technology, 7491 Trondheim, Norway

E-mail address: markus.grasmair@math.ntnu.no