1 Optimization Algorithms

Suppose X, Y are Banach spaces, $U \subset X$ is open and $F : U \subset X \to \mathbb{R}$ is differentiable on U. If F has a local minimum at $x \in U$ then $F'(x) = 0$. There are two approaches to fininding local minima of F. By creating a minimizing sequence or by locating the roots of $F'(x)$. We breifly give an overview of both approaches. We start by locating the roots of $F'(x)$. The method that we have in mind is, of course, Newtoon's method. In what follows we study a class of generalized Newton methods.For this purpose we suppose that $F : U \subset X \to Y$ and $x \in U$ is a solution of the equation

$$F(x) = 0. \tag{1}$$

The classical Newton method uses the iterations

$$x_{k+1} = x_k - F'(x_k)^{-1} F(x_k),$$

starting with an initial guess x_0 to approximate x. A generalization of this method is the following iterations

$$x_{k+1} = x_k - A_k^{-1} F(x_k), \tag{2}$$

where $\{A_k\}$ is a sequence of operators in $\mathcal{L}(X,Y)$ such that $A_k \in B_{\mathcal{L}(X,Y)}(F'(x), \rho)$ for some $\rho > 0$ and all k. For simplicity, we will set $A = F'(x)$. If A is bijective and ρ is sufficiently small then one can show, using Neumann's series, that $A_k^{-1} \in \mathcal{L}(Y,X)$ and

$$\left\|A_k^{-1} - A^{-1}\right\| \leq \frac{\rho \left\|A^{-1}\right\|^2}{1 - \rho \left\|A^{-1}\right\|}. \tag{3}$$

In particular the sequence $\left\{A_k^{-1}\right\}$ is uniformly bounded. Furthermore, the assumption that A is bijective implies that x is an isolated solution of (1). Thus, there exists a $r > 0$ such that x is the unique solution of (1) in $B_X(x, r)$.

Exercise 1 *Prove formula 3.*

Proposition 2 *Under the foregoing assumptions, if $r > 0$, $\rho > 0$ are sufficiently small then all the elelments x_k produced by the iterations (2) are in $B_X(x, r)$. Moreover, $x_k \to x$ and*

$$\|x_k - x\| \leq \left(\frac{1}{2}\right)^k \|x_0 - x\|$$

Proof. Subtracting x from both sides of (2) we get

$$\begin{aligned}
x_{k+1} - x &= x_k - x - A_k^{-1} F(x_k) \\
&= x_k - x - A_k^{-1} A (x_k - x) + o((x_k - x)) \\
&= x_k - x - \left(A^{-1} + \triangle_k\right) A (x_k - x) + o((x_k - x)) \\
&= -\triangle_k A (x_k - x) + o((x_k - x)),
\end{aligned}$$

where $\triangle_k = A_k^{-1} - A^{-1}$. Then

$$\begin{aligned}\|x_{k+1} - x\| &\leq \|\triangle_k\| \|A\| \|x_k - x\| + o(\|x_k - x\|) \\ &\leq \gamma \|A\| \|x_k - x\| + o(\|x_k - x\|)\end{aligned}$$

where $\gamma = \frac{\rho\|A^{-1}\|^2}{1-\rho\|A^{-1}\|}$. We will show by induction that for sufficiently small $r > 0,\ \rho > 0$, we have $x_k \in B_X(x, r)$.

$$\begin{aligned}\|x_1 - x\| &\leq \gamma \|A\| \|x_0 - x\| + o(\|x_0 - x\|) \\ &\leq \left(\gamma \|A\| + \frac{o(\|x_0 - x\|)}{\|x_0 - x\|}\right) r\end{aligned}$$

Since $\frac{o(\|y-x\|)}{\|y-x\|} \to 0$ as $r \to 0$, for sufficiently small $r > 0$ we get $\frac{o(\|x_0-x\|)}{\|x_0-x\|} \leq \frac{1}{4}$. Since $\gamma \to 0$ as $\rho \to 0$, for sufficiently small $\rho > 0$ we get $\gamma \|A\| \leq \frac{1}{4}$. Hence, $\|x_1 - x\| \leq \frac{1}{2} r$. i.e., $x_1 \in B_X(x, r)$. A similar argument then shows that if $x_k \in B_X(x, r)$ then $x_{k+1} \in B_X(x, r)$ for the same values of r, ρ above. Furthermore,

$$\begin{aligned}\frac{\|x_{k+1} - x\|}{\|x_k - x\|} &\leq \gamma \|A\| + \frac{o(\|x_k - x\|)}{\|x_k - x\|} \\ &\leq \frac{1}{2}.\end{aligned}$$

This gives

$$\|x_k - x\| \leq \left(\frac{1}{2}\right)^k \|x_0 - x\|,$$

which implies that $x_k \to x$ as $k \to \infty$.

Remark 3 *In the proof of the above proposition, the value of ρ is explicitly available. The value of r however depends on the behaviour of the derivative of F at x. For example, if F is C^1 on U then the remainder term $o(\cdot)$ takes the form*

$$o(x_k - x) = \int_0^1 (F'(x + t(x_k - x) - F'(x))(x_k - x)\, dt$$

from which we obtain the estimate

$$\frac{o(\|x_k - x\|)}{\|x_k - x\|} \leq \int_0^1 \|F'(x + t(x_k - x) - F'(x)\|\, dt$$

and, in this case, r is the δ needed in the definition on the continuity of F' when $\varepsilon = \frac{1}{4}$.

■

Recall that if $F : U \subset X \to \mathbb{R}$ has two derivatives then $F'(x) : V \subset X \to X^*$ and $F''(x) \in \mathcal{L}(X, X^*)$. As a consequence to this proposition we get

Corollary 4 *If $F : U \subset X \to \mathbb{R}$ is twice differentiable at x, has a local minimum at $x \in U$ and $F''(x)$ is bijective then there exist $r > 0$, $\rho > 0$ such that, for any choice of $x_0 \in B_X(x, r)$ and any choice of a sequence $\{A_k\} \in B_{\mathcal{L}(X,X^*)}(F''(x), \rho)$, the sequence*

$$x_{k+1} = x_k - A_k^{-1} F'(x_k)$$

belongs to $B_X(x, r)$ and converges to x greometrically. i.e.,

$$\|x_k - x\| \leq \left(\frac{1}{2}\right)^k \|x_0 - x\|.$$

2 Descent Methods

In this section we assume that $F : U \subset X \to \mathbb{R}$ has a Gateaux derivative on all of U. Suppose F has a minimum $x \in U$. An algorithm to find the minimizer x consists of generating a sequence $\{x_k\}$ in U such that $x_k \to x$. Descent methods generate x_{k+1} from x_k by moving a distance ρ_k in a direction w_k in which F decreases at x_k. The direction of maximum descent at x_k is the direction opposite to the gradient $F'(x_k)$. i.e., we may take $w_k = F'(x_k) / \|F'(x_k)\|$ provided that $F'(x_k) \neq 0$. We may then find x_{k+1} by moving in this direction as far as possible. In other words we set ρ_k to be the value at which

$$\inf_{\rho > 0} F(x_k - \rho w_k)$$

is attained and then define x_{k+1} by

$$x_{k+1} = x_k - \rho_k w_k$$

This method is called the maximum the method of maximum descent with optimal choice of parameters. Evidently this strategy generates a sequence $\{x_k\}$ such that $F(x_{k+1}) \leq F(x_k)$. As for the convergenc of this method, we have the following theorem.

Theorem 5 *Suppose H is a Hilbert space and $F : H \to \mathbb{R}$ is coercive, continuous with a continuous Gateaux derivative. Then the sequence $\{x_k\}$ generated by the method of maximum descent with optimal choice of papameters is a minimizing sequence. i.e., $x_k \to x$ where x is a local minimum of F.*

Proof. Omitted. ■

A variant of this method is called the method of conjugate gradients. To describe the method we assume H is a Hilbert space and $F : H \to \mathbb{R}$ has a positive definite second Gateaux derivative. Let $x_0 \in H$ be arbitrary and set $w_0 = F'(x_0) / \|F'(x_0)\|$ provided that $F'(x_0) \neq 0$ (otherwise F has a local minimum at x_0.) Suppose that x_k, w_k have been determined. Set $\rho_k > 0$ to be a point of minimum of $F(x_k - \rho w_k)$. i.e.,

$$F(x_k - \rho_k w_k) = \inf_{\rho > 0} F(x_k - \rho w_k)$$

This occures at the point ρ_k such that

$$\frac{d}{d\rho}F\left(x_k-\rho w_k\right)|_{\rho=\rho_k}=0.$$

Set

$$x_{k+1}=x_k-\rho_k w_k.$$

Then

$$\left(F'\left(x_{k+1}\right),w_k\right)=0.$$

Define a vector $\widetilde{w}_{k+1}\in H$ by

$$\widetilde{w}_{k+1}=F'\left(x_{k+1}\right)+\lambda_{k+1}w_k$$

where $\lambda_{k+1}\in\mathbb{R}$ is chosen such that

$$\left(F''\left(x_{k+1}\right)\widetilde{w}_{k+1},w_k\right)=0.$$

Hence, λ_{k+1} is given by

$$\lambda_{k+1}=-\frac{\left(F''\left(x_{k+1}\right)F'\left(x_{k+1}\right),w_k\right)}{\left(F''\left(x_{k+1}\right)w_k,w_k\right)}$$

Notice that the positive definiteness of F'' means that the denominator above is nonzero. The direction w_{k+1} is defined by

$$w_{k+1}=\widetilde{w}_{k+1}/\left\|\widetilde{w}_{k+1}\right\|.$$

Theorem 6 *If $F:H\to\mathbb{R}$ is coercive and has a positive definite second Gateaux derivative then the conjugate gradient method converges to the unique point of minimum x.*

Exercise 7 *Prove Thoerems 5, 6 for the case*

$$F\left(x\right)=\frac{1}{2}a\left(x,x\right)-\langle\eta,x\rangle$$

where $a\left(\cdot,\cdot\right)$ is a coercive continuous symetric bilinear form on H and $\eta\in H$.