Investigating Iterative Methods in Optimization

Shuvomoy Das Gupta

April 5, 2017

In this note, we will state some results on investigating different iterative schemes. In optimization theory, nonlinear analysis we often construct iteration scheme in the hope that the scheme will eventually produce a desired point, *e.g.*, fixed point of some operator. There are four main methods of investigating iteration schemes: Lyapunov's first method, Lyapunov's second method, Banach-Picard iteration and Krasnosel'skii-Mann iteration for nonexpansive operator.

Notation and notions.

• $\mathbf{N}$: The set of natural numbers $0, 1, \ldots$

• $\rho(A)$: Spectral radius of a square matrix $A \in \mathbf{R}^{n\times n}$, where

$$\rho(A) = \max_{1\leq i\leq n} |\lambda_i|.$$

• $\mathrm{Fix}T$: the set of fixed points of the operator $T : \mathbf{R}^n \to \mathbf{R}^n$, *i.e.*, if $x^* \in \mathrm{Fix}T$ then $Tx^* = x^*$. Often the optimal solution of an optimization problem is the fixed point of some operator.

• DTx : Derivative (or Jacobian) of the differentiable operator $T : \mathbf{R}^n \to \mathbf{R}^m$ evaluated at point x. $DTx \in \mathbf{R}^{m\times n}$ is a matrix, given by

$$(DTx)_{ij} = \frac{\partial T_i(x)}{\partial x_j} \quad i = 1, \ldots, m,\ j = 1, \ldots, n.$$

• $\|x\|$: Euclidean norm of the vector $x \in \mathbf{R}^n$.

• $I_n \in \mathbf{R}^{n\times n}$: Identity matrix.

• $0_n \in \mathbf{R}^{n\times n}$: a matrix containing all zeros.

• $x_n \to x^*$ means that the sequence $(x_n)_{n\in\mathbf{N}} \subseteq \mathbf{R}^n$ converges to x^*, *i.e.*,

$$\lim_{n\to\infty} \|x_n - x^*\| = 0.$$

Contents

Lyapunov's first method

Convergence of linear iteration scheme. The iteration scheme

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = \underbrace{A}_{\in\mathbf{R}^{n\times n}} x_n$$

converges to zero, *i.e.*, $x_n \to 0$ if and only if $\rho(A) < 1$.

Local convergence of nonlinear iteration scheme. Suppose

• T : some differentiable operator from $\mathbf{R}^n$ to $\mathbf{R}^n$,

• $x^* \in \mathrm{FixT}$,

• $\rho = \rho(DTx^*) < 1$.

Consider the iterative scheme

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = Tx_n.$$

Then,

• $x_n \overset{\text{local}}{\to} x^*$,[1] and

• More specifically,

$$\forall_{\epsilon\in]0,1-\rho[}\ \exists_{\delta>0: x_0\in B(x^*;\delta)}\ \exists_{\gamma\geq 0} \quad \|x_n - x^*\| \leq \gamma(\rho+\epsilon)^n.$$

[1] if x_0 is close to x^* then x_n will converge to x^*

Global convergence of nonlinear iteration scheme. Suppose,

- $A \in \mathbf{R}^{n\times n}$: $-A$ is stable
- $U \in \mathbf{R}^{n\times n}$: solution of the matrix equation

$$UA + A^T U = I_n,$$

- T : some nonlinear operator from $\mathbf{R}^n \to \mathbf{R}^n$, such that it satisfies

$$\|Tx\| \leq L\|x\|,\ L < \frac{1}{2\|U\|},$$

- γ : $\gamma \in]0, \frac{\|U\|^{-1}-2L}{(L+\|A\|)^2}[$,
- q :

$$q = 1 + \gamma L + \frac{\gamma^2(\|A\|+L)^2 - \gamma\|U\|^{-1}}{2}.$$

Consider the iteration scheme

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = x_n - \gamma\left(Ax_n + Tx_n\right).$$

Then,

- $x_n \to 0$ with rate of geometric progression for any $x_0 \in \mathbf{R}^n$,
- More specifically,

$$(\forall n \in \mathbf{N}) \quad \|x_n\|^2 \leq \|U^{-1}\|\|U\|q^n\|x_0\|^2.$$

Superlinear convergence of nonlinear iteration scheme.

- $T : \mathbf{R}^n \to \mathbf{R}^n$, nonlinear operator
- $x^* \in$ FixT,
- $S = \{x \in \mathbf{R}^n \mid \|x - x^*\| \leq \|x_0 - x^*\|\}$, where x_0 is the initial point of the iteration scheme,
- T : Frechet differentiable on S,
- DT is L-Lipschitz continuous on S *i.e.,*

$$(\forall x, y \in S) \quad \|DTx - DTy\| \leq L\|x - y\|,$$

- $DTx^* = 0_n$,
- $q = \frac{L}{2}\|x_0 - x^*\| < 1.$

Consider the iterative scheme

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = Tx_n.$$

Then,

- $x_n \to x^*$, and
- More specifically,

$$\|x_n - x^*\| \leq \frac{2}{L}q^{2^n}.$$

Algorithm 1 Investigating an iterative scheme using Lyapunov's second scheme

Input: An iterative scheme

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = Tx_n, \tag{1}$$

where $x_n \in \mathbf{R}^n$.

Output: A certificate that $x_n \to x^*$, where $x^* \in \mathrm{Fix}T$.

Scheme:

1. Construct a nonnegative scalar function $V : \mathbf{R}^n \to \mathbf{R}_+$, which is called the Lyapunov function, such that $V(x^*) = 0$,and if $V(x) \leq \delta$, then we have some estimate of how far is x is from x^*, *e.g.*, $\|x - x^*\| \leq \delta$.

2. Construct the nonnegative scalar sequence $(u_n)_{n\in\mathbf{N}}$ generated by the iterative scheme:

$$(\forall n \in \mathbf{N}) \quad u_n = V(x_n). \tag{2}$$

Find how u_n is related with its previous term u_{n-1}, *i.e.*, find a relation of type

$$u_{n+1} \leq \phi(u_n). \tag{3}$$

3. By analyzing (3)

- if we can show that $u_n \to 0$, then $x_n \to x^*$,
- if we can show that $u_n \leq \delta$ as $n \to \infty$, then we have an estimate how far x_n will be from x^*, *e.g.*, $\|x_n - x^*\| \leq \delta$. If we have some control over the parameter δ, then by making it arbitrarily small, we can reduce the gap between x_n and x^*, which may suffice for practical purpose.

Lyapunov's second method

Basic steps in Lyapunov's second method. The most commonly used method for proving the convergence of arbitrary iteration scheme is Lyapunov's second method. The basic steps are of the scheme is given by Algorithm (1).

So, essentially succeeding in proving the convergence of the original iteration scheme (1), is equivalent to showing the convergence of the scheme (2) from inequality (3). The success of Lyapunov's method hinges on whether we can guarantee the convergence of u_n in (2), which again depends of nature of the function ϕ in (3). Now we consider different types of ϕ below.

When ϕ is of affine type with constant coefficients. Suppose,

$$u_{n+1} \leq \alpha u_n + \beta,$$

where $\alpha \in [0, 1[$, and $\beta > 0$. Then,[2]

$$u_n \leq \frac{\beta}{1-\alpha} + \left(u_0 - \frac{\beta}{1-\alpha}\right)\alpha^n.$$

[2] Note that

$$u_n \leq \frac{\beta}{1-\alpha} + \left(u_0 - \frac{\beta}{1-\alpha}\right)\alpha^n$$

means that

$$u_n \leq \frac{\beta}{1-\alpha}, \text{ as n} \rightarrow \infty,$$

as $q \in [0, 1)$.

When ϕ is affine but with two other sequences as coefficients: Result 1. Suppose,

- $u_n \geq 0$,
- $u_{n+1} \leq (1 + \boxed{\alpha_n})u_n + \boxed{\beta_n}$,
- $\alpha_n \geq 0$, $\sum_{n=0}^{\infty} \alpha_n < \infty$,
- $\beta_n \geq 0$, $\sum_{n=0}^{\infty} \beta_n < \infty$.

Then, there exists a nonnegative number $u \geq 0$, *i.e.*, $u_n \rightarrow u \geq 0$.[3]

[3] Often α_n and β_n are associated with the step sizes in the related optimization problem.

When ϕ is affine but with two other sequences as coefficients: Result 2. Suppose,

- $u_{n+1} \leq \alpha_n u_n + \beta_n$,
- $\alpha_n \in [0, 1[$, $\sum_{n=0}^{\infty}(1 - \alpha_n) = \infty$,
- $\beta_n \geq 0$,
- $\frac{\beta_n}{1-\alpha_n} \rightarrow 0$.

Then, $\overline{\lim}_{n\rightarrow\infty} u_n \leq 0$ as $n \rightarrow \infty$. More specifically if $u_n \geq 0$, then $u_n \rightarrow 0$.

When ϕ is affine but with two other sequences as coefficients: Chung's 1st result. Suppose, [4]

[4] Chung's convergence theorem ascertains the convergence of u_n to 0. Moreover, it gives us at which rate u_n converge to 0.

- $u_n \geq 0$,
- $\boxed{u_{n+1} \leq \left(1 - \frac{c}{n}\right) u_n + \frac{d}{n^{p+1}}}$: $d > 0,\ p > 0,\ c > 0$.

Then,

- if $c > p$, then[5]
$$u_n \leq \frac{d}{(c-p)n^p} + o\left(\frac{1}{n^p}\right),$$
- if $c = p$, then
$$u_n = O\left(\frac{\log n}{n^c}\right),$$
- if $c < p$, then
$$u_n = O\left(\frac{1}{n^c}\right).$$

[5] Consider the sequences $(x_n)_{n\in\mathbf{N}}$, $(y_n)_{n\in\mathbf{N}}$.

Small o notation. We say that $x_n = o(y_n)$ if

$$\lim_{n\rightarrow\infty} \frac{\|x_n\|}{\|y_n\|} = 0.$$

Big O notation. We say that $x_n = O(y_n)$ if there exists some $\alpha > 0$ such that

$$\frac{\|x_n\|}{\|y_n\|} \leq \alpha \text{ as } n \rightarrow \infty.$$

Clearly the small o notation gives a more powerful statement in comparison with big O.

When ϕ is affine but with two other sequences as coefficients: Chung's 2nd result. Suppose,

- $u_n \geq 0$,
- $\boxed{u_{n+1} \leq \left(1 - \frac{c}{n^s}\right) u_n + \frac{d}{n^t}}$: $s \in]0, 1[, t > s$.

Then,

$$u_n \leq \frac{d}{cn^{t-s}} + o\left(\frac{1}{n^{t-s}}\right).$$

When ϕ is nonlinear: Result 1. Suppose,

- $u_n > 0$,
- $\boxed{u_{n+1} \leq u_n - \alpha_n u_n^{1+p}}$: $\alpha_n \geq 0$, $p > 0$.

Then,

- $u_n \leq u_0 \left(1 + p u_0^p \sum_{i=0}^{n-1} \alpha_i\right)^{-\frac{1}{p}}$.
- $(\alpha_n = \alpha, p = 1) \Rightarrow u_n \leq \frac{u_0}{1+\alpha n u_0}$

When ϕ is nonlinear: Result 2. Suppose,

- $u_n \geq 0$,
- $\boxed{u_{n+1} \leq (1 + \alpha_n) u_n + \beta_n - \gamma_n f(u_n)}$:
  - $\alpha_n, \beta_n, \gamma_n \geq 0$,
  - $\alpha_n \to 0$, $\beta_n \to 0$, $\frac{\alpha_n}{\gamma_n} \to 0$, $\frac{\beta_n}{\gamma_n} \to 0$,
  - $\sum_{n=0}^{\infty} \gamma_n = \infty$,
  - $f(0) = 0$,
  - $(\forall u > 0) \quad f(u) > 0$,
  - $(\forall v, u) \quad v \geq u \Rightarrow f(v) \geq f(u)$,
- Either $\alpha_n = 0$, or f is a convex function.

Then, $u_n \to 0$.

Banach-Picard iteration for contraction operator

Contraction mapping. An operator $T : \mathbf{R}^n \to \mathbf{R}^n$ is called a contraction operator if

$$(\forall x \in \mathbf{R}^n)\ (\forall y \in \mathbf{R}^n) \quad \|Tx - Ty\| \leq L\|x - y\|,$$

where $L \in [0, 1)$. Any contraction mapping has a unique fixed point.

Banach-Picard fixed point theorem. Suppose $T : \mathbf{R}^n \to \mathbf{R}^n$ is a contraction operator with Lipschitz constant $L \in [0,1)$, and consider the Banach-Picard iteration scheme

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = Tx_n.$$

Then the following hold:

- There exists some $x^* \in \mathbf{R}^n$ such that $\mathrm{Fix}T = \{x^*\}$,
- $(\forall n \in \mathbf{N}) \quad \|x_{n+1} - x^*\| \leq L\|x_n - x^*\|$,
- $(\forall n \in \mathbf{N})\ \|x_{n+1} - x^*\| \leq L^n\|x_0 - x^*\|$,
- $x_n \to x^*$,
- A priori error estimate: $(\forall n \in \mathbf{N}) \quad \|x_n - x^*\| \leq \frac{L^n}{1-L}\|x_0 - x_1\|$,
- A posteriori error estimate: $(\forall n \in \mathbf{N}) \quad \|x_n - x^*\| \leq \frac{1}{1-L}\|x_n - x_{n+1}\|$,
- $\frac{1}{1+L}\|x_0 - x_1\| \leq \|x_0 - x^*\| \leq \frac{1}{1-L}\|x_0 - x_1\|$.

A variant of Banach-Picard iteration. Suppose $T : \mathbf{R}^n \to \mathbf{R}^n$ is a mapping such that

- T satisfies

  $$(\forall n \in \mathbf{N})\ (\forall x \in \mathbf{R}^n)\ (\forall y \in \mathbf{R}^n) \quad \|T^n x - T^n y\| \leq \beta_n\|x - y\|,$$

  [6]where $(\beta_n)_{n=0}^{\infty}$ is a nonnegative sequence which is summable *i.e.*, $\sum_{n=0}^{\infty} \beta_n < \infty$.
- $(\forall n \in \mathbf{N}) \quad \alpha_n = \sum_{k=n}^{\infty} \beta_k$.

Consider the Banach-Picard iteration scheme,

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = Tx_n.$$

Then the following hold,

- There exists some $x^* \in \mathbf{R}^n$ such that $\mathrm{Fix}T = \{x^*\}$,
- $x_n \to x^*$,
- $(\forall n \in \mathbf{N}) \quad \|x_n - x^*\| \leq \alpha_n\|x_0 - x_1\|$.

[6] $T^0 x = Ix = x$, and for $n \geq 1$, $T^n x = \underbrace{T \cdots T}_{\substack{n\text{-fold}\\ \text{composition}}} x$.

Krasnosel'skii-Mann iteration for nonexpansive operator

Nonexpansive operator. Consider an operator

$$T : \overbrace{D}^{\text{nonempty},\subseteq \mathbf{R}^n} \to \mathbf{R}^n.$$

Then T is a nonexpansive operator if

$$(\forall x \in D)\ (\forall y \in D) \quad \|Tx - Ty\| \leq \|x - y\|.$$

This is a very important class of operators, as often a convex optimization problem can be reformulated as fixed point calculation problem of some nonexpansive operator. However, the set of fixed points of a nonexpansive operator may be empty. However there is Browder-Gohde-Kirk theorem that gives the condition of existence of fixed points for a nonexpansive operator.

Browder-Gohde-Kirk theorem for existence of fixed points of a nonexpansive operator. Suppose,

- D : nonempty, bounded, convex, subset of $\mathbf{R}^n$,
- $T : D \to D$: nonexpansive operator.

Then $\text{Fix}T \neq \emptyset$.

Banach-Picard iteration for finding fixed points of nonexpansive operators may fail. If we apply Banach-Picard iteration for finding the fixed point of a nonexpansive operator, it may fail. A simple example is when $T = -I$, which is a nonexpansive operator, a fixed point of which is the 0 vector. However if given Banach-Picard iteration $x_{n+1} = Tx_n$ and start with a nonzero vector $x_0 \neq 0$ as the initial iterate, then the scheme will fluctuate between x_0 and $-x_0$, and will never converge to zero. Now it turns out that, it fails because a property called *asymptotic regularity*[7] property does not hold in this case. When, the asymptotic regularity holds, then Banach-Picard iteration will indeed produce a fixed point of T.

[7] **Asymptotic regularity property:** If $x_n - Tx_n \to 0$, then we say asymptotic regularity property holds.

Banach-Picard iteration for finding fixed points of nonexpansive operators works when asymptotic regularity holds. Suppose,

- D : nonempty, closed, convex, subset of $\mathbf{R}^n$,
- $T : D \to D$: nonexpansive operator, $\text{Fix}T \neq \emptyset$,
- $x_0 \in D$.

Consider the Banach-Picard iteration scheme:

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = Tx_n,$$

where the asymptotic regularity property holds, *i.e.*, $x_n - Tx_n \to 0$. Then, $x_n \to x^*$, where $x^* \in \operatorname{Fix} T$.

Now we state Krasnosel'skii-Mann iteration for nonexpansive operator which always produces a fixed point of T.

Krasnosel'skii-Mann iteration for nonexpansive operators. Suppose,

- D : nonempty, closed, convex, subset of $\mathbf{R}^n$,
- $T : D \to D$: nonexpansive operator, $\operatorname{Fix} T \neq \emptyset$,
- $(\lambda_n)_{n=0}^{\infty}$: nonnegative sequence, $(\forall n \in \mathbf{N}) \quad \lambda_n \in [0,1]$, $\sum_{n=0}^{\infty} \lambda_n (1 - \lambda_n) = \infty$,
- $x_0 \in D$.

Consider the Krasnosel'skii-Mann iteration scheme:

$$(\forall n \in \mathbf{N}) \quad x_{n+1} = \lambda_n T x_n + (1 - \lambda_n) x_n.$$

Then the following hold:

- $(Tx_n - x_n)_{n=0}^{\infty}$ converges to 0,
- $(x_n)_{n=0}^{\infty}$ converges to a fixed point of T.