Chapter 2

New accelareted algorithm and its Application to regression problems

The purpose of this chapter is to investigate a new inertial iterative algorithm for finding the fixed points of a nonexpansive operator in the framework of Hilbert spaces. The first two sections of this chapter address the background and results related to this research work. We propose a novel accelerated iterative algorithm for finding fixed points of a nonexpansive mapping in Section 2.3. Since the presence of inertial terms in an iteration method increase its convergence rate, we use inertial terms to define the algorithm. We investigate the convergence behavior of the proposed algorithm and support it with a numerical example. Further in Section 2.4, we use the proposed algorithm to design a new accelerated proximal gradient algorithm. We also conduct the numerical experiments for regression problem and

This chapter is based on our published research work "**Dixit, A.**, Sahu, D. R., Singh, A. K., and Som, T. (2020). Application of a new accelerated algorithm to regression problems. Soft Computing, 24(2), 1539-1552."

compare the performances of the proposed algorithm with other already existing algorithms on the basis of their objective function values and accuracy.

2.1 Introduction

In 2008, Mainge [63] has combined the inertial type extrapolation algorithm with the classical Mann algorithm and has named it as inertial Mann iteration method for finding the fixed points of a nonexpansive mapping in a real Hilbert space. Inertial Mann algorithm is defined as follows:

$$\begin{cases} y_n = x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} = (1-\lambda_n)y_n + \lambda_n T y_n \quad \text{for all } n \in \mathbb{N}. \end{cases} \tag{2.1}$$

Mainge [63] has studied weak convergence of inertial Mann algorithm (2.1) under the following conditions:

(C_0) $\alpha_n \in [0, \alpha]$ for all $n \in \mathbb{N}$ and $\alpha \in [0, 1)$;

(C_1) $\sum_{n=1}^{\infty} \alpha_n \|x_n - x_{n-1}\|^2 < \infty$;

(C_2) $\inf_{n\geq 1} \lambda_n \geq 0$ and $\sup_{n\geq 1} \lambda_n \leq 1$.

The second condition (C_1) is very strong. It is not easy to verify the condition (C_1) in practical situations. In 2015, Bot and Csetneck [16] have proposed a modification in algorithm (2.1) for finding fixed points of nonexpansive mappings. They have proved the following weak convergence theorem for the sequence generated by the algorithm (2.1) by replacing the condition (C_1) with another condition.

Theorem 2.1.1. *([16], Theorem 5) Let C be nonempty closed affine subset of a real Hilbert space $\mathcal{H}$ and $T : C \to C$ be a nonexpansive mapping such that* $\mathrm{Fix}(T) \neq \emptyset$.

Consider the following iterative method:

$$\begin{cases} y_n = x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} = (1-\lambda_n)y_n + \lambda_n T y_n \quad \textit{for all } n \in \mathbb{N}, \end{cases}$$

where x_0, x_1 *are chosen arbitrarily from* C, *sequence* $\{\alpha_n\} \in [0, \alpha)$ *is nondecreasing with* $\alpha_1 = 0$ *and* $\alpha \in [0, 1)$ *and* $\lambda,\ \delta,\ \sigma\ > 0$ *such that*

$$\delta > \frac{\alpha^2(1+\alpha) + \alpha\sigma}{1-\alpha^2}, \qquad 0 < \lambda \leq \lambda_n \leq \frac{\delta - \alpha[\alpha(1+\alpha) + \alpha\delta + \sigma]}{\delta[1 + \alpha(1+\alpha) + \alpha\delta + \sigma]} \quad \textit{for all } n \in \mathbb{N}.$$

Then the sequence $\{x_n\}$ *converges weakly to a fixed point of* T.

In this chapter, we propose a novel iterative algorithm to solve the fixed point problem of a nonexpansive mapping. Our work is inspired by iterative methods developed by Bot and Csetnek [16] and Sahu [82].

2.2 Preliminary Results

In this section, we present some definitions and basic results useful for the chapter.

Let D be a nonempty subset of a real Hilbert space $\mathcal{H}$ and let $T : D \to \mathcal{H}$ be a mapping. Then T is said to be nonexpansive if

$$\|Tx - Ty\| \leq \|x - y\| \ \text{ for all } \ x, y \in D.$$

Let $f : \mathcal{H} \to (-\infty, \infty]$ be a function. f is said to be proper if $-\infty \notin f(\mathcal{H})$ and *dom* $f \neq \emptyset$. f is said to be lower semicontinuous at $x \in \mathcal{H}$ if, for every sequence $\{x_n\} \subseteq \mathcal{H}$,

$$x_n \to x \to f(x) \leq \underline{\lim} f(x_n).$$

The subdifferential of f is the set-valued operator $\partial f : \mathcal{H} \to 2^{\mathcal{H}}$ defined by

$$\partial f(x) = \{u \in \mathcal{H} : \langle y - x, u\rangle + f(x) \leq f(y) \text{ for all } y \in \mathcal{H}\}.$$

f is subdifferentiable at $x \in \mathcal{H}$ if $\partial f(x) \neq \emptyset$. The elements of $\partial f(x)$ are the subgradients of f at x.

Let $\mathit{\Gamma}_0(\mathcal{H})$ denotes the set of all proper lower semicontinuous convex functions from $\mathcal{H}$ to $(-\infty, +\infty]$.

Definition 2.2.1. ([9]) *Let $f \in \mathit{\Gamma}_0(\mathcal{H})$ and let $x \in \mathcal{H}$. Then $prox_f(x)$ is the unique point in $\mathcal{H}$ that satisfies*

$$f(x) = \min_{y \in \mathcal{H}} \ (f(y) + \frac{1}{2}\|x - y\|^2) = f(prox_f(x)) + \frac{1}{2}\|x - prox_f(x)\|^2.$$

The operator $prox_f : \mathcal{H} \to \mathcal{H}$ is the proximity operator or proximal mapping of f.

Remark 2.2.1. *The proximity operator for ℓ_1-norm is given by*

$$prox_{\lambda\|\cdot\|_1}(x) = (x - \lambda)_+ - (-x - \lambda)_+ = \begin{cases} x_i - \lambda & \text{if} \quad x_i \geq \lambda, \\ 0 & \text{if} \quad |x_i| \leq \lambda, \\ x_i + \lambda & \text{if} \quad x_i \leq -\lambda. \end{cases}$$

Lemma 2.2.1. [9, Corollary 2.14] *Let $z_1, z_2 \in \mathcal{H}$. Then the following identities hold for arbitrary $a \in \mathbb{R}$:*

(i) $\|z_1 - z_2\|^2 = \|z_1\|^2 + \|z_2\|^2 - 2\langle z_1, z_2\rangle$,

(ii) $\|az_1 + (1 - a)z_2\|^2 = a\|z_1\|^2 + (1 - a)\|z_2\|^2 - a(1 - a)\|z_1 - z_2\|^2$.

Lemma 2.2.2. *[37] Let ρ be a positive and α be a nonnegative real numbers. Then, for each $z_1,\ z_2 \in \mathcal{H}$,*

$$\|z_1 \pm \alpha z_2\|^2 \geq (1-\alpha\rho)\|z_1\|^2 + \alpha(\alpha - \frac{1}{\rho})\|z_2\|^2.$$

Lemma 2.2.3. [9, Corollary 4.18] *Let $\mathcal{C}$ be a nonempty closed convex subset of $\mathcal{H}$ and consider a nonexpansive mapping $T : \mathcal{C} \to \mathcal{H}$. Let $\{z_n\}$ be a sequence in $\mathcal{C}$ and $z \in \mathcal{H}$ be such that $z_n \rightharpoonup z$ and $z_n - Tz_n \to 0$ as $n \to \infty$. Then $z \in \mathrm{Fix}(T)$.*

Lemma 2.2.4. [5, Lemma 3] *Consider sequences $\{y_n\}, \{z_n\}$ and $\{\theta_n\}$ in $[0, \infty)$ such that*

$$y_{n+1} \leq y_n + \theta_n(y_n - y_{n-1}) + z_n \ \ \textit{for all} \ \ n \in \mathbb{N}, \quad \sum_{n=1}^{\infty} z_n < \infty$$

and let there exist a real number θ with $0 \leq \theta_n \leq \theta < 1$ for all $n \in \mathbb{N}$. Then the following hold:

(i) $\sum_{n=1}^{\infty}[y_n - y_{n-1}]_+ < \infty$, *where* $[t]_+ = \max\{t, 0\}$,

(ii) *there exists $y^* \in [0, \infty)$ such that $y_n \to y^*$.*

Lemma 2.2.5. *[74] Consider a nonempty subset $\mathcal{C}$ of $\mathcal{H}$. Let $\{\phi_n\}$ be a sequence in $\mathcal{H}$ such that the following two conditions hold:*

(i) *for all $\phi \in \mathcal{C}$, $\lim_{n\to\infty} \|\phi_n - \phi\|$ exists,*

(ii) *every sequential weak cluster point of $\{\phi_n\}$ is in $\mathcal{C}$.*

Then the sequence $\{\phi_n\}$ converges weakly to a point in $\mathcal{C}$.

2.3 Accelerated normal S-iteration method and its convergence analysis

In this section, we introduce a new accelerated fixed point iteration method and study the weak convergence analysis for finding fixed points of a nonexpansive mapping in the framework of real Hilbert spaces.

First, we introduce our accelerated iterative algorithm.

Algorithm 2.3.1. *Let C be a nonempty closed affine subset of a real Hilbert space $\mathcal{H}$ and $T : C \to C$ be a nonexpansive mapping with* $\mathrm{Fix}(T) \neq \emptyset$.
(1) *Initialization: Select $x_0, x_1 \in C$ arbitrarily.*
(2) *Iterative step: Select $\{\alpha_n\}$ and $\{\beta_n\}$ as iteration parameters in $[0, 1)$ and compute the $(n+1)^{th}$ iteration as follows:*

$$\begin{cases} y_n = x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} = T[(1-\beta_n)y_n + \beta_n T(y_n)] \quad \text{for all } n \in \mathbb{N}. \end{cases} \tag{2.2}$$

If $\alpha_n = 0$, then iteration method (2.2) reduces to the normal S-iteration method defined in (1.21). Thus, the iteration method (2.2) is an inertial form of normal S-iteration method. We call it inertial normal S-iteration method.

We assume that $\{\alpha_n\}$ and $\{\beta_n\}$ satisfy the following conditions:

(A1) $\{\alpha_n\} \subset [0, \alpha]$ is non-decreasing sequence with $\alpha \in [0, 1)$;

(A2) there exist constants $\beta,\ \sigma,\ \delta > 0$ satisfying

$$\delta > \frac{2q\alpha(\alpha(1+\alpha)+\sigma)}{1-\alpha^2(1-\beta)},\ 0 < \beta \leq \beta_n \leq \frac{\delta - \alpha(2q\alpha(1+\alpha)+\alpha\delta(1-\beta)+2q\sigma)}{\delta[1+2q\alpha(1+\alpha)+\alpha\delta(1-\beta)+2q\sigma]} \quad (2.3)$$

where $q = 1 + \frac{1}{\beta^2}$;

(A3) Define sequences $\{\xi_n\}$ and $\{\mu_n\}$ by

$$\mu_n = \alpha_n(1+\alpha_n) + \frac{\alpha_n(1-\beta_n)(1-\alpha_n\rho_n)}{2q\beta_n\rho_n}, \qquad \xi_n = -\frac{(1-\beta_n)(1-\alpha_n\rho_n)}{2q\beta_n}, \quad (2.4)$$

where $\rho_n = \frac{1}{\alpha_n + \delta\beta_n}$.

Before presenting our main convergence theorem, we need the following:

Proposition 2.3.1. *Let C be a nonempty closed affine subset of a real Hilbert space $\mathcal{H}$ and $T : C \to C$ be a nonexpansive mapping with $\mathrm{Fix}(T) \neq \emptyset$. Let $\{x_n\}$ be a sequence generated by Algorithm 2.3.1. Then*

$$\begin{aligned} \|x_{n+1} - p\|^2 \quad &\leq \quad (1+\alpha_n)\|x_n - p\|^2 - \alpha_n\|x_{n-1} - p\|^2 + \alpha_n(1+\alpha_n)\|x_n - x_{n-1}\|^2 \\ &\quad -\beta_n(1-\beta_n)\|y_n - T(y_n)\|^2 \ \ \textit{for all} \ \ p \in \mathrm{Fix}(T). \end{aligned}$$

Proof. From Algorithm 2.3.1 and Lemma 2.2.1, we have

$$\begin{aligned} \|x_{n+1} - p\|^2 &= \|T[(1-\beta_n)y_n + \beta_n T(y_n)] - p\|^2 \\ &\leq \|(1-\beta_n)y_n + \beta_n T(y_n) - p\|^2 \\ &= (1-\beta_n)\|y_n - p\|^2 + \beta_n\|T(y_n) - p\|^2 - \beta_n(1-\beta_n)\|y_n - T(y_n)\|^2 \\ &\leq (1-\beta_n)\|y_n - p\|^2 + \beta_n\|y_n - p\|^2 - \beta_n(1-\beta_n)\|y_n - T(y_n)\|^2 \\ &= \|y_n - p\|^2 - \beta_n(1-\beta_n)\|y_n - T(y_n)\|^2. \end{aligned} \quad (2.5)$$

Again, from (2.2) and Lemma 2.2.1, we have

$$\begin{aligned}\|y_n - p\|^2 &= \|x_n + \alpha_n(x_n - x_{n-1}) - p\|^2 \\ &= \|(1+\alpha_n)(x_n - p) - \alpha_n(x_{n-1} - p)\|^2 \\ &= (1+\alpha_n)\|x_n - p\|^2 - \alpha_n\|x_{n-1} - p\|^2 + \alpha_n(1+\alpha_n)\|x_n - x_{n-1}\|^2.\end{aligned} \tag{2.6}$$

Combining (2.5) and (2.6), we get

$$\begin{aligned}\|x_{n+1} - p\|^2 \leq\ & (1+\alpha_n)\|x_n - p\|^2 - \alpha_n\|x_{n-1} - p\|^2 + \alpha_n(1+\alpha_n)\|x_n - x_{n-1}\|^2 \\ & -\beta_n(1-\beta_n)\|y_n - T(y_n)\|^2.\end{aligned} \tag{2.7}$$

□

Proposition 2.3.2. *Let C be a nonempty closed affine subset of a real Hibert space $\mathcal{H}$ and $T : C \to C$ be a nonexpansive mapping with $\mathrm{Fix}(T) \neq \emptyset$. Let $\{\alpha_n\}$ and $\{\beta_n\}$ be sequences in $[0, 1)$ satisfying the conditions (A1) and (A2). Then the sequence $\{x_n\}$ generated by the Algorithm (2.2) satisfies the following inequality:*

$$\phi_{n+1} - (1+\alpha_n)\phi_n + \alpha_n\phi_{n-1} \leq \xi_n\|x_{n+1} - x_n\|^2 + \mu_n\|x_n - x_{n-1}\|^2, \tag{2.8}$$

where $\phi_n = \|x_n - p\|^2$.

Proof. Set $z_n = (1-\beta_n)y_n + \beta_n T(y_n)$. Then the Algorithm (2.2) can be written as:

$$\begin{cases} x_{n+1} = T(z_n), \\ z_n = (1-\beta_n)y_n + \beta_n T(y_n), \\ y_n = x_n + \alpha_n(x_n - x_{n-1}). \end{cases} \tag{2.9}$$

From (2.9), we have

$$\begin{aligned}\|y_n - T(y_n)\|^2 &= \frac{1}{\beta_n^2}\|z_n - y_n\|^2 \\ &\geq \frac{1}{\beta_n^2}\|T(z_n) - T(y_n)\|^2 \\ &= \frac{1}{\beta_n^2}\|x_{n+1} - T(y_n)\|^2 \\ &= \frac{1}{\beta_n^2}\|x_{n+1} - y_n + y_n - T(y_n)\|^2.\end{aligned}$$

Taking $\rho = \frac{1}{2}$ and using Lemma 2.2.2, we obtain

$$\|y_n - T(y_n)\|^2 \geq \frac{1}{\beta_n^2}\left\{\frac{1}{2}\|x_{n+1} - y_n\|^2 - \|y_n - T(y_n)\|^2\right\},$$

which implies that

$$(1 + \frac{1}{\beta_n^2})\|y_n - T(y_n)\|^2 \geq \frac{1}{2\beta_n^2}\|x_{n+1} - y_n\|^2.$$

Note $0 < \beta \leq \beta_n$ for all $n \in \mathbb{N}$, hence

$$\begin{aligned}(1 + \frac{1}{\beta^2})\|y_n - T(y_n)\|^2 &\geq \frac{1}{2\beta_n^2}\|x_{n+1} - y_n\|^2 \\ &= \frac{1}{2\beta_n^2}\|x_{n+1} - x_n - \alpha_n(x_n - x_{n-1})\|^2.\end{aligned}$$

Again using Lemma 2.2.2, we obtain

$$\begin{aligned}(1 + \frac{1}{\beta^2})\|y_n - T(y_n)\|^2 &\geq \frac{(1 - \alpha_n\rho_n)}{2{\beta_n}^2}\|x_{n+1} - x_n\|^2 + \frac{\alpha_n}{2\beta_n^2}(\alpha_n - \frac{1}{\rho_n})\|x_n - x_{n-1}\|^2 \\ &= \frac{(1 - \alpha_n\rho_n)}{2\beta_n^2}\|x_{n+1} - x_n\|^2 - \frac{\alpha_n(1 - \alpha_n\rho_n)}{2\beta_n^2\rho_n}\|x_n - x_{n-1}\|^2.\end{aligned} \tag{2.10}$$

Multiplying (2.10) by $-\beta_n(1-\beta_n)$, we obtain

$$\begin{aligned} -q\beta_n(1-\beta_n)\|y_n - T(y_n)\|^2 &\leq -\frac{(1-\beta_n)(1-\alpha_n\rho_n)}{2\beta_n}\|x_{n+1}-x_n\|^2 \\ &+\frac{\alpha_n(1-\beta_n)(1-\alpha_n\rho_n)}{2\beta_n\rho_n}\|x_n - x_{n-1}\|^2. \quad (2.11) \end{aligned}$$

From Proposition 2.3.1 and equation (2.11), we get

$$\begin{aligned} \|x_{n+1}-p\|^2 &\leq (1+\alpha_n)\|x_n-p\|^2 - \alpha_n\|x_{n-1}-p\|^2 + \alpha_n(1+\alpha_n)\|x_n-x_{n-1}\|^2 \\ &-\frac{(1-\beta_n)(1-\alpha_n\rho_n)}{2q\beta_n}\|x_{n+1}-x_n\|^2 + \frac{\alpha_n(1-\beta_n)(1-\alpha_n\rho_n)}{2q\beta_n\rho_n}\|x_n-x_{n-1}\|^2. \end{aligned}$$

Hence

$$\begin{aligned} \phi_{n+1} &\leq (1+\alpha_n)\phi_n - \alpha_n\phi_{n-1} + \alpha_n(1+\alpha_n)\|x_n-x_{n-1}\|^2 \\ &-\frac{(1-\beta_n)(1-\alpha_n\rho_n)}{2q\beta_n}\|x_{n+1}-x_n\|^2 + \frac{\alpha_n(1-\beta_n)(1-\alpha_n\rho_n)}{2q\beta_n\rho_n}\|x_n-x_{n-1}\|^2, \end{aligned}$$

which can be written as

$$\phi_{n+1} - (1+\alpha_n)\phi_n + \alpha_n\phi_{n-1} \leq \xi_n\|x_{n+1}-x_n\|^2 + \mu_n\|x_n-x_{n-1}\|^2.$$

□

Proposition 2.3.3. *Suppose $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences in $[0,1)$ satisfying the condition (A1) and (A2). Let $\{\xi_n\}$ and $\{\mu_n\}$ be sequences defined by (2.4). Then $\xi_n + \mu_{n+1} \leq -\sigma$ for all $n \in \mathbb{N}$.*

Proof. Note that $\mu_n = \alpha_n(1+\alpha_n) + \frac{\alpha_n(1-\beta_n)(1-\alpha_n\rho_n)}{2q\beta_n\rho_n} > 0$, since $\alpha_n\rho_n < 1$ and $\beta_n \in (0,1)$. Again, taking into account of choice of ρ_n, we have

$$\delta = \frac{1-\alpha_n\rho_n}{\rho_n\beta_n}.$$

Note

$$\mu_n = \alpha_n(1+\alpha_n) + \frac{\alpha_n(1-\beta_n)\delta}{2q} \leq \alpha(1+\alpha) + \frac{\alpha\delta(1-\beta)}{2q} \quad \text{for all } n \in \mathbb{N}. \tag{2.12}$$

For $n \in \mathbb{N}$, we have

$$\begin{aligned}
\xi_n + \mu_{n+1} \leq -\sigma \quad &\Leftrightarrow \quad \frac{(1-\beta_n)(\alpha_n\rho_n - 1)}{2q\beta_n} + (\mu_{n+1}+\sigma) \leq 0 \\
&\Leftrightarrow \quad (1-\beta_n)(\alpha_n\rho_n - 1) + 2q\beta_n(\mu_{n+1}+\sigma) \leq 0 \\
&\Leftrightarrow \quad -(1-\beta_n)\delta\rho_n\beta_n + 2q\beta_n(\mu_{n+1}+\sigma) \leq 0 \\
&\Leftrightarrow \quad -\frac{(1-\beta_n)\delta}{\alpha_n + \delta\beta_n} + 2q(\mu_{n+1}+\sigma) \leq 0 \\
&\Leftrightarrow \quad -(1-\beta_n)\delta + 2q(\mu_{n+1}+\sigma)(\alpha_n+\delta\beta_n) \leq 0 \\
&\Leftrightarrow \quad 2q(\mu_{n+1}+\sigma)(\alpha_n+\delta\beta_n) + \beta_n\delta \leq \delta.
\end{aligned}$$

By using (2.12), we obtain

$$2q(\mu_{n+1}+\sigma)(\alpha_n+\delta\beta_n) + \beta_n\delta \leq 2q(\alpha(1+\alpha) + \frac{\alpha\delta(1-\beta)}{2q} + \sigma)(\alpha+\delta\beta_n) + \beta_n\delta \leq \delta,$$

where the last inequality follows from the upper bound for $\{\beta_n\}$ in (2.3). Hence

$$\xi_n + \mu_{n+1} \leq -\sigma \text{ for all } n \in \mathbb{N}.$$

□

Now, we are ready to establish weak convergence of inertial normal S-iteration method defined by (2.2) for computation of fixed points of a nonexpansive mapping.

Theorem 2.3.1. *Let C be a nonempty closed affine subset of a real Hilbert space $\mathcal{H}$ and $T : C \to C$ be a nonexpansive mapping with* $\mathrm{Fix}(T) \neq \emptyset$*. Let $\{\alpha_n\}$ with $\alpha_1 = 0$ and $\{\beta_n\}$ be sequences in $[0, 1)$ satisfying the conditions (A1) and (A2). Then the sequence $\{x_n\}$ generated by Algorithm 2.2 converges weakly to a fixed point of T.*

Proof. Let $p \in \mathrm{Fix}(T)$ and set $\psi_n = \phi_n - \alpha_n\phi_{n-1} + \mu_n\|x_n - x_{n-1}\|^2$. We proceed with the following steps:

Step 1. First we show that $\{\psi_n\}$ is a nonincreasing sequence. Note

$$\begin{aligned}\psi_{n+1} - \psi_n &= \phi_{n+1} - \alpha_{n+1}\phi_n + \mu_{n+1}\|x_{n+1} - x_n\|^2 - \phi_n + \alpha_n\phi_{n-1} - \mu_n\|x_n - x_{n-1}\|^2 \\ &= \phi_{n+1} - (1 + \alpha_{n+1})\phi_n + \alpha_n\phi_{n-1} + \mu_{n+1}\|x_{n+1} - x_n\|^2 - \mu_n\|x_n - x_{n-1}\|^2.\end{aligned} \tag{2.13}$$

Using Proposition 2.3.2, we get

$$\begin{aligned}\psi_{n+1} - \psi_n &\leq \xi_n\|x_{n+1} - x_n\|^2 + \mu_{n+1}\|x_{n+1} - x_n\|^2 \\ &= (\xi_n + \mu_{n+1})\|x_{n+1} - x_n\|^2.\end{aligned} \tag{2.14}$$

Using Proposition 2.3.3, we have

$$\psi_{n+1} - \psi_n \leq -\sigma\|x_{n+1} - x_n\|^2 \text{ for all } n \in \mathbb{N}, \tag{2.15}$$

$\{\psi_n\}$ is nonincreasing sequence.

Step 2. Now, we show that $\sum_{n=1}^{\infty} \|x_{n+1} - x_n\|^2 < \infty$.

Since $\{\psi_n\}$ is nonincreasing and $\{\alpha_n\}$ is bounded, we get

$$-\alpha\phi_{n-1} \leq \phi_n - \alpha\phi_{n-1} \leq \psi_n \leq \psi_1.$$

Thus, we obtain

$$\begin{aligned}
\phi_n &\leq \alpha\phi_{n-1} + \psi_1, \\
&\leq \alpha(\alpha\phi_{n-2} + \psi_1) + \psi_1, \\
&\vdots \\
&\leq \alpha^n\phi_0 + \psi_1 \sum_{k=0}^{n-1} \alpha^k \leq \alpha^n\phi_0 + \frac{\psi_1}{1-\alpha}.
\end{aligned}$$

From (2.15), we conclude that

$$\begin{aligned}
\sigma \sum_{k=1}^{n} \|x_{k+1} - x_k\|^2 &\leq \psi_1 - \psi_{n+1} \\
&\leq \psi_1 + \alpha\phi_n \\
&\leq \psi_1 + \alpha(\alpha^n\phi_0 + \frac{\psi_1}{1-\alpha}) \\
&= \alpha^{n+1}\phi_0 + \frac{\psi_1}{1-\alpha}. \qquad (2.16)
\end{aligned}$$

Since $\alpha^{n+1} \to 0$ as $n \to \infty$, we obtain that

$$\sum_{n=1}^{\infty} \|x_{n+1} - x_n\|^2 < \infty. \qquad (2.17)$$

Step 3. Next, we show that $\lim_{n\to\infty} ||x_n - p||$ exists.

From (2.8), (2.12), (2.17) and Lemma 2.2.4, we obtain $\lim_{n\to\infty} ||x_n - p||$ exists.

Step 4. Finally to show that every sequential weak cluster point of sequence $\{x_n\}$ is in $\mathrm{Fix}(T)$.

From (2.17), we have

$$\lim_{n\to\infty} \|x_n - x_{n-1}\| = 0. \tag{2.18}$$

From Algorithm 2.3.1, we have

$$\begin{aligned}
\|y_n - x_{n+1}\| &\leq \|x_n - x_{n+1}\| + \alpha_n \|x_n - x_{n-1}\| \\
&\leq \|x_n - x_{n+1}\| + \alpha \|x_n - x_{n-1}\|,
\end{aligned}$$

using (2.18), we get that $\lim_{n\to\infty} \|y_n - x_{n+1}\| = 0$. From (2.2), we have

$$\begin{aligned}
\|Ty_n - y_n\| &= \|Ty_n - x_{n+1} + x_{n+1} - y_n\| \\
&\leq \|Ty_n - x_{n+1}\| + \|x_{n+1} - y_n\| \\
&= \|Ty_n - Tz_n\| + \|x_{n+1} - y_n\| \\
&\leq \|y_n - z_n\| + \|x_{n+1} - y_n\| \\
&= \|y_n - (1-\beta_n)y_n - \beta_n Ty_n\| + \|x_{n+1} - y_n\| \\
&= \beta_n \|y_n - Ty_n\| + \|x_{n+1} - y_n\|,
\end{aligned}$$

which implies that

$$(1-\beta_n)\|Ty_n - y_n\| \;\leq\; \|x_{n+1} - y_n\|. \tag{2.19}$$

From (2.19), we obtain

$$\|Ty_n - y_n\| \to 0 \;\text{ as }\; n \to \infty. \tag{2.20}$$

Let x be an arbitrary weak cluster point of sequence $\{x_n\}$. Then there exists a subsequence $\{x_{n_k}\}$ of sequence $\{x_n\}$ such that $x_{n_k} \rightharpoonup x \in C$. Using (2.20) and Lemma 2.2.3, we conclude that $x \in \text{Fix}(T)$. It follows from Lemma 2.2.5 that $\{x_n\}$ converges weakly to a point in $\text{Fix}(T)$.

□

Remark 2.3.1. *In order to proof Theorem 2.3.1, we need the nonnegativity of* μ*, which is accomplished by the condition* $\alpha_1 = 0$*. This condition can be removed by taking* $x_0 = x_1$ *in Algorithm 2.3.1.*

2.4 Numerical Example

In this section, we have to demonstrate behavior of the inertial normal S-iteration method (2.2), Mann iteration method (1.19), normal S-iteration method (1.21) and inertial Mann iteration method (2.1) by the following example.

Example 2.4.1. *Let* $\mathcal{H} = \mathbb{R}^2$ *with usual norm and* $T : \mathbb{R}^2 \to \mathbb{R}^2$ *be a mapping defined by*

$$T(u,v) = \left(sin\frac{u+v}{2}, sin\frac{u-v}{2}\right) \;\text{ for all }\; (u,v) \in \mathbb{R}^2.$$

Thus, finding the fixed point of T implies a solution to the following system:

$$\begin{cases} u = sin\frac{u+v}{2}, \\ v = sin\frac{u-v}{2}. \end{cases}$$

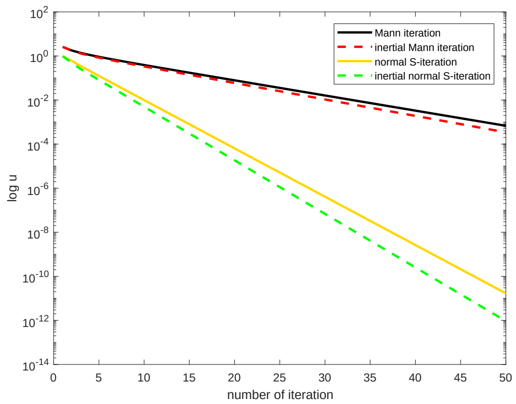

FIGURE 2.1: log u vs number of iteration.

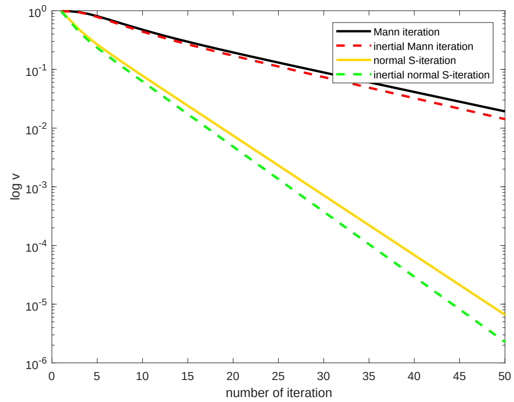

FIGURE 2.2: log v vs number of iteration .

FIGURE 2.3: Coordinatewise graph for different iteration methods.

First, we show that T is a nonexpansive mapping. Let $x = (u_1, v_1),\ y = (u_2, v_2) \in \mathbb{R}^2$, then

$$\begin{aligned} \|T(x) - T(y)\|^2 &= \left\| \left(sin\frac{u_1+v_1}{2}, sin\frac{u_1-v_1}{2} \right) - \left(sin\frac{u_2+v_2}{2}, sin\frac{u_2-y_2}{2} \right) \right\|^2 \\ &= \left| sin\frac{u_1+v_1}{2} - sin\frac{u_2+v_2}{2} \right|^2 + \left| sin\frac{u_1-v_1}{2} - sin\frac{u_2-v_2}{2} \right|^2 \\ &\leq \left| \frac{u_1+v_1}{2} - \frac{u_2+v_2}{2} \right|^2 + \left| \frac{u_1-v_1}{2} - \frac{u_2-v_2}{2} \right|^2 \\ &\leq \left| \frac{u_1-u_2}{2} + \frac{v_1-v_2}{2} \right|^2 + \left| \frac{u_1-u_2}{2} + \frac{v_2-v_1}{2} \right|^2. \end{aligned} \tag{2.21}$$

Note $(a+b)^2 \leq 2(a^2+b^2)$ for any $a, b \in \mathbb{R}$. Hence, from (2.21), we get

$$\begin{aligned} \|T(x)-T(y)\|^2 &\leq 2\left|\frac{u_1-u_2}{2}\right|^2 + 2\left|\frac{v_1-v_2}{2}\right|^2 + 2\left|\frac{u_1-u_2}{2}\right|^2 + 2\left|\frac{v_1-v_2}{2}\right|^2 \\ &= |u_1-u_2|^2 + |v_1-v_2|^2 \\ &= \|x-y\|^2. \end{aligned} \tag{2.22}$$

From (2.22), we get that T is a nonexpansive map. It is easy to see that $(0,0)$ is a fixed point of T. Hence Theorem 2.1.1 and Theorem 2.3.1 can be applied for computation of fixed point of T.

For numerical results, the initial values are taken as $(x_0, y_0) = (5, 1)$ and we have taken iteration parameter $\beta_n = 0.50$, inertial parameter $\alpha_n = \frac{1}{20}$ for Mann iteration method (1.19), inertial Mann iteration method (2.1), normal-S iteration method (1.21) and inertial normal S-iteration method (2.2). A graph is plotted between number of iteration versus the first coordinate in Figure 2.1 and versus second coordinate in Figure 2.2. From Figure 2.3, we observe the following:

- Convergence rate of inertial normal S-iteration method is faster than Mann iteration method (1.19), inertial Mann iteration method (2.1) and normal S-iteration method (1.21).
- Normal-S iteration method (1.21) converges rapidly than inertial Mann iteration method (2.1).
- Convergence rate of inertial Mann iteration method (2.1) is greater than Mann iteration method (1.19).

2.5 Numerical Experiment with Data Sets

In this section, we have conducted numerical experiments to compare the proposed algorithm with already existing algorithms on the basis of their efficiency. We used the Intel(R) Core(TM)i5-7200U CPU @2.50GHZ, 2.70GHZ processor with 8.0 GB RAM and 64-bit operating system in Matlab R2017a circumstance.

Let us consider the minimization problem with objective function as a sum of two convex function

$$\min_{x\in\mathbb{R}^d} \; F(x) = f(x) + g(x), \tag{2.23}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a smooth function but $g : \mathbb{R}^d \to \mathbb{R}$ is a subdifferentiable function. A point x^* is a solution of minimization problem (2.23) if and only if

$$0 \in \nabla f(x^*) + \partial g(x^*),$$

where ∇f and ∂g are gradient and subgradient of f and g respectively. From [76], for any $\lambda > 0$, we have

$$\begin{aligned} 0 \in \lambda\nabla f(x^*) + \lambda\partial g(x^*) \;&\Leftrightarrow\; 0 \in \lambda\nabla f(x^*) - x^* + x^* + \lambda\partial g(x^*) \\ &\Leftrightarrow\; (Id - \lambda\nabla f)(x^*) \in (Id + \lambda\partial g)(x^*) \\ &\Leftrightarrow\; x^* = (Id + \lambda\partial g)^{-1}(Id - \lambda\nabla f)(x^*) \\ &\Leftrightarrow\; x^* = prox_{\lambda g}(x^* - \lambda\nabla f)(x^*). \end{aligned}$$

Thus, x^* minimizes $(f + g)$ if and only if x^* is a fixed point of $prox_{\lambda g}(Id - \lambda\nabla f)$. Note that $prox_{\lambda g}$ is nonexpansive mapping for proper lower semicontinuous convex functions g [9].

Consider the Lasso problem

$$\min_{x\in\mathbb{R}^d} \; F(x) = \frac{1}{2m}\|Ax-b\|^2 + \rho\|x\|_1, \tag{2.24}$$

with $A \equiv [A_1, A_2, \cdots, A_i, \cdots, A_d] \in \mathbb{R}^{m\times d}$ data matrix having d-features and m-samples, each A_i is an m-dimensional vector $i = 1, 2, 3, \cdots, d$, b is the vector containing m responses and ρ is the sparsity controlling parameter. Note that x^* is the solution of minimization problem (2.24) if and only if it is the fixed point of the operator T, where $T : \mathbb{R}^d \to \mathbb{R}^d$ is nonexpansive operator defined by

$$T(x) = prox_{\rho\lambda\|.\|_1}(x - \lambda\partial(\frac{1}{2m}\{\|Ax-b\|_2\}^2)) \;\text{ for some }\; \lambda > 0,\; x \in \mathbb{R}^d. \tag{2.25}$$

Proximal gradient algorithms based on Mann iteration method (1.19), inertial Mann iteration method (2.1), normal S-iteration method (1.21) and Algorithm 2.3.1 for finding solution of minimization problem (2.24) are the following:

1. Mann proximal gradient (MPG) algorithm:

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n prox_{\lambda\rho\|.\|_1}(x_n - \frac{\lambda}{m}A^t(Ax_n - b)) \quad \text{for all } n \in \mathbb{N}, \tag{2.26}$$

2. Inertial Mann proximal gradient (IMPG) algorithm:

$$\begin{cases} y_n = x_n + \alpha_n(x_n - x_{n-1}), \\ x_{n+1} = (1-\beta_n)y_n + \beta_n prox_{\lambda\rho\|.\|_1}(y_n - \frac{\lambda}{m}A^t(Ay_n - b))) \quad \text{for all } n \in \mathbb{N}, \end{cases} \tag{2.27}$$

3. Normal S proximal gradient (NSPG) algorithm:

$$\begin{cases} v_n = (1-\beta_n)x_n + \beta_n prox_{\lambda\rho\|.\|_1}(x_n - \frac{\lambda}{m}A^t(Ax_n - b)), \\ x_{n+1} = prox_{\lambda\rho\|.\|_1}(v_n - \frac{\lambda}{m}A^t(Av_n - b)) \quad \text{for all } n \in \mathbb{N}, \end{cases} \tag{2.28}$$

4. Inertial normal S proximal gradient (INSPG) algorithm:

$$\begin{cases} y_n = x_n + \alpha_n(x_n - x_{n-1}), \\ u_n = prox_{\lambda\rho\|.\|_1}(y_n - \frac{\lambda}{m}A^t(Ay_n - b)), \\ z_n = (1-\beta_n)y_n + \beta_n u_n, \\ x_{n+1} = prox_{\lambda\rho\|.\|_1}(z_n - \frac{\lambda}{m}A^t(Az_n - b)) \quad \text{for all } n \in \mathbb{N}. \end{cases} \tag{2.29}$$

DATA SETS:

We have conducted our experiments on publicly available datasets [1]. We provide the description of the data sets as follows:

(i) **Colon-cancer dataset:** Colon cancer is the cancer of the large intestine (colon), which occurs due to the existence of anomalous cells in the last part of the digestive system i.e. large intestine or colon. This dataset is collected from 62 patients having 2000 gene expressions with the highest minimal intensity in decreasing order. It contains 40 tumor biopsies from tumors and 22 normal biopsies from healthy parts of the large intestine of a patient.

(ii) **Allaml dataset:** A molecular classification of cancer disease can be done on the basis of gene expressions. Gene expressions can be monitored by DNA chips and can be applied to human acute leukemias for test purposes. Using an automatically derived class predictor, we are able to classify into acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). This dataset collected has gene expressions of 7129 genes from 72 samples.

(iii) **Carcinom dataset:** Carcinoma is a cancer type, which occurs due to the damaged DNA of a cell and cell starts to grow uncontrollably. It begins in

[1] http://featureselection.asu.edu/datasets.php

The above information about datasets can be summerized as follows:

serial no.	dataset	samples	features	classes
1	colon	62	2000	2
2	Allaml	72	7129	2
3	Carcinom	174	9182	11
4	Lymphoma	96	4026	9
5	Nci9	60	9712	9
6	Lung discrete	73	325	7

Table 2.1: Information about datasets

a tissue that lines the inner or outer surfaces of the body. This dataset is collected from 174 samples having 9182 features.

(iv) **Lymphoma dataset:** Lymphoma is a broad term encompassing a variety of cancers of the lymphatic system. Total number of genes to be tested is 4026 and the number of samples to be tested is 96. There are 9 classes of Lymphoma.

(v) **Nci9 dataset:** A gene expression dataset with 9712 genes, 60 samples and 9 classes.

(vi) **Lung discrete dataset:** This dataset is of 73 instances for 325 features having 7 classes.

We use the proposed algorithm INSPG (2.29) to solve minimization problem (2.24) and compare it with already existing algorithms MPG (2.26), IMPG (2.27) and NSPG (2.28). To normalize the datasets, we applied z-score as a pre-processing and a column vector containing all entries 1 is added to the data-matrix A. We choose sequences $\{\alpha_n\} = \{\frac{n-1}{14n+a}\}$ with $a = 2.5$, similar to inertial term in [29] and $\{\beta_n\} = \{0.5 + \frac{1}{200n}\}$.

Clearly sequence $\{\alpha_n\}$ and $\{\beta_n\}$ satisfying the conditions (A1) and (A2) for the convergence of the Algorithm 2.3.1. The sparsity controlling parameter ρ is taken as $\theta \times \|A^T b\|_\infty$. We have tuned the parameter θ in the range [1,10e-10] in the multiple of 0.1. The maximum number of iteration is set to 1000 and to stop the procedure, we set the difference between consecutive iterations to be less than 0.001.

In the first experiment, we compare the proximal gradient algorithms on the basis of their convergence speed. We calculate the objective function value $F(x)$ defined in the minimization problems (2.24) at each iteration for all the datasets using different proximal gradient iterative methods. Figure 2.10 represents the graph between the number of iterations and corresponding function values. From Figure 2.10 we observe the following:

- For the initial iteration values, function value corresponding to INSPG algorithm (2.29) and NSPG algorithm (2.28) are nearly the same but as number of iteration increases difference between their function value increases.
- We can observe a similar pattern between MPG (2.26) and IMPG algorithm (2.27).
- For each data set, INSPG algorithm (2.29) has faster convergence rate than all other algorithms.
- In comparison to MPG algorithm (2.26) and IMPG algorithm (2.27), the performance of INSPG algorithm (2.29) is outstanding. For each dataset, the difference in their objective function value with INSPG algorithm (2.29) is significantly large. In the case of carcinom, lymphoma, Nci9 and Lung discrete dataset statement are more prominent.
- The objective function value corresponding to INSPG algorithm (2.29) is always less than NSPG algorithm (2.28). For carcinom, lymphoma, Nci9 and

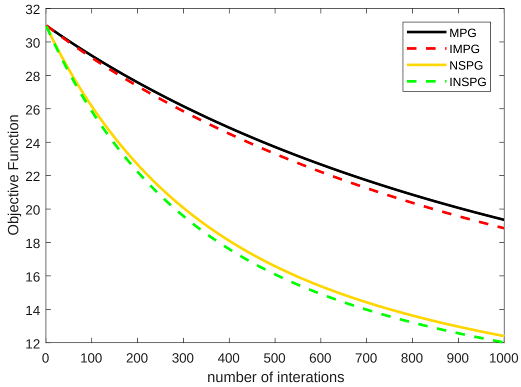

Figure 2.4: Colon.

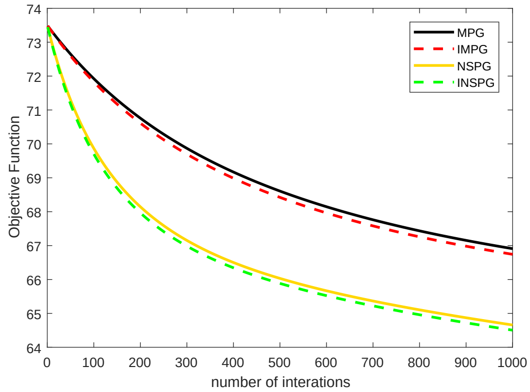

Figure 2.5: Allaml

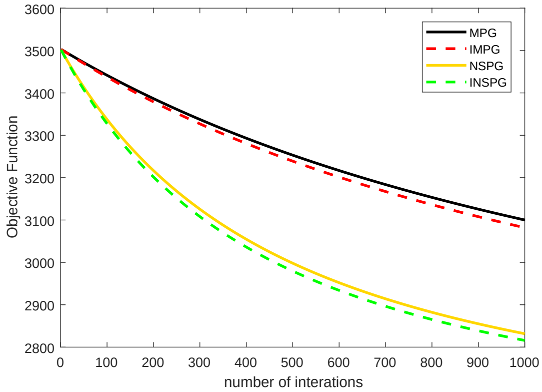

Figure 2.6: Carcinom.

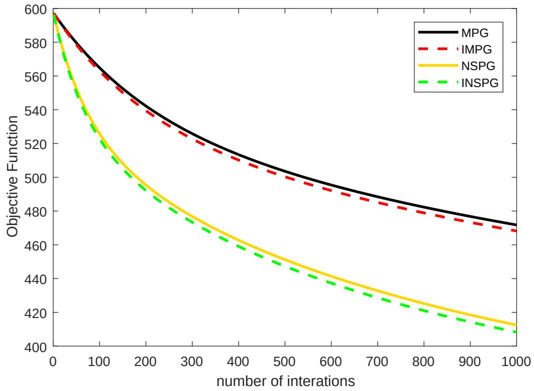

Figure 2.7: Lymphoma.

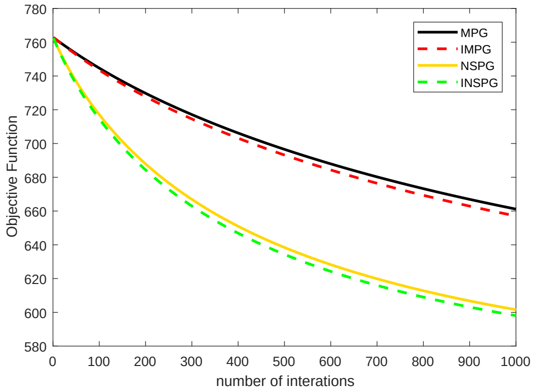

Figure 2.8: Nci9.

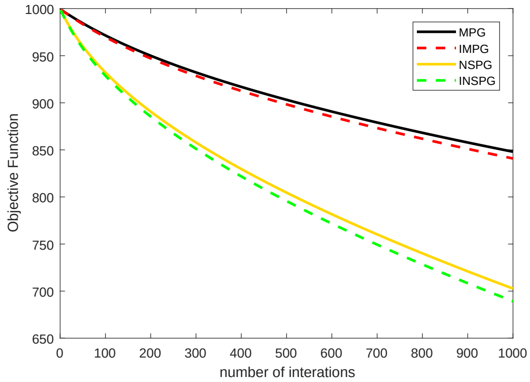

Figure 2.9: Lung discrete.

Figure 2.10: The graph is plotted between number of iteration vs corresponding objective Function value for different datasets.

lung discrete datasets the difference in their objective value is significantly large, which is 15.8 for cacinom, 13.4546 for lung discrete and 4.1947 for lymphoma datasets. It marks the applicability of INSPG algorithm (2.29) over NSPG algorithm (2.28).

- It is interesting to note that NSPG algorithm (2.28) has greater convergence rate than IMPG algorithm (2.27) in each data set.
- Convergence rate of IMPG algorithm (2.27) is better than MPG algorithm (2.26).

In experiment 2, we compare the algorithms on the basis of their regression accuracy. To compare regression accuracy, we consider the standard root mean square error (RMSE) using different iteration methods for all six datasets. Figure 2.17 represents the graph between the number of iteration and corresponding RMSE. From Figure 2.17, we can observe the following:

- For each dataset, RMSE value of INSPG algorithm (2.29) is less than those of all other algorithms which show INSPG algorithm (2.29) has better accuracy than MPG algorithm (2.26), IMPG algorithm (2.27) and NSPG algorithm (2.28).
- For each dataset, as the number of iteration increases, difference in the RMSE value of INSPG algorithm (2.29) with MPG algorithm (2.26), IMPG algorithm (2.27) and NSPG algorithm (2.28) increases. For MPG algorithm (2.26) and IMPG algorithm (2.27) this difference is large.
- INSPG algorithm (2.29) has better accuracy than NSPG algorithm (2.28) for each dataset which is more dominant in case of colon datasets and lung discrete datasets. For these datasets, difference in their accuracy is up to second decimal digit.

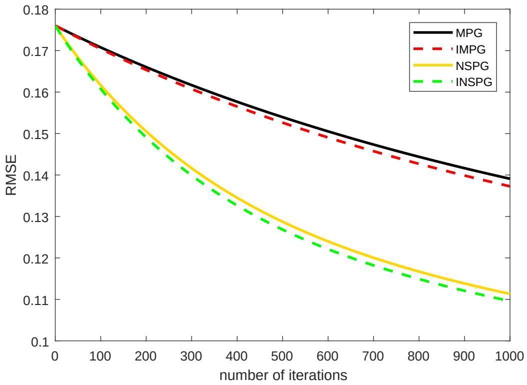

FIGURE 2.11: Colon.

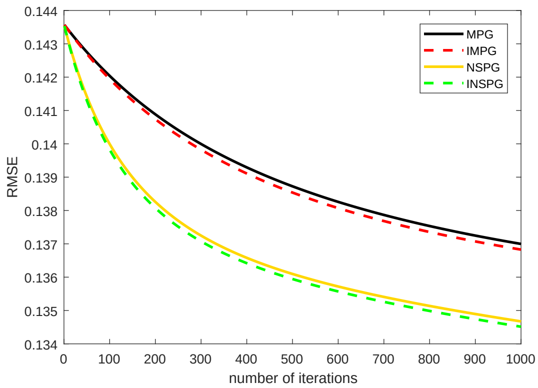

FIGURE 2.12: Allaml.

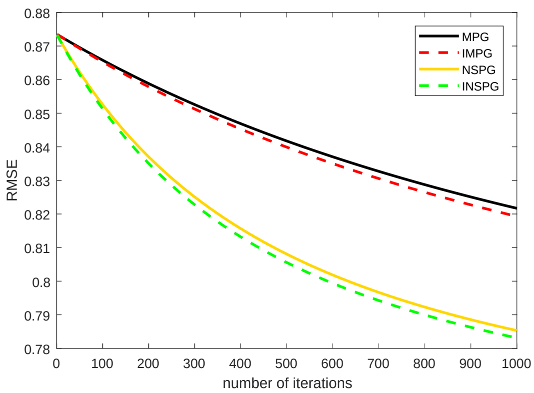

FIGURE 2.13: Carcinom.

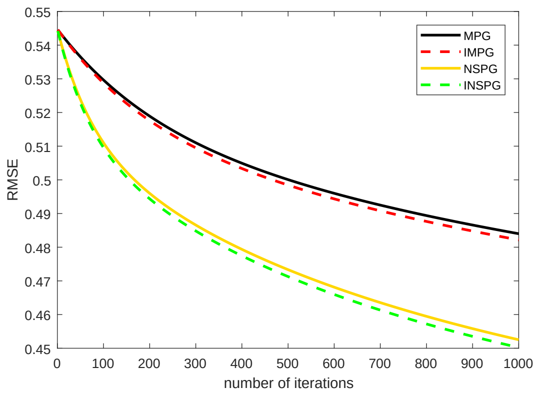

FIGURE 2.14: Lymphoma.

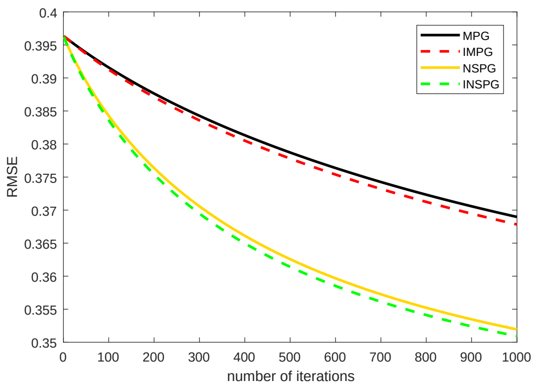

FIGURE 2.15: Nci9.

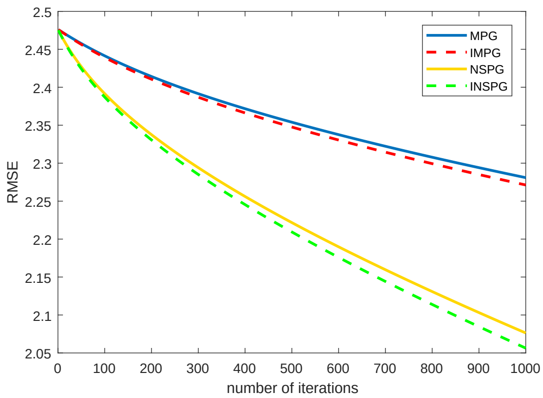

FIGURE 2.16: Lung discrete.

FIGURE 2.17: The graph is between number of iteration and corresponding root mean square error of the function.

- NSPG algorithm (2.28) is better than MPG algorithm (2.26) and IMPG algorithm (2.27) in terms of accuracy i.e. have less RMSE.

Datasets		MPG	IMPG	NSPG	INSPG
Colon	Obj fun	19.3569	18.8526	12.3988	**12.0180**
	RMSE	0.1391	0.1373	0.1113	**0.1096**
Allaml	Obj fun	66.9087	66.7423	64.6575	**64.5047**
	RMSE	0.1370	01368	0.1347	**0.1345**
Carcinom	Obj fun	3100.1	3081.9	2831.6	**2815.8**
	RMSE	0.8217	0.8193	0.7853	**0.7831**
Lymphoma	Obj fun	471.7750	468.1944	412.4782	**408.2835**
	RMSE	0.4840	0.4822	0.4525	**0.4502**
Nci9	Obj fun	661.1548	657.0767	601.5727	**598.0643**
	RMSE	0.3690	0.3678	0.3520	**0.3509**
Lung discrete	Obj fun	848.0305	840.8554	702.7026	**689.2480**
	RMSE	2.2809	2.2713	2.0763	**2.0563**

TABLE 2.2: Detailed analysis of proximal gradient algorithms. Objective function value and RMSE corresponding to different datasets at 1000 iteration. Best results are in bold letters.

- IMPG algorithm (2.27) has better accuracy than MPG algorithm (2.26) which is more effective in carcinom and lung discrete dataset.

Table 2.2 depicts objective function value and root mean square error at 1000^{th} iteration for all four algorithms and the best values are displayed in bold letters. Since all the bold letters are in the column of INSPG algorithm (2.29), clearly it explains the importance of the proposed algorithm over MPG algorithm (2.26), NSPG algorithm (2.28) and INSPG algorithm (2.29). NSPG algorithm (2.28) has better values than MPG algorithm (2.26) and IMPG algorithm (2.27) while MPG algorithm (2.26) performs worst for all datasets.

2.6 Conclusion

In this chapter, we have considered solving effectively the convex optimization problems which are frequently used in machine learning problems. First, we have proposed an accelerated iterative algorithm to solve fixed point problem of nonexpansive mappings and examined its detailed convergence behavior under mild conditions. Further, we have introduced a proximal gradient algorithm to solve the Lasso problem. Numerical experiments are conducted for the task of regression problem with high-dimensional datasets and we have compared the proposed proximal gradient algorithm with pre-existing algorithms on the basis of their convergence speed and accuracy. Numerical results have shown that the proposed algorithm surpasses the other algorithms in respect of performance.

In the future, we plan to study on different variants of S-iteration method and to apply it in various directions like image deblurring, signal processing and clustering problems etc.

****