Universidad Central de Venezuela
Facultad de Ciencias
Escuela de Computación

Lecturas en Ciencias de la Computación
ISSN 1316-6239

Unconstrained optimization techniques for the acceleration of alternating projection methods

L.M. HERNÁNDEZ, R. ESCALANTE y M. RAYDAN

RT 2011-01

Centro de Cálculo Científico y Tecnológico
CCCT-UCV
Caracas, Abril, 2011.

Unconstrained optimization techniques for the acceleration of alternating projection methods

Luis M. Hernández-Ramos * René Escalante† Marcos Raydan ‡

December 17, 2010

Abstract

Alternating projection methods have been extensively used to find the closest point, to a given point, in the intersection of several given sets that belong to a Hilbert space. One of the characteristics of these schemes is the slow convergence that can be observed in practical applications. To overcome this difficulty, several techniques, based on different ideas, have been developed to accelerate their convergence. Recently, a successful acceleration scheme was developed specially for Cimmino's method when applied to the solution of large-scale saddle point problems. This specialized acceleration scheme is based on the use of the well-known conjugate gradient method for minimizing a related convex quadratic map. In this work, we extend and further analyze this optimization approach for several alternating projection methods on different scenarios. In particular we include a specialized analysis and treatment for the acceleration of von Neumann-Halperin's method and Cimmino's method on subspaces, and Kaczmarz method on linear varieties. For some specific applications we illustrate the advantages of our acceleration schemes with encouraging numerical experiments.

Key words: Alternating projection methods, Cimmino's method, von Neumann-Halperin's method, conjugate gradient method.

1 Introduction

An important problem that appears quite frequently in many different applications is the following: Find the closest point, to a given point, in the intersection of several given sets that belong to a Hilbert space H, see, e.g., [13, 14, 27, 39, 42] and the references therein. Depending on the characteristics of the given sets different algorithms need to be used to solve this best approximation problem. We will consider the general family of Alternating Projection Methods (MAP) that consists, roughly speaking, in projecting onto each of the given sets individually to combine them in a suitable way to complete a cycle which is then repeated iteratively. One of the common characteristics of these alternating projection schemes is the slow convergence that can be observed in practice when the angles between the involved sets are small.

In this work, we will extend an optimization approach recently developed by Hernández-Ramos [29] to accelerate Cimmino's method [15] when applied to the solution of large-scale saddle point problems. For this special scenario, the well-known Conjugate Gradient (CG) method for minimizing a related convex quadratic map was combined with MAP to produce a very effective acceleration scheme. For a combination of Cimmino's method with CG for solving large and sparse linear systems, see also [2]. Taking advantage of the analysis developed by Kammerer and Nashed [33] for CG, when applied to singular linear operators, the optimization approach initiated in [29] will be adapted and analyzed for several different scenarios to produce specialized acceleration schemes.

*Departamento de Computación, Facultad de Ciencias, Universidad Central de Venezuela, Ap. 47002, Caracas 1041-A, Venezuela (luis.hernandez@ciens.ucv.ve). Partially supported by the Scientific Computing Center (CCCT) at UCV.

†Departamento de Cómputo Científico y Estadística, Universidad Simón Bolívar, Ap. 89000, Caracas 1080-A, Venezuela (rene@cesma.usb.ve). Partially supported by CESMA at USB.

‡Departamento de Cómputo Científico y Estadística, Universidad Simón Bolívar, Ap. 89000, Caracas 1080-A, Venezuela (mraydan@usb.ve). Partially supported by CESMA at USB.

The paper is organized as follows. In Section 2 we consider the case in which the given sets are subspaces of H. In that setting, we propose and analyze new acceleration schemes, for both Cimmino's method and von Neumann-Halperin's method, based on the conjugate gradient method applied to some equivalent optimization problems. In Section 3, we consider the case in which the given sets are linear varieties, and we propose and analyze acceleration schemes once again based on some low-cost optimization techniques. In this case, we concentrate on accelerating Cimmino's method due to its convenient properties. In Section 4, we present some numerical results to compare our proposed acceleration schemes with some well-known acceleration schemes that can be found in the literature, and also with the original alternating projection methods with no acceleration involved. Finally, in Section 5 we present our conclusions.

2 Acceleration of MAP onto subspaces

Let H be a Hilbert space with inner product $\langle .\,,. \rangle$ and associated norm $\|\,.\,\|$, and let $M_1, M_2, \dots M_n$ be closed (linear) subspaces of H with $M = \cap_{i=1}^n M_i$. For any closed subspace N of H, let P_N be the orthogonal projection onto N. The von Neumann-Halpering alternating projection method and the Cimmino's method are both members of the MAP family for determining the best approximation $P_M x$ to x in M. We now present, for the sake of completeness, two fundamental theorems that describe these well-known schemes and their convergence properties. The first one was established by von Neumann [43] for $n = 2$, and it was later extended by Halpering [26] for $n \geq 2$. The von Neumann-Halpering method is closely related to Kaczmarz alternating projection method [31] for solving linear systems of equations.

Theorem 2.1 *Let $M_1, M_2, \dots M_n$ be closed (linear) subspaces of a Hilbert space H with $M = \cap_{i=1}^n M_i$. Then,*

$$\lim_{k\to\infty} \|(P_{M_n} P_{M_{n-1}} \dots P_{M_1})^k x - P_M x\| = 0.$$

The second fundamental theorem is originally due to Cimmino [15] for solving systems of linear equalities, and was later extended by Kammerer and Nashed [32] for solving linear operator equations in Hilbert spaces. For additional extensions on several different scenarios see [7, 10, 11].

Theorem 2.2 *Let $M_1, M_2, \dots M_n$ be closed (linear) subspaces of a Hilbert space H with $M = \cap_{i=1}^n M_i$. Then,*

$$\lim_{k\to\infty} \|(\frac{1}{n}\sum_{i=1}^{n} P_{M_i})^k x - P_M x\| = 0.$$

Iterations of these methods can be viewed as fixed point schemes $x_k = T x_{k-1}$ for a bounded linear non-expansive operator T (i.e., $\|T\| \leq 1$, where $\|\,.\,\|$ is the linear-operators induced norm). For the von Neumann-Halpering method the operator is $T = P_{M_n} P_{M_{n-1}} \dots P_{M_1}$, and for the Cimmino's method the operator is $T = \frac{1}{n}\sum_{i=1}^n P_{M_i}$. In both cases, these schemes converge to the set of fixed points $Fix\ T = M = \cap_{i=1}^n M_i$, i.e.,

$$\lim_{k\to\infty} \|T^k x - P_{Fix\ T} x\| = 0.$$

When the angle between the subspaces involved is "small", both methods can present a slow rate of convergence [16, 17, 18, 30, 34], that implies a high computational cost because it requires many iterations of the alternating method to obtain a good approximation. Several acceleration schemes have been proposed for both methods, e. g. [1, 12, 23, 25, 30, 40].

In [6], Bauschke et al. propose a generalization of the accelerated schemes in [25] and [23]. This generalization is developed for non-expansive operators T. In this generalization, for a general bounded linear fixed point operator T such that

$$\lim_{k\to\infty} \|T^k x - P_{Fix\ T} x\| = 0,$$

the accelerated fixed point scheme, from $x_0 \in H$, can be written as follows:

$$x_{k+1} = A_T(x_k),$$

where

$$A_T(x) = t_x T x + (1 - t_x) x.$$

and

$$t_x = \begin{cases} \frac{\langle x, x-Tx\rangle}{\|x-Tx\|^2} & \text{if } Tx \neq x \\ 1 & \text{if } Tx = x. \end{cases}$$

With this choice of t_x, $A_T(x)$ is the point on the line through the points x and Tx closest to $P_M x$ [6].

In this work, we show that this accelerated scheme is related to the steepest descent method for solving singular linear equations in Hilbert spaces [38]. Taking into account that the steepest descent method is known for its slowness in the presence of ill-conditioned problems, we propose a new acceleration scheme based on the conjugate gradient method for solving singular linear equations in Hilbert spaces [33]. This conjugate gradient acceleration can be applied for accelerating any fixed point scheme defined by a linear closed range non expansive operator T, including the non expansive operators T related with the von Neumann-Halperin method and the Cimmino's method.

2.1 Acceleration of linear fixed point problems

Let T be a bounded linear operator in a Hilbert space H. The set of fixed points of T is:

$$Fix\ T = \{x \in H : Tx = x\}$$

The problem of finding the fixed points of T is equivalent to solving the linear system:

$$(I - T)x = 0.$$

In some cases, the operator $(I-T)$ is singular. For this case, an alternative is to solve the best approximation problem:

$$\min f(x) = \frac{1}{2}\|x - Tx\|^2. \tag{1}$$

If $Fix\ T$ is non empty, clearly the set of global minimizers of $f(x)$ is equal to $Fix\ T$. Therefore, the problem of finding the fixed points of T can be viewed as a particular case of the least-squares problem (in Hilbert spaces):

$$\min f(x) = \frac{1}{2}\|b - Ax\|^2, \tag{2}$$

where operator $A = I - T$ and $b = 0$.

Problem (2), has been treated previously by Nashed [38] for singular operators in Hilbert spaces using the steepest descent method, and later by Kammerer and Nashed [33] using the conjugate gradient method. In the following two theorems we summarize these fundamental results.

Theorem 2.3 (Nashed [38]) *Let H be a Hilbert space, $A : H \to H$ be a bounded linear operator such that its range is closed. The sequence $\{x_k\}$ generated by the steepest descent method applied to problem (2), converges to an element $u \in S = \{z : \inf \|Ax-b\| = \|Az-b\|,\ x \in H\}$ for any initial approximation $x_0 \in H$. Moreover, the sequence $\{x_k\}$ converges to $A^\dagger b$ if and only if $x_0 \in R(A^*)$.*

Theorem 2.4 (Kammered and Nashed [33]) *Let X and Y be two Hilbert spaces over the real field and let A be a bounded linear transformation mapping X to Y. If the range of A is closed, the conjugate gradient method, applied to problem (2) with initial approximation $x_0 \in H$, converges monotonically to the least squares solution $u = A^\dagger b + (I - P)x_0$ of $Ax = b$, where P denote the orthogonal projection of X onto $\overline{R(A^*)}$.*

A condition for applying these results to solve problem (1) is that the range of operator $I - T$ is closed. For finite dimensional spaces, this condition is satisfied automatically. However, for infinite dimensional spaces, the closed range condition for $I - T$ must be verified.

The iterates defined by the accelerated scheme A_T, from $x_0 \in H$, proposed in [6] are:

$$\begin{aligned} A_T(x_k) &= t_{x_k} Tx_k + (1 - t_{x_k})x_k \\ &= x_k + t_{x_k}(Tx_k - x_k) \\ &= x_k + t_k d_k, \end{aligned}$$

where $d_k = Tx_k - x_k$, and t_x is given by

$$t_{x_k} = \frac{\langle -x_k, d_k \rangle}{\langle d_k, d_k \rangle} \quad \text{for} \quad x_k \notin Fix\ T.$$

Therefore, this accelerated scheme correspond to the steepest descent sequence for solving the problem $(I-T)x = 0$, which is known for being a slow method. That explain the observation in [6] that the accelerated scheme A_T is not always faster that the original fixed point scheme when operator T is not a self-adjoint operator. That is the case when T is the operator associated with the von Neumman-Halperin alternating projection method [6]. Nevertheless, for symmetric versions, such as those associated with the self-adjoint operator T^*T, the convergence of the accelerated scheme is faster than their original versions [6]. Instead, for accelerating these fixed point schemes, we are more interested in the conjugate gradient sequence analyzed in [33] for singular linear systems in Hilbert spaces.

If the range $R(I-T)$ is closed, Theorem 2.4 implies that the conjugate gradient method applied to

$$\min f(x) = \frac{1}{2}\|(I-T)x\|^2, \tag{3}$$

converges to the least-squares solution

$$u = (I-T)^\dagger 0 + (I-P)x_0 = (I-P)x_0,$$

where P denotes the orthogonal projection of x onto $\overline{R(I-T)^*}$. Moreover,

$$\overline{R(I-T)^*} = \eta(I-T)^\perp = (Fix\ T)^\perp,$$

where $\eta(I-T)$ denotes the null of $(I-T)$.

Since $P_{Fix\ T} = I - P_{(Fix\ T)^\perp}$, the conjugate gradient method will converge to the projection of x_0 onto $Fix\ T$. Summarizing we obtain the following theorem.

Theorem 2.5 *Let T be a bounded non-expansive operator such as $R(I-T)$ is closed. Then, the conjugate gradient method for minimizing (1) with initial approximation $x_0 \in H$ converges monotonically to the projection of x_0 onto $Fix\ T$.*

However, to apply this result, it is necessary to show that $R(I-T)$ is a closed subspace. Below, we will characterize this closed range condition. For that, we will use the following result.

Theorem 2.6 (Kulkarni and Nair [35]) *Let $A : X \to Y$ be a nonzero bounded linear operator between two Hilbert spaces X and Y. The subspace R(A) is closed in Y if and only if there exists $\gamma > 0$ such that*

$$\sigma(A^*A\,|_{\eta(A)^\perp}) \subseteq \left[\gamma, \|A\|^2\right],$$

where $\sigma(A)$ denotes the spectrum of the operator A.

In other words, for $R(A)$ to be closed it suffices that the operator A verifies

$$0 < \gamma \leq \inf_{x \in \eta(A)^\perp} \frac{\langle x, A^*Ax\rangle}{\langle x, x\rangle} = \inf_{x \in \eta(A)^\perp} \frac{\|Ax\|^2}{\|x\|^2},$$

and

$$\sup_{x \in \eta(A)^\perp} \frac{\langle x, A^*Ax\rangle}{\langle x, x\rangle} = \sup_{x \in \eta(A)^\perp} \frac{\|Ax\|^2}{\|x\|^2} \leq \|A\|^2.$$

From properties of the norm, the sup condition is clearly verified. Since $\eta(I-T)^\perp = (Fix\ T)^\perp$, to prove that $A = I - T$ has closed range, it suffices to prove that there exists $\gamma > 0$ such that the Rayleigh quotient satisfies:

$$\inf_{x \in (Fix\ T)^\perp} \frac{\|x - Tx\|^2}{\|x\|^2} \geq \gamma > 0.$$

The following definition is a generalization of the definition of the cosine of the angle between subspaces $M_1, M_2, \ldots M_n$, given by Bauschke, Borwein and Lewis [5], and denoted by $\cos(M_1, M_2, \ldots M_n)$.

Definition 2.1 *Let T be a bounded non-expansive operator in a Hilbert space H. We define the cosine of T as*

$$\cos(T) = \|TP_{(Fix\ T)^\perp}\|.$$

Since T is a non-expansive operator ($\|T\| \leq 1$) then,

$$0 \leq \cos(T) = \|TP_{(Fix\ T)^\perp}\| \leq \|T\|\,\|P_{(Fix\ T)^\perp}\| \leq 1.$$

It is easy to show that $\cos(T) = \|T - P_{Fix\ T}\|$, i.e., the distance between operators T and the projection onto $Fix\ T$. In fact,

$$\cos(T) = \|TP_{(Fix\ T)^\perp}\| = \|T(I - P_{Fix\ T})\| = \|T - TP_{Fix\ T}\|.$$

Since $P_{Fix\ T}$ is a fixed point of T, we have that $TP_{Fix\ T} = P_{Fix\ T}$, and we obtain

$$\cos(T) = \|T - P_{Fix\ T}\|.$$

When T is the operator associated with the alternating projection method (i.e. $T = P_{M_n}P_{M_{n-1}}\dots P_{M_1}$), then $\cos(T) = \cos(M_1, M_2, \dots M_n)$, as defined in [5].

Lemma 2.1 *If $c = \cos(T) < 1$ then $R(I - T)$ is closed.*

Proof. Let $x \in (Fix\ T)^\perp$. By the Cauchy-Schwarz inequality,

$$\|Tx\| = \|TP_{(Fix\ T)^\perp}x\| \leq \|TP_{(Fix\ T)^\perp}\|\|x\| = c\|x\|.$$

By the triangular inequality,

$$\|x - Tx\| \geq \|x\| - \|Tx\|.$$

Hence, for all $x \in (Fix\ T)^\perp$ we have,

$$\|x - Tx\| \geq (1 - c)\|x\|$$

and

$$\inf_{x \in (Fix\ T)^\perp} \frac{\|x - Tx\|^2}{\|x\|^2} \geq (1 - c)^2 = \gamma > 0. \quad \blacksquare \tag{4}$$

Consequently, if $c = \cos(T) < 1$, we can apply the conjugate gradient acceleration. We summarize these results in the following theorem:

Theorem 2.7 *Let T be a bounded non-expansive operator such that $\cos(T) < 1$. Then, the conjugate gradient method for minimizing (1) with initial approximation $x_0 \in H$ converges monotonically to the projection of x_0 onto $Fix\ T$.*

Summing up, if $\cos(T) < 1$ and if we apply the conjugate gradient method for minimizing $f(x) = \frac{1}{2}\|x - Tx\|^2$, the iterations will converge to the fixed point of T closest to the initial approximation x_0, i. e., the sequence of iterates will converge to the projection of the initial approximation x_0, onto the fixed points $Fix\ T$.

To present the conjugate gradient acceleration algorithm, we need the gradient operator of $f(x) = \frac{1}{2}\|x - Tx\|^2$,

$$\nabla f(x) = (I - T)^*(I - T)x, \tag{5}$$

and its Hessian operator,

$$\nabla^2 f(x) = (I - T)^*(I - T). \tag{6}$$

Notice that the Hessian operator is constant and positive semi-definite, i.e., f is a convex quadratic map. In fact, for any $x \in H$ we have,

$$\langle x, \nabla^2 fx\rangle = \|(I - T)x\|^2 \geq 0.$$

In Algorithm 1 we show the conjugate gradient acceleration that computes $P_{Fix\ T}x_0$, i.e., the point in $Fix\ T$ closest to the initial approximation x_0. The rate of convergence of this algorithm is related to the spectral bounds of the restricted operator $U = \{(I-T)^*(I-T) \mid R(I-T)^*\}$. Let $\mu(x) = \langle Ux, x\rangle/\langle x, x\rangle,\ x \neq$

0, be the Rayleigh quotient of U. Since U is a symmetric positive definite linear operator on the Hilbert space $R(I-T)^* = (Fix\ T^*)^\perp = (Fix\ T)^\perp$, the spectral bounds:

$$s = \inf\{\mu(x) : x \in (Fix\ T)^\perp\}, \tag{7}$$

and

$$S = \sup\{\mu(x) : x \in (Fix\ T)^\perp\} \tag{8}$$

are positive and finite [33].

A bound for s is immediately given by (4),

$$s \geq (1-c)^2, \tag{9}$$

where $c = \cos(T)$.

Also, we have $\|x - Tx\| \leq \|x\| + \|Tx\|$. Since, for $x \in (FixT)^\perp$ we have $\|Tx\| \leq c$, hence

$$S \leq (1+c)^2. \tag{10}$$

Thus, a bound for the spectral condition number of the restricted operator U is given by

$$\mathbf{cond}\left((I-T)^*(I-T) \mid (Fix\ T)^\perp\right) \leq \left[\frac{1+c}{1-c}\right]^2.$$

Algorithm 1 Conjugate gradient acceleration for linear fixed point problems

INPUT: $x_0 \in H$
OUTPUT: $P_{Fix\ T}(x_0)$

$x = x_0$
$r = -(I-T^*)(I-T)x$
$rr = \langle r, r\rangle$
$u = r$
for $k = 1, 2, \ldots$ **do**
  if *convergence* **then**
    return $P_{Fix\ T}(x_0) = x$
  else
    $v = (I-T^*)(I-T)u$
    $\alpha = \frac{rr}{\langle u, v\rangle}$
    $x = x + \alpha u$
    $r0 = rr$
    $r = r - \alpha v$
    $rr = \langle r, r\rangle$
    $\beta = \frac{rr}{r0}$
    $u = r + \beta u$
  end if
end for

2.2 The self-adjoint case

An important case is when the operator T is self-adjoint. For this case, we will show that the condition $\cos(T) < 1$ is automatically satisfied under mild assumptions. First, some basic properties of self-adjoint operators will now be discussed in a more general setting: Normal operators. For that, let us denote $M := Fix\ T$. Clearly, If T commutes with P_M then T commutes with $P_{M^\perp}$.

Lemma 2.2 *If T commutes with P_M then $T^k - P_M = (TP_{M^\perp})^k$ for all k.*

Proof. Since T commutes with $P_{M^\perp}$, then,

$$(TP_{M^\perp})^k = T^k P_{M^\perp} = T^k(I - P_M) = T^k - T^k P_M.$$

Moreover, $TP_M = P_M$ because P_M is a fixed point of T, and so,

$$(TP_{M^\perp})^k = T^k - T^k P_M = T^k - P_M. \quad \blacksquare$$

Theorem 2.8 *Let T be a normal operator such that T commutes with P_M. Then, if $lim_{k\to\infty}\|T^k x - P_M x\| = 0$ for all $x \in H$, then $\cos(T) < 1$.*

Proof. If T commutes with P_M, using Lemma 2.2, we have for all $x \in H$

$$lim_{k\to\infty}\|T^k x - P_M x\| = lim_{k\to\infty}\|(TP_{M^\perp})^k x\| = 0.$$

Moreover, $lim_{k\to\infty}\|(TP_{M^\perp})^k x\| = 0$ for all $x \in H$ if and only if $lim_{k\to\infty}(TP_{M^\perp})^k = 0$ (see, e.g., [3]), which implies that $lim_{k\to\infty}\|(TP_{M^\perp})^k\| = 0$. Since $TP_{M^\perp}$ is also a normal operator then ([3])

$$\|(TP_{M^\perp})^k\| = \|(TP_{M^\perp})\|^k,$$

and hence

$$lim_{k\to\infty}\|(TP_{M^\perp})\|^k = 0.$$

Therefore $\|(TP_{M^\perp})\| < 1$, and so

$$\cos(T) = \|(TP_{M^\perp})\| < 1. \quad \blacksquare$$

Summing up, if T is normal and T commutes with $P_{Fix\,T}$, and the fixed point method converges, then the range of $I-T$ is closed an we can apply the conjugate gradient acceleration. For example, if $T = \frac{1}{n}\sum_{i=1}^n P_{M_i}$ (Cimmino's method), since the projection operators are self-adjoint then T is also a self-adjoint operator. Additionally, $P_M P_{M_i} = P_{M_i} P_M = P_M$, thus T commute with P_M, and the fixed point method converges ($lim_{k\to\infty}\|T^k x - P_M x\| = 0$), then $\cos(T) < 1$. In conclusion, we can apply the conjugate gradient acceleration directly to Cimmino's method.

2.2.1 Finite dimensional Case

The application of the conjugate gradient method to the related least-squares optimization problem (1) has the clear disadvantage of increasing the condition number of the involved linear operator. As we will discuss in our numerical results section, this in turn has an increasing effect in the number of required iterations for convergence when solving finite dimensional problems. If T is non-expansive the operator $(I - T)$ is clearly positive semi-definite. If the operator $(I - T)$, is symmetric and positive semi-definite then the proposed optimization approach could be applied directly to the system $(I-T)x = 0$, described above, for converge to the fixed points of T. An important case for which the operator $I-T$ is symmetric and positive semi-definite is Cimmino's method. Based on this important argument, for the rest of this work we will be interested in this method instead of other methods where the operator T is not symmetric.

The acceleration of Cimmino's method has been studied in [29], where a technique is proposed to find the projection onto the intersection of a finite number of closed subspaces of a Hilbert space. If $M_1, M_2, \ldots, M_r$ are closed subspaces of a Hlbert space H, this acceleration consists on minimizing, using the conjugate gradient method, the following quadratic function:

$$f(x) = \frac{1}{2}\sum_{i=1}^{r}\|x - P_{M_i}x\|^2 = \frac{1}{2}\sum_{i=1}^{r}\|(I - P_{M_i})x\|^2. \tag{11}$$

If a vector $x \in \cap_{i=1}^r M_i$, then $f(x) = 0$ and x minimizes f. For this function, the gradient $\nabla f(x) = (I-T_S)x$ and the Hessian is $\nabla^2 f(x) = I - T_S$, where $T_S = \frac{1}{n}\sum_{i=1}^n P_{M_i}(x)$, $I - T_S = \frac{1}{n}\sum_{i=1}^n (x - P_{M_i}(x))$ and M_i is a closed subspace for $i = 1, \ldots r$. This approach is equivalent to applying the conjugate gradient method to solve the linear symmetric semi-definite positive system $\nabla f(x) = 0$, i.e.,

$$(I - T_S)x = 0.$$

In Algorithm 2, Section 3, we present the acceleration scheme associated with Cimmino's method for the case when operator T is finite dimensional symmetric and positive semi-definite.

Algorithm 2 Conjugate gradient acceleration for linear symmetric and positive semi-definite fixed point problems

INPUT: $x_0 \in H$
OUTPUT: $P_{Fix\ T}(x_0)$

$x = x_0$
$r = -(I - T)x$
$rr = \langle r, r \rangle$
$u = r$
for $k = 1, 2, \ldots$ **do**
  if *convergence* **then**
    return $P_{Fix\ T}(x_0) = x$
  else
    $v = (I - T)u$
    $\alpha = \frac{rr}{\langle u, v \rangle}$
    $x = x + \alpha u$
    $r0 = rr$
    $r = r - \alpha v$
    $rr = \langle r, r \rangle$
    $\beta = \frac{rr}{r0}$
    $u = r + \beta u$
  end if
end for

3 Acceleration of MAP for linear varieties

In this section, we will extend the acceleration of Cimmino's method for the case of closed linear varieties. A linear variety is the translation of a subspace, i.e. any linear variety V can be written as $V = v_0 + S$ where S is a subspace. The subspace S is unique in this representation, but any vector in V can play the role of v_0.

Let H be a Hilbert space, $V_1, V_2, \ldots V_n$ closed linear varieties of H, $V = \cap_{i=1}^n V_i$, $V \neq \emptyset$, and $x \in H$. The problem to consider is: Find the projection of a vector x onto the linear variety V, $P_V(x)$, i.e. the closest point to x in V. A technique for solving this problem is Cimmino's method, that consists in the iteration:

$$x_{k+1} = \frac{1}{n} \sum_{i=1}^{n} P_{V_i}(x_k)$$

with $x_0 = x$. However, as explained before, Cimmino's method is known for being slow.

As shown in the previous section, a technique for accelerating Cimmino's method in subspaces is to apply directly the conjugate gradient method on the positive semi-definite linear system,

$$(I - T_S)x = 0,$$

where $T_S = \frac{1}{n} \sum_{i=1}^n P_{S_i}(x)$, $I - T_S = \frac{1}{n} \sum_{i=1}^n (x - P_{S_i}(x))$, and S_i is a closed subspace for $i = 1, \ldots n$. This accelerated iteration converges to $P_{\cap_{i=1}^n S_i} x_0$ where x_0 is the initial iterate [29].

Nevertheless, in the case of linear varieties, due to the involved translation, the operator $I - T_V = \frac{1}{n} \sum_{i=1}^n (x - P_{V_i}(x))$ is not a linear operator. Hence, we propose for that to solve the non linear equation $F(x) = 0$, where $F(x) = (I - T_V)x$, by a suitable technique for finding roots of a non-linear map. A suitable technique will be described below. First, let us recall that $V = \cap_{i=1}^n V_i$ is also a linear variety of H.

We now list some well-known identities that will be useful in our subsequent results.

Lemma 3.1 *Let $V = S + v_0$ be a closed linear variety with S a closed linear subspace. Let P_V be the projection onto V. Then*

1. *$P_V(x) = P_S(x) + P_{S^\perp}(v_0)$, for any $x \in H$.*

2. $x - P_V(x) = P_{S^\perp}(x - v_0)$*, for any* $x \in H$*.*

Where $P_{S^\perp} = I - P_S$ *is the projection onto the orthogonal complement of* S*.*

Let V_i be linear varieties for $i = 1, \ldots, n$. We now define $F : \Re^n \mapsto \Re^n$ as follows

$$F(x) = \frac{1}{n}\sum_{i=1}^{n}(x - P_{V_i}(x)). \tag{12}$$

Our next result establishes that the roots of $F(x)$ corresponds to points in $V = \cap_{i=1}^n V_i$.

Theorem 3.1 $F(x) = 0$ *if and only if* $x \in \cap_{i=1}^n V_i$.

Proof.

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} x - P_{V_i}(x) = \frac{1}{n}\sum_{i=1}^{n} P_{S_i{}^\perp}(x - v_0) = (x - v_0) - \frac{1}{n}\sum_{i=1}^{n} P_{S_i}(x - v_0),$$

where $v_0 \in \cap_{i=1}^n V_i$. Therefore,

$$F(x) = 0 \text{ if and only if } \frac{1}{n}\sum_{i=1}^{n} P_{S_i}(x - v_0) = x - v_0.$$

We have reduced the problem from linear varieties to subspaces. We can observe that $F(x) = 0$ if and only if Cimmino's method (for subspaces) has a solution in $x - v_0$ i. e.

$$x - v_0 \in \cap_{i=1}^n S_i = S,$$

however, as $v_0 \in V$ then $x \in S + v_0 = V = \cap_{i=1}^n V_i$. ∎

As a consequence, the solutions of $F(x) = 0$ are vectors in $V = \cap_{i=1}^n V_i$. We are not interested in any arbitrary vector in $\cap_{i=1}^n V_i$, but in the closest vector, in $\cap_{i=1}^n V_i$, to a given vector x. This special task can be accomplished if we use specialized methods for solving $F(x) = 0$. In other words, in general, standard methods (Newton-type methods) for solving $F(x) = 0$ do not necessarily have the optimality property described above. In contrast, any iterative method that uses plus or minus the residual direction, $F(x_k)$, as a search direction has automatically the desired optimality property. Fortunately, residual methods of this kind, with additional practical features, have been recently developed, and will be described and analyzed in the next subsection.

Before describing such a suitable method for solving the nonlinear system $F(x) = 0$ given by (12), we need to identify the Fréchet derivative of F, $F'(x)$, and some of its properties. For that we now take advantage of the equivalence of applying Cimmino's method on the linear varieties, V_i's, and Cimmino's method on the subspaces S_i's, assuming the knowledge of a hypothetical $v_0 \in V = \cap_{i=1}^n V_i$.

Lemma 3.2 *The equation* $F(x) = 0$ *is equivalent to* $(I - T_S)x = b$ *with* $b = (I - T_S)v_0$ *and* $T_S = \frac{1}{n}\sum_{i=1}^n P_{S_i}$*, for any* $v_0 \in V$*. Moreover, the Fréchet derivative of* $F(x)$ *is given by* $F'(x) = (I - T_S)$*.*

Proof.

$$F(x) = \frac{1}{n}\sum_{i=1}^{n}(x - P_{V_i}(x)) = (x - v_0) - \frac{1}{n}\sum_{i=1}^{n} P_{S_i}(x - v_0) = \left(\frac{1}{n}\sum_{i=1}^{n} P_{S_i}(v_0) - v_0\right) - \left(\frac{1}{n}\sum_{i=1}^{n} P_{S_i}(x) - x\right),$$

then

$$F(x) = 0 \Longleftrightarrow \left(x - \frac{1}{n}\sum_{i=1}^{n} P_{S_i}(x)\right) = \left(v_0 - \frac{1}{n}\sum_{i=1}^{n} P_{S_i}(v_0)\right),$$

and therefore

$$F(x) = 0 \Longleftrightarrow (I - T_S)x = b \equiv (I - T_S)v_0.$$

Clearly, from this equivalence, the Fréchet derivative of $F(x)$ is given by $F'(x) = (I - T_S)$. ∎

Theorem 3.2 *The Fréchet derivative of F, defined by (12), is self-adjoint and positive semi-definite.*

Proof. From Lemma 3.2, $F'(x) = (I - T_S)$, and the operator $(I - T_S)$ is clearly self-adjoint. To establish that it is positive semi-definite, consider the following calculations, for any $x \neq 0$:

$$\langle \frac{1}{n}\sum_{i=1}^{n}(I - P_{S_i})x, x\rangle = \frac{1}{n}\sum_{i=1}^{n}\langle (I - P_{S_i})x, x\rangle,$$

but $(I - P_{S_i}) = P_{S_i{}^{\perp}}$ is also a projection. Hence,

$$\langle (I - T_S)x, x\rangle = \frac{1}{n}\sum_{i=1}^{n}\langle P_{S_i{}^{\perp}}x, x\rangle = \frac{1}{n}\sum_{i=1}^{n}\|P_{S_i{}^{\perp}}x\|^2 \geq 0. \quad \blacksquare$$

If we known a vector $v_0 \in \cap_{i=1}^{n} V_i$ then the problem can be seen again as a subspace problem. In that case, $F(x) = 0$ becomes a positive semi-definite linear system. In fact, the problem to be solved would be $(I - T_S)x = b$, where $T_S = \frac{1}{n}\sum_{i=1}^{n} P_{S_i}(x)$ and $b = (I - T_S)v_0$, and we could use the conjugate gradient acceleration as described in the previous section.

We would like to stress out that the Fréchet derivative, $F'(x) = (I - T_S)$, of the map $F(x)$ defined in (12) cannot be explicitly known since it depends on the operator T_S which, in general, is not known when dealing with linear varieties or it could be computationally expensive to obtain. Therefore, the conjugate gradient method cannot be used in the presence of linear varieties. However, the properties of $F'(x)$ that will be analyzed in our next subsection guarantees convergence of a specialized residual scheme to be described for solving nonlinear systems of equations (i.e., in the finite dimensional case).

3.1 Residual methods for nonlinear systems

Residual methods that systematically use plus or minus the residual direction, $F(x_k)$, as a search direction for solving large-scale nonlinear system of equations $F(x) = 0$, have been recently proposed and analyzed by La Cruz et al. [36, 37]. As we will discuss later on this subsection, these schemes have the advantage for our special problem, that they automatically converge to the desired closest point to the given initial point.

The residual iterations (DFSANE and SANE) proposed and analyzed in [36, 37], for finite-dimensional nonlinear problems, are defined as

$$x_{k+1} = x_k \pm \alpha_k\, F(x_k), \tag{13}$$

where $\alpha_k > 0$ is the step-length and the search direction is either $F(x_k)$ or $-F(x_k)$ depending on which one is a descent direction for the merit function

$$f(x) = \|F(x)\|^2 = \langle F(x), F(x)\rangle. \tag{14}$$

These ideas have become effective, and competitive schemes for solving large-scale nonlinear systems, when the step lengths are chosen in a suitable way. The convergence of (13) is attained, for general nonlinear problems, when it is associated with a free-derivative non-monotone line search, fully described in [36]. Fortunately, since the Jacobian of $F(x)$, given by (12), is symmetric and positive semi-definite (Lemma 3.2 and Corollary 3.2) we will have convergence for the pure method, i.e., without the globalization strategy, and moving along the direction $-F(x_k)$ which is always a descent direction for $f(x)$, in our case.

For the choice of the step-length $\alpha_k > 0$, there are many options for which convergence is guaranteed. A well-known choice is the non-monotone spectral choice that has interesting properties, and is defined as the absolute value of

$$\alpha_k = \frac{\langle s_{k-1}, s_{k-1}\rangle}{\langle s_{k-1}, y_{k-1}\rangle}, \tag{15}$$

where $s_{k-1} = x_k - x_{k-1}$, and $y_{k-1} = F(x_k) - F(x_{k-1})$. Obtaining the step length using (15) requires a reduced amount of computational work, accelerates the convergence of the process, and involves the last two iterations in such a way that incorporates first order information into the search direction [4, 9, 41, 21].

Algorithm 3 DFSANE for $F(x)$ in its pure form

Let $X_0 \in H, \alpha_0 \in \Re$, $\alpha_0 \neq 0$

for $k = 0, 1, \cdots$ **do**

$x_{k+1} = x_k - \alpha_k F(x_k)$

$s_k = x_{k+1} - x_k$

$y_k = F(x_{k+1}) - F(x_k)$

$\alpha_{k+1} = \langle s_k, s_k \rangle / \langle s_k, y_k \rangle$

end for

Remark 2. Algorithm 3 produces a sequence $\{x_k\}$ that converges globally, i.e., from any initial guess x_0, to a vector $\bar{x}$ such that $F(\bar{x}) = 0$. This theoretical result follows directly from [22, Theorem 2.1], applied to the equivalent formulation (11), since the Jacobian of F is symmetric and positive semi-definite.

To close this section we now establish that the limit point of the sequence $\{x_k\}$ generated by Algorithm 3 has the optimality property of being the closest vector, in $\cap_{i=1}^n V_i$, to the given initial vector x_0. First, we show two required results.

Lemma 3.3 *Let $V = \cap_{i=1}^n V_i$, $v_0 \in V$, $V_i = S_i + v_0$, $V = S + v_0$, $S = \cap_{i=1}^n S_i$, then*

$$P_S F(x) = 0.$$

Proof.

$$P_S F(x) = P_S \frac{1}{n} \sum_{i=1}^{n} (x - P_{V_i}(x)) = \frac{1}{n} \sum_{i=1}^{n} P_S \left(x - v_0 - P_{S_i}(x - v_0)\right) = \frac{1}{n} \sum_{i=1}^{n} \left[P_S(x - v_0) - P_S P_{S_i}(x - v_0)\right],$$

but $P_S P_{S_i} = P_S$ because $S \subseteq S_i, i = 1, \ldots, n$.

Hence,

$$P_S F(x) = \frac{1}{n} \sum_{i=1}^{n} \left[P_S(x - v_0) - P_S(x - v_0)\right] = 0. \quad \blacksquare$$

Theorem 3.3 *The iteration $x_{k+1} = x_k - \alpha_k F(x_k)$ with $x_0 \in H$ verifies,*

$$P_V(x_{k+1}) = P_V(x_k), \quad k = 0, \ldots,$$

Proof. Combining Lemma 3.1 and Lemma 3.3, we have

$$\begin{aligned} P_V(x_{x+1}) &= P_S(x_{k+1}) + P_{S^\perp}(v_0) \\ &= P_S(x_k - \alpha_k F(x_k)) + P_{S^\perp}(v_0) \\ &= P_S(x_k) - \alpha_k P_S F(x_k) + P_{S^\perp}(v_0) \\ &= P_S(x_k) + P_{S^\perp}(v_0) \\ &= P_V(x_k). \quad \blacksquare \end{aligned}$$

By induction, all the iterates of $\{x_k\}$ have the same projection into the linear variety V. As the iteration converges to a vector $v \in V$, we have,

$$v = P_V(v) = P_V(x_0).$$

We summarize this in the following theorem.

Theorem 3.4 *The iteration $x_{k+1} = x_k - \alpha_k F(x_k)$ generated by Algorithm 3, from $x_0 \in H$, converges to $P_V(x_0)$.*

4 Numerical experiments

We compare our acceleration schemes, based on low-cost optimization techniques, with some well-know acceleration schemes that can be found in the literature, and also with the original alternating projection methods with no acceleration involved. We will present experiments for some selected subspace problems and also on some selected linear variety problems. All our experiments were run on an Intel Core 2 Duo using MATLAB 7.6. In all the experiments and for all the considered methods, we stop the process when the norm of the residual is less than or equal to 10^{-8}. For each scenario (subspaces or linear varieties) we use a different norm that will be specified below.

4.1 Projecting onto subspaces

We compare the methods of Cimmino and von Neumann-Halperin (MAP) with several different accelerated versions:

1. von-Neumann-Halperin type methods:
   (a) Clasical von-Neumann-Halperin method (MAP without acceleration).
   (b) Gearhart and Koshy acceleration [23] (see also Bauschke et al. in the finite dimensional case [6]).
   (c) (New) Conjugate gradient method to minimize $f(x) = \frac{1}{2}\|x - Tx\|^2$ where T is the operator associated with the von-Neumann-Halperin method (MAP-CG).
2. Cimmino type methods:
   (a) Clasical Cimmino's method (without acceleration).
   (b) (New) Conjugate gradient method to minimize $f(x) = \frac{1}{2}\|x - Tx\|^2$ where T is the operator associated with the Cimmino's method (CIM-CG-LS).
   (c) (New) Conjugate gradient method to solve $(I - T)x = 0$ where T is the symmetric positive semi-definite operator associated with the Cimmino's method (Cimmino-CG).

In our experiments alternating projection methods are used for solving saddle point problems, which appear in a wide range of applications (see e.g., [8]).

$$\begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix}\begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix}, \tag{16}$$

where

$(H1)$ $B \in \Re^{m\times n}$ is a matrix with $rank(B) = m, m \leq n$,

$(H2)$ $A \in \Re^{n\times n}$ is a symmetric positive definite matrix.

Under hypotheses (H1,H2), if $(x, \lambda)^T$ solves (16), then the vector x corresponds to the orthogonal projection of $x_u = A^{-1}f$ onto $\ker B$, in the scalar product $\langle ., . \rangle_A$. To be precise, the m involved subspaces, in this case, are given by: $\{x : \langle b_i, x\rangle_A = 0\}$, for $1 \leq i \leq m$, where b_i represents the i-th row of B [28, 29].

We compare these methods for solving saddle point problems from:

- A selection of saddle point problems from the CUTEr collection [24].
- A set of Stokes saddle point problems generated using IFISS incompressible flow software associated described in the book by Elman et al. [20] (stokes_testproblems):
  - STOKES1: Channel domain with natural outflow boundary.
  - STOKES2: Flow over a backward facing step.
  - STOKES3: Lid driven cavity.
  - STOKES4: Colliding flow.

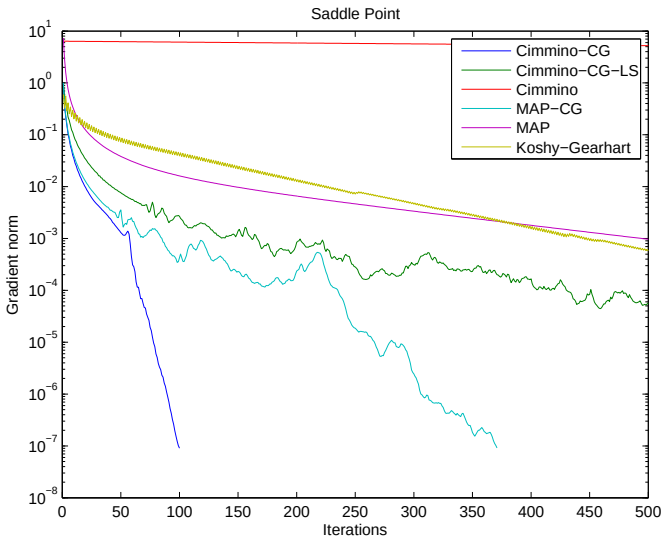

Figure 1: Different accelerations for a CUTEr problem (AUG2DCQP)

- Several domain decomposition saddle point problems discussed in [28] (DD1, DD2, DD3 and DD4).

For these subspace problems, we use the norm defined by $\|x\|_A = \langle x, Ax\rangle$. For all sets of problems, the matrix $A \in \Re^{n\times n}$ is symmetric and positive definite, and $B \in \Re^{m\times n}$ is a full row rank matrix. For the CUTEr problems, and also for Stokes problems we set the block $(2,2)$ of the saddle point matrix to be the zero matrix, for obtaining a system like (16). It is important to note that an augmented Lagrangean technique has been used to improve the behavior of all considered methods, and also to guarantee that the A matrix is positive definite. To be precise, the A matrix in (16) has been replaced by $A + \gamma B^T B$, where $\gamma = \|A\|_F / \|B\|_F^2$ [8]. Moreover, for the problems coming from the CUTEr Library, we also added to the left upper block of (16) the scalar 0.01 times the identity matrix, to improve the condition number of that block. All experiments were obtained by a one-row block partition of B (number of subspaces equal to m, where m is the number of rows of the matrix B).

4.1.1 Numerical results

- Accelerations based on MAP:

  Table 1 shows the convergence behavior of the different accelerations of von Neumann-Halperin type of methods. Among the methods based on the method of von-Neumann-Halperin, there is no one that can be declared the best for all the considered problems. The conjugate gradient acceleration of the MAP method has the disadvantages caused by the increase of conditioning resulting from the resolution of a least-squares problem; and the higher cost per iteration is due to the projection onto each subspace which is made twice per iteration. However, it was quite consistent in all problems. The acceleration proposed by Gearhart and Koshy was inconsistent in several of these problems, and required a large number of iterations for convergence. This inconsistent behavior of the Gearhart and Koshy acceleration was already commented in [6].

- Accelerations based on Cimmino's method:

  Table 2 shows the convergence behavior of the different accelerations for Cimmino's method. The Cimmino-CG acceleration applied to the system $(I - T)x = 0$ is clearly superior than the other accelerations, including the acceleration based in the method of von-Neumann. Cimmino-CG was in general much better than the others in number of iterations, and also in CPU time, required for convergence. Since $(I - T)$ is a symmetric operator, the ill-conditioning effect that comes from solving

Problem			MAP		Gearhart-Koshy		MAP-CG	
Name	m	n	iter	time	iter	time	iter	time
AUG2DCQP	1600	3280	*	*	*	*	371	407.81
CVXQP1S	50	100	652	1.42	312	1.70	53	0.39
CVXQP2S	25	100	13	0.03	11	0.03	39	0.09
CVXQP3S	75	100	*	*	*	*	206	1.20
DUALC1	215	223	2	0.06	*	*	2	0.13
DUALC5	278	285	2	0.08	*	*	2	0.16
DUALC8	503	510	1	0.30	*	*	2	0.47
GOULDQP2S	349	659	*	*	*	*	1	0.38
KSIP	1001	1021	*	*	*	*	21	20.06
MOSARQP1	700	3200	2	6.28	*	*	4	6.97
PRIMAL1	85	410	20	0.39	21	0.39	14	0.56
PRIMAL2	96	745	19	0.66	19	0.74	14	0.72
PRIMAL3	111	856	33	1.75	31	1.75	19	1.81
PRIMAL4	75	1564	17	1.19	18	1.23	14	1.94
PRIMALC1	9	239	7	0.03	*	*	4	0.03
PRIMALC2	7	238	6	0.03	*	*	3	0.03
QGROW15	300	645	6	0.53	*	*	6	0.59
QGROW22	440	946	6	0.84	*	*	9	1.41
QSCFXM3	990	1800	*	*	*	*	293	86.98
STOKES1	1024	2178	*	*	*	*	256	173
STOKES2	2816	5890	*	*	*	*	247	94.52
STOKES3	256	578	306	7.31	199	4.84	109	5.45
STOKES4	1024	2178	264	7.25	187	5.14	96	5.36
DD1	80	1600	199	1.66	156	1.32	102	1.80
DD2	85	1925	546	8.81	*	*	134	4.36
DD3	120	3600	215	6.75	151	4.31	135	9.56
DD4	125	3975	*	*	*	*	209	13.73

Table 1: Comparison between different accelerated von Neumann-Halperin methods for subspaces

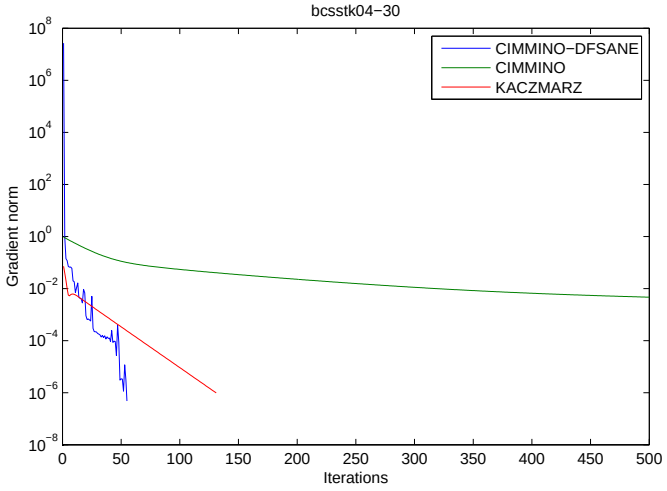

Figure 2: Different accelerations for the 30-rows matrix bcsstk04 (linear varieties)

Problem			Cimmino's		CIM-CG-LS		Cimmino-CG	
Name	m	n	iter	time	iter	time	iter	time
AUG2DCQP	1600	3280	*	*	*	*	100	50.73
CVXQP1S	50	100	*	*	104	1.70	54	0.33
CVXQP2S	25	100	*	*	25	0.09	17	0.06
CVXQP3S	75	100	*	*	605	3.80	133	0.53
DUALC1	215	223	*	*	2	0.19	2	0.09
DUALC5	278	285	*	*	2	0.16	2	0.16
DUALC8	503	510	*	*	2	0.48	2	0.38
GOULDQP2S	349	659	*	*	2	0.38	2	0.38
KSIP	1001	1021	*	*	11	11.37	6	3.92
MOSARQP1	700	3200	*	*	4	7.05	2	6.38
PRIMAL1	85	410	*	*	21	0.72	10	0.28
PRIMAL2	96	745	*	*	21	1.00	10	0.45
PRIMAL3	111	856	*	*	32	2.98	13	0.88
PRIMAL4	75	1564	*	*	21	1.94	10	1.08
PRIMALC1	9	239	133	0.33	4	0.03	3	0.03
PRIMALC2	7	238	101	0.23	4	0.03	3	0.03
QGROW15	300	645	*	*	7	0.70	5	0.47
QGROW22	440	946	*	*	10	1.61	7	0.91
QSCFXM3	990	1800	*	*	*	*	149	32.16
STOKES1	1024	2178	*	*	111	82.94	72	27.38
STOKES2	2816	5890	*	*	259	99.55	101	22.89
STOKES3	256	578	*	*	139	7.17	44	1.30
STOKES4	1024	2178	*	*	97	5.64	35	1.28
DD1	80	1600	*	*	179	3.34	62	0.63
DD2	85	1925	*	*	182	6.23	79	1.36
DD3	120	3600	*	*	240	16.28	72	2.72
DD4	125	3975	*	*	209	13.73	102	3.80

Table 2: Comparison between different accelerated Cimmino's methods for subspaces

a least-squares problem is avoided, and the cost per iteration does not increase because a projection onto each subspace is performed only once per iteration.

4.2 Projecting onto linear varieties

Our numerical experiments consist in finding the projection of x_0 onto the linear variety $V = \cap_{i=1}^{m} H_i$ where H_i is the hyperplane defined, for $i = 1, \dots m$, by:

$$H_i = \{x : \langle x, a_i \rangle = i\}.$$

In all the experiments of this section, $x_0 = (1, 1, ..., 1)^T$. The vectors a_i come from the first m rows of matrices chosen from the Hilbert matrix (MATLAB routine `hilb`), and from matrices coming from the Harwell-Boeing collection [19]. We compare the performance of the following methods: Cimmino's method, Kaczmarz's method [31], and the proposed acceleration for Cimmino's method combined with the DFSANE method (Cimmino-DFSANE).

4.2.1 Numerical results

In Table 3 we can observe that for the selected test problems the acceleration Cimmino-DFSANE strongly accelerates the classical Cimmino's method, and in almost all test problems it has a much better performance

than the Kaczmarz's method. As well as for Cimmino's method, the Cimmino-DFSANE acceleration has advantages in parallel machines because all the projections onto each linear variety can be made independently (in sharp contrast to Kaczmarz's method).

We can also observe that as the number of linear varieties increases, the relative improvement in performance also increases.

5 Concluding remarks

We have presented and analyzed acceleration schemes for some well-known alternating projection methods. Our acceleration schemes are based on the application of low-cost numerical optimization techniques over some equivalent formulations of the original problem: Find the closest point, to a given point, in the intersection of several given sets that belong to a Hilbert space H.

When the involved sets are subspaces, we combine the conjugate gradient (CG) method with MAP and also with Cimmino's method, and produce a significant acceleration in both cases, according to our preliminary and encouraging numerical results. For MAP we apply CG on the standard least-squares approach; and for Cimmino's method we apply CG directly on the nonlinear system $(I-T)x = 0$, where $T = \frac{1}{n}\sum_{i=1}^{n} P_{M_i}$, taking advantage of the fact that $(I-T)$ is a self-adjoint and positive semi-definite operator. This direct approach combined with Cimmino's method produces the most effective choice of the ones considered in our numerical results for subspaces.

When the involved sets are linear varieties we concentrate our attention on Cimmino's method. Unfortunately, the associated operator $I - T_V = \frac{1}{n}\sum_{i=1}^{n}(x - P_{V_i}(x))$ is not a linear operator, and so the use of CG is not possible, for general problems. For that case, we propose and analyze an acceleration technique that combines Cimmino's method with a simplified version of DF-SANE, a low-cost derivative-free technique for solving nonlinear systems of equations. This combined scheme produces a significant acceleration when compared to previously known techniques, including the well-known Kaczmarz method. An interesting line of research is to extend this machinery to accelerate the alternating projection methods specially designed for solving convex feasibility problems.

References

[1] G. APPLEBY and D. SMOLARSKI. A linear acceleration row action method for projecting onto subspaces. *Electronic Transactions on Numerical Analysis (ETNA)*, 20:253–275, 2005.

[2] M. ARIOLI, I. S. DUFF, J. NOAILLES, and D. RUIZ. A block projection method for sparse matrices. *SIAM J. Sci. Statist. Comput.*, 13:47–70, 1992.

[3] G. BACHMAN and L. NARICI. *Functional Analysis*. Dover, 2000.

[4] J. BARZILAI and J. M. BORWEIN. Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8:141–148, 1988.

[5] H. H. BAUSCHKE, J. M. BORWEIN, and A. S. LEWIS. On the method of cyclic projections for convex sets in Hilbert space. *Recent Developments in Optimization Theory and Nonlinear Analysis, Contemporary Mathematics*, 204:1–38, 1997.

[6] H. H. BAUSCHKE, F. DEUTSCH, H. HUNDAL, and S.-H. PARK. Accelerating the convergence of the method of alternating projections. *Transactions of the American Mathematical Society*, 355(9):3433–3461, 2003.

[7] M. BENZI. Gianfranco Cimmino's contributions to numerical mathematics. *Atti del Seminario di Analisi Matematica, Dipartimento di Matematica dell'Universita di Bologna. Special Volume: Ciclo di Conferenze in Memoria di Gianfranco Cimmino, March-April*, pages 87–109, 2004.

[8] M. BENZI, G. GOLUB, and J. LIESEN. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.

[9] E. G. BIRGIN, J. M. MARTINEZ, and M. RAYDAN. Spectral projected gradient methods. In C. A. Floudas and P. M. Pardalos, editors, *Encyclopedia of Optimization, Second Ed., Part 19*, pages 3652–3659. 2009.

[10] D. BUTNARIU and Y. CENSOR. Strong convergence of almost simultaneous projection methods in Hilbert spaces. *Journal of Computational and Applied Mathematics*, 53:33–42, 1994.

[11] D. BUTNARIU and S. J. FLAM. Strong convergence of expected projection methods in Hilbert spaces. *Numerical Functional Analysis and Optimization*, 16:601–636, 1995.

[12] Y. CENSOR. Computational acceleration of projection algorithms for the linear best approximation problem. *Linear Algebra and its Applications*, 416(1):111–123, 2006.

[13] Y. CENSOR, M. D. ALTSCHULER, and W. D. POWLIS. On the use of Cimmino's simultaneous projection method for computing a solution of the inverse problem in radiation therapy treatment planning. *Inverse Problems*, 4:607–623, 1988.

[14] Y. CENSOR and S.A. ZENIOS. *Parallel Optimization: Theory, Algorithms, and Applications*. Oxford University Press, USA, 1997.

[15] G. CIMMINO. Calcolo approsimato per le soluzioni dei sistemi di equazioni lineari. *La Ricerca Scientifica, II*, 9:326–333, 1938.

[16] F. DEUTSCH. Rate of convergence of the method of alternating projections. *ISNM*, 72, Birkhäuser Verlag, Basel:96–107, 1984.

[17] F. DEUTSCH. The angle between subspaces of a Hilbert space. In *Approximation theory, wavelets and applications (Maratea, 1994)*, pages 107–130. Kluwer Acad. Publ., Dordrecht, 1995.

[18] F. DEUTSCH and H. HUNDAL. The rate of convergence for the cyclic projections algorithm I: Angles between convex sets. *Journal of Approximation Theory*, 142(1):36–55, 2006.

[19] I. S. DUFF, R. G. GRIMES, and J. G. LEWIS. *User's guide for the Harwell-Boeing sparse matrix collection*. Research and Technology Division, Boeing Computer Services, Wasington, USA, 1992.

[20] H. C. ELMAN, D. J. SILVESTER, and A. J. WATHEN. *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Oxford University Press, USA, 2005.

[21] R. FLETCHER. On the Barzilai-Borwein method. In L.Qi, K. L. Teo, and X. Q. Yang, editors, *Optimization and Control with Applications*, pages 235–256. Springer, 2005.

[22] A. FRIEDLANDER, J. M. MARTINEZ, and M. RAYDAN. A new method for large-scale box constrained convex quadratic minimization problems. *Opt. Methods and Software*, 5:57–74, 1995.

[23] W. B. GEARHART and M. KOSHY. Acceleration schemes for the method of alternating projections. *J. Comp. Appl. Math.*, 26:235–249, 1989.

[24] N. I. M. GOULD, D. ORBAN, and P. L. TOINT. CUTEr and SifDec: A constrained and unconstrained testing environment, revisited. *ACM Transactions on Mathematical Software (TOMS)*, 29(4):394, 2003.

[25] L. G. GUBIN, B. T. POLYAK, and E. V. RAIK. The method of projections for finding the common point of convex sets. *USSR Comp. Math. Phys.*, 7:1–24, 1967.

[26] I. HALPERIN. The product of projection operators. *Acta Sci. Math. (Szeged)*, 23:96–99, 1962.

[27] G. T. HERMAN. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*. 2nd ed., Springer, 2009.

[28] L. M. HERNANDEZ-RAMOS. Alternating oblique projections for coupled linear systems. *Numerical Algorithms*, 38(4):285–303, 2005.

[29] L. M. HERNANDEZ-RAMOS. Accelerated Cimmino's method for saddle point problems. Technical Report RT-2008-06, Escuela de Computación. Universidad Central de Venezuela, 2008. Submitted.

[30] A. N. IUSEM and A. R. DE PIERRO. Convergence results for an accelerated nonlinear Cimmino algorithm. *Numerische Mathematik*, 49(4):367–378, 1986.

[31] S. KACZMARZ. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. Internat. Acad. Pol. Sci. Lett.*, A 35:355–357, 1937.

[32] W. J. KAMMERER and M. Z. NASHED. A generalization of a matrix iterative method of G. Cimmino to best approximate solutions to linear integral equations of the first kind. *Rendiconti della Classe di Scienze Fisiche, Matematiche e Naturali dell'Accademia Nazionale dei Lincei, VIII*, 51:20–25, 1971.

[33] W. J. KAMMERER and M. Z. NASHED. On the convergence of the conjugate gradient method for singular linear operator equations. *SIAM J. Numerical Analysis*, 9(1):165–181, 1972.

[34] S. KAYALAR and H. L. WEINERT. Error bounds for the method of alternating projections. *Math. Control Signals Systems*, 1:43–59, 1988.

[35] S. H. KULKARNI and M. T. NAIR. A characterization of closed range operators. *Indian Journal of Pure and Applied Mathematics*, 31(4):353–362, 2000.

[36] W. LA CRUZ, J. M. MARTINEZ, and M. RAYDAN. Spectral residual method without gradient information for solving large-scale nonlinear systems of equations. *Mathematics of Computations*, 75:1429–1448, 2006.

[37] W. LA CRUZ and M. RAYDAN. Nonmonotone spectral methods for large-scale nonlinear systems. *Optimization Methods and Software*, 18:583–599, 2003.

[38] M. Z. NASHED. Steepest descent for singular linear operator equations. *SIAM Journal on Numerical Analysis*, 7:358, 1970.

[39] M. Z. NASHED. Continuous and semicontinuous analogues of iterative methods of Cimmino and Kaczmarz with applications to the inverse Radon transform. *Proceedings of the Oberwolfach Conference on Mathematical Aspects of Computerized Tomography, Lecture Notes in Medical Informatics*, 8:160–178, 1981.

[40] G. PIERRA. Decomposition through formalization in a product space. *Math. Prog.*, 28:96–118, 1984.

[41] M. RAYDAN. On the Barzilai and Borwein choice of the steplength for the gradient method. *IMA Journal on Numerical Analysis*, 13:321–326, 1993.

[42] K. T. SMITH, D. C. SOLMON, and S. L. WAGNER. Practical and mathematical aspects of the problem of reconstructing objects from radiographs. *Bull. Amer. Math. Soc.*, 83:1227–1270, 1977.

[43] J. VON NEUMANN. Functional operators vol. II. The geometry of orthogonal spaces. *Annals of Math. Studies*, 22, 1950. Princeton University Press. This is a reprint of mimeographed lecture notes first distributed in 1933.

Matrix			Cimmino-DFSANE		Kaczmarz		Cimmino	
Name	Dimension	m	iter	time	iter	time	iter	time
Hilbert	5	2	8	0.01	407	0.28	1576	1.03
Hilbert	5	3	32	0.04	*	*	*	*
Hilbert	5	4	264	0.25	*	*	*	*
Hilbert	10	2	8	0.01	304	0.17	1176	0.59
Hilbert	10	3	29	0.03	*	*	*	*
Hilbert	20	2	6	0.01	256	0.22	991	0.41
Hilbert	20	3	23	0.03	*	*	*	*
bcspwr02	49	20	37	0.26	37	0.31	1661	11.89
bcspwr02	49	40	337	4.59	1413	19.70	*	*
bcspwr02	49	49	3668	63.33	*	*	*	*
bcspwr03	118	10	49	0.21	63	0.22	1086	3.93
bcspwr03	118	20	66	0.51	101	0.70	4240	30.05
bcspwr03	118	30	103	1.13	185	2.06	*	*
bcspwr03	118	50	165	2.97	264	4.84	*	*
bcspwr03	118	100	585	20.83	4009	146.42	*	*
bcspwr03	118	118	3834	166.28	*	*	*	*
bcspwr04	274	10	16	0.10	12	0.19	253	1.11
bcspwr04	274	20	34	0.29	38	0.31	1652	12.12
bcspwr04	274	30	49	0.67	59	0.69	3103	34.93
bcspwr04	274	50	545	10.54	*	*	*	*
bcspwr04	274	274	*	*	*	*	*	*
mcca	180	10	29	0.16	60	0.30	1235	4.31
mcca	180	20	37	0.28	42	0.33	1674	11.53
mcca	180	50	40	0.93	40	0.61	4794	87.45
mcca	180	100	298	11.98	1323	51.23	*	*
mcca	180	180	307	22.36	1101	81.18	*	*
mcfe	765	10	10	0.04	5	0.05	157	0.73
mcfe	765	20	137	1.13	565	4.63	*	*
mcfe	765	50	164	3.46	502	10.42	*	*
mcfe	765	100	152	6.64	412	17.71	*	*
mcfe	765	200	125	11.75	246	22.74	*	*
mcfe	765	500	383	114.23	*	*	*	*
mcfe	765	765	404	210.65	*	*	*	*
Sherman2	1080	10	634	2.83	3100	15.38	*	*
Sherman2	1080	20	410	3.53	2955	28.33	*	*
Sherman2	1080	30	3595	46.75	*	*	*	*
bcsstk01	48	10	39	0.20	41	0.12	966	3.67
bcsstk01	48	30	405	4.43	1137	12.82	*	*
bcsstk02	66	10	62	0.43	91	0.59	1641	7.71
bcsstk02	66	20	119	1.07	337	3.12	*	*
bcsstk02	66	30	231	3.24	1254	18.22	*	*
bcsstk02	66	50	520	13.62	4306	121.01	*	*
bcsstk03	112	10	1217	4.57	*	*	*	*
bcsstk03	112	20	*	*	*	*	*	*
bcsstk04	132	10	15	0.12	10	0.11	450	3.67
bcsstk04	132	30	54	0.66	131	1.61	6470	81.78
bcsstk04	132	50	157	4.04	2018	44.20	*	*
bcsstk04	132	100	1088	48.82	*	*	*	*

Table 3: Comparison between different methods for linear varieties