Linear Algebra and its Applications 509 (2016) 191–205

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Projection on the intersection of convex sets

Marko Stošić [a,b,*], João Xavier [c], Marija Dodig [d,b]

[a] CAMGSD, Departamento de Matemática, Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
[b] Mathematical Institute SANU, Knez Mihajlova 36, 11000 Belgrade, Serbia
[c] Institute for Systems and Robotics, Instituto Superior Técnico, University of Lisbon, Lisbon 1600-011, Portugal
[d] CEAFEL, Departamento de Matématica, Universidade de Lisboa, Edificio C6, Campo Grande, 1749-016 Lisbon, Portugal

A R T I C L E I N F O

Article history:
Received 30 October 2015
Accepted 25 July 2016
Available online 28 July 2016
Submitted by B. Lemmens

MSC:
90C53
15A09

Keywords:
Projections
Semi-smooth Newton algorithm
Generalized Jacobian

A B S T R A C T

In this paper, we give a solution of the problem of projecting a point onto the intersection of several closed convex sets, when a projection on each individual convex set is known. The existing solution methods for this problem are sequential in nature. Here, we propose a highly parallelizable method. The key idea in our approach is the reformulation of the original problem as a system of semi-smooth equations. The benefits of the proposed reformulation are twofold: (a) a fast semi-smooth Newton iterative technique based on Clarke's generalized gradients becomes applicable and (b) the mechanics of the iterative technique is such that an almost decentralized solution method emerges. We proved that the corresponding semi-smooth Newton algorithm converges near the optimal point (quadratically). These features make the overall method attractive for distributed computing platforms, e.g. sensor networks.

* Corresponding author at: CAMGSD, Departamento de Matemática, Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal.
E-mail address: mstosic@isr.ist.utl.pt (M. Stošić).

http://dx.doi.org/10.1016/j.laa.2016.07.023
0024-3795/

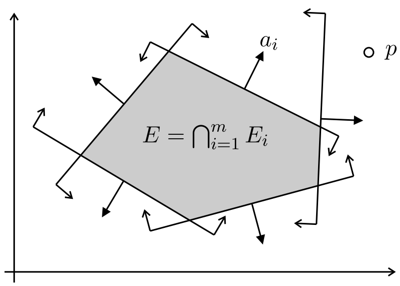
1. Introduction

Let E_i, $i = 1, \ldots, n$, be convex, closed subsets of $\mathbb{R}^m$. Suppose that $E = \bigcap_{i=1}^{n} E_i$ is non-empty. Let $p \in \mathbb{R}^m$ be a point.

By $P_{E_i}(p)$ we denote a projection of a point p to a set E_i. It is well-known that the projection of p on the set E exists and is unique. Thus, we have the following problem:

Problem 1. Find a projection of the point p on the set E.

We assume that the projections of a point p, $P_{E_i}(p)$ on each individual set E_i, $i = 1, \ldots, n$, are well known and easy to handle. However, the projection of a point p on the intersection E $(P_E(p))$ is very hard to compute:



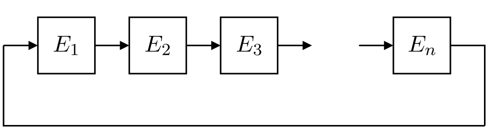
Problem 1 can be alternatively written as the following optimization problem:

$$\underset{\text{subject to } x \in E}{\text{minimize}} \; ||x - p||^2 \tag{1}$$

Problem 1 finds many applications in practice, e.g. in medical imaging, computerized tomography, stat fusion architecture, solving convex problems with strong duality, etc. – see [3] and the references therein.

There already exist many algorithms for resolving this problem. All of them are based on alternating or cyclic projection onto each set E_i. Von Neumann [10] studied the special case $n = 2$, where each E_i is affine subspace, and Halperin [8] analyzed the case $n > 2$. See [4] for more exhaustive results on the affine case. The non-affine case was considered in [6,9]. By reinterpreting the former cyclic projection methods in a suitable Cartesian product space one can obtain iterative simultaneous projection methods [11].

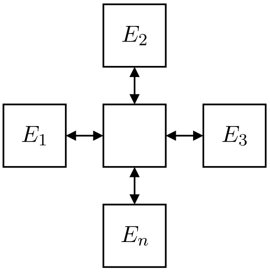
In all existing methods the convergence rate is linear, and they can be presented by the following scheme:



In this paper, we address Problem 1 for the general case: any finite number n of closed convex sets, and any closed convex sets E_i (i.e. not necessarily affine). Our approach consists in reformulating the optimization Problem 1 as a system of nonsmooth equations, which is then tackled by a semi-smooth Newton iterative method. We prove that in the generic case this algorithm converges near the optimal point. A semi-smooth Newton method for computation in certain projections related problems was also used in [1,2,7].

The main advantage is that, the convergence rate of the resulting Newton method can be super-linear. Also, the structure of our approach is such that an almost decentralized solution method emerges.

Our algorithm is particularly suitable for sensor networks, where the network communication is costly, and in our novel method the number of iterations is much smaller. In fact, our method can be presented by the following scheme:



Hence, we give an algorithm that has quadratic convergence rate in the number of iterations, and that performs better than all the existing ones.

2. Problem reformulation

Denote by x the projection of a point p on the set E, i.e.

$$P_E(p) = x.$$

In other words, x is a solution of the optimization problem (1).

Let $N_{E_i}(x)$ be the normal cone of E_i at the point x:

$$N_{E_i}(x) = \{s \in \mathbb{R}^m | \langle s, y - x\rangle \le 0, \quad \text{for all } y \in E_i\}.$$

Here, $\langle x, y\rangle$ denotes the inner-product of the vectors $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$ and $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$, i.e. $\langle x, y\rangle = \sum_{i=1}^m x_i y_i$. Suppose that the following constraint qualification holds: $\bigcap_{i=1}^n \operatorname{int}(E_i) \neq \emptyset$, where $\operatorname{int}(E_i)$ denotes the relative interior of the convex set $E_i \subset \mathbb{R}^m$. Then we have

$$N_E(x) = N_{E_1}(x) + \cdots + N_{E_n}(x),$$

and the optimality condition in (1) can be written as:

$$x + s_1 + s_2 + \ldots + s_n = p,$$
$$s_i \in N_{E_i}(x), \quad i = 1, \ldots, n,$$

or equivalently

$$x + s_1 + s_2 + \ldots + s_n = p, \tag{2}$$
$$P_{E_i}(x + s_i) = x, \quad i = 1, \ldots, n. \tag{3}$$

All variables x, s_1, $s_2, \ldots, s_n$ are from $\mathbb{R}^m$.

We eliminate x in (2) and (3), and set

$$z_i := \sum\nolimits_{j \neq i} s_j, \quad i = 1, \ldots, n.$$

The Problem 1 then becomes to resolve the following system of equations:

$$\begin{cases} P_{E_1}(p - z_1) + \frac{1}{n-1} \sum_{i=1}^n z_i = p, \\ P_{E_2}(p - z_2) + \frac{1}{n-1} \sum_{i=1}^n z_i = p, \\ \qquad \vdots \\ P_{E_n}(p - z_n) + \frac{1}{n-1} \sum_{i=1}^n z_i = p. \end{cases} \tag{4}$$

The solution $z = (z_1, z_2, \ldots, z_n) \in \mathbb{R}^m \times \mathbb{R}^m \times \cdots \times \mathbb{R}^m$, of (4) gives the wanted projection by:

$$x = p - \frac{1}{n-1} \sum_{i=1}^n z_i. \tag{5}$$

The problem (4) can be written as

$$F(z) = 0, \tag{6}$$

where $F = (F_1, \ldots, F_n)$, and $F_i : (\mathbb{R}^m)^{\times n} \to \mathbb{R}^m$, $i = 1, \ldots, n$, are given by

$$F_i(z) = P_{E_i}(p - z_i) + \frac{1}{n-1} \sum_{j=1}^n z_j - p, \qquad i = 1, \ldots, n.$$

Thus, in order to solve (1), it is enough to resolve the system (6), which then by (5), gives the wanted projection x.

3. Semi-smooth Newton algorithm

We shall solve the system (6) by using Newton-like algorithm.

In particular, if F was smooth, we could use the iterative Newton method

$$z^{(k+1)} = z^{(k)} - (DF(z^{(k)}))^{-1}F(z^{(k)}),$$

to solve (6), where $DF(z)$ denotes the derivative of F at the point z.

However, in general F is nonsmooth due to the presence of the projectors P_{E_i}. But it is semismooth, and there is an analogous Newton-like method for this case.

Definition 1. [5] A function $F : \mathbb{R}^N \to \mathbb{R}^N$ is semismooth at x if F is locally Lipschitzian at x, i.e. if

$$\|F(y) - F(z)\| \leq L\,\|y - z\| \qquad \text{for } y, z \text{ around } x,$$

for some $L > 0$, and if for any $d^\star \in \mathbb{R}^N$ the following limit exists:

$$\lim_{d \to d^\star, t \downarrow 0} \nabla F(x + td)d.$$

The second condition implies that there also exists the derivative

$$F'(x; d^\star) = \lim_{t \downarrow 0} \frac{F(x + td^\star) - F(x)}{t}.$$

The generalization of the differential for semismooth functions is given by the Clarke's generalized Jacobian [5]:

$$\partial F(x) = \text{convex hull} \left\{ \lim_{x_i \to x,\, DF(x_i) \text{ exists}} DF(x_i) \right\}.$$

In particular, Clarke's generalized Jacobian is a multi-valued function.

Example 1. Let $P : \mathbb{R} \to \mathbb{R}$, be a projection on the positive real axis $\mathbb{R}_+$, i.e.

$$P(x) = \begin{cases} x & , x \geq 0, \\ 0 & , x < 0 \end{cases}$$

This function is not differentiable at $x = 0$. However it is semismooth and its generalized Jacobian is given by:

$$\begin{aligned} \partial P(x) &= \{1\}, \quad x > 0, \\ \partial P(x) &= \{0\}, \quad x < 0, \\ \partial P(0) &= [0, 1]. \qquad \square \end{aligned}$$

In [12], the following iterative method for solving $F(z) = 0$ when F is semismooth is given:

$$z^{(k+1)} = z^{(k)} - \left(\partial F(z^{(k)})\right)^{-1} F\left(z^{(k)}\right). \tag{7}$$

Here, the Clarke's generalized derivative $\partial F(z^{(k)})$ is involved. Note that in general this derivative is multi-valued, and one can put any value in (7).

This algorithm converges quadratically, if $\partial F(z^*)$ is invertible at the optimal point z^*.

In our case:

$$\begin{bmatrix} z_1^{(k+1)} \\ z_2^{(k+1)} \\ \vdots \\ z_n^{(k+1)} \end{bmatrix} = \begin{bmatrix} z_1^{(k)} \\ z_2^{(k)} \\ \vdots \\ z_n^{(k)} \end{bmatrix} + \left(\partial F(z^{(k)})\right)^{-1} \begin{bmatrix} p - P_{E_1}\left(p - z_1^{(k)}\right) - \frac{1}{n-1}\sum_j z_j^{(k)} \\ p - P_{E_2}\left(p - z_2^{(k)}\right) - \frac{1}{n-1}\sum_j z_j^{(k)} \\ \vdots \\ p - P_{E_n}\left(p - z_n^{(k)}\right) - \frac{1}{n-1}\sum_j z_j^{(k)} \end{bmatrix} \tag{8}$$

where

$$\partial F(z^{(k)}) = \begin{bmatrix} -\partial P_{E_1}\left(p - z_1^{(k)}\right) & & \\ & \ddots & \\ & & -\partial P_{E_n}\left(p - z_n^{(k)}\right) \end{bmatrix} + \frac{1}{n-1} L L^T, \tag{9}$$

with $L = \underbrace{[I_m \cdots I_m]}_{n}^T$, where I_m is the $m \times m$ identity matrix.

In Theorem 2 below we prove that the function F from our main problem (6) satisfies the condition for the convergence of the semismooth Newton method, which thus can be applied to solve it.

4. Main theorem

In order to apply Newton-like algorithm given in (7), we need to prove that $\partial F(z^*)$ is invertible at the optimal point z^*.

Due to technical reasons we assume that all sets E_i, $i = 1, \ldots, n$, have smooth boundary.

Let $z^* = (z_1^*, \ldots, z_n^*)$. Let $x^* \in E = \bigcap_{i=1}^n E_i$, and let $s_i \in N_{E_i}(x^*)$, $i = 1, \ldots, n$. As in Section 2, we have

$$x^* + s_i = p - z_i^*, \quad i = 1, \ldots, n.$$

Condition 1. *Let $v_1, \ldots, v_n$ be vectors from $\mathbb{R}^m$. If*

$$v_1 + \cdots + v_n = 0$$

and

$$v_i \in \operatorname{span} N_{E_i}(x^*),$$

then

$$v_1 = \cdots = v_n = 0.$$

We note that in the generic case, the Condition 1 holds.

Theorem 2. *Let the Condition 1 be valid. Then the matrix*

$$\begin{bmatrix} -\partial P_{E_1}(x^* + s_1) & & \\ & \ddots & \\ & & -\partial P_{E_n}(x^* + s_n) \end{bmatrix} + \frac{1}{n-1} \begin{bmatrix} I_m & I_m & \dots & I_m \\ I_m & I_m & \dots & I_m \\ \vdots & \vdots & \ddots & \vdots \\ I_m & I_m & \dots & I_m \end{bmatrix}$$

is invertible.

Before giving the proof of Theorem 2, we shall give two auxiliary lemmas.

4.1. Auxiliary lemmas

Lemma 1. *Let P denote a projection on a convex closed set $\mathcal{C} \subset \mathbb{R}^m$, and let $b \in \mathbb{R}^m$ be an arbitrary point. Let $M = \partial P(b)$. Then*

$$x^T M x \geq ||Mx||^2, \qquad \textit{for all} \quad x \in \mathbb{R}^m. \tag{10}$$

Proof. If $b \in \operatorname{int} \mathcal{C}$, then $M = I_m$, and thus (10) is trivially valid.

If $b \notin C$, then we have the following situation:

By the definition, we have

$$Mx = \frac{d}{dt}(P(b + tx)) \mid_{t=0} = \lim_{t \to 0} \frac{P(b + tx) - P(b)}{t}.$$

Thus, since $x = \lim_{t \to 0} \frac{(b+tx)-b}{t}$, in order to prove the lemma, it is enough to prove the following:

$$((b + tx) - b)^T (P(b + tx) - P(b)) \geq ||P(b + tx) - P(b)||^2. \tag{11}$$

Now, let O be the origin of $\mathbb{R}^m$. Let A, B, C and D, be points in $\mathbb{R}^m$, such that $\overrightarrow{OC} = b+tx$, $\overrightarrow{OB} = b$, $\overrightarrow{OD} = P(b+tx)$, and $\overrightarrow{OA} = P(b)$. Let $s = \overrightarrow{BC} = \overrightarrow{OC} - \overrightarrow{OB} = (b+tx) - b$, and $w = \overrightarrow{AD} = \overrightarrow{OD} - \overrightarrow{OA} = P(b + tx) - P(b)$.

Then, (11) is equivalent to:

$$s^T w \geq ||w||^2. \tag{12}$$

Finally, (12) is true since $\mathcal{C}$ is convex and so we have:

$$\langle \overrightarrow{AB}, \overrightarrow{AD} \rangle \leq 0 \Rightarrow \langle \overrightarrow{BA}, \overrightarrow{AD} \rangle \geq 0,$$
$$\langle \overrightarrow{DC}, \overrightarrow{DA} \rangle \leq 0 \Rightarrow \langle \overrightarrow{DC}, \overrightarrow{AD} \rangle \geq 0.$$

Therefore

$$s^T w = \langle \overrightarrow{BA} + \overrightarrow{AD} + \overrightarrow{DC}, \overrightarrow{AD} \rangle \geq \langle \overrightarrow{AD}, \overrightarrow{AD} \rangle = ||w||^2,$$

as wanted.

Finally, we are left with the case $b \in \partial\mathcal{C}$. Then any matrix M from the set $\partial P(b)$ is, by definition, a convex combination of matrices, that we have already shown above to satisfy (10). Therefore, $M = \sum_{i=1}^k c_i M_i$, for some matrices M_i that satisfy (10) and numbers $c_i \geq 0$, $i = 1, \ldots, k$, with $\sum_{i=1}^k c_i = 1$. Then we have:

$$x^T M x = \sum_{i=1}^k c_i x^T M_i x \geq \sum_{i=1}^k c_i ||M_i x||^2 \geq ||\sum_{i=1}^k c_i M_i x||^2 = ||Mx||^2,$$

where the second inequality holds because the function $f(x) = ||x||^2$ is convex. □

Lemma 2. *Let P be a projection on a convex closed set $\mathcal{C} \subset \mathbb{R}^m$ with a smooth boundary, and let $b \in \mathbb{R}^m$ be an arbitrary point. Let $M = \partial P(b)$. Then*

$$Mx = 0 \Rightarrow x \in \operatorname{span} N_{\mathcal{C}}(P(b)). \tag{13}$$

Proof. If $b \in \operatorname{int} \mathcal{C}$, (13) trivially holds since $M = I_m$ and $N_{\mathcal{C}}(P(b)) = \{0\}$.

The most important case is when $b \notin \mathcal{C}$. In that case, we choose coordinates in $\mathbb{R}^m$ such that $P(b)$ is at the origin, such that the tangent space $T_{\mathcal{C}}(P(b))$ is $\mathbb{R}^{m-1} \times \{0\}$, and such that $\mathcal{C}$ is in the upper half space. Then locally near the origin $\partial\mathcal{C}$ consists of the points of the form

$$(y_1, y_2, \ldots, y_{m-1}, f(y_1, \ldots, y_{m-1})),$$

for some $f : \mathbb{R}^{m-1} \to \mathbb{R}$. Also, $N_{\mathcal{C}}(P(b))$ is the negative y_m-axis, and b is mapped to a point $(0, \ldots, 0, -d)$, for some $d \geq 0$.

Now, let $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$ be arbitrary. Since $M = \partial P(b)$, we have

$$Mx = \partial P(b)x = \frac{d}{dt} P(b + tx) \mid_{t=0} = \frac{d}{dt} (P(tx_1, \ldots, tx_{m-1}, -d + tx_m)) \mid_{t=0} .$$

Now, $P(tx_1, \ldots, tx_{m-1}, -d + tx_m) = (y_1(t), y_2(t), \ldots, y_{m-1}(t), f(y_1(t), \ldots, y_{m-1}(t)))$ is a curve on $\partial\mathcal{C}$ that tends to $P(b) = 0$ as $t \to 0$. Thus,

$$Mx = (\dot{y}_1(0), \ldots, \dot{y}_{m-1}(0), 0).$$

Also, if we denote by $A_t = (tx_1, \ldots, tx_{m-1}, -d + tx_m)$, we have that $A_t - P(A_t)$ is orthogonal to the tangent space $T_{\mathcal{C}}P(A_t)$, i.e. to all $m-1$ vectors $(0, \ldots, 1, \ldots, 0, f_i(y_1(t), \ldots, y_{m-1}(t)))$, $i = 1, \ldots, m-1$, where 1 is on the ith position, and $f_i = \frac{\partial f}{\partial y_i}$.

Thus, for all $i = 1, \ldots, m-1$, we have that

$$y_i(t) - x_i t + f_i(y_1(t), \ldots, y_{m-1}(t))(f(y_1(t), \ldots, y_{m-1}(t)) + d - x_m t) = 0.$$

By taking derivative $\frac{d}{dt}$, and setting $t = 0$, and by using $f(0) = f_i(0) = 0$ (since the tangent space of $\mathcal{C}$ at the origin is $\mathbb{R}^{m-1} \times \{0\}$), we have:

$$\dot{y}_i(0) - x_i + d \sum_{j=1}^{m-1} f_{ij}(0)\dot{y}_j(0) = 0, \quad i = 1, \ldots, m-1.$$

Here, $\dot{y}_i = \frac{d}{dt}(y_i)$, and $f_{ij}(0) = \frac{\partial^2 f}{\partial y_i \partial y_j}(0)$. In other words:

$$(I_{m-1} + dH_f(0)) \begin{pmatrix} \dot{y}_1(0) \\ \dot{y}_2(0) \\ \vdots \\ \dot{y}_{m-1}(0) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{m-1} \end{pmatrix}, \tag{14}$$

where $H_f(0) = \left[\frac{\partial^2 f}{\partial y_i \partial y_j}(0)\right]_{i,j=1}^{m-1}$ is the Hessian of f at the origin.

Now, since $\mathcal{C}$ is convex, by assumption $f \geq 0$, and since $f(0) = f_i(0) = 0$, we have that the Hessian $H_f(0)$ is positive semi-definite. Since $d \geq 0$, we have that the matrix $I_{m-1} + dH_f(0)$ is invertible, and so

$$\begin{pmatrix} \dot{y}_1(0) \\ \dot{y}_2(0) \\ \vdots \\ \dot{y}_{m-1}(0) \end{pmatrix} = (I_{m-1} + dH_f(0))^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{m-1} \end{pmatrix}.$$

Going back to the definition we have

$$Mx = \frac{d}{dt}(y_1(t), y_2(t), \ldots, y_{m-1}(t), f(y_1(t), \ldots, y_{m-1}(t))) \,|_{t=0} =$$

$$= (\dot{y}_1(0), \dot{y}_2(0), \ldots, \dot{y}_{m-1}(0), 0) = \begin{bmatrix} (I_{m-1} + dH_f(0))^{-1} & 0 \\ 0 & 0 \end{bmatrix} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{m-1} \\ x_m \end{bmatrix}}_{=x}.$$

So, we have $Mx = 0 \Leftrightarrow x_1 = \cdots = x_{m-1} = 0 \Leftrightarrow x \in \operatorname{span} N_{\mathcal{C}}(0)$, as wanted.

Finally, let $b \in \partial\mathcal{C}$. Then any matrix $M \in \partial P(b)$ is a convex combination of the identity matrix I_m and the matrix $\begin{bmatrix} I_{m-1} & 0 \\ 0 & 0 \end{bmatrix}$ (since $d = 0$ in this case). Therefore, in the same coordinates as above, any such matrix can be written in the following form for some $0 \le c \le 1$:

$$Mx = \left(cI_m + (1-c) \begin{bmatrix} I_{m-1} & 0 \\ 0 & 0 \end{bmatrix} \right) x = \begin{bmatrix} I_{m-1} & 0 \\ 0 & c \end{bmatrix} x,$$

and so again $Mx = 0 \Rightarrow x \in \operatorname{span} N_{\mathcal{C}}(0)$. $\square$

Now, we can give a proof of Theorem 2:

Proof of Theorem 2. Let $M_i = \partial P_{E_i}(x^* + s_i)$, $i = 1, \dots, n$. Note that $x^* \in \operatorname{int} E_i \Rightarrow M_i = I_m$.

Denote by M the matrix in Theorem 2, i.e.

$$M = \begin{bmatrix} -M_1 & & \\ & \ddots & \\ & & -M_n \end{bmatrix} + \frac{1}{n-1} \begin{bmatrix} I_m & I_m & \dots & I_m \\ I_m & I_m & \dots & I_m \\ \vdots & \vdots & \ddots & \vdots \\ I_m & I_m & \dots & I_m \end{bmatrix}.$$

It is well known, that the matrix M is invertible if and only if the following is true: If $v_1, \dots, v_n \in \mathbb{R}^m$, are such that

$$M \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = 0, \tag{15}$$

then

$$v_i = 0, \text{ for all } i = 1, \dots, n. \tag{16}$$

Thus, we are left with proving that (15) implies (16).

Equation (15) is equivalent to

$$M_i v_i = \frac{1}{n-1}(v_1 + \dots + v_n), \quad i = 1, \dots, n,$$

i.e. there exists $w \in \mathbb{R}^m$, such that

$$v_1 + \dots + v_n = (n-1)w \tag{17}$$

$$M_i v_i = w, \quad i = 1, \dots, n. \tag{18}$$

Relation (17) can be written as:

$$(v_1 - M_1 v_1) + (v_2 - M_2 v_2) + \cdots + (v_n - M_n v_n) = -w. \tag{19}$$

After taking scalar product with $w = M_i v_i$ in (19), we obtain

$$-||w||^2 = -w^T w = w^T \sum_{i=1}^{n} (v_i - M_i v_i) = \sum_{i=1}^{n} (M_i v_i)^T (v_i - M_i v_i).$$

Since ${v_i}^T M_i v_i \geq ||M_i v_i||^2$ is equivalent to $(v_i - M_i v_i)^T M_i v_i \geq 0$, we have that Lemma 1 gives:

$$(M_i v_i)^T (v_i - M_i v_i) \geq 0, \quad i = 1, \ldots, n.$$

Thus, we have

$$-||w||^2 = \sum_{i=1}^{n} (M_i v_i)^T (v_i - M_i v_i) \geq 0,$$

i.e.

$$w = 0.$$

Finally, the last with (17) and (18) gives

$$(*) \quad \begin{cases} v_1 + \cdots + v_n = 0 \\ M_i v_i = 0, \quad i = 1, \ldots, n. \end{cases}$$

Now, by Lemma 2, we have that $M_i v_i = 0$ if and only if $v_i \in \operatorname{span} N_{E_i}(x^*)$.

Finally, the Condition 1 from our theorem implies that $(*)$ is satisfied only for zero vectors, i.e. $v_i = \cdots = v_n = 0$, as wanted. □

Let F be as in (6). Then Theorem 2 gives our main result:

Theorem 3. *$\partial F(z^*)$ is invertible at the optimal point z^*.* □

Example 2. For $i = 1, \ldots, n$, let $E_i = \{x \in \mathbb{R}^m | a_i^T x \leq b_i\}$, with $a_i \in \mathbb{R}^m$, $|a_i| = 1$, $b_i \in \mathbb{R}$, be half-spaces in $\mathbb{R}^m$. Then $E = \cap_{i=1}^{n} E_i$ is a polyhedron.

The projection of a point $z \in \mathbb{R}^m$ on E_i is given by:

$$P_{E_i}(z) = \begin{cases} (I_m - a_i a_i^T) z + b_i a_i, & z \notin E_i \\ z, & z \in E_i. \end{cases} \tag{20}$$

Hence, the Clarke's generalized Jacobian of p_{E_i} is given by:

$$\partial P_{E_i}(z) = \begin{cases} I_m - a_i a_i^T, & z \notin E_i \\ I_m, & z \in \operatorname{int} E_i \\ \{I_m - k a_i a_i^T \mid k \in [0,1]\}, & z \in \partial E_i. \end{cases} \tag{21}$$

In particular, it is straightforward to see that $\partial P_{E_i}(z)$, for $z \notin E_i$, satisfies Lemmas 1 and 2.

5. Fast computation of $(\partial F(z^{(k)}))^{-1}$

In the previous section we have proved that the $\partial F(z^*)$ is invertible at the optimal point z^*, and that therefore we can use the semi-smooth Newton algorithm for computation of the projection point. The computation is now done in parallel and the number of iterations is drastically reduced comparing to existing algorithms. However, there is a time-consuming operation involved in this algorithm – the computation of the inverse of the matrix $\partial F(z^{(k)})$ in (8). This is a matrix of the size $mn \times mn$, and for large m and n it can be slow to compute its inverse. However, by exploring the special form of this matrix we can make the computations much faster by performing only inverses of $m \times m$ matrices.

As it can be seen from the form of the matrix (9), it is a sum of a block-diagonal matrix and a low-rank matrix. In the block-diagonal matrix there are n diagonal blocks of size $m \times m$, but they are not all invertible.

To resolve this we do the following: for every $i = 1, \dots, n$, denote by M_i, the i-th block, i.e.

$$M_i = -\partial P_{E_i}(p - z_i^{(k)}), \quad \text{for} \quad i = 1, \dots, n.$$

Now, we shall define certain auxiliary numbers c_i and vectors w_i, such that the matrices $\overline{M}_i$, given by $\overline{M}_i = M_i - c_i w_i w_i^T$ are invertible.

If $p - z_i^{(k)} \in E_i$ then we have that $M_i = -I_m$. In this case we also set the auxiliary number and vector $c_i := 0$ and $w_i := 0$, and so $\overline{M}_i = -I_m$.

But if $p - z_i^{(k)} \notin E_i$, then as we have shown in Lemma 2, the matrix M_i is not invertible, but rather $\operatorname{rank} M_i = m - 1$. However, as shown in the proof of Lemma 2, the matrix M_i acts in the hyperspace orthogonal to the line going from $p - z_i^{(k)}$ to its projection on E_i. Therefore, if we define the matrix

$$\overline{M}_i = M_i - c_i w_i w_i^T, \tag{22}$$

where

$$w_i = \frac{p - z_i^{(k)} - P_{E_i}(p - z_i^{(k)})}{||p - z_i^{(k)} - P_{E_i}(p - z_i^{(k)})||} \in \mathbb{R}^{m \times 1}.$$

Then $\overline{M}_i$ is invertible for any $c_i > 0$.

Now, for any $i = 1, \dots, n$, let $\bar{w}_i \in \mathbb{R}^{mn\times 1}$ be given by

$$\bar{w}_i = [\underbrace{0_m^T \ 0_m^T \ \dots \ 0_m^T}_{i-1} \ w_i^T \ 0_{(n-i)m}^T]^T,$$

where $0_m \in \mathbb{R}^{m\times 1}$ is the column vector of length m consisting only of zeros. Then, the matrix $\partial F(z^{(k)})$ can be written as

$$\partial F(z^{(k)}) = \begin{bmatrix} \overline{M}_1 & & \\ & \ddots & \\ & & \overline{M}_n \end{bmatrix} + \frac{1}{n-1} LL^T + \sum_{i=1}^{n} c_i \bar{w}_i \bar{w}_i^T. \tag{23}$$

Denote by

$$D := \operatorname{diag}(\overline{M}_1, \dots, \overline{M}_n).$$

Since all $\overline{M}_i$, $i = 1, \dots, n$, are $m \times m$ invertible matrices, so is D, and $D^{-1} = \operatorname{diag}(\overline{M}_1^{-1}, \dots, \overline{M}_n^{-1})$.

As we can see from (23), the matrix $\partial F(z^{(k)})$ is obtained by perturbing D by two low-rank matrices, $\frac{1}{n-1} LL^T$ and $\sum_{i=1}^{n} c_i \bar{w}_i \bar{w}_i^T$.

Denote by

$$W := \operatorname{diag}(\overline{M}_1, \dots, \overline{M}_n) + \frac{1}{n-1} LL^T.$$

The matrix LL^T is of rank m, and we can compute the inverse of the matrix W by using the Woodbury matrix identity:

Theorem 4 *((Sherman–Morrison–)Woodbury matrix identity). Suppose A is an invertible square $n \times n$ matrix and U is $n \times k$, C is $k \times k$ and V is $k \times n$ matrix. Then*

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}. \qquad \square \tag{24}$$

In our case, we set $A = D$, $C = I_m$, $U = \frac{1}{\sqrt{n-1}} L$ and $V = U^T$. Since D and $D + \frac{1}{n-1} LL^T$ are invertible (note that $D + \frac{1}{n-1} LL^T$ is invertible for a generic c_i), by (24) we have

$$\left(D + \frac{1}{n-1} LL^T\right)^{-1} = D^{-1} - \frac{1}{n-1} D^{-1} L \left(I_m + \frac{1}{n-1} L^T D^{-1} L\right)^{-1} L^T D^{-1}.$$

Note that on the right-hand-side we have involved inverses only of matrices of sizes $m \times m$.

Finally, in order to get $(\partial F(z^{(k)}))^{-1}$, by (23) we are left with perturbing the matrix $D + \frac{1}{n-1} LL^T$ by the matrix $\sum_{i=1}^{n} c_i \bar{w}_i \bar{w}_i^T$ whose rank is at most n.

By using Woodbury's identity again, we obtain the wanted inverse. In this case we put $A = D + \frac{1}{n-1} LL^T$, $C = I_m$, $U = \begin{bmatrix} \sqrt{c_1}\bar{w}_1 & \cdots & \sqrt{c_n}\bar{w}_n \end{bmatrix}$, and $V = U^T$.

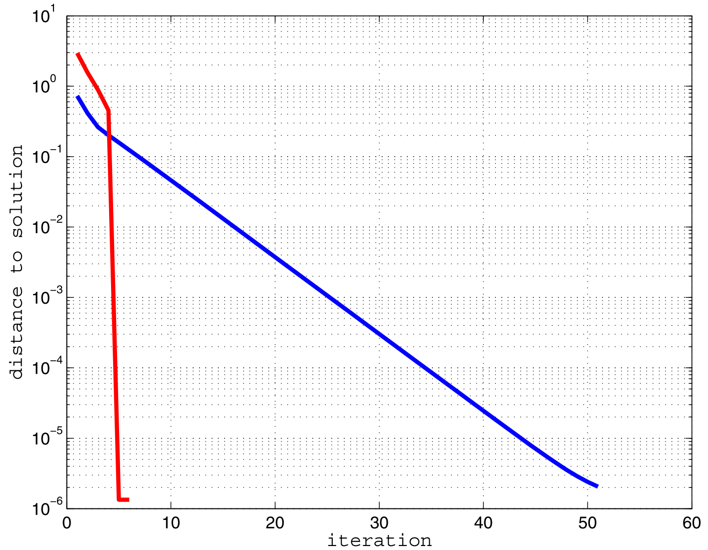


Fig. 1. Blue line represents the cyclical projections method, while red line represents the results of our semi-smooth Newton algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This algorithm significantly improves the calculations of the inverse of the matrix $\partial F(z^{(k)})$, making our approach faster and even more appealing for applications, especially in sensor network.

6. Applications

The method we described for computing the projections is very useful in applications, and particularly in the cases of distributed computations. Our method uses parallel, simultaneous computations of projections on each of particular convex sets. We can describe the setting by having n nodes, with i-th one knowing the projection on the set E_i, for each $i = 1, \ldots, n$. Then, in each iteration, one additional, central node sends $z_i^{(k)}$ to the i-th node. Then, all n nodes perform simultaneous computations: the i-th node computes $p_{E_i}\left(p - z_i^{(k)}\right)$ and $\partial p_{E_i}\left(p - z_i^{(k)}\right)$, and sends the answer to the central node. Then central node computes the updates $z_i^{(k+1)}$ according to the formulas from the Newton method, and sends them back to the nodes.

This parallel computation is particularly relevant in the case when communications between the nodes are costly. In addition the number of iterations is drastically reduced comparing to the standard methods.

Example 3. In the following example we computed the projecting of a point on the intersection of $n = 15$ affine half-spaces in $\mathbb{R}^8$. In Fig. 1 we plot the distance to the optimum as a function of number of iterations. The slow, linear, dependence is obtained for the case of cyclical projections, where as one iteration is counted the whole sequential cycle of projections onto each of the n sets. In contrast, in our approach the convergence is very fast and quadratic, and very few iterations are needed to get to the optimum.

6.1. Conclusions

We presented a semi-smooth Newton-type method for projecting a point on the intersection of arbitrary number of convex sets in the Euclidean space, when the projection on each of these convex sets is known. This gave a novel algorithm that explores parallel, simultaneous computation and is perfectly suited for distributed computation.

The strengths of the approach is that it works for arbitrary closed convex sets, that it significantly reduces the number of iterations needed to reach the optimal point, and also a particular form of the matrix from the Newton algorithm enables fast computation of its inverse. The parallel computation is particularly well-suited for the applications in distributed systems, like e.g. sensor networks.

Acknowledgements

This work was supported by Fundação para a Ciência e a Tecnologia (ISR/IST pluri-anual funding) through the POS Conhecimento Program that includes FEDER funds. This work was done within the activities of CEAFEL and was partially supported by FCT, project ISFL-1-1431. This work was also partially supported by the Ministry of Science, Technology and Development of Republic of Serbia, projects No. 174020 (M. Dodig) and No. 174012 (M. Stošić). The work of J. Xavier was supported by the FCT grants PTDC/EMS-CRO/2042/2012 and FCT [UID/EEA/5009/2013].

References

[1] J.G. Barrios, J.Y. Bello Cruz, O.P. Ferreira, S.Z. Németh, A semi-smooth Newton method for a special piecewise linear system with application to positively constrained convex quadratic programming, J. Comput. Appl. Math. 301 (2016) 91–100, http://dx.doi.org/10.1016/j.cam.2016.01.040.
[2] J.G. Barrios, O.P. Ferreira, S.Z. Németh, Projection onto simplicial cones by Picard's method, Linear Algebra Appl. 480 (1) (2015) 27–43.
[3] H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Rev. 38 (3) (1996) 367–426.
[4] H.H. Bauschke, J.M. Borwein, A.S. Lewis, The methods of cyclic projections for closed convex sets in Hilbert space, Contemp. Math. 204 (1997) 138.
[5] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[6] R.L. Dykstra, An algorithm for restricted least squares regression, J. Amer. Statist. Assoc. 78 (1983) 837–842.
[7] O.P. Ferreira, S.Z. Németh, Projection onto simplicial cones by a semi-smooth Newton method, Optim. Lett. 9 (4) (2015) 731–741.
[8] I. Halperin, The product of projection operators, Acta Sci. Math. 23 (1962) 96–99.
[9] S.P. Han, A successive projection method, Math. Program. 40 (1) (1988) 114.
[10] J. von Neumann, Functional Operators, vol. II, Princeton University Press, Princeton, 1950.
[11] G. Pierra, Decomposition through formalization in a product space, Math. Program. 28 (1) (1984) 96–115.
[12] L. Qi, J. Sun, A nonsmooth version of Newton's algorithm, Math. Program. 58 (1–3) (1993) 353–367.