Journal of Nonlinear and Convex Analysis
Volume 15, Number 2, 2014, 345–353

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

SUCCESSIVE LINEAR PROGRAMING APPROACH FOR SOLVING THE NONLINEAR SPLIT FEASIBILITY PROBLEM

A. GIBALI, K.-H. KÜFER, AND P. SÜSS

Dedicated to Professor Simeon Reich's 65th birthday

Abstract. The Split Feasibility Problem (SFP), which was introduced by Censor and Elfving, consists of finding a point in a set C in one space such that its image under a linear transformation belongs to another set Q in the other space. This problem was well studied both theoretically and practically as it was also used in practice in the area of Intensity-Modulated Radiation Therapy (IMRT) treatment planning. Recently Li et. al. extended the SFP to the non-linear framework. Their algorithm tries to follow the algorithm for the linear case. But, unlike the linear case, the involved proximity function is not necessarily convex. Therefore in order to use Baillon-Haddad and Dolidze Theorems, the authors assume convexity in order to prove convergence of the projected gradient method. Since convexity of the proximity function is too restrictive, we consider here a Successive Linear Programing (SLP) approach in order to obtain local optima for the non-convex case. We also aim to introduce a non-linear version of the Split Variational Inequality Problem (SVIP).

1. Introduction

In this paper we consider the *Nonlinear Split Feasibility Problem* (NLSFP) in Euclidean space and present an algorithm based on the Successive Linear Programing (SLP) approach in order to obtain convergence to a local solution. Our goal is to deal with a generalization of the NLSFP, which is the *Nonlinear Split Variational Inequality Problem* (NSVIP).

Let us first recall the *Split Feasibility Problem* (SFP). Given two non-empty, closed and convex sets, $C \subseteq \mathbb{R}^n$ and $Q \subseteq \mathbb{R}^m$ and an $m \times n$ matrix A. The SFP is formulated as follows:

$$\text{find a point } x^* \in C \text{ and } Ax^* \in Q. \tag{1.1}$$

This problem was first introduced by Censor and Elfving in [10]; it was later used in the area of Intensity-Modulated Radiation Therapy (IMRT) treatment planning; see [11, 9]. In case where there are multiple sets $C_i \subseteq \mathbb{R}^n$, $Q_j \subseteq \mathbb{R}^m$ for $i = 1, 2, \dots, p$ and $j = 1, 2, \dots, r$ this problem is called the *Multiple-Sets Split Feasibility Problem* (MSSFP), the approach will lead to a similar method and therefore our focus here will be in case where one set in each space is involved. Several generalizations of the MSSFP were introduced and studied in finite and infinite dimensional spaces.

2010 *Mathematics Subject Classification.* 47H05, 47H09, 47J20, 47J25.

Key words and phrases. Split feasibility problem, projected gradient, inverse strongly monotone, successive linear programing, variational inequality, trust-region method.

This work was partially funded by the Fraunhofer Institute for Industrial Mathematics (ITWM), Kaiserslautern, Germany. The first author was also supported by the Minerva Foundation.

For example, see Masad and Reich [24], Censor, Gibali and Reich [12], and Byrne, Censor, Gibali and Reich [6] and the references therein.

In a recent paper, Li et. al. [21] introduced the *Nonlinear Multiple-Sets Split Feasibility Problem* (NLMSSFP), where a general vector valued function $F:\mathbb{R}^n\rightarrow\mathbb{R}^m$ is taken instead of A. Again, for simplicity we consider the Nonlinear Multiple-Sets Split Feasibility Problem with two sets in each space, that is

$$\text{find a point } x^*\in C \text{ and } F(x^*)\in Q. \tag{1.2}$$

In order to establish an algorithm for solving (1.2), the authors [21] followed a similar technique as in [11, 9]. This idea is based on the introduction of a proximity function

$$p(x)=\frac{1}{2}\left\|P_Q(F(x))-F(x)\right\|^2. \tag{1.3}$$

Then (1.2) is considered as the following minimization problem.

$$\min_{x\in C}1/2\left\|P_Q(F(x))-F(x)\right\|^2. \tag{1.4}$$

Assuming that F is continuously differentiable, we get that

$$\nabla p(x)=(\nabla F(x))^t\left(F(x)-P_Q(F(x))\right) \tag{1.5}$$

where $\nabla F(x):\mathbb{R}^m\rightarrow\mathbb{R}^n$ is the Jacobian of F at a point $x\in\mathbb{R}^n$ and t denotes its transpose. It was proven in [21, Lemma 2.4] that when F and ∇F are Lipschitz continuous the operator ∇p is also Lipschitz. In the linear case where $F=A$, it is shown in [9] that p is convex and therefore by the Baillon-Haddad Theorem [1, Corollaire 10] ∇p is inverse strongly monotone (ISM). Then, using the Dolidze Theorem [13] their projected gradient method (see Goldstein [16] and Levitin and Polyak [20]) converges to the solution of the SFP. The classical convergence result for the projected gradient method establishes that cluster points of (if any) are stationary points, i.e., they satisfy the first order optimality conditions, but in general neither existence nor uniqueness of cluster points is guaranteed. In case the objective function is convex a local optimum is also a global optimum.

For the non-linear case, i.e., NLSFP, convexity of p is not guaranteed and therefore iterates of the projected gradient method might converge to a stationary point which is only a local minimizer. Therefore, convexity of p is assumed in [21] and the algorithm converges to the solution of the NLSFP. Since this assumption is quite strong and might be too hard to verify in practice, we suggest here an algorithm which does not require convexity of p, therefore only local convergence is guaranteed. The method we use here ([27]) is based on the Successive Linear Programing Approach (see e.g., [3, Subsection 10.3] and [25]) also known as the projected gradient trust-region method [22].

Remark 1.1. Consider the *Nonlinear Multiple-Sets Split Feasibility Problem* (NMSSFP), which is formulated as follows. Given non-empty, closed and convex subsets $C_i\subseteq\mathbb{R}^n$, $Q_j\subseteq\mathbb{R}^m$ for $i=1,2,\ldots,p$ and $j=1,2,\ldots,r$, respectively, and a mapping $F:\mathbb{R}^n\rightarrow\mathbb{R}^m$ the NMSSFP is formulated as follows:

$$\text{find a point } x^*\in C:=\cap_{i=1}^{p}C_i \text{ such that } F(x^*)\in Q:=\cap_{i=1}^{r}Q_j. \tag{1.6}$$

Define the proximity function

$$\tilde{p}(x) = \frac{1}{2}\sum_{i=1}^{p}\alpha_i \|P_{C_i}(x) - x\|^2 + \frac{1}{2}\sum_{j=1}^{r}\beta_j \left\|P_{Q_j}(F(x)) - F(x)\right\|^2 \tag{1.7}$$

and then apply the suggested method with $\tilde{p}(x)$.

The paper is organized as follows. In Section 2 we list several known facts about functions, operators and mappings that we need in the sequel. In Section 3 we present Toint's [27] projected gradient trust-region method for solving (1.2). Finally, in Section 4 we discuss further research directions and propose a generalization of the NLSFP, which is the *Nonlinear Split Variational Inequality Problem* (NLSVIP).

2. Preliminaries

In this section we present several definitions and notations that play a central role in the sequel.

We recall several definitions and properties of operators.

Definition 2.1. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be an operator and let $D \subseteq \mathbb{R}^n$.

(i) f is called `α-inverse strongly monotone (α-ISM)` on D if

$$\langle f(x) - f(y), x - y\rangle \geq \alpha\|f(x) - f(y)\|^2 \text{ for all } x, y \in D \tag{2.1}$$

this property is also known as the `Dunn property` or `cocoercivity`.

(ii) f is called `Lipschitz continuous` on $D \subseteq \mathbb{R}^n$ if there exists a constant $\kappa > 0$ such that

$$\|f(x) - f(y)\| \leq \kappa\|x - y\| \text{ for all } x, y \in D. \tag{2.2}$$

(iii) f is called `monotone` on $D \subseteq \mathbb{R}^n$ if

$$\langle f(x) - f(y), x - y\rangle \geq 0 \text{ for all } x, y \in D. \tag{2.3}$$

Definition 2.2. Given an operator $f : \mathbb{R}^n \to \mathbb{R}^n$, denote by $\mathrm{Fix}(f)$ the fixed point set of f, i.e.,

$$\mathrm{Fix}(f) := \{x \in H \mid f(x) = x\}. \tag{2.4}$$

The following theorem is known as Baillon-Haddad Theorem [1, Corollary 10].

Theorem 2.3. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be convex, continuously differentiable on $\mathbb{R}^n$, and such that ∇h is β-Lipschitz continuous for some $\beta \in (0, \infty)$. Then ∇h is $1/\beta$-ISM.*

Let D be a non-empty, closed and convex subset of $\mathbb{R}^n$. For every point $x \in \mathbb{R}^n$, there exists a unique nearest point in D, denoted by $P_D(x)$. This point satisfies

$$\|x - P_D(x)\| \leq \|x - y\| \text{ for all } y \in D. \tag{2.5}$$

The mapping P_D is called the *metric projection* of $\mathbb{R}^n$ onto D. We know that P_D is a nonexpansive operator of $\mathbb{R}^n$ onto D, i.e.,

$$\|P_D(x) - P_D(y)\| \leq \|x - y\| \text{ for all } x, y \in \mathbb{R}^n. \tag{2.6}$$

The metric projection P_D is characterized by the following two properties (see, e.g., Goebel and Reich [15, Section 3]):

$$P_D(x) \in D \tag{2.7}$$

and

$$\langle x - P_D(x), P_D(x) - y\rangle \geq 0 \text{ for all } x \in \mathbb{R}^n,\ y \in D. \tag{2.8}$$

If D is a hyperplane, then (2.8) becomes an equality. It also follows that

$$\|x - y\|^2 \geq \|x - P_D(x)\|^2 + \|y - P_D(x)\|^2 \text{ for all } x \in \mathbb{R}^n,\ y \in D. \tag{2.9}$$

Definition 2.4 ([4, Proposition 2.1.2, p. 194])**.** Given a continuously differentiable function $h : \mathbb{R}^n \to \mathbb{R}$ and a set $D \subseteq \mathbb{R}^n$, a vector x^* that satisfies the condition

$$\langle \nabla h(x^*), x - x^*\rangle \geq 0 \text{ for all } x \in D, \tag{2.10}$$

is called a `stationary point` of h over D. Condition (2.10) is an `optimality condition`. It is a necessary condition for x^* to be a local minimum of h over D, and if h is convex over D then it is also sufficient (see, e.g., [4, Proposition 2.1.1, p. 193]).

Remark 2.5. Following (2.8), one can see that x^* is stationary point of h over D if and only if $x^* = P_D(x^* - t\nabla h(x^*))$ for all $t \geq 0$, that is $x^* \in \mathrm{Fix}(P_D(I - t\nabla h))$.

The theorem of Dolidze [13], as presented and proven in Byrne [5, Theorem 2.3], can also be found in [17] and is as follows.

Theorem 2.6. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be α-`ISM` and let $\gamma \in (0, 2\alpha)$. Then, for any $x \in \mathbb{R}^n$, the sequence $\{(P_\Omega(I-\gamma f))x^k\}_{k=0}^\infty$ converges to a point x^* which is a solution (whenever a solution exists) of the following variational inequality problem.*

$$\langle f(x^*), x - x^*\rangle \geq 0 \textit{ for all } x \in \Omega. \tag{2.11}$$

3. The algorithm

In this section we formulate (1.2) as a minimization problem and show how an algorithm which is based on Toint's [27] trust region method can be applied. Although several improvements have been presented (see e.g., Zhu [28] and Jia and Zhu [22]), our goal here is to present a general scheme for which can be applied for solving (1.2).

Consider the problem (1.2); we define the operator $G(x) := P_Q(F(x)) - F(x)$. Then one can verify that (1.2) reduces to the problem of finding a zero of G which is also in C, i.e.,

$$\text{find a point } x^* \in C \text{ such that } G(x^*) = 0. \tag{3.1}$$

So, the function p in (1.3) is actually

$$p(x) = \frac{1}{2}\|G(x)\|^2 = \frac{1}{2}\|P_Q(F(x)) - F(x)\|^2. \tag{3.2}$$

Our goal is to solve the following optimization problem.

$$\min_{x \in C} p(x). \tag{3.3}$$

The function p is continuously differentiable and by [4, Proposition 2.1.2, p. 194], a stationary point $x^* \in C$ satisfies the condition

$$\langle \nabla p(x^*), x - x^*\rangle \geq 0 \text{ for all } x \in C. \tag{3.4}$$

The projected gradient trust-region method is based on the idea that at each iterate x^k, an approximation model of the objective function is constructed, denoted by m_k in a trust-region of x^k, where this approximation is an adequate approximation of the objective. The model m_k satisfies $m_k(x^k) = p(x^k)$ and there exist nonnegative constants k_1 and k_2 such that $\left\|\nabla m_k(x^k) - \nabla p(x^k)\right\| \leq \min\{k_1\triangle_k, k_2\}$ for all k. This provides some assurance that the first-order information on the objective is reasonably accurate.

In addition, a step s^k is taken which guarantees a "sufficient" decrease of the approximation inside the intersection of this trust-region and the feasible set.

In order to describe the algorithm in detail we introduce several notations and definitions following Toint [27]. At the k-th iterate the trust-region is the closed ball

$$B(x^k, \triangle_k) := \{y \mid \left\|y - x^k\right\| \leq \triangle_k\}. \tag{3.5}$$

For $t \in \mathbb{R}$, define the *arc*

$$d_k(t) := P_C(x^k - t\nabla m_k(x^k)) - x^k. \tag{3.6}$$

Now the technique of choosing the step s^k is presented. Following [27], this choice is divided to two procedures, the first ensures "sufficient" decrease and the second improvement.

Algorithm 3.1. Step determination

1: Find $t_k^A \in \mathbb{R}$ such that

$$m_k(x^k + d_k(t_k^A)) \leq p(x^k) + \mu_1 \left\langle \nabla m_k(x^k), d_k(t_k^A)\right\rangle \tag{3.7}$$

and $\left\|d_k(t_k^A)\right\| \leq \nu_1\triangle_k$, and $t_k^A \geq \nu_2 t_k^B$ or $t_k^A \geq \min\{\nu_3\triangle_k/\left\|\nabla m_k(x^k)\right\|, \nu_4\}$, where t_k^B (if required) satisfies

$$m_k(x^k + d_k(t_k^B)) \geq p(x^k) + \mu_2 \left\langle \nabla m_k(x^k), d_k(t_k^B)\right\rangle. \tag{3.8}$$

2: Choose s^k such that

$$p(x^k) - m_k(x^k + s^k) \geq \mu_3 \left(p(x^k) - m_k(x^k + d_k(t_k^A))\right) \tag{3.9}$$

with $\left\|s^k\right\| \leq \nu_5\triangle_k$, and $x^k + s^k \in C$.

The constants should satisfy $0 < \mu_1 < \mu_2 < 1$, $\mu_3 \in (0, 1]$, $0 < \nu_3 < \nu_1 \leq \nu_5$, $\nu_2 \in (0, 1]$ and $\nu_4 > 0$. Now assume that the constants η_1, η_2, γ_1, γ_2 and γ_3 are given and satisfy $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 \leq \gamma_2 < 1 \leq \gamma_3$. Next is the second procedure to guarantee decrease of the objective function.

Algorithm 3.2. Minimization

Initialization: Choose a starting point $x^0 \in C$, a trust-region radius $\triangle_0$ and set $k = 0$.

Iterative step: Given the current iterate x^k

(i) compute the step s^k by using Algorithm 3.1 above.

(ii) Compute the ratio

$$\rho_k = \frac{p(x^k) - p(x^k + s^k)}{p(x^k) - m_k(x^k + s^k)}. \tag{3.10}$$

(iii) If $\rho_k > \eta_1$, set $x^{k+1} = x^k + s^k$ and

$$\triangle_{k+1} \in \begin{cases} [\triangle_k, \gamma_3 \triangle_k] & \text{if } \rho_k \geq \eta_2, \\ [\gamma_2 \triangle_k, \triangle_k] & \text{if } \rho_k < \eta_2, \end{cases} \tag{3.11}$$

otherwise, set $x^{k+1} = x^k$ and $\triangle_{k+1} \in [\gamma_1 \triangle_k, \gamma_2 \triangle_k]$.

(iv) Define m_{k+1} in the neighborhood of x^{k+1} and go to (i).

For the convergence of the algorithm see [27, Sections 3 and 4]. Toint's proposed method is quite general since it allows different choices of the model m_k, for example the objective function itself or the linear model, that is

$$m_k(x^k + s) = p(x^k) + \left\langle \nabla p(x^k), s \right\rangle. \tag{3.12}$$

In case where m_k is chosen as the first-order approximation of the objective this is also called Successive Linear Programing Approach.

Remark 3.3. 1. Observe that if $m_k(x^k + s) = p(x^k + s)$ then $\rho_k > \eta_1$ for all k. In this case the step search of the algorithm is identical to the one in [7].

2. If $s^k = d_k(t_k^A)$ then the classical projected gradient method is obtained.

3. In practice if p happens to be convex then we are able to obtain a global optimum.

4. Summary and further discussion

In this paper we follow similar techniques as in [11] for solving the Nonlinear Split Feasibility Problem (NLSFP). Due to the absence of convexity of the proximity function we apply Toint's [27] scheme which leads to a projected gradient trust-region method. In the linear case if the model m_k is chosen as the first-order approximation of the objective this reduces to the method established in [11] (for two sets also known as Byrne's CQ algorithm [5]).

As explained in the introduction, moving from linear to non-linear SFP does not preserve the convexity property of the proximity function and therefore without this assumption, only local minima is guaranteed. Since the projected gradient trust-region method can be seen as a sequential linearization method, it can also be considered as a sequence of operators with "good" properties that eventually converge to a solution. Since in the linear case Censor et. al. [12] and Byrne et. al. [6] prove convergence by applying the classical Krasnosel'skiĭ-Mann-Opial [19, 23, 26] and Baillon, Bruck and Reich [2] Theorems, our goal is to generate a sequence of operators that converges eventually to the desired solution. This idea is inspired by the recent work of Cegielski and Censor [8].

Following the Nonlinear Split Feasibility Problem (NLSFP) we could also phrase next the *Nonlinear Split Variational inequality Problem* (NLSVIP).

4.1. **The Split Variational Inequality Problem.** Let us recall a generalization of the SFP, that is the *Split Variational Inequality Problem* (SVIP) introduced by Censor et. al. in [12]. This problem allows, for example, to formulate the split minimization problem as a special case. Consider the Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$. Given operators $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^m \to \mathbb{R}^m$, an $m \times n$ matrix A, and non-empty, closed and convex subsets $C \subseteq \mathbb{R}^n$ and $Q \subseteq \mathbb{R}^m$, the SVIP is formulated as follows:

$$\text{find a point } x^* \in C \text{ such that } \langle f(x^*), x - x^* \rangle \geq 0 \text{ for all } x \in C \tag{4.1}$$

and such that

$$\text{the point } y^* = Ax^* \in Q \text{ and solves } \langle g(y^*), y - y^* \rangle \geq 0 \text{ for all } y \in Q. \tag{4.2}$$

When viewed separately, (4.1) is the classical *Variational Inequality Problem* (VIP) introduced by Hartman and Stampacchia [18]; we denote the solution set of (4.1) by $SOL(f, C)$. Based on the well-known result due to Eaves [14] for any $\lambda \geq 0$

$$x^* \in SOL(f, C) \Leftrightarrow x^* = P_C(x^* - \lambda f(x^*)). \tag{4.3}$$

With this idea, the SVIP can be written as the following minimization problem.

$$\min_{x \in SOL(f,C)} \frac{1}{2} \left\| P_Q(I - \lambda g)A(x) - A(x) \right\|^2. \tag{4.4}$$

Using the abbreviations $T := P_Q(I - \lambda g)$ and $U := P_C(I - \lambda f)$, the proposed algorithm, which is close to the projected gradient method is introduced [12, Algorithm 6.1].

Algorithm 4.1.

Initialization: Let $\lambda > 0$ and select an arbitrary starting point $x^0 \in \mathbb{R}^n$.

Iterative step: Given the current iterate x^k, compute

$$x^{k+1} = U(x^k + \gamma A^t(T - I)(Ax^k)), \tag{4.5}$$

where $\gamma \in (0, 1/L)$, L is the spectral radius of the operator A^tA, and A^t is the transpose of A.

Convergence of the algorithm is guaranteed if f and g are α_1-ISM and α_2-ISM operators on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, the solution set of (4.1)-(4.2) is non-empty, $\gamma \in (0, 1/L)$ and $\lambda \in [0, 2\alpha]$ where $\alpha := \min\{\alpha_1, \alpha_2\}$.

Remark 4.2. 1. A multiple set split variational inequality problem is also considered in [12], and it is shown how to transform it to (4.1)-(4.2) using a product space formulation.

2. Observe that by setting $f \equiv g \equiv 0$ in (4.1)–(4.2) we obtain the Split Feasibility Problem (SFP). This problem was used in the area of intensity-modulated radiation therapy (IMRT) treatment planning; see [11, 9].

3. The convergence of Algorithm 4.1 is based on the classical Krasnosel'skiĭ-Mann-Opial Theorem [19, 23, 26] and it is related to the projected gradient method, therefore, we believe that similar techniques to the ones presented here can be applied for solving the Nonlinear Split Variational Inequality Problem (NLSVIP) ((4.1)–(4.2) where A is replaced by a general vector valued function $F : \mathbb{R}^n \to \mathbb{R}^m$).

Acknowledgments

We thank an anonymous referee for constructive comments.

References

[1] J.-B. Baillon and G. Haddad, *Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones*, Israel Journal of Mathematics **26** (1977), 137–150.
[2] J.-B. Baillon, R. E. Bruck and S. Reich, *On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces*, Houston Journal of Mathematics **4** (1978), 1–9.
[3] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, Wiley, 2006.
[4] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, MA, USA, 1995.
[5] C. Byrne, *A unified treatment of some iterative algorithms in signal processing and image reconstruction*, Inverse Problems **20** (2004), 103–120.
[6] C. Byrne, Y. Censor, A. Gibali and S. Reich, *The split common null point problem*, Journal of Nonlinear and Convex Analysis **13** (2012), 759–775.
[7] P. H. Calamai and J. J. Moré, *Projected gradient methods for linearly constrained problems*, *Math Programming* **30** (1987), 93–116.
[8] A. Cegielski and Y. Censor, *Opial-type theorems and the common fixed point problem*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke and H. Wolkowicz (Editors), Springer-Verlag, New York, NY, USA, 2011, pp. 155–183.
[9] Y. Censor, T. Bortfeld, B. Martin and A. Trofimov, *A unified approach for inversion problems in intensity-modulated radiation therapy*, Physics in Medicine and Biology **51** (2006), 2353–2365.
[10] Y. Censor and T. Elfving , *A multiprojection algorithm using Bregman projections in product space*, Numerical Algorithms **8** (1994), 221–239.
[11] Y. Censor, T. Elfving, N. Kopf and T. Bortfeld, *The multiple-sets split feasibility problem and its applications for inverse problems*, Inverse Problems **21** (2005), 2071–2084.
[12] Y. Censor, A. Gibali and S. Reich, *Algorithms for the split variational inequality problem*, Numerical Algorithms **59** (2012), 301–323.
[13] Z. O. Dolidze, *On convergence of an analog of the gradient method*, Ekonomika i Matematicheskie Metody **20** (1984), 755–758.
[14] B. C. Eaves, *On the basic theorem of complementarity*, Mathematical Programming **1** (1971), 68–75.
[15] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Marcel Dekker, New York and Basel 1984.
[16] A. A. Goldstein, *Convex programming in Hilbert space*, Bulletin of the American Mathematical Society **70** (1964), 709–710.
[17] E. G. Golshtein and N. Tretyakov, *Modified Lagrangians and Monotone Maps in Optimization*, John Wiley & Sons, New York, NY, USA, 1996.
[18] P. Hartman and G. Stampacchia, *On some non-linear elliptic diferential-functional equations*, Acta Mathematica **115** (1966), 271–310.
[19] M. A. Krasnosel'skiĭ, *Two remarks on the method of successive approximations*, Uspekhi Matematicheskikh Nauk **10** (1955), 123–127 (in Russian).
[20] E. S. Levitin and B. T. Polyak, *Constrained minimization problems*, USSR Computational Mathematics and Mathematical Physics **6** (1966), 1–50.
[21] Z. Li, D. Han and W. Zhang, *A self-adaptive projection-type method for non-linear multiple-sets split feasibility problem*, Inverse Problems in Science and Engineering iFirst (2012), 1–16.
[22] C.-X. Jia and D. T. Zhu, *Projected gradient trust-region method for solving non-linear systems with convex constraints*, Applied Mathematics-A Journal of Chinese Universities **26** (2011), 57–69.
[23] W. R. Mann, *Mean value methods in iteration*, Proceedings of the American Mathematical Society **4** (1953), 506–510.

[24] E. Masad and S. Reich, *A note on the multiple-set split convex feasibility problem in Hilbert space. Journal of Nonlinear and Convex Analysis* **8** (2007), 367–371.

[25] P. Marcotte and J.-P. Dussault, *A sequential linear programming algorithm for solving monotone variational inequalities*, SIAM Journal on Control and Optimization **27** (1989), 1260–1278.

[26] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bulletin of the American Mathematical Society **73** (1967), 591–597.

[27] P. L. Toint, *Global convergence of a class of trust region algorithms for nonconvex minimization in Hilbert space*, IMA Journal of Numerical Analysis **8** (1988), 231–252.

[28] D. T. Zhu, *Convergence properties of projected gradient methods with nonmonotonic back tracking technique for convex constrained optimization*, Systems Science and Mathematical Sciences **13** (2000), 407–424.

Manuscript received December 17, 2013
revised May 26, 2013

A. Gibali
Optimization Department, Fraunhofer - ITWM, Kaiserslautern, Germany
E-mail address: gibali@itwm.fraunhofer.de

K.-H. Küfer
Optimization Department, Fraunhofer - ITWM, Kaiserslautern, Germany
E-mail address: karl-heinz.kuefer@itwm.fraunhofer.de

P. Süss
Optimization Department, Fraunhofer - ITWM, Kaiserslautern, Germany
E-mail address: philipp.suess@itwm.fraunhofer.de