The Multiple-Sets Split Feasibility Problem and Its Applications for Inverse Problems*

Yair Censor[1], Tommy Elfving[2], Nirit Kopf[1]
and Thomas Bortfeld[3]

[1]Department of Mathematics, University of Haifa
Mt. Carmel, Haifa 31905, Israel
(yair@math.haifa.ac.il, nirit_kopf@hotmail.com)

[2]Department of Mathematics, Linköping University,
SE-581 83 Linköping, Sweden (toelf@mai.liu.se).

[3]Department of Radiation Oncology
Massachusetts General Hospital (MGH)
and Harvard Medical School
Boston, MA 02114, USA (tbortfeld@hms.harvard.edu)

May 27, 2005. Revised: October 6, 2005.

Abstract

The multiple-sets split feasibility problem requires to find a point closest to a family of closed convex sets in one space such that its image under a linear transformation will be closest to another family of closed convex sets in the image space. It can be a model for many

*Accepted for publication in the journal *Inverse Problems*.

inverse problems where constraints are imposed on the solutions in the domain of a linear operator as well as in the operator's range. It generalizes the convex feasibility problem as well as the two-sets split feasibility problem. We propose a projection algorithm that minimizes a proximity function that measures the distance of a point from all sets. The formulation, as well as the algorithm, generalize earlier work on the split feasibility problem. We offer also a generalization to proximity functions with Bregman distances. Application of the method to the inverse problem of intensity-modulated radiation therapy (IMRT) treatment planning is studied in a separate companion paper and is here only briefly described.

1 Introduction

In this, somewhat lengthy, introduction we define the new multiple-sets split feasibility problem as a generalization of the well-known convex feasibility problem and as a generalization of the two-sets split feasibility problem. We briefly explain the, in principle, advantage of applying projection methods to such problems and mention the inverse problem of intensity-modulated radiation therapy as the real-world application that inspired the present work. We must emphasize though that, in spite of the general statements that we make below about the computational advantages of projection algorithms in other fields where huge-size real-world problems need to be solved, our development of the multiple-sets split feasibility problem formulation and algorithm are not computation cost driven. Even in the inverse problem of intensity-modulated radiation therapy (IMRT) that inspired our work (see Section 3) we do not yet have enough computational experience that would support computational cost-effectiveness claims. What our study does, therefore, offer, both in IMRT and for other inverse problems, is a mathematically valid framework for applying projection algorithms to inverse problems where constraints are imposed on the solutions in the domain of a linear operator as well as in the operator's range.

1.1 Feasibility problems

Given closed convex sets $C_i \subseteq R^N$, $i = 1, 2, \ldots, t$, and closed convex sets $Q_j \subseteq R^M$, $j = 1, 2, \ldots, r$, in the N- and M-dimensional Euclidean spaces, respectively, the *multiple-sets split feasibility problem,* proposed and studied here, is to find a vector x^* for which

$$x^* \in C := \cap_{i=1}^{t} C_i \ \text{ such that } \ Ax^* \in Q := \cap_{j=1}^{r} Q_j, \tag{1}$$

where A is a given $M \times N$ real matrix. This can serve as a model for many inverse problems where constraints are imposed on the solutions in the domain of a linear operator as well as in the operator's range. The multiple-sets split feasibility problem extends the well-known *convex feasibility problem* which is obtained from (1) when there are no matrix A and sets Q_j present at all, or put differently, when $Q = R^M$. For information on the convex feasibility problem see, e.g., Bauschke and Borwein [4], Combettes [20], or Censor and Zenios [19]. Systems of linear equations, linear inequalities, or convex inequalities are all encompassed by the convex feasibility problem which has broad applicability in many areas of mathematics and the physical and engineering sciences. These include, among others, optimization theory (see, e.g., Eremin [23] and Censor and Lent [18]), approximation theory (see, e.g., Deutsch [21] and references therein) and image reconstruction from projections in computerized tomography (see, e.g., Herman [25, 26], Censor [13]).

1.2 Projection methods and their advantage

Projections onto sets are used in a wide variety of methods in optimization theory but not every method that uses projections really belongs to the class of projection methods. *Projection methods* are iterative algorithms that use projections onto sets while relying on the general principle that when a family of (usually closed and convex) sets is present then projections onto the given individual sets are easier to perform then projections onto other sets (intersections, image sets under some transformation, etc.) that are derived from the given individual sets.

A projection algorithm reaches its goal that is related to the whole family of sets by performing projections onto the individual sets. Projection algorithms employ projections onto convex sets in various ways. They may use different kinds of projections and, sometimes, even use different projections within the same algorithm. They serve to solve a variety of problems which are either of the feasibility or the optimization types. They have different algorithmic structures, of which some are particularly suitable for parallel computing, and they demonstrate nice convergence properties and/or good initial behavior patterns. This class of algorithms has witnessed great progress in recent years and its member algorithms have been applied with success to fully-discretized models of problems in image reconstruction and image processing, see, e.g., Stark and Yang [32], Bauschke and Borwein [4] and Censor and Zenios [19].

Apart from theoretical interest, the main advantage of projection methods which makes them successful in real-world applications is computational. They commonly have the ability to handle huge-size problems of dimensions beyond which other, more sophisticated currently available, methods cease to be efficient. This is so because the building bricks of a projection algorithm are the projections onto the given individual sets (assumed and actually easy to perform) and the algorithmic structure is either sequential or simultaneous (or in-between). Sequential algorithmic structures cater for the row-action approach (see Censor [13]) while simultaneous algorithmic structures favor parallel computing platforms, see, e.g., Censor, Gordon and Gordon [17].

1.3 The split feasibility problem

The special case when there is only one set in each space, i.e., $t = r = 1$ in (1), was proposed by Censor and Elfving in [16] and termed the *split feasibility problem* (because of the limitation to one set in each space we will call this from now on the *two-sets split feasibility problem).* There we used our simultaneous multiprojections algorithm (see also [19, Subsection 5.9.3]) to obtain an iterative algorithm whose iterative step has the form

$$x^{k+1} = A^{-1}(I + AA^T)^{-1}(AP_C\left(x^k\right) + AA^T P_Q(Ax^k)) \tag{2}$$

to solve the two-sets split feasibility problem. Here x^k and x^{k+1} are the current and the next iteration vectors, respectively, T stands for matrix transposition, I is the unit matrix and P_C and P_Q denote the orthogonal projections onto C and Q, respectively. That solution was restricted to the case when $M = N$ and to the feasible case, i.e., when $Q \cap A(C) \neq \emptyset$. Byrne and Censor investigated this further in [11, Section 5]. Recognizing the potential difficulties with calculating inverses of matrices, or, equivalently, solving a linear system in each iterative step, particularly when the dimensions are large, Byrne [9] devised the CQ-algorithm which uses the iterative step

$$x^{k+1} = P_C\left(x^k + \gamma A^T\left(P_Q - I\right)Ax^k\right), \tag{3}$$

where $\gamma \in (0, 2/L)$ and L is the largest eigenvalue of the matrix A^TA.

One might wonder why not solve one of the convex feasibility problems of finding a point in $Q \cap A(C)$ or of finding a point in $C \cap A^{-1}(Q)$ instead of using the CQ-algorithm? Examples of situations when this would not be recommended can occur when, due to the underlying specific data of the real-world problem, it is not easy to perform projections onto the sets $A(C)$ and/or $A^{-1}(Q)$. Other examples might occur when the dimensions M and N are very different from each other and choosing one of those convex feasibility problems would cost us in calculating projections for one of the sets in a much larger dimensional space then if we perform projections in each space separately onto the given individual sets. Similar arguments apply to the multiple-sets split feasibility problem.

Our aim in this paper is to present, motivate and study the multiple-sets split feasibility problem. We devise, in Section 2, a projection method that obeys the general paradigm of projection algorithms, described above, namely, it performs projections onto the given individual sets to reach the overall goal of the problem, and which reduces precisely to Byrne's CQ-algorithm in the two-sets split feasibility situation. In order to cover the feasible and the infeasible cases for our problem, we handle it with a proximity function minimization approach. We apply to this proximity function a gradient projection algorithmic scheme and study conditions that guarantee its convergence.

While the multiple-sets split feasibility problem is potentially useful for a variety of inversion problems that can be formulated so, we describe briefly, in Section 3, a specific application in intensity-modulated radiation therapy (IMRT) that motivates our interest in the multiple-sets split feasibility problem and which is presented in detail in our separate companion article to this one [14]. In the Appendix we bring a telegraphic list of definitions and results that we use in our work. The recent work of Yang [36] on the two-sets split feasibility problem and the CQ-algorithm will be mentioned in the sequel. Related to the two-sets split feasibility problem are also the recent papers of Zhao and Yang [37] and Qu and Xiu [30]. The linear case for the two-sets split feasibility problem is discussed by Cegielski [12].

2 A Projection Algorithm for the Multiple-Sets Split Feasibility Problem

2.1 The algorithm

Consider the multiple-sets split feasibility problem defined in Section 1. For notational convenience reasons we consider an additional closed convex set $\Omega \subseteq R^N$ and further define the *constrained multiple-sets split feasibility problem* as the problem of finding

$$x^* \in \Omega \ \text{ such that } x^* \ \text{ solves (1).} \tag{4}$$

Denoting by P the orthogonal projection onto the closed convex set appearing in its subscript, we define a proximity function on R^N for this problem by

$$p(x) := (1/2)\sum_{i=1}^{t} \alpha_i ||P_{C_i}(x) - x||^2 + (1/2)\sum_{j=1}^{r} \beta_j ||P_{Q_j}(Ax) - Ax||^2, \tag{5}$$

where $\alpha_i > 0$ for all i, $\beta_j > 0$ for all j. An additional condition like $\sum_{i=1}^{t} \alpha_i + \sum_{j=1}^{r} \beta_j = 1$ is sometimes very useful in practical application to real-world problems when the α_i's and β_j's are *weights of importance* attached to the constraints. But this condition is not necessary for our analysis below since $p(x)$ is convex by being a linear positive combination of convex terms. This proximity function "measures" the "distance" of a point to all sets of (1) for which the coefficient α_i or β_j is positive. If the problem (1) is feasible then unconstrained minimization of $p(x)$ will yield the value zero, otherwise it will, in the infeasible (i.e., inconsistent) case find a point which is least violating the feasibility in the sense of being "closest" to all sets, as "measured" by $p(x)$. Note that with the choice $t = r = 1$, $C_1 = R^N$ and $Q_1 = \{b\}$ we retrieve the classical Tikhonov regularization, namely, $p(x) = (1/2)\alpha_1||x||^2 + (1/2)\beta_1||b - Ax||^2$. In order to find a solution of the constrained multiple-sets split feasibility problem we consider the minimization problem

$$\min\{p(x) \mid x \in \Omega\} \tag{6}$$

and propose the following algorithm.

Algorithm 1

Initialization: *Let x^0 be arbitrary.*

Iterative step: *For $k \geq 0$, given the current iterate x^k calculate the next iterate x^{k+1} by*

$$x^{k+1} = P_\Omega \left\{ x^k + s \left(\sum_{i=1}^{t} \alpha_i (P_{C_i}(x^k) - x^k) + \sum_{j=1}^{r} \beta_j A^T (P_{Q_j}(Ax^k) - Ax^k) \right) \right\}, \tag{7}$$

where s is a positive scalar such that $0 < s < 2/L$ and L is the Lipschitz constant of the gradient $\nabla p(x)$ of the proximity function in (5).

2.2 Convergence analysis

We address the convergence question of Algorithm 1 with two different tools. One is based on the constant stepsize lemma for gradient projection methods (Lemma 12 in the Appendix). The other is based on Dolidze's theorem (Theorem 13 in the Appendix) and is inspired by the work of Byrne [10].

Theorem 2 *Let $C = \cap_{i=1}^{t} C_i$ and $Q = \cap_{j=1}^{r} Q_j$ be intersections of nonempty closed convex sets in R^N and R^M, respectively, let $\Omega \subseteq R^N$ be a nonempty closed convex set, let A be an $M \times N$ real matrix and let $p(x)$ be as in (5) with α_i and β_j positive scalars. Then*

(i) the gradient $\nabla p(x)$ of the proximity function (5) is Lipschitz continuous and

$$L = \sum_{i=1}^{t} \alpha_i + \rho(A^T A) \sum_{j=1}^{r} \beta_j, \tag{8}$$

is a Lipschitz constant for it, where $\rho(A^T A)$ is the spectral radius of $A^T A$, and

(ii) if s is a positive scalar such that $0 < s < 2/L$, where L is a Lipschitz constant of $\nabla p(x)$, then every limit point of any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 1, is a stationary point of the function $p(x)$ over Ω.

Proof. (i) If $F(x) = (1/2)||P_\Theta(x) - x||^2$ where $P_\Theta(x)$ is the projection of x onto some closed convex set Θ then, by Aubin and Cellina ([2, Proposition 1, p. 24]), $\nabla F(x) = x - P_\Theta(x)$. Using the chain rule (see, e.g., [31, Theorem

23.9]) $\nabla_x(F(Ax)) = A^T \nabla_y F(y) \mid_{y=Ax}$, where ∇_x and ∇_y are the gradients with respect to the subscript variable, respectively, we obtain

$$\nabla \left((1/2) \parallel P_\Theta(Ax) - Ax \parallel^2\right) = A^T(I - P_\Theta)Ax. \tag{9}$$

Applying this to $p(x)$ we get

$$\nabla p(x) = \sum_{i=1}^{t} \alpha_i (I - P_{C_i})x + \sum_{j=1}^{r} \beta_j A^T (I - P_{Q_j})Ax. \tag{10}$$

Let $T_i := I - P_{C_i}$ and $S_j := I - P_{Q_j}$. Since an orthogonal projector is firmly nonexpansive, see, e.g., [4, Fact 1.5], both T_i and S_j are firmly nonexpansive by Lemma 10 in the Appendix, thus, nonexpansive. From (10) we have

$$\nabla p(x) - \nabla p(y) = \sum_{i=1}^{t} \alpha_i (T_i x - T_i y) + \sum_{j=1}^{r} \beta_j A^T (S_j Ax - S_j Ay). \tag{11}$$

Hence,

$$||\nabla p(x) - \nabla p(y)|| \leq \sum_{i=1}^{t} \alpha_i ||x - y|| + \sum_{j=1}^{r} \beta_j ||A^T|| \cdot ||A|| \cdot ||x - y||. \tag{12}$$

By choosing the two-norm, and observing that $||A^T||_2 \cdot ||A||_2 = ||A^T A||_2 = \rho(A^T A)$ the expression (8) follows.

(ii) Algorithm 1 is of the form (40)–(41) for $r_k = 1$ and $s_k = s$, for all $k \geq 0$, with the proximity function $p(x)$ of (5) playing the role of $f(x)$. Using Lemma 12 in the Appendix we get the required result. ■

This theorem does not guarantee convergence of sequences generated by the algorithm though. Therefore, our second convergence result is as follows.

Theorem 3 *If the assumptions of Theorem 2 hold then any sequence* $\{x^k\}_{k=0}^{\infty}$, *generated by Algorithm 1, converges to a solution of the constrained multiple-sets split feasibility problem, if a solution exists.*

Proof. Since $p(x)$ is convex and its gradient has a Lipschitz constant L (Theorem 2) ∇p is a ν-ism (see Definition 8 in the Appendix) with

$$\nu = 1/L = 1/\left(\sum_{i=1}^{t} \alpha_i + \rho(A^T A) \sum_{j=1}^{r} \beta_j\right). \tag{13}$$

This follows from Baillon and Haddad [3, Corollary 10] and can also be deduced from [24, Lemma 6.7, p. 98]. Using Dolidze's theorem (Theorem 13 in the Appendix) it follows that for the gradient in (10) and for $\gamma \in (0, 2/L)$, any sequence generated by the iterative step

$$x^{k+1} = P_\Omega(I - \gamma G)x^k, \tag{14}$$

converges to a solution of the variational inequality problem $VIP(G, \Omega)$ (see Problem 7 in the Appendix), if a solution exists. Since $\gamma \in (0, 2/L)$ the operator $B = P_\Omega(I - \gamma G)$ is averaged and, by Dolidze's theorem, the orbit sequence $\{B^k x\}_{k=0}^{\infty}$ converges to a fixed point of B, whenever such points exists. If z is a fixed point of B, then $z = P_\Omega(z - \gamma G z)$. Therefore, for any $c \in \Omega$,

$$\langle c - z, z - (z - \gamma G z) \rangle \geq 0. \tag{15}$$

which means that

$$\langle c - z, Gz \rangle \geq 0, \tag{16}$$

implying that z minimizes $p(x)$ over the set Ω. ■

Remark 4 *Byrne's CQ-algorithm (3) and its convergence results follow from the above analysis by taking $\Omega = C$, no sets C_i at all and a single set $Q_1 = Q$. A further potentially useful modification that we proposed in [14], but which does not yet have a mathematical validation, is the replacement of the orthogonal projections onto the sets in each of the spaces by subgradient projections, see, e.g., Censor and Lent [18] or [19, Subsection 5.3]. These are "projections" which do not require the iterative minimization of distance between the point and the set but are rather given by closed-form analytical expressions. Recently, Yang [36] proved that replacement of orthogonal projections by subgradient projections is permissible, without ruining the convergence of the CQ-algorithm, for the two-sets split feasibility problem under the assumption of consistency.*

2.3 A generalized proximity function

The formula $\nabla F(x) = x - P_\Theta(x)$, mentioned above from ([2, Proposition 1, p. 24]), has been recently generalized to cover Bregman functions, distance and projections by Censor, De Pierro and Zaknoon [15] after an earlier generalization to the entropy case was done by Butnariu, Censor and Reich [8]. The *Bregman directed distance* $d_{\Theta}^{f}(x)$ *of a point* x *from a set* Θ *with respect to the Bregman function* f is defined [15, Equation (44)] by

$$d_{\Theta}^{f}(x) := D_f\left(P_{\Theta}^{f}(x), x\right) = f\left(P_{\Theta}^{f}(x)\right) - f(x) - \left\langle \nabla f\left(x\right), P_{\Theta}^{f}(x) - x\right\rangle \quad (17)$$

where $P_{\Theta}^{f}(x)$ is the Bregman projection of x onto Θ with respect to the Bregman function f and $D_f\left(y, x\right)$ is the Bregman distance between y and x. See, e.g., [19, Chapter 2] for definitions, basic properties and references. Proposition 12 in [15] gives precise conditions under which the formula

$$\nabla\left(d_{\Theta}^{f}(x)\right) = \nabla^2 f(x)\left(x - P_{\Theta}^{f}(x)\right) \quad (18)$$

holds. In this formula $\nabla^2 f(x)$ is the Hessian matrix of f at x. The availability of this formula enables us to calculate the gradient of a *generalized proximity function* $p_f(x)$, with respect to a Bregman function f, for the multiple-sets split feasibility problem, that will have the form

$$p_f(x) := \sum_{i=1}^{t} \alpha_i D_f(P_{C_i}^{f}(x), x) + \sum_{j=1}^{r} \beta_j D_f(P_{Q_j}^{f}(Ax), Ax). \quad (19)$$

Using (18) and the chain rule in the proof of Theorem 2, we find that

$$\nabla p_f(x) = \sum_{i=1}^{t} \alpha_i \nabla^2 f(x)(I - P_{C_i}^{f})x + \sum_{j=1}^{r} \beta_j A^T \nabla^2 f(Ax)(I - P_{Q_j}^{f})Ax. \quad (20)$$

For the special Bregman function $f(x) = (1/2)\|x\|^2$ (see, e.g., [19, Example 2.1.1]) the Hessians are $\nabla^2 f(x) = \nabla^2 f(Ax) = I$, Bregman projections are orthogonal projections and the proximity function (5) is recovered. For the entropy case the Bregman function is $f\left(x\right) = -\operatorname{ent} x$, where $\operatorname{ent} x$ is Shannon's entropy function which maps the nonnegative orthant R_{+}^{n} into R according to

$$\operatorname{ent} x := -\sum_{j=1}^{n} x_j \log x_j. \quad (21)$$

Here "log" denotes the natural logarithms and, by definition, $0 \log 0 = 0$. See [19, Example 2.1.2 and Lemma 2.13] for a verification that $f(x)$ is a Bregman function with zone

$$S_e := \{x \in R^n \mid x_j > 0, \text{ for all } 1 \leq j \leq n\} \tag{22}$$

and that

$$D_f(x, y) = \sum_{j=1}^{n} x_j (\log(x_j/y_j) - 1) + \sum_{j=1}^{n} y_j. \tag{23}$$

The Hessian of f

$$\nabla^2 f(x) = \nabla^2 \left(\sum_{j=1}^{n} x_j \log x_j \right) = \begin{pmatrix} \frac{1}{x_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_n} \end{pmatrix}, \tag{24}$$

is always a positive semi-definite matrix on $S_e \times S_e$ [15, Lemma 22].

In the general case, we make the additional assumption of boundedness of the Hessians, i.e., that $\|\nabla^2 f(x)\| \leq \chi_1$ for all $x \in \Omega$ and that $\|\nabla^2 f(Ax)\| \leq \chi_2$ for all $Ax \in A(\Omega)$, for some constants χ_1 and χ_2. Then we can show, in a similar manner to what has been done in the proof of Theorem 2, that

$$\begin{aligned} &||\nabla p_f(x) - \nabla p_f(y)|| \\ &\leq \left(\sum_{i=1}^{t} \alpha_i \|\nabla^2 f(x)\| + \sum_{j=1}^{r} \beta_j ||A^T|| \cdot ||A|| \cdot \|\nabla^2 f(Ax)\| \right) ||x - y||, \end{aligned} \tag{25}$$

so that

$$L = \chi_1 \sum_{i=1}^{t} \alpha_i + \rho(A^T A) \chi_2 \sum_{j=1}^{r} \beta_j \tag{26}$$

is a Lipschitz constant for $\nabla p_f(x)$ and a theorem that generalizes Theorem 2 to the case of generalized proximity functions follows. The generalized projection algorithm, for the multiple-sets split feasibility problem, that uses Bregman projections takes the following form.

Algorithm 5

Initialization: *Let x^0 be arbitrary.*

Iterative step: *For $k \geq 0$, given the current iterate x^k calculate the next iterate x^{k+1} by*

$$x^{k+1} = P_{\Omega}\left(x^k + s\Gamma(x^k)\right), \tag{27}$$

where

$$\Gamma(x^k) = \sum_{i=1}^{t} \alpha_i \nabla^2 f\left(x^k\right)\left(P_{C_i}^f(x^k) - x^k\right) + \sum_{j=1}^{r} \beta_j A^T \nabla^2 f\left(Ax^k\right)\left(P_{Q_j}^f(Ax^k) - Ax^k\right) \tag{28}$$

and s is a positive scalar such that $0 < s < 2/L$ and L is the Lipschitz constant of the gradient $\nabla p_f(x)$ of the generalized proximity function in (19).

3 The multiple-sets split feasibility problem in intensity-modulated radiation therapy

In our companion article to the present one [14] the multiple-sets split feasibility problem is applied to the inverse problem of intensity-modulated radiation therapy. In *intensity-modulated radiation therapy* (IMRT), see, e.g., Palta and Mackie [29], beams of penetrating radiation are directed at the lesion (tumor) from external sources. Based on understanding of the physics and biology of the situation, there are two principal aspects of radiation teletherapy that call for computational modeling.

The first aspect is the calculation of dose. The dose is a measure of the actual energy absorbed per unit mass everywhere in the irradiated tissue. This yields a *dose function* (also called *dose map* or *dose distribution*) whose values are the dose absorbed as a function of location inside the irradiated body. This dose calculation is the *forward problem* of IMRT.

The second aspect is the *inverse problem* of the first. In addition to the physical and biological parameters of the irradiated object that were assumed known for the dose calculation, we assume here that information about the capabilities and specifications of the available *treatment machine* (i.e., radiation source) is given. Based on medical diagnosis, knowledge, and experience, the physician prescribes *desired upper and lower dose bounds* to the treatment planning case. The output of a solution method for the inverse problem is a *radiation intensity function* (also called *intensity map*). Its values are the radiation intensities at the sources, as a function of source location, that would result in a dose function which agrees with the prescribed dose bounds.

To be of practical value, this radiation intensity function must be *deliverable* on the available treatment machine. The set of deliverable intensity maps is convex because for any two deliverable intensity maps (intensity vectors in the fully-discretized model) x^1 and x^2 the (nonnegative) linear combination $a_1x^1 + a_2x^2$ is also deliverable: simply deliver x^1 for a time fraction a_1 and x^2 for a time fraction a_2 (here we disregard the fact that in practice the treatment time often has to be an integer multiple of a "monitor unit"). An important delivery constraint is nonnegativity, namely, we can never deliver negative intensities. Hence, all deliverable intensity vectors must belong to the nonnegative orthant. Another physical/technical issue that needs to be considered in this context is leakage radiation, which always accompanies any primary radiation. Depending on the technical equipment used to deliver the treatment, there are also other delivery constraints. An orthogonal projector can easily be found for simple delivery constraints such as nonnegativity. For other delivery constraints, calculating projections might involve an inner-loop optimization process of minimizing the distance from a point to the set in each iteration.

An example of a constraint set C_i in the space of radiation intensity vectors is the smoothness constraint. Smoothness of intensity maps is desirable because it permits more efficient and robust delivery of IMRT with a multileaf collimator (MLC), see Webb, Convery and Evans [34], Alber and Nüsslin [1] and Kessen, Grosser and Bortfeld [28]. Smoothness constraints may be either convex or non-convex. One example of a convex smoothness constraint is the bandlimited constraint. Its associated orthogonal projector is an ideal low-pass filter with a given cut-off frequency.

For the sets Q_j, in the dose space, we already mentioned the commonly-used minimum and maximum dose constraints as examples. More recently the concept of *equivalent uniform dose* (EUD) was introduced to describe dose distributions with a higher clinical relevance, see, e.g., [14] for references. It has been used in IMRT optimization by Thieke *et al.* [33] and by Wu *et al.* [35]. EUD constraints are defined for tumors as the biological equivalent dose that, if given uniformly, will lead to the same cell-kill in the tumor volume as the actual non-uniform dose distribution. They could also be defined for normal tissues. Following the recent work of Thieke *et al.* [33] who derived an approximately orthogonal EUD projector, we develop, study and test experimentally in [14] a unified theory that enables treatment of both EUD constraints and physical dose constraints.

The unified new model relies on the multiple-sets split feasibility problem formulation, developed here, and it accommodates the specific IMRT situation. The constraints are formulated in two different Euclidean vector spaces. The delivery constraints are formulated as sets in the Euclidean vector space of radiation intensity vectors (i.e., vectors whose components are radiation intensities) and the dimensionality of this space equals the total number of discretized radiation sources.

The basic linear feasibility problem associated with recovering the radiation intensities vector x is the following.

$$0 \leq l_\nu \leq \sum_{j=1}^{N} d_{ij} x_j, \quad \text{for all} \quad i \in S_\nu^{PTV}, \ \nu = 1, 2, \cdots, T, \tag{29}$$

$$0 \leq \sum_{j=1}^{N} d_{ij} x_j \leq u_\nu, \quad \text{for all} \quad i \in S_\nu^{OAR}, \ \nu = 1, 2, \cdots, Q, \tag{30}$$

$$x_j \geq 0, \quad \text{for all} \ \ j = 1, 2, \cdots, N, \tag{31}$$

where S_ν^{PTV} are the planning target volumes (PTVs), S_ν^{OAR} are organs at risk (OARs) and l_ν and u_ν are lower and upper bounds, respectively, on the required or permitted doses to organs. The system (29)–(31) can be rewritten in the general form

$$0 \leq l_i \leq \sum_{j=1}^{M} d_{ij} x_j \leq u_i, \quad \text{for all} \ \ i = 1, 2, \cdots, t, \tag{32}$$

where l_i and u_i are correctly identified with the l_ν and u_ν, according to the organ to which the i-th voxel belongs, and by appropriately defining additional lower and upper bounds.

The EUD constraints are formulated in the Euclidean vector space of dose vectors (i.e., vectors whose components are the doses in each voxel) and the dimensionality of this space equals the total number of voxels. The EUD constraints refer to individual organs, thus to individual subsets of voxels. To simplify notations in the sequel let us count all PTVs and OARs sequentially by S_μ, $\mu = 1, 2, \cdots, T, T+1, \cdots, T+Q$, where the first T structures S_μ represent the sets S_ν^{PTV}, $\nu = 1, 2, \cdots, T$, and the next Q structures S_μ represent the sets S_ν^{OAR}, $\nu = 1, 2, \cdots, Q$. Also, let us denote for now by J_μ the number of voxels in structure S_μ. With these notations let $h^{(\mu)} = (h_i)_{i \in S_\mu}$, be the J_μ-th dimensional (i.e., $h^{(\mu)} \in R^{J_\mu}$) sub-vector (of the vector h) whose coordinates are *the doses absorbed in the voxels of the μ-th structure S_μ*. Alternatively, we say that $h^{(\mu)}$ is the μ-th block of the vector h. For each structure S_μ, $\mu = 1, 2, \cdots, T, T+1, \cdots, T+Q$, we define a real-valued function $E_{\mu,\alpha} : R^{J_\mu} \to R$, called *the EUD function*, by

$$E_{\mu,\alpha}(h^{(\mu)}) = \left((1/J_\mu) \sum_{i \in S_\mu} (h_i)^\alpha \right)^{1/\alpha}. \tag{33}$$

Each $E_{\mu,\alpha}$ maps the dose sub-vector of the μ-th structure S_μ into a single real number via (33). The parameter α is a tissue-specific number which is negative for PTVs and positive for OARs. For $\alpha = 1$ the EUD function is precisely the mean dose of the organ for which it is calculated.

For each PTV structure S_μ, $\mu = 1, 2, \cdots, T$, the parameter α is chosen negative and the EUD constraint is described by the set

$$\Omega_\mu = \{h^{(\mu)} \in R^{J_\mu} \mid E_{\mu,\alpha}^{\min} \le E_{\mu,\alpha}(h^{(\mu)}), \text{ and } \alpha < 0\}, \tag{34}$$

where $E_{\mu,\alpha}^{\min}$ is given, for each PTV structure, by the treatment planner. For each OAR S_μ, $\mu = T+1, T+2, \cdots, T+Q$, the parameter is chosen $\alpha \ge 1$ and the EUD constraint can be described by the set

$$\Gamma_\mu = \{h^{(\mu)} \in R^{J_\mu} \mid E_{\mu,\alpha}(h^{(\mu)}) \le E_{\mu,\alpha}^{\max}, \text{ and } \alpha \ge 1\}, \tag{35}$$

where $E_{\mu,\alpha}^{\max}$ is given, for each OAR, by the treatment planner. These sets have been shown to be convex. Thus, our *unified model for physical dose and EUD constraints* takes the form of a multiple-sets split feasibility problem where some constraints are formulated in the radiation intensities space R^N and other constraints are formulated in the dose space R^M and the two spaces are related by a (known) linear transformation D (the dose matrix). The problem then becomes

$$\text{find } x^* \in \cap_{i=1}^t C_i \text{ such that } h^* = Dx^* \text{ and } h^* \in \cap_{\mu=1}^{T+Q} \Theta_\mu \tag{36}$$

where C_i represent the hyperslabs of (32) and Θ_μ are generically describing the EUD constraints of (34) and (35) in the dose space R^M. The work presented here allows us to accommodate such constraints in a valid logical framework for performing an iterative solution process that iterates in each of the two spaces and correctly passes back and forth between the spaces during iterations. There are other inversion problems within IMRT and in other fields of applications that can be cast into a multiple-sets split feasibility problem and treated by the projection algorithmic approach presented here.

4 Conclusions

We propose the multiple-sets split feasibility problem as a generalization of both the convex feasibility problem and the two-sets split feasibility problem. This constitutes a mathematically valid framework for applying projection algorithms to inverse problems where constraints are imposed on the solutions in the domain of a linear operator as well as in the operator's range. We explain, in general terms, the advantages of projection methods and develop a simultaneous projection algorithm that minimizes a proximity function in order to reach a solution of the multiple-sets split feasibility problem problem. We offer an additional extension of the theory via using Bregman distances and Bregman projections in the proximity function and in the algorithm.

A specific inverse problem in intensity-modulated radiation therapy (IMRT), where both physical dose constraints, EUD (nonlinear) constraints and non-negativity constraints must all be satisfied to obtain a solution, is our inspiration for the work presented here. Our companion IMRT-oriented paper will be published elsewhere [14].

Acknowledgments. We thank three anonymous referees whose constructive comments helped us improve the paper. The work of T. Bortfeld and Y. Censor was supported by grant No. 2003275 of the United States-Israel Binational Science Foundation (BSF). The work of Y. Censor on this research was supported by a National Institutes of Health (NIH) grant No. HL70472 and by grant No. 522/04 of the Israel Science Foundation (ISF) and was partially done at the Center for Computational Mathematics and Scientific Computation (CCMSC) in the University of Haifa. This study was presented at "The Interdisciplinary Experts' Workshop on Intensity-Modulated Radiation Therapy (IMRT), Medical Imaging, and Optimization

Theory", June 6–11, 2004, the University of Haifa, Haifa, Israel.

References

[1] M. Alber and F. Nüsslin, Intensity modulated photon beams subject to a minimal surface smoothing constraint, *Physics in Medicine and Biology* **45,** pp. N49–N52, (2000).

[2] J.P. Aubin and A. Cellina, *Differential Inclusions: Set-Valued Maps and Viability Theory,* Springer-Verlag, Berlin, Heidelberg, Germany, 1984.

[3] J. Baillon and G. Haddad, Quelques propriétés des opérateurs angel-bornés et n-cycliquement monotones, *Israel Journal of Mathematics* **26**, pp. 137–150, (1977).

[4] H.H. Bauschke and J.M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review* **38**, pp. 367–426, (1996).

[5] D.P. Bertsekas, *Nonlinear Programming,* Athena Scientific, Belmont, MA, USA, 1995.

[6] D.P. Bertsekas, *Convex Analysis and Optimization,* Athena Scientific, Belmont, MA, USA, 2003.

[7] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.

[8] D. Butnariu, Y. Censor and S. Reich, Iterative averaging of entropic projections for solving stochastic convex feasibility problems, *Computational Optimization and Applications* **8**, pp. 21–39, (1997).

[9] C. Byrne, Iterative oblique projection onto convex sets and the split feasibility problem, *Inverse Problems* **18,** pp. 441–453, (2002).

[10] C. Byrne, A unified treatment of some iterative algorithms in signal processing and image reconstruction, *Inverse Problems* **20,** pp. 103–120, (2004).

[11] C. Byrne and Y. Censor, Proximity function minimization using multiple Bregman projections, with applications to split feasibility and Kullback-Leibler distance minimization, *Annals of Operations Research* **105**, pp. 77–98, (2001).

[12] A. Cegielski, Convergence of the projected surrogate constraints method for the linear split feasibility problems, Technical Report, September 14, 2004, Institute of Mathematics, University of Zielona Góra, Zielona Góra, Poland.

[13] Y. Censor, Row-action methods for huge and sparse systems and their applications, *SIAM Review* **23,** pp. 444–466, (1981).

[14] Y. Censor, T. Bortfeld, B. Martin and A. Trofimov, The split feasibility model leading to a unified approach for inversion problems in intensity-modulated radiation therapy, Technical Report, Department of Mathematics, University of Haifa, Haifa, Israel. April 20, 2005.

[15] Y. Censor, A.R. De Pierro and M. Zaknoon, Steered sequential projections for the inconsistent convex feasibility problem, *Nonlinear Analysis: Theory, Methods & Applications (Series A: Theory and Methods)* **59**, pp. 385–405, (2004).

[16] Y. Censor and T. Elfving, A multiprojection algorithm using Bregman projections in a product space, *Numerical Algorithms* **8,** pp. 221–239, (1994).

[17] Y. Censor, D. Gordon and R. Gordon, BICAV: A block-iterative, parallel algorithm for sparse systems with pixel-related weighting, *IEEE Transactions on Medical Imaging* **20**, pp. 1050–1060, (2001).

[18] Y. Censor and A. Lent, Cyclic subgradient projections, *Mathematical Programming* **24**, pp. 233–235, (1982).

[19] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, NY, USA, 1997.

[20] P.L. Combettes, The convex feasibility problem in image recovery, *Advances in Imaging and Electron Physics* **95**, pp. 155–270, (1996).

[21] F. Deutsch, The method of alternating orthogonal projections, in: S.P. Singh (Editor), *Approximation Theory, Spline Functions and Applications*, Kluwer Academic publishers, Dordrecht, The Netherlands, 1992, pp. 105–121.

[22] Z.O. Dolidze, On convergence of an analog of the gradient method, *Ekonomika i Matem. Metody* **20**, pp. 755–758, (1984).

[23] I.I. Eremin, Fejér mappings and convex programming, *Siberian Mathematical Journal* **10**, pp. 762–772, (1969).

[24] E.G. Golshtein and N. Tretyakov, *Modified Lagrangians and Monotone Maps in Optimization*, John Wiley & Sons, New York, NY, USA, 1996.

[25] G.T. Herman, *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, USA, 1980.

[26] G.T. Herman and L.B. Meyer, Algebraic reconstruction techniques can be made computationally efficient, *IEEE Transactions on Medical Imaging* **12**, pp. 600–609, (1993).

[27] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer-Verlag, Berlin, Heidelberg, Germany, 2001.

[28] A. Kessen, K. Grosser and T. Bortfeld, Simplification of IMRT intensity maps by means of 1-D and 2-D median-filtering during the iterative calculation, in: W. Schlegel and T. Bortfeld (Editors), *Proceedings of the XIIIth International Conference on The Use of Computers in Radiation Therapy,* Springer-Verlag, Heidelberg, Germany, 2000, pp. 545–547.

[29] J.R. Palta and T.R. Mackie (Editors), *Intensity-Modulated Radiation Therapy: The State of The Art,* American Association of Physicists in Medicine, Medical Physics Monograph No. 29, Medical Physics Publishing, Madison, WI, USA, 2003.

[30] B. Qu and N. Xiu, A note on the CQ algorithm for the split feasibility problem, *Inverse Problems*, **21** (2005), 1–11.

[31] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, USA, 1970.

[32] Stark, H., and Y. Yang. *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics.* New York, NY, USA, John Wiley & Sons, 1998.

[33] C. Thieke, T. Bortfeld, A. Niemierko and S. Nill, From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning, *Medical Physics* **30**, pp. 2332–2339, (2003).

[34] S. Webb, D.J. Convery and P.M. Evans, Inverse planning with constraints to generate smoothed intensity-modulated beams, *Physics in Medicine and Biology* **43,** pp. 2785–2794, (1998).

[35] Q. Wu, R. Mohan, A. Niemierko and R. Schmidt-Ullrich, Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose, *International Journal Radiation Oncology Biology Physics* **52,** pp. 224–235, (2002).

[36] Q. Yang, The relaxed CQ algorithm solving the split feasibility problem, *Inverse Problems*, **20** (2004), 1261–1266.

[37] J. Zhao and Q. Yang, Several solution methods for the split feasibility problem, *Inverse Problems*, **21** (2005), 1791–1799.

Appendix

In this appendix we give some definitions and mathematical tools that we use. We denote the inner product in R^n by $\langle x, y\rangle$ and the Euclidean norm by $\|x\|$. Given a nonempty closed convex set $\Omega \subseteq R^n$ and a point $x \in R^n$, an *orthogonal projection of* x *onto* Ω, denoted $P_\Omega(x)$, is a point in Ω which is closest to x, i.e., $P_\Omega(x) \in \Omega$, such that

$$\|x - P_\Omega(x)\| = \min\{\|x - y\| \mid y \in \Omega\}. \tag{37}$$

When the set Ω is closed and convex then existence and uniqueness of $P_\Omega(x)$ are guaranteed. Projections belong to the broader class of nonexpansive operators. A (possibly nonlinear) operator T on a closed convex set $\Omega \subseteq R^N$ is called *nonexpansive* if, for all x and y in Ω

$$||Tx - Ty|| \leq ||x - y||. \tag{38}$$

If Ω is a nonempty closed convex subset of R^N, then for all $x, y \in R^N$ we have $\|P_\Omega(x) - P_\Omega(y)\| \leq \|x - y\|$, see, e.g., Bertsekas [6, Proposition 2.2.1, p.

88]. Combining nonexpansive operators is done by composition or by convex combination as the following well-known result states.

Proposition 6 *If $T_1, T_2, \cdots, T_m$ are nonexpansive operators then the composition $T_m \cdots T_2 T_1$ is nonexpansive. If $w \in R^m$ is a weight vector (i.e., $w_i \geq 0$ and $\sum_{i=1}^{m} w_i = 1$) then $\sum_{i=1}^{m} w_i T_i$ is nonexpansive.*

Given a nonexpansive operator U, the operator $T := (1-\alpha)I + \alpha U$, for some $\alpha \in (0,1)$, where I is the unit operator, is called *averaged* or *averaging*. Such an operator is obviously also nonexpansive. A condition of the form

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \text{,for all } x, y \in R^n, \tag{39}$$

for some constant $L > 0$, is called a *Lipschitz continuity* condition on ∇f. The projection operators P_Ω are averaged, as are the operators of the form $(I - \gamma \nabla f)$ if ∇f is Lipschitz continuous and the parameter γ is appropriately chosen. The product of finitely many averaged operators is averaged, so the operators $P_{\Omega_2} P_{\Omega_1}$ and $P_\Omega (I - \gamma \nabla f)$ are also averaged.

The gradient projection method is a feasible directions method whose iterative step has the form

$$x^{k+1} = x^k + r_k(\overline{x}^k - x^k), \tag{40}$$

$$\overline{x}^k = P_\Omega(x^k - s_k \nabla f(x^k)), \tag{41}$$

where s_k is a positive scalar, called stepsize, and the numbers $r_k \in (0,1]$ are relaxation parameters.

An operator G on a closed convex set $\Omega \subseteq R^N$ is *monotone* (see, e.g., [27, Chapter A, Definition 4.1.3]) if for all x and y in Ω

$$\langle Gx - Gy, x - y \rangle \geq 0. \tag{42}$$

For example, if $g(\cdot)$ is a convex differentiable real-valued function on Ω then the gradient $\nabla g(\cdot)$ is a monotone operator.

Problem 7 ***(The Variational Inequality Problem).*** *Let G be a monotone operator with respect to a closed convex set $\Omega \subseteq R^N$. The* ***variational inequality problem with respect to** G **and** Ω*, *denoted by $VIP(G, \Omega)$, is to find a point $x^* \in \Omega$ for which $\langle Gx^*, x - x^* \rangle \geq 0$ for all $x \in \Omega$.*

Subject to certain restrictions on G and γ, a sequence $\left\{x^k\right\}_{k=0}^{\infty}$, defined by the iterative step

$$x^{k+1} = P_\Omega(I - \gamma G)x^k, \tag{43}$$

will converge to a solution of the $VIP(G,\Omega)$, if a solution exists. To see this observe [10, Theorem 2.1] that if G is an averaging operator on Ω and its fixed points set $Fix(G)$ is nonempty then the sequence $\left\{G^k x\right\}_{k=0}^{\infty}$ converges to a member of $Fix(G)$, for any $x \in \Omega$. The projection operator P_Ω is averaging, and for each $x \in \Omega$ the projection $P_\Omega(x)$ is characterized by (see, e.g., [27, Theorem 3.1.1, p. 47])

$$\langle y - P_\Omega(x), P_\Omega(x) - x\rangle \geq 0, \quad \text{for all} \;\; y \in \Omega. \tag{44}$$

Therefore, $x^* = P_\Omega(I - \gamma G)x^*$ if and only if

$$\langle y - x^*, x^* - (x^* - \gamma G x^*)\rangle = \gamma \langle y - x^*, Gx^*\rangle \geq 0, \quad \text{for all} \;\; y \in \Omega. \tag{45}$$

Consequently, the vector x^* solves the $VIP(G,\Omega)$ if and only if x^* is a fixed-point of the operator $P_\Omega(I - \gamma G)$.

Definition 8 *(See, e.g., Golshtein and Tretyakov [24, p. 256]). An operator G on a closed convex set $\Omega \subseteq R^N$ is called ν-`inverse strongly monotone` (ν-`ism`) if there is a $\nu > 0$ such that*

$$\langle Gx - Gy, x - y\rangle \geqslant \nu ||Gx - Gy||^2, \quad \textit{for all} \;\; x, y \in \Omega. \tag{46}$$

Definition 9 *An operator G on $\Omega \subseteq R^N$ is called `firmly nonexpansive` if it is a 1-ism, i.e., if*

$$\langle Gx - Gy, x - y\rangle \geq ||Gx - Gy||^2, \quad \textit{for all} \;\; x, y \in \Omega. \tag{47}$$

Lemma 10 *[10, Lemma 2.3] An operator F is firmly nonexpansive if and only if its complement $I - F$ is firmly nonexpansive. If F is firmly nonexpansive then F is averaged.*

It is well-known that every firmly nonexpansive operator is nonexpansive and that a convex combination of firmly nonexpansive operators is also a firmly nonexpansive operator.

Definition 11 *[5, p. 177]. Given a function $f : R^n \to R$ and a set $\Omega \subseteq R^n$, a vector x^* that satisfies the condition*

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0, \quad \textit{for all} \;\; x \in \Omega, \tag{48}$$

is referred to as ***a stationary point of*** f ***over*** Ω.

Condition (48) is an *optimality condition.* It is a necessary condition for x^* to be a local minimum of f over Ω, and if f is convex over Ω then it is also sufficient for local minimum at x^* (see, e.g., [5, Proposition 2.1.2, p. 176]). It is known (see, e.g., [7, Proposition 3.3(b), p. 213]) that x^* is a fixed point of the gradient projection mapping with constant stepsize $P_\Omega(x - s\nabla f(x))$ if and only if it is a stationary point of f over Ω. Furthermore, if f is convex on the set Ω then the latter guarantees that x^* minimizes f over Ω. Note that if $\Omega = R^n$ or if x^* is an interior point of Ω then (48) reduces to the stationarity condition $\nabla f(x^*) = 0$.

Lemma 12 *[5, Proposition 2.3.2, pp. 215–216] (**Constant Stepsize**). Let $\{x^k\}_{k=0}^\infty$ be a sequence, generated by the gradient projection method (40)–(41) with $r_k = 1$ and $s_k = s$, for all $k \geq 0$. Assume that for some constant $L > 0$, the gradient ∇f is Lipschitz continuous on Ω. If $0 < s < (2/L)$ then every limit point of $\{x^k\}_{k=0}^\infty$ is a stationary point of f.*

The theorem of Dolidze [22], as presented and proven in Byrne [10, Theorem 2.3], can also be found in [24] and is as follows.

Theorem 13 *(**Dolidze's Theorem**). Let G be ν-ism and let $\gamma \in (0, 2\nu)$. Then, for any $x \in R^n$, the sequence $\{(P_\Omega(I - \gamma G))^k x\}_{k=0}^\infty$ converges to a solution of $VIP(G, \Omega)$, whenever a solution exists.*