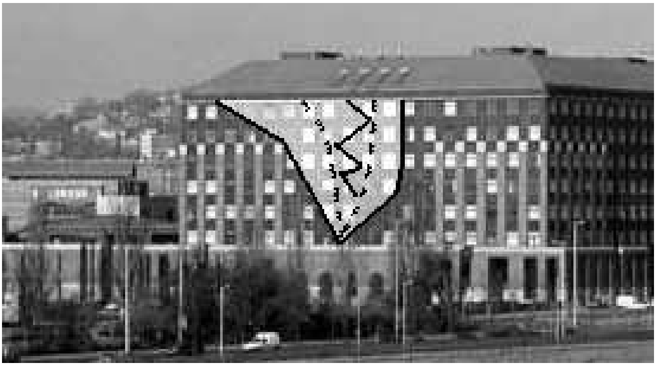
Operations Research
Report 2019-02



A new primal-dual algorithm for separable Bregman divergence minimization under linear constraints

Levente Mályusz, Anita Varga

June 2019

Eötvös Loránd University of Sciences
Department of Operations Research

`http://www.cs.elte.hu/opres/orr/`

ISSN 1215 - 5918

A new primal-dual algorithm for separable Bregman divergence minimization under linear constraints

L. Mályusz, A. Varga

Abstract

A transformed version of the separable Bregman divergence minimization problem under linear inequality constraints has been introduced by Kas et al. in [23], the Young programming problem. Their formulation led to a natural, symmetric duality theorem. They applied a row action method previously introduced by Censor et al. [8] to solve the problem and proved convergence under additional assumptions. In this paper we revisit the problem and present a weakened assumption that guarantees the convergence of the row action algorithm, furthermore give a new primal-dual iterative algorithm that can be used to solve the Young programming problem. The convergence of the new algorithm is provided without any further assumptions.

1 Introduction

Our paper presents a convergent algorithm to a symmetric primal-dual convex optimization problem pair. The Young programming problem was introduced in 1998 by Mályusz [26] and in 2000 by Kas et al [23]. The authors chose this name to emphasize the fundamental role of the Young inequality [32] in their formulation. This problem is a transformed version of the separable Bregman divergence minimization problem (under linear constraints) and is deeply related to this problem class. Therefore the problems introduced in [30], [2], [1], [4] have a strong connection to the Young programming problem, but in the previous case the objective function is a sum of separable Bregman functions. The approach of Mályusz and Kas et al. follows the methods of Klafszky [24] and Terlaky [31], and the symmetric primal-dual

forms obtained by them also show some relation to these papers.
Mályusz [26] and Kas et al [23] introduced the dual of the Young programming problem, proved the duality theorem and showed that under some additional constraints, both a primal and a dual row action method can be constructed to solve it. The new formulation lead to a very natural, symmetric dual problem. Preliminaries of the duality theorem were also present in their articles [22] and [21]. Since the problem is a reformulation of the separable Bregman divergence minimization problem under linear constraints, there are numerous related classical results in connection with the duality theorem and the presented algorithms as well. The same row-action type algorithm was previously used by Censor [8] and Csiszár [13] to solve linear inverse problems. The relation of the problem to other convex optimization problem classes was examined in [5] by Boratas et al. They showed that there exists a Young programming problem instance which does not belong to the extended entropy programming problem class. A generalized problem class was examined by Csiszár et al. [16], [17].
Pietra et al. introduced a different formulation of the problem and also developed the duality theory [18]. Their result was further improved in [4], and applied in [12]. Recent results regarding to the dual of this problem can be found in [29], [28]. The formulation used by Kas et al. gives a more natural, symmetric formulation of the problem.
Bregman divergences were introduced in [6]. An extensive bibliography on different divergence functions is presented in [3]. A new, detailed summary on Bregman divergence functions can be found in [30].
Row action methods belong to the class of alternating projection methods. An introduction to alternating projection methods including row action methods can be found in [19] and in [9]. A survey on projection methods for convex feasibility problems was presented in [10]. Bregman's work in the area [6] was followed by Censor et al. [8], [11], after this, the application of row action methods became widespread in convex optimization.
Since the Young programming problem class contains various frequently used divergence minimization problems as a subclass, possible applications of this theory and algorithms arise from machine learning [27], [12], information theory, architecture [26], economy etc. Linear inverse problems are frequently used in image processing ([15], [22], [7]). Young programming can be applied to approximate linear programming problems [23]. Several applications can be found in clustering [1], [2] as well.
Here we would like to point out that the known algorithms to the poblem in ([12], [15]) all use further assumptions to provide convergence, meanwhile the algorithm introduced in this article converges without any further assump-

tions. In next two chapters, in order to present the problem, we follow the main steps of [23].

2 Divergence and inaccuracy functions

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous, strictly decreasing function, and consider the curve $\Gamma_\varphi = \{(x, \varphi(x)) : x \in \mathbb{R}\}$. The following definition offers a way to describe how much an arbitrary point (u, v) of the plane is "away" from the curve Γ_φ.

Remark 2.1. A similar description can be introduced if $\varphi : \mathbb{R} \to \mathbb{R}$ is a continuous, strictly increasing function. Our choice of strictly decreasing φ functions was motivated by a possible application, namely the approximation of linear programming, see Mályusz [26].

Definition 2.1 (Inaccuracy function). Let $(u, v) \in \mathbb{R}^2$ and denote $\psi = \varphi^{-1}$, then $S_\varphi : \mathbb{R} \to \mathbb{R}$

$$S_\varphi(u, v) = (u - \psi(v))v - \int_{\psi(v)}^{u} \varphi(t)dt \quad (= (v - \varphi(u))u - \int_{\varphi(u)}^{v} \psi(t)dt).$$

is called the inaccuracy of (u, v) with respect to Γ_φ.

Geometrically $S_\varphi(u, v)$ is the area of the shaded region in Figure 1.

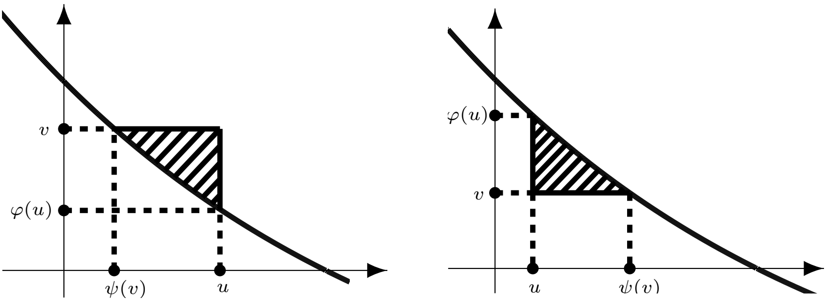


Figure 1: Geometric interpretation of $S_\varphi(u, v)$

From the Young inequality [32] it follows that

$$S_\varphi(u, v) \geq 0 \quad \text{for every} \quad (u, v) \in \mathbb{R}^2$$

and equality holds if and only if $v = \varphi(u)$. The geometric interpretation or a straightforward computation also proves the statement. The inaccuracy can be defined in $\mathbb{R}^{2n}$ as follows.

Definition 2.2 (Inaccuracy function of two vectors). Let $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^{2n}$, $\mathbf{u} = (u_1, u_2, \ldots, u_n)^T$, $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T$, then

$$\mathbf{S}_\varphi : \mathbb{R}^{2n} \to \mathbb{R} : \quad \mathbf{S}_\varphi(\mathbf{u}, \mathbf{v}) := \sum_{j=1}^{n} S_\varphi(u_j, v_j)$$

is called the inaccuracy of $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^{2n}$ with respect to Γ_φ, i.e. the inaccuracy is computed coordinatewisely.

The separable Bregman divergence function in $\mathbb{R}^{2n}$ can be defined in the following way.

Definition 2.3 (Separable Bregman divergence function). Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. The separable Bregman divergence function $D_F : \mathbb{R}^{2n} \to \mathbb{R}$ is defined as

$$\begin{aligned} D_F(\mathbf{x} \| \mathbf{y}) &= \sum_{j=1}^{n} \left(f(y_j) + f'(y_j)(x_j - y_j) - f(x_j)\right) = \\ &= \sum_{j=1}^{n} \left(f'(y_j)(x_j - y_j) + \int_{x_j}^{y_j} f'(t)dt \right). \end{aligned}$$

Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f' = \varphi$. It is clear that the inaccuracy function is a transformed version of the separable Bregman divergence function, i.e.

$$S_\varphi(u, \varphi(v)) = D_f(u||v), \quad \text{for all } u, v \in \mathbb{R},$$

or alternatively

$$S_\varphi(u, v) = D_f(u||\psi(v)), \quad \text{for all } u, v \in \mathbb{R}.$$

Some special cases, divergence functions known in the literature, are obtained by choosing

- $\varphi(t) = -\ln t - 1$, then $S_\varphi(u, \varphi(v)) = u \ln \frac{u}{v} - u + v$ known as I-divergence, introduced by Kullback [25],
- $\varphi(t) = \frac{1}{t}$, then $S_\varphi(u, \varphi(v)) = \ln \frac{v}{u} - \frac{u}{v} - 1$ known as Itakura-Saito divergence [20],

- $\varphi(t) = t^{\alpha-1}$, $\alpha < 1$, $\alpha \neq 0$, then $S_\varphi(u, \varphi(v)) = \frac{1}{\alpha}(v^\alpha - u^\alpha + \alpha v^{\alpha-1}(u - v))$ known as Csiszar's α-divergence [13].

Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous, strictly decreasing function with $\lim_{x\to 0^+} \varphi(x) = \infty$, $\lim_{x\to\infty} \varphi(x) = 0$. Let $\Gamma_\varphi = \{(x, \varphi(x)) : x \in \mathbb{R}_+\}$ and denote $\psi = \varphi^{-1}$. According to Definition 2.2, the inaccuracy of $(x, z) \in \mathbb{R}^2$, $x > 0$, $z > 0$ with respect to Γ_φ is

$$S_\varphi(x, z) = (x - \psi(z))z - \int_{\psi(z)}^{x} \varphi(t)dt.$$

Remark 2.2. Elementary computation proves the following properties of $S_\varphi(x, z)$.

(i) $S_\varphi(x, z)$ is strictly convex function both in $x > 0$ and in $z > 0$.

(ii) $S_\varphi(x, z) \geq 0$ for every $x > 0, z > 0$, and $S_\varphi(x, z) = 0$ if and only if $x = \psi(z)$.

(iii) $\lim_{x\to 0^+} \frac{\partial}{\partial x} S_\varphi(x, z) = -\infty$ for every $z > 0$.

(iv) $\lim_{z\to 0^+} \frac{\partial}{\partial z} S_\varphi(x, z) = -\infty$ for every $x > 0$.

(v) $\lim_{z\to\infty} S_\varphi(x, z) = \infty$ for every $x > 0$.

(vi) $S_\varphi(x, z) = S_\psi(z, x)$ for every $x > 0$ and $z > 0$.

3 Duality

Let us consider the following feasibility problem, which is in fact the Young programming problem in an equilibrium form.

Problem 3.1. Let $A \in \mathbb{R}^{m\times n}$ be a full row rank matrix, $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$ be arbitrary but fixed vectors with $\hat{\mathbf{x}} > \mathbf{0}, \hat{\mathbf{z}} > \mathbf{0}$. Find a feasible solution to the set of constraints below.

$$A\mathbf{x} = A\hat{\mathbf{x}}, \quad \mathbf{z} = \hat{\mathbf{z}} + A^T\mathbf{y} \tag{1}$$

$$\mathbf{x} > \mathbf{0}, \qquad \mathbf{z} > \mathbf{0} \tag{2}$$

$$x_j = \psi(z_j),\ j = 1, 2, \ldots, n. \tag{3}$$

The next lemma states some elementary but crucial observations about Problem 3.1.

Lemma 3.1. *Let* $\hat{\mathbf{x}} > \mathbf{0}, \hat{\mathbf{z}} > \mathbf{0}$ *be arbitrary but fixed vectors,* $\hat{\mathbf{x}}, \hat{\mathbf{z}} \in \mathbb{R}^n$.

(i) If $\mathbf{x}$ *and* $\mathbf{z}$ *satisfy (1) and (2) then* $\mathbf{S}_\varphi(\mathbf{x}, \hat{\mathbf{z}}) + \mathbf{S}_\varphi(\hat{\mathbf{x}}, \mathbf{z}) = \mathbf{S}_\varphi(\mathbf{x}, \mathbf{z}) + \mathbf{S}_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}})$.

(ii) If $\mathbf{x}$ *and* $\mathbf{z}$ *satisfy (1) and (2) then* $\mathbf{S}_\varphi(\mathbf{x}, \hat{\mathbf{z}}) + \mathbf{S}_\varphi(\hat{\mathbf{x}}, \mathbf{z}) \geq \mathbf{S}_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}})$.

(iii) If $\mathbf{x}$ *and* $\mathbf{z}$ *satisfy (1), (2) and (3) then* $\mathbf{S}_\varphi(\mathbf{x}, \hat{\mathbf{z}}) + \mathbf{S}_\varphi(\hat{\mathbf{x}}, \mathbf{z}) = \mathbf{S}_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}})$.

Elementary computation proves (i), (ii) and (iii).
Let $\mathcal{P}^+ = \{\mathbf{x} : A\mathbf{x} = A\hat{\mathbf{x}},\ \mathbf{x} > \mathbf{0}\}$ and $\mathcal{D}^+ = \{\mathbf{z} :\ \exists \mathbf{y} \text{ s.t. } \mathbf{z} = \hat{\mathbf{z}} + A^T\mathbf{y},\ \mathbf{z} > \mathbf{0}\}$ be the set of primal and dual feasible solutions, respectively. The following definition presents three equivalent settings of the Young programming problem.

Definition 3.1. Let $\hat{\mathbf{x}} > \mathbf{0}, \hat{\mathbf{z}} > \mathbf{0}$ be arbitrary but fixed vectors.

(i) (Equilibrium form) Find a feasible solution to 1,2 and 3.

(ii) (Optimization form) Find a feasible solution to 1,2 such that $\sum_{j=1}^n S_\varphi(x_j, z_j)$ is minimal.

(iii) (Primal-Dual form) Find solutions to both problems below

$$\begin{array}{cc} \text{primal} & \text{dual} \\ \min\left\{\sum_{j=1}^n S_\varphi(x_j, \hat{z}_j) : \mathbf{x} \in \mathcal{P}^+\right\} & \min\left\{\sum_{j=1}^n S_\varphi(\hat{x}_j, z_j) : \mathbf{z} \in \mathcal{D}^+\right\}. \end{array}$$

The next theorem presents three equivalent forms of the duality theorem corresponding to the problem settings in Definition 3.1 .

Theorem 3.2. (Duality theorem) *Let* $\hat{\mathbf{x}} > \mathbf{0}, \hat{\mathbf{z}} > \mathbf{0}$ *be arbitrary but fixed vectors.*

(i) The system 1,2, 3 is feasible.

(ii) There are unique $\mathbf{x}^* \in \mathcal{P}^+$ *and* $\mathbf{z}^* \in \mathcal{D}^+$ *vectors such that* $\sum_{j=1}^n S_\varphi(x_j^*, z_j^*) = 0$.

(iii) There exist $\mathbf{x}^* \in \mathcal{P}^+$ *and* $\mathbf{z}^* \in \mathcal{D}^+$ *unique optimal solutions to the primal and the dual problems, respectively. Furthermore* $x_j^* = \psi(z_j^*),\ j = 1, 2, \ldots, n$.

The proof of Theorem 3.2 can be found in [23].
The next corollary points out that the optimal solutions of the primal and dual problems do not depend on the choice of parameter vectors $\hat{\mathbf{x}} \in \mathcal{P}_+$ and $\hat{\mathbf{z}} \in \mathcal{D}_+$.

Corollary 3.1. There exist unique $\mathbf{x}^* \in \mathcal{P}^+, \mathbf{z}^* \in \mathcal{D}^+$ such that $x_j^* = \psi(z_j^*),\ j = 1, 2, \ldots, n$ and

$$\mathbf{S}_\varphi(\mathbf{x}^*, \mathbf{z}) + \mathbf{S}_\varphi(\mathbf{x}, \mathbf{z}^*) = \mathbf{S}_\varphi(\mathbf{x}, \mathbf{z}), \quad \forall \mathbf{x} \in \mathcal{P}^+,\ \forall \mathbf{z} \in \mathcal{D}^+.$$

The proof is obvious from Theorem 3.2.
Another important implication of the duality theorem is noted in the following corollary.

Corollary 3.2. Let $\hat{\mathbf{x}}, \hat{\mathbf{z}}$ be given and denote $\mathbf{x}^* = \operatorname{argmin}\{\mathbf{S}_\varphi(\mathbf{x}, \hat{\mathbf{z}}) : \mathbf{x} \in \mathcal{P}^+\}$ and $\mathbf{z}^* = \operatorname{argmin}\{\mathbf{S}_\varphi(\hat{\mathbf{x}}, \mathbf{z}) : \mathbf{z} \in \mathcal{D}^+\}$. Let us denote the rowspace of A by $\mathcal{L}$, the nullspace of A by $\mathcal{L}^\perp$. Suppose that $\mathcal{L}' \subseteq \mathcal{L}$ and denote $\mathbf{z}'^* = \operatorname{argmin}\{\mathbf{S}_\varphi(\hat{\mathbf{x}}, \mathbf{z}) : \mathbf{z} \in \hat{\mathbf{z}} \oplus \mathcal{L}',\ \mathbf{z} > \mathbf{0}\}$. Then

$$\mathbf{x}^* = \operatorname{argmin}\{\mathbf{S}_\varphi(\mathbf{x}, \mathbf{z}'^*) : \mathbf{x} \in \hat{\mathbf{x}} \oplus \mathcal{L}^\perp,\ \mathbf{x} > \mathbf{0}\}.$$

Since $\mathbf{z}'^* \in \hat{\mathbf{z}} \oplus \mathcal{L}' \subseteq \hat{\mathbf{z}} \oplus \mathcal{L}$, the statement follows from Corollary 3.1.

Corollary 3.3. Let $\hat{\mathbf{x}} \in P^+$, $\hat{\mathbf{z}} \in D^+$, $\mathbf{a}^{(k)}$ denotes the k^{th} row of A and $\mathbf{b}^{(r)}$ represents the r^{th} row of $null(A)^T$. Let $\mathbf{a}^{(k)}\mathbf{x}^{(k)} = \mathbf{a}^{(k)}\hat{\mathbf{x}}$, $\mathbf{x}^{(k)} > 0$ and $\mathbf{b}^{(r)}\mathbf{z}^{(r)} = \mathbf{b}^{(r)}\hat{\mathbf{z}}$, $\mathbf{z}^{(r)} > 0$, then

$$\mathbf{S}_\varphi(\mathbf{x}^{(k)}, \hat{\mathbf{z}}) + \mathbf{S}_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(r)}) = \mathbf{S}_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}}) + \mathbf{S}_\varphi(\mathbf{x}^{(k)}, \mathbf{z}^{(r)}). \tag{4}$$

Observations of Corollary 3.1, 3.2 and 3.3 give rise to the row-action method proposed in Section 5.

4 Convergence of the row action algorithm

In [26] and [23] a primal row action method has been introduced to provide a solution to the problem. The row action method for separable objective functions without duality was introduced earlier by Csiszďż˝r [15] in context with linear inverse problems. By introducing duality, Kas et al. presented a new sufficient condition that provides the convergence of the row action method.

In contrast with the usual notation, here the iterates given by the primal algorithm are dual feasible solutions, and the vectors given by the dual version are primal feasible solutions, to be consistent with the notation of the previous articles.

Algorithm 1 Csiszár's algorithm

Input: $A \in R^{nxm}$ with full row rank, $\hat{\mathbf{z}} \in \mathbf{R^n}$, $\hat{\mathbf{x}} \in \mathbf{R^n}$, $\epsilon > 0$
$\mathbf{x}^0 := \hat{\mathbf{x}}$
$k := 1$
$i := k \pmod m$
repeat
  $\mathbf{x}^{(k)} = \text{argmin}\{S_\varphi(\mathbf{x}, \mathbf{z}^{k-1}) : \mathbf{a}^{(i)}\mathbf{x} = \mathbf{a}^{(i)}\hat{\mathbf{x}}\ \}$
  $z_j^{(k)} = \varphi(x_j^k) \quad (\forall j = 1, 2, \ldots, n)$
  $k := k + 1$
  $i := k \pmod m$
until $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| < \epsilon$

As it can be seen from the description of the algorithm, the kth step - where $i = k \pmod m$ - is a minimization of the primal objective function along the ith row of matrix A. This in dual terms is a one-dimensional minimization of the dual objective function along the row vector $\mathbf{a}^{(i)}$ for all k. In dual terms, this algorithm can be viewed as a dir action method.
To make our implementations more efficient, it is important to notice that the primal feasible set can be transformed to the form given in the the dual problem, and this way the primal iterations can also be computed as one dimensional minimizations along the row vectors of $\text{null}(A)^T$. It was showed by Mályusz [26] and Kas et al. in [23] that Csiszár's algorithm can be considered as a primal algorithm and using the duality theorem the convergence result of the algorithm can be extended to the dual side.

$$\left.\begin{array}{rl} A\mathbf{x} & = A\hat{\mathbf{x}} \\ \mathbf{x} & > \mathbf{0} \end{array}\right\} \Rightarrow \qquad \left.\begin{array}{rl} \mathbf{x} & = \hat{\mathbf{x}} + \text{null}(A)^T\mathbf{y} \\ \mathbf{x} & > \mathbf{0} \end{array}\right\}$$

The following theorem and proof is a modified version of the results from [13] and [14] by Csiszár. It provides the convergence of the row action algorithm under weakened sufficient conditions, meaning that the convergence of the algorithm is provided for a larger function class.

Theorem 4.1. *The row action method is convergent if any of the following two assumptions hold:*

P1. the set $\{\mathbf{x} : \mathbf{a}^{(i)}\mathbf{x} = \mathbf{a}^{(i)}\hat{\mathbf{x}}, \mathbf{x} \geq \mathbf{0}\}$ *is bounded for at least one* i *index* $(1 \leq i \leq m)$.

P2. $\lim_{z_j \to 0^+} S_\varphi(\hat{x}_j, z_j) > S_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}})\ \forall j = 1 \dots n.$

Proof. Here we follow the main steps of Csiszár's proof. From Lemma 3.1:

$$S_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}}) = S_\varphi(\mathbf{x}^{(1)}, \hat{\mathbf{z}}) + S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(1)}),$$

$$S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(1)}) = S_\varphi(\mathbf{x}^{(2)}, \mathbf{z}^{(1)}) + S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(2)}),$$

$$\vdots$$

$$S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(k-1)}) = S_\varphi(\mathbf{x}^{(k)}, \mathbf{z}^{(k-1)}) + S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(k)}),$$

$$\vdots$$

Summing up the first N equations:

$$S_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}}) = \sum_{k=1}^{N} S_\varphi(\mathbf{x}^{(k)}, \mathbf{z}^{(k-1)}) + S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(N)})$$

It implies that the sequence $S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(k)})$ is strictly decreasing for every $\hat{\mathbf{x}} \in \mathcal{P}^+$, and $\lim_{k\to\infty} S_\varphi(\mathbf{x}^{(k)}, \mathbf{z}^{(k-1)}) = 0$, therefore

$$x_j^{(k)} \to \psi\left(z_j^{(k-1)}\right), \ \forall j = 1, 2, \dots n. \tag{5}$$

Furthermore (if $N \to \infty$):

$$S_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}}) = \sum_{k=1}^{\infty} S_\varphi(\mathbf{x}^{(k)}, \mathbf{z}^{(k-1)}) + \lim_{k\to\infty} S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(k)}) \tag{6}$$

Since the sequence $S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(k)})$ is strictly decreasing and has a lower bound, the $\{z_j^{(k)}\}$ sequence has an accumulation point for all $j = 1, 2, \dots, n$ in $[0; K] \subset \mathbb{R}$ for a suitable large K. We have to ensure that $\lim_{k\to\infty} z_j^{(k)} \neq 0\ \forall j = 1, 2, \dots n$. If any of the two conditions hold:

- in case of $P1$.: x_j^k cannot converge to infinity, meaning that $\lim_{k\to\infty} z_j^{(k)} \neq 0\ \forall j = 1, 2, \dots n$.
- in case of $P2$.: z_j^k cannot converge to 0, since the sequence $S_\varphi(\hat{\mathbf{x}}, \mathbf{z}^{(k)})$ is strictly decreasing.

Let $\mathbf{z}^*$ be an accumulation point of $\{\mathbf{z}^{(k)}\}$, and $\{\mathbf{z}^{k_r}\}$ be a subsequence converging to $\mathbf{z}^*$. $\mathbf{z}^* \in D^+$ because of the construction of the algorithm. Let $x_j^* = \psi(z_j^*), \forall j = 1 \dots n$. Then, because of (5) $\mathbf{x}^{(k_r)} \to \mathbf{x}^*$. The sequence $\mathbf{z}^{(k-1)}$ is from a compact subset of D^+ and because of $\lim_{k\to\infty} S_\varphi(\mathbf{x}_k, \mathbf{z}^{(k-1)}) = 0$, $||\mathbf{x}^{(k)} - \mathbf{z}^{(k-1)}|| \to 0$. Therefore $\mathbf{x}^{(k_r)} \to \mathbf{x}^*$ implies that $\mathbf{x}^{k_r+1} \to \mathbf{x}^*, \dots, \mathbf{x}^{(k_r+m-1)} \to \mathbf{x}^*$. Since $\mathbf{x}^{k_r}, \mathbf{x}^{(k_r+1)}, \dots \mathbf{x}^{(k_r+m-1)}$ are from a cyclic permutation of the affine spaces $\mathcal{L}_i = \{\mathbf{x} : \mathbf{a}^{(i)}\mathbf{x} = \mathbf{a}^{(i)}\hat{\mathbf{x}}\}$, $\mathbf{x}^* \in \cap_{i=1}^m \mathcal{L}_i$, meaning that $\mathbf{x}^* \in \mathcal{P}$. We have seen that $\mathbf{x}^* \in \mathcal{P}$, $\mathbf{z}^* \in \mathcal{D}$ and $x_j^* = \psi(z_j^*), j = 1, 2, \dots n$. Then due to the duality theorem, $\mathbf{x}^*$ and $\mathbf{z}^*$ are the unique optimal solutions to the primal, and dual problems, respectively. □

It can be proved by using Remark 2.2, that the new condition is indeed a weakened version of the previous one, meaning that if the constraint $\int_0^a \psi(t)dt = \infty$, for some $a > 0$ is true for the $\psi(t)$ function, then the new constraint $\lim\limits_{z_j\to 0^+} S_\varphi(\hat{x}_j, z_j) > S_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}})\ \forall j = 1 \dots n$ is also satisfied.

As it was pointed out by Kas et al. in [23], using the duality theorem, the following sufficient condition can be derived:

D2. $\lim\limits_{x\to 0^+} S_\varphi(x, z) = \infty$, for all $z \in \mathbb{R}_+$.

Using the dual problem pair we can also derive a weakened sufficient condition *(D2)* to the convergence of the algorithm:

D1. the set $\{\mathbf{z} : \mathbf{b}^{(r)}\mathbf{z} = \mathbf{b}^{(r)}\hat{\mathbf{z}}, \mathbf{z} \geq \mathbf{0}\}$ is bounded for at least one i index $(1 \leq r \leq m)$, where $\mathbf{b}^{(r)}$ is the r^{th} a row of matrix $\mathbf{B} = null(\mathbf{A})^T$.

D2. $\lim\limits_{x_j\to 0^+} S_\varphi(x_j, \hat{z}_j) > S_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}})\ \forall j = 1 \dots n$.

he main difference between *P2* and *P2'* is that *P2* only depends on the choice of the ψ function, meanwhile *P2'* depends on the starting feasible vectors $\hat{\mathbf{x}}, \hat{\mathbf{z}}$ as well.

5 A new primal-dual algorithm

In this section $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ denote primal and dual feasible vectors, respectively. i denotes the iteration counter of the row action algorithm, $\mathbf{a}^{(k)}$ is the k^{th} row of matrix A, $\mathbf{b}^{(r)}$ is the r^{th} row of matrix $B = null(A)^T$ and j denotes j^{th} element of vector $\mathbf{x}$ or $\mathbf{z}$.

Row-action type algorithms as defined in [8] are the ones that use only the previous iterate in each iterative step, and access is required to only one row of the system of equations of the constraint set. The following algorithm is consecutively a row action method on the primal side and a row action method on the dual side for the solution of Problem 3.1 as it is stated in Problem 3.1(iii). Kas et. al [23] showed that a row action step in primal side is a dir action step on the dual side and in each iteration we have a dual feasible solution.

Algorithm 2 Primal-Dual Algorithm

Input: Let $A \in \mathbb{R}^{m\times n}$ be a full row rank matrix, $\hat{\mathbf{x}} \in \mathcal{P}^+$, $\hat{\mathbf{z}} \in \mathcal{D}^+$, $\epsilon > 0$
$m := rank(A), q := n - m,\ B \in \mathbb{R}^{q\times n} = null(A)$
$\hat{\mathbf{x}}_1 := \hat{\mathbf{x}},\ \hat{\mathbf{z}}_{D_1} := \hat{\mathbf{z}}$
$i := 1$
repeat
Take m primal row action (dual dir action) steps starting from $(\hat{\mathbf{x}}_i, \hat{\mathbf{z}}_{D_i})$ to obtain $\mathbf{z}_{D_i}^{(k)} \in \mathcal{D}^+$:
for $k = 1..m$ **do**
Calculate $(\mathbf{x}_i^{(k)}, \mathbf{z}_{D_i}^{(k)})$, where

$$\mathbf{x}_i^{(k)} = \text{argmin}\{S_\varphi(\mathbf{x}_i^{(k)}, \mathbf{z}_{D_i}^{(k-1)}) : \mathbf{a}^{(k)}\mathbf{x}_i^{(k)} = \mathbf{a}^{(k)}\hat{\mathbf{x}}_i\}$$

$$(z_{D_i})_j^{(k)} = \varphi(x_{i_j}^{(k)}), \quad j = 1, 2, ..., n.$$

end for
$\hat{\mathbf{z}}_i := \mathbf{z}_{D_i}^{(m)}$
Take r dual row action steps (primal dir action) starting from $(\hat{\mathbf{x}}_{P_{i-1}}, \hat{\mathbf{z}}_{D_i})$ to obtain $\mathbf{x}_{P_i}^{(r)} \in \mathcal{P}^+$:
for $r = 1..q$ **do**
Calculate $(\mathbf{x}_{P_i}^{(r)}, \mathbf{z}_i^{(r)})$, where

$$\mathbf{z}_i^{(r)} = \text{argmin}\{S_\varphi(\mathbf{x}_{P_i}^{(r)}, \mathbf{z}_i^{(r)})) : \mathbf{b}^{(r)}\mathbf{z}_i^{(r)} = \mathbf{b}^{(r)}\hat{\mathbf{z}}_i\}$$

$$(x_{P_i})_j^{(r)} = \psi(z_{i_j}^{(r)}), \quad j = 1, 2, ..., n.$$

end for
$\hat{\mathbf{x}}_{P_i} := \mathbf{x}_{P_i}^{(r)}$
$i := i + 1$
until $S_\varphi(x_{P_i}^r, z_{D_i}^m) < \epsilon$

Two important observations referring to the primal and dual algorithms coming from the transitivity property of Bregman functions (see Csiszár [15] in Section 4) are summarized in the following remarks:

Remark 5.1. If we exchange the $\hat{\mathbf{x}}$ vector to a primal feasible $\tilde{\mathbf{x}}$ vector, the $\mathbf{z}^k$ iterates given by the dual algorithm do not change.

Remark 5.2. We can exchange $\hat{\mathbf{z}}$ to a dual feasible $\tilde{\mathbf{z}}$ without changing the primal optimal solution of the Young programming problem.

Theorem 5.1. *The primal-dual algorithm converges to the optimal solution of the Young programming problem.*

Proof. Let $\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}$ and $\hat{\mathbf{z}}_{D_1} = \hat{\mathbf{z}}$ and $i = 1...N$. The i^{th} row action iteration on the primal side gives the following:

$$\begin{aligned}
S_\varphi(\hat{\mathbf{x}}_i, \hat{\mathbf{z}}_{D_i}) &= S_\varphi(\mathbf{x}_i^{(1)}, \hat{\mathbf{z}}_{D_i}) + S_\varphi(\hat{\mathbf{x}}_i, \mathbf{z}_{D_i}^{(1)}) \\
S_\varphi(\hat{\mathbf{x}}_i, \mathbf{z}_{D_i}^{(1)}) &= S_\varphi(\mathbf{x}_i^{(2)}, \mathbf{z}_{D_i}^{(1)}) + S_\varphi(\hat{\mathbf{x}}_i, \mathbf{z}_{D_i}^{(2)}) \\
&\vdots \\
S_\varphi(\hat{\mathbf{x}}_i, \mathbf{z}_{D_i}^{(k-1)})) &= S_\varphi(\mathbf{x}_i^{(k)}, \mathbf{z}_{D_i}^{(k-1)}) + S_\varphi(\hat{\mathbf{x}}_i, \mathbf{z}_{D_i}^{(k)}) \\
&\vdots \\
S_\varphi(\hat{\mathbf{x}}_i, \mathbf{z}_{D_i}^{(m-1)})) &= S_\varphi(\mathbf{x}_i^{(m)}, \mathbf{z}_{D_i}^{(m-1)}) + S_\varphi(\hat{\mathbf{x}}_i, \mathbf{z}_{D_i}^{(m)})
\end{aligned}$$

Let $\hat{\mathbf{z}}_i = \mathbf{z}_{D_i}^{(m)}$ and $\hat{\mathbf{x}}_{P_i} = \hat{\mathbf{x}}_i$, the row action iteration on the dual side consists of the following steps:

$$\begin{aligned}
S_\varphi(\hat{\mathbf{x}}_{P_i}, \hat{\mathbf{z}}_i) &= S_\varphi(\hat{\mathbf{x}}_{P_i}, \mathbf{z}_i^{(1)}) + S_\varphi(\mathbf{x}_{P_i}^{(1)}, \hat{\mathbf{z}}_i) \\
S_\varphi(\mathbf{x}_{P_i}^{(1)}, \hat{\mathbf{z}}_i) &= S_\varphi(\mathbf{x}_{P_i}^{(1)}, \mathbf{z}_i^{(2)}) + S_\varphi(\mathbf{x}_{P_i}^{(2)}, \hat{\mathbf{z}}_i) \\
&\vdots \\
S_\varphi(\mathbf{x}_{P_i}^{(r-1)}, \hat{\mathbf{z}}_i) &= S_\varphi(\mathbf{x}_{P_i}^{(r-1)}, \mathbf{z}_i^{(r)}) + S_\varphi(\mathbf{x}_{P_i}^{(r)}, \hat{\mathbf{z}}_i) \\
&\vdots \\
S_\varphi(\mathbf{x}_{P_i}^{(q-1)}, \hat{\mathbf{z}}_i) &= S_\varphi(\mathbf{x}_{P_i}^{(q-1)}, \mathbf{z}_i^{(q)}) + S_\varphi(\mathbf{x}_{P_i}^{(q)}, \hat{\mathbf{z}}_i).
\end{aligned}$$

These equations can be summarized as follows when $N \to \infty$.

$$S_\varphi(\hat{\mathbf{x}}, \hat{\mathbf{z}}) = \sum_{i=1}^{\infty} \sum_{k=1}^{m-1} S_\varphi(\mathbf{x}_i^{(k)}, \mathbf{z}_{Di}^{(k-1)}) + \sum_{i=1}^{\infty} \sum_{r=1}^{q-1} S_\varphi(\mathbf{x}_{Pi}^{(r-1)}, \mathbf{z}_i^{(r)}) +$$
$$+ \sum_{i=1}^{\infty} \left\{ S_\varphi(\mathbf{x}_i^{(m)}, \mathbf{z}_{Di}^{(m-1)}) + S_\varphi(\mathbf{x}_{Pi}^{(q-1)}, \mathbf{z}_i^{(q)}) \right\} + \lim_{i \to \infty} S_\varphi(\mathbf{x}_{Pi}^{(q)}, \hat{z}_i)$$

Since the left hand side of the equation is a finite number and the elements of the numerical series on the right hand side are nonnegative, the three sums on the right side must be finite and the limit of the sequence is 0. Moreover Corollary 3.3 provides that:

$$S_\varphi(\mathbf{x}_i^{(m)}, \mathbf{z}_{Di}^{(m-1)}) + S_\varphi(\mathbf{x}_{Pi}^{(q-1)}, \mathbf{z}_i^{(q)}) = S_\varphi(\mathbf{x}_i^{(m)}, \mathbf{z}_i^{(q)}) + S_\varphi(\mathbf{x}_{Pi}^{(q-1)}, \mathbf{z}_{Di}^{(m-1)}) \tag{7}$$

where $\mathbf{x}_{Pi}^{(q-1)} \in P$, $\mathbf{z}_{Di}^{(m-1)} \in D$, and

$$\lim_{i \to \infty} S_\varphi(\mathbf{x}_i^{(m)}, \mathbf{z}_i^{(q)}) = 0$$

$$\lim_{i \to \infty} S_\varphi(\mathbf{x}_{P_i}^{(q-1)}, \mathbf{z}_{D_i}^{(m-1)}) = 0 \tag{8}$$

Since $S_\varphi(\mathbf{x}_{P_i}^{(q-1)}, \mathbf{z}_{D_i}^{(m-1)}))$ is strictly monotonically decreasing, the sequences $\mathbf{z}_{D_ij}^{(m-1)}$ and $\mathbf{x}_{P_ij}^{(q-1)}$ have accumulation points on $[0; K]$ for an arbitrary large K for all j indices. Because of 8 and

$$\lim_{z_j \to 0} S_\varphi(\hat{x}_j, z_j) > S_\varphi(\mathbf{x}_{P_i}^{(q-1)}, \mathbf{z}_{D_i}^{(m-1)}) \ \forall j = 1 \dots n$$

and

$$\lim_{x_j \to 0} S_\varphi(x_j, z_j) > S_\varphi(\mathbf{x}_{P_i}^{(q-1)}, \mathbf{z}_{D_i}^{(m-1)}) \ \forall j = 1 \dots n$$

$(\lim_{i \to \infty} \mathbf{z}_{D_ij}^{(m-1)} \neq 0 \ \forall j = 1, 2, \dots n)$ and $(\lim_{i \to \infty} \mathbf{x}_{P_ij}^{(q-1)} \neq 0 \ \forall j = 1, 2, \dots n)$ holds, and from Theorem 3.2,

$$\mathbf{x}^* = \lim_{i \to \infty} \mathbf{x}_{P_i}^{(q-1)}$$

and

$$\mathbf{z}^* = \lim_{i \to \infty} \mathbf{z}_{D_i}^{(m-1)}$$

are optimal solutions of the Young programming problem. □

By applying the new primal-dual algorithm, we can generate a primal-dual feasible pair that satisfies the new sufficient condition $D2$, independently from the properties of matrix A. It means that the new algorithm can also be used as a first phase before applying the primal or dual algorithms from [23], to generate a pair that satisfies the weakened sufficient conditions.

6 Numerical results

6.1 A function that doesn't satisfy the original conditions

As it was mentioned in [23] and [13], many of the famous divergence functions or their Fenchel conjugate - as a dual objective function - satisfy the $P2$ or $D2$ conditions. However it is not the case for all functions. For example, if we choose φ to be the function

$$\varphi(x) = \ln\left(\frac{1}{x^2} + 1\right),$$

the inverse function ψ is

$$\psi(z) = \frac{1}{\sqrt{e^z - 1}}.$$

This $\varphi(x)$ has all the properties listed in the definition of the Young programming problem. In this case neither $A2$ nor $A2'$ can be satisfied since

$$\int_0^\infty \psi(t) = \int_0^\infty \frac{1}{\sqrt{e^t - 1}} dt = \pi < \infty.$$

As it was mentioned, the fulfilment of $B2$ and $B2'$ depends on the choice of $\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ as well.

Example 6.1. Let $A = \begin{bmatrix} 2 & 3 & 5 & -1 \\ 4 & 2 & 6 & -3 \end{bmatrix}$. In this case the $P1$ condition doesn't hold.

Let $\hat{\mathbf{x}} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, $\hat{\mathbf{z}}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}$, $\hat{\mathbf{z}}_2 = \begin{pmatrix} 2 \\ 3 \\ 2 \\ 1 \end{pmatrix}$. In case of $\hat{\mathbf{z}}_1$, $P2$ is met, but it is not satisfied for $\hat{\mathbf{z}}_2$.

To solve the problem with the input parameters $A, \hat{\mathbf{x}}, \hat{\mathbf{z}}_2$ none of the previous algorithms can be used in their original form, but we can apply the new primal-dual algorithm, because in this case it is not necessary to satisfy any of the above conditions.

6.2 Implementation, numerical results

To be able to compare the performance of the new primal-dual algorithm to the primal and dual row action methods, we implemented the three algorithms in Matlab.
The following table shows the results for the data given in Example 6.1 with $\hat{\mathbf{z}}_2$ and $\epsilon = 10^{-8}$. As it can be seen, the desired precision was achieved in 8 iterations, in 6 seconds.

algorithms	*primal*	*dual*	*primal − dual*
epsilon	$\epsilon = 10^{-8}$	$\epsilon = 10^{-8}$	$\epsilon = 10^{-8}$
iteration number	2	5	2
running time (sec)	0,01	0,004	0,002
epsilon	$\epsilon = 10^{-10}$	$\epsilon = 10^{-10}$	$\epsilon = 10^{-10}$
iteration number	19	14711	2
running time (sec)	0,01	9,322	0,002
epsilon	$\epsilon = 10^{-12}$	$\epsilon = 10^{-12}$	$\epsilon = 10^{-12}$
iteration number	99	>200.000	3
running time (sec)	0,05	137,8	0,006

Based on the results of the numerical tests, it can be stated the primal-dual algorithm is more stable than the other two. There were many cases where either the primal or the dual algorithm could find the solution faster than the primal-dual method, but in these cases, the third algorithm (the primal, if the dual was the fastest, and conversely) could not find the solution in reasonable time (and under a limited iteration number). The primal-dual algorithm always gave the correct solution. If we orthogonalized the row vectors of A by using the Gram-Schmidt orthogonalization process, generally the algorithm performed better.

7 Conclusions, further research

In this paper, we revisited the Young programming problem, discussed its relationship to other results from different areas of operations research as well as possible application areas.
We extended the class of functions for which the convergence of the primal and dual row action methods given by Kas et al. can be assured by proving a weakened sufficient condition. We also showed that our new algorithm can be used as a first phase to construct feasible solutions that satisfy the new, weakened sufficient condition before using these algorithms as a second phase.

The main result of this paper is the introduction of the new primal-dual projection algorithm for which the convergence to the optimal solution of the Young programming problem (or alternatively to the separable Bregman divergence minimization problem under linear constraints) can be provided without any further assumptions. We also carried out preliminary numerical tests to examine the efficiency of the algorithm in practice.
In our future work, we would like to examine and improve the efficiency of the algorithm. To further our research, we intend to examine the complexity of the Young programming problem for different inaccuracy functions.

Acknowledgements

This research has been partially supported by the Hungarian Research Fund, OTKA (grant no. NKFIH 125700).

The research of A. Varga has been partially supported by the Higher Education Excellence Program of the Ministry of Human Capacities in the frame of Artificial Intelligence research area of Budapest University of Technology and Economics (BME FIKP-MI/FM).

References

[1] Arindam Banerjee, Inderjit Dhillon, Joydeep Ghosh, Srujana Merugu, and Dharmendra S Modha.
A generalized maximum entropy approach to Bregman co-clustering and matrix approximation.
Journal of Machine Learning Research, 8(Aug):1919–1986, 2007.

[2] Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh.
Clustering with Bregman divergences.
Journal of machine learning research, 6(Oct):1705–1749, 2005.

[3] Michele Basseville.
Divergence measures for statistical data processing – An annotated bibliography.
Signal Processing, 93(4):621–633, 2013.

[4] Heinz H Bauschke.
Duality for Bregman projections onto translated cones and affine subspaces.
Journal of Approximation Theory, 121(1):1–12, 2003.

[5] Zehra Boratas, Tibor Illés, and Péter Kas.
Entropy and Young programs: relations and self-concordance.
Central European Journal of Operations Research, 10(4), 2002.

[6] Lev M Bregman.
The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming.
USSR computational mathematics and mathematical physics, 7(3):200–217, 1967.

[7] Christoph Brune, Alex Sawatzky, and Martin Burger.
Primal and dual Bregman methods with application to optical nanoscopy.
International Journal of Computer Vision, 92(2):211–229, 2011.

[8] Yair Censor.
Row-action methods for huge and sparse systems and their applications.
SIAM review, 23(4):444–466, 1981.

[9] Yair Censor and Andrzej Cegielski.
Projection methods: an annotated bibliography of books and reviews.
Optimization, 64(11):2343–2358, 2015.

[10] Yair Censor, Wei Chen, Patrick L Combettes, Ran Davidi, and Gabor T Herman.
On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints.
Computational Optimization and Applications, 51(3):1065–1088, 2012.

[11] Yair Censor and Arnold Lent.
An iterative row-action method for interval convex programming.
Journal of Optimization theory and Applications, 34(3):321–353, 1981.

[12] Michael Collins, Robert E Schapire, and Yoram Singer.

Logistic regression, AdaBoost and Bregman distances.

Machine Learning, 48(1-3):253–285, 2002.

[13] Imre Csiszár.

I-divergence geometry of probability distributions and minimization problems.

The Annals of Probability, pages 146–158, 1975.

[14] Imre Csiszár.

Entrópiamaximalizálás és rokon módszerek: Axiomatika és algoritmusok.

Szigma, XXIV(3-4), 1993.

[15] Imre Csiszár et al.

Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems.

The annals of statistics, 19(4):2032–2066, 1991.

[16] Imre Csiszár and František Matúš.

Generalized minimizers of convex integral functionals, Bregman distance, Pythagorean identities.

arXiv preprint arXiv:1202.0666, 2012.

[17] Imre Csiszár and František Matúš.

Convergence of generalized entropy minimizers in sequences of convex problems.

In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pages 2609–2613. IEEE, 2016.

[18] Stephen Della Pietra, Vincent Della Pietra, and John Lafferty.

Duality and auxiliary functions for Bregman distances.

Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE, 2001.

[19] Renâ Escalante and Marcos Raydan.

Alternating projection methods, volume 8.

SIAM, 2011.

[20] Fumitada Itakura.
Analysis synthesis telephony based on the maximum likelihood method.
In *The 6th international congress on acoustics*, pages 280–292, 1968.

[21] Péter Kas and Emil Klafszky.
On the duality of the mixed entropy programming.
Optimization, 27(3):253–258, 1993.

[22] Péter Kas and Emil Klafszky.
On the dual of linear inverse problems.
European Journal of Operational Research, 91(3):634–639, 1996.

[23] Péter Kas, Emil Klafszky, Levente Mályusz, and Gökhan Izbirak.
Approximation of linear programs by Bregman's DF projections.
European Journal of Operational Research, 126(1):69–79, 2000.

[24] Emil Klafszky.
Geometric programming and its applications.
MTA SZTAKI Tanulmányok, 8:in Hungarian, 1974.

[25] Solomon Kullback.
Information theory and statistics.
Courier Corporation, 1997.

[26] Levente Mályusz.
Young programming and its application in engineering practice.
PhD thesis, Budapest University of Technology and Economics, 1998.
Supervisor: Emil Klafszky, in Hungarian.

[27] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka.
f-gan: Training generative neural samplers using variational divergence minimization.
arXiv preprint arXiv:1606.00709, 2016.

[28] Roman Polyak.
Lagrangian transformation and interior ellipsoid methods in convex optimization.
Journal of Optimization Theory and Applications,
164(3):966–992, 2015.

[29] Roman A Polyak.
The Legendre transformation in modern optimization.
In *Optimization and Its Applications in Control and Data Sciences*, pages 437–507. Springer, 2016.

[30] Daniel Reem, Simeon Reich, and Alvaro De Pierro.
Re-examination of Bregman functions and new properties of their divergences.
arXiv preprint arXiv:1803.00641, 2018.

[31] Tamás Terlaky.
On lp programming.
European Journal of Operational Research, 22(1):70–100, 1985.

[32] William Henry Young et al.
On classes of summable functions and their Fourier series.
Proc. R. Soc. Lond. A, 87(594):225–229, 1912.

Selected Operations Research Reports

2019-01 Petra-Renáta Rigó, Eszter Szénási: Interior-point algorithm for linear optimization based on a new search direction

2018-02 Zsolt Darvay, Tibor Illés, Csilla Majoros: Interior-point algorithm for sufficient LCPs based on the technique of algebraically equivalent transformation

2016-03 Petra-Renáta Takács, Zsolt Darvay: Infeasible interior-point method for symmetric optimization using a positive-asymptotic barrier

2016-01 Zsolt Darvay, Petra-Renáta Takács: New interior-point algorithm for symmetric optimization based on a positive-asymptotic barrier function

2015-02 Tibor Illés, Richárd Molnár-Szipai: Strongly polynomial primal monotonic build-up simplex algorithm for maximal flow problems

2014-01 Tibor Illés, Adrienn Nagy: Finiteness of the quadratic primal simplex method when **s**-monotone index selection rules are applied

2013-02 Tibor Illés, Gábor Lovics: Approximation of the Whole Pareto-Optimal Set for the Vector Optimization Problem

2013-01 Zsolt Csizmadia, Tibor Illés, Adrienn Nagy: The s-Monotone Index Selection Rule for Criss-Cross Algorithms of Linear Complementarity Problems

2012-02 Tibor Illés, Adrienn Nagy: Computational aspects of simplex and MBU-simplex algorithms using different anti-cycling pivot rules

2012-01 Tibor Illés, Richárd Molnár-Szipai: On Strongly Polynomial Variants of the MBU-Simplex Algorithm for a Maximum Flow Problem with Nonzero Lower Bounds