

# GENERALIZED DERIVATIVES OF THE OPTIMAL VALUE OF A LINEAR PROGRAM WITH RESPECT TO MATRIX COEFFICIENTS

Daniel De Wolf, Yves Smeers

REPRINT | 3140

## CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: [lidam-library@uclouvain.be](mailto:lidam-library@uclouvain.be)

<https://uclouvain.be/en/research-institutes/lidam/core/core-reprints.html>



# Generalized derivatives of the optimal value of a linear program with respect to matrix coefficients

Daniel De Wolf<sup>a,b,\*</sup>, Yves Smeers<sup>b</sup>

<sup>a</sup> TVES (EA 4477), Université du Littoral Côte d'Opale, 220 avenue de l'Université, F-59140 Dunkerque, France

<sup>b</sup> CORE, Université Catholique de Louvain, Voie du Roman Pays 34, 1348 Louvain-La-Neuve, Belgique

## ARTICLE INFO

### Article history:

Received 15 February 2019

Accepted 12 November 2019

Available online 18 November 2019

### Keywords:

Linear programming

Parametric linear programming

Nondifferentiable programming

## ABSTRACT

We present here a characterization of the Clarke subdifferential of the optimal value function of a linear program as a function of matrix coefficients. We generalize the result of Freund (1985) to the cases where derivatives may not be defined because of the existence of multiple primal or dual solutions.

© 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

In the framework of linear programming, we consider the problem of estimating the variation of the objective function resulting from changes in some matrix coefficients. Our objective is to extend results already available for the right-hand side to this more general problem.

The interpretation of the dual variables as derivatives of the optimal value of the objective function with respect to the elements of the right-hand side is well known in mathematical programming. This result can be extended to the case of multiple dual solutions. The set of all dual solutions is then the subdifferential of the optimal value of the objective function, seen as a convex function of the right-hand side. The object of this paper is to extend these well known results to the derivative of the optimal value of the objective function with respect to matrix coefficients.

It is easy to show on a simple example that the objective function value of a linear program is not a convex function of the matrix coefficients. The subdifferential concept is thus inappropriate here. One must therefore resort to Clarke's notion of a generalized gradient. A characterization of this generalized gradient will be derived and sufficient conditions of existence of the generalized gradient will be given for this particular application of nonsmooth analysis.

The paper is organized as follows. Section 2 presents a study of the literature about generalized derivatives and subdifferentials of optimal function value. Then Section 3 presents the basic definitions and main properties of nonsmooth analysis that will be useful for our application. Then In Section 4 we recall the result of Freund (1985) in the smooth case, namely the gradient of the optimal value function in the case where the optimal primal and dual solutions of the linear problem are unique. Then in Section 5, a complete characterization of the generalized gradient for the case where the primal or dual solutions are not unique is established. Since the Locally Lipschitz property plays an essential role in this characterization, we will give sufficient conditions to prove the Local Lipschitz property. Section 6 presents a practical application of this characterization coming from the gas industry, namely, the two-stage problem of optimal dimensioning and operating of a gas transmission network. Then Section 7 presents some conclusions.

## 2. Generalized derivatives and subdifferentials of optimal value functions

Before establishing the main result concerning the generalized gradient of the optimal value function of a linear program with respect to matrix coefficients, let us say a word about the literature on the problem of the computation of the generalized derivatives and subdifferentials of the optimal value function in general optimization problems.

One of the first papers on the subject was the paper of Gauvin (1979) who considered a general mathematical programming problem with equality and inequality constraints and perturbation of the right-hand side of the constraints, noted  $u_i$  for constraint  $i$ . He estimated the generalized gradient of the optimal value of the

\* Corresponding author at: TVES (EA 4477), Université du Littoral Côte d'Opale, 220 avenue de l'Université, F-59140 Dunkerque, France.

E-mail addresses: [daniel.dewolf@univ-littoral.fr](mailto:daniel.dewolf@univ-littoral.fr) (D. De Wolf), [yves.smeers@uclouvain.be](mailto:yves.smeers@uclouvain.be) (Y. Smeers).

problem considered as a function of the right-hand side perturbation:

$$\begin{aligned} z(u) &= \max_x f(x), x \in \mathbb{R}^n \\ \text{s.t. } & g_i(x) \leq u_i, i = 1, \dots, m \\ & h_i(x) = u_i, i = m + 1, \dots, p \end{aligned}$$

Note that this case corresponds, for the particular case of a linear program, to a perturbation of the right-hand side  $b$  of the linear programming problem since we consider a perturbation (noted  $A$  in our case) that appears as a coefficient of the variables  $x$  of the problem.

Several developments concerning this case (i.e. the perturbation only in the right-hand side of the constraints) where made since that time. For example, the regularity properties of the optimal value function in nonlinear programs where the perturbation parameters appears only in the right-hand side was done by Craven and Janin (1993). For affine case, namely when  $g(x) = Ax + b$ , an expression is given for its directional derivative, not assuming the optimum to be unique. Recently, Höffner, Khan, and Barton (2016) consider the computation of generalized derivatives of dynamic systems with a linear program embedded. They consider the optimal value of a linear program as a function of the right-hand side of the constraints and present an approach to compute an element of the generalized gradient. The approach is illustrated through a large-scale dynamic flux balance analysis example. Last year, Gomez, Höffner, Khan, and Barton (2018) studied the generalized derivatives problem of parametric Lexicographic Linear programs using the lexicographic directional derivative (See Barton, Khan, Stechliniski, and Watson (2018) for a survey on the lexicographic directional derivative). The paper derives generalized derivatives information.

Unfortunately, this results can not be applied to our problem since the perturbation appears only in the right-hand side of the constraints and doesn't appear as a coefficient of the decision's variables.

Rockafellar (1984) considered the directional derivative of the optimal value function in nonlinear programming problem with a perturbation  $u$  that appears in the left-hand side of the constraints:

$$\begin{aligned} z(u) &= \max_x f(x, u), x \in \mathbb{R}^n \\ \text{s.t. } & g_i(x, u) \leq 0, i = 1, \dots, m \\ & h_i(x, u) = 0, i = m + 1, \dots, p \end{aligned} \quad (1)$$

Under the assumption that every optimal solution  $x^*$  satisfies the second order constraints qualification condition, Rockafellar proved that the function  $z(u)$  is locally Lipschitz and finite. Rockafellar also gives an upper bound on the generalized derivative of Clarke of the function  $z(u)$ .

Note also that many developments were made from this original paper for the general case where the perturbation appears in the left-hand side of the constraints. For example, Thibault (1991) considered a general mathematical programming problem in which the constraints are defined by multifunctions and depend on a parameter  $u$ . A special study is done of problems in which the multifunctions defining the constraints take convex values. For these problems, generalized gradients of  $z(u)$  are given in terms of the generalized gradients of the support functions of the multifunctions. Bonnans and Shapiro (2000) studied the first order differentiability analysis of the optimal value function as a function of a parameter that appears in the objective function and in the left-hand side of the constraints. Under a constraint qualification condition, they give an upper bound on the directional derivative. Note that in our case of a linear program, we will give a complete characterization of the generalized gradient, not only upper bound on the directional derivative. Penot (2004) considers the differentia-

bility properties of optimal value functions for the particular case where the perturbation parameter only appears in the objective function. More recently, Mordukhovich, Nam, and Yen (2007) consider the subgradient of marginal functions in parametric mathematical programming. The authors show that the subdifferential obtained for the corresponding marginal value function are given in terms of Lagrange multipliers. Last year, Im (2018) studied the sensitivity analysis for the special case of linear optimization. In particular, he gives conditions for the objective function value of a linear problem to be a Locally Lipschitz function of matrix coefficients. We will use these conditions in our main characterization of the generalized gradient.

In the present paper, we shall give a complete characterization of the generalized gradient for a particular case, namely the linear case, and not only upper bound on directional derivative.

### 3. Basic definitions and properties

This section recalls some basic concepts and properties of nonsmooth optimization useful for our application. An introduction to the first-order generalized derivative can be found in Clarke (1990) for the case of a locally Lipschitz function.

**Definition 3.1.** A function  $f$  from  $\mathbb{R}^n$  (or a subset of  $\mathbb{R}^n$ ) into  $\mathbb{R}$  is locally Lipschitz if for any bounded set  $B$  from the interior of the domain of  $f$  there exists a positive scalar  $K$  such that

$$|f(x) - f(y)| \leq K\|x - y\| \quad \forall x, y \in B$$

where  $|\cdot|$  denotes the absolute value and  $\|\cdot\|$  the usual Euclidian norm.

The locally Lipschitz property can be interpreted as a finite bound on the variation of the function. It is well known that the locally Lipschitz property implies the continuity of  $f$ .

The Rademacher theorem says that a locally Lipschitz function  $f$  has a gradient almost everywhere (i.e. everywhere except on a set  $Z_f$  of zero (Lebesgue) measure on  $\mathbb{R}^n$ ).

**Definition 3.2.** In the locally Lipschitz case, the generalized gradient is defined as the convex hull of all the points  $\lim \nabla f(x^k)$  where  $\{x^k\}$  is any sequence which converges to  $x$  while avoiding the points where  $\nabla f(x)$  does not exist:

$$\partial f(x) = \text{conv} \left\{ \lim_{k \rightarrow \infty} \nabla f(x_k) : x^k \rightarrow x, \nabla f(x^k) \text{ exists} \right\} \quad (2)$$

where  $\text{conv}$  denotes the convex hull.

Another essential concept in nonsmooth optimization is the directional derivative. This notion can also be generalized to the non-convex case.

**Definition 3.3.** The generalized directional derivative of  $f$  evaluated at  $x$  in the direction  $d$  is defined (using the notation of Clarke) as

$$f^\circ(x; d) = \limsup_{y \rightarrow x; t \downarrow 0} \frac{f(y + td) - f(y)}{t}$$

In the convex case, this notion reduces to the classical notion of directional derivative

$$f'(x; d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$

We shall also use the following proposition for the proof of our characterization of the generalized gradient:

**Proposition 3.1.** Let  $f$  be a function from  $\mathbb{R}^n$  into  $\mathbb{R}$  almost everywhere continuously differentiable. Then  $f$  is continuously differentiable at  $x$  if and only if  $\partial f(x)$  reduces to a singleton.

**Proof.** See Clarke (1990).  $\square$

We have the following general remark. The definition of the generalized gradient (2) is only valid for the locally Lipschitz case. If the function is simply almost everywhere differentiable, one can construct examples for which the generalized gradient is not defined. A more general definition based on the cone of normals is given by Clarke (1990) in the case of a lower semi-continuous function.

#### 4. Gradient of the optimal value function

Returning now to our problem, we consider the optimal value of a linear problem as a function of the matrix coefficients:

$$\begin{aligned} z(A) &= \max_x c^T x \\ \text{subject to } &\begin{cases} Ax = b \\ x \geq 0 \end{cases} \end{aligned} \quad (3)$$

where  $c$  is the  $n$ -column vector of objective coefficients,  $x$  is the  $n$ -column vector of variables,  $A$  is the  $m \times n$  matrix of left-hand side coefficients and  $b$  is the  $m$ -column vector of right-hand side coefficients.

We first recall the result for the smooth case. It is established in Freund (1985) under the two following assumptions:

- (H1) The optimal solution of the primal problem (3) is unique;
- (H2) The optimal solution of the dual problem of (3) is unique.

The result in the smooth case can then be written as follows:

**Proposition 4.1.** (Freund (1985)). *If assumptions (H1) and (H2) are both satisfied for  $A$ , then  $z(A)$  is continuously differentiable and we have that*

$$\frac{\partial z}{\partial a_{ij}} = -u_i^* x_j^* \quad (4)$$

where  $u_i^*$  is the optimal dual variable associated to row  $i$  and  $x_j^*$  is the optimal primal variable associated to column  $j$ .

A complete analysis of the subject in the differentiable case can be found in Gal (1995).

#### 5. Generalized gradient characterization.

Before examining the case where the optimal basis is not unique, we show on an example that  $z(A)$  does not enjoy any convexity property. Consider the following linear problem with a single parametric matrix coefficient:

$$\begin{aligned} z(a) &= \min_x \quad x_1 + x_2 \\ \text{s.t. } &\begin{cases} ax_1 + x_2 = 1 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned}$$

Using the constraint,  $x_2$  can be substituted:

$$\begin{aligned} z(a) &= \min_x \quad 1 + (1-a)x_1 \\ \text{s.t. } &0 \leq x_1 \leq \frac{1}{a} \end{aligned}$$

The optimal objective function can thus be written explicitly as:

$$z(a) = \begin{cases} 1 & \text{if } a < 1 \\ 1/a & \text{if } a \geq 1 \end{cases}$$

It is clear that  $z(a)$  is neither convex nor concave. Because of this lack of convexity, the notion to be used is the Clarke's generalized gradient.

If  $A$  is such that the linear program is infeasible, we define  $z(A) = -\infty$ . Denote by  $\text{dom}(z)$ , the domain where  $z(A)$  is finite. Before stating the characterization of the generalized gradient, we

first recall the following propositions which result from Renegar (1994).

**Proposition 5.1.** *If the set of optimal primal solutions for  $A$  is unbounded, then  $A$  is not an interior point of  $\text{dom}(z)$ .*

**Proposition 5.2.** *If the set of optimal dual solutions for  $A$  is unbounded, then  $A$  is not an interior point of  $\text{dom}(z)$ .*

We will use the following notation  $u \times x$  for the outer product of an  $n$ -column vector  $u$  by the  $n$ -row vector  $x^T$ . The following theorem states a characterization of the generalized gradient.

**Theorem 5.1.** *If  $A$  is an interior point of  $\text{dom}(z)$  and if  $z(A)$  is locally Lipschitz in a neighborhood of  $A$ , then*

$$\partial z(A) = \text{conv}\{-u \times x \text{ where } u \text{ is any optimal dual solution and } x \text{ is any primal optimal solution of (3)}\}$$

**Proof.**

1. Suppose first that there is a single optimal basis. Since  $A$  is an interior point of  $\text{dom}(z)$ , we know by Propositions 5.1 and 5.2 that there are no extreme rays of primal or dual optimal solutions. In this case, a single optimal basis is a sufficient condition to have primal and dual nondegeneracy. We know from Proposition 4.1 that  $\partial z(A)$  reduces to a single matrix, which can be computed by the following formula:

$$\partial z(A) = \{-u \times x\}$$

where  $u$  and  $x$  are the dual and primal solutions associated to the unique optimal basis for (3). This proves the theorem.

2. Suppose next that there are several optimal basis. Let's first introduce some useful notation from linear programming. For a particular optimal basis, we denote by  $B$  the columns of matrix  $A$  corresponding to the basic variables noted  $x_B$  and we denote by  $N$  the columns of matrix  $A$  corresponding to non basic variables, noted  $x_N$ . The constraints can be rewritten as follows:

$$Ax = b \Leftrightarrow (B, N)(x_B, x_N)^T = b \Leftrightarrow Bx_B = b \Leftrightarrow x_B = B^{-1}b$$

Denote  $c_b$  the objective coefficient corresponding to basic variables, it follows that the objective function can be rewritten for a particular basis as follows:

$$z(A) = c_b^T B^{-1}b$$

We first prove the following inclusion:

$$\text{conv}\{-u \times x \text{ such that } u^T = c_b^T B^{-1}, x_B = B^{-1}b, x_N = 0$$

and  $B$  corresponds to an optimal basis of (3)}  $\subset \partial z(A)$

Let  $B$  corresponds to an optimal basis of (3). Since this optimal basis is not unique, there must be at least one non-basic variable  $x_j$  with a zero reduced cost:

$$c_j - c_b^T B^{-1}a_j = 0$$

where  $a_j$  denotes the column  $j$  of matrix  $A$ . Using the definition of the vector of dual variables  $u^T = c_b^T B^{-1}$ , this condition can be rewritten:

$$c_j - u^T a_j = 0$$

We can exclude the pathological case where  $u = 0$ . In fact, this case can be treated by a perturbation of the objective coefficients. This only requires to consider  $A$  as the extended matrix of the system where the objective is added as a constraint.

We can thus take  $u_i$  different from zero and define the following perturbation of the column  $a_j$  for any column with zero reduced cost: if  $u_i > 0$ , subtract  $\epsilon > 0$  ( $\epsilon < 0$  if  $u_i < 0$ ) from the  $i$ th component of  $a_j$ . For the perturbed problem, all the reduced costs are strictly positive, and therefore the optimal solution becomes unique.



More specifically, let a sequence of matrices  $A(\epsilon)$  where the objective function is differentiable converging towards matrix  $A$  where it is not differentiable. The primal and dual optimality conditions hold for each point of the sequence and the optimal primal and dual variables are continuous functions of the perturbed matrix since this latter is a perturbation of an invertible matrix at the limit point. The limits of the primal and dual variables at each point of the sequence thus exist and satisfy the primal dual relations at the limit point.

It can be concluded that all the points associated with all the optimal basis for the problem (3) belong to the generalized gradient, since they can be written as the limit of perturbed problems where the optimal basis is unique and so where the function  $z$  is continuously differentiable in  $A(\epsilon)$  for all  $\epsilon > 0$ .

Now consider the inverse inclusion:

$$\partial z(A) \subset \text{conv}\{-u \times x \text{ such that } u^T = c_B^T B^{-1}, x_B = B^{-1}b, x_N = 0 \text{ and } B \text{ corresponds to an optimal basis of (3)}\}$$

Let  $\{A_k\}$  be any sequence such that  $\nabla z(A_k)$  exists and converges. The optimal basis associated with each  $A_k$  does not have to be unique. As shown by Freund (1985), we can have, for a given matrix  $A$ , several degenerate optimal basis although  $z$  is continuously differentiable for matrix  $A$ . We will show, in this case, that any optimal basis associated with  $A_k$  must give the same point  $(\dots, -u_i x_j, \dots)$ .

Suppose the opposite, that is, there are two different optimal basis for  $A_k$ , giving two different points  $(\dots, -u_i^1 x_j^1, \dots)$  and  $(\dots, -u_i^2 x_j^2, \dots)$  respectively. As done in part i) of the proof, the matrix  $A_k$  can be perturbed in order to have no more than one of the two basis optimal. By taking this limit, we obtain that the first point is in the generalized gradient. By applying the same procedure of perturbation to the second basis, we show that the second point is also in the generalized gradient. We can therefore conclude that  $\partial z(A_k)$  is not a singleton. Applying Proposition 3.1, this contradicts the fact that  $z$  is continuously differentiable in  $A_k$ .

The gradient can therefore be associated with any of the optimal basis. Note by  $\{\beta_k\}$  a sequence of optimal basis for  $A_k$  (i.e.  $\beta_k$  is an optimal basis for  $A_k$ ). By a basis  $\beta$ , we mean here a partition of the variables between basic and non-basic variables. As  $\{\beta_k\}$  is an infinite sequence of basis and as there is only a finite choice of  $m$  columns among the  $n$  columns of the  $A$  matrix, so there must be a special basis  $\beta$  which is repeated infinitely often in the sequence. Let  $\{B_l\}$  be the subsequence corresponding to this basis which is repeated infinitely often. The corresponding subsequence  $\{(\dots, -x_j u_i, \dots)_l\}$  of gradients associated with this basis converges to the same point as the original sequence. As

$$c_N^T - c_B^T(B_l)^{-1} \geq 0 \\ (B_l)^{-1}b \geq 0$$

for all  $l$ , these inequalities remain true for  $l \rightarrow \infty$  and so  $\{B_l\}$  converges to an optimal basis for (3). This completes the proof of the reverse inclusion.

3. We finally show that

$$\partial z(A) = \text{conv}\{-u \times x \text{ where } u \text{ is any optimal dual solution and } x \text{ is any primal optimal solution of (3)}\}$$

Because the sets of primal and dual optimal solutions corresponding to point  $A$  are bounded by Propositions 5.1 and 5.2,  $u$  and  $x$  are convex combinations of extreme dual and primal solutions respectively. Let

$$u = \sum_k \mu_k u^k \text{ where } \sum_k \mu_k = 1 \text{ and } \mu_k \geq 0 \\ x = \sum_l \lambda_l x^l \text{ where } \sum_l \lambda_l = 1 \text{ and } \lambda_l \geq 0$$

Suppose first that  $u$  is a convex combination of extreme  $u^k$  while  $x$  is an extreme optimal point. One has

$$u_i x_j = \sum_k \lambda_k u_i^k x_j$$

for a given set of  $\lambda_k$  and for all  $i$  and  $j$ . Therefore

$$-u \times x = - \sum_k \lambda_k u^k \times x^k$$

This implies that

$$\text{conv}\{-u \times x \text{ where } u \text{ is any optimal dual solution and } x_B = B^{-1}b, x_N = 0, \text{ where } B \text{ is the optimal basis}\} \\ = \text{conv}\{-u \times x \text{ where } u = c_B^T B^{-1}, x_B = B^{-1}b, x_N = 0 \text{ and } B \text{ is any optimal basis of (3)}\}$$

The same reasoning can be made in order to relax the requirement that  $x$  is an extreme solution into the weaker one that  $x$  is any optimal solution of problem (3).  $\square$

Before illustrating the theorem on an example, let us say a few words about the requirements for  $A$  to be an interior point of  $\text{dom}(z)$  and for  $z(A)$  to be Lipschitz in a neighborhood of  $A$ . Im (2018) proves that these two requirements holds true if the following conditions are satisfied:

**Assumption 5.1.** The matrix  $A$  is of full rank and the Slater constraints qualification is satisfied.

**Proposition 5.3.** If matrix  $A$  is a full rank matrix and if the Slater constraints qualification is satisfied, then

- $A$  is an interior point of  $\text{dom}(z)$  and
- the function  $z(A)$  is Lipschitz in a neighborhood of  $A$ .

**Proof.** See Im (2018), pages 74–76.

As indicated by Höffner et al. (2016), any linear program can be reduced to an equivalent linear program that satisfies the full rank property for  $A$  by removing linearly dependent rows.

The following simple example illustrates Theorem 5.1:

$$z(a) = \max_x x_1 + x_2 \\ \text{s.t.} \begin{cases} x_1 + 2x_2 \leq 3 \\ x_1 + ax_2 \leq 2 \\ x_1, x_2 \geq 0 \end{cases}$$

The feasible region and the objective function are represented in Fig. 1 for the particular choice of  $a = 1$ .

For  $a = 1$ , there exists two different basic solutions. The first one is obtained with  $x_1$  and  $x_2$  in the basis:  $(x_1, x_2) = (1, 1)$  and the reduced cost of  $s_1$ , the first slack variable, is zero. The second solution is obtained by taking  $x_1$  and  $s_1$  in the basis:  $(x_1, x_2) = (2, 0)$  and the reduced cost of  $x_2$  is zero. In both cases, the optimal dual values are given by  $(u_1, u_2) = (0, 1)$ .

Take  $a = 1 - \epsilon$  and let  $\epsilon$  go to zero. We obtain the first solution and the reduced cost associated to  $s_1$  is strictly negative. Take  $a = 1 + \epsilon$  and let  $\epsilon$  go to zero. We obtain the second solution and the reduced cost associated to  $x_2$  is strictly negative. The extreme points of the generalized gradient are thus:

$$-u_2 x_2 = -1 \text{ (first case)}$$

$$-u_2 x_2 = 0 \text{ (second case).}$$

One therefore obtains:

$$\partial z(1) = [-1, 0]$$

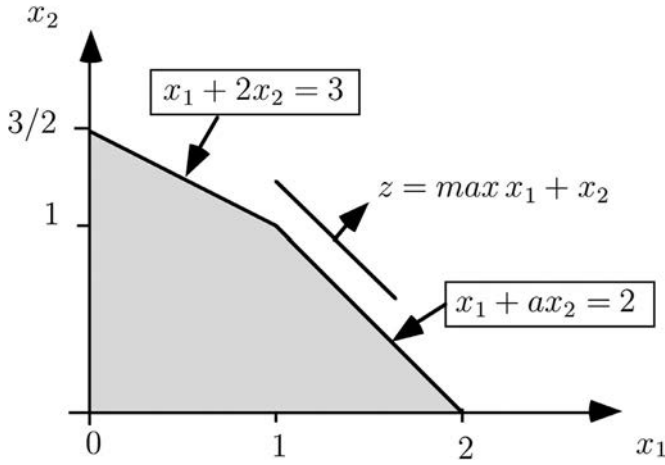


Fig. 1. Illustrative example.

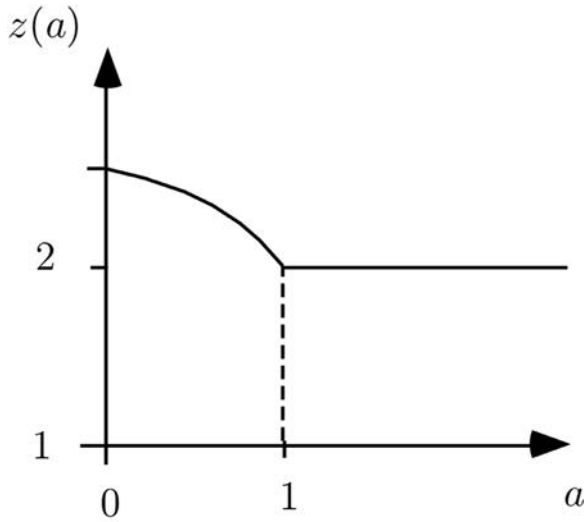


Fig. 2. Graph of the optimal value of the objective function.

In fact, the general expression of  $z(a)$  can be computed explicitly as:

$$z(a) = \begin{cases} \frac{3a-5}{a-2} & \text{if } a < 1 \\ 2 & \text{if } a \geq 1 \end{cases}$$

The graph of the optimal value of the function  $z(a)$  is represented in Fig. 2 as a function of parameter  $a$ . The two points  $-1$  and  $0$  correspond thus to the left- and right-derivatives of  $z(a)$  at point  $a = 1$  respectively.  $\square$

## 6. Pratical application

The motivation for considering the computation of the generalized gradient of the objective function of a linear problem with respect to the matrix coefficients is the general problem where there is a two-stage problem where at the first stage a capacity investment decision is made and at the second stage the operating of the system is optimized taking into account this investment decision. In some cases, the investment decision appears in the right-hand side (such as capacity level decision). We consider the case where the investment decision appears as *coefficient of the second-stage decision variables*. Let us illustrate this fact with an example from the gas industry.

Consider the problem of the optimal dimensioning of pipe networks for the transmission of natural gas. See, for example, De Wolf and Smeers (1996) who consider the following two-stage problem for investment and exploitation of gas transmission networks. At the first stage, the *optimal diameters of pipe lines*, denoted  $D$ , must be determined in order to minimize the sum of the *direct investment cost function*, denoted  $C(D)$ , and  $Q(D)$ , the *future operating cost function*.

$$\min_D \quad F(D) = C(D) + Q(D) \quad (5)$$

$$\text{s.t.} \quad D_{ij} \geq 0, \quad \forall (i, j) \in SA$$

where  $SA$  denotes the set of arcs of the network.

The *operations problem* for a given choice of the diameters can thus be formulated as follows:

$$Q(D) = \min_{f, s, p} \sum_{j \in N_s} c_j s_j$$

$$\text{s.t.} \quad \begin{cases} \sum_{j|(i,j) \in A} f_{ij} - \sum_{j|(j,i) \in A} f_{ji} = s_i & \forall i \in N \\ \text{sign}(f_{ij}) f_{ij}^2 = K_{ij}^2 D_{ij}^5 (p_i^2 - p_j^2) & \forall (i, j) \in SA \\ \underline{s}_i \leq s_i \leq \bar{s}_i & \forall i \in N \\ \underline{p}_i \leq p_i \leq \bar{p}_i & \forall i \in N \end{cases} \quad (6)$$

where the variables of the problem are  $f_{ij}$ , the flow in the arc  $(i, j)$ ,  $s_i$ , the net supply at the node  $i$  and  $p_i$ , the pressures at node  $i$ . The set of nodes is denoted  $N$ . For simplicity of notation, we define the variable  $\pi_i$  as the square of the pressure at node  $i$ :

$$\pi_i = p_i^2.$$

Let us replace in the only nonlinear relation of exploitation problem (6), the  $D_{ij}$  variable by the following substitute:

$$x_{ij} = D_{ij}^5$$

In fact, taking the  $x_{ij}$  as parameters, we find that they appear as linear coefficients of the squared pressure variables in the equation:

$$\text{sign}(f_{ij}) f_{ij}^2 - K_{ij}^2 D_{ij}^5 (\pi_i - \pi_j) = 0. \quad (7)$$

De Wolf and Smeers (2000) solve the problem of the gas transmission problem by an extension of the Simplex algorithm using piecewise linearisation of the first term  $\text{sign}(f_{ij}) f_{ij}^2$ . We are thus back in the linear case. Let  $w_{ij}^*$  be the optimal value of the dual variable associated with constraint (7). Applying Theorem 5.1, one obtains an element of the generalized gradient by:

$$\frac{\partial Q}{\partial x_{ij}} = w_{ij}^* K_{ij}^2 (\pi_i^* - \pi_j^*) \quad (8)$$

Now, to obtain an element of the generalized gradient with respect to the original variables  $(D_{ij})$ , one uses the chain rule for the composition of derivative with:

$$\frac{\partial x_{ij}}{\partial D_{ij}} = 5D_{ij}^4.$$

It is then easy to prove that the following expression gives an element of the generalized gradient of  $Q(D)$ :

$$\frac{\partial Q(D)}{\partial D_{ij}} = w_{ij}^* (\pi_i^* - \pi_j^*) 5K_{ij}^2 D_{ij}^4 \quad (9)$$

This formula gives thus an element of the generalized gradient, which is the only information required by the bundle method (See Lemaréchal (1989)) used to solve the two stage problem. See De Wolf and Smeers (1996) for the application of the bundle method to this two-stage problem.

## 7. Conclusions

It has been shown in this paper how the first-order derivatives of the optimal solution of a linear program with respect to matrix coefficients can be generalized to the nonsmooth case, even when the optimal function as a function of matrix coefficients admits breakpoints. Our result, [Theorem 5.1](#), emphasizes the fundamental role played by bases in this respect. The extreme points of the generalized gradient correspond to all the different optimal basis. A practical application to gas transmission network optimization, which was in fact the motivation for considering such a formula, was then presented.

## Acknowledgments

We would like to thank Professor Guy de Ghellinck for valuable discussions and suggestions. We would also like to thank Professors Jean-Jacques Strodiot and Claude Lemaréchal for insightful discussions on the state of the art in nonsmooth optimization. Finally, we would like to thank the two anonymous reviewers for valuable suggestions to improve the paper.

## References

- Barton, P. I., Khan, K. A., Stechlin, P., & Watson, H. A. J. (2018). Computationally relevant generalized derivatives: theory, evaluation and applications. *Optimization Methods & Software*, 33(4–6), 1030–1072.
- Bonnans, J. F., & Shapiro, A. (2000). Perturbation analysis of optimization problems. In *Springer series in operations research*. New York: Springer-Verlag.
- Clarke, F. H. (1990). Optimization and nonsmooth analysis. *Classics in Applied Mathematics* 5, Society for Industrial and Applied Mathematics.
- Craven, B. D., & Janin, R. (1993). Regularity properties of the optimal value function in non linear programming. *Optimization*, 28, 1–7.
- De Wolf, D., & Smeers, Y. (1996). Optimal dimensioning of pipe networks with application to gas transmission networks. *Operations Research*, 44(4), 596–608.
- De Wolf, D., & Smeers, Y. (2000). The gas transmission problem solved by an extension of the simplex algorithm. *Management Science*, 46(11), 1454–1465.
- Freund, R. M. (1985). Postoptimal analysis of a linear program under simultaneous changes in matrix coefficients. In R. W. Cottle (Ed.), *Mathematical programming essays in honor of George B. Dantzig part I, mathematical programming studies: vol. 24* (pp. 1–13).
- Gal, T. (1995). Postoptimal analyses. Berlin, New York. Parametric Programming and Related Topics, Walter de Gruyter.
- Gauvin, J. (1979). The generalized gradient of a marginal function in mathematical programming. *Mathematics of Operations Research*, 4, 458–463.
- Gomez, J. A., Höffner, K., Khan, K. A., & Barton, P. I. (2018). Generalized derivatives of lexicographic linear programs. *Journal of Optimization Theory and Applications*, 178, 477–501.
- Höffner, K., Khan, K. A., & Barton, P. I. (2016). Generalized derivatives of dynamic systems with a linear program embedded. *Automatica*, 63, 198–208.
- Im, J. (2018). Sensitivity analysis and robust optimization: A geometric approach for the special case of linear optimization. University of Waterloo, Ontario, Canada. Master's thesis, 1–97.
- Lemaréchal, C. (1989). Nondifferentiable optimization. In G. L. Nemhauser (Ed.), *Handbooks in operations research and management science: vol. 1* (pp. 529–572). Optimization, Elsevier Science Publishers.
- Mordukhovich, B. S., Nam, N. M., & Yen, N. D. (2007). Subgradients of marginal functions in parametric mathematical programming. *Mathematical Programming, Series B*, 116, 369–396.
- Penot, J. P. (2004). Differentiability properties of optimal value functions. *Canadian Journal of Mathematics*, 56(4), 825–842.
- Renegar, J. (1994). Some perturbation theory for linear programming. *Mathematical Programming*, 65, 73–91.
- Rockafellar, R. T. (1984). Directional differentiability of the optimal value function in a nonlinear programming problem. In A. V. Fiacco (Ed.), *Sensitivity, stability and parametric analysis, mathematical programming studies: vol. 21* (pp. 213–226). North-Holland.
- Thibault, L. (1991). On subdifferentials of optimal value functions. *SIAM Journal of Control and Optimization*, 29(5), 1019–1036.