A 3D model for Optimizing Infrastructure Costs in Road Design

Miguel E. Vázquez-Méndez*, Gerardo Casal, Duarte Santamarina

Departamento de Matemática Aplicada, Universidade de Santiago de Compostela, Escola Politécnica Superior, Benigno Ledo, 27002 Lugo, Spain

&

Alberte Castro

Departamento de Enxeñería Agroforestal, Universidade de Santiago de Compostela, Escola Politécnica Superior, Benigno Ledo, 27002 Lugo, Spain

Abstract: *In this paper the optimal design of a road joining two terminals is investigated. A geometric model is proposed including horizontal transition curves and vertical curves, obtaining parametrizations for the central axis of the road as well as for its entire surface. These parametrizations allow to express and compute, with great simplicity, the major infrastructure costs, including land acquisition, clearance, pavement, maintenance and earthwork, where multiple layers of materials with different costs can be handled. The road design problem is formulated as a smooth constrained optimization problem and a two-stage algorithm is suggested for its numerical resolution. Finally, numerical results are presented in an academic test and in a case study that propose designing a by-pass in a Spanish national road (N-640) to avoid crossing Monterroso's town center.*

1 INTRODUCTION

Road design is a complex and deeply studied problem in Civil Engineering. The basic idea, however, is very simple: to join two points by means of a three-dimensional surface (road) fulfilling a set of predefined requirements (constraints) aiming at previously marked objectives. The natural formulation of the problem is, therefore, framed into a constrained optimization problem, and in this scope it has been studied in the last decades by different authors. Thus, the fundamental elements are the design variables (determining unequivocally the layout), the mathematical model (providing the alignment from those variables), the constraints (conditions that any admissible route must fulfill), and the objective functionals (measuring the achievement of those objectives).

The design variables depend on the specific problem and on the mathematical (geometric) model. In this sense, works published so far can be classified into three main groups: those focusing on the horizontal alignment (Casal et al. (2017), Easa and Mehmood (2008), Lee et al. (2009), Mondal et al. (2015)), those centering their attention into the vertical alignment (see Fwa et al. (2002), Hare et al. (2014), Hare et al. (2015), Beiranvand et al. (2017)), and those studying both alignments simultaneously (Bosurgi et al. (2013), Chew et al. (1989), Hirpa et al. (2016), Jong and Schonfeld (2003), Li et al. (2013), Shafahi and Bagherian (2013)). For each of them, the chosen variables depend on the geometric model used to define the layout: within the horizontal alignment it is possible to consider, only main axes (Lee et al. (2009)), axes and circular curves (Hirpa et al. (2016), Mondal et al. (2015), Jong and Schonfeld (2003), Shafahi and Bagherian (2013)), or axes, circular curves and transition curves (Casal et al. (2017), Kang et al. (2012)); within the vertical alignment, vertical curves may be considered (Jong and Schonfeld (2003), Shafahi and Bagherian (2013)), or may not (Hirpa et al. (2016)). In addition, both the horizontal and vertical transition curves can be modeled differently (Bosurgi and D'Andrea (2012) and Kobryń (2017)), which also leads to distinct design variables.

For the constraints, there are the mandatory ones, regarding geometrical criteria depending on the road type (design speed, one or two lines, etc.) which are given by each country's legislation (AASHTO (2011)), but there are also many others, such as those concerning the topography (Li et al.

(2016)), environmental aspects (Bosurgi et al. (2013) and Davey et al. (2017)), or measures to improve the road safety (Easa and Mehmood (2008)).

Objective functions may also be different, and in the literature (see, for instance, Jong and Schonfeld (1999) or Jha et al. (2006)) the major layout costs are classified into constructions costs, maintenance costs, user costs and social and environmental costs. From the formulation point of view, the main difference lies in whether all the considered costs can be gathered into a single functional to minimize, leading to a single-objective problem (see, for example, Chew et al. (1989), Jong and Schonfeld (2003) or Kang et al. (2012)), or if the goal is to minimize different functionals simultaneously, leading to a multi-objective problem (see Hirpa et al. (2016)).

Finally, once the problem has been correctly stated, it is necessary to employ a numerical method to solve it. The search of suitable algorithms to solve these problems is another current topic. Kang et al. (2012), and more recently Li et al. (2016), give a classification of the papers on this topic, based on the optimization approach used. Nowadays, the most common are genetic algorithms (Kang et al. (2012), Davey et al. (2017), Li et al. (2017)), derivative-free optimization algorithms (Mondal et al. (2015), Hirpa et al. (2016)), shortest path (Pushak et al. (2016)) and distance transform methods (Li et al. (2016)). In any case, the selection of a suitable numerical method must be done according to the characteristics of the stated problem.

In conclusion, there is no uniformity in literature when it comes to choosing the design elements, and authors adapt the geometry of the model to their own needs and interests. With respect to this point, in this paper we use a 3D geometric model obtained by extension to three dimensions of the 2D model proposed by the authors for a different problem, where the main objective was to develop a mathematical tool for designing horizontal alignments in road reconstruction projects (Casal et al. (2017)). The design variables used and the 3D geometric model are detailed in Section 2. It considers all road alignment elements, including horizontal (clothoids) and vertical (parabolic) transition curves, and provides an algorithm to obtain, from the design variables, a parametrization for the road's central axis and for the whole road surface. In Section 2, typical constraints are also derived from compliance of current legislation (taking as example the Spanish standard (SMPW (2016))) and they are formulated from design variables by using previous parametrizations. The computational novelty of using these parametrizations are showed in Section 3, where major infrastructure costs are studied and formulated in a different mathematical framework, which is so flexible and simple, allowing a quick mathematical analysis of the problem. This new formulation is especially useful for land acquisition costs (allowing to include forbidden passage areas) and earthwork costs, which can be easily computed, and where sections with variable widths depending on the terrain topography are included. Transition cross-sections (*hillside* sections) are also taken into account and additionally the possibility to deal with different materials, with different excavation costs, in earthwork is also a notable advance. All this leads (Section 4) to a novel computational model for designing new alignments. It consists of a single-objective, differentiable, non-linear and non-convex optimization problem that is solved with a two-stage method, combining global optimization techniques with a gradient type algorithm. This method is used (Section 5) to solve an academic problem and a case study consisting in designing a by-pass layout in a Spanish national road (N-640) avoiding Monterroso's town center. The paper finishes (Section 6) with some conclusions and future research proposals.

2 MATHEMATICAL MODEL FOR A 3D ROAD ALIGNMENT. DESIGN VARIABLES AND CONSTRAINTS

The main goal of this work is the design of a road joining two given points $a = (x_a, y_a, z_a) \in \mathbb{R}^3$ and $b = (x_b, y_b, z_b) \in \mathbb{R}^3$. Geometrically, a road is a 3D surface fulfilling a set of requirements. We devote this section to set the variables determining that surface (design variables) and to give an algorithm providing, for each set of values of those variables, a parametrization of the corresponding surface. This parametrization will be the key point in writing and computing, in Section 3, the cost function to optimize.

A road path is completely defined giving the horizontal alignment, the vertical alignment and the cross-sections. In the following, a deep analysis of each of them is performed.

2.1 HORIZONTAL ALIGNMENT

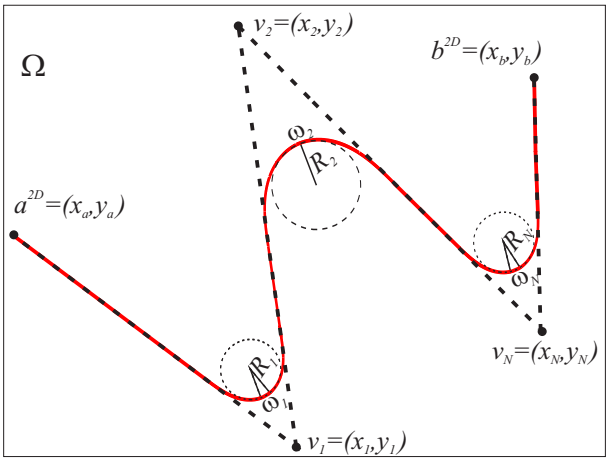

Figure 1: Horizontal alignment ($C^{2D}_{\mathbf{x}^N}$) connecting two terminals a^{2D} and b^{2D}.

Let us consider the points $a^{2D} = (x_a, y_a) \in \mathbb{R}^2$ and $b^{2D} = (x_b, y_b) \in \mathbb{R}^2$. The horizontal alignment is a 2D curve formed by a suitable combination of straight segments, circular curves and transition curves, joining a^{2D} and b^{2D}. Taking clothoids as transition curves, if the path consists of $N+1$ tangents, the horizontal alignment is unequivocally determined by the vertices ($v_i = (x_i, y_i)$, $i = 1, \ldots, N$) where these tangents intersect, the radii ($R_i \geq 0$, $i = 1, \ldots, N$) and the angles ($\omega_i \geq 0$, $i = 1, \ldots, N$) of circular curves, (see Figure 1). Thus, for each $N \in \mathbb{N}$ we denote

$$\mathbf{x}^N = (x_1, y_1, R_1, \omega_1, \ldots, x_N, y_N, R_N, \omega_N) \in \mathbb{R}^{4N},$$

the vector of the horizontal design variables and by $C^{2D}_{\mathbf{x}^N} \subset \mathbb{R}^2$ the 2D curve (horizontal alignment) determined by $\mathbf{x}^N$. This curve can be expressed in terms of the arc length parameter. If $L(\mathbf{x}^N)$ denotes the length of the horizontal alignment corresponding to $\mathbf{x}^N$, this parametrization is given by function

$$\sigma^{2D}_{\mathbf{x}^N} : s \in [0, L(\mathbf{x}^N)] \longmapsto \sigma^{2D}_{\mathbf{x}^N}(s) = (\sigma_1(s), \sigma_2(s))$$

detailed in Casal et al. (2017). For self-containedness, the algorithm to compute $\sigma^{2D}_{\mathbf{x}^N}$ is described in Appendix A. Other parametrization of the horizontal alignment in terms of different design variables can be seen, for example, in Lovell (1999).

The horizontal design variables should verify some constraints to guarantee that the corresponding curve $C^{2D}_{\mathbf{x}^N}$ is an appropriate horizontal alignment (Casal et al. (2017)). Moreover, each country usually has legal constraints over the elements of the layout: bounds over the length of the clothoids and/or straight segments, proper relationship between radii of two consecutive curves, etc. (see, for example, AASHTO (2011)). As an example, in this paper we consider the constraints that are collected in the following admissible set:

$$X^N_{ad} = \left\{ \begin{array}{l} \mathbf{x}^N \in \mathbb{R}^{4N} \text{ such that } R_{min} \leq R_i, 0 \leq \omega_i, \\ \omega_i \leq \theta_i(\mathbf{x}^N), L^C_{min} \leq L^C_i(\mathbf{x}^N) \leq L^C_{max}, \\ L^T_{min} \leq L^T_j(\mathbf{x}^N) \leq L^T_{max} \end{array} \right\},$$

where $\theta_i(\mathbf{x}^N)$ is the difference of azimuth between tangents i and $i+1$, $L^C_i(\mathbf{x}^N)$ is the length of each clothoid in turn i, and $L^T_j(\mathbf{x}^N)$ is the length of straight segment j (see Appendix A).

2.2 VERTICAL ALIGNMENT

Lets suppose that we have a vector $\mathbf{x}^N \in \mathbb{R}^{4N}$ giving the horizontal alignment between points $a^{2D} \in \mathbb{R}^2$ and $b^{2D} \in \mathbb{R}^2$. The aim is to define a z-coordinate to get a 3D curve properly linking points $a \in \mathbb{R}^3$ and $b \in \mathbb{R}^3$. Consequently, we seek for a function

$$\sigma_3 : s \in [0, L(\mathbf{x}^N)] \in \mathbb{R} \longmapsto \sigma_3(s) \in \mathbb{R},$$

verifying $\sigma_3(0) = z_a$, $\sigma_3(L(\mathbf{x}^N)) = z_b$ whose graphic is a concatenation of straight segments joined by parabolic curves (see Figure 2(a)) such that in each polygonal vertex (S_j, Z_j), the input and output tangents (S^{in}_j and S^{out}_j, respectively), hold

$$S_j - S^{in}_j = S^{out}_j - S_j.$$

Thus, once the horizontal distance between the input and output tangents $L^S_j > 0$ has been set, it must satisfy that

$$S^{in}_j = S_j - \frac{L^S_j}{2}, \quad S^{out}_j = S_j + \frac{L^S_j}{2}. \tag{1}$$

The computation of vertical alignment involves to obtain the parabolic arc connecting two consecutive straight segments, each is determined by the tangents (given by vertex (S_j, Z_j) and slopes m_j and m_{j+1}) and the two tangency points (given by S^{in}_j and S^{out}_j). These parabolic curves obey

$$f_j(s) = a_j s^2 + b_j s + c_j, \quad s \in [S^{in}_j, S^{out}_j],$$

where the coefficients a_j, b_j, and c_j are determined, from problem data, as follows:

1. In tangency points, the slope must match with the corresponding vertical segment, therefore,

$$f'_j(S^{in}_j) = m_j, \qquad f'_j(S^{out}_j) = m_{j+1},$$

leading to

$$2a_j(S_j - \frac{L^S_j}{2}) + b_j = m_j,$$
$$2a_j(S_j + \frac{L^S_j}{2}) + b_j = m_{j+1},$$

and, consequently,

$$a_j = \frac{m_{j+1} - m_j}{2L^S_j}, \tag{2}$$

$$b_j = \frac{m_{j+1} + m_j}{2} - \frac{S_j(m_{j+1} - m_j)}{L^S_j}. \tag{3}$$

2. Parabolic curves must go through tangency points, hence,

$$f_j(S^{in}_j) = Z_j - \frac{L^S_j}{2} m_j,$$
$$f_j(S^{out}_j) = Z_j + \frac{L^S_j}{2} m_{j+1},$$

and, taking into account (2) and (3),

$$c_j = Z_j - a_j\left(S_j^2 + \frac{(L^S_j)^2}{4}\right) - b_j S_j + \frac{L^S_j(m_{j+1} - m_j)}{4}. \tag{4}$$

According to what we have just seen, once the number $M \in \mathbb{N}$ of slope changes has been set, the vertical alignment is unambiguously determined by the following vector, called vector of the vertical design variables,

$$\mathbf{y}^M = (s_1, m_1, K_{v1}, \ldots, s_M, m_M, K_{vM}) \in \mathbb{R}^{3M},$$

where (see Figure 2):

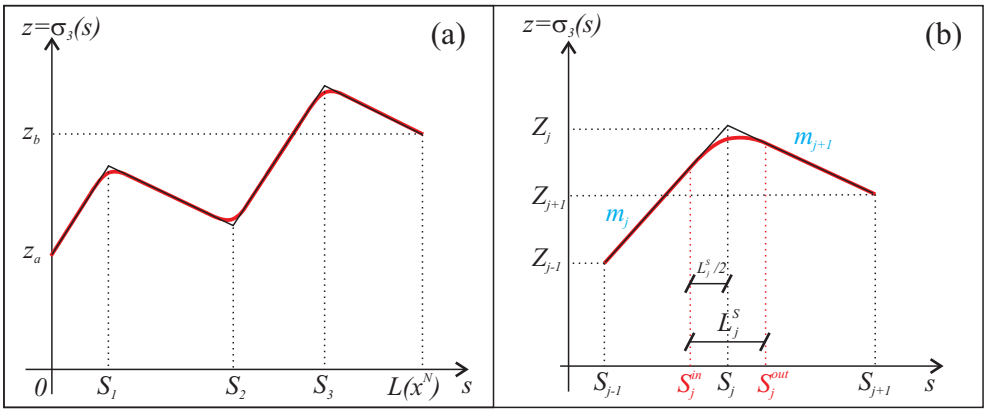

Figure 2: (a) Vertical alignment with $M = 3$ vertices. (b) Parameters defining the vertical alignment around vertex j.

- $0 < s_1 < \ldots < s_M < 1$ are the values that, once multiplied by the length of the horizontal alignment, provide the coordinates in s of the vertical alignment vertices, i.e., $s_j = S_j/L(\mathbf{x}^N)$.

- $m_j \in \mathbb{R}$ is the slope of the j-th straight segment in the vertical alignment. It should be noted that, although the vertical alignment consists of $M+1$ straight segments, only the first M are controlled, since the latter is fully determined by the height in both ends (z_a and z_b).

- $K_{vj} = \dfrac{L_j^S}{|m_{j+1} - m_j|} > 0$ gives the ratio between the length of the parabolic section and the difference between two consecutive uniform slopes.

If we take for the 3D layout the decision variable given by vector $\mathbf{u}^{N,M} = (\mathbf{x}^N, \mathbf{y}^M) \in \mathbb{R}^{4N+3M}$, then the vertical alignment is given by function

$$\sigma_{3\mathbf{u}^{N,M}} : s \in [0, L(\mathbf{x}^N)] \longmapsto \sigma_{3\mathbf{u}^{N,M}}(s) = \sigma_3(s)$$

whose definition is detailed in the following algorithm, where computations of its first derivative are also included:

Algorithm 1: Computation of $\sigma_{3\mathbf{u}^{N,M}}(s)$ and $\sigma'_{3\mathbf{u}^{N,M}}(s)$

- *Initial data:* `Compute` $S_0 = 0$, $S_0^{out} = 0$, $Z_0 = Z_a$.

- `For j=1,...,M`

  - *Parameters definition:*

    * `Compute` $S_j = s_j L(\mathbf{x}^N)$, $Z_j = Z_{j-1} + m_j(S_j - S_{j-1})$.

    * `If` $j = M$ `compute`

$$m_{M+1} = \frac{Z_b - Z_M}{L(\mathbf{x}^N) - S_M}.$$

    * `Compute` $L_j^S = K_{vj}|m_{j+1} - m_j|$,

$$S_j^{in} = S_j - \frac{L_j^S}{2}, \quad S_j^{out} = S_j + \frac{L_j^S}{2}.$$

  - *Straight segment:* `If` $s \in [S_{j-1}^{out}, S_j^{in})$, `compute`

$$\sigma_3(s) = Z_{j-1} + m_j(s - S_{j-1}),$$
$$\sigma_3'(s) = m_j.$$

  - *Parabola:* `If` $s \in [S_j^{in}, S_j^{out})$, `compute`

$$\sigma_3(s) = a_j s^2 + b_j s + c_j,$$
$$\sigma_3'(s) = 2a_j s + b_j.$$

    where a_j, b_j and c_j are given by (2), (3) and (4).

- *Last straight segment:* `If` $s \in [S_M^{out}, L(\mathbf{x}^N)]$, `compute`

$$\sigma_3(s) = Z_M + m_{M+1}(s - S_M),$$
$$\sigma_3'(s) = m_{M+1}.$$

In order to have a truly vertical alignment $\sigma_{3\mathbf{u}^{N,M}}(s)$ given by the previous algorithm it is mandatory to fulfill, for $j = 2, \ldots, M$,

$$S_1^{in} \geq 0, \quad S_M^{out} \leq L(\mathbf{x}^N), \quad S_j^{in} \geq S_{j-1}^{out}. \tag{5}$$

These constrains can be easily expressed in terms of the problem decision variables, requiring that the values giving the horizontal distances between vertical curves, functions $L_j^{3D}(\mathbf{u}^{N,M})$, must be positive. Those are given by

- $L_1^{3D}(\mathbf{u}^{N,M}) = s_1 L(\mathbf{x}^N) - \dfrac{K_{v1}|m_2 - m_1|}{2}$,

- $L_j^{3D}(\mathbf{u}^{N,M}) = (s_j - s_{j-1})L(\mathbf{x}^N) - \dfrac{K_{vj}|m_{j+1} - m_j| + K_{vj-1}|m_j - m_{j-1}|}{2}$, for $j = 2, \ldots, M$,

- $L^{3D}_{M+1}(\mathbf{u}^{N,M}) = (1-s_M)L(\mathbf{x}^N) - \dfrac{K_{vM}|m_{M+1}-m_M|}{2}$,

and, obviously, (5) is equivalent to

$$L^{3D}_j(\mathbf{u}^{N,M}) \geq 0 \quad \text{for } j = 1,\ldots,M+1. \tag{6}$$

In addition, slopes must be upper and lower bounded and, in order to ensure appropriate sight distances, the values of K_{vj} must be greater than a lower threshold. This leads to define the following admissible set of vertical design variables

$$Y^M_{ad} = \left\{ \begin{array}{l} \mathbf{y}^M \in \mathbb{R}^{3M} \text{ such that } K_{vmin} \leq K_{vj} \\ 0 < s_1 < \ldots < s_M < 1, \\ m_{min} \leq |m_j| \leq m_{max}. \end{array} \right\}.$$

Joining these restrictions with the horizontal design variables constraints and with the requirement of a minimum value for the horizontal distance between two vertical curves, the following admissible set of design variables is obtained:

$$U^{N,M}_{ad} = \left\{ \begin{array}{l} \mathbf{u}^{N,M} = (\mathbf{x}^N, \mathbf{y}^M) \text{ such that} \\ \mathbf{x}^N \in X^N_{ad}, \quad \mathbf{y}^M \in Y^M_{ad}, \\ L^{3D}_{min} \leq L^{3D}_j(\mathbf{u}^{N,M}). \end{array} \right\}.$$

2.3 ROAD CROSS-SECTION

Fixed $N, M \in \mathbb{N}$, in previous sections we have seen how to compute, for each $\mathbf{u}^{N,M} \in U^{N,M}_{ad}$, function

$$\sigma^{3D}_{\mathbf{u}^{N,M}} : s \in [0, L(\mathbf{x}^N)] \in \mathbb{R} \longmapsto \sigma^{3D}_{\mathbf{u}^{N,M}}(s) \in \mathbb{R}^3,$$

defined by $\sigma^{3D}_{\mathbf{u}^{N,M}}(s) = (\sigma_1(s), \sigma_2(s), \sigma_3(s))$. This function gives a parametrization of the road central axis (curve $C^{3D}_{\mathbf{u}^{N,M}} \in \mathbb{R}^3$). Our ultimate goal is to obtain a complete parametrization of the road surface ($S_{\mathbf{u}^{N,M}} \in \mathbb{R}^3$) and, to do so, we will see how to parametrize all of its cross-sections. To lighten notation, even if there is dependence on $\mathbf{u}^{N,M}$, from now on, we will omit that subscript.

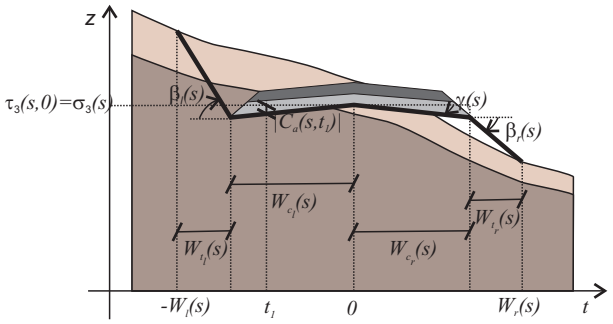

Figure 3: Transition (cut and fill) cross-section of a road profile.

For each $s \in [0, L]$ (with $L = L(\mathbf{x}^N)$), the cross-section at point $\sigma^{3D}(s)$ is defined as the intersection between the road and a vertical plane, orthogonal to the horizontal alignment at point $\sigma^{2D}(s)$. A road surface parametrization is a function

$$\tau : (s,t) \in [0,L] \times [-W_l(s), W_r(s)] \longmapsto \tau(s,t) \in \mathbb{R}^3,$$

with $\tau(s,t) = (\tau_1(s,t), \tau_2(s,t), \tau_3(s,t))$. The definition of τ_3 is obtained from Figure 3. In fact, for every $s \in [0,L]$, we have:

- If $t \in [-W_l, -W_{c_l}]$,

$$\tau_3(s,t) = \sigma_3(s) - C_a(s, -W_{c_l}) - \tan(\beta_l)(t + W_{c_l}).$$

- If $t \in (-W_{c_l}, W_{c_r})$,

$$\tau_3(s,t) = \sigma_3(s) - C_a(s,t),$$

- If $t \in [W_{c_r}, W_r]$,

$$\tau_3(s,t) = \sigma_3(s) - C_a(s, -W_{r_l}) + \tan(\beta_r)(t - W_{c_r}).$$

where:

- $\beta_l(s), \beta_r(s) \in (-\pi/2, \pi/2)$ are, respectively, the left and right side slope angles (positive angles for cut and negative for fill).
- $W_{c_l}(s), W_{c_r}(s) > 0$ width of left and right road pavement, respectively.
- $W_{t_l}(s), W_{t_r}(s) > 0$ width of left and right side slope, respectively.
- $W_l(s) = W_{c_l}(s) + W_{t_l}(s)$, $W_r(s) = W_{c_r}(s) + W_{t_r}(s)$.
- $C_a(s,.) : t \in [-W_{c_l}(s), W_{c_r}(s)] \longmapsto C_a(s,t) \in \mathbb{R}$ function giving the distance between the height of the road at point t and the height of the road central axis ($t = 0$). Explicitily, $C_a(s,t) = |t|\tan(\gamma(s))$, where $\gamma(s) \in (-\pi/2, \pi/2)$ represents the superelevation (or crossfall) of the cross-section, which, in turn, depends on the radius of curvature of the central axis.

On the other hand, $\mathbf{v_t}(s) = (\sigma_1'(s), \sigma_2'(s))$ is the tangent unit vector to the horizontal alignment at point $\sigma^{2D}(s)$. Therefore, $\mathbf{v_o}(s) = (-\sigma_2'(s), \sigma_1'(s))$ is the orthogonal unit vector at that point, and hence, (see Figure 4),

$$\tau_1(s,t) = \sigma_1(s) + t\sigma_2'(s), \tag{7}$$
$$\tau_2(s,t) = \sigma_2(s) - t\sigma_1'(s). \tag{8}$$

3 COST FUNCTIONS

Many cost components affect the construction of new roads. In literature (see, for instance, Chew et al. (1989), Jha et al. (2006), Kang et al. (2012)) major costs are classified in: constructions costs (earthwork, pavement, right-of-way, structures), maintenance costs (pavement moving, lighting), user

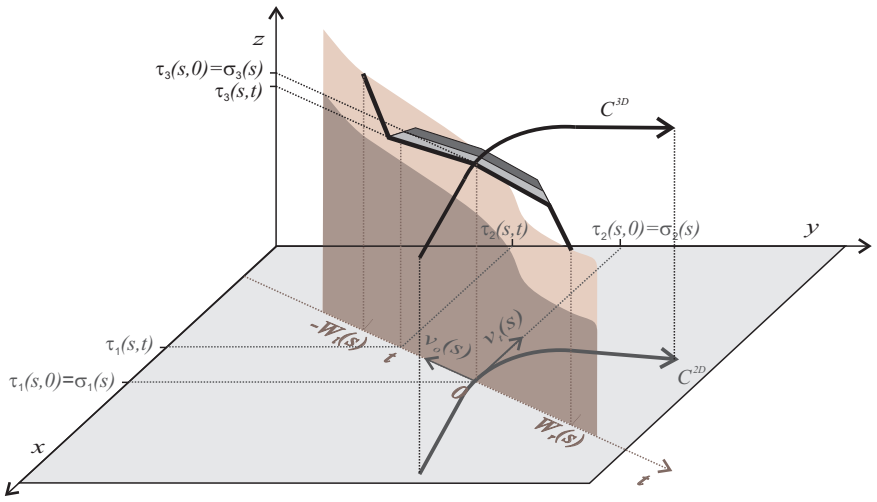

Figure 4: Road surface parametrization.

costs (fuel, travel time, accident risk) and social and environmental costs (noise, air pollution, loss of environmentally sensitive areas, loss of social-economic areas). Within these costs there are some which are purely economic and can be added on and brought all together into a unique functional to minimize, but there are also costs that include many other different aspects (social, environmental, travel time, accident risk, ...), which can not be combined and they give a strong multi-objective character to the optimization problem. In this work focus is set only on economical objectives, postponing to forthcoming research the multi-objective problem approach.

Leaving aside the structures (whose costs could be characterized similarly as done in Li et al. (2016)), the economic costs (construction and maintenance, mainly), can be gathered into three different types: location dependent costs, length-area dependent costs and earthwork costs (see, for instance, Jong and Schonfeld (1999), Jong and Schonfeld (2003)). As we will next see, parametrizations obtained in previous section allow to formulate and compute those three costs in an accurate and simple way.

3.1 LOCATION DEPENDENT COSTS

Land acquisition and also the preference/appropriateness for crossing or avoiding certain areas are included in this type of costs. If the horizontal layout is carried out on a domain $\Omega \subset \mathbb{R}^2$, we require a function $p : (x,y) \in \Omega \longmapsto p(x,y) \in \mathbb{R}$ gathering up the price for crossing through point (x,y). Initially we can think that we have a $p_a(x,y)$ ($[€/m^2]$) function giving the land acquisition price, and set $p = p_a$ (function p_a can be obtained thanks to a *geographic information system* (GIS), depending on the current land purpose). In addition, if the road has to avoid a particular region $R \subset \Omega$, then we should re-define function p, as $p = p_a + \chi_R$, where χ_R is the characteristic function for region R, given by

$$\chi_R(x,y) = \begin{cases} 0 & \text{if } (x,y) \in \Omega \setminus R, \\ p_R & \text{if } (x,y) \in R, \end{cases}$$

being $p_R >> 0$ a very large value in comparison with the land acquisition price. Function χ_R is clearly discontinuous but it can be replaced by any smooth approximation (see Vázquez-Méndez and Casal (2016)) if continuous rather than discrete algorithms are going to be used.

Once p is defined, it is also required to purchase w_a meters on both sides beyond the road limits and this acquisition is computed on the horizontal projection. In this case, location dependent costs will consist of the sum (integral) over the horizontal alignment (curve $C^{2D}_{\mathbf{x}^N}$) of function g compiling the cost of every cross-section, given by

$$g(s) = \int_{-W_l(s)-w_a}^{W_r(s)+w_a} p\left(\tau_1(s,t), \tau_2(s,t)\right) dt.$$

Thus, taking into account that $\sigma^{2D}_{\mathbf{x}^N}(s) = (\sigma_1(s), \sigma_2(s))$ is the parametrization of $C^{2D}_{\mathbf{x}^N}$ with respect to the arc-length parameter (hence $\sigma_1'(s)^2 + \sigma_2'(s)^2 = 1$), the localization costs turns out to be

$$J_A(\mathbf{u}^{N,M}) = \int_{C^{2D}_{\mathbf{x}^N}} g\, d\sigma = \int_0^{L(\mathbf{x}^N)} \left(\int_{-W_l(s)-w_a}^{W_r(s)+w_a} p\left(\tau_1(s,t), \tau_2(s,t)\right) dt \right) ds.$$

3.2 LENGTH-AREA DEPENDENT COSTS

By means of the appropriate curve or surface integral, depending on each case, the following costs are obtained:

1. Clearance cost: Since computations are on the horizontal projection, this cost is given by

$$J_C(\mathbf{u}^{N,M}) = p_d \times \text{Clearance Area} = p_d \int_0^{L(\mathbf{x}^N)} (W_l(s) + W_r(s)) \sqrt{1+\sigma_3'(s)^2}\, ds$$

with p_d ($[€/m^2]$) the price for ground preparation.

2. Pavement cost: Neglecting superelevation and crossfall, this cost is given by

$$J_P(\mathbf{u}^{N,M}) = p_p \times \text{Pavement Area} = p_p \int_0^{L(\mathbf{x}^N)} (W_{c_l}(s) + W_{c_r}(s)) \sqrt{1+\sigma_3'(s)^2}\, ds,$$

with p_p ($[€/m^2]$) paving price (including sub-grade preparation).

3. Maintenance cost:

$$J_M(\mathbf{u}^{N,M}) = p_{m_1} \times \text{Length of the road} + p_{m_2} \times \text{Pavement Area} = p_{m_1} \int_0^{L(\mathbf{x}^N)} \sqrt{1+\sigma_3'(s)^2}\, ds) + p_{m_2} \int_0^{L(\mathbf{x}^N)} (W_{c_l}(s) + W_{c_r}(s)) \sqrt{1+\sigma_3'(s)^2}\, ds,$$

where p_{m_1} ($[€/m]$) and p_{m_2} ($[€/m^2]$) are estimated road maintenance prices. These values should include preventive maintenance costs (as repairing pavements, guardrails, ...) and road rehabilitation costs. The road life at an appropriate interest rate should be also taken into account (see Kang et al. (2012)).

3.3 EARTHWORK COSTS

Earthwork cost is one of the major outlay in road construction. Currently, optimization research taking into account these expenses use approximations derived from the computation of the road central axis (see, for instance, Chew et al. (1989), Hirpa et al. (2016), Li et al. (2016)). These types of approximations may be useful for achieving an optimal design, but they turn out to be inaccurate in comparison with those where the shapes of each cross-section are taken into account. We present below an approximation of the earthwork obtained from the central axis of the road (useful for approaching the optimal alignment) and a more accurate computation, using the parametrization of the road surface given in Section 2.3. In both cases, we have not consider material transportation costs (these costs are deeply study, for example, in Hare et al. (2014)) or Beiranvand et al. (2017).

3.3.1 Earthwork over the road axis

As a first approximation, taking into account that the road cross-sections are the intersections between the road and the vertical planes, orthogonal to the horizontal alignment, we assume that, at each point, the earthwork is directly related to the height difference between the road central axis and the terrain. Under this hypothesis, the use of the road surface parametrization can be avoided, and the earthwork costs can be approached by

$$J_{EW}^{app}(\mathbf{u}^{N,M}) = p_{EW} \int_0^{L(\mathbf{x}^N)} (W_l(s) + W_r(s)) \, |\sigma_3(s) - H(\sigma_1(s), \sigma_2(s))| \, ds, \tag{9}$$

where p_{EW} ($[€/m^3]$) is the estimated price for earthwork, and function $H(x,y)$, set in $\Omega \subset \mathbb{R}^2$, gives the terrain height. This rough approximation can be improved as in Hare et al. (2015) at the cost of an increased number of the variables. However, in this paper we do not use that technique to keep the number of variables as low as possible in order to speed up the global search.

Remark 1 *Earthwork is realized in each cross-section (all defined in a vertical and orthogonal plane in relation to the horizontal alignment). This leads to establish its value as the integral given in (9), and not as a line integral over* $C^{3D}_{\mathbf{u}^{N,M}} \in \mathbb{R}^3$ *(see Chew et al. (1989)).*

3.3.2 Earthwork over the cross-section

Taking into account the parametrization of the road surface given in the previous section, we compute:

1. The cut volume:

$$V_c(\mathbf{u}^{N,M}) = \int_0^{L(\mathbf{x}^N)} \int_{-W_l(s)}^{W_r(s)} (H(\tau_1(s,t), \tau_2(s,t)) - \tau_3(s,t))_+ \, dt\, ds. \tag{10}$$

where $(.)_+$ positive part function given as $(x)_+ = \max(x, 0)$.

If additional information about ground composition is available, the cut volume can be classified according to the different encountered materials. Lets suppose that we have m different materials and that functions $0 = H_0(x,y) \leq H_1(x,y) \leq ... \leq H_m(x,y) = H(x,y)$ are giving the height distribution for those different appearances, where material i is located between H_i and H_{i-1} (see Figure 5). For $i = 1, \ldots, m$, we define

$$V^i(\mathbf{u}^{N,M}) = \int_0^{L(\mathbf{x}^N)} \int_{-W_l(s)}^{W_r(s)} (H_i(\tau_1(s,t), \tau_2(s,t)) - \tau_3(s,t))_+ \, dt\, ds, \tag{11}$$

and the cut volume for material i is given by

$$V_c^1(\mathbf{u}^{N,M}) = V^1, \;\; V_c^i(\mathbf{u}^{N,M}) = V^i - V^{i-1}.$$

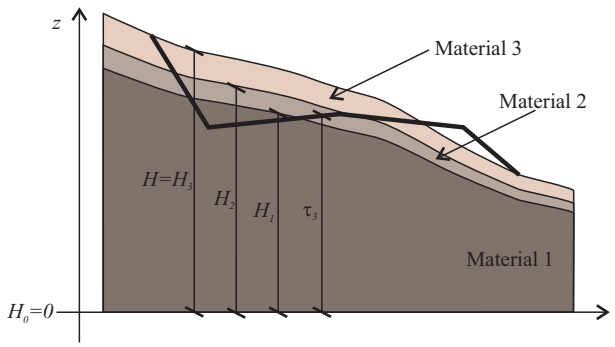

Figure 5: Material height distribution in a cross-section.

2. The ground fill volume is given by:

$$V_f(\mathbf{u}^{N,M}) = \int_0^{L(\mathbf{x}^N)} \int_{-W_l(s)}^{W_r(s)} (\tau_3(s,t) - H(\tau_1(s,t),\tau_2(s,t)))_+ \, dt\, ds. \tag{12}$$

3. The ground cut volume that can be re-used for filling

$$V_r(\mathbf{u}^{N,M}) = \min\left(V_f, \sum_{i=1}^{m} c_i r_i V_c^i\right),$$

where $c_i \in [0,1]$ is the efficiency (percentage of exploitable material i) and $r_i > 0$ is the shrinkage parameter of each material.

4. Material volume acquired for filling (borrow volume):

$$V_b(\mathbf{u}^{N,M}) = V_f - V_r,$$

5. Ground cut volume to dump (waste volume):

$$V_w(\mathbf{u}^{N,M}) = \sum_{i=1}^{m} (1-c_i) s_i V_c^i + \frac{\bar{s}}{\bar{r}} \left(\sum_{i=1}^{m} c_i r_i V_c^i - V_r\right)_+,$$

where s_i is the swell factor for material i, and $\bar{r}$ and $\bar{s}$ are mean values of parameters r_i and s_i, respectively.

From those volumes, the earthwork total cost is given by:

$$J_{EW}(\mathbf{u}^{N,M}) = \sum_{i=1}^{m} p_i V_c^i + p_r V_r + p_b V_b + p_w V_w, \tag{13}$$

where p_i ($[€/m^3]$) is the digging price for material i, p_r ($[€/m^3]$) is the filling price with re-used material, p_b ($[€/m^3]$) is the filling price with borrowed material and p_w ($[€/m^3]$) the price for ground waste management.

4 THE OPTIMIZATION PROBLEM. NUMERICAL RESOLUTION

Now, we are in the position to formulate the problem for road design minimizing construction and maintenance economic costs. It consists of considering the objective function $J : \mathbb{R}^{4N+3M} \longmapsto \mathbb{R}$ defined by $J = J_A + J_C + J_P + J_M + J_{EW}$, solving the problem

$$\min_{\mathbf{u}^{N,M} \in U_{ad}^{N,M}} J(\mathbf{u}^{N,M}) \tag{14}$$

for each $N, M \in \mathbb{N}$, and choosing the surface $S_{\mathbf{u}^{N,M}}$ corresponding with the lowest value of the objective function J.

Remark 2 *In a more complete formulation, variables N and M should be also decision variables and the problem should be studied in the framework of Mixed Integer Nonlinear Programming (MINLP). If these variables have to take low values (as we are assuming) an exhaustive search can be carried out on them. In this case, problem (14) should be solved for all possible combinations of N and M, and the solution corresponding with the lowest value of the objective function J should be chosen. On the contrary, if these variables can be high, an exhaustive search could require too much time and a more sophisticate technique can be necessary. Even so, in this case the dimension of problem (14) can be very high ($4N+3M$), and the main difficulty remains in obtaining a computational efficient method to solve it.*

For the numerical resolution of problem (14), it is worthwhile to remark that if smooth approximations of functions $(.)_+$ and $\min(.,.)$ are taken, and if W_l, W_r, H_i and p are smooth too, then function J will also be, and gradient-type algorithms can be used to solve the problem. On the other hand, keeping in mind that (14) is a non-convex problem, and that many local minima are expected, the use of a global optimization techniques becomes essential. Global optimization algorithms often require a high number of evaluations. In our cases, the most expensive calculation in order to evaluate functional J are all cross-section computations (especially the computation of width of side slopes, $W_{t_l}(s)$ and $W_{t_r}(s)$). Those computations can be avoided if we assume constant width ($W_l(s) = \bar{W}_l \in \mathbb{R}, W_r(s) = \bar{W}_r \in \mathbb{R}$) and we compute earthwork based on the central axis of the road, i.e., if we take $J_{app} = J_A + J_C + J_P + J_M + J_{EW}^{app}$ as objective function. This leads, once width $\bar{W}_l, \bar{W}_r \in \mathbb{R}$ are fixed, to consider the following approximated problem

$$\min_{\mathbf{u}^{N,M} \in U_{ad}^{N,M}} J_{app}(\mathbf{u}^{N,M}). \tag{15}$$

The computation of function $J_{app}(\mathbf{u}^{N,M})$ is fast (the road surface parametrization is no needed) and problem (15) can be solved using global optimization techniques. In addition, it is expected that solution of problem (15) will be close to the solution of problem (14) and therefore, the latter can be achieved employing a gradient-type algorithm using as initial step the solution of the former. Following these ideas, in this work we propose to solve problem (14) into two stages:

- **Stage 1:** Consider constant width $\bar{W}_l, \bar{W}_r \in \mathbb{R}$ and take as $\mathbf{u}_{app}^{N,M}$ the solution obtained solving problem (15) with

Constraints		Prices	Other parameteres
$R_{min} = 250\,m$	(minimum radius)	$p_d = 0.6€/m^2$	$\beta_l(s)$, $\beta_r(s) = 0.78\,rad$ if cut
$L^C_{min} = 70\,m$	(minimum length for clothoids)	$p_p = \frac{550}{15.15}€/m^2$	$\beta_l(s)$, $\beta_r(s) = -0.59\,rad$ if fill
$L^C_{max} = 108\,m$	(maximum length for clothoids)	$p_{m_1} = 0€/m$	$W_{c_l}(s) = W_{c_r}(s) = \frac{15.15}{2}\,m$
$L^T_{min} = 222\,m$	(minimum length for horizontal tangents)	$p_{m_2} = \frac{100}{15.15}€/m^2$	$\gamma(s) = 0$
$L^T_{max} = 1321\,m$	(maximum length for horizontal tangents)	$p_1 = 5€/m^3$	$w_a = 3\,m$
$m_{min} = 0.5/100$	(minimum slope)	$p_2 = 2€/m^3$	$H_1(x,y) = H(x,y) - 3$
$m_{max} = 5/100$	(maximum slope)	$p_r = 1.1€/m^3$	c_1 =0.97, $c_2 = 0.95$
$K_{v_{min}} = 5.4\,10^3\,m$	(minimum value for K_v parameter)	$p_b = 4.5€/m^3$	$r_1 = 1$, $r_2 = 0.8$
$L^{3D}_{min} = 225\,m$	(minimum horizontal distances between vertical curves)	$p_w = 2.0€/m^3$	$s_1 = s_2 = 1.25$
		$p_{EW} = 20€/m^3$	

Table 1: Constraints, prices and parameters used in the academic test (Section 5.1) and in the case study (Section 5.2).

global optimization techniques. Different stochastic, determinist or hybrid method can be used. In this work we compare:

- The genetic algorithm (GA) included in the Global Optimization Toolbox of MATLAB R2012a.
- The particle swarm pattern search method (PSwarm) suggested in Vaz and Vicente (2009). We have used the PSwarm MATLAB code, freely available from the PSwarm Home Page (http://www.norg.uminho.pt/aivaz/pswarm/).
- A multi-start application of NOMAD (Nonlinear Optimization with MADS) designed by Audet and Dennis, Jr. (2006) and already used in road alignment optimization by Mondal et al. (2015). NOMAD is included into the OPTI free MATLAB Toolbox, available at https://www.inverseproblem.co.nz/OPTI/.

At this stage, the main difficulty is to generate the initial population to start the algorithm (because *random* generation usually gives no admissible alignments) and this fact is being subject of active investigation.

- **Stage 2:** Assume variable width and solve problem (14) by a gradient type method, taking the solution obtained in Stage 1 ($\mathbf{u}^{N,M}_{app}$) as initial seed. In this work we have chosen the Sequential Quadratic Programming (SQP) algorithm (see Nocedal and Wright, 2006), which has been already successfully used by authors in horizontal alignment optimization (see Casal et al. (2017)).

5 NUMERICAL RESULTS

Numerical results obtained by solving problem (14) are now shown with the methodology described in the previous section, into two different situations. Firstly, in order to be able to analyse the *good behaviour* of the obtained solution, we will consider the academic example introduced in Casal et al. (2017), where it is relatively simple to understand the suitability of the layout. Secondly, we analyse a case study: a project for a by-pass of a Spanish road (N-640) avoiding the urban area and the industrial zone of Monterroso (small town in Northwest Spain).

To solve both problems we have developed an own MATLAB *script* where Algorithm A (Appendix A) and Algorithm 1 (Section 2.2) are implemented. In this *script*, for each input vector $\mathbf{u}^{N,M} \in U_{ad}$, the parametrization (for the central axis of the road or for its entire surface) is computed and, from it, the value of the objective function ($J_{app}(\mathbf{u}^{N,M})$ or $J(\mathbf{u}^{N,M})$) is also computed by means of numerical integration. Taking this script as the core, problem (15) has been solved with a global optimization technique (Stage 1), and the obtained solution has been taken as seed to solve problem (14) with the SQP method (Stage 2), computing the gradient by a finite-difference approximation, by using the Optimization Toolbox of MATLAB R2012a. All calculations were carried out with an Intel(R) Core™ i5-7200U CPU 2.50GHz PC.

5.1 AN ACADEMIC EXAMPLE

We take the following academic example: a 2D domain $\Omega = [-1,6] \times [-2,6]$ where the topography, given by a known function $H(x,y)$, can be seen in Figure 6(a)-(b) (the analytic description of function H can be found at Appendix A of Casal et al. (2017)). The aim is to look for the cheapest alignment connecting two terminals a=(0.25,1,H(0.25,1)) and b=(4,2,H(4,2)), taking into account restrictions corresponding to a $80\,km/h$ speed project road and prices shown in Table 1. We will assume (for sake of simplicity for the analysis of the obtained solution) that the land acquisition price has a constant value ($p_a(x,y) = 7€/m^2$).

In this situation, taking $N = 3$ turns and $M = 4$ slope changes, the problem has been solved by the two-stage

Optimization Method	J_{opt} (€)	Comp. Time (s)
two-stage methods		
GA+SQP	3.848.263	1.250
PSwarm+SQP	3.627.956	1.023
NOMAD+SQP	3.628.025	4.036
Multi-start applications on problem (14)		
NOMAD	3.924.969	6.094
SQP	3.689.451	4.956

Table 2: Numerical results and computation times for the different methods used in the academic example (Section 5.1).

Length of the layout	$4.187\,m$
Land acquisition cost(J_A)	798.089 €
Clearance cost (J_C)	53.361 €
Pavement cost (J_P)	2.302.758 €
Maintenance cost (J_M)	418.683 €
Cut volume cost ($\sum_{i=1}^{2} p_i V_c^i$)	37.228 €
Fill volume cost (re-used material) ($p_r V_r$)	15.441 €
Fill volume cost (acquired material) ($p_b V_b$)	97 €
Waste volume cost ($p_w V_w$)	2.299 €
Total cost (J)	3.627.956 €

Table 3: Length and costs breakdown for optimal alignment in academic example (Section 5.1).

method described in Section 4. In order to look for solutions in different corridors, at Stage 1 we have considered seven ad-hoc alignments to start NOMAD (a more general method to generate dissimilar alignments can be seen in Pushak et al. (2016)) and a population of 20 individuals for executing GA and PSwarm (the seven used in the NOMAD multi-start and other thirteen random layouts). In this numerical experiment, the combination of any of these three algorithms for solving (15), with the SQP method for solving (14), leads to the same layout. To numerically analyze the value of the global minimum in this experiment, we also have solved directly the problem (14) with a multi-start application of NOMAD and SQP methods. After larger computational times (see Table 2), SQP gives the same solution than our two-stage method, while NOMAD (which does not use gradient information) gives one slightly more expensive. The best solution, obtained by the combination of PSwarm with SQP, is

$$\begin{aligned}\mathbf{u}_{opt}^{3,4} = (&1.04,0.59,0.29,0.80,2.20,1.51,1.08,0.28,3.64,\\ &1.93,1.03,0.0003,0.19,0.04,6.23,0.54,0.03,\\ &5.78,0.64,0.009,14.45,0.90,-0.015,37.63),\end{aligned}$$

and the corresponding layout is shown in Figure 6 (the horizontal alignment is detailed in Figure 6(b) and Figure 6(c) shows the vertical alignment vs. the terrain height on the road central axis). The cost breakdown for this route is detailed in Table 3.

Taking into account that land acquisition price is constant, the cheapest alignment should be a short path and suitable for the land topography. In the absence of being able to prove that the obtained solution is the global optimum (due to the non-convexity of the problem), Figure 6 seems to show that it provides a suitable layout according to those two aspects.

5.2 CASE STUDY

In this section, we deal with the project for a by-pass road in N-640 avoiding the urban area and the industrial zone of Monterroso (Spain). We consider a square domain of about $16\,km^2$ centred on Monterroso town (see Figure 7). The intersections between the new detour and the existing road are already fixed (points a and b are known by UTM coordinates a =(593832,4738293) and b =(597295,4738941)). Terrain (function $H(x,y)$) is obtained by numerical interpolation from the contour grid lines data cut off every meter (see Figure 7(b)). The land acquisition price is also derived by GIS, by means of the plots rateable value. Function $p_a(x,y)$ is an approximation (obtained by spline interpolation), of the linear combination of characteristic functions corresponding to different priced areas (see Figure 7(a)). In addition, we set $p_R = 10^4 €/m^2$ as a penalty cost associated to prevent the layout crossing a restricted (urban or industrial) area. Other prices and parameter values considered, including alignment restrictions to fulfill Spanish road construction legislation (see SMPW (2016)), can be seen in Table 1 (all used data is public information and, if necessary, it is available upon request). Finally, in the definition of the admissible set X_{ad} we include bound constraints for the North-South coordinates of the horizontal alignment vertices (variables y_i, $i = 1,\ldots,N$, of the horizontal design vector $\mathbf{x}^N$), which allow to choose whether the by-pass path will go North or South of the current road.

We have used the two-step method (combining PSwarm and SQP) to obtain a North and a South by-pass alternatives. The horizontal alignment for both can be seen in Figure 7. Notice that the North variant seems to try to minimize the length, avoiding the urban area and a hill located between the two restricted zones. The southern alternative gives a lengthy detour to save the urban area and ends with a wide curve, adjusting to the industrial zone. The graphical interpretation is harder than in the academic example, but the initial detour seems to be due to the different acquisition prices and to the need to fit the terrain (see Figure 9). In any case, the ground adaptation in either two variants is pleasantly surprising, as

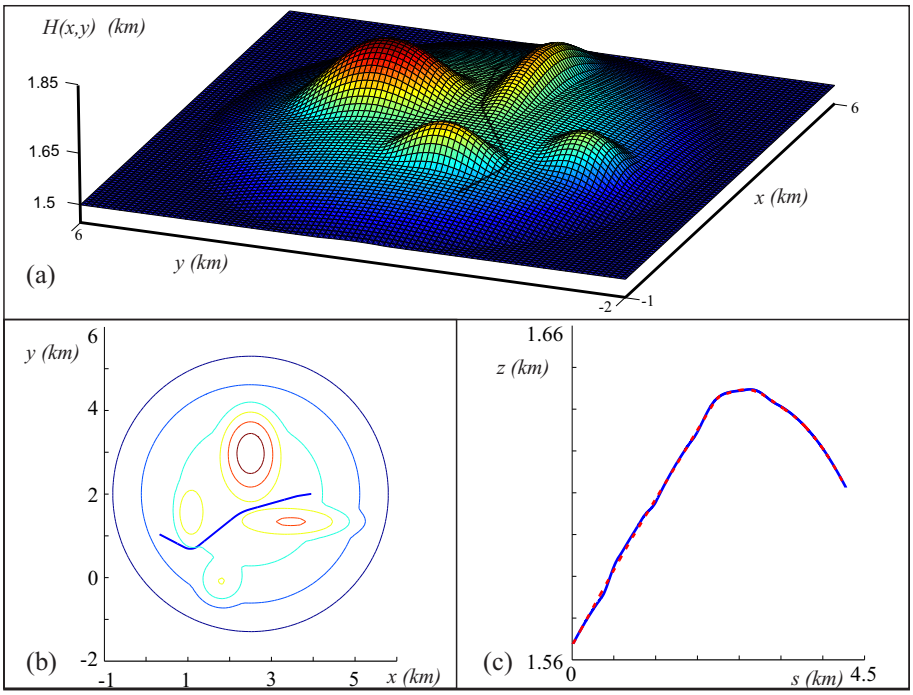

Figure 6: Optimal layout for the academic example detailed in Section 5.1: (a) 3D view of the road; (b) Horizontal alignment and some contour lines of height $H(x,y)$; (c) Comparison between the vertical alignment (dashed) and terrain height on the road central axis (solid).

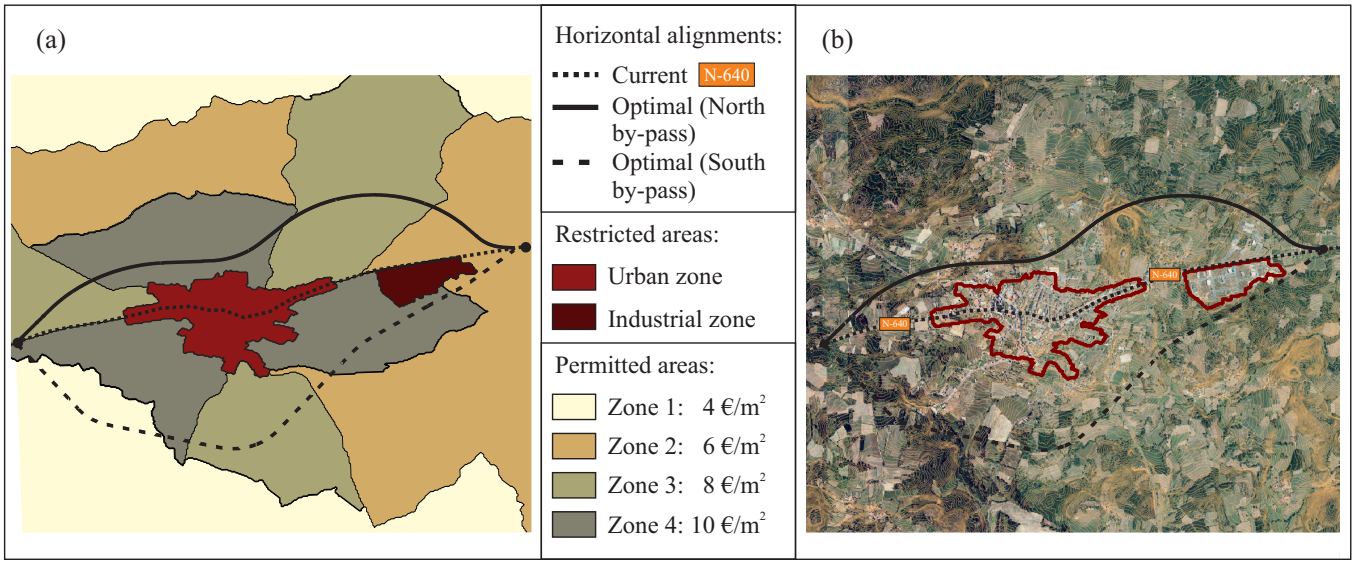

Figure 7: 2D view of Case Study: (a) Land cost map distribution (b) Orthophoto including terrain topography. Highlighted the computed horizontal alignment of both alternatives (North and South), the existing one (N-640) and Monterroso's site (Spain).

it can be seen in Figure 8 and 9. Table 4 shows that such good fitting generates very low earthwork costs in both cases, resulting in the North variant being cheaper than the South one.

6 CONCLUSIONS AND FUTURE WORK

In this paper we choose a set of design variables that unequivocally characterize the road alignment. From these design variables an algorithm is proposed to compute, not only the

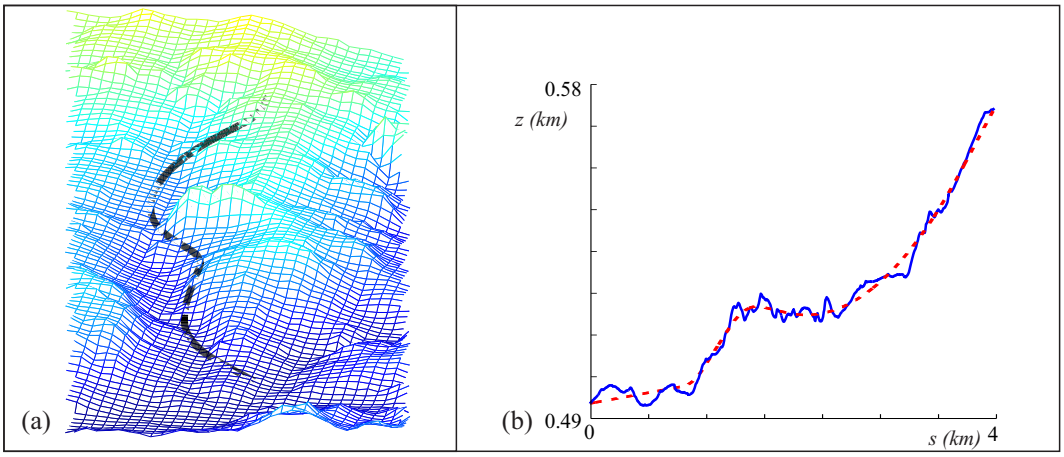

Figure 8: Terrain fitting of optimal North alternative: (a) 3D view of the road; (b) Comparison between the vertical alignment (dashed) and ground profile (solid).

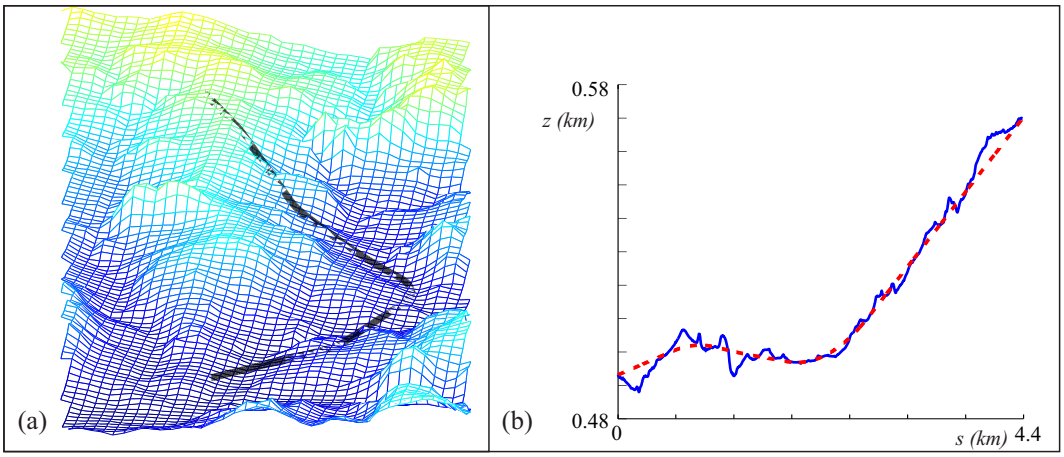

Figure 9: Terrain fitting of optimal South alternative: (a) 3D view of the road; (b) Comparison between the vertical alignment (dashed) and ground profile (solid).

parametrization of the road central axis, but also the road surface parametrization. These are useful because they are the fundamental basis for developing any layout software, but they are also very useful for expressing and computing, easily, the major infrastructure costs. From those parametrizations, the optimal alignment design is stated as a constrained, differentiable, non-linear and non-convex optimization problem. To solve it, we propose a two-stage algorithm, which combines global optimization techniques with a classical differentiable optimization method. The goodness of this algorithm is tested on an academic problem where it is relatively easy to understand the results. Finally, this methodology is applied to a case study for designing a by-pass detour on a Spanish national road, where it is noticed that the proposed technology turns out to be a very useful tool in this type of projects. In comparison with other recent works on the same topic, we highlight the following:

1. Regarding the layout:
   - Clothoids are include as transition between tangents and circular curves in the horizontal alignment.
   - Parabolic curves are included between uniform slopes.
   - Related to the cross-section, it should be noted that superelevation and crossfall are allowed, transition profiles (from cut to fill) are tolerated, and even the cut or fill embankment angles may depend on the soil material type.
   - Detailed parametrizations of the road central axis as well as the whole road surface are given.

	North detour	South detour
Length of the layout	3.990 m	4.449 m
Land acquisition cost (J_A)	977.285 €	1.469.676 €
Clearance cost (J_C)	55.473 €	63.704 €
Pavement cost (J_P)	2.194.500 €	2.446.929 €
Maintenance cost (J_M)	399.000 €	444.896 €
Cut volume cost ($\sum_{i=1}^{2} p_i V_c^i$)	179.275 €	311.704 €
Fill cost (re-used material) ($p_r V_r$)	60.351 €	92.521 €
Fill volume cost (acquired material) ($p_b V_b$)	37.129 €	19.883 €
Waste volume cost ($p_w V_w$)	7.818 €	10.714 €
Total cost (J)	3.910.831 €	4.860.027 €

Table 4: Length and costs breakdown for optimal alignments (North and South detours) in the Case Study (Section 5.2).

2. Regarding the optimization problem:

   (a) **Costs**: Only one objective function, embracing the main economic costs, is used. Working with a multi-objective problem taking into account social and environmental aspects, travel-time, accident risk, vehicle-operating, etc., is one of the future research lines.

   Regarding the costs dealt in this work, we underline the following:

   - Different acquisition price are admitted. In particular, this allows to include restricted areas where the alignment cannot cross.
   - Earthwork cost covers the possibility of different layer materials, with different digging prices. However, material transportation costs have not been optimized. In a forthcoming work, following the model of Hare et al. (2014), we will try to complete our formulation by including this type of costs in our objective function.
   - Structure costs such as tunnels, bridges or intersection with existing roads were not taken into account. They will be also included in a forthcoming work.

   (b) **Constraints**: Geometric restrictions over the elements of both horizontal and vertical alignments are included. For instance, in addition to bounds restrictions, also non-linear constraints are included over minimal and maximal length of clothoids, turns and tangents, as well as for horizontal distances between vertical curves. Other type of constrains (such as interrelations among vertical and horizontal alignments specified by national geometric design standards, or minimum distance to cross current existing infrastructures) can be included in the model and it will be studied in a further research.

   (c) **Resolution algorithm**: A two-stage method, combining global optimization techniques (to deal with the non-convexity of the problem) with a gradient type algorithm (to make the most of the smoothness of objective and constraint functions) is proposed. The obtained numerical results provide suitable layouts, but the generation of the initial population to start the global optimization is not simple and it is being subject of actual investigation (see Pushak et al. (2016)). This encourages searching *ad-hoc* methods preventing to sink in local minima and explode the advantages of the gradient type methods (see, for example, Yu et al. (2012)). The improvement and implementation of this kind of methods will also be a further research topic.

ACKNOWLEDGEMENTS

First author thanks the support given by Project MTM2015-65570-P of MEC (Spain) and FEDER. Third author thanks the support given by Project ENE2013-47867-C2-1-R. The authors are also very grateful to the Editor and the six anonymous reviewers for their very valuable suggestions.

APPENDIX A: COMPUTATION OF THE HORIZONTAL ALIGNMENT

We introduce the following functions and notation (see Figure 10):

- Unit vector giving direction (and sense) of tangent j,

$$\mathbf{u}_j(\mathbf{x}^N) = \frac{(x_j - x_{j-1}, y_j - y_{j-1})}{\sqrt{(x_j - x_{j-1})^2 + (y_j - y_{j-1})^2}}.$$

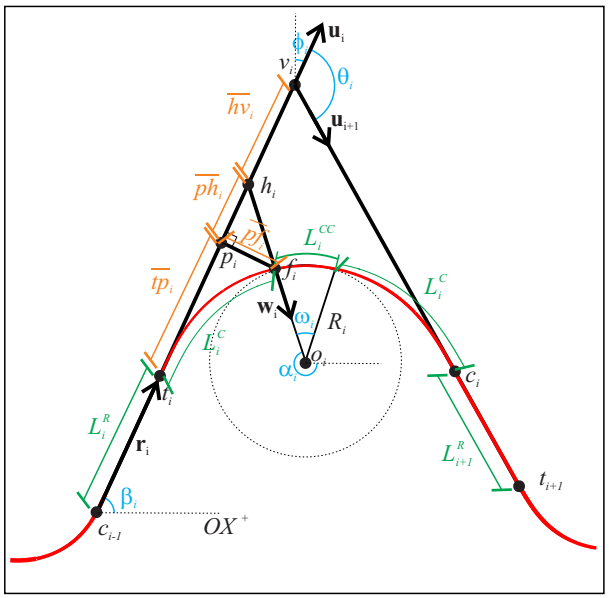

Figure 10: Naming convention of functions and variables involved in the horizontal alignment.

- Azimuth of tangent j, $\phi_j(\mathbf{x}^N)$ given by:
  If $x_j - x_{j-1} \geq 0$,

$$\phi_j(\mathbf{x}^N) = -\text{acos}\left(\frac{y_j - y_{j-1}}{\sqrt{(x_j - x_{j-1})^2 + (y_j - y_{j-1})^2}}\right).$$

  If $x_j - x_{j-1} < 0$,

$$\phi_j(\mathbf{x}^N) = 2\pi - \text{acos}\left(\frac{y_j - y_{j-1}}{\sqrt{(x_j - x_{j-1})^2 + (y_j - y_{j-1})^2}}\right).$$

- Difference of the azimuth between tangents i and $i+1$,

$$\theta_i(\mathbf{x}^N) = \left|\phi_{i+1}(\mathbf{x}^N) - \phi_i(\mathbf{x}^N)\right|.$$

- Length of circular curve i,

$$L_i^{CC}(\mathbf{x}^N) = R_i\omega_i. \tag{16}$$

- Length of each clothoid in turn i,

$$L_i^C(\mathbf{x}^N) = R_i(\theta_i(\mathbf{x}^N) - \omega_i). \tag{17}$$

- Junction between the straight segment i and the beginning of turn i,

$$t_i(\mathbf{x}^N) = v_i - \left(\overline{tp}_i(\mathbf{x}^N) + \overline{ph}_i(\mathbf{x}^N) + \overline{hv}_i(\mathbf{x}^N)\right)\mathbf{u}_i(\mathbf{x}^N).$$

  It can be clearly seen that (regard Vázquez-Méndez and Casal, 2016),

$$\overline{tp}_i = \int_0^{L_i^C} \cos\left(\frac{\tau^2}{2R_iL_i^C}\right) d\tau,$$

$$\overline{pf}_i = \int_0^{L_i^C} \sin\left(\frac{\tau^2}{2R_iL_i^C}\right) d\tau,$$

$$\overline{ph}_i = \overline{pf}_i \tan\left(\frac{\theta_i - \omega_i}{2}\right),$$

$$\overline{hv}_i = \left(R_i + \frac{\overline{pf}_i}{\cos(\frac{\theta_i - \omega_i}{2})}\right)\frac{\sin(\frac{\omega_i}{2})}{\sin(\frac{\pi - \theta_i}{2})}.$$

- Junction between the end of turn i and the beginning of the straight segment $i+1$,

$$c_i(\mathbf{x}^N) = v_i + \left(\overline{tp}_i(\mathbf{x}^N) + \overline{ph}_i(\mathbf{x}^N) + \overline{hv}_i(\mathbf{x}^N)\right)\mathbf{u}_{i+1}(\mathbf{x}^N).$$

- Length of straight segment j (*oriented* distance from the end of turn $j-1$ to the beginning of j),

$$L_j^T(\mathbf{x}^N) = \mathbf{r}_j(\mathbf{x}^N).\mathbf{u}_j(\mathbf{x}^N),$$

  where $\mathbf{r}_j$ is the vector starting at point c_{j-1} and ending at point t_j.

The length of the horizontal alignment prior to the beginning of turn j, is given by,

$$L_j(\mathbf{x}^N) = L_1^T(\mathbf{x}^N) + \sum_{k=1}^{j-1} 2L_k^C(\mathbf{x}^N) + L_k^{CC}(\mathbf{x}^N) + L_{k+1}^T(\mathbf{x}^N),$$

and, consequently, the total length of the horizontal alignment verifies $L(\mathbf{x}^N) = L_{N+1}(\mathbf{x}^N)$. The parametrization of the horizontal alignment (curve $C_{\mathbf{x}^N}^{2D} \subset \mathbb{R}^2$) in terms of the arc length parameter is the function

$$\sigma_{\mathbf{x}^N}^{2D} : s \in [0, L(\mathbf{x}^N)] \longmapsto \sigma_{\mathbf{x}^N}^{2D}(s) = (\sigma_1(s), \sigma_2(s)),$$

given by the next algorithm. Ingoing and outgoing clothoids therein are computed by an alternative method based on numerical integration (see Vázquez-Méndez and Casal, 2016), which results more efficient than classical method based on high degree Taylor's expansions of functions sin and cos. The greater computational efficiency lies in the fact that the calculation of each point of the spiral is made from the previous one, rather than evaluate those polynomials on every points.

Algorithm A: Computation of $\sigma_{\mathbf{x}^N}^{2D}(s)$

- *Initial straight segment:* `If` $s \in [0, L_1^T]$, `compute`

$$\sigma^{2D}(s) = a + s\mathbf{u}_1.$$

- `For i=1,...,N`
  - *Ingoing clothoid:* `If` $s \in (L_i, L_i + L_i^C]$, `compute`

$$\sigma^{2D}(s) = t_i + \left(\int_0^{\tilde{s}} \cos\left(\frac{\lambda_i\tau^2}{2R_iL_i^C} + \beta_i\right) d\tau, \int_0^{\tilde{s}} \sin\left(\frac{\lambda_i\tau^2}{2R_iL_i^C} + \beta_i\right) d\tau\right)$$

    where

- $\tilde{s} = s - L_i$,
- $\beta_i \in [0, 2\pi)$ is the angle between $\mathbf{u}_i$ and OX^+ (see Figure 10),
- $\lambda_i \in \{-1, 1\}$ gives the orientation of the ingoing clothoid (the clothoid starting at t_i): $\lambda_i = -1$ if it is traversed in the clockwise direction and $\lambda_i = 1$ in counter clockwise.

– *Circular curve:* `If` $s \in (L_i + L_i^C, L_i + L_i^C + L_i^{CC}]$, `compute`

$$\sigma^{2D}(s) = o_i + R_i \left(\cos\left(\alpha_i + \frac{\tilde{s}}{R_i}\right), \lambda_i \sin\left(\alpha_i + \frac{\tilde{s}}{R_i}\right) \right)$$

where

- $\tilde{s} = s - (L_i + L_i^C)$, $h_i = v_i - \overline{hv}_i \mathbf{u}_i$, $f_i = \sigma(L_i + L_i^C)$, $\mathbf{w}_i = \frac{f_i - h_i}{||f_i - h_i||}$, $o_i = f_i + R_i \mathbf{w}_i$, and
- $\alpha_i \in [0, 2\pi)$ the angle between $\mathbf{w}_i$ and OX^+ (see Figure 10).

– *Outgoing clothoid:* `If` $s \in (L_i + L_i^C + L_i^{CC}, L_i + 2L_i^C + L_i^{CC}]$, `compute`

$$\sigma^{2D}(s) = c_i + \left(\int_0^{\tilde{s}} \cos\left(-\frac{\lambda_i \tau^2}{2R_i L_i^C} + \beta_i\right) d\tau, \int_0^{\tilde{s}} \sin\left(-\frac{\lambda_i \tau^2}{2R_i L_i^C} + \beta_i\right) d\tau \right)$$

with $\tilde{s} = L_i + 2L_i^C + L_i^{CC} - s$.

– *Straight segment:* `If` $s \in (L_{i+1} - L_{i+1}^T, L_{i+1}]$, `compute`

$$\sigma^{2D}(s) = c_i + \tilde{s}\mathbf{u}_{i+1},$$

with $\tilde{s} = s - (L_{i+1} - L_{i+1}^T)$.

APPENDIX B: LIST OF SYMBOLS

a, b: Initial and final points of the road.
a^{2D}, b^{2D}: Initial and final points of the horizontal alignment.
N: Number of curves of the horizontal alignment.
M: Number of slope changes of the vertical alignment.
v_i, R_i, ω_i: Vertex, radius and angle of the horizontal circular curve i.
$\mathbf{x}^N$, $\mathbf{y}^M$: Horizontal and vertical design variable.
$C^{2D}_{\mathbf{x}^N}$: Horizontal alignment determined by $\mathbf{x}^N$.
L: Length of the horizontal alignment.
$\sigma^{2D}_{\mathbf{x}^N}$: Parametrization of $C^{2D}_{\mathbf{x}^N}$.
s: Arc length parameter.
θ_i: Difference of the azimuth between tangents i and $i+1$.
L_i^C: Length of each clothoid in turn i.
L_j^T: Length of straight segment j.
(S_j, Z_j): Parabolic curve vertex j of the vertical alignment.
S_j^{in}, S_j^{out}: Junctions between the input, output, tangents and the parabolic curve j.
L_j^S: Horizontal distance between the input and output tangents on the parabolic curve j.
m_j: Slope of the straight segment j of the vertical alignment.
s_j: Coordinate in s of the parabolic vertex j divided by the length of the horizontal alignment.
K_{vj}: Ratio between the length of the parabolic section and the difference of two consecutive slopes on vertex j.
$\mathbf{u}^{N,M}$: Design variable.
L_j^{3D}: Horizontal distance between vertices $j-1$ and j in the vertical alignment.
X_{ad}^N, Y_{ad}^M: Admissible sets of horizontal and vertical design variables, respectively.
$U_{ad}^{N,M}$: Admissible set of design variables.
$C^{3D}_{\mathbf{u}^{N,M}}$: road alignment determined by $\mathbf{u}^{N,M}$.
$\sigma^{3D}_{\mathbf{u}^{N,M}}$: Parametrization of $C^{3D}_{\mathbf{u}^{N,M}}$.
$S_{\mathbf{u}^{N,M}}$: Road surface determined by $\mathbf{u}^{N,M}$.
$\tau = (\tau_1, \tau_2, \tau_3)$: Parametrization of $S_{\mathbf{u}^{N,M}}$.
β_l, β_r: Left and right side slope angles.
W_{c_l}, W_{c_r}: Width of left and right road pavement.
W_{t_l}, W_{t_r}: Width of left and right slope.
W_l, W_r: Width of left and right road surface.
C_a: Distance between the height of the road and the height of its central axis.
γ: Angle given superelevation or crossfall.
p_a: Acquisition price.
p_R, χ_R: Penalty price and characteristic function of region R.
w_a: Extra width to be acquired.
J_A: Location dependent cost.
p_d, p_p: Ground preparation and paving prices, respectively.
p_{m_1}, p_{m_2}: Estimated road maintenance prices (per meter and square meter, respectively).
J_C, J_P, J_M: Clearance, pavement and maintenance costs, respectively.
J_{EW}^{app}: Approximate earthwork cost.
p_{EW}: Estimated price for earthwork.
H: Function giving the terrain height.
H_i: Function giving the terrain height of material i.
V_c^i, V_c: Cut volume for material i and total ground cut volume.
V_f: Ground fill volume.
V_r: Ground cut volume that can be re-used for filling.
V_b: Material volume acquired for filling (borrow).
V_w: Ground cut volume to dump (waste).
J_{EW}: Earthwork total cost.
s_i, c_i, r_i: Swell, efficiency and shrinkage parameters of material i.
$\bar{s}$, $\bar{r}$: Swell and shrinkage mean values.
J, J_{app}: Objective and approximated objective functions.

References

AASHTO (2011), *A Policy on Geometric Design of Highways and Streets, 6th Edition*, American Association of State Highway and Transportation Officials, Washington D.C.

Audet, C. and Dennis Jr., J.E. (2006), Mesh adaptive direct search algorithms for constrained optimization, *SIAM J. Optim.*, **17**(1), 188-217.

Beiranvand, V., Hare, W., Lucet, Y. and Hossain, S. (2017), Multi-haul quasi network flow model for vertical alignment optimization, *Eng. Optimiz.*, **49**(10), 1777-1795.

Bosurgi, G. and D'Andrea, A. (2012), A polynomial parametric curve (PPC-CURVE) for the design of horizontal geometry of highways, *Comput.-Aided Civ. Infrastruct. Eng.*, **27**(4), 304-312.

Bosurgi, G., Pellegrino, O. and Sollazzo, G. (2013), A PSO highway alignment optimization algorithm considering environmental constraints, *Adv. Transp. Stud.*, **31**, 63-80.

Casal, G., Santamarina, D. and Vázquez-Méndez, M.E. (2017), Optimization of horizontal alignment geometry in road design and reconstruction, *Transp. Res. Pt. C-Emerg. Technol.*, **74**, 261-174.

Chew, E.P., Goh, C.J. and Fwa, T.F. (1989), Simultaneous optimization of horizontal and vertical alignments for highways, *Transp. Res. Pt. B-Methodol.*, **23**(5), 315-329.

Davey, N., Dunstall, S. and Halgamuge, S. (2017), Optimal road design through ecologically sensitive areas considering animal migration dynamics, *Transp. Res. Part C: Emerging Technologies.*, **77**, 478-494.

Easa, S.M. and Mehmood, A. (2008), Optimizing design of highway horizontal alignments: new substantive safety approach, *Comput.-Aided Civil Infrastruct. Eng.*, **23**, 560-573.

Fwa, T. F., Chan, W. T. and Sim, Y. P. (2002), Optimal vertical alignment analysis for highway design, *J. Transp. Eng.*, **128**(5), 395-402.

Hare, W., Hossain, S., Lucet, Y. and Rahman, F. (2014), Models and strategies for efficiently determining an optimal vertical alignment of roads, *Comput. Oper. Res.*, **44**, 161-173.

Hare, W., Lucet, Y. and Rahman, F. (2015), A mixed-integer linear programming model to optimize the vertical alignment considering blocks and side-slopes in road construction, *Eur. J. Oper. Res.*, **241**(3), 631-641.

Hirpa, D., Hare, W., Lucet, Y., Pushak, Y. and Tesfamariam, S. (2016), A bi-objective optimization framework for three-dimensional road alignment design, *Transp. Res. Pt. C-Emerg. Technol.*, **65**, 61-78.

Jha, M.K., Schonfeld, P., Jong, J.-C. and Kim, E. (2006), *Intelligent Road Design*, WIT Press, Southampton.

Jong, J.-C. and Schonfeld, P. (1999), Cost functions for optimizing highway alignments, *Transp. Res. Record.*, **1659**, 58-67.

Jong, J.-C. and Schonfeld, P. (2003), An evolutionary model for simultaneously optimizing three-dimensional highway alignments, *Transp. Res. Pt. B-Methodol.*, **37**(2), 107-128.

Kang, M.-W., Jha, M.K. and Schonfeld, P. (2012), Applicability of highway alignment optimization models, *Transp. Res. Pt. C-Emerg. Technol.*, **21**(1), 257-286.

Kobryń, A. (2017), *Transition curves for highway geometrix design*. Springer Tracts on Transportation and Traffic, Vol. 14. New York.

Lee, Y., Tsou, Y. and Liu, H. (2009), Optimization Method for Highway Horizontal Alignment Design, *J. Transp. Eng.*, **4**, 217-224.

Li, W., Pu, H., Zhao, H. and Liu, W. (2013), Approach for optimizing 3D highway alignments based on two-stage dynamic programming, *J. Software*, **8**(11), 2967-2973.

Li, W., Pu, H., Schonfeld, P., Zhang, H. and Zheng, X. (2016), Methodology for optimizing constrained 3-dimensional railway alignments in mountainous terrain, *Transp. Res. Pt. C-Emerg. Technol.*, **68**, 549-565.

Li, W., Pu, H., Schonfeld, P., Yang, J., Zhang, H., Wang, L. and Xiong, J. (2017), Mountain Railway Alignment Optimization with Bidirectional Distance Transform and Genetic Algorithm. *Comp.-Aided Civil and Infrastruct. Engineering*, **32**, 691709.

Lovell, D.J. (1999), Automated calculation of sight distance from horizontal geometry, *J. Transp. Eng.*, **125**(4), 297-304.

Mondal, S., Lucet, Y. and Hare, W. (2015), Optimizing horizontal alignment of roads in a specified corridor, *Comput. Oper. Res.*, **64**, 130-138.

Nocedal, J. and Wright, S.J. (2006), *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering, Springer Science+Business Media, New York.

Pushak, Y., Hare, W. and Lucet, Y. (2016). Multiple-path selection for new highway alignments using discrete algorithms *Eur. J. Oper. Res.*, **248**, 415-427.

Shafahi, Y. and Bagherian, M. (2013), A Customized Particle Swarm Method to Solve Highway Alignment Optimization Problem, *Comp.-Aided Civil and Infrastruct. Engineering*, **28**, 52-67.

Spanish Ministry of Public Works, (2016), Order/FOM/273/2016 by adopting the Standard 3.1 IC of sections of the Road Instruction, *Spanish Oficial State Gazette BOE*, **55**, 17657-17893.

Vaz, A.I.F. and Vicente, L.N. (2009), PSwarm: a hybrid solver for linearly constrained global derivative-free optimization, *Optim. Methods Softw.*, **24**, 669-685.

Vázquez-Méndez, M.E. and Casal, G. (2016), The clothoid computation: a simple and efficient numerical algorithm, *J. Surv. Eng.-ASCE*, **142**(3), 04016005.

Yu, M.T., Lin, T.Y. and Hung, C. (2012), Sequential quadratic programming method with a global search strategy on the cutting-stock problem with rotatable polygons, *J. Intell. Manuf.*, **23**, 787-795.