



HAL
open science

Sparse Surrogate Model for Optimization: Example of the Bus Stops Spacing Problem

Valentin Vendi, Sébastien Verel, Cyril Fonlupt

► **To cite this version:**

Valentin Vendi, Sébastien Verel, Cyril Fonlupt. Sparse Surrogate Model for Optimization: Example of the Bus Stops Spacing Problem. Evolutionary Computation in Combinatorial Optimization Conference (evoCOP), Apr 2024, Aberystwyth, Wales, United Kingdom. hal-04502943

HAL Id: hal-04502943

<https://ulco.hal.science/hal-04502943v1>

Submitted on 13 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Sparse Surrogate Model for Optimization: Example of the Bus Stops Spacing Problem

Valentin Vendi¹[0009–0004–6687–2386], Sébastien Verel¹[0000–0003–1661–4093], and
Cyril Fonlupt¹[0000–0003–4729–4715]

Univ. Littoral Côte d’Opale, LISIC
Calais, France, F-62100
`valentin.vendi@univ-littoral.fr`

Abstract. Combinatorial optimization problems can involve computationally expensive fitness function, making their resolution challenging. Surrogate models are one of the effective techniques used to solve such black-box problems by guiding the search towards potentially good solutions. In this paper, we focus on the use of surrogate based on multinomial approaches, particularly based on Walsh functions, to tackle pseudo-Boolean problems. Although this approach can be effective, a potential drawback is the growth of the polynomial expansion with problem dimension. We introduce a method for analyzing real-world combinatorial black-box problems defined through numerical simulation. This method combines Walsh spectral analysis and polynomial regression. Consequently, we propose a sparse surrogate model that incorporates selected, relevant terms and is simpler to optimize. To demonstrate our approach, we apply it to the bus stop spacing problem, an exemplary combinatorial pseudo-Boolean challenge.

Keywords: Surrogate · Sparse model · Mobility problem · Walsh basis.

1 Introduction

In industry 4.0 or in academic research such as in chemistry science, ocean science, energy system or transportation system [30], digital twins have evolved into tools for modeling systems and analyzing them through numerical simulations. This evolution gave birth to Simulation-Based Optimization (SBO) [16] which aims to solve optimization problems based on numerical simulation. However, SBO leads to additional challenges. Mainly, evaluating a candidate solution may be highly time-consuming ranging from a few seconds to hours [8, 2]. Furthermore, the optimization problems are often considered as black-box problems where no algebraic definition is available. To address this problem, three main strategies can be used. Parallel approaches benefit from larger computational resources to reduce evaluation time. As the number of evaluated candidate solutions is limited, researchers also define optimization algorithms to increase the convergence rate toward the most promising solutions [24]. Lastly, Surrogate-Assisted Optimization (SAO) builds an algebraic model from evaluated solutions,

substituting the original optimization function with the surrogate to guide the search.

Although most research dealing with surrogate focus on numerical optimization [7], this article emphasizes on surrogate model in the case of pseudo-boolean optimization. Several surrogate models (see Section 2.1) such as neural network [15], random forest [17], kernel-based methods [42] and multinomial approaches are used, the latter being the most efficient [25]. The multinomial approaches decompose the original function into a multivariate polynomial. However, a significant limitation of multinomial approaches is the exponential increase in polynomial terms with order, which is a crucial parameter of the surrogate. Usually, this parameter is tuned by an expert based on the knowledge of interactions between variables. As the "shape", *i.e.* the algebraic structure of the model, of combinatorial real-world problems remains elusive, adding more knowledge into a multinomial surrogate model remains challenging even for an expert. Moreover, the primary goal of SAO is not to propose the most accurate machine learning model to approximate the original function, but to guide the search toward better candidate solutions. Thus, the surrogate model should be easy to optimize, and the optimal solution of the surrogate should guide the search to the optimal solutions of the original function.

The first objective of this paper is to propose a sparse surrogate model based on Fourier (Walsh) expansion, built with expert knowledge and easy to optimize using dynamic programming approach. The model is applied to the bus stop spacing problem, an optimization challenge involving selecting optimal bus stop positions in a city (see Section 2.3). This problem is defined using open source data, and the open source simulator MATSim for reproducibility. The second objective is to demonstrate how spectral Fourier/Walsh analysis of the real-world problem can be used to support hypothesis proposed by an expert in the application domain. To the best of our knowledge, this is the first spectral analysis of a real-world combinatorial optimization problem used to explore the algebraic properties of the fitness function. This work is a first step in turning a black-box optimization problem into a more transparent white/gray-box optimization problem [10].

The remainder paper is organized as follows. Section 2 reviews main works for combinatorial surrogate model, spectral analysis, and presents the bus stop spacing problem. Section 3 defines the sparse Walsh model. Section 4 presents the main experimental results on the spectral analysis, and the potential benefits of the proposed sparse model for optimization. The final section opens discussion, and perspectives of this work.

2 Related work

This section outlines the main elements of our work. Starting from the definition of surrogate models and their application in combinatorial optimization, we then presents the use of spectral analysis in a context of explaining a surrogate Walsh

model. Finally, we describe the bus stops spacing problem which serves as an illustration of our work.

2.1 Surrogate model for combinatorial optimization

Surrogate-Assisted Optimization (SAO) uses a surrogate model to approximate the original, computationally expensive fitness function. In its most basic form, an offline version builds a surrogate model from an initial sample of solutions and optimizes an acquisition function to generate promising solutions. This acquisition function could be the surrogate function itself or a criterion such as expected improvement, etc. that is guiding the search by selecting a new candidate solution according to the surrogate model. The online SAO algorithm is based on the two same main components (learning a surrogate model and optimizing the acquisition function) but updates the surrogate model with generated solutions to refine it over iterations. (See Algorithm. 1).

Algorithm 1 Surrogate-Assisted Algorithm.

```

 $X \leftarrow$  initial sample
repeat
   $M \leftarrow$  Learn surrogate model of  $f$  from  $X$ 
   $x^* \leftarrow$  Optimize w.r.t. an acquisition function based on  $M$ 
   $y^* \leftarrow f(x^*)$  using the numerical simulation
   $X \leftarrow X \cup \{(x^*, y^*)\}$ 
until stopping criterium

```

Surrogate models for combinatorial optimization have gained more and more attention [7] in recent years due to the progress of supervised machine learning techniques for learning heterogeneous data. Besides classical approaches from continuous optimization that replace the Euclidean distance by Hamming or other discrete distances in Kriging method [41] or kernel-based methods (Radial basis function) [27], advanced schemes use dedicated combinatorial structures. The BOCS method [6] uses multivariate polynomial of Boolean variables, estimated via Bayesian regression. This technique has been improved in COMBO [29] which uses a Cartesian product of graphs to represent discrete categorical variable in the framework of Bayesian optimization. More recently, to reduce the computation time of the parameters estimation of the multilinear polynomial model used for example in BOCS method, COMEX method [12] uses exponential weight updates from reinforcement learning. This approach, initially developed for pseudo-Boolean functions, has been extended to include categorical variables. [11].

Surrogate model based on Walsh functions [38], building multivariate polynomial models (see next Section 2.2), have proven to be effective in various contexts including interpolation where precise learning of Walsh coefficients is

necessary [14], numerical simulations with some noise [25], and in multiobjective combinatorial optimization [13].

However, the surrogate part of BOCS, COMBO, or COMEX shows accuracy for combinatorial optimization for problem dimensions up to 50 and is usually optimized by a basic Simulated Annealing algorithm. In contrast, models based on Walsh functions take care of the optimization part and the properties of Walsh functions through efficient local search or evolutionary computation algorithms.

2.2 Spectral analysis of Walsh model

The Walsh basis, sometime also called Fourier basis, of pseudo-boolean functions space is an orthogonal, normal, and finite basis [28]. Any pseudo-boolean function $f : \{0, 1\}^n \rightarrow \mathbb{R}$ can be uniquely represented as a linear combination of Walsh functions:

$$\forall x \in \{0, 1\}^n, f(x) = \sum_{I \subset [n]} \beta_I \varphi_I(x) \quad \text{with} \quad \varphi_I(x) = \prod_{i \in I} (-1)^{x_i}$$

$\beta_I \in \mathbb{R}$ is the coefficient associated to the Walsh function φ_I . Each Walsh function can be indexed by the subset of variables $I \subset [n]$. The order of a Walsh function φ_I is the size of the set I , *i.e.* the number of binary variables. As such, the Walsh expression can be written as a multilinear polynomial expression ranked by the order of terms from the constant term to the highest order which is the degree of the Walsh expansion:

$$f(\sigma) = \beta_\emptyset + \sum_{i \in [n]} \beta_i \sigma_i + \sum_{i < j \in [n]} \beta_{i,j} \sigma_i \sigma_j + \sum_{i < j < k \in [n]} \beta_{i,j,k} \sigma_i \sigma_j \sigma_k + \dots$$

Here $\sigma_i \in \{-1, 1\}$ corresponds to the binary variable x_i such that $\sigma_i = (-1)^{x_i}$.

The constant term β_\emptyset represents the average value of f across all solutions in $\{0, 1\}^n$: $\mathbb{E}_x[f] = \beta_\emptyset$. The orthogonal and normal properties of the basis imply that the total variance of f is equal to the sum of square of its non-constant terms: $\text{Var}_x[f] = \sum_{I \neq \emptyset} \beta_I^2$. As a consequence, each term can be easily understood. Each term $\beta_I \prod_{i \in I} \sigma_i$ indicates the interaction between the binary variables x_i for $i \in I$. The sign of β_I shows the sign of this interaction, while β_I^2 represents the part of f total variance of f explained by this interaction (*i.e.* the strength of interaction).

For instance, the linear terms β_i give the individual contribution of variables x_i to the fitness function, the terms $\beta_{i,j}$ is the quadratic contribution of variables x_i , and x_j , etc. From a geometric perspective, the variance of f is the square euclidean norm, and each β_I^2 is the squared length the projection onto the φ_I axis in the orthonormal Walsh basis. The square value β_I^2 is also called Fourier weight and defines the *spectral sample* of f (def. 1.17, and 1.18 [28]) as the probability distribution on $I \subset [n]$ is proportional to β_I^2 . Then, we can define the weighted degree of f which is the average degree weighted by β_I^2 : $\text{deg}_\beta(f) = \frac{1}{\sum_{I \in [n]} \beta_I^2} \sum_{I \in [n]} \beta_I^2 |I|$.

With the goal of understanding the main interactions between variables which contributes to fitness function, we use a spectral analysis of pseudo-boolean functions [28]. Basically, we analyze Walsh coefficients normalized by the total variance:

$$\bar{\beta}_I = \frac{\beta_I}{\sqrt{\sum_{J \neq \emptyset} \beta_J^2}}$$

The normalized coefficients explain the importance of the corresponding interaction by the ratio of the total variance explained: $\bar{\beta}_I^2 = \frac{\beta_I^2}{\text{var}_x[f]}$. The coefficient $\bar{\beta}_I^2$ is also called normalized amplitude spectrum [18]. Notice that the analysis of the amplitude spectrum has mainly been used to analyze the ruggedness of the fitness landscape [18, 35]. Actually, the ruggedness which is linked to non-linearity of the fitness function can be measured by the autocorrelation coefficient, and can be deduced from the Walsh coefficients [34]. In this work, we propose to analyze directly the normalized spectrum. The strength of interaction can be compared across different functions, surrogate scenario, etc. and can be represented according to the order of terms or other special properties of the terms. The spectrum also helps to understand the "shape" of a real-world optimization function which also contribute to design better benchmarks of combinatorial optimization problems.

2.3 Bus stops spacing problem

The bus stops spacing problem is a challenge in the field of transportation modeling [23, 43, 44, 31, 32, 22]. It involves determining the optimal placement of bus stops in a public transportation system within a certain area. Many approaches to this problem exist in order to optimize different criteria of public transportation such as passengers travel time [23, 43, 44], economic costs [31, 22], or environmental impact [32].

Various modeling approaches have been applied to this problem. Some utilize Thiessen polygons [43], or Voronoi diagrams [44] that use static properties of sub-divisions of the full road network for identifying optimal bus stop locations. Another approach uses a static model that incorporates multiple physical and economic constraints to compute the optimal bus stops spacing [31]. In [22], a bi-level optimization method aiming to minimize social costs of the transport system by finding the best positioning of bus stops evenly spaced along the road.

Agent-based simulators have emerged as another approach to design complex models of transportation systems. In [23] following [1], the authors use MATSim [21] as a simulation environment to compute bus passengers travel time according to their artificial scenario of mobility plan called SIALAC [25]. This method enhances the precision of the evaluation function, at the cost of an expensive computation time. In contrast to the milliseconds required for algebraic models, a single simulation of a scenario can take up to 1 minute to compute.

New opportunities have emerged with the introduction of the Eqasim [19] pipeline that allows to create mobility scenarios for MATSim based on open-

source publicly available data [20]. This pipeline has brought reproducible experiments in urban mobility context. It has been used to model various cities or regions such as Ile-de-France [20], Sao Paolo [3], or California [5]. The workflow of Eqasim from raw data to simulation is shown in Figure 1. For the sake of reproducibility, all data and code of this article are available: <https://gitlab.com/vvendi/offline-wsao>. Given the accuracy of such model based on real-world data, we use this pipeline, and MATSim in the context of bus stops spacing problem resolution as described in next section 4.

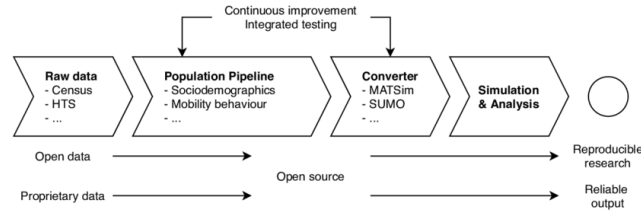


Fig. 1. Eqasim pipeline. Picture from the original article [19].

3 Sparse Walsh model

In Surrogate Assisted Optimization (SAO), a surrogate model is learned from a sample of evaluated solutions in order to approximate the original fitness function and to guide the search towards better solutions. Thus, the surrogate model has to be frugal to estimate using a small sample of solutions, and easy to optimize in order to guide the search efficiently. One main drawback of multi-linear polynomial surrogate such as Walsh expansion is the increase in the number of terms with the degree. The number of terms of order k is $\binom{n}{k}$ and the number of terms of degree d is then: $\sum_{k=0}^d \binom{n}{k}$. For example, for a degree 3 Walsh expansion, the number of terms is $1 + n + \frac{n(n-1)}{2} + \frac{n(n-1)(n-2)}{6}$. Indeed, many works try to use sparse surrogate models with a low number of terms (see Section 2.1) in order to increase the accuracy of the surrogate given the small sample size. Here, in this work we also propose to use a sparse model, but on the contrary to previous works mainly based on data-guided method, we will use expert knowledge to select the most relevant terms of the surrogate model and consider the difficulty of optimization of the surrogate model.

The mean square error of a surrogate function $\hat{f}(x) = \sum_I \hat{\beta}_I \varphi_I(x)$ to approximate the fitness function $f(x) = \sum_I \beta_I \varphi_I(x)$ is given by the distance between functions:

$$\text{mse}(\hat{f}) = \mathbb{E}_x[(\hat{f}(x) - f(x))^2] = \sum_{I \subset [n]} (\hat{\beta}_I - \beta_I)^2$$

For instance, when the surrogate model is a truncation \hat{f}_p of the original function f with p terms from a set $P \subset 2^{[n]}$ where: $\begin{cases} \hat{\beta}_I = \beta_I \forall I \in P \\ \hat{\beta}_I = 0 \quad \forall I \notin P \end{cases}$. Then, the mean square error of the truncation is: $\text{mse}(\hat{f}_p) = \sum_{I \notin P} \beta_I^2$. As a consequence, the truncation with p terms minimizing the mean square error is composed by the p terms with the highest β_I^2 -values. The design of a surrogate by an expert would push to select the most important variables interactions which influence the objective function. Heuristically based on knowledge of the optimization problems, some researchers propose to bound the degree of the Walsh expansion [6, 14, 13], but the number of terms can still be too large compare to the sample size. In feature selection machine learning problems [26], where the goal is to select the most important predictors for a machine learning algorithm, block model [33, 4] is used which supposes that the features are clustered into families of variables. In this work, we go in this direction by hypothesizing that the order of variables in the binary string representation of a solution is not random, and gives useful information to exploit. We suppose that variables close in the representation impact the fitness function, and those far in the representation do no interact. Let's define the distance between variable x_i , and x_j in the binary string x as the distance between indexes: $|i - j|$, and the diameter of the set of variables I by: $D(I) = \max\{|i_1 - i_2| : (i_1, i_2) \in I^2\}$. The sparse surrogate model of degree d , and lag ℓ is defined by:

$$\hat{f}_{d,\ell}(x) = \sum_{\substack{I \subset [n] \\ \text{s.t. } |I| \leq d, D(I) \leq \ell}} \hat{\beta}_I \varphi_I(x)$$

The number of terms is reduced compared to a simple full expansion of degree d . For a degree d , we got $(n - \ell) \binom{d-1}{\ell} + \sum_{k=d-1}^{\ell-1} \binom{d-1}{k}$ terms. For example, for a problem dimension n , the sparse model of degree 2 with lag ℓ has $1 + n + (n - \ell)\ell + 1$ terms, which is linear with problem dimension, and not quadratic. Indeed, in evolutionary computation, this sparse model is known as k -bounded Walsh model [36], where each sub-function of the model depends only on k others variables.

Moreover, for k -bounded functions, a dynamic programming approach have been proposed to find in polynomial time the global minimum [40, 39]. The same algorithm can be used for the sparse model of degree d and lag ℓ . First, a Walsh function f_n on $\{0, 1\}^n$ can be split in two parts, the terms that do not contain the variable x_n , and the terms that contain x_n : $\forall x \in \{0, 1\}^n$, $f_n(x) = f_{n-1}(x) + F_n(x)$ with $F_n = \sum_{I \subset [n] \text{ s.t. } n \in I} \beta_I \varphi_I(x)$. The lag ℓ ensures that F_n depends only on variables $x_{n-\ell}, \dots, x_n$. So, the common variables between f_{n-1} , and F_n are the ℓ variables $x_{n-\ell}, \dots, x_{n-1}$:
 $f_n(x_1, \dots, x_{n-\ell}, \dots, x_n) = f_{n-1}(x_1, \dots, x_{n-\ell}, \dots, x_{n-1}) + F_n(x_{n-\ell}, \dots, x_{n-1}, x_n)$

When those variables are fixed, the two parts f_{n-1} , and F_n are independent, and the global minimum of f_n is the sum of the minima:

$$\begin{aligned} \min_{x \in \{0,1\}^n \text{ s.t. } (x_{n-\ell}, \dots, x_{n-1})=s} f_n(x_1, \dots, x_{n-\ell}, \dots, x_{n-1}, x_n) = \\ \min_{x \in \{0,1\}^{n-1} \text{ s.t. } (x_{n-\ell}, \dots, x_{n-1})=s} f_{n-1}(x_1, \dots, x_{n-\ell}, \dots, x_{n-1}) + \min_{x_n \in \{0,1\} \text{ s.t. } (x_{n-\ell}, \dots, x_{n-1})=s} F_n(x_{n-\ell}, \dots, x_{n-1}, x_n) \quad (1) \end{aligned}$$

Defining the state S of the dynamic programming algorithm as the possible values of the variables $x_{n-\ell}, \dots, x_{n-1}$, the Equation 1 defines the recurrence formula to update the state of the algorithm. The last step of the algorithm selects the state value with the minimum fitness value. Notice that the state size is 2^ℓ , and the complexity is bounded by $\Theta(n2^\ell)$ which is linear with problem dimension.

4 Experimental analysis

4.1 Bus stop spacing problem design

In this section, we focus on an instance of the bus stops problem in the city of Calais, France. We make use of the MATSim simulator [21] coupled with the Eqasim pipeline [19] to run simulations from real-world data [20], which makes our simulation of the city of Calais very precise and reproducible, but with the drawback of a very expensive computation time, around 30 minutes per execution. We treat the bus stops spacing problem as a pseudo-boolean problem by modeling it as follows: All possible locations for bus stops are defined and represented as a binary variable representing the state of activation of the according bus stop. A solution is thus a binary string that represents the bus stops to activate in the simulation. In our case, we work on the bus line nÂ°2 in Calais, which contains 76 bus stops, and we arbitrarily add 32 additional bus stops that could be potential candidates to improve the existing route for a total of 108 bus stops. Notice that this problem dimension is large in comparison with previous artificial benchmarks for surrogate-based combinatorial optimization [6, 14]. Figure 2 shows the outward route and the return route of bus. A few of these stops perfectly fit with the original bus route, while others diverge, leading to detours. The purpose of these off-route stops is to introduce diverse options that our algorithm must identify and exclude from the final selection of bus stops. The 108 bus stops are ordered in a binary string solution such that the stop represented by the next bit is the next stop geographically. Our fitness function defines a pseudo-Boolean problem: $f : \{0, 1\}^n \rightarrow \mathbb{R}$ with a search space of binary strings of dimension $n = 108$, and the fitness function computes the average travel time of people using bus expressed in seconds. In practice, the 3-steps process to evaluate a candidate solution is the following:

- Edit relevant Eqasim input files, *i.e.* GTFS files which describe the public transportation system (schedule), according to the solution to evaluate
- Run Eqasim pipeline to generate MATSim input files
- Run MATSim simulation and get quality criterium *i.e.* average travel time of people using bus

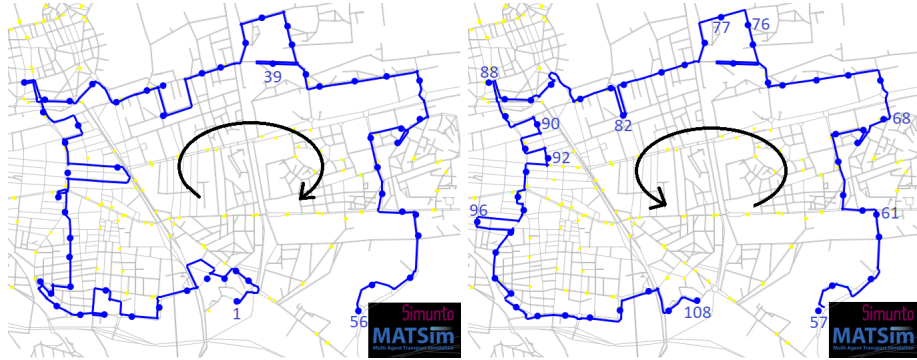


Fig. 2. Bus stops on the outward route and on the return route

4.2 Spectral analysis

In this section, the Walsh/Fourier spectral analysis is computed to analyze (i) the most relevant degree of Walsh expansion which is the main parameter used in polynomial surrogate models (See Section 2.1), (ii) the relevance of the lag parameter of the proposed sparse model. The search space dimension of bus stop spacing problem is too large ($n = 108$) to compute exactly the Walsh coefficients. So, we compute the exact value of Walsh coefficients using full enumeration of solutions from a family of sub-spaces. In our work, we used sub-spaces composed by 8 contiguous bits in the binary string x which correspond to the 8 contiguous bus stops: from bit x_i to bit x_{i+7} for $i \in \{1, 5, 9, \dots, n - 7\}$. In each of these sub-spaces, the bits that were not in the range of x_i to x_{i+7} were set to 1 *i.e.* the bus stop associated was turned on. Of course, this choice introduces bias in the estimation of Walsh coefficients, but it seems to be more meaningful (instead of a random binary value) for the bus stop problem. A same Walsh coefficient may appear in several sub-spaces. To estimate the weights, we compute the average of the normalized Fourier weight $\bar{\beta}_I^2$ across all sub-spaces which contains the coefficient $I \subset [n]$.

Figure 3 (left side) shows the distribution of the normalized Fourier weights $\bar{\beta}_I^2$ according to the order. The values are heterogeneous, so notice the log-scale of values. The terms of order 1, and 2 are the most important terms in the expansion. The median of order 1 weights is 17 times larger than order 2, and the median of order 2 weights is 1.7 times larger than order 3. From order 4 to 8, the ratio is much smaller. The sum of normalized Fourier weights of order 1 is 0.544, so 54.4% of the variance is explained by order 1 coefficients. Similarly, 79% of the variance is explained by order 1 and 2 combined. The weighted degree of the Walsh expansion (See definition in Section 2.2) is 1.66 between the order 1 and 2. Those first results suggest that a degree 2 Walsh expansion should bring an relevant surrogate model.

First, we analyze more precisely the linear impact of binary variables. Figure 4 shows the distribution of the logarithm of the normalized Fourier weights

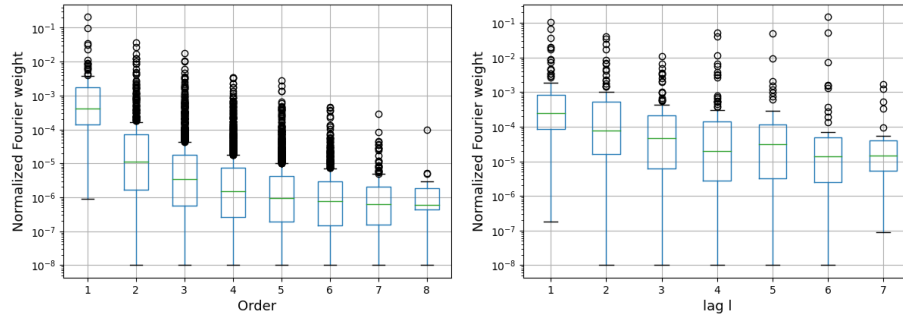


Fig. 3. Normalized Fourier weight. Left: as a function of order. Right: only for order 2 terms as a function of lag. Notice the y-log scale.

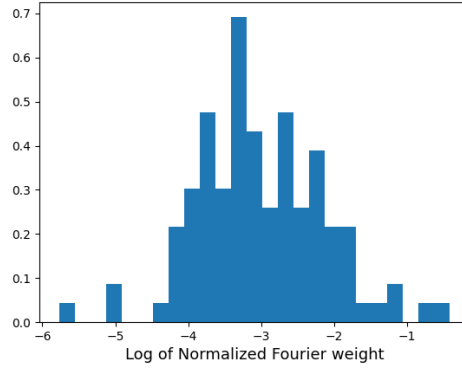


Fig. 4. Distribution of the logarithm of normalized Fourier weights $\bar{\beta}_I^2$ of order 1.

$\bar{\beta}_I^2$ of order 1. The distribution shapes an unimodal distribution. The median is around 0.7×10^{-3} , and the standard deviation of the log-values is 0.877. The distribution shows the Fourier weights are highly heterogeneous. Some linear contributions of bus stop are negligible, and at the opposite, some bus stops highly impact the fitness function. The 10 most important variables in decreasing rank of importance have identifier 96, 88, 82, 92, 90, 39, 76, 61, 77 and 68. Most of them match some extrapolated bus stops on the map that we can see in the Figure 2. Indeed, the Fourier/Walsh analysis is easily explainable on the map for an expert in mobility transportation. Focusing on Walsh coefficients of order 2, Figure 3 (right side) shows the distribution using boxplot of normalized Fourier weights of order 2 coefficients according to the lag parameter which is the distance $|i - j|$ in binary string between variables i , and j . The weights importance decreases with the lag. The median of the lag 1 weights is 3.24 times larger than the median of lag 2, and the lag 2 median is 1.67 times larger than the median of lag 3 coefficients. From lag 4 to 8, the ratio of medians is much

smaller. Coefficients with lag 1 and 2 explain 51.4% of the order 2 coefficients variance. Indeed, some very specific coefficients with lag 6 impact the fitness function as detailed in the next paragraph.

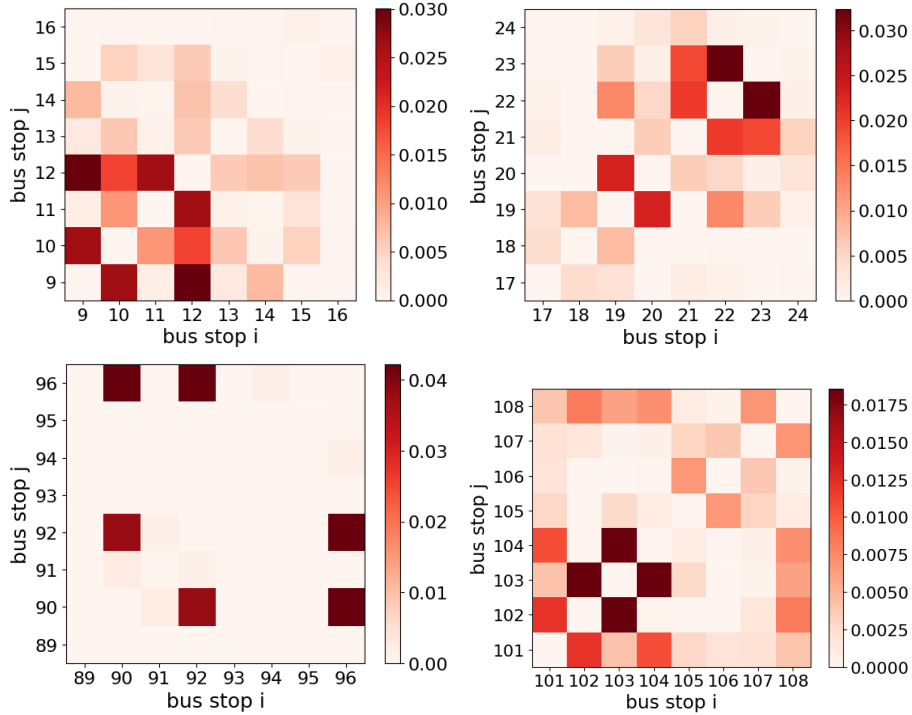


Fig. 5. Fourier spectrum of order 2 coefficients for 4 sub-spaces. The matrix value (i, j) displays the normalized Fourier weight $\bar{\beta}_{i,j}^2$ of the quadratic term which corresponds to bus stops with ids i , and j . Color is the intensity of $\bar{\beta}_{i,j}^2$ -values.

Figure 5 shows the Fourier spectrum of order 2 terms for 4 specific sub-spaces with variables (x_i, \dots, x_{i+7}) . The normalized Fourier weight $\bar{\beta}_{i,j}^2$ of term for variables i, j is displayed in matrix at position (i, j) . Thus, the matrix is symmetric, and the values on the diagonal are set to zero. All sub-spaces can not be displayed, so 4 representatives examples are selected. The one on the top-left shows a generic case, the highest weights are close to the diagonal *i.e.* are close to each others, with the highest weight representing the interaction between two stops at distance 3. The top-right with sub-space $(17, \dots, 24)$ is another example very similar to the top-left one. Once again, the highest weights are close to the diagonal with the maximum weight between the two stops 22 and 23 at distance 1. The sub-space with variables $(89, \dots, 96)$ at the bottom-left is a specific case where the terms with the highest weights are not the ones

with the smallest lag distance near the diagonal. Indeed, bus stops 90, 92 and 96 are among those with the highest linear importance and also show quadratic importance even though the lag is equal to 6. These bus stops are displayed on Figure 2 on the return route. Intuitively, we can make the hypothesis that this combination of "side-roads" would have an impact on the travel time, which is confirmed with the Fourier analysis of the fitness function. The sub-space with variables (101, . . . , 108) at the bottom-right is an example where all weights are pretty low leading to an uniform distribution of weights, but the most important quadratic terms are still close to the diagonal. The spectral analysis explains the degree of a relevant Walsh expansion, the main interaction between variables, and highlight the sparsity of an efficient surrogate model that can be used. The next step is to compare the candidate sparse models with the previous proposed models in the literature.

4.3 Sparse Walsh model quality

In this section, the quality of the sparse Walsh surrogate models is compared to the quality of the full Walsh surrogate [25]. We follow the sparse regression method based on the classical LASSO-LARS algorithm from [25, 23, 13] to estimate the parameters of the Walsh expansion which contains all terms of order below a given degree k , so called full Walsh expansion of degree k in this article. Figure 6 (left) compares the R^2 coefficient of determination (part of explained variance) of the different surrogate models: full Walsh expansion of degree 1, and of degree 2, and the sparse Walsh model with lag $\ell = 4$ of degree 2, and of degree 3. The R^2 coefficient is estimated on an independent random test set of size 400, and the surrogate is trained on random sample size up to 3,600 solutions. As in previous studies [38], for small sample size, linear Walsh expansion with only $n + 1$ terms is more accurate than quadratic Walsh expansion with $1 + n + n(n - 1)/2$ terms, but becomes more accurate for larger sample size of 2,000 solutions. However, sparse Walsh model of degree 2 outperforms full Walsh expansion of degree 1 and 2 for almost any training sample size, but also the sparse model of degree 3 until the training set reach a size of 2700 solutions which confirms the previous analysis of Section 4.2 on the importance of low order terms. Remember that computation time to evaluate a single solution is about 30 minutes making the maximum training sample size very large compared to typical computation effort of an optimization algorithm (which is often $10n$), in this regard, we can say that the sparse model of order 2 is better than the other tested models. Figure 6 (right) compares the R^2 regression quality of sparse model of degree 2 for different lag parameter values. For small train sample until size 700, all sparse models have approximately the same quality. Between sample size 700, and 2,200, the sparse with lag 2 slightly outperforms the other ones, and for the largest sample size, sparse model with lag 1 is outperformed by all the other ones. Overall, the sparse model of degree 2 with lag $\ell = 2$ seems to be efficient according to R^2 quality.

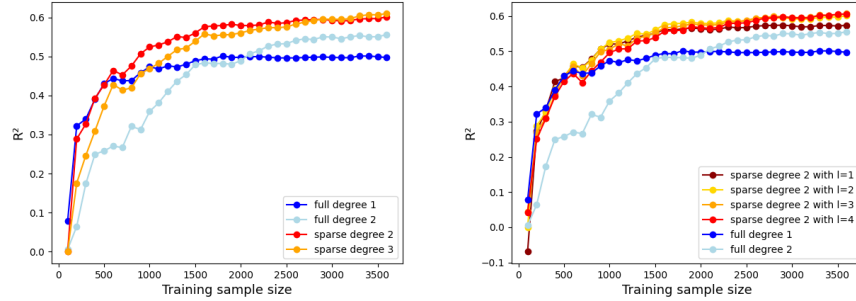


Fig. 6. R^2 estimated on a test set according to train sample size for different surrogate models. Left: comparison of degrees with sparse models of lag $\ell = 4$. Right: comparison of different lags with degree 2 models.

4.4 Optimization with the sparse surrogate model

The sparse Walsh model is designed to incorporate expert knowledge in the surrogate model for which the relevance is supported by the previous analysis, but also to be efficient to optimize. Indeed, in Section 3, we show that the optimal solution of the sparse surrogate can be found in linear time with problem dimension. In addition to the regression quality of the sparse surrogate, in this section we show the ability to find solutions with low fitness value using sparse surrogate model for the bus stop spacing problem instance.

In this preliminary study, we do not follow the anytime surrogate-assisted algorithm 1 and only use the offline version. Only the original fitness quality of the solutions which are the minimum of the surrogate function is analyzed. We compared two different surrogates models: the first one is a full Walsh model of degree 2, and the other one is a sparse Walsh model of degree 2 with lag $\ell = 2$ (see Section 4.3). For each surrogate model, we learn 30 surrogate models trained on independent random samples of medium size $5n = 540$, large size $10n = 1080$, and very large size of 3000 solutions. Medium, and large size are typical sizes used in the literature [6, 12]. Random solutions are sampled from the original random sample of 4,000 solutions by random sub-sampling technique. To optimize each surrogate models, we used the efficient iterated local search DRILS that uses partition cross-over, and the variable interaction graph perturbation¹ [10, 37] for the full Walsh surrogate models; and dynamic programming (Section 2.1) for the sparse Walsh surrogate models. Table 1 shows the average fitness obtained, and the Mann-Whitney statistical test at the level of 5%. As a comparison, the average fitness of the 4,000 random solutions is 2,746 with a standard deviation of 13.1 and a minimum of 2,712. The fitness of the existing bus stops in Calais is 2,698. The sparse model with dynamic programming outperformed the full Walsh model with DRILS for medium, and large sizes with a larger difference for

¹ recommended parameters values, and stopping criterion at 2 seconds of computation

the medium size. For very large size, both methods have similar performances, however notice that the variance of sparse models is always smaller which suggests a more robust optimization process. The sparse model of degree 2 and lag $\ell = 2$ is able to guide the search toward promising good solutions, and seems to converge quicker *w.r.t.* to sample size than the full model of degree 2.

Table 1. Average, and standard deviation of original fitness obtained from the optimization of different surrogate models for different size of training sample size $|X|$. Bold highlights statistical difference according to Mann-Whitney test with level of 5%.

$ X $	Full Walsh model	Sparse Walsh model
540	2,715 (7.5)	2,697 (4.8)
1,080	2,704 (4.5)	2,694 (3.6)
3,000	2,692 (3.7)	2,691 (1.9)

5 Conclusions, and perspectives

In this paper, inspired by block-model, we propose a sparse Walsh surrogate model which incorporates expert knowledge based on the hypothesis that the representation, *i.e.* the encoding, of solution is not random: close variables in the encoding could interact more than the other ones. Moreover, the sparse model is dedicated for efficient optimization. We propose a Walsh/Fourier spectral analysis of the fitness function of a real-world problem. This analysis shows that it is possible to use surrogate models based on Walsh expansion to help the expert to understand the real-world problem in addition to good optimal solutions given by the optimization process. As such, this work is a step forward to explainable optimization to bring decision supports around an optimization problem.

This initial works including the spectral analysis could be extended to other black-box combinatorial problems either to problems with binary representations, or to more complex representation such as permutations space [9]. In this work, as a first step for the optimization process, we use an offline optimization scenario. Anytime surrogate-assisted optimization which updates the sample of solutions during the optimization process has to be tested. Obviously, we plan to deal with other public transportation plans for the city of Calais with other criteria of interest, but also for larger cities, where the number of decision variables is much larger than the state-of-the-art surrogate-assisted algorithms, and requires parallel optimization algorithms.

Acknowledgments

Experiments presented in this paper were carried out using the CALCULCO computing platform, supported by DSI / ULCO. We thank Sebastian Hörl for his help in setting up the Eqasim environment and the modification he made for our research project.

References

1. Armas, R., Aguirre, H., Tanaka, K.: Multi-objective optimization of level of service in urban transportation. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 1057–1064 (2017)
2. Armas, R., Aguirre, H., Zapotecas-Martínez, S., Tanaka, K.: Traffic signal optimization: minimizing travel time and fuel consumption. In: International Conference on Artificial Evolution (Evolution Artificielle). pp. 29–43. Springer (2015)
3. Aurore Sallard, M.B., Hörl, S.: An open data-driven approach for travel demand synthesis: an application to são paulo. *Regional Studies, Regional Science* **8**(1), 371–386 (2021). <https://doi.org/10.1080/21681376.2021.1968941>
4. Bai, Z., Nguyen, H., Davidson, I.: Block model guided unsupervised feature selection. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1201–1211 (2020)
5. Balac, M., Hörl, S.: Synthetic population for the state of california based on open data: Examples of the san francisco bay area and san diego county (02 2021)
6. Baptista, R., Poloczek, M.: Bayesian optimization of combinatorial structures. In: International Conference on Machine Learning. pp. 462–471. PMLR (2018)
7. Bartz-Beielstein, T., Zaefferer, M.: Model-based methods for continuous and discrete global optimization. *Applied Soft Computing* **55**, 154–167 (2017)
8. Branke, J.: Simulation optimisation: tutorial. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. pp. 862–889 (2019)
9. Chicano, F., Derbel, B., Verel, S.: Fourier transform-based surrogates for permutation problems. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 275–283 (2023)
10. Chicano, F., Whitley, D., Ochoa, G., Tinós, R.: Optimizing one million variable nk landscapes by hybridizing deterministic recombination and local search. In: Proceedings of the genetic and evolutionary computation conference. pp. 753–760 (2017)
11. Dadkhahi, H., Rios, J., Shanmugam, K., Das, P.: Fourier representations for black-box optimization over categorical variables. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 10156–10165 (2022)
12. Dadkhahi, H., Shanmugam, K., Rios, J., Das, P., Hoffman, S.C., Loeffler, T.D., Sankaranarayanan, S.: Combinatorial black-box optimization with expert advice. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1918–1927 (2020)
13. Derbel, B., Pruvost, G., Liefoghe, A., Verel, S., Zhang, Q.: Walsh-based surrogate-assisted multi-objective combinatorial optimization: A fine-grained analysis for pseudo-boolean functions. *Applied Soft Computing* **136**, 110061 (2023)
14. Dushatskiy, A., Alderliesten, T., Bosman, P.A.: A novel approach to designing surrogate-assisted genetic algorithms by combining efficient learning of walsh coefficients and dependencies. *ACM Transactions on Evolutionary Learning and Optimization* **1**(2), 1–23 (2021)
15. Dushatskiy, A., Mendrik, A.M., Alderliesten, T., Bosman, P.A.: Convolutional neural network surrogate-assisted gomea. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 753–761 (2019)
16. Gosavi, A., et al.: Simulation-based optimization. Springer (2015)
17. Han, L., Wang, H.: A random forest assisted evolutionary algorithm using competitive neighborhood search for expensive constrained combinatorial optimization. *Memetic Computing* **13**, 19–30 (2021)

18. Hordijk, W., Stadler, P.F.: Amplitude spectra of fitness landscapes. *Advances in Complex Systems* **1**(01), 39–66 (1998)
19. Hörl, S., Balac, M.: Introducing the eqasim pipeline: From raw data to agent-based transport simulation. *Procedia Computer Science* **184**, 712–719 (2021), the 12th International Conference on Ambient Systems, Networks and Technologies (ANT) / The 4th International Conference on Emerging Data and Industry 4.0 (EDI40) / Affiliated Workshops
20. Hörl, S., Balac, M.: Synthetic population and travel demand for paris and ile-de-france based on open and publicly available data. *Transportation Research Part C: Emerging Technologies* **130**, 103291 (2021)
21. Horni, A., Nagel, K., Axhausen, K. (eds.): *Multi-Agent Transport Simulation MATSim*. Ubiquity Press, London (8 2016)
22. Ibeas, A., della Olio, L., Alonso, B., Sainz, O.: Optimizing bus stop spacing in urban areas. *Transportation Research Part E: Logistics and Transportation Review* **46**(3), 446–458 (2010)
23. Leprêtre, F., Fonlupt, C., Verel, S., Marion, V.: Combinatorial Surrogate-Assisted Optimization for Bus Stops Spacing Problem. In: *Biennial International Conference on Artificial Evolution (EA 2019)*. Mulhouse, France (Oct 2019)
24. Leprêtre, F., Fonlupt, C., Verel, S., Marion, V., Armas, R., Aguirre, H., Tanaka, K.: Fitness landscapes analysis and adaptive algorithms design for traffic lights optimization on sialac benchmark. *Applied Soft Computing* **85**, 105869 (2019)
25. Leprêtre, F., Verel, S., Fonlupt, C., Marion, V.: Walsh functions as surrogate model for pseudo-boolean optimization problems. In: *The Genetic and Evolutionary Computation Conference (GECCO 2019)*. pp. 303–311. *Proceedings of the Genetic and Evolutionary Computation Conference, ACM, Prague, Czech Republic (Jul 2019)*
26. Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H.: Feature selection: A data perspective. *ACM Computing Surveys (CSUR)* **50**(6), 94 (2018)
27. Moraglio, A., Kattan, A.: Geometric generalisation of surrogate model based optimisation to combinatorial spaces. In: *European conference on evolutionary computation in combinatorial optimization*. pp. 142–154. Springer (2011)
28. O’Donnell, R.: *Analysis of Boolean Functions*. Cambridge University Press (2014)
29. Oh, C., Tomczak, J., Gavves, E., Welling, M.: Combo: Combinatorial bayesian optimization using graph representations. In: *ICML Workshop on Learning and Reasoning with Graph-Structured Data (2019)*
30. Pires, F., Cachada, A., Barbosa, J., Moreira, A.P., Leitão, P.: Digital twin in industry 4.0: Technologies, applications and challenges. In: *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*. vol. 1, pp. 721–726. IEEE (2019)
31. Saka, A.A.: Model for determining optimum bus-stop spacing in urban areas. *Journal of Transportation Engineering* **127**(3), 195–199 (2001)
32. Saka, A.A.: Effect of bus-stop spacing on mobile emissions in urban areas (2003)
33. Saltiel, D., Benhamou, E.: Feature selection with optimal coordinate ascent (oca). arXiv preprint arXiv:1811.12064 (2018)
34. Stadler, P.F.: Landscapes and their correlation functions. *Journal of Mathematical chemistry* **20**(1), 1–45 (1996)
35. Stadler, P.F.: Spectral landscape theory. *Evolutionary dynamics: Exploring the interplay of selection, accident, neutrality, and function* pp. 221–272 (2003)
36. Sutton, A.M., Whitley, L.D., Howe, A.E.: Computing the moments of k-bounded pseudo-boolean functions over hamming spheres of arbitrary radius in polynomial time. *Theoretical Computer Science* **425**, 58–74 (2012)

37. Tinós, R., Przewozniczek, M.W., Whitley, D.: Iterated local search with perturbation based on variables interaction for pseudo-boolean optimization. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 296–304 (2022)
38. Verel, S., Derbel, B., Liefoghe, A., Aguirre, H., Tanaka, K.: A surrogate model based on Walsh decomposition for pseudo-boolean functions. In: PPSN 2018 - International Conference on Parallel Problem Solving from Nature. Lecture Notes in Computer Science, vol. 11102, pp. 181–193. Coimbra, Portugal (Sep 2018)
39. Whitley, L.D., Chicano, F., Goldman, B.W.: Gray box optimization for nk landscapes (nk landscapes and max-ksat). *Evolutionary Computation* **24**(3), 491–519 (2016)
40. Wright, A.H., Thompson, R.K., Zhang, J.: The computational complexity of nk fitness functions. *IEEE Transactions on Evolutionary Computation* **4**(4), 373–379 (2000)
41. Zaefferer, M., Horn, D.: A first analysis of kernels for kriging-based optimization in hierarchical search spaces. In: Auger, A., Fonseca, C.M., Lourenço, N., Machado, P., Paquete, L., Whitley, D. (eds.) *Parallel Problem Solving from Nature – PPSN XV*. pp. 399–410. Springer International Publishing, Cham (2018)
42. Zaefferer, M., Stork, J., Friese, M., Fischbach, A., Naujoks, B., Bartz-Beielstein, T.: Efficient global optimization for combinatorial problems. In: Proceedings of the 2014 annual conference on genetic and evolutionary computation. pp. 871–878 (2014)
43. Zheng, C., Zheng, S., Ma, G.: The bus station spacing optimization based on game theory. *Advances in Mechanical Engineering* **7**(2), 453979 (2015). <https://doi.org/10.1155/2014/453979>
44. Zhu, Z., Guo, X., Chen, H., Zeng, J., Wu, J.: Optimization of urban mini-bus stop spacing: A case study of shanghai (china). *Tehnicki Vjesnik* **24**, 949–955 (06 2017)