Distributed Optimization for Railway Track Maintenance Operations Planning

Muhammad Faris

Master of Science Thesis

Delft University of Technology

Delft Center for Systems and Control

Distributed Optimization for Railway Track Maintenance Operations Planning

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

Muhammad Faris

January 11, 2018

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

Abstract

An essential component of railway infrastructures is track ballast. As railway track is used frequently by passing rolling stocks, its performance degrades over time. At certain degradation levels, maintenance interventions must be carried out to improve the track performance so to meet technical and safety regulations. In this way, the risk of accident or derailment can be minimized and the railway interoperability is ensured. Furthermore, the responsibility of designing maintenance plan belongs to infrastructure managers. To help them, predictive strategies based on optimization can suggest the optimal schedule to maintain the track over a certain time period. In this way, track performance and maintenance costs can be explicitly optimized over the whole life cycle of the track.

However, a railway network typically consists of multiple track sections, each of them with different degradation level and parameters. Hence, the optimization of track maintenance can be considered as a large-scale problem which has a large number of decision variables. For such kind of problem, the conventional centralized optimization is very difficult or even not tractable to solve due to limitations on the computational time and resources. One way to overcome this issue is by applying the so-called distributed optimization scheme. In such approach, the original optimization problem is partitioned into multiple smaller, tractable subproblems. Therefore, the optimization is tractable and more preferred for real-life implementations.

This thesis develops distributed optimization approaches for track maintenance operations planning problem. Three different schemes are compared: Parallel Augmented Lagrangian Relaxation (PALR), Alternating Direction Method of Multipliers (ADMM), and Distributed Robust Safe But Knowledgeable (DRSBK). As these distributed approaches basically designed for convex problems, extension techniques to handle non-convex nature of the proposed optimization problem are implemented. Furthermore, some case studies are defined to evaluate the algorithms from both performance and numerical perspectives.

In simulations of small, medium, and large-scale instances, it is shown that in most cases, DRSBK can outperform the other distributed approaches, by providing the closest-to-optimum solution to the centralized optimization problem while having the shortest computation time.

Table of Contents

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Dr. Ir. Bart De Schutter for his guidance during the work of this thesis. Secondly, I would also like to thank Dr. Alfredo Núñez, Zhou Su, and Ali Jamshidi as my daily supervisors. Alfredo has been able to patiently supervise me and show me the path to finish this thesis. Our discussions have provided me with many insights and inspirations. Zhou and Ali have also taught me with some theories and practical knowledge in optimization and railway engineering. Their suggestions help me a lot while encountering problems.

I would like to acknowledge Lembaga Dana Pengelolaan Pendidikan (LPDP) for providing me the scholarship for my master study at Delft University of Technology. Also, I want to thank my friends and roommates in Delft for accompanying and supporting me during my master study. I owe you a lot.

I would like to express my gratitude towards my mother, father, and my brother who always support and pray for me from far away. Finally, I am very grateful for God to have "allowed" me to pursue my master study. This has been a tough yet wonderful journey.

Thank you.

Delft, University of Technology — Muhammad Faris
January 11, 2018

"Knowing is not enough; we must apply. Being willing is not enough; we must do"
— *Leonardo da Vinci*

Chapter 1

Introduction

1-1 Background

Nowadays, railway transportation across the world experiences a significant increase in the number of passengers and goods. Since operated first in the sixteenth century, trains have already played an important role in public transportation systems. Compared to the other means of transportation, trains or rolling stocks offer a series of advantages: high reliability, capability, and safety [1]. Nevertheless, the rising use of railway transportation comes along with various problems. On the one hand, the users demand for less expensive fares, which can lead to a lower revenue [2]. On the other hand, due to the increased usage of the infrastructure, maintenance costs increase accordingly [3]. These circumstances create challenges for the railway traffic operations, and also for the management of railway assets, including both development and maintenance.

Railway infrastructure consists of different assets, comprising railway tracks, electrical systems, signaling devices, switches, stations, and so forth. All assets are interconnected and work together, forming the entire railway system. Among those components, ballast is a vital component as it is used to support the track level and alignment at the designated positions [1]. In the ideal condition, ballast can properly bear the loads when trains are rolling over as such the rail can stay still. Due to regular usage of tracks, ballast nonetheless suffers quality degradation over time. The degradation further causes track misalignment, and so the so-called track performance deteriorates. If the degradation level exceeds certain thresholds, the railway operations safety could not be guaranteed anymore. In order to avoid such unexpected conditions, ballast must be maintained so that its performance could meet technical and safety criterion. Accordingly, there are a number of maintenance options for ballast: tamping, ballast cleaning, stoneblowing, partial and full renewal operations.

Current practice in maintenance operations is carried out with two different policies. Whenever there is an emergency, reactive maintenance policy is undertaken to immediately fix it. Due to its uncertain nature, this maintenance operation disrupts the railway operations. The other policy is called preventive strategy, where the maintenance interventions are scheduled

Figure 1-1: Railway maintenance (image by: strukton.com)

within certain interval periodically. One drawback of both strategies is that they have no prediction capability for estimating track condition [4]. Consequently, data of asset condition could not be fully exploited and so there is no guarantee that their operations will be efficient.

Apart from maintenance interventions, inspection and measurement activities are frequently conducted to collect asset data. Measurements are useful to design Key Performance Indicator (KPI) and asset degradation model. By utilizing KPI and degradation model, the evolution of degradation process can be estimated and this also leads to the concept of a third type of policy: condition-based predictive maintenance [5]. This strategy can suggest timely maintenance operations depending on the track condition. In this way, unnecessary intervention can be avoided. Recent developments in information technology also introduce a new strategy called cloud-based maintenance. Various measurements collected by either frequent measurement or deployed sensors can be stored and analyzed by artificial intelligence algorithms. This gives rise to big data and *internet of things* (IoT) frameworks to improve the role of decision support systems in railway maintenance, as studied in [6, 7]. Nowadays, big data will be an interesting topic to develop in railway industry.

There are several challenges in track maintenance operations planning nowadays. A railway track maintenance operation is expensive, from thirty thousands up to a hundred thousands euro per track kilometer per year in Europe [8]. Moreover, maintenance work typically involves some personnel along with maintenance machines as depicted in Figure 1-1. This kind of resources along with the allocated budget and closure times to maintain the railway assets are limited [3]. In order to tackle these challenges, railway infrastructure managers can consider solutions from a decision support system. Such a system basically can be developed by implementing optimization techniques into the predictive maintenance strategy. It is notable that the solutions must have an optimal balance between acceptable track performance and limited maintenance costs. Correspondingly, a number of researches in the maintenance optimization field have been studied in this thesis [4, 2, 9, 10, 11].

However, one problem that might arise from the nature of track maintenance is computational complexity. It is known that a typical railway network is very long, which can be up to hundred kilometers [3]. Furthermore, it consists of multiple track sections in which each of them can have different degradation level and dynamics. Therefore, the optimization of track maintenance operations planning is a large-scale problem. Unfortunately, most of the related studies in the past made use of centralized optimization structure, that theoretically is not tractable if the problem size goes up [12]. This can be an issue if, for instance, infrastructure managers require a quick solution in the decision-making process. Thus, the notion of deploying distributed optimization approaches for dealing with the large-scale track maintenance planning arises.

In this thesis, distributed optimization is developed and applied for the scheduling of maintenance operations, of a large-scale railway network. Three distributed optimization approaches are compared: Parallel Augmented Lagrangian Relaxation (PALR), Alternating Direction Method of Multipliers (ADMM), and Distributed Robust Safe But Knowledgeable (DRSBK). These approaches are analyzed, from both railway performance and numerical perspectives.

1-2 Problem statement

In this thesis, the following main research question is raised:

Is it possible to improve the computational performance while maintaining a good quality of the solutions for large-scale railway track maintenance operations planning problem solved by a distributed optimization approach?

In the literature, most of the optimization methods in railway maintenance operations planning used centralized optimization schemes, which are not tractable for large-scale problems. Thus, this thesis investigates the distributed optimization framework in order to improve the scalability, while reaching an acceptable solution for the optimization of track maintenance operations planning for large-scale railway network. Based on the main question, several sub-questions can be derived as follows:

1. *Which optimization problem can capture characteristics like degradation of track performance, maintenance interventions and closure time for large-scale maintenance operations in railway tracks?*

   The proposed problem should include only the most relevant characteristics of the maintenance operation planning. Chapter 3 discusses the problem formulation to answer this question.

2. *Which distributed optimization approaches can deal with the proposed optimization problem and reach feasible global solutions?*

   Techniques that can deal with the proposed optimization problem will be taken into account and analyzed. Furthermore, different distributed approaches will be tested. This question is addressed in Chapter 4.

3. *How is the performance of distributed approach in comparison with the centralized approaches in terms of the solution quality and computation time?*

   Benchmarks with a different number of track sections and degradation regimes will be defined. The comparison will be conducted for the distributed approaches and small-medium scale problem with the centralized approach. Moreover, Chapter 5 provides the analysis of study cases to answer this question.

1-3 Structure of the thesis

The remaining chapters of this report are explained as follows. Chapter 2 contains literature survey from related studies including maintenance optimization and distributed optimization. In Chapter 3, the optimization problem is formulated and explained. This includes the definition of the proposed objective function along with the degradation model and constraints. Next, Chapter 4 covers the theory and development of distributed approaches for the proposed optimization problems. Next, Chapter 5 discusses case studies to evaluate the optimization problem, along with a comparison between distributed approaches. Finally, Chapter 6 includes conclusions, remarks, and further research.

Chapter 2

Literature review

2-1 Overview

In this chapter, related literature including optimization in railway maintenance and distributed optimization, are studied and reviewed. This chapter consists of the following parts: Section 2-2 first discusses optimization approaches in the field of railway maintenance. Next, distributed optimization approaches for particularly MILP problems are explained in Section 2-3. Finally, this chapter is closed with conclusions in Section 2-4.

2-2 Optimization in railway maintenance

Railway infrastructure managers nowadays face a number of challenges when designing a maintenance schedule. To obtain an efficient and cost-effective plan, the use of decision support systems are necessary. Such support system usually relies on optimization methods. The goals and limitations in track maintenance can be mathematically modeled into an objective function along with a set of constraints. Moreover, it is common to have discrete decisions when deciding, for instance, whether maintenance needs to be performed or not and which type of maintenance to perform. Due to this nature, it is not surprising that scheduling tasks contain binary or integer decision variables. In such cases, the optimization problems can be formulated into either Mixed-Integer Linear Programming (MILP) or Mixed-Integer Quadratic Programming (MIQP).

One important intervention to correct the track alignment is tamping. There are a number of researches on the optimization of tamping operations. As a consequence, there are different objective functions developed for tamping. Vale et al. [9] proposed an objective function consisting of the total number of performed tamping over a planning horizon on the whole track. Each intervention is carried out at a time step k and at a track section i. The corresponding objective function is written as follows:

$$J_1(\bar{U}) = \sum_{i=1}^{N}\sum_{k=1}^{T} u_i(k) \qquad u_i(k) \in \{0,1\} \tag{2-1}$$

where

$$\bar{U} = \begin{bmatrix} u_1(1) & ... & u_1(T) & ... & u_N(1) & ... & u_N(T) \end{bmatrix}^T$$

Variable $u_i(k)$ is a binary variable to perform tamping $u_i(k) = 1$ or not $u_i(k) = 0$ in track section i at a time step k. N and T are number of track sections and planning or prediction horizon, respectively. Based on J_1, in [11] an objective function by multiplying the maintenance indicator variable with the defined cost for each work is proposed. In this way, the total costs can be straightforwardly seen. Moreover, the preparation cost for performing the tamping operation in the whole track is added to the objective function.

Other form of the binary optimization problem of tamping is suggested in [13, 8]. A decision support system is proposed to minimize the tamping total cost instead of the number of tamping operations. The cost function is the total tamping cost over the planning horizon or known as total tamping closure time, consisting of the following sub-costs: intervention $t_{i,1}(k)$, machine driving $t_3(k)$, and preparation / ramp down cost $t_{i,2}(k)$. Moreover, the weights c_1, c_2, c_3 are described as maintenance closure time cost for each phase. The expression of objective function is shown below:

$$\begin{aligned} &J_2(\bar{T_1}, \bar{T_2}, \bar{T_3}) = \\ &\sum_{k=1}^{T}(\sum_{i=1}^{N} c_1 t_{1,i}(k) + \sum_{i=1}^{N} c_2 t_{2,i}(k) + c_3 t_3(k))r(k) \qquad t_{1,i}(k), t_{2,i}(k), t_3(k) \in \{0,1\} \end{aligned} \tag{2-2}$$

where

$$\bar{T_1} = \begin{bmatrix} t_{1,1}(1) & ... & t_{1,1}(T) & ... & t_{1,N}(1) & ... & t_{1,N}(T) \end{bmatrix}^T$$
$$\bar{T_2} = \begin{bmatrix} t_{2,1}(1) & ... & t_{2,1}(T) & ... & t_{2,N}(1) & ... & t_{2,N}(T) \end{bmatrix}^T$$
$$\bar{T_3} = \begin{bmatrix} t_3(1) & ... & t_3(k) & ... & t_3(T) \end{bmatrix}^T$$

Variable $r(k)$ is discount rate. The objective function on the basis of a number of actions (J_1) is considered to be less effective to reduce the total tamping work, as it did not count the loss due to closure time of maintenance works. That is correct but only in reactive cases. In preventive and predictive strategies, the plan is designed long before the maintenance operations are carried out, such that the closure times would not disturb the railway operations. Furthermore, Famurewa et al. [10] utilized both types of cost function above (the total number of actions and total tamping closure time) and summed them up in the end to get the total maintenance costs. Nonetheless, this kind of time-based formulation is only suitable for reactive or short-term planning cases, where disruptions should be minimized. If the closure time for tamping is minimized in the preventive maintenance, the maintenance team could not maximize the handled track section given the maintenance slot time. Thus, in [14], the typical objective function J_2 is expressed as maximization instead of minimization problem.

Additionally, it is difficult to find the exact representation of closure times into money loss such that they can be added to the maintenance action cost.

In representations above, the asset performance is not yet included in the optimization. The study by Arasteh-khouy et al. [3] considers the performance variable, along with tamping intervention indicator cost, in their objective function. This variable describes the train capacity lost and reduced reliability due to the degradation of track performance. However, the variable representation in binary condition (switched by means of a limit) might not be optimal as the track performance can be affected differently by different types of maintenance actions. If the performance variable is continuous, as used in [14], the optimizer will have more freedom to decide the timing of maintenance. Another type of maintenance actions, such as renewal can also be included, so more options are available. In addition, the relation between both component performance and maintenance costs are not explained. This can be included by, for instance, adding a weighting term in the objective function.

Aside from single control action-based optimization, other studies take into account different maintenance actions for one certain asset Su et al. [4]. This research deals with the case of integer optimization above by employing the Model Predictive Control (MPC) methodology. One advantage of this method is that the flexibility offered during the maintenance operation to adapt to several different conditions, based on measurement data. Furthermore, the objective function of track performance and maintenance cost is explicitly included in the optimization problem. The objective function is calculated in each time step for all track sections as shown in the following expression:

$$J_3(k) = \sum_{i=1}^{N} \sum_{t=k}^{k+T} \bar{x}_i^T(t) P x_i(t) + \lambda Q \bar{V}_i(t) \tag{2-3}$$

where $\bar{x}_i^T(k)$ and $\bar{V}_i(k)$ stands for the track performance and aggregated integer-binary input for track section i at time step k. P and Q are positive definite weighting matrices. However, the closure time owing to maintenance actions is not considered yet. The final problem is formulated into Mixed-Integer Quadratic Programming (MIQP). If the variable is defined with positive-definite constraints, the problem can be formulated as a less complex MILP instead of MIQP problem. Likewise, the study did not evaluate the proposed model with the other previous optimization models. Thus the improvement in terms of solution quality cannot be ensured.

Beside tamping, optimization of renewal operations in railway is studied in [15, 16]. The objects are different types of components. Therefore, the integer decision variables can be the combination of various maintenance actions, depending on such factors: location, possibility of joint operation in the same maintenance time slot. In this way, the total closure time due to multiple renewal operations within planning horizons can be minimized. The final problem is also expressed in MILP and solved via heuristics. Furthermore, in [17], the use of degradation models enable to account for corrective maintenance. Event though conducting joint maintenance operation is promising, the application is not trivial, as each component has different degradation models. Moreover, various equipment is required to carry out multiple interventions at the same time, which is complex and might take more resources than individual maintenance.

A number of studies in the past have optimized large-scale maintenance scheduling. In [18], the rolling-horizon optimization is performed to minimize the maintenance time of large-scale drainage networks, through scheduling of preventive and corrective maintenance on a daily basis. Meanwhile, the research in [19] deals with the optimization of large-scale construction projects in railway tracks. The objective is to minimize the construction costs by, particularly clustering working time based on some scenarios. Even though these studies mentioned the large-scale nature of their problems, they still utilize the so-called centralized optimization approach. From the computational perspective, the centralized approach is unattractive. In large-scale problems, the number of decision variables in centralized approach is increased linearly, but the computational time to solve the problem can be increased exponentially [12, 20]. This implies that the maintenance plan for large-scale problem cannot be generated by the centralized optimization in acceptable time. At a certain point, the centralized approach is no longer tractable. In [19], the study considers an optimization result to be not optimal if it takes more than two hours to solve. Therefore, particular approaches in optimization are required when dealing with large-scale problem size.

Research carried out by Ferrario et al. [21] introduces the concept of distributed strategy, that is specifically designed to solve maintenance problems in large-scale areas. The approach exploits a software that enables the multiple registered users to give reviews about a maintenance case. In other words, a maintenance case can be distributed separately to be analyzed locally by each user. The main coordinator in the server then only needs to check the submitted reviews and conclude the main solution which is aggregated from the user's review. Nevertheless, the proposed method lack of optimization processes, meaning that the result cannot be ensured to be optimal with respect to a set of constraints.

The other approach to be evaluated is the hierarchical scheme. In [14, 22] a multi-level or hierarchical MPC is proposed to optimize squat defect maintenance scheduling. The proposed approach is divided into the high-level controller and low-level optimizer. The high-level MPC controller works to optimize the railway track quality condition while minimizing maintenance efforts cost, whereas the low-level optimizer uses clustering technique to generate an optimal grinding and replacement schedule, based on MPC suggestions, which also reduces track possession time. One prominent advantage offered by this approach is that a lower computational effort for larger systems compared to centralized scheme due to the separation of optimization problems. However, note that the hierarchical computational burden reduction in handling very large-scale system is not as effective as decentralized or distributed schemes [23]. Based on the nature of the problem, decentralized or distributed approaches can decompose the main objective function into multiple (dozens or even hundreds) subproblems, while hierarchical typically only has two or three stages. Meanwhile, the hierarchical approach is effective when the whole problem consists of different heterogeneous subsystems, for example having different time scales, type of actions or objectives. They can be divided into different modules accordingly.

One of the state-of-the-art approaches in maintenance optimization which merges the hierarchical and decentralized optimization schemes is presented by Verbert et al. [24]. The research particularly addresses a multi-component system which has to be tackled by different types of maintenance works and timing. This motivates the utilization of a two-stage hierarchical scheme with different objective functions. The lower component-level maximizes the balance between accuracy and timeliness, while upper system-level considers the economic and structural costs to explicitly show the trade-off between component quality and maintenance cost.

The main advantage of exploiting decentralized and hierarchical approach is computational tractability when optimizing large-scale maintenance problems. The approach can also decouple any degradation model. Consequently, minor and medium maintenance interventions, which are prominent in cost reduction, cannot be accounted. Only renewal operation for different components is accounted in this research. Additionally, this study did not explore on computational efforts and comparison with respect to the centralized optimization scheme. Besides, the couplings among components, which usually exist in interconnected components as in railway infrastructure, are not addressed explicitly.

2-3 Distributed optimization methodology

The main goal of this thesis is to deal with the computational issue from the maintenance operations for large-scale railway track networks. To achieve that goal, various distributed optimization methods developed since last decades can be evaluated. Techniques, such as decomposition and communication will also be discussed.

Lagrangian-based decomposition methods are one of the basic techniques for applying distributed optimization [25]. The two basic techniques are called primal and dual methods. In general, they are used to decompose the centralized optimization problem into multiple subproblems. Basically, an optimization problem that contains multiple subproblems, can be categorized as either problem with Decoupled Cost but Coupled Constraint (DCCC) or Coupled Cost but Decoupled Constraint (CCDC). The study in [26] shows via study cases, that it is possible to convert from one type to the other one, i.e., from DCCC to CCDC or the other way around. The primal method is basically more fitted to CCDC, while the dual method can conveniently handle the DCCC. Both decomposition methods are proven to handle convex optimization problems in various engineering fields. From [14, 10], it can be indicated that the coupling constraint can be either the closure time or a maximum number of treated track section, which fits the dual-based method. However, the dual technique has a major drawback, that is premature termination when dealing with non-convex problems [27]. This phenomenon can lead to suboptimal solutions or known as dual gap.

Thus, a decomposition methodology known as augmented-Lagrangian relaxation or method of multipliers rises to solve the premature termination. Augmented Lagrangian employs penalty terms of the couplings in its Lagrangian equation such that the premature termination of iteration process can be prevented [25]. However, the existence of quadratic penalty terms also implies that the Lagrange equation cannot be decomposed directly. Thus, to solve this problem, the research in [28] uses Alternating Direction Method of Multipliers (ADMM). Through alternating technique, such algorithm enables the decomposition of the augmented Lagrangian equation, while keeping strong convergence property of the method of multipliers. The concept behind this is the alternating update between its primal variables, followed by the dual update.

To facilitate the update of decision variables and avoid conflicting objectives from the subproblems, communication links are established between subproblems [26]. The communication is typically conducted in parallel way. Therefore, In [29, 30] serial communication scheme is introduced and compare it with the existing parallel scheme and ADMM. Both communication techniques are tested with augmented Lagrangian on a convex problem. The comparison of both, with ADMM as well, shows that they have their own strength and weakness, i.e., ADMM has the shortest computation time and serial augmented Lagrange has the least iterations.

It is notable that in the first place, Lagrangian-based decomposition methods are intended only for a convex problem. An issue arises when dealing with non-convex integer programming. As mentioned in [25], the dual function must be differentiable in order for primal iterates to converge toward the optimal solution. In other words, the function must be smooth, otherwise, the optimal point cannot be guaranteed. Unfortunately, mixed-integer problems are non-smooth in nature. In addition, the drawback of this Lagrangian-based method is that the feasibility of each iteration result cannot be guaranteed, hence this might be an additional issue when implementing dual decomposition, augmented Lagrangian or ADMM in non-convex problems.

Fortunately, previous studies have developed various extension methods for the Lagrangian-based approaches to deal with integer problems. They made use of different decomposition methods. First, the dual decomposition method is used in [31, 32, 33]. Due to the non-convex nature of binary or integer variables, those variables are relaxed in continuous way and afterward, the resulting optimal objective can be set as the lower bound of the integer problem. The resulting decision variables can also be used to warm start the next MILP optimization. Moreover, different modification to dual decomposition by adding quadratic terms of the respective decision variables to regularize the objective function is applied in [34, 25]. This leads to the concepts of augmented Lagrangian and ADMM.

The extension of ADMM algorithm to handle mixed-integer problem are also available [35, 36, 37, 38, 39, 40]. In this regard, ADMM can be seen as a heuristic method [28]. In other words, there is no guarantee that the standard ADMM algorithm can converge to a feasible optimal solution. Binary variables can make the couplings become non-smooth, which lead to a non-convex problem. Such binary variables hence require particular handling. In [37], a continuous relaxation technique of binary variables is used. The resulting solution is used as a bound and to warm start the optimization of binary variables in the next original binary optimization. Study in [35] adds auxiliary variables to substitute the coupling primal variables. It turns out that the coupling constraints did not contain binary variables, thus ADMM can converge with ease. In addition, all of the mentioned literature of dual-based and ADMM decomposition uses subgradient or its projected version as dual variables update technique. A heuristic extension to ADMM is developed in [36], called Tailored ADMM, basically has the same concept of continuous relaxation pre-process as in [37]. This method is also applied in integer programming to overcome non-convexity [41]. The continuous relaxation is again used in this paper as a mean to reduce the search space for the next integer programming. Moreover, any inequality constraint must be transformed into equality, in ADMM. However, this paper did not explicitly explore the feasibility of the solution with respect to the couplings among subproblems.

Apart from Lagrangian-based methods, a distributed algorithm called Distributed Robust Safe But Knowledgeable (DRSBK) [42] also has been applied to MILP problems with hard non-convex coupling constraints. This algorithm is utilized by applying coupling constraint tightening approach when solving one subproblem. In other words, the decisions from other variables are fixed so that the coupling constraints can be individually solved. Hence, the subproblems are solved sequentially and they can be coordinated according to the couplings.

In this way, the feasibility of the solutions can be guaranteed. The drawback, however, is that the solution might only be suboptimal. Moreover, it appears that the order of sequence might affect the optimal solution, even feasibility of the solution. To solve those issues, the cooperative version of DRSBK is also available [43]. The idea is that one subproblem involves other subproblems decisions in its individual objective function, treated as perturbations. In this way, it is expected to get closer to optimal or even global solution. However, clustering methods are required to avoid the exploding number of perturbation variables due to increasing number of subproblems. For instance, in the case of multi-vehicle coordination, only nearby vehicles are considered, which might lead to into suboptimal solutions. Apart from that, it is interesting to note that compared to the Lagrangian-based, these methods are not much used in distributed MILP or MIQP domains.

Other known decomposition methods that have been applied for MILP or MIQP as well are Benders decomposition [44], and Dantzig-Wolfe decomposition [38, 45]. Similar to the relation between primal and dual decompositions, Benders decomposition is preferred for coupled variables, while the Dantzig-Wolfe decomposition can handle constraint coupling. In [46], it is mentioned that this method guarantee feasibility in each of its iteration, unlike the Lagrangian-based methods. Even though in general Dantzig-Wolfe decomposition also has better convergence than Lagrangian-based methods, it can take longer processing time [38].

2-4 Conclusions

From the literature survey, it can be concluded that there are a number of studies that explicitly have handled large-scale railway maintenance. They also mentioned about the long computation time needed to solve their respective problems. Furthermore, various schemes such as decentralized and hierarchical optimization approaches for railway infrastructure maintenance have been developed. However, the distributed optimization for track maintenance operations planning has never been explicitly implemented and addressed, to the best of authors knowledge. Furthermore, computation issue arising from a large number of decision variables in maintenance optimization has not been thoroughly identified and analyzed yet.

Moreover, there are distributed optimization approaches that work based on Lagrangian theory and constraint tightening concept, which can be applied to either MILP or MIQP problems. In this thesis, PALR, ADMM and DRSBK approaches are used for distributed optimization and will be explained in Chapter 4.

Chapter 3

Optimization problem of track maintenance operations planning

In this chapter, the optimization problem for the planning of large-scale track maintenance operations is discussed. The proposed formulation is developed based on [4, 14], which are designated for the case of large-scale track maintenance operations. The formulations of the optimization problems, mixed-integer linear or quadratic problems (MILP/MIQP), have been only solved via centralized and hierarchical schemes.

This chapter begins with a brief discussion of the operations of the track maintenance and its characteristics in Section 3-1. Next, Section 3-2 explains the system description, which includes the definition of the dynamics, constraints, and objective function. It is followed by the formulation of the optimization problem in Section 3-3. Finally, Section 3-4 presents the conclusions of the works in this chapter.

3-1 Track maintenance operations

When it is installed for the first time, ballast has sharp-edged stones which form a foundation for rail. The form can be seen in Figure 3-1. As the tracks are regularly used by rolling stocks, the ballast condition deteriorates over time and the track alignment is affected. Ballast stones reach the end of the useful life when their shape is rounded. In such state, it would not be able to hold the track properly. Maintenance operations are therefore required to solve this issue and keep the track fully operational [8].

Before the degradation level reaches the safety limits, maintenance operations are normally undertaken. The responsibility of maintenance operations planning belongs to infrastructure managers. Except for the emergency cases, preventive and predictive maintenance in principle are carefully scheduled by the managers long before the operations are carried out. This is because they consider prominent knowledge, such as the inspection results, assessment of collected measurements, as well as economical matters [1]. The decisions or known as the

Figure 3-1: Ballasted track (image by: scienceabc.com)

maintenance schedule include the maintenance timing, location (track section), and type of interventions to be performed. This sort of decision-making process can be aided by an optimization technique, which takes into account necessary characteristics and limitation in track maintenance, e.g., maintenance costs, closure times, maximum degradation level thresholds and so forth [14]. In this way, inefficient and costly decisions can be avoided. The major factors in track maintenance can be mathematically formulated as an optimization problem, which includes an objective function, a set of constraints, and dynamics of track degradation. It is expected that the optimization problem could reflect and obtain the optimal balance between maintenance costs and acceptable component performance [4].

In addition, two major different ballast maintenance options are considered in this thesis as integer inputs to the system: tamping and renewal. Tamping is performed by inserting a pair of claws into the ballast and use vibration technique to squeeze the stones, which is illustrated in Figure 3-2. In this way, it can reset the track degradation level to some extent. Due to the effect of track degradation memory (such as roundness of the stones) and the accumulated offset from the previous maintenance actions, tamping would reach some point where it will not be effective anymore. This point in practice can be indicated by either high degradation level or the short time gap between two consecutive tamping operations. In such condition, renewal operation is the only remaining option to reset the degradation level and memory to be as-good-as-new. Additionally, the machine employed for renewal is depicted in Figure 3-3, where old ballast is replaced by new ballast. Note that, this activity is very expensive and requires a longer time of track closure. The minor maintenance options, such as ballast cleaning and stoneblowing are not studied in this thesis [1]. For the future research, additional track maintenance interventions can be easily incorporated into the dynamic model.

Practically, a single railway track consists of multiple track sections since it is typically very long. For instance in [14], the track studied from Eindhoven to Weert in The Netherlands is 25 km. Even in [10], their study case has 130 km, which is the track from Kiruna to Riksgransen in Sweden. If a single track section is defined as 200 m and each of it has different dynamics, such study cases are obviously having a large number of decision variables, designated for each track section. Therefore, the optimization process of track maintenance operations planning can be considered as a large-scale problem. The next section will explain the optimization problem used in this research.

Figure 3-2: Heavy tamping machine (image by: powertransmissionworld.com)

Figure 3-3: Heavy renewal machine (image by: plasseramerican.com)

3-2 Problem description

There are a number of characteristics and limitation in track maintenance operations that can be incorporated into the optimization problem. Some of them are explained below accordingly.

3-2-1 Degradation dynamics

The dynamic of ballast degradation represents the track performance level and it can be mathematically modeled in a state-space model. By using this model, the evolution of track degradation level and offset memory can be incorporated. Moreover, the use of discrete dynamics or exponential degradation model has more practical advantage over linear degradation model [11]. The deterministic linear discrete-time dynamic (or also called as exponential degradation model) with integer inputs for each track section is expressed as follows:

$$\begin{aligned} x_{1,i}(k+1) &= a_{1,i}x_{1,i}(k) + f_{1,i}(x_i(k), u_i(k)) \\ x_{2,i}(k+1) &= a_{2,i}x_{2,i}(k) + f_{2,i}(x_{i,2}(k), u_i(k)) \end{aligned} \tag{3-1}$$

where the state variables $x_{1,i}$ and $x_{2,i}$ are track degradation level which is represented by standard deviation of longitudinal level and track recorder of ballast degradation. Compactly, the state vector can be defined as $x_i(k) = \begin{bmatrix} x_{1,i}(k) & x_{2,i}(k) \end{bmatrix}$. Moreover, the state transition matrix is time-invariant that contains the degradation rates $a_{1,i}$ and $a_{2,i}$ for the degradation level and offset recorder. Index $i \in \mathbb{N}$ is the number of track sections, k is the intervention time step. It is worthy to point out that this dynamic equation is slightly different from that in [4], in the sense that, the degradation memory variable is set to have a multiplier constant $a_{2,i}$. This parameter enables the deterioration exponentially rather than linearly, according to [47]. Moreover, every time tamping is undertaken, it brings back the degradation level to $x_{2,i}(k+1)$, instead of $x_{2,i}(k)$. In this way, in the time step $(k+1)$, both variables will start from an exactly similar value.

The integer input u has three different options to be decided in each maintenance time step, as can be seen in Table 3-1. Each option also has different characteristic and cost. Likewise, the discontinuous functions have the following representation:

Table 3-1: Maintenance option as system input

Input $u_i(k)$	**Decision**
1	Doing nothing
2	Tamping
3	Renewal

$$f_i^1(x_i(k), u_i(k)) = \begin{cases} 0, & \text{if } u_i(k) = 1 \\ -a_{1,i}x_i^1(k) + a_{2,i}x_i^2(k) + \alpha, & \text{if } u_i(k) = 2 \\ -a_{1,i}x_i^1(k) + h_{\text{min}}, & \text{if } u_i(k) = 3 \end{cases}$$

and

$$f_i^2(x_i(k), u_i(k)) = \begin{cases} 0, & \text{if } u_i(k) = 1 \\ \alpha, & \text{if } u_i(k) = 2 \\ -a_{2,i}x_i^2(k) + h_{\text{min}}, & \text{if } u_i(k) = 3 \end{cases}$$

where α and h_{min} stand for offset constant and minimum track performance limit, respectively. The offset constant, which bears small constant value, comes from the tamping operation side effect. This means that the effectivity of the current operation is reduced compared to the previous operation, by the offset and its memory degradation.

Such piece-wise affine model (3-1) is basically non-linear. Therefore, it must be transformed into the so-called Mixed-Logical Dynamical (MLD) [48], which can describe a linear dynamics with mixed-ineteger variables. According to [4], three options of maintenance input above can be represented by two binary variables δ_1 and δ_2. The conversion table is given in Table 3-2.

Table 3-2: Conversion of system input and binary variables

$u_i(k)$	$\delta_{1,i}(k)$	$\delta_{2,i}(k)$
1	0	0
2	0	1
3	1	0

Additionally, the option to perform both tamping and renewal at the same time step is not logical, so it is eliminated by using the following constraint:

$$\delta_{1,i}(k) + \delta_{2,i}(k) \leq 1 \tag{3-2}$$

Furthermore, by taking into account the binary variables above, the non-linear model is then reformulated as follows:

$$\begin{aligned} x_{1,i}(k+1) &= a_{1,i}x_{1,i}(k) + \delta_{1,i}(k)(-a_{1,i}x_i^1(k) + h_{\min}) + \delta_{2,i}(k)(-a_{1,i}x_{1,i}(k) + a_{2,i}x_{2,i}(k) + \alpha) \\ &= a_{1,i}x_{1,i}(k) - \delta_{1,i}(k)a_{1,i}x_i^1(k) - \delta_{2,i}(k)a_{1,i}x_{1,i}(k) + \delta_{2,i}(k)a_{2,i}x_{2,i}(k) + \delta_{1,i}(k)h_{\min} \\ &+ \delta_{2,i}(k)\alpha \end{aligned}$$

$$\begin{aligned} x_{2,i}(k+1) &= a_{2,i}x_{2,i}(k) + \alpha\delta_{2,i}(k) + \delta_{1,i}(k)(-a_{2,i}x_{2,i}(k) + h_{\min}) \\ &= a_{2,i}x_{2,i}(k) - \delta_{1,i}(k)a_{2,i}x_{2,i}(k) + \delta_{1,i}(k)h_{\min} + \delta_{2,i}(k)\alpha \end{aligned}$$

Since the model is non-linear, the auxiliary variables are introduced to linearize the system, in the sense of MLD system. It can be written in the following matrix form:

$$\begin{bmatrix} x_{1,i}(k+1) \\ x_{2,i}(k+1) \end{bmatrix} = \begin{bmatrix} a_{1,i} & 0 \\ 0 & a_{2,i} \end{bmatrix} \begin{bmatrix} x_{1,i}(k) \\ x_{2,i}(k) \end{bmatrix} + \begin{bmatrix} h_{\min} & \alpha \\ h_{\min} & \alpha \end{bmatrix} \begin{bmatrix} \delta_{1,i}(k) \\ \delta_{2,i}(k) \end{bmatrix} + \begin{bmatrix} -a_{1,i} & -a_{1,i} & a_{2,i} & 0 \\ 0 & 0 & 0 & -a_{2,i} \end{bmatrix} \begin{bmatrix} z_{1,i}(k) \\ z_{2,i}(k) \\ z_{3,i}(k) \\ z_{4,i}(k) \end{bmatrix} \tag{3-3}$$

where the auxiliary variables are defined as:

$$\begin{aligned} z_{1,i}(k) &= \delta_{1,i}(k)x_{1,i}(k) \qquad & z_{2,i}(k) &= \delta_{2,i}(k)x_{1,i}(k) \\ z_{3,i}(k) &= \delta_{2,i}(k)x_{2,i}(k) \qquad & z_{4,i}(k) &= \delta_{1,i}(k)x_{2,i}(k) \end{aligned} \tag{3-4}$$

and the auxiliary vector can be defined as $z_i(k) = \begin{bmatrix} z_{1,i}(k) & z_{2,i}(k) & z_{3,i}(k) & z_{4,i}(k) \end{bmatrix}^T$. Additionally, the degradation dynamics can be included within a set of constraints.

3-2-2 System constraints

Alongside the degradation dynamics, a set of constraints are defined in the following subsections.

Initialization

The initial condition and degradation rate for each track section are defined according to various case studies. In simulations, they are taken from look-up tables. The degradation rate is assumed to be constant within the prediction horizon in every simulation. The corresponding constraint is:

$$\begin{aligned} x_{1,i}(0) &= x^1_{i,0} \\ x_{2,i}(0) &= x^2_{i,0} \\ a^j_i \lambda &\geq 0 \end{aligned} \qquad (3\text{-}5)$$

where $x^1_{i,0}$ and $x^2_{i,0}$ are the initial values from look-up tables for state variable degradation level and offset memory,respectively. Moreover, $j \in 1, 2$ indicates the state variable. λ is the trade-off parameter for the objective function.

Auxiliary constraints

Each auxiliary $z_{p,i}$ variables introduce the following four constraints [48]:

$$\begin{aligned} z_{p,i}(k) &\leq h_{\max}\delta_{l,i}(k) \\ z_{p,i}(k) &\geq h_{\min}\delta_{l,i}(k) \\ z_{p,i}(k) &\leq x_{j,i}(k) - h_{\min}(1 - \delta_{l,i}(k)) \\ z_{p,i}(k) &\geq x_{j,i}(k) - h_{\max}(1 - \delta_{l,i}(k)) \end{aligned} \qquad (3\text{-}6)$$

where $h_{\max}$ is the maximum degradation level constraint. $p \in \{1, 2, 3, 4\}$ and $l \in \{1, 2\}$ indicate the auxiliary and binary decision variables, respectively. In this way, there will be sixteen equations for each track section.

Early renewal prevention

Renewal operation in practice is allowed once the offset memory has been considered high. In long-term planning, it is much more costly to perform renewal when tamping operation is still effective. To prevent performing renewal at the early stage of degradation, this constraint is added:

$$x_{2,i}(k) - h_{\mathrm{r}} \geq (r_i - 1)h_{\max} \qquad (3\text{-}7)$$

$$r_i - \delta_{1,i}(k) \geq 0 \qquad (3\text{-}8)$$

where r_i is the binary indicator for the *if-then* condition. Both equations basically say that if the track memory degradation level is below the renewal threshold h_{r}, the renewal could not be carried on. When the level passes the threshold, the renewal decision variable $\delta_{1,i}(k)$ is no longer restricted to any value.

Degradation level thresholds

In any condition, the decision support system must be able to prevent the degradation curve exceeds the safety limit. Therefore, to keep the track degradation level in acceptable conditions, the constraints on degradation level limits are defined as follows:

$$h_{\min} \leq x_i(k) < h_{\max} \tag{3-9}$$

this constraint is also applied to the offset memory state.

Maximum number of maintenance interventions

From railway perspective, the maintenance budget is limited [2]. Therefore, the following constraints make sure that the number of interventions, both tamping, and renewal, over the prediction horizon is restricted by thresholds:

$$\sum_{k=1}^{T} \delta_{2,i}(k) \leq g_{\mathrm{t}} \tag{3-10}$$

$$\sum_{k=1}^{T} \delta_{1,i}(k) \leq g_{\mathrm{r}} \tag{3-11}$$

where g_{t} and g_{r} are maximum numbers of allowed tamping and renewal operations over the prediction horizon, respectively. Likewise, this constraint applies to each track section individually, hence it is an individual or non-coupled constraint.

Maximum closure time each time step

The maintenance time slot is less than six or seven hours. The preventive or predictive maintenance operation is only allowed during night time, at weekends [1]. This applies for both tamping and renewal, respectively. Based on [14, 10], this constraint can be written as follows:

$$\sum_{i=1}^{N_{\mathrm{t}}(k)} t_{\mathrm{t1}} \delta_{2,i}(k) + \sum_{j=1}^{N-N_{\mathrm{t}}(k)} t_{\mathrm{t2}} \delta_{2,j}(k) < t_{\max} \tag{3-12}$$

$$\sum_{i=1}^{N_{\mathrm{r}}(k)} t_{\mathrm{r1}} \delta_{1,i}(k) + \sum_{j=1}^{N-N_{\mathrm{r}}(k)} t_{\mathrm{r2}} \delta_{1,j}(k) < t_{\max} \tag{3-13}$$

where t_{t1} and t_{t2} are maintenance operation and traveling times for tamping, respectively. The same representations also hold for t_{r1} and t_{r2} for renewal. $N_{\mathrm{t}}(k)$ and $N_{\mathrm{r}}(k)$ is the total number of track sections that receive tamping and renewal, respectively, at time step k. The maintenance time already includes the intervention and machine switching time. The renewal closure time constraint is also expressed in the same way, with different time value. Moreover, it is assumed that the machines move in one direction in each time step, from a starting point toward an endpoint at the other end of the track. Hence, the position of the maintained track section does not matter in the operation time. In addition, tamping and renewal are performed at different time slots.

3-2-3 Objective function

The objective function used in this thesis is defined based on [4], but with linear expression. This model is inherently different from the models given in [13, 2, 10, 11, 9], in the sense that the track performance term is included in this thesis. This theoretically enables the optimizer to look for the optimal balance between track performance and maintenance costs. Furthermore, the optimal state variables and decision variables for all track sections over the prediction horizon, $\bar{X}$ and $\bar{V}$ can be obtained by solving the following open-loop objective function:

$$J(\bar{X}, \bar{V}) = \sum_{i=1}^{N} \sum_{t=k}^{k+T} Q x_i(t) + \lambda R V_i(t) \tag{3-14}$$

where

$$\bar{X} = \begin{bmatrix} x_1(k) & ... & x_1(k+T) & ... & x_N(k) & ... & x_N(k+T) \end{bmatrix}^T$$
$$\bar{V} = \begin{bmatrix} V_1(k) & ... & V_1(k+T) & ... & V_N(k) & ... & V_N(k+T) \end{bmatrix}^T$$

The control action for track section i at time step k is $V_i(k) = \begin{bmatrix} \delta_i(k) & z_i(k) \end{bmatrix}^T$.$Q$ and R are matrices with only positive entries and appropriate dimension. This equation holds for all sections. Note that, due to the fact that renewal is more costly, it has a higher value in matrix R than tamping operation.

Likewise, the objective function must be converted in more compact form to be implemented [4]. Based on [49], the state variables $x_i(k)$ can be represented by the following data equation:

$$x_i(k) = A_i^k x_0 + \sum_{n=1}^{k-1} A^{k-n-1} B V_i(n)$$

and if the equations are expressed for the state variables of the entire prediction horizon, this would end up in the following augmented matrix form:

$$\tilde{X}_i = M_i x_i(0) + O_i \tilde{V}_i$$

where $\tilde{X}_i$ and $\tilde{V}_i$ stand for state and decision variables for track section i over the prediction horizon or

$$\tilde{X}_i = \begin{bmatrix} x_i(k) & ... & x_i(k+t) & ... & x_i(k+T) \end{bmatrix}^T$$
$$\tilde{V}_i = \begin{bmatrix} V_i(k) & ... & V_i(k+t) & ... & V_i(k+T) \end{bmatrix}^T$$

Moreover, M_i and O_i are parameter matrices for initial values and decision variables for track section i, respectively. By substituting $\tilde{X}_i$, the objective function can be reformulated as follows:

$$J(\bar{V}) = \sum_{i=1}^{N} (QO_i(k) + \lambda R)\tilde{V}_i + M_i x_i(0) = \sum_{i=1}^{N} C_i \tilde{V}_i \tag{3-15}$$

Moreover, it is notable that linear objective function is chosen in this research since the state variables $x_i(k)$ and $V_i(k)$ are vectors with only positive entries.

3-3 Problem formulation

Having expressed the system dynamics, constraints, and objective function, the optimization problem can be written in the following compact form:

$$\underset{\bar{V}}{\text{minimize}} \quad J(\bar{V}) = \sum_{i=1}^{N} J_{\text{ind}}(\tilde{V}_i) \tag{3-16}$$

$$\begin{aligned} \text{subject to} \qquad & E\bar{V} \leq g_{\text{ind}} \\ & \sum_{i=1}^{N} F_i \tilde{V}_i \leq g_{\text{coup}} \end{aligned} \tag{3-17}$$

where on one hand, E and g_{ind} are the parameter matrix and right-hand-side vector associated with all individual or non-coupled constraints, respectively. On the other hand, F_i and g_{coup} are parameter matrix and right-hand side vector associated with coupling constraints. This separation aims at preparing for the development of distributed approaches in the next chapter. However, in the centralized optimization, both constraints can be aggregated.

The existence of the binary variables in the degradation dynamics causes the proposed problem to be formulated into Mixed-Integer Linear Programming (MILP). In general, such problem can also be categorized as an NP-hard problem [50], which will increase the computational complexity exponentially with the higher number of decision variables. Hence, it is difficult to solve and requires longer computation time.

3-4 Conclusions

In this chapter, an optimization problem for track maintenance operations planning is formulated. It can be formulated as a Mixed-Integer Linear Programming (MILP) problem. The degradation dynamic is defined for each track section, thus each of them can have different initial conditions and degradation rates. As the problem is NP-hard, it might influence the complexity of the problem if the number of track sections increases. Such a problem having a large number of decision problem can be categorized as a large-scale problem.

Furthermore, the problem also includes some constraints that can be grouped into individual/non-coupling and coupling constraint. Individual constraints are those related to each track section, such as degradation dynamic, early renewal prevention, degradation limits, and maximum number of interventions. Coupling constraints are, on the other hand, those who affect multiple track sections. In this thesis, they correspond to tamping and renewal closure times. Moreover, it can be observed that the objective function is decoupled. Besides, a set of coupling and individual constraints are included in the problem. Hence, the problem can be defined as a Decoupled Cost but Coupled Constraint (DCCC) problem [26].

Chapter 4

Distributed optimization

4-1 Overview

The main contribution of this thesis is solving the optimization problem of track maintenance operations planning using distributed approaches. A number of advantages are offered: low computational cost and problem scalability for large-scale instances [26]. This chapter hence discusses the development of distributed optimization approaches.

Three distributed optimization approaches are discussed: Parallel Augmented Lagrangian Relaxation (PALR), Alternating Direction Method of Multipliers (ADMM), and Distributed Robust Safe But Knowledgeable (DRSBK). The first two are iterative methods based on Lagrangian duality theory, which are usually used to solve convex smooth problems [25, 28]. It is worthy to mention that the proposed optimization problem is a MILP. The basic form of those decomposition methods could not guarantee the zero duality gap for such problem and the global feasibility. Thus, extension methods to solve this issue will also be discussed. Moreover, DRSBK is a non-iterative method which is originally designed to deal with MILP problems by exploiting constraint tightening techniques [43].

The remaining of this chapter is organized as follows: Section 4-2 will firstly address coupling constraints in the proposed optimization problem. Next, the Lagrangian-based (PALR and ADMM) and constraint tightening approach DRSBK algorithm will be extensively discussed in Section 4-3 and 4-4, respectively. Finally, Section 4-5 provides the conclusions from this chapter.

4-2 Identification of problem couplings

A distributed optimization approach is particularly required to handle the couplings between subproblems. Therefore, coupling constraints in the proposed optimization problem are first identified. In general, the coupling can exist either in the form of a variable in the objective function or in the constraint [26]. Thus, a smaller-size subproblem which is partitioned from the large original main problem would inherit the couplings. A coupled subproblem usually has uncoupled constraints and vice versa. In this research, a subproblem is defined to be an optimization problem of one single track section, which has its own individual objective function and a set of individual constraints.

In the proposed optimization problem from Chapter 3, the centralized objective function comprises the sum of the objective functions of all subproblems. In the equation, there is no common variable that is shared between subproblems. In other words, there is no explicit coupling variable and thus the main objective function can be conveniently split into N sub-objective functions. Furthermore, the track degradation dynamic, maximum and minimum degradation level thresholds, the level threshold for renewal, auxiliary variables constraints, and a maximum number of efforts in the set of constraints are already designated for every track section. Since those constraints have no direct effect on the other subproblems, they can be grouped into individual constraints.

Moreover, the other constraints, namely maximum closure time for tamping and renewal operations (3-13), can be considered as coupling constraints, since they affect across multiple subproblems. Thus, such constraints basically can be reformulated as follows:

$$\sum_{i=1}^{N} F_i^t \tilde{V}_i < t_{\max}$$

where F_i^t is the parameter matrix of tamping coupling constraint for each subproblem. Alongside this, the renewal parameter matrix is F_i^r defined in the same way. Coupling constraints typically establish dependencies between subproblems, which can complicate the partition of the original centralized problem. Therefore, distributed optimization techniques are required to decouple them. Having identified the couplings, the main goal is to find and implement distributed approaches that do not only split the centralized problem but also enable the coordinator to communicate the coupling constraints among subproblems such that the global optimal point can be reached or at least find a suboptimal solution close to it.

Before coming to the main discussions, some assumptions are made in this thesis. First, due to the slow nature of the track degradation dynamics as well as the scheduling of railway maintenance operations, simulations are conducted offline. Moreover, it is noteworthy that since in practice the optimization is processed in a single computer, it is clear that the full communication between subproblems is guaranteed. This implies that decisions from all subproblems are completely shared. In such setting, a coordinator is typically required to aggregate the decisions from all suproblems and use it to update the price or dual variable. This leads to the development of a coordinator-based distributed optimization. The corresponding scheme of the algorithm is depicted in Figure 4-1.

4-3 Lagrangian-based decomposition methods

This section addresses the implementation of two different Lagrangian-based decomposition methods applied to handle the proposed optimization problem, namely Parallel Augmented Lagrangian Relaxation (PALR) and Alternating Direction Method of Multipliers (ADMM). As the problem has coupling constraints, it is more suitable to use methods that work based

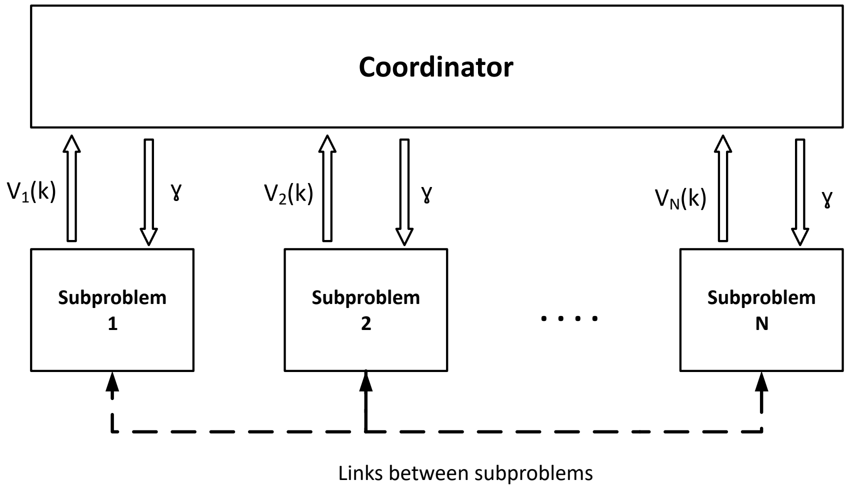

Figure 4-1: Schematic of distributed optimization with a central coordinator

on dual decomposition. These methods exploit pricing strategy to modify the augmented Lagrangian of subproblems so that coupling constraints can be incorporated in the individual computation. In other words, the corresponding dual function is decomposed instead of the original problem.

4-3-1 Parallel augmented lagrangian relaxation

This method hails from the dual decomposition and is called augmented Lagrangian. This method basically adds a quadratic term consisting of coupling constraints into the augmented Lagrangian. This term allows for solving the convergence issue in dual decomposition [27]. Thus, a combination of auxiliary problem principle and block coordinate descent are used in this method. The combination of both methods approximates the non-separable quadratic terms by linearizing the quadratic term and adding an individual separable term [51]. Doing so enables the solving of subproblems in parallel, which then leads to the concept of Parallel Augmented Lagrangian Relaxation (PALR) [30]. In this way, all subproblems can practically be solved at the same time.

In order to cope with the requirement for implementing this approach, the centralized problem in equation (3-16) has to be rewritten according to the Lagrangian dual theory [25]. This can be done by removing the couplings from the constraint set and putting them into the objective function in such a way that the augmented Lagrangian form is constructed. Moreover, another requirement in order to use augmented Lagrangian-like method is that any inequality coupling constraint in the proposed problem must be converted into equality form [41]. Thus, a slack variable is added to each row of coupling constraint, the number of which in total equals to the number of time step over the prediction horizon. Thus, $s_t(k)$ and $s_r(k)$ are defined as slack variables for tamping and renewal works for each decision time step, respectively. Variables $\bar{S}_t$ and $\bar{S}_r$ are also defined as tamping and renewal slack variables over the prediction horizon. Next, the augmented Lagrangian formulation of the centralized objective function (3-15) can be written as follows:

$$
\begin{aligned}
L_{\text{PALR}}(\bar{V}, \bar{S}_t, \bar{S}_r, \gamma^t, \gamma^r) = \sum_{i=1}^{N} J_{\text{ind}}(\tilde{V}_i) + \sum_{k=1}^{T} J_t(s_t(k)) + \sum_{k=1}^{T} J_r(s_r(k)) + \\
\gamma^t(\sum_{i=1}^{N} F_i^t \tilde{V}_i + \sum_{k=1}^{T} F_{st} s_t(k) - t_{\max}) + \gamma^r(\sum_{i=1}^{N} F_i^r \tilde{V}_i + \sum_{k=1}^{T} F_{sr} s_r(k) - t_{\max}) + \\
\frac{\rho}{2} || \sum_{i=1}^{N} F_i^t \tilde{V}_i + \sum_{k=1}^{T} F_{st} s_t(k) - t_{\max}||_2^2 + \frac{\rho}{2} || \sum_{i=1}^{N} F_i^r \tilde{V}_i + \sum_{k=1}^{T} F_{sr} s_r(k) - t_{\max}||_2^2
\end{aligned} \tag{4-1}
$$

$$
\text{subject to} \quad E\bar{V} \leq g_{\text{ind}} \tag{4-2}
$$

where γ^t and γ^r are Lagrange multiplier or dual variables defined for each row of tamping and renewal closure time couplings, respectively. Beside, F_{st} and F_{sr} are the parameter matrices associated with slack variables. These matrices theoretically contain resources to cover the unused allocation from the original inequality condition. Furthermore, the objective functions of the slack variables $J_t(s_t(k))$ and $J_r(s_r(k))$ are defined to be linear problem with the multiplier of 1. Adding such slack variables theoretically will not change the original centralized objective function of $\bar{V}$. Note that, the constraint expression in equation (4-2) serves as the remaining non-coupling constraints, which can be solved individually by each respective subproblem. The Lagrangian equation for the dual problem or known as the dual function can hence be written as:

$$
\begin{aligned}
& q(\gamma^t, \gamma^r) = \\
& \inf_{\bar{V}, \bar{S}} \left(L_{\text{PALR}}(\bar{V}, \bar{S}_t, \bar{S}_r, \gamma^t, \gamma^r) | \sum_{i=1}^{N} F_i^t \tilde{V}_i + \sum_{k=1}^{T} F_{st} s_t(k) - t_{\max}, F_i^r \tilde{V}_i + \sum_{k=1}^{T} F_{sr} s_r(k) - t_{\max} \right)
\end{aligned} \tag{4-3}
$$

To make equations (4-1) and (4-3) more compact, the tamping and renewal coupling constraints can be combined into one single parameter matrix F_i since their expression of couplings are similar. Thus, let $\gamma = [\gamma^t \quad \gamma^r]$ and $F_i = [F_i^t \quad F_i^r]^T$. The objective functions and the parameter matrices of slack variables associated with tamping and renewal constraints are combined into one single problem $J_t(\tilde{S})$ and $F_s = [F_{st} \quad F_{sr}]^T$ as well. This implies that $s(k) = [s_t(k) \quad s_r(k)]^T$ and $\bar{S}$ is slack vector over the prediction horizon. Since the value of $t_{\max}$ are similar for all rows of coupling constraints, it can be icluded within the column vector of g_{coup}. Meanwhile, as denoted by Negenborn et al. [23], the dual problem can be interpreted as a maximization problem of dual variables:

$$
\begin{aligned}
\underset{\gamma}{\text{maximize}} \quad & q(\gamma) \\
\text{subject to} \quad & \gamma \geq 0
\end{aligned}
$$

.

In each iteration, each subproblem runs in parallel, meaning that this problem uses the results from the last iterations. Once all subproblems have been solved, the results are collected by a coordinator to be included in the update of the dual variable. The existence of the coordinator also implies that one dual variable is used to determine the common price for all subproblems. Each iteration consists of the following steps:

1. Calculating the optimal value of subproblem, along with primal decision variables V_i

$$\tilde{V}_i(j+1) = \arg\min \quad L_{\text{PALR}}(\tilde{V}_i(j), \sum_{n\neq i}^{N-1} \tilde{V}_n(j), \bar{S}(j), \gamma(j))$$

2. Calculating the optimal value of each slack variable

$$s(k)(j+1) = \arg\min \quad L_{\text{PALR}}(\bar{V}(j), s(k)(j), \sum_{m\neq i}^{T-1} s(m)(j), \gamma(j))$$

3. Updating the residual vector P_{res}

$$P_{\text{res}} = \sum_{i=1}^{N} F_i \tilde{V}_i(j+1) + \sum_{k=1}^{T} F_s s(k)(j+1) - g_{\text{coup}}$$

4. Updating dual variables γ

$$\gamma(j+1) = \gamma(j) + \alpha P_{\text{res}}$$

where j is the iteration and the parameter matrix for tamping and renewal closure times constraint $F_i = [F_i^t \quad F_i^t]^T$. The iteration stops whenever the feasible condition is fulfilled or the maximum number of iteration is reached, as explained later on how to handle the proposed MILP problem.

4-3-2 Alternating direction method of multipliers

Basically, ADMM shares the similar augmented Lagrangian equation than PALR. The difference lies in the way of dealing with the quadratic term. Instead of linearizing it, ADMM uses the so-called alternating technique. This technique enables the separation of the quadratic terms to be determined individually by fixing the decisions coming from the other subproblems. This also implies that the algorithm runs in sequence instead of parallel. In this way, ADMM can exploit the latest decisions from the other subproblems. Moreover, it is shown in [30] that ADMM can outperform PALR in terms of number of iterations. The corresponding Lagrangian and constraint expressions are basically similar with PALR, written as follows:

$$\begin{aligned} L_{\text{ADMM}}(\bar{V}, \bar{S}, \gamma) = &\sum_{i=1}^{N} J_{\text{ind}}(\tilde{V}_i) + \sum_{k=1}^{T} J(\tilde{S}) + \gamma(\sum_{i=1}^{N} F_i \tilde{V}_i + \sum_{k=1}^{T} F_s s(k) - g_{\text{coup}}) + \\ &\frac{\rho}{2} || \sum_{i=1}^{N} F_i \tilde{V}_i + \sum_{k=1}^{T} F_s s(k) - g_{\text{coup}} ||_2^2 \end{aligned} \tag{4-4}$$

$$\text{subject to} \quad \text{equation } (4-2) \tag{4-5}$$

where F_s stands for the parameter matrix for the slack variables. The remaining parts, including the iteration, have the same structure than its PALR counterpart. Practically, in ADMM the inner terms of decision variables will have no place in the linear equation of objective function, unlike the PALR. The unscaled form of ADMM, as provided in [28] is chosen to be implemented. Furthermore, the ADMM algorithm comprises the following iterations:

1. Calculating the optimal value of subproblem, along with primal decision variables V_i

$$\tilde{V_i}(j+1) = \text{arg min} \quad L_{\text{ADMM}}(\tilde{V_i}(j), \sum_{n \neq i}^{N-1} \tilde{V_n}(j+1), \bar{S}(j), \gamma(j))$$

2. Calculating the optimal value of each slack variable

$$s(k)(j+1) = \text{arg min} \quad L_{\text{ADMM}}(\bar{V}(j+1), s(k)(j), \sum_{m \neq i}^{T-1} s(m)(j+1), \gamma(j))$$

3. Updating the residual vector P_{res}

$$P_{\text{res}} = \sum_{i=1}^{N} F_i \tilde{V_i}(j+1) + \sum_{k=1}^{T} F_s s(k)(j+1) - g_{\text{coup}}$$

4. Updating dual variables γ

$$\gamma(j+1) = \gamma(j) + \alpha P_{\text{res}}$$

4-3-3 Extension for the lagrangian-based methods

To deal with the proposed MILP problem, some modifications to the original PALR and ADMM are required. Since the problem has non-convex non-smooth coupling constraint, this causes the subgradient dual update might be unable to converge. One way to solve this problem is by applying continuous relaxation of the binary decision variables, such that the MILP becomes a less complex linear programming problem [41, 37]. The main goal is to preserve the convergence properties of the dual variables. On top of that, the generated objective function from solving LP can be used as a lower bound for the next MILP optimization and the decision variables could be the warm start vector as well. In other words, each decomposition runs two times, for linear programming relaxation and original MILP.

4-3-4 Stopping criterion

As the optimization problem is MILP, the convergence of primal residuals cannot be guaranteed [28]. In [1], it is shown that practically the maintenance intervention for one single track section is not always performed in each month. Thus, there is no way the residuals can go to zero in this regard. Therefore, rather than observing the convergence, the residuals are only checked whether they have been reaching all negative values. Negative values for all rows of residuals vector means that the problem is already feasible from input perspective. In other words, an input feasibility checker is added in within the iteration [41].

Additionally, the best objective value in each iteration is also checked as well, after it is guaranteed that it is feasible. A simple terminating technique is implemented. The complete algorithm containing the extension technique and stopping criterion is shown in Flowchart 4-2.

4-3-5 Input feasibility compensator

In this thesis, the maximum number of performed tamping and renewal is adapted such that one can see how the distributed approaches comply with the couplings. However, if the limitation is very strict, these Lagrangian-based approaches will very likely fail to follow the coupling, resulting in an infeasible result from the input perspective. In practice, this can happen when there is not enough machine or personnel to treat the severed track sections at one time step. In this regard, they only handle the most severe track sections and leave the rest as it is. As a result, the degradation level of untreated track sections can go above the limit and enters the high-risk region. However this condition practically can be fixed by e.g., manual spot tamping machine or renewal in the next control horizon.

In order to accommodate such condition, a heuristic compensator is designed. It is included right after decisions $\bar{V}$ coming from the optimizer, before entering the system. When the block detects any input is violating the time limit at one time step, it limits the allowed number of interventions according to the threshold. On one hand, the violation of the coupling is prevented. On the other hand, one or more track sections can have degradation level above the limit. For the next optimization cycle, the initialization of degradation level taken from current cycle must below the limit, otherwise, the optimization is not feasible. Thus, additional tamping or renewal is undertaken to improve the track degradation level and/or offset memory. The resulting state variables thus can be used for the optimization.

4-4 Constraint tightening approaches

4-4-1 Distributed robust safe but knowledgeable

The last algorithm implemented in this thesis is Distributed Robust Safe But Knowledgeable (DRSBK) approach that was originally used in [42, 43] to handle the distributed optimization with hard non-convex coupling constraints. Unlike its Lagrangian-based counterparts, this algorithm is originally devoted to the MILP problem. The idea is as follows: instead of putting the coupling constraints into the Lagrangian form objective function, this algorithm applies tightening resource allocation in the coupling constraints for each subproblem computation. This can be illustrated by the following objective function for each subproblem:

$$\underset{\tilde{V}_i}{\text{minimize}} \quad J_{\text{ind}}(\tilde{V}_i) \tag{4-6}$$

$$\begin{aligned} \text{subject to} \quad & \text{equation } (4-2) \\ & F_i\tilde{V}_i < g_{\text{coup}} - \sum_{j=1}^{N-1} F_j\tilde{V}_j \quad \forall j \neq i \end{aligned} \tag{4-7}$$

the second set of constraints in equation (4-7) is the coupling in which the total resources or g_{coup} have been reduced by the other previous subproblems interventions. This can be done by freezing the decisions from the others. In this way, this coupling can be reformulated into the non-coupled ones. The computation can then be solved individually by each subproblem in the sequential and non-iterative way. One advantage of assigning the couplings into individual constraints is that the solution is guaranteed to be feasible from the input perspective.

DRSBK is developed based on receding horizon control or MPC with coupling constraints [42]. The decisions for all subproblems are calculated for the entire prediction horizon. Therefore, the original approach can be directly applied to the optimization problem of maintenance planning over the entire prediction horizon. Likewise, unlike the coordinator in Lagrangian-based methods, the job of coordinator in DRSBK is only checking the feasibility of the generated solution. The decision from one subproblem is communicated to the other remaining subproblems. This enables the calculation of the remaining allocation individually, and thus the coordinator does not have to aggregate them.

4-4-2 Random sequence generator and stopping criterion

It is mentioned in Chapter 2 that the basic version of DRSBK might be stuck within suboptimal or even not feasible solutions from output perspective. Hence, the algorithm is modified such that the sequence of subproblems to be processed in each iteration is generated randomly. If the output from the solver indicates that the result from an iteration is not feasible, the sequence is generated randomly, which is very likely to be different from the previous sequence. The feasibility checking technique is different from the Lagrangian-based algorithms, in the sense that it sums up the total feasible solutions given by all subproblems. The result is feasible if the total solved subproblems are the same with N. Meanwhile, the stopping criterion is designed such that if in two consecutive cycles the difference of objective value is below some optimality threshold and the result from the previous iteration is feasible, the iterations are stopped. In addition, the typical feasibility compensator block similar to the one applied in Lagrangian-based methods is also used. The complete algorithm can be seen in Algorithm 4-3.

4-5 Conclusions

This chapter describes the implemented distributed optimization methods for the proposed problem. First, the coupling constraints are identified as maximum maintenance closure

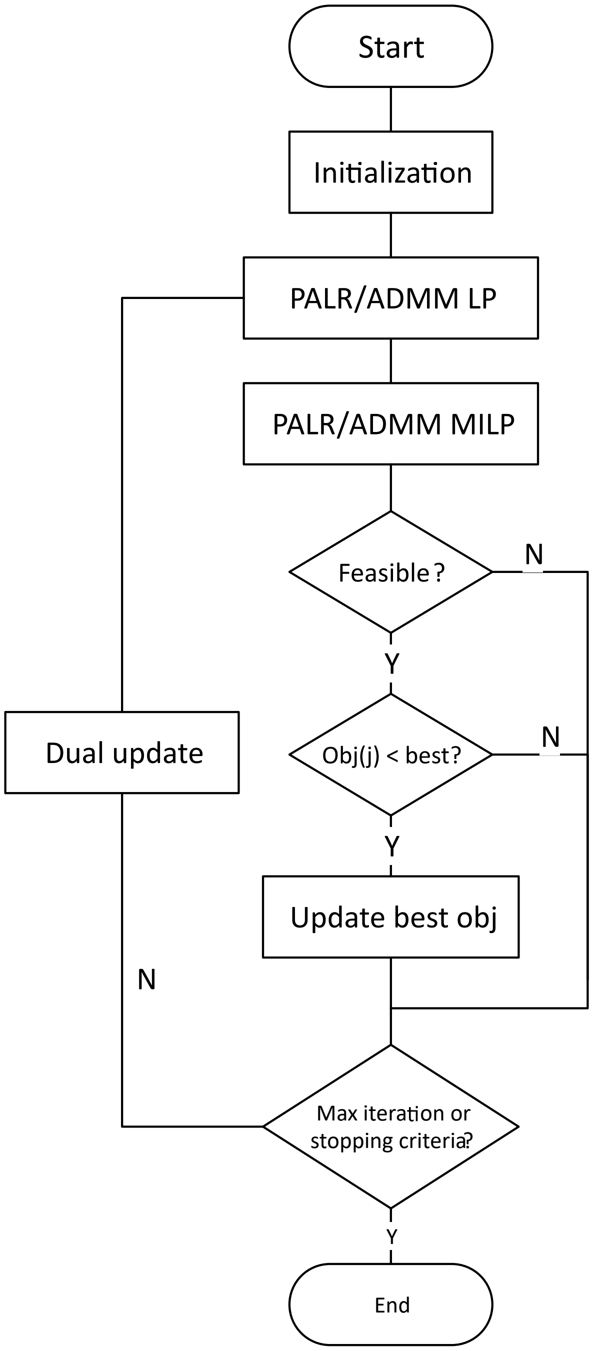

Figure 4-2: Flowchart of Lagrangian-based approaches(PALR and ADMM)

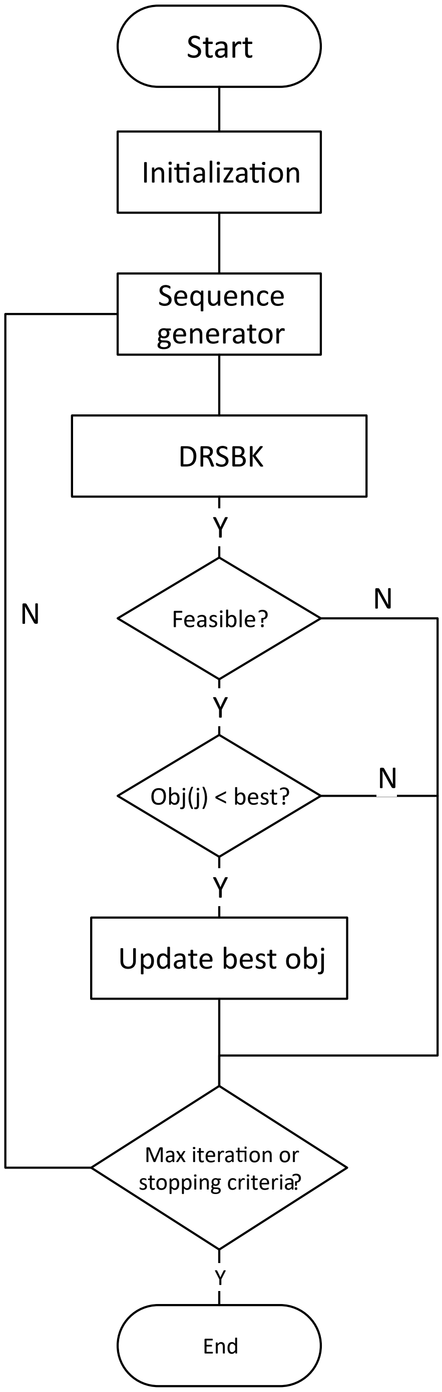

Figure 4-3: Flowchart of DRSBK

times due to tamping and renewal respectively. Since the maintenance options are integer, the coupling constraints are inherently non-smooth. This issue complicates the approaches when trying to retrieve feasible results.

The coordinated distributed optimization schemes are explained in this chapter. PALR and ADMM are decomposition methods which are based on Lagrangian dual theory. In such approaches, the augmented Lagrangian of the objective function is modified, such that it can accommodate the coupling constraints. Hence, it is also known as pricing strategy. The third algorithm, DRSBK, works based on constraint tightening, which modifies the resources of the coupling constraints and solves subproblems sequentially. A coordinator is also employed to update the dual variables (PALR and ADMM) or update the sequence (DRSBK).

In general, all algorithms require extension techniques to retrieve feasible and solutions close to the global optimal. In Lagrangian-based algorithms, the extension technique basically consists of two step methods where the first uses the continuous relaxation of binary variables to provide the warm start variables and objective function bounds for the second MILP step. The input-feasibility checking module examines whether the generated decisions over the prediction horizon violate the coupling constraints. Furthermore, a stopping criterion is designed, so that input-feasible and suboptimal solutions can be retrieved during the iterations, within reasonable computation time.

The extension technique in DRSBK is applied to check whether the solution is feasible from the output perspective. This is done by summing up the feasibility indicator from all subproblems. The error between the current and the previous iteration is used as one stopping criteria. As a result, output-feasible and optimal solutions can be retrieved. In the next chapter, performance of the implemented distributed optimization methods will be presented and analyzed.

Chapter 5

Case study and comparison

In this chapter, the centralized and distributed optimization approaches (PALR, ADMM, and DRSBK) are compared. Case studies, consisting of different numerical experiments, are considered. Results then will be analyzed from both performance and numerical point of views.

This chapter comprises the following sections: First, Section 5-1 explains the performance criterion that is used to analyze in case studies. Second, the setup and assumption are described in Section 5-2. Next, evaluation of the centralized optimization approach is presented in Section 5-3. Section 5-4 presents numerical simulations results along with the analysis between centralized and distributed optimization approaches. Finally, findings from this chapter are concluded in Section 5-5.

5-1 Performance criterion

In order to analyze the performance from the numerical perspective, different optimization approaches will be compared on the basis of the defined criteria. This enables us to review the results from the performance and computational perspectives.

Objective function

All simulations in case studies are carried out in closed-loop. The total objective function value over the simulation horizon is compared for each algorithm, which is calculated as:

$$J_{\text{cl}} = \sum_{t=1}^{T_{\text{sim}}} \sum_{i=1}^{N} C_i \tilde{V}_i \tag{5-1}$$

where T_{sim} are the simulation horizon. C_i and $\tilde{V}_i$ are previously defined in Chapter 3. The performance of distributed approaches will also be calculated based on the same objective functions. In this case, the decisions from all subproblems are aggregated such that the centralized objective function is computed. To compare the solution given by centralized optimization with the solutions from other distributed schemes, the following objective function is evaluated:

$$J_{\text{norm}} = \frac{J_{\text{meth}} - J_{\text{cent}}}{J_{\text{cent}}} \tag{5-2}$$

the J_{meth} is the objective function from the distributed optimization approach.

Computational time

To see the effectiveness of the distributed optimization approaches, the computational costs of centralized and distributed optimization are compared. This is done by calculating the required computation time to finish the entire simulation, from initialization until a solution is generated. The time is expressed in seconds.

Number of performed tamping and renewal

The number of performed tamping and renewal over the simulation horizon will be summed up, respectively. In general, the lower the number of interventions, the more cost-effective the algorithm. This criterion belongs to the railway performance perspective.

Total track performance

The total track performance or degradation level over the prediction horizon is to be evaluated as well. This criterion also belongs to the railway performance perspective.

5-2 Setup

All simulations in this research are conducted on Lenovo Thinkpad X260 with an Intel core-i5 processor and 8GB of RAM. All the LP and MILP problems are solved by Gurobi optimizer 7.5, called from MATLAB R2017a. Moreover, the following assumptions and general settings underpin the simulation of case studies:

- The control horizon is six months, which is inspired by the measurement from Dutch railway case [1]. This means that optimization is only conducted once for every six months. The time step of maintenance intervention is one month, which leads to six consecutive decisions for six months. The default prediction horizon is set to be nine. But, it can also vary, either with similar or higher period than the control horizon.
- Initial condition and degradation rate for each track section can be different, depending on the case studies. Different degradation rates for each track section are randomly generated as a Gaussian distribution. This distribution is adapted from the practical study of [3]. Also, the degradation rate is assumed to be known and constant within the simulation horizon, or in other words, the system is time invariant.

- The cost of renewal operation is set to be 30 times that of than tamping. This is to indicate that renewal is a very expensive operation. Other parameter values can be checked in Table 5-1.

Table 5-1: Fixed parameter data

Parameter	**value**
Offset (α)	0.001
Maximum level [mm] (h_{max})	1.8
Minimum level [mm] (h_{min})	0.01
Machine (tamping and renewal) travelling time/ track section [min.] ($t_{\mathrm{t2}}/t_{\mathrm{r2}}$)	0.5
Tamping intervention time/ track section [min.] (t_{t1})	33
Renewal intervention time/ track section [min.] (t_{r1})	60
Renewal/tamping cost	30
Maximum number of tamping/ track section over T (g_t)	4
Maximum number of tamping/ renewal over T (g_r)	1
Renewal allowance threshold (h_r)	1.53

- The system is deterministic, meaning that no stochasticity or any perturbation involved. This also implies that reactive maintenance is not considered.

- Several case studies with different settings and conditions are designed. Four different experiments in Section 5-3 are used to evaluate the centralized optimization. The other three experiments in Section 5-4 are used to compare and analyze distributed optimization approaches.

- Practical considerations, such as maintenance machine and personnel, are assumed to be always available, but their capability to maintain the track is adjusted according to case studies. Moreover, the asset of study case is a single railway track, consisting of multiple track sections. Each track section is 200 m. The total length of the track is adjusted according to the case studies. Two stations are situated at the start and end points of the track. When performing any maintenance on the track, the tamping or renewal machine goes from one station to the other. The illustration is depicted in Figure 5-1.

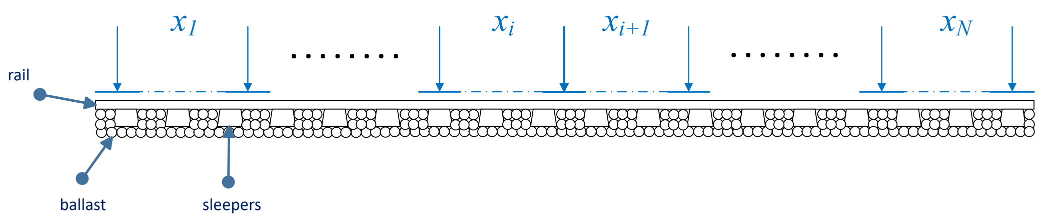

Figure 5-1: Illustration of track sections

5-3 Evaluation of centralized optimization

This section discusses case studies for the centralized optimization. The goal is to analyze the behavior of the proposed optimization problem in the centralized scheme, mainly from railway perspective. This includes the effect of maintenance operations on track performance and interventions costs. Furthermore, numerical performance can also be examined through the computational time of each case.

5-3-1 Experiment 1: similar initial condition with random degradation rates

First of all, a case study with similar initial degradation level and Gaussian random degradation rates are discussed. The Gaussian-random function represents natural distribution of slow, medium, and fast degradation rates. The detail can be seen in Appendix A. The track consists of eight track sections. The simulation horizon is five years. It is chosen since it allows the behavior of track degradation and maintenance operations to be completely shown.

Figures 5-2 and 5-3 depicts simulation results of the first experiment. In these figures, it can be observed that the track has various degradation curves. In general, different degradation rates leads to different number and timing of maintenance intervention. First, a pattern in the maintenance operations for the whole track can be observed in Figure 5-4. In the beginning, from the month 19 to 40, the optimizer suggests that the machine and personnel should treat more than one track section often, afterward they do nothing for a few months. This is because the offset memory is still moderate, thus it is more efficient to tamp multiple track sections at one time step if the closure time allows. This pattern is changed when the degradation level is approaching the limit. Thus, tamping is performed more frequently, with no more than two track sections at once. The time gap with the previous tamping is also reduced. At some point where tamping is not efficient anymore, it is necessary to perform renewal operation. Once the renewal is carried out, the track performance improves to the best condition. This might also imply that no interventions are required until the end of the simulations, as can be seen in Figure 5-2.

Apart from intervention timing, the number of required interventions for various degradation rates might also be different. For instance, in track section 2, it only requires three tamping works, while in track section 4, five tamping is required. This is because track section 2 degrades faster, so the renewal is performed earlier, rather than performing another tamping. On the other hand, track section 1 degrades slower, thus the algorithm waits until tamping is no longer effective. Additionally, the closure time constraint works perfectly in simulations. The maximum possible number of treated track sections are set to four. None of the suggested interventions exceeds four track sections. Furthermore, the maintenance map to see in which track section, maintenance actions are carried out at one time step, is presented in Figure 5-5. This kind of map can facilitate infrastructure managers to precisely schedule maintenance operations over the control horizon.

5-3-2 Experiment 2: different initial condition and degradation rates

Next, the second experiment provides a simulation with different degradation rates (slow and fast) and different initial values (low and high). Hence, this experiment results in four different

Table 5-2: Combination of parameters for the Experiment 2

		Degradation rate	
		Slow	**Fast**
Initial value	**High**	Section 3 - 4	Section 7 - 8
	Slow	Section 1 - 2	Section 5 - 6

combinations. Each combination has two track sections. Table 5-2 shows the corresponding track sections and their combinations. Meanwhile, simulation results can be seen in Figure 5-6 and 5-7. First, from the initial condition perspective, when the degradation level is started from a low degradation level, tamping is conducted just one time until the end of the simulation. This is shown in track sections 1 and 2. On the contrary, the high initial condition can lead to higher number of tamping, as shown in track sections 3 and 4. Thus, it will have a lower number of total interventions compared to the low initial degradation level, within the simulation horizon. But this is not the case for the track sections 7 and 8, where they perform less tamping than track sections 5 and 6. Track sections 5 and 6 have faster degradation rate, thus tamping is more required. Meanwhile, all track sections started from low degradation level do not require renewal since their offset memory states are below the limit. Conversely, renewal is needed in track sections 3, 4, 7, and 8. Track sections 3 and 4 perform it later than track sections 7 and 8. With low degradation rates, optimizer tries to delay renewal since it is very expensive. However, if it is seen from the entire simulation horizon, track sections 7 and 8 will spend fewer maintenance costs. In general, these decisions and responses have a similar pattern to the first experiment and they are expected from such a decision support tool.

Another potential problem coming from this kind of experiment relates to the closure time constraint. As the parameters are exactly the same, the occurrence of infeasible result can be high if the closure time is stricter. A high degree of parameter similarity increases the number of track sections that must be treated in a time step, while in the other time step, none of them are treated. This pattern can be observed in Figure 5-8 and 5-9, where every time a maintenance operation is undertaken, exactly two track sections are intervened. This is because there are two track sections that have exactly the same parameters in this test.

5-3-3 Experiment 3: different trade-off in the objective function

In the third experiment, the effect of different weights λ in the objective function are addressed. The weight is increased with the difference of the power of ten. Also, the other parameters are fixed. Table 5-3 depicts the numerical performance results. From the table, several remarks can be made. First, from the mathematical perspective, higher weights correspond to higher optimal objective function values. It is obvious since λ basically acts as a multiplier to the maintenance intervention costs. This condition also implies that the term which contains the input is treated more costly compared to the track performance term. Hence, the optimizer will theoretically prioritize the saving of interventions over the track degradation level [52]. Moreover, higher λ also means that in general, the optimizer is able to quickly generate the solution. On the contrary, lower values of λ mean that the track performance and interventions are on the same level of importance. Thus, the optimizer might require a longer time to look for an optimal solution. However, it can be seen that

Table 5-3: Numerical performance of various weights

Parameter / value	0.01	0.1	1	10	100	1000
Optimum J(V)	104.46	111.97	133.82	455.59	3695.60	36096.00
Time (sec.)	0.617	0.518	0.590	0.533	0.415	0.363
Number of tamping	23	15	7	6	6	6
Number of renewal	1	1	1	1	1	1

Table 5-4: Numerical performance of various prediction horizon

Parameter / value	6	9	12	15	18
Optimum J(V)	1692.2	1671.1	1604	1519.4	1462.5
Time (sec.)	0.8415	3.2213	5.1850	12.226	18.853
Number of tamping	16	16	13	8	5
Number of renewal	4	4	4	4	4
Total track degradation	332.248	311.122	273.995	239.401	212.528

the computational times from all settings are hardly different. Only the computational time of the instance highest weight has around half value of that of the instance with the lowest weight.

Furthermore, Figure 5-10 shows multiple degradation curves associated with different weights. As discussed before, lower value of λ means that the interventions are not expensive, thus the curve of the lowest weight has the highest number of tamping among others. As the weight is increased, maintenance interventions become more costly, thus the number of suggested tamping is reduced. However, as shown in Figure 5-11, the last three instances with the highest weights have a similar total number of tamping. This means that the effect of increasing weight is already saturated in this case. Moreover, all of them perform renewal as indicated by Figure 5-12. It can be that the first two lowest weights perform renewal earlier. This is likely due to the similar reason as discussed in Experiment 1 and 2.

5-3-4 Experiment 4: different prediction horizon

Theoretically, longer prediction horizon in a MPC-based approach leads to a solution closer to the global optimal [53]. The fourth experiment is defined to examine the effect of different prediction horizon. The result can be seen in Table 5-4. First, from the table, it can be observed that longer horizon is associated with longer computation time. It is expected since the optimizer must look further into the future, thus more decision variables are added to facilitate this. The computation time increases exponentially, as expected from NP-hard problem. In the large-scale instance, long prediction horizon might not be feasible to apply. Second, this also means that the optimal objective function value over the control horizon is reduced. Since the longer the prediction, the more knowledge needs to be considered by the optimizer. This also comes with the lower number of tamping and total track degradation.

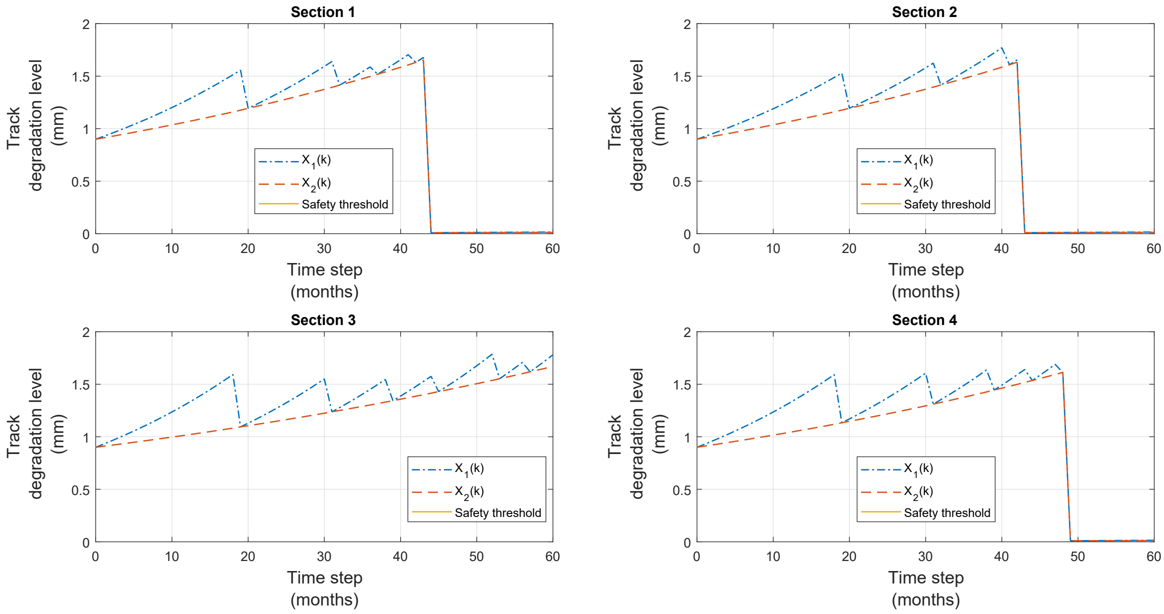

Figure 5-2: Track degradation curve for track section 1-4 in the Experiment 1

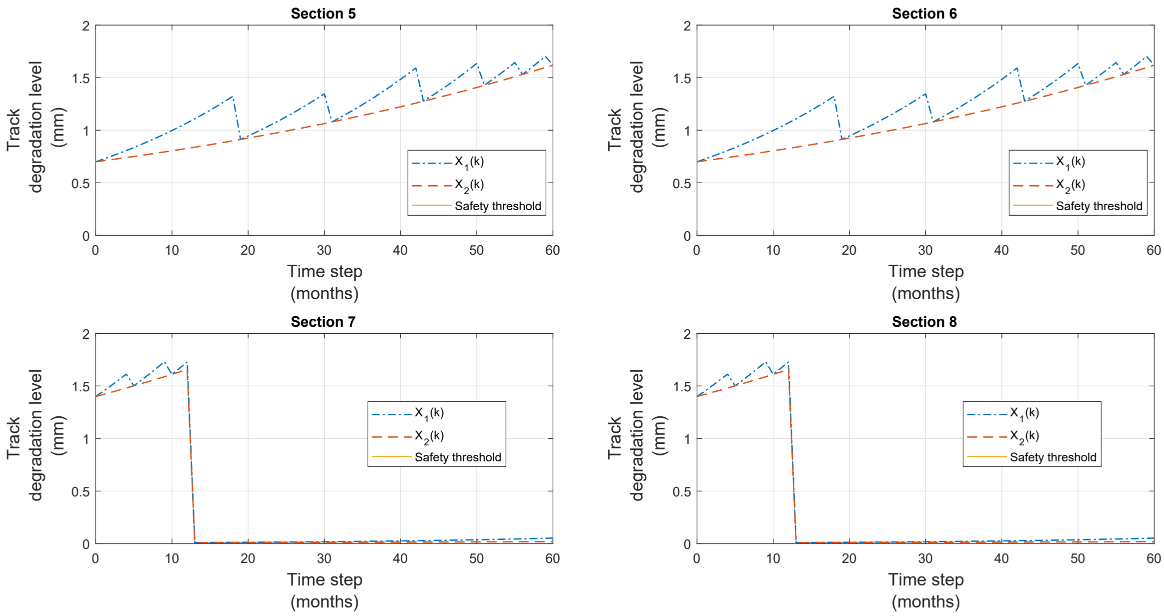

Figure 5-3: Track degradation curve for track section 5-8 in the Experiment 1

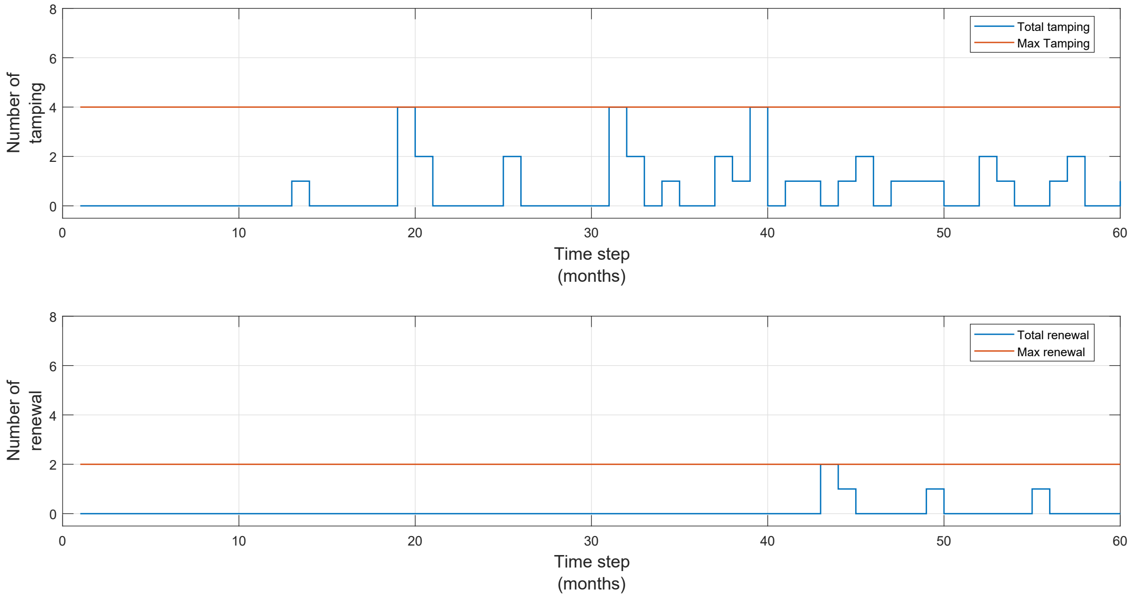

Figure 5-4: Total intervention (tamping and renewal) in the first Experiment 1

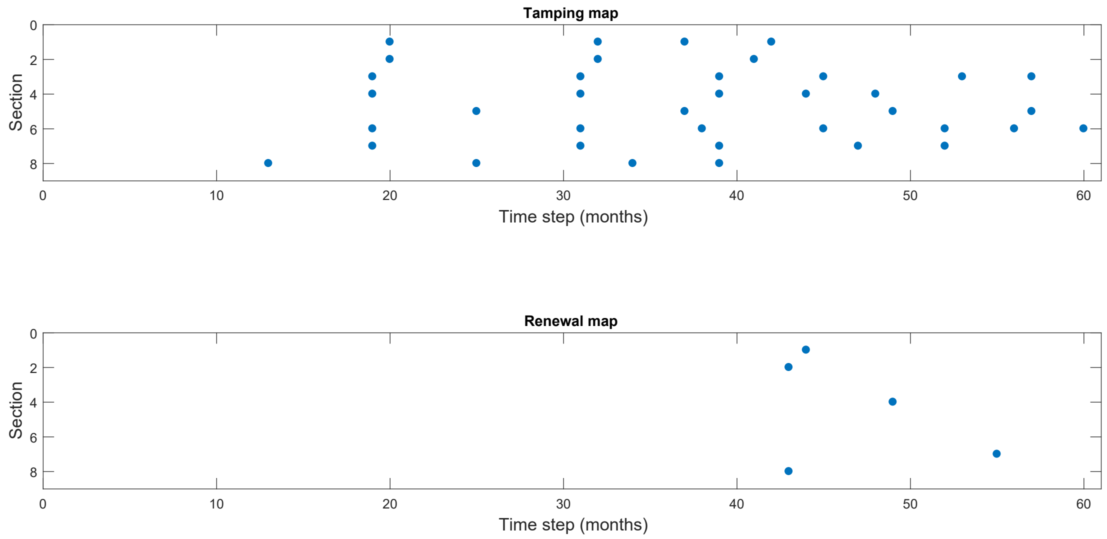

Figure 5-5: Mapping of intervention in the Experiment 1

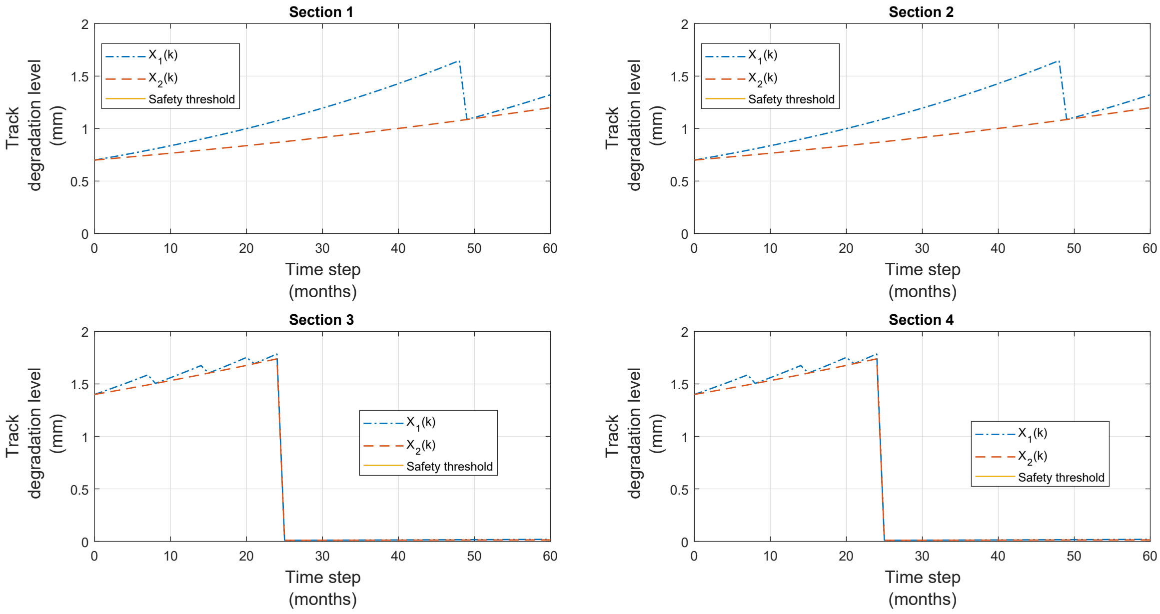

Figure 5-6: Track degradation curve for track section 1-4 in the Experiment 2

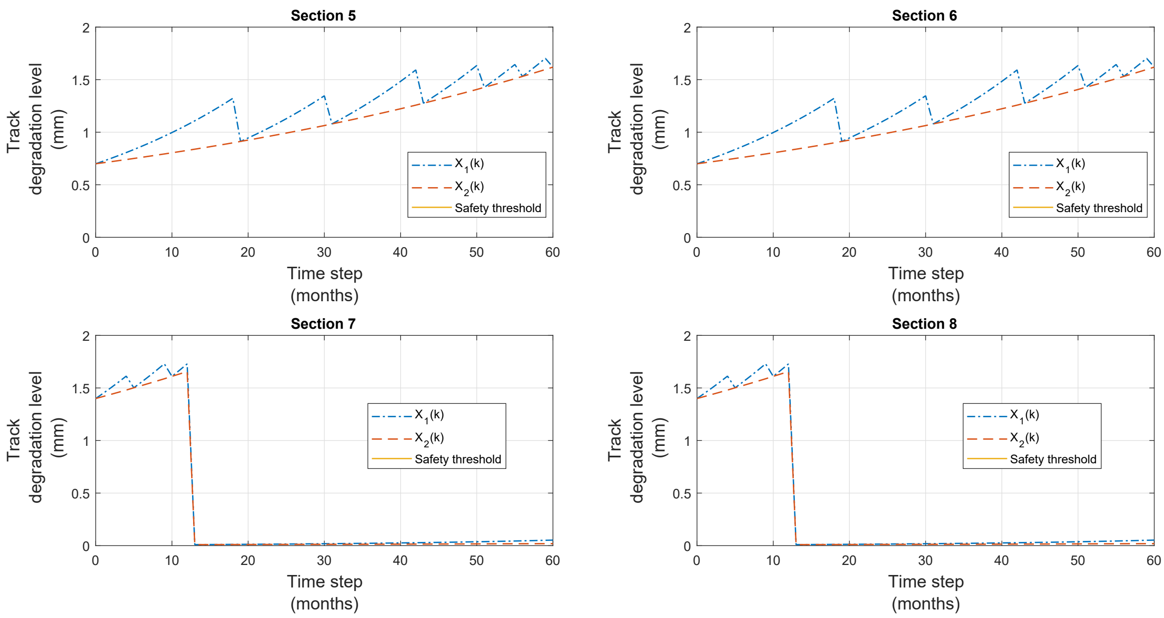

Figure 5-7: Track degradation curve for track section 5-8 in the Experiment 2

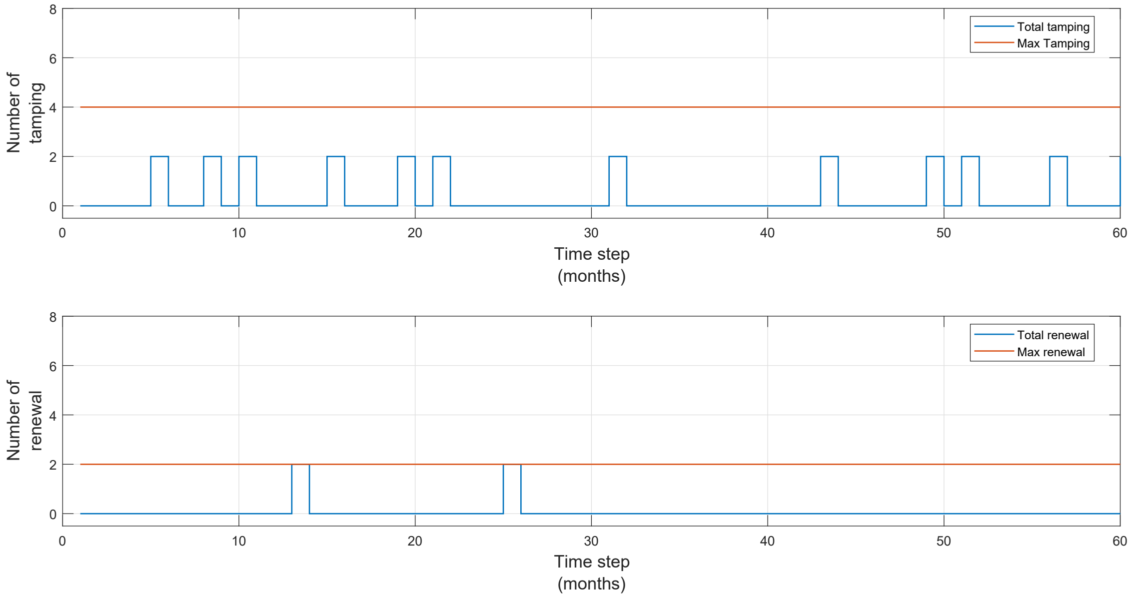

Figure 5-8: Total intervention (tamping and renewal) in the Experiment 2

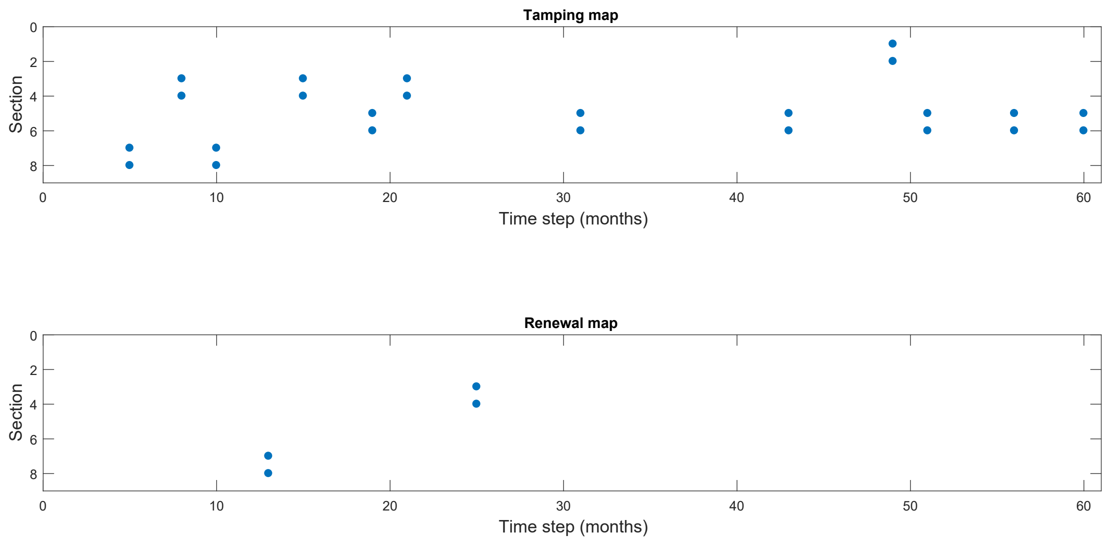

Figure 5-9: Mapping of intervention in the Experiment 2

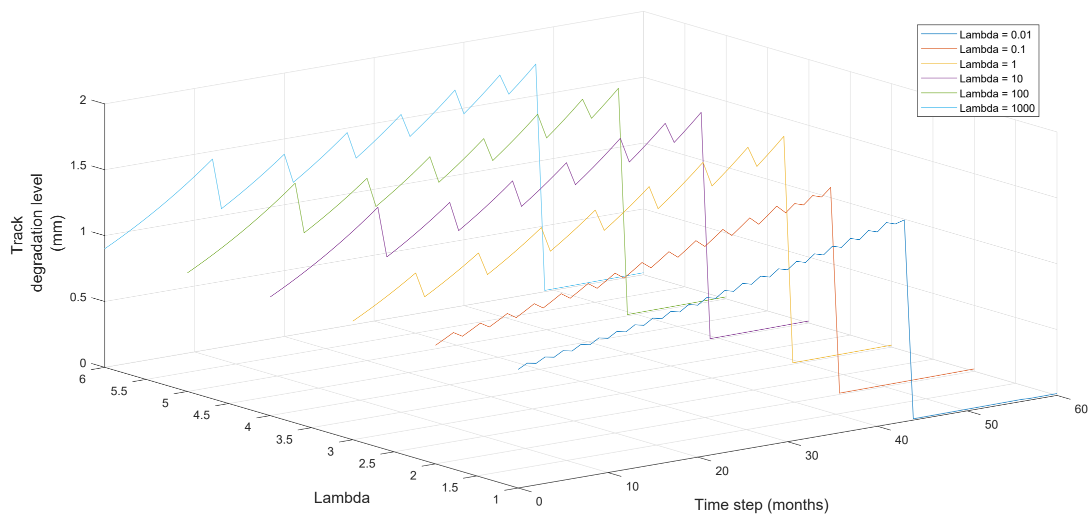

Figure 5-10: Comparison of degradation curves with different weight λ

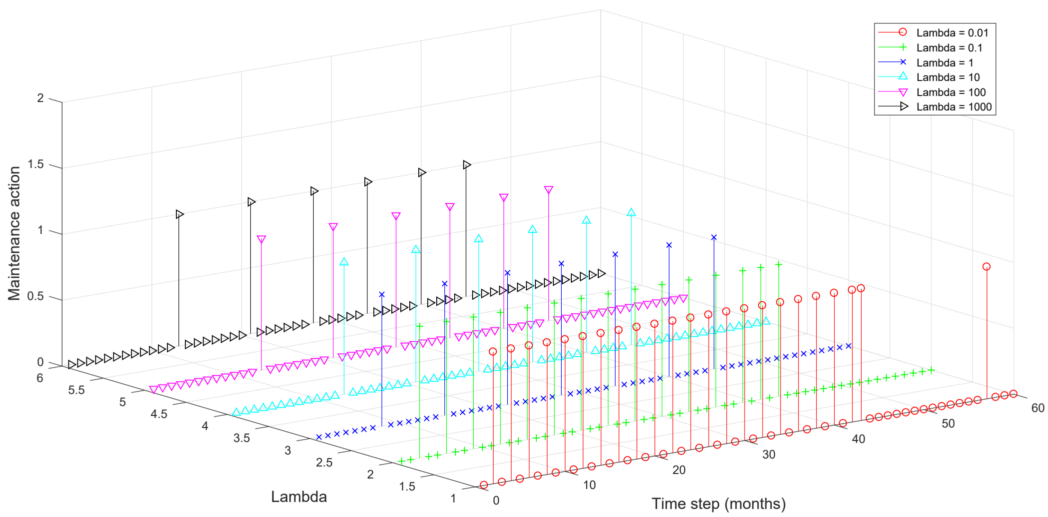

Figure 5-11: Comparison of tamping operations with different weight λ

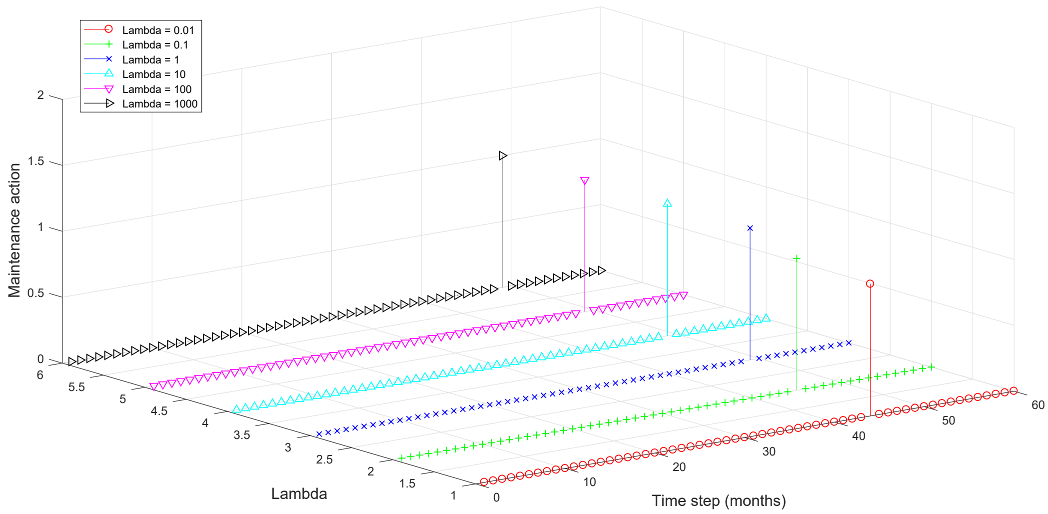

Figure 5-12: Comparison of renewal operations with different weight λ

5-4 Evaluation of distributed optimization

Next, the limitation of the centralized optimization and the advantages offered by distributed approaches are analyzed. In this section, distributed optimization approaches are evaluated from both the performance and numerical point of views.

Different case studies are considered. In general, the evaluation tests are categorized into two types. First, three different problem scales are established: small, medium and large. Each of them consists of 4, 50, and 150 track sections, respectively. To further thoroughly observe the behavior of distributed approaches, ten different simulations are carried out for each problem scale. The mean and standard deviation of the retrieved output is then calculated. Second, the distributed approaches are compared by increasing number of track sections gradually. A threshold is established based on the limitation of the centralized approach. In this way, each distributed approach can be pushed to their respective limit defined by the threshold. Likewise, it is notable that all simulations are conducted in the closed-loop. To further see how the algorithms comply with the coupling constraint, the following expression can be used to calculate the maximum treated track sections at one time step:

$$t_{\max} = t_{t1} * (N - c_{\max} * N/10) \tag{5-3}$$

where t_{t1} and $c_{\max}$ is a tamping intervention time and constant factor, respectively. N is the number of track sections. The limit is then decided right after observing the result of the loose coupling constraint or unconstrained instances, thus the c_{max} can be adjusted accordingly.

Apart from the settings, the parameter, which consists of the initial condition and degradation rates, is generated randomly using Gaussian distribution, as had been done in the previous section. It is particularly designated for medium and large-scale settings. The parameter data

Table 5-5: Performance comparison in small-scale problem

Parameter / algorithm	Centralized		PALR		ADMM		DRSBK	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
Total J(V)	1286.31	630.21	1354.96	596.87	1362	596.78	1292.35	636.27
Comp. time (sec.)	1.70	0.68	22.29	12.51	6.92	2.08	3.40	0.74
Number of tamping	15.90	6.35	22.30	6.93	25	8.25	16.30	6.75
Number of renewal	2.40	2.07	2.50	1.96	2.50	1.96	2.40	2.07
Total track performance	407.31	108.01	381.96	110.01	362.14	121.12	409.35	109.48
Normalized J(V)	-	-	-8.58%	0.13	-9.43%	0.14	-0.33%	0.01

Table 5-6: Performance comparison in medium-scale problem

Parameter / algorithm	Centralized		PALR		ADMM		DRSBK	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
Total J(V)	11365.43	5897.54	18542.59	1634.98	13816	5798.2	11975.95	6169.34
Comp. time (sec.)	107.94	92.1	145.48	56.29	54.69	25.20	34.78	5.60
Number of tamping	172.60	82.47	300.50	88.25	293.5	84.05	172.90	82.69
Number of renewal	16.30	16.96	30.50	4.50	20.10	18.66	16.30	16.96
Total track performance	5357.03	1021.20	6387.58	187.49	4851.30	1050.08	5356.96	1041.11
Normalized J(V)	-	-	-112.89%	1.37	-22.15%	0.16	-0.02%	0

Table 5-7: performance comparison in large-scale problem

Parameter / algorithm	Centralized		PALR		ADMM		DRSBK	
	mean	**stdev**	**mean**	**stdev**	**mean**	**stdev**	**mean**	**stdev**
Total J(V)	36200.18	12619.83	54137.37	5856.02	45174	10102.62	36205.13	12619.1
Comp. time (sec.)	871.84	512.43	453.74	117.02	160.12	21.48	104.67	15.15
Number of tamping	493.00	105.77	688.00	213.42	946.00	35.35	493.5	105.58
Number of renewal	43.5	43.17	94.78	22.06	58.30	41.05	43.5	43.17
Total track performance	18220.18	2361.42	18824.03	1106.48	15300.91	3977.68	18220.13	2361.36
Normalized J(V)	-	-	-52.00%	0.63	-31.00%	0.15	-0.02%	0

distribution and allocation over the entire track sections used in case studies are presented in Appendix A.

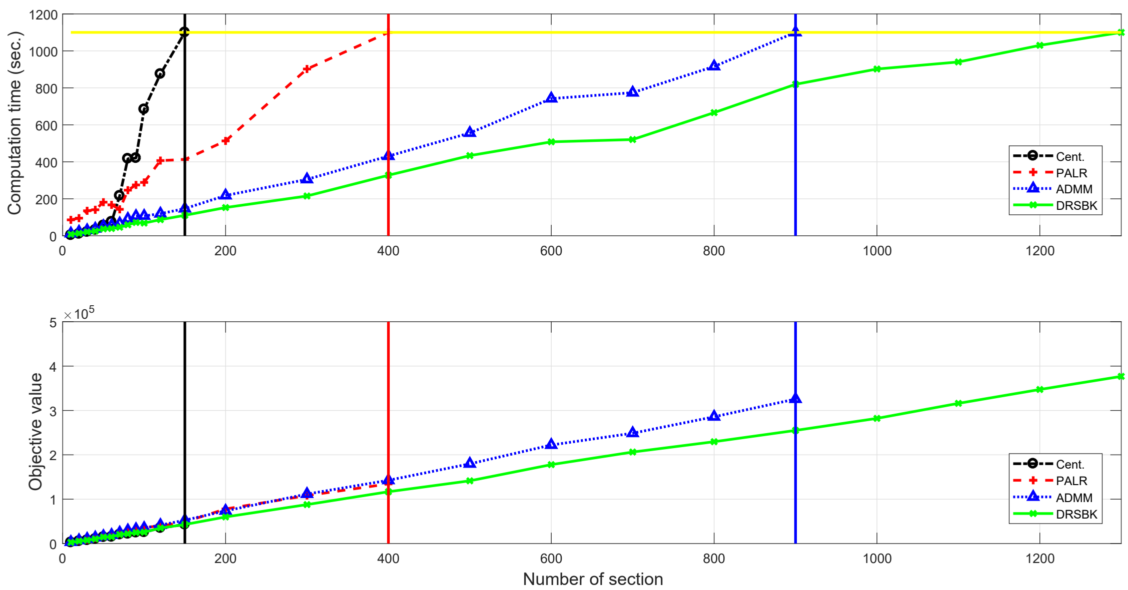

Figure 5-13: Comparison of computation time and objective function value

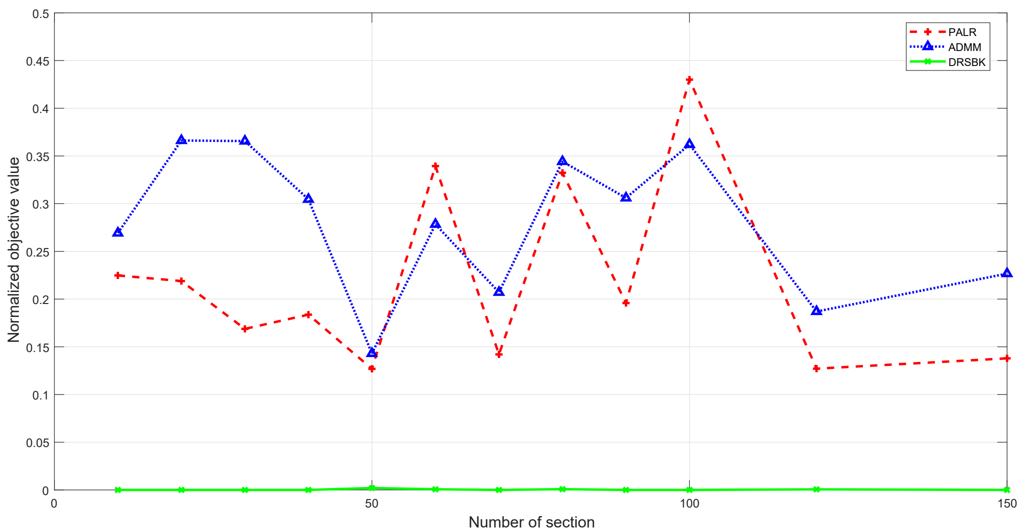

Figure 5-14: Comparison of normalized objective function value

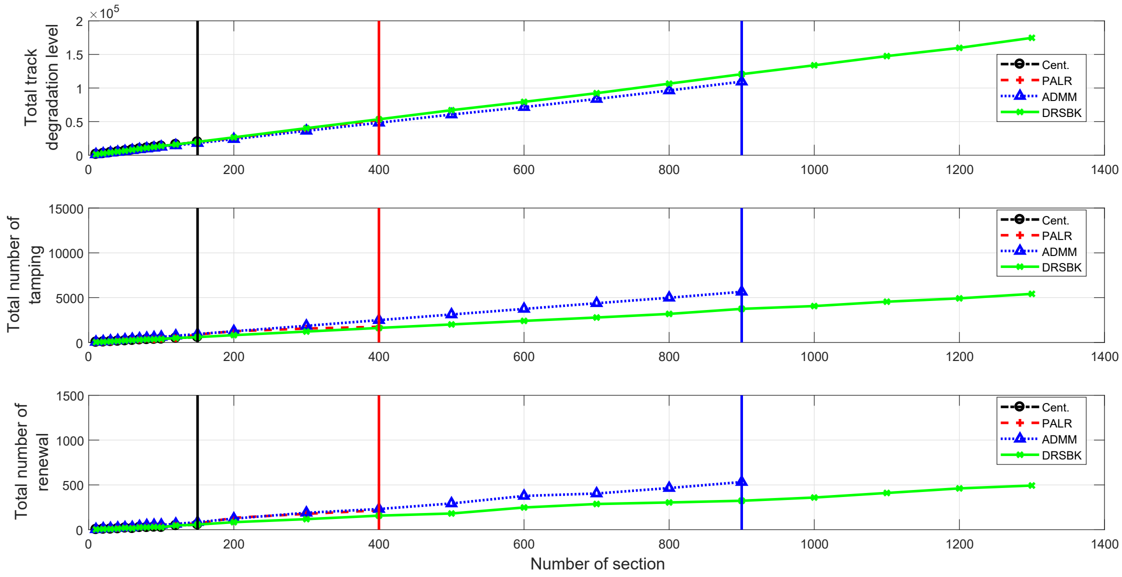

Figure 5-15: Comparison of total degradation level, number of tamping and renewal over the simulation horizon

5-4-1 Experiment 1: multiple tests on small, medium, and large-scale settings

The comparison for small, medium, and large-scale are presented in tables 5-5, 5-6, 5-7, respectively. Each table presents the average and standard deviation values from overall ten different tests for every problem scale. First of all, in the small-scale tests, it can be seen that the centralized optimization approach has the fastest computation time among all of the approaches. This is expected since in small-scale, the number of variables is relatively small. Hence, the centralized optimization approach is still tractable and distributed approaches are not urgently required yet.

Next, among distributed approaches, DRSBK is the fastest algorithm. This is expected from such simple method that modifies the couplings allocation instead of the augmented Lagrangian of the objective function. On top of that, it generates the closest objective function value to the centralized problem. This also implies that the average number of suggested tamping can be exactly the same as the one suggested by the centralized approach. Likewise, the difference in the number of renewal is only one. It can even have exactly the same solution with the centralized approach in some tests, as can be seen in Appendix B. Apart from DRSBK, the solutions given from Lagrangian-based algorithms are not close to the centralized solution. Since no convergence guarantee to the proposed MILP problem, they have to a spend longer time in their iterations, looking for feasible results. The two-step computation also implies longer processing time. Besides, they suggest a higher number of tamping and renewal than the centralized approach. ADMM has a better average solution yet shorter processing time than PALR. This is because ADMM uses the latest decisions to take into account, thus it is easier to get into feasible regions than PALR. However, it does not mean that the result is global optimal or close to it. Actually, it might be possible to improve both Lagrangian-based algorithms to find a solution closer to the global optimal, but this requires more iterations, sacrificing the computation time. Furthermore, since they suggest more interventions, the track degradation level can be lower than the centralized problem and DRSBK. In some conditions where the track is busy, these results are preferred. However, of course, it is not cost-effective. Compared to Lagrangian-based methods, DRSBK is much simpler to develop yet it could perform remarkably well in this research. Additionally, all algorithms are able to give a feasible solution without violating any coupling constraints. This is due to the extension methods in Lagrangian-based algorithms that drive the iterations into feasible results. The original DRSBK is by default designed to follow couplings. Thus, all of the distributed optimization approaches are comparable. Additionally, the examples of simulation results from all approaches in the first test are given in Appendix A. All of them show typical responses and no couplings were violated.

The advantage of distributed approach in computation time is started to rise in medium-scale comparison, especially DRSBK and ADMM. The table shows a significant improvement of computation time offered by both approaches, which surpasses the centralized optimization. However, the solutions given by other two Lagrangian-based algorithms PALR and ADMM are far from global optimal. Since the problem size is bigger, it might be more difficult to find the feasible regions, in the sense that more iterations are required. More number of iterations means increasing computation time. Furthermore, DRSBK again has the closest value to the centralized approach. The objective function values of PALR and ADMM now have greater difference with the centralized approach compared to their solutions in small-scale instances. ADMM is better than PALR in all criterion, even the track performance is the most minimal among all algorithms. From both performance and numerical perspectives, PALR has the worst overall performance among the three.

Finally, in the large-scale instance, it is shown that all distributed schemes have significantly faster processing time than the centralized. In such large-scale problem, the number of variables is really big, leading to the exponential computation time and centralized optimization might no longer be tractable. Thus, distributed optimization approaches can start to play a role. PALR can halve the centralized processing time. Likewise, the same pattern for both Lagrangian-based algorithms is again noticed. Again, ADMM is better at all criterion than PALR. Also, DRSBK stands out as the best distributed algorithm from the numerical perspective. Among ten tests in each problem scale, there is an issue of some distributed algorithms having a lower objective function value than the centralized optimization in a few tests. For instance in the third test in small-scale, ADMM outperforms the centralized approach. From the observation, it is shown that ADMM decides to conduct renewal operation earlier than its centralized counterpart. At that moment, ADMM spends more maintenance cost than the centralized approach but afterward, ADMM does not have to conduct tamping anymore until the end of the simulation. The centralized approach still suggests some tamping and it performs renewal when the degradation level already high. This also implies the total degradation level of the ADMM can be far lower than the centralized approach. This issue is always started by the centralized having lower objective function value at one cycle, which makes sense because theoretically, the solutions of distributed approaches are the same or worse than the centralized problem. The constraint 3-8 is defined to prevent this issue to arise, but the threshold can only be set up to 85%. If the threshold is higher than that, the computation process is untractable. Additionally, the responses can be observed in Appendix A

5-4-2 Experiment 2: gradual increment tests

The second experiment presents a comparison of performance criterion with gradual increment in the number of track section. With the higher number of variables taken into account, the problem sizes will be bigger and so typically the computation process will take much longer time [44]. For the NP-hard problems, this can be worse, since the problem complexity might increase exponentially [50].

The simulation settings are as follows. The initial degradation level and rates are Gaussian-randomized for all track sections. This is done in the same way as before, but the mean and standard deviation remain the same for all tests. Likewise, the number of track sections is increased with the difference of 10 in the beginning and after $N = 150$, the gap is set to be 100. The rule for all algorithms are once the threshold for centralized is reached (in this experiment, it is 1100 seconds), the test of the corresponding approach is stopped. In this way, each algorithm will have different data length due to their different capabilities in handling large-scale problem. The limitation of each algorithm can be compared as well. Moreover, other parameters are fixed.

Next, an illustration which shows the simulation results of the first criteria (computation time) against a number of track sections from the proposed optimization problem is depicted in the first plot of Figure 5-13. From the figure, it can be observed that the computation cost increases with the higher number of track sections. This issue is suffered not only by centralized optimization but also the distributed approaches as well. However, the curve of the centralized approach is exponential, which can be expected from such centralized NP-hard problem.

It stops the experiment earlier than the other algorithms, at $N = 150$. PALR can continue to perform computationally reasonable until it reaches 400 track sections. ADMM can prolong its experiment until $N = 900$. ADMM can outperform PALR possibly due to the use of current iteration data rather than the previous iterations. In this way, the algorithm can quickly find feasible solutions, in early iterations. On top of that, DRSBK can treat up to 1300 track sections, which proves itself to be the most tractable among all algorithms. Moreover, it is notable that during simulations in large-scale instances (with $N > 150$) the simulations of centralized optimization is sometimes not possible to be carried out due to lack of available memory in the computer. This issue never happens during the simulations of distributed optimization algorithms. Furthermore, the second plot in Figure 5-13 shows that in general, the curve of objective function values for all algorithms are linear. From start to the end of the centralized approach computation, all of them look coinciding with each other. After $N > 150$, PALR and ADMM have diverted curve due to their higher objective function values. They are very likely trapped in local optimal points. On the other hand, the curve of DRSBK is linear which means that its solution is close to the global optimal. In Figure 5-14, the distance between centralized and distributed approaches are shown. PALR and ADMM could only retrieve local solutions, which are far from the benchmark. On the other hand, DRSBK solution can manage to stay close to the global optimal.

From the railway performance perspective, the total track degradation level and the number of tamping and renewal operations are also discussed. In the first plot, all algorithms have linear curves. A small gap can be noticed where PALR and ADMM have smaller values than DRSBK. This is due to the fact that they suggest more number of interventions. This can be confirmed in the second and third plots, where both Lagrangian-based algorithms have higher curves. Theoretically, lower value of track degradation level is good to reduce the risk of unavailability, especially in highly occupied tracks, but doing so requires a higher number of interventions, which is not preferable from the economical point of view. On top of that, it can be inferred that the number of track sections does not change the working characteristics of all approaches.

5-4-3 Experiment 3: feasibility compensator

In the practice of track maintenance, the machines and personnel required to maintain the ballast is limited or not always available [1]. Besides, the track possession time for maintenance can also be reduced due to the fact of, for instance, busy network. In this case, the target track sections cannot be fully handled at one time slot. On the other hand, this leads to unfeasible results from the output perspective despite the coupling constraints are fulfilled. Hence, the untreated tracks will be fixed later with spot tamping or renewal.

To deal with such conditions, this experiment will test the heuristic compensator algorithms applied in distributed approaches. The data from test 3 of the small-scale instance is used. The resource of couplings is reduced so that only one tamping is allowed and renewal could not be conducted at all. The simulation results for PALR, ADMM, DRSBK are depicted in Figures 5-16, 5-17, 5-18, respectively. First, it can be observed that all of them suffers unfeasible results which can be indicated by the degradation level violating the maximum limit. This condition can be handled by spot tamping and renewal operations, which are suggested by the compensator. As can be seen in the figures, the passing degradation level can be brought back to the safe region by spot interventions. With the imposed limitation, DRSBK can properly assign tamping operation thus, it only requires spot renewal, as shown in Table 5-8. PALR and ADMM on the other hand could not get feasible solutions for tamping. In addition to spot renewal, they consequently need to perform spot tamping. With stricter coupling, the feasible region might be more difficult for Lagrangian-based algorithms to reach it. Moreover, In this experiment, it is assumed that the spot interventions only available by the end of control horizon. Therefore, the dangerous condition might last for a couple of months. For instance, it is shown in the curve of the track section 1 in PALR result where it lasts for four months. In the end, spot tamping is performed accordingly.

Table 5-8: Comparison of required spot intervention

Parameter / Algorithm	PALR	ADMM	DRSBK
Spot tamping	3	2	0
Spot renewal	4	4	4

5-5 Conclusions

In this chapter, various experiments with different settings and conditions are designed. In general, the experiments can be grouped into two categories. The first one is used to evaluate the behaviour of centralized optimization. This consists of four different experiments. The

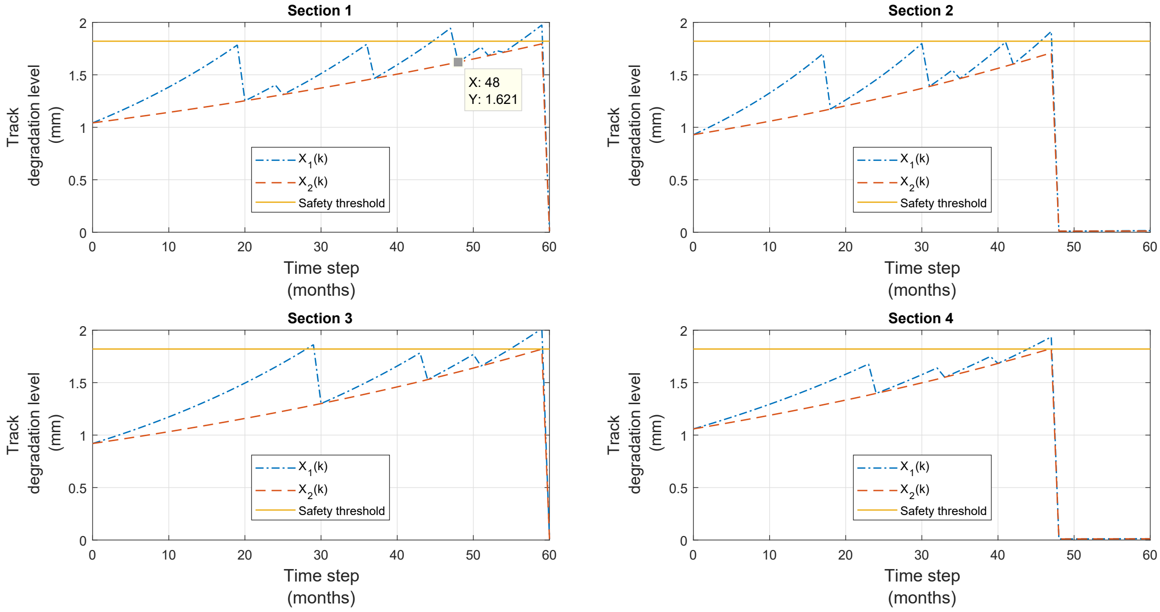

Figure 5-16: Compensator output in PALR

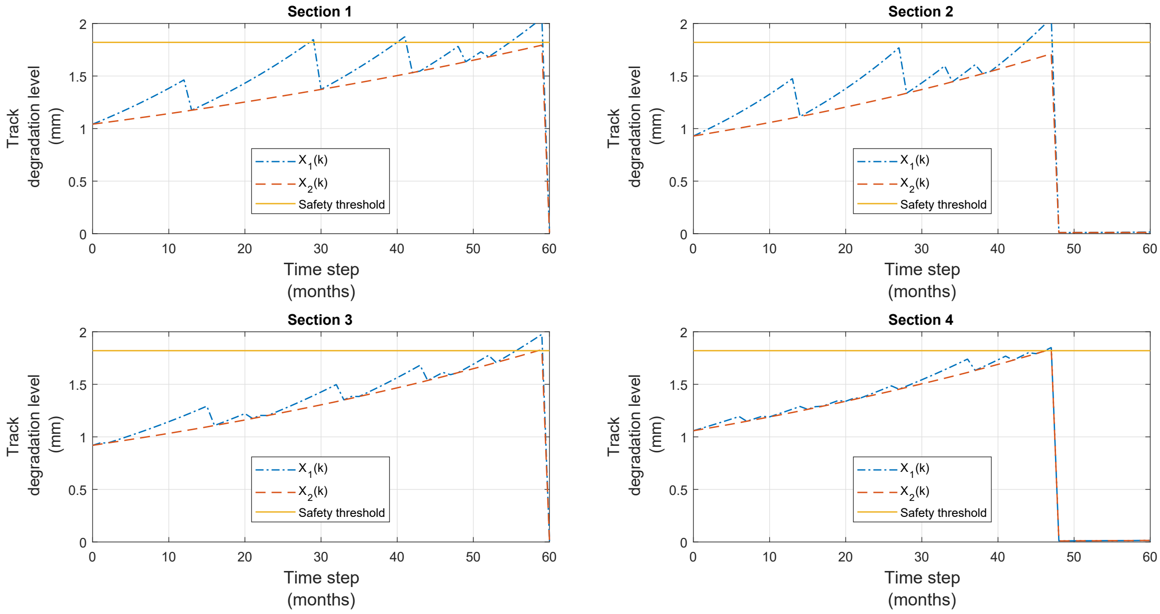

Figure 5-17: Compensator output in ADMM

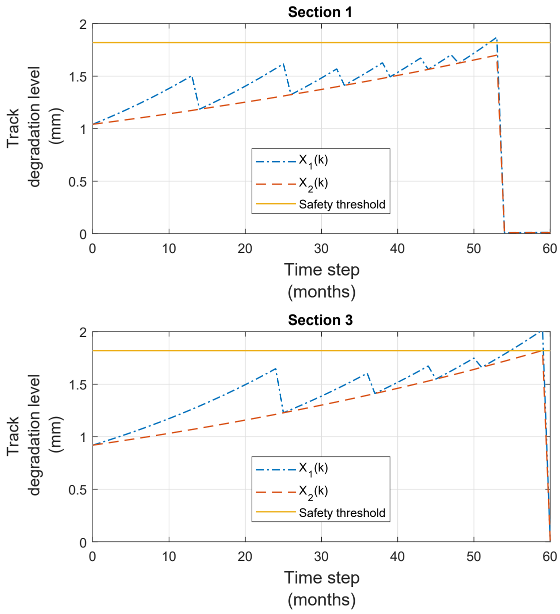

Figure 5-18: Compensator output in DRSBK

second category aims at comparing and analyzing the performance of distributed optimization approaches. Two different experiments are carried out under this category.

In the first category, it is shown that the proposed optimization problem formulation works properly. The centralized optimization can suggest timely maintenance interventions. Generally, tamping is suggested when the degradation level is high but the offset memory is low. When offset memory is approaching the limit, renewal is performed. The trade-off variable λ and prediction horizon can also be tuned to find a good balance between quality of the solutions and maintenance or computational costs. In the second category, it is shown that the centralized optimization approach encounters an enormous computation time for large-scale settings and is often not feasible in practice due to the other issues, such as computer memory limitation. It was shown that given a limitation in computational time, the centralized optimization is not scalable.

Furthermore, two Lagrangian-based approaches (PALR and ADMM) and a constraint-tightening approach (DRSBK) can solve the formulated problem with more flexibility. The extension methods enable the Lagrangian-based algorithms to get feasible solutions. In general ADMM, can generate solutions slightly closer to the global optimal yet quicker than PALR. However, both algorithms performance are still below DRSBK, which outperforms them in term of computational time and the normalized objective function value. However, such random sequence-based algorithm is not guaranteed to convergence to the same solution in each run. This is because neither the objective functions from the other subproblems taken into account nor common price (like dual variables in Lagrangian-based approaches) are used by DRSBK. Additionally, when the coupling constraint is stricter, the comparison algorithm is able to handle unfeasible track performance by suggesting spot interventions by the end of control horizon.

Finally, the experiment results and analysis show that the Lagrangian-based approaches do not perfectly suited to handle distributed optimization of MILP problems with integer inputs, like the optimization problem in this thesis. Instead, a simple constraint tightening approach as used by DRSBK could work better by reducing computation time while maintaining good solutions in the distributed optimization of such NP-hard problem.

Chapter 6

Concluding Remarks

This chapter provides the conclusions obtained from conducting this research and direction for future research of the current methods. The implementation of distributed optimization in railway maintenance operations planning field is novel, thus there are some open issues and challenges yet to be explored. This chapter begins with the explanation of the conclusions in Section 6-1. Afterwards, the research questions are answered in Section 6-2. This is followed by the discussion of future works in Section 6-3.

6-1 Conclusions

In this research, three different distributed optimization approaches have been developed for large-scale railway track maintenance operations planning. The main objective is to reduce computation costs imposed by the optimization of large-scale maintenance problem while maintaining the solution close to the global optimal. To that end, several procedures are defined and carried out in this thesis.

First of all, the formulation of an optimization problem is done. The objective function applied contains the track performance and maintenance costs terms. The merit of exploiting such objective function is that the balance between the two terms can be optimized. Moreover, the formulation includes prominent characteristics of track maintenance operations such as degradation dynamic, early renewal prevention, degradation limits, maximum number of interventions, and maintenance closure times. In general, the constraints can be categorized into individual and coupling constraints. The formulated optimization problem can be categorized as Mixed-Integer Linear Programming (MILP), which is also a non-convex NP-hard problem.

Three distributed approaches are designed for the proposed optimization problem. The couplings constraints, which are tamping and renewal closure times, are firstly identified and reformulated to cope with the distributed approaches. The first two distributed optimization approaches work based on Lagrangian duality theory: Parallel Augmented Lagrangian Relaxation (PALR) and Alternating Direction Method of Multipliers (ADMM). These approaches are modified with extension techniques such that they can handle the non-convex mixed-integer problem. The extension technique basically consists of a two-step method where the first uses the continuous relaxation to provide the warm start variables and bounds for the next MILP step. Furthermore, a stopping criterion is designed, such that input-feasible and suboptimal solutions can be retrieved within a reasonable time.

Alongside the Lagrangian-based approaches, Distributed Robust Safe But Knowledgeable (DRSBK) is implemented. Instead of modifying the objective equation through dual function, this approach exploits constraint tightening concept. In this way, the couplings can be reformulated as an individual constraint with diminishing allocations. The subproblems are hence solved sequentially. To avoid infeasible solutions, A random sequence process algorithm is also added. This makes the algorithm become iterative. Likewise, a stopping criterion is used such that the solution is feasible from the output perspective. Furthermore, a coordinator is also employed for each approach to update the dual variables (PALR and ADMM) or update the sequence (DRSBK).

In the case studies, it is shown that the distributed optimization approaches are able to solve the proposed problem quicker than centralized optimization in large-scale instances. DRSBK rises as the fastest yet still able to generate the closest solution to the centralized problem. ADMM is quicker than PALR. However, solutions from both Lagrangian-based methods are suboptimum. Therefore, a Lagrangian dual-based approach might not be a suitable option to be developed for the future research of the proposed optimization problem.

6-2 Research questions

In this thesis, the following main research question is defined:

Is it possible to improve the computational performance while maintaining a good quality of the solutions for large-scale railway track maintenance operations planning problem solved by a distributed optimization approach?

To deal with this research question, complete explanations regarding the proposed optimization problem, distributed optimization, and case studies are given in previous chapters. Moreover, brief answers to each subquestion are discussed below:

1. *Which optimization problem can capture characteristics like degradation of track performance, maintenance interventions and closure time for large-scale maintenance operations in railway tracks?*

   The proposed optimization problem includes the objective function that can show the trade-off between track performance and maintenance costs. The track degradation dynamics with integer inputs are also accounted to estimate the track condition, based on measurement. Moreover, the constraints are in general grouped into two, individual and coupling constraints. The individual constraints include degradation dynamic, early renewal prevention, degradation limits, and maximum number of interventions. The coupling constraints are tamping and renewal closure times or down times. The problem is deterministic. Hence, the proposed optimization problem already considers the prominent characteristics of the track maintenance operation planning.

2. *Which distributed optimization approaches can deal with the proposed optimization problem and reach feasible global solutions?*

   The proposed optimization problem contains non-smooth couplings problem which might not be able to converge. Based on literature survey, three distributed approaches which can deal with such issue, are chosen. The Lagrangian-based algorithms, PALR and ADMM, are modified with two-steps extension techniques to handle the non-smooth coupling constraints such that feasible solutions can be obtained and none of the couplings are violated. The third approach, DRSBK, equipped with an extension technique, is also able to get feasible solutions.

3. *How is the performance of distributed approach in comparison with the centralized approaches in terms of the quality of global solutions and computation time?*

   In the study cases, it is shown that the distributed optimization approaches are able to outperform the centralized optimization in large-scale instances in computation time. DRSBK rises as the fastest approaches among the others. This is followed by ADMM and the worst is PALR. From performance perspective, DRSBK is able to generate closest solutions to the centralized problem. On the other hand, solutions from PALR and ADMM are suboptimum, which is far from the global solution.

Having the answers from all subquestions, the answer to the main research questions can be written as follows:

Three distributed optimization approaches, (PALR, ADMM, and DRSBK) can be implemented to the track maintenance operations planning problem to reduce computation time when dealing with the large-scale problem. However, only DRSBK can maintain a good quality of the solutions.

6-3 Future work

Several challenges can be found in conducting further research and development in the field of distributed optimization of railway maintenance operations, among others:

Adjacent track scenario

At the current research, the number of maximum maintained track sections are assumed to be constant. However, in practice, they can vary according to the position of the treated track sections. If the switching times between adjacent tracks are eliminated, this enables the machine to treat more track sections. Nevertheless, this makes the feasibility checking of solutions become more complicated.

Real-life measurement for case studies and system identification

The use of real data is important to evaluate the effectiveness of the distributed optimization approaches to handle track maintenance planning in practice. Moreover, new challenges can also come into play. First, this enables the use of practical constraints, such as track curve limitations. Moreover, the other maintenance options for ballast can also be added to the options. More options might give rise to more practical constraints. In practice, there is also the possibility that some assumed parameters are no longer fixed, such as degradation rates and offset from tamping. Thus, system identification can also be deployed to accurately model the practical parameters. The degradation dynamics can also be improved to be time-varying instead of time-invariant.

Maximum number of maintenance interventions as coupling constraint

The constraint for maximum number of maintenance can also be assigned to the couplings as well. In this way, the problem will be a bit more complicated since the optimizer must allocate the allowed number of interventions for each track sections properly, with respect to the total budget. It can also be multiplied with some constant to reflect the maintenance cost for each type of intervention.

Stochastic model and perturbation

The current deterministic model can be developed into a stochastic model, which can take into account uncertainty coming from the practical knowledge in track maintenance, such as degradation dynamics, measurement error, ineffective tamping, and so forth. Moreover, perturbation due to the updated track degradation level after reactive intervention or machine unavailability can also be considered as well. This model further encourages the development of robust distributed optimization approaches.

Column generation-based distributed approach

The improvement of current model and addition of more constraints in the future research can increase the complexity of the problem and so it will be more difficult to obtain the global solutions. At some point, the current methods might not be able to handle the problem anymore. Thus, improvements from methodology perspective are required. A distributed approach based on column generation technique under the framework of Dantzig-Wolfe decomposition can be applied. Unlike the Lagrangian-based methods, this method guarantees feasibility. Moreover, this algorithm is also suited for coupling constrained problem.

Hierarchical distributed optimization approach

The other potential approach to be implemented is distributed hierarchical structure. This scheme facilitates different time-scales of maintenance planning and scenarios that exist in the railway industry. Moreover, the development of maintenance scheduling for different types of assets is also challenging to develop with this scheme. The clustering technique can also be added to the lower level system to categorize the track sections into multiple smaller groups. In this regard, subproblems will not only handle single track section but can also be multiple track sections at once and the size of each might be different. In this kind of scheme, the so-called cooperative DRSBK, which includes perturbations from other subproblems in the individual objective functions, can be applied to improve the performance of current distributed optimization approaches.

Appendix A

Table A-1, A-2 and Figure A-1 presented below contain initial track degradation level and degradation rates data used in case studies. For the sake of simplicity, the parameters for medium and large-scale experiments along with gradual experiment are not presented. Moreover, Figures A-2 - A-17 depict the simulations from test 1 and test 3 in small-scale experiments.

Table A-1: Parameter for the second experiment in centralized evaluation

Parameter	**Low**	**High / fast**
Initial condition X_0	0.7	1.4
Degradation rate a_1	1.018	1.036
Memory rate a_2	1.009	1.014

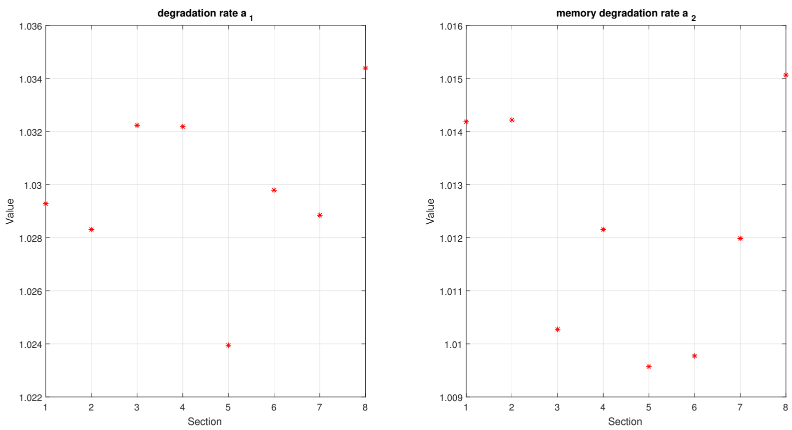

Figure A-1: Distribution of a_1 and a_2 for the first experiment in centralized evaluation

Table A-2: Parameters for small scale tests

		Test									
		1	2	3	4	5	6	7	8	9	10
Section 1	x0	1.3	0.2938	1.0419	1.3722	0.6655	0.887	1.366	1	0.9	0.9
	a1	1.04	1.0224	1.0287	1.0244	1.035	1.0388	1.0305	1.04	1.06	1.03
	a2	1.012	1.0151	1.0092	1.0124	1.0124	1.0101	1.0141	1.014	1.012	1.012
Section 2	x0	1.3	0.635	0.9292	1.5585	0.5308	0.9184	1.2932	1.1	1.1	1
	a1	1.02	1.0302	1.0362	1.0349	1.0495	1.0455	1.0432	1.02	1.05	1.04
	a2	1.01	1.0122	1.013	1.0097	1.0113	1.0137	1.0151	1.012	1.01	1.01
Section 3	x0	1.3	0.3384	0.9198	1.2333	0.3743	0.8524	1.2805	1.2	1.3	1.1
	a1	1.03	1.0339	1.0246	1.0358	1.0384	1.0307	1.0341	1.03	1.04	1.05
	a2	1.009	1.009	1.0116	1.0097	1.0108	1.0138	1.0108	1.01	1.009	1.009
Section 4	x0	1.3	0.4748	1.0588	1.3187	0.4135	0.9862	1.2782	1.3	1.5	1.2
	a1	1.035	1.0377	1.0202	1.0283	1.0359	1.0237	1.0377	1.02	1.03	1.06
	a2	1.011	1.0105	1.0116	1.0122	1.0181	1.0139	1.0155	1.008	1.011	1.011

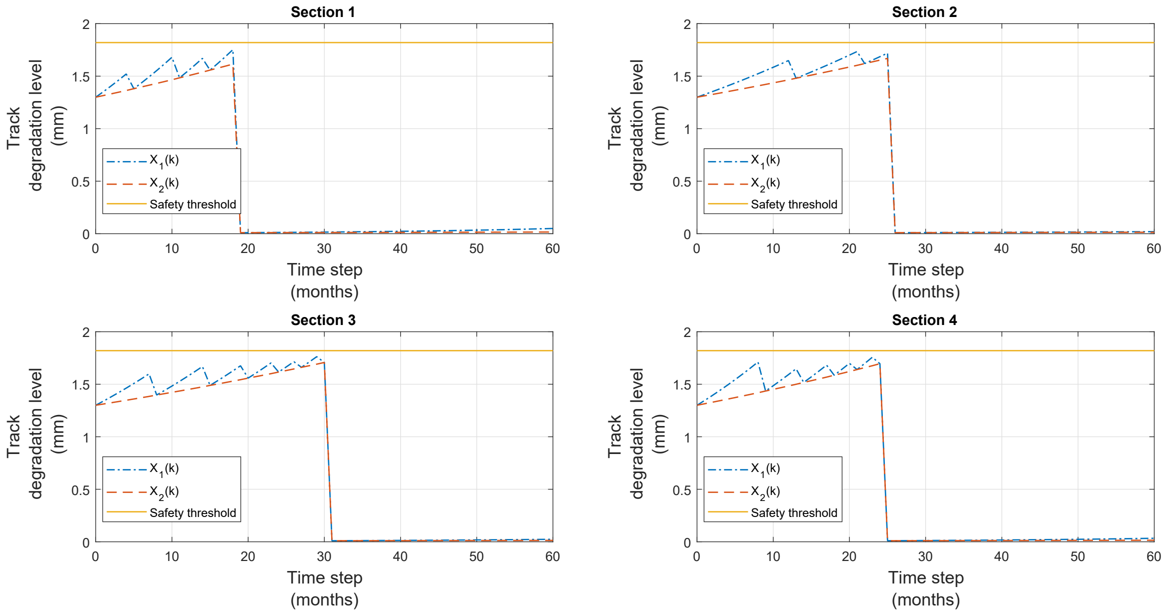

Figure A-2: Track degradation curves of the centralized in test 1

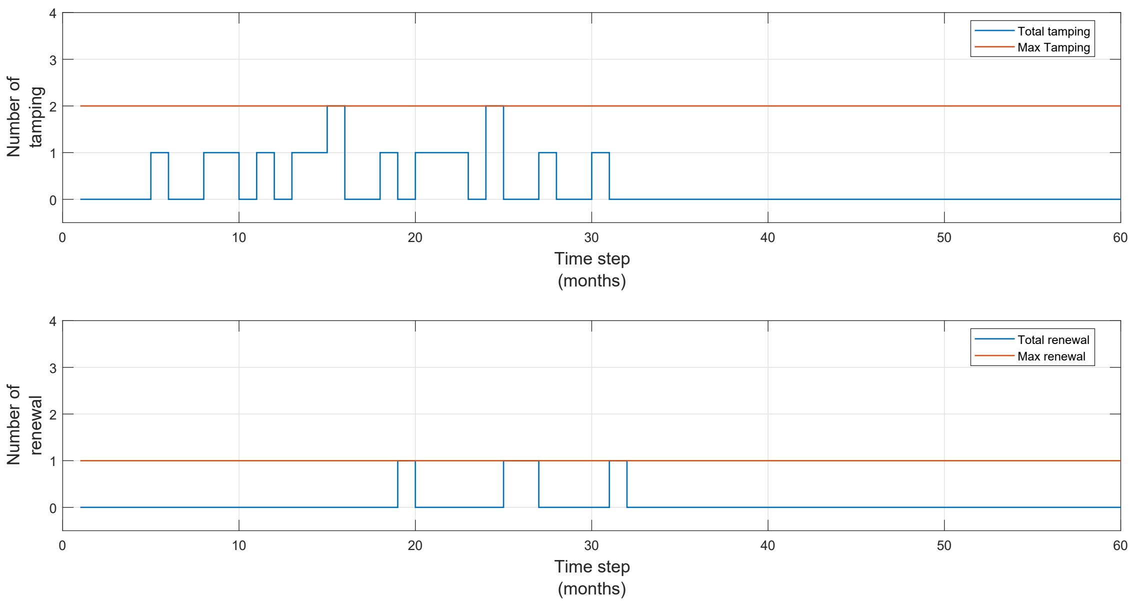

Figure A-3: Total intervention (tamping and renewal) of the centralized in test 1

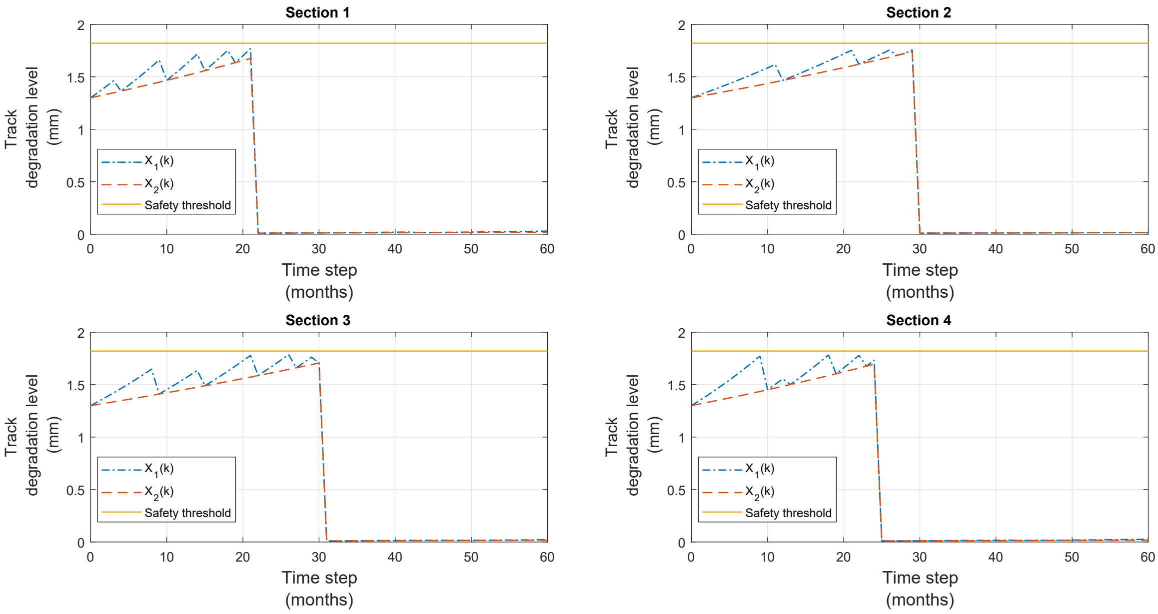

Figure A-4: Track degradation curves of the PALR in test 1

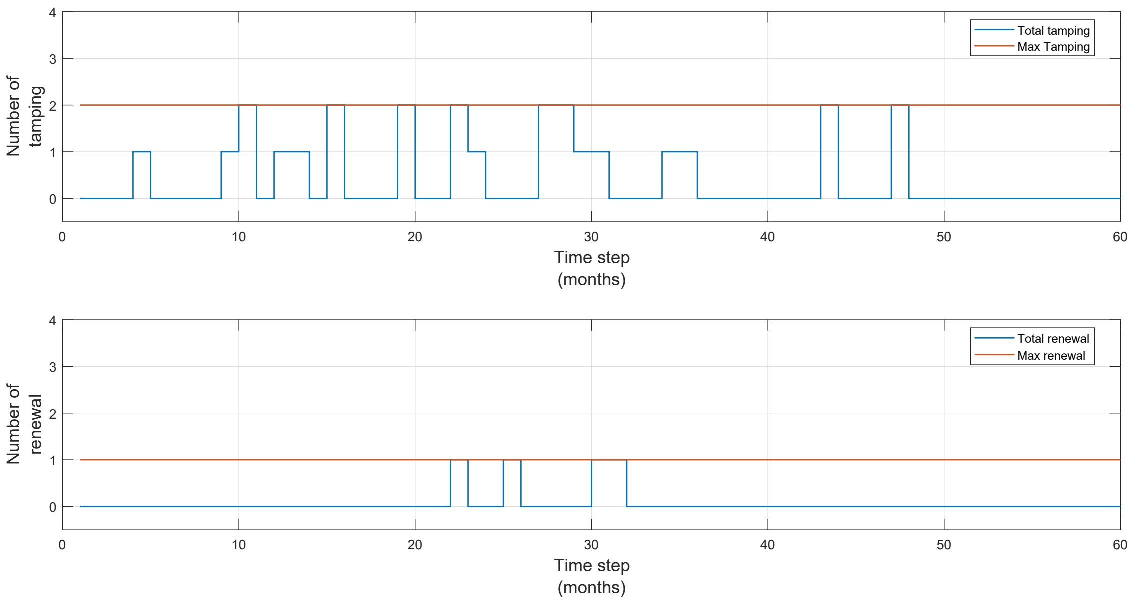

Figure A-5: Total intervention (tamping and renewal) of the PALR in test 1

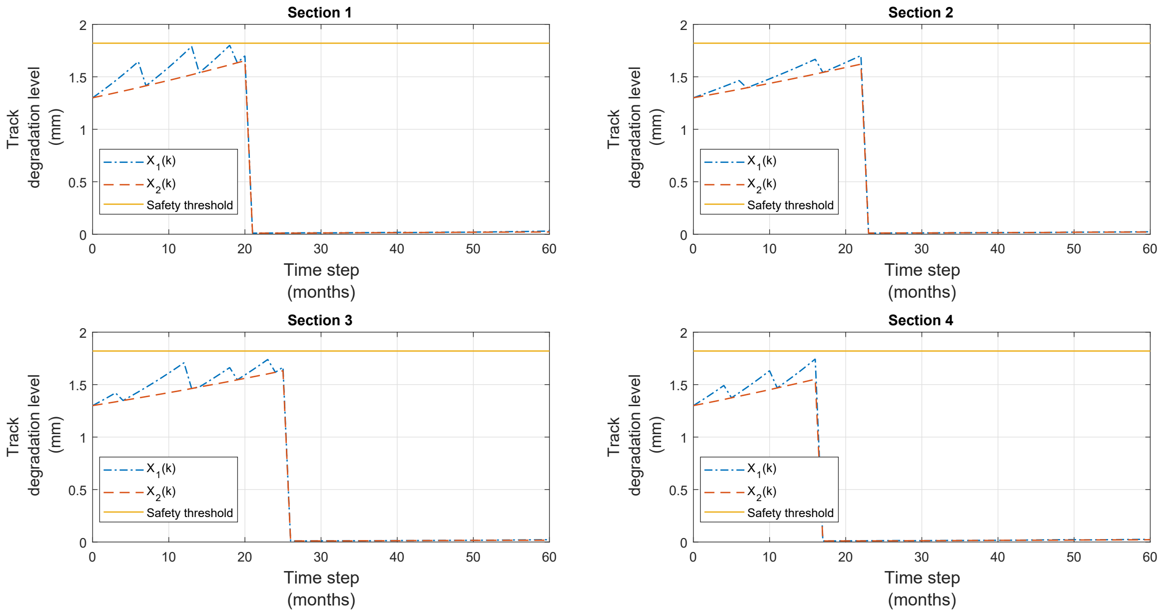

Figure A-6: Track degradation curves of the ADMM in test 1

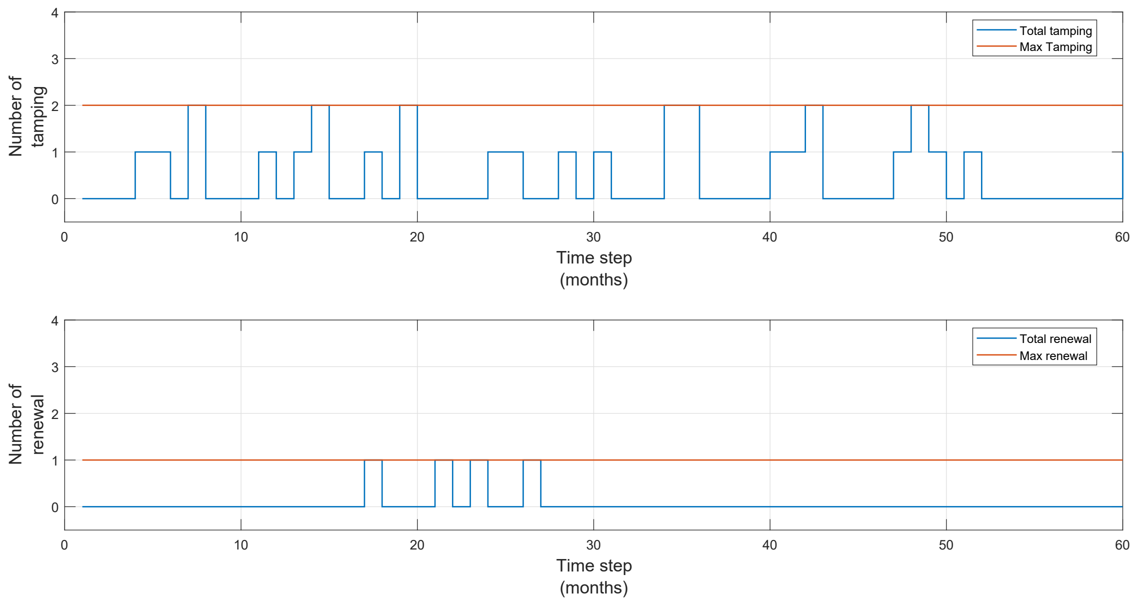

Figure A-7: Total intervention (tamping and renewal) of the ADMM in test 1

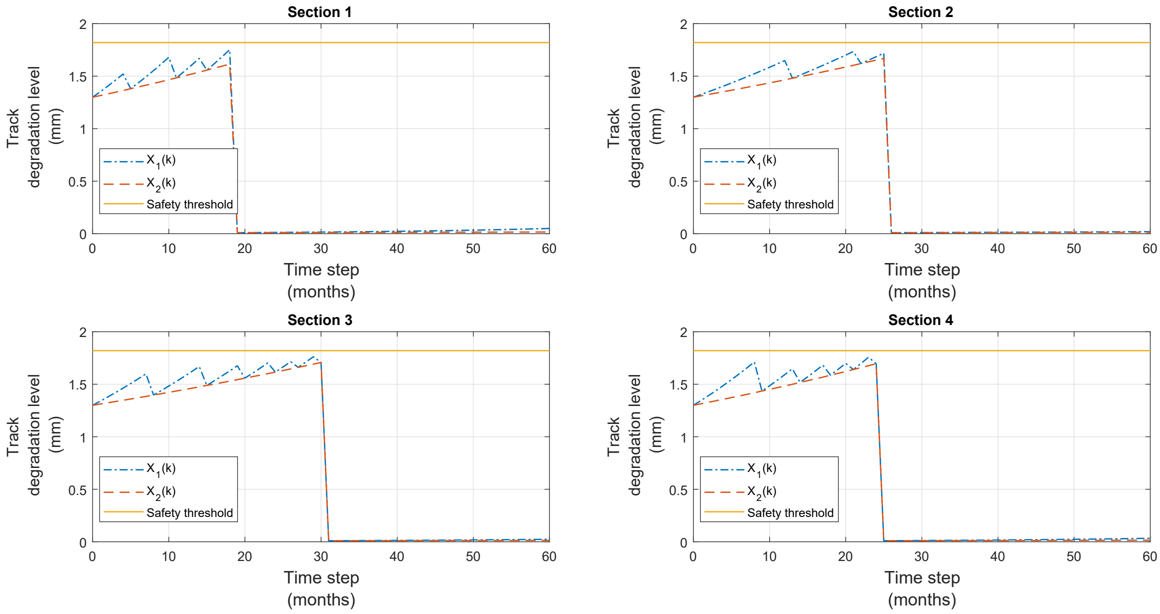

Figure A-8: Track degradation curves of the DRSBK in test 1

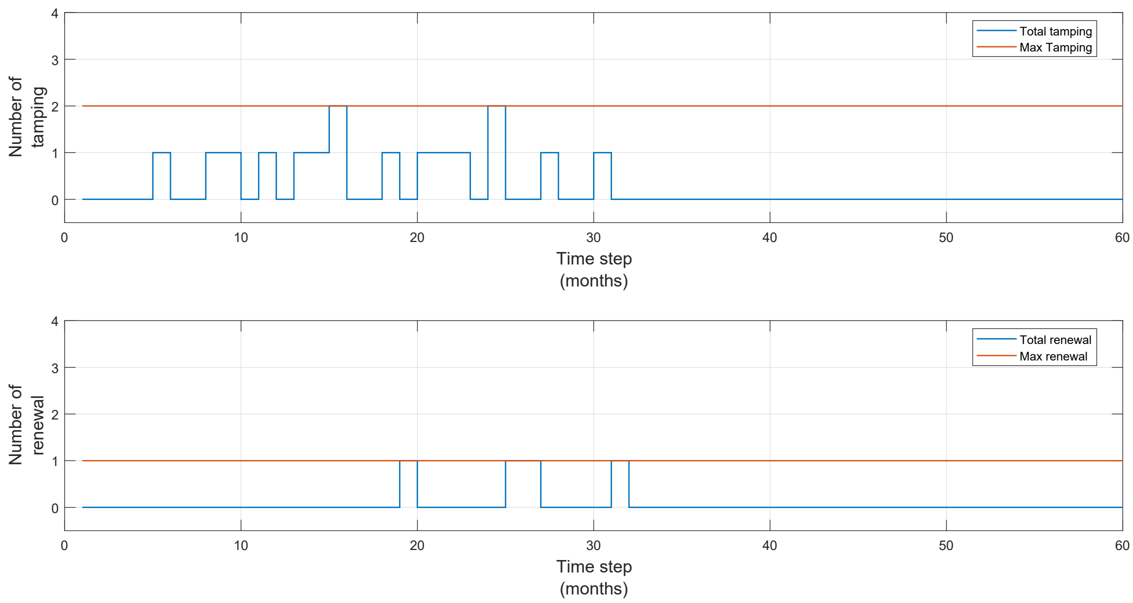

Figure A-9: Total intervention (tamping and renewal) of the DRSBK in test 1

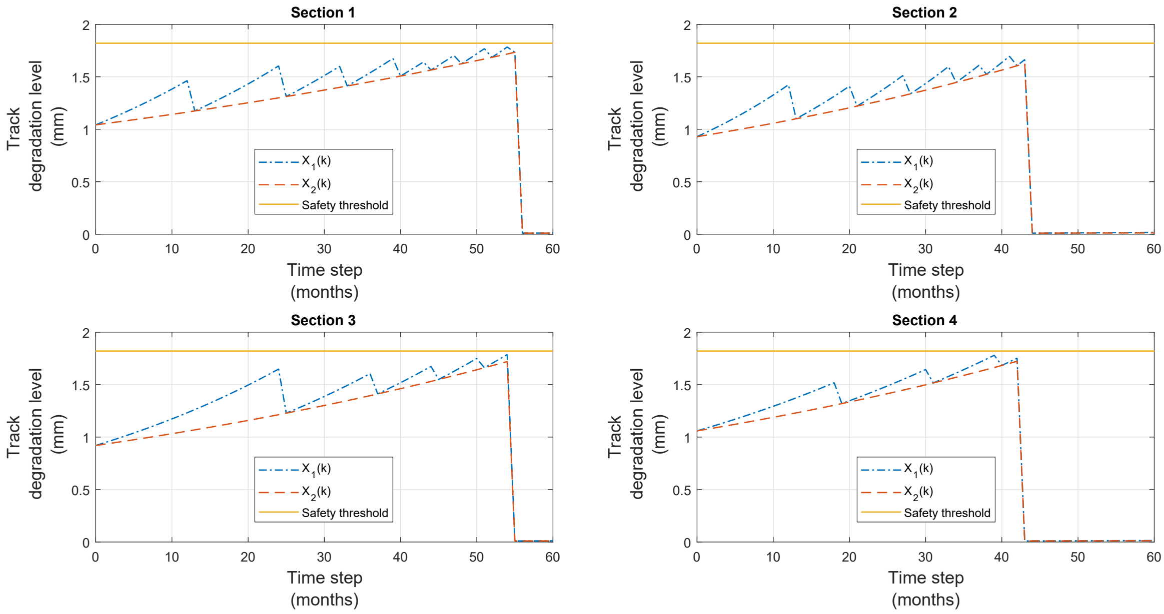

Figure A-10: Track degradation curves of the centralized in test 3

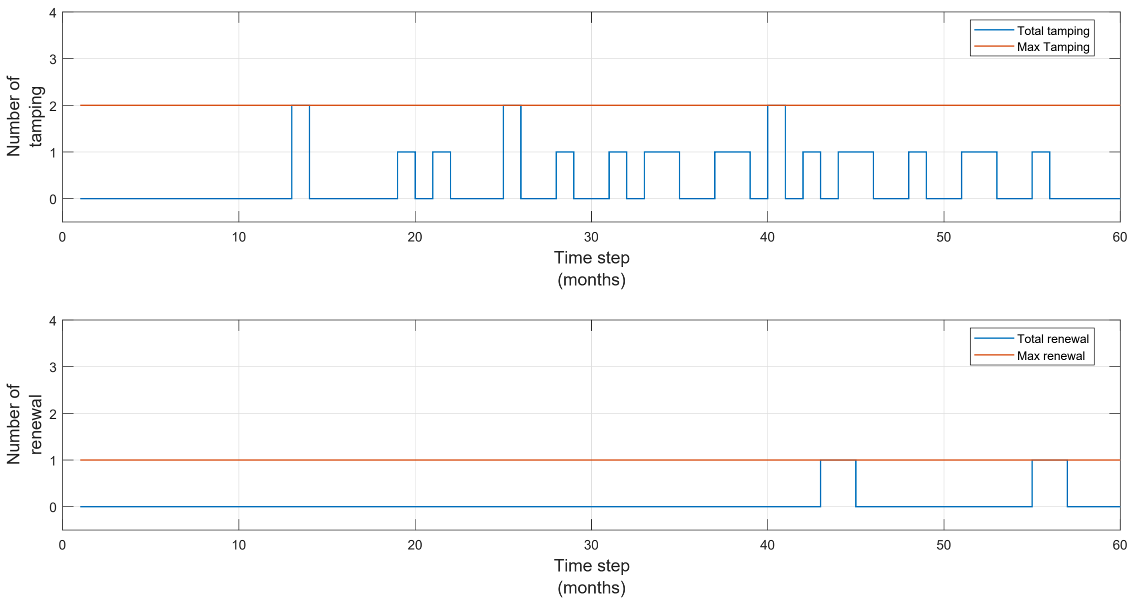

Figure A-11: Total intervention (tamping and renewal) of the centralized in test 3

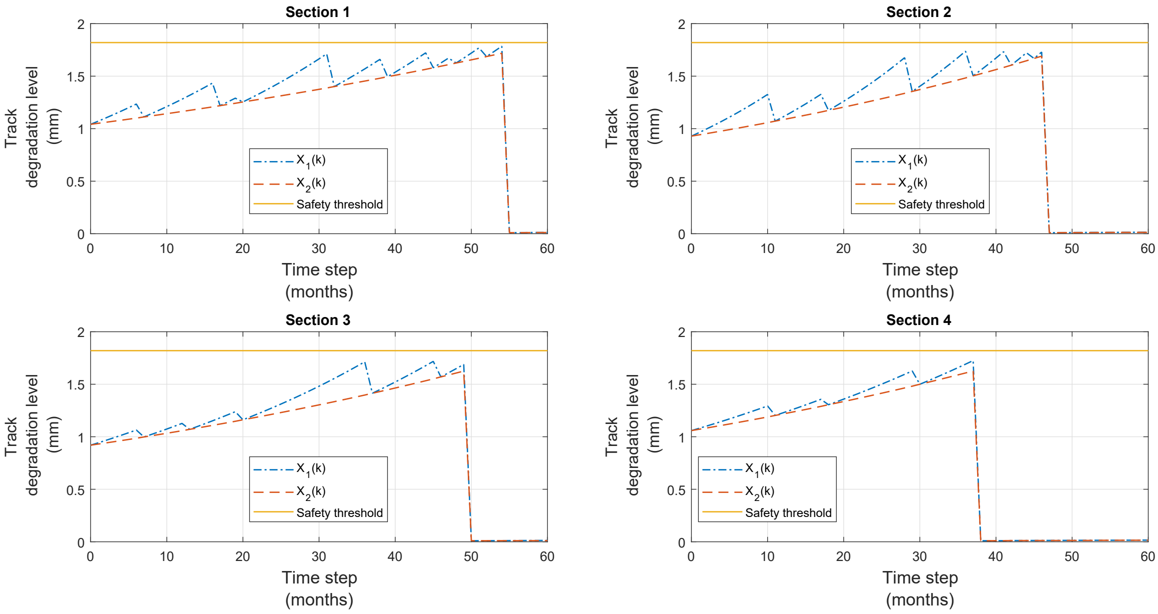

Figure A-12: Track degradation curves of the PALR in test 3

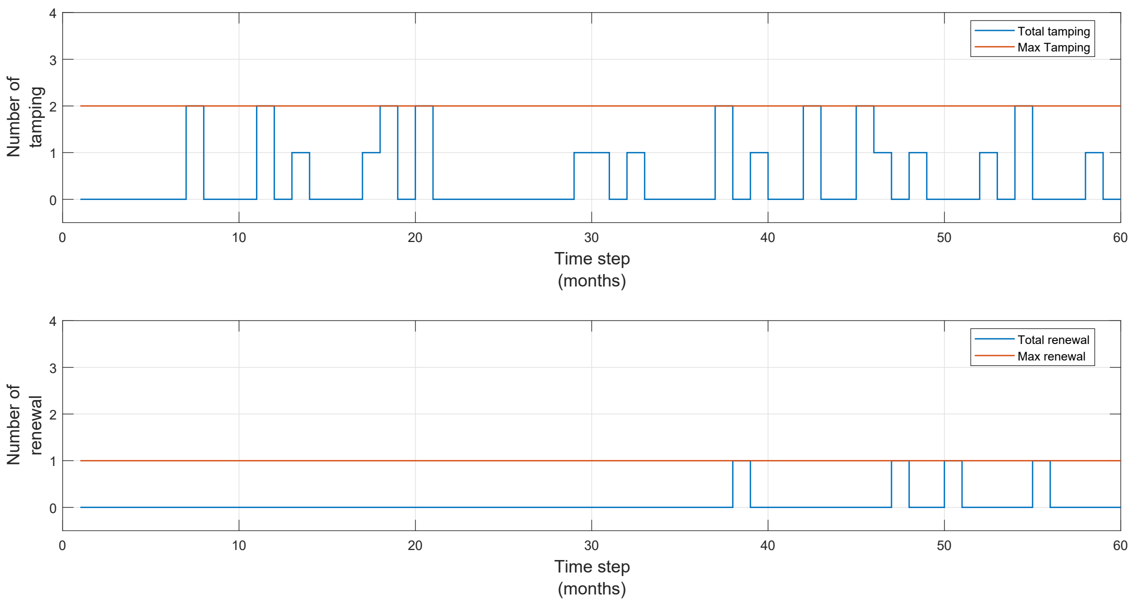

Figure A-13: Total intervention (tamping and renewal) of the PALR in test 3

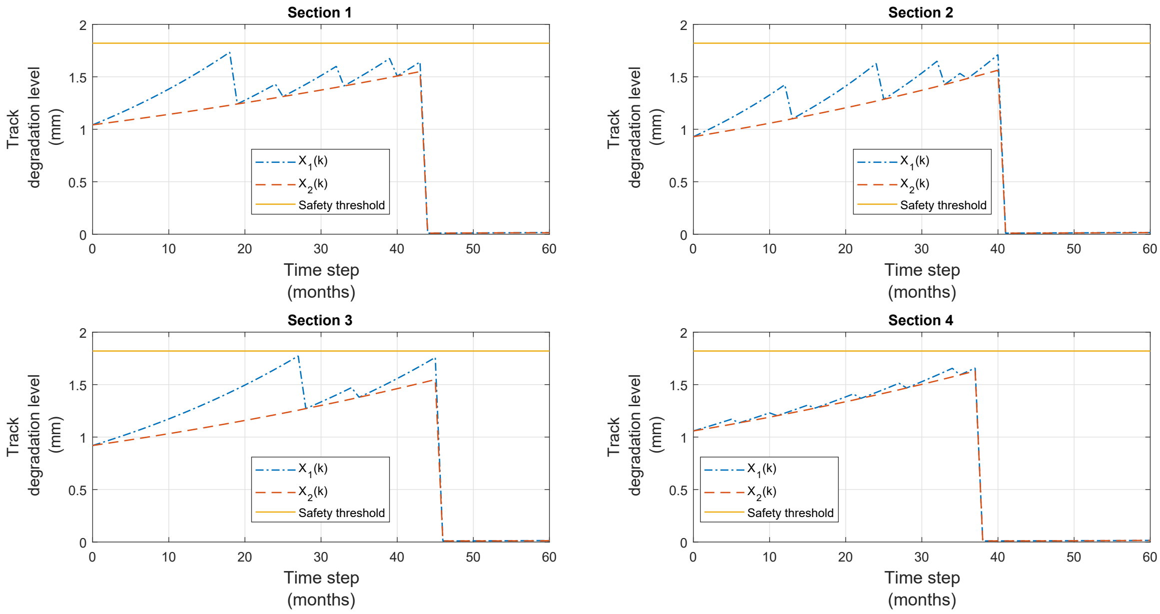

Figure A-14: Track degradation curves of the ADMM in test 3

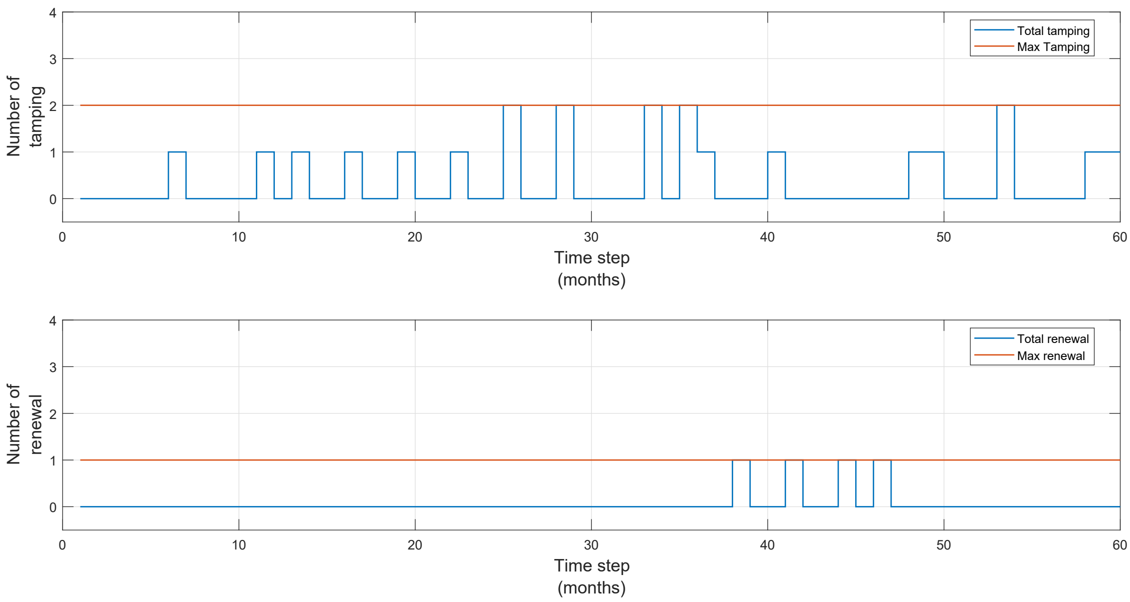

Figure A-15: Total intervention (tamping and renewal) of the ADMM in test 3

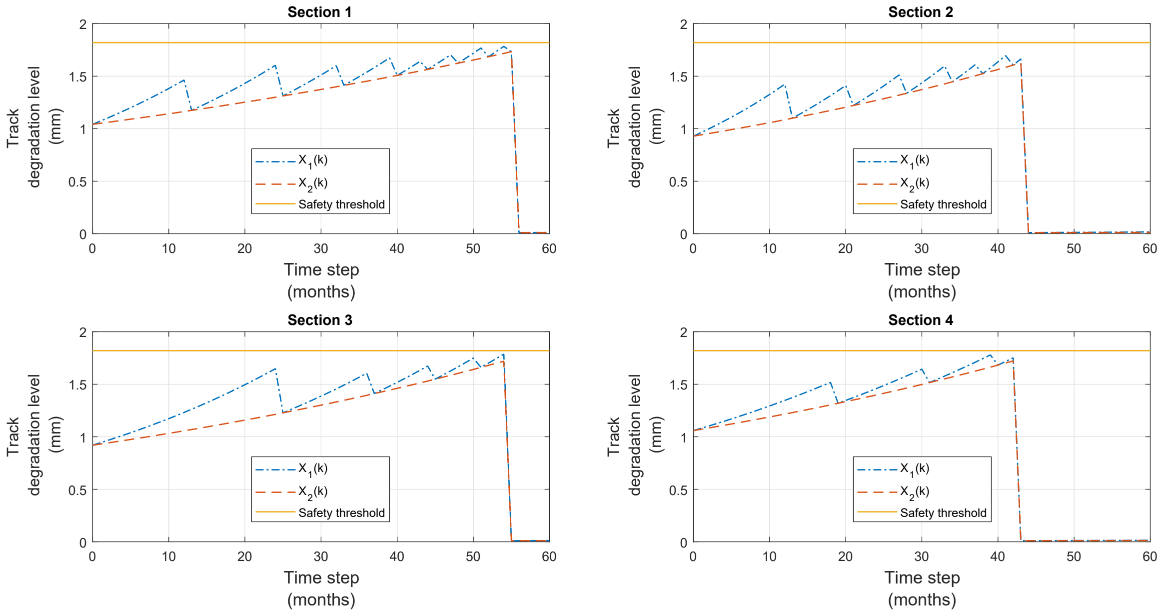

Figure A-16: Track degradation curves of the DRSBK in test 3

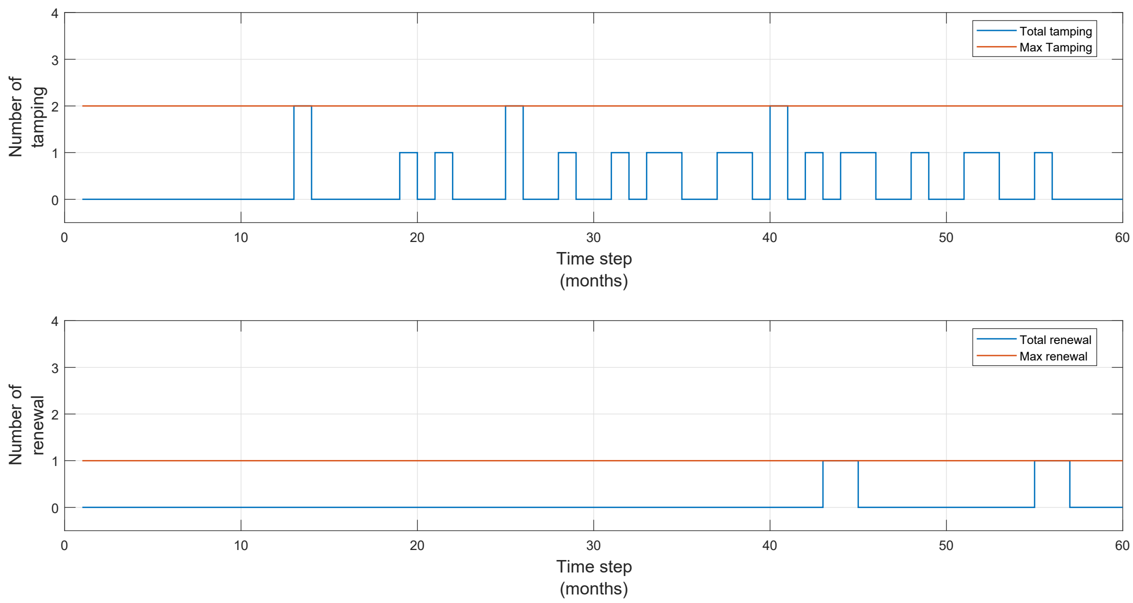

Figure A-17: Total intervention (tamping and renewal) of the DRSBK in test 3

Appendix B

Appendix B

The performance data from the ten tests for small, medium, and large-scale experiments are presented in Table B-1, B-2, and B-3, respectively.

Table B-1: Performance comparison for small-scale tests

Approach	Criterion	Test									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Centralized	**Total J(V)**	1671.12	413.21	1950.84	1550.05	593.79	747.31	1687.57	535.26	1801.01	1913.06
	Comp. time (sec.)	2.71	0.81	2.00	1.39	0.98	1.56	1.80	0.97	2.05	2.69
	Number of tamping	16	5	21	11	10	19	19	11	23	24
	Number of renewal	4	0	4	4	0	0	4	0	4	4
	Total track performance	311.12	363.21	540.84	240.05	493.76	557.31	297.53	425.27	371.01	473.06
PALR	**Total J(V)**	1694.47	420.86	1970.34	1636.33	707.30	1061.23	1736.44	596.41	1769.62	1956.59
	Comp. time (sec.)	24.27	9.10	25.18	13.29	45.27	24.81	15.18	40.99	14.67	10.14
	Number of tamping	22	5	26	22	24	22	25	19	27	31
	Number of renewal	4	0	4	4	0	1	4	0	4	4
	Total track performance	274.47	370.86	510.34	216.33	467.30	541.23	286.44	406.41	299.63	446.59
	Normalized J(V)	-1.40%	-1.85%	-1.00%	-5.57%	-19.12%	-42.01%	-2.90%	-11.42%	1.74%	-2.28%
ADMM	**Total J(V)**	1788.78	452.06	1876.37	1628.64	651.58	1092.99	1757.86	609.75	1782.53	1980.87
	Comp. time (sec.)	5.47	12.44	7.17	7.07	5.88	7.40	6.24	5.12	6.00	6.39
	Number of tamping	36	10	23	26	16	25	30	20	28	36
	Number of renewal	4	0	4	4	0	1	4	0	4	4
	Total track performance	228.78	352.06	446.37	168.64	491.58	542.99	257.86	409.75	302.53	420.87
	Normalized J(V)	-7.04%	-9.40%	3.82%	-5.07%	-9.74%	-46.26%	-4.17%	-13.92%	1.03%	-3.54%
DRSBK	**Total J(V)**	1671.12	413.21	1950.84	1560.27	593.79	747.42	1687.57	535.66	1801.01	1962.64
	Comp. time (sec.)	3.24	2.51	3.32	3.74	2.80	3.27	3.99	2.36	4.04	4.71
	Number of tamping	16	5	21	12	10	19	19	11	23	27
	Number of renewal	4	0	4	4	0	0	4	0	4	4
	Total track performance	311.12	363.21	540.84	240.27	493.76	557.42	297.57	425.66	371.01	492.64
	Normalized J(V)	0.00%	0.00%	0.00%	-0.66%	0.00%	-0.01%	0.00%	-0.07%	0.00%	-2.59%

Table B-2: Performance comparison for medium-scale tests

Approach	Criterion	Test									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Centralized	**Total J(V)**	4798.86	15000.9	20372.51	11164.69	20723.48	8539.25	13890.93	15738.79	4570.44	4930.44
	Comp. time (sec.)	33.27	151.28	145.35	52.88	284.90	38.41	234.03	72.89	34.05	32.32
	Number of tamping	52	221	180	196	285	163	265	217	94	53
	Number of renewal	0	21	47	11	40	3	17	24	0	0
	Total track performance	4278.86	6490.90	4472.51	5904.69	5873.48	6009.25	6140.93	6368.79	3630.44	4400.44
PALR	**Total J(V)**	16907.37	16254.85	17535.27	19150.68	19170.92	19783.79	19585.68	17673.33	21730.27	17633.75
	Comp. time (sec.)	298.88	149.57	120.03	144.28	99.35	130.73	119.85	116.08	151.11	124.97
	Number of tamping	377	196	266	295	237	324	417	226	448	219
	Number of renewal	22	26	28	33	35	34	31	30	37	29
	Total track performance	6537.35	6494.85	6475.26	6300.67	6300.91	6343.78	6115.68	6413.32	6150.26	6743.73
	Normalized J(V)	-252.32%	-8.36%	13.93%	-71.53%	7.49%	-131.68%	-41.00%	-12.29%	-375.45%	-257.65%
ADMM	**Total J(V)**	6050.95	18432.76	19485.36	13074.2	21851.25	11405.99	16062.27	17975.32	6570.71	7254.24
	Comp. time (sec.)	125.70	49.02	52.04	44.10	46.95	44.43	52.58	46.82	43.02	42.28
	Number of tamping	201	327	86	335	321	336	340	313	341	335
	Number of renewal	0	31	50	14	45	8	23	30	0	0
	Total track performance	4040.95	5862.76	3625.36	5524.20	5141.25	5645.99	5762.27	5845.32	3160.71	3904.24
	Normalized J(V)	-26.09%	-22.88%	4.35%	-17.10%	-5.44%	-33.57%	-15.63%	-14.21%	-43.77%	-47.13%
DRSBK	**Total J(V)**	4798.86	15010.74	20372.68	11174.39	20723.52	8538.84	13900.85	15738.78	4570.44	4930.44
	Comp. time (sec.)	29.33	36.13	40.19	32.79	46.77	29.56	36.99	34.90	32.21	28.88
	Number of tamping	52	222	180	197	285	163	266	217	94	53
	Number of renewal	0	21	47	11	40	3	17	24	0	0
	Total track performance	4278.86	6490.74	4472.68	5904.39	5873.52	6008.84	6140.85	6368.78	3630.44	4400.44
	Normalized J(V)	0.00%	-0.07%	0.00%	-0.09%	0.00%	0.00%	-0.07%	0.00%	0.00%	0.00%

Table B-3: Performance comparison for large-scale tests

Approach	Criterion	Test									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Centralized	**Total J(V)**	18405.94	49237.36	62972.27	25855.37	42935.24	34365.01	32995.82	31585.9	33307.13	30341.79
	Comp. time (sec.)	341.42	1523.59	1871.33	651.47	1105.33	941.34	501.43	415.76	942.20	424.50
	Number of tamping	242	641	527	448	597	507	487	4.62E+02	514	5.05E+02
	Number of renewal	2	77	150	12	57	33	30	26	29	19
	Total track performance	15385.94	19727.36	12702.27	17775.37	19865.24	19395.01	19125.82	19165.9	19467.13	19591.79
PALR	**Total J(V)**	55458.73	46622.42	62650.87	58006.38	48856.31	48998.48	52609.6	50343.33	59941.88	59207.04
	Comp. time (sec.)	410.4266	607.2614	527.9807	659.9002	412.59	318.7711	393.1536	410.5224	358.6506	394.8306
	Number of tamping	974	450	423	693	889	480	758	652	1059	788
	Number of renewal	89	78	138	114	67	83	85	81	100	107
	Total track performance	19018.72	18722.41	17020.86	16876.37	19866.3	19298.48	19529.6	19523.31	19351.86	19227.04
	Normalized J(V)	-201.31%	5.31%	0.51%	-124.35%	-13.79%	-42.58%	-59.44%	-59.39%	-79.97%	-95.13%
ADMM	**Total J(V)**	25750.05	60292.71	58975.83	39154.84	52667.13	44103.7	41689.69	42147.38	44389.7	42571.84
	Comp. time (sec.)	140.646	186.3321	180.8575	193.6696	146.82	145.798	134.5131	144.8122	143.06	151.1301
	Number of tamping	985	891	866	944	937	959	951	958	955	957
	Number of renewal	6	118	150	47	84	55	48	49	56	49
	Total track performance	14100.05	15982.72	5315.841	15614.84	18097.13	18013.7	17779.7	17867.38	18039.7	18301.85
	Normalized J(V)	-39.90%	-22.45%	6.35%	-51.44%	-22.67%	-28.34%	-26.35%	-33.44%	-33.27%	-40.31%
DRSBK	**Total J(V)**	18415.79	49237.35	62982.29	25856.03	42935.3	34364.93	32995.8	31585.9	33307.1	30370.79
	Comp. time (sec.)	86.70398	126.9894	130.8357	110.5493	110.61	95.30206	87.28103	98.8019	102.7722	96.84716
	Number of tamping	243	641	528	448	597	507	487	462	514	508
	Number of renewal	2	77	150	12	57	33	30	26	29	19
	Total track performance	15385.79	19727.35	12702.29	17776.03	19865.3	19394.93	19125.8	19165.9	19467.1	19590.79
	Normalized J(V)	-0.05%	0.00%	-0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	-0.10%

Appendix C

The draft of the paper written from this thesis is presented in this Appendix. The paper template used is IEEE journal.

Distributed Optimization for Railway Track Maintenance Operations Planning

Author 1, *Member, IEEE,* Author 2, *Fellow, OSA,* and Author 3, *Life Fellow, IEEE*

***Abstract*—In this paper, distributed optimization approaches are developed for maintenance operations planning of large-scale railway track network. A Mixed Integer Linear programming (MILP) problem of the maintenance planning is formulated and solved with three different distributed optimization schemes: Parallel Augmented Lagrangian Relaxation (PALR), Alternating Direction Method of Multipliers (ADMM), and Distributed Robust Safe But Knowledgeable (DRSBK). Extension techniques to handle the non-convex nature of the optimization problem and improve the solution quality are implemented. In the computational experiments of large-scale test instances, it is shown that DRSBK can outperform the other distributed approaches, by providing the closest-to-optimum solution while having the least computation time.**

Keywords*—track maintenance planning, railway engineering, mixed-integer programming, distributed optimization.

I. Introduction

Railway infrastructure consists of different assets, comprising railway tracks, electrical systems, signaling devices, switches, stations, and so forth. All assets are interconnected and work together. Among those components, ballast is a vital component as it is used to support the track level and alignment at the designated positions [3]. Due to regular usage of tracks, ballast suffers quality degradation over time. In order to avoid unexpected dangerous situations, ballast must be maintained so that its performance could meet the technical and safety criterion. Maintenance plan is usually formulated by infrastructure managers to decide the maintenance timing, location, and type of intervention.

Since there are various aspects to consider, railway infrastructure managers nowadays face a number of challenges when designing a maintenance schedule. To obtain a cost-effective and safe schedule, the use of decision support system is necessary. Such support system usually relies on optimization methods. There are a number of studies that apply optimization methods to track maintenance planning [20]. A railway track is typically very long, consisting of a long number of track sections. A single track section is defined as 200 m and each of it has independent dynamics, such studies case hence have a large number of decision variables. Indeed, the optimization problem of track maintenance operations planning can be considered as a large-scale problem. Aside from different optimization problems that they consider, all of them use a centralized optimization approach, where all information processing and computation of all decision variables are conducted in a single centralized node. From the computational perspective, the centralized structure is unattractive as the number of variables is increased linearly, but the computation burden might be heavier exponentially [8, 15].

A number of studies states that they were interested to handle large-scale maintenance scheduling problem [2, 14]. Even though these studies mentioned the large-scale nature of their case studies, they still utilize the traditional centralized scheme. Fortunately, a couple of studies have attempted to use non-centralized schemes in maintenance optimization. Su et al. [18] proposed a multi-level optimization for rail maintenance, which is motivated by different time sampling in maintenance planning. The proposed approach is divided into the high, middle and low level optimizer, each with different tasks. This kind of hierarchical approach might be able to reduce computation cost. However, hierarchical scheme is limited by the problem structures. In [21], they deal with the optimization of multiple assets renewal. The combination of decentralized and hierarchical scheme is applied. However, this research lacks investigation on computational efforts. Thus the main goal of this paper is to deal with the computational issue from the maintenance operations for large-scale railway track networks. To achieve that goal, various distributed optimization methods developed since last decades are evaluated.

It is common to have discrete decisions when deciding, for instance, whether or not to perform maintenance and which type of intervention to perform. Due to this nature, it is not surprising that scheduling tasks contain to integer decision variables. In such cases, the optimization problems can be formulated in the resulting problem as Mixed-Integer Programming (MIP). Lagrangian-based decomposition methods are one of the basic techniques for applying distributed optimization [23]. However their standard algorithms are intended only for dealing with convex problems. Previous studies have developed various extension methods for the Lagrangian-based approaches to deal with MIP problems. They made use of different decomposition methods. First, the dual decomposition method is used in [7, 22, 17]. Moreover, different modification to dual decomposition is used by [5, 23] by adding quadratic terms of the respective decision variables to regularize the objective function, which leads to the notion of augmented Lagrangian or Alternating Direction Method of Multipliers (ADMM).

The extension of ADMM algorithm to handle mixed-integer problem are also available. In this regard, ADMM can be seen as a heuristic method [1]. In [6, 16], a continuous relaxation technique of binary variables is used. The resulting

A. 1 is with the Delft Center for Systems and Control, Delft University of Technology, Delft, the Netherlands. e-mail:-

A. 2 and A. 3 are with Section of Railway Engineering, Delft University of Technology, Delft, the Netherlands.

Manuscript received April 19, 2005; revised January 11, 2007.

solution is used as a bound to the objective function and to warm start the optimization of the MIP problem. Apart from Lagrangian-based methods, a distributed algorithm called Distributed Robust Safe But Knowledgeable (DRSBK) [10] also has been applied to an MIP problem with hard non-convex coupling constraints. This algorithm applies coupling constraint tightening approach when solving one subproblem. Hence, all subproblems are solved sequentially. In this way, the global feasibility of the solution can be guaranteed. The drawback, however, is that the solution might be suboptimal. To the best of the authors knowledge, it is interesting to note that compared to the Lagrangian-based, this approach is not much used in the distributed MIP domains.

To address the large problem size in track maintenance problem, this paper develops distributed optimization approaches. The focus lies in reducing computation time and increase the scalability of the optimization method to handle large-scale instances, while maintaining good solutions. To that end, two Lagrangian-based decomposition approaches along with a constraint tightening distributed optimization approach are implemented to solve the large-scale track maintenance operations planning.

This paper is organized as follows: the optimization problem is firstly described in Section II. Afterward, Section III addresses the development of distributed optimization approaches. Case studies are discussed in Section IV. Finally, Section V provides the conclusions and future work of this research.

II. Problem definition

First, the optimization problem of the track maintenance problem is explained. The proposed formulation is based on [19], which is adapted for the case of large-scale track maintenance operations planning.

There are a number of characteristics and limitation in track maintenance operations planning that can be incorporated into the optimization problem. First, the dynamics of ballast degradation representing the track performance level x_i^1 and offset recorder x_i^2, can be mathematically modeled as follows,

$$\begin{aligned} x_{1,i}(k+1) &= a_{1,i}x_{1,i}(k) + f_{1,i}(x_i(k), u_i(k)) \\ x_{2,i}(k+1) &= a_{2,i}x_{2,i}(k) + f_{2,i}(x_i(k), u_i(k)) \end{aligned} \tag{1}$$

with $a_{1,i}$ is the track degradation rate and $a_{2,i}$ is the offset memory rate. Functions $f_{1,i}$ and $f_{2,i}$ are discontinuous functions. This model is slightly different than in [19], in the sense that, the degradation memory variable is set to have a multiplier constant $a_{2,i}$, which enables exponential degradation of the second state. Furthermore, this system has three types of inputs, as depicted in Table I.

TABLE I: System input

Input $u_i(k)$	Decision
1	Doing nothing
2	Tamping
3	Renewal

Due to the multiplications by discrete variables, this system is basically non-linear. Therefore, one way to linearize the system is by using the so called Mixed Logical Dynamical (MLD) framework. The representation of the discontinuous functions along with transformation of this system into the MLD form are provided in [19]. The resulting state-space model is of the form:

$$x_i(k+1) = A_i x_i(k) + B_i V_i(k) \tag{2}$$

where $V_i(k)$ contains the binary and auxiliary variables. Meanwhile, the bounding constraints for the auxiliary variables can be found in [19].

Renewal operation in practice is allowed once the degradation level has been considered high. In a long-term planning, it is much more costly to perform renewal when tamping operation is still effective. To prevent unnecessary renewal at early stage, this constraint is added:

$$x_{2,i}(k) - h_{\mathrm{r}} \geq (r_i - 1)h_{\max} \tag{3}$$

$$r_i - \delta_1(k) \geq 0 \tag{4}$$

where r_i is the binary indicator for the switching between stages. This constraint is applied for every track section.

In any condition, the decision support system must be able to prevent the degradation curve exceeds the safety limit. Therefore, the following constraint ensures the track degradation level always in an acceptable condition:

$$h_{\min} \leq x_i(k) < h_{\max} \tag{5}$$

this constraint is also applied to the offset memory state. The other limitation is that the maintenance budget for such complex railway system is limited [12]. The following constraints make sure that the number of interventions, both tamping and renewal, over the prediction horizon is restricted by the thresholds:

$$\sum_{k=1}^{T} \delta_{2,i}(k) \leq g_{\mathrm{t}} \tag{6}$$

$$\sum_{k=1}^{T} \delta_{1,i}(k) \leq g_{\mathrm{r}} \tag{7}$$

where g_{t} and g_{r} are maximum numbers of allowed tamping and renewal operations over the prediction horizon, respectively.

The previous defined constraints can be categorized as individual constraints. Alongside them, coupling constraints which affect multiple track sections, exist in maintenance planning. One of them is track closure time due to maintenance operations. Typically, the time slot is less than 7 hours. The operation is only allowed during night time, at weekends [3]. This constraints applies for both tamping and renewal, respectively. Based on [18], [4], this constraint can be written as follows:

$$\sum_{i=1}^{N_t(k)} t_{t1}\delta_{2,i}(k) + \sum_{j=1}^{N-N_t(k)} t_{t2}\delta_{2,j}(k) < t_{\max} \quad (8)$$

$$\sum_{i=1}^{N_r(k)} t_{r1}\delta_{1,i}(k) + \sum_{j=1}^{N-N_r(k)} t_{r2}\delta_{1,i}(k) < t_{\max} \quad (9)$$

where t_{t1} and t_{t2} are maintenance operation and traveling times for tamping, respectively. The same representations also hold for t_{r1} and t_{r2} for renewal. $N_t(k)$ and $N_r(k)$ is the total number of track sections that receive tamping and renewal, respectively, at time step k. The maintenance time already includes the intervention and machine switching time. The renewal closure time constraint is also expressed in the same way, with different time value. Moreover, it is assumed that the machines move in one direction at each time step, from a starting point toward an endpoint at the other end of the track. Hence, the position of the maintained track section does not matter in the operation time. Tamping and renewal are performed at different time slots. In addition, since such constraints affect the maintenance schedule across multiple track sections, they can be considered as coupling constraints.

The objective function used in this paper is defined based on [19], but with linear rather than quadratic expression. The optimal state variables and decision variables for all track sections over the prediction horizon, $\bar{X}$ and $\bar{V}$, can be obtained through solving the following equation:

$$J(\bar{X}, \bar{V}) = \sum_{i=1}^{N}\sum_{k=1}^{T} Qx_i(k) + \lambda RV_i(k) \quad (10)$$

where the decision variable for one section at a time step $V_i(k) = \begin{bmatrix}\delta_i(k) & z_i(k)\end{bmatrix}^T$ along with Q and R with only positive entries matrices with appropriate dimension. The state variables can be substituted in the same way as in [19], leaving the input $V_i(k)$ as the decision variables. Finally, The optimization problem can be compactly written in following forms:

$$\underset{\bar{V}}{\text{minimize}} \quad J(\bar{V}) = \sum_{i=1}^{N} J_{\text{ind}}(\tilde{V}_i) \quad (11)$$

$$\text{subject to} \quad E\bar{V} \leq g_{\text{ind}}$$
$$\sum_{i=1}^{N} F_i\tilde{V}_i \leq g_{\text{coup}} \quad (12)$$

where on one hand, E and G_{ind} are the paremeter matrix and right-hand side vector associated with all individual constraints, respectively. On the other hand, F_i and G_{coup} are parameter matrix and right-hand side vector associated with coupling constraints. This separation aims at preparing for the development of distributed approaches.

III. Distributed optimization

Three distributed optimization approaches are discussed in this work: Parallel Augmented Lagrangian Relaxation (PALR), Alternating Direction Method of Multipliers (ADMM), and Distributed Robust Safe But Knowledgeable (DRSBK). The first two are iterative methods based on Lagrangian duality theory, which are usually used to solve convex smooth problems [23] [1]. It is worthy to mention that the proposed optimization problem is non-convex. Since the basic form of those decomposition methods could not guarantee the zero duality gap for such problem or the global solution feasibility. Thus, extension methods to solve this issue will also be discussed. Moreover, DRSBK is a non-iterative method which is originally designed to deal with MILP problem by exploiting constraint tightening techniques [10].

A. *Parallel augmented lagrangian relaxation*

In order to cope with the requirement for implementing the Parallel Augmented Lagrangian relaxation (PALR), the centralized problem in equation (11) has to be transformed into augmented Lagrangian form [23]. Prior to that, the other requirement in order to use augmented Lagrangian-like method is that any inequality coupling constraint in the proposed problem must be converted into the equality form [16]. variable $\bar{S}$ is defined as slack variables for the couplings, for both tamping and renewal, over the prediction horizon. The augmented Lagrangian can be written as follows:

$$\begin{aligned} L_{\text{PALR}}(\bar{V}, \bar{S}, \gamma) = \sum_{i=1}^{N} J_{\text{ind}}(\tilde{V}_i) + \sum_{k=1}^{T} J(\tilde{S}) + \\ \gamma(\sum_{i=1}^{N} F_i\tilde{V}_i + \sum_{k=1}^{T} F_s s(k) - g_{\text{coup}}) + \\ \frac{\rho}{2}|| \sum_{i=1}^{N} F_i\tilde{V}_i + \sum_{k=1}^{T} F_s s(k) - g_{\text{coup}}||_2^2 \end{aligned} \quad (13)$$

$$\text{subject to} \quad E\bar{V} \leq g_{\text{ind}} \quad (14)$$

where F_i and F_s are the parameter matrices of coupling constraints for the input and slack variables. Moreover, $s(k)$ is slack variable for time step k. For simplicity, the dual variables γ is devoted to both tamping and renewal closure times over the prediction horizon. Furthermore, the Lagrangian equation for the dual problem can be written as:

$$q(\gamma) = \inf_{\bar{V}, \bar{S}} \left(L(\bar{V}, \bar{S}, \gamma)| \sum_{i=1}^{N} F_i\tilde{V}_i + \sum_{k=1}^{T} F_s s(k) - t_{\max} \right) \quad (15)$$

which an be interpreted as a maximization problem of dual variables [13]:

$$\underset{\gamma}{\text{maximize}} \quad q(\gamma)$$
$$\text{subject to} \quad \gamma \geq 0$$

In each iteration, each subproblem runs in parallel, meaning that this problem uses the results from the last iterations. Once all subproblems have been solved, the results are collected by a coordinator to be included in the update of the dual variable. The existence of the coordinator also implies that one dual variable γ is used to determine the common price for all subproblems.

B. Alternating direction method of multipliers

Basically, Alternating Direction Method of Multipliers (ADMM) shares the similar augmented Lagrangian equation than PALR. The difference lies in the way of decomposing the quadratic term. Instead of linearizing it, ADMM uses the so-called alternating technique. This technique enables the separation of the quadratic terms to be determined individually by fixing the decisions coming from the other subproblems [11]. This also implies that the approach runs in sequence. In this way, ADMM can exploit the latest decisions from the other subproblems. The unscaled form of ADMM, as provided in [1], is chosen to be implemented.

C. Extension for the lagrangian-based methods

To deal with the proposed MILP problem, some modifications of the original PALR and ADMM approaches are required. Since the problem has non-convex non-smooth coupling constraints, the subgradient dual update might be unable to converge or even drive to feasible regions. One way to solve this problem is by applying continuous relaxation of the binary decision variables, such that the MILP becomes a less complex linear programming [16, 6]. On top of that, the generated objective function from solving LP can be used as a lower bound for the MILP optimization in the next step and the decision variables can be used as a warm start vector.

D. Stopping criterion

In the non-convex optimization problem, the convergence of primal residual cannot be guaranteed. Therefore, rather than observing the primal residuals until it convergences [1], the primal residuals are only checked whether they have reached the feasibility condition. In this sense, an input feasibility checker is added in the iteration [16]. Additionally, the best objective value in each iteration is also checked, after it is guaranteed that it is feasible. Simple terminating technique is implemented. The complete algorithm containing the extension technique and stopping criterion is shown in Algorithm 1.

E. Distributed robust safe but knowledgeable

The last algorithm implemented in this research is Distributed Robust Safe But Knowledgeable (DRSBK) approach that is originally used in [10] to implement optimization for multi-vehicle or agent coordination. This algorithm is originally devoted to the non-convex MILP problem. The concept is as follows: instead of putting the coupling constraints into the Lagrangian form objective function, this algorithm applies tightening resource allocation in the coupling constraints for each subproblem computation. This can be illustrated by the following expressions:

$$\text{minimize} \quad J_{\text{ind}}(\tilde{V}_i) \tag{16}$$

$$\text{subject to} \quad \text{equation } (14)$$

$$F_i\tilde{V}_i < g_{\text{coup}} - \sum_{j=1}^{N-1} F_j\tilde{V}_j \quad \forall j \neq i \tag{17}$$

where $\tilde{V}_i$ is decision variables for track section i over the prediction horizon. The second set of constraints in equation (17) is the coupling in which the total resources or g_{coup} have been reduced by the other previous subproblems interventions. This can be done by fixing the decisions from other subproblems. In this way, the coupling constraints can be decoupled. The computation can then be solved individually by each subproblem in a sequential and non-iterative way. One advantage of assigning the couplings into individual constraints is that the solution is guaranteed to be feasible from the input perspective.

DRSBK is developed based on receding horizon control with coupling constraints [10]. The decisions for all subproblems are calculated for the entire prediction horizon. Therefore, the original approach can be directly applied to the optimization problem of maintenance planning over the entire prediction horizon. Likewise, unlike the coordinator in Lagrangian-based methods, the job of coordinator in DRSBK is only checking the feasibility of the generated solution. The decision from one subproblem are communicated to the other remaining subproblems. This enables the calculation of the remaining allocation individually.

F. Random sequence generator and stopping criterion

It is mentioned previously that the basic version of DRSBK might be stuck in local optimum or even not feasible. Hence, the algorithm is modified such that the sequence of subproblems to be processed in each iteration is generated randomly. If the output from the solver indicates that the result from an iteration is not feasible, the sequence is generated again randomly, which is in general different from the previous sequence. The feasibility checking technique is different from the Lagrangian-based algorithms, in the sense that it sum up the total feasible solutions given by all subproblems. The result is feasible if the total solved subproblems is the same with N. Meanwhile, the stopping criterion is designed so that if in two consecutive cycles the difference of objective value is below some optimality tolerance and the result from the last iterations is feasible, the iterations are stopped. In addition, the complete algorithm can be seen in Algorithm 2.

IV. Case studies

In this section, distributed optimization approaches (PALR, ADMM, and DRSBK) are compared and analyzed. Case studies consisting of different numerical experiments, are performed. Results will then be analyzed from both performance and numerical point of views. All simulations in this research are conducted on Lenovo Thinkpad X260 with an Intel core-i5

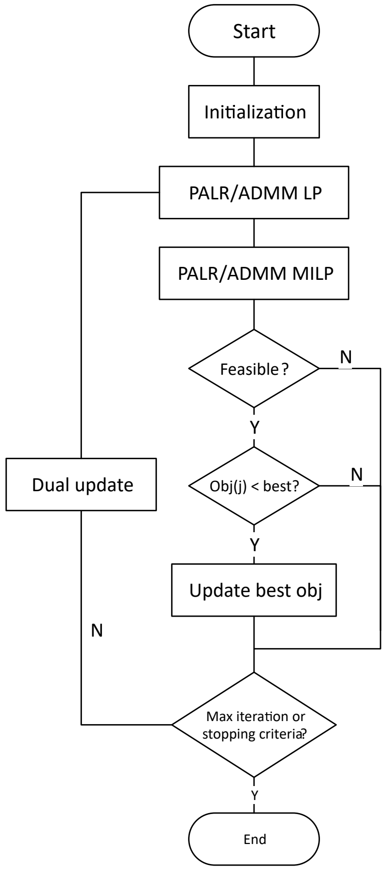

Fig. 1: Flowchart of Lagrangian-based approaches(PALR and ADMM)

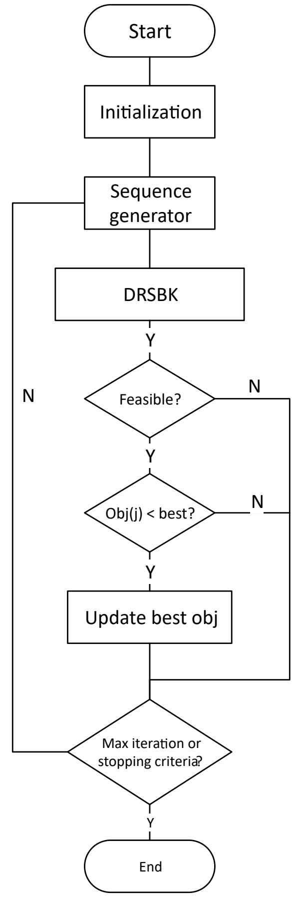

Fig. 2: Flowchart of DRSBK

processor and 8GB of RAM. All the LP and MILP problems are solved by Gurobi optimizer 7.5, called from MATLAB R2017a. Moreover, the following assumptions and general settings underpin the simulation of case studies in this section:

- The time step of maintenance intervention is one month. The control horizon is six months [3], which leads to six consecutive decisions for six months. The default prediction horizon is set to be nine.
- Initial condition and degradation rate for each track section can be different, according to the case studies. Different degradation rates for each track section are randomly generated as a Gaussian distribution [9]. Also, the degradation rate is assumed to be known and constant within the simulation horizon.
- The system is deterministic, meaning that no stochasticity or any perturbation from reactive maintenance involved.
- Practical considerations, such as maintenance machine and personnel, are assumed to be always available. Moreover, the object of study case is a single railway track, consisting of a number of track sections. Each track section is 200 m. Two stations are situated in the start and end points of the track. When performing any maintenance on the track, the tamping or renewal machine goes from one station to the other. The illustration is depicted in Figure 3.

A. *Experiment 1: multiple tests on large-scale settings*

The comparison for large-scale are presented in tables II, respectively. Each table presents the average and standard deviation values from overall ten different tests for every problem scale. First, among distributed approaches, DRSBK is the fastest algorithm. This is expected from such simple method that modifies the couplings allocation instead of the augmented Lagrangian of cost function. On top of that, it generates the closest objective function value to the centralized problem. This also implies that the average number of suggested tamping can be exactly the same as the centralized problem and the difference in the number of renewal is only one. Apart from DRSBK, the solutions given from Lagrangian-based algorithms are not close to the centralized solution. Since no convergence guarantee to the non-convex problem, they have to spend longer time in their iterations, looking for feasible solutions. The two-step computation also implies longer processing time. Besides, they suggest a higher number

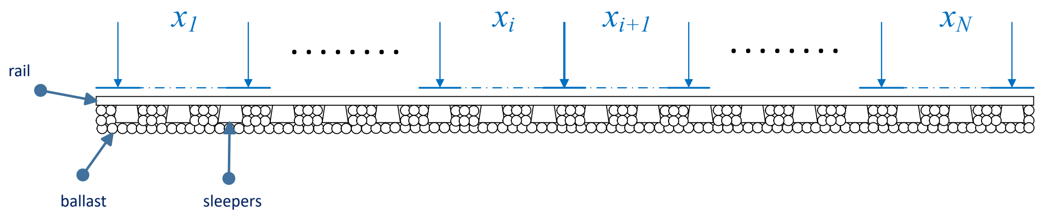

Fig. 3: Illustration of track sections

of tamping and renewal than the centralized approach. ADMM has a better average solution yet shorter processing time than PALR. This is because ADMM uses the latest decisions, thus it is easier to get into feasible regions than PALR. However, ADMM solutions is not necessarily global optimal or close to it. It might be possible to improve both Lagrangian-based algorithms to find a solution closer to the global optimum, but this requires more iterations, sacrificing the computation time. Furthermore, since they suggest more interventions, the track degradation level can be lower than the centralized optimization and DRSBK. In some conditions where the track is busy, these results are preferred. However, it is not cost-effective.

It is shown that all distributed schemes have significant faster processing time than the centralized approach. In such large-scale problem, the number of variables is really big, leading to the exponential computation cost and centralized approach might no longer be tractable. PALR can halve almost the centralized processing time. Likewise, the same pattern for both Lagrangian-based algorithms is again noticed. Again, ADMM is better at all criterion than PALR. Also, DRSBK stands out as the best distributed algorithm from the computational perspective.

B. *Experiment 2: gradual increment tests*

The second experiment presents comparison of performance criterion with gradual increment in the number of track section. The simulation settings are as follows. The initial degradation level and rates are Gaussian-randomized for all track sections. Likewise, the number of track sections is increased with the difference of 10 in the beginning until $N = 150$, the gap is set to be 100. The rule for all algorithms are once the threshold for centralized approach is reached (in this experiment, it is 1100 seconds), the test of corresponding approach is stopped.

Next, an illustration which shows the simulation results of the first criteria (computation time) against a number of track sections from the proposed optimization problem is depicted in the first plot of Figure 4. From the figure, it can be observed that the computation cost increases with the number of track sections. This issue is suffered not only by centralized approach, but also the distributed approaches as well. However, the curve of centralized optimization is exponential, which can be expected from such NP-hard problem. The centralized approach stops the experiment earlier than the other algorithms, at $N = 150$. PALR can continue to perform computationally reasonable until it reaches 400 track sections. ADMM can prolong its experiment until $N = 900$. ADMM can outperform PALR possibly due to the use of current iteration data rather than the previous iterations. In this way, the algorithm can quickly find feasible solutions in early iterations. On top of that, DRSBK can treat up to 1300 track sections, which proves itself to be the fastest among all algorithms. Moreover, it is notable that during simulations in large-scale instances (with $N > 150$) the simulations of centralized optimization approach is sometimes not possible to be carried out due to lack of available memory in the computer. This issue never happens during the simulations of distributed optimization algorithms. Furthermore, the second plot in Figure 4 shows that in general, the curve of objective function values for all algorithms are linear. From start to the end of the centralized approach computation, all of them look coinciding with each other. After $N > 150$, PALR and ADMM have diverted curve due to their higher objective function values. They are very likely trapped in local optimum. On the other hand, the curve of DRSBK coincides with the curve of the centralized. This means that DRSBK solutions close to the global optimum.

V. Conclusions

In this research, three different distributed optimization approaches have been developed for large-scale railway track maintenance operations planning. First of all, the formulation of an optimization problem is done. The formulation includes an objective functions and a series of prominent characteristics of track maintenance operations. Afterwards, three distributed approaches are designed for the optimization problem. The first two distributed optimization approaches work based on Lagrangian duality theory: Parallel Augmented Lagrangian Relaxation (PALR) and Alternating Direction Method of Multipliers (ADMM). These approaches are modified with extension techniques such that they can handle the non-convex mixed-integer problem. Furthermore, a stopping criterion is designed, such that input-feasible and suboptimal solutions can be retrieved within reasonable time. Alongside the Lagrangian-based approaches, DRSBK is implemented. To avoid infeasible solutions, A random sequence process algorithm is also added. This makes the algorithm become iterative. Likewise, a stopping criteria is used such that the solution is feasible from the output perspective.

In case studies, it is shown that the distributed optimization approaches are able to solve the proposed problem quicker than

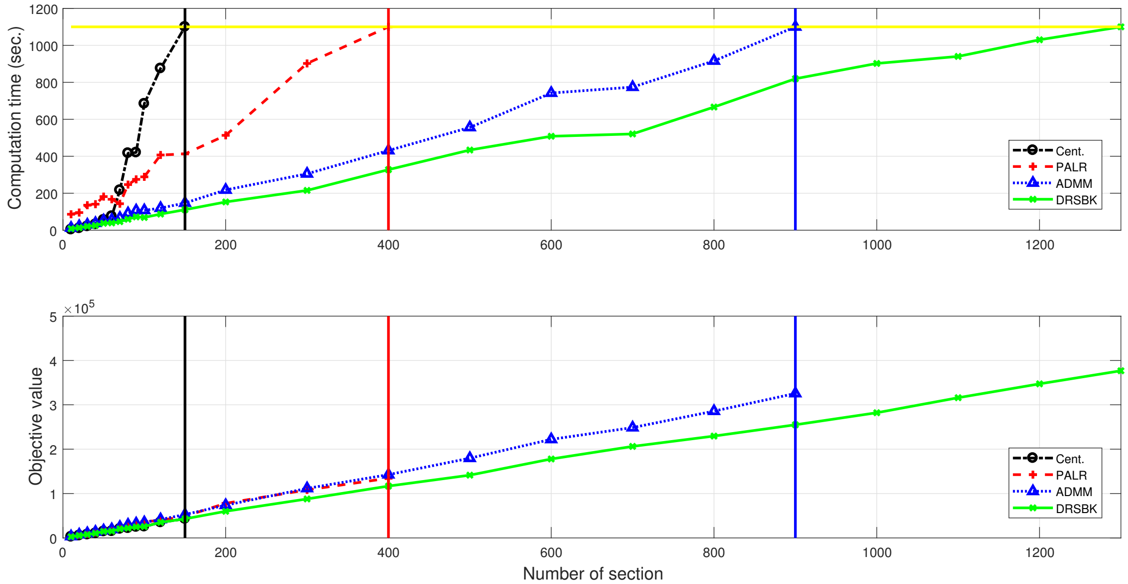

Fig. 4: Comparison of computation time and objective function value

TABLE II: performance comparison in large-scale problem

Parameter / algorithm	Centralized		PALR		ADMM		DRSBK	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev
Total J(V)	36200.18	12619.83	54137.37	5856.02	45174	10102.62	36205.13	12619.1
Comp. time (sec.)	871.84	512.43	453.74	117.02	160.12	21.48	104.67	15.15
Number of tamping	493.00	105.77	688.00	213.42	946.00	35.35	493.5	105.58
Number of renewal	43.5	43.17	94.78	22.06	58.30	41.05	43.5	43.17
Total track performance	18220.18	2361.42	18824.03	1106.48	15300.91	3977.68	18220.13	2361.36
Normalized J(V)	-	-	-52.00%	0.63	-31.00%	0.15	-0.02%	0

centralized approach in large-scale instances. DRSBK rises as the fastest yet being able to generate the closest solution to the centralized problem. Furthermore, ADMM is quicker than PALR. However, solutions from both Lagrangian-based methods are only suboptimals. Therefore, a Lagrangian dual-based approach might not be a suitable option to be developed for the future research of the proposed optimization problem.

Various future works are available for improving the current methods and results. The use of real-life data is useful to evaluate the effectiveness of the distributed approaches. Model identification can also be carried out. Furthermore, the stochastic degradation model can also be utilized to consider uncertainty and perturbation in railway maintenance. Finally, the current scheme can be extended into distributed hierarchical to facilitate different time-scales of planning.

References

[1] S. Boyd et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". In: *Foundations and Trends in Machine Learning* 3.1 (2011), pp. 1–122. ISSN: 1935-8237. DOI: 10.1561/2200000016. URL: papers3://publication/uuid/4BDE54D0-4DE4-4136-BF94-8E1201C74798.

[2] Y. Chen et al. "Dynamic optimisation of preventative and corrective maintenance schedules for a large scale urban drainage system". In: *European Journal of Operational Research* 257.2 (2017), pp. 494–510. ISSN: 03772217. DOI: 10.1016/j.ejor.2016.07.027. URL: http://www.sciencedirect.com/science/article/pii/S0377221716305641.

[3] C. Esveld. *Modern railway track*. MRT-Productions, 2001. ISBN: 9789080032415. URL: https://books.google.nl/books?id=DYaQQgAACAAJ.

[4] S. M. Famurewa et al. "Optimisation of maintenance track possession time: a tamping case study". In: *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 229.1 (2015), pp. 12–22. DOI: 10.1177/0954409713495667. URL: http://dx.doi.org/10.1177/0954409713495667.

[5] C. Fang et al. "Distributed energy consumption management in green content-centric networks via dual decomposition". In: *IEEE Systems Journal* 11.2 (2017),

pp. 625–636. ISSN: 19379234. DOI: 10.1109/JSYST.2015.2454231.

[6] M. J. Feizollahi et al. "Large-scale decentralized unit commitment". In: *International Journal of Electrical Power and Energy Systems* 73 (2015), pp. 97–106. ISSN: 01420615. DOI: 10.1016/j.ijepes.2015.04.009.

[7] A. Firooznia, R. Bourdais, and B. De Schutter. "A distributed algorithm to determine lower and upper bounds in branch and bound for hybrid Model Predictive Control". In: *2015 54th IEEE Conference on Decision and Control (CDC)*. 2015, pp. 1736–1741. DOI: 10.1109/CDC.2015.7402461.

[8] F. Guo et al. "A distributed hierarchical algorithm for multi-cluster constrained optimization". In: *Automatica* 77 (2017), pp. 230–238. ISSN: 00051098. DOI: 10.1016/j.automatica.2016.11.029. URL: http://www.sciencedirect.com/science/article/pii/S0005109816304642.

[9] I. A. Khouy et al. "Cost-effective track geometry maintenance limits". In: *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 230.2 (2016), pp. 611–622. DOI: 10.1177/0954409714542859. URL: http://dx.doi.org/10.1177/0954409714542859.

[10] Y. Kuwata et al. "Distributed robust receding horizon control for multivehicle guidance". In: *IEEE Transactions on Control Systems Technology* 15.4 (2007), pp. 627–641. ISSN: 10636536. DOI: 10.1109/TCST.2007.899152.

[11] L. Li, R. R. Negenborn, and B. De Schutter. "Distributed model predictive control for cooperative synchromodal freight transport". In: *Transportation Research Part E: Logistics and Transportation Review* (2016). DOI: 10.1016/j.tre.2016.08.006.

[12] S. Li, R. R. Negenborn, and G. Lodewijks. "Distributed constraint optimization for addressing vessel rotation planning problems". In: *Engineering Applications of Artificial Intelligence* 48 (2016), pp. 159–172. ISSN: 0952-1976. DOI: 10.1016/j.engappai.2015.11.001. URL: http://dx.doi.org/10.1016/j.engappai.2015.11.001.

[13] R.R. Negenborn, B. De Schutter, and J. Hellendoorn. "Multi-agent model predictive control for transportation networks: Serial versus parallel schemes". In: *Engineering Applications of Artificial Intelligence* 21.3 (2008), pp. 353–366. ISSN: 09521976. DOI: 10.1016/j.engappai.2007.08.005. URL: http://linkinghub.elsevier.com/retrieve/pii/S0952197607001157.

[14] A. Ortega Hortelano, S. Blainey, and J. Preston. "Reducing maintenance costs for ballasted track: evidence from Britain". 2016. URL: https://eprints.soton.ac.uk/391238/.

[15] A. Reina et al. "Maintenance decision support for manufacturing systems based on the minimization of the life cycle cost". In: *Procedia CIRP* 57 (2016), pp. 674–679. ISSN: 22128271. DOI: 10.1016/j.procir.2016.11.117. URL: http://linkinghub.elsevier.com/retrieve/pii/S221282711631277X.

[16] Stefano Sebastio, Giorgio Gnecco, and Alberto Bemporad. "Optimal Distributed Task Scheduling in Volunteer Clouds". In: *Comput. Oper. Res.* 81.C (May 2017), pp. 231–246. ISSN: 0305-0548. DOI: 10.1016/j.cor.2016.11.004. URL: https://doi.org/10.1016/j.cor.2016.11.004.

[17] K. Singh, A. Gupta, and T. Ratnarajah. "A utility-based joint subcarrier and power allocation for green communications in multi-user two-way regenerative relay networks". In: *IEEE Transactions on Communications* 65.9 (Sept. 2017), pp. 3705–3722. ISSN: 0090-6778. DOI: 10.1109/TCOMM.2017.2713382.

[18] Zhou Su et al. "Multi-level condition-based maintenance planning for railway infrastructures – A scenario-based chance-constrained approach". In: *Transportation Research Part C: Emerging Technologies* 84 (2017), pp. 92–123. ISSN: 0968090X. DOI: 10.1016/j.trc.2017.08.018.

[19] Z. Su et al. "Model predictive control for maintenance operations planning of railway infrastructures". In: *Computational Logistics: 6th International Conference, ICCL 2015, Delft, The Netherlands, September 23-25, 2015, Proceedings*. Ed. by Francesco Corman, Stefan Voß, and Rudy R Negenborn. Cham: Springer International Publishing, 2015, pp. 673–688. ISBN: 978-3-319-24264-4. DOI: 10.1007/978-3-319-24264-4_46. URL: http://dx.doi.org/10.1007/978-3-319-24264-4%7B%5C_%7D46.

[20] C. Vale, I .M. Ribeiro, and R. Calcada. "Integer programming to optimize tamping in railway tracks as preventive maintenance". In: *Journal of Transportation Engineering* 138.1 (2012), pp. 123–131.

[21] K. Verbert, B. De Schutter, and R. Babuška. "Timely condition-based maintenance planning for multi-component systems". In: *Reliability Engineering & System Safety* 159 (2017), pp. 310–321. ISSN: 0951-8320. DOI: http://dx.doi.org/10.1016/j.ress.2016.10.032. URL: http://www.sciencedirect.com/science/article/pii/S0951832016307128.

[22] M. Wang et al. "Joint channel allocation, mode selection and power control in D2D-enabled femtocells". In: *MILCOM 2016 - 2016 IEEE Military Communications Conference*. Nov. 2016, pp. 454–459. DOI: 10.1109/MILCOM.2016.7795369.

[23] B. Yang and M. Johansson. "Distributed optimization and games: a tutorial overview". In: *Networked Control Systems*. Ed. by Alberto Bemporad, Maurice Heemels, and Mikael Johansson. London: Springer London, 2010, pp. 109–148. ISBN: 978-0-85729-033-5. DOI: 10.1007/978-0-85729-033-5_4. URL: http://dx.doi.org/10.1007/978-0-85729-033-5%7B%5C_%7D4.

Bibliography

[1] C. Esveld, *Modern railway track.* MRT-Productions, 2001.

[2] S. Li, R. R. Negenborn, and G. Lodewijks, "Distributed constraint optimization for addressing vessel rotation planning problems," *Engineering Applications of Artificial Intelligence*, vol. 48, pp. 159–172, 2016.

[3] I. A. Khouy, P. Larsson-Kråik, A. Nissen, and U. Kumar, "Cost-effective track geometry maintenance limits," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 2, pp. 611–622, 2016.

[4] Z. Su, A. Núñez, A. Jamshidi, S. Baldi, Z. Li, R. Dollevoet, and B. De Schutter, "Model predictive control for maintenance operations planning of railway infrastructures," in *Computational Logistics: 6th International Conference, ICCL 2015, Delft, The Netherlands, September 23-25, 2015, Proceedings* (F. Corman, S. Voß, and R. R. Negenborn, eds.), (Cham), pp. 673–688, Springer International Publishing, 2015.

[5] A. Jamshidi, A. Núñez, R. Dollevoet, and Z. Li, "Robust and predictive fuzzy key performance indicators for condition-based treatment of squats in railway infrastructures," *Journal of Infrastructure Systems*, vol. 23, no. 3, p. 4017006, 2017.

[6] A. Jamshidi, S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, R. Dollevoet, Z. Li, and B. De Schutter, "A big data analysis approach for rail failure risk assessment," *Risk Analysis*, vol. 37, no. 8, pp. 1495–1507, 2017.

[7] J. Chen, Z. Liu, H. Wang, A. Núñez, and Z. Han, "Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. PP, no. 99, pp. 1–13, 2017.

[8] R. Li, K. B. Salling, A. Zoeteman, and A. R. M. Wolfert, "Preventive condition-based tamping for railway tracks: a decision support model," in *Life-Cycle of Engineering Systems: Emphasis on Sustainable Civil Infrastructure: Proceedings of the Fifth International Symposium on Life-Cycle Civil Engineering (IALCCE 2016), 16-19 October 2016, Delft, The Netherlands*, p. 355, CRC Press, 2016.

[9] C. Vale, I. M. Ribeiro, and R. Calcada, “Integer programming to optimize tamping in railway tracks as preventive maintenance,” *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 123–131, 2012.

[10] S. M. Famurewa, T. Xin, M. Rantatalo, and U. Kumar, “Optimisation of maintenance track possession time: a tamping case study,” *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 229, no. 1, pp. 12–22, 2015.

[11] E. Gustavsson, “Scheduling tamping operations on railway tracks using mixed integer linear programming,” *EURO Journal on Transportation and Logistics*, vol. 4, no. 1, pp. 97–112, 2015.

[12] F. Guo, C. Wen, J. Mao, G. Li, and Y. D. Song, “A distributed hierarchical algorithm for multi-cluster constrained optimization,” *Automatica*, vol. 77, pp. 230–238, 2017.

[13] M. Wen, R. Li, and K. Salling, “Optimization of preventive condition-based tamping for railway tracks,” *European Journal of Operational Research*, vol. 252, no. 2, pp. 455–465, 2016.

[14] Z. Su, A. Núñez, S. Baldi, and B. De Schutter, “Model predictive control for rail condition-based maintenance: a multilevel approach,” in *Proceedings of the 19th IEEE Intelligent Transportation Systems Conference (ITSC 2016)*, (Rio de Janeiro, Brazil), pp. 354–359, 2016.

[15] J. Zhao, A. H. C. Chan, and M. P. N. Burrow, “A genetic-algorithm-based approach for scheduling the renewal of railway track components,” *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 223, no. 6, pp. 533–541, 2009.

[16] R. Li and R. Roberti, “Optimal scheduling of railway track possessions in large-scale projects with multiple construction works,” *Journal of Construction Engineering and Management*, vol. 143, no. 6, p. 4017007, 2017.

[17] L. F. Caetano and P. F. Teixeira, “Optimisation model to schedule railway track renewal operations: a life-cycle cost approach,” *Structure and Infrastructure Engineerinmg*, vol. 11, no. 11, pp. 1524–1536, 2015.

[18] Y. Chen, P. Cowling, F. Polack, S. Remde, and P. Mourdjis, “Dynamic optimisation of preventative and corrective maintenance schedules for a large scale urban drainage system,” *European Journal of Operational Research*, vol. 257, no. 2, pp. 494–510, 2017.

[19] A. Ortega Hortelano, S. Blainey, and J. Preston, “Reducing maintenance costs for ballasted track: evidence from Britain.” 2016.

[20] A. Reina, Á. Kocsis, A. Merlo, I. Németh, and F. Aggogeri, “Maintenance decision support for manufacturing systems based on the minimization of the life cycle cost,” *Procedia CIRP*, vol. 57, pp. 674–679, 2016.

[21] M. A. Ferrario and B. Smyth, “Distributing case-base maintenance: the collaborative maintenance approach,” *Computational Intelligence*, vol. 17, no. 2, pp. 315–330, 2001.

[22] Z. Su, A. Jamshidi, A. Núñez, S. Baldi, and B. De Schutter, "Multi-level condition-based maintenance planning for railway infrastructures – A scenario-based chance-constrained approach," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 92–123, 2017.

[23] R. R. Negenborn and J. M. Maestre, "On 35 approaches for distributed MPC made easy," in *Distributed Model Predictive Control Made Easy* (J. M. Maestre and R. R. Negenborn, eds.), pp. 1–37, Dordrecht: Springer Netherlands, 2014.

[24] K. Verbert, B. De Schutter, and R. Babuška, "Timely condition-based maintenance planning for multi-component systems," *Reliability Engineering & System Safety*, vol. 159, pp. 310–321, 2017.

[25] B. Yang and M. Johansson, "Distributed optimization and games: a tutorial overview," in *Networked Control Systems* (A. Bemporad, M. Heemels, and M. Johansson, eds.), pp. 109–148, London: Springer London, 2010.

[26] I. Necoara, V. Nedelcu, and I. Dumitrache, "Parallel and distributed optimization methods for estimation and control in networks," *Journal of Process Control*, vol. 21, no. 5, pp. 756–766, 2011.

[27] F. Farokhi, I. Shames, and K. H. Johansson, "Distributed MPC via dual decomposition and alternative direction method of multipliers," in *Distributed Model Predictive Control Made Easy* (J. M. Maestre and R. R. Negenborn, eds.), pp. 115–131, Dordrecht: Springer Netherlands, 2014.

[28] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[29] R. Negenborn, B. De Schutter, and J. Hellendoorn, "Multi-agent model predictive control for transportation networks: Serial versus parallel schemes," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 3, pp. 353–366, 2008.

[30] L. Li, R. R. Negenborn, and B. De Schutter, "Distributed model predictive control for cooperative synchromodal freight transport," *Transportation Research Part E: Logistics and Transportation Review*, 2016.

[31] A. Firooznia, R. Bourdais, and B. De Schutter, "A distributed algorithm to determine lower and upper bounds in branch and bound for hybrid Model Predictive Control," in *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 1736–1741, 2015.

[32] M. Wang, H. Gao, X. Su, and T. Lv, "Joint channel allocation, mode selection and power control in D2D-enabled femtocells," in *MILCOM 2016 - 2016 IEEE Military Communications Conference*, pp. 454–459, nov 2016.

[33] K. Singh, A. Gupta, and T. Ratnarajah, "A utility-based joint subcarrier and power allocation for green communications in multi-user two-way regenerative relay networks," *IEEE Transactions on Communications*, vol. 65, pp. 3705–3722, sep 2017.

[34] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, “Distributed energy consumption management in green content-centric networks via dual decomposition,” *IEEE Systems Journal*, vol. 11, no. 2, pp. 625–636, 2017.

[35] S. Timotheou, C. G. Panayiotou, and M. M. Polycarpou, “Network traffic signal control with nonconvex alternating direction method of multipliers formulations,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2490, pp. 11–20, 2015.

[36] C. He, L. Wu, T. Liu, and M. Shahidehpour, “Robust co-optimization scheduling of electricity and natural gas systems via ADMM,” *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 658–670, 2017.

[37] M. J. Feizollahi, M. Costley, S. Ahmed, and S. Grijalva, “Large-scale decentralized unit commitment,” *International Journal of Electrical Power and Energy Systems*, vol. 73, pp. 97–106, 2015.

[38] V. Gunnerud and B. Foss, “Oil production optimization-a piecewise linear model, solved with two decomposition strategies,” *Computers and Chemical Engineering*, vol. 34, no. 11, pp. 1803–1812, 2010.

[39] O. Miksik, V. Vineet, P. Pérez, and P. Torr, “Distributed non-convex ADMM-inference in large-scale random fields,” in *British Machine Vision Conference (BMVC)*, pp. 1–12, 2014.

[40] A. Alavian and M. C. Rotkowitz, “Improving ADMM-based optimization of mixed integer objectives,” in *2017 51st Annual Conference on Information Sciences and Systems, CISS 2017*, 2017.

[41] S. Sebastio, G. Gnecco, and A. Bemporad, “Optimal Distributed Task Scheduling in Volunteer Clouds,” *Comput. Oper. Res.*, vol. 81, pp. 231–246, may 2017.

[42] Y. Kuwata, A. Richards, T. Schouwenaars, and J. P. How, “Distributed robust receding horizon control for multivehicle guidance,” *IEEE Transactions on Control Systems Technology*, vol. 15, no. 4, pp. 627–641, 2007.

[43] Y. Kuwata and J. P. How, “Cooperative distributed robust trajectory optimization using receding horizon MILP,” *IEEE Transactions on Control Systems Technology*, vol. 19, no. 2, pp. 423–431, 2011.

[44] Z. Yang, K. Long, P. You, and M. Chow, “Joint scheduling of large-scale appliances and batteries via distributed mixed optimization,” in *2015 IEEE Power Energy Society General Meeting*, p. 1, 2015.

[45] F. Vanderbeck and M. Savelsbergh, “A generic view of Dantzig-Wolfe decomposition in mixed integer programming,” *Operations Research Letters*, vol. 34, no. 3, pp. 296–306, 2006.

[46] J. E. Contreras-Ocana, M. R. Sarker, and M. A. Ortega-Vazquez, “Decentralized Coordination of a Building Manager and an Electric Vehicle Aggregator,” *IEEE Transactions on Smart Grid*, vol. PP, no. 99, p. 1, 2017.

[47] B. Lichtberger, "The track system and its maintenance," *Railway Technical Review (RTR) Maintenance & Renewal*, pp. 14–19, 2007.

[48] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, no. 3, pp. 407–427, 1999.

[49] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach.* New York, NY, USA: Cambridge University Press, 1st ed., 2007.

[50] P. V., *Combinatorial optimization volume 1, concepts of combinatorial optimization.* Weinheim -: Wiley-VCH, 2010.

[51] C. B. Royo, *Generalized unit commitment by the radar multiplier method.* PhD thesis, Technical University of Catalonia, Barcelona, Spain, 2001.

[52] K. J. Aström and B. Wittenmark, *Computer-controlled Systems (3rd Ed.).* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.

[53] E. F. Camacho and C. Bordons, *Model predictive control.* Advanced Textbooks in Control and Signal Processing, Springer London, 2004.

Glossary

List of Acronyms

KPI	Key Performance Indicator
MPC	Model Predictive Control
MLD	Mixed-Logical Dynamical
MILP	Mixed-Integer Linear Programming
MIQP	Mixed-Integer Quadratic Programming
DCCC	Decoupled Cost but Coupled Constraint
CCDC	Coupled Cost but Decoupled Constraint
PALR	Parallel Augmented Lagrangian Relaxation
ADMM	Alternating Direction Method of Multipliers
DRSBK	Distributed Robust Safe But Knowledgeable

List of Symbols

$x_{1,i}(\mathrm{k})$	Track degradation level represented by standard deviation of longitudinal level [mm]
$x_{2,i}(k)$	Track offset memory
$a_{1,i}$	Track degradation rate [mm/MGT]
$a_{2,i}$	Offset memory rate
$u_i(k)$	Maintenance action (input)
$\delta_{1,i}(k)$	Renewal binary indicator input
$\delta_{2,i}(k)$	Tamping binary indicator input
$z_{p,i}(k)$	Auxiliary variables

λ	Trade-off variable
$\mathrm{h}_{\mathrm{min}}$	Minimum level threshold
$\mathrm{h}_{\mathrm{max}}$	Maximum level threshold
$\mathrm{t}_{\mathrm{max}}$	Maximum closure time for tamping or renewal work
T	Prediction horizon
N	Number of track sections
$V_i(k)$	Decision variables for track section i at time step k
$\tilde{V}$	Decision variables for track section i over the prediction horizon
$\bar{V}$	Decision variables for all track sections over the prediction horizon
$s(k)$	Aggregated slack vector at time step k
$\bar{S}$	Aggregated slack vector over the prediction horizon
γ^t	Dual variables for tamping closure time coupling over the prediction horizon
γ^r	Dual variables for renewal closure time coupling over the prediction horizon
γ	Aggregated dual variables