A sparse optimization approach for energy-efficient timetabling in metro railway systems

Xiaoyu Li[a], Ziyan Luo*[b], Naihua Xiu[a]

[a]*Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, China*
[b]*State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China*

Abstract

In this paper we propose a sparse optimization approach to maximize the utilization of regenerative energy produced by baking trains for energy-efficient timetabling in metro railway systems. By introducing the cardinality function and the square of the Euclidean norm function as the objective function, the resulting sparse optimization model can characterize the utilization of the regenerative energy appropriately. A two-stage alternating direction method of multipliers is designed to efficiently solve the convex relaxation counterpart of the original NP-hard problem and then to produce an energy-efficient timetable of trains. The resulting approach is applied to Beijing Metro Yizhuang Line with different instances of service for case study. Comparison with the approach proposed by Das Gupta et al. [Transportation Research Part B 93 (2016): 57-74] is also conducted which illustrates the effectiveness of our proposed sparse optimization model and the efficiency of our numerical optimization algorithm.

Keywords: Energy-efficient timetable, regenerative braking energy, sparse optimization model, alternative direction method of multipliers

1. Introduction

With the rapid increase in industry and population in cities, the metro railway system has been playing a more and more important role in urban economic and social development for metropolis due to its high loading capacity and low pollution. Despite the inherent efficiency, the energy consumption of metro railway systems is of a huge magnitude and the efficient energy management for such systems is then of global importance. It is known that the energy consumption of trains is largely influenced by the driving strategy and the arrivals and departures in the timetable, which inspires the extensive study on energy-efficient train control and energy-efficient train timetabling. A careful treatment on such topics can be found in the nice survey paper by Scheepmaker et al. in [15], with focuses on reviewing the mathematical problem formulations and solution approaches in the literature.

Among all the involved energy saving strategies in metro railway systems, an energy-efficient timetable with regenerative braking which schedules the arrival time and the departure time of each train to and from the platforms it visits for the minimization of energy consumption and the maximization of produced regenerative energy utilization, has gained considerable academical attention in recent years. In 2012, Peña-Alcaraz et al. [14] proposed a large-scale mixed integer programming model to maximize the utilization of regenerative braking energy by means of maximizing the total duration of all possible synchronization processes of acceleration and regenerative braking between all possible train pairs. The simulation model is applied to line three of the Madrid metro system and the energy savings of the optimized timetable is reported to be about 7% on average. By merely choosing those train pairs suitable for regenerating energy transfer other than all possible train pairs in [14], Das Gupta et al. developed a more tractable mixed integer programming model in [9]. In 2013, Yang et al. [22] presented a cooperative integer programming model to maximize the time overlaps of nearby accelerating and braking trains and applied a genetic algorithm to

Email addresses: 16121607@bjtu.edu.cn (Xiaoyu Li), zyluo@bjtu.edu.cn, corresponding author (Ziyan Luo*), nhxiu@bjtu.edu.cn (Naihua Xiu)

approximately solve the resulting optimization model. For the case study on Beijing Metro Yizhuang Line, they reported 22.1% and 15.2% increase of effective time overlaps during peak hours and off-peak hours, respectively. A two-objective timetable optimization model for maximizing the regenerative braking energy with consideration of the passenger waiting time was developed by Yang et al. in [23] where the genetic algorithm was again adopted to solve the optimization model. An integrated energy-efficient operation model was proposed by Li & Lo in [10] with the speed profile part to minimize the net energy consumption while the timetabling part with headway constraints to synchronize the accelerating and regenerative braking for the reuse of regenerative energy. About 25% energy savings was reported for Beijing Metro Yizhuang Line with the headway between trains being 90 seconds. Li & Lo further developed an integrated energy-efficient timetable and speed profile optimization model to determine the cycle time, the headway time and the speed profiles for a metro line to minimize the energy consumption [17]. The resulting mathematical model is a strictly convex quadratic programming problem with simple bound constraints and an analytical approach based on the first-order optimality condition (i.e., the KKT condition). An integrated timetable and speed profile optimization model with Taylor approximation was also considered by Yang et al. in [19] where the active set method was employed to handle the resulting strictly convex quadratic programming with linear constraints. Under the assumption that trains are operating in the optimal speed profiles, Yang et al.[21] proposed a mixed integer programming model aiming at synchronizing the arrivals and departures of the trains such that regenerative braking energy can be used as much as possible during the whole day service. A two-step linear programming model with the first step to minimize the total energy consumed by all trains and the second step to maximize the utilization of regenerative energy produced by braking trains for a whole day service in metro railway networks was proposed by Das Gupta et al. in [4]. To ensure the transfer of maximum possible regenerative energy from braking trains to accelerating trains, a time overlap vector was introduced which requires to be as close as possible to the zero vector. To effectively quantify such a requirement, the ℓ_0-norm and the ℓ_1-norm were employed in [4]. By adopting the ℓ_1-norm heuristic for the original ℓ_0-norm (i.e., the number of nonzero components in a vector), together with epigraph approach that equivalently transforms the convex ℓ_1-norm minimization to a linear programming, the resulting relaxation problem in the second step was efficiently solved by Gurobi Optimizer. Numerical experiments on different instances of service PES2-SFM2 of line 8 of Shanghai Metro network spanning a full service period of one day were conducted and a significant reduction in effective energy consumption with the worst case being 19.27% was reported by applying the proposed approach.

The original ℓ_0-norm minimization problem proposed in [4] is well-known to be NP-hard due to the discontinuity and non-convexity of the ℓ_0-norm. Such a type of optimization problems is coined by sparse optimization in the optimization community and has a wide range of applications in compressed sensing, signal processing, statistical regression, financial portfolio, and so on. Various tractable relaxation strategies are propose to replace the ℓ_0-norm, such as the popular convex surrogate ℓ_1-norm [5] and some weighted ℓ_1-norm variants [24]. Numerical algorithms are also developed for the resulting relaxation models, such as the greedy-type methods [3], the first-order optimization methods [1], and second-order methods for some special structured models [11].

Inspired by the development of sparse optimization methods, in this paper, a sparse optimization model with the ℓ_0-norm and the squared Euclidean norm as the objective function is built to maximize the utilization of the regenerative braking energy. Motivated by the adaptive LASSO model in the high-dimensional statistical regression [24], we propose a weighted ℓ_1-norm relaxation scheme. A two-stage alternating direction method of multipliers (TSADMM) is then proposed with the first stage for calculating an appropriate weight vector and an initial point for the iterative algorithm in the second stage. To illustrate the effectiveness of our new mathematical model, the linear programming model in [4], a quadratic model merely considering the squared Euclidean norm, and our weighted-ℓ_1 companied with the squared Euclidean norm are applied to Beijing Metro Yizhuang Line for comparison. The numerical results on energy savings shows that our model is the best among them. To demonstrate the efficiency of our proposed TSADMM, the academical version of CVX is called to solve the our model as well. The numerical results in computation time shows that our algorithm outperforms that used in CVX.

The rest of this paper is organized as follows. In Section 2, the sparse optimization model for energy-efficient timetabling in metro railway systems is built. In Section 3, the weighted ℓ_1 norm relaxation model is proposed and a two-stage alternating direction method of multipliers (TSADMM) is designed to get an approximate optimal solution. Theoretical global convergence is established as well. By applying our proposed approach to Beijing Metro Yizhuang Line, we demonstrate the effectiveness of our proposed model and the efficiency of our proposed TSADMM by the numerical results in Section 4. Conclusions and future research are given in Section 5. For reference, Appendix A

provides the proof of the global convergence theorem for our algorithm.

2. Model formulation

2.1. Notation

The notation that will be used throughout the paper is listed in Table 1 for the convenience of the subsequent model formulation.

Table 2.1: A list of notation.

a_i^t	The arrival time of train t at platform i
d_i^t	The departure time of train t from platform i
$\mathcal{T}$	The set of all trains
$\lvert\cdot\rvert$	The cardinality of a set, i.e., the number of elements in a set
$[\underline{tr}_{ij}^t, \overline{tr}_{ij}^t]$	The trip time window for train t from platform i to j
ϕ	The set of all the crossing-overs
O_{ij}	The set of all train pairs associated with the crossing-over (i, j)
$[\underline{\kappa}_{ij}^{tt'}, \overline{\kappa}_{ij}^{tt'}]$	The trip window for train t on the crossing-over (i, j)
$[\underline{dw}_i^t, \overline{dw}_i^t]$	The dwell time window for train t at platform i
$\mathcal{S}$	The set of all platform pairs at the same interchange station
C_{ij}	The set of all connection train pairs for a platform pair (i, j)
$[\underline{c}_{ij}^{tt'}, \overline{c}_{ij}^{tt'}]$	The connection window between the train t at the platform i and the train t' at platform j
$\mathcal{H}_{i,j}$	The set of train-pairs that move along the track (i, j)
$[\underline{h}_i^{tt'}, \overline{h}_i^{tt'}]$	The headway window between train t and t' arrival at or departure from platform i
$[\underline{tt}^t, \overline{tt}^t]$	The window of total travel time constraint for train t
f_{ij}	Energy consumption associated with the trip $(i, j) \in \mathcal{A}_{tr}$
$\mathcal{P}$	The set of all platform pairs powered by a same electrical substation
$\overrightarrow{t}$	Temporally closest train to the right of train t
$\overleftarrow{t}$	Temporally closest train to the left of train t
∇_i^t	The relative distance from a_i^t to regenerative alignment point
$\triangle_i^t$	The relative distance from d_i^t to the consumption alignment point
$\overrightarrow{\varepsilon}$	A set of ε containing elements of the form $(i, j, t, \overrightarrow{t})$
$\overleftarrow{\varepsilon}$	A set of ε containing elements of the form $(i, j, t, \overleftarrow{t})$
$\mathcal{N}$	The set of all platforms
$\mathcal{K}$	The set of all tracks
$\mathcal{K}^t$	The set of tracks visited by a train t
$\mathcal{N}^t$	The set of platforms visited by a train t
$\mathcal{K}_{tr}$	The set of all arcs associated with trip time constraints
$\delta_S(\cdot)$	The indicator function with respect to the set S
$\Pi_S(\cdot)$	The projection operator onto the set S
$sign(\cdot)$	The sign function
$\circ$	The componentwise product, also known as the Hadamard product
A^T	The transpose of the matrix A
$vec(X)$	The column vector generated by stacking all columns of the matrix X
$\mathbf{R}^n$	The n-dimensional Euclidean space
$\mathbf{R}_+^n$	The set of all nonnegative vectors in the n-dimensional Euclidean space
$\mathbf{R}_{++}^n$	The set of all positive vectors in the n-dimensional Euclidean space
$\mathbf{R}^{n\times m}$	The set of all n-by-m real matrices

2.2. Constraints

The set of constraints in the metro train networks including constraints of the trip time, the dwell time, the headway time, the total travel time and the domain of event times will be described in this subsection, followed from [4].

- Trip time constraint: It includes the trip time constraints with tracks and with crossing-overs. For the former one, When a train $t \in \mathcal{T}$ has a trip from platform $i \in \mathcal{N}^t$ to platform $j \in \mathcal{N}^t$ along the track $(i, j) \in \mathcal{K}^t$, it departs from platform i at time d_i^t and arrives at platform j at time a_j^t, and its trip time is required to be bounded by $\underline{tr}_{ij}^t$ and $\overline{tr}_{ij}^t$. Thus, the corresponding constraints can be described as follows:

$$\forall t \in \mathcal{T}, \forall (i, j) \in \mathcal{K}^t, \underline{tr}_{ij}^t \leq a_j^t - d_i^t \leq \overline{tr}_{ij}^t. \tag{1}$$

  For the latter one, one knows that a crossing-over connecting two train-lines is a special track, where a train-line is a path with some non-opposite platforms and tracks. The crossing-over is to connect the terminal platform

of a train-line with the first platform of another. A train t turns around by a crossing-over and departs from the first platform to travel through the train-line after it arrives at the final platform at another train-line, and then the same train is labelled as t' in order to treat it as two different trains[13]. Let ϕ represent the set of all crossing-overs. Consider a crossing-over $(i, j) \in \phi$, we denote the set of all train pairs associated with a same physical train through the crossing-over (i, j) by O_{ij}. The trip time to travelling through the crossing-over $(i, j) \in \phi$ is the time from the departure from the platform i(train is labelled as t) to the arrival at the platform j(train is labelled as t'), which has to stay within a window $[\underline{\kappa}_{ij}^{tt'}, \overline{\kappa}_{ij}^{tt'}]$. And the corresponding constraints can be described as below:

$$\forall (i, j) \in \phi, \forall (t, t') \in O_{i,j}, \underline{\kappa}_{ij}^{tt'} \leq a_j^{t'} - d_i^t \leq \overline{\kappa}_{ij}^{tt'}. \tag{2}$$

- Dwell time constraint: When a train $t \in \mathcal{T}$ arrives at a platform $i \in \mathcal{N}^t$, it should dwell there in a time interval denoted by $[\underline{dw}_i^t, \overline{dw}_i^t]$ in order to make sure that passengers can get off and get on the train before its departure from the platform i. The dwell time constraint can be written as:

$$\forall t \in \mathcal{T}, \forall i \in \mathcal{N}^t, \underline{dw}_i^t \leq d_i^t - a_i^t \leq \overline{dw}_i^t. \tag{3}$$

- Connection constraint: An interchange station used by connecting trains is to make passengers transfer between trains on different lines. Any platform pair $(i, j) \in \mathcal{S}$ consists of two platforms at a same interchange station. For a connection train pair $(t, t') \in C_{ij}$, we assume that the train t arrives at platform i and another train t' departs from platform j. The difference of both times can be in a window $[\underline{c}_{ij}^{tt'}, \overline{c}_{ij}^{tt'}]$ to maintain that passengers can get off from the former train and get on the latter. The constraint of connection can be described as:

$$\forall (i, j) \in \mathcal{S}, \forall (t, t') \in C_{ij}, \underline{c}_{ij}^{tt'} \leq d_j^{t'} - a_i^t \leq \overline{c}_{ij}^{tt'}. \tag{4}$$

- Headway time constraint: Similar to the trip time constraints, the headway time constraints includes those with tracks and with crossing-overs. In any subway system, a minimum interval between the departures and arrivals of sequenced trains on the same track is maintained, called headway time. To meet the needs of the passengers, railway networks in many cities designate an upper bound between the departures and arrivals of consecutive trains, in order to ensure passengers do not wait for a long time before the arrival of next train. Define $\mathcal{H}_{ij}$ be the set of pairs of successive trains that pass by track $(i, j) \in \mathcal{K}$. Consider $(t, t') \in \mathcal{H}_{ij}$, and let $[\underline{h}_i^{tt'}, \overline{h}_i^{tt'}]$ and $[\underline{h}_j^{tt'}, \overline{h}_j^{tt'}]$ be the time intervals that are maintained the departures and arrivals of consecutive trains t and t' from and to platform i platform j respectively. For any $(i, j) \in \mathcal{K}$ between platform i and platform j, headway time constraint with tracks can be described as follows:

$$\forall (i, j) \in \mathcal{K}, (t, t') \in \mathcal{H}_{ij}, \quad \begin{cases} \underline{h}_i^{tt'} \leq d_i^t - d_i^{t'} \leq \overline{h}_i^{tt'}, \\ \underline{h}_j^{tt'} \leq a_j^t - a_j^{t'} \leq \overline{h}_j^{tt'}. \end{cases} \tag{5}$$

In the other case, consider two consecutive trains t_1 and t_2 that go through a crossing-over $(i, j) \in \phi$ after departing from the final platform i of a train-line. They will be labelled as t_1' and t_2' when they arrive at the first platform j of some other train-line. Let $\widetilde{O}_{ij}$ be the set of all such two-train pairs $\big((t_1, t_1'), (t_2, t_2')\big)$. There is an interval time window $[\underline{h}_i^{t_1t_2}, \overline{h}_i^{t_1t_2}]$ between the departures of train t_1 and t_2 from platform i and a window $[\underline{h}_j^{t_1't_2'}, \overline{h}_j^{t_1't_2'}]$ between the arrivals of trains t_1' and t_2' at platform j. We can then write the headway constraints with crossing-overs as follows:

$$\forall (i, j) \in \phi, \forall ((t_1, t_1'), (t_2, t_2')) \in \widetilde{O}_{ij}, \quad \begin{cases} \underline{h}_i^{t_1t_2} \leq d_i^{t_1} - d_i^{t_2} \leq \overline{h}_i^{t_1t_2}, \\ \underline{h}_j^{t_1't_2'} \leq a_j^{t_1'} - a_j^{t_2'} \leq \overline{h}_j^{t_1't_2'}. \end{cases} \tag{6}$$

Besides, the headway time is related to the passenger demand and the number of trains. Let $\bar{m}'$ be the number

of trains that are in service per hour, c be the train capacity, u be the train utility rate and D be the passenger demand. Then, $D = c \times u \times \bar{m}'$ [10]. Since $h = \frac{3600}{\bar{m}'}$, we have

$$h = \frac{3600 \times c \times u}{D}. \tag{7}$$

Here we assume that u and c are constant while the passenger demand varies. In other words, the headway time of trains changes at different periods in a day.

- Total travel time constraint: When a train $t \in \mathcal{T}$ has a trip, traversing all platforms and tracks in chronological order, it can have the total travel time between $\underline{tt}^t$ and $\overline{tt}^t$, to meet the needs of passengers in subway systems. Total travel time constraints can be written as:

$$\forall t \in \mathcal{T}, \underline{tt}^t \leq d^t_{\mathcal{N}^t(|\mathcal{N}^t|)} - a^t_{\mathcal{N}^t(1)} \leq \overline{tt}^t, \tag{8}$$

- Domain of event times: We set zero second as the time of the departure of the first train from a platform in a day when the railway is in service. By adding the maximum possible values of all trip times and dwell times, we can set an upper bound for arrival of the last train at the final platform. And we denote the upper bound by $m_T \in Z_{++}$. So, we can write the domain of the event time constraints as follows:

$$\forall t \in \mathcal{T}, \forall i \in \mathcal{N}^t, 0 \leq a^t_i \leq m_T, 0 \leq d^t_i \leq m_T. \tag{9}$$

2.3. Minimizing the energy consumption

The major energy consumption of a train is from the acceleration phase. To efficiently minimize the total energy consumption of trains in a metro railway network, Das Gupta [4] proposed an optimization model with linear inequality constraints to get a feasible timetable. The involved optimization model takes the form of

$$\begin{aligned} \min \quad & \sum_{(i,j)\in\mathcal{K}_{tr}, t\in\mathcal{T}} f_{ij}(a^t_j - d^t_i) \\ s.t. \quad & (1) - (9) \end{aligned} \tag{10}$$

where $f_{ij} : \mathbf{R}_{++} \to \mathbf{R}_{++}$ with argument $(a^t_j - d^t_i)$ is the energy consumption function is unknown in advance and some best possible affine approximate function in the sense of least-squares by using the measured energy in practice is then utilized. By denoting $x := \mathrm{vec}(X^T)$ with

$$X = \begin{pmatrix} a^1_1 & a^2_1 & \cdots & a^{|\mathcal{T}|-1}_1 & a^{|\mathcal{T}|}_1 \\ d^1_1 & d^2_1 & \cdots & d^{|\mathcal{T}|-1}_1 & d^{|\mathcal{T}|}_1 \\ a^1_2 & a^2_2 & \cdots & a^{|\mathcal{T}|-1}_2 & a^{|\mathcal{T}|}_2 \\ d^1_2 & d^2_2 & \cdots & d^{|\mathcal{T}|-1}_2 & d^{|\mathcal{T}|}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a^1_{|\mathcal{N}|} & a^2_{|\mathcal{N}|} & \cdots & a^{|\mathcal{T}|-1}_{|\mathcal{N}|} & a^{|\mathcal{T}|}_{|\mathcal{N}|} \\ d^1_{|\mathcal{N}|} & d^2_{|\mathcal{N}|} & \cdots & d^{|\mathcal{T}|-1}_{|\mathcal{N}|} & d^{|\mathcal{T}|}_{|\mathcal{N}|} \end{pmatrix}$$

the approximate counterpart of (10) can be formulated by the following linear programming

$$\begin{aligned} \min \quad & \sum_{(i,j)\in\overline{\mathcal{K}}_{tr}} c_{ij}(x_j - x_i) \\ s.t. \quad & l_{ij} \leq x_j - x_i \leq u_{ij}, (i,j) \in \overline{\mathcal{K}} \\ & 0 \leq x_i \leq m_T, i = 1, 2, \ldots, n, \end{aligned} \tag{11}$$

where $n = 2|\mathcal{T}||\mathcal{N}|$, $\overline{\mathcal{K}}$ is a subset of $\{(i, j) \mid i, j = 1, 2, \ldots, n\}$ that collects all those indices according to the constraints (1)-(8), and $\overline{\mathcal{K}}_{tr}$ is a subset of $\overline{\mathcal{K}}$ that collects all those indices according to the constraints (1)-(2). By utilizing its optimal solution $\bar{x}$, we can get $\overline{a}_j^t$'s and $\overline{d}_j^t$'s such that

$$\forall t \in \mathcal{T}, \forall (i, j) \in \mathcal{K}^t, a_j^t - d_i^t = \overline{a}_j^t - \overline{d}_i^t. \tag{12}$$

and

$$\forall (i, j) \in \phi, \forall (t, t') \in O_{i,j}, a_j^{t'} - d_i^t = \bar{a}_j^{t'} - \bar{d}_i^t. \tag{13}$$

For all other constraints, lower and upper bounds are allowed to vary as described by inequalities (3)-(9).

2.4. Maximizing the energy regeneration: A sparse optimization model

We adopt the speed profile by maximum accelerating - speed holding - maximum braking strategy from [18] with a multi-phase-speed-limit section, as described in Figs. 2.1 and 2.2.

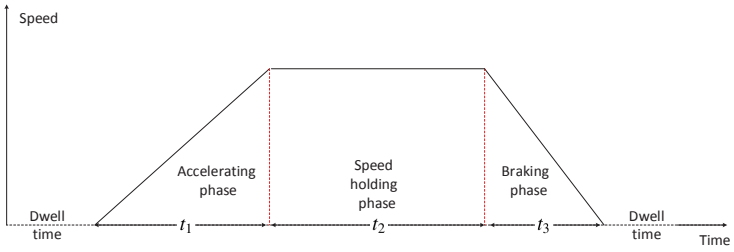

Figure 2.1: Three working phases(accelerating-cruising-braking)of speed profile.

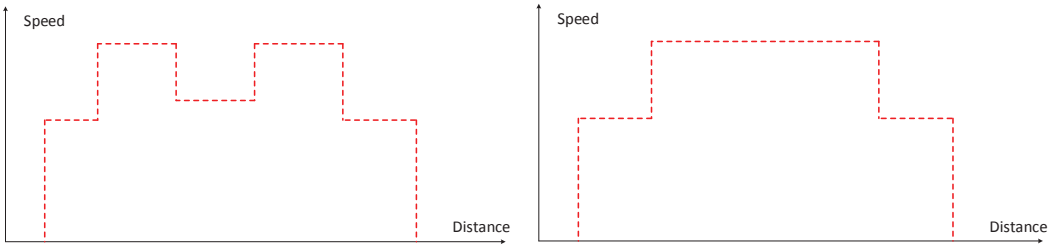

Figure 2.2: The speed limit of a train with a multi-phase section

Let t_1, t_2, and t_3 be the time of maximum accelerating, speed hold and maximum braking phases, and a_{acc} and a_{bra} are the accelerating rate of the accelerating and braking phase. It is known that with regenerative braking, kinetic energy can be converted into electricity which can be fed back to the power supply system to be used by other nearby trains in the same electric substation. Denote the conversion factor from electricity to kinetic energy as θ_{acc} and the conversion factor from kinetic energy to regenerative electricity as θ_{bra}, and the resistance rate at the speed holding phase as γ. For any two adjacent stations, we adopt the formula from [18] to calculate the energy consumption for accelerating and the regenerative energy produced during braking, denoted by E^{con} and E^{reg} respectively, in the way as below:

$$E^{con} = \frac{a_{acc}^2 t_1^2}{2\theta_{acc}} + \int_{t_1}^{t_1+t_2} \frac{\gamma^2 (t - t_1)}{\theta_{acc}} dt = \frac{a_{acc}^2 t_1^2}{2\theta_{acc}} + \frac{\gamma^2 t_2^2}{2\theta_{acc}} \quad \text{and} \quad E^{reg} = \frac{\theta_{bra} a_{bra}^2 t_3^2}{2}.$$

Here E^{con} and E^{reg} are both for the unit mass. A simple heuristic method is applied to approximate the power graph as shown in Fig.2.3, where c and $\tilde{c}$ be the width of the rectangles for accelerating and braking and h and $\tilde{h}$ are heights

of the rectangles. Similar to those in [4], the midpoint of the width in the rectangle is called the regenerative or consumptive alignment point, where Δ_i^t represents the distance between d_i^t and the regenerative alignment point and ∇_i^t represents the distance between a_i^t and the consumption alignment point.

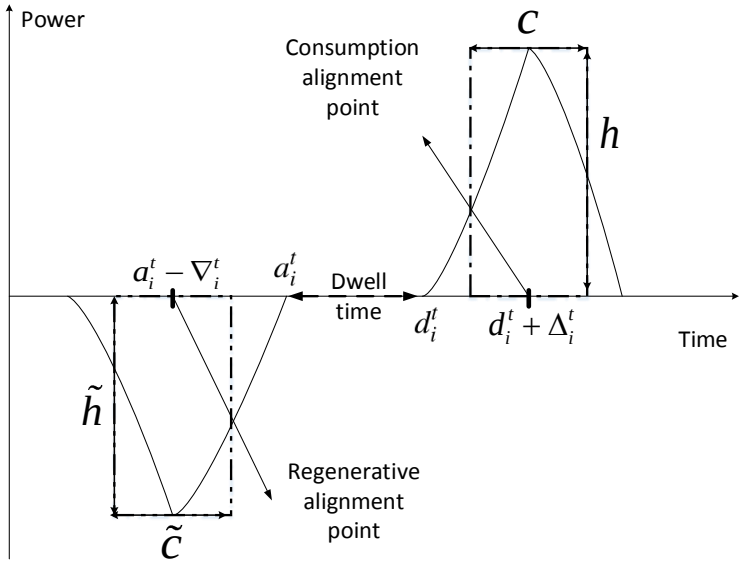

Figure 2.3: Use rectangles instead energy consumption and regenerative energy.

Here, we align the distance between regenerative and consumption alignment points to maximum the utilization. Before introducing such a distance, we recall the suitable train pairs as introduced in [4]. Assume that platform pairs which are opposite to each other at the same electrical substation. Let $\mathcal{P}$ be the set of all platform pairs. For any pair $(i, j) \in \mathcal{P}$ and the set $\mathcal{T}_i \subseteq \mathcal{T}$ contains all trains that arrives at, dwells, departs from platform i. When the train $t \in \mathcal{T}_i$ departs from the platform i, it is a best choice to find a train $\overrightarrow{t}$ at the platform j who are producing the regenerative braking energy using the following define:

Definition 1. [4] Consider any $(i, j) \in \mathcal{P}$. For every train $t \in \mathcal{T}_i$,the train $\overrightarrow{t} \in \mathcal{T}_j$ is called the temporally closest train to the right of t if

$$\overrightarrow{t} = \underset{t' \in \left\{x \in \mathcal{T}_j : 0 \le \frac{\overline{a}_j^x + \overline{d}_j^x}{2} - \frac{\overline{a}_i^t + \overline{d}_i^t}{2} \le \beta\right\}}{\arg\min} \left\{ \left| \frac{\overline{a}_i^t + \overline{d}_i^t}{2} - \frac{\overline{a}_j^{t'} + \overline{d}_j^{t'}}{2} \right| \right\}$$

where β is an empirical parameter determined by the timetable designer and is much smaller than the time horizon of the entire timetable [4]. Similarly, when a train $t \in \mathcal{T}_i$ arrives at platform i, it produces the regenerative energy. To find a suitable train $\overleftarrow{t}$ departing from platform j for transferring the regenerative braking energy, we use a define as follows:

Definition 2. [4] Consider any $(i, j) \in \mathcal{P}$. For every train $t \in \mathcal{T}_i$,the train $\overleftarrow{t} \in \mathcal{T}_j$ is called the temporally closest train to the left of t if

$$\overleftarrow{t} = \underset{t' \in \left\{x \in \mathcal{T}_j : 0 \le \frac{\overline{a}_i^t + \overline{d}_i^t}{2} - \frac{\overline{a}_j^x + \overline{d}_j^x}{2} \le \beta\right\}}{\arg\min} \left\{ \left| \frac{\overline{a}_i^t + \overline{d}_i^t}{2} - \frac{\overline{a}_j^{t'} + \overline{d}_j^{t'}}{2} \right| \right\}.$$

And we denote the sets $\overleftarrow{\varepsilon}$ and $\overrightarrow{\varepsilon}$ whose components are $(i, j, t, \overleftarrow{t})$ and $(i, j, t, \overrightarrow{t})$ respectively. Consider any $(i, j, t, \overleftarrow{t}) \in \overleftarrow{\varepsilon}$. To maximize the transfer of the regenerative energy from the braking train $\overleftarrow{t}$ by the the accelerating train t, we attempt to make the term $(d_i^t + \Delta_i^t - a_j^{\overrightarrow{t}} + \nabla_j^{\overrightarrow{t}})$ to be zero or as close to zero as possible. Similarly, for

$(i, j, t, \overrightarrow{t}) \in \overrightarrow{\varepsilon}$, our goal is to make the term $(d_j^{\overleftarrow{t}} + \Delta_j^{\overleftarrow{t}} - a_i^t + \nabla_i^t)$ to be zero or as close to zero as possible. Followed from [4], we define an auxiliary variable $y \in \mathbf{R}^n$ with $n = 2\bar{n}\bar{m}$, $\bar{n} = |\mathcal{N}|$ and $\bar{m} = |\mathcal{T}|$ by

$$y = \left((d_i^t + \Delta_i^t - a_j^{\overrightarrow{t}} + \nabla_j^{\overrightarrow{t}})_{(i,j,t,\overrightarrow{t})\in\overrightarrow{\varepsilon}}, (d_j^{\overleftarrow{t}} + \Delta_j^{\overleftarrow{t}} - a_i^t + \nabla_i^t)_{(i,j,t,\overleftarrow{t})\in\overleftarrow{\varepsilon}} \right). \tag{14}$$

The components of y represent difference of the time between the alignment points of accelerating and braking. And once the running time is fixed, Δ_i^t and ∇_i^t are two constants.

The ultimate energy consumption can be approximately calculated as follows (see Fig. 2.4)

$$E_{(i,j,t,\tilde{t})} = \begin{cases} E_k^{t,con}, & \text{if } |y_k| \geq \left(\tilde{c}_j^{\tilde{t}} + c_i^t\right)/2, \\ E_k^{t,con} - min\left\{\tilde{c}_j^{\tilde{t}}, c_i^t\right\} * min\left\{h_i^t, \tilde{h}_j^{\tilde{t}}\right\}, & \text{if } |y_k| \leq \left|\left(c_i^t - \tilde{c}_j^{\tilde{t}}\right)/2\right|, \\ E_k^{t,con} - \left(\left(\tilde{c}_j^{\tilde{t}} + c_i^t\right)/2 - y_k\right) * min\left\{h_i^t, h_j^{\tilde{t}}\right\}, & \text{otherwise.} \end{cases} \tag{15}$$

where train t is accelerating and train $\tilde{t}$ is braking.

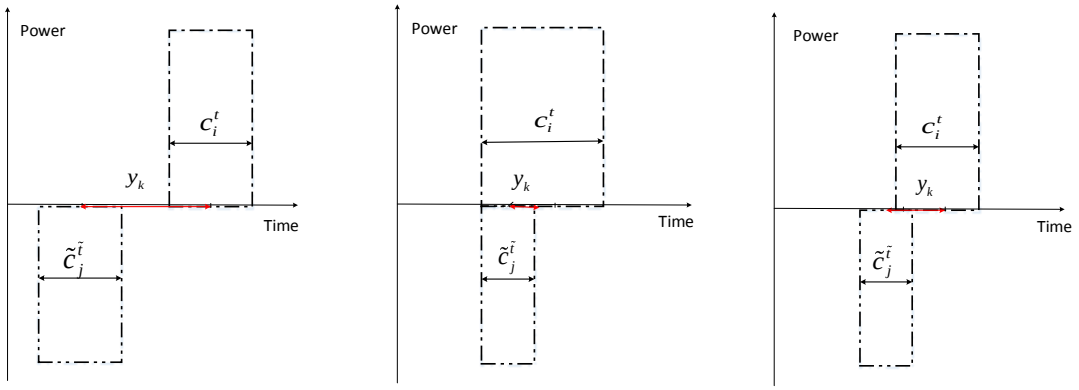

Figure 2.4: Three cases for net energy consumption function.

By sorting all possible consecutive platforms for braking energy absorbing, a 0-1 vector $\alpha \in \mathbf{R}^n$ is introduced to indicate all those locations in y with regeneration energy usage when the components of α take the value 1. In order to make the utilization of the regenerative energy as much as possible, we then turn to minimizing the number of nonzero elements of $\alpha \circ y$ which corresponds to minimizing $\|\alpha \circ y\|_0$ and making nonzero elements in $\alpha \circ y$ as close to zero as possible which corresponds to minimizing $\|\alpha \circ y\|_2^2$. The resulting sparse optimization model turns out to be

$$\begin{aligned} \min_{x,y} \quad & \lambda_1 \|\alpha \circ y\|_0 + \frac{\lambda_2}{2} \|\alpha \circ y\|_2^2 \\ s.t. \quad & (3) - (9), (12), (13), (14). \end{aligned} \tag{16}$$

where $y \in \mathbf{R}^n$ and $x \in \mathbf{R}^n$ are decision variables. Similar to (11), (16) can be rewritten in the following condensed form

$$\min_{x,y} \left\{ \lambda_1 \|\alpha \circ y\|_0 + \frac{\lambda_2}{2} \|\alpha \circ y\|_2^2 : y - Ax - b = 0, Fx - g = 0, f - Ex \geq 0. \right\} \tag{17}$$

where $A \in \mathbf{R}^{n\times n}$, $E \in \mathbf{R}^{m\times n}$, $F \in \mathbf{R}^{l\times n}$ are

$$E = \begin{pmatrix} E_1 \\ -E_1 \\ E_2 \\ -E_2 \\ E_3 \\ -E_3 \\ E_4 \\ -E_4 \end{pmatrix}, \qquad F = \begin{pmatrix} O_{\bar{m}\times\bar{m}} & \hat{E}_1 & O_{\bar{m}\times 2\bar{m}} & \dots & O_{\bar{m}\times 2\bar{m}} & O_{\bar{m}\times\bar{m}} \\ O_{\bar{m}\times\bar{m}} & O_{\bar{m}\times 2\bar{m}} & \hat{E}_2 & \dots & O_{\bar{m}\times 2\bar{m}} & O_{\bar{m}\times\bar{m}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O_{\bar{m}\times\bar{m}} & O_{\bar{m}\times 2\bar{m}} & \dots & O_{\bar{m}\times 2\bar{m}} & \hat{E}_{\bar{n}-1} & O_{\bar{m}\times\bar{m}} \end{pmatrix}$$

with

$$E_1 = \begin{pmatrix} \hat{E}_1 & O_{\bar{m}\times 2\bar{m}} & \dots & O_{\bar{m}\times 2\bar{m}} \\ O_{\bar{m}\times 2\bar{m}} & \hat{E}_2 & \dots & O_{\bar{m}\times 2\bar{m}} \\ \vdots & \vdots & \ddots & \vdots \\ O_{\bar{m}\times 2\bar{m}} & O_{\bar{m}\times 2\bar{m}} & \dots & \hat{E}_{\bar{n}} \end{pmatrix}, \quad \hat{E}_i = \begin{pmatrix} -I_{\bar{m}\times\bar{m}} & I_{\bar{m}\times\bar{m}} \end{pmatrix},$$

$$E_2 = \begin{pmatrix} -I_{\bar{m}\times\bar{m}} & O_{\bar{m}\times(2\bar{n}\bar{m}-2\bar{m})} & I_{\bar{m}\times\bar{m}} \end{pmatrix}, \quad E_3 = I_{2\bar{n}\bar{m}\times 2\bar{n}\bar{m}},$$

$$E_4 = \begin{pmatrix} \tilde{E}_1 & O_{(\bar{m}-1)\times\bar{m}} & \dots & O_{(\bar{m}-1)\times\bar{m}} \\ O_{(\bar{m}-1)\times\bar{m}} & \tilde{E}_2 & \dots & O_{(\bar{m}-1)\times\bar{m}} \\ \vdots & \vdots & \ddots & \vdots \\ O_{(\bar{m}-1)\times\bar{m}} & O_{(\bar{m}-1)\times\bar{m}} & \dots & \tilde{E}_{2\bar{n}} \end{pmatrix}, \quad \tilde{E}_i = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix},$$

$m = 2(2\bar{n}(\bar{m}-1) + \bar{n}\bar{m} + \bar{m} + n)$ and $l = (\bar{n}-1)\bar{m}$, b, g, and f are the corresponding column vectors from constraints (14), (12), (13), (3)-(9). Apparently, $y - Ax - b = 0$ is the reformulation of the constraint (14), $Fx - g = 0$ is the condensed form of constraints (12) and (13), and $f - Ex \geq 0$ is the representation of all the inequality constraints distributed in (3)-(9).

3. Model relaxation and solution algorithm

3.1. Relaxation Model

Problem (17) is generally NP-hard due to the combinatorial property of the involved ℓ_0-norm [12]. By adopting the well-known convex surrogate ℓ_1-norm initiated in compressed sensing [2, 5], we get the following $\ell_1 + \ell_2^2$ relaxation model:

$$\min_{x,y} \left\{ \lambda_1 \|\alpha \circ y\|_1 + \frac{\lambda_2}{2} \|\alpha \circ y\|_2^2 : y - Ax - b = 0, Fx - g = 0, f - Ex \geq 0. \right\} \tag{18}$$

Inspired by Oracle property beneficial from the adaptive lasso strategy proposed by Zou et al [24], we will further utilize a re-weighted ℓ_1-norm term instead of the ℓ_1-norm in (18) and get the following relaxation counterpart:

$$\min_{x,y} \left\{ \lambda_1 \|\alpha \circ w \circ y\|_1 + \frac{\lambda_2}{2} \|\alpha \circ y\|_2^2 : y - Ax - b = 0, Fx - g = 0, f - Ex \geq 0. \right\} \tag{19}$$

where w is some appropriate positive column vector to encourage a low sparsity of y.

3.2. A two-stage alternating direction method of multipliers

By introducing an auxiliary vector $z \in \mathbf{R}^m$, (19) can be rewritten as the following equality constrained optimization problem

$$\min_{x,y,z} \left\{ \lambda_1 \|\alpha \circ w \circ y\|_1 + \frac{\lambda_2}{2} \|\alpha \circ y\|_2^2 + \delta_{\mathbf{R}_+^m}(z) : y - Ax - b = 0, Fx - g = 0, z - f + Ex = 0, \right\} \tag{20}$$

where $\delta_{\mathbf{R}^m_+}(h)$ is the indicator function with respect to $\mathbf{R}^m_+$ which takes the value 0 if $h \in \mathbf{R}^m_+$, and $+\infty$ otherwise. The augmented Lagrangian function of (20) takes the form of

$$\begin{aligned} L^A_\sigma(x,y,z;u,v,s) &= \lambda_1 \sum_{i=1}^{n} w_i|\alpha_i y_i| + \frac{\lambda_2}{2}\|\alpha \circ y\|_2^2 + \delta_{\mathbf{R}^m_+}(z) - \langle u, y - Ax - b\rangle - \langle v, Fx - g\rangle \\ &\quad - \langle s, z - f + Ex\rangle + \frac{\sigma}{2}(\|y - Ax - b\|_2^2 + \|Fx - g\|_2^2 + \|z - f + Ex\|_2^2), \end{aligned}$$

where $\sigma > 0$ is the penalty parameter, and $u \in \mathbf{R}^n, v \in \mathbf{R}^l, s \in \mathbf{R}^m$ are Lagrangian multipliers corresponding to those three equality constraint systems in (20). The well-known alternating direction method of multipliers (ADMM) proposed by Glowinski and Marroco [8] and Gabay and Mercier [7] is then adopted for handling (20) with the iterative scheme for $k = 0, 1, \ldots,$

$$\begin{cases} (y^{k+1}, z^{k+1}) = \arg\min\left\{L^A_\sigma(x^k, y, z; u^k, v^k, s^k) | y \in \mathbf{R}^n, z \in \mathbf{R}^m\right\}; \\ x^{k+1} = \arg\min\left\{L^A_\sigma(x, y^{k+1}, z^{k+1}; u^k, v^k, s^k) | x \in \mathbf{R}^n\right\}; \\ \left(u^{k+1}, v^{k+1}, s^{k+1}\right) = \left(u^k, v^k, s^k\right) - \tau\sigma\left(y^{k+1} - Ax^{k+1} - b, Fx^{k+1} - g, z^{k+1} - f + Ex^{k+1}\right); \end{cases} \tag{21}$$

where $\tau \in (0, \frac{1+\sqrt{5}}{2})$ is the dual stepsize. Note that $L^A_\sigma(x, y, z; u, v, s)$ is separable of y and z. Thus the first subproblem in (21) can be easily solved by the corresponding first-order optimality condition which leads to the following closed-form solution

$$y_i^{k+1} = \begin{cases} (Ax^k + b + \frac{u^k}{\sigma})_i & \text{if } \alpha_i = 0, \\ sign(q_i)\max\{|q_i| - \frac{\lambda_1 w_i}{\lambda_2 + \sigma}, 0\} & \text{otherwise,} \end{cases} \tag{22}$$

with $q = \frac{1}{\lambda_2+\sigma}(u^k + \sigma Ax^k + \sigma b)$ for all $i = 1, \ldots, n$, and

$$z^{k+1} = \Pi_{\mathbf{R}^m_+}(f - Ex^k + \frac{1}{\sigma}s^k). \tag{23}$$

The update for x^{k+1} from the second subproblem can be easily obtained by solving the following linear system

$$Hx = rhs, \tag{24}$$

where $rhs = A^T(y^{k+1} - b - \frac{1}{\sigma}u^k) + F^T(g + \frac{1}{\sigma}v^k) + E^T(f - z^{k+1} + \frac{1}{\sigma}s^k)$ and $H = A^TA + F^TF + E^TE$. Note that E is full column rank. Thus, E^TE is positive define and also is H. This indicates that (24) admits a unique solution

$$x^{k+1} = H^{-1}rhs. \tag{25}$$

To effectively reduce the computation cost resulting from the large size of the involved matrices in H, the preconditioned conjugate gradient(PCG) method is employed to numerically solve the linear system (24) instead of the direct way as in (25).

It is worth mentioning that we can utilize the above ADMM method to solve an ℓ_1 problem which is exactly (20) by taking $w = e$ and $\lambda_2 = 0$ to a relatively low accuracy. The resulting numerical solution $(\hat{x}, \hat{y}, \hat{z})$ can not only be served as the initial point, but also produce the weight vector w in the way that $w = \left(\frac{1}{|\hat{y}_i|^r + \epsilon}\right)$ with $r \in (0, +\infty)$ and $\epsilon > 0$ sufficiently small for a low sparsity promotion when solving (20). The two-stage ADMM algorithmic framework then can be summarized in Algorithm 1.

3.3. Global convergence

Algorithm 1 admits a global convergence property as stated in the following theorem whose proof will be established in Appendix A.

Theorem 3.1. *Let* $\left\{(x^k, y^k, z^k, w)\right\}$ *be generated from the Algorithm 1. If* $\tau \in (0, \frac{1+\sqrt{5}}{2})$*, then the sequence* $\{(x^k, y^k, z^k)\}$ *converges to an optimal solution to (20).*

Algorithm 1 A two-stage alternating direction methods of multipliers (TSADMM)

Require: Choose an initial point $\left(x_0^0, y_0^0, z_0^0, u_0^0, v_0^0, s_0^0\right)$, accuracy parameters $\varepsilon_1, \varepsilon_2 > 0$, the parameters λ_1^0, λ_2^0, σ, r, ϵ.
Ensure: (x^*, y^*, z^*) ;

Stage I Compute $(\hat{x}, \hat{y}, \hat{z}, \hat{u}, \hat{v}, \hat{s})$.

Step 1.1 Set $\lambda_1 = \lambda_1^0$, $\lambda_2 = 0$, $\left(x^0, y^0, z^0, u^0, v^0, s^0\right) = \left(x_0^0, y_0^0, z_0^0, u_0^0, v_0^0, s_0^0\right)$, $w = e$ and $k = 0$;

Step 1.2 Compute $\left(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1}, v^{k+1}, s^{k+1}\right)$ by (24), (22), (23) and (21);

Step 1.3 Set $k = k + 1$. If some criterion is satisfied to the accuracy ε_1, then stop. Otherwise, go to Step 1.2.

Set $(\hat{x}, \hat{y}, \hat{z}, \hat{u}, \hat{v}, \hat{s}) = \left(x^k, y^k, z^k, u^k, v^k, s^k\right)$ and update $w = \left(\frac{1}{|\hat{y}|^r + \epsilon}\right)$;

Stage II Solve Problem (20) and output (x^*, y^*, z^*).

Step 2.1 Set $\lambda_1 = \lambda_1^0$, $\lambda_2 = \lambda_2^0$, $(x^0, y^0, z^0) = (\hat{x}, \hat{y}, \hat{z})$, $(u^0, v^0, s^0) = 0$ and $k = 0$;

Step 2.2 Compute $\left(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1}, v^{k+1}, s^{k+1}\right)$ by (24), (22), (23) and (21);

Step 2.3 Set $k = k + 1$. If some criterion is satisfied to the accuracy ε_2, then set $(x^*, y^*, z^*) = \left(x^k, y^k, z^k\right)$ and stop. Otherwise, go to Step 2.2.

4. Numerical experiments

Numerical experiments on Beijing Metro Yizhuang Line will be conducted in this section to evaluate our proposed model and to illustrate the performance our designed TSADMM algorithm. The procedure of the numerical study is shown in Fig.4.2. All the computational results are obtained by running Matlab(version 2016b) on a windows desktop(Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz 3.40GHz RAM 16.0 G).

Beijing Metro Yizhuang Line, opened on December in 2010, is one of the 20 lines in Beijing Metro Networks. It has two train-lines with 14 stations as shown in Fig 4.1. The total length of Yizhuang Line is 23.23 km, with the average distance of two station being 1.8 km, the minimum distance being 1 km, and the maximum being 2.6 km. In the section, we calculate the energy conversion when a train has a trip under unit mass.

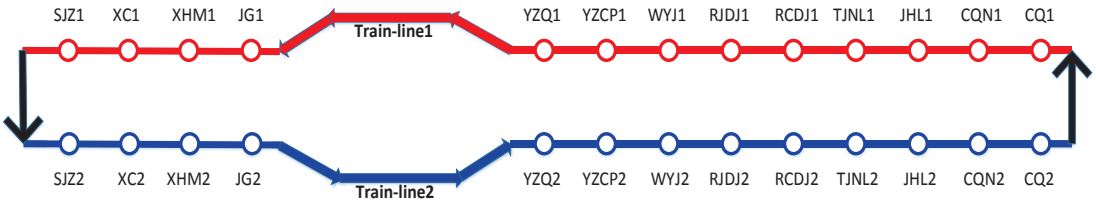

Figure 4.1: Railway network considered for numerical experiment

4.1. Inputs and algorithm initialization

For (20), $\alpha \in \mathbf{R}^n$ in the objective function is a 0-1 vector defined in Section 2.4. And the weight vector is denoted as $w = \left(\frac{1}{|\hat{y}|^r + \epsilon}\right)$, where $\hat{y}$ is an approximate solution generated by Stage I in Algorithm 1. Here we set $r = 0.5$ and $\epsilon = 10^{-20}$. In the equality constraint $y - Ax - b = 0$, $A \in \mathbf{R}^{n \times n}$ with $n = 2\bar{n}\bar{m}$, $\bar{n} = |\mathcal{N}| = 28$, $\bar{m}$ the number of trains in one day service, $b \in \mathbf{R}^n$ with components $(\Delta_i^t + \nabla_j^{\tilde{t}})_{(i,j,t,\tilde{t})}$ sorting in the appropriate way, where train t is accelerating and $\tilde{t}$ is braking, and platforms i and j are opposite to each other, satisfying the equation $i + j = 29$. $Fx - g = 0$ in (20) is an equality constraint about trip times. $F \in \mathbf{R}^{l \times m}$ is the coefficient matrix in (20), $l = (\bar{n} - 1)\bar{m}$. The components of $g \in \mathbf{R}^l$ is the trip times. $E \in \mathbf{R}^{m \times n}$ in $z - f + Ex = 0$ is the coefficient matrix of inequality constraints, and $m = 2(2\bar{n}(\bar{m} - 1) + \bar{n}\bar{m} + \bar{m} + n)$. The elements in the vector $f \in \mathbf{R}^m$ are upper and lower bounds of

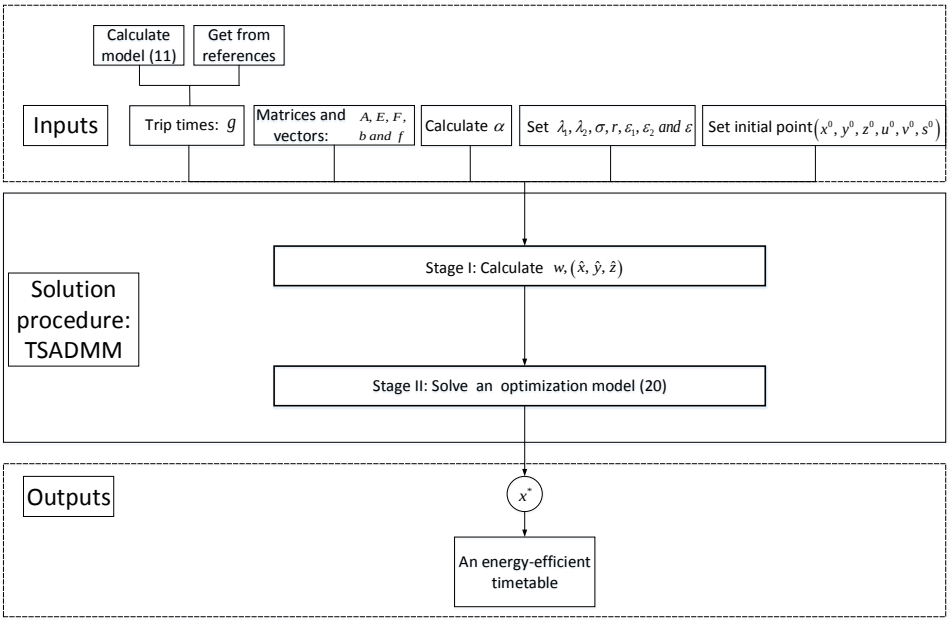

Figure 4.2: The procedure of the numerical experiments

Table 4.1: Speed limits of every section on Beijing Metro Yizhuang Line followed from [18].

Sections	Segment (m/s)	(m)	Segment (m/s)	(m)	Segment (m/s)	(m)	Segment (m/s)	(m)	Segment (m/s)	(m)
SJZ-XC	13.89	0-150	23.61	150-480	18.06	480-1161	23.61	1161-2501	16.67	2501-2631
XC-XH	16.67	2631-2643	23.61	2643-2797	20.83	2797-3534	23.61	3534-3780	16.67	3780-3905
XH-JG	16.67	3905-3918	23.61	3918-5808	20.83	5808-6141	23.61	6141-6271	-	-
JG-YZQ	16.67	6271-6281	23.61	6281-8122	16.67	8122-8254	-	-	-	-
YZQ-WH	16.67	8254-8265	23.61	8265-9116	16.67	9116-9246	-	-	-	-
WH-WY	16.67	9246-9256	23.61	9259-10655	16.67	10655-10785	-	-	-	-
WY-RJ	16.67	10785-10797	23.61	10797-11933	16.67	11933-12065	-	-	-	-
RJ-RC	16.67	12065-12077	23.61	12077-13289	16.67	13289-13419	-	-	-	-
RC-TJ	16.67	13419-13431	23.61	13419-14649	20.83	14649-15426	23.61	15426-15624	16.67	15624-15756
TJ-JH	16.67	15756-15768	23.61	15768-17891	16.67	17891-18021	-	-	-	-
JH-CQN	16.67	18021-18033	23.61	18033-19982	16.67	19982-20107	-	-	-	-
CQN-CQ	16.67	20107-20120	23.61	20120-21264	16.67	21264-21394	-	-	-	-
CQ-YZ	16.67	21394-21406	23.61	21406-22569	16.67	22569-22728	-	-	-	-

dwell time constraints, headway time constraints, total travel time constraints and domain of event times constraints. We calculate these bounds by the speed limit in Table 4.1 from [18] and Eq. (7) as references.

The initial point (u^0, v^0, s^0) in Algorithm 1 is chosen to be 0. To calculate the energy consumption and regenerative energy, we set the maximum acceleration $a_{acc} = 0.5m/s^2$ in accelerating, $a_{bra} = -0.8m/s^2$ for braking, the resistance acceleration $\gamma = -0.05m/s^2$ during the speed holding phase, the conversion factor from electricity to kinetic energy $\theta_{acc} = 0.7$ and the ratio of kinetic energy converting to regenerative energy $\theta_{bra} = 0.5$ as used in [18].

4.2. Stopping Criteria

To measure the accuracy of an approximate optimal solution (x, y, z) of Problem (20) from (x, y, z, u, v, s) generated by Stage II in Algorithm 1, the relative primal and dual infeasibilities will be adopted. Specifically, denote

$$\eta_p = \max\left\{\frac{\|y - Ax - b\|}{1 + \|y\|}, \frac{\|Fx - g\|}{1 + \|g\|}, \frac{\|z - f + Ex\|}{1 + \|z\|}\right\}, \quad \eta_d = \frac{\|A^T u - F^T w - E^T s\|}{1 + \|A^T u\| + \|F^T w\| + \|E^T s\|}. \tag{26}$$

For a given tolerance $\varepsilon_2 > 0$, we will stop the tested algorithms when both η_p and η_d are less than ε_2. The algorithms will also be terminated when they reach a given maximum number of iterations $MaxIter_2$. Here we set $\varepsilon_2 = 10^{-3}$ and

$MaxIter_2 = 10^4$. For Stage I in Algorithm 1, it will be stopped when both η_p and η_d are less than ε_1 or the maximum iteration reaches $MaxIter_1$. Here we set $\varepsilon_1 = 10^{-1}$ and $MaxIter_1 = 10^3$.

4.3. Experiment results

To evaluate our model with ℓ_0-norm plus ℓ_2^2-norm as the object, the other two possible models are computed for comparison. The first model is the ℓ_0-norm plus ℓ_1-norm proposed by Das Gupta et al. in [4] which takes the form of

$$\min_x \{\|y\|_0 + \lambda\|y\|_1 : y - Ax - b = 0, Fx - g = 0, Ex \le f\} \tag{27}$$

which is handled by approximately solving its linear programming relaxation problem

$$\min_x \left\{e^T s : y \le s, y \ge -s, y - Ax - b = 0, Fx - g = 0, Ex \le f\right\}. \tag{28}$$

The second possible model is the squared ℓ_2-norm related quadratic programming

$$\min_x \left\{\frac{1}{2}\|y\|_2^2 : y - Ax - b = 0, Fx - g = 0, Ex \le f\right\}, \tag{29}$$

where only the squared Euclidean norm is employed to measure the vanishing of y. Since we focus on the reduction of the effective energy consumption, we define the energy saving rate η as follows:

$$\eta = \frac{\sum_{t=1}^{|\mathcal{T}||\mathcal{N}|} E_{(i,j,t,\tilde{t})}}{\sum_{t=1}^{|\mathcal{T}||\mathcal{N}|} E_t^{con}}.$$

The energy savings by using these three different models are listed in Table 4.2 with different choices of $\bar{m}$.

Table 4.2: Results of numerical experiments on Beijing Metro Yizhuang Line: A1 stands for the linear programming model (28) solved by Gurobi; A2 stands for the quadratic programming (29) solved by CVX; A3 stands for our proposed model (20) solved by Algorithm 1.

	$\eta(\%)$		
$\bar{m}$	*A*1	*A*2	*A*3
357	25.54	24.52	**29.92**
368	23.67	23.44	**28.36**
380	25.56	23.01	**28.54**
391	24.50	23.80	**28.49**
403	24.78	24.17	**27.47**
414	27.02	26.57	**30.05**
426	27.73	27.07	**30.52**
437	27.59	26.94	**30.46**
449	28.26	27.54	**30.84**

As shown in Table 4.2, the energy savings by using our model are better than the other two in all the testing instances with at least 2.58 percentage higher in the energy saving rate η.

Other choices of the initial trip times from the references [16, 17, 20, 21, 22] with $\bar{m} = 368$ are also considered as inputs. The comparison results in terms of the energy saving rate η is stated in Table 4.3.

Table 4.3: $\bar{m} = 368$: A1, A2, A3 are the same as in Table 4.2

	$\eta(\%)$		
Ref.	*A*1	*A*2	*A*3
[16]	19.82	21.34	**24.94**
[17]	18.85	18.83	**23.11**
[20]	18.99	18.88	**23.11**
[21]	19.02	18.85	**23.12**
[22]	18.97	18.78	**23.29**

As one can see in Table 4.3, the utilization of regenerative energy of our approach is the best among these three models as well with the energy saving rate at least 3.60 percentage higher than the other two. These evidently show the effectiveness of our proposed model.

In order to demonstrate the efficiency of our proposed TSADMM algorithm, we call the academical CVX to solve ours model for comparison. The time comparisons for all the aforementioned instances are presented in Tables 4.4 and 4.5, respectively.

Table 4.4: Time comparison between CVX and TSADMM for solving ours model (20) with different choices of $\bar{m}$.

time(s) \ $\bar{m}$ / Alg.	357	368	380	391	403	414	426	437	449
CVX	47.17	47.06	54.33	56.90	31.09	50.63	56.18	70.72	71.14
TSADMM	24.35	12.80	31.06	21.26	25.39	19.36	27.61	30.80	23.66

Table 4.5: Time comparison between CVX and TSADMM for solving (20) with different trip times from [16, 17, 20, 21, 22] .

time(s) Ref. / Alg.	[16]	[17]	[20]	[21]	[22]
CVX	55.61	57.86	55.45	51.25	47.81
TSADMM	20.36	29.22	28.21	27.64	31.84

As can be seen in Tables 4.4 and 4.5, our proposed TSADMM outperforms the CVX solver for the model (20) in terms of computation time. Besides the theoretical global convergence of our TSADMM, the numerical convergence behaviors in terms of the infeasibility measure for all the above 14 instances are presented in Figs. 4.3 and 4.4. One can see that for each instance, the infeasibility decreases rapidly as the number of iterations increases and it meets the required accuracy 10^{-3} within less than 4000 iterations.

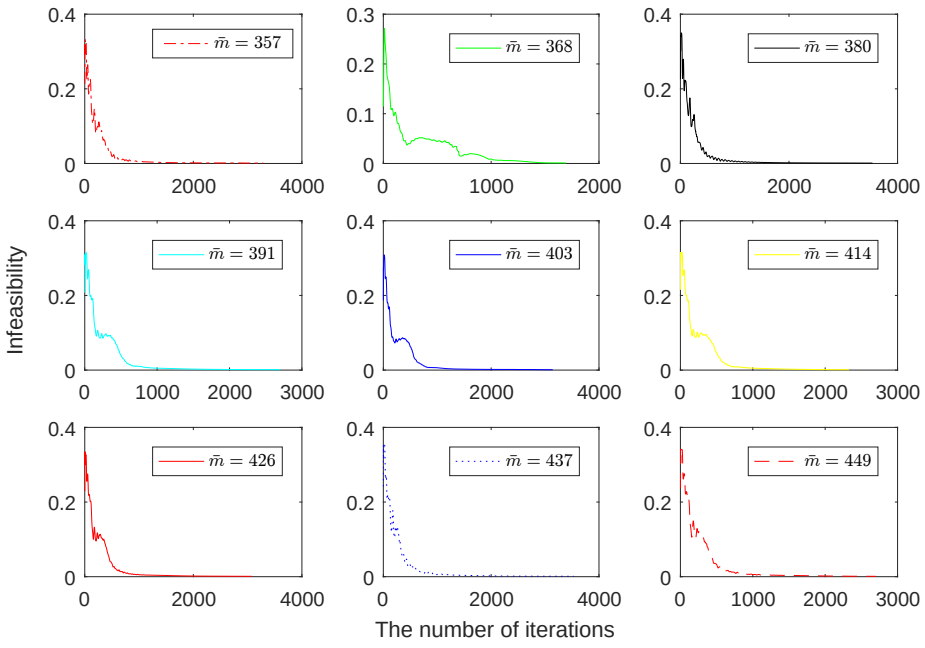

Figure 4.3: The numerical convergence behavior of TSADMM for solving (20) with different choices of $\bar{m}$

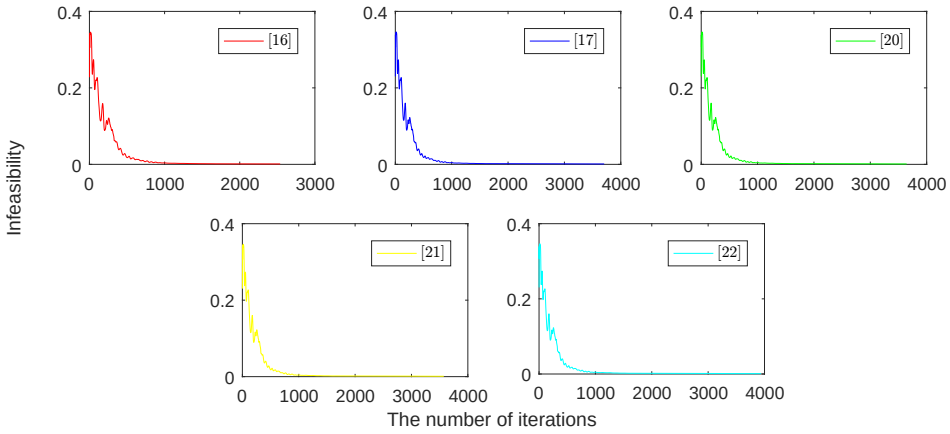

Figure 4.4: The numerical convergence behavior of TSADMM for solving our model (20) for different trip times from [16, 17, 20, 21, 22]

5. Conclusion

In this paper we have proposed a sparse optimization model using ℓ_0-norm and squared ℓ_2-norm as objective function so as to maximize the utilization of regenerative energy in metro railway system. To overcome the NP-hardness resulting from the ℓ_0-norm, the weighted ℓ_1-norm by invoking the adaptive lasso strategy in statistical regression has been introduced to make the relaxation counterpart computationally tractable and numerically effective. A two-stage alternating direction method of multipliers (TSADMM) has been designed to efficiently solve the proposed mathematical model and the global convergence of the algorithm has been established as well. Numerical experiments have been conducted on Beijing Metro Yizhuang Line for case study. The comparison results with the linear programming proposed by Das Gupta et al. and a quadratic programming with least-squares loss function have illustrated the effectiveness of our proposed model in terms of the energy saving rate, and the timing comparison results with the CVX toolbox have demonstrated the efficiency of our proposed TSADMM algorithm.

Since the weighted ℓ_1-norm is a convex approximation of the ℓ_0-norm, the optimal solution to the (20) is a proximal optimal solution to the the original sparse optimization model. How to design some more effective non-convex relaxation scheme to get a better proximal optimal solution so that to get a better energy saving rate will be one of our future work.

Acknowledgment

This work was partly supported by the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University (RCS2017ZJ001,RCS2016ZZ01), and National Natural Science Foundation of China (11771038, 11431002).

Appendix A. The Proof of Theorem 3.1

Proof. Let $\bar{x} = \left([x^{(1)}]^T, \ldots, [x^{(2\bar{n})}]^T\right)^T \in \mathbf{R}^n$ be an optimal solution of Problem (11). Then, the optimality and the feasibility of $\bar{x}$ imply that $F\bar{x} - g = 0$ and $f - E\bar{x} \geq 0$. By virtue of the expressions of E and F in Subsection 2.4, together with the time window constraints as stated in f, we can find some $\tilde{x}$ in a neighborhood $N(\bar{x}, \delta)$ with $\delta < \min_{i:f_i-(E\bar{x})_i>0} f_i - (E\bar{x})$, such that $F\tilde{x} - g = 0$ and $f - E\tilde{x} > 0$. Set $\tilde{y} = A\tilde{x} + b$ and $\tilde{z} = f - E\tilde{x}$. It is easy to verify that $(\tilde{x}, \tilde{y}, \tilde{z}) \in (\mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}^m_{++}) \cap P$ with P being the feasible region of Problem (20). By employing [6, Theorem B.1], we can get the desired global convergence result in Theorem 3.1.

References

[1] Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. Siam Journal on Imaging Sciences 2 (1), 183–202.

[2] Candes, E. J., Tao, T., 2005. Decoding by linear programming. IEEE Transactions on Information Theory 51 (12), 4203–4215.

[3] Chen, S. S., Donoho, D. L., Saunders, M. A., 2001. Atomic decomposition by basis pursuit. In: Siam J. Sci. Comput. pp. 33–61.

[4] Das Gupta, S., Tobin, J. K., Pavel, L., 2016. A two-step linear programming model for energy-efficient timetables in metro railway networks. Transportation Research Part B 93, 57–74.

[5] Donoho, D. L., 2006. Compressed sensing. IEEE Transactions on Information Theory 52 (4), 1289–1306.

[6] Fazel, M., Pong, T. K., Sun, D., Tseng, P., 2013. Hankel matrix rank minimization with applications to system identification and realization. SIAM Journal on Matrix Analysis and Applications 34 (3), 946–977.

[7] Gabay, D., Mercier, B., 1976. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Computers & Mathematics with Applications 2 (1), 17–40.

[8] Glowinski, R., Marroco, A., 1975. Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualité, d'une classe de problems de dirichlet non lineares. Revue Francaise d'Automatique, Informatique, et Recherche Opérationelle 9, 41–76.

[9] Gupta, S. D., Pavel, L., Tobin, J. K., 2015. An optimization model to utilize regenerative braking energy in a railway network. In: American Control Conference. pp. 5919–5924.

[10] Li, X., Hong, K. L., 2014. An energy-efficient scheduling and speed control approach for metro rail operations. Transportation Research Part B Methodological 64 (4), 73–89.

[11] Li, X., Sun, D., Toh, K. C., 2017. A highly efficient semismooth newton augmented lagrangian method for solving lasso problems. Siam Journal on Optimization 28 (1), 433–458.

[12] Natarajan, B. K., 1995. Sparse approximate solutions to linear systems. SIAM Journal on Computing 24 (2), 227–234.

[13] Peeters, L. W. P., 2003. Cyclic Railway Timetable Optimization. Erasmus University Rotterdam.

[14] Pena-Alcaraz, M., Fernandez, A., Cucala, A. P., Ramos, A., Pecharroman, R. R., 2011. Optimal underground timetable design based on power flow for maximizing the use of regenerative-braking energy. Proceedings of the Institution of Mechanical Engineers Part F Journal of Rail and Rapid Transit 226 (4), 397–408.

[15] Scheepmaker, G. M., Goverde, R. M. P., Kroon, L. G., 2017. Review of energy-efficient train control and timetabling. European Journal of Operational Research 257 (2), 355–376.

[16] Tian, Z., Weston, P., Zhao, N., Hillmansen, S., Roberts, C., Chen, L., 2017. System energy optimisation strategies for metros with regeneration. Transportation Research Part C Emerging Technologies 75, 120–135.

[17] Xiang, L., Hong, K. L., 2014. Energy minimization in dynamic train scheduling and control for metro rail operations. Transportation Research Part B 70, 269–284.

[18] Yang, S., Wu, J., Sun, H., Yang, X., Gao, Z., Chen, A., 2017. Bi-objective nonlinear programming with minimum energy consumption and passenger waiting time for metro systems, based on the real-world smart-card data. Transportmetrica B 2017 (4), 1–18.

[19] Yang, S., Wu, J., Yang, X., Sun, H., Gao, Z., 2018. Energy-efficient timetable and speed profile optimization with multi-phase speed limits: Theoretical analysis and application. Applied Mathematical Modelling 56, 32–50.

[20] Yang, S., Wu, J., Yang, X., Sun, H., Gao, Z., 2018. Energy-efficient timetable and speed profile optimization with multi-phase speed limits: Theoretical analysis and application. Applied Mathematical Modelling 56, 32–50.

[21] Yang, X., Chen, A., Li, X., Ning, B., Tang, T., 2015. An energy-efficient scheduling approach to improve the utilization of regenerative energy for metro systems. Transportation Research Part C Emerging Technologies 57, 13–29.

[22] Yang, X., Li, X., Gao, Z., Wang, H., Tang, T., 2013. A cooperative scheduling model for timetable optimization in subway systems. IEEE Transactions on Intelligent Transportation Systems 14 (1), 438–447.

[23] Yang, X., Ning, B., Li, X., Tang, T., 2014. A two-objective timetable optimization model in subway systems. IEEE Transactions on Intelligent Transportation Systems 15 (5), 1913–1921.

[24] Zou, H., 2006. The adaptive lasso and its Oracle properties. Publications of the American Statistical Association 101 (476), 1418–1429.