IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 14, 2019, accepted May 3, 2019, date of publication May 8, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915597

An ℓ_0-Norm Minimization for Energy- Efficient Timetabling in Subway Systems

ZIYAN LUO[1], XIAOTONG YU[2], AND XIAOYU LI[2]
[1]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
[2]Department of Mathematics, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Xiaotong Yu (18121627@bjtu.edu.cn)

This work was supported by the Fundamental Research Funds for the Central Universities of China (2019JBM078).

ABSTRACT In this paper, a sparse optimization model with ℓ_0-norm and the squared ℓ_2-norm as the objective function is proposed for energy-efficient timetabling in subway systems by means of improving the regenerative braking energy utilization. Optimality analysis is addressed for the proposed sparse optimization problem. Specifically, the local minimizer is shown to be a KKT point without any additional constraint qualification. Moreover, based on the hard-thresholding operator, we yield an explicit formula for the Lagrangian dual problem for the proposed non-convex discontinuous optimization problem and achieve the strong duality under the stationarity condition. To evaluate the effectiveness of our proposed model for the energy-efficient timetabling, we design a hard-thresholding based alternating direction method of multipliers for solving the proposed model. The case study on Beijing Metro Yizhuang Line is conducted and the comparison to some recent existing approaches illustrates the effectiveness of our model in terms of energy saving rate and the efficiency of our proposed algorithm in terms of computation time.

INDEX TERMS ℓ_0-norm minimization, hard-thresholding operator, energy-efficient timetabling, optimality analysis, alternating direction method of multipliers.

I. INTRODUCTION

With ever-increasing energy consumption in metro railway systems, energy-efficient timetabling has attracted more and more attention from both academic and industrial communities, see several survey papers [1], [27], [30], [39] and a more recent paper by Wang and Goverde [29]. It is known that the regenerative braking energy will be fed back into the overhead contact line and then be utilized for adjacent accelerating trains in a short period before being wasted by heating resistors installed on the overhead contact line [9], [21]. The net energy consumption turns out to be the difference between the total required tractive energy and the total utilization of regenerative braking energy. Thus, a special treatment on the utilization of regenerative energy has then been performed in the last decade, see, e.g., [7], [8], [16], [20], [23], [24], [37], [38], [40]. In 2008, Ramos *et al.* [24] aims to maximizing the overlapping time between speed-up and slow-down actions of all the trains circulating at any time and located in the same electrical section. Later, by maximizing the total duration of synchronization for effective acceleration and regenerative braking train pairs, Peña-Alcaraz *et al.* [23] and Das Gupta *et al.* [7] have built mixed integer programming models for energy-saving metro timetabling; By maximizing the time overlaps of nearby accelerating and braking trains, Yang *et al.* [38] have proposed an integer programming problem which was approximately solved by a genetic algorithm. Huang *et al.* [13], [14] proposed timetabling method by optimizing passenger trip time and operational energy consumption including the utilization of regenerative braking energy. Particularly, 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 braking energy was proposed by Das Gupta *et al.* [8]. The resulting linear programming problem for the second step is actually the tightest convex relaxation for the original ℓ_0-ℓ_1 norm minimization problem. By replacing the involved ℓ_0-norm with the squared ℓ_2-norm in the original model in [8], Luo *et al.* [20] proposed a two-stage alternating direction method of multipliers for the resulting convex programming problem and performed the case study on

The associate editor coordinating the review of this manuscript and approving it for publication was Huaqing Li.

TABLE 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
$\mathcal{C}$	The set of all the crossing-overs
$\mathcal{O}_{ij}$	The set of all train pairs associated with the crossing-over (i,j)
$\widetilde{\mathcal{O}}_{ij}$	The set of all two-train pairs $\big((t_1,t_1'),(t_2,t_2')\big)$ for two consecutive trains t_1 and t_2 going through a crossing-over $(i,j)\in\mathcal{C}$
$[\underline{\kappa}_{ij}^{tt'}, \overline{\kappa}_{ij}^{tt'}]$	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
$\mathcal{C}_{ij}$	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
∇_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
$\mathcal{N}$	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
$\boldsymbol{A}^\top$	The transpose of the matrix A
$vec(\boldsymbol{X})$	The column vector generated by stacking all columns of the matrix X
$\mathbb{R}_+^n$	The set of all n-dimensional nonnegative vectors
$N_{\mathbb{R}_+^m}(x)$	The normal cone at x with respect to $\mathbb{R}_+^m$
$[n]$	the set $\{1,\ldots,n\}$

Beijing Metro Yizhuang Line to illustrate the effectiveness of the proposed approach. A natural question arises: can we formulate a more appropriate ℓ_0-norm involved optimization model to improve the utilization of regenerative braking energy in metro railway system, and handle the resulting sparse optimization problem directly rather than relaxing it in an undirect way.

The main challenge comes from the discontinuity and non-convexity inherited in the ℓ_0-norm. Problems with such a norm as objective functions or constraint functions are in the range of sparse optimization. Such a type of optimization problems was popularized by the well-known compressed sensing [5], [6], [10], and found numerous applications in signal and image processing, high-dimensional statistical regression and machine learning, sparse portfolio in finance, network design, data and knowledge intelligence, etc, see [2], [10], [11], [15], [18], [28], [31] and references therein. Considerable concentrations have been attracted from theoretical and algorithmic perspectives due to the non-convexity and discontinuity inherited by ℓ_0-norm. From the algorithmic perspective, the iterative hard thresholding (IHT) methods are one of those methods that handle the ℓ_0-norm directly, rather than those relaxation schemes. The IHT-based algorithms have been proposed for solving sparsity constrained problems (i.e., the ℓ_0-norm acts as a constraint function) and ℓ_0-norm regularized optimization problems, see for example [3], [19], [44] and references therein.

Inspired by the effectiveness of IHT-based methods for ℓ_0-norm related problems, we aim to design an efficient hard-thresholding based iterative algorithm for solving a special type of ℓ_0-norm coupled with the squared ℓ_2-norm minimization problems over a convex polyhedral set, arising from the energy-efficient timetabling. This specific sparse optimization takes the form of

$$\begin{aligned} \min\ & \frac{1}{2}\|\boldsymbol{A}x-b\|_2^2+\lambda\|\boldsymbol{A}x-b\|_0 \\ \text{s.t. } & \boldsymbol{F}x-g=0 \\ & \boldsymbol{E}x-f\le 0, \end{aligned} \tag{1}$$

where $\boldsymbol{A}\in\mathbb{R}^{n\times n}$, $\boldsymbol{E}\in\mathbb{R}^{m\times n}$, $\boldsymbol{F}\in\mathbb{R}^{l\times n}$. As can be seen in (1), it is not covered by the standard form of the ℓ_0-norm regularized optimization, and hence optimality analysis in the theoretical aspect and algorithm design in the numerical aspect are required. Additionally, with the application purpose, we also need to verify the effectiveness of the model and the efficiency of the proposed algorithm, in terms of the

energy saving rate and the computation time. All of these contribute the main contents in this paper.

By employing the tools in variational analysis, along with the careful exploration to the ℓ_0-norm, we propose a first-order necessary optimality conditions in terms of KKT conditions without any constraint qualification (CQ), and a sufficient first-order optimality condition for both the sparse optimization problem and its Lagrangian dual with zero duality gap property in terms of stationarity. It is worth pointing out that even for continuous nonlinear constrained problems, the (local) minimizer is not always guaranteed to be accessible to the KKT conditions unless some CQs are satisfied, such as the Robinson CQ, or more specifically the MFCQ or LICQ (see, [4]). All of these theoretical results tailored for problem (1) are then reasonably regarded as a supplement or a refinement in non-convex optimization theory. By taking the benefits of the hard-thresholding operator and the related scheme in IHT, we propose a hard-thresholding based alternating direction method of multipliers. For the purpose of effectiveness evaluation of our approach for energy-efficient timetabling, we benchmark it against several recent related approaches, including the two-step linear programming approach in [8] with Gurobi optimizer and the ℓ_1-ℓ_2^2 model with the proposed TSADMM in [20]. The non-convex half-thresholding relaxation approach (i.e., the square root $\ell_{1/2}$-norm, which is shown to be more sparsity promoting than ℓ_1-norm for ℓ_0-norm relaxation [34]) is also proposed for (1) for comparison. Numerical experiments are conducted as case study for the Beijing Yizhuang Line timetabling, in which the effectiveness of our approach is illustrated in terms of the energy saving rates and computation time for different service instances.

The paper is organized as follows. In Section II, an ℓ_0-ℓ_2^2 sparse optimization model is introduced for maximizing the utilization of the regenerating braking energy in subway systems. In Section III, the optimality conditions are established for the proposed ℓ_0-norm minimization problem. A hard-thresholding based alternating direction method of multipliers (HTADMM) is proposed to get an approximate optimal solution in Section IV. As a case study, our proposed approach is applied to Beijing Metro Yizhuang Line for illustrating the effectiveness and the efficiency of our proposed HTADMM in Section V. Conclusions and future research are given in Section VI.

II. SPARSE OPTIMIZATION MODELLING

This section is devoted to mathematical modeling for the energy-efficient timetable for metro railway system, with a special treatment on how to maximize the utilization of the regenerative braking energy in the system. Before proceeding, some notations are listed in Table 1 for the convenience of the subsequent model formulation.

The set of basic 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 are listed in Table 2, followed from [8].

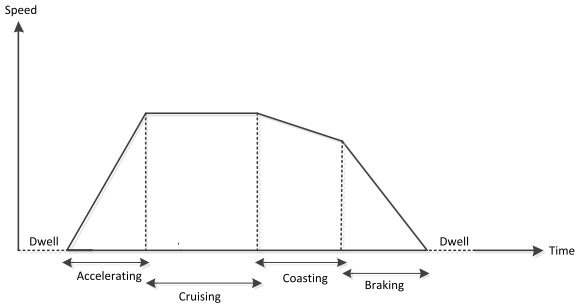

FIGURE 1. The accelerating-cruising-coasting-braking speed profile.

Similar to [43], we adopt the speed profile which involves four steps: accelerating, cruising, coasting and braking with a multi-phase-speed-limit section from [35], as described in Figures 1 and 2.

As the majority of power is consumed in the acceleration phase, Das Gupta [8] proposed a nonlinear programming problem with basic constraints distributed in Table 2 to efficiently minimize the total energy consumption of trains in a metro railway network. The corresponding optimization model takes the form of

$$\begin{aligned} \min \quad & \sum_{(i,j)\in\mathcal{K}_{tr},t\in\mathcal{T}} f_{ij}(a_j^t - d_i^t) \\ \text{s.t.} \quad & \text{all constraints in Table 2} \end{aligned} \tag{2}$$

where the objective function $f_{ij}:\mathbb{R}_{++}\to\mathbb{R}_{++}$ is the energy consumption function. Some appropriate estimation for such an unknown function is obtained by sampling in the sense of least-squares. We can introduce a matrix $\boldsymbol{X}$ of size $2\mathcal{N}\times\mathcal{T}$ with its (i,j)-th entry defined as

$$X_{ij} = \begin{cases} a^j_{(i+1)/2}, & \text{if } i \text{is} odd; \\ d^j_{i/2}, & \text{otherwise.} \end{cases} \tag{3}$$

An approximate counterpart of (2) can be then obtained in terms of $\sum_{(i,j)\in\overline{\mathcal{K}}_{tr}} c_{ij}(x_j - x_i)$ where c_{ij}'s are obtained from least-squares estimator and x is the vectorization of the matrix $\boldsymbol{X}$, i.e., $x = vec(\boldsymbol{X})$. By rewriting all the involved constraints in terms of x, a linear programming is built up to achieve a feasible timetable,

$$\begin{aligned} \min \quad & \sum_{(i,j)\in\overline{\mathcal{K}}_{tr}} c_{ij}(x_j - x_i) \\ s.t. \quad & l_{ij} \le x_j - x_i \le u_{ij}, \quad (i,j)\in\overline{\mathcal{K}} \\ & 0 \le x_i \le m_T, \quad i = 1,2,\ldots,n, \end{aligned} \tag{4}$$

where $n = 2|\mathcal{T}||\mathcal{N}|$, $\overline{\mathcal{K}}$ is the index set that collects all those indices according to all the constraints in Table 2 except the domain of event time constraints as listed, and $\overline{\mathcal{K}}_{tr}$ is a subset of $\overline{\mathcal{K}}$ that collects all those indices according to the trip time constraints. If $\bar{x}$ is the optimal solution to problem (4), then by the definition of $\boldsymbol{X}$ as shown in (3), we can obtain $\overline{a}_j^t$'s and $\overline{d}_j^t$'s such that

$$\forall t\in\mathcal{T}, \quad \forall (i,j)\in\mathcal{K}^t, \quad a_j^t - d_i^t = \overline{a}_j^t - \overline{d}_i^t \tag{5}$$

TABLE 2. The set of constraints.

Trip time constraint:	$\forall t \in \mathcal{T}, \forall (i,j) \in \mathcal{K}^t, \underline{tr}_{ij}^t \le a_j^t - d_i^t \le \overline{tr}_{ij}^t.$ $\forall (i,j) \in \mathcal{C}, \forall (t,t') \in \mathcal{O}_{i,j}, \underline{\kappa}_{ij}^{tt'} \le a_j^{t'} - d_i^t \le \overline{\kappa}_{ij}^{tt'}.$
Dwell time constraint:	$\forall t \in \mathcal{T}, \forall i \in \mathcal{N}^t, \underline{dw}_i^t \le d_i^t - a_i^t \le \overline{dw}_i^t.$
Connection constraint:	$\forall (i,j) \in \mathcal{S}, \forall (t,t') \in \mathcal{C}_{ij}, \underline{c}_{ij}^{tt'} \le d_j^{t'} - a_i^t \le \overline{c}_{ij}^{tt'}.$
Headway time constraint:	$\forall (i,j) \in \mathcal{K}, (t,t') \in \mathcal{H}_{ij}, \begin{cases} \underline{h}_i^{tt'} \le d_i^t - d_i^{t'} \le \overline{h}_i^{tt'}, \\ \underline{h}_j^{tt'} \le a_j^t - a_j^{t'} \le \overline{h}_j^{tt'}. \end{cases}$ $\forall (i,j) \in \mathcal{C}, \forall ((t_1,t_1'),(t_2,t_2')) \in \widetilde{\mathcal{O}}_{ij}, \begin{cases} \underline{h}_i^{t_1 t_2} \le d_i^{t_1} - d_i^{t_2} \le \overline{h}_i^{t_1 t_2}, \\ \underline{h}_j^{t_1' t_2'} \le a_j^{t_1'} - a_j^{t_2'} \le \overline{h}_j^{t_1' t_2'}. \end{cases}$
Total travel time constraint:	$\forall t \in \mathcal{T}, \underline{tt}^t \le d^t_{\mathcal{N}^t(\lvert\mathcal{N}^t\rvert)} - a^t_{\mathcal{N}^t(1)} \le \overline{tt}^t$
Domain of event times:	$\forall t \in \mathcal{T}, \forall i \in \mathcal{N}^t, 0 \le a_i^t \le m_T, 0 \le d_i^t \le m_T.$

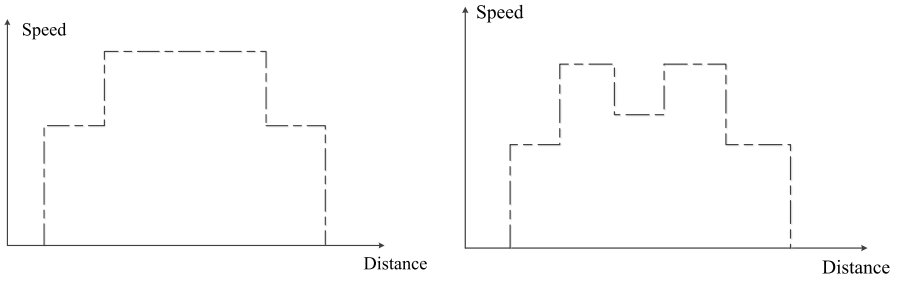

FIGURE 2. The speed limit of a train with a multi-phase section.

$$\forall (i,j) \in \mathcal{C}, \quad \forall (t,t') \in \mathcal{O}_{i,j}, \quad a_j^{t'} - d_i^t = \bar{a}_j^{t'} - \bar{d}_i^t. \quad (6)$$

This provides us a feasible timetable which admits a small energy consumption in the accelerating phases, regardless of the utilization of regenerative braking energy.

Following from the modelling approach as introduced in [20], we use t_1, t_2, t_3 and t_4 to be the time of maximum accelerating, cruising, coasting and maximum braking phases, respectively. In the numerical experiments, we will assign t_i's the percentages to get the corresponding time by multiplying the percentage with the running time in the corresponding period. a_{acc} and a_{bra} are given which stand for the accelerating rates of the accelerating and braking phase. The conversion factor from electricity to kinetic energy and the conversion factor from kinetic energy to regenerative electricity are denoted by θ_{acc} and θ_{bra}, respectively. The resistance rate at the cruising phase is denoted by γ. For any two adjacent stations, the energy consumption for accelerating and the regenerative energy produced during braking, denoted by E^{con} and E^{reg} respectively, are calculated with the following formulas for unit mass [35]

$$E^{con} = \frac{a_{acc}^2 t_1^2}{2\theta_{acc}} + \int_{t_1}^{t_1+t_2} \frac{\gamma^2 (t - t_1)}{\theta_{acc}} dt = \frac{a_{acc}^2 t_1^2}{2\theta_{acc}} + \frac{\gamma^2 t_2^2}{2\theta_{acc}}$$

and

$$E^{reg} = \frac{\theta_{bra} a_{bra}^2 t_4^2}{2}.$$

We adopt a simple heuristic method to approximately calculate the power, see Figure 3. Here c and c' are the widths of the rectangles for accelerating and braking phases, h and h' are heights of the rectangles. Following from [8], the midpoint of the width in the first rectangle is called the regenerative alignment point (RAP), and the midpoint of the width in the second rectangle is called the consumption alignment point (CAP), termed as $a_i^t - \nabla_i^t$ and $d_i^t + \Delta_i^t$ respectively, with ∇_i^t representing the distance between a_i^t and CAP and Δ_i^t representing the distance between d_i^t and RAP.

In electric railway transportation systems, electric current is transmitted through the overhead contact line to the train and returned to the traction power substation through running rails [42] as shown in Figure 4. Due to the electrical resistance of the overhead contact line [9], [21] and imperfect insolution to the ground [22], a part of the current drawn by the vehicles from the traction power substation will leak out of the current return circuit. Based on this, we simply use

$$\bar{E}_t^{con} = \bar{\eta} E_t^{con}, \quad \text{with } \bar{\eta} = 110\% \quad (7)$$

to characterize the total energy consumption with considering the factor of overhead contact line loss and stray current.

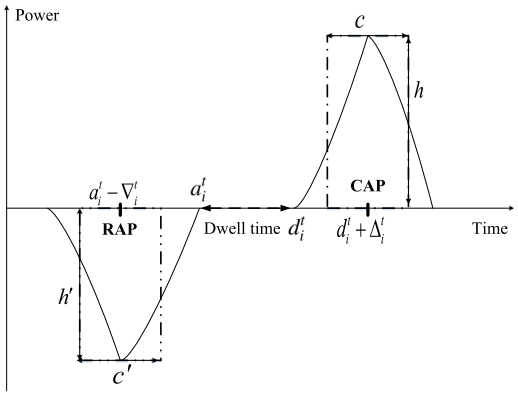

FIGURE 3. Approximate calculation of energy.

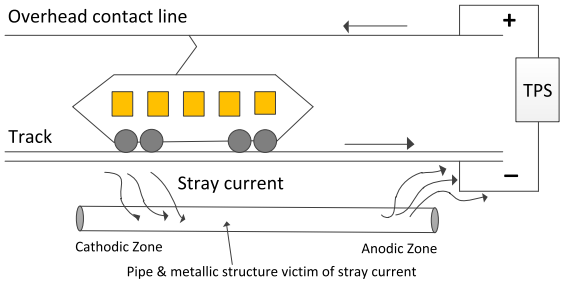

FIGURE 4. Electric traction system.

It is known that, with regenerative braking, kinetic energy can be converted into electricity which can be fed back to the power supply system so that can be used by other nearby trains in the same electric substation. Then to characterize the utilization of the regenerative braking energy, it is reasonable to consider the distance between regenerative and consumption alignment points for possible train pairs. Recalling the suitable train pairs as introduced in [8], we assume that platform pairs which are opposite to each other are 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 arrive at, dwell, depart from platform i. When the train $t \in \mathcal{T}_i$ departs from (or arrives at) the platform i, it is a best choice to find a train $\overrightarrow{t}$ at the platform j which is producing (or absorbing) the regenerative braking energy in the following way.

Definition 1: (See [8, Definitions 1,2]) Consider any $(i,j) \in \mathcal{P}$.

(i) 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

$$\overrightarrow{t} = \arg\min_{t' \in \overrightarrow{\Omega}_t} \left\{ \left| \frac{\overline{a}_i^t + \overline{d}_i^t}{2} - \frac{\overline{a}_j^{t'} + \overline{d}_j^{t'}}{2} \right| \right\}$$

where $\overrightarrow{\Omega}_t = \left\{ x \in \mathcal{T}_j : 0 \le \frac{\overline{a}_j^x + \overline{d}_j^x}{2} - \frac{\overline{a}_i^t + \overline{d}_i^t}{2} \le \beta \right\}$, β is an empirical parameter determined by the timetable designer and is much smaller than the time horizon of the entire timetable.

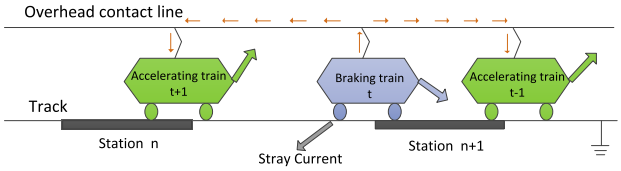

FIGURE 5. Trains operation scenario.

(ii) 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

$$\overleftarrow{t} = \arg\min_{t' \in \overleftarrow{\Omega}_t} \left\{ \left| \frac{\overline{a}_i^t + \overline{d}_i^t}{2} - \frac{\overline{a}_j^{t'} + \overline{d}_j^{t'}}{2} \right| \right\}.$$

$$\text{with } \overleftarrow{\Omega}_t = \left\{ x \in \mathcal{T}_j : 0 \le \frac{\overline{a}_i^t + \overline{d}_i^t}{2} - \frac{\overline{a}_j^x + \overline{d}_j^x}{2} \le \beta \right\}.$$

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 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 [8], we define an auxiliary variable $y \in \mathbb{R}^n$ with $n = 2\bar{n}\bar{m}$, $\bar{n} = |\mathcal{N}|$ and $\bar{m} = |\mathcal{T}|$ by

$$y = \begin{pmatrix} (d_i^t + \Delta_i^t - a_j^{\overrightarrow{t}} + \nabla_j^{\overrightarrow{t}}) \\ (d_j^{\overleftarrow{t}} + \Delta_j^{\overleftarrow{t}} - a_i^t + \nabla_i^t) \end{pmatrix}_{(i,j,t,\overrightarrow{t}) \in \overrightarrow{\varepsilon}, (i,j,t,\overleftarrow{t}) \in \overleftarrow{\varepsilon}}. \quad (8)$$

Here components of y represent difference of the time between the alignment points in accelerating and braking phases. Once the running time is fixed, Δ_i^t and ∇_i^t are then determined.

It is reasonable to approximately calculate such a regenerative energy in terms of y as follows: Given stations i, j, and trains t, $\tilde{t}$, $\hat{E}_{reg}(i,j,t,\tilde{t}) = 0$ if $|y_k| \ge \left(\tilde{c}_j^{\tilde{t}} + c_i^t\right)/2$; $\hat{E}_{reg}(i,j,t,\tilde{t}) = \min\left\{\tilde{c}_j^{\tilde{t}}, c_i^t\right\} * min\left\{h_i^t, \tilde{h}_j^{\tilde{t}}\right\}$ if $|y_k| \le \left|\left(c_i^t - \tilde{c}_j^{\tilde{t}}\right)/2\right|$; and $\hat{E}_{reg}(i,j,t,\tilde{t}) = \left(\left(\tilde{c}_j^{\tilde{t}} + c_i^t\right)/2 - y_k\right) * min\left\{h_i^t, h_j^{\tilde{t}}\right\}$ otherwise. Generally, we will ignore the transmission losses of electricity since the transmission distance is short between the successive trains (see, e.g., [38]). In addition, with the consideration of stray current (see, Figure 5), we assume that 95% of the regenerative braking current is transmitted to the overhead contact line for nearby accelerating train, and 5% of them becomes stray current, following from [22], [42]. Thus, the regenerative braking energy that can be utilized turns out to be

$$\bar{E}_{reg}(i,j,t,\tilde{t}) = \bar{\tau}\hat{E}_{reg}(i,j,t,\tilde{t}), \quad \text{with } \bar{\tau} = 95\%. \quad (9)$$

Then the energy saving rate r, adopted as a criterion to evaluate the approach for energy-efficient timetabling, is defined

by

$$r = \frac{\sum_{(i,j,t,\tilde{t})\in \overrightarrow{\mathcal{E}}\cup\overleftarrow{\mathcal{E}}} \bar{E}_{reg}(i,j,t,\tilde{t})}{\sum_{t=1}^{|\mathcal{T}||\mathcal{N}|}\bar{E}_t^{con}} \times 100\%, \tag{10}$$

where $\bar{E}_t^{con}$ and $\bar{E}_{reg}(i,j,t,\tilde{t})$ are defined as in (7) and (9), respectively.

In order to maximize the utilization of the regenerative braking energy for a higher energy saving rate, we adopt the following ℓ_0-ℓ_2^2 optimization model from [20]:

$$\begin{aligned} \min\ & \frac{1}{2}\|y\|_2^2 + \lambda\|y\|_0 \\ \text{s.t. } & y - \boldsymbol{A}x + b = 0 \\ & \boldsymbol{F}x - g = 0 \\ & \boldsymbol{E}x - f \le 0. \end{aligned} \tag{11}$$

Here $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{E} \in \mathbb{R}^{m\times n}$, $\boldsymbol{F} \in \mathbb{R}^{l\times n}$ with $m = 2(2\bar{n}(\bar{m}-1)+\bar{n}\bar{m}+\bar{m}+n)$ and $l = (\bar{n}-1)\bar{m}$. Specifically,

$$\boldsymbol{A} = \left(\boldsymbol{A^{(1)}}^T, \boldsymbol{A^{(2)}}^T, \cdots, \boldsymbol{A^{(k)}}^T, \cdots, \boldsymbol{A}^{(2\bar{n}-1)^T}, \boldsymbol{A}^{(2\bar{n})^T}\right)^T \tag{12}$$

where $\boldsymbol{A^{(k_i)}} \in \mathbb{R}^{\bar{m}\times 2\bar{n}\bar{m}}$, with all odd integers k_0 and all even integers k_1 in $\{1,\ldots,2\bar{n}\}$, defined as below : for any given $i = (k_0-1)\bar{m}+t, t = 1,\cdots,\bar{m}$,

$$A_{ij}^{(k_0)} = \begin{cases} -1 & \text{if } j = i, \\ 1 & \text{if } j = (2\bar{n}-k_0)\bar{m} + \overleftarrow{t}, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

with $\overleftarrow{t}$ as defined in Definition 1; and for any given $i = (k_1-1)\bar{m}+t, t = 1,2,\ldots,\bar{m}$,

$$A_{ij}^{(k_1)} = \begin{cases} 1 & \text{if } j = i, \\ -1 & \text{if } j = (2\bar{n}-k_1)\bar{m} + \overrightarrow{t}, \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

with $\overrightarrow{t}$ as defined in Definition 1.

$$\boldsymbol{E} = \left(\boldsymbol{E_1}^\top, -\boldsymbol{E_1}^\top, \boldsymbol{E_2}^\top, -\boldsymbol{E_2}^\top, \boldsymbol{E_3}^\top, -\boldsymbol{E_3}^\top, \boldsymbol{E_4}^\top, -\boldsymbol{E_4}^\top\right)^\top, \tag{15}$$

$$\boldsymbol{F} = \begin{pmatrix} O_{\bar{m}\times\bar{m}} & \hat{E}_1 & O_{\bar{m}\times 2\bar{m}} & \cdots & O_{\bar{m}\times 2\bar{m}} & O_{\bar{m}\times\bar{m}} \\ O_{\bar{m}\times\bar{m}} & O_{\bar{m}\times 2\bar{m}} & \hat{E}_2 & \cdots & O_{\bar{m}\times 2\bar{m}} & O_{\bar{m}\times\bar{m}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O_{\bar{m}\times\bar{m}} & O_{\bar{m}\times 2\bar{m}} & \cdots & O_{\bar{m}\times 2\bar{m}} & \hat{E}_{\bar{n}-1} & O_{\bar{m}\times\bar{m}} \end{pmatrix} \tag{16}$$

with

$$\boldsymbol{E_1} = \begin{pmatrix} \hat{E}_1 & O_{\bar{m}\times 2\bar{m}} & \cdots & O_{\bar{m}\times 2\bar{m}} \\ O_{\bar{m}\times 2\bar{m}} & \hat{E}_2 & \cdots & O_{\bar{m}\times 2\bar{m}} \\ \vdots & \vdots & \ddots & \vdots \\ O_{\bar{m}\times 2\bar{m}} & O_{\bar{m}\times 2\bar{m}} & \cdots & \hat{E}_{\bar{n}} \end{pmatrix},$$

$$\hat{\boldsymbol{E}}_i = \left(-I_{\bar{m}\times\bar{m}}\ I_{\bar{m}\times\bar{m}}\right),$$

$$\boldsymbol{E_2} = \left(-I_{\bar{m}\times\bar{m}}\ O_{\bar{m}\times(2\bar{n}\bar{m}-2\bar{m})}\ I_{\bar{m}\times\bar{m}}\right),$$

$$\boldsymbol{E_3} = \boldsymbol{I}_{2\bar{n}\bar{m}\times 2\bar{n}\bar{m}},$$

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

$$\tilde{\boldsymbol{E}}_i = \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -1 & 1 \end{pmatrix}.$$

b, g, and f are the corresponding column vectors from constraints (8), (5), (6) and all the constraints listed in Table 2. Indeed, the constraint $y - \boldsymbol{A}x + b = 0$ represents the constraint (8), $\boldsymbol{F}x - g = 0$ collects the constraints (5) and (6), and $f - \boldsymbol{E}x \ge 0$ stands for all the inequality constraints distributed in Table 2.

As above, we will use the model (11) to improve the utilization of regenerative braking energy and adopt the energy saving rate r defined in (10) to measure the energy saving efficiency.

III. OPTIMALITY ANALYSIS

Before proposing the solution approach, the optimality analysis for the non-convex discontinuous optimization problem (11) is elaborated in this section.

A. KKT CONDITIONS AND STATIONARITY

Problem (11) can be equivalently rewritten as

$$\begin{aligned} \min\ & \frac{1}{2}\|y\|_2^2 + \lambda\|y\|_0 + \delta_{\mathbb{R}_+^m}(z) \\ \text{s.t. } & y - \boldsymbol{A}x + b = 0 \\ & \boldsymbol{F}x - g = 0 \\ & \boldsymbol{E}x - f + z = 0. \end{aligned} \tag{17}$$

The Lagrangian function associated with problem (17) is defined by

$$\begin{aligned} & L(x,y,z;u,v,w): \\ & \quad = \frac{1}{2}\|y\|^2 + \lambda\|y\|_0 + \delta_{\mathbb{R}_+^m}(z) + u^\top(y - \boldsymbol{A}x + b) \\ & \qquad + v^\top(\boldsymbol{F}x - g) + w^\top(\boldsymbol{E}x - f + z), \end{aligned} \tag{18}$$

for all $(x,y,z,u,v,w) \in \mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^l\times\mathbb{R}^m$. The KKT conditions for problem (17) can be expressed as

$$\begin{cases} 0 = \nabla_x L(x,y,z,u,v,w) = -\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w, \\ 0 \in \partial_y L(x,y,z,u,v,w) = y + u + \lambda\partial\|y\|_0, \\ 0 \in \partial_z L(x,y,z,u,v,w) = N_{\mathbb{R}_+^m}(z) + w, \\ y - \boldsymbol{A}x + b = 0, \\ \boldsymbol{F}x - g = 0, \\ \boldsymbol{E}x - f + z = 0, \end{cases} \tag{19}$$

where $\partial\|y\|_0$ is the subdifferential of $\|\cdot\|_0$ at y which takes the form of

$$\partial\|y\|_0 = \left\{h \in \mathbb{R}^n \mid h_i = 0, \text{if } y_i \ne 0\right\} =: \mathbb{R}^n_{[n]\setminus I_y}, \tag{20}$$

with $I_y := \{i \in [n] \mid y_i \neq 0\}$ being the supporting set of y. A first-order necessary optimality condition via KKT conditions can then be established in the following theorem.

Theorem 1: If $(\bar{x}, \bar{y}, \bar{z})$ is a local minimizer of problem (17), then there exists $(\bar{u}, \bar{v}, \bar{w})$ such that $(\bar{x}, \bar{y}, \bar{z}, \bar{u}, \bar{v}, \bar{w})$ satisfies the KKT conditions in (19).

Proof. The optimality of $(\bar{x}, \bar{y}, \bar{z})$ leads to its feasibility to problem (17), i.e.,

$$(\bar{x}, \bar{y}) \in \Omega := \left\{ (x, y) \middle| \begin{array}{l} y - \boldsymbol{A}x + b = 0 \\ \boldsymbol{F}x - g = 0 \\ \boldsymbol{E}x - f \leq 0 \end{array} \right\}, \text{ and } \bar{z} = f - \boldsymbol{E}\bar{x}.$$

Next, we claim that for any $(x, y) \in \Omega$, the regular subdifferential (see, [26, Definition 8.3])

$$\hat{\partial}\left(\lambda\|y\|_0 + \delta_\Omega(x, y)\right) = \{0\}_{\mathbb{R}^m} \times (\lambda\partial\|y\|_0) + N_\Omega(x, y), \quad (21)$$

where $N_\Omega(x, y)$ is the normal cone to Ω at (x, y). By virtue of [26, Corollary 10.9], we get the inclusion

$$\hat{\partial}\left(\lambda\|y\|_0 + \delta_\Omega(x, y)\right) \supseteq \{0\}_{\mathbb{R}^m} \times (\lambda\hat{\partial}\|y\|_0) + N_\Omega(x, y). \quad (22)$$

It then remains to show

$$\hat{\partial}\left(\lambda\|y\|_0 + \delta_\Omega(x, y)\right) \subseteq \{0\}_{\mathbb{R}^m} \times (\lambda\hat{\partial}\|y\|_0) + N_\Omega(x, y). \quad (23)$$

For any given $(v_1, v_2) \in \hat{\partial}\left(\lambda\|y\|_0 + \delta_\Omega(x, y)\right)$, it follows from the defintion of regular subdifferential as stated in [26, Definition 8.3]) that for any $(x', y') \to (x, y)$, we have

$$\begin{aligned} o(\|(x', y') - (x, y)\|) \geq\ & \|y'\|_0 + \delta_\Omega(x', y') \\ & - \|y\|_0 - \delta_\Omega(x, y) \\ & - v_1^\top(x' - x) - v_2^\top(y' - y). \end{aligned} \quad (24)$$

Particularly, if we choose any $(\tilde{x}, \tilde{y}) \to (x, y)$ satisfying $(\tilde{x}, \tilde{y}) \in \Omega$ and $\tilde{y} \in \mathbb{R}^n_{I_y}$, where $I_y := \{i \mid y_i \neq 0\}$ and $\mathbb{R}^n_{I_y} := \{y \in \mathbb{R}^n \mid y_i = 0, \text{if } i \notin I_y.\}$. Then Equation (24) should also hold for all $(\tilde{x}, \tilde{y}) \in \Omega \cap (\mathbb{R}^n \times \mathbb{R}^n_{I_y})$. Noting that $\tilde{y} \to y$ and $\tilde{y} \in \mathbb{R}^n_{I_y}$, we have $\|\tilde{y}\|_0 = \|y\|_0$. Thus, (24) turns out to be

$$o(\|(\tilde{x}, \tilde{y}) - (x, y)\|) \leq -v_1^\top(\tilde{x} - x) - v_2^\top(\tilde{y} - y), \quad (25)$$

which indicates that $(v_1, v_2) \in N_{\Omega\cap\left(\mathbb{R}^n\times\mathbb{R}^n_{I_y}\right)}(x, y)$. Note that Ω and $\mathbb{R}^n \times \mathbb{R}^n_{I_y}$ are convex polyheral sets. It follows from [25, Corollary 23.8.1] that

$$N_{\Omega\cap\left(\mathbb{R}^n\times\mathbb{R}^n_{I_y}\right)}(x, y) = N_\Omega(x, y) + N_{\mathbb{R}^n\times\mathbb{R}^n_{I_y}}(x, y).$$

In virtue of the observation $\lambda\hat{\partial}\|y\|_0 = \lambda\partial\|y\|_0$, together with (20), the claim is proved. By reformulating problem (17) as

$$\min_{x,y} \frac{1}{2}\|y\|^2 + \lambda\|y\|_0 + \delta_\Omega(x, y),$$

we can apply [26, Theorem 10.1] to get the following first-order optimality condition

$$\begin{aligned} (0, 0) &\in \hat{\partial}\left(\frac{1}{2}\|\bar{y}\|^2 + \lambda\|\bar{y}\|_0 + \delta_\Omega(\bar{x}, \bar{y})\right) \\ &= (0, \bar{y}) + \hat{\partial}\left(\lambda\|\bar{y}\|_0 + \delta_\Omega(\bar{x}, \bar{y})\right) \\ &= (0, \bar{y}) + N_\Omega(\bar{x}, \bar{y}) + N_{\mathbb{R}^n\times\mathbb{R}^n_{I_y}}(\bar{x}, \bar{y}) \\ &= (0, \bar{y}) + \{(\bar{\xi}, \bar{u}) \mid \boldsymbol{A}^\top\bar{u} + \bar{\xi} = \boldsymbol{E}^\top\bar{w} + \boldsymbol{F}^\top\bar{v}, \\ &\quad \exists\bar{w} \in N_{\mathbb{R}^m_-}(\boldsymbol{E}\bar{x} - f), \bar{v} \in \mathbb{R}^l\} + \{0\}_{\mathbb{R}^n} \times \mathbb{R}^n_{[n]\setminus\bar{I}}, \end{aligned} \quad (26)$$

where $\bar{I} := \{i \mid \bar{y}_i \neq 0\}$. Here the first equality follows from [26, Exercise 8.8(c)], the second equality follows from the above claim. To get the last equality in (26), we know that

$$\begin{aligned} & N_\Omega(\bar{x}, \bar{y}) \\ &= \left\{(\bar{\xi}, \bar{u}) \mid \bar{\xi}^\top(x - \bar{x}) + \bar{u}^\top(y - \bar{y}) \leq 0, \forall (x, y) \in \Omega\right\} \\ &= \left\{(\bar{\xi}, \bar{u}) \mid \bar{\xi}^\top(x - \bar{x}) + \bar{u}^\top\boldsymbol{A}(x - \bar{x}) \leq 0, \forall x \in \Omega_1 \cap \Omega_2\right\} \\ &= \left\{(\bar{\xi}, \bar{u}) \mid \boldsymbol{A}^\top\bar{u} + \bar{\xi} \in N_{\Omega_1\cap\Omega_2}(\bar{x})\right\} \\ &= \left\{(\bar{\xi}, \bar{u}) \mid \boldsymbol{A}^\top\bar{u} + \bar{\xi} \in N_{\Omega_1}(\bar{x}) + N_{\Omega_2}(\bar{x})\right\}, \end{aligned} \quad (27)$$

where $\Omega_1 := \{x \in \mathbb{R}^n \mid \boldsymbol{E}x \leq f\}$ and $\Omega_2 := \{x \in \mathbb{R}^n \mid \boldsymbol{F}x = g\}$. Together with $N_{\Omega_1}(\bar{x}) = \boldsymbol{E}^\top N_{\mathbb{R}^m_-}(\boldsymbol{E}\bar{x} - f)$ and $N_{\Omega_2}(\bar{x}) = \boldsymbol{F}^\top\mathbb{R}^l$, the desired equality in (26) follows. Utilizing the fact

$$\bar{w} \in N_{\mathbb{R}^m_-}(\boldsymbol{E}\bar{x} - f) \Leftrightarrow 0 \in N_{\mathbb{R}^m_+}(\bar{z}) + \bar{w}, \quad \boldsymbol{E}\bar{x} - f + \bar{z} = 0,$$

the desired result can then be obtained immediately. □

It is known from [26, Example 10.2] that the inclusion condition $0 \in y + u + \lambda\partial\|y\|_0$ in the above KKT system is necessary for

$$y \in \mathrm{Prox}_{\lambda\|\cdot\|_0}(-u), \quad (28)$$

where the proximal mapping

$$\mathrm{Prox}_{\lambda\|\cdot\|_0}(u) := \arg\min_{y\in\mathbb{R}^n} \lambda\|y\|_0 + \frac{1}{2}\|y - u\|^2$$

is the so-called hard-thresholding operator in [12] and [32] and it has the following explicit expression:

$$\mathrm{Prox}_{\lambda\|\cdot\|_0}(u) = \left\{ y \in \mathbb{R}^n \middle| y_i = \left\{ \begin{array}{ll} u_i & \text{if } |u_i| > \sqrt{2\lambda}, \\ u_i \text{ or } 0 & \text{if } |u_i| = \sqrt{2\lambda}, \\ 0 & \text{otherwise}, \end{array} \right. \right\}. \quad (29)$$

Thus, by replacing this inclusion condition in (19) with (28), we can get a stronger condition system than (19) for optimality analysis. Any $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ satisfying such a new condition system is called a stationary point of problem (17). Note that the condition in the third line in (19) is equivalent to

$$w - \Pi_{\mathbb{R}^m_+}(w - z) = 0, \quad (30)$$

where $\Pi_{\mathbb{R}^m_+}$ is the projection operator onto $\mathbb{R}^m_+$. We can get an equivalent reformulation for the stationary point system

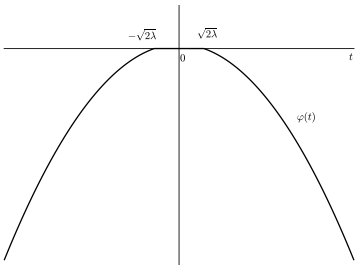

FIGURE 6. The graph of $\varphi(t)$.

as follows

$$F(x,y,z,u,v,w) := \begin{pmatrix} -\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w \\ \min\limits_{\vartheta \in \mathrm{Prox}_{\lambda g}(-u)} \|y - \vartheta\| \\ w - \Pi_{\mathbb{R}^m_+}(w - z) \\ y - \boldsymbol{A}x + b \\ \boldsymbol{F}x - g \\ \boldsymbol{E}x - f + z \end{pmatrix} = 0. \tag{31}$$

We call the involved mapping F the residual function for problem (17). The stationarity condition can imply the optimality of the primal problem (17) and its dual problem. This will be discussed in the coming subsection.

B. DUAL OPTIMALITY

The Lagrangian dual problem of (17) is defined by

$$\sup_{w,u,v} \inf_{x,y,z} L(x,y,z;u,v,w). \tag{32}$$

Note that

$$\begin{aligned} &\inf_{x,y,z} L(x,y,z;u,v,w) \\ &= u^\top b - v^\top g - w^\top f + \inf_y \left\{ \frac{1}{2}\|y\|_2^2 + \lambda\|y\|_0 + u^\top y \right\} \\ &\quad + \inf_x \left\{ \left(-\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w\right)^\top x \right\} \\ &\quad + \inf_z \left\{ \delta_{\mathbb{R}^m_+}(z) + w^\top z \right\} \\ &= \begin{cases} \varphi(u) + u^\top b - v^\top g - w^\top f, & \text{if } \boldsymbol{A}^\top u = \boldsymbol{F}^\top v + \boldsymbol{E}^\top w, \\ & \text{and } w \geq 0; \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

where

$$\varphi(u) = -\frac{1}{2}\|u\|^2 + \sum_{i=1}^n \min\left\{\lambda, \frac{1}{2}u_i^2\right\} \tag{33}$$

is a concave function in u, following from [25, Th. 5.5]. The graph of the univariate version of φ is plotted as in Figure 6.

Thus, the dual problem has the following explicit form

$$\begin{aligned} \max \quad & \varphi(u) + u^\top b - v^\top g - w^\top f \\ \text{s.t.} \quad & -\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w = 0 \\ & w \geq 0. \end{aligned} \tag{34}$$

Problem (34) with $\boldsymbol{A}$, $\boldsymbol{E}$, $\boldsymbol{F}$ defined as above is a convex program and admits the following necessary and sufficient optimality condition.

Theorem 2: Let $\boldsymbol{A}$, $\boldsymbol{E}$ and $\boldsymbol{F}$ be defined as in (12)-(16). Then $(\bar{u}, \bar{v}, \bar{w})$ is an optimal solution to (34) if and only if there exists $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{cases} -\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w = 0, \\ y \in \mathrm{conv}\left(\mathrm{Prox}_{\lambda\|\cdot\|_0}(-u)\right), \\ y - \boldsymbol{A}x + b = 0, \\ \boldsymbol{F}x - g = 0, \\ \boldsymbol{E}x - f + z = 0, z \geq 0, w \geq 0, w^\top z = 0, \end{cases} \tag{35}$$

Proof. Problem (34) can be rewritten as

$$\begin{aligned} \min \quad & -\varphi(u) - u^\top b + v^\top g + w^\top f + \delta_{\mathbb{R}^m_+}(w) \\ \text{s.t.} \quad & -\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w = 0. \end{aligned} \tag{36}$$

The corresponding Lagrangian function with respect to (36) is defined by

$$\begin{aligned} L^D(u,v,w;x) := & -\varphi(u) - u^\top b + v^\top g + w^\top f + \delta_{\mathbb{R}^m_+}(w) \\ & - x^\top\left(-\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w\right), \end{aligned}$$

and hence the KKT conditions for (36) turn out to be

$$\begin{cases} 0 = \partial_u L^D(u,v,w;x) = \partial\left(-\varphi(u)\right) + (\boldsymbol{A}x - b), \\ 0 = \nabla_v L^D(u,v,w;x) = g - \boldsymbol{F}x, \\ 0 \in \partial_w L^D(u,v,w;x) = \partial\delta_{\mathbb{R}^m_+}(w) + f - \boldsymbol{E}x, \\ -\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w = 0, \end{cases} \tag{37}$$

From the nonnegativity of the function $\|\cdot\|_0$, it is known from [26, p. 21] that $\|\cdot\|_0$ is prox-bounded with the $\lambda_g = +\infty$. By applying the subsmoothness of Moreau envelopes addressed in [26, Example 10.32], we have

$$\begin{aligned} \partial\left(-\varphi(u)\right) &= u - \left(\mathrm{conv}\left(\mathrm{Prox}_{\lambda\|\cdot\|_0}(-u)\right) - (-u)\right) \\ &= -\mathrm{conv}\left(\mathrm{Prox}_{\lambda\|\cdot\|_0}(-u)\right). \end{aligned}$$

By employing the fact $\partial\delta_{\mathbb{R}^m_+}(w) = N_{\mathbb{R}^m_+}(w)$, the condition in the third line in system (37) is equivalent to the following complementarity system

$$w \in \mathbb{R}^m_+, f - \boldsymbol{E}x \in \mathbb{R}^m_+, \quad w^\top(f - \boldsymbol{E}x) = 0.$$

Thus, we conclude that the KKT system (37) can be equivalently reformulated as (35) by setting $y = \boldsymbol{A}x - b$. By virtue of the structure in $\boldsymbol{F}$ as defined in (16), we know that $\boldsymbol{F}$ is full row rank. Thus, the generalized Slater condition, i.e., $\exists w > 0$ such that $-\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w = 0$, holds automatically. Due to the convex programming theory, we can conclude the desired equivalence. □

A sufficient optimality condition via stationarity is then obtained to ensure the solvability of both Problem (17) and its dual (34) with zero duality gap, i.e., the strong duality holds.

Corollary 1: If $(x^*, y^*, z^*, u^*, v^*, w^*)$ is a stationary point of (17), then (x^*, y^*, z^*) is an optimal solution of (17) and

(u^*, v^*, w^*) is an optimal solution to (34) and the duality gap vanishes.

Proof. Invoking [25, Example 10.2], we know that $(x^*, y^*, z^*, u^*, v^*, w^*)$ satisfies the KKT conditions of problem (17) as presented in (19). It also satisfies the conditions in (35) by the property of convex hull in [25, Th. 2.3]. Thus, (u^*, v^*, w^*) is an optimal solution to the dual problem (34) and (x^*, y^*, z^*) is a feasible solution to Problem (17). Denote the optimal values of problems (17) and (34) by p^* and d^* respectively. The feasibility implies that

$$p^* \le \frac{1}{2}\|y^*\|_2^2 + \lambda\|y^*\|_0 = L(x^*, y^*, z^*; u^*, v^*, w^*). \quad (38)$$

On the other hand, by the observations

$$\begin{cases} \phi_1(y^*, u^*) = \inf\limits_y \phi_1(y, u^*), \\ 0 = \phi_2(x^*, u^*, v^*, w^*) = \inf\limits_x \phi_2(x, u^*, v^*, w^*), \\ \delta_{\mathbb{R}^m_+}(z^*) + (w^*)^\top z^* = \inf\limits_z \left\{\delta_{\mathbb{R}^m_+}(z) + (w^*)^\top z\right\}, \end{cases}$$

where $\phi_1(y, u) := \frac{1}{2}\|y\|_2^2 + \lambda\|y\|_0 + u^\top y$, and $\phi_2(x, u, v, w) := x^\top\left(-\boldsymbol{A}^\top u + \boldsymbol{F}^\top v + \boldsymbol{E}^\top w\right)$, we can immediately verify that

$$\begin{aligned} L(x^*, y^*, z^*; u^*, v^*, w^*) &= \inf_{x,y,z} L(x, y, z; u^*, v^*, w^*) \\ &= \varphi(u^*) + b^\top u^* - g^\top v^* - f^\top w^* \\ &= d^*. \end{aligned} \quad (39)$$

Combining with the weak duality, i.e., $p^* \ge d^*$, and (38), we have the duality gap $p^* - d^* = 0$, and hence (x^*, y^*, z^*) is an optimal solution of Problem (17). The proof is complete. □

To summarize, the first-order optimality conditions for the primal and dual problems are established. Particularly, the stationarity is shown to be a strong duality condition. All of these enrich the contents for non-convex optimization theory and hence of theoretical significance.

IV. A HARD-THRESHOLDING BASED ADMM APPROACH

The augmented Lagrangian function associated with the sparse model (17) can be written as

$$\begin{aligned} &L_\sigma(x, y, z; u, v, w) \\ &= \lambda\|y\|_0 + \frac{1}{2}\|y\|_2^2 + \delta_{\mathbb{R}^m_+}(z) + u^\top(y - \boldsymbol{A}x + b) \\ &\quad + v^\top(\boldsymbol{F}x - g) + w^\top(\boldsymbol{E}x - f + z) \\ &\quad + \frac{\sigma}{2}\left(\|y - \boldsymbol{A}x + b\|_2^2 + \|\boldsymbol{F}x - g\|_2^2 + \|\boldsymbol{E}x - f + z\|_2^2\right) \end{aligned}$$

where $\sigma > 0$ is the penalty parameter, $u \in \mathbb{R}^n$,$v \in \mathbb{R}^l$ and $w \in \mathbb{R}^m$ are Lagrangian multipliers. The iterative framework of the hard-thresholding based alternating direction method of multipliers (HTADMM) can be described as

$$\begin{cases} (y^{k+1}, z^{k+1}) \in \arg\min\limits_{y\in\mathbb{R}^n, z\in\mathbb{R}^m} \left\{L_\sigma(x^k, y, z; u^k, v^k, w^k)\right\}; \\ x^{k+1} = \arg\min\limits_{x\in\mathbb{R}^n} \left\{L_\sigma(x, y^{k+1}, z^{k+1}; u^v, v^k, w^k)\right\}; \\ u^{k+1} = u^k - \tau\sigma(y^{k+1} - \boldsymbol{A}x^{k+1} + b); \\ v^{k+1} = v^k - \tau\sigma(\boldsymbol{F}x^{k+1} - g); \\ w^{k+1} = w^k - \tau\sigma(\boldsymbol{E}x^{k+1} - f + z^{k+1}), \end{cases} \quad (40)$$

with $\tau > 0$ as the dual stepsize. For the first subproblem in (40), by employing the hard-thresholding operator (29), together with the sparsity consideration, we can get a closed-form expression for y^{k+1} as

$$y_i^{k+1} = \begin{cases} 0, & \text{if } |u^k + \sigma(\boldsymbol{A}x^k - b)|_i \\ & \le \sqrt{2\lambda(1+\sigma)}; \\ \frac{1}{1+\sigma}(u^k + \sigma(\boldsymbol{A}x^k - b))_i, & \text{otherwise.} \end{cases} \quad (41)$$

and the first-order optimality condition for z yields the update

$$z^{k+1} = \Pi_{\mathbb{R}^m_+}\left(-\boldsymbol{E}x^k + f - \frac{w^k}{\sigma}\right). \quad (42)$$

For the second subproblem in (40), we can also employ the first-order optimality condition to get the x-update by solving the following linear system

$$\boldsymbol{H}x = Rhs, \quad (43)$$

where $\boldsymbol{H} \mathcal{D} \boldsymbol{A}^\top\boldsymbol{A} \mathcal{C} \boldsymbol{F}^\top\boldsymbol{F} \mathcal{C} \boldsymbol{E}^\top\boldsymbol{E}$ and $Rhs = \boldsymbol{A}^\top\left(\frac{u^k}{\sigma} - b + y^{k+1}\right) + \boldsymbol{F}^\top\left(g - \frac{v^k}{\sigma}\right) + \boldsymbol{E}^\top\left(f - z^{k+1} - \frac{w^k}{\sigma}\right)$. Note that matrix $\boldsymbol{E}$ is full column rank. Thus $\boldsymbol{H}$ is positive definite and hence nonsingular, which leads to a unique solution to System (43). For large scale problems, this linear system can be approximately solved by adopting preconditioned conjugate gradient (PCG) method. The detailed framework of our HTADMM is presented in Algorithm 1.

Algorithm 1 HTADMM for Solving Problem (17)

Require: Choose an initial point $(x_0^0, y_0^0, z_0^0, u_0^0, v_0^0, w_0^0)$, accuracy parameters $\epsilon > 0$, the parameters $\lambda, \sigma > 0$.

Ensure: (x^*, y^*, z^*);

Step 1 Set $(x^0, y^0, z^0) = (x_0^0, y_0^0, z_0^0)$, $(u^0, v^0, w^0) = (0, 0, 0)$ and $k = 0$;

Step 2 Compute $\left(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1}, v^{k+1}, w^{k+1}\right)$ by (43), (41), (42) and (40);

Step 3 Set $k = k + 1$. If some criterion is satisfied to the accuracy ε, then set $(x^*, y^*, z^*) = \left(x^k, y^k, z^k\right)$ and stop. Otherwise, go to Step 2.

V. A CASE STUDY

In this section we conduct numerical experiments on Beijing Metro Yizhuang Line spanning full service period of one day to evaluate our proposed model. The numerical study is executed by running Matlab (version 2016b) on a windows laptop (Intel(R) Core(TM) i5-5250U CPU@ 1.60GHZ 1.60GHZ RAM 4.0G)

TABLE 3. The windows of the dwell time constraints.

Platforms	SJZ1	XC1	XHM1	JG1	YZQ1	YZWHY1	WYJ1	RJDJ1	RCDJ1	TJNL1	JHL1	CQN1	CQ1	YZHCZ1
Lower bounds	25	25	25	25	25	25	25	25	25	25	25	25	25	25
Upper bounds	45	45	45	45	45	45	45	45	45	45	45	45	45	45
Platforms	SJZ2	XC2	XHM2	JG2	YZQ2	YZWHY2	WYJ2	RJDJ2	RCDJ2	TJNL2	JHL2	CQN2	CQ2	YZHCZ2
Lower bounds	25	25	25	25	25	25	25	25	25	25	25	25	25	25
Upper bounds	45	45	45	45	45	45	45	45	45	45	45	45	45	45

TABLE 4. The train utility rate and distribution of D for one-day(16 hours) services.

	Hours															
Periods	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
u	0.3	0.8	0.8	0.6	0.4	0.4	0.5	0.5	0.3	0.6	0.6	0.8	0.8	0.4	0.3	0.3
u^D	0.03	0.11	0.13	0.07	0.05	0.05	0.04	0.05	0.04	0.05	0.07	0.13	0.12	0.03	0.02	0.02

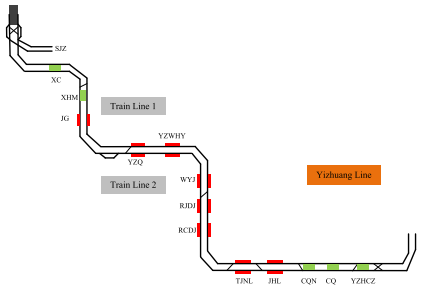

FIGURE 7. Beijing metro yizhuang line.

A. DATA PREPARATION

We provide some revelent information with regard to Beijing Metro Yizhuang Line. Beijing Metro Yizhuang Line, opened on December in 2010, is a railway extended to southeast direction with 14 stations as shown in Figure 7. The total length of Yizhuang Line is 23.23 km and the average distance of two stations is 1.8 km, with the minimum distance being 1 km (between YZQ and YZWHY) and maximum distance being 2.8 km (between SJZ and XC) respectively.

The number of departure trains, says $\bar{m}$, is closely related to the average headway time in the way that the number of trains increases when the headway time decreases. Six different choices of $\bar{m}$ are given according to the passenger flow volume in non-peak hours and peak hours, which are $\bar{m} = 248, 259, 271, 282, 293, 305$ with the passenger flows varying from 215000 to 265000. Set $\bar{n} = 28$, which is the number of stations (two lines) in Beijing Yizhuang Line. The inputs in the equality constraint $y - \boldsymbol{A}x + b = 0$ are $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with $n = 2\bar{n}\bar{m}$, where each row contains one -1, one 1 and the rest is 0, determined by (12), (13) and (14). The component of $b \in \mathbb{R}^n$ is $-(\Delta_i^t + \nabla_j^{\tilde{t}})_{(i,j,t,\tilde{t})}$, where train t is accelerating and $\tilde{t}$ is braking, and the opposite platforms i and j satisfy the equation $i + j = 29$. In the equality constraint $\boldsymbol{F}x - g = 0$, $\boldsymbol{F} \in \mathbb{R}^{l\times m}$ is a coefficient matrix with $l = (\bar{n} - 1)\bar{m}$ and $g \in \mathbb{R}^l$ is composed of trip time. In the equality constraint $z - f + \boldsymbol{E}x = 0$, $\boldsymbol{E} \in \mathbb{R}^{m\times n}$ is the coefficient matrix of inequality constraints with $m = 2(2\bar{n}(\bar{m} - 1) + \bar{n}\bar{m} + \bar{m} + n)$. The elements in the vector of $f \in \mathbb{R}^m$ are upper and lower bounds of dwell time constraints, headway time constraints, total travel time constraints, and the domain of event time constraints.

Following from [20], the windows of dwell time constraints including lower bounds and upper bounds are presented in Table 3. Let D be the passenger demand for one-day service, $u^D \in \mathbb{R}^{16}$ be the distribution of D for each hour and $u \in \mathbb{R}^{16}$ be the train utility rate for different hours which are both given in Table 4. By using $h = \frac{3600\times C_t\times u}{D}$, we can get the headway time for each hour. The bounds of trip times are calculated by Table 5 as reference. For each train t, the bounds of trip time are more than $\sum_{q=1}^{5} \frac{s_q}{v_q}$ and the total travel time constraints are related to the dwell time and trip time, we can calculate the windows of total travel time constraints by using

$$\underline{tt} = r_1 \times (\sum_{i=1}^{28} \underline{dw}_i + \sum_{\substack{i=1 \\ i\neq 14}}^{27} \underline{tr}_{ij} + \sum_{i=14} \underline{\kappa}_{ij})$$

$$\overline{tt} = r_2 \times (\sum_{i=1}^{28} \overline{dw}_i + \sum_{\substack{i=1 \\ i\neq 14}}^{27} \overline{tr}_{ij} + \sum_{i=14} \overline{\kappa}_{ij})$$

where $r_1 = 1.1$ and $r_2 = 0.85$.

To calculate the energy consumption and regenerative energy, we set the maximum acceleration $a_{acc} = 0.5m/s^2$ at accelerating phrase, and $a_{bra} = -0.8m/s^2$ at braking phrase. The resistance acceleration during the coasting phrase $r = -0.05m/s^2$, the conversion factor from electricity to kinetic energy $\theta_{acc} = 0.7$ and the conversion factor from kinetic energy to regenerative energy $\theta_{bra} = 0.5$. As presented in Section II, we set parameters $t_1 = 20\%$, $t_2 = 25\%$, $t_3 = 20\%$ and $t_4 = 35\%$ as the percentages of the running time to approximately calculate the energy consumption for accelerating E^{con} and the regenerative energy produced during braking E^{bra}. The parameters c and c' in Figure 3 which represent the width of the rectangles for E^{con} and E^{bra} are chosen to be 14% and 21% respectively.

TABLE 5. Speed limits of every section on beijing metro yizhuang line.

Sections	Segment v_1(m/s)	Segment s_1(m)	Segment v_2(m/s)	Segment s_2(m)	Segment v_3(m/s)	Segment s_3(m)	Segment v_4(m/s)	Segment s_4(m)	Segment v_5(m/s)	Segment s_5(m)
SJZ-XC	13.89	0-150	23.61	150-480	18.06	480-1161	23.61	1161-2501	16.67	2501-2631
XC-XHM	16.67	2631-2643	23.61	2643-2797	20.83	2797-3534	23.61	3534-3780	16.67	3780-3905
XHM-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-YZWHY	16.67	8254-8265	23.61	8265-9116	16.67	9116-9246	-	-	-	-
YZWHY-WYJ	16.67	9246-9256	23.61	9259-10655	16.67	10655-10785	-	-	-	-
WYJ-RJDJ	16.67	10785-10797	23.61	10797-11933	16.67	11933-12065	-	-	-	-
RJDJ-RCDJ	16.67	12065-12077	23.61	12077-13289	16.67	13289-13419	-	-	-	-
RCDJ-TJNL	16.67	13419-13431	23.61	13419-14649	20.83	14649-15426	23.61	15426-15624	16.67	15624-15756
TJNL-JHL	16.67	15756-15768	23.61	15768-17891	16.67	17891-18021	-	-	-	-
JHL-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-YZHCZ	16.67	21394-21406	23.61	21406-22569	16.67	22569-22728	-	-	-	-

TABLE 6. Five different approaches for comparison.

Approach No.	Optimization Model	Algorithm/Solver
$A1$	$\min_{x,y,s}\left\{e^\top s \middle\vert -s \le y \le s, (x,y)\in\mathcal{F}_1\right\}$	Gurobi [7]
$A2$	$\min_{x,y}\left\{\frac{1}{2}\|y\|_2^2 \middle\vert (x,y)\in\mathcal{F}_1\right\}$	CVX
$A3$	$\min_{x,y}\left\{\frac{1}{2}\|y\|_2^2+\lambda\|y\|_1+\delta_{\mathbb{R}^m_+}(z) \middle\vert (x,y,z)\in\mathcal{F}\right\}$	TSADMM [20]
$A4$	$\min_{x,y}\left\{\frac{1}{2}\|y\|_2^2+\lambda\|y\|_0+\delta_{\mathbb{R}^m_+}(z) \middle\vert (x,y,z)\in\mathcal{F}\right\}$	HTADMM
$A5$	$\min_{x,y}\left\{\frac{1}{2}\|y\|_2^2+\lambda\|y\|_{1/2}^{1/2}+\delta_{\mathbb{R}^m_+}(z) \middle\vert (x,y,z)\in\mathcal{F}\right\}$	HaTADMM

B. INITIALIZATION AND STOPPING CRITERIA

Invoking the optimality analysis in Section III, we will use the relative primal infeasibility and relative dual infeasibility, defined as below, to measure the quality of the approximate solution:

$$\eta_p = \left\{\frac{\|y_k - \boldsymbol{A}x_k + b\|}{1+\|y_k\|}, \frac{\|\boldsymbol{F}x_k - g\|}{1+\|x_k\|}, \frac{\|\boldsymbol{E}x_k - f + z_k\|}{1+\|\boldsymbol{E}x_k\|+\|z_k\|}\right\},$$

$$\eta_d = \frac{\|\boldsymbol{A}^T u_k - \boldsymbol{F}^T w_k - \boldsymbol{E}^T s_k\|}{1+\|\boldsymbol{A}^T u_k\|+\|\boldsymbol{F}^T w_k\|+\|\boldsymbol{E}^T s_k\|}.$$

The algorithm will be terminated once η_p and η_d is less than some prescribed accuracy parameter ϵ, or when the iteration number k reaches the maximal iteration number Maxiter. Here we choose $\epsilon = 10^{-3}$ and Maxiter $= 10^4$. Set the dual stepsize $\tau = 1.618$ in HTADMM.

In attempt to get a better performance of Algorithm 1, we choose the initial point (x_0^0, y_0^0, z_0^0) by solving the following convex problem

$$\min_{x,y,z}\left\{\lambda\|y\|_1 + \frac{1}{2}\|y\|_2^2 + \delta_{\mathbb{R}^m_+}(z) \mid (x,y,z)\in\mathcal{F}\right\}$$

with

$$\mathcal{F}: =\{(x,y,z) \mid y-\boldsymbol{A}x+b=0,\ \boldsymbol{F}x-g=0,\ \boldsymbol{E}x-f+z=0\}, \tag{44}$$

using the ADMM framework with the accuracy $\epsilon_0 = 10^{-1}$. The computation time for the initialization is also counted in the sequent analysis on time comparison.

C. NUMERICAL RESULTS

To illustrate the effectiveness of our proposed approach for energy-efficient timetabling on the case study for Beijing Metro Yizhuang Line, the comparison among the following five approaches, as listed in Table 6, will be conducted in terms of the energy saving rate and the computation time. In this table, the set $\mathcal{F}$ is defined as in (44), and the set $\mathcal{F}_1$ is defined by

$$\mathcal{F}_1 := \{(x,y) \mid y - \boldsymbol{A}x + b = 0,\ \boldsymbol{F}x - g = 0,\ \boldsymbol{E}x - f \le 0\}.$$

Specifically, the first approach $A1$ in Table 6 is the approach proposed in [8], the second and the third approaches are from [20], the fourth approach $A4$ is the proposed one in this paper, and the last approach $A5$ is the square root of $\ell_{1/2}$-norm relaxation approach where the involved $\|y\|_{1/2}^{1/2} := \sum_{i=1}^{n}|y_i|^{1/2}$. We design an efficient ADMM algorithm based on the so-called half-thesholding (HaTADMM), in which the updating formula for y in (41) is replaced by

$$y^{k+1} = \left(h_{\frac{\lambda_1}{\lambda_2+\sigma},\frac{1}{2}}\left(\frac{\sigma(\boldsymbol{A}x^k - b) - u^k}{1+\sigma}\right)\right)_{i=1}^{n}. \tag{45}$$

Here, $h_{\tau,\frac{1}{2}}$ is the half-thresholding operator [34] which takes the form of

$$h_{\tau,\frac{1}{2}}(t_i) := \begin{cases} f_{\tau,\frac{1}{2}}(t_i), & \text{if } |t_i| > \frac{\sqrt[3]{54}}{4}\tau^{2/3}; \\ 0, & \text{otherwise} \end{cases}$$

TABLE 7. Energy saving rates with different $\bar{m}$.

	$r(\%)$				
$\bar{m}$	$A1$	$A2$	$A3$	$A4$	$A5$
248	18.18	17.70	27.21	**28.07**	**28.08**
259	19.60	19.05	25.89	**27.93**	**27.99**
271	21.65	19.82	27.60	**31.30**	**30.80**
282	19.44	18.50	25.67	**30.76**	**31.01**
293	19.98	18.66	24.33	**28.55**	**28.67**
305	18.57	18.88	23.83	**25.27**	**25.25**

TABLE 8. Energy saving rates for $\bar{m} = 293$.

	$r(\%)$				
Ref. No.	$A1$	$A2$	$A3$	$A4$	$A5$
[17]	23.44	18.90	24.98	**27.56**	**27.56**
[36]	22.90	18.54	24.80	**27.40**	**27.42**
[37]	24.67	19.82	25.58	**28.98**	**29.03**
[38]	27.06	21.53	27.45	**31.25**	**31.30**
[33]	21.15	18.20	25.58	**28.63**	**28.64**

TABLE 9. Timetable generated by *A4* for the beijing metro yizhuang line with $\bar{m} = 293$.

Station	SJZ	XC	XHM	JG	YZQ	YZWHY	WYJ
Arrival Time(s)	0	192	337	537	712	844	977
Dwell Time(s)	28	30	28	25	23	24	24
Station	RJDJ	RCDJ	TJNL	JHL	CQN	CQ	YZHCZ
Arrival Time(s)	1120	1243	1406	1589	1757	1901	2044
Dwell Time(s)	24	23	24	24	24	24	-

with $f_{\tau,\frac{1}{2}}(t_i) = \frac{2}{3}t_i\left(1+\cos\left(\frac{2\pi}{3}-\frac{2}{3}\zeta_\tau(t_i)\right)\right)$, and $\zeta_\tau(t_i) = \arccos\left(\frac{\tau}{8}\left(\frac{|t_i|}{3}\right)^{-3/2}\right)$.

Energy saving rates calculated by aforementioned models are listed in Table 7 with different choices of $\bar{m}$. As shown in Table 7, for all the instances, approaches $A4$ and $A5$ have a quite competitive performance in terms of the energy saving rate, both of which outperform the other three approaches. More specifically, for our proposed $A4$, the saving rates are at least 6.39 percent higher than those generated by $A1$ and $A2$, both of which lack the consideration of the sparsity in y, and with around 2.89 percentage in average higher than those of $A3$ where the ℓ_1-norm is employed as a surrogate of the sparsity characterization ℓ_0-norm. Similar results can be obtained by using different initial trip times from references [17], [33], [36]–[38] with $\bar{m} = 293$, as shown in Table 8.

According to the optimal timetable generated by $A4$, the total travel time is $T = 2044s$ and the arrival time and dwell time at each station are shown in Table 9. We also illustrate the timetable of $A1$ in Table 10. These figures indicate that the trip time and dwell time in two timetables on each station is different and therefore consumption energy consumed by two trains is different.

One-way train diagram on Beijing Metro Yizhaung Line for $\bar{m} = 293$ is illustrated in figure 8. It is shown that, all the trains depart from the original station at different timestamps one by one and then arrive at final station. The departure

TABLE 10. Timetable generated by *A1* for the beijing metro yizhuang line with $\bar{m} = 293$.

Station	SJZ	XC	XHM	JG	YZQ	YZWHY	WYJ
Arrival Time(s)	0	185	320	515	690	825	961
Dwell Time(s)	25	25	25	25	25	25	25
Station	RJDJ	RCDJ	TJNL	JHL	CQN	CQ	YZHCZ
Arrival Time(s)	1106	1231	1396	1581	1751	1896	2041
Dwell Time(s)	25	25	25	25	25	25	-

TABLE 11. Time comparisons with different $\bar{m}$.

	Time(s)				
$\bar{m}$	$A1$	$A2$	$A3$	$A4$	$A5$
248	11.37	4.86	8.06	7.77	8.93
259	7.84	6.42	25.17	16.34	21.68
271	7.49	8.05	93.47	19.05	22.42
282	8.76	8.06	72.64	11.51	12.92
293	8.03	7.75	45.21	10.89	14.04
305	14.38	16.15	22.59	7.49	9.68

TABLE 12. Time comparison for $\bar{m} = 293$.

	Time(s)				
Ref. No.	$A1$	$A2$	$A3$	$A4$	$A5$
[17]	7.82	7.56	24.27	15.67	18.03
[36]	7.87	9.90	27.42	15.82	18.44
[37]	6.30	6.87	18.29	16.28	18.98
[38]	6.74	8.47	16.85	16.42	18.85
[33]	8.58	7.00	33.87	16.39	19.04

frequencies are essentially consistent with time, it can be seen that the departure frequency is higher and the departure headway is shorter during the morning and evening rush hours.

Using timetable and train diagram, we can calculate the average total energy and net energy consumed by all trains passing through each platform for $A4$ and $A1$ with $\bar{m} = 293$ during the day as shown in Figures 9 and 10. We can see that, the energy consumption after regulation decreased significantly in Figure 9, which partially reflects the reason why approach $A4$ outperforms $A1$ in terms of the energy saving rate.

Besides the energy saving rate, we also evaluate five approaches in terms of the computation time. The time comparisons for all the testing instances are listed in Table 11 with different choices of $\bar{m}$. Compared with $A3$ and $A5$, the computation time for $A4$ is the least. Particularly, for the case of $\bar{m}$ equal to 305, the computation time of $A4$ is the least among all involved approaches.

The time comparisons for testing instances with different choices of initial trip times from reference [17], [33], [36]–[38] are listed in Table 12. Compared with $A3$ and $A5$, the computation time for $A4$ is the least.

Although the computation time of $A4$ is longer than that of $A1$ and $A2$, it superiorly outperforms $A1$ and $A2$ in terms of energy saving rate as one can see in Tables 7 and 8. Meanwhile, as shown in Figures 11 and 12, the number of iterations of $A4$ is less than that of $A3$ and $A5$, which partially explains the reason why the computation time of $A4$ is the least among these three methods.

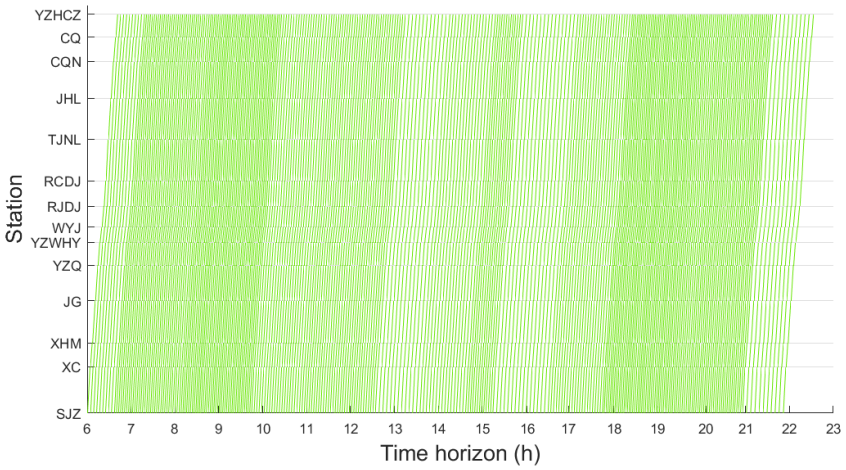

FIGURE 8. One-way train diagram on beijing metro yizhuang line for $A4$ with $\bar{m} = 293$.

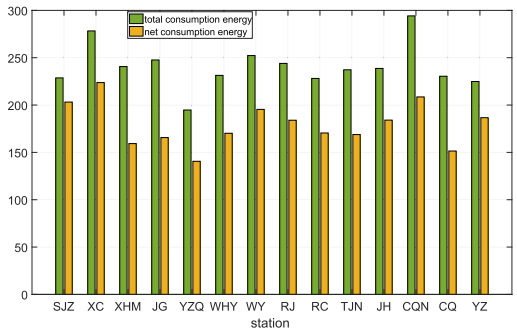

FIGURE 9. Comparison of total energy consumption and net energy consumption for $A4$ with $\bar{m} = 293$.

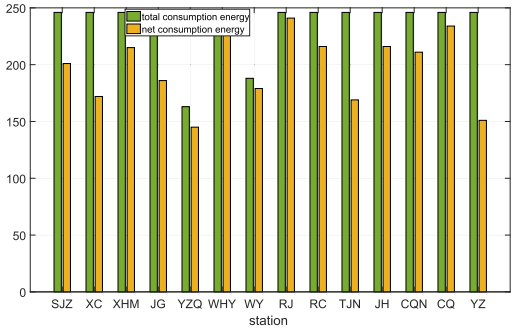

FIGURE 10. Comparison of total energy consumption and net energy consumption for $A1$ with $\bar{m} = 293$.

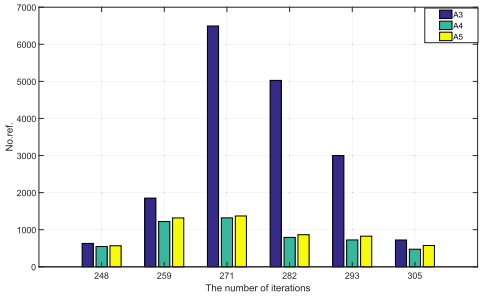

FIGURE 11. Number of iterations with different choices of $\bar{m}$.

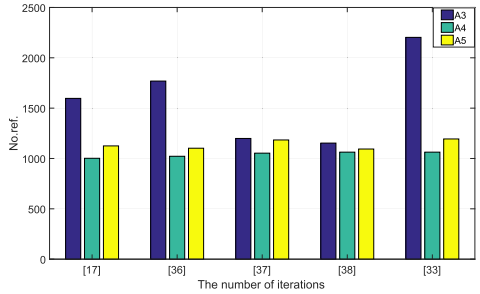

FIGURE 12. Number of iterations for different trip time.

It is known that the theoretical convergence of non-convex ADMM approaches is not always guaranteed, unless some strong conditions are imposed to the objective functions and feasible regions of the underlying models (see, e.g., [41] and references therein). The existing conditions fail to hold for our proposed non-convex model. Nevertheless, the numerical convergence behaviors in terms of the infeasibility measure for the above 11 instances are presented in the following Figures. We 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 1500 iterations.

VI. CONCLUSION

An ℓ_0-norm minimization model has been built and a hard-thresholding based alternating direction method of multipliers (HTADMM) has been proposed for the energy-efficient timetabling in subway systems. From the theoretical perspective, a first-order necessary optimality condition in terms of the KKT condition, and a first-order sufficient optimality condition in terms of the stationarity, have been proposed for the proposed ℓ_0-norm minimization problem with no additional constraint qualifications. This, to some extent, provides an enrichment or a supplement to the optimization theory. From the practical perspective, our proposed approach has shown to be efficient in terms of the

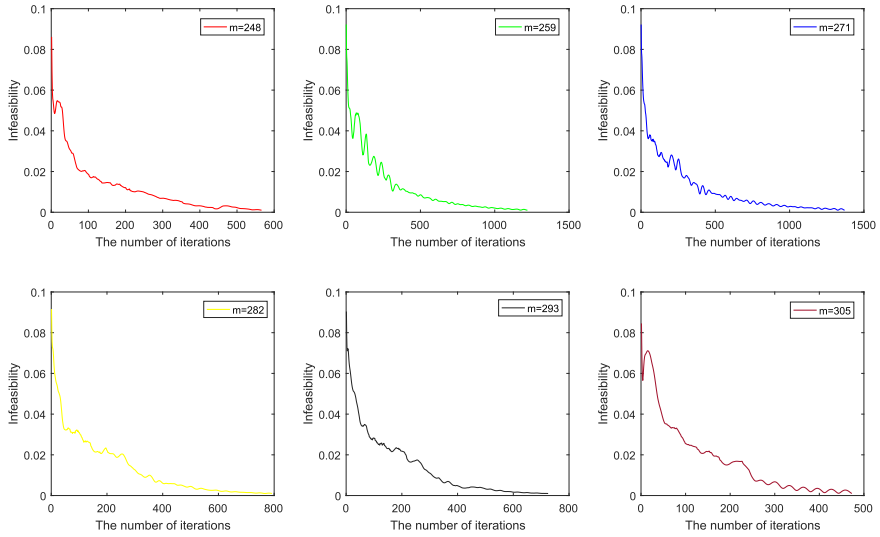

FIGURE 13. The numerical convergence behavior of A4 with different choices of $\bar{m}$.

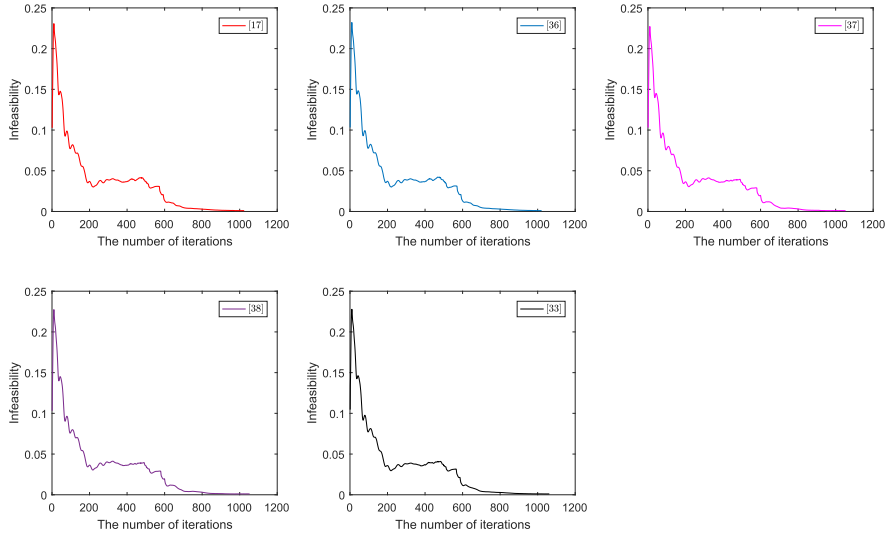

FIGURE 14. The numerical convergence behavior of A4 for different trip time.

energy saving rate and the computation time in the case study of Beijing Metro Yizhuang Line. It is worth mentioning that the HTADMM is a first-order since we only use the first-order (sub)-derivative information of the involved functions, future research would be in considering some second-order methods based on the optimality analysis to further accelerate the iterative process.

ACKNOWLEDGMENT

The authors would like to thank Prof. Naihua Xiu and Dr. Lili Pan for their helpful suggestions on optimality analysis, and the anonymous reviewers for their helpful comments and suggestions to improve the quality of the paper.

REFERENCES

[1] A. Albrecht, P. Howlett, P. Pudney, V. Xuan, and P. Zhou, "The key principles of optimal train control—Part 1: Formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points," *Transp. Res. B, Methodol.*, vol. 94, no. 2, pp. 482–508, 2016.

[2] D. Bertsimas, A. King, and R. Mazumder, "Best subset selection via a modern optimization lens," *Ann. Statist.*, vol. 44, no. 2, pp. 813–852, Apr. 2016.

[3] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, nos. 5–6, pp. 629–654, Dec. 2008.

[4] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*. New York, NY, USA: Springer, 2000.

[5] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[7] S. D. Gupta, L. Pavel, and J. K. Tobin, "An optimization model to utilize regenerative braking energy in a railway network," in *Proc. Amer. Control Conf.*, Jul. 2015, pp. 5919–5924.

[8] S. Das, J. Tobin, and L. Pavel, "A two-step linear programming model for energy-efficient timetables in metro railway networks," *Transp. Res. B, Methodol.*, vol. 93, pp. 57–74, Nov. 2016.

[9] X. Dong *et al.*, "Estimating life-cycle energy payback ratio of overhead transmission line toward low carbon development," *J. Mod. Power Syst. Clean Energy*, vol. 3, no. 1, pp. 123–130, 2015.

[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[11] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.
[12] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, 2011.
[13] Y. Huang, L. Yang, T. Tang, F. Cao, and Z. Gao, "Saving energy and improving service quality: Bicriteria train scheduling in urban rail transit systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3364–3379, Dec. 2016.
[14] Y. Huang, L. Yang, T. Tao, Z. Gao, and C. Fang, "Joint train scheduling optimization with service quality and energy efficiency in urban rail transit networks," *Energy*, vol. 138, pp. 1124–1147, Nov. 2017.
[15] D. Li, X. Sun, and J. Wang, "Optimal lot solution to cardinality constrained mean–variance formulation for portfolio selection," *Math. Finance*, vol. 16, no. 1, pp. 83–101, 2006.
[16] X. Li and H. K. Lo, "An energy-efficient scheduling and speed control approach for metro rail operations," *Transp. Res. B, Methodol.*, vol. 64, pp. 73–89, Jun. 2014.
[17] X. Li and H. K. Lo, "Energy minimization in dynamic train scheduling and control for metro rail operations," *Transp. Res. B, Methodol.*, vol. 70, pp. 269–284, Dec. 2014.
[18] J. Liu, J. Chen, and Y. Jieping, "Large-scale sparse logistic regression," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 547–556.
[19] Z. Lu, "Iterative hard thresholding methods for ℓ_0 regularized convex cone programming," *Math. Program.*, vol. 147, no. 1, pp. 125–154, Oct. 2014.
[20] Z. Luo, X. Li, and N. Xiu, "A sparse optimization approach for energy-efficient timetabling in metro railway systems," *J. Adv. Transp.*, vol. 2018, Nov. 2018, Art. no. 1784789.
[21] L. F. Ochoa and G. P. Harrison, "Minimizing energy losses: Optimal accommodation and smart operation of renewable distributed generation," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 198–205, Feb. 2011.
[22] A. Ogunsola, L. Sandrolini, and A. Mariscotti, "Evaluation of stray current from a DC-electrified railway with integrated electric–electromechanical modeling and traffic simulation," *IEEE Trans. Ind. Appl.*, vol. 51, no. 6, pp. 5431–5441, Nov. 2015.
[23] M. Peña-Alcaraz, A. Fernández, A. P. Cucala, A. Ramos, and R. R. Pecharromán, "Optimal underground timetable design based on power flow for maximizing the use of regenerative-braking energy," *Proc. Inst. Mech. Eng., F, J. Rail Rapid Transit*, vol. 226, no. 4, pp. 397–408, 2011.
[24] A. Ramos, M. T. Peña, A. Fernández, and P. Cucala, "Mathematical programming approach to underground timetabling for maximizing the use of regenerative braking power," in *Proc. Int. Workshop Oper. Res.*, Madrid, Spain, Jun. 2008, pp. 88–95.
[25] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.
[26] R. T. Rockafellar and J.-B. R. Wets, *Variational Analysis*. New York, NY, USA: Springer-Verlag, 1997.
[27] G. M. Scheepmaker, R. M. Goverde, and L. G. Kroon, "Review of energy-efficient train control and timetabling," *Eur. J. Oper. Res.*, vol. 257, no. 2, pp. 355–376, 2017.
[28] C. Sun, R. Dai, and M. Mesbahi, "Weighted network design with cardinality constraints via alternating direction method of multipliers," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 4, pp. 173–189, Dec. 2018.
[29] P. Wang and R. M. P. Goverde, "Multi-train trajectory optimization for energy-efficient timetabling," *Eur. J. Oper. Res.*, vol. 272, no. 2, pp. 621–635, 2019.
[30] Y. Wang, B. Ning, F. Cao, B. D. Schutter, and J. J. Ton van den Boom, "A survey on optimal trajectory planning for train operations," in *Proc. IEEE Int. Conf. Service Oper., Logistics, Inform. (SOLI)*, Beijing, China, Jul. 2011, pp. 589–594.
[31] Z. Wei and S. Link, "Embedded cardinality constraints," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.* Cham, Switzerland: Springer, 2018, pp. 523–538.
[32] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.
[33] X. Xu, K. Li, and X. Li, "A multi-objective subway timetable optimization approach with minimum passenger time and energy consumption," *J. Adv. Transp.*, vol. 50, no. 1, pp. 69–95, 2015.
[34] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$L_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, Jul. 2012.
[35] S. Yang, J. Wu, H. Sun, X. Yang, Z. Gao, and A. Chen, "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, Transp. Dyn.*, vol. 6, no. 4, pp. 302–319, 2017.
[36] S. Yang, J. Wu, X. Yang, H. Sun, and Z. Gao, "Energy-efficient timetable and speed profile optimization with multi-phase speed limits: Theoretical analysis and application," *Appl. Math. Model.*, vol. 56, pp. 32–50, Apr. 2018.
[37] X. Yang, A. Chen, X. Li, B. Ning, and T. Tang, "An energy-efficient scheduling approach to improve the utilization of regenerative energy for metro systems," *Transp. Res. C, Emerg. Technol.*, vol. 57, pp. 13–29, Aug. 2015.
[38] X. Yang, X. Li, Z. Gao, H. Wang, and T. Tang, "A cooperative scheduling model for timetable optimization in subway systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 438–447, Mar. 2013.
[39] X. Yang, X. Li, B. Ning, and T. Tao, "A survey on energy-efficient train operation for urban rail transit," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 2–13, Jan. 2016.
[40] X. Yang, B. Ning, X. Li, and T. Tang, "A two-objective timetable optimization model in subway systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1913–1921, Oct. 2014.
[41] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, Jan. 2019.
[42] A. Zaboli, B. Vahidi, S. Yousefi, and M. M. Hosseini-Biyouki, "Evaluation and control of stray current in DC-electrified railway systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 974–980, Feb. 2017.
[43] H. Zhang, L. Jia, L. Wang, and X. Xu, "Energy consumption optimization of train operation for railway systems: Algorithm development and real-world case study," *J. Cleaner Prod.*, vol. 214, pp. 1024–1037, Mar. 2019.
[44] W. Zhu, Z. Dong, Y. Yu, and J. Chen, "Lagrange dual method for sparsity constrained optimization," *IEEE Access*, vol. 6, pp. 28404–28416, 2018.

ZIYAN LUO was born in Hunan, China, in 1983. She received the Ph.D. degree in operations research and cybernetics from Beijing Jiaotong University, Beijing, China, in 2010.

She was a Research Associate with The Hong Kong Polytechnic University, in 2010, 2015, 2018, and 2019, and a Visiting Scholar with Stanford University, from 2011 to 2012 and also with the National University of Singapore, from 2015 to 2016. She is currently an Associate Professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. She has coauthored the book *Tensor Analysis: Spectral Theory and Special Tensors* (SIAM Press, 2017). Her research interests include sparse and low-rank optimization methods and applications, tensor analysis, and computation.

XIAOTONG YU was born in Shandong, China, in 1996. She is currently pursuing the degree with the Department of Mathematics, Beijing Jiaotong University.

XIAOYU LI was born in Inner Mongolia, China, in 1994. She is currently pursuing the degree with the Department of Mathematics, Beijing Jiaotong University.