Research Article

A Sparse Optimization Approach for Energy-Efficient Timetabling in Metro Railway Systems

Ziyan Luo,1 Xiaoyu Li,2 and Naihua Xiu2

1State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
2Department of Mathematics, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Xiaoyu Li; 16121607@bjtu.edu.cn

Received 5 July 2018; Accepted 8 October 2018; Published 1 November 2018

Academic Editor: Luca Pugi

Copyright © 2018 Ziyan Luo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we propose a sparse optimization approach to maximize the utilization of regenerative energy produced by braking 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 existing two-step linear program approach is also conducted which illustrates the effectiveness of our proposed sparse optimization model in terms of the energy saving rate and the efficiency of our numerical optimization algorithm in terms of computational time.

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 [1], with focuses on reviewing the mathematical problem formulations and solution approaches in the literature. Related work on energy optimization for rail traffic systems such as tramway systems has also been well studied; see, e.g., [2, 3].

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. [4] 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 are 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 [4], Das Gupta et al. developed a more tractable mixed integer programming model in [5]. In 2013, Yang et al. [6] presented a cooperative integer programming model to maximize the time overlaps of nearby accelerating and braking trains and applied a genetic algorithm to 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 [7] where the genetic algorithm was again adopted to solve the optimization model. An integrated energy-efficient operation model was proposed by Li & Lo in [8] with the speed profile part to minimize the net energy consumption and the timetable part with headway constraints to synchronize the accelerating and regenerative braking for the reuse of regenerative energy. About 25% energy savings were 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 [9]. 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 were also considered by Yang et al. in [10] 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. [11] 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 [12]. 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 \( \ell_0 \)-norm and the \( \ell_1 \)-norm were employed in [12]. By adopting the \( \ell_1 \)-norm heuristic for the original \( \ell_0 \)-norm (i.e., the number of nonzero components in a vector), together with epigraph approach that equivalently transforms the convex \( \ell_1 \)-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 \( \ell_0 \)-norm minimization problem proposed in [12] is well-known to be NP-hard due to the discontinuity and nonconvexity of the \( \ell_0 \)-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 proposed to replace the \( \ell_0 \)-norm, such as the popular convex surrogate \( \ell_1 \)-norm [13], and some weighted \( \ell_1 \)-norm variants [14]. Numerical algorithms are also developed for the resulting relaxation models, such as the greedy-type methods [15], the first-order optimization methods [16], and second-order methods for some special structured models [17].

Inspired by the development of sparse optimization methods, in this paper, a sparse optimization model with the \( \ell_0 \)-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 methods, in this paper, a sparse optimization model with the \( \ell_0 \)-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 methods [12], a quadratic model merely considering the squared Euclidean norm, and our weighted-\( \ell_1 \) companied with the squared Euclidean norm are applied to Beijing Metro Yizhuang Line for comparison. The numerical results on energy savings show that our model is the best among them. To demonstrate the efficiency of our proposed TSADMM, the academical version of CVX (http://cvxr.com/cvx/) is called to solve our model as well. The numerical results in computation time show 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 \( \ell_1 \) 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, the appendix 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.

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, following from [12].

(i) Trip time constraint: it includes the trip time constraints with tracks and with crossing-overs. For the former one, when a train \( t \in T \) has a trip from platform \( i \in N^s \) to platform \( j \in N^s \) along the track \( (i, j) \in T \), the departure time \( d^t_i \), and the arrival
time $a'_i$ are bounded by $\mathbf{1}^t_{ij}$ and $\mathbf{1}^t_{ij}$, respectively. The corresponding constraints are

$$\mathbf{1}^t_{ij} \leq a'_i - d'_i \leq \mathbf{1}^t_{ij}, \quad \forall t \in T, \quad \forall (i, j) \in \mathcal{H}.$$  \hspace{1cm} (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 nonopposite 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 [20]. Let $\phi$ 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 $\mathcal{O}_{ij}$. 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 $[\mathbf{1}^t_{ij}, \mathbf{1}^t_{ij}]$. And the corresponding constraints can be described as below:

$$\mathbf{1}^t_{ij} \leq a'_i - d'_i \leq \mathbf{1}^t_{ij}, \quad \forall (i, j) \in \phi, \quad \forall (t, t') \in \mathcal{O}_{ij}. \hspace{1cm} (2)$$

(ii) Dwell time constraint: when a train $t \in \mathcal{T}$ arrives at a platform $i \in \mathcal{N}$, it should dwell there in a time interval denoted by $[\mathbf{1}^w_{ij}, \mathbf{1}^w_{ij}]$ to make sure that passengers can get off and get on the train before

<table>
<thead>
<tr>
<th>Table 1: A list of notations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a'_i$</td>
</tr>
<tr>
<td>$d'_i$</td>
</tr>
<tr>
<td>$\mathcal{T}$</td>
</tr>
<tr>
<td>$</td>
</tr>
<tr>
<td>$[\mathbf{1}^t_{ij}, \mathbf{1}^t_{ij}]$</td>
</tr>
<tr>
<td>$\phi$</td>
</tr>
<tr>
<td>$\mathcal{O}_{ij}$</td>
</tr>
<tr>
<td>$[\mathbf{1}^t_{ij}, \mathbf{1}^t_{ij}]$</td>
</tr>
<tr>
<td>$[\mathbf{1}^w_{ij}, \mathbf{1}^w_{ij}]$</td>
</tr>
<tr>
<td>$\delta$</td>
</tr>
<tr>
<td>$\mathcal{S}$</td>
</tr>
<tr>
<td>$\mathcal{E}_{ij}$</td>
</tr>
<tr>
<td>$\mathcal{P}_{ij}$</td>
</tr>
<tr>
<td>$\mathcal{K}$</td>
</tr>
<tr>
<td>$\mathcal{N}$</td>
</tr>
<tr>
<td>$\mathcal{A}$</td>
</tr>
<tr>
<td>$\delta_{\phi}(\cdot)$</td>
</tr>
<tr>
<td>$\Pi_{\mathcal{S}}(\cdot)$</td>
</tr>
<tr>
<td>$\text{sign}(\cdot)$</td>
</tr>
<tr>
<td>$\circ$</td>
</tr>
<tr>
<td>$\mathbf{A}^T$</td>
</tr>
<tr>
<td>$\text{vec}(X)$</td>
</tr>
</tbody>
</table>
it departs from the platform \( i \). The dwell time constraint can be written as follows:

\[
\frac{d_{ij}^t}{\theta_i} \leq d_i^t - a_i^t \leq \frac{d_{ij}^t}{\theta_i}, \quad \forall t \in T, \forall i \in N^i. \tag{3}
\]

(iii) 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 \((t, t') \in \mathcal{C}_{ij}\), 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 \([c_j^i, c_{ij}^t]\) to maintain that passengers can get off from the former train and get on the latter. The constraint of connection can be described as follows:

\[
c_{ij}^t \leq d_j^t - a_j^t \leq c_{ij}^t, \quad \forall (i, j) \in \mathcal{S}, \forall (t, t') \in \mathcal{C}_{ij}. \tag{4}
\]

(iv) 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 \( H_{ij} \) as the set of pairs of successive trains that pass by track \((i, j) \in \mathcal{H}\).

Consider \((t, t') \in \mathcal{H}_{ij}\), and let \([h_{ij}^t, h_{ij}^t]\) and \([h_{ij}^t, h_{ij}^t]\) be the time intervals that maintained the departures and arrivals of consecutive trains \( t \) and \( t' \) from and to platform \( i \) and platform \( j \), respectively. For any \((i, j) \in \mathcal{H}\) between platform \( i \) and platform \( j \), headway time constraint with tracks can be described as follows:

\[
h_{ij}^t \leq d_j^t - a_j^t \leq h_{ij}^t, \quad h_{ij}^t \leq d_j^t - a_j^t \leq h_{ij}^t, \quad \forall (i, j) \in \mathcal{H}, \forall (t, t') \in \mathcal{H}_{ij}. \tag{5}
\]

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 \( \mathcal{O}_{ij} \) be the set of all such two-train pairs \(\{(t_1, t'_1), (t_2, t'_2)\}\). There is an interval time window \([h_{ij}^t, h_{ij}^t]\) between the departures of trains \( t_1 \) and \( t_2 \) from platform \( i \) and a window \([h_{ij}^t, h_{ij}^t]\) 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:

\[
\begin{align*}
&h_{ij}^t \leq d_j^t - a_j^t \leq h_{ij}^t, \\
&h_{ij}^t \leq d_j^t - a_j^t \leq h_{ij}^t, \\
&\forall (i, j) \in \phi, \forall (t_1, t'_1, t_2, t'_2) \in \mathcal{O}_{ij}.
\end{align*} \tag{6}
\]

Besides, the headway time is related to the passenger demand and the number of trains. Let \( m' \) 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, \( D = c \times u \times m' \) [8]. Let \( h \) be the headway time. Since \( h = 3600 / m' \), 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.

(v) Total travel time constraint: when a train \( t \in T \) has a trip, traversing all platforms and tracks in chronological order, it can have the total travel time between \( T^t \) and \( T' \), to meet the needs of passengers in subway systems. Total travel time constraints can be written as follows:

\[
\frac{d_{ij}^t}{\theta_i} \leq d_i^t - a_i^t \leq \frac{d_{ij}^t}{\theta_i}, \quad \forall t \in T, \forall i \in N^i. \tag{8}
\]

(vi) 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:

\[
0 \leq a_i^t \leq m_T, \quad 0 \leq d_i^t \leq m_T, \quad \forall t \in T, \forall i \in N^i. \tag{9}
\]

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 [12] proposed an optimization model with linear inequality constraints to get a feasible timetable. The involved optimization model takes the form of

\[
\min \sum_{(i, j) \in \mathcal{H}, t \in T} f_{ij}(d_i^t - a_i^t) \tag{10}
\]

s.t. (1)-(9)
where $f_{ij}: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ with argument $(a_i^t - d_i^t)$ being the energy consumption function that 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 $x = \text{vec}(X^T)$ with

$$X = \begin{pmatrix}
a_1^1 & a_1^2 & \cdots & a_1^{|\mathcal{F}|-1} & a_1^{|\mathcal{F}|} \\
a_2^1 & a_2^2 & \cdots & a_2^{|\mathcal{F}|-1} & a_2^{|\mathcal{F}|} \\
a_3^1 & a_3^2 & \cdots & a_3^{|\mathcal{F}|-1} & a_3^{|\mathcal{F}|} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_i^1 & a_i^2 & \cdots & a_i^{|\mathcal{F}|-1} & a_i^{|\mathcal{F}|} \\
ad_1^1 & ad_1^2 & \cdots & ad_2^{|\mathcal{F}|-1} & ad_2^{|\mathcal{F}|} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
ad_i^1 & ad_i^2 & \cdots & ad_i^{|\mathcal{F}|-1} & ad_i^{|\mathcal{F}|}
\end{pmatrix}$$

(11)

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

$$
\begin{align*}
\text{min} & \quad \sum_{(i,j) \in \mathcal{K}_p} c_{ij} (x_j - x_i) \\
\text{s.t.} & \quad l_{ij} \leq x_j - x_i \leq u_{ij}, \quad (i, j) \in \mathcal{K} \\
& \quad 0 \leq x_i \leq m_T, \quad i = 1, 2, \ldots, n,
\end{align*}
$$

(12)

For all other constraints, lower and upper bounds are allowed to vary as described in (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 Figures 1 and 2.

Let $t_1$, $t_2$, and $t_3$ be the time of maximum accelerating, speed hold, and maximum braking phases, and $a_{\text{acc}}$ and $a_{\text{bra}}$ 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 $\theta_{\text{acc}}$ and the conversion factor from kinetic energy to regenerative electricity as $\theta_{\text{bra}}$ and the resistance rate at the speed holding phase as $\gamma$. 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 $E^{\text{con}}$ and $E^{\text{reg}}$, respectively, in the way as below:

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

$$= \frac{a_{\text{acc}}^2 t_i^2}{2 \theta_{\text{acc}}} + \frac{\gamma^2 t_2^2}{2 \theta_{\text{acc}}}$$

(15)

and $E^{\text{reg}} = \frac{\theta_{\text{bma}} a_{\text{bma}} t_i^2}{2}$.

Here $E^{\text{con}}$ and $E^{\text{reg}}$ are both for the unit mass. A simple heuristic method is applied to approximate the power graph as shown in Figure 3, where $c$ and $\bar{c}$ are the width of the rectangles for accelerating and braking and $h$ and $\bar{h}$ are heights of the rectangles. Similar to those in [12], the midpoint of the width in the rectangle is called the regenerative or consumptive alignment point, where $\Delta_i^j$ represents the distance between $d_i^j$ and the regenerative alignment point and $\bar{V}_i^j$ represents the distance between $a_i^j$ and the consumption alignment point.

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 [12]. Assume that platform pairs are opposite to each other at the same electrical substation. Let $P$ be the set of all platform pairs. For any pair $(i, j)$ in $P$ and the set $T_i \subseteq T$ contains all trains that arrives at, dwells, departs from platform $i$. When the train $t \in T_i$ departs from the platform $i$, it is a best choice to find a train $\bar{t}$ at the platform $j$ which are producing the regenerative braking energy using the following definition.

**Definition 1 (see [12]).** Consider any $(i, j) \in P$. For every train $t \in T_j$, the train $\bar{t} \in T_i$ is called the temporally closest train to the right of $t$ if

$$\bar{t} = \arg \min_{t' \in \{x \in T_j : d_{\bar{t}} < d_{\bar{t'}} \}} \left\{ \frac{\gamma_{\bar{t}}' + d_{\bar{t}}'}{2} \right\}$$

(16)

where $\beta$ is an empirical parameter determined by the timetable designer and is much smaller than the time horizon of the entire timetable [12]. Similarly, when a train $t \in T_i$ arrives at platform $i$, it produces the regenerative energy. To find a suitable train $\bar{t}$ departing from platform $j$ for transferring the regenerative braking energy, we recall the following definition from [12].

**Definition 2 (see [12]).** Consider any $(i, j) \in P$. For every train $t \in T_i$, the train $\bar{t} \in T_j$ is called the temporally closest train to the left of $t$ if

$$\bar{t} = \arg \min_{t' \in \{x \in T_j : d_{\bar{t}} < d_{\bar{t'}} \}} \left\{ \frac{\gamma_{\bar{t}} + d_{\bar{t}}}{2} \right\}$$

(17)

Denote the sets $\mathcal{E}$ and $\mathcal{E}$ whose components are $(i, j, t, \bar{t})$ and $(i, j, t, \bar{t})$, respectively. Consider any $(i, j, t, \bar{t})$ $\in \mathcal{E}$. To maximize the transfer of the regenerative energy from the braking train $\bar{t}$ by the accelerating train $t$, we attempt to make the term $(d_{\bar{t}} + \Delta_{\bar{t}}^t - a_{\bar{t}}^j + V_{\bar{t}}^j)$ to be zero or as close to zero as possible. Similarly, for $(i, j, t, \bar{t})$ $\in \mathcal{E}$, our goal is to make the term $(d_{\bar{t}} + \Delta_{\bar{t}}^t - a_{\bar{t}}^j + V_{\bar{t}}^j)$ to be zero or as close to zero as possible. Following from [12], we define an auxiliary variable $y \in \mathbb{R}^n$ with $n = 2|\mathcal{P}|, \mathcal{P} = |A|$ and $\mathcal{P} = |\mathcal{T}|$ by

$$y = \left( \left( d_{\bar{t}}^j + \Delta_{\bar{t}}^t - a_{\bar{t}}^j + V_{\bar{t}}^j \right)_{(i,j,t) \in \mathcal{E}} \right),$$

(18)

The components of $y$ represent difference of the time between the alignment points of accelerating and braking. Once the running time is fixed, $\Delta_{\bar{t}}^t$ and $V_{\bar{t}}^j$ are two constants.

The ultimate energy consumption can be approximately calculated as follows (see Figure 4):
By sorting all possible consecutive platforms for braking energy absorbing, a 0-1 vector $\alpha \in \mathbb{R}^n$ is introduced to indicate all those locations in $y$ with regeneration energy usage when the components of $\alpha$ 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 in $\alpha \ast y$ which corresponds to minimizing $\|\alpha \ast y\|_0$ and making nonzero elements in $\alpha \ast y$ as close to zero as possible which corresponds to minimizing $\|\alpha \ast y\|_2^2$. The resulting sparse optimization model turns out to be

$$
\min_{x,y} \lambda_1 \|\alpha \ast y\|_0 + \frac{\lambda_2}{2} \|\alpha \ast y\|_2^2
$$

s.t. (3)–(9), (13), (14), (18).

where $x, y \in \mathbb{R}^n$ are decision variables. Similar to (12), (20) can be rewritten as

$$
\min_{x,y} \left\{ \frac{\lambda_1}{2} \|\alpha \ast y\|_2^2 : y - Ax - b = 0, \; Fx - g = 0, \; f - Ex \geq 0 \right\}
$$

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

$$
E = \begin{pmatrix}
E_1 \\
\vdots \\
E_l
\end{pmatrix},
$$

$$
F = \begin{pmatrix}
O_{m \times m} & \tilde{E}_1 & O_{m \times m} & \cdots & O_{m \times m} \\
O_{m \times m} & O_{m \times m} & \tilde{E}_2 & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
O_{m \times m} & O_{m \times m} & \cdots & O_{m \times m} & \tilde{E}_{l-1}
\end{pmatrix},
$$

with

$$
E_1 = \begin{pmatrix}
\tilde{E}_1 & \cdots & \cdots & O_{m \times m} \\
O_{m \times m} & \tilde{E}_2 & \cdots & O_{m \times m} \\
\vdots & \vdots & \ddots & \vdots \\
O_{m \times m} & O_{m \times m} & \cdots & \tilde{E}_{l-1}
\end{pmatrix},
$$

$$
E_2 = (-I_m \otimes I_m),
$$

$$
E_3 = I_{m \times m}.
$$
strategy proposed by Zou et al. [14], we will further utilize a
rewritten as the following equality constrained optimization
problem:

$$\begin{align*}
\min \{ & \lambda_1 \|\alpha \circ y\|_1 + \frac{\lambda_2}{2} \|\alpha \circ y\|_2^2 : y - Ax - b = 0, \quad Fx - g = 0, \quad f - Ex \geq 0 \}\end{align*}$$

(24)

Inspired by Oracle property beneficial from the adaptive lasso strategy proposed by Zou et al. [14], we will further utilize a
rewighted $\ell_1$-norm term instead of the $\ell_1$-norm in (24) and
get the following relaxation counterpart:

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

(25)

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 $z \in \mathbb{R}^m$, (25) can be
rewritten as the following equality constrained optimization
problem:

$$\begin{align*}
\min_{x,y,z} \left\{ \lambda_1 \|\alpha \circ w \circ y\|_1 + \frac{\lambda_2}{2} \|\alpha \circ y\|_2^2 + \delta_{\mathbb{R}^m} (z) : y - Ax - b = 0, \quad Fx - g = 0, \quad z - f + Ex = 0 \right\}
\end{align*}$$

(26)

where $\delta_{\mathbb{R}^m} (h)$ is the indicator function with respect to
$\mathbb{R}^m$, which takes the value 0 if $h \in \mathbb{R}^m$, and $+\infty$ otherwise. The augmented Lagrangian function of (26) takes the form of

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

(27)

where $\sigma > 0$ is the penalty parameter, and $u \in \mathbb{R}^n, v \in \mathbb{R}^l, s \in \mathbb{R}^m$ are Lagrangian multipliers corresponding to
those three equality constraint systems in (26). The well-
known alternating direction method of multipliers (ADMM)
proposed by Glowinski and Marroco [23] and Gabay and
Mercier [24] is then adopted for handling (26) with the iterative scheme for $k = 0, 1, \ldots$

$$\begin{align*}
\left\{ y^{k+1}, z^{k+1} \right\} &= \arg \min \left\{ L^A_{\sigma}(x^k, y^k, z^k; u^k, v^k, s^k) \mid y \in \mathbb{R}^n, \quad z \in \mathbb{R}^m \right\} \\
x^{k+1} &= \arg \min \left\{ L^A_{\sigma}(x^k, y^{k+1}, z^{k+1}; u^k, v^k, s^k) \mid x \in \mathbb{R}^l \right\} \\
u^{k+1} &= (u^k, v^k, s^k) - \tau \sigma (y^{k+1} - Ax^{k+1} - b, Fx^{k+1} - g, z^{k+1} - f + Ex^{k+1} \rangle
\end{align*}$$

(28)

where $\tau \in (0, (1+\sqrt{5})/2)$ 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 (28) can be easily solved by the corresponding
first-order optimality condition which leads to the following
closed-form solution:

$$\begin{align*}
y_i^{k+1} &= \begin{cases} 
\left( Ax_i + b + \frac{u_i^k}{\sigma} \right)_i \quad \text{if } \alpha_i = 0, \\
\text{sign}(q_i) \max \left\{ |q_i| - \frac{\lambda_1 w_i}{\lambda_2 + \sigma}, 0 \right\} \quad \text{otherwise},
\end{cases}
\end{align*}$$

(29)

with $q_i = (1/(\lambda_2 + \sigma))(u_i + \sigma Ax_i + \sigma b)$, for all $i = 1, \ldots, n$, and

$$z^{k+1} = \Pi_{\mathbb{R}^m} \left( f - Ex^{k+1} + \frac{1}{\sigma} s^k \right),$$

(30)

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

$$Hx = rhs,$$

(31)

where $rhs = A^T (y^{k+1} - b - (1/\sigma)u^k) + F^T (g + (1/\sigma)v^k) + E^T (f - z^{k+1} + (1/\sigma)s^k)$ and $H = A^T A + F^T F + E^T E$. Note that $E$ is full
column rank. Thus, $E^T E$ is positive definition and also is $H$. This indicates that (31) admits a unique solution:

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

(32)

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 (31) instead of the direct way as in (32).

It is worth mentioning that we can utilize the above ADMM method to solve an $\ell_1$ problem which is exactly (26) by taking $w = e$ and $\lambda_2 = 0$ to a relatively low accuracy. The resulting numerical solution $(\tilde{x}, \tilde{y}, \tilde{z})$ can not only be served as the initial point, but also produce the weight vector $w$ in the way that $w = (1/(\|j^T\| + e))$ with $r \in (0, +\infty)$ and $e > 0$ sufficiently small for a low sparsity promotion when solving (26). 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 the appendix.

**Theorem 3.** Let $\{x^k, y^k, z^k, w\}$ be generated from Algorithm 1. If $\tau \in (0, (1 + \sqrt{5})/2)$, then the sequence $\{x^k, y^k, z^k\}$ converges to an optimal solution to (26).

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 Figure 6. 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 Figure 5. 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.

4.1. Inputs and Algorithm Initialization. For (26), $\alpha \in \mathbb{R}^n$ in the objective function is a 0-1 vector defined in Section 2.4. And the weight vector is denoted by $w = (1/(\|j^T\| + e))$, where $\tilde{y}$ is an approximate solution generated by Stage I in Algorithm 1. Here we set $r = 0.5$ and $e = 10^{-20}$. In the equality constraint $y - Ax - b = 0$, $A \in \mathbb{R}^{m \times n}$ with $m = 2n \overline{m}$, $\overline{m} = |X| = 28$, $\overline{m}$ the number of trains in one day service, $b \in \mathbb{R}^n$ with components $(\Delta_i^T + V_i^T j_{i,j}^T)$ 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 $i + j = 29$. $Ex - g = 0$ in (26) is an equality constraint about trip times, $F \in \mathbb{R}^{m \times n}$ is the coefficient matrix in (26), $l = (\overline{T} - 1)\overline{m}$. The components of $g \in \mathbb{R}^l$ are the trip times. $E \in \mathbb{R}^{m \times n}$ in $z - f + Ex = 0$ is the coefficient matrix of inequality constraints, and $m = 2(2n(\overline{m} - 1) + n\overline{m} + n\overline{m} + n)$. The elements in the vector $f \in \mathbb{R}^m$ are upper and lower bounds of dwell time constraints, headway time constraints, total travel time constraints, and domain of event times constraints. For testing purpose, elements of $f$ are given in the following way. The bounds of dwell time constraints are given in Table 2. Let $D$ be the passenger demand for one-day service (16 hours) and $u^{(k)} \in \mathbb{R}^D$ be the distribution of $D$ for each hour which is presented in Table 3. Let $c_i$ be the capacity of a train, which takes the value 1440, and $\bar{u} \in \mathbb{R}^D$ be the train utility rate for different hours as stated in Table 3 as well. By utilizing formula (7), we can get the headway time for each different period of time (per hour). Specifically, denote the headway time vector by $h \in \mathbb{R}^D$. Then $h_k = (3600 \times 1440 \times u_k)/(D \times u_k)$ for all $k = 1, 2, \ldots, 16$. Set the windows for the headway time constraints to be $h_k \pm 8$ for all possible train pairs in the $i^{th}$ hour. The total travel time constraints are related
Table 2: The windows of the dwell time constraints.

<table>
<thead>
<tr>
<th>Platforms</th>
<th>SJZ1</th>
<th>XCl</th>
<th>XHM1</th>
<th>JG1</th>
<th>YZQ1</th>
<th>YZWHY1</th>
<th>WYJ1</th>
<th>RJDJ1</th>
<th>RCDJ1</th>
<th>TJNL1</th>
<th>JHL1</th>
<th>CQN1</th>
<th>CQ1</th>
<th>YZHCZ1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper bounds</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Platforms</th>
<th>SJZ2</th>
<th>XCl</th>
<th>XHM2</th>
<th>JG2</th>
<th>YZQ2</th>
<th>YZWHY2</th>
<th>WYJ2</th>
<th>RJDJ2</th>
<th>RCDJ2</th>
<th>TJNL2</th>
<th>JHL2</th>
<th>CQN2</th>
<th>CQ2</th>
<th>YZHCZ2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper bounds</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
</tbody>
</table>
Table 3: The train utility rate and distribution of $D$ for one-day (16 hours) services.

<table>
<thead>
<tr>
<th>Periods</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u$</td>
<td>0.3</td>
<td>0.8</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td>0.6</td>
<td>0.6</td>
<td>0.8</td>
<td>0.8</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>$u^D$</td>
<td>0.03</td>
<td>0.11</td>
<td>0.13</td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
<td>0.13</td>
<td>0.12</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>
to the dwell time and trip time. The bounds of trip times are calculated by Table 4 as reference. For each train \( t \), the bounds of trip time are more than \( \sum_{q=1}^{5} (s_q/v_q) \) for all sections \( (ij) \), \( i = 1, 2, \ldots , 27, j = i + 1 \). The bounds of trip time constraints are presented in Table 5. The following formulae are adopted for calculating the windows of total travel time constraints,

\[
\begin{align*}
TT &= r_1 \times \left( \sum_{i=1}^{28} dw_i + \sum_{i=1, i \neq 14}^{27} tr_{ij} + \sum_{i=14}^{26} k_i \right) \\
\overline{TT} &= r_2 \times \left( \sum_{i=1}^{28} dw_i + \sum_{i=1, i \neq 14}^{27} tr_{ij} + \sum_{i=14}^{26} k_i \right)
\end{align*}
\]

(33)

where \( r_1 = 1.1 \) and \( r_2 = 0.85 \). For the domain of event times constraints, we set \( m_r = 16 \times 3600 \).

The initial point \((u^0, v^0, d^0)\) in Algorithm 1 is chosen to be 0. To calculate the energy consumption and regenerative energy, we set the maximum acceleration \( a_{acc} = 0.5 m/s^2 \) in accelerating, \( a_{braking} = -0.8 m/s^2 \) for braking, the resistance acceleration \( g = -0.05m/s^2 \) 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 \( \theta_{braking} = 0.5 \) as used in [18]. For testing purpose, we approximately calculate the energy consumption for accelerating and the regenerative energy produced during braking in the following way. As presented in Section 2.4, the maximum accelerating, speed holding, maximum braking strategy from [18] is adopted and the parameters \( t_1, t_2, t_3 \) are set in Table 6. The parameters \( c \) and \( \bar{c} \) in Figure 3 which represent
Table 4: Speed limits of every section on Beijing Metro Yizhuang Line following from [18].

<table>
<thead>
<tr>
<th>Sections</th>
<th>Segment</th>
<th>$v_1$(m/s)</th>
<th>$s_1$(m)</th>
<th>$v_2$(m/s)</th>
<th>$s_2$(m)</th>
<th>$v_3$(m/s)</th>
<th>$s_3$(m)</th>
<th>$v_4$(m/s)</th>
<th>$s_4$(m)</th>
<th>$v_5$(m/s)</th>
<th>$s_5$(m)</th>
<th>$v_6$(m/s)</th>
<th>$s_6$(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJZ-XC</td>
<td>0-150</td>
<td>13.89</td>
<td></td>
<td>23.61</td>
<td></td>
<td>150-480</td>
<td></td>
<td>18.06</td>
<td></td>
<td>480-1161</td>
<td></td>
<td>1161-2501</td>
<td></td>
</tr>
<tr>
<td>XC-XHM</td>
<td>16.67</td>
<td>2631-2643</td>
<td></td>
<td>23.61</td>
<td></td>
<td>2643-2797</td>
<td></td>
<td>20.83</td>
<td></td>
<td>2797-3534</td>
<td></td>
<td>3534-3780</td>
<td></td>
</tr>
<tr>
<td>XHM-JG</td>
<td>16.67</td>
<td>3905-3918</td>
<td></td>
<td>23.61</td>
<td></td>
<td>3918-5808</td>
<td></td>
<td>20.83</td>
<td></td>
<td>5808-6141</td>
<td></td>
<td>6141-6271</td>
<td></td>
</tr>
<tr>
<td>JG-YZQ</td>
<td>16.67</td>
<td>6271-6281</td>
<td></td>
<td>23.61</td>
<td></td>
<td>6281-8122</td>
<td></td>
<td>16.67</td>
<td></td>
<td>8122-8254</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>YZQ-YZWHY</td>
<td>16.67</td>
<td>8254-8265</td>
<td></td>
<td>23.61</td>
<td></td>
<td>8265-9216</td>
<td></td>
<td>16.67</td>
<td></td>
<td>9116-9246</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>YZWHY-WYJ</td>
<td>16.67</td>
<td>9246-9256</td>
<td></td>
<td>23.61</td>
<td></td>
<td>9259-10655</td>
<td></td>
<td>16.67</td>
<td></td>
<td>10655-10785</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>WYJ-RJDJ</td>
<td>16.67</td>
<td>10785-10797</td>
<td></td>
<td>23.61</td>
<td></td>
<td>10797-18933</td>
<td></td>
<td>16.67</td>
<td></td>
<td>11933-12065</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>RJDJ-RCDJ</td>
<td>16.67</td>
<td>12065-12077</td>
<td></td>
<td>23.61</td>
<td></td>
<td>12077-13289</td>
<td></td>
<td>16.67</td>
<td></td>
<td>13289-13419</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>RCDJ-TJNL</td>
<td>16.67</td>
<td>13419-13431</td>
<td></td>
<td>23.61</td>
<td></td>
<td>13419-14649</td>
<td></td>
<td>20.83</td>
<td></td>
<td>14649-15426</td>
<td></td>
<td>15426-15624</td>
<td></td>
</tr>
<tr>
<td>TJNL-JHL</td>
<td>16.67</td>
<td>15756-15768</td>
<td></td>
<td>23.61</td>
<td></td>
<td>15768-17891</td>
<td></td>
<td>16.67</td>
<td></td>
<td>17891-18021</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>JHL-CQN</td>
<td>16.67</td>
<td>18021-18033</td>
<td></td>
<td>23.61</td>
<td></td>
<td>18033-19982</td>
<td></td>
<td>16.67</td>
<td></td>
<td>19982-20107</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CQN-CQ</td>
<td>16.67</td>
<td>20107-20120</td>
<td></td>
<td>23.61</td>
<td></td>
<td>20120-21264</td>
<td></td>
<td>16.67</td>
<td></td>
<td>21264-21394</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CQ-YZHCZ</td>
<td>16.67</td>
<td>21394-21406</td>
<td></td>
<td>23.61</td>
<td></td>
<td>21406-22569</td>
<td></td>
<td>16.67</td>
<td></td>
<td>22569-22728</td>
<td></td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Table 5: The windows of the trip time constraints.

<table>
<thead>
<tr>
<th>Sections</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower bounds</td>
<td>160</td>
<td>92</td>
<td>150</td>
<td>130</td>
<td>80</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>140</td>
<td>130</td>
<td>130</td>
<td>100</td>
<td>100</td>
<td>170</td>
</tr>
<tr>
<td>Upper bounds</td>
<td>190</td>
<td>110</td>
<td>170</td>
<td>150</td>
<td>110</td>
<td>125</td>
<td>120</td>
<td>120</td>
<td>170</td>
<td>160</td>
<td>145</td>
<td>120</td>
<td>120</td>
<td>200</td>
</tr>
<tr>
<td>Sections</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>–</td>
</tr>
<tr>
<td>Lower bounds</td>
<td>100</td>
<td>100</td>
<td>130</td>
<td>130</td>
<td>140</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>80</td>
<td>130</td>
<td>150</td>
<td>92</td>
<td>160</td>
<td>–</td>
</tr>
<tr>
<td>Upper bounds</td>
<td>120</td>
<td>120</td>
<td>145</td>
<td>160</td>
<td>170</td>
<td>120</td>
<td>120</td>
<td>125</td>
<td>110</td>
<td>150</td>
<td>170</td>
<td>110</td>
<td>190</td>
<td>–</td>
</tr>
</tbody>
</table>
the width of the rectangles for $E^\text{com}$ and $E^\text{bra}$ (in percentage of the trip time) are also given in Table 6 for approximately calculating the energy consumption for accelerating $E^\text{com}$ and the regenerative energy produced during braking $E^\text{bra}$.

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

$$\eta_p = \max \left\{ \frac{\|y\|_2 + \|Fx - g\|_2 + \|z - f + Ex\|_2}{1 + \|y\|_2 + \|g\|_2 + \|z\|_2} \right\},$$

$$\eta_d = \frac{|A^T u - F^T w - E^T s|}{1 + |A^T u| + |F^T w| + |E^T s|}.$$ (34)

For a given tolerance $\varepsilon_2 > 0$, we will stop the tested algorithms when both $\eta_p$ and $\eta_d$ are less than $\varepsilon_2$. The algorithms will also be terminated when they reach a given maximum number of iterations $\text{MaxIter}_2$. Here we set $\varepsilon_2 = 10^{-3}$ and $\text{MaxIter}_2 = 10^3$. For Stage I in Algorithm 1, it will be stopped when both $\eta_p$ and $\eta_d$ are less than $\varepsilon_1$ or the maximum iteration reaches $\text{MaxIter}_1$. Here we set $\varepsilon_1 = 10^{-1}$ and $\text{MaxIter}_1 = 10^3$.

4.3. Experiment Results. To evaluate our model with $\ell_0$-norm plus $\ell_2$-norm as the object, the other two possible models are computed for comparison. The first model is the $\ell_0$-norm plus $\ell_1$-norm proposed by Das Gupta et al. in [12] which takes the form of

$$\min_x \left\{ \|y\|_0 + \lambda \|y\|_1 : y - Ax - b = 0, \ Fx - g = 0, \ Ex \leq f \right\},$$ (35)

which is handled by approximately solving its linear programming relaxation problem

$$\min_x \left\{ \frac{1}{2} \|y\|_2^2 : y - Ax - b = 0, \ Fx - g = 0, \ Ex \leq f \right\},$$ (37)

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 $\eta$ as follows:

$$\eta = \frac{\sum_{i=1}^{\#\ell_i} \|E^\text{con}_{ij}\|_2}{\sum_{i=1}^{\#\ell_i} \|E^\text{con}_{ij}\|_2}.$$ (38)

Table 6: Parameters for calculating $E^\text{com}$ and $E^\text{bra}$.

<table>
<thead>
<tr>
<th>$t_1$ (%)</th>
<th>$t_2$ (%)</th>
<th>$t_3$ (%)</th>
<th>$c$ (%)</th>
<th>$\varepsilon$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>45</td>
<td>35</td>
<td>14</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 7: Results of numerical experiments on Beijing Metro Yizhuang Line. $A_1$ stands for the linear programming model (36) solved by Gurobi; $A_2$ stands for the quadratic programming (37) solved by CVX; $A_3$ stands for our proposed model (26) solved by Algorithm 1.

Table 8: Energy saving rates for $\bar{m} = 368$: $A_1$, $A_2$, and $A_3$ are the same as in Table 7.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>$A_1$ (%)</th>
<th>$A_2$ (%)</th>
<th>$A_3$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[9]</td>
<td>18.85</td>
<td>18.83</td>
<td>23.11</td>
</tr>
<tr>
<td>[10]</td>
<td>18.99</td>
<td>18.88</td>
<td>23.11</td>
</tr>
<tr>
<td>[6]</td>
<td>18.97</td>
<td>18.78</td>
<td>23.29</td>
</tr>
</tbody>
</table>

As shown in Table 7, the energy savings by using our model are better than those by the other two in all the testing instances with at least 2.58 percentage higher in the energy saving rate $\eta$.

Other choices for the initial trip times from [6, 9–11, 19] with $\bar{m} = 368$ are also considered as inputs. The comparison results in terms of the energy saving rate $\eta$ are stated in Table 8.
In order to demonstrate the efficiency of our proposed TSADMM algorithm, we call the academic CVX to solve ours model for comparison. The time comparisons for all the aforementioned instances are presented in Tables 9 and 10, respectively.

As can be seen in Tables 9 and 10, our proposed TSADMM outperforms the CVX solver for the model (26) in terms of computation time. Particularly, for the cases of \( \overline{m} \) equaling 368 and 449 in Table 9, the computation time for TSADMM is less than one-third of those for CVX. 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 Figures 7 and 8. 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.

**5. Conclusion**

In this paper we have proposed a sparse optimization model using \( \ell_0 \)-norm and squared \( \ell_2 \)-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 \( \ell_0 \)-norm, the weighted \( \ell_1 \)-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.
Table 9: Time comparison between CVX and TSADMM with different choices of $m$.

<table>
<thead>
<tr>
<th>Alg.</th>
<th>$m$</th>
<th>time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>357</td>
<td>47.17</td>
</tr>
<tr>
<td></td>
<td>368</td>
<td>47.06</td>
</tr>
<tr>
<td></td>
<td>380</td>
<td>54.33</td>
</tr>
<tr>
<td></td>
<td>391</td>
<td>56.90</td>
</tr>
<tr>
<td></td>
<td>403</td>
<td>31.09</td>
</tr>
<tr>
<td></td>
<td>414</td>
<td>50.63</td>
</tr>
<tr>
<td></td>
<td>426</td>
<td>56.18</td>
</tr>
<tr>
<td></td>
<td>437</td>
<td>70.72</td>
</tr>
<tr>
<td></td>
<td>449</td>
<td>71.14</td>
</tr>
<tr>
<td>CVX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSADMM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24.35</td>
<td>12.80</td>
</tr>
<tr>
<td></td>
<td>31.06</td>
<td>21.26</td>
</tr>
<tr>
<td></td>
<td>25.39</td>
<td>19.36</td>
</tr>
<tr>
<td></td>
<td>27.61</td>
<td>30.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23.66</td>
</tr>
</tbody>
</table>
Table 10: Time comparison between CVX and TSADMM with different trip times from [6, 9–11, 19].

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CVX</td>
<td>55.61</td>
<td>57.86</td>
<td>55.45</td>
<td>51.25</td>
<td>47.81</td>
</tr>
<tr>
<td>TSADMM</td>
<td>20.36</td>
<td>29.22</td>
<td>28.21</td>
<td>27.64</td>
<td>31.84</td>
</tr>
</tbody>
</table>

![Graphs](image.png)

Figure 8: The numerical convergence behavior of TSADMM for different trip times: tr1, ..., tr5 stand for the cases for different trip times proposed in [6, 9–11, 19].

Since the weighted \( \ell_1 \)-norm is a convex approximation of the \( \ell_0 \)-norm, the optimal solution to the (26) is a proximal optimal solution to the original sparse optimization model. How to design some more effective nonconvex relaxation scheme to get a better proximal optimal solution so as to get a better energy saving rate will be one of our future work. Meanwhile, the delay perturbations of the metro trains and the randomness of passenger demands have not been considered in our static timetabling. How to use advanced optimization techniques for handling such cases will also be our future research issue.

Appendix

The Proof of Theorem 3

**Proof.** Let \( \mathbf{x} = ([x^{(1)}]^T, \ldots, [x^{(\infty)}]^T)^T \in \mathbb{R}^n \) be an optimal solution of Problem (12). Then, the optimality and the feasibility of \( \mathbf{x} \) imply that \( F\mathbf{x} - g = 0 \) and \( f - E\mathbf{x} \geq 0 \). By virtue of the expressions of \( E \) and \( F \) in Section 2.4, together with the time window constraints as stated in \( f \), we can find some \( \mathbf{z} \in \mathbb{R}^m \) in a neighborhood \( N(\mathbf{x}, \delta) \) with \( \delta \) such that \( F\mathbf{x} - g = 0 \) and \( f - E\mathbf{x} > 0 \). Set \( \mathbf{y} = A\mathbf{x} + b \) and \( \mathbf{z} = f - E\mathbf{x} \). It is easy to verify that \( (\mathbf{x}, \mathbf{y}, \mathbf{z}) \in (\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+^m) \cap P \) with \( P \) being the feasible region of Problem (26). By employing [25, Theorem B.1], we can get the desired global convergence result in Theorem 3.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was partly supported by the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University [Grants nos. RCS2017J001 and RCS2016ZZ01], and National
Natural Science Foundation of China [Grants nos. 11771038 and 11431002].

References


