Separation of rail and wheel contributions to pass-by noise with sparse regularization methods

Zea, Elias; Fernandez-Grande, Efren; Lopez Arteaga, Ines

Published in:
Journal of Sound and Vibration

Link to article, DOI:
10.1016/j.jsv.2020.115627

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Zea, E., Fernandez-Grande, E., & Lopez Arteaga, I. (2020). Separation of rail and wheel contributions to pass-by noise with sparse regularization methods. Journal of Sound and Vibration, 487, [115627]. https://doi.org/10.1016/j.jsv.2020.115627

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Separation of rail and wheel contributions to pass-by noise with sparse regularization methods

Elias Zea\textsuperscript{a,1,*}, Efren Fernandez-Grande\textsuperscript{b}, Ines Lopez Arteaga\textsuperscript{a,c}

\textsuperscript{a} KTH Royal Institute of Technology, Department of Engineering Mechanics, The Marcus Wallenberg Laboratory for Sound and Vibration Research, Stockholm SE-10044, Sweden
\textsuperscript{b} Technical University of Denmark, Department of Electrical Engineering, Acoustic Technology, Lyngby 2800 Kgs., Denmark
\textsuperscript{c} Eindhoven University of Technology, Department of Mechanical Engineering, MB Eindhoven 5600, The Netherlands

\section*{Article Info}

\begin{itemize}
  \item Article history:
  \item Received 19 February 2020
  \item Revised 26 June 2020
  \item Accepted 8 August 2020
  \item Available online 8 August 2020
\end{itemize}

\section*{Abstract}

This paper proposes a method for separating rail and wheel noise contributions via sparse regularization of microphone array data. The underlying idea is to promote sparse solutions, which jointly approximate the two noise contributions with few non-zero coefficients. The main hypothesis is that the acoustic field radiated by the rail is sparse in a dictionary of plane waves, that the acoustic field radiated by the wheels is sparse in a dictionary of moving sources, and that both acoustic fields are dense in the opposite dictionaries. How well this happens is studied with the coherence between the plane waves and the moving sources. The strength of the proposed approach is that it does not require static tests prior to the pass-by. The separation is performed in three main steps, executed in the time-frequency domain. First, the rail contribution is separated from the total pass-by noise using matching pursuit optimization, promoting solutions with a limited number of plane waves per frequency. Second, the residual between the total pass-by noise and the estimated rail noise is calculated. And third, the wheel contribution is separated from the residual via $\ell_1$-norm minimization, promoting solutions that are row-sparse at all frequencies. The separation performance is investigated with synthetic data, and validated with experimental data against reference predictions with the TWINS software, for pass-by noise measurements of trains running at 40, 80 and 160 km/h.

\copyright 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license. (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

There is a present urge for quantifying railway pass-by noise regardless of the track on which the train is running. Specially so for running speeds under 300 km/h, a regime in which rolling noise is a dominant source. In this regime, the vehicle noise—the target component for the homologation of European trains [1], is often masked by the noise radiated by the track in the frequency range between 500 and 1600 Hz; leading to geographically-dependent sound emissions of the...
same vehicle. This dependency is traditionally taken care of by means of transporting the vehicle and measuring the pass-by noise in so-called “silent” tracks.

A number of methods based on microphone array data have been proposed as an alternative to this traditional solution. In 2006, Kitagawa and Thompson reported the use of 1D beamforming arrays to quantify the wheel and rail noise contributions. The approach resulted in an under-estimation of the rail contribution, attributed to the normal orientation of the beamformer towards the track [2]. A few years later, the problem was revisited by steering the beamformer towards a radiation angle that is dictated by the ratio between structural and acoustic wavelengths [3]. Reasonable agreement could be found by tuning the beamformer angles; however, no clear strategy for choosing such angles could be provided. Similar conclusions were drawn in 2012 by Le Courtois et al. with 2D beamforming array data [4], and further examinations on more complex source directivity were recently carried out by Zhang et al. [5]. Another approach, proposed by Faure et al. in 2015, is the SWEAM method [6], which approximates the sound radiation of the track with cylindrical waves, and performs an inverse estimation of the structural wavenumbers and decay rates of the waves. More recently, Zea et al. introduced the wave signature extraction (WSE) method [7], whereby using wavenumber band-pass filters the rail contribution is extracted from microphone array measurements of the pass-by near the track. As the WSE method is tailored for the rail contribution, its obvious limitation is the inability to estimate the wheel contribution. For a recent round robin of state-of-the-art separation methods the reader is referred to the paper by Thompson et al. [8].

A different perspective to the separation problem is related to the concept of sparsity. A signal that is sparse has few non-zero active coefficients, meaning that only such coefficients should ideally allow for the recovery of the entire signal. This rationale was first applied in the 1970s by Claerbout and Muir [9], who were concerned with the problem of statistical regression of geophysical data. Since then, mathematicians put a tremendous effort in understanding the machinery underlying sparse signal recovery, eventually leading to the sampling paradigm of compressed sensing, formalized in 2006 by Candès et al. and Donoho [10,11]. In particular, numerous applications of sparse regularization methods to the problem of sound source quantification have appeared in the recent years. The inverse solution is then enforced to meet a sparsity requirement, such that it can be represented with a few coefficients in some vector space: known as dictionary. In acoustics, such dictionaries are typically composed of elementary functions that faithfully resemble vibro-acoustic fields: e.g. plane waves for the vibrations of plates [12], or equivalent sources for multipole-like sound radiation [13]. Further studies have also considered a combination of dictionaries and solution approaches. For example, Fernandez-Grande and Daudet have used block-sparse representations of the sound field and its derivatives in order to reconstruct both compact and spatially extended sources [14]. In addition, the concept of sparse regularization lies in close connection to the problem of structural (geometric) separation of signals [15]. For instance, in cosmological data analysis, galaxies are sparsely structured into point-like singularities and filaments [16]; and in material sciences, microscopic images of composites tend to consist of granular and fibrous structures [17]. Consequently, enforcing sparsity on the different classes of signals facilitates the separation of such signals.

On the basis of these works and ideas, we propose a method to separate the contribution of the rail and the wheel to pass-by noise by means of sparse regularization methods. In essence, the noise contributions are found by projecting the measured pressure field onto different dictionaries: plane waves and moving equivalent sources. To the best of our knowledge, such an application has not been explored yet. The rail contribution is enforced to have a sparse representation in the wavenumber–frequency domain, that is, in a dictionary of plane waves with given incidence angles. And the wheel contribution is enforced to have a sparse representation into equivalent sources that move at the speed of the train. Such dictionary of moving sources can be considered an extension of the sparse equivalent source method [13].

The structure of the paper is as follows. Section 2 introduces a sparse model for the rail and the wheel noise contributions, including the dictionary representations—plane waves and moving equivalent sources, followed by an analysis of the sensing coherence between them. The separation method is then introduced in Section 3, describing the separation process step by step. A numerical assessment is included in Section 4, demonstrating the robustness of the method applied to synthetic data. Section 5 presents experimental validations with pass-by recordings of trains running at 40, 80, and 160 km/h, comparing the separated results with reference predictions by TWINS as well as with estimations via the WSE method; and including an analysis of multipolar sound radiation of the wheels. Section 6 entails a discussion regarding state-of-the-art separation methods. Conclusions are presented in Section 7.

2. A sparse model for rail and wheel noise

The motivation to exploit sparsity in the context of rolling noise comes from the existing knowledge about the sound directivity of the track and the wheels [18]. The dispersion of a railway track consists of a finite number of bending wave families that spread across frequencies [19]. In fact, several studies suggest that the rail sound directivity can be approximated with such wave families, which radiate at an angle depending on the ratio of the acoustic and structural wavelengths [3,5,7]. These arguments set grounds for choosing a finite number of plane waves—or wavenumbers—to sparsely represent the rail contribution.

Similarly, studies on the sound directivity of the wheel [18,20] reveal that it radiates nearly equally in all directions—like a monopole—when it is subject to radial motion, whereas it radiates efficiently (inefficiently) in the direction parallel (orthogonal) to the wheel axis—like a dipole—when it is subject to axial motion. In addition, the spatial distribution of wheels is intrinsically sparse throughout the length of the train, typically consisting of two sets per bogie (i.e. per 26 m).
Using these arguments it seems reasonable to represent the sound field radiated by the moving wheels with a sparse set of moving equivalent sources. (Fig. 1).

2.1. Dictionary representations

The following definitions hold throughout the remainder of the paper: \((x_m, y_m, z_m)\) is the position of the \(m\)-th microphone of the array, \((\hat{x}_n(t), \hat{y}_n, \hat{z}_n)\) is the position of the \(n\)-th moving equivalent source, \(t\) is a given time instant, \(\omega\) is a given angular frequency in rad/s, \(c\) and \(V\) are respectively the speed of sound in air and the speed of the train (both assumed constant in time), \(j\) is the imaginary unit, and \(\| \cdot \|_p = (\sum | |^p)^{1/p}\) is the \(\ell_p\)-norm for \(0 < p < \infty\).

2.1.1. Band-limited plane waves

Consider a set of \(L\) plane waves at frequency \(\omega\), whose complex amplitudes at time \(t\) are collected in the column vector \(\alpha = \alpha(\omega, t) = [\alpha_1, \ldots, \alpha_L]^T \in \mathbb{C}^L\). Then, the sound pressure at the \(m\)-th microphone position due to the plane waves is defined as

\[
\begin{align*}
    r_m & = \sum_{l=1}^{L} \psi_{lm} \alpha_l, \quad \text{with} \quad \psi_{lm} = e^{i(\omega t - k_l x_m)},
\end{align*}
\]

where \(k_l\) is the spatial frequency associated with the \(l\)-th plane wave. As the strongest contribution of rail noise is band-limited to a portion of the propagating waves in the wavenumber domain [7], we band-limit the spatial frequencies to the parabolic region \(|k| \leq k_{mw} = 50\sqrt{\omega/c}\) rad/m. This region can be physically understood as a set of bending wavenumbers \(k\) whose phase velocities resemble the parabolic-like dispersion of the waves propagating on the track \(k^2 \propto \omega^2\) [21]. The need for band-limitedness will later be supported with the coherence analysis in Section 2.2, as well as with numerical results in Section 4.

The pressure field \(r = r(\omega, t) \in \mathbb{C}^M\) at the \(M\) microphone positions due to the \(L\) plane waves can be written in matrix form as

\[
\begin{align*}
    r & = \Psi \alpha,
\end{align*}
\]

where

\[
\begin{align*}
    \Psi & = \Psi(\omega, t) = \begin{bmatrix}
         \psi_{11} & \ldots & \psi_{1L} \\
         \vdots & \ddots & \vdots \\
         \psi_{1M} & \ldots & \psi_{LM}
    \end{bmatrix} \in \mathbb{C}^{M \times L}
\end{align*}
\]

is the plane wave dictionary. This dictionary executes a similar operation to that of an inverse spatial Fourier transform, synthesizing back the pressure from a collection of spatial frequency components with harmonic time-dependence \(e^{i\omega t}\).
2.1.2. Moving equivalent sources

Consider now a set of $N$ monopoles at frequency $\omega$, moving parallel to the $x$-axis with constant subsonic speed $V < c$, as shown in Fig. 2. The source strengths at time $t$ are collected in the vector $\mathbf{\beta} \equiv \mathbf{\beta}(\omega, t) = [\beta_1, \ldots, \beta_N]^T \in \mathbb{C}^N$.

Then, the sound pressure at the $m$-th microphone position due to the moving sources including Doppler effect is defined as

$$w_m = \sum_{n=1}^{N} \phi_{nm} \beta_n, \text{ with } \phi_{nm} = \frac{\exp(\text{j} \omega \tau_{nm})}{4\pi |1 - \mathcal{M} \cos \theta_{nm}|^2} \left[ \frac{\text{j} \omega}{d_{nm}} + V (\cos \theta_{nm} - \mathcal{M}) \right],$$

where the two terms of $\phi_{nm}$ correspond to the far- and near-field components of the sound pressure [22]. Here $\tau_{nm}$ denotes the time elapsed since the sound field is emitted from the $n$-th source until it is recorded at the $m$-th microphone position. This emission time is earlier than the measurement time $t$, and its value is found from the solution to the (quadratic) equation [23]

$$c(t - \tau_{nm}) = d_{nm}$$

provided

$$d_{nm}(t) = \sqrt{(x_m - \bar{x}_n(t))^2 + (y_m - \bar{y}_n)^2 + (z_m - \bar{z}_n)^2}. \quad (6)$$

$\bar{x}_n(t) = d_0 + V \tau_{nm}$ is the position of the $n$-th source at instant $t$, the Mach number $\mathcal{M} = V/c$, and

$$\cos \theta_{nm}(t) = \frac{x_m - \bar{x}_n(t)}{d_{nm}(t)}. \quad (7)$$

The pressure field $\mathbf{w} \equiv \mathbf{w}(\omega, t) \in \mathbb{C}^M$ at the $M$ microphone positions due to the $N$ moving equivalent sources can be written in matrix form as

$$\mathbf{w} = \mathbf{\Phi} \mathbf{\beta}. \quad (8)$$

where

$$\mathbf{\Phi}(\omega, t) = \begin{bmatrix} \phi_{11} & \cdots & \phi_{1n} \\ \vdots & \ddots & \vdots \\ \phi_{Nm} & \cdots & \phi_{NN} \end{bmatrix} \in \mathbb{C}^{M \times N} \quad (9)$$

is the moving equivalent source dictionary. This dictionary executes a similar operation to the equivalent source dictionary [13], with the difference that the effect of source motion is taken into account.

The transfer functions $\phi_{nm}$ in Eq. (4) are defined in free space. This means that effects such as ground/ballast reflection and scattering due to the vehicle body are not accounted for, which might cause discrepancies in the wheel estimates. In principle it is possible to measure or model other kinds of transfer functions and account for these effects. However, the numerical estimation of the phased sound field is not trivial in the presence of complex boundaries and/or scatterers. The transfer function would require precise knowledge of phase shifts between transducers (as in e.g. matched field processing), which in the presence of complex boundaries of heterogeneous or unknown properties, is out of reach. Therefore, it is sensible to utilize only the direct field, knowing that the reflections will have less amplitude/carry less energy, leading to robust estimations—this is for example the rationale typically used in sound field control in rooms, and similar problems, where phased propagation is not trivial to estimate.
Fig. 3. Absolute value of the blocks of the Gram matrix formed between dictionaries [see Eq. (11)], at 500 Hz and \( t = 0.5 \) seconds. (a) Block formed between plane waves \( \psi^i\psi \). (b) Block formed between moving sources and plane waves \( \psi^i\Phi \). (c) Block formed between plane waves and moving sources \( \Phi^i\psi \). (d) Block formed between moving sources \( \Phi^i\Phi \).

2.2. Sensing coherence

Ideally we want a dense representation of the rail noise into moving sources (i.e. a non-sparse representation, requiring a large number of coefficients), as much as a dense representation of the wheel noise into plane waves. In this way the chances of successful separation increase. As the coherence is an inner product between the columns of both dictionaries concatenated \( \mathbf{D} = [\Psi \Phi] \in \mathbb{C}^{M \times (L+N)} \), the linear dependency between such columns must be as close to zero as possible. For this one can calculate the coherence metric \([24]\)

\[
\mu(\mathbf{D}) = \max_{j \neq i} \left( \frac{|\langle \mathbf{d}_i, \mathbf{d}_j \rangle|}{\|\mathbf{d}_i\|_2 \|\mathbf{d}_j\|_2} \right),
\]

at fixed time and frequency, where \( \langle \cdot, \cdot \rangle \) denotes the inner product, and \( \mathbf{d}_i \) is the \( i \)-th column of \( \mathbf{D} \). Note that the above coherence metric is a single number, while what one wants is that the inner product between all possible column combinations is as low as possible—to verify that there is maximal dissimilarity. One way to examine this situation is to look at the Gram matrix of the concatenated dictionary,

\[
\mathbf{G} = \mathbf{D}^H\mathbf{D} = \begin{bmatrix} \psi^i\psi & \psi^i\Phi & \Psi^H\Phi & \Phi^H\Phi \end{bmatrix} \in \mathbb{C}^{(L+N) \times (L+N)},
\]

which, when plotted, illustrates the linear dependency between all combinations of columns \( \{\mathbf{d}_i, \mathbf{d}_j\} \). The largest off-diagonal element of \( \mathbf{G} \) is then equivalent to \( \mu(\mathbf{D}) \).

Figs. 3 and 4 show examples of the blocks of \( \mathbf{G} \) at frequencies of 500 and 1250 Hz, respectively. As Eq. (11) suggests, these blocks result from the four possible combinations: (a) plane waves with plane waves, (b) moving sources with plane waves, (c) plane waves with moving sources, and (d) moving sources with moving sources. Because \( \mathbf{G} \) is symmetric, (b) and (c) are each other’s transpose. The plane wave and moving source dictionaries have \( L = N = 300 \) atoms. The microphone array is located such that \( x_1 = 0 \), and \( y_m = -9 \) cm and \( z_m = 1.2 \) m \( \mathrm{Vm} \). The array has \( M = 64 \) microphones spaced by \( s = 16 \) cm. The equivalent sources are located at \( y_n = 38.1 \) cm and \( z_n = 0 \) m \( \forall n \), uniformly distributed in a 26 m long segment with \( d_0 = 0 \) m. The time stamp corresponds to \( t = 0.5 \) s, and the sources move with speed \( V = 160 \) km/h.

The off-diagonal coefficients in the blocks (a) \( \psi^i\psi \) and (d) \( \Phi^i\Phi \) should be as close to zero as possible. Otherwise, the sparse regularization applied to the individual noise contributions may result in allocating weights to wrong coefficients. It
can be seen in Figs. 3(a) and 4(a) that the off-diagonal coefficients follow a rather fast decay for the wavenumbers considered. Such a decay increases with frequency as the sound field becomes more directional. In contrast to this, the off-diagonal elements in Figs. 3(d) and 4(d) exhibit a fast decay only within the distance spanned between 0 and about 10 m, which in fact corresponds to the array aperture 64 × 0.16 = 10.24 m. As a consequence, sources located outside the array aperture are highly coherent with each other, indicated by the non-zero off-diagonal elements in Figs. 3(d) and 4(d). Nevertheless, as it will later be shown in Section 5.1, the source positions appear to be correctly estimated even when they lie outside the aperture.

Similarly, the coefficients in the blocks (b) $\Psi^{ij}\Phi$ and (c) $\Phi^{ij}\Psi$ should ideally be as close to zero as possible. Otherwise, the sparse regularization applied to the combined noise contributions may result in wrongly allocating weights to plane waves which should be allocated to moving sources, and vice versa. As it can be seen in Figs. 3(b)–(c) and 4(b)–(c), there is little linear dependency between plane waves and moving sources—the coherence is no more than 0.7 at the largest wavenumbers; see e.g. the coherence values at $|k| \approx 13$ rad/m in Fig. 4(c). The coefficients in the blocks $\Psi^{ij}\Phi$ and $\Phi^{ij}\Psi$ can then be regulated by means of band-limiting the plane wave dictionary.

2.2.1. Coherence and band-limitedness of plane waves

This relationship can be understood in the wavenumber (spatial Fourier) domain. Propagating plane waves are defined with wavenumbers $|k| \leq \omega/c$, which corresponds to real incidence angles $\theta \in [-\pi/2, \pi/2]$. What happens is that sources moving outside the array aperture, $x_n(t) \in \mathbb{R} \setminus [0, s(M - 1)]$, radiate waves incident with grazing angle $|\theta| = \pi/2$, i.e. nearly parallel to the array. This results in an increased coherence between (i) plane waves propagating with grazing incidence angle and (ii) moving sources outside the array aperture. In other words, plane waves and moving sources are increasingly linearly dependent and can be mistaken by one another. An increased coherence can lead to an over-estimation of the plane waves—the pressure due to the moving sources is also collected in the plane wave dictionary; potentially resulting in an under-estimation of the moving source amplitudes.

The main purpose of band-limiting the plane wave dictionary is then to remove wavenumbers corresponding to grazing incidence angles, thus minimizing the coherence with moving sources that are outside the array aperture. As it will later be shown in Section 4, band-limitedness establishes a compromise between identifying the plane wave families present in the sound field, while keeping the coherence—and separation errors—small.
3. Separation method

Consider the pressure data measured with the microphone array at fixed frequency $\omega$ and time $t$, denoted with $p \equiv p(\omega, t) \in \mathbb{C}^M$. This is simply the result of applying spectral density estimation to the time-domain microphone signals. Assume that (i) the microphone array is close and parallel to the track, and (ii) the train speed is below 300 km/h, such that rolling noise is the dominant source of noise during the pass-by. The measurement model is written as

$$ p = r + w + z = [\Psi \Phi] [\alpha \beta] + z, \quad (12) $$

where the entries of $z \in \mathbb{C}^M$ account for modelling errors and/or contributions of other sources of noise: e.g. sleepers, ground scattering, etc. The above linear system of equations becomes under-determined when the number of unknowns $L + N$ exceeds the number of observations $M$.

3.1. Step 1: Rail separation

This first step involves solving the non-convex optimization problem

$$ \min_{\alpha} \frac{1}{2} \| \Psi \alpha - p \|_2^2 \quad \text{subject to} \quad \| \alpha \|_0 \leq K \quad (13) $$

individually for every frequency $\omega_b$, $b = 1, \ldots, B$, in the 1/3 octave band. Here the positive integer $K < M$, and the $\ell_0$ pseudonorm $\| \alpha \|_0 = (j | \alpha_j \neq 0)$ counts the number of non-zero wavenumber atoms. To solve this optimization problem one can use for instance the orthogonal matching pursuit (OMP) [25], which seeks for a solution $\alpha$ that is K-sparse. Then, the rail contribution to the array pressure at the $b$-th frequency $r(\omega_b)$ follows from Eq. (2). In contrast to the linear filtering of the wave signature extraction method [7], the proposed separation strategy for the rail contribution entails a nonlinear filtering of the measured data, in which the Fourier coefficients of the filter correspond exactly to the K sparse plane wave atoms. As a consequence, the Doppler effect can be accounted for since the OMP program finds the strongest plane waves per time instant, even when the wavenumbers are shifted.

3.2. Step 2: Residual calculation

Once the rail contribution at the array has been estimated in the whole frequency band, the residual with respect to the total pass-by noise is computed,

$$ \Delta(\omega_b) = p(\omega_b) - r(\omega_b), \quad (14) $$

for all frequencies $b = 1, \ldots, B$ in the 1/3 octave band. The residuals at all frequencies and microphone positions are then stacked into the column vector $\Delta^{(\text{all})} = [\Delta(\omega_1), \ldots, \Delta(\omega_B)]^T \in \mathbb{C}^{MB}$. 

3.3. Step 3: Wheel separation

Intuitively, the wheels should be localized at the same position regardless of the frequency. This means that the array pressure should be row-sparse at multiple frequencies; that is, the same spatial sparsity profile of moving sources is shared across frequencies. A multiple measurement vector approach [26,27] can be suitable for this task. One can then cast the problem under the framework of the least absolute shrinkage and selection operator (LASSO) as

$$ \min_{\beta^{(\text{all})}} \frac{1}{2} \| \Theta \beta^{(\text{all})} - \Delta^{(\text{all})} \|_2^2 \quad \text{subject to} \quad \| \beta^{(\text{all})} \|_1 \leq \epsilon \quad (15) $$

for a suitable regularization parameter $\epsilon > 0$ that can be related to the noise floor. A suitable choice can be found with a parameter selection method, e.g. the Pareto frontier curve [28]. Then, the corner of the logarithmic plot $\| \beta^{(\text{all})} \|_1$ versus $\| \Theta \beta^{(\text{all})} - \Delta^{(\text{all})} \|_2$ corresponds to the optimal regularization parameter. The block diagonal matrix

$$ \Theta = \begin{bmatrix}
\Phi(\omega_1) & 0 & \ldots & 0 \\
0 & \Phi(\omega_2) & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \Phi(\omega_B)
\end{bmatrix} \in \mathbb{C}^{MB \times NB} \quad (16) $$

stacks the dictionary matrices $\Phi(\omega_b)$ for all frequencies $b = 1, \ldots, B$ in the 1/3 octave band. Then, the wheel contribution at the array is obtained for all frequencies as

$$ w^{(\text{all})} = [w(\omega_1), \ldots, w(\omega_B)]^T = \Theta \beta^{(\text{all})}. \quad (17) $$
Fig. 5. Block diagram illustrating the signal processing chain of the proposed method. The processes inside the dashed box involve processing the pressure field individually per each sub-frequency \( b = 1, \ldots, B \) in the 1/3 octave band.

3.4. Sound pressure level spectra

The sound pressure level (SPL) spectra per 1/3 octave frequency band centered at \( f_c \), at the \( m \)-th microphone in the array, are obtained as

\[
\text{SPL}_{\text{rail}}(f_c) = 20 \log_{10} \left( \frac{\sqrt{\frac{\delta_B}{\pi B} \sum_{b=1}^{B} |f_m(\omega_b)|^2}}{p_{\text{ref}}} \right) + A(f_c)
\]

for the rail contribution; and similarly for the wheel contribution as

\[
\text{SPL}_{\text{wheel}}(f_c) = 20 \log_{10} \left( \frac{\sqrt{\frac{\delta_B}{\pi B} \sum_{b=1}^{B} |w_m(\omega_b)|^2}}{p_{\text{ref}}} \right) + A(f_c),
\]

where \( \delta_b \) is the 1/3 octave bandwidth per unit Hz, \( p_{\text{ref}} = 20 \mu\text{Pa} \), and \( A(f_c) \) is the A-weight curve coefficient at the center frequency \( f_c \).

3.5. Processing chain

Fig. 5 shows a block diagram of the signal processing chain of the proposed approach. The input consists of the raw time-series data measured with the microphones, which is first filtered into 1/3 octave bands. Thereafter, an iterative time-domain process is executed, beginning with spectral density estimation via the Welch’s method [29]. The array calibration data is taken from [7], and it is used to account for the sound scattering introduced by the frame and supports of the microphone array. Up to this point there is no difference in the processing with respect to the wave signature extraction method [7].

Next, Step 1 runs the OMP individually per sub-frequency in the 1/3 octave band, leading to an estimate of the rail contribution at each sub-frequency \( r(\omega_b) \), as described in Section 3.1. Once the field is processed at all sub-frequencies \( \omega_b, \ldots, \omega_B \), the residuals are calculated and stacked altogether in \( \Delta^{(\text{all})} \), as described in Step 2 in Section 3.2. Step 3 involves executing the LASSO optimization routine for the whole frequency band, leading to the estimate of the wheel contribution at all frequencies \( w^{(\text{all})} \). The algorithm is repeated until the last block of time data is processed, and the calculation of the rail and wheel contributions entails averaging the SPL values obtained per block of time data.

4. Separation of synthetic data

The purpose of this section is to assess the performance of the method using synthetic data, given various choices of dictionary atoms \( L \) and \( N \), as well as varying bandwidth of the plane wave dictionary \( k_{\text{BW}} \). The time-harmonic pressure is synthesized at narrow-band frequency \( f \) as the sound generated by 6 plane waves and 4 moving sources: using Eqs. (2) and (8) with uniformly random plane wave amplitudes \( \alpha \) and moving source strengths \( \beta \). The six plane waves consist of three families with wavenumbers \( [\pm 12 \sqrt{\omega/c}, \pm 33.5 \sqrt{\omega/c}, \pm 44 \sqrt{\omega/c}] \) rad/m; all within the bandwidth of the plane wave dictionary \( 50 \sqrt{\omega/c} \) rad/m. The four moving sources are initially located at \( [-1, -3.6, -8.4, -11] \) m, and they move with speed \( V = 160 \) km/h. The \( N \) moving sources are distributed over a length of 12 m, with \( \delta_0 = 0 \) m, spanning the locations of the 4 moving sources. Gaussian noise is added to the total field, with signal-to-noise ratio \( \text{SNR} = 15 \) dB. The pressure is synthesized at \( M = 42 \) positions, spaced by 8 cm, uniformly distributed at \( x_m \in [0, 3.36] \) m. The time series at each array position is 7.8 ms long, sampled at 6.4 kHz.
Fig. 6. Performance of the separation method using synthetic data, for varying (a)-(c) number of dictionary atoms L and N, and (d) bandwidth of the plane wave dictionary $k_{bw}$. The synthetic pressure consists of the sound field due to 6 plane waves and 4 moving equivalent sources. The separation is performed for a hundred realizations of random plane wave amplitudes and moving source strengths, and the results are presented in terms of mean error $\mu$ and 95% confidence intervals $\pm 2\sigma$. Gaussian noise is added to the total field, with signal-to-noise ratio $\text{SNR} = 15$ dB. (a) Separation error of plane waves $\epsilon_{pw}$ at frequencies $\omega$ for $f = 500$ Hz, and $\times$ for $f = 1250$ Hz, versus no. plane wave atoms $L \in [10, 100]$, using N = 12 moving source atoms. (b) Separation error of moving sources $\epsilon_{mes}$ at frequencies $\omega$ for $f = 500$ Hz, and $\times$ for $f = 1250$ Hz, versus no. plane wave atoms $L \in [10, 100]$, using N = 12 moving source atoms. (c) Separation error of moving sources $\epsilon_{mes}$ at frequencies $\omega$ for $f = 500$ Hz, and $\times$ for $f = 1250$ Hz, versus no. moving source atoms $N \in [8, 80]$, using L = 50 plane wave atoms. (d) Separation error of plane waves $\epsilon_{pw}$ and moving sources $\epsilon_{mes}$, at $f = 1250$ Hz, versus bandwidth of plane wave dictionary $k_{bw} \in [30\alpha / \sqrt{c}, c / \alpha]$ rad/m. The bandwidth $k_{bw}$ in (d) is expressed as percentage of the acoustic wavenumber $k = \omega / c$.

4.1. Separation errors

To assess the separation performance we compute relative separation errors of the time series at the $m$-th array position

$$
\epsilon_{pw}(x_m, f) = 10 \log_{10} \frac{\|r_m(f)^{\text{(sep)}} - r_m(f)^{\text{(ref)}}\|_2^2}{\|r_m(f)^{\text{(ref)}}\|_2^2}
$$

(20)

for the plane waves, and, similarly,

$$
\epsilon_{mes}(x_m, f) = 10 \log_{10} \frac{\|w_m(f)^{\text{(sep)}} - w_m(f)^{\text{(ref)}}\|_2^2}{\|w_m(f)^{\text{(ref)}}\|_2^2}
$$

(21)

for the moving equivalent sources. Here the superscripts (sep) and (ref) denote separated and reference pressure. The errors are evaluated at the center of the array $m = 21$. For each changing parameter ($L, N$ or $k_{bw}$), the separation is executed for a hundred realizations of random plane wave amplitudes and moving source strengths; and the separation errors are presented in terms of mean errors with 95% confidence intervals. These error metrics can be understood as the negative of the SNR for each of the sound field components. That is, error values $\epsilon < -\text{SNR}$ correspond to a likely accurate separation; and, conversely, errors $\epsilon > -\text{SNR}$ correspond to a likely poor separation. Fig. 6 summarizes the results.

The sparsity of the plane wave solution, $K$ in Eq. (13), should if known be chosen as the number of plane waves present in the pressure field, i.e. $K = 6$. From looking at Fig. 6(a), the more plane wave atoms $L$ the lower the separation error of the
plane waves $\epsilon_{\text{pw}}$. This can be understood in the wavenumber domain. If the number of plane wave atoms $L$ is too small, the dictionary has too coarse a wavenumber resolution and will likely misplace the $K$ plane waves present in the pressure field. The denser the dictionary resolution, the more likely it will resolve the $K$ plane waves. After about $L = 50$ plane wave atoms, the separation error at both frequencies seems to converge: using more atoms does not significantly improve the separation. The separation error of the moving sources $\epsilon_{\text{mes}}$, shown in Fig. 6(b), appears to be relatively constant with respect to the number of plane wave atoms. The results in Fig. 6(a) and (b) suggest that as one chooses $L \geq 50$ plane wave atoms, both separation errors $\epsilon_{\text{pw}}$ and $\epsilon_{\text{mes}}$ become smaller.

In these simulations a total of 4 moving sources is considered, mimicking four wheels of a train passing by the array. Because of the sequential nature of the separation method (first plane waves, then moving sources), the choice of moving source atoms $N$ does not influence the separation error of plane waves $\epsilon_{\text{pw}}$. On the other hand, as shown in Fig. 6(c), the separation error of moving sources $\epsilon_{\text{mes}}$ does not significantly improve when the number of atoms $N$ is greater than three times the actual number of moving sources ($4N$). In fact, there seems to be a small increase in error when $N$ is too large, e.g. see $\epsilon_{\text{mes}}$ for $N \geq 30$ at 1250 Hz in Fig 6(c). Overall, these results suggest that as one chooses $N \geq 12$ moving source atoms, the separation error $\epsilon_{\text{mes}}$ becomes smaller.

As explained earlier in Section 2.2.1, the plane wave dictionary ought to be band-limited. Fig. 6(d) demonstrates how the band-limitedness of the plane waves plays an important role in the separation of synthetic data. The bandwidth of the plane wave dictionary $k_{\text{bw}}$ is varied in the set $\left[ 50 \sqrt{\omega}/c, \omega/c \right]$ rad/m, and expressed as percentage of the acoustic wavenumber $k = \omega/c$. The dictionaries have $L = 50$ plane wave atoms and $N = 12$ moving source atoms. Both errors increase the more the bandwidth approaches the acoustic wavenumber. In particular, the separation error of plane waves $\epsilon_{\text{pw}}$ is relatively constant when the bandwidth is below 65% of $k$; thereafter it increases nearly linearly with 0.3 dB per % of $k$. Similarly, the separation error of moving sources $\epsilon_{\text{mes}}$ increases rather slowly when the bandwidth is below 90% of $k$; thereafter it increases nearly linearly with 1.2 dB per % of $k$. These results suggest that the bandwidth $k_{\text{bw}}$ should be large enough to identify the plane waves present in the sound field. But not too large either, otherwise it increases the coherence between (i) plane waves at grazing incidence and (ii) moving sources outside the array aperture, and, as a consequence, a likely poorer separation performance.

4.2. Separated SPL at the array

This section includes results of the separated sound fields at the array for various parameters: number of dictionary atoms $L$ and $N$, and the bandwidth of the plane wave dictionary $k_{\text{bw}}$. Fig. 7 shows the separated and reference SPL at the array, for both plane waves and moving sources. In this case we examine one realization of random plane wave amplitudes and moving source strengths. Time is fixed at $t = 3.9$ ms and the harmonic frequency is $f = 1250$ Hz. The separated SPLs are compared with the reference SPL.

Fig. 7 (a) shows how the choice of $L$ influences the accuracy of separating the plane waves: the larger the choice of $L$, the closer the separated field is to the reference. For $L = \{10, 50, 100\}$ atoms, the separation errors at the array center are $\epsilon_{\text{pw}} = \{-12.7, -18.8, -19\}$ dB. In a similar manner, Fig. 7(b) suggests that a larger choice of $L$ makes the separated SPL at the array due to moving sources closer to the reference. In this case, the separation errors at the array center are $\epsilon_{\text{mes}} = \{-18.2, -17.4, -17\}$ dB for the three values of $L$. Too small a choice of $N$ can deteriorate the separation of moving sources, as shown in Fig. 7(c) for $N = 8$. In particular, the sound field due to moving sources using $N = 12$ and 30 atoms is rather similar, suggesting that a larger $N$ might not significantly improve the separation of moving sources. In this case, the separation errors at the array center are $\epsilon_{\text{mes}} = \{-8, -17.4, -17.2\}$ dB for the three values of $N$. The SPL curves in Fig. 7(b) and (c) indicate that the sound field is strongest toward the left hand side of the array, which is expected considering the position of the moving sources with respect to the array.

Furthermore, as Fig. 7(d) demonstrates, the choice of bandwidth $k_{\text{bw}}$ influences the separation of both sound fields. In this case the dictionaries have $L = 50$ and $N = 12$ atoms. The separated SPL due to plane waves is closer to the reference SPL when the plane waves are band-limited (+), compared with the separated SPL when they are not band-limited (○). The same happens with the separated SPL due to moving sources: it is closer to the reference SPL with band-limited plane waves (+○) than with non band-limited plane waves (-○-). The results from this Section follow the separation performance observed in Fig. 6.

5. Separation of experimental data

The experimental data used in this paper is taken from a measurement campaign executed during the H2020-636032- ROLL2RAIL project in 2016 (more details in [7,8]). Pass-bys at three different speeds are considered: 40, 80, and 160 km/h. These speeds correspond to Mach numbers $M = 0.03, 0.06$, and 0.13, when $c = 340$ m/s. Fig. 8 shows a photograph of the microphone array setup at the test site near Munich. The array consists of $M = 42$ microphones spaced by $s = 8$ cm, uniformly distributed on a line parallel to the track. The microphone spacing allows for aliasing-free sampling up to 2.1 kHz.

See Table 1 for a summary of the relevant parameters describing the train pass-bys, the data acquisition, and the post-processing. The dashed box in Fig. 9 shows the segment of the train that is used for the analyses. This segment is approximately 26 m long, corresponding to a bogie and two wheel-sets (i.e. four wheels).
Fig. 7. Separated and reference sound pressure level (SPL) at $t = 3.9$ ms and $f = 1250$ Hz, versus array position, for various atom choices $L$ and $N$, and bandwidth choices $k_{bw}$. (a) Separated SPL due to plane waves using (□) $L = 15$, (○) $L = 50$, and (∗) $L = 100$ atoms, compared with (solid line) reference SPL. (b) Separated SPL due to moving sources using (□) $L = 15$, (○) $L = 50$, and (∗) $L = 100$ atoms, compared with (solid line) reference SPL. (c) Separated SPL due to moving sources using □ $N = 8$, (○) $N = 12$, and (∗) $N = 30$ atoms; compared with (solid line) reference SPL. (d) Separated SPL due to plane waves (PW) using ( + ) $k_{bw} = 50\sqrt{\omega/c}$, and (○) $k_{bw} = \omega/c$, compared with (solid line) reference SPL; and separated SPL due to moving sources (MES) using (-□-) $k_{bw} = 50\sqrt{\omega/c}$, and (-○-) $k_{bw} = \omega/c$, compared with (dotted line) reference SPL.

Fig. 8. Photograph of the experimental setup at the test site near Munich in 2016, H2020-636032-ROLL2RAIL [8].
Based on the results from Section 4, the dictionaries have $L = 50$ and $N = 12$ atoms. The moving sources span a total length of 12 m, i.e. one source per meter. The offset $d_0$ is chosen according to the time analysis window of interest: in this case $d_0 = -9$ m guarantees that the $N$ moving source atoms span the region where the wheels are located in the train (see minima in the dashed box of Fig. 9). The spacing between moving sources results then from dividing the length of the equivalent source distribution (12 m) by the number of dictionary atoms $N$. Readings from one accelerometer in the vertical direction of the track can be used to estimate the passage of the wheels.

The sparsity of plane waves in the OMP program of Eq. (13) is set to $K = 8$ atoms. This choice is motivated with the number of bending wave families propagating in the track. In the frequency range up to 2 kHz there are typically up to four families that cut on: vertical bending, lateral bending, torsional waves, and web bending [19]. Each of these families propagate along the positive and negative directions of the track, hence $K = 2 \times 4$ wavenumbers.

The processing is performed in MATLAB 2019b. The plane wave amplitudes that optimize Eq. (13) are found with the toolbox CoSaMP and OMP for sparse recovery [30], whereas the moving source strengths that optimize Eq. (15) are found with the toolbox SPGL1 [28,31].

5.1. Time-domain separation results

This section includes results of the iterative time-domain processes in Fig. 5. The results are shown for the trains running at 80 and 160 km/h in Figs. 10 and 11, for the frequency bands centered at 500 Hz and 1.25 kHz, respectively. Each of these figures show plane wave amplitudes $\alpha$ and moving source strengths $\beta$ within the entire frequency band, at three different time instants.
Fig. 10. Separation results at three time instants (rows), for the 1/3 octave band $f_s = 500$ Hz, and a train running at $V = 80$ km/h. Left column: analysis time window showing the band-pass filtered acceleration (solid curve) and the microphone array aperture (vertical lines). The minima correspond to the passage of the wheels above the accelerometer. Middle column: absolute value of plane wave amplitudes $\alpha$ versus frequency and wavenumber. The arrows on the bottom of each plot correspond to the wavenumbers of the free Timoshenko beam (see Table 2). Right column: absolute value of equivalent source strengths $\mathcal{B}$ versus frequency and source position. The dotted vertical lines indicate the position of the wheels as estimated from the acceleration signals in (a), (d) and (g).

Table 2

| Frequency (Hz) | Vertical bending | Lateral bending |
|----------------|------------------|-----------------|
| 500 Hz         | ± 3.33 rad/m     | ± 5.00 rad/m    |
| 1250 Hz        | ± 5.82 rad/m     | ± 8.27 rad/m    |

It can be seen that the wavenumbers of the plane waves estimated with the OMP program closely resemble those of a free Timoshenko beam. These Timoshenko wavenumbers are included in Table 2 below. The resemblance is clear at 500 Hz, shown in Fig. 10(b), (e), and (h), in which two wave families can be seen, indicated with the arrows: one with wavenumber of approximately ± 3 rad/m and another of ± 5 rad/m.

It must be stressed that the rail contribution is stronger than that of the wheel at such a low frequency, and, as a consequence, the associated plane waves can be detected even when the array is in front of the wheels; see e.g. Fig. 10(e) and (h). The situation changes as frequency increases, because the wheel contribution to the pass-by noise increases. Inspection of Fig. 11 suggests that the wave families can be better detected when the array is far from the wheels. That is, the wave families centered at approximately ± 6 rad/m and ± 8 rad/m are more clearly visible in Fig. 11(b) than in Fig. 11(e) and (h).

Regarding the moving sources, Fig. 10(c), (f), and (i), as well as in Fig. 11(c), (f), and (i), demonstrate that the solution is row-sparse as intended: the sparsity profile is shared across the frequency band. This corroborates the intuition of using the multiple-measurement vector optimization described in Section 3.3. One key difference between Figs. 10 and 11 is the precision of localizing individual moving sources, which appears to be frequency-dependent. We suspect this is due to the difference in spherical spreading of the sound field. At lower frequencies, such as in Fig. 10, it results more challenging to distinguish moving sources from one another. In particular, the localization of the sources is more accurate when the array...
aperture is closer to the wheels: e.g. comparing Fig. 10(d) and (g) with Fig. 10(f) and (i). When the array is far from the wheels the sources are not accurately localized: as seen by comparing Fig. 10(a) with Fig. 10(c). This result is in agreement with the increase in coherence between moving sources that lie outside the array aperture, as described in Section 2.2.

On the other hand, moving sources at higher frequencies, such as in Fig. 11, are localized and distinguished more clearly. Surprisingly so even when the array aperture is far from the sources: see e.g. Fig. 11(c). Furthermore, the estimated locations of the sources agree reasonably well with the actual positions of the wheels (indicated with the dashed lines). Overall, the estimated moving sources seem to be in fairly good agreement with the moving wheels, justifying the proposed physical model.

5.2. Separated SPL spectra

The separated spectra are evaluated at the center-most microphone in the array m = M/2 + 1, and compared against reference predictions obtained with TWINS [8], as well as with the results obtained using the wave signature extraction (WSE) method [7]. Figs. 12–14 show the SPL spectra for the rail and the wheel, for train speeds of 40, 80, and 160 km/h, respectively. The wheel contribution is also inferred with the WSE method from the energetic difference between the total pass-by and the estimated rail contribution, shown in Figs. 12(b), 13(b) and 14(b).

As regards the rail contribution, it can be seen that the spectra estimated with the proposed approach agree well with the reference data for the three train speeds investigated. The differences are in the order of ± 2 dBA, comparable with the performance of the WSE method up to 1.25 kHz. As it was found in [7], the WSE method under-estimates the rail contribution above 1.25 kHz. In [7], the authors attributed such an under-estimation to the presence of higher-order bending waves that were unaccounted for in the WSE filters, and the letter [32] further verified this by means of modifying the filter shapes. The spectra in Figs. 12(a) and 13(a) indicate that the proposed approach slightly under-estimates the rail contribution in the higher frequency end; however, not as much as the WSE method does. Unlike the WSE method, the proposed method does not have a fixed wavenumber filter shape, but, rather, it finds the strongest active wavenumbers at every time instant.
Fig. 12. Sound pressure level (SPL) spectra of (a) rail contribution and (b) wheel contribution in dBA, for a train running at 40 km/h. Solid lines: reference spectra predicted by TWINS. Solid/dotted lines and triangles: estimated spectra via the WSE method [7]. Solid lines and circles: estimated spectra via the proposed method. The wheel contribution estimated with the WSE method in (b), marked with \((\Delta E)\) in the legend, is inferred from the energetic difference between the total pass-by and the estimated rail contribution.

Fig. 13. Sound pressure level (SPL) spectra of (a) rail contribution and (b) wheel contribution in dBA, for a train running at 80 km/h. Solid lines: reference spectra predicted by TWINS. Solid/dotted lines and triangles: estimated spectra via the WSE method [7]. Solid lines and circles: estimated spectra via the proposed method. The wheel contribution estimated with the WSE method in (b), marked with \((\Delta E)\) in the legend, is inferred from the energetic difference between the total pass-by and the estimated rail contribution.

Fig. 14(a) shows that the rail contribution estimated by the proposed approach closely follows the reference prediction—and slightly over-estimates it at 2 kHz.

As regards the wheel contribution, the spectra separated with the proposed method agree reasonably well for the three train speeds, most noticeably at 800 Hz and 1.25 kHz. Figs. 12(b), 13(b), and 14(b) show that the wheel is over-estimated by the proposed approach at 1 kHz by about 5 dBA; considerably less, however, than the over-estimation resulting from the energetic difference estimate via the WSE method. For frequencies above 1.25 kHz, the proposed approach under-estimates the wheel contribution, particularly for the train running at 40 km/h shown in Fig. 12(b). The under-estimation at frequencies above 1.6 kHz seems to be the result of having too few equivalent sources to represent the sound radiation of the wheels. If the number of equivalent sources \(N\) is increased with frequency, the estimated wheel spectra approaches the ref-
ference spectra more closely. Another reason one could consider is the multipolar sound radiation of the wheels, which is investigated in more detail in Section 5.3 below.

On the other hand, Figs. 13(b) and 14(b) demonstrate excellent agreement at 630 and 800 Hz between the proposed approach and the reference prediction, with differences below 1 dBA. At 500 Hz, however, the wheel contribution is underestimated in the three cases, and by as much as 15 dBA in Fig. 12(b) for the train running at 40 km/h. As discussed in Section 5.1, such an under-estimation is at present attributed to the difference in spherical spreading of the sound field, which complicates the task of distinguishing and localizing individual sources in motion—particularly when the array is far from the wheels, and its aperture is small compared with the wavelength. Another possible explanation to the under-estimation is the effect of ground reflection, which is not accounted for.

Lastly, it is worth mentioning that the proposed approach separates the wheel contribution fairly accurately even when its spectrum is far below the rail spectrum—compare ordinate axes between e.g. Fig. 13(a) and (b) at 630 Hz. This strongly suggests that the use of sparse regularization methods may lead to accurate separation even in cases of poor signal-to-noise ratio.

5.3. Multipolar wheel sound radiation

One may wonder whether simple moving monopoles are sufficient to model the radiation of wheels, specially at higher frequencies since more complex modal patterns arise. Naturally, more complex modal patterns would result in a more complex sound field, of potentially multipolar nature. In this section we address this by appending a dictionary of moving dipoles to represent the wheel contribution. In mathematical terms, define the pressure field at the $m$-th microphone due to $N$ moving dipoles as [22]

$$
g_m = \sum_{n=1}^{N} u_{nm} \gamma_n, \quad \text{with} \quad u_{nm} = \frac{(z_m - \bar{z}_n)}{4\pi |1 - m \cos \theta_{nm}|^2} \left[ \frac{j \omega}{c d_{mn}^2} + \frac{1}{|1 - m \cos \theta_{nm}| d_{mn}^2} \right],$$

where, similarly to Eq. (4), the two terms of $u_{nm}$ correspond to the far- and near-field components of the sound pressure. The pressure contribution of the $N$ moving dipoles at the $M$ microphones is written in matrix form as

$$
g = \Upsilon \gamma,$$

where

$$
\Upsilon \equiv \Upsilon(\omega, t) = \begin{bmatrix} u_{11} & \cdots & u_{N1} \\
\vdots & \ddots & \vdots \\
u_{1M} & \cdots & u_{NM} \end{bmatrix} \in \mathbb{C}^{M \times N}
$$

is the moving dipole dictionary. Appending the sound field due to the moving dipoles means the pressure field at fixed frequency $\omega$ and time $t$, recorded at the microphone array, is redefined as

---

**Fig. 14.** Sound pressure level (SPL) spectra of (a) rail contribution and (b) wheel contribution in dBA, for a train running at 160 km/h. Solid lines: reference spectra predicted by TWINS. Solid/dotted lines and triangles: estimated spectra via the WSE method [7]. Solid lines and circles: estimated spectra via the proposed method. The wheel contribution estimated with the WSE method in (b), marked with ($\Delta E$) in the legend, is inferred from the energetic difference between the total pass-by and the estimated rail contribution.
\[ \hat{w} \equiv \hat{w}(\omega, t) = w + g = \left[ \begin{array}{c} \Phi \\ \Upsilon \end{array} \right] \left[ \begin{array}{c} \beta \\ \gamma \end{array} \right]. \]  

(25)

Then, the matrix \( \Theta \) defined in Eq. (16) is redefined as \( \hat{\Theta} \equiv \hat{\Theta}(\omega, t) \in \mathbb{C}^{MB \times (N + \hat{R})B} \), the concatenation of monopole and dipole dictionaries,

\[ \hat{\Theta} = \left[ \begin{array}{cccccccc} \Phi(\omega_1) & \cdots & 0 & \Upsilon(\omega_1) & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \Phi(\omega_B) & 0 & \cdots & \Upsilon(\omega_B) \end{array} \right]. \]

(26)

and the solution vector to the LASSO program (15) is modified accordingly,

\[ \hat{\beta}^{(all)} = \left[ \begin{array}{c} \beta^{(all)} \\ \gamma^{(all)} \end{array} \right] \in \mathbb{C}^{N + \hat{R}}. \]

(27)

Lastly, the wheel contribution to the sound pressure at the array is obtained for all frequencies as \( \hat{w}^{(all)} = \hat{\Theta}\hat{\beta}^{(all)} \).

In what follows we set \( \hat{N} = N \), and the problem parameters presented in Table 1 remain unchanged. Table 3 lists the SPL spectra of the wheel using only moving monopoles, using both moving monopoles and moving dipoles, and the reference spectra predicted by TWINS.

The most noticeable differences between the two cases—using only monopoles versus using both monopoles and dipoles—appear at 500 Hz and 2 kHz, with an effective increase of the estimated wheel contribution by 0.5 dBA. Such an increase indeed makes the estimated wheel contribution closer to the reference predictions. In the remainder of the frequency range, however, the differences are rather negligible, in the order of 0.1 dBA. Similar results are obtained for the two other train speeds of 40 and 160 km/h, not reported here for the sake of brevity. From these findings, we suspect that using higher-order multipole expansions will not result in considerable differences.

6. Discussion

It is straightforward to compare the proposed method with the wave signature extraction (WSE) method [7]. On the one hand, the method proposed here estimates the bending wavenumbers during the pass-by with the matching pursuit optimization program, whereas such estimates of the WSE method rely on sufficiently long recordings of the sound field before and after the pass-by. Thus, the proposed approach circumvents altogether the dependency of available recording time (before and after the pass-by) on the rail contribution estimation, a crucial limitation of the WSE method for measurements on tracks with higher decay rates. Yet another disadvantage of the WSE method is its inability to estimate the wheel contribution: it can only provide an estimate of the energetic difference between the total pass-by noise and the rail noise. Such an estimate most often leads to an over-estimation of the wheel noise, as shown in Figs. 12–14. In contrast to this, the method proposed here performs a (sparse) linear regression to the residual between the total pass-by noise and the estimated rail contribution, using a physical model of the moving wheels into moving equivalent sources.

A few words can be said about other state-of-the-art methods. According to the report by Thompson et al. [8], only two methods have been proposed to identify the wheel component: beamforming and pass-by analysis (PBA). Beamforming consistently over-estimates the wheel contribution, with differences of up to 20 dBA at 1 kHz (see Fig. 19(a) in [8]). The two variants of the PBA method studied in that paper either over-estimate the wheel contribution significantly at frequencies below 1.6 kHz, or produce inconsistent estimates with respect to the reference results (see Fig. 19(b) in [8]). One must however be careful in comparing the data obtained here with the data reported there, as the measurement distances are different. The application of the proposed method at other measurement distances should be performed in order to make a fair comparison. It is at any rate promising that the wheel contribution estimated with the proposed method is in fairly good agreement with the reference predictions by TWINS.

7. Conclusions

This paper introduces a method to separate the rail and wheel contributions to pass-by noise, by means of sparse regularization of microphone array data. The main contribution of this work is to enable the joint separation of the two components without requiring static tests prior to the pass-by, using a collection of microphones near the track. Because sparse
regularization methods are used to separate the two noise components, requirements regarding signal sparsity and sensing coherence must be met. Based on the existing knowledge about the sound radiation of the rail and the wheels, the sound field radiated by the rail is assumed to be sparse in a dictionary of band-limited plane waves; and the field radiated by the wheels is assumed to be sparse in a dictionary of equivalent sources that move with the speed of the train. An analysis of the sensing coherence between the two dictionaries is included, showing that band-limited plane waves and moving sources are incoherent within the array aperture. A numerical assessment is also included, showing the performance of the proposed method to varying dictionary sizes; as well as the bandwidth of the plane wave dictionary, demonstrating the need for band-limited plane waves. Lastly, the proposed method is experimentally tested with measurements taken during the European project H2020-636032-ROLL2RAIL. Overall, the separated spectra of the rail and the wheel are in good agreement with the reference data predicted by the TWINS software.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement
Elias Zea: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. Efren Fernandez-Grande: Methodology, Validation, Formal analysis, Writing - review & editing. Ines Lopez Arteaga: Methodology, Formal analysis, Writing - review & editing, Project administration.

Acknowledgments
The authors would like to acknowledge the anonymous reviewers for their valuable comments and suggestions. This work is financially supported by the Swedish Research Council (Vetenskapsrådet), grant agreement No. 2016-04366, and the H2020 ROLL2RAIL project, grant agreement No. 636032. Fernandez-Grande is partly supported by the VILLUM Foundation, grant no. 19179. Acknowledgments are given to Giacomo Squicciarini and David Thompson for the reference data obtained with the TWINS software. Thanks to Nicolas Pignier for sharing the code to estimate emissions times of a moving monopole, and to Luca Manzari and Leping Feng for the acquisition of the measurement data. Thanks to Anton Shipsha for the discussions on microscopic imaging of composites.

References
[1] Commission Regulation (EU) No 1304/2014 of 26 November 2014 on the technical specification for interoperability relating to the subsystem ‘rolling stock noise’ amending Decision 2008/232/EC and repealing Decision 2011/229/EU Text with EEA relevance, 2014.
[2] T. Kitagawa, D.J. Thompson, Comparison of wheel/rail noise radiation on Japanese railways using the TWINS model and microphone array measurements, J. Sound Vib. 293 (3–5) (2006) 496–509. doi:10.1016/j.jsv.2005.08.037.
[3] T. Kitagawa, D.J. Thompson, The horizontal directivity of noise radiated by a rail and implications for the use of microphone arrays, J. Sound Vib. 329 (2) (2010) 202–220. doi:10.1016/j.jsv.2009.09.002.
[4] F. Le Courtois, J.-H. Thomas, F. Poisson, J.-C. Pascal, Identification of the rail radiation using beamforming and a 2 D array, in: Acoustics, 2012, pp. 3745–3750, Nantes, France.
[5] J. Zhang, G. Squicciarini, D.J. Thompson, Implications of the directivity of railway noise sources for their quantification using conventional beamforming, J. Sound Vib. 459 (2019) 114841, doi:10.1016/j.jsv.2019.07.007.
[6] B. Faure, O. Chiello, M.-A. Pallas, C. Sirvĕre, Characterisation of the acoustic field radiated by a rail with a microphone array: the SWEAM method, J. Sound Vib. 346 (2015) 165–190, doi:10.1016/j.jsv.2015.02.022.
[7] E. Zea, L. Manzari, G. Squicciarini, L. Feng, D. Thompson, I. Lopez Arteaga, Wavenumber–domain separation of rail contribution to pass-by noise, J Sound Vib 490 (2017) 24–42, doi:10.1016/j.jsv.2017.07.040.
[8] D. Thompson, G. Squicciarini, J. Zhang, I. Lopez Arteaga, E. Zea, M. Dittrich, E. Jansen, K. Arcas, E. Cierco, F. Magrans, A. Malkoun, E. Iturritxa, A. Guiral, M. Stangl, G. Schleinzer, B. Martin Lopez, C. Chaufour, J. Wändell. Assessment of measurement-based methods for separating wheel and track contributions to railway rolling noise, Appl. Acoust. 140 (2018), doi:10.1016/j.apacoust.2018.05.012.
[9] J.F. Claerbout, F. Muir, Robust modeling with erratic data, Geophysics 38 (5) (1973) 826–844.
[10] E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun Pure Appl Math 59 (8) (2006) 1207–1223, doi:10.1002/cpa.20124.
[11] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306 doi:10.1109/TIT.2006.871582.
[12] G. Chardon, L. Daudet, A. Peillot, F. Ollivier, N. Bertin, R. Gribonval, Near-field acoustic holography using sparse regularization and compressive sampling principles., J. Acoust. Soc. Am. 122 (3) (2012) 1521–1534. doi:10.1121/1.4740476.
[13] E. Fernandez-Grande, A. Xenaki, P. Gerstoft, A sparse equivalent source method for near-field acoustic holography, J. Acoust. Soc. Am. 141 (1) (2017) 532–542, doi:10.1121/1.4974047.
[14] E. Fernandez-Grande, L. Daudet, Compressive acoustic holography with block-sparse regularization, J. Acoust. Soc. Am. 143 (6) (2018) 3737–3746, doi:10.1121/1.5042412.
[15] E. Fernandez-Grande, L. Daudet, Compressive acoustic holography with block-sparse regularization, J. Acoust. Soc. Am. 143 (6) (2018) 3737–3746, doi:10.1121/1.5042412.
[16] J. Lopez Arteaga, I. Lopez Arteaga, J.-L. Starck, M. Elad, D. Donoho, Redundant multiscale transforms and their application for morphological component separation, Adv. Imag. Elect. Phys. 132 (2004) 287–348, doi:10.1002/SI01657-5607/04(32006-9).
[17] V. Alonso de la Peña, I.L. Darriba, M. Caserío Valea, F. Guitián Rivera, Mechanical properties related to the microstructure of seven different fiber reinforced composite posts, J. Adv. Prosthodont, 8 (6) (2016) 433–438.
[18] D. Thompson, Sound radiation from wheels and track, in: Railway Noise and Vibration, Elsevier, Oxford, UK, 2009, pp. 175–222, doi:10.1016/B978-0-08-045147-3.00006-2.
[19] D.J. Thompson, Experimental analysis of wave propagation in railway tracks, J. Sound Vib. 203 (5) (1997) 867–888, doi:10.1006/jsvi.1997.0903.
[20] D.J. Thompson, M.G. Dittrich, Wheel response and radiation – laboratory measurements of five types of wheel and comparisons with theory, Technical Report, ORE Technical Document DT249 (C163), Utrecht, The Netherlands, 1991.

[21] D. Thompson, Track Vibration, in: Railway Noise and Vibration, Elsevier, Oxford, UK, 2009, pp. 29–95, doi:10.1016/B978-0-08-045147-3.00003-7.

[22] P.M. Morse, K.U. Ingard, Theoretical Acoustics, McGraw-Hill, New York, USA, 1968, doi:10.1119/1.1976432.

[23] M. Åbom, The effect of source motion, in: An Introduction to Flow Acoustics, 4, KTH, Stockholm, Sweden, 2017, pp. 121–123.

[24] D.L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inf. Theory 47 (7) (2001) 2845–2862, doi:10.1109/18.959265.

[25] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, 1993, pp. 40–44, doi:10.1109/ACSSC.1993.342465, vol.1, Pacific Grove, USA.

[26] S.F. Cotter, B.D. Rao, K. Engan, K. Kreutz-Delgado, Sparse solutions to linear inverse problems with multiple measurement vectors, IEEE Trans. Signal Process. 53 (7) (2005) 2477–2488, doi:10.1109/TSP.2005.849172.

[27] Y.C. Eldar, M. Mishali, Robust recovery of signals from a structured union of subspaces, IEEE Trans. Inf. Theory 55 (11) (2009) 5302–5316, doi:10.1109/TIT.2009.2030471.

[28] E. van den Berg, M.P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2) (2008) 890–912, doi:10.1137/080714488.

[29] P. Welch, The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. 15 (2) (1967) 70–73, doi:10.1109/TAU.1967.1161901.

[30] S. Becker, CoSaMP and OMP for sparse recovery, 2020, MATLAB Central File Exchange.

[31] E. van den Berg, M.P. Friedlander, SPGL1: A solver for large-scale sparse reconstruction, 2019, http://www.cs.ubc.ca/labs/scl/spgl1

[32] E. Zea, I. Lopez Arteaga, On modified wavenumber filters for rail contribution estimations, J. Acoust. Soc. Am. 144 (4) (2018) EL286–EL289, doi:10.1121/1.5063453.