Linear Model and Regularization for Transient Wave–Based Pipeline-Condition Assessment

Xun Wang¹; Mohamed S. Ghidaoui, M.ASCE²; and Pedro J. Lee³

Abstract: Condition assessment or defect detection of a pipeline is a difficult inverse problem. This paper proposes a general linear model framework that can approximately describe a wide range of pipeline condition assessment and defect detection problems. More specifically, the system response is governed by a linear function of a pipe property at discrete locations along a pipe, such that the pipe property can be reconstructed via a least-squares fit to the measured response. Real pipe systems in general involve a large number of uncertain pipe characteristics, limited data, and a very high level of noise, such that the inverse problem is ill-posed. The well-known Tikhonov regularization scheme is employed on the linear model to provide a general solution for the ill-posed inverse problem. The optimal regularization parameter, which is crucial and problem-dependent such that no universal approach always generates satisfactory results, is decided via the generalized cross validation (GCV) and L-curve approaches. The proposed general linear model and inverse problem methodologies are illustrated via two application examples: time-domain impulse response function extraction using least-squares deconvolution and leakage detection based on a frequency-domain linearized model. In both examples, numerical and experimental results demonstrate the significance of the regularization parameter and the merits of the GCV and L-curve methods in the pipeline condition assessment problems. DOI: 10.1061/(ASCE)WR.1943-5452.0001205. © 2020 American Society of Civil Engineers.

Author keywords: Linear model; Least squares; Pipeline condition assessment; Leak detection; Tikhonov regularization; Generalized cross validation; L-curve.

Introduction

Urban water supply systems are the lifeline of three billion people in the world. Despite their importance, these vital systems are aging and fraught with deficiencies and inefficiencies, resulting in huge financial costs from the wastage of resources, paralysis of businesses, and devastating urban floods. Accurate, timely, and cost-effective approaches of pipe condition assessment and defect detection are critical for the long-term operation of these networks. Since most water supply systems are widely distributed and buried underground, external monitoring of pipes, such as by visual inspection, is very difficult. Alternatively, transient wave-based methods have been used that involve sending active waves, collecting data from field instruments, and analyzing the information mathematically or statistically to estimate defects in pipes (Colombo et al. 2009; Mpsha et al. 2001; Lee et al. 2005a; Duan et al. 2010, 2011; Meniconi et al. 2010). The transient waves can be generated either by a sudden change of flow, such as rapid closing or opening of a valve (Liggett and Chen 1994; Brunone 1999; Lee et al. 2006; Meniconi et al. 2013, 2015; Jing et al. 2018), or by a designed (sharp) input signal sent by a transducer or control valve (Liou 1998; Lee et al. 2008a; Nguyen et al. 2018). The propagating waves are partially reflected and attenuated when they encounter a defect (e.g., leakage or blockage), and these changes to the signal are used to determine the properties of the defects.

With these measurements of waves, condition assessment or defect detection of pipeline is carried out by estimating the key system parameters. This is usually a difficult inverse problem, partially because of the nonlinearity of a piping system with respect to these parameters. In many pipe problems, however, the pipe system can be approximately assumed to be linear such that solutions of the inverse problems can be largely simplified. A pipeline system can be treated as an approximate linear time-invariant (LTI) system if the magnitude of transient excitation is appropriate (Lee and Vítkovský 2010; Lee 2013; Gong et al. 2013). Then the time-domain impulse response function (IRF) can be represented in linear form Nguyen et al. (2018). As soon as a pipe system’s IRF is obtained, its anomalies, such as leakage, blockage, junction, or pipe wall thickness, can be estimated (Nguyen et al. 2018; Zeng et al. 2018). In a frequency-domain wave propagation model (Chaudhry 2014), the nonlinear convective terms are neglected because they are of the order of the Mach number, which is very small in practice (Ghidaoui 2004; Ghidaoui et al. 2005). Pipe friction, valves, pumps, and other hydraulic elements are also linearized, which is reasonable since the injected waves for pipe condition assessment are generally small to avoid compromising the structural integrity of the system. Furthermore, it is shown in Lee et al. (2005b) that the location of a leak determines the shape of the frequency response function (FRF), and its amplitude varies linearly with leak size. This is mathematically justified in Wang and Ghidaoui (2018b), where a linear expression of the FRF is explicitly given. It is further found in Wang and Ghidaoui (2018a) that the linear model in Wang and Ghidaoui (2018b) can be generalized to the case of multiple leaks if the effect of leak–leak interaction is neglected, which is reasonable in the case of small leaks. A similar linear behavior...
is also found in problems of extended blockage (Zouari et al. 2017) and discrete blockage (Lee et al. 2008b). With these linear model forms, pipeline condition assessment or defect detection solutions are more simplified (Wang and Ghaidawi 2018a, b; 2019; Wang et al. 2019c; Zouari et al. 2017; Nguyen et al. 2018; Zeng et al. 2018) than previous approaches (Lee et al. 2005b; Rubio Scola et al. 2016). This usually implies a more accurate and robust estimation of pipe properties because the model linearization significantly reduces the complexity of inverse problems.

In this paper, a general linear model framework is proposed that describes a wide range of pipeline condition assessment problems, including identification of leakage, (discrete and extended) blockage, branches, and pipe wall thickness (Lee et al. 2008b; Zouari et al. 2017; Meniconi et al. 2011, 2016; Wang and Ghaidawi 2018a, b; Zeng et al. 2018; Nguyen et al. 2018; Wang et al. 2019a, b). Based on this model, pipeline properties can be reconstructed via a least-squares (LS) solution. However, owing to the high level of noise, uncertainties with respect to the system parameters and limited data, inverse problems are often ill-posed, in the sense that a slight perturbation in measurement leads to a very large perturbation in the solution. The well-known Tikhonov regularization (Tikhonov 1963) scheme is employed to solve the aforementioned ill-posed inverse problems. Generalized cross validation (GCV) (Golub et al. 1979) and L-curve methods (Hansen 1992; Hansen and O’Leary 1993) are tested to determine the optimal regularization parameter, which is crucial and problem-dependent such that no universal method can produce satisfactory results for all regularization problems (Bauer and Lukas 2011). The proposed methodologies are illustrated with two application examples. The first example is time-domain IRF extraction using a LS deconvolution. The Tikhonov regularization solution for this problem is proposed in Nguyen et al. (2018), but determining the regularization parameter is not investigated, which is essential in this inverse problem. The second example is pipeline leakage detection based on a frequency-domain sparse linearized model (Wang and Ghaidawi 2018a; Zhou et al. 2018), where some discrete locations in a pipe are assumed to be potential leaks. The LS and Tikhonov regularization solutions for leakage detection based on this model have never been investigated. In both examples, numerical and experimental results illustrate the significance of the regularization parameter and the efficiency of the proposed methods.

This paper is an extension of a conference paper by the authors (Wang et al. 2018). In the next section, the general framework of the proposed linear model for transient wave–based pipe condition assessment is introduced, followed by a discussion of the LS solution and regularization methods. Then two application examples are introduced, and finally conclusions are drawn.

### Linear Model for Pipe Condition Assessment

This paper proposes a general configuration for transient wave–based pipe condition assessment problems. To generalize the approach, a spatially distributed pipe property is denoted by \( f_x(x), x \in [0, l] \), where \( x \) is the one-dimensional coordinate along the pipe from the upstream node \( x = 0 \) to the downstream node \( x = l \). \( M \)-dimensional data (system output or measurement) \( y = (y_1, \ldots, y_M)^\top \) are used to estimate the pipeline properties \( f_x(x), x \in [0, l] \). Note that \( y \) can be a time signal, a frequency signal, or a combination of both, measured at a single or many locations along the pipe.

It is assumed that each system output \( y_m (m \in \{1, \ldots, M\}) \) can be represented by a linear summation form, where each term corresponds to the contribution from (a property, for example defect, of) a specific location \( x \in [0, l] \). More precisely, the system outputs are decided by the spatial pipe property in a linear way. The transfer function \( G(y_m; x) \) between the pipe property \( f_x(x) \) at \( x \) and the \( m \)-th measurement \( y_m \) is assumed known, either analytically or numerically, so the system output can be represented by the integral equation with an additive random noise term \( n_m \):

\[
y_m = \int_0^l G(y_m; x)f_x(x)dx + n_m, \quad m = 1, \ldots, M
\]  

In pipe condition assessment problems, the purpose is to reconstruct the pipeline properties \( f_x(x), x \in [0, l] \) using the measurements \( y = (y_1, \ldots, y_M)^\top \). In practice, the pipe is often discretized by \( N \) discrete points \( x = (x_1, \ldots, x_N) \) along the pipe (Liggett and Chen 1994; Vitkovský et al. 2000; Zhou et al. 2018; Keramat et al. 2019a, b); the corresponding pipeline property is denoted by \( f(x) = (f_x(x_1), \ldots, f_x(x_N))^\top \). Furthermore, let the \( M \times N \) transfer matrix \( G = (G(y_m; x_n))_{m=1}^M{,n=1}^N \) and \( n = (n_1, \ldots, n_M)^\top \) be the random vector of measurement noise; then one has a discrete version of Eq. (1):

\[
y = Gf(x) + n
\]

Note that the general linear model Eq. (2) can describe a wide range of transient wave–based pipe condition assessment problems, including leakage detection (Nguyen et al. 2018; Wang and Ghaidawi 2018a, b), extended blockage detection (Zouari et al. 2017), discrete blockage detection (Lee et al. 2008b), and junction or pipe wall thickness estimation (Zeng et al. 2018). The present paper specifies the general linear model Eq. (2) with two illustrative examples, after the introduction of an inverse problem solution for the general model Eq. (2) in the next section.

### Regularization for Ill-Posed Inverse Problems

#### Tikhonov Regularization

The pipeline properties at the discrete points, i.e., \( f \) in Eq. (2), can be estimated by LS. When \( N \leq M \), i.e., the number of unknowns is not greater than the number of measurements, the LS solution is

\[
\hat{f} = \arg \min_f \|y - Gf\|^2 = (G^H G)^{-1} G^H y
\]

in which \( \| \cdot \|_2 \) stands for the \( L_2 \)-norm and the superscript \( H \) represents the operation of conjugate transpose of vector. The Gauss–Markov theorem illustrates that the LS solution in Eq. (3) is optimal in the meaning of the minimum variance unbiased estimator, provided the error terms in \( y \) are uncorrelated and identically distributed. However, even when \( N \leq M \), \( G \) may be ill-conditioned, such that the solution in Eq. (3) is ill-posed in the sense that a small perturbation of \( y \) and \( G \) leads to a very large perturbation of the solution \( \hat{f} \). Singular value decomposition (SVD) of the matrix \( G \) can be employed to verify its wellposedness (Hansen 1990):

\[
G = U \Sigma V^H
\]

where \( U \) and \( V \) are \( M \times M \) and \( N \times N \) left and right orthogonal matrices, \( \Sigma \) is an \( M \times N \) rectangular diagonal matrix with nonnegative singular values, denoted by \( \sigma_1, \ldots, \sigma_N \), of \( G \) on the diagonal. Let \( \sigma_1 \geq \cdots \geq \sigma_r = \sigma_r = 0 \), where \( r \) is the rank of \( G \). When \( r \neq N \) or some singular values are very close to zero, the solution of Eq. (3), which involves a matrix inversion operation, is very sensitive to perturbations of \( y \) and \( G \).
In the case that \( \mathbf{G} \) is ill-conditioned, a well-known method is the Tikhonov regularization (Tikhonov 1963), which modifies Eq. (3) by adding a regularization term:

\[
\hat{f}_\lambda = \arg \min_f \left[ \| \mathbf{y} - \mathbf{G}f \|^2_2 + \lambda \| f \|^2_2 \right]
\]

in which \( \lambda \) is regularization parameter. The solution of Eq. (5) is

\[
\hat{f}_\lambda = (\mathbf{G}^H\mathbf{G} + \lambda \mathbf{I})^{-1}\mathbf{G}^H\mathbf{y}
\]

Note that Eq. (5) is different from Eq. (3), so the properties of LS no longer hold. The determination of \( \lambda \) is critical and problem-dependent, i.e., a universal approach that can always generate satisfactory results for all the regularization problems does not exist (Bauer and Lukas 2011). In what follows, GCV and L-curve methods are introduced and tested in connection with this issue.

**Determination of Regularization Parameter**

GCV (Golub et al. 1979) originates from the leave-one-out (LOO) cross-validation approach. The basic idea is to use one observation \( y_m \) as the validation set and the remaining observations as data for estimating \( y_m \). The regularization parameter is decided by minimizing the average estimation error of \( y_m \) (Allen 1974). GCV is a rotation-invariant extension of the LOO cross-validation approach, where \( \lambda \) is obtained by minimizing

\[
GCV(\lambda) = \frac{\| (\mathbf{I} - \mathbf{A}(\lambda))\mathbf{y} \|^2_2}{\text{tr}(\mathbf{I} - \mathbf{A}(\lambda))}
\]

where

\[
\mathbf{A}(\lambda) = \mathbf{G}^H(\mathbf{G}^H\mathbf{G} + \lambda \mathbf{I})^{-1}\mathbf{G}^H
\]

\( \text{tr} \) = trace of a matrix; and \( \mathbf{I} \) = identity matrix.

L-curve (Hansen 1992; Hansen and O’Leary 1993) is a graphical approach that plots \( \log \eta(\lambda) \) versus \( \log \rho(\lambda) \) for all possible \( \lambda \), where \( \eta(\lambda) = \| \hat{f}_\lambda \|_2 \) and \( \rho(\lambda) = \| \mathbf{y} - \mathbf{G}\hat{f}_\lambda \|_2 \). As this curve often has a clear L-shape (the vertical part and horizontal part of the curve) concerning over- and underregularization, respectively, so the corner of the L-shape curve is identified as an optimal solution. The \( \lambda \) corresponding to the corner of the curve can also be derived by maximizing the curvature \( \kappa(\lambda) \) of the curve (log \( \rho(\lambda) \), log \( \eta(\lambda) \)), which results in

\[
\hat{\lambda} = \arg \max_{\lambda} \kappa(\lambda) = \arg \max_{\lambda} \frac{\tilde{\rho}'\tilde{\eta}'' - \tilde{\rho}''\tilde{\eta}'}{(\tilde{\rho}')^2 + (\tilde{\eta}')^2)^{3/2}}
\]

where \( \tilde{\rho} = \log \rho \) and \( \tilde{\eta} = \log \eta \).

In the following two sections, the general model Eq. (2) is specified by two application examples and the corresponding inverse problems are considered. In both examples, the importance of regularization and the determination of the regularization parameter are demonstrated with numerical and experimental data.

**Example 1: Time-Domain Impulse Response Function Extraction**

**Model Description**

IRF is widely used in pipeline condition assessment or defect detection problems (e.g., estimation of leakage, blockage, wall thickness, junction) (Vitkovský et al. 2003; Lee et al. 2007; Nguyen et al. 2018; Zeng et al. 2018). Here, a typical reservoir-pipe-valve system is considered, as shown in Fig. 1. An input signal \( I(t) \) is used to excite the pipe system where the location of the wave generator is at the downstream end of the pipe \( x = l \). Let \( O(t) \) denote the system output signal at \( x = l \) captured by the sensor. The system input and output have the following relationship provided the pipe system is assumed to be approximately a LTI system:

\[
O(t) = \int_{-\infty}^{+\infty} g(\tau)I(t - \tau)d\tau
\]

in which \( g(t) \) is the IRF of the system. Note that this LTI assumption is reasonable only if the magnitude of excited waves is appropriate (Lee and Vitkovský 2010; Lee 2013; Gong et al. 2013). In fact, \( g(t) \) represents the pipeline defect (reflectivity) property; if a reflected signal is received at time \( t \in (0, 2l/a) \), then a reflection happens at \( x = l - at/2 \), where \( a \) is the wave speed. Thus, \( g(t) \) relates to the pipe coordinate by

\[
f_x(x) = g\left(\frac{l - x}{a}\right), \quad x \in (0, l)
\]

or

\[
g(t) = f_x\left(\frac{l - at}{2}\right), \quad t \in (0, 2l/a)
\]

Let \( I[n], O[n], g[n], n = 1, \ldots, N \), be the discretized \( I(t), O(t), g(t) \); then a discrete version of the convolution Eq. (10) is obtained:

\[
O[n] = \sum_{k=1}^{N} g[k]I[N - k + 1]
\]

Therefore, a linear model Eq. (2) is obtained (Nguyen et al. 2018), in which

\[
y = (O[1], \ldots, O[N])^\top
\]
The signal emission and the measurement both last \( \Delta \) s. A PRBS input signal is emitted by the transmitter. The wave propagation in the pipe is realized by steady-state discharge; and (b) output pressure head perturbation at downstream end of pipe.

**Numerical Results**

A numerical example is studied in this section. The pipe length is \( l = 1,000 \) m and the cross-sectional diameter of the pipe is \( D = 0.2 \) m. The wave speed is \( a = 1,000 \) m/s. The Darcy–Weisbach friction factor is \( f_{DW} = 0.02 \). The pressure heads at the upstream reservoir and the downstream end are 20 and 0 m, respectively. A leak is located at \( x = 300 \) m, and the effective leak size is \( s^2 = C_l A_L^2 = 3 \times 10^{-4} \) m², where \( C_l \) is the discharge coefficient of the leak and \( A_L \) the orifice area of the leak. A transmitter and a sensor are both located at \( x = 1,000 \) m (the downstream end of the pipe). A PRBS input signal is emitted by the transmitter. The wave propagates along the pipe and the pressure is measured by the sensor. The signal emission and the measurement both last \( T = 80 \) s. Numerical simulation of wave propagation in the pipe is realized using the method of characteristics (MOC) (Chaudhry 2014). The grid lengths in the simulation are \( \Delta x = 10 \) m in space and \( \Delta t = \Delta x/a = 0.01 \) s in time, respectively. This implies that the sampling frequency is 100 Hz and signal length is \( N = 8,000 \). Fig. 2(a) shows the input signal (more precisely, the transient discharge at the valve normalized by the steady-state discharge) for 0–2 s, and Fig. 2(b) plots the output pressure head where \( t \in [0, 20] \) s.

SVD is carried out for \( G \) where \( N = 8,000 \). Fig. 3 shows the singular values (in descending order) normalized by their maximum value \( \sigma_1 \), in the numerical case of Example 1.
method determines the IRF by Liou (1998) and Beck et al. (2005):

$$R_{oi}(t) = \frac{1}{N} \sum_{n=1}^{N-t} I[n] O[n + t]$$  \hspace{1cm} (19)

Independent and identically distributed Gaussian-distributed random noise is added to the measurements. The signal-to-noise ratio (SNR) is 10 dB, where the reference pressure level is the averaged output signal, rather than the signal drop due to leak reflection in Ferrante et al. (2007), Wang and Ghidaoui (2018a, b), i.e.

$$\text{SNR} = 20 \log_{10} \left( \frac{1}{N} \sum_{n=1}^{N} |y[n]/\sigma| \right)$$  \hspace{1cm} (20)

in which $\sigma$ = standard deviation of Gaussian random noise. Fig. 4 shows the normalized reconstructed IRF. It indicates that the LS solution (without regularization) leads to a very noisy IRF estimate such that the leak cannot be identified. In contrast, the jump in the IRF due to leak reflection can be clearly found using the three other methods and GCV returns an IRF estimate with least noise. Fig. 5 plots $f(x)$ in Eq. (18) using the IRF estimation results in Fig. 4, which displays the leakage localization results. Again, this shows the importance of regularization and the GCV method returns a best leak localization result in this case.

Fig. 6 shows the average root-mean-square error (RMSE) of the normalized IRF estimate, which is defined by

$$\text{RMSE} = \frac{1}{T} \sqrt{\int_{0}^{T} (\text{IRF}(t) - \text{IRF}(t))^2}$$  \hspace{1cm} (21)

with respect to different SNRs, along with the corresponding 95% confidence intervals of the RMSE. Each average RMSE is obtained from 100 simulations, from data generation (with different

![Fig. 4. Estimation of IRF normalized by its maximum value. Simulated data with SNR = 10 dB are used. (Data from Wang et al. 2018.)](image-url)
Fig. 5. Plot of \( \hat{f}(x) \) for leak localization using IRF estimation results in Fig. 4; dashed line: actual leak location at \( x_L = 300 \) m; simulated data with SNR = 10 dB are used.

Fig. 6. Average and 95% confidence interval of RMSE of normalized IRF estimate with different SNRs.

Fig. 7. Input (designed signal) and output (experimental measurement) with PRBS excitation. (Data from Lee et al. 2008a.)
realization of random noise) to IRF extraction. Here, IRF stands for the estimate of IRF. At a relatively high SNR, it is clear that GCV has the lowest error. Note that it has been shown in the literature (Bauer and Lukas 2011; Hanke 1996) that the L-curve method is not convergent as the noise level tends to 0 and so is not recommended in this case. At a relatively low SNR, however, L-curve performs best. Note that the cross-correlation method has an error similar to that of L-curve. However, the cross-correlation method has much lower computational costs than GCV and L-curve, particularly at a large dimension of $G$ (the latter two methods need to compute the matrix inversion of $G$), and a large range of $\lambda$ needed to be tested.

To localize a defect, the RMSE of the estimated IRF must be much lower than the jump (signature of the defect) in the IRF. In Fig. 4, the effective leak size is $s_L = 3 \times 10^{-4}$ m$^2$, and the jump corresponding to the reflection at the leak is 0.2. For a larger leak, the reflection in the IRF increases; for example, if $s_L = 10^{-3}$ m$^2$, then the height of the jump is around 0.8, meaning that a higher

![Fig. 8. Singular values $\sigma_n$ of $G$ in descending order, normalized by their maximum value $\sigma_1$, in experimental case of Example 1.](image)

![Fig. 9. Impulse response function obtained from Eq. (5) with different values of Tikhonov regularization parameter $\lambda$ (from $10^{-4}$ to $10^5$). The experimental data in Fig. 7 are used.](image)
Experimental Results

In this section, experimental data in Lee et al. (2008a) are used, where a copper intact pipe with length \( l = 37.5 \text{ m} \) and diameter \( D = 0.022 \text{ m} \) is considered. A customized side-discharge valve is set at the downstream end of the pipe to excite the system. The design of the valve can be found in Fig. 1 in Lee et al. (2008a). A PRBS signal is generated by customized controls of the valve opening. The input PRBS and measurements are shown in Fig. 7, which is a reproduction of Fig. 4 in Lee et al. (2008a). The sampling rate is 2,000 Hz, the duration is 5 s, and thus the number of data is \( N = 10^5 \). Fig. 8 shows the singular values of \( G \) in descending order normalized by their maximum, which illustrates the ill-posedness of \( G \). Fig. 9 shows the IRF estimate obtained using Eq. (5) with various \( \lambda \) from \( 10^{-4} \) to \( 10^{5} \). This figure shows that the IRF estimation is sensitive to \( \lambda \); thus, its determination is crucial for obtaining an IRF with least perturbation.

Then IRF estimates using the cross-correlation, GCV, and L-curve approaches are plotted in Fig. 10. In this case, the cross-correlation method generates a very noisy result. The regularization parameters \( \lambda \) determined by GCV [via Eq. (7)] and L-curve [via Eq. (9)] are 20 and 210, respectively. The fluctuations in the GCV and L-curve results are relatively small in comparison to the reflection at the upstream boundary. L-curve returns almost optimal results in the sense of showing the least fluctuation in the estimated IRF, which can be found by comparing Fig. 10(c) with Fig. 9. For the purpose of pipe condition assessment, Fig. 11 plots the normalized \( \hat{f}(x) \) from Eq. (18) using the IRF estimation obtained from Fig. 10, which shows that no significant defect exists in the pipe. Note that, other than the random noise in the previous numerical example, model uncertainties exist in real experiments. For instance, the designed PRBS input signal is sharp at the corners, but in reality that sharpness cannot be achieved. The Tikhonov regularization can better cope with uncertainties by suppressing overfitting, while the cross-correlation approach seems to only cancel the random noise.

In this section, experimental results in an intact pipe are presented. However, it is emphasized that for pipe condition assessment problems, when reflection at a defect (e.g., a leakage) is higher than the perturbation of \( \hat{f}(x) \), it can be identified. For example, as indicated in Fig. 11(c), if reflection at a defect is higher than approximately 10% of boundary reflection at pipe ends, the L-curve method is definitely able to localize the defect. Here, we take leakage detection as an example. Let \( h_{\text{inc}} \) and \( h_{\text{ref}} \) denote the amplitudes of the incident transient waves and reflected waves at a leak. They have the relationship (Contractor 1965)

\[
h_{\text{ref}} = \frac{Z_C/Z_L}{2 + Z_C/Z_L} h_{\text{inc}} \approx \frac{Z_C}{2Z_L} h_{\text{inc}}
\]

where \( Z_C = a/(gA) \) = pipe impedance; \( Z_L = H_C^0/Q_L^0 \) = leak impedance; \( g = \) gravitational acceleration; \( A = \) area of pipeline; \( H_C^0 \) and \( Q_L^0 \) = steady-state head and discharge at leak; and \( \approx \) is due to the fact that in general \( Z_C/Z_L < 2 \) (Lin et al. 2019). According to Eq. (22), a 10% reflection implies that \( Z_C/Z_L \approx 0.2 \), which is indeed a small leak (Lin et al. 2019).

**Fig. 10.** Impulse response function estimated using the cross-correlation, GCV, and L-curve methods. The experimental data in Fig. 7 are used.

**Fig. 11.** Plot of normalized \( \hat{f}(x) \) for defect detection using IRF estimation obtained from Fig. 10. The experimental data in Fig. 7 are used.
Example 2: Leakage Detection Based on Frequency-Domain Linearized Model

Model Description

A linearized model of frequency-domain transient wave propagation in pipes is first introduced in Wang and Ghidaoui (2018a). In this section, a discrete version of the model in Wang and Ghidaoui (2018a) is considered. More specifically, some discrete points in the pipe are assumed to be potential leaks. Then the leak identification problem is solved by estimating the leak sizes of these potential leaks, which requires only solving the LS with Tikhonov regularization. On the other hand, using the model in Wang and Ghidaoui (2018a), the locations, sizes, and number of leaks need to be estimated separately (Wang and Ghidaoui 2018a, 2019; Wang et al. 2019a).

A single pipeline with leaks is considered, as shown in Fig 12. It is assumed that N discrete points in the pipe are potential leaks, located at

\[ x = (x_1, \ldots, x_N), \quad 0 < x_1 < \ldots < x_N < l \]  

(23)

The corresponding sizes of the N potential leaks are

\[ f(x) = (f_s(x_1), \ldots, f_s(x_N))^T \]  

(24)

in which each leak size is represented by \( f_s(x_n) = s^{ln} = C_d A^{ln} \), where \( A^{ln} \) is the orifice area of the nth leak. The steady-state discharge from a leak is related to the lumped leak parameter by \( Q_s^{ln} = f_s(x_n) \sqrt{2g(H_s^{ln} - z^{ln})} \), in which \( z^{ln} \) denotes the elevation of the pipe, and \( Q_s^{ln} \) and \( H_s^{ln} \) are the steady-state discharge and head of the nth leak.

The leaks are estimated by the pressure head difference at S sensors and J frequencies, denoted by

\[ y = (y_1, \ldots, y_M)^T \]  

(25)

Here, \( M = JS, y_m = h_{js}^M - h_{js}^NL [j = 1, \ldots, J \) and \( s = 1, \ldots, S, m = j + (s-1)J] \) denotes the head difference between the head measurement \( h_{js}^M \) (in the presence of leaks) and the theoretical head, which does not include the leak terms

\[ h_{js}^NL = h_{js}^NL(\omega_j, x_s) = -Z_j \sinh(\mu_j x_s)q(x^U) + \cosh(\mu_j x_s)h(x^U) \]  

(26)

at the frequency \( \omega_j \) and at the \( s \)th measurement station \( x_s \). Physically, \( y_m \) stands for the scattering of a wave at leaks. In Eq. (26), \( x^U \) is the upstream node of the pipe, \( Z_j = \mu_j a^2 / (i \omega_j g A) \) is the characteristic impedance, \( \mu_j = a^{-1} \sqrt{-\omega_j^2 + i g A \omega} \) is the propagation function, and \( R \) is the frictional resistance term. For turbulent flows, \( R = (f_{DW} Q_0) / (gDA^2) \), where \( Q_0 \) is the steady-state discharge in the pipe.

It is justified in Wang and Ghidaoui (2018a) that the theoretical expression of the data model approximately follows the linearized model:

\[ y = Gf(x) + n = \sum_{m=1}^{N} G_m(x_m) f_s(x_m) + n \]  

(27)

The matrix \( G \) in Eq. (27) is an \( M \times N \)-dimensional matrix whose \( nth \) column is

\[ G_m(x^{ls}) = (G(\omega_1, x^{ls}, x_1), \ldots, G(\omega_j, x^{ls}, x_1), \ldots, G(\omega_j, x^{ls}, x_s), \ldots, G(\omega_j, x^{ls}, x^L))^T \]  

(28)
in which
\[ G(\omega_j, x^L, x_s) = -\frac{\sqrt{Z} \sinh(\mu(x_s - x^L))}{\sqrt{2(H_0^L - z^L)}} \left( Z \sinh(\mu x^L) q(x^L) \right) \\
- \cosh(\mu x^L) h(x^L) \] (29)

The discharge \( q(x^U) \) in Eqs. (26) and (29) can be estimated (Kashima et al. 2012, 2013; Wang and Ghidaoui 2018a) by
\[ \hat{q}(x^U) = \frac{\cosh(\mu x^U) h(x^U) - h(x_0)}{Z \sinh(\mu(x_0 - x^U))} \] (30)
if the pressure at \( x_0 (x_0 > x^U) \) is measured, boundary condition \( h(x^U) \) is applied, and \( x_0 - x^U \) is assumed to be so small that no leak is between \( x^U \) and \( x_0 \). If the upstream of the pipe is connected to a reservoir, it is assumed that \( h(x^U) = 0 \).

Note that “≈” in Eq. (27) holds because the multiple reflections between leaks are neglected, which is a higher-order effect of the main reflection (the wave of a single reflection at a leak). This assumption is proven to be reasonable in the case of small leaks (Wang and Ghidaoui 2018a), which is the main concern of leak detection problems. Furthermore, it is shown in Zhou et al. (2018) that the sparse representation of leaks in Eq. (27) is reasonable in the sense that the spatial sampling mismatch converges to zero as the sample size of uniformly distributed assumed potential leaks tends to infinity. However, the LS solution and Tikhonov regularization for leakage detection based on the model Eq. (27) have never been investigated previously. The following two subsections examine these issues via numerical and experimental data; GCV and L-curve are tested for leakage detection in pipes.

**Numerical Results**

A numerical example is first considered. A pipe with length \( l = 2,000 \) m and diameter \( D = 0.5 \) m is assumed. The wave speed is \( a = 1,200 \) m/s. The Darcy–Weisbach friction factor is \( f_{DW} = 0.02 \). The pressure head at the upstream and downstream

![Fig. 14. Estimates of leak sizes \( f(x) \) at potential leaks with different regularization parameter \( \lambda \). The circles stand for actual leaks. The pipe length is 2,000 m. The two leaks are located at 300 and 800 m.](image-url)
boundary is 25 and 20 m, respectively. The pipe has two leaks; their locations are \( x_L^1 = 300 \text{ m} \) and \( x_L^2 = 800 \text{ m} \), and their sizes are \( s_{L1} = 10^{-4} \text{ m}^2 \) and \( s_{L2} = 1 \times 10^{-4} \text{ m}^2 \). The assumed potential leaks are located at \( x = \{50n:n = 1, 2, \ldots, 39\} \text{ m} \), i.e., \( N = 39 \).

Transient waves are generated by a fast valve operation at the downstream end of the pipe. The numerical simulation of wave generation and propagation in the frequency domain (more precisely, FRF) is carried out using the transfer matrix method (Chaudhry 2014). Gaussian-distributed random noise with \( \text{SNR} = 10 \text{ dB} \) is added to the pressure head numerical simulation results.

Here, the reference for defining SNR is the average pressure head difference (Wang and Ghidaoui 2018a, b; Wang et al. 2019c). One measurement station located at \( x_0 = 50 \text{ m} \) is used to estimate the boundary condition of flow discharge \( \hat{q}(x) \) via Eq. (30). Two other measurement stations at \( x_1 = 2000 \text{ m} \) and \( x_2 = 1800 \text{ m} \) are used to estimate the sizes of the potential leaks \( f(x) \). Thirty-one frequencies are selected for leak detection, and the corresponding angular frequencies are \( \omega = j\omega_h j = 1, 2, \ldots, 31 \), where \( \omega_h = \pi a/2l \) is the fundamental frequency, which implies 16 resonant frequencies and 15 antiresonant frequencies are used. Therefore, the number of data is \( M = 62 \).

Fig. 13 shows the normalized singular values \( \sigma_n/\sigma_1 \) of \( G \) in descending order, normalized by their maximum value \( \sigma_1 \), in numerical case of Example 2.

Fig. 17. Leak localization using MFP method. The dashed line and crosses indicate the locations of actual leak and sensors, respectively. The experimental data in Fig. 16 are used.

Fig. 18. Singular values \( \sigma_n \) of \( G \) in descending order, normalized by their maximum value \( \sigma_1 \), in numerical case of Example 2.
cannot be used for leak detection. Furthermore, it also shows that a good regularization parameter in the sense of clear leak detection is about \( \lambda = 10^{11} \), which is much larger than those \( \lambda \) in Example 1 in the previous section. Such a large value would be difficult to arrive at empirically. The very large value reflects the small size of the leak being identified.

Fig. 15 shows the estimates of leak sizes \( f(x) \) at the potential leaks \( x \) using LS without regularization \( [\lambda = 0 \text{ in Eq. (6)}] \), GCV \( [\lambda \text{ is decided by Eq. (7)}] \), and L-curve \( [\lambda \text{ is decided by Eq. (9)}] \). The regularization parameter decided by GCV and L-curve is \( \lambda = 9.8 \times 10^{11} \) and \( \lambda = 1.7 \times 10^{13} \), respectively. It is clear that both methods can accurately determine the locations of both leaks.

**Experimental Results**

In this section, experimental data in Lee (2005) are used to test the proposed methods. A copper pipe with length \( l = 37.4 \text{ m} \) and diameter \( D = 0.022 \text{ m} \) is used. The wave speed is \( a = 1,328 \text{ m/s} \). The pipe has one leak at the location \( x_L = 28.1 \text{ m} \) and size \( s_L = 1.77 \times 10^{-6} \text{ m}^2 \). Three sensors at \( x_0 = 6.7 \text{ m}, x_1 = 37.4 \text{ m}, \) and \( x_2 = 18.2 \text{ m} \) are set in the pipe; the first sensor is used to estimate the boundary condition of flow discharge \( \hat{q}(x_U) \) via Eq. (30), while the latter two are used to estimate the sizes of the potential leaks \( f(x) \). The wave is generated by a fast closing and opening operation of an in-line ball valve at the downstream end of the pipe. More details of the experimental setup can be found in Chapter 4 of Lee (2005).

The pressure heads measured from the three sensors in both the time domain and the frequency domain (FRF) are shown in Fig. 16. The selected frequencies for leak detection are \( f_\omega = j\omega_j : j = 1, 2, \ldots, 23 \). Thus, the number of data is \( M = 46 \). Fig. 17 displays the leak localization result using the matched-field processing (MFP) method (Wang and Ghidaoui 2018b, Wang et al. 2019b), which assumes a single leak in the pipe and determines the maximum point of the MFP objective function (solid curve in Fig. 17) as the leak estimate, i.e., 29.5 m (the actual leak location is 28.1 m).
Then the linear model with discrete potential leaks [Eq. (27)] is used and the regularization solution [Eq. (6)] is tested. Along the pipe, 17 potential leaks are assumed that are uniformly distributed between $x_0$ and the downstream end of the pipe (it is assumed that no leak exists between the upstream end of the pipe and $x_0$), i.e., $N = 17$. Fig. 18 shows the normalized singular values of $G$, which becomes almost 0 after the 12th singular value. Fig. 19 displays the results of leak size estimates with different regularization parameter $\lambda \in [10^{13}, 10^{17}]$, which clearly shows that $\lambda$ is crucial for regularization and for leak localization. Fig. 20 shows the results using LS without regularization ($\lambda = 0$). GCV [$\lambda$ is determined by Eq. (7)], and L-curve [$\lambda$ is determined by Eq. (9)]; in the latter two cases the corresponding value of $\lambda$ is $10^{12.65}$ and $10^{12.86}$. It is clear that both GCV and L-curve can localize the leak where the potential leak with the highest size estimate is at $x = 29.2$ m (the actual leak location is 28.1 m). In this case with experimental data, GCV and L-curve cannot very accurately estimate the leak size; the highest leak size estimate is not equal to the size of an actual leak, and the leak size estimates at other potential leaks are nonzero. As shown in Fig. 21, there exists a value of regularization parameter $\lambda$ (approximately $10^{13.7}$) that returns a better result in the sense of leak size estimation, and it is not captured by GCV or L-curve. However, GCV and L-curve are satisfactory since in practice leak localization is much more important than estimating leak size.

**Conclusion**

This paper examines the problems of condition assessment and defect detection of a pipeline using transient waves. A general linear model is proposed to describe a wide range of pipeline condition assessment problems. The system outputs are determined by a pipe property at discrete points along the pipe in a linear way. This general model is specified via two application examples: time-domain impulse response function extraction of a pipe system with pseudo-random binary sequence signals and pipeline leakage detection using frequency response functions.

With the proposed linear model, the pipeline properties can be replicated via LS. However, these inverse problems are usually ill-posed because real pipe systems involve many uncertain parameters, limited data sets, and very high noise. Therefore, the Tikhonov regularization scheme is employed. Generalized cross validation and L-curve methods are used to obtain the regularization parameter, which is crucial and strongly problem-dependent. In both

![Fig. 20. Leak size estimates of $f(x)$ at potential leaks using (a) LS without regularization ($\lambda = 0$); (b) GCV; and (c) L-curve. The experimental data in Fig. 16 are used. The circles stand for the actual leak. The pipe length is 37.4 m and the leak is located at 28.1 m.](image)

![Fig. 21. Leak size estimates of $f(x)$ at potential leaks with regularization parameter $\lambda = 10^{13.5}, 10^{13.625}, 10^{13.75}, 10^{13.875}$. The circles stand for the actual leak. The experimental data in Fig. 16 are used. The pipe length is 37.4 m, and the leak is located at 28.1 m.](image)
examples, numerical and experimental results illustrate the importance of regularization parameter and the efficiency of the proposed methods.

Data Availability Statement

All data, models, or code generated or used during the study are available from the corresponding author by request.

Acknowledgments

This work has been supported by research grants from the Research Grant Council of the Hong Kong SAR, China (Project T21-602/15R).

Notation

The following symbols are used in this paper:

- $A$ = area of pipe;
- $a$ = wave speed;
- $D$ = internal diameter of pipe;
- $f(x)$ = pipe property of concern at $x$;
- $G$ = transfer matrix from $f(x)$ to $y$;
- $l$ = pipe length;
- $x = (x_1, \ldots, x_N)^T$ = discrete points in pipe;
- $y = (y_1, \ldots, y_M)^T$ = measurement data;
- $n$ = measured noise; and
- $\lambda$ = regularization parameter.

References


Lee, P. J., J. P. Vitkovsky, M. F. Lambert, A. R. Simpson, and J. A. Liggett. 2008a. “Leak location using the pattern of the frequency response...