2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Bilinear parameter estimation with application in water leak localization

Victor Molnö, Henrik Sandberg*

Abstract— **In this paper, we present a novel statistical convergence analysis for bilinear parameter estimators. We account for two variations of a two-stage separation technique introduced by Bai [1], where the variations differ in the second stage. It turns out for both estimators that the probability of a large error decreases as the inverse square root of the number of measurements. We numerically demonstrate the estimators' performance by solving a water leak localization problem involving bilinear parameter estimation.**

I. Introduction

This paper deals with a two-stage approach to solve a bilinear parameter estimation problem. Bilinear equations, that depend linearly on each of two sub-parameters, $x \in \mathbb{R}^r$ and $y \in \mathbb{R}^m$, however, nonlinearly on the total parameter (x, y), are abundant throughout physics and engineering. This type of nonlinear equation appears in various parts of physics and engineering. Historical examples can be found in economic models [2], fluid dynamics [3], biological models [4], and mixture analysis in chemistry [5]. At the end of our paper, we present an application in water distribution system leak localization with uncertain pipe parameters.

Basic theory for systems of bilinear equations, including a general formulation and an account for the existence of solutions, is found in [6], [7].

When the system of bilinear equations is overdetermined and inconsistent, as is the case in a noisy estimation setting, least squares constitute a natural solution approach. In the control systems literature, such a setting arises in Hammerstein-Wiener system identification. With the Hammerstein-Wiener problem as motivation, bilinear parameter estimation is treated in [1], and a two-stage algorithm is suggested. It is shown that the two-stage algorithm estimates converge to the true parameters.

Further analysis of the bilinear estimation problem is conducted in [8], [9], [10], [11]. In [11], the two-stage method from [1] is analyzed for weighted and unweighted least squares. In [9], the convergence of an iterative alternative to the two-stage approach is analyzed. In [10], the two-stage approach is compared numerically to an iterative alternative and a third algorithm based on elimination theory. In [8], various iterative algorithms for the bilinear parameter estimation problem are compared numerically.

In this work, we return to the two-stage approach presented by Bai [1]. We suggest a variation of the two-stage approach with an alternative second stage. Our main contribution is the convergence rates that we derive for Bai's original two-stage algorithm, as well as for our variation. Our convergence rates in probability are more explicit than the convergence results in [1], which do not include any rate, and all other analyses of the bilinear parameter estimation we have seen.

In water distribution systems engineering, leak localization is an essential maintenance task. Automation of this task using smart sensors, is an active research field. See for example, the recent survey [12]. For a long transportation pipe, with flow and pressure sensors at the inlet and outlet of the pipe, it is possible to infer the leak location using sensor measurements along with a model of the head-loss function of the pipe, i.e., the relation between flow and pressure drop, due to friction losses. This problem is treated in [13]. However, uncertainty in the knowledge of the head-loss function makes leak localization difficult. Uncertain head-loss functions are analyzed in [14]. In that work, there are assumed bounds for pipe parameters. Leak locations that disagree with the model and these bounds are ruled out. An alternative in the case of the long transport pipe in [13], under pipe head-loss uncertainty, is to estimate the leak location and the pipe parameters simultaneously. We model the head-loss function so that this task takes the form of a bilinear parameter estimation problem, as discussed above. We use an EPANET [15] simulation of a leaky water pipe problem to verify the convergence rates we have derived for the two-stage approach presented by Bai [1] and for our variation of this.

The paper is organized as follows. In Section II, we state the bilinear parameter estimation Problem 1. In Section III, we show how Problem 1 can be solved via a two-stage procedure, as done in [1]. In Section IV, we propose an alternative solution to the second stage. We show that the probability of a large error decreases as the inverse square root of the number of measurements. In Section V, we find a similar convergence rate for the SVD method used for the second stage in [1]. The probability of a large error decreases as the inverse square root of the number of measurements with this approach as well. In Section VI, we apply both alternative estimators in a water leak localization scenario with data from EPANET [15]. Finally, in Section VII, we summarize our findings and present future research paths.

II. Problem formulation

In this section, we present a bilinear parameter estimation problem, which includes the problem in [1].

We consider a set of N equations

$$g_t = x^T A^{(t)} y + \varepsilon_t, \quad t = 1, \ldots, N, \tag{1}$$

*V. Molnö and H. Sandberg are with the Division of Decision and Control Systems, and Digital Futures (project DEMOCRITUS), KTH Royal Institute of Technology, Stockholm, Sweden. {vmolno,hsan}@kth.se.

where $g_t \in \mathbb{R}$, $x \in \mathbb{R}^r$, $y \in \mathbb{R}^m$, $A^{(t)} \in \mathbb{R}^{r \times m}$, and $\varepsilon_t \in \mathbb{R}$. The right-hand side of (1) is linear in x and y both. However, it is not linear in the total parameter (x, y), It is bilinear.The matrices $A^{(t)}$ describe the relation between the parameter (x, y) and the observations g_t. In [1], the equivalent, $\phi(k)$, (although formulated slightly differently) of $A^{(t)}$, contains transformations of the past inputs and outputs of the Hammerstein-Wiener dynamical system. In the current work, we do not make specific assumptions on the origin of $A^{(t)}$, but take them as given. The vector $\varepsilon = \begin{bmatrix}\varepsilon_1 & \cdots & \varepsilon_N\end{bmatrix}^T \sim \rho$ is i.i.d. measurement noise, from some distribution ρ, with expectation $\mathbb{E}_\rho[\varepsilon_t] = 0$, and variance $\mathbb{E}_\rho[\varepsilon_t^2] = \sigma^2$, $\sigma > 0$. The noise is independent of the $A^{(t)}$.

It should be noted that while we use the index t, which is common for correlated time series, we do not consider a dynamical system, and each instance of (1) is independent of the other.

Our problem is to estimate x and y, given $g = \begin{bmatrix}g_1 & \cdots & g_N\end{bmatrix}^T$ and $\{A^{(t)}\}_{t=1}^N$. However, for every $\gamma \in \mathbb{R} \setminus \{0\}$, it follows that $x'^T A^{(t)} y' = x A^{(t)} y$, where $x' = \gamma x$ and $y' = \gamma y$. Thus the parameters x and y that solve (1) are not unique unless we add further constraints. Therefore, we will assume a constraint $x^T b = 1$, for a known $b \in \mathbb{R}^r$. We note that there are equally valid alternative constraints, such as $\|x\| = 1$, with $x^T \mathbf{1} > 0$, etc. We summarize the estimation task in Problem 1.

Problem 1: Given $g = \begin{bmatrix}g_1 & \ldots & g_N\end{bmatrix}^T$ and $\{A^{(t)}\}_{t=1}^N$, from (1), estimate x and y, where $x^T b = 1$.

III. Over-parameterization

In this section, we show how Problem 1 can be solved via a two-stage procedure and give a convergence rate for the first stage.

We repeat the over-parameterization procedure in [1], rewriting (1) as a linear equation system

$$g = A(x \otimes y) + \varepsilon. \tag{2}$$

Here

$$x \otimes y = \begin{bmatrix} x_1 y \\ \vdots \\ x_r y \end{bmatrix} = \mathrm{vec}(y x^T) \in \mathbb{R}^{rm}$$

is the Kronecker product between x and y, and

$$A = \begin{bmatrix} a_{11}^{(1)} & \ldots & a_{1m}^{(1)} & \cdots & a_{r1}^{(1)} & \cdots & a_{rm}^{(1)} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{11}^{(N)} & \ldots & a_{1m}^{(N)} & \cdots & a_{r1}^{(N)} & \cdots & a_{rm}^{(N)} \end{bmatrix},$$

$A \in \mathbb{R}^{N \times rm}$, is a matrix where each row is a flattened $A^{(t)}$, $t = 1, \ldots, N$.

Based on the reformulated equation (2), the bilinear parameter estimation problem can be solved via a two-stage procedure:

1) finding an estimate $\widehat{x \otimes y}$ of $x \otimes y$,
2) finding estimates $\hat{x}$ and $\hat{y}$ of x and y through an approximation $\hat{x} \otimes \hat{y}$ of $\widehat{x \otimes y}$.

Remark 1: Stage 2) above is essentially a rank-one approximation problem, where $\hat{y}\hat{x}^T$ approximates

$$\begin{aligned} \widehat{yx^T} &:= \mathrm{vec}^{-1}(\widehat{x \otimes y}) \\ &:= \begin{bmatrix}(\widehat{x \otimes y})_{1:m} & \cdots & (\widehat{x \otimes y})_{(r-1)m+1:rm}\end{bmatrix} \in \mathbb{R}^{m \times r}. \end{aligned}$$

Now if both stages 1) and 2) can be solved with small errors, then Lemma 1 ensures that $\hat{x}$ and $\hat{y}$ estimate x and y with small errors.

Lemma 1: For $x, x' \in \mathbb{R}^r$, $y, y' \in \mathbb{R}^m$ and $\delta > 0$, under the uniqueness condition $x^T b = x'^T b = 1$,

$$\|x' \otimes y' - x \otimes y\| < \delta \tag{3}$$

implies

$$\|y' - y\| < \delta \|b\|. \tag{4}$$

Furthermore, for $\delta < \|y\|/\|b\|$, (3) and (4) imply

$$\|x' - x\| < \delta \frac{1 + \|b\|\|x\|}{\|y\| - \delta\|b\|}. \tag{5}$$

Proof: We have

$$\begin{aligned} \|y' - y\| &= \|y' 1 - y 1\| = \|y' x'^T b - y x^T b\| \\ &\le \|y' x'^T - y x^T\|_F \|b\| \\ &= \|x' \otimes y' - x \otimes y\| \|b\| \end{aligned}$$

Thus (3) implies (4). Now assuming (3) and (4),

$$\begin{aligned} \|x' - x\|(\|y\| - \delta\|b\|) &< \|x' - x\|(\|y\| - \|y - y'\|) \\ &\le \|(x' - x)\|\|y'\| \\ &= \|(x' - x) \otimes y'\| \\ &\le \|x' \otimes y' - x \otimes y\| + \|x\|\|y - y'\| \\ &< \delta + \delta\|x\|\|b\|. \end{aligned}$$

Therefore, if $\delta < \|y\|/\|b\|$, (5) follows. ∎

Remark 2: Regarding exact approximation, under $x^T b = x'^T b = 1$, $x' \otimes y' = x \otimes y$ if and only if $y' = y$, $x' = x$.

We now start with stage 1) of the two-stage procedure, which can be efficiently solved with the linear least squares estimator.

Lemma 2: Suppose A has full column rank ($\dim(\mathrm{range}(A)) = rm$). Let $\widehat{x \otimes y} = A^+ g = (A^T A)^{-1} A^T g$. Then

$$\mathbb{E}_\rho[\widehat{x \otimes y}] = x \otimes y,$$

and

$$\mathbb{E}_\rho[(\widehat{x \otimes y} - x \otimes y)(\widehat{x \otimes y} - x \otimes y)^T] = \sigma^2 (A^T A)^{-1}.$$

Lemma 2 is a standard linear least squares result, so the proof is omitted.

We now introduce some notation to describe how $(A^T A)^{-1}$ can approach zero.

Definition 1: The set $\{A^{(t)}\}_{t=1}^\infty$ is *PE*(c, n) (persistently exciting of order (c, n)) if for $c \in \mathbb{R}$, $c > 0$, $n \in \mathbb{N}$ and every $t \ge 1$, $(A^{(t:t+n-1)})^T A^{(t:t+n-1)} - cI \ge 0$, where ≥ 0 means positive semidefinite and

$$A^{(t:t+n-1)} = \begin{bmatrix} a_{11}^{(t)} & \ldots & a_{rm}^{(t)} \\ \vdots & & \vdots \\ a_{11}^{(t+n-1)} & \ldots & a_{rm}^{(t+n-1)} \end{bmatrix} \in \mathbb{R}^{n \times rm},$$

are rows t through $t+n-1$ of A. We say that $\{A^{(t)}\}_{t=1}^{\infty}$ is PE (persistently exciting) if it is PE(c,n) for some c and n.

Remark 3: In simpler terms, $\{A^{(t)}\}_{t=1}^{\infty}$ being PE(c,n) implies that for every n new measurements we get a new full column rank $A^{(t:t+n-1)}$. Thus the PE property says something about the variation in $\{A^{(t)}\}_{t=1}^{\infty}$. In Section VI, we will explain how we can make sure that $\{A^{(t)}\}_{t=1}^{\infty}$ is PE in a water leak localization scenario.

Lemma 3: If $\{A^{(t)}\}_{t=1}^{\infty}$ is PE(c,n), then the smallest singular value of the full A-matrix satisfies $s^2_{\min}(A) \geq c\lfloor N/n\rfloor$.

Proof: We notice that

$$\begin{aligned} A^TA &= \sum_{t=1}^{N} \text{vec}((A^{(t)})^T)\text{vec}((A^{(t)})^T)^T \\ &= \sum_{k=1}^{\lfloor N/n\rfloor} \left((A^{(1+(k-1)n:kn)})^T A^{(1+(k-1)n:kn)}\right) \\ &\quad + (A^{(\lfloor N/n\rfloor n+1:N)})^T A^{(\lfloor N/n\rfloor n+1:N)}. \end{aligned}$$

We multiply from left and right with the unit eigenvector ξ associated with the smallest eigenvalue $\lambda_{\min}(A^TA) = s^2_{\min}(A)$.

$$\begin{aligned} \xi^TA^TA\xi &\geq \sum_{k=1}^{\lfloor N/n\rfloor} \xi^T(A^{(1+(k-1)n:kn)})^T A^{(1+(k-1)n:kn)}\xi \\ &\geq \sum_{k=1}^{\lfloor N/n\rfloor} c\xi^T\xi \geq c\lfloor N/n\rfloor \end{aligned}$$

■

We can now give a convergence rate for the error of stage 1).

Theorem 1: If $\{A^{(t)}\}_{t=1}^{\infty}$ is PE(c,n), then

$$\mathbb{E}_\rho[\|\widehat{x\otimes y} - x\otimes y\|^2] \leq \sigma^2 rm/(c\lfloor N/n\rfloor).$$

Proof: Using Lemma 2, we get $\mathbb{E}_\rho[\|\widehat{x\otimes y} - x\otimes y\|^2] = \text{trace}(\mathbb{E}_\rho[(\widehat{x\otimes y} - x\otimes y)(\widehat{x\otimes y} - x\otimes y)^T]) = \sigma^2\text{trace}((A^TA)^{-1}) \leq \sigma^2 rm/s^2_{\min}(A) \leq \sigma^2 rm/(c\lfloor N/n\rfloor)$, where the last part follows from Lemma 3. ■

According to Lemma 1 and Theorem 1, if we can solve stage 2), we can accurately estimate x and y (assuming $\{A\}_{t=1}^{\infty}$ is PE). In Sections IV and V, we present and show convergence for two alternative solutions to stage 2).

IV. Separated solution

In this section, we present our new approach to calculate the estimates $\hat{x}$ and $\hat{y}$ as an approximation $\hat{x}\otimes\hat{y}$ of $\widehat{x\otimes y}$. The approach consists of two steps:

a) estimating y via $\hat{y} = (b^T\otimes I)\widehat{x\otimes y} = \begin{bmatrix} b_1I_m & \cdots & b_rI_m\end{bmatrix}\widehat{x\otimes y}$.
b) estimating x via the least squares estimator $\hat{x} = \arg\min_z \|z\otimes\hat{y} - \widehat{x\otimes y}\|^2$.

For $\hat{y}$ of step a), we have a result similar to Lemma 2 and Theorem 1.

Lemma 4: $\mathbb{E}_\rho[\hat{y}] = y$. Further, if A is PE(c,n), $\mathbb{E}_\rho[\|\hat{y} - y\|^2] \leq \sigma^2\|b\|^2rm^2/(c\lfloor N/n\rfloor)$.

Proof:

$$\begin{aligned} \mathbb{E}_\rho[\hat{y}] &= (b^T\otimes I)\mathbb{E}_\rho[\widehat{x\otimes y}] \\ &= (b^T\otimes I)(x\otimes y) = \sum_{i=1}^{r} b_ix_iy = (b^Tx)y = y. \end{aligned}$$

For the variance, we can expand $y = (b^T\otimes I_m)(x\otimes y)$. So $\hat{y} - y = (b^T\otimes I_m)(\widehat{x\otimes y} - x\otimes y)$, which gives

$$\mathbb{E}_\rho[(\hat{y}-y)(\hat{y}-y)^T] = (b^T\otimes I_m)\sigma^2(A^TA)^{-1}(b^T\otimes I_m)^T$$

Therefore

$$\begin{aligned} \mathbb{E}_\rho[\|\hat{y}-y\|^2] &= \text{trace}(\mathbb{E}_\rho[(\hat{y}-y)(\hat{y}-y)^T]) \\ &\leq \sigma^2\text{trace}((b^T\otimes I_m)^T(b^T\otimes I_m)) \\ &\quad\times \text{trace}((A^TA)^{-1}) \\ &\leq \sigma^2 m\|b\|^2 rm/(c\lfloor N/n\rfloor) \end{aligned}$$

■

Finding $\hat{x}$ in step b) is easy. We notice that $\arg\min_z \|z\otimes\hat{y} - \widehat{x\otimes y}\|^2 = \arg\min_z \sum_{i=1}^r \|z_i\sum_{j=1}^m y_j - \sum_{j=1}^m(\widehat{x\otimes y})_{(i-1)m+j}\|^2$ separates into one problem for each component z_i. The solution is

$$\hat{x} = \frac{1}{\hat{y}^T\hat{y}}(I_r\otimes\hat{y}^T)\widehat{x\otimes y} = \frac{1}{\|\hat{y}\|^2}\begin{bmatrix}\hat{y}^T & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \hat{y}^T\end{bmatrix}\widehat{x\otimes y}.$$

This estimator fulfills the constraint

$$\begin{aligned} \hat{x}^Tb = b^T\hat{x} &= \frac{1}{\|\hat{y}\|^2}\begin{bmatrix}b_1 & \cdots & b_r\end{bmatrix}\begin{bmatrix}\hat{y}^T & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \hat{y}^T\end{bmatrix}\widehat{x\otimes y} \\ &= \frac{1}{\|\hat{y}\|^2}\begin{bmatrix}b_1\hat{y}^T & \cdots & b_r\hat{y}^T\end{bmatrix}\widehat{x\otimes y} \\ &= \frac{1}{\|\hat{y}\|^2}\hat{y}^T\begin{bmatrix}b_1I_m & \cdots & b_rI_m\end{bmatrix}\widehat{x\otimes y} \\ &= \frac{\|\hat{y}\|^2}{\|\hat{y}\|^2} = 1. \end{aligned}$$

However, $\hat{x}$ does not depend linearly on ε, making it difficult to analyze the mean and variance without auxiliary information about the distribution ρ. Instead, we give a convergence bound in probability.

Theorem 2: For every $\delta' > 0$,

$$\mathbb{P}_\rho(\|\hat{x} - x\| > \delta') \leq \frac{\sigma\hat{f}(\|x\|,\|y\|,\|b\|,m,r,\delta')}{\delta'\sqrt{c\lfloor N/n\rfloor}},$$

where $\hat{f}(\|x\|,\|y\|,\|b\|,m,r,\delta') = \sqrt{rm}(\|x\|\|b\|\sqrt{m} + 2)(1+\|b\|\|x\| + \delta'\|b\|)/\|y\|$.

Proof: According to Lemma 1, for every δ such that $\|y\|/\|b\| > \delta > 0$, if $\|\hat{x}\otimes\hat{y} - x\otimes y\| < \delta$ then $\|\hat{x} - x\| < \delta\dfrac{1+\|b\|\|x\|}{\|y\| - \delta\|b\|}$. That is,

$$\mathbb{P}_\rho\left(\|\hat{x}-x\| > \delta\frac{1+\|b\|\|x\|}{\|y\|-\delta\|b\|}\right) \leq \mathbb{P}_\rho(\|\hat{x}\otimes\hat{y} - x\otimes y\| > \delta).$$

We will upper bound this probability.

$$\begin{aligned}\|\hat{x}\otimes\hat{y}-x\otimes y\| &\le \|\hat{x}\otimes\hat{y}-\widehat{x\otimes y}\|+\|\widehat{x\otimes y}-x\otimes y\|\\ &\le \|x\otimes\hat{y}-\widehat{x\otimes y}\|+\|\widehat{x\otimes y}-x\otimes y\|\\ &= \|x\|\|\hat{y}-y\|+2\|\widehat{x\otimes y}-x\otimes y\|.\end{aligned}$$

Therefore

$$\begin{aligned}&\mathbb{P}_\rho(\|\hat{x}\otimes\hat{y}-x\otimes y\|>\delta)\\ &\quad\le \mathbb{P}_\rho(\|x\|\|\hat{y}-y\|+2\|\widehat{x\otimes y}-x\otimes y\|>\delta)\\ &\quad\le \frac{\|x\|\mathbb{E}_\rho[\|\hat{y}-y\|]+2\mathbb{E}_\rho[\|\widehat{x\otimes y}-x\otimes y\|]}{\delta}\\ &\quad\le \frac{\|x\|\sqrt{\mathbb{E}_\rho[\|\hat{y}-y\|^2]}+2\sqrt{\mathbb{E}_\rho[\|\widehat{x\otimes y}-x\otimes y\|^2]}}{\delta}.\end{aligned}$$

Using Theorem 1 and Lemma 4 we get

$$\begin{aligned}&\mathbb{P}_\rho(\|\hat{x}-x\|>\delta\frac{1+\|b\|\|x\|}{\|y\|-\delta\|b\|})\\ &\quad\le \frac{\|x\|\sqrt{\sigma^2\|b\|^2rm^2/(c\lfloor N/n\rfloor)}+2\sqrt{\sigma^2rm/(c\lfloor N/n\rfloor)}}{\delta}.\end{aligned}$$

Now if we let $\delta'=\delta\frac{1+\|b\|\|x\|}{\|y\|-\delta\|b\|}$, we get $\delta=\delta'\frac{\|y\|}{1+\|b\|\|x\|+\delta'\|b\|}$, and the result follows. What remains to show is that $\delta<\|y\|/\|b\|$ for every $\delta'>0$. To this end, see that $\delta\|b\|=\delta'\frac{\|y\|}{1/\|b\|+\|x\|+\delta'}=\frac{\|y\|}{1/(\delta'\|b\|)+\|x\|/\delta'+1}<\|y\|$. ■

Remark 4: The key element in the proof of Theorem 2 is noticing that the least square property of $\hat{x}$ gives $\|\hat{x}\otimes\hat{y}-\widehat{x\otimes y}\|\le\|x\otimes\hat{y}-\widehat{x\otimes y}\|$. The SVD estimators in the coming Section V are $(\tilde{x},\tilde{y})=\arg\min_{z,w}\|z\otimes w-\widehat{x\otimes y}\|^2$. Therefore $\|\tilde{x}\otimes\tilde{y}-\widehat{x\otimes y}\|\le\|\hat{x}\otimes\hat{y}-\widehat{x\otimes y}\|$ and so the bounds we derive there (Theorem 3) are necessarily tighter. However, as we will see in Example 1, this does not mean that $\|\tilde{x}-x\|\le\|\hat{x}-x\|$ always.

V. SVD SOLUTION

In this section, we present the SVD solution to the approximation stage 2) that Bai [1] uses. While there is a convergence proof (Theorem 2.1) in the original paper, no specific rate is given. We give such a rate in our Theorem 3.

The estimators in this section are based on the SVD of $\widehat{yx^T}$. In particular, we let $\tilde{x}=\hat{v}_1/\hat{v}_t^Tb$ and $\tilde{y}=\hat{u}_1\hat{v}_t^Tb$, be the singular vectors associated with the largest singular value $\hat{s}_1$, where $\sum_{i=1}^p\hat{s}_i\hat{u}_i\hat{v}_i^T=\widehat{yx^T}$ ($p=\min(m,r)$, $\hat{s}_1\ge\hat{s}_2\ge\ldots\hat{s}_p\ge0$) is the SVD.

We know from the proof of Lemma 2 that $\widehat{x\otimes y}=x\otimes y+A^+\varepsilon$. Therefore $\widehat{yx^T}=yx^T+\text{vec}^{-1}(A^+\varepsilon)$. We let $\sum_{i=1}^p s_i^{(\varepsilon)}u_i^{(\varepsilon)}v_i^{(\varepsilon)T}=\text{vec}^{-1}(A^+\varepsilon)$ be the SVD of the error $\text{vec}^{-1}(A^+\varepsilon)$. Lemma 5 relates $\hat{s}_2,\ldots,\hat{s}_p$ and $s_1^{(\varepsilon)},\ldots,s_p^{(\varepsilon)}$.

Lemma 5: $\hat{s}_i\le s_{i-1}^{(\varepsilon)}$, $i=2,\ldots p$. Therefore $\|\tilde{x}\otimes\tilde{y}-\widehat{x\otimes y}\|\le\text{trace}(A^+\varepsilon\varepsilon^T(A^+)^T)$.

Proof: We rewrite $yx^T=\sum_{i=1}^p s_iu_iv_i^T$, where clearly $s_1=\|x\|\|y\|$, $u_1=y/\|y\|$, $v=x/\|x\|$, $s_i=0,i=2,\ldots,p$. Assume for the rest of the proof that $i\ge2$. We let

$$\begin{aligned}V_2&=\text{span}(\{v_j\}_{j=2}^r),\\ V_{i-1}^{(\varepsilon)}&=\text{span}(\{v_j^{(\varepsilon)}\}_{j=i-1}^r).\end{aligned}$$

Then $\dim(V_2\cap V_{i-1}^{(\varepsilon)})\ge r-i+1$. Therefore

$$\begin{aligned}\hat{s}_i&=\min_{(\hat{V}_i\in\mathbb{R}^r:\dim(\hat{V}_i)=r-i+1)}\max_{(\hat{v}\in\hat{V}_i:\|\hat{v}\|=1)}\|\widehat{yx^T}\hat{v}\|\\ &\le\max_{\hat{v}\in(V_2\cap V_{i-1}^{(\varepsilon)}):\|\hat{v}\|=1}\|\widehat{yx^T}\hat{v}\|\\ &\le\max_{\hat{v}\in(V_2\cap V_{i-1}^{(\varepsilon)}):\|\hat{v}\|=1}\|(yx^T+\text{vec}^{-1}(A^+\varepsilon))\hat{v}\|\\ &\le\max_{\hat{v}\in(V_2\cap V_{i-1}^{(\varepsilon)}):\|\hat{v}\|=1}(\|yx^T\hat{v}\|+\|\text{vec}^{-1}(A^+\varepsilon)\hat{v}\|)\\ &\le\max_{v\in V_2:\|v\|=1}\|yx^T\hat{v}\|\\ &\qquad+\max_{v^{(\varepsilon)}\in V_{i-1}^{(\varepsilon)}:\|v^{(\varepsilon)}\|=1}\|\text{vec}^{-1}(A^+\varepsilon)\hat{v}\|\\ &=s_2+s_{i-1}^{(\varepsilon)}=0+s_{i-1}^{(\varepsilon)}.\end{aligned}$$

Using this first result, we get $\|\tilde{x}\otimes\tilde{y}-\widehat{x\otimes y}\|=\|\tilde{y}\tilde{x}^T-\widehat{yx^T}\|_F^2=\|\hat{s}_1\hat{u}_1\hat{v}_1^T-\sum_{i=1}^p\hat{s}_i\hat{u}_i\hat{v}_i^T\|=\|\sum_{i=2}^p\hat{s}_i\hat{u}_i\hat{v}_i^T\|=\sqrt{\sum_{i=2}^p\hat{s}_i^2}\le\sqrt{\sum_{i=1}^p(s_i^{(\varepsilon)})^2}=\|\text{vec}^{-1}(A^+\varepsilon)\|_F=\|A^+\varepsilon\|=\sqrt{\text{trace}(A^+\varepsilon\varepsilon^T(A^+)^T)}$. ■

Remark 5: Lemma 5 is a special case of Weyl's inequality. A full proof can be found in the book [16].

Following Lemma 5, we get a result with a probability convergence rate bound on $\|\tilde{x}-x\|$, similar to Theorem 2.

Theorem 3: For every $\delta'>0$,

$$\mathbb{P}_\rho(\|\tilde{x}-x\|>\delta')\le\frac{\sigma\tilde{f}(\|x\|,\|y\|,\|b\|,m,r,\delta')}{\delta'\sqrt{c\lfloor N/n\rfloor}},$$

where $\tilde{f}(\|x\|,\|y\|,\|b\|,m,r,\delta')=2\sqrt{rm}(1+\|b\|\|x\|+\delta'\|b\|)/\|y\|$.

Proof: The proof is similar to that of Theorem 2. We will use Lemma 1 and 5. $\|\tilde{x}\otimes\tilde{y}-x\otimes y\|\le\|\tilde{x}\otimes\tilde{y}-\widehat{x\otimes y}\|+\|\widehat{x\otimes y}-x\otimes y\|=\|\tilde{y}\tilde{x}^T-\widehat{yx^T}\|_F+\|\widehat{x\otimes y}-x\otimes y\|\le2\sqrt{\text{trace}(A^+\varepsilon\varepsilon^T(A^+)^T)}$, where we used Lemma 5 for the last inequality. Now using Lemma 1, $\mathbb{P}_\rho(\|\tilde{x}-x\|>\delta\frac{1+\|b\|\|x\|}{\|y\|-\delta\|b\|})\le\mathbb{P}_\rho(\|\tilde{x}\otimes\tilde{y}-x\otimes y\|>\delta)\le\frac{2\sqrt{\mathbb{E}_\rho[\text{trace}(A^+\varepsilon\varepsilon^T(A^+)^T)]}}{\delta}=\frac{2\sqrt{\sigma^2rm/(c\lfloor N/n\rfloor)}}{\delta}$. Now we let $\delta'=\delta\frac{1+\|b\|\|x\|}{\|y\|-\delta\|b\|}$ like in the proof of Theorem 2 and the result follows. ■

Remark 6: Theorem 2 and Theorem 3 display similarity between $\hat{x}$ and $\tilde{x}$ (same dependence on σ, c, N, n), but differ in the functions $\hat{f}$ and $\tilde{f}$. In particular, the ratio $\hat{f}(\|x\|,\|y\|,\|b\|,m,r,\delta')/\tilde{f}(\|x\|,\|y\|,\|b\|,m,r,\delta')=1+\|x\|\|b\|\sqrt{m}/2>1$. As we mentioned in Remark 4, this means that Theorem 3 provides a tighter bound than Theorem 2. However, we want to emphasize again that this

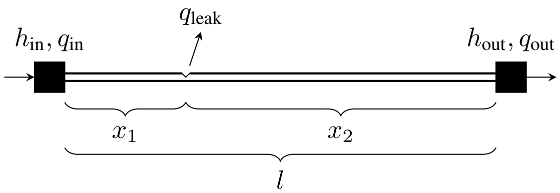

Fig. 1. A water pipe with a leak, as described in Section VI.

does not necessarily mean that $\|\tilde{x} - x\| < \|\hat{x} - x\|$, in all problem instances.

Remark 7: Lemma 4 in Section IV provides a bound for the $\hat{y}$ error. Using Lemma 1, and the reasoning in the proof of Theorem 3, it is possible to prove a similar bound for the $\tilde{y}$ error. Namely, $\mathbb{P}_\rho(\|\tilde{y} - y\| > \delta') \leq \frac{\sigma}{\delta' \sqrt{c\lfloor N/n \rfloor}} 2\|b\|\sqrt{rm}$. However, we refrain from formally articulating this for two reasons. First, it is not necessary for deriving Theorem 3 (unlike the necessity of Lemma 4 for deriving Theorem 2). Second, our focus within the water leak localization problem does not concern the estimation of y but solely centers on the estimation of x.

VI. Water leak localization application

In this section, we apply the estimators $\hat{x}$ and $\tilde{x}$ in a water leak localization scenario. The problem concerns a leaking water pipe, as shown in Fig. 1. At the inlet (left), there is a flow q_{in} and a hydraulic head h_{in}. At the outlet (right), there is a flow q_{out} and a hydraulic head h_{out}. At position x_1 between the inlet and the outlet, there is a leak, with outflow $q_{\text{leak}} = q_{\text{in}} - q_{\text{out}}$. The problem at hand is to estimate the leak position x_1. The sum $x_1 + x_2 = l$, where l is the (known) length of the pipe.

There is a steady-state head-loss relation

$$h_{\text{in}} - h_{\text{out}} = x_1 U(q_{\text{in}}) + x_2 U(q_{\text{out}}), \tag{6}$$

where the head-loss function $U(q)$ describes the rate per unit length at which the hydraulic head decreases along a pipe section with flow q.

Estimation of x_1, has been treated in [13]. In that work, the water system operator is assumed to know the head-loss function $U(q)$. Under this assumption, can be readily found given measurements of h_{in}, h_{out}, q_{in} and q_{out}. Unfortunately, in practice, due to aging pipes, effects of changing temperature, etc. When a-priori unknown, $U(q)$ and x_1 have to be estimated simultaneously.

We consider a parameterization of $U(q)$ for which the simultaneous estimation of the leak position and the head-loss function is an instance of our bilinear parameter estimation problem.

Specifically, we assume a model $U(q) = \phi(q)y$ where $\phi(q)^T = \begin{bmatrix}\phi_1(q) & \cdots & \phi_m(q)\end{bmatrix}^T \in \mathbb{R}^m, y \in \mathbb{R}^m$, and an observation model

$$\begin{aligned}\hat{h}_{\text{in},t} - \hat{h}_{\text{out},t} &= x_1 \phi(q_{\text{in},t}) y + x_2 \phi(q_{\text{out},t}) y + \varepsilon_t \qquad (7)\\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix}\phi(q_{\text{in},t}) \\ \phi(q_{\text{out},t})\end{bmatrix} y + \varepsilon_t, t = 1, \ldots, N.\end{aligned}$$

Here, $\hat{h}_{\text{in},t}$ and $\hat{h}_{\text{out},t}$ are sensor readings of the hydraulic head, corrupted by noise (of which the difference is ε_t). For simplicity, we assume perfect measuring of flows, i.e., we know $q_{\text{in},t}$ and $q_{\text{out},t}$. However, we plan to extend our analysis to an error-in-variables model, where $q_{\text{in},t}$ and $q_{\text{out},t}$ are also corrupted by noise. We recognize that $\hat{h}_{\text{in},t} - \hat{h}_{\text{out},t} = g_t$, $\begin{bmatrix} x_1 & x_2 \end{bmatrix} = x^T$ and $\begin{bmatrix}\phi(q_{\text{in},t}) \\ \phi(q_{\text{out},t})\end{bmatrix} = A^{(t)}$ puts (7) in the form of (1).We can apply the estimators described throughout Sections III, IV and V.

The convergence results of Theorems 2 and 3, require that $\{A^{(t)}\}_{t=1}^\infty$ is PE. This is a property of $\{(q_{\text{in},t}, q_{\text{out},t})\}_{t=1}^N$ and $\phi(q)$. We have

$$A = \begin{bmatrix} \phi(q_{\text{in},1}) & \phi(q_{\text{out},1}) \\ \vdots & \vdots \\ \phi(q_{\text{in},N}) & \phi(q_{\text{out},N}) \end{bmatrix}.$$

If for example $q_{\text{in},1} = q_{\text{in},2} = \cdots = q_{\text{in},N}$ (and $m > 1$), we get

$$A \begin{bmatrix} 1 \\ -\phi_1(q_{\text{in},1})/\phi_2(q_{\text{in}}, 1) \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0,$$

i.e., A is not even full rank. Informally, for $\{A\}_{t=1}^\infty$ to be PE, the flows need to fulfill some notion of sufficient variation. We will not investigate such a notion any further. Instead, we present two cases where $\{A^{(t)}\}_{t=1}^\infty$ is indeed PE.

The first case describes a general result for a water system where the operator has control capabilities. According to Proposition 1, a water system operator with access to pumps and valves to control the flow in the pipes can repeat a set of flows to attain a matrix A that is PE.

Proposition 1: Repeated measurements. If there is a set $\{(q_{\text{in},t}, q_{\text{out},t})\}_{t=1}^n$ such that $s_{\min}(A^{(1:n)}) = c > 0$, that is repeated over and over: $\{(q_{\text{in},t}, q_{\text{out},t})\}_{t=kn+1}^{(k+1)n} = \{(q_{\text{in},t}, q_{\text{out},t})\}_{t=1}^n$, $k = 1, \ldots$, then A is PE(c, n).

Proof: Follows immediately from Definition 1. ■

The second case is a scenario where we do not show formally that $\{A\}_{t=1}^\infty$ is PE, but where the leak position estimates $\hat{x}_1$ and $\tilde{x}_1$ (first elements of $\hat{x}$ and $\tilde{x}$) converge in practice.

Example 1: Numerical simulation. We generate data mimicking a leaking pipe using the water network simulation software EPANET [15]. We define pipe parameters such as length, diameter, friction factor, and leak position, as well as a pressure dependent leak function $q_{\text{leak}} = C\sqrt{h_{\text{leak}}}$, based on Bernoulli's principle. We simulate the leaking pipe with q_{out} drawn uniformly between 10 and 20 liters per second, and h_{in} drawn uniformly between 100 and 110 meters elevation, independently of q_{out}. Given q_{out} and h_{in}, EPANET calculates the hydraulic state of the pipe. The sample means and standard deviations for the variables available to the system operator are shown in Table I. The values in Table I are presented excluding the sensor noise ε, which is normal with $\sigma = 0.1$ meter.

TABLE I
DATA DISTRIBUTION FOR EXAMPLE 1.

variable	sample mean	sample standard deviation
q_{in} (l/s)	22.1	2.9
q_{out} (l/s)	15.0	2.9
h_{in} (m)	105.0	2.9
h_{out} (m)	98.4	3.5

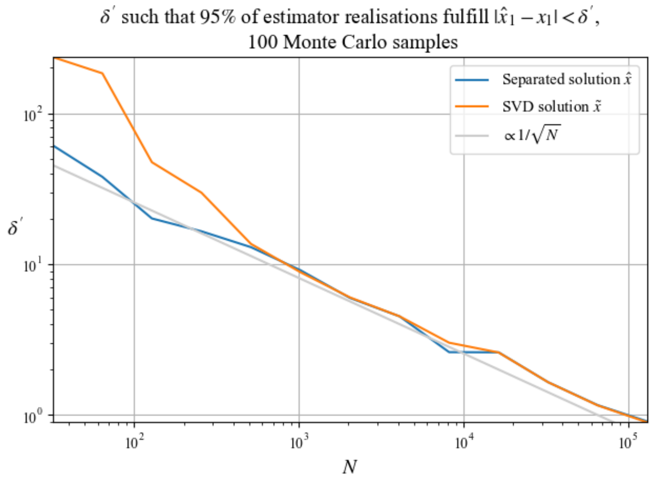

Fig. 2. Convergence of $\hat{x}_1$ and $\tilde{x}_1$ in Example 1. 95%-quantile for errors.

EPANET uses the Darcy-Weisbach head-loss function in the simulation. We approximate this function with $\phi(q)y = \left[q/q_0 \quad (q/q_0)^2\right] y$, where y $(m = 2)$ is to be determined (feature scaling by the typical flow q_0 improves numerical stability). Then, we calculate estimators $\hat{x}$ and $\tilde{x}$ using the procedures described in Section IV and V. Fig. 2 shows the convergence of the first elements $\hat{x}_1$ and $\tilde{x}_1$, representing the leak position estimates. The convergences are as we expect from Theorem 2 and Theorem 3 inversely proportional to the square root of the number of measurements. Furthermore, we see that, while the probability error upper bound we managed to derive for the separated solution (Theorem 2) is larger than that of the SVD solution (Theorem 3), the separated estimator performs better for small numbers of measurements N. For large N, the estimators are comparable.

Remark 8: Lemmas and theorems in this paper do not take into account the distribution of the flows q_{in} and q_{out} (essentially the distribution of A). In the theoretical results we explicate the noise distribution ρ ($\mathbb{E}_\rho$ and $\mathbb{P}_\rho$), and assume A as given.

VII. CONCLUSION

In this paper, we have derived convergence rates for two bilinear parameter estimators. We have demonstrated their performance in a water leak localization scenario. In future work, we will incorporate an error-in-variables analysis to deal with imperfect flow measurements. We will also investigate the difference between the separated and SVD estimators for small numbers of measurements.

ACKNOWLEDGEMENT

We are grateful to the anonymous reviewers for their valuable feedback, which has significantly contributed to enhancing the clarity of our paper.

REFERENCES

[1] Er-Wei Bai. An optimal two-stage identification algorithm for Hammerstein–Wiener nonlinear systems. Automatica, Volume 34, Issue 3, 1998, pages 333-338.

[2] Masanao Aoki. Some Examples of Dynamic Bilinear Models in Economics. In book: Antonio Ruberti, Ronald R. Mohler. Variable Structure Systems with Application to Economics and Biology. Lecture Notes in Economics and Mathematical Systems, vol 111, 1975, pages 163–169.

[3] Hans Heinrich Bleich, Ivan S. Sandler, Interaction between structures and bilinear fluids, International Journal of Solids and Structures, Volume 6, Issue 5, 1970, pages 617-639.

[4] Yves Pagnotte, Alain Pavé, Application of Bilinear Models to Biological Problems. An Approach to Mechanism Inference and Analysis, IFAC Proceedings Volumes, Volume 10, Issue 12, 1977, pages 165-170.

[5] Nematollah Omidikia, Mahdiyeh Ghaffari, Jeroen Jansen, Lutgarde Buydens, Roma Tauler, Bilinear model factor decomposition: A general mixture analysis tool, Chemometrics and Intelligent Laboratory Systems, Volume 240, 2023.

[6] Scott Cohen and Carlo Tomasi. Systems of bilinear equations. Technical report, Stanford, CA, USA, 1997.

[7] Charles R. Johnson, Helena Šmigoc, Dian Yang, Solution Theory for Systems of Bilinear Equations, Linear and Multilinear Algebra, Volume 62, 2013, Pages 1553-1566.

[8] Lars Eldén, Salman Ahmadi-Asl, Solving bilinear tensor least squares problems and application to Hammerstein identification, Numerical Linear Algebra with Applications, 2019.

[9] Er-Wai Bai, Duan Li, Convergence of the iterative Hammerstein system: identification algorithm, IEEE. Trans. Automat. Control, volume 49, 2004, pages 1929 –1940.

[10] Er-Wei Bai, Yun Liu, Least squares solutions of bilinear equations, Systems & Control Letters, Volume 55, Issue 6, 2006, Pages 466-472.

[11] Jiandong Wang, Qinghua Zhang, Lennart Ljung, Revisiting Hammerstein system identification through the Two-Stage Algorithm for bilinear parameter estimation, Automatica, Volume 45, Issue 11, 2009, Pages 2627-2633.

[12] Luis Romero-Ben, Débora Alves, Joaquim Blesa, Gabriela Cembrano, Vicenç Puig, and Eric Duviella. Leak detection and localization in water distribution networks: Review and perspective. Annual Reviews in Control, volume 55, 2023.

[13] Ludvig Lindström, Sebin Gracy, Sindri Magnússon, Henrik Sandberg, Leakage Localization in Water Distribution Networks: A Model-Based Approach, in 2022 European Control Conference (ECC), 2022, pages 1515–1520.

[14] Stelios G. Vrachimis, Stelios Timotheou, Demetrios G. Eliades, Marios M. Polycarpou, Leakage detection and localization in water distribution systems: A model invalidation approach, Control Engineering Practice, Volume 110, 2021.

[15] Lewis A. Rossman, EPANET Users Manual : Project Summary. Cincinnati, OH: U.S. Environmental Protection Agency, Risk Reduction Engineering Laboratory, 1994.

[16] Roger A. Horn, Charles R. Johnson. Topics in matrix analysis. Cambridge: Cambridge University Press, 1991.