Experiments in Fluids (2019) 60:126
https://doi.org/10.1007/s00348-019-2768-4

RESEARCH ARTICLE

Singular value decomposition of noisy data: noise filtering

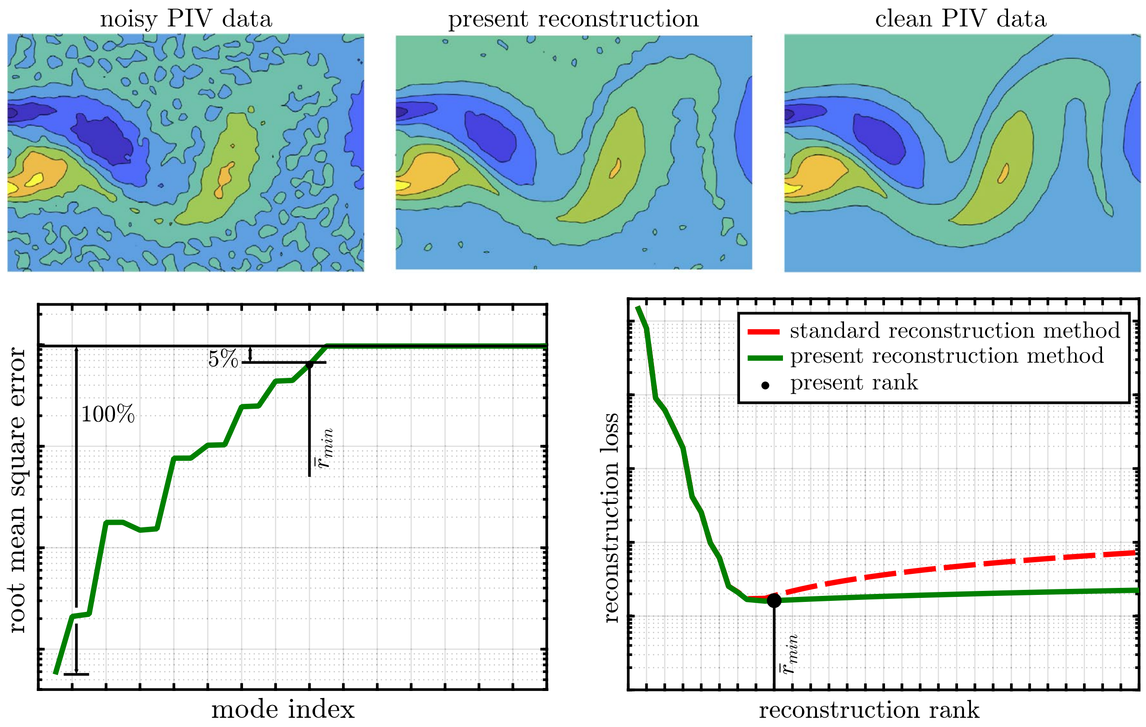
Brenden P. Epps[1] · Eric M. Krivitzky[1]

Received: 22 January 2019 / Revised: 20 May 2019 / Accepted: 24 June 2019 / Published online: 16 July 2019

Abstract

The *singular value decomposition* (SVD) and *proper orthogonal decomposition* are widely used to decompose velocity field data into spatiotemporal modes. For noisy experimental data, the lower SVD modes remain relatively clean, which suggests the possibility for data filtering by retaining only the lower modes. Herein, we provide a method to (1) estimate the noise level in a given noisy dataset, (2) estimate the root mean square error (rmse) of the SVD modes, and (3) filter the noise using only the SVD modes that have low enough rmse. We show through both analytic and PIV examples that this method yields nearly the most accurate possible SVD-based reconstruction of the clean data. Moreover, we provide an analytic estimate of the accuracy of this reconstruction.

Graphic abstract



✉ Brenden P. Epps
brenden.epps@dartmouth.edu

[1] Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA

1 Introduction

In fluid mechanics, *singular value decomposition* (SVD) and *proper orthogonal decomposition* (POD)[1] are used for particle image velocimetry (PIV) data regression (Sherry et al. 2013; Raiola et al. 2015; Mendez et al. 2017; Brindise and Vlachos 2017); identification of coherent structures (Kriegseis et al. 2009; Druault et al. 2012; Kourentis and Konstantinidis 2012; Marié et al. 2013; Xu et al. 2013; Gandhi et al. 2015); low-order modeling, possibly for flow control (Ma et al. 2003; Feng et al. 2011; Feng and Wang 2014); optimal sensor placement (Cohen et al. 2004); dynamic mode decomposition (Schmid 2010; Dawson et al. 2016); and more. In (Epps and Krivitzky 2019), we considered mode corruption due to noise present in the data, building on the works of Kato (1976), Breuer and Sirovich (1991), Venturi (2006) and Epps and Techet (2010). Herein, we consider use of the SVD/POD for data de-noising.

The SVD and POD are very attractive for de-noising experimental velocity field data, because no other rank r approximation captures more of the kinetic energy[2] in the data as the sum of the first r SVD modes (Schmidt 1907). However, when the data are corrupted by noise, it is a priori unclear how badly the SVD modes are corrupted by the noise and which modes to use when noise filtering.

Motivating questions for this article include the following: How effective is the SVD for filtering out noise and reconstructing the clean data? What method yields the most accurate reconstruction, and what are the limits to its accuracy? Can the magnitude of the noise be inferred from the noisy data themselves?

To define our goals mathematically, consider the following definitions: Let $\mathbf{A} \in \mathbb{R}^{T\times D}$ be a matrix of *clean data*, with T time steps and D data sites ($T < D$). The reduced *singular value decomposition* of $\mathbf{A}$ is

$$\mathbf{A} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^\intercal = \textstyle\sum_{k=1}^{T} \mathbf{u}_k s_k \mathbf{v}_k^\intercal\ , \tag{1}$$

where matrices $\mathbf{U} \in \mathbb{R}^{T\times T}$ and $\mathbf{V} \in \mathbb{R}^{D\times T}$ are each *orthogonal* and $\mathbf{S} \in \mathbb{R}^{T\times T}$ is diagonal. The kth SVD mode consists of a *left singular vector* (temporal mode shape) $\mathbf{u}_k \equiv U_{1:T,k}$, a *singular value* $s_k \equiv S_{kk}$, and a *right singular vector* (spatial mode shape) $\mathbf{v}_k \equiv V_{1:D,k}$. Let $\tilde{\mathbf{A}} \in \mathbb{R}^{T\times D}$ be a matrix of *noisy data:*

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}\ , \tag{2}$$

where $\mathbf{E} \in \mathbb{R}^{T\times D}$ contains random noise with zero mean and standard deviation ϵ.[3] Let $\tilde{\mathbf{A}} = \tilde{\mathbf{U}}\,\tilde{\mathbf{S}}\,\tilde{\mathbf{V}}^\intercal = \sum_{k=1}^{T} \tilde{\mathbf{u}}_k \tilde{s}_k \tilde{\mathbf{v}}_k^\intercal$ and $\mathbf{E} = \acute{\mathbf{U}}\,\acute{\mathbf{S}}\,\acute{\mathbf{V}}^\intercal = \sum_{k=1}^{T} \acute{\mathbf{u}}_k \acute{s}_k \acute{\mathbf{v}}_k^\intercal$.

A *reconstruction* of the clean data can be formed by summing the first r SVD modes:

$$\bar{\mathbf{A}}_r \equiv \sum_{k=1}^{r} \bar{\mathbf{u}}_k \bar{s}_k \bar{\mathbf{v}}_k^\intercal\ , \tag{3}$$

where these barred variables could be the noisy (tilde) variables or any estimate of the clean ones. To evaluate such a reconstruction, we use the *reconstruction loss*

$$\varDelta_r \equiv \|\mathbf{A} - \bar{\mathbf{A}}_r\|_F^2 = \sum_{t=1}^{T} \sum_{d=1}^{D} (A_{td} - (\bar{A}_r)_{td})^2\ , \tag{4}$$

where $\|\cdot\|_F$ is the *Frobenius norm*, so (4) is a measure of the difference between the clean data and the reconstruction.

Herein, we consider being given noisy data $\tilde{\mathbf{A}}$ but not knowing $\mathbf{A}$ and ϵ. Our goals are (1) to make an accurate estimate $\bar{\epsilon}$ of the measurement error; and (2) to form a minimum-loss estimate $\bar{\mathbf{A}}_r$ of the clean data.

1.1 Existing noise filtering methods

A number of methods exist to choose which SVD modes to use for noise filtering. Epps and Techet (2010) reconstruct an estimate of the clean data using all noisy SVD modes for which

$$\tilde{s}_k > \epsilon\sqrt{TD}\ , \tag{5}$$

where $\tilde{s}_k$ is the singular value of noisy mode k, ϵ is the root mean square error in the data, T and D are the number of timesteps and data sites, respectively.

Raiola et al. (2015) use all noisy SVD modes for which

$$F_k \equiv \frac{s_k^2 + \epsilon^2 D}{s_{k-1}^2 + \epsilon^2 D} < 0.999\ . \tag{6}$$

Note that (6) can not be used in practice, since one does not know the clean singular values (if only given a noisy dataset). However, (6) can be put into usable form by noting that Raiola et al. (2015) derived (6) assuming $\tilde{s}_k^2 \approx s_k^2 + \epsilon^2 D$, so the spirit of (6) is equivalent to

$$\tilde{F}_k \equiv \frac{\tilde{s}_k^2}{\tilde{s}_{k-1}^2} < 0.999\ . \tag{7}$$

[1] In (Epps and Krivitzky 2019), we showed that for discrete data, the SVD and POD yield identical results. Thus, our results apply to both the SVD and POD.

[2] Since the *Frobenius norm* of the kth SVD mode is s_k, then for velocity field data, the kinetic energy (per unit mass) of the mode is $\frac{1}{2}s_k^2$.

[3] Ideally, $\mathbf{E}$ contains i.i.d. noise drawn from a Gaussian distribution, but herein we also consider $\mathbf{E}$ containing spatially-correlated noise, as occurs in PIV data.

Brindise and Vlachos (2017) do not sum noisy modes $k = 1, \dots, r$ but instead choose the r "smoothest" modes, as identified by the *Shannon entropy*. Their so called *entropy line fit* (ELF) method is a six step procedure:

1. Compute the *discrete cosine transform* (DCT-II) of each right singular vector,

$$\hat{V}_{dk} = \sqrt{\frac{2/N}{1+\delta_{d1}}} \sum_{i=1}^{D} V_{ik} \cos\left[\frac{\pi}{D}\left(i - \tfrac{1}{2}\right)(d-1)\right], \qquad (8)$$

for $d = 1, \dots, D$ and $k = 1, \dots, T$.

2. Square and normalize the DCT coefficients:

$$X_{dk} = \hat{V}_{dk}^2 \;/\; \textstyle\sum_{d=1}^{D} \hat{V}_{dk}^2 \,. \qquad (9)$$

3. Compute the *Shannon entropy* of the normalized DCT signal[4]

$$Y_k = -\sum_{d=1}^{D} X_{dk} \log_2 X_{dk} \,. \qquad (10)$$

4. Sort the minimum entropy into ascending order, $\bar{Y}_k$.
5. Use a "two-line fit" procedure to find the the optimal reconstruction rank, r_B, where the sorted entropy $\bar{Y}_k$ nearly reaches a plateau.
6. Use *sorted* modes $1, \dots, r_B$ for data reconstruction.

Note that while the *Schmidt theorem* guarantees that this reconstruction $\bar{\mathbf{A}}_r$ will yield more loss to the noisy data $\|\tilde{\mathbf{A}} - \bar{\mathbf{A}}_r\|_F^2$ than that using *unsorted* modes $1, \dots, r$, this reconstruction might (or might not) yield less loss to the clean data $\|\mathbf{A} - \bar{\mathbf{A}}_r\|_F^2$.

Shabalin and Nobel (2013) reconstruct the clean data using all the modes, but they discount the singular values in such a way as to suppress the more noisy modes.

Note that the above references (as well as this manuscript) assume a complete dataset with no large outliers. Other papers have considered POD for reconstruction of gappy data or data with extreme outliers (Venturi and Karniadakis 2004; Wang et al. 2015; Higham et al. 2016).

2 Proposed 'E15' noise filtering method

The present work builds upon (Epps and Krivitzky 2019), wherein we derived and validated a theoretical prediction (13) of the *root mean square error* of the SVD modes, which is defined as

$$\operatorname{rmse}(\tilde{\mathbf{v}}_k) \equiv \left[\frac{1}{D}\sum_{i=1}^{D}\left(\tilde{V}_{ik} - V_{ik}\right)^2\right]^{\frac{1}{2}} . \qquad (11)$$

Further, we showed that for modes constituting random noise, the $\operatorname{rmse}(\tilde{\mathbf{v}}_k)$ reaches a maximum value of $\sqrt{2/D}$. The key idea of our proposed noise filtering method is to use only the modes for which the $\operatorname{rmse}(\tilde{\mathbf{v}}_k)$ is sufficiently below this noise ceiling.

Herein, we propose the following method for minimum-loss noise filtering.

1. Given a noisy data matrix $\tilde{\mathbf{A}}$, perform the SVD (Matlab `svd` command) to obtain $\tilde{\mathbf{u}}_k$, $\tilde{s}_k$, and $\tilde{\mathbf{v}}_k$.
2. Estimate the measurement error $\bar{\epsilon}$ and the 'spatial correlation parameter' f by fitting a *Marchenko–Pastur distribution* to the tail of the noisy singular values. See "Appendix 1" for details.
3. (Optional) Infer the 'effective smoothing window width', w, from the curve fit in Fig. 14b, which reads

$$w = 1 + \left(2f - \tfrac{3}{2}\right)\left(1 - \mathrm{e}^{-20(f-1)}\right) . \qquad (12)$$

See "Appendix 1" for discussion of w. For PIV data, however, we recommend that (12) *not* be used and instead to set $w = 1$ (see PIV examples in Sects. 4–6).

4. Estimate the *root mean square error* of the modes:

$$\operatorname{rmse}(\tilde{\mathbf{v}}_k) \approx \min\left\{ \sqrt{2/D}\,,\; \frac{\bar{\epsilon}}{\tilde{s}_k}\left[\frac{D-w}{D} + \frac{w}{D}\sum_{\substack{m=1\\ m\neq k}}^{T} \frac{\tilde{\lambda}_m(3\tilde{\lambda}_k - \tilde{\lambda}_m)}{(\tilde{\lambda}_m - \tilde{\lambda}_k)^2}\right]^{\frac{1}{2}} \right\}, \qquad (13)$$

where $\tilde{\lambda}_k \equiv \tilde{s}_k^2$. This formula was derived and validated in (Epps and Krivitzky 2019).

5. Estimate the rank for minimum-loss reconstruction as follows:

$$t_k \equiv \frac{\log(\operatorname{rmse}(\tilde{\mathbf{v}}_k)) - \log(\sqrt{2/D})}{\log(\operatorname{rmse}(\tilde{\mathbf{v}}_1)) - \log(\sqrt{2/D})}, \qquad (14)$$

$$\boxed{\bar{r}_{\min} \equiv \text{maximum } k \text{ such that } \quad t_k > 5\%} \,. \qquad (15)$$

The parameter t_k in (14) quantifies the cleanliness of a mode, where $t_1 = 1$ for the first (cleanest) mode, and $t_k = 0$ for modes at the noise ceiling ($\operatorname{rmse}(\tilde{\mathbf{v}}_k) = \sqrt{2/D}$). Modes that are sufficiently below this noise ceiling (i.e. that have a large enough t_k) are deemed clean enough to be useful for data reconstruction. The threshold in (15)

[4] If the data $\mathbf{A}$ are 2D-2C velocity fields, then compute the 2D DCT-II and entropy of each velocity component separately, and then take Y_k as the minimum entropy.

was set at 5%, because we empirically found that level to yield accurate and robust results.

6. Reconstruct an estimate of the clean singular values:

$$\bar{s}_k = \begin{cases} \sqrt{\tilde{s}_k^2 - (\epsilon'\hat{s}_k)^2} & (k < k_c) \\ 0 & (k_c \le k) \end{cases}, \qquad (16)$$

where $\epsilon'\hat{s}_k$ is a *Marchenko–Pastur distribution* (see "Appendix 1"), and

k_c is the minimum index k such that $\tilde{s}_k < \epsilon'\hat{s}_k$ (Epps 2015). This cutoff index ensures that (16) yields a real number, and it sets $\bar{s}_k$ to zero for modes in the tail of the distribution, which are obliterated by noise. Equation (16) follows from the observation that $\tilde{s}_k^2 \approx s_k^2 + (\epsilon'\hat{s}_k)^2$.

7. Reconstruct an estimate of the clean data via

$$\bar{\mathbf{A}}_r = \sum_{k=1}^{r} \tilde{\mathbf{u}}_k \bar{s}_k \tilde{\mathbf{v}}_k^{\mathsf{T}}, \qquad (17)$$

with $r = \bar{r}_{\min}$ from (15) and $\bar{s}_k$ from (16). We will refer to (15)/(16)/(17) as the *'E15' reconstruction.*

2.1 Illustrative example

To illustrate this procedure by way of example, consider a laminar, unsteady, 2D flow past a cylinder at Reynolds number $Re = V_\infty d/\nu = 100$. A clean (noise-free) dataset $\mathbf{A}$ is extracted from CFD simulations, with $D = 33{,}284$ data sites (x and y velocities at 16,642 grid points) and $T = 455$ timesteps. The time-mean were *not* removed from these data. A noisy dataset $\tilde{\mathbf{A}}$ is created by adding *independent, identically-distributed (i.i.d.)* Gaussian noise with standard deviation $\epsilon = 10^{-4}$. Since the noise is *i.i.d.*, $w = f = 1$.

Figure 1 provides a snapshot in time of the velocity magnitude, showing the clean data, noisy data, and a data set formed by the 'E15' reconstruction procedure (15)/(16)/(17). This reconstruction filters out much of the noise and faithfully represents the clean data.

Figure 2a illustrates the singular values of the clean and noisy datasets, as well as the 'E15' reconstruction (16) and a best-fit Marchenko–Pastur distribution. Equation (31) uses this best-fit Marchenko–Pastur distribution to predict $\bar{\epsilon} = \epsilon' = 1.0019 \times 10^{-4}$, just 0.19 percent greater than the true ϵ.

Figure 2b shows the root mean square error of the spatial mode shapes. The theoretical prediction (13) agrees well with the numerically-computed rmse. The rmse is very low for the first mode and reaches $\sqrt{2/D}$ for modes that are saturated with noise. The optimum reconstruction rank $\bar{r}_{\min} = 9$ (15) is illustrated as the index for which the rmse approaches 5% of this noise ceiling, since modes with higher rmse constitute noise.

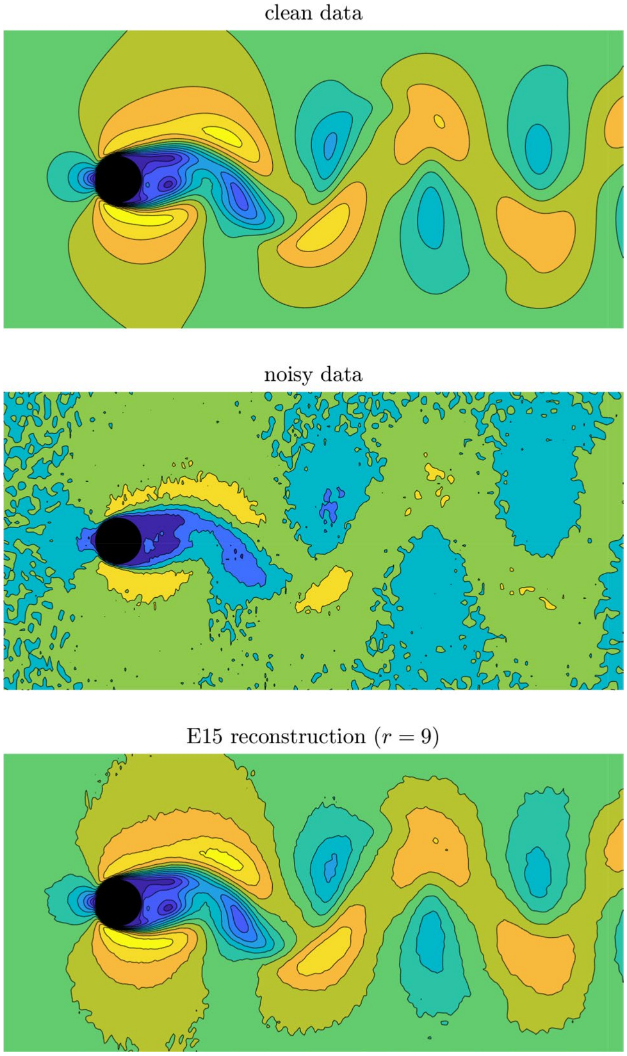
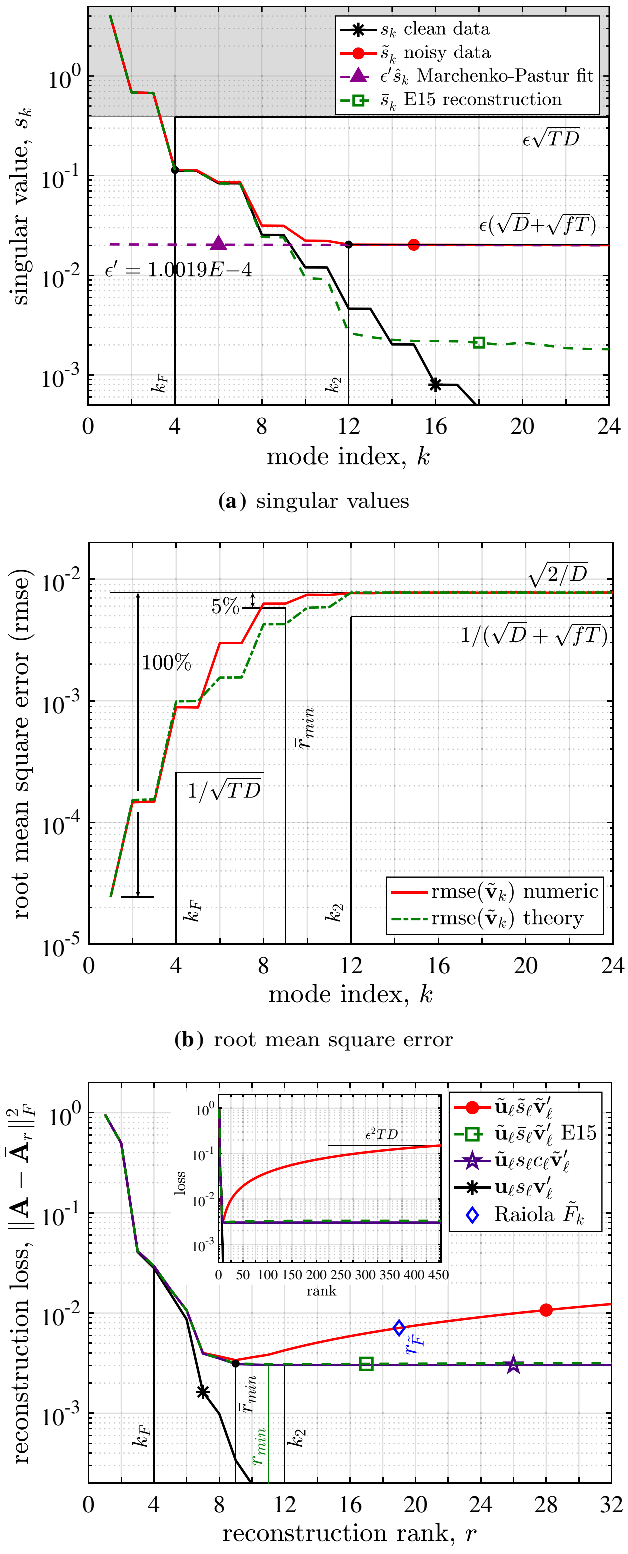


Fig. 1 (Sect. 2.1 example) Snapshot of velocity magnitude, showing the clean data (top), noisy data (middle), and the 'E15' reconstruction (bottom). Flow is left to right

Figure 2c presents the loss Δ_r (4) versus rank r for several reconstruction methods. For example, using the noisy singular values and vectors, $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathsf{T}}$, is labeled '$\tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathsf{T}}$'. As expected, the loss for the '$\tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathsf{T}}$' reconstruction (red curve) has a pronounced minimum, because the lower modes contain most of the signal, and the higher modes contain mostly noise.

The proposed reconstruction method (15)–(17) is labeled '$\tilde{\mathbf{u}}_\ell \bar{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathsf{T}}$ E15' in Fig. 2c. Figure 2c shows the loss for the 'E15' reconstructions at all ranks (green dashed curve), but note that the 'E15' method selects rank $\bar{r}_{\min} = 9$ (per 15). There are two key points to emphasize: First, the 'E15' loss at $\bar{r}_{\min}$ is slightly lower than the minimum loss achievable using the standard '$\tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathsf{T}}$' reconstruction (red curve).



(a) singular values

(b) root mean square error

(c) reconstruction loss

Fig. 2 (Sect. 2.1 example) **a** Singular values. **b** Root mean square error, comparing theory (13) versus numerical results, and illustrating the optimum rank prediction $\bar r_{\min}$ (15). **c** Loss (4) for reconstructions using the summands shown in the legend

Moreover, the 'E15' loss at $\bar r_{\min}$ is only 1.7% higher than the minimum 'E15' loss (which actually occurs at $r_{\min} = 11$).

Second, the main advantage of the '$\tilde{\mathbf u}_\ell \bar s_\ell \tilde{\mathbf v}_\ell^\intercal$ E15' reconstruction, which uses reconstructed singular values $\bar s_\ell$ per (16), is that its loss is much less sensitive to the choice of rank $\bar r_{\min}$ (for ranks greater than the actual $r_{\min}$), because the singular values for the higher (noisier) modes are suppressed. In contrast, the loss of the standard '$\tilde{\mathbf u}_\ell \tilde s_\ell \tilde{\mathbf v}_\ell^\intercal$' approach increases (dramatically in some cases) for ranks $r > r_{\min}$. Thus, we will see in Sect. 3 that the 'E15' method achieves lower loss *and* less variability than the standard '$\tilde{\mathbf u}_\ell \tilde s_\ell \tilde{\mathbf v}_\ell^\intercal$' method.

Figure 2c also shows the $\bar{\mathbf A}_r = \sum_{\ell=1}^{r} \tilde{\mathbf u}_\ell s_\ell c_\ell \tilde{\mathbf v}_\ell^\intercal$ reconstruction loss, which we show in "Appendix 2" to be the minimum possible loss if reconstructing with the noisy singular vectors. Note that the 'E15' loss nearly overlays on this theoretical minimum.

Reconstruction methods from the prior literature result in higher losses than that achieved herein. Raiola et al. (2015) suggest the $\tilde{\mathbf u}_\ell \tilde s_\ell \tilde{\mathbf v}_\ell^\intercal$ reconstruction at rank $r_F = 12$ (per 6) or $r_{\tilde F} = 19$ (per 7). Epps and Techet (2010) suggest rank $r = 3$. The ELF method can not be used in this example, because the CFD grid was unstructured, and the 2D DCT-II step in the ELF method requires data on a plaid grid.

2.2 Outline

The following sections provide examples that illustrate additional details of the proposed methods. In Sect. 3, we consider analytic examples in order to provide a "controlled environment" within which to present and validate some finer details of the present theory. In Sects. 4 and 5, we consider application to synthetic PIV data, which raises "real world" issues yet still affords us the ability to validate our results using clean data. Finally, in Sect. 6, we demonstrate the methods on real PIV data.

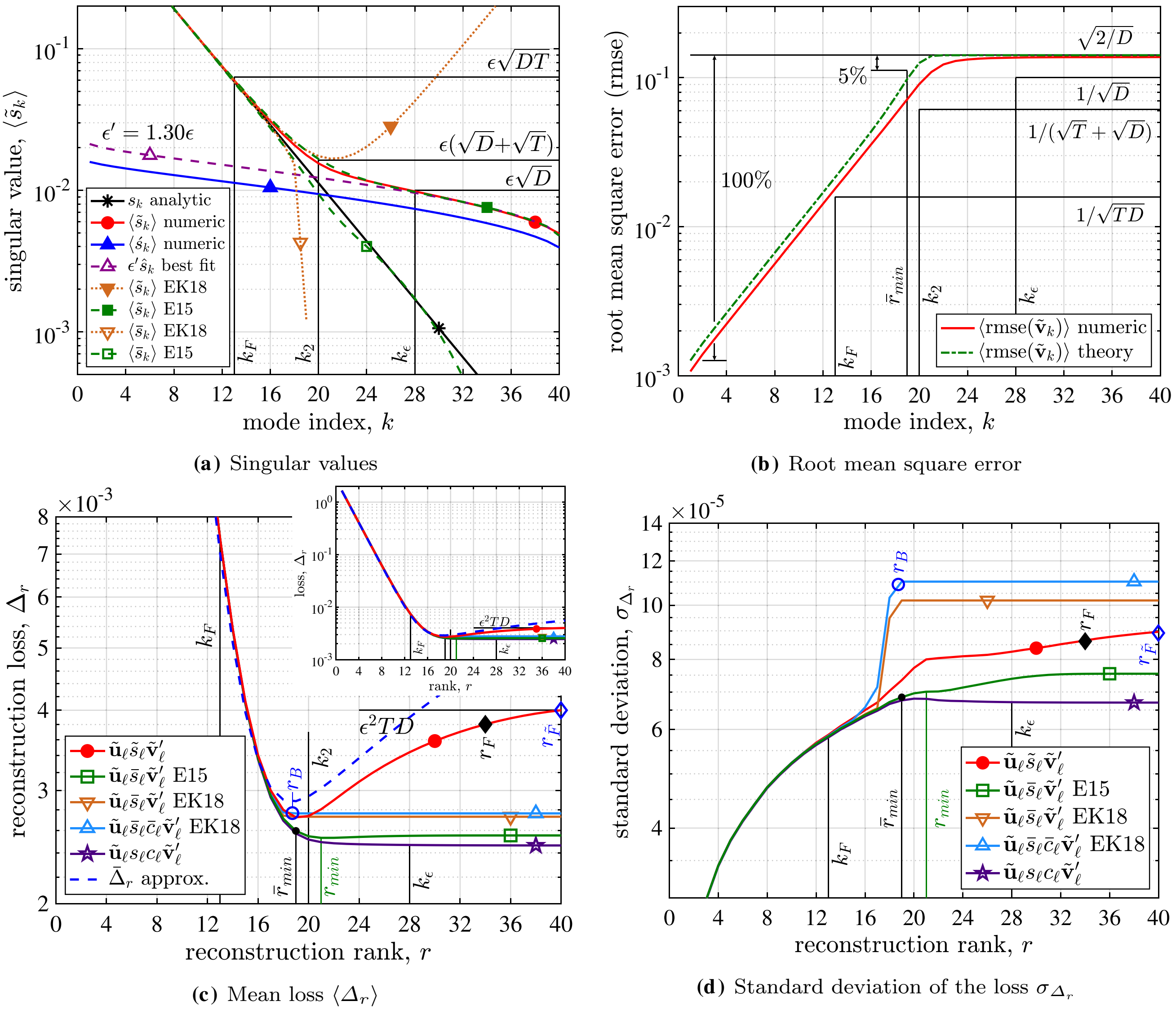
3 Analytic examples and discussion

3.1 Analytic minimal working example

Consider data constructed via $\mathbf A = \mathbf U\,\mathbf S\,\mathbf V^\intercal$, with

$$\begin{aligned} U_{tk} &= \cos\left(\tfrac{2\pi}{T}(t-1)(k-1) - \tfrac{\pi}{4}\right)\sqrt{2/T} \\ S_{kk} &= 10^{-4(k-1)/(T-1)} \\ V_{dk} &= \sin\left(\tfrac{\pi}{D+1} d\,k\right)\sqrt{2/(D+1)}\,, \end{aligned} \tag{18}$$

with $T = 40$ and $D = 100$.



(a) Singular values

(b) Root mean square error

(c) Mean loss $\langle \Delta_r \rangle$

(d) Standard deviation of the loss σ_{Δ_r}

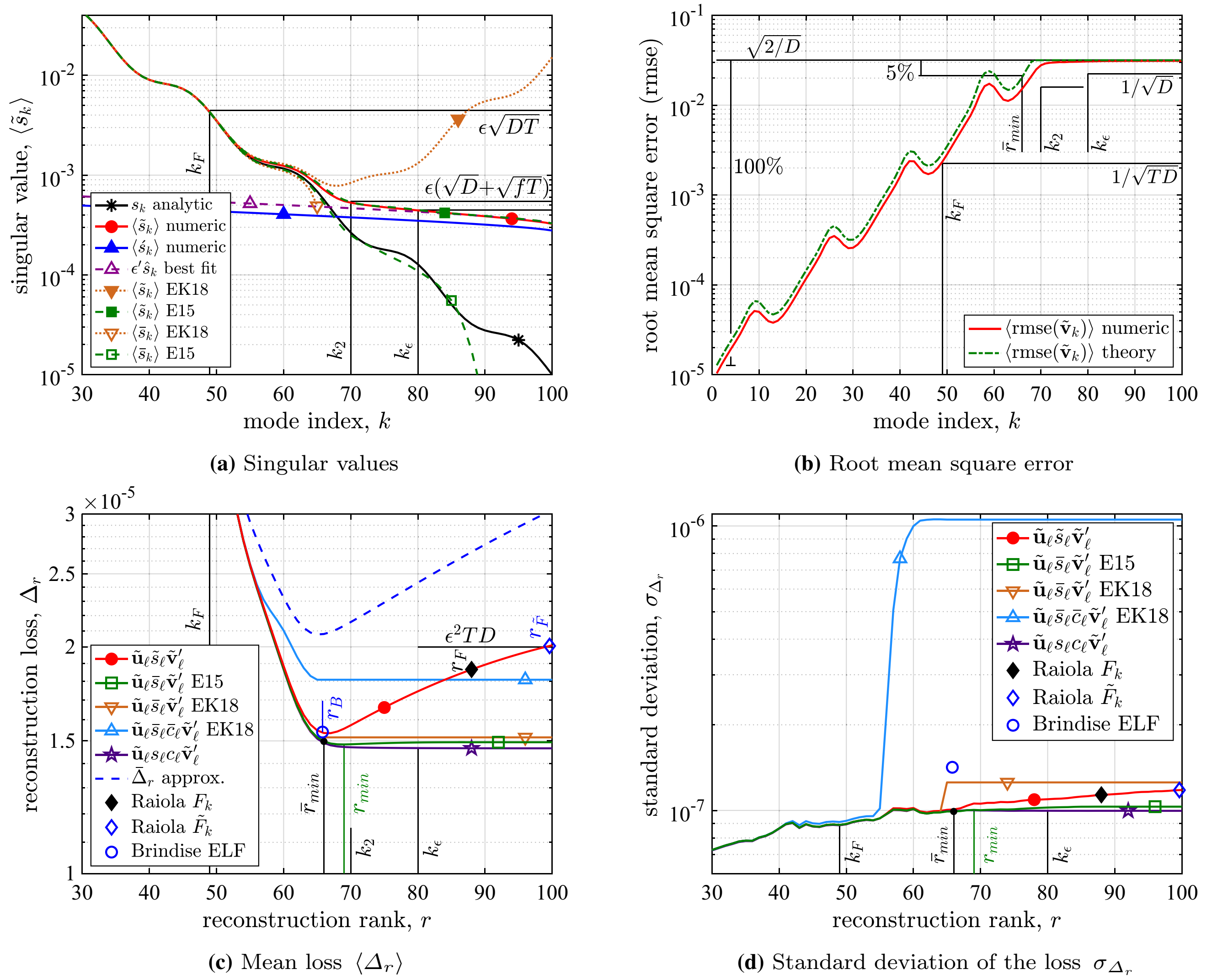
Fig. 3 (Sect. 3.1 example) Monte Carlo results: **a** singular values showing approximation of the noisy singular values $\langle \tilde{s}_k \rangle$ (filled symbols) and reconstruction of the clean singular values $\langle \bar{s}_k \rangle$ (open symbols). **b** Root mean square error of the spatial modes. **c**, **d** Mean and standard deviation of the reconstruction loss. Loss estimate (22) is marked "$\bar{\Delta}_r$ approx.". The symbols marked r_B, r_F, and $r_{\tilde{F}}$ indicate the Brindise ELF method and the Raiola criteria (6) and (7), respectively. Here, $T = 40$, $D = 100$, $\epsilon = 10^{-3}$, $\epsilon' = 1.30\epsilon$, $\epsilon'' = 0.97\epsilon$, and $\bar{\epsilon} = 1.18\epsilon$

In this example, a Monte Carlo simulation of $N = 10{,}000$ trials was used to evaluate the performance of various noise filtering schemes. In each trial, we generated a noisy dataset $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$ (with $\mathbf{E}$ containing *i.i.d.* Gaussian noise with $\epsilon = 10^{-3}$), made estimates $\bar{\epsilon}$, $\bar{s}_k$, and $\bar{\mathbf{A}}_r$, and evaluated the reconstruction loss. Figure 3 shows the mean and standard deviation of the results for the N trials.

3.1.1 Estimates of the clean singular values

First, consider estimates of the clean singular values. Figure 3a shows good agreement between the clean singular values (black curve) and the 'E15' reconstruction $\bar{s}_k$ (16) (green dashed curve). The reason for this good agreement is that the noisy singular values (red curve) are well approximated by $\tilde{s}_k \approx \sqrt{s_k^2 + (\epsilon'\hat{s}_k)^2}$ (green dashed curve, labeled '$\langle \tilde{s}_k \rangle$ E15'), which is the theoretical basis of the 'E15' reconstruction. This approximation works well here, because the clean singular values decay rapidly as compared to the noisy ones.

A more theoretically-sophistocated estimate of the clean singular values will be referred to as the *'EK18' reconstruction*



(a) Singular values

(b) Root mean square error

(c) Mean loss $\langle \Delta_r \rangle$

(d) Standard deviation of the loss σ_{Δ_r}

Fig. 4 (Sect. 3.2 example) Monte Carlo results: **a** singular values showing approximation of the noisy singular values $\langle \tilde{s}_k \rangle$ (filled symbols) and reconstruction of the clean singular values $\langle \bar{s}_k \rangle$ (open symbols). **b** Root mean square error. (c & d) Mean and standard deviation of the loss. Loss estimate (22) is marked "$\bar{\Delta}_r$ approx.". The symbols marked r_B, r_F, and $r_{\tilde{F}}$ indicate the Brindise ELF method and the Raiola criteria (6) and (7), respectively. Here, $T = 100$, $D = 2000$, $\epsilon = 10^{-5}$, $\epsilon' = 1.23\epsilon$, $\epsilon'' = 0.96\epsilon$, and $\bar{\epsilon} = 1.23\epsilon$

$$\bar{s}_k = \begin{cases} \frac{1}{2}\left(\tilde{s}_k + \sqrt{\tilde{s}_k^2 - 2\bar{\epsilon}^2 D}\right) & (k < k_c) \\ 0 & (k_c \le k) \end{cases}, \tag{19}$$

where cutoff index k_c is the minimum k such that $\tilde{s}_k < \max\{\bar{\epsilon}\sqrt{2D}, \bar{\epsilon}(\sqrt{D} + \sqrt{fT})\}$ (Krivitzky and Epps 2017). This cutoff ensures both that (19) yields a real number and that perturbation theory is accurate. Equation (19) follows from perturbation theory, which predicts $\tilde{s}_k \approx s_k + \frac{1}{2}\epsilon^2 D / s_k$ (Epps and Krivitzky 2019). For the Sect. 3.2 example (Fig. 4a), neither reconstruction is perfect, but again the simple 'E15' method (16) works well.

3.1.2 Root mean square error (rmse)

In order to help orient the reader as to the levels of noise in the modes, we find it useful to define the following mode indices:

$$\begin{aligned} k_F &= \min k \text{ such that } \tilde{s}_k < \bar{\epsilon}\sqrt{TD} \\ k_2 &= \min k \text{ such that } \tilde{s}_k < \bar{\epsilon}(\sqrt{D} + \sqrt{fT}) \, . \\ k_\epsilon &= \min k \text{ such that } \tilde{s}_k < \bar{\epsilon}\sqrt{D} \end{aligned} \tag{20}$$

Since rmse($\tilde{\mathbf{v}}_k$) $\approx \bar{\epsilon}/\bar{s}_k$, these three indices correspond to modes with rmse($\tilde{\mathbf{v}}_k$) $\approx 1/\sqrt{TD}, 1/(\sqrt{D} + \sqrt{fT})$, and

$1/\sqrt{D}$, respectively. Moreover, index k_F marks the first mode that fails the (Epps and Techet 2010) criterion (5), k_2 provides a rough approximation for the minimum-loss reconstruction rank, and k_ϵ is the minimum index for which the noisy singular values overlay on the Marchekno–Pastur noise distribution.

Figure 3b shows the root mean square error of the spatial modes, comparing the theoretical prediction (13) to numerical results. Here, the theory slightly over-predicts the rmse, because the rmse is proportional to the noise level $\bar{\epsilon}$, and the estimated $\bar{\epsilon} = 1.18\epsilon$ is too large. Nevertheless, the form of the rmse is correct, and the optimum index $\bar{r}_{\min}$ (15) is predicted within ± 2 indices of what it would be if $\bar{\epsilon}$ was more accurate.

3.1.3 Reconstruction loss

Figure 3c shows the reconstruction loss for the 'E15' method (16)/(17) and the 'EK18' method (19)/(17) with and without $\bar{c}_\ell$ (41). For reference, we also show the $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell s_\ell c_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$ and $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$ methods, which are the best- and worse-case scenario reconstructions. The variable c_ℓ is derived in "Appendix 2", where we show that among all reconstructions that use the noisy singular vectors, the reconstruction that minimizes the loss Δ_r is the one with $\bar{s}_\ell = s_\ell c_\ell$, where

$$c_\ell \equiv \sum_{k=1}^{T} (\tilde{\mathbf{u}}_\ell \cdot \mathbf{u}_k)\,(\tilde{\mathbf{v}}_\ell \cdot \mathbf{v}_k)\, s_k / s_\ell \ . \tag{21}$$

This c_ℓ accounts for the projections of the noisy singular vectors in each of the clean singular vector directions.

The inset of Fig. 3c shows that (for all five methods) the loss is very sensitive to choice of rank for $r < k_2$ but is relatively insensitive to rank for $r > k_2$. The reason for this behavior is that modes $k < k_2$ contain most of the signal content, whereas modes $k > k_2$ contain mostly noise.

Note that for the '$\tilde{\mathbf{u}}_\ell \bar{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$ E15' method, the rank $\bar{r}_{\min} = 19$ (from 15) yields $\Delta_{\bar{r}_{\min}} = 2.60 \times 10^{-3}$, which is just 2.5% larger than the minimum of the 'E15' loss curve $\Delta_{r_{\min}} = 2.53 \times 10^{-3}$, where $r_{\min} = 21$. Also, note that at $\bar{r}_{\min}$ the 'E15' loss is nearly equal to the loss of the '$\tilde{\mathbf{u}}_\ell s_\ell c_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$' method, which is the theoretical minimum.

As shown in Fig. 3d, the 'E15' method also has the least variation in loss, almost as low as the hypothetical '$\tilde{\mathbf{u}}_\ell s_\ell c_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$' method.

3.1.4 Estimate of reconstruction loss

One motivating question of this article is *What is the limit to the accuracy of SVD-based data reconstruction?* We can answer this question mathematically with the following estimate of the loss:

$$\boxed{\bar{\Delta}_r = \bar{\epsilon}^2 D r \ + \sum_{k=r+1}^{T} \bar{s}_k^2}\ , \tag{22}$$

where the $\bar{s}_k$ are the reconstructed singular values (16). Equation (22) is an approximation of the hypothetical loss prediction (61) derived in "Appendix 3". This approximation (22) has the correct asymptotic behavior at $k \approx 1$ and correctly equals $\bar{\epsilon}^2 TD$ at $k = T$ (since $\epsilon^2 TD$ is the loss of the original noisy data). Figure 3c shows this "$\bar{\Delta}_r$ approx." is reasonable; at rank $\bar{r}_{\min}$, Eq. (22) predicts $\bar{\Delta}_{\bar{r}_{\min}} = 2.89 \times 10^{-3}$, which is just 11% higher than the actual 'E15' loss $\Delta_{\bar{r}_{\min}}$.

Clearly, the loss predicted by Eq. (22) has a minimum, since $\sum_{k=r+1}^{T} \bar{s}_k^2$ monotonically decreases with r whereas the $\bar{\epsilon}^2 Dr$ monotonically increases with r. Thus, the minimum $\bar{\Delta}_r$ forms a rough limit to the accuracy of the data reconstruction. More usefully, $\bar{\Delta}_{\bar{r}_{\min}}$ forms an estimate of the actual loss of the 'E15' data reconstruction.[5]

3.1.5 Figure of merit for reconstruction accuracy

A figure of merit for the reconstruction can be formed by comparing the estimated reduction in loss to the loss of the original noisy data:

$$\boxed{M \equiv \frac{\bar{\epsilon}^2 TD - \bar{\Delta}_{\bar{r}_{\min}}}{\bar{\epsilon}^2 TD}}\ . \tag{23}$$

This non-dimensional ratio can be used to gauge the improvement in accuracy.[6]

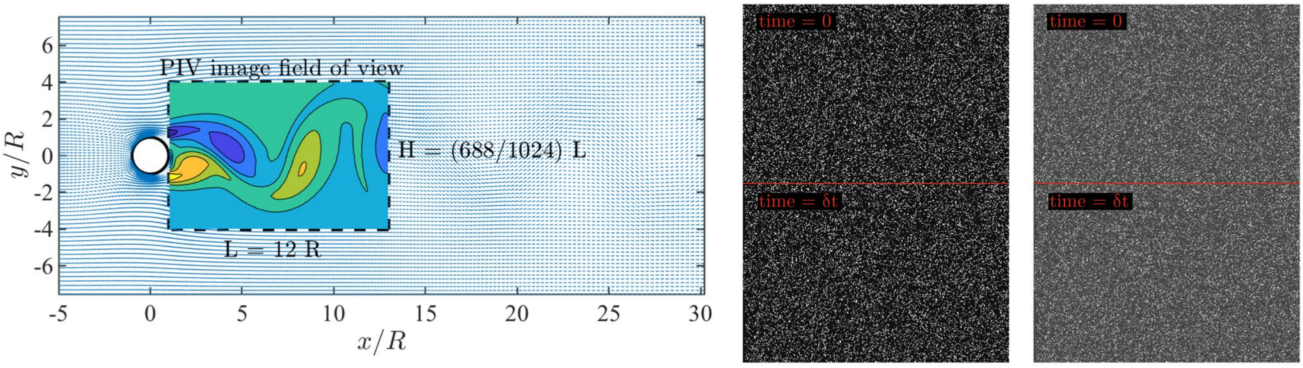
The parameter M can also be loosely interpreted as the fraction of noise that has been removed. For the cylinder example of Sect. 2.1, $M = 98\%$ (also $\bar{\Delta}_{\bar{r}_{\min}} = 0.0034$ and $\Delta_{\bar{r}_{\min}} = 0.0031$). For the present analytic example (Sect. 3.1), $M = 48\%$.

3.1.6 Comparison to literature methods

Other reconstruction methods were examined as well. The (Brindise and Vlachos 2017) ELF method on average yields low loss, $\langle \Delta_r \rangle = 2.77E-3$, although it has a much larger variability in loss ($\sigma_{\Delta_r} = 1.09E-4$) than the 'E15' method (see Fig. 4c, d). The reason for this larger variability is that the ELF method has a larger variability in rank ($\sigma_{r_B} = 0.905$) than the 'E15' method ($\sigma_{\bar{r}_{\min}} = 0.494$), so the ELF method

[5] Evidently, one could choose $\bar{r}_{\min}$ as the index for which $\bar{\Delta}_r$ (22) is a minimum. In the present examples, this scheme yields a $\bar{r}_{\min}$ within ± 2 indices of the $\bar{r}_{\min}$ from (15).

[6] Values of M less than zero indicate that the reconstruction is less accurate than the original noisy data.



(a) PIV image field of view

(b) Clean image pair

(c) Noisy image pair

Fig. 5 (Sect. 4 cylinder PIV example) Generation of synthetic PIV datasets from clean CFD data

more often chooses the rank too low, resulting in large losses.

The (Raiola et al. 2015) F and $\tilde{F}$ methods yield significantly higher ranks and losses than the present ‘E15’ method. For this example, the $\tilde{F}_k$ method always predicted $\tilde{r}_F = 40$, because the $\tilde{F}_k$ parameter was always less than 0.999. The reason that $\tilde{F}_k < 0.999$ here is because $\tilde{F}_k$ asymptotes to 1 only when the data matrix has a very high aspect ratio ($D \gg T$) such that the tail of the noisy singular values is nearly constant.

Not shown in the figures are the results of the (Shabalin and Nobel 2013) method, since they nearly overlay on the present ‘$\tilde{\mathbf{u}}_\ell \bar{s}_\ell \bar{c}_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$ EK18’ method, although their theoretical development and final equations differ from those presented herein.

3.2 Effect of the distribution of singular values

In order to illustrate the effect of the distribution of singular values, consider constructing the analytic data using the singular vectors from (18) but now using the following singular values:

$$\log_{10}(s_k) = -5\frac{k-1}{T-1} - 0.1\sin\left(12\pi\frac{k-1}{T-1}\right), \tag{24}$$

The noise data were constructed by first drawing from a normal distribution with standard deviation $\sqrt{w\epsilon}$ and then performing uniform spatial smoothing over a window of width w. This two-step process yields spatially-correlated noise with standard deviation ϵ. Results of a Monte Carlo simulation with $\epsilon = 10^{-5}$, $w = 5$, $T = 200$, $D = 2000$, and $N = 1000$ are shown in Fig. 4.

The results in Fig. 4 generally agree with those of in Fig. 3. One key difference in the results of these examples is that here the ‘$\tilde{\mathbf{u}}_\ell \bar{s}_\ell \bar{c}_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$ EK18’ reconstruction has unacceptably poor performance in both loss (Fig. 4c) and variability (Fig. 4d). The reason is that the estimated $\bar{c}_\ell$ attenuates modes 56 and higher much more than the exact c_ℓ, because the root mean square errors of these modes are large (see Fig. 4b). Since using $\bar{c}_\ell$ might lead to a *more* lossy reconstruction than the original noisy data, we recommend *not* using $\bar{c}_\ell$ in data reconstruction.

Although the two ‘EK18’ methods have a more theoretically-sophisticated foundation than the ‘E15’ method, they again are found to be less accurate and more variable than the ‘E15’ method; therefore, the ‘EK18’ methods are *not* recommended.

Figures 3 and 4 show that the ‘E15’ method (15)/(16)/(17) yields the least mean loss $\langle \Delta_r \rangle$ and the least variation σ_{Δ_r} of all the practical methods shown. Moreover, the loss from the ‘E15’ method is very close to that from the theoretical minimum-loss method ‘$\tilde{\mathbf{u}}_\ell s_\ell c_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$’. Here, the optimum reconstruction rank is predicted to be $\bar{r}_{\min} = 66$, while the actual minimum of the ‘E15’ loss curve occurs at $r_{\min} = 69$. However, the ‘E15’ loss curve is very flat, and the loss $\Delta_{\bar{r}_{\min}} = 1.50 \times 10^{-5}$ is just 0.9% higher than $\Delta_{r_{\min}} = 1.48 \times 10^{-5}$.

Again, Eq. (22) provides an acceptable estimate of the loss. Here, (22) predicts $\bar{\Delta}_{\bar{r}_{\min}} = 2.08 \times 10^{-5}$, about 40% larger than the actual $\Delta_{\bar{r}_{\min}}$. Here, the reconstruction merit (23) is $M = 32\%$.

4 Synthetic PIV: cylinder flow

4.1 Generation of datasets

In this section, we consider a synthetic PIV data set generated using Sect. 2.1 CFD data. As illustrated in Fig. 5, a clean image pair was created for each time step, with particles placed randomly in the initial image, and then advected with the local flowfield (as interpolated from the clean CFD data) for the second image. The noisy image color intensity was obtained by reducing that of the corresponding clean image and adding uniformly-distributed noise to each pixel. PIV processing was performed in DaVis 8.3.1, using several multi-pass approaches as summarized in Table 1. For the figures, PIV processing was performed with an initial interrogation window size of 64 × 64 px with 50% overlap, then three passes on a final window size of 32 × 32 px with 75% overlap. Since the noisy data at the edge of the PIV domain have non-Gaussian error, these data were cropped prior to SVD analysis. The resulting vector field contained 126 × 84 vectors ($D = 21,168$) and $T = 455$ timesteps. Further details regarding generation and processing of the synthetic PIV images are given in (Epps and Krivitzky 2019).

Here, we introduce the following notation: $\mathbf{A}$ = clean CFD data, $\mathbf{A}^*$ = clean PIV data, $\tilde{\mathbf{A}}^*$ = noisy PIV data, and $\bar{\mathbf{A}}^*$ = reconstructed PIV data (which should agree with $\mathbf{A}^*$). We also redefine the *reconstruction loss* from (4) as $\Delta_r \equiv \|\mathbf{A}^* - \bar{\mathbf{A}}_r^*\|_F^2$.

4.2 Results overview

As in previous examples, the proposed methods prove successful:

- The error estimation procedure of "Appendix 1" produces an accurate estimate $\bar{\epsilon} = 1.85 \times 10^{-5}$ of the actual rms error $\epsilon = 1.90 \times 10^{-5}$, which is the rms difference between the clean and noisy PIV data.
- Equation (16) yields an accurate reconstruction of the clean singular values for $k < \bar{r}_{\text{min}}$, which are the important modes (see Fig. 6a).
- The 'E15' noise filtering method (16)/(17) yields losses nearly as low as the theoretical minimum-loss '$\tilde{\mathbf{u}}_\ell s_\ell c_\ell \tilde{\mathbf{v}}_\ell^{\mathsf{T}}$' curve (see Fig. 6c).
- The Brindise method yields rank and loss similar to those of the present approach, whereas the Raiola methods yield much higher rank and loss.
- The estimated loss $\bar{\Delta}_{\bar{r}_{\text{min}}} = 3.0 \times 10^{-4}$ from (22) is within a factor of 2 of the actual $\Delta_{\bar{r}_{\text{min}}} = 1.6 \times 10^{-4}$.
- Here, the reconstruction merit (23) is $M = 91\%$.

Figure 7 shows a snapshot of the vorticity field of the clean PIV data, noisy PIV data, 'E15' reconstruction at $\bar{r}_{\text{min}} = 16$, and the standard '$\tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathsf{T}}$' reconstruction at $r = k_F - 1 = 7$. Note that while the original noisy-data vorticity field contains significant noise, the character of the vorticity field is restored by the 'E15' reconstruction. For example, the 'E15' reconstruction recovers the thin blue 'stripe' in the middle of the frame and the green 'spike' near the right side. Choosing a lower rank, such as $r = 7$, results in additional smoothing but loss of some of these fine details of the flowfield.

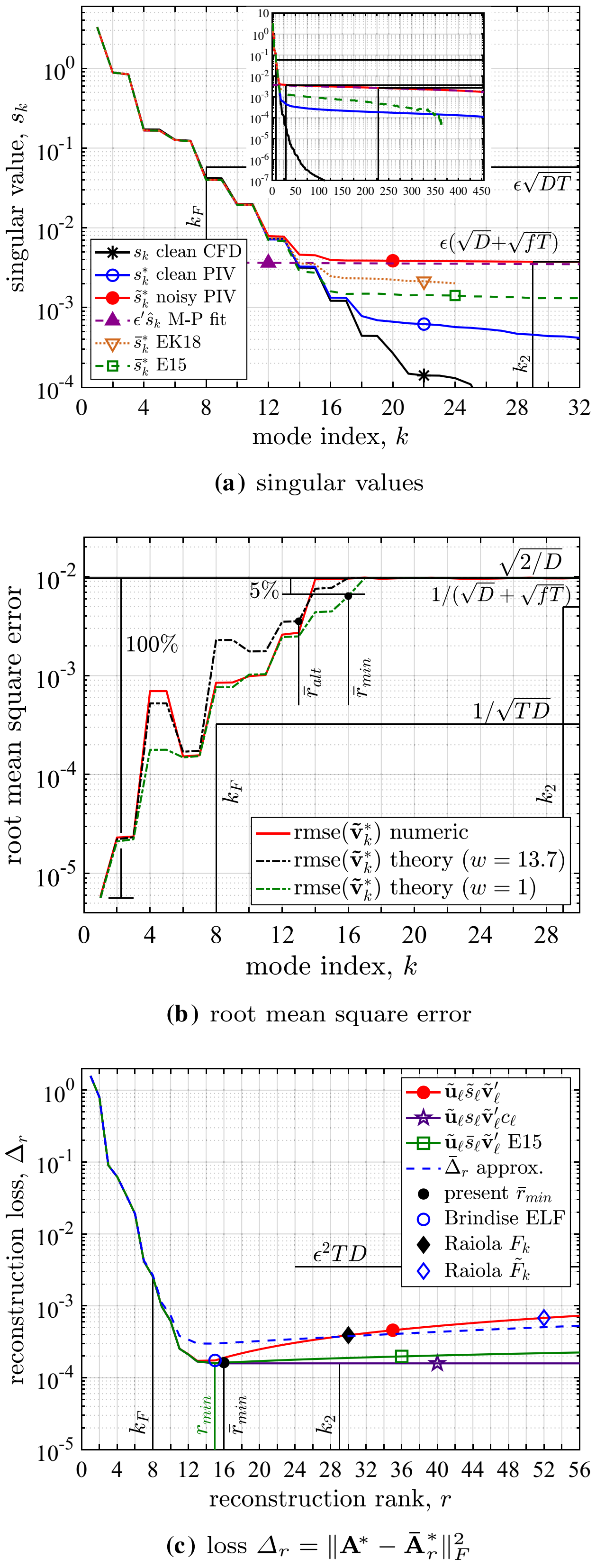
4.3 On the use of $w = 1$ to predict $\bar{r}_{\text{min}}$

The important new feature of this example is that it uses PIV data, which contain spatially-correlated noise. Consequently, the noisy singular values in Fig. 6a are best fit by at Marchenko–Pastur distribution that has spatial-correlation parameter $f > 1$ (here, $f = 7.1$). Inserting this $f = 7.1$ into (12) yields an effective smoothing window width of $w = 13.7$, which is reasonable considering that the data were smoothed three times using the 9 nearest

Table 1 (Sect. 4 cylinder PIV example) Results for selected PIV processing schemes

PIV processing approach	$\epsilon \times 10^5$	$\bar{\epsilon} \times 10^5$	r_{min}	$\bar{r}_{\text{min}}$	r_B	r_F	$r_{\tilde{F}}$	k_2	$\Delta_{r_{\text{min}}} \times 10^3$	$\Delta_{\bar{r}_{\text{min}}} \times 10^3$	$\Delta_{r_B} \times 10^3$	$\Delta_{r_F} \times 10^3$	$\Delta_{r_{\tilde{F}}} \times 10^3$
○ 32 px 75% → 8 px 0%	6.13	6.08	13	13	13	14	32	21	1.16	1.16	1.30	1.41	3.22
○ 32 px 75% → 16 px 50%	3.22	3.19	15	15	13	25	27	42	0.75	0.75	0.77	1.12	1.19
○ 32 px 75% → 32 px 75%	1.72	1.72	19	19	15	50	59	49	0.47	0.47	0.48	0.80	0.88
○ 64 px 50% → 8 px 0%	6.06	6.08	13	13	13	14	21	14	1.10	1.10	1.23	1.33	2.03
○ 64 px 50% → 16 px 50%	3.06	3.01	15	15	13	24	56	35	0.35	0.35	0.37	0.67	1.55
○ 64 px 50% → 32 px 75%	1.43	1.41	17	21	13	56	58	54	0.27	0.28	0.28	0.57	0.58
○ 64 px 50% → 3 × (32 px 75%)	**1.90**	**1.85**	**15**	**16**	**15**	**30**	**52**	**29**	**0.16**	**0.16**	**0.17**	**0.39**	**0.68**
□ 32 px 75% → 8 px 0%	6.58	6.48	13	13	13	16	19	14	2.87	2.87	3.04	3.38	3.72
□ 32 px 75% → 16 px 50%	3.77	3.74	15	13	13	27	34	36	1.04	1.04	1.07	1.60	1.88
□ 32 px 75% → 32 px 75%	2.09	2.30	378	17	13	49	90	41	1.07	1.24	1.27	1.59	2.01
□ 64 px 50% → 8 px 0%	6.51	6.40	13	12	13	18	24	14	2.78	2.80	2.94	3.50	4.16
□ 64 px 50% → 16 px 50%	3.61	3.56	15	13	13	24	27	25	0.52	0.52	0.54	0.93	1.05
□ 64 px 50% → 32 px 75%	1.71	1.71	16	17	13	50	52	60	0.45	0.45	0.46	0.76	0.78
□ 64 px 50% → 3 × (32 px 75%)	2.14	2.10	15	15	15	27	32	29	0.20	0.20	0.22	0.43	0.51

ϵ actual rms error between noisy and clean PIV data, $\bar{\epsilon}$ error estimate ("Appendix 1"), r_{min} *and* $\Delta_{r_{min}}$ actual rank and value of minimum 'E15' loss, $\bar{r}_{min}$ *and* $\Delta_{\bar{r}_{min}}$ estimated optimum rank (15) [with $w = 1$ in (13)] and resulting loss, r_B *and* Δ_{r_B} Brindise ELF rank and loss, r_F *and* Δ_{r_F} Raiola F rank and loss, $r_{\tilde{F}}$ *and* $\Delta_{r_{\tilde{F}}}$ Raiola $\tilde{F}$ rank and loss



(a) singular values

(b) root mean square error

(c) loss $\Delta_r = \|\mathbf{A}^* - \bar{\mathbf{A}}_r^*\|_F^2$

Fig. 6 (Sect. 4 cylinder PIV example) Singular values, root mean ▶ square error, and reconstruction loss. In **a**, threshold levels are shown with $\epsilon = 1.90 \times 10^{-5}$, and the M–P fit is made with $\epsilon' = 1.87 \times 10^{-5}$ and $f = 7.1$. In **b**, $\bar{r}_{\min} = 16$ is the index for which the rmse is within 5% of the ceiling $\sqrt{2/D}$. In **c**, $r_{\min} = 15$ is the index for minimum '$\tilde{\mathbf{u}}_\ell \bar{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$ E15' loss

neighbors. Also, the data contain spatial correlation due to being acquired in overlapping interrogation windows.

A question arises as to whether the value $w = 13.7$ should be used to estimate the root mean square error (rmse) of the SVD modes (and subsequently the optimal reconstruction rank). Figure 6b compares the rmse predictions (13) using both $w = 13.7$ and $w = 1$, as well as the numerically-computed rmse (which is that between the noisy and clean PIV modes). The optimal reconstruction rank (15) predicted using the $w = 1$ rmse is marked $\bar{r}_{\min} = 16$, and an alternative rank (predicted using the $w = 13.7$ rmse) is marked $\bar{r}_{\mathrm{alt}} = 13$. In this case, both ranks are nearly the same as the actual rank $r_{\min} = 15$ that yields minimum 'E15' loss. Moreover, the difference in loss between these three ranks is negligible, so use of $w = 13.7$ or 1 is somewhat of a moot point in this example.

However, in other examples (Sects. 5, 6), we have found that using the rmse predicted with $w = 1$ yields a more accurate estimate of the optimal reconstruction rank. Using $w = 1$ drives the rmse down, which drives the $w = 1$ rank $\bar{r}_{\min}$ lower than the $w > 1$ rank $\bar{r}_{\mathrm{alt}}$. Considering the shape of the loss curve (see Fig. 6c), it is much better to estimate the rank too high than too low. Therefore, it is recommended that $w = 1$ should be used to predict rmse and subsequently $\bar{r}_{\min}$.

4.4 Effect of PIV processing scheme

Another question arises as to the efficacy of the proposed noise filtering method with respect to different PIV processing schemes. Table 1 compares results for several PIV processing approaches. Interestingly, the minimum loss $\Delta_{r_{\min}}$ decreases as the final interrogation window size *increases*. This result is explained by the fact that $\Delta_{r_{\min}}$ is some fraction of $\epsilon^2 TD$, and ϵ also decreases with increasing interrogation window size.

For all cases in Table 1, the tail fit method of "Appendix 1" produces an accurate estimate $\bar{\epsilon}$ of the actual error ϵ, and the rank $\bar{r}_{\min}$ (predicted using $w = 1$) is very close to the actual optimum rank $r_{\min}$. Moreover, the resulting loss $\Delta_{\bar{r}_{\min}}$ is nearly as low as the minimum loss $\Delta_{r_{\min}}$ for most cases. Again, the Brindise ELF method yields slightly higher losses, and the Raiola F and $\tilde{F}$ methods yield significantly higher losses.

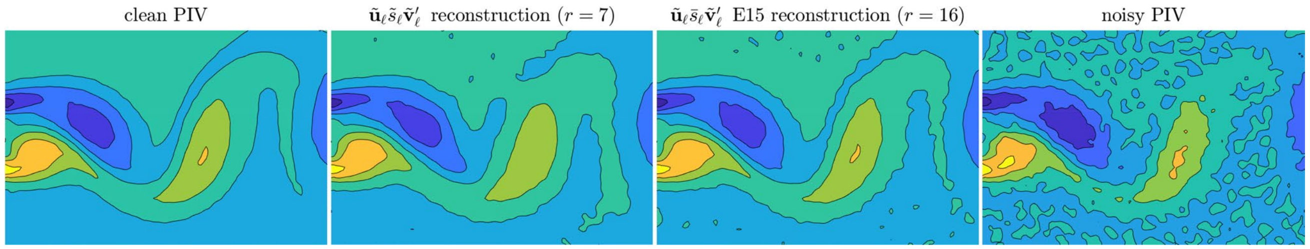


Fig. 7 (Sect. 4 cylinder PIV example) Vorticity field for a single timestep, illustrating the 'clean PIV' and 'noisy PIV' data, as well as two reconstructions: the $\tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^\mathsf{T}$ reconstruction at rank $r = k_F - 1 = 7$, and the $\tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^\mathsf{T}$ E15 reconstruction at rank $\bar{r}_{\min} = 16$

5 Synthetic PIV: channel flow

In this example, we consider synthetic PIV of a much more complex flowfield, fully-developed turbulent channel flow. This example provides further validation of the proposed methods, and it allows us to illustrate the effects of sampling timestep and number of snapshots on the proposed methods.

5.1 Generation of data sets

Synthetic PIV images were generated as in Sect. 4, with particles placed randomly in the initial image and advected with the local flowfield for the second image. Here, the flowfield was interpolated from direct numerical simulation (DNS) data of turbulent channel flow (Kim et al. 1987; Lee et al. 2013) obtained from the Johns Hopkins turbulence database (Li et al. 2008; Graham et al. 2016). The image size was 1024 × 1024 px, with 20,480 particles ($\approx$ 20 particles per 32 × 32 px interrogation window). Image pairs were generated corresponding to DNS domain $0 \le x \le 1, -1 \le y \le 0$ and the first 2000 timesteps of the DNS database.

PIV processing was performed in DaVis 8.3.1, using an initial interrogation window size of 64 × 64 px with 50% overlap, then three passes on a final window size of 32 × 32 px with 75% overlap. After cropping the edges, the resulting vector field contained 126 × 126 vectors ($D = 31,752$) and 2000 timesteps.

5.2 Baseline case (δt, $T = 500$)

In this subsection, we consider only the first $T = 500$ timesteps of PIV data (spaced δt apart, where δt is the DNS database timestep). Here, the rms error between the noisy and clean PIV data is $\epsilon = 0.0105$, and the Marchenko–Pastur tail fit predicts $\bar{\epsilon} = 0.0109$ and $f = 7.9$ ($w = 15.3$).

The key difference between this turbulent flow example and the previous ones is that here, the clean singular values decay much more slowly. As a result, there is less separation between the clean and noisy PIV singular values (see Fig. 8a). For example, here $\tilde{s}^*_T / s^*_T = 7.7$, whereas in the PIV cylinder example $\tilde{s}^*_T / s^*_T = 15.3$, and in the analytic examples of Sects. 3.2 and 3.1, $\tilde{s}_T / s_T = 32$ and 49, respectively.

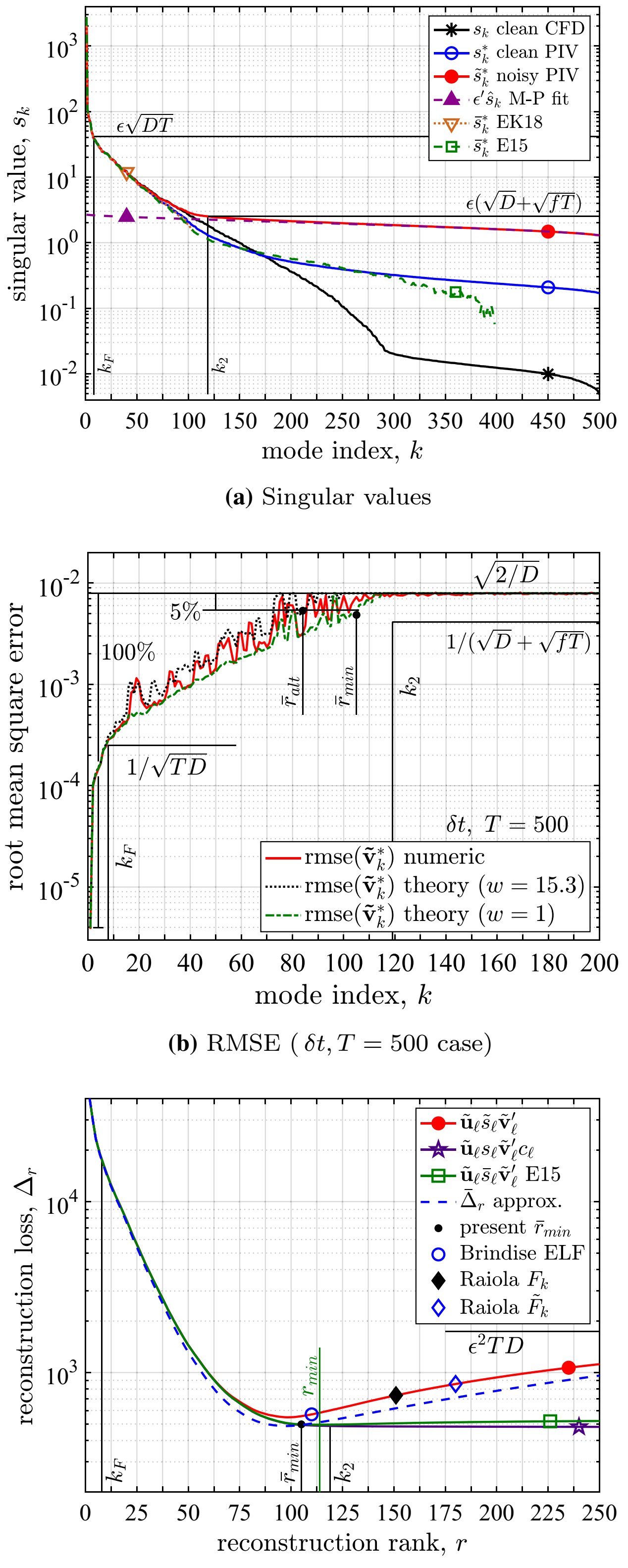
This reduced separation between the clean and noisy singular values makes it challenging to correctly estimate the error level $\bar{\epsilon}$ and effective smoothing window width w.[7] As a result, the rmse predictions end up too large when using $w > 1$ (c.f. compare the $w = 15.3$ theory curve and the numeric rmse curve in Fig. 8b). Instead, the rmse should be estimated using $w = 1$, which provides a conservative estimate of the rmse that is useful for predicting the optimum reconstruction rank $\bar{r}_{\min}$.

Figure 8c shows the reconstruction loss for this channel flow case: The estimated reconstruction rank $\bar{r}_{\min} = 105$ yields a loss of $\Delta_{\bar{r}_{\min}} = 497$, which is nearly as low as the minimum 'E15' loss $\Delta_{r_{\min}} = 495$ (at $r_{\min} = 114$). As in previous examples, the Brindise ELF method has slightly higher loss ($\Delta_{r_B} = 569$ with $r_B = 110$), and the Raiola methods have higher losses ($\Delta_{r_{\tilde{F}}} = 857$ at $r_{\tilde{F}} = 180$). Equation (22) again provides a very good loss estimate: $\bar{\Delta}_{\bar{r}_{\min}} = 494$ (just 0.8% lower than $\Delta_{\bar{r}_{\min}}$). The reconstruction merit (23) is $M = 74\%$.

5.3 Effect of timestep and number of snapshots

Two important questions to address at this point are: (1) how do the various noise filtering methods perform as the separation between the noisy and clean singular values varies?, and (2) how do the timestep and number of frames in the dataset affect the singular values? To address these questions, we performed the SVD on six datasets subsampled with various

[7] Consider the limit of $\epsilon \to 0$, where the noisy singular values would overlay on the clean ones. If the clean singular values were to decay slowly, as in Fig. 8a, then a Marchenko–Pastur distribution would be able to be fit, and some $\bar{\epsilon} > 0$ and $w > 1$ would erroneously be predicted.



(a) Singular values

(b) RMSE ($\delta t, T = 500$ case)

(c) Reconstruction loss ($T = 500$)

Fig. 8 (Sect. 5 Channel flow PIV example) Singular values, rms error, and reconstruction loss

timesteps (δt, $2\delta t$, and $4\delta t$, where δt is the DNS database timestep) and number of snapshots ($T = 500, 1000, 2000$).

Figure 9a shows the clean singular values for these six cases. When plotted versus normalized mode index k / T, the cases with the same timestep (but different T) overlay, indicating that the number of snapshots T has little effect on the clean singular values. Rather, the shape (slope and magnitude) of the tail of the clean singular values is dictated by the timestep.

By contrast, the noisy singular values are relatively insensitive to the timestep (Fig. 9b), so a larger timestep causes there to be less separation between clean and noisy singular values.

Figure 9c shows the loss for the three $T = 500$ cases (with the δt case carried over from Sect. 5.2). As expected, a larger timestep results in more loss, and less of a "bucket" below the noisy dataset loss $\epsilon^2 TD$ within which to advantageously filter noise. Consequently, the reconstruction merit (23) decreases with increasing timestep; for the δt, $2\delta t$, and $4\delta t$ cases, we find $M = 74\%$, 54%, and 9%, respectively. In all three cases, the present methods yield reasonable predictions of the optimal reconstruction rank $\bar{r}_{\text{min}}$. The ELF method performs well at δt and $2\delta t$ but not $4\delta t$. In contrast, the Raiola method performs best at $4\delta t$.

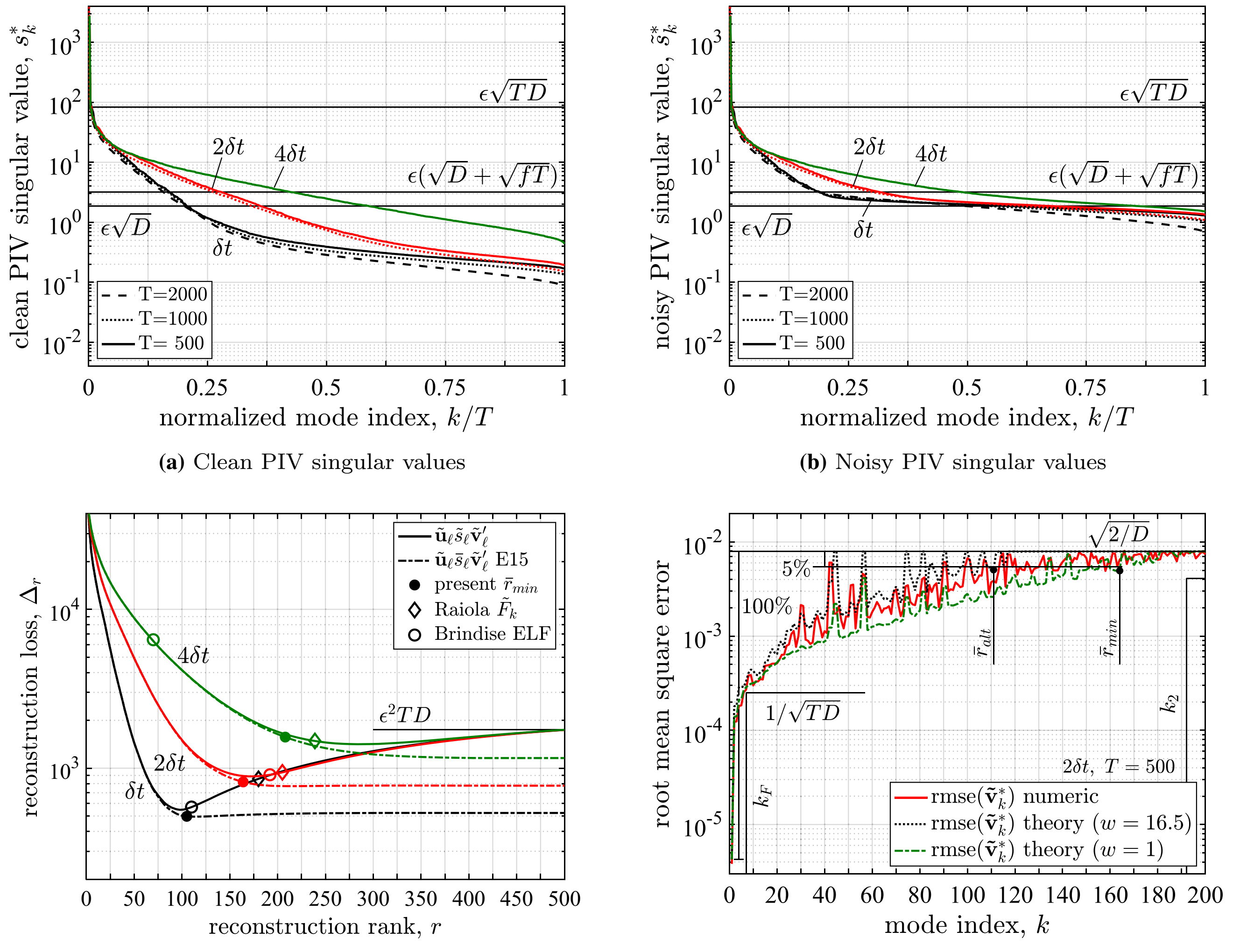
For reference, the rmse for the $\{2\delta t, T = 500\}$ case are shown in Fig. 9d. Here, the effect of using $w > 1$ versus $w = 1$ is very pronounced, with the $w = 1$ rmse predictions yielding a much better optimal rank estimate $\bar{r}_{\text{min}} = 164$. The $w > 1$ rmse predictions yield $\bar{r}_{\text{alt}} = 111$, which results in significantly more loss than using $\bar{r}_{\text{min}}$.

5.4 Comment regarding optimal rank criterion

Some readers might find it controversial to use the root mean square error of the spatial modes (rmse($\tilde{\mathbf{v}}_k$)) in order to predict the optimal reconstruction rank. The reasoning is that modes with very similar energy might switch order, so their computed rmse might be very large, even if the modes themselves did not change.

Regarding modes with similar energy (i.e. small gap between neighboring singular values), there are two important cases to consider: First, consider "paired modes", which are modes with similar energy that are well below the noise ceiling. Equation (13) shows that the rmse of each of these modes will indeed be higher than the rmse of "isolated modes" (i.e. those with well-separated singular values). The reason is that paired modes can and do mix upon perturbation. However, *Wedin's theorem* ensures that the subspace spanned by these modes, as a group, is only slightly perturbed. Thus, paired modes below the noise ceiling should be and are used in our data reconstruction method.

The second case to consider is "noise modes", which are modes with rmse at the noise ceiling (rmse($\tilde{\mathbf{v}}_k$) $\approx \sqrt{2/D}$).



(a) Clean PIV singular values

(b) Noisy PIV singular values

(c) Reconstruction loss ($T = 500$ cases)

(d) RMSE ($2\,\delta t, T = 500$ case)

Fig. 9 (Sect. 5.3 Channel flow PIV example) Effect of timestep and number of snapshots

These modes have similar energy, since their singular values fall in the tail of the distribution. In (Epps and Krivitzky 2019), we show that modes with $\mathrm{rmse}(\tilde{\mathbf{v}}_k) \approx \sqrt{2/D}$ constitute random noise. Thus, these modes should be discarded for noise filtering.

A subtle point is that Eq. (15) keeps all the paired modes and only discards the noise modes, because $\bar{r}_{\min}$ is set as the "maximum k such that $t_k > 5\%$". In other words, there might be some paired modes with $t_k > 5\%$, but as long as there are some isolated modes with $t_k < 5\%$ for larger k, then those paired modes are kept. For example, note in Fig. 9d that modes $k = 134$–135 and 142–143 have rmse above the 5% cutoff, but these paired modes are kept, because there are other isolated modes between those and $\bar{r}_{\min} = 164$, which is the last mode with rmse below the 5% cutoff.

6 Laminar jet PIV example

In this final example, we apply the present noise filtering methods to a real PIV dataset. Here, we consider a publicly-available PIV dataset that captures the Kelvin–Helmholz instabilities of a laminar jet in a quiescent fluid (Neal et al. 2015). This set of PIV images is unique in that it is accompanied by 'clean' vector fields that have been obtained using high-dynamic-range PIV (PIV-HDR), which has higher accuracy than standard PIV because it uses multiple cameras with very high image resolution (Neal et al. 2015).

6.1 Generation of datasets

A sample PIV image is shown in Fig. 10; the red solid rectangle highlights the cropped region used for PIV processing, and the green dash-dot rectangle represents the region where the PIV-HDR data are available. Since the PIV and

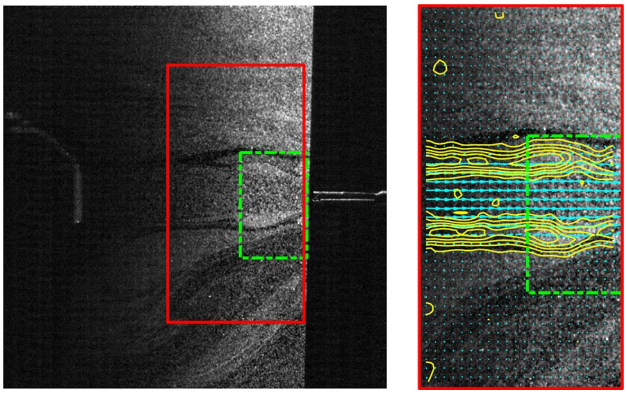


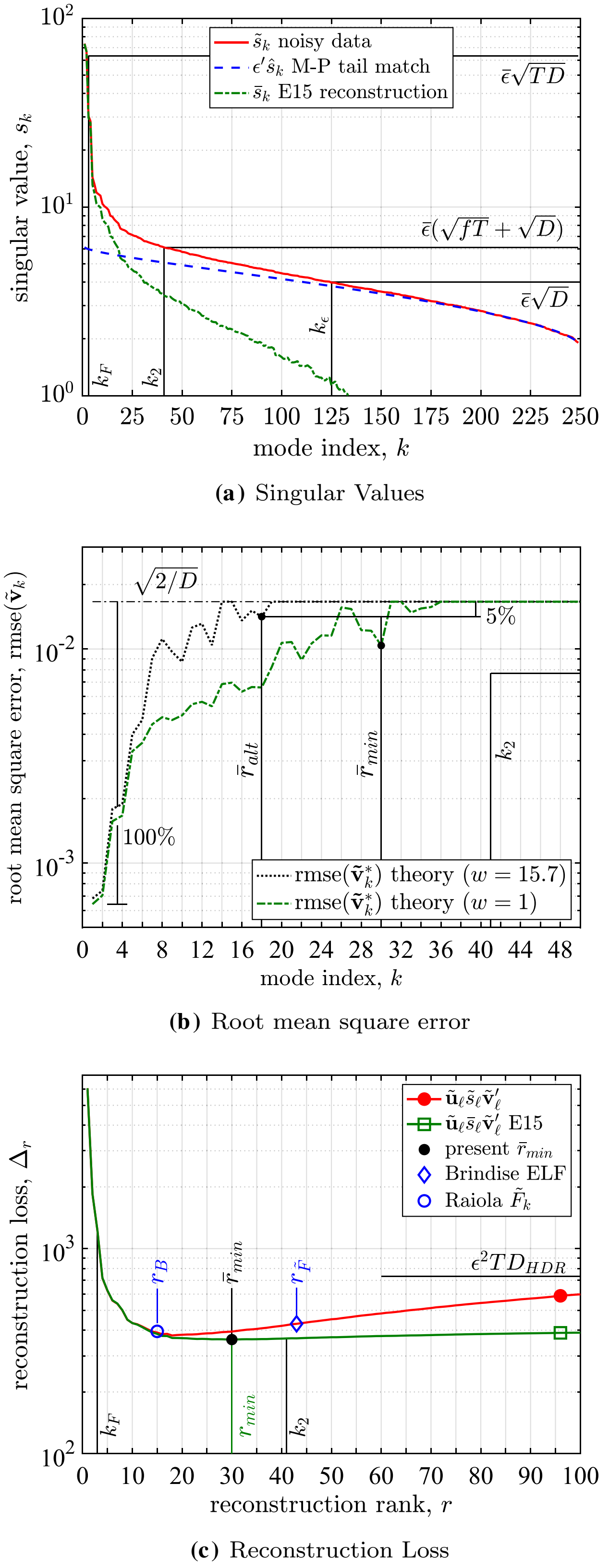
Fig. 10 (Sect. 6 Laminar Jet PIV example). (Left) Sample PIV image (Neal et al. 2015), with cropped region used for PIV processing (red box), and region where the PIV-HDR data are available (green dash-dot box). (Right) Resulting PIV velocity vectors (every fourth vector shown) and vorticity contours

PIV-HDR data are not available on the same grid, the SVDs of these two datasets cannot be compared directly, and the reconstruction loss must be computed only over the PIV-HDR region. For the reconstruction loss, the PIV-HDR data are down-sampled to the standard PIV grid locations. PIV processing was performed using DaVis 8.3.1, with a final window size of 16 × 16 px with 75% overlap, leading to a 43 × 84 vector grid. Figure 10 displays a sample vector field and corresponding vorticity contours; flow is left to right.

6.2 Results and discussion

Similar to the previous example, singular values of the noisy data (Fig. 11a) are very-well fit by a Marchenko–Pastur distribution with $f > 1$ (here $f = 8.1$, $w = 15.7$). In this example, a long ‘transition region’ $k_F < k < k_\epsilon$ exists, wherein the modes transition from fairly clean ($k < k_F = 3$) to complete noise ($k > k_\epsilon = 125$). With the ‘E15’ reconstruction (16), the singular values in this transition region are reduced. Although ‘clean’ data exists, their singular values are not directly comparable to those of the PIV dataset due to the different region sizes, so ‘clean’ singular values are not shown.

Reconstruction loss (Fig. 11c) was determined by sampling both the ‘clean’ PIV-HDR data and the PIV reconstruction datasets at points within the HDR region. The results generally follow the previous examples: The standard ‘$\tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^\intercal$’ reconstruction has a minimum (here, $\Delta_r = 377$ at $r = 18$) and then rises for higher ranks. The ELF method selects $r_B = 15$ modes for reconstruction, resulting in a loss of $\Delta_{r_B} = 395$. The Raiola $\tilde{F}$ method selects rank $r_{\tilde{F}} = 43$, yielding loss $\Delta_{r_{\tilde{F}}} = 430$. The ‘E15’ reconstruction selects rank $\bar{r}_{\min} = 30$ and has loss $\Delta_{\bar{r}_{\min}} = 360$, which happens to



(a) Singular Values

(b) Root mean square error

(c) Reconstruction Loss

Fig. 11 (Sect. 6 Laminar Jet PIV example) Singular values, rmse, and reconstruction loss

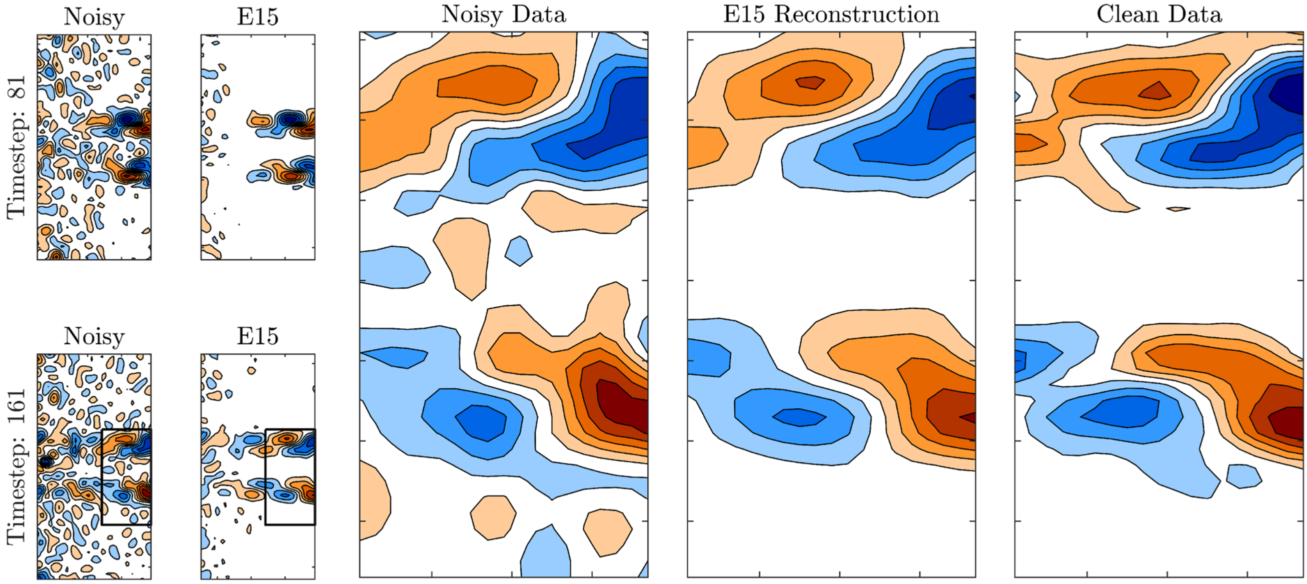


Fig. 12 (Sect. 6 Laminar Jet PIV example) Snapshots of vorticity (with time-mean flow removed). The small panels show the noisy data and '$\tilde{\mathbf{u}}_\ell \bar{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathrm{T}}$ E15' reconstruction ($\bar{r}_{\min} = 30$) at two times ($t = 81$ and 161). The large panels blow up the PIV-HDR region at time $t = 161$, showing the noisy PIV flowfield, E15 reconstruction, and 'clean' PIV-HDR data

be the minimum E15 loss. Of the three methods, again the E15 reconstruction has the minimum loss. The Raiola F method was not applicable, because the clean singular values were unknown for this real PIV dataset. The loss estimate (22) and reconstruction merit (23) also are not applicable here, because the loss is not computed over the entire PIV domain.[8]

Snapshots of vorticity for the noisy data compared to the 'E15' reconstruction (at $\bar{r}_{\min} = 30$) are shown on the left of Fig. 12 at two different times for the entire PIV region. The time-mean velocity field has been removed, revealing significant noise in the measurements of unsteady components. The E15 reconstruction filters out much of the noise, leaving a much clearer view of the important flow structures. The three large panels of Fig. 12 show the noisy data, E15 reconstruction, and 'clean' PIV-HDR data on the HDR region. The E15 reconstruction matches the prominent features of the 'clean' data quite well.

7 Conclusions

This paper addresses several questions regarding noise filtering via the singular value decomposition:

How effective is the SVD for filtering out the noise and reconstruction the clean data?

The effectiveness of the SVD for noise filtering is dependent on the decay rate of the clean singular values and the noise level of the data. Recall, the reconstruction loss was reasonably estimated by Eq. (22), which predicted

$$\bar{\Delta}_{\bar{r}_{\min}} = \bar{\epsilon}^2 D \bar{r}_{\min} + \sum_{k=\bar{r}_{\min}+1}^{T} \bar{s}_k^2 \,. \tag{25}$$

Considering (25), it is clear that the lowest losses can be achieved in problems where the optimum reconstruction rank $\bar{r}_{\min}$ is small and the clean singular values in the tail $k > \bar{r}_{\min}$ are negligible. The examples from Sects. 2.1, 4, and 5.2 have such character, so their reconstruction loss $\Delta_{\bar{r}_{\min}}$ was much less than the loss of the original noisy data $\epsilon^2 TD$.

The non-dimensional ratio

$$M \equiv \frac{\bar{\epsilon}^2 TD - \bar{\Delta}_{\bar{r}_{\min}}}{\bar{\epsilon}^2 TD} \,. \tag{23}$$

can be used to gauge the improvement in loss. Values greater than zero suggest that the reconstruction is more accurate than the original noisy data, with values approaching unity suggesting a highly accurate reconstruction.

The character of having negligible clean singular values and a low reconstruction rank is akin to having a large separation between the tails of the clean and noisy singular values. With large separation, the filtering capacity of the SVD is quite good. However, in cases with little separation, an SVD-based reconstruction might filter some noise but not necessarily improve the accuracy (loss) over that of the

[8] For the inquiring mind, $\bar{\Delta}_{\bar{r}_{\min}} \cdot (D_{HDR}/D) = 198$, just 45% lower than $\Delta_{\bar{r}_{\min}}$, and $M = 73\%$.

original noisy data (*c.f.* the $4\delta t$ curve in Fig. 9c). In Sect. 5.3, we showed that the separation between clean and noisy singular values can be increased by using a decreased timestep.

What method yields the most accurate data reconstruction, and what are the limits to its accuracy? While a number of practical reconstruction approaches were investigated, the 'E15' method (15)/(16)/(17) proved to be most robust, with the lowest mean loss $\langle \Delta_r \rangle$ and the least variation σ_{Δ_r}. Moreover, choosing the rank $\bar{r}_{\min}$ yields nearly the minimum loss possible when reconstructing with the noisy singular *vectors*.

In "Appendix 3", we show that even more accurate reconstructions could be formed if the clean singular vectors were known (or correctly estimated somehow). Thus, one focus of future work is to develop a practical method for estimating the clean singular vectors from the noisy ones.

Can the magnitude of the noise be inferred from the noisy data themselves? A method to estimate the RMS noise was presented in "Appendix 1". The approach is to fit a Marchenko–Pasteur distribution to the tail of the singular value distribution, with the noise level and the choice of index defining the tail determined so as to minimize the least square error of this fit. This method is sufficiently robust so as to enable noise estimation (and subsequently noise filtering) in an automated data processing code.

Collectively, this body of work provides a thorough analysis of the effects of noise on the SVD of noisy data, the potential for noise estimation using the SVD, and the capabilities of the SVD for noise filtering.

Acknowledgements B. P. Epps dedicates this paper to Julian D. Cole, who likely would have enjoyed working on this problem.

Appendix 1: Estimation of the 'measurement error' $\bar{\epsilon}$, 'spatial correlation factor' *f*, and 'effective smoothing window width' *w*

Here we describe a procedure that can be used to estimate the measurement error $\bar{\epsilon}$, 'spatial correlation factor' *f*, and 'effective smoothing window width' *w* for a given noisy dataset. The procedure involves fitting the tail of the noisy singular values with a *Marchenko–Pastur distribution*, $\epsilon' \hat{s}_\ell$. The inputs to the procedure are the noisy singular values $\tilde{s}_\ell$ and the size, T and D, of the dataset. The outputs are $\bar{\epsilon}$, f, and w, as well as the noise level of the fit ϵ', and the index $\bar{k}_\epsilon$ marking the start of the fit.

In (Epps and Krivitzky 2019), we provide the formula for the *Marchenko–Pastur distribution* $\epsilon' \hat{s}_\ell$. We use the notation $\epsilon' \hat{s}_\ell$ to emphasize that these singular values are linearly proportional to the noise level ϵ' (such that $\hat{s}_\ell$ corresponds to a unit-noise distribution). These $\hat{s}_\ell$ are a function of T, D, and f.

In (Epps and Krivitzky 2019), we showed that the tail of the noisy singular values $\tilde{s}_\ell$ follows a *Marchenko–Pastur distribution* $\epsilon' \hat{s}_\ell$. That is, $\tilde{s}_\ell \approx \epsilon' \hat{s}_\ell$ in the tail $\ell \geq k_\epsilon$, where k_ϵ is defined as the minimum index for which $\tilde{s}_\ell < \epsilon\sqrt{D}$. This suggests a noise estimation procedure: Fit the tail of the data ($\tilde{s}_\ell$ for $\ell = k_\epsilon, \ldots, T$) with $\epsilon' \hat{s}_\ell$, and then upon setting $\epsilon' \hat{s}_{k_\epsilon} = \epsilon\sqrt{D}$ at the tail-start index k_ϵ, find the estimate $\bar{\epsilon} = \epsilon' \hat{s}_{k_\epsilon} / \sqrt{D}$. One difficulty with this approach is that k_ϵ is unknown *a priori*, because it depends on ϵ. Therefore, a slightly more elaborate procedure is required.

Given a guess of the tail-start index k, we use least squares to fit a Marchenko–Pastur distribution $\epsilon' \hat{s}_\ell$ to the tail of the noisy singular values $\tilde{s}_{\ell\,(\ell=k,\ldots,T)}$. The *mean square error* between $\log_{10}(\epsilon' \hat{s}_\ell)$ and $\log_{10} \tilde{s}_\ell$ is:

$$L = \frac{1}{T+1-k} \sum_{\ell=k}^{T} (\log_{10} \epsilon' + \log_{10} \hat{s}_\ell - \log_{10} \tilde{s}_\ell)^2 \,. \tag{26}$$

The ϵ' that yields the minimum L is found by solving $\frac{\mathrm{d}L}{\mathrm{d}(\log_{10} \epsilon')} = 0$ for ϵ', which yields

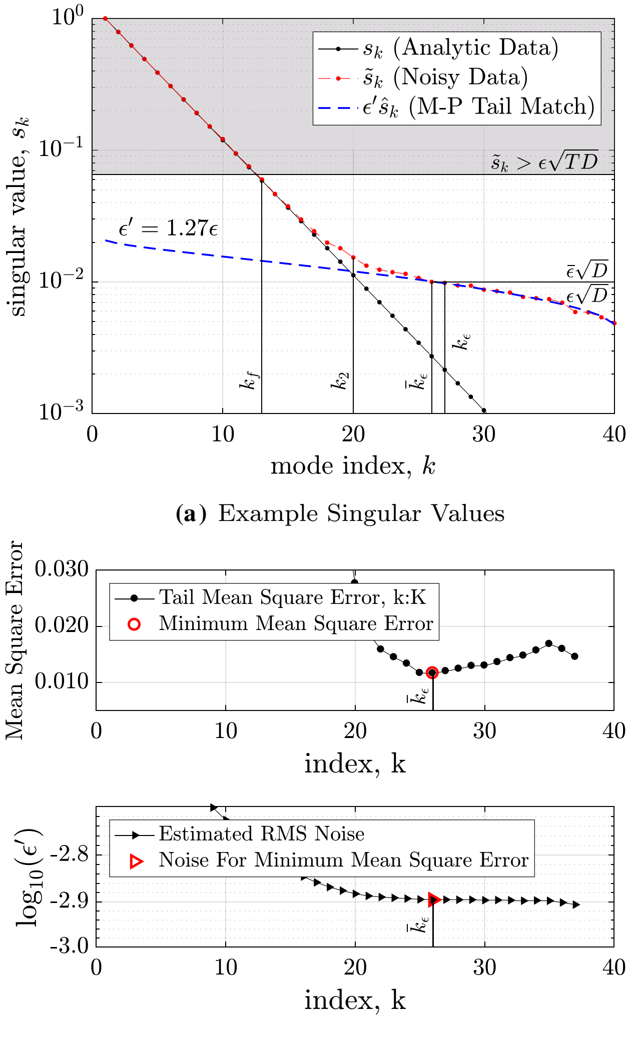
$$\log_{10} \epsilon' = \frac{1}{T+1-k} \sum_{\ell=k}^{T} (\log_{10} \tilde{s}_\ell - \log_{10} \hat{s}_\ell) \,. \tag{27}$$

Thus, for each choice of tail-start index k, Eq. (27) can be used to find the best fit $\epsilon'(k)$, and (26) can be used to evaluate the associated error $L(k)$.

The actual tail-start index k_ϵ can then be estimated as the index for which the fit error $L(k)$ is a minimum. This assumption is reasonable, because k_ϵ marks a rapid departure of the noisy singular values from the fitted Marchenko–Pastur distribution.

For example, Fig. 13a shows a distribution of noisy singular values and the resulting best-fit Marchenko–Pastur distribution. Figure 13b shows the mean square error $L(k)$ (26) and the fitted noise level, $\log_{10} \epsilon'(k)$ (27) versus tail start index k. Observe that the mean square error $L(k)$ has a clear minimum (at the index marked $\bar{k}_\epsilon$), which suggests that the overall best fit is formed with $k = \bar{k}_\epsilon$. Moreover, note that there is very low sensitivity of the inferred noise level $\epsilon'(k)$ to the particular choice of index $\bar{k}_\epsilon$, so this procedure provides a robust way to determine ϵ'.

For *i.i.d.* noise, $w = f = 1$, and the above methods can be used to find ϵ' and $\bar{k}_\epsilon$. For spatially-correlated noise, w and



(a) Example Singular Values

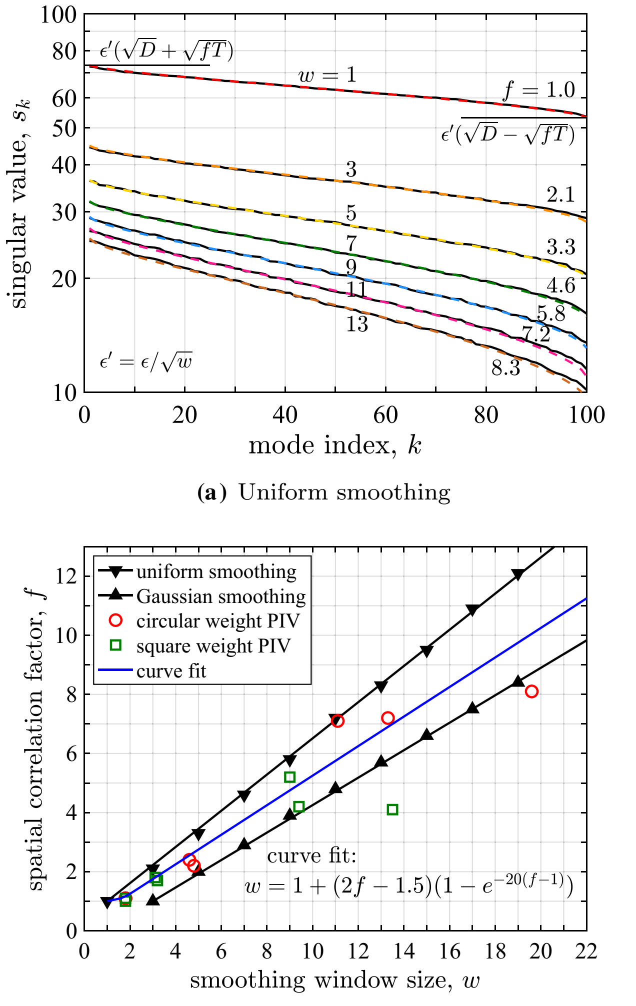
(b) Tail Matching Metrics

Fig. 13 Example Marchenko–Pastur tail fit: **a** singular values; **b** (top) *mean square error*, $L_{(k)}$ (26) and (bottom) fitted noise level, $\log_{10}\epsilon'_{(k)}$ (27). Here, $T = 40$, $D = 100$, $\epsilon = 0.001$, $k_\epsilon = 27$, $\bar{k}_\epsilon = 26$, $\epsilon'_{\bar{k}_\epsilon} = 1.272\epsilon$, and $\bar{\epsilon} = 1.003\epsilon$

f need to be determined as part of the fit. In our procedure, we determine the best fit using (26)/(27) for a series of *f*, so that we build a table of $L_{(k,f)}$. The best fit is that corresponding to the *k* and *f* for which the fit error $L_{(k,f)}$ is a minimum.

Spatially-correlated Noise

Spatially-correlated noise can occur in experimental data that are spatially smoothed during collection or processing. For example, PIV data are typically collected from overlapping interrogation windows, and processing typically includes smoothing by a weighted average over the nine nearest neighbors. Such a dataset effectively has fewer than *D independent* data sites. Thus, it is reasonable to expect that



(a) Uniform smoothing

(b) Spatial correlation factor f versus window width w

Fig. 14 **a** Singular values of random data smoothed over selected window widths *w*. Each dataset is fit with a Marchenko–Pastur distribution, with the correlation factor *f* as shown. Here, $T = 100$, $D = 4000$, $\epsilon = 1$. For uniform smoothing, $\epsilon' = \epsilon/\sqrt{w}$. **b** Spatial correlation factor *f* plotted versus smoothing window width *w*; the uniform smoothing data are from **a**, and the Gaussian smoothing data and PIV data are given in (Epps and Krivitzky 2019)

the singular values of spatially-correlated random data still follow a Marchenko–Pastur distribution, but with parameter $y = T/D$ replaced by $fT\,/\,D$ (see Epps and Krivitzky 2019, Appendix D). Indeed, we have empirically found this approximation to work well when $D/T \gtrsim 20$ and $D/fT \gtrsim 5$. The 'spatial-correlation factor' *f* represents the ratio

$$f = D/(\text{effective number of independent data sites})\,, \qquad (28)$$

so $f > 1$ indicates effectively-fewer independent data sites due to spatial correlation.

For example, consider random data with spatial correlation that is produced by taking a moving average of i.i.d. random data. Such an average could either have uniform weighting 1 / *w* or Gaussian weighting (see Matlab `gausswin` function), where w is the width of the smoothing window. Figure 14a shows the singular values of random data with uniform spatial smoothing. The $w = 1$ curve corresponds to the original i.i.d. data (no smoothing) and is well represented by the original Marchenko–Pastur distribution ($f = 1$). Clearly, the 'spatial-correlation factor' f increases with increasing 'smoothing window width' w.

Figure 14b shows these f versus w data for uniform smoothing and Gaussian smoothing, as well as the f versus $\tilde{w}$ data from the synthetic PIV example for several PIV processing schemes. These $\tilde{w}$ data were determined in (Epps and Krivitzky 2019) so as to match the theoretical prediction of rmse($\tilde{\mathbf{u}}_1$) to the numerical value. Equation (12) is a curve fit to these synthetic PIV data and provides an empirical relationship between w and f.

The precise value of w is not critically important for noise filtering, since it is recommended that $w = 1$ be used in (13) to the evaluate rmse($\tilde{\mathbf{v}}_k$) for PIV data. However, it is conceivable that one might be interested in knowing the precise value of w for a particular flow problem or PIV processing scheme. Under those circumstances, it is recommended to use a known dataset to determine the f–w relation for the new PIV processing scheme before post-processing the target data.

Summary of error estimation procedure

To summarize, the following procedure is used to determine f, $\bar{k}_\epsilon$, ϵ', $\bar{\epsilon}$, and w:

1. For each $f = 1.0, 1.1, \dots, \text{floor}(D/T)$[9] and tail-start index $k = 1, \dots,$ `floor`$(0.8\,T)$,[10] compute the unit Marchenko–Pastur distribution $\hat{s}_\ell$, evaluate (27) to determine the best-fit noise level $\epsilon'(f,k)$, and compute the associated mean square error $L(f,k)$ via (26).
2. Find the f and index k for which $L(f,k)$ is a minimum. Set $\bar{k}_\epsilon$ to this selected k.
3. Form a preliminary estimate of the measurement error:

$$\epsilon'' = \frac{\epsilon'(f,\bar{k}_\epsilon)\,\hat{s}_{\bar{k}_\epsilon}}{\sqrt{D}}\,. \tag{29}$$

Equation (29) follows from the observation that the noisy singular values $\tilde{s}_k$ rapidly depart from the Marchenko–Pastur distribution $\epsilon'\hat{s}_\ell$ near the value $\epsilon'\hat{s}_\ell = \epsilon\sqrt{D}$. Note, however, that typically ϵ' is greater than the true ϵ while ϵ'' is less than ϵ, so a weighted average of ϵ' and ϵ'' typically forms the best estimate of ϵ.

4. Find index

$$k_{\epsilon'} = \text{minimum } k \text{ such that } \tilde{s}_k < \epsilon'\sqrt{D}\,. \tag{30}$$

5. Estimate the measurement error by

$$\bar{\epsilon} = \min\left\{\epsilon',\ \epsilon'' + (\epsilon' - \epsilon'')\frac{\bar{k}_\epsilon - k_{\epsilon'}}{\text{floor}(0.8T) - k_{\epsilon'}}\right\}. \tag{31}$$

6. Using the best fit f, evaluate (12) to find w.

This approach has proven to be straightforward and robust under an array of singular value distributions, error levels, and datasets. This procedure provides a robust alternative to the traditional "scree test" (Cattell 1966), which can be foiled by closely spaced singular values, such as those in Sect. 3.2 example.

Appendix 2: Derivation of optimum reconstruction when using noisy modes

Consider reconstruction using the noisy mode shapes $\{\tilde{\mathbf{u}}_\ell, \tilde{\mathbf{v}}_\ell\}$ and some optimum singular values $\bar{s}_\ell$ to be determined:

$$\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell \bar{s}_\ell \tilde{\mathbf{v}}_\ell^{\mathsf{T}}\,, \tag{32}$$

We now derive the $\bar{s}_\ell$ that minimize the *reconstruction loss* $\Delta_r \equiv \|\mathbf{A} - \bar{\mathbf{A}}_r\|_F^2$ (4). In general, the loss can be written as:

$$\Delta_r = \|\mathbf{A}\|_F^2 + \|\bar{\mathbf{A}}_r\|_F^2 - 2A_{ij}\bar{A}_{ij}\,, \tag{33}$$

$$\Delta_r = s_k^2\delta_{kk} + \bar{s}_\ell^2\delta_{\ell\ell} - 2U_{ik}\tilde{U}_{i\ell}\, s_k\bar{s}_\ell\, V_{jk}\tilde{V}_{j\ell}\,, \tag{34}$$

with implied summation over $k = 1, \dots, T$ and $\ell = 1, \dots, r$, $i = 1, \dots, T$, and $j = 1, \dots, D$. The minimum loss can be found by taking partial derivatives of (34) with respect to each $\bar{s}_\ell$ and setting them to zero

$$\frac{\partial \Delta_r}{\partial \bar{s}_\ell} = 2\bar{s}_\ell - 2U_{ik}\tilde{U}_{i\ell}\, s_k\, V_{jk}\tilde{V}_{j\ell} = 0\,, \tag{35}$$

now with no implied sum over ℓ. Clearly, the optimum $\bar{s}_\ell$ is $\boxed{\bar{s}_\ell = s_\ell c_\ell}$ with c_ℓ defined as

$$c_\ell \equiv \sum_{i=1}^{T}\sum_{j=1}^{D}\sum_{k=1}^{T} U_{ik}\tilde{U}_{i\ell}\, V_{jk}\tilde{V}_{j\ell}\, s_k/s_\ell\,,$$

$$\boxed{c_\ell \equiv \sum_{k=1}^{T} (\mathbf{u}_k \cdot \tilde{\mathbf{u}}_\ell)\,(\mathbf{v}_k \cdot \tilde{\mathbf{v}}_\ell)\, s_k/s_\ell}\,. \tag{21}$$

[9] For *i.i.d.* data, only use $f = 1$.

[10] We cap k at $0.8\,T$ in order to leave sufficient singular values in the tail to yield a reliable and meaningful tail fit.

With a similar development, the expected value of the loss $\langle \Delta_r \rangle$ is found to be minimized by setting $\boxed{\bar{s}_\ell = s_\ell \langle c_\ell \rangle}$. This expected value can be evaluated using the perturbation theory results from (Epps and Krivitzky 2019) as follows:

$$\begin{aligned}\langle c_\ell \rangle = \langle U_{ik}(U_{i\ell} + \epsilon\, W^{(1)}_{im\ell} U_{m\ell} + \epsilon^2\, W^{(2)}_{im\ell} U_{m\ell} + \cdots) \\ \cdot\, V_{jk}(V_{j\ell} + \epsilon\, N^{(1)}_{jm\ell} V_{m\ell} + \epsilon^2\, N^{(2)}_{jm\ell} V_{m\ell} + \cdots)\rangle\, s_k / s_\ell\ .\end{aligned} \tag{36}$$

Note that the expected values of the resulting $\mathcal{O}(\epsilon)$ terms are zero, because $\langle W^{(1)}_{im\ell} \rangle = \langle N^{(1)}_{jm\ell} \rangle = 0$. Also note $U_{ik} U_{i\ell} = \delta_{k\ell}$ and $V_{jk} V_{j\ell} = \delta_{k\ell}$. Thus, (36) reduces to

$$\langle c_\ell \rangle = 1 + \epsilon^2 U_{i\ell} \langle W^{(2)}_{im\ell} U_{m\ell} \rangle + \epsilon^2 V_{j\ell} \langle N^{(2)}_{jm\ell} V_{m\ell} \rangle + \mathcal{O}(\epsilon^4) \tag{37}$$

The remaining two expected value terms are given in (Epps and Krivitzky 2019). To $\mathcal{O}(\epsilon^4)$, we find

$$\begin{aligned}\langle c_\ell \rangle = 1 - \frac{\epsilon^2 w}{2} \frac{\lambda_m + \lambda_\ell}{(\lambda_m - \lambda_\ell)^2} (1 - \delta_{m\ell}) U^2_{i\ell} \\ - \frac{\epsilon^2}{2\lambda_\ell} \left[D - w + w \frac{\lambda_m (3\lambda_\ell - \lambda_m)}{(\lambda_m - \lambda_\ell)^2} (1 - \delta_{m\ell}) \right] V^2_{j\ell}\ .\end{aligned} \tag{38}$$

Upon summing over i and j, the result simplifies to

$$\langle c_\ell \rangle = 1 - \tfrac{T}{2} [\mathrm{rms}(\sigma_{\tilde{\mathbf{u}}_\ell})]^2 - \tfrac{D}{2} [\mathrm{rms}(\sigma_{\tilde{\mathbf{v}}_\ell})]^2 + \mathcal{O}(\epsilon^2)\ . \tag{39}$$

Finally, in terms of the canonical angles, we have

$$\boxed{\langle c_\ell \rangle = \langle \cos\phi_\ell \rangle \langle \cos\theta_\ell \rangle} + \mathcal{O}(\epsilon^2)\ . \tag{40}$$

where $\cos\phi_\ell \equiv \tilde{\mathbf{u}}_\ell \cdot \mathbf{u}_\ell$ and $\cos\theta_\ell \equiv \tilde{\mathbf{v}}_\ell \cdot \mathbf{v}_\ell$. Upon regrouping, this optimum reconstruction can be interpreted as $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} (\tilde{\mathbf{u}}_\ell \langle \cos\phi_\ell \rangle) s_\ell (\tilde{\mathbf{v}}^{\mathsf{T}}_\ell \langle \cos\theta_\ell \rangle)$. In words, the most accurate reconstruction is the one that uses the correct singular values (mode amplitudes) and the projections of the clean singular vectors into the noisy singular vector directions.

An estimate of the canonical angle term $\langle c_\ell \rangle$ from (40) can be made as follows:

$$\boxed{\bar{c}_k \equiv \overline{\langle \cos\phi_k \rangle}\ \overline{\langle \cos\theta_k \rangle}} + \mathcal{O}(\epsilon^2)\ , \tag{41}$$

where

$$\begin{aligned}\overline{\langle \cos\phi_k \rangle} = 1 - \tfrac{T}{2} \left[\overline{\mathrm{rms}(\sigma_{\tilde{\mathbf{u}}_k})} \right]^2 \\ \overline{\langle \cos\theta_k \rangle} = 1 - \tfrac{D}{2} \left[\overline{\mathrm{rms}(\sigma_{\tilde{\mathbf{v}}_k})} \right]^2\ ,\end{aligned} \tag{42}$$

and

$$\overline{\mathrm{rms}(\sigma_{\tilde{\mathbf{u}}_k})} \approx \min\left\{ \sqrt{2/T}\ ,\ \ \frac{\bar{\epsilon}}{\tilde{s}_k} \left[\frac{w}{T} \sum_{\substack{m=1 \\ m \neq k}}^{T} \frac{\tilde{\lambda}_k (\tilde{\lambda}_m + \tilde{\lambda}_k)}{(\tilde{\lambda}_m - \tilde{\lambda}_k)^2} \right]^{\frac{1}{2}} \right\} \tag{43}$$

$$\begin{aligned}\overline{\mathrm{rms}(\sigma_{\tilde{\mathbf{v}}_k})} \approx \min\Bigg\{ \sqrt{2/D}\ , \\ \frac{\bar{\epsilon}}{\tilde{s}_k} \left[\frac{D - w}{D} + \frac{w}{D} \sum_{\substack{m=1 \\ m \neq k}}^{T} \frac{\tilde{\lambda}_m (3\tilde{\lambda}_k - \tilde{\lambda}_m)}{(\tilde{\lambda}_m - \tilde{\lambda}_k)^2} \right]^{\frac{1}{2}} \Bigg\}\ ,\end{aligned} \tag{44}$$

where $\tilde{\lambda}_k \equiv \tilde{s}^2_k$ (Epps and Krivitzky 2019).

Appendix 3: Hypothetical data reconstruction methods

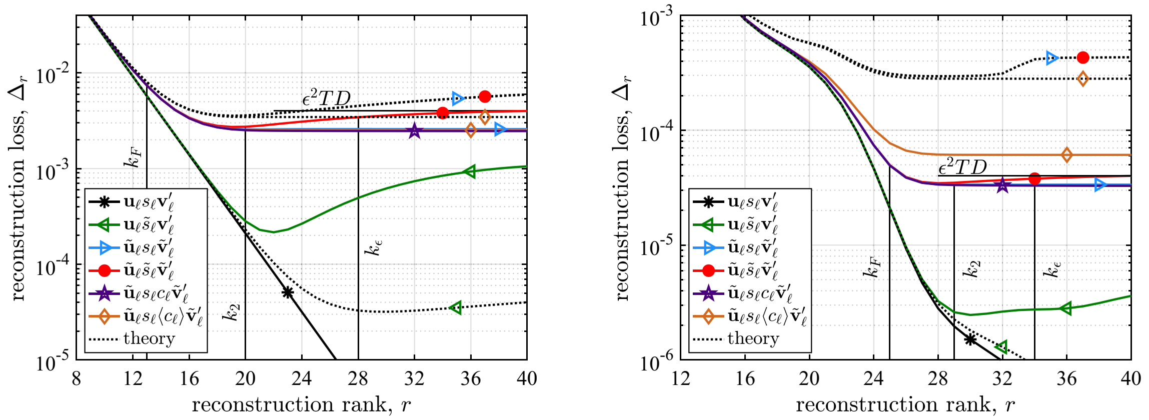
In general, the reconstruction $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \bar{\mathbf{u}}_\ell \bar{s}_\ell \bar{\mathbf{v}}^{\mathsf{T}}_\ell$ from (3) depends on the choice of rank, as well as estimates of the clean singular values and vectors. In this section, we consider hypothetical reconstruction methods that use various combinations of the noisy and clean variables: $\bar{\mathbf{u}}_\ell = \{\tilde{\mathbf{u}}_\ell, \mathbf{u}_\ell\}$, $\bar{\mathbf{v}}_\ell = \{\tilde{\mathbf{v}}_\ell, \mathbf{v}_\ell\}$, and $\bar{s}_\ell = \{\tilde{s}_\ell, s_\ell, s_\ell c_\ell\}$, where c_ℓ is defined below. Obviously, the methods that use the clean variables ($\mathbf{u}_\ell$,s_ℓ,$\mathbf{v}_\ell$) are only hypothetical, because these clean variables are unknown if just given $\tilde{\mathbf{A}}$; nevertheless, these methods are interesting, because they illustrate the best-case scenarios for reconstruction losses. Comparing these hypothetical data reconstruction methods allows us to answer two important questions:

Which has more of an effect on the reconstruction loss, perturbations to the singular values or to the singular vectors?

Figure 15 shows that reconstruction loss is affected much more by perturbations to the singular *vectors* than by perturbations to the singular *values*. For example, the reconstruction loss due to perturbed singular vectors $\tilde{\mathbf{u}}_\ell s_\ell \tilde{\mathbf{v}}^{\mathsf{T}}_\ell$ (blue curve) is much larger than that due to perturbed singular values $\mathbf{u}_\ell \tilde{s}_\ell \mathbf{v}^{\mathsf{T}}_\ell$ (green curve).

How important to the reconstruction accuracy is it to estimate the clean singular values?

Observe that the '$\tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}^{\mathsf{T}}_\ell$' reconstruction (red curve) has a local minimum (near $r = k_2$), whereas the '$\tilde{\mathbf{u}}_\ell s_\ell \tilde{\mathbf{v}}^{\mathsf{T}}_\ell$' reconstruction (blue curve) has nearly constant loss for $r > k_2$. Thus, the advantage of estimating the clean singular values is that the resulting reconstruction loss will be much less sensitive to the choice of rank.



(a) §3.1 example

(b) §3.2 example

Fig. 15 Mean reconstruction loss $\langle \mathit{\Delta}_r \rangle \equiv \langle \|\mathbf{A} - \bar{\mathbf{A}}_r\|_F^2 \rangle$, comparing results from Monte Carlo simulations (solid curves) to perturbation theory (dashed curves, formulae in Table 2)

Assessment of reconstruction error predictions from perturbation theory

In this section, we use perturbation theory to make theoretical predictions of the expected reconstruction losses $\langle \mathit{\Delta}_r \rangle$ (4) for each of the hypothetical reconstruction schemes shown in Fig. 15. This analysis is relegated to the appendix, because these perturbation theory predictions are relatively poor. Indeed, Fig. 15 shows poor agreement between the mean reconstruction loss determined from Monte Carlo simulations (solid curves) and the perturbation theory predictions derived herein (dashed curves, formulae in Table 2). The reason for the poor agreement is that the perturbation theory predictions are not very accurate for $k > k_2$, which is the region of interest in Fig. 15.

One interesting result is that perturbation theory predicts that most of the loss is due to the perturbations of the singular vectors, not the singular values. Indeed, comparing Eqs. (61) and (63), we find that if the clean singular values are used for reconstruction (instead of the noisy ones), then the loss is only improved by $\epsilon^2 r$.

Table 2 summarizes the perturbation theory predictions of the expected loss from reconstructions of the form:

$$\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \bar{\mathbf{u}}_\ell \bar{s}_\ell \bar{\mathbf{v}}_\ell^\intercal \,. \tag{3}$$

The entries of Table 2 are derived as follows. In general, the expected loss $\langle \mathit{\Delta}_r \rangle \equiv \langle \|\mathbf{A} - \bar{\mathbf{A}}_r\|_F^2 \rangle$ is

$$\langle \mathit{\Delta}_r \rangle = \|\mathbf{A}\|_F^2 + \langle \|\bar{\mathbf{A}}_r\|_F^2 \rangle - 2\langle A_{ij} \bar{A}_{ij} \rangle \,, \tag{45}$$

$$\langle \mathit{\Delta}_r \rangle = \lambda_k \delta_{kk} + \langle \bar{\lambda}_\ell \rangle \delta_{\ell\ell} - 2\langle U_{ik} \bar{U}_{i\ell}\, s_k \bar{s}_\ell\, V_{jk} \bar{V}_{j\ell} \rangle \,, \tag{46}$$

where $\lambda = s^2$, and summation is implied over $k = 1, \ldots, T$ and $\ell = 1, \ldots, r$, $i = 1, \ldots, T$, and $j = 1, \ldots, D$. The crux of

Table 2 Expected *reconstruction loss* $\langle \mathit{\Delta}_r \rangle \equiv \langle \|\mathbf{A} - \bar{\mathbf{A}}_r\|_F^2 \rangle$ from several reconstruction methods

$\mathbf{A}_r = \sum_{\ell=1}^{r} \mathbf{u}_\ell s_\ell \mathbf{v}_\ell^\intercal$	(56)	$\langle \mathit{\Delta}_r \rangle = \sum_{k=r+1}^{T} \lambda_k$	(57)
$\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \mathbf{u}_\ell \tilde{s}_\ell \mathbf{v}_\ell^\intercal$	(58)	$\langle \mathit{\Delta}_r \rangle = \epsilon^2 r + \sum_{k=r+1}^{T} \lambda_k + \mathcal{O}(\epsilon^4)$	(59)
$\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell s_\ell \tilde{\mathbf{v}}_\ell^\intercal$	(60)	$\langle \mathit{\Delta}_r \rangle = \epsilon^2 (D-T) r + \sum_{k=r+1}^{T} \lambda_k + 2\epsilon^2 \sum_{\ell=1}^{r} \sum_{\substack{p=1 \\ p\neq\ell}}^{T} \frac{\lambda_\ell(\lambda_p+\lambda_\ell)}{(\lambda_p-\lambda_\ell)^2} + \mathcal{O}(\epsilon^4)$	(61)
$\tilde{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^\intercal$	(62)	$\langle \mathit{\Delta}_r \rangle = \epsilon^2 r + \epsilon^2 (D-T) r + \sum_{k=r+1}^{T} \lambda_k + 2\epsilon^2 \sum_{\ell=1}^{r} \sum_{\substack{p=1 \\ p\neq\ell}}^{T} \frac{\lambda_\ell(\lambda_p+\lambda_\ell)}{(\lambda_p-\lambda_\ell)^2} + \mathcal{O}(\epsilon^4)$	(63)
$\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell s_\ell \tilde{\mathbf{v}}_\ell^\intercal \bar{c}_\ell$	(64)	$\langle \mathit{\Delta}_r \rangle = \sum_{k=r+1}^{T} \lambda_k + \sum_{\ell=1}^{r} \lambda_\ell (1-\bar{c}_\ell)^2 + \epsilon^2 (D-T) \sum_{\ell=1}^{r} \bar{c}_\ell + 2\epsilon^2 \sum_{\ell=1}^{r} \sum_{\substack{p=1 \\ p\neq\ell}}^{T} \frac{\lambda_\ell(\lambda_p+\lambda_\ell)}{(\lambda_p-\lambda_\ell)^2} \bar{c}_\ell + \mathcal{O}(\epsilon^4)$	(65)

estimating $\langle \varDelta_r \rangle$ is in evaluating $\langle A_{ij}\bar{A}_{ij}\rangle = \langle U_{ik}\bar{U}_{i\ell}\, s_k\bar{s}_\ell\, V_{jk}\bar{V}_{j\ell}\rangle$ using perturbation theory.

For method (56), $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \mathbf{u}_\ell s_\ell \mathbf{v}_\ell^\intercal$, we have $\langle \bar{\lambda}_\ell\rangle = \lambda_\ell$ and $\langle A_{ij}\bar{A}_{ij}\rangle = \langle U_{ik}U_{i\ell}\, s_k s_\ell\, V_{jk}V_{j\ell}\rangle = s_k s_\ell \delta_{k\ell} = \lambda_\ell \delta_{\ell\ell}$. Trivially, the loss (46) then is $\langle \varDelta_r\rangle = \sum_{k=r+1}^{T} \lambda_k$ (57). This is the minimum possible loss for a rank r reconstruction, per the *Schmidt theorem* (1907).

For method (58), $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \mathbf{u}_\ell \tilde{s}_\ell \mathbf{v}_\ell^\intercal$, we have $\langle \bar{\lambda}_\ell\rangle = \langle \tilde{\lambda}_\ell\rangle$, which we found in (Epps and Krivitzky 2019) to be:

$$\langle \tilde{\lambda}_\ell\rangle = \lambda_\ell + \epsilon^2\left(D - \sum_{\substack{p=1\\p\neq\ell}}^{T} \frac{\lambda_p+\lambda_\ell}{\lambda_p-\lambda_\ell}\right) + \mathcal{O}(\epsilon^4)\,. \tag{47}$$

We have $\langle A_{ij}\bar{A}_{ij}\rangle = \langle U_{ik}U_{i\ell}\, s_k\tilde{s}_\ell\, V_{jk}V_{j\ell}\rangle = s_\ell\langle\tilde{s}_\ell\rangle$. In (Epps and Krivitzky 2019), we derived $\langle\tilde{s}_\ell\rangle$, from which directly follows

$$\langle A_{ij}\bar{A}_{ij}\rangle = \lambda_\ell\delta_{\ell\ell} + \frac{\epsilon^2}{2}\left(D - 1 - \sum_{\substack{p=1\\p\neq\ell}}^{T} \frac{\lambda_p+\lambda_\ell}{\lambda_p-\lambda_\ell}\right)\delta_{\ell\ell} + \mathcal{O}(\epsilon^4)\,, \tag{48}$$

with implied sum over $\ell = 1,\ldots,r$. With these results, Eq. (46) yields $\langle\varDelta_r\rangle = \sum_{\underline{k}=r+1}^{T} \lambda_k + \epsilon^2 r + \mathcal{O}(\epsilon^4)$ (59).

For method (60) $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell s_\ell \tilde{\mathbf{v}}_\ell^\intercal$, we have $\langle\bar{\lambda}_\ell\rangle = \lambda_\ell$ and $\langle A_{ij}\bar{A}_{ij}\rangle = \langle U_{ik}\tilde{U}_{i\ell}\, s_k s_\ell\, V_{jk}\tilde{V}_{j\ell}\rangle$. Working through the algebra, we find (with $\mathcal{O}(\epsilon^4)$ accuracy)

$$\langle A_{ij}\bar{A}_{ij}\rangle = \lambda_\ell\delta_{\ell\ell} + \epsilon^2\lambda_\ell\left\langle U_{i\ell}W^{(2)}_{im\ell}U_{m\ell} + V_{j\ell}N^{(2)}_{jm\ell}V_{m\ell}\right\rangle\,. \tag{49}$$

With some effort, we have (with $\mathcal{O}(\epsilon^4)$ accuracy)

$$\langle A_{ij}\bar{A}_{ij}\rangle = \lambda_\ell\delta_{\ell\ell} - \epsilon^2\tfrac{1}{2}(D-T)r - \epsilon^2\frac{\lambda_\ell(\lambda_p+\lambda_\ell)}{(\lambda_p-\lambda_\ell)^2}(1-\delta_{\ell p})\,, \tag{50}$$

with implied sum over $\ell = 1,\ldots,r$ and $p = 1,\ldots,T$ but $p\neq\ell$. Inserting (50) into (46) and simplifying, the result is Eq. (61) listed in Table 2.

For method (62) $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r} \tilde{\mathbf{u}}_\ell \tilde{s}_\ell \tilde{\mathbf{v}}_\ell^\intercal$, we have $\langle\bar{\lambda}_\ell\rangle = \langle\tilde{\lambda}_\ell\rangle$, which was given in Eq. (47). We also have $\langle A_{ij}\bar{A}_{ij}\rangle = \langle U_{ik}\tilde{U}_{i\ell}\, s_k\tilde{s}_\ell\, V_{jk}\tilde{V}_{j\ell}\rangle$. Using perturbation theory results from (Epps and Krivitzky 2019), we find

$$\begin{aligned}\langle A_{ij}\bar{A}_{ij}\rangle = \Bigg\langle & U_{ik}(U_{i\ell} + \epsilon\, W^{(1)}_{im\ell}U_{m\ell} + \epsilon^2\, W^{(2)}_{im\ell}U_{m\ell} + \cdots)\\ &\cdot s_k\left(s_\ell + \epsilon\left(\frac{\hat{\lambda}^{(1)}_\ell}{2s_\ell}\right) + \epsilon^2\left(\frac{\hat{\lambda}^{(2)}_\ell}{2s_\ell} - \frac{(\hat{\lambda}^{(1)}_\ell)^2}{8s_\ell^3}\right)\right)\\ &\cdot V_{jk}(V_{j\ell} + \epsilon\, N^{(1)}_{jm\ell}V_{m\ell} + \epsilon^2\, N^{(2)}_{jm\ell}V_{m\ell} + \cdots)\Bigg\rangle\,.\end{aligned} \tag{51}$$

Note that upon expanding and collecting terms, the expected values of the resulting $\mathcal{O}(\epsilon)$ terms will be zero, because $\langle W^{(1)}_{im\ell}\rangle = \langle N^{(1)}_{jm\ell}\rangle = \langle\lambda^{(1)}_\ell\rangle = \cdots = 0$. Also note $U_{ik}U_{i\ell} = \delta_{k\ell}$ and $V_{jk}V_{j\ell} = \delta_{k\ell}$. Thus, we are left with

$$\begin{aligned}\langle A_{ij}\bar{A}_{ij}\rangle = s_k s_\ell\delta_{k\ell} + \epsilon^2\Bigg\langle & (W^{(2)}_{im\ell}U_{m\ell}U_{ik})(s_k s_\ell)\delta_{k\ell}\\ &+ \delta_{k\ell}\left(\tfrac{s_k}{2s_\ell}\hat{\lambda}^{(2)}_\ell - \tfrac{s_k}{8s_\ell^3}(\hat{\lambda}^{(1)}_\ell)^2\right)\delta_{k\ell}\\ &+ \delta_{k\ell}(s_k s_\ell)(N^{(2)}_{jm\ell}V_{m\ell}V_{jk})\\ &+ \delta_{k\ell}\left(\tfrac{s_k}{2s_\ell}\hat{\lambda}^{(1)}_\ell\right)(N^{(1)}_{jm\ell}V_{m\ell}V_{jk})\\ &+ (W^{(1)}_{im\ell}U_{m\ell}U_{ik})(s_k s_\ell)(N^{(1)}_{jm\ell}V_{m\ell}V_{jk})\\ &+ (W^{(1)}_{im\ell}U_{m\ell}U_{ik})\left(\tfrac{s_k}{2s_\ell}\hat{\lambda}^{(1)}_\ell\right)\delta_{k\ell}\Bigg\rangle\,.\end{aligned} \tag{52}$$

The remaining terms can be evaluated using the methods in (Epps and Krivitzky 2019). With some effort, we find

$$\langle A_{ij}\bar{A}_{ij}\rangle = \lambda_\ell\delta_{\ell\ell} - \epsilon^2\sum_{\substack{p=1\\p\neq\ell}}^{T}\frac{2\lambda_p\lambda_\ell}{(\lambda_p-\lambda_\ell)^2} + \mathcal{O}(\epsilon^4)\,, \tag{53}$$

with implied sum over $\ell = 1,\ldots,r$. Inserting (53) and (47) into (46), we have

$$\begin{aligned}\langle\varDelta_r\rangle = \lambda_k\delta_{kk} + &\left(\lambda_\ell\delta_{\ell\ell} + \epsilon^2 D\delta_{\ell\ell} - \epsilon^2\frac{\lambda_p+\lambda_\ell}{\lambda_p-\lambda_\ell}(1-\delta_{\ell p})\right)\\ &- 2\left(\lambda_\ell\delta_{\ell\ell} - \epsilon^2\frac{2\lambda_p\lambda_\ell}{(\lambda_p-\lambda_\ell)^2}(1-\delta_{\ell p})\right),\end{aligned} \tag{54}$$

which simplifies to the final result, Eq. (63).

For method (64), $\bar{\mathbf{A}}_r = \sum_{\ell=1}^{r}\tilde{\mathbf{u}}_\ell s_\ell\tilde{\mathbf{v}}_\ell^\intercal\bar{c}_\ell$, we effectively have $\bar{s}_\ell = s_\ell\bar{c}_\ell$ so $\sum_{\ell=1}^{r}\langle\bar{\lambda}_\ell\rangle = \sum_{\ell=1}^{r}\lambda_\ell\bar{c}_\ell^2$. We also have $\langle A_{ij}\bar{A}_{ij}\rangle = \langle U_{ik}\tilde{U}_{i\ell}\, s_k s_\ell\, V_{jk}\tilde{V}_{j\ell}\rangle\bar{c}_\ell$, which is $\bar{c}_\ell$ times the summand in method (60). Working through the algebra as in method (60), we find (with $\mathcal{O}(\epsilon^4)$ accuracy)

$$\langle A_{ij}\bar{A}_{ij}\rangle = \lambda_\ell\bar{c}_\ell - \epsilon^2\tfrac{1}{2}(D-T)\bar{c}_\ell\delta_{\ell\ell} - \epsilon^2\bar{c}_\ell\frac{\lambda_\ell(\lambda_p+\lambda_\ell)}{(\lambda_p-\lambda_\ell)^2}(1-\delta_{\ell p})\,. \tag{55}$$

Again, the sums involving $\bar{c}_\ell$ and $\bar{c}_\ell^2$ do not combine neatly, and the loss (46) works out to be that in Eq. (65).

References

Breuer K, Sirovich L (1991) The use of the Karhunen–Loève procedure for the calculation of linear eigenfunctions. J Comput Phys 96:277–296

Brindise MC, Vlachos PP (2017) Proper orthogonal decomposition truncation method for data denoising and order reduction. Exp Fluids 58(4):28

Cattell RB (1966) The scree test for the number of factors. Multivar Behav Res 1:245–276

Cohen K, Siegel S, Wetlesen D, Cameron J, Sick A (2004) Effective sensor placements for the estimation of proper orthogonal decomposition mode coefficients in von Karman vortex street. J Vib Control 10:1857–1880. https://doi.org/10.1177/1077546304046035

Dawson STM, Hemati MS, Williams MO, Rowley CW (2016) Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition. Exp Fluids 57(3):42

Drauault P, Bouhoubeiny E, Germain G (2012) POD investigation of the unsteady turbulent boundary layer developing over porous moving flexible fishing net structure. Exp Fluids 53:277–292

Epps B (2015) On the singular value decomposition of measured data. In: 68th annual meeting of the APS Division of Fluid Dynamics

Epps BP, Krivitzky EM (2019) Singular value decomposition of noisy data: mode corruption. Exp Fluids 60:121. https://doi.org/10.1007/s00348-019-2761-y

Epps BP, Techet AH (2010) An error threshold criterion for singular value decomposition modes extracted from PIV data. Exp Fluids 48:355–367

Feng LH, Wang JJ (2014) Modification of a circular cylinder wake with synthetic jet: vortex shedding modes and mechanism. Eur J Mech B Fluids 43:14–32

Feng LH, Wang JJ, Pan C (2011) Proper orthogonal decomposition analysis of vortex dynamics of a circular cylinder under synthetic jet control. Phys Fluids 23(1):014106

Gandhi V, Bryant DB, Socolofsky SA, Stoesser T, Kim JH (2015) Concentration-based decomposition of the flow around a confined cylinder in a UV disinfection reactor. J Eng Mech 141(12):04015050

Graham J, Kanov K, Yang XIA, Lee M, Malaya N, Lalescu CC, Burns R, Eyink G, Szalay A, Moser RD, Meneveau C (2016) A web services accessible database of turbulent channel flow and its use for testing a new integral wall model for les. J Turbul 17(2):181–215

Higham J, Brevis W, Keylock CJ (2016) A rapid non-iterative proper orthogonal decomposition based outlier detection and correction for PIV data. Meas Sci Technol 27(12):125303

Kato T (1976) Perturbation theory for linear operators. Springer, Berlin

Kim J, Moin P, Moser R (1987) Turbulence statistics in fully developed channel flow at low Reynolds number. J Fluid Mech 177:133–166

Kourentis L, Konstantinidis E (2012) Uncovering large-scale coherent structures in natural and forced turbulent wakes by combining PIV, POD, and FTLE. Exp Fluids 52:749–763

Kriegseis J, Dehler T, Pawlik M, Tropea C (2009) Pattern-identification study of the flow in proximity of a plasma actuator. In: 47th AIAA Aerospace Sciences Meeting, p 1001

Krivitzky E, Epps B (2017) Uncertainty propagation in the singular value decomposition of measured data. In: 70th annual meeting of the APS Division of Fluid Dynamics

Lee M, Malaya N, Moser RD (2013) Petascale direct numerical simulation of turbulent channel flow on up to 786k cores. In: Supercomputing (SC13), Denver, CO

Li Y, Perlman E, Wan M, Yang Y, Meneveau C, Burns R, Chen S, Szalay A, Eyink G (2008) A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. J Turbul 9:31

Ma X, Karniadakis GE, Park H, Gharib M (2003) DPIV-driven flow simulation: a new computational paradigm. Proc R Soc Lond Ser A Math Phys Sci 459:547–565. https://doi.org/10.1098/rspa.2002.0981

Marié S, Druault P, Lambaré H, Schrijer F (2013) Experimental analysis of the pressure–velocity correlations of external unsteady flow over rocket launchers. Aerosp Sci Technol 30:83–93

Mendez M, Raiola M, Masullo A, Discetti S, Ianiro A, Theunissen R, Buchlin JM (2017) Pod-based background removal for particle image velocimetry. Exp Therm Fluid Sci 80:181–192

Neal DR, Sciacchitano A, Smith BL, Scarano F (2015) Collaborative framework for PIV uncertainty quantification: the experimental database. Meas Sci Technol 26(7):074003. http://stacks.iop.org/0957-0233/26/i=7/a=074003

Raiola M, Discetti S, Ianiro A (2015) On piv random error minimization with optimal pod-based low-order reconstruction. Exp Fluids 56(4):75

Schmid PJ (2010) Dynamic mode decomposition of numerical and experimental data. J Fluid Mech 656:5–28

Schmidt E (1907) Zur theorie der linearen und nichtlinearen integralgleichungen. i teil. entwicklung willkurlichen funktionen nach system vorgeschriebener. Math Annal 63:433–476

Shabalin AA, Nobel AB (2013) Reconstruction of a low-rank matrix in the presence of Gaussian noise. J Multivar Anal 118:67–76

Sherry M, Nemes A, Jacono DL, Blackburn HM, Sheridan J (2013) The interaction of helical tip and root vortices in a wind turbine wake. Phys Fluids 2013:25

Venturi D (2006) On proper orthogonal decomposition of randomly perturbed fields with applications to flow past a cylinder and natural convection over a horizontal plate. J Fluid Mech 559:215–254

Venturi D, Karniadakis GE (2004) Gappy data and reconstruction procedures for flow past a cylinder. J Fluid Mech 519:315–336

Wang H, Gao Q, Feng L, Wei R, Wang J (2015) Proper orthogonal decomposition based outlier correction for PIV data. Exp Fluids 56(2):43

Xu Y, Feng LH, Wang JJ (2013) Experimental investigation of a synthetic jet impinging on a fixed wall. Exp Fluids 2013:54