A robust computational framework for variational data assimilation of mean flows with sparse measurements corrupted by strong outliers

Souvik Ghosh, Vincent Mons, Denis Sipp, Peter J. Schmid

Journal of Computational Physics 508 (2024) 113008

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

A robust computational framework for variational data assimilation of mean flows with sparse measurements corrupted by strong outliers

Souvik Ghosh [a,b,*], Vincent Mons [b], Denis Sipp [b], Peter J. Schmid [c]

[a] Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom
[b] DAAA, ONERA, Institut Polytechnique de Paris, 8 Rue des Vertugadins, 92190 Meudon, France
[c] Department of Mechanical Engineering, King Abdullah University of Science and Technology, 23955 Thuwal, Saudi Arabia

ARTICLE INFO

Keywords:
Outliers
Adjoints
ADMM
Huber loss
Hybrid loss

ABSTRACT

A variational framework for the reconstruction of time-averaged mean flows using a sparse set of observations with large magnitudes of noise (referred to as outliers), is presented. The observations constitute a set of point-wise measurements of the mean flow with outliers at certain measurement locations and are incorporated into a numerical simulation governed by the two-dimensional, incompressible Reynolds-averaged Navier-Stokes (RANS) equations with an unknown momentum forcing. This forcing, which corresponds to the divergence of the Reynolds stress tensor, is calculated from a direct-adjoint optimization procedure to reduce the deviation between the measured and estimated mean velocities. ℓ_2, ℓ_1, Huber, and hybrid loss functions are used to represent the discrepancy in the mean velocity field between the measurements and the predictions. A variety of algorithms are considered to solve the optimization problem with these loss functions and a performance comparison in terms of the quality and physical features of the recovered mean flow is presented. The Huber loss function performed best as it remained robust to strong outliers in the measurements with its ℓ_1 contribution and also ensured the uniqueness of the optimal solution with its ℓ_2 contribution. Huber loss functions restrict the effect of outliers at the local measurement locations, thereby not affecting the quality of the high-dimensional reconstructed mean flow field. The hybrid loss function, a modified form of the continuous Huber loss function, also recovered the mean flow with high accuracy. We demonstrate the performance of the data assimilation framework for the case of two-dimensional laminar flow around a circular cylinder at $Re = 100$. We then extend the analysis to the case of two-dimensional laminar flow over a backward-facing step at $Re = 500$, to further assess the efficacy and robustness of the data assimilation framework.

1. Introduction

Computations and experiments constitute the two principal means of investigating the mechanisms and transport processes governing complex fluid flows. Both techniques have undergone remarkable progress over the past decades and currently produce data of impressive detail and accuracy. While many studies focus on one of the two approaches, a combined perspective that

* Corresponding author.
E-mail address: s.ghosh17@alumni.imperial.ac.uk (S. Ghosh).

https://doi.org/10.1016/j.jcp.2024.113008
Received 15 October 2023; Received in revised form 9 April 2024; Accepted 10 April 2024

Available online 16 April 2024

leverages the advantages of both methods often yields a more complete and robust picture of the observed flow behavior. Merging information from experiments and computations is traditionally achieved by data assimilation – a technique that injects experimental measurements into numerical simulations to eliminate uncertainty and increase predictability. In this process, computational model predictions are matched to experimental observations, and corrections within the variance of the measurements are directly applied to the prescribed model. This procedure can be used, among other applications, to extend the predictability of flow features beyond the reach of sensors. One of the most challenging tasks in data assimilation is the handling of unavoidable errors (in particular, outliers many standard deviations away from the mean) in the measurements and the restricting of their influence on the remaining data set. The development of an effective computational framework to accomplish this objective is the goal of this article.

Data assimilation is an established technique of estimating a complete flow state from limited observations under the constraint of a model equation. It has a long and distinguished history, with much of the initial progress driven by the recovery of weather and climate patterns from local and sparse measurements using a circulation model (see, e.g., Le Dimet and Talagrand, 1986 [1]). These identified global patterns are then advanced in time to provide forecasts, and similar attempts have been made to determine dominant ocean currents from a localized buoy network or regional satellite data.

The data used in data-assimilation computations are varied in quality and call for different techniques to incorporate them into the underlying model to maximum effect and benefit. Experimental data is commonly time-accurate, but spatially sparse and localized; it is also high in model fidelity since it is the experimental truth, but often noisy, depending on the sensor's capabilities. Computational data, on the other hand, is spatially global and commonly noise-free but suffers from a far lower model fidelity. For example, flows in the turbulent regime often misrepresent quantities of interest due to the shortcomings of the used turbulence or subgrid-scale closure model. A promising direction of pursuit in data assimilation is thus a hybridized approach that combines complementary strengths, with a mutual compensation of known inaccuracies.

Data assimilation techniques can be roughly divided into sequential and variational. The sequential approach is based on the Kalman Filter and alternates forecast and analysis steps (see Kalman, 1960 [2]). While successful for systems with lower degrees of freedom, it becomes intractable when the state is high-dimensional. Even using ensemble averages of the statistics, a technique referred to as the Ensemble Kalman Filter (for details, see Evensen, 1994 [3] and Evensen, 2009 [4]) can only partially alleviate the limitations of the initial approach for large-scale dynamical systems, due to the inherent cost of generating and propagating the ensembles.

Variational data assimilation has its roots in optimal control theory (see, for example, Zhou et al., 1995 [5], Gunzburger [6], Kim and Bewley, 2007 [7]) where the objective is to minimize a loss functional subject to satisfying governing equations. In the context of variational data assimilation applied to fluid mechanics, this involves minimizing the discrepancy between the experimental observations and numerical simulations, while considering the governing equations as constraints. Put differently, the aim is to use a sparse data set of measurements and recover flow fields by optimally fitting an underlying model to the measurements. Under the label of 3D-Var/4D-Var problems, variational data assimilation has routinely and successfully been applied to weather forecasting, see, e.g., Talagrand and Courtier, 1987 [8]. Apart from meteorological applications, it has been used to determine boundary and initial conditions from PIV measurements (Gronskis et al., 2013 [9]), to reconstruct mean velocity fields from sparse measurements (Foures et al., 2014 [10], Symon et al., 2017 [11]), or higher-order dynamics from localized input (see Mons et al., 2016 [12]), where it surpassed sequential methods. Further examples include the data-based recovery of scalar transport (see Cerizza et al., 2016 [13], Wang et al., 2019 [14]) and the blending of variational and ensemble-based techniques in a series of efforts by Mons et al., 2019 [15], Mons et al., 2021 [16], and Buchta et al., 2022 [17].

Key challenges in variational data assimilation stem from the nature of experimental data. Measurements from experiments in fluid dynamics are commonly underresolved in space (and often also in time) and contain only partial information about the full state of the flow. In mathematical terms, the sparse data constitute a low-dimensional representation of an actual high-dimensional flow field. In particular, the recovery of mean flows from sparse measurements – a common requirement for the design of engineering fluid systems – is an example of this highly restricted representation. In addition to this representation, experimental data are often contaminated by statistical outliers, i.e., data points that, due to measurement or sensor errors, depart by multiple standard deviations from the mean value. Moreover, noise, again from faulty or low-quality sensors, presents a challenge for the recovery of meaningful and physically relevant results. The reader can refer Wang and Zaki, 2021 [18] and Zaki and Wang, 2021 [19], to get an insight into the difficulties associated with variational data assimilation of turbulent flows.

A computational framework for the robust recovery of mean-flow fields has to account for these obstacles. More specifically, we do not wish to follow the common procedure in dealing with outliers, sparsity, and noise by using pre-processing steps, such as filtering or outlier detection. Rather, we attempt to incorporate a judicious selection step, distinguishing between valuable information and compromised data, directly into the variational formulation. In this way, we aim to include *all* data points but limit their negative effect on an accurate recovery when their contribution exceeds a statistical threshold given by a particular loss functional. Consequently, we use advanced methodology from mixed loss function optimization, concepts from robust statistics, and techniques from inverse problems to design a computational platform for variational data assimilation in the presence of sparse, noisy, and outlier-contaminated data.

Our approach goes beyond the traditional approach of treating the error deviation in the variational formulation in a quadratic or ℓ_2 form, yielding a least-square problem from a mathematical point of view. Data sets with high kurtosis are particularly misrepresented by this choice. Instead, an ℓ_1-based loss functional in the variational formulation, i.e., an absolute-value-based error measure is far more suited to capture large standard deviations from the mean and thus yield more robust mean-flow recoveries. This type of variational problem, however, suffers from convergence problems in the optimization step owing to the singularity for vanishing errors and the lack of a proper definition of gradients at these points. In this present study, various hybrid ℓ_2/ℓ_1 loss functionals are

proposed, based on earlier work (see Huber, 1972 [20]; Huber, 1973 [21]; Bube and Langan, 1997 [22]; Tavolato and Isaksen, 2015 [23]; Zadorozhnyi et al., 2016 [24] for details) which allows us to account for the above-mentioned inaccuracies in the processed data and significantly robustify the traditional variational data assimilation approach. The application of the Huber loss function to simple data assimilation models using data with outliers has been investigated by Rao et al., 2017 [25]. In this paper, we demonstrate the application of the mixed loss function framework to complex flows. The use of this mixed loss function framework leads to complex optimization problems and sophisticated algorithms have to be employed to solve the resulting PDE-constrained minimization problem. We will showcase the approach in low-Reynolds number configurations, such as two-dimensional flow around a cylinder at $Re = 100$ and two-dimensional flow over a backward-facing step at $Re = 500$. Yet, extension to higher-number Reynolds numbers, which require a turbulence model [26], or to compressible flows does not pose any problem in principle.

The present paper is thus organized as follows. The general data assimilation problem is formulated in §2, which is followed by the introduction of different loss functions in §3. The numerical methods for optimization are reported in §4 while the error quantification is reported in §5. Results for flow around a circular cylinder and flow over a backward-facing step, together with discussions of the obtained results, are presented in §6. Conclusions are presented in §7.

2. Problem statement

2.1. Reynolds-Averaged Navier-Stokes (RANS) framework

Any unsteady flow $[\mathbf{u}, p]$ in a configuration statistically steady can be decomposed into a time-averaged steady mean flow $\left[\overline{\mathbf{u}}, \overline{p}\right]$ and fluctuating perturbations $\left[\mathbf{u}', p'\right]$, using the Reynolds decomposition. We represent $\mathbf{u} = [u, v, w]$, $\overline{\mathbf{u}} = \left[\overline{u}, \overline{v}, \overline{w}\right]$, and $\mathbf{u}' = \left[u', v', w'\right]$ respectively. Time-averaging of the three-dimensional incompressible Navier-Stokes equations results in the Reynolds-Averaged Navier-Stokes (RANS) equations:

$$\overline{\mathbf{u}} \cdot \nabla \overline{\mathbf{u}} + \nabla \overline{p} - \frac{1}{Re} \nabla^2 \overline{\mathbf{u}} - \mathbf{f} = 0 \tag{1a}$$

$$\nabla \cdot \overline{\mathbf{u}} = 0, \tag{1b}$$

which are valid in a two-dimensional domain.

The true (experimental or computational) mean flow is assumed to satisfy the two-dimensional incompressible RANS equations, considered the reduced-order model for data assimilation. In the above equation (1a), $\mathbf{f} = \left[f_u, f_v, f_w\right]$ represents the unknown volumetric momentum forcing term which can be modeled using a turbulence closure model or can be obtained using a data-driven approach. This forcing term accounts for the divergence of the Reynolds stresses, as seen from the Reynolds averaging of the Navier-Stokes equations. It is given as

$$\mathbf{f} = -\nabla \cdot \underline{\boldsymbol{\tau}} \tag{2}$$

where τ is the Reynolds-stress tensor.

The boundary conditions for the RANS equations are discussed later in section 6 for each flow case.

2.2. Data assimilation problem

Into the data assimilation framework, we embed methodology for effectively handling outliers in the measurements. Outliers are measurements that deviate from the mean value by abnormal multiples of the standard deviation. The precise definition of 'abnormal' is subjective and is left for the experimentalist to quantify. In the present work, we focus on variational data assimilation only and the purpose of data assimilation algorithms is thus to obtain an optimal estimate of the design control vector such that the error between the measured and reconstructed states, represented by an objective functional $\mathcal{M}$, becomes minimal. The forcing term $\mathbf{f}$ is considered as the design control vector, and the RANS equations are used as constraints for the inverse problem. Different types of loss functions to represent the deviation error between the measurements and the model predictions will be presented and discussed in the subsequent sections. The goal of the inverse problem can therefore be stated as follows

$$\min_{\mathbf{f}} \ \mathcal{M} \tag{3a}$$

$$\text{subject to} : \overline{\mathbf{u}} \cdot \nabla \overline{\mathbf{u}} + \nabla \overline{p} - \frac{1}{Re} \nabla^2 \overline{\mathbf{u}} - \mathbf{f} = 0 \tag{3b}$$

$$\nabla \cdot \overline{\mathbf{u}} = 0 \tag{3c}$$

Observations are measurements of the real mean flow field, available at discrete sensor locations and often contaminated by noise. In general, measurements $\widetilde{\mathbf{u}}$ of the true mean velocity field $\overline{\mathbf{u}}$ are expressed as

$$\mathcal{H}\left(\overline{\mathbf{u}}\right) = \widetilde{\mathbf{u}} + \mathbf{r} \tag{4}$$

where

$$\widetilde{\mathbf{u}} = \begin{bmatrix} \widetilde{u} \\ \widetilde{v} \end{bmatrix} = \begin{bmatrix} \left\{\widetilde{u}_i\right\}_{i=1}^{n} \\ \left\{\widetilde{v}_i\right\}_{i=1}^{n} \end{bmatrix} = \begin{bmatrix} \left[\widetilde{u}_1 \ \widetilde{u}_2 \ \dots \ \widetilde{u}_n\right]^{\mathrm{T}} \\ \left[\widetilde{v}_1 \ \widetilde{v}_2 \ \dots \ \widetilde{v}_n\right]^{\mathrm{T}} \end{bmatrix}, \text{ and} \tag{5a}$$

$$\mathbf{r} = \left\{r_i\right\}_{i=1}^{2n} = \left[\widetilde{r}_1\ \widetilde{r}_2\ \ldots\ \widetilde{r}_n\ \widetilde{r}_{n+1}\ \ldots\ \widetilde{r}_{2n}\right]^{\mathrm{T}} \tag{5b}$$

with i representing each discrete sensor location and n denoting the total number of measurement sensors. We perform two-dimensional mean flow reconstruction, and therefore, we set up the data assimilation problem for two components of mean velocity in equations (3). $\mathbf{r}$ represents the vector of observation errors in $\widetilde{u}$ and $\widetilde{v}$ respectively. $\mathcal{H} : V\left(\Omega^F\right) \rightarrow \mathbb{R}^{2n}$ is the representation of a general operator that projects the high-dimensional $V\left(\Omega^F\right)$ space of functions for the mean velocity field onto a discrete real space, compatible with the locations of the measurement sensors.

When measurements from experiments or simulations are limited, and the reconstructed field is higher-dimensional, the inverse problem may not yield smooth solutions. Furthermore, the presence of strong levels of noise in the measurements can render the inverse problems rather ill-posed. For an inverse problem to revert to well-posedness, a regularization function $\mathcal{N}(\mathbf{f})$ must be added to the objective functional. The inverse problem stated in equation (3) is a classic example of a constrained optimization problem which can be simplified into an unconstrained optimization problem with the introduction of Lagrangian multipliers that eliminate the implicit dependence of the state variable on the control vector and therefore, henceforth $\mathcal{M}$ will be referred to as an objective function in the Lagrangian as shown in equation (6). The goal is to determine the critical points of this Lagrangian functional

$$\mathcal{L}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],\mathbf{f}\right) = \mathcal{M}\left(\overline{\mathbf{u}}\right) + \mathcal{N}(\mathbf{f}) - \int_{\Omega^F} \overline{p}^{\dagger}\left\{\nabla\cdot\overline{\mathbf{u}}\right\}\, d\Omega - \int_{\Omega^F}\left(\overline{\mathbf{u}}^{\dagger}\right)^{\mathrm{T}}\left\{\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}} + \nabla\overline{p} - \frac{1}{Re}\nabla^2\overline{\mathbf{u}} - \mathbf{f}\right\}\, d\Omega \tag{6}$$

where the spatial integration is defined on the high-dimensional flow domain.

Lagrange multipliers or adjoint variables $\left(\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right)$ are introduced in equation (6) to enforce adherence to the governing equations. The regularization function defined on the high-dimensional assimilated space is a function of the design control vector ($\mathbf{f}$) and is of the following form

$$\mathcal{N}(\mathbf{f}) = \frac{\beta}{2}\int_{\Omega^F} |\nabla\mathbf{f}|^2\, d\Omega \tag{7}$$

The regularization parameter also referred to as the balancing parameter ($\beta > 0$), is taken as sufficiently small to produce a proper weighting between the deviation error and the balancing error (arising from the regularization function).

The first-order necessary conditions for the optimal solution of a non-linear optimization problem are obtained by solving the Karush–Kuhn–Tucker (KKT) system of the above equations (Kuhn and Tucker, 1951 [27]). Setting the first variation of the augmented cost functional to zero results in the set of KKT equations that are then solved iteratively. The adjoint equations (obtained from the partial differential of the Lagrangian with respect to the direct variables) are presented in equations (8a). The partial differential of the Lagrangian with respect to the design control vector $\mathbf{f}$ is presented in equation (8b) which provides the necessary gradient information for the optimization algorithm.

$$\frac{\partial\mathcal{L}}{\partial\overline{\mathbf{u}}} = 0 \Rightarrow -\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}}^{\dagger} + \overline{\mathbf{u}}^{\dagger}\cdot\left(\nabla\overline{\mathbf{u}}\right)^{\mathrm{T}} - \nabla\overline{p}^{\dagger} - \frac{1}{Re}\nabla^2\overline{u}^{\dagger} = \frac{\partial\mathcal{M}}{\partial\overline{\mathbf{u}}}, \text{ and } \frac{\partial\mathcal{L}}{\partial\overline{p}} = 0 \Rightarrow \nabla\cdot\overline{\mathbf{u}}^{\dagger} = 0 \tag{8a}$$

$$\nabla_{\mathbf{f}}\mathcal{L} = \overline{\mathbf{u}}^{\dagger} + \beta\nabla^{\mathrm{T}}\nabla\mathbf{f} \tag{8b}$$

The boundary conditions for the direct and adjoint equations are discussed in detail for each flow case in Section 6.

3. Objective function

This section provides details on the different types of loss functions used to quantify the point-wise deviation error between the model predicted and the observed mean velocities. Before introducing the different loss functions, we provide a quick introduction to the Bayesian framework of deterministic data assimilation.

3.1. Deterministic formulation for a Bayesian perspective on variational data assimilation

A Bayesian approach to data assimilation (Wikle and Berliner, 2007 [28]) employs Bayes' theorem to obtain an optimal estimate from observations $\widetilde{\mathbf{u}}$ resulting in an associated assimilated state $\overline{\mathbf{u}}$. Following this approach, we can define a continuous form of the assimilated state that can be viewed as a random variable drawn from a probability distribution, defined by its probability density function (PDF) and conditioned on observations according to

$$\mathbb{P}\left(\overline{\mathbf{u}} \mid \widetilde{\mathbf{u}}\right) = \frac{\mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}\right)\ \mathbb{P}\left(\overline{\mathbf{u}}\right)}{\mathbb{P}\left(\widetilde{\mathbf{u}}\right)} \propto \mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}\right)\ \mathbb{P}\left(\overline{\mathbf{u}}\right) \tag{9}$$

From RANS equation (1), we see that the assimilated state $\overline{\mathbf{u}}$ is an implicit function of the control variable $\mathbf{f}$, denoted as $\overline{\mathbf{u}}(\mathbf{f})$. We can write equation (9) as

$$\mathbb{P}\left(\overline{\mathbf{u}}(\mathbf{f}) \mid \widetilde{\mathbf{u}}\right) = \frac{\mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}(\mathbf{f})\right)\ \mathbb{P}(\mathbf{f})}{\mathbb{P}\left(\widetilde{\mathbf{u}}\right)} \propto \mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}(\mathbf{f})\right)\ \mathbb{P}(\mathbf{f}) \tag{10}$$

where $\mathbb{P}(\mathbf{f})$ represents the prior PDF (i.e., knowledge of the control state before new observations), $\mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}(\mathbf{f})\right)$ denotes the likelihood of observations (i.e., the PDF of observations, conditioned on the assimilated state), $\mathbb{P}\left(\overline{\mathbf{u}}(\mathbf{f}) \mid \widetilde{\mathbf{u}}\right)$ stands for the posterior PDF (i.e., the updated PDF of the assimilated state, after the analysis), and $\mathbb{P}\left(\widetilde{\mathbf{u}}\right)$ represents the marginal PDF of the observations, a normalizing constant which ensures that the total probability sums to one.
Equation (10) states that the prior probability is updated after observing the likelihood of the measurement and assigning an update to the posterior probability. The Bayesian viewpoint consists of processing the measurement data to maximize the probability of an assimilated state that is most likely to corroborate the measured data. This implies that the Bayesian approach determines argmax of $\mathbb{P}\left(\overline{\mathbf{u}}(\mathbf{f}) \mid \widetilde{\mathbf{u}}\right)$ in equation (10), i.e., we seek a $\mathbf{f}$ which yields $\overline{\mathbf{u}}$ that is most probable, given the measurement data $\widetilde{\mathbf{u}}$. This step is referred to as a maximum likelihood estimation (MLE), and it is achieved by maximizing a likelihood function such that, under the assumed probability distribution, the observed (measured) data emerge as most probable. The maximum likelihood hypothesis is given by the following form

$$\mathbf{f}_{\text{opt}} = \underset{\mathbf{f}}{\text{argmax}} \ \mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}(\mathbf{f})\right) \ \mathbb{P}(\mathbf{f}) \tag{11}$$

Considering equation (11), the most likely assimilated state $\mathbf{f}_{\text{opt}}$ is the one for which the conditional probability of the measurement data (given the assimilated state) reaches a maximum. The Bayesian perspective thus allows the modeling of the observation error based on the nature of a probability distribution. Since the observation error (noise) can be taken as a random variable, we can consider different loss functions for modeling the noise based on the associated probability distribution.
We can use the log-likelihood transformation of the maximum likelihood estimate in equation (11) as

$$\mathbf{f}_{\text{opt}} = \underset{\mathbf{f}}{\text{argmax}} \ \ln \mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}(\mathbf{f})\right) + \ln \mathbb{P}(\mathbf{f}) \tag{12}$$

where maximization of $\mathbb{P}$ is equivalent to maximization of $\ln(\mathbb{P})$ as $\ln(\mathbb{P})$ constitutes a monotonic function of $\mathbb{P}$.
The maximization problem in equation (12) can also be written as a minimization problem of the corresponding negative value of the log-likelihood estimate as

$$\mathbf{f}_{\text{opt}} = \underset{\mathbf{f}}{\text{argmin}} \ -\ln \mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}(\mathbf{f})\right) \ -\ln \mathbb{P}(\mathbf{f}), \tag{13}$$

where the two terms respectively correspond to the $\mathcal{M}$ and $\mathcal{N}$ parts of the objective function in (6).
We denote the general expression for $\mathbb{P}\left(\widetilde{\mathbf{u}} \mid \overline{\mathbf{u}}(\mathbf{f})\right)$ as $\mathbb{F}(\mathbf{r})$ for the observation error in the measurements. The general expression for the observation loss function is therefore based on the PDF of the observation error distribution (Lorenc, 1986 [29]) denoted as

$$\mathbf{f}_{\text{opt}} = \underset{\mathbf{f}}{\text{argmin}} \ -\ln \mathbb{F}(\mathbf{r}) - \ln \mathbb{P}(\mathbf{f}) \tag{14}$$

Based on the above definition, the measurement error part of the loss function for different error distributions is discussed in what follows. While defining the error distribution, we consider independent measurements to neglect the covariance between them and only consider the auto-covariance of different measurements.

3.2. Quadratic loss function

For a more general quadratic loss, the observation error is given by a density function as

$$\frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{|r_i|^2}{{\sigma_i}^2}\right) \tag{15}$$

where σ_i is the standard deviation of the observation error at any particular measurement location. In our analysis, we consider the scaled form of the random observation error at each measurement point such that the value of σ_i is normalized to 1. When we assume that the observation error's likelihood at each measurement location follows a Gaussian distribution, the objective is represented in terms of a ℓ_2 loss function, see Table 1. However, we will present a test case with Gaussian error in section 6.1.4 wherein the observation error has a standard deviation of σ_i at a particular measurement location.
A ℓ_2 loss function performs well when the minimization function has deviation errors within three standard deviations from the mean. The quadratic loss function is used to minimize the mean square error between the model output and the observation, and it is convex with respect to the direct variable $\overline{\mathbf{u}}$. However, in our case, the control variable is $\mathbf{f}$ which differs from the direct variable used in the objective function. Hence, no definitive statement on the convexity of the inverse problem in equation (6) can be made.

3.3. Absolute loss function

When there are large deviation errors, we assume that the observation's likelihood at each measurement location follows a Laplace distribution, and the objective is represented in terms of a ℓ_1 loss function, see Table 1. The partial differential with respect to the direct variable is represented by a sign function denoted by sgn( ) and is a piece-wise function.
A Laplace distribution follows a double exponential distribution with fatter tails than a Gaussian distribution, which allows efficient

Table 1
Error density function, objective and partial differential to the direct variable, for different loss functions.

Loss function	Error density function $\mathbb{F}(r_i)$	Objective $\mathcal{M}$	Partial differential $\frac{\partial \mathcal{M}}{\partial \overline{\mathbf{u}}}$
Quadratic	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\lvert r_i\rvert^2\right)$	$\sum_{i=1}^{2n} \frac{1}{2}\lvert r_i\rvert^2$	$\mathcal{H}^{\mathrm{T}}\mathbf{r}$
Absolute	$\frac{1}{2}\exp\left(-\lvert r_i\rvert\right)$	$\sum_{i=1}^{2n} \lvert r_i\rvert$	$\mathcal{H}^{\mathrm{T}}\left[\mathrm{sgn}(\mathbf{r})\right]$ with $\mathrm{sgn}(\mathbf{r}) = \begin{cases} -1, & r_i < 0 \\ 0, & r_i = 0 \\ 1, & r_i > 0 \end{cases}$
Huber	$\begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\lvert r_i\rvert^2\right), & \lvert r_i\rvert \le \epsilon_i \\ \frac{1}{\sqrt{2\pi}} \exp\left(\frac{(\epsilon_i)^2}{2} - \epsilon_i \lvert r_i\rvert\right), & \lvert r_i\rvert > \epsilon_i \end{cases}$	$\sum_{i=1}^{2n} \begin{cases} \frac{1}{2}\lvert r_i\rvert^2, & \lvert r_i\rvert \le \epsilon_i \\ \epsilon_i \lvert r_i\rvert - \frac{(\epsilon_i)^2}{2}, & \lvert r_i\rvert > \epsilon_i \end{cases}$	$\mathcal{H}^{\mathrm{T}}\widetilde{\mathbf{r}}$ with $\widetilde{\mathbf{r}} = \begin{cases} r_i, & \lvert r_i\rvert \le \epsilon_i \\ \epsilon_i \,\mathrm{sgn}(r_i), & \lvert r_i\rvert > \epsilon_i \end{cases}$
Hybrid	$\frac{1}{2\mathcal{X}\epsilon_i} \exp\left(-\left\{\left[1 + \frac{\lvert r_i\rvert^2}{\epsilon_i^2}\right]^{1/2} - 1\right\}\right)$	$\sum_{i=1}^{2n}\left[\left(1 + \frac{\lvert r_i\rvert^2}{\epsilon_i^2}\right)^{1/2} - 1\right]$	$\mathcal{H}^{\mathrm{T}}\left[\frac{\lvert\mathbf{r}\rvert}{\epsilon^2}\left(1 + \frac{\lvert\mathbf{r}\rvert^2}{\epsilon^2}\right)^{-1/2}\right]$

handling of observation points that are many standard deviations from the mean. The absolute loss function is used to minimize the mean absolute error between the model output and the observation. The absolute loss function is more robust to outliers as it weighs the deviation error on a linear scale as compared to square values for the quadratic loss function. However, the ℓ_1 loss function will not be as effective in handling outliers, as large errors from outliers are weighed the same as lower errors. Moreover, the absolute loss function suffers from a singularity at the point of vanishing deviation error and hence is not differentiable at that point. Due to the complexity of non-smoothness, it poses challenges to conventional optimization methods. There have been various approaches to solving this kind of non-smooth optimization problems such as linear programming (Barrodale and Roberts, 1980 [30]) or iterative smoothing (Scales et al., 1988 [31]), but they require a considerable amount of mathematical tampering to eliminate the singularity. This issue prompted the development of hybrid ℓ_2/ℓ_1 loss functions which are characterized by a robust treatment of large deviation errors while following a Gaussian distribution for small deviation errors.

3.4. Huber loss function

The observation error's likelihood at each measurement location may follow a piece-wise mixed Gaussian and Laplace distribution. In this case, the parameter ϵ_i denotes the epsilon parameter at each measurement sensor which dictates the switch from a ℓ_1 to a ℓ_2 loss function.

The Huber loss function was first proposed by Huber, 1964 [32] as one of the hybrid ℓ_1/ℓ_2 estimators for contaminated normal distributions. It enables the treatment of observations with large deviations (outliers) at the expense of minimal risk of corrupting the analysis locally, as shown by Huber, 1972 [20] and Tavolato & Isaksen, 2009 [33]. The asymptotic properties, robustness, stability, and convexity of the piece-wise Huber loss function are discussed in detail in Huber, 1972 [20] and Huber, 1973 [21]. The Huber loss function follows a Laplace distribution in the limit of large deviation errors (see Table 1), blended into a Gaussian distribution in the vicinity of vanishing deviation errors. The Huber loss function has two advantages: first, it maintains a quadratic function near the mean and evaluates the solution with minimal variance around the mean with its ℓ_2 part, and, second, it mitigates the weights assigned to the outliers for large error values with its ℓ_1 part. The value where the Huber function switches from ℓ_1 to ℓ_2 is given by the epsilon parameter which needs to be estimated *a priori*. We will present a numerical methodology to estimate the value of this epsilon parameter in Appendix D.

While the Huber function is continuous as well as differentiable at the $\ell_1 - \ell_2$ interface, however, it is only first-order differentiable and hence cannot be subjected to a numerical algorithm that requires second-moment information (Fountoulakis and Gondzio, 2016 [34]). The Huber loss function, like the absolute function, is not strictly convex to the direct variable $\overline{\mathbf{u}}$ (Bube and Nemeth, 2007 [35]).

3.5. Hybrid loss function

If we assume that the observation error's likelihood at each measurement location follows a continuous mixed Gaussian and Laplace distribution, the objective is represented in terms of a continuous and differentiable function. In Table 1, the value of $\mathcal{X}$ for the hybrid loss density function is given as

$$\mathcal{X} = \int_0^\infty \exp\left(-t_i\right) \frac{t_i + 1}{\sqrt{t_i^2 + 2t_i}}\, dt_i \approx 1.636, \text{with } t_i = \left[1 + \frac{\lvert r_i\rvert^2}{\epsilon_i^2}\right]^{1/2} - 1 \tag{16}$$

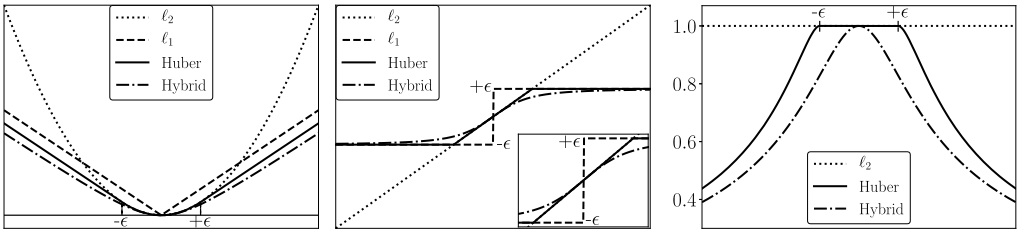

Fig. 1. (a) Objective, (b) gradient, and (c) weight for different loss functions used to represent the point-wise deviation error between the measurements and the predictions. The weights in (c) are obtained by normalizing the objective of each loss function with respect to the objective of the ℓ_2 loss function. The inset in (b) shows the zoomed-in view of the gradients in the vicinity of the epsilon/cut-off parameter.

Table 2
Relative weight for Huber and hybrid loss functions as shown in Fig. 1c. The weight function is obtained by normalizing the objective of each loss function with respect to the objective of the ℓ_2 loss function.

Loss function	ℓ_2	Huber	Hybrid
Weight function, $\mathcal{W}(r_i)$	1	$1;\ \lvert r_i\rvert \le \epsilon_i$ $2\frac{\epsilon_i}{\lvert r_i\rvert};\ \lvert r_i\rvert > \epsilon_i$	$\frac{1}{\epsilon_i^2};\ \lvert r_i\rvert \ll \epsilon_i$ $\frac{2}{\epsilon_i\lvert r_i\rvert};\ \lvert r_i\rvert \gg \epsilon_i$

The hybrid loss function is called the pseudo-Huber loss function (Bube and Langan, 1997 [22]). It is a smooth approximation of the Huber loss function and approximates the Huber function in different asymptotic limits of the deviation error. In Table 1, $\boldsymbol{\epsilon} = \{\epsilon_i\}_{i=1}^{2n}$ represents the vector of cut-off parameters where ϵ_i denotes the ℓ_1/ℓ_2 switch at a particular measurement location. We will present a numerical methodology to estimate the value of this cut-off parameter in Appendix D.
The hybrid loss function in Table 1 reduces to ℓ_2 and ℓ_1 functions in different asymptotic limits as follows.

$$\mathcal{M} = \sum_{i=1}^{2n} \begin{cases} \frac{1}{2}\frac{\lvert r_i\rvert^2}{\epsilon_i^2} & , \frac{\lvert r_i\rvert}{\epsilon_i} \ll 1 \\ \frac{\lvert r_i\rvert}{\epsilon_i} & , \frac{\lvert r_i\rvert}{\epsilon_i} \gg 1 \end{cases} \tag{17}$$

The problem of the non-existence of higher-order derivatives for the Huber loss function has been circumvented in the hybrid loss function as it can furnish derivatives of any order. Fig. 1a shows the different loss functions used to approximate the deviation error between the measurements and the model predictions. For the cases of Huber and hybrid loss functions, there is an epsilon/cut-off regime within which the function behaves in an ℓ_2 manner and ℓ_1 everywhere else. The associated gradient for the different loss functions is presented in Fig. 1b where we notice a discontinuity for the ℓ_1 loss function at the minimal point. The discontinuity in the ℓ_1 loss function has been removed with the use of the Huber or hybrid loss function where the gradient near the point of epsilon/cut-off ($\pm\epsilon$) is far smoother. Finally, the weights of the different loss functions for each measurement point are displayed by normalizing the objective of each loss function with respect to the objective of the ℓ_2 loss function, see Fig. 1c. The weights vary between zero and one; they quantify how an impact of a strong deviation in the measurement is de-emphasized (or even discarded) in the data assimilation process. A mathematical representation of the objective and the associated weights for each loss function are represented in Table 2. The quadratic loss function assigns equal weights (= 1) to every point in the measurement data set irrespective of outliers. Fig. 1c shows that the Huber and hybrid loss functions assign reduced weights to the model-predicted values with large departures from the mean observation value, in contrast to the full weights assigned by the ℓ_2 loss function throughout the range of the deviation error. The Huber loss function gives full weights to the predicted values near the minimum and gradually decreasing weights away from the minimum, as shown in Table 2. The weighting distribution for the hybrid loss function remains the same – except near the minimum where it narrows to very few predicted values around the mean – with full weights (see Table 2 for details).

4. Numerical methods for optimization

This section presents a class of algorithms employed for numerical optimization. The inverse problem involving ℓ_2, ℓ_1, Huber, and hybrid loss functional are solved using a variational formulation with a gradient-based iterative scheme, very similar to a 3D-Var problem as shown by Lewis et al., 2006 [36]. The iterative optimization scheme has to be supplemented with an initial guess, and for all cases of optimization, the initial guess is taken as $\mathbf{f} = \vec{0}$, unless stated otherwise.

4.1. Quasi-Newton Method (L-BFGS) for ℓ_2 differentiable loss functions

Quasi-Newton methods are a good alternative to the Newton methods and have been used widely for solving non-linear unconstrained optimization problems. Limited-memory BFGS (L-BFGS) has been one of the most popular quasi-Newton methods. It has been shown by Franceschini et al., 2020 [26] that the Euclidean inner product defined for L-BFGS (Nocedal and Wright, 2006 [37]) is inconsistent with the inner product defined for the inverse problem, and therefore a symmetric positive weighting matrix based on the metric of the finite-element spatial discretization is considered. A Cholesky decomposition of this mass matrix $\mathscr{M}$ is performed as $\mathscr{M} = \mathscr{L}\mathscr{L}^{\mathrm{T}}$, and the Cholesky factor $\mathscr{L}$ is used to precondition the design vector as well as the gradient of the L-BFGS optimization problem. For the L-BFGS optimization algorithm, a weighted form of the design vector $\left(\widetilde{\mathbf{f}} = \mathscr{L}\mathbf{f}\right)$ and the gradient, $\left(\widetilde{\mathscr{G}} = \mathscr{L}^{\mathrm{T}}\mathscr{G}\right)$ are considered. This weighting technique has been proven to be more efficient, see Franceschini et al., 2020 [26] for more details.
The L-BFGS method has been shown to work rather well for the non-linear least-squares problems (Murea, 2005 [38]), and the efficiency of this method has been extended to the case of an IRLS formulation by Bube and Langan, 1997 [22]. The utility of the L-BFGS method for a Huber loss function has been demonstrated by Guitton and Symes, 2003 [39] and for a hybrid loss function by Bube and Nemeth, 2007 [35].

4.2. Iteratively Reweighted Least-Squares (IRLS) for ℓ_1 loss functions

An efficient way to remove the singularity problem in ℓ_1 loss function is to cast the original problem into a form of weighted least-squares problem which can then be solved efficiently as an Iteratively Reweighted Least-Squares (IRLS) problem, see Bube and Langan, 1997 [22]. The equivalent objective function for the IRLS form of a ℓ_1 minimization problem and the partial differential with respect to the direct variable can be written as

$$\mathcal{M} = \sum_{i=1}^{2n} \frac{1}{2} r_i W_{i,i} r_i, \text{ and } \frac{\partial \mathcal{M}}{\partial \overline{\mathbf{u}}} = \mathcal{H}^{\mathrm{T}}[\mathbf{W}\mathbf{r}] \tag{18}$$

where $\mathbf{W}_{(2n\times 2n)}$ is a diagonal matrix of weights with $W_{i,i} = \dfrac{1}{|r_i|}$.
The steps for solving the IRLS problem in equation (18) in an iterative manner are as follows:
1. Set $k := 0$ and initialize the weighting matrix as $\mathbf{W}^{\{k=0\}} = \mathbb{I}_{(2n\times 2n)}$ where $\mathbb{I}$ represents an identity matrix.
2. Solve the IRLS problem in equation (18) for the $(k)^{th}$ iteration with $\mathbf{W}^{\{k\}}$ using the quasi-Newton L-BFGS algorithm.
3. Compute the residual at the k^{th} iteration as $\mathbf{r}^{\{k\}} = \mathcal{H}\left(\overline{\mathbf{u}}^{\{k\}}\right) - \widetilde{\mathbf{u}}$.
4. Compute the modified weighting matrix $\mathbf{W}^{\{k+1\}}$ using $\mathbf{r}^{\{k\}}$ with weights as $W_{i,i}^{\{k+1\}} = \dfrac{1}{\left|r_i^{\{k\}}\right|}$.
5. Set $k := k+1$ and go to step 2.

The above formulation can lead to singularities when the absolute error deviation is zero. To avoid this, a stabilizing parameter is considered to keep the elements of the weighting matrix finite. We have

$$W_{i,i}^{\{k+1\}} = \frac{1}{\max\left\{\eta_i, \left|r_i^{\{k\}}\right|\right\}} \quad \text{or} \quad W_{i,i}^{\{k+1\}} = \frac{1}{\left\{\eta_i^2 + \left|r_i^{\{k\}}\right|^2\right\}^{1/2}} \tag{19}$$

where $\eta_i > 0$ is a small positive value, and $\boldsymbol{\eta} = \left\{\eta_i\right\}_{i=1}^{2n}$ is the vector of positive values.

4.3. Alternating Direction Method of Multipliers (ADMM) for ℓ_2, ℓ_1 and Huber loss functions

The Alternating Direction Method of Multipliers (ADMM) (Nocedal and Wright, 2006 [37], and Boyd and Vandenberghe, 2004 [40]) is an effective approach that sequentially performs the minimization, thereby allowing a better treatment of large-scale complex non-linear constrained optimization problems (Dhingra et al., 2014 [41]). The advantage of this method is the breakdown of the multi-variable minimization problem into separable unconstrained primal variable problems with the specification of the dual variables which establishes the connection between the different primal variables (see Boyd et al., 2010 [42], and Parikh and Boyd, 2013 [43] for further details).
The optimization problem using ℓ_2, ℓ_1 and Huber loss functions can be solved using ADMM by introducing a transformed variable vector $\mathbf{z} = \left\{z_i\right\}_{i=1}^{2n}$ accounting for point-wise deviation and formulating it as a primal-dual problem given as

$$\min_{\mathbf{f}} \ \mathcal{M}(\mathbf{z}) \tag{20a}$$

$$\text{subject to} : \mathbf{z} = \mathcal{H}\left(\overline{\mathbf{u}}\right) - \widetilde{\mathbf{u}} \tag{20b}$$

$$\overline{\mathbf{u}} \cdot \nabla\overline{\mathbf{u}} + \nabla\overline{p} - \frac{1}{Re}\nabla^2\overline{\mathbf{u}} - \mathbf{f} = 0 \tag{20c}$$

$$\nabla \cdot \overline{\mathbf{u}} = 0 \tag{20d}$$

The Lagrangian form for equations (20a) - (20d) can be written as,

$$\begin{aligned}\mathcal{L}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],[\mathbf{z},\lambda],\mathbf{f}\right)=&\mathcal{M}(\mathbf{z})+\mathcal{N}(\mathbf{f})-\lambda^{\mathrm{T}}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right)-\widetilde{\mathbf{u}}-\mathbf{z}\right]-\int_{\Omega^{F}}\left(\overline{\mathbf{u}}^{\dagger}\right)^{\mathrm{T}}\left\{\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}}+\nabla\overline{p}-\frac{1}{Re}\nabla^{2}\overline{\mathbf{u}}-\mathbf{f}\right\}\,d\Omega\\&-\int_{\Omega^{F}}\overline{p}^{\dagger}\left\{\nabla\cdot\overline{\mathbf{u}}\right\}\,d\Omega+\frac{\mu}{2}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right)-\widetilde{\mathbf{u}}-\mathbf{z}\right]^{\mathrm{T}}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right)-\widetilde{\mathbf{u}}-\mathbf{z}\right]\end{aligned}\tag{21}$$

Upon simplification, the above equation (21) can be re-written as,

$$\begin{aligned}\mathcal{L}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],[\mathbf{z},\lambda],\mathbf{f}\right)=&\mathcal{M}(\mathbf{z})+\mathcal{N}(\mathbf{f})-\int_{\Omega^{F}}\left(\overline{\mathbf{u}}^{\dagger}\right)^{\mathrm{T}}\left\{\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}}+\nabla\overline{p}-\frac{1}{Re}\nabla^{2}\overline{\mathbf{u}}-\mathbf{f}\right\}\,d\Omega-\int_{\Omega^{F}}\overline{p}^{\dagger}\left\{\nabla\cdot\overline{\mathbf{u}}\right\}\,d\Omega\\&+\frac{\mu}{2}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right)-\widetilde{\mathbf{u}}-\mathbf{z}-\frac{\lambda}{\mu}\right]^{\mathrm{T}}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right)-\widetilde{\mathbf{u}}-\mathbf{z}-\frac{\lambda}{\mu}\right]-\frac{1}{2\mu}\lambda^{\mathrm{T}}\lambda\end{aligned}\tag{22}$$

where $\lambda=\left\{\lambda_i\right\}_{i=1}^{2n}$ represents the Lagrange multiplier, $\mu>0$ is a scalar penalty parameter, and the last term in equation (21) represents a facilitating term that helps to make the dual function differentiable under mild conditions for the optimization problem. The convergence of ADMM is assured under mild conditions on $\mathcal{M}$ and $\mathcal{N}$ and for all values of μ (see Boyd et al., 2010 [42] for further details).

The ADMM algorithm involves inner and outer iterations. While the inner iterations find an optimal iterate that minimizes the objective function for a specific value of μ, the outer iterations modify μ and update the Lagrange multiplier λ based on specific conditions. The Lagrangian in the above equation (22) can be solved using ADMM, and the basic structure of the ADMM iterative optimization is presented in equations (23a) - (23c).

$$\mathbf{f}^{\{k+1\}}=\underset{\mathbf{f}}{\operatorname{argmin}}\ \mathcal{L}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],\left[\mathbf{z}^{\{k\}},\lambda^{\{k\}}\right],\mathbf{f}\right)\tag{23a}$$

$$\mathbf{z}^{\{k+1\}}=\underset{\mathbf{z}}{\operatorname{argmin}}\ \mathcal{L}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],\left[\mathbf{z},\lambda^{\{k\}}\right],\mathbf{f}^{\{k+1\}}\right)\tag{23b}$$

$$\lambda^{\{k+1\}}=\lambda^{\{k\}}-\mu^{\{k\}}\left(\mathcal{H}\left(\overline{\mathbf{u}}\left(\mathbf{f}^{\{k+1\}}\right)\right)-\widetilde{\mathbf{u}}-\mathbf{z}^{\{k+1\}}\right)\tag{23c}$$

The iterative optimization problem in equation (23a) is solved using the quasi-Newton L-BFGS algorithm. The steps of ADMM iterations are mentioned in Algorithm 1. The convergence of ADMM is guaranteed by the primal and dual conditions as stated in equations (24) and (25), see Appendix C for the details on the derivation.

$$\text{Primal condition}:\mathcal{H}\left(\overline{\mathbf{u}}\left(\mathbf{f}^{\{k+1\}}\right)\right)-\widetilde{\mathbf{u}}-\mathbf{z}^{\{k+1\}}\tag{24}$$

$$\text{Dual condition}:\overline{\mathbf{u}}_{-}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right)\tag{25}$$

The primal and dual conditions are not enforced explicitly in the ADMM iterative algorithm. The primal-dual condition residuals will approach zero when the ADMM algorithm has achieved the convergence i.e. found the optima of the equations (23a) - (23c).

4.4. Solvability of the inverse problem

In the case where the available measurements are of the same dimension as the reconstructed mean flow field, the inverse problem can be solved with ease. However, when measurements are limited, and the reconstructed mean flow field is higher-dimensional than the available measurements, the inverse problem may not yield smooth solutions as a function of the control variable. In addition, the loss function used to minimize the observation error may not be convex as a function of the control variable $\mathbf{f}$, and hence, the inverse problem can yield mean flows that can be a local minimum solution. For the inverse problem to facilitate a smooth solution, we include the information from the prior PDF (see section 3.1) in the form of a regularization function. Hence we manage to regularize the smoothness of the solution of the inverse problem but we do not manage to make it convex.

Furthermore, the presence of outliers in the measurements can render the inverse problem rather ill-posed. Even though the Lagrangian formulation ensures that the RANS equations along with the boundary conditions are obeyed throughout the flow domain and the observation error at each measurement location is minimized, it cannot be guaranteed that the assimilated mean flow is unique. We did not perform a parametric study to analyze the effect of different initial conditions on the convergence to a unique mean flow state, and neither did we study the effect of stochastic gradients in the minimization algorithm (see [44] for further details). However, we considered an initial guess for the inverse problem to be the base flow which is very far from the true mean flow, and we managed to revert the inverse problem to well-posedness.

In terms of the existence of the solution of the inverse problem, we will show that with the introduction of an appropriate loss function, we could recover mean flow with sufficient accuracy in the presence of measurements with outliers, in section 6.

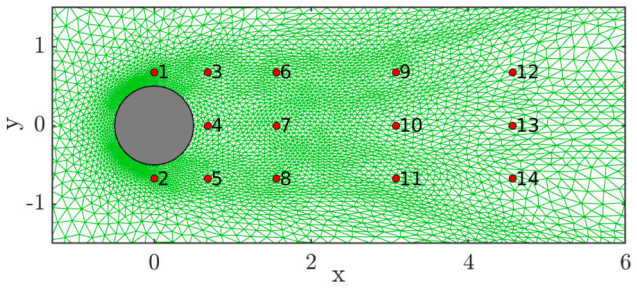

Fig. 2. Part view of the mesh for two-dimensional flow around a circular cylinder at $Re = 100$. The measurement locations are denoted by •. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

5. Error quantification

For the introduction of outliers at any true measurement $\widehat{a}_i$, we simply chose the actual measurement $\widetilde{a}_i$ to be:

$$\widetilde{a}_i = \left(1 + \delta_i\right) \widehat{a}_i \tag{26}$$

where δ_i is the fractional uncertainty level, whose value will be arbitrarily chosen.
Here $\widetilde{a}_i$ and $\widehat{a}_i$ represent the horizontal or vertical component of the mean velocity measurement at a particular sensor location with and without outliers, respectively. $\boldsymbol{\delta} = \left\{\delta_i\right\}_{i=1}^{m}$ is the vector of fractional uncertainties where m represents the number of sensor locations with outliers. In addition, we also present the absolute uncertainty which is defined as, $\Delta_i = \left|\widetilde{a}_i - \widehat{a}_i\right|$ with the vector of absolute uncertainties denoted as $\boldsymbol{\Delta} = \left\{\Delta_i\right\}_{i=1}^{m}$ respectively.
To assess the performance of the data assimilation procedure, we define a parameter $\left(\mathcal{G}_r^{\ell_2}\right)$ and call it the normalized global assimilated error. We define $\mathcal{G}_r^{\ell_2}$ as

$$\mathcal{G}_r^{\ell_2} = \frac{\mathcal{G}^{\ell_2}}{\mathcal{G}_0^{\ell_2}} = \left[\frac{\displaystyle\int_{\Omega^{\mathcal{E}}} \left|\overline{\mathbf{u}} - \widehat{\mathbf{u}}\right|^2 \, d\Omega}{\displaystyle\int_{\Omega^{\mathcal{E}}} \left|\overline{\mathbf{u}}^{\{0\}} - \widehat{\mathbf{u}}\right|^2 \, d\Omega}\right]^{1/2} \tag{27}$$

where, $\mathcal{G}^{\ell_2}$ denotes the norm of the global assimilated mean flow error, and $\mathcal{G}_0^{\ell_2}$ denotes the norm of the global base flow error. $\overline{\mathbf{u}}^{\{0\}}$ denotes the base flow with $\overline{\mathbf{u}}\left(\mathbf{f}^{\{0\}}\right)$ and $\widehat{\mathbf{u}}$ denotes the true mean flow field. The control volume considered for integration in the above equation is defined on a part of the flow domain $\left(\Omega^{\mathcal{E}} \subset \Omega^{\mathcal{T}}\right)$ that encompasses all the measurement sensors as well as the essential mean flow features, and this is defined explicitly for each flow case in section 6.
Similarly, we define another parameter $\mathcal{M}_r$ for each loss function and call it the normalized local measurement error, and define it as $\mathcal{M}_r = \dfrac{\mathcal{M}}{\mathcal{M}_0}$, where $\mathcal{M} = \mathcal{M}\left(\overline{\mathbf{u}}\right)$ denotes the objective evaluated with the assimilated mean flow, and $\mathcal{M}_0 = \mathcal{M}\left(\overline{\mathbf{u}}^{\{0\}}\right)$ denotes the objective evaluated with base flow respectively.
Finally, we define a parameter $\mathcal{M}_r^{\ell_2}$ shown in equation (28), similarly to that of $\mathcal{M}_r$ except that the numerator and the denominator are evaluated considering the ℓ_2 norm equivalent of a particular loss function representing the mean velocity error.

$$\mathcal{M}_r^{\ell_2} = \sqrt{\frac{\mathcal{M}^{\ell_2}}{\mathcal{M}_0^{\ell_2}}} \tag{28}$$

The parameter $\mathcal{M}_r^{\ell_2}$ will be compared with $\mathcal{G}_r^{\ell_2}$ to check for the quality of the assimilated mean flow.

6. Applications

We present two different numerical experiment flow cases to illustrate the performance and features of our data assimilation framework. We first consider the reconstruction of mean flow around a circular cylinder at $Re = 100$, which is followed by the assimilation of mean flow over a backward-facing step at $Re = 500$. All computations are performed on the FreeFem++ package (Hecht, 2012 [45]), based on the finite-element method for spatial discretization. Taylor-Hood finite elements are used, more specifically quadratic $\left(\mathcal{P}^2\right)$ elements for the velocity and linear $\left(\mathcal{P}^1\right)$ elements for the pressure. The sparse mean flow measurements, as inputs for the data assimilation, are obtained from these two-dimensional direct numerical simulations on FreeFem++.
The direct and adjoint equations are solved in weak form, obtained by multiplying the Navier-Stokes equations by a test function

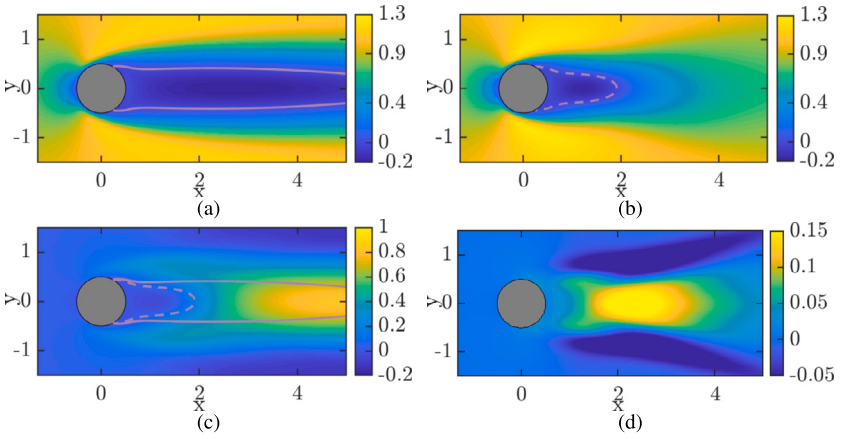

Fig. 3. Flow around a circular cylinder at $Re = 100$: Stream-wise component of (a) the base flow, (b) the time-averaged mean flow, (c) the deviation between the base flow and the time-averaged mean flow, and (d) the solenoidal part of the forcing from DNS. The solid and the dashed lines denote the re-circulation region for the base flow and the true mean flow, respectively.

and employing integration by parts. The direct equations are solved using a Newton-Raphson method which requires the linearized RANS equations. The adjoint equations in the discrete form are solved in a computationally efficient way by using the linearized RANS equations from the direct method. The transpose of the Jacobian matrix (also required for the direct solution) is used to obtain the discrete solution of the adjoint equations (8a) and (8b). The consistency of the discrete adjoint within a finite-element formalism has been well established by Hartmann, 2006 [46]. The optimization problem is solved within FreeFem++ using the NLopt nonlinear optimization package; for details on NLopt, the reader can refer to Johnson [47]. Throughout this study, the outliers are only introduced along the horizontal component of the mean velocity measurements, unless specified otherwise.

6.1. Flow around a circular cylinder

We consider two-dimensional flow around a circular cylinder at $Re = 100$. A view of the mesh surrounding the cylinder surface is shown in Fig. 2. The RANS equations (1) are non-dimensionalized by the diameter of the cylinder D and the inlet free-stream velocity u_∞. The Reynolds number is hence defined as $Re = {}^{u_\infty D}/_{\nu}$, with ν as the kinematic viscosity of the fluid. The computations are carried out on a square mesh with coordinates $x \in [-30, 30] \times y \in [-30, 30]$, producing 10^4 degrees of freedom. The initial guess of $\mathbf{f} = \vec{0}$ satisfies the divergence-free condition, as well as the zero Reynolds stress conditions on the solid walls of the cylinder.
The base flow for the circular cylinder at $Re = 100$ is calculated using the initial guess of the control vector, and the stream-wise component of the base flow is represented in Fig. 3a respectively. The mean flow field is obtained by time averaging the instantaneous fields from the direct numerical simulation of flow around a circular cylinder at $Re = 100$. The stream-wise component of the mean flow is displayed in Fig. 3b. Finally, the initial difference of the stream-wise component between the base flow and the mean flow is plotted in Fig. 3c. The difference in the length of the re-circulation region as predicted from the base and the mean flows around the circular cylinder, is significant. The length of the re-circulation region $\mathbf{l}_r$ denotes the extent downstream of the cylinder where the mean velocity vector is in the opposite direction of the incident mean flow, and hence $\mathbf{l}_r$ is calculated at the transition point $(\overline{u} = 0)$. This shows that the inverse problem is characterized by a significant amount of control on the base flow to predict the mean flow accurately. We present the solenoidal part of the true forcing in Fig. 3d as the formulation of the data assimilation problem in equations (3) facilitates the recovery of only this part of the forcing, see Foures et al., 2014 [10] for further details.
The boundary conditions for the direct system are given as, Inlet: $\overline{u} = 1, \overline{v} = 0$; Cylinder surface: $\overline{u} = 0, \overline{v} = 0$; Top & bottom symmetry: ${}^{\partial\overline{u}}/_{\partial y} = 0, \overline{v} = 0$; Outlet: $Re^{-1}\left({}^{\partial\overline{u}}/_{\partial x}\right) - \overline{p} = 0$, ${}^{\partial\overline{v}}/_{\partial x} = 0$, and the boundary conditions for the adjoint (obtained via integration by parts) system are given as, Inlet: $\overline{u}^\dagger = 0, \overline{v}^\dagger = 0$; Cylinder surface: $\overline{u}^\dagger = 0, \overline{v}^\dagger = 0$; Top & bottom symmetry: ${}^{\partial\overline{u}^\dagger}/_{\partial y} = 0, \overline{v}^\dagger = 0$; Outlet: $Re^{-1}\left({}^{\partial\overline{u}^\dagger}/_{\partial x}\right) + \overline{p}^\dagger = -\overline{u}\,\overline{u}^\dagger$, $Re^{-1}\left({}^{\partial\overline{v}^\dagger}/_{\partial x}\right) = -\overline{u}\,\overline{v}^\dagger$ respectively. The spatial domain for evaluating the global error of the assimilated mean flow around the circular cylinder is considered as a rectangular section $\Omega^{\mathcal{E}} : x \in [-1, 5] \times y \in [-3, 3]$ in equation (27). The sensor locations are considered in the most sensitive flow regions: close to the point of separation, and inside the re-circulation region of the mean flow around the circular cylinder. After a parametric study, we consider a set of 14 sparse measurements (unless stated otherwise), as shown in Fig. 2, and introduce outliers artificially at some of the sensitive regions of the flow, as discussed earlier in section 5. The number and locations of the sensors can be freely chosen; nonetheless, a detailed mathematical analysis has been presented by Mons and Marquet, 2021 [48] to analyze the effect of different sensor placement strategies on the data assimilation technique.

6.1.1. Data assimilation without noise

First, we perform data assimilation in the absence of outliers. A ℓ_2 loss function is used to represent the deviation error between the model-predicted and true measured mean velocities. The stream-wise component of the assimilated mean flow is reconstructed

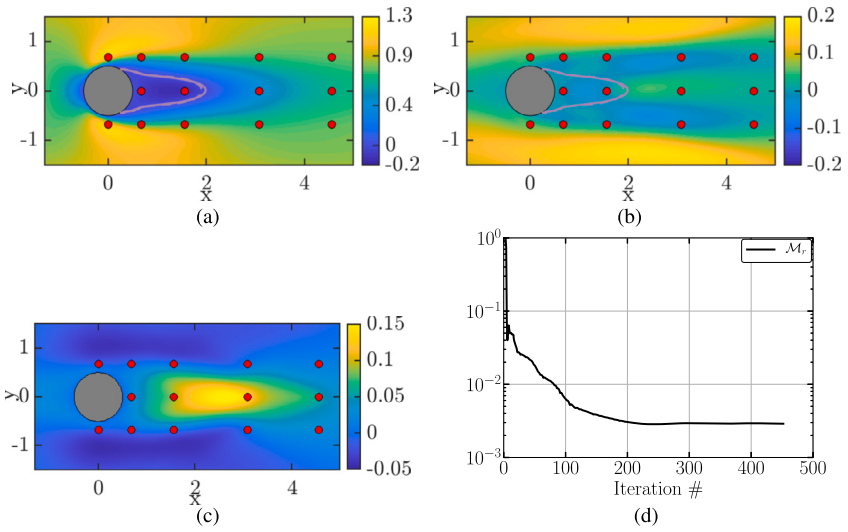

Fig. 4. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using a ℓ_2 loss function to represent the error in measurements, solved with L-BFGS: Stream-wise component of (a) the assimilated mean flow, (b) the deviation between the assimilated and the true mean flows, and (c) the assimilated forcing. (d) denotes the objective function as mentioned in Table 1. The solid and the dashed lines denote the re-circulation region for the assimilated and the true mean flows.

Table 3
Performance of ℓ_2 and hybrid loss functions, for data assimilation of mean flow around a circular cylinder at $Re = 100$.

Loss function	Δ	$\delta(\%)$	$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	l_r
DNS	—	—	—	—	1.9125
ℓ_2: L-BFGS	—	—	0.3242	0.0538	1.9891
Hybrid: L-BFGS	—	—	0.2932	0.0882	2.0329

with good accuracy using the L-BFGS method, see Fig. 4a. The discrepancy between the assimilated and true mean velocities is rather low in the region around the circular cylinder where measurements are available, as shown in Fig. 4b. The ℓ_2 deviation error approaches order 10^{-3}, as shown in Fig. 4d, and the length of the re-circulation region is within a 4% uncertainty limit of the true length; see Table 3 for details. The normalized global error $\mathcal{G}_r^{\ell_2}$ (refer to equation (27)) between the true and assimilated mean velocities over the control volume $\Omega^{\mathcal{E}}$ is sufficiently small. The stream-wise component of the assimilated forcing becomes significant downstream of the cylinder near the re-circulation region, see Fig. 4c. Even though the assimilated forcing does not completely represent the solenoidal part of the true forcing, however, a significant portion of it is recovered (see Fig. 3d for comparison).

6.1.2. *Performance of ℓ_2 and ℓ_1 loss function*

We next consider the same measurement locations, but with outliers at sensor locations 1 and 7, amounting to a percent uncertainty of $[100\%, 100\%]$ as estimated from equation (26). First, the inverse problem is solved using the multi-step ADMM algorithm, and the results are presented in Fig. 5. One of the major inferences is the loss of symmetry of the re-circulation region due to the presence of an outlier inside it, as can be seen from Fig. 5a. The deviation between the assimilated and the true mean flows in Fig. 5b clearly shows that the assimilated mean velocity field in the lower half of the control volume still preserves the symmetry to a certain extent while the distortion is mainly in the upper half of the control volume. The loss of symmetry is further aggravated due to the presence of the outlier at location 1 which shifts the point of mean flow separation further downstream of the cylinder surface. However, the mean flow separation point close to sensor location 2 remains less affected and matches more closely with the true one, see Fig. 5b. The magnitude of the primal and dual conditions for the ADMM algorithm attain constant values, as seen from Fig. 5c, which shows the convergence of the algorithm as the objective function also converges to a fixed value (of order 10^{-5}), see Fig. 5d. As discussed earlier, the ℓ_2 loss function gives equal weight to all points in the measurement data set, and hence the presence of an outlier at any sensor location is not restricted locally.

The same data assimilation, in Fig. 5, is performed using the quasi-Newton method L-BFGS algorithm; the assimilated mean flow is very similar to the results obtained with ADMM, as shown in Fig. 6. The objective function value settles down to a value of order 10^{-5} for both cases of ADMM and L-BFGS algorithm, which demonstrates the independence of the numerical algorithm employed for solving the inverse problem. In addition, the assimilated forcing is highly distorted (see Fig. 6c). These results reveal that the failure of the ℓ_2 loss function to recover the mean flow with sufficient accuracy in the presence of outliers, is an artifact of the loss function rather than the optimization algorithm. The length of the re-circulation region remains the same, up to the 1^{st} decimal place with both ADMM and L-BFGS formulations, which is around 14% (ADMM) and 11% (L-BFGS) larger than the true length; see Table 4.

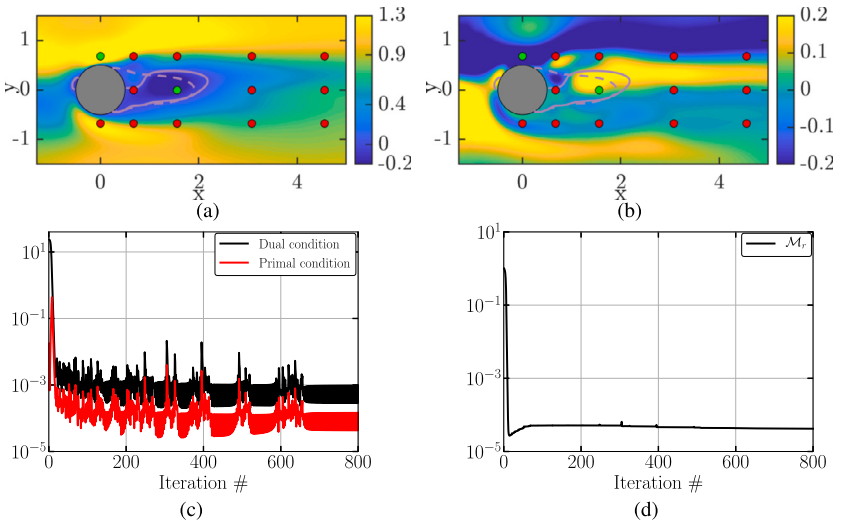

Fig. 5. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using a ℓ_2 loss function to represent the error in measurements, solved with ADMM: Stream-wise component of (a) the assimilated mean flow and (b) the error between the assimilated and the true mean flows. (c) denotes the magnitude of the primal and dual conditions of the optimization as stated in equations (C.14) and (C.15), and (d) denotes the objective function as mentioned in Table 1. The measurements with outliers along the horizontal velocity component $(\overline{u})$ at sensor locations $[1, 7]$ amounting to percent uncertainty $[100, 100]$, are denoted by • respectively. For details on the re-circulation region, see the caption of Fig. 4.

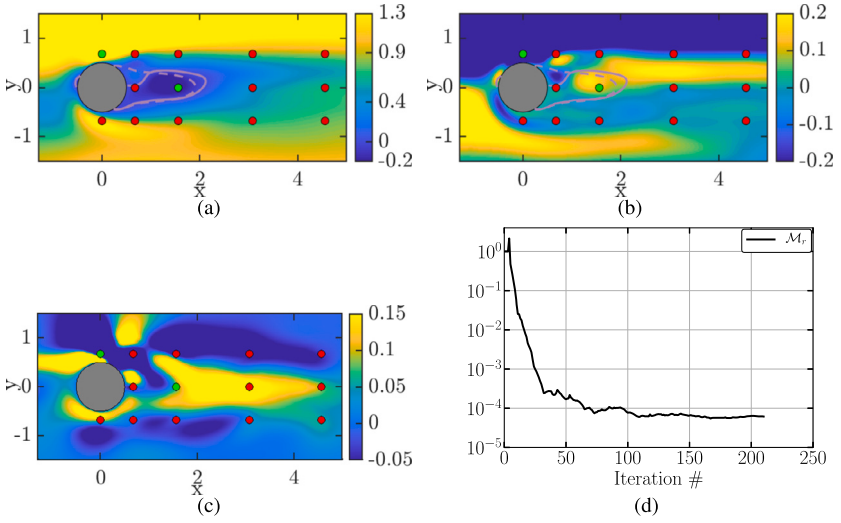

Fig. 6. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using a ℓ_2 loss function to represent the error in measurements, solved with L-BFGS: Stream-wise component of (a) the assimilated mean flow, (b) the error between the assimilated and the true mean flows, and (c) the assimilated forcing. (d) denotes the objective function as mentioned in Table 1. For details on the terminology of sensor locations, and the re-circulation region, see the caption of Fig. 5, and the percent uncertainty, see Table 5.

To compare the performance of other loss functions, data assimilation is performed for the same case as mentioned above but using a ℓ_1 loss function for the deviation error between the model-predicted and measured mean velocities. The assimilated mean flow obtained using the ADMM algorithm exhibits a similar level of accuracy to that of the ℓ_2 loss function, as shown in Fig. 7. A striking observation for the ℓ_1 loss function is that it furnishes different optimal solutions when initialized with different values of the penalty parameter (μ) in the ADMM algorithm. This behavior may point to the fact that the optimization algorithm is trapped in different local minima. Two different cases are presented in Table 4: with the same configuration of the sensors, percent uncertainty level, and convergence conditions, but initialized with different values of μ, and they give rise to two different mean flows. The relative uncertainty for predicting the length of the re-circulation region changes from around 16% to 21% for the two different minima of the ℓ_1 loss function, see Table 4 for the values corresponding to case 1 and case 2 of the ℓ_1 loss function.

6.1.3. *Performance of IRLS*

We present an IRLS form of the loss function, which facilitates the use of a mixed ℓ_1/ℓ_2 function. The IRLS method solves a modified form of a ℓ_1 function as a weighted least-squares in an iterative framework. The IRLS parameter $(\eta_i \to 0)$ avoids a

Table 4
Performance of different loss functions in the presence of outliers along the horizontal velocity component $(\overline{u})$ at different sensor locations, for the data assimilation of mean flow around a circular cylinder at $Re = 100$. The percent uncertainty in measurements at each sensor location with an outlier is calculated using equation (26) and denoted as $\widetilde{\boldsymbol{\delta}} = [100, 100, 100, 100, 100, 100, 100]$, and the absolute uncertainty is given as $\widetilde{\boldsymbol{\Delta}} = [0.8057, 0.8067, 0.4405, 0.0259, 0.1422, 0.4179, 0.6637]$ respectively. For the case of $*[1,7]\,(\overline{u})$, the outliers are introduced at the sensor locations $[1,7]$ along with Gaussian errors at other sensor locations.

Outlier locations	Loss function	$\boldsymbol{\Delta}$	$\boldsymbol{\delta}(\%)$	$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	l_r
—	DNS	—	—	—	—	1.9125
$[1,7]\,(\overline{u})$	ℓ_2: ADMM	$[0.8057, 0.1422]$	$[100, 100]$	0.5655	4.1932e-05	2.1818
$[1,7]\,(\overline{u})$	ℓ_2: L-BFGS	$[0.8057, 0.1422]$	$[100, 100]$	1.2045	6.1347e-05	2.1220
$[1,7]\,(\overline{u})$	ℓ_1: ADMM (Case 1)	$[0.8057, 0.1422]$	$[100, 100]$	1.3513	0.0047	2.2292
$[1,7]\,(\overline{u})$	ℓ_1: ADMM (Case 2)	$[0.8057, 0.1422]$	$[100, 100]$	1.6927	0.0021	2.3168
$[1,7]\,(\overline{u})$	IRLS $(\eta_i = 10^{-2})$	$[0.8057, 0.1422]$	$[100, 100]$	1.1676	0.0725	2.1652
$[1,7]\,(\overline{u})$	IRLS $(\eta_i = 10^{-6})$	$[0.8057, 0.1422]$	$[100, 100]$	0.1572	0.3789	2.1394
$[1,7]\,(\overline{u})$	Huber: ADMM	$[0.8057, 0.1422]$	$[100, 100]$	0.2063	0.4576	2.0716
$[1,7]\,(\overline{u})$	Huber: L-BFGS	$[0.8057, 0.1422]$	$[100, 100]$	0.2102	0.4654	2.0360
$[1,7]\,(\overline{u})$	Hybrid: L-BFGS	$[0.8057, 0.1422]$	$[100, 100]$	0.2101	0.4657	2.0360
$*[1,7]\,(\overline{u})$	Hybrid: L-BFGS	$[0.8057, 0.1422]$	$[100, 100]$	0.2062	0.4767	1.9710
$[1,2,3,4,7,10,13]\,(\overline{u})$	Huber: L-BFGS	$\widetilde{\boldsymbol{\Delta}}$	$\widetilde{\boldsymbol{\delta}}$	0.1735	0.5548	2.0237
$[1,2,3,4,7,10,13]\,(\overline{u})$	Hybrid: L-BFGS	$\widetilde{\boldsymbol{\Delta}}$	$\widetilde{\boldsymbol{\delta}}$	0.1777	0.5557	2.0238

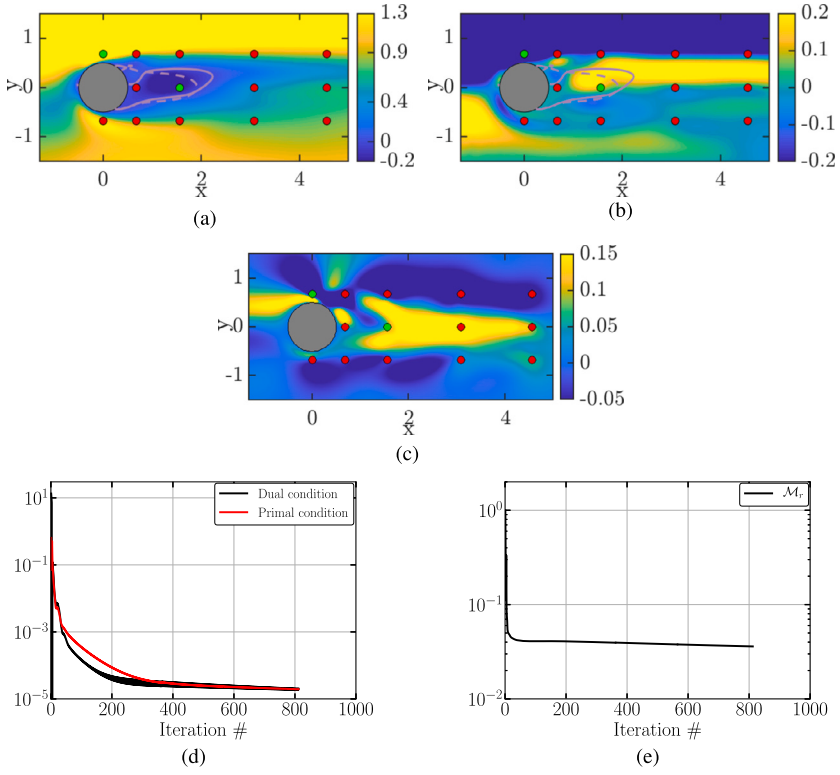

Fig. 7. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using a ℓ_1 loss function to represent the error in measurements, solved with ADMM: Stream-wise component of (a) the assimilated mean flow, (b) the error between the assimilated and the true mean flows, and (c) the assimilated forcing. (d) denotes the magnitude of the primal and the dual conditions of the optimization, and (e) denotes the objective function as mentioned in Table 1. For details on the terminology of sensor locations, and the re-circulation region, see the caption of Fig. 5, and the percent uncertainty, see Table 5.

singularity at the points of vanishing deviation error. Fig. 8a illustrates the effect of different limiting values of (η_i) on the accuracy of the assimilated mean flow, solved with an L-BFGS solver for the same case of data assimilation with outliers at locations 1 and 7 respectively. For larger values of η_i, the assimilated mean flow field shown in Fig. 8a suffers from the same issues as observed for a ℓ_2 or ℓ_1 loss function. Lowering the value of η_i not only improves the quality of the assimilated mean flow but also helps to restore the symmetry of the re-circulation region when η_i changes from order 10^{-2} to 10^{-4} (not shown here). Finally, with $\eta_i = 10^{-6}$ which accounts for the addition of a constant value of 10^{-12} to the entries of the weight matrix, as denoted in equation (19), the assimilated mean flow closely matches the true mean flow, as shown in Fig. 8b. The symmetry of the re-circulation region is restored for the

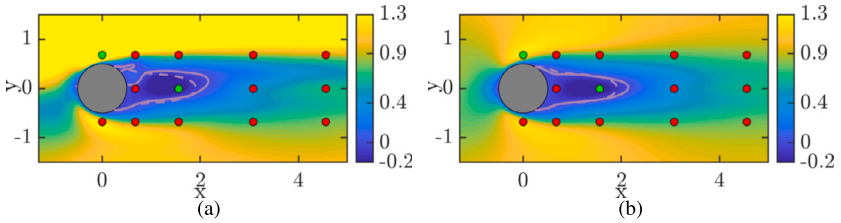

Fig. 8. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using an IRLS form of the ℓ_1 loss function to represent the error in measurements, solved with L-BFGS: Stream-wise component of the assimilated mean flow for (a) $\eta_i = 10^{-2}$, and (b) $\eta_i = 10^{-6}$. For details on the terminology of sensor locations, and the re-circulation region, see the caption of Fig. 5, and the percent uncertainty, see Table 5.

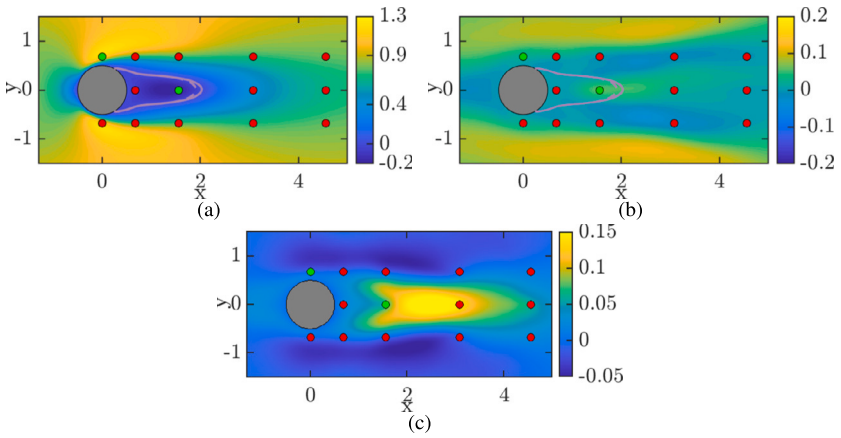

Fig. 9. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using the Huber loss function to represent the error in measurements, solved with L-BFGS: Stream-wise component of (a) the assimilated mean flow, (b) the error between the assimilated and the true mean flows, and (c) the assimilated forcing. For details on the terminology of sensor locations, and the re-circulation region, see the caption of Fig. 5, and the percent uncertainty, see Table 5.

case of $\eta_i = 10^{-6}$, except for a slight distortion at the upper half of the control volume $\Omega_{\mathcal{E}}$ which is evident from the mean flow in Fig. 8b. However, the improvement in the relative uncertainty in the re-circulation length changes slightly from around 13% to 12%. For even lower values of η_i, the assimilated mean flow did not change significantly and thus has not been presented here.

6.1.4. *Performance of Huber and hybrid loss functions*

With the introduction of the Huber loss function for the same case of data assimilation with outliers at sensor locations 1 and 7, the symmetry of the re-circulation region is completely recovered using the L-BFGS method, see Fig. 9a. We observe that the effect of the outliers at the point of separation and within the re-circulation region is restricted locally – without influencing the accuracy of the global mean flow around the circular cylinder. The length of the re-circulation region with a Huber loss function is within a relative uncertainty of 6%, which is half of the relative uncertainty for the best case of the IRLS function $\left(\eta_i = 10^{-6}\right)$. The accuracy of the reconstructed mean flow based on a Huber loss function can be attributed to the similarity of the deviation field, shown in Fig. 9b, to the case of data assimilation without outliers using a ℓ_2 loss function, in Fig. 4b. Similar behavior is seen in the stream-wise component of the assimilated forcing, see Figs. 9c and 4c. We solved the above case of the Huber loss function with an ADMM formulation, to check for the consistency of the reconstructed mean flow, see Table 4. We thus establish that the Huber loss function succeeds in reconstructing the mean flow, with good accuracy and without affecting the quality of the mean flow restoration around the locations of the outliers. We also recover the mean flow using the hybrid loss function, to a similar level of accuracy as that of the Huber loss function and therefore has not been presented here. In addition, we consider a case of data assimilation with outliers at sensor locations 1 and 7 and Gaussian noise at the other 12 measurement locations. At a particular measurement location, the Gaussian noise magnitude is selected randomly from a normal distribution with a mean of zero and a standard deviation $\left(\sigma_i = 0.1\right)$ respectively. With a hybrid loss representation of the deviation error between the measured and the assimilated mean velocities, the mean flow is reconstructed accurately and the length of the re-circulation region is recovered with sufficient accuracy (see Table 4) as compared to the case of data assimilation with no Gaussian noise. Comparing the two cases of data assimilation with outliers at sensor locations 1 and 7 in the presence and absence of Gaussian noise at other measurement locations, we see that the hybrid loss function can recover the mean flow with a similar level of accuracy. Furthermore, we see the relative uncertainty in recovering the length of the re-circulation region has reduced by 3% as compared to the case without Gaussian noise. However, we cannot guarantee a reduction in uncertainty in the recovery of the re-circulation region in the cases of data assimilation for different Gaussian noise intensities generated using a different seed, as compared to the cases of data assimilation without Gaussian noise.

As a test of robustness, a configuration comprising a measurement data set with outliers at seven locations is considered for the reconstruction of mean flow around a circular cylinder at $Re = 100$. The quantitative values of the local and global error associated

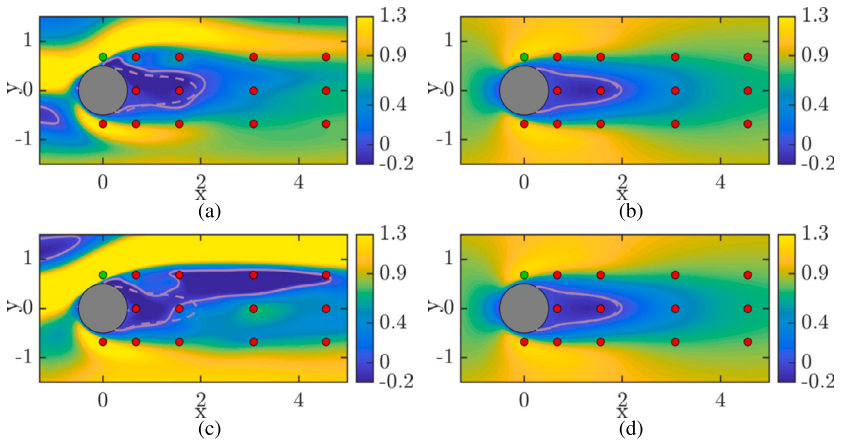

Fig. 10. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using the (a,c) ℓ_2 and (b,d) hybrid loss functions to represent the error in measurements, solved with L-BFGS: Stream-wise component of the assimilated mean flow. The measurement with outliers along (a)-(b) the vertical component $(\overline{v})$, and (c)-(d) the horizontal and vertical components $(\overline{u}, \overline{v})$ of the mean velocity at sensor location [1], is denoted by • respectively. For details on the re-circulation region, see the caption of Fig. 5 and the percent uncertainty, see Table 5.

Table 5
Performance of the ℓ_2 and hybrid loss functions in the presence of outliers along different components of the mean velocity at different sensor locations, for the data assimilation of mean flow around a circular cylinder at $Re = 100$.

Outlier locations	Δ	$\delta(\%)$	ℓ_2 loss function			Hybrid loss function		
			$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	I_r	$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	I_r
[1] $(\overline{u})$	[0.8178]	[101.5]	1.1247	0.0078	2.0228	0.2289	0.4696	1.9829
[1] $(\overline{v})$	[0.8185]	[563]	1.7030	0.0674	2.0831	0.2007	0.4807	1.9802
[1] $(\overline{u}, \overline{v})$	[0.8178, 0.8185]	[101.5, 563]	2.1402	0.9006	4.6627	0.1892	0.5988	1.9889
[2] $(\overline{u})$	[0.8178]	[101.5]	1.1253	0.0075	1.9823	0.2290	0.4722	1.9833
[2] $(\overline{v})$	[0.8184]	[562]	1.7454	0.0649	2.0396	0.2465	0.4814	1.9790
[2] $(\overline{u}, \overline{v})$	[0.8178, 0.8184]	[101.5, 562]	1.9012	0.9099	2.7653	0.2160	0.6114	1.9806
[4] $(\overline{u})$	[0.8185]	[3160]	0.6044	0.0046	1.9105	0.2864	0.4831	2.0313
[4] $(\overline{v})$	[0.8180]	[18400]	0.4063	0.0005	2.1379	0.2093	0.4772	1.9793
[4] $(\overline{u}, \overline{v})$	[0.8185, 0.8180]	[3160, 18400]	0.7406	0.0038	1.9189	0.2253	0.6175	2.0790
[US] $(\overline{u})$	[0.8209]	[461.5]	1.2535	0.0064	2.0656	0.3041	0.4769	2.0137
[US] $(\overline{v})$	[0.8206]	[117500]	0.5556	0.0006	2.0122	0.2084	0.4769	1.9800
[US] $(\overline{u}, \overline{v})$	[0.8209, 0.8206]	[461.5, 117500]	0.9626	0.0086	2.0078	0.2907	0.5963	2.0584

with the assimilated mean flow are presented in Table 4. We did not see a significant difference in the quality of the assimilated mean flow while using the Huber or hybrid loss function to represent the deviation error.

Furthermore, we consider a configuration comprising a measurement data set with an outlier at the point of detachment located on the upper side of the circular cylinder. Initially, the outlier is introduced along the vertical component of the mean velocity for which the stream-wise component of the assimilated mean flow is well reconstructed with a hybrid loss function (see Fig. 10b) as compared to the one with a ℓ_2 loss function (see Fig. 10a). Moreover, we considered another case where the outliers are introduced along the horizontal as well as the vertical components of the mean velocity at the point of detachment located on the upper side of the circular cylinder, and again, the mean flow is well recovered with a hybrid loss function (see Fig. 10d) as compared to the one with a ℓ_2 loss function (see Fig. 10c). We considered outliers along the horizontal component, vertical component, and the horizontal and vertical components of the mean velocity, for each of the sensors at the point of detachment located on the upper and lower sides of the circular cylinder as well as inside the re-circulation region, for which the local and global errors of reconstruction are presented in Table 5. We can see that for every case presented in Table 5, the global error reduces significantly when considering the hybrid loss function over the ℓ_2 loss function.

Lastly, we consider a case of data assimilation where the outliers are introduced along the horizontal component, vertical component, and the horizontal and vertical components of the mean velocity, at the stagnation point upstream (denoted by US) of the circular cylinder. When the outlier is entirely along the vertical mean velocity component at the sensor location US, the reconstructed mean flow loses its symmetry with the ℓ_2 loss function (see Fig. 11a) while the hybrid loss function remains robust to the outliers (see Fig. 11b). Similarly, we can see that the hybrid loss function (see Fig. 11d) preserves the symmetry of the mean flow around the circular cylinder as compared to the ℓ_2 loss function (see Fig. 11c) when the outliers are introduced along the horizontal and vertical components of the mean velocity at the sensor location US. From Table 5, we can confirm that the hybrid loss function always supersedes ℓ_2 loss function in reconstructing the mean flow with a desirable accuracy.

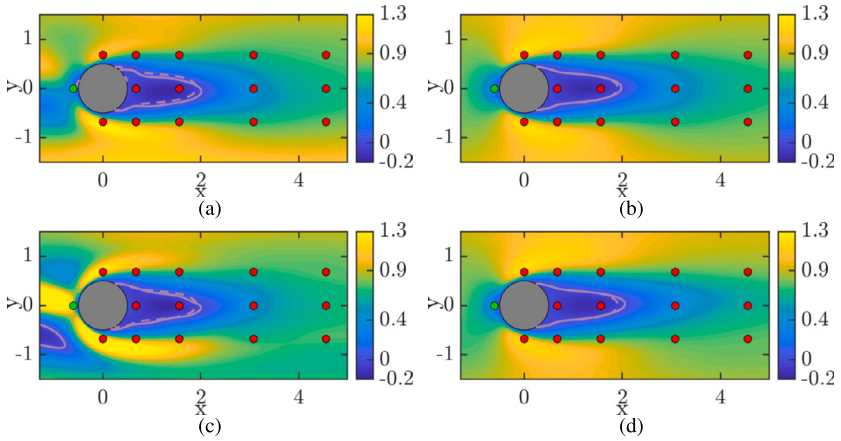

Fig. 11. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using the (a,c) ℓ_2 and (b,d) hybrid loss functions to represent the error in measurements, solved with L-BFGS: Stream-wise component of the assimilated mean flow. The measurement with outliers along (a)-(b) the vertical component $(\overline{v})$, and (c)-(d) the horizontal and vertical components $(\overline{u}, \overline{v})$ of the mean velocity at sensor location [US], is denoted by • respectively. For details on the re-circulation region, see the caption of Fig. 5 and the percent uncertainty, see Table 5.

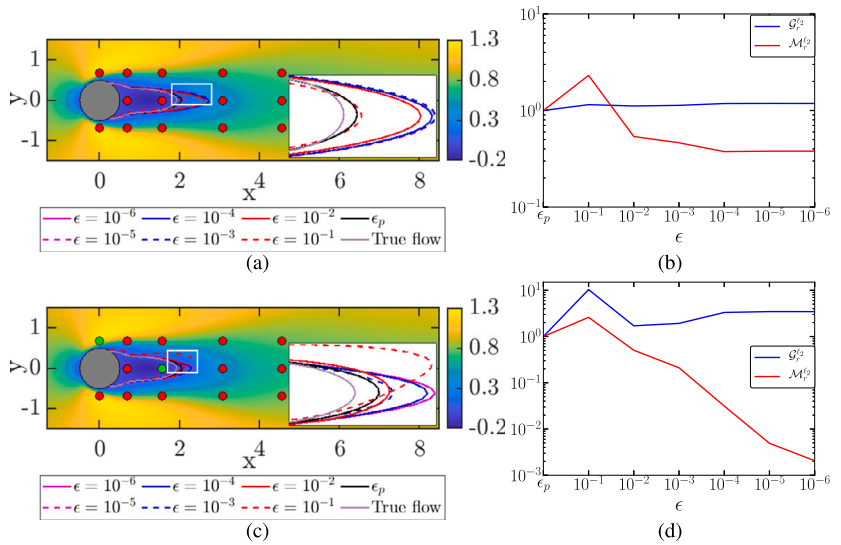

Fig. 12. Sensitivity of the assimilated mean flow around a circular cylinder at $Re = 100$ with respect to the cut-off parameter of the hybrid loss function, for a set of measurements without (top) and with (bottom) outliers. (a,c) shows the stream-wise component of the true mean flow. The inset in (a,c) shows the variation of the re-circulation region with the variation of the cut-off parameter. The black line indicates the cut-off parameter (ϵ_p) estimated using our proposed numerical framework in Appendix D. (b,d) shows the variation of local measurement and global assimilation errors for each value of the cut-off parameter (shown along the horizontal axis), obtained by normalizing the values of $\mathcal{M}_r^{\ell_2}$ and $\mathcal{G}_r^{\ell_2}$ for each ϵ with respect to the values of $\mathcal{M}_r^{\ell_2}$ and $\mathcal{G}_r^{\ell_2}$ for ϵ_p. For the terminology of sensor locations, see the caption of Fig. 5, and for the percent uncertainty, see Table 4.

6.1.5. *Sensitivity of epsilon/cut-off parameter*

The epsilon/cut-off parameter in the Huber and hybrid loss functions is critical for the minimization problem to reach optimal conditions. This section presents the effect of the cut-off parameter on the quality of the assimilated mean flow. Firstly, we consider measurements without any outliers and perform a parametric study to assess the effect of cut-off parameter variations on the assimilated mean flow. We consider a range of values of ϵ ranging from 10^{-6} to 10^{-1}. From Fig. 12a, the length of the re-circulation region is far longer than the true length, for rather low values of ϵ. As ϵ increases to the order of 10^{-1}, the uncertainty in the prediction of the re-circulation region reduces. We denote ϵ_p as the value of the cut-off parameter determined by the mathematical framework discussed in Appendix D. In this case, the re-circulation region with ϵ_p matches more closely the true re-circulation region, among all values of the cut-off parameter considered. The results are displayed in Fig. 12b, where the values of the global and local error for different values of ϵ are normalized with respect to the values of $\mathcal{G}_r^{\ell_2}$ and $\mathcal{M}_r^{\ell_2}$ for ϵ_p. The local measurement error $\mathcal{M}_r^{\ell_2}$ increases for $\epsilon = 10^{-1}$ and then diminishes quickly with a decreasing ϵ while the value of the global assimilated error $\mathcal{G}_r^{\ell_2}$ initially increases and then reduces, but ultimately does not fall below the global error for ϵ_p. This kind of objective function, where the local measurement error reduces rapidly without a decrease in the corresponding global error, is referred to as a bad objective

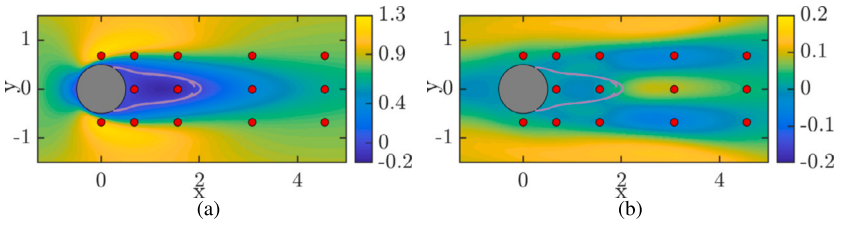

Fig. 13. Data assimilation of mean flow around a circular cylinder at $Re = 100$ using a hybrid loss function to represent the error in measurements, solved with L-BFGS: Stream-wise components of (a) the assimilated mean flow, and (b) the error between the assimilated and the true mean flows. For details on the re-circulation region, see the caption of Fig. 5.

Table 6
Performance of the ℓ_2 and hybrid loss functions in the presence of outliers along the horizontal component of the mean velocity $(\overline{u})$ at different sensor locations, for the data assimilation of mean flow around a circular cylinder at $Re = 100$.

Outlier locations	Δ	$\delta(\%)$	ℓ_2 loss function			Hybrid loss function		
			$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	l_r	$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	l_r
[3] $(\overline{u})$	[0.8194]	[186]	0.7326	0.0061	1.7391	0.2294	0.4509	1.9927
[5] $(\overline{u})$	[0.8179]	[185.4]	2.3404	0.0190	2.0441	0.2149	0.4487	1.9910
[6] $(\overline{u})$	[0.8182]	[144.8]	2.0962	0.0014	2.5185	0.2782	0.4075	2.0407
[7] $(\overline{u})$	[0.8176]	[575]	0.6019	0.0011	2.2544	0.2710	0.4865	2.0442
[8] $(\overline{u})$	[0.8179]	[140.3]	2.8883	0.0086	2.4445	0.2725	0.4098	2.0337
[9] $(\overline{u})$	[0.8179]	[127.1]	1.4927	0.0008	2.0865	0.3207	0.3500	2.0354
[10] $(\overline{u})$	[0.8192]	[196]	1.9446	0.0005	1.9661	0.2569	0.4349	2.0010
[11] $(\overline{u})$	[0.8185]	[129.6]	1.4559	0.0007	2.0676	0.3113	0.3472	2.0294
[12] $(\overline{u})$	[0.8182]	[109.9]	2.2647	0.0004	1.9179	0.2967	0.3358	1.9929
[13] $(\overline{u})$	[0.8164]	[123]	2.3041	0.0005	2.0126	0.2949	0.3794	1.9837
[14] $(\overline{u})$	[0.8182]	[110.15]	1.9169	0.0007	1.9290	0.3097	0.3344	1.9911
[1,2] $(\overline{u})$	[0.5785, 0.5792]	[71.8, 71.8]	1.6444	0.0084	1.9052	0.2464	0.4699	1.9936
[1,3] $(\overline{u})$	[0.7171, 0.3921]	[89, 89]	1.2162	0.0033	1.8289	0.1896	0.4300	1.9683
[1,4] $(\overline{u})$	[0.8178, 0.0263]	[101.5, 101.5]	1.1218	0.0074	2.0119	0.2357	0.4708	1.9801
[1,5] $(\overline{u})$	[0.7171, 0.3929]	[89, 89]	1.1441	0.0065	1.8207	0.1985	0.4562	1.9732
[2,3] $(\overline{u})$	[0.7179, 0.3921]	[89, 89]	1.1278	0.0060	1.7927	0.2079	0.4582	1.9756
[2,4] $(\overline{u})$	[0.8188, 0.0263]	[101.5, 101.5]	1.0449	0.0073	1.9773	0.2197	0.4728	1.9853
[2,5] $(\overline{u})$	[0.7179, 0.3929]	[89, 89]	1.2190	0.0033	1.8050	0.1762	0.4292	1.9689
[3,4] $(\overline{u})$	[0.8173, 0.0480]	[185.5, 185.5]	0.6415	0.0057	1.7465	0.2416	0.4492	1.9850
[4,5] $(\overline{u})$	[0.0479, 0.8166]	[185, 185]	0.5252	0.0062	1.7698	0.2294	0.4471	1.9859
[4,10] $(\overline{u})$	[0.0506, 0.8163]	[195.3, 195.3]	2.0246	0.0006	1.9700	0.2624	0.4346	1.9996

function (see Symon et al., 2017 [11] for further details). We repeat the same analysis to study the effect of the cut-off parameter on the quality of the assimilated mean flow, but with measurements accompanied by outliers at locations 1 and 7 for a percent uncertainty level of $[100\%, 100\%]$. We observe a similar kind of behavior in the reconstruction of the re-circulation region, but now the re-circulation region breaks its symmetry for specific values of ϵ, see Fig. 12c. The re-circulation region with cut-off parameter ϵ_p not only is near the true value but also preserves symmetry, in contrast to all other values of ϵ considered in our study. We normalize the values of $\mathcal{M}_r^{\ell_2}$ and $\mathcal{G}_r^{\ell_2}$ across the entire range of ϵ considered, with respect to the values of $\mathcal{M}_r^{\ell_2}$ and $\mathcal{G}_r^{\ell_2}$ for ϵ_p. The global error is again the lowest for ϵ_p, see Fig. 12d. This analysis is in line with the principle that, with the overestimation of the cut-off parameter, the loss function performs a quadratic fit over a large deviation from the mean value. With the underestimation of the cut-off parameter, on the other hand, the loss function performs an absolute fit over a large deviation from the mean value. In both cases of over- and under-estimation of the cut-off parameter, the loss function performs either a ℓ_2 or ℓ_1 fit over a large deviation from the mean value, which distorts the symmetry of the re-circulation region as observed in the earlier cases of data assimilation. Furthermore, we perform a comparison of the ℓ_2 and hybrid loss functions for a measurement data set in the absence of outliers. The accuracy of the global assimilated mean flow is shown in Fig. 13a. The performance of the hybrid loss function to recover the global mean flow is markedly better than an equivalent ℓ_2 counterpart, as can be deduced from a comparison of the deviation fields in Fig. 13b, and Fig. 4b. The relative uncertainty in the length of the re-circulation region goes from 6% for a hybrid function to 4% for a ℓ_2 function, see Table 3.

Finally, we provide a comparison between the ℓ_2 and hybrid loss functions in the presence of one and two outliers at different sensor locations, in Table 6. For all cases, the quantitative values for $\mathcal{M}_r^{\ell_2}$, $\mathcal{G}_r^{\ell_2}$, and the length of the re-circulation region $\mathbf{l}_r$, are provided in Table 6 to support that hybrid loss function remains robust to outliers irrespective of their magnitude and location.

6.2. Flow over a backward-facing step

We consider the two-dimensional flow over a backward-facing step at $Re = 500$. A close view of the mesh is shown in Fig. 14. This is an example of an amplifier flow, a type of flow that is extremely sensitive to any kind of disturbance (noise or perturbation)

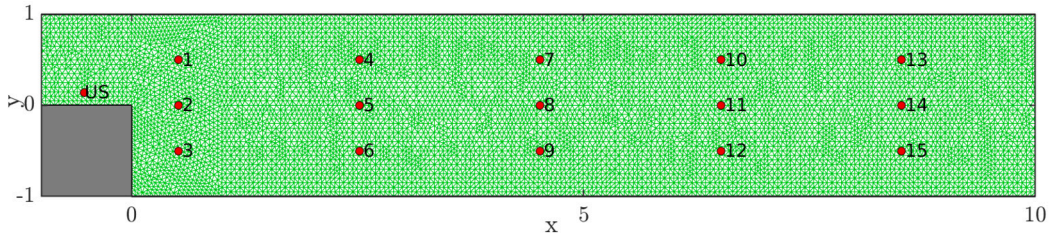

Fig. 14. Part view of the mesh for two-dimensional flow over a backward-facing step at $Re = 500$. The measurement locations are denoted by •.

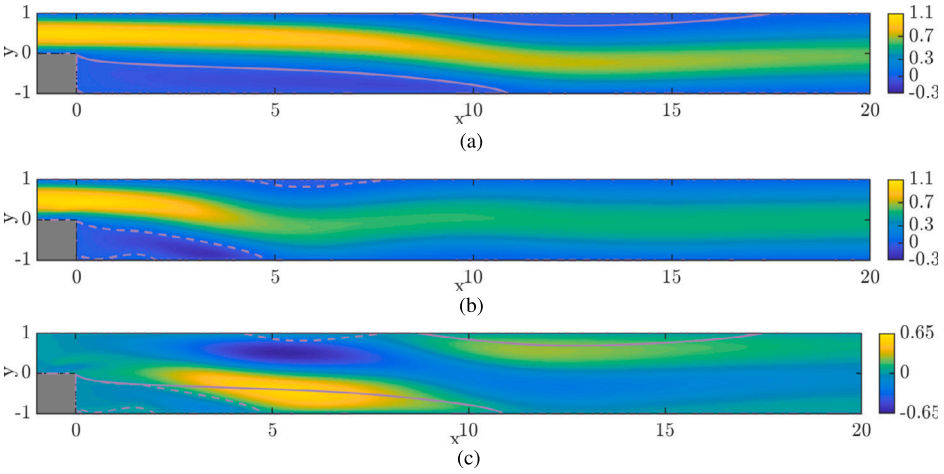

Fig. 15. Flow over a backward-facing step at $Re = 500$: Stream-wise component of (a) the base flow, (b) the time-averaged mean flow, and (c) the error for approximation of the mean flow by the base flow. The solid and the dashed lines denote the re-circulation region for the base flow and the true mean flow.

upstream of the step, as shown by Hervé et al., 2012 [49], and Sipp and Schmid, 2016 [50]. The RANS equations (1) are non-dimensionalized with the height of the step h and the inlet center-line velocity $u_{\max}$. The Reynolds number is defined as $Re = {}^{u_{\max}h}/_{\nu}$, with ν as the kinematic viscosity of the fluid. The computations are carried out on a rectangular mesh with coordinates $x \in [-5, 35] \times y \in [-1, 1]$ with 10^4 degrees of freedom. The base flow for the backward-facing step at $Re = 500$ is calculated using the initial guess of the control vector $\mathbf{f} = \vec{0}$, and the stream-wise component of the base flow is presented in Fig. 15a respectively. The mean flow field is obtained by time averaging the instantaneous fields from the direct numerical simulation of flow over a backward-facing step at $Re = 500$, and the stream-wise component of the mean flow is displayed in Fig. 15b. Finally, the initial deviation of the stream-wise component between the base flow and the mean flow is plotted in Fig. 15c. Similar to the circular cylinder case, the difference in the re-circulation region $(\overline{u} = 0)$, as predicted from the base flow and the mean flow over the backward-facing step, is rather high. This demonstrates that the inverse problem has to produce a significant amount of control on the base flow to predict the mean flow accurately.

The boundary conditions for the direct system are given as, Inlet: $\overline{u} = 4 \cdot y \cdot (1 - y), \overline{v} = 0$; Top and bottom walls: $\overline{u} = 0, \overline{v} = 0$; Outlet: $Re^{-1}\left({}^{\partial\overline{u}}/_{\partial x}\right) - \overline{p} = 0$, ${}^{\partial\overline{v}}/_{\partial x} = 0$, and the boundary conditions for the adjoint (obtained via integration by parts) system are given as, Inlet: $\overline{u}^{\dagger} = 0, \overline{v}^{\dagger} = 0$; Top and bottom walls: $\overline{u}^{\dagger} = 0, \overline{v}^{\dagger} = 0$; Outlet: $Re^{-1}\left({}^{\partial\overline{u}^{\dagger}}/_{\partial x}\right) + \overline{p}^{\dagger} = -\overline{u}\,\overline{u}^{\dagger}$, $Re^{-1}\left({}^{\partial\overline{v}^{\dagger}}/_{\partial x}\right) = -\overline{u}\,\overline{v}^{\dagger}$ respectively.

The spatial domain for evaluating the global error of the assimilated mean flow over a backward-facing step is considered as a rectangular section $\Omega^{\mathcal{E}} : x \in [-1, 15] \times y \in [-1, 1]$ in equation (27). We now consider a measurement set with 15 sensor locations in the presence of outliers at sensor locations 2, 6, and 8, amounting to a percent uncertainty of $[300\%, 300\%, 300\%]$ as estimated from equation (26). The inverse problem is solved using the quasi-Newton L-BFGS algorithm for a ℓ_2 representation of the deviation error. The results are presented in Fig. 16. One of the main observations is the over-predicted length of the re-circulation region on the top wall as well as the bottom wall in front of the step, as can be seen from Fig. 16a. The poor performance of the ℓ_2 loss function can be attributed to the large magnitudes of the deviation error between the assimilated and the true mean flows in Fig. 16b. As this is an amplifier flow, any noise at the sensor locations is amplified owing to the equal weights applied by the ℓ_2 loss function to all measurements, which results in poor accuracy of the reconstructed mean flow. The objective function value attains a constant value of order 10^{-3} after about 100 iterations (not shown here). The length of the assimilated re-circulation region exceeds the true length by about 31%, see Table 7.

With the introduction of a Huber representation of the deviation error for the same case, we see that the recovered mean flow is in good agreement with the true mean flow, see Fig. 17a. The relative uncertainty in the length of the re-circulation region on the bottom wall reduces to 6% with the Huber loss function as compared to the ℓ_2 loss function, however, the global reconstruction error remains very high at 60% for the Huber loss function, see Table 7 for more details. The deviation in the mean flow is again mainly

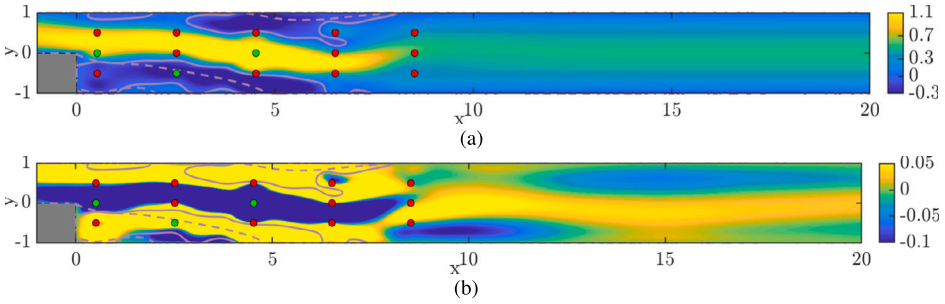

Fig. 16. Data assimilation of mean flow over a backward-facing step at $Re = 500$ using a ℓ_2 loss function to represent the error in measurements, solved with L-BFGS: Stream-wise component of (a) the assimilated mean flow and (b) the error between the assimilated and the true mean flows. The measurements with outliers (denoted by •) along the horizontal mean velocity component $(\overline{u})$ at sensor locations $[2, 6, 8]$ amount to percent uncertainty $[300, 300, 300]$. For details on the re-circulation region, see the caption of Fig. 15.

Table 7
Performance of ℓ_2, Huber, and hybrid loss functions in the presence of outliers along different components of the mean velocity at multiple sensor locations for data assimilation of mean flow over a backward-facing step at $Re = 500$. The percent uncertainty is given as $\widetilde{\delta} = [147.8, 147.8, 147.8, 147.8, 147.8, 147.8, 147.8]$ and the absolute uncertainty is given as $\widetilde{\mathbf{\Delta}} = [1.4717, 0.3491, 0.0667, 1.1528, 0.9333, 0.1289]$ respectively.

Outlier locations	$\mathbf{\Delta}$	$\delta(\%)$	DNS			ℓ_2 loss function		
			$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	l_r	$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	l_r
$[2, 6, 8]$ $(\overline{u})$	$[0.7086, 0.2617, 1.9848]$	$[300, 300, 300]$	—	—	4.8000	1.5105	0.0335	6.2800

Outlier locations	$\mathbf{\Delta}$	$\delta(\%)$	Huber loss function			Hybrid loss function		
			$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	l_r	$\mathcal{G}_r^{\ell_2}$	$\mathcal{M}_r^{\ell_2}$	l_r
$[2, 6, 8]$ $(\overline{u})$	$[0.7086, 0.2617, 1.9848]$	$[300, 300, 300]$	0.5918	0.8696	4.5200	0.1009	0.7882	4.9200
$[1, 2, 3, 4, 5, 6]$ $(\overline{u})$	$\widetilde{\mathbf{\Delta}}$	$\widetilde{\delta}$	0.0933	0.8559	4.7373	0.2627	0.8574	5.1600
[US] $(\overline{u})$	[2.1872]	[460]	0.4269	0.8774	4.9200	0.2790	0.8587	5.0800
[US] $(\overline{v})$	[1.5108]	[9900]	0.0967	0.7869	4.5600	0.0966	0.7870	4.5600
[US] $(\overline{u}, \overline{v})$	$[2.1872, 1.5108]$	$[460, 9900]$	0.3759	0.8665	4.8400	0.3237	0.9210	5.6800

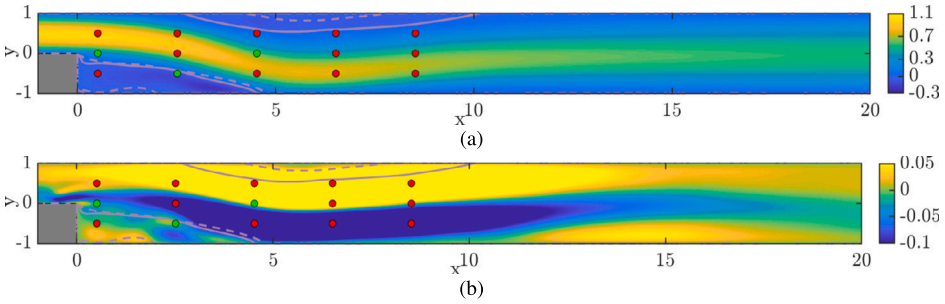

Fig. 17. Data assimilation of mean flow over a backward-facing step at $Re = 500$ using a Huber loss function to represent the error in measurements, solved with L-BFGS: Stream-wise component of (a) the assimilated mean flow and (b) the error between the assimilated and the true mean flows. For details on the terminology of sensor locations, and the re-circulation region, see the caption of Fig. 16, and the percent uncertainty, see Table 7.

concentrated in the region around the outlier locations, and the noise from the outliers affects the quality of the reconstructed mean flow downstream of the step, as shown in the Fig. 17b.

We also consider the hybrid representation for the same case, and see that the length of the re-circulation region on the bottom wall closely matches the true mean flow, as shown in Fig. 18a. While the length of the re-circulation region on the top wall is under-predicted, however, the relative uncertainty in the length of the re-circulation region on the bottom wall falls sharply to 2.5%. With the hybrid loss function representation of the deviation error, the global reconstruction error reduces drastically to around 10% (see Table 7 for details), and this can be seen from Fig. 18b.

Next, we consider a case of data assimilation, where we introduce outliers at six sensor locations in the region downstream of the step keeping the percent uncertainty level at 147.8% for each outlier location. While using the Huber loss function representation of the deviation error, the length of the re-circulation region on the bottom wall is under-predicted with a relative uncertainty of 1.3%. However, with the hybrid loss function representation of the deviation error, the length of the re-circulation region on the bottom wall is over-predicted and the relative uncertainty increases to 7.5%, see Table 7 for further details. Furthermore, the global reconstruction error increases from 9% for the Huber loss function to around 26% for the hybrid loss function case, see Table 7.

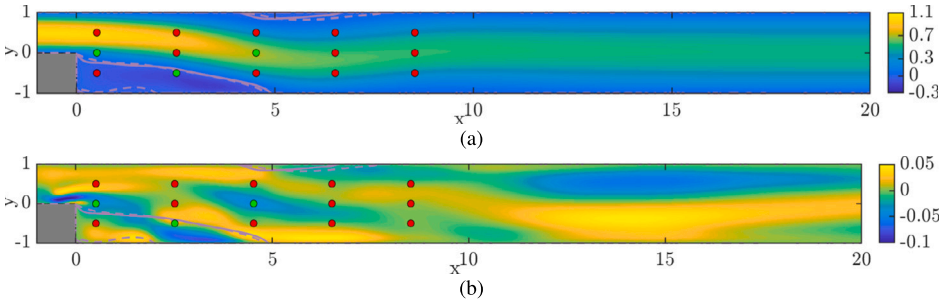

Fig. 18. Data assimilation of mean flow over a backward-facing step at $Re = 500$ using a hybrid loss function to represent the error in measurements, solved with L-BFGS: Stream-wise component of (a) the assimilated mean flow, and (b) the error between the assimilated and the true mean flows. For details on the terminology of sensor locations, and the re-circulation region, see the caption of Fig. 16, and the percent uncertainty, see Table 7.

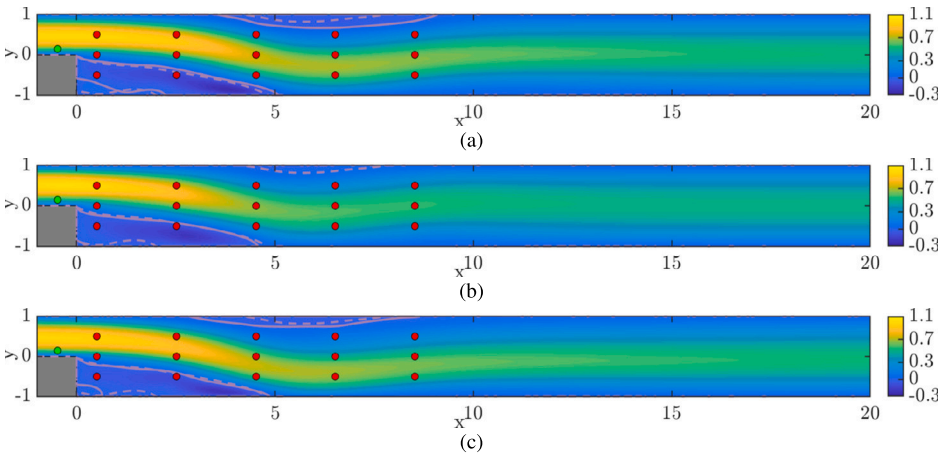

Fig. 19. Data assimilation of mean flow over a backward-facing step at $Re = 500$ using the (a,b) hybrid and (c) Huber loss functions to represent the error in measurements, solved with L-BFGS: Stream-wise component of the assimilated mean flow. The outliers (denoted by •) are introduced along (a) the horizontal component $(\overline{u})$, (b) the vertical component $(\overline{v})$, and (c) the horizontal and vertical components $(\overline{u}, \overline{v})$ of the mean velocity at the sensor location [US] upstream of the step amounting to percent uncertainty: (a,b) [460], (c,d) [9900], and (e,f) [460, 9900]. For details on the re-circulation region, see the caption of Fig. 16.

However, considering the criticality of the location of the outliers, the performance of the Huber loss function in recovering the mean flow field with reduced uncertainty is certainly striking.

As a check of the robustness of the reconstruction framework for an amplifier flow, we consider the cases of data assimilation where we introduce outliers at a sensor (denoted by US) located upstream of the step. Initially, we introduce an outlier only along the horizontal mean velocity component at the sensor US for which the stream-wise component of the reconstructed mean flow using the hybrid loss function is shown in Fig. 19a, respectively. The reconstructed mean flow accounts for a global error of 28% using the hybrid loss function as compared to 42% using the Huber loss function while the local error remains almost the same for both loss functions, see Table 7 for further details. However, the length of the re-circulation region is over-predicted by 2.5% using the Huber loss function and 6% using the hybrid loss function. When the outlier is introduced only along the vertical mean velocity component at the sensor US as shown in Fig. 19b, the performance of both the Huber and hybrid loss functions is similar with a global reconstruction error of around 9% (see Table 7), and the length of the re-circulation region is under-predicted by 5%. Finally, we consider the case of data assimilation where the outliers are introduced along the horizontal and vertical mean velocity components at the sensor US for which the reconstructed mean flow using the Huber loss function is shown in Fig. 19c, respectively. From Table 7, we can see that the uncertainty in the prediction of the length of the re-circulation region increases from 1% using the Huber loss function to around 18% using the hybrid loss function while the global reconstruction error differs by around 6%. We notice the inconsistency in the performance of Huber and hybrid loss functions for the mean flow reconstruction over a backward-facing step, and the reason for this behavior remains to be investigated.

7. Conclusions

We introduced a robust numerical framework for the accurate reconstruction of mean flows. The methodology adds a volume forcing term to the momentum equation to account for the divergence of the Reynolds stresses in the RANS equations, similar to the study of Foures et al., 2014 [10]. Flow around a circular cylinder demonstrated the performance of different loss functions, used to represent the point-wise error between the measured and the assimilated mean velocities. With sparsity in the measurement data,

we demonstrated that the quadratic representation of the error is sufficient to recover the mean flow with acceptable accuracy, once a regularization function, which promotes a smooth mean flow field, is introduced.

With the introduction of outliers in the sparse measurement data set, the quadratic loss function is no longer able to reconstruct the mean flow accurately. This failure is intrinsic to the choice of the loss function and is unrelated to the algorithm employed to solve the data assimilation problem. An absolute (ℓ_1) function, assigning weights to the point-wise error in a linear rather than quadratic fashion, has no significant effect on the accuracy of the assimilated mean flow. Moreover, the ℓ_1 loss function caused strong variations in the mean flow due to parameter fluctuations in the employed algorithm. With an IRLS transformation of the original ℓ_1 loss function, the quality of the reconstruction improved markedly in the limit of a vanishing IRLS parameter. Yet, an outlier near the point of separation on the circular cylinder surface caused a deterioration of the reconstruction of the mean flow re-circulation region. Finally, we demonstrated that the Huber loss function gracefully and effectively handles outliers in the measurement data. The Huber function combines the uniqueness from the ℓ_2 part and the robustness to large deviation errors from the ℓ_1 part. We also used a hybrid loss function as a continuous and differentiable proxy that approximates the combined effects of ℓ_1/ℓ_2 functions in limiting cases of the cut-off parameter. The hybrid loss function has the additional advantage, over the Huber loss function, of access to second-order (Hessian) information. The effective estimation of the epsilon/cut-off parameter is rather critical for the data assimilation process, as this parameter determines the switch between ℓ_1 and ℓ_2 behavior. We provide a mathematical framework to predict an approximate value of this parameter based on a few initial iterations using a quadratic loss function. In addition, we also observed that the presence of Gaussian noise at the measurement locations does not alter the performance of the hybrid loss function in the recovery of the mean flow. Furthermore, the performance of the hybrid loss function to recover the mean flow is markedly better than an equivalent ℓ_2 counterpart and this is reflected in the global mean flow assimilation error.

The effect of localized outliers on the recoverability of the mean flow around the circular cylinder has been explored as well. When the outlier location is asymmetrical with respect to the symmetry axis of the circular cylinder, the recovered mean flow also loses symmetry. Similarly, when the outliers are more localized along the symmetry axis of the circular cylinder, the recovered mean flow is symmetrical but the re-circulation region is strongly over/under-predicted by a quadratic loss function, but under no circumstances do we recover the true re-circulation zone using a ℓ_2 loss function. The Huber/hybrid loss function recovers the mean flow with diminished relative uncertainty, irrespective of the location and intensity of the outliers.

We demonstrated the robustness of the data assimilation framework through the reconstruction of the mean flow for an amplifier backward-facing step flow using a noisy sparse measurement data set. We considered strong levels of outliers (many standard deviations from the mean) very close to and upstream of the backward-facing step and still recover the mean flow with sufficient accuracy. All these findings therefore strongly encourage the application of the present methodologies to actual experimental measurements in future work. Furthermore, while this study focused on mean flow reconstruction, it may be worth noting that the present data assimilation framework could be straightforwardly extended to the assimilation of instantaneous measurements, changing the interpretation of the optimized volume forcing term as an unknown Eulerian acceleration [51].

CRediT authorship contribution statement

Souvik Ghosh: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Vincent Mons:** Writing – review & editing, Supervision, Software. **Denis Sipp:** Writing – review & editing, Supervision. **Peter J. Schmid:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by the European Union's Framework Programme for Research and Innovation "Horizon 2020" under the Marie Skłodowska-Curie Innovative Training Programme grant agreement no. 675008. S. Ghosh acknowledges computational resources provided by the HPC service at Imperial College London.

Appendix A. Validation of optimization framework

We perform data assimilation for mean flow around a circular cylinder at $Re = 150$, with the same parameter values studied by Foures et al., 2014 [10]. We use the complete information about the mean flow as input to the data assimilation framework, where the optimization procedure is initialized by the base flow. The stream-wise component of the mean velocity is shown in Fig. A.20a. It matches well the result of Foures et al., 2014 [10]. The normalized objective function $\mathcal{M}_r$ in Fig. A.20b clearly shows the efficiency of the present data assimilation framework to reduce the relative error to 10^{-6} within 500 iterations. We also display the stream-wise

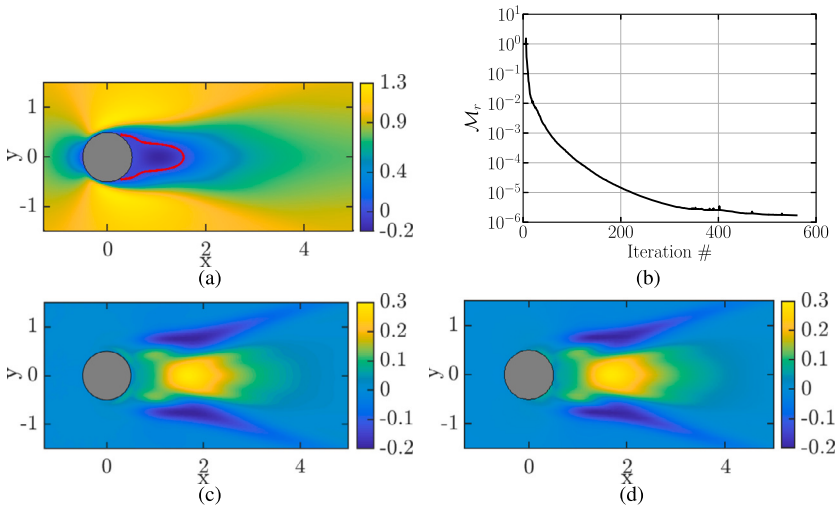

Fig. A.20. Data assimilation of mean flow around a circular cylinder at $Re = 150$ using a ℓ_2 loss function to represent the error in measurements, solved with L-BFGS: Stream-wise component of (a) the assimilated mean flow, (c) the assimilated forcing, and (d) the solenoidal part of the forcing from DNS. (b) shows the convergence of the normalized objective function. The solid and the dashed lines refer to the re-circulation region for the assimilated and the true mean flows, and they are not distinguishable as they are overlapping.

component of the recovered forcing in Fig. A.20c and the solenoidal part of the true forcing from DNS in Fig. A.20d. Again, we obtain a good match with the result of Foures et al., 2014 [10].

Appendix B. Sub-gradient calculation for different loss functions

Any function is called sub-differentiable at a point if at least one sub-gradient exists at that point. The sub-differential of a function g is denoted by $\partial g(x)$ and consists of the set of all sub-gradients of g at the point x, see Boyd and Vandenberghe, 2004 [40] for more details.

B.1. Sub-gradient for ℓ_2 loss function

For a function that is differentiable at a point, there exists a unique sub-gradient given by the gradient of the function at that point. An exact gradient of the Lagrangian with $\mathcal{M}(\mathbf{z}) = \sum_i |z_i|^2 \Big/ 2$ in Step - 2 of Algorithm 1 can be calculated by taking the variation of the Lagrangian with respect to the variation in $\mathbf{z}$ and is given as

$$\nabla_{\mathbf{z}} \mathcal{M}(\mathbf{z}) = \mathbf{z} - \mu^{\{k\}} \left[\mathbf{r}^{\{k+1\}} - \mathbf{z} - \frac{\lambda^{\{k\}}}{\mu^{\{k\}}} \right] \tag{B.1}$$

On setting the gradient in the above equation (B.1) equal to zero, the update for the variable $\mathbf{z}$ is obtained as

$$\mathbf{z}^{\{k+1\}} = \frac{\mu^{\{k\}}}{1 + \mu^{\{k\}}} \left[\mathbf{r}^{\{k+1\}} - \frac{\lambda^{\{k\}}}{\mu^{\{k\}}} \right] \tag{B.2}$$

B.2. Sub-gradient for ℓ_1 loss function

For an absolute value function ($g(z) = |z|$), a unique sub-gradient exists for the intervals $z < 0$ and $z > 0$, and their respective sub-differentials are given as $\partial g(z)|_{(z<0)} = -1$ and $\partial g(z)|_{(z>0)} = +1$. For $z = 0$, the set of all sub-gradients is given by the inequality function $|z| = hz$ where $h \in [-1, 1]$. Hence the sub-differential at $z = 0$ is given by $\partial g(z)|_{(z=0)} = [-1, 1]$.
The sub-differential for the Lagrangian with $\mathcal{M}(\mathbf{z}) = \sum_i |z_i|$ in Step - 2 of Algorithm 1, is calculated by considering all the set of sub-gradients and that is obtained by setting $z_i = 0$ at each measurement location. We obtain the interval

$$\left[-1 - \mu^{\{k\}} \left(r_i^{\{k+1\}} - \frac{\lambda_i^{\{k\}}}{\mu^{\{k\}}} \right), 1 - \mu^{\{k\}} \left(r_i^{\{k+1\}} - \frac{\lambda_i^{\{k\}}}{\mu^{\{k\}}} \right) \right] ; \ \forall r_i \in \mathbf{r}, \ \forall \lambda_i \in \lambda \tag{B.3}$$

If $\zeta_i = 0$ is a sub-gradient in the above interval, then $z_i = 0$ is optimal. This occurs precisely when

$$\frac{1}{\mu^{\{k\}}} \geq \left| r_i^{\{k+1\}} - \frac{\lambda_i^{\{k\}}}{\mu^{\{k\}}} \right| ; \ \forall r_i \in \mathbf{r}, \ \forall \lambda_i \in \lambda \tag{B.4}$$

Algorithm 1 ADMM.

Input: Initialize $\mathbf{f}^{\{0\}} = \vec{\mathbf{0}}$, and find first guess for the velocity field $\overline{\mathbf{u}}^{\{0\}} = \overline{\mathbf{u}}\left(\mathbf{f}^{\{0\}}\right)$

Put $\mathbf{z}^{\{0\}} = \mathcal{H}\left(\overline{\mathbf{u}}^{\{0\}}\right) - \widetilde{\mathbf{u}},\ \lambda^{\{0\}} = \vec{\mathbf{0}},\ \mu^{\{0\}} > 0,\ \delta > 0$

Output: Optimum values of $\mathbf{f}$, $\mathbf{z}$

Procedure:

while $\sqrt{[\mathbf{r}-\mathbf{z}]^{\mathrm{T}}[\mathbf{r}-\mathbf{z}]} > \delta$ and $\sqrt{\left[\overline{\mathbf{u}}_{-}^{\dagger}(\mathbf{f})\right]^{\mathrm{T}}\left[\overline{\mathbf{u}}_{-}^{\dagger}(\mathbf{f})\right]} > \delta$ **do**

Step - 1. Fix $\mathbf{z}^{\{k\}}, \lambda^{\{k\}}$ and solve for $\mathbf{f}^{\{k+1\}}$:

$$\mathbf{f}^{\{k+1\}} = \underset{\mathbf{f}}{\operatorname{argmin}} \frac{\mu^{\{k\}}}{2}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right) - \widetilde{\mathbf{u}} - \mathbf{z}^{\{k\}} - \frac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]^{\mathrm{T}}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right) - \widetilde{\mathbf{u}} - \mathbf{z}^{\{k\}} - \frac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right] + \mathcal{N}(\mathbf{f}) - \int_{\Omega^r} \overline{p}^{\dagger}\left\{\nabla\cdot\overline{\mathbf{u}}\right\}\, d\Omega - \int_{\Omega^r}\left(\overline{\mathbf{u}}^{\dagger}\right)^{\mathrm{T}}\left\{\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}} + \nabla\overline{p} - \frac{1}{Re}\nabla^2\overline{\mathbf{u}} - \mathbf{f}\right\}\, d\Omega$$

Step - 2. Fix $\mathbf{f}^{\{k+1\}}, \lambda^{\{k\}}$ and solve for $\mathbf{z}^{\{k+1\}}$:

$$\mathbf{z}^{\{k+1\}} = \underset{\mathbf{z}}{\operatorname{argmin}}\ \mathcal{M}(\mathbf{z}) + \frac{\mu^{\{k\}}}{2}\left[\mathbf{r}^{\{k+1\}} - \mathbf{z} - \frac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]^{\mathrm{T}}\left[\mathbf{r}^{\{k+1\}} - \mathbf{z} - \frac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]$$

where $\mathbf{r}^{\{k+1\}} = \mathcal{H}\left(\overline{\mathbf{u}}\left(\mathbf{f}^{\{k+1\}}\right)\right) - \widetilde{\mathbf{u}}$

The update for this step for different loss functions is provided in Appendix B.

Step - 3. Fix $\mathbf{f}^{\{k+1\}}, \mathbf{z}^{\{k+1\}}$ and update $\lambda^{\{k+1\}}$:

$$\lambda^{\{k+1\}} = \lambda^{\{k\}} - \mu^{\{k\}}\left(\mathbf{r}^{\{k+1\}} - \mathbf{z}^{\{k+1\}}\right)$$

Step - 4. Update $\mu^{\{k+1\}}$:

$$\mu^{\{k+1\}} = \begin{cases} \rho\mu^{\{k\}}, & \textbf{if } \sqrt{\left[\mathbf{r}^{\{k+1\}} - \mathbf{z}^{\{k+1\}}\right]^{\mathrm{T}}\left[\mathbf{r}^{\{k+1\}} - \mathbf{z}^{\{k+1\}}\right]} > \left(\eta \times \sqrt{\left[\overline{\mathbf{u}}_{-}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right)\right]^{\mathrm{T}}\left[\overline{\mathbf{u}}_{-}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right)\right]}\right) \\ \mu^{\{k\}}/\rho, & \textbf{if } \sqrt{\left[\overline{\mathbf{u}}_{-}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right)\right]^{\mathrm{T}}\left[\overline{\mathbf{u}}_{-}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right)\right]} > \left(\eta \times \sqrt{\left[\mathbf{r}^{\{k+1\}} - \mathbf{z}^{\{k+1\}}\right]^{\mathrm{T}}\left[\mathbf{r}^{\{k+1\}} - \mathbf{z}^{\{k+1\}}\right]}\right) \\ \mu^{\{k\}}, & \textbf{otherwise} \end{cases}$$

where $\rho > 1,\ \eta > 1$

The primal and the dual condition residuals at $(k+1)^{th}$ iteration of the ADMM algorithm are denoted as $\sqrt{\left[\mathbf{r}^{\{k+1\}} - \mathbf{z}^{\{k+1\}}\right]^{\mathrm{T}}\left[\mathbf{r}^{\{k+1\}} - \mathbf{z}^{\{k+1\}}\right]}$ and $\sqrt{\left[\overline{\mathbf{u}}_{-}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right)\right]^{\mathrm{T}}\left[\overline{\mathbf{u}}_{-}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right)\right]}$, and are discussed in detail in Appendix C respectively.

end while

If $\left(r_i^{\{k+1\}} - \dfrac{\lambda_i^{\{k\}}}{\mu^{\{k\}}}\right) > \dfrac{1}{\mu^{\{k\}}}$ then all the sub-gradients at $z_i = 0$ are negative, so the minimizer is achieved at some positive z_i and this occurs at

$$z_i = r_i^{\{k+1\}} - \frac{\lambda_i^{\{k\}}}{\mu^{\{k\}}} - \frac{1}{\mu^{\{k\}}};\ \forall r_i \in \mathbf{r},\ \forall \lambda_i \in \lambda,\ \forall z_i \in \mathbf{z} \tag{B.5}$$

Similarly, if $\left(r_i^{\{k+1\}} - \dfrac{\lambda_i^{\{k\}}}{\mu^{\{k\}}}\right) < -\dfrac{1}{\mu^{\{k\}}}$ then all the sub-gradients at $z_i = 0$ are positive, so the minimizer is achieved at some negative z_i and this occurs at

$$z_i = r_i^{\{k+1\}} - \frac{\lambda_i^{\{k\}}}{\mu^{\{k\}}} + \frac{1}{\mu^{\{k\}}};\ \forall r_i \in \mathbf{r},\ \forall \lambda_i \in \lambda,\ \forall z_i \in \mathbf{z} \tag{B.6}$$

A soft-shrinkage operator is defined as

$$S_{\{1/\mu\}}\left(\mathbf{r} - \frac{\lambda}{\mu}\right) = \begin{cases} r_i - \dfrac{\lambda_i}{\mu} - \dfrac{1}{\mu}, & \left(r_i - \dfrac{\lambda_i}{\mu}\right) > \dfrac{1}{\mu} \\ 0, & -\dfrac{1}{\mu} \le \left(r_i - \dfrac{\lambda_i}{\mu}\right) \le \dfrac{1}{\mu} \\ r_i - \dfrac{\lambda_i}{\mu} + \dfrac{1}{\mu}, & \left(r_i - \dfrac{\lambda_i}{\mu}\right) < -\dfrac{1}{\mu} \end{cases} \tag{B.7}$$

which can be written in compact notation as

$$\begin{aligned} \mathcal{S}_{\{1/\mu\}}\left(\mathbf{r}-\frac{\lambda}{\mu}\right) &= \left(1-\frac{\frac{1}{\mu}}{\sqrt{\left[\mathbf{r}-\frac{\lambda}{\mu}\right]^{\mathrm{T}}\left[\mathbf{r}-\frac{\lambda}{\mu}\right]}}\right)_{+}\left(\mathbf{r}-\frac{\lambda}{\mu}\right) \\ &= \max\left\{\sqrt{\left[\mathbf{r}-\frac{\lambda}{\mu}\right]^{\mathrm{T}}\left[\mathbf{r}-\frac{\lambda}{\mu}\right]}-\frac{1}{\mu},0\right\}\cdot\frac{\mathbf{r}-\frac{\lambda}{\mu}}{\sqrt{\left[\mathbf{r}-\frac{\lambda}{\mu}\right]^{\mathrm{T}}\left[\mathbf{r}-\frac{\lambda}{\mu}\right]}} \end{aligned} \tag{B.8}$$

where $(\cdot)_+$ denotes the positive part of its argument.
Hence the soft-shrinkage operator for an update of $\mathbf{z}$ can be obtained as

$$\begin{aligned} \mathbf{z}^{\{k+1\}} &= \ell_1\text{SHRINKAGE}\left(\mu^{\{k\}},\mathbf{r}^{\{k+1\}},\lambda^{\{k\}}\right) \\ &= \max\left\{\sqrt{\left[\mathbf{r}^{\{k+1\}}-\frac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]^{\mathrm{T}}\left[\mathbf{r}^{\{k+1\}}-\frac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]}-\frac{1}{\mu^{\{k\}}},0\right\}\cdot\frac{\mathbf{r}^{\{k+1\}}-\frac{\lambda^{\{k\}}}{\mu^{\{k\}}}}{\sqrt{\left[\mathbf{r}^{\{k+1\}}-\frac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]^{\mathrm{T}}\left[\mathbf{r}^{\{k+1\}}-\frac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]}} \end{aligned} \tag{B.9}$$

B.3. Sub-gradient for Huber loss function

The sub-gradient for the Huber loss function can be derived following along similar lines of ℓ_2 and ℓ_1 norms and can be written as

$$\begin{aligned} \mathbf{z}^{\{k+1\}} &= \text{HUBERSHRINKAGE}\left(\mu^{\{k\}},\mathbf{r}^{\{k+1\}},\lambda^{\{k\}},\epsilon\right) \\ &= \begin{cases} \dfrac{\mu^{\{k\}}}{1+\mu^{\{k\}}}\left[r_i^{\{k+1\}}-\dfrac{\lambda_i^{\{k\}}}{\mu^{\{k\}}}\right] & ,\left|r_i^{\{k+1\}}\right|\le\epsilon_i \\ \max\left\{\sqrt{\left[\mathbf{r}^{\{k+1\}}-\dfrac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]^{\mathrm{T}}\left[\mathbf{r}^{\{k+1\}}-\dfrac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]},0\right\}\cdot\dfrac{r_i^{\{k+1\}}-\dfrac{\lambda_i^{\{k\}}}{\mu^{\{k\}}}}{\sqrt{\left[\mathbf{r}^{\{k+1\}}-\dfrac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]^{\mathrm{T}}\left[\mathbf{r}^{\{k+1\}}-\dfrac{\lambda^{\{k\}}}{\mu^{\{k\}}}\right]}} & ,\left|r_i^{\{k+1\}}\right|>\epsilon_i \end{cases} \end{aligned} \tag{B.10}$$

The Huber shrinkage operator follows a ℓ_2 sub-gradient calculation within and at the epsilon limit, and a ℓ_1 sub-gradient calculation strictly outside the epsilon limit, as shown in equation (B.10).

Appendix C. Optimal conditions for ADMM

The unconstrained Lagrangian in equation (21) without the penalty regularization term can be written as

$$\begin{aligned} \mathcal{G}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],[\mathbf{z},\lambda],\mathbf{f}\right) =& \mathcal{M}(\mathbf{z})+\mathcal{N}(\mathbf{f})-\lambda^{\mathrm{T}}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right)-\widetilde{\mathbf{u}}-\mathbf{z}\right] \\ &-\int_{\Omega^F}\overline{p}^{\dagger}\left\{\nabla\cdot\overline{\mathbf{u}}\right\}\,d\Omega-\int_{\Omega^F}\left(\overline{\mathbf{u}}^{\dagger}\right)^{\mathrm{T}}\left\{\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}}+\nabla\overline{p}-\frac{1}{Re}\nabla^2\overline{\mathbf{u}}-\mathbf{f}\right\}\,d\Omega \end{aligned} \tag{C.1}$$

The original Lagrangian in equation (21) can then be written as

$$\mathcal{L}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],[\mathbf{z},\lambda],\mathbf{f}\right)=\mathcal{G}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],[\mathbf{z},\lambda],\mathbf{f}\right)+\frac{\mu}{2}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right)-\widetilde{\mathbf{u}}-\mathbf{z}\right]^{\mathrm{T}}\left[\mathcal{H}\left(\overline{\mathbf{u}}\right)-\widetilde{\mathbf{u}}-\mathbf{z}\right] \tag{C.2}$$

C.1. Primal feasibility condition

The primal feasibility condition is given by

$$\mathcal{H}\left(\overline{\mathbf{u}}\left(\mathbf{f}^*\right)\right)-\widetilde{\mathbf{u}}-\mathbf{z}^*=0 \tag{C.3}$$

where variables with * denote the optimum for the vector variables $\mathbf{f}$ and $\mathbf{z}$.

C.2. Dual feasibility condition

The dual feasibility condition is obtained by taking the sub-differential of equation (C.1) with respect to variations in the variables $\mathbf{f}$ and $\mathbf{z}$. Taking the sub-differential of equation (C.1) with respect to $\mathbf{f}$ gives

$$\partial\mathcal{G} = 0 \in \partial\mathcal{N}\left(\mathbf{f}^*\right) + \overline{\mathbf{u}}_*^{\dagger}\left(\mathbf{f}^*\right) \tag{C.4}$$

where $\overline{\mathbf{u}}_*^{\dagger}$ is the solution to equations (C.5) given below.

$$-\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}}_*^{\dagger} + \overline{\mathbf{u}}_*^{\dagger}\cdot\left(\nabla\overline{\mathbf{u}}\right)^{\mathrm{T}} - \nabla\overline{p}_*^{\dagger} - \frac{1}{Re}\nabla^2\overline{\mathbf{u}}_*^{\dagger} = -\mathcal{H}^{\mathrm{T}}\lambda^*;\ \nabla\cdot\overline{\mathbf{u}}_*^{\dagger} = 0 \tag{C.5}$$

Taking the sub-differential of equation (C.1) with respect to $\mathbf{z}$ again leads to

$$\partial\mathcal{G} = 0 \in \partial\mathcal{M}(\mathbf{z}^*) + \lambda^* \tag{C.6}$$

Since $\mathbf{z}^{\{k+1\}}$ minimizes $\mathcal{L}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],\left[\mathbf{z},\lambda^{\{k\}}\right],\mathbf{f}^{\{k+1\}}\right)$ by definition, we have from equations (C.1) and (C.2)

$$\begin{aligned} 0 &\in \partial\mathcal{M}(\mathbf{z}^{\{k+1\}}) + \lambda^{\{k\}} - \mu\left[\mathcal{H}\left(\overline{\mathbf{u}}\left(\mathbf{f}^{\{k+1\}}\right)\right) - \widetilde{\mathbf{u}} - \mathbf{z}^{\{k+1\}}\right] \\ &\in \partial\mathcal{M}(\mathbf{z}^{\{k+1\}}) + \lambda^{\{k+1\}} \end{aligned} \tag{C.7}$$

This means $\mathbf{z}^{\{k+1\}}$ and $\lambda_i^{\{k+1\}}$ always satisfy equation (C.6), so attaining optimality comes down to satisfying equation (C.3) and equation (C.4).

Since $\mathbf{f}^{\{k+1\}}$ minimizes $\mathcal{L}\left(\left[\overline{\mathbf{u}},\overline{p}\right],\left[\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right],\left[\mathbf{z}^{\{k\}},\lambda^{\{k\}}\right],\mathbf{f}\right)$ by definition, we have from equations (C.1) and (C.2)

$$0 \in \mathcal{N}\left(\mathbf{f}^{\{k+1\}}\right) + \overline{\mathbf{u}}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right) \tag{C.8}$$

where $\overline{\mathbf{u}}^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right)$ is the solution to equations (C.9) given as

$$-\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}}^{\dagger} + \overline{\mathbf{u}}^{\dagger}\cdot\nabla\overline{\mathbf{u}}^{T} - \nabla\overline{p}^{\dagger} - \frac{1}{Re}\nabla^2\overline{\mathbf{u}}^{\dagger} = -\mathcal{H}^{\mathrm{T}}\lambda^{\{k\}} + \mu\mathcal{H}^{\mathrm{T}}\left[\mathcal{H}\left(\overline{\mathbf{u}}\left(\mathbf{f}^{\{k+1\}}\right)\right) - \widetilde{\mathbf{u}} - \mathbf{z}^{\{k\}}\right];\ \nabla\cdot\overline{\mathbf{u}}^{\dagger} = 0 \tag{C.9}$$

The right-hand side forcing term in equation (C.9) can be simplified as follows

$$\begin{aligned} &-\mathcal{H}^{\mathrm{T}}\lambda^{\{k\}} + \mu\mathcal{H}^{\mathrm{T}}\left[\mathcal{H}\left(\overline{\mathbf{u}}\left(\mathbf{f}^{\{k+1\}}\right)\right) - \widetilde{\mathbf{u}} - \mathbf{z}^{\{k+1\}} + \mathbf{z}^{\{k+1\}} - \mathbf{z}^{\{k\}}\right] \\ &= -\mathcal{H}^{\mathrm{T}}\lambda^{\{k+1\}} + \mu\mathcal{H}^{\mathrm{T}}\left[\mathbf{z}^{\{k+1\}} - \mathbf{z}^{\{k\}}\right] \end{aligned} \tag{C.10}$$

Since the adjoint equations (C.9) are linear with respect to the direct variables $\left(\overline{\mathbf{u}},\overline{p}\right)$, the adjoint variables can be decomposed as $\left(\overline{\mathbf{u}}^{\dagger},\overline{p}^{\dagger}\right) = \left(\overline{\mathbf{u}}_*^{\dagger},\overline{p}_*^{\dagger}\right) + \left(\overline{\mathbf{u}}_-^{\dagger},\overline{p}_-^{\dagger}\right)$. The decomposed adjoint equations with the simplified forcing from equation (C.9) can be written as

$$-\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}}_*^{\dagger} + \overline{\mathbf{u}}_*^{\dagger}\cdot\left(\nabla\overline{\mathbf{u}}\right)^{\mathrm{T}} - \nabla\overline{p}_*^{\dagger} - \frac{1}{Re}\nabla^2\overline{\mathbf{u}}_*^{\dagger} = -\mathcal{H}^{\mathrm{T}}\lambda^{\{k+1\}};\ \nabla\cdot\overline{\mathbf{u}}_*^{\dagger} = 0 \tag{C.11}$$

$$-\overline{\mathbf{u}}\cdot\nabla\overline{\mathbf{u}}_-^{\dagger} + \overline{\mathbf{u}}_-^{\dagger}\cdot\left(\nabla\overline{\mathbf{u}}\right)^{\mathrm{T}} - \nabla\overline{p}_-^{\dagger} - \frac{1}{Re}\nabla^2\overline{\mathbf{u}}_-^{\dagger} = \mu\mathcal{H}^{\mathrm{T}}\left[\mathbf{z}^{\{k+1\}} - \mathbf{z}^{\{k\}}\right];\ \nabla\cdot\overline{\mathbf{u}}_-^{\dagger} = 0 \tag{C.12}$$

It can be verified that equations (C.11) are already satisfied by the sub-differential equations (C.5). This means the quantity

$$\mathbf{d}^{\{k+1\}} = \overline{\mathbf{u}}_-^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right) \tag{C.13}$$

which is given by the solution to equations (C.12), can be viewed as a residual for the dual feasibility condition. Hence the primal-dual condition residuals at $(k+1)^{th}$ iteration step of the optimization are given as

$$\text{Primal condition}: \mathbf{p}^{\{k+1\}} = \mathcal{H}\left(\overline{\mathbf{u}}\left(\mathbf{f}^{\{k+1\}}\right)\right) - \widetilde{\mathbf{u}} - \mathbf{z}^{\{k+1\}} \tag{C.14}$$

$$\text{Dual condition}: \mathbf{d}^{\{k+1\}} = \overline{\mathbf{u}}_-^{\dagger}\left(\mathbf{f}^{\{k+1\}}\right) \tag{C.15}$$

Appendix D. Optimal value of cut-off parameter

Guitton and Symes, 2003 [39] took the value of the epsilon/cut-off parameter $\left(\epsilon_i\right)$ as the 98th percentile of the data or 0.01 times the maximum deviation in error. However, these estimates lack proper analytical support or justification. A more effective approach for selecting this parameter is based on the estimate of the deviation error of the data, as proposed by Bube and Langan, 1997 [22]. Following this general strategy, we evaluate the standard deviation for the hybrid distribution and obtain an *a priori* estimate of the cut-off parameter $\left(\epsilon_i\right)$. With $[\mathbf{r},\boldsymbol{\epsilon}]$ representing the vector of all measurement sensor locations, the mathematical approach described below is presented for a particular measurement location, denoted by $\left[r_i,\epsilon_i\right]$.

D.1. Hybrid distribution

Under the assumption that the noise obeys a hybrid distribution at a sensor location, the associated loss function is expressed in hybrid form. The standard deviation (S.D.) of r_i for a hybrid distribution can be calculated using the density function, $\mathbb{F}\left(r_i\right)$ as

$$\text{S.D.} = \left[\int_{-\infty}^{+\infty} r_i^2 \mathbb{F}\left(r_i\right)\, dr_i - \int_{-\infty}^{+\infty} r_i \mathbb{F}\left(r_i\right)\, dr_i\right]^{1/2} = \left[\frac{\mathcal{Y}+\mathcal{Z}}{\mathcal{X}}\right]^{1/2} \epsilon_i \approx 1.643\, \epsilon_i$$
$$\text{where, } \mathcal{Y} = \int_0^{\infty} r_i \exp\left(-r_i\right)\sqrt{{r_i}^2 + 2r_i}\, dr_i \approx 2.78, \text{ and } \mathcal{Z} = \int_0^{\infty} \exp\left(-r_i\right)\sqrt{{r_i}^2 + 2r_i}\, dr_i \approx 1.636. \tag{D.1}$$

In the above equation (D.1), the value of $\mathcal{X}$ is taken from equation (16). From the above formulation, an optimal value for the epsilon/cut-off parameter (ϵ_i) at a particular measurement location can be obtained as follows.

- A few initial iterations using a ℓ_2 loss function are performed, which do not require knowledge of the cut-off parameter;
- Then, the deviation error using equation (4) is calculated as r_i;
- Finally, the cut-off parameter $\left(\epsilon_i\right)$ is chosen as $0.6 (= 1/1.643)$ times the absolute deviation error, $\left|r_i\right|$. Using this value of ϵ_i and dividing by $\left|r_i\right|$ in the above equation (D.1), the standard deviation attains a value of 1 (same as that of a standard normal distribution for the quadratic loss function, see Table 1).

This cut-off/epsilon limit denotes the deviation at which the loss function can switch from ℓ_1 to ℓ_2.

References

[1] F.X.L. Dimet, O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, Tellus 38A (2) (1986) 97–110.
[2] E.R. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 82 (Ser. D) (1960) 35–45.
[3] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., Oceans 99 (C5) (1994) 10143–10162.
[4] G. Evensen, Data Assimilation: The Ensemble Kalman Filter, 2nd edition, Springer Science & Business Media, Berlin, 2009.
[5] K. Zhou, J.C. Doyle, K. Glover, Robust and Optimal Control, Prentice Hall, 1995.
[6] M.D. Gunzburger, Perspectives in Flow Control and Optimization, SIAM, 2002.
[7] J. Kim, T.R. Bewley, A linear systems approach to flow control, Annu. Rev. Fluid Mech. 39 (2007) 383–417.
[8] O. Talagrand, P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: theory, Tellus 113 (478) (1987) 1311–1328.
[9] A. Gronskis, D. Heitz, E. Mémin, Inflow and initial conditions for direct numerical simulation based on adjoint data assimilation, J. Comput. Phys. 242 (2013) 480–497.
[10] D.P.G. Foures, N. Dovetta, D. Sipp, P.J. Schmid, A data-assimilation method for Reynolds-averaged Navier-Stokes-driven mean flow reconstruction, J. Fluid Mech. 759 (2014) 404–431.
[11] S. Symon, N. Dovetta, B.J. McKeon, D. Sipp, P.J. Schmid, Data assimilation of mean velocity from 2D PIV measurements of flow over an idealized airfoil, Exp. Fluids 58 (2017) 61.
[12] V. Mons, J.C. Chassaing, T. Gomez, P. Sagaut, Reconstruction of unsteady viscous flows using data assimilation schemes, J. Comput. Phys. 316 (2016) 255–280.
[13] D. Cerizza, W. Sekiguchi, T. Tsukahara, T.A. Zaki, Y. Hasegawa, Reconstruction of scalar source intensity based on sensor signal in turbulent channel flow, Flow Turbul. Combust. 97 (2016) 1211–1233.
[14] Q. Wang, Y. Hasegawa, T.A. Zaki, Spatial reconstruction of steady scalar sources from remote measurements in turbulent flow, J. Fluid Mech. 870 (2019) 316–352.
[15] V. Mons, Q. Wang, T.A. Zaki, Kriging-enhanced ensemble variational data assimilation for scalar-source identification in turbulent environments, J. Comput. Phys. 398 (2019) 108856–1–108856–28.
[16] V. Mons, Y. Du, T.A. Zaki, Ensemble-variational assimilation of statistical data in large-eddy simulation, Phys. Rev. Fluids 6 (2021) 104607–1–104607–35.
[17] D.A. Buchta, S.J. Laurence, T.A. Zaki, Assimilation of wall-pressure measurements in high-speed flow over a cone, J. Fluid Mech. 947 (2022) R2–1–R2–13.
[18] M. Wang, T.A. Zaki, State estimation in turbulent channel flow from limited observations, J. Fluid Mech. 917 (2021) A9–1–A9–29.
[19] T.A. Zaki, M. Wang, From limited observations to the state of turbulence: fundamental difficulties of flow reconstruction, Phys. Rev. Fluids 6 (2021) 100501–1–100501–20.
[20] P.J. Huber, Robust statistics: a review, Ann. Math. Stat. 43 (4) (1972) 1041–1067.
[21] P.J. Huber, Robust regression: asymptotics, conjectures and Monte Carlo, Ann. Stat. 1 (5) (1973) 799–821.
[22] K.P. Bube, R.T. Langan, Hybrid l^1/l^2 minimization with applications to tomography, Geophysics 62 (4) (1997) 1183–1195.
[23] C. Tavolato, L. Isaksen, On the use of a Huber norm for observation quality control in the ECMWF 4D-Var, Q. J. R. Meteorol. Soc. 141 (690) (2015) 1514–1527.
[24] O. Zadorozhnyi, G. Benecke, S. Mandt, T. Scheffer, M. Kloft, Huber-norm regularization for linear prediction models, in: P. Frasconi, N. Landwehr, G. Manco, J. Vreeken (Eds.), Machine Learning and Knowledge Discovery in Databases, ECML PKDD 2016, in: Lecture Notes in Computer Science, vol. 9851, Springer, Cham, 2016, pp. 714–730.
[25] V. Rao, A. Sandu, M. Ng, E.D. Nino-Ruiz, Robust data assimilation using L_1 and Huber norms, SIAM J. Sci. Comput. 39 (3) (2017) B548–B570.
[26] L. Franceschini, D. Sipp, O. Marquet, Mean-flow data assimilation based on minimal correction of turbulence models: application to turbulent high Reynolds number backward-facing step, Phys. Rev. Fluids 5 (2020) 094603–1–094603–23.
[27] H.W. Kuhn, A.W. Tucker, Nonlinear programming, in: Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, 1951, pp. 481–492.
[28] C.K. Wikle, L.M. Berliner, A Bayesian tutorial for data assimilation, Phys. D: Nonlinear Phenom. 230 (1–2) (2007) 1–16.
[29] A.C. Lorenc, Analysis methods for numerical weather prediction, Q. J. R. Meteorol. Soc. 112 (474) (1986) 1177–1194.
[30] I. Barrodale, F.D.K. Roberts, Algorithm 552: solution of the constrained ℓ_1 linear approximation problem, ACM Trans. Math. Softw. 6 (2) (1980) 231–235.

[31] J.A. Scales, A. Gersztenkorn, S. Treitel, L.R. Lines, Robust optimization methods in geophysical inverse theory, in: SEG Technical Program Expanded Abstracts, 1988, pp. 827–830.
[32] P.J. Huber, Robust estimation of a location parameter, Ann. Math. Stat. 1 (53) (1964) 73–101.
[33] C. Tavolato, L. Isaksen, Huber norm quality control in the IFS, Meteorol. Sect. ECMWF Newslett. 122 (2009/10) 27–31.
[34] K. Fountoulakis, J. Gondzio, A second-order method for strongly convex ℓ_1-regularization problems, Math. Program. 156 (2016) 189–219.
[35] K.P. Bube, T. Nemeth, Fast line searches for the robust solution of linear systems in the hybrid l^1/l^2 and Huber norms, Geophysics 72 (2) (2007) A13–A17.
[36] J.M. Lewis, S. Lakshmivarahan, S.K. Dhall, Dynamic Data Assimilation: A Least Squares Approach, vol. 104, Cambridge University Press, 2006.
[37] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, NY, 2006.
[38] C.M. Murea, The BFGS algorithm for a nonlinear least squares problem arising from blood flow in arteries, Comput. Math. Appl. 49 (2005) 171–186.
[39] A. Guitton, W.W. Symes, Robust inversion of seismic data using the Huber norm, Geophysics 68 (4) (2003) 1310–1319.
[40] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.
[41] N.K. Dhingra, M.R. Jovanović, Z.Q. Luo, An ADMM algorithm for optimal sensor and actuator selection, in: 53rd IEEE Conference on Decision and Control, Los Angeles, California, USA, December 15-17, 2014, vol. 9851, 2014, pp. 4039–4044.
[42] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2010) 1–122.
[43] N. Parikh, S. Boyd, Proximal algorithms, Found. Trends Optim. 1 (3) (2013) 123–231.
[44] D.P. Kingma, J.L. Ba, ADAM: a method for stochastic optimization, in: International Conference on Learning Representations (ICLR), San Diego, USA, May 7-9, 2015.
[45] F. Hecht, New development in FreeFem++, J. Numer. Math. 20 (3–4) (2012) 251–265, https://freefem.org/.
[46] R. Hartmann, Derivation of an adjoint consistent discontinuous Galerkin discretization of the compressible Euler equations, in: G. Lube, G. Rapin (Eds.), International Conference on Boundary and Interior Layers, BAIL 2006, Göttingen, Germany, 2006, pp. 1–8.
[47] S.G. Johnson, The nlopt nonlinear-optimization package, http://github.com/stevengj/nlopt.
[48] V. Mons, O. Marquet, Linear and nonlinear sensor placement strategies for mean-flow reconstruction via data assimilation, J. Fluid Mech. 923 (2021) A1–1–A1–39.
[49] A. Hervé, D. Sipp, P.J. Schmid, M. Samuelides, A physics-based approach to flow control using system identification, J. Fluid Mech. 702 (2012) 26–58.
[50] D. Sipp, P.J. Schmid, Linear closed-loop control of fluid instabilities and noise-induced perturbations: a review of approaches and tools, ASME Appl. Mech. Rev. 68 (2) (2016) 020801.
[51] V. Mons, O. Marquet, B. Leclaire, P. Cornic, F. Champagnat, Dense velocity, pressure and Eulerian acceleration fields from single-instant scattered velocities through Navier–Stokes-based data assimilation, Meas. Sci. Technol. 33 (2022) 124004.