Bayesian Inference and Data Assimilation

Prof. Dr.-Ing. Sebastian Reich

Universität Potsdam

12 April 2021

7 McKean approach to sequential data assimilation

We start by recalling that a Markov process gives rise to families of random variables

$$Z_{t_{0:N}} := (Z^0, Z^1, \ldots, Z^N) : \Omega \to \mathbb{R}^{N_z \times N+1},$$

with joint PDFs given by

$$\pi_{Z_{t_{0:N}}}(z_{t_{0:N}}) = \pi_Z(z^0, t_0)\pi(z^1|z^0)\cdots\pi(z^N|z^{N-1}),$$

where $\pi(z'|z)$ denotes the transition kernel of the Markov process and $\pi_Z(z, t_0)$ is the initial PDF. The marginal PDFs $\pi_Z(z, t_n)$ for the random variables Z^n, $n = 1, \ldots, N$, are recursively determined by the Chapman-Kolmogorov equation

$$\pi_Z(z', t_n) = \int_{\mathbb{R}^{N_z}} \pi(z'|z)\pi_Z(z, t_{n-1})\mathrm{d}z.$$

Ensemble prediction methods produce realisations (or samples) of $Z_{t_{0:N}}$, based on the recursive structure of the underlying Markov process. More precisely, given a set of M ensemble members z_i^{n-1} at t_{n-1}, this ensemble is subsequently transformed into an ensemble at t_n according to a (stochastic) model which we write as

$$z_i^n = Z_i^n(\omega_i), \qquad Z_i^n \sim \pi(\cdot|z_i^{n-1}),$$

for $i = 1, \ldots, M$, *i.e.*, z_i^n is taken as the realisation of a random variable with PDF $\pi(\cdot|z_i^{n-1})$.

After having completed this process for $n = 1, \ldots, N$ we may formally write

$$Z_{t_{0:N}}(\omega_i) := (z_i^0, z_i^1, \ldots, z_i^N),$$

for $i = 1, \ldots, M$.

Bayesian data assimilation algorithms combine ensemble prediction for Markov processes with a recursive application of Bayes' formula

$$\pi_{Z^a}(z|y_{\rm obs}) = \frac{\pi_Y(y_{\rm obs}|z)\pi_{Z^f}(z)}{\pi_Y(y_{\rm obs})}, \tag{1}$$

which transforms a forecast PDF $\pi_{Z^f}(z)$ into an analysis PDF $\pi_{Z^a}(z|y_{\rm obs})$ under the likelihood $\pi_Y(y_{\rm obs}|z)$ of an observation $y_{\rm obs}$.

One of the challenges of merging ensemble prediction methods and Bayesian data assimilation algorithms lies in the fact that Bayes' formula does not come in the form of a Chapman-Kolmogorov equation. The SIR particle filter works around that problem by implementing Bayes' formula as an importance resampling step in order to obtain an analysis ensemble with uniform weights.

The McKean approach[1] to data assimilation pushes importance resampling a step further by putting Bayes' formula directly into the framework of Markov chains.

The Markov chain defines the transition from a forecast ensemble into an analysis ensemble at an observation time t_k, with all ensemble members keeping equal weights. The desired Markov chain will only be applied once and has to produce an analysis which is consistent with Bayes' formula (1).

[1] McKean pioneered the study of stochastic processes which are generated by stochastic differential equations for which the diffusion term depends on the time-evolving marginal distributions $\pi_Z(z,t)$. Such stochastic differential equations lead to interacting particle/ensemble approximations. More precisely, Monte Carlo sampling paths $\{z_i(t)\}_{t\in[0,T]}$, $i = 1,\ldots,M$, are no longer independent if samples $\{z_i(t)\}$ at time t are used to approximate the marginal PDF $\pi_Z(z,t)$ by, *e.g.*, the ensemble induced empirical measure. Here we utilise a generalisation of this idea which allows for Markov transition kernels $\pi(z', t_n|z)$ that depend on the marginal distribution $\pi_Z(z,t_n)$.

Definition (McKean approach)

Let $b(z'|z)$ be a transition kernel such that Bayes' formula (1) becomes equivalent to a Chapman-Kolmogorov equation,

$$\pi_{Z^a}(z^a|y_{\rm obs}) = \int_{\mathbb{R}^{N_z}} b(z^a|z^f)\pi_{Z^f}(z^f){\rm d}z^f.$$

Given an ensemble z_i^f which follows the forecast PDF, an analysis ensemble can be obtained from

$$z_i^a = Z_i^a(\omega), \qquad Z_i^a \sim b(\cdot|z_i^f),$$

$i = 1,\ldots,M$, *i.e.*, the analysis samples z_i^a are realisations of random variables Z_i^a with PDFs $b(z|z_i^f)$. This Monte Carlo approach to Bayesian inference is called the McKean approach.

Remark

The transition kernel $b(z'|z)$ will, in general, depend on both the observed value $y_{\rm obs}$ as well as the forecast PDF π_{Z^f}. In this sense, the induced Markov process will be non-autonomous.

Example (Resampling)

Consider the following transition kernel:

$$b(z^a|z^f) = \varepsilon\pi_Y(y_{\rm obs}|z^f)\delta(z^a - z^f) + (1 - \varepsilon\pi_Y(y_{\rm obs}|z^f))\pi_{Z^a}(z^a|y_{\rm obs}).$$

Here $\varepsilon \geq 0$ is chosen such that

$$1 - \varepsilon\pi_Y(y_{\rm obs}|z) \geq 0,$$

for all $z \in \mathbb{R}^{N_z}$. Indeed we find that

$$\begin{aligned}\int_{\mathbb{R}^{N_z}} b(z^a|z^f)\pi_{Z^f}(z^f){\rm d}z^f &= \varepsilon\pi_Y(y_{\rm obs}|z^a)\pi_{Z^f}(z^a) + \pi_{Z^a}(z^a|y_{\rm obs}) - \\ &\qquad \varepsilon\pi_{Z^a}(z^a|y_{\rm obs}) \int_{\mathbb{R}^{N_z}} \pi_Y(y_{\rm obs}|z^f)\pi_{Z^f}(z^f){\rm d}z^f, \\ &= \pi_{Z^a}(z^a|y_{\rm obs}) + \\ &\qquad \varepsilon\left\{\pi_Y(y_{\rm obs}|z^a)\pi_{Z^f}(z^a) - \pi_{Z^a}(z^a|y_{\rm obs})\pi_Y(y_{\rm obs})\right\} \\ &= \pi_{Z^a}(z^a|y_{\rm obs}).\end{aligned}$$

In order to apply the McKean approach to data assimilation, we need to find an appropriate transition kernel $b(z^a|z^f)$. There are many possibilities for doing so. The kernels considered in this lecture are based on the concept of coupling two probability measures.

Definition (Coupling for Bayesian inference)

A Bayes' formula coupling between the forecast and the analysis PDFs is a joint measure $\mu_{Z^f Z^a}$ such that the marginals satisfy

$$\pi_{Z^f}(z^f)\mathrm{d}z^f = \int_{z^a \in \mathbb{R}^{N_z}} \mu_{Z^f Z^a}(\mathrm{d}z^f, \mathrm{d}z^a),$$

and

$$\pi_{Z^a}(z^a|y_{\mathrm{obs}})\mathrm{d}z^a = \int_{z^f \in \mathbb{R}^{N_z}} \mu_{Z^f Z^a}(\mathrm{d}z^f, \mathrm{d}z^a),$$

respectively.

Once such a coupling is available we obtain the desired transition PDF $b(z^a|z^f)$ from the disintegration formula

$$b(z^a|z^f)\,\pi_{Z^f}(z^f)\,\mathrm{d}z^f\mathrm{d}z^a = \mu_{Z^f Z^a}(\mathrm{d}z^f, \mathrm{d}z^a).$$

Such couplings can often be made deterministic and are then given in the form of a transport map, *i.e.*, there exists a transformation $z^a = T(z^f)$ such that

$$\mu_{Z^f Z^a}(\mathrm{d}z^f, \mathrm{d}z^a) = \delta(z^a - T(z^f))\, \pi_{Z^f}(z^f)\, \mathrm{d}z^f \mathrm{d}z^a,$$

in which case we have

$$b(z^a|z^f) = \delta(z^a - T(z^f)),$$

formally.

A Monte Carlo implementation of the McKean approach to Bayesian inference (online learning) then reduces to the simple mathematical formula

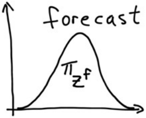
$$z_i^a = T(z_i^f), \qquad i = 1, \ldots, M. \tag{2}$$

In addition to finding appropriate couplings there is another major challenge for implementing the McKean approach in practice.

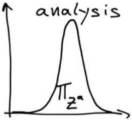
This challenge is related to the fact that the forecast PDF is often not available explicitly. Indeed, an ensemble prediction method will only provide us with a forecast ensemble z_i^f, $i = 1, \ldots, M$, from which we will have to estimate a statistical model in the form of a forecast PDF (or measure) π_{Z^f}.

See the following figure for a graphical presentation of the complete ensemble-based implementation of a McKean approach for Bayesian inference.

PDFs



$$\xrightarrow[\pi_{Z^a} \propto \pi_Y(y|\cdot)\,\pi_{Z^f}]{\text{Bayes}}$$



RVs

$$Z^f \qquad \text{law}(Z^f) = \pi_{Z^f}$$

$$\xrightarrow[\pi_{Z^f Z^a}]{\text{coupling}}$$

$$Z^a \qquad \text{law}(Z^a) = \pi_{Z^a}$$

MC

$$z_i^f = Z^f(\omega_i), \quad i = 1, \ldots, M$$

$$\xrightarrow[\text{(Markov process)}]{\text{McKean}}$$

$$z_i^a \sim \frac{\pi_{Z^f Z^a}(z_i^f, \cdot)}{\pi_{Z^f}(z_i^f)}, \quad i = 1, \ldots, M$$

A statistical model for the forecast PDF π_{Z^f} can be either parametric, semiparametric, or non-parametric.

A simple parametric model is the Gaussian $\mathrm{N}(\bar{z}_M^f, P_M^f)$ with its mean $\bar{z}_M^f$ and covariance matrix P_M^f estimated from the prior or forecast ensemble z_i^f, $i = 1, \ldots, M$, *i.e.*,

$$\bar{z}_M^f = \frac{1}{M} \sum_{i=1}^{M} z_i^f \tag{3}$$

and

$$P_M^f = \frac{1}{M-1} \sum_{i=1}^{M} (z_i^f - \bar{z}_M^f)(z_i^f - \bar{z}_M^f)^T \in \mathbb{R}^{N_z \times N_z}. \tag{4}$$

If the likelihood function $\pi_Y(y_{\mathrm{obs}}|z)$ is also Gaussian, then the posterior is Gaussian as well with its mean and covariance matrix given by Kalman's formulas.

However, unlike the classical Kalman filter, we need to establish a coupling between the prior and posterior Gaussians in order to derive an appropriate map (2) for ensemble-based data assimilation algorithms.

We will see later in this chapter that the associated data assimilation algorithms give rise to the popular family of ensemble Kalman filters.

On the other hand, a non-parametric model for the forecast measure is provided by the empirical measure

$$\mu_M^f(\mathrm{d}z) = \frac{1}{M}\sum_{i=1}^{M} \delta(z - z_i^f)\mathrm{d}z. \tag{5}$$

The associated analysis measure is then

$$\mu_M^a(\mathrm{d}z) = \sum_{i=1}^{M} w_i \delta(z - z_i^f)\mathrm{d}z \tag{6}$$

with weights $w_i \propto \pi_Y(y_{\mathrm{obs}}|z_i^f)$.

The problem of finding couplings for such measures has been discussed in Chapter 2 and can be used to construct alternative sequential Monte Carlo methods. See Chapter 7 in the textbook for more details.

One of the key findings of this chapter will be that Monte Carlo implementations of the McKean approach lead to linear ensemble transformations of the form

$$z_j^a = \sum_{i=1}^{M} z_i^f d_{ij}, \tag{7}$$

regardless of the chosen prior approximation.

Data assimilation algorithms from this chapter will therefore only differ through their coefficients d_{ij} in (7), which have some (possibly random) functional dependence on the forecast ensemble. These coefficients can be collected in an $M \times M$ matrix D.

Definition (Linear ensemble transform filters)

A particle or ensemble filter with the Bayesian inference step implemented in the form of a linear ensemble transformation (7) is called a linear ensemble transform filter (LETF).

7.1 Ensemble Kalman filters

The classical Kalman filter formulas provide a transition from the forecast mean $\bar{z}^f$ and covariance matrix P^f to the analysis mean $\bar{z}^a$ and covariance matrix P^a.

An ensemble Kalman filter (EnKF) pulls this transition back onto the level of forecast ensembles z_i^f, $i = 1, \ldots, M$, and analysis ensembles z_i^a, $i = 1, \ldots, M$, respectively. Therefore it provides an implicit coupling between the underlying forecast and analysis random variables.

Definition (EnKF update)

(i) Compute the empirical forecast mean (3) and covariance matrix (4).

(ii) The Kalman update formulas result in the analysis mean

$$\bar{z}^a = \bar{z}^f_M - K(H\bar{z}^f_M - y_{\rm obs}), \tag{8}$$

and the analysis covariance matrix

$$P^a = P^f_M - KHP^f_M. \tag{9}$$

with Kalman gain matrix given by

$$K = P^f_M H^T (HP^f_M H^T + R)^{-1}, \tag{10}$$

(iii) The analysis ensemble $\{z_i^a\}$ is defined by a linear transformation (7) with the transformation matrix D chosen such that

$$\frac{1}{M}\sum_{i=1}^M z_i^a = \bar{z}^a, \quad \text{and} \quad \frac{1}{M-1}\sum_{i=1}^M (z_i^a - \bar{z}^a)(z_i^a - \bar{z}^a)^T = P^a. \tag{11}$$

Remark (Transformation matrix)

The transformation matrix D of an EnKF is not uniquely determined by the constraints (11); this allows for different implementations of an EnKF.

We start with the oldest EnKF formulation, which is based on a non-deterministic coupling between the forecast and analysis ensemble.

In this case, D is a random matrix and (11) is to hold in expectation.

Definition (EnKF with perturbed observations)

Given a prior ensemble z_i^f, $i = 1, \ldots, M$, with associated empirical mean (3) and covariance matrix (4), we first compute the Kalman gain matrix (10). The *EnKF with perturbed observations* is then based on the coupling

$$Z^a = Z^f - K(HZ^f + \Xi - y_{\rm obs}), \tag{12}$$

where Ξ is a Gaussian random variable with mean zero and covariance matrix R.

The posterior ensemble $\{z_i^a\}$ is obtained accordingly via

$$z_i^a = z_i^f - K(Hz_i^f + \xi_i - y_{\rm obs}), \qquad i = 1, \ldots, M,$$

where the variables $\{\xi_i\}$ are realisations of M i.i.d. Gaussian random variables with PDF $\mathrm{N}(0, R)$.

Remark

The McKean transition kernel $b(z^a|z^f)$ of the EnKF is given by

$$b(z^a|z^f) = \mathrm{n}(z^a; z^f - K(Hz^f - y_{\rm obs}), KRK^T).$$

Lemma (Consistency of EnKF update step)

The EnKF with perturbed observations is consistent with the standard Kalman update step if the prior distribution is Gaussian.

Proof.

We demonstrate that (12) is consistent with the Kalman update formulas provided Z^f has mean $\bar{z}^f$ and covariance matrix P^f_M. We first verify that

$$\bar{z}^a = \mathbb{E}[Z^a] = \bar{z}^f_M - K(H\bar{z}^f_M - y_{\rm obs}),$$

since Ξ is centred. Next, we find that

$$\begin{aligned} P^a &= \mathbb{E}[(Z^a - \bar{z}^a)(Z^a - \bar{z}^a)^T] \\ &= \mathbb{E}[(Z^f - K(HZ^f + \Xi - y_{\rm obs}) - \bar{z}^a)(Z^f - K(HZ^f + \Xi - y_{\rm obs}) - \bar{z}^a)^T], \\ &= KRK^T + P^f_M - P^f_M H^T K^T - KHP^f_M + KHP^f_M H^T K^T = P^f_M - KHP^f_M. \end{aligned}$$

Here we have used the property that Ξ and Z^f are independent, and the identity

$$KRK^T + KHP^f_M H^T K^T = P^f_M H^T (HP^f_M H^T + R)^{-1} HP^f_M = P^f_M H^T K^T.$$

□

Remark (EnKF as linear ensemble transform filter)

The EnKF with perturbed observations fits into the framework of LETFs (7) and the associated matrix-valued random variable $D : \Omega \to \mathbb{R}^{M\times M}$ *has realisations with entries*

$$d_{ij} = \delta_{ij} - \frac{1}{M-1}(z_i^f - \bar{z}_M^f)^T H^T (HP_M^f H^T + R)^{-1}(Hz_j^f + \xi_j - y_{\mathrm{obs}}), \qquad (13)$$

where δ_{ij} *denotes the Knonecker delta, i.e.,* $\delta_{ii} = 1$, $\delta_{ij} = 0$ *for* $i \neq j$.

The claim can be verified by direct calculations making use of

$$P_M^f = \frac{1}{M-1}\sum_{i=1}^{M} z_i^f (z_i^f - \bar{z}_M^f)^T.$$

Example (Analysis of EnKF in terms of implied coupling)

The EnKF with perturbed observations for a scalar state variable and forward operator $H = 1$ becomes

$$Z^a = Z^f - K(Z^f - y_{\rm obs} + \Xi), \tag{14}$$

with Kalman gain factor

$$K = \frac{(\sigma^f)^2}{(\sigma^f)^2 + R},$$

where we have replaced the prior covariance matrix P_M^f by $(\sigma^f)^2$ for notational convenience. Let us interpret this update in terms of an induced coupling between Gaussian Z^f and Z^a subject to

$$\bar{z}^a = \bar{z}^f - \frac{(\sigma^f)^2}{(\sigma^f)^2 + R}(\bar{z}^f - y_{\rm obs}),$$

and variance

$$(\sigma^a)^2 = \left(1 - \frac{(\sigma^f)^2}{(\sigma^f)^2 + R}\right)(\sigma^f)^2 = \frac{R}{(\sigma^f)^2 + R}(\sigma^f)^2.$$

Example (continued)

Any Gaussian coupling must have mean

$$(\bar{z}^f, \bar{z}^a) \in \mathbb{R}^2,$$

and covariance matrix

$$\Sigma = \begin{pmatrix} (\sigma^f)^2 & \rho\sigma^f\sigma^a \\ \rho\sigma^f\sigma^a & (\sigma^a)^2 \end{pmatrix},$$

with the correlation ρ taking values between minus one and one. In the case of the EnKF with perturbed observations, the induced coupling the correlation between Z^f and Z^a is given by

$$\begin{aligned}\mathbb{E}[(Z^f - \bar{z}^f)(Z^a - \bar{z}^a)] &= \mathbb{E}[(Z^f - \bar{z}^f)(Z^f - \bar{z}^f - K(Z^f - \bar{z}^f + \Xi))], \\ &= (\sigma^f)^2 - K(\sigma^f)^2 = (\sigma^a)^2,\end{aligned}$$

and, therefore,

$$\Sigma = \begin{pmatrix} (\sigma^f)^2 & (\sigma^a)^2 \\ (\sigma^a)^2 & (\sigma^a)^2 \end{pmatrix} \quad \text{and} \quad \rho = \frac{\sigma^a}{\sigma^f} = \sqrt{\frac{R}{(\sigma^f)^2 + R}} < 1.$$

Definition (EnKF with perturbed observations)

We draw M independent realisations $z_i^0 = z_i(0)$ from a given initial distribution $\pi_Z(z,0)$. Observations are given in intervals of $\Delta t_{\text{out}} = \delta t N_{\text{out}}$.
The following steps are performed recursively for $k = 1,\ldots,N_{\text{obs}}$:

(i) In between observations, the initial ensemble $\{z_i(\mathfrak{t}_{k-1})\}$ at $\mathfrak{t}_{k-1} = (k-1)\Delta t_{\text{out}}$ is propagated under the given evolution model in order to produce a forecast ensemble $\{z_i^f\}$ at the next observation time $\mathfrak{t}_k = k\Delta t_{\text{out}}$.

(ii) The forecast ensemble is transformed into an analysis ensemble $\{z_i^a\}_{i=1}^M$ according to (7). The coefficients d_{ij} are defined by (13) with observed $y_{\text{obs}} = y_{\text{obs}}(\mathfrak{t}_k)$ and M independent realisations ξ_i of a Gaussian random variable with distribution $\mathrm{N}(0,R)$. Finally, we return to (i) with new initial conditions

$$z_i(\mathfrak{t}_k) := z_i^a, \qquad i = 1,\ldots,M,$$

and the index k increased by one.

Remark (Deterministic couplings)

Given the theoretical results on couplings, it is rather natural to look for deterministic couplings between the forecast and the analysis. Indeed, a suitable analysis ensemble could, for example, be defined by the linear transformation

$$z_i^a = \bar{z}^a + (P^a)^{1/2}(P_M^f)^{-1/2}(z_i^f - \bar{z}_M^f), \qquad i = 1, \ldots, M, \tag{15}$$

with $\bar{z}^a$ *given by (8) and* P^a *by (9), respectively.*

This formulation requires the computation of the square root of two $N_z \times N_z$ *matrices, which can be computationally demanding if* N_z *is large.*

In order to derive an alternative update formula, we write the empirical forecast covariance matrix (4) as

$$P_M^f = \frac{1}{M-1} A^f (A^f)^T, \tag{16}$$

with the $N_z \times M$ matrix of forecast *ensemble perturbations* (also called *ensemble anomalies*)

$$A^f := \left[\; (z_1^f - \bar{z}_M^f) \quad (z_2^f - \bar{z}_M^f) \quad \cdots \quad (z_M^f - \bar{z}_M^f) \;\right]. \tag{17}$$

We next seek a matrix $S \in \mathbb{R}^{M\times M}$ such that

$$\frac{1}{M-1} A^f S S^T (A^f)^T = P^a = P_M^f - KHP_M^f. \tag{18}$$

A set of analysis ensemble anomalies is then provided by

$$A^a := A^f S \in \mathbb{R}^{N_z \times M}. \tag{19}$$

More specifically, Eq. (18) together with (16) imply

$$P^a = \frac{1}{M-1} A^f \left\{ I - \frac{1}{M-1} (HA^f)^T \left[HP^f_M H^T + R \right]^{-1} HA^f \right\} (A^f)^T,$$

and hence

$$S := \left\{ I - \frac{1}{M-1} (HA^f)^T \left[HP^f_M H^T + R \right]^{-1} HA^f \right\}^{1/2}.$$

Recall that the square root of a positive semi-definite matrix U is the unique symmetric matrix $U^{1/2}$ such that $U^{1/2} U^{1/2} = U$.

An application of the Sherman-Morrison-Woodbury formula leads to the equivalent formulation

$$S = \left\{ I + \frac{1}{M-1}(HA^f)^T R^{-1} HA^f \right\}^{-1/2}. \tag{20}$$

Formulation (20) is preferable computationally whenever the ensemble size M is smaller than the number of observations N_y, and R is diagonal. Furthermore, the equivalent Kalman update formula

$$\begin{aligned} \bar{z}^a &= \bar{z}^f_M - P^a H^T R^{-1}(H\bar{z}^f_M - y_{\mathrm{obs}}), \\ &= \bar{z}^f_M - \frac{1}{M-1} A^f S^2 (A^f)^T H^T R^{-1}(H\bar{z}^f_M - y_{\mathrm{obs}}), \end{aligned}$$

can be used to bring (8) into the form

$$\bar{z}^a = \sum_{i=1}^M z_i^f w_i, \tag{21}$$

with the weights w_i defined as the ith entry of the column vector

$$w = \frac{1}{M}\mathit{1} - \frac{1}{M-1} S^2 (A^f)^T H^T R^{-1} \left(H\bar{z}^f_M - y_{\mathrm{obs}} \right). \tag{22}$$

Here $\mathit{1} := (1, 1, \ldots, 1)^T \in \mathbb{R}^M$.

Since $S\mathit{1} = \mathit{1}$ and $A^f\mathit{1} = 0$ it holds that

$$A^a\mathit{1} = A^f S\mathit{1} = 0, \tag{23}$$

and the weights w_i satisfy

$$\sum_{i=1}^{M} w_i = 1.$$

The weights can therefore be interpreted as "importance" weights similar to the weights in a SIS or SIR particle filter.

A complete ensemble update is given by

$$z_j^a = \sum_{i=1}^{M} w_i z_i^f + \sum_{i=1}^{M} (z_i^f - \bar{z}_M^f) s_{ij} = \sum_{i=1}^{M} z_i^f \left\{ w_i + s_{ij} - \frac{1}{M} \right\}, \tag{24}$$

where $s_{ij} = (S)_{ij}$ denotes the (i,j)-th entry of the transform matrix S which satisifies $\sum_{i=1}^{M} s_{ij} = 1$. This formula forms the basis of the popular *ensemble square root filter* (ESRF). We can bring (24) into the form (7) with coefficients

$$d_{ij} = w_i - \frac{1}{M} + s_{ij}.$$

Remark (Optimal deterministic couplings)

We have already discussed that not all deterministic couplings are optimal in the sense of the Monge-Kantorovitch transportation problem. The same applies to the ESRF update (24). An optimal update in the sense of Monge-Kantorovitch is

$$z_i^a = \bar{z}^a + \frac{1}{\sqrt{M-1}} A^a \left[(A^a)^T P_M^f A^a \right]^{-1/2} (A^a)^T (z_i^f - \bar{z}_M^f), \tag{25}$$

where A^a is defined by (19).

Note that (25) requires an additional computation of an $M \times M$ matrix square root.

Definition (ensemble square root filter (ESRF))

Given an initial distribution $\pi_Z(z,0)$, we draw M independent realisations $z_i^0 = z_i(0)$ from this distribution. Observations are given in intervals of $\Delta t_{\mathrm{out}} = \delta t N_{\mathrm{out}}$.

The following steps are performed recursively for $k = 1, \dots, N_{\mathrm{obs}}$:

(i) In between observations, the initial ensemble $\{z_i(\mathfrak{t}_{k-1})\}$ at $\mathfrak{t}_{k-1} = (k-1)\Delta t_{\mathrm{out}}$ is propagated under the given evolution model in order to produce a forecast ensemble $\{z_i^f\}$ at the next observation time $\mathfrak{t}_k = k\Delta t_{\mathrm{out}}$.

(ii) A Gaussian $\mathrm{N}(\bar{x}_M^f, P_M^f)$ is fit to the forecast ensemble z_i^f, $i = 1, \dots, M$. The observation $y_{\mathrm{obs}} = y_{\mathrm{obs}}(\mathfrak{t}_k)$ leads to a posterior Gaussian $\mathrm{N}(\bar{z}^a, P^a)$. Using either (24) or (25) together with either (8) or (21) we define a posterior ensemble z_i^a, $i = 1, \dots, M$. Finally, we return to (i) with new initial conditions

$$z_i(\mathfrak{t}_k) := z_i^a, \qquad i = 1, \dots, M,$$

and the index k increased by one.

Example (Lorenz-63)

We return to the stochastically perturbed Lorenz-63 model from Section 4. The EnKF with perturbed observations and the ESRF are implemented for the data assimilation setting of Section 6. Note the the EnKF/ESRF lead to smaller or comparable errors to the SIR implementation.