RESEARCH

Open Access

Optimal measurement budget allocation for Kalman prediction over a finite time horizon by genetic algorithms

Antoine Aspeel[*] , Axel Legay, Raphaël M. Jungers and Benoit Macq

*Correspondence:
antoine.aspeel@uclouvain.be
ICTEAM Institute, UCLouvain
University,
Ottignies-Louvain-la-Neuve,
Belgium

Abstract

In this paper, we address the problem of optimal measurement budget allocation to estimate the state of a linear discrete-time dynamical system over a finite horizon. More precisely, our aim is to select the measurement times in order to minimize the variance of the estimation error over a finite horizon. In addition, we investigate the closely related problem of finding a trade-off between number of measurements and signal to noise ratio.

First, the optimal measurement budget allocation problem is reduced to a deterministic combinatorial program. Then, we propose a genetic algorithm implementing a count preserving crossover to solve it. On the theoretical side, we provide a one-dimensional analysis that indicates that the benefit of using irregular measurements grows when the system is unstable or when the process noise becomes important. Then, using the duality between estimation and control, we show that the problem of selecting optimal control times for a linear quadratic regulator can be reduced to our initial problem. Finally, numerical implementations demonstrate that using measurement times optimized by our genetic algorithm gives better estimate than regularly spaced measurements. Our method is applied to a discrete version of a continuous-time system and the impact of the discretization time step is studied. It reveals good convergence properties, showing that our method is well suited to both continuous-time and discrete-time setups.

Keywords: Kalman filtering, Optimal sampling, Genetic algorithms, Budget allocation

1 Introduction

Kalman filtering [1] is an algorithm that provides state estimation of a stochastic linear system over time from noisy measurements. It has been applied for many estimation or prediction problems [2–4]. For example, a Kalman filter is used in [5] to estimate the position of a camera when the noise statistics are unknown. They propose an adaptive scheme (called adaptative non-linear Kalman filtering) to estimate the noise statistics via sampling the noises. They progressively adjust the filter parameters thanks to these samples. The estimation of the parameters improves with the number of samples. On the other hand, the computational complexity increases with the number of samples. The Kalman

filter is also used for compressed sensing. In [6], the authors use an extended linearized Kalman filter and the Steffensen's acceleration method to optimally reconstruct sparse data from noisy measurements. They show that their Kalman-based approach gives similar or better results than traditional methods for ℓ_1 minimization such as the primal-dual algorithm or the Orthogonal Matching Pursuit method. The interest of their method is particularly important in the case of high dimensional signals. However, this method is restricted to the case where the signal is sufficiently sparse (the sparsity must be less than half the dimension of the measurements). In most of these applications, measurements are acquired in a regular way, i.e., measurements are equally spaced in time.

Several authors have studied the optimization of Kalman prediction in the presence of a variable rate of measurements or in the presence of unknown statistical properties of the measurement noise. For example, [7] have proposed solutions for the case of irregular sampling rate based on a track to track Kalman filter fusion method. They combine the information from two sensors, the first one is fast rate, regular, delay free but less accurate, and the second one is slow rate, irregular, delayed but more accurate. Two Kalman filters are used to estimate the states based on each type of measurement. Then, the estimates are fused. They show on simulations and on a laboratory experiment that the fused estimation is more precise than the individual ones. In these previous works, variable rate and unknown noise statistics were considered as constraints to be solved while optimizing Kalman prediction.

In this paper, on the contrary, we aim to optimally choice the measurement sampling times and their noise level to minimize a measurement budget for operating an optimal Kalman prediction. Sampling times and measurement noise levels are, for our case, parameters to be optimized.

Reducing the number of measurements often has significant advantages. The first one is to reduce the energy consumption that is linked to measurement acquisition. This is essential for mobile applications. A second advantage could be economic if, for example, each measurement acquisition is expensive. A third advantage is linked to acquisition safety issues. That is the reason for relying on Kalman prediction to manage the radiotherapy of mobile tumors, in which the goal is to target a tumor with ionizing radiation. In the case of lung tumors, the target is constantly moving due to the patient's breathing. One option is to track the tumor with imaging X-rays. Unfortunately, with this option each image acquisition also irradiates the patient, including health tissues [8]. The total number of X-ray acquisitions used for tumor tracking must thus be kept below a cutoff to prevent secondary cancer induction. In most cases, it is possible to optimize the measurement times in order to maximize healing while staying below the maximal total irradiation dose.

The claim of this paper is that when the number of measurements is restricted, Kalman prediction will provide better results if the measurements can be selected at the optimal times instead of a regular sampling. This paper addresses the problem of selecting the optimal measurement times to predict the state of a stochastic linear dynamical system from noisy measurements, over a finite time horizon, when the number of measurements is fixed. The optimal measurement times are obtained by minimizing the mean prediction error variance over the complete horizon. This paper focuses on prediction instead of filtering and addresses real-time applications. Indeed, in our targeted applications, prediction allows real-time tracking by compensating for the measurement delay.

However, the proposed method can be adapted straightforwardly for filtering or smoothing applications.

The optimal measurement times can be formulated as the solution of a combinatorial optimization problem. Evolutionary algorithms are a potential effective approach to solve that kind of problem [9]. We propose different variants of genetic algorithms (GA) to solve this problem and demonstrate their efficiency compared to a random trial approach. GA provides an effective approach because they are able to sample a population broadly.

1.1 Related works

There is a quite large body of work in the literature, focusing on measurements failures arriving randomly according to a Bernouilli distribution [10–15]. In other papers, the missing measurements can only arrive according to certain patterns, and the objective is to design an estimator that is robust to all these patterns [16]. Under these assumptions, different problems have been studied. For example, [11] is interested in estimating the state of a multi-rate multi-sensors system with random missing measurements. They assume that they do not know when a dropout has occurred (when there is one dropout, the received measurement is pure noise). Consequently, they propose a method to detect dropouts. In [17], a system with multiple sensors is subject to denial-of-service (DoS) and false data injection (deception) attacks. The proposed method first detects the attack and isolates the concerned subsystems, then identifies the kind of attack, and uses a resilient observer subsystem.

In [12], a distributed Kalman filter is used in the case of large-scale power systems with random missing measurements. Kalman filtering with random missing measurements has also been studied when the noise variances are only approximately known [13, 14]. Using a Lyapunov-based approach, they provide guarantees on the estimation variance.

In the above cited papers, measurement times (or equivalently, the dropout times) are *not* design variables: they are suffered and not chosen. On the contrary, sensor scheduling problems consist in designing the behavior of the sensors. Typically, a set of sensors is available but only a few of them can be used at the same time. This problem has been studied in the case of an infinite horizon in both the discrete [18–23] and the continuous-time settings [24]. Most of the time, the objective is to minimize the average of the variance of the estimation error. In the continuous-time case, Ha et al. show in [24] that the infinite horizon optimal cost and optimal scheduling are independent of the initial error covariance. In addition, they prove that a periodic scheduling can approximate arbitrarily the cost of an optimal scheduling. Similar results are obtained in the discrete-time case in [20, 22]. Some works consider a cost associated to the use of the sensors [19, 25]. Other works dealing with periodic scheduling associate a usage budget per period to each sensor [21, 23].

The finite horizon problem has also received attention in both discrete [25, 26] and continuous-time [27–29] settings. In 1972, Athans [27] studied the best prediction at the end of an horizon (and not on average over the horizon). In their framework, different sensors are available and a cost is associated with each of them. The goal is to select sensors (one at a time) in order to minimize a trade-off between the sensors' cost and the prediction error. In [28], Lee et al. consider that different sensors are available for measurements and the user has to choose only a few of them at each time. Optionally, each sensor switch can be associated with a cost. In the work described in Woon et al.'s

[29] published in 2010, different sensors are available but only one can be chosen at the same time. There is no budget constraint on the use of each measurement process. In 2012, Vitus et al. [26] proposed an efficient algorithm to choose a sensor among a given set at each discrete time step. No measurement cost or budget constraint on the number of uses of each sensor is considered.

Due to the diverse potential applications mentioned above, several contributions related to the problem of optimal selection of the measurement times have appeared in different fields of the engineering literature. In 1970, Sano et al. [30] proposed a solution to the problem of selecting optimal measurement times in the continuous-time one-dimensional case. They proposed an explicit formula when there was only one measurement and proposed a numerical method when the number of measurements was greater than one. Still in the one-dimensional case, Tanaka et al. [31] selected the measurement times to minimize the maximum over time of the error variance. They provided an explicit formula for the measurement times. More recently, Aksenov et al. studied the Brownian motion over similar assumptions [32, 33].

1.2 Contributions
Despite the attention paid to these related problems for decades, the problem of selecting a given number of measurement times to minimize the average over a finite horizon of the estimation error variance is, to the best of our knowledge, not resolved in the multivariate case. This paper covers that topic. To summarize, the contributions of this paper are the following: (i) three different GAs are proposed and compared to efficiently solve the combinatorial optimization problem; (ii) an analysis of the impact of the model (stability, process noise and measurement noise variances) on the obtained solutions in the one-dimensional case is proposed and interpreted; (iii) links are illustrated between the solutions of the problem and the rank of the observability matrix of the system; (iv) the discretization of a continuous-time system is considered and numerical experiments show that our method is well suited to both discrete-time and continuous-time frameworks; (v) a related problem concerning the trade-off between quality and quantity of measurements is mathematically formalized and illustrated by an example; and (vi) the optimal intermittent linear quadratic regulator (LQR) problem is handled through the duality between estimation and control.

In the conference proceedings [34], we proposed an initial method for optimal intermittent prediction. However, contributions (ii)–(vi) are completely new. In addition, we propose here two new GAs and one of them significantly outperforms the previous one. Finally, a more general definition of the model is used: measurements are not directly related to the quantity to be estimated.

1.3 Paper outline
The rest of the paper is organized as follows: Section 2 formalizes the problem mathematically (Sections 2.1 to 2.3) and presents different algorithms to solve it (Section 2.4). Then, a continuous-time version of the problem is presented (Section 2.5). The problem of optimal intermittent LQR is addressed and we show that it can be solved by reduction to the problem of the optimal intermittent Kalman predictor (Section 2.6). Finally, a related problem concerning the compromise between the quality and the quantity of measurements is presented (Section 2.7). Section 3 compares the different proposed

algorithms (Section 3.1); numerical examples are extensively studied (Section 3.2); links with the observability matrix are illustrated (Section 3.3); a continuous-time example is presented with a particular focus on the impact of the discretization time step (Section 3.4); then, the impact of the system characteristics (stability, variances of the process noise and the measurement noise) on the obtained solution is illustrated (Section 3.5); and finally, the compromise between quality and quantity is illustrated by an example (Section 3.6).

Finally, Section 4 presents our conclusions and proposes further avenues of work.

2 Materials and methods

2.1 Problem description

We consider the discrete-time framework: $t = 0, \ldots, T$. The set of measurement times is denoted $\mathcal{M} \subset \{0, \ldots, T-1\}$ and is constrained to contain only N measurements, i.e., $|\mathcal{M}| = N$ with $N \leq T$. The evolution dynamic, quantity to estimate and measurements are described by the following equations,

$$x(t+1) = Ax(t) + b + Gw(t) \qquad t = 0, \ldots, T-1, \tag{1}$$

$$y(t) = Bx(t) \qquad\qquad t = 0, \ldots, T, \tag{2}$$

$$z(t) = Cx(t) + d + v(t) \; t \in \mathcal{M}, \tag{3}$$

$$x(0) \sim \mathcal{N}(\bar{x}_0, \bar{P}_0), \tag{4}$$

where $w(t)$ and $v(t)$ are independent zero-mean white Gaussian noises with covariances $\mathbb{E}[w(t_1)w(t_2)^\top] = Q\delta(t_1 - t_2)$ and $\mathbb{E}[v(t_1)v(t_2)^\top] = R\delta(t_1 - t_2)$, and $\delta(\cdot)$ is Kronecker's delta. Column vectors $x(t)$, $y(t)$, $z(t)$ and $w(t)$ have sizes m, n, p and q respectively. The dimensions of other quantities are deduced from compatibility. Matrices and vectors A, b, G, B, C, d, $\bar{x}_0$, $\bar{P}_0$, Q and R are known and could be time-dependent. Equation (1) describes the dynamic of internal state x, Eq. (2) gives the quantity to estimate y and Eq. (3) expresses the measurement z. Note that the last holds only for $t \in \mathcal{M}$, where $\mathcal{M}$ is the set of measurement times. Relation (4) indicates that the initial state follows a known normal distribution with mean $\bar{x}_0$ and covariance matrix $\bar{P}_0$.

Let us introduce the best a priori mean squared estimator of $y(t)$ according to $\mathcal{M}$, namely, $\hat{y}_{\mathcal{M}}(t|t-1) := \mathbb{E}[y(t)|z(\tau) : \tau \in \mathcal{M}, \tau < t]$. The set $\mathcal{M}$ is chosen to minimize the variance of the prediction error norm averaged for each t from 1 to T, it is

$$\min_{\mathcal{M} \subset \{0, \cdots, T-1\}} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\|y(t) - \hat{y}_{\mathcal{M}}(t|t-1)\|^2\right] \text{ subject to } |\mathcal{M}| = N, \tag{5}$$

where $\| \cdot \|$ is the Euclidean norm.

2.2 Intermittent Kalman predictor

In this section, we show that problem (5) can be reformulated in a more explicit way, thanks to the Kalman formalism. The latter provides explicit formula to compute, for a given measurement set $\mathcal{M}$, the best mean squared estimator $\hat{x}_{\mathcal{M}}(t|t-1)$ of $x(t)$ from which one can compute $\hat{y}_{\mathcal{M}}(t|t-1)$.

Let us introduce $\hat{x}_{\mathcal{M}}(t|t-1) := \mathbb{E}[x(t)|z(\tau) : \tau \in \mathcal{M}, \tau < t]$ as the best a priori mean squared estimator of $x(t)$ and $\hat{x}_{\mathcal{M}}(t|t) := \mathbb{E}[x(t)|z(\tau) : \tau \in \mathcal{M}, \tau \leq t]$ as the best a posteriori mean squared estimator of $x(t)$. In addition, define the a priori and the a posteriori covariance matrices as $P(t|t-1) := \mathbb{E}[(x(t)-\hat{x}_{\mathcal{M}}(t|t-1))(x(t)-\hat{x}_{\mathcal{M}}(t|t-1))^\top]$

and $P(t|t) := \mathbb{E}[(x(t) - \hat{x}_{\mathcal{M}}(t|t))(x(t) - \hat{x}_{\mathcal{M}}(t|t))^{\top}]$, respectively. The classical Kalman filtering theory [1] states how to update these four quantities recursively in the case where a measurement is acquired at each time step, i.e., when $N = T$. In addition, by the linear relation (2) between $x(t)$ and $y(t)$, it holds that

$$\hat{y}_{\mathcal{M}}(t|t - 1) = B\hat{x}_{\mathcal{M}}(t|t - 1).$$

We consider the intermittent case by replacing the measurement matrix C with the null matrix when no measurement is available, i.e., when $t \notin \mathcal{M}$. In view of Eq. (3), it models the fact that $z(t)$ does not contain any information about the state $x(t)$. Then the equations of the Intermittent Kalman predictor are the *time update* equations

$$P(t + 1|t) = AP(t|t)A^{\top} + GQG^{\top}, \tag{6}$$
$$\hat{x}_{\mathcal{M}}(t + 1|t) = A\hat{x}_{\mathcal{M}}(t|t) + b, \tag{7}$$
$$\hat{y}_{\mathcal{M}}(t + 1|t) = B\hat{x}_{\mathcal{M}}(t + 1|t), \tag{8}$$

for $t = 0, \ldots, T - 1$. In addition, the *measurement update* equations are

$$K(t) = \begin{cases} P(t|t - 1)C^{\top}[CP(t|t - 1)C^{\top} + R]^{-1} & \text{if } t \in \mathcal{M} \\ 0 & \text{else,} \end{cases} \tag{9}$$

$$P(t|t) = (I - K(t)C)P(t|t - 1), \tag{10}$$
$$\hat{x}_{\mathcal{M}}(t|t) = \hat{x}_{\mathcal{M}}(t|t - 1) + K(t)[z(t) - C\hat{x}_{\mathcal{M}}(t|t - 1) - d], \tag{11}$$
$$\hat{y}_{\mathcal{M}}(t|t) = B\hat{x}_{\mathcal{M}}(t|t), \tag{12}$$

for $t = 0, \ldots, T$. Note that when $t \notin \mathcal{M}$, $z(t)$ is not defined in Eq. (11) but can be set to any arbitrary value because it is multiplied by $K(t) = 0$. In addition, one can see from (9)-(11) that when no measurement is available, i.e., when $t \notin \mathcal{M}$, each a posteriori quantity is simply the a priori one, which could have been expected intuitively. The *initialization* of these recurrence equations are

$$P(0| - 1) = \bar{P}_0, \ \hat{x}_{\mathcal{M}}(0| - 1) = \bar{x}_0, \ \hat{y}_{\mathcal{M}}(0| - 1) = B\bar{x}_0. \tag{13}$$

The intermittent Kalman predictor is summarized by Eqs. (6) to (13).

We call a Kalman predictor for which the N measurement times are selected as equally spaced as possible a *Regular Kalman predictor*. It is

$$\mathcal{M}_{\text{REG}} := \left\{ \text{Round}\left[\frac{kT}{N}\right] \middle| k = 0, \ldots, N - 1 \right\}, \tag{14}$$

where Round[$\cdot$] is the rounding operator.

2.3 Optimal intermittent Kalman predictor

The optimal intermittent Kalman predictor is the intermittent Kalman predictor for which the set of measurement times $\mathcal{M}$ is chosen to minimize the mean error variance, i.e., it is the solution of problem (5).

Let us denote the a priori error $\tilde{y}(t) := y(t) - \hat{y}_{\mathcal{M}}(t|t-1)$. Its covariance matrix can be written

$$
\begin{aligned}
S(t) &:= \mathbb{E}\left[\tilde{y}(t)\tilde{y}(t)^{\top}\right] \\
&= \mathbb{E}\left[(y(t) - \hat{y}_{\mathcal{M}}(t|t-1))(y(t) - \hat{y}_{\mathcal{M}}(t|t-1))^{\top}\right] \\
&= \mathbb{E}\left[(Bx(t) - B\hat{x}_{\mathcal{M}}(t|t-1))(Bx(t) - B\hat{x}_{\mathcal{M}}(t|t-1))^{\top}\right] \\
&= B\mathbb{E}\left[(x(t) - \hat{x}_{\mathcal{M}}(t|t-1))(x(t) - \hat{x}_{\mathcal{M}}(t|t-1))^{\top}\right]B^{\top} \\
&= BP(t|t-1)B^{\top}.
\end{aligned}
$$

Then, the variance of the prediction error norm $\|\tilde{y}(t)\|$ can be written in terms of $P(t|t-1)$ as

$$
\mathbb{E}[\|\tilde{y}(t)\|^2] = \mathrm{Tr}[S(t)] = \mathrm{Tr}[BP(t|t-1)B^{\top}],
$$

where $\mathrm{Tr}[\cdot]$ is the trace operator.

Thanks to the previous equations, problem (5) can be reformulated as

$$
\min_{\mathcal{M} \subset \{0,\ldots,T-1\}} \frac{1}{T} \sum_{t=1}^{T} \mathrm{Tr}[BP(t|t-1)B^{\top}] \text{ subject to}
$$

$$
|\mathcal{M}| = N, \text{Eqs. (6), (9) and (10) with } P(0|-1) = \bar{P}_0, \tag{15}
$$

where constraint (6) holds for $t = 0, \ldots, T-1$ and constraints (9) and (10) hold for $t = 0, \ldots, T$.

Remark 1 *The equations that govern covariance matrices, i.e., Eqs. (6), (9) and (10), are independent of the measurements $z(t)$. Consequently, the optimization problem (15) can be solved before measurements are made, i.e., offline. In addition, they are independent of b and d.*

Remark 2 *Contrary to Eqs. (1) through (4) which are stochastic, problem (15) is deterministic.*

In addition, the equations involving $\hat{x}_{\mathcal{M}}$ and $\hat{y}_{\mathcal{M}}$ can be ignored during selection of the measurement times, i.e., the resolution of problem (15), and used only for prediction. Consequently, using the optimal intermittent Kalman predictor consists in firstly computing the optimal measurement times offline and then, doing online prediction.

Remark 3 *Note that $x(t)$ and $y(t)$ are linearly related by relation (2). However an optimal set of measurement times given for estimating the $y(t)$, i.e., an optimal solution of problem (15), is not necessarily optimal for estimating the state $x(t)$. Indeed, the objective function of the problem depends on the matrix B that connects $y(t)$ to $x(t)$.*

Remark 4 *It is possible to compute the distribution of the norm of the squared prediction error $\|\tilde{y}(t)\|^2$. First, consider the following eigendecomposition, $S(t) = \Sigma \Lambda \Sigma^{\top}$, where Λ is a diagonal matrix whose diagonal entries are the eigenvalues $\lambda_1(t), \ldots, \lambda_n(t)$ of the symmetric matrix $S(t) = BP(t|t)B^{\top}$, and where Σ is a unitary matrix, i.e., $\Sigma^{\top}\Sigma = I$.*

Define $\zeta := \Lambda^{-1/2}\Sigma^{\top}\tilde{y}(t)$. Note that Σ, Λ and ζ depend on t but this dependency is omitted for clarity. Each random variable ζ follows a standard normal distribution $\mathcal{N}(0, I)$. Indeed, it is a linear transformation of a centered Gaussian random variable so it is also a

central Gaussian random variable. In addition, its covariance matrix is identity. It can be computed using the commutativity of diagonal matrices and the fact that Σ is unitary, so that we obtain,

$$
\begin{aligned}
\mathbb{E}\left[\zeta\zeta^\top\right] &= \mathbb{E}\left[\left(\Lambda^{-1/2}\Sigma^\top\tilde{y}(t)\right)\left(\Lambda^{-1/2}\Sigma^\top\tilde{y}(t)\right)^\top\right] \\
&= \Lambda^{-1/2}\Sigma^\top\mathbb{E}\left[\tilde{y}(t)\tilde{y}(t)^\top\right]\Sigma\Lambda^{-1/2} \\
&= \Lambda^{-1/2}\Sigma^\top\Sigma\Lambda\Sigma^\top\Sigma\Lambda^{-1/2} \\
&= I.
\end{aligned}
$$

The random variable $\|\tilde{y}(t)\|^2$ can be expressed in terms of ζ, yielding

$$
\begin{aligned}
\|\tilde{y}(t)\|^2 &= \tilde{y}(t)^\top\tilde{y}(t) \\
&= (\Sigma\Lambda^{1/2}\zeta)^\top(\Sigma\Lambda^{1/2}\zeta) \\
&= \zeta^\top\Lambda^{1/2}\Sigma^\top\Sigma\Lambda^{1/2}\zeta \\
&= \zeta^\top\Lambda\zeta \\
&= \sum_{i=1}^{n}\lambda_i(t)\zeta_i^2,
\end{aligned}
$$

where all ζ_i^2 follow an independent chi-squared distribution with one degree of freedom, i.e., $\zeta_i^2 \sim \chi^2$. It shows that $\|\tilde{y}(t)\|^2$ is a linear combination with non-negative coefficients of independent random variables following a chi-squared distribution with one degree of freedom. Then, a closed form formula can be obtained for the cumulative distribution function $\mathbb{P}[\|\tilde{y}(t)\|^2 < \xi]$ using the theorem of [35]. From a practical point of view, this quantity can be efficiently estimated using the method proposed in [36].

2.4 Optimization algorithms

Solving the combinatorial optimization problem (15) by means of an exhaustive search would require evaluating $T!/(N!\,(T-N)!)$ times the cost function. That is computationally intractable. In this section, we propose a random trial (RT) algorithm and various genetic algorithms (GAs) to tackle this problem.

The RT algorithm randomly samples a given number of admissible solutions $\mathcal{M}$, computes their cost and keeps the best one.

In the GA nomenclature, a feasible solution $\mathcal{M} = \{t_1, \ldots, t_N\}$ of the optimization problem is called an individual and its corresponding measurement times t_i are called its genes. A set of several individuals is called a population.

The GA [37] implements the following steps.

1. **Initialization:** An initial population is uniformly sampled on the admissible set of problem (15).
2. **Evaluation:** For each individual in the population, the cost is evaluated according to the objective function of problem (15).
3. **Selection:** Individuals with low costs are preferably selected according to stochastic universal sampling [37].
4. **Crossover:** Selected individuals are mixed to produce new individuals called the offspring. This reconstitutes a complete population. Three different types of

crossover are considered: the shuffle crossover (SC), replace crossover (RC) and count preserving crossover (CPC). They are described below.

5. **Mutation:** Each gene of each individual mutates according to a probability. When a mutation occurs, some measurement times are replaced by random times. This step could create duplicates, i.e., an individual could have repeated times $t_i = t_j$ with $i \neq j$. To avoid this situation, the random times are selected uniformly in $\{0, \ldots, T-1\} \backslash \mathcal{M}$.

6. **Repeat:** Come back to step 2 until a convergence criterion is reached.

One pass of steps 2 to 5 is called a generation.

Let us present the three different crossover operators. Each of them gives a variant of the genetic algorithm. They are described and a brief example is given.

Shuffle crossover (SC) For each parent pair, the SC [38] method picks the genes to exchange (each gene has the same probability to be selected) randomly. For example, let two parents (that is, two sets of measurement times $\mathcal{M}$),

$$P_1 = 0 \ \underline{1} \ 3 \ \underline{5} \ \underline{6} \ 7 \text{ and } P_2 = 0 \ \underline{1} \ 2 \ \underline{3} \ \underline{5} \ 8, \tag{16}$$

where _ indicates the location of genes selected for the crossover. The obtained offspring are

$$O_1 = 0 \ 1 \ 3 \ 3 \ 5 \ 7 \text{ and } O_2 = 0 \ 1 \ 2 \ 5 \ 6 \ 8. \tag{17}$$

First, one can observe that O_1 contains a duplication of 3, which corresponds to choosing the same measurement time twice. That is not allowed in problem (15) and has to be considered a wasted measurement. In other words, fewer measurements are acquired, which is suboptimal, because adding a new distinct measurement can only decrease the cost.

A second observation is that the second offspring O_2 does not contain a 3, while both parents P_1 and P_2 contain one. That means that common heritage is lost during the crossover, which possibly deteriorates the convergence of the algorithm.

These two observations motivate a careful choice of the crossover method and leads naturally to the next two crossover methods.
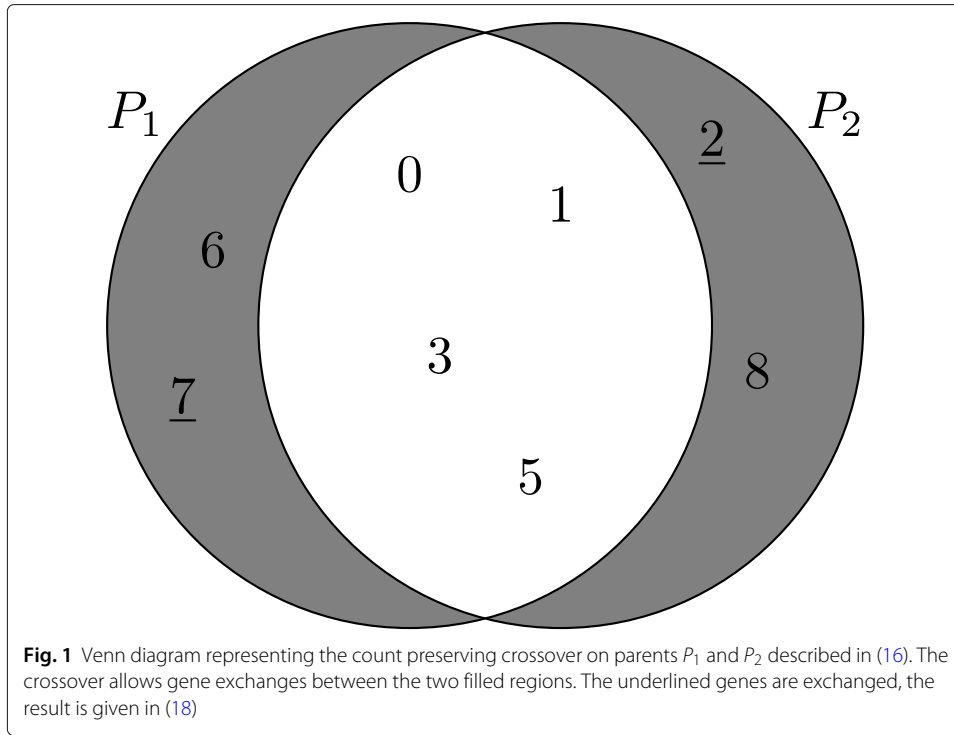
Replace crossover (RC) The RC method is an SC followed by a random replacement of duplicates, taking care to avoid new duplicates. Duplicates are replaced by picking genes uniformly at random in $\{0, \ldots, T-1\} \backslash \mathcal{M}$. The offspring (17) become, for example

$$O_1 = 0 \ 1 \ 3 \ 5 \ 7 \ \underline{8} \text{ and } O_2 = 0 \ 1 \ 2 \ 5 \ 6 \ 8,$$

where _ indicates the new gene. Note that the second offspring remains unchanged because it does not contain duplicates.

With this crossover, the obtained offspring have no duplicates. However, the second offspring O_2 does not contain a 3 while this gene was common to both parents. This illustrates that, like SC, this RC can also suffer from a loss of common heritage.

Count preserving crossover (CPC) The CPC [38, 39] represents each individual by a set (unordered) instead of a vector (ordered). It implements an SC between subsets $P_1 \backslash P_2$ and $P_2 \backslash P_1$. It allows transmission of the common measurement times $P_1 \cap P_2$ to both

Fig. 1 Venn diagram representing the count preserving crossover on parents P_1 and P_2 described in (16). The crossover allows gene exchanges between the two filled regions. The underlined genes are exchanged, the result is given in (18)

offspring. For parents in example (16), the SC is executed between subsets $P_1 \setminus P_2 = \{6, 7\}$ and $P_2 \setminus P_1 = \{2, 8\}$ (see Fig. 1). Possible offspring are then

$$O_1 = 0 \ 1 \ \underline{2} \ 3 \ 5 \ 6 \text{ and } O_2 = 0 \ 1 \ 3 \ 5 \ \underline{7} \ 8. \tag{18}$$

This crossover method makes it possible both to transmit the common genes to all the offspring and to avoid duplicates.

All GA implementations use sigma scaling [37] with the sigma factor equal to 1 (standard value). The crossover operation is applied systematically and mutation occurs with a probability by gene of 0.003 (standard value). The population size is set to 100 individuals and the algorithm stops after 100 generations.

2.5 Discretization of a continuous-time system

Many real-world problems are modeled as continuous-time systems. For this reason, we study the continuous-time equivalent of problem (5). It is[1],

$$\inf_{0 \le \tau_0 < \tau_1 < \cdots < \tau_{N-1} \le \bar{\tau}} \frac{1}{\bar{\tau}} \int_0^{\bar{\tau}} \mathbb{E}[\, \|y(\tau') - \hat{y}(\tau')\|^2]\, d\tau', \tag{19}$$

such that

$$
\begin{aligned}
\frac{dx}{d\tau}(\tau) &= A^c x(\tau) + b^c + G^c w^c(\tau) & \tau \in [0, \bar{\tau}], \\
y(\tau) &= B^c x(\tau) & \tau \in [0, \bar{\tau}], \\
z(\tau_k) &= C x(\tau_k) + d + v(\tau_k) & k = 0, \ldots, N-1, \\
x(0) &\sim \mathcal{N}(\bar{x}_0, \bar{P}_0),
\end{aligned}
\tag{20}
$$

---

[1]An infimum is used because the admissible set of the optimization problem is not compact due to constraint $\tau_k < \tau_{k+1}$. This constraint models the fact that two measurements can not be acquired at the same time. When the time is discretized at time step δ, this constraint becomes $\tau_k + \delta \le \tau_{k+1}$ which makes the set compact. Then, the corresponding infimum exists and is a minimum.

where $w^c(\tau) \sim \mathcal{N}(0, Q^c)$ and $v(\tau_k) \sim \mathcal{N}(0, R)$ are independent Gaussian white noises. Finally, $\hat{y}(\tau) = \mathbb{E}[y(\tau)|z(\tau_k) : \tau_k < \tau]$ is the best mean squared estimator of $y(\tau)$. The exponent ^{c} emphasizes that time is continuous. Note that this is a stochastic differential equation for which derivatives have a non-classical sense. The reader is referred to [40] for details.

This problem can be exactly discretized at a time step $\delta = \bar{\tau}/T$ to fit the formalism of Eqs. (1) through (5). The corresponding quantities are [41],

$$A = e^{A^c \delta}, \ B = B^c, \ b = \int_0^\delta e^{A^c \tau} d\tau b^c, \tag{21}$$

$$Q = \int_0^\delta e^{A^c \tau} G^c Q^c (G^c)^\top e^{(A^c)^\top \tau} d\tau, \ G = 1, \tag{22}$$

where $e^\cdot$ is the matrix exponential operator.

In Section 3.4, we study the impact of the discretization time step on the obtained solution.

2.6 Optimal intermittent linear quadratic regulator

Duality between estimation and control problems has been known for decades [42]. Roughly, it states that instead of solving a control problem directly, one can solve a related estimation problem and deduce the solution to the control problem from the solution of the estimation problem. The converse is also possible.

In this section, we formalize the problem of optimal intermittent LQR and use duality to show that it can be handled by computing an optimal intermittent Kalman predictor.

Let us consider a controlled linear system with known initial state,

$$x(t+1) = \tilde{A}x(t) + \tilde{\sigma}(t)\tilde{B}u(t), \ \text{for } t = 0, \dots, T-1, \tag{23}$$

$$x(0) = \tilde{x}_0, \tag{24}$$

where $\tilde{\sigma}(t) \in \{0, 1\}$. This control signal $\tilde{\sigma}(\cdot)$ determines the times at which the system is controlled. Note that $\tilde{A}$ and $\tilde{B}$ can be time-dependent.

The *optimal intermittent LQR problem* consists of the following minimization,

$$\tilde{\sigma}^*(\cdot) = \arg\min_{\tilde{\sigma}(\cdot)} V(\tilde{\sigma}(\cdot)) \text{ such that } \sum_{t=0}^{T-1} \sigma(t) = N, \tag{25}$$

with

$$V(\tilde{\sigma}(\cdot)) = \min_{u(\cdot)} \left\{ x(T)^\top \tilde{Q}_f x(T) + \sum_{t=0}^{T-1} x(t)^\top \tilde{Q}x(t) + u(t)^\top \tilde{R}u(t) \right\}$$

$$\text{subject to (23) and (24)},$$

where $\tilde{Q}$ and $\tilde{R}$ can be time-dependent. This problem has been studied in [43] where a closed form formula is provided under some restrictive conditions.

The following theorem formalizes how to compute this optimal LQR by computing an optimal intermittent Kalman predictor. Note that at each optimal set of measurement times $\mathcal{M}$ can be associated with a signal $\sigma(\cdot)$ such that $\sigma(t) = 1$ if $t \in \mathcal{M}$ and 0 otherwise.

Theorem 1 *The optimal solution* $\tilde{\sigma}^*(\cdot)$ *of the problem* (25) *is given by* $\tilde{\sigma}^*(t) = \sigma^*(T-t)$ *where* $\sigma^*(\cdot)$ *is the optimal solution of the estimation problem* (15) *solved for parameters*

$$A(t) = \tilde{A}(T-t)^\top, \ Q(t) = \tilde{Q}(T-t), \ R(t) = \tilde{R}(T-t), \tag{26}$$

$$b(t) = 0, \ G = I, \ C(t) = \tilde{B}(T-t)^\top, \ \bar{P}_0 = \tilde{Q}_f, \ d(t) = 0, \ and \tag{27}$$

$$B(t) = \delta(T-t)\tilde{x}_0^\top, \tag{28}$$

where $\delta(T-t)$ *is 1 if* $t = T$, *and 0 otherwise.*

In addition, the optimal control law is

$$u(t) = -K_{\sigma^*(\cdot)}(t)^\top A(t)^\top x(t), \tag{29}$$

where $K_{\sigma^*(\cdot)}(t)$ *is given by* (9).

Proof We want to use Theorem 2.1 from [42]. However, these authors' formalism has two differences from ours.

First, they do not consider missing control times (or missing measurement times). However, this is not restrictive because for a given $\tilde{\sigma}(t)$, Eq. (23) can reduce to the form of [42] by considering $\tilde{\sigma}(t)\tilde{B}$ as a known time-varying control matrix $\bar{B}(t)$. Similarly, the measurement matrix C in Eq. (3) can be seen as the null matrix when no measurement is acquired, i.e., when $\sigma(t) = 0$ or equivalently $t \notin \mathcal{M}$.

The second difference with [42] is that they limit the estimation problem to the case where $B = I$, i.e., $y(t) = x(t)$. However, as shown by Eqs. (6), (9), and (10), the prediction error covariance matrix $P(t|t-1)$ does not depend on B. Consequently, results from Theorem 2.1 from [42] that concern $P(t|t-1)$ can be used.

Thus, for a given signal $\tilde{\sigma}(\cdot)$, Theorem 2.1 from [42] states that $V(\tilde{\sigma}(\cdot)) = \tilde{x}_0^\top \tilde{P}_{\tilde{\sigma}(\cdot)}\tilde{x}_0$ where $\tilde{P}_{\tilde{\sigma}(\cdot)} = P_{\sigma(\cdot)}(T|T-1)$ and the last is given by Eqs. (6), (9), and (10), under transformations $\sigma(t) = \tilde{\sigma}(T-t)$, (26) and (27).

Then, under transformation (28), the objective function of the estimation problem (15) is

$$\sum_{t=1}^{T} \mathrm{Tr}[B(t)P_{\sigma(\cdot)}(t|t-1)B(t)^\top] = \sum_{t=1}^{T} \mathrm{Tr}[\delta(T-t)\tilde{x}_0^\top P_{\sigma(\cdot)}(t|t-1)\tilde{x}_0]$$
$$= \tilde{x}_0^\top P_{\sigma(\cdot)}(T|T-1)\tilde{x}_0$$
$$= \tilde{x}_0^\top \tilde{P}_{\tilde{\sigma}(\cdot)}(0)\tilde{x}_0$$
$$= V(\tilde{\sigma}(\cdot)),$$

where the trace operator has been removed because its argument is one-dimensional. Taking the minimum overall admissible $\sigma(\cdot)$ (on the left-hand side) and the corresponding $\tilde{\sigma}(\cdot)$ (on the right-hand side) gives equality between problems (15) and (25).

Once the signal $\tilde{\sigma}^*(\cdot)$ is fixed, finding the optimal command $u(t)$ is a classic LQR problem. Theorem 2.1 form [42] gives the relation (29). $\square$

2.7 Quality or quantity of measurements

Instead of considering a fixed number of measurements, an alternative problem is to reach a trade-off between many low-quality measurements and a small number of high-quality measurements. This problem can be handled in the presented formalism.

Allowing the choice of the quantity of measurements corresponds to making the number of measurements N a variable to optimize and no longer a given of the problem. However, with only this modification, the optimal solution will always be to take a maximum number of measurements, i.e., $N = T$. To model the compromise between measurement quality and quantity, an increase in the number of measurements N implies an increase in the noise of measurements, i.e., an increase in R.

In other words, the measurement covariance matrix R must depends on N. We assume $R = f(N)$ for a given function $f : \{1, \ldots, T\} \to \mathcal{S}_+^{p \times p}$ where $\mathcal{S}_+^{p \times p}$ is the set of $p \times p$ positive semi-definite matrices. Then, the problem can be formalized similarly to (15), it gives,

$$\min_{1 \leq N \leq T} \min_{\mathcal{M} \subset \{0, \ldots, T-1\}} \frac{1}{T} \sum_{t=1}^{T} \mathrm{Tr}[\, BP(t|t-1)B^\top \,] \ \text{ subject to}$$

$$|\mathcal{M}| = N, \text{Eqs. (6), (9) and (10)}, P(0|-1) = \bar{P}_0 \text{ and } R = f(N), \tag{30}$$

where, as previously, constraint (6) holds for $t = 0, \ldots, T - 1$ and constraints (9) and (10) hold for $t = 0, \ldots, T$.

This problem is studied in Section 3.6.

3 Results and discussion

Several results will be illustrated on a discretized spring-mass system. The derivation of these equations will be detailed in Section 3.4. For readability, we anticipate the system finally obtained; it is defined by

$$A = \begin{pmatrix} \cos \delta & \sin \delta \\ -\sin \delta & \cos \delta \end{pmatrix}, \ \ B = (1 \quad 0), \ \ b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \ \ G = 1, \tag{31}$$

$$Q = \frac{1}{80} \begin{pmatrix} \delta - \sin \delta \cos \delta & \sin^2 \delta \\ -\sin^2 \delta & \delta + \sin \delta \cos \delta \end{pmatrix}, \tag{32}$$
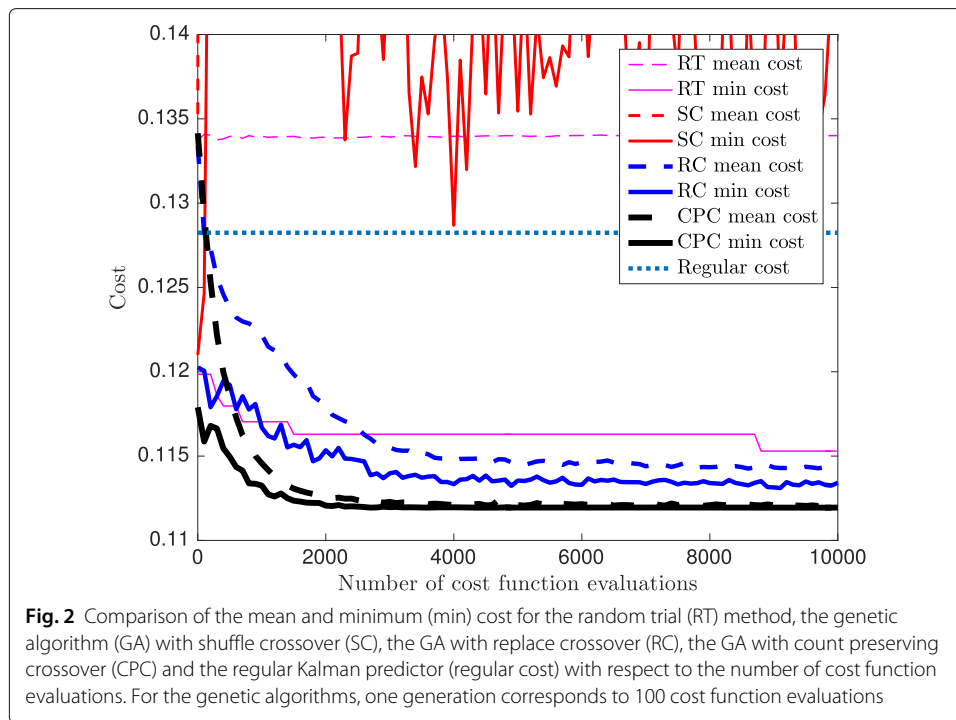
where δ is the discretization time step.

3.1 Comparison of optimization methods

In this section, the different optimization algorithms presented in Section 2.4 are compared. The random trial (RT) method and the three variants of Genetic Algorithms (GA) are compared, i.e., Shuffle Crossover (SC), Replace Crossover (RC), and Count Preserving Crossover (CPC).

Figure 2 presents the mean cost and the minimum cost obtained by the four algorithms with respect to the number of cost function evaluations in the case of the spring-mass system described in Section 3.4 by Eqs. (31) and (32) with a discretization time step $\delta = 0.1$ [s]. The number of measurements is set to $N = 70$ and the number of time steps is $T = 100$. The mean and minimum costs obtained with RT are also printed. In addition, the cost for regularly spaced measurement times is printed.

Firstly, one can observe that the minimum and average of the GA with SC has bad convergence behavior (its average cost quickly rises out of the figure). This confirms our previous comments in Section 2.4 about the flaws of this crossover operator (creation of duplicates and loss of common heritage). All other algorithms found a solution better than the regular one in few generations. The regular cost is close to the average RT cost, meaning that regular measurement times are 'typical' random measurement times

Fig. 2 Comparison of the mean and minimum (min) cost for the random trial (RT) method, the genetic algorithm (GA) with shuffle crossover (SC), the GA with replace crossover (RC), the GA with count preserving crossover (CPC) and the regular Kalman predictor (regular cost) with respect to the number of cost function evaluations. For the genetic algorithms, one generation corresponds to 100 cost function evaluations

in this example. The GAs with RC and CPC outperform the RT algorithm. Finally, the GA with CPC quickly outperforms all other proposed algorithms. In addition, the average cost converges to the minimal cost only for the CPC. This means that the entire population converges to a single individual—assumed to be optimal—which is the desired behavior for a GA. Table 1 prints the averages and standard deviations of the optimal costs found by the different algorithms over 100 resolutions. The advantage of the CPC crossover is confirmed and the small standard deviations indicate the reproducibility of the results.

In the following, all experiments use the GA that implements CPC.

3.2 Numerical examples

To continue the analysis of our method, predictions using our method are compared with the ones obtained with a Kalman predictor with regularly spaced measurement times (see Eq. (14)). After an analysis on the spring-mass system, we will apply our method to a 50-dimensional system. To conclude this section, some links with the observability matrix will be highlighted.

Table 1 Average and standard deviation of the optimal costs found by the different algorithms on 100 resolutions. The algorithms are the random trial (RT) method, the genetic algorithm (GA) with shuffle crossover (SC), the GA with replace crossover (RC), the GA with count preserving crossover (CPC). The cost obtained with regularly spaced measurements (see (14)) is also indicated

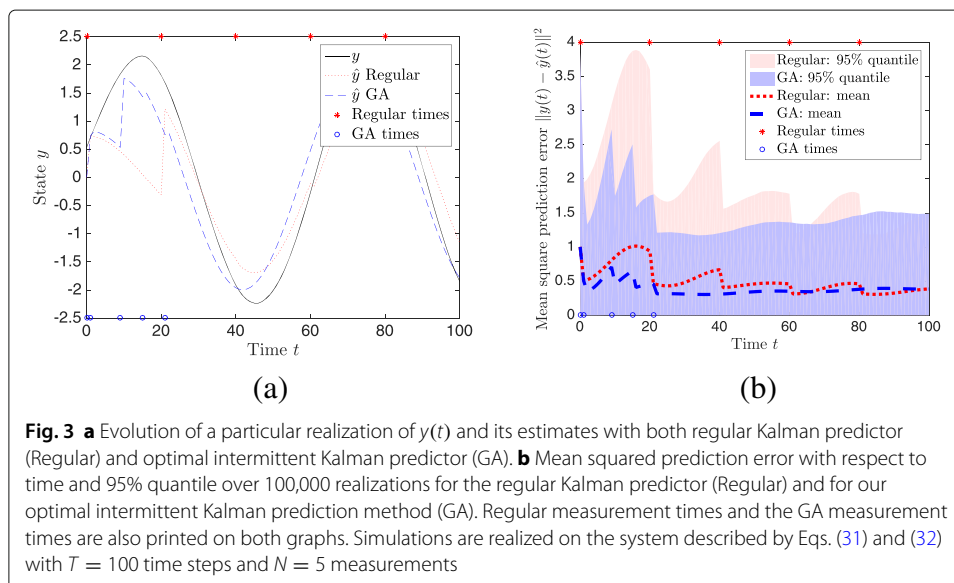| RT	SC	RC	CPC	Regular
$0.115 \pm 6 \cdot 10^{-4}$	0.120 ± 10^{-3}	0.113 ± 10^{-4}	$0.112 \pm 2 \cdot 10^{-6}$	0.128

3.2.1 Spring-mass system

The set of measurement times found by the genetic algorithm is denoted $\mathcal{M}_{\mathrm{GA}}$ and the set of regularly spaced measurements is denoted $\mathcal{M}_{\mathrm{REG}}$. The corresponding mean squared tracking errors are denoted, MSE($\mathcal{M}_{\mathrm{GA}}$) and MSE($\mathcal{M}_{\mathrm{REG}}$), respectively.
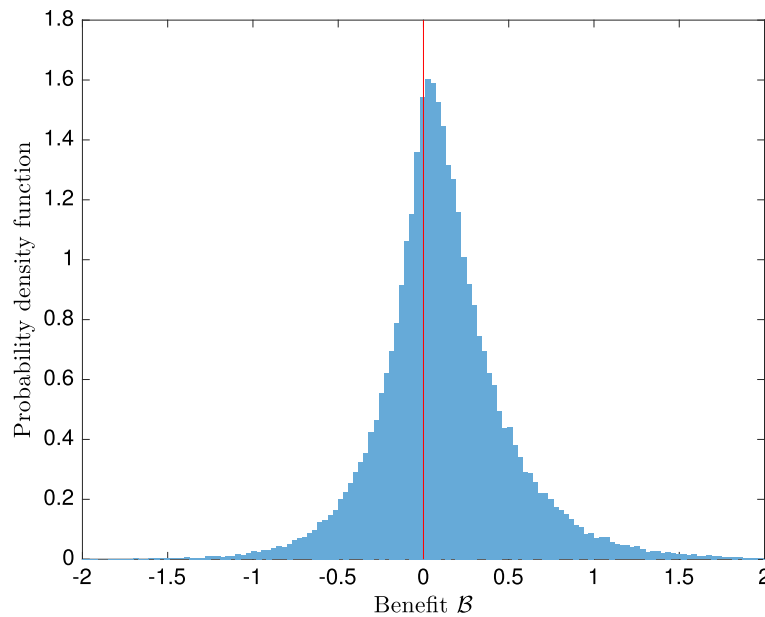
For the same system as in the previous section, i.e., Eqs. (31) and (32) with a discretization time step $\delta = 0.1\,[s]$ and with $T = 100$ time steps and $N = 5$ measurements, Fig. 3a presents the real $y(t)$ and the predictions obtained with the two predictors. In additions, the regular measurement times $\mathcal{M}_{\mathrm{REG}}$ and the GA measurement times $\mathcal{M}_{\mathrm{GA}}$ are printed. One can see that the main difference occurs at the beginning of the tracking. Indeed, our method gets closer faster. Intuitively, this can be understood by the fact that for the intermittent predictor, all measurement times are selected close to the beginning. Over the complete time horizon, the regular approach has a mean squared error of MSE($\mathcal{M}_{\mathrm{REG}}$) = 0.46 whereas the mean squared error with our method is MSE($\mathcal{M}_{\mathrm{GA}}$) = 0.36. This is a significant improvement. However, these results are valid only for this particular realization of the dynamic system.

On the contrary, Fig. 3b presents the squared error with respect to time over 100,000 realizations. The mean and the 95% quantile are indicated for the regular Kalman predictor and for our optimal intermittent Kalman predictor method. One can see that the remarks about Fig. 3a hold for the mean and the 95% quantile behavior: our method rapidly produces a better estimate than the Kalman predictor with regular measurements.

Let us introduce the *benefit* as $\mathcal{B} = \mathrm{MSE}(\mathcal{M}_{\mathrm{REG}}) - \mathrm{MSE}(\mathcal{M}_{\mathrm{GA}})$. A positive benefit indicates that our method outperforms the regular Kalman predictor. Figure 4 presents the histogram of the benefit $\mathcal{B}$ computed over 100,000 realizations. It shows that the mean benefit is 0.12 and the benefit is positive in 64% of the cases, i.e., our method outperforms the regular Kalman predictor. The results are summarized in the second column of Table 2. Note that the optimal costs correspond to the average mean squared errors.

Fig. 3 a Evolution of a particular realization of $y(t)$ and its estimates with both regular Kalman predictor (Regular) and optimal intermittent Kalman predictor (GA). **b** Mean squared prediction error with respect to time and 95% quantile over 100,000 realizations for the regular Kalman predictor (Regular) and for our optimal intermittent Kalman prediction method (GA). Regular measurement times and the GA measurement times are also printed on both graphs. Simulations are realized on the system described by Eqs. (31) and (32) with $T = 100$ time steps and $N = 5$ measurements

Fig. 4 Histogram of the benefit $\mathcal{B} = \text{MSE}(\mathcal{M}_{\text{REG}}) - \text{MSE}(\mathcal{M}_{\text{GA}})$ computed over 100,000 realizations for system (31) and (32) with $T = 100$ time steps and $N = 5$ measurements. The histogram approximates the probability density function of the benefit. The red vertical line indicates a null benefit. The mean and standard deviation of the benefit are 0.12±0.38 and the benefit is positive in 64% of the realizations

3.2.2   A large dimensional system

We propose here to apply our method to a 50-dimensional system. To this end, we generate a random system as follows: each entry of the matrix $A \in \mathbb{R}^{50 \times 50}$ is picked at random independently from a Gaussian distribution with standard deviation 0.25. We consider $b = 0$ and $G = Q = I$. Ten components of the state $x(t)$ are uniformly picked at random to form the objective vector $z(t)$. In other words, each row of the matrix $B \in \{0, 1\}^{10 \times 50}$ contains only one 1 and all lines are different. The measurement matrix $C \in \{0, 1\}^{10 \times 50}$ is picked at random in the same way but independently of B. We consider $d = 0$ and the covariance matrix of the process noise is the identity, i.e., $R = I$. Finally, we consider $\bar{x}_0 = 0$ and $\bar{P}_0 = I$. This system is studied over $T = 50$ time steps and $N = 25$ measurements are allowed.

Table 2 Results for the spring-mass system (Section 3.2.1) and the random system (Section 3.2.2). The cost using regularly spaced measurements and the optimal cost found by the GA are indicated. The mean square errors are indicated (mean and standard deviation over 100,000 realizations). The benefit $\mathcal{B} = \text{MSE}(\mathcal{M}_{\text{REG}}) - \text{MSE}(\mathcal{M}_{\text{GA}})$ is also indicated (mean, standard deviation and proportion of positive)

	Spring-mass system	Random system
Cost regular	0.52	16,817.80
Cost GA	0.39	10,621.31
$\text{MSE}(\mathcal{M}_{\text{REG}})$ (mean ± std)	0.51 ± 0.42	$16,506.12 \pm 3212.83$
$\text{MSE}(\mathcal{M}_{\text{GA}})$ (mean ± std)	0.39 ± 0.33	$10,421.00 \pm 1866.53$
Benefit $\mathcal{B}$ (mean ± std)	0.12 ± 0.38	6085.12 ± 3391.64
Proportion of positive benefit $\mathcal{B}$	64%	97%

The eigenvalues of the randomly sampled matrix A are depicted in Fig. 5a. Among the 50 eigenvalues, 17 are stable and 33 are unstable. Their modulus varies from 0.12 to 2.09.

The optimal measurement times for this problem are computed using the GA with the CPC. Then, the mean squared prediction error using the optimal intermittent Kalman predictor is compared to the one with regular measurements on 100,000 realizations. The average and standard deviation of the mean squared error in the regular case $\text{MSE}(\mathcal{M}_{\text{REG}})$ are $10,374.91 \pm 1655.75$. Using the optimal measurement times, the average and standard deviation of $\text{MSE}(\mathcal{M}_{\text{GA}})$ are 8947.08 ± 1785.55. The benefit $\mathcal{B} = \text{MSE}(\mathcal{M}_{\text{REG}}) - \text{MSE}(\mathcal{M}_{\text{GA}})$ is computed and its distribution is depicted in Fig. 5b. The mean benefit is 1427.83 and the benefit is positive for 76% of the realizations. These results are summarized in the third column of Table 2.

3.3 Links with observability matrix

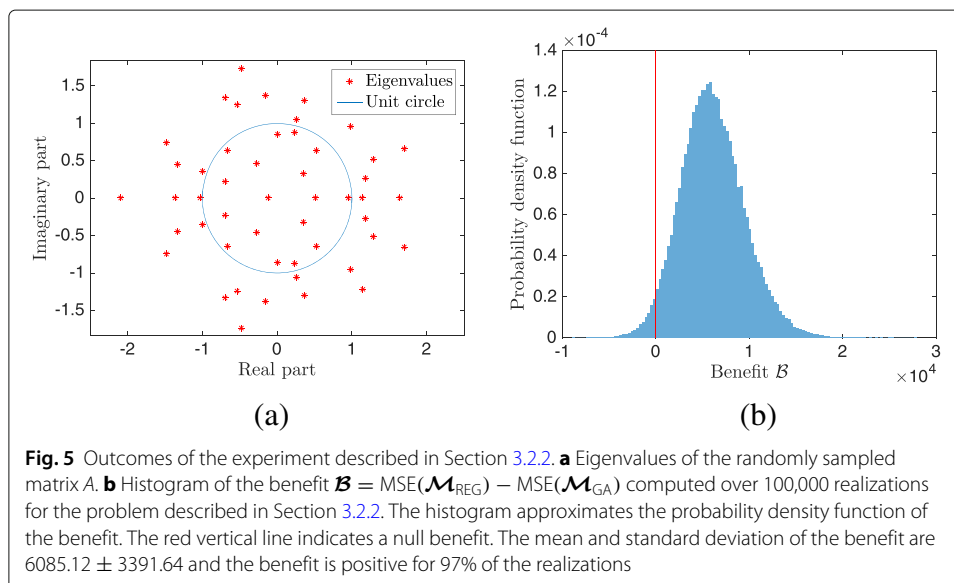To make the results more intuitive, let us present a simple system defined by

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \ B = Q = \bar{P}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ C = (1 \quad 0), \ R = G = 1, \ b = d = 0, \quad (33)$$

and consider $T = 20$ time steps and $N = 10$ measurements.

Jungers et al. show in [44] that the observability matrix of a linear system with missing measurements is given by

$$\mathcal{O}_T(A, C, \mathcal{M}) = \begin{pmatrix} \sigma(0)C \\ \sigma(1)CA \\ \vdots \\ \sigma(T-1)CA^{T-1} \end{pmatrix},$$

where $\sigma(t) = 1$ if $t \in \mathcal{M}$ and 0 otherwise. For system (33), because the transition matrix A is a 90 degree rotation matrix, the observability matrix has the following structure,

Fig. 5 Outcomes of the experiment described in Section 3.2.2. **a** Eigenvalues of the randomly sampled matrix A. **b** Histogram of the benefit $\mathcal{B} = \text{MSE}(\mathcal{M}_{\text{REG}}) - \text{MSE}(\mathcal{M}_{\text{GA}})$ computed over 100,000 realizations for the problem described in Section 3.2.2. The histogram approximates the probability density function of the benefit. The red vertical line indicates a null benefit. The mean and standard deviation of the benefit are 6085.12 ± 3391.64 and the benefit is positive for 97% of the realizations

$$\mathcal{O}_T(A, C, \mathcal{M}) = \begin{pmatrix} \sigma(0) & 0 & -\sigma(2) & 0 & \sigma(4) & \cdots & 0 \\ 0 & -\sigma(1) & 0 & \sigma(3) & 0 & \cdots & \sigma(19) \end{pmatrix}^\top. \qquad (34)$$
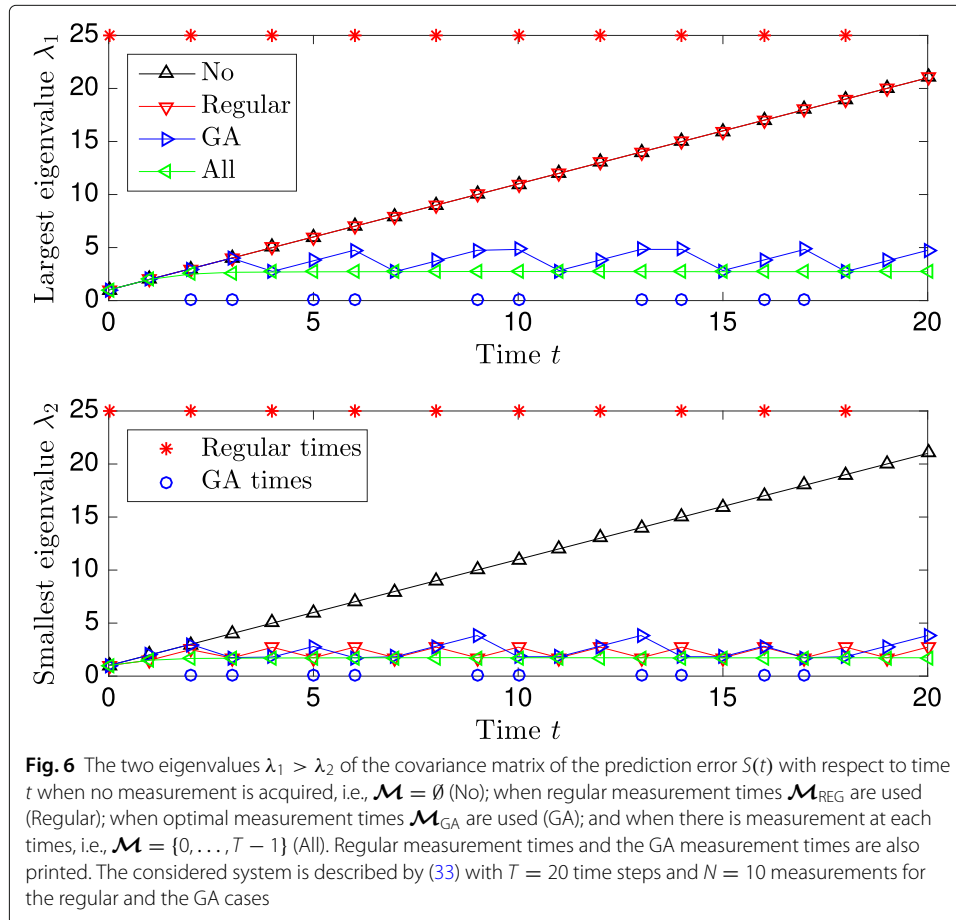
From definition (14), a regular measurement set is $\mathcal{M}_{\mathrm{REG}} = \{0, 2, \ldots, 18\}$, which leads to the following observability matrix

$$\mathcal{O}_T(A, C, \mathcal{M}_{\mathrm{REG}}) = \begin{pmatrix} 1 & 0 & -1 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}^\top. \qquad (35)$$

With such regular set of measurement times, the observability matrix is rank deficient. That means that even in the absence of any noise, the state can not be completely determined.

Bearing in mind that the trace of a matrix is the sum of its eigenvalue, then the objective function of problem (15) can be rewritten $\sum_{t=1}^{T}(\lambda_1(t) + \lambda_2(t))/T$. Figure 6, presents the eigenvalues $\lambda_1(t) > \lambda_2(t)$ of the prediction error covariance matrix $S(t) = BP(t|t-1)B^\top$ with respect to time. They are presented for regular measurements $\mathcal{M}_{\mathrm{REG}}$, for the optimal measurements found by the genetic algorithm but also when a measurement is acquired at each time, i.e., $\mathcal{M} = \{0, \ldots, T-1\}$ and when no measurement is acquired, i.e., $\mathcal{M} = \emptyset$. In addition, the measurement times are printed for both the regular and the intermittent cases.

One can observe that the largest eigenvalue λ_1 has the same evolution with regular measurements as when no measurement are used. This shows that the regular measurements

Fig. 6 The two eigenvalues $\lambda_1 > \lambda_2$ of the covariance matrix of the prediction error $S(t)$ with respect to time t when no measurement is acquired, i.e., $\mathcal{M} = \emptyset$ (No); when regular measurement times $\mathcal{M}_{\mathrm{REG}}$ are used (Regular); when optimal measurement times $\mathcal{M}_{\mathrm{GA}}$ are used (GA); and when there is measurement at each times, i.e., $\mathcal{M} = \{0, \ldots, T-1\}$ (All). Regular measurement times and the GA measurement times are also printed. The considered system is described by (33) with $T = 20$ time steps and $N = 10$ measurements for the regular and the GA cases

have no effect on the largest eigenvalue λ_1. This observation is compatible with the rank deficiency of the observability matrix (35). On the contrary, in the case of the measurement times found by the genetic algorithm, both eigenvalues remain significantly smaller than the ones without measurements. In addition, the measurement times selected by the GA are composed of five pairs of successive measurements. The structure of the observability matrix (34) ensures that each line is non zero.

3.4 Continuous-time example

This subsection presents an example of the continuous-time case presented in Section 2.5. More precisely, we are interested in the impact of the number of discretization steps T (or equivalently, the impact of the discretization step size δ) on the obtained solution.

We want to solve the problem defined by Eqs. (19) and (20) for the spring-mass system subject to a random force w^c,

$$\frac{d^2 x}{d\tau^2}(\tau) = -\frac{k}{m} x(\tau) + \frac{1}{m} w^c(\tau) \text{ with } w^c(\tau) \sim \mathcal{N}(0, 1), \tag{36}$$

$$y(\tau) = x(\tau), \tag{37}$$

$$z(\tau_k) = x(\tau_k) + v(\tau_k) \text{ with } v(\tau_k) \sim \mathcal{N}(0, 1), \tag{38}$$

$$x(0) \sim \mathcal{N}(0, 1), \; \frac{dx}{d\tau}(0) \sim \mathcal{N}(0, 1), \tag{39}$$

where m is the mass, expressed in $[kg]$ and k is the stiffness of the spring, expressed in $[N/m]$. $x(\tau)$ is the elongation of the mass (in $[m]$) at time τ (in $[s]$). In the formalism of (20), it corresponds to

$$A^c = \begin{pmatrix} 0 & 1 \\ -k/m & 0 \end{pmatrix}, \; B^c = C = (1 \quad 0), \; b^c = \bar{x}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; G^c = \begin{pmatrix} 0 & 0 \\ 0 & 1/m \end{pmatrix},$$

$$Q^c = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \; \bar{P}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \; d = 0, \; R = 1.$$
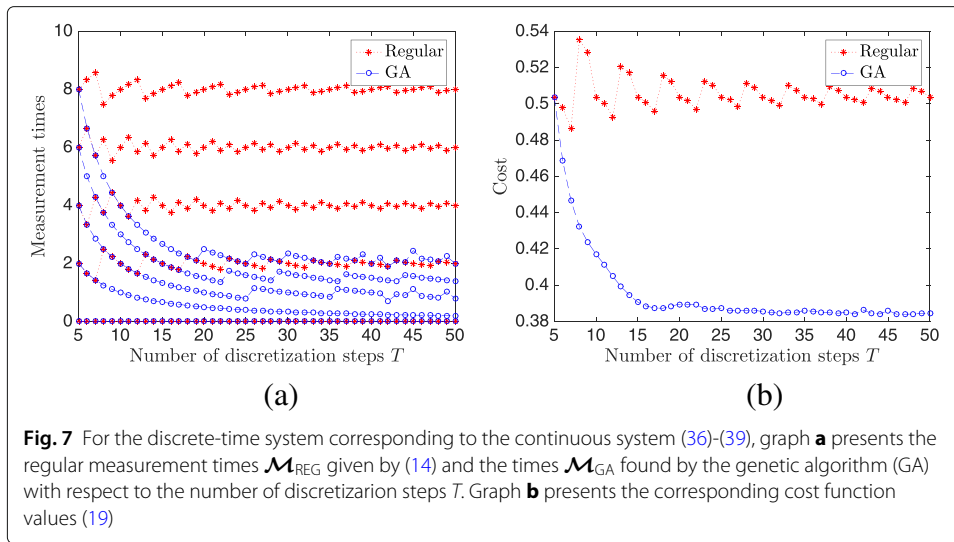
The corresponding discrete system can be obtained thanks to (21) and (22) and, in the particular case $k = m$, it gives Eq. (31) and

$$Q = \int_0^\delta \frac{1}{m} \begin{pmatrix} \sin^2 \tau & \sin \tau \cos \tau \\ -\sin \tau \cos \tau & \cos^2 \tau \end{pmatrix} d\tau = \frac{1}{2m} \begin{pmatrix} \delta - \sin \delta \cos \delta & \sin^2 \delta \\ -\sin^2 \delta & \delta + \sin \delta \cos \delta \end{pmatrix}.$$

We set $m = 40 \, [kg]$ and $k = 40 \, [N/m]$ and finally obtain Eqs. (31) and (32), which had been anticipated at the beginning of Section 3.

In this subsection, the time horizon is $\bar{\tau} = 100 \, [s]$ and the number of measurements is fixed to $N = 5$.

The measurement times are computed by solving the discretized problem using the GA. Then, the quality of the obtained solution is assessed by looking at the cost function of the continuous-time problem (19). It is done thanks to the continuous-discrete Kalman filter [45]. The results are presented in Fig. 7. Figure 7a presents the measurement times found by the GA and the measurement times for the regular Kalman predictor.

Fig. 7 For the discrete-time system corresponding to the continuous system (36)-(39), graph **a** presents the regular measurement times $\mathcal{M}_{REG}$ given by (14) and the times $\mathcal{M}_{GA}$ found by the genetic algorithm (GA) with respect to the number of discretizarion steps *T*. Graph **b** presents the corresponding cost function values (19)

Firstly, the measurement times found by the GA seem to converge when the number of discretization steps *T* increases (or equivalently, when δ decreases). In addition, they are more concentrated at the beginning. Furthermore, one can observe that even for the regular Kalman predictor, the measurement times vary with the number of discretization steps *T*. This is due to the fact that the measurement times are constraints to belong in the discrete-time set. Algebraically, it is the Round operator in Eq. (14). When the number of discretization steps *T* increases, the rounding effect decreases.

Figure 7b depicts the corresponding cost values for the continuous-time problem (19). It shows that the cost for optimal intermittent measurements is systematically smaller than in the regular case. This should be expected because the regular measurement times are in the admissible domain of problem (19)–(20). If the regular measurements had smaller cost function value, the GA would probably have selected it. The roughly decreasing and converging behavior of the cost function with respect to the number of discretization steps shows that our method can be used for continuous-time systems using an appropriate discretization scheme.

3.5 Impact of stability, process noise variance and measurement noise variance
In this section, the impact of the system's stability, process noise variance Q and measurement noise variance R on the solutions are studied in the one-dimensional case. More precisely, we compare the intermittent Kalman predictor to the regular Kalman predictor for different values of Q and R. The experiments are restricted to one-dimensional system, for the stable case $|A| < 1$, the marginally stable case $|A| = 1$ and the unstable case $|A| > 1$. Finally, we comment on the performance in the different cases.

The studied system is a simple version of (1)–(4) where all quantities are one-dimensional and with $b = 0$, $G = 1$, $B = 1$, $C = 1$, $d = 0$. This problem is studied for all $A \in \{1/2, 1, 2\}$ and for Q and R in the range $[\,0.01, 100\,]$. Results are presented in Fig. 8. For graphs a–c, $A = 1/2$; for graphs d–f, $A = 1$, and for graphs g–i, $A = 2$. In graphs a, d, and g, Q and R vary. In graphs b, e, and h, $R = 50$ and Q varies. Finally, in graphs c, f, and i, $Q = 50$ and R varies.

Fig. 8 Cost function of (15) for both regular measurement times and times found by the GA for varying process noise variance Q and/or measurement noise variance R. It corresponds to a one-dimensional version of model (1)-(4) with $b = 0, G = 1, B = 1, C = 1, d = 0$. For **a**–**c**, $A = 1/2$; for **d**–**f**, $A = 1$; for **g**–**i**, $A = 2$. In **a**, **d**, and **g**, Q and R vary. In **b**, **e**, and **h**, $R = 50$ and Q varies. In **c**, **f**, and **i**, $Q = 50$ and R varies
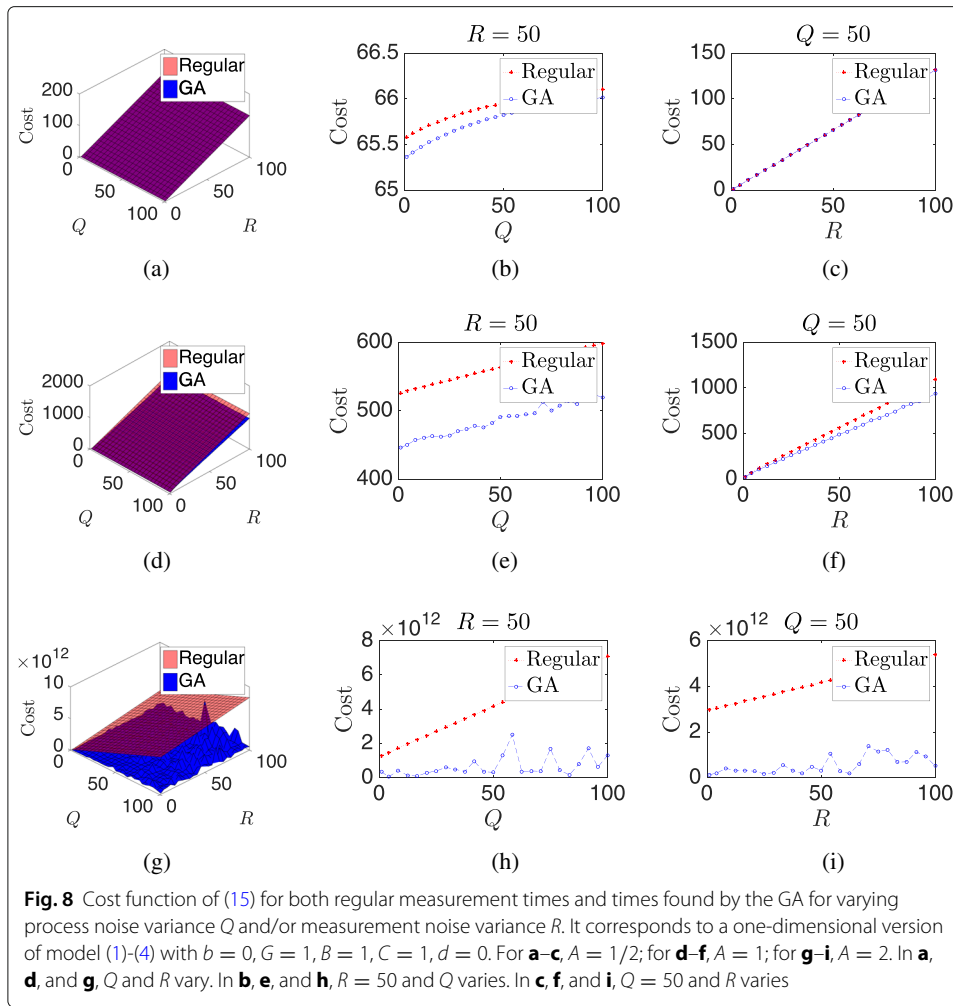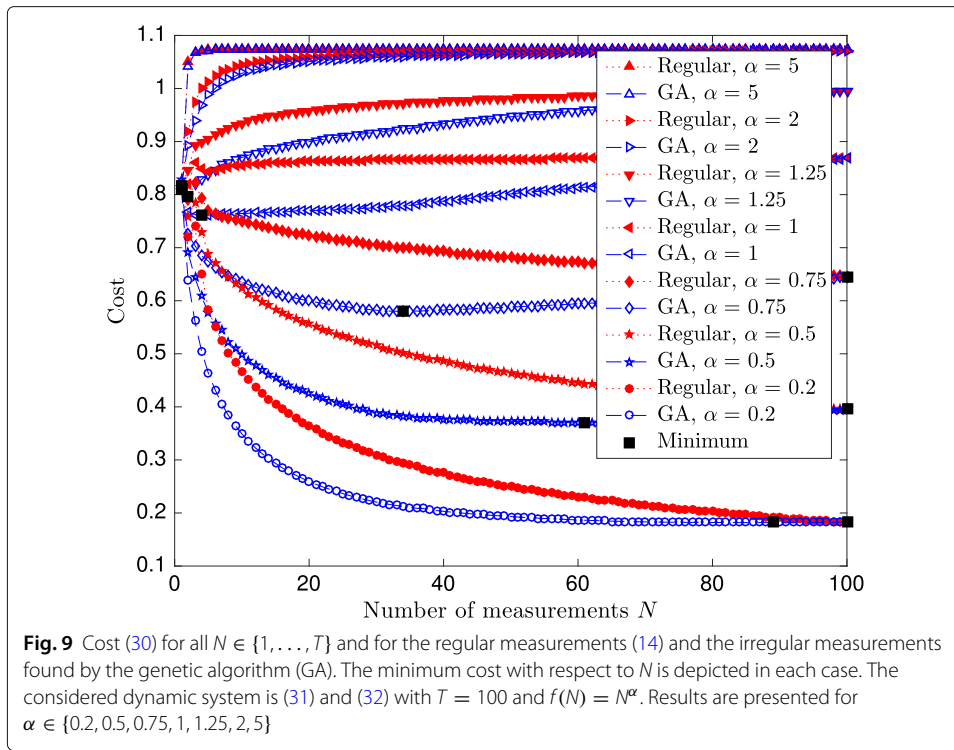
In all cases, the optimal intermittent Kalman predictor gives a smaller cost than with regular measurements. For the stable system, i.e., $A = 1/2$ in graphs a–c, the intermittent Kalman predictor gives results very similar to the regular Kalman predictor. In other cases, i.e., $A \in \{1, 2\}$ in graphs d–i, the improvement due to the use of intermittent measurements instead of regular ones is greater. This improvement increases when R increases. The unstable case, i.e., $A = 2$ in graphs g–i, is the one in which the improvement is the most significant. This improvement increases when either Q or R increase.

3.6 Quality or quantity of measurements: an example

In this section, we study the compromise between few precise measurements and many more noisy measurements presented in Section 2.7. The problem (30) is studied for the dynamical system (31) and (32) for $T = 100$ time steps. The measurement noise variance is related to the number of measurements as $R = f(N) = N^\alpha$ for different $\alpha > 0$. The larger the α, the more an additional measurement increases the measurement noise.

Figure 9 presents the cost (30) for all $N \in \{1, \dots, T\}$ and for the regular measurements (14) and the irregular measurements found by the genetic algorithm. In each case, the minimum is indicated. Results are presented for all values of $\alpha \in$

Fig. 9 Cost (30) for all $N \in \{1, \ldots, T\}$ and for the regular measurements (14) and the irregular measurements found by the genetic algorithm (GA). The minimum cost with respect to N is depicted in each case. The considered dynamic system is (31) and (32) with $T = 100$ and $f(N) = N^{\alpha}$. Results are presented for $\alpha \in \{0.2, 0.5, 0.75, 1, 1.25, 2, 5\}$

$\{0.2, 0.5, 0.75, 1, 1.25, 2, 5\}$. Firstly, observe that the irregular measurements found by the GA give smaller cost than regular measurements for all α and all N.

For the tested values of α, the optimal number of measurement times for the regular measurements is $N = 1$ when $\alpha \geq 1.25$ and $N = 20 = T$ when $\alpha \leq 1$. In comparison, for the intermittent measurements the optimal number of measurements is $N = 1$ only for $\alpha \geq 2$ and increases progressively when α decreases. These observations illustrate that when measurements are expensive, i.e., α is high, only few measurements will be acquire. Conversely, when measurements are cheap, i.e., α is small, many measurements will be acquired.

Another observation is that the cost difference between the regular measurements and the intermittent varies with N. When $N = T$, the regular and intermittent measurements sets are the same because the only admissible set is the set with all measurement times, i.e., $\mathcal{M} = \{0, \ldots, T-1\}$. More generally, the size of the admissible domain of the optimization problem (30) is smaller when N is close to 1 or T. Using intermittent measurements is most justified when N is between 10 and 60.

4   Conclusion

This paper addresses the problem of selecting the optimal measurement times for Kalman prediction over a finite time horizon. A random trial algorithm and three variants of genetic algorithms are proposed to solve it and the genetic algorithm that implements a count preserving crossover is shown to outperform the others. The tracking performances are extensively demonstrated on a numerical example, showing an improvement in 64% of the cases in comparison with regularly spaced measurement times. Then, the case of continuous-time systems is considered through a spring-mass system. The

solution is shown to converge with respect to the discretization step size, making our method suitable for both discrete-time and continuous-time systems. Finally, the influence of noise variance is studied in the one-dimensional case, showing that the improvement is the most significant in the unstable case and when the process noise becomes important.

Thanks to duality, we have shown that the problem of selecting optimal control times for a LQR can be reduced to the proposed problem. Then, we studied the optimal compromise between a lot of noisier measurements and less more precise measurements.

Further work will consider extending this work to the infinite time-horizon case and to the non-linear case. A reinforcement learning approach will be explored. From an experimental point of view, we will apply our method to the problem of tumor tracking from X-ray images, as suggested in [34].

Abbreviations
GA: Genetic algorithm; LQR: Linear quadratic regulator; RT: Random trial; SC: Shuffle crossover; RC: Replace crossover; CPC: Count preserving crossover

Authors' contributions
All the concepts in the paper have been discussed collectively by all the authors. A.A. implemented all the simulations shown in the paper. All the authors contributed to the writing of the paper. The author(s) read and approved the final manuscript.

Availability of data and materials
Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Competing interests
The authors declare that they have no competing interests.

References
1. R. E. Kalman, A new approach to linear filtering and prediction problems. J. Basic Eng. **82**(1), 35–45 (1960)
2. S.-K. Weng, C.-M. Kuo, S.-K. Tu, Video object tracking using adaptive kalman filter. J. Vis. Commun. Image Represent. **17**(6), 1190–1208 (2006)
3. S. Chen, Kalman filter for robot vision: a survey. IEEE Trans. Ind. Electron. **59**(11), 4409–4420 (2011)
4. Y. Yang, W.-g. Gao, X.-d. Zhang, Robust kalman filtering with constraints: a case study for integrated navigation. J. Geodesy. **84**(6), 373–381 (2010)
5. A. Assa, F. Janabi-Sharifi, K. N. Plataniotis, Sample-based adaptive kalman filtering for accurate camera pose tracking. Neurocomputing. **333**, 307–318 (2019)
6. D. A. Hage, M. H. Conde, O. Loffeld, Sparse signal recovery via kalman-filter-bases l1 minimization. Signal Process. **171**, 107487 (2020)
7. A. Fatehi, B. Huang, Kalman filtering approach to multi-rate information fusion in the presence of irregular sampling rate and variable measurement delay. J. Process Control. **53**, 15–25 (2017)
8. G. C. Sharp, S. B. Jiang, S. Shimizu, H. Shirato, Prediction of respiratory tumour motion for real-time image-guided radiotherapy. Phys. Med. Biol. **49**(3), 425 (2004)
9. S. Khuri, T. Bäck, J. Heitkötter, in *ACM Conference on Computer Science*, An evolutionary approach to combinatorial optimization problems (Association for Computing Machinery, New York, 1994), pp. 66–73
10. B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, S. S. Sastry, Kalman filtering with intermittent observations. IEEE Trans. Autom. Control. **49**(9), 1453–1464 (2004)
11. M. Kordestani, M. Dehghani, B. Moshiri, M. Saif, A new fusion estimation method for multi-rate multi-sensor systems with missing measurements. IEEE Access. **8**, 47522–47532 (2020)
12. Z. Cheng, H. Ren, B. Zhang, R. Lu, Distributed kalman filter for large-scale power systems with state inequality constraints. IEEE Trans. Ind. Electron. **68**, 6238–6247 (2020)
13. C. Yang, Z. Deng, Guaranteed cost robust weighted measurement fusion kalman estimators with uncertain noise variances and missing measurements. IEEE Sensors J. **16**(14), 5817–5825 (2016)
14. C. Yang, J. Ji, S. Chen, X. Wang, in *2020 39th Chinese Control Conference (CCC)*, Guaranteed cost robust centralized fusion kalman estimators with uncertain noise variances and missing measurements (IEEE, New York, 2020), pp. 2752–2757

15. H. S. Karimi, B. Natarajan, Kalman filtered compressive sensing with intermittent observations. Signal Process. **163**, 49–58 (2019)
16. K. J. Rutledge, S. Z. Yong, N. Ozay, Optimization-based design of bounded-error estimators robust to missing data. IFAC-PapersOnLine. **51**(16), 157–162 (2018)
17. M. Kordestani, A. Chibakhsh, M. Saif, in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, A control oriented cyber-secure strategy based on multiple sensor fusion (IEEE, New York, 2019), pp. 1875–1881
18. V. Gupta, T. H. Chung, B. Hassibi, R. M. Murray, On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage. Automatica. **42**(2), 251–260 (2006)
19. J. Le Ny, E. Feron, M. A. Dahleh, Scheduling continuous-time kalman filters. IEEE Trans. Autom. Control. **56**(6), 1381–1394 (2010)
20. Y. Mo, E. Garone, B. Sinopoli, On infinite-horizon sensor scheduling. Syst. Control Lett. **67**, 65–70 (2014)
21. L. Shi, P. Cheng, J. Chen, Optimal periodic sensor scheduling with limited resources. IEEE Trans. Autom. Control. **56**(9), 2190–2195 (2011)
22. L. Zhao, W. Zhang, J. Hu, A. Abate, C. J. Tomlin, On the optimal solutions of the infinite-horizon linear sensor scheduling problem. EEE Trans. Autom. Control. **59**(10), 2825–2830 (2014)
23. S. Liu, M. Fardad, E. Masazade, P. K. Varshney, Optimal periodic sensor scheduling in networks of dynamical systems. IEEE Trans. Signal Process. **62**(12), 3055–3068 (2014)
24. J.-S. Ha, H.-L. Choi, On periodic optimal solutions of persistent sensor planning for continuous-time linear systems. Automatica. **99**, 138–148 (2019)
25. L. Meier, J. Peschon, R. Dressler, Optimal control of measurement subsystems. IEEE Trans. Autom. Control. **12**(5), 528–536 (1967)
26. M. P. Vitus, W. Zhang, A. Abate, J. Hu, C. J. Tomlin, On efficient sensor scheduling for linear dynamical systems. Automatica. **48**(10), 2482–2493 (2012)
27. M. Athans, On the determination of optimal costly measurement strategies for linear stochastic systems. Automatica. **8**(4), 397–412 (1972)
28. H. J. Lee, K. L. Teo, A. E. Lim, Sensor scheduling in continuous time. Automatica. **37**(12), 2017–2023 (2001)
29. S. F. Woon, V. Rehbock, R. Loxton, Global optimization method for continuous-time sensor scheduling. Nonlinear Dyn. Syst. Theory. **10**, 175–188 (2010)
30. A. Sano, M. Terao, Measurement optimization in optimal process control. Automatica. **6**(5), 705–714 (1970)
31. S. Tanaka, T. Okita, Optimal timing of observations for state estimation in a one-dimensional linear continuous system. Automatica. **21**(3), 329–331 (1985)
32. A. Aksenov, P.-O. Amblard, O. Michel, C. Jutten, in *International Conference on Latent Variable Analysis and Signal Separation*, Optimal measurement times for observing a brownian motion over a finite period using a kalman filter (Springer, Gewerbestrasse, 2017), pp. 509–518
33. A. Aksenov, P.-O. Amblard, O. Michel, C. Jutten, Optimal measurement times for a small number of measures of a brownian motion over a finite period. arXiv preprint arXiv:1902.06126 (2019). https://arxiv.org/abs/1902.06126. Accessed: 16 Feb 2019
34. A. Aspeel, D. Dasnoy, R. M. Jungers, B. Macq, Optimal intermittent measurements for tumor tracking in x-ray guided radiotherapy. Int. Soc. Opt. Photonics. **10951**, 109510C (2019)
35. J. L. Fleiss, On the distribution of a linear combination of independent chi squares. J. Am. Stat. Assoc. **66**(333), 142–144 (1971)
36. J. Bausch, On the efficient calculation of a linear combination of chi-square random variables with an application in counting string vacua. J. Phys. A Math. Theor. **46**(50), 505202 (2013)
37. M. Mitchell, *An Introduction to Genetic Algorithms*. (MIT press, 1998)
38. A. Umbarkar, P. Sheth, Crossover operators in genetic algorithms: A review. ICTACT J. Soft Comput. **6**(1), 1083–1092 (2015)
39. S. J. Hartley, A. H. Konstam, in *Proceedings of the 1993 ACM Conference on Computer Science*, Using genetic algorithms to generate steiner triple systems (ACM, New York, 1993), pp. 366–371
40. B. Oksendal, *Stochastic Differential Equations: an Introduction with Applications*. (Springer, Gewerbestrasse, 2013)
41. P. Dorato, A. Levis, Optimal linear regulators: The discrete-time case. IEEE Trans. Autom. Control. **16**(6), 613–620 (1971)
42. X. Song, X. Yan, X. Li, Survey of duality between linear quadratic regulation and linear estimation problems for discrete-time systems. Adv. Differ. Equ. **2019**(1), 90 (2019)
43. L. Shi, Y. Yuan, J. Chen, Finite horizon lqr control with limited controller-system communication. IEEE Trans. Autom. Control. **58**(7), 1835–1841 (2012)
44. R. M. Jungers, A. Kundu, W. Heemels, Observability and controllability analysis of linear systems subject to data losses. IEEE Trans. Autom. Control. **63**(10), 3361–3376 (2017)
45. F. L. Lewis, L. Xie, D. Popa, *Optimal and Robust Estimation: with an Introduction to Stochastic Control Theory*. (CRC press, London, 2017)

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.