Automatica 90 (2018) 130–137

Contents lists available at ScienceDirect

Automatica

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

Brief paper

The generalized cross validation filter☆

Giulio Bottegal [a], Gianluigi Pillonetto [b,*]

[a] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[b] Department of Information Engineering, University of Padova, Padova, Italy

ARTICLE INFO

Article history:
Received 1 September 2016
Received in revised form 19 September 2017
Accepted 28 November 2017
Available online 15 February 2018

Keywords:
Kalman filtering
Generalized cross-validation
On-line system identification
Inverse problems
Regularization
Smoothness parameter
Splines

ABSTRACT

Generalized cross validation (GCV) is one of the most important approaches used to estimate parameters in the context of inverse problems and regularization techniques. A notable example is the determination of the smoothness parameter in splines. When the data are generated by a state space model, like in the spline case, efficient algorithms are available to evaluate the GCV score with complexity that scales linearly in the data set size. However, these methods are not amenable to on-line applications since they rely on forward and backward recursions. Hence, if the objective has been evaluated at time $t-1$ and new data arrive at time t, then $O(t)$ operations are needed to update the GCV score. In this paper we instead show that the update cost is $O(1)$, thus paving the way to the on-line use of GCV. This result is obtained by deriving the novel *GCV filter* which extends the classical Kalman filter equations to efficiently propagate the GCV score over time. We also illustrate applications of the new filter in the context of state estimation and on-line regularized linear system identification.

1. Introduction

Linear state space models assume the form

$$x_{k+1} = A_k x_k + \omega_k$$
$$y_k = C_k x_k + e_k$$

where x_k is the state at instant k, y_k is the output, while ω_k and e_k are random noises. The matrices A_k and C_k regulate the state transition and the observation model at instant k. This kind of models plays a central role in the analysis and design of discrete-time systems (Kalman, 1960). Applications abound and include tracking, navigation and biomedicine.

In *on-line state estimation*, the problem is the reconstruction of the values of x_k from measurements of y_k collected over time. When the matrices A_k and C_k and the noises covariances are known, the optimal linear estimates are efficiently returned by the classical Kalman filter (Anderson & Moore, 1979). However, in many circumstances there can be unknown model parameters that also need to be inferred from data in an on-line manner, e.g. variance components or entries of the transition/observation matrices. One can interpret such parameters as additional states. Then, the extended Kalman filter (Jazwinski, 1970) or more sophisticated stochastic techniques, such as particle filters and Markov chain Monte Carlo (Andrieu, Doucet, & Holenstein, 2010; Frigola, Lindsten, Schon, & Rasmussen, 2013; Gilks, Richardson, & Spiegelhalter, 1996; Ninness & Henriksen, 2010), can be used to track the filtered posterior. Another technique consists of propagating the marginal likelihood of the unknown parameters via a bank of filters (Anderson & Moore, 1979, Ch. 10). In this paper, we will show that another viable alternative is the use of an approach known in the literature as generalized cross validation (GCV) (Golub, Heath, & Wahba, 1979).

In the literature of statistics and inverse problems, GCV is widely used in off-line contexts to estimate unknown parameters entering regularized estimators (Bertero, 1989; Tarantola, 2005; Wahba, 1990). This approach was initially used to tune the smoothness parameter in ridge regression and smoothing splines (Golub et al., 1979; Hoerl & Kennard, 1970; Rice, 1986). GCV is now also popular in machine learning, used to improve the generalization capability of regularized kernel-based approaches (Evgeniou, Pontil, & Poggio, 2000; Schölkopf & Smola, 2001), such as regularization networks, which contain spline regression as special case (Girosi, Jones, & Poggio., 1995; Poggio & Girosi, 1990).

To introduce GCV in our state space context, we first recall that smoothing splines are closely linked to state space models of

☆ This research has been partially supported by the MIUR FIRB project RBFR12M3AC-Learning meets time: a new computational approach to learning in dynamic systems and by the Progetto di Ateneo CPDA147754/14-New statistical learning approach for multi-agents adaptive estimation and coverage control. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Wei Xing Zheng under the direction of Editor Torsten Söderström.

* Corresponding author.
E-mail addresses: g.bottegal@tue.nl (G. Bottegal), giapi@dei.unipd.it (G. Pillonetto).

https://doi.org/10.1016/j.automatica.2017.12.054
0005-1098/

m-fold integrated Wiener processes (Pillonetto & Chiuso, 2009); then it appears natural to extend GCV to general state space models. To this end, assume that measurements y_k have been collected up to instant t and stacked in the vector Y_t. Denote with $\hat{Y}_t$ the vector containing the optimal linear output estimate[1] and use H_t to denote the *influence matrix* satisfying

$$\hat{Y}_t = H_t Y_t.$$

Then, the parameter estimates achieved by GCV minimize

$$\text{GCV}_t = \frac{S_t}{t(1-\delta_t/t)^2}, \quad (2)$$

where S_t is the sum of squared residuals, i.e.

$$S_t = \|\hat{Y}_t - Y_t\|^2,$$

and δ_t are the *degrees of freedom* given by the trace of H_t, i.e.

$$\delta_t = \text{Tr}(H_t).$$

In the objective (2), the term S_t accounts for the goodness of fit while δ_t assumes values on $[0, t]$ and measures model complexity. In fact, in nonparametric regularized estimation, the degrees of freedom δ_t can be seen as the counterpart of the number of parameters entering a parametric model (Hastie, Tibshirani, & Friedman, 2001; MacKay, 1992; Pillonetto & Chiuso, 2015).

GCV is supported by important asymptotic results. Also, for finite data set size it turns often out a good approximation of the output mean squared error (Craven & Wahba, 1979). It is worth stressing that such properties have been derived without postulating the correctness of the prior models describing the output data (Wahba, 1983, 1985). In a system identification perspective, this means that GCV can compensate for possible modeling mismatch affecting the state space description (Ljung, 2000).

Despite these nice features, the use of GCV within the control community appears limited, in particular in on-line contexts. One important reason is the following one. For state space models, there exist efficient algorithms which, for a given parameter vector, return its GCV score with $O(t)$ operations (Ansley & Kohn, 1987; Kohn & Ansley, 1989), see also De Nicolao, Trecate, and Sparacino (2000), Hutchinson and De Hoog (1985) and Silverman (1985) for procedures dedicated to smoothing splines. But all of these techniques are not suited to on-line computations since they involve forward and backward recursions. Hence, if GCV_{t-1} is available and new data arrive at time t, other $O(t)$ operations are needed to achieve GCV_t. In this paper, we will instead show that the update cost is $O(1)$, thus paving the way to a more pervasive on-line use of GCV. This result is obtained by deriving the novel *GCV filter* which consists of an extension of the classical Kalman equations. Thanks to it, one can run a bank of filters (possibly in parallel) to efficiently propagate GCV over a grid of parameter values. This makes the proposed GCV filter particularly suitable for applications where a measurement model admits a state space description with dynamics depending on few parameters, see e.g. the next section for an application in numerical differentiation. In this framework, an implementation of the GCV filter via a bank of parallel filters turns out computationally attractive.

The paper is organized as follows. In Section 2, first some additional notation is introduced. Then, the GCV filter is presented. Its asymptotic properties are then discussed in Section 3. In Section 4 we illustrate some applications, including also smoothing splines and on-line regularized linear system identification with the stable spline kernel used as stochastic model for the impulse response (Pillonetto & De Nicolao, 2010; Pillonetto, Dinuzzo, Chen, Nicolao, & Ljung, 2014). Conclusions end the paper while the correctness of the GCV filter is shown in Appendix.

[1] The components of $\hat{Y}_t$ are thus given by $C\hat{x}_{k|t}$, where the smoothed state $\hat{x}_{k|t}$ can be obtained for any t with $O(t)$ operations by a fixed-interval Kalman smoothing filter (Ljung & Kailath, 1976; Rauch, Tung, & Striebel, 1965).

2. The GCV filter

2.1. State space model

First, we provide full details about our measurements model. We use $x \sim (a, b)$ to denote a random vector x with mean a and covariance matrix b. Then, our state space model is defined by

$$x_{k+1} = A_k x_k + \omega_k \quad (3a)$$
$$y_k = C_k x_k + e_k, \quad k = 1, 2, \ldots \quad (3b)$$
$$x_1 \sim (\mu, P_0) \quad (3c)$$
$$\omega_k \sim (0, Q_k) \quad (3d)$$
$$e_k \sim (0, \gamma) \quad (3e)$$

where the initial condition x_1 and all the noises $\{\omega_k, e_k\}_{k=1,2,\ldots}$ are mutually uncorrelated. We do not specify any particular distribution for these variables, since the GCV score does not depend on the particular noise distribution.[2] If x_1, ω_k, e_k are Gaussian, then the Kalman filter provides the optimal state estimate in the mean-square sense. In the other cases, the Kalman filter corresponds to the best linear state estimator (Anderson & Moore, 1979). In addition, just to simplify notation the measurements y_k are assumed scalar, so that γ represents the noise variance.

We assume that some of the parameters in (3) may be unknown, or could enter A_k, B_k, Q_k and P_0; however, we do not stress this possible dependence to make the formulas more readable. The matrix P_0 is assumed to be independent of γ. Such parameter is typically unknown, being connected to the ratio between the measurement noise variance and the variance of the driving noise. It corresponds to the regularization parameter in the smoothing-splines context described in the example below.

Example 1 (*Smoothing Splines (Pillonetto & Saccomani, 2006)*). Function estimation and numerical differentiation are often required in various applications. These include also input reconstruction in nonlinear dynamic systems as described e.g. in Pillonetto and Saccomani (2006). Assume that one is interested in determining the first m derivatives of a continuous-time signal measured with non-uniform sampling periods T_k. Modeling the signal as an m-th fold integrated Wiener process one obtains the stochastic interpretation of the mth order smoothing splines (Wahba, 1990). In particular, one can use (3) to represent the signal dynamics as follows

$$A_k = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ T_k & 1 & 0 & \ldots & 0 \\ \frac{T_k^2}{2} & T_k & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \frac{T_k^m}{m!} & \frac{T_k^{m-1}}{(m-1)!} & \ldots & T_k & 1 \end{pmatrix}, \quad C_k = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}^T,$$

$$[Q_k]_{ij} = \frac{T_k^{i+j-1}}{(i-1)!(j-1)!(i+j-1)}.$$

Such model depends on the measurement noise variance γ, making this application particularly suited for the GCV filter.

2.2. The GCV filter

The GCV filter equations are now reported. Below, $\hat{x}_k$ denotes the optimal linear one-step ahead state prediction having covari-

[2] Of course, GCV may result not effective if the noises are highly non-Gaussian. Different approaches, like particle filters, should instead be used if linear estimators perform poorly due e.g. to multimodal noise distributions.