LOCAL LIKELIHOOD ESTIMATION OF TIME-VARIANT HAWKES MODELS

Boris I. Godoy, Victor Solo, Jason Min

School of Electrical Eng. & Telecommunications
The University of New South Wales
Sydney, Australia
email: {b.godoy,v.solo}@unsw.edu.au

Syed Ahmed Pasha

Department of Electrical Engineering
Air University, Pakistan
s.pasha@mail.au.edu.pk

ABSTRACT

The Hawkes process is the workhorse of dynamic point process modelling - the point process version of the autoregression. It has been applied, for example, in high frequency finance, electricity price spike modelling and gene regulatory network modelling. But, in all these and other applications, it is assumed the parameters are time invariant. However, it is becoming clear that in many applications the parameters vary with time. Here, we develop for the first time, a very simple local likelihood approach to estimation of time-variant Hawkes processes. The new algorithm is tested on simulations and then applied to data from the Australian electricity market.

Index Terms**—** Point processes, time-variant parameters, system identification, local likelihood, EM algorithm.

1. INTRODUCTION

Point processes have gained increased attention from the system identification community in the last few decades. They have been used to describe random phenomena in seismology [1], finance [2, 3], neuroscience [4, 5], and genomics [6]. This wide range of applications was only possible due to the rigorous development of the concept of *intensity function*, see e.g [7, 8]. Further extension was done by [9] in the early 70s, where the likelihood function was derived based on the intensity function concept. For more applications of point processes, the reader can refer to e.g [10, chap.1], [11, 12, 13].

There are two assumptions that dominate the applied literature on point processes. One is no-memory or independent increments i.e. no dynamics. The other is that most of the models assume that the intensity function is parametrized by time-invariant parameters. We briefly discuss each of these in turn.

In electricity markets, the spot price can be subject to sudden changes due to unexpected increases in demand, unexpected shortfalls in supply and failures of transmission infrastructure, see [14]. Most of the literature in forecasting spikes for the electricity spot price is based on traditional autoregressive time-series models, Bernoulli and Poisson jump processes, and a variety of heavy-tailed error processes. They have in common that no memory (i.e. no dynamics) is considered in the modelling. However, accumulating evidence shows that there is an historical component (i.e. memory) in explaining spiking processes, see [15]. The most common approach to modelling point processes with memory is via Hawkes processes [16].

More recently two groups [17, 18] have combined history and time-variant behaviour by developing time-variant Hawkes process models for modelling electricity spot prices. The approach in [17] is based on so-called duration models and is a kind of inverse model of the stochastic intensity. The approach in [18] models the parameter time-variation as regressions on covariates. Both these approaches are valid but cumbersome, and a comparison with each of them goes beyond the purpose of this paper. Here, we develop a new and simple approach based on local likelihood.

Local likelihood [19] is a development of the sliding window approach. This is a very old idea which has been repeatedly rediscovered. Some history can be found in [19] but the approach goes back at least to the 1930s. However a thorough theoretical understanding of the method was not developed until the 1990s in the statistics literature [19, 20] under the name *local least squares* or *local polynomial modelling*. This theory has not, however, diffused outside the statistics literature and so a number of basic insights are ignored in other literatures.

The idea is to take a window of data and fit a time-invariant parameter model in the window. Then the window is moved along by a fixed amount and the modelling is repeated and so on. In this way time-variant parameter estimates are traced out. One of the fundamental insights from the statistical literature is that rectangular windows induce a kind of Gibbs-ringing (familiar from Fourier analysis) which produces (severe) estimator bias. Thus, only smooth windows should be used. Unfortunately, in a lot of applied literature this basic insight is unknown and rectangular windows are still very common.

To estimate the parameters of time-variant Hawkes models, at each time t, we optimize a windowed predictable likelihood (see e.g. [13, chap.7]) to obtain the maximum likelihood estimates for the unknown parameters.

We illustrate our new approach on both simulated data and real electricity price data.

The remainder of this paper is as follows: in section 2 we give background on modelling of Hawkes processes; In section 3 we pose the time-variant maximum likelihood estimation problem; in section 4, we give the updates for the parameters. Section 5 presents numerical examples. Finally, section 6 gives conclusions.

This work was supported by the Australian Research Council.

2. TIME-VARIANT HAWKES MODELS

2.1. Hawkes

The intensity function for the time-invariant Hawkes model - see e.g [16], can be converted into a time-variant intensity function as

$$\lambda(t) = c(t) + \int_{-\infty}^{t} g(t-u)dN_u \tag{1}$$

where $g(u) \geq 0$, $c > 0$, N_u: number of counts up to u, $t \in \mathbb{R}^+$, and (1) is an integral of the Riemann-Stieltjes type.

For practical use, the integral (1) must be truncated; the simplest truncation gives the *simple finite Hawkes process* (sFHP).

$$\lambda(t) = c(t) + \int_{0}^{t} g(t-u)dN_u \tag{2}$$

For the case of time-variant parameters, there are two common ways to represent the impulse response (IR) $h(u)$:

- Hawkes-exponential (HE)

$$g(t-u) = \sum_{j=1}^{p} \alpha_j(t)\phi_j(t-u), \quad \phi_j(u) = e^{-\beta_j u}\beta_j \tag{3}$$

  One of the advantages of this representation is that simulation is easy, but model fitting is difficult since β_j appears non linearly.

  There is an alternative, and it is the so-called HaL model[1], which is given in (4).

- Hawkes-Laguerre (HaL), first used by [21]

$$g(t-u) = \sum_{j=1}^{p} \alpha_j(t)\phi_j(t-u), \ \phi_j(u) = e^{-\beta_o u}\frac{(\beta_o u)^{j-1}}{(j-1)!}\beta_o \tag{4}$$

The advantage of this representation is that the stochastic intensity function is linear in the parameters (aside from β_o), but the likelihood function is nonlinear, thus requiring nonlinear optimization methods. The basis is valid for any $\beta_o > 0$, where $\tau_o = 1/\beta_o$ and τ_o : time constant of the system. This value β_o can, in principle, be chosen by the user. However, as we will show later in this work, this parameter can also be estimated by selecting a grid, and then optimizing a marginalized maximum likelihood problem.

Remark 1 *Note that our parametrization differs from that in [21]. Our scaling ensures the stability (or stationarity) condition discussed below has a very simple form.*

3. LOCAL MAXIMUM LIKELIHOOD ESTIMATION FOR HAWKES MODELS WITH TIME-VARIANT PARAMETERS

We introduce the 'predictable' likelihood function for time-invariant parameters given by [10]

$$\mathcal{L}(\theta) = \int_0^T \ln\lambda(t,\theta)dN_t - \int_0^T \lambda(t,\theta)dt + T \tag{5}$$

This function can be optimized, and an algorithm can be derived to estimate the parameters of the intensity function.

However, in this work, we are interested to optimize a likelihood function which is time-variant, that is, the parameter vector θ in (5) has the form $\theta = \theta(t)$.

One way to optimize this time-variant likelihood function is by using a local (or windowed) likelihood centred at time t. This is carried out by introducing a weight $K(\frac{u-t}{h})$. Here, the idea is to use a window (of nominated span and time point) to estimate the parameters that maximize the log-likelihood function within the window. We should note that the parameter estimates correspond to the middle of the window. The estimation algorithm continues when the time points are increased by one unit. The estimation algorithm finishes when the last data point is included in the window.

Different windows can be used. However, it is important to use smooth windows to minimize the effects caused by the Gibbs ringing phenomenon, which is achieved by creating a window with *rounded* edges, see e.g. [23]. Examples of such windows are given in (6) and (7).

$$K(v) = \begin{cases} \frac{3}{4}(1-|v|^2), & \text{if} |v| \leq 1, \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

$$K(v) = \begin{cases} \frac{5}{8}(1-|v|^4), & \text{if} |v| \leq 1, \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $K(v) \geq 0$ with $\int_{-\infty}^{\infty} K(v)dv = 1$

We also introduce a window width $h = h_o T$ where $h_o << 1$. Some trial and error is needed in order to choose h_o. For example, $h_o = 0.1$ may be a good starting point.

3.1. Maximum Likelihood (ML) equations at time t

The windowed log-likelihood function is essentially the point wise product of the log-likelihood and the window function in time domain, thus, the log-likelihood function centred at time t is given by:

$$\begin{aligned} \mathcal{L}_t(\theta) = & \int_0^T K(\frac{u-t}{h})\ln\lambda(u,\theta)dN_u \\ & - \int_0^T K(\frac{u-t}{h})(\lambda(u,\theta)-1)du \end{aligned}$$

where the subscript t in $\mathcal{L}_t$ refers to the time dependancy of the log-likelihood. Note that the idea here is that the parameter is fixed within the window. But the imposition of the kernel delivers a time-variant parameter estimate.

If we now differentiate $\mathcal{L}_t(\theta)$ with respect to θ, and define $v = (u-t)/h$, we obtain

$$\frac{\partial\mathcal{L}_t}{\partial\theta} = \int_0^T \frac{K(v)}{\lambda(u)}\frac{\partial\lambda(u,\theta)}{\partial\theta}dN_u - \int_0^T K(v)\frac{\partial\lambda(u,\theta)}{\partial\theta}du \tag{8}$$

For the time-variant HaL (tv-HaL) model in (4), we have that the intensity function can be written as:

$$\lambda(t,\theta) = \zeta^T(t)\theta \tag{9}$$

where $\theta = [c\ \alpha_1 \dots\ \alpha_p]^T$, $\zeta(t) = [1, x_1(t), \dots, x_p(t)]^T$, and

$$x_l(t) = \int_0^t \phi_l(t-u)dN_u = \sum_{j:T_j<t-} \phi_l(t-T_j). \tag{10}$$

[1] introduced in [21] much earlier than in the system identification community [22]

We can omit the dependance of λ on θ, and given that within the window $K(v)$, parameter θ is considered to be constant, the derivatives of $\lambda(t,\theta)$ take the form

$$\frac{\partial\lambda(t)}{\partial c}=1, \qquad \frac{\partial\lambda(t)}{\partial\alpha_l}=x_l(t) \tag{11}$$

Replacing (11) in (8), and setting to zero, we obtain

$$\begin{aligned} A_t &= \int_0^T K(\frac{u-t}{h})du \qquad (12)\\ &= \int_0^T K(\frac{u-t}{h})\frac{dN_u}{\lambda(u)} = \sum_{i=1}^{N_T}\frac{1}{\lambda(T_i)}K(\frac{T_i-t}{h}) \\ B_t^l &= \int_0^T K(\frac{u-t}{h})x_l(u)du = \int_0^T K(\frac{u-t}{h})\frac{x_l(u)}{\lambda(u)}dN_u \qquad (13)\\ &= \sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\frac{x_l(T_i)}{\lambda(T_i)} \end{aligned}$$

where N_T is the number of events in the interval $(0,T]$

4. THE EM ALGORITHM

The Expectation Maximization (EM) algorithm is an iterative algorithm that has been used to optimize the (log-) likelihood function. For this kind of processes, the EM algorithm can intuitively be derived from the ML equations (12) and (13), which can now be rewritten as:

$$1=\frac{1}{A_t}\sum_{1}^{N_T}K(\frac{T_i-t}{h})\frac{1}{\lambda(T_i)} \tag{14}$$

$$1=\frac{1}{B_t^l}\sum_{1}^{N_T}K(\frac{T_i-t}{h})\frac{x(T_i)}{\lambda(T_i)}. \tag{15}$$

Thus, from (14) and (15), we can find that the EM iterations at time t are given by:

$$c^{k+1}(t)=c^k(t)\frac{1}{A_t}\sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\frac{1}{\lambda_k(T_i)} \tag{16}$$

$$\alpha_l^{k+1}(t)=\alpha_l^k(t)\frac{1}{B_t^l}\sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\frac{x_l(T_i)}{\lambda_k(T_i)} \tag{17}$$

$$\lambda^k(u)=c^k(t)+\sum_{l=1}^{p}\alpha_l^k(t)x_l(u), \quad 0\le u\le T \tag{18}$$

These iterations ensure that both $\alpha_k(t)$ and $c_k(t)$ are positive. They can be computed on a grid of time points e.g., we can pick n, set $\delta=T/n$, then we have the grid $k\delta$, $k=1,\ldots,n$.

The integrals A_t and B_t can be pre-computed at each time t being approximated as Riemann sums on a similar (preferably finer) grid.

4.1. Likelihood at time t

Iterations should be monitored by plotting the log-likelihood, which is given by the following expression:

$$\begin{aligned} \mathcal{L}_t^{k+1} &= \sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\ln(\lambda^{k+1}(T_i)) \\ &\quad -\int_0^T K(\frac{u-t}{h})\lambda^{k+1}(u)du + A_t \\ &= \sum_{1}^{N_T}\ln(\lambda^{k+1}(T_i)) - c^{k+1}(t)A_t - \sum_{l=1}^{p}\alpha_l^{k+1}(t)B_t^l + A_t \end{aligned} \tag{19}$$

Substituting the updates (16) and (17) in (19), we have

$$\begin{aligned} \mathcal{L}_t^{k+1} &= \sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\ln(\lambda^{k+1}(T_i)) - c^k(t)\sum_{1}^{N_T}K(\frac{T_i-t}{h})\times \\ &\frac{1}{\lambda_k(T_i)} - \sum_{l=1}^{p}\alpha_l^k(t)\sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\frac{x_l(T_i)}{\lambda^k(T_i)} + A_t \\ &= \sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\ln(\lambda^{k+1}(T_i)) \\ &- \sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\frac{c^k(t)+\sum_{l=1}^{p}\alpha_l^k(t)x_l(T_i)}{\lambda^k(T_i)} + A_t \\ &= \sum_{i=1}^{N_T}K(\frac{T_i-t}{h})\ln(\lambda_{k+1}(T_i)) - \bar{A}_t + A_t \end{aligned} \tag{20}$$

where $\bar{A}_t := \sum_1^{N_t}K(\frac{T_i-t}{h})$ can be pre-computed.

5. EXAMPLES

We apply our algorithm to two cases. The first one is on simulated data, where we exactly know the parameters that generate the data; and the second application of the algorithm is on modelling of the spot price spikes in the Australian electricity market.

5.1. Simulated data

We consider the system given by the following intensity function:

$$\lambda(t)=c(t)+\alpha(t)\sum_{T_i<t}e^{-\beta(t-T_i)}\beta,$$

where $c(t)=c_0+c_1(1-\cos(\omega_0 t))$, $\alpha(t)=\alpha_0+\alpha_1(1-\cos(\omega_1 t))$ with $\omega_0=2\pi/T_0$, and $\omega_1=2\pi/T_1$. The value for T_0 and T_1 are chosen to be $T_0=T_1=0.5T$, where T is the time interval where the process occurs, and $c_o=0.2$, $c_1=0.5c_o$, $\alpha_o=0.1562$, $\alpha_1=0.5\alpha_o$.

To generate the data, we run the *thinning* algorithm [24], and to check we generated the correct data, we rescale it [25]. After rescaling the data, and compare it to data from a exponential distribution, we obtain a quantile-quantile (Q-Q) plot of the two distributions. For lack of space, this is not shown here. The results of a 100 replicates of the estimation are given below in Fig. 1. Here, we can see that

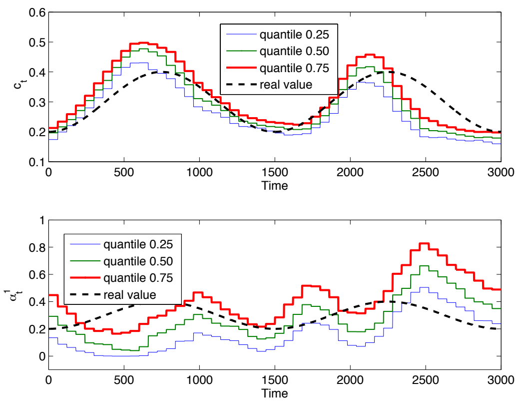

Fig. 1. Parameter estimation of time-variant Hawkes-Laguerre model. Upper plot shows the comparison of lower, median, and upper quartile with the real value of $c(t)$. Lower plot shows the comparison of lower, median, and upper quartile with the real value of $\alpha(t)$.

the algorithm can estimate the time-variant nature of the parameters. In this figure, we compare the lower quartile, the median, and the upper quartile with the real value for each of the parameters. The time-variant estimation of β_o is given in Fig. 2.

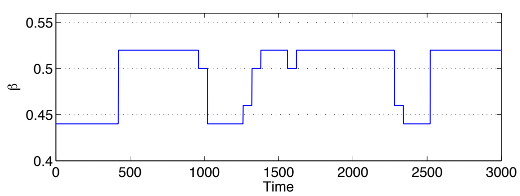

Fig. 2. Estimation of β_o

5.2. Spikes in electricity spot price data

We now use data for the spikes in Australian electricity spot price. In particular, we use data from the state of South Australia. Spikes are produced when the price of 1 MWh increases over a threshold previously defined. In particular, we define that threshold to be \$ AUD 300 per MWh. A helpful and interesting discussion on determining appropriate threshold levels can be found in [26].

The result of the time-variant parameter estimation of a first order model is given in Fig. 3. With these parameters, we calculate the QQ-plot of the estimated model. We then compare it to the QQ-plot obtained by estimating time-invariant parameters[2]. The parameters for a first order time-invariant model are given by: $\hat{c} = 1.21$, $\hat{\alpha}_1 = 0.7135$, and $\hat{\beta} = 66$. In Fig. 4, we show the comparison of both QQ- plots. As we observe, the time-variant case (Fig. 4 - right plot) overperforms the fitting of a time-invariant parameter estimation algorithm (Fig, 3 - left plot). The latter is an indication of the potential of the approach described in this paper.

[2]See e.g [27] for details about time-invariant parameter estimation in Hawkes models.

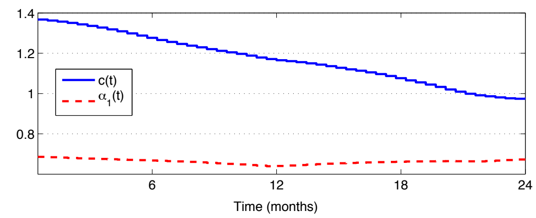

Fig. 3. Estimation of time-variant parameters for modelling spikes in the South Australian electricity spot price.

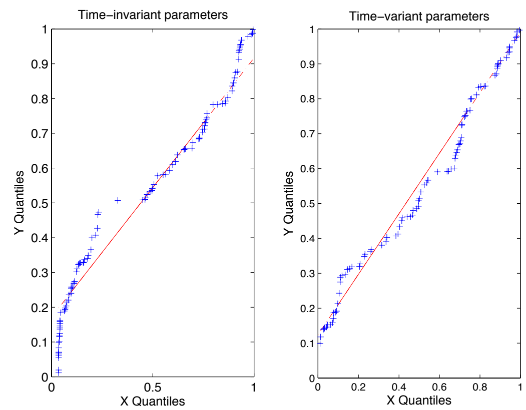

Fig. 4. Left: QQ-plot for time-invariant parameters. Right:QQ-plot for the time-variant parameters.

6. CONCLUSIONS

In this work, we have proposed a time-variant parameter estimation algorithm for Hawkes processes. To the best of the authors' knowledge, this is the first time such an algorithm is proposed for point processes.

Our algorithm works by selecting a window of the data, and estimating the parameters within that window, in which parameters are considered to be *constant*. The estimation process is repeated until the last data point is included in the window.

Simulation results have shown that the estimation algorithm performs well to recover the parameters of interest.

The proposed algorithm has also shown to have a comparative advantage for modelling spikes in the Australian electricity market with respect to our earlier time-invariant estimation algorithm. In particular, we have used data from the state of South Australia to obtain time-variant parameters. When comparing both kind of models, a better fit is obtained for the time-variant case. The use of more complex models to model the Australian electricity market is a topic of current research.

7. REFERENCES

[1] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Stat. Association*, vol. 83, no. 401, pp. 9–27, 1988.

[2] N. Hautsch, *Modelling irregularly spaced financial data*, MIT Press, Boston, MA, 1997.

[3] V. Solo and B. Godoy, "Time-varying vector Poisson processes with coincidences," in *40th IEEE Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, Brisbane, Australia, 2015.

[4] P. Dayan and L. Abbott, *Theoretical Neuroscience*, MIT Press, Cambridge, MA, 2001.

[5] F. Rieke, D. Warland, R. de Ruyter van Stevenink, and W. Bialek, *Spike: Exploring the neural code*, MIT Press, Boston, MA, 1997.

[6] S.A. Pasha and V. Solo, "Topology identification for sparse dynamic point process networks," in *53rd IEEE Conf. on Decision and Control (CDC)*, Los Angeles, CA, 2014.

[7] D.P. Gaver, "Random hazard in reliability problems," *Tecnometrics*, vol. 5, pp. 211–216, 1963.

[8] J.A. McFadden, "The entropy of a point process," *J. Soc. Indust. Appl. Math.*, vol. 13, pp. 988–994, 1965.

[9] I. Rubin, "Regular point processes and their detection," *IEEE Trans. Inf. Theory*, vol. 18, no. 5, pp. 547–557, 1972.

[10] D.L. Snyder, *Random Point Processes*, John Wiley & Sons, New York, 1975.

[11] P. Bremaud, *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, Berlin, 1981.

[12] R.-D. Reiss, *A course on Point Processes*, Springer-Verlag, Berlin, 1993.

[13] D.J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, vol. I: Elementary theory and methods, Springer-Verlag, New York, 2nd ed. edition, 2003.

[14] H. German, "Understanding the fine structure of electricity prices," *The Journal of Business*, vol. 79, no. 3, pp. 1225–1261, 2006.

[15] T.M. Christensen, A.S. Hurn, and K.A Lindsay, "It never rains but it pours: Modelling the persistence of spikes in electricity prices," *The Energy Journal*, vol. 30, no. 1, pp. 25–48, 2009.

[16] A.G Hawkes, "Spectra of some self-exciting point processes," *Biometrika*, vol. 58, pp. 83–90, 1971.

[17] T.M. Christensen, A.S. Hurn, and K.A Lindsay, "Forecasting spikes electricity prices," *Int. Journal of Forecasting*, vol. 28, no. 2, pp. 400–411, 2012.

[18] R. Becker, A.E. Clements, and W.N.R.A. Zainudin, "Modeling electricity price events as point processes," *Journal of Energy Markets*, vol. 6, pp. 99140, 2013.

[19] C. Loader, *Local Regression and Likelihood*, Springer, 1999.

[20] J. Fan and I. Gijbels, *Local Polynomial modelling and its Applications*, Chapman& Hall, CRC, 1996.

[21] Y. Ogata and H. Akaike, "On the linear intensity models for mixed doubly stochastic Poisson and self-exciting processes," *J. Royal Stat. Society, Series B*, pp. 102–107, 1982.

[22] B. Wahlberg, "System identification using Laguerre models," *IEEE Trans. on Aut. Control*, vol. 36, no. 5, pp. 551–562, 1991.

[23] E. Hewitt and R.E. Hewitt, "The Gibbs-Wilbraham phenomenon: An episode in Fourier analysis," *Archive of history of exact sciences*, vol. 21, no. 2, pp. 129–160, 1979.

[24] P. Lewis and G. Shedler, "Simulation of nonhomogeneous Poisson processes by thinning," *Naval Research Logistics Quarterly*, , no. 26, pp. 403–413, 1979.

[25] E.N. Brown, R. Barbieri, V. Ventura, R.E. Kass, and L.M. Frank, "The time-rescaling theorem and its applications to neural spike train," *Neural Computation*, vol. 14, pp. 325–346, 2001.

[26] S. Forrest and I. MacGill, "Assessing the impact of wind generation on wholesale prices and generator dispatch in the Australian electricity market," *Energy Policy*, vol. 59, pp. 120–132, 2013.

[27] B.I. Godoy, V. Solo, and S.A. Pasha, "Maximum likelihood identification of Hawkes-Pham models with a guaranteed stability condition," in *54th IEEE Conf. on Decision and Control (CDC)*, Osaka, Japan, 2015, pp. 3373–3378.