[8] M. Galicki, "Task space control of mobile manipulators," *Robotica*, vol. 29, pp. 221–232, 2010.
[9] M. Galicki, "Collision-free control of mobile manipulators in a task space," *Mech. Syst. Signal Process.*, vol. 25, pp. 2766–2784, 2010.
[10] A. Mazur, "Trajectory tracking control in workspace-defined tasks for nonholonomic mobile manipulators," *Robotica* , vol. 28, pp. 57–68, 2010.
[11] J. H. Chung, S. A. Velinsky, and R. A. Hess, "Interaction control of a redundant mobile manipulator," *Int. J. Robot. Res.*, vol. 17, no. 12, pp. 1302–1309, 1998.
[12] A. Mazur, "Hybrid adaptive control laws solving a path following problem for nonholonomic mobile manipulators," *Int. J. Control*, vol. 77, no. 15, pp. 1297–1306, 2004.
[13] S. Lin and A. A. Goldenberg, "Robust damping control of mobile manipulators," *IEEE Trans. Syst., Man, Cybern. B*, vol. 32, no. 1, pp. 1–6, Feb. 2002.
[14] M. Mailah, E. Pitowarno, and H. Jamaluddin, "Robust motion control for mobile manipulator using resolved acceleration and proportional-integral active force control," *Int. J. Adv. Robot. Syst.*, vol. 2, no. 2, pp. 125–134, 2005.
[15] M. Galicki, "Motion control of robotic manipulators in task space," in *Proc. Int. Symp. Intell. Robot. Syst.*, Lausanne, Switzerland, 2002, pp. 2061–2066.
[16] M. Galicki, "An adaptive regulator of robotic manipulators in the task space," *IEEE Trans. Autom. Control*, vol. 28, no. 4, pp. 1058–1062, Apr. 2008.
[17] G. Campion, G. Bastin, and B. D. Andrea-Novel, "Structural properties and classification of kinematic and dynamic models of wheeled mobile robots," *IEEE Trans. Robot. Autom.*, vol. 12, no. 1, pp. 47–62, Jan. 1996.
[18] F. L. Lewis, C. T. Abdallach, and D. M. Dawson, *Control of Robotic Manipulators*. New York: Macmillan, 1993.
[19] I. Duleba, "Modeling and control of mobile manipulators," in *Proc. IFAC Symp. SYROCO*, 2000, pp. 687–692.
[20] T. Yoshikawa, "Manipulability of robotic mechanisms," *Int. J. Robot. Res.*, vol. 4, no. 2, pp. 3–9, 1985.
[21] E. G. Gilbert and D. W. Johnson, "Distance functions and their application to robot path planning," *IEEE J. Robot. Autom.*, vol. RA-1, no. 1, pp. 21–30, Mar. 1985.
[22] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile manipulators," *Int. J. Robot. Res.*, vol. 5, no. 1, pp. 90–98, 1986.
[23] W. Rubinowicz and W. Krolikowski, *Theoretical Mechanics*. Warszawa, Poland: PWN, 1980.
[24] M. Galicki, "Inverse kinematics solution to mobile manipulators," *Int. J. Robot. Res.*, vol. 22, no. 12, pp. 1041–1064, 2003.
[25] Y. Nakamura, H. Hanafusa, and T. Yoshikawa, "Task-priority based redundancy control of robot manipulators," *Robot. Res.*, vol. 6, no. 2, pp. 3–15, 1987.
[26] M. Krstic, I. Kanellakopoulos, and P. Kokotovic, *Nonlinear and Adaptive Control Design*. New York: Wiley, 1995.
[27] M. Galicki, "Collision-free control of robotic manipulators in the task space," *J. Robot. Syst.*, vol. 22, no. 8, pp. 439–455, 2005.
[28] M. Galicki and A. Morecki, "Finding collision-free trajectory for redundant manipulator by local information available," in *Proc. 9th CISM-IFToMM Symp. Theor. Practice Robots Manip.*, 1992, pp. 61–71.

Kalman Filtering for a Generalized Class of Nonlinear Systems and a New Gaussian Quadrature Technique

Alexandros C. Charalampidis and George P. Papavassilopoulos

***Abstract*—The class of nonlinear systems treated in this technical note consists of the discrete time nonlinear systems that are formed by the interconnection of linear systems through static nonlinearities with few inputs. This special structure is exploited to reduce the dimension of the integrals involved in the propagation of mean values and covariances, thus permitting accurate calculations. Furthermore, a new quadrature scheme suitable for nonlinear Kalman filtering is introduced. The proposed techniques are applied to a seven-dimensional numerical example. The results show that they can increase the performance significantly.**

***Index Terms*—Covariance matrices, Kalman filtering, nonlinear filters, numerical methods.**

I. Introduction

The Kalman filter (KF) [1]–[4] provides an exact solution to the recursive state estimation problem for linear systems with Gaussian noise and initial condition. In that case the state distribution is also Gaussian and it can be fully described by its mean and covariance matrix. For nonlinear systems there exist cases, for example when the state probability density function (pdf) is multimodal, where more computationally intensive techniques [5]–[9] have to be used. However, there also many practical situations in which it is possible to use a Gaussian to approximate the state distribution, and KF-based techniques (in which this approximation is made) have been widely studied and used in applications [10]–[14].

The nonlinear KF variants incorporate a technique to compute the expected values of nonlinear functions of the state, as is explained in more detail in the next section. Thus n-dimensional quadrature is needed, which can be very computationally demanding in order to be accurate when n is not small. In the Extended Kalman Filter (EKF) [2], [10] the functions are linearized, but this approximation can be very poor and may lead even to instability. The Unscented Kalman Filter (UKF) [15], [16] is more robust, and its computational cost is small. However, its performance is not as high as that of other filters in which more accurate approximations are made.

In [17] the use of Gauss-Hermite quadrature [17]–[19] has been suggested, but this approach is infeasible for high-dimensional systems. Furthermore, Gauss-Hermite quadrature has not been designed for the nonlinear recursive state estimation problem, in which neither the exact distribution nor its parameters are known exactly. An interesting class of systems consists of the systems formed by the interconnection of dynamical systems through nonlinear static characteristics [20], [21]. In [19] it was shown that for the case of linear dynamical systems and

Manuscript received December 01, 2010; revised December 02, 2010, June 06, 2011; accepted January 26, 2012. Date of publication July 11, 2012; date of current version October 24, 2012. This work was supported in part by the National Technical University of Athens. Recommended by Associate Editor C.-H. Chen.

A. Charalampidis is with the Swiss Federal Institute of Technology, Lausanne (EPFL), Lausanne CH-1015, Switzerland (e-mail: alexandros.charalampidis@epfl.ch).

G. P. Papavassilopoulos is with the National Technical University of Athens, School of Electrical and Computer Engineering, Athens 157 80, Greece (e-mail: yorgos@netmode.ntua.gr).

Digital Object Identifier 10.1109/TAC.2012.2208289

single-input nonlinear characteristics it is possible to substitute integration in n-dimensions by solving a number of linear systems in n-dimensions and integration in one and two dimensions.

In this technical note, using analogous techniques it is shown how the dimensionality of the integration can be reduced when the nonlinear characteristics have many inputs. Furthermore, an alternative quadrature scheme suitable for the recursive estimation is introduced. The two proposed techniques can be applied combined or separately.

The technical note is structured as follows. Section II presents the problem formulation and some basic results. In Section III the class under study is defined and it is shown how the dimensionality of the integration can be reduced. Section IV presents the new quadrature technique, while Section V presents a numerical example comparing the performance of the proposed techniques with the standard ones. Section VI contains the conclusion.

II. Background

The general nonlinear filtering problem for dynamic systems with additive noise is to estimate the state of systems of the form

$$x_{k+1} = f(x_k) + w_k, \quad k = 0, 1, 2, \dots \tag{1}$$

$$y_k = h(x_k) + v_k, \quad k = 1, 2, \dots \tag{2}$$

where x_k is the state of the system and y_k the measured output at time k. w_k is the disturbance, also referred to as process noise, and v_k is the measurement noise. In this technical note it is assumed that the random variables (r.v.) x_0, w_k, $k = 0, 1, 2, \dots$ and v_k, $k = 1, 2, \dots$ are mutually independent and normally distributed with known parameters. Furthermore, w_k and v_k have zero mean.

Using Bayes Rule [22], [23] it is possible [19], [22] to obtain recursive integral equations. However, the integrals involved cannot be evaluated analytically. As stated in the introduction, the problem can be solved exactly when the system is linear. This is done as follows.

Suppose that a dynamical system is described by $x_{k+1} = A_k x_k + B_k + w_k$, $y_k = C_k x_k + D_k + v_k$ where w_k and v_k are normally distributed with zero mean, while their covariance matrices are Q and R respectively. Suppose also that it is known that x_k follows the normal distribution with mean $\hat{x}_k$ and covariance P_{x_k}. Then, a priori with respect to y_{k+1}, x_{k+1} follows the normal distribution with mean $\hat{x}^-_{k+1}$ and covariance $P^-_{x_{k+1}}$ given by $\hat{x}^-_{k+1} = A_k \hat{x}_k + B_k$, $P^-_{x_{k+1}} = A_k P_{x_k} A_k^T + Q$. The predicted value of y_{k+1} is then $\hat{y}^-_{k+1} = C_{k+1}\hat{x}^-_{k+1} + D_k$ while its covariance is $P^-_{y_{k+1}} = C_{k+1} P^-_{x_{k+1}} C^T_{k+1} + R$, and the cross covariance of x_{k+1} and y_{k+1} is equal to $P^-_{x_{k+1}y_{k+1}} = P^-_{x_{k+1}} C^T_{k+1}$. The value of y_{k+1} can be then used to refine the distribution, according to the following equations:

$$\begin{aligned} &K_{k+1} = P^-_{x_{k+1}y_{k+1}} P^{-1}_{y_{k+1}},\ \hat{x}_{k+1} = \hat{x}^-_{k+1} + K_{k+1}\left(y_{k+1} - \hat{y}^-_{k+1}\right), \\ &P_{x_{k+1}} = P^-_{x_{k+1}} - K_{k+1} P^-_{y_{k+1}} K^T_{k+1}. \end{aligned} \tag{3}$$

Remark 1: If the Gaussian assumption is removed, KF gives the linear minimum covariance estimate of the state [1], [3].

In the case that the system is nonlinear, the above equations cannot be applied. For example, instead of $\hat{x}^-_{k+1} = A_k \hat{x}_k + B_k$ it holds $\hat{x}^-_{k+1} = \mathbb{E}[x_{k+1}|y_{1:k}] = \mathbb{E}[f(x_k) + w_k|y_{1:k}] = \mathbb{E}[f(x_k)|y_{1:k}]$, where $y_{1:k} = \{y_1, y_2, \cdots, y_k\}$. Instead of $P^-_{x_{k+1}} = A_k P_{x_k} A_k^T + Q\mathrm{V}[x_{k+1}|y_{1:k}] = \mathbb{E}[(x_{k+1} - \hat{x}^-_{k+1})(x_{k+1} - \hat{x}^-_{k+1})^T|y_{1:k}] = \mathbb{E}[(f(x_k) + w_k - \hat{x}^-_{k+1})(f(x_k) + w_k - \hat{x}^-_{k+1})^T|y_{1:k}] = \mathbb{E}[(f(x_k) f(x_k)^T|y_{1:k}] - \hat{x}^-_{k+1}\hat{x}^{-T}_{k+1} + Q$. Similar expressions are derived for the correction step. Therefore the nonlinear KF variants must incorporate a numerical technique approximating expected values of the form $\mathbb{E}[g(X)]$. Then, they apply (3).

If X has mean μ and covariance P, then it holds

$$\mathbb{E}[g(X)] = \int g(x) \frac{1}{(2\pi)^{\frac{n}{2}} |P|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T P^{-1}(x-\mu)} dx \tag{4}$$

thus approximating $\mathbb{E}[g(X)]$ is a numerical integration problem in n-dimensions, which is computationally demanding for high n, considering that it has to be done on-line and in every time step.

III. Analysis for the Class Under Study

A. Description of the Class

Let us consider a set of N_s linear dynamical systems described by $S^{(i)} : x^{(i)}_{k+1} = A^{(i)} x^{(i)}_k + B^{(i)} u^{(i)}_k$, $y^{(i)}_k = C^{(i)} x^{(i)}_k$ for $i = 1, \dots, N_s$. Suppose that these systems are interconnected linearly as well as through nonlinear functions, so that

$$u^{(i)}_k = \sum_{j=1}^{N_s} \lambda_{ij} y^{(j)}_k + \sum_{j=1}^{N_c} \mu_{ij} g_j(D_j x_k). \tag{5}$$

Thus, the matrix $\Lambda = (\lambda_{ij})$ describes the linear interconnection of the subsystems, while the nonlinear interconnection is made through the nonlinear functions g_i, $i = 1, \dots, N_c$. The matrix D_i has l_i rows and without loss of generality it is assumed of full row rank. The gain from the output of g_j to the input of $S^{(i)}$ is equal to the (i, j)th element of the matrix $M = (\mu_{ij})$. In this situation, the following proposition holds.

Proposition 1: Consider the N_s linear systems defined above, whose interconnection is described by M, Λ and g_i, D_i for $i = 1, \dots, N_c$. Suppose further that the total system S with state $x = [x^{(1)T} \cdots x^{(N_s)T}]^T$ is subject to additive disturbance w_k. Define $\mathrm{diag}(A^{(i)})$ the block diagonal matrix with diagonal entries $A^{(1)}, \dots, A^{(N_s)}$ and define also similarly $\mathrm{diag}(B^{(i)})$ and $\mathrm{diag}(C^{(i)})$. Then the state of the total system S is described by the following dynamics:

$$x_{k+1} = A x_k + \sum_{i=1}^{N_c} B_i g_i(D_i x_k) + w_k \tag{6}$$

where B_i is the i-th column of $B = \mathrm{diag}(B^{(i)})M$ and $A = \mathrm{diag}(A^{(i)}) + \mathrm{diag}(B^{(i)})\Lambda\mathrm{diag}(C^{(i)})$.

Proof: Substituting (5) and $y^{(i)}_k = C^{(i)} x^{(i)}_k$ in $x^{(i)}_{k+1} = A^{(i)} x^{(i)}_k + B^{(i)} u^{(i)}_k$ yields $x^{(i)}_{k+1} = A^{(i)} x^{(i)}_k + B^{(i)} \sum_{j=1}^{N_s} \lambda_{ij} C^{(j)} x^{(j)}_k + B^{(i)} \sum_{j=1}^{N_s} \mu_{ij} g_j(D_j x_k)$. By rewriting these relations in matrix form and taking also into account the additive noise w_k, (6) is obtained. ∎

The measured output of the total system S may contain both a linear and a nonlinear part. It is described by

$$y_k = \begin{bmatrix} y^L_k \\ y^N_k \end{bmatrix} = \begin{bmatrix} C^L x_k \\ (h_i(C^N_i x_k)) \end{bmatrix} + v_k. \tag{7}$$

$(h_i(C^N_i x_k))$ is a column vector containing the N_o nonlinear outputs. C_i has r_i rows and, like D_i, it is assumed of full row rank. If only linear outputs exist, then $N_o = 0$ and $y_k = y^L_k$. If only nonlinear outputs exist, then C^L is the empty matrix and $y_k = y^N_k$.

The class described above is strictly larger than the class studied in [19], because in that class D_i and C^N_i (D^T_i and C^T_i in the notation of that paper) were row vectors, while now they may have more rows. Equivalently, now the static nonlinearities may have more than one

real-valued inputs, while in that paper they had only one real-valued input.

Finally, we note that the system under study can be time-variant since all the analysis is done separately for each time step. This includes both the dynamical systems as well as their interconnection. The linear systems can be multi-input or multi-output but then some of the matrices have to be block matrices (such as M and Λ).

B. Building Blocks of the Algorithm

As will be made apparent in the next subsection, the algorithm needs to calculate expectations of the three following forms: $\mathbb{E}[f(Lx)]$, $\mathbb{E}[f(L_1x, L_2x)]$ and $\mathbb{E}[xf(Lx)]$, where x is n-dimensional and follows the Gaussian distribution with mean $\hat{x}$ and covariance matrix P, L, L_1 and L_2 are $m \times n$, $n_1 \times n$ and $n_2 \times n$ full row rank matrices, and f is a nonlinear function. In the following, it is explained how each of them is computed.

1) $\mathbb{E}[f(Lx)]$*:* The r.v. $\xi = Lx$ is m-dimensional and Gaussian, with mean $L\hat{x}$ and covariance matrix LPL^T, thus the computation of $\mathbb{E}[f(Lx)] = \mathbb{E}[f(\xi)]$ is an m-dimensional integration problem, and any quadrature technique can be used to yield an approximation.

2) $\mathbb{E}[f(L_1x, L_2x)]$*:* Let us define the $n_1 + n_2$-dimensional r.v. $z = \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} L_1x \\ L_2x \end{smallmatrix}\right] = \left[\begin{smallmatrix} L_1 \\ L_2 \end{smallmatrix}\right] x$. It holds

$$\hat{z} = \mathbb{E}[z] = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \hat{x}, \quad P_z = \mathrm{V}[z] = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} P \begin{bmatrix} L_1^T & L_2^T \end{bmatrix} \tag{8}$$

therefore $\mathbb{E}[f(L_1x, L_2x)] = \mathbb{E}[f(z_1, z_2)]$ can be approximated using $n_1 + n_2$-dimensional quadrature. Possibly, however, $\operatorname{rank} \left[\begin{smallmatrix} L_1 \\ L_2 \end{smallmatrix}\right] = r < n_1 + n_2$. In that case, it is possible to find matrices C and R such that $\operatorname{rank} \left[\begin{smallmatrix} L_1 \\ L_2 \end{smallmatrix}\right] = CR$. Indeed, if R is composed from r linearly independent rows of $\left[\begin{smallmatrix} L_1 \\ L_2 \end{smallmatrix}\right]$, then every row of $\left[\begin{smallmatrix} L_1 \\ L_2 \end{smallmatrix}\right]$ is a linear combination of the rows of R, therefore it suffices to set the elements of C equal to the corresponding coefficients.

Let us now define the r-dimensional r.v. $\zeta = Rx$. ζ is Gaussian, with $\hat{\zeta} = \mathbb{E}[\zeta] = R\hat{x}$ and $P_\zeta = \mathrm{V}[\zeta] = RPR^T$. It holds $z = C\zeta$, thus $\mathbb{E}[f(z)] = \mathbb{E}[f(C\zeta)]$. If the function $f_C(\zeta) = f(C\zeta)$ is defined, combining all the above results yields $\mathbb{E}[f(L_1x, L_2x)] = \mathbb{E}[f_C(\zeta)]$, thus the computation of $\mathbb{E}[f(L_1x, L_2x)]$ is a problem of r-dimensional quadrature.

3) $\mathbb{E}[xf(Lx)]$*:* In this case the computation of $\mathbb{E}[xf(Lx)]$ is reduced to m dimensions using the following proposition, whose proof is given in the Appendix:

Proposition 2: Suppose that x is a normally distributed random variable with values in $\mathbb{R}^n$, mean $\hat{x}$ and covariance matrix $P > 0$, L is a $m \times n$ full row rank matrix and $f : \mathbb{R}^n \to \mathbb{R}$ is a Borel-measurable function such that $f(Lx)$ has finite variance. Suppose also that $L^\perp$ is a $(n - m) \times n$ full row rank matrix such that $L^\perp PL^T = 0$. Let $S = \mathbb{E}[xf(Lx)]$. Then S is the unique solution of the linear system

$$\begin{bmatrix} L \\ L^\perp \end{bmatrix} S = \begin{bmatrix} \mathbb{E}\left[Lxf(Lx)\right] \\ L^\perp \hat{x} \mathbb{E}\left[f(Lx)\right] \end{bmatrix}. \tag{9}$$

Remark 2: Such an $L^\perp$ always exists, because P is positive definite and L is of full row rank. In fact, if R^n is endowed with the inner product $\langle v_1, v_2 \rangle = v_1^T P v_2$, then $\mathrm{Row}(L^\perp) = (\mathrm{Row}L)^\perp$.

Remark 3: Proposition 2 holds for normal random variables, while the true state distribution is not normal. However, in the nonlinear Kalman filters the expectations are computed for a Gaussian distribution even without the use of Proposition 2. Thus, if the quadrature is accurate, the usage of the proposition does not affect the results.

Remark 4: Proposition 1 of [19] is a special case of the above proposition, covering the case of L having one row.

The numerical approximation of the integrals can be made with any standard quadrature technique, such as the Gauss-Hermite quadrature, or with the technique described in Section IV.

C. Proposed Filtering Algorithm

The analysis is similar to that of [19], especially for the prediction step. The key idea is that all expectations which must be calculated have one of the above three forms.

1) Prediction Step: Assume that x_k has mean $\hat{x}_k$ and covariance matrix P_{x_k}. Then (6) yields

$$\mathbb{E}[x_{k+1}] = A\hat{x}_k + \sum_{i=1}^{N_c} B_i \mathbb{E}\left[g_i(D_i x_k)\right]. \tag{10}$$

$\mathbb{E}[g_i(D_i x_k)]$ is of the first form of the previous subsection as seen in (11), shown at the bottom of the page.

Remark 5: All expectations in this paragraph are conditioned upon $\hat{x}_k$ and P_{x_k}, but the conditioning is omitted for brevity. For example, $\mathbb{E}[x_{k+1}]$ is written instead of $\mathbb{E}[x_{k+1}|\hat{x}_k, P_{x_k}]$ and $\mathrm{V}[x_{k+1}]$ instead of $\mathrm{V}[x_{k+1}|\hat{x}_k, P_{x_k}]$.

The prediction step includes also calculation of the a priori covariance matrix of x_{k+1}. It holds

$$\begin{aligned} P_{x_{k+1}}^- &= \mathrm{V}[x_{k+1}] = \mathrm{V}\left[Ax_k + \sum_{i=1}^{N_c} B_i g_i(D_i x_k) + w_k\right] \\ &= \mathrm{V}[Ax_k] + \mathrm{V}\left[\sum_{i=1}^{N_c} B_i g_i(D_i x_k)\right] + \mathrm{V}[w_k] \\ &\quad + \mathrm{Cov}\left(Ax_k, \sum_{i=1}^{N_c} B_i g_i(D_i x_k)\right) \\ &\quad + \mathrm{Cov}\left(\sum_{i=1}^{N_c} B_i g_i(D_i x_k), Ax_k\right). \end{aligned} \tag{12}$$

It is known that $\mathrm{V}[w_k] = Q$ and $\mathrm{V}[Ax_k] = AP_{x_k}A^T$.

The second term of (12) is equal to

$$\begin{aligned} &\mathbb{E}\left[\sum_{i=1}^{N_c} B_i g_i(D_i x_k) \left(\sum_{i=1}^{N_c} B_i g_i(D_i x_k)\right)^T\right] \\ &\quad - \mathbb{E}\left[\sum_{i=1}^{N_c} B_i g_i(D_i x_k)\right] \mathbb{E}\left[\sum_{i=1}^{N_c} B_i g_i(D_i x_k)\right]^T \\ &= \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} B_i \mathbb{E}\left[g_i(D_i x_k) g_j(D_j x_k)\right] B_j^T \\ &\quad - \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} B_i \mathbb{E}\left[g_i(D_i x_k)\right] \mathbb{E}\left[g_j(D_j x_k)\right] B_j^T. \end{aligned} \tag{13}$$

$$P_{y_{k+1}}^- = \begin{bmatrix} C^L P_{x_{k+1}}^- (C^L)^T & C^L \left(\mathbb{E}\left[x_{k+1} h_i\left(C_i^N x_{k+1}\right)\right] - \hat{x}_{k+1}^- \mathbb{E}\left[h_i\left(C_i^N x_{k+1}\right)\right]\right) \\ \star & \left(\mathbb{E}\left[h_i\left(C_i^N x_{k+1}\right) h_j\left(C_j^N x_{k+1}\right)\right] - \mathbb{E}\left[h_i\left(C_i^N x_{k+1}\right)\right] \mathbb{E}\left[h_j\left(C_j^N x_{k+1}\right)\right]\right) \end{bmatrix} + R \tag{11}$$

For $i=j$, $\mathbb{E}[g_i(D_ix_k)g_j(D_jx_k)]=\mathbb{E}[g_i^2(D_ix_k)]$ is of the first form of the previous subsection, while for $i\neq j$ it is of the second form.

The fourth term of (12) is the transpose of the third term, so it suffices to approximate the third term only. It holds

$$\begin{aligned}\mathrm{Cov}\left(Ax_k,\sum_{i=1}^{N_c}B_ig_i(D_ix_k)\right)&=\mathbb{E}\left[Ax_k\left(\sum_{i=1}^{N_c}B_ig_i(D_ix_k)\right)^T\right]\\&\quad-\mathbb{E}[Ax_k]\mathbb{E}\left[\sum_{i=1}^{N_c}B_ig_i(D_ix_k)\right]^T\\&=A\sum_{i=1}^{N_c}\mathbb{E}[x_kg_i(D_ix_k)]B_i^T\\&\quad-A\hat{x}_k\sum_{i=1}^{N_c}\mathbb{E}[g_i(D_ix_k)]B_i^T.\end{aligned}\tag{14}$$

$\mathbb{E}[x_kg_i(D_ix_k)]$ has the third form of the the previous subsection.

2) Correction Step: Again the expectations involved in the analysis have one of the three forms of the previous analysis. The quantities that must be computed for the correction step are $\hat{y}_{k+1}^-$, $P_{y_{k+1}}^-$ and $P_{x_{k+1},y_{k+1}}^-$.

With respect to $\hat{y}_{k+1}^-$, from (7) it follows that:

$$\hat{y}_{k+1,i}^-=\begin{bmatrix}C^L\hat{x}_{k+1}^-\\(\mathbb{E}\left[h_i\left(C_i^Nx_{k+1}\right)\right])\end{bmatrix}.\tag{15}$$

Remark 6: All expectations in this paragraph are conditioned upon $\hat{x}_{k+1}^-$ and $P_{x_{k+1}}^-$, but the conditioning is omitted for brevity. For example, $\mathbb{E}[h_i(C_i^Nx_{k+1})]$ is written instead of $\mathbb{E}[h_i(C_i^Nx_{k+1})|\hat{x}_{k+1}^-,P_{x_{k+1}}^-]$.

$P_{x_{k+1},y_{k+1}}^-$ is given by

$$P_{x_{k+1},y_{k+1}}^-=\begin{bmatrix}P_{x_{k+1}}^-(C^L)^T\\(\mathbb{E}\left[x_{k+1}h_i\left(C_i^Nx_{k+1}\right)\right])-\hat{x}_{k+1}^-\left(\mathbb{E}\left[h_i\left(C_i^Nx_{k+1}\right)\right]\right)\end{bmatrix}\tag{16}$$

where $(\mathbb{E}[x_{k+1}h_i(C_i^Nx_{k+1})])$ is an $n\times N_o$ matrix whose i-th column is $\mathbb{E}[x_{k+1}h_i(C_i^Nx_{k+1})]$, and similarly $(\mathbb{E}[h_i(C_i^Nx_{k+1})])$ is a row vector.

$P_{y_{k+1}}^-$ is given by (11) where $\star$ is the transpose of the matrix in the symmetric position.

After the calculation of $\hat{y}_{k+1}^-$, $P_{y_{k+1}}^-$ and $P_{x_{k+1},y_{k+1}}^-$, the correction step is completed using (3).

Finally, we state without detailed proof the following proposition, analogous to Proposition 2 of [19]. Its proof, whose key idea is that the expectations of the nonlinear functions do not change (namely $\bar{g}_i'=\bar{g}_i$, etc.), follows the same lines with the proof of Proposition 2 in [19] and is omitted.

Proposition 3: Consider the system described by (6), (7) and the coordinates transformation $x'=Tx$. If $\hat{x}_k$ and P_{x_k} are the estimates produced by the proposed filter for the initial coordinates while $\hat{x}_k'$ and $P_{x_k'}$ are those for the transformed coordinates, then for every $k\in\mathbb{N}$ it holds

$$\hat{x}_k'=T\hat{x}_k,\quad P_{x_k'}=TP_{x_k}T^T.\tag{17}$$

IV. Novel Gaussian Quadrature Technique

As in [17], the one-dimensional quadrature is used as a basis for the multidimensional quadrature. This is done as follows.

Suppose that $\sum_{i=1}^m w_i\gamma(x_i)$ is the approximation made for $\mathbb{E}[\gamma(\chi)]$ when χ follows the standard normal distribution. Let us now consider the approximation of $\mathbb{E}[g(X)]$ when X is n-dimensional Gaussian with mean μ and covariance matrix P. Let S_i the ith column of $S=\sqrt{P}$, where S is a matrix such that $P=SS^T$. S is obtained using Cholesky decomposition. Then if v_i, $i=1\ldots n_i$, are n independent r.v. following the standard normal distribution, $\mu+\sum S_iv_i$ is equidistributed with X. Thus it is reasonable to make the following approximation:

$$\mathbb{E}\left[g(X)\right]\approx\sum_{i_1=1}^m\ldots\sum_{i_n=1}^m w_{i_1}\cdots w_{i_n}g\left(\mu+x_{i_1}S_1+\ldots+x_{i_n}S_n\right).\tag{18}$$

However, in the nonlinear filtering problem, the parameters of the state distribution are not known exactly. Thus, it is advantageous to use an approximation which is robust with respect to parameter deviations. The sensitivity of the approximation with respect to μ is a sum of the form $\sum_{i=1}^M W_i(\partial g/\partial x)(X_i)$, where $M=m^n$ is the total number of the points, whose locations are given by X_i and weights by W_i. This sum is the inner product $(W_1,\ldots,W_M)\cdot((\partial g/\partial x)(X_1),\ldots,(\partial g/\partial x)(X_M))$. Since $((\partial g/\partial x)(X_1),\ldots,(\partial g/\partial x)(X_M))$ is not known a priori (in fact even g is unknown since the algorithm does not depend on it), it is reasonable to minimize the norm of $(W_1,\ldots,W_M)$.

$\|(W_1,\ldots,W_M)\|_p$, $1<p\leq\infty$, is minimized when all the weights are equal to $1/M$ (note that their sum must be equal to 1). This choice would also minimize the variance of the inner product if $(\partial g/\partial x)(X_i)$ were i.i.d. random variables. The weights W_i, $i=1\ldots M$, are all equal if and only if the weights w_i, $i=1\ldots m$, are all equal. The proposed algorithm, thus, chooses $w_i=1/m$ for the one-dimensional case, in order to get a robust estimate.

Therefore, the approximation made for $\mathbb{E}[\gamma(\chi)]$ is $\sum_{i=1}^m\frac{1}{m}\gamma(x_i)$ which is, in fact, equal to $\mathbb{E}[\gamma(\chi)]$ when χ follows a discrete distribution, whose possible values are equiprobable and equal to x_i, $i=1\ldots m$. It is reasonable, therefore, to choose x_i, $i=1\ldots m$, so that the corresponding distribution is close to the Gaussian distribution. The values of x_i, $i=1\ldots m$, are chosen using the following proposition. Its proof is given in the Appendix.

Proposition 4: Let $F(x)=(1/\sqrt{2\pi})\int_{-\infty}^x e^{-u^2/2}du$ be the cumulative distribution function (cdf) of the standard normal distribution. If $G(x)$ is the cdf of the point set, then for every $1\leq p\leq\infty$, the criterion $C_p=\|F-G\|_p$ is minimized for the point set

$$\left\{l_n=F^{-1}(p_n),p_n=\frac{1}{2m}+\frac{n-1}{m},n=1\ldots m\right\}.\tag{19}$$

An interesting question is whether the approximation convergences to the true expectation when $m\to\infty$. The answer is affirmative for a rather large class of functions:

Proposition 5: Suppose that γ is an almost everywhere continuous function of exponential order, namely there exist $M,a>0$ such that $|\gamma(x)|\leq Me^{a|x|}$. Then

$$\lim_{m\to\infty}\frac{1}{m}\sum_{n=1}^m\gamma(l_n)=\mathbb{E}\left[\gamma(x)\right].\tag{20}$$

Proof: See the Appendix. ∎

V. Numerical Example

The system under study consists of three nonlinearly interconnected linear systems. The first has a real pole at 0.9 and a complex pair of poles at $0.9(\cos\frac{\pi}{6}\pm i\sin\frac{\pi}{6})$. The second has a real pole at 0.9 and a complex pair of poles at $0.9(\cos\frac{\pi}{8}\pm i\sin\frac{\pi}{8})$. The third has one real pole at 0.9. They have no zeros and they are scaled so that when driven by Gaussian white noise (GWN) with unit variance, the variance of

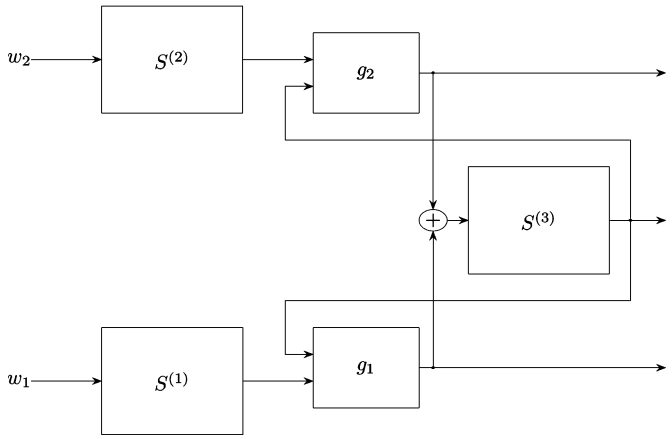

Fig. 1. Block diagram of the example systems.

their output is also equal to 1. The above systems are described in state space by the following (rounded) matrices:

$$A^{(1)} = \begin{bmatrix} 0.9 & 1 & 0 \\ 0 & 0.7794 & 1 \\ 0 & -0.2025 & 0.7794 \end{bmatrix},$$
$$A^{(2)} = \begin{bmatrix} 0.9 & 1 & 0 \\ 0 & 0.8315 & 1 \\ 0 & -0.1186 & 0.8315 \end{bmatrix},\ B^{(1)} = B^{(2)} = \begin{bmatrix} 0 \\ 0 \\ 0.25 \end{bmatrix},$$
$$C^{(1)} = [\,0.373 \quad 0 \quad 0\,],\ C^{(2)} = [\,0.2277 \quad 0 \quad 0\,],\ A^{(3)} = 0.9,$$
$$B^{(3)} = 0.5,\ C^{(3)} = 0.8718. \tag{21}$$

These linear systems are interconnected only through nonlinear characteristics, namely $\Lambda = 0_{3\times 3}$. With respect to the nonlinear characteristics, there are two of them, g_1 and g_2. For g_1 it holds $g_1(D_1 x) = (y_1^S + 0.2(y_1^S)^2)(\pi + \arctan(5y_3^S))$, while for g_2 it holds $g_2(D_2 x) = (y_2^S + 0.2(y_2^S)^2)(\pi + \arctan(5y_3^S))$, where y_i^S is the output of the ith linear system. The outputs of g_1 and g_2 are driven to the input of the third system, namely $M = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$. Each of the first and second linear system is driven by GWN with variance equal to 10. The initial state is supposed to have zero mean and covariance matrix equal to $I_{7\times 7}$. The output of the system has both a linear and a nonlinear part. The linear part consists of y_3^S. The nonlinear output variables are equal to the outputs of the two nonlinear characteristics, namely $h_i = g_i$, $C_i^N = D_i, i = 1, 2$. All three measurement sequences are corrupted by additive GWN with variance equal to 10, or equivalently $R = 10 I_{3\times 3}$. The block diagram of the system is presented in Fig. 1.

Suppose that the goal is to estimate the linear part of the output, namely y_3^S. This estimation problem can be dealt with effectively using the proposed techniques. Indeed, the total order of the system is 7, thus 7-dimensional quadrature would be needed for the direct computation of the involved expectations. Since $\operatorname{rank}\begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = 3$, only up to 3-dimensional quadrature will have to be deployed with the proposed algorithm.

Both for Gauss-Hermite as well as for the novel quadrature introduced in this technical note, m in equation (18) has been set equal to 5. UKF has been also applied to the problem for comparison purposes. Direct quadrature was too slow even for $m = 3$ (about 100 times slower), while its results were very close to those of the reduced-order quadrature.

The results from 100 runs are presented in Table I. The Gauss-Hermite KF with reduced integration order (GHKF-RO) yields significantly better results, especially for the worst case. The order reduction enables to get the improved results in reasonable time. Usage of the novel quadrature (NQKF-RO) further improves the results, and this improvement comes at no cost in comparison with GHKF-RO. The computational demands presented are per time step and have been recorded on a 64-bit PC clocked at 2.9 GHz and running MATLAB 7.2 for Linux. It is noted that the values of l_n in (19) are computed off-line, since they do not change in each time step.

TABLE I
TABLE OF RMS ESTIMATION ERROR

Estim. Technique	Mean Value	Max. Value	Comput. Time (ms)
UKF	4.30	8.68	2.42
GHKF-RO	3.60	6.18	13.03
NQKF-RO	3.16	4.94	13.05

VI. CONCLUSION

This technical note generalized the results of [19] by showing that the nonlinear filtering problem can be solved without using computationally intensive high-order quadrature when the system under study consists of a number of linear subsystems interconnected through nonlinear characteristics. These characteristics, in contrast with [19], can potentially be multi-input. Additionally, a novel quadrature technique is presented which is designed using norm minimization concepts. The proposed techniques are shown to be effective in a seven-dimensional example, in which they yield better results than UKF. In the same example, the proposed quadrature technique outperforms the Gauss-Hermite quadrature.

APPENDIX

Proof of Propostion 2: It is easy to prove that $\operatorname{rank}\begin{bmatrix} L \\ L^\perp \end{bmatrix} = n$. Indeed, in different case there exists a non-zero vector $\lambda = [\lambda_1\ \lambda_2], \lambda_1 \in \mathbb{R}^m, \lambda_2 \in \mathbb{R}^{n-m}$ such that $\lambda \begin{bmatrix} L \\ L^\perp \end{bmatrix} = 0 \Leftrightarrow \lambda_1 L + \lambda_2 L^\perp = 0$. Postmultiplying the last equation by PL^T yields $\lambda_1 LPL^T + \lambda_2 L^\perp PL^T = 0 \Rightarrow \lambda_1 LPL^T = 0$. But since $P > 0$ and $\operatorname{rank} L = m$, $LPL^T > 0$ thus $\lambda_1 = 0$. This implies that $\lambda_2 L^\perp = 0$. Since $\operatorname{rank} L^\perp = n - m$, it follows that $\lambda_2 = 0$. This completes the argument and shows that (9) has a unique solution.

Let us now show that S satisfies (9). Obviously $LS = L\mathbb{E}[xf(Lx)] = \mathbb{E}[Lxf(Lx)]$. To prove the second vector equation, the fact that Lx and $L^\perp x$ are independent is used. Indeed, $\operatorname{Cov}(L^\perp x, Lx) = L^\perp PL^T = 0$, and since x is Gaussian this implies that Lx and $L^\perp x$ are independent. Then $f(Lx)$ and $L^\perp x$ are independent, therefore $\mathbb{E}[L^\perp x f(Lx)] = \mathbb{E}[L^\perp x]\mathbb{E}[f(Lx)] = L^\perp \hat{x}\mathbb{E}[f(Lx)]$, and the proof is completed. ∎

Proof of Propostion 4: Let us consider first the case $1 \le p < \infty$. Then, for any choice of $x_1 \le \ldots \le x_k \le \ldots \le x_m$, let $h(x_1, \ldots, x_m)$ be the associated value of the criterion $C_p = \|F - G\|_p$. It holds

$$h(x_1, \ldots, x_m) = \int_{-\infty}^{x_1} |F(x)|^p\, dx + \int_{x_1}^{x_2} \left|F(x) - \frac{1}{m}\right|^p dx + \ldots + \int_{x_k}^{x_{k+1}} \left|F(x) - \frac{k}{m}\right|^p dx + \ldots + \int_{x_m}^{\infty} |F(x) - 1|^p\, dx. \tag{22}$$

Then

$$\frac{\partial h}{\partial x_k} = \left|F(x_k) - \frac{k-1}{m}\right|^p - \left|F(x_k) - \frac{k}{m}\right|^p. \tag{23}$$

Thus $\partial h/\partial x_k > 0$ for $F(x_k) > \frac{1}{2}(((k-1)/m) + (k/m))$ while $\partial h/\partial x_k < 0$ for $F(x_k) < \frac{1}{2}(\frac{k-1}{m} + (k/m))$. Note that $\frac{1}{2}(((k-1)/m) + (k/m)) = \frac{1}{2m} + ((k-1)/m) = p_k = F(l_k)$. Since F is

strictly increasing, this implies that $\partial h/\partial x_k > 0$ for $x_k > l_k$ while $\partial h/\partial x_k < 0$ for $x_k < l_k$.

Using the last statement it is easy to prove that for any $y_1 \le \ldots \le y_k \le \ldots \le y_m$ it holds $h(l_1,\ldots,l_m) \le h(y_1,\ldots,y_m)$. Indeed, if $y_m \le l_m$ set $y_k^{(1)} = y_k$ for $k < m$ and $y_m^{(1)} = l_m$, while if $y_m > l_m$, let r the least index such that $y_r > l_m$ and set $y_k^{(1)} = y_k$ for $k < r$ and $y_k^{(1)} = l_m$ for $k \ge r$. Then, in any case, $h(y_1^{(1)},\ldots,y_m^{(1)}) \le h(y_1,\ldots,y_m)$ and $y_m^{(1)} = l_m$. In the same way, an m-tuple $y_1^{(2)} \le \ldots \le y_k^{(2)} \le \ldots \le y_m^{(2)}$ can be constructed such that $h(y_1^{(2)},\ldots,y_m^{(2)}) \le h(y_1^{(1)},\ldots,y_m^{(1)})$ and $y_{m-1}^{(2)} = l_{m-1}$, $y_m^{(2)} = l_m$. After m steps, an m-tuple $y_1^{(m)} \le \ldots \le y_k^{(m)} \le \ldots \le y_m^{(m)}$ is constructed with the properties $h(y_1^{(m)},\ldots,y_m^{(m)}) \le h(y_1^{(m-1)},\ldots,y_m^{(m-1)})$ and $y_k^{(m)} = l_k$, $k = 1\ldots m$. Combining all the inequalities obtained at these steps yields $h(l_1,\ldots,l_m) \le h(y_1,\ldots,y_m)$.

Let us now consider the case $p = \infty$. It can be checked that for the proposed points the value of the criterion is $C_\infty = 1/2m$. For any m-tuple $x_1 \le \ldots \le x_k \le \ldots \le x_m$, the following $2m$ intervals can be defined: $I_l = [0, F(x_1))$, $I_r = [F(x_m), 1]$, $I_k^{(l)} = [F(x_k), (F(x_k) + F(x_{k+1}))/2)$ and $I_k^{(r)} = [(F(x_k) + F(x_{k+1}))/2, F(x_{k+1}))$ for $k = 1\ldots m-1$. Since their union is the unit interval, at least one of them has length greater than or equal to $1/2m$. But if $G(x)$ is the cdf of the m-tuple $x_1 \le \ldots \le x_k \le \ldots \le x_m$, the value of $C_\infty = \|F - G\|_\infty$ is greater than or equal to the maximum length of these intervals. Indeed, $\lim_{x\to x_1^-}(F(x) - G(x)) = F(x_1) = \lambda(I_l)$ and $\lim_{x\to x_m^+}(G(x) - F(x)) = 1 - F(x_m) = \lambda(I_r)$. Furthermore, for $k = 1\ldots m-1$, G is constant on $[x_k, x_{k+1})$. Let G_k its value on this interval (it is supposed that $x_k < x_{k+1}$, otherwise neither $I_k^{(l)}$ nor $I_k^{(r)}$ is the interval with the maximum length). Then obviously $\max\{|G_k - F(x_k)|, |G_k - F(x_{k+1})|\} \ge (F(x_{k+1}) - F(x_k))/2 = \lambda(I_k^{(l)}) = \lambda(I_k^{(r)})$. Noting that $\lim_{x\to x_k^+}(G(x) - F(x)) = G_k - F(x_k)$ and $\lim_{x\to x_{k+1}^-}(G(x) - F(x)) = G_k - F(x_{k+1})$ completes the proof. ∎

Remark 7: It is easy to see that the above arguments also show that the choice of x_i, $i = 1\ldots m$ is unique. Additionally, the proof holds verbatim for any continuous and strictly increasing F. For $p < \infty$ it is also required that C_p is finite. It is easy to verify that if $\mathbb{E}[|X|] < \infty$ where X is a random variable with cdf F, then $C_p < \infty \forall p \in [1,\infty)$.

Proof of Propostion 5: Consider first that γ is continuous and let any $\varepsilon > 0$. It must be proved that there exists m_0 such that for every $m > m_0$, $|\mathbb{E}[\gamma(x)] - \frac{1}{m}\sum_{n=1}^m \gamma(l_n)| < \varepsilon$. Since $(1/\sqrt{2\pi})\int_{-\infty}^{\infty} Me^{a|x|}e^{-x^2/2}dx < \infty$, it is possible to find $T_1 > 0$ such that $(1/\sqrt{2\pi})(\int_{-\infty}^{-T_1} Me^{a|x|}e^{-x^2/2}dx + \int_{T_1}^{\infty} Me^{a|x|}e^{-x^2/2}dx) < \varepsilon/3$. Similarly, it is possible to find $T_2 > 0$ such that $(1/\sqrt{2\pi})M\int_{T_2}^{\infty} e^{-x^2/4}dx < \varepsilon/12$. Let us set $T = \max\{T_1, T_2, 4a\}$ and define $\gamma_T = \gamma \cdot \chi_{[-T,T]}$, where $\chi_{[-T,T]}$ is the characteristic function of the set $[-T,T]$, and $\gamma_T^c = \gamma - \gamma_T = \gamma \cdot \chi_{[|x|>T]}$.

Then $|\mathbb{E}[\gamma(x)] - \frac{1}{m}\sum_{n=1}^m \gamma(l_n)|$ is equal to

$$\begin{aligned}&\left|\mathbb{E}[\gamma(x)] - \mathbb{E}[\gamma_T(x)] + \mathbb{E}[\gamma_T(x)] - \frac{1}{m}\sum_{n=1}^m \gamma_T(l_n) + \right.\\ &\qquad \left.\frac{1}{m}\sum_{n=1}^m \gamma_T(l_n) - \frac{1}{m}\sum_{n=1}^m \gamma(l_n)\right|\\ &\le \left|\mathbb{E}\left[\gamma_T^C(x)\right]\right| + \left|\mathbb{E}[\gamma_T(x)] - \frac{1}{m}\sum_{n=1}^m \gamma_T(l_n)\right|\\ &\qquad + \left|\frac{1}{m}\sum_{n=1}^m \gamma_T^c(l_n)\right|. \qquad (24)\end{aligned}$$

For the first term of the right-hand part of (24) it holds

$$\begin{aligned}&\left|\mathbb{E}\left[\gamma_T^C(x)\right]\right| \le \mathbb{E}\left[\left|\gamma_T^C(x)\right|\right]\\ &= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{-T} |\gamma(x)|\, e^{-x^2/2}dx + \int_{T}^{\infty} |\gamma(x)|\, e^{-x^2/2}dx\right)\\ &\le \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{-T_1} Me^{a|x|}e^{-x^2/2}dx + \int_{T_1}^{\infty} Me^{a|x|}e^{-x^2/2}dx\right) < \frac{\varepsilon}{3}. \qquad (25)\end{aligned}$$

The third term of the right-hand part of (24) is also bounded by $\varepsilon/3$. This can be proved as follows. Let us denote by n_T the least index n such that $l_n > T$. Then

$$\begin{aligned}&\left|\frac{1}{m}\sum_{n=1}^m \gamma_T^c(l_n)\right| \le \frac{1}{m}\sum_{n=1}^m |\gamma_T^c(l_n)| \le \frac{1}{m}\sum_{n:|l_n|>T} |\gamma(l_n)| \le\\ &\le M\frac{1}{m}\sum_{n:|l_n|>T} e^{a|l_n|} = 2M\frac{1}{m}\sum_{n:l_n>T} e^{al_n} = 2M\frac{1}{m}\sum_{n=n_T}^{m} e^{al_n}\\ &= 2M\sum_{n=n_T}^{m-1}\frac{1}{m}e^{al_n} + 4M\frac{e^{al_m}}{2m} \le 4M\left(\sum_{n=n_T}^{m-1}\frac{1}{m}e^{al_n} + \frac{e^{al_m}}{2m}\right). \qquad (26)\end{aligned}$$

From the definition of l_n, for $n = n_T\ldots m-1$ it holds $1/m = (1/\sqrt{2\pi})\int_{l_n}^{l_{n+1}} e^{-x^2/2}dx$. Thus

$$\begin{aligned}\frac{1}{m}e^{al_n} &= \frac{1}{\sqrt{2\pi}}\int_{l_n}^{l_{n+1}} e^{-x^2/2}e^{al_n}dx\\ &= \frac{1}{\sqrt{2\pi}}\int_{l_n}^{l_{n+1}} e^{-x^2/4}e^{-x^2/4}e^{al_n}dx\\ &\le \frac{1}{\sqrt{2\pi}}\int_{l_n}^{l_{n+1}} e^{-x^2/4}e^{-l_n^2/4}e^{al_n}dx\,\frac{1}{\sqrt{2\pi}}\int_{l_n}^{l_{n+1}} e^{-x^2/4}dx \qquad (27)\end{aligned}$$

because $n \ge n_T$ implies $l_n > T \ge 4a$ which in turn implies $l_n^2/4 \ge al_n$. Furthermore, $1/2m = (1/\sqrt{2\pi})\int_{l_m}^{\infty} e^{-x^2/2}dx$. Therefore

$$\begin{aligned}\frac{e^{al_m}}{2m} &= \frac{1}{\sqrt{2\pi}}\int_{l_m}^{\infty} e^{-x^2/2}e^{al_m}dx = \frac{1}{\sqrt{2\pi}}\int_{l_m}^{\infty} e^{-x^2/4}e^{-x^2/4}e^{al_m}dx\\ &\le \frac{1}{\sqrt{2\pi}}\int_{l_m}^{\infty} e^{-x^2/4}dx. \qquad (28)\end{aligned}$$

The above arguments yield

$$\begin{aligned}&\left|\frac{1}{m}\sum_{n=1}^m \gamma_T^c(l_n)\right|\\ &\le 4M\left(\sum_{n=n_T}^{m-1}\frac{1}{\sqrt{2\pi}}\int_{l_n}^{l_{n+1}} e^{-x^2/4}dx + \frac{1}{\sqrt{2\pi}}\int_{l_m}^{\infty} e^{-x^2/4}dx\right)\\ &= 4M\frac{1}{\sqrt{2\pi}}\int_{l_{n_T}}^{\infty} e^{-x^2/4}dx \le 4M\frac{1}{\sqrt{2\pi}}\int_{T_2}^{\infty} e^{-x^2/4}dx < 4\frac{\varepsilon}{12} = \frac{\varepsilon}{3}. \qquad (29)\end{aligned}$$

The second term in the right-hand part of (24), $|\mathbb{E}[\gamma_T(x)] - \frac{1}{m}\sum_{n=1}^m \gamma_T(l_n)|$, can be rewritten as $|\int \gamma_T d\mu - \int \gamma_T d\mu_m|$ where μ is the standard normal distribution while μ_m is the distribution of the set $\{l_1,\ldots,l_m\}$. γ is continuous on the compact interval $[-T,T]$, thus it is bounded on it, therefore γ_T is bounded. Furthermore, γ_T

is μ-almost everywhere continuous (the only possible discontinuity points are $-T$ and T). Therefore (see the proof of Theorem 25.8 of [22]) $\lim_{m\to\infty}\int \gamma_T d\mu_m = \int \gamma_T d\mu$ if μ_m is shown to converge weakly to μ. Since F is everywhere continuous, to that end it must be shown that $\forall x \in \mathbb{R} \lim_{m\to\infty} F_m(x) = F(x)$, where F_m is the cdf of $\{l_1, \ldots, l_m\}$. If the cardinal number of a set A is denoted by $|A|$, then $F_m(x) = \frac{1}{m}|\{n : l_n \leq x\}| = \frac{1}{m}|\{n : F(l_n) \leq F(x)\}| = \frac{1}{m}|\{n : \frac{1}{2m} + \frac{n-1}{m} \leq F(x)\}| = \frac{1}{m}|\{n \leq mF(x) + 1/2\}|$. But $|\{n \leq mF(x) + 1/2\}| = mF(x) + \Delta_m$, where $|\Delta_m| \leq 1/2$, thus it follows that $F_m(x) \to F(x)$ as $m \to \infty$. Therefore it is possible to choose an integer m_0 such that for every $m > m_0$ it holds $|\mathbb{E}[\gamma_T(x)] - \frac{1}{m}\sum_{n=1}^{m} \gamma_T(l_n)| < \varepsilon/3$. Then, from (24) it follows that $|\mathbb{E}[\gamma(x)] - \frac{1}{m}\sum_{n=1}^{m} \gamma(l_n)| < \varepsilon$ for every $m > m_0$. The fact that it suffices for γ_T to be almost everywhere and bounded implies that the convergence holds true even if γ has a measure-zero set of discontinuity points. In this case the boundedness of γ_T follows from the fact that γ is of exponential order. ■

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problem," *ASME J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[3] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. New York: Dover, 2009.

[4] D. Simon, *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches*. Hoboken, NJ: Wiley, 2006.

[5] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.

[6] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian state estimation," *Proc. Inst. Elect. Eng. F*, vol. 140, no. 2, pp. 107–113, Apr. 1993.

[7] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space modules," *J. Computat. Graphical Stat.*, vol. 5, no. 1, pp. 1–25, Mar. 1996.

[8] M. S. Arulampalam, S. Maskel, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[9] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proc. IEEE*, vol. 95, no. 5, pp. 899–924, May 2007.

[10] G. L. Smith, S. F. Schmidt, and L. A. McGee, Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle NASA, Tech. Rep. TR R-135, 1962.

[11] T. Lefebvre, H. Bruyninckx, and J. Schutter, "Kalman filters for nonlinear systems: A comparison of performance," *Int. J. Control*, vol. 77, no. 7, pp. 639–653, May 2004.

[12] M. Simandl and J. Duník, "Derivative-free estimation methods: New results and performance analysis," *Automatica*, vol. 45, no. 7, pp. 1749–1757, Jul. 2009.

[13] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*,, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[14] M. St-Pierre and D. Gingras, "Comparison between the unscented Kalman filter and the extended Kalman filter for the position estimation module of an integrated navigation information system," in *Proc. 2004 IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 831–835.

[15] S. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Trans. Autom. Control*, vol. 45, no. 3, pp. 477–482, Mar. 2000.

[16] S. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[17] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Autom. Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[18] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*. New York: Academic Press, 1975.

[19] A. C. Charalampidis and G. P. Papavassilopoulos, "Computationally efficient Kalman filtering for a class of nonlinear systems," *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 483–491, Mar. 2011.

[20] K. Hsu, K. Poolla, and T. Vincent, "Identification of structured nonlinear systems," *IEEE Trans. Autom. Control*, vol. 53, no. 11, pp. 2497–2513, Dec. 2008.

[21] K. Wen and Z. Geng, "Stability of distributed heterogeneous systems with static nonlinear interconnections," *Syst. Control Lett.*, vol. 59, no. 11, pp. 680–686, 2010.

[22] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ: Wiley-Interscience, 1995.

[23] R. B. Ash, *Basic Probability Theory*. Mineola, NY: Dove, 2008 [Online]. Available: www.math.uiuc.edu/~r-ash/BPT.html