The universality of the Kalman filter: a conditional characteristic function perspective

Sandhya Rathore\(^1\), Shambhu N. Sharma\(^2\) and Shaival H. Nagarsheth\(^3\)
Electrical Engineering Department,
S.V. National Institute of Technology, Surat, 395007, Gujarat, India.
E-mail: \(^1\)sandhya.rathore.svnit@gmail.com, \(^2\)snsvolterra@gmail.com
\(^3\)shn411@gmail.com

Abstract
The universality of the celebrated Kalman filtering can be found in control theory. The Kalman filter has found its striking applications in sophisticated autonomous systems and smart products, which are attributed to its realization in a single complex chip. In this paper, we revisit the Kalman filter from the perspective of ‘conditional characteristic function evolution and Itô calculus’ and develop three Kalman filtering Theorems and their formal proof. Most notably, this paper reveals the following: (i) Kalman filtering equations are a consequence of the ‘evolution of conditional characteristic function’ for the linear stochastic differential system coupled with the linear discrete measurement system. (ii) The Kalman filtering is a consequence of the ‘stochastic evolution of conditional characteristic function’ for the linear stochastic differential system coupled with the linear continuous measurement system. (iii) The structure of the Kalman filter remains invariant under two popular stochastic interpretations, the Itô vs Stratonovich.

Keywords: Conditional characteristic function, Kalman filtering, the Itô vs Stratonovich, vec function.

Introduction
Norbert Wiener in 1921 proved that the Brownian motion has continuous sample path and non-differentiable with probability one [1]. Norbert Wiener developed the filtering equation using the convolution integral equation for linear observation equations. The Wiener filter is a signal estimation method. Suppose the Wiener filter is the LTI system, the Wiener filtering can be regarded as an infinite-dimensional LTI system, the transfer function would be irrational. In 1960, Kalman sowed the seed of the filtering of dynamic systems in which he considered linear stochastic differential equation coupled with linear observation equation [2]. The result was published in the Journal of Basic Engineering. This attracted attention in greater circles via its Apollo mission application. The Kalman filtering has found appealing applications in smartphones as well. Thus, the Kalman filtering is ubiquitous and is in everybody’s life. Suppose the Kalman filter is an LTI system, where the observation is the input and the filtered estimate is the output, that can be regarded as the finite-dimensional and the transfer function of the Kalman filter would be rational. The Kalman filtering and its discoverer’s legacy can be found in appealing compilations and papers, e.g. [3]. Kalman meets Shannon [4] in noisy Gaussian channels, the ubiquitous Riccati equation is a part of the coupled Kalman filtering equations for the continuous state-continuous measurement systems. That reveals the conceptual depth of the Kalman filter. After combining the celebrated Wiener-Höpf equation, Wiener filter,
linear stochastic differential equation, linear measurement equation coupled with the orthogonal projection Lemma, we arrive at the Kalman gain and coupled filtering equations. Furthermore, the notion of non-linear Kalman filtering was introduced. The non-linear filtering has greater conceptual depth as well as involves greater mathematical rigour. Non-linear filtering theory hinges on the filtering density evolution equation and the stochastic evolution of the conditional expectation of the scalar function. A remarkable paper on non-linear filtering can be traced back to a pioneering work of 1967, H J Kushner. That is popularly known the Kushner-Stratonovich equation, see [5]. An alternative interpretation of the Kushner-Stratonovich is FKK filtering [6]. In this paper, we restrict our discussions to the celebrated Kalman filtering.

This paper is inspired from the following two papers: (i) A simple concept with new perspectives leads to surprising results ([7]) (ii) Moore and Anderson published an interesting paper [8], the Kalman filtering: whence, what and whither. In their paper, they explained ‘the Kalman filtering by extending it in non-linear filtering perspective with some discussions ([8], p. 52). In this paper, we answer the question, whither of the Kalman filtering? in greater detail. This paper sketches the answer to the question in greater detail with conditional characteristic perspective. The conditional characteristic function evolution, a new perspective in the Kalman filtering setting, is not available yet in literature. Secondly, the Kalman filtering can be sketched from the stochastic evolution of the conditional characteristic function for the linear Stochastic Differential Equation (SDE) with linear observation equation as well.

Notation: For the action of the conditional expectation operator, we adopt three notations, i.e. $\tilde{x}_t = E(x_t|Y_t) = \langle x_t \rangle$. A choice of the notation hinges on the structure of conditional expectation evolution equations.

The Kalman filter: a conditional characteristics function perspective

In this paper, we wish to construct the Kalman filter from the conditional characteristic function perspective by exploiting the Itô differential rule for the Wiener process [9], which is not available yet in literature. This paper is inspired from the fact that the older, but the celebrated result with new perspectives, produces greater insights. In this paper, the Itô calculus [10] and conditional characteristic function [5] are the cornerstone formalisms of the construction of the coupled Kalman filtering equations. Thus, we take a little pause and explain very briefly the influence of the Itô calculus and conditional characteristic function [2] in stochastic processes. The Itô calculus, a 1945 discovery, sowed the seed of stochastic calculus, a new branch of mathematics. The beauty, power and universality of the Itô calculus can be found in a famous paper of Kunita [10]. Later that the Itô calculus was adjudged for the inaugural Gauss prize. Secondly, the conditional characteristic function is a conditional moment generating function that forms the cornerstone formalism for the Central Limit Theorem of stochastic processes.

Consensus on the Itô-Stratonovich dilemma recommends ‘the Stratonovich SDE is expected for stochastic processes in typical, continuous and real physical systems’. On the other hand, the Itô SDE is the correct stochastic calculus for stochastic finance and theoretical biology. The non-linear filtering would be different for two perspectives, the
Itô and the Stratonovich. Interestingly, the Kalman filtering is invariant under the Itô and Stratonovich perspectives, see [11]. In this paper, we construct three Theorems, the first is about the evolution of conditional characteristic function for the Kalman filtering for the continuous state-discrete measurement system. The second Theorem is about the stochastic evolution of the conditional characteristic function [5] for the Kalman filtering for the continuous state-continuous measurement systems. The third Theorem is about the stochastic evolution of the bilinear Kalman filter for the continuous state-continuous measurement system, where the Kalman filtering is a special case of the bilinear Kalman filtering. This paper chiefly hinges on the evolution of conditional characteristic function and the evolution of conditional characteristic function holds for the continuous state. Thus, the continuous state obeying the Itô SDE is the subject of investigations in lieu of the stochastic difference equation.

**Theorem 1.** Consider the filtering model for the continuous state-discrete measurement system is

\[
dx_i(t) = \sum_a A_{ia}(t)x_a(t)dt + \sum_i G_{i\phi}(t)dW_{i}(t)
\]

\[
y_i(t_k) = \sum_i C_{i\phi}(t_k)x_i(t_k) + V_i(t_k),
\]

where the vector Brownian motion process \( W = \{W_t \in R^n, 0 \leq t < \infty \} \). Consider \( x_i \) and \( A_i \) denote the state vector and system state matrix respectively. The term \( G(t)dW_i \) denotes the vector stochastic correction term with var \( (dW_i) = Idt \). The term \( y_{t_k} \) indicates a noisy observation vector. The observation noise \( V_i \) is \( N(0, R_k) \), where the term \( R_k \) is a real-valued observation noise variance matrix. Then, the evolution of conditional characteristic function of the above filtering model between the observation instants, \( \tau \), where \( t_{k-1} \leq \tau < t_k \) becomes

\[
d\langle \exp(s^T x_\tau) \rangle = \left\langle \sum_p \sum_a A_{pa}(\tau)x_a \frac{\partial \exp(s^T x_\tau)}{\partial x_p} + \frac{1}{2} \sum_{p,q} (GG^T)_{pq}(\tau) \frac{\partial^2 \exp(s^T x_\tau)}{\partial x_p \partial x_q} \right\rangle d\tau.
\]

Alternatively,

\[
d\langle \exp(s^T x_\tau) \rangle = \left( \sum_p \sum_a A_{pa}(\tau) \left\langle x_a(\tau) \frac{\partial \exp(s^T x_\tau)}{\partial x_p} \right\rangle \right.
\]

\[
+ \left. \frac{1}{2} \sum_{p,q} (GG^T)_{pq}(\tau) \left\langle \frac{\partial^2 \exp(s^T x_\tau)}{\partial x_p \partial x_q} \right\rangle \right) d\tau,
\]

where the characteristic function \( \langle \exp(s^T x_\tau) \rangle \) is \( E(\exp(s^T x_\tau) | Y_{t_{k-1}}) \). The conditional expectation operator \( E \) accounts for the accumulated observation \( Y_{t_{k-1}} = \{y_n, 0 \leq n \leq k - 1\} \), \( t_{k-1} < \tau \leq t_k \) and \( t_{k-1}, t_k \) are the observation instants.

The Kalman filtering between the observations becomes a special case of the conditional characteristic function evolution. On the other hand, the Kalman filtering at the observation is not a consequence of the evolution of conditional characteristic function.
Proof. The proof of the conditional characteristic function evolution unaccounting for noisy observables can be sketched by exploiting the property of the differential and the expectation operators, Itô calculus. For the action of the conditional expectation operator $E(.)$ on lengthier terms, we adopt another notation $\tau$ as well.

\[ \langle d \exp(s^T x_i) \rangle = \langle d \varphi(x_i,s) \rangle = \langle \varphi(x_{i+dt},s) - \varphi(x_i,s) \rangle = \psi(t + dt, s) - \psi(t, s). \]

In the sense of characteristic function, the variable $s$ becomes $j\omega$, the input argument of the Fourier Transform. Now,

\[ \langle \exp(s^T x_i) \rangle = \psi(t, s) = \langle \varphi(s,x_i) \rangle, \]

\[ d\langle \exp(s^T x_i) \rangle = \psi(t + dt, s) - \psi(t, s) = \langle d \exp(s^T x_i) \rangle. \]

(2)

The above relation reveals the fact that the differential and the conditional expectation operator can be interchanged. Since we wish to develop the conditional characteristic function evolution at the instant $\tau$, where observations are not available, we adopt the notation $\tau$ in place of the time variable $t$. Thus,

\[ d \exp(s^T x_i) = \left( \sum_p \sum_{\alpha} A_{\alpha p}(\tau)x_{\alpha}(\tau) \frac{\partial \exp(s^T x_i)}{\partial x_p} + \frac{1}{2} \sum_{p,q} (GG^T)_{pq}(\tau) \frac{\partial^2 \exp(s^T x_i)}{\partial x_p \partial x_q} \right) d\tau \]

\[ + \sum_{p,\phi} G_{p\phi}(\tau) \frac{\partial \exp(s^T x_i)}{\partial x_p} dW_\phi. \]

(3)

The above equation is a consequence of the Itô differential rule (Karatzas and Shreve [9]) for the scalar function $\exp(s^T x_i)$. From equation (2) and equation (3), we have

\[ d\langle \exp(s^T x_i) \rangle = \left\{ \sum_p \sum_{\alpha} A_{\alpha p}(\tau)x_{\alpha}(\tau) \frac{\partial \exp(s^T x_i)}{\partial x_p} + \frac{1}{2} \sum_{p,q} (GG^T)_{pq}(\tau) \frac{\partial^2 \exp(s^T x_i)}{\partial x_p \partial x_q} \right\} d\tau \]

\[ + \left\{ \sum_{p,\phi} G_{p\phi}(t) \frac{\partial \exp(s^T x_i)}{\partial x_p} dW_\phi \right\} \]

(4)

Since the conditional expectation operator is a linear operator and

\[ \left\{ \sum_{p,\phi} G_{p\phi}(t) \frac{\partial \exp(s^T x_i)}{\partial x_p} dW_\phi \right\} = 0. \]

Equation (4) boils down to

\[ d\langle \exp(s^T x_i) \rangle = \left( \sum_p \sum_{\alpha} A_{\alpha p}(\tau)x_{\alpha}(\tau) \frac{\partial \exp(s^T x_i)}{\partial x_p} + \frac{1}{2} \sum_{p,q} (GG^T)_{pq}(\tau) \frac{\partial^2 \exp(s^T x_i)}{\partial x_p \partial x_q} \right) d\tau \]

\[ = \left( \sum_p \sum_{\alpha} A_{\alpha p}(\tau)x_{\alpha}(\tau) \frac{\partial \exp(s^T x_i)}{\partial x_p} \right) d\tau + \frac{1}{2} \sum_{p,q} (GG^T)_{pq}(\tau) \left\{ \frac{\partial^2 \exp(s^T x_i)}{\partial x_p \partial x_q} \right\} d\tau. \]

Thus, we arrive at equation (1) of the Theorem 1 of the paper.

\[ QED \]

Now, consider $\exp(s^T x_i) = x_i(t)$. Then, the first term of the right-hand side will
contribute and the last term of the right-hand side will vanish. As a result of this, equation (1) becomes
\[ d\langle x_i(\tau) \rangle = \sum_{a} A_{ia}(\tau)\langle x_a(\tau) \rangle d\tau, \]
Alternatively
\[ d\langle x(\tau) \rangle = A_{\tau} \langle x_i \rangle d\tau. \] (5)
Consider \( \exp(s^T x_i) = x_i x_j \), we have
\[ d\langle x_i x_j \rangle = \left( \sum_{a} A_{ia}(\tau)\langle x_a(\tau) x_j(\tau) \rangle + \sum_{a} A_{ja}(\tau)\langle x_a(\tau)x_i(\tau) \rangle \right) \]
\[ + \frac{1}{2} \sum_{p,q}(GG^T)_{pq}(\tau) \left( \frac{\partial^2 x_j}{\partial x_p \partial x_q} \right) d\tau. \]
\[ = \left( \sum_{a} A_{ia}(\tau)\langle x_a(\tau) x_j(\tau) \rangle + \sum_{a} A_{ja}(\tau)\langle x_a(\tau)x_i(\tau) \rangle \right) \]
\[ + \frac{1}{2} \sum_{p,q}(GG^T)_{pq}(\tau) \left( \delta_{ip}\delta_{jq} + \delta_{iq}\delta_{jp} \right) d\tau. \]
\[ = \left( \sum_{a} A_{ia}(\tau)\langle x_a(\tau) x_j(\tau) \rangle + \sum_{a} A_{ja}(\tau)\langle x_a(\tau)x_i(\tau) \rangle \right) + (GG^T)_{ij}(\tau) d\tau. \]
After combining the above equation with equation (5), we get
\[ dp_{ij} = d\langle x_i x_j \rangle - \langle x_i \rangle d\langle x_j \rangle - \langle x_j \rangle d\langle x_i \rangle - d\langle x_j \rangle d\langle x_i \rangle \]
\[ = \left( \sum_{a} A_{ia}(\tau)\langle x_a(\tau) x_j(\tau) \rangle + \sum_{a} A_{ja}(\tau)\langle x_a(\tau)x_i(\tau) \rangle + (GG^T)_{ij}(\tau) \right) \]
\[ - \sum_{a} A_{ia}(\tau)\langle x_j(\tau) x_a(\tau) \rangle - \sum_{a} A_{ja}(\tau)\langle x_i(\tau)x_a(\tau) \rangle \right) d\tau \]
\[ = \left( \sum_{a} A_{ia}(\tau)P_{ja}(\tau) + \sum_{a} P_{ia}(\tau)A_{ja}(\tau) + (GG^T)_{ij}(\tau) \right) d\tau. \]
Alternatively,
\[ dp_{\tau} = (A_{\tau}P_{\tau} + P_{\tau}A_{\tau}^T + (GG^T)_{ij}(\tau)) d\tau. \] (6)
Equation (1) describes the evolution of characteristic function for the Kalman filtering between the observations. Equations (5)-(6) are the coupled Kalman filtering equations between the observations. The Kalman filtering at the observation can be sketched using the conditional probability density and conditional expectation at the observations instants \( t_{k-1} \) and \( t_k \). A proof of the Kalman filtering equations at the observation can be found in a famous book authored by Andrew H Jazwinski on Stochastic Processes and Non-linear filtering, see [12] as well. That is not a consequence of the evolution of conditional characteristic function. Here, we omit the greater detail. For the vector case, we have equations (7)-(8).
\[ \tilde{x}_{t_i} = \tilde{x}_{t_{i-1}} + P_{t_i}^{t_{i-1}}C_{t_i}^{T}(C_{t_i}P_{t_i}^{t_{i-1}}C_{t_i}^T + R_{t_i})^{-1}(y_{t_i} - C_{t_i}\tilde{x}_{t_{i-1}}), \] (7)
\[ P_{t_i} = P_{t_{i-1}} - P_{t_{i-1}}^{t_{i-1}}C_{t_i}^{T}(C_{t_i}P_{t_i}^{t_{i-1}}C_{t_i}^T + R_{t_i})^{-1}C_{t_i} P_{t_{i-1}}^{t_{i-1}}. \] (8)
Note that \( \tilde{x}_{t_i} = E(x_{t_i} | Y_{t_{i-1}}) \), \( \bar{x}_{t_i} = E(x_{t_i} | Y_{t_i}) \). The Kalman filtering for the continuous state-discrete measurement system is given by the following set of four equations by adopting the vec function setting as well, i.e.
\[ d\vec{\langle x(\tau) \rangle} = \text{vec}(A_{\tau}\{x_{\tau}\}) d\tau = (\{x_{\tau}\}^T \otimes I)\text{vec}(A_{\tau}) d\tau, \quad (9a) \]
\[ d\vec{\langle P_{\tau} \rangle} = ((I \otimes A_{\tau} + A_{\tau} \otimes I)\text{vec}(P_{\tau}) + \text{vec}(GG^T(\tau))) d\tau, \quad (9b) \]
\[ \text{vec}(\tilde{x}_{i_k}) = \text{vec}(\tilde{x}_{i_{k-1}}) + \begin{pmatrix} (y_{i_k} - C_{i_k}\tilde{x}_{i_k})^T \end{pmatrix} (C_{i_k}P_{i_{k-1}}C_{i_k}^T + R_k)^{-1}C_{i_k} \otimes I)\text{vec}(P_{i_{k-1}}). \quad (9c) \]
\[ \text{vec}(P_{i_k}) = \text{vec}(P_{i_{k-1}}) - (P_{i_{k-1}}C_{i_k}^T(C_{i_k}P_{i_{k-1}}C_{i_k}^T + R_k)^{-1}C_{i_k} \otimes I)\text{vec}(P_{i_{k-1}}). \quad (9d) \]

Equations (9a)-(9d) describe the Kalman filtering for the continuous state-discrete measurement system using the vec function interpretation. This paper is intended to develop the Kalman filtering in the conditional characteristic function evolution setting. On the other hand, the conditional characteristic function evolution setting is not applicable to the discrete state-discrete measurement system. Thus, in this paper, the Kalman filtering for the discrete state-discrete measurement system is not the subject of investigations. A good source on the Kalman filtering for the discrete state-discrete measurement can be found in Moore and Anderson [3].

**Remark 1.** The stationary Kalman filtering for the continuous state-discrete measurement has algebraic variance equations and constant coefficients. Conditional variance equations for the between the observations case and at the observation case become

\[ 0 = (I \otimes A + A \otimes I)\text{vec}(P_{i_{k-1}}) + \text{vec}(GG^T), \]
\[ \text{vec}(P_{i_k}) = \text{vec}(P_{i_{k-1}}) - (P_{i_{k-1}}C_{i_k}^T(C_{i_k}P_{i_{k-1}}C_{i_k}^T + R_k)^{-1}C_{i_k} \otimes I)\text{vec}(P_{i_{k-1}}) \]
respectively.

**Theorem 2.** Consider the filtering model for the continuous state-continuous measurement

\[ dx_{i}(t) = \sum_{\alpha} A_{\alpha}(t)x_{\alpha}(t)dt + \sum_{\alpha} G_{\alpha}(t)dW_{\alpha}(t) \]
\[ dz_{i}(t) = \sum_{\alpha} C_{\alpha}(t)x_{\alpha}(t)dt + d\eta_{i}, \]
where the vector Brownian motion process \( W = \{W_{i} \in R^{n}, 0 \leq t < \infty\} \). Consider \( x_{i} \) and \( A_{i} \) denote the state vector and system state matrix respectively. The term \( G(t)dW_{i} \) denotes the vector stochastic correction term with \( \text{var}(dW_{i}) = Idt \). The term \( z_{i} \) indicates the noisy observation vector that is available continuously. The observation noise statistics is \( N(0, \phi_{q}, dt) \). Then, the stochastic evolution of conditional characteristic function of the above filtering model assumes the structure of an SDE, i.e.

\[ d\exp(s^T x_{i}) = \left\{ \sum_{p \alpha} A_{p\alpha}(t)x_{\alpha} \frac{\partial \exp(s^T x_{i})}{\partial x_{p}} + \frac{1}{2} \sum_{p \alpha q \alpha} \{GG^T\}_{pq}(t) \frac{\partial^2 \exp(s^T x_{i})}{\partial x_{p}\partial x_{q}} \right\} dt \]
\[ + (\{\exp(s^T x_{i})x_{i}C_{i}^T\} - \{\exp(s^T x_{i})\{x_{i}C_{i}^T\}\}\phi^{-1}_{q}(dz_{i} - C_{i}\{x_{i}\}dt). \]
\[ \]
\[ = \left( \sum_{p \alpha} A_{p\alpha}(t)\{x_{\alpha}(\tau) \frac{\partial \exp(s^T x_{i})}{\partial x_{p}} \} + \frac{1}{2} \sum_{p \alpha q \alpha} \{GG^T\}_{pq}(t) \frac{\partial^2 \exp(s^T x_{i})}{\partial x_{p}\partial x_{q}} \right) dt \]
\[ + \sum_{\alpha, \beta} \left( \text{exp}(s^T x_{i})\sum_{\gamma} C_{\alpha\gamma}(t)x_{\gamma}(t) - \text{exp}(s^T x_{i})\sum_{\gamma} C_{\alpha\gamma}(t)x_{\gamma}(t) \right) \]
\[
\times (\varphi^{-1}_{\eta})_{ab} (dz_{\beta} - \sum_{\gamma} C_{\rho_{\gamma}}(t) \langle x_{\gamma} \rangle dt),
\]

where the conditional characteristic function \( \langle \exp(s^{T} x_{i}) \rangle = E(\exp(s^{T} x_{i}) | Y_t) \). The conditional expectation operator \( E \) accounts for the accumulated observation \( Y_t = \{ z_{\tau}, t_0 \leq \tau \leq t \} \). The Kalman filtering becomes a special case of stochastic evolution of the conditional characteristic function, equation (10).

**Proof.** Here, we explain succinctly the proof of Theorem 2 of the paper. For the filtering model of Theorem 2, the stochastic evolution of conditional characteristic function assumes the structure of a stochastic integro-differential equation. One can construct the stochastic evolution of conditional characteristic function using Itô differential rule and the multiplication theorem of densities. As a result of this, we get equation (10) of Theorem 2 of the paper.

Notably, we can construct the stochastic evolution of conditional characteristic function by an alternative backward method. First, the method involves the construction of the stochastic evolution of conditional characteristic function for the Itô SDE using Itô differential rule and the multiplication theorem of densities (Liptser and Shiryaev [5]). As a result of this, the stochastic evolution of conditional characteristic function for the non-linear Itô SDE becomes

\[
d\langle \exp(s^{T} x_{i}) \rangle = \left( \sum_{i} f_{i}(t, x_{i}) \frac{\partial \exp(s^{T} x_{i})}{\partial x_{i}} \right) dt + \frac{1}{2} \left( \sum_{i} (GG^{T})_{i}(t, x_{i}) \frac{\partial^{2} \exp(s^{T} x_{i})}{\partial x_{j}\partial x_{j}} \right) dt + \langle \varphi h^{T} \rangle - \langle \varphi \rangle \langle h^{T} \rangle \varphi^{-1}_{\eta} (dz_{i} - \langle h \rangle dt),
\]

and the stochastic evolution for the linear Itô SDE is

\[
d\langle \exp(s^{T} x_{i}) \rangle = \left( \sum_{p} \sum_{a} A_{pa}(t) \left[ \frac{\partial \exp(s^{T} x_{i})}{\partial x_{p}} \right] + \frac{1}{2} \left( \sum_{p,q} (GG^{T})_{pq}(t) \frac{\partial^{2} \exp(s^{T} x_{i})}{\partial x_{p}\partial x_{q}} \right) \right) dt + \sum_{a,\beta} \left( \exp(s^{T} x_{i}) \sum_{\gamma} C_{a\gamma}(t) x_{\gamma}(t) - \exp(s^{T} x_{i}) \left( \sum_{\gamma} C_{a\gamma}(t) x_{\gamma}(t) \right) \right) \times (\varphi^{-1}_{\eta})_{ab} (dz_{\beta} - \sum_{\gamma} C_{\rho_{\gamma}}(t) \langle x_{\gamma} \rangle dt).
\]

Thus, we arrive at equation (10) of Theorem 2 of the paper.

**QED**

That is a special case of equation (11) of the paper. Equation (10) can be regarded as the stochastic evolution of conditional characteristic function for the Kalman filtering for the continuous state-continuous measurement system. The term

\[
(\langle \exp(s^{T} x_{i}) x_{i}^{T} C_{i}^{T} \rangle - \langle \exp(s^{T} x_{i}) \rangle \langle x_{i}^{T} C_{i}^{T} \rangle) \varphi^{-1}_{\eta}
\]

of equation (10) can be regarded as the gain of the Kalman filter from the conditional characteristic function perspective.

Consider \( \exp(s^{T} x_{i}) = x_{i}(t) \). Then, the first term of the right-hand side will contribute and the second term of the right-hand side will vanish. The last term of the above will contribute. Thus, equation (10) reduces to
\[ d\langle x_i \rangle = \left( \sum_p \sum_a A_{p\alpha}(t) \left( x_{a}(t) \frac{\partial x_i}{\partial x_p} \right) + \frac{1}{2} \sum_{p,q} (GG^T)_{pq}(t) \left( \frac{\partial^2 x_i}{\partial x_p \partial x_q} \right) \right) dt \]

\[ \quad + \sum_{a,\beta} \left( \left( x_i \sum_{\gamma} C_{a\gamma}(t)x_{\gamma}(t) \right) - \langle x_i \rangle \left( \sum_{\gamma} C_{a\gamma}(t)x_{\gamma}(t) \right) \right) \times (\varphi^{-1}_{\eta})_{a\beta} (dz_\beta - \sum_{\gamma} C_{\beta\gamma}(t)x_{\gamma}(t) dt) \]

Consider \( \exp(s^T x_i) = x_i, x_j \), equation (10) boils down to

\[ d\langle x_i, x_j \rangle = \left( \sum_{\alpha} A_{i\alpha}(t) \langle x_{\alpha}(t)x_j(t) \rangle + \sum_{\alpha} A_{j\alpha}(t) \langle x_{\alpha}(t)x_i(t) \rangle \right) \]

\[ \quad + \frac{1}{2} \sum_{p,q} (GG^T)_{pq}(t) \left( \frac{\partial^2 x_i, x_j}{\partial x_p \partial x_q} \right) dt \]

\[ \quad + \sum_{a,\beta,\gamma} \left( \left( x_i, x_j \sum_{\gamma} C_{a\gamma}(t)x_{\gamma}(t) \right) - \langle x_i, x_j \rangle \left( \sum_{\gamma} C_{a\gamma}(t)x_{\gamma}(t) \right) \right) \times (\varphi^{-1}_{\eta})_{a\beta} (dz_\beta - \sum_{\gamma} C_{\beta\gamma}(t)x_{\gamma}(t) dt). \] (13)

After embedding equation (A.2) in equation (13), the above boils down to

\[ \left( \sum_{\alpha} A_{i\alpha}(t) \langle x_{\alpha}(t)x_j(t) \rangle + \sum_{\alpha} A_{j\alpha}(t) \langle x_{\alpha}(t)x_i(t) \rangle \right) \]

\[ \quad + (GG^T)_{ij}(t) dt + \sum_{a,\beta,\gamma} \left( (P_{ij} \langle x_j \rangle + P_{ji} \langle x_i \rangle) \right) C_{a\gamma}(t) \]

\[ \times (\varphi^{-1}_{\eta})_{a\beta} (dz_\beta - \sum_{\gamma} C_{\beta\gamma}(t)x_{\gamma}(t) dt), \]

\[ \langle x_i(t) \rangle d\langle x_j(t) \rangle = \left( \sum_{\alpha} A_{j\alpha}(t) \langle x_{\alpha}(t) \rangle \right) \]

\[ \quad + \sum_{a,\beta,\gamma} \left( \sum_{\gamma} C_{a\gamma}(t)P_{ji} \langle x_j(t) \rangle \right) (\varphi^{-1}_{\eta})_{a\beta} (dz_\beta - \sum_{\gamma} C_{\beta\gamma}(t)x_{\gamma}(t) dt), \] (14)

\[ \langle x_j(t) \rangle d\langle x_i(t) \rangle = \left( \sum_{\alpha} A_{i\alpha}(t) \langle x_{\alpha}(t) \rangle \right) \]

\[ \quad + \sum_{a,\beta,\gamma} \left( \sum_{\gamma} C_{a\gamma}(t)P_{ij} \langle x_j(t) \rangle \right) (\varphi^{-1}_{\eta})_{a\beta} (dz_\beta - \sum_{\gamma} C_{\beta\gamma}(t)x_{\gamma}(t) dt), \] (15)
\[ d\langle x(t) \rangle d\langle x(t) \rangle = \left( \sum_{\alpha} A_{\alpha} (t) \langle x_{\alpha}(t) \rangle \right) dt + \sum_{\alpha, \beta, \gamma} \left( \sum_{\gamma} C_{\alpha \gamma}(t) P_{\gamma \beta} \right) (\varphi_{\eta}^{-1})_{\beta \gamma} \left( d\langle x_{\gamma}(t) \rangle - \sum_{\gamma} C_{\beta \gamma}(t) \langle x_{\gamma}(t) \rangle dt \right) \times \left( \sum_{\alpha} A_{\alpha \beta}(t) \langle x_{\alpha}(t) \rangle \right) dt + \sum_{\alpha, \beta, \gamma} \left( \sum_{\gamma} C_{\alpha \gamma}(t) P_{\gamma \beta} \right) (\varphi_{\eta}^{-1})_{\beta \gamma} \left( d\langle x_{\gamma}(t) \rangle - \sum_{\gamma} C_{\beta \gamma}(t) \langle x_{\gamma}(t) \rangle dt \right) \times \left( \sum_{\alpha} A_{\alpha \beta}(t) \langle x_{\alpha}(t) \rangle \right) dt \]

After combining equations (13)-(16), we have

\[ dP_{\eta} = d\langle x_{x}(t) \rangle - \langle x_{i}(t) \rangle d\langle x_{i}(t) \rangle - \langle x_{j}(t) \rangle d\langle x_{j}(t) \rangle - d\langle x_{i}(t) \rangle d\langle x_{j}(t) \rangle = \left( \sum_{\alpha} A_{\alpha}(t) \langle x_{\alpha}(t) x_{\beta}(t) \rangle \right) + \sum_{\alpha} A_{\alpha}(t) \langle x_{\alpha}(t) x_{\beta}(t) \rangle \]

\[ + (GG^T)_{\beta \gamma} (t) dt + \sum_{\alpha, \beta, \gamma} \left( \sum_{\gamma} (P_{\gamma \beta} \langle x_{\gamma}(t) \rangle + P_{\gamma \beta} \langle x_{\gamma}(t) \rangle \right) C_{\alpha \gamma}(t) \]

\[ \times (\varphi_{\eta}^{-1})_{\beta \gamma} \left( d\langle x_{\gamma}(t) \rangle - \sum_{\gamma} C_{\beta \gamma}(t) \langle x_{\gamma}(t) \rangle dt \right) \]

\[ - \sum_{\alpha} A_{\alpha \beta}(t) \langle x_{\alpha}(t) x_{\beta}(t) \rangle \]

\[ - \sum_{\alpha, \beta, \gamma} \left( \sum_{\gamma} C_{\alpha \gamma}(t) P_{\gamma \beta} \langle x_{\gamma}(t) \rangle \right) (\varphi_{\eta}^{-1})_{\beta \gamma} \left( d\langle x_{\gamma}(t) \rangle - \sum_{\gamma} C_{\beta \gamma}(t) \langle x_{\gamma}(t) \rangle dt \right) \]

\[ - \sum_{\alpha, \gamma} \left( \sum_{\gamma} C_{\alpha \gamma}(t) P_{\gamma \beta} \langle x_{\gamma}(t) \rangle \right) \langle x_{\beta}(t) \rangle \forall \alpha \beta \gamma \]

\[ = \sum_{\alpha} A_{\alpha}(t) \left( \langle x_{\alpha}(t) x_{\beta}(t) \rangle - \langle x_{\beta}(t) \rangle \langle x_{\gamma}(t) \rangle \right) \]

\[ + \sum_{\alpha} A_{\alpha \beta}(t) \left( \langle x_{\alpha}(t) x_{\beta}(t) \rangle - \langle x_{\beta}(t) \rangle \langle x_{\gamma}(t) \rangle \right) + (GG^T)(t) dt \]

\[ + \sum_{\alpha, \beta, \gamma} \left( \sum_{\gamma} (P_{\gamma \beta} \langle x_{\gamma}(t) \rangle + P_{\gamma \beta} \langle x_{\gamma}(t) \rangle \right) C_{\alpha \gamma}(t) (\varphi_{\eta}^{-1})_{\beta \gamma} \left( d\langle x_{\gamma}(t) \rangle - \sum_{\gamma} C_{\beta \gamma}(t) \langle x_{\gamma}(t) \rangle dt \right) \]

\[ - \sum_{\alpha, \beta, \gamma} \left( \sum_{\gamma} C_{\alpha \gamma}(t) P_{\gamma \beta} \langle x_{\gamma}(t) \rangle \right) (\varphi_{\eta}^{-1})_{\beta \gamma} \left( d\langle x_{\gamma}(t) \rangle - \sum_{\gamma} C_{\beta \gamma}(t) \langle x_{\gamma}(t) \rangle dt \right) \]

\[ = \sum_{\alpha} A_{\alpha}(t) \left( \langle x_{\alpha}(t) x_{\beta}(t) \rangle - \langle x_{\beta}(t) \rangle \langle x_{\gamma}(t) \rangle \right) + \sum_{\alpha} A_{\alpha \beta}(t) \left( \langle x_{\alpha}(t) x_{\beta}(t) \rangle - \langle x_{\beta}(t) \rangle \langle x_{\gamma}(t) \rangle \right) + (GG^T)(t) dt \]

\[ + \sum_{\alpha, \beta, \gamma} \left( \sum_{\gamma} (P_{\gamma \beta} \langle x_{\gamma}(t) \rangle + P_{\gamma \beta} \langle x_{\gamma}(t) \rangle \right) C_{\alpha \gamma}(t) (\varphi_{\eta}^{-1})_{\beta \gamma} \left( d\langle x_{\gamma}(t) \rangle - \sum_{\gamma} C_{\beta \gamma}(t) \langle x_{\gamma}(t) \rangle dt \right) \]

\[ - \sum_{\alpha, \beta, \gamma} \left( \sum_{\gamma} C_{\alpha \gamma}(t) P_{\gamma \beta} \langle x_{\gamma}(t) \rangle \right) (\varphi_{\eta}^{-1})_{\beta \gamma} \left( d\langle x_{\gamma}(t) \rangle - \sum_{\gamma} C_{\beta \gamma}(t) \langle x_{\gamma}(t) \rangle dt \right) \]

9
\[
- \sum_{\alpha, \beta} \sum_{i} C_{i\alpha}(t) P_{i\beta}(t) \langle x_{i}(t) \rangle (\varphi_{i\beta}^{-1})_{\alpha \theta}(dz_{\beta} - \sum_{\gamma} C_{\beta\gamma}(t) \langle x_{\gamma} \rangle dt)
\]
\[
- \sum_{\alpha, \beta} \sum_{i} C_{i\alpha}(t) P_{i\beta}(t) \left( \sum_{\gamma} C_{\beta\gamma}(t) P_{\gamma\beta}(t) \right) (\varphi_{i\beta}^{-1})_{\alpha \theta} dt
\]
\[
= \left( \sum_{i} A_{i\alpha}(t) P_{i\alpha}(t) + \sum_{i} P_{i\alpha}(t) A_{i\alpha}(t) + (GG^{T})_{ij}(t) \right)
\]
\[
- \sum_{\alpha, \beta} \sum_{i} C_{i\alpha}(t) P_{i\beta}(t) \left( \sum_{\gamma} C_{\beta\gamma}(t) P_{\gamma\beta}(t) \right) (\varphi_{i\beta}^{-1})_{\alpha \theta} dt.
\]

Alternatively, the standard Kalman filtering equations, the above equation coupled with equation (12), are recast as
\[
d\langle x_{i} \rangle = A_{i} \langle x_{i} \rangle dt + P_{i} C_{i}^{T} \varphi_{i}^{-1}(dz_{i} - C_{i} \langle x_{i} \rangle dt),
\]
\[
dP_{i} = (A_{i}P_{i} + P_{i}A_{i}^{T} + (GG^{T})(t) - P_{i}C_{i}^{T} \varphi_{i}^{-1}C_{i}P_{i})dt.
\]

The Kalman filtering for the continuous state-continuous state measurement system is given by the following set of two equations by adopting the vec function setting, i.e.
\[
dvec(\langle x(t) \rangle) = ((\langle x_{i} \rangle^{T} \otimes I)vec(A_{i}))dt + ((dz_{i}^{T} - \langle x_{i} \rangle C_{i}^{T} dt)\varphi_{i}^{-1}C_{i} \otimes I)vec(P_{i}), \quad (17a)
\]
\[
dvec(P_{i}) = ((I \otimes A_{i} + A_{i} \otimes I)vec(P_{i}) + vec((GG^{T})(t)) - (P_{i}C_{i}^{T} \varphi_{i}^{-1}C_{i} \otimes I)vec(P_{i})). \quad (17b)
\]

Equations (17a)-(17b) describe the Kalman filtering for the continuous state-continuous measurement system using the vec function interpretation.

**Remark 2.** The stationary Kalman filtering for the continuous state-continuous measurement has algebraic variance equation and constant coefficients. Alternatively, the conditional variance equation is
\[
0 = (I \otimes A + A \otimes I)vec(P_{i}) + vec(GG^{T}) - (PC_{i}^{T} \varphi_{i}^{-1}C_{i} \otimes I)vec(P_{i}).
\]

The following table 1 sums up five systems-theoretic properties of the Kalman filtering. Some of them are scattered in the literature and some of them are relatively very less known.

| Properties | Kalman filtering for the continuous state-continuous measurement system | Kalman filtering for the continuous state-discrete measurement system |
|------------|------------------------------------------------------------------------|---------------------------------------------------------------------|
| Evolution of conditional characteristic function [5] | Ordinary differential equation | Stochastic integro-differential equation |
| Riccati equation | Does not hold | Matrix ordinary differential equation |
| Lyapunov equation | Matrix ordinary differential equation | Does not hold |
| The Itô Vs Stratonovich [11] | Invariant | Invariant |
| Vec function [13] | \[dvec(P_{i}) = ((I \otimes A_{i} + A_{i} \otimes I)vec(P_{i}) + vec((GG^{T})(\tau)))d\tau,\] | \[dvec(P_{i}) = ((I \otimes A_{i} + A_{i} \otimes I) \times vec(P_{i}))\] |
Theorem 3. Suppose the filtering model of a vector time-varying bilinear Stratonovich stochastic differential system is the following:

\[
\begin{align*}
    dx_i &= (A_i(t) + A_{ix_i}) dt + (G(t) + x_i B_i^T) \cdot dW_t, \\
    dz_i &= C x_i dt + d\eta_i,
\end{align*}
\]

where the solution \( x_i \in U \), the phase space \( U \subset R^n \). Note that the vector Brownian motion process \( W = \{W_t(t), W_{x(t)}, \ldots, W_{x_j(t)}(t) \} \in R^d, \forall 0 \leq t < \infty \). The former part of the filtering model denotes the vector Brownian motion-driven vector time-varying Stratonovich bilinear stochastic differential equation. The latter denotes the noisy observation equation.

In the component-wise description, the coupled bilinear filtering equations for the filtering model are the following:

\[
\begin{align*}
    d\hat{x}_i(t) &= (A_i(t) + \sum_a A_{ia}(t) \hat{x}_a(t)) + \frac{1}{2} \sum_{\phi} (G_{i\phi}(t)B_{\phi}(t) + B_{i\phi}(t)\hat{x}_i(t)))dt \\
    &\quad + \sum_{a,\beta} (\sum_p P_{ip}(t)C_{ap}(t)) (\theta^{-1}_{a})_{ap} (dz_\beta - \sum C_{\beta p}(t) \hat{x}_p(t))dt, \\
    dP_{ij} &= \sum_p P_{ip}(t) A_{pj}(t) + \sum_p P_{jp}(t) A_{pi}(t) + \sum_{\phi} G_{i\phi}(t)G_{j\phi}(t) + \hat{x}_i \sum_{\phi} G_{i\phi}(t)B_{\phi}(t) \\
    &\quad + \sum_{\phi} G_{i\phi}(t)B_{\phi}(t)\hat{x}_j(t) + \hat{x}_i \hat{x}_j \sum P_{\phi} B_{\phi}(t) + P_{ij} \sum_{\phi} B_{\phi}(t) \\
    &\quad - \sum_{a,\beta} (\sum_p P_{ip}(t)C_{ap}(t)) (\theta^{-1}_{a})_{ap} (\sum P_{jp}(t)C_{\beta p}(t))dt.
\end{align*}
\]

The Kalman filtering becomes a special case of the coupled bilinear filtering equations, i.e. (18a)-(18b),

\[
\begin{align*}
    d\hat{x}_i(t) &= (A_i(t) + \sum_a A_{ia}(t) \hat{x}_a(t))dt + \sum_{a,\beta} (\sum_p P_{ip}(t)C_{ap}(t)) (\theta^{-1}_{a})_{ap} (dz_\beta - \sum C_{\beta p}(t) \hat{x}_p(t))dt, \\
    dP_{ij} &= \sum_p P_{ip}(t) A_{pj}(t) + \sum_p P_{jp}(t) A_{pi}(t) + \sum_{\phi} G_{i\phi}(t)G_{j\phi}(t) - \sum_{a,\beta} (\sum_p P_{ip}(t)C_{ap}(t)) \\
    &\quad - \sum_{a,\beta} (\sum_p P_{ip}(t)C_{ap}(t)) (\theta^{-1}_{a})_{ap} (\sum P_{jp}(t)C_{\beta p}(t))dt.
\end{align*}
\]

Proof. The proof of the coupled bilinear filtering equations involves the following steps: First, construct the stochastic evolution of conditional characteristic function for the bilinear Stratonovich stochastic filtering model of Theorem 3 of the paper. Then, consider \( \exp(s^T x_i) = x_i(t) \) and \( \exp(s^T x_i) = x_i(t), x_j(t) \) to compute the component-wise bilinear stochastic filtering equations, equations (18a)-(18b). Since a procedure to construct the proof of the bilinear stochastic filtering equations is similar to the proof of the Theorem 2, we omit the detail. The following is worth to mention: (i) after ignoring the term,

\[
\frac{1}{2} \sum_{\phi} (G_{i\phi}(t)B_{\phi}(t) + B_{i\phi}(t)\hat{x}_i(t)))dt,
\]
from the conditional mean equation (18a), we get the conditional mean equation of the Kalman filter. Secondly, after ignoring five correction terms from the conditional variance equation of the bilinear filter, equation (18b),

\[ \hat{x}_t = \sum G_{j\delta}(t)B_{j\delta}(t), \sum G_{j\delta}(t)B_{j\delta}(t)\hat{x}_j(t), \]

\[ \sum G_{j\delta}(t)B_{j\delta}(t)\hat{x}_j(t), \hat{x}_t\sum B_{j\delta}(t), P_{\delta}\sum B_{j\delta}(t), \]

we arrive at the conditional variance evolution equation of the celebrated Kalman filter.

**Remark 3.** Theorem 3 is about the filtering equations for the bilinear Stratonovich stochastic differential equation. An alternative appealing interpretation is the bilinear Itô stochastic differential equation. The structure of the bilinear filtering equations hinges on stochastic interpretations. On the other hand, the Kalman filtering equations do not hinge on Itô and Stratonovich stochastic interpretations. The Itô-Stratonovich dilemma does not influence the Kalman filtering since the Itô and Stratonovich stochastic integrals coincide for linear stochastic differential equations.

**Conclusion**

In this paper, we have developed three Kalman filtering Theorems and their formal proof for the continuous state-discrete measurement system and continuous state-continuous measurement system. The paper exploits conditional characteristic function evolution equation and the Kiyoshi Itô calculus to achieve that. Most notably, this paper chiefly reveals ‘how Kalman filtering equations become a special case of the stochastic evolution of conditional characteristic function for linear stochastic differential systems and linear measurement systems.

For the first time, the celebrated Kalman filtering was introduced in 1960. A philosophical contribution of the paper is to reveal connections between three influential results of stochastic processes, i.e. Kalman filtering, conditional characteristic function and the Itô calculus. This paper unifies them by adopting formal and systematic frameworks that were not available previously in the literature.

This paper differentiates between the Kalman filtering for the continuous state-continuous measurement system as well as the continuous state-discrete measurement system in table (1). Table (1) displays five major ingredients of the Kalman filter, conditional characteristic function evolution, the Riccati equation, the Lyapunov matrix equation, the vec function, the Itô-Stratonovich dilemma.

**Appendix**

Calculation of the term \( \exp(s^T x_t)\sum_{\gamma} C_{\gamma\gamma}(t)x_{\gamma}(t) - \left( \exp(s^T x_t)\right) \left( \sum_{\gamma} C_{\gamma\gamma}(t)x_{\gamma}(t) \right) \).

Here, we prove that

\[ \exp(s^T x_t)\sum_{\gamma} C_{\gamma\gamma}(t)x_{\gamma}(t) - \left( \exp(s^T x_t)\right) \left( \sum_{\gamma} C_{\gamma\gamma}(t)x_{\gamma}(t) \right) = \sum_{\gamma} C_{\gamma\gamma}(t)\left( P_{i\gamma} \langle x_j \rangle + P_{j\gamma} \langle x_i \rangle \right), \]

where the terms of the above relation are associated with the filtering model of the Theorem 2 of the paper.
The term \( \langle \exp(s^T x_t) \sum_{\gamma} C_{\alpha\gamma}(t) x_{\gamma}(t) \rangle - \langle \exp(s^T x_t) \rangle \langle \sum_{\gamma} C_{\alpha\gamma}(t) x_{\gamma}(t) \rangle \) arises in the stochastic evolution of conditional characteristic function for the linear stochastic differential system coupled with the continuous linear measurement system. We wish to compute the term, where \( \exp(s^T x_t) = x_j x_j \). Then, the resulting expression is embedded in the conditional variance evolution of the standard Kalman filter. Thus, we have

\[
\begin{align*}
\langle \exp(s^T x_t) \sum_{\gamma} C_{\alpha\gamma}(t) x_{\gamma}(t) \rangle - \langle \exp(s^T x_t) \rangle \langle \sum_{\gamma} C_{\alpha\gamma}(t) x_{\gamma}(t) \rangle \\
= \langle x_j x_j \sum_{\gamma} C_{\alpha\gamma}(t) x_{\gamma}(t) \rangle - \langle x_j x_j \rangle \langle \sum_{\gamma} C_{\alpha\gamma}(t) x_{\gamma}(t) \rangle.
\end{align*}
\]

Since the conditional expectation operator is linear, the right-hand side of the above becomes

\[
\begin{align*}
\langle \exp(s^T x_t) \sum_{\gamma} C_{\alpha\gamma}(t) x_{\gamma}(t) \rangle - \langle \exp(s^T x_t) \rangle \langle \sum_{\gamma} C_{\alpha\gamma}(t) x_{\gamma}(t) \rangle \\
= \sum_{\gamma} C_{\alpha\gamma}(t) \langle x_j x_j x_{\gamma} \rangle - \sum_{\gamma} C_{\alpha\gamma}(t) \langle x_j x_{\gamma} \rangle \langle x_j \rangle.
\end{align*}
\]

After considering the Gaussian assumption, even powers will contribute and odd power contributions will vanish. Thus, the right-hand side of equation (A.1) becomes

\[
\sum_{\gamma} C_{\alpha\gamma}(t) \langle x_j x_j x_{\gamma} \rangle - \sum_{\gamma} C_{\alpha\gamma}(t) \langle x_j x_{\gamma} \rangle \langle x_j \rangle
\]

\[
= \sum_{\gamma} C_{\alpha\gamma}(t) \left( \frac{1}{2} \sum_{p,q} P_{pq} \frac{\partial^2}{\partial x_p \partial x_q} \langle x_j x_j x_{\gamma} \rangle - \frac{1}{2} \sum_{p,q} P_{pq} \frac{\partial^2}{\partial x_p \partial x_q} \langle x_j \rangle \langle x_j \rangle \right)
\]

\[
= \frac{1}{2} \sum_{\gamma} C_{\alpha\gamma}(t) \left( \sum_{p,q} P_{pq} \frac{\partial^2}{\partial x_p \partial x_q} \langle x_j x_j x_{\gamma} \rangle - \sum_{p,q} P_{pq} \frac{\partial^2}{\partial x_p \partial x_q} \langle x_j \rangle \langle x_j \rangle \right)
\]

\[
= \frac{1}{2} \sum_{\gamma} C_{\alpha\gamma}(t) \left( \sum_{p,q} P_{pq} \left( \frac{\partial}{\partial x_p} \langle x_j x_j \rangle + \delta_{pq} \langle x_j \rangle \langle x_j \rangle + \delta_{qj} \langle x_j \rangle \langle x_j \rangle \right)
\]

\[
- \delta_{pq} \delta_{jq} \langle x_j \rangle - \delta_{jp} \delta_{iq} \langle x_j \rangle \right) \langle x_j \rangle.
\]

The right-hand side of the above expression is further simplified to

\[
\frac{1}{2} \sum_{\gamma} C_{\alpha\gamma}(t) \left( \sum_{p,q} P_{pq} \left( \delta_{pq} \delta_{jq} \langle x_j \rangle + \delta_{pq} \delta_{jp} \langle x_j \rangle + \delta_{jq} \delta_{jp} \langle x_j \rangle + \delta_{qj} \delta_{jp} \langle x_j \rangle + \delta_{pq} \delta_{jq} \langle x_j \rangle + \delta_{jp} \delta_{iq} \langle x_j \rangle \right)
\]

\[
- \delta_{pq} \delta_{jq} \langle x_j \rangle - \delta_{jp} \delta_{iq} \langle x_j \rangle \right) \langle x_j \rangle
\]
\[
= \frac{1}{2} \sum_{y} C_{ay}(t) \sum_{p,q} P_{pq}(\delta_{iq} \delta_{jp} \langle x_{y} \rangle + \delta_{jq} \delta_{ip} \langle x_{j} \rangle + \delta_{jq} \delta_{ip} \langle x_{j} \rangle + \delta_{iq} \delta_{jp} \langle x_{i} \rangle - \\
- \delta_{iq} \delta_{jp} \langle x_{i} \rangle - \delta_{jq} \delta_{ip} \langle x_{j} \rangle) \\
= \sum_{y} C_{ay}(t) \sum_{p,q} P_{pq}(\delta_{iq} \delta_{jp} \langle x_{j} \rangle + \delta_{jq} \delta_{ip} \langle x_{j} \rangle) \\
= \sum_{y} C_{ay}(t)(P_{iy} \langle x_{j} \rangle + P_{iy} \langle x_{i} \rangle).
\]

Thus,
\[
\left\langle \exp(s^{T} x_{i}) \sum_{y} C_{ay}(t)x_{y}(t) \right\rangle - \left\langle \exp(s^{T} x_{i}) \right\rangle \left\langle \sum_{y} C_{ay}(t)x_{y}(t) \right\rangle \\
= \left\langle x_{i}x_{j} \sum_{y} C_{ay}(t)x_{y}(t) \right\rangle - \left\langle x_{i}x_{j} \right\rangle \left\langle \sum_{y} C_{ay}(t)x_{y}(t) \right\rangle \\
= \sum_{y} C_{ay}(t)(P_{iy} \langle x_{j} \rangle + P_{iy} \langle x_{i} \rangle). \quad (A.2)
\]

**QED**

**REFERENCES**

1. Sharma, S. N., and Gawalwad, B. G., 2016, “Wiener meets Kolmogorov,” *Norbert Wiener in the 21st Century* (Thinking Machines in the Physical World), Jul 13-Jul 15, University of Melbourne, Australia.
2. Kalman, R. E., 1960, “A New Approach to Linear Filtering and Prediction Problems,” The ASME Transactions, Journal of Basic Engineering, 82, pp. 34-45.
3. Anderson, B. D. O., and Moore, J. B., 1991, “Kalman Filtering: Whence, What and Whither?,” Mathematical System Theory, Ac Antoulas, ed., Springer, Berlin Heidelberg, pp. 41-54.
4. Gattami, A., 2014, “Kalman meets Shannon,” *19th IFAC World Congress*, IFAC Proceedings, 47(3), pp. 2376–2381.
5. Liptser, R. S., and Shiryaev, A. N., 1977, *Statistics of Random Processes I*, Springer, Berlin.
6. Fujisaki, M., Kallianpur, G., and Kunita, H., 1972, “Stochastic Differential Equations for the Nonlinear Filtering Problem,” Osaka Journal of Mathematics, 9(1), pp. 19-40.
7. Holbrook, J., Bhatia, R., 2006, “An Old Question Asked in a New Context Presents Strange Aspects,” Mathematical Intelligencer, 28 (1), pp. 32-39.
8. Anderson, B. D. O., and Moore, J. B., 1991, “Kalman Filtering: Whence, What and Whither? Mathematical System Theory “, Ac Antoulas, ed., Springer, Berlin Heidelberg, pp. 41-54.
9. Karatzas, I., and Shreve, S. E., 1991, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, Berlin.
10. Kunita, H., 2010, “Itô’s Stochastic Calculus: Its Surprising Power for Applications, Stochastic Processes and their Applications,” 120, pp.622-652.
11. Mannella, R., and McClintock, P. V. E., 2012, “Itô versus Stratonovich: 30 Years Later,” Fluctuation and Noise Letters, 11(1), pp.1240010-1240019.
12. Patel, H. G., and Sharma, S. N., 2014, “Third-order Continuous-discrete Filtering Equations for a Non-linear Dynamical System: The ASME Transactions, Journal of Computational and Non-linear Dynamics, 9(3), pp. 034502-9.
13. Brewer, J.W., 1978, “Kronecker Products and Matrix Calculus in System Theory,” IEEE Transactions on Circuits and Systems, CAS-25 (9), pp. 772-781.