The Cauchy-Schwarz divergence for Poisson point processes

Hung Gia Hoang, Ba-Ngu Vo, Ba-Tuong Vo, and Ronald Mahler

Abstract—Information theoretic divergences are fundamental tools used to measure the difference between the information conveyed by two random processes. In this paper, we show that the Cauchy-Schwarz divergence between two Poisson point processes is the half the squared $L^2$-distance between their respective intensity functions. Moreover, this can be evaluated in closed form when the intensities are Gaussian mixtures. We illustrate the result via a sensor management application involving the Probability Hypothesis Density (PHD) filter.

Index Terms—Poisson point process, information theoretic divergence, random finite sets

I. INTRODUCTION

The Poisson point process, which models “no interaction” or “complete spatial randomness” in spatial point patterns, is arguably one of the best known and most tractable of point processes [1], [2], [3], [4], [5]. Point process theory is the study of random counting measures with applications spanning numerous disciplines, see for example [2], [1], [5], [6], [7]. The Poisson point process itself arises in forestry [8], geology [9], biology [10], particle physics [11], communication networks [12], [13], [14] and signal processing [15], [16], [17].

The role of the Poisson point process in point process theory, in most respects, is analogous to that of the normal distribution in random vectors [18]. Information theoretic divergences measure the “difference” in information between two random variables, and are fundamental in information theory and statistical analysis [19].

The better known divergences include Kullback-Leibler, Rényi (or $\alpha$-divergence) and their generalization Csiszár-Morimoto (or Ali-Silvey), as well as Jensen-Rényi and Cauchy-Schwarz.

The Kullback-Leibler divergence can be evaluated analytically for Gaussians (random vectors), but for the more versatile class of Gaussian mixture [20], [21], only Jensen-Rényi and Cauchy-Schwarz divergences can be evaluated in closed forms [22], [23]. The Kullback-Leibler and Rényi divergences have also been studied for point processes or random finite sets, [24]–[26], [27]–[29]. However, so far except for trivial special cases, these divergences cannot be computed analytically and requires expensive approximations such as Monte Carlo.

In this paper, we show that for Poisson point processes, the Cauchy-Schwarz divergence between their probability densities is given by the square of the $L^2$-distance between their intensity functions. Geometrically, this result relates the angle subtended by the probability densities to the $L^2$-distance between the corresponding intensity functions. Moreover, for Gaussian mixture intensity functions, the $L^2$-distance, and hence the Cauchy-Schwarz divergence, can be evaluated in closed form. The Poisson point process enjoys numerous nice properties [1], [2], [3], and our result is an interesting and useful addition. We demonstrate our result by using the closed form Cauchy-Schwarz divergence in a sensor management application for multi-target tracking involving the Probability Hypothesis Density (PHD) filter.

The organization of the paper is as follows. Section II presents the main results of the paper that establish the analytical formulation for the Cauchy-Schwarz divergence between two Poisson point processes. In Section III the application of the Cauchy-Schwarz divergence to sensor management, including numerical examples, is studied. Finally, Section IV concludes the paper.

II. CAUCHY-SCHWARZ DIVERGENCE FOR POISSON POINT PROCESSES

In this work we consider a state space $\mathcal{X} \subseteq \mathbb{R}^d$, and adopt the inner product notation $\langle f, g \rangle \triangleq \int f(x)g(x)dx$; the $L^2$-norm notation $\|f\| \triangleq \sqrt{\langle f, f \rangle}$; the multi-target exponential notation $h^A \triangleq \prod_{x \in \mathcal{X}} h(x)$, where $h$ is a real-valued function, with $h^B \triangleq 1$ by convention; and the indicator function notation

$$
1_B(x) \triangleq \begin{cases} 1, & \text{if } x \in B \\ 0, & \text{otherwise} \end{cases}.
$$

A. Point processes

This section briefly summarizes concepts in point process theory needed for the exposition of our result. Point process theory, in general, is concerned with random counting measures. Our result is restricted to simple-finite point processes, which can be regarded as random finite sets. For simplicity, we omit the prefix “simple-finite” in the rest of the paper. For an introduction to the subject we refer the reader to the article [6], and for detailed treatments, textbooks such as [1], [2], [4], [5].

A point process or random finite set (RFS) $X$ on $\mathcal{X}$ is random variable taking values in $\mathcal{P}(\mathcal{X})$, the space of finite subsets of $\mathcal{X}$. Let $|X|$ denotes the number of elements in a
set \( X \). A point process \( X \) on \( \mathcal{X} \) is said to be Poisson with a given intensity function \( u \) (defined on \( \mathcal{X} \)) if \([1, 2]\):

1) for any \( B \subseteq \mathcal{X} \) such that \( \langle u, 1_B \rangle < \infty \), the random variable \([X \cap B] \) is Poisson distributed with mean \( \langle u, 1_B \rangle \).

2) for any disjoint \( B_1, ..., B_l \subseteq \mathcal{X} \), the random variables \([X \cap B_1], ..., [X \cap B_l] \) are independent.

Since \( \langle u, 1_B \rangle \) is the expected number of points of \( X \) in the region \( B \), the intensity value \( u(x) \) can be interpreted as the instantaneous expected number of points per unit hyper-volume at \( x \). Consequently, \( u(x) \) is not dimensionless in general. If hyper-volume (on \( \mathcal{X} \)) is measured in units of \( K \) (e.g. \( m^d \), \( cm^d \), \( in^d \), etc.) then the intensity function \( u \) has unit \( K^{-1} \).

The number of points of a Poisson point process \( X \) is Poisson distributed with mean \( \langle u, 1 \rangle \), and condition on the number of points the elements \( x \) of \( X \) are independently and identically distributed (i.i.d.) according to the probability density \( u(\cdot)/\langle u, 1 \rangle \) \([1, 2, 3, 4, 5]\). It is implicit that \( \langle u, 1 \rangle \) is finite since we only consider simple-finite point processes.

The probability distribution of a Poisson point process \( X \) with intensity function \( u \) is given by \([5]\) pp. 15

\[
\Pr(X \in \mathcal{T}) = \sum_{i=0}^{\infty} \frac{e^{-\langle u, 1 \rangle}}{i!} \int_{\mathcal{T}^i} \mathcal{P}^{x_1, ..., x_i} d(x_1, ..., x_i), \tag{1}
\]

for any (measurable) subset \( \mathcal{T} \) of \( \mathcal{P}(\mathcal{X}) \), where \( \mathcal{X}^i \) denotes an \( i \)-fold Cartesian product of \( \mathcal{X} \), with the convention \( \mathcal{X}^0 = \{\emptyset\} \), and the integral over \( \mathcal{X}^0 \) is \( \mathcal{P}(\emptyset) = 1 \). A Poisson point process is completely characterized by its intensity function (or more generally the intensity measure).

Probability densities of point processes considered in this work are defined with respect to the reference measure \( \mu \) given by

\[
\mu(\mathcal{T}) = \sum_{i=0}^{\infty} \frac{1}{i! K^i} \int_{\mathcal{T}^i} \mathcal{P}^{x_1, ..., x_i} d(x_1, ..., x_i), \tag{2}
\]

for any (measurable) subset \( \mathcal{T} \) of \( \mathcal{P}(\mathcal{X}) \). The measure \( \mu \) is analogous to the Lebesgue measure on \( \mathcal{X} \) (indeed it is the normalized distribution of a Poisson point process with unit intensity \( u = 1/K \) when the state space \( \mathcal{X} \) is bounded). Moreover, it was shown in \([30]\) that for this choice of reference measure, the integral of a function \( f : \mathcal{P}(\mathcal{X}) \rightarrow \mathbb{R} \), given by

\[
\int f(X) \mu(dX) = \sum_{i=0}^{\infty} \frac{1}{i! K^i} \int_{\mathcal{X}^i} f(x_1, ..., x_i) d(x_1, ..., x_i), \tag{3}
\]

is equivalent to Mahler’s set integral \([15]\). Note that the reference measure \( \mu \), and the integrand \( f \) are all dimensionless.

Our main result involves Poisson point processes with probability densities of the form

\[
f(X) = K^{X|} e^{-\langle u, 1 \rangle} u^X. \tag{4}
\]

Note that for any (measurable) subset \( \mathcal{T} \) of \( \mathcal{P}(\mathcal{X}) \)

\[
\int_{\mathcal{T}} f(X) \mu(dX) = \int_{\mathcal{T}} 1_{\mathcal{T}}(X)f(X) \mu(dX) = \sum_{i=0}^{\infty} \frac{e^{-\langle u, 1 \rangle}}{i!} \int_{\mathcal{T}^i} 1_{\mathcal{T}^i}(x_1, ..., x_i) u^{(x_1, ..., x_i)} d(x_1, ..., x_i).
\]

Thus, comparing with \([1, 2]\), \( f \) is indeed a probability density (with respect to \( \mu \)) of a Poisson point process with intensity function \( u \).

### B. Cauchy-Schwarz divergence

The Cauchy-Schwarz divergence is based on the Cauchy-Schwarz inequality for inner products, and is defined for two random vectors with probability densities \( f \) and \( g \) by \([23]\)

\[
D_{CS}(f, g) = -\ln \frac{\langle f, g \rangle}{\|f\| \|g\|}, \tag{5}
\]

The argument of the logarithm in \([5]\) is non-negative (since probability densities are non-negative) and does not exceed one (by the Cauchy-Schwarz inequality). Moreover, it can be interpreted as the cosine of the angle subtended by \( f \) and \( g \) in \( L^2(\mathcal{X}, \mathbb{R}) \), the space of square integrable functions taking \( \mathcal{X} \) to \( \mathbb{R} \). Note that \( D_{CS}(f, g) \) is symmetric and positive unless \( f = g \), in which case \( D_{CS}(f, g) = 0 \).

Geometrically, the Cauchy-Schwarz divergence determines the information “difference” between random vectors from the angle between their probability densities. The Cauchy-Schwarz divergence can also be interpreted as an approximation to the Kullback-Leibler divergence \([23]\). Hence, the Cauchy-Schwarz divergence between two densities of random variables has been employed in many information theoretic applications, especially in machine learning and pattern recognition \([23, 31, 34]\).

For point processes, the Csiszár-Morimoto divergence, which includes Kullback-Leibler and Rényi, were formulated in \([25]\) by replacing standard (Lebesque) integrals with set integrals \([15]\). However, the Cauchy-Schwarz divergence cannot be extended to point processes by simply replacing the standard integral with the set integral since the latter is not unit compatible for defining inner products and norms of set derivatives.

Using the standard notion of density and integration summarized in Section \([12, 14]\) we can define the inner product \( \langle f, g \rangle_{\mu} \triangleq \int f(X) g(X) \mu(dX) \), and corresponding norm \( \|f\|_{\mu} \triangleq \sqrt{\langle f, f \rangle_{\mu}} \) on \( L^2(\mathcal{P}(\mathcal{X}), \mathbb{R}) \), the space of square integrable functions taking \( \mathcal{P}(\mathcal{X}) \) to \( \mathbb{R} \). Interestingly, the inner product between multi-target exponentials is given by the following result.

**Lemma 1.** Let \( f(X) = K^{X|} u^X \) and \( g(X) = K^{X|} v^X \) with \( u, v \in L^2(\mathcal{X}, \mathbb{R}) \) (measured in units of \( K^{-1} \)). Then \( \langle f, g \rangle_{\mu} = e^{K\langle u, v \rangle} \).
Proof:

\[
(f,g)_\mu = \int K^{2\mathcal{N}}[v](u)\mu(dx)
= \sum_{i=0}^{\infty} \frac{K_i}{i!} \left[ \int \mathcal{X}(x)\nu(x)dx \right]^i \quad \text{(using \(\mathcal{F}\))}
= \sum_{i=0}^{\infty} \frac{K_i}{i!} (u,v)^i = e^{\mathcal{F}(u,v)} \quad \Box
\]

In the spirit of using the angle between probability densities to determine the information “difference”, the Cauchy-Schwarz divergence can be extended to point processes as follows.

Definition 2. The Cauchy-Schwarz divergence between the probability densities \(f\) and \(g\) of two point processes with respect to the reference measure \(\mu\) is defined by

\[
D_{CS}(f,g) = -\ln \frac{(f,g)_\mu}{\|f\|_\mu \|g\|_\mu}. \quad (6)
\]

The following Proposition asserts that the Cauchy-Schwarz divergence between two Poisson point process is half the squared distance between their intensity functions.

Proposition 3. The Cauchy-Schwarz divergence between the probability densities \(f\) and \(g\) of two Poisson point processes with respective intensity functions \(u\) and \(v\) in \(L^2(\mathcal{X},\mathbb{R})\) (measured in units of \(K^{-1}\)), is

\[
D_{CS}(f,g) = \frac{K}{2} \|u - v\|^2. \quad (7)
\]

Proof: Applying Lemma[I] to the numerator and denominator of (6) and cancelling out the constants \(e^{-\langle u,1 \rangle}, e^{-\langle v,1 \rangle}\) we have

\[
D_{CS}(f,g) = -\ln \left( e^{\mathcal{F}(u,v)-\frac{K}{2}(u,v)} \right)
= \frac{1}{2} [K \langle u,u \rangle - 2K \langle u,v \rangle + K \langle v,v \rangle]
= K \frac{1}{2} \|u - v\|^2 \quad \Box
\]

Note that since the intensity functions have units of \(K^{-1}\), \(\|u - v\|^2\) also has units of \(K^{-1}\) and hence \(K \|u - v\|^2\) is unitless. Moreover, \(K \|u - v\|^2\), referred to as the squared distance between the intensity functions \(u\) and \(v\), takes on the same value regardless of the choice of measurement unit. Indeed, suppose that the unit of the hyper-volume in the intensity function space has been change from \(K\) to \(\rho K\) (for example, from \(dm^3\) to \(m^3 = 10^3 dm^3\)) as illustrated in Fig. 1. Note that the change of unit inevitably leads to the change in values of the two intensity functions (for example, the intensity measured in \(m^{-3}\), which is the expected number of points per cubic meter, is one thousand times the intensity measured in \(dm^{-3}\)). However, these changes cancel each other in the product \(\frac{K}{2} \|u - v\|^2\) such that the squared distance remains unchanged.

The above result has a nice geometric interpretation that relates the angle subtended by the probability densities in \(L^2(\mathcal{X},\mathbb{R})\) to the distance between the corresponding intensity functions in \(L^2(\mathcal{X},\mathbb{R})\) as depicted in Fig. 2. More specifically, for Poisson point processes: the secant of the angle between their probability densities equals the exponential of half the squared distance between their intensity functions.

Proposition 3 has important implications in the approximation of Poisson point processes through their intensity functions. It is intuitive that the “difference” between the Poisson distributions vanishes as the distance between their intensity functions tends to zero. However, it was not clear that a reduction in the error between the intensity functions necessarily implies a reduction in the “difference” between the associated Poisson distributions. Our result not only verifies that the “difference” between the distributions is reduced, it also quantifies the reduction.

In general, the \(L^2\)-norm and hence the Cauchy-Schwarz divergence above cannot be numerically evaluated in closed form. However, for Poisson point processes with Gaussian mixture intensity function, applying the Gaussian identity \(\mathcal{N}(\mu_0, \Sigma_0) = \mathcal{N}(\mu_1, \Sigma_1)\) to (7) yields an analytic expression for the Cauchy-Schwarz divergence. This is stated more concisely in the following result.
Corollary 4. The Cauchy-Schwarz divergence between two Poisson point processes with Gaussian mixture intensities:

\[ u(x) = \sum_{i=0}^{N_k} w_u^{(i)} \mathcal{N}(x; m_u^{(i)}, p_u^{(i)}), \quad (8a) \]

\[ v(x) = \sum_{i=0}^{N_k} w_v^{(i)} \mathcal{N}(x; m_v^{(i)}, p_v^{(i)}), \quad (8b) \]

(measured in units of \( K^{-1} \)) is given by

\[ D_{CS} = \frac{1}{2} \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} w_u^{(i)} w_v^{(j)} \mathcal{N}(m_u^{(i)}; m_v^{(j)}, p_u^{(i)} + p_v^{(j)}) \]

\[ + \frac{1}{2} \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} w_v^{(i)} w_v^{(j)} \mathcal{N}(m_v^{(i)}; m_v^{(j)}, p_v^{(i)} + p_v^{(j)}) \]

\[ - \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} w_u^{(i)} w_v^{(j)} \mathcal{N}(m_u^{(i)}; m_v^{(j)}, p_u^{(i)} + p_v^{(j)}) \]

This Corollary has important implications in Gaussian mixture reduction for intensity functions. The result provides mathematical justification for Gaussian mixture intensity reduction based on \( L^2 \)-error. Furthermore, since Gaussian mixtures can approximate any density to any desired accuracy [36], Corollary 4 enables the Cauchy-Schwarz divergence between two Poisson point processes to be approximated to any desired accuracy.

III. MULTI-TARGET SENSOR CONTROL

In this section, we present an application of our result to a sensor management problem for multi-target systems. Multi-target systems are an established area of research in which the system states can be naturally modeled as point processes [15], [30], [37], [38].

A multi-target system is fundamentally different from a single-target system in that the number of states changes with time due to births and deaths of targets. The hidden state of a multi-target system at time \( k \) is a finite set \( X_k \) which is partially observed as another finite set \( Z_k \) by a sensor with control signal \( a_{k-1} \) applied at time \( k-1 \). Multi-target sensor management is a stochastic control problem which involves:

1) Propagating the multi-target posterior density, or alternatively a tractable approximation, recursively in time;
2) At each time, determining the control/action by optimizing the reward function over a set of admissible actions.

Denote by \( \mathcal{R}(a, Z_k) \) the reward function if we were to apply the control vector \( a \) and subsequently were to observe the measurement sequence \( Z_{k+1} \). Notice that the reward function depends on \( p+1 \) future measurements which could be taken to \( \infty \). For the purpose of illustrating the result in the previous section, we focus on a single step ahead (myopic) policy. The best control vector \( a_{k-1} \) is chosen to maximize the expected reward \( E[\mathcal{R}(a, Z_k)] \), where the expectation is taken over all possible values of the future measurement \( Z_k \) [28], [39], [40]. The expected reward is generally intractable and a number of approximation techniques have been proposed in [28], [29], [39], [40]. In this paper, we follow the computationally cheaper strategy that select the control \( a \) to optimize \( \mathcal{R}(a, Z^*) \), where \( Z^* \) is the ideal predicted measurement [28], [39].

The form of the reward function \( \mathcal{R}(a, Z_k) \) plays a crucial role in sensor control problems as it determines the manoeuvre of the sensor. In [28], [29], [40], the reward function \( \mathcal{R}(a, Z_{k+1}) \) is the Rényi divergence between the approximate prediction and posterior probability densities. In this paper, we propose the Cauchy-Schwarz divergence between the prediction and posterior probability densities as the reward function. The implementation of the Cauchy-Schwarz divergence in multi-target sensor control is described in the following subsections.

A. System model

We consider the following linear Gaussian multi-target model for the illustration of our result. Each constituent element \( x_{k-1} \) of a multi-target state \( X_k \) at time \( k-1 \), either continues to exist at time \( k \) with probability \( p_{S,k} \) (or dies with probability \( 1 - p_{S,k} \)) and conditional on the existence at time \( k \), the transition from \( x_{k-1} \) to \( x_k \) is governed by transition density

\[ f(x_k|x_{k-1}) = \mathcal{N}(x_k; F_{k-1}x_{k-1}, Q_{k-1}). \]

New targets can arise at time \( k \) either by spontaneous births, or by spawning from targets at time \( k-1 \). The set of birth targets and spawned targets are modelled as RFSs with respective Gaussian mixture intensity functions

\[ \gamma_k(x) = \sum_{i=1}^{J} w^{(i)}_{\gamma,k} \mathcal{N}(x; m_{\gamma,k}^{(i)}, P_{\gamma,k}^{(i)}), \]

\[ \beta_{k-1}(x|\xi) = \sum_{i=1}^{J} w^{(i)}_{\beta,k-1} \mathcal{N}(x; m_{\beta,k-1}^{(i)} \xi + d_{\beta,k-1}^{(i)} \xi, Q_{\beta,k-1}^{(i)} \xi), \]

The multi-target state is hidden and is partially observed by a sensor. Each target evolves and generates observations independently of one another. A target with state \( x_k \) is detected by the sensor given the control vector \( a_{k-1} \) at time \( k-1 \), with probability \( p_{D,k}(x_k; a_{k-1}) \) (or missed with probability \( 1 - p_{D,k}(x_k; a_{k-1}) \)) and conditional on detection generates a measurement \( z_k \) according to the probability density

\[ g_k(z_k|x_k) = \mathcal{N}(z_k; H_k x_k, R_k). \]

The sensor also registers a set of spurious measurements (clutter), independent of the detections, modeled as a Poisson point process with intensity \( \kappa_k \). Thus, at each time step the measurement is collection of detections \( Z_k \), only some of which are generated by targets.

B. Posterior intensity propagation

Since computing the full posterior of the multi-target state \( X_k \) is generally intractable, an approximation known as the Probability Hypothesis Density (PHD) filter has been proposed to recursively compute the posterior intensity function as a first order approximation [13]. For the linear Gaussian multi-target model, the Gaussian Mixture (GM-PHD) filter propagates the Gaussian mixture intensity functions forward in time [41].
GM-PHD filter consists of the following prediction and update steps: 

**Prediction:** If the posterior intensity at time $k - 1$ is a Gaussian mixture of the form 
\[
v_{k-1}(x) = \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} N \left( x; m_{k-1}^{(i)}, P_{k-1}^{(i)} \right)
\]
then the predicted intensity at time $k$ is also a Gaussian mixture and is given by 
\[
v_{k|k-1}(x) = v_{S,k|k-1}(x) + v_{\beta,k|k-1}(x) + \gamma_k(x)
\]
where 
\[
v_{S,k|k-1}(x) = \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} N \left( x; m_{S,k|k-1}^{(i)}, P_{S,k|k-1}^{(i)} \right)
\]
\[
v_{\beta,k|k-1}(x) = \sum_{j=1}^{J_{k-1}} \sum_{i=1}^{J_{k-1}} w_{k-1}^{(i)} w_{k}^{(j)} N \left( x; m_{\beta,k|k-1}^{(ij)}, P_{\beta,k|k-1}^{(ij)} \right)
\]
\[
m_{S,k|k-1}^{(i)} = F_k - m_{k|k-1}^{(i)}
\]
\[
p_{\beta,k|k-1}^{(ij)} = Q_k - F_k - P_{\beta,k|k-1}^{(ij)} F_k + P_{\beta,k|k-1}^{(ij)}
\]

**Update:** If predicted intensity and detection probability are Gaussian mixtures of the form 
\[
v_{k|k-1}(x) = \sum_{i=1}^{J_{k|1}} w_{k|k-1}^{(i)} N \left( x; m_{k|k-1}^{(i)}, P_{k|k-1}^{(i)} \right),
\]
\[
p_{D,k}(x;a_{k-1}) = \sum_{j=0}^{J_{D,k}} w_{D,k}^{(j)} N \left( x; m_{D,k}^{(j)}, P_{D,k}^{(j)} \right)
\]
then, the posterior intensity at time $k$ is given by 
\[
v_k(x;a_{k-1}, Z_k) = v_{k|k-1}(x) - v_{D,k}(x;a_{k-1}) + \sum_{\chi \in \mathcal{A}} v_{D,k}(x;a_{k-1}, \chi)
\]

where Gaussian mixtures $v_{D,k}(x;a_{k-1})$ and $v_{D,k}(x;a_{k-1}, \chi)$ are computed respectively as follows 
\[
v_{D,k}(x;a_{k-1}) = \sum_{i=1}^{J_{k-1}} \sum_{j=0}^{J_{D,k}} w_{D,k}^{(i,j)} (a_{k-1}) \cdot N \left( x; m_{D,k}^{(i,j)}, P_{D,k}^{(i,j)} \right),
\]
\[
w_{D,k}^{(i,j)} (a_{k-1}) = w_{D,k}^{(i)} w_{D,k}^{(j)} (a_{k-1}),
\]
\[
q_{k|k-1}^{(i,j)} (a_{k-1}) = m_{k|k-1}^{(i)} - m_{D,k}^{(i,j)},
\]
\[
m_{k|k-1}^{(i,j)} (a_{k-1}) = m_{D,k}^{(i,j)} + P_{\tilde{D},k}^{(i,j)} (a_{k-1}) - m_{D,k}^{(i,j)},
\]
\[
P_{k|k-1}^{(i,j)} = \left( I - K_{i|k-1}^{(i,j)} \right)^{-1},
\]
\[
K_{i|k-1}^{(i,j)} = \sum_{i=1}^{J_{k-1}} \sum_{j=0}^{J_{D,k}} w_{D,k}^{(i,j)} (a_{k-1})
\]

and 
\[
v_{D,k}(x;a_{k-1}, \chi) = \sum_{i=1}^{J_{k-1}} \sum_{j=0}^{J_{D,k}} w_{D,k}^{(i,j)} (a_{k-1}) \cdot N \left( x; m_{D,k}^{(i,j)}, P_{D,k}^{(i,j)} \right),
\]
\[
w_{D,k}^{(i,j)} (a_{k-1}) = w_{D,k}^{(i)} w_{D,k}^{(j)} (a_{k-1}),
\]
\[
q_{k|k-1}^{(i,j)} (a_{k-1}) = m_{k|k-1}^{(i)} - m_{D,k}^{(i,j)},
\]
\[
m_{k|k-1}^{(i,j)} (a_{k-1}) = m_{D,k}^{(i,j)} + P_{\tilde{D},k}^{(i,j)} (a_{k-1}) - m_{D,k}^{(i,j)},
\]
\[
P_{k|k-1}^{(i,j)} = \left( I - K_{i|k-1}^{(i,j)} \right)^{-1},
\]
\[
K_{i|k-1}^{(i,j)} = \sum_{i=1}^{J_{k-1}} \sum_{j=0}^{J_{D,k}} w_{D,k}^{(i,j)} (a_{k-1})
\]

Exact implementation of the Gaussian mixture update step encounters difficulties due to negative weights in the misdetection component, $v_{k|k-1}(x) - v_{D,k}(x;a_{k-1})$. This issue with negative weights arises in general for Gaussian mixture implementations, and for tractability it is desirable to approximate the misdetection term with a Gaussian mixture of non-negative weights. Simple approaches to mitigating this difficulty involve linearizing the state dependent detection probability $\beta_{k|k-1}$.

We propose an alternative but more accurate scheme in which the misdetection term is approximated by a Gaussian mixture based on the predicted intensity $v_{k|k-1}(x)$ where the non-negative predicted weights are modulated by the non-negative function $\left[ 1 - p_{D,k}(x,a_{k-1}) \right]$ and normalized in order to preserve the total mass of the exact misdetection term. Specifically, the first term in the right hand side of (37) is approximated by 
\[
v_{\text{miss},k}(x,a_{k-1}) = \sum_{i=1}^{J_{k}} w_{i}^{(i)} (a_{k-1}) \cdot N \left( x; m_{k}^{(i)}, P_{k}^{(i)} \right)
\]
\[
w_{\text{miss},k}^{(i)} (a_{k-1}) = \frac{w_{i}^{(i)} (a_{k-1}) T_{\text{miss},k}^{(i)} (a_{k-1})}{\sum_{i=1}^{J_{k}} w_{i}^{(i)} (a_{k-1})},
\]
\[
th_{D,k}^{(i)} (a_{k-1}) = \sum_{i=1}^{J_{k}} w_{i}^{(i)} (a_{k-1}) - \sum_{i=1}^{J_{k}} \sum_{j=0}^{J_{D,k}} w_{D,k}^{(i,j)} (a_{k-1})
\]

C. Cauchy-Schwarz divergence based reward function

Since the PHD filter propagates the posterior intensity of the multi-target state, one natural choice of reward function is the Rényi divergence between the Poisson point processes associated with the predicted intensity $v_{k|k-1}(x)$ and posterior intensity $v_{k|k-1}(\alpha, Z_k)$. The main drawback of the Rényi divergence based reward function is that it involves computation
of integrals in infinite dimensional spaces, which is intractable in general.

In this work, we propose to replace the Rényi divergence in the reward function of [28] by the Cauchy-Schwarz divergence. Using Proposition 4, the reward function reduces to the squared $L^2$-distance between the predicted and posterior intensity:

$$ R(a, Z_k) = \frac{K}{2} \left\| v_{k-1}(\cdot) - v_k(\cdot; a, Z_k) \right\|^2 $$

(42)

This choice of divergence reduces the evaluation of integrals in the infinite dimensional space $\mathcal{F}(\mathcal{X})$ to standard integrals on the finite dimensional space $\mathcal{F}$. Moreover, when the GM-PHD filter is used for the propagation of the Gaussian mixture posterior intensity, the reward function $R(a, Z_k)$ can be evaluated in closed form using Corollary 4.

Following [28], [39], we select the control $a$ to optimize $R(a, Z^*)$, where $Z^*$ is the ideal predicted measurement from the predicted intensity $v_{k-1}$, assuming zero clutter and unity detection rate.

D. Numerical example

This case study is based on an adapted from [28], where a mobile robot is tracking a varying number of moving targets. The surveillance area is a square of dimensions $1000m \times 1000m$. Each target in this area is characterized by a single-target state of the form $x_k = [p_k^T, p_k^T]^T$, where $p_k$ is the position vector and $v_k$ is the velocity vector. If at time $k$ the sensor is driven by the control vector $a_k$ to position $s_k(a_k)$, it detects a target at position $p$ with probability

$$ p_{D,k}(x_k; a_k) = \sqrt{2\pi|S|} \mathcal{N}(s_k(a_k); Hs_k, S) $$

(43)

where

$$ H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, S = 10^6 \begin{bmatrix} 3 & 2.4 \\ -2.4 & 3.6 \end{bmatrix}. $$

The detection profile is illustrated in Fig. 3.

The single-target transition density is $f(x_k|x_{k-1}) = \mathcal{N}(x_k; Fx_{k-1}, Q)$, where

$$ F = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, Q = 27 \begin{bmatrix} T^3 & 0 & T^2 & 0 \\ 0 & T^3 & 0 & T^2 \\ T^2 & 0 & T & 0 \\ 0 & T^2 & 0 & T \end{bmatrix} $$

with $T = 1s$.

Measurements are noisy position returns according to the single-target likelihood $g(z_k|x_k) = \mathcal{N}(z_k; Hs_k, R_k)$, where

$$ R_k = \sigma_{e,k}^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} $$

with $\sigma_{e,k} = 3m$.

Clutter is modeled by a Poisson RFS with intensity $\kappa(z) = \lambda \cdot c(z)$ with $\lambda = 20$ and $c(z) = U([0, 1000] \times [0, 1000])$ is the uniform density over the surveillance area.

The set of admissible control vectors $a_k$ is computed as follows. If the current position of the sensor is $s_k = [s_k^{(s)}, s_k^{(v)}]^T$, the set of all possible one-step ahead control actions are:

$$ a_k = \{ s_k^{(s)} + j \Delta R \cos(\ell \Delta \theta), s_k^{(v)} + j \Delta R \sin(\ell \Delta \theta) \}_{\ell=0,...,N_k} $$

where $\Delta \theta = 2\pi/N_k$ and $\Delta R = 50m$ is the radial step size. Other parameters are $N_k = 2$, $N_\theta = 8$, and $\Delta R = 50m$. We thus have $17$ control options in total. The observer is always kept inside the surveillance area by setting the reward function associated with control vectors outside the area to $-\infty$.

For this scenario, and for the purposes of performing sensor control, it is expected that a “good” control policy should, intuitively speaking, move the sensor towards the targets, and remain in their vicinity in order to obtain high detection probabilities. Fig. 4 illustrates a typical sensor trajectory that appears to be consistent with the expected behaviour of a “good” control policy as hypothesized above.
We proceed to illustrate and compare the performance of the proposed Cauchy-Schwarz divergence based control strategy with an existing Rényi divergence based control strategy [23]. In the case of the Cauchy-Schwarz divergence, in addition to the GM-PHD implementation which allows exact computation of the reward, we also compare with an SMC-PHD implementation which entails approximate computation of the reward via equation (42). In the case of the Rényi divergence, the SMC-PHD based implementation presented in [23] is used. All algorithms were implemented in MATLAB R2010b on a laptop with Intel Core i5-3360 CPU and 8GB of RAM. The average run time for the Rényi divergence based strategy is 10.62 seconds while those for Cauchy-Schwarz based strategies are 10.68 seconds (SMC-PHD implementation) and 3.21 seconds (GM-PHD implementation). As expected, the closest form Cauchy-Schwarz divergence based strategy is the fastest by roughly an order of magnitude.

Fig. 5 shows the averaged Optimal SubPattern Assignment (OSPA) metric [43] (with parameters \( p = 2, c = 100m \)) over 200 Monte Carlo runs for various control strategies. The OSPA curves in Figure 5 suggest that the GM-PHD based strategy outperforms its SMC-PHD counterparts, while the performance of the two approximate SMC-PHD based strategies are virtually identical. These results indicate that the closed form GM Cauchy-Schwarz based control strategy results in superior performance in terms of both miss distance and computational savings as compared to the Rényi based strategy. When paired with an SMC implementation, the multi-target Cauchy-Schwarz divergence has the same performance as the multi-target Rényi divergence, even though the former has the advantage of admitting a closed form solution with superior performance.

![Averaged OSPA](image)

**Fig. 5:** Averaged OSPA

### IV. CONCLUSIONS

In this paper, we have extended the Cauchy-Schwarz divergence to point processes by using the inner product of their probability densities and showed that for Poisson point processes, this divergence is half the squared distance between the intensity functions. Moreover, for Gaussian mixture intensity functions this divergence can be evaluated analytically. Our result is an addition to the list of interesting properties of Poisson point processes and has important implications in numerical approximations. An application of this result to sensor control for multi-target tracking has been studied.

### REFERENCES


