Indoor WLAN localization via adaptive Lasso Bayesian inference and convex optimization

Osamah A. Abdullah¹*, Mohammed Aal-nouman² and Aya Al-joudi³

Abstract: There has been growing interest in indoor positioning system technologies due to the important role of real-time indoor positioning services in modern technologies such as security services and emergency healthcare. Currently, many large companies such as Apple, Microsoft, and Google have researched location-based services (LBS), as they are key for network optimization and extensive computing applications. Multiple techniques were proposed using fingerprinting-based location methods due to their ability to obtain accurate results within several meters. However, their major drawback is that the received signal strength (RSS) can fluctuate with time and different environment, giving RSS distribution a multimodal distribution. Thus, in this paper, we established a framework consisting of k-mean-symmetrical-Hölder-divergence, a statistical model that encapsulates Cauchy-Schwarz divergence and skews Bhattacharyya divergence, to measure dissimilarities among signals that have multivariate distributions. In other words, the traditional k-mean is extended to meta-algorithms to detect the cluster that RSS is related to. Our second approach was hierarchical Bayesian model based on adaptive lasso criterion to recover sparse signals to optimize the accuracy of the indoor location estimation by solving the l1-minimization problem. The experimental results showed that the proposed system had substantially improved localization estimation accuracy compared to traditional fingerprinting-based localization methods.

ABOUT THE AUTHOR

I recently completed my PhD in electrical and computer engineering at Western Michigan University in December 2016. My teaching experience and practical skills in communication systems and signal processing, alongside with my research background, My current research focuses machine learning algorithms, UAV channel modeling, Brainwave signals detections, mobile communication, signal processing, and peer-to-peer/ad hoc networking, especially with applications to group localization, tracking, and monitoring. I’m a member of the Institute of Electrical and Electronics Engineers (IEEE).

PUBLIC INTEREST STATEMENT

There has been growing interest in indoor positioning system technologies due to the important role of real-time indoor positioning services in modern technologies such as security services and emergency healthcare. Currently, many large companies such as Apple, Microsoft, and Google have researched location-based services (LBS), as they are key for network optimization and extensive computing applications. Multiple techniques were proposed using fingerprinting-based location methods due to their ability to obtain accurate results within several meters. However, their major drawback is that the received signal strength (RSS) can fluctuate with time and different environment, giving RSS distribution a multimodal distribution. Thus many algorithms were proposed to improved localization estimation accuracy compared to traditional fingerprinting-based localization methods.
1. Introduction

Recently, location-based services (LBS) have garnered great interest due to the strong drive of location-based technologies (McGuire et al., 2005). They are considerably used for healthcare monitoring, emergency personnel navigation, personalized information delivery, network management and security, and context awareness. The Global Positioning System (GPS) and Global Navigation Satellite System (GNSS) (Patterson et al., 2003) are superior for outdoor positioning, but cannot provide indoor positioning navigation for two reasons: they require a direct line-of-sight between the user and the satellite; and they cannot locate the floor inside a building. Researchers have proposed many approaches for indoor localization, which can be divided into two categories: (1) non-radio-based positioning such as ultrasonic techniques, infrared (IR) waves, visible light, magnetic field exploitation, and inertial systems; and (2) radio-based positioning techniques such as ultra-wide band, radio-frequency identification, ZigBee-based techniques, Bluetooth-based techniques, frequency modulation techniques, and IEEE 802.11 wireless local area network (WLAN)-based techniques. IEEE 802.11 WLANs have widespread technology in indoor positioning techniques as they can be found in almost every building and do not need extra infrastructure in contrast to many other technologies that have high maintenance costs and need a massive amount of transceivers. IEEE 802.11 WLANs operate in the 2.4 GHz (IEEE 802.11b/g) and 5 GHz (IEEE 802.11a) unlicensed bands. However, other technologies transit simultaneously, which can lead to interference distortion.

Indoor positioning approaches for WLAN are categorized into two main categories: (1) lateration methods that consist of direction of arrival, angle of arrival, time of arrival, and time difference of arrival techniques; and (2) RSS exploitation (fingerprinting-based localization) (Abdullah & Abdel-Qader, 2016; Abdullah et al., 2016a). However, the lateration method suffers from poor localization accuracy in indoor environments. It was reported in one study (B. Wang et al., 2015) that localization distance error in a typical office environment is about 24.73 ft with a length of 200 ft and width 80 ft; this inaccurate estimation is due to two main reasons. First, the localization distance estimation is based on the propagation model, which is not accurate due to the complexity of the indoor environment and has a non-line-of-sight problem that can cause a multi-path signal. Second, the localization distance error can significantly increase if one or more access points (APs) are not correctly estimated (Abdullah & Abdel-Qader, 2018; Ji et al., 2006).

Fingerprinting radio map techniques do not need an infrastructure as they use APs in indoor environments. The fingerprinting radio map consists of two phases: the offline and online phases. During the offline phase, a multiset of predefined points referred to as landmarks or reference points (RPs) are determined in the area of interest (AoI). At each RP, a survey is conducted and the RSS value of available APs in the AoI is recorded for a specific time interval to generate a database of RSS values. During the online phase, the user collects the RSS values at his location and sends them to the server where specific algorithms are applied to find similar measurements in radio map fingerprints using the RPs to estimate the user’s location (Abdullah et al., 2016b).

In this paper, RSS fingerprinting database were recorded into four different direction during the offline phase. After that k-mean-symmetrical-Hölder-divergence was applied on the fingerprinting database in order categorized the cluster with the most similarity head cluster head. During the online phase, the similarity was computed between the online measurement and the cluster head and then categorized and selected with the least distance measurement as coarse location of the object (Abdullah, 2018). However, each cluster has a subset of RPs that is confined the location of the object called Region Of Interest (ROI) (Abdullah et al., 2017). Contrasting the traditional indoor
localization techniques which select the RPs and the APs based on the fingerprinting database, three different schemes were proposed to select the APs, random selection scheme, strongest APs scheme, and the Fisher criterion and then the comparison were made between these three different schemes. We utilize that the random selection scheme was the optimum scheme between them. After that, a novel framework that provide a Bayesian inference incorporate with adaptive Least Absolute Shrinkage and Selection Operator (LASSO) were proposed for fine localization, that formulate the sparse user location to find the position vector by finding the similar RPs and minimize the dissimilarity between the online measurement and the fingerprinting database. Table 1 summarizes the notation of this work.

2. Related work

Significant advances have been made in the area of wireless localization using RF signals. RSS maps were employed by Radar, in which the RSS values of the APs were recorded with their

| Symbol | Meaning |
|--------|---------|
| $q_i^j(\tau)$ | Direction of the orientation |
| $t$ | Time sample |
| $\Delta^i_j$ | Variance vector |
| $L$ | Number of aps |
| $N$ | Number of rps. |
| $\Delta^i_j$ | Variance for AP $i$ at RP $j$ |
| $\alpha, \theta$ | Conjugate exponents |
| $y$ | Absolute Lebesgue integrable |
| $p(x), q(x)$ | Scalar value |
| $\theta, \alpha$ | Conjugate exponents |
| IPS | |
| $c_i$ | Center |
| $C_i$ | Cluster |
| $\left| C_i \right|$ | Cardinality of $C_i$ |
| $c_l$ | Cluster centers |
| $l \in \{1, \ldots, L\}$ | Cluster labels |
| $\ell_1$ | Sparse recovery |
| $\theta = [0,0,1,\ldots,0]^T$ | Sparse location vector |
| $\psi$ | Fingerprinting database |
| $\epsilon$ | Error vector |
| $y$ | The RSS measurement |
| $\epsilon$ | Noise measurement |
| $\hat{\theta} = \arg \min ||\theta||_1$ | Convex optimization |
| $s : t \cdot \psi = \Phi \psi \theta$ | |
| $\lambda$ | Complexity parameter |
| $\Lambda^{-1}$ | Diagonal of $y$ and $y = [y_1, y_2, \ldots, y_N]^T$ |
| $\gamma$ | Hyperparameters |
| $\Phi$ | AP selection matrix |

Table 1. Work notation

Abdullah et al., Cogent Engineering (2020), 7: 1813677
https://doi.org/10.1080/23311916.2020.1813677

Page 3 of 16
coordinates at different locations. Bahl et al. (Feng et al., 2009) enhanced the performance of Radar by exploiting RSS properties such as scattering and multipath. In one study (Orr & Abowd, 2000), an approach was proposed to reduce the localization distance error by estimating the conditional property of different locations from nine APs, and then employing the speed of the user’s mobile to refine the speed. Kung et al. (2009) utilized the RSS power of APs on localization distance error and categorized the APs as “good” to represent the maximum power of the APs and “de-emphasizing” to represent the bad role of the APs. In one study (Youssef & Agrawala, 2005), the accuracy was significantly improved by using an autoregressive model that autocorrelated the RSS with the same AP. The RP was created based on the autocorrelation of APs, and then the variance and mean were estimated based on that RP. In most previous works, object location has been estimated using a probabilistic framework with the interpolation method. Chai and Yang (2007) proposed the statistical Markov model, termed the hidden Markov Model (HMM), by utilizing the trace of the user by developing a patch on the radio map. The proposed algorithm compensated between the accuracy for further improvement on the radio map by exploiting the motion constraint. Then a multivariate Gaussian model-based algorithm was used to estimate the location of the object. In general, probabilistic approaches have better accuracy than deterministic approaches as they provide one point for estimating the user’s location. To improve accuracy and reduce complexity, (Miliaris et al., 2014) used multivariate Gaussian by exploiting the mean vector and covariance matrix and employing them in Kullback–Leibler divergence, especially its one-iteration method compared HMM-based localization algorithm that needs multiple iterations to obtain the converge. Recently the CS algorithm has motivated researchers to apply many emerging applications; nevertheless, CS-LPS is still in the early stage. Orr and Abowd (2000) proposed CS using a framework and taking advantage of the spatial sparsity by using constrained l1 norm based on a learning dictionary by taking the RSS average measurement and considering it as a signature for the WLAN. Then the dictionary learning was established by concatenating each cell and the unknown cell was projected to the dictionary-learning algorithm to form the vector of measurement. Several studies (Zhao et al., 2017) have proposed a novel approach by leveraging the sparse signal of the RSS measurement. These techniques open a new ear for deterministic approaches by dividing the RSS measurement into a subset of small vectors containing indices of each nonzero measurement. So the object location can be estimated by locating the sparse position by solving the l-minimization through the RSS measurement in the online phase and the RSS measurement of the offline phase.

(Kalistatov, 2019) investigated in magnetic transport construction through wireless video monitoring of traffic flows through calculating radio channel and find the advantage and disadvantage of the infrastructure. (Astaneh & Gheisari, 2018) proposed about cognitive network and Its routing parameter and showed not efficient methods of allocating static spectrum due to lack of spectrum, they compared routing metrics to analyze its challenges in multi-route routing and in one-way routing. (X. Wang et al., 2020) investigated the effect of the passenger flowing in urban road and how the weather effect on daily ridership rate. While in (Çokl et al., 2019) investigated about the effect of stink bug on multimodal signals that have been transmitted the substrate and air, and how that can affect hormone aggregation pheromones as uni- or multicomponent signals.

3. Indoor positioning system
Fingerprinting data collecting is considered the main task in IPS in the offline phase. A person holding the mobile phone has WLAN access to collect RSS data for different locations from various APs. This process was performed at the College of Engineering and Applied Sciences at Western Michigan University (Kalamazoo, MI, USA). It was observed that RSS data have multimodal distribution due to many reasons such as multipath signal and scattering. For confirmation, an experiment was performed in which the signal-to-noise ratio (SNR) was collected for 35 minutes for a single AP in a long corridor by using a mobile robot. It was found that the SNR value differed by 10 dBm as shown in Figure 1.
4. Offline phase

4.1. IPS data collection

The human body of the person who holds the mobile phone is an obstacle for the signal as well as for the pedestrians, so the radio map fingerprint is recorded into four directions within the time sample where the directions are 45°, 135°, 225°, and 315°, and the time sample is denoted as \( \{ q_{ij}^{(r)}, r = 1, \ldots, t, t = 100 \} \). \(^{(r)}\) represents the direction of the orientation and \( t \) is the time sample. For four directions, the average and covariance matrix were calculated to generate the fingerprinting database represented as \( Q^{(1)} \)

\[
Q^{(1)} = \begin{pmatrix}
q_{11}^{(1)} & q_{12}^{(1)} & \cdots & q_{1N}^{(1)} \\
q_{21}^{(1)} & q_{22}^{(1)} & \cdots & q_{2N}^{(1)} \\
\vdots & \vdots & \ddots & \vdots \\
q_{L1}^{(1)} & q_{L2}^{(1)} & \cdots & q_{LN}^{(1)}
\end{pmatrix}
\]

(1)

where \( t = 10 \) and \( q_{ij}^{(r)} = \frac{1}{t} \sum_{r=1}^{t} q_{ij}^{(r)} \) randomly chosen for 100 time samples. The variance vector of each RP was calculated as

\[
\Delta_{ij}^{(1)} = \left[ \Delta_{ij}^{(1)}, \Delta_{ij}^{(2)}, \Delta_{ij}^{(3)}, \ldots, \Delta_{ij}^{(4)} \right]
\]

(2)

where \( L \) is the number of APs and \( N \) is the number of RPs.

\[
\Delta_{ij}^{(1)} = \frac{1}{t-1} \sum_{r=1}^{t} (q_{ij}^{(r)} - q_{ij}^{(1)})^2
\]

(3)

where \( \Delta_{ij}^{(1)} \) is the variance for AP \( i \) at RP \( j \) with orientation \( \theta_{ij}^{(1)} \); so the radio map of the offline phase is

\((x_{ij}, y_{ij}, q_{ij}^{(1)}, \Delta_{ij}^{(1)})\) with \( q_{ij}^{(1)} \) defined as:

\[
q_{ij}^{(1)} = \begin{pmatrix}
q_{i1}^{(1)} & q_{i2}^{(1)} & q_{i3}^{(1)} & \cdots & q_{iN}^{(1)}
\end{pmatrix}
\]

(4)

and the RSS measurement is represented as

\( p_{r} = [p_{1}, p_{2}, \ldots, p_{L}] \)

(5)
4.2. Clustering based on K-MEAN Hölder divergence

In this section, symmetric Hölder Divergence was produced, to calculate the dissimilarity between \( p \) and \( q \), by using bi-parametric inequalities Lhs (\( p, q \)) \( \leq \) rhs (\( q, p \)), where Lhs and rhs denote the left-hand side and right-hand side, respectively, and use the log ratio gap:

\[
D(p : q) = - \log \left( \frac{\text{Lhs}(p, q)}{\text{rhs}(p, q)} \right) = \log \left( \frac{\text{rhs}(p, q)}{\text{Lhs}(p, q)} \right) \geq 0
\]  

(6)

The Hölder divergence measurement between \( p(x) \) and \( q(x) \) is

\[
D_H^\alpha(p : q) = - \log \left( \frac{\int p(x)q(x)^{\frac{1}{\alpha}}dx}{\left( \int p(x)^{\frac{1}{\beta}}dx \right)^{\frac{\alpha}{\beta}} \left( \int q(x)^{\frac{1}{\alpha}}dx \right)^{\frac{\beta}{\alpha}}} \right)
\]

(7)

where \( \alpha, \beta \) represents the conjugate exponents and \( \gamma \) represents the value of absolute Lebesgue integrable, \( p(x) \) and \( q(x) \) are the scalar value. Hölder divergence is an indiscernible law, which means the yield of \( p(x) = q(x) \) is not equal to zero. To solve this issue, the Cauchy-Schwarz divergence skew Bhattacharyya divergence are encapsulated together which yield to Cauchy-Schwarz divergence if \( \gamma, \beta, \alpha \) where set to two:

\[
D_{2,2}^H(p : q) = CS(p : q) := - \log \left( \frac{\int p(x)q(x)dx}{\left( \int p(x)^{2}dx \right)^{\frac{1}{2}} \left( \int q(x)^{2}dx \right)^{\frac{1}{2}}} \right)
\]

(8)

If we set \( \gamma = 1 \) the Hölder divergence will produce skew Bhattacharyya divergence:

\[
D_{1,1}^H(p : q) = B_{1/1}(p : q) := - \log \left( \frac{\int p(x)^{1/\alpha}q(x)^{1/\beta}dx}{\left( \int p(x)^{1/\alpha}dx \right)^{\frac{1}{\beta}} \left( \int q(x)^{1/\beta}dx \right)^{\frac{1}{\alpha}}} \right)
\]

(9)

The relationship between the divergence families is illustrated in Figure 2.

The Hölder divergence satisfies the conjugate exponents \( \theta \) and \( \alpha \).

\[
D_{\alpha,\beta}^H(p : q) = D_{\beta,\alpha}^H(p : q)
\]

(10)

The symmetrized Hölder divergence is:

\[
S_{a,\alpha}^H(p : q) = \frac{1}{2} \left( D_{a,\alpha}^H(p : q) + D_{a,\alpha}^H(q : p) \right)
\]

(11)

\[
= \frac{1}{2} \left[ F(\gamma p) + F(\gamma q) - F(\frac{\gamma}{\alpha} p + \frac{\gamma}{\beta} q) - F(\frac{\gamma}{\beta} p + \frac{\gamma}{\alpha} q) \right]
\]

(12)

Figure 2. Hölder divergence encompasses the skew Bhattacharyya divergence and the Cauchy-Schwarz divergence.
To improve the IPS accuracy, we proposed that k-mean incorporated symmetrized Hölder divergence. The k-mean algorithms first introduced by Lloyd in 1957 to solve vector quaternion problems by choosing seeds from data points as initial centroid and each point will be associated to the closet center. Each center will be reiterated and updated with each calculation step until the difference between each successive calculation will be below the threshold value. In general, k-mean uses squared Euclidean distance to calculate the dissimilarity between the center and the other points. The center $c_i$ and cluster $C_r$s is defined as follows:

$$c_i = \arg \min_{p_i \in C_i} \sum_{p_i \in C_i} \|p_j - c_i\|$$

(13)

$$= \arg \min \text{AVG}_{L_j}(C_i, c)$$

(14)

$$c_i = \frac{1}{|C_i|} \sum_{p_j \in C_i} p_j$$

(15)

where $|C_i|$ denotes the cardinality of $C_i$.

Given the data of the fingerprinting radio map database k-mean Hölder divergence was performed based on $S^{a,(p:q)}(\tilde{p} : \tilde{q})$ (Nielsen et al., 2017). The cost function of k-mean Hölder divergence is

$$E = \sum_{i=1}^{n} S^{a,(p_i : c_i)}$$

(16)

where $c_i$ are the cluster centers, and $l_i \in \{1, \ldots, L\}$ are the cluster labels.

5. Model of sparse recovery modification

Conventional fingerprinting techniques have high computationally complexity; nevertheless, CS has introduced a new methodology in wireless localization with high accuracy and acceptable time. In previous works, the online phase has been considered one of the RPs of the radio map. Wireless localization with sparse information can be programed as an indicator to find the pose vector, and the element can be recast as a one-sparse vector $\ell_1$-minimization problem. The sparse recovery in wireless indoor localization can be represented as:

$$y = \Phi \Psi \theta + \varepsilon$$

(17)

where $\theta = [0, 0, 1, \ldots, 0]^T$ represents the sparse location vector where 1 indicates the index where the is the nearest; $\Phi$ represents the selection matrix of APs, from the fingerprinting database $\Psi$, $\varepsilon$ represents the error vector and $y$ represents the RSS measurement of the online phase from specific APs as:

$$y = \Phi \gamma$$

(18)

In general the CS does not count the noise measurement $\varepsilon$ into the calculation, and under-determines the problem due to the fact that the size of $\theta$ is greater than the size of $y$. Nevertheless, $\theta$ only uses one RP location and is solved by using convex optimization

$$(\hat{\theta}) = \arg \min_{\theta} ||\theta||_1$$

(19)
is the \( \ell_1 \)-norm of \( \theta \). Still, this optimization algorithm does not have the error measurement \( \epsilon \).

Many algorithms have been proposed to deal with error measurement such as basis pursuit (Jin et al., 2010) and greedy algorithms (Zou, 2006). The previous techniques had a fine localization method accompanied by offline clustering and other optimization algorithms. In general, offline clustering has a number of a prior clustering that enforce the data to be divided into a specific number of clusters, and only one or more will be chosen even if the user between them. Moreover, the \( \Psi \) matrix has to obey two conditions to obtain optimal results: first, the coherence between \( \Psi \) and \( \Phi \) must be small; second, the product of \( \Psi \) and \( \Phi \) has to be orthogonal (Alhamzawi & Ali, 2018). Nevertheless, the orthogonality cannot be completely achieved since is not a square.

In general, the main goal of indoor localization is to estimate the location of the object using WLAN signal by using the RSS reading of online phase. Essentially, we are looking for an algorithm that can map the fingerprinting database. However, the sparse vector minimization has tremendous zeroes on less probable RPs, but the similarity between the radio map fingerprint and the online phase is quite precise, with small residuals. Henceforth, we propose a convex algorithm the formulate \( l_1 \) norm and \( l_2 \) as

\[
\hat{\theta} = \arg\min_{\theta} \left[ \frac{1}{|Z|} \| Y - H \theta \|^2 + P_\alpha \right] \tag{20}
\]

\[
P_\alpha = \lambda \left( (1 - \alpha) \| \theta \|^2 + \alpha \| \theta \|_1 \right) \tag{21}
\]

where \( 0 \leq \alpha \leq 1 \) adjustment between Lasso and ridge regression and \( \lambda \) is a complexity parameter. Ridge regression intend to reduce the coefficient of correlated fingerprinting database by using \( \| \theta \|^2 \). If the \( \alpha \) set 0 the above equation will be ridge regression, while if \( \alpha \) was set between 0 and 1 the \( \lambda \) will have sparsity monotonically lead to LASSO solution.

6. Bayesian adaptive Lasso

The penalize of \( \ell_1 \)-norm is proportional to the size of the signal component, and that will lead to suboptimal solutions. (Alhamzawi & Ali, 2018) proposed that the different weighting coefficients of the \( \ell_1 \) is considered as different entries of the sparse vector \( \gamma \). After that Largrangian argument was used to improve the accuracy and the adaptive Lasso is expressed as:

\[
\gamma_{\text{odlasso}} = \arg\min_{w} \left\{ \| Y - \Phi w \|^2 + \sum_{i=1}^{N} \mu_i |w_i| \right\} \tag{22}
\]

where \( \mu_i > 0, i = \) Since the Lasso and adaptive Lasso both are sparse, so choosing the suitable parameters of the penalization will lead to a better performance. (Themelis et al., 2012) developed a hierarchical Bayesian model through independent Laplace priors.

Since the independence of \( w_i \) is among of \( \gamma \), \( \gamma \) can be expressed as multivariate Gaussian distribution i.e.,

\[
p(w|\gamma) = \prod_{i=1}^{N} N(w_i|0, \gamma_i) \tag{23}
\]

\[
= (2\pi)^{-\frac{N}{2}} |\Lambda|^{\frac{1}{2}} \exp \left[ -\frac{1}{2} w^T \Lambda w \right] = N(w|0, \Lambda^{-1}) \tag{24}
\]
where $\Lambda^{-1}$ is the diagonal of $\gamma$, and $\gamma = [\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n]^T$, furthermore, the conjucate of Gamma distribution is expressed as:

$$p(\beta; k, \theta) = G(\beta|k, \theta) = \frac{\theta^k}{\Gamma(k)} \beta^{k-1} \exp[-\theta\beta]$$  \hspace{1cm} (25)

The hyperparameters $\gamma$ is defined as:

$$p(\gamma|\lambda) = \prod_{i=1}^{N} p(\gamma_i|\lambda_i) = \prod_{i=1}^{N} \left[ \frac{1}{2\pi} \lambda_i \exp \left[ -\frac{1}{2\lambda_i} \gamma_i \right] \right]$$  \hspace{1cm} (26)

$$= \left( \frac{1}{2} \right)^N |\Psi| \exp \left[ -\frac{1}{2\lambda} \sum_{i=1}^{N} \lambda_i \gamma_i \right]$$  \hspace{1cm} (27)

Where $N$ represents the independent of exponential distributions, $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_N]^T$, $\lambda_i > 0$ and $\Psi = \text{diag}(\lambda)$.

In conclusion, to define the hierarchical Bayesian algorithm with adaptive Lasso with multivariate Laplace distribution as prior of $\gamma$ is defined as:

$$p(w|\lambda) = \prod_{i=1}^{N} p(w_i|\gamma_i)p(\gamma_i|\lambda_i)d\gamma_i$$  \hspace{1cm} (28)

$$= 2^{-N|\Psi|^{1/2}} \exp \left[ -\sum_{i=1}^{N} \sqrt{\lambda_i} |w_i| \right]$$  \hspace{1cm} (29)

The MAP estimator is defined as:

$$w_{\text{MAP}} = \arg \min_w \left[ -\log p(y|w, \beta)p(w|\lambda) \right]$$  \hspace{1cm} (30)

$$= \arg \min_w \left\{ \|y - \Phi w\|^2 + \frac{2}{\beta} \sum_{i=1}^{N} \sqrt{\lambda_i} |w_i| \right\}$$  \hspace{1cm} (31)

**7. Maximum likelihood approach using adaptive Lasso Bayesian inference**

In general, Bayesian inference is depend on posterior distribution that is defined as:

$$p(w, \gamma, \lambda, \beta|y) = \frac{p(w, \gamma, \lambda, \beta|y)}{p(y)}$$  \hspace{1cm} (32)

By maximizing the marginal likelihoood we maximize the likelihood approach by integrating $w$ with respect to the other unknown parameters. Therefore we can make the posterior distribution as multivariate Gaussian.
where

\[ \Sigma = [\beta \Phi^T \Phi + \lambda]^{-1} \quad \text{and} \quad \mu = \beta \Sigma \Phi^T y \]

By using Bayes Law \( p(y, \gamma, \lambda, \beta) \) is maximized by integration \( p(w, y, \gamma, \lambda, \beta) \) out of \( w \) i.e.

\[
p(y, \gamma, \lambda, \beta) = \int p(w, y, \gamma, \lambda, \beta, y) dw = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y^T C^{-1} y \right) p(y|\lambda)p(\lambda)p(\beta)
\]

where \( C = \beta^{-1} I_m + \Phi^T \Lambda^{-1} \Phi \)

### 8. Experimental results

In this section experimental evaluation is elaborated for the proposed algorithm on real data environment. Large number of smartphones devices are working on Linux-based android operating system, the RSS data of fingerprinting database were collected using a Samsung S5 mobile phone functioning on Android 4.4.2 operating system through an android application by using the inherent chip of Wi-Fi package. However, to create the fingerprinting database data (the MAC address and RSS value) within time sample frame from 1s to 100 s for 84 RPs for four direction within a grid of 1 m. 47 active APs were detected over the layout shown in Figure 3. Furthermore, to evaluate the performance of the proposed algorithms, the online data were collected in different environment during different times of a day for 65 unknown locations as test points.

Most of the previous work worked on uniform granularity of RPs. In general Wi-Fi signal were not created to be used in localization, which means even though high dense granularity RPs can lead to low localization error nevertheless, a dense granularity RPs have high computational cost with great redundancy. But on the other hand, due to the 1-sparse vector property, a dense granularity RPs is beneficial in sparsity algorithms which could lead to high precision positions system. The evaluation of the proposed algorithms depend is measured by using the average Euclidean distance between the estimated location of the test point and the true location of it known as Mean Average Error (MAE).

### 9. A clustering-based approach for improving the accuracy of LPS

Sparse recovery localization technique performance is decidedly depending on the number of the fingerprinting database and the number of the APs that have been used in the measurement techniques. For instance, the number of measurements in CS-based localization method has to be

![Figure 3. The layout used in the experimental work in the College of Engineering and Applied.](image-url)
in order of log $N$ to imply the sparse recovery of the positioning vector. Nevertheless, the fingerprinting database has to be correlated which demand large number of measurements and the have a large number of APs to have a sparse recovery which considered a big problem in LPS. However, this problem was alleviated in adaptive Lasso Bayesian inference because it has its own recovery correlated scheme. Another approach was proposed to enhance the localization quality, by clustering the fingerprinting database by using K-mean Hölder Divergence, since the clustering the database will decrease the number of RPs that have to process with proposed technique. Figure 4 shows the outperformed comparison of K-mean Hölder Divergence compared to traditional K-mean with respect to localization distance error with respect to the number of APs. The proposed algorithms with clustering methods show lower distance error with lower number of APs. For instance, the lowest localization distance error was obtained with 22 APs for adaptive Lasso Bayesian inference, with proved that the clustering methods can improve the localization quality with small number of APs.

Figure 5 shows the localization distance error (m) of the adaptive Lasso Bayesian inference and CS with the increasing of APs. When the number of APs were low the difference in localization distance error is high, which shows the high impact of the clustering in adaptive Lasso Bayesian inference and CS. the best results were obtained when 22 or more APs were used.

10. AP selection methods
AP selection have different methods that can be performed in both online and offline phase. If it performed in the offline phase, then the AP selection will be only depend on the fingerprinting database regardless the data of the online measurement. Conversely the way of selection it may have a big fail if the online characteristics environment was different from the offline characteristics environment. Henceforth the APs selections during the online phase utilizing implicitly or
explicitly characterization. During the implicit utilization the APs selection will be exploiting in offline and online phases, while in the explicit utilization the APs selection will be using only during the online measurements. This process was formulated as fellows:

$$\Phi^i = [\ldots, 1, \ldots], i = 1, \ldots, L'$$

where $\Phi$ is the AP selection matrix, of the APs $\Gamma \subseteq \Gamma'$, $L' \leq L$ be the cardinality of $\Gamma'$.

### 10.1. Strongest APs

Early techniques advocate that the APs selections should be on their signal strength in fingerprinting database as well as in online phase. The perception behind is that the APs with the strongest RSS have the most coverage and accurate measurements. Nevertheless, this assumption may not suitable for a render criterion (Feng et al., 2012).

### 10.2. Fisher criterion

Fisher criterion is a metric that computes the discrimination capacity of the AP across the fingerprinting database by considering the stability of the AP across the RPs. Fisher criterion uses the statistical properties of the RPs by depending on the performance of the APs during the offline and the online phases which the score can be calculated as fellows:

$$\zeta_i = \frac{\sum_{j=1}^{N} (q_j^{(0)} - \bar{q})^2}{\sum_{j=1}^{N} (q_j^{(0)})}$$

$$\bar{q} = \frac{1}{\text{N}} \sum_{j=1}^{\text{N}} q_j^{(0)}$$

This technique is based on the fact if the APs has a smaller score only if it has a higher variance value and that would indicate that the APs are less reliable to consider. Nevertheless the fisher criterion may lead to faulty measurement if one of the APs was appeared in fingerprinting database only and didn’t appear in online phase (Youssef et al., 2003).

### 10.3. Random selection

Contrasting the other two schemes, in this scheme the APs were arbitrarily selected despite the performance of the APs. In this scheme the computational complexity is less than the other two schemes that didn’t need to calculate the variance, by running a matrix to select different AP at each run and then select the subset with high performance.

Figure 6 shows the comparison between different AP selections scheme with respect the localization distant error with adaptive Lasso Bayesian inference where the x-axis represents the number of APs and the y-axis is the localization distance error. The better results were obtained with random scheme. The performance of Fisher criterion and the strongest AP where so close when the number of APs increased. The best results were obtained when 22 APs were used. Thus, we conclude the selection scheme of APs affect the performance of the of the proposed system that not only the number.

### 11. Online computing and prior work comparison

The fine estimation of object location is highly depending on the number of the APs that were used in the proposed algorithm. Different technique and numbers of APs were used for our proposed algorithms to estimate the localization distance error. The results were compared with different prior techniques, such as, CS-based localization method (Feng et al., 2012). Kernel Density Estimation (KDE) (Youssef et al., 2003) and Weighted KNN (WKNN) (Themelis et al., 2012). Our implementation of CS-based localization contained all the steps that were implemented in (Han...
et al., 2015) expect the coarse schemes which have been replaced by offline clustering. Figure 7 shows the localization distance error of different proposed algorithm compared to the proposed algorithms, KDE, WKNN have the highest localization distance error while the lowest localization error was with the proposed algorithm at 22 APs. The results showed that the proposed algorithms have. It was observed the localization error was high with CS-based localization even though large number of APs were used. Adaptive LASSO showed a better performance when large number of APs were used.

Figure 8 depicts the Cumulative Distribution Function (CDF) of the localization error for the proposed algorithm scheme along with some will known localization algorithms scheme. Our proposed algorithm surpasses the other schemes which give around 0.9 m at % 90 or the error in this range while the CS-localization is more than 1.2 m at % 90.

12. Conclusion
Nowadays, the WLAN LPS has gained a great attention due to the existing infrastructure, low cost positioning, and easy to implement. The WLAN fingerprinting-based localization methods became very famous due to the extraordinary performance in real environments. The WLAN propagation has a very complex phenomenon that the signal is distorted by scattering, reflection and multipath, which make the traditional techniques such, AOA, TOA will not lead to good performance. To reduce the affect of multipath signal and time variation we used a symmetrical Hölder divergence clustering method. The fingerprinting radio map were created in CEAS and many algorithms were used to investigate the localization distance error, such as, CS, kNN, and PNN. We proposed a novel hierarchical Bayesian by depending on adaptive lasso criterion that is based on type-II maximum likelihood...
methodology. The results showed an adequate result that the localization distance error was less than 1 m. For accurate results we have used different AP selections schemes to reduce the multimodal distribution of the RSS signal. Furthermore, the adaptive Lasso was also used to obtain the best results through small number of RPs. The feasibility of using adaptive Lasso to obtain the location of the object from a small number of RPs. Now, we are in the process to investigate to make the object localization within smaller cluster with more stable localization mechanism with less number of APs. Also, new algorithms were built for more accurate results, and robust system that can keep and update the database automatically without manual work needed.

Funding
The authors received no direct funding for this research.

Author details
Osamah A. Abdullah1
E-mail: osamah.abdullah@wmich.edu
ORCID ID: http://orcid.org/0000-0002-6953-7346
Mohammed Aal-nouman2
E-mail: mi.aal-nouman@edu.salford.ac.uk
Aya Al-joudi3
E-mail: aya_ahmed@yahoo.com
1 Department of Information and communications engineering, College of Information Engineering, Al-Ma’moon University College, Baghdad, Iraq.
2 Department of Information and communications engineering, College of Information Engineering, Al-Nahrain University, Baghdad, Iraq.
3 Central Organization for Standardization and Quality Control (COSQC), Baghdad, Iraq.

Citation information
Cite this article as: Indoor WLAN localization via adaptive Lasso Bayesian inference and convex optimization, Osamah A. Abdullah, Mohammed Aal-nouman & Aya Al-joudi, Cogent Engineering (2020), 7: 1813677.

References
Abdullah, O. (2018, August). Convex optimization via symmetrical h"{o}lder divergence for a WLAN indoor positioning system. Entropy, 20(9), 639. https://doi.org/10.3390/e20090639
Abdullah, O., & Abdel-Qader, I. (2016, May), A PNN- Jensen-Bregman divergence symmetrization for a WLAN indoor positioning system. 2016 IEEE international conference on electro information technology (EIT), Grand Forks, North Dakota, USA.
Abdullah, O., Abdel-Qader, I., & Bazuin, B. (2016a, October). Fingerprint-based technique for indoor positioning system via machine learning and convex optimization. 2016 IEEE 7th annual ubiquitous computing, electronics & mobile communication conference (UEMCN), New York City, USA.
Abdullah, O., Abdel-Qader, I., & Bazuin, B. (2016b, October). K-means-Jensen-Shannon divergence for a WLAN indoor positioning system. 2016 IEEE 7th annual ubiquitous computing, electronics & mobile communication conference (UEMCN), New York City, USA.
Abdullah, O. A., & Abdel-Qader, I. (2018, September). Machine learning algorithm for wireless indoor localization. Machine Learning - Advanced Techniques and Emerging Applications. DOI: 10.5772/intechopen.74754
Abdullah, O. A., Abdel-Qader, I., & Bazuin, B. (2017, January). Convex optimization via Jensen-Bregman divergence for WLAN indoor positioning system. International Journal of Handheld Computing Research, 8(1), 29–41. https://doi.org/10.4018/IJHCR.2017010103
Alhamzawi, R., & Ali, H. T. M. (2018, September). The Bayesian adaptive lasso regression. Mathematical Biosciences, 303, 75–82. https://doi.org/10.1016/j.mbs.2018.06.004
Astaneh, A. A., & Gheisari, S. (2018, September). Review and comparison of routing metrics in cognitive radio networks. Emerging Science Journal, 2(4). https://doi.org/10.28991/esj-2018-01143
Chai, X., & Yang, Q. (2007, June). Reducing the calibration effort for probabilistic indoor location estimation. IEEE Transactions on Mobile Computing, 6(6), 649–662. https://doi.org/10.1109/TMC.2007.1025
Čokl, A., Žunič-Kosi, A., & Laumann, R. A. (2019, December). Stink Bug communication with multimodal signals transmitted through air and substrate. Emerging Science Journal, 3(6), 407–424. https://doi.org/10.28991/esj-2019-01203
Feng, C., Au, W. S. A., Valae, S., & Tan, Z. (2009, December). Orientation-aware indoor localization using affinity
propagation and compressive sensing. 2009 3rd IEEE international workshop on computational advances in multi-sensor adaptive processing (CAMSAP), Radisson Aruba Resort, Casino & Spa Aruba, Dutch Antilles.

Feng, C., Au, W. S. A., Valsee, S., & Tan, Z. (2012, December). Received-signal-strength-based indoor positioning using compressive sensing. IEEE Transactions on Mobile Computing, 11(12), 1983–1993. https://doi.org/10.1109/TMC.2011.216

Han, S., Zhao, C., Meng, W., & Li, C. (2015, June). Cosine similarity based fingerprinting algorithm in WLAN indoor positioning against device diversity. 2015 IEEE international conference on communications (ICC), London.

Ji, Y., Biaz, S., Pandey, S., & Agrawal, P. (2006). ARIADNE. Proceedings of the 4th international conference on mobile systems, applications and services - MobiSys 2006, Uppsala, Sweden.

Jin, J., Gu, Y., & Mei, S. (2010, March). An introduction to compressive sampling and its applications. Journal of Electronics Information & Technology, 32 (2), 470–475. https://doi.org/10.3724/SP.J.1146.2009.00497

Kalistatov, K. D. (2019, May). Wireless video Monitoring of the Megacities Transport Infrastructure. Civil Engineering Journal, 5(5), 1033–1040. https://doi.org/10.28991/cej-2019-03091309

Kung, H. T., Lin, C.-K., Lin, T.-H., & Vlah, D. (2009). Localization with snap-inducing shaped residuals (SISR). Proceedings of the 15th annual international conference on mobile computing and networking - MobiCom ’09, Beijing, China.

Kushki, A., Plataniotis, K. N., & Venetsanopoulos, A. N. (2007, June). Kernel-based positioning in wireless local area networks. IEEE Transactions on Mobile Computing, 6(6), 689–705. https://doi.org/10.1109/TMC.2007.1017

Ladd, A. M., Bekris, K. E., Rudys, A., Kavaki, L. E., & Wallach, D. S. (2005, January). Robotics-based location sensing using wireless ethernet. Wireless Networks, 11(1–2), 189–204. https://doi.org/10.1007/s11276-004-4755-8

McGuire, M., Plataniotis, K. N., & Venetsanopoulos, A. N. (2005, March). Data fusion of power and time measurements for mobile terminal location. IEEE Transactions on Mobile Computing, 4(2), 142–153. https://doi.org/10.1109/TMC.2005.24

Millioris, D., Tzagkarakis, G., Papakonstantinou, A., Papadopouli, M., & Tsakalides, P. (2014, January). Low-dimensional signal-strength fingerprint-based positioning in wireless LANs. Ad Hoc Networks, 12, 100–114. https://doi.org/10.1016/j.adhoc.2011.12.006

Nielsen, F., Sun, K., & Marchand-Maillet, S. (2017). k-Means Clustering with Hölder Divergences. Geometric Science of Information, 856–863. doi:10.1007/978-3-319-68445-1_98

Orr, R. J., & Abowd, G. D. (2000). The smart floor. CHI ’00 extended abstracts on human factors in computing systems - CHI ’00, Netherlands.

Patterson, C. A., Munz, R. R., & Pancake, C. M. (2003, April). Challenges in location-aware computing. IEEE Pervasive Computing, 2(2), 80–89. https://doi.org/10.1109/MPRV.2003.1203757

Themelis, K. E., Rontogiannis, A. A., & Koutroumbas, K. D. (2012, February). A novel hierarchical bayesian approach for sparse semisupervised hyperspectral unmixing. IEEE Transactions on Signal Processing, 60(2), 585–599. https://doi.org/10.1109/TSP.2011.2174052

Wang, B., Zhou, S., Liu, W., & Mo, Y. (2015, January). Indoor localization based on curve fitting and location search using received signal strength. IEEE Transactions on Industrial Electronics, 62(1), 572–582. https://doi.org/10.1109/TIE.2014.2327595

Wang, X., Guo, Y., Bai, C., Liu, S., Liu, S., & Han, J. (2020, January). The effects of weather on passenger flow of urban rail transit. Civil Engineering Journal, 6(1), 11–20. https://doi.org/10.28991/cej-2020-03091449

Youssef, M., & Agrawala, A. (2005). The horus WLAN location determination system. Proceedings of the 3rd international conference on mobile systems, applications, and services - MobiSys ’05, Seattle Washington.

Youssef, M. A., Agrawala, A., & Udoaya Shankar, A. (2003). WLAN location determination via clustering and probability distributions. Proceedings of the first IEEE international conference on pervasive computing and communications, Fort Worth, Texas, USA. (PerCom 2003).

Zhao, Y., Li, W., Wu, J., Lu, S., & Chen, B. (2017). Compressed RSS measurement for communication and sensing in the Internet of Things. Wireless Communications and Mobile Computing, 2017, 1–11. doi:10.1155/2017/6345316

Zou, H. (2006, December). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418–1429. https://doi.org/10.1198/016214506000000735
