ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10094833

ADAPTIVE FILTERING ALGORITHMS FOR SET-VALUED OBSERVATIONS–SYMMETRIC MEASUREMENT APPROACH TO UNLABELED AND ANONYMIZED DATA

Vikram Krishnamurthy

Cornell University, vikramk@cornell.edu

ABSTRACT

Suppose L simultaneous independent stochastic systems generate observations, where the observations from each system depend on the underlying parameter of that system. The observations are unlabeled (anonymized), in the sense that an analyst does not know which observation came from which stochastic system. How can the analyst estimate the underlying parameters of the L systems? Since the anonymized observations at each time are an unordered set of L measurements (rather than a vector), classical stochastic gradient algorithms cannot be directly used. By using symmetric polynomials, we formulate a symmetric measurement equation that maps the observation set to a unique vector. We then construct an adaptive filtering algorithm that yields a statistically consistent estimate of the underlying parameters.

1. INTRODUCTION

The classical stochastic gradient algorithm operates on a *vector-valued* observation process that is inputted to the algorithm at each time instant. Suppose due to anonymization, the observation at each time is a *set* (i.e., the elements are unordered rather than a vector). Given these anonymized observation sets over time, how to construct a stochastic gradient algorithm to estimate the underlying parameter?

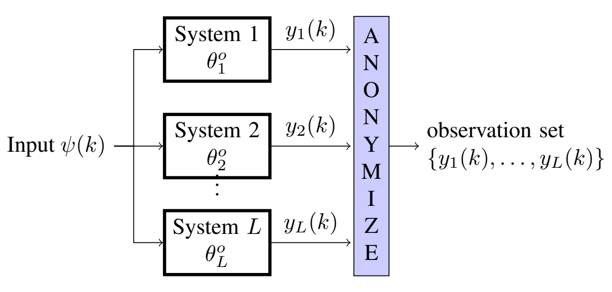


Fig. 1: Schematic setup comprising L stochastic systems. Given the sequence of anonymized observation sets $(\{y_1(k), \ldots, y_L(k)\}, k = 1, 2, \ldots)$, the aim is to estimate the underlying parameter set $\theta^o = \{\theta_1^o, \ldots \theta_L^o\}$ of the L systems.

This research was supported by the National Science Foundation grant CCF-2112457, U.S. Army Research Office grant W911NF-21-1-0093 and US. Air Force Office of Scientific Research grant FA9550-22-1-0016.

Figure 1 shows the schematic setup comprising L simultaneous independent stochastic systems indexed by $l = 1, \ldots, L$, evolving over discrete time $k = 1, 2, \ldots$. Each stochastic system l is parametrized by true model $\theta_l^o \in \mathbb{R}^D$ and generates observations $y_l(k) \in \mathbb{R}^D$ given input signal $D \times D$ dimensional matrix $\psi(k)$:

$$y_l(k) = \psi(k)\, \theta_l^o + v_l(k), \quad l \in [L] \stackrel{\text{defn}}{=} \{1, \ldots, L\} \tag{1}$$

We assume that $v_l(k) \in \mathbb{R}^D$ is an iid random sequence with bounded second moment. We (the analyst) know the input signal sequence $(\psi(k), k = 1, 2, \ldots)$. For convenience, assume that elements of $(\psi(k), k = 1, 2, \ldots)$ are zero mean iid sequences of random variables. Thus the output of the L stochastic systems at time k is the observation **matrix**

$$\mathbf{y}(k) = [y_1(k), \ldots, y_L(k)]' \in \mathbb{R}^{L \times D}$$

where a' denotes transpose of matrix a.

The analyst observes at each time k the anonymized (unlabeled) observation **set**

$$y(k) = \sigma_k(\mathbf{y}(k)) = \{y_1(k), \ldots, y_L(k)\} \tag{2}$$

The anonymization map σ_k is a permutation over the set $\{1, 2, \ldots, L\}$. By anonymization we mean that the row indices l of the matrix $\mathbf{y}(k)$ are hidden. Thus $y(k)$ is an unordered set of L row vectors. The time dependence of σ_k emphasizes that the permutation map operating on $\mathbf{y}(k)$ changes at each time k.

Aim. The analyst only sees the anonymized observation set $y(k)$ at each time k. Given the time sequence of observation sets $(y(k), k = 1, 2, \ldots)$, the aim of the analyst is to estimate the underlying set of true parameters $\theta^o = \{\theta_1^o, \ldots, \theta_L^o\}$ of the L stochastic systems. Note that the analyst aims to estimate the *set* θ^o; due to the anonymization (unknown permutation map), in general, it is impossible to assign which parameter belongs to which individual system.

Remarks: (i) Another way of viewing the estimation objective is: Given noisy measurements of unknown permutations of the rows a matrix, how to estimate the elements of the matrix? Our main result below is to propose a symmetric transform framework that circumvents modeling the permutations σ_k and is agnostic to the probabilistic structure of σ_k. In our extended paper [1], we assign a Markovian process to the permutations σ_k yielding a permutation mixture model.

(ii) Since the ordering of the elements of the set $y(k)$ is arbitrary, we cannot use the classical LMS algorithm. If we naively choose a random permutation of the set $y(k)$ as the observation vector, and feed this L-dimensional observation vector into L LMS algorithms, then the estimates will not in general converge to $\theta^o_l, l = 1, \dots, L$.

Applications. We classify applications of the observation model (1), (2) into two types: (i) Due to sensing limitations, the sensor provides noisy measurements from multiple processes, and there is uncertainty as to which measurement came from which process and (ii) examples where the identities of the processes generating the measurements are purposefully hidden to preserve anonymity. Anonymization of trajectories arises in several applications including health care where wearable monitors generate time series of data uniquely matched to an individual, and connected vehicles, where location traces are recorded over time.

The concept of k-anonymity[1] (we will call this L-anonymity since we use k for time) was proposed by [2]. It guarantees that there are at least L identical records in a data set that are indistinguishable. In our formulation, due to the anonymization step (2), the identities (indexes) l of the L processes are indistinguishable. More generally, in the model (1), (2), the identity l of each target itself can be a categorical vector $[l_1, \dots, l_N]$. For example if each process models GPS data trajectories of individuals [3], the categorical data $\psi_l(k)$ records discrete-valued variables such as individuals identity, specific locations visited, etc. To ensure L-anonymity, these categorical vectors are all allocated a single vector, thereby maintaining anonymity of the categorical data. Thus the analyst only sees the anonymized observation set $y(k)$.

In our formulation, the input signal matrices $\psi(k)$ are the same for all L processes; preserving L-anonymity. If the analyst could specify a different input signal ψ_l to each system l, then the analyst can straightforwardly estimate θ^o_l for each target process l, thereby breaking anonymity.

2. MAIN IDEA. SYMMETRIC TRANSFORMS & ADAPTIVE FILTERING

A remarkable approach developed in the 1990s by Kamen and coworkers [4, 5, 6] in the context of Bayesian estimation, involves using symmetric transforms. In this paper we extend this idea of symmetric transforms to stochastic optimization. Specifically, we show that the symmetric transform approach preserves convexity. Since [4] deals with Bayesian filtering for estimating the state, convexity is irrelevant. In comparison, preservation of convexity is crucial in stochastic optimization problems to ensure that the estimates of a stochastic gradient algorithm converge to the global minimum.

To explain our main ideas, suppose there are $L = 3$ scalar-valued random processes, so each observation $y_l(k)$ is scalar-valued. For simplicity assume the input signal $\psi(k) = 1$; so the observations are $y_l(k) = \theta^o_l + v_l(k)$. Given the anonymized observation set $y(k) = \{y_1(k), \dots, y_3(k)\}$ at each time k, how to estimate the parameters $\theta^o_1, \theta^o_2, \theta^o_3$? Our main idea is to use the *set* $y(k)$ to construct a pseudo-measurement *vector* $z(k) \in \mathbb{R}^3$. Suppressing the time dependency (k), we construct pseudo-measurements z_1, z_2, z_3 via a symmetric transform as follows:

$$\begin{aligned} z_1 &= S_1\{y_1, y_2, y_3\} = y_1 + y_2 + y_3 \\ z_2 &= S_2\{y_1, y_2, y_3\} = y_1\, y_2 + y_1\, y_3 + y_2\, y_3 \\ z_3 &= S_3\{y_1, y_2, y_3\} = y_1\, y_2\, y_3 \end{aligned} \tag{3}$$

The key point is that the pseudo-observations z_l are symmetric in y_1, y_2, y_3. Any permutation of the elements of $\{y_1, \dots, y_3\}$ does not affect z_l. In this way, there is no need to assign (classify) an observation to a specific process. But we have introduced a new problem: estimating θ^o using the pseudo-observations is no longer a convex stochastic optimization problem. To estimate θ^o we minimize the second order moments to compute:

$$\begin{aligned} \theta^* = \arg\min_\theta \{ & \mathbb{E}\{(z_1 - (\theta_1 + \theta_2 + \theta_3))^2\} \\ & + \mathbb{E}\{(z_2 - (\theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3))^2\} + \mathbb{E}\{(z_3 - \theta_1\theta_2\theta_3)^2\}\} \end{aligned} \tag{4}$$

Clearly the multi-linear objective (4) is non-convex in $\theta_1, \theta_2, \theta_3$. However, the problem is convex in the symmetric transformed variables (denoted as λ below), and the original variables θ can be evaluated by inverting the symmetric transform. We formalize this as follows:

Result 1. *(Informal) The global minimum θ^* of the non-convex objective* (4) *can be computed in three steps:*
(i) Given the observations $y(k)$, compute the pseudo-observations $z(k)$ using (3).
(ii) Using these pseudo-observations, estimate the pseudo parameters $\lambda_1 = \theta_1 + \theta_2 + \theta_3$, $\lambda_2 = \theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3$, $\lambda_3 = \theta_1\theta_2\theta_3$. Clearly (4) *is a stochastic convex optimization problem in pseudo-parameters $\lambda_1, \lambda_2, \lambda_3$. Let $\lambda^*_1, \lambda^*_2, \lambda^*_3$ denote the estimates.*
(iii) Finally, solve the polynomial equation $s^3 + \lambda^*_1 s^2 + \lambda^*_2 s + \lambda^*_3 = 0$. Then the roots[2] are θ^*. Computing the roots of a polynomial is equivalent to computing the eigenvalues of the companion matrix.

Put simply, Result 1 says that while (4) is non-convex in the roots of a polynomial, it is convex in the coefficients of the polynomial! To explain Step (ii), clearly (4) is convex in the pseudo-parameters $\lambda_1, \lambda_2, \lambda_3$. We can straightforwardly compute the global minimum in terms of these pseudo parameters as $\lambda^*_1 = \mathbb{E}\{z_1\}$, $\lambda^*_2 = \mathbb{E}\{z_2\}$, $\lambda^*_3 = \mathbb{E}\{z_3\}$.

[1] Data anonymity is mainly studied under two categories: k-anonymity and differential privacy. Differential privacy methods add noise to trajectory data providing a provable privacy guarantee for the data set. Even though our model has additive noise v and this can be motivated in terms of differential privacy; we will not discuss differential privacy in this paper.

[2] Strictly speaking $\theta_1, \theta_2, \theta_3$ are factors. The root is the negative of factor.

To explain Step (iii) of the above result, we use a crucial property of symmetric functions. The reader van verify that the following monic polynomial in variable s satisfies

$$(s+\theta_1)(s+\theta_2)(s+\theta_3) = s^3 + \lambda_1 s^2 + \lambda_2 s + \lambda_3$$

The above equation states that a monic polynomial with pseudo-parameters $\lambda_1, \lambda_2, \lambda_3$ as coefficients has the parameters $\theta_1, \theta_2, \theta_3$ as roots of the polynomial. By the fundamental theorem of algebra, there is a unique invertible map between the coefficients of a monic polynomial and the set of roots of the polynomial. As a result, we can first compute the global minimum λ^* of the above objective (4) (since it is convex in λ), and then compute the unique parameter set θ^*, which is the set of roots of the corresponding polynomial. Thus we have computed the global minimum θ^* of the non-convex objective (4). Thus, Result 1 estimates the true parameter set θ^o given anonymized observations (in a simplified setting).

3. ADAPTIVE FILTERING WITH SCALAR ANONYMIZED OBSERVATIONS

Due to page restrictions, we discuss the problem of estimating the true parameter θ^o when the observation $y_l(k)$ of each process l is a scalar; so $D = 1$ in (1) and $\psi(k)$ is a scalar. The vector observation case is discussed in [1]; see Conclusions. Since there are L independent scalar processes in (1), the parameters generating these L processes is $\theta^o = \{\theta^o_1, \ldots, \theta^o_L\}$.

Given the anonymized observation set $y(k) = \{y_1(k), \ldots, y_L(k)\}$ at each time k, our main idea is to construct a pseudo-measurement vector $z(k) \in \mathbb{R}^L$. Suppressing the time dependency (k), we construct the L pseudo-measurements $z_l, l \in [L]$ via a symmetric transform[3] [7] as

$$\begin{aligned} z = S\{y\} \iff z_l &= S_l\{y_1, \ldots, y_L\} \\ &\overset{\text{defn}}{=} \sum_{i_1 < i_2 < \cdots < i_l} y_{i_1}\, y_{i_2} \cdots y_{i_l},\ l \in [L] \end{aligned} \tag{5}$$

where $[L] = \{1, \ldots, L\}$. Using the classical Vieta's formulas [8], that the pseudo-measurements $z_l, l \in [L]$ in (5) are the coefficients of the following L-order polynomial in variable s:

$$S\{y\}(s) \overset{\text{defn}}{=} \prod_{l=1}^{L}(s + y_l) = s^L + \sum_{l=1}^{L} z_l\, s^{L-l} \tag{6}$$

As an example, consider $L = 3$ independent scalar processes. Then the pseudo-observations using (5) are given by (3). Indeed, the pseudo-observations z_1, z_2, z_3 are the coefficients of the polynomial $(s + y_1)(s + y_2)(s + y_3)$.

Note that each z_l is permutation invariant: any permutation of the elements of $\{y_1, \ldots, y_L\}$ does not affect z_l. That is why our notation above involves the set $\{y_1, y_2, \ldots, y_L\}$.

[3] By symmetric transform S_l, we mean $S_l\{y_1, \ldots, y_L\} = S_l\{P \cdot \{y_1, \ldots, y_L\}\}$ for any permutation P of $\{y_1, \ldots, y_L\}$. Thus while the elements $\{y_1, \ldots, y_L\}$ are arbitrarily ordered, the value of $S_l\{\cdot\}$ is unique.

Symmetric Transform and Estimation Objective. Given the set valued sequence of anonymized observations, $y(1), y(2), \ldots y(k), \ldots$ generated by (1), our aim is to estimate the true parameter set $\theta^o = \{\theta^o_1, \ldots, \theta^o_L\}$. To do so, we first construct the pseudo measurement vectors $z(1), z(2), \ldots, z(k)$ via (5). Denoting $\theta = \{\theta_1, \ldots, \theta_L\}$, our objective is to estimate the set $\theta^* = \{\theta^*_1, \ldots, \theta^*_L\}$:

$$\begin{aligned} \theta^* = \arg\min_{\theta} \sum_{l \in [L]} \mathbb{E}\,|z_l - S_l\{\psi\,\theta_1, \psi\,\theta_2, \ldots, \psi\,\theta_L\}|^2 \\ \text{where } z_l = S_l\{\psi\,\theta^o_1 + v_1, \ldots, \psi\,\theta^o_L + v_L\} \end{aligned} \tag{7}$$

Recall the symmetric transform S_l is defined in (5). Finally, define the symmetric transforms on the model parameters as

$$\lambda = S\{\theta\} \iff \lambda_l = S_l\{\theta_1, \ldots, \theta_L\},\ l \in [L]. \tag{8}$$

Note that $\lambda = [\lambda_1, \ldots, \lambda_L]'$ is an L-dimension vector whereas θ is a set with L (unordered) elements.

From (7), we see that θ^* is a second order method of moments estimate of θ^o wrt pseudo observations. Importantly, this estimate is independent of the anonymization map σ.

3.1. Main Result. Consistent Estimator for θ^o.

We are now ready to state our main result, namely an adaptive filtering algorithm to estimate θ^o given anonymized scalar observations. The result says that while objective (7) is non-convex in θ, we can reformulate it as a convex optimization problem in terms of λ defined in (8). The intuition is that the objective (7) is non-convex in the roots of the polynomial (namely, θ) , but is convex in the *coefficients* of the polynomial (namely, λ); and by the fundamental theorem of algebra there is a one-to-one map from the coefficients λ to the roots θ. Therefore, by mapping observations to pseudo observations, we can construct the optimal estimate of (7).

Theorem 1. *Consider the sequence of anonymized observation sets $(y(k), k \geq 1)$ generated by (1) and (2), where $\psi(k)$ is a known iid scalar sequence. Then*

1. *The objective (7) can be expressed as L decoupled convex optimization problems in terms of λ defined in (8):*

$$\min_{\lambda_l} \mathbb{E}|z_l - \psi^l \lambda_l|^2,\ z_l(k) = \big(\psi(k)\big)^l\, \lambda^o_l + w_l(k) \tag{9}$$

   The zero mean process $w(k)$ is defined in [1].
2. *The global minimizer θ^* of objective (7) is consistent in the sense that $\theta^* = \theta^o$.*
3. *With pseudo observations $z(k) = S\{y(k)\}$ (5), consider L decoupled adaptive filtering algorithms operating on $z(k)$: Choose $\lambda(0) \in \mathbb{R}^L$. Then for $l \in [L]$,*

$$\begin{aligned} \lambda_l(k+1) &= \lambda_l(k) + \epsilon\, \psi^l(k)\, \big(z_l(k) - \psi^l(k)\, \lambda_l(k)\big) \\ \theta(k+1) &= \mathfrak{Re}\big(S^{-1}(\lambda(k+1))\big) \end{aligned} \tag{10}$$

   Here S^{-1} is defined in (11) and $\mathfrak{Re}$ denotes the real part of the complex vector. The estimates $\theta(k)$ converge in probability and mean square to θ^*.

Discussion: 1. Theorem 1 gives a tractable and consistent method for estimating the parameter set θ^o of the L stochastic systems given set valued anonymized observations $y(1), y(2), \ldots$. We emphasize that since the observations $y(k)$ are set-valued, the ordering of the elements of θ^o cannot be recovered; Statement 1 of the theorem asserts that the set-valued estimate θ^* converges to θ^o. Statement 2 of the theorem gives an adaptive filtering algorithm (10) that operates on the pseudo observation vector $z(k)$. Applying the transform S^{-1} to the estimates $\lambda(k)$ generated by (10) yields estimates $\theta(k)$ that converge to the global minimum θ^*. Since $\theta^o \in \mathbb{R}^L$, the second step of (10) chooses the real part of possibly complex valued roots.

2. The symmetric operator S is uniquely invertible since an L-th degree polynomial has a unique set of at most L roots. Given $\lambda = S\{\theta\}$, $\theta = S^{-1}(\lambda)$ are the unique set of roots $\{\theta_1, \ldots, \theta_L\}$ of the polynomial with coefficients $\lambda_l, l \in [L]$:

$$\theta = S^{-1}(\lambda) \iff s^L + \sum_{l=1}^{L} \lambda_l \, s^{l-1} = \prod_{l=1}^{L} (s + \theta_l) \quad (11)$$

Note that $S^{-1}(\cdot)$ maps the vector λ to unique set θ.

3. The adaptive filtering algorithm (10) uses a constant step size; hence it converges weakly (in distribution) to the true parameter θ^o [9, 10]. Since we assumed θ^o is a constant, weak convergence is equivalent to convergence in probability.

4. A stochastic gradient algorithm operating on (7) is

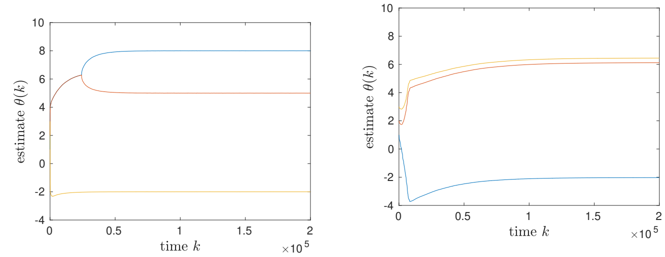
$$\theta(k+1) = \theta(k) - \epsilon \, \nabla_\theta \sum_{l \in [L]} |z_l(k) - S_l\{\psi(k)\, \theta_1(k), , \ldots, \psi(k)\, \theta_L(k)\}|^2 \quad (12)$$

However, objective (7) has local minima and stochastic gradient algorithm (12) can get stuck at these local minima (see numerical example). In comparison, the formulation involving pseudo-measurements yields a convex (quadratic) objective and algorithm (10) provably converges to the global minimum. There is also another problem with (12). If initial condition $\theta(0)$ is chosen with equal elements, then since gradient ∇_θ is symmetric (wrt y and θ), all the elements of the estimate $\theta(k)$ have equal elements at each time k, regardless of the choice of θ^o, so algorithm (12) will never converge to θ^o.

4. NUMERICAL EXAMPLE

We show that objective (7) has local minima wrt θ; so the classical stochastic gradient algorithm (12) gets stuck in a local minimum. In comparison, the objective (9) using pseudo-measurements is convex (quadratic) wrt λ and so the adaptive filtering algorithm (10) converges to global minimum θ^*.

We consider $L = 3$ independent scalar processes ($D = 1$) with anonymized observations generated as in (2). The true model that generates the observations is $\theta^o = [-2,\ 5,\ 8]'$. The regression signal $\psi(k) \sim \mathbf{N}(0, \sigma^2)$ where $\sigma = 1$. The noise error $v(k) \sim \mathbf{N}(0, \sigma_v^2)$ where $\sigma_v = 10^{-2}$.



(a) Algorithm (10) converges to global optimum θ^o.

(b) Classical stochastic gradient algorithm (12) gets stuck in local minimum.

Fig. 2: Fig.2a shows that the parameter estimates generated by Algorithm (10) converge to θ^o. Fig.2b shows that the parameter estimates generated by stochastic gradient algorithm (12) operating on (7) do not converge to θ^o.

We ran the adaptive filtering algorithm (10) on a sample path of 2×10^5 anonymized observations generated by the above model with step size $\epsilon = 10^{-4}$. For initial condition $\theta(0) = [1, 2, 3]'$, Figure 2a shows that the estimates generated by Algorithm (10) converges to θ^o.

We also ran the classical stochastic gradient algorithm (12) on the anonymized observations. Recall this algorithm minimizes (7) directly. The step size chosen was $\epsilon = 10^{-7}$ (larger step sizes led to instability). For initial condition $\theta(0) = [1, 2, 3]'$, Figure 2(b) shows that the estimates converge to a local stationary point $[-2.02, 6.12, 6.45]'$ which is not θ^o. For initial condition $\theta(0) = [3, 6, 9]'$, we found that the estimates converged to θ^o. This provides numerical verification that objective (7) is non-convex. Besides the non-convex objective, another problem with the algorithm (12) is that if we choose $\theta(0) = [c, c, c]$ for any $c \in \mathbb{R}$, then all elements of $\theta(k)$ are identical, regardless of θ^o.

5. CONCLUSIONS

We proposed a symmetric transform based adaptive filtering algorithm for parameter estimation when the observations are a set (unordered/anonymized) rather than a vector. Such observation sets arise due to uncertainty in sensing or deliberate anonymization of data. By exploiting the uniqueness of factorization, Theorem 1 showed that the adaptive filtering algorithms converge to the true parameter (global minimum).

Extension. In [1], we extend the paper to vector observation sets by developing vector symmetric transforms using a two-variable polynomial transform. In general the fundamental theorem of algebra, does not extend to polynomials in two variables. By exploiting that fact that the algebraic ring of multi-variable polynomials is a unique factorization domain over the ring of one-variable polynomials, we construct an adaptive filtering algorithm that yields consistent estimates of the underlying parameters thereby extending Theorem 1 to vector observations. [1] also has several numerical examples.

6. REFERENCES

[1] V. Krishnamurthy, "Adaptive filtering algorithms for set-valued observations – symmetric measurement approach to unlabeled and anonymized data," 2022, under review, IEEE Trans Signal Processing. [Online]. Available: https://arxiv.org/abs/2208.14576

[2] L. Sweeney, "k-anonymity: A model for protecting privacy," *International journal of uncertainty, fuzziness and knowledge-based systems*, vol. 10, no. 05, pp. 557–570, 2002.

[3] T. Hara, A. Suzuki, M. Iwata, Y. Arase, and X. Xie, "Dummy-based user location anonymization under real-world constraints," *IEEE Access*, vol. 4, pp. 673–687, 2016.

[4] E. W. Kamen, "Multiple target tracking based on symmetric measurement equations," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 371–374, 1992.

[5] E. W. Kamen and C. R. Sastry, "Multiple target tracking using products of position measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 2, pp. 476–493, 1993.

[6] U. D. Hanebeck, M. Baum, and P. Willett, "Symmetrizing measurement equations for association-free multi-target tracking via point set distances," in *Signal Processing, Sensor/Information Fusion, and Target Recognition XXVI*, vol. 10200. SPIE, 2017, pp. 48–61.

[7] I. G. Macdonald, *Symmetric functions and Hall polynomials*. Oxford university press, 1998.

[8] N. Jacobson, *Basic algebra I*. Courier Corporation, 2012.

[9] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Recursive Algorithms and Applications*, 2nd ed. Springer-Verlag, 2003.

[10] A. Sayed, *Adaptive Filters*. Wiley, 2008.