TDOA-based Localization via Stochastic Gradient Descent Variants

Luis F. Abanto-Leon

Co-authors: Arie Koppelaar
Sonia Heemstra de Groot

Department of Electrical Engineering
Eindhoven University of Technology

IEEE 88th Vehicular Technology Conference (VTC 2018-Fall)

Contents

Background

- Source localization is of pivotal importance in several areas such as WSN and Internet of Things (IoT).

Background

- Source localization is of pivotal importance in several areas such as WSN and Internet of Things (IoT).
- Location information can be used for a variety of purposes, e.g. surveillance, monitoring, tracking, etc.

Background

- Source localization is of pivotal importance in several areas such as WSN and Internet of Things (IoT).
- Location information can be used for a variety of purposes, e.g. surveillance, monitoring, tracking, etc.
- TDOA is one of the well-known localization approaches, where a source broadcasts a signal and a number of receivers record the arriving time of the transmitted signal.

Background

- Source localization is of pivotal importance in several areas such as WSN and Internet of Things (IoT).
- Location information can be used for a variety of purposes, e.g. surveillance, monitoring, tracking, etc.
- TDOA is one of the well-known localization approaches, where a source broadcasts a signal and a number of receivers record the arriving time of the transmitted signal.
- By means of computing the time difference from various receivers, the source location can be estimated.

Background

- On the other hand, in the recent few years novel optimization algorithms have emerged for (i) processing big data and for (ii) training deep neural networks.

Background

- On the other hand, in the recent few years novel optimization algorithms have emerged for (i) processing big data and for (ii) training deep neural networks.
- Most of these techniques are enhanced variants of the classical stochastic gradient descent (SGD) but with additional features that promote faster convergence.

Background

- On the other hand, in the recent few years novel optimization algorithms have emerged for (i) processing big data and for (ii) training deep neural networks.
- Most of these techniques are enhanced variants of the classical stochastic gradient descent (SGD) but with additional features that promote faster convergence.
- We propose an optimization procedure called RMSProp+AF, which is based on RMSProp algorithm but incorporating adaptation of the decaying factor.

Background

- On the other hand, in the recent few years novel optimization algorithms have emerged for (i) processing big data and for (ii) training deep neural networks.
- Most of these techniques are enhanced variants of the classical stochastic gradient descent (SGD) but with additional features that promote faster convergence.
- We propose an optimization procedure called RMSProp+AF, which is based on RMSProp algorithm but incorporating adaptation of the decaying factor.
- We show through simulations that all of these techniques can also be successfully applied to source localization.

TDOA Model

- Consider a system consisting of a set of receivers $\mathcal{R} = \{r_1, r_2, \ldots, r_N\}$

TDOA Model

- Consider a system consisting of a set of receivers $\mathcal{R} = \{r_1, r_2, \ldots, r_N\}$
- The receivers are located at known positions $\tilde{\mathbf{p}}_i = [\tilde{x}_i, \tilde{y}_i]^T$, $i = 1, 2, \cdots, N$.

TDOA Model

- Consider a system consisting of a set of receivers $\mathcal{R} = \{r_1, r_2, \ldots, r_N\}$
- The receivers are located at known positions $\tilde{\mathbf{p}}_i = [\tilde{x}_i, \tilde{y}_i]^T$, $i = 1, 2, \cdots, N$.
- There is a single transmitter at the unknown location $\mathbf{p}$, which is actively broadcasting beacon signals $s(t)$ that are not necessarily known by the receivers.

TDOA Model

- Let $z_i(t) = h_i \cdot s(t - \tau_i) + \eta_i(t)$ denote the received signal at receiver $r_i \in \mathcal{R}$.

TDOA Model

- Let $z_i(t) = h_i \cdot s(t - \tau_i) + \eta_i(t)$ denote the received signal at receiver $r_i \in \mathcal{R}$.
- τ_i represents the time of arrival at the receiver r_i.

TDOA Model

- Let $z_i(t) = h_i \cdot s(t - \tau_i) + \eta_i(t)$ denote the received signal at receiver $r_i \in \mathcal{R}$.
- τ_i represents the time of arrival at the receiver r_i.
- The channel gain at receiver r_i is denoted by h_i whereas η_i represents Gaussian noise.

TDOA Model

- Let $z_i(t) = h_i \cdot s(t - \tau_i) + \eta_i(t)$ denote the received signal at receiver $r_i \in \mathcal{R}$.
- τ_i represents the time of arrival at the receiver r_i.
- The channel gain at receiver r_i is denoted by h_i whereas η_i represents Gaussian noise.
- When $s(t)$ is unknown by the receivers, the incognito signal $s(t)$ can be removed by means of correlation analysis.

TDOA Model

- Let $z_i(t) = h_i \cdot s(t - \tau_i) + \eta_i(t)$ denote the received signal at receiver $r_i \in \mathcal{R}$.
- τ_i represents the time of arrival at the receiver r_i.
- The channel gain at receiver r_i is denoted by h_i whereas η_i represents Gaussian noise.
- When $s(t)$ is unknown by the receivers, the incognito signal $s(t)$ can be removed by means of correlation analysis.
- Thus, the TDOA measurements $\Delta\tau_{ij}$ are indirectly estimated by computing the normalized cross-correlation (NCC) between every pair of signals.

Normalized Cross-Correlation (NCC)

$$
\begin{aligned}
\Delta\hat{\tau}_{ij} &= \arg\max_{\Delta\tau_{ij}} \frac{\sum_u \bar{z}_i(u)\bar{z}_j(u-\Delta\tau_{ij})}{\sqrt{\sum_u \bar{z}_i^2(u)}\sqrt{\sum_u \bar{z}_j^2(u)}} \\
&= \arg\max_{\Delta\tau_{ij}} \frac{\sum_u \left(s(u-\tau_i)+\frac{h_i^*}{|h_i|^2}\eta_i(t)\right)\left(s(u-\Delta\tau_{ij}-\tau_j)+\frac{h_j^*}{|h_j|^2}\eta_j(t)\right)}{\sqrt{\sum_u \left(s(u-\tau_i)+\frac{h_i^*}{|h_i|^2}\eta_i(t)\right)^2}\sqrt{\sum_u \left(s(u-\tau_j)+\frac{h_j^*}{|h_j|^2}\eta_j(t)\right)^2}} \\
&= (\tau_i-\tau_j)+\pi_{ij}
\end{aligned} \tag{1}
$$

Problem Formulation

Because the underlying location estimation problem requires using distances, all TDOAs $\Delta\hat{\tau}_{ij}$ will be converted from time to range differences as shown in (2).

$$\begin{aligned} \Delta\hat{d}_{ij} &= c \cdot \Delta\hat{\tau}_{ij} \\ &= c \cdot (\tau_i - \tau_j) + c \cdot \pi_{ij} \\ &= d_i - d_j + \epsilon_{ij} \\ &= \|\mathbf{p} - \tilde{\mathbf{p}}_i\|_2 - \|\mathbf{p} - \tilde{\mathbf{p}}_j\|_2 + \epsilon_{ij} \\ &= g(\mathbf{p}, \tilde{\mathbf{p}}_i, \tilde{\mathbf{p}}_j) + \epsilon_{ij} \end{aligned} \qquad (2)$$

Equivalently,

$$\Delta\hat{d}_m = g(\mathbf{p}, \tilde{\mathbf{p}}_i, \tilde{\mathbf{p}}_j) + \epsilon_m \qquad (3)$$

Problem Formulation

Given the observed measurements $\mathbf{\Delta \hat{d}} = [\Delta\hat{d}_{1,2}, \Delta\hat{d}_{1,3}, \cdots, \Delta\hat{d}_{N-1,N}]^T$, the objective is to estimate—with the least uncertainty—the true position $\mathbf{p}$ of the transmitter. This can be formulated as maximizing the likelihood function

$$p(\mathbf{\Delta \hat{d}} \mid \mathbf{p}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \exp\left(-\frac{1}{2}(\mathbf{\Delta \hat{d}} - \mathbf{g})^T \mathbf{C}^{-1} (\mathbf{\Delta \hat{d}} - \mathbf{g}) \right) \tag{4}$$

where

$$\mathbf{g} = \begin{bmatrix} \|\mathbf{p} - \tilde{\mathbf{p}}_1\|_2 - \|\mathbf{p} - \tilde{\mathbf{p}}_2\|_2 \\ \|\mathbf{p} - \tilde{\mathbf{p}}_1\|_2 - \|\mathbf{p} - \tilde{\mathbf{p}}_3\|_2 \\ \vdots \\ \|\mathbf{p} - \tilde{\mathbf{p}}_{N-1}\|_2 - \|\mathbf{p} - \tilde{\mathbf{p}}_N\|_2 \end{bmatrix}. \tag{5}$$

Problem Formulation

Maximizing (4) is equivalent to minimizing (5)

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \underbrace{(\boldsymbol{\Delta}\hat{\mathbf{d}} - \mathbf{g})^T \mathbf{C}^{-1} (\boldsymbol{\Delta}\hat{\mathbf{d}} - \mathbf{g})}_{J:\text{ cost function}} . \quad (6)$$

We determine $\mathbf{p}$ iteratively using a gradient approach

$$\hat{\mathbf{p}}^{(k+1)} = \hat{\mathbf{p}}^{(k)} - \mu \nabla_{\mathbf{p}}^{(k)} J, \quad (7)$$

$$\nabla_{\mathbf{p}}^{(k)} J = -2\boldsymbol{\epsilon}^{(k)} \begin{bmatrix} \frac{\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_1}{\|\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_1\|_2} - \frac{\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_2}{\|\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_2\|_2} \\ \frac{\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_1}{\|\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_1\|_2} - \frac{\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_3}{\|\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_3\|_2} \\ \vdots \\ \frac{\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_{N-1}}{\|\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_{N-1}\|_2} - \frac{\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_N}{\|\mathbf{p}^{(k)} - \tilde{\mathbf{p}}_N\|_2} \end{bmatrix} \quad (8)$$

Proposed Algorithm: RMSProp + AF

Algorithm 1: Proposed RMSProp with Adaptive Decaying Factor (RMSProp+AF)

Input: The gradient $\nabla_{\mathbf{p}}^{(k)} J$ of cost function J

Output: The estimated position $\hat{\mathbf{p}}$

begin

Step 1: Initialize the FIFO buffers b_x and b_y of size L.

Step 2: Initialize $\boldsymbol{\rho}^{(0)} = [0.99\ 0.99]^T$.

Step 3: Initialize $\mathbf{r}^{(0)} = [0\ 0]^T$.

Step 4: Define the vectors $\mathbf{u}_x = [1\ 0]^T$ and $\mathbf{u}_y = [0\ 1]^T$.

for $k = 1 : K$ **do**

Step 5a: Compute the circular buffer index k'

$$k' = k - L \left\lfloor \frac{k-1}{L} \right\rfloor$$

Step 5b: Store the square of gradient $\nabla_{\mathbf{p}}^{(k)} J$ at index k' on each of the buffers

$$b_x(k') = \mathbf{u}_x^T \left(\nabla_{\mathbf{p}}^{(k)} J \odot \nabla_{\mathbf{p}}^{(k)} J \right)$$
$$b_y(k') = \mathbf{u}_y^T \left(\nabla_{\mathbf{p}}^{(k)} J \odot \nabla_{\mathbf{p}}^{(k)} J \right)$$

Step 5c: Define the vectors $\mathbf{v}_{max}$ and $\mathbf{v}_{min}$

$$\mathbf{v}_{max} = \begin{bmatrix} \max\{b_x\} \\ \max\{b_y\} \end{bmatrix}, \mathbf{v}_{min} \begin{bmatrix} \min\{b_x\} \\ \min\{b_y\} \end{bmatrix}$$

Step 5d: Compute the adaptive decaying factor $\boldsymbol{\rho}^{(k)}$

$$\boldsymbol{\gamma}^{(k)} = (\mathbf{v}_{max} - \mathbf{v}_{min}) \oslash (\mathbf{v}_{max} + \mathbf{v}_{min} + \mathbf{1}_{2\times 1})$$
$$\boldsymbol{\rho}^{(k)} = \begin{bmatrix} \max\left\{\mathbf{u}_x^T \boldsymbol{\rho}^{(0)}, \mathbf{u}_x^T \boldsymbol{\gamma}^{(k)}\right\} \\ \max\left\{\mathbf{u}_y^T \boldsymbol{\rho}^{(0)}, \mathbf{u}_y^T \boldsymbol{\gamma}^{(k)}\right\} \end{bmatrix}$$

Step 5e: Accumulate the squared gradient

$$\mathbf{r}^{(k)} = \boldsymbol{\rho}^{(k)} \odot \mathbf{r}^{(k-1)} + (\mathbf{1}_{2\times 1} - \boldsymbol{\rho}^{(k)}) \odot \nabla_p^{(k)} J \odot \nabla_p^{(k)} J$$

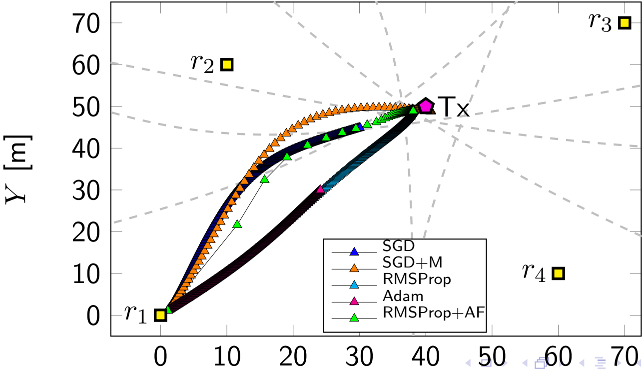
Step 5f: Update the position

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} - \frac{\mu}{\delta + \sqrt{\mathbf{r}^{(k)}}} \odot \nabla_{\mathbf{p}} J$$

Step 6: Output $\hat{\mathbf{p}} = \mathbf{p}^{(K+1)}$

Simulation Results: Case I

Scenario 1: *Consider that* $N = 4$ *receivers are located at positions* $\tilde{\mathbf{p}}_1 = [0\ 0]^T$, $\tilde{\mathbf{p}}_2 = [10\ 60]^T$, $\tilde{\mathbf{p}}_3 = [70\ 70]^T$ *and* $\tilde{\mathbf{p}}_4 = [60\ 10]^T$. *In addition, the unknown position of the transmitter is* $\mathbf{p} = [40\ 80]^T$.



Simulation Results: Case I

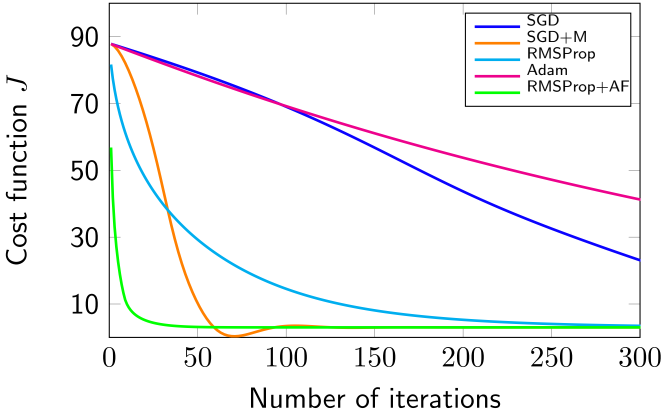
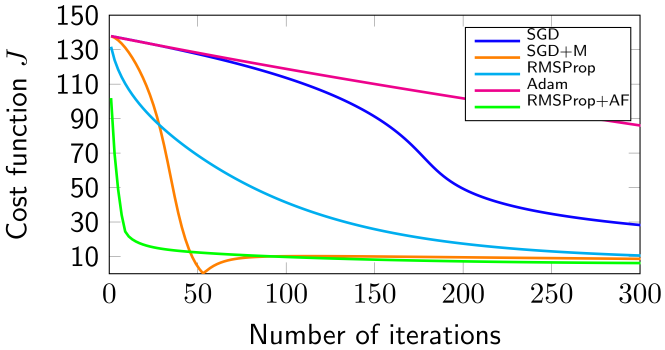


Figure 3:Scenario 1 - Convergence of algorithms

Simulation Results: Case II

Scenario 2: *Consider that $N = 4$ receivers are located at positions $\tilde{\mathbf{p}}_1 = [0\ 0]^T$, $\tilde{\mathbf{p}}_2 = [10\ 60]^T$, $\tilde{\mathbf{p}}_3 = [70\ 70]^T$ and $\tilde{\mathbf{p}}_4 = [60\ 10]^T$. In addition, the unknown position of the transmitter is $\mathbf{p} = [75\ 65]^T$.*



Conclusions

- In this work we have presented a comparison of different optimization techniques—commonly used in the machine learning realm—to solve TDOA-based localization.
- We conclude that most of the approaches can be successfully applied and can outperform classical methods such as stochastic gradient descent.
- In addition, we presented an improved version named RMSProp+AF, which is capable of providing enhanced convergence in comparison to state–of–the–art approaches.
- We showed that the proposed scheme outperforms other competing approaches (i) when the transmitter is inside and (ii) when it is outside the convex hull.

Questions

Email: l.f.abanto@ieee.org