Distributed SPS Algorithms for Non-Asymptotic Confidence Region Evaluation

Vincenzo Zambianchi [1]

[1] CNIT-Italian National Consortium in Telecommunications
University of Bologna, Italy

July 2014, Brixen

Introduction and Motivation

- *Context:* Parameter estimation distributed over a wireless sensor network.
- Normally, accent on estimation.
- Confidence regions assess quality for estimates.
- Classical methods (Cramer-Rao bound) in the asymptotic domain.
- Non asymptotic methods are of interest $\rightarrow$ SPS algorithm

Introduction and Motivation

- *Context:* Parameter estimation distributed over a wireless sensor network.
- Normally, accent on estimation.
- Confidence regions assess quality for estimates.
- Classical methods (Cramer-Rao bound) in the asymptotic domain.
- Non asymptotic methods are of interest $\rightarrow$ SPS algorithm

Introduction and Motivation

- *Context:* Parameter estimation distributed over a wireless sensor network.
- Normally, accent on estimation.
- Confidence regions assess quality for estimates.
- Classical methods (Cramer-Rao bound) in the asymptotic domain.
- Non asymptotic methods are of interest $\rightarrow$ SPS algorithm

Introduction and Motivation

- *Context:* Parameter estimation distributed over a wireless sensor network.
- Normally, accent on estimation.
- Confidence regions assess quality for estimates.
- Classical methods (Cramer-Rao bound) in the asymptotic domain.
- Non asymptotic methods are of interest $\rightarrow$ SPS algorithm

Introduction and Motivation

- *Context:* Parameter estimation distributed over a wireless sensor network.
- Normally, accent on estimation.
- Confidence regions assess quality for estimates.
- Classical methods (Cramer-Rao bound) in the asymptotic domain.
- Non asymptotic methods are of interest $\rightarrow$ SPS algorithm

The centralized sign perturbed sums (SPS) algorithm

Method developed in Csàji et al. 2012 [1].

Network of N nodes taking linear observations $y_i = \psi_i^T \mathbf{p}^* + w_i$.

Measurement noise is assumed **independent** and **symmetrically distributed** with respect to zero. Compute

$$z_0(\mathbf{p}) = \left\| \sum_{i=1}^{N} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 \qquad (1)$$

and the $m-1$ norms of the sign perturbed sums

$$z_j(\mathbf{p}) = \left\| \sum_{i=1}^{N} \alpha_{j,i} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 , \qquad (2)$$

where $j = 1, \ldots, m-1$ and $\alpha_{j,i} \in \{\pm 1\}$ are *independent, identically distributed and equiprobable random signs.*

The centralized sign perturbed sums (SPS) algorithm

Method developed in Csàji et al. 2012 [1].
Network of N nodes taking linear observations $y_i = \psi_i^T \mathbf{p}^* + w_i$.
Measurement noise is assumed **independent** and **symmetrically distributed** with respect to zero. Compute

$$z_0(\mathbf{p}) = \left\| \sum_{i=1}^{N} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 \quad (1)$$

and the $m-1$ norms of the sign perturbed sums

$$z_j(\mathbf{p}) = \left\| \sum_{i=1}^{N} \alpha_{j,i} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 , \quad (2)$$

where $j = 1, \ldots, m-1$ and $\alpha_{j,i} \in \{\pm 1\}$ are *independent, identically distributed and equiprobable random signs.*

The centralized sign perturbed sums (SPS) algorithm

Method developed in Csàji et al. 2012 [1].
Network of N nodes taking linear observations $y_i = \psi_i^T \mathbf{p}^* + w_i$.
Measurement noise is assumed **independent** and **symmetrically distributed** with respect to zero. Compute

$$z_0(\mathbf{p}) = \left\| \sum_{i=1}^{N} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 \tag{1}$$

and the $m-1$ norms of the sign perturbed sums

$$z_j(\mathbf{p}) = \left\| \sum_{i=1}^{N} \alpha_{j,i} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 , \tag{2}$$

where $j = 1, \dots, m-1$ and $\alpha_{j,i} \in \{\pm 1\}$ are *independent, identically distributed and equiprobable random signs.*

The centralized sign perturbed sums (SPS) algorithm

Method developed in Csàji et al. 2012 [1].
Network of N nodes taking linear observations $y_i = \psi_i^T \mathbf{p}^* + w_i$.
Measurement noise is assumed **independent** and **symmetrically distributed** with respect to zero. Compute

$$z_0(\mathbf{p}) = \left\| \sum_{i=1}^{N} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 \tag{1}$$

and the $m-1$ norms of the sign perturbed sums

$$z_j(\mathbf{p}) = \left\| \sum_{i=1}^{N} \alpha_{j,i} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 , \tag{2}$$

where $j = 1, \ldots, m-1$ and $\alpha_{j,i} \in \{\pm 1\}$ are *independent, identically distributed and equiprobable random signs.*

The centralized sign perturbed sums (SPS) algorithm

Method developed in Csàji et al. 2012 [1].
Network of N nodes taking linear observations $y_i = \psi_i^T \mathbf{p}^* + w_i$.
Measurement noise is assumed **independent** and **symmetrically distributed** with respect to zero. Compute

$$z_0(\mathbf{p}) = \left\| \sum_{i=1}^{N} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2 \tag{1}$$

and the $m-1$ norms of the sign perturbed sums

$$z_j(\mathbf{p}) = \left\| \sum_{i=1}^{N} \alpha_{j,i} \psi_i \left(y_i - \psi_i^T \mathbf{p} \right) \right\|_2^2, \tag{2}$$

where $j = 1, \ldots, m-1$ and $\alpha_{j,i} \in \{\pm 1\}$ are *independent, identically distributed and equiprobable random signs.*

The centralized SPS algorithm

The confidence region $\boldsymbol{\Sigma}_q$ for the true parameter value $\mathbf{p}^*$ is defined as

$$\boldsymbol{\Sigma}_q = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \sum_{j=1}^{m-1} \tau_j(\mathbf{p}) \geq q \right\}, \tag{3}$$

where $\tau_j(\mathbf{p}) = 1$ if $z_j(\mathbf{p}) - z_0(\mathbf{p}) > 0$ and 0 otherwise. In Csàji et al. 2012 [1] it was proven that one has

$$\mathrm{Prob}(\mathbf{p}^* \in \boldsymbol{\Sigma}_q) = 1 - \frac{q}{m}. \tag{4}$$

The centralized SPS algorithm

The confidence region $\mathbf{\Sigma}_q$ for the true parameter value $\mathbf{p}^*$ is defined as

$$\mathbf{\Sigma}_q = \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \sum_{j=1}^{m-1} \tau_j(\mathbf{p}) \geq q \right\}, \tag{3}$$

where $\tau_j(\mathbf{p}) = 1$ if $z_j(\mathbf{p}) - z_0(\mathbf{p}) > 0$ and 0 otherwise. In Csàji et al. 2012 [1] it was proven that one has

$$\mathrm{Prob}(\mathbf{p}^* \in \mathbf{\Sigma}_q) = 1 - \frac{q}{m}. \tag{4}$$

Centralized vs Distributed Approach

- The centralized version requires **all** regressors ψ_i and measurements y_i.
- Goal: develop a distributed version ensuring **every** node can compute the confidence region, merging the **local** information and keeping the same level of confidence.

Centralized vs Distributed Approach

- The centralized version requires **all** regressors ψ_i and measurements y_i.
- Goal: develop a distributed version ensuring **every** node can compute the confidence region, merging the **local** information and keeping the same level of confidence.

Proposed distributed approaches

Information diffusion via

- Flooding
- Consensus algorithm
- Mixed flooding + consensus scheme

Comparison in terms of *traffic load*.

Proposed distributed approaches

Information diffusion via

- Flooding
- Consensus algorithm
- Mixed flooding + consensus scheme

Comparison in terms of *traffic load*.

Flooding

- To evaluate z_j in (2), node i starts transmitting message $\mathbf{m}_i = \left[\psi_i^T, y_i\right]$.
- Each node initially transmits $D_f^{(0)} = n_p + 1$ values, and $D_f^{(\text{last})} = N(n_p + 1)$ values, in the end.

Flooding

- To evaluate z_j in (2), node i starts transmitting message $\mathbf{m}_i = \left[\psi_i^T, y_i\right]$.
- Each node initially transmits $D_{\mathrm{f}}^{(0)} = n_p + 1$ values, and $D_{\mathrm{f}}^{(\mathrm{last})} = N(n_p + 1)$ values, in the end.

Consensus

- In eq. (2), we have sums of local quantities.
- Average consensus algorithm is well suited.
- Node i must start transmitting $\mathbf{x}_i^{(0)} = \left[(\psi_i y_i)^T, \left\{ \psi_i \psi_i^T \right\}, \{(\alpha_{j,i} \psi_i y_i)^T\}_j, \left\{ \alpha_{j,i} \psi_i \psi_i^T \right\}_j \right]$, with $j = 1, 2, \ldots, m-1$.
- The state dimension is $D_c = m(3n_p + n_p^2)/2$.
- The state dimension is *constant*, but is *larger* than the one initially required by flooding.

Consensus

- In eq. (2), we have sums of local quantities.
- Average consensus algorithm is well suited.
- Node i must start transmitting $\mathbf{x}_i^{(0)} = \left[(\psi_i y_i)^T, \left\{ \psi_i \psi_i^T \right\}, \{(\alpha_{j,i} \psi_i y_i)^T\}_j, \left\{ \alpha_{j,i} \psi_i \psi_i^T \right\}_j \right]$, with $j = 1, 2, \ldots, m-1$.
- The state dimension is $D_c = m(3n_p + n_p^2)/2$.
- The state dimension is *constant*, but is *larger* than the one initially required by flooding.

Consensus

- In eq. (2), we have sums of local quantities.
- Average consensus algorithm is well suited.
- Node i must start transmitting
  $\mathbf{x}_i^{(0)} = \left[(\psi_i y_i)^T, \left\{ \psi_i \psi_i^T \right\}, \{(\alpha_{j,i} \psi_i y_i)^T\}_j, \left\{ \alpha_{j,i} \psi_i \psi_i^T \right\}_j \right]$, with
  $j = 1, 2, \ldots, m-1$.
- The state dimension is $D_c = m(3n_p + n_p^2)/2$.
- The state dimension is *constant*, but is *larger* than the one initially required by flooding.

Consensus

- In eq. (2), we have sums of local quantities.
- Average consensus algorithm is well suited.
- Node i must start transmitting $\mathbf{x}_i^{(0)} = \left[(\psi_i y_i)^T, \left\{\psi_i \psi_i^T\right\}, \{(\alpha_{j,i}\psi_i y_i)^T\}_j, \left\{\alpha_{j,i}\psi_i\psi_i^T\right\}_j\right]$, with $j = 1, 2, \ldots, m-1$.
- The state dimension is $D_c = m(3n_p + n_p^2)/2$.
- The state dimension is *constant*, but is *larger* than the one initially required by flooding.

Consensus

- Node i updates its state as

$$\mathbf{x}_i^{(k+1)} = \sum_{j=1}^{N} w_{i,j} \mathbf{x}_j^{(k)} \,. \tag{5}$$

- A possible choice for the coefficients: Metropolis matrix.

Consensus

- Node i updates its state as

$$\mathbf{x}_i^{(k+1)} = \sum_{j=1}^{N} w_{i,j}\mathbf{x}_j^{(k)} \,. \tag{5}$$

- A possible choice for the coefficients: Metropolis matrix.

Mixed Approach

- Mixed strategy performs flooding until data dimension exceeds D_c.
- When this happens, switch to the consensus strategy.
- The initial consensus state $\mathbf{x}_i^{(0)}$, must be set to the **average** of the already received quantities.

Mixed Approach

- Mixed strategy performs flooding until data dimension exceeds D_c.
- When this happens, switch to the consensus strategy.
- The initial consensus state $\mathbf{x}_i^{(0)}$, must be set to the **average** of the already received quantities.

Effect of information diffusion truncation

- Due to traffic load limitations, truncation of information diffusion may occur.
- The truncation yields a **loss of performance** (larger regions).
- The **confidence level** is instead kept the same as in the centralized case.

Effect of information diffusion truncation

- Due to traffic load limitations, truncation of information diffusion may occur.
- The truncation yields a **loss of performance** (larger regions).
- The **confidence level** is instead kept the same as in the centralized case.

Numerical Results Setup

- Simulations were performed using Matlab alongside the Intlab package [2].
- True parameter: $\mathbf{p}^* = [p_1^*, p_2^*, p_3^*]^T = [0.2, 0.3, 0.4]^T$.
- White Gaussian measurement noise, with variance $\sigma^2 = 115$.
- Regressors taking values in $\{-1, 1\}$.
- 90% *outer approximation* of confidence regions ($q = 1$, $m = 10$), computed at node $\#1$.
- Computational efficiency via interval analysis techniques Kieffer et al. 2014 [3].

Numerical Results Setup

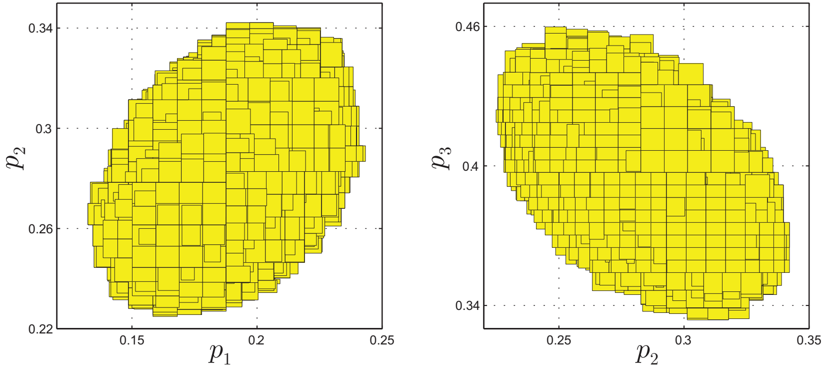
- Simulations were performed using Matlab alongside the Intlab package [2].
- True parameter: $\mathbf{p}^* = [p_1^*, p_2^*, p_3^*]^T = [0.2, 0.3, 0.4]^T$.
- White Gaussian measurement noise, with variance $\sigma^2 = 115$.
- Regressors taking values in $\{-1, 1\}$.
- 90% *outer approximation* of confidence regions ($q = 1$, $m = 10$), computed at node $\#1$.
- Computational efficiency via interval analysis techniques Kieffer et al. 2014 [3].

Numerical Results Setup

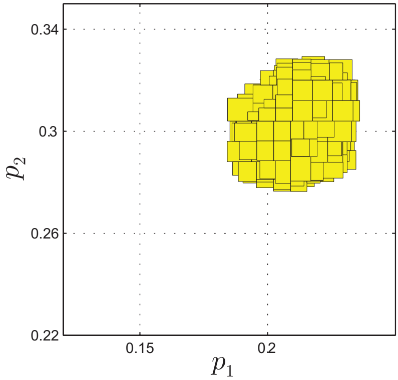
- Simulations were performed using Matlab alongside the Intlab package [2].
- True parameter: $\mathbf{p}^* = [p_1^*, p_2^*, p_3^*]^T = [0.2, 0.3, 0.4]^T$.
- White Gaussian measurement noise, with variance $\sigma^2 = 115$.
- Regressors taking values in $\{-1, 1\}$.
- 90% *outer approximation* of confidence regions ($q = 1$, $m = 10$), computed at node $\#1$.
- Computational efficiency via interval analysis techniques Kieffer et al. 2014 [3].

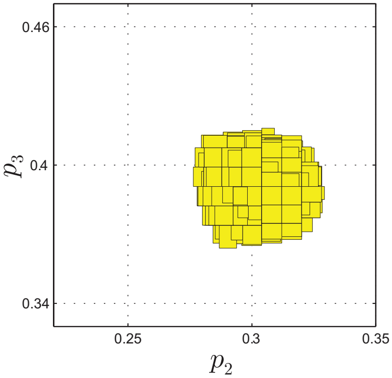
Numerical Results Setup

- Simulations were performed using Matlab alongside the Intlab package [2].
- True parameter: $\mathbf{p}^* = [p_1^*, p_2^*, p_3^*]^T = [0.2, 0.3, 0.4]^T$.
- White Gaussian measurement noise, with variance $\sigma^2 = 115$.
- Regressors taking values in $\{-1, 1\}$.
- 90% *outer approximation* of confidence regions ($q = 1$, $m = 10$), computed at node $\#1$.
- Computational efficiency via interval analysis techniques Kieffer et al. 2014 [3].

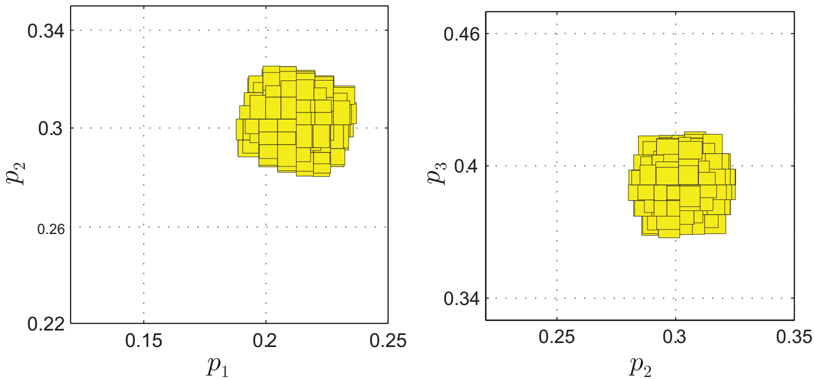


Outer approximation of the 90% confidence region at node #1 after 4 consensus iterations, when $N = 100$.

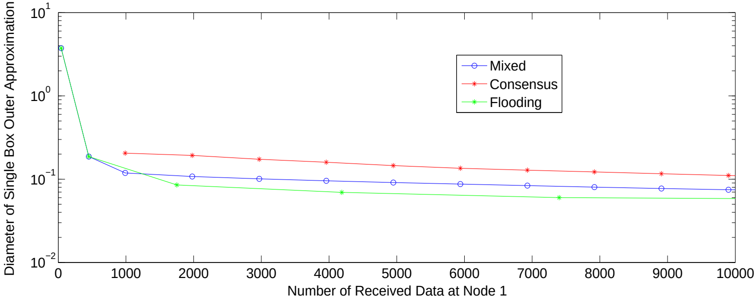




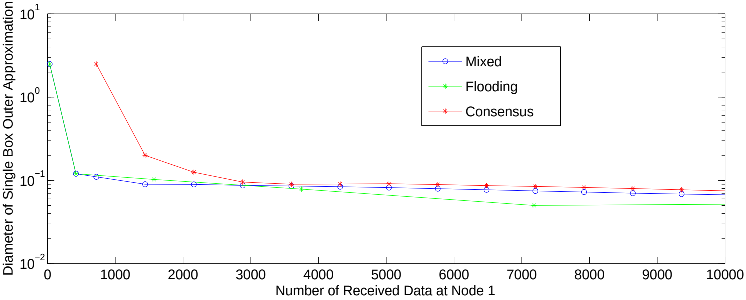
Outer approximation of the 90% confidence region at node #1 after 30 consensus iterations, when $N = 100$.



Outer approximation of the 90% confidence region at node #1 after a complete flooding, when $N = 100$.



Comparison of diameters of the single box outer approximation as a function of the amount of received data at node $\#1$, when $N = 100$ nodes.



Comparison of diameters of the single box outer approximation as a function of the amount of received data at node $\#1$, when $N = 250$ nodes.

Final Remarks

1. Distributed non-asymptotic confidence regions have the same level of confidence as centralized ones.
2. The price for information shortage is larger confidence regions.
3. Flooding is good at diffusing information, when no limitations on traffic load are present. The proposed mixed approach is advantageous when a limit is instead present.

Final Remarks

1. Distributed non-asymptotic confidence regions have the same level of confidence as centralized ones.
2. The price for information shortage is larger confidence regions.
3. Flooding is good at diffusing information, when no limitations on traffic load are present. The proposed mixed approach is advantageous when a limit is instead present.

Final Remarks

1. Distributed non-asymptotic confidence regions have the same level of confidence as centralized ones.
2. The price for information shortage is larger confidence regions.
3. Flooding is good at diffusing information, when no limitations on traffic load are present. The proposed mixed approach is advantageous when a limit is instead present.

References

[1] B. C. Csàji, M. C. Campi, and E. Weyer, “Non-asymptotic confidence regions for the least-squares estimate,” in *Proc. IFAC SYSID*, Brussels, Belgium, 2012, pp. 227–232.

[2] S. Rump, “INTLAB - INTerval LABoratory,” in *Developments in Reliable Computing*, T. Csendes, Ed. Dordrecht: Kluwer Academic Publishers, 1999, pp. 77–104. [Online]. Available: http://www.ti3.tu-harburg.de/rump/

[3] M. Kieffer and E. Walter, “Guaranteed characterization of exact non-asymptotic confidence regions as defined by {LSCR} and {SPS},” *Automatica*, vol. 50, no. 2, pp. 507–512, February 2014.