

Distributed Weighted Least-Squares and Gaussian Belief Propagation: An Integrated Approach

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

14-06-2021 / 17-06-2021

CITATION

Zivojevic, Dino; Delalic, Muhamed; Raca, Darijo; Vukobratovic, Dejan; Cosovic, Mirsad (2021): Distributed Weighted Least-Squares and Gaussian Belief Propagation: An Integrated Approach. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.14781489.v2>

DOI

[10.36227/techrxiv.14781489.v2](https://doi.org/10.36227/techrxiv.14781489.v2)

Distributed Weighted Least-Squares and Gaussian Belief Propagation: An Integrated Approach

Dino Zivojevic, Muhamed Delalic, Darijo Raca, Dejan Vukobratovic, Mirsad Cosovic

Abstract—The purpose of a state estimation (SE) algorithm is to estimate the values of the state variables considering the available set of measurements. The centralised SE becomes impractical for large-scale systems, particularly if the measurements are spatially distributed across wide geographical areas. Dividing the large-scale systems into clusters (i.e., subsystems) and distributing the computation across clusters, solves the constraints of centralised based approach. In such scenarios, using distributed SE methods brings numerous advantages over the centralised ones. In this paper, we propose a novel distributed approach to solve the linear SE model by combining local solutions obtained by applying weighted least-squares (WLS) of the given subsystems with the Gaussian belief propagation (GBP) algorithm. The proposed algorithm is based on the factor graph operating without a central coordinator, where subsystems exchange only “beliefs”, thus preserving privacy of the measurement data and state variables. Further, we propose an approach to speed-up evaluation of the local solution upon arrival of a new information to the subsystem. Finally, the proposed algorithm provides results that reach accuracy of the centralised WLS solution in a few iterations, and outperforms vanilla GBP algorithm with respect to its convergence properties.

Index Terms—Distributed Weighted Least-Squares, Gaussian Belief Propagation, Factor Graphs, Recursive Least-Squares

I. INTRODUCTION

Solving an *overdetermined* system of *linear* equations arises in numerous technical fields such as statistics, signal processing, and control [1]. Common approach to obtain a solution of such system is the least-squares method, or generally *weighted least-squares* (WLS), typically performed in a centralised manner. The WLS method can be applied in engineering problems such as state estimation (SE) in power systems [2], localisation [3], parameter estimation [4], target tracking or monitoring in wireless sensor networks [5].

Finding a WLS solution in a *large-scale* system is a challenging task because of its inherent complexity as a consequence of spatial distributions of measurements across wide geographical areas. For such scenarios, finding WLS solution by applying a centralised approach is impractical. As a result, recent years have witnessed a lot of effort by the research community in development of the novel methods for solving large-scale systems in a *distributed* manner [6]–[10]. Distributed algorithms find a solution in two steps. First, they allow every node, or a subsystem, to calculate

estimation locally using the subset of the state variables and the corresponding measurements. Finally, neighbours and a possible central coordinator, exchange local estimations to obtain global optimal solution. Instead of performing collection and estimation at the centralised control center, the distributed approach relies on exploitation of the individual subsystems’ computational capacities. The main target is to reduce communication and storage demands, as well as computational complexity, while preserving the same accuracy as the centralised WLS approach.

Traditional approaches for solving systems of linear equations rely on various centralised matrix decomposition techniques such as LU or QR factorisation [11, Ch. 4]. While the centralised matrix decomposition techniques provide reliable and fast solution, they become impractical for large-scale systems, thus calling for various iterative methods [12]–[14]. To speed up computing by taking the advantage of the system architecture, iterative methods can be applied in parallel [15]–[17] or distributed [18]–[22] fashion.

Distributed algorithms imply system partitioning, where each subsystem holds a particular subset of the state variables and corresponding local measurements. Every subsystem is able to compute local estimates while exchanging information with immediate neighbours in order to reach a globally optimal solution. In [23] authors discuss various types of distributed least-squares algorithms implemented over the mesh network, comparing them with respect to the communication cost and time-to-completion criteria. In consensus-based algorithms, every subsystem calculates a global estimate of the state variables based on local measurements. With this approach, a global solution of the least-squares problem is obtained via the agreement reached through the communication between neighbouring subsystems [18]. In [19] authors present accelerated projection-based consensus algorithm in which all subsystems, interested in reaching consensus, communicate to a central coordinator, which averages all current estimates to obtain globally optimal estimates. Authors in [10] propose a distributed algorithm for solving WLS estimation problem, where every subsystem aims to estimate one component of the state variable in a globally optimal sense. Even though the basic form of the algorithm works for general networks, authors provide different iterative algorithm to compute the global optimal estimate which converges in a finite number of steps for acyclic graphs.

Recently, *probabilistic inference* methods have been recognised as the attractive approach for solving distributed WLS systems due to their underlying distributed architecture. The

D. Zivojevic, M. Delalic, D. Raca and M. Cosovic are with Faculty of Electrical Engineering, University of Sarajevo, Bosnia and Herzegovina (e-mail: dino.zivojevic@gmail.com, muha.delalic@gmail.com, draca@etf.unsa.ba, mirsad.cosovic@gmail.com), D. Vukobratovic is with Faculty of Technical Sciences, University of Novi Sad, Serbia, (email: dejanv@uns.ac.rs).

WLS can be formulated as a probabilistic inference problem and solved using *Gaussian belief propagation* (GBP) algorithm. The GBP algorithm implemented over probabilistic graphical models called *factor graphs*, represents a powerful tool for distributed inference [24, Sec. 8.4] [25]. Assuming that state variables follow a set of jointly Gaussian distributions, authors in [26] obtain a solution of the linear system of equations by applying GBP. In that case messages between graph nodes are also Gaussian, and therefore can be fully represented by corresponding means and variances. In addition, authors in [27] use the GBP to establish certain convergence behaviour of the distributed WLS algorithm presented in [10]. Also, the GBP can be successfully used for solving SE problems in power systems with both linear and non-linear measurement models. The work in [28] applies the GBP to a linear approximation of the non-linear functions, while authors in [29] present distributed Gauss-Newton method based on the GBP for the non-linear SE.

In this paper we propose GBP-WLS, a novel distributed algorithm for solving the system of linear equations. The key novelty comes from combining local WLS solutions between neighbouring clusters obtained directly across subsystems of linear equations through GBP messages. Unlike approaches proposed in [30], [31], where a solution depends on the assumption that only pairwise connections between vector variables exist, we consider a more general measurement model. Also, by implementing adaptive mechanism via *recursive least-squares* (RLS), we introduce a standardised framework for swift *updates* of the existing local solutions when subsystems are modified by incoming messages. Finally, we analyse performances of the resulting GBP-WLS algorithm showing several advantages over the vanilla (i.e., classical) GBP algorithm used to solve WLS problems. Namely, encapsulating parts of the system into clusters eliminates impact of the local loops and therefore manifests higher rate and probability of convergence compared to the vanilla GBP.

II. CENTRALISED WEIGHTED LEAST-SQUARES METHOD

Consider the following measurement model described by the system of linear equations:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{u}, \quad (1)$$

where $\mathbf{x} = [x_1, \dots, x_n]^T$ represents the vector of state variables, $\mathbf{z} = [z_1, \dots, z_m]^T$ is observation vector, while $\mathbf{u} = [u_1, \dots, u_m]^T$ is the vector of uncorrelated measurement errors, where $u_i \sim \mathcal{N}(0, v_i)$ represents a zero-mean Gaussian distribution with variance v_i . Considering that vector $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \dots, h_m(\mathbf{x})]^T$ comprises linear measurements functions, then (1) represents a linear model and can be presented in the matrix form:

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{u}, \quad (2)$$

where $\mathbf{H} \in \mathbb{R}^{m \times n}$ is the observation matrix. Assuming an overdetermined system ($m > n$), the WLS estimator $\hat{\mathbf{x}}$ can be

obtained solving the following optimisation problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} (\mathbf{z} - \mathbf{H}\mathbf{x})^T \mathbf{W} (\mathbf{z} - \mathbf{H}\mathbf{x}), \quad (3)$$

where $\mathbf{W} \in \mathbb{R}^{m \times m}$ is diagonal weight matrix, obtained as observation error covariance matrix inverse, $\mathbf{W} = \mathbf{\Sigma}^{-1}$.

Assuming that \mathbf{H} has full column rank, a centralised solution of (3) is given by:

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{z}. \quad (4)$$

The estimation error covariance matrix can be calculated using:

$$\mathbf{P} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1}, \quad (5)$$

where diagonal elements of the matrix \mathbf{P} correspond to the variances of the state variables estimate $\hat{\mathbf{x}}$.

III. GAUSSIAN BELIEF PROPAGATION

Under assumption that each measurement error u_i follows zero-mean Gaussian distribution, measurement z_i can be associated with the following probability density function:

$$\mathcal{N}(z_i | \mathbf{x}, v_i) = \frac{1}{\sqrt{2\pi v_i}} \exp \left\{ -\frac{[z_i - h_i(\mathbf{x})]^2}{2v_i} \right\}, \quad (6)$$

where v_i is the variance associated with the i -th measurement error. It can be shown that WLS estimation from (3) is equivalent to the following maximum likelihood solution [32, Ch. 2.4]:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} \mathcal{L}(\mathbf{z} | \mathbf{x}) = \underset{\mathbf{x}}{\operatorname{argmax}} \prod_{i=1}^m \mathcal{N}(z_i | \mathbf{x}, v_i). \quad (7)$$

The lack of a priori knowledge about the state variables implies its uniform distribution. In that case, the maximum likelihood and WLS are equivalent to maximum a posteriori solution [33, Ch. 4.2.5]. Hence, converting the WLS problem to the probabilistic domain leads to a representation of a system of equations using a graphical model. This transformation allows us to solve the maximum a posteriori problem in a distributed manner by applying the GBP algorithm.

The factor graph describes a factorisation of the likelihood function $\mathcal{L}(\mathbf{z} | \mathbf{x})$. It comprises the set of *factor nodes* \mathcal{F} and the set of *variable nodes* \mathcal{X} . The vector of state variables \mathbf{x} determines the set of variable nodes $\mathcal{X} = \{x_1, \dots, x_n\}$, while the vector of measurement functions $\mathbf{h}(\mathbf{x})$ defines the set of factor nodes $\mathcal{F} = \{f_1, \dots, f_m\}$. A factor node $f_i \in \mathcal{F}$ connects to a variable node $x_s \in \mathcal{X}$ if and only if the state variable x_s is an argument of the matching measurement function $h_i(\cdot)$. The GBP algorithm efficiently calculates marginal distributions of state variables by passing two types of messages along the edges of the factor graph: i) a variable node to a factor node, and ii) a factor node to a variable node messages. The marginal inference provides marginal probability distributions that is used to estimate values $\hat{\mathbf{x}}$ of state variables \mathbf{x} .

Message from a variable node to a factor node: Consider a part of a factor graph shown in Fig. 1(a) with a group of factor nodes $\mathcal{F}_s = \{f_i, f_w, \dots, f_W\} \subseteq \mathcal{F}$ that are neighbours

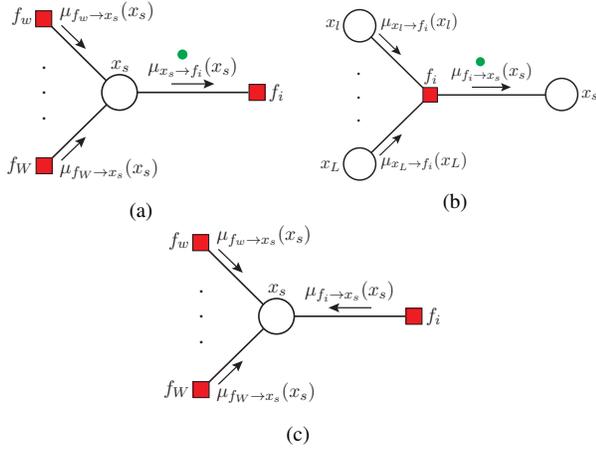


Fig. 1. Message $\mu_{x_s \rightarrow f_i}(x_s)$ from variable node x_s to factor node f_i (subfigure a), message $\mu_{f_i \rightarrow x_s}(x_s)$ from factor node f_i to variable node x_s (subfigure b), and marginal inference of the variable node x_s (subfigure c).

of the variable node $x_s \in \mathcal{X}$. It can be shown that the message $\mu_{x_s \rightarrow f_i}(x_s)$ from the variable node x_s to the factor node f_i is proportional (i.e., \propto) to the Gaussian function:

$$\mu_{x_s \rightarrow f_i}(x_s) \propto \mathcal{N}(x_s | z_{x_s \rightarrow f_i}, v_{x_s \rightarrow f_i}), \quad (8)$$

with mean $z_{x_s \rightarrow f_i}$ and variance $v_{x_s \rightarrow f_i}$ obtained as:

$$z_{x_s \rightarrow f_i} = \left(\sum_{f_a \in \mathcal{F}_s \setminus f_i} \frac{z_{f_a \rightarrow x_s}}{v_{f_a \rightarrow x_s}} \right) v_{x_s \rightarrow f_i} \quad (9a)$$

$$\frac{1}{v_{x_s \rightarrow f_i}} = \sum_{f_a \in \mathcal{F}_s \setminus f_i} \frac{1}{v_{f_a \rightarrow x_s}}. \quad (9b)$$

Calculation of the message $\mu_{x_s \rightarrow f_i}(x_s)$ depends on messages received from all the neighbouring factor nodes of variable node x_s from the set $\mathcal{F}_s \setminus f_i$. Upon receiving all messages, variable node x_s evaluates the message $\mu_{x_s \rightarrow f_i}(x_s)$ by applying (9) and sends it to the factor node f_i .

Message from a factor node to a variable node: Consider a part of a factor graph shown in Fig. 1(b) consisting of a group of variable nodes $\mathcal{X}_i = \{x_s, x_l, \dots, x_L\} \subseteq \mathcal{X}$ that are neighbours of the factor node $f_i \in \mathcal{F}$. Similar to the calculation of the message $\mu_{x_s \rightarrow f_i}(x_s)$, the message $\mu_{f_i \rightarrow x_s}(x_s)$ can be computed only when all other incoming messages (variable to factor node messages) are known. It can be shown that the message $\mu_{f_i \rightarrow x_s}(x_s)$ from the factor node f_i to the variable node x_s is represented by the Gaussian function:

$$\mu_{f_i \rightarrow x_s}(x_s) \propto \mathcal{N}(x_s | z_{f_i \rightarrow x_s}, v_{f_i \rightarrow x_s}), \quad (10)$$

with mean $z_{f_i \rightarrow x_s}$ and variance $v_{f_i \rightarrow x_s}$ obtained as:

$$z_{f_i \rightarrow x_s} = \frac{1}{K_{x_s}} \left(z_i - \sum_{x_b \in \mathcal{X}_i \setminus x_s} K_{x_b} z_{x_b \rightarrow f_i} \right) \quad (11a)$$

$$v_{f_i \rightarrow x_s} = \frac{1}{K_{x_s}^2} \left(v_i + \sum_{x_b \in \mathcal{X}_i \setminus x_s} K_{x_b}^2 v_{x_b \rightarrow f_i} \right). \quad (11b)$$

where coefficients are defined according to the linear function:

$$h_i(x_s, x_l, \dots, x_L) = K_{x_s} x_s + \sum_{x_b \in \mathcal{X}_i \setminus x_s} K_{x_b} x_b, \quad (12)$$

where $\mathcal{X}_i \setminus x_s$ is the set of variable nodes incident to the factor node f_i , excluding the variable node x_s . Hence, after the factor node f_i receives messages from all of the neighbouring variable nodes from the set $\mathcal{X}_i \setminus x_s$, it evaluates the message $\mu_{f_i \rightarrow x_s}(x_s)$ according to (11), and sends it to the variable node x_s .

Marginal inference: It can be shown that the marginal of the variable node x_s , illustrated in Fig. 1(c), is represented by the Gaussian function:

$$p(x_s) \propto \mathcal{N}(x_s | \hat{x}_s, \hat{v}_s), \quad (13)$$

with the mean value \hat{x}_s and variance \hat{v}_s :

$$\hat{x}_s = \left(\sum_{f_a \in \mathcal{F}_s} \frac{z_{f_a \rightarrow x_s}}{v_{f_a \rightarrow x_s}} \right) \hat{v}_s \quad (14a)$$

$$\frac{1}{\hat{v}_s} = \sum_{f_a \in \mathcal{F}_s} \frac{1}{v_{f_a \rightarrow x_s}}. \quad (14b)$$

Finally, the mean-value \hat{x}_s is adopted as the estimated value of the state variable x_s .

IV. DISTRIBUTED WEIGHTED LEAST-SQUARES METHOD

Distributed architecture of GBP allow us to run GBP inference over the sub-parts of connected factor graph defined by (1). This approach results in better convergence performance compared to the traditional approach of running the GBP inference over the entire factor graph, as we demonstrate in this paper.

We start by dividing the factor graph of the global system (1) into a disjoint set of clusters $\mathcal{C} = \{c_1, \dots, c_p\}$, where the k -th cluster c_k encapsulates n_{c_k} variable nodes $\mathcal{X}_{c_k} \neq \emptyset \subset \mathcal{X}$. Each cluster c_k contains m_{c_k} local factor nodes $\mathcal{F}_{c_k} \subset \mathcal{F}$, where factor nodes from \mathcal{F}_{c_k} form connections exclusively between variable nodes from \mathcal{X}_{c_k} (i.e., we can view each cluster as an isolated island). Therefore, the following set of linear system of equations represents each cluster c_k :

$$\mathbf{z}_{c_k} = \mathbf{H}_{c_k} \mathbf{x}_{c_k} + \mathbf{u}_{c_k}, \quad (15)$$

where $\mathbf{x}_{c_k} \in \mathbb{R}^{n_{c_k}}$ represents the local vector of the state variables, $\mathbf{H}_{c_k} \in \mathbb{R}^{m_{c_k} \times n_{c_k}}$ is the local regular observation matrix, $\mathbf{z}_{c_k} \in \mathbb{R}^{m_{c_k}}$ represents the local measurement values, while $\mathbf{u}_{c_k} \in \mathbb{R}^{m_{c_k}}$ corresponds with the local measurements errors.

Note that, during the clustering process, we did not include all factor nodes into the clusters. The omitted set of factor nodes, denoted as $\mathcal{F}^{\text{tie}} \subset \mathcal{F}$, represent connections that *tie* (i.e., connect) state variables, denoted as $\mathcal{X}^{\text{tie}} \subset \mathcal{X}$, that belong to different clusters. This separation imposes the introduction of an additional centralised coordinator for managing factor nodes \mathcal{F}^{tie} . Without loss of generality, factor nodes from \mathcal{F}^{tie} can be assigned to a particular cluster $\mathcal{F}_{c_k}^{\text{tie}} \subset \mathcal{F}^{\text{tie}}$, removing

the requirement for the centralised coordinator. However, the set of factor nodes \mathcal{F}^{tie} links with a certain subset of variable nodes $\mathcal{X}_{c_k}^{\text{tie}}$ inside each of the clusters c_k . Factor nodes from the \mathcal{F}^{tie} preserve communication between clusters according to the following system of equations¹:

$$\mathbf{z}_{\text{tie}} = \mathbf{H}_{\text{tie}}\mathbf{x}_{\text{tie}} + \mathbf{u}_{\text{tie}}, \quad (16)$$

where $\mathbf{x}_{\text{tie}} \in \mathbb{R}^{n_{\text{tie}}}$ represents the vector of the tie state variables, $\mathbf{H}_{\text{tie}} \in \mathbb{R}^{m_{\text{tie}} \times n_{\text{tie}}}$ is the tie observation matrix, $\mathbf{z}_{\text{tie}} \in \mathbb{R}^{m_{\text{tie}}}$ represents the tie measurement values, while $\mathbf{u}_{\text{tie}} \in \mathbb{R}^{m_{\text{tie}}}$ corresponds with the tie measurements errors. Dimensions $m_{\text{tie}} = m - m_{c_1} - \dots - m_{c_p}$ and n_{tie} represent the number of omitted equations and the number of the corresponding state variables, respectively.

By applying (4) to (15), we obtain the local estimate, where messages from the set of factor nodes \mathcal{F}^{tie} to variable nodes $\mathcal{X}_{c_k}^{\text{tie}}$ are missing to find the exact solution according to (14). More precisely, the equation (14) requires knowledge of means $z_{f_i \rightarrow x_s}$ and variances $v_{f_i \rightarrow x_s}$, $f_i \in \mathcal{F}^{\text{tie}}$, $x_s \in \mathcal{X}_{c_k}^{\text{tie}}$, for exact marginals calculation. Hence, we integrate the mean $z_{f_i \rightarrow x_s}$ and variance $v_{f_i \rightarrow x_s}$ into the corresponding cluster that contains the state variable x_s , thus completing the system defined in (15). This information can either be observed as a singly-connected factor node at the variable node x_s , or an artificial measurement with the mean $s_s \equiv z_{f_i \rightarrow x_s}$, and error proportional to the variance $q_s \propto v_{f_i \rightarrow x_s}$. As a result, the system of equations (15) expands to:

$$\begin{bmatrix} \mathbf{z}_{c_k} \\ \mathbf{s}_{c_k} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{c_k} \\ \mathbf{G}_{c_k} \end{bmatrix} \mathbf{x}_{c_k} + \begin{bmatrix} \mathbf{u}_{c_k} \\ \mathbf{q}_{c_k} \end{bmatrix}, \quad (17)$$

where vectors $\mathbf{s}_{c_k} \in \mathbb{R}^{d_{c_k}}$ and $\mathbf{q}_{c_k} \in \mathbb{R}^{d_{c_k}}$ contain artificial measurement means and errors, respectively. Further, $\mathbf{G}_{c_k} \in \mathbb{E}^{d_{c_k} \times n_{c_k}}$ is an artificial matrix, where $\mathbb{E} = \{0, 1\}$, and d_{c_k} is the total degree number of variable nodes $\mathcal{X}_{c_k}^{\text{tie}}$ in the cluster c_k .

Finally, by solving (17) according to (4), we obtain the mean and variance set $(\hat{\mathbf{x}}_{c_k}, \hat{\mathbf{v}}_{c_k})$ that corresponds to (14). After obtaining these marginals, the last step includes defining messages from variable nodes $\mathcal{X}_{c_k}^{\text{tie}}$ to factor nodes \mathcal{F}^{tie} . Using (9) and (14), we obtain the mean $z_{x_s \rightarrow f_i}$ and variance $v_{x_s \rightarrow f_i}$, $f_i \in \mathcal{F}^{\text{tie}}$, $x_s \in \mathcal{X}_{c_k}^{\text{tie}}$ as:

$$z_{x_s \rightarrow f_i} = \left(\frac{\hat{x}_s}{\hat{v}_s} - \frac{z_{f_i \rightarrow x_s}}{v_{f_i \rightarrow x_s}} \right) v_{x_s \rightarrow f_i} \quad (18a)$$

$$\frac{1}{v_{x_s \rightarrow f_i}} = \frac{1}{\hat{v}_s} - \frac{1}{v_{f_i \rightarrow x_s}}, \quad (18b)$$

where \hat{x}_s and \hat{v}_s are obtained by solving (17). The mean $z_{f_i \rightarrow x_s}$ and variance $v_{f_i \rightarrow x_s}$ are calculated as defined in (11).

A. Vanilla distributed WLS algorithm

We formalise the proposed method *without* the central coordinator in Algorithm 1. First, we divide the system into

¹Note that this system of equations complete our full measurement model defined by (1)

a set of clusters $\mathcal{C} = \{c_1, \dots, c_p\}$, where the k -th cluster contains $\mathcal{X}_{c_k} \subset \mathcal{X}$ and $\mathcal{F}_{c_k} \subset \mathcal{F}$ variable and factor nodes, respectively. After the initialisation step, where default values for means and variances are assigned to all messages from tie factor nodes to tie variable nodes (lines 3-8), the main loop of the algorithm is initiated (lines 10-27). In every iteration, all clusters are updated (line 13) and WLS is computed over the clusters (line 14). Next, means and variances from tie variable nodes to tie factor nodes are calculated using the GBP (lines 15-18). Upon the completion of previous lines, all means and variances from tie factor nodes to tie variable nodes can be calculated (lines 20-25). The main loop is active until the stopping criterion is met, which can be defined as either maximum number of iterations or minimum residual value.

Algorithm 1 Vanilla distributed WLS algorithm

```

1: procedure INITIALISATION
2:   divide global system into disjoint set of clusters  $\mathcal{C} = \{c_1, \dots, c_p\}$ 
3:   for each  $c_k \in \mathcal{C}$  do
4:     for each  $f_i \in \mathcal{F}_{c_k}^{\text{tie}}$  do
5:       initialise mean values  $z_{f_i \rightarrow x_s}$ ,  $x_s \in \mathcal{X}^{\text{tie}}$ 
6:       initialise variance values  $v_{f_i \rightarrow x_s}$ ,  $x_s \in \mathcal{X}^{\text{tie}}$ 
7:     end for
8:   end for
9: end procedure
10: procedure ITERATIVE UPDATE
11:   while stopping criterion is not met do
12:     for each  $c_k \in \mathcal{C}$  do
13:       update  $\mathbf{s}_{c_k}$  and  $\mathbf{q}_{c_k}$ 
14:       compute the WLS solution of (17)
15:       for each  $x_s \in \mathcal{X}_{c_k}^{\text{tie}}$  do
16:         compute mean values  $z_{x_s \rightarrow f_i}$ ,  $f_i \in \mathcal{F}^{\text{tie}}$ , using (18a)
17:         compute variance values  $v_{x_s \rightarrow f_i}$ ,  $f_i \in \mathcal{F}^{\text{tie}}$ , using (18b)
18:       end for
19:     end for
20:     for each  $c_k \in \mathcal{C}$  do
21:       for each  $f_i \in \mathcal{F}_{c_k}^{\text{tie}}$  do
22:         compute mean values  $z_{f_i \rightarrow x_s}$ ,  $x_s \in \mathcal{X}^{\text{tie}}$ , using (11a)
23:         compute variance values  $v_{f_i \rightarrow x_s}$ ,  $x_s \in \mathcal{X}^{\text{tie}}$ , using (11b)
24:       end for
25:     end for
26:   end while
27: end procedure

```

In the next subsection, we propose different approach to solve the system of linear equations (17), alternative to solving the system in each iteration from the scratch (line 14). Note that for the system defined in (17), only vectors \mathbf{s}_{c_k} and \mathbf{q}_{c_k} dynamically change their values through iterations according to messages from tie factor nodes to tie variable nodes.

B. State information update via RLS

When the system is modified by a new data, we can update the existing WLS solution by using RLS algorithm. Let us assume that we have the state variable estimate $\hat{\mathbf{x}}_{c_k}$ and estimation error covariance matrix \mathbf{P}_{c_k} from WLS estimation

applied to the cluster c_k . After the cluster c_k receives a set of incoming messages from tie factor nodes to tie variable nodes, we update only vectors \mathbf{s}_{c_k} and \mathbf{q}_{c_k} . This allows us to efficiently update the existing WLS solution by exploiting only a new data and previously calculated estimates. We start with the update of the error covariance matrix within cluster c_k as:

$$(\mathbf{P}_{c_k}^{\text{new}})^{-1} = \mathbf{P}_{c_k}^{-1} + \mathbf{G}_{c_k}^T \mathbf{R}_{c_k} \mathbf{G}_{c_k}, \quad (19)$$

where $(\mathbf{P}_{c_k}^{\text{new}})^{-1}$ is a new updated covariance matrix based on \mathbf{G}_{c_k} which is equivalent to a new observation matrix, and \mathbf{R}_{c_k} represents the inverse error covariance matrix (e.g., weighted matrix). With the updated covariance matrix, a new state variable estimate can be calculated as follows:

$$\hat{\mathbf{x}}_{c_k}^{\text{new}} = \hat{\mathbf{x}}_{c_k} + \mathbf{K}(\mathbf{s}_{c_k} - \mathbf{G}_{c_k} \hat{\mathbf{x}}_{c_k}), \quad (20)$$

where \mathbf{K} represents the gain matrix defined by:

$$\mathbf{K} = \mathbf{P}_{c_k}^{\text{new}} \mathbf{G}_{c_k}^T \mathbf{R}_{c_k}, \quad (21)$$

and vector \mathbf{s}_{c_k} represents a new measurement (information) with $\mathbf{s}_{c_k} - \mathbf{G}_{c_k} \hat{\mathbf{x}}_{c_k}$ indicating the error in our prediction of \mathbf{s}_{c_k} . With the RLS approach, the basis of the cluster c_k is easily updated after reception of incoming messages. If the cluster is updated by p additional rows, the complexity of the RLS algorithm is $\mathcal{O}(pn^2)$, resulting in a reduced computational complexity. As a result, the GBP-WLS algorithm can be accelerated.

V. NUMERICAL RESULTS

We evaluate the performance of the proposed method using IEEE 30-bus test case divided into 4 clusters $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$, which correspond to resulting factor graph obtained according to the set of measurements, shown in Fig. 2. First, we apply the power flow analysis to generate the exact solution on the IEEE 30-bus test case. Further, we corrupt the exact solution with the additive white Gaussian noise, and linearize the non-linear model to obtain the DC SE problem that ignores the reactive powers and transmission losses, and takes into account only the active powers. Finally, we observe a randomly selected set of measurements that includes injection, power flows, and bus voltage angle measurements. In order to get statistically significant results, we generate 9000 random measurement configurations, where each set contains 10 bus voltage angle measurements, while injection and power flows measurements are generated according to measurement redundancy $\gamma \in \{2, 3, 4, 5\}$.

We start by comparing the convergence performance as the function of the number of measurements of our proposed GBP-WLS algorithm against the traditional scalar (vanilla) GBP. For the convergence metrics, we adopt for the weighted residual sum of squares (WRSS) [29]. The algorithm is said to converge if $(\text{WRSS}_{\text{GBP-WLS}}/\text{WRSS}_{\text{WLS}}) \rightarrow 1$, where the WRSS_{WLS} represents the WRSS value for the centralised WLS solution applied over the global system. We use the same performance metric for the GBP as well. Fig. 3 shows the aforementioned convergence comparison between GBP-

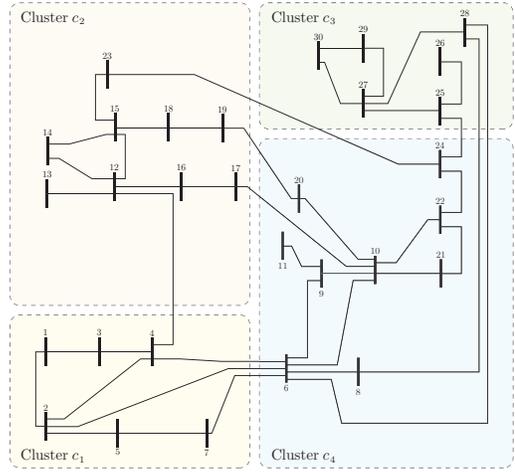


Fig. 2. The IEEE 30-bus test case segmented into four areas which correspond to resulting factor graph obtained according to the set of measurements.

WLS and GBP algorithms as redundancy increases. It is evident that the GBP-WLS expresses high convergence ability which is maintained as the number of measurements increases. However, GBP algorithm shows worse convergence performance in all scenarios compared to GBP-WLS. This difference increases proportional to the number of measurements. Even for the $\gamma = 2$, GBP-WLS has 33.6% higher convergence rate than vanilla GBP. The behaviour of GBP algorithm for larger number of measurements is expected since clusters mostly absorb the influence of loops when the graph becomes more complex. Unlike vanilla GBP algorithm, the number of measurements have no significant impact on the proposed GBP-WLS convergence performance.

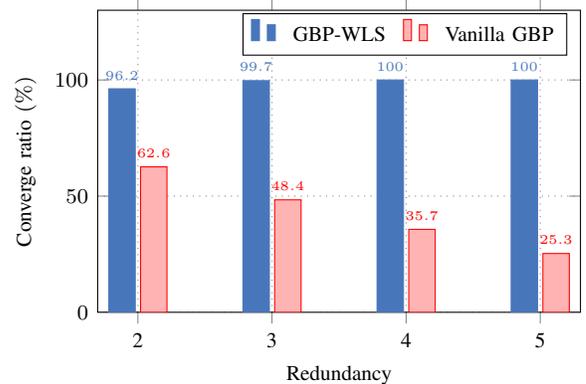


Fig. 3. Converge ratio of instances of the GBP-WLS and GBP algorithms for various redundancy $\gamma \in \{2, 3, 4, 5\}$.

Further, the convergence rate of the GBP-WLS algorithm for the first few iterations is shown in Fig. 4. We use root mean square error (RMSE) metric for convergence rate analysis. Fig. 4 indicates that the GBP-WLS algorithm expresses a relatively high convergence rate, reaching the precision of the centralised WLS estimator in only a few iterations. Note that by applying RLS for updating local estimates, the complexity

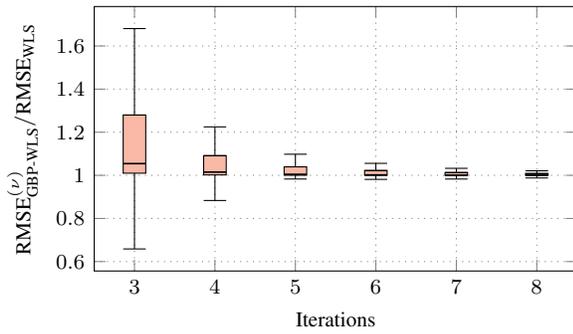


Fig. 4. Convergence rate of the GBP-WLS algorithm.

diminishes as the number of iterations increases.

VI. CONCLUSIONS

Over the last few decades many systems (e.g., power systems, wireless sensor networks) grew from just a few nodes to thousands and even millions of densely interconnected nodes. Solving such large-scale systems with a traditional centralised algorithms is unfeasible, paving the way for the novel distributed techniques that can successfully solve these massive systems. In this paper we presented how encapsulating parts of the factor graph into clusters and combining WLS solutions obtained locally with GBP messages results in better convergence properties compared to the ordinary scalar GBP, additionally emphasising the high rate of convergence of the GBP-WLS. We consider a general approach for the update of existing solutions within individual clusters of the system, and introduce different approach to update local solutions by implementing RLS. Future work includes establishing convergence conditions for the iterative GBP-WLS algorithm, as well as implementation of a special version of the GBP-WLS for tree structured graphs that converges in a fixed number of steps.

VII. ACKNOWLEDGEMENTS

This paper has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement number 856967. Also, authors acknowledge the support of the Ministry of Education, Science and Youth of Sarajevo Canton.

REFERENCES

- [1] Å. Björck, *Numerical methods for least squares problems*. SIAM, 1996.
- [2] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.
- [3] Y. Shang and W. Ruml, "Improved MDS-based localization," in *IEEE INFOCOM 2004*, vol. 4, 2004, pp. 2640–2651.
- [4] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4583–4588, 2009.
- [5] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh, "Monitoring volcanic eruptions with a wireless sensor network," in *Proceedings of the Second European Workshop on Wireless Sensor Networks, 2005*. IEEE, 2005, pp. 108–120.
- [6] A. J. Conejo, S. de la Torre, and M. Canas, "An optimization approach to multiarea state estimation," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 213–221, 2007.
- [7] T. C. Aysal and K. E. Barner, "Constrained decentralized estimation over noisy channels for sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1398–1410, 2008.
- [8] S. Nabavi, J. Zhang, and A. Chakraborty, "Distributed optimization algorithms for wide-area oscillation monitoring in power systems using interregional PMU-PDC architectures," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2529–2538, 2015.
- [9] U. A. Khan, S. Kar, and J. M. Moura, "Distributed sensor localization in random environments using minimal number of anchor nodes," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 2000–2016, 2009.
- [10] D. E. Marelli and M. Fu, "Distributed weighted least-squares estimation with fast convergence for large-scale systems," *Automatica*, vol. 51, pp. 27–39, 2015.
- [11] P. C. Hansen, V. Pereyra, and G. Scherer, *Least squares data fitting with applications*. JHU Press, 2013.
- [12] Y. Saad, *Iterative methods for sparse linear systems*. SIAM, 2003.
- [13] L. A. Hageman and D. M. Young, *Applied iterative methods*. Courier Corporation, 2012.
- [14] M. Cosovic and D. Vukobratovic, "Fast real-time DC state estimation in electric power systems using belief propagation," in *Proc. IEEE SmartGridComm*, 2017, pp. 207–212.
- [15] M. T. Heath, E. Ng, and B. W. Peyton, "Parallel algorithms for sparse linear systems," *SIAM review*, vol. 33, no. 3, pp. 420–460, 1991.
- [16] K. E. Prikopa, W. N. Gansterer, and E. Wimmer, "Parallel iterative refinement linear least squares solvers based on all-reduce operations," *Parallel Computing*, vol. 57, pp. 167–184, 2016.
- [17] R. A. Renaut, "A parallel multisplitting solution of the least squares problem," *Numerical linear algebra with applications*, vol. 5, no. 1, pp. 11–31, 1998.
- [18] X. Wang, J. Zhou, S. Mou, and M. J. Corless, "A distributed algorithm for least square solutions of linear equations," *arXiv preprint arXiv:1709.10157*, 2017.
- [19] N. Azizan-Ruhi, F. Lahouti, A. S. Avestimehr, and B. Hassibi, "Distributed solution of large-scale linear systems via accelerated projection-based consensus," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3806–3817, 2019.
- [20] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Consensus-based distributed least-mean square algorithm using wireless ad hoc networks," in *Proc. Allerton Conf*, vol. 2. Citeseer, 2007, pp. 568–574.
- [21] G. Mateos and G. B. Giannakis, "Distributed recursive least-squares: Stability and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3740–3754, 2012.
- [22] A. H. Sayed and C. G. Lopes, "Distributed recursive least-squares strategies over adaptive networks," in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*. IEEE, 2006, pp. 233–237.
- [23] L. Shi, L. Zhao, W.-Z. Song, G. Kamath, Y. Wu, and X. Liu, "Distributed least-squares iterative methods in networks: A survey," *arXiv preprint arXiv:1706.07098*, 2017.
- [24] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [25] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [26] D. Bickson, "Gaussian belief propagation: Theory and application," *arXiv preprint arXiv:0811.2518*, 2008.
- [27] Q. Yang, Z. Zhang, and M. Fu, "Distributed weighted least-squares estimation for networked systems with edge measurements," *Automatica*, vol. 120, p. 109091, 2020.
- [28] Y. Hu, A. Kuh, T. Yang, and A. Kavcic, "A belief propagation based power distribution system state estimator," *IEEE Comput. Intell. Mag.*, vol. 6, no. 3, pp. 36–46, 2011.
- [29] M. Cosovic and D. Vukobratovic, "Distributed Gauss-Newton method for state estimation using belief propagation," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 648–658, 2018.
- [30] X. Tai, Z. Lin, M. Fu, and Y. Sun, "A new distributed state estimation technique for power networks," in *Proc. IEEE ACC*, 2013, pp. 3338–3343.
- [31] T. Sui, D. E. Marelli, and M. Fu, "Convergence analysis of Gaussian belief propagation for distributed state estimation," in *Proc. IEEE CDC*, 2015, pp. 1106–1111.
- [32] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.
- [33] S. V. Vaseghi, *Advanced digital signal processing and noise reduction*. John Wiley & Sons, 2008.