Supplementary Materials of Distributed Probabilistic Learning for Camera Networks with Missing Data

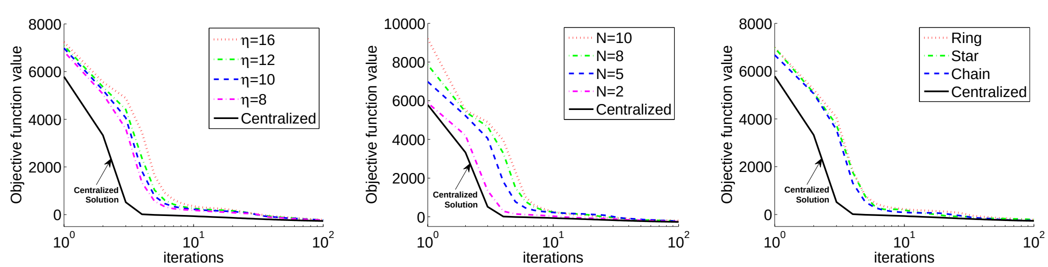
Sejong Yoon
Department of Computer Science
Rutgers University
Piscataway, NJ 08854
sjyoon@cs.rutgers.edu

Vladimir Pavlovic
Department of Computer Science
Rutgers University
Piscataway, NJ 08854
vladimir@cs.rutgers.edu

1 Synthetic Data Result

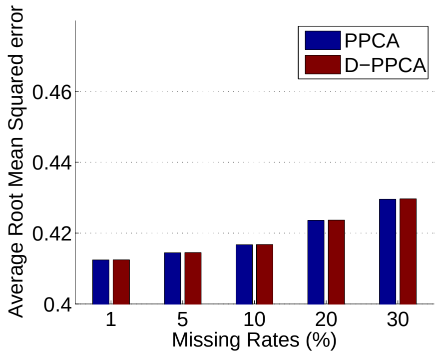
We first demonstrate the empirical convergence properties of the D-PPCA. Note that the general convergence properties are implied by the Augmented Lagrangian optimization algorithm. Additionally, in a distributed network setting the convergence will depend on the connectivity structure of the network, which in turn depends on the spectral properties of its graph Laplacian. We generated 50 dimensional 100 random samples from $\mathcal{N}(0, 0.2 \cdot \mathbf{I})$. We assigned 20 samples equally to each node in a 5-nodes network connected with ring topology to find a 5 dimensional subspace. Our convergence criterion is the relative change in objective of (1) and we stop when it is smaller than 10^{-5}. In real settings, one can monitor local parameter updates instead. We initialized parameters with random values from a uniform distribution. Alternative choices of starting points may lead to faster convergence. If not explicitly mentioned otherwise, all our results are averaged over 20 independent random initializations.

Fig. 1a shows the convergence curve of D-PPCA for various η values. As one can easily see, all η values lead to convergence within 10^2 iterations. Moreover, the value they converge to is equivalent to centralized solution meaning we can achive the same global solution using the distributed algorithm. This behavior matches results reported in [1]. Fig. 1b shows convergence curve as a function of the number of nodes in a network. In all cases, D-PPCA successfully converged within 10^2 iterations. Similar trends were observed with networks of more than 10 nodes. We also conducted experiments to test the effects of network topology on the parameter convergence. Fig. 1c depicts the result for three simple network types. In all cases we considered, D-PPCA reached near the stationary point within only 10 iterations regardless of any of the aforementioned factors.

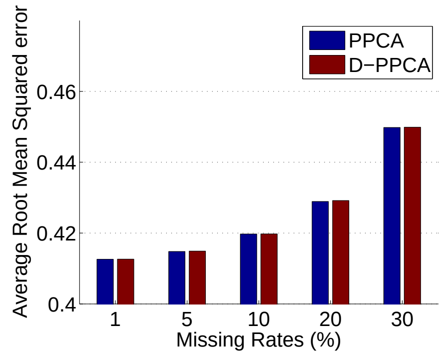


(a) Impact of η (b) Impact of number of nodes (c) Impact of topologies

Figure 1: Convergence trends of D-PPCA.



(a) Missing At Random



(b) Missing Not At Random

Figure 2: Average root mean squared error of reconstructions based on PPCA and D-PPCA results.

One of the main benefits of employing probabilistic formulation of PCA is the flexibility of allowing missing values. Here, we consider two possibilities of missing values; the case when values missing at random (MAR) and the case when values missing not at random (MNAR). In [2], it has been shown that probabilistic formulations of PCA can deal with missing values, particularly well in the MAR setting. The same conclusion holds for D-PPCA. As shown in Fig. 2a, D-PPCA can effectively reconstruct the original measurement comparable to its centralized counterpart under different amounts of missing values. This fact also holds for MNAR case although the error tends to be slightly larger than in the MAR case.

2 Full Derivation of Distributed Probabilistic PCA

2.1 Quick Reference to Notations

- $G = (V, E)$: Undirected connected graph with vertices in V and edges in E
- $i, j \in V$: Node index
- $e_{ij} = (i, j) \in E$: Edge connecting node i and node j
- $\mathcal{B}_i = \{j; e_{ij} \in E\}$: Set of neighbor nodes directly connected to i-th node
- N_i: The number of samples collected in i-th node
- $\mathbf{z}_{in}$: n-th $M \times 1$ dimensional latent variable at node i where $n = 1, ..., N_i$
- $\mathbf{x}_{in}$: n-th $D \times 1$ dimensional column vector at node i where $n = 1, ..., N_i$
- $\mathbf{Z}_i = \{\mathbf{z}_{in}; n = 1, ..., N_i\}$
- $\mathbf{X}_i = \{\mathbf{x}_{in}; n = 1, ..., N_i\}$
- $\mathbf{W}_i$, $\boldsymbol{\mu}_i$, a_i: Model parameters

2.2 Distributed Algorithm for Probabilistic PCA

In a distributed probabilistic model setting, we impose consensus constraints on parameters for each node. With the introduction of auxiliary variables, we can assure that all parameters reach consensus only by local optimizations. Using this idea, Forero, et al. proposed an iterative EM algorithm for the Gaussian mixture model [3]. Using a similar approach, we derive an iterative EM algorithm for PPCA. In the centralized setting, the local optimization problem using expectation on the complete data log likelihood with respect to the posterior of the latent variable is

$$\min_{\{f_{\mathbf{Z}_i}, \mathbf{W}_i, \boldsymbol{\mu}_i, a_i : i \in V\}} -F(f_{\mathbf{Z}_i}, \mathbf{W}_i, \boldsymbol{\mu}_i, a_i) = -\mathbb{E}_{f_{\mathbf{Z}_i}} \left[\sum_{n=1}^{N_i} \log P(\mathbf{x}_{in}, \mathbf{z}_{in} | \mathbf{W}_i, \boldsymbol{\mu}_i, a_i^{-1}) \right]$$

where $f_{\mathbf{Z}_i} = p(\mathbf{Z}_i|\mathbf{X}_i)$[1]. If we impose the consensus constraints on this, the constrained local optimization problem becomes

$$\min_{\{f_{\mathbf{Z}_i}, \mathbf{W}_i, \boldsymbol{\mu}_i, a_i : i \in V\}} -F(f_{\mathbf{Z}_i}, \mathbf{W}_i, \boldsymbol{\mu}_i, a_i) \tag{1}$$
$$\begin{aligned} s.t. \quad & \mathbf{W}_i = \boldsymbol{\rho}_{ij}, \quad \boldsymbol{\rho}_{ij} = \mathbf{W}_j \quad i \in V, j \in \mathcal{B}_i \\ & \boldsymbol{\mu}_i = \boldsymbol{\phi}_{ij}, \quad \boldsymbol{\phi}_{ij} = \boldsymbol{\mu}_j \quad i \in V, j \in \mathcal{B}_i \\ & a_i = \psi_{ij}, \quad \psi_{ij} = a_j \quad i \in V, j \in \mathcal{B}_i \end{aligned}$$

where $\boldsymbol{\rho}_{ij}, \boldsymbol{\phi}_{ij}, \psi_{ij}$ are auxiliary variables. If we solve this local optimization problem, we also solve global optimization since global optimization is simply the summation of local ones given consensus constraints meet. The augmented Lagrangian of (1) is

$$\begin{aligned} \mathcal{L}(\Phi_i) = & -F(f_{\mathbf{Z}_i}, \mathbf{W}_i, \boldsymbol{\mu}_i, a_i) \\ & + \sum_{i \in V} \sum_{j \in \mathcal{B}_i} \left(\boldsymbol{\lambda}_{ij1}^{\mathrm{T}} (\mathbf{W}_i - \boldsymbol{\rho}_{ij}) + \boldsymbol{\lambda}_{ij2}^{\mathrm{T}} (\boldsymbol{\rho}_{ij} - \mathbf{W}_j) \right) \\ & + \sum_{i \in V} \sum_{j \in \mathcal{B}_i} \left(\boldsymbol{\gamma}_{ij1}^{\mathrm{T}} (\boldsymbol{\mu}_i - \boldsymbol{\phi}_{ij}) + \boldsymbol{\gamma}_{ij2}^{\mathrm{T}} (\boldsymbol{\phi}_{ij} - \boldsymbol{\mu}_j) \right) \\ & + \sum_{i \in V} \sum_{j \in \mathcal{B}_i} \left(\beta_{ij1} (a_i - \psi_{ij}) + \beta_{ij2} (\psi_{ij} - a_j) \right) \\ & + \frac{\eta}{2} \sum_{i \in V} \sum_{j \in \mathcal{B}_i} (||\mathbf{W}_i - \boldsymbol{\rho}_{ij}||^2 + ||\boldsymbol{\rho}_{ij} - \mathbf{W}_j||^2) \\ & + \frac{\eta}{2} \sum_{i \in V} \sum_{j \in \mathcal{B}_i} (||\boldsymbol{\mu}_i - \boldsymbol{\phi}_{ij}||^2 + ||\boldsymbol{\phi}_{ij} - \boldsymbol{\mu}_j||^2) \\ & + \frac{\eta}{2} \sum_{i \in V} \sum_{j \in \mathcal{B}_i} ((a_i - \psi_{ij})^2 + (\psi_{ij} - a_j)^2) \end{aligned} \tag{2}$$

where $\Phi_i = \{f_{\mathbf{Z}_i}, \mathbf{W}_i, \boldsymbol{\mu}_i, a_i, \boldsymbol{\rho}_{ij}, \boldsymbol{\phi}_{ij}, \psi_{ij}; i \in V, j \in \mathcal{B}_i\}$ and $\{\boldsymbol{\lambda}_{ijk}\}, \{\boldsymbol{\gamma}_{ijk}\}, \{\beta_{ijk}\}$ with $k = 1, 2$ are the Lagrange multipliers. η is a positive scalar and $|| \cdot ||$ denotes the induced norm. We cyclically minimize $\mathcal{L}(\Phi_i)$ over its parameters, then follow a gradient ascent step over the Lagrange multipliers. The iterates, using t as iteration index, are

$$f_{\mathbf{Z}}(t+1) = \arg\min_{f_{\mathbf{Z}}} \mathcal{L}(f_{\mathbf{Z}}, \mathbf{W}_i(t), \boldsymbol{\mu}_i(t), a_i(t), \boldsymbol{\rho}_{ij}(t), \boldsymbol{\phi}_{ij}(t), \psi_{ij}(t)), \tag{3}$$
$$\mathbf{W}_i(t+1) = \arg\min_{\mathbf{W}_i} \mathcal{L}(f_{\mathbf{Z}}(t+1), \mathbf{W}_i, \boldsymbol{\mu}_i(t), a_i(t), \boldsymbol{\rho}_{ij}(t), \boldsymbol{\phi}_{ij}(t), \psi_{ij}(t)), \tag{4}$$
$$\boldsymbol{\mu}_i(t+1) = \arg\min_{\boldsymbol{\mu}_i} \mathcal{L}(f_{\mathbf{Z}}(t+1), \mathbf{W}_i(t+1), \boldsymbol{\mu}_i, a_i(t), \boldsymbol{\rho}_{ij}(t), \boldsymbol{\phi}_{ij}(t), \psi_{ij}(t)), \tag{5}$$
$$a_i(t+1) = \arg\min_{a_i} \mathcal{L}(f_{\mathbf{Z}}(t+1), \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i, \boldsymbol{\rho}_{ij}(t), \boldsymbol{\phi}_{ij}(t), \psi_{ij}(t)), \tag{6}$$
$$\begin{aligned} \boldsymbol{\rho}_{ij}(t+1) = \arg\min_{\boldsymbol{\rho}_{ij}} & \mathcal{L}(f_{\mathbf{Z}}(t+1), \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i(t+1), \\ & \boldsymbol{\rho}_{ij}, \boldsymbol{\phi}_{ij}(t), \psi_{ij}(t)), \end{aligned} \tag{7}$$
$$\begin{aligned} \boldsymbol{\phi}_{ij}(t+1) = \arg\min_{\boldsymbol{\phi}_{ij}} & \mathcal{L}(f_{\mathbf{Z}}(t+1), \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i(t+1), \\ & \boldsymbol{\rho}_{ij}(t+1), \boldsymbol{\phi}_{ij}, \psi_{ij}(t)), \end{aligned} \tag{8}$$
$$\begin{aligned} \psi_{ij}(t+1) = \arg\min_{\psi_{ij}} & \mathcal{L}(f_{\mathbf{Z}}(t+1), \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i(t+1), \\ & \boldsymbol{\rho}_{ij}(t+1), \boldsymbol{\phi}_{ij}(t+1), \psi_{ij}), \end{aligned} \tag{9}$$
$$\boldsymbol{\lambda}_{ij1}(t+1) = \boldsymbol{\lambda}_{ij1}(t) + \eta \left[\mathbf{W}_i(t+1) - \boldsymbol{\rho}_{ij}(t+1) \right], \forall i \in V, j \in \mathcal{B}_i, \tag{10}$$
$$\boldsymbol{\lambda}_{ij2}(t+1) = \boldsymbol{\lambda}_{ij2}(t) + \eta \left[\boldsymbol{\rho}_{ij}(t+1) - \mathbf{W}_j(t+1) \right], \forall i \in V, j \in \mathcal{B}_i, \tag{11}$$
$$\boldsymbol{\gamma}_{ij1}(t+1) = \boldsymbol{\gamma}_{ij1}(t) + \eta \left[\boldsymbol{\mu}_i(t+1) - \boldsymbol{\phi}_{ij}(t+1) \right], \forall i \in V, j \in \mathcal{B}_i, \tag{12}$$
$$\boldsymbol{\gamma}_{ij2}(t+1) = \boldsymbol{\gamma}_{ij2}(t) + \eta \left[\boldsymbol{\phi}_{ij}(t+1) - \boldsymbol{\mu}_j(t+1) \right], \forall i \in V, j \in \mathcal{B}_i, \tag{13}$$
$$\beta_{ij1}(t+1) = \beta_{ij1}(t) + \eta \left[a_i(t+1) - \psi_{ij}(t+1) \right], \forall i \in V, j \in \mathcal{B}_i, \tag{14}$$
$$\beta_{ij2}(t+1) = \beta_{ij2}(t) + \eta \left[\psi_{ij}(t+1) - a_j(t+1) \right], \forall i \in V, j \in \mathcal{B}_i. \tag{15}$$

[1]In the manuscript, we defined the optimization using marginal distribution to make it consistent with our general distributed probabilistic model. However, one can optimize the expected value of the completed log likelihood with respect to posterior as shown here.

Computing (3): Omitting t notations, the first term of (2)

$$-F(f_{\mathbf{Z}_i}, \mathbf{W}_i, \boldsymbol{\mu}_i, a_i) = \sum_{n=1}^{N_i} \Bigg\{ \frac{M}{2} \log(2\pi) + \frac{1}{2} tr \left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] \right] - \frac{D}{2} \log(2\pi a_i) + \frac{a_i}{2} ||\mathbf{x}_{in} - \boldsymbol{\mu}_i||^2 + \frac{a_i}{2} tr \left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}]\mathbf{W}_i^{\mathrm{T}}\mathbf{W}_i \right] - a_i \mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}} \mathbf{W}_i^{\mathrm{T}} (\mathbf{x}_{in} - \boldsymbol{\mu}_i) \Bigg\} \tag{16}$$

is the only term dependent on $f_{\mathbf{Z}_i}$. Thus, if all other parameters and multipliers are fixed at the t-th iteration, (3) can be computed using the expected values of the latent variables as we did in the E-step of the centralized setting. By using the posterior distribution of the centralized PPCA ((4) in the manuscript), we compute

$$\mathbb{E}[\mathbf{z}_{in}] = \mathbf{L}_i^{-1}\mathbf{W}_i^{\mathrm{T}}(\mathbf{x}_{in} - \boldsymbol{\mu}_i) \tag{17}$$

$$\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] = a_i^{-1}\mathbf{L}_i^{-1} + \mathbb{E}[\mathbf{z}_{in}]\mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}} \tag{18}$$

and we plug (17) and (18) into (16).

Computing (7)-(15): Auxiliary variables, $\boldsymbol{\rho}_{ij}, \boldsymbol{\phi}_{ij}, \psi_{ij}$ are independent from (16). Thus (7)-(9) are linear-quadratic optimization with respect to these variables and we can find a closed-form solution for these variables. For (9), this yields

$$\frac{\partial \mathcal{L}(\Phi_i)}{\partial \psi_{ij}} = \frac{\partial}{\partial \psi_{ij}} \Bigg\{ \sum_{i \in V} \sum_{j \in \mathcal{B}_i} (\beta_{ij1}(t)(a_i(t+1) - \psi_{ij}) + \beta_{ij2}(t)(\psi_{ij} - a_j(t+1))) + \frac{\eta}{2} \sum_{i \in V} \sum_{j \in \mathcal{B}_i} ((a_i(t+1) - \psi_{ij})^2 + (\psi_{ij} - a_j(t+1))^2) \Bigg\}$$

$$0 = \sum_{i \in V} \sum_{j \in \mathcal{B}_i} \Bigg\{ -(\beta_{ij1}(t) - \beta_{ij2}(t)) - \eta\,(a_i(t+1) + a_j(t+1)) + 2\eta\psi_{ij} \Bigg\}$$

Since the Lagrange multipliers, parameters and η are all zero or positive values, we get

$$\psi_{ij}(t+1) = \frac{1}{2\eta}(\beta_{ij1}(t) - \beta_{ij2}(t)) + \frac{1}{2}(a_i(t+1) + a_j(t+1)) \tag{19}$$

Using the same technique on (7) and (8), we get

$$\boldsymbol{\rho}_{ij}(t+1) = \frac{1}{2\eta}(\boldsymbol{\lambda}_{ij1}(t) - \boldsymbol{\lambda}_{ij2}(t)) + \frac{1}{2}(\mathbf{W}_i(t+1) + \mathbf{W}_j(t+1)), \tag{20}$$

$$\boldsymbol{\phi}_{ij}(t+1) = \frac{1}{2\eta}(\boldsymbol{\gamma}_{ij1}(t) - \boldsymbol{\gamma}_{ij2}(t)) + \frac{1}{2}(\boldsymbol{\mu}_i(t+1) + \boldsymbol{\mu}_j(t+1)) \tag{21}$$

respectively. Plugging (20) into (10) and (11), (21) into (12) and (13), (19) into (14) and (15), we get

$$\boldsymbol{\lambda}_{ij1}(t+1) = \frac{1}{2}(\boldsymbol{\lambda}_{ij1}(t) + \boldsymbol{\lambda}_{ij2}(t)) + \frac{\eta}{2}(\mathbf{W}_i(t+1) - \mathbf{W}_j(t+1)), \tag{22}$$

$$\boldsymbol{\lambda}_{ij2}(t+1) = \frac{1}{2}(\boldsymbol{\lambda}_{ij1}(t) + \boldsymbol{\lambda}_{ij2}(t)) + \frac{\eta}{2}(\mathbf{W}_i(t+1) - \mathbf{W}_j(t+1)), \tag{23}$$

$$\boldsymbol{\gamma}_{ij1}(t+1) = \frac{1}{2}(\boldsymbol{\gamma}_{ij1}(t) + \boldsymbol{\gamma}_{ij2}(t)) + \frac{\eta}{2}(\boldsymbol{\mu}_i(t+1) - \boldsymbol{\mu}_j(t+1)), \tag{24}$$

$$\boldsymbol{\gamma}_{ij2}(t+1) = \frac{1}{2}(\boldsymbol{\gamma}_{ij1}(t) + \boldsymbol{\gamma}_{ij2}(t)) + \frac{\eta}{2}(\boldsymbol{\mu}_i(t+1) - \boldsymbol{\mu}_j(t+1)), \tag{25}$$

$$\beta_{ij1}(t+1) = \frac{1}{2}(\beta_{ij1}(t) + \beta_{ij2}(t)) + \frac{\eta}{2}(a_i(t+1) - a_j(t+1)), \tag{26}$$

$$\beta_{ij2}(t+1) = \frac{1}{2}(\beta_{ij1}(t) + \beta_{ij2}(t)) + \frac{\eta}{2}(a_i(t+1) - a_j(t+1)) \tag{27}$$

where $\forall i \in V$ and $j \in \mathcal{B}_i$. As in Appendix B of [3], we observe that at iteration t, $\boldsymbol{\lambda}_{ij1}(t+1) = \boldsymbol{\lambda}_{ij2}(t+1), \boldsymbol{\gamma}_{ij1}(t+1) = \boldsymbol{\gamma}_{ij2}(t+1), \beta_{ij1}(t+1) = \beta_{ij2}(t+1)$ if we assume initial value of each Lagrange multiplier was set to zero. Thus, it suffices to find only one of the two. We define this one value as $\boldsymbol{\lambda}_{ij}(t) := \boldsymbol{\lambda}_{ij1}(t) = \boldsymbol{\lambda}_{ij2}(t)$, $\boldsymbol{\gamma}_{ij}(t) := \boldsymbol{\gamma}_{ij1}(t) = \boldsymbol{\gamma}_{ij2}(t)$ and $\beta_{ij}(t) := \beta_{ij1}(t) = \beta_{ij2}(t)$. Moreover, if we define

$$\boldsymbol{\lambda}_i(t) := \sum_{j \in \mathcal{B}_i} \boldsymbol{\lambda}_{ij}(t), \quad \boldsymbol{\gamma}_i(t) := \sum_{j \in \mathcal{B}_i} \boldsymbol{\gamma}_{ij}(t), \quad \beta_i(t) := \sum_{j \in \mathcal{B}_i} \beta_{ij}(t),$$

then (22)-(27) reduce to

$$\boldsymbol{\lambda}_i(t+1) = \boldsymbol{\lambda}_i(t) + \frac{\eta}{2} \sum_{j \in \mathcal{B}_i} \left\{ \mathbf{W}_i(t+1) - \mathbf{W}_j(t+1) \right\}, \tag{28}$$

$$\boldsymbol{\gamma}_i(t+1) = \boldsymbol{\gamma}_i(t) + \frac{\eta}{2} \sum_{j \in \mathcal{B}_i} \left\{ \boldsymbol{\mu}_i(t+1) - \boldsymbol{\mu}_j(t+1) \right\}, \tag{29}$$

$$\beta_i(t+1) = \beta_i(t) + \frac{\eta}{2} \sum_{j \in \mathcal{B}_i} \left\{ a_i(t+1) - a_j(t+1) \right\}. \tag{30}$$

Computing (6): We tackle a_i first. Omitting t temporarily for notational brevity, the derivate of $\mathcal{L}(\boldsymbol{\Phi}_i)$ with respect to a_i is

$$\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\Phi}_i)}{\partial a_i} = \frac{\partial}{\partial a_i} \Bigg[\sum_{n=1}^{N_i} \Bigg\{ & -\frac{D}{2} \log(2\pi a_i) + \frac{a_i}{2} ||\mathbf{x}_{in} - \boldsymbol{\mu}_i||^2 + \frac{a_i}{2} tr \left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] \mathbf{W}_i^{\mathrm{T}} \mathbf{W}_i \right] \\
& - a_i \mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}} \mathbf{W}_i^{\mathrm{T}} (\mathbf{x}_{in} - \boldsymbol{\mu}_i) \Bigg\} \\
& + \sum_{i \in V} \sum_{j \in \mathcal{B}_i} \left(\beta_{ij1}(a_i - \psi_{ij}) + \beta_{ij2}(\psi_{ij} - a_j) \right) \\
& + \frac{\eta}{2} \sum_{i \in V} \sum_{j \in \mathcal{B}_i} \left((a_i - \psi_{ij})^2 + (\psi_{ij} - a_j)^2 \right) \Bigg]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\Phi}_i)}{\partial a_i} &= \sum_{n=1}^{N_i} \Bigg\{ -\frac{D}{2} a_i^{-1} - \mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}} \mathbf{W}_i^{\mathrm{T}} (\mathbf{x}_{in} - \boldsymbol{\mu}_i) \\
&\qquad + \frac{1}{2} \left\{ ||\mathbf{x}_{in} - \boldsymbol{\mu}_i||^2 + tr \left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] \mathbf{W}_i^{\mathrm{T}} \mathbf{W}_i \right] \right\} \Bigg\} \\
&\quad + \sum_{j \in \mathcal{B}_i} (\beta_{ij1} - \beta_{ji2}) + \frac{\partial}{\partial a_i} \left[\eta \sum_{j \in B_i} \left\{ (a_i - \psi_{ij})^2 + (\psi_{ji} - a_i)^2 \right\} \right] \\
&= \mathcal{Q}_{a_i}(f_{\mathbf{z}}, \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i) \\
&\quad + \sum_{j \in \mathcal{B}_i} (\beta_{ij1} - \beta_{ji2}) + 2\eta \sum_{j \in B_i} (a_i - \psi_{ij}) \quad (\because \psi_{ij} = \psi_{ji})
\end{aligned} \tag{31}$$

where we temporarily suppressed the first summation term as $\mathcal{Q}_{a_i}(f_{\mathbf{z}}, \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i)$ for clearer presentation. Putting t back while plugging (19), the closed form solution of ψ_{ij}, into (31), we get

$$\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\Phi}_i)}{\partial a_i} &= \mathcal{Q}_{a_i}(f_{\mathbf{z}}, \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i) + \sum_{j \in \mathcal{B}_i} (\beta_{ij1}(t) - \beta_{ji2}(t)) \\
&\quad + 2\eta \sum_{j \in B_i} \left(a_i - \left\{ \frac{1}{2\eta} \left(\beta_{ij1}(t-1) - \beta_{ij2}(t-1) \right) + \frac{1}{2} \left(a_i(t) + a_j(t) \right) \right\} \right) \\
&= \mathcal{Q}_{a_i}(f_{\mathbf{z}}, \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i) + \sum_{j \in \mathcal{B}_i} (\beta_{ij1}(t) - \beta_{ji2}(t)) \\
&\quad + 2\eta a_i^2 |\mathcal{B}_i| - \eta \sum_{j \in B_i} \left(a_i(t) + a_j(t) \right)
\end{aligned} \tag{32}$$

since $\beta_{ij1}(t-1) = \beta_{ij2}(t-1)$ if we assume their initial values are zeros. Here again, we can further simplify (32) using the following property.

Proposition 1. *(Forero, et al. [3]) Given* $\beta_{ij}(0) = 0$*,* $\beta_{ij}(t) = -\beta_{ji}(t)$*.*

Proof. We have observed that $\beta_{ij1}(0) = \beta_{ij2}(0) = 0 \implies \beta_{ij1}(t) = \beta_{ij2}(t)$ by applying induction on (26) and (27). Thus, if we define $\beta_{ij}(t) := \beta_{ij1}(t) = \beta_{ij2}(t)$,

$$\begin{aligned}
\beta_{ij1}(t+1) &= \frac{1}{2}(\beta_{ij1}(t) + \beta_{ij2}(t)) + \frac{\eta}{2}(a_i(t+1) - a_j(t+1)) \\
&= \frac{1}{2}(\beta_{ij1}(t) + \beta_{ij1}(t)) + \frac{\eta}{2}(a_i(t+1) - a_j(t+1)) \\
&= \beta_{ij1}(t) + \frac{\eta}{2}(a_i(t+1) - a_j(t+1)) \\
\therefore \beta_{ij}(t+1) &= \beta_{ij}(t) + \frac{\eta}{2}(a_i(t+1) - a_j(t+1)).
\end{aligned}$$

Therefore, given $\beta_{ij}(0) = \beta_{ji}(0) = 0$, $\beta_{ij}(1) = -\beta_{ji}(1)$. If we continue iterations,

$$\begin{aligned}
\beta_{ij}(2) &= \beta_{ij}(1) + \frac{\eta}{2}(a_i(t+1) - a_j(t+1)) \\
\beta_{ji}(2) &= \beta_{ji}(1) + \frac{\eta}{2}(a_j(t+1) - a_i(t+1)) \\
&= -\beta_{ij}(1) - \frac{\eta}{2}(a_i(t+1) - a_j(t+1)) \\
&= -\beta_{ij}(2).
\end{aligned}$$

By induction, $\beta_{ij}(t) = -\beta_{ji}(t)$. □

Therefore, if we define $\beta_i(t) = \sum_{j \in \mathcal{B}_i} \beta_{ij}(t)$, (32) becomes

$$\frac{\partial \mathcal{L}(\mathbf{\Phi}_i)}{\partial a_i} = \mathcal{Q}_{a_i}(f_{\mathbf{z}}, \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i) + 2\beta_i(t) + 2\eta a_i |\mathcal{B}_i| - \eta \sum_{j \in B_i} (a_i(t) + a_j(t))$$

Setting this to zero while substituting $\mathcal{Q}_{a_i}(f_{\mathbf{z}}, \mathbf{W}_i(t+1), \boldsymbol{\mu}_i(t+1), a_i)$ back into its original form and simplifying for a_i, we get

$$\begin{aligned}
0 &= \sum_{n=1}^{N_i} \left\{ -\frac{D}{2}(a_i)^{-1} - \mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}} \mathbf{W}_i^{\mathrm{T}} (\mathbf{x}_{in} - \boldsymbol{\mu}_i) + \frac{1}{2} \left\{ ||\mathbf{x}_{in} - \boldsymbol{\mu}_i||^2 + tr\left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] \mathbf{W}_i^{\mathrm{T}} \mathbf{W}_i \right] \right\} \right\} \\
&\quad + 2\beta_i(t) + 2\eta a_i |\mathcal{B}_i| - \eta \sum_{j \in B_i} (a_i(t) + a_j(t)) \\
&= -\frac{N_i D}{2}(a_i)^{-1} - \sum_{n=1}^{N_i} \mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}} \mathbf{W}_i^{\mathrm{T}} (\mathbf{x}_{in} - \boldsymbol{\mu}_i) + 2\beta_i(t) + 2\eta a_i |\mathcal{B}_i| - \eta \sum_{j \in B_i} (a_i(t) + a_j(t)) \\
&\quad + \frac{1}{2} \sum_{n=1}^{N_i} \left\{ ||\mathbf{x}_{in} - \boldsymbol{\mu}_i||^2 + tr\left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] \mathbf{W}_i^{\mathrm{T}} \mathbf{W}_i \right] \right\} \\
&= -\frac{N_i D}{2} + 2\eta |\mathcal{B}_i| (a_i)^2 \\
&\quad + a_i \left\{ 2\beta_i(t) - \eta \sum_{j \in B_i} (a_i(t) + a_j(t)) \right. \\
&\quad \left. - \sum_{n=1}^{N_i} \mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}} \mathbf{W}_i^{\mathrm{T}} (\mathbf{x}_{in} - \boldsymbol{\mu}_i) + \frac{1}{2} \sum_{n=1}^{N_i} \left\{ ||\mathbf{x}_{in} - \boldsymbol{\mu}_i||^2 + tr\left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] \mathbf{W}_i^{\mathrm{T}} \mathbf{W}_i \right] \right\} \right\}. \qquad (33)
\end{aligned}$$

as we omitted t in $\mathbf{z}_{in}(t+1)$, $\mathbf{W}_i(t+1)$ and $\boldsymbol{\mu}_i(t+1)$ for notational brevity. This is a quadratic function of a_i for which we can find an algebraic solution.

Computing (5): We cannot simply use the closed form solution of the centralized PPCA, i.e. $\boldsymbol{\mu} = \bar{\boldsymbol{\mu}}$. We could use this result in the centralized setting but we cannot use it in the distributed model due to the auxiliary variable constraints. Again, we omit t temporarily for brevity.

$$
\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\Phi}_i)}{\partial \boldsymbol{\mu}_i} &= \frac{\partial}{\partial \boldsymbol{\mu}_i}\Bigg[\sum_{n=1}^{N_i}\Bigg\{-\frac{D}{2}\log(2\pi a_i) + \frac{a_i}{2}||\mathbf{x}_{in} - \boldsymbol{\mu}_i||^2 + \frac{a_i}{2} tr\left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}]\mathbf{W}_i^{\mathrm{T}}\mathbf{W}_i\right] \\
&\qquad - a_i\mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}}\mathbf{W}_i^{\mathrm{T}}(\mathbf{x}_{in} - \boldsymbol{\mu}_i)\Bigg\} \\
&\qquad + \sum_{i\in V}\sum_{j\in\mathcal{B}_i}\left(\boldsymbol{\gamma}_{ij1}^{\mathrm{T}}(\boldsymbol{\mu}_i - \boldsymbol{\phi}_{ij}) + \boldsymbol{\gamma}_{ij2}^{\mathrm{T}}(\boldsymbol{\phi}_{ij} - \boldsymbol{\mu}_j)\right) \\
&\qquad + \frac{\eta}{2}\sum_{i\in V}\sum_{j\in\mathcal{B}_i}\left(||\boldsymbol{\mu}_i - \boldsymbol{\phi}_{ij}||^2 + ||\boldsymbol{\phi}_{ij} - \boldsymbol{\mu}_j||^2\right)\Bigg] \\
&= \sum_{n=1}^{N_i}\Bigg\{-a_i(\mathbf{x}_{in} - \boldsymbol{\mu}_i) + a_i\mathbf{W}_i\mathbb{E}[\mathbf{z}_{in}]\Bigg\} \\
&\qquad + \sum_{j\in\mathcal{B}_i}(\boldsymbol{\gamma}_{ij1} - \boldsymbol{\gamma}_{ji2}) + \frac{\partial}{\partial \boldsymbol{\mu}_i}\Bigg[\eta\sum_{j\in B_i}\left\{||\boldsymbol{\mu}_i - \boldsymbol{\phi}_{ij}||^2 + ||\boldsymbol{\phi}_{ji} - \boldsymbol{\mu}_i||^2\right\}\Bigg] \\
&= -a_i\sum_{n=1}^{N_i}\mathbf{x}_{in} + N_i a_i\boldsymbol{\mu}_i + a_i\sum_{n=1}^{N_i}\mathbf{W}_i\mathbb{E}[\mathbf{z}_{in}] \\
&\qquad + \sum_{j\in\mathcal{B}_i}(\boldsymbol{\gamma}_{ij1} - \boldsymbol{\gamma}_{ji2}) + 2\eta\sum_{j\in B_i}(\boldsymbol{\mu}_i - \boldsymbol{\phi}_{ij})
\end{aligned} \tag{34}
$$

If we substitute the closed form solution of $\boldsymbol{\phi}_{ij}$, i.e. (21) into (34) while taking t notations back, we get

$$
\begin{aligned}
0 &= -a_i\sum_{n=1}^{N_i}\mathbf{x}_{in} + N_i a_i\boldsymbol{\mu}_i + a_i\sum_{n=1}^{N_i}\mathbf{W}_i\mathbb{E}[\mathbf{z}_{in}] + \sum_{j\in\mathcal{B}_i}(\boldsymbol{\gamma}_{ij1} - \boldsymbol{\gamma}_{ji2}) \\
&\quad + 2\eta\sum_{j\in B_i}\left(\boldsymbol{\mu}_i - \left\{\frac{1}{2\eta}(\boldsymbol{\gamma}_{ij1}(t-1) - \boldsymbol{\gamma}_{ij2}(t-1)) + \frac{1}{2}(\boldsymbol{\mu}_i(t) + \boldsymbol{\mu}_j(t))\right\}\right) \\
&= -a_i\sum_{n=1}^{N_i}\mathbf{x}_{in} + N_i a_i\boldsymbol{\mu}_i + a_i\sum_{n=1}^{N_i}\mathbf{W}_i\mathbb{E}[\mathbf{z}_{in}] + \sum_{j\in\mathcal{B}_i}(\boldsymbol{\gamma}_{ij1} - \boldsymbol{\gamma}_{ji2}) \\
&\quad + 2\eta\sum_{j\in B_i}\left(\boldsymbol{\mu}_i - \frac{1}{2}(\boldsymbol{\mu}_i(t) + \boldsymbol{\mu}_j(t))\right) \quad (\because \boldsymbol{\gamma}_{ij1}(t) = \boldsymbol{\gamma}_{ij2}(t)) \\
&= (N_i a_i + 2\eta|\mathcal{B}_i|)\,\boldsymbol{\mu}_i \\
&\quad - a_i\sum_{n=1}^{N_i}\mathbf{x}_{in} + a_i\sum_{n=1}^{N_i}\mathbf{W}_i\mathbb{E}[\mathbf{z}_{in}] + \sum_{j\in\mathcal{B}_i}(\boldsymbol{\gamma}_{ij1} - \boldsymbol{\gamma}_{ji2}) - \eta\sum_{j\in B_i}(\boldsymbol{\mu}_i(t) + \boldsymbol{\mu}_j(t)))
\end{aligned}
$$

$$
\begin{aligned}
\therefore \boldsymbol{\mu}_i(t+1) &= (N_i a_i + 2\eta|\mathcal{B}_i|)^{-1} \\
&\quad\cdot\left\{a_i\sum_{n=1}^{N_i}(\mathbf{x}_{in} - \mathbf{W}_i\mathbb{E}[\mathbf{z}_{in}]) - 2\boldsymbol{\gamma}_i(t) + \eta\sum_{j\in B_i}(\boldsymbol{\mu}_i(t) + \boldsymbol{\mu}_j(t)))\right\}
\end{aligned} \tag{35}
$$

if we define $\boldsymbol{\gamma}_i = \sum_{\mathcal{B}_i}\boldsymbol{\gamma}_{ij}$ and apply Proposition 1.

Computing (4): We can apply the same approach to find the expression for $\mathbf{W}_i$.

$$\frac{\partial \mathcal{L}(\mathbf{\Phi}_i)}{\partial \mathbf{W}_i} = \frac{\partial}{\partial \mathbf{W}_i}\Bigg[\sum_{n=1}^{N_i}\left\{\frac{a_i}{2} tr\left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}]\mathbf{W}_i^{\mathrm{T}}\mathbf{W}_i\right] - a_i\mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}}\mathbf{W}_i^{\mathrm{T}}(\mathbf{x}_{in} - \boldsymbol{\mu}_i)\right\}$$
$$+ \sum_{i\in V}\sum_{j\in\mathcal{B}_i}\left(\boldsymbol{\lambda}_{ij1}^{\mathrm{T}}(\mathbf{W}_i - \boldsymbol{\rho}_{ij}) + \boldsymbol{\lambda}_{ij2}^{\mathrm{T}}(\boldsymbol{\rho}_{ij} - \mathbf{W}_j)\right)$$
$$+ \frac{\eta}{2}\sum_{i\in V}\sum_{j\in\mathcal{B}_i}(||\mathbf{W}_i - \boldsymbol{\rho}_{ij}||^2 + ||\boldsymbol{\rho}_{ij} - \mathbf{W}_j||^2)\Bigg]$$

By using the derivative of the trace

$$\frac{\partial tr[ABA^{\mathrm{T}}]}{\partial A} = A(B + B^{\mathrm{T}}) \text{ and } \frac{\partial tr[A^{\mathrm{T}}B]}{\partial A} = B,$$

we get

$$\begin{aligned} 0 &= \frac{\partial}{\partial \mathbf{W}_i}\left[\sum_{n=1}^{N_i}\left\{\frac{a_i}{2} tr\left[\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}]\mathbf{W}_i^{\mathrm{T}}\mathbf{W}_i\right] - a_i\mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}}\mathbf{W}_i^{\mathrm{T}}(\mathbf{x}_{in} - \boldsymbol{\mu}_i)\right\}\right] \\ &\quad + \sum_{j\in\mathcal{B}_i}(\boldsymbol{\lambda}_{ij1} - \boldsymbol{\lambda}_{ji2}) + 2\eta\sum_{j\in B_i}\left(\mathbf{W}_i - \frac{1}{2}(\mathbf{W}_i(t) + \mathbf{W}_j(t))\right) \\ &= \sum_{n=1}^{N_i}\left\{a_i\mathbf{W}_i\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] - a_i(\mathbf{x}_{in} - \boldsymbol{\mu}_i)\mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}}\right\} \\ &\quad + \sum_{j\in\mathcal{B}_i}(\boldsymbol{\lambda}_{ij1} - \boldsymbol{\lambda}_{ji2}) + 2\eta\sum_{j\in B_i}\left(\mathbf{W}_i - \frac{1}{2}(\mathbf{W}_i(t) + \mathbf{W}_j(t))\right) \end{aligned} \tag{36}$$

$$\begin{aligned} \therefore \mathbf{W}_i(t+1) &= \left\{a_i\sum_{n=1}^{N_i}(\mathbf{x}_{in} - \boldsymbol{\mu}_i)\mathbb{E}[\mathbf{z}_{in}]^{\mathrm{T}} - 2\boldsymbol{\lambda}_i(t) + \eta\sum_{j\in B_i}((\mathbf{W}_i(t) + \mathbf{W}_j(t)))\right\} \\ &\quad \cdot\left(a_i\sum_{n=1}^{N_i}\mathbb{E}[\mathbf{z}_{in}\mathbf{z}_{in}^{\mathrm{T}}] + 2\eta|\mathcal{B}_i|\mathbf{I}\right)^{-1} \end{aligned} \tag{37}$$

as we define $\boldsymbol{\lambda}_i = \sum_{\mathcal{B}_i}\boldsymbol{\lambda}_{ij}$ and apply Proposition 1.

The overall algorithm for the Distributed Probabilistic Principal Component Analysis (D-PPCA) is summarized as Algorithm 1 in the manuscript.

3 Supplementary Video Description

3.1 Videos of Caltech Dataset

Under the directory named caltech, we provide videos of reconstructed 3D structures of all 3D objects we used from the Caltech dataset [4]. In the video, red crosses at the center of the scene are the estimated 3D structure. Green points are the input image points and magenta points are reprojected image points.

3.2 Videos of Hopkins Dataset

Under the directory named hopkins, we provide videos of reconstructed 3D structures of 3D objects we used from the Hopkins155 dataset [5]. We picked examples both from the calibration grids and traffic videos. The filename is the ID number of the object. Objects 109 and 132 represent the objects that yielded large subspace angles because their non rigid structure while others represent the ones with small angles as mentioned in our manuscript.

References

[1] Pedro A. Forero, Alfonso Cano, and Geogios B. Giannakis. Consensus-Based Distributed Support Vector Machines. *Journal of Machine Learning Research*, 11:1663–1707, 2010.

[2] Alexander Ilin and Tapani Raiko. Practical Approaches to Principal Component Analysis in the Presence of Missing Values. *Journal of Machine Learning Research*, 11:1957–2000, 2010.

[3] P.A. Forero, A. Cano, and G.B. Giannakis. Distributed Clustering Using Wireless Sensor Networks. *Selected Topics in Signal Processing, IEEE Journal of*, 5(4):707 –724, aug. 2011.

[4] Pierre Moreels and Pietro Perona. Evaluation of Features Detectors and Descriptors based on 3D Objects. *International Journal of Computer Vision*, 73(3):263–284, July 2007.

[5] Roberto Tron and Rene Vidal. A Benchmark for the Comparison of 3-D Motion Segmentation Algorithms. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2007.