Supplementary Material

In the supplementary materials, we include all the proofs for the proposed theorems and the detailed derivations for the formulation of the crowdsourcing task assignment problem.

1. Proof of Lemma 1

Firstly we show how to convert the projection problem (2) to a support set selection problem. For any vector $\mathbf{w}$, let vector $\mathbf{x} \in \{0,1\}^p$ indicate the nonzero positions of $\mathbf{w}$, then we can claim that

$$\|\mathbf{v} - \mathbf{v_x}\|^2 \leq \|\mathbf{v} - \mathbf{w}\|^2$$

where $\mathbf{v_x}$ is a vector having same dimension with $\mathbf{v}$, and it keeps elements at positions where $\mathbf{x}$ has "1", and fills zeros at positions where $\mathbf{x}$ has "0". In addition, vector $\mathbf{w} \in \Omega(\mathcal{G}, \mathbf{s})$ if and only if its support set indicator vector $\mathbf{x}$ satisfies $A\mathbf{x} \leq \mathbf{s}$, given A is defined in (4).

So the problem (2) can be converted to integer programming:

$$\begin{aligned} \min_{\mathbf{x} \in \{0,1\}^p} \quad & \|\mathbf{v} - \mathbf{v_x}\|^2 \\ \text{subject to} \quad & A\mathbf{x} \leq \mathbf{s} \end{aligned} \tag{9}$$

and the objective can be further simplified:

$$\begin{aligned} \|\mathbf{v} - \mathbf{v_x}\|^2 &= \langle \mathbf{v}^2, \mathbf{1} - \mathbf{x} \rangle \\ &= \|\mathbf{v}\|^2 - \langle \mathbf{v}^2, \mathbf{x} \rangle \end{aligned}$$

Since $\mathbf{v}$ is constant here, then the problem (9) is equivalent to the ILP (3), which means problem (2) is equivalent to ILP (3). Then we complete the proof.

2. Proof of Theorem 2

To prove Theorem 2, we use the concept of totally unimodular matrix.

Definition 2. Totally Unimodular (TU) Matrix. *An integer matrix is TU, if the determinant of any square submatrices*[3] *is in the set* $\{-1, 0, 1\}$.

Proposition 1. *If A is TU, then $A^\top$ is TU, and their concatenations with identity matrices (i.e. $[A, I], [A^\top, I]^\top$) are still TU.*

Proof. Since transposing matrix will not change the determinant, so it is obvious $A^\top$ is TU.
Then we prove stacking with identity matrix I preserves TU property. We prove it by induction. Firstly, we show that submatrix with size 1 always has determinant in $\{-1, 0, 1\}$, because any element from I is either 1 or 0. Now consider a submatrix with size k having determinant in $\{-1, 0, 1\}$, then submatrix with size $k+1$ will still have determinant in $\{-1, 0, 1\}$. To show this, we only need to prove that adding a new row/column from I will not change the determinant out of set $\{-1, 0, 1\}$. Since any row/column from I only has one nonzero element "1", we can eliminate other elements in the same position by subtracting a multiple of this row/column to other rows/columns. After that, we can remove this row and column, and the determinant can only change its sign. So we know that submatrix with size $k+1$ has determinant in $\{-1, 0, 1\}$ if submatrix with size k has determinant in $\{-1, 0, 1\}$. $\square$

Lemma 4. *If A is TU, $\mathbf{s}$ is an integer vector, then all vertices of the following polytope are integer points:*

$$\{\mathbf{x} \mid A\mathbf{x} \leq \mathbf{s}, \mathbf{x} \in [0,1]^p\} \tag{10}$$

[3] Submatrix here is a square smaller matrix obtained by removing certain rows and columns

Proof. We can have an equivalent form of this polytope:

$$\left\{\mathbf{x}\middle|\begin{bmatrix}A\\I\end{bmatrix}\mathbf{x}\leq\begin{bmatrix}\mathbf{s}\\\mathbf{1}\end{bmatrix},\mathbf{x}\geq\mathbf{0}\right\}\tag{11}$$

From Proposition 1, we know that matrix

$$\begin{bmatrix}A\\I\end{bmatrix}$$

is TU if A is TU, so this meets the case in Theorem 13.2 (see Papadimitriou & Steiglitz, 1982, chap 13). Then we complete the proof. □

Lemma 5. *If C is the matrix whose each row is the indicator vector of a group $g\in\mathcal{G}_1\cup\mathcal{G}_2$ in our TVCS model, then C is a TU matrix.*

Proof. Since C's rows are the indicator vectors of groups in $\mathcal{G}$, so $C_{ij}\in\{0,1\}$. From the definition 1, we know that there are at most two "1"s in each column. For the column which has two "1"s, the two corresponding groups are from $\mathcal{G}_1$ and $\mathcal{G}_2$ respectively. (Because we know that groups within $\mathcal{G}_1$ or $\mathcal{G}_2$ do not overlap.)

By this way, our matrix C meets the case in Theorem 13.3 (see Papadimitriou & Steiglitz, 1982, chap 13), and it is a TU matrix. □

Lemma 6. *Recall the matrix A in equation (4):*

$$A=\begin{bmatrix}\mathbf{1}^\top\\C\end{bmatrix}$$

If it is constructed from the TVCS model, then A is a TU matrix.

Proof. From Lemma 5, we know C is a TU matrix for any $\mathcal{G}$ of our TVCS model. In other words, any submatrix restricted in C has determinant -1, 0, or 1. Therefore, we only need to consider the submatrix of A' has overlaps with the first row $\mathbf{1}^\top$. There are only three possible forms of such submatrix S. We will show all of their determinants are in $\{-1,0,1\}$.

1) At least one column of S has a single "1", so it must appear in the first row $\mathbf{1}^\top$. By exchanging such column with the last column (which can only influence the sign of determinant), we can transform it with form:

$$\begin{bmatrix}\mathbf{1}^\top & 1\\\bar{C} & \mathbf{0}\end{bmatrix}$$

where $\bar{C}$ is any submatrix of C. From the matrix determinant property, we have $|S|=\pm|\bar{C}|\in\{-1,0,1\}$. Therefore, submatrix S in such form have determinants in $\{-1,0,1\}$.

2) All columns of S have three "1" elements (the last row has "1" for every column). For the rows which are from C, we can sum all the rows to a certain row (this will not change the determinant). By this way we transform S to the following form:

$$\begin{bmatrix}\mathbf{1}^\top\\\mathbf{2}^\top\\\bar{C}\end{bmatrix}$$

where $\bar{C}$ is a submatrix of C. In this case, S is not full rank, so its determinant is 0.

3) Each column in S contains at least two "1" elements, and there exists one column which has exactly two "1"s. By exchanging it to the last column, we can transform it to be:

$$\begin{bmatrix}\mathbf{1}^\top & 1\\\cdots & \mathbf{0}\\\bar{C}_i & 1\\\cdots & \mathbf{0}\end{bmatrix}$$

This means that one "1" is in the first row, and another is in the row from C, let us say it's $\bar{C}_i$. Since subtracting one row from another row will not change the determinant, we can let the first row subtract $\bar{C}_i$:

$$\begin{bmatrix} \mathbf{1}^\top & 1 \\ \cdots & \mathbf{0} \\ \bar{C}_i & 1 \\ \cdots & \mathbf{0} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{1}^\top - \bar{C}_i & 0 \\ \cdots & \mathbf{0} \\ \bar{C}_i & 1 \\ \cdots & \mathbf{0} \end{bmatrix}$$

Now the last column only has a single "1" in the i-th row. We can generate a smaller matrix S' by removing the i-th row and the last column, and if S' has determinant in $\{-1, 0, 1\}$, so does S.

$$\begin{bmatrix} \mathbf{1}^\top - \bar{C}_i & 0 \\ \cdots & \mathbf{0} \\ \bar{C}_i & 1 \\ \cdots & \mathbf{0} \end{bmatrix} \rightarrow S' = \begin{bmatrix} \mathbf{1}^\top - \bar{C}_i \\ \cdots \\ \cdots \end{bmatrix}$$

If $\bar{C}_i \neq \mathbf{0}^\top$, then there are some positions (including j-th column) in the first row will become zeros. For any column of matrix S' which has "0" element in the first row, there are two cases:

(a) This column only contains zeros, i.e. S' has zero determinant.

(b) This column contains a single "1", we can generate a smaller matrix S'' by removing this column and the row where this "1" sits. If S'' has determinant in $\{-1, 0, 1\}$, so does S'.

Notice that it is impossible for the case that such column has two "1"s. (Since each column can have at most three "1"s, and we already remove the "1" in the first row by subtraction, and discard another "1" by removing $\bar{C}_i$.) In the above case (b), we can repeat removing columns and rows until we get a degenerate matrix (has 0 determinant), or a matrix whose first row does not have zeros. For the later situation, we can process it by same procedures as the original matrix S unless it only has one row and one column, i.e. a matrix having single element "1" (has determinant 1).

If $\bar{C}_i = \mathbf{0}^\top$, we can also process it by same procedures as the original matrix S.

Therefore, we have proved that any square submatrix S in A' has determinant in $\{-1, 0, 1\}$, which means A' is TU, and hence A is TU. □

Applying Lemma 4 and Lemma 6, we complete the proof of Theorem 2.

3. Proof of Theorem 3

To prove Theorem 3, we start with several lemmas.

Lemma 7. *Formulate the feasibility problem as problem (7), let f be the objective of the formulation in Theorem 3. If $f^* = 0$, there exists a λ such that*

$$f(\mathbf{z}) - f^* \geq \frac{\lambda}{2} \|\mathbf{z} - P_{\mathbf{z}^*}(\mathbf{z})\|^2, \forall \mathbf{z} \in \Omega$$

where $P_{\mathbf{z}^*}(\mathbf{z})$ is the optimal point which is closet to $\mathbf{z}$.

Proof. Since $f^* = 0$, there exists at least an $\mathbf{z}^*$ such that

$$\begin{aligned} A\mathbf{z}^* - a &\leq 0 \\ B\mathbf{z}^* &= b \\ C\mathbf{z}^* &\leq c \end{aligned}$$

From Hoffman's Theorem (Hoffman, 2003), we know that there exists a $\lambda > 0$, such that

$$\frac{\lambda}{2} \|\mathbf{z} - P_{\mathbf{z}^*}(\mathbf{z})\|^2 \leq \|[A\mathbf{z} - a]_+\|^2 + \|B\mathbf{z} - b\|^2 + \|[C\mathbf{z} - c]_+\|^2$$

Therefore, we know for any $\mathbf{z}$ in Ω,

$$C\mathbf{z} \leq c$$

and

$$\frac{\lambda}{2}\|\mathbf{z} - P_{\mathbf{z}^*}(\mathbf{z})\|^2 \leq \|[A\mathbf{z} - a]_+\|^2 + \|B\mathbf{z} - b\|^2$$

□

Using the lemma above, we can prove the Theorem 3 now.

Proof. Denote by $\Delta_t := \|\mathbf{z}^t - P_{\mathbf{z}^*}(\mathbf{z}^t)\|^2$. We have

$$\begin{aligned}
\Delta_{t+1} &= \|\mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^{t+1})\|^2 \\
&\leq \|\mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^t)\|^2 \\
&\leq \|\mathbf{z}^t - P_{\mathbf{z}^*}(\mathbf{z}^t) + \mathbf{z}^{t+1} - \mathbf{z}^t\|^2 \\
&= \|\mathbf{z}^t - P_{\mathbf{z}^*}(\mathbf{z}^t)\|^2 + \|\mathbf{z}^{t+1} - \mathbf{z}_t\|^2 + 2\langle \mathbf{z}^{t+1} - \mathbf{z}^t, \mathbf{z}^t - P_{\mathbf{z}^*}(\mathbf{z}^t)\rangle \\
&= \Delta_t - \|\mathbf{z}^{t+1} - \mathbf{z}^t\|^2 + 2\langle \mathbf{z}^{t+1} - \mathbf{z}^t, \mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^t)\rangle \\
&\leq \Delta_t - \|\mathbf{z}^{t+1} - \mathbf{z}^t\|^2 - 2\gamma\langle \nabla f(\mathbf{z}^t), \mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^t)\rangle
\end{aligned}$$

Let $T = \langle \nabla f(\mathbf{z}^t), \mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^t)\rangle$. Then we have

$$\begin{aligned}
T &= \langle \nabla f(\mathbf{z}^t), \mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^t)\rangle \\
&= \langle \nabla f(\mathbf{z}^t), \mathbf{z}^t - P_{\mathbf{z}^*}(\mathbf{z}^t)\rangle + \langle \nabla f(\mathbf{z}^t), \mathbf{z}^{t+1} - \mathbf{z}^t\rangle \\
&\geq -f^* + f(\mathbf{z}^t) + f(\mathbf{z}^{t+1}) - f(\mathbf{z}^t) - \frac{L}{2}\|\mathbf{z}^{t+1} - \mathbf{z}^t\|^2 \\
&= f(\mathbf{z}^{t+1}) - f^* - \frac{L}{2}\|\mathbf{z}^{t+1} - \mathbf{z}^t\|^2
\end{aligned}$$

where L is the Lipschitz continuous gradient constant. Back to the original inequality, we have

$$\begin{aligned}
\|\mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^{t+1})\|^2 &\leq \Delta_t - \|\mathbf{z}^{t+1} - \mathbf{z}^t\|^2 - 2\gamma\langle \nabla f(\mathbf{z}^t), \mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^t)\rangle \\
&\leq \Delta_t - (1 - L\gamma)\|\mathbf{z}^{t+1} - \mathbf{z}^t\|^2 - 2\gamma(f(\mathbf{z}^{t+1}) - f^*) \\
&\leq \Delta_t - (1 - L\gamma)\|\mathbf{z}^{t+1} - \mathbf{z}^t\|^2 - 2\gamma\frac{\lambda}{2}\|\mathbf{z}^{t+1} - P_{\mathbf{z}^*}(\mathbf{z}^{t+1})\|^2
\end{aligned}$$

where the last inequality comes from Lemma 7.

Let $\gamma = \frac{1}{L}$ and we have

$$\begin{aligned}
(1 + \frac{\lambda}{L})\Delta_{t+1} &\leq \Delta_t \\
\Delta_{t+1} &\leq \frac{1}{1 + \frac{\lambda}{L}}\Delta_t
\end{aligned}$$

which shows the linear convergence rate $\alpha = \frac{1}{1+\frac{\lambda}{L}}$, then it completes the proof. □

4. Formulation of the Expected Accuracy in Crowdsourcing Task Assignment

In crowdsourcing task assignment problem, recall the objective function of problem (8):

$$\frac{1}{m}\sum_{j=1}^{m}\mathcal{E}_{\text{acc}}(Q_{\cdot,j}, X_{\cdot,j})$$

For the j-th task, $\mathcal{E}_{\text{acc}}(Q_{\cdot,j}, X_{\cdot,j})$ is defined in the following:

$$\begin{aligned}\mathcal{E}_{\text{acc}}(Q_{\cdot,j}, X_{\cdot,j}) =& \mathbb{P}(\hat{\mathbf{y}}_j = 1|\mathbf{y}_j = 1)\mathbb{P}(\mathbf{y}_j = 1) + \mathbb{P}(\hat{\mathbf{y}}_j = 0|\mathbf{y}_j = 0)\mathbb{P}(\mathbf{y}_j = 0) \\ =& \mathbb{E}_{\hat{Y}_{\Omega_j,j}|\mathbf{y}_j=1}[\mathbf{I}(\hat{\mathbf{y}}_j = 1)]\mathbb{P}(\mathbf{y}_j = 1) + \mathbb{E}_{\hat{Y}_{\Omega_j,j}|\mathbf{y}_j=0}[\mathbf{I}(\hat{\mathbf{y}}_j = 0)]\mathbb{P}(\mathbf{y}_j = 0)\end{aligned} \tag{12}$$

where $\mathbf{I}(\cdot)$ is the indicator function. We can further specify this formulation by considering the equivalent forms for $\hat{\mathbf{y}}_j = 1$ and $\hat{\mathbf{y}}_j = 0$:

$$\begin{aligned}&\hat{\mathbf{y}}_j = 1 \\ \Leftrightarrow&\mathbb{P}(\mathbf{y}_j = 1|\hat{Y}_{\Omega_j,j}) \geq \mathbb{P}(\mathbf{y}_j = 0|\hat{Y}_{\Omega_j,j}) \\ \Leftrightarrow&\frac{\mathbb{P}(\hat{Y}_{\Omega_j,j}|\mathbf{y}_j = 1)}{\mathbb{P}(\hat{Y}_{\Omega_j,j}|\mathbf{y}_j = 0)} \geq \frac{\mathbb{P}(\mathbf{y}_j = 0)}{\mathbb{P}(\mathbf{y}_j = 1)} \\ \Leftrightarrow&\prod_{i\in\Omega_j}\left(\frac{Q_{ij}}{1-Q_{ij}}\right)^{2\hat{Y}_{i,j}-1} \geq \frac{\mathbb{P}(\mathbf{y}_j = 0)}{\mathbb{P}(\mathbf{y}_j = 1)} \\ \Leftrightarrow&\sum_{i\in\Omega_j}(2\hat{Y}_{i,j}-1)\log\left(\frac{Q_{ij}}{1-Q_{ij}}\right) \geq \log\left(\frac{\mathbb{P}(\mathbf{y}_j = 0)}{\mathbb{P}(\mathbf{y}_j = 1)}\right) \\ \Leftrightarrow&\sum_{i=1}^{n}X_{ij}(2\hat{Y}_{i,j}-1)\log\left(\frac{Q_{ij}}{1-Q_{ij}}\right) - \log\left(\frac{\mathbb{P}(\mathbf{y}_j = 0)}{\mathbb{P}(\mathbf{y}_j = 1)}\right) \geq 0\end{aligned}$$

Similar derivation can be applied to $\hat{\mathbf{y}}_j = 0$ (change "$\geq$" to "$<$"). Here we substitute the indicator function $\mathbf{I}(t \geq 0)$ as sigmoid function $S(t) = \frac{1}{1+\exp(-t)}$ to obtain a smooth approximation. Denote by $Z_{ij} := (2\hat{Y}_{i,j} - 1)\log(Q_{ij}/(1 - Q_{ij}))$ and $\mathbf{r}_j := \log(\mathbb{P}(\mathbf{y}_j = 0)/\mathbb{P}(\mathbf{y}_j = 1))$ for short. The (smooth) objective turns out to be:

$$\frac{1}{m}\sum_{j=1}^{m}\left(\mathbb{E}_{\hat{Y}|\mathbf{y}_j=1}\left[S\left(\sum_{i=1}^{n}Z_{ij}X_{ij} - \mathbf{r}_j\right)\right]\mathbb{P}(\mathbf{y}_j = 1) + \mathbb{E}_{\hat{Y}|\mathbf{y}_j=0}\left[S\left(\mathbf{r}_j - \sum_{i=1}^{n}Z_{ij}X_{ij}\right)\right]\mathbb{P}(\mathbf{y}_j = 0)\right)$$

and its stochastic gradient is:

$$\begin{aligned}\mathbf{g}(X)_{\cdot,j} =& \frac{1}{m}\mathbb{P}(\mathbf{y}_j = 1)(1 - S(\sum_{i=1}^{n}Z_{ij}^{\mathbf{y}_j=1}X_{ij} - \mathbf{r}_j))S(\sum_{i=1}^{n}Z_{ij}^{\mathbf{y}_j=1}X_{ij} - \mathbf{r}_j)Z_{\cdot,j}^{\mathbf{y}_j=1} + \\ & \frac{1}{m}\mathbb{P}(\mathbf{y}_j = 0)(1 - S(\sum_{i=1}^{n}Z_{ij}^{\mathbf{y}_j=0}X_{ij} - \mathbf{r}_j))S(\sum_{i=1}^{n}Z_{ij}^{\mathbf{y}_j=0}X_{ij} - \mathbf{r}_j)Z_{\cdot,j}^{\mathbf{y}_j=0}\end{aligned}$$

where $Z^{\mathbf{y}_j=1}$ (or $Z^{\mathbf{y}_j=0}$) is generated by sampling $\hat{Y}$ given $\mathbf{y}_j = 1$ (or $\mathbf{y}_j = 0$).