Optimal Client Sampling With Partial Training

Ahmad Dabaja[1] and Rachid El-Azouzi[1]

[1]LIA, University of Avignon

Problem Formulation

We now consider the optimization problem:

$$\begin{aligned} \min_{\{p_k^t\}} \quad & \mathbb{E}_{A^t}\left[\left\|\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t-\sum_{k=1}^{K}n_k\,U_k^t\right\|^2\right] \\ \text{such that} \quad & 0\le p_k^t\le 1, \quad \forall\, k=1,\dots,N, \\ & \sum_{k=1}^{K} r_k\,p_k^t = m. \end{aligned}$$

where:

- p_k^t is the probability of sampling client k at round t,
- A^t is the set of clients selected at round t,
- n_k is the number of data samples for client k,
- U_k^t is the gradient transferred from client k ar round t,
- r_k is a ratio in $[0,1]$ representing the training proportion of client's k model.

Problem Simplification

Observe that

$$\mathbb{E}_{A^t}\left[\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t\right]=\sum_{k=1}^{K}n_k\,U_k^t.$$

Thus, the objective function

$$\mathbb{E}_{A^t}\left[\left\|\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t-\sum_{k=1}^{K}n_k\,U_k^t\right\|^2\right]$$

is the variance of the first term. By the variance identity,

$$\mathbb{E}_{A^t}\left[\left\|\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t-\sum_{k=1}^{K}n_k\,U_k^t\right\|^2\right]=\mathbb{E}_{A^t}\left[\left\|\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t\right\|^2\right]-\left\|\sum_{k=1}^{K}n_k\,U_k^t\right\|^2.$$

By expanding the squared norms into inner products and distributing the transpose over the sum, we obtain

$$\left\|\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t\right\|^2=\left(\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\sum_{j\in A^t}\frac{n_j}{p_j^t}\,U_j^t\right)=\left(\sum_{k\in A^t}\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\right)\left(\sum_{j\in A^t}\frac{n_j}{p_j^t}\,U_j^t\right).$$

Thus, the squared norm can be written as

$$\left\|\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t\right\|^2=\sum_{k\in A^t}\sum_{j\in A^t}\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_j}{p_j^t}\,U_j^t\right).$$

Next, we split the double summation into two terms: one for the case $k=j$ and one for $k\neq j$:

$$\sum_{k\in A^t}\sum_{j\in A^t}\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_j}{p_j^t}\,U_j^t\right)=\underbrace{\sum_{k\in A^t}\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_k}{p_k^t}\,U_k^t\right)}_{\text{terms with }k=j}+\underbrace{\sum_{k\in A^t}\sum_{\substack{j\in A^t\\ j\neq k}}\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_j}{p_j^t}\,U_j^t\right)}_{\text{terms with }k\neq j}.$$

Taking the expectation over A^t and noting that each client k is selected with probability p_k^t, we have

$$\begin{aligned}
&\mathbb{E}_{A^t}\left[\sum_{k\in A^t}\sum_{j\in A^t}\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_j}{p_j^t}\,U_j^t\right)\right]\\
&\qquad=\sum_{k=1}^{K}\mathbb{E}[I_k]\,\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_k}{p_k^t}\,U_k^t\right)+\sum_{k=1}^{K}\sum_{\substack{j=1\\ j\neq k}}^{K}\mathbb{E}[I_kI_j]\,\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_j}{p_j^t}\,U_j^t\right)\\
&\qquad=\sum_{k=1}^{K}p_k^t\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_k}{p_k^t}\,U_k^t\right)+\sum_{k=1}^{K}\sum_{\substack{j=1\\ j\neq k}}^{K}(p_k^tp_j^t)\left(\frac{n_k}{p_k^t}\,U_k^t\right)^T\left(\frac{n_j}{p_j^t}\,U_j^t\right)\\
&\qquad=\sum_{k=1}^{K}\frac{n_k^2}{p_k^t}\,\|U_k^t\|^2+\sum_{k=1}^{K}\sum_{\substack{j=1\\ j\neq k}}^{K}n_k\,K_j\left(U_k^t\right)^T\left(U_j^t\right).
\end{aligned}$$

Similarly, for the second term we have

$$\left\|\sum_{k=1}^{K}n_k\,U_k^t\right\|^2=\left(\sum_{k=1}^{K}n_k\,U_k^t\right)^T\left(\sum_{j=1}^{K}n_j\,U_j^t\right)=\sum_{k=1}^{K}\sum_{j=1}^{K}\left(n_k\,U_k^t\right)^T\left(n_j\,U_j^t\right).$$

Dividing the double summation into the diagonal and off-diagonal parts, we obtain

$$\sum_{k=1}^{K}\sum_{j=1}^{K}\left(n_k\,U_k^t\right)^T\left(n_j\,U_j^t\right)=\underbrace{\sum_{k=1}^{K}\left(n_k\,U_k^t\right)^T\left(n_k\,U_k^t\right)}_{k=j}+\underbrace{\sum_{k=1}^{K}\sum_{\substack{j=1\\ j\neq k}}^{K}\left(n_k\,U_k^t\right)^T\left(n_j\,U_j^t\right)}_{k\neq j}.$$

Thus, from the previous reduction, we have

$$\mathbb{E}_{A^t}\left[\left\|\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t\right\|^2\right]=\sum_{k=1}^{K}\frac{n_k^2}{p_k^t}\,\|U_k^t\|^2+\sum_{\substack{k,j=1\\ i\neq j}}^{K}n_k\,K_j\,\left(U_k^t\right)^T\left(U_j^t\right),$$

$$\left\|\sum_{k=1}^{K}n_k\,U_k^t\right\|^2=\sum_{k=1}^{K}n_k^2\,\|U_k^t\|^2+\sum_{\substack{k,j=1\\ i\neq j}}^{K}n_k\,K_j\,\left(U_k^t\right)^T\left(U_j^t\right).$$

By rewriting the variance in terms of the separated summations, the total variance is given by

$$\begin{aligned}
\mathrm{Var}&=\mathbb{E}_{A^t}\left[\left\|\sum_{k\in A^t}\frac{n_k}{p_k^t}\,U_k^t\right\|^2\right]-\left\|\sum_{k=1}^{K}n_k\,U_k^t\right\|^2\\
&=\left[\sum_{k=1}^{K}\frac{n_k^2}{p_k^t}\,\|U_k^t\|^2+\sum_{\substack{k,j=1\\ i\neq j}}^{K}n_k\,K_j\,\left(U_k^t\right)^T\left(U_j^t\right)\right]-\left[\sum_{k=1}^{K}n_k^2\,\|U_k^t\|^2+\sum_{\substack{k,j=1\\ i\neq j}}^{K}n_k\,K_j\,\left(U_k^t\right)^T\left(U_j^t\right)\right]\\
&=\sum_{k=1}^{K}\left(\frac{n_k^2}{p_k^t}-n_k^2\right)\|U_k^t\|^2\\
&=\sum_{k=1}^{K}n_k^2\,\|U_k^t\|^2\left(\frac{1}{p_k^t}-1\right).
\end{aligned}$$

Thus, the original optimization problem can be rewritten as

$$\min_{\{p_k^t\}}\sum_{k=1}^{K}n_k^2\,\|U_k^t\|^2\left(\frac{1}{p_k^t}-1\right)$$

subject to

$$0\leq p_k^t\leq 1,\quad\forall\,i,\quad\text{and}\quad\sum_{k=1}^{K}r_k\,p_k^t=m.$$

Variable Substitution

We perform the substitutions:

$$q_k=\frac{1}{p_k^t}-1\quad\Longrightarrow\quad p_k^t=\frac{1}{q_k+1},$$

and define

$$x_k = n_k^2 \, \|U_k^t\|^2.$$

Then, the objective becomes

$$\min_{\{q_k\}} \sum_{k=1}^{K} x_k \, q_k,$$

with the constraints:

$$q_k \geq 0, \quad \forall \, i,$$

and the modified equality constraint is now

$$\sum_{k=1}^{K} \frac{r_k}{q_k + 1} = m.$$

Lagrangian Formulation and KKT Conditions

Define the Lagrangian with Lagrange multiplier λ for the equality constraint and multipliers $\mu_k \geq 0$ for the non-negativity constraints on q_k:

$$\mathcal{L}(\{q_k\}, \lambda, \{\mu_k\}) = \sum_{k=1}^{K} x_k \, q_k + \lambda \left(\sum_{k=1}^{K} \frac{r_k}{q_k + 1} - m \right) - \sum_{k=1}^{K} \mu_k \, q_k.$$

The KKT conditions are:

1. **Stationarity:** For each $k = 1, \ldots, K$,

$$\frac{\partial \mathcal{L}}{\partial q_k} = x_k - \lambda \, \frac{r_k}{(q_k + 1)^2} - \mu_k = 0.$$

2. **Primal Feasibility:**

$$q_k \geq 0, \quad \forall \, i, \quad \text{and} \quad \sum_{k=1}^{K} \frac{r_k}{q_k + 1} = m.$$

3. **Dual Feasibility:**

$$\mu_k \geq 0, \quad \forall \, i.$$

4. **Complementary Slackness:**

$$\mu_k \, q_k = 0, \quad \forall \, i.$$

These conditions will be used to solve for the optimal $\{q_k\}$, from which the original probabilities are recovered via

$$p_k^t = \frac{1}{q_k + 1}.$$

Solving the KKT Conditions

The stationarity condition is

$$x_k - \frac{\lambda\, r_k}{(q_k+1)^2} - \mu_k = 0.$$

Isolating μ_k yields

$$x_k - \frac{\lambda\, r_k}{(q_k+1)^2} = \mu_k.$$

We now consider two cases:

Case 1: $q_k > 0$**.** Then complementary slackness implies $\mu_k = 0$, so that

$$x_k = \frac{\lambda\, r_k}{(q_k+1)^2}.$$

Solving for q_k gives

$$q_k = \sqrt{\frac{\lambda\, r_k}{x_k}} - 1.$$

Case 2: $q_k = 0$**.** In this case, the stationarity condition becomes

$$x_k - \lambda\, r_k = \mu_k \geq 0,$$

which implies

$$x_k \geq \lambda\, r_k.$$

Thus, we have:

- For $q_k > 0$: $q_k = \sqrt{\frac{\lambda\, r_k}{x_k}} - 1$.
- For $q_k = 0$: $x_k \geq \lambda\, r_k$.

Determining λ via the Equality Constraint

The modified equality constraint is

$$\sum_{k=1}^{K} \frac{r_k}{q_k+1} = m.$$

Define the index set

$$\mathcal{S} = \{k \in \{1, \dots, K\} : x_k < \lambda\, r_k \},$$

so that for $k \in \mathcal{S}$ (by Case 1) we have

$$\frac{1}{q_k+1} = \sqrt{\frac{x_k}{\lambda\, r_k}},$$

and for indices $k \in \mathcal{S}^c$ (where $q_k = 0$ by Case 2) we have

$$\frac{1}{q_k + 1} = 1.$$

Thus, the equality constraint becomes

$$\sum_{k \in \mathcal{S}} r_k \sqrt{\frac{x_k}{\lambda\, r_k}} + \sum_{k \in \mathcal{S}^c} r_k = m.$$

note that

$$\sum_{k \in \mathcal{S}^c} r_k = \sum_{k=1}^{K} r_k - \sum_{k \in \mathcal{S}} r_k.$$

Thus, the equality constraint becomes

$$\frac{1}{\sqrt{\lambda}} \sum_{k \in \mathcal{S}} \sqrt{r_k\, x_k} + \left(\sum_{k=1}^{K} r_k - \sum_{k \in \mathcal{S}} r_k \right) = m.$$

Isolating λ yields

$$\lambda = \left(\frac{\sum_{k \in \mathcal{S}} \sqrt{r_k\, x_k}}{m - \sum_{k=1}^{K} r_k + \sum_{k \in \mathcal{S}} r_k} \right)^2.$$

Finding the Explicit Solution for q_k

For $k \in \mathcal{S}$ (i.e. where $q_k > 0$), substituting the expression for λ into

$$q_k = \sqrt{\frac{\lambda\, r_k}{x_k}} - 1,$$

yields

$$q_k = \frac{\sqrt{r_k}}{m - \sum_{j=1}^{K} r_j + \sum_{j \in \mathcal{S}} r_j} \cdot \frac{\sum_{j \in \mathcal{S}} \sqrt{r_j\, x_j}}{\sqrt{x_k}} - 1.$$

For $k \in \mathcal{S}^c$, we set $q_k = 0$.

Thus, the piecewise definition is:

$$q_k = \begin{cases} \dfrac{\sqrt{r_k}}{m - \sum_{j=1}^{K} r_j + \sum_{j \in \mathcal{S}} r_j} \cdot \dfrac{\sum_{j \in \mathcal{S}} \sqrt{r_j\, x_j}}{\sqrt{x_k}} - 1, & \text{if } k \in \mathcal{S}, \\ 0, & \text{if } k \in \mathcal{S}^c. \end{cases}$$

Returning to the Original Variable p_k^t

Recall that

$$p_k^t = \frac{1}{q_k + 1}.$$

Thus, for $k \in \mathcal{S}$ (where $q_k > 0$) we have

$$p_k^t = \frac{m - \sum_{j=1}^{K} r_j + \sum_{j \in \mathcal{S}} r_j}{\sqrt{r_k}} \cdot \frac{\sqrt{x_k}}{\sum_{j \in \mathcal{S}} \sqrt{r_j\, x_j}},$$

and for $k \in \mathcal{S}^c$ (where $q_k = 0$) we obtain

$$p_k^t = 1.$$

Substituting the expression and recalling that $x_k = n_k^2 \, \|U_k^t\|^2$, the final solution is

$$p_k^t = \begin{cases} \dfrac{m - \sum_{j=1}^{K} r_j + \sum_{j \in \mathcal{S}} r_j}{\sqrt{r_k}} \cdot \dfrac{n_k \, \|U_k^t\|}{\sum_{j \in \mathcal{S}} \sqrt{r_j}\, K_j \, \|U_j^t\|}, & \text{if } k \in \mathcal{S}, \\ 1, & \text{if } k \in \mathcal{S}^c. \end{cases}$$

Where

$$\mathcal{S} = \left\{ k \in \{1, \ldots, K\} : \sqrt{r_k}\, K_k \, \|U_k^t\| < \frac{\sum_{j \in \mathcal{S}} \sqrt{r_j}\, K_j \, \|U_j^t\|}{m - \sum_{j=1}^{K} r_j + \sum_{j \in \mathcal{S}} r_j} \right\}.$$

Determining the Index Set $\mathcal{S}$

Since for indices in $\mathcal{S}^c$ we have $p_k^t = 1$, the equality constraint

$$\sum_{k=1}^{K} r_k \, p_k^t = m$$

becomes

$$\sum_{k \in \mathcal{S}} r_k \, p_k^t + \sum_{k \in \mathcal{S}^c} r_k = m.$$

and by splitting the second term, we got:

$$\sum_{k \in \mathcal{S}} r_k \, p_k^t + \sum_{k=1}^{K} r_k - \sum_{k \in \mathcal{S}} r_k = m.$$

equivalent to

$$\sum_{k \in \mathcal{S}} r_k \, \left(1 - p_k^t\right) = \sum_{k=1}^{K} r_k - m.$$

Thus, the indices in $\mathcal{S}$ must satisfy

$$\sum_{k \in \mathcal{S}} r_k > \sum_{k=1}^{K} r_k - m.$$

That is, the total r-mass in $\mathcal{S}$ must exceed $\sum_{k=1}^{K} r_k - m$.

The procedure to determine $\mathcal{S}$ and its complement $\mathcal{S}^c$ is as follows:

1. **Initialization:** Order the indices in increasing order of $\sqrt{r_k}\, K_k^t\, \|U_k^t\|$. Begin by selecting the indices corresponding to the smallest values until the cumulative sum $\sum_{k\in\mathcal{S}} r_k$ exceeds $\sum_{k=1}^{K} r_k - m$.

2. **Iteration:** Let $\mathcal{S}^c = \{1,\ldots,N\} \setminus \mathcal{S}$. For the next candidate index k in $\mathcal{S}^c$ (with the next smallest $\sqrt{r_k}\, d_k^t\, \|U_k^t\|$), check whether adding k to $\mathcal{S}$ maintains the inequality

$$\sqrt{r_k}\, d_k^t\, \|U_k^t\| < \frac{\sum_{j\in\mathcal{S}\cup\{k\}} \sqrt{r_j}\, K_j^t\, \|U_j^t\|}{m - \sum_{j=1}^{K} r_j + \sum_{j\in\mathcal{S}} r_j}$$

3. **Update:**

   - If the inequality holds, update $\mathcal{S} \leftarrow \mathcal{S} \cup \{k\}$ and $\mathcal{S}^c \leftarrow \mathcal{S}^c \setminus \{k\}$, and then repeat the iteration.
   - If the inequality fails for the candidate k, NO further indices can be added to $\mathcal{S}$; then $\mathcal{S}$ and $\mathcal{S}^c$ are finalized.