QUIC-FL: Quick Unbiased Compression for Federated Learning

Ran Ben Basat, Shay Vargaftik, Amit Portnoy, Gil Einziger, Yaniv Ben-Itzhak, Michael Mitzenmacher

Gautam Singh

Indian Institute of Technology Hyderabad
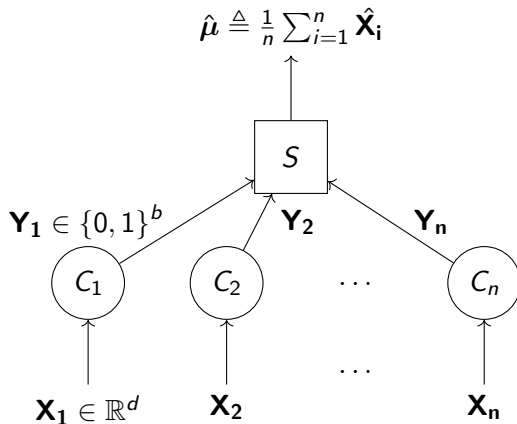
November 22, 2023

The DME Problem

Figure 1: Illustration of the DME Problem. Here, $\hat{\mathbf{X}}_i$ denotes the server estimate for $\mathbf{X}_i$.

vNMSE and NMSE

Definition 1 (vNMSE)

The *vector Normalized Mean Square Error* of **x** is defined as

$$vNMSE \triangleq \frac{\mathbb{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2\right]}{\|\mathbf{x}\|_2^2}. \tag{1}$$

Definition 2 (NMSE)

The *Normalized Mean Square Error* in the case of the DME problem is defined as

$$NMSE \triangleq \frac{\mathbb{E}\left[\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2^2\right]}{\frac{1}{n}\sum_{i=1}^n \|\mathbf{x_i}\|_2^2} = \frac{\mathbb{E}\left[\|\hat{\boldsymbol{\mu}} - \frac{1}{n}\sum_{i=1}^n \mathbf{x_i}\|_2^2\right]}{\frac{1}{n}\sum_{i=1}^n \|\mathbf{x_i}\|_2^2}. \tag{2}$$

An Important Lemma

Lemma 3

For unbiased algorithms and independent estimates, we have

$$NMSE = \frac{vNMSE}{n} \qquad (3)$$

From the above lemma, we need the *vNMSE* to be a constant so that the *NMSE* is $\mathcal{O}\left(1/n\right)$ for QUIC-FL.

Randomized Hadamard Transform

Definition 4 (Walsh-Hadamard Matrix)

The Walsh-Hadamard matrix $\mathbf{H}_{2^k}$ with $\mathbf{H}_1 = \begin{pmatrix} 1 \end{pmatrix}$ is recursively defined as

$$\mathbf{H}_{2^k} = \begin{pmatrix} \mathbf{H}_{2^{k-1}} & \mathbf{H}_{2^{k-1}} \\ \mathbf{H}_{2^{k-1}} & -\mathbf{H}_{2^{k-1}} \end{pmatrix}. \tag{4}$$

Definition 5 (Randomized Hadamard Transform)

Let $\mathbf{H} \in \{-1, +1\}^{d \times d}$ be a Walsh-Hadamard Matrix and $\mathbf{D}$ be a diagonal matrix with uniform iid Rademacher entries *i.e.*, entries that are ± 1 with equal probability. Then, the *randomized Hadamard transform* (RHT) of $\mathbf{x} \in \mathbb{R}^d$ is given by

$$\mathcal{R}_{\mathbf{H}} (\mathbf{x}) \triangleq \left(\frac{1}{\sqrt{d}} \mathbf{H} \mathbf{D} \right) \mathbf{x}. \tag{5}$$

Properties of RHT

Definition 6

Inverse RHT The *inverse randomized Hadamard transform* of **x** is given by

$$\mathcal{R}_{\mathbf{H}}^{-1}(x) \triangleq \left(\frac{1}{\sqrt{d}}\mathbf{D}\mathbf{H}\right)\mathbf{x}. \tag{6}$$

Lemma 7

Define $F_{i,d}(x)$ to be the CDF of $\frac{1}{\sigma}\mathcal{R}_{\mathbf{H}}(\mathbf{x})_i$ and $\Phi(x)$ the CDF of the standard normal distribution. Then, as $d \to \infty$, $\forall\ 1 \le j \le d$,

$$F_{i,d}(x_j) \xrightarrow{\mathrm{d}} \Phi(x_j). \tag{7}$$

Design Goals of QUIC-FL

1. Less computational complexity compared to other methods, at *both* client and server.

2. Lesser amount of shared global as well as private randomness.

3. Same (asymptotic) NMSE as other methods: $\mathcal{O}\left(\frac{1}{n}\right)$.

BSQ Algorithm

1. Uses a parameter $p \in (0, 1]$, based on which a threshold $[-t_p, t_p]$ is chosen such that at most dp values lie outside this interval.

2. **x** separated into two parts: *large* values that lie outside this interval and *small* values that lie inside this interval.

3. To encode **x**,
   1. *Large* values are sent exactly, along with their indices. Costs $\log \binom{d}{dp} \approx dp \log \left(\frac{1}{p} \right)$ bits. Can use delta encoding or no encoding at cost of $dp \lceil \log d \rceil$ bits (Authors assume $p \log d << 1$).
   2. *Small* values are stochastically quantized and sent using b bits per coordinate. *This implies that BSQ is unbiased.*

4. To decode and compute $\hat{\mathbf{x}}$, decode the *large* values along with their indices. Then, estimate the quantized *small* values.

Main Result of BSQ

Lemma 8

BSQ, without further assumptions, admits a worst-case NMSE of $\frac{1}{np\left(2^b-1\right)^2}$ with $\mathcal{O}\left(d\right)$ and $\mathcal{O}\left(nd\right)$ complexity for encoding and decoding respectively.

Arbitrarily Large NMSE

Notice that the constant in the NMSE can be made arbitrarily large. For example, $p = 2^{-10}$, $b = 1$.

What QUIC-FL Does

The **random rotation preprocessing** applied by QUIC-FL guarantees smaller constant values for small p.

Distribution-Aware Unbiased Quantization
Key Idea

We have $\forall\ 1 \leq i \leq d$, where $\mathbf{R}$ denotes a random rotation matrix [1],

$$\frac{\sqrt{d}}{\|\mathbf{x}\|_2} (\mathbf{Rx})_i \sim \mathcal{N}(0, 1). \tag{8}$$

Thus, we only need to optimize the reconstruction values for when the data is sampled from the *standard normal distribution*. This will yield a *near-optimal* quantization for the $\mathbf{x}_i$.

Notice that this is possible since

1. We know the distribution of the transformed vector by (8).
2. We know the range $[-t_p, t_p]$ of $\mathbf{x}_i$.

Distribution-Aware Unbiased Quantization

Definitions

Notation	Definition
p	BSQ parameter
t_p	BSQ threshold
b	Number of bits per quantized value
$\mathcal{M}$	Message space $\{0, 1, \ldots, 2^b - 1\}$
$R(x)$	Reconstruction point associated with x at the receiver
$\mathcal{Q}_{b,p}$	Set of all reconstruction points $\{R(x) \mid x \in \mathcal{M}\}$
$\mathcal{I}_m$	$\{0, 1, \ldots, m-1\}$
$\mathcal{A}_{p,m}(i)$	$\Pr\left(Z \leq \mathcal{A}_{p,m}(i) \mid Z \in [-t_p, t_p], \ Z \sim \mathcal{N}(0,1)\right) = \frac{i}{m-1}$
$\mathcal{A}_{p,m}$	Right endpoints of quantiles $\{\mathcal{A}_{p,m}(i) \mid i \in \mathcal{I}_m\}$
$S(z, x)$	$\Pr(z \text{ is quantized to } R(x))$
$S'(i, x)$	$S(\mathcal{A}_{p,m}(i), x)$

Table 1: Notation for Distribution-Aware Unbiased Quantization.

Distribution-Aware Unbiased Quantization

Optimization Problem

We want to minimize the MSE of this scheme. The optimization problem becomes

$$\min_{S,R} \int_{-t_p}^{t_p} \sum_{x \in \mathcal{M}} S(z,x)(z - R(x))^2 e^{-\frac{z^2}{2}} dz \text{ s.t.} \tag{9}$$

$$(\text{Probability}) \sum_{x \in \mathcal{M}} S(z,x) = 1 \tag{10}$$

$$(\text{Unbiasedness}) \sum_{x \in \mathcal{M}} S(z,x) R(x) = z. \tag{11}$$

This is not easy to solve!

Distribution-Aware Unbiased Quantization

Discretization

Can now use solvers like APOPT and IPOPT to solve for $\mathcal{Q}_{b,p}$.

$$\min_{S',R} \sum_{\substack{i \in \mathcal{I}_m \\ x \in \mathcal{M}}} S'(i,x)\left(\mathcal{A}_{p,m}(i) - R(x)\right)^2 \text{ s.t.} \tag{12}$$

$$\text{(Probability)} \sum_{x \in \mathcal{M}} S'(i,x) = 1 \tag{13}$$

$$\text{(Unbiasedness)} \sum_{x \in \mathcal{M}} S'(i,x) R(x) = \mathcal{A}_{p,m}(i). \tag{14}$$

The reconstruction points $\mathcal{Q}_{b,p}$ obtained correspond to a *near-optimal* solution.

1. Convergence of randomly rotated coordinates to normal distribution, from (8).
2. As $m \to \infty$, the discrete problem converges to the continuous version.

QUIC-FL Client

Algorithm 1 QUIC-FL Client

Require: BSQ parameter p and threshold t_p, bits per symbol b, global randomness T, precomputed reconstruction values $\mathcal{Q}_{b,p}$, stochastic quantization algorithm Q.

1: **procedure** $\mathrm{QUICFLCLIENT}(\mathbf{x}_c)$
2:      $\mathbf{Z}_c \leftarrow \frac{\sqrt{d}}{\|\mathbf{x}_c\|_2} T(\mathbf{x}_c)$
3:      $\mathbf{U}_c \leftarrow \{Z_{ci} \mid |Z_{ci}| > t_p\}$
4:      $\mathbf{I}_c \leftarrow \{i \mid |Z_{ci}| > t_p\}$
5:      $\mathbf{V}_c \leftarrow \{Z_{ci} \mid |Z_{ci}| \leq t_p\}$
6:      $\mathbf{X}_c \leftarrow Q(\mathbf{V}_c)$ using $\mathcal{Q}_{b,p}$
7:      Send $(\|\mathbf{x}_c\|_2, \mathbf{U}_c, \mathbf{I}_c, \mathbf{X}_c)$ to server

---

QUIC-FL Server

Algorithm 2 QUIC-FL Server

Require: BSQ parameter p and threshold t_p, bits per symbol b, global randomness T, precomputed reconstruction values $\mathcal{Q}_{b,p}$, stochastic quantization algorithm Q, subroutine MERGE to reconstruct quantized vector given its BSQ parts.

1: **procedure** QUICFLSERVER
2:     **for all** $1 \leq c \leq n$ **do**
3:         $\hat{\mathbf{V}}_c \leftarrow \{\mathcal{Q}_{b,p}(x) \text{ for } x \text{ in } \mathbf{X}_c\}$
4:         $\hat{\mathbf{Z}}_c \leftarrow \text{MERGE}\left(\hat{\mathbf{V}}_c, \mathbf{U}_c, \mathbf{I}_c\right)$
5:     $\hat{\boldsymbol{\mu}}_Z \leftarrow \frac{1}{n} \sum_{c=1}^{n} \frac{\|\mathbf{x}_c\|_2}{\sqrt{d}} \hat{\mathbf{Z}}_c$
6:     $\hat{\boldsymbol{\mu}} \leftarrow T^{-1}(\hat{\boldsymbol{\mu}}_Z)$

NMSE of QUIC-FL

Theorem 9

Let $Z \sim \mathcal{N}(0,1)$ and $\hat{Z}$ be its estimate using the distribution-aware unbiased quantization scheme. Then, for any number of clients n and any set of input vectors $\{\mathbf{x}_i \in \mathbb{R}^d \mid i \in \{0, 1, \ldots, n-1\}\}$, the NMSE of QUIC-FL is given by

$$NMSE = \frac{1}{n}\mathbb{E}\left[\left(Z - \hat{Z}\right)^2\right] + \mathcal{O}\left(\frac{1}{n}\sqrt{\frac{\log d}{d}}\right) \qquad (15)$$

When $d \to \infty$, the NMSE respects $\mathcal{O}\left(\frac{1}{n}\right)$.

Optimizations

Client-Specific Shared Randomness

Some amount of random bits are shared between each client and the server. This can reduce the MSE at the cost of generating more random bits.

Define the following quantities.

1. l, the number of random bits used.

2. $\mathcal{H}_l \triangleq \left\{0, 1, \ldots, 2^l - 1\right\}$, the set of all possible random bitstrings of length l.

3. $H \sim \mathcal{U}\left[\mathcal{H}_l\right]$, the random variable representing the shared randomness, selected uniformly from $\mathcal{H}_l$.

4. $S'(h, i, x) \triangleq \Pr\left(\mathcal{A}_{p,m}(i) \text{ is quantized to } R(H, x) \mid H = h\right)$.

5. $R(h, x)$, the reconstruction point for $x \in \mathcal{M}$ given shared randomness h.

Optimizations

Client-Specific Shared Randomness

The new optimization problem becomes the following.

$$\min_{S',R} \sum_{\substack{h \in \mathcal{H}_l \\ i \in \mathcal{I}_m \\ x \in \mathcal{M}}} S'\left(h, i, x\right) \left(\mathcal{A}_{p,m}\left(i\right) - R\left(h, x\right)\right)^2 \text{ s.t.} \tag{16}$$

$$\text{(Probability)} \sum_{x \in \mathcal{M}} S'\left(h, i, x\right) = 1 \tag{17}$$

$$\text{(Unbiasedness)} \frac{1}{2^l} \sum_{\substack{h \in \mathcal{H}_l \\ x \in \mathcal{M}}} S'\left(h, i, x\right) R\left(h, x\right) = \mathcal{A}_{p,m}\left(i\right). \tag{18}$$

Optimizations

Using the RHT

1. Accelerates both client and server algorithms.
2. Same asymptotic guarantees but with larger constants.
   1. For the RHT, $\mathbb{E}\left[\left(Z - \hat{Z}\right)^2\right] = \mathcal{O}(p)$.
   2. Can compensate by slight increase in b.
3. Fastest complexity in comparison to other state-of-the-art methods, at both client and server.

Complexity Analysis

Algorithm	Enc. Complexity	Dec. Complexity	NMSE
QSGD	$\mathcal{O}(d)$	$\mathcal{O}(nd)$	$\mathcal{O}(d/n)$
Hadamard	$\mathcal{O}(d\log d)$	$\mathcal{O}(nd + d\log d)$	$\mathcal{O}(\log d/n)$
Kashin	$\mathcal{O}(d\log d\log(nd))$	$\mathcal{O}(nd + d\log d)$	$\mathcal{O}(1/n)$
EDEN	$\mathcal{O}(d\log d)$	$\mathcal{O}(nd\log d)$	$\mathcal{O}(1/n)$
QUIC-FL	$\mathcal{O}(d\log d)$	$\mathcal{O}(nd + d\log d)$	$\mathcal{O}(1/n)$

Table 2: Complexities and NMSE of Unbiased DME Algorithms.
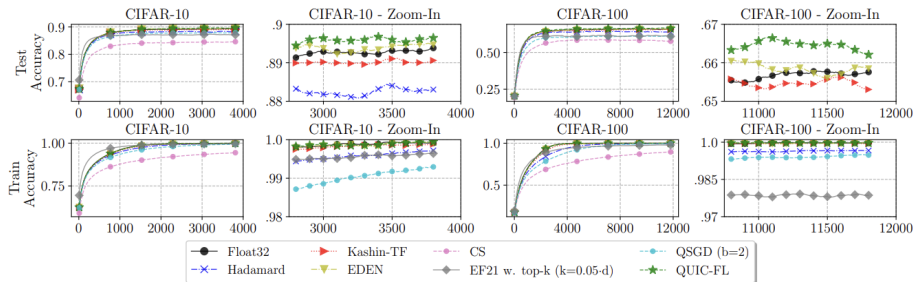
Accuracy

Figure 2: Cross-device Federated Learning with Various Methods of Unbiased Quantization. [2]

Further Improvements in QUIC-FL

1. (Open problem) Using more than one RHT can lead to better bounds?
2. QUIC-FL for cases using limited bits b and dimensions d.
3. Space complexity per client? Better compression ratio?
   1. Depends on d, b and p.
4. Can the NMSE be lowered using a client-server feedback system?
5. Security concerns
   1. Does the random transformation T act as a "secret key"?
   2. What security guarantees does QUIC-FL have against various adversarial attacks?

References

📄 S. Vargaftik, R. B. Basat, A. Portnoy, G. Mendelson, Y. Ben-Itzhak, and M. Mitzenmacher, "DRIVE: one-bit distributed mean estimation," *CoRR*, vol. abs/2105.08339, 2021. [Online]. Available: https://arxiv.org/abs/2105.08339

📄 R. B. Basat, S. Vargaftik, A. Portnoy, G. Einziger, Y. Ben-Itzhak, and M. Mitzenmacher, "QUIC-FL: Quick unbiased compression for federated learning," 2023. [Online]. Available: https://arxiv.org/abs/2205.13341