Konstantin Mishchenko, Bokun Wang, Dmitry Kovalev, Peter Richtárik

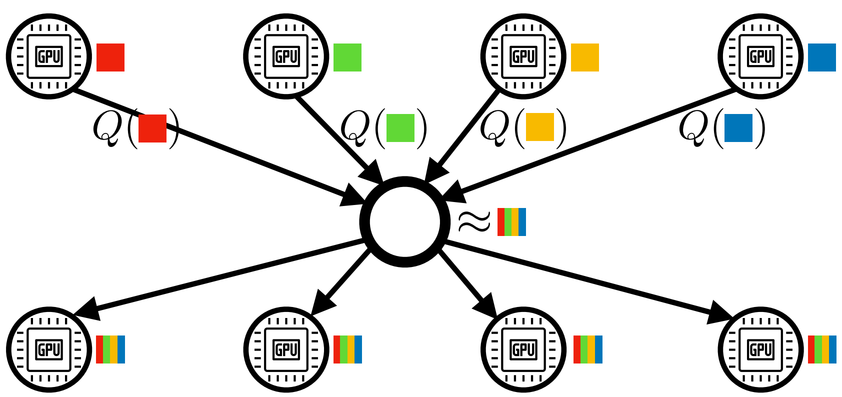
# Problem

$$\min_{x\in\mathbb{R}^d} f(x)$$

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x)$$

$$f_i(x) = \mathbb{E}_\xi[f_i(x;\xi)]$$

# Idea: Compression that supports all-reduce




# IntSGD

$$\alpha_k = \frac{\sqrt{d}}{\sqrt{2n\|x^k - x^{k-1}\|^2/\eta_k^2 + \varepsilon^2}}$$

$$x^{k+1} = x^k - \frac{\eta_k}{n}\sum_{i=1}^{n}\frac{1}{\alpha_k}\mathcal{I}nt(\alpha_k \circ g_i^k)$$

$$Q(g_i^k) = \frac{1}{\alpha_k}\mathcal{I}nt(\alpha_k \circ g_i^k)$$

$$\mathcal{I}nt(t) \overset{\text{def}}{=} \begin{cases} [t]+1, & \text{with probability } p_t \overset{\text{def}}{=} t - [t],\\ [t], & \text{with probability } 1 - p_t, \end{cases}$$

# Example

$$\begin{pmatrix}0.089\\-0.01\\0.05\\0.023\end{pmatrix} \xrightarrow{\alpha_k} \begin{pmatrix}8.9\\-1\\5\\2.3\end{pmatrix} \xrightarrow{\mathcal{I}nt} \begin{pmatrix}9\\-1\\5\\2\end{pmatrix} \xrightarrow{\frac{1}{\alpha_k}} \begin{pmatrix}0.09\\-0.01\\0.05\\0.02\end{pmatrix}$$

# Properties

$$\mathbb{E}_Q[Q(g_i^k)] = g_i^k$$

$$\mathbb{E}_\xi[g_i^k] = \nabla f_i(x^k)$$

# Convergence

**Theorem.**

**Smooth nonconvex:**

$$\mathbb{E}\left[\|\nabla f(\hat{x}^k)\|^2\right] = \mathcal{O}\left(\frac{\sigma+\varepsilon}{\sqrt{kn}} + \frac{f(x^0)-f^{\inf}}{k}\right)$$

**Smooth convex:**

$$\mathbb{E}\left[f(\hat{x}^k) - f(x^*)\right] = \mathcal{O}\left(\frac{\sigma_*+\varepsilon}{\sqrt{kn}} + \frac{\|x^0-x^*\|}{k}\right)$$

**Non-smooth convex:**

$$\mathbb{E}\left[f(\hat{x}^k) - f(x^*)\right] = \mathcal{O}\left(\frac{\sigma+\varepsilon}{\sqrt{kn}} + \frac{G}{\sqrt{k}}\right)$$

| Algorithm | Supports all-reduce | Supports switch | Provably works | Fast compression | Works without error-feedback | Adaptive | Reference |
|---|---|---|---|---|---|---|---|
| IntSGD | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | **Ours** |
| Heuristic IntSGD | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ | Sapio et al. (2021) |
| PowerSGD (theoretical) | ✓ | ✗ | ✓ | ✗(1) | ✗ | ✗(2) | Vogels et al. (2019) |
| PowerSGD (practical) | ✓ | ✗ | ✗ | ✓(1) | ✗ | ✗(2) | Vogels et al. (2019) |
| NatSGD | ✗ | ✓ | ✓ | ✗ | ✓ | N/A | Horváth et al. (2019) |
| QSGD | ✗ | ✗ | ✓ | ✓ | ✓ | N/A | Alistarh et al. (2017) |
| SignSGD | ✗ | ✗ | ✓ | ✓ | ✗ | N/A | Karimireddy et al. (2019) |