Algorithm 1 Original Gradient sampling

Step 0: Set $k = 0$, initial point $x_0 \in \mathcal{D}$, initial sampling radius $\epsilon_0 \in (0, \infty)$, initial stationarity target $\nu_0 \in [0, \infty)$, sample size $m \geq n + 1$, line search parameters $(\beta, \gamma) \in (0, 1) \times (0, 1)$, termination tolerances $(\epsilon_{\text{opt}}, \nu_{\text{opt}}) \in [0, \infty) \times [0, \infty)$, and reduction factors $(\theta_\epsilon, \theta_\nu) \in (0, 1] \times (0, 1]$;

Step 1: Choose $\{x_{k1}, \ldots, x_{km}\} \in \mathcal{B}_{\epsilon_k}(x^k)$ with randomly, independently, and uniformly sampled elements.

Step 2: Compute g_k as the solution of $\min_{\mathbf{g} \in \mathcal{G}_k} \frac{1}{2}\|\mathbf{g}\|^2$, where

$$\mathcal{G}_k := \operatorname{conv}\{\nabla f(x_k), \nabla f(x_{k,1}), \ldots, \nabla f(x_{k,m})\}$$

Step 3: If $\|g_k\| \leq \nu_{\text{opt}}$ and $\epsilon_k \leq \epsilon_{\text{opt}}$, then STOP!

Otherwise, if $\|g_k\| \leq \nu_k$, then $\epsilon_{k+1} = \theta_\epsilon \epsilon_k$, $\nu_{k+1} = \theta_\nu \nu_k$, $x_{k+1} = x_k$ and go to Step 6.

Step 4: Do a backtracking line search and find the maximum $t_k \in \{1, \gamma, \gamma^2, \ldots\}$ such that

$$f(x_k - t_k g_k) < f(x_k) - \beta t_k ||g_k||^2.$$

Step 5: **(D.C.)** If f is differentiable at $x_k - t_k g_k$, set $x_{k+1} \leftarrow x_k - t_k g_k$.

Otherwise, set x_{k+1} randomly as any point where f is differentiable such that $f(x_{k+1}) < f(x_k) - \beta t_k \|g_k\|^2$ and $\|x_k - t_k g_k - x_{k+1}\| \leq \min\{t_k, \epsilon_k\}\|g_k\|$

Step 6: Set $k \leftarrow k + 1$ and go back to Step 1.