

**M.A. Butakova, O.O. Kartashov (Southern Federal University, Rostov-on-Don, Russia). Integral Gradient Descent Method in Deep Learning.**<sup>1</sup> The boosted optimization problem as many variables function  $F(x)$  based on the integral gradient descent compiled by stochastic and online gradient featuring.

**Theorem.** Let the variables be convex functions and functions satisfying Lipschitz conditions, then the integral gradient method is convergent. The integral gradient estimation uses a cache (dynamic memory) storing stochastic or online gradient directions. The cache size can be fixed or variable length. The cache has the form  $\{\nabla f_{\xi}(x_{\eta})\}$ , where  $\xi$  and  $\eta$  are independent random values distributed uniformly on the target function terms and iteration position number set:

$$\nabla_u f(X) = \sum_i \sum_j \nabla f_i(x_j).$$

At each iteration there is a stochastic change in the cache according to the random values distribution  $\xi$  and  $\eta$ .

---

<sup>1</sup>Research was financially supported by the Ministry of Science and Higher Education of the Russian Federation (State assignment in the field of scientific activity, № FENW-2023-0019).