Theorem

Филатов Андрей

21 Апреля 2021

Theorem 1. *Every accumulation point of sequence* $x_k = [z_k, \theta_k^1, \ldots, \boldsymbol{\theta}_k^t]_{k=1}^{\infty}$ *produced by gradient descent with following Armijo-rules is a Pareto critical point:*

$$\mathcal{L}^t(z - \eta d_z, \boldsymbol{\theta}^t - \eta \nabla_{\boldsymbol{\theta}^t} \mathcal{L}^t) \leq \mathcal{L}^t - \eta\beta \left\| \frac{\partial \mathcal{L}^t}{\partial \boldsymbol{\theta}^t} \right\|^2 \quad \forall t \tag{1}$$

$$\mathcal{L}^t(z - \eta d_z, \boldsymbol{\theta}^t - \eta \nabla_{\boldsymbol{\theta}^t} \mathcal{L}^t) \leq \mathcal{L}^t - \eta\beta \frac{\partial \mathcal{L}^t}{\partial \boldsymbol{\theta}^t}^\top d_z \quad \forall t \tag{2}$$

$$\mathcal{L}^t(z - \eta d_z, \boldsymbol{\theta}^t - \eta \nabla_{\boldsymbol{\theta}^t} \mathcal{L}^t) \leq \mathcal{L}^t - \eta\beta \frac{\partial \mathcal{L}^t}{\partial \boldsymbol{\theta}^t}^\top d_z - \eta\beta \left\| \frac{\partial \mathcal{L}^t}{\partial \boldsymbol{\theta}^t} \right\|^2 \quad \forall t \tag{3}$$

Proof.

1. Let $u^t = \nabla_{\boldsymbol{\theta}^t} L^t$. Let $x = [\bar{z}, \bar{\theta^1}, \ldots, \bar{\theta^n}]$ be an accumulation point of $x_k = [z_k, \theta_k^1, \ldots, \theta_k^n]_{k=1}^{\infty}$. Then, there exists a subsequence $x_{k_j} = [z_{k_j}, \theta_{1k_j}, \ldots, \theta_{nk_j}]_{j=1}^{\infty}$ converging to x. As L continious: $L(x_{k_j}) \to L(\bar{x})$. Consequently, $\eta\beta\|u^t\|^2 \to 0$. There is an alternative:

   - $\limsup \eta > 0$
   - $\lim \eta = 0$

   In the first case $\|u^t\|^2 \to 0$, hence, $\frac{\partial L^t}{\partial z} = u^t \frac{\partial \boldsymbol{\theta}^t}{dz} \to 0$ and this point is Pareto stationary.

   In the second case suppose that accumulation point x is not Pareto critical. Then, there exists d_z a minimizing direction. As $\lim \eta = 0$ then for every constant step size η_n from some j_0 we can't satisfy Armijo condition at least for one function:

   $$\forall \eta_n = \frac{1}{n} \exists j_0 : \forall j \geq j_0 \; \exists t_n \; L^{t_n}(z_{k_j} - \eta_n s_{k_j}, \theta_{k_j} - \eta_n u_j^{t_n}) \geq L^{t_n}(z_{k_j}) - \eta_n \beta \|u_j^{t_n}\|^2$$

   As consequence of indices $\{t_n\}$ is bounded by number of tasks T, one can get sub-sequence of indices $\{t_{n_m}\}$, which converges to some index t_0.

   Hence, starting from some j_0 we get that for L^{t_0} is true

   $$\forall j \geq j_0 \quad L^{t_0}(z_{k_j} - \eta_n s_{k_j}, \theta_{k_j}^{t_0} - \eta_n u_j^{t_0}) \geq L^{t_0}(z_{k_j}) - \eta_n \beta \|u_j^{t_0}\|^2$$

As $L^{t_0} \in C^1$ and $x_{k_j} \to x$ so $s_{k_j} \to s$ and we get

$$L^{t_0}(\bar{z} - \eta_n \bar{s}, \bar{\theta}^{t_0} - \eta_n \bar{u}^{t_0}) \geq L^{t_0}(\bar{z}) - \eta_n \beta \|\bar{u}^{t_0}\|^2$$

As it is true $\forall \eta_n = \frac{1}{n}$, where $n \in \mathbb{N}$ we get contradiction with Armijo rule: because If $\left[\frac{\partial L}{\partial \theta}, \frac{\partial L}{\partial z}\right]\left[\frac{\partial L}{\partial \theta}, s\right]^T \preceq 0$, then $\exists \bar{\eta}: \ \forall t \ L^t(\bar{z} - \eta \bar{s}, \bar{\theta}^t - \eta \bar{u}^t) \leq L^t(\bar{z}) - \eta\beta\|u^t\|^2 - \eta\beta \frac{\partial L^t}{\partial z}^T s$. So $\bar{s} = 0$ and $\bar{x}$ — Pareto critical point.

2. For this Armijo-rule proof can be gained by following modifications. As $L(x_{k_j}) \to L(x)$ we have $\eta\beta \frac{\partial L^t}{\partial z}^T \bar{s} \to 0$.

   In the first variant of alternative we have $\forall t \quad \frac{\partial L^t}{\partial z}^T \bar{s} = 0$. As $\frac{\partial L^t}{\partial z}$ isn't singular matrix, then $\bar{s} = 0$ and this is Pareto stationary point.

   In the second variant of alternative we can change all $\eta\beta\|u_{t_0}\|^2$ on $\eta\beta \frac{\partial L^t}{\partial z}^T s$ and proof doesn't change.

3. For this Armijo-rule proof can be gained by following modifications. As $L(x_{k_j}) \to L(x)$ we have $\eta\beta \frac{\partial L^t}{\partial z}^T \bar{s} + \eta\beta\|u^t\|^2 \to 0$. Both terms are non-negative so we have same alternative.

   In the first variant of alternative we have that $s = 0, u^t = 0$, so x — Pareto stationary point.

   In the second variant of alternative we can change all $\eta\beta\|u^t\|^2$ on $\eta\beta \frac{\partial L^t}{\partial z}^T \bar{s} + \eta\beta\|u^t\|^2$ and proof doesn't change.

$\square$