最优化方法　第九讲

Weiwen Wang(王伟文)

暨南大学

2025 年 5 月 13 日

目录

定理 1

设 f 为凸函数，$\mathrm{dom}\ f$ 非空且 $f > -\infty$，则

$$\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x} \in \mathrm{dom}\ f} f(\boldsymbol{x})$$

当且仅当 $\mathbf{0} \in \partial f(\boldsymbol{x}^*)$.

考虑优化问题

$$\min_{\boldsymbol{x}\in\mathbb{R}^d}\left\{f(\boldsymbol{x})\triangleq\max_{i\in[m]}\{\boldsymbol{a}_i^T\boldsymbol{x}+b_i\}\right\},$$

其中 $\boldsymbol{a}_i\in\mathbb{R}^d$, $b_i\in\mathbb{R}$. 记 $I(\boldsymbol{x})=\{i\in[m]:f(\boldsymbol{x})=\boldsymbol{a}_i^T\boldsymbol{x}+b_i\}$, 已知

$$\partial f(\boldsymbol{x})=\left\{\sum_{k\in I(\boldsymbol{x})}\lambda_i\boldsymbol{a}_k:\lambda_k\in\Delta_{|I(\boldsymbol{x}^*)|}\right\}.$$

由定理 1, $\boldsymbol{x}^*$ 为最优解当且仅当 $\mathbf{0}\in\partial f(\boldsymbol{x}^*)$, 即

$$\mathbf{0}\in\left\{\sum_{k\in I(\boldsymbol{x})}\lambda_i\boldsymbol{a}_k:\lambda_k\in\Delta_{|I(\boldsymbol{x}^*)|}\right\}.$$

$$\mathbf{0} \in \left\{ \sum_{k \in I(\boldsymbol{x})} \lambda_i \boldsymbol{a}_k : \lambda_k \in \Delta_{|I(\boldsymbol{x}^*)|} \right\}.$$

此条件可改写为存在 $\boldsymbol{\lambda} = (\lambda_i)_{i \in [m]}$ 满足以下为方程组

$$\begin{aligned} &\boldsymbol{A}\boldsymbol{\lambda} = 0, \\ &\boldsymbol{\lambda} \in \Delta_m, \\ &\lambda_j = 0, \text{if } j \notin I(\boldsymbol{x}^*). \end{aligned}$$

其中 $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m] \in \mathbb{R}^{d \times m}$.

定理 2

设 f 为凸函数, $\mathrm{dom}\, f$ 非空且 $f > -\infty$, C 为凸集, 则

$$\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x}\in C} f(\boldsymbol{x})$$

当且仅当存在 $\boldsymbol{g} \in \partial f(\boldsymbol{x}^*)$, 使得

$$\langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x}^* \rangle \geq 0 \quad \forall \boldsymbol{y} \in C.$$

命题 3

设 f 为凸函数, $\operatorname{dom} f$ 非空且 $f > -\infty$, 则

$$\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x}\in\Delta_d} f(\boldsymbol{x})$$

当且仅当存在 $\boldsymbol{g} \in \partial f(\boldsymbol{x}^*)$ 和 $\mu \in \mathbb{R}$, 使得

$$g_i \begin{cases} = \mu, & x_i^* > 0; \\ \geq \mu, & x_i^* = 0. \end{cases}$$

考虑优化问题

$$\min_{\boldsymbol{x}\in\Delta_n}\left\{\sum_{i=1}^{n}x_i\log x_i-\sum_{i=1}^{n}y_ix_i\right\}$$

其中 $\boldsymbol{y}=(y_i)_{i\in[m]}$ 为给定向量. 假设存在最优解 $\boldsymbol{x}^*$ 每一个分量为正数, 此时

$$\frac{\partial f(\boldsymbol{x}^*)}{\partial x_i}=1+\log x_i^*-y_i$$

由命题 3, 存在 $\mu\in R$, 使得

$$1+\log x_i^*-y_i=\mu\quad\forall i\in[m]$$

解得

$$x_i^*=e^{\mu-1}e^{y_i}.$$

记 $\alpha = e^{\mu-1}$, 因为 $\sum\limits_{i=1}^{n} x_i^* = 1$, 故 $\alpha = \frac{1}{\sum\limits_{i=1}^{n} e^{y_i}}$, 因此

$$x_i^* = \frac{e^{y_i}}{\sum\limits_{i=1}^{n} e^{y_i}}.$$

再次应用命题 3, 上述 $\boldsymbol{x}^*$ 满足最优性条件, 因此为优化问题的最优解.

考虑一般凸优化问题

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$\text{s.t. } \boldsymbol{x} \in \mathsf{C}$$

其中 f 为凸函数, 但不一定可微, C 为非空闭凸集.

次梯度法

设步长为 γ_k, 迭代格式

- $\boldsymbol{x}_{k+1} = \Pi_C(\boldsymbol{x}_k - \gamma_k \boldsymbol{g}_k), \quad \boldsymbol{g}_k \in \partial f(\boldsymbol{x}_k).$

沿着次梯度反方向移动不一定保证目标函数值下降, 因此次梯度法不是下降方法.

步长选择

- 常数步长 $\gamma_k = \gamma > 0$;
- 规范化步长 $\gamma_k = \frac{\gamma}{\|\boldsymbol{g}_k\|_2}$;
- $\sum\limits_{k=0}^{\infty} = \infty, \lim\limits_{k\to\infty} \gamma_k = 0$;
- Robbins-Monro 步长

$$\sum_{k=0}^{\infty} \gamma_k = 0, \sum_{k=0}^{\infty} \gamma_k^2 < \infty;$$

  (e.g., $\gamma_k = \frac{1}{k+1}$)

- Polyak 步长. 设 f^* 为优化问题最优值,

$$\gamma_k = \frac{f(\boldsymbol{x}_k) - f^*}{\|\boldsymbol{g}_k\|_2^2}.$$

定理 4

次梯度法满足

$$\min_{k=0,1,\ldots,K} f(\boldsymbol{x}_k) - f^* \le \left(\sum_{k=0}^{K} \gamma_k\right)^{-1} \left(\frac{1}{2}\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2 + \frac{1}{2}\sum_{k=0}^{K} \gamma_k^2 \|\boldsymbol{g}_k\|_2^2\right)$$

证明.

由次梯度定义

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \le \langle \boldsymbol{g}_k, \boldsymbol{x}_k - \boldsymbol{x}^* \rangle$$

不等式两端同时乘上 γ_k

$$\gamma_k \left(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \right) \le \langle \gamma_k \boldsymbol{g}_k, \boldsymbol{x}_k - \boldsymbol{x}^* \rangle$$

又因为

$$\langle \gamma_k \boldsymbol{g}_k, \boldsymbol{x}_k - \boldsymbol{x}^* \rangle = \frac{1}{2}\gamma_k^2 \|\boldsymbol{g}_k\|_2^2 + \frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - \frac{1}{2}\|\boldsymbol{x}_k - \gamma_k \boldsymbol{g}_k - \boldsymbol{x}^*\|_2^2$$

由到闭凸集上的正交投影非扩张性

$$\|\Pi_C(\boldsymbol{x}_k - \gamma_k \boldsymbol{g}_k) - \boldsymbol{x}^*\|_2^2 \le \|\boldsymbol{x}_k - \gamma_k \boldsymbol{g}_k - \boldsymbol{x}^*\|_2^2$$

续.

$$
\begin{aligned}
\langle \gamma_k \boldsymbol{g}_k, \boldsymbol{x}_k - \boldsymbol{x}^* \rangle &= \frac{1}{2}\gamma_k^2 \|\boldsymbol{g}_k\|_2^2 + \frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - \frac{1}{2}\|\boldsymbol{x}_k - \gamma_k \boldsymbol{g}_k - \boldsymbol{x}^*\|_2^2 \\
&\le \frac{1}{2}\gamma_k^2 \|\boldsymbol{g}_k\|_2^2 + \frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - \frac{1}{2}\|\Pi_C(\boldsymbol{x}_k - \gamma_k \boldsymbol{g}_k) - \boldsymbol{x}^*\|_2^2 \\
&= \frac{1}{2}\gamma_k^2 \|\boldsymbol{g}_k\|_2^2 + \frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - \frac{1}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_2^2
\end{aligned}
$$

因此

$$
\gamma_k \left(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)\right) \le \frac{1}{2}\gamma_k^2 \|\boldsymbol{g}_k\|_2^2 + \frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - \frac{1}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_2^2
$$

取 $k = 0, 1, \ldots, K$, 不等式求和得到

$$
\sum_{k=0}^{K} \gamma_k \left(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*)\right) \le \frac{1}{2}\sum_{k=0}^{K} \gamma_k^2 \|\boldsymbol{g}_k\|_2^2 + \frac{1}{2}\left(\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2 - \|\boldsymbol{x}_{K+1} - \boldsymbol{x}^*\|_2^2\right)
$$

续.

不等式

$$\sum_{k=0}^{K}\gamma_k\left(f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\right)\le\frac{1}{2}\sum_{k=0}^{K}\gamma_k^2\|\boldsymbol{g}_k\|_2^2+\frac{1}{2}\left(\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2^2-\|\boldsymbol{x}_{K+1}-\boldsymbol{x}^*\|_2^2\right)$$

可进一步改写为

$$\left(\min_{k=0,1,\ldots,K}f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\right)\sum_{k=0}^{K}\gamma_k\le\frac{1}{2}\sum_{k=0}^{K}\gamma_k^2\|\boldsymbol{g}_k\|_2^2+\frac{1}{2}\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2^2$$

变换得到

$$\min_{k=0,1,\ldots,K}f(\boldsymbol{x}_k)-f^*\le\left(\sum_{k=0}^{K}\gamma_k\right)^{-1}\left(\frac{1}{2}\|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2^2+\frac{1}{2}\sum_{k=0}^{K}\gamma_k^2\|\boldsymbol{g}_k\|_2^2\right)$$

□

若 $D := \mathtt{diam}(\mathsf{C}) = \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathsf{C}} \|\boldsymbol{x}-\boldsymbol{y}\|_2^2$, 且存在 $B>0$, 使得对任意 $\boldsymbol{x}\in\mathsf{C}$,

$$\|\boldsymbol{g}\|_2 \le B \quad \forall \boldsymbol{g}\in\partial f(\boldsymbol{x}).$$

则由定理 4,

$$\min_{k=0,1,\ldots,K} f(\boldsymbol{x}_k) - f^* \le \left(\sum_{k=0}^{K}\gamma_k\right)^{-1}\left(\frac{1}{2}D^2 + \frac{1}{2}B^2\sum_{k=0}^{K}\gamma_k^2\right)$$

若取 $\gamma_k = \gamma$

$$
\begin{aligned}
\min_{k=0,1,\ldots,K} f(\boldsymbol{x}_k) - f^* &\le \left(\sum_{k=0}^{K} \gamma_k\right)^{-1} \left(\frac{1}{2}D^2 + \frac{1}{2}B^2 \sum_{k=0}^{K} \gamma_k^2\right) \\
&= \frac{\frac{1}{2}D^2 + \frac{1}{2}B^2(K+1)\gamma^2}{(K+1)\gamma} \\
&= \frac{D^2}{2(K+1)\gamma} + \frac{B^2\gamma}{2} \to \frac{B^2\gamma}{2} (K \to \infty)
\end{aligned}
$$

不等式最右端关于 γ 最小化, 得到最优步长

$$
\gamma^* = \frac{D}{B\sqrt{K+1}}.
$$

将 $\gamma^* = \frac{D}{B\sqrt{K+1}}$ 代入

$$
\begin{aligned}
\min_{k=0,1,\dots,K} f(\boldsymbol{x}_k) - f^* &\leq \frac{D^2}{2(K+1)} \frac{B\sqrt{K+1}}{D} + \frac{B^2}{2} \frac{D}{B\sqrt{K+1}} \\
&\leq \frac{DB}{2\sqrt{K+1}}.
\end{aligned}
$$

若取 Polyak 步长

$$\gamma_k = \frac{f(\boldsymbol{x}_k) - f^*}{\|\boldsymbol{g}_k\|_2^2}.$$

由正交投影的非扩张性及次梯度定义

$$\begin{aligned}
\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|_2^2 &= \|\Pi_C(\boldsymbol{x}_k - \gamma_k \boldsymbol{g}_k) - \boldsymbol{x}^*\|_2^2 \\
&\leq \|\boldsymbol{x}_k - \gamma_k \boldsymbol{g}_k - \boldsymbol{x}^*\|_2^2 \\
&= \|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - 2\langle \gamma_k \boldsymbol{g}_k, \boldsymbol{x}_k - \boldsymbol{x}^* \rangle + \gamma_k^2 \|\boldsymbol{g}_k\|_2^2 \\
&= \|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - 2\gamma_k (f(\boldsymbol{x}_k) - f^*) + \gamma_k^2 \|\boldsymbol{g}_k\|_2^2 \\
&= \|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - \frac{(f(\boldsymbol{x}_k) - f^*)^2}{\|\boldsymbol{g}_k\|_2^2}
\end{aligned}$$

$$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|_2^2 \leq \|\boldsymbol{x}_k-\boldsymbol{x}^*\|_2^2-\frac{(f(\boldsymbol{x}_k)-f^*)^2}{\|\boldsymbol{g}_k\|_2^2}$$

可变换为

$$\frac{(f(\boldsymbol{x}_k)-f^*)^2}{B^2} \leq \frac{(f(\boldsymbol{x}_k)-f^*)^2}{\|\boldsymbol{g}_k\|_2^2} \leq \|\boldsymbol{x}_k-\boldsymbol{x}^*\|_2^2-\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|_2^2$$

取 $k=0,1,\ldots,K$, 求和得到

$$\sum_{k=0}^{K}\frac{(f(\boldsymbol{x}_k)-f^*)^2}{B^2} \leq \|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2^2-\|\boldsymbol{x}_{K+1}-\boldsymbol{x}^*\|^2 \leq R^2.$$

故

$$\sum_{k=0}^{K}(f(\boldsymbol{x}_k)-f^*)^2 \leq R^2B^2 < \infty,$$

因此

$$(f(\boldsymbol{x}_K)-f(\boldsymbol{x}^*))^2 \xrightarrow{K\to\infty} 0 \Longrightarrow f(\boldsymbol{x}_K) \xrightarrow{K\to\infty} f(\boldsymbol{x}^*).$$

定义 5 (非光滑 μ-强凸函数)

设 f 为凸集 C 上的凸函数, 若存在 $\mu > 0$ 使得对任意 $\boldsymbol{x} \in C$ 有

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{\mu}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \quad \forall \boldsymbol{y} \in C, \forall \boldsymbol{g} \in \partial f(\boldsymbol{x}).$$

则称 f 为 C 上的 μ-强凸函数.

定理 6

若 f 为凸集 C 上的 μ-强凸函数, 则步长 $\gamma_k = \frac{1}{\mu(k+1)}$ 的次梯度法满足

$$\min_{k=0,1,\ldots,K} f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \le \frac{B^2(\ln(K+1)+1)}{2\mu(K+1)},$$

$$f(\hat{\boldsymbol{x}}_K) - f(\boldsymbol{x}^*) \le \frac{B^2(\ln(K+1)+1)}{2\mu(K+1)},$$

其中 $\hat{\boldsymbol{x}}_K = \frac{1}{K+1}\sum_{k=0}^{K} \boldsymbol{x}_k$, $K \ge 0$.

证明.

由定理 4 的证明有

$$\begin{aligned}\langle \boldsymbol{g}_k, \boldsymbol{x}_k - \boldsymbol{x}^* \rangle &\le \frac{1}{2}\gamma_k \|\boldsymbol{g}_k\|_2^2 + \frac{1}{2\gamma_k}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - \frac{1}{2\gamma_k}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_2^2 \\ &= \frac{1}{2\mu(k+1)}\|\boldsymbol{g}_k\|_2^2 + \frac{\mu(k+1)}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - \frac{\mu(k+1)}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_2^2\end{aligned}$$

因为 f 为 μ-强凸函数,

$$f(\boldsymbol{x}^*) \ge f(\boldsymbol{x}_k) + \langle \boldsymbol{g}_k, \boldsymbol{x}^* - \boldsymbol{x}_k \rangle + \frac{\mu}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2,$$

移项得到

$$\begin{aligned}f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) &\le \langle \boldsymbol{g}_k, \boldsymbol{x}_k - \boldsymbol{x}^* \rangle - \frac{\mu}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 \\ &\le \frac{1}{2\mu(k+1)}\|\boldsymbol{g}_k\|_2^2 + \frac{\mu}{2}\left(k\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_2^2 - (k+1)\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_2^2\right)\end{aligned}$$

续.

$$f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\le \frac{1}{2\mu(k+1)}\|\boldsymbol{g}_k\|_2^2+\frac{\mu}{2}\left(k\|\boldsymbol{x}_k-\boldsymbol{x}^*\|_2^2-(k+1)\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|_2^2\right)$$

取 $k=0,1,\ldots,K$, 求和得到

$$\sum_{k=0}^{K}\left(f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\right)\le \frac{1}{2\mu}\sum_{k=0}^{K}\|\boldsymbol{g}_k\|_2^2\frac{1}{k+1}\le \frac{B^2(\ln(K+1)+1)}{2\mu}$$

故

$$(K+1)\min_{k=0,1,\ldots,K} f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\le \sum_{k=0}^{K}\left(f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\right)\le \frac{B^2(\ln(K+1)+1)}{2\mu}$$

□

续.

$$\sum_{k=0}^{K}\left(f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\right)\le\frac{B^2(\ln(K+1)+1)}{2\mu}$$

不等式两端同时除以 $K+1$

$$\frac{1}{K+1}\sum_{k=0}^{K}\left(f(\boldsymbol{x}_k)-f(\boldsymbol{x}^*)\right)\le\frac{B^2(\ln(K+1)+1)}{2\mu(K+1)}$$

取 $\hat{\boldsymbol{x}}_K=\frac{1}{T}\sum\limits_{k=0}^{K}\boldsymbol{x}_k$, 由凸函数性质

$$f(\hat{\boldsymbol{x}}_K)\le\frac{1}{K+1}\sum_{k=0}^{K}f(\boldsymbol{x}_k)$$

故

$$f(\hat{x}_K)-f(\boldsymbol{x}^*)\le\frac{B^2(\ln(K+1)+1)}{2\mu(K+1)}$$

次梯度
最优性条件
次梯度法
迭代格式
收敛性分析
收敛阶

定理 7

若 f 为凸集 C 上的 μ-强凸函数，则步长 $\gamma_k = \frac{2}{\mu(k+2)}$ 的次梯度法满足

$$\min_{k=0,1,\dots,K} f(\boldsymbol{x}_k) - f(\boldsymbol{x}^*) \le \frac{2B^2}{\mu(K+2)},$$

$$f(\hat{\boldsymbol{x}}_K) - f(\boldsymbol{x}^*) \le \frac{2B^2}{\mu(K+2)},$$

其中 $\hat{\boldsymbol{x}}_K = \sum_{k=0}^{K} \frac{2(k+1)}{(K+1)(K+2)} \boldsymbol{x}_k$, $K \ge 0$.

凸函数

- 次梯度法 $\mathcal{O}\left(\frac{BR}{\sqrt{k}}\right)$
- 加速梯度下降 $\mathcal{O}(\frac{LR^2}{k^2})$

μ-强凸函数

- 次梯度法 $\mathcal{O}(\frac{B^2}{\mu k})$
- 加速梯度下降 $\mathcal{O}\left(\left(\frac{1-\sqrt{\kappa}}{1+\sqrt{\kappa}}\right)^{2k}\right)$

- 次梯度法总比加速梯度下降慢;
- 在非光滑强凸函数下, 次梯度法次线性收敛; 而光滑强凸函数的梯度下降法线性收敛;
- 关于次梯度法的复杂度是最优的 (Nemirovski & Yudin 1979).

定理 8 (Nemirovski & Yudin 1979)

对任意 $1 \le k \le d$, $\boldsymbol{x}_1 \in \mathbb{R}^d$, 任意一阶方法产生迭代序列 $\{\boldsymbol{x}_k\} \subset \mathrm{C}$ 满足

$$\boldsymbol{x}_k \in \boldsymbol{x}_1 + \mathrm{span}(\boldsymbol{g}_1, \ldots, \boldsymbol{g}_{k-1})$$

其中 $\boldsymbol{g}_i \in \partial f(\boldsymbol{x}_i), i = 1, \ldots, k-1$, $\mathrm{C} \subseteq \mathbb{R}^d$ 为直径 D 的凸集.

(1) 存在 B-*Lipschitz* 连续的凸函数, 使得

$$\min_{s \in [k]} f(\boldsymbol{x}_s) - f(\boldsymbol{x}^*) \ge \frac{BD}{4(1+\sqrt{k})}.$$

(2) 存在 B-*Lipschitz* 连续的凸函数的 μ 强凸函数, 使得

$$\min_{s \in [k]} f(\boldsymbol{x}_s) - f(\boldsymbol{x}^*) \ge \frac{B^2}{8\mu k}.$$