الگوریتم های بهینه‌سازی

الگوریتم

* الگوریتم های کاهشی
* الگوریتم غیرساختاری
* الگوریتم های مبتنی بر ساختار ←

(وقتی که توابع هدف و قید دارای ساختار مشخصی باشند)

الگوریتم غیرساختاری

اوراکل

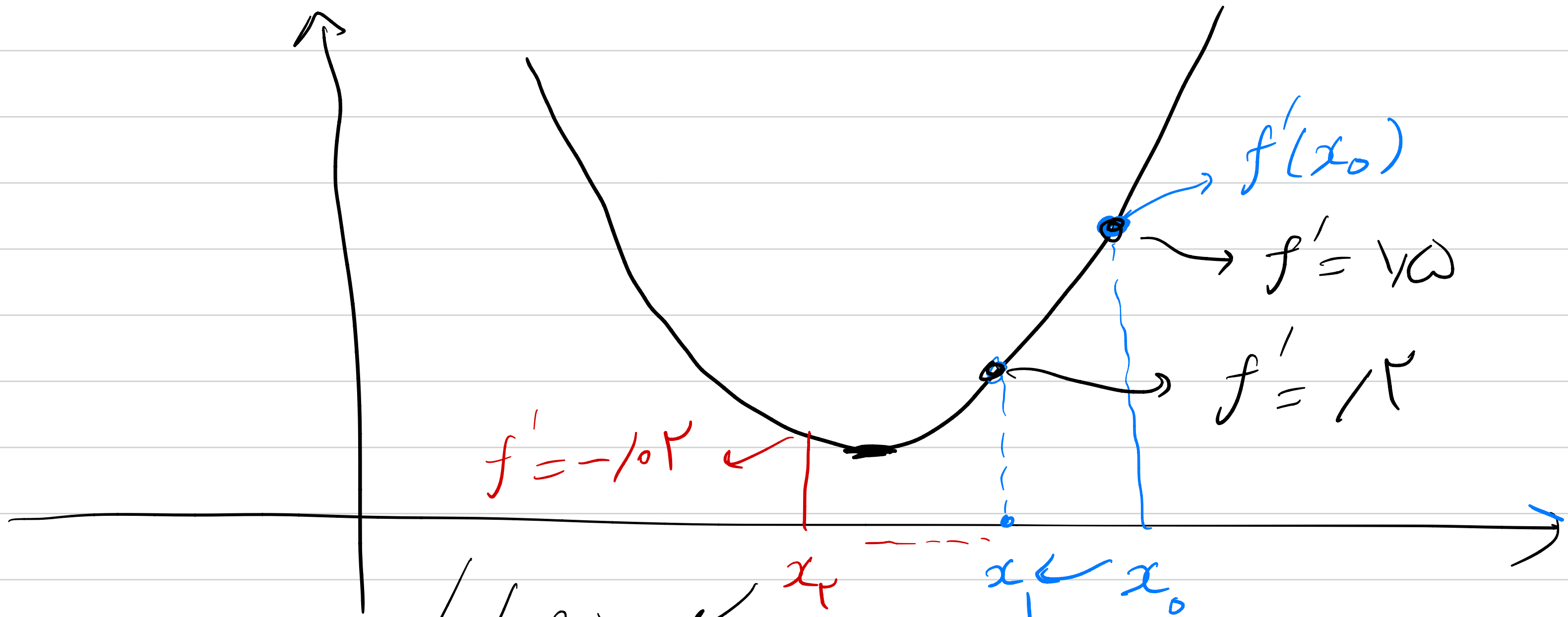
- مقدار تابع در نقطه x ؟
- مقدار مشتق تابع در x ؟ ← مرتبه اول
- مشتق دوم در x ؟

مرتبه دوم

• الگوریتم کاهش گرادیان .

$$\min_{x \in \text{کل فضا}} f(X)$$

ایده کلی ← مقدار تابع در x و گرادیان در x .



چون شیب کوچک متوقف می‌شوم!

• f قویاً محدب با پارامتر m.

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle + \frac{m}{۲} \| y - x \|^{۲}$$

$$\min_y f(y) \geqslant \min_y \text{-------}$$

$$f^* \geqslant f(x) + \min_y \langle \nabla f(x), y - x \rangle + \frac{m}{2} \| y - x \|^2$$

$$\nabla f(x) + m (y - x) = ۰$$

$$\longrightarrow y_0 = x - \frac{1}{m} \nabla f(x)$$

$$f^* \geqslant f(x) - \frac{1}{۲m} \| \nabla f(x) \|^{۲}$$

$$\frac{1}{2m}\|\nabla f(x)\|^2 \geq f(x)-f^* \geq 0$$

$$\longrightarrow \frac{1}{2m}\|\nabla f(x)\|^2 \leq \varepsilon \Longrightarrow f(x)-f^* < \varepsilon$$

$$\|\nabla f(x)\|^2 \leq 2m\varepsilon$$

محک فوق در مورد نزدیکی x به x^* صحبتی نمی‌کند.

شرایط معادل قویاً محدب

$$\nabla^2 f(x) \geq mI$$

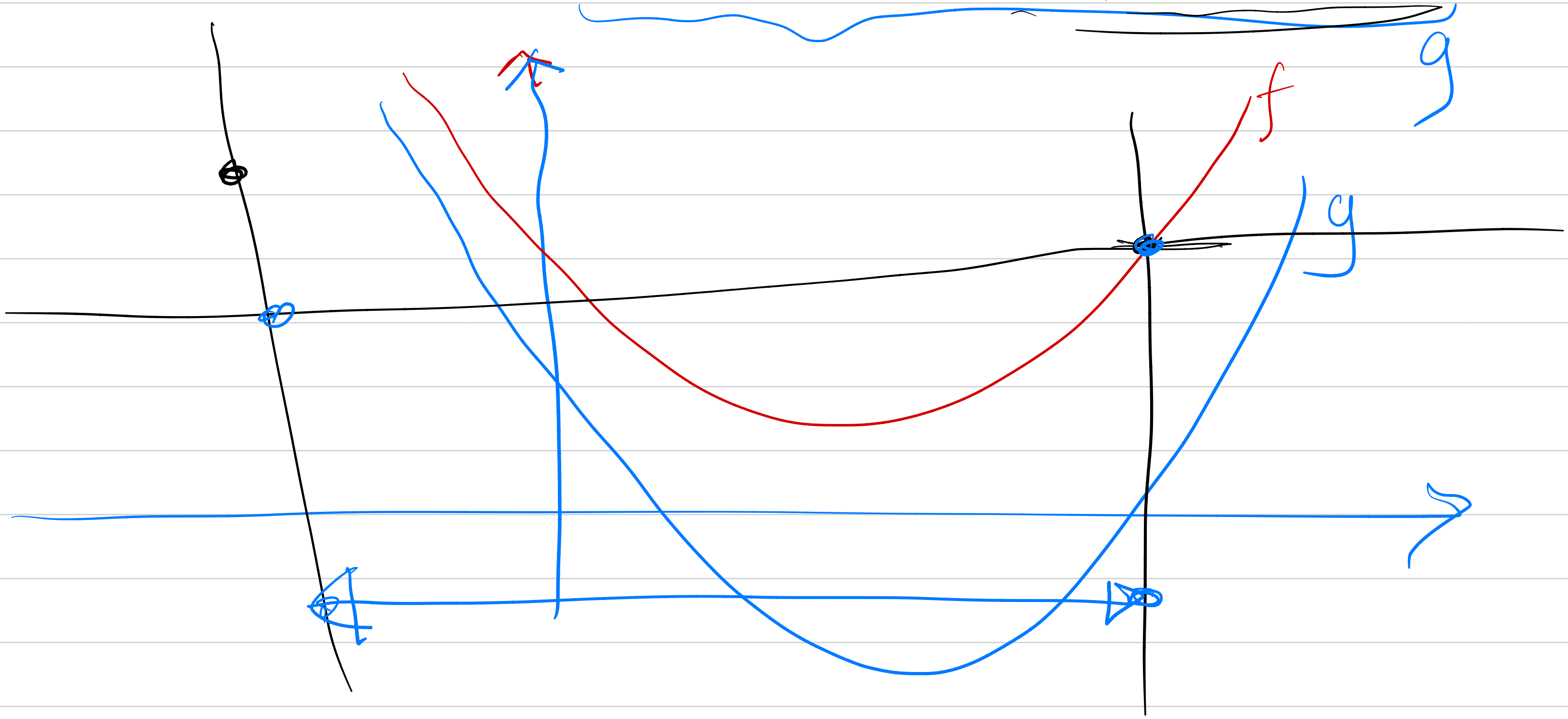
$$\langle \nabla f(y) - \nabla f(x),\ y-x \rangle \geq m\|y-x\|^2$$

$$\|\nabla f(y) - \nabla f(x)\| \geq m\|x-y\|$$

کافی ✓

لازم ؟

$$f(y) \geq \underbrace{f(x) + \cdots + \frac{m}{2}\|y-x\|^2}_{g}$$



هموار بودن : M ـ هموار

$$f(y) \le f(x) + \langle \nabla f(x), y-x \rangle + \frac{M}{2}\|y-x\|^2$$

• لیپ شیتز بودن گرادیان - (M - هموار)

$$\begin{cases} \nabla^2 f \preccurlyeq M I \\ f(y) \leq f(x) + \langle \nabla f(x), y-x \rangle + \frac{M}{2} \|y-x\|^2 \\ \langle \nabla f(y) - \nabla f(x), y-x \rangle \leq M \|y-x\|^2 \\ \|\nabla f(y) - \nabla f(x)\| \leq M \|x-y\| \end{cases}$$

$$f^* \leq f(x) - \frac{1}{۲M} \|\nabla f(x)\|^۲$$

$$\frac{1}{۲m} \|\nabla f(x)\|^۲ \geq f(x) - f^* \geq \frac{1}{۲M} \|\nabla f(x)\|^۲$$

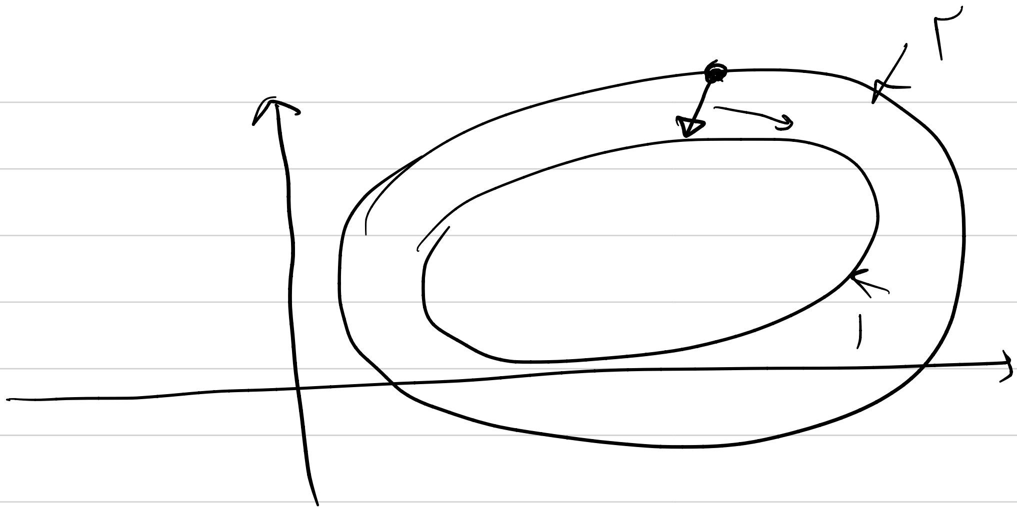
عدد حالت :

$$\text{Cond}(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$$

$$mI \preceq \nabla^2 f \preceq MI$$

$$\text{Cond}(\nabla^2 f) \leq \frac{M}{m}$$

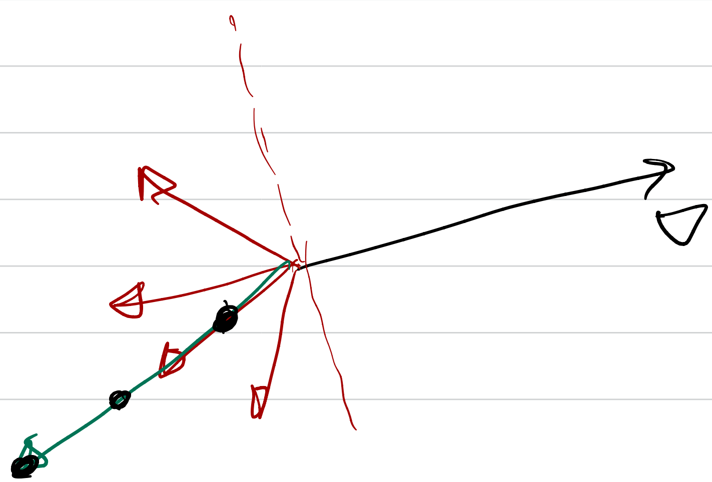
• جهت کاهش.

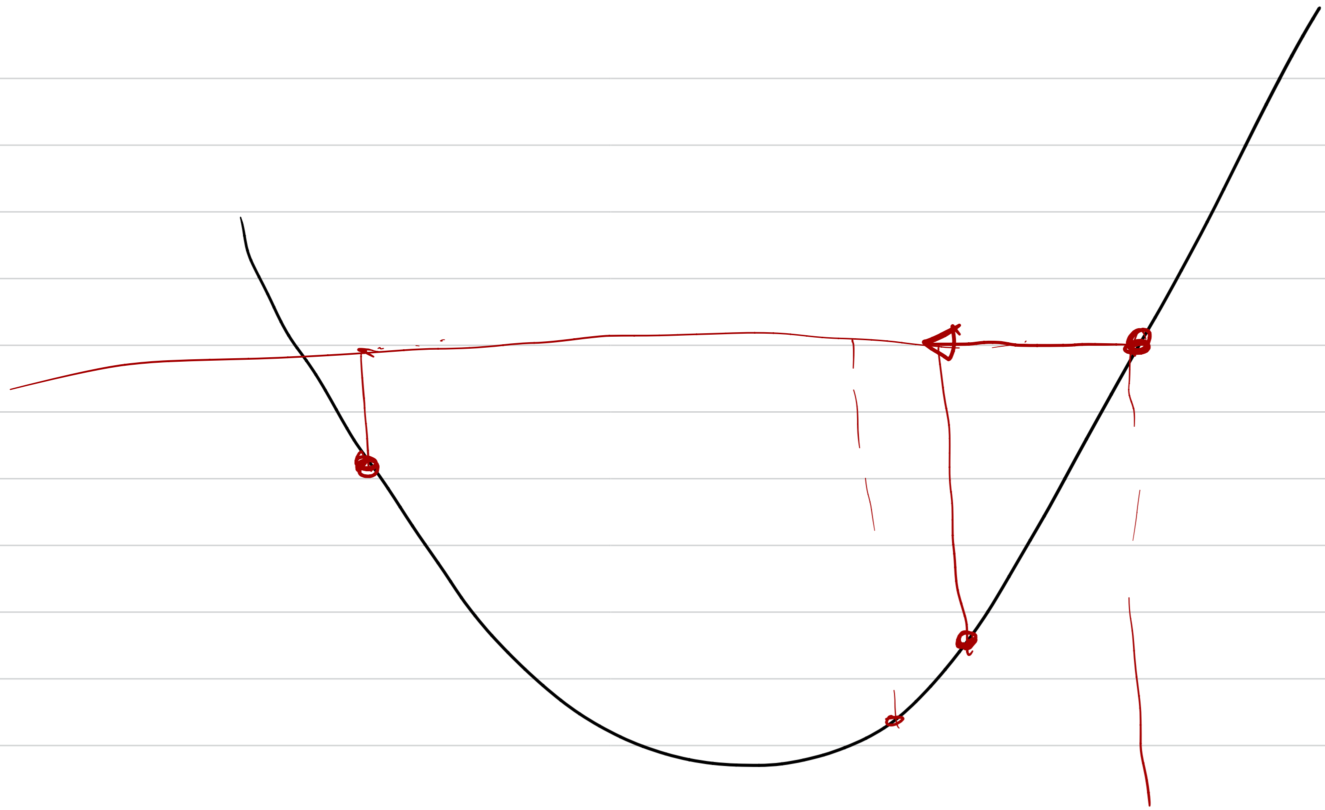


$$f(x+\varepsilon v) \approx f(x) + \varepsilon \langle v, \nabla f(x) \rangle$$

جهت کاهش: جهتی $\langle v, \nabla f(x) \rangle < 0$

مقدار حرکت ε چقدر باشد؟





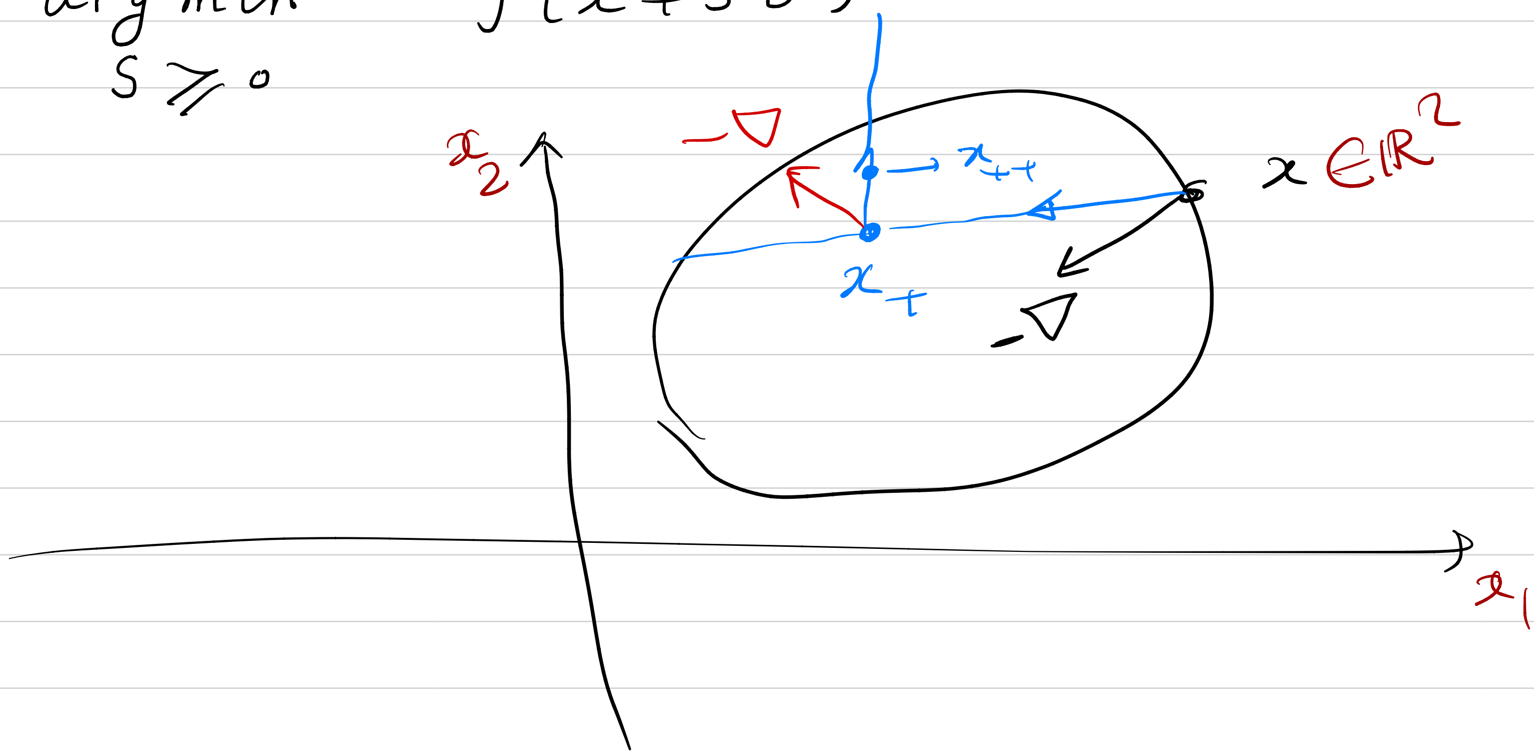
• روش اول: جستجوی خط •

جهت کاهش v :

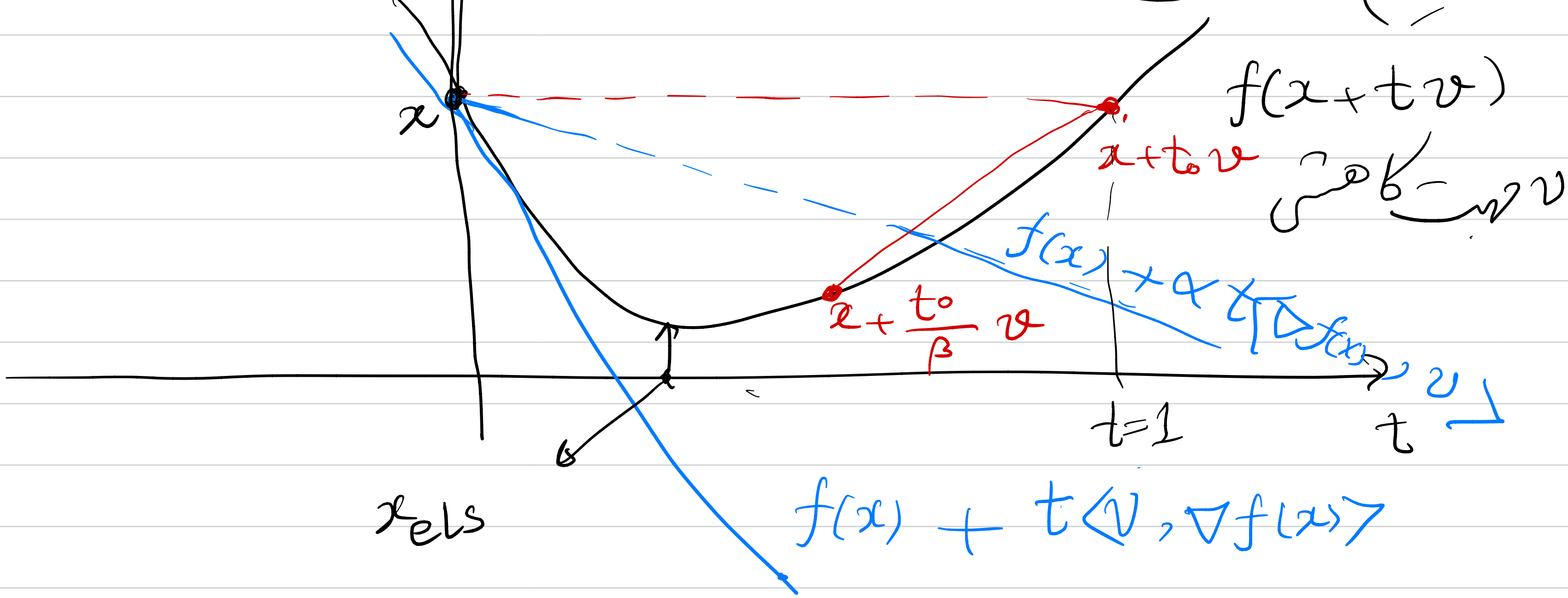
$$f(x + s v)$$

$$x_+ = x + s v$$

$$s = \underset{s \geq 0}{\arg\min} \; f(x + s v)$$



الگوریتم Backtracking



Alg. $\quad t_0,\ \alpha\in(0,\tfrac{1}{2}),\ \beta<1,\ x_0$ $\qquad \alpha\in(0,\tfrac{1}{2})$

در ازای i

$$z = t_0$$

while

$$f(x_i - \underbrace{z\nabla f(x_i)}) > f(x) - \alpha z\,\|\nabla f(x_i)\|^2$$

$$z = \beta z$$

$$x_{i+1} = x_i - z\nabla f(x_i)$$

معیار T ثابت است !

$\{\ \nabla f(x_i)<\delta\ !\ \}$

الگوریتم گرادیان کاهشی :

• جهت کاهشی : $v = -\nabla f(x)$

نقطه شروع x_0

در مرحله kام $x_k = x_{k-1} - t_k \nabla f(x_{k-1})$

T بار انجام می‌گیرد.

t_k ثابت

از els , backtracking

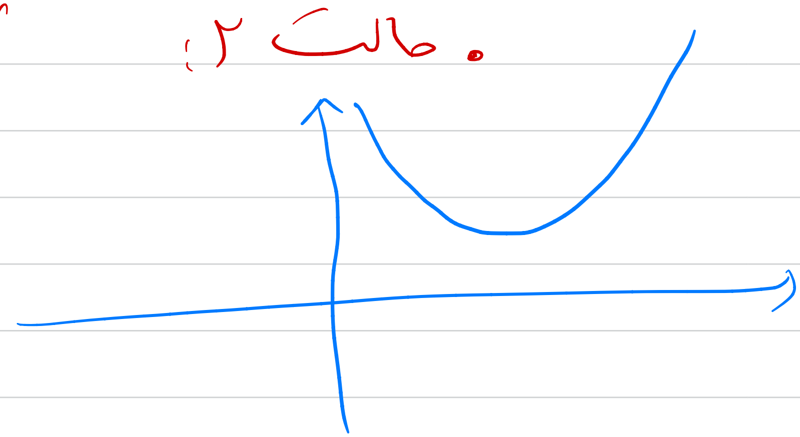
• تحلیل الگوریتم • اجرای کاهش گرادیان با اندازه قدم ثابت .

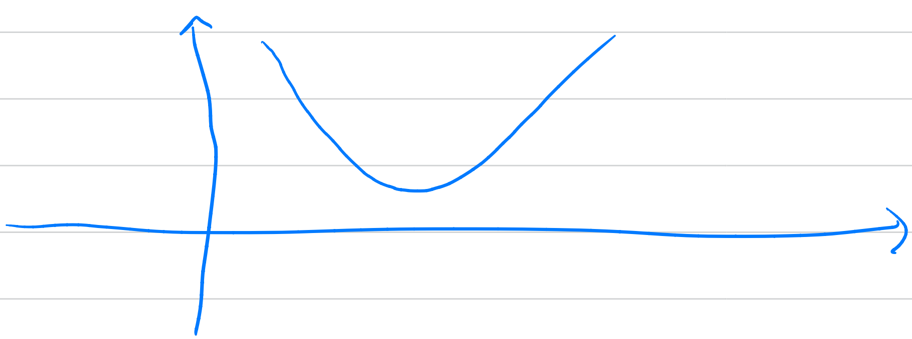
$$\boxed{\begin{array}{l} x_0, \eta \\ x_{i+1} = x_i - \eta \nabla f(x_i) \\ \text{for } t = 1, \dots, T \end{array}}$$

• سوال : $f(x_T) - f^* \leq \;?$

• حالت ۱ : f محدب ، M هموار

• حالت ۲ : f قویا محدب ، m ، M هموار





حالت ۲: m قویاً محدب ، M هموار

$$f(x_{i+1}) - f(x_i) \leq \langle \nabla f(x_i), x_{i+1} - x_i \rangle + \frac{M}{۲} \| x_{i+1} - x_i \|^{۲}$$

$$= -\eta \| \nabla f(x_i) \|^2 + \frac{M\eta^2}{2} \| \nabla f(x_i) \|^2$$

$$\eta = \frac{1}{M}$$

$$= -\frac{1}{2M} \| \nabla f(x_i) \|^2$$

$$f(x_i) - f^* \leq \frac{1}{2m} \| \nabla f(x_i) \|^2 \;;$$

قبلاً دیدیم

$$\Delta_i = f(x_i) - f^*$$

$$\Delta_{i+1} - \Delta_i \leq -\frac{1}{2M} \|\nabla f(x_i)\|^2$$

$$\Delta_i \leq \frac{1}{2m} \|\nabla f(x_i)\|^2 \rightarrow \| \quad \|^2 \leq -2m\Delta_i$$

$$\rightarrow \Delta_{i+1} - \Delta_i \leq -\frac{m}{M} \Delta_i$$

$$\rightarrow \Delta_{i+1} \leq \Delta_i \Big(1 - \underbrace{\frac{m}{M}}_{=\frac{1}{K}}\Big) \leq \Delta_i e^{-\frac{1}{K}}$$

$$\rightarrow \Delta_T \leq \Delta_0 e^{-\frac{T}{K}}$$

$$; \quad \Delta_0 = f(x_0) - f^* \leq \langle \underbrace{\nabla f(x^*)}_{=0}, x_0 - x^* \rangle + \frac{M}{2}\|x_0 - x^*\|^2$$

$$\Delta_0 \le \frac{M}{2}\|x_0 - x^*\|^2$$

$$\Delta_T \le \frac{M}{2}\|x_0 - x^*\|^2 e^{-\frac{T}{K}} \le \varepsilon$$

$T = O(\log \frac{1}{\varepsilon})$ ؛ الگوریتم ما سرعت خطی.

$0/00001\boxed{1}0 \ldots$ میزان تفاضل برحسب بیت بنویسید

$$\varepsilon = 2^{-b} \longrightarrow T = c \cdot b$$

سؤال: $\|x_T - x^*\|$ ؟

$$\le f(x_T) - f(x^*) \le \ldots$$

$$\ldots \ge f(x_T) - f(x^*) \ge \underbrace{\cancel{\langle \nabla f(x^*), x_T - x^* \rangle}}_{0} + \frac{m}{2}\|x_T - x^*\|^2$$

• حالت اول: f محدب، هموار با M. (f در x^* به کمینه خود می‌رسد)

الگوریتم $\eta = \frac{1}{M}$.

$$f(x_{i+1}) - f(x_i) \leq -\frac{1}{2M} \|\nabla f(x_i)\|^2$$

$$f(\underset{x^*}{y}) \geq f(x) + \langle \nabla f(\overset{x_i}{x}), \underset{x^*}{y} - x \rangle$$

$$f(x_i) - f^* \leq \langle \nabla f(x_i), x_i - x^* \rangle$$

$$f(x_{i+1}) - f^* \leq -\frac{1}{2M} \|\nabla f(x_i)\|^2 + \langle \nabla f(x_i), x_i - x^* \rangle$$

$$= \frac{1}{2M} \langle \nabla f(x_i), -\nabla f(x_i) + 2M(x_i - x^*) \rangle$$

$$= \frac{M}{2} \left(\|x_i - x^*\|^2 - \|x_i - x^* - \frac{1}{M} \nabla f(x_i)\|^2 \right)$$

$$= \frac{M}{2} \left(\|x_i - x^*\|^2 - \|x_{i+1} - x^*\|^2 \right)$$

$$\sum_{i=0}^{T-1} f(x_{i+1}) - f^* \le \frac{M}{2}\left(\|x_0 - x^*\|^2 - \|x_T - x^*\|^2\right)$$

$$\le \frac{M}{2}\|x_0 - x^*\|^۲$$

$$T\left(f(x_T) - f^*\right) \le \frac{M}{2}\|x_0 - x^*\|^2$$

$$\longrightarrow \quad f(x_T) - f^* \le \frac{M}{2T}\|x_0 - x^*\|^2 = O\left(\frac{1}{T}\right)$$

$$T = O\left(\frac{1}{\epsilon}\right)$$

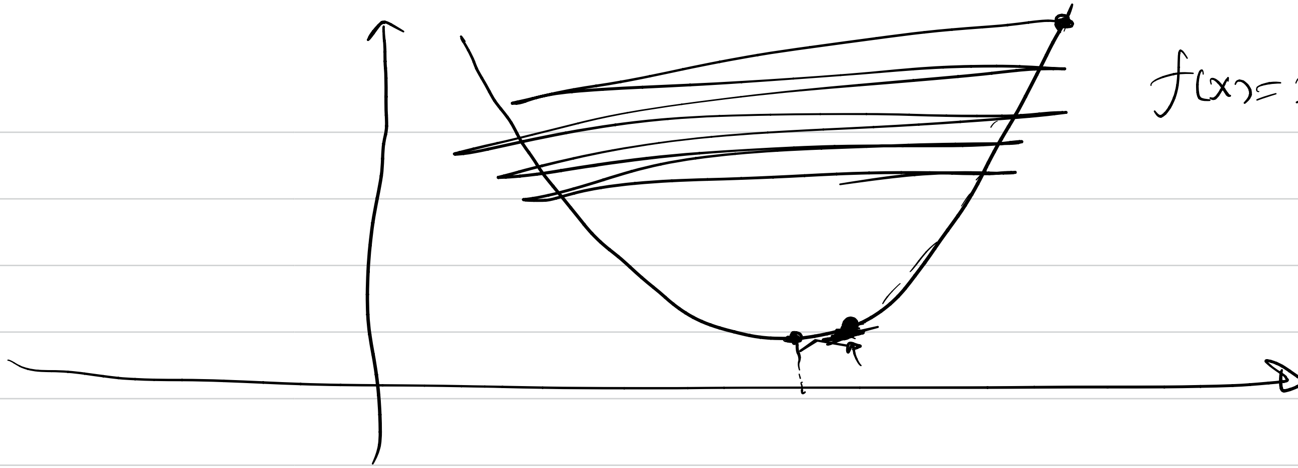
قضیه Nestrov : اگر یک تابع هموار f وجود دارد که هر الگوریتم بهینه‌سازی

که نقاط به $\nabla f(x_0), \dots, \nabla f(x_i)$ ،

$$f(x_i) - f^* \ge \omega\left(\frac{1}{T}\right) \iff T = \omega\left(\frac{1}{\epsilon}\right)$$

x_0

در قدم i ام $x_{i+1} = x_i +$ < یک نقطه‌ای در پوشش خطی $\nabla f(x_0), \dots, \nabla f(x_i)$ >



$f(x) = x^4$

$$\nabla f(x_i) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_i) \\ \vdots \\ \frac{\partial f}{\partial x_{100}}(x_i) \end{bmatrix}$$

یک درایه به صورت تصادفی میزان گرادیان را در

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 100\,\frac{\partial f}{\partial x_j}(x_i) \\ \vdots \\ 0 \end{bmatrix}$$