GRADIENT DESCENT

$x_{t+1} = x_t - \gamma \nabla f(x_t)$

$f(x_t) - f(x^*) \le \nabla f(x_t)^T (x_t - x^*) \rightarrow f(x)$ IS ABOVE TANGENT AT x^*

EQUIVALENT TO CONVEX.

→ f CONVEX, DIFFERENTIABLE, B-LIPSCHITZ

$\|x_0 - x^*\| \le R$, $\|\nabla f(x)\| \le B$, $\gamma = \frac{R}{B\sqrt{T}}$ (STEP SIZE)

$$\boxed{\left(\frac{1}{T}\sum_{t=0}^{T-1} f(x_t)\right) - f(x^*) \le \frac{RB}{\sqrt{T}}}$$

$$\left(\min_{t\in\{0..T-1\}} f(x_t)\right) - f(x^*) < \varepsilon \quad \text{IF } T = \frac{R^2B^2}{\varepsilon^2}$$

→ f CONVEX, DIFFERENTIABLE, L-SMOOTH, $\gamma = \frac{1}{L}$

$$\boxed{\frac{1}{2L}\|\nabla f(x_t)\|^2 \le f(x_t) - f(x_{t+1})}$$

PUT ε IN LHS $\rightarrow f(x_T) - f(x^*) \le \frac{LR^2}{2T}$

$$\boxed{f(x_T) - f(x^*) < \varepsilon, \quad T \ge \frac{R^2L}{2\varepsilon}}$$

→ f DIFFERENTIABLE, L SMOOTH, $\mu > 0$ STRONGLY CONVEX $\gamma = \frac{1}{L}$

$$\boxed{\|x_{t+1} - x^*\|^2 \le \left(1 - \frac{\mu}{L}\right)\cdot\|x_t - x^*\|^2, t \ge 0}$$

$$\boxed{f(x_t) - f(x^*) \le \frac{L}{2}\left(1 - \frac{\mu}{L}\right)^T R^2, T > 0}$$

$$\boxed{f(x_T) - f(x^*) \le \varepsilon, \quad T \ge \frac{L}{\mu}\ln\left(\frac{R^2L}{2\varepsilon}\right)}$$

SUBGRADIENT DESCENT (NO NEED FOR CONVEXITY)

$g_t \in \partial f(x_t)$

$x_{t+1} = x_t - \gamma_t g_t$

IF $0 \in \partial f(x)$, x IS GLOBAL MINIMUM

→ B-LIPSCHITZ (SAME AS BEFORE)

→ $\mu > 0$ STRONGLY CONVEX $\gamma_t = \frac{2}{\mu(t+1)}$, $t > 0$

$$f\left(\frac{2}{T(T+1)}\sum_{t=1}^{T} t x_t\right) - f(x^*) \le \gamma_t B^2 \quad \left(B = \max_{t=1..T}\|\overset{\partial f(x_t)\ \ni}{g_t}\|\right)$$

ACCELERATED GRADIENT DESCENT : (L-SMOOTH)

$y_{t+1} = x_t - \frac{1}{L}\nabla f(x_t)$

$z_{t+1} = z_t - \frac{t+1}{2L}\nabla f(x_t)$

$x_{t+1} = \frac{t+1}{t+3} y_{t+1} + \frac{2}{t+3} z_{t+1}$

CONVEX: $f(y) \ge f(x) + \nabla f(x)^T(y - x)$

$(\nabla f(y) - \nabla f(x))^T(y - x) \ge 0$

$\nabla^2 f(x) \succeq 0$

L-SMOOTH: f CONVEX, DIFFERENTIABLE

$f \le$ TANGENT + QUADRATIC

f BETWEEN TANGENT AND QUADRATIC

$(\forall x, y)\ f(y) \le f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|x - y\|^2$

HAS AT MOST QUADRATIC GROWTH

IFF ∇f IS L-LIPSCHITZ

$\|\nabla f(x) - \nabla f(y)\| \le L \cdot \|x - y\|$

MUST BE DIFFER.

STRONG CONVEXITY: f CONVEX, DIFFERENTIABLE

$(\forall x, y)\ f(y) \ge f(x) + \nabla f(x)^T(y - x) + \frac{\mu}{2}\|x - y\|^2$

f ABOVE THE QUADRATIC TANGENT

A FUNCTION CANNOT BE L-LIPSCHITZ AND STRONGLY CONVEX ON THE PLANE, ONLY IN A BOUNDED REGION

$$\boxed{\partial f(x) = \mathrm{conv}\left(\left\{\lim_{n\to\infty}\nabla f(x_n) \,\middle|\, \lim_{n\to\infty} x_n = x\right\}\right)}$$

PROJECTED GRADIENT DESCENT (GO IN CONSTRAINED SETTING)

$x_{t+1} = \underbrace{\Pi_X(y_{t+1})}_{\text{PROJECTION}} = \arg\min_{x\in X}\|x - y_{t+1}\|$

$(x - \Pi_X(y))\cdot(y - \Pi_X(y)) \le 0$ (NEXT ITER. IS AN IMPROVEMENT)

$\|x - \Pi_X(y)\|^2 + \|y + \Pi_X(y)\|^2 \le \|x - y\|^2$

IT'S ENOUGH FOR THE FUNCTION TO BE LIPSCHITZ IN THE REGION OF INTEREST

→ f DIFFERENT., B-LIPSCHITZ (SAME AS STOCHASTIC)

→ SMOOTH (CONVEX DIFFERENTIABLE L-SMOOTH), $\gamma = \frac{1}{L}$

$$\boxed{\frac{1}{2L}\|\nabla f(x_t)\|^2 - \frac{L}{2}\|y_{t+1} - x_{t-1}\|^2 \le f(x_t) - f(x_{t+1}), \quad t \ge 0}$$

$\boxed{f(x_T) - f(x^*) \le \frac{LR^2}{2T}, T > 0} \rightarrow$ SAME AS SUBGRADIENT

→ L-SMOOTH AND μ-STRONGLY CONVEX

x^* UNIQUE MINIMIZER

$\|x_{t+1} - x^*\|^2 \le \left(1 - \frac{\mu}{L}\right)\|x_T - x^*\|^2, t \ge 0$

$f(x_T) - f(x^*) \le \|\nabla f(x^*)\|\cdot\left(1 - \frac{\mu}{L}\right)^{\frac{T}{2}}\|x_0 - x^*\| + \frac{L}{2}\left(1 - \frac{\mu}{L}\right)^T\|x_0 - x^*\|^2$

FOR $T > 0$

LOWER BOUNDS

NESTEROV: $T < d - 1$, $\exists f$ ST. B-LIP. IN $\mathbb{R}^d$ WITH (SUB)GRADIENT HAVING ERROR $f(x_T) - f(x^*) \ge \frac{RB}{2(1 + \sqrt{T+1})}$

USING $\partial f(x)$

NEMIROVSKI, YUDIN: $T \le \frac{1}{2}(d - 1)$, $\exists f$ L-SMOOTH, FIRST ORDER METHOD HAS $f(x_T) - f(x^*) \ge \frac{3L\|x_0 - x^*\|^2}{32T^2}$

STOCHASTIC GRADIENT DESCENT

g_t UNBIASED ESTIMATOR FOR $\nabla f(x_t)$

$f(x_t) - f(x^*) \leq g_t^T (x_t - x^*)$ IN EXPECTATION

→ f CONVEX, DIFFERENTIABLE, B-LIPSCHITZ

$$\gamma = \frac{R}{B\sqrt{T}} \quad \frac{1}{T}\sum_{t=0}^{T-1}\left(E[f(x_t)] - f(x^*)\right) \leq \frac{RB}{\sqrt{T}}$$

SUBGRADIENT CONVERGES IN $O\left(\frac{1}{\varepsilon^2}\right)$

→ TAKE STRONG CONVEXITY (S. CONVEXITY WITH BOUNDED GRADIENTS)

$$\gamma_t = \frac{2}{\mu(t+1)}, \quad B^2 = \max_{t=1..T} E(\|g_t\|^2)$$

$$E\left[f\left(\frac{2}{T(T+1)}\sum_{t=1}^{T} t \cdot x_t\right) - f(x^*)\right] \leq \gamma_t B^2$$

SUBGRADIENT CONVERGES IN $O\left(\frac{1}{\varepsilon}\right)$

→ MINIBATCH (REDUCES VARIANCE $\rightarrow V_{NEW} = \frac{V_{OLD}}{m}$)

PICK $i_1 \dots i_m \sim U(1 \dots n)$

$$\tilde{g}_t = \frac{1}{m}\sum_{j=1}^{m} g_t^j, \quad g_t^j = \nabla f_{i_j}(x_t)$$

↳ HAS A SMALLER VARIANCE

$$x_{t+1} = x_t - \gamma \tilde{g}_t$$

- $m = 1 \rightarrow$ STOCASTIC
- $m = n \rightarrow$ GRADIENT DESCENT