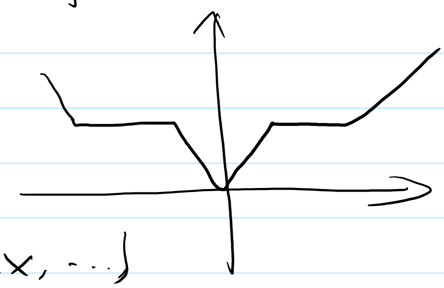
Lecture 15 Non-convex Optimization I Local Analysis

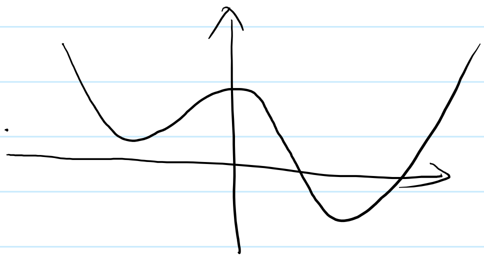
Sunday, October 23, 2016 10:59 PM

- Non-convex optimization

- what can a non-convex function look like?
  - simpler case
    still has a unique minimum.
    (quasi-convex, pseudo-convex, ...)



  - complicated case
    multiple local optima..



- optimality conditions
  - first order optimality condition

$$\nabla f(x) = 0$$

    - such points are called critical points.
    - for (strongly) convex function, $\nabla f(x) = 0 \Rightarrow x$ is optimal.

  - second order condition

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

e.g. $f(x) = x_1^2 + x_2^2$

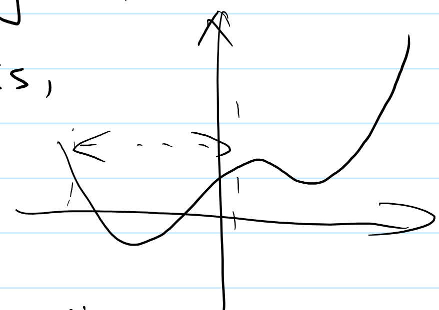
$$\nabla^2 f(x) = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \succeq 0.$$

  - saddle points
    $\nabla^2 f(x)$ is not positive semidefinite or negative semidefinite.

e.g. $f(x) = x_1^2 - x_2^2$

$$\nabla^2 f(x) = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$$

- Local convergence vs global convergence.
  - when multiple local minima exists, can hope to converge to global minimum with good initialization.



  - local structure of a non-convex function can behave like a convex function!

- Approximate Gradient Descent.
  - idea: after a good initialization, maybe the function is very similar to convex.
  - how to measure "similarity" to convex?
  - Consider gradient descent, initial z^0, goal z^*

    z^* is the optimum for convex function $f(z)$

    however, only has non-convex function $g(z)$

    hope: $g(z)$ close to $f(z)$

$$z^{t+1} = z^t - \eta \nabla g(z^t)$$

- Def: g is $(\alpha, \beta, \varepsilon)$-correlated if

$$\langle \nabla g(z^t), z^t - z^* \rangle \geq \alpha \|z^t - z^*\|^2 + \beta \|\nabla g(z^t)\|^2 - \varepsilon$$

note: if f is μ-strongly convex and L-smooth

$$\langle \nabla f(z^t), z^t - z^* \rangle \geq \frac{\mu L}{\mu + L} \|z^t - z^*\|^2 + \frac{1}{\mu + L} \|\nabla g(z^t)\|^2$$

$$\langle \nabla f(z^t), z^t - z^* \rangle \geq \frac{\mu L}{\mu + L} \|z^t - z^*\|^2 + \frac{1}{\mu + L} \|\nabla g(z^t)\|^2$$

$\left(\frac{\mu L}{\mu + L}, \frac{1}{\mu + L}, 0\right)$ - correlated.

Theorem: $\|z^{t+1} - z^*\|^2 \leq (1 - 2\alpha\eta) \|z^t - z^*\|^2 + 2\eta\varepsilon$, (when $\eta \leq 2\beta$), in particular

$$\|z^t - z^*\|^2 \leq (1 - 2\alpha\eta)^t \|z^0 - z^*\|^2 + \frac{\varepsilon}{\alpha}$$

Proof: $\|z^{t+1} - z^*\|^2 = \|z^t - z^*\|^2 - 2\eta \langle \nabla g(z^t), z^t - z^* \rangle + \eta^2 \|\nabla g(z^t)\|^2$

$$= \|z^t - z^*\|^2 - \eta\left(2\langle \nabla g(z^t), z^t - z^* \rangle - \eta \|\nabla g(z^t)\|^2\right)$$

$$\leq \|z^t - z^*\|^2 - \eta\left(2\alpha \|z^t - z^*\|^2 + (2\beta - \eta)\|\nabla g(z^t)\|^2 - 2\varepsilon\right)$$

$$\leq (1 - 2\alpha\eta) \|z^t - z^*\|^2 + 2\eta\varepsilon$$