- 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
    \[ \nabla^2 f(x) \succeq 0 \]
    - \( f(x) = x_1^2 + x_2^2 \)
    \[ \nabla^2 f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \succeq 0. \]

  - Saddle points
    \( \nabla^2 f(x) \) is not positive semidefinite or negative semidefinite.
\[ f(x) = x_1^2 - x_2^2 \]
\[ \nabla^2 f(x) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \]

- Local convergence vs. global convergence.
  - When multiple local minima exist, 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 \( \mu \)-strongly convex and \( \beta \)-smooth
  \[ \langle \nabla f(z^t), z^t - z^* \rangle \geq \frac{\mu L}{\mu + \beta} \| z^t - z^* \|^2 + \frac{1}{\mu + \beta} \| \nabla g(z^t) \|^2 \]
\[
\langle \nabla f(z^{t+1}), z^t - z^* \rangle \geq \frac{\mu L}{M+L} \| z^t - z^* \|^2 + \frac{1}{M+L} \| \nabla g(z^t) \|^2,
\]

\[
(\frac{\mu L}{M+L}, \frac{1}{M+L}, 0) \text{- correlated.}
\]

Theorem: \( \| z^{t+1} - z^* \|^2 \leq (1 - 2\alpha \eta) \| z^t - z^* \|^2 + 2\eta \zeta, \)

(when \( \eta \leq 2\beta \)), in particular

\[
\| z^t - z^* \|^2 \leq (1 - 2\alpha \eta)^t \| z^0 - z^* \|^2 + \frac{\zeta}{2}
\]

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

\[
= \| z^t - z^* \|^2 - \eta (\langle \nabla g(z^t), z^t - z^* \rangle - \eta \| \nabla g(z^t) \|^2) \]

\[
\leq \| z^t - z^* \|^2 - \eta (2\alpha \| z^t - z^* \|^2 + (2\beta - \eta) \| \nabla g(z^t) \|^2) \]

\[
\leq (1 - 2\alpha \eta) \| z^t - z^* \|^2 + 2\eta \zeta.
\]