Unconstrained optimization

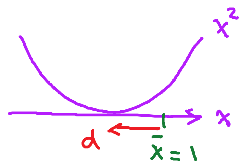
Let $f: \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Consider

$$(P) \qquad \inf_{x \in \mathbb{R}^n} f(x)$$

Recall: If $n=1$, then $\frac{df}{dx} = 0$ is a necessary condition for optimality.
How about general n?

Proposition: Let f be as above, $\bar{x} \in \mathbb{R}^n$ be arbitrary. If there exists a $d \in \mathbb{R}^n \setminus \{0\}$ s.t. $\nabla f(\bar{x})^T d < 0$, then $\exists\, \alpha_0 > 0$ s.t. $f(\bar{x} + \alpha d) < f(\bar{x})$ for all $\alpha \in (0, \alpha_0)$.

Here, d is called a descent direction of f at $\bar{x}$.



Corollary (First-order necessary optimality condition) $\nabla f(\bar{x}) = \frac{df}{dx}\Big|_{\bar{x}=1} = 2$

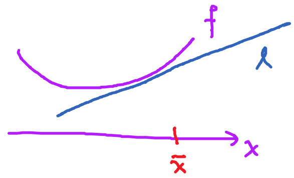
$d = -1$

Under the above setting, if $\bar{x}$ is a local minimum of f, then $\nabla f(\bar{x}) = 0$.

Proof: If $\nabla f(\bar{x}) \neq 0$, then take $d = -\nabla f(\bar{x})$ in the proposition.

Proposition: Let $f: \mathbb{R}^n \to \mathbb{R}$ be a continuous differentiable and convex function. Then, $\bar{x}$ is global minimizer $\iff \nabla f(\bar{x}) = 0$.

Proof: $(\Rightarrow)$ Follow directly from the corollary



$(\Leftarrow)$ Suppose that $\nabla f(\bar{x}) = 0$. Since f is convex,

$$f(x) \geq \underbrace{f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})}_{\ell} \quad \forall x$$

$$= f(\bar{x}).$$

This implies that $\bar{x}$ is a global minimizer.

Proposition (Second-order sufficient optimality condition)

Proposition (Second-order sufficient optimality condition)

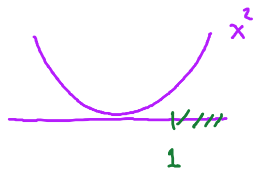
Let $f: \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. If $\nabla f(\bar{x}) = 0$, $\nabla^2 f(\bar{x}) \succ 0$, then $\bar{x}$ is a local minimizer.

Constrained Optimization

Examples:

1) $\inf \; f(x) = x^2$

s.t. $x \geq 1$

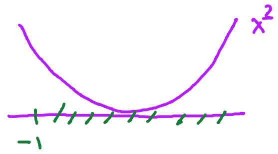


$\left.\frac{df}{dx}\right|_{\bar{x}=1} = 2 \neq 0$

$\frac{df}{dx} = 0 \Rightarrow x = 0$ infeasible

2) $\inf \; f(x) = x^2$

s.t. $x \geq -1$



$\bar{x} = 0$ optimal

$\left.\frac{df}{dx}\right|_{\bar{x}=0} = 0$

Let $f, g_1, \cdots, g_{m_1}, h_1, h_2, \cdots, h_{m_2} : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable functions. Consider

$$
(Q) \quad \begin{array}{ll} \inf & f(x) \\ \text{s.t.} & g_i(x) \leq 0, \quad i = 1, \cdots, m_1, \; (v_i) \\ & h_j(x) = 0, \quad j = 1, \cdots, m_2, \; (w_j) \\ & x \in \mathbb{R}^n \end{array} \Bigg\} \; S: \text{feasible region}
$$

Theorem (Karush-Kuhn-Tucker (KKT) necessary optimality condition)

Let $\bar{x} \in \mathbb{R}^n$ be a local minimum of (Q). Let

$$I = \{ i : g_i(\bar{x}) = 0 \}$$

be the active index set. Suppose that $\{\nabla g_i(\bar{x})\}_{i \in I} \cup \{\nabla h_j(\bar{x})\}_{j=1}^{m_2}$ are linearly independent. Then, $\exists$ Lagrange multipliers $v_1, \cdots, v_{m_1}$; $w_1, \cdots, w_{m_2} \in \mathbb{R}$ s.t.

$i \notin I \Rightarrow g_i(\bar{x}) < 0 \Rightarrow v_i = 0$

$$(*) \quad \nabla f(\bar{x}) + \sum_{i=1}^{m_1} v_i \nabla g_i(\bar{x}) + \sum_{j=1}^{m_2} w_j \nabla h_j(\bar{x}) = 0,$$

$\sum_{i \in I}$

$$(*) \qquad \nabla f(\bar{x}) + \sum_{i=1}^{m_1} v_i \nabla g_i(\bar{x}) + \sum_{j=1}^{m_2} w_j \nabla h_j(\bar{x}) = 0, \qquad \Rightarrow v_i = 0$$

$$(**) \qquad v_i \geq 0, \quad \underbrace{v_i\, g_i(\bar{x}) = 0}_{\text{complementarity}} \quad \forall i.$$

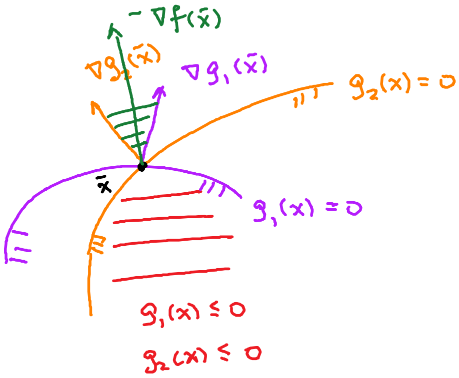
Illustration: For simplicity, suppose that there is no equality constraint.

Consider

$m_1 = 2$,

$m_2 = 0$

$I = \{1, 2\}$



From (*), (**),

$$-\nabla f(\bar{x}) = v_1 \nabla g_1(\bar{x}) + v_2 \nabla g_2(\bar{x})$$

$$v_1, v_2 \geq 0$$

Idea:

1) If $\bar{x}$ is a local minimum, then $\nexists\, d$ s.t. d is simultaneously a descent direction and a feasible direction.

2) To stay feasible, intuitively we need to decrease the value of g_1, g_2; e.g., find $d \in \mathbb{R}^n$ s.t. $\nabla g_i(\bar{x})^T d < 0$, $i = 1, 2$.

But for such d, from (*),

$$-\nabla f(\bar{x})^T d = \underbrace{v_1}_{\geq 0} \underbrace{\nabla g_1(\bar{x})^T d}_{< 0} + \underbrace{v_2}_{\geq 0} \underbrace{\nabla g_2(\bar{x})^T d}_{< 0} < 0 \qquad \text{assume } (v_1, v_2) \neq 0$$

$\Rightarrow \quad \nabla f(\bar{x})^T d > 0 \;\Rightarrow\; d$ is not a descent direction of f.