REVIEW OF CONVEX OPTIMIZATION*

Concepts from convex optimization such as Karush-Kuhn-Tucker (KKT) conditions were used in the previous sections of this lecture. In this section we give a brief introduction and derivation of these conditions.

Definition. *A set* $\mathcal{X} \in \mathbb{R}^n$ *is convex if*

$$\forall x_1, x_2 \in \mathcal{X},\ \forall \lambda \in [0,1],\ \lambda x_1 + (1-\lambda) x_2 \in \mathcal{X}.$$

A set is convex if, given any two points in the set, the line segment connecting them lies entirely inside the set.

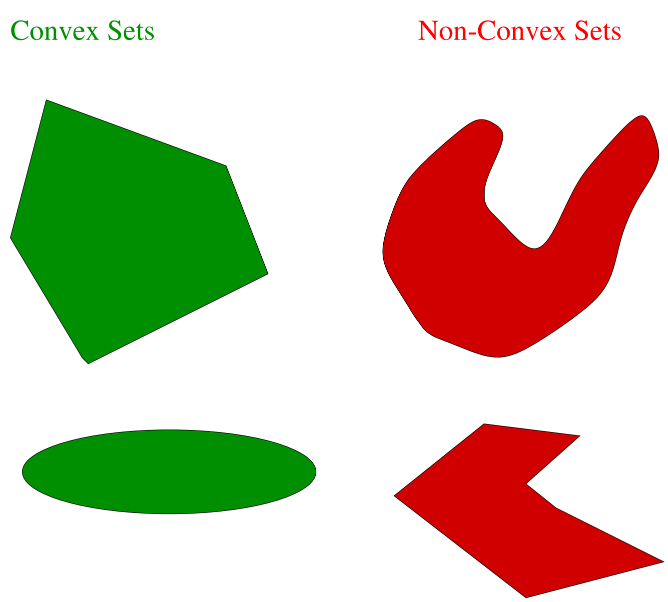

Figure 1. Examples of convex and nonconvex sets in $\mathbb{R}^2$.

Definition. *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is convex if:*

(1) *For any* x_1 *and* x_2 *in the domain of* f*, for any* $\lambda \in [0,1]$*,*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

(2) *The line segment connecting two points $f(x_1)$ and $f(x_2)$ lies entirely on or above the function f.*

(3) *The set of points lying on or above the function f is convex.*

A function is strictly convex if we replace "on or above" with "above", or replace "$\leq$" with $<$.

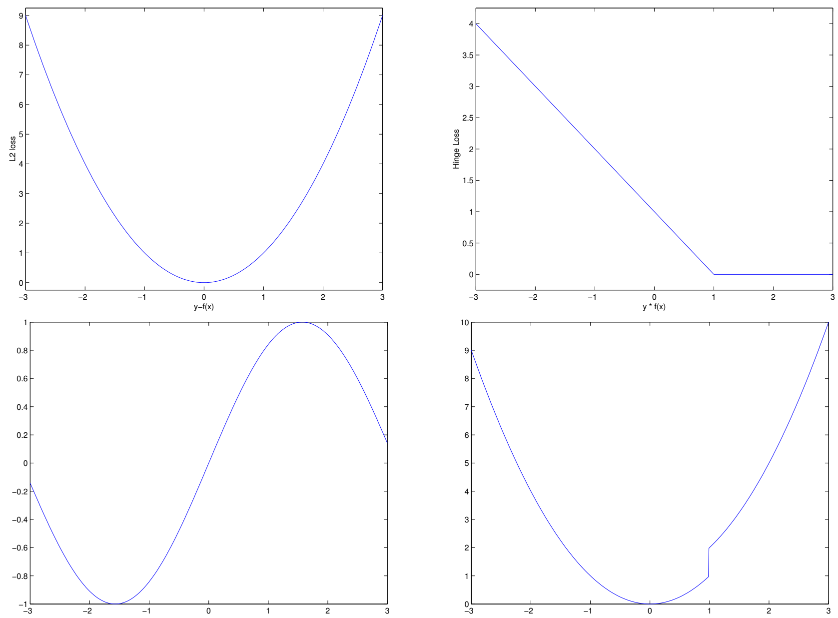

Figure 2. The top two figures are convex functions. The first function is strictly convex. Bottom figures are nonconvex functions.

Definition. *A point $\mathbf{x}^*$ is called a local minimum of f if there exists $\varepsilon > 0$ such that $f(\mathbf{x}^*) < f(\mathbf{x})$ for all $\|\mathbf{x} - \mathbf{x}^*\| \leq \varepsilon$.*

Definition. *A point $\mathbf{x}^*$ is called a global minimum of f if $f(\mathbf{x}^*) < f(\mathbf{x})$ for all feasible $\mathbf{x}$.*

Unconstrained convex functions (convex functions where the domain is all of $\mathbb{R}^n$) are easy to minimize. Convex functions are differentiable almost everywhere. Directional derivatives always exist. If we cannot improve our solution by moving locally, we are at the optimum. If we cannot find a direction that improves our solution, we are at the optimum.

Convex functions over convex sets (a convex domain) are also easy to minimize. If the set and the functions are both convex, if we cannot find a direction which we are able to move in which decreases the function, we are done. Local optima are global optima.

Example. *Linear programming is always a convex problem*

$$\begin{aligned} \min_{c} \quad & \langle c, x \rangle \\ \text{subject to:} \quad & Ax = b \\ & Cx \leq d. \end{aligned}$$

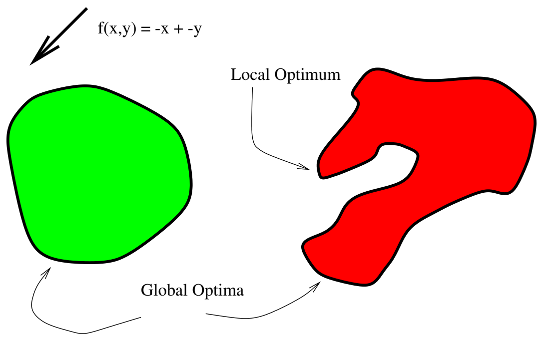

Figure 3. Optimizing a convex function of convex and nonconvex sets. In the example on the left the set is convex and the function is convex so a local minima corresponds to a global minima. In the example on the right the set is nonconvex and the function is convex one can find local minima that are not global minima.

Example. *Quadratic programming is a convex problem iff the matrix Q is positive semidefinite*

$$\begin{aligned} \min \quad & x'Qx + \langle c, x\rangle \\ \text{subject to :} \quad & Ax = b \\ & Cx \leq d. \end{aligned}$$

Definition. *The following constrained optimization problem P will be called the primal problem*

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{subject to :} \quad & g_i(\mathbf{x}) \geq 0 \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0 \quad i = 1, \ldots, n \\ & x \in \mathcal{X}. \end{aligned}$$

Here, f is our objective function, the g_i are inequality constraints, the h_i are equality constraints, and $\mathcal{X}$ is some set.

Definition. *We define a Lagrangian dual problem D:*

$$\begin{aligned} \max \quad & \Theta(\mathbf{u}, \mathbf{v}) \\ \text{subject to :} \quad & \mathbf{u} \geq 0 \end{aligned}$$

where $\Theta(\mathbf{u}, \mathbf{v}) := \inf \left\{ f(\mathbf{x}) - \sum_{i=1}^m u_i g_i(\mathbf{x}) - \sum_{j=1}^n v_i h_i(\mathbf{x}) : x \in \mathcal{X} \right\}$.

Theorem (Weak Duality)**.** *Suppose x is a feasible solution of P. Then $x \in \mathcal{X}, g_i(\mathbf{x}) \leq 0\ \forall i, h(\mathbf{x}) = 0\ \forall i$. Suppose $\mathbf{u}, \mathbf{v}$ are a feasible solution of D. Then for all $\mathbf{u} \geq 0$*

$$f(\mathbf{x}) \geq \Theta(\mathbf{u}, \mathbf{v}).$$

Proof.

$$
\begin{aligned}
\Theta(\mathbf{u},\mathbf{v}) &= \inf\left\{ f(\mathbf{y}) - \sum_{i=1}^{m} u_i g_i(\mathbf{y}) - \sum_{j=1}^{n} v_i h_i(\mathbf{y}) : y \in \mathcal{X} \right\} \\
&\leq f(\mathbf{x}) - \sum_{i=1}^{m} u_i g_i(\mathbf{x}) - \sum_{i=1}^{n} v_i h_i(\mathbf{x}) \\
&\leq f(\mathbf{x}).
\end{aligned}
$$

Weak duality says that every feasible solution to P is at least as expensive as every feasible solution to D. It is a very general property of duality, and we did not rely on any convexity assumptions to show it.

Definition. *Strong duality holds when the optima of the primal and dual problems are equivalent $Opt(P) = Opt(D)$.*

If strong duality, does not hold, we have the possibility of a duality gap. Strong duality is very useful, because it usually means that we may solve whichever of the dual or primal is more convenient computationally, and we can usually obtain the solution of one from the solution of the other.

Proposition. *If the objective function f is convex, and the feasible region is convex, under mild technical we have strong duality.*

We now look at a what are called saddle points of the Lagrangian function. We defined the Lagrangian function as the dual problem

$$
L(\mathbf{x},\mathbf{u},\mathbf{v}) = f(\mathbf{x}) - \sum_{i=1}^{m} \mathbf{u}_i \mathbf{g}_i(\mathbf{x}) - \sum_{j=1}^{n} \mathbf{v}_j \mathbf{h}_j(\mathbf{x}).
$$

A set $(\mathbf{x}^*,\mathbf{u}^*,\mathbf{v}^*)$ of feasible solutions to P and D is called a saddle point of the Lagrangian if

$$
L(\mathbf{x}^*,\mathbf{u},\mathbf{v}) \leq L(\mathbf{x},\mathbf{u},\mathbf{v}) \leq L(\mathbf{x},\mathbf{u}^*,\mathbf{v}^*) \ \forall x \in \mathcal{X}, \quad \forall \mathbf{u} \geq 0
$$

$\mathbf{x}^*$ minimizes L if $\mathbf{u}$ and $\mathbf{v}$ are fixed at $\mathbf{u}^*$ and $\mathbf{v}^*$, and $\mathbf{u}^*$ and $\mathbf{v}^*$ maximize L if $\mathbf{x}$ is fixed at $\mathbf{x}^*$.

Definition. *The points $(\mathbf{x}^*,\mathbf{u}^*,\mathbf{v}^*)$ satisfy the Karush Kuhn Tucker (KKT) conditions or are KKT points if they are feasible to P and D and*

$$
\begin{aligned}
\nabla f(\mathbf{x}^*) - \nabla \mathbf{g}(\mathbf{x}^*)'\mathbf{u}^* - \nabla \mathbf{h}(\mathbf{x}^*)'\mathbf{v} &= 0 \\
\mathbf{u}^* \mathbf{g}(\mathbf{x}^*) &= 0.
\end{aligned}
$$

In a convex, differentiable problem, with some minor technical conditions, points that satisfy the KKT conditions are equivalent to saddle points of the Lagrangian.