Convex Optimization
(for CS229)

Erick Delage, and Ashutosh Saxena

October 20, 2006

1 Convex Sets

Definition: A set $G \subseteq \Re^n$ is convex if every pair of point $(x, y) \in G$, the segment between x and y is in A. More formally:

$$\theta x + (1-\theta)y \in G \ , \ \forall \ x, y \in G \text{ and } 0 \leq \theta \leq 1 \ . \tag{1}$$

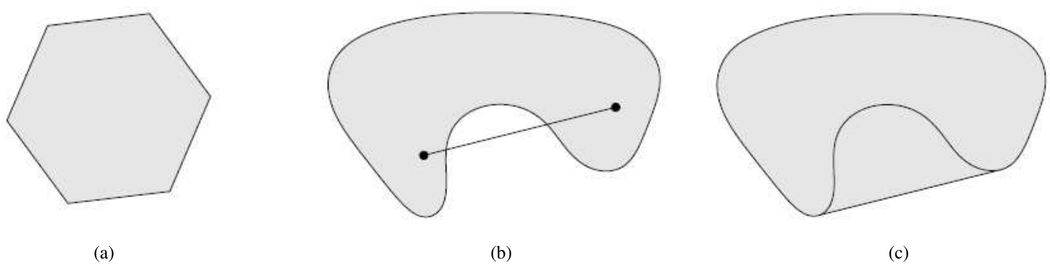


Figure 1: Examples of sets in 2d. (a) Convex polyotope. (b) Non-convex bean shape set. (c) Convex hull of bean shape set. Images taken from [1].

2 Convex Functions

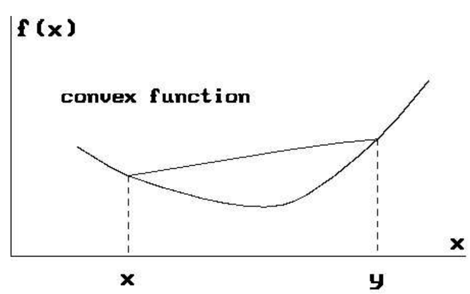


Figure 2: Convex Function

Definition: $f : \Re^n \rightarrow \Re$ is convex if domain of f is a convex set and

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \tag{2}$$

for all $x, y \in dom f$, $0 \leq \theta \leq 1$

A consequence of this definition is that any **Local minimum is Global minimum** for convex functions.

Examples of convex functions:

- affine: $ax + b$ on $\Re$
- affine: $a^T x + b$ on $\Re^n$
- exponential: e^{ax}
- powers: x^α on $x > 0$, for $\alpha \geq 1$ or $\alpha \leq 0$
- negative logarithm: $-\log x$, on $x \geq 0$
- norm: $||x||^2 = \sum_{i=1}^{n} x_i$
- quadratic: $x^2 + bx + c$

3 Conditions for Convexity

3.1 First order condition

For differentiable f, f is convex if and only if dom f is convex and

$$f(y) \geq f(x) + \frac{df(x)}{dx}(y - x) \tag{3}$$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \tag{4}$$

holds for all $x, y \in dom f$. This means the tangent (tangent plane for more than one dimensions) at any point is a lower bound for the function.

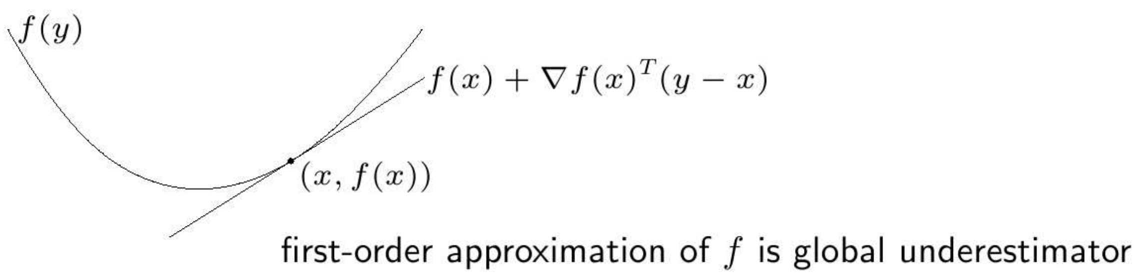


Figure 3: First Order Condition.

3.2 Second order condition

Hessian is defined as

$$H_{i,j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \tag{5}$$

f is convex **if and only if Hessian H is positive semidefinite** for all x. One dimensional equivalent says that curvature should be always positive. f(x) is convex if and only if

$$\frac{\partial^2 f(x)}{\partial x^2} \geq 0 \tag{6}$$

Examples: $f(x_1, x_2) = (x_1 - 1)^2 + x_2^2$ has Hessian, $H = [2,\ 0; 0,\ 2]$, therefore $f(x_1, x_2)$ is convex. $f(x_1, x_2) = x_1^2 + x_2^2 - 4x_1x_2$ is not convex, because Hessian is not positive semi-definite.

4 Maximizing Convex Functions

An optimization problem is said to be convex if it is equivalent to minimizing a convex objective function subject to the variable lying in a convex set. Equivalently, a Convex Optimization problem can be described in the following form:

$$\begin{aligned} \min_x \quad & f(x) && x \in \Re^n \\ \text{s.t.} \quad & g_i(x) \leq 0, && i = 1, ..., m \\ & Ax = b && \end{aligned} \tag{7}$$

where, $f, g_1, ..., g_m$ are convex.

The locally optimal point of a convex problem is (globally) optimal. E.g., minimizing with Stochastic Gradient descent or Newton's method, will give converge to optimal point for convex problems.

Examples:

$$\begin{aligned} \min_{x_1, x_2} \quad & (x_1 - 2)^2 + x_2^2 \\ \text{s.t.} \quad & x_2 \leq 0 \end{aligned} \tag{8}$$

Off the shelf software exist to solve some Convex Optimization problems. For example, if the functions are quadratic, a Quadratic Program (QP) solver maybe used. Similary, there is Linear Program (LP) for linear functions. (quadprog() and linprog() in Matlab)

5 Duality

The original problem can be transferred into another form, which is sometimes easier to solve.

5.1 Lagrangian

Consider the primal problem to be

$$\begin{aligned} \min_w \quad & f(w) && \\ \text{s.t.} \quad & g_i(w) \leq 0, && i = 1, ..., m \\ & h_i(w) = 0, && i = 1, ..., l \end{aligned} \tag{9}$$

Let the optimal value of the problem be p^*, at point w^* ($p^* = f(w^*)$). Lagrangian for this problem is defined as

$$L(w, \alpha, \beta) = f(w) + \sum_{i=1}^{m} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w) \tag{10}$$

where, α_i's and β_i's are the Lagrange multipliers. Lagrangian can be viewed as a weighted sum of objective and constraint functions. Note also that the following problem is equivalent to Problem (9):

$$\min_{w} \max_{\alpha \geq 0, \beta} L(w, \alpha, \beta) \ . \tag{11}$$

Define the Lagrange dual function as

$$\theta_D(\alpha, \beta) = \min_{w} L(w, \alpha, \beta) \tag{12}$$

5.2 Lower Bounds on Optimal value

$$\theta_D(\alpha, \beta) \leq p^* \tag{13}$$

when $\alpha \geq 0$. (Because from Eq. 10, g_i are negative).

Now, although θ_D is a lower bound, how good is it? We can maximize θ_D to get the best lower bound (possibly tight).

5.3 Lagrange Dual problem

To find the tightest bound, we can solve the Lagrange Dual problem:

$$\begin{aligned} \max_{\alpha, \beta} \quad & \theta_D(\alpha, \beta) \\ s.t. \quad & \alpha \geq 0 \end{aligned} \tag{14}$$

The optimal value of this problem is $d^* = \theta_D(\alpha^*, \beta^*)$. Since, $\theta_D(\alpha, \beta) \leq p^*$, therefore, $d^* \leq p^*$, this is referred as weak duality. If the optimization problem respects Slater's condition, we are assured that $d^* = p^*$ (strong duality). A convex problem respects Slater's condition if there exists a point $w \in dom f$ such that w is strictly feasible : $g_i(w) < 0$ for all i, and $h_i(w) = 0$ for all i.

5.4 KKT Conditions

Given that a convex optimization problem conforms with Slater's condition, we can find an optimal 3-tuple (w^*, α^*, β^*) using the Karush-Kuhn-tucker (KKT) conditions:

$$\frac{\partial}{\partial w_i} L(w^*, \alpha^*, \beta^*) = 0, \qquad i = 1, ...n \tag{15}$$

$$\frac{\partial}{\partial \beta_i} L((w^*, \alpha^*, \beta^*) = 0, \qquad i = 1, ...l \tag{16}$$

$$\alpha_i^* g_i(w^*) = 0, \qquad i = 1, ..., m \tag{17}$$

$$g_i(w^*) \leq 0, \qquad i = 1, ..., m \tag{18}$$

$$\alpha^* \geq 0, \qquad i = 1, ..., m \tag{19}$$

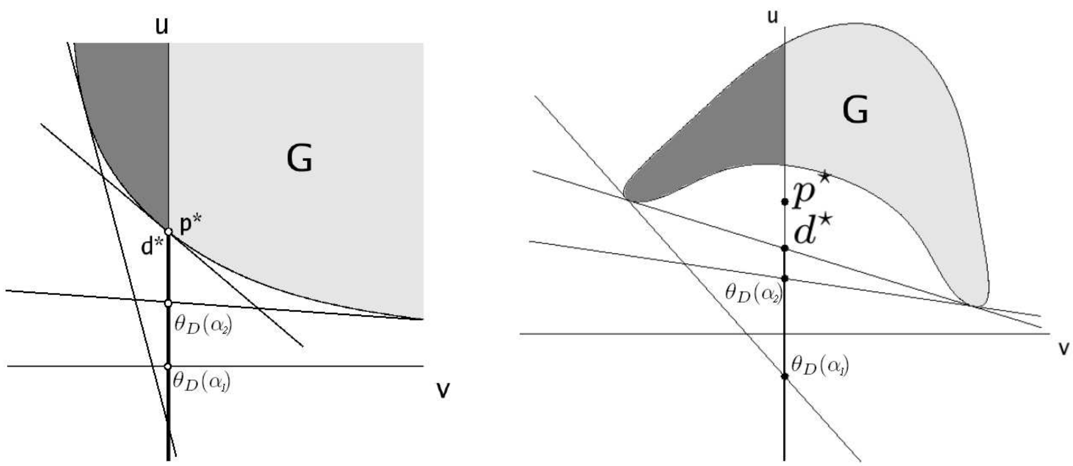
If some w^*, α^*, β^* satisfy KKT conditions, then w^* is an optimal point for the primal problem and (α^*, β^*) is optimal for the dual problem.

5.5 Understanding with a example

This section describes a simple case of minimizing $f(w)$ with constraint that $g(w) \leq 0$, where $w \in \Re^n$. Now,

$$L(w, \alpha) = f(w) + \alpha g(w) \quad \text{and} \quad \theta_D(\alpha) = \min_w L(w, \alpha) = \min_{(u,v) \in G} u + \alpha v \ ,$$

where $G = \{(u, v) | u = f(w), v = g(w) \text{ for some } w \in \Re^n\}$. Figure 5.5(a) and (b) show two examples of G, with v on x-axis and u on y-axis. Now, $-\alpha$ gives the slope of the line $\theta_D(\alpha) = \min_{(u,v) \in G} u + \alpha v$. The Dual problem is to change this line (by changing α) to find the maximum value of $\theta_D(\alpha)$. For example 1, $d^* = p^*$. However, for example 2, $d^* < p^*$.



(a) (b)

Figure 4: Examples of set G. The darker region in G indicates where $v \leq 0$ and therefore w is feasible. (a) Here, f and g are convex and $d^* = p^*$. (b) Here, either f or g is non-convex and $d^* < p^*$.

References

1. Convex Optimization by Boyd & Vandenberghe (Cambridge University Press 2004). Available Online: http://www.stanford.edu/ boyd/cvxbook, Online notes: http://www.stanford.edu/class/ee364

2. Convex Analysis by R.T. Rockarfeller (1970), Princeton University Press.