MATH 612
Computational methods for equation solving and function minimization – Week # 13

F.J.S.

Spring 2014 – University of Delaware

CONVEX FUNCTIONS ON CONVEX SETS

Convex sets and convex functions

A set C is convex when

$$x_0, x_1 \in C \qquad \Longrightarrow \qquad (1-\tau)x_0 + \tau x_1 \in C \quad \forall \tau \in (0,1).$$

We will care about closed convex sets. A function $f : C \to \mathbb{R}$ is convex in C when

$$f((1-\tau)x_0+\tau x_1) \le (1-\tau)f(x_0)+\tau f(x_1), \qquad x_0, x_1 \in C, \quad 0 < \tau < 1.$$

(The definition of strict convexity is not repeated.) Note that it does not make that much sense to talk about convex functions on non-convex domains (unless they are defined in a larger convex set). Why?

A new definition of stationary points

Let f be a $\mathcal{C}^1$ function on a convex set $C \subset \mathbb{R}^n$. If $\overline{x} \in C$ is a minimum, then

$$\nabla f(\overline{x}) \cdot (x - \overline{x}) \geq 0 \qquad \forall x \in C.$$

Proof. Consider the function

$$[0, 1] \ni \tau \mapsto \varphi(\tau) = f(\overline{x} + \tau(x - \overline{x})).$$

Then φ has a local minimum at $\tau = 0$ which implies that $\varphi'(0) \geq 0$. This is the condition in the statement.

A new definition of stationary points (cnt'd)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex $\mathcal{C}^1$ function. Then

$$f(x) - f(\overline{x}) \geq \nabla f(\overline{x}) \cdot (x - \overline{x}) \qquad \forall x, \overline{x} \in \mathbb{R}^n.$$

Assume now that $\overline{x} \in C$ (C being a convex set) satisfies

$$\nabla f(\overline{x}) \cdot (x - \overline{x}) \geq 0 \qquad \forall x \in C.$$

Then $\overline{x}$ is a global minimum of f in C.

One result

Let f be a convex function in a convex set C. If $\overline{x}$ is a local minimum, it is a global minimum. (Therefore, all local minima give the same optimal value.)

Proof. Let $x \in C$ and take $0 < \tau < 1$ small enough so that

$$f(\overline{x}) \leq f((1-\tau)\overline{x} + \tau x) \leq (1-\tau) f(\overline{x}) + \tau f(x).$$

(The smallness of τ is used in the first inequality. How?) Simplifying, it is clear that

$$f(\overline{x}) \leq f(x).$$

Another result

Let f be a convex function in a convex set C. The set of all minima is convex.

Proof. Let x_0, x_1 be two minima. The previous result shows that

$$f(x_0) = f(x_1).$$

Let then $x = (1-\tau)x_0 + \tau x_1$ and note that

$$\begin{aligned} f(x_0) &\le f(x) = f((1-\tau)x_0 + \tau x_1) \\ &\le (1-\tau)f(x_0) + \tau f(x_1) = f(x_0). \end{aligned}$$

Then

$$f(x_0) = f((1-\tau)x_0 + \tau x_1) \qquad \forall \tau \in (0,1),$$

so the points in the segment connecting two minima are also minima.

A consequence

If f is a strictly convex function on a convex set C, then the minimum of f is either unique or not attained. If the set C is closed and bounded, then the minimum is attained.

THE GRADIENT PROJECTION METHOD

Ingredients

- A closed bounded convex set C
- **The projection onto the set C**. This is a **nonlinear** operator $\mathcal{P} : \mathbb{R}^n \to C$ that solves the minimization problem:

$$\mathcal{P}x \in C, \qquad \|x - \mathcal{P}x\|_2 \leq \|x - z\|_2 \quad \forall z \in C.$$

- A convex function f.

The idea of the algorithm

If we are at the point x, we choose to descend in the gradient direction $w = -\nabla f(x)$. It's likely this will take us outside the feasible set C. Therefore, instead of doing a line search for the function

$$\varphi(t) = f(x + \tau w)$$

we do it for

$$\varphi(t) = f(\mathcal{P}(x + \tau w)).$$

We will use backtracking. Instead of trying to get

$$\varphi(\tau) < \varphi(0) + \beta\tau\varphi'(0)$$

(we might not have a differentiable φ because of $\mathcal{P}$), we will need to modify the right-hand side of this inequality.

Adapting the backtracking algorithm

Without restrictions, the inequality is

$$f(x+\tau w) < f(x) + \beta \nabla f(x) \cdot (x + \tau w - x)$$

or

$$f(y) < f(x) + \beta \nabla f(x) \cdot (y - x) \qquad y = x + \tau w.$$

Instead, we do

$$f(y) < f(x) + \beta \left(-\frac{\|y - x\|_2^2}{\tau} \right) \qquad y = \mathcal{P}(x + \tau w).$$

Gradient descent with projection and backtracking

for $\nu \geq 1$

$w = -\nabla f(x)$

$\varphi_0 = f(x)$

$\tau = \gamma$

$y = \mathcal{P}(x + \tau w)$

$\varphi_1 = f(y)$

$\psi_1 = \|y - x\|^2/\tau^2$

while $\varphi_1 > \varphi_0 + \tau\beta\psi_1$

$\tau = \tau\gamma$

$y = \mathcal{P}(x + \tau w)$

$\varphi_1 = f(y)$

$\psi_1 = \|y - x\|^2/\tau^2$

end

$x = x + \tau w$

stopping criterion

end

For you to think...

- If $C = \{x : \|x - x_0\| \leq R\}$, it is easy to compute the operator $\mathcal{P}$. How?
- If $C = [a_1, b_1] \times \ldots \times [a_n, b_n]$ is a bounded closed box, it is also easy to find $\mathcal{P}x$. How?
- If C is a rotated box, how would you compute $\mathcal{P}x$?

CONVEX CONSTRAINTS

Convex sets through constraints

Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be convex. Then the level sets

$$\{x \in \mathbb{R}^n \,:\, f_i(x) \leq b_i\}$$

are convex. Therefore, if $f_1, \ldots, f_k$ are convex, the set

$$\{x \in \mathbb{R}^n \,:\, f_i(x) \leq b_i \quad i = 1, \ldots, k\}$$

is convex. A particular case of convex set defined through constraints happens when

$$f_i(x) = a_i \cdot x \qquad a_i \in \mathbb{R}^n.$$

We saw how inequalities $Ax \leq b$ include also equalities, box-constraints, etc.

Active constraints

Let us consider the closed convex set

$$C = \{x \, : \, Ax \leq b\} = \{x \, : \, a_i \cdot x \leq b_i \quad i = 1, \ldots, k\}.$$

We say that the i-th constraint is active at a point $\overline{x} \in C$, when

$$a_i \cdot \overline{x} = b_i$$

and it is inactive when

$$a_i \cdot \overline{x} < b_i.$$

A mental exercise

In $\mathbb{R}^2$, find feasible closed convex polyhedral sets defined as

$$C = \{x \,:\, Ax \leq b\} = \{x \,:\, a_i \cdot x \leq b_i \quad i = 1, \ldots, k\}.$$

and points in the following situations:

- A feasible point where all constraints are inactive.
- A feasible point where all constraints are active.
- A feasible set where it is impossible to have a point with all constraints inactive.
- A feasible set where there is a constraint that is always inactive. (What happens to this constraint?)
- A feasible set with a constraint that is always active.

LAGRANGE MULTIPLIERS

Feasible sets through equality constraints

We consider the set

$$\begin{aligned} C &= \{x \in \mathbb{R}^n \,:\, f_i(x) = 0, \quad i = 1, \ldots, s\} \\ &= \{x \in \mathbb{R}^n \,:\, F(x) = 0\} \qquad F(x) = (f_1(x), \ldots, f_s(x))^T. \end{aligned}$$

The **standard constraint qualification** at $\overline{x} \in C$ is the condition:

$$\sum_i y_i \nabla f_i(\overline{x}) = 0 \qquad \Longrightarrow \qquad y_i = 0 \quad \forall i,$$

that is,

$$\{\nabla f_1(\overline{x}), \ldots, \nabla f_s(\overline{x})\} \quad \text{is linearly independent,}$$

that is,

$$\operatorname{rank} DF(\overline{x}) = s.$$

Necessary condition for optimality

Let $C = \{x \in \mathbb{R}^n \; : \; f_i(x) = 0 \quad i = 1, \ldots, s\}$ with $f_i \in \mathcal{C}^1$. Assume that $\overline{x}$ is a local minimum of $f \in \mathcal{C}^1$ in C and that the standard constraint qualification at $\overline{x}$ holds. Then there exists a vector $\overline{y} \in \mathbb{R}^s$ such that

$$\nabla f(\overline{x}) + \sum_{j=1}^{s} \overline{y}_j \nabla f_j(\overline{x}) = 0.$$

The components of the vector $\overline{y}$ are called **Lagrange multipliers**.

Sufficient condition and the Lagrangian

Let $(\overline{x}, \overline{y}) \in \mathbb{R}^n \times \mathbb{R}^s$ satisfy:

$$
\begin{aligned}
\nabla f(\overline{x}) + \sum_{j=1}^{s} \overline{y}_j \nabla f_j(\overline{x}) &= 0 \\
f_i(\overline{x}) &= 0 \qquad i = 1, \ldots, s.
\end{aligned}
$$

If f and $C = \{x \in \mathbb{R}^n : f_i(x) = 0 \quad i = 1, \ldots, s\}$ are convex, then $\overline{x}$ is a global minimum of f in C.
The Lagrangian is the function

$$L(x, y) = f(x) + \sum_{i=1}^{s} y_i f_i(x).$$

The above conditions are equivalent to

$$\nabla_x L(x, y) = 0, \qquad \nabla_y L(x, y) = 0.$$

A quadratic example

Minimize

$$f(x) = \frac{1}{2} x \cdot Ax - x \cdot b$$

subject to $Bx = c$.

- $\nabla f(x) = Ax - b$ (if A is symmetric, which we will assume)
- $f_i(x) = b_i \cdot x - c_i$, where b_i are the rows of B treated as colum vectors
- $\nabla f_i(x) = b_i$

The Lagrangian is

$$L(x, y) = \frac{1}{2} x \cdot Ax - x \cdot b + \sum_i y_i (b_i \cdot x - c_i) = \frac{1}{2} x \cdot Ax - x \cdot b + y \cdot (Bx - c)$$

A quadratic example (2)

$$L(x,y) = \frac{1}{2}x\cdot Ax - x\cdot b + \sum_i y_i(b_i\cdot x - c_i) = \frac{1}{2}x\cdot Ax - x\cdot b + y\cdot(Bx - c)$$

The Lagrangian equations are:

$$Ax - b + \sum_i y_i b_i = 0 \qquad \Longleftrightarrow \qquad Ax + B^T y = b$$

and

$$b_i \cdot x - c_i = 0 \quad \forall i \qquad \Longleftrightarrow \qquad Bx = c$$

or as a system

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix}$$

Checking the conditions

- The set $C = \{x : Bx = c\}$ is convex.
- Constraint qualification: the rank of B is s, that is, B has full rank by rows
- Convexity of f. If A is positive semidefinite, then f is convex. This is not necessary though, since we only need f to be convex over C. What is needed is the following property:

$$w \cdot Aw \geq 0 \qquad \text{for all } w \text{ satisfying } Bw = 0$$

Rmk. In these conditions, the matrix

$$\begin{bmatrix} A & B^T \\ B & 0 \end{bmatrix}$$

is symmetric, but it is indefinite.

KUHN-TUCKER CONDITIONS

Feasible sets through inequality constraints

Now

$$C = \{x \in \mathbb{R}^n \;:\; f_i(x) \leq 0, \quad i = 1, \ldots, s\}$$

Given $x \in C$, we define the active and inactive sets:

$$\begin{aligned} \mathcal{A}(x) &= \{i \;:\; f_i(x) = 0\} \\ \mathcal{I}(x) &= \{i \;:\; f_i(x) < 0\} \end{aligned}$$

The standard constraint qualification at x is slightly more difficult to define:

$$\left.\begin{array}{r} \displaystyle\sum_{i \in \mathcal{A}(x)} y_i \nabla f_i(x) = 0 \\ y_i \geq 0 \quad \forall i \end{array}\right\} \implies \quad y_i = 0 \qquad \forall i$$

For instance, linear independence of groups of gradients of the constraints that can be active at the same time implies the result.

Sufficient conditions

For $\mathcal{C}^1$ objective function and constraints, if $\bar{x}$ is a minimum at which the standard constraint qualification holds, then there exists a vector $\bar{y} \in \mathbb{R}^s$ satisfying

$$\begin{aligned} \nabla f(\bar{x}) + \sum_{i=1}^{s} \bar{y}_i \nabla f_i(\bar{x}) &= 0 \\ \bar{y}_i &\geq 0 \qquad \forall i \\ \bar{y}_i &= 0 \qquad i \in \mathcal{I}(\bar{x}) \end{aligned}$$

Another way of writing this is: there exist $\bar{y}_i, i \in \mathcal{A}(\bar{x})$ such that

$$\nabla f(\bar{x}) + \sum_{i \in \mathcal{A}(\bar{x})} \bar{y}_i \nabla f_i(\bar{x}) = 0, \qquad \bar{y}_i \geq 0.$$

The Kuhn-Tucker conditions

$$\nabla f(\overline{x}) + \sum_i \overline{y}_i \nabla f_i(\overline{x}) = 0$$

$$\overline{y}_i \geq 0 \qquad \forall i$$

$$f_i(\overline{x}) \leq 0 \qquad \forall i$$

$$\overline{y}_i f_i(\overline{x}) = 0 \qquad \forall i$$

The last group of inequalities are called complementarity conditions. They mean that for each index i either $f_i(\overline{x}) = 0$ (active constraint) or $\overline{y}_i = 0$, so Lagrange multipliers are only active on active constraints.

These conditions are also known as the Karush-Kuhn-Tucker (KKT) conditions. *KKT+ convexity implies optimality.*

A quadratic example again

The KKT conditions for the problem of minimizing

$$\tfrac{1}{2}x \cdot Ax - x \cdot b \qquad A \text{ is symmetric}$$

subject to $Bx \leq c$, are

$$Ax + B^T y = b,$$

and

$$y \geq 0, \qquad Bx \geq c \qquad y \odot (Bx - c) = 0$$

where $\odot$ is the element by element product of two vectors. Constraint qualification *is implied* by linear independence of groups of rows of B such that the planes $b_i \cdot x = c_i$ intersect.