| | |
|---|---|
| **Mathematical Programming II** | **Lecture 1** |
| **ORIE 6310 Spring 2014** | **January 23, 2011** |
| **Scribe: Mike Todd** | |

In 1963, G.B. Dantzig wrote his classic text entitled "Linear Programming and Extensions." ORIE 6300 covered linear programming: this course treats a variety (different from those addressed by Dantzig) of extensions.

Topics will probably include the following (but this will possibly change!):

- A very brief review/introduction to the KKT conditions for $\min\{f(x) : g(x) \leq 0\}$, $f$, $g$ smooth;
- The linear complementarity problem (LCP):

$$w = Mz + q, \quad w \geq 0, \quad z \geq 0, \quad w \cdot z = 0$$

  - applications to linear and quadratic programming, and to Nash equilibria of bimatrix games
  - complementary pivot algorithms
- Efficiency of pivoting algorithms:
  - diameter of polyhedra
  - expected number of pivots;
- Informational complexity of convex programming $\min\{f(x) : x \in C\}$, $f$, $C$ convex, $f$ nonsmooth:
  - nonsmooth convex optimization, Lagrangian relaxation
  - lower bounds
  - the method of centers of gravity, the ellipsoid method, and methods for high dimensions
- Applications of the ellipsoid method:
  - semidefinite programming
  - equivalence of separation and optimization
- Some examples of modelling using structured nonlinear programming.

Throughout we will be interested in efficiency/complexity issues.

**Work involved**: There will be occasional homework sets (a total of around 4) and a final exam (probably 24-hour take-home). Each student in this course is expected to abide by the Cornell University Code of Academic Integrity: see `http://cuinfo.cornell.edu/Academic/AIC.html`. In particular, I will provide strict guidelines for how much you can consult with other students on the homeworks, and of course the final can only be discussed with me.

**The KKT optimality conditions: introduction/review**

We are interested in the constrained nonlinear programming problem:

$$
\begin{array}{lrl}
 & \min_x & f(x) \\
(P) & g(x) \leq & 0,
\end{array}
$$

where $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^m$ are continuously differentiable. This means that, for every $x \in \mathbf{R}^n$,

$$
\begin{aligned}
\nabla f(x) &:= \left( \frac{\partial f(x)}{\partial x_1}; \frac{\partial f(x)}{\partial x_2}; \ldots; \frac{\partial f(x)}{\partial x_n} \right), \\
\nabla g_i(x) &:= \left( \frac{\partial g_i(x)}{\partial x_1}; \frac{\partial g_i(x)}{\partial x_2}; \ldots; \frac{\partial g_i(x)}{\partial x_n} \right),
\end{aligned}
$$

for each $i$ exist and vary continuously with $x$. (Here we use the Matlab notation that $(u_1; u_2; \ldots; u_k)$ denotes a column vector; if we use commas instead of semicolons, we obtain a row vector.) The key consequence is that then $f(x + \tau d)$ and $g_i(x + \tau d)$ are continuously differentiable functions of $\tau \in \mathbf{R}$ for fixed $x, d \in \mathbf{R}^n$, with

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau} f(x + \tau d)|_{\tau=0} &= \nabla f(x)^T d, \\
\frac{\mathrm{d}}{\mathrm{d}\tau} g_i(x + \tau d)|_{\tau=0} &= \nabla g_i(x)^T d
\end{aligned}
$$

for each $i$. We write $\nabla g(x) = [\nabla g_1(x), \nabla g_2(x), \ldots, \nabla g_m(x)] \in \mathbf{R}^{n \times m}$.

Suppose we are given a feasible point $\bar{x}$ for $(P)$. We wish either to "improve" it (find another feasible point with lower objective) or to show that it satisfies some sort of optimality condition. Moreover, we should do this based on quantities we can easily evaluate. We take these to be $f(\bar{x})$, $\nabla f(\bar{x})$, $g(\bar{x})$, $\nabla g(\bar{x})$. We suppose that an "oracle" (like a black box) gives us these values. Hence we have some local information about the functions defining $(P)$, but not global information. This is a key difference between (general) nonlinear programming problems and their linear counterparts. With this information we can build 1st-order Taylor approximations of the functions. We need an extra condition for these data to provide an accurate local picture of the feasible region.

**Definition 1** *Given $\bar{x} \in \mathbf{R}^n$ with $g(\bar{x}) \leq 0$, the* **active constraint indices** *are those of the tight constraints, i.e., those $i$ with $g_i(\bar{x}) = 0$. We let $I(\bar{x}) := \{i : 1 \leq i \leq m, g_i(\bar{x}) = 0\}$ be the set of these. We say the* **Mangasarian-Fromovitz constraint qualification (MFCQ) holds at** $\bar{x}$ *if there is no nontrivial nonnegative linear dependence among the $\nabla g_i(\bar{x}), i \in I(\bar{x})$.*

Here is the important consequence:

**Lemma 1** *Suppose the MFCQ holds at $\bar{x}$. Then there is some $\hat{d} \in \mathbf{R}^n$ with $\nabla g_i(\bar{x})^T \hat{d} < 0$ for all $i \in I(\bar{x})$.*

**Proof:** Consider the linear programming problem

$$\begin{array}{rrcl} \min_\lambda & -\sum_{i\in I(\bar{x})} \lambda_i & & \\ & \sum_{i\in I(\bar{x})} \nabla g_i(\bar{x})\lambda_i & = & 0, \\ & \lambda_i & \geq & 0,\, i \in I(\bar{x}). \end{array}$$

By the MFCQ, there is no feasible solution with negative objective value, while $\lambda = 0$ is a feasible solution with zero objective. Hence there is an optimal solution, and by linear programming duality, there is a feasible solution $\hat{d}$ to the dual problem. Hence

$$\nabla g_i(\bar{x})^T \hat{d} \leq -1,\, i \in I(\bar{x}).$$

This completes the proof. □

**Definition 2** *We say that $\bar{x}$ is a* **local minimizer** *for $(P)$ if it is feasible and, for some positive $\epsilon$, $\|x - \bar{x}\| \leq \epsilon$ and $g(x) \leq 0$ imply that $f(x) \geq f(\bar{x})$, i.e., no sufficiently close feasible point has a better objective value.*

To try to find such an improved point, consider the direction-finding subproblem

$$\begin{array}{lrrcl} & \min_d & \nabla f(\bar{x})^T d & & \\ (DFSP) & & \nabla g(\bar{x})^T d & \leq & -g(\bar{x}). \end{array}$$

Note that, except for the omission of the constant term $f(\bar{x})$, we are trying to minimize the 1st-order Taylor approximation of $f(\bar{x} + d)$ subject to requiring that the 1st-order Taylor approximations of all $g_i(\bar{x} + d)$'s be nonpositive. This problem is a linear programming problem, a linearization of $(P)$. (If we wanted to use this to generate a good direction, we might want to add extra constraints like $-1 \leq d_j \leq 1$ for all $j$ to ensure boundedness, but we don't need this for our purposes.) By considering $(DFSP)$, we get the fundamental result:

**Theorem 1** *(First-order necessary conditions for local optimality: Karush (1939) and Kuhn and Tucker (1951)) Suppose that $\bar{x}$ is a local minimizer for $(P)$ and that MFCQ holds at $\bar{x}$. Then there are Lagrange multipliers $\bar{u} \in \mathbb{R}^m$ such that the following (KKT) conditions hold:*

$$\nabla f(\bar{x}) + \nabla g(\bar{x})\bar{u} = 0, \quad g(\bar{x}) \leq 0, \quad \bar{u} \geq 0, \quad \bar{u}_i g_i(\bar{x}) = 0,\, i = 1, \ldots, m. \tag{1}$$

(The last conditions ensure that $\bar{u}_i > 0$ only if $i$ is the index of an active constraint. So the first condition says that the gradient of the objective function plus a nonnegative combination of the gradients of the active constraints gives zero.)

**Proof:** Consider $(DFSP)$. Suppose there is a feasible solution $\bar{d}$ with a negative objective value. Using $\hat{d}$ from the lemma, choose $\delta > 0$ so that $d := \bar{d} + \delta\hat{d}$ satisfies $\nabla f(\bar{x})^T d < 0$. We also have $\nabla g_i(\bar{x})^T d < 0$ for each $i \in I(\bar{x})$. Then it is easy to see that $x := \bar{x} + \tau d$, for all sufficiently small positive $\tau$, is feasible with $f(x) < f(\bar{x})$. This contradicts $\bar{x}$ being a local minimizer. (Note that, because of nonlinearity of the constraints, this may not hold if we only have $\nabla g_i(\bar{x})^T d \leq 0$: we see why a constraint qualification may be necessary.)

Thus there is no such $\bar{d}$, so that $d = 0$ is optimal in

$$
\begin{array}{llcl}
 & \min_d & \nabla f(\bar{x})^T d & \\
(DFSP) & & -\nabla g(\bar{x})^T d & \geq \; g(\bar{x}).
\end{array}
$$

(Notice that the constraint has been multiplied by $-1$.) Hence there is an optimal solution $\bar{u}$, with objective value 0, to its dual. So we have

$$
-\nabla g(\bar{x})\bar{u} = \nabla f(\bar{x}), \quad \bar{u} \geq 0, \quad \bar{u}^T g(\bar{x}) = 0.
$$

Since $\bar{x}$ is feasible, $g(\bar{x}) \leq 0$. The final equation then implies complementary slackness, i.e., $\bar{u}_i g_i(\bar{x}) = 0$, all $i$. Thus we have the KKT conditions. $\square$

**Remark 1** *It is unpleasant to have to require a constraint qualification. Without this requirement, a weaker set of conditions, the Karush-John conditions, hold at a local minimizer. For these, an additional nonnegative multiplier $\bar{v}$ is introduced multiplying the objective function gradient in the first condition, and we insist that $\bar{v}$ and $\bar{u}$ are not both zero.* **Exercise***: Construct an example ($\mathbf{R}^2$ is sufficient) where the unique global (and local) minimizer does not satisfy the KKT conditions.*

**Remark 2** *In the case of linear constraints, no constraint qualification is necessary, as can be seen from the proof: we can move a small distance in any direction $d$ with $\nabla g_i(\bar{x})^T d \leq 0$ for all $i \in I(\bar{x})$.* **Exercise***: formulate a weakened form of the MFCQ that can be used when some but not all of the constraints are linear, and check that when this holds, the KKT conditions are necessary for local optimality.*

**Remark 3** *Clearly, if $(P)$ is a linear programming problem, the KKT conditions are a combination of linear equations and inequalities and complementary slackness conditions. But this also holds for quadratic programming problems, where the constraints are linear but the objective function may be quadratic.*