LP. Kap. 17: Interior-point methods

- the simplex algorithm moves along the boundary of the polyhedron P of feasible solutions
- an alternative is interior-point methods
- they find a path in the interior of P, from a starting point to an optimal solution
- for large-scale problems interior-point methods are usually faster
- we consider the main idea in these methods

1. The barrier problem

Consider the LP problem

$$\begin{array}{lrcl} \max & c^T x & & \\ \text{s.t.} & & & \\ & Ax & \leq & b, \\ & x & \geq & O \end{array}$$

and its dual

$$\begin{array}{lrcl} \min & b^T y & & \\ \text{s.t.} & & & \\ & A^T y & \geq & c, \\ & y & \geq & O \end{array}$$

Introduce slack variables w in the primal and (negative) slack z in the dual, which gives

Primal (P):

$$\begin{array}{rrcl} \max & c^T x & & \\ \text{s.t.} & & & \\ & Ax + w & = & b, \\ & x, w & \geq & O \end{array}$$

Dual (D):

$$\begin{array}{rrcl} \min & b^T y & & \\ \text{s.t.} & & & \\ & A^T y - z & = & c, \\ & y, z & \geq & O \end{array}$$

- We want to rewrite the problems such that we eliminate the constraints $x, w \geq O$ og $y, z \geq O$, but still avoid negative values (and 0) on the variables!!
- This is achieved by a logarithmic barrier function, and we get the following *modified* primal problem

The barrier problem:

$$(P_\mu): \quad \begin{array}{ll} \max & c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i \\ \text{s.t.} & \\ & Ax + w = b \end{array}$$

- (P_μ) is not equivalent to the original problem (P), but it is an approximation
- it contains a parameter $\mu > 0$.
- remember: $x_j \rightarrow 0^+$ implies that $\log x_j \rightarrow -\infty$.
- (P_μ) is a nonlinear optimization problem
- interpretation/geometry: see Figure 17.1 in Vanderbei: level curves for f_μ, polyhedron P, central path when $\mu \rightarrow 0$.
- Goal: shall see that (P_μ) has a unique optimal solution $x(\mu)$ for each $\mu > 0$, and that $x(\mu) \rightarrow x^*$ when $\mu \rightarrow 0^+$, where x^* is the unique optimal solution of (P). (Note: w is uniquely determined by x)

2. Lagrange multiplier

From (for instance) T. Lindstrøm, "Optimering av funksjoner av flere variable", MAT1110, multivariable calculus) we have the following Lagrange multiplier rule:

Theorem *Assume $U \subseteq \mathbb{R}^n$ is open, and that $f, g_i : U \to \mathbb{R}$ are functions with continuous partial derivatives ($i \leq m$), and let $b_1, \ldots, b_m \in \mathbb{R}$. Assume that x^* is a local maximum (or minimum) for f on the set $S = \{x \in \mathbb{R}^n : g_i(x) = b_i \ \ (i \leq m)\}$, and that $\nabla g_1(x^*), \ldots, \nabla g_m(x^*)$ are linearly independent. Then there are constants $\lambda_1, \ldots, \lambda_m$ such that*

$$(*) \qquad \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*).$$

- λ_i's are called Lagrange multipliers
- This is a necessary optimality condition and leads to $n + m$ equations for finding x and λ ($n + m$ variables).

This can also be expressed by the Lagrange function (we redefine the function g_i by $g_i := g_i - b_i$, such that we now consider $g_i(x) = 0$):

$$L(x, y) = f(x) - \sum_{i=1}^{m} y_i g_i(x).$$

Then $(*)$ says that

$$\nabla_x L(x^*, y) = O$$

while the constraints $g_i(x^*) = 0$ $(i \leq m)$ become (where $y = \lambda$)

$$\nabla_y L(x^*, y) = O.$$

These equations are called the first-order optimality conditions and a solution x^* is called a critical point.

Are these conditions also sufficient for optimality?

Consider the Hessian matrix

$$H_f(x) = [\frac{\partial^2 f(x)}{\partial x_i \partial x_j}] \in \mathbb{R}^{n \times n}$$

Note: $f(x) = o(g(x))$ when $x \to 0$ means $\lim_{x \to 0} f(x)/g(x) = 0$

Theorem 17.1 *Let the g_i's be linear functions, and assume x^* is a critical point. Then x^* is a local maximum if*

$$z^T H_f(x^*) z < 0$$

for each $z \neq O$ satisfying $z^T \nabla g_i(x^*) = 0$ $(i \leq m)$.

Proof: Second order Taylor formula gives

$$f(x^* + z) = f(x^*) + \nabla f(x^*)^T z + (1/2) z^T H_f(x^*) z + o(\|z\|^2)$$

where z is a perturbation from the point x^*. To preserve feasibility z must be chosen such that $x^* + z$ satisfies the constraints, i.e., $z^T \nabla g_i(x^*) = 0$ $(i \leq m)$. But, since x^* is a critical point

$$\nabla f(x^*)^T z = \left(\sum_{i=1}^{m} \lambda_i \nabla g_i(x^*)\right)^T z = 0$$

so the assumption $(z^T H_f(x^*) z < 0$ for ...$)$ and Taylor's formula give that $f(x^* + z) \leq f(x^*)$, so x^* is a local maximum. $\square$

3. Lagrange applied to the barrier problem

The barrier problem (P_μ):

$$\begin{aligned} \max \quad & c^T x + \mu \textstyle\sum_j \log x_j + \mu \sum_i \log w_i \\ \text{s.t.} \quad & \\ & Ax + w = b \end{aligned}$$

Introduce the Lagrange function

$$L(x, w, y) = c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i + y^T (b - Ax - w)$$

First-order optimality condition becomes

$$\frac{\partial L}{\partial x_j} = c_j + \mu \frac{1}{x_j} - \textstyle\sum_i y_i a_{ij} = 0 \quad (j \leq n)$$

$$\frac{\partial L}{\partial w_i} = \mu \frac{1}{w_i} - y_i = 0 \quad (i \leq m)$$

$$\frac{\partial L}{\partial y_i} = b_i - \textstyle\sum_j a_{ij} x_j - w_i = 0 \quad (i \leq m)$$

Notation: write X for the diagonal matrix with the vector x on the diagonal. e is the vector with only 1's.

Then first order optimality conditions become, in matrix form:

$$\begin{aligned} A^T y - \mu X^{-1} e &= c \\ y &= \mu W^{-1} e \\ Ax + w &= b \end{aligned}$$

Introduce $z = \mu X^{-1} e$ and we obtain (1.OPT)

$$\begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ z &= \mu X^{-1} e \\ y &= \mu W^{-1} e \end{aligned}$$

We had:

$$\begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ z &= \mu X^{-1} e \\ y &= \mu W^{-1} e \end{aligned}$$

Multiply the third equation by X and the fourth with W and we get

$$(**) \qquad \begin{aligned} Ax + w &= b \\ A^T y - z &= c \\ XZe &= \mu e \\ YWe &= \mu e \end{aligned}$$

The last two equations say: $x_j z_j = \mu$ $(j \leq n)$ and $y_i w_i = \mu$ $(i \leq m)$ which is μ-complementarity (approximative complementary slack). These are nonlinear. In total we have $2(n+m)$ equations and the same number of variables.

It is quite simple:

- Interior-point methods (at least this type) consist in solving the equations $(**)$ approximately using Newton's method for a sequence of μ's (converging to 0).

4. Second order information

We show: *if* there is a solution of the opt. condition $(**)$, then it must be unique! We use Theorem 17.1 and consider the barrier function $f(x, w) = c^T x + \mu \sum_j \log x_j + \mu \sum_i \log w_i$

First derivative:

$$\frac{\partial f}{\partial x_j} = c_j + \frac{\mu}{x_j} = 0 \quad (j \le n)$$

$$\frac{\partial f}{\partial w_i} = \frac{\mu}{w_i} \qquad (i \le m)$$

Second derivative:

$$\frac{\partial^2 f}{\partial x_j^2} = -\frac{\mu}{x_j^2} \quad (j \le n)$$

$$\frac{\partial^2 f}{\partial w_i^2} = -\frac{\mu}{w_i^2} \quad (i \le m)$$

So the Hessian matrix is a diagonal matrix with negative diagonal elements: this matrix is negative definite. Uniqueness then follows from Teorem 17.1.

5. Existence

Theorem 17.2 *There is a solution of the barrier problem if and only if both the primal feasible region and the dual feasible region have a nonempty interior.*

Proof: Shall show the "if"-part.

- Assume there is a $(\bar{x}, \bar{w}) > O$ such that $A\bar{x} + \bar{w} = b$ (relative interior point in the (x, w)-space), and $(\bar{y}, \bar{z}) > O$ with $A^T\bar{y} - \bar{z} = c$.
- Let (x, w) be primal feasible. Then

$$\bar{z}^T x + \bar{y}^T w = (A^T y - c)^T x + \bar{y}^T (b - Ax) = b^T \bar{y} - c^T x.$$

so

$$c^T x = -\bar{z}^T x - \bar{y}^T w + b^T \bar{y}$$

- The barrier function f becomes

$$\begin{aligned} f(x, w) &= c^T x + \mu \textstyle\sum_j \log x_j + \mu \sum_i \log w_i \\ &= \textstyle\sum_j (-\bar{z}_j x_j + \mu \log x_j) + \sum_i (-\bar{y}_i w_i + \mu \log w_i) + b^T \bar{y} \end{aligned}$$

- The terms in each sum has the form $h(v) = -av + \mu \log v$ where $a > 0$ and $0 < v < \infty$ and this function has a unique maximum in μ/a and tends to $-\infty$ as $v \to \infty$. This implies that the set $\{(x, w) : f(x, w) \geq \delta\}$ is bounded for each δ.
- Let now $\delta = \bar{f} = f(\bar{x}, \bar{w})$ and define the set

$$\begin{aligned} \bar{P} &= \{(x, w) : Ax + w = b,\ x \geq O,\ w \geq O, \} \\ &\quad \cap \{(x, w) : x > O,\ w > O,\ f(x, w) \geq \bar{f}\}. \end{aligned}$$

  Then $\bar{P}$ is closed. Because: $\bar{P}$ is an intersection between two closed sets; the last set is closed as f is continuous (that the domain $\{(x, w) : x > O,\ w > O\}$ is not closed does not matter here.)
- Therefore $\bar{P}$ is closed and bounded, i.e., compact. $\bar{P}$ is also nonempty (it contains $(\bar{x}, \bar{w})$). By the extreme value theorem f attains its supremum on $\bar{P}$, and therefore also on $\{(x, w) : Ax + w = b,\ x > O,\ w > O\}$ as desired.

$\square$

We then obtain (using an exercise saying that the dual has an interior point when the primal feasible region is bounded):

Corollary 17.3 *If the primal feasible region has interior points and is bounded, then for each $\mu > 0$ there exists a unique solution*

$$(x(\mu), w(\mu), y(\mu), z(\mu))$$

of $(**)$.

We then get a path (curve) $p(\mu) := \{(x(\mu), w(\mu), y(\mu), z(\mu)) : \mu > 0\}$ in $\mathbb{R}^{2(m+n)}$ which is called the primal-dual central path.

In the primal-dual path following method one computes a sequence $\mu^{(1)}, \mu^{(2)}, \ldots$ converging to 0, and for each $\mu^{(k)}$ one *approximately* solves the nonlinear system of equations $(**)$ using Newton's method. The corresponding sequence $p(\mu^{(k)})$ will then converge towards an optimal optimal primal-dual solution.
A more precise result on this convergence, and more details, are found in Chapter 18 and 19 (not syllabus).

Example: A problem with $m = 40$ and $n = 100$. We show ℓ_2-norm of the the residuals for each iteration: (primal) $\rho = b - Ax - w$; (dual) $\sigma = c - A^T y + z$; (compl.slack.) $\gamma = z^T x + y^T w$. We find an optimal solution.

Iter.	primal	dual	KS
2	189.61190	124.81236	103.89923
4	117.87500	77.59142	49.26126
6	81.95498	53.94701	30.11503
8	55.11458	36.27926	18.64561
10	30.92967	20.35951	9.75917
12	10.05169	6.61654	3.24588
14	4.37507	2.87990	1.52481
16	1.62442	1.06928	0.59844
18	0.64896	0.42718	0.25285
20	0.37908	0.24953	0.15657
22	0.14284	0.09402	0.06359
24	0.11378	0.07490	0.05191
26	0.00145	0.00095	0.00413
28	0.00000	0.00000	0.00005
30	0.00000	0.00000	0.00000