Modern Optimization

Jean-Luc Bouchot

School of Mathematics and Statistics
Beijing Institute of Technology
jlbouchot@bit.edu.cn

Spring 2021

Outline

Definition 2.1

Let $f_0 : \mathbb{R}^d_{>0} \to \mathbb{R}$ *be defined as*

$$f_0(x) := \sum_{j=1}^{m} c_j \left(\prod_{i=1}^{d} x_i^{a_{ij}} \right)$$

for some coefficients $c_j > 0$ *and some real exponents* a_{ij}. f_0 *is called a* **posynomial**.

Definition 2.2

A **geometric programming problem** *is an minimization problem in which the objective function is a posynomial and constrained to positive variables.*

Remark 2.1

For the minimization problem to be meaningful, we require some of the a_{ij} to be negative.

Definition 2.3 (DGP)

The **dual geometric program** *to a GP problem is the following optimization problem*

$$\text{Maximize}_{\delta \in \mathbb{R}^m}\, g_0(\delta) \tag{2.1}$$

$$\textit{subject to } \sum_{j=1}^{m} \delta_j = 1 \tag{2.2}$$

$$\textit{and } \sum_{j=1}^{m} a_{ij}\delta_j = 0 \tag{2.3}$$

where we define

$$g_0(x) := \prod_{j=1}^{m} \left(\frac{c_j}{\delta_j} \right)^{\delta_j}.$$

If we cannot find such a δ we say that the DGP is **inconsistent**.

Definition 2.4 (Duality gap)

Assume both the primal problem (GP) and the dual problem (DGP) have solutions and let p^* and d^* denote these solutions respectively. The **duality gap** *is the difference between the two optimal values: $p^* - d^*$.*

Theorem 2.1

If $x^* > 0$ minimizes $f_0(x)$, then DGP is consistent. Moreover

$$\delta_j^* := \frac{u_j(x^*)}{f_0(x^*)}$$

is feasible and solves the DGP.
Finally, the duality gap is 0, i.e.

$$f_0(x^*) = g_0(\delta^*).$$

Remark 2.2

The solution is implemented in two steps:

1. Find a solution to the dual problem (MART: Multiplicative algebraic reconstruction technique)
2. Remembering that $\delta_j^* = \frac{u_j(x^*)}{g_0(\delta^*)}$, we can solve for x^* (probably use some logs to help here!)

Remark 2.3 (MART Algo – setup)

The MART algorithm is used to find nonnegative solutions to linear systems

$$Ax = b,$$

where $A \in \mathbb{R}^{I \times J}$ has nonnegative entries and b has positive entries. Moreover, we request that

$$s_j = \sum_{i=1}^{I} A_{i,j} > 0.$$

Definition 2.5 (The MART Algo)

The MART algorithm is an iterative algo with the following steps:

1. *Start with a positive vector* x^0
2. *Iterate starting with* $k = 0$*, setting* $i = k(\text{mod } I) + 1$.
3. *Iterate* $x_j^{k+1} = x_j^k \left(\frac{b_i}{(Ax^k)_i}\right)^{A_{i,j}/m_i}$*, for all* $1 \le j \le J$ *and where* $m_i = \max\{A_{i,j} | 1 \le j \le J\}$.
4. *Increment* k*:* $k \leftarrow k + 1$

Definition 2.6 (KL Divergence)

The (non-symmetric) Kullback-Leibler divergence between two positive numbers a and b is defined as

$$KL(a, b) = a \log(a/b) + b - a.$$

It is extended to 0 numbers as $KL(a, 0) = +\infty$ and $KL(0, b) = b$.
For J dimensional vectors, it is defined as

$$KL(a, b) = \sum_{j=1}^{J} KL(a_j, b_j).$$

Proposition 2.1 (admitted)

Let x^0 be a given non negative vector and A be a consistent matrix for the MART algorithm. Then the sequence of iterates $\{x^k\}_k$ converges to the nonnegative solution of $b = Ax$ which minimizes $KL(x, x^0)$.

Remark 2.4

Since the a_{ij} may be negative, we cannot apply MART directly:

1. *Lift* all the equations in the system as much as needed adding a constant times the last row of the matrix. This generates a system $B\delta = \widetilde{b}$.
2. Solve $B\delta = \widetilde{b}$ which is equivalent to finding a solution to $A\delta = b$. (why?)
3. Set the starting point $\delta^0 = c = (c_1, \cdots, c_m)^T$.

Remark 2.5

δ^k converges to a solution which minimizes

$$\begin{aligned} KL(\delta, c) &= \sum_{i=1}^{m} \delta_i \log\left(\frac{\delta_i}{c_i}\right) + c_i - \delta_i \\ &= -\sum_{i=1}^{m} \delta_i + \sum_i c_i - \sum_{i=1}^{m} \delta_i \log\left(\frac{c_i}{\delta_i}\right) \\ &= -1 + \sum_i c_i - \log\left(g_0(\delta)\right). \end{aligned}$$

Example 2.1 (in-class)

Find the minimum to the following function in 3 dimensions
$f_0(x) = \frac{40}{x_1 x_2 x_3} + 20x_1x_3 + 10x_1x_2 + 40x_2x_3$ for strictly positive entries.

1. Compute the first 2-3 iterations of MART
2. Assuming MART converges to $\delta^* = (0.4, 0.2, 0.2, 0.2)^T$, what are the optimal values for x^* and the minimal cost?

Exercise 2.1

Minimize the function $f_0(x, y) = \frac{1}{xy} + xy + x + y$ for $x > 0$ and $y > 0$. (You might need some help of a computer.)