Numerical Optimization – Tutorial on ADMM

L. Desbat & F. Iutzeler

Let f and g be respectively $\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ and $\mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ convex functions. We are interested in the problem

$$(1) \qquad \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^m} f(x) + g(z) \quad \text{s.t.} \quad Mx = z$$

with M a full rank $m \times n$ real matrix. We assume that this problem has a solution i.e. $\exists x^\star : f(x^\star) + g(x^\star) < +\infty$. The objective of this session is to investigate a method, called the *Alternating Direction Method of Multpliers* (ADMM) to solve (1) by using only f OR g at once and not $f + g$. This method has grown tremendously popular in distributed computing and in image or signal processing.

A. Augmented Lagrangian and Method of Multipliers

First, to simplify let us take $M = I$. The *augmented Lagrangian* for this problem writes

$$\mathcal{L}_\rho(x, z; \lambda) = f(x) + g(z) + \langle\lambda; x - z\rangle + \frac{\rho}{2}\|x - z\|^2$$

for some positive parameter ρ.

Exercise 1 (Lagrangian Properties)**.**

a. Show that $\mathcal{P}_\rho(x, z) = \sup_{\lambda\in\mathbb{R}^n} \mathcal{L}_\rho(x, z; \lambda) < +\infty$ if and only if $x = z$.
b. Deduce that $(x^\star, z^\star) \in \arg\min_{x,z} \mathcal{P}_\rho(x, z)$ are solutions of Prob (1). Is minimizing $\mathcal{P}_\rho$ easier than solving the original problem?
c. Let $\mathcal{D}_\rho(\lambda) = \min_{x,z} \mathcal{L}_\rho(x, z; \lambda)$. Let $\lambda^\star = \arg\sup_\lambda \mathcal{D}_\rho(\lambda) \in \mathbb{R}\cup\{-\infty, +\infty\}$, show that $(x^\star, z^\star) \in \arg\min_{x,z} \mathcal{L}_\rho(x, z; \lambda^\star)$ are solutions of Prob (1). Is minimizing $\mathcal{L}_\rho(\cdot, \cdot; \lambda^\star)$ easier than solving the original problem?

This first considerations told us that completely solving *primal* $\mathcal{P}_\rho$ or *dual* $\mathcal{D}_\rho$ lead to problems as hard as the original one; this is due to the fact the sought point $(x^\star, z^\star; \lambda^\star)$ is a *saddle point* of $\mathcal{L}_\rho$. To find such a point, the *Method of Multipliers* was introduced; it consists in a gradient ascent over $\mathcal{D}_\rho$:

$$(x_{k+1}, z_{k+1}) \in \arg\min_{x,z} \mathcal{L}_\rho(x, z; \lambda_k)$$
$$\lambda_{k+1} = \lambda_k + \rho\nabla\mathcal{D}_\rho(\lambda_k)$$

Exercise 2 (Method of Multipliers)**.**

a. How does $\nabla\mathcal{D}_\rho(\lambda_k)$ writes in function of x_{k+1}, z_{k+1}?
b. Show that for λ fixed, $\mathcal{L}_\rho(x, z; \lambda)$ is strongly convex. Which is the strong convexity parameter?
c. It is possible to show that if $\mathcal{L}_\rho(x, z; \lambda)$ is μ-strongly convex in (x, y) then $\mathcal{D}_\rho$ is $1/\mu$-smooth. Deduce that the Method of Multipliers converge to a point that solves Problem (1).
d. Conclude regarding the interest of augmenting the Lagrangian.

B. Alternating Direction Method of Multipliers

We saw that the Method of Multipliers was a big step forward as it enables to minimize constrained problems more easily; however, it does not solve our problem of minimizing $f+g$ by using only f OR g at once. That is the interest of the *Alternating Direction Method of Multipliers* (ADMM). It consists in approximating the arg min step of the Method of Multpliers into two separate arg mins:

$$\begin{aligned} x_{k+1} &\in \arg\min_x \mathcal{L}_\rho(x, z_k; \lambda_k) \\ z_{k+1} &\in \arg\min_z \mathcal{L}_\rho(x_{k+1}, z; \lambda_k) \\ \lambda_{k+1} &= \lambda_k + \rho \nabla_\lambda \mathcal{L}_\rho(x_{k+1}, z_{k+1}; \lambda_k) \end{aligned}$$

The ADMM converges to a point $(x^\star, z^\star; \lambda^\star)$ such that $(x^\star, z^\star)$ is solution to Problem (1) for any convex, lower-semicontinuous functions f, g and any $\rho > 0$.

Exercise 3 (ADMM)**.** Show that the iterations of ADMM for Problem (1) write

$$\begin{aligned} x_{k+1} &= \arg\min_x \left\{ f(x) + \frac{\rho}{2}\|Mx - z_k + \lambda_k/\rho\|^2 \right\} \\ z_{k+1} &= \arg\min_z \left\{ g(z) + \frac{\rho}{2}\|Mx_{k+1} - z + \lambda_k/\rho\|^2 \right\} \\ \lambda_{k+1} &= \lambda_k + \rho(Mx_{k+1} - z_{k+1}) \end{aligned}$$

Exercise 4 (Parallel Computing)**.** Suppose that you want to solve $\min_x F(x) := \sum_{i=1}^n f_i(x)$ over n machines that can communicate through a master, each of which have the knowledge of only one f_i (F is known nowhere).

a. Reformulate the problem as a constrained problem where each f_i is linked to a copy x_i of a global variable z. Put it in the form of Problem 1.
b. Apply ADMM to the previous problem. Do you have a parralel algorithm?