CONSERVATIVE SIGNAL PROCESSING ARCHITECTURES RESULTING IN THE METHOD OF MULTIPLIERS

Thomas A. Baran

Initial draft: November 2, 2014
Latest revision: December 5, 2015

ABSTRACT

In this memo we demonstrate that the framework for generating optimization algorithms described in [2]-[3] can be used to automatically generate the method of multipliers. The method of multipliers is thereby one specific class of algorithms among the more general classes of algorithms that the framework is known to generate.

1. INTRODUCTION

The method of multipliers describes a class of algorithms used to solve optimization problems of the following form:

$$\begin{aligned} \underset{\underline{x}}{\text{minimize}} \quad & f(\underline{x}) \\ \text{subject to} \quad & A\underline{x} = \underline{b}. \end{aligned} \tag{1}$$

The method of multipliers specifically describes an associated algorithm that may be written in imperative pseudocode as:

$$\underline{x}[k+1] \quad := \quad \underset{\underline{x}}{\operatorname{argmin}}\, L_\rho\left(\underline{x}, \underline{y}[k]\right) \tag{2}$$

$$\underline{y}[k+1] \quad := \quad \underline{y}[k] + \rho\left(A\underline{x}[k+1] - \underline{b}\right), \tag{3}$$

where $L_\rho(\underline{x}, \underline{y})$ indicates a so-called *augmented Lagrangian*, defined for this class of problems as

$$L_\rho(\underline{x}, \underline{y}) = f(\underline{x}) + \underline{y}^T(A\underline{x} - \underline{b}) + (\rho/2)||A\underline{x} - \underline{b}||_2^2. \text{ [1]} \tag{4}$$

In this memo we show that the framework for generating optimization algorithms described in [2]-[3] can be used to automatically generate the method of multipliers exactly. The method of multipliers is thereby one specific class of algorithms among the more general classes of algorithms that the framework is known to generate.

2. KEY RESULT

The basic strategy used in [2]-[3] is to generate optimization algorithms by beginning with associated stationarity conditions as in [4], and applying a change of coordinates to these conditions, admitting an iteration that is known to converge. In [2]-[3] one such set of transformations is selected for which an arbitrary strongly-convex cost function with bounded gradient was known to result in an algorithm having linear convergence.

The key result described in the abstract is depicted graphically in Fig. 1. The basic idea is to create an iteration that couples the stationarity conditions associated with the following two problems:

$$\begin{aligned} \underset{\underline{x},\underline{q},\underline{e}}{\text{minimize}} \quad & f(\underline{x}) + \frac{1}{2}||\underline{q}||_2^2 \\ \text{subject to} \quad & \underline{q} = A\underline{x} - \underline{b} \\ & \underline{e} = A\underline{x} - \underline{b} \end{aligned} \tag{5}$$

and

$$\begin{aligned} \underset{\underline{w},\underline{e}}{\text{minimize}} \quad & 0 \\ \text{subject to} \quad & \underline{w} = \underline{e} \\ & \underline{w} = 0. \end{aligned} \tag{6}$$

The coupling of separable stationarity conditions associated with various separable components of optimization problems is a recurring theme in [2]-[3].

Referring to Fig. 1, the associated stationarity conditions for problems (5) and (6) are depicted graphically in Fig. 1(a), using the notation outlined in [4]. Given a sufficiently smooth function $f(\cdot)$ in (5), the function $g(\cdot)$ depicted in Fig. 1(a) is known to be the gradient of the convex conjugate of $f(\cdot)$, although smoothness of $f(\cdot)$ is not specifically required in justifying the key result. The transformed stationarity conditions, transformed using the specific transformations described in [2]-[3], are depicted in Fig. 1(b), where they have been coupled together with a delay element consistent with [2]-[3].

Fig. 1(c) depicts the result of performing straightforward algebraic manipulations to the system illustrated in Fig. 1(b). Referring to Fig. 1(c), the stationarity conditions associated with minimizing the augmented Lagrangian in Eq. 4 become readily apparent, resulting in an overall iteration that reduces exactly to the iteration described in Eqns. 2-3, for $\rho = 2$. Alternative values of ρ may effectively be realized by appropriately scaling $f(\underline{x})$ in the original formulation of the problem (1).

3. REFERENCES

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, 2011.

[2] T. A. Baran and T. A. Lahlou, “Conservative signal processing architectures for asynchronous, distributed optimization part I: General framework,” in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, Dec 2014, pp. 35–39.

[3] T. A. Baran and T. A. Lahlou, “Conservative signal processing architectures for asynchronous, distributed optimization part II: Example systems,” in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, Dec 2014, pp. 40–44.

[4] T. A. Baran, *Conservation in Signal Processing Systems*, Ph.D. thesis, Massachusetts Institute of Technology, 2012.

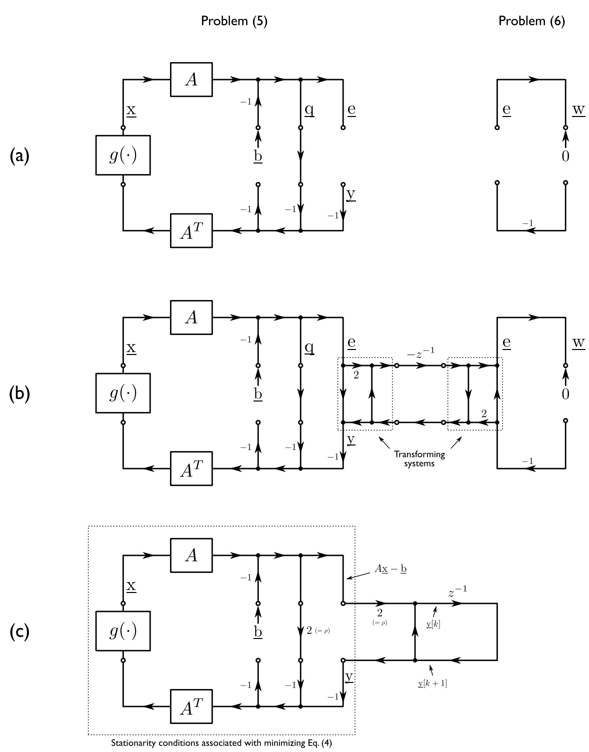

Fig. 1. Manipulations used in illustrating the key result.