

# Corrections

Trung Vu

Department of CSEE, University of Maryland, Baltimore County, MD 21250, USA

trungvv@umbc.edu

May 4, 2024

## 1 Equation (18)

In the paper, the augmented Lagrangian is given by

$$\mathcal{J}_n^c(\mathbf{W}) \triangleq H(y_n) - 2 \log |\mathbf{d}_n^H \mathbf{w}_n| - \frac{1}{2\gamma_n} \left( (\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\})^2 - \mu_n^2 \right), \quad (1)$$

where  $\mu_n$  is a Lagrangian multiplier and  $\gamma_n$  is a **positive** scalar learning parameter.  $h_n(\mathbf{w}_n, \mathbf{r}_n)$  corresponds to the inequality constraint  $h_n(\mathbf{w}_n, \mathbf{r}_n) = \rho_n - |\mathbf{w}_n^\top \mathbf{r}_n| \leq 0$ . Also, we note that the goal here is to minimize  $\mathcal{J}_n^c(\mathbf{W})$  *w.r.t.*  $\mathbf{W}$ . In the following, we review the augmented Lagrangian method and show that there is a **sign issue** for the Lagrangian term in (1).

### 1.1 Augmented Lagrangian Method

The augmented Lagrangian method, a.k.a. the method of multipliers, is used to handle the inequality constraints as follows. Consider the general setting of a constrained minimization problem

$$\min f(\mathbf{x}) \quad \text{subject to } g_j(\mathbf{x}) \leq 0, \text{ for } j = 1, \dots, m, \quad (2)$$

where  $f(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $g_j(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}$ . Let us define the augmented Lagrangian as

$$\mathcal{L}_\gamma(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \frac{1}{2\gamma} \sum_{j=1}^m \left( (\max(0, \mu_j + \gamma g_j(\mathbf{x})))^2 - \mu_j^2 \right), \quad (3)$$

where  $\boldsymbol{\mu} \in \mathbb{R}^m$  is the Lagrange multiplier and  $\gamma > 0$  is a scalar penalty parameter. The iterative equations to minimize (3) are given by

$$\begin{cases} \mathbf{x}^{i+1} = \operatorname{argmin}_{\mathbf{x}} \mathcal{L}_\gamma(\mathbf{x}, \boldsymbol{\mu}^i) \\ \boldsymbol{\mu}^{i+1} = \max(\mathbf{0}, \boldsymbol{\mu}^i + \gamma \mathbf{g}(\mathbf{x}^{i+1})) \end{cases}, \quad (4)$$

where the operators in the second update are element-wise. It can be shown [1] that for sufficiently large  $\gamma$ , the solution of (3) coincides with the solution of (2).

### 1.2 Correction of the Sign Issue in Equation (18)

Comparing the minimization in (3) versus the minimization in (1), we see that the sign of the last term with  $\gamma_n$  is incorrect in (1). If one would like to minimize  $\mathcal{J}_n^c(\mathbf{W})$ , the augmented Lagrangian function should be defined as

$$\min \mathcal{J}_n^c(\mathbf{W}) \triangleq H(y_n) - 2 \log |\mathbf{d}_n^H \mathbf{w}_n| + \frac{1}{2\gamma_n} \left( (\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\})^2 - \mu_n^2 \right). \quad (5)$$

## 2 Equation (19)

In the paper, the gradient of  $\mathcal{J}_n^c(\mathbf{W})$  *w.r.t.*  $\mathbf{w}_n$  is given by

$$\frac{\partial \mathcal{J}_n^c}{\partial \mathbf{w}_n^*} = \frac{\partial H(y_n)}{\partial \mathbf{w}_n^*} - \frac{\mathbf{d}_n^H}{\mathbf{d}_n^H \mathbf{w}_n} - \mu_n (\mathbf{r}_n (\mathbf{w}_n^H \mathbf{r}_n)^* h'_n(\mathbf{w}_n, \mathbf{r}_n)), \quad (6)$$

where  $h_n(\mathbf{w}_n, \mathbf{r}_n) = \rho_n - |\mathbf{w}_n^H \mathbf{r}_n|^2$  and  $h'_n$  is the derivative of  $h_n$  *w.r.t.*  $(\mathbf{w}_n^H \mathbf{r}_n)$ . Ignoring the sign issue mentioned in the previous section and the transposition of the column vector to row vector, we focus on the derivation of the gradient of the Lagrangian term:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}_n} \left( \frac{1}{2\gamma_n} \left( (\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\})^2 - \mu_n^2 \right) \right) &= \frac{1}{2\gamma_n} \frac{\partial}{\partial \mathbf{w}_n} (\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\})^2 \\ &= \frac{1}{2\gamma_n} (2 \max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\}) \frac{\partial}{\partial \mathbf{w}_n} (\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\}) \\ &= \frac{\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\}}{\gamma_n} \mathbb{I}_{\gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n > 0} \frac{\partial}{\partial \mathbf{w}_n} (\gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n) \quad (\text{since } (\partial \max\{0, x\})/\partial x = \mathbb{I}_{x>0}) \\ &= \frac{\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\}}{\gamma_n} \frac{\partial h_n(\mathbf{w}_n, \mathbf{r}_n)}{\partial \mathbf{w}_n} \quad (\text{absorbing } \mathbb{I}_{\gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n > 0} \text{ into the max}) \\ &= \max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\} \frac{\partial h_n(\mathbf{w}_n, \mathbf{r}_n)}{\partial (\mathbf{w}_n^H \mathbf{r}_n)} \frac{\partial (\mathbf{w}_n^H \mathbf{r}_n)}{\partial \mathbf{w}_n} \\ &= \max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\} h'(\mathbf{w}_n, \mathbf{r}_n) \mathbf{r}_n. \end{aligned} \quad (7)$$

Note that one key difference between (6) and (7) is the term  $\mu_n$  is replaced by its updated value in the next iteration  $\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\}$  (see 4). This typo, however, does not affect the correctness of the implementation since  $\mu_n$  is updated before  $\Delta \mathbf{w}_n$  (see Steps 7 and 8 of Algorithm 2). Another mistake in (6) is that the derivative of  $h_n$  *w.r.t.*  $(\mathbf{w}_n^H \mathbf{r}_n)$  already includes the term  $(\mathbf{w}_n^H \mathbf{r}_n)^*$ . Thus, it should not appear in the final expression.

## References

- [1] Dimitri P Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic press, 2014.