Corrections

Trung Vu

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

trungvv@umbc.edu

May 4, 2024

1 Equation (6)

In the paper, the augmented Lagrangian is given by

$$\mathcal{J}(\boldsymbol{W}) \triangleq \mathcal{J}_{\text{IVA}} - \sum_{l=1}^{L} \frac{1}{2\gamma_l} \sum_{m=1}^{M} \Big\{ \big[\max\{0, \mu_l^{[m]} + \gamma_l g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{d}_l)\}\big]^2 - (\mu_l^{[m]})^2 \Big\}, \tag{1}$$

where $\mu_l^{[m]}$ is a Lagrangian multiplier and γ_l is a positive scalar penalty parameter. $g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{d}_l)$ corresponds to the inequality constraint $g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{d}_l) = \rho_l - \epsilon(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{d}_l) \leq 0$. Also, we note that the goal here is to minimize $\mathcal{J}(\boldsymbol{W})$ *w.r.t.* $\boldsymbol{W}$. 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

$$\min f(\boldsymbol{x}) \quad \text{subject to } g_j(\boldsymbol{x}) \leq 0, \text{ for } i = j, \ldots, m, \tag{2}$$

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

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

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

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

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

1.2 Correction of the Sign Issue in Equation (6)

Comparing the minimization in (3) versus the minimization in (1), we see that the sign of the last term with γ_l is incorrect in (1) and there should be only one penalty parameter γ instead of L parameters $\gamma_1, \ldots, \gamma_L$. If one would like to minimize $\mathcal{J}(\boldsymbol{W})$, the augmented Lagrangian function should be defined as

$$\min \mathcal{J}(\boldsymbol{W}) \triangleq \mathcal{J}_{\text{IVA}} + \frac{1}{2\gamma} \sum_{l=1}^{L} \sum_{m=1}^{M} \Big\{ \big[\max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{d}_l)\}\big]^2 - (\mu_l^{[m]})^2 \Big\}. \tag{5}$$

2 Step 10 in Algorithm 1

In the paper, the gradient of $\mathcal{J}(\boldsymbol{W})$ *w.r.t.* $\boldsymbol{w}_l^{[m]}$ is given by

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{w}_l^{[m]}} = \frac{\partial J_{\mathrm{IVA}}}{\partial \boldsymbol{w}_l^{[m]}} - \frac{1}{\gamma_n}\Big\{\big[\max\{0, \gamma\big(\hat{\rho}_n - \epsilon(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{d}_l)\big) + \mu_n^{[m]}\}\big]^2 - (\mu_n^{[m]})^2\Big\}, \tag{6}$$

Ignoring the sign issue mentioned in the previous section, we focus on the derivation of the gradient of the Lagrangian term:

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{w}_l^{[m]}}\left(\frac{1}{2\gamma}\big[\max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{d}_l)\}\big]^2 - (\mu_l^{[m]})^2\right) &= \frac{1}{2\gamma}\frac{\partial}{\partial \boldsymbol{w}_l^{[m]}}\big(\max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)\}\big)^2 \\
&= \frac{1}{2\gamma}(2\max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)\})\frac{\partial}{\partial \boldsymbol{w}_l^{[m]}}\big(\max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)\}\big) \\
&= \frac{\max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)\}}{\gamma}\mathbb{I}_{\mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l) > 0}\frac{\partial}{\partial \boldsymbol{w}_l^{[m]}}\big(\mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)\big) \qquad (\text{since } (\partial \max\{0, x\}/\partial x = \mathbb{I}_{x>0}) \\
&= \frac{\max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)\}}{\gamma}\gamma\frac{\partial g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)}{\partial \boldsymbol{w}_l^{[m]}} \qquad (\text{absorbing } \mathbb{I}_{\mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l) > 0} \text{ into the max}) \\
&= \max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)\}\frac{\partial g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)}{\partial \hat{\boldsymbol{s}}_l^{[m]}}\frac{\partial \hat{\boldsymbol{s}}_l^{[m]}}{\partial \boldsymbol{w}_l^{[m]}} \\
&= \max\{0, \mu_l^{[m]} + \gamma g(\hat{\boldsymbol{s}}_l^{[m]}, \boldsymbol{r}_l)\}g'(\boldsymbol{w}_l^{[m]}, \boldsymbol{r}_l)\boldsymbol{r}_l. \qquad (7)
\end{aligned}$$

Note the significant difference between (6) and (7).

References

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