

Proof of Theorem

I. THE SOLUTION FOR (7)

The constrained optimization problem of the MGB is:

$$\begin{aligned} \mathbf{g}^*(t) &= \min_{\mathbf{g}(t)} \frac{1}{2} \|\hat{\mathbf{g}}_k(t) - \mathbf{g}(t)\|_2^2 \\ \text{s.t. } &\tilde{\mathbf{G}}(t)^T \mathbf{g}(t) \geq 0, \end{aligned} \quad (1)$$

where k is the ordinal number of the complex language, $\hat{\mathbf{g}}_k(t) = w_k(t+1)\mathbf{g}_k(t)$ is the normalized gradient, $\mathbf{g}^*(t)$ is the balancing gradient, and $\tilde{\mathbf{G}}(t)$ is the set of the rest normalized gradients. We can obtain the lagrangian function of (1), which is shown as follows:

$$L(\mathbf{g}(t), \gamma) = \frac{1}{2} \mathbf{g}(t)^T \mathbf{g}(t) - \hat{\mathbf{g}}_k(t)^T \mathbf{g}(t) - \gamma \tilde{\mathbf{G}}(t)^T \mathbf{g}(t), \quad (2)$$

where $L(\mathbf{g}(t), \gamma)$ is the lagrangian function and γ is the lagrange multiplier.

The lagrangian dual form of (1) is shown as follows:

$$\begin{aligned} f(\gamma) &= \min_{\mathbf{g}(t)} L(\mathbf{g}(t), \gamma) \\ \text{s.t. } &\gamma \geq 0, \end{aligned} \quad (3)$$

where $f(\gamma)$ is the lagrangian dual function.

We next find $\mathbf{g}^*(t)$ that minimizes (2) by setting the derivative of (2) to zero:

$$\begin{aligned} \frac{L(\partial \mathbf{g}(t), \gamma)}{\partial \mathbf{g}(t)} &= \mathbf{g}(t)^T - \hat{\mathbf{g}}_k(t)^T - \gamma \tilde{\mathbf{G}}(t)^T = 0, \\ \implies \mathbf{g}^*(t) &= \hat{\mathbf{g}}_k(t) + \tilde{\mathbf{G}}(t) \gamma^T. \end{aligned} \quad (4)$$

We further put $\mathbf{g}^*(t)$ in (3), and it can be written as:

$$\begin{aligned} f(\gamma) &= \frac{1}{2} (\hat{\mathbf{g}}_k(t)^T \hat{\mathbf{g}}_k(t) + 2\gamma \tilde{\mathbf{G}}(t)^T \hat{\mathbf{g}}_k(t) + \gamma \tilde{\mathbf{G}}(t)^T \tilde{\mathbf{G}}(t) \gamma^T) \\ &\quad - \hat{\mathbf{g}}_k(t)^T \hat{\mathbf{g}}_k(t) - 2\gamma \tilde{\mathbf{G}}(t)^T \hat{\mathbf{g}}_k(t) - \gamma \tilde{\mathbf{G}}(t)^T \tilde{\mathbf{G}}(t) \gamma^T \\ &= -\frac{1}{2} \hat{\mathbf{g}}_k(t)^T \hat{\mathbf{g}}_k(t) - \gamma \tilde{\mathbf{G}}(t)^T \hat{\mathbf{g}}_k(t) - \frac{1}{2} \gamma \tilde{\mathbf{G}}(t)^T \tilde{\mathbf{G}}(t) \gamma^T \end{aligned} \quad (5)$$

The solution $\gamma^* = \max_{\gamma; \gamma > 0} f(\gamma)$ to the dual form is obtained by:

$$\begin{aligned} \frac{\partial f(\gamma)}{\partial \gamma} &= -\hat{\mathbf{g}}_k(t)^T \tilde{\mathbf{G}}(t) - \gamma \tilde{\mathbf{G}}(t)^T \tilde{\mathbf{G}}(t) = 0, \\ \implies \gamma^* &= -\frac{\hat{\mathbf{g}}_k(t)^T \tilde{\mathbf{G}}(t)}{\tilde{\mathbf{G}}(t)^T \tilde{\mathbf{G}}(t)}. \end{aligned} \quad (6)$$

Finally, because γ^* must be grater than zeros, we obtain the balancing gradient by putting γ^* in (4):

$$\mathbf{g}^*(t) = \hat{\mathbf{g}}_k(t) - \tilde{\mathbf{G}}(t) \left[\frac{\hat{\mathbf{g}}_k(t)^T \tilde{\mathbf{G}}(t)}{\tilde{\mathbf{G}}(t)^T \tilde{\mathbf{G}}(t)} \right]^T. \quad (7)$$