Last Iterate Convergence in No-regret Learning: Constrained Min-max Optimization for Convex-concave Landscapes

1 Equations of the Jacobian of OMWU

$$
\begin{aligned}
\frac{\partial g_{1,i}}{\partial x_i} &= \frac{e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}}{S_x} + x_i\frac{1}{S_x^2}\left(e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}(-2\eta\frac{\partial^2 f}{\partial x_i^2})S_x - e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}\frac{\partial S_x}{\partial x_i}\right)\\
&\text{where } \frac{\partial S_x}{\partial x_i} = e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}} - 2\eta\sum_k x_k e^{-2\eta\frac{\partial f}{\partial x_k}+\eta\frac{\partial f}{\partial z_k}}\frac{\partial^2 f}{\partial x_i^2}\\
\frac{\partial g_{1,i}}{\partial x_j} &= x_i\frac{1}{S_x^2}\left(e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}(-2\eta\frac{\partial^2 f}{\partial x_i\partial x_j})S_x - e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}\frac{\partial S_x}{\partial x_j}\right)\\
&\text{where } \frac{\partial S_x}{\partial x_j} = e^{-2\eta\frac{\partial f}{\partial x_j}+\eta\frac{\partial f}{\partial z_j}} - 2\eta\sum_k x_k e^{-2\eta\frac{\partial f}{\partial x_k}+\eta\frac{\partial f}{\partial z_k}}\frac{\partial^2 f}{\partial x_j\partial x_k}\\
\frac{\partial g_{1,i}}{\partial y_j} &= x_i\frac{1}{S_x^2}\left(e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}(-2\eta\frac{\partial^2 f}{\partial x_i\partial y_j})S_x - e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}\frac{\partial S_x}{\partial y_j}\right)\\
&\text{where } \frac{\partial S_x}{\partial y_j} = \sum_k x_k e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}(-2\eta\frac{\partial^2 f}{\partial x_k\partial y_j})\\
\frac{\partial g_{1,i}}{\partial z_j} &= x_i\frac{1}{S_x^2}\left(e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}(\eta\frac{\partial^2 f}{\partial z_j\partial x_i})S_x - e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}\frac{\partial S_x}{\partial z_j}\right)\\
&\text{where } \frac{\partial S_x}{\partial z_j} = \eta\sum_k x_k e^{-2\eta\frac{\partial f}{\partial x_k}+\eta\frac{\partial f}{\partial z_k}}\frac{\partial^2 f}{\partial z_k\partial z_j}\\
\frac{\partial g_{1,i}}{\partial w_j} &= x_i\frac{1}{S_x^2}\left(e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}\eta\frac{\partial^2 f}{\partial z_i\partial w_j}S_x - e^{-2\eta\frac{\partial f}{\partial x_i}+\eta\frac{\partial f}{\partial z_i}}\frac{\partial S_x}{\partial w_j}\right)\\
&\text{where } \frac{\partial S_x}{\partial w_j} = \sum_k x_k e^{-2\eta\frac{\partial f}{\partial x_k}+\eta\frac{\partial f}{\partial z_k}}\eta\frac{\partial f}{\partial z_k\partial w_j}\\
\frac{\partial g_{2,i}}{\partial x_j} &= y_i\frac{1}{S_y^2}\left(e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}(2\eta\frac{\partial^2 f}{\partial x_j\partial y_i})S_y - e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}\frac{\partial S_y}{\partial x_j}\right)\\
&\text{where } \frac{\partial S_y}{\partial x_j} = \sum_k y_k e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}2\eta\frac{\partial^2 f}{\partial x_j\partial y_k}\\
\frac{\partial g_{2,i}}{\partial y_i} &= \frac{e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}}{S_y} + y_i\frac{1}{S_y^2}\left(e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}2\eta\frac{\partial^2 f}{\partial y_i^2}S_y - e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}\frac{\partial S_y}{\partial y_i}\right)\\
&\text{where } \frac{\partial S_y}{\partial y_i} = e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}} + 2\eta\sum_k y_k e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}\frac{\partial^2 f}{\partial y_i\partial y_k}\\
\frac{\partial g_{2,i}}{\partial z_j} &= y_i\frac{1}{S_y^2}\left(e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}(-\eta\frac{\partial^2 f}{\partial w_i\partial z_j})S_y - e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}\frac{\partial S_y}{\partial z_j}\right)\\
&\text{where } \frac{\partial S_y}{\partial z_j} = \sum_k y_k e^{2\eta\frac{\partial f}{\partial y_i}-\eta\frac{\partial f}{\partial w_i}}(-\eta\frac{\partial^2 f}{\partial w_k\partial z_j})
\end{aligned}
$$

$$\frac{\partial g_{2,i}}{\partial w_j} = y_i \frac{1}{S_y^2}\left(e^{2\eta\frac{\partial f}{\partial y_i} - \eta\frac{\partial f}{\partial w_i}}(-\eta\frac{\partial^2 f}{\partial w_i \partial w_j}) - e^{2\eta\frac{\partial f}{\partial y_i} - \eta\frac{\partial f}{\partial w_i}}\frac{\partial S_y}{\partial w_j}\right)$$
$$\text{where } \frac{\partial S_y}{\partial w_j} = \sum_k y_k e^{2\eta\frac{\partial f}{\partial y_i} - \eta\frac{\partial f}{\partial w_i}}(-\eta\frac{\partial^2 f}{\partial w_k \partial w_j})$$

1.1 Jacobian matrix at $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*, \mathbf{w}^*)$

This section serves for the "Spectral Analysis" of Section 3. The Jacobian matrix of g at the fixed point is obtained based on the calculations above. We refer the main article for the subscript indicating the size of each block matrix.

$$J = \begin{bmatrix} \mathbf{I} - D_{\mathbf{x}^*}\mathbf{1}\mathbf{1}^\top - 2\eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xx}} f & -2\eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xy}} f & \eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xx}} f & \eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xy}} f \\ 2\eta D_{\mathbf{y}^{*\top}}(\mathbf{I} - \mathbf{1}\mathbf{y}^*)\nabla^2_{\mathbf{yx}} f & \mathbf{I} - D_{\mathbf{y}^*}\mathbf{1}\mathbf{1}^\top + 2\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yy}} f & -\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yx}} f & -\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yy}} f \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

By acting on the tangent space of each simplex, we observe that $D_{\mathbf{x}^*}\mathbf{1}\mathbf{1}^\top \mathbf{v} = 0$ for $\sum_k v_k = 0$, so each eigenvalue of matrix J is an eigenvalue of the following matrix

$$J_{\text{new}} = \begin{bmatrix} \mathbf{I} - 2\eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xx}} f & -2\eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xy}} f & \eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xx}} f & \eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xy}} f \\ 2\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yx}} f & \mathbf{I} + 2\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yy}} f & -\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yx}} f & -\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yy}} f \\ \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

The characteristic polynomial of J_{new} is $\det(J_{new} - \lambda I)$ that can be computed as the determinant of the following matrix:

$$\begin{bmatrix} (1-\lambda)\mathbf{I} + (\frac{1}{\lambda} - 2)\eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xx}} f & (\frac{1}{\lambda} - 2)\eta D_{\mathbf{x}^*}(\mathbf{I} - \mathbf{1}\mathbf{x}^{*\top})\nabla^2_{\mathbf{xy}} f \\ (2 - \frac{1}{\lambda})\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yx}} f & (1-\lambda)\mathbf{I} + (2 - \frac{1}{\lambda})\eta D_{\mathbf{y}^*}(\mathbf{I} - \mathbf{1}\mathbf{y}^{*\top})\nabla^2_{\mathbf{yy}} f \end{bmatrix} \tag{1}$$