Appendix: On a Strictly Convex IBM Model 1

Appendix A: Convex and Concave functions.

For what follows, we denote **dom** h by the domain of h.

Definition 1. *A set S is convex if and only if all $x, y \in S$ and all $\theta \in [0,1]$, we have $\theta x + (1-\theta)y \in S$.*

Definition 2. *A function $h : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if **dom** h is convex and for all $x, y \in$ **dom** h and all $\theta \in [0,1]$, Jensen's inequality holds:*

$$h(\theta x + (1-\theta)y) \leq \theta h(x) + (1-\theta)h(y) .$$

Definition 3. *A function $h : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if and only if **dom** h is convex and for all $x \neq y \in$ **dom** h and all $\theta \in (0,1)$, Jensen's inequality holds:*

$$h(\theta x + (1-\theta)y) < \theta h(x) + (1-\theta)h(y) .$$

Definition 4. *A function $h : \mathbb{R}^n \to \mathbb{R}$ is concave if and only if **dom** h is convex and for all $x, y \in$ **dom** h and all $\theta \in [0,1]$, Jensen's inequality holds:*

$$h(\theta x + (1-\theta)y) \geq \theta h(x) + (1-\theta)h(y) .$$

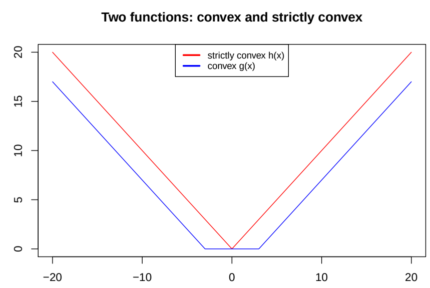
Definition 5. *A function $h : \mathbb{R}^n \to \mathbb{R}$ is strictly concave if and only if **dom** h is convex and for all $x \neq y \in$ **dom** h and all $\theta \in (0,1)$, Jensen's inequality holds:*

$$h(\theta x + (1-\theta)y) > \theta h(x) + (1-\theta)h(y) .$$

For visual reference, the graph below shows a convex function

$$g(x) = \begin{cases} -x-3, & : x \leq -3 \\ 0, & : -3 < x < 3 \\ x-3, & : 3 \leq x \end{cases}$$

and strictly convex function $h(x) = |x|$. Notice that g has multiple minimization points while h has a unique minimizer. For IBM Model 1, the log-likelihood's ”flat parts” are much more complicated than that of the simple graph above. However, this graph captures the main idea of why a convex function has multiple optima points.



Appendix B: Convex Optimization.

Definition 6. *A minimization optimization problem*

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & h_0(x) \\ \text{subject to} \quad & h_i(x) \leq 0, \; i = 1, \ldots, m. \\ & a_j^T x = b_j, \; j = 1, \ldots, l. \end{aligned}$$

is said to be convex if h_i are convex $\forall i$.

Definition 7. *A maximization optimization problem*

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & h_0(x) \\ \text{subject to} \quad & h_i(x) \geq 0, \; i = 1, \ldots, m. \\ & a_j^T x = b_j, \; j = 1, \ldots, l. \end{aligned}$$

is said to be convex if h_i are concave $\forall i$.

We note that the main issue with the above is that the equality constraints have to be linear.

Under the above setup, it can be shown that the feasible set (the set of points that satisfy the constraints) is convex and that any local optimum for the problem is a global optimum. If h_0 is strict then any local optimum is actually then the unique global optimum.