1 bLAT Layer

Definition 1. *The bi-Lipschitz affine transformation (bLAT) layer is given by the transformation, $f : \mathbb{R}^N \to \mathbb{R}^M$, defined by*

$$f(x) = U\Sigma V^{\mathrm{T}} x + b, \tag{1}$$

where $U \in \mathbb{R}^{M\times r}$, $V \in \mathbb{R}^{r\times N}$, $\Sigma \in \mathbb{R}^{r\times r}$, $b \in \mathbb{R}^M$, $r = \min(M, N)$, U and V are orthogonal matrices satisfying $U^{\mathrm{T}}U = V^{\mathrm{T}}V = I$, and

$$\Sigma = \mathrm{diag}\left[\begin{array}{ccc} \sigma_1 & \dots & \sigma_r \end{array}\right], \tag{2}$$

where $\frac{1}{L} \le \sigma_i \le L$ for $i = 1, \dots, r$ and some $L \ge 1$. When $M = N$, the mapping is invertible.

$$f^{-1}(x) = V\Sigma^{-1}U^{\mathrm{T}}(x - b). \tag{3}$$

Our new contribution is to include the matrix term $U\Sigma V^T$, which is written as a singular value decomposition (SVD), which exists for every matrix in $\mathbb{R}^{M\times N}$. One advantage of the bLAT layer is the following theorem.

Theorem 1. *The forward and inverse bLAT layers are both L-bi-Lipschitz.*

This follows from the fact that $\frac{1}{L} \le \|\nabla f(x)\|_2 \le L$ and $\frac{1}{L} \le \|\nabla f^{-1}(x)\|_2 \le L$. Another advantage of using the SVD is that, by construction, the evaluation of $f(x)$ and $f^{-1}(x)$ have an equivalent computational cost. In our implementation of the $M = N$ case, the orthogonal matrices of the bLAT layers are parameterized using the matrix exponential of a skew-symmetric matrix input, i.e. $U = \mathrm{expm}(S)$, where $S = -S^T$. When $M \ne N$, the orthogonal matrices are parameterized using the Householder factorization.

2 Orthogonal/Unitary Parameterizations

2.1 Matrix Exponential

Let $S \in \mathbb{C}^{M\times M}$ be a skew symmetric matrix ($S = -S^T$), then

$$U = \mathrm{expm}(S) \tag{4}$$

is unitary. (Note: when $S \in \mathbb{R}^{M\times M}$, then U is orthogonal.)

2.2 Householder

The notes in this section were originally written to help with implementation of code. Consider an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$. The Householder decomposition is defined by

$$Q = Q_n \dots Q_1, \tag{5}$$
$$Q^T = Q_1^T \dots Q_n^T = Q_1 \dots Q_n, \tag{6}$$

Where $Q_1, \ldots, Q_n$ are symmetric elementary orthogonal matrices defined by

$$Q_k = \begin{bmatrix} I_{N-k \times N-k} & 0 \\ 0 & F_k \end{bmatrix}, \qquad F_k = I_{k \times k} - 2\frac{v_k v_k^T}{v_k^T v_k}, \qquad v_k \in \mathbb{R}^k. \tag{7}$$

We can also define Q_k using

$$Q_k = \tilde{F}_k = I_{N \times N} - 2\frac{\tilde{v}_k \tilde{v}_k^T}{\tilde{v}_k^T \tilde{v}_k}, \qquad \tilde{v}_k = \begin{bmatrix} 0 \\ v_k \end{bmatrix} \in \mathbb{R}^N. \tag{8}$$

Note that Q_k is symmetric.

$$\tilde{F}_k^T = I_{N \times N} - 2\frac{\tilde{v}_k \tilde{v}_k^T}{\tilde{v}_k^T \tilde{v}_k} \tag{9}$$

Matrix action of Q_k:

$$Q_k y = y - 2\frac{\tilde{v}_k \tilde{v}_k^T}{\tilde{v}_k^T \tilde{v}_k} y, \tag{10}$$

Consider $y \in \mathbb{R}^M$ $Q \in \mathbb{R}^{M \times N}$, $\tilde{v}_k \in \mathbb{R}^N$, then

$$y^T Q_k = y^T - 2\frac{\tilde{v}_k^T}{\tilde{v}_k^T \tilde{v}_k}(y^T \tilde{v}_k). \tag{11}$$

2.3 Polcari

Orthogonal matrices can be parameterized using the Polcari decomposition, which was originally introduced in [**?**]. Given an orthogonal matrix, $Q \in \mathbb{R}^{N \times N}$, then Q can be decomposed as

$$Q = \Psi(w_{N-1}) \ldots \Psi(w_1) S(\varphi_1, \ldots, \varphi_N), \tag{12}$$

where $w_j \in \{x : x \in \mathbb{R}^j,\ \|x\|_2 \leq 1\}$, for $j = 1, \ldots, N-1$, and $\varphi_1, \ldots, \varphi_N \in \{1, -1\}$, and

$$\Psi(w_j) = \begin{bmatrix} \mathrm{I}_j - \frac{w_j w_j^{\mathrm{T}}}{1+\sqrt{1-w_j^{\mathrm{T}} w_j}} & w_j & \\ -w_j^{\mathrm{T}} & \sqrt{1 - w_j^{\mathrm{T}} w_j} & \\ & & \mathrm{I}_{N-j-1} \end{bmatrix}, \qquad i = 1, \ldots, N-1,$$

$$S(\varphi_1, \ldots, \varphi_N) = \begin{bmatrix} \varphi_1 & & \\ & \ddots & \\ & & \varphi_N \end{bmatrix}.$$

Here, I_d represents a $d \times d$ identity matrix. Note that a vector, $v_j \in \mathbb{R}^i$, can be projected inside the unit ball, $w_j \in \{x : x \in \mathbb{R}^j,\ \|x\|_2 \leq 1\}$, using a transformation such as

$$w_j = \rho_j \frac{v_j}{\|v_j\|}, \qquad \rho_j = \tanh(\|v_j\|) \in [0, 1].$$

The decomposition in (12) can be used as a parameterization to express any orthogonal matrix. Given an orthogonal matrix, $Q \in \mathbb{R}^{N\times N}$, the parameters of the Polcari decomposition can be uniquely determined. Consider the following properties.

1. $\Psi(w_j)$ for $j = 1, \dots, N-1$ and S are orthogonal matrices and therefore, by construction, the right-hand-side of (12) is orthogonal.
2. $\Psi(w_j)\mathrm{e}_k = \mathrm{e}_k$ for $k > j+1$, where e_k is a vector with a 1 in the k^{th} component and 0's elsewhere.
3. $\mathrm{P}_{(k-1)N}\Psi(w_{k-1})\mathrm{e}_k = w_{k-1}$, for $k = 2, \dots N$, where $\mathrm{P}_{(k-1)N}$ is an identity matrix of size $(k-1)\times N$.
4. $\mathrm{e}_k^T\Psi(w_{k-1})\mathrm{e}_k = \sqrt{1 - w_{k-1}^{\mathrm{T}} w_{k-1}}$, for $k = 2, \dots N$.
5. $\mathrm{P}_{(k-1)N}\Psi(w_{k-1})\dots\Psi(w_1)S\mathrm{e}_k = \varphi_k w_{k-1}$, for $k = 2, \dots N$. This follows from (2) and (3).
6. $\mathrm{e}_k^T\Psi(w_{k-1})\dots\Psi(w_1)S\mathrm{e}_k = \varphi_k\sqrt{1 - w_{k-1}^{\mathrm{T}} w_{k-1}}$ for $k = 2, \dots N$. This follows from (2) and (4).

Using properties (5) and (6) for $k = N$, we have

$$P_{N-1,N}Q\mathrm{e}_N = \varphi_N w_{N-1}, \qquad \mathrm{e}_N^{\mathrm{T}} Q\mathrm{e}_N = \varphi_N\sqrt{1 - w_{N-1}^{\mathrm{T}} w_{N-1}},$$

and therefore

$$w_{N-1} = \varphi_N P_{N-1,N}Q\mathrm{e}_N, \qquad \varphi_N = \operatorname{sign}\left(\mathrm{e}_N^{\mathrm{T}} Q\mathrm{e}_N\right).$$

Suppose that $w_k, \dots w_{N-1}$ is known. We can obtain relations for w_{k-1} and φ_k using properties (5) and (6).

$$\mathrm{P}_{(k-1)N}\Psi(w_k)^{\mathrm{T}}\dots\Psi(w_{N-1})^{\mathrm{T}} Q\mathrm{e}_k = \varphi_k w_{k-1},$$
$$\mathrm{e}_k^{\mathrm{T}}\Psi(w_k)^{\mathrm{T}}\dots\Psi(w_{N-1})^{\mathrm{T}} Q\mathrm{e}_k = \varphi_k\sqrt{1 - w_{k-1}^{\mathrm{T}} w_{k-1}}.$$

Using this, general relations for w_k and φ_k for all k can be obtained.

$$w_k = \varphi_{k+1}\mathrm{P}_{k,N}\Psi(w_{k+1})^{\mathrm{T}}\dots\Psi(w_{N-1})^{\mathrm{T}} Q\mathrm{e}_{k+1}, \qquad k = 1, \dots, N-1, \tag{13a}$$
$$\varphi_k = \operatorname{sign}\left(\mathrm{e}_k^{\mathrm{T}}\Psi(w_k)^{\mathrm{T}}\dots\Psi(w_{N-1})^{\mathrm{T}} Q\mathrm{e}_k\right), \qquad k = 1, \dots, N. \tag{13b}$$

An algorithm for determining the Polcari parameters would use (13) recursively, decrementing k from N to 1. Since the Polcari parameters can be uniquely determined, the following theorem can be stated.

Theorem 2. *The decomposition given by* (12) *is a parameterization for any orthogonal matrix in* $\mathbb{R}^{N\times N}$.

Note that since $S(\varphi_1, \dots, \varphi_N)US(\varphi_1, \dots, \varphi_N)^T = U$, $A(\theta_A)$ is invariant to the choice of $\varphi_1, \dots, \varphi_N$. For simplicity, we choose $\varphi_1 = \dots = \varphi_N = 1$ in our parameterization, since $S(1, \dots, 1) = I$.

2.4 Normalization

Note that Lipschitz control can be obtained through the use of layers of the form

$$\sigma(Ax + b), \tag{14}$$

where $\|A\|_2 \leq L$. However, we have the bound $\|A\|_2 \leq \sqrt{\|A\|_\infty \|A\|_1}$ and therefore we can use the parameterization

$$A = L \frac{B}{\max(1, \sqrt{\|B\|_\infty \|B\|_1})} \tag{15}$$