1 Image Embedding Regularizer

We will regularize our image embeddings by penalizing the mismatch between the empirical distribution of embedding increments across an image and a target distribution. Intuitively, we want our target distribution to model locally smooth changes punctuated by occasional discontinuities, while remaining analytically and computationally tractable. Therefore we choose a Lévy sheet with Gaussian and jump components.

1.1 Lévy sheet definition

This is a process $Z : \mathbb{R}_+^2 \to \mathbb{R}^d$ which can be decomposed as

$$Z(x, y) = \sigma W(x, y) + J(x, y),$$

where $W(x, y)$ is a standard Brownian sheet, $\sigma \geq 0$, and $J(x, y)$ is a pure-jump finite-activity Lévy sheet with independent increments over rectangles. The jump component is specified by a spherically symmetric Lévy measure μ on jump sizes with density

$$\frac{d\mu}{dz}(z) = \lambda \int_0^\infty (2\pi\tau)^{-d/2} \exp\left(-\frac{\|z\|^2}{2\tau}\right) \rho(d\tau),$$

with finite intensity $0 < \lambda < \infty$ and probability measure ρ on the random variance τ, corresponding to generating a (zero-mean) jump size $\gamma = \sqrt{\tau} v$ by sampling $\tau \sim \rho$ and $v \sim N(0, I_d)$. These choices ensure $Z(x, y)$ is isotropic, i.e., $\{QZ(x, y) : (x, y) \in \mathbb{R}_+^2\} \stackrel{d}{=} \{Z(x, y) : (x, y) \in \mathbb{R}_+^2\}$ for any orthogonal matrix $Q \in O(d)$.

1.2 Characteristic Function

If Z is observed on a regular square integral grid of side length s, a discrete second difference is

$$\Delta Z(x, y) = Z(x, y) - Z(x, y - 1) - Z(x - 1, y) + Z(x - 1, y - 1),$$

with $Z(x, y) \doteq 0$ whenever $x \leq 0$ or $y \leq 0$. Using convolution notation,

$$\Delta Z = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \star Z. \tag{1}$$

These ΔZ are independent and each has the same infinitely-divisible law with characteristic function

$$\begin{aligned} \phi(u) &= \mathbb{E}\left[e^{iu^\top \Delta Z}\right], \\ &= \exp\left[\left(-\frac{\sigma^2}{2}\|u\|^2 + \lambda \int_0^\infty \left(e^{-\frac{\tau}{2}\|u\|^2} - 1\right) \rho(d\tau)\right)\right]. \end{aligned}$$

The empirical characteristic function of observed increments is

$$\hat{\phi}(u) = \frac{1}{s^2} \sum_{x=1}^{s} \sum_{y=1}^{s} \exp\left(iu^\top \Delta Z(x,y)\right).$$

More generally, if the increments are computed over disjoint cells of area ϵ^2, the log characteristic function is scaled by ϵ^2.

1.3 Epps-Pulley based regularizer

Our regularizer will be based upon the Epps-Pulley statistic for embedding differences across a single image, based upon the independent increments and characteristic function from above. The Epps-Pulley statistic can be defined and computed for an $\mathbb{R}^d$ valued Lévy sheet, but it is unwieldy computationally, so we will reduce to $\mathbb{R}$. Let $v \in S^{d-1}$ be any given unit-norm vector, then

$$\begin{aligned}\phi_v(u) &= \mathbb{E}\left[e^{iuv^\top \Delta Z}\right] \\ &= \exp\left(-\frac{\sigma^2}{2}u^2 + \lambda\left(\mathbb{E}_\rho\left[e^{-\frac{\tau}{2}u^2}\right] - 1\right)\right),\end{aligned}$$

where table 1 contains $\mathbb{E}_\rho\left[e^{-\frac{\tau}{2}u^2}\right]$ in closed-form for some choices of ρ. Assuming Z is observed on regular square integral grid of side length s, the Epps-Pulley statistic on the projected increments is

$$\begin{aligned}\mathrm{EP} &= s^2 \int_{-\infty}^{\infty} \left|\hat{\phi}_v(u) - \phi_v(u)\right|^2 w(u) du, \\ \hat{\phi}_v(u) &= \frac{1}{s^2} \sum_{x=1}^{s} \sum_{y=1}^{s} e^{iuv^\top \Delta Z(x,y)},\end{aligned}$$

where $w(u)$ is a symmetric non-negative function with $\int_{-\infty}^{\infty} w(u)du < \infty$.

1.4 Incorporating other orientations

The discrete difference operator from section 1.2 has an orientation, fundamentally due to the Markov property of the Lévy sheet over a rectangular filtration: the value $Z(x,y)$ is a sufficient statistic for the entire "southwest" rectangle $\{Z(R) : R \subseteq [0,x] \times [0,y]\}$. This preferred direction is a mathematical consequence of inducing a partial ordering on the plane and is necessary to define the Lévy sheet, but the difference operator might be a poor fit for certain images. We will mitigate this by using multiple Lévy sheets with different orientations.

For axis-oriented filtrations, suppose for $(x,y) \in [0,s]^2$, we observe $A(x,y) = Z(s-x,y)$. Up to a sign, the difference operator acting on A values is the same as the difference operator acting on Z values. Repeating the argument for $A(x,y) = Z(x,s-y)$ and $A(x,y) = Z(s-x,s-y)$ reveals

$$\Delta A = \pm \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \star A.$$

ρ	$\mathbb{E}_\rho\left[e^{-\frac{\tau}{2}u^2}\right]$	**Notes**
$\mathrm{Exp}\left(\frac{\alpha^2}{2}\right)$	$\frac{\alpha^2}{\alpha^2+u^2}$	Laplace
$\Gamma(k,\beta)$	$\left(\frac{\beta}{\beta+u^2/2}\right)^k$	Variance-Gamma
$\mathrm{IG}\left(\frac{\nu}{2},\frac{\nu}{2}\right)$	$\frac{2K_{\frac{\nu}{2}}\left(\sqrt{\nu}\,\lvert u\rvert\right)}{\Gamma(\frac{\nu}{2})}\left(\frac{\sqrt{\nu}\,\lvert u\rvert}{2}\right)^{\nu/2}$	t with ν d.o.f.

Table 1: For certain jump size distributions the contribution to the characteristic function has closed-form. Both $\mathrm{Exp}(\alpha^2/2)$ and $\Gamma(k,\beta)$ use their respective rate parameterizations.

Since the distribution of Z is sign invariant, this means increments using this one difference operator covers all 4 cases.

For a 45 degree rotated process, suppose for $0 \leq x \leq s$, $0 \leq y \leq s$, we observe $A(x,y) = Z\left(t_0(x,y), t_1(x,y)\right)$ where

$$\begin{pmatrix} t_0(x,y) \\ t_1(x,y) \end{pmatrix} = \begin{pmatrix} s \\ 0 \end{pmatrix} + \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \tag{2}$$

Then A will have a “southern” cone filtration. The resulting discrete difference operator for computing independent increments of A is

$$\Delta A(x,y) = A(x-1,y) - A(x,y-1) - A(x,y+1) + A(x+1,y), \tag{3}$$

with $A(x,y) \doteq 0$ for $x \leq 0$, $x > s$, $y \leq 0$, or $y > s$. Using convolution notation,

$$\Delta A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \star A, \tag{4}$$

whose preimage under the inverse 45 degree rotation is an axis-aligned square of area 2. The other diagonal orientations utilize the same difference operator to compute increments up to a sign, therefore, this one difference operator covers all 4 diagonal cases.

We will model our observations $A : [0,s]^2 \to \mathbb{R}^{2d}$, where the first d observation coordinates are produced by an untransformed Lévy sheet and hence use eq. (1) to compute increments, and the last d observation components are produced by a rotated Lévy sheet as in eq. (2) and hence use eq. (4) to compute increments.