Variational Methods for Computer Vision: Solution Sheet 1

Exercise: 23 October 2014

Part I: Theory

1. A metric expresses an intuitive notion of distance on an abstract set X.

   (a) We verify all four conditions:

   - Summands are all positive due to the absolute value $\Rightarrow d(x,y) \geq 0$.
   - $x = y \Rightarrow d(x,y) = 0$ follows directly by substitution. For the other direction we assume $d(x,y) = 0$ which implies $|x_i - y_i| = 0$ for all $1 \leq i \leq n$, which in turn implies $x_i = y_i$.
   - Symmetry: Follows directly from symmetry of absolute value function.
   - Subadditivity: Follows directly from the basic triangle inequality ($|x+y| \leq |x|+|y|$):

   $$|x_i - z_i| = |x_i - y_i + y_i - z_i| \leq |x_i - y_i| + |y_i - z_i|.$$

   $\Rightarrow$ We have verified that the Manhattan (or Taxicab) distance is a metric. Note that the Manhattan distance is induced by the standard ℓ^1 norm on $\mathbb{R}^n$: $d(x,y) = \|x-y\|_1$.

   (b) Again we check the individual conditions:

   - Follows from Q being positive definite. $\langle x, Qx\rangle \geq 0, \forall x \in \mathbb{R}^n$.
   - $x = y \Rightarrow d(x,y) = 0$ again follows directly. Assume $\langle x-y, Q(x-y)\rangle = 0$ and $x \neq y$. Then we have $\langle z, Qz\rangle = 0$ for some $z = x - y \neq 0$. This violates positive definiteness of Q.
   - Symmetry: $d(x,y) = \langle x-y, Q(x-y)\rangle = \langle y-x, Q(y-x)\rangle = d(y,x)$
   - For subadditivity, let us start with the following:

   $$\begin{aligned} \|x+y\|_Q^2 &:= \langle x+y, Q(x+y)\rangle = \left\langle Q^{\frac{1}{2}}(x+y), Q^{\frac{1}{2}}(x+y)\right\rangle \\ &= \|Q^{\frac{1}{2}}x\|^2 + 2\left\langle Q^{\frac{1}{2}}x, Q^{\frac{1}{2}}y\right\rangle + \|Q^{\frac{1}{2}}y\|^2 \\ &\leq \|Q^{\frac{1}{2}}x\|^2 + 2\|Q^{\frac{1}{2}}x\|\|Q^{\frac{1}{2}}y\| + \|Q^{\frac{1}{2}}y\|^2 \\ &= (\|Q^{\frac{1}{2}}x\| + \|Q^{\frac{1}{2}}y\|)^2 = (\|x\|_Q + \|y\|_Q)^2. \end{aligned}$$

   This implies $\|x+y\|_Q \leq \|x\|_Q + \|y\|_Q$. Now we have

   $$\begin{aligned} d(x,y) &= \|x-y\|_Q = \|x-z+z-y\|_Q \\ &\leq \|x-z\|_Q + \|z-y\|_Q = d(x,z) + d(z,y). \end{aligned}$$

   Note that it would also suffice to show that $\langle x, Qy\rangle = \langle x, y\rangle_Q$ defines a valid inner product on $\mathbb{R}^n$ which in turn induces a norm $\|x\|_Q = \langle x, x\rangle_Q^{1/2}$ which in turn induces the Mahalanobis distance $d(x,y) = \|x-y\|_Q$.

   The following relationships hold: Inner Product $\xrightarrow{\text{induces}}$ Norm $\xrightarrow{\text{induces}}$ Metric.

   (c) The Kullback-Leibler divergence can be interpreted as a measure of dissimilarity between two probability distributions. It is however not symmetric, and we show that by constructing a counterexample. Let p_1, p_2 be probability distributions with

$$p_1(x) = \begin{cases} 1 & \text{if } -0.5 \le x \le 0.5, \\ 0 & \text{else,} \end{cases} \quad \text{and} \quad p_2(x) = \begin{cases} 0.5 & \text{if } -0.5 \le x \le 0, \\ 1.5 & \text{if } 0 \le x \le 0.5, \\ 0 & \text{else.} \end{cases}$$

It can be verified that these are indeed probability distributions. Then we have

$$d(p_1, p_2) = 0.5(\ln 2 + \ln \frac{2}{3}) \approx 0.144,$$
$$d(p_2, p_1) = 0.5(0.5 \ln \frac{1}{2} + 1.5 \ln \frac{3}{2}) \approx 0.131.$$

Furthermore the Kullback-Leibler divergence does not satisfy the triangle inequality. Note that it can be shown that it still satisfies

- $d(x, y) \ge 0$,
- $d(x, y) = 0 \Leftrightarrow x = y$.

A function $d : X \times X \to \mathbb{R}$ which satisfies only these two conditions is sometimes called a *premetric*.

2. (a) Let us prove associativity of convolution first:

$$\begin{aligned}
((f * g) * h)(u) &= \int_{\mathbb{R}} (f * g)(x)\, h(u - x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(y) g(x - y)\, \mathrm{d}y \right) h(u - x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(y) g(x - y) h(u - x)\, \mathrm{d}y\, \mathrm{d}x \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} f(y) g(x - y) h(u - x)\, \mathrm{d}x\, \mathrm{d}y && \text{(Fubini's theorem)} \\
&= \int_{\mathbb{R}} f(y) \int_{\mathbb{R}} g(x - y) h(u - x)\, \mathrm{d}x\, \mathrm{d}y \\
&= \int_{\mathbb{R}} f(y) \int_{\mathbb{R}} g((x + y) - y) h(u - (x + y))\, \mathrm{d}x\, \mathrm{d}y && \text{(Translation invariance)} \\
&= \int_{\mathbb{R}} f(y) \int_{\mathbb{R}} g(x) h((u - y) - x)\, \mathrm{d}x\, \mathrm{d}y \\
&= \int_{\mathbb{R}} f(y) (g * h)(u - y)\, \mathrm{d}y \\
&= (f * (g * h))(u).
\end{aligned}$$

(b) For distributivity we have:

$$\begin{aligned}
f * (g + h)(u) &= \int_{\mathbb{R}} f(x)(g + h)(u - x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}} f(x) g(u - x) + f(x) h(u - x)\, \mathrm{d}x \\
&= \int_{\mathbb{R}} f(x) g(u - x)\, \mathrm{d}x + \int_{\mathbb{R}} f(x) h(u - x)\, \mathrm{d}x \\
&= (f * g + f * h)(u).
\end{aligned}$$

(c) We start with the definition of the Fourier transform:

$$\begin{aligned}
\mathcal{F}\{f * g\}(\nu) &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(y) g(x - y)\, \mathrm{d}y \right) e^{-2\pi i x \nu}\, \mathrm{d}x \\
&= \int_{\mathbb{R}} f(y) \left(\int_{\mathbb{R}} g(x - y) e^{-2\pi i x \nu}\, \mathrm{d}x \right) \mathrm{d}y.
\end{aligned}$$

Introducing the substitution $z = x - y$, $\mathrm{d}z = \mathrm{d}x$ we arrive at

$$\begin{aligned}
\int_{\mathbb{R}} f(y) \left(\int_{\mathbb{R}} g(x-y) e^{-2\pi i x \nu} \, \mathrm{d}x \right) \mathrm{d}y &= \int_{\mathbb{R}} f(y) \left(\int_{\mathbb{R}} g(z) e^{-2\pi i (z+y) \nu} \, \mathrm{d}z \right) \mathrm{d}y \\
&= \int_{\mathbb{R}} f(y) e^{-2\pi i y \nu} \int_{\mathbb{R}} g(z) e^{-2\pi i z \nu} \, \mathrm{d}z \, \mathrm{d}y \\
&= \underbrace{\int_{\mathbb{R}} f(y) e^{-2\pi i y \nu} \, \mathrm{d}y}_{=:\mathcal{F}\{f\}(\nu)} \underbrace{\int_{\mathbb{R}} g(z) e^{-2\pi i z \nu} \, \mathrm{d}z}_{=:\mathcal{F}\{g\}(\nu)} .
\end{aligned}$$

As the Fourier transform can be implemented to run in $\mathcal{O}(n \log n)$ time, convolutions can be computed efficiently by exploiting this property:

$$f * g = \mathcal{F}^{-1}\{\mathcal{F}\{f\} \cdot \mathcal{F}\{g\}\}.$$

(d) Let us consider the difference quotient

$$\frac{(f*g)(x+t) - (f*g)(x)}{t} = \int_{\mathbb{R}} f(y) \frac{g(x+t-y) - g(x-y)}{t} \, \mathrm{d}y.$$

For $t \to 0$ it follows that

$$\frac{g(x+t-y) - g(x-y)}{t} \longrightarrow \frac{dg}{dx},$$

which in turn yields

$$\frac{d}{dx}(f*g) = f * \frac{dg}{dx}.$$

The remaining equality can be shown analogously using commutativity of convolution.