Variational Methods for Computer Vision: Solution Sheet 2

Exercise: October 31, 2016

Part I: Theory

1. This follows from a direct calculation:

$$
\begin{aligned}
((f * k_1) * k_2)(x, y) &= \int \left(\int k_1(s) f(x - s, y - t) \, \mathrm{ds} \right) k_2(t) \, \mathrm{dt} \\
&= \int \int f(x - s, y - t) k_1(s) k_2(t) \, \mathrm{ds} \, \mathrm{dt} \\
&= \int \int f(x - s, y - t) \frac{1}{2\pi\sigma^2} \exp(-\frac{s^2 + t^2}{2\sigma^2}) \, \mathrm{ds} \, \mathrm{dt} \\
&= \int \int f(x - s, y - t) K(s, t) \, \mathrm{ds} \, \mathrm{dt} \\
&= (f * K)(x, y)
\end{aligned}
$$

2. (a) To compute the gradient, we take a look at the directional derivative of $\tilde{f}$ in an arbitrary direction $R^\top h \in \mathbb{R}^2$.

$$
\begin{aligned}
\langle \nabla (f \circ R)(x), R^\top h \rangle &= \lim_{\varepsilon \to 0} \frac{(f \circ R)(x + \varepsilon R^\top h) - (f \circ R)(x)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{f(Rx + \varepsilon R R^\top h) - f(Rx)}{\varepsilon} \\
&= \lim_{\varepsilon \to 0} \frac{f(Rx + \varepsilon h) - f(Rx)}{\varepsilon} \\
&= \langle \nabla f\big|_{Rx}, h \rangle \\
&= \langle \nabla f\big|_{Rx}, R R^\top h \rangle \\
&= \langle R^\top \nabla f\big|_{Rx}, R^\top h \rangle.
\end{aligned}
$$

Since the above holds for all $R^\top h$ we can conclude that $\nabla (f \circ R)(x) = R^\top \nabla f\big|_{Rx}$, using the following argument:

$$
\begin{aligned}
& \langle x, h \rangle = \langle y, h \rangle, \forall h \\
\Leftrightarrow \; & \langle x - y, h \rangle = 0, \forall h \\
\Rightarrow \; & \langle x - y, x - y \rangle = 0, \\
\Rightarrow \; & x = y.
\end{aligned}
$$

Thus we know that $\nabla (f \circ R) = R^\top \circ \nabla f \circ R$.
Another way of showing this is to use the multivariate chain-rule for the Jacobian

$$J_{f \circ g}(a) = J_f(g(a)) \circ J_g(a).$$

Since by definition the gradient of a function is the transpose of the Jacobian we have:

$$[\nabla (f \circ R)](x) = J_{f \circ R}(x)^\top = (J_f(Rx) \circ R)^\top = R^\top \circ J_f(Rx)^\top = R^\top (\nabla f)(Rx).$$

(b) We show the identity directly:

$$\begin{aligned}\|\nabla(f\circ R)(x)\| &\overset{(a)}{=} \|R^\top \nabla f\big|_{Rx}\| \\ &= \sqrt{\langle R^\top \nabla f\big|_{Rx}, R^\top \nabla f\big|_{Rx}\rangle} \\ &= \sqrt{\langle \nabla f\big|_{Rx}, RR^\top \nabla f\big|_{Rx}\rangle} \\ &= \sqrt{\langle \nabla f\big|_{Rx}, \nabla f\big|_{Rx}\rangle} \\ &= \|\nabla f\big|_{Rx}\|,\end{aligned}$$

and hence $\|\nabla(f\circ R)\| = \|(\nabla f)\circ R\|$.

(c) Denoting the partial derivatives as $\partial_x f = f_x$, $\partial_y f = f_y$, etc., (with the convention $f_x(R\mathbf{x}) := ((\partial_x f)\circ R)(x)$ and using short hand notation $\cos(\alpha) = c$, $\sin(\alpha) = s$, $\mathbf{x} = (x, y)$ we compute the following:

$$\begin{aligned}\left[\Delta(f\circ R)\right](x,y) &= \text{div}(R^T \begin{bmatrix} f_x(R\mathbf{x}) \\ f_y(R\mathbf{x})\end{bmatrix}) \\ &= \text{div}\begin{pmatrix} cf_x(R\mathbf{x}) + sf_y(R\mathbf{x}) \\ -sf_x(R\mathbf{x}) + cf_y(R\mathbf{x})\end{pmatrix} \\ &= \partial_x(cf_x(R\mathbf{x}) + sf_y(R\mathbf{x})) + \partial_y(-sf_x(R\mathbf{x}) + cf_y(R\mathbf{x}))\end{aligned}$$

Now we have that

$$\nabla f_x(R\mathbf{x}) = \begin{bmatrix}\partial_x f_x(R\mathbf{x}) \\ \partial_y f_x(R\mathbf{x})\end{bmatrix} = R^\top \begin{bmatrix} f_{xx}(R\mathbf{x}) \\ f_{xy}(R\mathbf{x})\end{bmatrix} = \begin{bmatrix} cf_{xx}(R\mathbf{x}) + sf_{xy}(R\mathbf{x}) \\ -sf_{xx}(R\mathbf{x}) + cf_{xy}(R\mathbf{x})\end{bmatrix},$$

$$\nabla f_y(R\mathbf{x}) = \begin{bmatrix}\partial_x f_y(R\mathbf{x}) \\ \partial_y f_y(R\mathbf{x})\end{bmatrix} = R^\top \begin{bmatrix} f_{yx}(R\mathbf{x}) \\ f_{yy}(R\mathbf{x})\end{bmatrix} = \begin{bmatrix} cf_{yx}(R\mathbf{x}) + sf_{yy}(R\mathbf{x}) \\ -sf_{yx}(R\mathbf{x}) + cf_{yy}(R\mathbf{x})\end{bmatrix},$$

and thus:

$$\begin{aligned}&\partial_x(cf_x(R\mathbf{x}) + sf_y(R\mathbf{x})) + \partial_y(-sf_x(R\mathbf{x}) + cf_y(R\mathbf{x})) = \\ &c\partial_x f_x(R\mathbf{x}) + s\partial_x f_y(R\mathbf{x}) - s\partial_y f_x(R\mathbf{x}) + c\partial_y f_y(R\mathbf{x}) = \\ &c(cf_{xx}(R\mathbf{x}) + sf_{xy}(R\mathbf{x})) + s(cf_{yx}(R\mathbf{x}) + sf_{yy}(R\mathbf{x})) - \\ &s(-sf_{xx}(R\mathbf{x}) + cf_{xy}(R\mathbf{x})) + c(-sf_{yx}(R\mathbf{x}) + cf_{yy}(R\mathbf{x})) = \\ &c^2 f_{xx}(R\mathbf{x}) + csf_{xy}(R\mathbf{x}) + scf_{yx}(R\mathbf{x}) + s^2 f_{yy}(R\mathbf{x}) + \\ &s^2 f_{xx}(R\mathbf{x}) - scf_{xy}(R\mathbf{x}) - csf_{yx}(R\mathbf{x}) + c^2 f_{yy}(R\mathbf{x}) = \\ &\underbrace{(c^2+s^2)}_{=1} f_{xx}(R\mathbf{x}) + \underbrace{(c^2+s^2)}_{=1} f_{yy}(R\mathbf{x}) = ((\Delta f)\circ R)(\mathbf{x}),\end{aligned}$$

which shows $(\Delta f)\circ R = \Delta(f\circ R)$.

Remark: Schwarz's theorem (or Clairaut's theorem) states that if $f : \mathbb{R}^n \to \mathbb{R}$ has continuous second partial derivatives at a then

$$f_{xy}(a) = \frac{\partial^2 f}{\partial x \partial y}(a) = \frac{\partial^2 f}{\partial y \partial x}(a) = f_{yx}(a).$$

This condition is fulfilled since $f \in C^2(\Omega; \mathbb{R})$ by assumption.

3. (a) This immediately follows from the linearity of the divergence operator:

$$\begin{aligned}\text{div}\left(g \cdot \nabla u\right)(x) &= g\text{div}\left(\nabla u\right)(x) \\ &= g\Delta u(x)\end{aligned}$$

(b) We compute:

$$\begin{aligned}\text{div}\left(g\nabla u\right)(x) &= \frac{\partial}{\partial x_1}\left(g\frac{\partial}{\partial x_1}u\right)(x) + \frac{\partial}{\partial x_2}\left(g\frac{\partial}{\partial x_2}u\right)(x) \\ &= \frac{\partial^2 u}{\partial x_1^2}(x)g(x) + \frac{\partial^2 u}{\partial x_2^2}(x)g(x) + \frac{\partial g}{\partial x_1}(x)\frac{\partial u}{\partial x_1}(x) + \frac{\partial g}{\partial x_2}(x)\frac{\partial u}{\partial x_2}(x) \\ &= g(x)\Delta u(x) + \langle \nabla g(x), \nabla u(x)\rangle\end{aligned}$$