Integration by Parts on $\mathbb{R}^d$ and Stein's Identity

Qiang Liu

Stokes' theorem, $\int_{\mathcal{D}} \mathrm{d}\omega = \int_{\partial\mathcal{D}} \omega$, is the generalized fundamental theorem of calculus. It allows us to convert between functions and their derivatives, and underlies many key tools, including integration by parts in $\mathbb{R}^d$ and Stein's identity.

In textbooks, this is typically presented for a bounded domain $\mathcal{D} \subset \mathbb{R}^d$ with a smooth boundary $\partial\mathcal{D}$:

$$\int_{\mathcal{D}} \nabla f(x)\,\mathrm{d}x = \int_{\partial\mathcal{D}} f(x)\,\vec{n}(x)\,\mathrm{d}S(x), \tag{1}$$

where $\vec{n}(x)$ is the outward unit normal to $\partial\mathcal{D}$ and $\mathrm{d}S(x)$ is the surface area element.

However, we often work on the whole space $\mathbb{R}^d$ with a vanishing boundary condition:

$$\int_{\mathbb{R}^d} \nabla f(x)\,\mathrm{d}x = 0. \tag{2}$$

This result can be viewed as the limit of the bounded form (1) as the domain expands. A simple sufficient condition is $\int_{\mathbb{R}^d} |\nabla f(x)|\,\mathrm{d}x < \infty$, provided $f \in L^1(\mathbb{R}^d)$ and is weakly differentiable.

Here, we give a quick derivation of (2). The idea is to simply leverage the *translation invariance* of the Lebesgue integral: if $f \in L^1(\mathbb{R}^d)$, then

$$\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} f(x+y)\,\mathrm{d}x, \qquad \forall\, y \in \mathbb{R}^d.$$

Fix a unit vector $v \in \mathbb{S}^{d-1}$ and a scalar $\epsilon \neq 0$. Setting the displacement $y = \epsilon v$ gives

$$\int_{\mathbb{R}^d} \frac{f(x+\epsilon v) - f(x)}{\epsilon}\,\mathrm{d}x = 0. \tag{3}$$

Taking $\epsilon \to 0$, if we can exchange the limit with the integral, we obtain

$$\begin{aligned}
\int_{\mathbb{R}^d} \langle \nabla f(x),\ v\rangle\,\mathrm{d}x &= \int_{\mathbb{R}^d} \lim_{\epsilon\to 0} \frac{f(x+\epsilon v) - f(x)}{\epsilon}\,\mathrm{d}x \\
&= \lim_{\epsilon\to 0} \int_{\mathbb{R}^d} \frac{f(x+\epsilon v) - f(x)}{\epsilon}\,\mathrm{d}x \\
&= 0.
\end{aligned}$$

Since this holds for any direction v, we conclude $\int_{\mathbb{R}^d} \nabla f(x)\,\mathrm{d}x = 0$.

To make this rigorous, we need a condition that justifies passing the limit through the integral.

Definition 1 (L^1-Taylor Approximation (Definition 1.1 of Spector [2])). A function $f\colon \mathbb{R}^d \to \mathbb{R}$ is said to have a first order L^1-Taylor approximation if $f \in L^1(\mathbb{R}^d)$ and there exists a function $g \in L^1(\mathbb{R}^d; \mathbb{R}^d)$ such that

$$\lim_{\epsilon \to 0} \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} \left| \frac{f(x+\epsilon v) - f(x)}{\epsilon} - \langle g(x),\, v\rangle \right| \mathrm{d}x\, \mathrm{d}S(v) = 0, \tag{4}$$

where $\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d \colon |v| = 1\}$ is the unit sphere equipped with surface measure $\mathrm{d}S(v)$. In this case, we write $\nabla f = g$.

Theorem 1. If $f \in L^1(\mathbb{R}^d)$ is L^1-differentiable, then

$$\int_{\mathbb{R}^d} \nabla f(x)\, \mathrm{d}x = 0.$$

Proof. Let $I = \int_{\mathbb{R}^d} \nabla f(x)\, \mathrm{d}x$. For every unit vector $v \in \mathbb{S}^{d-1}$ and $\epsilon \neq 0$, we have by (3):

$$\int_{\mathbb{R}^d} \frac{f(x+\epsilon v) - f(x)}{\epsilon}\, \mathrm{d}x = 0.$$

We introduce this zero term into the projection of I onto v:

$$\begin{aligned}\langle I, v\rangle &= \int_{\mathbb{R}^d} \langle \nabla f(x), v\rangle\, \mathrm{d}x \\ &= \int_{\mathbb{R}^d} \left(\langle \nabla f(x), v\rangle - \frac{f(x+\epsilon v) - f(x)}{\epsilon} \right) \mathrm{d}x.\end{aligned}$$

Now, consider the integral over the unit sphere. We bound the magnitude of the projection:

$$\begin{aligned}\int_{\mathbb{S}^{d-1}} |\langle I, v\rangle|\, \mathrm{d}S(v) &= \int_{\mathbb{S}^{d-1}} \left| \int_{\mathbb{R}^d} \left(\langle \nabla f(x), v\rangle - \frac{f(x+\epsilon v) - f(x)}{\epsilon} \right) \mathrm{d}x \right| \mathrm{d}S(v) \\ &\leq \int_{\mathbb{S}^{d-1}} \int_{\mathbb{R}^d} \left| \langle \nabla f(x),\, v\rangle - \frac{f(x+\epsilon v) - f(x)}{\epsilon} \right| \mathrm{d}x\, \mathrm{d}S(v).\end{aligned}$$

Taking $\epsilon \to 0$, the right-hand side vanishes by definition (4). Thus,

$$\int_{\mathbb{S}^{d-1}} |\langle I, v\rangle|\, \mathrm{d}S(v) = 0.$$

Since the function $v \mapsto |\langle I, v\rangle|$ is continuous and non-negative, it must be identically zero on the sphere. This implies $I = 0$. □

Sobolev space $W^{1,1}$. [2] proved that the L^1 Taylor approximation condition is equivalent to the requirement that f admits weak first derivatives belonging to $L^1(\mathbb{R}^d)$, namely that $f \in W^{1,1}(\mathbb{R}^d)$. Hence, the condition above may simply be replaced by $f \in W^{1,1}(\mathbb{R}^d)$.

Definition 2 (Weak Derivative and $W^{1,p}(\mathbb{R}^d)$). A function $v_i \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ is the *i-th weak derivative* of $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ if

$$\int_{\mathbb{R}^d} f(x)\, \partial_i \varphi(x)\, \mathrm{d}x = -\int_{\mathbb{R}^d} v_i(x)\, \varphi(x)\, \mathrm{d}x, \qquad \forall\, \varphi \in C_c^\infty(\mathbb{R}^d).$$

We write $\partial_i f = v_i$, and $\nabla f = (\partial_1 f, \ldots, \partial_d f)$.

For $1 \le p \le \infty$, the *Sobolev space* $W^{1,p}(\mathbb{R}^d)$ consists of all functions $f \in L^p(\mathbb{R}^d)$ whose weak partial derivatives $\partial_i f$ exist and belong to $L^p(\mathbb{R}^d)$. The $W^{1,p}$ norm is defined as:

$$\|f\|_{W^{1,p}} := \|f\|_{L^p} + \sum_{i=1}^{d} \|\partial_i f\|_{L^p}.$$

Theorem 2 (Theorem 1.3 of [2])**.** Suppose $f \in L^1(\mathbb{R}^d)$. Then $f \in W^{1,1}(\mathbb{R}^d)$ if and only if f has a first order L^1-Taylor approximation.

Theorem 1 therefore implies the standard result:

Theorem 3. If $f \in W^{1,1}(\mathbb{R}^d)$, then $\int_{\mathbb{R}^d} \nabla f(x)\, \mathrm{d}x = 0$.

Stein's Identity. Applying the Leibniz product rule $\nabla(pg) = (\nabla p)g + p(\nabla g)$ yields integration by parts (with zero boundary) and Stein's identity.

Theorem 4. Let $p, g \colon \mathbb{R}^d \to \mathbb{R}$. Suppose $p, g \in W^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$. Then the product $pg \in W^{1,1}(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ and:

$$\int_{\mathbb{R}^d} (\nabla p(x)g(x) + p(x)\nabla g(x))\, \mathrm{d}x = 0.$$

In particular, if p is a probability density function, this yields Stein's identity:

$$\mathbb{E}_{X\sim p}[\nabla \log p(X)g(X) + \nabla g(X)] = 0.$$

Proof. The assumption $p, g \in W^{1,1}(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ ensures that the product rule $\nabla(pg) = (\nabla p)g + p(\nabla g)$ holds, as stated in Proposition 9.4 of [1]. The conclusion then follows directly.

$\square$

References

[1] Brezis, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer.

[2] Spector, D. E. (2015). l^p-taylor approximations characterize the Sobolev space $w^{1,p}$. *Comptes Rendus Mathématique*, 353(4):327–332.