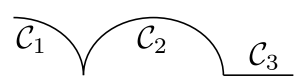
More About Which Vector Fields Obey $\nabla \times \mathbf{F} = 0$

Preliminaries

We already know that if a vector field $\mathbf{F}$ passes the screening test $\nabla \times \mathbf{F} = 0$ on *all* of $\mathbb{R}^2$ or $\mathbb{R}^3$, then there is a function φ with $\mathbf{F} = \nabla\varphi$ (that is, $\mathbf{F}$ is conservative). We are now going to take a first look at what happens when $\mathbf{F}$ passes the screening test $\nabla \times \mathbf{F} = 0$ on only on some proper subset $\mathcal{D}$ of $\mathbb{R}^n$. We will just scratch the surface of this topic — there is a whole subbranch of Mathematics (cohomology theory, which is part of algebraic topology) concerned with a general form of this question. We shall imagine that we are given a vector field $\mathbf{F}$ that is only defined on $\mathcal{D}$ and we shall assume

- that $\mathcal{D}$ is a connected, open subset of $\mathbb{R}^n$ (defined in Definition 1, below)
- that all first order derivatives of all vector fields and functions that we consider are continuous and
- that all curves we consider are piecewise smooth. A curve is piecewise smooth if it is a union of a finite number of smooth curves $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_m$ with the end point of $\mathcal{C}_i$ being the beginning point of $\mathcal{C}_{i+1}$ for each $1 \le i < m$. A curve is smooth if it has a parametrization $\mathbf{r}(t)$, $a \le t \le b$, whose first derivative $\mathbf{r}'(t)$ exists, is continuous and is nonzero everywhere.



Definition 1 Let $n \ge 1$ be an integer.

(a) Let $\mathbf{a} \in \mathbb{R}^n$ and $\varepsilon > 0$. The open ball of radius ε centred on $\mathbf{a}$ is

$$B_\varepsilon(\mathbf{a}) = \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; |\mathbf{x} - \mathbf{a}| < \varepsilon \right\}$$

Note the strict inequality in $|\mathbf{x} - \mathbf{a}| < \varepsilon$.

(b) A subset $\mathcal{O} \subset \mathbb{R}^n$ is said to be an "open subset of $\mathbb{R}^n$" if, for each point $\mathbf{a} \in \mathcal{O}$, there is an $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{a}) \subset \mathcal{O}$. Equivalently, $\mathcal{O}$ is open if and only if it is a union of open balls.

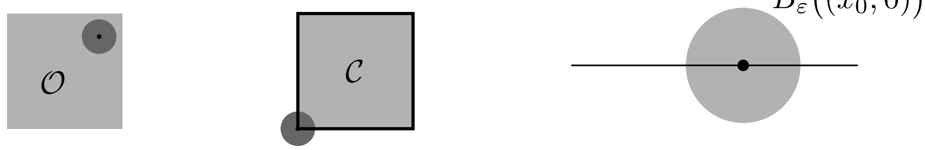
(c) A subset $\mathcal{D} \subset \mathbb{R}^n$ is said to be (pathwise) connected if every pair of points in $\mathcal{D}$ can be joined by a piecewise smooth curve in $\mathcal{D}$.

Example 2

(a) The "open rectangle" $\mathcal{O} = \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; 0 < x < 1,\ 0 < y < 1 \right\}$ is an open subset of $\mathbb{R}^2$ because any point $\mathbf{a} = (x_0, y_0) \in \mathcal{O}$ is a nonzero distance, namely

$d = \min\{x_0, 1-x_0, y_0, 1-y_0\}$ away from the boundary of $\mathcal{O}$. So the open ball $B_{d/2}(\mathbf{a})$ is contained in $\mathcal{O}$.

(b) The "closed rectangle" $\mathcal{C} = \{ (x,y) \in \mathbb{R}^2 \mid 0 \le x \le 1,\ 0 \le y \le 1 \}$ is an *not* open subset of $\mathbb{R}^2$. For example, $\mathbf{0} = (0,0)$ is a point in $\mathcal{C}$. No matter what $\varepsilon > 0$ we pick, the open ball $B_\varepsilon(\mathbf{0})$ is not contained in $\mathcal{C}$ because $B_\varepsilon(\mathbf{0})$ contains the point $(-\frac{\varepsilon}{2}, 0)$, which is not in $\mathcal{C}$.



(c) The x–axis, $\mathcal{X} = \{ (x,y) \in \mathbb{R}^2 \mid y = 0 \}$, in $\mathbb{R}^2$ is *not* an open subset of $\mathbb{R}^2$ because for any point $(x_0, 0) \in \mathcal{X}$ and any $\varepsilon > 0$, the ball $B_\varepsilon\big((x_0,0)\big)$ contains points with nonzero y–coordinates and so is not contained in $\mathcal{X}$.

Review

Many of the basic facts that we developed about conservative fields in $\mathbb{R}^n$ also applies (with the same proofs) to fields on $\mathcal{D}$. In particular, for a vector field $\mathbf{F}$ on $\mathcal{D} \subset \mathbb{R}^n$,

$$\begin{aligned}
\mathbf{F} \text{ is conservative on } \mathcal{D} &\iff \mathbf{F} = \boldsymbol{\nabla}\varphi \text{ on } \mathcal{D}, \text{ for some function } \varphi \text{ (again the definition)} \\
&\iff \text{for each } P, Q \in \mathcal{D} \text{ the work integral } \int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} \text{ takes} \\
&\qquad\qquad \text{the same value for all curves } \mathcal{C} \text{ from } P \text{ to } Q \\
&\iff \int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = 0 \text{ for all closed curves } \mathcal{C} \text{ in } \mathcal{D} \\
&\implies \boldsymbol{\nabla} \times \mathbf{F} = 0 \text{ on } \mathcal{D}
\end{aligned}$$

Combining this with Stokes' theorem gives us the following two consequences.

- If $\mathcal{D}$ has the property that

  $$\text{every closed curve } \mathcal{C} \text{ in } \mathcal{D} \text{ is the boundary of a bounded oriented surface, } \mathcal{S}, \text{ in } \mathcal{D} \tag{H}$$

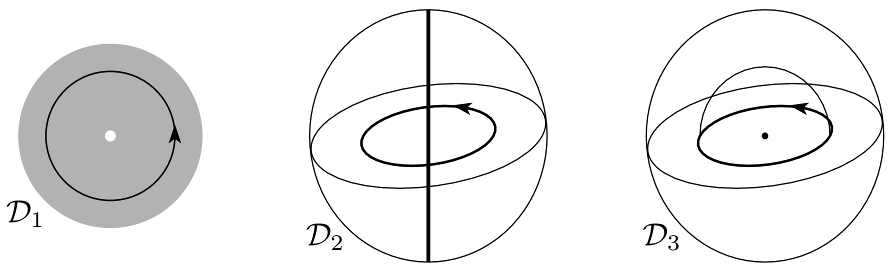
  then

  $$\mathbf{F} \text{ is conservative on } \mathcal{D} \iff \boldsymbol{\nabla} \times \mathbf{F} = 0 \text{ on } \mathcal{D}$$

  This is just because if $\boldsymbol{\nabla} \times \mathbf{F} = 0$ on $\mathcal{D}$ and if the curve $\mathcal{C} = \partial\mathcal{S}$, with $\mathcal{S}$ an oriented surface in $\mathcal{D}$, then Stokes' theorem gives $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \int_{\partial\mathcal{S}} \mathbf{F} \cdot d\mathbf{r} = \iint_{\mathcal{S}} \boldsymbol{\nabla} \times \mathbf{F} \cdot \hat{\mathbf{n}}\, dS = 0$.
- For any $\mathcal{D}$, if $\boldsymbol{\nabla} \times \mathbf{F} = 0$ on $\mathcal{D}$, then $\mathbf{F}$ is locally conservative. This means that for each point $\mathbf{x}_0 \in \mathcal{D}$, there is an $\varepsilon > 0$ and a function φ such that $\mathbf{F} = \boldsymbol{\nabla}\varphi$ on $B_\varepsilon(\mathbf{x}_0)$. This is true just because $B_\varepsilon(\mathbf{x}_0)$ satisfies property (H).

Example 3 Here are some examples of $\mathcal{D}$'s that violate (H).

- When $\mathcal{D} = \mathcal{D}_1 = \big\{\ (x, y) \in \mathbb{R}^2\ \big|\ 0 < |(x, y)| < 3\ \big\}$, the circle $x^2 + y^2 = 4$ is a curve in $\mathcal{D}$ that is not the boundary of a surface in $\mathcal{D}$. The circle $x^2 + y^2 = 4$ is the boundary of the disk $x^2 + y^2 < 4$, but the disk $x^2 + y^2 < 1$ is not contained in $\mathcal{D}$ because the point $(0, 0)$ is in the disk and not in $\mathcal{D}$. See the figure on the left below.
- When $\mathcal{D} = \mathcal{D}_2 = \big\{\ (x, y, z) \in \mathbb{R}^3\ \big|\ |(x, y, z)| < 2, |(x, y)| > 0\ \big\}$, the circle $x^2 + y^2 = 1$, $z = 0$ is a curve in $\mathcal{D}$ that is not the boundary of a surface in $\mathcal{D}$. The circle is the boundary of many different surfaces in $\mathbb{R}^3$, but each contains a point on the z–axis and so is not contained in $\mathcal{D}$. See the figure in the centre below.



On the other hand, here is an example which does satisfy (H).

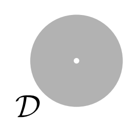
- $\mathcal{D} = \mathcal{D}_3 = \big\{\ (x, y, z) \in \mathbb{R}^3\ \big|\ 0 < |(x, y, z)| < 2\ \big\}$. For example the circle $x^2 + y^2 = 1$, $z = 0$ is the boundary of $\big\{\ (x, y, z) \in \mathbb{R}^3\ \big|\ x^2 + y^2 + z^2, z > 0\ \big\} \subset \mathcal{D}$. See the figure on the right above.

This leaves the question of what happens when (H) is violated. We'll just look at one example, which however gives the flavour of the general theory.

The Punctured Disk in $\mathbb{R}^2$

The punctured disk is

$$\mathcal{D} = \big\{\ (x, y) \in \mathbb{R}^2\ \big|\ 0 < |(x, y)| < 1\ \big\}$$



We'll start by looking at one particular vector field, which passes the screening test, but which cannot possibly be conservative. The field is

$$\boldsymbol{\Theta} = -\tfrac{y}{x^2+y^2}\,\hat{\boldsymbol{\imath}} + \tfrac{x}{x^2+y^2}\,\hat{\boldsymbol{\jmath}}$$

with domain of definition $\mathcal{D}$. We'll first check that it passes the screening test:

$$\begin{aligned}\boldsymbol{\nabla} \times \boldsymbol{\Theta} &= \Big\{\tfrac{\partial}{\partial x}\big(\tfrac{x}{x^2+y^2}\big) - \tfrac{\partial}{\partial y}\big(-\tfrac{y}{x^2+y^2}\big)\Big\}\hat{\mathbf{k}} = \Big\{\big(\tfrac{1}{x^2+y^2} - \tfrac{2x^2}{(x^2+y^2)^2}\big) + \big(\tfrac{1}{x^2+y^2} - \tfrac{2y^2}{(x^2+y^2)^2}\big)\Big\}\hat{\mathbf{k}}\\ &= 0\end{aligned}$$

Next we'll check that it cannot be conservative. Denote by C_ε the circle $x^2+y^2=\varepsilon^2$, with counterclockwise orientation. Parametrize C_ε by $\mathbf{r}(\theta)=\varepsilon\cos\theta\,\hat{\boldsymbol{\imath}}+\varepsilon\sin\theta\,\hat{\boldsymbol{\jmath}}$ with $0\le\theta\le 2\pi$. Then

$$\begin{aligned}\int_{C_\varepsilon}\boldsymbol{\Theta}\cdot d\mathbf{r} &= \int_0^{2\pi}\boldsymbol{\Theta}\big(\mathbf{r}(\theta)\big)\cdot\tfrac{d\mathbf{r}}{d\theta}(\theta)\ d\theta\\ &= \int_0^{2\pi}\big(-\tfrac{1}{\varepsilon}\sin\theta\,\hat{\boldsymbol{\imath}}+\tfrac{1}{\varepsilon}\cos\theta\,\hat{\boldsymbol{\jmath}}\big)\cdot\big(-\varepsilon\sin\theta\,\hat{\boldsymbol{\imath}}+\varepsilon\cos\theta\,\hat{\boldsymbol{\jmath}}\big)\ d\theta\\ &= \int_0^{2\pi}d\theta\\ &= 2\pi\end{aligned}\tag{1}$$

is not zero, so that $\boldsymbol{\Theta}$ cannot be conservative on the punctured disk.

Next we'll check that it is locally conservative. That is, it can be written in the form $\boldsymbol{\nabla}\theta(x,y)$ near any point (x_0,y_0) in its domain. Define $\theta(x,y)$ to be the polar angle of (x,y) with, for example, $-\pi<\theta<\pi$. This θ is defined on all of $\mathcal{D}$, *except* for the negative real axis. The domain of definition, $\mathcal{D}_\pi$, is sketched on the left below. If (x_0,y_0) happens to lie

on the negative real axis, just replace $-\pi<\theta<\pi$ by a different interval of length 2π, like $0<\theta<2\pi$. The domain of definition of θ would then change to the $\mathcal{D}_0$, sketched on the right above. It's now a simple matter to check that $\boldsymbol{\nabla}\theta(x,y)=\boldsymbol{\Theta}(x,y)$ on the domain of definition of θ. If $x\ne 0$, then $\tan\theta(x,y)=\frac{y}{x}$, and

$$\begin{aligned}\tfrac{\partial}{\partial x}\tan\theta(x,y)=-\tfrac{y}{x^2}\quad&\Longrightarrow\quad\big[\tfrac{\partial}{\partial x}\theta(x,y)\big]\ \sec^2\theta(x,y)=-\tfrac{y}{x^2}\\ &\Longrightarrow\quad\tfrac{\partial}{\partial x}\theta(x,y)=-\tfrac{y}{x^2}\cos^2\theta(x,y)=-\tfrac{y}{x^2}\tfrac{x^2}{x^2+y^2}=-\tfrac{y}{x^2+y^2}\\ \tfrac{\partial}{\partial y}\tan\theta(x,y)=\tfrac{1}{x}\quad&\Longrightarrow\quad\big[\tfrac{\partial}{\partial y}\theta(x,y)\big]\ \sec^2\theta(x,y)=\tfrac{1}{x}\\ &\Longrightarrow\quad\tfrac{\partial}{\partial y}\theta(x,y)=\tfrac{1}{x}\cos^2\theta(x,y)=\tfrac{1}{x}\tfrac{x^2}{x^2+y^2}=\tfrac{x}{x^2+y^2}\end{aligned}$$

If $x=0$, then we must have $y\ne 0$ and we can use $\cot\theta(x,y)=\frac{x}{y}$ instead.

We are now ready to consider any vector field $\mathbf{F}$ on $\mathcal{D}$ that passes the screening test $\boldsymbol{\nabla}\times\mathbf{F}=0$ on $\mathcal{D}$. I claim that there is a function φ on $\mathcal{D}$ such that

$$\mathbf{F}=\alpha_{\mathbf{F}}\boldsymbol{\Theta}+\boldsymbol{\nabla}\varphi\qquad\text{where}\qquad\alpha_{\mathbf{F}}=\tfrac{1}{2\pi}\oint_{C_\varepsilon}\mathbf{F}\cdot d\mathbf{r}\tag{2}$$

The significance of this claim is that it says that if a vector field on $\mathcal{D}$ passes the screening test on $\mathcal{D}$ then either it is conservative (that's the case if and only if $\alpha_{\mathbf{F}}=0$) or, if it fails to be conservative, then it differs from a conservative field (namely $\boldsymbol{\nabla}\varphi$) only by a constant (namely $\alpha_{\mathbf{F}}$) times the fixed vector field $\boldsymbol{\Theta}$. That is, there is only one nonconservative vector field on $\mathcal{D}$ that passes the screening test, up to multiplication by constants and addition of conservative fields.

Observe that in the definition of $\alpha_{\mathbf{F}}$, I did not specify the radius ε of the circle C_ε to be used for the integration curve. That's because the answer to the integral does not depend on the choice of ε. To see this, take any $0 < \varepsilon' < \varepsilon < 1$. Then the curve $C_\varepsilon - C_{\varepsilon'}$ is the boundary of $S = \big\{ (x, y) \in \mathbb{R}^2 \bigm| \varepsilon' < |(x, y)| < \varepsilon \big\}$, which is completely contained in $\mathcal{D}$. So, by Stokes' theorem,

$$\oint_{C_\varepsilon} \mathbf{F} \cdot d\mathbf{r} - \oint_{C_{\varepsilon'}} \mathbf{F} \cdot d\mathbf{r} = \oint_{C_\varepsilon - C_{\varepsilon'}} \mathbf{F} \cdot d\mathbf{r} = \oint_{\partial S} \mathbf{F} \cdot d\mathbf{r} = \iint_S \boldsymbol{\nabla} \times \mathbf{F} \cdot \hat{\mathbf{n}}\, dS = 0$$

Finally to verify the claim (2), we check that the vector field $\mathbf{G} = \mathbf{F} - \alpha_{\mathbf{F}} \boldsymbol{\Theta}$ is conservative on $\mathcal{D}$. To do so, it suffices to check that $\oint_{\mathcal{C}} \mathbf{G} \cdot d\mathbf{r} = 0$ for any closed curve $\mathcal{C}$ in $\mathcal{D}$. In fact we can restrict our attention to curves $\mathcal{C}$ that are simple, closed, counterclockwise oriented curves on $\mathcal{D}$. A curve is called simple if it does not cross itself. Closed curves which are not simple can be split up into simple closed subcurves. And changing the orientation of $\mathcal{C}$ just changes the sign of $\oint_{\mathcal{C}} \mathbf{G} \cdot d\mathbf{r} = 0$, which does not affect whether it is zero or not. So let $\mathcal{C}$ be a simple, closed, counterclockwise oriented curve in $\mathcal{D}$. We need to verify that $\oint_{\mathcal{C}} \mathbf{G} \cdot d\mathbf{r} = 0$. Any simple closed curve in $\mathbb{R}^2$ divides $\mathbb{R}^2$ into three mutually disjoint subsets — $\mathcal{C}$ itself, the set of points inside $\mathcal{C}$ and the set of points outside $\mathcal{C}$. Since $(0, 0)$ is not on $\mathcal{C}$, it must be either inside $\mathcal{C}$ or outside $\mathcal{C}$.

- *Case 1:* $(0, 0)$ *outside* $\mathcal{C}$. In this case $\mathcal{C}$ is the boundary of a set, S, which is completely contained in $\mathcal{D}$, namely all of the points inside $\mathcal{C}$. So, by Stokes' theorem,

$$\oint_{\mathcal{C}} \mathbf{G} \cdot d\mathbf{r} = \oint_{\partial S} \big(\mathbf{F} - \alpha_{\mathbf{F}} \boldsymbol{\Theta}\big) \cdot d\mathbf{r} = \iint_S \boldsymbol{\nabla} \times \mathbf{F} \cdot \hat{\mathbf{n}}\, dS - \alpha_{\mathbf{F}} \iint_S \boldsymbol{\nabla} \times \boldsymbol{\Theta} \cdot \hat{\mathbf{n}}\, dS = 0$$

- *Case 2:* $(0, 0)$ *inside* $\mathcal{C}$. Since $(0, 0)$ is not on $\mathcal{C}$, we can choose ε small enough that the circle C_ε lies completely inside $\mathcal{C}$. Then the curve $\mathcal{C} - C_\varepsilon$ is the boundary of a set, S, which is completely contained in $\mathcal{D}$, namely the part of $\mathcal{D}$ that is between C_ε and $\mathcal{C}$. So, by Stokes' theorem,

$$\oint_{\mathcal{C}} \mathbf{G} \cdot d\mathbf{r} - \oint_{C_\varepsilon} \mathbf{G} \cdot d\mathbf{r} = \oint_{\mathcal{C} - C_\varepsilon} \mathbf{G} \cdot d\mathbf{r} = \oint_{\partial S} \mathbf{G} \cdot d\mathbf{r} = \iint_S \boldsymbol{\nabla} \times \mathbf{G} \cdot \hat{\mathbf{n}}\, dS = 0$$

  since $\boldsymbol{\nabla} \times \mathbf{G} = \boldsymbol{\nabla} \times \mathbf{F} - \alpha_{\mathbf{F}} \boldsymbol{\nabla} \times \boldsymbol{\Theta} = 0$ on $\mathcal{D}$. Hence

$$\oint_{\mathcal{C}} \mathbf{G} \cdot d\mathbf{r} = \oint_{C_\varepsilon} \mathbf{G} \cdot d\mathbf{r} = \oint_{C_\varepsilon} \mathbf{F} \cdot d\mathbf{r} - \alpha_{\mathbf{F}} \oint_{C_\varepsilon} \boldsymbol{\Theta} \cdot d\mathbf{r} = 2\pi\alpha_{\mathbf{F}} - \alpha_{\mathbf{F}}(2\pi) = 0$$

  by the definition, (2), of $\alpha_{\mathbf{F}}$ and (1).

So $\mathbf{G}$ is conservative on $\mathcal{D}$ and $\mathbf{F}$ is of the form (2) on $\mathcal{D}$.