Infinite Dimensional Gradient Flow

John McCuan

February 27, 2021

I want to paint for you a picture of gradient flow in infinite dimensions. In order for you to discern that picture you will, first of all, need to know what to look for. To prepare our eyes, let us consider the "players" or components in finite dimensional gradient flow where things are clear—and even in the simplest of examples.

We are first given a set $\mathcal{U} \subset \mathbb{R}^n$. In our simplest example we take $\mathcal{U} = \mathbb{R}^2$. We postpone any lengthy discussion of characteristics or properties of the flow domain $\mathcal{U}$, but let us certainly assume $\mathcal{U}$ is open. This means there is enough structure to make sense of open sets. We know one minimal such structure is that of a metric space, and $\mathbb{R}^n$ is certainly a metric space.

Perhaps the next player to consider is the function which drives the flow. In finite dimensions this is a real valued function $u : \mathcal{U} \to \mathbb{R}$ defined on the flow domain. For our simplest example, we take $u : \mathbb{R}^2 \to \mathbb{R}$ by

$$u(\mathbf{x}) = u(x, y) = \frac{1}{4}(x^2 + y^2) + 1.$$

At this point we can introduce some subsidiary pictures to illustrate these players in the finite dimensional case; see Figure 1.
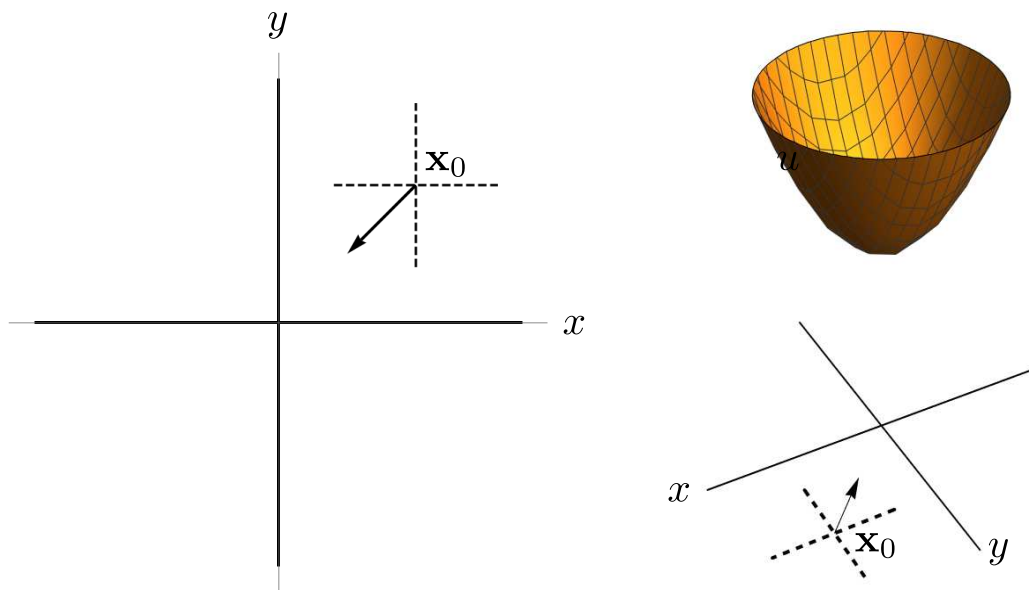

Figure 1: The domain for a gradient flow (left) and the graph of the function which drives the flow with its gradient.

It should be no surprise that subsidiary pictures like these will not be available, strictly speaking, in the infinite dimensional case. The illustration of the graph on the right in Figure 1, for example, already has no comparable version when $n = 3$. (Try to draw the graph of $u : \mathbb{R}^3 \to \mathbb{R}$ by $u(x, y, z) = x^2 + y^2 + z^2$.)

Nevertheless, for what they illustrate these subsidiary pictures are helpful, and one would do well to pay close attention to them.

Let us give names to the three players introduced so far:

(b) The **background** (vector) **space** $\mathbb{R}^n$.

(c) The **domain** of the flow $\mathcal{U} \subset \mathbb{R}^n$.

(d) The **driving function**(al) $u : \mathcal{U} \to \mathbb{R}$.

Perhaps the next player to introduce is

(e) The **equation** of the flow:

$$\frac{d}{dt}\mathbf{x} = -Du(\mathbf{x}).$$ (1)

This equation involves two additional players worth note on their own:

(f) The **moving point** $\mathbf{x} : I \to \mathcal{U}$.

(g) The **gradient field** $Du : \mathcal{U} \to \mathbb{R}^n$.

In our simplest example, the interval I may be taken to be $\mathbb{R}$. For gradient flow in finite dimensions, the interval I is most often taken to be of the form $[0, T)$ where T is some right extremity for existence; see (3) below. In some cases, as in our simplest example, $T = \infty$. It is also usual to assume the function $\mathbf{x}$ can be extended (uniquely) according to (1) to an interval of the form $(-\epsilon, \epsilon)$ for some $\epsilon > 0$, and this may be considered a kind of minimal form for I. In point of fact, the actual value $\mathbf{x}(0)$ for $0 \in I$ is entirely a matter of convention, and the interval of definition for $\mathbf{x}$ can always be shifted to an (open) interval of the form $(t_0 - \epsilon, t_0 + \epsilon)$.

The most important thing to note about the moving point $\mathbf{x}$, which moves as specified by the flow equation (1), is that the domain of $\mathbf{x}$ is an **interval** in $\mathbb{R}$, interpreted as a "time" interval, identifying $\mathbf{x}$ fundamentally as an object from the study of ODEs and, in the finite dimensional case, precisely the object of study in elementary ODEs, namely a solution of an initial value problem for a system of ordinary differential equations. In particular, the standard interpretation of $\mathbf{x}$ as the parameterization of a **curve** in the domain of the flow and having a well-defined **tangent velocity vector**

$$\mathbf{x}' = \frac{d}{dt}\mathbf{x} : I \to \mathbb{R}^n$$ (2)

applies. Figure 2 supplies some additional subsidiary pictures intended to aid in visualization of the moving point $\mathbf{x}$. Having recognized $\mathbf{x}$ in the gradient flow equation as the solution of an autonomous ODE, it is natural to introduce

(h) The **initial point** $\mathbf{x}_0 \in \mathcal{U}$

and recognize $\mathcal{U}$ as a **phase space** for the gradient flow. Thus, we are led to the initial value problem (IVP)

$$\begin{cases} \mathbf{x}' = -Du(\mathbf{x}) \\ \mathbf{x}(0) = \mathbf{x}_0. \end{cases}$$ (3)

Let us now return to discuss the centrally important component of the gradient flow equation, namely,

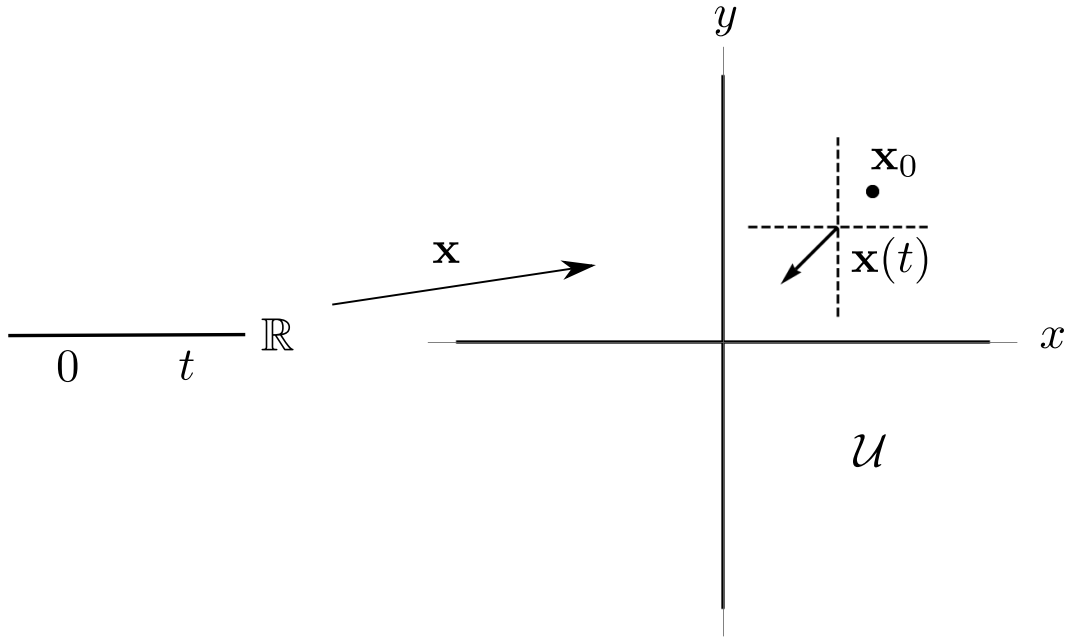**(g)** The **gradient field** $Du : \mathcal{U} \to \mathbb{R}^n$ of the driving function.

Figure 2: The mapping $\mathbf{x} : I \to \mathcal{U}$ and its tangent mapping $bx' : I \to \mathbb{R}^n$.

The central problem, or at least one of the central problems, in defining a gradient flow in infinite dimensions is the **identification of the gradient**. In finite dimensions we do not have this problem due to the existence of readily available coordinates. In the finite dimensional case, for example, there are standard coordinate directions $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ in the background space $\mathbb{R}^n$ for which the associated **directional derivatives** are partial derivatives and

$$Du = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \ldots, \frac{\partial u}{\partial x_n} \right). \tag{4}$$

In our infinite dimensional setting, we will have a kind of directional derivative, but no specific coordinate directions to use to find a gradient. To highlight this difficulty and emphasize the need for a gradient field independent of coordinates, let us introduce the notation

$$\operatorname{grad} u : \mathcal{U} \to \mathbb{R}^n$$

for the gradient field (without coordinates). This designation is used especially to denote a gradient when coordinates are not specified. The initial contemplation of such a possibility may seem a strange or difficult (or unnecessary) abstraction to the student of elementary calculus. How else might one think of a gradient other than (4)? We are about to see, and we are about to see the necessity of such a thing as well.

We have essentially completed out list of active characters participating in gradient flow, but we will add one more playing a kind of supporting role:

(i) The **inequality** on directional derivatives:

$$D_{\mathbf{v}} u(\mathbf{x}) \geq -|\operatorname{grad} u(\mathbf{x})| \qquad \text{with equality for } \mathbf{v} = -\frac{\operatorname{grad} u(\mathbf{x})}{|\operatorname{grad} u(\mathbf{x})|}. \tag{5}$$

Note that the equality condition is degenerate (and not well-defined) when $\operatorname{grad} u = \mathbf{0}$. Indeed one role played by the inequality is to highlight the hope (and objective) of **finding a critical point $\mathbf{x}_* \in \mathcal{U}$** for the driving function u, and perhaps even a minimizer as a limit

$$\mathbf{x}_* = \lim_{t \nearrow T} \mathbf{x}(t).$$

3

In the degenerate case $\operatorname{grad} u = \operatorname{grad} u(\mathbf{x}_0) = \mathbf{0}$ we have evidently already found/started at a critical point. Indeed the ODE (1), as mentioned above, is **autonomous** and a point $\mathbf{x}_*$ for which $Du(\mathbf{x}_*) = \mathbf{0}$ is, by definition, an **equilibrium point** in phase space. Thus, the inequality (5) identifies gradient flow as specifying a motion/evolution of the moving point $\mathbf{x} = \mathbf{x}(t)$ in the direction of maximum decrease of the function(al) u and with speed $|\mathbf{x}'| = |Du(\mathbf{x})|$ the magnitude of the gradient.

1   A coordinate free gradient in finite dimensions

The discussion of the inequality (5) above suggests one immediate coordinate free definition of the gradient of a function $u : \mathcal{U} \to \mathbb{R}$ where $\mathcal{U} \subset \mathbb{R}^n$. Namely, we can consider the values of the directional derivative (restricted to the unit sphere/circle $\partial B_1(\mathbf{0}) \subset \mathbb{R}^n$.

Before we discuss briefly the details of this approach to defining the gradient, let us attempt to put a little more notational difference between the gradient and the directional derivative. It is usual in calculus to denote the gradient and the directional derivative, as we have done above, by Du and $D_{\mathbf{v}}u$ respectively. While these notations look very similar, the objects denoted by them are quite different. We have (ex)changed the notation for the gradient field $Du : \mathcal{U} \to \mathbb{R}^n$ by writing

$$\operatorname{grad} u : \mathcal{U} \to \mathbb{R}^n.$$

Let us also introduce an alternative (though quite standard) notation for the directional derivative function. A directional derivative at a point is most naturally thought of as a (linear) function on vectors (or even unit vectors) in the background space $\mathbb{R}^n$. That is, for each $\mathbf{x} \in \mathcal{U}$ we have a linear functional

$$du_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R} \qquad \text{by} \qquad du_{\mathbf{x}}(\mathbf{v}) = D_{\mathbf{v}}u(\mathbf{x}).$$

Hopefully, this notation creates a clear and strong distinction between the gradient vector $\operatorname{grad} u(\mathbf{x})$ and a directional derivative $du_{\mathbf{x}}(\mathbf{v})$ at $\mathbf{x} \in \mathcal{U}$.

Returning to the idea of defining the gradient by maximizing the directional derivative, if there is a unique unit vector $\mathbf{u}$ for which

$$du_{\mathbf{x}}(\mathbf{u}) = \max_{\mathbf{v} \in \partial B_1(\mathbf{0})} |du_{\mathbf{x}}(\mathbf{v})|, \tag{6}$$

then $\mathbf{u}$ gives us a good candidate for the gradient direction, and we can define

$$\operatorname{grad} u(\mathbf{x}) = |du_{\mathbf{x}}(\mathbf{u})|\, \mathbf{u}.$$

This approach works and is relatively easy in the finite dimensional case. Unfortunately, the minimization problem (6) can be difficult in infinite dimensions, and we will use a different definition.

Another possible approach in the finite dimensional case, which we will use in the infinite dimensional case, is to define the gradient to be the unique vector $\mathbf{w}$ for which the directional derivative is given by an **inner product** in the form

$$du_{\mathbf{x}}(\mathbf{v}) = \mathbf{w} \cdot \mathbf{v} \qquad \text{for every } \mathbf{v} \in \mathbb{R}^n.$$

Then we set $\operatorname{grad} u(\mathbf{x}) = \mathbf{w}$. This does not, in principle, require coordinates. In general writing a linear functional in terms of an inner product with a fixed vector like this is called **Riesz representation**. In the finite dimensional case, we can obtain existence by simply writing down the vector $\mathbf{w} = Du(\mathbf{x})$ in coordinates and using the chain rule. It should be noted that whenever one has a representation like this, uniqueness and the inequality always hold. Let me briefly explain why. First of all, whenever you have a fixed vector $w \in X$ where X is an inner product space, then $L : X \to \mathbb{R}$ by

$$Lx = \langle w, x \rangle \tag{7}$$

defines a linear functional. The

Morever, this functional defined by (7) determines the vector w uniquely because if

$$\langle w, x \rangle = \langle \tilde{w}, x \rangle \qquad \text{for all } x \in X,$$

then by the (bi)linearity of the inner product

$$\langle w - \tilde{w}, x \rangle = 0.$$

Now we can take $x = w - \tilde{w}$ to conclude $\|w - \tilde{w}\|^2 = 0$. Thus, $\tilde{w} = w$ because the inner product is also positive definite.

Such a functional also satisfies a kind of Lipschitz continuity estimate which is called "boundedness" by the Cauchy-Schwarz inequality:

$$|Lx| = |\langle w, x \rangle| \le \|w\| \, \|x\|.$$

In this estimate, think of $\|w\|$ as a non-negative constant.

Finally, for the maximizing property, just imagine $\|x\| = 1$. Then

$$|Lx| \le \|w\|.$$

But also, if $w \ne \mathbf{0}$, then $v = w/\|w\|$ is a unit vector and

$$Lv = \langle w, w/\|w\| \rangle = \langle w, w \rangle / \|w\| = \|w\|^2 / \|w\|.$$

Thus, we know the maximizing direction is in the direction $\mathbf{w}$.

In infinite dimensions, one theorem giving existence is called the **Riesz representation theorem**. In our situation, we will be able to obtain the representation directly without coordinates.

2 Infinite Dimensional Gradient Flow

Let us begin by recording the players listed above:

(b) The **background** (vector) **space** $\mathbb{R}^n$.

(c) The **domain** of the flow $\mathcal{U} \subset \mathbb{R}^n$.

(d) The **driving function**(al) $u : \mathcal{U} \to \mathbb{R}$.

(e) The **equation** of the flow:
$$\frac{d}{dt}\mathbf{x} = -Du(\mathbf{x}).$$

(f) The **moving point** $\mathbf{x} : I \to \mathcal{U}$.

(g) The **gradient field** $Du : \mathcal{U} \to \mathbb{R}^n$.

(h) The **initial point** $\mathbf{x}_0 \in \mathcal{U}$.

(i) The **inequality** on directional derivatives:
$$D_{\mathbf{v}}u \geq -|\operatorname{grad} u|$$

In the infinite dimensional case, we start with the driving function. This will be a Lagrangian integral functional
$$\mathcal{F} : \mathcal{A} \to \mathbb{R}$$
defined on an admissible class $\mathcal{A}$. This choice should bring to mind the framework of the calculus of variations and the objective of finding a minimizer (or extremal) $u_* \in \mathcal{A}$ for $\mathcal{F}$. In particular, we can start to fill in a preliminary version of the list above for the infinite dimensional case:

(a) The **background space** for admissible functions in $\mathcal{A}$.

(b) The **background space** for perturbations in $\mathcal{V}$.

(c) The **domain** of the flow $\mathcal{A}$.

(d) The **driving functional** $\mathcal{F} : \mathcal{A} \to \mathbb{R}$.

(e) The **equation** of the flow (?)

(f) The **moving function** (??)

(g) The **gradient field** $\operatorname{grad} \mathcal{F}(u)$ (???)

(h) The **initial function** $u_0 \in \mathcal{A}$.

(i) The **inequality** on directional derivatives:
$$\delta \mathcal{F}_u[\phi] \geq -|\operatorname{grad} \mathcal{F}|$$

Many blanks need to be filled in, and our preliminary players will require some modification as well. Nevertheless, this is a good start with a solid framework involving a functional $\mathcal{F} : \mathcal{A} \to \mathbb{R}$ from the calculus of variations with a well-defined notion of directional derivative given from the first variation $\delta\mathcal{F} : C_c^\infty(\mathcal{U}) \to \mathbb{R}$.

Perhaps one of the first and easiest topics to address is that of the **moving function**. It should be natural to contemplate a parameterized curve of functions $u : I \to \mathcal{A}$. If each such function $u = u(t)$ is a function of several variables $\mathbf{x} \in \mathcal{U} \subset \mathbb{R}^n$, then it is also quite natural to introduce the dependence on $\mathbf{x}$ directly into an argument in u and consider $U : \mathcal{U} \times I \to \mathbb{R}$. This brings up certain issues of regularity with respect to the **spatial variables x** versus the **time variable** t, but we will mostly set these aside for this discussion.

The important idea is that our gradient flow equation in this case will be a PDE of the form

$$\frac{\partial U}{\partial t} = -\operatorname{grad}\mathcal{F}[u]$$

where $U = U(x, t)$ and $u : \mathcal{U} \to \mathbb{R}$ on the right is defined by $u(x) = U(x, t)$ for t fixed, so u is a function on which the gradient can make sense. This is, on the face of it, quite different from the ODE $\mathbf{x}' = -Du(\mathbf{x})$, but the generalization is rather natural. One big question, of course, is

$$\text{What is } \operatorname{grad}\mathcal{F}?$$

As mentioned above, we will get at $\operatorname{grad}\mathcal{F}$ using an **inner product**. The choice of this inner product is not entirely obvious, but it should be an inner product on functions, and this suggests (perhaps) the choice of $L^2(\mathcal{U})$ as the background space for the perturbations $\mathcal{V}$. The space $L^2(\mathcal{U})$ consists of those functions $v : \mathcal{U} \to \mathbb{R}$ which are square integrable:

$$\int_{\mathcal{U}} |v|^2 < \infty.$$

On this space

$$\langle v, w \rangle_{L^2} = \int_{\mathcal{U}} vw$$

defines an inner product. We also have

$$C_c^\infty(\mathcal{U}) \subset \mathcal{V} \subset L^2(\mathcal{U}).$$

With this choice, we can ask the question:

$$\text{Is there a particular vector } w \in L^2(\mathcal{U}) \text{ for which } \delta\mathcal{F}_u[\phi] = \langle w, \phi \rangle_{L^2} = \int_{\mathcal{U}} w\phi ?$$

We do know a viable candidate.

Recall that for the functional $\mathcal{F} : \mathcal{A} \to \mathbb{R}$ by

$$\mathcal{F}[u] = \int_{\mathcal{U}} F(x, u, Du)$$

where

$$\mathcal{A} = \left\{ u \in C^1(\overline{\mathcal{U}}) : u\big|_{\partial\mathcal{U}} = g \right\}$$

we have

$$\delta\mathcal{F}_u[\phi] = \int_{\mathcal{U}} \left(\sum_{j=1}^n \frac{\partial F}{\partial p_j} \frac{\partial \phi}{\partial x_j} + \frac{\partial F}{\partial z} \right).$$

For $u \in C^2(\overline{\mathcal{U}})$ we can obtain the Euler-Lagrange operator

$$\operatorname{grad} \mathcal{F}[u] = -\sum_{j=1}^{n} \frac{\partial}{\partial x_j} \left(\frac{\partial F}{\partial p_j} \right) + \frac{\partial F}{\partial z}$$

which, at least nominally, looks like it does what we want. We know that for this operator to make sense classically, we need some additional regularity, and there is another minor problem as well. We will make an attempt to smooth these difficulties over below, but we can write down our flow equation at this point. This is (a big part of) what gradient flow in infinite dimensions looks like:

$$\frac{\partial U}{\partial t} = -\operatorname{grad} \mathcal{F}[u] = \sum_{j=1}^{n} \frac{\partial}{\partial x_j} \left(\frac{\partial F}{\partial p_j}(x, U, DU) \right) - \frac{\partial F}{\partial z}(x, U, DU)$$

where it is understood that

$$DU = \left(\frac{\partial U}{\partial x_1}, \frac{\partial U}{\partial x_2}, \ldots, \frac{\partial U}{\partial x_n} \right).$$

This is called the **spatial gradient**. Note there is no partial derivative with respect to t.

In the special case of Dirichlet energy, this gives

$$U_t = \Delta U \qquad \text{(the heat equation)}$$

where Δ is the **spatial Laplacian** given by

$$\Delta U = \sum_{j=1}^{n} \frac{\partial^2 U}{\partial x_j^2}.$$

Thus, the heat equation is not properly variational in our previously discussed sense, but it is (part of) a gradient flow driven by the Dirichlet energy.

3   More Details

Let us specialize to Dirichlet energy for this discussion.

Recall that functions in our admissible class $\mathcal{A}$ have fixed boundary values given by a function g. This function g may be defined only on $\partial \mathcal{U}$ with $g : \partial \mathcal{U} \to \mathbb{R}$, but it must have enough regularity to allow it to be the boundary values of a C^1 function in $\mathcal{A}$. The main point, however, is that if we are going have a flow of functions $U = U(x, t)$ with $u \in \mathcal{A}$ where $u : \mathcal{U} \to \mathbb{R}$ by $u(x) = U(x, t)$ for each fixed t, then we must have

$$\frac{\partial U}{\partial t} \bigg|_{x \in \partial \mathcal{U}} \equiv 0.$$

If it is the case that $u \in C^2(\overline{\mathcal{U}})$, then this means we must also have $\Delta u = 0$ on $\partial \mathcal{U}$. It turns out that gradient flow (the heat equation) in this case, does allow for this condition, but to make everything classically viable, it's easiest to introduce a good deal of regularity.

To this end, let us introduce the smooth admissible class

$$\mathcal{A}^\infty = \left\{ u \in C^\infty(\overline{\mathcal{U}}) : u\big|_{\partial \mathcal{U}} = g \right\}$$

and the restricted admissible class

$$\mathcal{A}_0^\infty = \left\{ u \in C^\infty(\overline{\mathcal{U}}) : u\big|_{\partial \mathcal{U}} = g \text{ and } \Delta u\big|_{\partial \mathcal{U}} \equiv 0 \right\}.$$

We also impose the following restriction on the boundary values: There exists a function $u \in C^2(\overline{\mathcal{U}})$ with

$$u\big|_{\partial\mathcal{U}} = g \qquad \text{and} \qquad \Delta u\big|_{\partial\mathcal{U}} \equiv 0.$$

With these restrictions, we can fill in our list of players a little more properly and completely (at least for Dirichlet energy):

(a) The **background space** for admissible functions in $\mathcal{A}_0^\infty$: $C^\infty(\overline{\mathcal{U}})$.

(b) The **background space** for perturbations in $\mathcal{V}$: $L^2(\mathcal{U})$.

(c) The **domain** of the flow $\mathcal{A}_0^\infty$.

(d) The **driving functional** $\mathcal{F} : \mathcal{A}_0^\infty \to \mathbb{R}$.

(e) The **equation** of the flow:
$$U_t = \Delta U \qquad \text{on } \mathcal{U} \times [0, \infty).$$

(f) The **moving function** $U : \overline{\mathcal{U}} \times [0, \infty)$.

(g) The **gradient field** $\operatorname{grad} \mathcal{F}[U] = -\Delta U$.

(h) The **initial function** $u_0 \in \mathcal{A}_0^\infty$ and the **boundary function** $g \in C^\infty(\partial\mathcal{U})$.

(i) The **inequality** on directional derivatives:

$$\delta\mathcal{F}_u[\phi] \geq -\|\operatorname{grad}\mathcal{F}(u)\|_{L^2} = -\|\Delta u\|_{L^2}.$$

We finish this discussion with the associated **initial/boundary value problem** corresponding to (3):

$$\begin{cases} U_t = \Delta U & \text{on } \mathcal{U} \times [0, \infty), \\ U\big|_{\mathcal{U}\times\{0\}} = u_0, & U\big|_{\partial\mathcal{U}\times[0,\infty)} = g. \end{cases}$$

In practice much much less regularity is required for, say having a well-defined heat flow, than is used in our discussion above. But if one wants purely classical solutions, probably something like the very restrictive hypotheses above are necessary (or at least convenient). At least we have illustrated how the heat equation can arise from a gradient flow in infinite dimensions. Similar considerations are at the foundation of constructions like *mean curvature flow*, but very few books even mention this, much less give a treatment with any details.