1 Course overview

The Internet of Things (IoT) is a term that represents a huge technological trend that is taking place: almost every device is being imbued with the intelligence of a microprocessor and an Internet connection. The interconnection in IoT promises an infrastructure that can drastically change how consumers live their day-to-day lives, with huge gains in efficiency, value, and possibility due to the shared knowledge and autonomy allowed. In profound ways, as the technology develops, the modalities of existence people experience will grow and shift.

However, the scale and scope of IoT raises new problems for engineers to consider. These problems are significantly different from ones previously explored in the design of comparatively isolated systems, and require a new theoretical underpinning to analyze IoT with models that capture all salient facets of these new technologies. This textbook contains a handful of theoretical frameworks, and their applications, as a first step into this new research frontier.

The goal of this course is to cover the theoretical foundations for the analysis of such systems.

This course will cover some of the new problems that arise due to: a) the scale, complexity, and homogeneity of multiple devices, b) the interactions between new service models and human agents, c) the new vulnerabilities that arise due to new interconnections, and d) the structure of these new disruptive markets. In the process, we’ll introduce the theoretical background underpinning many of the different formulations of these novel problems; researchers from many different fields have approached different facets of the new problems in cyberphysical systems, and the goal of this course is to provide the student with enough resources and context to be able to understand the cutting-edge research in several of these fields.

One of the main focuses of this course will be the study of the role of information in these cyberphysical systems. First, we must consider what data is transmitted. Then, we ask ourselves the legitimate value of the data, i.e. how can data be leveraged for effective system operation? Once this data is being used in closed-loop system operation, be it through classical control methods, incentive schemes, or some other new service model, we have to analyze the effect of data in closed-loop. Particularly, once the data is used in system operation, do data sources have incentives to modify or obfuscate their data? What statistical information is contained in the data and does it raise privacy concerns?

Another main focus will be the interaction between humans and the system. We will discuss different methods for modeling humans and how we can learn parameters of these models from data.

This course will not cover how to design physical devices or the protocols for interfacing different devices. This course will not deeply treat machine learning qua machine learning, but will touch on various aspects of machine learning as it relates to these new service models. For example, we will consider how machine learning must change when data is not drawn from a distribution but, rather, is the reported values of strategic agents.

2 Continuous optimization

In this section, we’ll do an overview of some core concepts to optimization. The theory underlying this will be very relevant in a lot of our applications later on in the course, as optimization plays a key role in a lot of different facets of cyberphysical systems. This section will be heavily focused on optimality conditions and duality.

Optimality conditions have been commonly used to verify optimality of a potential optimizer, but has also more recently been used to try and backtrack parameters of the optimization, e.g. given an optimizer, what cost function does it minimize?

Relatively, duality has a deep theoretical literature and has a variety of nice interpretations. For example, dual variables often are seen as shadow prices in optimizations that arise in economics. In fact, in many mathematical formulations of economics, dual variables do the work of the ‘invisible hand’ or arbitrage. As another example, dual variables can be used to coordinate multiple agents in distributed optimization.
2.1 Unconstrained optimization

Let’s consider the unconstrained optimization problem. Suppose we’ve given some function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \). Then, in optimization, we want to find \( \inf_{x \in \mathbb{R}^n} f(x) \). Additionally, we sometimes want to find a \( x^{\text{opt}} \in \mathbb{R}^n \) such that \( f(x^{\text{opt}}) = \inf_x f(x) \). Note that \( \inf_x f(x) \) always exists, while, in general, \( x^{\text{opt}} \) might not always exist and might not be unique if it exists.

We’ll begin with some definitions of the general concepts in unconstrained optimization. Much of this content is drawn from [Bertsekas, 2016, Chapter 1], with some of the notation also borrowed from [Rockafellar, 1997].

**Definition 2.1** (Topological properties). For a point \( x \in \mathbb{R}^n \) and radius \( r > 0 \), we define the open ball \( B_r(x) = \{ y : \|x - y\| < r \} \).

For a set \( A \subset \mathbb{R}^n \), we say \( A \) is open if for every \( x \in A \), there exists \( \epsilon > 0 \) such that \( B_\epsilon(x) \subset A \).

For a set \( A \subset \mathbb{R}^n \), we say \( A \) is closed if \( \mathbb{R}^n \setminus A \) is open.

**Definition 2.2** (Minima). Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \).

A vector \( x \) is an unconstrained local minimum of \( f \) if there exists some \( \epsilon > 0 \) such that \( f(x) \leq f(y) \) for all \( y \in B_\epsilon(x) \). This is strict if the inequality holds strictly for all \( y \neq x \).

A vector \( x \) is an unconstrained global minimum of \( f \) if \( f(x) \leq f(y) \) for all \( y \in \mathbb{R}^n \). This is strict if the inequality holds strictly for all \( y \neq x \).

**Definition 2.3** (Differentiation and stationary points). Let \( f : \mathbb{R}^n \rightarrow \mathbb{R} \).

The directional derivative of \( f \) is defined as:

\[
f'(x; y) = \lim_{\lambda \to 0} \frac{f(x + \lambda y) - f(x)}{\lambda}
\]

We say \( f'(x; y) \) is the directional derivative of \( f \) at \( x \) in direction \( y \).

Recall \( e_i \) denotes the unit vector in the \( i \)th coordinate. If \( f'(x; e_i) = f'(x; -e_i) \), we define the partial derivative at \( x \) with respect to \( x_i \) as:

\[
\frac{\partial f}{\partial x_i}(x) = f'(x; e_i)
\]

If \( f'(x; e_i) \neq f'(x; -e_i) \), this partial derivative is not defined.

If there exists a vector \( x^* \) such that \( \langle x^*, y \rangle = f'(x; y) \) for all \( y \), then we say \( f \) is differentiable at \( x \). If such an \( x^* \) exists, it must necessarily equal the partial derivatives with respect to each coordinate. Thus, when \( f \) is differentiable at \( x \), we define the gradient \( \nabla f \) at \( x \) as:

\[
\nabla f(x) = x^* = \left( \frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)
\]

We say \( f \) is differentiable if \( f \) is differentiable at \( x \) for all \( x \in \mathbb{R}^n \). If \( \nabla f \) is continuous, we say \( f \) is continuously differentiable.

We can also define partial second derivatives. If we introduce the shorthand \( g = \frac{\partial f}{\partial x_j} \), then:

\[
\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_j}(x) \right) = \frac{\partial g}{\partial x_i}(x)
\]

Note that this definition requires the existence of all the partial derivatives.

If all the partial second derivatives exist, then we say \( f \) is twice differentiable at \( x \), and we define the Hessian \( \nabla^2 f \) element-wise:

\[
(\nabla^2 f(x))_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)
\]

We say \( f \) is twice differentiable if \( f \) is twice differentiable at \( x \) for all \( x \in \mathbb{R}^n \). If \( \nabla^2 f \) is continuous, we say \( f \) is twice continuously differentiable.

We say \( x \) is a stationary point of \( f \) if \( \nabla f(x) = 0 \). Note that this implicitly requires that \( \nabla f(x) \) exists.

With these definitions, we note some basic properties.

**Proposition 2.4.** If \( f \) is twice-differentiable at \( x \), then \( \nabla^2 f(x) \) is symmetric, i.e. \( \nabla^2 f(x) = (\nabla^2 f(x))^\top \).

We can also write out necessary and sufficient conditions for optimality.
Proposition 2.5 (Necessary optimality conditions). Let \( x^{opt} \) be an unconstrained local minima of \( f : \mathbb{R}^n \to \mathbb{R} \). Suppose \( f \) is continuously differentiable in an open set \( A \) containing \( x^{opt} \). Then \( \nabla f(x^{opt}) = 0 \).

Furthermore, if \( f \) is twice continuously differentiable within \( A \), then \( \nabla^2 f(x^{opt}) \succeq 0 \).

Proof. Since \( x^{opt} \) is an unconstrained local minimum, we have that \( f'(x^{opt}; y) \geq 0 \) for all \( y \). Fix any \( y \neq 0 \), note that \( \langle \nabla f(x^{opt}), y \rangle = f'(x^{opt}; y) \geq 0 \) by Proposition 2.4, and \( -\langle \nabla f(x^{opt}), y \rangle = \langle \nabla f(x^{opt}), -y \rangle \geq 0 \), which implies that \( \langle \nabla f(x^{opt}), y \rangle = 0 \). This held for any \( y \), which implies that \( \nabla f(x^{opt}) = 0 \).

For the second half of this proof, suppose \( f \) is twice continuously differentiable. Fixing a vector \( y \), we have the Taylor series expansion for any \( \alpha \in \mathbb{R} \):

\[
 f(x^{opt} + \alpha y) = f(x^{opt}) + \alpha \langle \nabla f(x^{opt}), y \rangle + \frac{\alpha^2}{2} y^\top \nabla^2 f(x^{opt}) y + o(\alpha^2)
\]

By the first half of this proof, we know \( \nabla f(x^{opt}) = 0 \), so, using optimality:

\[
 0 \leq \frac{f(x^{opt} + \alpha y) - f(x^{opt})}{\alpha^2} = \frac{1}{2} y^\top \nabla^2 f(x^{opt}) y + \frac{o(\alpha^2)}{\alpha^2}
\]

Taking the limit \( \alpha \to 0 \) yields \( y^\top \nabla^2 f(x^{opt}) y \geq 0 \), and this held for arbitrary \( y \), so \( \nabla^2 f(x^{opt}) \succeq 0 \). \( \square \)

Proposition 2.6 (Sufficient optimality conditions). Let \( f : \mathbb{R}^n \to \mathbb{R} \) be twice continuously differentiable in some open set \( A \). Suppose a vector \( x \in S \) satisfies \( \nabla f(x) = 0 \) and \( \nabla^2 f(x) \succeq 0 \). Then, \( x \) is a strict unconstrained local minimum of \( f \).

Even stronger, there exists scalars \( \gamma > 0 \) and \( \epsilon > 0 \) such that:

\[
 f(y) \geq f(x) + \frac{\gamma}{2} \| y - x \|_2^2 \quad \text{for all } y \in B_\epsilon(x)
\]

Proof. Let \( \lambda \) be the smallest eigenvalue of \( \nabla^2 f(x) \). We know that \( \lambda > 0 \) since \( \nabla^2 f(x) \succeq 0 \). Furthermore, from the eigenvalue decomposition of \( \nabla^2 f(x) \), we can see that \( y^\top \nabla^2 f(x) y \geq \lambda \| y \|_2^2 \) for any \( y \in \mathbb{R}^n \). The second order Taylor series is, for any \( y \):

\[
 f(x + y) - f(x) = \langle \nabla f(x), y \rangle + \frac{1}{2} y^\top \nabla^2 f(x) y + o(\| y \|_2^2)
\]

\[
 \geq \frac{\lambda}{2} \| y \|_2^2 + o(\| y \|_2^2)
\]

\[
 = \left( \frac{\lambda}{2} + o(\| y \|_2^2) \right) \| y \|_2^2
\]

Since \( \lim_{\| y \|_2 \to 0} o(\| y \|_2^2)/\| y \|_2^2 = 0 \), we can pick \( \epsilon > 0 \) and \( \gamma > 0 \) such that, for all \( \| y \|_2 < \epsilon \), we have:

\[
 \left( \frac{\lambda}{2} + o(\| y \|_2^2) \right) \| y \|_2^2 \geq \frac{\gamma}{2}
\]

This is our desired result. \( \square \)

References
