Outline

1. Fundamentals of Unconstrained Optimization
   - Quick Review...
   - Characterizing the Solution
   - Some Fundamental Theorems and Definitions...

2. Optimality
   - Necessary vs. Sufficient Conditions; Convexity
   - From Theorems to Algorithms...
We established that our “favorite problem” for the semester will be of the form

$$\min_{\bar{x} \in \mathbb{R}^n} f(\bar{x}),$$

where

- $f(\bar{x})$ the objective function
- $\bar{x}$ the vector of variables (a.k.a. unknowns, or parameters.)

The problem is unconstrained since all values of $\bar{x} \in \mathbb{R}^n$ are allowed.

Further, we established that our initial approach will focus on problems where we do not have any extra factors working against us, i.e. we are considering local optimization, continuous variables, and deterministic techniques.
A solution to the unconstrained optimization problem is a point \( \bar{x}^* \in \mathbb{R}^n \) such that

\[
f(\bar{x}^*) \leq f(\bar{x}), \quad \forall \bar{x} \in \mathbb{R}^n,
\]
such a point is called a global minimizer.

In order to find a global optimizer we need information about the objective on a global scale.

— Unless we have special information (such as convexity of \( f \)), this information is “expensive” since we would have to evaluate \( f \) in (infinitely?) many points.
Most algorithms will take a starting point $\bar{x}_0$ and use information about $f$, and possibly its derivative(s) in order to compute a point $\bar{x}_1$ which is “closer to optimal” than $\bar{x}_0$, in the sense that

$$f(\bar{x}_1) \leq f(\bar{x}_0).$$

Then the algorithm will use information about $f +$ derivative(s) in $\bar{x}_1$ (and possibly in $\bar{x}_0$ — this increases the storage requirement) to find $\bar{x}_2$ such that

$$f(\bar{x}_2) \leq f(\bar{x}_1) \leq f(\bar{x}_0).$$

An algorithm of this type will only be able to find a **local minimizer**.
A point $\bar{x}^* \in \mathbb{R}^n$ is a **local minimizer** if there is a neighborhood $N$ of $\bar{x}^* \in \mathbb{R}^n$ such that $f(\bar{x}^*) \leq f(\bar{x})$, $\forall \bar{x} \in N$.

**Note:** A neighborhood of $\bar{x}^*$ is an open set which contains $\bar{x}^*$.

**Note:** A local minimizer of this type is sometimes referred to as a **weak local minimizer**. A **strict** or **strong** local minimizer is defined as —

A point $\bar{x}^* \in \mathbb{R}^n$ is a **strict local minimizer** if there is a neighborhood $N$ of $\bar{x}^* \in \mathbb{R}^n$ such that $f(\bar{x}^*) < f(\bar{x})$, $\forall \bar{x} \in N - \{\bar{x}^*\}$.
Definition (Isolated Local Minimizer)

A point $\bar{x}^* \in \mathbb{R}^n$ is an **isolated local minimizer** if there is a neighborhood $N$ of $\bar{x}^* \in \mathbb{R}^n$ such that $\bar{x}^*$ is the only local minimizer in $N$.

**Figure:** The objective $f(x) = x^2(2 + \cos(1/x))$ has a strict local minimizer at $x = 0$, however there are strict local minimizers at infinitely many neighboring points. $x^* = 0$ is not an isolated minimizer.
Recognizing A Local Minimum

If we are given a point $\bar{x} \in \mathbb{R}^n$ how do we know if it is a (local) minimizer?? — Do we have to look at all the points in the neighborhood?

If/when the objective function $f(\bar{x}) \in \mathbb{R}$ is **differentiable** we can recognize a minimum by looking at the first and second derivatives — the **gradient** $\nabla f(\bar{x}) \in \mathbb{R}^n$, and
— the **Hessian** $\nabla^2 f(\bar{x}) \in \mathbb{R}^{n \times n}$.

The key tool is the multi-dimensional version of **Taylor’s Theorem** (Taylor expansions/series).

Illustration: The Gradient ($\nabla f$) and the Hessian ($\nabla^2 f$)

Example: Let $\bar{x} \in \mathbb{R}^3$, i.e.

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

then

$$\nabla f(\bar{x}) = \begin{bmatrix} \frac{\partial f(\bar{x})}{\partial x_1} \\ \frac{\partial f(\bar{x})}{\partial x_2} \\ \frac{\partial f(\bar{x})}{\partial x_3} \end{bmatrix}, \quad \nabla^2 f(\bar{x}) = \begin{bmatrix} \frac{\partial^2 f(\bar{x})}{\partial x_1^2} & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\bar{x})}{\partial x_2^2} & \frac{\partial^2 f(\bar{x})}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f(\bar{x})}{\partial x_1 \partial x_3} & \frac{\partial^2 f(\bar{x})}{\partial x_2 \partial x_3} & \frac{\partial^2 f(\bar{x})}{\partial x_3^2} \end{bmatrix}.$$
Taylor’s Theorem

Theorem (Taylor’s Theorem)

Suppose that \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is continuously differentiable, and that \( \bar{p} \in \mathbb{R}^n \). Then,

\[
f(\bar{x} + \bar{p}) = f(\bar{x}) + \nabla f(\bar{x} + t\bar{p})^T \bar{p},
\]

for some \( t \in (0, 1) \). Moreover, if \( f \) is twice continuously differentiable, then

\[
\nabla f(\bar{x} + \bar{p}) = \nabla f(\bar{x}) + \int_0^1 \nabla^2 f(\bar{x} + t\bar{p})\bar{p} \, dt
\]

and

\[
f(\bar{x} + \bar{p}) = f(\bar{x}) + \nabla f(\bar{x})^T \bar{p} + \frac{1}{2} \bar{p}^T \nabla^2 f(\bar{x} + t\bar{p})\bar{p}
\]

for some \( t \in (0, 1) \).
Theorem (First-order Necessary Conditions)

If $\bar{x}^*$ is a local minimizer and $f$ is continuously differentiable in an open neighborhood of $\bar{x}^*$, then $\nabla f(\bar{x}^*) = 0$. 
Proof (By contradiction).

Suppose $\nabla f(\bar{x}^*) \neq 0$. Let $\bar{p} = -\nabla f(\bar{x}^*)$ and realize that $\bar{p}^T \nabla f(\bar{x}^*) = -\|\nabla f(\bar{x}^*)\|^2 < 0$. 

Peter Blomgren, ⟨blomgren.peter@gmail.com⟩
Proof (By contradiction).

Suppose $\nabla f(\bar{x}^*) \neq 0$. Let $\bar{p} = -\nabla f(\bar{x}^*)$ and realize that

$$\bar{p}^T \nabla f(\bar{x}^*) = -\|\nabla f(\bar{x}^*)\|^2 < 0.$$  

By continuity of $\nabla f$, there is a scalar $T > 0$ such that

$$\bar{p}^T \nabla f(\bar{x}^* + t\bar{p}) < 0, \quad \forall t \in [0, T]$$
Optimality: First Order Necessary Conditions (Proof)

Proof (By contradiction).

Suppose $\nabla f(\bar{x}^*) \neq 0$. Let $\bar{p} = -\nabla f(\bar{x}^*)$ and realize that $\bar{p}^T \nabla f(\bar{x}^*) = -\|\nabla f(\bar{x}^*)\|^2 < 0$. By continuity of $\nabla f$, there is a scalar $T > 0$ such that

$$\bar{p}^T \nabla f(\bar{x}^* + t\bar{p}) < 0, \quad \forall t \in [0, T]$$

Further, for any $s \in (0, T]$, by Taylor’s theorem:

$$f(\bar{x}^* + s\bar{p}) = f(\bar{x}^*) + s \bar{p}^T \nabla f(\bar{x}^* + t\bar{p}), \quad \text{for some } t \in (0, s).$$
Proof (By contradiction).

Suppose $\nabla f(\bar{x}^*) \neq 0$. Let $\bar{p} = -\nabla f(\bar{x}^*)$ and realize that

$$
\bar{p}^T \nabla f(\bar{x}^*) = -\|\nabla f(\bar{x}^*)\|^2 < 0.
$$

By continuity of $\nabla f$, there is a scalar $T > 0$ such that

$$
\bar{p}^T \nabla f(\bar{x}^* + t\bar{p}) < 0, \quad \forall t \in [0, T]
$$

Further, for any $s \in (0, T]$, by Taylor’s theorem:

$$
f(\bar{x}^* + s\bar{p}) = f(\bar{x}^*) + s \bar{p}^T \nabla f(\bar{x}^* + t\bar{p}), \quad \text{for some } t \in (0, s).
$$

Therefore $f(\bar{x}^* + s\bar{p}) < f(\bar{x}^*)$, which contradicts the fact that $\bar{x}^*$ is a local minimizer. Hence, we must have $\nabla f(\bar{x}^*) = 0$. 

Peter Blomgren, (blomgren.peter@gmail.com)
If $\nabla f(\bar{x}^*) = 0$, then we call $\bar{x}^*$ a **stationary point**.

Recall from linear algebra —

**Definition (Positive Definite Matrix)**

An $n \times n$-matrix $A$ is **Positive Definite** if and only if

$$\forall \bar{x} \neq 0, \quad \bar{x}^t A \bar{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j > 0.$$  

**Definition (Positive Semi-Definite Matrix)**

An $n \times n$-matrix $A$ is **Positive Semi-Definite** if and only if

$$\forall \bar{x} \neq 0, \quad \bar{x}^t A \bar{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \geq 0.$$
Theorem (Second-Order Necessary Conditions)

If \( \bar{x}^* \) is a local minimizer of \( f \) and \( \nabla^2 f \) is continuous in an open neighborhood of \( \bar{x}^* \), then \( \nabla f(\bar{x}^*) = 0 \) and \( \nabla^2 f(\bar{x}^*) \) is positive semi-definite.

Proof.

\( \nabla f(\bar{x}^*) = 0 \) follows from the previous proof. We show that \( \nabla^2 f(\bar{x}^*) \) is positive semi-definite by contradiction:
Optimality: Second-Order Necessary Conditions

Theorem (Second-Order Necessary Conditions)

If $\bar{x}^*$ is a local minimizer of $f$ and $\nabla^2 f$ is continuous in an open neighborhood of $\bar{x}^*$, then $\nabla f(\bar{x}^*) = 0$ and $\nabla^2 f(\bar{x}^*)$ is positive semi-definite.

Proof.

$\nabla f(\bar{x}^*) = 0$ follows from the previous proof. We show that $\nabla^2 f(\bar{x}^*)$ is positive semi-definite by contradiction: Assume that $\nabla^2 f(\bar{x}^*)$ is not positive semi-definite. Then there must exist a vector $\bar{p}$ such that $\bar{p}^t \nabla^2 f(\bar{x}^*) \bar{p} < 0$. 

Peter Blomgren, ⟨blomgren.peter@gmail.com⟩
Optimality: Second-Order Necessary Conditions

Theorem (Second-Order Necessary Conditions)

If $\bar{x}^*$ is a local minimizer of $f$ and $\nabla^2 f$ is continuous in an open neighborhood of $\bar{x}^*$, then $\nabla f(\bar{x}^*) = 0$ and $\nabla^2 f(\bar{x}^*)$ is positive semi-definite.

Proof.

$\nabla f(\bar{x}^*) = 0$ follows from the previous proof. We show that $\nabla^2 f(\bar{x}^*)$ is positive semi-definite by contradiction: Assume that $\nabla^2 f(\bar{x}^*)$ is not positive semi-definite. Then there must exist a vector $\bar{p}$ such that $\bar{p}^t \nabla^2 f(\bar{x}^*) \bar{p} < 0$. By continuity of $\nabla^2 f$ there is a $T > 0$ such that $\bar{p}^t \nabla^2 f(\bar{x}^* + t\bar{p}) \bar{p} < 0 \forall t \in [0, T]$. 

Peter Blomgren, (blomgren.peter@gmail.com)
Theorem (Second-Order Necessary Conditions)

If $\bar{x}^*$ is a local minimizer of $f$ and $\nabla^2 f$ is continuous in an open neighborhood of $\bar{x}^*$, then $\nabla f(\bar{x}^*) = 0$ and $\nabla^2 f(\bar{x}^*)$ is positive semi-definite.

Proof.

$\nabla f(\bar{x}^*) = 0$ follows from the previous proof. We show that $\nabla^2 f(\bar{x}^*)$ is positive semi-definite by contradiction: Assume that $\nabla^2 f(\bar{x}^*)$ is not positive semi-definite. Then there must exist a vector $\bar{p}$ such that $\bar{p}^T \nabla^2 f(\bar{x}^*) \bar{p} < 0$. By continuity of $\nabla^2 f$ there is a $T > 0$ such that $\bar{p}^T \nabla^2 f(\bar{x}^* + t\bar{p}) \bar{p} < 0$ $\forall t \in [0, T]$. Now, the Taylor expansion around $\bar{x}^*$, shows that $\forall s \in (0, T]$ there exists $t \in (0, T)$ such that

$$f(\bar{x}^* + s\bar{p}) = f(\bar{x}^*) + s\bar{p}^T \nabla f(\bar{x}^*) + \frac{1}{2} s^2 \bar{p}^T \nabla^2 f(\bar{x}^* + t\bar{p}) \bar{p}.$$

$$= 0 + \frac{1}{2} s^2 \bar{p}^T \nabla^2 f(\bar{x}^* + t\bar{p}) \bar{p} < 0$$
Theorem (Second-Order Necessary Conditions)

If $\bar{x}^*$ is a local minimizer of $f$ and $\nabla^2 f$ is continuous in an open neighborhood of $\bar{x}^*$, then $\nabla f(\bar{x}^*) = 0$ and $\nabla^2 f(\bar{x}^*)$ is positive semi-definite.

Proof.

$\nabla f(\bar{x}^*) = 0$ follows from the previous proof. We show that $\nabla^2 f(\bar{x}^*)$ is positive semi-definite by contradiction: Assume that $\nabla^2 f(\bar{x}^*)$ is not positive semi-definite. Then there must exist a vector $\bar{p}$ such that $\bar{p}^T \nabla^2 f(\bar{x}^*) \bar{p} < 0$. By continuity of $\nabla^2 f$ there is a $T > 0$ such that $\bar{p}^T \nabla^2 f(\bar{x}^* + t\bar{p}) \bar{p} < 0 \forall t \in [0, T]$. Now, the Taylor expansion around $\bar{x}^*$, shows that $\forall s \in (0, T]$ there exists $t \in (0, T)$ such that

$$f(\bar{x}^* + s\bar{p}) = f(\bar{x}^*) + s\bar{p}^T \nabla f(\bar{x}^*) + \frac{1}{2} s^2 \bar{p}^T \nabla^2 f(\bar{x}^* + t\bar{p}) \bar{p}.$$

Hence $f(\bar{x}^* + s\bar{p}) < f(\bar{x}^*)$, which is a contradiction.
The conditions we have outlined so far are necessary; hence if $\bar{x}^*$ is a minimum, then the conditions must hold.

It is more useful to have a set of sufficient conditions, so that if the conditions are satisfied (at $\bar{x}^*$), then $\bar{x}^*$ is a minimum.

The second order sufficient conditions guarantee that $\bar{x}^*$ is a strict local minimizer of $f$, and the convexity of $f$ guarantees that any local minimizer is a global minimizer...
Suppose that $\nabla^2 f$ is continuous in an open neighborhood of $\bar{x}^*$ and that $\nabla f(\bar{x}^*) = 0$ and $\nabla^2 f(\bar{x}^*)$ is positive definite. Then $\bar{x}^*$ is a strict local minimizer of $f$. 
Optimality: Second-order Sufficient Conditions (Proof)

Proof.

Peter Blomgren, ⟨blomgren.peter@gmail.com⟩
Proof.

Since the Hessian $\nabla^2 f(\bar{x}^*)$ is positive definite, we can find a open ball of positive radius $r$, $D(r; \bar{x}^*) = \{\bar{y} \in \mathbb{R}^n : \|\bar{x}^* - \bar{y}\| < r\}$, so that $\nabla^2 f(\bar{y})$ is positive definite $\forall \bar{y} \in D$. 
Proof.

Since the Hessian $\nabla^2 f(\bar{x}^*)$ is positive definite, we can find an open ball of positive radius $r$, $D(r; \bar{x}^*) = \{ \bar{y} \in \mathbb{R}^n : \|\bar{x}^* - \bar{y}\| < r \}$, so that $\nabla^2 f(\bar{y})$ is positive definite $\forall \bar{y} \in D$. Now, for any vector $\bar{p}$ such that $\|\bar{p}\| < r$, we have $\bar{x}^* + \bar{p} \in D$ and therefore (by Taylor)

$$f(\bar{x}^* + \bar{p}) = f(\bar{x}^*) + \bar{p}^T \nabla f(\bar{x}^*) + \frac{1}{2} \bar{p}^T \nabla^2 f(\bar{x}^* + t\bar{p})\bar{p}$$

for some $t \in (0, 1)$. Hence it follows that $f(\bar{x}^*) < f(\bar{x}^* + \bar{p})$, and so $\bar{x}^*$ must be a strict local minimizer.
Theorem

When the objective function $f$ is convex, any local minimizer $\bar{x}^*$ is also a global minimizer of $f$. If in addition $f$ is differentiable, then any stationary point $\bar{x}^*$ is a global minimizer of $f$. 
Proof (part-1).
Proof (part-1).

Suppose that $\bar{x}^*$ is a local, but not a global minimizer. Then there must exist a point $\bar{z} \in \mathbb{R}^n$ such that $f(\bar{z}) < f(\bar{x}^*)$. 
Proof (part-1).

Suppose that $\bar{x}^*$ is a local, but not a global minimizer. Then there must exist a point $\bar{z} \in \mathbb{R}^n$ such that $f(\bar{z}) < f(\bar{x}^*)$. Consider the line-segment that joins $\bar{x}^*$ and $\bar{z}$:

$$\bar{y}(\lambda) = \lambda \bar{z} + (1 - \lambda) \bar{x}^*, \quad \lambda \in [0, 1]$$
Proof (part-1).

Suppose that $\bar{x}^*$ is a local, but not a global minimizer. Then there must exist a point $\bar{z} \in \mathbb{R}^n$ such that $f(\bar{z}) < f(\bar{x}^*)$. Consider the line-segment that joins $\bar{x}^*$ and $\bar{z}$:

$$\bar{y}(\lambda) = \lambda \bar{z} + (1 - \lambda) \bar{x}^*, \quad \lambda \in [0, 1]$$

Since $f$ is convex we must have [by definition]

$$f(\bar{y}(\lambda)) \leq \lambda f(\bar{z}) + (1 - \lambda)f(\bar{x}^*) < f(\bar{x}^*), \quad \lambda \in (0, 1]$$

Every neighborhood of $\bar{x}^*$ will contain a piece of the line-segment, hence $\bar{x}^*$ cannot be a local minimizer.
Optimality: Convexity

Proof (part-2).
Proof (part-2).

Suppose that $\bar{x}^*$ is a local but not a global minimizer, and let $\bar{z}$ be such that $f(\bar{z}) < f(\bar{x}^*)$. 
Proof (part-2).

Suppose that $\bar{x}^*$ is a local but not a global minimizer, and let $\bar{z}$ be such that $f(\bar{z}) < f(\bar{x}^*)$. Using convexity, and the definition of a directional derivative (NW$^{2nd}$ p-628), we have

$$
\nabla f(\bar{x}^*)^T(\bar{z} - \bar{x}^*) = \frac{d}{d\lambda} f(\bar{x}^* + \lambda(\bar{z} - \bar{x}^*)) \bigg|_{\lambda=0}
$$

$$
= \lim_{\lambda \to 0} \frac{f(\bar{x}^* + \lambda(\bar{z} - \bar{x}^*)) - f(\bar{x}^*)}{\lambda}
$$

$$
\leq \lim_{\lambda \to 0} \frac{\lambda f(\bar{z}) + (1 - \lambda)f(\bar{x}^*) - f(\bar{x}^*)}{\lambda}
$$

$$
= f(\bar{z}) - f(\bar{x}^*) < 0.
$$
Proof (part-2).

Suppose that $\bar{x}^*$ is a local but not a global minimizer, and let $\bar{z}$ be such that $f(\bar{z}) < f(\bar{x}^*)$. Using convexity, and the definition of a directional derivative ($NW^{2nd}$ p-628), we have

$$\nabla f(\bar{x}^*)^T (\bar{z} - \bar{x}^*) = \frac{d}{d\lambda} f(\bar{x}^* + \lambda(\bar{z} - \bar{x}^*)) \bigg|_{\lambda=0}$$

$$= \lim_{\lambda \downarrow 0} \frac{f(\bar{x}^* + \lambda(\bar{z} - \bar{x}^*)) - f(\bar{x}^*)}{\lambda}$$

$$\leq \lim_{\lambda \downarrow 0} \frac{\lambda f(\bar{z}) + (1 - \lambda)f(\bar{x}^*) - f(\bar{x}^*)}{\lambda}$$

$$= f(\bar{z}) - f(\bar{x}^*) < 0.$$

Therefore, $\nabla f(\bar{x}^*) \neq 0$, so $\bar{x}^*$ cannot be a stationary point. This contradicts the supposition that $f$ is a local minimum.
Optimality: Theorems and Algorithms

The theorems we have shown — all of which are based on elementary (vector) calculus — are the backbone of unconstrained optimization algorithms.

Since we usually do not have a global understanding of $f$, the algorithms will seek stationary points, i.e. solve the problem

$$\nabla f(\bar{x}) = 0.$$ 

When $\bar{x} \in \mathbb{R}^n$, this is a system of $n$ (generally) non-linear equations.

**Hence, there is a strong connection between the solution of non-linear equations and unconstrained optimization.**

— We will focus on developing an optimization framework, and in the last few weeks of the semester we will use it to solve non-linear equations.
Algorithms — An Overview

The algorithms we study start with an initial (sub-optimal) guess \( \overline{x}_0 \), and generate a sequence of iterates \( \{\overline{x}_k\}_{k=1}^N \).

The sequence is terminated when either

- [success] We have approximated a solution up to desired accuracy.
- [failure] No more progress can be made.

Different algorithms make different decisions in how to move from \( \overline{x}_k \) to the next iterate \( \overline{x}_{k+1} \).

Many algorithms are monotone, \( i.e. f(\overline{x}_{k+1}) < f(\overline{x}_k), \forall k \geq 0 \), but there exist non-monotone algorithms. Even a non-monotone algorithm is required to eventually decrease — how else can we reach a minimum? Typically \( f(\overline{x}_{k+m}) < f(\overline{x}_k) \) is required for some fixed value \( m > 0 \) and \( \forall k \geq 0 \).
Most optimization algorithms use one of two fundamental strategies for finding the next iterate: —

1. **Line search** based algorithms reduce the $n$-dimensional optimization problem

   \[
   \min_{\bar{x} \in \mathbb{R}^n} f(\bar{x}),
   \]

   with a one-dimensional problem:

   \[
   \min_{\alpha > 0} f(\bar{x}_k + \alpha \bar{p}_k),
   \]

   where $\bar{p}_k$ is a chosen search direction. Clearly, how cleverly we select $\bar{p}_k$ will affect how much progress we can make in each iteration.

   — The intuitive choice gives a slow scheme!
Moving from $\bar{x}_k$ to $\bar{x}_{k+1}$

2. **Trust region** based methods take a completely different approach. — Using information gathered about the objective $f$, i.e. function values, gradients, Hessians, etc. during the iteration, a simpler **model function** is generated.

A good model function $m_k(\bar{x})$ approximates the behavior of $f(\bar{x})$ in a neighborhood of $\bar{x}_k$, e.g. Taylor expansion

$$m_k(\bar{x}_k + \bar{p}) = f(\bar{x}_k) + \bar{p}^T \nabla f(\bar{x}_k) + \frac{1}{2} \bar{p}^T H_k \bar{p},$$

where $H_k$ is the full Hessian $\nabla^2 f(\bar{x}_k)$ (expensive) or a clever approximation thereof.
The model is chosen simple enough that the optimization problem

\[
\min_{p \in N(\bar{x}_k)} m_k(\bar{x}_k + \bar{p}),
\]

can be solved quickly. The neighborhood \( N(\bar{x}_k) \) of \( \bar{x}_k \) specifies the region in which we trust the model.

A simple model can only capture the local behavior of \( f \) — think about how the Taylor expansion approximates a function well close to the expansion point, but not very well further away.

Usually the trust region is a ball in \( \mathbb{R}^n \), i.e.

\[
N(\bar{x}_k) = \{ \bar{p} : \| \bar{p} - \bar{x}_k \| \leq r \},
\]

but elliptical or box-shaped trust regions are sometimes used.
### Line Search vs. Trust Region

<table>
<thead>
<tr>
<th>Step</th>
<th>Line Search</th>
<th>Trust Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Choose a search direction $\bar{p}_k$.</td>
<td>Establish the maximum distance — the size of the trust region.</td>
</tr>
<tr>
<td>2</td>
<td>Identify the distance, e.g. the step length in the search direction.</td>
<td>Find the direction in the trust region.</td>
</tr>
</tbody>
</table>

**Table:** Line search and trust region methods handle the selection of direction and distance in opposite order.

Next time:

— Rate of Convergence.
— Line search methods, detailed discussion.

Peter Blomgren, ⟨blomgren.peter@gmail.com⟩