Module 04
Optimization Problems

Ahmad F. Taha

CE 5999-02 Special Topics — Intro to Optimization

Email: ahmad.taha@vanderbilt.edu

Webpage: http://lab.vanderbilt.edu/taha

VANDERBILT UNIVERSITY

February 15, 2022

In this module

- Optimization problems in standard form
- Optimality conditions
- Quasi-convex optimization problems and the bisection method
- Application: Maximum Likelihood Estimation
- Application: Capacity of parallel Gaussian channels
- Application: Economic Dispatch

Optimization problem in standard form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \quad (\textbf{inequality constraints}) \\ & h_i(x) = 0, \quad i = 1, \ldots, p \quad (\textbf{equality constraints}) \end{aligned}$$

- $x = [x_1, x_2, \ldots, x_n]^\top$ is a vector of **unknowns**, also called **variables**
- $f_0(x) : \mathbb{R}^n \to \mathbb{R}$ is the **objective function**, also called **cost function**
- $f_i(x) : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, m)$ and $h_i(x) : \mathbb{R}^n \to \mathbb{R}$ are the **constraint functions**
- The set of points that satisfy the constraints is called the **feasible set**

$$C = \{x \in \mathbb{R}^n \mid f_i(x) \le 0 \; (i = 1, \ldots, m) \text{ and } h_i(x) = 0 \; (i = 1, \ldots, p)\}$$

- If there are no constraints ($m = p = 0$), the problem is called **unconstrained**

Implicit and explicit constraints

Explicit constraints: $f_i(x) \leq 0$ $(i = 1, \dots, m)$ and $h_i(x) = 0$ $(i = 1, \dots, p)$
Implicit constraints: $x \in D$, where D is the *domain* of the problem

$$D = \mathrm{dom}(f_1) \cap \dots \cap \mathrm{dom}(f_m) \cap \mathrm{dom}(h_1) \cap \dots \cap \mathrm{dom}(h_p)$$

For example, the problem

$$\begin{array}{ll} \text{minimize} & -\log x_1 - \log x_2 \\ \text{subject to} & x_1 + x_2 - 1 \leq 0 \end{array}$$

has implicit constraint

$$x \in D = \{(x_1, x_2) \in \mathbb{R}^2 \,|\, x_1 > 0, x_2 > 0\}$$

Feasibility and optimal value

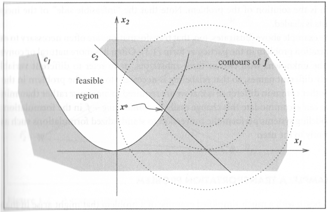
The **optimal value** of the problem is

$$f^* = \inf\{f_0(x) \mid f_i(x) \leq 0 \,(i = 1, \ldots, m) \text{ and } h_i(x) = 0 \,(i = 1, \ldots, p)\}$$

- If one can find at least one x satisfying the constraints (that is, $f_i(x) \leq 0$ for all $i = 1, \ldots, m$, and $h_i(x) = 0$, for all $i = 1, \ldots, p$), the problem is called **feasible**.
- If no such x can be found, then the problem is called **infeasible**.
  In this case, the feasible set is empty and we can write $f^* = \infty$.
- If the problem is feasible, there are two possibilities for f^*.
  1. $f^* = -\infty$. This means that we can stay in the feasible set and move to a direction that $f_0(x)$ becomes smaller without bound.
     We say the problem is **unbounded**.
  2. f^* is a number.

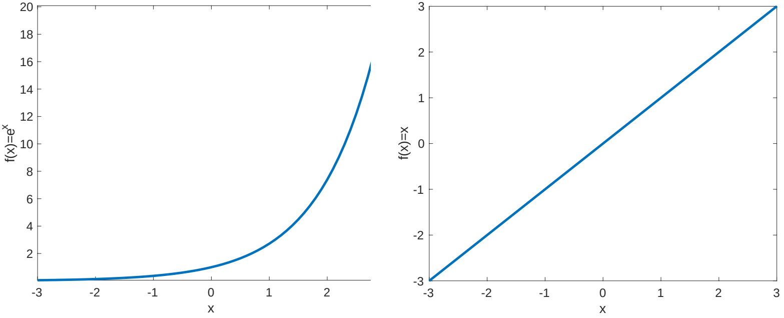
Example

$$\begin{aligned} \text{minimize} \quad & (x_1-2)^2+(x_2-1)^2 \\ \text{subject to} \quad & x_1^2 - x_2 \leq 0 \\ & x_1 + x_2 - 2 \leq 0 \end{aligned}$$



- Contours of $f(x_1,x_2)=(x_1-2)^2+(x_2-1)^2$ plot the points (x_1,x_2) so that $f(x_1,x_2)=\text{constant}$, for different values
- Contours are circles around $(2,1)$, where $f(2,1)=0$
- White area shows the feasible set—the optimal solution must be on a contour of f as close to $(2,1)$ as possible **without** leaving the feasible set
- The optimal point is where the parabola $x_1^2=x_2$ meets the line $x_1+x_2-2=0$
- At the solution: $x_1^2=2-x_1$, and we get $x_1=1$, $x_2=1$ or $x_1=-2$, $x_2=4$
- The point $(x_1^*,x_2^*)=(1,1)$ gives the smallest objective, $f^*=1$

Other examples



- Optimization problem $f^* = \min e^x$
- Optimal value $f^* = 0$
- There is no x such that $f(x) = f^*$

- Optimization problem $f^* = \min x$
- Optimal value $f^* = -\infty$
- The problem is unbounded

Optimal points

$$C = \{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \; (i = 1, \ldots, m) \text{ and } h_i(x) = 0 \; (i = 1, \ldots, p)\}$$
$$f^* = \inf\{f_0(x) \mid f_i(x) \leq 0 \; (i = 1, \ldots, m) \text{ and } h_i(x) = 0 \; (i = 1, \ldots, p)\}$$

- Even if f^* is a number, we don't know if there is an x^* that gives $f_0(x^*) = f^*$
- Consider minimizing $f_0(x) = e^{-x}$; we have $f^* = 0$, but there is no x^*
- If there an $x^* \in C$ that gives $f_0(x^*) = f^*$, we say that x^* is an **optimal point** and that the optimal value is *achieved*
- There may be more than one optimal point
- The set of all optimal points is called the **optimal set** and is

$$X_{\text{opt}} = \{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \; (i = 1, \ldots, m) \text{ and } h_i(x) = 0 \; (i = 1, \ldots, p), f_0(x) = f^*\}$$

Feasibility problems

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

The problem is really to find an x that satisfies the system of inequalities and equalities given by the constraints.

Convex optimization problem in standard form

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& a_i^\top x - b_i = 0, \quad i = 1, \ldots, p
\end{aligned}
$$

- $f_0, f_1, \ldots, f_m$: *convex* functions
- Linear equality constraints
- The feasible set of a convex optimization problem is a convex set
  - Intersection of $m + p$ convex sets: sublevel sets of convex functions $\{x \mid f_i(x) \leq 0\}$ and hyperplanes $a_i^\top x - b_i = 0$
- The equality constraints can be organized in matrix form $Ax = b$ where

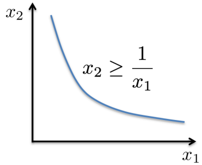
$$
A = \begin{bmatrix} a_1^\top \\ a_2^\top \\ \ldots \\ a_p^\top \end{bmatrix}, \; b = \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ b_p \end{bmatrix}
$$

Example: An optimization problem in standard form

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 \\ \text{subject to} \quad & -x_1 \leq 0 \\ & -x_2 \leq 0 \\ & 1 - x_1 x_2 \leq 0 \end{aligned}$$

- The example is an optimization problem in standard form
- Is it a convex optimization problem in standard form?

- No: We saw that $x_1 x_2$ is not convex and not concave—the same follows for $-x_1 x_2$
- But the the objective is linear, and the feasible set $C = \{(x_1, x_2) \,|\, x_1 > 0, x_2 > 0, x_2 \geq 1/x_1\}$ is convex



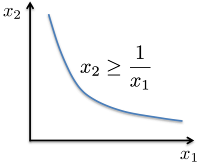
- Can we cast it as a convex optimization problem in standard form?

Example: An optimization problem in standard form

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 \\ \text{subject to} \quad & -x_1 \leq 0 \\ & -x_2 \leq 0 \\ & 1 - x_1 x_2 \leq 0 \end{aligned}$$

- The example is an optimization problem in standard form
- Is it a convex optimization problem in standard form?

- No: We saw that $x_1 x_2$ is not convex and not concave—the same follows for $-x_1 x_2$
- But the the objective is linear, and the feasible set $C = \{(x_1, x_2) \,|\, x_1 > 0, x_2 > 0, x_2 \geq 1/x_1\}$ is convex



- Can we cast it as a convex optimization problem in standard form?

Example: Finding the standard form of a convex optimization problem

Here are two ways:

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 \\ \text{subject to} \quad & -x_1 \leq 0 \\ & -x_2 \leq 0 \\ & 1 - \sqrt{x_1 x_2} \leq 0 \end{aligned}$$

- $1 - x_1x_2 \leq 0 \Leftrightarrow 1 \leq x_1x_2 \Leftrightarrow 1 \leq \sqrt{x_1x_2} \Leftrightarrow 1 - \sqrt{x_1x_2} \leq 0$
- In the above, we used $x_1 > 0$, $x_2 > 0$
- $\sqrt{x_1x_2}$ is concave, $-\sqrt{x_1x_2}$ is convex
- Check the Hessian

$$\begin{aligned} \text{minimize} \quad & x_1 + x_2 \\ \text{subject to} \quad & -x_1 \leq 0 \\ & -x_2 \leq 0 \\ & -\log x_1 - \log x_2 \leq 0 \end{aligned}$$

- $1 - x_1x_2 \leq 0 \Leftrightarrow 1 \leq x_1x_2 \Leftrightarrow \log 1 \leq \log(x_1x_2) \Leftrightarrow -\log x_1 - \log x_2 \leq 0$
- In the above, we used $x_1 > 0$, $x_2 > 0$
- $\log x_1$ and $\log x_2$ are concave functions
- $-\log x_1 - \log x_2$ is a convex function

Equivalent problems

Two optimization problems are equivalent if from a solution of one, a solution of the other can be found, and vice versa.

For example, the following two problems are equivalent:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

$$\begin{array}{ll} \text{minimize} & e^{f_0(x)} \\ \text{subject to} & e^{f_i(x)} - 1 \leq 0, \quad i = 1, \ldots, m \\ & e^{h_i(x)} - 1 = 0, \quad i = 1, \ldots, p \end{array}$$

The essential reason why the problems are equivalent is that the function e^z is strictly increasing [we have that $f_i(x) \leq 0 \Leftrightarrow e^{f_i(x)} \leq e^0 \Leftrightarrow e^{f_i(x)} - 1 \leq 0$].

Maximization problems

$$
\begin{aligned}
\text{maximize} \quad & g_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p
\end{aligned}
$$

The problem is equivalent to minimizing $-g_0(x)$ over the feasible set.
The problem will be convex if

- $g_0(x)$ is a concave function
- $f_1, \ldots, f_m$ are convex functions
- $h_i(x) = a_i^\top x - b_i$

Likewise, a constraint $g(x) \geq 0$ can be equivalently written as $-g(x) \leq 0$.

Epigraph problem form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & f_0(x) - t \leq 0 \\ & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{array}$$

- Problem on the right: epigraph problem form with variables $(x, t) \in \mathbb{R}^{n+1}$
- $m + 1$ inequality constraints; objective is linear
- At the optimal point, we must have $f_0(x^*) = t^*$
- The two problems are **equivalent**: The left problem minimizes $f_0(x)$, while the right problem minimizes t, with the property that at the end (upon solution), $f_0(x) = t$
- If original problem is convex, then so is epigraph form
- Specifically, if $f_0(x)$ is convex, then $g(x, t) = f_0(x) - t$ is convex (convex plus linear)

Example

- Consider the following optimization problem:

$$\min ||x||_1 \quad \text{subject to} \quad Ax = b$$

- How do you pose this as a convex optimization problem using the epigraph problem form?
- Well, first recall that $||x||_1 = \sum_{i=1} |x_i|$
- Hence setting $|x_i| \leq t_i$, one could write

$$\min \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} t = \mathbf{1}^\top t \quad \text{subject to} \quad Ax = b, \ |x_i| \leq t_i \ \forall i = 1, \ldots, n$$

which is equivalent to

$$\min \mathbf{1}^\top t \quad \text{subject to} \quad Ax = b, \ x_i \leq t_i, \ x_i \geq -t_i \ \forall i = 1, \ldots, n$$

which is a linear program (LP)

- Note that this reformulated problem now has twice as many variables

Eliminating linear equality constraints

- Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax - b = 0 \end{array}$$

- Suppose A is $m \times n$ with $\operatorname{rank}(A) = m$ (more columns than rows)
- Remember that when A has linearly indep. columns the pseudo inverse is

$$A^\dagger = \left(A^\top A\right)^{-1} A^\top \quad \text{this is the left inverse}$$

- when A has linearly indep. rows the pseudo inverse is

$$A^\dagger = A^\top \left(AA^\top\right)^{-1} \quad \text{this is the right inverse because } AA^\top = I$$

- We can find $B \in \mathbb{R}^{n \times (n-m)}$ so that $\operatorname{range}(B) = \operatorname{nullspace}(A)$.
- Change of variables $x = A^\top (AA^\top)^{-1} b + Bz$
- Equivalent problem with variable $z \in \mathbb{R}^{n-m}$

$$\begin{array}{ll} \text{minimize} & f_0(A^\top (AA^\top)^{-1} b + Bz) \\ \text{subject to} & f_i(A^\top (AA^\top)^{-1} b + Bz) \leq 0, \quad i = 1, \ldots, m \end{array}$$

Locally and globally optimal points

Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in C \end{array}$$

A point $x^* \in C$ is locally optimal if there is an $R > 0$ such that

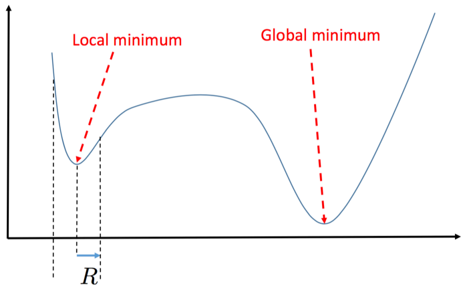
$$y \in C, \ \|y - x^*\| \leq R \Longrightarrow f_0(x^*) \leq f_0(y)$$

A point $x^* \in C$ is (globally) optimal if

$$y \in C \Longrightarrow f_0(x^*) \leq f_0(y)$$

"Locally optimal" means that x^* minimizes $f_0(x)$ only over nearby points, not over the entire C.

Locally and globally optimal points: Example



- Function f with one local minimum and one global minimum
- The figure illustrates the concept of ball with radius R for local minimum
- We can see from the figure that an even larger R is valid for this local minimum

Optimality properties of convex problems

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in C \end{array}$$

Suppose $f_0(x)$ is convex and C is a convex set. Then the following hold:

1. Any locally optimal point is also globally optimal.
2. If in addition $f_0(x)$ is strictly convex, there can be at most one global minimum.
3. The optimal set X_{opt} is a convex set.

Optimality criterion for unconstrained optimization

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathbb{R}^n \end{array}$$

Suppose f_0 is differentiable.

- If x^* is a local minimum of f_0, then

$$\nabla f_0(x^*) = 0$$

- If in addition f_0 is convex, then the previous condition is sufficient for optimality

This is a system of n equations with n unknowns. When are these equations linear?

Necessary and sufficient condition: Word of caution

- First-order necessary optimality condition **(when C is a convex set)**:

$$\nabla f_0(x^*)^\top (x - x^*) \geq 0 \text{ for all } x \in C$$

- ... means that local minima are guaranteed to satisfy the condition
- But the converse is not true in general: We may have points that satisfy the first-order optimality condition, but are not local minima
- Example: $f(x) = x^3$, $C = \mathbb{R}$, $x = 0$
- **Convexity in the objective function** guarantees the sufficiency of the condition
- All points $x^* \in C$ satisfying $\nabla f_0(x^*)^\top (x - x^*) \geq 0$ for all $x \in C$ will be (global) minimizers of $f_0(x)$ over C, when f_0 is a convex function and C a convex set

Quasiconvex optimization problems

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax - b = 0 \end{aligned}$$

- f_0: *quasi-convex* function
- $f_1, \ldots, f_m$: *convex* functions
- Linear equality constraints
- The feasible set of a quasiconvex optimization problem is a convex set

Example: Linear fractional programming

$$\begin{aligned} \text{minimize} \quad & (c^\top x + d)/(f^\top x + e) \\ \text{subject to} \quad & Ax - b = 0, Gx \preceq h, f^\top x + e > 0 \end{aligned}$$

Representation via family of convex functions

Sublevel set of quasiconvex function $f_0(x) \leq t$ can be expressed as sublevel set of a convex function $\phi_t(x)$

$$f_0(x) \leq t \Leftrightarrow \phi_t(x) \leq 0$$

where $\phi_t(x) : \mathbb{R}^n \to \mathbb{R}$ satisfies two properties

1. $\phi_t(x)$ is convex in x for any t
2. $\phi_t(x)$ is nonincreasing in t for any x

Example: For the linear-fractional objective $f_0(x) = (a^\top x + b)/(c^\top x + d)$, choose

$$\phi_t(x) = (a^\top x + b) - t(c^\top x + d) = (a - tc)^\top x + b - td$$

Solving quasiconvex problems via convex feasibility

Consider the quasiconvex problem with optimal value f^*

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0,\ i = 1, \ldots, m, \quad Ax - b = 0 \end{array}$$

and the convex feasibility problem

$$\phi_t(x) \leq 0, \quad f_i(x) \leq 0,\ i = 1, \ldots, m, \quad Ax - b = 0$$

If the previous problem is feasible, then $f^* \leq t$. If not, then $f^* > t$.
We can solve the quasiconvex problem by solving a series of convex feasibility problems.

Bisection method

Given an interval $[l, u]$ so that $l \leq f^* \leq u$, and a tolerance $\epsilon > 0$, repeat the following steps until $u - l \leq \epsilon$

1. $t := (l + u)/2$
2. Solve the convex feasibility problem

$$\phi_t(x) \leq 0, \quad f_i(x) \leq 0, \; i = 1, \ldots, m, \quad Ax - b = 0$$

3. If the problem is feasible, set $u := t$, $x :=$ any solution of the feasibility problem; else set $l := t$

The algorithm divides the interval $[l, u]$ in two at every iteration. It requires $\log_2[(u - l)/\epsilon]$ iterations to converge.

Abstracting constraints

- Sometimes we may write an optimization problem as

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \\ & x \in X \end{array}$$

- We call X an *abstract constraint set*, because we don't describe it explicitly with inequalities and equalities
- It is sometimes useful to keep certain constraints in the form $x \in X$
- An example is when it is numerically efficient to project onto X (e.g., X is a box)
  - A large class of algorithms are constructed relying on projection onto the feasible set

Restriction and relaxation

Original problem with optimal value f^*

$$f^* = \text{minimize} \quad f_0(x)$$
$$\text{subject to} \quad x \in C$$

New problem, with optimal value $\tilde{f}^*$

$$\tilde{f}^* = \text{minimize} \quad f_0(x)$$
$$\text{subject to} \quad x \in \tilde{C}$$

New problem is

- *Relaxation* of original if $\tilde{C} \supset C$, in which case, $\tilde{f}^* \leq f^*$
- *Restriction* of original if $\tilde{C} \subset C$, in which case, $\tilde{f}^* \geq f^*$

Example: If $f_0(x)$ is convex, C is nonconvex, and $\tilde{C} = \text{conv}(C)$, then the relaxation is a convex problem, and gives a lower bound for the original one.

Application: Maximum Likelihood Estimation (MLE)

- Suppose we have m independent and identically distributed (i.i.d.) measurements $y_1, \ldots, y_m$ of a random variable Y with pdf $p(y;\theta)$
- Remember that $\Pr[a \leq Y \leq b] = \int_a^b p(y;\theta)\, dy$ (this probability is the cumulative distribution function)
- PDFs of random variables model the tiny probability of Y falling within the infinitesimal interval $[y, y+dy]$
- Parameter $\theta \in \mathbb{R}^n$ is unknown—we want to estimate it using data
- Example: Measuring a constant signal $\theta \in \mathbb{R}$ in noise.

$$y_i = \theta + v_i, \quad i = 1, \ldots, m$$

  - Engineering problem: Given measurements $y_1, \ldots, y_m$, estimate θ
  - The random quantities $v_1, \ldots, v_m$ are i.i.d. noise samples
  - If the noise pdf is $f(v)$, then y_i is a sample of the pdf $p(y;\theta) = f(y-\theta)$
  - For instance, if the noise is Gaussian $\mathcal{N}(0, \sigma^2)$, then we have

$$p(y;\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\theta)^2}{2\sigma^2}}$$

- The notation $(y;\theta)$ emphasizes that y is a measurement (known), and θ is a parameter (unknown)

Likelihood Function and the MLE

- The pdf of the measurements viewed as a function of θ is called the *likelihood function*
- The measurements are i.i.d., so the likelihood function is $\prod_{i=1}^{m} p(y_i;\theta)$
- The maximum likelihood (ML) estimate of θ is the value of θ that maximizes the likelihood function
- So, the ML estimate $\hat{\theta}_{\mathrm{ML}}$ is the solution to the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \prod_{i=1}^{m} p(y_i;\theta) \\ \text{subject to} \quad & \theta \in \mathbb{R}^n \end{aligned}$$

- Sometimes there are known constraints on θ, so $\theta \in \mathbb{R}^n$ is replaced by $\theta \in C$

Convexity of the MLE

- Finding the MLE is equivalent to maximizing the *log-likelihood* function, defined as $\log \prod_{i=1}^{m} p(y_i; \theta)$. (Why are they equivalent?)

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{i=1}^{m} \log p(y_i; \theta) \\ \text{subject to} & \theta \in C \end{array}$$

When is this a convex optimization problem?

1. The log-likelihood function $\log p(y; \theta)$ is concave in θ
   - In this case, $p(y; \theta)$ is called *log-concave* in θ
   - Note that y takes specific values (measurements $y_1, \ldots, y_m$), it is not a variable
2. The set C is convex

Convexity of the MLE

- Finding the MLE is equivalent to maximizing the *log-likelihood* function, defined as $\log \prod_{i=1}^{m} p(y_i;\theta)$. (Why are they equivalent?)

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{m} \log p(y_i;\theta) \\ \text{subject to} \quad & \theta \in C \end{aligned}$$

When is this a convex optimization problem?

1. The log-likelihood function $\log p(y;\theta)$ is concave in θ
   - In this case, $p(y;\theta)$ is called *log-concave* in θ
   - Note that y takes specific values (measurements $y_1, \ldots, y_m$), it is not a variable
2. The set C is convex

Finding the MLE

- Suppose $C = \mathbb{R}^n$ and $\log p(y_i;\theta)$ is differentiable with respect to θ (usual case)
- The first order optimality condition yields a system of equations for the optimal θ

$$\nabla_\theta \sum_{i=1}^{m} \log p(y_i;\theta) = 0$$

- One approach to finding the MLE is to numerically solve the previous system using a method like Newton-Raphson (we will learn this method later in class)
- Another approach is the *expectation-maximization* algorithm—this algorithm yields a local maximum under certain conditions
- When the problem is convex, the previous system is sufficient for optimality, and any local maximum will be global maximum

Back to the example of a constant signal in noise: Log-concave noise pdfs

$$y_i = \theta + v_i, \quad i = 1, \ldots, m$$

Finding the MLE is a convex optimization problem if the noise pdf $f(v)$ is

1. Gaussian
$$f(v) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{v^2}{2\sigma^2}}$$
2. Laplacian $(a > 0)$
$$f(v) = \frac{1}{2a} e^{-\frac{|v|}{a}}$$
3. Uniform on $[-a, a]$
$$f(v) = \frac{1}{2a}$$

Gaussian noise: Recovering the sample mean

$$y_i = \theta + v_i, \quad i = 1, \ldots, m$$

Suppose the noise is Gaussian. The log-likelihood becomes

$$\sum_{i=1}^{m} \log p(y_i; \theta) = -m \log \left(\sigma \sqrt{2\pi}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (y_i - \theta)^2$$

The ML estimate $\hat{\theta}_{\mathrm{ML}}$ is the solution of the following (unconstrained) least-squares problem:

$$\text{minimize} \quad \sum_{i=1}^{m} (y_i - \theta)^2$$

The solution is the *sample mean* of the measurements:

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{m} \sum_{i=1}^{m} y_i$$

A more general example

- Consider that $y_i = a_i^\top x + v_i,\ \forall i = 1, \ldots, m$
- Problem: estimating parameter $x \in \mathbb{R}^n$ from available measurements $y_i \in \mathbb{R}$
- Again, the problem can be written as

$$\text{maximize} \quad \sum_{i=1}^{m} \log p(y_i; x) = \sum_{i=1}^{m} \log p(y_i - a_i^\top x)$$

- As before, we do not know distributions of y or x but we know that of noise v_i
- PDF of v:

$$f(v) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2}$$

- How do you proceed? First, consider zero mean noise (fair assumption)
- Then what?
- You should end up with:

$$\hat{x}_{\text{ML}} = \text{argmin}_x ||Ax - y||_2^2$$

  where A is a matrix of rows $a_i^\top$
- Prove that result
- What if the noise isn't Gaussian? Check Example 7.1 in the book

Laplacian noise

$$y_i = \theta + v_i, \quad i = 1, \ldots, m$$

- Suppose the noise is Laplacian. The log-likelihood becomes

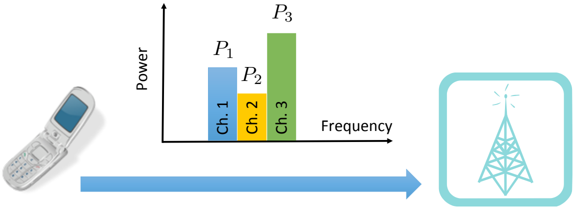
$$\sum_{i=1}^{m} \log p(y_i; \theta) = -m \log(2a) - \frac{1}{a} \sum_{i=1}^{m} |y_i - \theta|$$

- The ML estimate $\hat{\theta}_{\mathrm{ML}}$ is the solution of the following (unconstrained) least absolute deviations problem:

$$\text{minimize} \quad \sum_{i=1}^{m} |y_i - \theta|$$

- Why is this a convex optimization problem?
- The objective is not differentiable, which means that we cannot use the first order optimality condition to write a condition for the unknown θ
- We will see that this problem can be converted to a linear program

Application: Capacity of parallel Gaussian channels



- Transmitter communicates with a receiver
- Available bandwidth is divided into n channels
- Transmission power at each channel is P_i, $i = 1, \ldots, n$
- Noise power at each channel is N_i, $i = 1, \ldots, n$ (known)
- Transmission rate (capacity) of each channel is $C_i = \dfrac{1}{\log 2} \log(1 + P_i/N_i)$ (bits/sec) (expression assumes the noise is Gaussian)

Maximizing the transmission rate: Power allocation

The transmitter has a power budget $P_{\max}$ that can be split among the channels. How should be power be allocated?
To maximize the total communication rate, solve the following optimization problem with variables $P_1, \ldots, P_n$:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} \log(1 + P_i/N_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} P_i = P_{\max} \\ & P_i \geq 0, \quad i = 1, \ldots, n \end{aligned}$$

The feasible set is a polyhedron and the objective is concave in $P_1, \ldots, P_n$. This is a convex optimization problem.
The traditional solution method is the *waterfilling algorithm* (later in class).

Maximizing the transmission rate: Power allocation

The transmitter has a power budget $P_{\max}$ that can be split among the channels. How should be power be allocated?
To maximize the total communication rate, solve the following optimization problem with variables $P_1, \ldots, P_n$:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} \log(1 + P_i/N_i) \\ \text{subject to} \quad & \sum_{i=1}^{n} P_i = P_{\max} \\ & P_i \geq 0, \quad i = 1, \ldots, n \end{aligned}$$

The feasible set is a polyhedron and the objective is concave in $P_1, \ldots, P_n$. This is a convex optimization problem.
The traditional solution method is the *waterfilling algorithm* (later in class).

Questions And Suggestions?

Thank You!

Please visit
https://lab.vanderbilt.edu/taha/
IFF you want to know more ☺