ORF522 – Linear and Nonlinear Optimization

12. Introduction to nonlinear optimization
Homogeneous self-dual embedding
Optimality conditions

### Primal

- **minimize**: $c^T x$
- **subject to**: $Ax + s = b$
  - $s \geq 0$

### Dual

- **maximize**: $-b^T y$
- **subject to**: $A^T y + c = 0$
  - $y \geq 0$

**Optimality conditions**

$$
\begin{bmatrix}
0 \\
s \\
0
\end{bmatrix} =
\begin{bmatrix}
0 & A^T \\
-A & 0 \\
c^T & b^T
\end{bmatrix}
\begin{bmatrix}
x \\
y
\end{bmatrix} +
\begin{bmatrix}
c \\
b \\
0
\end{bmatrix}
$$

$s, y \geq 0$

Any $(x^*, s^*, y^*)$ satisfying these conditions is **optimal**

What happens if the problem is infeasible?
How do you detect infeasibility/unboundedness?

<table>
<thead>
<tr>
<th>Primal</th>
<th>Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>minimize $c^T x$</td>
<td>maximize $-b^T y$</td>
</tr>
<tr>
<td>subject to $Ax + s = b$</td>
<td>subject to $A^T y + c = 0$</td>
</tr>
<tr>
<td>$s \geq 0$</td>
<td>$y \geq 0$</td>
</tr>
</tbody>
</table>

Alternatives (Farkas lemma) Write feasibility problem and dualize…

- **primal feasible:** $Ax + s = b$, $s \geq 0$
- **primal infeasible:** $A^T y = 0$, $b^T y < 0$, $y \geq 0$ (primal infeasibility certificate)
- **dual feasible:** $A^T y + c = 0$, $y \geq 0$
- **dual infeasible:** $Ax \leq 0$, $c^T x < 0$ (dual infeasibility certificate)
The homogeneous self-dual embedding

Derivation

Introduce two new variables $\kappa, \tau \geq 0$

Homogeneous self-dual embedding

\[
\begin{bmatrix}
  0 \\
  s \\
  \kappa
\end{bmatrix} =
\begin{bmatrix}
  0 & A^T & c \\
  -A & 0 & b \\
  -c^T & -b^T & 0
\end{bmatrix}
\begin{bmatrix}
  x \\
  y \\
  \tau
\end{bmatrix}
\]

$s, y, \kappa, \tau \geq 0$

\[
Q =
\begin{bmatrix}
  0 & A^T & c \\
  -A & 0 & b \\
  -c^T & -b^T & 0
\end{bmatrix}
\]

\[
Qu = v
\]

\[
u, v \geq 0
\]

$u = (x, y, \tau)$

$v = (0, s, \kappa)$
The homogeneous self-dual embedding

Properties

\[ Qu = v \]
\[ u, v \geq 0 \]

\[
Q = \begin{bmatrix}
0 & A^T & c \\
-A & 0 & b \\
-c^T & -b^T & 0
\end{bmatrix}
\]

\[ u = (x, y, \tau) \]
\[ v = (0, s, \kappa) \]

Matrix

- \( Q \) is skew-symmetric: \( Q^T = -Q \) \( \Rightarrow \) \( u^T Qu = 0 \)
- \( u \perp v \) \hspace{1em} \text{proof} \hspace{1em} Qu - v = 0 \hspace{1em} \Rightarrow \hspace{1em} u^T Qu - u^T v = 0 \hspace{1em} \Rightarrow \hspace{1em} u^T v = 0 \]

Homogeneous

\((u, v)\) satisfy \( Qu = v, (v, u) \geq 0 \) \( \Rightarrow \hspace{1em} \alpha(u, v) \) with \( \alpha \geq 0 \) feasible

Always feasible

\( \alpha = 0 \) \( \Rightarrow \hspace{1em} (0, 0) \) is feasible
The homogeneous self-dual embedding

Outcomes

Find $x, s, y, \kappa, \tau$ such that

$$
\begin{bmatrix}
0 \\
-\kappa \\
-\tau
\end{bmatrix} =
\begin{bmatrix}
0 & A^T & c \\
-A & 0 & b \\
-c^T & -b^T & 0
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
\tau
\end{bmatrix}
$$

$s, y, \kappa, \tau \geq 0$

**Note.** By strict complementarity, we can ensure $\kappa + \tau > 0$

**Case 1: feasibility**

$\tau > 0, \kappa = 0$ define $(\hat{x}, \hat{s}, \hat{y}) = (x^*/\tau, s^*/\tau, y^*/\tau)$

$$
0 = A^T \hat{y} + c \\
\hat{s} = -A\hat{x} + b
$$

$\hat{s} \geq 0, \quad \hat{y} \geq 0, \quad \hat{s}^T \hat{y} = 0$

$\longrightarrow$ $(\hat{x}, \hat{s}, \hat{y})$ is a **solution** to the original problem
The homogeneous self-dual embedding

Outcomes

Find $x, s, y, \kappa, \tau$ such that

$$
\begin{bmatrix}
0 \\
s \\
\kappa
\end{bmatrix} =
\begin{bmatrix}
A^T & c \\
-A & 0 & b \\
-c^T & -b^T & 0
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
\tau
\end{bmatrix}
$$

$s, y, \kappa, \tau \geq 0$

Case 2: infeasibility

$\tau = 0, \kappa > 0 \quad \rightarrow \quad c^T x + b^T y < 0$ (impossible). Must have infeasibility

If $b^T y < 0$ then $\hat{y} = y / (-b^T y)$ is a certificate of primal infeasibility

$$
A^T \hat{y} = 0, \quad b^T \hat{y} = -1 < 0, \quad \hat{y} \geq 0
$$

If $c^T x < 0$ then $\hat{x} = x / (-c^T x)$ is a certificate of dual infeasibility

$$
A\hat{x} \leq 0, \quad c^T \hat{x} = -1 < 0
$$
Interior-point method for homogeneous self-dual embedding

Linear complementarity problem

\[ Qu = v \]
\[ u^T v = 0 \]
\[ u, v \geq 0 \]

Equations

\[ h(u, v) = \begin{bmatrix} Qu - v \\ UV1 \end{bmatrix} = 0 \]
\[ u, v \geq 0 \]

Directions

\[
\begin{bmatrix} Q & -I \\ V & U \end{bmatrix}
\begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} =
\begin{bmatrix} -r_e \\ -UV1 + \sigma \mu 1 \end{bmatrix}
\]
\[ r_e = Qu - v \]
\[ \mu = (u^T v) / d \]

Line search to enforce \( u, v > 0 \)

\[ (u, v) \leftarrow (u, v) + \alpha (\Delta u, \Delta v) \]

Interior-point methods can solve linear complementarity problems
Today’s lecture
[Chapter 2-4 and 6, CO] [Chapter A and B, FCA]

- Nonlinear optimization
- Examples
- Convex analysis review
- Convex optimization
What if the problem is no longer linear?
Nonlinear optimization

minimize \( f(x) \)
subject to \( g_i(x) \leq 0, \quad i = 1, \ldots, m \)

\( x = (x_1, \ldots, x_n) \) Variables
\( f : \mathbb{R}^n \rightarrow \mathbb{R} \) Nonlinear objective function
\( g_i : \mathbb{R}^n \rightarrow \mathbb{R} \) Nonlinear constraints functions

Feasible set
\[
C = \{ x \mid g_i(x) \leq 0, \quad i = 1, \ldots, m \}
\]
Small example

minimize \[ 0.5x_1^2 + 0.25x_2^2 \]
subject to
\[ e^{x_1} - 2 - x_2 \leq 0 \]
\[ (x_1 - 1)^2 + x_2 - 3 \leq 0 \]
\[ x_1 \geq 0 \]
\[ x_2 \geq 1 \]

Contour plot has curves
(no longer lines)

Feasible set is
no longer a polyhedron
Integer optimization
It’s still nonlinear optimization

minimize \( f(x) \)
subject to \( x \in \mathbb{Z} \)

minimize \( f(x) \)
subject to \( \sin(\pi x) = 0 \)
We cannot solve most nonlinear optimization problems
Examples of (solvable) nonlinear optimization
Regression

Fit affine function $f(z) = \alpha + \beta z$ to $m$ points $(z_i, y_i)$

Approximation problem $Ax \approx b$ where

$$A = \begin{bmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_m \end{bmatrix}, \quad x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad b = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

Goal

minimize $\|Ax - b\|$ 

1-norm or $\infty$-norm $\implies$ linear optimization

2-norm $\implies$ least-squares

$$\|Ax - b\|_2^2 = \sum_i (f(z_i) - y_i)^2$$
Sparse regression

Regressor selection

\[
\begin{align*}
\text{minimize} & \quad \|Ax - b\|_2^2 \\
\text{subject to} & \quad \text{card}(x) \leq k \\
\end{align*}
\]

(very hard)

Add regularization to the objective

Regularized regression (ridge)

\[
\text{minimize} \quad \|Ax - b\|_2^2 + \gamma\|x\|_2^2
\]

Regularized regression (lasso)

\[
\text{minimize} \quad \|Ax - b\|_2^2 + \gamma\|x\|_1
\]

Sparse $x$ is more robust and interpretable
Lasso vs ridge regression

Regularized regression (ridge)

\[
\text{minimize } \|Ax - b\|_2^2 + \gamma \|x\|_2^2
\]

Regularized regression (lasso)

\[
\text{minimize } \|Ax - b\|_2^2 + \gamma \|x\|_1
\]

Regularization paths
Portfolio optimization

We have a total of $n$ assets

$x_i$ is fraction of money invested in asset $i$

$p_i$ is the relative price change of asset $i$

Returns

$p^T x$

$p$ random variable: mean $\mu$, covariance $\Sigma$

Portfolio optimization

Expected return

maximize $\mu^T x - \gamma x^T \Sigma x$

subject to $1^T x = 1$

$x \geq 0$

Risk

Risk-aversion parameter
Convex analysis review
Extended real-value functions

\[ f(x) \text{ on } \text{dom} f \]

Extended-value extension

\[ \tilde{f}(x) = \begin{cases} 
  f(x) & x \in \text{dom} f \\
  \infty & x \notin \text{dom} f 
\end{cases} \]

Always possible to evaluate functions

\[ \text{dom} \tilde{f} = \{ x \mid \tilde{f}(x) < \infty \} \]
Indicator functions

Indicator function

\[ I_C(x) = \begin{cases} 
0 & x \in C \\
\infty & x \notin C 
\end{cases} \]

Constrained form
minimize \( f(x) \)
subject to \( x \in C \)

Unconstrained form
minimize \( f(x) + I_C(x) \)
Convex set

**Definition**

For any $x, y \in C$ and any $\alpha \in [0, 1]$,

$$\alpha x + (1 - \alpha)y \in C$$

**Examples**

- $\mathbb{R}^n$
- Hyperplanes
- Hyperspheres
- Polyhedra

**Examples**

- Cardinality constraint $\text{card}(x) \leq k$
- $\mathbb{Z}^n$
- Any disjoint set
Convex combinations

Convex combination

\[ \alpha_1 x_1 + \cdots + \alpha_k x_k \] for any \( x_1, \ldots, x_k \) and \( \alpha_1, \ldots, \alpha_k \) such that \( \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i = 1 \)

Convex hull

\[ \text{conv } C = \left\{ \sum_{i=1}^{k} \alpha_i x_i \mid x_i \in C, \quad \alpha_i \geq 0, \quad i = 1, \ldots, k, \quad 1^T \alpha = 1 \right\} \]
Cones

Cone
\[ x \in C \implies tx \in C \quad \text{for all} \quad t \geq 0 \]

Convex cone
\[ x_1, x_2 \in C \implies t_1 x_1 + t_2 x_2 \in C \quad \text{for all} \quad t_1, t_2 \geq 0 \]
Conic combinations

Conic combination
\[ \alpha_1 x_1 + \cdots + \alpha_k x_k \] for any \( x_1, \ldots, x_k \) and \( \alpha_1, \ldots, \alpha_k \) such that \( \alpha_i \geq 0 \)

Conic hull
\[ \left\{ \sum_{i=1}^{k} \alpha_i x_i \mid x_i \in C, \quad \alpha_i \geq 0, \quad i = 1, \ldots, k \right\} \]
Cones

Examples

Nonnegative orthant
\[ \mathbb{R}_+^n = \{ x \in \mathbb{R}^n \mid x \geq 0 \} \]

Norm-cone
\[ \{ (x, t) \mid \|x\| \leq t \} \] (if 2-norm, second-order cone)

Positive semidefinite cone
\[ \mathcal{S}_+^n = \{ X \in \mathcal{S}^n \mid z^T X z \geq 0, \text{ for all } z \in \mathbb{R}^n \} \]
Normal cone

For any set $C$ and point $x \in C$, we define

$$\mathcal{N}_C(x) = \left\{ g \mid g^T(y - x) \leq 0, \text{ for all } y \in C \right\}$$

$\mathcal{N}_C(x)$ is always convex

What if $x \in \text{int} S$?
Gradient

Derivative
If \( f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m \) continuously differentiable, we define

\[
Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n
\]

Gradient
If \( f : \mathbb{R}^n \rightarrow \mathbb{R} \), we define

\[
\nabla f(x) = Df(x)^T
\]

Example
\[
f(x) = (1/2)x^T P x + q^T x
\]
\[
\nabla f(x) = Px + q
\]

First-order approximation
\[
f(y) \approx f(x) + \nabla f(x)^T (y - x)
\]
(affine function of \( y \))
Hessian

Hessian matrix (second derivative)

If \( f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) second-order differentiable, we define

\[
\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \quad i, j = 1, \ldots, n
\]

**Example**

\[
f(x) = (1/2)x^T Px + q^T x
\]

\[
\nabla^2 f(x) = P
\]

**Second-order approximation**

\[
f(y) \approx f(x) + \nabla f(x)^T (y - x) + (1/2)(y - x)^T \nabla^2 f(x)(y - x)
\]

(quadratic function of \( y \))
Convex optimization
Convex functions

Convex function
For every $x, y \in \mathbb{R}^n$, $\alpha \in [0, 1]$  \[ f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) \]

Concave function
$f$ is concave if and only if $-f$ is convex
Convex conditions

First-order
Let $f$ be a continuous differentiable function, then it is convex if and only if $\text{dom } f$ is convex and

$$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

for all $x, y \in \text{dom } f$

Second-order
If $f$ is twice differentiable, then $f$ is convex if and only if $\text{dom } f$ is convex and

$$\nabla^2 f(x) \succeq 0$$

for all $x \in \text{dom } f$
Verifying convexity

Basic definition (inequality)

First and second order conditions (gradient, hessian) → Hard!

Convex calculus (directly construct convex functions)

- Library of basic functions that are convex/concave
- Calculus rules or transformations that preserve convexity → Easy!
Disciplined Convex Programming
Convexity by construction

General composition rule

\[ h(f_1(x), f_2(x), \ldots, f_k(x)) \] is convex when \( h \) is convex and for each \( i \)

- \( h \) is nondecreasing in argument \( i \) and \( f_i \) is convex, or
- \( h \) is nonincreasing in argument \( i \) and \( f_i \) is concave, or
- \( f_i \) is affine

Check your functions at [https://dcp.stanford.edu/](https://dcp.stanford.edu/)
More details and examples in ORF523
Convex optimization problems

minimize \( f(x) \)
subject to \( g_i(x) \leq 0, \quad i = 1, \ldots, m \)

\( f : \mathbb{R}^n \rightarrow \mathbb{R} \) Convex objective function
\( g_i : \mathbb{R}^n \rightarrow \mathbb{R} \) Convex constraints functions

Convex feasible set
\[ C = \{ x \mid g_i(x) \leq 0, \quad i = 1, \ldots, m \} \]
Modelling software for convex optimization

Modelling tools simplify the formulation of convex optimization problems

- **Construct problems** using library of basic functions
- **Verify convexity** by general composition rule
- Express the problem in input format required by a specific solver

Examples

- CVX, YALMIP (Matlab)
- CVXPY (Python)
- Convex.jl (Julia)
Solving convex optimization problems

CVXPY

minimize $\|Ax - b\|_2$
subject to $0 \leq x \leq 1$

```python
x = cp.Variable(n)
objective = cp.Minimize(cp.norm(A*x - b))
constraints = [0 <= x, x <= 1]
problem = cp.Problem(objective, constraints)

# The optimal objective value is returned by `problem.solve()`.
result = problem.solve()

# The optimal value for x is stored in `x.value`
print(x.value)
```
Local vs global minima (optimizers)

\[
\text{minimize } f(x) \\
\text{subject to } x \in C
\]

Local optimizer \( x \)
\[
f(y) \geq f(x), \quad \forall y \text{ such that } \|x - y\|_2 \leq R
\]

Global optimizer \( x \)
\[
f(y) \geq f(x), \quad \forall y \in C
\]
Optimality and convexity

Theorem

For a convex optimization problem, any local minimum is a global minimum.

Local optimizer $x$

$$f(y) \geq f(x), \quad \forall y$$

such that $\|x - y\|_2 \leq R$

Global optimizer $x$

$$f(y) \geq f(x), \quad \forall y \in C$$
Optimality and convexity

Proof (contradiction)

Suppose that $f$ is convex and $x$ is a local (not global) minimum for $f$, i.e.,

$$f(y) \geq f(x), \quad \forall y \text{ such that } \|x - y\|_2 \leq R.$$ 

Therefore, there exists a feasible $z$ such that $\|z - x\| > R$ and $f(z) < f(x)$.

Consider $y = (1 - \alpha)x + \alpha z$ with $\alpha = \frac{R}{2\|z - x\|_2}$.

Then, $\|y - x\|_2 = \alpha\|z - x\|_2 = \frac{R}{2} < R$, and by convexity of the feasible set, $y$ is feasible.

By convexity of $f$ we have $f(y) \leq (1 - \alpha)f(x) + \alpha f(z) < f(x)$, which contradicts the local optimum definition.

Therefore, $x$ is globally optimal.
"...in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."

Nonlinear optimization
Topics of this part of the course

**Conditions to characterize minima**

**Algorithms to find** (local) minima

(if applied to **convex problems**, they find **global minima**)
Introduction to nonlinear optimization

Today, we learned to:

• **Define** nonlinear optimization problems

• **Understand** convex analysis fundamentals (sets, cones, functions, and gradients)

• **Verify** convexity and **construct** convex optimization problems

• **Define** convex optimization problems in CVXPY

• **Understand** the importance of *convexity vs nonconvexity* in optimization
Next lecture

- Optimality conditions in nonlinear optimization