ECE 595: Machine Learning I
Tutorial 04: Constrained Optimization

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

PURDUE
UNIVERSITY

Outline

Outline

- Equality Constrained Optimization (Same as Lecture 4)
- Inequality Constrained Optimization

Reference

- Nocedal-Wright, Numerical Optimization. (Chapter 12.3, 12.4, 12.5)
- Boyd-Vandenberghe, Convex Optimization. (Chapter 9.1, 10.1, 11.1)

Constrained Optimization

Equality Constrained Optimization:

$$\begin{aligned} \underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & h_j(\boldsymbol{x}) = 0, \qquad j = 1, \ldots, k. \end{aligned}$$

Requires a function: Lagrangian function

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\nu}) \stackrel{\text{def}}{=} f(\boldsymbol{x}) - \sum_{j=1}^{k} \nu_j h_j(\boldsymbol{x}).$$

$\boldsymbol{\nu} = [\nu_1, \ldots, \nu_k]$: **Lagrange multipliers** or the **dual variables**.

Solution $(\boldsymbol{x}^*, \boldsymbol{\nu}^*)$ satisfies

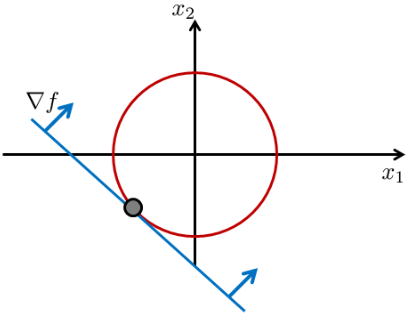
$$\begin{aligned} \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\nu}^*) &= \mathbf{0}, \\ \nabla_{\boldsymbol{\nu}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\nu}^*) &= \mathbf{0}. \end{aligned}$$

Example

- Consider the problem

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad x_1 + x_2$$
$$\text{subject to} \quad x_1^2 + x_2^2 = 2.$$

- Minimizer is $\boldsymbol{x} = (-1, -1)$.



- Objective gradient

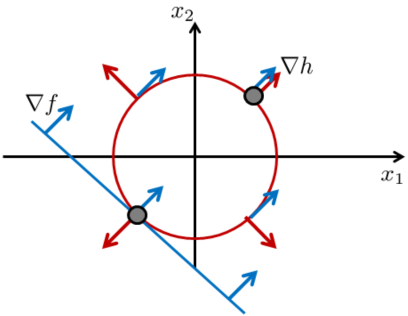
$$\nabla f(\boldsymbol{x}^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- Constraint gradient

$$\nabla h(\boldsymbol{x}^*) = \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

x_2

First Order Optimality

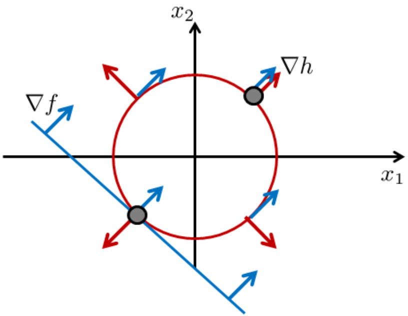


$$\nabla f(\boldsymbol{x}^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{and} \qquad \nabla h(\boldsymbol{x}^*) = \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

- Lagrangian condition holds: Put $\nu^* = -\frac{1}{2}$. Then,

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\nu}^*) = \nabla f(\boldsymbol{x}^*) - \sum_{j=1}^{k} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}.$$

Second Order Optimality



- First Order Condition: Find stationary point:

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\nu}^*) = \nabla f(\boldsymbol{x}^*) - \sum_{j=1}^{k} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0}.$$

- Second Order Condition: Determine maxima / minima:

$$\nabla_{\boldsymbol{x}\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\nu}^*) \geq 0$$

Example: ℓ_2-minimization with constraint

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_0\|^2, \qquad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}.$$

The Lagrangian function of the problem is

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\nu}) = \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_0\|^2 - \boldsymbol{\nu}^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}).$$

The first order optimality condition requires

$$\begin{aligned}
\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\nu}) &= (\boldsymbol{x}-\boldsymbol{x}_0) - \boldsymbol{A}^T\boldsymbol{\nu} = \boldsymbol{0}\\
\nabla_{\boldsymbol{\nu}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\nu}) &= \boldsymbol{A}\boldsymbol{x}-\boldsymbol{y} = \boldsymbol{0}.
\end{aligned}$$

Multiply the first equation by $\boldsymbol{A}$ on both sides:

$$\begin{aligned}
& & \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}_0) - \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} &= \boldsymbol{0}\\
&\Rightarrow & \underbrace{\boldsymbol{A}\boldsymbol{x}}_{=\boldsymbol{y}} - \boldsymbol{A}\boldsymbol{x}_0 &= \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu}\\
&\Rightarrow & \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_0 &= \boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu}\\
&\Rightarrow & (\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0) &= \boldsymbol{\nu}
\end{aligned}$$

Example: ℓ_2-minimization with constraint

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_0\|^2, \qquad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}.$$

The first order optimality condition requires

$$\begin{aligned}
\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\nu}) &= (\boldsymbol{x}-\boldsymbol{x}_0)-\boldsymbol{A}^T\boldsymbol{\nu}=\boldsymbol{0}\\
\nabla_{\boldsymbol{\nu}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\nu}) &= \boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}=\boldsymbol{0}.
\end{aligned}$$

We just showed: $\boldsymbol{\nu}=(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0)$. Substituting this result into the first order optimality yields

$$\begin{aligned}
\boldsymbol{x} &= \boldsymbol{x}_0+\boldsymbol{A}^T\boldsymbol{\nu}\\
&= \boldsymbol{x}_0+\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0)
\end{aligned}$$

Therefore, the solution is $\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0)$.

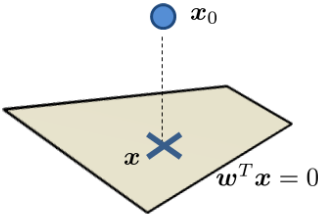
Special Case

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{x}_0\|^2, \qquad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}.$$

Special case: When $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$ is simplified to $\boldsymbol{w}^T\boldsymbol{x}=0$.

- $\boldsymbol{w}^T\boldsymbol{x}=0$ is a line.
- Find a point $\boldsymbol{x}$ on the line that is closest to $\boldsymbol{x}_0$.
- Solution is

$$\begin{aligned}\boldsymbol{x} &= \boldsymbol{x}_0 + \boldsymbol{w}(\boldsymbol{w}^T\boldsymbol{w})^{-1}(0-\boldsymbol{w}^T\boldsymbol{x}_0)\\ &= \boldsymbol{x}_0 - \left(\frac{\boldsymbol{w}^T\boldsymbol{x}_0}{\|\boldsymbol{w}\|^2}\right)^T\boldsymbol{w}.\end{aligned}$$



In practice ...

- Use CVX to solve problem
- Here is a MATLAB code
- Exercise: Turn it into Python.

```
% MATLAB code: Use CVX to solve min ||x - x0||, s.t. Ax = y
m = 3; n = 2*m;
A     = randn(m,n); xstar = randn(n,1);
y     = A*xstar;
x0    = randn(n,1);
cvx_begin
   variable x(n)
   minimize( norm(x-x0) )
   subject to
        A*x == y;
cvx_end
% you may compare with the solution x0 + A'*inv(A*A')*(y-A*x0).
```

ℓ_1-minimization with constraint

Solve the ℓ_1 problem:

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \quad \|\boldsymbol{x}\|_1,$$
$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}.$$

```
% MATLAB code: Use CVX to solve min ||x||_1, s.t. Ax <= y
m = 100; n = 50;
A  = randn(m,n);
x0 = randn(n,1);
y  = A*x0 + rand(m,1);
cvx_begin
  variable x_l1(n)
  minimize( norm( x_l1, 1 ) )
  subject to
    A*x_l1 == y;
cvx_end
```

Outline

- Equality Constrained Optimization (Same as Lecture 4)
- Inequality Constrained Optimization

Reference

- Nocedal-Wright, Numerical Optimization. (Chapter 12.3, 12.4, 12.5)
- Boyd-Vandenberghe, Convex Optimization. (Chapter 9.1, 10.1, 11.1)

Inequality Constrained Optimization

Inequality constrained optimization:

$$
\begin{aligned}
\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f(\boldsymbol{x}) \\
\text{subject to} \quad & g_i(\boldsymbol{x}) \geq 0, \qquad i = 1, \ldots, m \\
& h_j(\boldsymbol{x}) = 0, \qquad j = 1, \ldots, k.
\end{aligned}
$$

Requires a function: Lagrangian function

$$
\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\nu}) \overset{\text{def}}{=} f(\boldsymbol{x}) - \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) - \sum_{j=1}^{k} \nu_j h_j(\boldsymbol{x}).
$$

$\boldsymbol{\mu} \in \mathbb{R}^m$ and $\boldsymbol{\nu} \in \mathbb{R}^k$ are called the **Lagrange multipliers** or the **dual variables**.

Karush-Kahn-Tucker Conditions

If $(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\nu}^*)$ is the solution to the constrained optimization, then all the following conditions should hold:

(i) $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\mu}^*, \boldsymbol{\nu}^*) = \mathbf{0}$.

- **Stationarity.**
- The primal variables should be stationary.

(ii) $g_i(\boldsymbol{x}^*) \geq 0$ and $h_j(\boldsymbol{x}^*) = 0$ for all i and j.

- **Primal Feasibility.**
- Ensures that constraints are satisfied.

(iii) $\mu_i^* \geq 0$ for all i and j.

- **Dual Feasibility.**
- Require $\mu_i^* \geq 0$; but no constraint on ν_i^*.

(iv) $\mu_i^* g_i(\boldsymbol{x}^*) = 0$ for all i and j.

- **Complementary Slackness**
- Either $\mu_i^* = 0$ or $g_i(\boldsymbol{x}^*) = 0$ (or both).

KKT Condition is a first order **necessary** condition.

Example: ℓ_2-minimization with two constraints

Solve the following least squares over positive quadrant problem.

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{b}\|^2, \\ \text{subject to} \quad & \boldsymbol{x}^T\mathbf{1} = 1, \quad \text{and} \quad \boldsymbol{x} \geq \mathbf{0}. \end{aligned} \tag{1}$$

```
%MATLAB code: Use CVX to solve min ||x-b|| s.t. sum(x) = 1, x >= 0.
cvx_begin
   variable x(n)
   minimize( norm(x-b, 2) )
   subject to
        sum(x) == 1;
        x      >= 0;
cvx_end
```

Analytic Solution

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \gamma) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2 - \boldsymbol{\lambda}^T\boldsymbol{x} - \gamma(1 - \boldsymbol{x}^T\mathbf{1}).$$

Stationarity suggests that:

$$\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \gamma) = \boldsymbol{x} - \boldsymbol{b} - \boldsymbol{\lambda} + \gamma\mathbf{1} = \mathbf{0}$$

This means

$$\boldsymbol{x} = \boldsymbol{b} + \boldsymbol{\lambda} - \gamma\mathbf{1} \quad \text{or} \quad x_i = b_i + \lambda_i - \gamma.$$

The complementary slackness implies $\lambda_i x_i = 0$.

- Case 1: If $\lambda_i = 0$, then
  - $x_i = b_i + \cancelto{0}{\lambda_i} - \gamma = b_i - \gamma$.
  - Since constraint requires $x_i \geq 0$, this means $b_i \geq \gamma$.
- Case 2: If $\lambda_i > 0$, then $x_i = 0$.
  - $\cancelto{0}{x_i} = b_i + \lambda_i - \gamma$.
  - This implies $b_i + \lambda_i = \gamma$.
  - Since $\lambda_i > 0$, this implies $b_i < \gamma$.

These three cases can be re-written as:

- If $b_i > \gamma$, then $x_i = b_i - \gamma$;
- If $b_i = \gamma$, then $x_i = 0$;
- If $b_i < \gamma$, then $x_i = 0$.

Compactly written as

$$\boldsymbol{x} = \max(\boldsymbol{b} - \gamma \mathbf{1}, 0).$$

Primal feasibility implies that

$$\boldsymbol{x}^T \mathbf{1} = 1, \qquad \Leftrightarrow \qquad \sum_{i=1}^{n} x_i = 1.$$

Therefore, γ needs to satisfy the equation

$$\sum_{i=1}^{n} \max(b_i - \gamma, 0) = 1,$$

which can be found by doing a root-finding of

$$g(\gamma) = \sum_{i=1}^{n} \max(b_i - \gamma, 0) - 1.$$

Non-CVX Implementation

```
%MATLAB code: solve min ||x-b|| s.t. sum(x) = 1, x >= 0.
n = 10;
b = randn(n,1);
fun = @(gamma) myfun(gamma,b);
gamma = fzero(fun,0);
x = max(b-gamma,0);
```

where the function `myfun` is defined as

```
function y = myfun(gamma,b)
y = sum(max(b-gamma,0))-1;
```

Equivalence between Problems

Consider three optimization problems

$$
\begin{aligned}
\boldsymbol{x}^*_\lambda &= \underset{\boldsymbol{x}}{\text{argmin}} \;\; \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda\|\boldsymbol{x}\|^2 \\
\boldsymbol{x}^*_\alpha &= \underset{\boldsymbol{x}}{\text{argmin}} \;\; \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 \qquad \text{subject to} \;\; \|\boldsymbol{x}\|^2 \le \alpha \\
\boldsymbol{x}^*_\epsilon &= \underset{\boldsymbol{x}}{\text{argmin}} \;\; \|\boldsymbol{x}\|^2 \qquad \text{subject to} \;\; \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 \le \epsilon.
\end{aligned}
$$

They are equivalent when $\alpha = \|\boldsymbol{x}^*_\lambda\|^2$, $\epsilon = \|\boldsymbol{A}\boldsymbol{x}^*_\lambda - \boldsymbol{y}\|^2$.