ECE 595: Machine Learning I

Lecture 4.3: Intro to Optimization - Constrained Optimization

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering Purdue University

Outline

Mathematical Background

- Lecture 4: Intro to Optimization
- Lecture 5: Gradient Descent

Lecture 4: Intro to Optimization

- Unconstrained Optimization
  - First Order Optimality
  - Second Order Optimality
- Convexity
  - What is convexity?
  - Convex optimization
- Constrained Optimization
  - Lagrangian
  - Examples

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

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

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

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

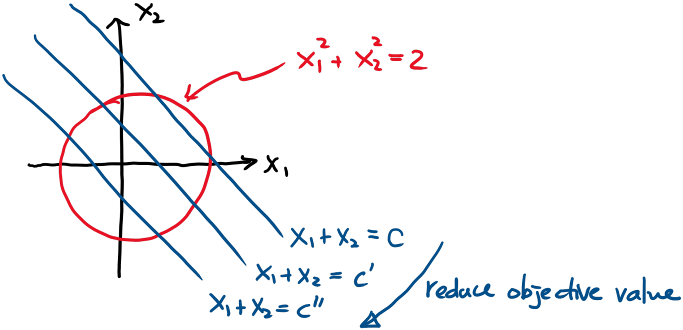
$$\begin{aligned} \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\nu}^*) &= \mathbf{0}, \\ \nabla_{\boldsymbol{\nu}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\nu}^*) &= \mathbf{0}. \end{aligned}$$

Example: Illustrating Lagrangian

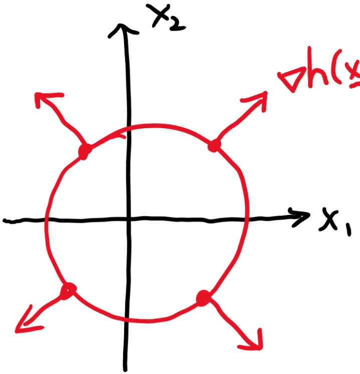
- Consider the problem

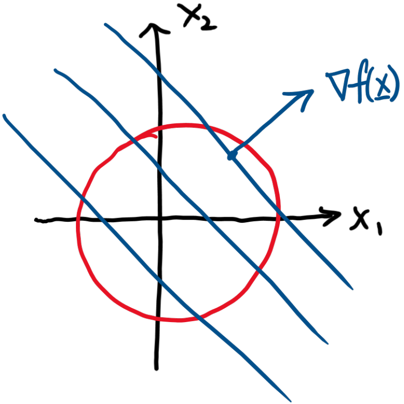
$$\underset{\boldsymbol{x}}{\text{minimize}} \quad x_1 + x_2$$
$$\text{subject to} \quad x_1^2 + x_2^2 = 2.$$

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

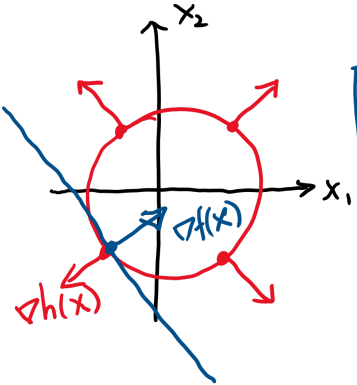


Example: Illustrating Lagrangian





Example: Illustrating Lagrangian



$$\boxed{\nabla f(x^*) = \lambda \nabla h(x^*)}$$

$$\Rightarrow \nabla \mathcal{L}(x^*, \lambda^*)$$

$$= \nabla f(x^*) - \lambda \nabla h(x^*) = 0$$

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:

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

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 $\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0)$.

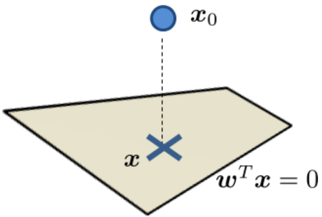
Special Case

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

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).
```

Reading List

Unconstrained Optimality Conditions

- Nocedal-Wright, Numerical Optimization. (Chapter 2.1)
- Boyd-Vandenberghe, Convex Optimization. (Chapter 9.1)

Convexity

- Nocedal-Wright, Numerical Optimization. (Chapter 1)
- Boyd-Vandenberghe, Convex Optimization. (Chapter 2 and 3)
- CMU, Convex Optimization (Lecture 2 and 4)
  https://www.stat.cmu.edu/~ryantibs/convexopt-F18/
- Stanford CS 229 (Tutorial)
  http://cs229.stanford.edu/section/cs229-cvxopt.pdf
- UCSD ECE 273 (Tutorial)
  http://eceweb.ucsd.edu/~gert/ECE273/CvxOptTutPaper.pdf

Constrained Optimization

- Nocedal-Wright, Numerical Optimization. (Chapter 12.1)