ECE 586: Vector Space Methods
Lecture 23: Constrained Optimization

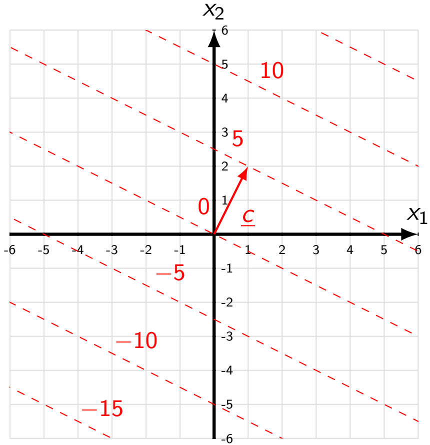
Henry D. Pfister
Duke University

Unconstrained Linear Optimization

Consider the unconstrained linear optimization problem:

$$\min_{\underline{x} \in \mathbb{R}^n} \underline{c}^T \underline{x} = \begin{cases} 0 & \text{if } \underline{c} = \underline{0} \\ -\infty & \text{otherwise.} \end{cases}$$

Figure shows labeled level sets of $\underline{c}^T \underline{x}$ for $n = 2$ and $\underline{c} = (1, 2)$.



Constrained Linear Optimization

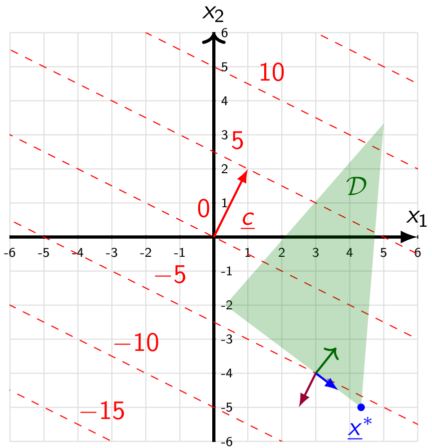
Now, consider the constrained linear optimization problem:

$$\min_{\underline{x} \in \mathcal{D}} \underline{c}^T \underline{x}.$$

Figure shows $\mathcal{D}$ with level sets of $\underline{c}^T \underline{x}$ for $n = 2$ and $\underline{c} = (1, 2)$.

◦ The optimal point $\underline{x}^*$ can be found using gradient descent from any initial feasible point.

◦ The negative gradient is reduced by the constraint normal to give a feasible descent direction.



Linear Programs

The optimization of a linear function with arbitrary affine equality and inequality constraints is called a linear program (LP).

LPs have many equivalent forms because:

- $x_1 = 0$ is the same as $(x_1 \leq 0) \wedge (x_1 \geq 0)$
- $x_1 \leq 0$ is the same as $(x_1 + x_2 = 0) \wedge (x_2 \geq 0)$ for slack variable x_2
- x_1 free is the same as $x_1 = x_2 - x_3$ with slack vars $x_2 \geq 0$ and $x_3 \geq 0$
- negation swaps: min $\leftrightarrow$ max for objective and $\geq \leftrightarrow \leq$ for constraints

Definition

Any LP can be transformed into one of the standard min forms:

$$\begin{aligned} \text{minimize} \quad & \underline{c}^T \underline{x} \\ \text{subject to} \quad & A\underline{x} = \underline{b} \\ & \underline{x} \succeq \underline{0} \end{aligned}$$	$$\begin{aligned} \text{minimize} \quad & \underline{c}^T \underline{x} \\ \text{subject to} \quad & A\underline{x} \succeq \underline{b} \\ & \underline{x} \succeq \underline{0} \end{aligned}$$

Duality in Linear Programs

Let p^* be the value of the linear program:

$$\begin{aligned} \text{minimize} \quad & x_1 + 2x_2 \\ \text{subject to} \quad & 18x_1 + 25x_2 \geq -47 \\ & 96x_1 - 87x_2 \geq 190 \\ & -75x_1 + 6x_2 \geq -355 \end{aligned}$$

Since $(x_1, x_2) = (4, -4)$ satisfies the constraints, we know $p^* \leq 4 - 8 = -4$.

One can lower bound p^* by finding a positive linear combination of the constraints that equals the objective.

Combining the constraints with coefs $(\lambda_1, \lambda_2, \lambda_3) = \left(\frac{52}{661}, 0, \frac{11}{1983}\right)$ gives

$$x_1 + 2x_2 \geq -\frac{17}{3} \approx -5.66.$$

In essence, this is how Lagrangian duality works and implies $p^* \geq -5.66$.

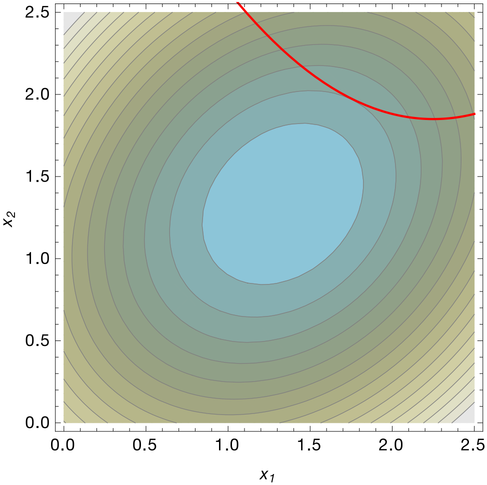
5.4: Constrained Non-Linear Optimization

Consider a constrained non-linear optimization problem over $\mathcal{D} \subseteq \mathbb{R}^n$ in the following standard form. Let $f_i : \mathcal{D} \to \mathbb{R}$ and $h_j : \mathcal{D} \to \mathbb{R}$ be real functionals for $i = 0, 1, \ldots, m$ and $j = 1, 2, \ldots, p$. Then, we write

$$\begin{array}{rll} \text{minimize} & f_0(\underline{x}) & \\ \text{subject to} & f_i(\underline{x}) \leq 0, & i = 1, 2, \ldots, m \\ & h_j(\underline{x}) = 0, & j = 1, 2, \ldots, p. \end{array}$$

- the function f_0 is called the objective function
- the functions $f_1, \ldots, f_m$ define inequality constraints
- the functions $h_1, \ldots, h_p$ define equality constraints
- feasible points in $\mathcal{F} \triangleq \{\underline{x} \in \mathcal{D} \,|\, f_i(\underline{x}) \leq 0, i \in [m], h_j(\underline{x}) = 0, j \in [p]\}$ satisfy all constraints and the problem is feasible if $\mathcal{F} \neq \emptyset$.
- the optimal value is $p^* \triangleq \inf \{f_0(\underline{x}) \,|\, \underline{x} \in \mathcal{F}\}$
- called convex if $\mathcal{D} = \mathbb{R}^n$, all f_i convex, and all h_j affine: $h_j(\underline{x}) = \underline{a}_j^T \underline{x} - b_j$

General Example



Contour plot of $f_0(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 1)^2 - x_1 x_2/2$ whose minimum occurs at $(4/3, 4/3)$ (i.e., center of blue ellipse). The red line shows the inequality constraint $f_1(x_1, x_2) = 1.85 + (x_1 - 2.25)^2/2 - x_2 \leq 0$

Langrangian Formulation

Used to transform from constrained to unconstrained optimization

Definition

For the standard optimization, the Lagrangian $L\colon \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ is

$$L(\underline{x}, \underline{\lambda}, \underline{\nu}) = f_0(\underline{x}) + \sum_{i=1}^{m} \lambda_i f_i(\underline{x}) + \sum_{j=1}^{p} \nu_j h_j(\underline{x}),$$

where the Lagrange multipliers λ_i and ν_j define penalties associated with violating the i-th inequality and j-th equality constraints, respectively.

Theorem (5.4.7) (Karush-Kuhn-Tucker)

If $\underline{x}^*$ is a constrained local optimum that satisfies a constraint qualification (e.g., mild technical conditions) and $A = \{i \in [m] \mid f_i(\underline{x}^*) = 0\}$ is the set of active constraints at $\underline{x}^*$, then there exist $\underline{\lambda}^* \geq 0$ and $\underline{\nu}^*$ such that

$$\nabla_{\underline{x}} L(\underline{x}, \underline{\lambda}, \underline{\nu}) = \nabla f_0(\underline{x}^*) + \sum_{i \in A} \lambda_i^* \nabla f_i(\underline{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\underline{x}^*) = \underline{0}.$$

Langrangian Duality

Definition

The Lagrangian dual function is defined to be

$$g(\underline{\lambda}, \underline{\nu}) \triangleq \inf_{\underline{x} \in \mathcal{D}} L(\underline{x}, \underline{\lambda}, \underline{\nu}) = \inf_{\underline{x} \in \mathcal{D}} \left(f_0(\underline{x}) + \sum_{i=1}^{m} \lambda_i f_i(\underline{x}) + \sum_{j=1}^{p} \nu_j h_j(\underline{x}) \right).$$

Lemma

The Lagrangian dual function is concave and the Lagrangian dual problem,

$$\begin{aligned} \text{maximize} \quad & g(\underline{\lambda}, \underline{\nu}) \\ \text{subject to} \quad & \underline{\lambda} \geq 0, \end{aligned}$$

has a unique max value $d^* \leq p^*$. This property is known as weak duality.

Definition

If $d^* = p^*$, then one says that strong duality holds for the problem.

Weak Duality Proof

Lagrangian dual is concave because pointwise infimum of affine functions:

$$\begin{aligned}
& g(\alpha\underline{\lambda}+(1-\alpha)\underline{\lambda}', \alpha\underline{\nu} + (1-\alpha)\underline{\nu}') \\
&\quad = \inf_{\underline{x}\in\mathcal{D}} L(\underline{x}, \alpha\underline{\lambda} + (1-\alpha)\underline{\lambda}', \alpha\underline{\nu} + (1-\alpha)\underline{\nu}') \\
&\quad = \inf_{\underline{x}\in\mathcal{D}} \left(\alpha L(\underline{x}, \underline{\lambda}, \underline{\nu}) + (1-\alpha) L(\underline{x}, \underline{\lambda}', \underline{\nu}')\right) \\
&\quad \geq \inf_{\underline{x}\in\mathcal{D}} \alpha L(\underline{x}, \underline{\lambda}, \underline{\nu}) + \inf_{\underline{x}'\in\mathcal{D}} (1-\alpha) L(\underline{x}', \underline{\lambda}', \underline{\nu}') \\
&\quad = \alpha g(\underline{\lambda}, \underline{\nu}) + (1-\alpha) g(\underline{\lambda}', \underline{\nu}').
\end{aligned}$$

Concavity implies unique maximum value d^* upper bounded by

$$\begin{aligned}
g(\underline{\lambda}, \underline{\nu}) = \inf_{\underline{x}\in\mathcal{D}} L(\underline{x}, \underline{\lambda}, \underline{\nu}) &\overset{(a)}{\leq} \inf_{\underline{x}\in\mathcal{F}} L(\underline{x}, \underline{\lambda}, \underline{\nu}) \\
&\overset{(b)}{=} p^* + \sum_{i=1}^{m} \lambda_i f_i(\underline{x}) \overset{(c)}{\leq} p^*,
\end{aligned}$$

where (a) is implied by $\mathcal{F} \subseteq \mathcal{D}$, (b) follows from $h_j(\underline{x}) = 0$ for $\underline{x} \in \mathcal{F}$, and (c) holds by combining $f_i(\underline{x}) \leq 0$ for $\underline{x} \in \mathcal{F}$ and $\lambda_i \geq 0$.

The Lagrangian Dual Problem

While $g(\underline{\lambda}, \underline{\nu})$ can be $-\infty$, this is avoided by defining the dual feasible set:

$$\mathcal{C} \triangleq \{(\underline{\lambda}, \underline{\nu}) \in \mathbb{R}^m \times \mathbb{R}^p \,|\, \underline{\lambda} \succeq \underline{0}, g(\underline{\lambda}, \underline{\nu}) > -\infty\} \, .$$

The value of the dual optimization problem is

$$d^* = \sup_{(\underline{\lambda}, \underline{\nu}) \in \mathcal{C}} g(\underline{\lambda}, \underline{\nu}).$$

If $\mathcal{C} \neq \emptyset$, then the dual problem is feasible and, by definition, $d^* > -\infty$.

Consider the LP defined by

$$\begin{aligned} \text{minimize} \quad & \underline{c}^T \underline{x} \\ \text{subject to} \quad & A\underline{x} = \underline{b} \\ & \underline{x} \succeq \underline{0}. \end{aligned}$$

Lagrangian Duality for Linear Programs

For the previous LP, the Lagrangian is given by

$$L(\underline{x}, \underline{\lambda}, \underline{\nu}) = \underline{c}^T \underline{x} + \underline{\nu}^T (\underline{b} - A\underline{x}) - \underline{\lambda}^T \underline{x},$$

where $-\underline{\lambda}^T \underline{x}$ corresponds to $\underline{x} \succeq \underline{0}$ and the Lagrangian dual function is

$$g(\underline{\lambda}, \underline{\nu}) = \inf_{\underline{x} \in \mathcal{D}} L(\underline{x}, \underline{\lambda}, \underline{\nu}) = \begin{cases} \underline{b}^T \underline{\nu} & \text{if } \underline{c} - A^T \underline{\nu} - \underline{\lambda} = \underline{0} \\ -\infty & \text{otherwise.} \end{cases}$$

Adding the implied constraint and using $\underline{\lambda} \succeq \underline{0}$, one gets the dual LP problem

$$\begin{aligned} \text{maximize} \quad & \underline{b}^T \underline{\nu} \\ \text{subject to} \quad & A^T \underline{\nu} \preceq \underline{c}. \end{aligned}$$

Strong duality for linear programs says that, if the original LP has an optimal solution (i.e., it is neither unbounded nor infeasible), then the dual LP has an optimal solution of the same value.

Next Steps

- To continue studying after this video –
  - Try the required reading: Course Notes EF 5.4 - 5.4.3
  - Also, look at the problems in Assignment 9