Optimization III: Constrained Optimization

CS 205A:
Mathematical Methods for Robotics, Vision, and Graphics

Doug James (and Justin Solomon)

Announcements

- HW6 due today
- HW7 out
- HW8 (last homework) out next Thursday

Constrained Problems

$$\begin{aligned} \text{minimize} \quad & f(\vec{x}) \\ \text{such that} \quad & g(\vec{x}) = \vec{0} \\ & h(\vec{x}) \geq \vec{0} \end{aligned}$$

Really Difficult!

Simultaneously:

- Minimizing f
- Finding roots of g
- Finding feasible points of h

Implicit Projection

Implicit surface: $g(\vec{x}) = 0$

Implicit Projection

Implicit surface: $g(\vec{x}) = 0$

Example: Closest point on surface

$$\begin{array}{l} \text{minimize}_{\vec{x}} \; \|\vec{x} - \vec{x}_0\|_2 \\ \text{such that} \;\; g(\vec{x}) = 0 \end{array}$$

Nonnegative Least-Squares

$$\text{minimize}_{\vec{x}} \; \|A\vec{x} - \vec{b}\|_2^2$$
$$\text{such that} \;\; \vec{x} \geq \vec{0}$$

Manufacturing

- m materials
- s_i units of material i in stock
- n products
- p_j profit for product j
- Product j uses c_{ij} units of material i

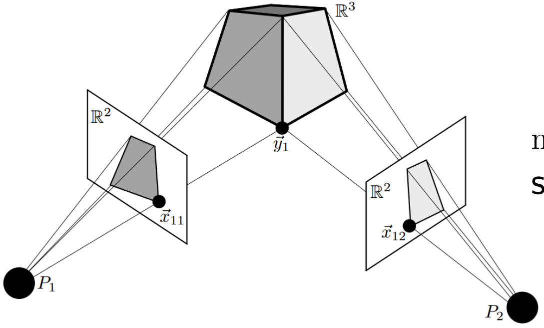
Manufacturing

Linear programming problem:

$$\begin{aligned} \text{maximize}_{\vec{x}} \quad & \sum_j p_j x_j \\ \text{such that} \quad & x_j \geq 0 \ \forall j \\ & \sum_j c_{ij} x_j \leq s_i \ \forall i \end{aligned}$$

"Maximize profits where you make a positive amount of each product and use limited material."

Bundle Adjustment



$$\min_{\vec{y}_j, P_i} \sum_{ij} \|P_i \vec{y}_j - \vec{x}_{ij}\|_2^2$$
$$\text{s.t.} \quad P_i \text{ orthogonal } \forall i$$

Applications:

- **Bundler**
- **Building Rome in a Day**

Constrained Problems

$$\begin{aligned} \text{minimize} \quad & f(\vec{x}) \\ \text{such that} \quad & g(\vec{x}) = \vec{0} \\ & h(\vec{x}) \geq \vec{0} \end{aligned}$$

Basic Definitions

Feasible point and feasible set

A *feasible point* is any point $\vec{x}$ satisfying $g(\vec{x}) = \vec{0}$ and $h(\vec{x}) \geq \vec{0}$. The *feasible set* is the set of all points $\vec{x}$ satisfying these constraints.

Basic Definitions

Feasible point and feasible set

A *feasible point* is any point $\vec{x}$ satisfying $g(\vec{x}) = \vec{0}$ and $h(\vec{x}) \geq \vec{0}$. The *feasible set* is the set of all points $\vec{x}$ satisfying these constraints.

Critical point of constrained optimization

A critical point is one satisfying the constraints that also is a local maximum, minimum, or saddle point of f within the feasible set.

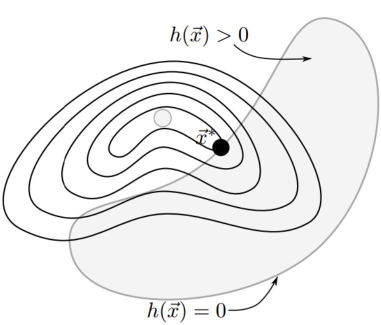
Differential Optimality

Without h:

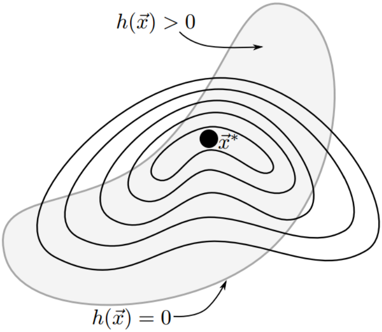
$$\Lambda(\vec{x}, \vec{\lambda}) \equiv f(\vec{x}) - \vec{\lambda} \cdot g(\vec{x})$$

Lagrange Multipliers

Inequality Constraints at $\vec{x}^*$



Active constraint
$h(\vec{x}^*) = 0$



Inactive constraint
$h(\vec{x}^*) > 0$

Inequality Constraints at $\vec{x}^*$

Two cases:

- **Active:** $h_i(\vec{x}^*) = 0$
  Optimum might change if constraint is removed
- **Inactive:** $h_i(\vec{x}^*) > 0$
  Removing constraint does not change $\vec{x}^*$ locally

Idea

Remove inactive constraints and make active constraints equality constraints.

Lagrange Multipliers

$$\Lambda(\vec{x}, \vec{\lambda}, \vec{\mu}) \equiv f(\vec{x}) - \vec{\lambda} \cdot g(\vec{x}) - \vec{\mu} \cdot h(\vec{x})$$

No longer a critical point! But if we ignore that:

$$\vec{0} = \nabla f(\vec{x}) - \sum_i \lambda_i \nabla g_i(\vec{x}) - \sum_j \mu_j \nabla h_j(\vec{x})$$

Lagrange Multipliers

$$\Lambda(\vec{x}, \vec{\lambda}, \vec{\mu}) \equiv f(\vec{x}) - \vec{\lambda} \cdot g(\vec{x}) - \vec{\mu} \cdot h(\vec{x})$$

No longer a critical point! But if we ignore that:

$$\vec{0} = \nabla f(\vec{x}) - \sum_i \lambda_i \nabla g_i(\vec{x}) - \sum_j \mu_j \nabla h_j(\vec{x})$$

$$\mu_j h_j(\vec{x}) = 0$$

Zero out inactive constraints!

Inequality Direction

So far: Have not distinguished between $h_j(\vec{x}) \geq 0$ and $h_j(\vec{x}) \leq 0$

Inequality Direction

So far: Have not distinguished between $h_j(\vec{x}) \geq 0$ and $h_j(\vec{x}) \leq 0$

- Direction to decrease f: $-\nabla f(\vec{x}^*)$
- Direction to decrease h_j: $-\nabla h_j(\vec{x}^*)$

Inequality Direction

So far: Have not distinguished between
$h_j(\vec{x}) \geq 0$ and $h_j(\vec{x}) \leq 0$

- Direction to decrease f: $-\nabla f(\vec{x}^*)$
- Direction to decrease h_j: $-\nabla h_j(\vec{x}^*)$

$$\nabla f(\vec{x}^*) \cdot \nabla h_j(\vec{x}^*) \geq 0$$

Dual Feasibility

$$\mu_j \geq 0$$

KKT Conditions

Theorem (Karush-Kuhn-Tucker (KKT) conditions)

$\vec{x}^* \in \mathbb{R}^n$ is a critical point when there exist $\vec{\lambda} \in \mathbb{R}^m$ and $\vec{\mu} \in \mathbb{R}^p$ such that:

- $\vec{0} = \nabla f(\vec{x}^*) - \sum_i \lambda_i \nabla g_i(\vec{x}^*) - \sum_j \mu_j \nabla h_j(\vec{x}^*)$ *("stationarity")*
- $g(\vec{x}^*) = \vec{0}$ *and* $h(\vec{x}) \geq \vec{0}$ *("primal feasibility")*
- $\mu_j h_j(\vec{x}^*) = 0$ *for all* j *("complementary slackness")*
- $\mu_j \geq 0$ *for all* j *("dual feasibility")*

KKT Example from Book

Example 10.6 (KKT conditions)**.** Suppose we wish to solve the following optimization (proposed by R. Israel, UBC Math 340, Fall 2006):

$$\begin{aligned} \text{maximize } & xy \\ \text{subject to } & x + y^2 \leq 2 \\ & x, y \geq 0. \end{aligned}$$

In this case we will have no λ's and three μ's. We take $f(x,y) = -xy$, $h_1(x,y) \equiv 2-x-y^2$, $h_2(x,y) = x$, and $h_3(x,y) = y$. The KKT conditions are:

$$\begin{aligned} \text{Stationarity: } & 0 = -y + \mu_1 - \mu_2 \\ & 0 = -x + 2\mu_1 y - \mu_3 \\ \text{Primal feasibility: } & x + y^2 \leq 2 \\ & x, y \geq 0 \\ \text{Complementary slackness: } & \mu_1(2 - x - y^2) = 0 \\ & \mu_2 x = 0 \\ & \mu_3 y = 0 \\ \text{Dual feasibility: } & \mu_1, \mu_2, \mu_3 \geq 0 \end{aligned}$$

KKT Example from Book

Example 10.7 (Linear programming)**.** Consider the optimization:

$$\begin{aligned} &\text{minimize}_{\vec{x}}\ \vec{b} \cdot \vec{x} \\ &\text{subject to } A\vec{x} \geq \vec{c}. \end{aligned}$$

Example 10.2 can be written this way. The KKT conditions for this problem are:

$$\begin{aligned} \text{Stationarity: } & A^\top \vec{\mu} = \vec{b} \\ \text{Primal feasibility: } & A\vec{x} \geq \vec{c} \\ \text{Complementary slackness: } & \mu_i(\vec{a}_i \cdot \vec{x} - c_i) = 0\ \forall i, \text{ where } \vec{a}_i^\top \text{ is row } i \text{ of } A \\ \text{Dual feasibility: } & \vec{\mu} \geq \vec{0} \end{aligned}$$

Physical Illustration of KKT

Example: Minimal gravitational-potential-energy position $\vec{x} = (x_1, x_2)^T$ of a particle attached to inextensible rod (of length ℓ), and above a hard surface.

$$\begin{aligned} \text{minimize}_{\vec{x}} \quad & x_2 && \text{(Minimize gravitational potential energy)} \\ \text{such that} \quad & \|\vec{x} - \vec{c}\|_2 - \ell = 0 && \text{(rod of length } \ell \text{ attached at } \vec{c}\text{)} \\ & x_2 \geq 0 && \text{(height} \geq 0\text{)} \end{aligned}$$

Physical interpretation of f, g, h, λ and μ?
Physical interpretation of stationarity, primal feasibility, complementary slackness and dual feasibility?

Sequential Quadratic Programming (SQP)

$$\vec{x}_{k+1} \equiv \vec{x}_k + \arg\min_{\vec{d}} \left[\frac{1}{2} \vec{d}^\top H_f(\vec{x}_k)\vec{d} + \nabla f(\vec{x}_k) \cdot \vec{d} \right]$$

$$\text{such that } g_i(\vec{x}_k) + \nabla g_i(\vec{x}_k) \cdot \vec{d} = 0$$

$$h_i(\vec{x}_k) + \nabla h_i(\vec{x}_k) \cdot \vec{d} \geq 0$$

Equality Constraints Only

$$\begin{pmatrix} H_f(\vec{x}_k) & [Dg(\vec{x}_k)]^\top \\ Dg(\vec{x}_k) & 0 \end{pmatrix} \begin{pmatrix} \vec{d} \\ \vec{\lambda} \end{pmatrix} = \begin{pmatrix} -\nabla f(\vec{x}_k) \\ -g(\vec{x}_k) \end{pmatrix}$$

- Can approximate H_f
- Can limit distance along $\vec{d}$

Inequality Constraints

Active set methods:

Keep track of active constraints and enforce as equality, update based on gradient

Barrier Methods: Equality Case

$$f_\rho(\vec{x}) \equiv f(\vec{x}) + \rho \|g(\vec{x})\|_2^2$$

Unconstrained optimization, crank up ρ until $g(\vec{x}) \approx \vec{0}$

Caveat: H_{f_ρ} becomes poorly conditioned

Barrier Methods: Inequality Case

$$\text{Inverse barrier: } \frac{1}{h_i(\vec{x})}$$

$$\text{Logarithmic barrier: } -\log h_i(\vec{x})$$

To Read: Convex Programming

A ray of hope:

Minimizing convex functions with convex constraints

Next