Optimization III: Constrained Optimization


Doug James (and Justin Solomon)
Announcements

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

minimize \( f(\vec{x}) \)
such that \( g(\vec{x}) = 0 \)
\( h(\vec{x}) \geq 0 \)
Really Difficult!

Simultaneously:

- Minimizing $f$
- Finding roots of $g$
- Finding feasible points of $h$
Implicit surface: \( g(\vec{x}) = 0 \)
Implicit surface: \( g(\vec{x}) = 0 \)

Example: Closest point on surface

\[
\begin{align*}
\text{minimize } \vec{x} & \quad \| \vec{x} - \vec{x}_0 \|_2 \\
\text{such that } & \quad g(\vec{x}) = 0
\end{align*}
\]
Nonnegative Least-Squares

$$\text{minimize } \| A\vec{x} - \vec{b} \|_2^2$$

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:

$$\text{maximize } \vec{x} \sum_j p_j x_j$$

such that

$$x_j \geq 0 \ \forall j$$

$$\sum_j c_{ij} x_j \leq s_i \ \forall i$$

“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

minimize \( f(\vec{x}) \)

such that \( g(\vec{x}) = \vec{0} \)

\( h(\vec{x}) \geq \vec{0} \)
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!
So far: Have not distinguished between
\[ h_j(\vec{x}) \geq 0 \text{ and } h_j(\vec{x}) \leq 0 \]
Inequality Direction

So far: Have not distinguished between

\[ h_j(\vec{x}) \geq 0 \text{ and } h_j(\vec{x}) \leq 0 \]

- 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

\[ h_j(x) \geq 0 \text{ and } h_j(x) \leq 0 \]

- Direction to decrease \( f \): \(-\nabla f(x^*)\)
- Direction to decrease \( h_j \): \(-\nabla h_j(x^*)\)

\[ \nabla f(x^*) \cdot \nabla h_j(x^*) \geq 0 \]
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")
Example 10.6 (KKT conditions). Suppose we wish to solve the following optimization (proposed by R. Israel, UBC Math 340, Fall 2006):

maximize $xy$

subject to $x + y^2 \leq 2$

$x, y \geq 0$.

In this case we will have no $\lambda$’s and three $\mu$’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:

Stationarity: $0 = -y + \mu_1 - \mu_2$

$0 = -x + 2\mu_1 y - \mu_3$

Primal feasibility: $x + y^2 \leq 2$

$x, y \geq 0$

Complementary slackness: $\mu_1 (2 - x - y^2) = 0$

$\mu_2 x = 0$

$\mu_3 y = 0$

Dual feasibility: $\mu_1, \mu_2, \mu_3 \geq 0$
Example 10.7 (Linear programming). Consider the optimization:

$$\text{minimize}_x \ b \cdot \ x$$
subject to $A\bar{x} \geq \bar{c}$.

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

Stationarity: $A^\top \bar{\mu} = \bar{b}$

Primal feasibility: $A\bar{x} \geq \bar{c}$

Complementary slackness: $\mu_i (\bar{a}_i \cdot \bar{x} - c_i) = 0 \ \forall i$, where $\bar{a}_i^\top$ is row $i$ of $A$

Dual feasibility: $\bar{\mu} \geq \bar{0}$
Example: Minimal gravitational-potential-energy position $\vec{x} = (x_1, x_2)^T$ of a particle attached to inextensible rod (of length $\ell$), and above a hard surface.

$$\begin{align*}
\text{minimize } & \quad x_2 \\
\text{such that } & \quad \|\vec{x} - \vec{c}\|_2^2 - \ell = 0 \\
& \quad x_2 \geq 0
\end{align*}$$

(Minimize gravitational potential energy)

(rod of length $\ell$ attached at $\vec{c}$)

(height $\geq 0$)

Physical interpretation of $f$, $g$, $h$, $\lambda$ and $\mu$?
Physical interpretation of stationarity, primal feasibility, complementary slackness and dual feasibility?
Sequential Quadratic Programming (SQP)

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

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 $\vec{d}$
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 \( \rho \) until

\[ g(\vec{x}) \approx \vec{0} \]

Caveat: \( H_{f_\rho} \) becomes poorly conditioned
Barrier Methods: Inequality Case

Inverse barrier: \( \frac{1}{h_i(\bar{x})} \)

Logarithmic barrier: \( -\log h_i(\bar{x}) \)
A ray of hope: Minimizing convex functions with convex constraints