Optimization III: Constrained Optimization

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

Doug James (and Justin Solomon)
Constrained Problems

\[
\text{minimize } \quad f(\vec{x})
\]
\[
\text{such that } \quad g(\vec{x}) = \vec{0}
\]
\[
\quad h(\vec{x}) \geq \vec{0}
\]
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

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

\[
\minimize_{\vec{x}} \| A\vec{x} - \vec{b} \|_2^2
\]
\[
such \text{ that } \vec{x} \geq \vec{0}
\]
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

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

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

\[ \min_{\vec{y}_j, P_i} \sum_{ij} \| P_i \vec{y}_j - \vec{x}_{ij} \|^2_2 \]

s.t. \( P_i \) orthogonal \( \forall i \)
Basic Definitions

Feasible point and feasible set

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

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

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

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

- 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

\[ \mu_j \geq 0 \]
Theorem (Karush-Kuhn-Tucker (KKT) conditions)

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

1. \( \nabla \mathbf{0} = \nabla f(\mathbf{x}^*) - \sum_i \lambda_i \nabla g_i(\mathbf{x}^*) - \sum_j \mu_j \nabla h_j(\mathbf{x}^*) \) ("stationarity")
2. \( g(\mathbf{x}^*) = \mathbf{0} \) and \( h(\mathbf{x}) \geq \mathbf{0} \) ("primal feasibility")
3. \( \mu_j h_j(\mathbf{x}^*) = 0 \) for all \( j \) ("complementary slackness")
4. \( \mu_j \geq 0 \) for all \( j \) ("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

\[
\begin{align*}
g_i(\vec{x}_k) + \nabla g_i(\vec{x}_k) \cdot \vec{d} &= 0 \\
h_i(\vec{x}_k) + \nabla h_i(\vec{x}_k) \cdot \vec{d} &\geq 0
\end{align*}
\]
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

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

Unconstrained optimization, crank up \( \rho \) until
\[ g(\vec{x}) \approx 0 \]

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

Inverse barrier: \[ \frac{1}{h_i(\vec{x})} \]

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