Optimization III: Constrained Optimization


Justin Solomon
Constrained Problems

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

such that
\[
\begin{align*}
g(\vec{x}) &= \vec{0} \\
h(\vec{x}) &\geq \vec{0}
\end{align*}
\]
Really Difficult!

Simultaneously:

- Minimizing $f$
- Finding roots of $g$
- Finding feasible points of $h$
Implicit Projection

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

minimize \( \| \vec{x} - \vec{x}_0 \|_2 \)

such that \( g(\vec{x}) = 0 \)
Nonnegative Least-Squares

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

\begin{align*}
\text{maximize } & \quad \sum_j p_j x_j \\
\text{such that } & \quad x_j \geq 0 \ \forall j \\
& \quad \sum_j c_{ij} x_j \leq s_i \ \forall i \\
\end{align*}

“Maximize profits where you make a positive amount of each product and use limited material.”
Bundle Adjustment

\[
\min_{\tilde{y}_j, P_i} \sum_{i,j} \left\| P_i \tilde{y}_j - \tilde{x}_{ij} \right\|_2^2
\]

s.t. \( P_i \) orthogonal \( \forall i \)
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 $\mathbf{x}^*$

Two cases:

- **Active:** $h_i(\mathbf{x}^*) = 0$
  Optimum might change if constraint is removed

- **Inactive:** $h_i(\mathbf{x}^*) > 0$
  Removing constraint does not change $\mathbf{x}^*$ 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:

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

\[ \mu_j h_j(\vec{x}) = \vec{0} \]

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

\[ \mu_j \geq 0 \]
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:

- \( \mathbf{0} = \nabla f(\mathbf{x}^*) - \sum_i \lambda_i \nabla g_i(\mathbf{x}^*) - \sum_j \mu_j \nabla h_j(\mathbf{x}^*) \)  
  (“stationarity”)

- \( g(\mathbf{x}^*) = \mathbf{0} \) and \( h(\mathbf{x}) \geq \mathbf{0} \)  
  (“primal feasibility”)

- \( \mu_j h_j(\mathbf{x}^*) = 0 \) for all \( j \)  
  (“complementary slackness”)

- \( \mu_j \geq 0 \) for all \( j \)  
  (“dual feasibility”)

KKT Conditions

Theorem (Karush-Kuhn-Tucker (KKT) conditions)
Sequential Quadratic Programming (SQP)

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

such that
\[ g_i(\bar{x}_k) + \nabla g_i(\bar{x}_k) \cdot \bar{d} = 0 \]
\[ h_i(\bar{x}_k) + \nabla h_i(\bar{x}_k) \cdot \bar{d} \geq 0 \]
Equality Constraints Only

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

- Can approximate \( H_f \)
- Can limit distance along \( d \)
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 \( \rho \) until
\[ g(\vec{x}) \approx \vec{0} \]

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

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

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