Topics in Applied Optimization

Optimization for ML and Data Sciences

Dr. Pawan Kumar

September 30, 2020

IIIT, Hyderabad

Equivalent Convex Problems: Eliminating Equality Constraints

Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Convex Optimization Problem (Standard Form):

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& a_i^T x = b_i, \quad i = 1, \ldots, p
\end{array}
$$

where $f_0, \ldots, f_m$ are convex functions.

Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Convex Optimization Problem (Standard Form):

$$\begin{array}{rl} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p \end{array}$$

where $f_0, \ldots, f_m$ are convex functions.

Eliminating Equality Constrains:

Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Convex Optimization Problem (Standard Form):

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& a_i^T x = b_i, \quad i = 1, \ldots, p
\end{aligned}
$$

where $f_0, \ldots, f_m$ are convex functions.

Eliminating Equality Constrains:

- Equality constraints are of the form $Ax = b$. To eliminate it, we do

Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Convex Optimization Problem (Standard Form):

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& a_i^T x = b_i, \quad i = 1, \ldots, p
\end{aligned}
$$

where $f_0, \ldots, f_m$ are convex functions.

Eliminating Equality Constrains:

- Equality constraints are of the form $Ax = b$. To eliminate it, we do
  - Find a particular solution x_0 of $Ax = b$

Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Convex Optimization Problem (Standard Form):

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p \end{aligned}$$

where $f_0, \ldots, f_m$ are convex functions.

Eliminating Equality Constrains:

- Equality constraints are of the form $Ax = b$. To eliminate it, we do
  - Find a particular solution x_0 of $Ax = b$
  - Consider a matrix whose range is nullspace of A. A general solution to $Ax = b$ is then given by $x = Fz + x_0$

Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Convex Optimization Problem (Standard Form):

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p \end{aligned}$$

where $f_0, \ldots, f_m$ are convex functions.

Eliminating Equality Constrains:

- Equality constraints are of the form $Ax = b$. To eliminate it, we do
  - Find a particular solution x_0 of $Ax = b$
  - Consider a matrix whose range is nullspace of A. A general solution to $Ax = b$ is then given by $x = Fz + x_0$

$$\begin{aligned} \text{minimize} \quad & f_0(Fz + x_0) \\ \text{subject to} \quad & f_i(Fz + x_0), \quad i = 0, \ldots, m \quad \leftarrow \text{No equality constraints!} \end{aligned}$$

Equivalent Convex Problems: Eliminating Equality Constraints

Recall the convex optimzation problem in standard form:

Convex Optimization Problem (Standard Form):

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p \end{array}$$

where $f_0, \ldots, f_m$ are convex functions.

Eliminating Equality Constrains:

- Equality constraints are of the form $Ax = b$. To eliminate it, we do
  - Find a particular solution x_0 of $Ax = b$
  - Consider a matrix whose range is nullspace of A. A general solution to $Ax = b$ is then given by $x = Fz + x_0$

$$\begin{array}{ll} \text{minimize} & f_0(Fz + x_0) \\ \text{subject to} & f_i(Fz + x_0) \leq 0, \quad i = 0, \ldots, m \end{array}$$

$\leftarrow$ No equality constraints!

Hence, equality constraints are now eliminated (but it may destroy sparsity!)

Linear Optimization Problems

Linear Optimization Problems

$f_0(x)$ $f_i(x)$, $h_i(x)$ $i \geq 1$

optimization problem

When the objective and constraint functions are all affine, the problem is called linear program.

Linear Optimization Problems

When the objective and constraint functions are all affine, the problem is called linear program.

Linear Program (General Form):

$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{subject to} \quad & Gx \preceq h, \\ & Ax = b, i = 1, \ldots, p, \end{aligned}$$

affine

affine

where $G \in \mathbb{R}^{m\times n}$ and $A \in \mathbb{R}^{p\times n}$

Linear Optimization Problems

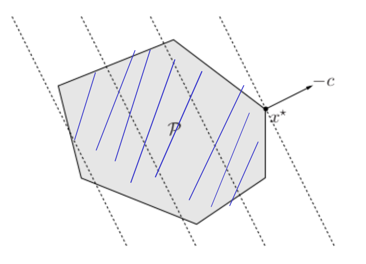
When the objective and constraint functions are all affine, the problem is called linear program.

Linear Program (General Form):

$$
\begin{aligned}
\text{minimize} \quad & c^T x + d \\
\text{subject to} \quad & Gx \leq h, \\
& Ax = b, i = 1, \ldots, p,
\end{aligned}
$$

where $G \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{p \times n}$

Figure 1: Linear program



Linear Optimization Problems

When the objective and constraint functions are all affine, the problem is called linear program.

Linear Program (General Form):

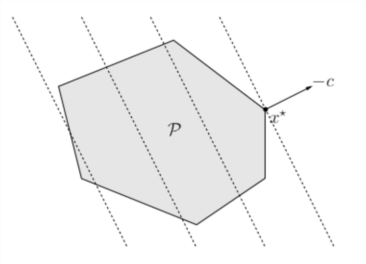
$$
\begin{aligned}
\text{minimize} \quad & c^T x + d \\
\text{subject to} \quad & Gx \leq h, \\
& Ax = b, i = 1, \ldots, p,
\end{aligned}
$$

constant
can be dropped

$x^2 + 2$ ← does not affect optimal point x^*.
x^2

where $G \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{p \times n}$

Figure 1: Linear program



- common to omit d in objective function
- can maximize an affine objective

$$c^T x + d$$

by minimizing

$$-c^T x - d$$

Quadratic and Quadratically Optimization Problems

Quadratic and Quadratically Optimization Problems

Quadratic Program:

$$
\begin{aligned}
\text{minimize} \quad & (1/2)x^T P x + q^T x + r \\
\text{subject to} \quad & Gx \leq h, \\
& Ax = b, \quad i = 1, \ldots, p,
\end{aligned}
$$

where $P \in S^n_+$, $G \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{p \times n}$.

General form of a quadratic fn. or form c f(x)? f(x)

- $f_0(x)$ is quadr.
- ineq. constr. fn is affine
- eq. constr. f_i is aff.

Quadratic and Quadratically Optimization Problems

Quadratic Program:

ineq.
affine, constr't.

$$
\begin{aligned}
\text{minimize} \quad & (1/2)x^T Px + q^T x + r \\
\text{subject to} \quad & Gx \le h, \\
& Ax = b, \quad i = 1, \ldots, p,
\end{aligned}
$$

where $P \in S_+^n$, $G \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{p \times n}$.

quadratic. quadratic constraint

Quadratically Constrained Quadratic Program:

$$
\begin{aligned}
\text{minimize} \quad & (1/2)x^T P_0 x + q^T x + r \\
\text{subject to} \quad & (1/2)x^T P_i x + q_i^T x + r_i \le 0, \quad i = 1, \ldots, m \\
& Ax = b, \qquad i = 1, \ldots, p,
\end{aligned}
$$

affine

where $P_i \in S_+^n$, $\quad i = 0, \ldots, m$.

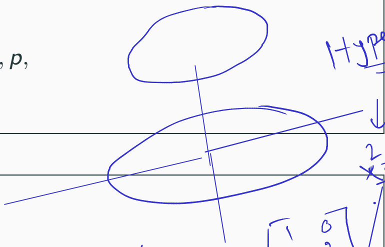
Quadratic and Quadratically Optimization Problems

(LP)

Quadratic Program: (QP)

$$
\begin{aligned}
\text{minimize} \quad & (1/2)x^T P x + q^T x + r \\
\text{subject to} \quad & Gx \preceq h, \\
& Ax = b, \quad i = 1, \ldots, p,
\end{aligned}
$$

where $P \in S^n_+$, $G \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{p \times n}$.



Hyperplane

$x^2 + 2y^2$

$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$

Quadratically Constrained Quadratic Program: (QCQP)

$$
\begin{aligned}
\text{minimize} \quad & (1/2)x^T P_0 x + q^T x + r \\
\text{subject to} \quad & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\
& Ax = b, \qquad i = 1, \ldots, p,
\end{aligned}
$$

where $P_i \in S^n_+$, $\quad i = 0, \ldots, m$.

$\begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}$

positive definite

- In QCQP, we minimize over a region that is intersection of ellipsoids (when $P_i \succ 0$)

Quadratic and Quadratically Optimization Problems

$x^T P x$

$(\sqrt{P}x)^T(\sqrt{P}x)$

$\|\sqrt{P}x\|_2$

Quadratic Program:

$P=0$

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Gx \preceq h, \\ & Ax = b, \quad i = 1, \dots, p, \end{array}$$

where $P \in S_+^n$, $G \in \mathbb{R}^{m\times n}$ and $A \in \mathbb{R}^{p\times n}$.

$x^T x$

affine if $P_0 = 0$

affine if $P_i = 0$ $\forall i = 1, \dots, m$

Quadratically Constrained Quadratic Program:

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q^T x + r \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \le 0, \quad i = 1, \dots, m \\ & Ax = b, \qquad i = 1, \dots, p, \end{array}$$

$\rightarrow$ LP

where $P_i \in S_+^n, \quad i = 0, \dots, m$.

- In QCQP, we minimize over a region that is intersection of ellipsoids (when $P_i > 0$)
- Linear programs are special case with $P_i = 0$ $\forall i = 0, 1, 2 \dots$

Second Order Cone Programming

Second Order Cone Programming

Second Order Cone Programming (SOCP):

affine

cone

$$\begin{aligned} \text{minimize} \quad & f^T x \\ \text{subject to} \quad & \|A_i x + b_i\| \le c_i^T x + d_i,\ i = 1, \ldots, m \\ & Fx = g, \end{aligned}$$

where $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

Second Order Cone Programming

Second Order Cone Programming (SOCP):

$$
\begin{aligned}
\text{minimize} \quad & f^T x \\
\text{subject to} \quad & \|A_i x + b_i\| \leq c_i^T x + d_i, \; i = 1, \ldots, m \\
& Fx = g,
\end{aligned}
$$

where $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

- We call a constraint of the form

$$\|Ax + b\|_2 \leq c^T x + d$$

Second Order Cone Programming

Second Order Cone Programming (SOCP):

$$\begin{aligned} \text{minimize} \quad & f^T x \\ \text{subject to} \quad & \|A_i x + b_i\| \leq c_i^T x + d_i, \; i = 1, \dots, m \\ & Fx = g, \end{aligned}$$

where $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

- We call a constraint of the form

$$\|Ax + b\|_2 \leq c^T x + d$$

a second order cone constraint.

Second Order Cone Programming

Second Order Cone Programming (SOCP):

$$\begin{aligned} \text{minimize} \quad & f^T x \\ \text{subject to} \quad & \|A_i x + b_i\| \leq c_i^T x + d_i, \; i = 1, \ldots, m \\ & Fx = g, \end{aligned}$$

where $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

- We call a constraint of the form

$$\|Ax + b\|_2 \leq c^T x + d$$

  a second order cone constraint.
- When $c_i = 0$ for $i = 1, \ldots, m$

Second Order Cone Programming

Second Order Cone Programming (SOCP):

$$\begin{aligned} \text{minimize} \quad & f^T x \\ \text{subject to} \quad & \|A_i x + b_i\| \leq c_i^T x + d_i, \; i = 1, \ldots, m \\ & Fx = g, \end{aligned}$$

where $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

- We call a constraint of the form

$$\|Ax + b\|_2 \leq c^T x + d$$

  a second order cone constraint.
- When $c_i = 0$ for $i = 1, \ldots, m$ then SOCP is equivalent to QCQP

Second Order Cone Programming

Second Order Cone Programming (SOCP):

$$
\begin{aligned}
\text{minimize} \quad & f^T x \\
\text{subject to} \quad & \|A_i x + b_i\| \leq c_i^T x + d_i, \; i = 1, \ldots, m \\
& Fx = g,
\end{aligned}
$$

where $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

- We call a constraint of the form

$$\|Ax + b\|_2 \leq c^T x + d$$

  a second order cone constraint.
- When $c_i = 0$ for $i = 1, \ldots, m$ then SOCP is equivalent to QCQP
- If $A_i = 0$ for $i = 1, \ldots, m$,

Second Order Cone Programming

Second Order Cone Programming (SOCP):

$$\begin{aligned} \text{minimize} \quad & f^T x \\ \text{subject to} \quad & \|A_i x + b_i\| \leq c_i^T x + d_i, \ i = 1, \dots, m \\ & Fx = g, \end{aligned}$$

where $x \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{n_i \times n}$, and $F \in \mathbb{R}^{p \times n}$.

LP < QP < QCQP < SOCP

- We call a constraint of the form

$$\|Ax + b\|_2 \leq c^T x + d$$

  a second order cone constraint.
- When $c_i = 0$ for $i = 1, \dots, m$ then SOCP is equivalent to QCQP
- If $A_i = 0$ for $i = 1, \dots, m$, then the SOCP reduces to a LP.

Robust Linear Programming

Robust Linear Programming

Robust Linear Program:

$$
\begin{array}{rl}
\text{minimize} & c^T x \\
\text{subject to} & a_i^T x \le b_i, i = 1, \ldots, m,
\end{array}
$$

Robust Linear Programming

Robust Linear Program:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \; i = 1, \dots, m, \end{array}$$

where there are uncertainty in c, a_i, b_i.

Robust Linear Programming

Robust Linear Program:

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \; i = 1, \ldots, m, \end{array}$$

where there are uncertainty in c, a_i, b_i.

- To simplify assume c and b_i are fixed

Robust Linear Programming

Robust Linear Program:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, i = 1, \dots, m, \end{array}$$

where there are uncertainty in c, a_i, b_i.

- To simplify assume c and b_i are fixed
- Assume a_i lies in the given ellipsoids:

Robust Linear Programming

Robust Linear Program:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \le b_i,\ i = 1, \ldots, m, \end{array}$$

where there are uncertainty in c, a_i, b_i.

- To simplify assume c and b_i are fixed
- Assume a_i lies in the given ellipsoids:

$$a_i \in \mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \le 1\},$$

$P_i \in \mathbb{R}^{n \times n}$.

Robust Linear Program:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \le b_i \quad \text{for all } a_i \in \mathcal{E}, \quad i = 1, \ldots, m \end{array}$$

Semidefinite Programming (SDP)

Semidefinite Programming (SDP)

$c^T x + \cancel{d}$

Semidefinite Programming:

$$
\begin{aligned}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n \leq 0 \\
& Ax = b,
\end{aligned}
$$

where $\cancel{G}, F_1, \ldots, F_n \in S^k$, and $A \in \mathbb{R}^{p \times n}$.

Semidefinite Programming (SDP)

$$x_1 \begin{bmatrix} f_1^1 & \\ & f_2^1 \end{bmatrix} + x_2 \begin{bmatrix} f_1^2 & 0 \\ 0 & f_2^2 \end{bmatrix} \le 0$$

$$x_1 f_1^1 + x_2 f_1^2 + g_1 \le 0$$

$$x_1 f_2^1 + x_2 f_2^2 + g_2 \le 0$$

Semidefinite Programming:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n \le 0 \quad +G \\ & Ax = b, \end{aligned}$$

where $G, F_1, \ldots, F_n \in S^k$, and $A \in \mathbb{R}^{p \times n}$.

- If the matrices $\cancel{G}$ G, $F_1, \ldots, F_n$ are diagonals then LMI (Linear Matrix Inequality) reduces to a set of n linear inequalities, and SDP becomes LP

Duality: Introduction

- Transform the (primal) problem into a dual problem!

Duality history...

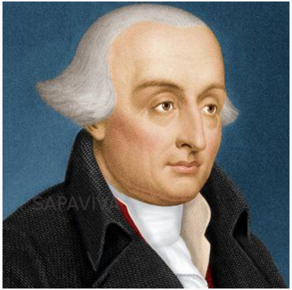



Figure 1: Left: Lagrange, Right: Fenchel

- Lagrange: Introduced Lagrange multiplies for equality constrainted problems
- Fenchel: Introduced conjugates functions

Geometrical Intuition

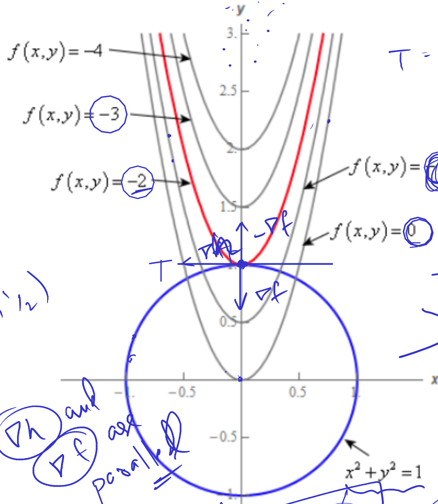
Consider the following problem

$$\begin{aligned} \text{minimize} \quad & f(x,y) = 8x^2 - 2y \\ \text{subject to} \quad & h(x,y) : x^2 + y^2 = 1 \end{aligned}$$

$8x^2 - 2y = c$

$8x^2 - 2y = -1$

$0 - 2 \cdot \frac{1}{2} = -1$



$(0, \frac{1}{2})$

∇h and ∇f are parallel

$T = \begin{bmatrix} T_x \\ T_y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}$

$\nabla f : \begin{bmatrix} 0 \\ y \end{bmatrix}$

$\nabla h = \lambda \nabla f$

$x^2 + y^2 = 1$

cond

Geometrical Intuition

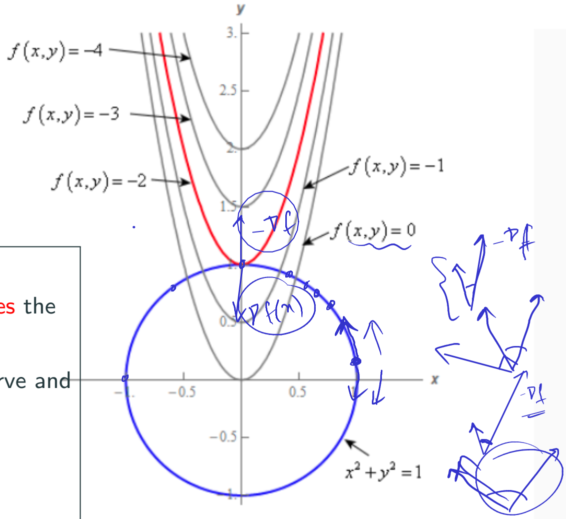
Consider the following problem

$$
\begin{aligned}
\text{minimize} \quad & f(x, y) = 8x^2 - 2y \\
\text{subject to} \quad & h(x, y) : x^2 + y^2 = 1
\end{aligned}
$$

- Draw level curves and contraint set
- Function takes minimum where it just touches the constraint
- At the touch point, normals to constraint curve and level curve are parallel (Why?)
- That is, there exists $\lambda \in \mathbb{R}$, s.t.,

$$\nabla f(x, y) = \lambda \nabla h(x, y),$$

λ is called Lagrange multiplier



Another Example of Lagrange Multipliers...

Consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & h(x, y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

Another Example of Lagrange Multipliers...

Consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & h(x, y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

- The constraint $h(x, y)$ defines a curve

Another Example of Lagrange Multipliers...

Consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & h(x, y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

$$\frac{d}{dx} f(g_1(x), g_2(x), \ldots, g_k(x)) = \frac{\partial f}{\partial g_1}\frac{dg_1}{dx} + \frac{\partial f}{\partial g_2}\frac{dg_2}{dx} + \cdots$$

$$\frac{d}{dx} h(x, y) = \frac{\partial h}{\partial x} \cdot 1 + \frac{\partial h}{\partial y} \cdot \frac{dy}{dx}.$$

- The constraint $h(x, y)$ defines a curve
- Differentiating w.r.t. x

$$\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\frac{dy}{dx} = 0$$

Another Example of Lagrange Multipliers...

Consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x,y) \\ \text{subject to} \quad & h(x,y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

$$\begin{bmatrix} \frac{\partial h}{\partial x}, & \frac{\partial h}{\partial y} \end{bmatrix} \underbrace{\begin{bmatrix} 1 \\ \frac{dy}{dx} \end{bmatrix}}_{} = 0$$

$\Longleftarrow$

must be Tangent vector.

- The constraint $h(x,y)$ defines a curve
- Differentiating w.r.t. x

$$\frac{\partial h}{\partial x} \cdot 1 + \frac{\partial h}{\partial y}\frac{dy}{dx} = 0$$

- We know that gradient of the curve is $\nabla h = (\partial h/\partial x, \partial h/\partial y)$

Another Example of Lagrange Multipliers...

Consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & h(x, y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

- The constraint $h(x, y)$ defines a curve
- Differentiating w.r.t. x

$$\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y} \frac{dy}{dx} = 0$$

- We know that gradient of the curve is $\nabla h = (\partial h / \partial x, \partial h / \partial y)$
- Since tangent is always perpendicular to ∇h: Tangent of the curve is

$$T(x, y) = (1, dy/dx)$$

Another Example of Lagrange Multipliers...

Consider the following minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & h(x, y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

- The constraint $h(x, y)$ defines a curve
- Differentiating w.r.t. x

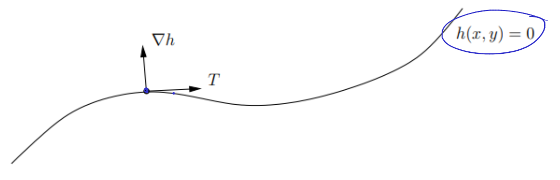
$$\frac{\partial h}{\partial x} + \frac{\partial h}{\partial y}\frac{dy}{dx} = 0$$

- We know that gradient of the curve is $\nabla h = (\partial h/\partial x, \partial h/\partial y)$
- Since tangent is always perpendicular to ∇h: Tangent of the curve is

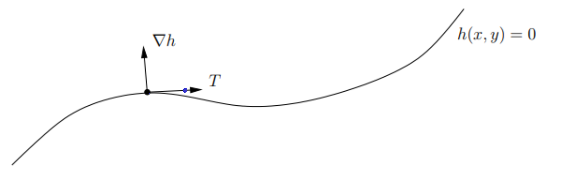
$$T(x, y) = (1, dy/dx)$$

- That is, $T \cdot \nabla h = 0$, i.e., tangent to the curve must be normal to gradient

Another Example for Lagrange's multiplier

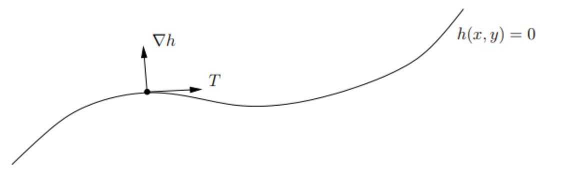


Another Example for Lagrange's multiplier

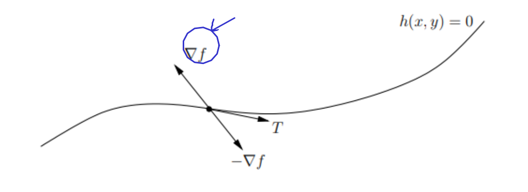


- To stay on the curve, (infinitesimal) motion must be along tangent T

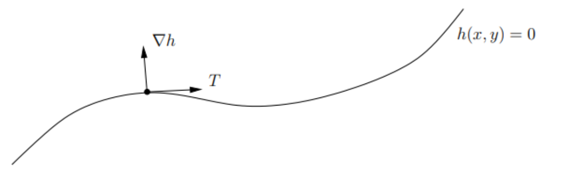
Another Example for Lagrange's multiplier



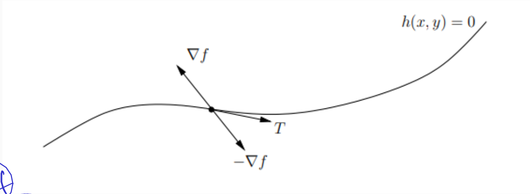
- To stay on the curve, (infinitesimal) motion must be along tangent T



Another Example for Lagrange's multiplier



- To stay on the curve, (infinitesimal) motion must be along tangent T



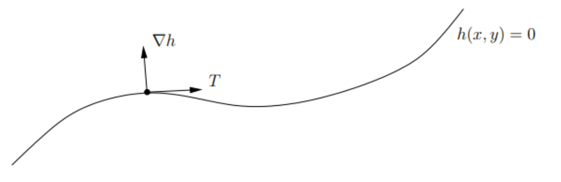
∇f

$-\nabla f$

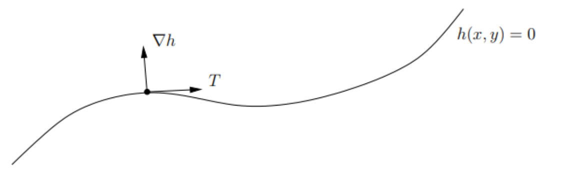
- To be able to increase or decrease $f(x,y)$: motion along constraint curve must have a component along the gradient of f, i.e.,

$$\nabla f \cdot T \neq 0$$

Another Example for Lagrange's multiplier

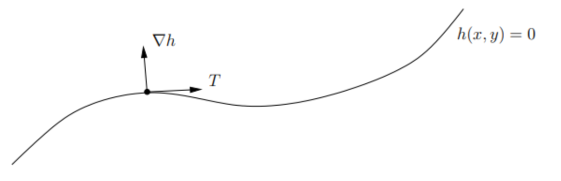


Another Example for Lagrange's multiplier



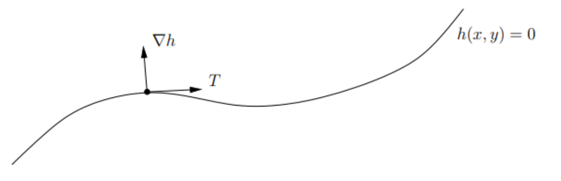
- At extremum of f, an instantenous motion should not yield component of motion along ∇f,

Another Example for Lagrange's multiplier



- At extremum of f, an instantenous motion should not yield component of motion along ∇f, otherwise, the function can still decrease or increase!

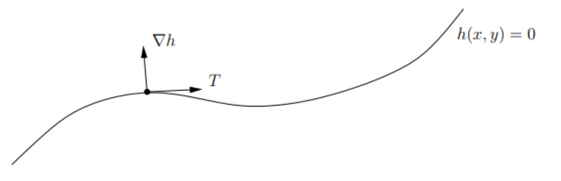
Another Example for Lagrange's multiplier



- At extremum of f, an instantenous motion should not yield component of motion along ∇f, otherwise, the function can still decrease or increase!
- Hence at extremum, we must have

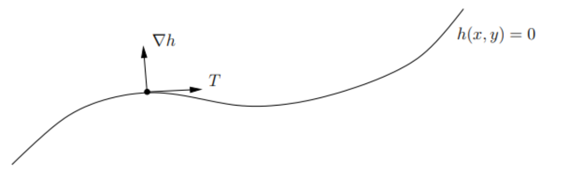
$$\nabla f \cdot T = 0$$

Another Example for Lagrange's multiplier



- At extremum of f, an instantenous motion should not yield component of motion along ∇f, otherwise, the function can still decrease or increase!
- Hence at extremum, we must have
$$\nabla f \cdot T = 0$$
- T is orthogonal to ∇h always and now T is orthogonal to ∇f

Another Example for Lagrange's multiplier



- At extremum of f, an instantenous motion should not yield component of motion along ∇f, otherwise, the function can still decrease or increase!
- Hence at extremum, we must have

$$\nabla f \cdot T = 0$$

- T is orthogonal to ∇h always and now T is orthogonal to ∇f
- That is, ∇h is parallel or anti-parallel to ∇f at extrema
- This means there exists $\lambda \in \mathbb{R}$ such that

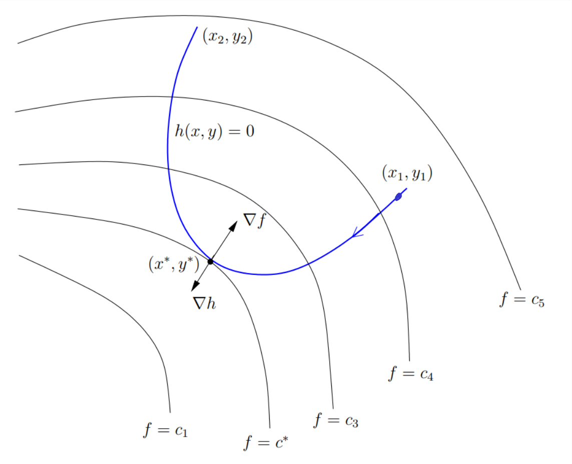
$\nabla f = -\lambda \nabla h$

$$\nabla f + \lambda \nabla h = 0$$

Another Example for Lagrange Multiplier

- Superimpose the curve $h(x, y) = 0$ onto the level curves of $f(x, y) = c$

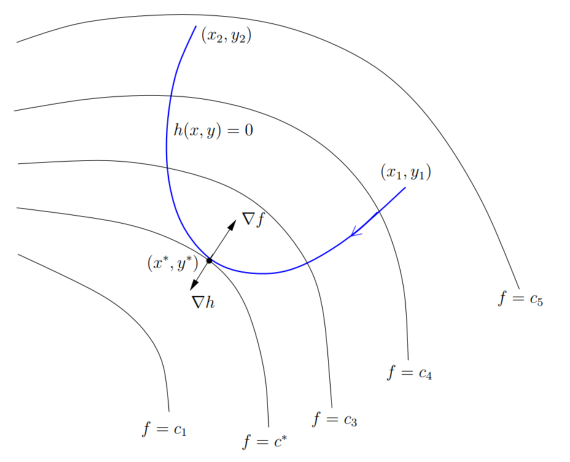
Figure 3: Motion



Another Example for Lagrange Multiplier

- Superimpose the curve $h(x, y) = 0$ onto the level curves of $f(x, y) = c$
- $c_5 > c_4 > \cdots > c_1$

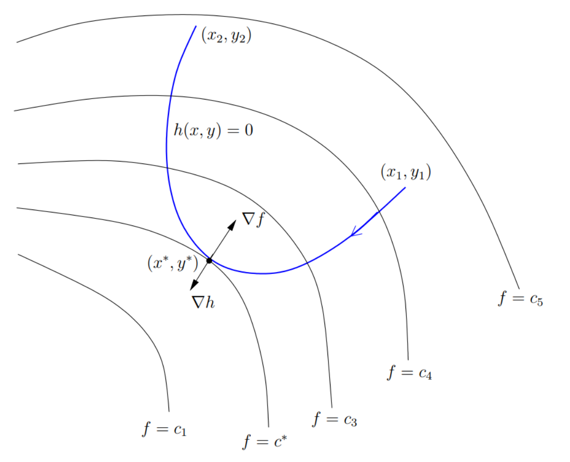
Figure 3: Motion



Another Example for Lagrange Multiplier

- Superimpose the curve $h(x, y) = 0$ onto the level curves of $f(x, y) = c$
- $c_5 > c_4 > \cdots > c_1$
- Tangent of a level curve is always orthogonal to gradient ∇f

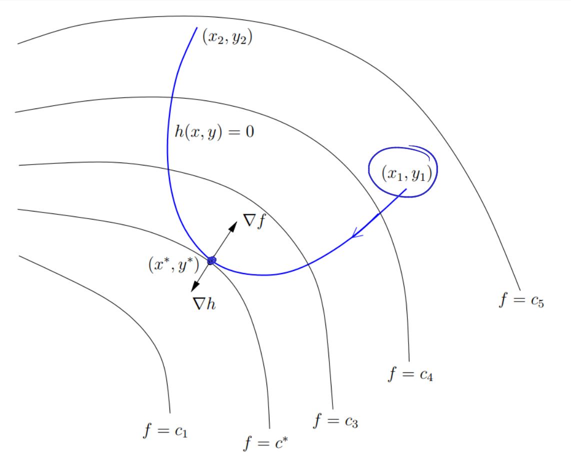
Figure 3: Motion



Another Example for Lagrange Multiplier

- Superimpose the curve $h(x, y) = 0$ onto the level curves of $f(x, y) = c$
- $c_5 > c_4 > \cdots > c_1$
- Tangent of a level curve is always orthogonal to gradient ∇f
- Imagine a point moving along curve $h(x, y)$ from (x_1, y_1) to (x_2, y_2)

Figure 3: Motion



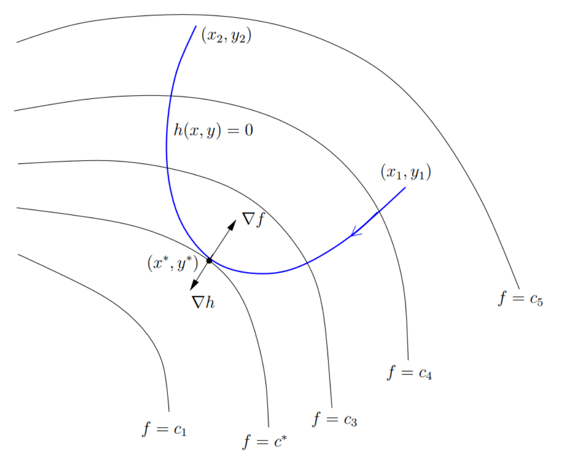
Another Example for Lagrange Multiplier

- Superimpose the curve $h(x, y) = 0$ onto the level curves of $f(x, y) = c$
- $c_5 > c_4 > \cdots > c_1$
- Tangent of a level curve is always orthogonal to gradient ∇f
- Imagine a point moving along curve $h(x, y)$ from (x_1, y_1) to (x_2, y_2)
- Initially, the motion has component along the negative gradient direction

$$-\nabla f$$

resulting in the decrease in the value of f

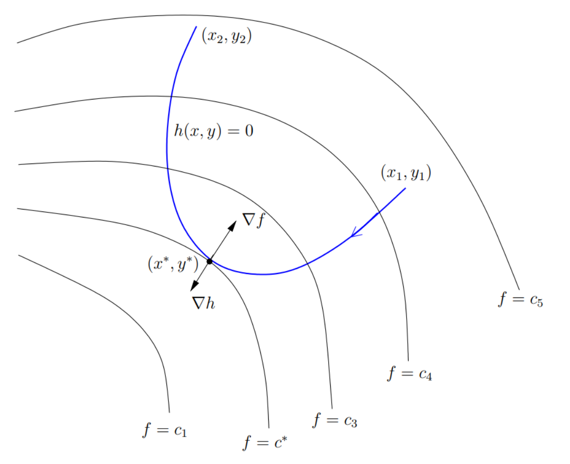
Figure 3: Motion



Another Example for Lagrange Multiplier

Figure 4: Motion

- Component of the motion along $-\nabla f(x)$ gets smaller



Another Example for Lagrange Multiplier

- Component of the motion along $-\nabla f(x)$ gets smaller
- When the motion point reaches (x^*, y^*), the motion is orthogonal to gradient $\nabla f(x)$

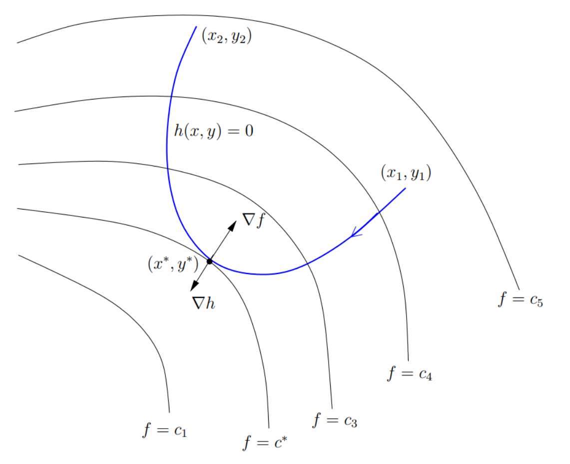


Figure 4: Motion

Another Example for Lagrange Multiplier

- Component of the motion along $-\nabla f(x)$ gets smaller
- When the motion point reaches (x^*, y^*), the motion is orthogonal to gradient $\nabla f(x)$
- After (x^*, y^*), if it continues along $h(x, y)$, the motion has component along ∇f, i.e., the function value increases

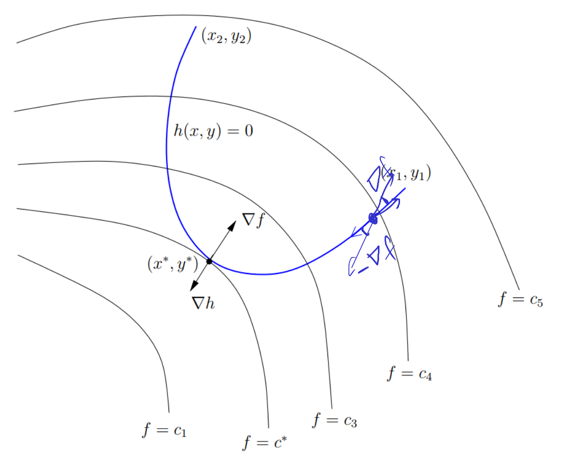
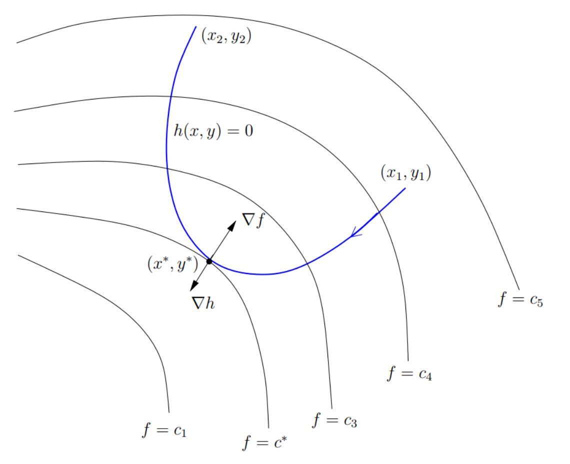


Figure 4: Motion

Another Example for Lagrange Multiplier

- Component of the motion along $-\nabla f(x)$ gets smaller
- When the motion point reaches (x^*, y^*), the motion is orthogonal to gradient $\nabla f(x)$
- After (x^*, y^*), if it continues along $h(x, y)$, the motion has component along ∇f, i.e., the function value increases
- Thus, at (x^*, y^*) function achieves a local minimum, which is c^*

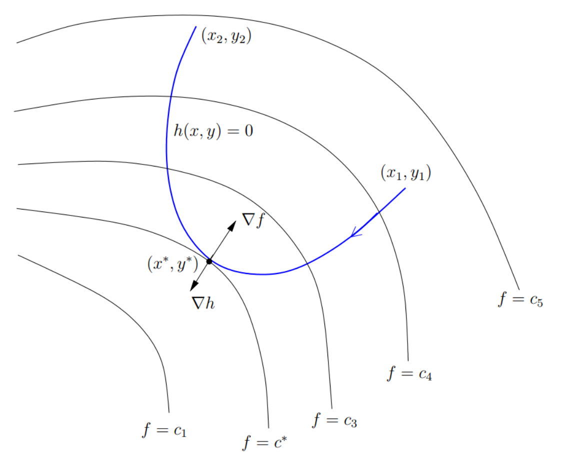
Figure 4: Motion



Another Example for Lagrange Multiplier

- Component of the motion along $-\nabla f(x)$ gets smaller
- When the motion point reaches (x^*, y^*), the motion is orthogonal to gradient $\nabla f(x)$
- After (x^*, y^*), if it continues along $h(x, y)$, the motion has component along ∇f, i.e., the function value increases
- Thus, at (x^*, y^*) function achieves a local minimum, which is c^*
- At (x^*, y^*) the two gradients ∇f and ∇h are opposite

Figure 4: Motion



Motivation for Lagrange Multiplier

The problem of solving the minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x,y) \\ \text{subject to} \quad & h(x,y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

c

now reduces to solving a nonlinear system of equation in variables x, y, λ

$$\begin{aligned} h(x,y) &= 0, \\ \nabla f + \lambda \nabla h &= 0, \quad \text{for some } \lambda. \end{aligned}$$

The system of equations above can be described using so called Lagrangian

$L = f + \lambda h$

$\nabla L = \nabla f + \lambda \nabla h$

$\nabla L = 0$

$\begin{bmatrix} \nabla f + \lambda \nabla h \\ h \end{bmatrix} = 0$

Motivation for Lagrange Multiplier

Given the minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x, y) \\ \text{subject to} \quad & h(x, y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

Motivation for Lagrange Multiplier

Given the minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x,y) \\ \text{subject to} \quad & h(x,y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

Define Lagrangian

$$\ell(x,y,\lambda) = f(x,y) + \lambda h(x,y)$$

Motivation for Lagrange Multiplier

Given the minimization problem

$$\begin{aligned} \text{minimize} \quad & f(x,y) \\ \text{subject to} \quad & h(x,y) = 0, \quad x, y \in \mathbb{R} \end{aligned}$$

Define Lagrangian

$$\ell(x,y,\lambda) = f(x,y) + \lambda h(x,y)$$

We have

$$\nabla \ell = \begin{pmatrix} \partial f/\partial x + \lambda \partial h/\partial x \\ \partial f/\partial y + \lambda \partial h/\partial y \\ h \end{pmatrix} = (\nabla f + \lambda \nabla h, h)$$

Motivation for Lagrange Multiplier

Given the minimization problem

$$
\begin{array}{ll}
\text{minimize} & f(x, y) \\
\text{subject to} & h(x, y) = 0, \quad x, y \in \mathbb{R}
\end{array}
$$

Define Lagrangian

$$\ell(x, y, \lambda) = f(x, y) + \lambda h(x, y)$$

We have

$$\nabla \ell = \begin{pmatrix} \partial f/\partial x + \lambda \partial h/\partial x \\ \partial f/\partial y + \lambda \partial h/\partial y \\ h \end{pmatrix} = (\nabla f + \lambda \nabla h, h)$$

$\partial \ell / \partial x$

$\partial \ell / \partial y$

$\partial \ell / \partial \lambda$

Setting $\nabla \ell = 0$ yields the nonlinear equation

Motivation for Lagrange Multiplier

Given the minimization problem

$$\boxed{\begin{array}{rl} \text{minimize} & f(x,y) \\ \text{subject to} & h(x,y) = 0, \quad x, y \in \mathbb{R} \end{array}}$$

Define Lagrangian

$$\ell(\underline{x, y}, \lambda) = f(x,y) + \lambda h(x,y)$$

We have

$$\nabla h = \begin{pmatrix} \partial f/\partial x + \lambda \partial h/\partial x \\ \partial f/\partial y + \lambda \partial h/\partial y \\ h \end{pmatrix} = (\underbrace{\nabla f + \lambda \nabla h}_{\nabla \ell}, h)$$

Setting $\nabla \ell = 0$ yields the nonlinear equation

$$\begin{cases} h(x,y) = 0, & \\ \nabla f + \lambda \nabla h = 0, & \text{for some } \lambda. \end{cases}$$

Example of nonconvex optimization problem with constraint

Problem. Solve the following.

$$\text{minimize} \quad f(x,y) = xy$$

$$\text{subject to} \quad h(x,y) = \frac{x^2}{8} + \frac{y^2}{2} = 1$$

Solution. The Lagrangian is

$$\ell(x,y,\lambda) = xy + \lambda\left(\frac{x^2}{8} + \frac{x^2}{2} - 1\right) \implies \nabla \ell(x,y,\lambda) = \begin{pmatrix} y + \frac{\lambda x}{4} \\ x + \lambda y \\ \frac{x^2}{8} + \frac{y^2}{2} - 1 \end{pmatrix}^T = 0$$

$f(x,y) + \lambda\,(h(x,y))$

$\partial \ell / \partial x$ $\quad \frac{\partial \ell}{\partial y}$ $\quad \frac{\partial \ell}{\partial \lambda}$

This leads to the system of equations

$$\left.\begin{aligned} y + \frac{\lambda x}{4} &= 0 \\ x + \lambda y &= 0 \\ x^2 + 4y^2 &= 8 \end{aligned}\right\}$$

Example of nonconvex optimization with equality constraint

To solve the system:

$$\begin{aligned} y + \frac{\lambda x}{4} &= 0 \\ x + \lambda y &= 0 \\ x^2 + 4y^2 &= 8 \end{aligned}$$

Example of nonconvex optimization with equality constraint

To solve the system:

$$y + \frac{\lambda x}{4} = 0$$

$$x + \lambda y = 0 \qquad x = -\lambda y$$

$$x^2 + 4y^2 = 8$$

Eliminate x from 1st and 2nd eqution

$$y - \frac{\lambda^2}{4} y = 0$$

Example of nonconvex optimization with equality constraint

To solve the system:

$$y + \frac{\lambda x}{4} = 0$$

$$x + \lambda y = 0$$

$$x^2 + 4y^2 = 8$$

Eliminate x from 1st and 2nd eqution

$$y - \frac{\lambda^2}{4} y = 0$$

Here y cannot be zero, because then x will be zero.

and it violates eq. constr.

Example of nonconvex optimization with equality constraint

To solve the system:

$$y + \frac{\lambda x}{4} = 0$$

$$x + \lambda y = 0$$

$$x^2 + 4y^2 = 8$$

Eliminate x from 1st and 2nd eqution

$$y - \frac{\lambda^2}{4} y = 0$$

Here y cannot be zero, because then x will be zero. Hence $\lambda^2 = 4$, $\lambda = \pm 2$. From 2nd equation, $x = \pm 2y$.

Example of nonconvex optimization with equality constraint

To solve the system:

$$y + \frac{\lambda x}{4} = 0$$
$$x + \lambda y = 0$$
$$x^2 + 4y^2 = 8$$

Eliminate x from 1st and 2nd eqution

$$y - \frac{\lambda^2}{4} y = 0$$

Here y cannot be zero, because then x will be zero. Hence $\lambda^2 = 4$, $\lambda = \pm 2$. From 2nd equation, $x = \pm 2y$. So there are four extremal points of f subject to

Example of nonconvex optimization with equality constraint

To solve the system:

$$\left.\begin{aligned} y + \frac{\lambda x}{4} &= 0 \\ x + \lambda y &= 0 \\ x^2 + 4y^2 &= 8 \end{aligned}\right\}$$

Eliminate x from 1st and 2nd eqution

$$y - \frac{\lambda^2}{4} y = 0$$

Here y cannot be zero, because then x will be zero. Hence $\lambda^2 = 4$, $\lambda = \pm 2$. From 2nd equation, $x = \pm 2y$. So there are four extremal points of f subject to

$$h : (2, 1), (-2, -1), (2, -1), \text{ and } (-2, -1)$$

Example of nonconvex optimization with equality constraint

To solve the system:

$$y + \frac{\lambda x}{4} = 0$$

$$x + \lambda y = 0$$

$$x^2 + 4y^2 = 8$$

Eliminate x from 1st and 2nd eqution

$$y - \frac{\lambda^2}{4} y = 0$$

Here y cannot be zero, because then x will be zero. Hence $\lambda^2 = 4$, $\lambda = \pm 2$. From 2nd equation, $x = \pm 2y$. So there are four extremal points of f subject to

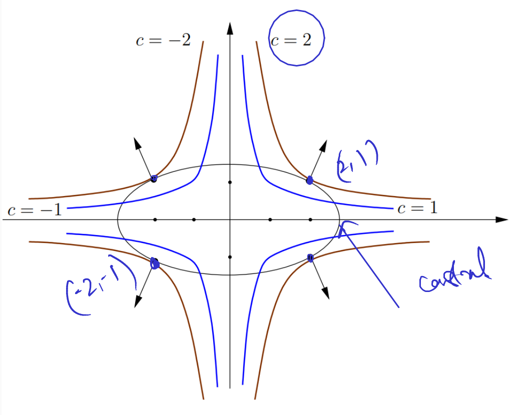
$$h : (2, 1), (-2, -1), (2, -1), \text{ and } (-2, +1)$$

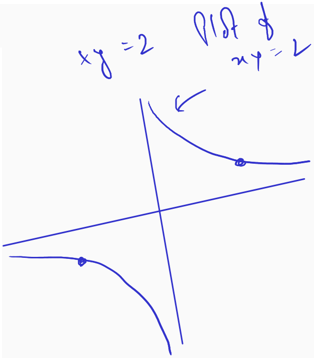
2 2

- Max value is achieved at $(2, 1), (-2, -1)$
- Min value is achieved at $(2, -1), (-2, +1)$

-2 -2

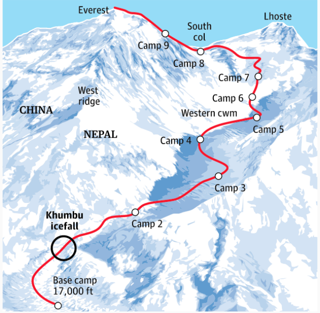
Example of nonconvex optimization with equality constraint



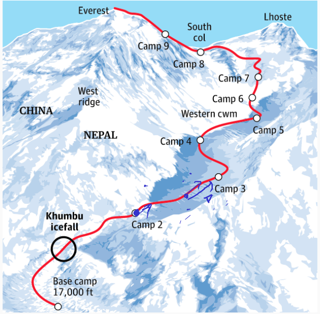


- constraint h defines an ellipse
- level contours of f are hyperbolas $xy = c$

Summary

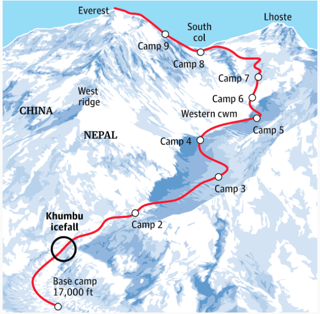

Everest
South col
Lhoste
Camp 9
Camp 8
Camp 7
Camp 6
West ridge
CHINA
Western cwm
Camp 5
NEPAL
Camp 4
Camp 3
Khumbu icefall
Camp 2
Base camp 17,000 ft

Summary



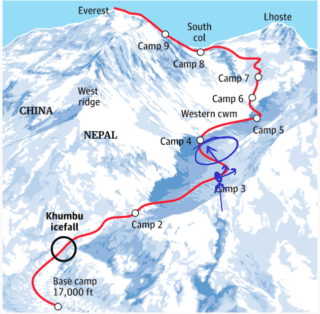
- Keep climbing if the gradient has non-zero component along the climb direction (climb direction (equality constraint) is along red path)!

Summary



- Keep climbing if the gradient has non-zero component along the climb direction (climb direction (equality constraint) is along red path)!
- Can get stuck in local maxima! Because at local maxima, gradient=0 $\implies$ graient has no component along climb, and you stop, even though maxima was further ahead on climb path! Unfortunate case!

Summary

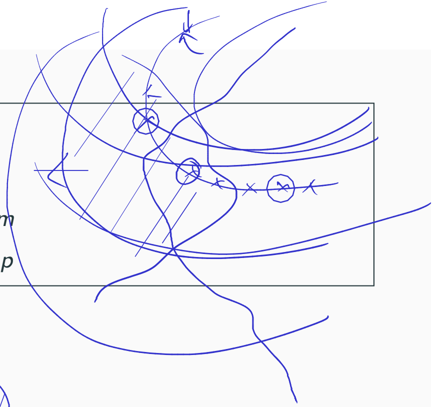


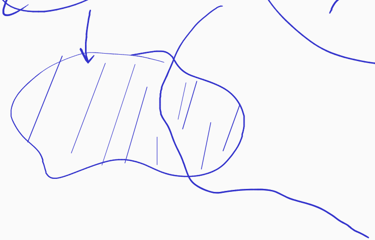
- Keep climbing if the gradient has non-zero component along the climb direction (climb direction (equality constraint) is along red path)!
- Can get stuck in local maxima! Because at local maxima, gradient=0 $\implies$ graient has no component along climb, and you stop, even though maxima was further ahead on climb path! Unfortunate case!
- As humans, we have much wider view than local gradient view! Hence, we dont get stuck at local maxima.

Lagrangian Dual Function

Optimization problem (Standard Form):

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\
& h_i(x) = 0, i = 1, \ldots, p
\end{aligned}
$$





Lagrangian Dual Function

Optimization problem (Standard Form):

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, i = 1, \ldots, m \\
& h_i(x) = 0, i = 1, \ldots, p
\end{aligned}
$$

Idea: Augment the objective $f_0(x)$ with a weighted sum of the constraint functions.

Lagrangian Dual Function

Duality

Primal $\rightarrow$ dual

$\boxed{x}$ $\quad$ $\boxed{\lambda_i, \nu_i}$

eliminate x

$\boxed{\nu}$

Dual problem

Optimization problem (Standard Form):

$$
\begin{array}{rl}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m \\
 & h_i(x) = 0, i = 1, \ldots, p
\end{array}
$$

Idea: Augment the objective $f_0(x)$ with a weighted sum of the constraint functions.

Lagrangian: Define Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

with dom $L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$. Here λ_i, ν_i are called Lagrange multipliers. Here λ and ν are called dual variables or Lagrange multiplier vectors.